Generation of a *de novo* intronic junction in the *DMD* gene through CRISPR/Cas genome editing as a potential therapy for a high proportion of Duchenne muscular dystrophy patients.

THESIS SUBMITTED FOR THE DEGREE OF DOCTOR OF PHILOSOPHY

IN THE ROYAL HOLLOWAY UNIVERSITY OF LONDON.

By M. Rebeca Gil Garzón.

Student No: 100908617.

School of Biological Sciences.

Supervisors: Dr. Linda Popplewell & Dr. Alberto Malerba.

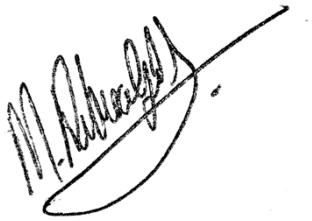
Date: May 2023.

Generation of a *de novo* intronic junction in the *DMD* gene through CRISPR/Cas genome editing as a potential therapy for a high proportion of Duchenne muscular dystrophy patients.

TABLE OF CONTENTS.

DECLARATION OF AUTHORSHIP.

I Monica Rebeca Gil Garzón hereby declare that this thesis and the work presented in it is entirely my own. Where I have consulted the work of others, this is always clearly stated.



Signed: M. Rebeca Gil Garzón.

Date: 30th May 2023.

ABSTRACT.

Duchenne Muscular Dystrophy (DMD) is caused by mutations across the *DMD* gene. The subsequent absence of dystrophin protein compromises muscle stability and contractility and gives rise to progressive muscle wasting. Different gene therapies are being investigated, such as AAV micro-dystrophin delivery, premature termination codon read-through, exon-skipping and utrophin upregulation. Nevertheless, these therapies would require repeated administration, could carry an adverse immunological risk and some are restricted by mutation specificity. Such problems may be circumvented with genome editing.

The aim of this project is to create a *de novo* junction between introns 18 and 55, using a CRISPR/Cas system, to express a truncated functional dystrophin from the endogenous *DMD* locus. The gRNAs designed to target intron 18 and 55 would produce a near 800 kbp deletion. It is estimated that this strategy would eliminate approximately 81% of total DMD mutations.

This novel approach would produce a new truncated dystrophin. To assess potential functionality, a cDNA construct expressing Del19-55 dystrophin was generated and tested *in-vitro* and *in-vivo* in *mdx* mice. Results indicate that Del19-55 dystrophin has potential functionality and could have beneficial effects when expressed in sufficient levels.

To produce the deletion of exons 19 to 55 that would drive expression of Del19-55 dystrophin, gRNAs for *Staphylococcus aureus* (*Sa*)Cas9 were designed targeting introns 18 and 55 of the *DMD* gene to human and mouse sequences. Individual gRNA efficiency to induce site-specific cleavage was analysed *in-vitro* by TIDE analysis. The optimal gRNAs for each intronic site in murine *Dmd* were multiplexed into an AAV9-*Sa*Cas9 construct. Multiplex construct and co-delivery of top candidate gRNAs were assessed *in-vitro* by transfection of N2A cells and nucleofection of C2C12 cells. Deletion of exons 19 to 55 was confirmed at DNA level by end-point PCRs and sequencing on both cell lines.

Multiplex construct and co-delivery of gRNAs, alongside the positive control (plasmid expressing Del19-55 dystrophin), were assessed *in-vivo* by plasmid injections. The positive control plasmid significantly increased dystrophin positive fibres. However, no significant difference was observed from the other groups. To increase delivery efficiency, constructs were packaged into AAV9 vectors. 2-months old *mdx* mice were treated with our multiplex gRNAS and co-transduced with individual gRNAs. No beneficial effects were observed on muscle physiology analysis and it was not possible to detect a deletion from treated samples.

This study shows the development of a universal genome editing strategy from theory to *in-vivo* proof of concept. From gRNA design, *in-vitro* screening, development and

assessment of an *Sa*Cas9 multiplex system *in-vitro*, to *in-vivo* assessment by AAV9 delivery in a single vector. This thesis explores limitations of achieving a large deletion *in-vivo* and highlights potential functionality of a new truncated dystrophin.

ACKNOWLEDGEMENTS.

I want to thank my family for their unconditional support. Mum, dad and David, I wouldn't be here today without your help, encouragement and words of wisdom.

I want to thank my supervisors Dr. Linda Popplewell and Dr. Alberto Malerba for their guidance and support.

Linda, thank you for accepting me as your student and for the opportunities you allowed me to have, for the constant support throughout 4 years and all your help.

Alberto, thank you for all your help and encouragement, particularly in the last couple of years.

I want to thank my lab colleagues: Dr. James March, Dr. Ngoc Lu-Nguyen, Dr. Jamuna Selvakumaran, Dr Marc Moore, Jess and my PhD twin Arjun, for their help and good times while trying to do science and for always sharing their scientific knowledge with me.

Thank you to my partner Dominic and my friends for keeping me sane these last four years. To my friends in the UK, Chiara, Leoni, Julia & Aisha, for their company and to my friends back in Mexico, Ana, Ivana, Ceci, Ale, Adri & Mariela, for being so present despite the distance.

Life is not about the destination, but about enjoying the road. Thanks to my family, friends & colleagues for accompanying me in my road and making it such a happy one.

ABBREVIATIONS.

6MWT – 6 Minute walk test

AAV – Adeno-associated virus

AO - Antisense oligonucleotide

BMD – Becker muscular dystrophy

cDNA – complementary/coding DNA

CMV – Cytomegalovirus

CRISPR – Clustered regularly interspaced palindromic repeats

crRNA – CRISPR RNA

DAPC – Dystrophin associated protein complex

DNA – Deoxyribonucleic acid

DMD – Duchenne muscular dystrophy

DMEM – Dulbecco's modified Eagle´s medium

DSB - Double stranded break

ECM – Extracellular matrix

EMA – European Medicines Agency

FACS – Fluorescent activated cell sorting

FCS – Foetal calf serum

FDA – U.S. Food and Drug Administration

GFP – Green fluorescent protein

gRNA – guide RNA for CRISPR

HDR – Homology directed repair

HEK293T – Human embryonic kidney cell line containing Sv40 large T antigen

hDys – human native dystrophin cDNA

Indels – small insertions or deletions

MD – Muscular dystrophy

MD1 – micro-dystrophin 1

Mdx – X-chromosome linked muscular dystrophy in mouse model

mRNA – messenger RNA

N2A – Mouse Albino Neuroblastoma cells

NHEJ – Non-homologous end joining

nNOS – Neuronal nitric oxide synthase

PAM – Protospacer adjacent motif

PBS – Phosphate buffered saline

PCR – Polymerase chain reaction

PMO - Phsorphorodiamidate morpholino oligomer

qPCR – quantitative PCR

RT-qPCR – Reverse transcription quantitative PCR

RNA – Ribonucleic acid

Spc512 – Synthetic muscle specific promoter

*Sa*Cas9 – *Staphylococcus aureus* Cas9

*Sp*Cas9 – *Streptococcus pyogenes* Cas9

TA – Tibialis anterior muscle

TALENs – Transcriptional activator like effector nucleases

TIDE - Tracking of indels by decomposition

tracrRNA – Trans-activating CRISPR RNA

ZFNs – Zinc finger nucleases

1. INTRODUCTION.

1.1. GENE THERAPIES OVERVIEW.

The definition of gene therapies by scientists and regulatory bodies has evolved alongside developments in the field. In the 1970s, scientists referred to gene therapy as techniques requiring exogenous DNA to replace defective DNA in those with genetic diseases (Friedmann & Roblin, 1972). More than twenty years later, the U.S. Food and Drug Administration (FDA) finalised its position on the definition of gene therapies to those used to "modify or manipulate the expression of genetic material to alter the biological properties of living cells" (FDA, 1993). In 2018, FDA updated this definition by adding that gene therapies are "products that mediate their effects by transcription or translation of transferred genetic material or by specifically altering host (human) genetic sequences" (FDA, 2018). While the European Medicines Agency (EMA) latest update refers to gene therapy medicines as those that "contain genes that lead to a therapeutic, prophylactic or diagnostic effect. They work by inserting 'recombinant' genes into the body, usually to treat a variety of diseases, including genetic disorders, cancer or long-term diseases" and considers them a type of "Advanced therapy medicinal products" (ATMPs) alongside somatic-cell therapy medicines and tissue-engineered medicines (EMA, 2018).

An updated definition was proposed by Sherkow et al., (2018) referring to gene therapy as “the intentional, expected permanent, and specific alteration of the DNA sequence of the cellular genome, for a clinical purpose”, not necessarily being irreversible, just expected to persist in the cell during its life. Under this definition, gene therapy could be split into three main categories: direct (intentional modification of a somatic gene to correct a defect or fix alleles malign function), compensatory (induction of expression of genetic material to compensate for cellular malfunction) and augmenting (introduction of a function that is not present in the target cells). These categories would apply to *in-situ* and *ex-vivo* therapies, regardless of the delivery vector (Sherkow et al., 2018).

Remarkably, even though the term gene therapy was first defined in the 1970´s, it was not until 2003 that the first gene therapy was approved and since then, at least 40 products have been approved by the FDA, EMA and SFDA (Arabi et al., 2022) (approved products summarised in Table 1.1). Furthermore, between 2012 and 2020 there were 1,907 active records registered on Clinicaltrials.gov involving gene therapies (Arabi et al., 2022). The surge of approved gene therapy products in recent years, alongside the active clinical trials currently underway, signifies an exceptionally thrilling period for research and development in the field, conceding unprecedented opportunities for advancements.

Table 1.1. Gene therapy products approved (until 2022) by the EMA, FDA and/or SFDA. Adapted from (Arabi et al., 2022).

Year of approval	Trade name	Details/indication	Agency/Country	Reference
2003	Gendicine	AAV gene therapy product for head and neck cancer	SFDA	(Pearson et al., 2004) (Guo & Song, 2018)
2005	Oncorine	First oncolytic virus to treat nasopharyngeal carcinoma	SFDA	(Liang, 2018)
2012	Glybera	Gene therapy for familial lipoprotein lipase deficiency (withdrawn from market in 2017)	EMA	(Gruber, 2012)
2014	Translarna (Ataluren)	Small molecule for Duchenne muscular dystrophy (conditional approval)	EMA	(Haas et al., 2015)
2015	Imlygic	Gene therapy for melanoma	EMA, FDA, UK, Australia	(Poh, 2016), (Ott & Hodi, 2016)
2016	Strimvelis	First *ex-vivo* gene therapy product indicated for treatment of severe combined immunodeficiencies due to adenosine deaminase deficiency (ADA-SCID)	EMA, UK	(Schimmer & Breazzano, 2016), (Aiuti et al., 2017)
2016	Exondys 51 (Eteplirsen)	Antisense oligonucleotide for Duchenne muscular dystrophy	FDA	(FDA, 2016), (Aartsma-Rus & Goemans, 2019)
2016	Spinraza	Antisense oligonucleotide for spinal muscular atrophy	EMA, FDA, UK, Canada, Japan, Brazil, Switzerland, Australia, South Korea, SFDA,	(Ottesen, 2017)

			Argentina, Colombia, Taiwan, Turkey	
2017	Kymriah	Chimeric antigen receptor (CAR)-T cells to treat oncological diseases	EMA, FDA, UK, Japan, Australia, Canada, South Korea	(Seimetz et al., 2019)
2017	Yescarta	(CAR)-T cells to treat oncological diseases	EMA	(Seimetz et al., 2019)
2017	Luxturna	Adeno-associated virus (AAV) gene therapy product, indicated for Leber congenital amaurosis	EMA, FDA, UK, Australia, Canada, South Korea	(Padhy et al., 2020)
2018	Tegsedi	Antisense oligonucleotide for hereditary transthyretin-related amyloidosis	EMA, FDA, UK, Canada, Brazil	(Gales, 2019)
2018	Onpattro	siRNA for hereditary transthyretin-related amyloidosis	EMA, FDA, UK, Japan, Canada, Switzerland, Brazil, Taiwan, Israel, Turkey,	(Maurer et al., 2018)
2019	Zolgensma	The most expensive drug to date, an AAV vector indicated for paediatric spinal muscular atrophy	EMA, FDA, UK, Japan, Australia, Canada, Brazil, Israel, Taiwan, South Korea,	(Mahajan, 2019)
2019	Vyondys 53 (Golodirsen)	Antisense oligonucleotide for Duchenne muscular dystrophy	FDA	(Heo, 2020)
2019	Waylivra	Antisense oligonucleotide for adult familial chylomicronaemia syndrome	EMA, UK, Brazil	(Paik & Duggan, 2019)
2020	Tecartus	Ex-vivo gene therapy for relapse/refractory mantle cell lymphoma	EMA, FDA, UK	(FDA, 2021), (EMA, 2020)

2020	Libmeldy	Ex-vivo gene therapy for metachromatic leukodystrophy	EMA, UK	(EMA, 2020)
2020	Givlaari (givosiran)	siRNA for porphyria	EMA, FDA, UK, Canda, Switzerland, Brail, Israel, Japan	(Scott, 2020)
2020	Oxlumo (lumasiran)	siRNA for primary hyperoxaluria	EMA, FDA, UK, Brazil	(Scott & Keam, 2021)
2020	Viltepso (viltolarsen)	Antisense oligonucleotide for Duchenne muscular dystrophy	FDA, Japan	(FDA, 2020) , (Roshmi & Yokota, 2021)
2020	Leqvio	Antisense oligonucleotide for primary hypercholesterolemia	EMA, FDA	(Migliorati et al., 2022)
2020	Comirnaty	mRNA COVID-19 vaccines by BioNTech and Pfizer	EMA, FDA, Bahrain, Israel, Canada, Rwanda, Serbia, United Arab Emirates, Macao, Mexico, Kuwait, Singapore, Saudi Arabia, Chile, Switzerland, Colombia, Philippines, Australia, Hong Kong, Peru, South Korea, New Zealand, Japan, Brazil, Sri Lanka, Vietnam, South Africa, Thailand, Oman Egypt, Malaysia	(Lamb, 2021)

2020	Spikevax	mRNA COVID-19 vaccine by Moderna	EMA, FDA, Canada, Israel, Switzerland, Singapore, Qatar, Vietnam, UK, Philippines, Thailand, Japan, South Korea, Brunei, Paraguay, Taiwan, Botswana, India, Indonesia, Saudi Arabia, Mexico, Australia, Nigeria, Colombia	(EMA, 2021), (FDA, 2023)
2021	Breyanzi (lisocabtagene maraleucel)	Gene therapy with retroviral vector for relapsed or refractory diffuse large B cell lymphoma, follicular lymphoma and multiple myeloma	FDA	(FDA, 2021)
2021	Abecma (idecabtagene vicleuel)	Gene therapy with lentiviral vector for multiple myeloma	EMA, FDA	(EMA, 2021), (FDA, 2021)
2021	Amondys 45 (Casimersen/srp-4045)	Antisense oligonucleotide for Duchenne muscular dystrophy (Exon 45)	FDA	(FDA, 2021)
2022	Carvykti (ciltacabtagene autoleucel)	Gene therapy for relapsed or refractory multiple myeloma	FDA	(FDA, 2023)

1.1.1. Genome editing & the discovery of CRISPR/Cas9 and predecessors.

Genome editing technologies allow for permanent, highly specific, targeted modifications to the genome, which can be used to correct defective genes by precise removal or correction of a mutation or by insertion of a therapeutic gene (Cox et al., 2015).

Genome editing strategies allow to treat diseases that are refractory for gene addition and gene silencing therapies (Fridovich-Keil, Judith L., 2019). Gene addition can be achieved by delivering exogenous DNA that works as a transcriptional template for the expression of a protein (Kay, 2011), allowing for treatment of autosomal recessive or haploinsufficiency disorders that lead to a loss-of function. In contrast, genome editing allows correction of mutations in genomic DNA, hence it could be applicable to loss- and gain-of function mutations that lead to genetic disorders (Raguram et al., 2022). Gene knockdown or silencing, by RNA interference with silencing RNAs, micro RNAs or anti-sense oligonucleotides (synthetic single stranded molecules of nucleic acid), leads to transient silencing of the gene of interest (Kher et al., 2011, Lam et al., 2015). In contrast, if efficient enough, genome editing strategies would make a permanent correction and likely avoid the need of repeated administration, required by transient RNA modulation and gene augmentation strategies (Raguram et al., 2022).

Furthermore, targeted genome editing has become more efficient and easier to undertake thanks to the ability to generate artificial DNA endonucleases that induce a specific double-strand break in a sequence of choice (Naldini, 2015).

Genome editing started with the discovery of bacterial restriction enzymes (Meselson & Yuan, 1968, Smith & Welcox, 1970) and their ability to produce a double-strand break, that would then be repaired by homology-directed repair (HDR) or by non-homologous end joining (NHEJ) (Rouet et al., 1994). Then, studies showed that mammalian cells predominantly repair DNA double-strand breaks (DSBs) by NHEJ mechanisms that lead to insertion and deletion of nucleotides at the break site (Phillips & Morgan, 1994). It was shown that NHEJ or HDR mechanisms predominance is influenced by the stage of the cell cycle in which the DBS was induced (Moore & Haber, 1996) and could be independently modulated (Fishman-Lobell et al., 1992). It was also shown that the homologous integration via homologous direct repair in mammalian cells could be enhanced by having a double-strand break at a position of homology between transfected DNA (donor DNA) and the genomic target (Jasin & Berg, 1988).

This led to the first attempts of targeted *in-vivo* DNA manipulation. In the 1980´s, Mario Capecchi, Martin Evans and Oliver Smithies were independently researching the principles for introducing specific gene modifications in mice by the use of embryonic stem cells and transgenesis (Watts, 2007). The independent research of Capecchi (Folger et al., 1985; Frels et al., 1985; Capecchi, 1989), Smithies (Doetschman et al.,

1987; Koller et al., 1989) and Evans (Kuehn et al., 1987; Ratcliff et al., 1993) on targeted editing by homologous recombination, led to the development of what was then named "knockout mice" and to the award of the 2007 Nobel prize in Physiology or Medicine. They eventually showed that new genes could be eliminated and incorporated to the mouse genome (Watts, 2007).

In the course of time, engineered genome-editing meganucleases such as Zinc finger nucleases (ZFNs), transcription activator-like effector nucleases (TALENs) and Clustered regularly interspaced short palindromic repeats (CRISPR) became the most popular tools for genome engineering.

ZFNs were created by linking two different zinc finger proteins to the cleavage domain of *Fok*I endonuclease. The modular structure of *Fok*I endonuclease made it possible to construct chimeric restriction enzymes by linking other DNA-binding proteins to its cleavage domain. The modular structure of zinc finger proteins allows one to select the peptides that will bind to a specific site in the DNA under the right conditions, allowing the generation of artificial nucleases with tailor-made sequence specificities (Kim et al., 1996).

Repeat regions Transcription-activator-like effectors (TALE), originally found in plant pathogenic *Xanthomona* acting as transcriptional activators in the plant cell nucleus,

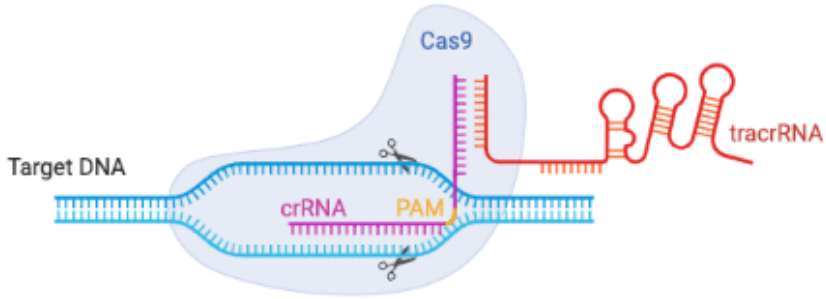
have a sequential nature that corresponds to a consecutive target DNA sequence, and can be used to construct artificial effectors with new specificities by binding them to tandem repeat domains (Boch et al., 2009; Zhang et al., 2011) that can then be fused to other proteins, like the restriction enzyme *Fok*I endonuclease domain (Li et al., 2011), creating TALENs.

CRISPR sequences were originally discovered in *E. Coli* genome as a safeguard against bacteriophages (Ishino et al., 1987). Since then CRISPR sequences have been identified in other species such as *Mycobacterium* tuberculosis strains, members of the Archaea family (Mojica et al., 1995), filamentous cyanobacteria (Masepohl et al., 1996) and *Streptococcus* strains (Hoe et al., 1999). Soon enough, CRISPR loci were recognised as a family of repeats in genomes of Archaea, Bacteria and mitochondria (Mojica et al., 2000). In nature, specialized CRISPR associated (Cas) proteins snip foreign viral DNA into small fragments of 20 bp and paste them into what is known as CRISPR arrays (Jansen et al., 2002). Different Cas proteins express and process the CRISPR loci to generate the crRNA (CRISPR RNA). Then, through sequence homology, the crRNA guides a Cas nuclease to the specific exogenous genetic material previously “saved” that contains a specific sequence called protospacer adjacent motif or PAM sequence (Mojica et al., 2009). This is how the CRISPR complex recognizes and binds to foreign DNA to destroy it. A couple of years later it was demonstrated that crRNA fused to trans-activating RNA (tracrRNA), crRNA:tracrRNA, was sufficient to direct the Cas9 protein to cleave to the target DNA sequence matching the crRNA-guide sequence, also referred to as single

guide RNA (sgRNA). After PAM sequence recognition, the Cas9 protein unwinds DNA and allows the Cas9-sgRNA complex to hybridize with the exposed DNA strain, if the DNA sequence matches the sgRNA target sequence, HNH and RuvC catalytic domains from the Cas cleave both strands of target DNA and generate a DSB (Jinek et al., 2012). Furthermore, it was shown that this system was programmable by changing the DNA target-binding sequence in the sgRNA, meaning that CRISPR could be programmed to introduce site-specific DSBs in target DNA (Jinek et al., 2012). Soon after, it was demonstrated that this system could be applied in eukaryotic DNA (Jinek et al., 2013). Native CRISPR/Cas9 system programmed by crRNA:tracrRNA duplex and CRISPR/Cas9 system programmed by a single chimeric guide RNA can be compared on Figure 1.1.

An illustrative summary of programable nucleases (ZFNs, TALENs and CRISPR/Cas9) commonly used for genome editing and the DNA repair pathways after a double-strand break can be found on Fig. 1.2.

A) Cas9 programmed by crRNA:tracrRNA duplex



B) Cas9 programmed by a single chimeric guide RNA

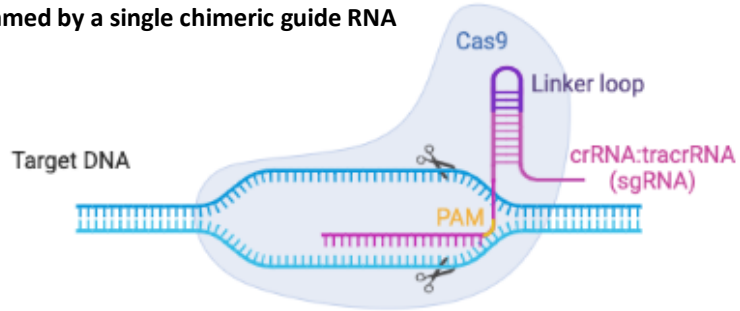


Figure 1.1. Comparison of type II CRISPR/Cas9 native system programmed by crRNA:tracrRNA duplex and CRISPR/Cas9 system programmed by a single chimeric guide RNA. A) CRISPR/Cas9 system programmed by a two-RNA structure formed by a targeting crRNA and an activating tracrRNA to cleave a specific site at target dsDNA. B) CRISPR/Cas9 system guided by a single chimeric guide RNA generated by fusing 3´end of crRNA to the 5´end of the tracrRNA with a linker loop. In both systems the CRISPR/Cas9 complex binds to the PAM site adjacent to the crRNA sequence. The Cas9 makes a double strand break 3 bp upstream of the PAM site. Adapted from (Jinek et al., 2012). Created with BioRender.com.

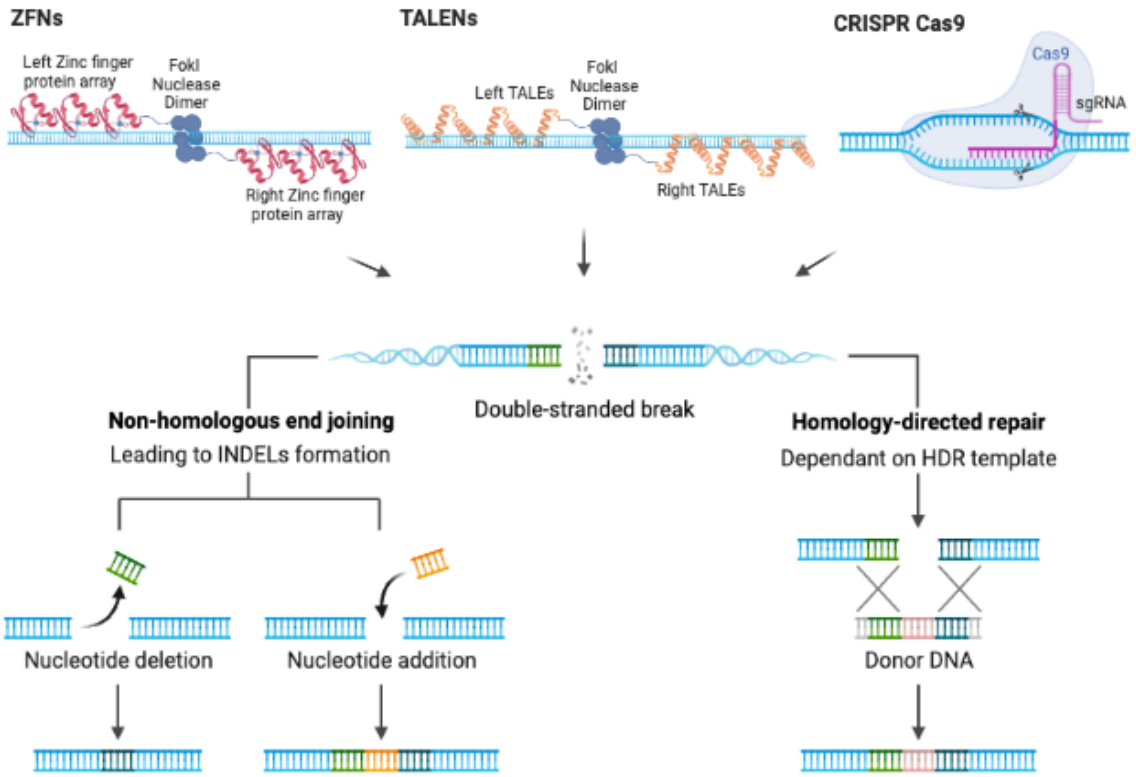


Figure 1.2. Illustrative summary of programmable nucleases commonly used for genome editing and DNA natural repair pathways after double-strand breaks (DSB). Zinc finger nucleases (ZFNs), transcription activator-like effector nucleases (TALENs) and CRISPR Cas9 systems (with a fused tracrRNA:crRNA = sgRNA in pink) create double strand cuts at specific locations in the genome. DNA DSBs are repaired by non-homologous end joining, a mechanism that leads to small deletions and insertions (INDELs) or by homology-direct repair (HDR), a mechanism that requires a DNA template. For genome editing applications an exogenous HDR DNA template can be supplied to produce a change or insertion at the target site. Created with BioRender.com.

The mechanisms and outcomes from the two major repair pathways are now better understood. NHEJ is initiated after a DSB by the binding of Ku70-Ku80 to the blunt DNA ends (Fell & Schild-Poulter, 2015). Then the KU heterodimer (Ku70-Ku80) recruits other factors to the DSB site, such as DNA-PKcs that form an active DNA-PKcs-KU Complex (Kragelund et al., 2016). DNA-PKcs then phosphorylates Artemis, X-ray repair cross complementing protein 4 (XRCC4), DNA ligase IV and XRCC4-like factor (XLF) (Davis et al., 2014). This promotes synapsis of DNA ends, if compatible, termini can be directly ligated, if the DNA termini are incompatible, the exonuclease Artemis binds with polymerases lambda and mu to prepare blunt ends for ligation. XRCC4-DNA ligase IV-XLF complex performs the ligation (Conlin et al., 2017, Stinson et al., 2020) and DNA bases are added randomly by the DNA polymerases or removed by the nucleases, leading to smalls indels (Yang et al., 2020).

In contrast to NHEJ, the HDR process occurs largely during the S/G2 phase when an undamaged sister chromatid or donor DNA is available and resection of DNA forms a 3´ single-stranded DNA overhang (Symington, 2016). HDR is initiated by the MRN complex (MRE11-RAD50-NBS1) (Symington & Gautier, 2011). This complex recruits C-terminal-binding interacting protein (CtIP) and initiates resection, generating short single-stranded tails (Huertas & Jackson, 2009). Then, exonuclease 1 (Exo1) and the DNA replication ATP-dependent helicase/nuclease DNA2/bloom syndrome protein (BLM) complex perform long-range DNA resection leading to a 3´ ssDNA tail (Garcia et al., 2011, Daley et al., 2017). This 3´ ssDNA overhang, which is unstable, is rapidly shielded by

replication protein A (RPA), which with mediators BRCA1, BRCA2 and partner and localizer of BRCA2 (PALB2), is replaced by DNA repair protein RAD51 homolog 1 (RAD51), leading to formation of extended nucleoprotein filaments (Renkawitz et al., 2014, Bhat & Cortez, 2018). The protein filaments at the 3´ search for homology and invade the strand of the homologous DNA generating a displacement loop (D-loop) (San Filippo et al., 2008). The resolution junction is processed by resolvases that terminate the repair process (Heyer et al., 2010, Symington & Gautier, 2011).

NHEJ and HDR repair mechanisms and their outcomes are depicted in Figure 1.3.

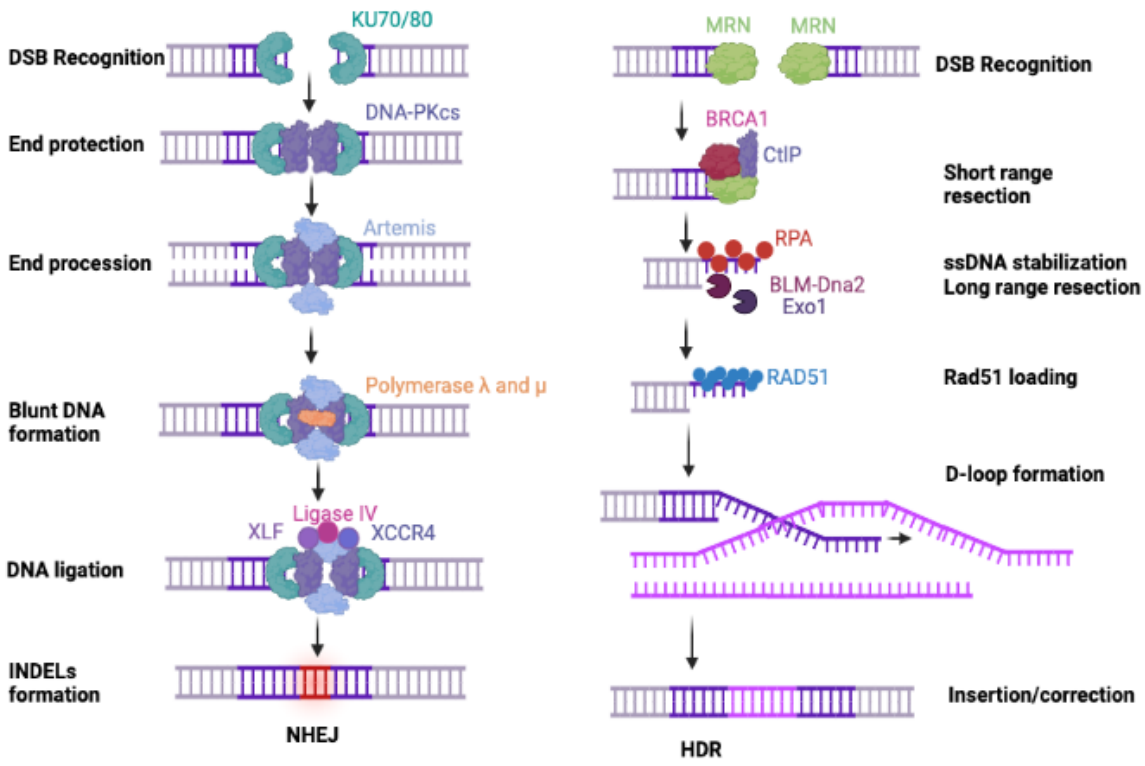


Figure 1.3. Mechanisms and outcomes of non-homologous end joining (NHEJ) and homology-directed repair pathways in mammalian cells. NHEJ initiates after a DSB by binding of Ku70-Ku80 to blunt DNA ends (Fell & Schild-Poulter, 2015). KU heterodimer recruits DNA-PKcs, forming a complex (Kragelund et al., 2016). DNA-PKcs phosphorylates Artemis, X-ray repair cross complementing protein 4 (XRCC4), DNA ligase IV and XRCC4-like factor (XLF) (Davis et al., 2014). Exonuclease Artemis binds with polymerases lambda and mu to prepare blunt ends for ligation. XRCC4-DNA ligase IV-XLF complex performs ligations (Conlin et al., 2017, Stinson et al., 2020) and DNA bases are added or removed randomly by DNA polymerases or nucleases, leading to indels formation (Yang et al., 2020). HDR process is initiated by the MRN complex (MRE11-RAD50-NBS1) (Symington & Gautier, 2011). MRN complex recruits C-terminal-binding interacting protein (CtIP) and initiates resection (Huertas & Jackson, 2009). Exonuclease 1 (Exo1) and the DNA replication ATP-dependent helicase/nuclease DNA2/bloom syndrome protein (BLM) complex perform long-range DNA resection leading to 3´ ssDNA tails (Garcia et al., 2011, Daley et al., 2017). 3´ ssDNA overhang is shielded by replication protein A (RPA), which with mediators BRCA1, BRCA2 and partner and localizer of BRCA2 (PALB2), is replaced by DNA repair protein RAD51 homolog 1 (RAD51), leading to formation of extended nucleoprotein filaments (Renkawitz et al., 2014, Bhat & Cortez, 2018). Protein filaments at the 3´ invade the strand of the homologous DNA forming a displacement loop (D-loop) (San Filippo et al., 2008). The resolution junction is processed by resolvases that terminate the repair process (Heyer et al., 2010, Symington & Gautier, 2011). Adapted from (Yang et al., 2020). Created with BioRender.com.

In the following years, thousands of CRISPR-related papers were published on PubMed detailing research improving CRISPR specificity and development of new applications. A relevant feature is the “multiplexing” of the gRNAs, demonstrated by a study that showed that CRISPR/Cas systems could be applied in mammalian cells and multiple gRNAs could be used in parallel to target multiple sites in the same cells (McCarty et al., 2020).

Additionally, various CRISPR/Cas systems have been found and classified into two classes, five types and 16 subtypes varying in the PAM sequence and types of Cas proteins (Makarova et al., 2015). A summary of the characteristics of each type can be found on Table 1.2.

Table 1.2. Summary of the classification of Cas proteins based on their class and type. Type classification depends on particular Cas proteins, which differ mainly in their distinct domain architecture. Adapted from (Makarova et al., 2015).

Class	Type	Details
1 (Multi-subunit crRNA-effector complex)	**I**	Type I systems express *cas3* gene, which encodes a ssDNA-stimulated superfamily 2 helicase with the capacity to unwind dsDNA and RNA-DNA duplexes (Mulepati & Bailey, 2011), (Gong et al., 2014), (Huo et al., 2014). They often contain an HD domain involved in cleavage of target DNA (Sinkunas et al., 2011) and express Cas5, Cas7 and proteins from the Cas8 family as part of their effector module.
	III	Type III systems express signature gene *cas10*, which encodes a multi-domain protein with a Palm domain (a variant of RNA recognition motif RRM), homologous to the core domain of various nucleic acid polymerases and cyclases. They also express Cas5 and Cas7 proteins, as part of their effector module and an HD nuclease (different from the type I one) (Makarova et al., 2002, Makarova et al., 2006).
	IV	Putative type IV systems lack *cas1* and *cas2* genes and encode a predicted minimal multi-subunit crRNA-effector complex consisting of Csf1 (signature gene for these systems), Cas5 and Cas7 (Makarova et al., 2011).
2 (Single Cas protein)	**II**	Type II systems express signature *cas9* gene, that encodes for the multidomain protein that cleaves target DNA via the tracrRNA:crRNA complex (Jinek et al., 2012). Additionally, type II systems contain the conserved *cas1* and *cas2* genes that form a complex that allows for spacer acquisition (Nuñez et al., 2014). Systems commonly used in the lab belong to type II CRISPR systems, such as *Sp*Cas9 and *Sa*Cas9
	V	Type V systems encode the *cpf1* gene, adjacent to *cas1* and *cas2,* which expresses Cpf1 (Schunder et al., 2013), a functional analogue of Cas9.

CRISPR/Cas9 programmable systems addressed some of the challenges presented by ZFNs and TALENs, such as the need to redesign proteins (ZF proteins and TALEs) for each target. Unlike ZFNs and TALENs, that recognise targets by protein-DNA interactions, CRISPR/Cas9 systems recognise targets by DNA and RNA base pairing (Gaj et al., 2013). Furthermore, CRISPR/Cas9 systems can be adapted to target a specific site containing a PAM sequence by changing the single guide RNA (sgRNA) sequence that binds to target DNA. Simplifying the process even further, designed sgRNA can be cloned into a plasmid backbone containing a tracrRNA:crRNA and a Cas protein component (Gupta & Musunuru, 2014). These findings led CRISPR systems to becoming the most easy and quick tool for genome editing.

More recently, new systems that facilitate point mutation corrections through single-nucleotide conversions have been developed: DNA base-editors. These systems have two main components, a catalytically impaired Cas protein (Cas nickase), that binds to DNA, and a single-stranded DNA modifying enzyme, for nucleotide alteration (Kantor et al., 2020). Two base-editing systems have been described, cytosine base-editors (CBEs) and adenine base-editors (ABEs), that allow for conversion of adenine and thymine to guanine and cytosine in genomic DNA (Komor et al., 2016, Gaudelli et al., 2017). An additional system was invented, referred to as prime editors. This system allows for replacement or insertion of any base pair, with an impaired Cas protein fused to a reverse transcriptase that can edit a sequence using the prime editor guide RNA as a

template (Anzalone et al., 2019). The three mentioned base-editing systems are depicted in Figure 1.4.

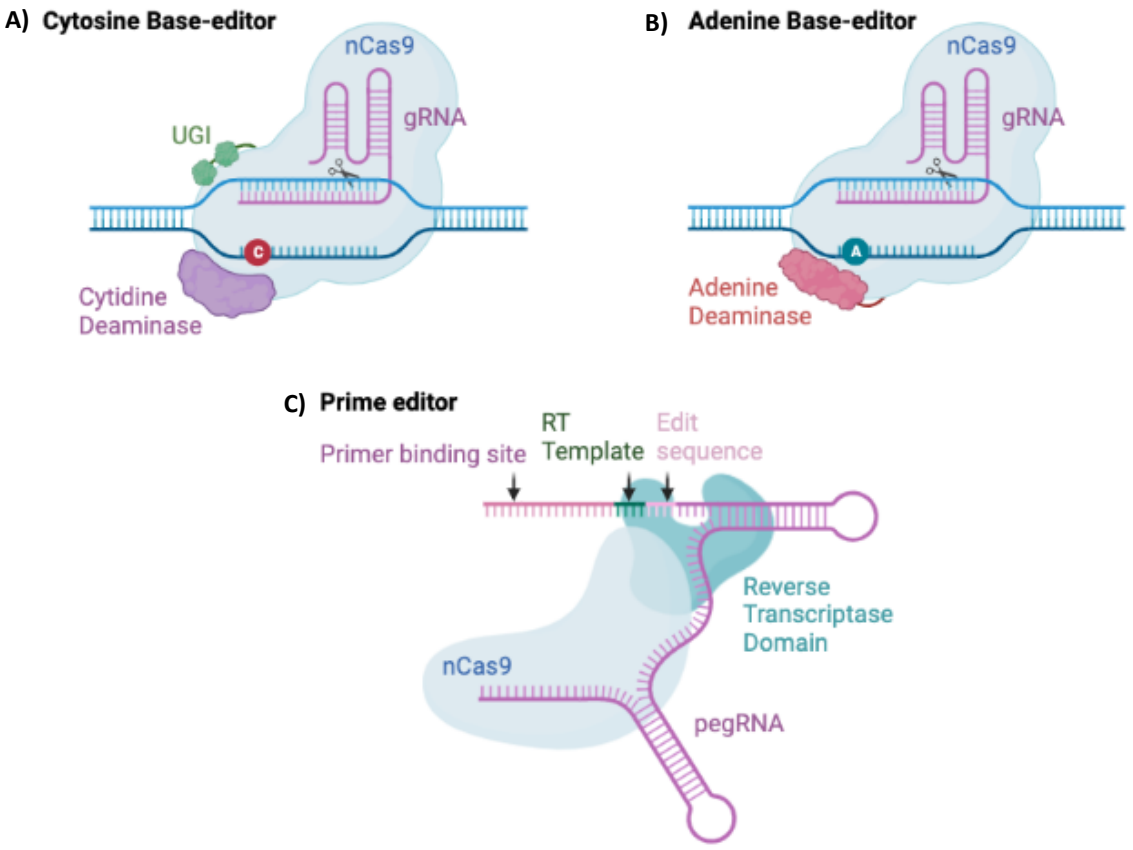


Figure 1.4. Illustrative summary of base-editing CRISPR systems. A) Cytosine base-editors: Cytidine deaminase generates an uracil by deamination in cytosine, which base pairs as thymidine in DNA. Fused uracil DNA glycosylate (UGI) inhibits activity of uracil N-glycosylate (UNG) increasing editing efficiency in human cells. B) Adenine base-editors: generates a inosine by adenosine deamination, which has the base pairing preferences of guanosine in DNA. C) Prime editors: composed of an engineered reverse transcriptase fused to a Cas9 nickcase and a prime-editing guide RNA (pegRNA), which contains the target complementary sequence (primer binding site), the reverse transcriptase (RT) template and the edit sequence. Adapted from (Kantor et al., 2020). Created with BioRender.com.

Therapeutic applications of ZFNs, TALENs and CRISPR systems include permanently correcting genetic mutations *in-vivo* that lead to inherited diseases, facilitating the generation of animal models and developing *ex-vivo* gene therapies, which have recently showed promising results on clinical trials.

ZFN systems have been used in clinical trials to treat human immunodeficiency virus (HIV) by modifying autologous CD4 T-cells *ex-vivo* to silence *CCR5* gene, the major coreceptor for HIV. Results after cell infusions showed reduced viral loads in some patients and proved safety (Tebas et al., 2014). In another clinical trial, ZNFs were used to edit autologous CD34+ cells to treat β-thalassemia. Modified cells infused to patients showed safety and improved haemoglobin levels (Thompson et al., 2018). ZFNs have also been used to create rat animal models for cystic fibrosis (Tuggle et al., 2014) and rat animal models for X-linked severe combined immunodeficiency (X-SCID) (Mashimo et al., 2010). Lastly, ZNFs were used to improve performance of immunotherapies for leukaemia by inactivating expression of endogenous T-cell receptor genes and enabling generation of tumour specific T-cells (Provasi et al., 2012).

TALENs were used to correct human β-globin (*HBB*) gene in disease-specific patient-derived human induced pluripotent stem cells (hiPSCs) to treat sickle cell anaemia (N. Sun & Zhao, 2014). TALENs were also used to generate autologous CAR T-cells to treat leukaemia, leading to remission in two infants (Qasim et al., 2017).

Similar therapeutic applications have now been achieved with CRISPR/Cas. CRISPR/Cas9 was used to target *BCL11A* transcription factor, that represses gamma-globin and foetal haemoglobin expression, in autologous CD34+ cells. Two patients infused with edited cells showed an increase in foetal haemoglobin and in the patient with sickle cell disease eliminated vaso-occlusive episodes (Frangoul et al., 2021). In a phase I clinical trial to assess safety and feasibility, three patients with advanced cancer were treated with CRISPR/Cas9 edited T-cells, to improve their antitumor immunity and to recognize tumours (by disruption of *TRAC*, *TRBC* and *PDCD1* genes and introduction of *NY-ESO-1* gene). Treatment was well tolerated and durable engraftment was observed for the duration of the study (Stadtmauer et al., 2020). CRISPR was recently used to create a pool of knockout mouse models that have led to the identification of a target that could improve cancer immunotherapy efficacy (X. Wang et al., 2021). Remarkably, a few CRISPR systems have reached clinical trials. The first-in-human *in-vivo* CRISPR/Cas9 phase I trial tested CRISPR-Cas9 as an *in-vivo* therapeutic agent to treat transthyretin amyloidosis by reducing accumulation of misfolded transthyretin protein and showed promising safety and efficacy results. This system was delivered with lipid nanoparticles encapsulating mRNA for the Cas9 protein and a single guide RNA (Gillmore et al., 2021). Other trials include the EDIT-101 trial for Leber Congenital Amaurosis Type 10 (NCT03872479), which previously demonstrated efficacious genome editing with an *Sa*Cas9 system in CEP290-associated Leber congenital amaurosis mouse model and safety in nonhuman primates with AAV5 vectors (Maeder et al., 2019) and the EBT-101

phase I/II clinical trial (NCT05144386) to treat HIV by genome editing. This *Sa*Cas9 system delivered by AAV9 vectors deletes the HIV-1 proviral DNA (C. Yin et al., 2017).

1.1.2. Delivery methods for genome editing.

Genome editing strategies require a delivery system for programmable endonucleases to reach target cells. Programmable nucleases can be delivered in the form of DNA, mRNA or protein (H.-X. Zhang et al., 2019). In this section, the focus will be on delivery methods for *in-vivo* genome editing strategies, that can be mainly classified into viral vectors, lipid nanoparticles and virus-like particles (VLP) (Raguram et al., 2022).

A popular non-viral method are synthetic lipid nanoparticles (LNPs), typically composed of a ionizable or cationic lipid, a helper lipid, cholesterol and polyethylene glycol (PEG)-lipid; varying these components leads to different pharmacokinetic profiles (Paunovska et al., 2022). LNPs have been adapted to deliver *Sp*Cas9 nuclease mRNA and protein, although the sgRNA expression cassette and DNA donor template were co-delivered in an AAV8 vector (H. Yin et al., 2016). Later on, a modified sgRNA expression cassette was co-delivered with *Sp*Cas9, both packaged into LNPs and achieved higher editing (80%) *in-vivo* in mice (H. Yin, Song, et al., 2017). However, most of these particles target the liver, thus research is being done to achieve non-liver delivery and is yet to be tested in genome editing applications (Raguram et al., 2022). Moreover, cationic lipids and polymers as non-viral delivery methods have some advantages such as non-

immunogenicity and relatively easy production processes, but toxicity is a common concern (Lv et al., 2006, Zhang et al., 2019).

VLPs are non-infectious assemblies of viral proteins that allow for packaging of mRNA, protein or ribonucleoproteins (Lyu et al., 2020). Most reported VLPs are based on retroviruses, hence they allow flexibility to package large cargos (W. Zhang et al., 2015). Recently, genome editing systems have been delivered with VLPs, such as delivery of *Sp*Cas9 in a non-integrating retrovirus all-in-one particle to attempt a targeted knockout *in-vitro* (Knopp et al., 2018); *in-vitro* delivery of *Sa*Cas9 mRNA in a lentivirus-like bio-nanoparticle (Lu et al., 2019); *in-vivo* delivery in mice via subretinal injection, with a similar system using HIV-1 VLPs and *Sp*Cas9, which achieved prevention of wet age-related macular degeneration (Ling et al., 2021) and on a separate study, similar particles delivered *in-vivo* via intracorneal injection cured herpetic stromal keratinitis in mice (D. Yin et al., 2021). A disadvantage of using VLPs to deliver CRISPR systems is that non-modified gRNAs are rapidly degraded (Allen et al., 2021) .

Another non-viral approach consists in delivering purified recombinant Cas9 protein packaged into a nanoparticle. *In-vitro* studies showed transient genome-editing with reduced off-target effects by delivering purified recombinant Cas9 protein complexed with *in-vitro* transcribed gRNA, also known as RNA guided engineered nuclease (RGEN) ribonucleoproteins complex (RNPs) (S. Kim et al., 2014). This approach works well for *ex-vivo* cell therapies but presents at least three limitations for *in-vivo* gene therapies.

Firstly, direct protein delivery can trigger immune responses (Chew, 2018). Since CRISPR-Cas systems are derived from bacteria, it is common to find pre-existing antibodies against Cas proteins in humans. *Sa*Cas9 and *Sp*Cas9 pre-existent antibodies were detected by Western Blot in 67% and 42% of 12 serum samples respectively from peripheral blood from healthy adults (Charlesworth et al., 2018) and in 10% and 2.5% respectively of 200 human serum samples analysed using an ELISA-based assay (Simhadri et al., 2018). Secondly, systemic delivery of Cas protein is challenging due to its large size (H. Yin, Kauffman, et al., 2017). Lastly, manufacturing and purification of large nucleases is a complex process and endotoxin contamination is concern (H.-X. Wang et al., 2017).

Viruses' natural infectivity and native tropism to different cell lines make them an attractive vehicle for genome editing agents. The most popular vectors for *in-vivo* genome editing are adeno-associated viruses (AAVs), lentiviruses or adenoviruses (Raguram et al., 2022).

Adenovirus (Ad) is a non-enveloped, double-stranded DNA virus with a large genome that is episomally maintained after transduction (C. S. Lee et al., 2017). Over 55 serotypes have been identified in humans and are grouped in species A to G based on phylogenetic, genome structure and hemagglutination criteria. Human adenoviruses can cause mild respiratory, gastrointestinal, urogenital and ocular disease. (Volpers & Kochanek, 2004, Gonçalves & de Vries, 2006). Furthermore, their prevalence in healthy

individuals leads to pre-existing immunity that hinders the potential adenovirus-derived vectors from most serotypes (Davison et al., 2003, Vannucci et al., 2013). Adenoviral vectors have been tested with genome editing systems mainly targeting stem cells and progenitor cells (Tasca et al., 2020). However, their strong immunogenicity makes them ideal for immunotherapy with oncolytic viruses (Choi & Yun, 2013).

Lentiviral vectors, derived from enveloped HIV-1 viruses, were made replication-incompetent by making deletions in the 3´ long terminal repeats and by splitting component for viral production into multiple plasmids (Naldini et al., 1996, Dull et al., 1998). Lentiviral vectors can transduce dividing and non-diving cells (Dull et al., 1998) and can package up to 10 kb of DNA (Sweeney & Vink, 2021). To avoid integration of the delivered transgene, integrase-deficient lentiviral vectors have been developed which are maintained as episomes (Wanisch & Yáñez-Muñoz, 2009). However, studies have shown residual genome integration from these “non-integrating lentiviral vectors” (Apolonia et al., 2007), hence most current applications are limited to gene augmentation not genome editing (Milone & O’Doherty, 2018).

AAV is a non-enveloped icosahedral virus (Chapman & Agbandje-McKenna, 2005) endogenous to various mammalian species, including humans (Carter, 1992). AAVs have been engineered as vectors to express genes of interests. Recombinant AAVs can package up to 5.2 kb of DNA, but studies have showed that the optimal size is less than 4.9 kb to avoid reduction of packaging efficiency (Dong et al., 1996). Different serotypes

have been identified and their natural tropism has been studied in mice; AAV serotypes 1 to 9 were detected in heart, lung, liver, kidney, testes, brain gastrocnemius and hamstrings. However, each serotype showed a higher tropism for certain tissues: AAV serotypes 1, 2, 5, 6, 7, and 9 target primarily the liver and hindlimbs, AAV4 showed higher tropism for the lung and heart, AAV6 showed a bias for the heart, while AAV7 showed a strong tropism for the liver, AAV 8 and 9 showed more ubiquitous and robust tissue expression and AAV9 also showed a high expression in brain and heart (Zincarelli et al., 2008). Over a 100 AAV variants of these serotypes have been found in human or nonhuman primates (G. Gao et al., 2005). Furthermore, AAV vectors have been genetically engineered to enhance their transduction efficiency and to overcome immunity barriers; mutants have been generated by rational design or directed evolution (C. Li & Samulski, 2020). A successful strategy is capsid engineering, which has led to the generation of vectors such as AAV2.5 (with residues from AAV1 into an AAV2 capsid) that possesses an increased muscle tropism and has been used to deliver micro-dystrophin as a potential treatment for Duchenne muscular dystrophy (Bowles et al., 2012) or AAV9.HR (mutated AAV9 capsid) with enhanced ability to cross the blood-brain barrier and transduce neurons in neonatal mice (D. Wang et al., 2018). Several gene therapies using AAVs as a delivery mechanism have reached clinical trials (J. R. Mendell et al., 2021) and a few have been approved by the FDA, such as Zolgensma to treat SMA using an AAV9 vector (Mendell et al., 2017; Mahajan, 2019) and Luxturna (Voretigene Neparvovec-rzyl) to treat inherited retinal dystrophy with AAV2 vector (Russell et al., 2017, Miraldi Utz et al., 2018). To address the limiting cargo capacity of AAVs when used for genome editing, alternative strategies include the use of dual vectors, in which the

transgene is split into two vectors and full-length expression is achieved after co-transduction (Tornabene & Trapani, 2020) or by replacing *Sp*Cas9 protein (4.10 kbp) for smaller orthologs, such as S*taphylococcus aureus* Cas9 (*Sa*Cas9) (Ran et al., 2015) or *Campolylobacter jejuni* Cas9 (*Cj*Cas9), which are ∼1 kilobase shorter than *Sp*Cas9 and hence fit within packaging limits of AAVs (E. Kim et al., 2017).

Even though viral vectors sound like a promising approach to deliver genome editing mechanisms, it is not straightforward to select the optimal vector. Selection will be dependent on the system being used, for example: the large size of TALENs limits delivery with size-restricted vectors such as AAV (Gaj et al., 2013), so an AAV dual-vector approach could be tested or a lentiviral vector could be used, as their plasmid vectors can accommodate full-length TALEN sequences. However, lentiviral vectors are prone to rearrangements after transduction (Holkers et al., 2013) and repeat sequences from the TALENS difficult cloning strategies.

Furthermore, it is relevant to consider that viral vectors still present some challenges, such as overcoming immune responses, the potential risk of random insertional mutagenesis (Bessis et al., 2004) and the challenge of reaching target tissues. In addition to the costs of production and the need for high amounts of vector per treatment.

1.1.3. CHALLENGES & FUTURE DIRECTIONS OF GENOME EDITING.

Genome editing strategies need to overcome major challenges before reaching the clinic. Some of these include increasing the efficiency of gene correction (considering levels of gene correction required for therapeutic effect differs based on the disease) and overcoming the challenge to induce HDR efficiency needed for strategies that involve the use of a repair template (Cox et al., 2015).

Specificity of editing nucleases needs to be improved: off-target mutations can create cells with oncogenic potential or functional impairment, CRISPR/Cas9 can lead to large deletions or complex rearrangements (Kosicki et al., 2018). Rapid screening methods should be developed to scan total genome mutations induced by CRISPR systems (Devkota, 2018).

One of the major challenges for *in-vivo* genome editing strategies to reach the clinic, is the need for efficient and safe delivery methods that can reach a large fraction of target cells or tissues (Raguram et al., 2022). Furthermore, there is a need to improve delivery systems to avoid an immune response: for *in-vivo* applications the most promising system are viral vectors, particularly AAV vectors. However, pre-existing immunity to delivery vehicles and pre-existing antibodies could neutralize viral vectors (Verdera et al., 2020; Weber, 2021). To overcome these challenges, efficient alternative delivery methods are needed. Some proposed strategies include the use of alternative AAV

capsids and immunomodulatory treatments (Duan, 2018). AAV capsid engineering to enhance a particular muscle tropism was briefly discussed in the previous section. However, AAV capsids have also been engineered to escape the immune system. A successful strategy is using directed evolution under selected pressure from neutralizing antibodies, leading to the generation of libraries with random mutations in the capsids (Maheshri et al., 2006, Waterkamp et al., 2006). This strategy resulted in the generation of AAV-DJ, composed of AAV2, AAV8 and AAV9 capsids, which transduces liver more efficiently than parental serotypes (Grimm et al., 2008). Other non-genetic approaches include the use of pharmacological agents, like rituximab and rapamycin to prevent production of neutralizing antibodies by B-cells (Mingozzi et al., 2012, Meliani et al., 2018), and coating the AAV surface with lipids or cell-derived extracellular vesicles to prevent their recognition (Meliani et al., 2017, Katrekar et al., 2018).

Lasty, another challenge is the pre-existing immunity to Cas9 proteins: studies have demonstrated human pre-existing immunity in 10%-67% of the population against *Sa*Cas9 and 2.5%-42% against *Sp*Cas9 (Charlesworth et al., 2018; Simhadri et al., 2018) depending on the study. Potential solutions include recoded versions of a Cas protein to evade the immune system or identify Cas protein from microorganisms that have not been in contact with humans (Devkota, 2018).

1.2. DUCHENNE MUSCULAR DYSTROPHY.

DMD belongs to a group of diseases known as the muscular dystrophies. These are a group of inherited disorders characterised by muscle wasting and weakness in variable degrees and distributions. A summary of the most common muscular dystrophies (according to Emery (2002)), their symptoms and affected proteins that lead to particular phenotypes are presented on Table 1.3.

Table 1.3. Summary of common forms of muscular dystrophies. Summarized description of phenotype and defective proteins leading to the disease.

Muscular Dystrophy Form	Description and Symptoms	Defective proteins
Congenital muscular dystrophy (Tomé FM et al., 1994), (Hayashi et al., 1998), (Kobayashi et al., 1998), (Norwood et al., 2009)	Heterogenous group of autosomal recessively inherited disorders leading to hypotonia and weakness within the first months of life, with or without mental retardation. Prevalence of 0.89/100,000 individuals. Muscle weakness is non-progressive, but patients develop feeding and respiratory problems.	Laminin α2 (muscle extracellular protein) Integrin α7 Fukutin
Duchenne (DMD) and Becker (BMD) (Becker & Kiener, 1955), (Becker, 1962), (Hoffman et al., 1987), (Grain et al., 2001), (Moat et al., 2013), (A. E. H. Emery et al., 2015).	X-linked disorders affecting approximately 1 in 5,000 male births. DMD has an early onset in childhood, showing difficulties running and climbing stairs. Some degree of mental impairments is common. Weakness is progressive. Most frequent cause of death is pneumonia with cardiac involvement around late 20s-early 30s. BMD symptoms can be similar to DMD symptoms. However, some patients are asymptomatic, and it has an estimated incidence of 1 individual per 30,000 male births. Onset is	Dystrophin (sarcolemmal protein)

	around 12 years or later, leading to death around 40-50 years old. In both dystrophies, 5-10% of female carriers show some degree of muscle weakness.	
Emery-Dreifuss muscular dystrophy (A. E. Emery, 1989), (Bione et al., 1994)	Patients start showing contractions until the entire spine becomes restricted. Then there is a slowly progressive muscle wasting and weakness and cardiomyopathy arises showing cardiac disease usually by the age of 30. There is a prevalence of 0.13/100,000 individuals.	Emerin (nuclear membrane protein)
Distal muscular dystrophy (Nonaka, 1999)	Mainly distal weakness. Presents late (over 40 years old) or early (less than 30 years old) onset. Can be considered a myopathy rather than a dystrophy. Incidence estimated to be 1 in 1,000 individuals.	Dysferlin (sarcolemmal associated protein)
Facioscapulohumeral muscular dystrophy (Tawil & Van Der Maarel, 2006), (Lemmers et al., 2010)	The affected muscle groups are facial and shoulder girdle, foot extensors and pelvic muscles. In most cases the heart is not implicated. Incidence is of 3.95/100,000 individuals and presents a varied onset. Most individuals notice symptoms by the age of 20.	Toxic gain of function caused by mutations in the DUX4 gene
Oculopharyngeal muscular dystrophy (Brais et al., 1999), (Brais et al., 1998)	Onset around the 3rd decade of life, affecting extraocular muscles, facial muscles neck and limb musculature. It has an estimated incidence of 1 in 100,000 people. Also, patients show dysphagia.	Poly-(A)-binding protein (PABP2 gene)
Limb-girdle muscular dystrophy (LGMD) (K. M. Bushby, 1999), (K. M. D. Bushby, 1999), (Johnson & Statland, 2022), (Broglio et al., 2010)	LGMDs are a group of inherited muscle disorder caused by over 29 individual genes. It has an incidence of 2.27/100,000 individuals. Common features are weakness affecting proximal limb-girdle musculature and progressive disability. There are two groups: autosomal dominant (AD) inherited or autosomal recessive (AR) inherited which are often caused by loss-of-function mutation in muscle structural or repair protein, these lead to younger ages of onset. Most of	AD: Myotilin (LGMD1A), Lamin A/C (LGMD1B), Caveloin 3 (LGMD1C), proteins unknown for LGMD1D, 1E, 1F, 1G. AR: Calpain-3 (LGMD2A), Dysferlin (LGMD2B), ɣ-sarcoglycan (LGMD2B), α-sarcoglycan

	them associated with significant cardiac involvement.	(adhalin) (LGMD2D), β-sarcoglycan (LGMD2E), δ-sacroglycan (LFMD2F), Telethonin (LGMD2G), Fukutin-related (LGMD2I), TRIM 32 (LGMD2H), Titin (LGMD2J), POMT1 (LGMD2K), POMT2 (LGMD2N), protein unknown for. LGMD2L.

The most common type of muscular dystrophy is Duchenne Muscular Dystrophy (DMD), an inherited, X-linked neuromuscular disease, resulting from mutations across the *DMD* gene that lead to the absence of dystrophin protein; compromising muscle stability and contractility, giving rise to progressive muscle wasting and loss of independent ambulation by the age of 13 years (Hoffman et al., 1987).

1.2.1. Clinical features and prevalence of Duchenne muscular dystrophy.

DMD is estimated to affect 1 in 5000 male births (Moat et al., 2013). The common onset age is between 3-5 years of age when symptoms include walking abnormalities, gross motor delays, difficulties rising from the ground, frequent falls and raised serum creatine kinase (CK) levels (Zatz et al., 1991, Emery et al., 2015, Yiu & Kornberg, 2015). Weakness

is seen in proximal lower limbs, followed by upper limbs and distal muscles (Darras et al., 2015). Most boys stop gaining motor skills around the age of 6, when progressive deterioration in strength starts, leading to the need of wheelchair by the age of 11-22 years (Darras et al., 2015). Clinical cardiomyopathy is evident after 10 years of age and incidence increases with age. By the age of 18, it is present in all patients (Nigro et al., 1990). Respiratory muscle decline starts around 12 years of age and decreases by 4-8% per year (Khirani et al., 2014), leading to the need of assisted ventilation at around 20 years of age (Mercuri et al., 2019). Intellectual impairment, particularly verbal is associated with DMD, but is non-progressive and does not affect all children (Leibowitz & Dubowitz, 1981). Boys with Duchenne have a higher prevalence of attention-deficit hyperactive disorders (ADHD) and autism spectrum disorder (Hendriksen & Vles, 2008). Scoliosis development is frequent, impacting respiratory vital capacity in untreated patients (A. D. Smith et al., 1989). Most patients with DMD, receiving optimal care, die between 20-40 years of age from respiratory or cardiac failure (Mercuri et al., 2019).

1.2.1.1. Genetic basis of Duchenne muscular dystrophy.

The *DMD* gene is one of the largest human genes spanning 2,200 kb, approximately 0.1% of the whole human genome (Koenig et al., 1987). Its coding sequence is distributed across 79 exons and encodes a protein called dystrophin (Roberts et al., 1992).

Dystrophin protein is localised at the sarcolemma of skeletal and cardiac muscle cells and is a component of the dystrophin-associated glycoprotein complex (DAPC), where it acts as a mechanical link between the intracellular cytoskeleton and the extracellular matrix (Rando, 2001, Gao & McNally, 2015). The absence of dystrophin protein prevents the correct formation of the DAPC. The dystrophin-associated proteins can be divided in three groups based on their cellular localization: extracellular (α-dystroglycan that functions as a receptor for extracellular ligands), proteins at the transmembrane (β-dystroglycan, sarcoglycans, sarcospan) and cytoplasmic proteins (dystrophin, dystrobrevin, syntrophins, neuronal nitric oxide synthase) (Q. Gao & McNally, 2015) (Fig. 1.5).

Dystrophin protein has four main functional domains, an actin-binding amino-terminal domain (ABD1), a central rod domain, a cystein-rich domain and a carboxyl terminus. ABD1 contains the calponin homology domains that bind directly to F-actin, allowing dystrophin to link to the subsarcolemmal actin network (Way et al., 1992). The central rod domain contains 24 spectrin-like repeats and harbours a second actin-binding motif that can interact with acidic actin filaments (Amann et al., 1998). The rod domain also mediates interaction between dystrophin and microtubules via spectrin-like repeats 20-23 (Belanto et al., 2014). The 24 spectrin like repeats are interrupted by four flexible hinges at precise positions that are relevant for the mechanical properties of the elongated dystrophin (Koenig & Kunkel, 1990). At the end of the rod domain, Hinge 4

that contains a WW domain, works as a protein-protein interaction module (Ilsley et al., 2002).

The cysteine-rich domain contains the EF-hand motifs, that consist of two α-helices linked by a loop region implicated in calcium binding (Koenig et al., 1988). The carboxy-terminal (CT) domain provides binding sites for dystrobrevin and syntrophin, mediating their sarcolemma localisation (Sadoulet-Puccio et al., 1997).

An illustration of the DAPC and the main elements of dystrophin can be seen on Fig 1.5.

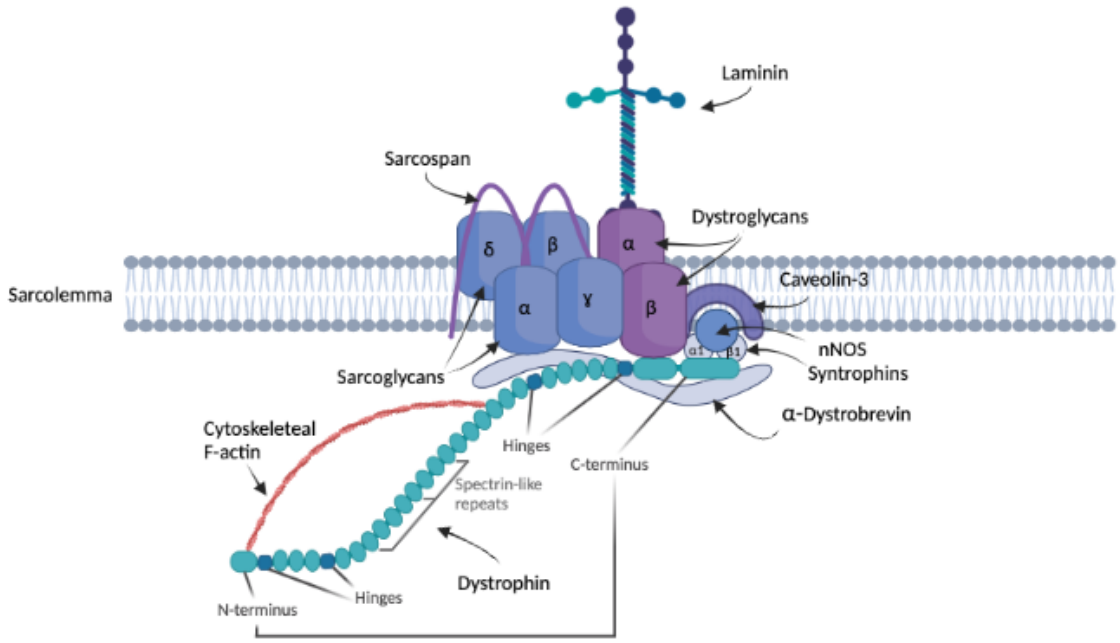


Figure 1.5. Dystrophin-associated protein complex (DAPC) localized at the sarcolemma in muscle cells. The DAPC provides a strong structural link between intracellular cytoskeleton and the extracellular matrix (ECM). Dystrophin binds to intracellular actin network and interacts with β-dystroglycan, to link the cytoskeleton with the DAPC. Components of the DAPC include dystroglycan, sarcolgycan and sarcospan. Dystroglycan, composed of subunits α and β, is a receptor for laminin, that links the DAPC to the ECM. Sarcoglycans form a complex with the sarcospan and strengthen connection between alpha and β-dystroglycans. The C-terminus of dystrophin recruits syntrophins, dystrobrevins and nNOS, which participate in signal transduction pathways, and the scaffolding protein caveolin-3 (Rando, 2001, Galbiati et al., 2001, Gao & McNally, 2015). Dystrophin domains can be observed: N-terminal actin-binding domain, central rod domain (with 24 spectrin-like repeats) with four hinge regions among them and the cysteine-rich domain next to the C-terminal domain. Adapted from (Gao & McNally, 2015). Created with BioRender.com.

DMD gene mutations that cause a shift in the open transcript reading frame generally lead to the lack of dystrophin expression, and the severe phenotype of DMD. Mutations that do not disrupt the transcript reading frame lead the expression of a truncated dystrophin protein and the less severe phenotype of BMD. This is “the reading frame rule” and it explains the majority of phenotypic differences between DMD and BMD

(Monaco et al., 1988), with a few exceptions to the rule (Tuffery-Giraud et al., 2009). Deletions and duplications present a non-random distribution, with a deletion hotspot between exons 45-55, representing 74% of identified deletions, and a second hotspot accounting for 15% of deletions between exons 2-20 (Tuffery-Giraud et al., 2009). Duplications occur in 11% of patients and the rest are point mutations (11%) or small mutations, including smalls deletions or insertions (20%) (Bladen et al., 2015). To date, eleven types of mutations over the *DMD* gene have been identified and categorised (Bladen et al., 2015) (Table 1.4).

Table 1.4. Type and frequency of mutations held within the TREAT-NMD DMD Global Database. Mutations are divided in three main groups: Large mutations that include large deletions and duplications; Small mutations including small deletions, duplications and point mutations (subdivided into nonsense mutations and missense mutations); Mid-intronic mutations. Obtained from (Bladen et al., 2015).

Total cases	**7,149**	**% of total mutations**
Large mutations	5,682	79
Large deletions (> 1 exon)	4,894	68
Large duplications (> 1 exon)	784	11
Small mutations	1,445	20
Small deletions (< 1 exon)	358	5
Small insertions (< 1 exon)	132	2
Splice sites (<10 bp from exon)	199	3
Point mutations	756	11
Nonsense	726	10
Missense	30	0.4
Mid-intronic mutations	22	0.3

1.2.1.2. CURRENT TREATMENTS AND STANDARD OF CARE FOR DUCHENNE MUSCULAR DYSTROPHY.

To date, DMD is an incurable disease that involves interdisciplinary management such as medical procedures, lifelong pharmacological treatments, and physical dependence from others. International standard of care were first published in 2010 and included mainly eight areas of management: diagnostics, rehabilitation, orthopaedic, psychosocial, cardiac, pulmonary, nutritional and corticosteroids management (K. Bushby et al., 2010). Details of tools, assessments and interventions recommended for each area can be found on Figure 1.6. Since then, they have updated and expanded to add care consideration for diagnosis, neuromuscular, rehabilitation, endocrine and gastrointestinal management (Birnkrant et al., 2018a); with recommendations for respiratory, cardiac, bone health, orthopaedic and surgical management (Birnkrant et al., 2018b) and a focus on primary care, emergency management, psychosocial care and transition of care across the lifespan (Birnkrant et al., 2018c).

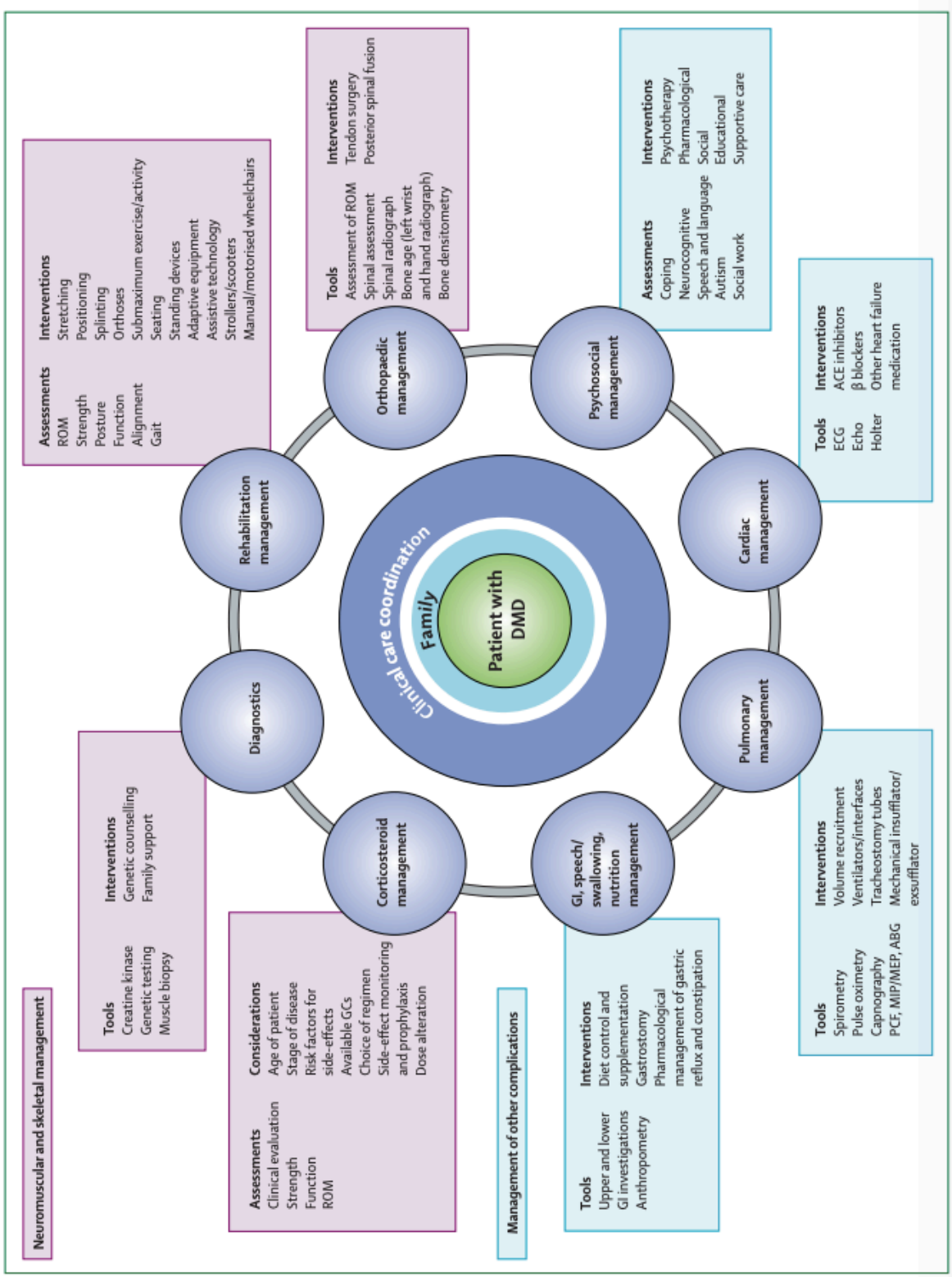


Figure 1.6. First international standards of care for DMD published in 2010. Interdisciplinary management of DMD involves a wide range of health-care professionals such as neurologists, rehabilitation specialists, neurogeneticists, paediatricians and primary care physicians. ABG =

arterial blood gas. ACE = angiotensin-converting enzyme. Echo = echocardiogram- ECM = electrocardiogram. GC = glucocorticoids. GI = gastrointestinal. MEP = maximum expiratory pressure. MIP = maximum inspiratory pressure. PCF = peak cough flow. ROM = range of motion. Obtained from (K. Bushby et al., 2010) .

However, DMD progression is mainly managed with glucocorticoids which aim to slow down the decline of muscle strength by reducing inflammation-induced muscle damage (S. Kim et al., 2015). Although, glucocorticoids are effective in the short term, they can cause clinically significant secondary effects in the long term (Manzur et al., 2008; Matthews et al., 2016). Glucocorticoids function by binding to the cytoplasmic nuclear hormone receptors (glucocorticoid receptors (GR)) and form a receptor-ligand complex (Oakley & Cidlowski, 2013). Glucocorticoid receptors suppresses the proinflammatory nuclear factor kappa B signalling pathway, leading to a potent anti-inflammatory effect (Reeves et al., 2013). Nuclear factor kappa B alongside other factors such as tumour necrosis factor (TNF)-alpha and interleukin-6 (IL-6), are chronically elevated in DMD and exacerbate oxidative stress and endogenous inflammatory response, that lead to muscle degeneration (Messina et al., 2011). However, the GR receptor-ligand complex can also bind to negative sites of the glucocorticoid response element and lead to cis-repression, a process that supresses transcription of genes that contribute to regulation of hypothalamic-pituitary-adrenal axis, bone and skin function, inflammation, angiogenesis, and lactation (Dostert & Heinzel, 2004). Their repression is associated with growth retardation, osteoporosis and skin fragility (Dostert & Heinzel, 2004). These are known secondary-effects of the long term use of glucocorticosteroids (i.e. prednisone and deflazacort), alongside more severe side effects, such as excessive weight gain

(Beenakker et al., 2005), behavioural issues (Sienko et al., 2016), cataracts (Balaban et al., 2005), osteoporosis (Bianchi et al., 2003) and delayed growth (Griggs et al., 1993).

Novel dissociative steroids, such as Vamorolone, are being explored as a superior substitute to corticosteroids (Kourakis et al., 2021). Vamorolone is an anti-inflammatory steroid analogue that also inhibits nuclear factor kappa B through interaction with the GR, but shows reduced cis-repression (Heier et al., 2013). In a Phase IIA study, over 18 months, Vamorolone was reported to be safe and well-tolerated, met the primary outcome of improved muscle function without evidence of severe secondary effects. Importantly, Vamorolone showed less incidence of weight gain and behavioural changes and did not repress growth (E. C. Smith et al., 2020). Based on available data, Vamorolone received orphan drug status in the US and Europe and will likely establish itself as a superior alternative to standard of care (Kourakis et al., 2021).

Nevertheless, current standards of care mainly focus on management of the symptoms, rather than the causes of the disease, which is why there is an urgent need to develop novel therapies to treat this disease.

1.3. NOVEL THERAPIES FOR DUCHENNE MUSCULAR DYSTROPHY.

Several therapeutic approaches are being developed targeting different elements of the DMD pathophysiology. These approaches can be mainly divided in those aiming to restore dystrophin expression and those aiming to reduce consequences of lack of dystrophin expression (Duan et al., 2021).

1.3.1. SMALL MOLECULES.

As mentioned earlier, some point mutations among the *DMD* gene lead to a premature stop codon in the mRNA. One strategy to restore the protein expression is to induce translational read-through of these mutations with antibiotics (Seto et al., 2014) or other small molecules. Approximately 10% of DMD patients have mutations that could be corrected with this approach (Bladen et al., 2015).

Gentamicin induces read-through of the premature termination codon in the mRNA and inserts a new amino acid allowing the continuation of the complete protein translation (Manuvakhova et al., 2000, Malik et al., 2010). Nevertheless, this molecule is known to cause toxicity (dose-dependent) (Balakumar et al., 2010). Research is being done to ameliorate nephrotoxicity from Gentamicin (Ibrahim et al., 2022).

A more potent molecule was identified; Ataluren (PTC124), a small molecule that induces ribosomal read-through. Systemic delivery was achieved without toxicity in *mdx* mice, possibly because its readthrough activity is specific for premature stop codons (Welch et al., 2007). This molecule reached clinical trials and was well tolerated for over 48 weeks in a double blind, placebo-controlled, longer term Phase II b study showing a safety profile and a trend of therapeutic effect with three daily doses of 10, 10 and 20mg/kg. However, clinical activity data was underpowered due to large standard deviation on the main assessment, the 6-Minute walk test (6MWT) (K. Bushby et al., 2014). In a phase III trial, Ataluren did not show a statistically significant change in the primary measure, the 6MWT, in intent-to-treat patient population. However, significant effects were observed in other measures (time function tests) and there was a statistically significant change in the 6MWT in a pre-specified subgroup of patients with a baseline >300 to <400m in the 6MTW (McDonald et al., 2017). Results led to conditional approval by the EMA (Haas et al., 2015). However, Ataluren was not approved by the FDA due to lack of persuasive positive results (Macdonald, 2017). Furthermore, the cost of the treatment is of $3,000 USD per gram, translating to approximately $300,000 USD per year of treatment for each DMD patient (Namgoong & Bertoni, 2016). The approximate cost of care per DMD patient, including expenses by health care providers and from patient´s families, is of $100,000-$120,000 USD per year (Ouyang et al., 2008, Landfeldt et al., 2014, Larkindale et al., 2014), which means that commercialization of Ataluren would bring the standard of care costs for qualifying patients to approximately $400,000 USD per year (Namgoong & Bertoni, 2016).

Alternative strategies, such as utrophin (an embryonic isoform of dystrophin) upregulation to compensate the lack of dystrophin, are being explored. Utrophin has a similar structure to dystrophin with 80% sequence similarity in the N- and C-terminal regions (Pearce et al., 1993). Utrophin is expressed in developing skeletal muscle around the entire sarcolemma instead of dystrophin (Takemitsu et al., 1991). However, in adult skeletal muscle utrophin expression becomes restricted to neuromuscular and myotendinous junctions (Khurana et al., 1991). A strategy to upregulate utrophin is by using small molecules to act at the utrophin A promoter (Soblechero-Martín et al., 2021), such as with 2-Arylbenzoxazole (Ezutromid or SMT C1100) utrophin modulator.

Ezutromid resulted from an exhaustive chemical screening and optimisation campaign and its use demonstrated an increase in utrophin expression that led to an increase in muscle strength and resistance to exercise in *mdx* mice (Tinsley et al., 2011). This molecule reached phase I clinical trials and was well tolerated by paediatric DMD patients (Ricotti et al., 2016). A follow-up study showed no adverse events and achieved plasma concentrations that should be able to modulate utrophin (Muntoni et al., 2019). However, the phase II clinical trial (NCT02858362) was terminated due to lack of efficacy (Summit Therapeutics, 2019).

A second generation compound (SMT022357), related to Ezutromid, with improved physicochemical properties showed an increase in utrophin expression in skeletal, respiratory and cardiac muscles and prevented dystrophic pathology in *mdx* mice (Guiraud et al., 2015), but has not reached clinical trials yet.

Other molecules involved in various pathways are being investigated to upregulate utrophin and have been tested in *mdx* mice, such as transcriptional upregulation with Heregulin (Krag et al., 2004) and post-transcriptional and translational events upregulation with Betaxolor and Pravastine, FDA approved molecules for blood pressure and high cholesterol respectively, they achieve utrophin upregulation through internal ribosome entry site (IRES) activation (Péladeau et al., 2020); or combinatorial therapies with Heparin (an anticoagulant used in clinic) (Péladeau et al., 2016), to name a few. However, the amount of utrophin levels required to achieve a clinical benefit in DMD patients remains to be determined (Soblechero-Martín et al., 2021).

1.3.2. CELL THERAPIES.

Cell therapies aim to transplant cells with a functional copy of the *DMD* gene into patients. There are several types of stem cells that could be used for *ex-vivo* gene therapies for DMD, such as satellite cells (myoblasts), pericytes and mesenchymal stem cells, bone marrow-derived cells or induced pluripotent stem cells (Duan et al., 2021).

Research using myoblasts dates back to 1989, when normal neonatal mouse myoblasts where transplanted into *mdx* mice via intramuscular injection (Partridge et al., 1989). However, this approach did no translate to clinic (C. Sun et al., 2020) as only small percentages of normal dystrophin were detected following healthy immune-compatible donor muscle stem cell transplantations (Karpati et al., 1993). Negative results using myoblasts are explained by immune-rejection, cell death after transplantation and scarce migration of injected cells (Skuk & Tremblay, 2003).

Satellite cells showed better muscle engraftment than myoblasts (Collins & Partridge, 2005). However, the use of satellite cells in clinic present various challenges, such as the difficulty to isolate them from a biopsy, reduced transplantation efficiency after culturing, death after transplantation and challenges in systemic delivery (C. Sun et al., 2020).

Other cells are being tested as an alternative to overcome some of these obstacles (Péault et al., 2007). Bone marrow derived myogenic cell treatments have not shown improved dystrophin production (Gussoni et al., 2002). Human pericytes in mouse (Dellavalle et al., 2011) and mouse mesoangioblasts in canine models can colonise the muscle (Sampaolesi et al., 2006). Intra-arterial transplantation of donor mesoangioblasts reached clinical trials and proved to be relatively safe, however patients showed no functional improvement (Cossu et al., 2015). Autologous CD133+ cells, muscle derived multipotent stem cells, were evaluated in clinical trials for DMD,

where they showed no side effects and led to an increase in muscle vascularization, but did not integrate into muscle fibres (Torrente et al., 2007).

iPSCs obtained by reprogramming strategies have also gained attention for muscular dystrophies, as transplantation of therapeutic cells derived from human iPSCs generated from the patient would avoid an immune response (C. Sun et al., 2020).

The use of CRIPSR systems has become increasingly popular in the context of cell therapies. CRISPR systems have been used to induce pluripotent genes and trigger reprogramming of mouse embryonic fibroblasts to establish pluripotent cell (iPSC) lines (Liu et al., 2018) and to correct DMD human iPSCs (hiPSCs); skeletal muscle myotubes and cardiomyocytes derived from these reframed hiPSCs showed restored dystrophin expression. Dystrophin restoration was also demonstrated *in-vivo* after engraftments of hiPSCs in *mdx* mice (Young et al., 2016).

Although the use of hiPSCs seems promising, further studies are needed to prove efficacy and safety of these cell therapy approaches (Seto et al., 2014). Additionally, there are technical hurdles to overcome before reaching clinical trials. There is a need to identify the best somatic cell type from which to generate hiPSCs, the route of delivery needs to be optimised, long-term stability and colonisation need to be achieved

and genome editing performed in *ex-vivo* cell therapies, must be achieved without any off-target effects (C. Sun et al., 2020).

1.3.3. Genome therapies.

1.3.3.1. Exon skipping for out-of-frame deletions.

Exon skipping aims to restore the mRNA reading frame with antisense oligonucleotides (AONs). AONs are chemically synthesised 20-30 bp single stranded nucleic acids (Brolin & Shiraishi, 2011). These molecules are designed to mask exonic splicing enhancer motifs on out-of-frame exons in pre-mRNA, leading to its exclusion in mature mRNA (so called exon skipping) and hence restoring the reading frame (Aartsma-Rus et al., 2009). This leads to the expression of a truncated but functional Becker-like dystrophin protein.

AONs can have different chemistries such as bicyclic-locked nucleic acid (LNA), ethylene-bridged nucleic acid (ENA), 2´-O-Methyl phosphorothioated (2OME-PS) AON, peptide nucleic acid (PNA), phosphorodiamidate morpholino oligomer (PMO) (Nakamura & Takeda, 2009) or tricyclo-DNA (tcDNA) (Goyenvalle et al., 2016). However, only a few types of AONs have been tested for exon skipping in DMD animal models, including PNAs, 2OME-PS AONs, PMOs and tcDNA AONs (Brolin & Shiraishi, 2011, Goyenvalle et al., 2015).

Clinical development is more advanced for antisense oligonucleotides targeting exons that affect the largest groups of patients, such as exon 51 (14% of patients), exon 53 (10%), exon 45 (8%), and exon 44 (6%) (Bladen et al., 2015; Duan et al., 2021). Four AONs have been granted conditional FDA approval: Eteplirsen (Aartsma-Rus & Goemans, 2019; Alfano et al., 2019), Golodirsen (Frank et al., 2020, Heo, 2020), Viltolarsen (Roshmi & Yokota, 2019) and Casimersen (Shirley, 2021).

A 31-mer PMO was designed against the splice enhancer sequence in exon 19 of the *DMD* gene and administered intraperitoneally to *mdx* mice without any carrier. Results showed exon 19 skipping and dystrophin recovery (Takeshima et al., 2005). This PMO was administered by intravenous infusion to a 10-year-old DMD patient in Japan and showed safety and some dystrophin expression recovery. However, the trial did not improve serum CK levels nor muscle strength (Takeshima et al., 2006).

Eteplirsen, developed by Sarepta Therapeutics is a 30-nucleotide PMO. This PMO hybridizes to exon 51 of *DMD* pre-mRNA and leads to its skipping during splicing, correcting the transcript reading frame and resulting in expression of a shortened functional dystrophin (Lim et al., 2017). In an open-label phase II study, dystrophin expression assessed by semiquantitative immunohistochemistry showed a significant average dystrophin increase from 8.9% to 16.4% in patients. Three patients that responded particularly well showed an increase in dystrophin expression varying from 0-2% to 7.7% and 17-18% when assessed by Western Blot (Cirak et al., 2011). A three-

year progression study in 12 patients showed a slower rate of decline in ambulation when compared to matched historical controls by 6MWT assessment (J. R. Mendell et al., 2016). The FDA granted accelerated approval in 2016 on the basis of results showing an increase in dystrophin levels in patients (FDA, 2019). However, this led to controversy due to low levels of dystrophin recovery (Kesselheim & Avorn, 2016). The FDA required an additional trial to demonstrate strong evidence of clinical benefit (Lim et al., 2017). The PROMOVI trial, a phase III, multi-centre, open label study evaluated efficacy and safety in a larger cohort for 96 weeks. Results, similar to previous ones, showed attenuation of decline on the 6MWT and significant attenuation of percent predicted forced vital capacity annual decline (PROMOVI) (McDonald et al., 2021).

To improve Eteplirsen´s efficacy Sarepta Therapeutics developed a peptide-conjugated Eteplirsen (PPMO), named SRP-5051. In July a phase I study on safety and tolerability was completed (NCT03375255) and phase II study is still active (NCT04004065). However, there was a clinical hold in 2022 following a serious adverse event of hypomagnesemia (Sarepta Therapeutics, 2022a), which led to changes in the protocol to include monitoring of additional biomarkers. A couple of months later the FDA lifted the clinical hold (Sarepta Therapeutics, 2022b). The phase II trial is estimated to be completed in 2025 (Sarepta Therapeutics, 2023).

A 2OME-PS AON called Drisapersen (PRO051), targeting exon 51, developed by Prosensa, also reached clinical trials. In a phase II study Drisapersen showed some

injection-site reactions and some renal events. There was non-statistically improved effect in the six-minute walk distance test (6MWT), even though there was a positive trend towards improvement. It was hypothesised that lack of statistical significance was due to greater data variability and subgroup heterogeneity (Voit et al., 2014). Therefore, a second analysis was performed in 80 subjects with a similar 6MWD baseline of 300-400 meters and ability to rise from the floor and there was indeed a statistically improvement in the 6MWT of 35.4 meters, suggesting a potential benefit in a less impaired population of DMD patients (Goemans et al., 2018). However, the FDA did not approve Drisapersen as the standard of substantial evidence of effectiveness had not been met (Andersone Pauline, 2016). Unlike Eteplirsen, Drisapersen internucleotide phosphorothioate linkages are negatively charged (Kole & Krieg, 2015). It has been shown that these negatively charged oligonucleotides interact with numerous proteins, including immune cell receptors like the toll-like receptors, which lead to inflammatory effects when activated (Lee et al., 2004, Henry et al., 2007).

Golodirsen is a PMO that targets exon 53 pre-mRNA, applicable to 7.7% of DMD patients (Aartsma-Rus et al., 2009). It was approved by the FDA in 2019 based on results from phase I/II clinical trials (FDA, 2019). This clinical trial showed skipping of exon 53 resulting in restoration of reading frame and expression of a truncated dystrophin localized at the sarcolemma that increased levels of regeneration in patient biopsy samples (Frank et al., 2020). Long term safety and efficacy was tested for over 3 years in

a phase I/II multi-centre trial. Results showed long-term biologic activity and safety (Servais et al., 2022).

Vitolarsen PMO was obtained through comprehensive sequence optimization and also leads to exon 53 skipping (Komaki et al., 2018). It differs from Golodirsen because of its size, Vitolarse has a 21-nucleotide sequence while Golodirsen has 25 nucleotides, which means that due to its shorter size at a per molecule level, an 80 mg/kg dose is more than threefold higher than the 30 mg/kg dose administered of Golodirsen (Aartsma-Rus & Corey, 2020). This PMO showed safety and efficacy over 24 weeks in phase II clinical trials (Clemens et al., 2020). This trial was extended to evaluate long-term functional outcomes for 2 years (NCT03167255) and results showed statistical significant improvements in timed function tests (Clemens et al., 2022).

Casimersen, a PMO targeting exon 45, was granted FDA approval in 2021 based on observed increase of dystrophin in skeletal muscle of treated patients (FDA, 2021d). Casimersen trial included a phase I dose escalation study (NCT02530905) in which treatment showed safety and tolerability (Wagner et al., 2021) and a phase I/II trial (NCT02500381) is estimated to be completed in 2025.

A mix of PMOs have been tested as a strategy to increase applicability. Multi-exon skipping with this strategy, has been achieved by skipping exons 45-55 in myotubes

derived from DMD patient fibroblasts (J. Lee et al., 2018). Systemic safety and efficacy of early multi-exon skipping was assessed in dystrophic dog neonates (with a mutation in exon 6). An intra-venous treatment with a 4-PMO cocktail was administered and resulted in 3-27% in-frame skipping of exons 6-9. Dystrophin was restored across skeletal muscles up to 14% of healthy levels. However, no dystrophin rescue was detected in the heart. After 7-8 weeks, treatment led to significant improvement in the standing test and there was no toxicity observed (Lim et al., 2019). This study was the first to demonstrate significant functional improvement by multi-exon skipping in dystrophic dogs. Furthermore, in a recent proof-of-concept study, peptide-conjugated PMOs were tested and achieved exon 45-55 skipping in immortalized patient myotubes. These PMOs, conjugated to DG9 cell-penetrating peptide, were further tested and showed skipping in hDMDdel52;*mdx* mice, restoring dystrophin from 2.8% to 3.9% of wild-type levels (Lim et al., 2022). Low levels of dystrophin recovery seem to be a challenge for these multi-exon strategies.

A new class of AONs made of tricyclo-DNA (tcDNA) have shown interesting results. Tc-DNA deviates from natural DNA by the addition of three carbon atoms, that result in increased RNA affinity, hydrophobicity and nuclease resistance (Renneberg et al., 2002). Systemic delivery of tcDNA-AONs targeting exon 23 showed dystrophin expression rescue of 20-30% in skeletal muscles and 50% in the heart in *mdx* mice (Goyenvalle et al., 2015). A later study evaluated efficacy and toxicology of this 13-mer tcDNA in *mdx* mice. Systemic delivery of the treatment resulted in dystrophin restoration in skeletal

muscles and to a lesser extent in the brain. Furthermore, treatment showed only a slight variation in toxicity biomarkers levels analysed, demonstrating an encouraging safety profile (Relizani et al., 2017).

Exon skipping with AAV viral vectors carrying modified AONs has been explored. Persistent exon skipping was achieved by a single dose administration of an AAV vector expressing an AON linked to a modified U7 small nuclear RNA in *mdx* mice. *Mdx* mice injected intramuscularly at the TA muscle, showed dystrophin restoration of 3% of normal levels 2 weeks after treatment. While a group of *mdx* mice that received treatment by intra-arterial perfusion of the lower limbs showed 80% positive fibres in most muscles of the perfused leg a month after treatment (Goyenvalle et al., 2004). A later study used a set of optimised U7snRNAs carrying AONs in AAV1 vectors to treat dystrophic golden retriever (GRMD) dogs. Results showed sustained correction of the phenotype in muscles and partial muscle strength recovery 4 months after treatment. A 5-year follow-up was done, and dystrophin positive fibres were detected at 2, 6, 18 and 56 months. However, there was a progressive decline of expression leading to an 8-fold decrease after 5 years. This led to the conclusion that recurrent treatments would be required to maintain therapeutic benefits (Vulin et al., 2012).

Other approaches being investigated to enhance delivery and increase efficacy of the skipping include conjugating AONs to molecules, such as various peptides derivates or dendrimeric octa-guanidine (Vivo-morpholino). In addition to previously discussed SRP-

5051, other PMOs have been conjugated to cell penetrating peptides (CPPs). A PMO targeting exons 6 and 8 of *DMD* conjugated to arginine-rich CPP (R-Ahx-R)$_4$ showed more efficiency than non-conjugated PMO and 2-OME-PS AON, in primary muscle cells isolated from a golden retriever muscular dystrophy dog (McClorey, Moulton, et al., 2006). A PMO targeting exon 18 to restore reading frame conjugated to CPP R_6Pen showed consistent levels of exon 18 skipping at day 14 after treatment in Del3-17 DMD human tissue explants (McClorey, Fall, et al., 2006). A PMO targeting exon 23 conjugated to CPP (RXR)$_4$XB led to dystrophin expression of 100% in the diaphragm and 3-8% in limb muscles in *mdx* mice (Fletcher et al., 2007). A Vivo-morpholino PMO showed a significantly increase in delivery compared to the unmodified PMO, leading to 50% and 10% dystrophin expression of normal levels in skeletal and cardiac muscles in *mdx* mice (Bo et al., 2009).

An alternative strategy attempted with exon skipping aims to increase muscle mass to counteract muscle wasting in DMD by knock down of myostatin, a negative regulator of skeletal muscle mass. A proof-of-principle study showed skipping of exon 2 of myostatin, which led to an out-of-phase splicing of exons 1 and 3 to knock down myostatin. A 2-OME-PS AON injected intramuscularly induced exon skipping but did not affect myostatin activity. In the same study, a PMO targeting the same sequence showed efficient skipping *in-vitro*. This PMO was then conjugated to an ocatguanidinie omiety (Vivo-PMO) and tested by systemic tail vein injection, which led to an significant increase in muscle mass of the soleous muscle in normal mice (Kang et al., 2011). In another

study, a PMO also targeting exon 2 of myostatin was conjugated to an arginine-rich cell-penetrating peptide (B-PMO). This B-PMO reached approximately 70% skipping after 4 weeks of treatment (based on densitometric analysis of RT-PCRs) and results showed a significant increase in muscle mass in *mdx* mice after 4 weeks of treatment. Furthermore, this B-PMO was co-administered with a B-PMO targeting dystrophin exon 23 and showed no detrimental interaction, showing potential for a dual antisense combination therapy (Malerba et al., 2012).

Although various exon skipping strategies have reached the market, these strategies still have the disadvantages of being mutation-specific, as different mutations would require skipping of different exons (Aartsma-Rus et al., 2009) and repeated administration is required (Duan et al., 2021).

1.3.3.2. GENE ADDITION.

Gene addition strategies aim to restore missing dystrophin by delivering a functional copy of *DMD* cDNA to affected tissues. These strategies generally use viral vectors as delivery systems. The most promising vectors are the recombinant adeno-associated virus (rAAV)-based vectors as serotypes AAV6, 8 and 9 show high muscle tropism, they have potential to be safe and lead to a long-term effect (Seto et al., 2014, Duan et al., 2021).

In addition, rAAV (referred to as AAV) vectors can persist in muscle cells for several years (Zincarelli et al., 2008) making them a convenient delivery method to treat muscular dystrophies. The main limitation of AAV vectors it the capacity of DNA they can package. To address this, abbreviated functional versions of the *DMD* gene had been made and successfully packaged into AAV vectors (Duan, 2018). These mini and micro-dystrophin genes were developed by deleting parts of the rod domain of the *DMD* gene, while retaining the most crucial domains and regions relevant for structural flexibility and stability of the expressed protein (Harper et al., 2002).

Research on micro-dystrophins dates back to 1990, when a highly functional truncated dystrophin (Δ17-48) was identified in a family presenting very mild BMD allowing a patient to be ambulant at the age of 61 and a second patient to be a body builder at the age of 25 (England et al., 1990). This discovery led to the development of synthetic micro-dystrophin of less than 4kb, such as the first published ΔDysM3 micro-gene encoding for the N-terminal domain, hinges 1, a single spectrin-like repeat, hinge 4, the cysteine-rich domain and the C-terminal domain (Yuasa et al., 1997). However, this micro-dystrophin did not show any effects on dystrophic phenotype of *mdx* mice (Takeda, 2001). The first functional micro-dystrophins were published by Wang et al., (2000); in this study a series of mini-dystrophin <4.2kb were created, driven by a muscle-specific promoter and delivered with AAVs into *mdx* mice. Two of the mini genes (named Δ3849 and Δ3990 which retains hinge 3) restored missing dystrophin and ameliorated dystrophic pathology in the muscle. Results in this study indicate that five rods and two

hinges were sufficient to provide length and flexibility to the central rod domain (B. Wang et al., 2000). These micro-dystrophins were followed by publication of MD1 (R4-R23/ΔCT), which resulted from the analysis of another series of micro-dystrophins. MD1 showed to be to most efficient from this series and reversed histopathological features of *mdx* mice (Harper et al., 2002).

Potency of micro-dystrophins was then significantly improved by codon optimization. Micro-dystrophin cDNA sequences were optimised to improve mRNA stability and translation efficiency by including a consensus Kozac sequence (Kozak, 2005) for optimal translation and by optimising codon usage to maintain a more stable and ordered mRNA secondary structure (Angellotti et al., 2007). This led to modification of 63% of codons in micro-dystrophin1. Furthermore, GC content was increased from 45% to 60%. Non-codon optimised micro-dystrophin1 delivered by AAV vector was compared to the codon optimised version in *mdx* mice. The latter showed increased number of dystrophin positive fibres (approximately from 0 in untreated samples to 230 positive fibres in treated samples within a field of myocardium), statistically significant improved muscle function (from 30% maximal force in untreated *mdx* mice to 90% in treated *mdx* mice) and amelioration of dystrophic pathology (Foster et al., 2008). This led to the establishment of protocols to optimise micro-dystrophins cDNA, which can be done by online tools using algorithms that generate optimised variants of a sequence in an evolutionary approach, with the following parameters: removal of introns, knockout of cryptic splice sites and RNA destabilizing sequence elements, adaption of codon usage,

extensive mutagenesis, flexible combination of functional domains, introduction of restriction sites, epitope shuffling and consideration of immune modulatory CpG motifs (Athanasopoulos et al., 2011).

Potency of micro-dystrophins was further improved with the inclusion of the neuronal nitric oxide synthase (nNOS) domain, in spectrin-like repeats 16 and 17 (Lai et al., 2009, Hakim et al., 2017). Hitherto, more than 30 micro-dystrophin configurations have been published and are reviewed by Duan (2018). Noteworthy ones based on (Duan, 2018) comparative analysis, given the extensive safety and efficacy in canine animal models (Kornegay et al., 2010, Shin et al., 2013, Yue et al., 2015, Le Guiner et al., 2017, Hakim et al., 2017), including previously discussed Δ3990, MD1 and µDys-5R (including the nNOS domain) are represented in Fig. 1.7 alongside the first micro-dystrophin ΔDysM3 and full-length dystrophin.

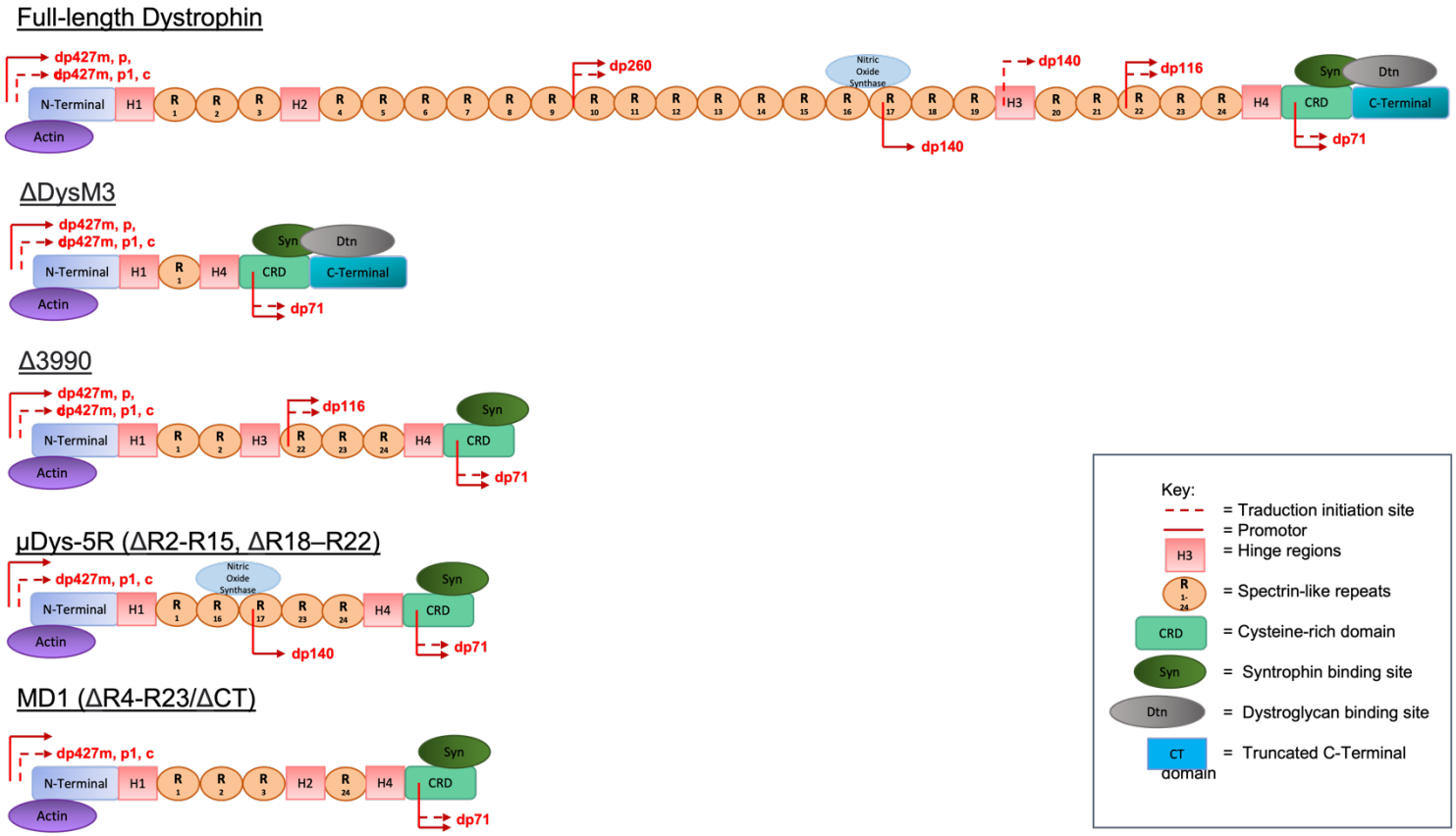


Figure 1.7. Full length dystrophin and representative micro-dystrophins. The proteins schematic shows a variety of dystrophin variants highlighting dystrophin domains and isoform promotors. Full-length dystrophin contains actin bound N-terminal domain, 24 spectrin-like repeats, four hinges, a cysteine-rich domain and a C-Terminal domain. ΔDysM3 mini-dystrophin, the firs synthetic micro-dystrophin contains one rod repeat and two hinges (Yuasa et al., 1997). Δ3990 mini-dystrophin (Wang et al., 2000), μDys-5R (ΔR2-R15, ΔR18-R22) (Hakim et al., 2017) and MD1 (R4-R23/ΔCT) (Harper et al., 2002) have been in clinical trials. These micro-dystrophins present common features, such a as n-terminal domain, cystein-rich domain, spectrin-like repeats 1 and 24, hinges 1 and 4. Differences are in central hinges and the nNOS domain present only in μDys-5R. Figure adapted from (Duan, 2018).

Improved versions are still being developed, nevertheless some effectiveness has been shown in mice models. Two studies delivered rAAV6 via intravascular administration carrying a micro-dystrophin and restored expression of dystrophin in respiratory, cardiac and limb musculature in mice (Gregorevic et al., 2006, Gregorevic et al., 2008). In a different study it was shown that mini- and micro-dystrophins are expressed and prevent fragmentation and loss of postsynaptic folds at the neuromuscular junction (NMJ), which are characteristic of impaired NMJ in *mdx* mice, as well as muscle degeneration (Banks et al., 2008, Banks et al., 2009). It was also shown that in truncated dystrophins, replacing hinge 2 with hinge 3 lead to better protection of skeletal muscles, larger muscle fibres and normal junction. This is explained by a polyproline site in hinge 2 that causes structural abnormalities when there is a highly truncated rod domain (Banks et al., 2010).

The first successful systemic delivery of micro-dystrophin with a modified AAV9 in adult dystrophic dog models was achieved in 2015, showing amelioration of muscle pathology and proving safety and effectiveness (Yue et al., 2015). Another study in canine animal model confirmed safety and durability of canine MD1 delivered in rAAV2/8 and demonstrated dose-dependent improvement (Le Guiner et al., 2017). Recently, results from a blinded, placebo-controlled 90-day study in dystrophic dogs were published. Canine micro-dystrophin-5 was administered with AAV9 and showed dose-dependent increase in micro-dystrophin expression in muscles including the heart and diaphragm.

This led to functional changes such as less impairment in respiratory muscles (Birch et al., 2023).

Alternative approaches to single AAV delivery, are: co-delivery of a truncated dystrophin divided in two or more rAAV vectors that would reconstitute in the muscle cells by protein trans-splicing (Li et al., 2008) or to deliver a high functional mini-dystrophin by co-infection of two independent vectors sharing a central homologous recombinogenic region or overlapping portions of the target gene. One vector providing the promoter with part of the mini/micro-dystrophin and the second one providing the remaining mini/micro-dystrophin and the polyadenylation signal; they would then be reconstituted by homologous recombination (Odom et al., 2011). The delivery of a full lenght *DMD* gene was achieved using a triple-AAV vector system and expression was achieved by trans-splicing events conjoining the three vectors, nevertheless the efficiency was low, as only 4.1% of total fibres expressed the exogenous dystrophin (Koo et al., 2014). Due to low efficiency and the need to inject 3 vectors in patients for this approach, translation to clinic is unlikely.

The first trial in human patients started on 2006, using micro-dystrophin Δ3990 delivered with AAV-2.5 (Bowles et al., 2012). However, levels of micro-dystrophin form this trial were not sufficient for a therapeutic effect. Some suggested reasons for the negative outcome included an immune response to the micro-dystrophin or the viral capsid (Duan, 2018).

Since then, there have been a few trials aiming to establish safety and efficiency: Solid Biosciences trial of micro-dystrophin SGT-001 driven by a muscle-specific promoter delivered in AAV9 vectors (NCT03368742), Sarepta Therapeutics trial of micro-dystrophin SRP-9001 (Delandistrogene Moxeparvovec) driven by MHCK7 muscle specific promoter an delivered with rAAV.rh74 vectors (NCT03375164) and Pfizer trial of micro-dystrophin PF-06939926 driven by a human muscle-specific promoter delivered with AAV9 vectors (NCT03362502).

Solid Biosciences recently shared a press release (Solid Biosciences, 2022) presenting positive one-year data from the IGNITE DMD Phase I/II clinical trial of its micro-dystrophin (SGT-001). Results at 1 year post treatment showed stabilization or improvement of motor function, pulmonary function and patient reported outcomes. Furthermore, micro-dystrophin expression was confirmed from patient biopsies. Three out of 9 patients presented severe adverse effects due to complement activation a few weeks after first dosage, but effects were resolved with no sequelae (Dreghici et al., 2022).

Sarepta presented 3-year safety results from Delandistrogene moxeparvovec (SRP-9001), showing improvements in functional measures over 3 years and long-term

acceptable safety with no severe adverse effects, clinical trial will carry on (J. Mendell et al., 2022).

Unfortunately, Pfizer´s clinical trial with candidate PF-069399206 (Dadistrogene movaparvovec) was put on hold by the FDA due to the death of a participant who presented hypovolemia and cardiogenic shock. The outcome is still being investigated (Philippidis, 2022a, Philippidis, 2022b).

One of the biggest hurdles of these strategies is that they can cause an immune response to the viral vector. Some attempts to ameliorate immune response include modifying the vectors capsids, for example: a chimeric AAV capsid variant (AAV2.5) was used to deliver a micro-dystrophin in a phase I clinical trial diminishing immune response to the vector (Bowles et al., 2012) and engineered muscle-tropic AAV capsid variants (MyoAAVs) were created via directed evolution (Tabebordbar et al., 2021). However, immune response against the foreign micro-dystrophins remains an obstacle to overcome. Furthermore, manufacturing of these therapies is another challenge. Although AAV-based vectors can be produced on an industrial scale (Wright, 2008), production and purification of good manufacturing practice (GMP) grade vectors for trials and commercialization can be difficult and is very expensive (Clément & Grieger, 2016, Kotin & Snyder, 2017), contributing to debates in elevated pricing issues of these therapies and how healthcare systems or patients will afford them (Brennan & Wilson, 2014).

1.3.3.3. GENOME EDITING.

Beside gene addition with mini- and micro-dystrophins and exon-skipping techniques, genome editing has become a promising approach to treat DMD that has evolved alongside genome editing tools.

A genome editing approach using oligonucleotide vectors has been investigated. Introduction of oligonucletides that contain one or more mismatches can activate the innate cellular repair mechanisms and induce a desired correction (Cole-Strauss et al., 1996). Self-pairing, chimeric RNA/DNA oligonucleotides have been shown to induce single base alteration correcting a point mutation in *mdx* mice, confirmed by wild-type dystrophin expression in treated mice. However, dystrophin positive fibres only reached 1-2% of total fibres (Rando et al., 2000). Furthermore, oligodeoxynucleotides (ODNs) have been used to induce gene repair and correct point mutations in the *DMD* gene. Oligonucleotide-mediated repair was demonstrated *in-vitro* and *in-vivo* in mdx^{5CV} mice. However, gene correction efficiency only reached 0.2-5% when determined by immunoblot analysis and quantitative RT-PCR (Bertoni et al., 2005).

To enhance efficacy, studies in DMD models have been done using targeting-specific meganucleases (Chapdelaine et al., 2010), that led to correction of human patient myoblasts using a target-specific meganuclease (MN) and a homologous recombination repair matrix. In this study, the MN was designed to target intron 44, upstream of a

deletion hotspot and was packaged into an integration-competent lentiviral vector. A homologous repair matrix carrying exons 45-52 was packaged into an integration-deficient lentiviral vector. Both vectors were co-transduced in DMD myoblasts carrying a deletion of exons 45 to 52. Results showed expression of full-length, correctly spliced wild-type dystrophin mRNA containing exons 45-52. However, it was not possible to demonstrate that corrected mRNA led to dystrophin protein expression. This study demonstrated that knock-in of missing exons can be achieved by homologous recombination but highlighted the low frequency of correction by the HDR pathway (Popplewell et al., 2013).

ZFNs have also been used in the DMD context. A study showed excision of exon 51 using ZFNs, which led to dystrophin expression in DMD patient myoblasts. A clonal edited population was isolated and transplanted into immunodeficient mice, which resulted in modest human dystrophin expression at the sarcolemma membrane (Ousterout, Kabadi, Thakore, Perez-Pinera, et al., 2015). Another study used engineered artificial zinc finger transcription factors (ZF-ATFs) to upregulate utrophin by targeting its “A” promoter in *mdx* mice. Results showed remarkable amelioration of the *mdx* phenotype (Pisani et al., 2018).

TALENs have been used to knock-in missing exon 44 in patient derived iPSCs. In this study, CRISPR/Cas9 was tested alongside for the same knock-in, both approaches restored full-length dystrophin expression and had a similar activity (H. L. Li et al., 2014).

However, is it relevant to note that to develop this study, 16 pairs of TALENs had to be screened, while only 5 gRNA were screened to find optimal candidates.

A proof-of-concept study used CRISPR/Cas9 to correct a germline of *mdx* mice, that produced genetically mosaic animals with 2-100% correction of mutations (Long et al., 2014). Another proof-of-concept study to support feasibility and efficacy of *in-vivo* genome editing to correct frame-disrupting mutations in *Dmd* was performed on *mdx* mice with a CRISPR-*Sa*Cas9 system targeting exon 23, delivered by an AAV vector. Successful exon removal and restoration of reading frame were achieved and led to protein expression in skeletal and cardiac muscles, resulting in partial recovery of function in dystrophic muscles (Tabebordbar, Zhu, Cheng, Widrick, et al., 2016).

It has also been shown, that single-stranded oligodeoxynucleotides (ssODNs) work as repair templates combined with CRISPR/Cas9 systems to induce HDR in a zebra fish model (Boel et al., 2018). This method has been used to correct the C-to-T mutations within *Dmd* exon 23 and restore dystrophin in *mdx* mice by HDR. A ssODN template was used with a *Lb*Cpf1 system. The ORF of the mouse *Dmd* gene was successfully restored and some of the characteristics of the dystrophic phenotype were rescued, such as fibrosis and inflammatory infiltration (Zhang et al., 2017).

In a different study, removal of exons 52-53 was preformed to restore the open reading frame (ORF) in *mdx*^{4CV} mouse model using two strategies: single AAV6 vector delivery of *Sa*Cas9 and dual AAV6 vector delivery of *Sp*Cas9, both with a muscle-specific cassette containing their respective gRNAs targeting introns flanking exon 52-53. This approach resulted in a deletion of approximately 45 kb of genomic DNA, hence successful removal of exons 52-53, with both single and dual vector approaches that induced dystrophin expression in similar levels, 0.8-18.6% and 1.5-22.9% of wild type dystrophin levels respectively. The second strategy involved delivering a DNA template with a homology region alongside, to allow potential homology-directed repair (HDR). An induction of HDR-mediated *DMD* gene correction was achieved in a fraction of myogenic cells in dystrophic muscles with the addition of the repair template, which led to 1.8-8.4% levels of dytrophin compared to wild type dystrophin. Results from both strategies showed dystrophin expression in skeletal and cardiac muscles resulting in increased force generation (Bengtsson et al., 2017).

Previously mentioned strategies with CRISPR systems and additional studies in cells and *mdx* mice are summarised on Table 1.5

Table 1.5. Overview of pre-clinical genome editing therapeutic strategies with CRISPR systems for treating DMD. Adapted from (Salmaninejad et al., 2021).

Therapy/Application	Model	References
Exon 44-55 skipping with CRISPR/Cas9	Humanized dystrophic mice (Del44)	(Young et al., 2017)
Exon 45-52 skipping with CRISPR/Cas9	DMD-derived muscle cells	(Maggio et al., 2016)
Exon 51 skipping by disruption of splice acceptor with CRISPR/Cpf1	iPSC and *mdx* mice	(Y. Zhang et al., 2017)
Exon 50-54 skipping by CRISPR/Cas9 induced deletion	Myoblasts and hDMD/*mdx* mice	(Iyombe-Engembe et al., 2016)
Exon 44 knock-in with TALEN and CRISPR/Cas9	iPSCs and fibroblasts	(H. L. Li et al., 2014)
Exon 23 skipping with CRISPR/*Sp*Cas9	*Mdx* mice	(Long et al., 2016), (Long et al., 2014), (Nelson et al., 2016), (Tabebordbar, Zhu, Cheng, Chew, et al., 2016)
Exon 23 skipping with CRISPR/*Sa*Cas9 and AAV	*Mdx/Utr*$^{+/-}$ mice	(El Refaey et al., 2017), (Hanson et al., 2022)
Exon 20-23 skipping with *Sa*Cas9 and AAVrh.74 vector led to life-long genome editing	*Mdx* mice	(Xu et al., 2016), (Xu et al., 2019)
Exon 3-9, 6-9, 7-11 skipping	iPSC	(Goyenvalle et al., 2015)
Exon 52-53 skipping with CRISPR/(*Sp* and *Sa*)Cas9 and AAV6	*Mdx*^{4cv} mice	(Bengtsson et al., 2017)
Exon 2 duplication skipping	Patient-derived myogenic cells	(Lattanzi et al., 2017)
Deletion of exons 45-55 with multiplex CRISPR/*Sp*Cas9 system	Human DMD myoblasts and hiPSCs	(Ousterout, Kabadi, Thakore, Majoros, et al., 2015) (Young et al., 2016)
Utrophin upregulation with a modified CRISPR/Cas9 and removal of a duplications of exons 18-30 in *DMD* with CRISPR/Cas9 in a lentiviral vector	Myoblasts from DMD patient	(Wojtal et al., 2016)
Skip of a frame-shifting deletion at exons 51 and germline editing with CRISPR/Cpf1	iPSCs and *mdx* mice	(Y. Zhang et al., 2017)

Deletion of exons 47-58 to restore reading frame and dystrophin expression	Patient derived myoblasts and del52h*DMD/mdx* mice	(Duchêne et al., 2018)

Additional research is being done with other CRISPR systems such as base editing, that would be applicable to 30% of DMD patients harbouring point mutations (Bladen et al., 2015). Base editing has been used to skip exon 23 by interrupting the splicing acceptor site in *mdx* mice (Ryu et al., 2018). Skipping of exon 51 was achieved in iPSC by targeting donor splice site (Chemello et al., 2021). Recently, a DMD hiPSC line was generated by deleting exons 48-54 with CRISPR/Cas9. Cells were derived into cardiomyocytes, that retained the dystrophin disruption of exons 48-54, and a base editor targeting the splice acceptor enabled skipping of exon 55 and restored dystrophin expression. In the same study, gRNAs targeting splice sites of exons 6, 7, 8, 43, 4, 46 and 53 induced exon skipping in DMD hiPSC-derived cardiomyocytes (P. Wang et al., 2023). The disadvantage of base editing strategies is that they would need specific gRNAs developed for each individual base to be edited.

There have also been advances in larger animal models, dystrophin expression was restored in a canine model missing exon 50. AAV9 vectors were used to deliver a CRISPR-*Sp*Cas9 system targeting exon 51 splice acceptor site to knock out exon 51, leading to an in-frame deletion. This resulted in dystrophin levels restoration to almost normal in some muscles and improved muscle histology (Amoasii et al., 2018).

Based on all these advances it can be concluded that some CRISPR genome editing strategies have potential to eventually enter clinical phase. However, certain challenges need to be overcome, such as increasing editing efficiency, avoiding off-target events (Happi Mbakam et al., 2022) and avoiding immune responses to viral vectors (Verdera et al., 2020, Weber, 2021) or Cas proteins (Simhadri et al., 2018, Charlesworth et al., 2018, Crudele & Chamberlain, 2018). Furthermore, it must be highlighted that there is a lack of cell division in skeletal muscle (Alberts et al., 2002) and therefore a need to rely on NHEJ repair pathways rather than HDR, difficulting strategies that require a repair template.

1.4. PROJECT OBJECTIVES & HYPOTHESIS.

1.4.1. RESEARCH PROJECT SCOPE.

Current genome editing strategies to treat DMD have limited patient applicability due to the targeting of specific mutations in certain exons. Alternative strategies being developed, like exon skipping and gene augmentation therapies, would involve repeated administration or could lead to an immune response of patients. This leaves us with the need to develop a more efficient treatment, applicable to a higher number of patients and that would require a single dose to show a beneficial effect on patients.

The aim of this project is to develop a genome editing strategy with CRISPR to delete *DMD* mutational hotspots and achieve the expression of a truncated dystrophin with near 100% functionality, that would potentially mimic micro-dystrophins that are performing well in clinical trials. This genome editing strategy would reduce the need for repeated administrations as the correction of the gene would be permanent, therefore the effect should persist for longer than strategies like exons skipping and gene augmentation therapies, and would have a high DMD patient applicability.

To decide which introns to target in this research project, two things were considered: the deletion had to remove as many mutational hotspots as possible and it had to be in-frame, so a potentially functional truncated dystrophin could be expressed. Deletion of exons 19 to 55 would result in an in-frame deletion that would eliminate mutational

hotspot of exons 45-55 and mutations related to exons 19 to 44, that account for ~81% of total *DMD* mutations (65% of mutations located in mutational hotspot of exons 45-55 (Béroud et al., 2007) plus 20.7% of mutations within exons 19 to 45 calculated from data of 2898 mutations registered on (*The DMD Mutations Database*, n.d.)).

An S*a*Cas9 system was selected for this project, as this system could be packaged into an AAV vector alongside both gRNAs required to achieve the deletion.

The selected CRISPR system would target intron-18 and intron-55, producing a near 800 kbp in-frame deletion. Since the constructs are designed to target introns, the guide RNAs (gRNAs) should have no detrimental effect on DNA coding regions or splice sites of flanking exons, 18 and 56.

1.4.2. HYPOTHESIS AND AIMS.

The hypothesis to be tested is that by using a CRISPR-*Sa*Cas9 system to produce an in-frame large deletion of exons 19-55, the *DMD* gene would repair itself through NHEJ after the double strand break, generating a *de novo* junction between introns 18 and 55, and the edited *DMD* gene would be capable of expressing a functional truncated dystrophin protein, as illustrated in Figure 1.8.

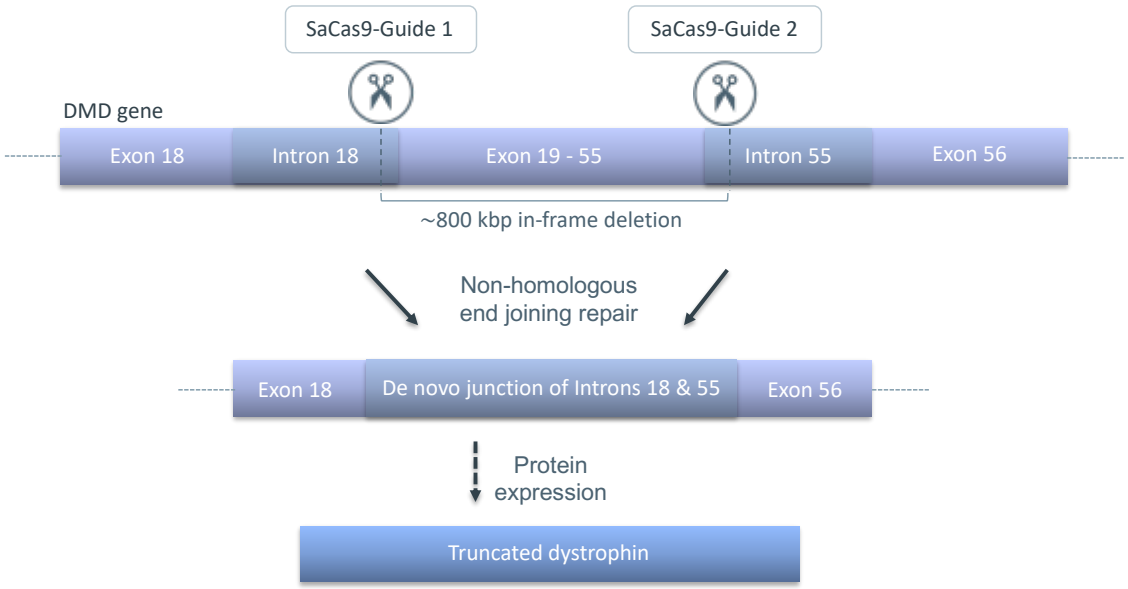


Figure 1.8. Genome editing deletion strategy with an *Sa*Cas9 system targeting introns 18 and 55 of the *DMD* gene. An in-frame deletion of exons 19 to 55 would produce a truncated functional dystrophin.

To develop a proof-of-principle and test the hypothesis the following research plan and aims were defined:

- Perform an *in-silico* analysis, modelling the protein that would be expressed after the in-frame deletion to evaluate its potential functionality.
- Validate *in-silico* findings by designing a positive control expressing the truncated Del19-55 dystrophin cDNA and test it by transfection on appropriate cell lines, then test the positive control *in-vivo* in *mdx* mice.
- Compare CRISPR/Cas systems and select the most appropriate one for the project aim. Due to the size of the Cas protein a *Staphylococcus aureus (Sa)*Cas9 system was picked, as the size of the Cas protein would allow packaging into an AAV vector.

- Design gRNAs for an *Sa*Cas9 system targeting introns 18 and 55 of the *DMD/Dmd* gene to human and mouse sequences.
- Establish an *Sa*Cas9 system. Perform appropriate dose responses. Then clone all gRNAs and screen *in-vitro* for cleavage efficiency by transfection, DNA harvest, PCR amplification, sequencing, and TIDE Analysis.
- Multiplex the most efficient gRNAs and test new constructs on appropriate cell lines.
- Produce AAV vectors with *Sa*Cas9 CRISPR constructs with multiplexed gRNAs and test *in-vivo* in *mdx* mice.

2. MATERIALS AND METHODS.

2.1. BIOINFORMATICS.

2.1.1. DYSTROPHIN PROTEIN SEQUENCES FOR IN-SILICO ANALYSIS.

The full-length dystrophin protein sequence was obtained for human and mouse versions from the Ensembl database (https://www.ensembl.org/index.html) with the following specifications:

For mouse:

Gene of interest: Dmd

Species: Mouse GRCm39

Transcript ID: ENSMUST00000114000.8

Protein length: 3678 aa

For human:

Gene of interest: DMD

Species: Human GRCh38.913

Transcript ID: ENST00000357033.9

Protein length: 3685 aa

Full length mouse dystrophin amino acid sequence:

MLWWEEVEDCYEREDVQKKTFTKWVNAQFSKFGKQHIENLFSDLQDGRRLLDLLEGLTGQKLPKEKGSTRVHALNNV
NKALRVLQNNNVDLVNIGSTDIVDGNHKLTLGLIWNIILHWQVKNVMKNIMAGLQQTNSEKILLSWVRQSTRNYPQV
NVINFTTSWSDGLALNALIHSHRPDLFDWNSVVCQQSATQRLEHAFNIARYQLGIEKLLDPEDVDTTYPDKKSILMYITS
LFQVLPQQVSIEAIQEVEMLPRPPKVTKEEHFQLHHQMHYSQQITVSLAQGYERTSSPKPRFKSYAYTQAAYVTTSDPT
RSPFPSQHLEAPEDKSFGSSLMESEVNLDRYQTALEEVLSWLLSAEDTLQAQGEISNDVEVVKDQFHTHEGYMMDLTA
HQGRVGNILQLGSKLIGTGKLSEDEETEVQEQMNLLNSRWECLRVASMEKQSNLHRVLMDLQNQKLKELNDWLTKTE
ERTRKMEEEPLGPDLEDLKRQVQQHKVLQEDLEQEQVRVNSLTHMVVVVDESSGDHATAALEEQLKVLGDRWANIC
RWTEDRWVLLQDILLKWQRLTEEQCLFSAWLSEKEDAVNKIHTTGFKDQNEMLSSLQKLAVLKADLEKKKQSMGKLYS
LKQDLLSTLKNKSVTQKTEAWLDNFARCWDNLVQKLEKSTAQISQAVTTTQPSLTQTTVMETVTTVTTREQILVKHAQ
EELPPPPPQKKRQITVDSEIRKRLDVDITELHSWITRSEAVLQSPEFAIFRKEGNFSDLKEKVNAIEREKAEKFRKLQDASRS
AQALVEQMVNEGVNADSIKQASEQLNSRWIEFCQLLSERLNWLEYQNNIIAFYNQLQQLEQMTTTAENWLKIQPTTP
SEPTAIKSQLKICKDEVNRLSDLQPQIERLKIQSIALKEKGQGPMFLDADFVAFTNHFKQVFSDVQAREKELQTIFDTLPP
MRYQETMSAIRTWVQQSETKLSIPQLSVTDYEIMEQRLGELQALQSSLQEQQSGLYYLSTTVKEMSKKAPSEISRKYQSE
FEEIEGRWKKLSSQLVEHCQKLEEQMNKLRKIQNHIQTLKKWMAEVDVFLKEEWPALGDSEILKKQLKQCRLLVSDIQTI
QPSLNSVNEGGQKIKNEAEPEFASRLETELKELNTQWDHMCQQVYARKEALKGGLEKTVSLQKDLSEMHEWMTQAE
EEYLERDFEYKTPDELQKAVEEMKRAKEEAQQKEAKVKLLTESVNSVIAQAPPVAQEALKKELETLTTNYQWLCTRLNG
KCKTLEEVWACWHELLSYLEKANKWLNEVEFKLKTTENIPGGAEEISEVLDSLENLMRHSEDNPNQIRILAQTLTDGGV
MDELINEELETFNSRWRELHEEAVRRQKLLEQSIQSAQETEKSLHLIQESLTFIDKQLAAYIADKVDAAQMPQEAQKIQS
DLTSHEISLEEMKKHNQGKEAAQRVLSQIDVAQKKLQDVSMKFRLFQKPANFEQRLQESKMILDEVKMHLPALETKSV
EQEVVQSQLNHCVNLYKSLSEVKSEVEMVIKTGRQIVQKKQTENPKELDERVTALKLHYNELGAKVTERKQQLEKCLKLS
RKMRKEMNVLTEWLAATDMELTKRSAVEGMPSNLDSEVAWGKATQKEIEKQKVHLKSITEVGEALKTVLGKKETLVE
DKLSLLNSNWIAVTSRAEEWLNLLLEYQKHMETFDQNVDHITKWIIQADTLLDESEKKKPQQKEDVLKRLKAELNDIRPK
VDSTRDQAANLMANRGDHCRKLVEPQISELNHRFAAISHRIKTGKASIPLKELEQFNSDIQKLLEPLEAEIQQGVNLKEED
FNKDMNEDNEGTVKELLQRGDNLQQRITDERKREEIKIKQQLLQTKHNALKDLRSQRRKKALEISHQWYQYKRQADDL
LKCLDDIEKKLASLPEPRDERKIKEIDRELQKKKEELNAVRRQAEGLSEDGAAMAVEPTQIQLSKRWREIESKFAQFRRLN
FAQIHTVREETMMVMTEDMPLEISYVPSTYLTEITHVSQALLEVEQLLNAPDLCAKDFEDLFKQEESLKNIKDSLQQSSG
RIDIIHSKKTAALQSATPVERVKLQEALSQLDFQWEKVNKMYKDRQGRFDRSVEKWRRFHYDIKIFNQWLTEAEQFLRK
TQIPENWEHAKYKWYLKELQDGIGQRQTVVRTLNATGEEIIQQSSKTDASILQEKLGSLNLRWQEVCKQLSDRKKRLEE
QKNILSEFQRDLNEFVLWLEEADNIASIPLEPGKEQQLKEKLEQVKLLVEELPLRQGILKQLNETGGPVLVSAPISPEEQDK
LENKLKQTNLQWIKVSRALPEKQGEIEAQIKDLGQLEKKLEDLEEQLNHLLLWLSPIRNQLEIYNQPNQEGPFDVKETEIA
VQAKQPDVEEILSKGQHLYKEKPATQPVKRKLEDLSSEWKAVNRLLQELRAKQPDLAPGLTTIGASPTQTVTLVTQPVV
TKETAISKLEMPSSLMLEVPALADFNRAWTELTDWLSLLDQVIKSQRVMVGDLEDINEMIIKQKATMQDLEQRRPQLE
ELITAAQNLKNKTSNQEARTIITDRIERIQNQWDEVQEHLQNRRQQLNEMLKDSTQWLEAKEEAEQVLGQARAKLES
WKEGPYTVDAIQKKITETKQLAKDLRQWQTNVDVANDLALKLLRDYSADDTRKVHMITENINASWRSIHKRVSEREAA
LEETHRLLQQFPLDLEKFLAWLTEAETTANVLQDATRKERLLEDSKGVKELMKQWQDLQGEIEAHTDVYHNLDENSQK
ILRSLEGSDDAVLLQRRLDNMNFKWSELRKKSLNIRSHLEASSDQWKRLHLSLQELLVWLQLKDDELSRQAPIGGDFPA
VQKQNDVHRAFKRELKTKEPVIMSTLETVRIFLTEQPLEGLEKLYQEPRELPPEERAQNVTRLLRKQAEEVNTEWEKLNL
HSADWQRKIDETLERLRELQEATDELDLKLRQAEVIKGSWQPVGDLLIDSLQDHLEKVKALRGEIAPLKENVSHVNDLAR
QLTTLGIQLSPYNLSTLEDLNTRWKLLQVAVEDRVRQLHEAHRDFGPASQHFLSTSVQGPWERAISPNKVPYYINHETQ
TTCWDHPKMTELYQSLADLNNVRFSAYRTAMKLRRLQKALCLDLLSLSAACDALDQHNLKQNDQPMDILQIINCLTTIY
DRLEQEHNNLVNVPLCVDMCLNWLLNVYDTGRTGRIRVLSFKTGIISLCKAHLEDKYRYLFKQVASSTGFCDQRRLGLLL
HDSIQIPRQLGEVASFGGSNIEPSVRSCFQFANNKPEIEAALFLDWMRLEPQSMVWLPVLHRVAAAETAKHQAKCNIC
KECPIIGFRYRSLKHFNYDICQSCFFSGRVAKGHKMHYPMVEYCTPTTSGEDVRDFAKVLKNKFRTKRYFAKHPRMGYL
PVQTVLEGDNMETPVTLINFWPVDSAPASSPQLSHDDTHSRIEHYASRLAEMENSNGSYLNDSISPNESIDDEHLLIQHY
CQSLNQDSPLSQPRSPAQILISLESEERGELERILADLEEENRNLQAEYDRLKQQHEHKGLSPLPSPPEMMPTSPQSPRD
AELIAEAKLLRQHKGRLEARMQILEDHNKQLESQLHRLRQLLEQPQAEAKVNGTTVSSPSTSLQRSDSSQPMLLRVVGS
QTSDSMGEEDLLSPPQDTSTGLEEVMEQLNNSFPSSRGRNTPGKPMREDTM

Full length human dystrophin amino acid sequence:

MLWWEEVEDCYEREDVQKKTFTKWVNAQFSKFGKQHIENLFSDLQDGRRLLDLLEGLTGQKLPKEKGSTRVHALNNV
NKALRVLQNNNVDLVNIGSTDIVDGNHKLTLGLIWNIILHWQVKNVMKNIMAGLQQTNSEKILLSWVRQSTRNYPQV
NVINFTTSWSDGLALNALIHSHRPDLFDWNSVVCQQSATQRLEHAFNIARYQLGIEKLLDPEDVDTTYPDKKSILMYITS
LFQVLPQQVSIEAIQEVEMLPRPPKVTKEEHFQLHHQMHYSQQITVSLAQGYERTSSPKPRFKSYAYTQAAYVTTSDPT
RSPFPSQHLEAPEDKSFGSSLMESEVNLDRYQTALEEVLSWLLSAEDTLQAQGEISNDVEVVKDQFHTHEGYMMDLTA
HQGRVGNILQLGSKLIGTGKLSEDEETEVQEQMNLLNSRWECLRVASMEKQSNLHRVLMDLQNQKLKELNDWLTKTE
ERTRKMEEEPLGPDLEDLKRQVQQHKVLQEDLEQEQVRVNSLTHMVVVVDESSGDHATAALEEQLKVLGDRWANIC
RWTEDRWVLLQDILLKWQRLTEEQCLFSAWLSEKEDAVNKIHTTGFKDQNEMLSSLQKLAVLKADLEKKKQSMGKLYS
LKQDLLSTLKNKSVTQKTEAWLDNFARCWDNLVQKLEKSTAQISQAVTTTQPSLTQTTVMETVTTVTTREQILVKHAQ
EELPPPPPQKKRQITVDSEIRKRLDVDITELHSWITRSEAVLQSPEFAIFRKEGNFSDLKEKVNAIEREKAEKFRKLQDASRS
AQALVEQMVNEGVNADSIKQASEQLNSRWIEFCQLLSERLNWLEYQNNIIAFYNQLQQLEQMTTTAENWLKIQPTTP
SEPTAIKSQLKICKDEVNRLSDLQPQIERLKIQSIALKEKGQGPMFLDADFVAFTNHFKQVFSDVQAREKELQTIFDTLPP
MRYQETMSAIRTWVQQSETKLSIPQLSVTDYEIMEQRLGELQALQSSLQEQQSGLYYLSTTVKEMSKKAPSEISRKYQSE
FEEIEGRWKKLSSQLVEHCQKLEEQMNKLRKIQNHIQTLKKWMAEVDVFLKEEWPALGDSEILKKQLKQCRLLVSDIQTI
QPSLNSVNEGGQKIKNEAEPEFASRLETELKELNTQWDHMCQQVYARKEALKGGLEKTVSLQKDLSEMHEWMTQAE
EEYLERDFEYKTPDELQKAVEEMKRAKEEAQQKEAKVKLLTESVNSVIAQAPPVAQEALKKELETLTTNYQWLCTRLNG
KCKTLEEVWACWHELLSYLEKANKWLNEVEFKLKTTENIPGGAEEISEVLDSLENLMRHSEDNPNQIRILAQTLTDGGV
MDELINEELETFNSRWRELHEEAVRRQKLLEQSIQSAQETEKSLHLIQESLTFIDKQLAAYIADKVDAAQMPQEAQKIQS
DLTSHEISLEEMKKHNQGKEAAQRVLSQIDVAQKKLQDVSMKFRLFQKPANFEQRLQESKMILDEVKMHLPALETKSV
EQEVVQSQLNHCVNLYKSLSEVKSEVEMVIKTGRQIVQKKQTENPKELDERVTALKLHYNELGAKVTERKQQLEKCLKLS
RKMRKEMNVLTEWLAATDMELTKRSAVEGMPSNLDSEVAWGKATQKEIEKQKVHLKSITEVGEALKTVLGKKETLVE
DKLSLLNSNWIAVTSRAEEWLNLLLEYQKHMETFDQNVDHITKWIIQADTLLDESEKKKPQQKEDVLKRLKAELNDIRPK
VDSTRDQAANLMANRGDHCRKLVEPQISELNHRFAAISHRIKTGKASIPLKELEQFNSDIQKLLEPLEAEIQQGVNLKEED
FNKDMNEDNEGTVKELLQRGDNLQQRITDERKREEIKIKQQLLQTKHNALKDLRSQRRKKALEISHQWYQYKRQADDL
LKCLDDIEKKLASLPEPRDERKIKEIDRELQKKKEELNAVRRQAEGLSEDGAAMAVEPTQIQLSKRWREIESKFAQFRRLN
FAQIHTVREETMMVMTEDMPLEISYVPSTYLTEITHVSQALLEVEQLLNAPDLCAKDFEDLFKQEESLKNIKDSLQQSSG
RIDIIHSKKTAALQSATPVERVKLQEALSQLDFQWEKVNKMYKDRQGRFDRSVEKWRRFHYDIKIFNQWLTEAEQFLRK
TQIPENWEHAKYKWYLKELQDGIGQRQTVVRTLNATGEEIIQQSSKTDASILQEKLGSLNLRWQEVCKQLSDRKKRLEE
QKNILSEFQRDLNEFVLWLEEADNIASIPLEPGKEQQLKEKLEQVKLLVEELPLRQGILKQLNETGGPVLVSAPISPEEQDK
LENKLKQTNLQWIKVSRALPEKQGEIEAQIKDLGQLEKKLEDLEEQLNHLLLWLSPIRNQLEIYNQPNQEGPFDVKETEIA
VQAKQPDVEEILSKGQHLYKEKPATQPVKRKLEDLSSEWKAVNRLLQELRAKQPDLAPGLTTIGASPTQTVTLVTQPVV
TKETAISKLEMPSSLMLEVPALADFNRAWTELTDWLSLLDQVIKSQRVMVGDLEDINEMIIKQKATMQDLEQRRPQLE
ELITAAQNLKNKTSNQEARTIITDRIERIQNQWDEVQEHLQNRRQQLNEMLKDSTQWLEAKEEAEQVLGQARAKLES
WKEGPYTVDAIQKKITETKQLAKDLRQWQTNVDVANDLALKLLRDYSADDTRKVHMITENINASWRSIHKRVSEREAA
LEETHRLLQQFPLDLEKFLAWLTEAETTANVLQDATRKERLLEDSKGVKELMKQWQDLQGEIEAHTDVYHNLDENSQK
ILRSLEGSDDAVLLQRRLDNMNFKWSELRKKSLNIRSHLEASSDQWKRLHLSLQELLVWLQLKDDELSRQAPIGGDFPA
VQKQNDVHRAFKRELKTKEPVIMSTLETVRIFLTEQPLEGLEKLYQEPRELPPEERAQNVTRLLRKQAEEVNTEWEKLNL
HSADWQRKIDETLERLRELQEATDELDLKLRQAEVIKGSWQPVGDLLIDSLQDHLEKVKALRGEIAPLKENVSHVNDLAR
QLTTLGIQLSPYNLSTLEDLNTRWKLLQVAVEDRVRQLHEAHRDFGPASQHFLSTSVQGPWERAISPNKVPYYINHETQ
TTCWDHPKMTELYQSLADLNNVRFSAYRTAMKLRRLQKALCLDLLSLSAACDALDQHNLKQNDQPMDILQIINCLTTIY
DRLEQEHNNLVNVPLCVDMCLNWLLNVYDTGRTGRIRVLSFKTGIISLCKAHLEDKYRYLFKQVASSTGFCDQRRLGLLL
HDSIQIPRQLGEVASFGGSNIEPSVRSCFQFANNKPEIEAALFLDWMRLEPQSMVWLPVLHRVAAAETAKHQAKCNIC
KECPIIGFRYRSLKHFNYDICQSCFFSGRVAKGHKMHYPMVEYCTPTTSGEDVRDFAKVLKNKFRTKRYFAKHPRMGYL
PVQTVLEGDNMETPVTLINFWPVDSAPASSPQLSHDDTHSRIEHYASRLAEMENSNGSYLNDSISPNESIDDEHLLIQHY
CQSLNQDSPLSQPRSPAQILISLESEERGELERILADLEEENRNLQAEYDRLKQQHEHKGLSPLPSPPEMMPTSPQSPRD
AELIAEAKLLRQHKGRLEARMQILEDHNKQLESQLHRLRQLLEQPQAEAKVNGTTVSSPSTSLQRSDSSQPMLLRVVGS
QTSDSMGEEDLLSPPQDTSTGLEEVMEQLNNSFPSSRGRNTPGKPMREDTM

Then, anticipated translation of Del19-55 dystrophin was obtained from Expasy (*ExPASy - Translate Tool*, n.d.) by inputting hDel19-55 cDNA sequence, the following aminoacidic sequence was obtained for *in-silico* protein analysis:

MLWWEEVEDCYEREDVQKKTFTKWVNAQFSKFGKQHIENLFSDLQDGRRLLDLLEGLTGQKLPKEKGSTRVHALNNV
NKALRVLQNNNVDLVNIGSTDIVDGNHKLTLGLIWNIILHWQVKNVMKNIMAGLQQTNSEKILLSWVRQSTRNYPQV
NVINFTTSWSDGLALNALIHSHRPDLFDWNSVVCQQSATQRLEHAFNIARYQLGIEKLLDPEDVDTTYPDKKSILMYITS
LFQVLPQQVSIEAIQEVEMLPRPPKVTKEEHFQLHHQMHYSQQITVSLAQGYERTSSPKPRFKSYAYTQAAYVTTSDPT
RSPFPSQHLEAPEDKSFGSSLMESEVNLDRYQTALEEVLSWLLSAEDTLQAQGEISNDVEVVKDQFHTHEGYMMDLTA
HQGRVGNILQLGSKLIGTGKLSEDEETEVQEQMNLLNSRWECLRVASMEKQSNLHRVLMDLQNQKLKELNDWLTKTE
ERTRKMEEEPLGPDLEDLKRQVQQHKVLQEDLEQEQVRVNSLTHMVVVVDESSGDHATAALEEQLKVLGDRWANIC
RWTEDRWVLLQDILLKWQRLTEEQCLFSAWLSEKEDAVNKIHTTGFKDQNEMLSSLQKLAVLKADLEKKKQSMGKLYS
LKQDLLSTLKNKSVTQKTEAWLDNFARCWDNLVQKLEKSTAQISQAVTTTQPSLTQTTVMETVTTVTTREQILVKHAQ
EELPPPPPQKKRQITVDSEIRKRLDVDITELHSWITRSEAVLQSPEFAIFRKEGNFSDLKEKVNDLQGEIEAHTDVYHNLDE
NSQKILRSLEGSDDAVLLQRRLDNMNFKWSELRKKSLNIRSHLEASSDQWKRLHLSLQELLVWLQLKDDELSRQAPIGG
DFPAVQKQNDVHRAFKRELKTKEPVIMSTLETVRIFLTEQPLEGLEKLYQEPRELPPEERAQNVTRLLRKQAEEVNTEWE
KLNLHSADWQRKIDETLERLQELQEATDELDLKLRQAEVIKGSWQPVGDLLIDSLQDHLEKVKALRGEIAPLKENVSHVN
DLARQLTTLGIQLSPYNLSTLEDLNTRWKLLQVAVEDRVRQLHEAHRDFGPASQHFLSTSVQGPWERAISPNKVPYYIN
HETQTTCWDHPKMTELYQSLADLNNVRFSAYRTAMKLRRLQKALCLDLLSLSAACDALDQHNLKQNDQPMDILQIINC
LTTIYDRLEQEHNNLVNVPLCVDMCLNWLLNVYDTGRTGRIRVLSFKTGIISLCKAHLEDKYRYLFKQVASSTGFCDQRR
LGLLLHDSIQIPRQLGEVASFGGSNIEPSVRSCFQFANNKPEIEAALFLDWMRLEPQSMVWLPVLHRVAAAETAKHQA
KCNICKECPIIGFRYRSLKHFNYDICQSCFFSGRVAKGHKMHYPMVEYCTPTTSGEDVRDFAKVLKNKFRTKRYFAKHPR
MGYLPVQTVLEGDNMETPVTLINFWPVDSAPASSPQLSHDDTHSRIEHYASRLAEMENSNGSYLNDSISPNESIDDEHL
LIQHYCQSLNQDSPLSQPRSPAQILISLESEERGELERILADLEEENRNLQAEYDRLKQQHEHKGLSPLPSPPEMMPTSPQ
SPRDAELIAEAKLLRQHKGRLEARMQILEDHNKQLESQLHRLRQLLEQPQAEAKVNGTTVSSPSTSLQRSDSSQPMLLR
VVGSQTSDSMGEEDLLSPPQDTSTGLEEVMEQLNNSFPSSRGRNTPGKPMREDTM

2.1.2. Protein analysis on Phyre2 Software.

PHYRE2 Protein Fold Recognition Server (Kelley et al., 2015) (http://www.sbg.bio.ic.ac.uk/~phyre2/html/page.cgi?id=index) was used to predict a protein model of Del19-55 truncated dystrophin based on its aminoacidic sequence.

2.1.3. GUIDE RNA DESIGN AND SCORING.

Guide RNAs targeting introns 18 and 55 for human and mouse *DMD* genes were designed using the following Software: Benchling (https://benchling.com), CRISPOR (http://crispor.tefor.net/) and The BROAD Institute Portal (https://portals.broadinstitute.org/gpp/public/analysis-tools/sgrna-design). All guide RNAs were designed to target intronic regions. Selection criteria included selecting guide RNAs that were outputs in at least two of the three softwares in use and with the higher scores for on- and off-targets. The selected guide RNAs were synthetized by IDT (https://eu.idtdna.com) and used for guide RNA cloning.

On- and off-targeting scores are defined as following: On-target scores refer to the activity or predicted efficiency of the gRNAs according to an algorithm designed by (Doench et al., 2014). Off-target scores refer to specificity of the gRNAs according to an algorithm designed by (Hsu et al., 2013). Higher MIT specificity scores indicate lower off-target effects. This score has been adapted for *Sa*Cas9 and based on the off-target scores shown on mouse-over. This algorithm by (Tycko et al., 2018b) is aggregated from all off-target scores and ranges 0-100. Finally, higher predicted efficiency scores indicate more likely cleavage at this position. This algorithm is a modified version of the Doench et al. (2016) score by (Najm et al., 2018) for *Sa*Cas9, with a range from 0-100.

2.2. GENERAL LABORATORY REAGENTS.

General reagents were purchased from Sigma, Invitrogen or VWR with standard chemically purity graded as analytical reagents for analysis applications (AnalaR). All reagents were dissolved in double distilled water (ddH_2O), unless stated otherwise. Solution used in tissue culture were autoclaved at 121ºC for 15 minutes, with exception of solutions containing proteins, detergents or glucose. These were filter sterilised with a 0.22 μm filter when needed. All solutions were stored at room temperature unless stated otherwise.

List of reagents and manufacturer:

- Acetic acid (CH3COOH) – VWR
- Agarose (molecular grade) – Invitrogen
- Foetal calf serum – Invitrogen
- Dimethyl sulphoxide (DMSO) - Sigma
- EDTA – Sigma
- Ethanol – VWR
- Glucose – Sigma
- Liquid broth (LB) – Sigma
- Methanol – Sigma
- Paraformaldehyde (PFA) – Sigma
- Potassium chloride (KCl) – Sigma

- Phosphate buffered saline (PBS) pH 7.3 – Gibco
- Sodium dodecylsulfate (SDS) – Sigma
- Trizma hydrochloride (HCl) – Sigma

The following kits were purchased from QIAGEN and manufacturer´s protocols were followed:

- DNA extraction: DNeasy Blood & Tissue Kit
- Gel extraction: QIAquick Gel Extraction Kit
- PCR purification: QIAquick PCR Purification Kit
- Mini-preps: QIAprep Spin Miniprep Kit
- Maxi-preps and giga-preps: EndoFree Plasmid Kit (RNAse free)
- RNA extraction: RNeasy kit

2.3. DNA CLONING AND ANALYSIS.

2.3.1. MATERIALS FOR BACTERIAL CULTURES AND MOLECULAR CLONING.

- Ampicillin from Sigma: stock prepared as 1000x in ddH_2O at 50 mg/mL concentration, filter sterilised (0.22 μm filter) and stored at -20ºC
- Lysogeny broth (LB) from Invitrogen
- LB agar from Invitrogen (10 gr Peptone, 5 gr Yeast Extract, 5 gr Sodium Chloride and 12 gr Agar)
- LB SOC medium (100 μ 1M $MgSO_4$ and 20 μL 1M Glucose)
- 100 mM $MgCl_2$ from BDH
- 100 mM $CaCl_2$ from Sigma
- Top 10 *E. coli* competent cells from NEB
- 85 mM $CaCl_2$/15% Glycerol
- Restriction enzymes and buffers from NEB
- 50X TAE: 242 gr Tris Base, 57.1 mL Glacial Acetic Acid, 200 mL of 0.5M EDTA pH 8 brought to a total volume of 1L with ddH_2O
- 5X loading dye from Bioline
- 1000X SYBR Safe DNA gel stain from Invitrogen
- DNA molecular weight markers: Hyperladder I, IV and V from Bioline
- T4 DNA ligase and buffer from Promega

2.3.2. Preparing chemically competent cells for cloning.

The protocol to prepare chemically competent Top 10 *E. Coli* to transform is a 3-day long protocol. On day one, LB agar was prepared and autoclaved with no antibiotics. Then, Top 10 *E. Coli* competent cells from a stock stored at -80ºC were streaked on 100 mm petri dishes with solidified LB agar and incubated overnight at 37ºC.

The next day the plate was stored at -4ºC while LB broth (20 gr/L) was being autoclaved. A colony from the plate was picked and used to inoculate 5 mL of LB broth without antibiotics. Sample was incubated at 200 rpm and 37ºC overnight.

On the third day, 2.5 mL of the starter culture were used to inoculate a second starter culture of 250 mL of LB without antibiotics. The culture was incubated at 37ºC and 200 rpm until the Optic Density (O.D.) at 260 nm was between 0.2-0.5 ODU. The cells were poured into a pre-chilled 250 mL centrifuge tube and the pellet was resuspended in 100 mL of pre-chilled 100 mM $MgCl_2$. The cells were centrifuged at 3,273 x g for 15 min at 4ºC. The pellet was resuspended in 50 mL of 100 mM $CaCl_2$ and incubated on ice for 20 min. The cells were centrifuged again at 3,273 x g for 15 min at 4ºC. The pellet was resuspended on 10 mL of ice-cold 85 mM $CaCl_2$/15% Glycerol. Cells were aliquoted in 250 μL in pre-chilled Eppendorf tubes on a -20ºC mini-cooler (alternatively dry ice) and stored at -80ºC.

2.3.3. Cloning.

Cloning protocols consisted of the following steps: vector preparation, a preparative restriction digestion, DNA ligation from CRISPR guide RNAs or oligonucleotides and vector backbone, bacterial transformation, cells counting and colony picking, plasmid miniprep, diagnostic restriction digestions and sequencing.

2.3.4. Plasmids.

The following plasmids were used for cloning:

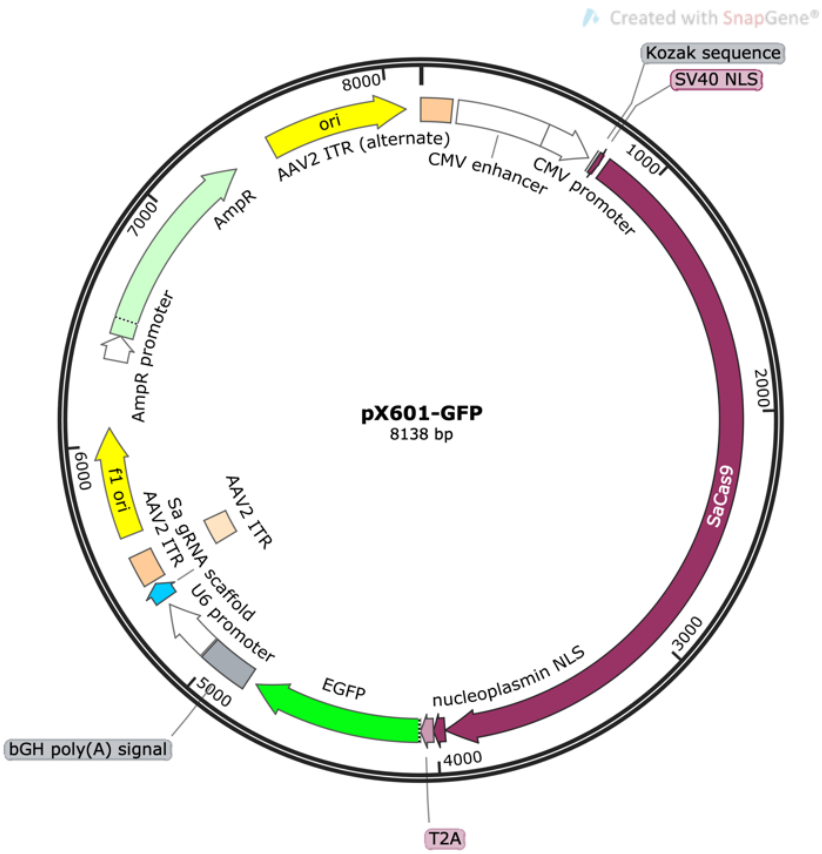


Figure 2.1. pX601-GFP, referred to as pX601-CMV-SaCas9-GFP. Plasmid with an AAV backbone, containing *Sa*Cas9 and a GFP marker driven by a CMV promoter. Plasmid was a gift from Yuet Wai Kan (Ye et al., 2016) (Addgene plasmid #84040).

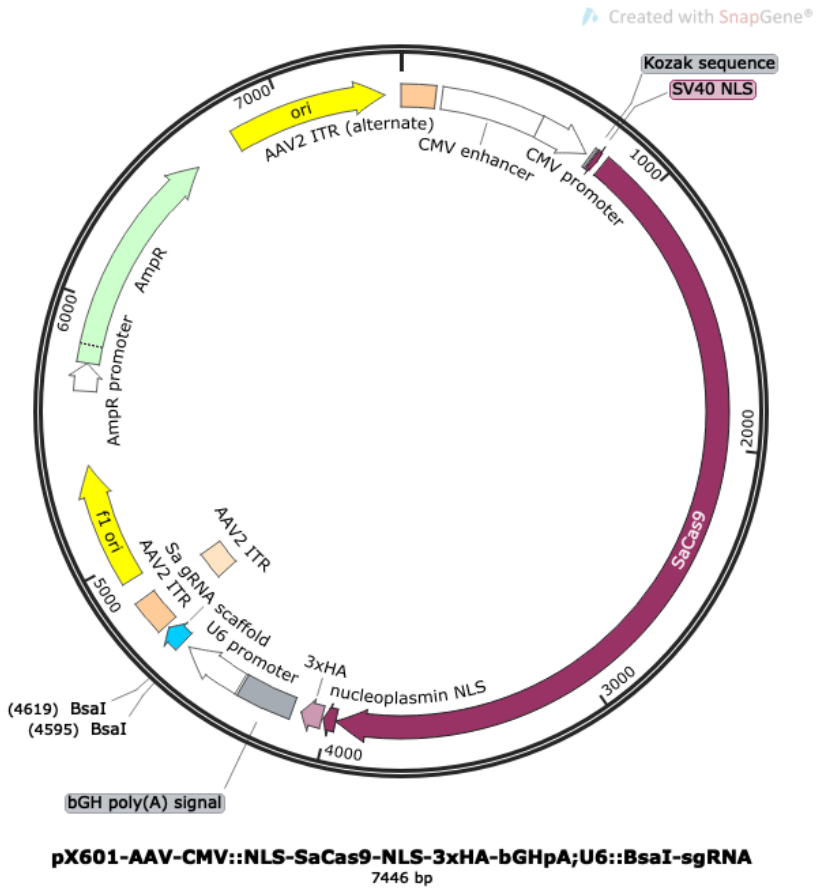


Figure 2.2. pX601-AAV-CMV::NLS-SaCas9-NLS-3xHA-bGHpA;U6::BsaI-sgRNA, referred to as pAAV-CMV-SaCas9. AAV backbone plasmid containing *Sa*Cas9 driven by a CMV promoter used to deliver CRISPR Guide RNAs. Plasmid was a gift from Feng Zhan (Ran et al., 2015) (Addgene plasmid #61591).

2.3.5. Vector preparation.

LB agar plates were prepared by autoclaving LB with agar in a concentration of 35 gr/L, then 1 µL of filtered ampicillin (1000X) per 1 mL of media was added before plating 25 - 27 mL of media per petri dish. Once the plate cooled down, the vector was taken from Glycerol stocks stored at -80ºC and streaked on the plates. The plates were incubated

overnight at 32ºC for 18-22 hrs. at 200 rpm, as indicated for AAV based vectors to avoid ITR mutations with higher temperatures.

2.3.6. Bacterial plasmid Miniprep protocol.

The miniprep protocol was performed to purify plasmid DNA from bacteria. A day prior to the protocol, colonies were picked into 5 μL LB broth containing the appropriate selective antibiotic and grown overnight in a shaking incubator at 37ºC or at 32ºC when working with AAV plasmids.

The culture was transferred to a 5 mL tube, then centrifuged for 15 minutes at 3,273 x g at 4ºC to pellet the bacteria and the supernatant was discarded. The pellet was resuspended in 250 μL of buffer P1 from QIAGEN – QIAprep Spin Miniprep Kit. 250 μL of buffer P2 was added to lyse the cells. Then, 350 μL of buffer N3 were added and samples were centrifuged at 12,470 x g for 10 minutes at room temperature. Supernatant (with plasmid DNA) was kept in a new Eppendorf tube and the pellet was discarded, since all the proteins and cell debris remain there, 800 μL were transferred to the spin column by pipetting and the column was centrifuged at 12,470 x g at room temperature for 60 seconds. The solution that came off the column was discarded. The column was washed with 500 μL of buffer PB and the column was centrifuged for another 60 seconds. Supernatant was discarded thoroughly, and the column washed with 750 μL of buffer PE and centrifuged for 60 seconds. The column was transferred to

a microcentrifuge tube and DNA was eluted with 50 μL of Buffer EB by centrifuging for 1 minute. Concentration of clean DNA was measured using the Nanodrop with an absorbance of 260 nm and samples were then stored at -20ºC.

2.3.7. SEQUENCING.

DNA samples were sent at 50-100 ng DNA/μL to Eurofins for sequencing with appropriate primers previously designed flanking the region of interest, using Eurofins SeqPrimer Design Tool (https://www.eurofinsgenomics.eu/en/ecom/tools/sequencing-primer-design/). For sequencing of PCR products, either the forward or reverse PCR primer was used to sequence its respective amplicon. The list of primers used to confirm gRNA insertion and plasmid integrity can be found in Table 2.1. It is important to consider that sequencing results from Eurofins cover a 700 bp long region without errors. Sequences longer than that are prone to show errors. While designing the primers, it is important to consider that the first 50 bp of the sequence, next to the primer, will not be clean, so the target region needs to be at least 50 bp downstream of the primer annealing site and within a region shorter than 700 bp.

Table 2.1. Sequencing primers to confirm correct guide RNA cloning and plasmids integrity. Primers designed on Eurofins SeqPrimer Design Tools.

Primer name (target)	Sequence (5´to 3´)
pAAV-SaCas9 Guides FW (to confirm guide RNA insertion)	CCGAGGGCCTATTTCCCATGATTC
pAAV Ampicilin Site FW	CTATGTGGCGCGGTATTATCC
pAAV Ampicilin Site RV	TTGCAAGCAGCAGATTACGC
Exon18-56 in pCI-CMV-Del19-55-hDys-GFP FW	AATGGAAACAGTAACTACGGTG
Exon18-56 in pAAV-spc512-Del19-55-hDys-GFP RV	AATACCGGTACAGCATGGTGGCGAAT
Spc512 promoter in pAAV-Spc512-Del19-55-hDys-GFP RV	TCATAACAGTCCTCTACTTCTTCC

2.3.8. Bacterial plasmid Maxiprep protocol.

Once the plasmid was analysed by sequencing or restriction digestion and confirmed to be as expected, it was maxiprepped to have clean DNA in high concentrations to perform other experiments such as transfections.

A day prior to the protocol, colonies were picked into 5 mL LB broth containing the appropriate selective antibiotic and grown for 8 hrs in a shaking incubator at 250 rpm and 37ºC or at 32ºC when working with AAV plasmids. Then, 500 μL of the starter culture were transferred to a second culture of 250 mL LB media with 1 μL/ml of the appropriate selective antibiotic (all plasmids in this project have an ampicillin antibiotic resistance cassette) and incubated overnight at 250 rpm.

The culture was transferred to a 250 mL flask, then centrifuged for 30 minutes at 3,273 x g at 4ºC to pellet the bacteria and the supernatant is discarded. The pellet was

resuspended in 10 mL of buffer P1 from EndoFree Plasmid Maxi Kit form from QIAGEN. 10 mL of buffer P2 was added to lyse the cells and mixed 4-6 times by inverting. Then, 10 mL of chilled buffer P3 were added, and samples were mixed by inverting. The lysate was poured into the barrel of the QIAfilter Cartridge and incubated at room temperature for 10 min. The lysate was then filtered into a 50 mL tube. 2.5 mL of ER buffer were added to the filtered lysate, mixed by inverting and incubated on ice for 30 min. The QIAGEN-tip 500 was equilibrated by applying 10 mL of QBT Buffer. The filtered lysate was applied in the tip and allowed to empty by gravity flow. The tip was then washed twice with 30 mL of QC Buffer. DNA was eluted with 15 mL of QN Buffer into a 50 mL tube. DNA was precipitated by adding 10.5 mL of isopropanol to the eluted DNA, then the mix was chilled for 20 min at -20ºC and centrifuged at 15,000 x g at 4ºC for 30 min. In a Laminar Flow Hood, the DNA pellet was washed with 5 mL of endotoxin-free 70% ethanol and centrifuged at 4ºC at 15,000 x g rpm for 10 min. The pellet was then air dried for 5 minutes and left overnight in a suitable volume of TE Buffer. The next day the samples were nanodropped at 260 nm, aliquoted at 1000 ng DNA/μL and stored at -20ºC.

2.3.9. Restriction Digestion.

Restriction digestions were performed as a diagnostic test or as a preparative procedure to obtain a certain DNA fragment that was then used for cloning. To select the appropriate enzymes, each plasmid was analysed on SnapGene and enzymes cutting

among each relevant region of the plasmid were selected to confirm plasmid integrity. For preparative restriction digestions, appropriate enzymes to recover a particular backbone were also selected on SnapGene. All enzymes used and their respective buffers are from NEB.

Samples were then prepared as following:

For 15 μL final volume: 12.5 μL of dH_2O, 1.5 μL 10X Enzyme digest buffer chosen according to preferences of the enzyme, 500 ng of DNA and 0.2 μL of enzyme.

A master mix was prepared by mixing dH_2O and the buffer, cooled on ice for 3 minutes before adding the enzyme and then the mix with the enzyme was added to each sample tube. DNA was added and the digest incubated for 1 to 3 hours at the optimal temperature for the enzyme.

Typically, 15 μL were prepared per sample for an analytical digest and 30 μL for preparative ones. Since after a preparative restriction digestion DNA will be extracted, 1000-4000 ng of DNA were used in these digests to assure enough DNA could be extracted afterwards. Digests were then analysed by gel electrophoresis.

2.3.10. AGAROSE GEL ELECTROPHORESIS.

Agarose gels were used to run DNA fragments (<15 kb) after a restriction digestion and separate them by size. Usually a 1% (w/v) agarose gel with 0.5X SYBR Safe in 1X Tris Borate EDTA (TAE) Buffer was run for 1 to 3 hours at 80-120V depending on the size of the gel. Gels were then visualized in a blue light transilluminator.

2.3.11. DNA EXTRACTION FROM AGAROSE GELS.

Specific fragments were recovered from agarose gels by gel extraction, such as vector backbones for CRISPR Guide RNA cloning. Gels were visualized in a blue light transilluminator and the target fragment was excised with a scalpel to be extracted with the QIAGEN Gel Extraction Kit as following:

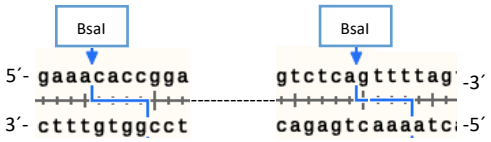
The excised band was weighed and 3 Volumes of Buffer QG to 1 Volume of gel were added. Samples were incubated at 50ºC for 10 minutes and vortexed every 3 minutes to dissolve the gel. It was checked that the buffer was yellow after the incubation to confirm that pH did not need to be adjusted. Then, 1 volume of isopropanol was added and mixed. The sample was transferred to the QIAquick column and centrifuged for 1 minute at 12,470 x g. Then, 500 μL of QG Buffer were added to the column and centrifuged again at the same speed and time. The column was incubated for 5 minutes with 750 μL of PE Buffer and then washed and centrifuged twice. The column was then

transferred to a new Eppendorf tube and DNA was eluted with 50 µL of EB Buffer after incubating the column with the buffer for 5 minutes.

2.3.12. Oligonucleotides annealing for CRISPR gRNA cloning.

CRISPR Cas9 sgRNAs targeting intron 18 and 55 in human and mouse genes were designed using Benchling and CRISPOR Software. Designs are presented in the results section 4.1.1. Before ordering the forward and reverse sequence of each gRNA, overhangs complementary to the cut site of the backbone were added to each strand. For example, Guide 14 targeting mouse intron 18 sgRNA forward sequence is 5´-ACTTTCAGGGAATAACGTAC-3´. Overhangs complementary to BsaI restriction site at the pAAV-CMV-SaCas9 plasmid were added:

BsaI restriction sites in pAAV-CMV-SaCas9 sequence:



Guide 14 oligonucletoide sequence with overhangs (indicated in red) complementary to pAAV-CMV-SaCas9 after digestion with BsaI:

CACCGACTTTCAGGGAATAACGTAC G14-FW

aaacGTACGTTATTCCCTGAAAGTC G14-RV

These overhangs were added to all *Sa*Cas9 gRNAs to be cloned into pAAV-CMV-SaCas9 backbone (digested with BsaI). Oligonucleotides (with respective overhangs) were ordered from IDT. Synthesised oligos were resuspended in dH_2O at 100 µM. To anneal the oligos, 10 µl of each (sense and antisense) were mixed with 5 µl 10x SuRE/Cut Buffer H (Roche; Sigma-Aldrich, St. Louis, MO, USA) and 75 µl DEPC treated water in each tube. Annealing reactions were heated to 95ºC for 10 minutes, then cooled by 2ºC/sec to 85ºC and held at this temperature for 1 minute, then cooled by 0.3ºC/sec to 75ºC and held at this temperature for 1 minute, a further 5 similar rounds of cooling were repeated for every 10ºC to 25ºC; reactions were then cooled to 4ºC and held at this temperature.

The process was followed by a ligation of the vector backbone, as described in the next section, and 2-4 µL of each annealed pair of oligos. Afterwards, ligations were transformed and resulting plasmids were analysed and sent for sequencing, each sample at 100 ng/µL to IDT.

2.3.13. Ligation of DNA fragments.

Ligations were performed with a 2:1 insert to vector ratio. In this protocol, vector backbone was recovered from pAAV-CMV-SaCas9 (digested with BsaI) and annealed oligonucleotides are the CRISPR gRNAs designed with an overhang complementary to BsaI restrictions sites in pAAV-CMV-SaCas9 sequence. Alternatively, backbones can be

ligated to other DNA pieces recovered from a restriction digestion if the same restriction enzyme is used to cut the ends of the fragments that will be ligated, so cut sites between vector and insert are complementary. The final volume of the reaction was set as following:

Table 2.2. Template for ligation reactions of backbone and annealed gRNAs. Amounts for water, 10X ligation buffer, vector, insert (annealed oligo) and T4 ligase indicates for ligations of a positive control, negative control and samples. T4 ligase and buffer from Promega.

Component	Positive Control: uncut Vector (1 in 100 maxiprep dilution); volume to add in µl	Negative control: cut vector; volume to add in µl	For each guide ligation reaction; volume to add in µl
H20	7.00	8.00	7.00
10X Ligation buffer	1.00	1.00	1.00
Vector	1.00	-	Between 25-50ng
Annealed oligo	0.00	0.00	1.00
T4 Ligase	1.00	1.00	1.00
Total	10.00	10.00	10.00

Respective ligation reactions were prepared according to Table 2.2, including T4-DNA Ligase and 10X buffer from Promega. Samples were incubated in the PCR machine with the following Program: 22°C for 1 hour, then 16°C for 10 hours, and then held at 4°C.

The negative control is a control for colonies that are product of vector re-ligation without the insert. The positive control confirms that the vector backbone should express correct antibiotic resistance markers.

2.3.14. BACTERIAL TRANSFORMATION BY HEAT SHOCK.

DNA from minipreps can be used to transform, usually using 1 μL from stock (approximately 100-250 ng of DNA). A ligation reaction can also be used directly to transform, usually 2-4 μL from the reaction are used.

50 μL of Top10 *E. Coli* were added to each DNA sample obtained from ligations and rested in ice for 30 minutes. The samples were transferred to 42°C for 45 seconds ('heat shock') and stacked in ice. 250 μL of SOC media from NEB (49.4 mL LB media, 0.5 mL 1M $MgSO_4$, 0.1 mL 20% (w/v) glucose) was added to each sample and then they were incubated at 37°C with shaking for one hour (or at 32°C AAV vectors), so the cells could grow and express the resistance gene. Samples were plated on LB agar plates with 1 μL of filtered ampicillin per 1 mL and left overnight to generate colonies.

2.3.15. G-BLOCKS RESUSPENSION.

Double stranded blocks of DNA (g-blocks) were designed using SnapGene. The DNA was synthetized by IDT and each sample containing a g-block was resuspended according to IDT specification sheet by centrifuging the tube with the sample to 3000 x g to ensure material was at the bottom. Then TE Buffer was added to each sample to reach a final concentration of 10 ng/μL and the samples were vortexed briefly. Samples were incubated at 50ºC for 20 minutes on a waterbath and vortexed and centrifuged briefly. G-block were then ready for further experiments, such as restriction digests and subsequent ligation and transformation.

2.4. CELL CULTURE.

2.4.1. MATERIALS FOR ADHERENT CELL CULTURE.

- Sterile PBS from Gibco: 1 PBS tablet dissolved in 500 mL ddH2O, autoclaved.
- Trypsin from Sigma (1X diluted in PBS).
- T175 cm^3 tissue culture flasks from Corning.
- DMSO from Sigma.
- Sterile glass stripettes (5, 10, 25 and 50 mL) from Starlabs.
- Falcon tubes (15 and 50 mL) from Corning.
- Screw top vials (15 mL) from Corning.
- Dulbecco modified Eagle medium (DMEM) from Gibco.
- Foetal Calf Serum (FCS) from Sigma.

2.4.2. MAINTENANCE OF ADHERENT CELLS.

2.4.2.1. ADHERENT CELL LINES.

Cells were maintained at 37ºC and 5% CO_2 and seeded on the appropriate density to achieve required confluence within a time of growth and perform different experiments such as transfections, nucleofections or transduction. Each cell line has different requirements, in this project the following adherent cell lines were used to screen gRNAs and test the established *Sa*Cas9 system:

- Human Embryonic Kidney cells (HEK293T) were used to screen gRNAs targeting human introns 18 and 55 of the human *DMD* gene.
- Mouse Albino Neuroblastoma cells (Neuro-2A or N2A) were used to screen gRNAs targeting introns 18 and 55 in the mouse *Dmd* gene.
- Mouse myoblasts (C2C12 cells) were used to test top pair of mouse gRNAs in a multiplexed construct and for transduction of AAV9 vectors.
- H2KB-mdx (*mdx* mouse myoblasts) cells were used for transduction of AAV9 vector.

2.4.2.2. MAINTENANCE CONDITIONS.

HEK293T, N2A and C2C12 cells were maintained in DMEM containing 10% FCS and 1% Penicillin/Streptomycin (prepared media). Depending on how confluent they looked under the microscope, usually aiming for 70-90% confluency, cells were passaged, counted and split twice a week. HEK293, N2A and C2C12 cells were incubated at 37ºC and 5% CO_2.

H2KB-mdx (*mdx* mouse myoblasts) cells were maintained in growth media (DMEM, 20% FCS, 0.5% chicken embryo extract, 20U/mL of interferon gamma from Gibco, 5 mL 1% Penicillin/Streptomycin) at 33ºC and 10% CO_2 and passaged when 60-70% confluent. To differentiate into myotubes, myoblasts were seeded in plates coated in 0.1mg/mL Matrigel, when cells were 80-90% confluent, growth media was changed to

differentiation media (DMEM, 10% horse serum, 0.5% chicken embryo extract, 20U/mL interferon gamma, 1% Penicillin/Streptomycin) and incubation conditions changed to 37ºC and 5% CO_2.

2.4.2.3. PASSAGING/SPLITTING.

All cell lines were passaged/split by sucking culture media out with a stripette, washing cells gently with PBS, adding 4 mL of 1X trypsin and incubating for 2-4 minutes to detach cells from the flask. 16 mL of prepared media were added to inactivate trypsin. Cells were then transferred to a 50 mL Falcon tube and centrifuged at 500 x g for 5 minutes at room temperature in a Beckmann coulter centrifuge. Cell pellets were resuspended in 10 mL of media. Cells were counted using an hemocytometer (Neubauer camera) and $1x10^6$ cells were seeded on each T175 Flasks with 25 mL of DMEM containing 10% foetal calf serum and 1% Penicillin/Streptomycin on every split.

2.4.2.4. THAWING CELLS & MAKING A CELL BANK.

To prepare a cell bank, $1x10^6$ cells in 1 mL of 90% FCS and 10% DMSO were aliquoted and frozen in 2 mL cryogenic vials at -80ºC in a cell freezing container (Mr. Frosty). After 1-3 days vials were transferred to liquid Nitrogen storage. When needed, cells were thawed in the waterbath, transferred to a 15 mL Falcon tube and 9 mL of prepared media were added. Cells were centrifuged at 500 x g for 5 minutes at room temperature

in a Beckmann coulter centrifuge. Supernatant was discarded to eliminate DMSO, cell pellet was resuspended in 5 mL of prepared media. Cells were transferred to T175 flask with 25 mL of pre-warmed prepared medium.

2.4.3. TRANSFECTION.

Transfection is the introduction of foreign DNA into cells, this can be achieved by different protocols, for this project Viafect was used in most experiments after being compared to Lipofectamine.

A day before transfection cells were seeded in a 6-well plate with a cell density of $5x10^5$ cells/well in 2 mL of DMEM 10% FBS and 1% Pen/Strep, to achieve a 60-70% confluence in HEKs and N2As before the transfection. It is recommended to first count cells and then prepare a mix of $5x10^5$ cells per 2 mL so the cells can be seeded while adding the 2 mL of media to the each well. This way cells are seeded in a more even way throughout each well and unnecessary shaking of the plate can be avoided, as this can cause cells to accumulate and grow more confluent in the centre of the well.

2.4.3.1. VIAFECT PROTOCOL.

Next day, media was changed to 2 mL of fresh prepared media 1 hour before transfection. Meanwhile, a transfection mix of: Viafect transfection reagent from

Promega and DNA of interest in a 4:1 Viafect to DNA ratio and the amount of Serum-Free DMEM needed to make the volume up to 700 μL per sample was prepared, before transfecting a 6-well plate (each sample transfected by triplicate). After a 20-minute incubation of the mix at room temperature, 200 μL of the mix was added to each well with the cells. Two days after transfection cells were harvested. DMEM media 10% FBS and 1% P/S was pre-warmed at 37ºC. Then media from seeded cells was aspirated, cells were washed with 2 mL PBS, 500 μL of 1X Trypsin Were added and incubated for 2 minutes or until cells had detached. Trypsin was then neutralised with 1.5 mL of supplemented DMEM and cells were transferred to a 15 mL falcon tube. Samples were spun for 5 minutes at 500 x g in a Beckmann coulter centrifuge. Supernatant was aspirated and pellet washed with 2 mL of PBS. Samples were centrifuged again, and the PBS was aspirated again. Samples were then ready to process for DNA extraction or to be stored at -80ºC.

A dose response for each plasmid was performed to find the optimal DNA dose for transfection. In the following table an example of the set up for the experiment can be observed. When running an experiment with different plasmids, the amount of DNA would be the same for all samples.

Table 2.3. Dose response for a transfection with Viafect and plasmid DNA 4:1 on HEK293T or N2A cells. Samples were transfected by triplicates and an extra amount of the mix was prepared to ensure having 200 μL per well. Viafect, plasmid DNA and Serum-free DMEM were added to each mix as calculated on the table.

Plasmid	DNA (ng/μl)	Number of Wells	Amount of DNA/μg	Viafect (μl)	Guide RNA (μl)	DMEM (μl)	Total Volume (μl)
Untreated	-	-	-	-	-	-	-
Mock	-	3.5	**0.0**	84.0	0	616.0	700
pY095	1000	3.5	**4.0**	56.0	14.0	630.0	700
pY095	1000	3.5	**6.0**	84.0	21.0	595.0	700
pY095	1000	3.5	**8.0**	112.0	28.0	560.0	700

2.4.3.2. LIPOFECTAMINE PROTOCOL.

Next day after seeding, media was changed 1 hour before transfection. Meanwhile, two mixes were prepared accordingly as shown on Table 2.6, before transfecting a 6-well plate (each sample transfected by triplicate). Then, Mix 1 and Mix 2 were mixed and incubated for 10 minutes before adding 250 μL of the final mix to each well. Cells were then incubated for 48 hours before harvesting them for further analysis.

Table 2.4. Dose response for a transfection with Lipofectamine and plasmid DNA 3:1 and 3:4 of P3000 on HEK293T cells. Samples were transfected by triplicates and an extra amount of the mix was prepared to ensure having 250 µL per well. Mix 1 contains lipofectamine and serum-free media while Mix 2 contains plasmid DNA, Serum-free DMEM and P300.

				Mix 1		Mix 2		
Condition	**Wells**	**Plasmid Concentration µg**	**Total plasmid µl**	**SF Media (µl)**	**Lipofectamine (µl)**	**SF Media (µl)**	**DNA (µl)**	**P3000 (µl)**
Untreated	3.5	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Lipofectamine 3000	3.5	0.0	0.0	546.8	15.8	423.5	0.0	14.0
CMV-eGFP (8ug)	3.5	1.0	28.0	395.5	42.0	353.5	28.0	56.0
empty pY095 (1ug)	3.5	1.0	3.5	432.3	5.3	427.0	3.5	7.0
empty pY095 (2ug)	3.5	1.0	7.0	427.0	10.5	416.5	7.0	14.0
empty pY095 (4ug)	3.5	1.0	14.0	416.5	21.0	395.5	14.0	28.0
empty pY095 (6ug)	3.5	1.0	21.0	406.0	31.5	374.5	21.0	42.0
empty pY095 (8ug)	3.5	1.0	28.0	395.5	42.0	353.5	28.0	56.0

2.4.4. MYOBLASTS REVERSE TRANSDUCTION AND DIFFERENTIATION TO MYOTUBES.

An hour before seeding, 6-well plates were coated in 0.1mg/mL Matrigel. Plates with Matrigel were incubated for an hour. C2C12 or H2KB-*mdx* cells were seeded with a cell density of $2x10^5$ cells/well for reverse transduction with AAV9 with an MOI of $1x10^6$. AAV vectors were added right after adding cells in suspension to each well. Cells were incubated with growth media at 33ºC and 10% CO_2 for 16-18 hours. Then media was changed to differentiation media and incubation to 37ºC and 5% CO_2. Cells were harvested on day 5 after reverse transduction for DNA and RNA extraction and cells were harvested on day 7 for protein extraction.

2.5. FLUORESCENCE MICROSCOPY.

2.5.1. MATERIALS.

- Zeiss microscope (Zeiss Axio Vision D1 with AxioCam MRm).
- Software ZEN 2012 for image acquisition.
- FIJI Software ("Fiji is just ImageJ" for mac users).

2.5.2. FLUORESCENCE MICROSCOPY OF CELLS AND TA MUSCLE SECTIONS.

Fluorescence microscopy was used in this project for a few experiments, including confirming GFP expression after transfection of cells with plasmids containing a GFP marker and analysing immunohistochemistry samples.

Before using the microscope, it was always confirmed that the fluorescence had not been used in the past half an hour. The microscope´s components were always switched on in the following order: stage controller, microscope and computer. Then the fluorescence box was switched on and then logged in on the computer screen. The microscope was set on 10X/0.25 Magnification Phase 1, 100X Magnification and 2.3 Voltage.

ZENpro Software was used to capture all images, once the program was ready and the plate was placed on the microscope stage, the “acquisition” tab was used on the screen on “live” to focus the cells and adjust the exposure. To adjust exposure and intensity: while observing the Brightfield phase, the intensity was set on 2.8 V and exposure on 10 lux-seconds. Once the image was satisfactorily adjusted, 5 pictures were taken per well and 6 fields were imaged per TA section, in a systematic way (Fig. 2.3) by clicking on “Snap”. All pictures were saved as .czi files and exported as .tiff files (including individual and merged channels). Images were then processed with FIJI Software.

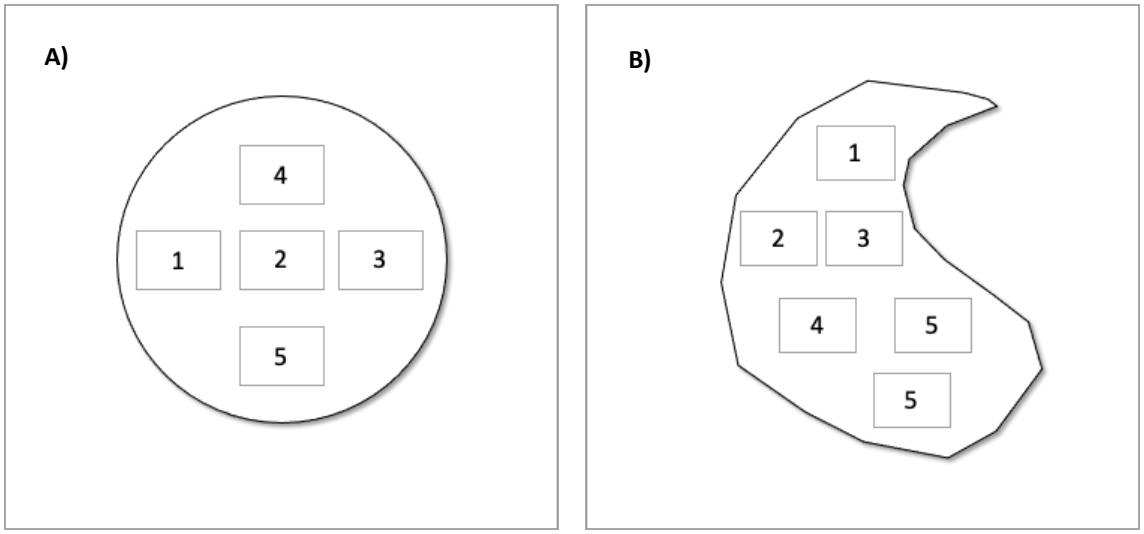


Figure 2.3. Systematic way to image cells in wells and TA sections with fluorescent Zeiss microscope. A) 5 images were acquired per well. B) 6 fields were imaged per TA section.

2.5.2.1. FIJI SOFTWARE.

FIJI Software allows processing of .czi files for Mac users. This tool also allows to arrange images and display them on individual or merged channels.

Once the Software is launched, a .czi file was opened (preferably in the order they wanted to be displayed), then the split channel was selected on the "Image" menu on "type", RGB was selected for both channels. Then on the same "Image" menu on "Colour", channels were merged. Images were then stacked, and the final montage was built for display (as shown on the results section).

2.6. FACS (FLUORESCENCE-ACTIVATED CELL SORTING).

Flow cytometric analysis was used in this project to determine subpopulation of cells expressing GFP. This was used as a proxy to determine transient transfection efficiency when using plasmids expressing a GFP marker, such as pX601-CMV-GFP.

2.6.1. MATERIALS.

- FACS Buffer: prepared by mixing 500 mL of autoclaved PBS, 2 mL of filter sterilized 0.5M EDTA (156.1 gr/L H_2O), 5 mL of 10 % NaN_3 and 10 mL of FCS (foetal calf serum previously thawed).

- Trypsin/EDTA.
- 5 mL round bottom snap cap FACS tubes from BD Falcon.
- 4% paraformaldehyde (PFA) from Sigma.
- CST FACS beads from BD Falcon.
- Clean and rinse solution from BD Falcon.
- FACS Canto II machine from BD Falcon.
- FlowJo Software from BD Falcon.

2.6.2. Cell Harvesting.

When cells were ready for harvesting (2 days after transfection), media was removed, each well was washed with 2 mL PBS and 500 μL of Trypsin/EDTA were added; after 2 minutes Trypsin was neutralized with 2 mL of FACS Buffer and cells were transferred to a 15 mL Falcon tube. Samples were spun at 500 x g for 7 min at room temperature and supernatant was disposed. 200 μL of 4% PFA previously thawed was added to each sample to fix the cells, samples were then vortexed and incubated at room temperature for 20 min. Samples were then washed with 3 mL of FACS Buffer and spun again at 500 x g for 7 min. Supernatant was disposed and cells were resuspended in 200 μL of FACS Buffer. Samples were covered in foil and stored at 4ºC in the cold room for no longer than 48 hours before being analysed.

2.6.3. FACS ANALYSIS.

Before running the samples, the FACS machine needs to be calibrated. The machine was first switched on and then the computer. The first step is to check the machine is connected to all the appropriate buffers and then to launch the FACSDiva Software. Then the fluidic start-up on the Cytometer tab was performed and the machine was calibrated with CST Beads (fluorospheres with stable size and fluorescence intensity) by running a sample of 800 μL of FACS Buffer with one drop of the CST Beads. The outcomes report should have <6% on all parameters.

Once the calibration was finished, a new experiment was set up on the global worksheet and 2 dot plots graphs and a histogram graph were drawn. Dot plot #1 had FSC-A (Forward Scatter Area) on x-axis and SSC-A (Side Scatter Area) on y-axis, Dot plot #2 had FSC-A on x-axis and FSC-H (Forward Scatter Height) on y-axis and the histogram had FITC-A on x-axis and cell count on y-axis. Parameters that were not required, were de-selected on the "Cytometer FACS Control" menu.

The first sample used for initial population gating was a non-transfected sample (mock), this was used as a negative control. Samples were acquired on "low rate" and parameters were adjusted depending on the cell line.

Parameters for HEKs:

Voltage:

FSC - 230

SSC - 370

FITC – 393

Set up threshold – 50,000 events

Parameters for N2As:

Voltage:

FSC - 242

SSC - 357

FITC – 393

Set up threshold – 50,000 events

Once the Mock was acquired, P1 Gate was drawn on Dot plot #1 with the "Polygon Gate" menu, selecting the live population of cells. Then on Dot plot #2 the selected live population was gated again to keep only single cells. On the histogram graph, an interval gate was added and named P3.

The second sample analysed was the positive control; after acquiring the sample P1 Gate was re-adjusted and then the rest of the samples were acquired on medium rate stopping at 50,000 events per sample. All data was stored and then analysed with FlowJo Software.

2.6.4. Analysis with FlowJo Software.

FlowJo Software was initiated and data obtained from the FACS machine was opened, then all graphs (Dot plots #1 and #2 and histogram) were dragged staggered to the "layout editor". Then "live", "single" and "GFP" events were dragged to the "table editor". Once this was done, one of the Mock samples was opened and samples were gated as following:

Dot plot #1: x-axis = FSC-A, y-axis = SSC-A -> Gate all live cells on P1

Population of interest should be in the middle of the graph.

Dot plot #2: x-axis = FSC-A, y-axis = FDC-H -> Gate all single cells on P2

The minimum number of cells needed on this plot are around 3,300 acquired cells.

Histogram: x-axis = FITC-A, y-axis = Histogram -> Gate fluorescent population on P3

Both, the positive control and the mock were displayed in the histogram before drawing P3 Gate. This gate should start were the Mock population ends on the x-axis.

Once all the gating was done and checked on all samples, data was exported as an excel file and analysed on Prism9 (for statistical analysis).

2.7. DNA/RNA EXTRACTION & cDNA SYNTHESIS.

2.7.1. MATERIALS FOR DNA AND RNA EXTRACTION.

- 1X sterile PBS
- Ethanol
- DNeasy Blood & Tissue kit from QIAGEN
- QIAshredder QIAGEN
- RNeasy kit from QIAGEN
- β-mercaptoethanol
- Heat block to 55ºC
- QuantiTect Reverse Transcription kit from QIAGEN

2.7.2. DNA EXTRACTION FROM CELLS.

DNA was harvested from frozen cells pellets using the "DNeasy Blood & Tissue Kit" from QIAGEN. Samples were taken out of -80ºC and pellets were left at room temperature to thaw. Then, in accordance to manufacturer´s protocol, pellets were resuspended with 200 μL of PBS, 20 μl of proteinase K were added, followed by 200 μL of AL Buffer before mixing the sample by vortexing. 200 μL of 100% ethanol were added and mixed by vortexing. The mix was then transferred with a micropipette into a DNeasy Mini spin column placed in a 2 mL collection tube. Samples were centrifuged at 6,000 x g for 1 minute. Flow-through and collection tubes were discarded, and columns were placed in

new 2 mL collection tubes. 500 μL of AW1 Buffer were added to each column and samples were centrifuged again at 6,000 x g for 1 minute. Columns were transferred again to a new collection tube and 500 μL of AW2 were added to each column before centrifuging samples at 20,000 x g for 3 minutes. Columns were then transferred to a new 1.5 mL centrifuge tube and DNA was eluted by adding 100 μL of AE Buffer to each column, incubating samples at room temperature for 1 minute and then centrifuging them at 6,000 x g for minute. DNA samples were then quantified with the nanodrop at 260 nm and stored at -20ºC for further analysis.

2.7.3. DNA EXTRACTION FROM TISSUE.

To extract DNA from tissue, "DNeasy Blood & Tissue Kit" from QIAGEN was used and manufacturer´s protocol was followed. Tissue samples were thawed on ice (from -80ºC), then 20 μL of proteinase K were added per sample (approximately 30 (30 μm) intersections from TA muscle), samples were vortexed and incubated at 56ºC in a heat block until the tissue was lysed. Samples were then vortexed and proceeded following the same protocol used for DNA extraction from cells.

2.7.4. RNA EXTRACTION FROM CELLS.

For RNA extraction, cells were harvested by aspirating culture medium and adding 350 μL of RLT lysis Buffer from QIAGEN. Cell lysate was transferred to a QIAshredder spin

column placed in a 2 mL collection tube. Samples were centrifuged 2 minutes at full speed at room temperature. Then 350 μL of 70% ethanol were added to the flow-though. The total 700 μL of samples were transferred to an RNeasy spin column placed at a 2 mL collection tube and centrifuged for 15 second at 8000 x g. Flow-through was discarded. Then 350 μL of RW1 Buffer were added to the column, samples were centrifuged again for 15 seconds at the same speed and flow-through was discarded. 80 μL of DNAse I incubation mix (10 μL DNase I and 70 μL RDD Buffer from RNase free DNase Set) were added to each column and incubated at room temperature for 15 minutes. Then 350 μL of RW Buffer were added to each column, samples were centrifuged at 8000 x g for 15 seconds, then 500 μL of Buffer RPE were added and samples were centrifuged at the same conditions. Flow-through was discarded. Then, 500 μL of RPE Buffer were added to each column and samples were centrifuged at 8000 x g for 2 minutes and then for an additional minute after discarding flow-through. RNeasy column was transferred to a collection Eppendorf tube and 30 μL of RNase free water were added to centre of each column. Samples were incubated at room temperature for 1 minute and the centrifuged for 1 minute at 8000 x g to elute RNA. RNA samples were then kept on ice and RNA was quantified with the nanodrop at 260 nm. Samples were stored at -80ºC for further analysis.

2.7.5. RNA EXTRACTION FROM TISSUE.

Tissue samples (TA muscle intersections) were thawed on ice. In the meantime, working solution was prepared by adding 10 μL of β-mercaptoethanol to 1 mL of RLT Buffer (from RNeasy Mini Kit from QIAGEN). Then, 300 μl of working solution (RLT Buffer with β-mercaptoethanol) and 1 metal bead (3 mm) were added per sample and tissue was disrupted and homogenized with TissueRuptor (25Hz speed) for 4 minutes at 4ºC. In the fume hood, 590 μl of RNase-free water and then 10 μl of proteinase K were added to each sample. Samples were mixed and incubate at 55ºC for 10 minutes in a heat block. Then, samples were centrifuged at 10,000 x g for 3 minutes at room temperature. Supernatant was transferred to a new tube. 0.5 volumes (450 μl) of 100% ethanol were added per sample and samples were mixed (not centrifuged!). 700 μl of each sample were transferred to RNeasy Mini Column (placed in 2 ml collection tubes). Lid was closed and samples centrifuged for 15 seconds at 8000 x g. Flow through was discarded. Remaining supernatant from each sample was added to their respective column and centrifuged again at the same condition. 350 μl of Buffer RW1 were added to each RNeasy column. Lid was closed and samples centrifuged for 15 seconds at 8000 x g. Flow through was discarded. 80 μl of DNase solution (10 μl of DNase and 70 μl of Buffer RDD) were added per sample directly to the column membrane and incubated at room temperature for 15 minutes. Afterwards, 350 μl of Buffer RW1 were added to each RNeasy column. Samples were centrifuged again at the same conditions and flow through was discarded. 500 μl of Buffer RPE were then added to each RNeasy column. Samples were centrifuged again at the same conditions and flow through discarded. 500

μl of Buffer RPE were added again to each column and samples were centrifuged at the same conditions. Lastly, RNeasy column was placed in new 1.5 mL tube. 50 μl of RNase-free water were added to each column and samples were centrifuged for 1 min at 8000 x g at room temperature. Samples were kept on ice afterwards; RNA was quantified with nanodrop at 260 nm. Samples were stored at -80ºC for further processing.

2.7.6. cDNA SYNTHESIS.

A mix of RNA (thawed on ice) and water was prepared in PCR tubes for cDNA synthesis. The mix had a final volume of 14 μL, with 1000 ng of RNA per reaction, 2 μL of gRNA wipeout and appropriate volume of RNase-free water. Samples were then incubated in a PCR machine for 2 minutes at 42ºC and held at 4ºC for 5 minutes to eliminate DNA. In the meantime, a master mix was prepared with: Quantiscript RT buffer 5X of (4μL/sample), Quantiscript RT (1μL/sample) and RT Primer mix (1 μL/sample) from the QuantiTect Reverse Transcription kit from QIAGEN. 6 μL of the master mix were added to each sample and samples were incubated in the PCR machine for 30 minutes at 42ºC, 3 minutes at 95ºC and held at 4ºC. Samples were stored at -20ºC or kept on ice for further analysis.

2.8. POLYMERASE CHAIN REACTION (PCR).

PCRs were performed extensively on this project, particularly to screen gRNAs cutting efficiency and to detect potential deletion of introns 19-55 on edited DNA and cDNA samples from cells and tissue. PCR reactions consisted of three standard stages: denaturation, primer annealing and extension.

2.8.1. MATERIALS FOR PCRS.

- PCR primers ordered from IDT.
- DEPC H_2O from ThermoFisher.
- Thermocycler (PCR machine).
- Q5 High fidelity polymerase kit (including Q5 HF master mix) from NEB.
- GoTaq G2 Flex from Promega.

2.8.2. PCR OPTIMIZATION.

PCR primers were designed on Primer3 adjusting the following parameters: primer size (18-23 bp), primer Tm (57-62ºC), product Tm (-1000-1000ºC, default setting), primer GC% (30-70%, optimal 50%) and the required product size ranges, depending on the product of interest. Primers were ordered and synthetised by IDT. Over 40 primer pairs were tested throughout this research project. Primer pairs that had one PCR product

(were target specific) and were used for gRNA screening and other experiments are presented in Table 2.5. The tubes with the lyophilized primers were spun for 1 minute at 8,000 rpm and then resuspended with DEPC H_2O to obtain 100 μM. After a 5-minute incubation, samples were mixed, and the primer stock was diluted 1:10 into aliquots and stored at -20ºC.

Table 2.5. PCR primers used for gRNA screening. Primers designed on Eurofins SeqPrimer Design Tools.

Primer name (target)	Sequence (5´to 3´)
PCR Primer #16 FW (Guides: 21, 22, 23, 24, 1)	CACTCTGTCAGCTTATCACGTG
PCR Primer #16 RV	ACCTTCTGCCTCAAATTCAAGAG
PCR Primer #17 FW (Guides: 2, 3, 4)	ACCTTCTGCCTCAAATTCAAGAG
PCR Primer #17 RV	TCGGATTACAGGCCTATCTCTT
PCR Primer #18 FW (Guides: 21, 22, 23, 24, 1, 41	TTTCTCGCTCTATGGCCTGC
PCR Primer#18 RV	TGGTGCAGACTGTCCATGTA
PCR Primer #19 FW (Guides: 2, 3, 4, 25)	CTTGAATTTGAGGCAGAAGGTTA
PCR Primer #19 RV	GTGGCGCAATGATAGTTCGT
PCR Primer #21 FW (Guides 7, 26, 27, 28, 29, 30)	GTATCACCAGACCTAACACCAC
PCR Primer #21 RV	TCAAATCACTCCCTTCCCTAATC
PCR Primer #24 FW (Guides: 11, 12, 13, 14, 15, 42)	CCCAGGCAAACATGATACAATTAG
PCR Primer #24 RV	AGCATGAGAGCAAAGGTGAG
PCR Primer #31 FW (Guides 16, 36, 40)	GAATCCCACTGAAGCAGTCTAA
PCR Primer #31 RV	CCTTTGAGACCTACGGAACTAC
PCR Primer #32 FW (Guides: 19, 20, 37)	AAATGGAATCATGTTCTGTAGTTCCG
PCR Primer #32 RV	TCAAATTACCTCCACAGGAGCA
PCR Primer #34 FW (Guides: 17, 18, 38)	GCTAATCAAATCTGTGCATGGT
PCR Primer #34 RV	ATATGGTTAGGCATGGACCAG

Specific PCR protocols had to be optimised for each primer pair by running a temperature gradient in order to confirm the optimal temperature to run the primer annealing stage (Tm).

A PCR Mix using Q5 was prepared on ice as following:

Total volume of 25 μL per sample:

- 12.5 μL of Q5 HF Master Mix
- 1.25 μL of 10 μM Forward primer
- 1.25 μL of 10 μM Reverse primer
- 1,000 - 2,000 ng of DNA
- Appropriate H_2O volume for a final volume of 25 μL

Then optimal annealing temperature for each pair of primer was calculated with NEB Tm Calculator (https://tmcalculator.neb.com/#!/main) and a temperature gradient was run with the following program:

1. Initial denaturation --- 98ºC for 45 seconds or 2 minutes when PCR product is longer than 1kb
2. 35 cycles ---------- 98ºC for 30 seconds

   Temperature gradient for Tm (range of T below 72ºC (i.e. 59-67º C) for 30 seconds

   72ºC for 45 seconds

3. Final extension --------- 72ºC for 2 minutes (when the product is longer than 1kb)
4. Hold ---------- 4ºC for infinite time

A PCR program for amplicons with AT rich regions (Dhatterwal et al., 2017) was used in PCRs targeting AT-rich regions, particularly in intron 55 of human and mouse *DMD* gene. This program uses a lower temperature and longer time for the extension stage. This PCR program was set up as following:

1. Initial denaturation --- 98ºC for 1.5 minutes
2. 35 cycles ---------- 98ºC for 30 seconds
   65º C for 3 min
   65ºC for 3 min
3. Final extension --------- 65ºC for 7 minutes
4. Hold ---------- 4ºC for infinite time

Amplification was then analysed by running the 10 μL of each sample with 2μL of 6X loading dye on a 1% (w/v) agarose gel at 80 V for 1.5 hrs (for PCR products of 200-10,000 bp). A higher percentage of agarose was used when smaller products were expected (2% agarose gel for products of 100-1,000 bp and 3% agarose for smaller products, 25-500 bp). The optimal Tm temperature from the temperature gradients was selected by analysing product bands. If the primers were specific, there should only be one PCR product, showing as a clean bright band on the agarose gel.

2.8.3. PCRs.

Once the optimal annealing temperature was selected for PCR primer pairs, DNA samples and a master mix were prepared as following and kept on ice:

Total volume of 50 μL per sample

- 25 μL of Q5 HF Master Mix
- 2.5 μL of 10 μM Forward primer
- 2.5 μL of 10 μM Reverse primer
- 250 - 350 ng of DNA
- H_2O volume needed to make up to 50 μL

A master mix was prepared for all samples before adding the DNA. DNA was added accordingly to each PCR tube. Then the appropriate program was set up on the PCR machine, as following:

1. Initial denaturation --- 98ºC for 45 seconds
2. 35 cycles ---------- 98ºC for 30 seconds

   Optimal T for PCR primers for 30 seconds

   72ºC for 45 seconds
3. Final extension --------- 72ºC for 1 minute
4. Hold ---------- 4ºC for infinite time

Once the PCR was ready, 10 μl of each sample were run in agarose gel as described earlier. The rest of the PCR samples were stored at -4ºC.

2.8.4. PCR PURIFICATION.

Once expected PCR product was confirmed by an agarose gel, the remaining PCR samples were purified with the QIAquick PCR Purification Kit according to manufacturer´s protocol, in order to send samples for sequencing.

PCR samples were transferred to a 1.5 mL Eppendorf tube (40 μl per sample). Then, 5 volumes of PB Buffer were added per each volume of PCR reaction (i.e. 200 μL of PB Buffer to 45 μL of PCR). The mix was transferred to a QIAquick column and centrifuged at 17,000 x g for 1 minute. Flow-through was discarded and each column was washed with 750 μL of PE Buffer and centrifuged for 1 minute. Flow-through was discarded and samples were centrifuged again. Then, columns were changed to a new 1.5 mL Eppendorf tube and DNA was eluted by adding 50 μL of EB Buffer (10 mM Tris-Cl, pH 8.5). DNA was quantified with a nanodrop at 260 nm. Samples were prepared (according to Eurofins requirements) using the forward or reverse PCR primer as the sequencing primer.

2.9. GUIDE RNA EFFICIENCY ASSESSMENT BY TIDE ANALYSIS.

All *Sa*Cas9 gRNAs were cloned into pAAV-CMV-SaCas9 plasmid. Each construct was transfected by triplicates into an appropriate cell line (HEK293T cells for human gRNAs and N2As for mouse gRNAs) with Viafect transfection reagent. Two days after transfection, cells were harvested and DNA was extracted as described in previous section. PCR products flanking the cut site were purified from treated and untreated samples (as a control) and sent for sequencing to Eurofins.

DNA sequence traces were analysed on the TIDE (Tracking of Indels by Decomposition) web tool. Its algorithm reconstructs the spectrum of indels from an “edited” sequencing trace based on a control (untreated) trace. The output reports identity and frequency of detected indels, as a percentage, generated in a pool of cells (Brinkman et al., 2014) and can be considered the “edited population” from a pool of cells, which is used as a proxy for editing efficiency of an individual gRNA assessed.

2.10. PROTEIN EXTRACTION.

2.10.1. MATERIALS FOR PROTEIN EXTRACTION.

- 1X PBS pre-chilled at 4ºC.
- RIPA Buffer: NaCl 0.15 M, HEPES 0.05m, np-40 1%, sodium deoxycholate (SOC) 0.5%, SDS 0.1%, EDTA 0.01M, protease inhibitor tablet.
- Pre-chilled Eppendorf tubes
- Cell scraper.
- 3 mm metal bead from QIAGEN.
- Tissue homogenizer.

2.10.2. PROTEIN EXTRACTION FROM CELLS.

Before protein extraction, cells were harvested from 6-well plates by aspirating media from each well, washing cells with cold PBS (1 mL per well), removing the PBS, adding 50 μL of RIPA buffer per well and incubating at room temperature for 5 minutes. Then, while holding the plate at a 45-degree angle, cells were accumulated on the bottom side of the well with a cell scraper. Cells were transferred to pre-chilled Eppendorf tubes, and each tube was vortexed for 30 seconds, three times. Samples were then centrifuged at maximum speed for 15 minutes at 4ºC. Supernatant (proteins) was recovered and kept at -20ºC until further analysis and protein quantification.

2.10.3. PROTEIN EXTRACTION FROM TISSUE.

TA muscles were sectioned on a cryostat before protein extraction and stored at -80ºC. Each sample consisted of approximately 30 intersections of 30 μm from TA muscle, samples were always kept on ice. 150 μL of RIPA buffer and 1 (3 mm) metal bead were added to each sample. Samples were processed on the tissue homogenizer for 4 minutes and were then centrifuged at 13,000 x g for 10 min at 4ºC. Supernatant (proteins) was kept and stored at -20ºC until further analysis.

2.11. PROTEIN QUANTIFICATION BY DC ASSAY.

2.11.1. MATERIALS.

- Bovine Serum Albumin (BSA) stock at 2 mg/ml (ampules available from Thermo Scientific, Cat no. 23209).
- Bio-Rad DC protein assay Reagent A (Bio-rad Cat no. 5000113).
- Bio-Rad DC protein assay Reagent S (Bio-rad Cat no. 5000115).
- Bio-Rad DC protein assay Reagent B (Bio-rad Cat no. 5000114).
- 96-well clear flat bottom plate.
- Multichannel pipette.

2.11.2. PROTEIN DC ASSAY.

To have a standard curve, protein standards were prepared with BSA (2 mg/kg) as following:

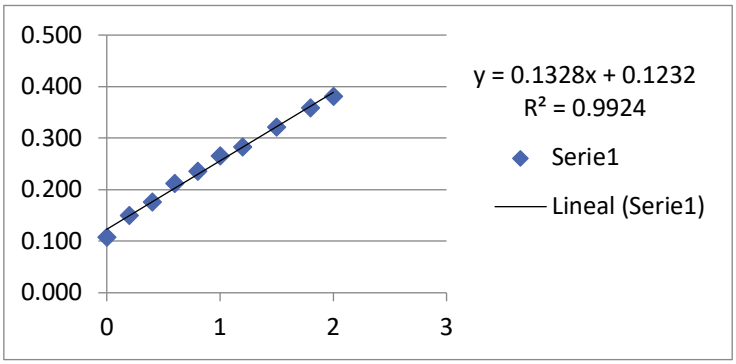
Final conc.	**2**	**1.8**	**1.5**	**1.2**	**1**	**0.8**	**0.6**	**0.4**	**0.2**	**0**
RIPA buffer	0	4	10	16	20	24	28	32	36	40
BSA (2 mg/ml)	40	36	30	24	20	16	12	8	4	0

Before starting the assay reagent A´ was prepared by adding 20 µl of Reagent S to each 1 ml of reagent A. Then, 5 µl of protein standards (in triplicate) and 0.5 µl of samples (in duplicates) were added into a clean, dry 96 well plate. 25 µl of reagent A' were added into each well. 200 µl of reagent B were added into each well with a multi-channel

pipette. Reagent B was also used as a blank in a couple of wells. Plates were gently agitated (~80 rpm) to mix the reagents. After 15 minutes, plates were read at 750 nm. Results output look like the following example (standards are in red):

	1	2	3	4	5	6	7	8	9	10	11	12
A	0.387	0.356	0.323	0.278	0.259	0.233	0.21	0.175	0.147	0.106	Blank	Blank
B	0.372	0.36	0.32	0.28	0.268	0.236	0.21	0.175	0.153	0.109		
C	0.383	0.362	0.323	0.293	0.269	0.238	0.218	0.177	0.152	0.107		
D	0.33	0.363	0.409	0.411	0.354	0.343	0.395	0.35	0.294	0.304		
E	0.331	0.304	0.248	0.238	0.376	0.362	0.227	0.266	0.244	0.277		
F	0.309	0.32	0.282	0.308	0.185	0.184	0.26	0.248	0.3	0.307		
G	0.277	0.265	0.273	0.31	0.363	0.363	0.264	0.288	0.271	0.282		
H	0.276	0.31	0.282	0.307	0.311	0.34	0.298	0.325				

To obtain protein concentration, values of each standard and samples were averaged. OD values of standards were used to make a standard curve in Excel by graphing data and obtaining its linear trend line and slope-intercept equation. Output should look similar to the following example:



To obtain protein concentration, calculations for each sample were done in excel using the obtained slope-intercept equation ($y = (m)(x) + (b)$, where x = concentration, y = OD reading, m = slope of the linear trend and b = y-axis intercept), as following:

$$y = (m)(x) + b$$

$$y - b = (m)(x)$$

Concentration: $x = (y - b)/m$

Dilution factor was taken into account (concentration obtained is 0.1X of the actual concentration as 0.5 µl of the samples were used vs 5 µl of the standards).

2.12. WESTERN BLOTS.

2.12.1. MATERIALS AND SOLUTIONS FOR WESTERN BLOTS.

- Running Buffer (1000 mL): prepared by adding 500 mL of distilled water, 50 mL of NuPAGE running buffer (Tris-Acetate or MOPS running buffer) and then remaining distilled water up to 1000 mL. Running buffer to be chosen accordingly to gel type: for dystrophin - Tris-Acetate running buffer with 3-8% Tris-Acetate gel and for Cas9 protein MOPS running buffer with 4-12% Bis-Tris gel into a 1L flask.
- Transfer Buffer (1000 mL): prepared by adding 400 mL of distilled water, 100 mL of methanol (10 %), 50 mL of NuPAGE transfer buffer (20X), 1 mL of antioxidant and then the remaining water up to 1000 mL into a 1L flask.
- 1X PBS (1L).
- 0.1% PBST: prepared by adding 1000 mL of 1X PBS and 1 mL of Tween 20 into a 1L flask.
- 0.2% PBST: prepared by adding 1000 mL of 1X PBST and 2mL of Tween 20 into a 1L flask.
- 5% Milk: prepared by adding 50 mL of 0.2 PBST and 2.5 gr Marvel Milk powder into a 50 mL flask.
- Blotting pads (4 for one gel or 5 for two gels).
- Filter paper cut to Blotting Pad size (1 per gel).
- Nitrocellulose membrane cut to Blotting pad size from Fisher Scientific.
- Blotting pads.

- NuPAGE 10X Reducing agent (ThermoFisher Cat no. NP0009).
- NuPAGE 4X LDS buffer (ThermoFisher Cat no. NP0007).
- HiMark pre-stained HMW ladder from Life Technologies.
- Chameleon Duo Pre-stained protein Ladder from LI-COR.

2.12.2. Sample preparation.

Protein samples were prepared to a final volume of 20 µl using 2 µl of NuPAGE sample reducing agent (10x), 5 µl of NuPAGE LDS sample buffer (4x), water and protein. The amount of protein loaded was 50 µg/well for protein samples obtained from cells and 30 µg/well for protein samples obtained from tissue (TA muscles), as these seemed to be thicker and 50 µg would make the well collapse. Prior to electrophoresis, samples were heated at 70 °C for 10 minutes to denature proteins.

2.12.3. Western Blotting protocol.

2.12.3.1. Electrophoresis.

Tanks were prepared with gel holders and transfer buffer was poured in the tank. Appropriate gel type and size was chosen (NuPAGE 3-8 % Tris-Acetate gel for dystrophin protein samples or 4-12% Bis-Tris gel for Cas9 protein samples) according to the size of the specific protein to be detected. Comb from the gel was removed gently, gels were

inspected and the wells rinsed with distilled water. Gels were inserted (if only one gel was running then an empty gel cassette was inserted as gel tanks hold two gels at a time) and samples were loaded (10 µL of protein lysate per well with 30-50 µg of protein) and two ladders for each gel. Ladder was chosen according to protein of interest size (i.e. for dystrophin, HiMark pre-stained HMW ladder was used and for Cas9 protein, Chameleon Duo Pre-stained protein Ladder). Tanks were filled up to 1 cm to the top edge with running buffer and 500 µl of antioxidant were added in the inner side of the chamber before starting the electrophoresis. Samples were run for 1 h and 15 minutes at 150V.

While the gels were running:

- Blotting pads were soaked in approximately 700 mL of transfer buffer.
- Filter papers and nitrocellulose membrane were cut.

2.12.3.2. TRANSFER.

Gel fasters were cracked open with a spatula and the upper side was gently removed. Wells and the red line at the bottom of the gel were cut off. Gel was rinsed in transfer buffer (poured in a tray) and then lifted with a filter paper to make up the “transfer sandwich” by following the scheme on Figure 2.4.

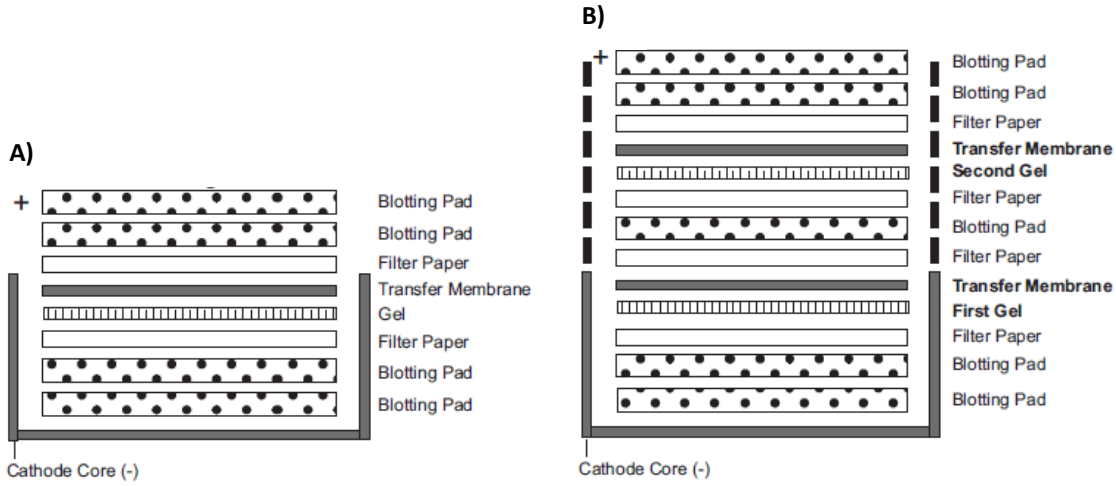


Figure 2.4. Transfer sandwich for Western Blots. A) Set up for one gel. B) Set up for two gels.

When making the "transfer sandwich" air bubbles between the gel and the membrane were squeezed gently with a roller to avoid them interfering with the transfer. Once the transfer cassette was assembled, transfer buffer was poured in the inner chamber only and chilled distilled water (to prevent the tank overheating) was poured in the outer chamber. Transfer was set up for 2 hours at 30V.

2.12.3.3. PONCEAU STAINING.

After transferring, it was proceeded with Ponceau staining to check quality of the transfer. Membrane was put on a weighing tray and rinsed gently with 1X PBS. A small amount of red Ponceau solution was poured in the weighing tray to dye all the protein in a non-specific way allowing to visually confirm if the transfer was successful. After a

few minutes membrane was rinsed with 0.1 % PBST until no red Ponceau was visible on the membrane.

2.12.3.4. BLOCKING.

Membranes were transferred to black boxes and proteins on the membrane were blocked with 25 mL/box of 5% milk for 1 hour at room temperature.

2.12.3.5. PREPARING MEMBRANES FOR ANTIBODIES.

Membranes were cut in two, one half containing protein of interest and bottom half containing reference protein (alpha-tubulin). A scalpel was used to cut through the membranes (ladders on both sides of the membrane were used as guides).

2.12.3.6. PRIMARY ANTIBODIES.

Appropriate primary antibodies were selected:

- For dystrophin: Manex1011C mouse primary antibody at a 1/100 dilution.
- For Cas9 protein: Anti-SaCas9 mouse primary antibody from Diagenode at 1/5000 dilution.

- Reference gene: alpha-tubulin rabbit primary antibody from Abcam at 1/2500 dilution.

Membranes with primary antibodies in 5 mL of 5 % milk were incubated in black boxes overnight at 4 °C on the orbital shaker (in the cold room).

2.12.3.7. SECONDARY ANTIBODIES.

Secondary antibodies were selected accordingly:

- Goat anti-mouse 800CW (green) from LI-COR at 1/10,000 dilution.
- Donkey anti-rabbit 680RD (red) from LI-COR at 1/10,000 dilution.

Before adding the secondary antibodies, membranes were washed 5 times for 5 minutes with 0.1 % PBST. Secondary antibodies were added and incubated for 1 hour at room temperature on the orbital shaker. Membranes were kept covered at all times (in black boxes). Afterwards, membranes were washed 5 times for 5 minutes with 0.1 % PBST and washed one last time with 1X PBS. Membranes were stored in PBS at 4ºC until imaging.

2.12.4. IMAGING.

Membranes were scanned with LI-COR Odyssey CLX machine and analysed with Image Studio Lite Software.

2.13. AAV PRODUCTION.

2.13.1. MATERIALS & SOLUTIONS FOR AAV PRODUCTION.

For giga-preps:

- EndoFree Plasmid mega and giga kit from QIAGEN.

For cells transfection:

- DMEM Glutamax/10%: 500 ml DMEM, 50 mL heat inactivated FCS, 5 mL Penicillin/Streptomycin.
- DMEM Glutamax/2%: 500 mL DMEM, 10 mL heat inactivated FCS.
- Polyethylenimine (PEI) (MW ~25,000 from Polysciences Inc. Cat. No. #23966): dissolved in water heated to 50°C at 1 mg/mL & pH 7.0. Filtered (0.22 µm) and aliquoted, stored at -20°C for up to 6 months.
- PEG supernatant AAV precipitation: 40% Polyethylene glycol (PEG) 8000 40% [w/v] PEG 8000 (Sigma #P2139) with 2.5 M NaCl (Sigma #S7653) in water. (For 500 mL: 200 gr PEG 8000, 73.05 gr NaCl). Autoclaved for 15 min at 121 °C. (After autoclaving, it will separate into two layers, while it's still warm. Allow to stir without heating until it has cooled down). Then stored at room temperature. 50 mL 40% PEG used for every 200 mL supernatant.

For AAV purification:

- Lysis buffer (500ml): 0.15M NaCl (Sigma #S7653), 25 mL 1M Tris HCl pH8.5 (50mM) (Sigma #3253), 1 ml 1M $MgCl_2$ (92 mM) (Sigma # 8266). Volume made up to 500 mL with ddH2O. Autoclaved for 15 min at 121 °C.
- 5X PBS-MK (500 ml): 450 mL distilled water added to 25 Phosphate Buffered Saline Tablets (Oxoid #BR14a). Autoclaved and cooled down. Then, 2.5 mL 1M $MgCl_2$ (5mM) and 6.25 ml 1M KCl (12.5mM) (Sigma #P9541) were added and volume was made up to 500 mL with sterile water. ($MgCl_2$ and KCl were added after autoclaving as these salts would precipitate out of solution if autoclaved, $MgCl_2$ and KCl salts were autoclaved separately (50 mL each).
- 1X PBS-MK (2L): 400 mL of 5X PBS-MK were added to 1600 mL of sterile water. 200 µl of 10% Pluronic F-68 (Gibco #24040-032) were added. Solution was filtered (0.22 µm) and stored at room temperature.
- Pierce Universal Nuclease for cell lysis (Thermo Fisher #88701)
- 0.1M Glycine pH 2.0 (Sigma #G7126): prepared by adding 7.5g of glycine to 1 L of water and adjusting pH to 2.0 using an acid (Sulfuric of hydrochloric acid). Solution filtered sterilized through a 0.2 µM filter unit.
- 25mM NaOH (Sigma #S8045): prepared by adding 1 gr of NaOH pellets to 1 L of water. Solution filter sterilized through a 0.2 µM filter unit.
- Tris-HCl, pH 8.5 (Sigma Cat no. #10812846001): prepared by adding 157.6 gr of Tris-HCl to 1 L of water and adjusting pH to 8.5 using NaOH. Solution filter sterilized through a through a 0.2 µM filter unit.

- 1X PBS: prepared by adding 10 Phosphate Buffered Saline Tablets (Oxoid #BR14a] to 1000 mL distilled water and filter sterilizing.
- 0.1M Citric acid: prepared by adding 19.21 gr of citric acid powder (Sigma #C2404-100G) to 1 L of distilled water. Solution filter sterilised through a 0.2 µM filter unit.

Other materials required:

- Slide-A-Lyzer Dialysis cassette from ThermoFisher, 10,000 MWCO, 12mL, (#66453).
- Syringes: 1 mL, 5 mL and 10 mL syringes.
- Needles: 18G x 1 ½" (from BD) and 21G x 4 ¾" (from Sterican, B. Braun).
- Bottle Top Filtration Unit: 500 mL Funnel Only 0.45 µM and 0.22 µM (#83.3941.100 and 83.3941.101 respectively from Starstedt).
- FACS Tubes: 5mL FACS tubes with caps (from Fisher #10186400).
- Duran Bottles: clean bottles for collection of flow through and for filtration of supernatants.
- Syringe Filters: 0.8 µM (Corning, #431221) 0.45 µM (Starstedt #83.1826), 0.22 µM (Starstedt #83.1826.001).
- 2 L plastic beaker with magnetic flea and stirrer for dialysing overnight.
- Virkon.

2.13.2. GIGA-PREPS.

Giga-preps of the following plasmids were prepared according to manufacturer´s protocol: pAAV-Spc512-GFP (Fig. 2.5), pAAV-Spc512-Multiplex-G14-G18 (Fig. 2.6), pAAV-Spc512-Multiplex-BsaI-BbsI (empty construct) (Fig. 2.7), pAAV-Spc512-Multiplex-G14-BbsI, pAAV-Spc512-Multiplex-BsaI-G18 and pDP9 (helper plasmid for AAV9 production) (Fig. 2.8.).

2.13.2.1. PLASMIDS USED FOR AAV9 PRODUCTION.

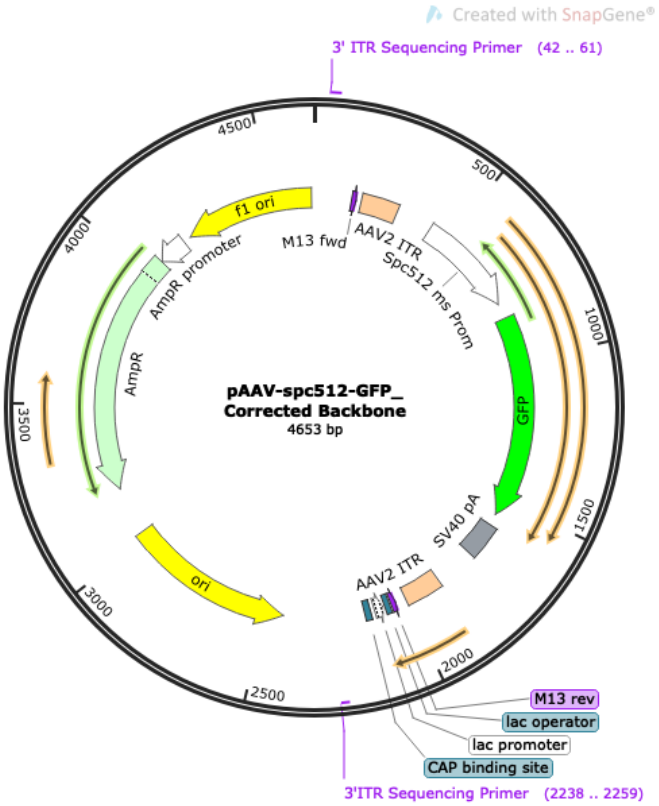


Figure 2.5. Plasmid map of pAAV-Spc512-GFP. Plasmid expressing a GFP under an Spc512 promoter, used as a control.

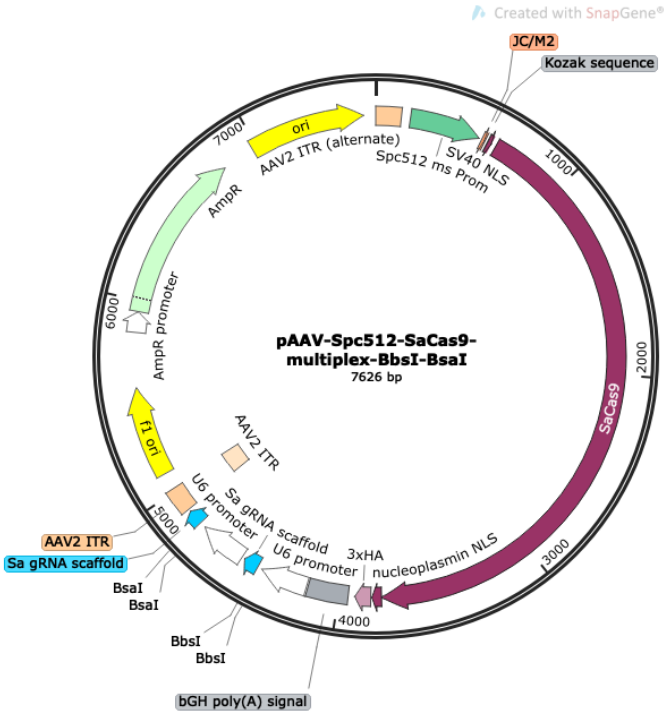


Figure 2.6. Plasmid map of pAAV-Spc512-SaCas9-multiplex-G14-G18. Multiplex construct expressing two gRNAs, G14- targeting intron 18 and G18 - targeting intron 55 of the mouse *DMD* gene. *Sa*Cas9 driven by an Spc512 promoter.

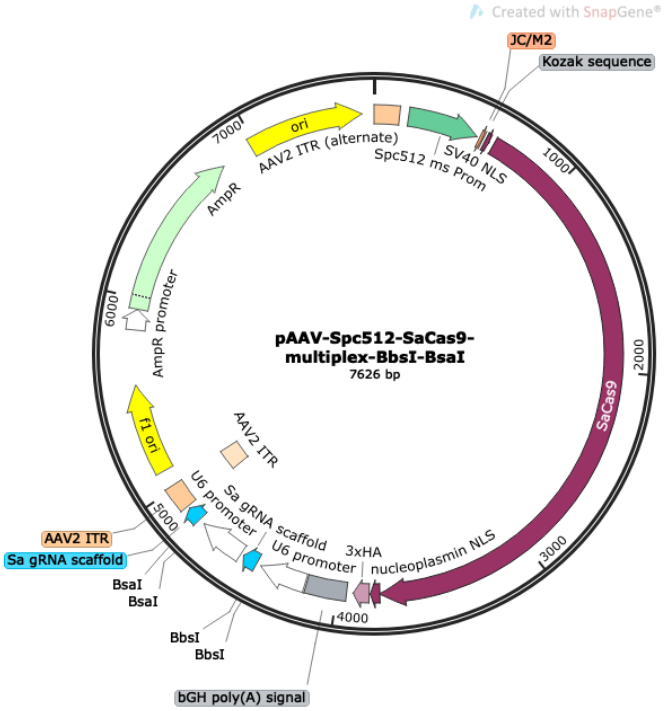


Figure 2.7. Plasmid map of pAAV-Spc515-SaCas9-BbsI-BsaI. Construct in which G14 and 18 were individually cloned into to generate of pAAV-Spc515-SaCas9-G14-BsaI and pAAV-Spc515-SaCas9-BbsI-G18, used as controls.

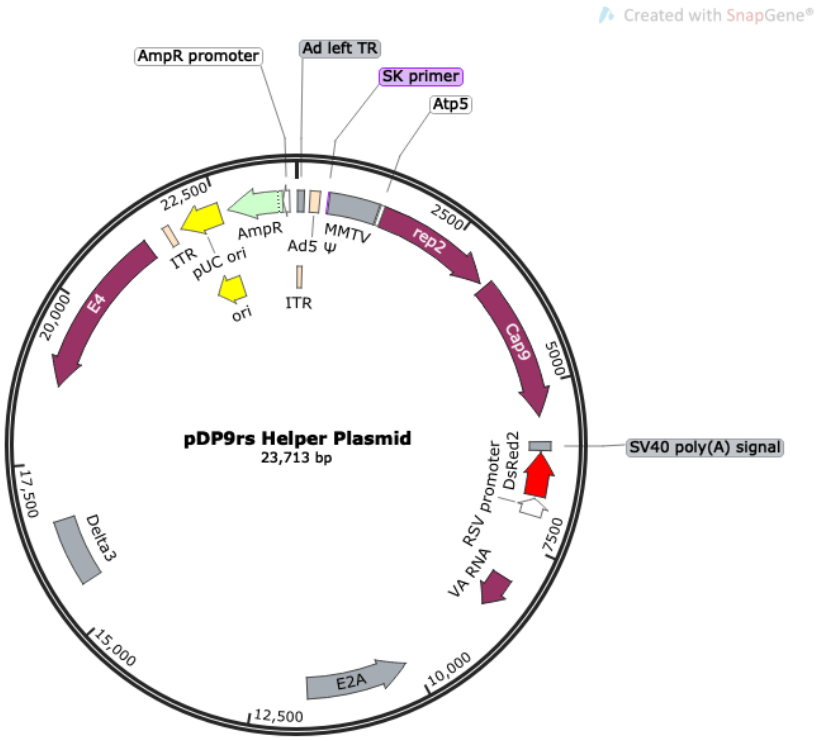


Figure 2.8. Plasmid map of pDP9 helper construct. Plasmid expressing rep (replication) and cap (capsid) genes for AAV9 vector production and E4 gene to stimulate replication.

2.13.3. Transfection of HEK293T/C17 cells in roller bottles with Polyethylenimine (PEI).

The Polyethylenimine (PEI) transfection method was used (1:4 DNA to PEI ratio) to transfect HEK293T/C17 cells. Cells were plated until 70–80% confluent in DMEM/10% (usually 3 days). Cells were then split and seeded at 5 x 10^7 cells per roller bottle, 200 mL of DMEM/10% were added per roller bottle with ventilated caps. Roller bottles were placed in incubator with rotor at 0.5 rpm overnight. Cells were monitored for contamination throughout the process. The next day, rotor was turned up to 1 rpm. Two

days later, 2 hours before transfection, medium on cells was changed to 180 mL DMEM/2% FCS. Serum free DMEM was pre-warmed to room temperature. 500 µg of plasmid (per roller bottle) were diluted in a total of 18 mL serum free DMEM in a universal tube and incubated for 5 minutes at room temperature. Then, 2 mL of PEI were added to the DNA/DMEM mix. Mix was mixed gently and incubated for 15 minutes at room temperature (not vortexed!). 20 mL of DNA/PEI/DMEM mix were added to each roller bottle. Roller bottle was gently tipped upright and DNA mix was added to medium at bottom of bottle to avoid contact with the plastic (as DNA will stick to plastic surfaces). Cells were cultured for 3 days.

2.13.4. SUPERNATANT HARVESTING & CELL LYSIS.

Following transfection after 3 days, cells were harvested as following:

Roller bottles were shaken to detach all cells. Cells were transferred to 500 mL Corning tubes. Roller bottles were washed with 20 mL of 1X PBS and added to their respective supernatant. Tubes were centrifuged at 4000 x rpm for 30 minutes. Supernatant was decanted to fresh 500 mL Corning tubes and frozen at -20°C. Cell pellet was resuspended in 10 mL/roller bottle of lysis buffer and transferred to a 50 mL Falcon tube. 2 mL of lysis buffer were used to wash out the centrifuge tube and were added to cell suspension. Tubes were vortexed for 1 minute and incubated at room temperature for 5 minutes. Samples were freeze/thaw from -80 °C to 37°C three times and were finally stored at -80ºC until purification.

2.13.5. AAV PURIFICATION BY LIQUID CHROMATOGRAPHY WITH THE AKTA GO SYSTEM.

Day 1:

Cell lysate and supernatant were thawed and 4 µl (50U/mL) (approx. 4 µL/18 mL of lysate) of Pierce Universal Nuclease were added to the lysate and 1 µL Pierce Universal Nuclease for each 10 mL of supernatant. Samples were incubated for 30 minutes at 37ºC in the shaking water bath. Cell lysate and supernatant were then clarified by centrifugation at 4000 x rpm for 30 minutes at 4°C. Cell lysate was poured off into a fresh tube. Using a syringe, the lysate was filtered through the following series of syringe filters: 0.8 µm, 0.45 µm and 0.2 µm. This is referred to as the crude lysate, which was store at 4 °C. The supernatant was filtered through a 0.45 µm and a 0.2 µm. Filtered lysate was added to the filtered supernatant and stored at 4 °C overnight.

Day 2:

The supernatant was allowed to come to room temperature while the equilibration of the HPLC machine (AKTA go) was carried out. If the machine had not been used for a while, any air from the piston pumps was removed by opening the inlet valve box (done from the control panel software, inlets/outlets are opened/closed by clicking on them in the control panel).

Line “A” and pump A were opened and set with a flow rate of 0mL/min (a dashed green line in the control panel indicated that the line was open but not running). Then, a syringe was inserted into pump A and the screw was turned 2 ½ times round to open the valve. 3 mL of liquid were removed with a syringe. Screw was returned to the original position and tightened securely. Then, line “B” and pump B were opened and the procedure was repeated. Pumps indicated in the following equipment image:

HPLC machine was prepared by equilibrating the machine and relevant column (Poros AAV9 SN 00068 from ThermoFisher) with 1X PBS. The flow rate was maintained between 3-4 mL/min to equilibrate the lines and at least 5X of column volumes (30 mL) to equilibrate the column. To change between the lines, the inlet valve box was opened and the relevant line was clicked on the control panel. Then it was confirmed that all the lines (line A, B, C, 2A, waste, outlet 1, fraction collector and sample line) were free from 20% EtOH and were in PBS.

Before equilibrating the column, it was confirmed that the system was clear of ethanol and had been flushed with PBS through all the lines. Flow rate was reduced to 1 mL/min. When fixing the column (the arrow on the column indicates the flow direction), the flow direction was followed. The column was connected to the pump tubing in the flow direction, drop-to-drop, to avoid introducing air into the system. 30 mL of PBS were passed through the column before proceeding. Once equilibration was completed, the supernatant was fed through the sample line and through the column at a rate suitable to the volume of the supernatant, 1-4 mL/min. The pressure gauge was checked to ensure that the pressure did not reach 2.8 mPA as this would damage the column. (If pressure is high, flow rate is lowered). Supernatant flow through was collected into a clean bottle by changing the Outlet Valve to Outlet line, connected to the clean bottle.

Once the supernatant had flowed through, the lines and column were washed with 1X PBS until the UV reading was back to baseline. Once UV readings were back at baseline, the flow rate was slowed down to 1 mL/min and the fractionating volume was changed to 3 mL/tube. The machine was paused at this stage. 10 FACS tubes (5 mL) were placed into the fractionator starting from position 1. Flow rate was set to 1 ml/min and the machine was unpaused. Virus was eluted from the column using 0.1M Glycine pH 2.0. An increase in the UV reading indicated virus is passing, when UV reading started to rise (roughly after 10 mL of flow) the 'start fractions' button was pressed in the control panel. A peak in the UV reading appeared and the elute was collected into the tubes, this was the eluted virus.

Tubes containing virus were marked. Once the UV fell back to baseline, the fractionation was stopped and the elution buffer was run through. Recovered fractions containing virus were neutralised with Tris-HCL pH 8.5 (30 µL of Tris-HCl neutralised 1 mL of eluent). Flow rate was increased to 3mL/min and 50 mL of elution buffer were run through the column, then the line was transferred to run 1X PBS through for 30 mL.

While the PBS was running through the column, all the fractions containing virus were collected with a syringe and needle. A Slide-A-Lyzer dialysis cassette (10, 000 Molecular weight cut off) was pre-wetted in a 2 L bucket containing 1.5 L of 1X PBS. The virus was injected into the Slide-A-Lyzer cassette and any air was removed by extraction using the needle and syringe. A floater device was placed onto the Slide-A-Lyzer cassette on the side that the virus was injected and this was placed in PBS. The Slide-A-Lyzer cassette was slowly stirred overnight at room temperature.

Lastly, the column was cleaned: following the PBS wash, the flow of the column was inverted and washed with 30 mL of PBS. Then all lines and column were washed with EtOH 20%. Following the 20% EtOH wash, the flow was paused, the column (filled with 20% EtOH) was removed and stored at 4ºC. All lines were washed again with EtOH 20% and left filled with 20% EtOH. The machine and computer were turned off.

Day 3:

On day 3, virus samples were desalted and concentrated using an Amicon Ultra-15 Centrifugal Filter Unit with Ultracel-100 membrane from Millipore (#UFC910024) for

concentrating samples. The filter was pre-rinsed by adding 15 mL of 1X PBS-MK (with 1:1000 of 10% Pluronic F-68 added) to the filter and centrifuged for 15 minutes at 5000 x g at 4ºC. The virus was removed from the Slide-A-Lyzer cassette using a syringe and needle and added to the filter device. The Slide-A-Lyzer cassette was washed with 5 mL of 1X PBS-MK, the solution used for the wash was added to filter device as well. Samples were centrifuged at 5000 x g for 15 minutes or until volume has been reduced to approximately 250 µL.

Once this volume was achieved, the sample was removed from the filter (with a short/ medium needle), rinsing the sides carefully. This contained the desalted AAV, which was aliquoted in 50 µL volumes and store at -80°C. A separate 10 µL volume was aliquoted in two PCR tubes (5 µL each) for viral DNA extraction and titration.

2.14. QUANTITATIVE POLYMERASE CHAIN REACTION (QPCR).

In this project qPCR protocols were used to titrate AAV vectors and to quantify dystrophin expression from cDNA obtained from treated cells and tissue samples.

2.14.1. MATERIALS FOR QPCRS.

- SYBR Green master mix (FastStart Universal SYBR Green Master mix 2X with FastStart Taq DNA Polymerase, Reaction Buffer, Nucleotides (dATP, dCTP, dGTP, dUTP), SYBR Green I and a reference dye) from Roche.
- Nuclease free water from QIAGEN.
- 96-well plates.
- LightCycler480 Instrument II from Roche.
- LightCycler480 Software.

2.14.2. AAV TITRATION BY QPCR.

A common method for AAV titration is by quantifying viral genome copy numbers by qPCR. To titre the AAV9 preps, a sample from each prep was digested with DNAse to eliminate any potential DNA outside of the viral capsids and then digested with Proteinase K to eliminate viral capsid and obtain packaged viral genomes.

These samples were then titred by qPCR with primers designed to bind on the *Sa*Cas9 sequence, present in: pAAV-Spc512-SaCas9-multiplex-G14-G18, pAAV-Spc512-SaCas9-BbsI-BsaI, pAAV-Spc512-SaCas9-G14-BsaI and pAAV-Spc512-SaCas9-BbsI-G18 and the GFP present in pAAV-Spc512-GFP control plasmid (Table 2.6).

Table 2.6. qPCR primer pairs used for AAV titration.

Primer name (target)	Sequence (5´to 3´)
SaCas9 Set 1 FW	CTGGAACGGCTGAAGAAAGA
SaCas9 Set 1 RV	GTCGATGTAGGTGTCGATGAAG
SaCas9 Set 2 FW	CAAGTGCTATGAGGAAGCTAAGA
SaCas9 Set 2 RV	GTTCACGCCGATCACTCTATAC
SaCas9 Set 3 FW	AACCGAGCAGGAGTACAAAG
SaCas9 Set 3 RV	GGAGTACAGGGTGTCGTTAATC
GFP FW	CAAGATCCGCCACAACATCG
GFP RV	GACTGGGTGCTCAGGTAGTG
Rplp0 (reference gene) FW	TTATAACCCTGAAGTGCTCGA
Rplp0 RV	CGCTTGTACCCATTGATGATG

Standard curves were set up appropriately with plasmids: pAAV-Spc512-SaCas9-multiplex-G14-G18 and pAAV-Spc512-GFP, from giga-preps which were nanodropped and diluted to obtain 1E+10 copy numbers in 40 µL. A g-block of Rplp0 was used to prepare the standard curve for the reference gene. Then appropriate standard curves were prepared by serial dilutions from 1E+10 to 1E+1 DNA copy numbers.

SYBR Green master mix (FastStart Universal SYBR Green Master mix 2X with FastStart Taq DNA Polymerase, Reaction Buffer, Nucleotides (dATP, dCTP, dGTP, dUTP), SYBR

Green I and a reference dye from Roche) was prepared to 1X mixed with 400 nM of each primer (forward and reverse); 6 µL of the mix and 4 µL of each sample were loaded per well on a 96-well plate by triplicates. Then, plates were processed on a LightCycler480 Instrument II from Roche and data was analysed on the LightCycler480 Software to obtain: the melting curve, the amplification curve of the standard curve samples and the amplification curve of all samples, the standard curve and its efficiency, Cp values and concentration of each sample calculated by the Software.

Based on the concentrations calculated by the LightCyler480 Software, titres were obtained by calculating the viral genome copy numbers (or viral particles) per reaction for each sample/prep (by triplicates) and averaged.

2.14.3. Dystrophin expression and deletion of exons 19-55 quantification by qPCR.

To detect deletion of exons 19-55, a primer pair binding to exons 20-21 was designed alongside a primer pair binding to exons 6-7, as a control. Rplp0 was used as a reference gene. Primer sequences are presented in Table 2.7. If exons 19-55 were deleted there should be a decrease in exons 20-21 expression in treated samples compared to control samples.

Table 2.7. qPCR primer pairs used for dystrophin quantification.

Primer name (target)	Sequence (5´to 3´)
Exon 6-7 FW	GTCATCAACTTCACCTCTAGCTG
Exon 6-7 RV	CCACACTATTCCAATCAAACAGG
Exon 20-21 FW	CAGATGACAACTACTGCCGAA
Exon 20-21 RV	GAAGAGCTGACAATCTGTTGAC
Rplp0 (reference gene) FW	TTATAACCCTGAAGTGCTCGA
Rplp0 RV	CGCTTGTACCCATTGATGATG

Standard curves were set up using g-blocks expressing: *Dmd* mouse gene exons 6-7 and exons 20-21 and Rplp0, with the following sequences:

- Rplp0 (reference gene):

  5´- TTA TAA CCC TGA AGT GCT CGA CAT CAC AGA GCA GGC CCT GCA CTC TCG CTT TCT GGA GGG TGT CCG CAA CGT GGC CAG TGT GTG TCT GCA GAT CGG GTA CCC AAC TGT TGC CTC GGT GCC ACA CTC CAT CAT CAA TGG GTA CAA GCG -3´

- Exon 6-7 *Dmd*:

  5´- TAT CCA CAG GTT AAC GTC ATC AAC TTC ACC TCT AGC TGG TCC GAC GGG TTG GCT TTG AAT GCT CTT ATC CAT AGT CAC AGG CCC GAC CTG TTT GAT TGG AAT AGT GTG GTT TCA CAG CAC TCA GC -3´

- Exon 20-21 *Dmd*:

  5´- GAA CAG ATG ACA ACT ACT GCC GAA AAC TTG TTG AAA ACC CAG TCT ACC ACC CTA TCA GAG CCA ACA GCA ATT AAA AGC CAG TTA AAA ATT TGT AAG GAT GAA GTC AAC AGA TTG TCA GCT CTT CAG C -3´

Standard curves were prepared by serial dilutions from 1E+10 to 1E+1 DNA copy numbers of each g-block.

A SYBR Green master 1X master mix was prepared for each primer pair as described in the previous section, with 400 nM of each primer (forward and reverse); 6 µL of the mix and 4 µL of each sample were loaded per well on a 96-well plate by triplicates. Then, plates were processed (LightCycler480 Instrument II from Roche) and data was analysed (LightCycler480 Software) as described in the previous section, to obtain: the melting curve, the amplification curve of the standard curve samples and the amplification curve of all samples, the standard curve and its efficiency, Cp values and concentration of each sample. Data was then analysed on Excel.

2.15. PROTOCOLS USED FOR *IN-VIVO* INJECTIONS AND TISSUE SAMPLES PROCESSING.

Two *in-vivo* experiments were performed in *mdx* mice in this research project, one experiment involved plasmid delivery with different doses and electro-transfer directly into TA muscles to assess expression and potential protein functionality of the positive control plasmid expressing Del19-55 dystrophin. The final *in-vivo* experiment was delivered by AAV9 vectors containing our CRISPR systems to assess their efficiency *in-vivo*.

All animal procedures in this project were performed in accordance with the UK Animals (Scientific Procedures) Act, 1986. *Mdx* mice (C57BL/1-ScSn-Dmd*mdx*) and C57/Bl10 mice (referred to as “wild type”) were bred in our animal facility and were maintained in a standard 12-hour light/dark cycle with free access to food and water.

2.15.1. MATERIALS.

For mice injections:

- Isoflurane-based anaesthesia system (Harvard apparatus - including isoflurane, isoflurane absorber, O_2, induction chamber).
- Disinfectant, i.e. 1% distil, 70% ethanol.
- Syringes and needles (0.3-0.5 ml syringes with 29G-30G needle size).
- Substances of interest.

For muscle electrophysiology:

- Black braided silk, non-sterile, non-absorbable surgical suture, 4-0 USP, 1.5 metric (Harvard Apparatus 633573 or 51-7615).
- Hypnorm (Vetapharma Vm 41760/4000).
- Hypnovel or Midazolam (Roche 10107972, 10 mg/2 mL).
- Syringes and needles.
- Scalpels.
- Cork board.
- Medical tapes.
- Cotton buds (Tesco).
- Saline or 1X PBS.

- Surgery tools (InterFocus Ltd., surgicaltools.co.uk): fine scissors martensitic stainless stell straight 10.5 cm (14094-11). Student Dumont #5 forceps standard inox (91150-20), Dumont medical #7 forceps curved inox (11273-20) and spring scissors straight sharp 8mm cutting edge (15024-10).

2.15.2. INTRAMUSCULAR TA INJECTIONS.

2.15.2.1. PLASMID DNA TRANSFER BY ELECTRO-TRANSFER.

Before starting the injections, the work area was cleaned with disinfectant. Before plasmid injections and electro-transfer, Tibialis anterior (TA) muscles were injected with hyaluronidase (1 hour before treatment) to enhance gene transfer (Gollins et al., 2003).

For plasmid injections and electro-transfer, mice were sedated as following: the isoflurane absorber was weighed to ensure its weight is <1400 gr. It was checked that there was backup O_2 available. Mouse to be injected was weighed. The volume of substance to be injected was previously calculated. The isoflurane system was started and O_2 flow rate was set at 2 L/min and isoflurane at level 5 (=5% in 100% O_2). The mouse was placed into the anaesthesia chamber and monitored until mouse is under deep anaesthesia (heart rate goes down and beats constantly). Plasmids were injected intramuscularly in (TA) muscles followed by electro-transfer. Muscles were harvested 7 or 14 days after injections and stored at -80ºC until sectioning.

2.15.2.2. AAV9 DELIVERY (TA MUSCLE TRANSDUCTIONS).

Mice were prepared and sedated as described in the previous section. Mice were injected on both TA muscles with a dose of $1x10^{11}$ vp / 30 µL of saline solution per TA. TA muscles were harvested 2 months after treatments and stored at -80ºC until further analysis.

2.15.3. ELECTRO-PHYSIOLOGY ANALYSIS.

2.15.3.1. PREPARATION.

Mice were weighed and weight was recorded. Anaesthesia reagent was prepared by mixing Dolethal (Vetoquinol) and Buprenodale (Dechra) in the following concentrations:

- Dolethal 200 mg/mL, diluted 1:10 in H_2O.
- Buprenodale 0.3 mg/mL, diluted 1:10 in H_2O.

Diluted reagents were mixed 1:1 and injected into mice at volume (µl) of 5-7X body weight (gr). After injection, mouse was put back in its box until asleep (5-10 mins). Withdrawal reflex should be absent before starting surgery. Animal's breathing was monitored and when needed anaesthesia was topped (with 50 µl if mouse started twitching and 100 µl if it started moving through subcutaneous injection). Total volume of anaesthesia should never go over 200-300 µl.

2.15.3.2. SURGERY.

1) To prepare the tendon:

The leg to be used was wetted with water and fur was shaved off, from foot to hind quarter. Excess water was dried. Mouse was placed on its back with the foot in use being taped to the board. Exposed tissue was kept moist with a saline soaked cotton wool throughout the surgery to prevent muscle drying. An incision was made over the TA tendon in the mouse's foot by lifting a bit of skin and extending it proximally to the myotendinous junction, if necessary, a small strip was cut away (avoiding blood vessels).

The TA tendon curves slightly, medial to the EDL tendon towards the little toe. Both are anchored down by a stiff cuff of fascia (retinaculum, this covers the tendons) that needed to be cut to release the tendons. The cuff was cut with a scalpel on the inner side and removed completely. The TA tendon should then spring out of its bindings and be a lot easier to manipulate. Excess connective tissue was removed as this might cause knot slips. The tendon of the small muscle to the little toe was identified and cut (it lies beneath the TA tendon). Two lengths of the silk thread were cut about 20 cm long. One was passed under the TA tendon with fine forceps and pulled halfway through very gently. A double knot was made (left as an untightened loop). The second piece of thread was passed through the loop just below the first thread. A single knot was tied as close to the first thread as possible. A thick needle/pin was inserted into the cork board just below the toes of the mouse. The second thread was tied into a loop around

this needle/pin (1 double knot and then 1 single knot) (knot lies lateral to the body = left side for right leg, right side for left leg). The ends of the second thread were cut, leaving ~3 mm. The needle/pin was unpinned. The TA tendon was cut as far distal as possible (towards the toes). The tendon was folded up over the second knot, passed through the loop of the first knot and a double knot was made right over the top of the folded piece of tendon. Then 2 single knots were tied on top of the double knot, every time in mirror image. These knots prevented slipping on the rig. The ends of the first thread were cut, leaving ~3 mm.

2) To prepare the sciatic nerve:

The mouse was turned onto its side keeping the foot stuck down. The tuber coxa was located and an incision was made just below it over the natural division between the gluteal muscles. The muscles were split to expose the nerve. There should be 2 nerve branches. The deep peroneal branch of the sciatic nerve was identified (the smaller, thinner, and more distal of the two visible branches, it stimulates the TA and EDL). Gently a sharp cut was made to disable the fat nerve branch (mouse will twitch). In order to avoid other muscles contracting and causing noise in the system when measuring the contractions of the TA and EDL muscles, it is important to cut through the upper fatter branch of the sciatic nerve and therefore disable it.

A ~15 cm piece of thread was cut and passed under the deep peroneal nerve as close to the spine as possible (avoiding blood vessels). A single loose knot was tied round the nerve and ends were cut to about 0.8 cm. The nerve proximal to the knot was cut as close to the spine as possible. The nerve was gently lifted up using the threads and freed from any connective tissue (fascia), it was then put back and bathed in saline, avoiding touching the nerve after this.

3) Preparing the patella:

Mouse was placed on its side. An incision over the patella tendon was made. The pin was passed from lateral to medial (body side to abdomen). Mouse was moved to the physiology rig.

2.15.3.3. MUSCLE PHYSIOLOGY.

1) Starting up the system:

PC was turned on and software (DMC v5.300) launched. The “Dual Mode Lever System” was switched on. The “Stimulators” were switched on (only after DMC is open). The know was set up to 20V and 1% (= 0.2 V stimulation). The S-hook was hung on the transducer.

2) Positioning of the mouse on the rig:

The lamp was switched on and the mouse was placed on its stomach on the cork board. The pin was hammered in a hole on the cork board to stabilise the patella and prevent the ankle from moving. Foot was taped to a side of the cork board. The loop was attached to the S-hook on the transducer so that there was a straight line between the muscle and the pin. The position of the knot was kept consistent to prevent twisting. The 2 silver dials on the device were used to adjust the position. The tension of the thread was adjusted (black dial on device) so that it was tense but not pulling on the muscle (approx. 1 gr). The threads attaching to the nerve were held and the nerve was placed over the electrode. The nerve was kept moist with saline (not in excess) with cotton buds without directly touching the nerve. The aim was to have a good contact between the nerve and the electrode without having the electrode touching the rest of the mouse.

3) Initial twitch:

On the PC screen, the menu **File → Live data monitor**, was clicked to check if the system was working and the surgery was correct, by using the **Manual Trigger** on the Dual Level. If there was a spike on Force, the system was working. To start the test the following menus were clicked: **Protocols → Open protocol → Protocols DMC/RHUL folder/Protocols → twitch → Load protocol → Start test.**

The first twitch was started (to confirm correct contact with the nerve and muscle contraction) and the positions of threads were adjusted if necessary. If the whole system was working properly, there would be a single spike, ~70 mN. Appropriate tension on the muscle was checked (start with about 1-1.3 g) and the nerve was kept moist. The voltage was increased slightly and twitch was checked again, repeating until the amplitude of the twitch stopped increasing (should not reach 2 V, if it did, it is was an indicator that there was something wrong with the surgical preparation). The folder to save the data was selected: **Setup → Autosave folder →** Folder where to save data → selected **Current Folder**. On the main screen, boxes for **Save on test completion** and **Open analysis on test completion** were clicked (**without selecting** "Enable autosave").

4) Warm up:

The following menus were clicked to start the warm-up: **Sequence setup → Open sequence → Protocols DMC/RHUL folder/Sequences.** The **warm-up sequence** protocol was selected (this is 5 protocols of warm-up 60 seconds apart from each other), needed to measure eccentric contraction following force frequency.

5) Repeating Twitch to determine optimal tension and voltage:

After the **Warm-up**, the following menus were clicked: **Protocols → Open protocol → Protocols DMC/RHUL folder/Protocols → twitch → Load protocol → Start test.** The tension, position, angle, etc. were adjusted and the twitch was repeated until the force

stopped increasing. This would define the best tension of the thread (maximum tension was avoided as this could break the tendon or the knots of the thread). To save the data, the twitch data for the optimal muscle tension (so called **best twitch**) was manually selected from the data directory and saved. Once optimal tension was defined, it was used for the rest of the contraction protocols for that mouse.

6) Tetanic contractions:

The following menus were clicked to start the tetanic contraction protocol: **Sequences → Open sequence → Protocols DMC/RHUL folder/Sequences → Force-frequency mod2 → Start sequence.** This is a 9-protocol sequence with different frequency of stimulation at 10, 30, 40, 50, 80, 100, 120, 150 and 180 Hz. The entire sequence lasted ~ 7 minutes. The nerve was kept moist and the tension optimal (~1.232 g) and tetanic contractions were measured and saved.

7) Eccentric contractions:

After the tetanic contraction protocol was done, a 5-minute rest period is required before starting the eccentric contraction protocol. During this period: the nerve was kept moist, units were set-up by clicking the following menus **Setup → Channel setup → Length in display units →** "ref" was selected instead of "mm" and settings were saved. The TA muscle was carefully measured a caliper. A measurement from the patella to the myotendinous junction was made and the length of the TA muscle was put in the

main window **"ref length"** in mm. Once the 5 minutes were completed the protocol was initiated by clicking on the following menus: **Sequences → Open sequence → Protocols DMC/RHUL folder/Sequences → ECC sequence relative 15%** (or ECC sequence def). The resting tension was checked before each eccentric contraction. Nerve was kept wet during the protocol (approximately 25 minutes). The mouse was detached and TA muscles were harvested.

8) Data extraction:

DMA v5.0 software was launched. The following menus were selected: "High Throughput" → "Force-Frequency Analysis". Then, selected "Pick Files" and selected the .ddf files need for analysis. Data was analysed by the Software and exported to an excel table.

9) Quantification:

After harvesting TA muscles, weight (mg) and length (mm) for each muscle were recorded and the following calculation were done in excel:

- TA mass was obtained by dividing TA over body weight (mg/g).
- TA Cross Sectional area was calculated: CSA (mm^2) = TA weight / (TA length x 0.6 x 1.067), where 1.067 (mg/mm^3) is the density of mammalian muscle and 0.6 is the optimum muscle length/fibre length ratio for TA muscle.

- Absolute force was measured in the 9-protocol sequence with different frequency of stimulation at 10, 30, 40, 50, 80, 100, 120, 150 and 180 Hz.
- Specific force (mN/mm^2) calculated as maximal force/CSA.
- Eccentric force calculated as percentage of force drop in Eccentric contraction (ECC) = $(ECC_n \times 100)/ECC_1$

2.15.4. MUSCLE HARVESTING.

TA muscles were harvested by cutting through the skin in the euthanised mouse, isolating the TA muscle with forceps. Then the tendon was cut at the base of the ankle and the TA muscle was separated from the other muscles. Once isolated the muscle was recovered, fixed on a labelled cork with OCT compound by the tendon and frozen in liquid nitrogen. Muscles were then wrapped in aluminium foil and stored at -80ºC until further analysis.

2.15.5. MUSCLE SECTIONING WITH CRYOSTAT.

The cryostat was always kept at the following temperatures:

- Quick freeze temperature: -35ºC.
- Specimen temperature: -20ºC.
- Chamber temperature: -22ºC.

TA muscles were transferred from -80ºC storage to the cryostat 15 minutes before starting the procedure, to allow sample temperature to drop. Tissue samples were always kept frozen. Each TA muscle was cut in half, one half was saved for protein extraction and the other one was used for sectioning (the half attached to the cork).

The cork was attached to a metal block with water (water was allowed to freeze to fix the cork against the metal without taking the samples out of the cryostat). Then the block was placed in the block holder of the cryostat and the blade distance adjusted. Each sample was cut in "3 levels", from each level sections of 10 µm were fixed on microscope glass slides (one section per slide) for immunohistochemistry analysis and intersections of 30 µm were placed in Eppendorf tubes for DNA and RNA extraction. Sections on slides and Eppendorf tubes with sections were stored at -80ºC until further analysis.

2.16. IMMUNOHISTOCHEMISTRY OF TISSUE SAMPLES.

Immunohistochemistry was used to detect GFP expressed from the positive control plasmid expressing Del-19-55 dystrophin (fused to a GFP), to detect dystrophin positive fibres co-localised with laminin at the sarcolemma after treatments with AAV9 vector and our CRISPR System and to confirm co-localization of Del-19-55 dystrophin fused to a GFP (positive control) with dystrophin and dystrophin associated complex proteins (α-sarcoglycan, β-dystroglycan and nNOS domain) to assess potential protein functionality.

2.16.1. MATERIALS.

- PBST: PBS + 0.05% Tween20.
- Biotin/Avidin blocking kit from Vector Lab (SP-2001).
- MOM fluorescein kit from Vector Lab (FMK-2201).
- Dako-pen.

2.16.2. LAMININ, EGFP & DAPI IMMUNOSTAINING.

Slides were taken out from -80ºC storage and aired at room temperature for 20 minutes. Limits around the sections were drawn with a Dako-pen to delimit the area for staining. Sections were rehydrated for 5 minutes in ice-cold 1X PBS. Slides were then fixed in ice-cold 4% PFA (in 1x PBS) for 15 minutes at room temperature. Slides were rinsed twice for 5 minutes in ice-cold 1X PBS. Sections were permeabilised in 0.3% Triton X-100, PBS for 10 minutes at room temperature and then rinsed with 1X PBS. Sections were blocked in 2% BSA, 5% goat serum, 0.1% triton X-100, 1X PBS, for 30 minutes at room temperature and then rinsed with 1X PBS.

Samples were stained with anti-GFP primary antibody (1:1000 rabbit polyclonal from Abcam, Ab6JJ6), that would bind to the GFP fused to Del19-55 dystrophin. Antibodies for alpha laminin (rat polyclonal from Sigma, L0663, at 1:1000) and DAPI (1:1000) were used for laminin and central nuclei staining. Anti-GFP and anti-laminin antibodies were

added (in blocking solution) and incubated for 2 hours at room temperature and then washed 3 times for 5 minutes in PBST (0.05% Tween-20).

Then, respective secondary antibodies goat-anti-rabbit Alexa 488 (Invitrogen, 1:500) and goat-anti-rat Alexa 568 (Invitrogen, 1:500), were added and incubated for 1 hour at room temperature. The samples were washed 3 times for 5 min in PBST (0.05% Tween-20). Samples were lastly incubated with DAPI 1:1000 in 1x PBS for 10 minutes and washed 3 times for 5 minutes in 1X PBS.

Slides with sections were then mounted with Mowiol with PDD solution (900 µL + 100 µL) and a cover slip and stored at 4ºC in the dark (wrapped in foil) until analysed by fluorescent microscopy. When imaging, 6 fields were pictured per section with the fluorescent microscope (Zeiss Axio Vision D1 with AxioCam MRm, images acquired with Software ZEN 2012).

2.16.3. DYSTROPHIN AND DPC PROTEINS IMMUNOSTAINING.

Sections were air dried for 30 minutes and then fixed in cold acetone for minutes at 4ºC. Area for staining was delimited with the Dako-pen. MOM blocking solution was added and incubated for 1 hour (2 drops in 2.5 mL of PBS). Afterwards, MOM diluent solution was added for 5 minutes (600 μL protein concentrate in 7.5 mL of PBS). Solution was tipped-off. Appropriate primary antibodies were added in diluent solution and incubated for 60 minutes:

- GFP: anti-GFP primary rabbit antibody (1:1000).
- Dystrophin: Manex1011C primary mouse antibody (1:50).
- α-sarcoglycan: anti-α-sarcoglycan primary mouse antibody (1:50) from Abcam (Ab1120A6).
- β-dystroglycan: anti-β-dystroglycan primary mouse antibody (1:50) from Sigma (11H6C4).
- nNOS domain: anti-nNOS primary mouse antibody (1:50) from BD Biosciences (Cat. No. 610308).
- α-laminin: anti-α-laminin rabbit polyclonal antibody (1:400).

Slides were then wash 3 times for 5 minutes in PBST. Secondary antibody anti-rabbit-488 (1:200) in MOM diluent solution was added for 60 minutes at room temperature. Slides were washed 3 times for 5 minutes in PBST. Then anti-mouse –IgG from the MOM kit was added for 10 minutes (10μl in 2.5 mL of diluent buffer) at room temperature.

Slides were washed 3 times for 5 minutes in PBST. Avidin-568 complex from MOM kit was added for 5 minutes (40 μL in 2.5mL of PBST). Slides were washed 3 times for 5 minutes in PBST. DAPI (1:1000) was added in PBS for 5 minutes at room temperature. Solution was tipped-off and washed with PBST. Slides were mounted in Mowiol/DDP (9/1 proportion) using 2-3 drops depending on the area covered by the sections. Slides were stored at 4ºC until needed for analysis.

2.16.4. MYOFIBRE ANALYSIS: TOTAL FIBRE COUNT WITH MUSCLEJ (FIJI).

Total myofibre count from immunohistochemistry samples was performed with the FIJI Software and the MuscleJ plugin.

Image files (.czi files generated in the Zeiss microscope) were opened after launching the FJII Software. Then the MuscleJ plugin was launched and the following criteria were selected on the “Data Acquisition” window:

- Microscopy: “Apotome/WideField”.
- Volume: “Single”.
- Scanned muscle area: “Crop”.
- Data format: “Original File Format”.
- Data analysis: “Fibre Morphology”.
- Data cartography: “Fibre area class” (for total fibre counting).

Appropriate channels were assigned and the Software made total fibre count of the file. Data was recorded and later analysed on excel once dystrophin positive fibres were counted.

2.16.5. DYSTROPHIN POSITIVE FIBRES COUNT.

From each muscle, one section was analysed for dystrophin positive fibres. From each section, 6 fields were analysed to account for total fibres and dystrophin positive fibres. Dystrophin positive results were evaluated as a percentage of the number of total fibres within the same image/field that were positive with laminin staining.

The “Cell counter” feature from the “Analyze” plugin of the FIJI Software was used to aid manual counting of dystrophin positive fibres. Fibres were considered dystrophin positive when >50% of the fibre showed recovered dystrophin. This criterion was kept consistent among all experiments. Data was gathered and analysed on Excel.

2.17. QUANTIFICATION OF INFECTIOUS PARTICLES BY INFECTIOUS CENTRE ASSAY (ICA).

The Infectious Centre Assay (ICA) allows the quantification of infectious particles in a recombinant AAV stock. This assay involves the infection of a permissive cell line stably carrying the AAV2 *rep* and *cap* sequences (HeLaRC32) with increasing serial dilutions of the AAV vectors to be assessed and with wild type Adenovirus. Thus, infectious AAV particles entering into the cells will be able to replicate. The replication events are then detected by chemiluminescence and quantified following hybridization with a transgene specific probe.

This assay was developed and published in Human Gene Therapy in 1998 (Salvetti et al., 1998). It has since been widely used in the pre-clinical vector core of the UMR1089 (where this assay was kindly performed by Dr. Veronique Blouin and Dr. Caroline Le Guiner) and the HeLaRC32 cells are available at the ATCC.

2.17.1. MATERIALS.

- HeLaRC32 cell line.
- AAV vectors preps.
- 48-well plates.

- Nylon membranes.

2.17.2. PROTOCOL FOR ICA.

The Infectious Centre Assay (ICA) consists of a co-infection of wild type adenovirus (type 5) and recombinant AAV vectors into Hela32RC cells. The HeLa32RC cells are transformed HeLa cells expressing the AAV2 *rep/cap* genes, therefore allowing the replication of rAAV in presence of adenovirus.

26 hours post-infection, the cells are harvested, lysed and blotted on a nylon membrane. A hybridization is performed with a specific transgene probe labelled with fluorescein. The signal is then amplified with an anti-fluorescein antibody coupled with Alkaline Phosphatase (chemiluminescence). Finally, the replication events are quantified by dot counting after revelation on a “radiographic film”.

Full test duration:

- Day 1: Cells seeding in 48 well plates.
- Day 2: Infection of the cells with adenovirus and serial dilutions of AAV vectors to be assessed.
- Day 3: Cells harvest, samples loading on membrane, pre-hybridization and hybridization with the transgene specific probe.

- Day 4: Membranes wash and saturation of non-specific sites, incubation with antibody, washing and chemiluminescence revelation.

TEST CONTROLS AND VALIDITY CRITERIA:

- Positive controls:

(1) HeLa32RC infected with an internal AAV2/8.GFP referent vector and the wild type adenovirus (wtAd5). The titer is expected within a **specific range.**

- Negative controls:

(2) HeLa32RC infected with AAV vector only (without wtAd5): **no replication is expected**. A detected replication event reveals a wtAd5 contamination.

(3) HeLa cells infected with wtAd5 and AAV vector: **no replication is expected** because the HeLa cells do not contain the *rep/cap* genes. A replication event reveals a *Rep+ particles* contamination.

If one of the controls is not conform to the validity criteria, the assay is deemed not valid and the sample is re-tested.

2.18. Statistical Analysis.

Statistical analysis was performed on the GraphPad Prism Software (Version 9.1.0, GraphPad Software INC. San Diego, CA, USA). Results in this thesis are presented as mean ± standard error of the mean (SEM). For multiple comparisons of non-normally distributed data sets or when normal distribution could not be determined, a Kruskal-Wallis test was performed, followed by a Dunn´s test (when comparing means to a control group) and for normally distributed data sets, a one-way or two-way ANOVA was performed (depending on the number of factors analysed), followed by a post-hoc Tukey´s test or Holm-Šídák´s test (for increased power) when making multiple group comparison and followed by a Dunnett´s test when comparing means to a control group. In all analysis $p<0.05$ was considered significant.

3. DESIGN & ANALYSIS OF DEL19-55 TRUNCATED DYSTROPHIN: *IN-SILICO, IN-VITRO* & *IN-VIVO* ASSESSMENT OF POTENTIAL PROTEIN FUNCTIONALITY.

The *DMD* gene is a large complex gene spanning more than 2 million base pairs of the human X chromosome. The genomic sequence is approximately 200 times larger than the final RNA transcript, resulting in a mean size of exons of 200 bp and a mean size of introns of 35,000bp (Koenig et al., 1987). This gene represents almost 0.1% of the whole genome and its large size might be a reason for the high frequency of mutations within its sequence (Koenig et al., 1987). Additionally, this gene displays mutational hotspots for two allelic diseases, Becker muscular dystrophy (BMD) and Duchenne muscular dystrophy (DMD).

The reading frame rule helps explain the clinical differences between Becker´s and Duchenne at a molecular level, showing that a shift in the reading frame of *DMD* mRNA, therefore an out-of-frame mutation, leads to a more severe DMD phenotype; while in-frame mutations lead to the expression of a truncated but functional dystrophin and therefore the milder Becker´s phenotype (Monaco et al., 1988). Nevertheless, according to a study in 2,405 patients from the UMD-DMD database, the reading frame rule

applies to approximately 96% of DMD patients and 93% of BMD patients (Tuffery-Giraud et al., 2009).

This study (Tuffery-Giraud et al., 2009) also showed that DMD mutations from the database encompassed 61% large deletions, 13% duplications and 26% point mutations. Presenting a similar deletion rate to the one reported by Baumbach et al. (1989).

In this research project, it was decided to target introns 18 and 55 of the *DMD* to remove as many mutational hotspots as possible while maintaining the reading frame in-frame, so a potentially functional truncated dystrophin could be expressed. Deletion of exons 19 to 55 would result in an in-frame deletion that would eliminate ~81% of total *DMD* mutations (65% located in mutational hotspot of exons 45-55 (Béroud et al., 2007) and 20.7% mutations within exons 19 to 45 (*The DMD Mutations Database*, n.d.)).

Nevertheless, such a large deletion (of approximately 800 kbp) had not been attempted previously for this gene. Considering this, the aims of this chapter were the following:

- To perform a literature review including patient databases to try to find if a similar deletion had occurred in clinic and led to a mild phenotype.

- To perform an *in-silico* protein analysis to predict a model of the truncated version of the dystrophin that would result from the deletion of exons 19-55 and assess its potential functionality.
- To confirm if Del19-55 *DMD* would express a truncated dystrophin and if this protein would be functional (relevant to highlight that deletion of exons 19 to 55 is an in-frame deletion). Therefore, a positive control was needed. A construct expressing Del19-55 dystrophin cDNA was designed and protein expression was assessed *in-vitro* by Western Blot and *in-vivo* by plasmid delivery, immunohistochemistry and Western Blot.

3.1. LITERATURE REVIEW OF CLINICALLY IDENTIFIED LARGE *DMD* DELETIONS.

It has been previously reported that patients with an in-frame deletion larger than 36 exons tend to show a severe phenotype (Fanin et al., 1996). Since the deletion of Exons 19-55 is a 36-exon deletion, literature and DMD/BMD patient databases (PubMed, The TREAT-NMD DMD global database and the LEIDEN DMD Mutation Database) were reviewed for existence of deletion of exons 19-55 or a similar one in patients and confirm the phenotype presented.

Thirty-three cases of deletions spanning from 15 to 42 exons from unrelated patients, incorporating deletion of exon 19, were identified from PubMed, The TREAT-NMD DMD global database and the LEIDEN DMD Mutation Database and are summarised in Table 3.1. From the 33 cases presented, 21 had in-frame deletions and ten of these 21 cases presented BMD; these patients had deletions ranging from 27 to 42 exons. The remaining 10 cases (from the 21 cases with in-frame deletions) presented a DMD phenotype (severity of the phenotype not indicated). The largest deletion found in a patient was from exons 13-55 (Dastur et al., 2008), spanning an in-frame 42 exon deletion; the patient presented a Becker´s phenotype and was considered an exception to the “>36-exon large deletion rule”. In addition, out of five cases of patients with Del19-51 (The DMD Mutations Database, Agarwal et al., 2017, Mohammed et al., 2018, Lim et al., 2020), three of them displayed a BMD phenotype.

Table 3.1. Summarised data of patients with large *DMD* gene deletions (>15 exons) incorporating deletion of exon 19. Table indicates exons deleted, in- or out-of-frame deletion, phenotype presented, isoforms affected by the deletion inferred from The DMD Mutations Database and references.

Deleted Exons	# Deleted Exons	In- or out-of-frame	Number of registered cases	Phenotype (DMD or BMD)	Isoforms affected	Reference
8-47	39	Out-of-frame	1	DMD	Dp427c, Dp427m, Dp427p, Dp260, Dp140.	(Vengalil et al., 2017)
10-42	32	In-frame	1	DMD	Dp427c, Dp427m, Dp427p, Dp260.	(Andrews et al., 2018)
10-43	33	Out-of-frame	2	DMD	Dp427c, Dp427m, Dp427p, Dp260.	(Vengalil et al., 2017)
13-53	40	In-frame	1	BMD	Dp427c, Dp427m, Dp427p, Dp260, Dp140.	(Lim, Nguyen and Yokota, 2020b)
13-55	42	In-frame	1	BMD	Dp427c, Dp427m, Dp427p, Dp260, Dp140.	(Dastur et al., 2008)
18-44	26	Out-of-frame	1	DMD	Dp427c, Dp427m, Dp427p, Dp260.	(Vieitez et al., 2017)
18-44	26	In-frame	2	DMD	Dp427c, Dp427m, Dp427p, Dp260.	(B. L. Lee et al., 2012)
18-45	27	In-frame	3	Unknown - most likely to be BMD	Dp427c, Dp427m, Dp427p, Dp260, Dp140.	(Vieitez et al., 2017)

19-34	15	In-frame	1	DMD	Dp427c, Dp427m, Dp427p, Dp260.	(Lim, Nguyen and Yokota, 2020b)
19-43	24	Out-of-frame	1	BMD (Unusual exception)	Dp427c, Dp427m, Dp427p, Dp260.	(Juan-Mateu et al., 2015)
19-44	25	In-frame	4	DMD	Dp427c, Dp427m, Dp427p, Dp260.	*(The DMD Mutations Database*, n.d.)
19-44	25	In-frame	1	DMD	Dp427c, Dp427m, Dp427p, Dp260.	(R. Guo et al., 2015)
19-46	27	In-frame	1	Unknown - most likely to be BMD	Dp427c, Dp427m, Dp427p, Dp260, Dp140.	(Vieitez et al., 2017)
19-48	29	In-frame	1	Unknown - no observation	Dp427c, Dp427m, Dp427p, Dp260, Dp140.	(Zimowski et al., 2014)
19-50	31	Out-of-frame	4	DMD	Dp427c, Dp427m, Dp427p, Dp260, Dp140.	(*The DMD Mutations Database*, n.d.)
19-51	32	In-frame	1	DMD	Dp427c, Dp427m, Dp427p, Dp260, Dp140.	(*The DMD Mutations Database*, n.d.)
19-51	32	In-frame	1	DMD	Dp427c, Dp427m, Dp427p, Dp260, Dp140.	(Lim et al., 2020)

19-51	32	In-frame	1	BMD	Dp427c, Dp427m, Dp427p, Dp260, Dp140.	(Mohammed et al., 2018)
19-51	32	In-frame	2	BMD	Dp427c, Dp427m, Dp427p, Dp260, Dp140.	(Agarwal et al., 2017)
20-53	33	Out-of-frame	1	DMD	Dp427c, Dp427m, Dp427p, Dp260, Dp140.	(Takeshima et al., 2010)
22-45	23	Out-of-frame	1	DMD	Dp427c, Dp427m, Dp427p, Dp260, Dp140.	(Vieitez et al., 2017)
24-43	19	Out-of-frame	1	DMD	Dp427c, Dp427m, Dp427p, Dp260.	(Servais et al., 2015)

The deletions from the cases summarised in Table 3.1 are depicted on a full exon schematic in Figure 3.1, indicating DMD and BMD cases by colour (blue and grey respectively).

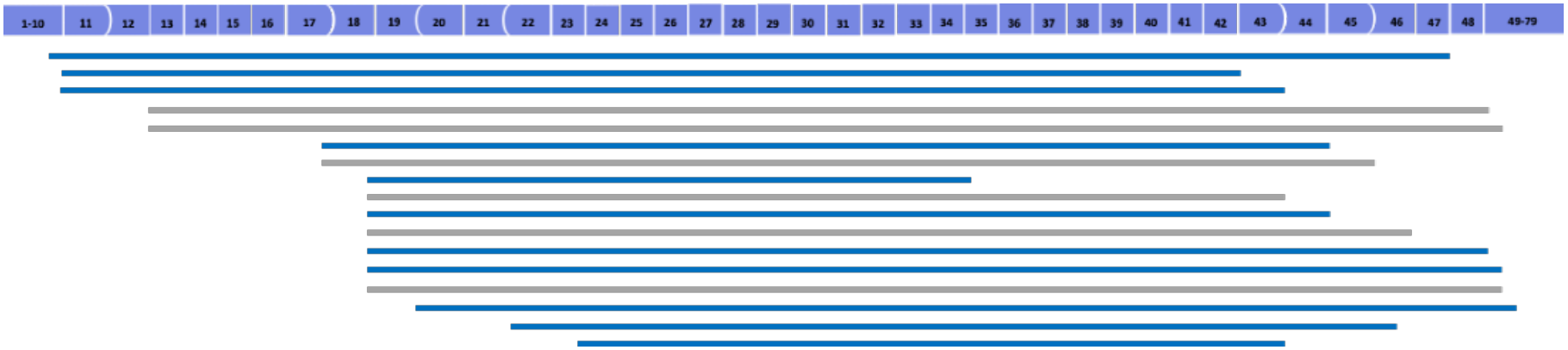


Figure 3.1. Summarised data of large *DMD* gene deletions (>15 exons) incorporating deletion of exon 19 (from Table 3.1) aligned to a full exon schematic. Blue bars indicate DMD phenotype and gray bars indicate BDM phenotype.

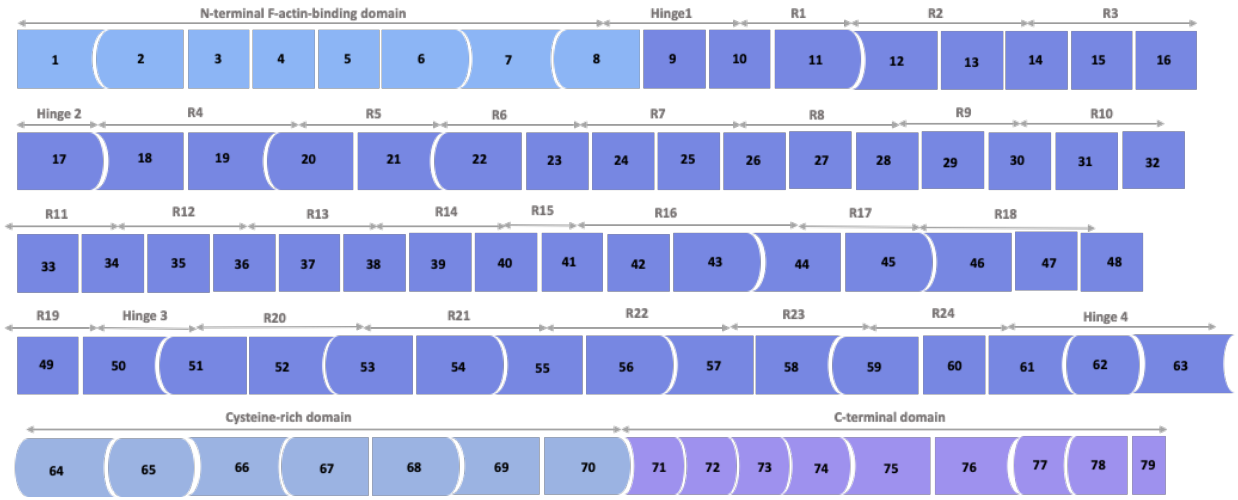
Even though the exact deletion of exons 19-55 has not been reported in clinic yet, it was encouraging to find similar large deletions leading to a BMD phenotype. The fact that such like deletion has not been reported in clinic yet, could imply that if a patient is carrying it, they might be asymptomatic. Based on these findings, the next aim of this chapter was to perform an *in-silico* analysis modelling the protein that would be expressed after the deletion of Exons 19-55 and evaluate its potential functionality as a truncated dystrophin.

3.2. *IN-SILICO* ANALYSIS OF TRUNCATED DYSTROPHIN AND *DE NOVO* JUNCTION FROM DELETION OF EXONS 19-55.

Deletion of exons 19 to 55 would result in a potential Becker-like in-frame deletion and would theoretically remove 81% of *DMD* mutations that result in DMD, including the deletion hotspot on exons 45-55.

Once the correct reading frame was confirmed based on exons phasing, as shown on Fig. 3.2, the predicted truncated protein structure was elucidated in Figure 3.3 and compared with that of full-length dystrophin and other micro-dystrophins. It must be highlighted that, unlike µDys-5R, the truncated Del19-55 form would not possess the nNOS domain. However, the truncated Del19-55 dystrophin possesses the features of MD1: Hinge 1, spectrin-like repeats 1-3, hinge 2, spectrin-like repeat 24 and Hinge 4. Based on successful results from canine MD1 studies (Le Guiner et al., 2017) and successful safety results from clinical trials (NCT03375164) with MD1 (Delandistrogene moxeparvovec (SRP-9001) from Sarepta), showing improvements in functional measures over 3 years (J. Mendell et al., 2022), Del19-55 dystrophin, with similar features, has potential to restore dystrophin expression and stabilize clinical symptoms as well.

A. Phasing of the 79 exons of the *DMD* gene.



B. Phasing of exons 1 to 18 and 55 to 59, confirming an in-frame deletion of exons 19-55.

Figure 3.2. Comparison of exon phasing between A) full-length dystrophin, with 79 exons. Domains indicated: N-terminal F-actin-binding domain (encoded by exons 1–8), rod (R; encoded by exons 8–64), cysteine-rich (CR; encoded by exons 64–70) and C-terminal (CT; encoded by exons 71–79) domains. The rod domain can be further divided into 24 spectrin-like repeats and four interspersed hinges. Adapted from (Duan et al., 2021). B) Truncated dystrophin with exons 19-55 deleted. Exons remain in-frame after deletion of E19-55.

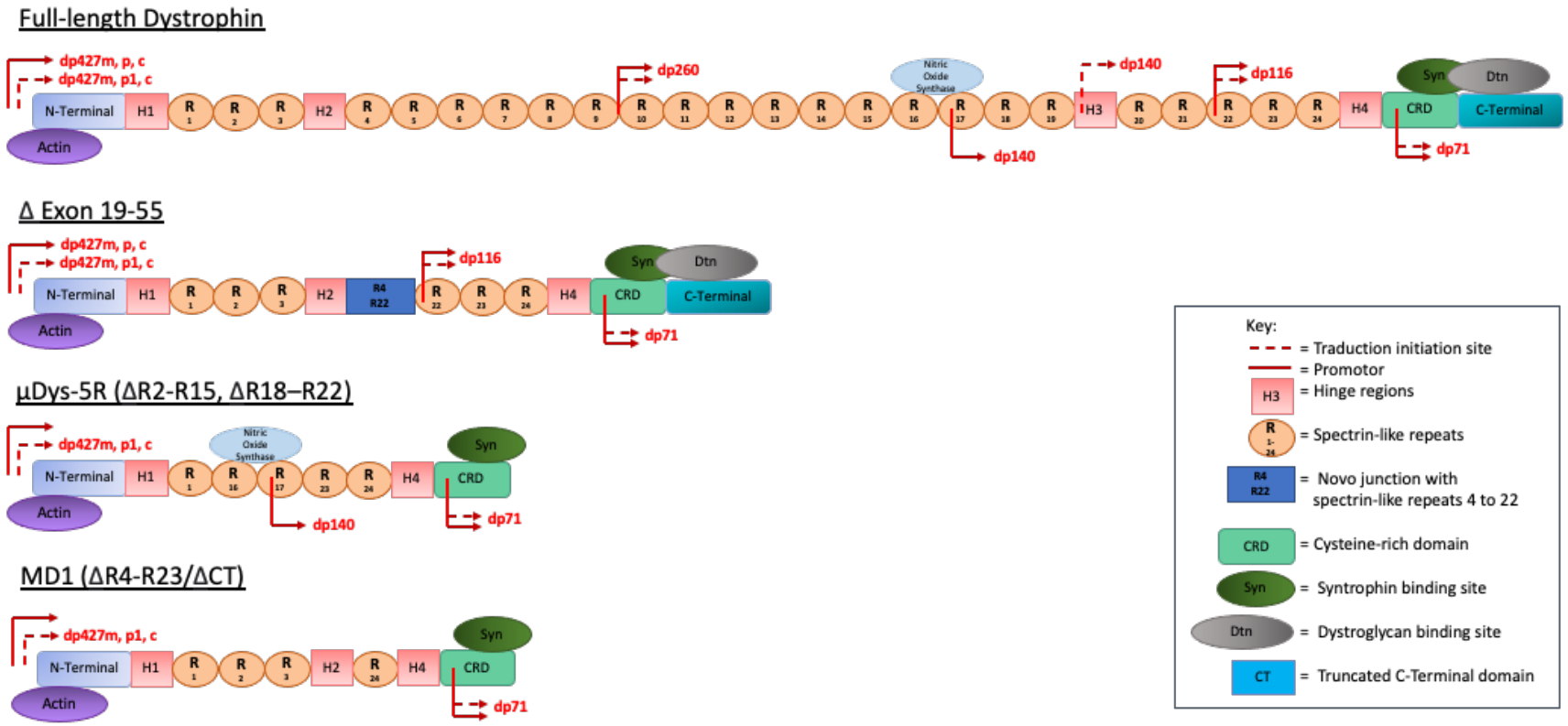


Figure 3.3. Full length dystrophin, truncated dystrophin (Del19-55) and representative micro-dystrophins. The proteins schematic shows a variety of dystrophin variants highlighting dystrophin domains and isoform promotors. Full-length dystrophin contains actin bound N-terminal domain, 24 spectrin-like repeats, four hinges, a cysteine-rich domain and a C-Terminal domain. Δ Exon 19-55 is the truncated dystrophin that would be express after deletion of Exons 19-55, including a de novo junction of spectrin-like repeats 4 and 22. µDys-5R (Hakim et al., 2017) and MD1 (Harper et al., 2002) present common features, such a as n-terminal domain, cystein-rich domain, spectrin-like repeats 1 and 24, hinges 1 and 4. Differences are in central hinges and the nNOS domain present only in µDys-5R. Promotors are displayed in red. Figure adapted from (Duan, 2018).

To have a more detailed analysis, the protein sequence of the hDel19-55 dystrophin (Section 2.1.1.) was modelled and analysed on The Phyre2 web portal for protein modelling, prediction and analysis, developed by (Kelley et al., 2015).

The output model has more than 90% confidence for 81% of the modelled residues and can thus be considered highly confident as shown in Fig. 3.4.

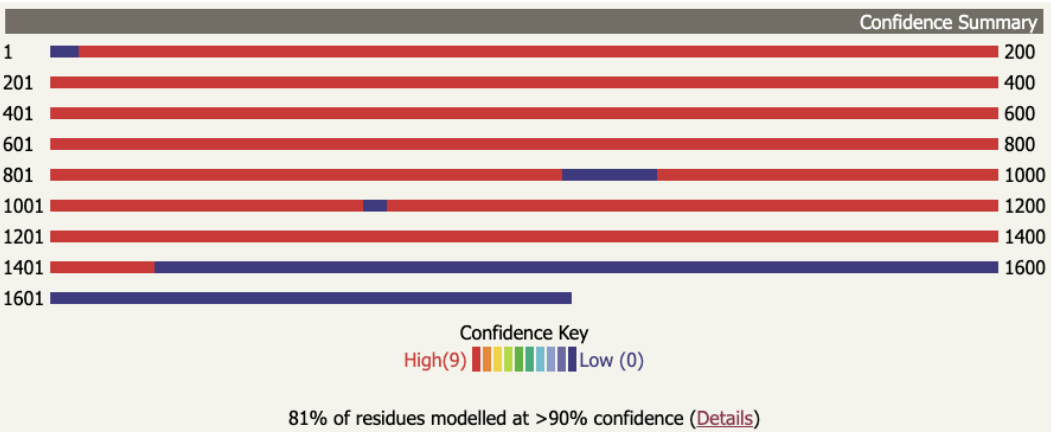
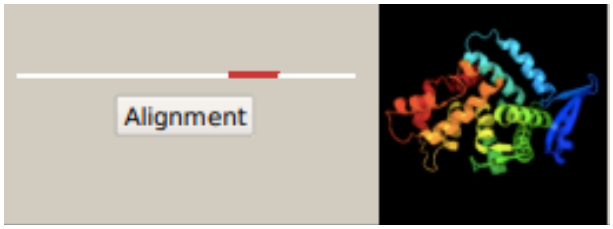
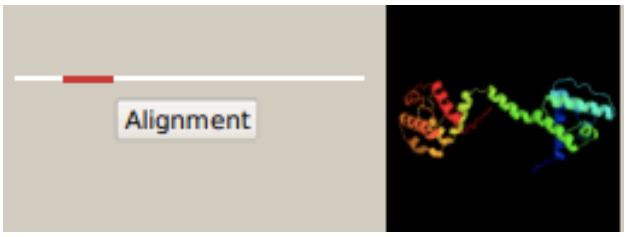
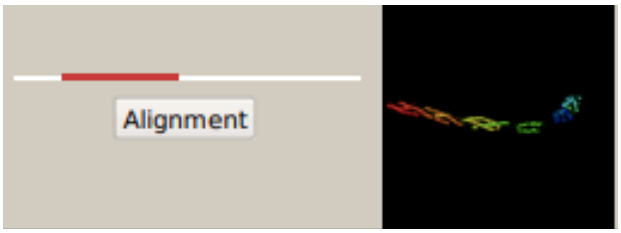
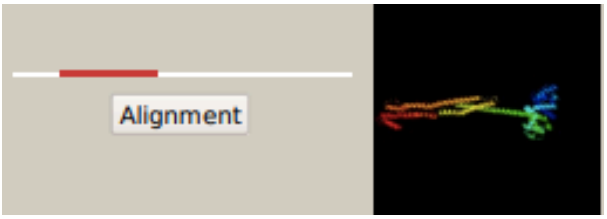
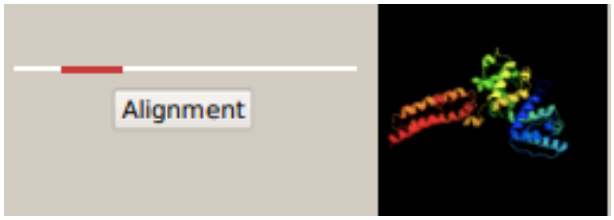
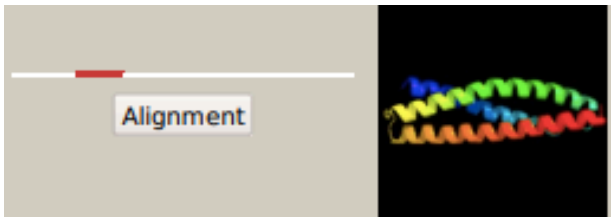


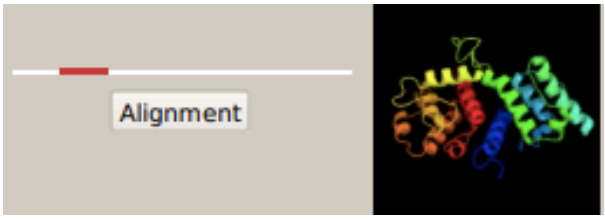
Figure 3.4 PHYRE2 output of Del19-55 protein model confidence, showing 81% of residues in red, indicating confidence >90% and low confidence highlighted in blue. Numbers denote amino acids (aa) positions. Loss of confidence starts at the 3´end of the sequence, after aa position 1401, which falls within the sequence of exon 70.

The Del19-55 dystrophin model was aligned against PHYRE2 protein database. The output indicates: “Confidence”, the probability that the match between the model and the template is a true homology and “% i.d.”, the percentage identity or accuracy between the model and the template. The 10 most confident alignments can be seen on Table 3.2 where Del19-55 model shows 100% confidence and 99-100% i.d. against full length dystrophin.

Table 3.2. Output of the 10 most confident alignments by PHYRE2 of Del19-55 dystrophin model against proteins from database. 3D models show the protein structure expressed by Del19-55 based on a template alignment. 3D models colour scheme draws each chain as a spectrum from blue, green, yellow and orange to red, where the n-termini of proteins are coloured in blue and the c termini red. Alignment shows coverage of Del19-55 dystrophin alignment (in red) against the template. Details of the protein/domain used as template are indicated in the "Template information" section. "Confidence" indicates homology between the model and the template. "ID%" indicates accuracy between the model and the template. Template name is a 6-character identifier (assigned by PHYRE2) where "c" indicates this protein is a whole chain from the protein data bank (PDB) (RCSB PDB) or "d" indicating that the template is domain entry from the SCOP database, followed by an alphanumeric identifier assigned by the PDB to experimental structures (i.e. 1ed4) followed by the chain identifier (instance level identifier assigned by the PDB to indicate a distinct copy of an entire molecule), i.e. A or B.

#	Alignment coverage & 3D Model	Template information
1		**Confidence: 100** **ID%: 100** **Template:** c1ed4A **PDB Header:** structural protein **Chain:** A **PDB Molecule:** dystrophin **PDB Title:** structure of a dystrophin ww domain fragment in complex2 with a beta-dystroglycan peptide
2		**Confidence: 100** **ID%: 99** **Template:** c1dxxB **PDB Header:** structural protein **Chain:** B **PDB Molecule:** dystrophin **PDB Title:** n-terminal actin-binding domain of human dystrophin
3		**Confidence: 100** **ID%: 25** **Template:** c1sjjB **PDB Header:** contractile protein **Chain:** B **PDB Molecule:** actinin **PDB Title:** cryo-em structure of chicken gizzard smooth muscle alpha-actinin
4		**Confidence: 99.9** **ID%: 99** **Template:** d1dxxa2 **PDB Header:** CH domain-like **Superfamily:** calponin-homology domain, CH domain

		Family: calponin-homology domain, CH domain
5		**Confidence: 100** **ID%: 100** **Template:** d1eg3a1 **Fold:** EF Hand-like **Superfamily:** EF-hand **Family:** EF-hand modules in multidomain proteins
6		**Confidence: 100** **ID%: 26** **Template:** c6sl2A **PDB Header:** structural protein **Chain:** A **PDB Molecule:** calponin homology domain protein putative **PDB Title:** alpha-actinin from entamoeba histolytica
7		**Confidence: 100** **ID%: 36** **Template:** c4z6gA **PDB Header:** cell adhesion **Chain:** A **PDB Molecule:** microtubule-actin cross-linking factor 1, isoforms 1/2/3/5 **PDB Title:** structure of nt domain
8		**Confidence: 98.8** **ID%: 100** **Template:** c3uunA **PDB Header:** structural protein **Chain:** A **PDB Molecule:** dystrophin **PDB Title:** crystal structure of n-terminal first spectrin repeat of dystrophin
9		**Confidence: 99.8** **ID%:99** **Template:** d1dxxa1 **Superfamily:** Calponin-homology domain, CH-domain **Family:** Calponin-homology domain, CH-domain

10		**Confidence: 100** **ID% 48** **Template:** c3f7dA **PDB Header:** structural protein/cell adhesion **Chain:** A **PDB Molecule:** plectin-1 **PDB Title:** crystal structure of a complex between integrin beta4 and 2 plectin

The superposition of dystrophin (Chain A) and the Del19-55 model (#1 from Table 3.2) showed a template modelling (TM) score of 1. TM-score is a normalized score from 0-1 representing overall similarity of protein. Identical structures score 1, scores above 0.5 indicate the same overall fold and scores under 0.2 indicate a similarity no better than random.

Then this model was analysed to predict whether missense mutations in the protein are likely to have a functional/phenotypic effect. Compared to dystrophin (Chain A), the model had low mutation sensitivity as observed on Fig. 3.5.

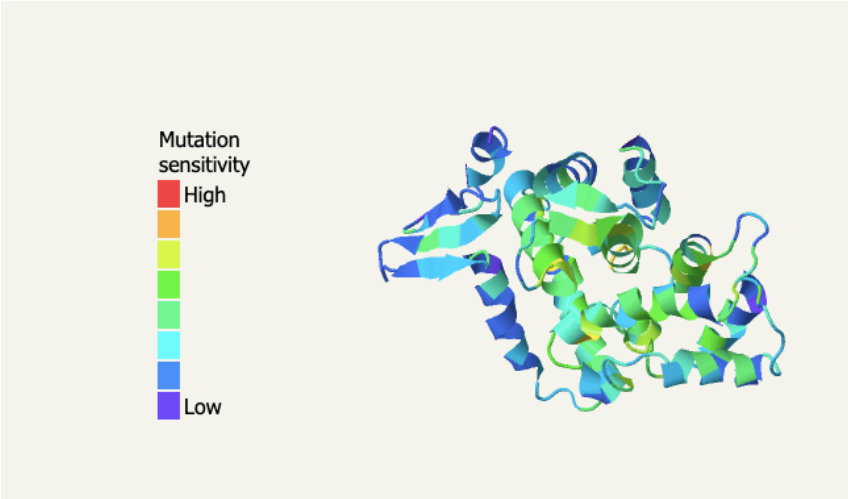


Figure 3.5. Protein structure mutational sensitivity analysis of the superposition of Del19-55 dystrophin model and dystrophin (Chain A). Colour code scale indicates mutation sensitivity, where red is high (mutations could cause a negative functional effect) and blue is low.

The superposition of dystrophin (Chain B) and the Del19-55 model (#2 from Table 3.2) had a TM-score of 1. Missense mutations analysis of the model showed low mutation sensitivity as observed on Fig. 3.6.

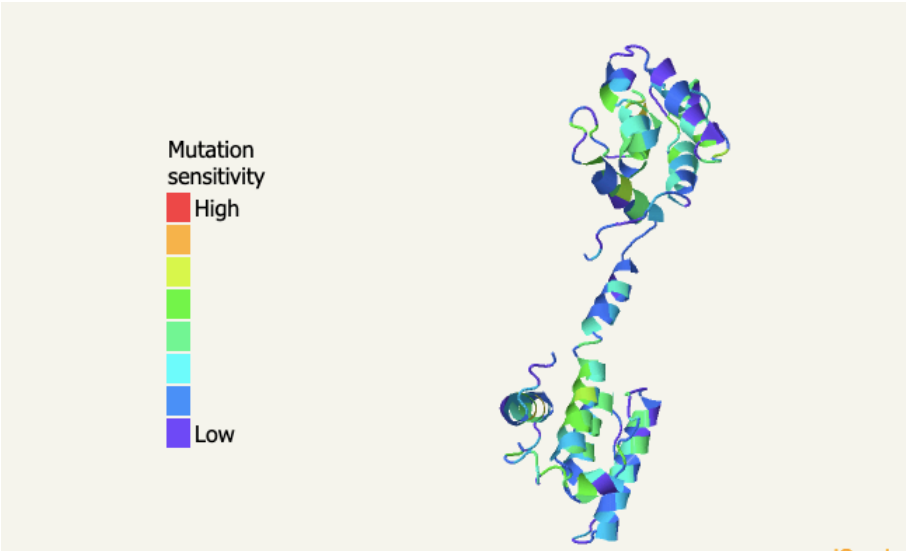


Figure 3.6. Protein structure mutational sensitivity analysis of the superposition of Del19-55 dystrophin model and dystrophin (Chain B). Colour code scale indicates mutation sensitivity, where red is high (mutations could cause a negative functional effect) and blue is low.

The predictive full model of Del19-55 dystrophin can be seen on Fig. 3.7. Based on PHYRE2 analysis, there are no obvious clashes in the protein structure that could interfere with folding or mutations that would affect functionality. The C-terminal of the protein is expressed, however it is important to highlight that the 3´ end of the model had a low confidence and structure might not be 100% accurate.

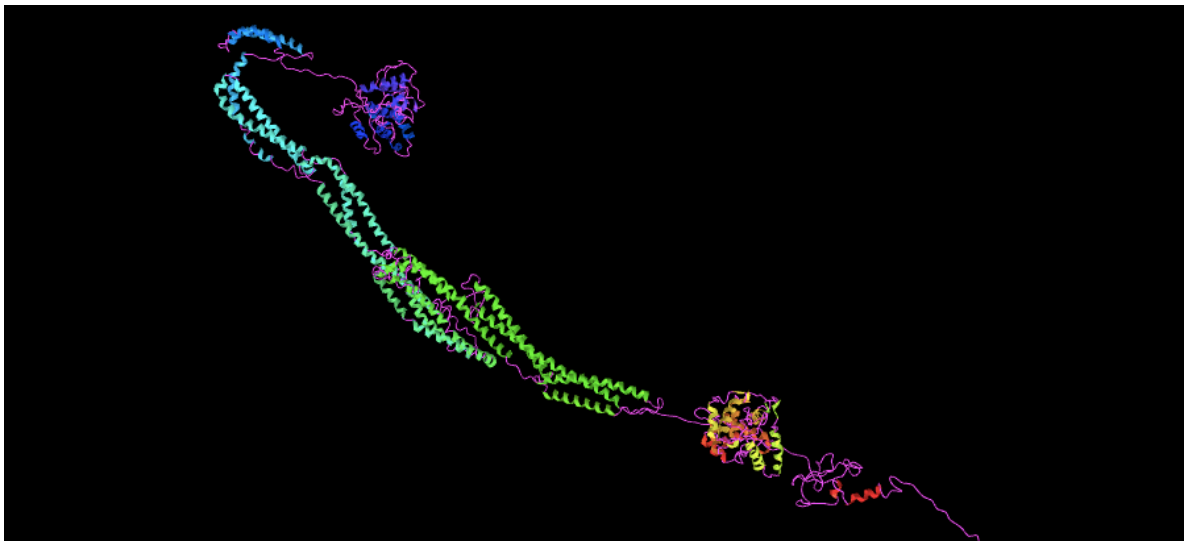


Figure 3.7. Predicted 3D model of Del19-55 dystrophin, modelled on intensive mode (de novo) on The Phyre2 web portal. Model presented with a colour scheme that draws protein as a spectrum from blue, green, yellow and orange to red, where the n-termini of proteins are coloured in blue and the c termini red.

Based on *in-silico* analysis, no obvious reasons were found to expect that the Del19-55 truncated dystrophin would not be expressed and possess functionality. Nevertheless, the only way to confirm this would be by developing a positive control expressing this protein to test *in-vitro* and *in-vivo*.

3.3. VALIDATION OF *IN-SILICO* PROTEIN ANALYSIS THROUGH DEVELOPMENT AND ASSESSMENT OF A DEL19-55 *DMD* CDNA CONSTRUCT.

The positive control would express Del19-55 h*DMD* cDNA emulating the truncated protein that would be expressed after deleting exons 19 to 55.

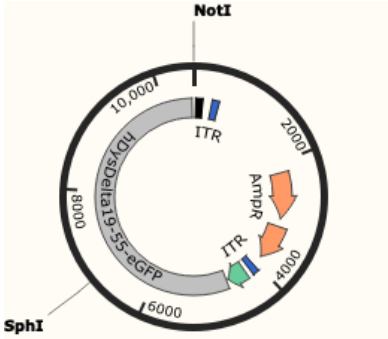
To build this construct, a g-bock with the following sequence (5´ to 3´) containing the *de novo* junction of exons 19 (blue) and 55 (green) was designed and ordered from IDT:

taatggaaacagtaactacggtgaccacaagggaacagatcctggtaaagcatgctcaagaggaacttccaccaccacct
ccccaaaagaagaggcagattactgtggattctgaaattaggaaaaggttggatgttgatataactgaacttcacagctgg
attactcgctcagaagctgtgttgcagagtcctgaatttgcaatctttcggaaggaaggcaacttctcagacttaaaagaaa
aagtcaatgacctccaaggtgaaattgaagctcacacagatgtttatcacaacctggatgaaaacagccaaaaaatcctga
gatccctggaaggttccgatgatgcagtcctgttacaaagacgtttggataacatgaacttcaagtggagtgaacttcggaa
aaagtctctcaacattaggtcccatttggaagccagttctgaccagtggaagcgtctgcacctttctctgcaggaacttctggt
gtggctacagctgaaagatgatgaattaagccggcaggcacctattggaggcgactttccagcagttcagaagcagaacg
atgta

Then, a plasmid expressing a Del44-55 human dystrophin with a fused GFP (pCI-CMV-hDysGFP-Del44-55) was digested with NaeI and SphI to recover the vector backbone and insert the g-block with exons 19 and 55 cDNA junction, as shown on Figure 3.8.

After cloning the g-block and confirming plasmid integrity by restriction digests (Fig. 3.9), the construct (pCI-CMV-hDysGFP-Del19-55) was transfected into HEK293T cells to

confirm expression, in parallel with a GFP positive control plasmid. Since this plasmid has a GFP fused to the h*DMD*, it was possible to confirm expression by fluorescence microscopy as confirmed on Fig. 3.11 and by FACS analysis. After confirming correct expression, the CMV promoter was swapped for an Spc512 muscle specific promoter lifted from pAAV-Spc512-hDys-Del44-55 plasmid (diagnostic restriction digest on Fig. 3.9). The complete cloning strategy can be observed in Figure 3.8. Correct insertion of the g-block and integrity of Sp512 promoter after cloning was confirmed by sequencing as shown on Fig. 3.10.

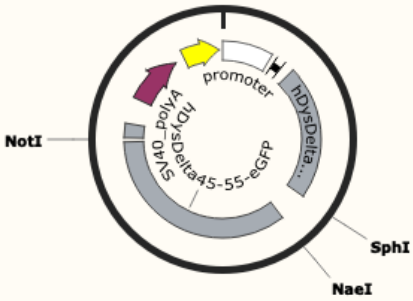


pAAV-Spc512-hDys-Del19-55-GFP
10,536 bp

INSERT FRAGMENT
Replace
SphI (6742) – NotI (0)

Insert
SphI (3279) – NotI (7073)

pAAV-hDys-Del45-55-Spc512
13,957 bp



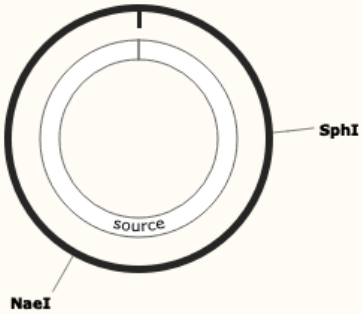
pCI-hDys-Del19-55-GFP
9444 bp

DELETE BASES
Remove
11,086 .. 12,867

INSERT FRAGMENT
Replace
SphI (3279) – NaeI (7893)

Insert
SphI (54) – NaeI (522)

Spc512-hDys
15,739 bp



pCI-hDysGFP delta 45-55
13,590 bp

G-block Exon Del19-55
575 bp

Figure 3.8. Cloning strategy to build pAAV-Spc512-hDys-Del19-55-GFP. Final construct pAAV-Sp512-hDys-Del19-55-GFP (on the top left) was built by ligating a backbone from pCI-hDysGFP-Del-44-55 digested with restriction enzymes NaeI and SphI and a g-block containing the junction of exon 18 and exon 56. Then, pCI-hDys-Del19-55-GFP CMV promoter was swapped for an Spc512 promoter from pAAV-hDys-Del45-55-Spc512 with SphI and NotI restriction enzymes.

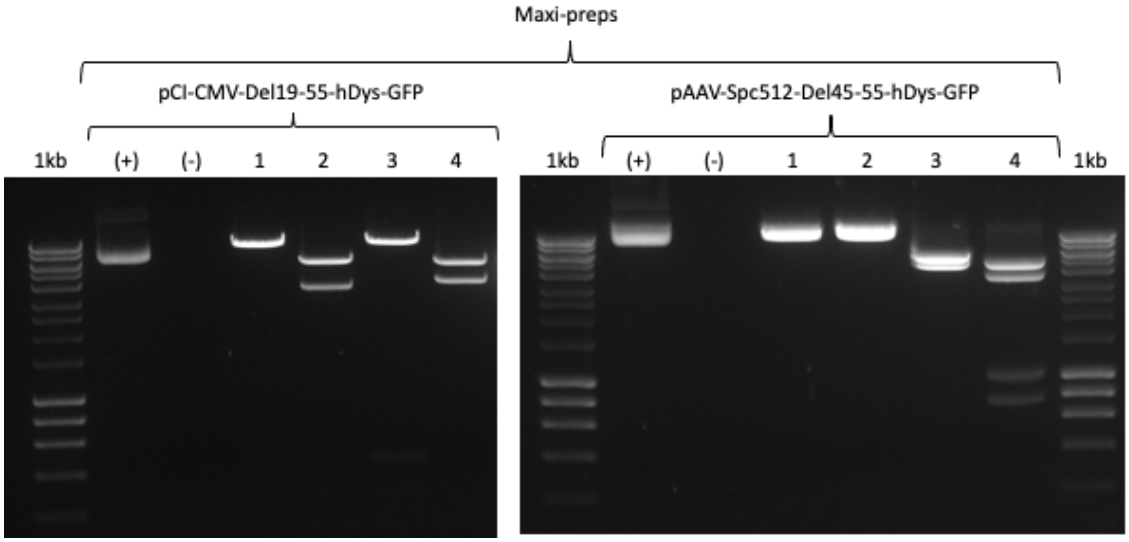


Figure 3.9. Gel Image from pCI-CMV-Del19-55-hDys-GFP and pAAV-Spc512-Del19-55-hDys-GFP (maxi-preps) restriction digestion. 1% (w/v) agarose gel with 0.5X SYBR Safe in 1X TAE (Tris-Acetate–EDTA Buffer). From left to right (pCI-CMV-Del19-55-hDys-GFP digest) bands matching expected sizes: Hyperladder I from Bioline, positive control (undigested plasmid), negative control (enzyme only), Lane 1 - MluI: 1. 9,444 bp. Lane 2 - HindIII: 1. 5,868 bp, 2. 3,576 bp, Lane 3 – SphI+NaeI: 1. 8,976 bp, 2. 468 bp, Lane 4 – SphI+NotI: 1. 5,650 bp, 2. 3,974. From left to right (pAAV-Spc512-Del19-55-hDys-GFP digest): Hyperladder I from Bioline, positive control (undigested plasmid), negative control (enzyme only), Lane 1 - NotI: 1. 10,536 bp. Lane 2 - SpeI: 1. 10,536 bp, Lane 3 – MfeI: 1. 5,673 bp, 2. 4,863 bp, Lane 4 – MscI: 1. 4,820 bp, 2. 3,803, 3. 1,105 bp, 4. 808 bp.

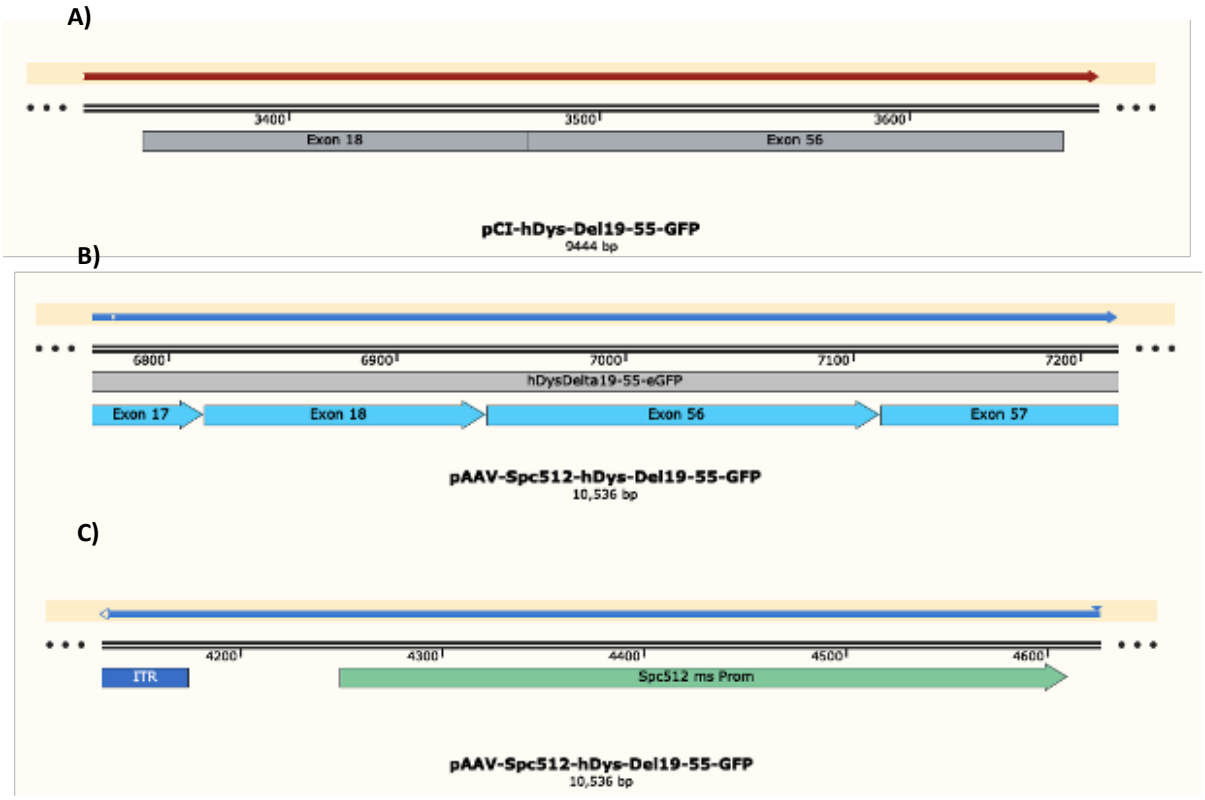


Figure 3.10. Alignment of plasmid sequencing and plasmid maps on SnapGene Software. Correct g-block insertion (exons 18 and exon 56) into A) pCI-CMV-Del19-55-hDys-GFP (sequencing trace from forward primer in red) and B) pAAV-CMV-Del19-55-hDys-GFP (sequencing trace from reverse primer in blue). C) Spc512 promoter cloned into pAAV-CMV-Del19-55-hDys-GFP (sequencing trace from reverse primer in blue). Numbers indicate bp position. Plasmid size indicated below plasmid name tag. Sequencing primers: Exons 18-56 FW: 5´-AAT GGA AAC AGT AAC TAC GGT G-3´, Exons18-56 RV: 5´-AAT ACC GGT ACA GCA TGG TGG CGA AT-3´, Spc512 promoter RV: 5´-TCA TAA CAG TCC TCT ACT TCT TCC-3´.

3.4. *In-vitro* assessment of positive controls: pCI-CMV-hDys-Del19-55-GFP and pAAV-Spc512-hDys-Del19-55-GFP.

3.4.1. Fluorescence microscopy and FACS Analysis to confirm GFP expression from positive control plasmid (pCI-CMV-hDys-Del19-55-GFP).

A CMV promoter drove expression of Del19-55-hDys, which is fused to a GFP, in transfected HEK293T cells. GFP expression was confirmed by fluorescence microscopy, as shown on Fig. 3.11 and compared to positive control plasmid pCI-CMV-GFP (expression by the same promoter is easily comparable).

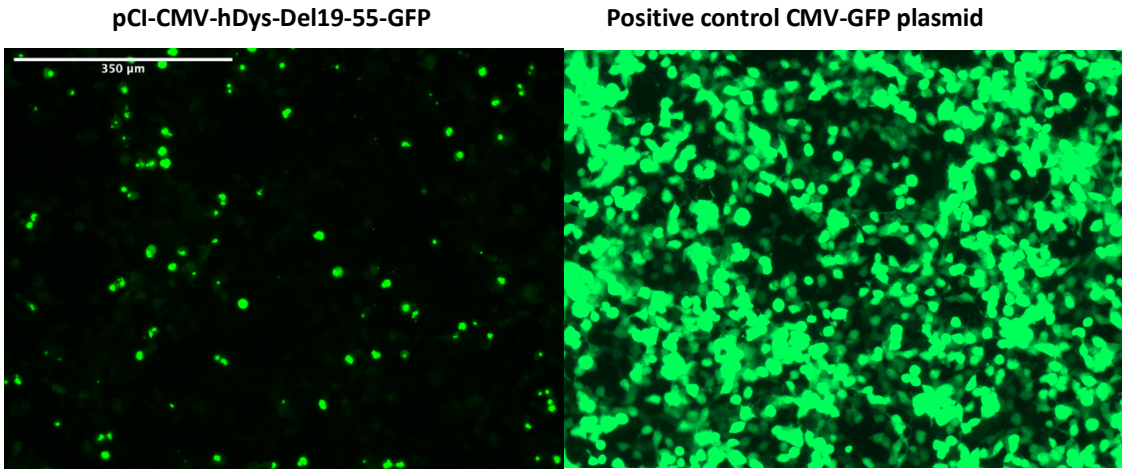


Figure 3.11. Representative fluorescence microscopy images of HEK293T cells transfected with a positive control GFP plasmid (pCMV-GFP) and pCI-CMV-hDys-Del19-55-GFP, confirming expression of the Del19-55 construct. Transfection performed with Viafect transfection reagent (from Promega) and 4 µg of plasmid DNA (4:1 Viafect to DNA). 48 hours after transfection cells were imaged with a Zeiss fluorescence microscope (Axio Vision D1 with AxioCam MRm) and 5 images per well were acquired with Software ZEN 2012. Magnification bar (top left corner) = 350 µm.

Transfection efficiency was then quantified by FACS Analysis on a FACS Canto II machine (from BD Biosciences). The gating of live cells, single cells and fluorescent cells used for the analysis was performed using FACSDiva Software and can be seen in Fig. 3.12. The final data analysis to quantify GFP positive cell populations was performed using FloJo Software and results are shown in Fig. 3.13. According to the analysis, ~8% of cells were expressing GFP after transfection with pCI-CMV-hDysGFP-Del19-55 compared with ~94% after transfection with positive control (pCI-CMV-GFP).

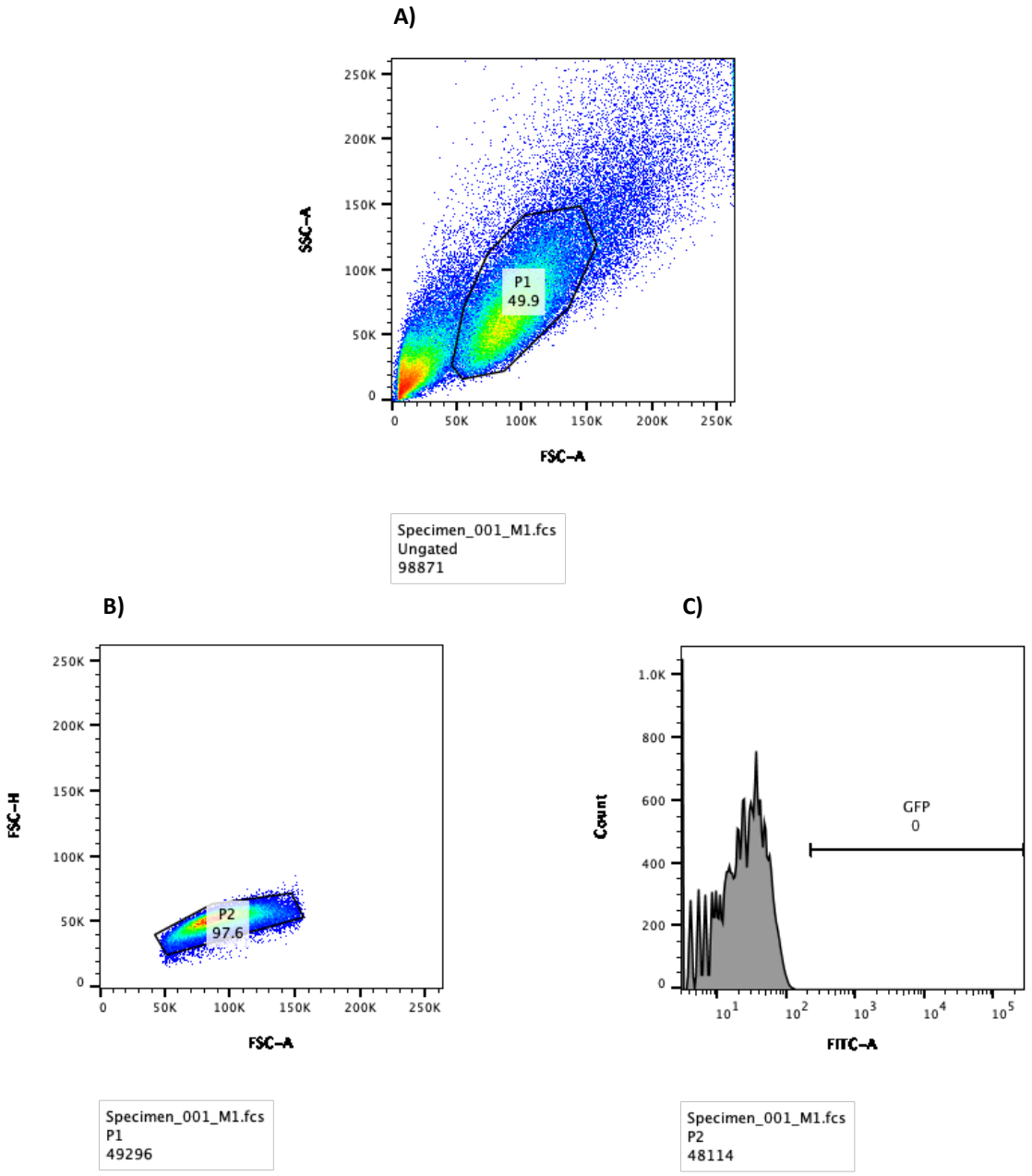


Figure 3.12. Gating for FACS Analysis of HEK293T cells mocks (from transfection with Viafect) performed on FACSDiva Sofware. Cells harvested 48 hours after transfection and processed on FACS Canto II machine (from BD Biosciences). A) Dot plot with gate from total cells to live cells (Population 1, P1), B) dot plot with gate for single cells (Population 2, P2) and C) Gate on Histogram to select fluorescent population (Population 3, P3). Below every graph, the cell count for each gated population is indicated.

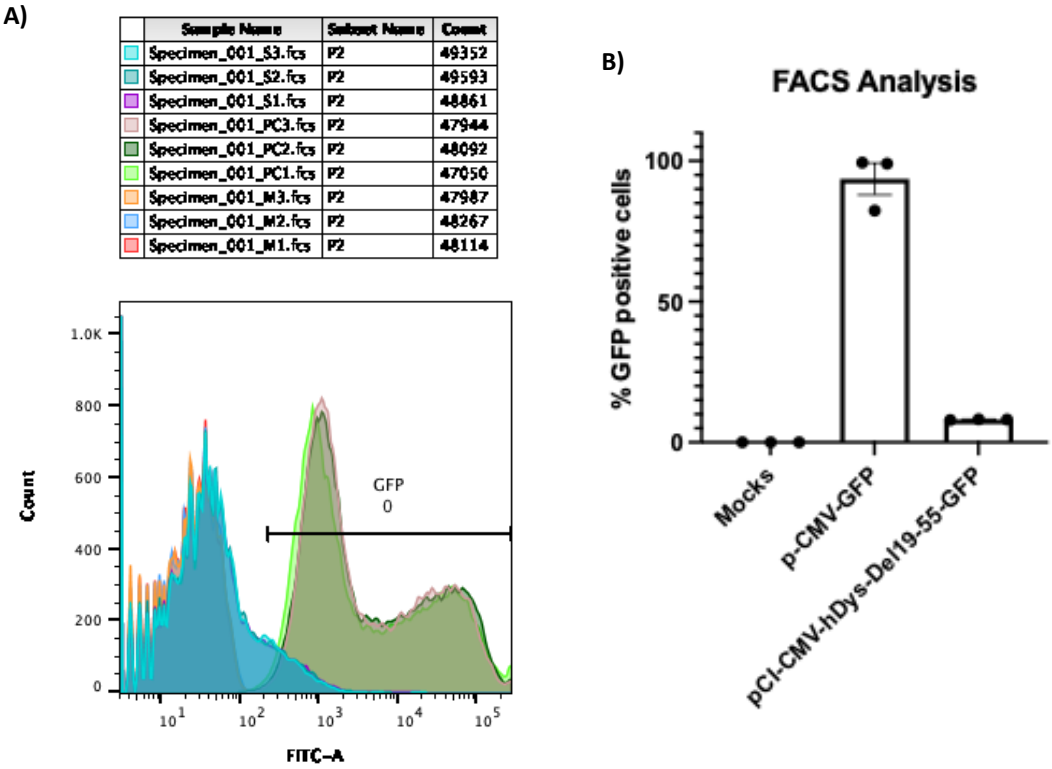


Figure 3.13. FACS Analysis from HEK293T cells transfected with pCI-CMV-hDys-Del19-55-GFP using Viafect (dose 4 µg of DNA 1:4 to Viafect, cells harvested 48 hours after transfection). A) Table and histogram overlay showing single cell counts for mock (Specimen_001_M1-M3), positive control p-CMV-GFP (Specimen_001_PC1-PC3) and samples from pCI-CMV-hDys-Del19-55-GFP transfection (n = 3 technical repeats) (Specimen_001_S1-S3), histograms under "GFP gate" show populations expressing GFP. B) Bar chart of GFP positive cells percentages. On average: 93.6% of cells expressed GFP after transfection with p-CMV-GFP and 8.1% of cells transfected with pCI-CMV-hDys-Del19-55-GFP expressed GFP. Analysis done on FloJo Software, percentages calculated on Excel and graph done on Prism 9.

Once expression of pCI-CMV-hDysGFP-DEL19-55 was established, the CMV promoter was swapped for the muscle-specific Spc512 promoter as described in Section 3.3.

3.4.2. Western Blotting to confirm Del19-55 dystrophin expression from pCI-CMV-hDys-Del19-55-GFP and pAAV-Spc512-hDys-Del19-55-GFP.

To confirm Del19-55 dystrophin expression from plasmids pCI-CMV-hDys-Del19-55-GFP and pAAV-Spc512-hDys-Del19-55-GFP and compare its size against full-length dystrophin and micro-dystrophin MD1, permissive cell line HEK293T (that expresses no dystrophin) was transfected with pCI-CMV-hDysGFP-DEL19-55, pCI-Spc512-hDysGFP-DEL19-55, a plasmid expressing MD1 under a CMV promoter and three constructs expressing full length dystrophin (full length under an Spc512 promoter, codon optimised full length dystrophin under a CMV promoter and codon optimized full length dystrophin under an Spc512 promoter). Cells were harvested for protein extraction and Western Blotting. Manex1011C primary antibody that binds to Exon 10 and 11 was used to detect expression of dystrophin. The anticipated translation of Del19-55 dystrophin was obtained from Expasy (*ExPASy - Translate Tool*, n.d.) and the protein weight calculation performed using the online Protein Weight Calculator (*Protein Molecular Weight*, n.d.), resulting in 224.04 kilodaltons from the 1952 residue sequence of Del19-55 dystrophin (sequence found in Section 2.1.1). Bands matching the expected protein size can be seen from transfected samples by duplicates (Fig. 3.14), confirming Del19-55 dystrophin expression from both constructs. Bands expressing full length dystrophin can be observed at the expected size of 427 kDa from the codon optimized full length dystrophin constructs (Meng et al., 2022) as well as smaller bands from MD1 at the expected size of 138 kDa (Le Guiner et al., 2017).

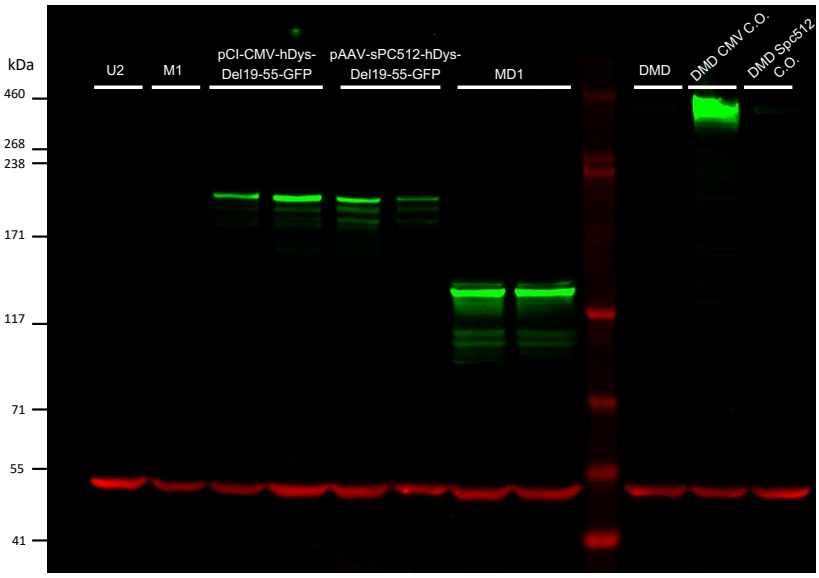


Figure 3.14. Western Blot to detect dystrophin from protein samples extracted from HEK293T cells transfected with pCI-CMV-hDysGFP-DEL19-55, pCI-Spc512-hDysGFP-DEL19-55, a plasmid expressing MD1 under a CMV promoter and three constructs expressing full length dystrophin (full length under an Spc512 promoter, codon optimised under a CMV promoter (Meng et al., 2022) and codon optimized under an Spc512 promoter donated from George Dickson lab). Translation of Del19-55 *DMD* protein was anticipated on Expasy (*ExPASy - Translate Tool*, n.d.), protein molecular weight was calculated on the online Protein Weight Calculator (*Protein Molecular Weight*, n.d.), resulting in an expected size of 224.06 kilodaltons. Bands from samples treated with pCI-CMV-hDys-Del19-55-GFP and pCI-Spc512-hDys-Del19-55-GFP match the expected size, indicating the expression of a truncated dystrophin from both constructs. MD1 bands can be seen at their expected size of 138 kDa and 427 kDa for full length dystrophin. 50 µg of protein lysate were loaded per well in a 3-8% Tris-Acetate gel with Tris-Acetate running buffer, alongside HiMark pre-stained HMW ladder from ThermoFisher and analysed with antibodies: Manex1011C (1:100, green) for dystrophin and α-tubulin (1:10,000, red) as a loading protein control.

Based on these encouraging *in-vitro* results confirming expression of Del19-55 dystrophin, it was decided to further test our construct *in-vivo*.

3.5. *IN-VIVO* ASSESSMENT OF DEL19-55 DYSTROPHIN EXPRESSION BY PLASMID INJECTION (PAAV-SPC512-HDYS-DEL19-55-GFP) AND ELECTRO-TRANSFER ON *MDX* MICE.

The construct driven by the muscle specific Spc512 promoter expressing Del19-55 dystrophin fused to a GFP was tested *in-vivo* in *mdx* mice. Plasmid pAAV-Spc512-hDys-Del19-55-GFP was delivered by intramuscular injection of Tibialis anterior (TA) muscles and electro-transfer on two separate experiments. In the first one, 6-months old *mdx* mice (n=4 muscles) were injected with 25 µg of plasmid per muscle, previously injected with hyaluronidase (1 hour before treatment) to enhance gene transfer (Gollins et al., 2003). TA muscles were harvested 2 weeks after treatment. Muscles were sectioned and stained for dystrophin, then fluorescence microscopy was performed. In the second experiment, 1-month old *mdx* mice (n=6 muscles) were injected with different doses (2 and 20 µg) of plasmid and muscles were harvested on two time points, 7 and 14 days after injections. Each harvested TA was cut in half, one half was used for protein extraction and the other half was sectioned on a cryostat for immunohistochemistry analysis.

3.5.1. Immunohistochemistry and fluorescence microscopy of mdx Tibialis Anterior muscles injected with pAAV-Spc512-DMD-Del19-55-GFP.

From each TA muscle, 10 µm sections were fixed on slides with acetone and stained with primary antibodies for GFP, laminin, and central nuclei as described in materials and method section 2.16.

On Fig. 3.15, GFP expression from pAAV-Spc512-hDys-Del19-55-GFP can be observed on samples treated with 20 and 25 µg doses and at the two harvesting time points of 7- and 14-days post treatment. A few GFP positive fibres can be observed with the 2 µg dose after 7 days of treatment, nevertheless 14 days later with the same dose, no fibres were visible. GFP positive fibres were co-localized with laminin, indicating potential correct distribution of our truncated dystrophin.

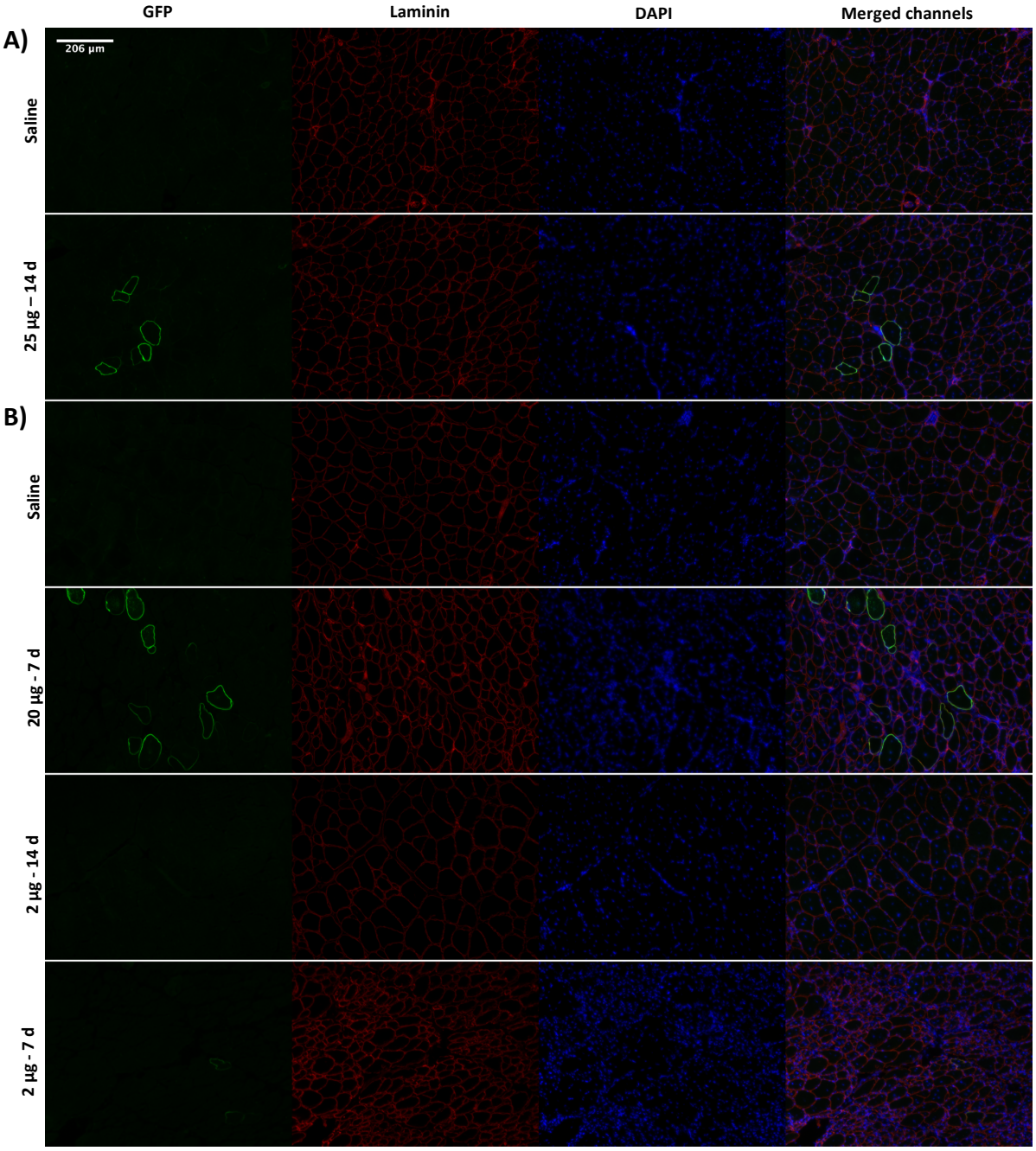


Figure 3.15. Representative images of immunohistochemistry of fixed TA muscle sections from *mdx* mice injected with pCI-Spc512-hDys-Del19-55-GFP using (from left to right) anti-GFP (1:1000, green) and laminin (1:1000, red); central nuclei counterstained with DAPI (blue) and merged channels. All sections are 10 µm thick. A) 6 months old *mdx* mice (n=4 TA muscles per group) were injected with 25 µg of plasmid per muscle, previously injected hyaluronidase (1 hour before treatment). TA muscles were harvested 14 days after treatment. B) 1 month old *mdx* mice (n=6 TA muscles per group) were injected different doses (2 and 20 µg) and muscles were harvested on two time points, 7 and 14 days after injection (shown saline harvested 7 days after treatment).

To further test correct localization of our truncated dystrophin and potential functionality, co-staining of GFP with dystrophin, α-sarcogylcan, β-dystroglycan and nNOS proteins, that would normally interact with dystrophin to form the dystrophin protein associated complex, were performed on samples treated with the 25 µg dose (harvested 14 days after treatment) of pAAV-Spc512-hDys-Del19-55-GFP.

Co-localization of GFP, dystrophin and α-sarcogylcan can be clearly observed on Figure 3.16. Co-localization of β-dystroglycan can be partially observed, but no expression of nNOS was detected. This could be due to the lack of interaction of nNOS with our truncated dystrophin, lacking spectrin-like repeats 16 and 17, which harbour the nNOS binding domain.

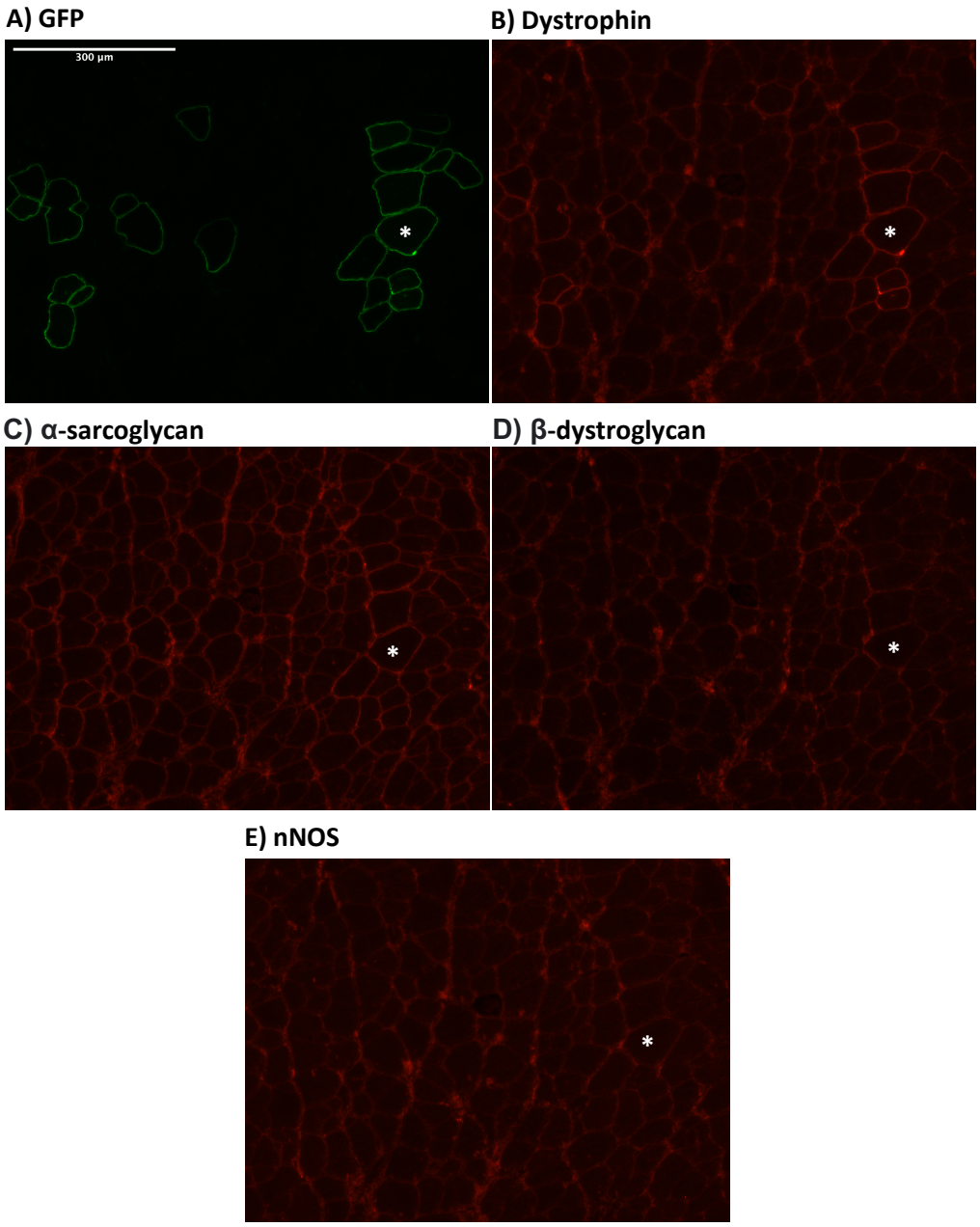


Figure 3.16. Representative fields of dystrophin-associated protein complex immunohistochemistry in TA sections from *mdx* mice after injection of 25 µg of pCI-Spc512-hDys-Del19-55-GFP; white asterisks identify the same fibres in serial sections that are stained for GFP (anti-GFP, 1:1000, green), dystrophin (Manex1011C, 1:50, red), α-sarcogylcan (α-SG, 1:50, red), β-dystroglycan (β -DG, 1:50, red) and nNOS (anti-nNOS, 1:50, red).

3.5.2. *Dystrophin positive fibres 14 days after pAAV-Spc512-DMD-Del19-55-GFP plasmid injection with a 25 µg DNA dose.*

Dystrophin positive fibres from sections treated with the 25 µg dose (harvested 14 days after treatment) were manually counted and percentage of dystrophin positive fibres was calculated based on total fibres (counted over whole section (n=2 muscles) with the FIJI Software/MuscleJ plugin, obtained percentages are presented on Fig. 3.17. Even though there was a slight increase in positive fibres (~0.5%) on samples injected with pAAV-Spc512-hDys-Del19-55-GFP, the levels of dystrophin expression were too low to expect any beneficial effect.

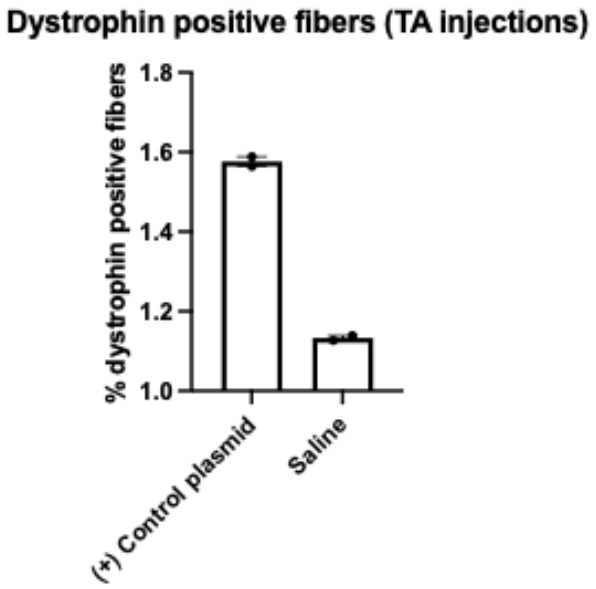


Figure 3.17. Dystrophin positive fibre percentages found on TA sections of 6-months old *mdx* mice injected with pAAV-Spc512-hDys-Del19-55-GFP and saline injections. Fibre count over whole sections (n=2 biological repeats per group). Fibres considered positive when >50% of the fibre was dystrophin positive. Positive control showed a slight increase in positive fibres when compared to saline group (~1.6% vs ~1.1%). Error bars represent standard error of the mean. Data was plotted on Prism9 Software.

3.5.3. *In-vivo Del19-55 dystrophin expression confirmation by Western Blot from sample injected with pAAV-Spc512-DMD-Del19-55-GFP at different doses.*

After plasmid injection on *mdx* mice with a high and a low dose and harvesting 7 and 14 days after treatment, protein was extracted from one half of the TA muscles and analysed by Western Blotting with Manex1011C primary antibody. Samples with high doses, of 20 and 25 µg, show bands of approximately 224 kilodaltons matching the control sample from pAAV-CMV-hDys-Del19-55-GFP transfection on HEK293T cells, on image from Western Blot in Fig. 3.18. A bright band can be observed in one of the samples with the 25 µg dose, a fainter band can be observed in the second sample and in one of the samples from muscle treated with the 20 µg dose. No bands can be observed in the protein extract from mice treated with the 2 µg dose of plasmid or the saline samples.

From this experiment it can be concluded that a dose of 20-25 µg is enough to see Del19-55 dystrophin expression on Western blots after plasmid injection and electro-transfer on *mdx* mice, when harvesting muscles 7 or 14 days later after treatment. But a low dose, of 2 µg, is not enough to detect protein expression from pAAV-Spc512-hDys-Del19-55-GFP by Western Blots.

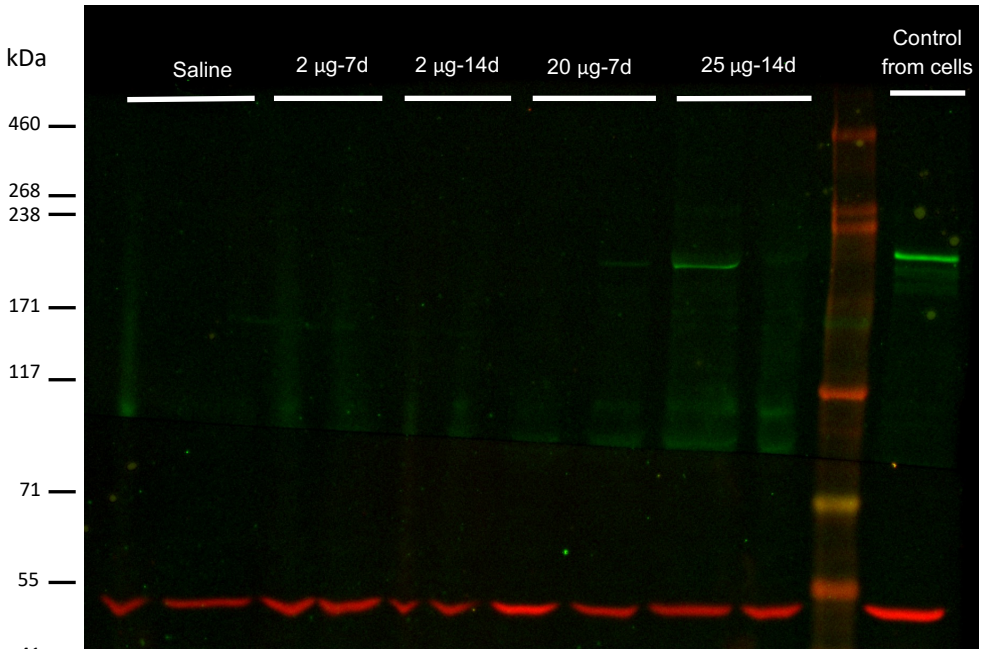


Figure 3.18. Western Blot to detect dystrophin from TA muscle samples treated with pAAV-Spc512-hDys-Del19-55-GFP. From left to right (n=2): Saline samples, 2 µg dose harvested 7 and 14 days later, 20 µg dose harvested 7 days later and 25 µg dose harvested 14 days later; HiMark pe-stained ladder from ThermoFisher and control from protein sample extracted from transfected HEK293T cells with pAAV-CMV-hDys-Del19-55-GFP. 30 µg of protein lysate per well were loaded and analysed with antibodies: Manex1011C (1:100, green) for dystrophin and α-tubulin (1:10,000, red) as a Loading control.

3.6. DISCUSSION.

The first part of this chapter included a review of clinical cases with large (>15 exons) deletions incorporating exon 19. Deletion of exons 19 to 55 has not been reported in clinic yet. Nevertheless, some cases of large (>31 exons) in-frame deletions of patients with BDM phenotype were identified, such as Del13-55 (Lim et al., 2020) and Del19-51 Agarwal et al., 2017). Is important to note that these deletions were not selected as a target in this project for mainly two reasons: exon 17 needed to be maintained to express hinge 2 and ideally mutational hotspot of exons 45-55 would be deleted. To achieve this while maintaining exon phasing in-frame, exons 19-55 were selected as a target.

Del 19-55 dystrophin would share the features expressed by MD1: Hinges 1, 2 and 4, spectrin-like repeats 1 to 3 and 24. Unlike MD1, our truncated dystrophin would still possess the CT-terminal domain, that mediates sarcolemma localization by binding to dystrobrevin and syntrophin (Sadoulet-Puccio et al., 1997) and spectrin-like repeats 22 and 23, allowing for expression of the dp116 dystrophin isoform. This non-muscle isoform expresses mainly in Schwann cells (Byers et al., 1993) and although it lacks actin-binding domains and hence a mechanical function, it assembles the dystrophin-glycoprotein complex and has shown to prevent severe aspects of *mdx:utrn*$^{-/-}$ mice phenotype by improving muscle mass increase through alternative interaction between the DGC and the cytoskeleton. However, the mechanism is not clear and its expression

by itself does not improve histopathology or mechanical properties of muscles (Judge et al., 2011).

Furthermore, it is relevant to note that some of the features lost in the truncated Del19-55 dystrophin resulting from our deletion, such as Hinge 3 and spectrin-like repeats 16-17, would improve functionality further. In a study, domain composition of micro-dystrophins was examined *in-vivo* on *mdx* mice and it was found that the hinge regions can profoundly influence functionality. It was concluded that even though micro-dystrophin with hinge 2 significantly prevented muscle degeneration, a similar version containing hinge 3 protected muscle more effectively from turnover (Banks et al., 2010). Additionally, the nNOS domain harboured in spectrin-like repeats 16-17 has shown to play a critical role in normal muscle physiology (Lai et al., 2009). Considering this, if expression of Del19-55 dystrophin shows no improvement in muscle physiology, an additional strategy that could be tested, is the delivery of a repair template expressing the nNOS binding domain and hinge 3 to potentially improve protein functionality.

An *in-silico* analysis of Del19-55 dystrophin was performed to predict the protein model and estimate potential functionality. From this analysis it was concluded that Del19-55 dystrophin has potential to be functional considering its protein structure. However, protein modelling is not always 100% accurate so results were confirmed with *in-vitro* and *in-vivo* analysis. Results from these experiments confirmed that Del19-55 truncated dystrophin is expressed and has potential to be functional as it co-localizes with α-

sarcogylcan and β-dystroglycan at the sarcolemma, as shown by the immunohistochemistry analysis.

Human full-length dystrophin has been delivered as a cDNA construct by intramuscular injection to *mdx* mice (Acsadi et al., 1991). Although this non-optimized construct expressed in ∼1% of myofibres, results suggested that an exogenous dystrophin could correct effects of dystrophin deficiency on *mdx* mice (Acsadi et al., 1991). This led to gene addition studies using mini and micro-dystrophins, that showed that low level expression of a functional truncated dystrophin can slow down muscle degeneration in *mdx* mice (Vincent et al., 1993). When delivered in a viral vector, Del17-48 dystrophin mini-gene expressed 5-20% dystrophin of control levels and this was enough to prevent development of dystrophic symptoms in *mdx* mice (Phelps et al., 1995). In a more recent study using exon skipping, it was shown that 15% of homogenous dystrophin expression was sufficient to protect against contraction-induced injuries in muscle and slow down disease progression (Godfrey et al., 2015). Furthermore, the same study established that changes in muscle strength in *mdx* mice are proportional to dystrophin expression levels. Another study exploring how dystrophin levels relate to neuromuscular junction (NMJ) function and morphology, in *mdx-Xist*$^{\Delta hs}$ mice (mouse model expressing variable low full-length dystrophin levels), established that 19% is the minimal dystrophin level required for normal NMJ function and morphology when dystrophin expression is not uniform (Van der Pijl et al., 2018). A recent publication suggests that approximately 20% expression of uniformly distributed dystrophin within skeletal muscles and the heart

may be sufficient to prevent disease progression, based on data review from humans and animal models studies (Wells, 2019). Based on these levels of expression, it was not expected to see a functional effect from pAAV-Spc512-DMD-Del19-55-GFP plasmid injection, as only 1.5% of dystrophin positive fibres were detected after treatment. It is relevant to note that the Del19-55 dystrophin construct used for the experiments presented in this chapter is not codon optimised. Regardless, it was possible to detect its signal *in-vitro* and *in-vivo*.

To evaluate potential effects of Del19-55 dystrophin in muscle functionality, an experiment with a different delivery system would be needed to increase delivery efficiency. Since pAAV-Spc512-DMD-Del19-55-GFP it too large to be packaged into an AAV vector, other delivery systems such as lentiviral vectors or nanoparticles could be considered. Lentiviral vectors with a muscle specific promoter (CK9) have been used to deliver codon optimized full-length dystrophin into myotubes. Successfully corrected myoblasts from this experiment were then grafted into *mdx* mice and restored dystrophin in donor-derived muscle fibres (Meng et al., 2022). An alternative approach to evaluate Del19-55 dystrophin effects on phenotype would be to create a mouse model with deletion of exons 19-55 and assess the phenotype in detail.

4. Design of SaCas9 Single gRNAs targeting mouse and human *Dmd/DMD* introns 18 and 55, *in-vitro* gRNA screening & assessment of genome editing efficiency for the creation of a *de novo* intronic junction.

In nature, CRISPR/Cas systems can express multiple CRISPR arrays and Cas proteins by acquiring new spacers. These spacers and their orientation are dependent on the PAM sequence. Thus, it can be said that native CRISPR/Cas systems are multiplexed by nature (F. J. M. Mojica et al., 2009). Multiplex CRISPR technologies allow for multi-locus editing, using a single construct expressing multiple gRNAs (McCarty et al., 2020). Many of the methods used to multiplex gRNAs in the lab are based on mechanisms found in native systems and for Cas9 proteins can be narrowed to two main approaches: expressing multiple gRNAs, each one in an individual cassette containing a promoter (i.e. U6), the gRNA and a terminator, or an array expressing multiple gRNAs linked by an appropriate spacer, expressing from an individual promoter (McCarty et al., 2020), as depicted in Fig. 4.1.

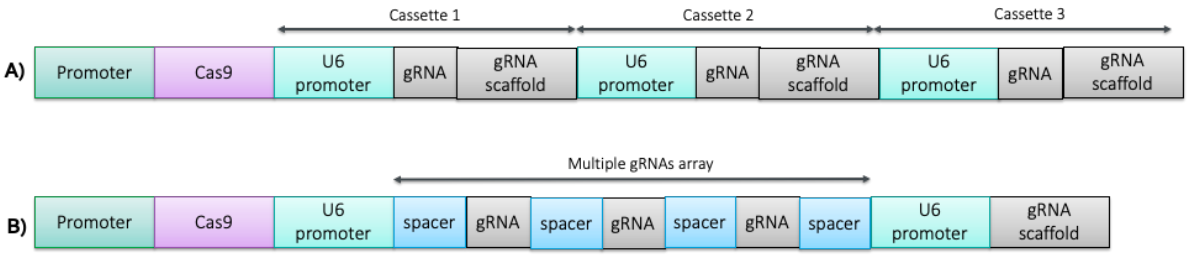


Figure 4.1. Mechanisms to multiplex various gRNAs, depicted with arrays expressing a Cas9 and three gRNAs. A) Each gRNA on an individual cassette driven by its own U6 promoter. B) A gRNA array linked by appropriate spacers driven by a U6 promoter and the gRNA scaffold expressed separately. Adapted from (McCarty et al., 2020).

Multiplexing kits are commercially available to clone multiple gRNAs in individual cassettes for *Sp*Cas9 systems, by modular assembly using golden gate cloning (Sakuma et al., 2014). However, commercial kits are not yet available for multiplexing gRNAs on an *Sa*Cas9 systems ready to be packaged into AAV vectors.

For this chapter the main objective was to establish a multiplex *Sa*Cas9 system with two gRNAs, targeting intron 18 and 55 to achieve the deletion of exons 19-55 and test it *in-vitro* and *in-vivo*. The experimental milestones required to achieve establishment of this system are summarized below:

- Design single guide RNAs (sgRNAs) for *Staphylococcus aureus (Sa)*Cas9 targeting introns 18 and 55 of the *DMD/Dmd* gene to human and mouse sequences by using on-line design tools: Benchling (*CRISPR GRNA Design Tool | Benchling*, n.d.), CRISPOR (Concordet & Haeussler, 2018) and The Broad Institute Online

Tool (*SgRNA Designer: CRISPRko*, n.d.). Then assess potential efficiency and off-target events as predicted by *in-silico* analysis.

- Clone designed sgRNAs into a plasmid expressing an *Sa*Cas9: gRNAs were cloned into plasmid pAAV-CMV-SaCas9 (pX601-AAV-CMV::NLS-SaCas9-NLS-3xHA-bGHpA;U6::BsaI-sgRNA, #6159 from Addgene). Confirm integrity of the constructs by restriction digests and sequencing.
- Optimise transfection protocols by performing a dose response experiment with an *Sa*Cas9 construct.
- Screen gRNAs *in-vitro* in appropriate human and mouse cell lines respectively and evaluate their cutting efficiency by transfecting gRNAs into HEK293T and N2A cell lines respectively, extracting DNA and performing PCRs using primers to amplify the region flanking each target. Then, sequence PCR products and assess cutting efficiency on the TIDE online tool based on the sequencing traces.
- Design a construct on an AAV plasmid backbone expressing an *Sa*Cas9 under the control of the synthetic muscle specific promoter (Spc512) (X. Li et al., 1999) to multiplex two gRNAs: design G-blocks to clone two cassettes for multiplexing, each one with a U6 promoter, a gRNA scaffold, a restriction site to clone a gRNA in and a terminator (similar to design depicted in Fig. 4.1.A). Each cassette with a unique restriction site, so gRNAs could be cloned in sequentially on the same construct. This construct would allow delivery of a selected pair of gRNAs in the same plasmid, rather than performing a co-transfection of plasmids with each gRNA, and its size would allow for packaging into AAV vectors.

- Test multiplexed construct for efficiency for creation of the desired *de novo* intronic junction *in-vitro* in N2A and C2C12 cells.

4.1. ESTABLISHING AN SACAS9 SYSTEM.

4.1.1. gRNA DESIGN TARGETING INTRONS 18 AND 55 AND PREDICTED OFF-TARGET ASSESSMENT.

Different CRISPR systems were compared (Table 4.1). *Sa*Cas9 and Cpf1 have translational potential for the deletion of Exons 19-55 since these Cas proteins are smaller than *Sp*Cas9 and their respective cDNAs would be packageable into an AAV vector, along with both gRNAs required for the deletion. Since our strategy did not include the use of a repair template, overhangs at the cut site were not needed, hence an *Sa*Cas9 system was selected for our deletion strategy.

Table 4.1. Comparisons of different CRISPR systems. Comparative of different Cas proteins including the size of their cDNAs, cut type and PAM sequences.

Cas9	Size	Cut	PAM
SpCas9	4 kb	Blunt End	NGG
SaCas9	3 kb	Blunt End	NGRRT
CjCas9	2.9 kb	Blunt End	NNNNACAC
Cpf1	3.9 kb	5´ Overhang	TTTN

*Sa*Cas9 gRNAs were designed to target intron 18 and intron 55 of the human and mouse *DMD/Dmd* genes. Originally the goal was to design gRNAs applicable for both human and mouse by targeting homologous sequences. To find these regions, human and mouse intron 18 and intron 55 sequences were aligned on the EMBOSS Online Tool (Madeira et al., 2022). Intron 18 only had three partially homologous regions ≥ 20 bp (Table 4.2) and only eight short homology regions ≥ 19 bp were found in Intron 55 (Table 4.3). Full alignments can be found on Appendix A.

These regions were screened for potential *Sa*Cas9 gRNA targets using online gRNA design tools: Benchling (*CRISPR GRNA Design Tool | Benchling*, n.d.), CRISPOR (Concordet & Haeussler, 2018) and The Broad Institute Online Tool (*SgRNA Designer: CRISPRko*, n.d.); alongside intronic regions located ≥200 bp upstream/downstream of the 3´-end/5´-end of introns 18 an 55 to avoid gRNA targets close to exonic sequences, that could potentially cause exon disruption.

Only one potential target for each intron was found within these homologous regions. The gRNAs designed for these targets and their respective PAM sequence are highlighted in Table 4.2-4.3 and presented alongside human and mouse gRNAs designed to target intron 18 and 55 in Table 4.4.

Table 4.2. Human and mouse intron 18 alignments showing partially homologous regions (≥ 20 bp) aligned on EMBOSS Online Tool. The numbers indicate bp position in the intronic sequence. Partially homologous regions highlighted in yellow. Target sites for *Sa*Cas9 sgRNAs within these regions are highlighted in red text and PAM sequence is indicated in bold text next to respective guide.

Homology region	Alignment of Intron 18
1	HUMAN DMD 1075 AA---AGCTATTTTAAATTACTTATTAGCTTTATA--AGACATGCTGTTG 1119 || ||.||||||.||||.|.||||||||.|||| |||.|||.|.||. MOUSE DMD 1249 AACATAGTTATTTTGAATTTCATATTAGCTGTATATTAGATATGATTTTA 1298
2	HUMAN DMD 3700 CTACTTTAG-TCGAAATAATATTTCTCAAATTGTGGGTATTTGTGCTCAT 3748 .|| ||||| ||.|||.|.||.||||.|||.|.|...||||.||.|.||| MOUSE DMD 4876 TTA-TTTAGATCAAAACAGTACTTCTAAAAGTATATATATTGGTACCCAT 4924
3	HUMAN DMD 16018 AAAAGTGTTGAGAAAA--AGTCTT-TAGATTCACGTGATAAGCTGA**CAGA** 16064 .|.|||||| |||| |||.|| .|||||.||.||||.||.|||.|.| MOUSE DMD 16887 --TATTGTTGA-AAAATCAGTATTAAAGATTTACATG**ATGAGT**TGATAAA 16933 HUMAN DMD 16065 **GT**GAAACATCTTAAGGCTTGAAAGGGCAAGTAGAAGTTATAATTATTGTG 16114 .||||..|||..|||..|||||||...||.||...|||||.|||||.|||. MOUSE DMD 16934 ATGAAGTATCAGAAGAATTGAAAAATCAGGTTACAGTTACAATTACTGTT 16983

Table 4.3. Human and mouse intron 55 alignments showing partially homologous regions (≥ 19 bp) aligned on EMBOSS Online Tool. The numbers indicate bp position in the intronic sequence. Partially homologous regions are highlighted in yellow. Target sites for *Sa*Cas9 sgRNAs within these regions are highlighted in red text and PAM sequence is indicated in bold text next to respective guide.

Homology region	Alignment of Intron 55
1	HUMAN DMD 346 TTACAGGGAAAGCATCTGTATGAATTGTCTGTTTTATTTAGCGTTGCTAA 395 ||||||||||||||||||||.|||.||||||||||||||||||||| MOUSE DMD 345 TTACAGGGAAAGCATCTGTAGGAACTGTCTGTTTTATTTAGCGTTGCTAA 394
2	HUMAN DMD 446 TGGCATTTTGTAGCTTTCTTCCTAACATGATCTGTGAAAAT**AAGAAT**GAG 495 ||.|||||||||| ||||||||||||||||||||||||||||||||| MOUSE DMD 445 TGACATTTTGTAG-TTTCTTCCTAACATGATCTGTGAAAAT**AAGAAT**GAG 493
3	HUMAN DMD 496 ATGGCTGAATTTGTCGTAGTTAATGATCAAACAATTTTCAGACAATTGTT 545 ||.|||.|||||||..||||||.||.|....|||||||||||||||| MOUSE DMD 494 ATTGCTAAATTTGTTATAGTTAGTGGTTGTGCAATTTTCAGACAATTGTT 543
4	HUMAN DMD 17073 CA--TTATAATCAATTTCTCAAAAGTAAAGTTAATCAAGAGAAGGAAAAA 17120 .| ||||||||||||||||||||.||.||||| ||| ||||||||| MOUSE DMD 16424 AAGCTTATAATCAATTTCTCAAAATTATAGTTA----AGA-AAGGAAAAA 16468
5	HUMAN DMD 31662 GCTTTTTGCTGATGGTTTCTCTCATTTTATTATAGCTTATAGCATTGTAA 31711 ||.||.|||||.||||||||||||||||||||||||.|||||.||||| MOUSE DMD 30685 GCCTTCTGCTGGTGGTTTCTCTCATTTTATTATAGCTTGTAGCACTGTAA 30734
6	HUMAN DMD 31712 ATTAATTTAACATGAAAGGATAAAAACGTTGCTTTTGAAATGTTTCTCAT 31761 |||||||||||||||||||||||||.|||||||||||||||||| MOUSE DMD 30735 ATTAATTTAACATGAAAGGATAAAAATGTTGCTTTTGAAATGTTTCTCAT 30784
7	HUMAN DMD 31762 TAAATTATGAAAAAATATTACACTAAATAAAAGAAAGGAATGCCTCTGGT 31811 |||||||||.|||||||||.|.||.|||||||||||||||||.| MOUSE DMD 30785 TAAATTATGGAAAAATATTATAATAGATAAAAGAAAGGAATGCCTCTGCT 30834
8	HUMAN DMD 31812 ACCAGCTTCTGTTTGCTCAATTATTGCAGTACCCAAAGTGAATTATTACA 31861 ||||||||||||||||||||.|.|.|..| ||.||.|||||||.|| MOUSE DMD 30835 ACCAGCTTCTGTTTGCTCAATTGTGGAATGA---AATGTAAATTATTTCA 30881

Table 4.4. summarizes the results from different online tools: Benchling (*CRISPR GRNA Design Tool | Benchling*, n.d.), CRISPOR (Concordet & Haeussler, 2018) and The Broad Institute Online Tool (*SgRNA Designer: CRISPRko*, n.d.). Outputs from these tools were compared and gRNAs were selected if the same sequence was suggested by two or more of these tools. Then, at least five gRNAs per target were selected for screening based on their efficiency and specificity, indicated by the following scores:

- **On-target score:** refers to the activity or predicted efficiency of the guides according to an algorithm designed by (Doench et al., 2014).
- **Off-target score:** refers to specificity of the guides according to an algorithm designed by (Hsu et al., 2013).
- **MIT Specificity Score:** higher MIT specificity score, lower off-target effects in the genome. This score has been adapted for *Sa*Cas9 and based on the off-target scores shown on mouse-over. This algorithm by (Tycko et al., 2018b) is aggregated from all off-target scores and ranges 0-100.
- **Predicted efficiency score:** higher efficiency score, more likely cleavage at this position. This is a modified version of the Doench et al. (2016) score by (Najm et al., 2018) for *Sa*Cas9, with a range from 0-100.

Selected *Sa*Cas9 gRNAs targeting Introns 18 and 55 of mouse and human *Dmd*/*DMD* genes (Table 4.4) include:

- Two equivalent gRNAs for *Sa*Cas9, one for human (Guide 41) and one for mouse (Guide 42), cutting in the same region of the mouse and human intron 18 (highlighted in red on Table 4.2).
- One gRNA targeting one homology region between mouse and human *DMD/Dmd* in Intron 55 (Table 4.3, Guide 10).

Guide RNA distribution within intronic sequences can be seen on Fig. 4.2.

Table 4.4. *Sa*Cas9 gRNAs targeting Introns 18 and 55 of mouse and human *Dmd*/*DMD* genes. A) gRNAs targeting intron 18. B) gRNAs targeting intron 55. On and Off-Target Scores were obtained from Benchling. MIT Specificity Score and Predicted efficiency were obtained from CRISPOR. **indicates position at a homologous/partially homologous region.

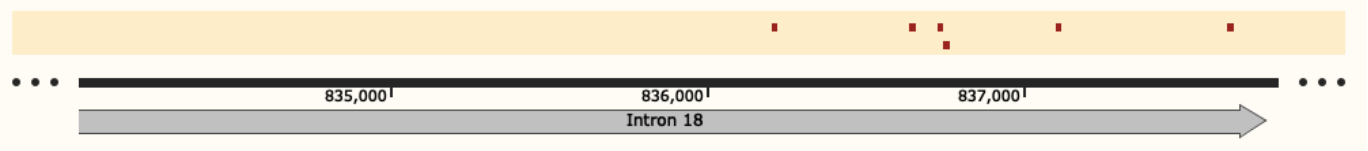
A)	Intron 18								
	GUIDE	**Position**	**Strand**	**Sequence**	**PAM**	**On-Target Score**	**Off-Target Score**	**MIT Specificity Score**	**Predicted efficiency**
Human	**1**	32502508	+	ATAGCCAGAAT TTCATACTA	TTGAGT	44.7	42.5	98	64
	2	32502954	-	TGCTGACCACCT TTCAAGTG	CTGAAT	25.2	80.7	99	65
	3	32502881	+	TGTTGAGTATA AATTTGTGC	AGGGAT	14.6	68.8	98	66
	4	32502863	+	ATGGACAGTCT GCACCACTG	TTGAGT	17.1	80.3	97	85
	5	32503389	-	GATATTGCCATA TTATATGA	AAGAGT	33.1	71.4	96	85
	41	**	+	AGATTCACGTG ATAAGCTGA	CAGAGT	47.1	88.9	100	86
Mouse	**11**	83771589	-	ACAGTATCTAGT CACTACAC	ATGAGT	29.2	84.6	100	83
	12	83772092	-	ATCACTGCCATA CTAACAGC	CTGAGT	39.3	86.7	100	86
	13	83771708	-	AGCATTCTATGA TTCAATAT	TAGAGT	13.7	41.1	99	77

	14	83772048	+	ACTTTCAGGGA ATAACGTAC	AGGAAT	22.7	93	99	78
	15	83772001	+	ATATGGGTATG AGTATACTA	CAGAAT	85	82.8	99	64
	42	**	+	CAGTATTAAAG ATTTACATG	ATGAGT	37.9	-	96	86

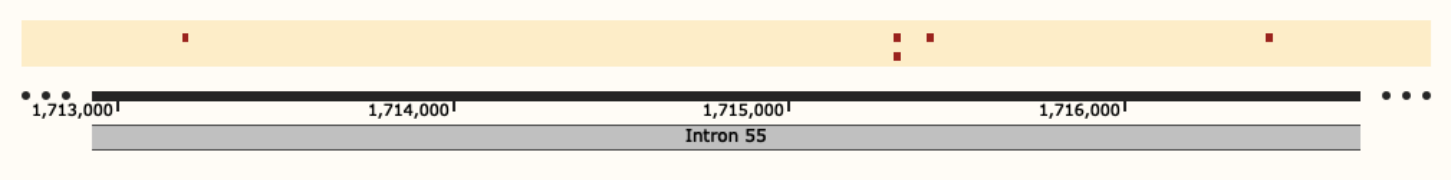
B)	**Intron 55**								
	GUIDE	**Position**	**Strand**	**Sequence**	**PAM**	**On-Target Score**	**Off-Target Score**	**MIT Specificity Score**	**Predicted efficiency**
Human	**6**	31626413	+	CATTGTCTAACC ATACATCG	AAGAGT	38.3	90.9	100	86
	7	31623167	-	TTAGTAGCACA ATTAGTACC	TTGAAT	40.8	91.8	100	68
	8	31624177	-	GAACGCCATAC AAAGCCTTT	AGGGGT	10.2	77.9	100	75
	9	31624276	-	TTACCGTCGTCC TTGTACTT	CAGGAT	52.1	90.9	100	42
	10	**	+	CTAACATGATCT GTGAAAAT	AAGAAT	33	60.7	95	64
Mouse	**16**	84647506	-	TAAACGCTGAA CTTACTTCT	CTGAGT	4.0	81.9	100	76
	17	84648943	-	GATGTCGAGCG GTTTATCAT	TGGAGT	27.1	94.3	100	77
	18	84649255	+	GTCTTAGTATAA AGTGACGA	GTGGAT	45.7	89.4	100	79
	19	84647948	-	AACCAAAAACT CAGGCGCAA	AAGAAT	15.5	87.399	99	85
	20	84647995	+	AAATGCACATC ATTGATATC	TAGAAT	5.6	78.1	99	44
	10	**	+	CTAACATGATCT GTGAAAAT	AAGAAT	33	60.7	95	64

A) Human *DMD* (full-length: 2´220,391 bp):

Intron 18 (821,602 - 837,767 bp)

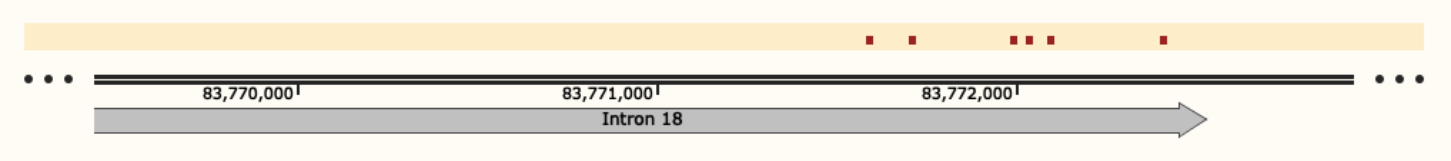


Intron 55 (1´711,937 - 1´832,156 bp)



B) Mouse *DMD:* (full-length: 2´390,387 bp):

Intron 18 (83´755,500 - 83´772,535 bp)



Intron 55 (84´647,227 - 84´773,582 bp)

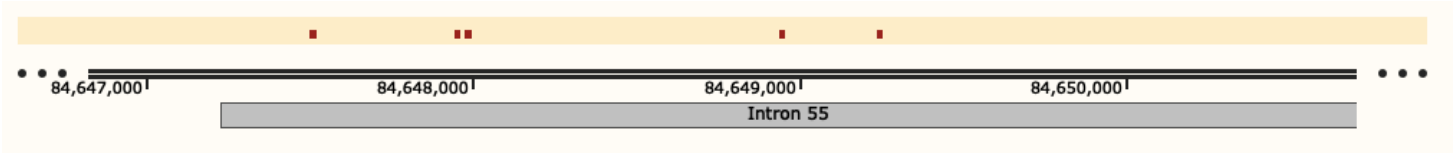


Figure 4.2. ***DMD*** **gene representative regions of intron 18 and intron 55 (5´ to 3´) aligned with gRNAs. A) Human version. B) Mouse version. With introns 18 and 55 highlighted in gray, position on the sequence indicated below black lines in number of base pairs and selected** ***Sa*****Cas9 gRNAs represented as red boxes.**

Potential off-target events with up to three mismatches on the target for each gRNA were retrieved from CRISPOR and are presented on Table 4.5 considering the following:

- For each number of bp mismatches on potential targets for each gRNA sequence (0-1-2-3), the number of off-targets is indicated. For example: 0-15-31-30- means 0 off-target with 0 mismatches, 15 off-target with 1 mismatch, 31 off-targets with 2 mismatches, etc.
- Off-targets included in Table 4.5 have no mismatches within the 12 bp adjacent to the PAM sequence, known as "PAM-proximal seed region", that has shown no mismatch tolerance in *Sp* and *Sa*Cas9 systems (Fu et al., 2014, Tycko et al., 2018).
- Potential matches for each off-target are results from Genome Browser (*UCSC Genome Browser Home*, n.d.) indicating: location in genome (intronic, exonic or intergenic region) and gene. Potential matches are ranked by Cutting Frequency Determination (CFD) Off-target Score (Doench et al., 2016) from most to least likely.

Table 4.5. Potential off-target events per *Sa*Cas9 gRNA retrieved from CRISPOR, evaluating potential off-targets with up to 3 mismatches. CRISPOR considers off-targets if they are flanked by one of these motifs (PAM sequence for SaCas9): NNGRRT, NNGRRN. In this table, (shown in grey) off-targets that have no mismatches in the 12 bp adjacent to the PAM are summarised. Off-targets for each number of potential bp mismatches on the target are indicated (0-1-2-3), i.e: 0-15-31-30- indicates 0 off-targets with 0 mismatches, 15 off-targets with 1 mismatch, 31 off-targets with 2 mismatches, etc. A) For Intron 18. B) For Intron 55. Potential matches are results from Genome Browser indicating locations in genome (intronic, exonic or intergenic region) and gene. Potential matches are ranked by Cutting Frequency Determination (CFD) Off-target Score (Doench et al., 2016) from most to least likely.

A)	Intron 18		
	GUIDE	**Off-targets for 0-1-2-3 bp mismatches**	**Potential matches**
Human	**1**	0 - 0 - 0 - 0 0 off-targets	None
	2	0 - 0 - 0 - 1 1 off-targets	Intron: TMCO3
	3	0 - 0 - 0 - 2 2 off-targets	Intergenic: between RNU6-754P and CLIC5 Intergenic: between PTH2R and RNA5SP117
	4	0 - 0 - 0 - 0 0 off-targets	None
	5	0 - 0 - 0 - 1 1 off-targets	Intron: HDAC9
	41	0 - 0 - 0 - 0 0 off-targets	None
Mouse	**11**	0 - 0 - 0 - 0 off-targets	None
	12	0 - 0 - 0 - 0 0 off-targets	None
	13	0 - 0 - 0 - 2 2 off-targets	Intron: Xpo5 Intergenic: between Gm5973 and Gm9915

	14	0 - 0 - 0 - 0 0 off-targets	None
	15	0 - 0 - 0 - 0 0 off-targets	None
	42	0 - 0 - 0 - 0 0 off-targets	None

B)	**Intron 55**		
	GUIDE	**Off-targets for 0-1-2-3-mismatches**	**Potential matches**
Human	**6**	0 - 0 - 0 - 0 0 off-targets	None
	7	0 - 0 - 0 - 0 0 off-targets	None
	8	0 - 0 - 0 - 4 4 off-targets	Intergenic: between Ttll7 and Gm23131 Intron: Tenm2 Intergenic: between Mir470 and Mir465c-1 Exon: Mir465
	9	0 - 0 - 0 - 0 0 off-targets	None
	10	0 - 0 - 0 - 4 0 off-targets	Intergenic: between Gm26166 and Tsc22d2 Intergenic: between Gm20386 and Grm7 Intergenic: between Gm26321 and Gm13597 Intergenic: between Gm23795 and Gm5342
Mouse	**16**	0 - 0 - 0 - 1 1 off-targets	Intergenic: between Flrt2 and 1700019M22Rik
	17	0 - 0 - 0 - 0 0 off-targets	None
	18	0 - 0 - 0 - 0 0 off-targets	None

	19	0 - 0 - 0 - 0 0 off-targets	None
	20	0 - 0 - 0 - 0 0 off-targets	None

It is relevant to note that all selected gRNAs showed no potential off-target events with 0 to up to 2 mismatches at the potential target site (that would mismatch within the PAM-distal end of the gRNA, consisting of 1-8 nucleotides) and in some cases even with 3 mismatches. It has been shown that only ≤4.2% of gRNAs remain able to bind targets with 2 mismatches (Anderson et al., 2015) and compared to on target cleavage (around 1 sec^{-1}), 3 mismatches at the distal PAM regions led to a 40-fold reduction in rate (Bravo et al., 2022). Furthermore, most off-targets are in intronic or intergenic regions, reducing the possibility of having a detrimental effect as no coding DNA would be targeted.

*Sa*Cas9 gRNAs were designed with a length of 20 nucleotides (nt), as gRNAs with this length have proven less tolerant to mismatches than 21 nt gRNAs, with 16% off:on target activity ratio vs 2% off:on target activity ratio (Tycko et al., 2018a).

All gRNAs listed in Table 4.4 were ordered from IDT to proceed with cloning and *in-vitro* screening.

4.1.2. Assessment of transient transfection efficiency in different cell lines.

Previous to cloning and *in-vitro* gRNA screening, transient transfection protocols were tested in a mouse and a human cell line to confirm optimal amount of DNA needed for an efficient transfection, hence a dose response was performed with pX601-CMV-SaCas9-GFP, an *Sa*Cas9 plasmid expressing GFP.

Viafect reagent was used to transfect N2A cells with pX601-CMV-SaCas9-GFP (Viafect to DNA 4:1 ratio). Cells were harvested 48 hours after transfection and analysed on a FACS Canto II machine (from BD Biosciences). The final data analysis to quantify GFP positive populations was performed using FloJo Software and plotted with Prism9 Software (Fig. 4.3). After Viafect transfection with 4 and 6 µg DNA doses, approximately 60 and 70% of N2A cells were GFP-positive, these results were used as proxy to determine transfection efficiency.

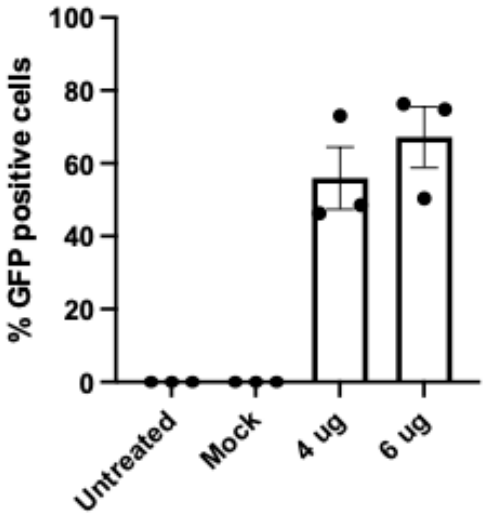


Figure 4.3. Bar chart of results from FACS Analysis of pX601-CMV-SaCas9-GFP dose response in transfected N2A cells. Bar charts show percentage of GFP positive cells according to different doses of plasmid DNA used per well on 6-well plates, seeded at $5x10^5$ cells per well (n = 3 technical repeats). Viafect to DNA ratio was 4:1. Error bars represent standard error of the mean.

The experiment was repeated on HEK293T cells alongside a positive control expressing a GFP (pCMV-GFP), to assess if Viafect was equally efficient in HEK cells as seen in N2A cells. Results from FACS analysis, can be observed in Fig. 4.4, showing ~65-75% GFP positive cells for both tested doses of the pX601-CMV-SaCas9-GFP plasmid, similar to results observed in N2A cells.

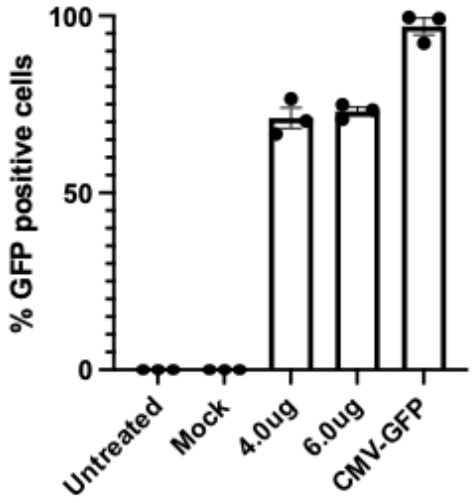


Figure 4.4. Bar chart of results from FACS Analysis of pX601-CMV-SaCas9-GFP dose response in transfected HEK293T cells. Bar chart shows the percentage of positive cells for 4 and 6 µg of plasmid DNA used per well, 1:4 to Viafect, on 6-well plates, seeded at $5x10^5$ cells per well (n = 3 technical repeats). CMV-eGFP is a plasmid with a CMV promoter expressing a GFP, used a positive control at a 4 µg dose. Error bars represent standard error of the mean.

Since Viafect transfections of pX601 at a 4 µg dose (4:1 Viafect to DNA ratio) showed similar efficiencies in N2A and HEK293T cells respectively, it was concluded that 4 µg was an adequate dose of pX601-CMV-SaCas9-GFP to transfect these cell lines. With this dose, ~60-65% of both cell types were GFP positive which should be sufficient for screening of sgRNA cleavage efficiency.

4.1.3. SaCas9 protein expression from pX601-CMV-SaCas9-GFP and pAAV-CMV-SaCas9 assessed by Western Blot.

Once the optimal dose for transfections was defined, *Sa*Cas9 plasmids (pX601-CMV-SaCas9-GFP and pAAV-CMV-SaCas9) were transfected into HEK293T cells in triplicate

using Viafect at a 4:1 ratio with DNA. Protein was extracted from harvested cells 48 hours after transfection and 50 µg of protein lysate loaded per well for Western Blotting on a 4-12% Bis-Tris Gel and analysed with an anti-*Sa*Cas9 primary antibody (monoclonal antibody raised in mouse against the N-terminus of the *S. Aureus* Cas9 nuclease, 1:5000, from Diagenode). Western Blot image confirming *Sa*Cas9 expression from both constructs in triplicate can be observed on Figure 4.5. The *Sa*Cas9 protein band matches the expected 127 kDa size.

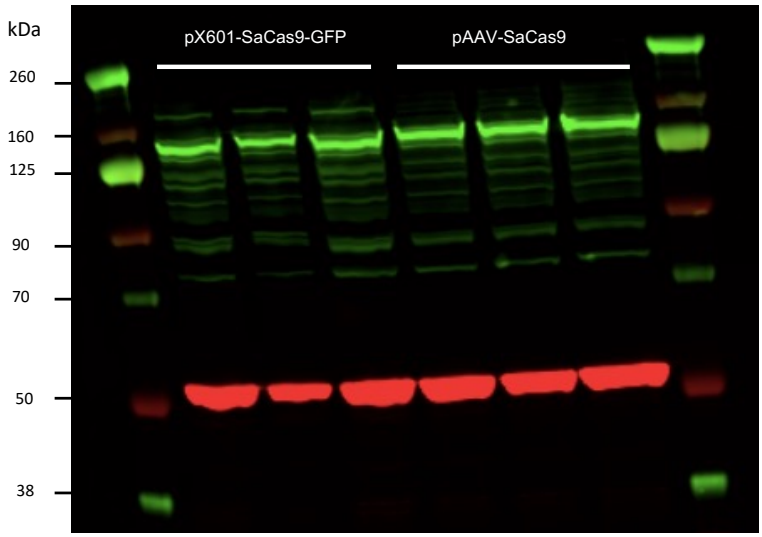


Figure 4.5. Western Blot to detect *Sa*Cas9 from HEK293T samples transfected with pX601-CMV-SaCas9-GFP and pAAV-CMV-SaCas9 in triplicate using Viafect. Chameleon Duo Pre-Stained Protein Ladder from Li-cor was used. 50 µg of protein lysate per well were loaded and analysed with antibodies: SaCas9 (1:5000, green) and α-tubulin (1:10,000, red) as a loading control.

4.2. SaCas9 gRNA cloning & *in-vitro* screening to determine cleavage efficiency.

4.2.1. SaCas9 gRNA cloning into pAAV-CMV-SaCas9.

Plasmid integrity of pAAV-CMV-SaCas9 was confirmed by restriction digestions before attempting to clone designed gRNAs. The expected band sizes produced by each selected restriction enzyme were obtained from SnapGene Software. Expected results were compared with results obtained from restriction digestions shown in Figure 4.6, confirming plasmid integrity. It is important to note that a GFP marker was not cloned into this plasmid to avoid exceeding the packaging size limit from AAV viral vectors.

*Sa*Cas9 gRNAs were cloned into pAAV-CMV-SaCas9 plasmid, at the BsaI restriction site located in a U6 expression cassette containing the gRNA scaffold, as following: a preparative restriction digestion of pAAV-CMV-SaCas9 plasmid was done with BsaI to recover plasmid backbone (Figure 4.7) which was then extracted and purified from the agarose gel. After recovering plasmid backbone, oligonucleotides with the forward and reverse sequence of each gRNA targeting human and mouse *DMD/Dmd* introns 18 and 55 were annealed to form a double stranded gRNA with extended sticky ends for ligation. Each annealed gRNA was ligated into the same plasmid backbone. The constructs were heat-shock transformed into *E. Coli* and plated on petri dishes with LB agar and 1% ampicillin. Four colonies were picked from each plate and mini-prepped. A sequencing primer targeting the U6 promoter upstream of the guide insertion site was designed (5´- CCG AGG GCC TAT TTC CCA TGA TTC -3´) and used to sequence and confirm

correct gRNA insertion. Sequencing traces of regions showing correct gRNA insertion are shown on Table 4.6.

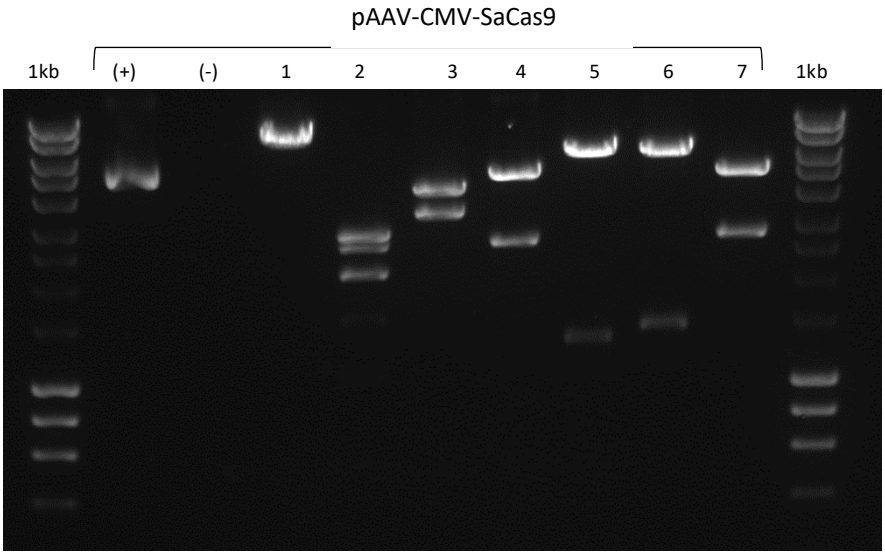


Figure 4.6. Gel Image from pAAV-CMV-SaCas9 restriction digests on 1% (w/v) agarose gel with 0.5X SYBR Safe in 1X TAE (Tris-Acetate–EDTA Buffer). From left to right: Hyperladder I from Bioline, positive control (undigested plasmid), negative control (enzyme only, to check for potential contamination). Lane 1 - BamHI: 1. 7446 bp. Lane 2 - MscI: 1. 2820 bp, 2. 2537 bp, 3. 2089 bp. Lane 3 - NdeI: 1. 4126 bp, 2. 3320 bp. Lane 4- SbfI: 1. 4841 bp, 2. 2605 bp. Lane 5 - SphI: 1. 6141 bp, 2. 1305 bp. Lane 6 - StuI: 1. 6047 bp, 2. 1399 bp. Lane 7 - XmaI: 1. 4754 bp, 2. 2681 bp, 3. 11 bp.

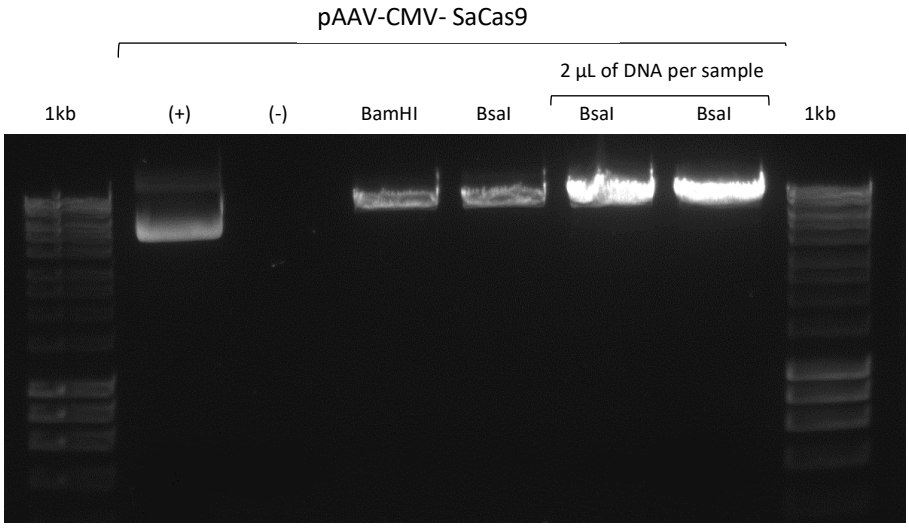
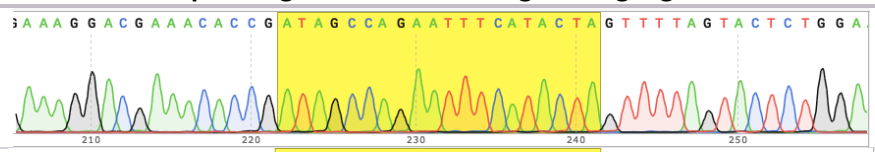
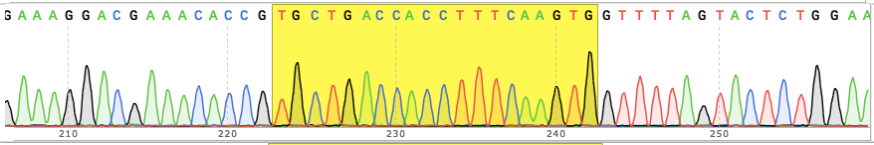
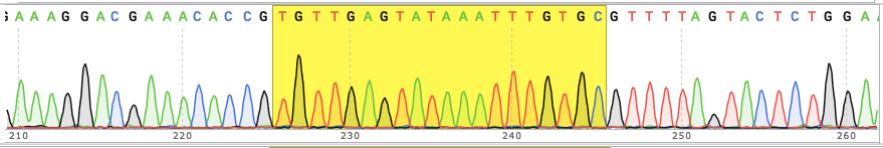
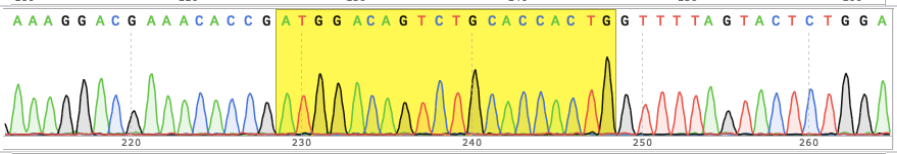
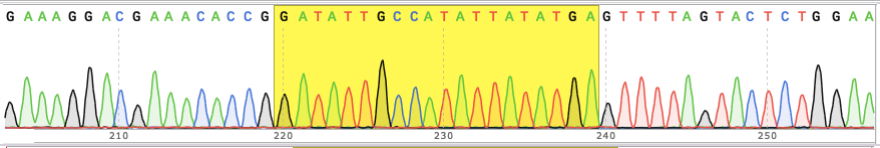
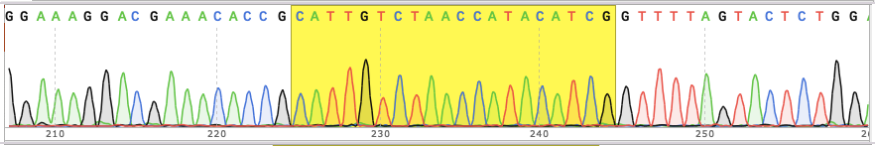
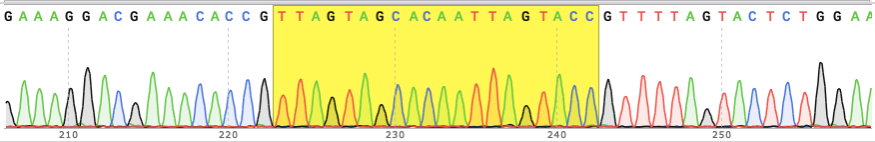
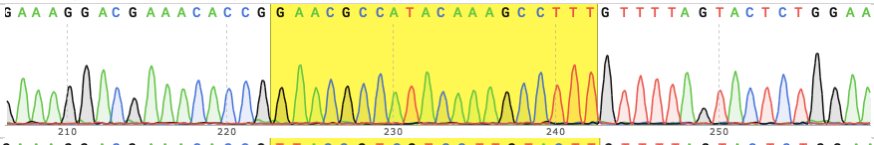
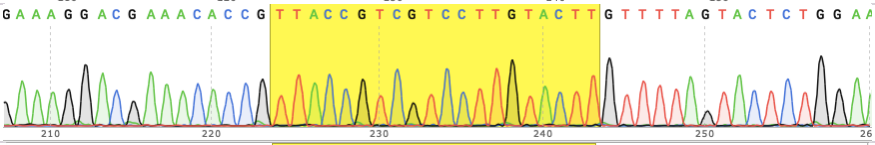
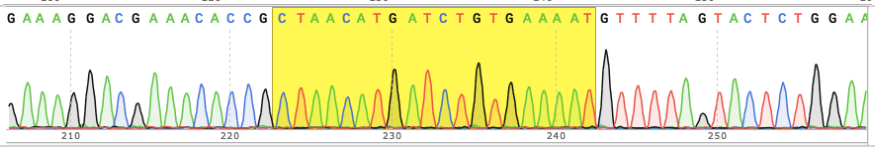
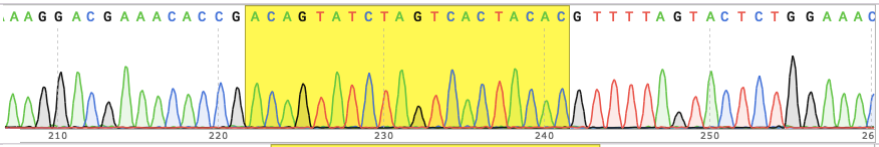
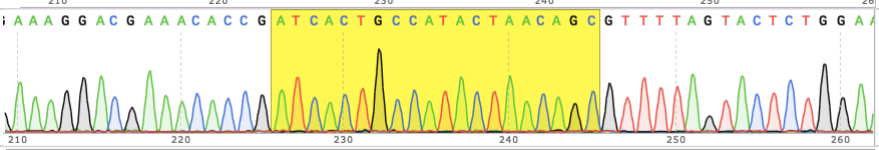
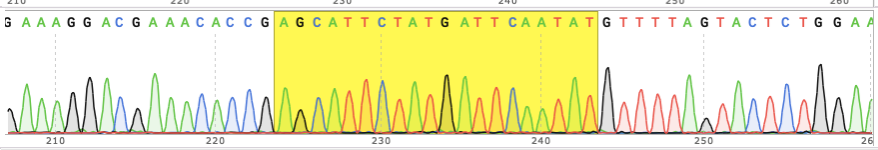
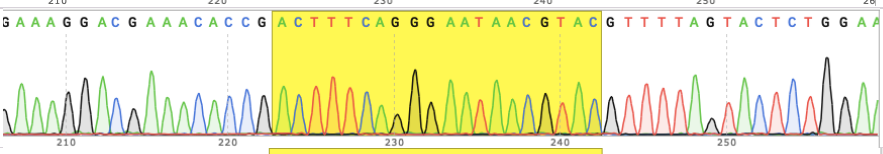
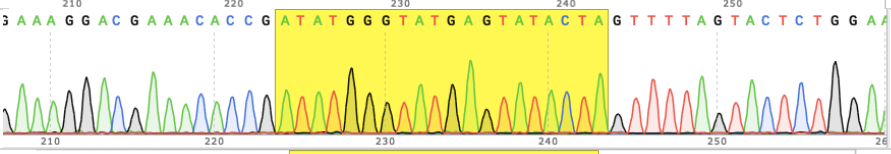
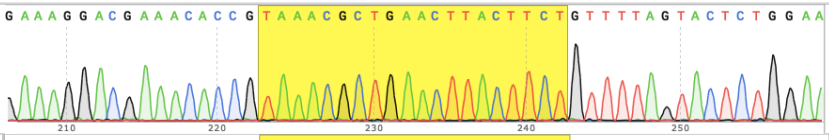
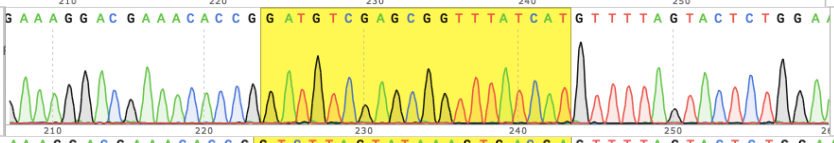
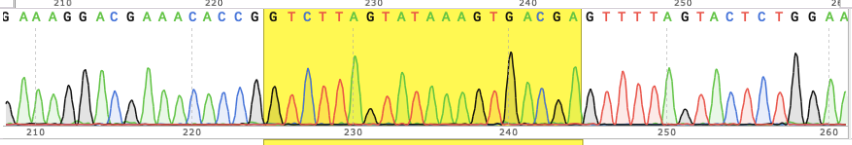
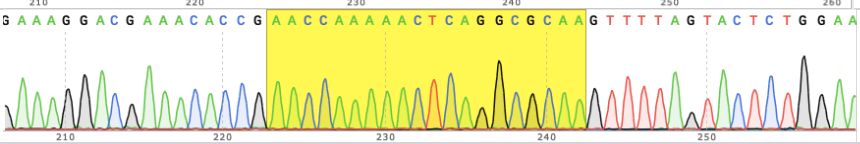
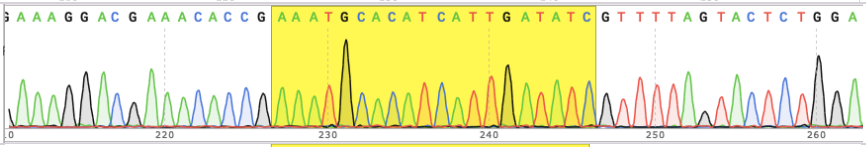
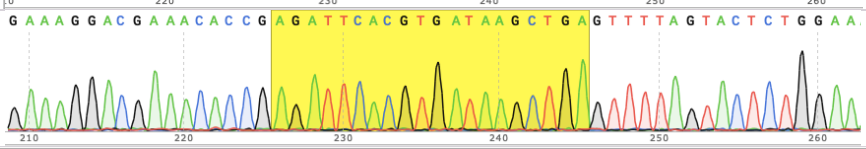
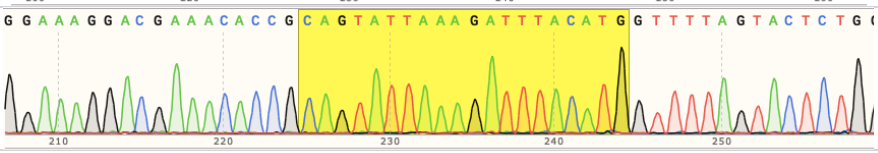


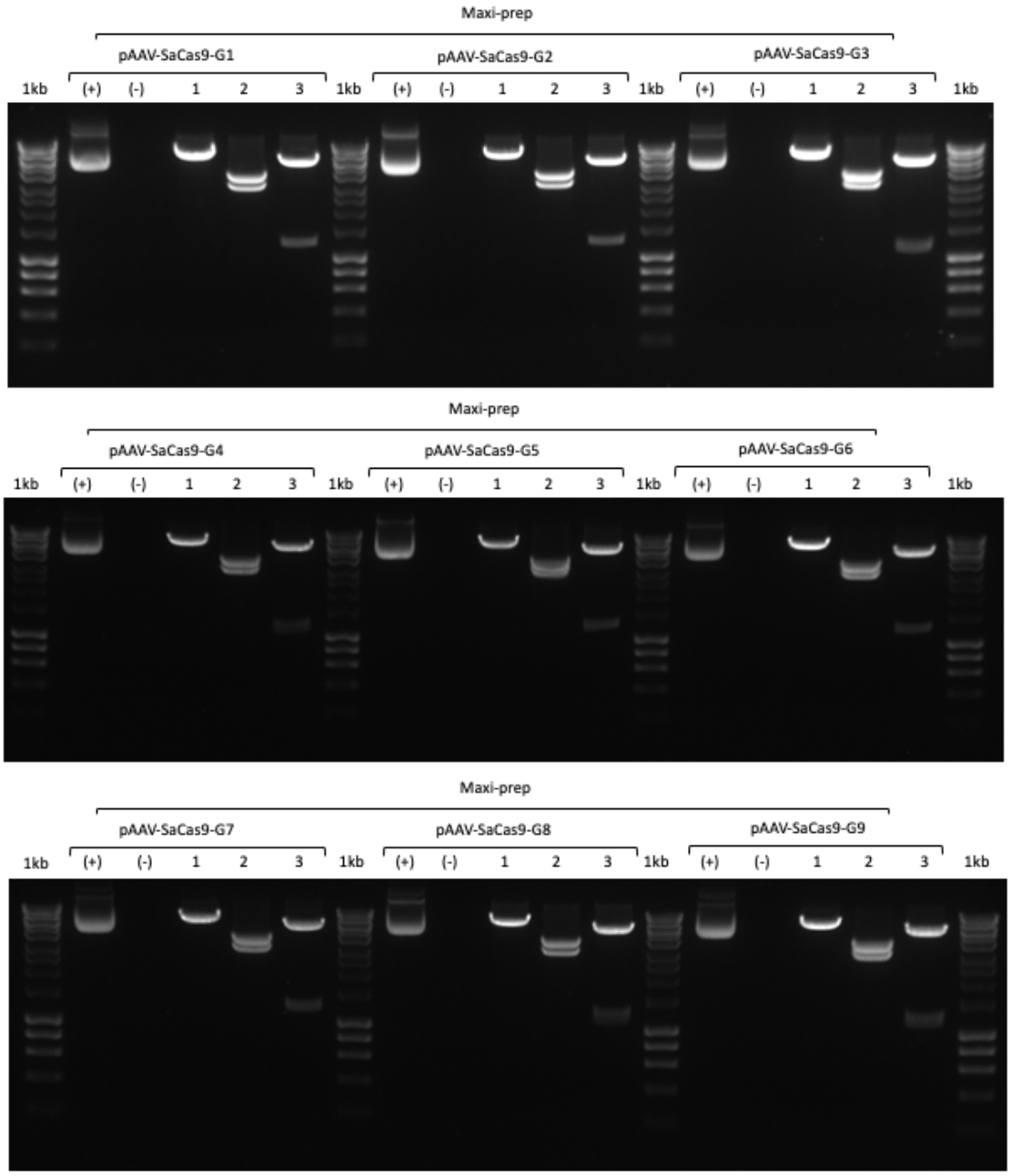
Figure 4.7. Gel Image from pAAV-CMV-SaCas9 preparative restriction digestion for band extraction. 1% (w/v) agarose gel with 0.5X SYBR Safe in 1X TAE (Tris-Acetate–EDTA Buffer). From left to right: Hyperladder I from Bioline, positive control (undigested plasmid), negative control (enzyme only), BamHI digest as an additional control (single cutter) and BsaI digest to recover vector backbone.

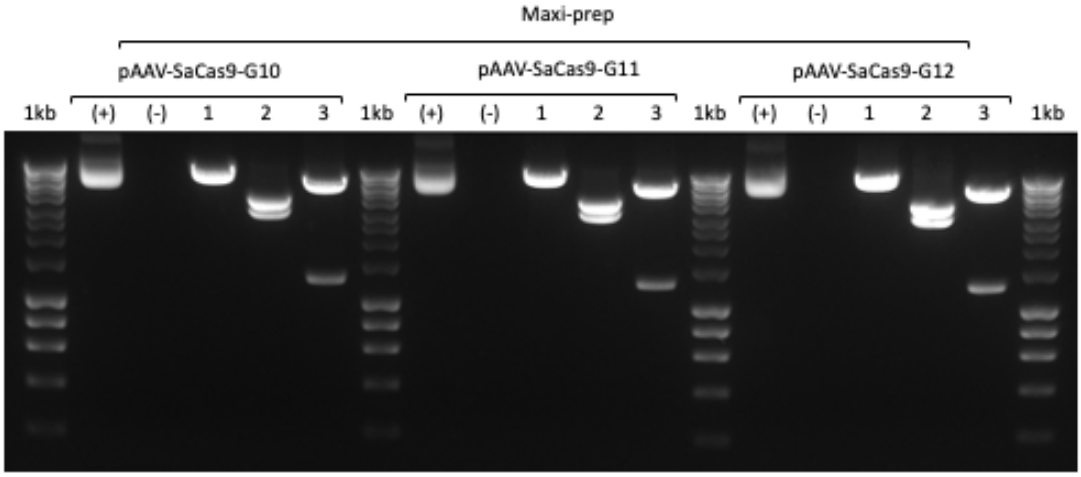
Table 4.6. Representative sequencing trace of approx. 55 bp showing correct gRNA insertion on pAAV-CMV-SaCas9 backbone. DNA obtained from mini-preps of colonies transformed with ligated constructs (each individual gRNA and a backbone). 2 µL of 10 pmol/µL stock of sequencing primer targeting U6 promoter upstream of guide cloning site (5´- CCG AGG GCC TAT TTC CCA TGA TTC -3´) were used per reaction with 50-100 ng of DNA/µL in a final volume of 20 µL for sequencing. Samples were sent and sequenced by Eurofins. Guide legend indicates human or mouse (H or M) – Intron (I)(18 or 55) – Guide label. Guide RNA sequence is highlighted in yellow on the sequencing trace.

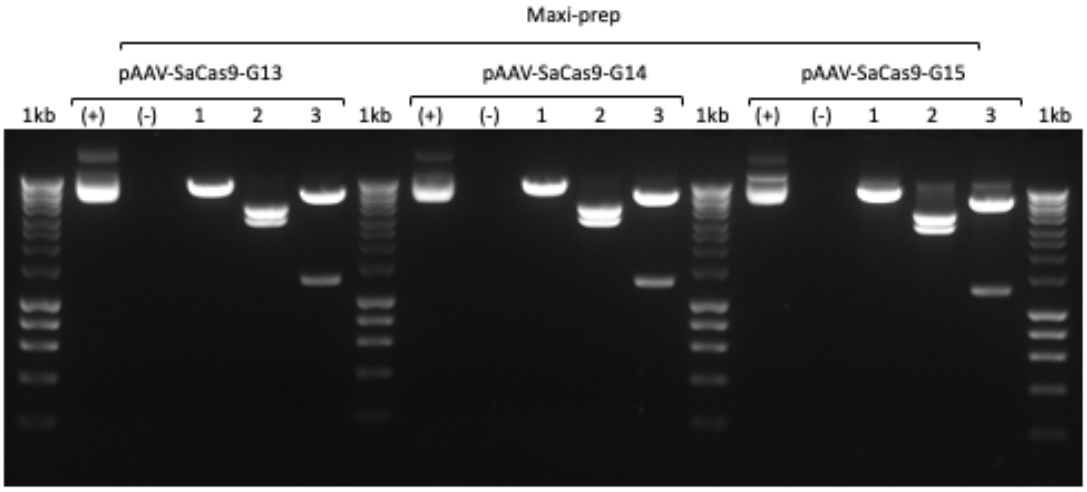
Guide	Sequencing trace with cloned gRNA highlighted
H-I-18-G1.2	
H-I-18-G2.2	
H-I-18-G3.2	
H-I-18-G4.3	
H-I-18-G5.1	
H-I-55-G6.1	
H-I-55-G7.1	
H-I-55-G8.3	
H-I-55-G9.2	
H-I-55-G10.4	

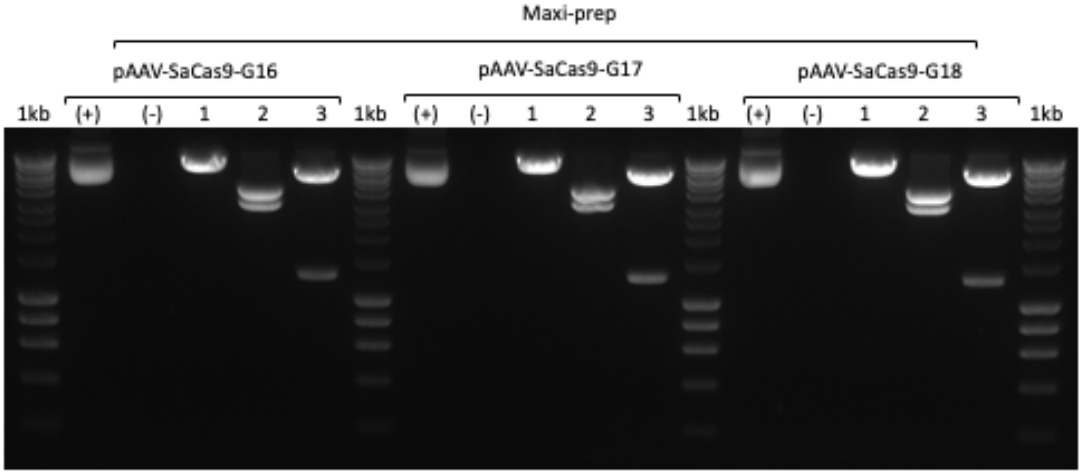
M-I-18-G11.1	
M-I-18-G12.2	
M-I-18-G13.1	
M-I-18-G14.1	
M-I-18-G15.1	
M-I-55-G16.1	
M-I-55-G17.1	
M-I-55-G18.1	
M-I-55-G19.1	
M-I-55-G20.1	
H-I-18-G41.1	
H-I-18-G42.4	

Once successful gRNA cloning was confirmed by sequencing, one clone per gRNA was selected and maxi-prepped to obtain clean DNA with higher concentrations. Plasmid integrity of each maxi-prep was confirmed by sequencing and restriction digestions, as seen on Fig. 4.8 and Table 4.7.


Maxi-prep
pAAV-SaCas9-G1
pAAV-SaCas9-G2
pAAV-SaCas9-G3
1kb (+) (-) 1 2 3 1kb (+) (-) 1 2 3 1kb (+) (-) 1 2 3 1kb
Maxi-prep
pAAV-SaCas9-G4
pAAV-SaCas9-G5
pAAV-SaCas9-G6
1kb (+) (-) 1 2 3 1kb (+) (-) 1 2 3 1kb (+) (-) 1 2 3 1kb
Maxi-prep
pAAV-SaCas9-G7
pAAV-SaCas9-G8
pAAV-SaCas9-G9
1kb (+) (-) 1 2 3 1kb (+) (-) 1 2 3 1kb (+) (-) 1 2 3 1kb


Maxi-prep
pAAV-SaCas9-G10
pAAV-SaCas9-G11
pAAV-SaCas9-G12
1kb (+) (-) 1 2 3 1kb (+) (-) 1 2 3 1kb (+) (-) 1 2 3 1kb


Maxi-prep
pAAV-SaCas9-G13
pAAV-SaCas9-G14
pAAV-SaCas9-G15
1kb (+) (-) 1 2 3 1kb (+) (-) 1 2 3 1kb (+) (-) 1 2 3 1kb


Maxi-prep
pAAV-SaCas9-G16
pAAV-SaCas9-G17
pAAV-SaCas9-G18
1kb (+) (-) 1 2 3 1kb (+) (-) 1 2 3 1kb (+) (-) 1 2 3 1kb

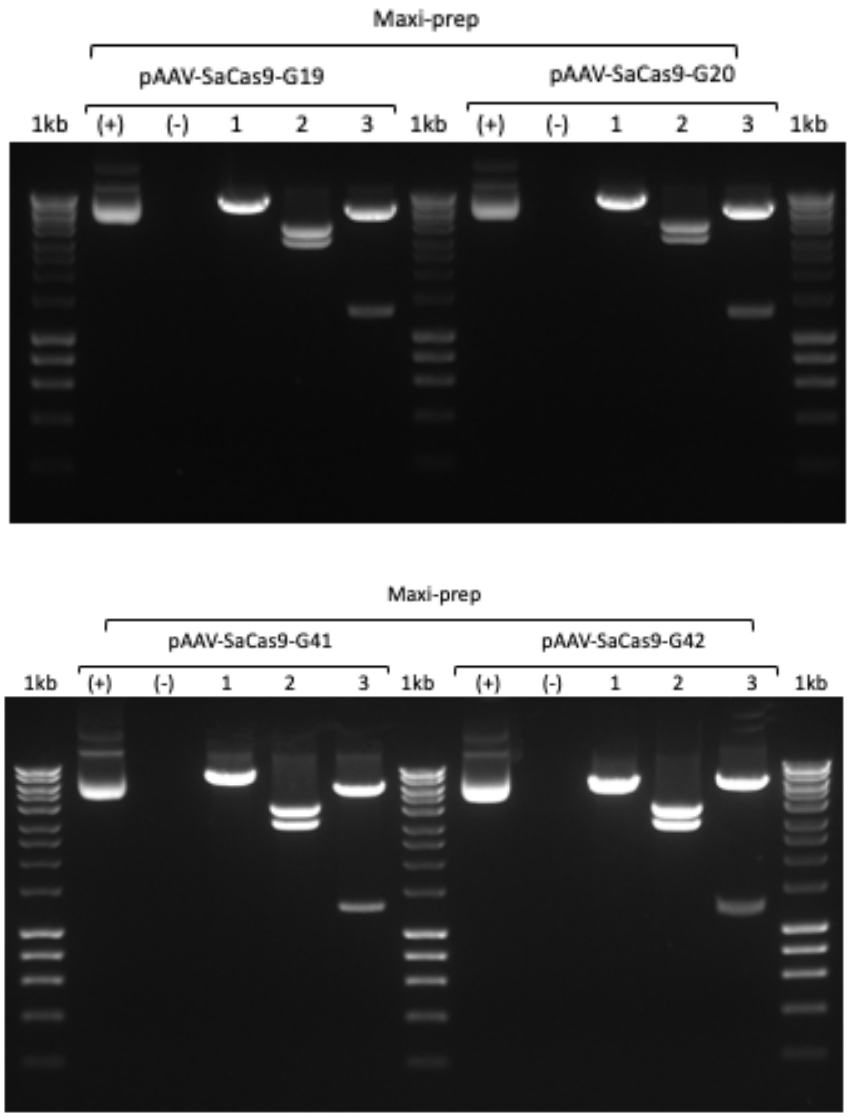
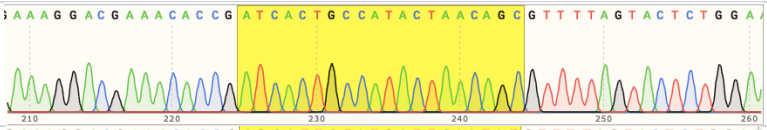
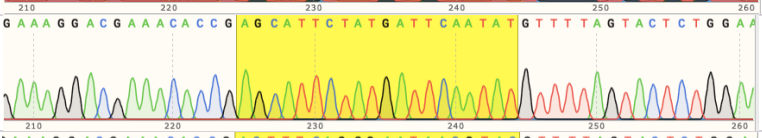
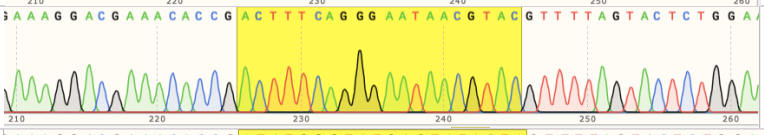
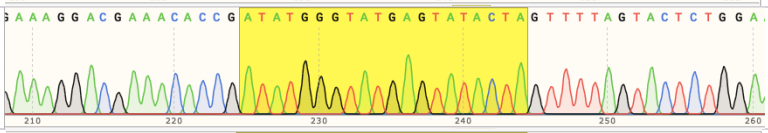
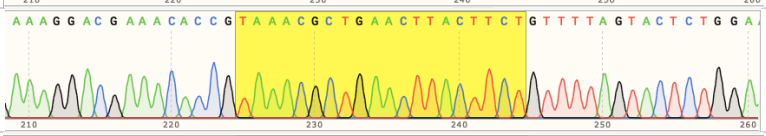
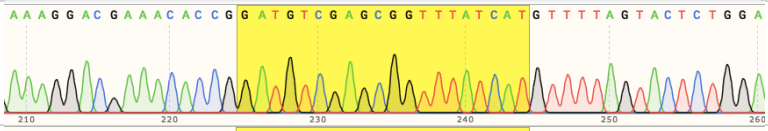
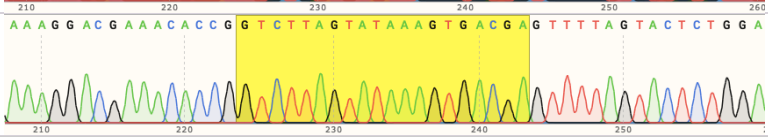
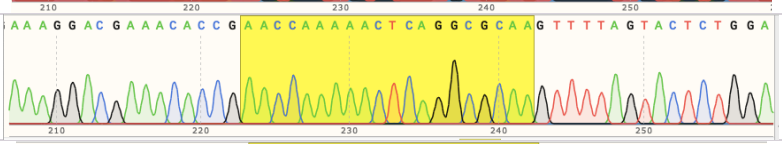
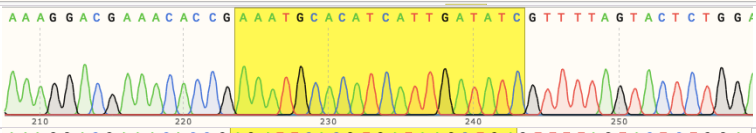
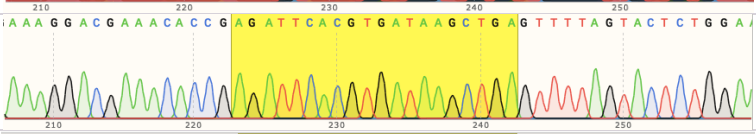
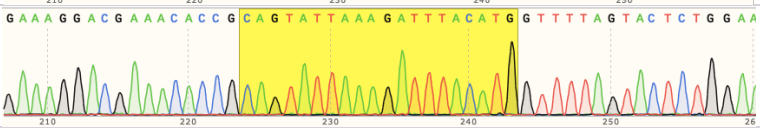


Figure 4.8. Gel Images from restriction digestions of maxi-preps from pAAV-CMV-SaCas9 with cloned gRNAs (G1-G20 and G41-G42). 1% (w/v) agarose gel with 0.5X SYBR Safe in 1X TAE Buffer. Positive control (+) is undigested plasmid with respective gRNA cloned in and negative (-) control is enzyme with no DNA (to test for contamination). Plasmids digested with 1=BamHI, 2=NdeI and 3=SphI. Fragments show the following expected band sizes: BamHI: 1. 7,446 bp. NdeI: 1. 4,126 bp. 2. 3,320 bp. and SphI: 1. 6,141 bp. 2.1,305 confirming plasmid integrity of each maxi-prep. Hyperladder I (1kb) was used.

Table 4.7. Representative sequencing trace of approx. 55 bp showing correct gRNA insertion on pAAV-CMV-SaCas9 backbone. DNA obtained from maxi-preps of colonies transformed with ligated constructs (each individual guide and a backbone). 2 µL of 10 pmol/µL stock of sequencing primer targeting U6 promoter upstream of guide cloning site (5´- CCG AGG GCC TAT TTC CCA TGA TTC -3´) were used per reaction with 50-100 ng of DNA/µL in a final volume of 20 µL for sequencing. Samples were sent and sequenced by Eurofins. Guide legend indicates human or mouse (H or M) – Intron (18 or 55) – Guide label. Guide RNA sequence highlighted in yellow on the sequencing trace.

Guide	Sequencing trace with cloned gRNA highlighted
H-I-18-G1	
H-I-18-G2	
H-I-18-G3	
H-I-18-G4	
H-I-18-G5	
H-I-55-G6	
H-I-55-G7	
H-I-55-G8	
H-I-55-G9	
H-I-55-G10	
M-I-18-G11	

M-I-18-G12	
M-I-18-G13	
M-I-18-G14	
M-I-18-G15	
M-I-55-G16	
M-I-55-G17	
M-I-55-G18	
M-I-55-G19	
M-I-55-G20	
H-I-18-G41	
H-I-18-G42	

4.2.2. *In-vitro gRNA screening by transfection, DNA extraction & Tracking of Indels by Decomposition (TIDE) Analysis of purified PCR products.*

All constructs with gRNAs were transfected on 6-well plates in triplicate, alongside pAAV-CMV-SaCas9 plasmid (with no gRNAs) as a negative control, using Viafect 4:1 to DNA and 4 µg of DNA per well, into HEK293T or N2A cells (for human and mouse gRNAs respectively) at a cell density of $5x10^5$ per well. 48 hours after transfection, cells were harvested and DNA extracted. PCRs were performed for each DNA sample (treated and untreated) with appropriate primers for amplifying around the predicted on-target site for each gRNA (primers used for PCRs can be found in Table 2.5 of Materials & Methods Section 2.8). PCR products were run on a 1% (w/v) agarose gel, PCR products would then be purified (using a QIAquick PCR Purification Kit) and sent for Sanger sequencing to Eurofins with appropriate primers.

DNA sequences were analysed using the TIDE (Tracking of Indels by Decomposition) web tool (Brinkman et al., 2014), as described in Materials & Methods Section 2.9, by comparing percentage of frequency of indels in "edited" populations vs. control (untreated) populations, used as a proxy for editing efficiency of each individually assessed gRNA.

A representative image of TIDE output can be seen on Fig. 4.9. The rest of the analysis by TIDE for each gRNA can be found on Appendix C.

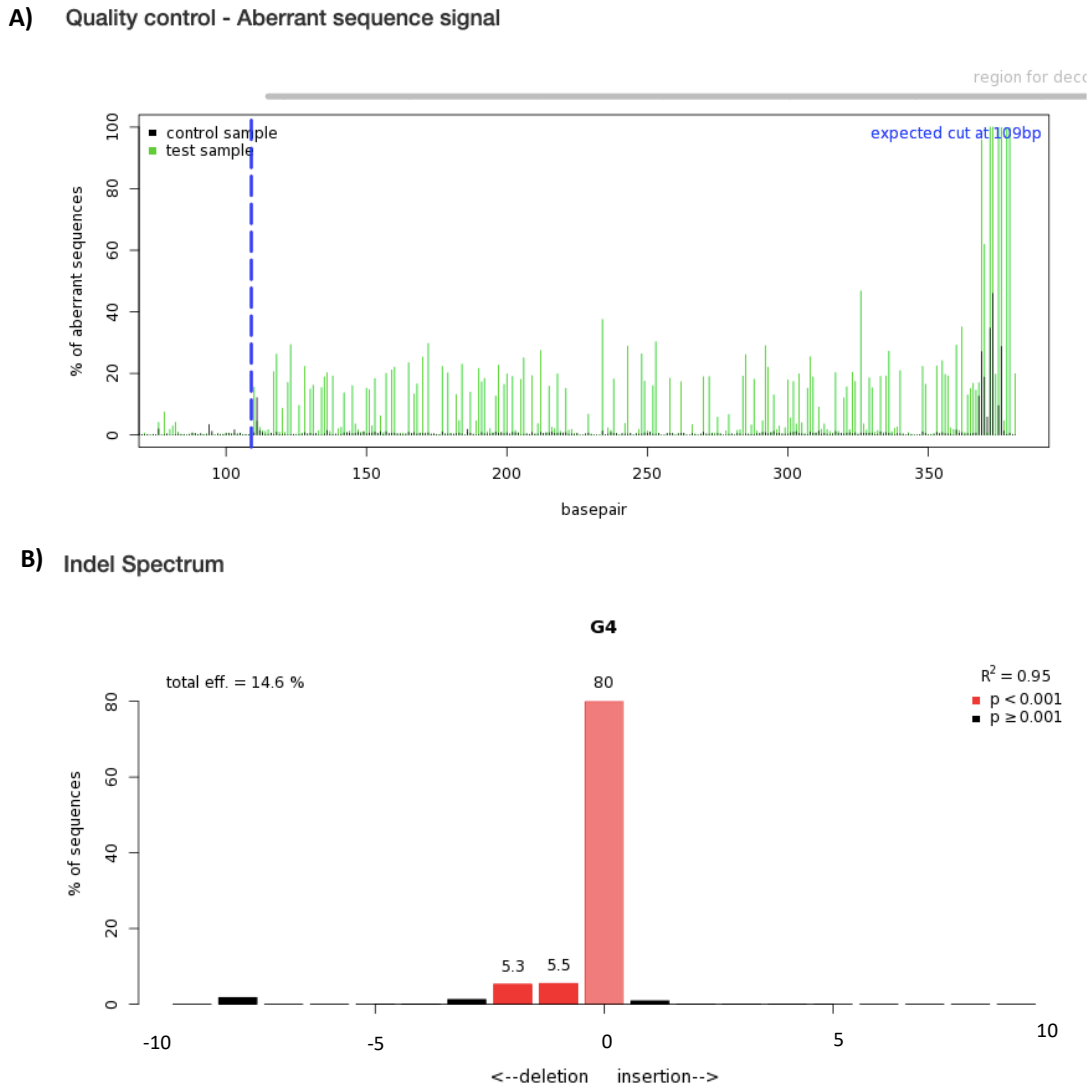


Figure 4.9. Representative images of outputs from TIDE analysis Software. Analysis of Guide 4 (H-I-18-G4) presented. A) Decomposition trace, aberrant sequence signal (green) compared to control trace (black). Dotted blue line indicates cut site. B) Bar chart indicating indel spectrum output. X-axis indicates small deletions of up to 10 base pairs on a negative scale (-10 to 0) and insertions on a positive scale (0 to 10). In this example, the red bars indicate 80% of traces had 0 deletions or insertions, 5.5% had 1 deleted bp and 5.3% had 2 deleted bp, lower percentages of edited populations (black bars) indicate some -3 and 8 bp deletions and 1 bp insertions. These percentages add up to a total efficiency of 14.6% from Guide 4, indicated at the top left corner of the graph. Numbers at the top right corner denote the coefficient of determination (R^2), a statistical measure to evaluate model accuracy with values from 0 to 1. A low R^2 can be caused by poor sequence quality or non-optimal setting. P-values indicate significance cutoff, set up at $p<0.001$. Significant outputs indicated in red, non-significant ($p \geq 0.001$) indicated in black.

Transfections with each gRNA were performed in triplicates, hence gRNA efficiency was calculated for each replicate, then averaged for each gRNA and analysed on Prism9 Software. The efficiencies of all *Sa*Cas9 gRNAs for human and mouse *DMD/Dmd* genes are presented in Fig. 4.10, compared to a negative control (pAAV-CMV-SaCas9 "empty", original construct with no gRNAs cloned in).

No efficient gRNAs were found for human intron 55 using this method of assessment; this could be due to the "AT" rich region that the gRNAs were targeting. For the same reason it was not possible to sequence the PCR product produced from amplification around the target site of Guide 7.

For human intron 18, Guide 4 seemed to be the most efficient one with approximately 12% editing efficiency. As a positive control, an *Sa*Cas9 gRNA previously shown to have an efficiency of 45% (efficiency assessed by T7 assay) in HEK293FT cells, when transfected with Lipofectamin 2000 in a construct with an EFS promoter (Kumar et al., 2018), was cloned into pX601-CMV-SaCas9-GFP and pAAV-CMV-SaCas9. Both plasmids with the positive control gRNA showed 20-23% efficiency. The difference in efficiency shown by the positive control gRNA compared to its previous 45% (Kumar et al., 2018), could be due to transfection protocols (Viafect instead of Lipofectamin 2000), the construct used to deliver the *Sa*Cas9 and gRNA, particularly the use of a different promoter to drive expression of the Cas9, the use of a different cell line for screening or the assessment method.

For mouse intron 55, Guides 16 and 18 showed the highest activity, with an average editing efficiency of 10.6% and 11.7% respectively. It was only possible to sequence one PCR product from Guide 19, as this gRNA targets an “AT” rich region that interfered with sequencing.

For mouse intron 18, Guide 12 and 14 showed the highest activity, with an average editing efficiency of 17.8% and 19.4% respectively. The rest of the gRNAs targeting this intron showed low to no activity.

Guide RNAs that could target both mouse and human sequenced showed low or no activity, Guide 41 (H-I-18-G41) showed ∼5% editing efficiency, Guide 42 (M-I-18-G42) showed no activity and it was not possible to recover a PCR product to assess Guide 10 (H/M-I55-G10), as the region that this gRNA targets is an “AT” rich region.

In summary, from the gRNAs targeting human intron 18, Guide 1 and 4 showed the highest editing efficiencies, ∼12.4% and ∼12.5% respectively. Guide RNAs targeting human intron 55 showed low to no activity, with Guide 8 showing the highest activity, ∼3% of editing efficiency. From the gRNAs targeting mouse *Dmd* gene, Guide 14 with an average of 20% and Guide 18 with an average of 12% editing efficiency, targeting intron 18 and 55 respectively, were the gRNAs that showed the highest activity. At this stage it was decided to continue further testing with the most efficient mouse gRNAs.

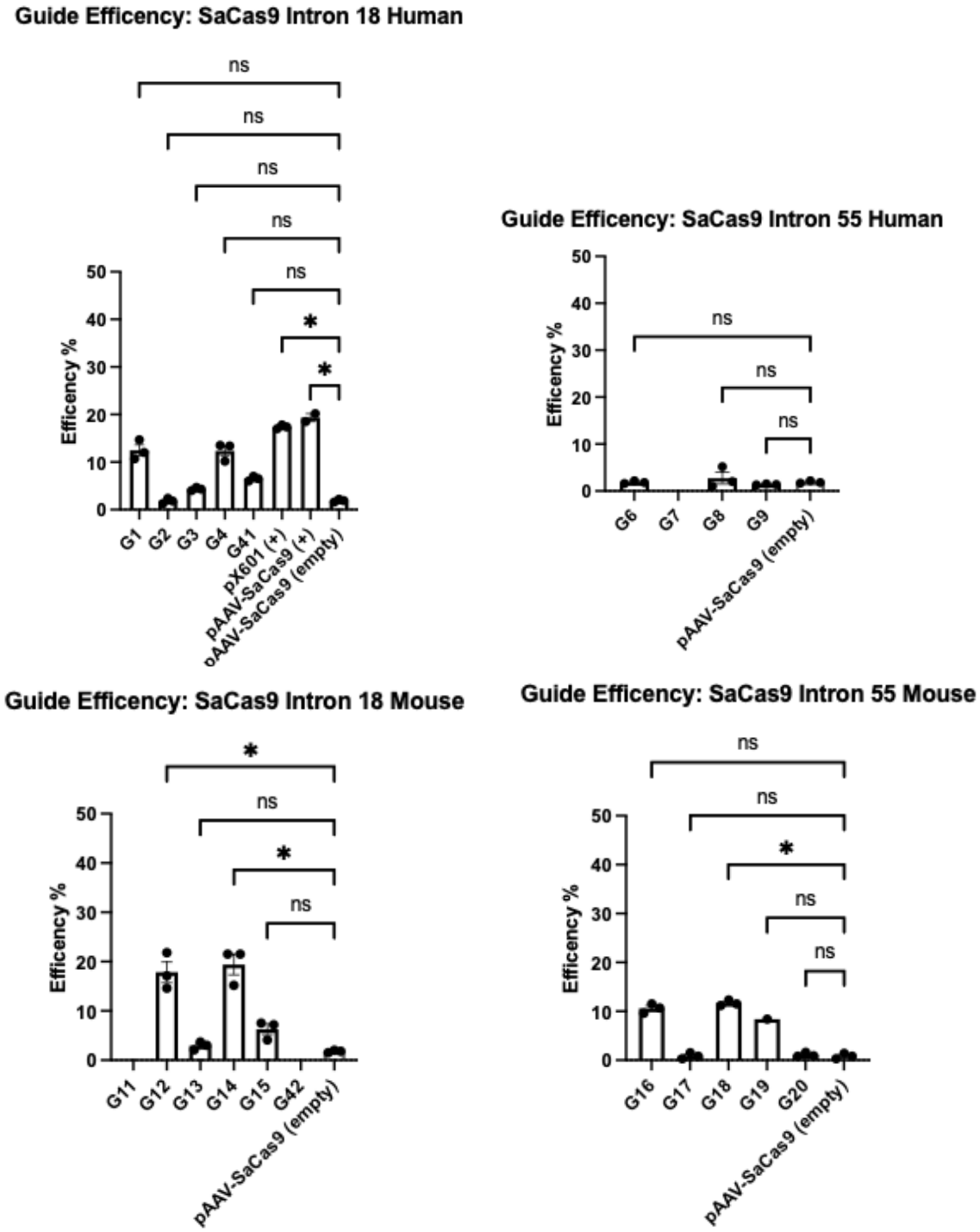


Figure 4.10. Graphical summary of *Sa*Cas9 gRNAs cutting efficiency based on TIDE Analysis. Bar charts show: A) gRNAs targeting intron 18 of human *DMD* gene. B) gRNAs targeting intron 55 of human *DMD* gene. C) gRNAs targeting intron 18 of mouse *Dmd* gene. D) gRNAs targeting intron 55 of mouse *Dmd* gene. All gRNAs were cloned into

pAAV-CMV-SaCas9, transfected with Viafect 4:1 to DNA (4 µg/transfection) on HEK293T cells for human gRNAs or N2A cells for mouse gRNAs. Cells were harvested 48 hrs. after transfection and DNA was extracted. Appropriate PCR primers were designed targeting the sequence flanking the target site of the sgRNAs. PCRs were performed for each sample and run on a 1% (w/v) agarose gel, PCR products were extracted, cleaned and sent for sequencing (to Eurofins) with appropriate primers. Sequence traces were then analysed by TIDE (Tracking of Indels by Decomposition). TIDE web tool algorithm reconstructs the spectrum of indels from two sequencing traces per guide (an edited vs. untreated trace). The output reports identity and frequency of detected indels, as a percentage, generated in a pool of cells (Brinkman et al., 2014). Each guide was transfected in triplicate. Data was plotted & analysed on Prism9 Software. Human gRNAs from A) Positive controls (pX601(+) adjusted p-value = 0.0223, pAAV-SaCas9(+) = 0.0165) and mouse gRNAs from C) G14 (adjusted p-value = 0.0138), G12 (adjusted p-value = 0.0185) and E) G18 (adjusted p-value = 0.0379), were found significant by mean comparison against negative control samples (transfected with pAAV-CMV-SaCas9-empty plasmid) with a Kruskal-Wallis test (95% confidence interval, p-value<0.05), followed by a post-hoc Dunn´s test. *indicates an adjusted p-value<0.05. Non-significant difference = ns. Error bars represent standard error of the mean. It was not possible to obtain PCR products from H-I-55-G7, M-I-18-G11 and M-I-18-G42. Only one sample was obtained to assess M-I-55-G19. The rest of the groups have an n = 3 technical repeats.

4.3. *IN-VITRO* ESTABLISHMENT OF CREATION OF *DE NOVO* INTRONIC JUNCTION AFTER DELETION OF EXONS 19 TO 55 BY CO-TRANSFECTION OF MOUSE GRNAS.

4.3.1. CO-TRANSFECTION OF N2A CELLS WITH GRNAS TARGETING INTRON 18 AND 55.

Based on the results of gRNA screening, Guide 14 (targeting Intron 18 of mouse *Dmd*) and Guide 18 (targeting Intron 55 of *Dmd* gene) were selected to attempt the deletion of exon 19 to exon 55 by plasmid co-transfection into N2A cells. Guides 14 and 18 were co-transfected (2 µg of each plasmid) with a 4 µg total DNA dose per well and a ratio of 1:4 to Viafect. Cells were harvested 48 hrs. after transfection, then genomic DNA and RNA were extracted for further analysis.

4.3.1.1. CONFIRMATION OF THE DE NOVO INTRONIC JUNCTION OF INTRONS 18 AND 55 BY SANGER SEQUENCING FROM DNA EXTRACTED FROM CO-TRANSDUCED N2A CELLS.

Extracted DNA was subjected to PCR amplification using a forward primer (5´-CCCAGGCAAACATGATACAATTAG-3´) targeting intron 18 and a reverse primer targeting intron 55 (5´-CTGGTCCATGCCTAACCATAT-3´). This primer pair would amplify a product of 757 bp if deletion had occurred as a result of NHEJ of DNA 3’ of the intron 18 gRNA cleavage site and DNA 5’ of the intron 55 gRNA cleavage site. Where deletion did not occur, either as result of lack of two DSBs or repair of each cleavage site with small InDels, the PCR product would be over 800 kbp and therefore too large to be amplified

with the conditions used for the PCR amplification. A deletion was detectable using this PCR as shown on Figure 4.11.

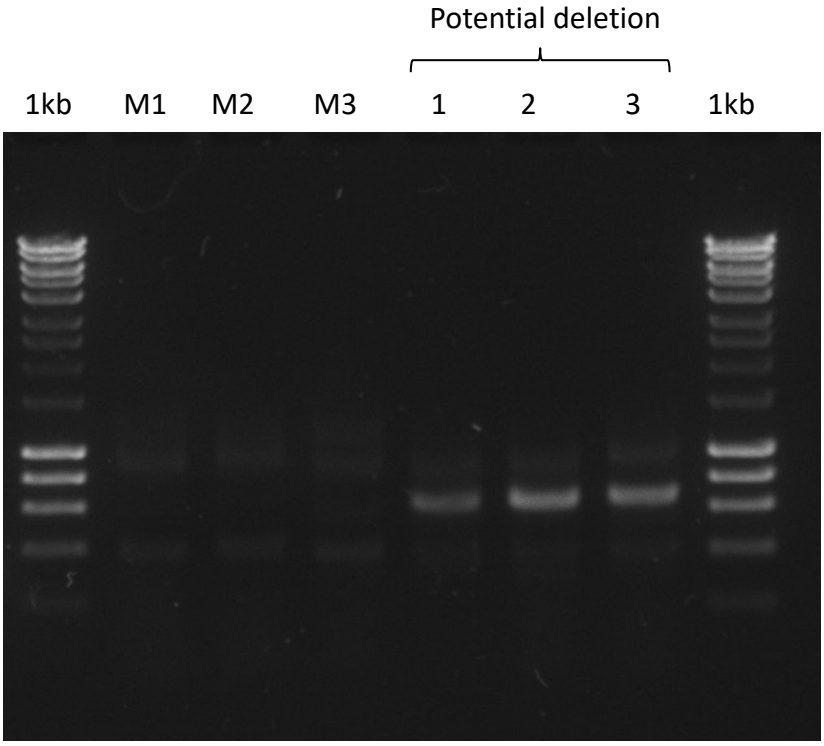


Figure 4.11. Gel image of PCR products resulting from deletion of exons 19-55. DNA extracted from N2A cells co-transfected with Guide 14 and Guide 18, showing a band on samples with a potential deletion of Exons 19-55. 1% (w/v) agarose gel with 0.5X SYBR Safe in 1X TAE Buffer. From left to right: Mocks (samples treated with Viafect and no DNA) showing very faint bands, PCR products of 3 samples co-transfected with guides 14 and 18 matching the expected band size of ~757 bp which would indicate deletion of exons 19-55. PCR primers used: 5´-CCCAGGCAAACATGATACAATTAG-3´ forward primer targeting intron 18 and 5´-CTGGTCCATGCCTAACCATAT-3´ reverse primer targeting intron 55.

The PCR products of ~757 bp were extracted from the gel and sent for Sanger sequencing with forward and reverse primers. All sequencing traces showed no background noise and confirmed the expected deletion, as shown on a representative

alignment in Figs. 4.12-13. Sequencing traces showed a *de novo* junction of intron 18 and intron 55.

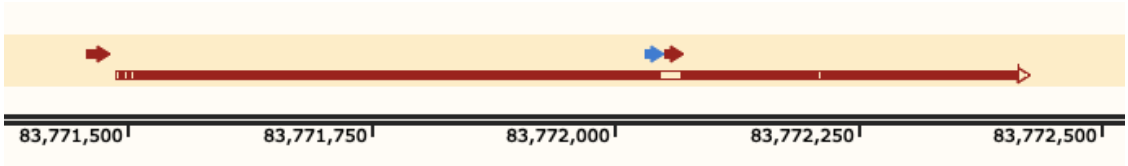


Figure 4.12. Representative alignment of Del19-55 mouse *Dmd* gene generated in SnapGene and sequencing trace from treated sample, at the junction of introns 18 and 55. Black lines indicate Del19-55. Base pair position indicated below with numbers. Sample sequencing trace (red arrow indicating no mismatches with aligned sequence) was obtained from sequencing a PCR product from DNA extracted from N2A cells co-transfected with Guides 14 and 18. Arrows from left to right indicate sequencing primer, Guide 14 (blue) and Guide 18 (red).

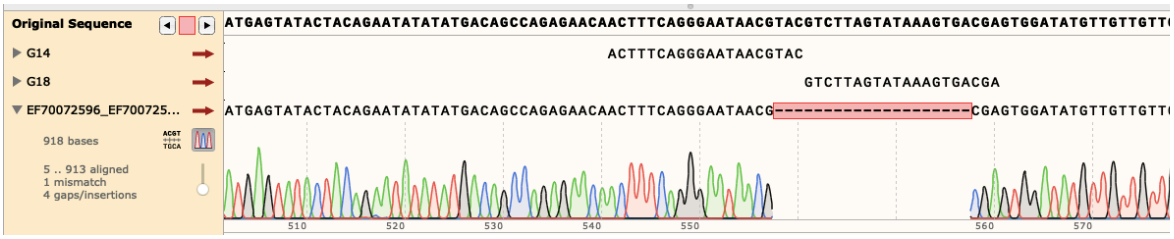


Figure 4.13. Zoom-in of Fig. 4.12 showing the 5´to 3´ forward sequence of the *de novo* junction of introns 18 and 55 of Del19-55 mouse *Dmd.* Del 19-55 sequence aligned on SnapGene with a sequencing trace from PCR product from DNA extracted from co-transfected N2A cells, confirming the *de novo* junction of the 3' end of intron 18, at the gRNA cut site, and the cut site at the 5' end of intron 55.

4.3.1.2. *Deletion of exons 19 to 55 confirmed by Sanger sequencing of cDNA obtained from RNA extracted from co-transduced N2A cells.*

RNA was extracted from N2A cells co-transduced with Guides 14 and 18 (targeting intron 18 and 55 of *Dmd* gene) and processed to make cDNA using reverse transcriptase with the QuantiTect Rev. Transcription Kit from QIAGEN. PCR primer pairs targeting E17-20 and E17-56 were designed as illustrated on Fig. 4.14. These primers would allow for a triple primer approach, where the second primer pair would only amplify using the PCR protocol of 1 minute final extension time (full protocol on Section 2.8.3) if there is a deletion, as illustrated on Fig. 4.14B.

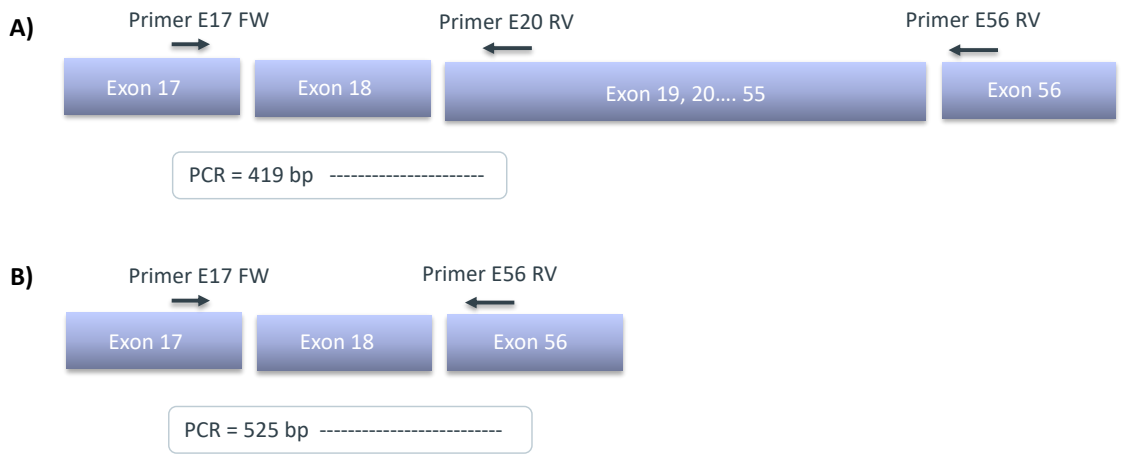


Figure 4.14. Illustration of the triple primer design to detect a deletion from cDNA PCR of co-transfected (G14 and G18) N2A cells. A) Forward primer targeting Exon 17 (5´-CAAGGGAACAGATCCTGGTAAA-3´) and reverse primer targeting Exon 20 (5´-CTGATACTCCAGCCAGTTAAGTC-3´). PCR product of 419bp. If a third primer was added targeting Exon 56 (5´-CTGGAAAGTCGCCTCCAATAG-3´), the PCR reaction from full length cDNA would be 6,200 bp long, therefore unlikely to be synthetized with the PCR conditions of this experiment. B) Forward primer targeting Exon 17 and reverse primer targeting Exon 56, if there is a deletion, this primer pair would produce a product of 525 bp.

Results from primer pairs targeting Exons 17-20 and Exons 17-56 run with cDNA from samples co-transfected with Guides 14 an 18 can be seen in Fig. 4.15.

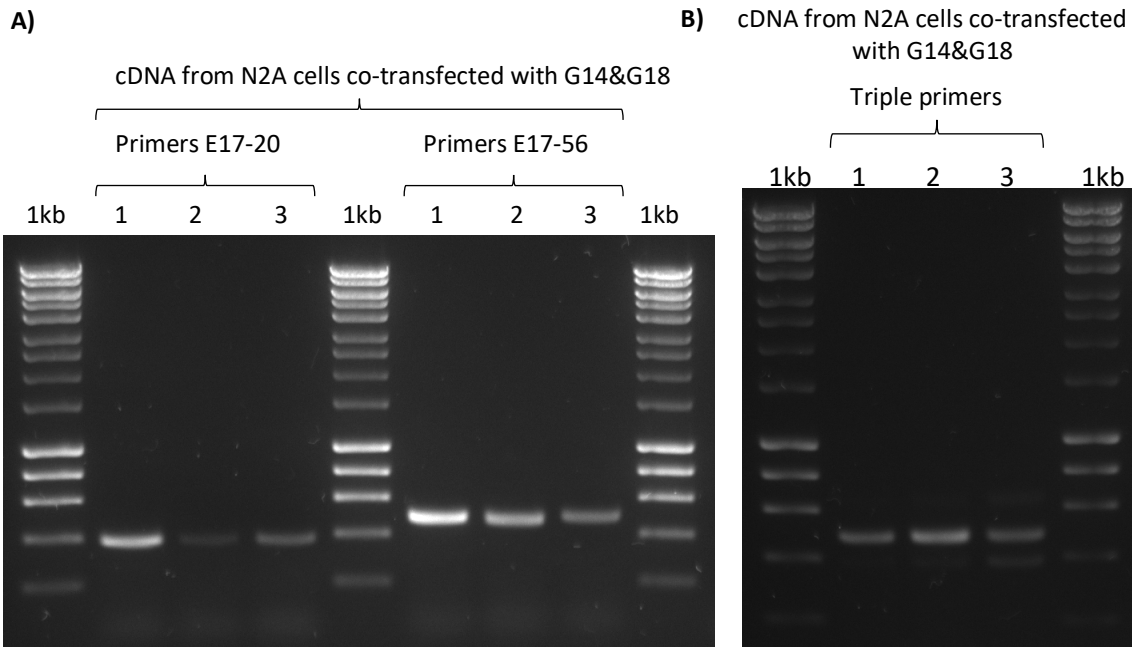
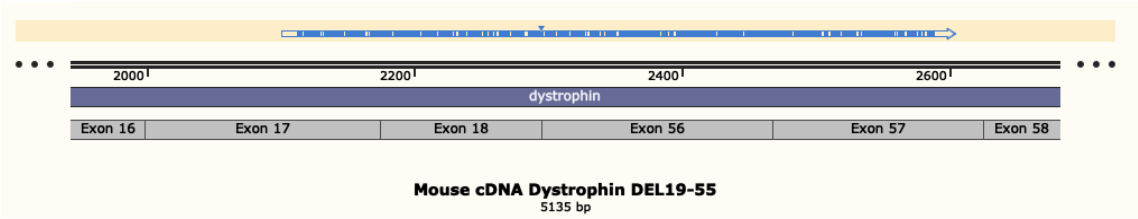


Figure 4.15. Images of agarose gels with PCR products from N2A cells cDNA co-transfected with Guide 14 & Guide 18 (n = 3 technical repeats). 1%(w/v) gels made in 1X TAE Buffer with 0.5X SYBR Safe. Hyperladder I used (1kb). A) PCR products from primer pairs targeting Exons 17-20: bands matching expected product size of 419 bp and PCR products from primers targeting Exons 17-56: bands matching expected product size expected of 525 bp, if exons 19-55 are deleted. B) PCR product from cDNA samples with forward primer targeting Exon 17 and two reverse primers, one targeting Exon 20 and the second one targeting Exon 56. Two products were amplified, a 419 bp product for non-deleted dystrophin cDNA and a 525 bp product representing deletion of exons 19 to 55.

PCR products from the reaction with primer pair targeting E17-56 (Fig. 4.15A) were purified and sent for Sanger sequencing with forward and reverse primers respectively. A representative alignment of Del19-55 mouse cDNA sequence and sequenced trace is presented in Fig. 4.16, showing a few mismatches of individual bases and some indels throughout the sequence and at the junction site. When comparing control trace (generated on SnapGene) and sample trace, it can be observed that mismatches in the cDNA sequences lead to mismatches in the aminoacidic sequence. It must be noted that no stop codons were generated. Further testing would be required in a muscle cell line to draw any conclusions regarding dystrophin protein expression. However, at this stage, deletion of exons 19 to 55 was confirmed on cDNA from samples treated with Guides 14 and 18.

A)



B)

Figure 4.16. Representative alignment of Del19-55 mouse cDNA and sequence trace from cDNA obtained from N2A cells co-transfected with G14 and G18. A) Alignment showing the whole sequenced trace in blue, alignment with Del 19-55 mouse cDNA sequence starting at Exons 17 and ending at Exon 57. B) Zoom-in to the alignment, to see expected sequence of exons 18 and 55 junction aligned to sequencing trace from sample, showing a few indels. From top to bottom: 5´ to 3´ double stranded Del 19-55 mouse cDNA sequence, position at the sequence indicated in bp, amino acids sequence, Exon 18 and Exon 55 indicated in grey boxes, forward strand of Del 19-55 mouse cDNA in bold text as reference for the alignment of sequence from sample, with indels indicated in red boxes, sequencing trace from sample, position in the sample trace in bp and amino acid sequence from sample highlighted in yellow. Red boxes indicate amino acid mismatch.

4.4. DESIGN OF AN AAV MULTIPLEX *SA*CAS9 CONSTRUCT TARGETING INTRON 18 AND 55 OF THE *DMD* GENE, ESTABLISHMENT BY CLONING AND *IN-VITRO ASSESSMENT.*

4.4.1. DESIGN & SUCCESSFUL CLONING OF AN AAV MULTIPLEX SACAS9 CONSTRUCT.

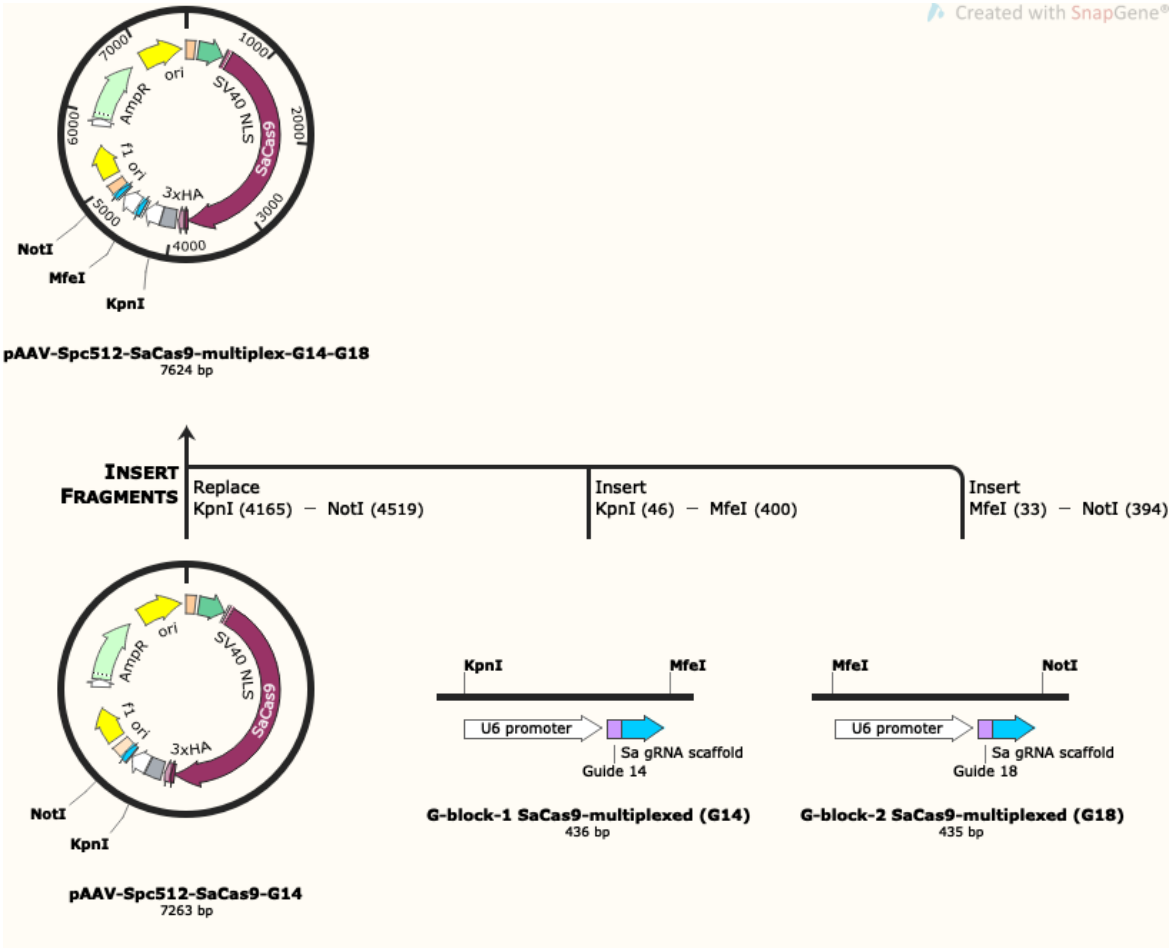
To deliver previously selected most efficient gRNAs (G14 and G18) in the same construct, a multiplex plasmid was designed and built with g-blocks so it would express both gRNAs, each one driven by a U6 promoter and an *Sa*Cas9 driven by an Spc512 promoter. This construct was built on an AAV plasmid backbone so it could then be packaged into AAV vectors for further testing.

To reduce cloning time, two set of g-blocks were designed. A pair already containing Guides 14 and 18 and a pair that would be the "empty" version where gRNAs could be cloned in with BsaI and BbsI restriction sites respectively for each gRNA. The aim of the second construct was to use it as a negative control (no gRNA expressed) and to clone each individual gRNA into a construct in its respective position, to test individual expression.

The cloning strategy of both constructs, from now on referred to as pAAV-Spc512-SaCas9-multiplex-G14-G18 and pAAV-Spc512-SaCas9-BbsI-BsaI (empty), can be seen on Fig. 4.17. Two g-blocks were digested with KpnI and MfeI or MfeI and NotI accordingly, the backbone was digested with KpnI and NotI and a triple-ligation was performed to clone final constructs. Constructs were then mini-prepped, plasmid integrity was

confirmed by restriction digests and sequencing. Both constructs were then maxi-prepped and plasmid integrity was confirmed again by restriction digests and sequencing of the multiplexed cassette and ITR regions, as shown in Fig. 4.18.

A)


Created with SnapGene®
ori
SV40 NLS
SaCas9
3xHA
f1 ori
AmpR
1000
2000
3000
4000
5000
6000
7000
NotI
MfeI
KpnI
pAAV-Spc512-SaCas9-multiplex-G14-G18
7624 bp
INSERT FRAGMENTS
Replace
KpnI (4165) – NotI (4519)
Insert
KpnI (46) – MfeI (400)
Insert
MfeI (33) – NotI (394)
pAAV-Spc512-SaCas9-G14
7263 bp
KpnI
MfeI
U6 promoter
Sa gRNA scaffold
Guide 14
G-block-1 SaCas9-multiplexed (G14)
436 bp
MfeI
NotI
U6 promoter
Sa gRNA scaffold
Guide 18
G-block-2 SaCas9-multiplexed (G18)
435 bp

B)

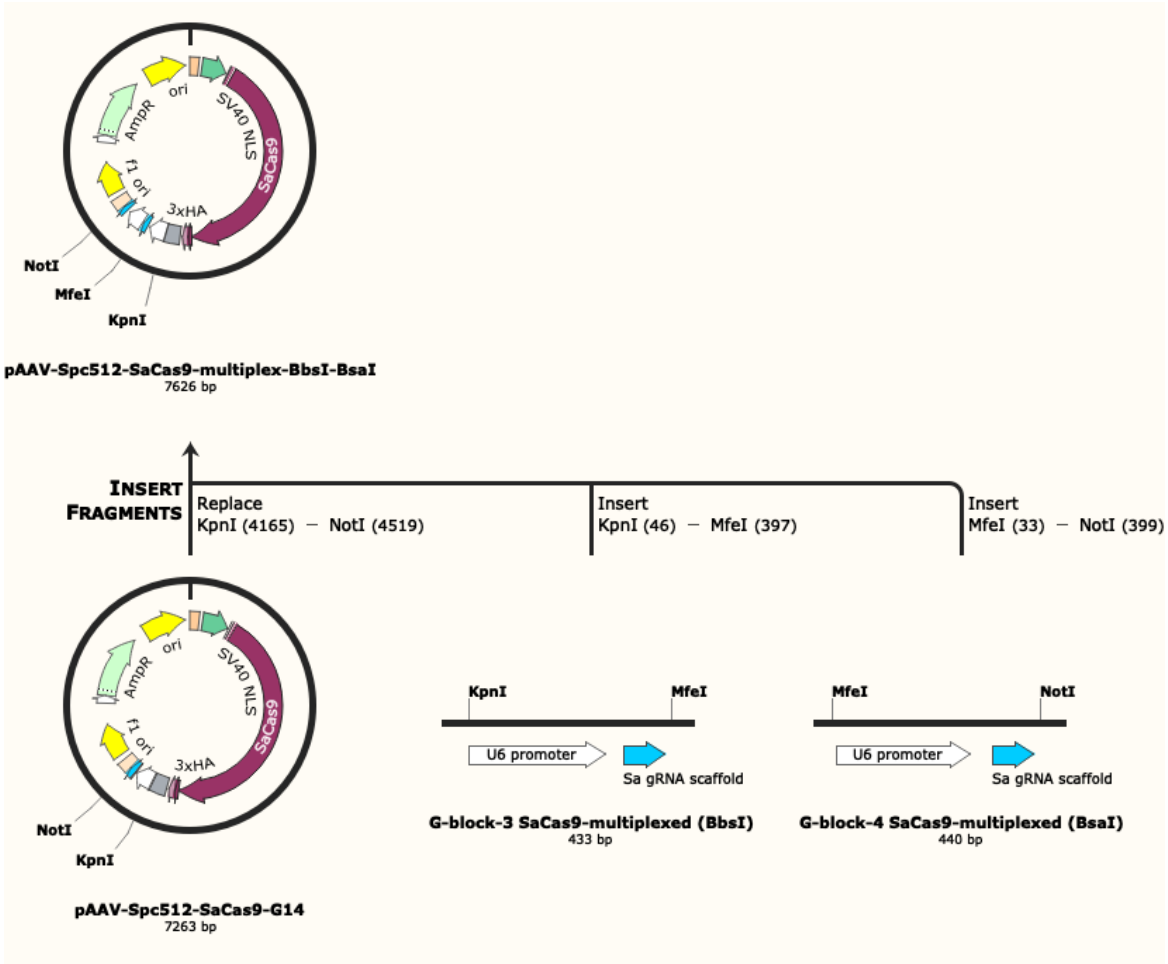


Figure 4.17 Cloning strategy to build AAV plasmids expressing two multiplex gRNAa, each plasmid expressing an *Sa*Cas9 under an Spc512 promoter and two cassettes, each one expressing: a U6 promoter, a gRNA scaffold and terminator. Vector backbone was digested with NotI and KpnI from pAAV-Spc512-SaCas9. G-blocks were amplified by PCR with appropriate primers, PCR products were cleaned and digested with KpnI and MfeI or MfeI and NotI accordingly. Finally, two g-blocks and a backbone were triple-ligated using a 1:2 backbone to insert ratio and 100 ng of backbone DNA per reaction. A) Cloning strategy to build pAAV-Spc512-SaCas9-Multiplex-G14-G18, multiplex construct expressing Guides 14 and 18. B) Cloning strategy of pAAV-Spc512-SaCas9-BbsI-BsaI. "Empty" plasmid that allows insertion of two gRNA, each one on a different restriction site, BbsI and BsaI.

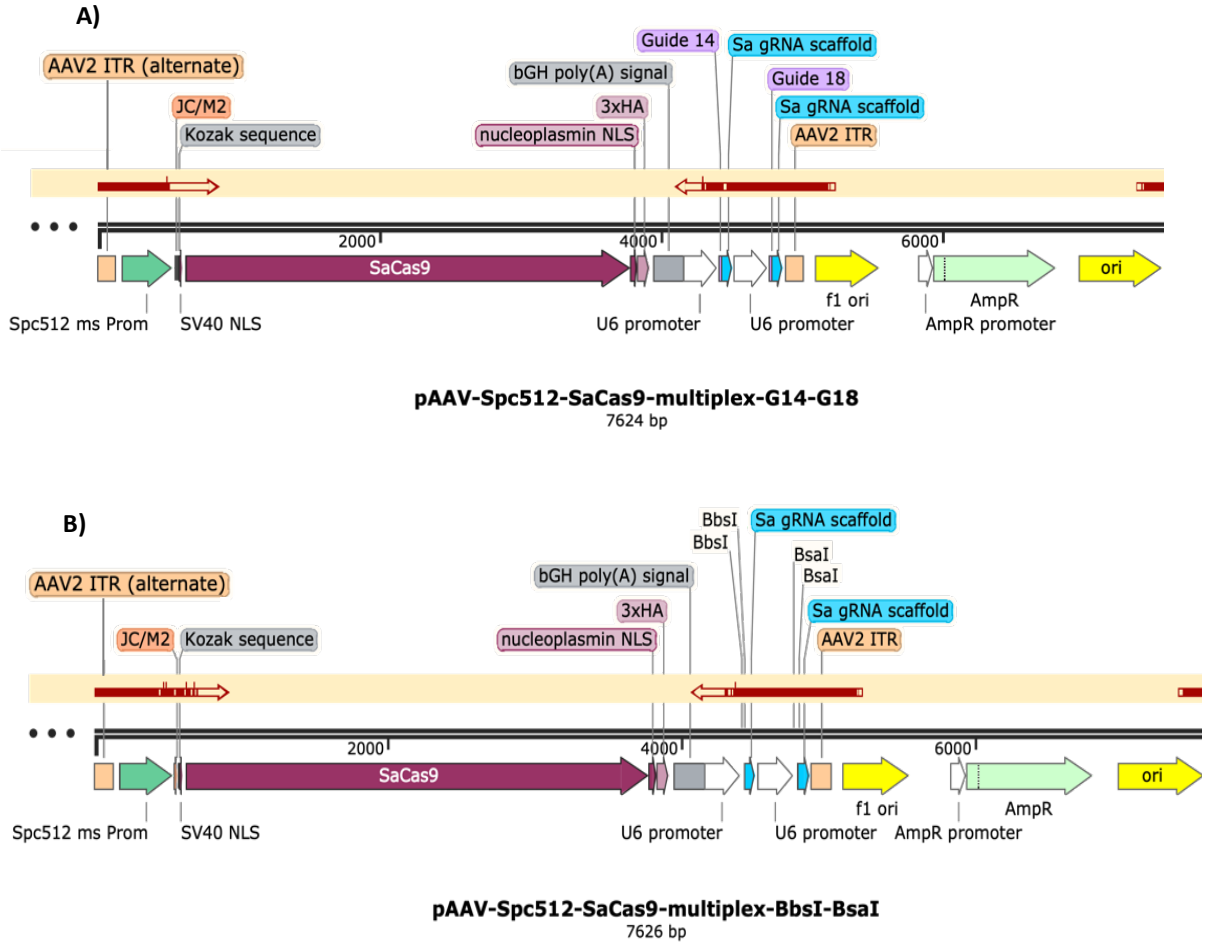


Figure 4.18. Alignments of plasmid maps and sequencing traces of ITRs and multiplex gRNA cassettes from plasmid maxi-preps. Plasmid maps and alignments generated with SnapGene Software. Plasmid sequencing trace indicated as a red arrow on top of plasmids map. Red filling indicates alignment with no mismatches. Samples were sent for ITR sequencing to GeneWiz (now Azenta Life Sciences) with the following primers: 5´-AGC GTG AGC TAT GAG AAA GC-3´ for the 5´ ITR region and 5´-CCG ATT TAG AGC TTG ACG GG-3´ for the 3´ ITR region. A) pAAV-Spc512-SaCas9-multiplex-G14-G18 map alignment with ITRs sequencing, showing no mismatches on ITRs nor in the multiplex gRNA cassettes. B) pAAV-Spc512-SaCas9-BbsI-BsaI alignment with ITRs sequencing, showing no mismatches on ITRs nor in the multiplex gRNA cassettes. The last segments of alignments tend to mismatch (after 800 bp) as the sequencing trace ends.

4.4.2. *In-vitro assessment of pAAV-Spc512-SaCas9-multiplex-G14-G18 construct by transient transfection on N2A cells alongside co-transfection of Guides 14 and 18.*

4.4.2.1. *Confirmation of the generation of a de novo intronic junction of introns 18 and 55.*

The multiplex construct (pAAV-Spc512-SaCas9-multiplex-G14-G18) was transfected into N2A cells with Viafect (4:1 ratio to DNA) with a 4 µg plasmid DNA dose alongside a co-transfection of the individual gRNAs (G14 and G18) cloned into pAAV-CMV-SaCas9 (2 µg of each plasmid per transfection reaction). DNA was extracted 48 hrs. after transfection and a PCR was performed with a similar primer pair to the one described in Section 4.3.1.1. This new optimised, more specific, primer pair also targets intron 18 (5´-CCCAGGCAAACATGATACAATTAG -3´) and intron 55 (5´- GAACCAGAGTACAGGGTGAAAG -3´) but produces no additional bands. If there was a successful deletion from intron 18 to intron 55, a PCR product of ~970 bp would be produced. PCR products of ~970 bp can be observed for samples treated with our multiplex construct and for co-transduced samples, in Fig. 4.19, confirming a successful deletion with both treatments.

PCR products were purified and sent for Sanger sequencing to confirm a deletion between introns 18 and 55. Both sets of samples, the ones transfected with the multiplex construct and the ones co-transfected with individual gRNAs showed the expected deletion, as shown with representative samples in Fig. 4.20.

Once it was confirmed that both approaches, the multiplex construct and co-transfection of individual gRNAs, could achieve the deletion between introns 18 and 55 on N2A cells, the next step was to test them on a mouse muscle cell line.

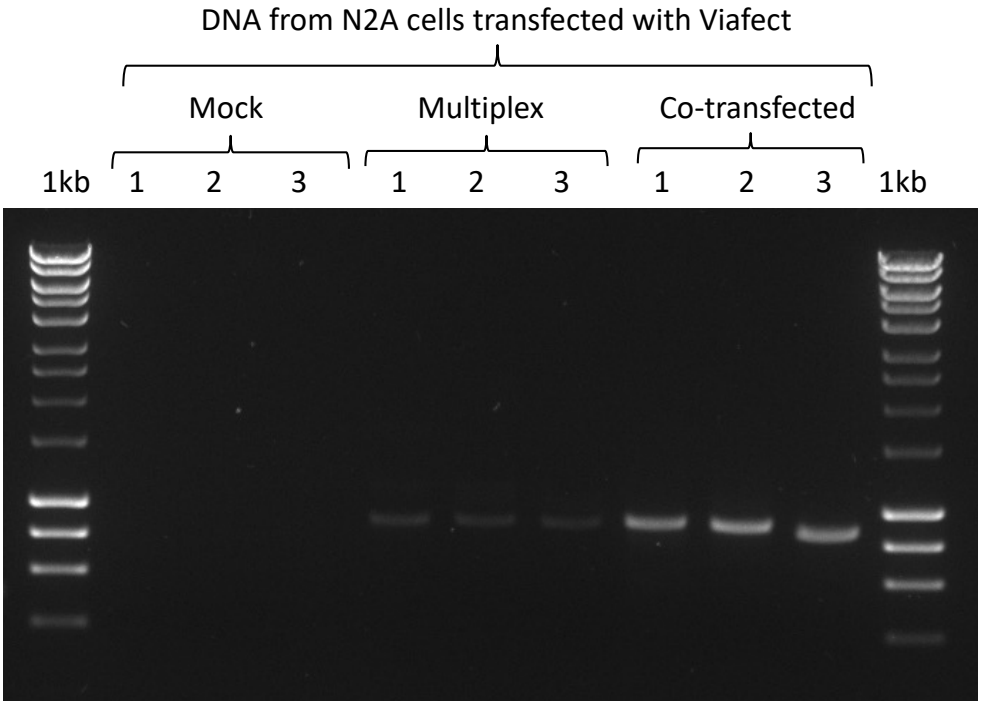


Figure 4.19. Gel image of PCR products resulting from a deletion between introns 18 and 55 in *Dmd*. PCR products from DNA extracted 48 hours after treatment from N2A cells transfected using Viafect 4:1 to DNA (4 µg/transfection) with pAAV-Spc512-SaCas9-multiplex-G14-G18 and co-transfected with Guide 14 and Guide 18 (2 µg each) (n = 3 technical repeats). Samples ran on a 1% (w/v) agarose gel with 0.5X SYBR Safe in 1X TAE Buffer. From left to right: Hyperlader I, Mocks (samples treated with Viafect and no DNA), PCR products of expected size (~970 bp) indicating a deletion caused by transfection of pAAV-Spc512-SaCas9-multiplex-G14-G18 or co-transfection of pAAV-CMV-SaCas9-G14 and pAAV-CMV-SaCas9-G18.

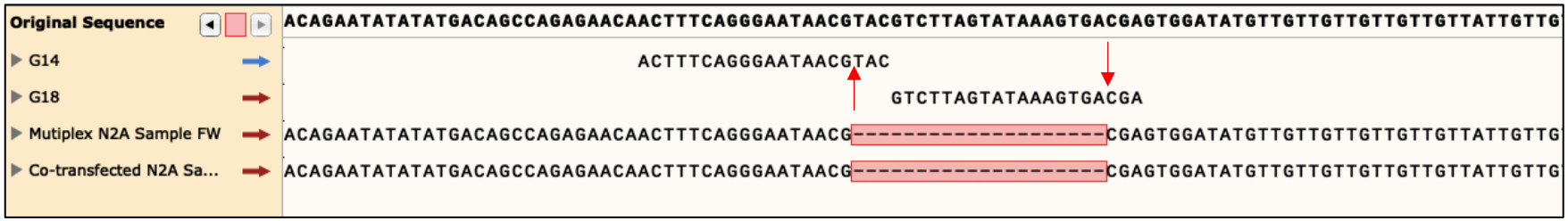


Figure 4.20. Representative alignment of samples with a deletion between introns 18 and 55. Alignment of: Del19-55 *Dmd* sequence (original sequence), Guide 14 sequence (G14), Guide 18 sequence (G18) and sequencing traces from DNA extracted of N2A cell samples transfected with pAAV-Spc512-SaCas9-multiplex-G14-G18 (Multiplex N2A Sample) and co-transfected with pAAV-CMV-SaCas9-G14 and pAAV-CMV-SaCas9-G18 (Co-transfected N2A Sample). Traces from both transfections show a junction of the cut site of Guide 14 and the cut site of Guide 18. Cut sites indicated on the guide sequence with red arrows. Dotted line highlighted in red indicates there is no alignment of those base pairs. Alignments generated on Snapgene Software.

4.4.3. Mouse muscle cell line (C2C12 cells) nucleofected with multiplex SaCas9 system and individual gRNAs targeting introns 18 and 55 of Dmd.

To test our multiplex *Sa*Cas9 constructs with the muscle specific Spc512 promoter, a nucleofection of mouse muscle cell line C2C12 was performed, since transient transfections in this cell line is of low efficiency. First, a dose response with pX601-CMV-SaCas9-GFP was performed to confirm optimal DNA dose. Then, C2C12 cells were nucleofected with the multiplex *Sa*Cas9 system alongside individual gRNAs and co-delivered gRNAs.

For both experiments a 4D-Nucleofector X Unit from Lonza was used following the 4D-Nucleofector Protocol for C2C12 cells. Cells were harvested 48 hours post-nucleofection for DNA and RNA extraction.

4.4.3.1. Plasmid DNA dose response on C2C12 cells delivered by nucleofection.

The 4D-Nucleofector Protocol for C2C12 cells recommends a dose of 1 to 5 µg of plasmid DNA per 100 µL cuvette. A dose response with 2 to 6 µg of plasmid DNA was set-up with pX601-CMV-SaCas9-GFP. Cells were assessed by fluorescence microscopy 24 and 48 hours after nucleofection to confirm GFP expression and were then harvested for FACS Analysis. Percentage of GFP positive cells was calculated on FloJo Software and plotted on Prism9 Software. Results are presented in Fig. 4.21, where it can be observed that

with a 4 µg DNA dose, ~55% of cells were nucleofected efficiently. Even though a higher efficiency was achieved with a 6 µg dose (~65%), it was decided to stick to the DNA range suggested by the protocol and use a 4 µg dose for future experiments.

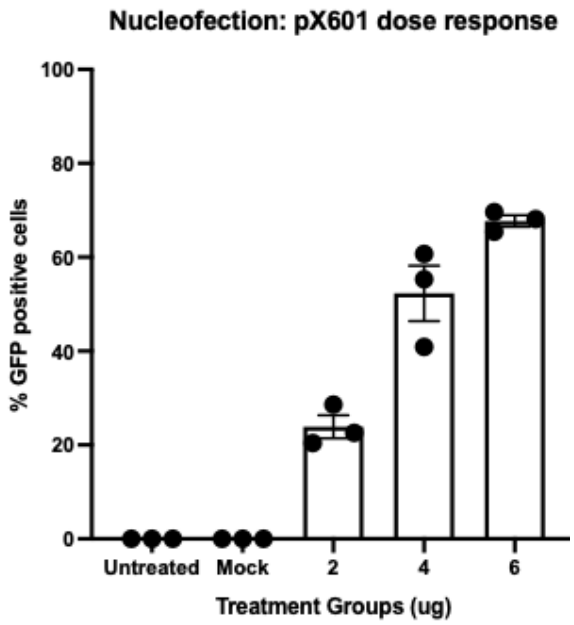


Figure 4.21. Dose response results from FACS Analysis of pX601-CMV-SaCAS9-GFP nucleofection into C2C12 cells. Bar chart shows GFP positive cells percentage according to different doses of plasmid. A 4D-Nucleofector X Unit from Lonza was used to nucleofect cells in suspension in 100 µL cuvettes following the 4D-Nucleofector Protocol for C2C12 cells. Post-nucleofection, cells were incubated on a 6-well plate at 37ºC/5% CO2 and harvested 48 hours later for FACS Analysis on a FACS Canto II machine (from BD Biosciences), FACSDiva Sofware used for acquisition and gating (n = 3 technical repeats). FloJo Software used for data analysis and Prism9 used to generate bar chart. Error bars represent standard error of the mean.

4.4.3.2. C2C12 CELLS NUCLEOFECTION AND CONFIRMATION OF A DELETION BETWEEN INTRONS 18 AND 55 AT GENOMIC DNA LEVEL.

A second nucleofection was set-up using 4 µg per reaction of the following plasmids:

- pAAV-Spc512-Del19-55-GFP (as a positive control)
- pAAV-Spc512-SaCas9-BbsI-BsaI (as a negative control)
- pAAV-Spc512-SaCas9-multiplex-G14-G18
- Individual gRNAs (pAAV-CMV-G14 and pAAV-CMV-G18)
- Co-nucleofection of pAAV-CMV-G14 + pAAV-CMV-G18 (2 µg of each plasmid)

Samples were harvested 48 hours after nucleofection and DNA was extracted. DNA was analysed by PCR to assess if a deletion between introns 18 and 55 was achieved by pAAV-Spc512-SaCas9-multiplex-G14-G18 and by the co-nucleofection of both gRNAs. Additionally, efficiency of individual gRNAs on C2C12 cells was assessed.

To assess the deletion, previously designed PCR primers (Section 4.4.2.1) binding to intron 18 and 55 (5´- CCCAGGCAAACATGATACAATTAG -3´ and 5´- GAACCAGAGTACAGG GTGAAAG -3´), which produce a 970 bp product if exons 19 to 55 are deleted, were used. Bands with low intensity matching the size of the expected product can be observed from samples co-nucleofected with G14 and G18 and one of the samples nucleofected with the multiplex construct (Fig. 4.22). Some smaller bands can be observed in these samples. PCR products (~970 bp) from samples “Multiplex”-2 and “Co-nucleofected”-2

were gel extracted, purified and sent for sequencing. It was not possible to extract smaller bands as the intensity was very low.

The alignment of the sequence from extracted samples against the sequence of Del19-55 *Dmd* can be seen on Fig. 4.23, showing achievement of the expected deletion on the co-nucleofected sample. Unfortunately, the trace for the “multiplex” sample showed a high background noise after the cut site of Guide 14. Even though it can be speculated that the noise was caused by successful editing, conclusions cannot be made based on this sequence alignment.

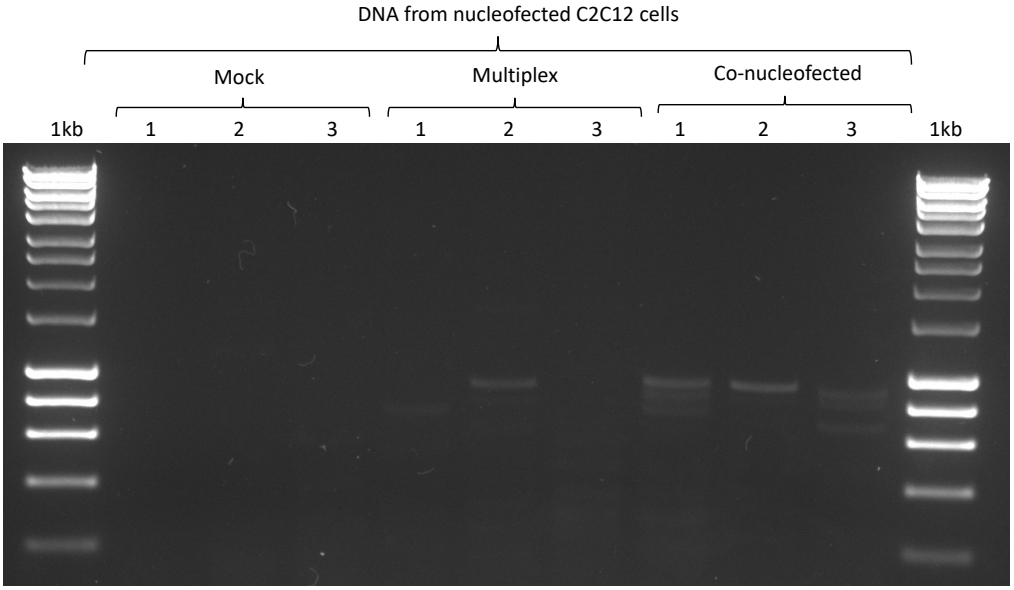


Figure 4.22. Gel image of PCR products from DNA extracted from C2C12 cells 48 hours after nuclefection with 4D-Nucleofector X Unit from Lonza. "Multiplex" samples were nucleofected with pAAV-Spc512-SaCas9-multiplex-G14-G18 and "Co-nucleofected" samples with pAAV-CMV-SaCas9-G14 and pAAV-CMV-SaCas9-G18 (n = 3 technical repeats per group). If Exons 19 to 55 were deleted, a PCR product of 970 bp was expected. "Co-nucleofected" samples showed expected bands and some unexpected additional smaller bands. Only one of the "Multiplex" samples showed the expected band. Gel was 1% agarose (w/v) with 0.5X SYBR Safe in 1X TAE Buffer. Hyperladder I was used.

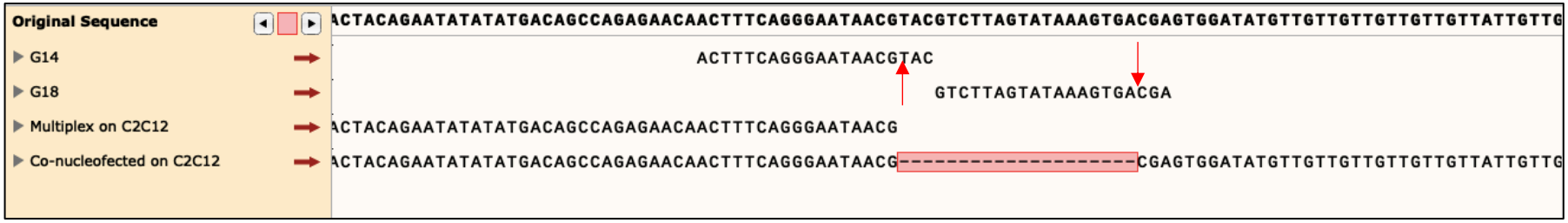


Figure 4.23. Alignment on SnapGene Software of Del19-55 *Dmd* mouse sequence and nucleofected C2C12 cell samples. From top to bottom: Del19-55 *Dmd* sequence (original sequence), Guide 14 (G14), Guide 18 (G18), sample nucleofected with pAAV-Spc512-SaCas9-multiplex-G14-G18 (multiplex on C2C12) and sample co-nucleofected with pAAV-CMV-SaCas9-G14 and pAAV-CMV-SaCas9-G18 (co-nucleofection on C2C12). Cut sites indicated on the guide sequence with red arrows. Dotted line highlighted in red indicates there is no alignment of those base pairs.

4.4.3.3. Assessment of individual gRNA editing efficiency on nucleofected C2C12 cells.

To assess individual gRNA efficiency on C2C12 cells, previously designed primers flanking the cut site for each gRNA were used for PCRs: for Guide 14 primers forward 5´- CCCAGGCAAACATGATACAATTAG -3´ and reverse 5´- AGCATGAGAGCAAAGGTGAG -3´ and for Guide 18 primers forward 5´- GCTAATCAAATCTGTGCATGGT -3´ and reverse 5´- CTGGTCCATGCCTAACCATAT -3´. PCR products with the expected size for each gRNA can be observed on Fig. 4.24.

After confirming a unique PCR product with the expected size, PCR samples were purified and sent for Sanger sequencing with the forward primers. Guide RNA efficiency was assessed by TIDE Analysis with TIDE web tool, results are presented on Fig. 4.25. Some editing was observed with both gRNAs. However, there was a lot of variability between samples.

One of the samples treated with G14 showed 5% editing while the two other samples showed ~1% editing, while editing with G18 varied from 9.7% to 25.7% to 31%. It must be noted that that the region downstream of Guide 18 cut site is an AT rich region that could interfere with Sanger sequencing and cause variations between samples (even mock samples). This can be observed on the representative output from TIDE Analysis web tool showed on Fig. 4.26.A where the control trace (in black) presents some

background, rather than a low equally distributed signal (for reference, Fig 4.9.A presents no background on control trace). Furthermore, the spread of indels from Guide 18 (Fig. 4.26.B) seems strange, as the majority of significant indels (11.6%) are a deletion of 9 bp, rather than the typical spread of 1-6 bp deletions and 1-2 bp insertions caused by a DSB with CRISPR/Cas9.

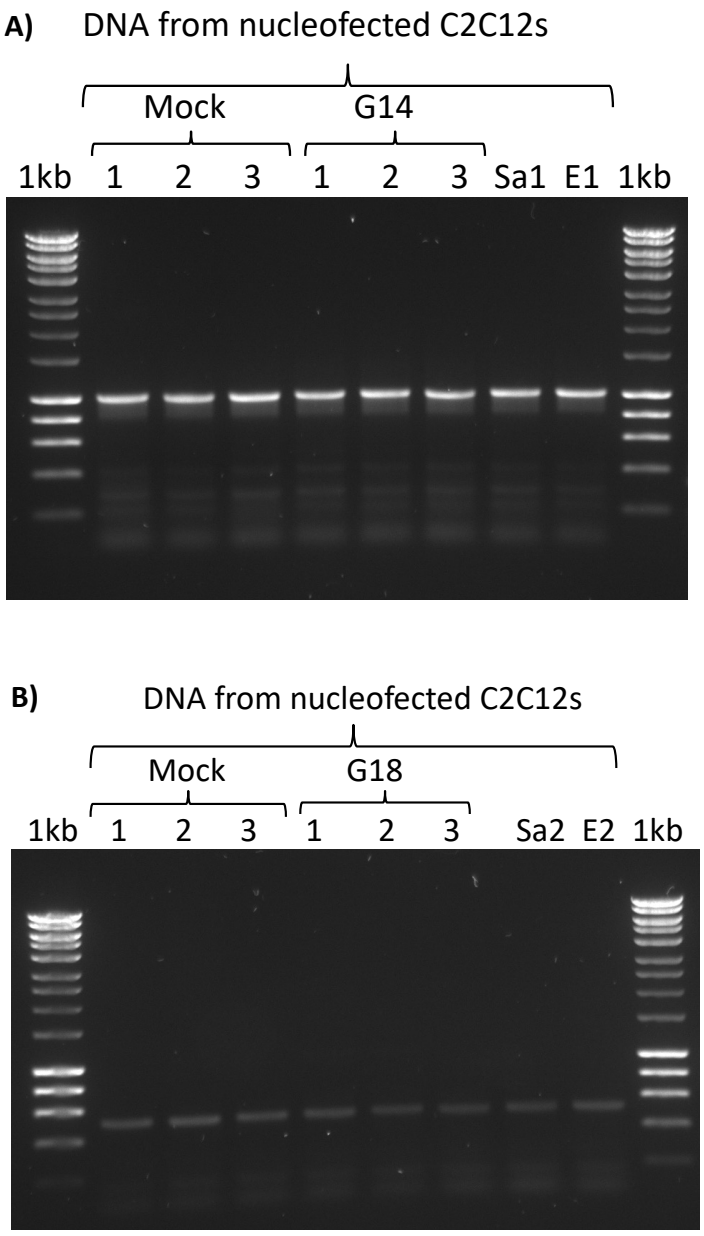


Figure 4.24. Gel images of PCR products from DNA extracted from C2C12 cells 48 hours after nuclefection with 4D-Nucleofector X Unit from Lonza. A) PCR products of expected size: 1073 bp from samples nucleofected with pAAV-CMV-SaCas9-G14 (by triplicates), pAAV-CMV-SaCas9 (Sa1) and pAAV-Spc512-SaCas9-BbsI-BsaI (E1). PCR primers used: forward 5´- CCC AGG CAA ACA TGA TAC AAT TAG -3´and reverse 5´- AGC ATG AGA GCA AAG GTG AG -3´. B) PCR products of expected size (548 bp) from samples nucleofected with pAAV-CMV-SaCas9-G18 (in triplicate), pAAV-CMV-SaCas9 (Sa2) and pAAV-Spc512-SaCas9-BbsI-BsaI (E2). PCR primers: forward 5´- GCT AAT CAA ATC TGT GCA TGG T -3´ and reverse 5´- CTG GTC CAT GCC TAA CCA TAT -3´. Both gels were 1% agarose (w/v) with 0.5X SYBR Safe in 1X TAE Buffer. Hyperladder I was used (1kb).

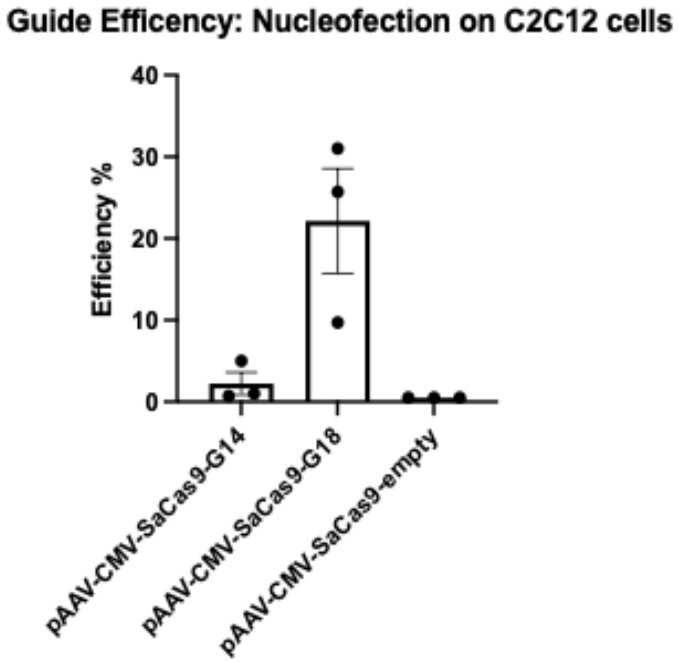


Figure 4.25. Individual gRNA cutting efficiency based on TIDE Analysis, bar chart shows efficiency in percentage of Guide 14 (targeting intron 18 of *Dmd* gene) and Guide 18 (targeting intron 55). Sequence traces of PCR products from amplification around the predicted target sites were analysed by TIDE web tool (Brinkman et al., 2014). Each guide was nucleofected in triplicate. Data was analysed on Prism9 Software. Error bars represent standard error of the mean.

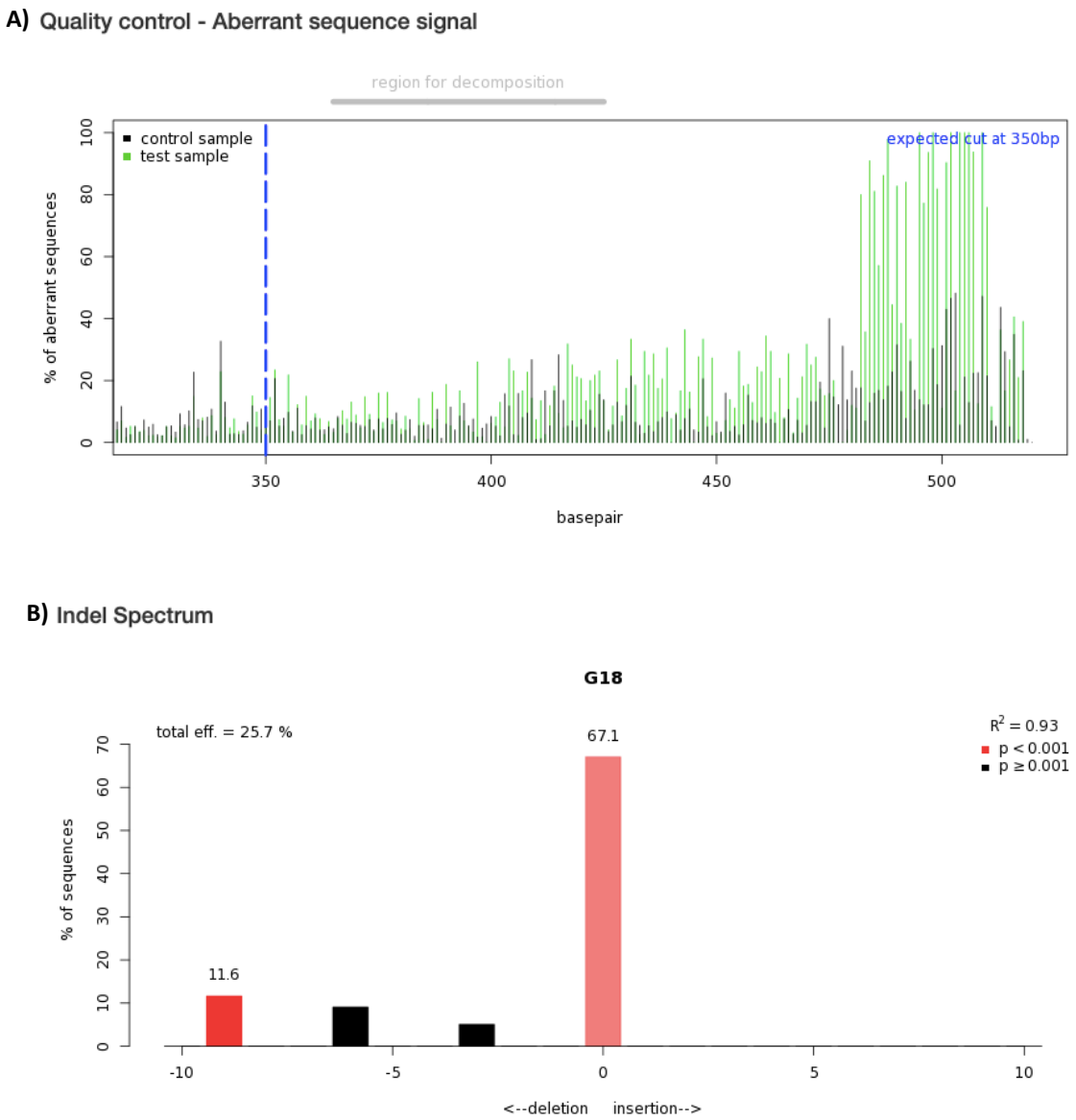


Figure 4.26. Representative images of outputs from TIDE Analysis Web Tool. Analysis of Guide 18 presented. A) Decomposition trace, aberrant sequence signal (green) compared to control trace (black). Dotted blue line indicates cut site. B) Bar chart indicating indel spectrum output. X-axis indicates small deletions of up to 10 base pairs on a negative scale (-10 to 0) and insertions on a positive scale (0 to 10). In this example, the red bars indicate 61.7% of traces had 0 deletions or insertions, 11.6% had -9 deleted bp and lower percentages of edited populations (black bars) indicate some -3 and -6 bp deletions. These percentages add up to a total efficiency of 25.7% from Guide 18, indicated at the top left of the graph. Numbers at the top right corner denote the coefficient of determination (R^2), to evaluate model accuracy (values from 0 to 1). Low R^2 can be due to poor sequence quality or non-optimal setting. P-values indicate significance cutoff, set up at $p<0.001$. Significant outputs in red, non-significant ($p \geq 0.001$) in black.

4.5. DISCUSSION.

In the first section of this chapter, gRNAs for *Sa*Cas9 targeting introns 18 and 55 were designed for mouse and human *Dmd*/*DMD*. The *in-silico* analysis of efficiency and specificity scores obtained from online tools when designing gRNA for CRISPR systems, do not always translate to *in-vitro* or *in-vivo* efficiency. Thus, to increase probabilities of finding a highly efficient gRNA, scores were considered but also at least five gRNAs per target were tested *in-vitro*.

Off-target events were only evaluated *in-silico*. Off-target events assessment could have been improved by an *in-vitro* evaluation performing PCRs with primers designed to target potential off-target sites. Nevertheless, considering 44 gRNAs would be screened, additionally screening off-targets per gRNA would have been a very long process. This could have been circumvented by whole genome sequencing to confirm any off-target activity of each individual gRNA, it might be relevant to consider this approach in the future for the gRNAs selected for further testing.

*Sa*Cas9 gRNAs targeting introns 18 and 55 of human and mouse *DMD/Dmd* genes were successfully cloned into pAAV-CMV-SaCas9, a plasmid expressing an *Sa*Cas9 driven by a CMV promoter. Guide RNAs were screened by transfection, each construct on triplicates, on appropriate cell lines, HEK293T cells for human and N2A cell for mouse gRNAs. DNA was extracted from harvested cells and appropriate PCR primers were

design to target the flanking sequence of each site targeted by a gRNA. PCR products were sequenced and analysed by TIDE assay.

It is important to note that multiple assays are available to determine level of activity of CRISPR/Cas9 gRNAs. Some of the frequently used assays for this purpose can be classified into two categories based on their main technique: enzyme mismatch cleavage (EMC) detection assays and detection of indels by sequencing of edited populations. The most common ones based on EMC are the T7 endonuclease 1 (T7E1) mismatch detection assay (Mashal et al., 1995) and the Surveyor EMC assay (Oleykowski et al., 1998). Both assays have been compared previously and it was found that T7E1 is more sensitive to detect deletions, while the Surveyor nuclease is better at detecting single nucleotide changes (Vouillot et al., 2015). However, authors (Vouillot et al., 2015) preferred T7E1 assay to scan for mutations caused by engineered nucleases as this method was more sensitive, with a detection limit of ∼5% mutant DNA, while Surveyor assay limit was ∼10%. The most common assays involving sequencing are: (i) targeted next-generation sequencing (NGS) (Bell et al., 2014), which involves high costs, (ii) Indel Detection by Amplicon Analysis (IDAA) assay (Z. Yang et al., 2015), a multiple step protocol involving PCR amplicon labelling and capillary electrophoresis, which can also be coupled to FACS analysis (Lonowski et al., 2017) and (iii) Tracking for Indel by Decomposition (TIDE) assay (Brinkman et al., 2014), which only involves sequencing of PCR products from edited and wild type populations.

Accuracy of four of these frequently used assays (T7E1, TIDE, IDAA and NGS) was compared in a study (Sentmanat et al., 2018) and it was demonstrated that the T7E1 assay often incorrectly reports gRNA activity due the low dynamic range and DNA heteroduplex formation requirement. Additionally, it has a low detection range that plateaus at 30-40% edited pools. In contrast, TIDE and IDAA assays showed reliable prediction of overall gRNA activity comparable to NGS (Sentmanat et al., 2018).

To assess gRNA efficiency in this project, TIDE analysis was the preferred method as it is a reliable cost-effective assay for screening multiple samples. However, this assay also has some limitations that need to be considered: TIDE relies on high quality sequencing traces, which are not always easily obtained from every target sequence and it most accurately predicts indels of a limited size (10 bp); this range can be adjusted but it would reduce confidence level (Brinkman et al., 2014). Furthermore, the intronic regions flanking the targets of G14 and G18 are AT rich regions with repeats that could affect the quality of Sanger sequencing and interfere with accurate assessment of gRNA activity. It should also be noted that all transfections were performed with an n = 3 technical repeats. To assess gRNA efficiency variability, more biological repeats could be performed including transfections of cell lines obtained from different sources or including other cell lines, such as C2C12 mouse muscle cell line for gRNA screening.

After analysing results from TIDE of all gRNAs, it was decided to proceed further testing with the most efficient gRNAs designed for mouse *Dmd*, as these could potentially be

tested *in-vivo*. The most efficient gRNA targeting mouse Intron 18 was Guide 14 (ACTTTCAGGGAATAACGTAC) and the most efficient one targeting Intron 55 was Guide 18 (GTCTTAGTATAAAGTGACGA). Constructs expressing these gRNAs were co-transfected on N2A cells to test if these gRNAs could achieve the deletion of exons 19 to 55 *in-vitro*. DNA and RNA were extracted from cells 48 hours after transfection and processed accordingly. Sequencing of PCR products from DNA and cDNA, obtained from the co-transfections, confirmed the deletion of exons 19 to 55.

A multiplex *Sa*Cas9 system that could be packaged into an AAV vector was designed. Two constructs were cloned, pAAV-Spc512-SaCas9-multiplex-G14-G18, expressing an *Sa*Cas9 from an Spc512 promoter and two multiplex gRNAs (Guide 14 and 18), each one in an individual cassette expressing a gRNA, gRNA scaffold and a terminator under a U6 promoter; the second construct was the "empty" version with no gRNAs cloned into the BbsI and BsaI restrictions sites. Constructs were transfected on N2A cells and nucleofected on C2C12 muscle cells, alongside co-delivery of individual gRNAs. Both approaches (multiplex and co-delivery), achieved a deletion between introns 18 and 55 on both cell lines, confirmed by Sanger sequencing of amplicons from DNA obtained from treated cells. However, it was only possible to confirm deletions of exons 19 to 55 from cDNA obtained from N2A cells. The primer pair that confirmed this deletion on N2A was not specific when used in cDNA from C2C12 cells. Integrity of the cDNA from C2C12s was confirmed by PCR with primer pairs targeting reference gene Rplp0. Then, additional approaches were attempted to assess deletion of exons 19 to 55: additional

primer pairs were screened and nested and semi-nested PCRs were also attempted. If time had permitted, these samples could have been further analysed by RT-qPCR to detected deletion of exons 19 to 55.

It is relevant to note that when nucleofected, gRNAs showed high variability in their efficiency. This could be due to variability in the delivery, variability in cell seeding or pipetting errors. It is relevant to highlight that GFP expression from the control plasmid had previously been demonstrated on N2A cells by transfection with Viafect. However, it was not possible to detect GFP expression form this plasmid by fluorescence microscopy in nucleofected C2C12s. It was expected that the Spc512 promoter from the control plasmid would express well in C2C12 cells. Nevertheless, it is possible that delivery efficiency was low, or the control had a weak expression. These theories could have been further investigated by harvesting protein from nucleofected C2C12 cells and assessing *Sa*Cas9 protein expression by Western Blots. Unfortunately, due to time concerns, it was not possible to repeat a nucleofection on C2C12 cells and harvest cells for protein extraction. For this reason it was not possible to assess Del19-55 dystrophin expression in edited cells.

Another limitation at this stage was the quantification of the deletion efficiency in edited populations. It can be estimated that the limiting factor was the efficiency achieved by Guide 14. The highest efficiency achieved by G14 in C2C12 cells was ~5%, therefore we can expect a deletion efficiency ≤5%. The difference in efficiency between both gRNAs

(20% on average for G18) could also lead to asymmetric cleavage, leading to cut, repair and formation on indels in one of the target sites before the second target is reached, which would make the locus refractory to further editing (Hanson et al., 2022). However, this was not confirmed experimentally. This could be assessed by performing PCRs of the target regions in intron 18 and intron 55 respectively in edited samples and compare individual gRNA activity on each end when co-delivered.

Nevertheless, once it was confirmed that the multiplex construct was expressing correctly and achieving deletion of exons 19-55, it was decided to package it into an AAV vector to increase delivery efficiency and further test the construct *in-vivo*.

5. AAV9 PRODUCTION & ASSESSMENT OF TRANSDUCED MULTIPLEX *SA*CAS9 CONSTRUCT & CO-TRANSDUCED GRNAS, TARGETING INTRONS 18 AND 55 IN *MDX* MICE.

Adeno-associated virus (AAV) vectors have been widely studied and have shown efficient delivery in gene therapy clinical trials (Kotterman & Schaffer, 2014). Multiple AAV serotypes have been identified and their natural tropism has been studied (Zincarelli et al., 2008). In the context of muscle delivery, AAV9 serotype has been widely used and shown robust tissue expression and a natural tropism for skeletal muscle and heart (Qiao et al., 2011, Gruntman et al., 2013). Thus, this serotype was selected as the delivery system for this project.

Furthermore, various studies have used AAV9 vectors to deliver *Sa*Cas9 CRISPR systems targeting the *DMD* gene in *mdx* mice and achieved successful excision of exon 23, restoring the reading frame and leading to dystrophin expression (Nelson et al., 2016, Tabebordbar et al., 2016, Hanson et al., 2022)

In the previous chapter, a multiplex *Sa*Cas9 construct expressing two gRNAs targeting introns 18 and 55 and an *Sa*Cas9 driven by an Spc512 promoter was established and assessed *in-vitro*; alongside the co-delivery of G14 and G18 in their respective plasmids

expressing an *Sa*Cas9 under a CMV promoter. Both strategies achieved a deletion and the generation of a *de novo* junction between introns 18 and 55. In order to assess efficiency of these strategies *in-vivo*, it was decided to package our constructs into AAV9 vectors to transduce *mdx* mice.

The milestones for this chapter are summarised below:

- Clone individual gRNAs (G14 and G18) into pAAV-Spc512-SaCas9-BbsI-BsaI (empty) construct to assess individual gRNA efficiency in muscle tissue, with the *Sa*Cas9 driven by an Spc512 promoter.
- Produce and titre AAV9 vectors packaging our multiplex *Sa*Cas9 construct and individual gRNAs targeting introns 18 and 55 of *Dmd*.
- Assess editing efficiency, protein expression and functionality, of AAV9 vectors with packaged constructs *in-vivo* in 2-month-old *mdx* mice by intramuscular injection of TA muscles.
- Transduce AAV9 vectors with packaged constructs *in-vitro* if further assessment is needed.

5.1. PRODUCTION OF AAV9 VECTORS PACKAGING MULTIPLEX *SA*CAS9 CONSTRUCTS AND PLASMIDS WITH INDIVIDUAL GRNAS.

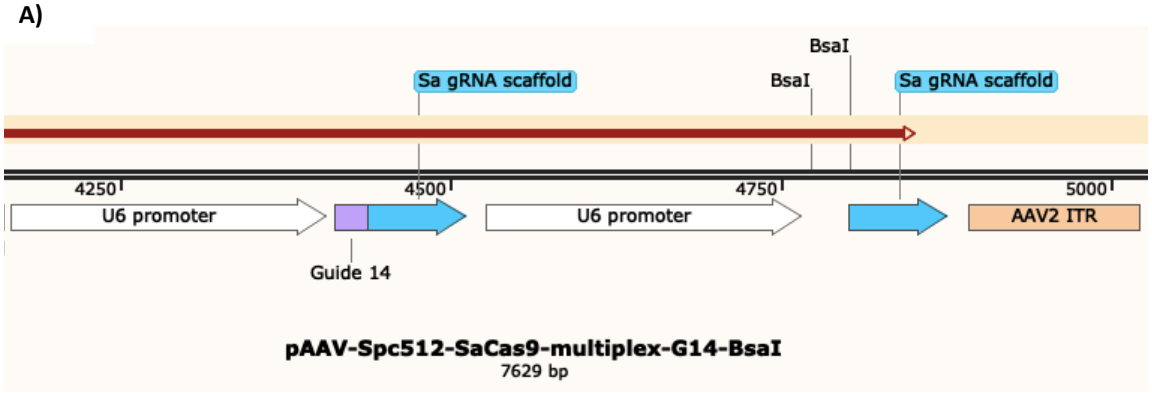
5.1.1. AAV9 VECTORS PRODUCTION: CLONING, CELL CULTURE & PURIFICATION BY LIQUID CHROMATOGRAPHY.

To test our multiplex *Sa*Cas9 construct (pAAV-Spc512-SaCas9-multiplex-G14-G18) *in-vivo* using AAV9 vectors, firstly two more constructs were cloned: pAAV-Spc512-SaCas9-G14-BsaI and pAAV-Spc512-SaCas9-BbsI-G18. Guide 14 was cloned with BbsI into a backbone from pAAV-Spc512-SaCas9-BbsI-BsaI digested with BbsI and Guide 18 was cloned with BsaI on a backbone from pAAV-Spc512-SaCas9-BbsI-BsaI digested with BsaI. These constructs were cloned to test individual gRNA efficiency considering their location in different cassettes of the multiplex construct and to co-deliver them to compare efficiency against multiplex gRNAs.

Once all constructs were giga-prepped, plasmid integrity was confirmed by restriction digest and correct gRNA insertion was confirmed by sequencing, shown in Fig. 5.1, and constructs were packaged into AAV9 vectors as described in Materials & Methods Section 2.13. The crude lysate was then purified by chromatography with the AKTA go protein purification system and a Poros AAV9 SN00068 Column from ThermoFisher. Eluted fractions containing virus were neutralized with Tris-HCL and injected into a dialysis cassette (10,000 Molecular weight cut off). Samples were left on the dialysis cassette overnight in 1X PBS. Next day, samples were recovered from dialysis cassettes

and concentrated with an Amicon Ultra-15 Centrifugal Filter Unit with Ultracel-100 membrane (Millipore UFC9110024) until left with ~450 µL of each prep.

A)



B)

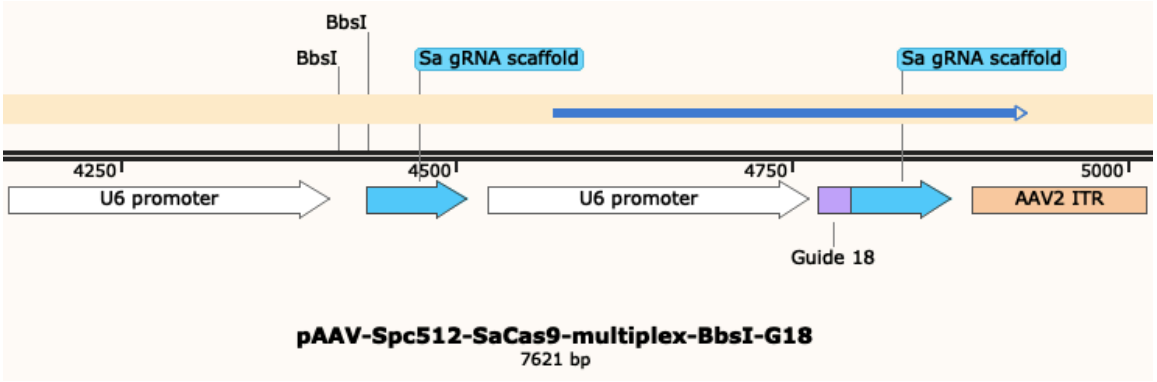


Figure 5.1. Alignments of plasmid maps and sequencing traces from samples confirming gRNA insertion. Plasmid maps and alignments generated with SnapGene Software. A) pAAV-Spc512-SaCas9-G14-BsaI plasmid map alignment with giga-prep sample sequencing trace (red box) confirming insertion of G14 (Sequencing primer: 5´- CAC TCC CAC TGT CCT TTC CT -3´). B) pAAV-Spc512-SaCas9-BbsI-G18 plasmid map alignment with giga-prep sample sequencing trace (blue box) confirming insertion of G18 (Sequencing primer: 5´- CCG AGG GCC TAT TTC CCA TGA TTC -3´).

5.1.2. Optimisation of primer pairs for AAV9 Titration by qPCR.

AAV titration was done by quantifying viral genome copy numbers by qPCR (Materials and Methods Section 2.14.2). To optimize the titration, three qPCR primer pairs were designed to target the *Sa*Cas9 sequence, present in all the constructs that were packaged into AAV9 vectors (except pAAV-Spc512-GFP control plasmid). A PCR gradient from 57ºC to 61ºC was set up to find optimal Tm. Primer pairs seemed to work well with temperatures in this range. Two temperatures were selected (57ºC and 60ºC, one on the lower end and one on the higher end of the range) to compare their binding efficiency by qPCR. A standard curve was set up with plasmid pAAV-Spc512-SaCas9-muliplex-G14-G18. The melting curve, amplification curves and standard curve for the three primer pairs targeting *Sa*Cas9 can be compared at 57ºC and 60ºC on Figs. 5.2 & 5.3.

All primers showed an efficiency of ~100% and a single main peak on the melt curve, confirming primer specificity. It was decided to use primer pair "*Sa*Cas9.1" (FW 5´- CTG GAA CGG CTG AAG AA GA -3´, RV 5´- GTC GAT GTA GGT GTC GAT GAA G -3´), at 57ºC for future experiments as the efficiency of this primer pair was the closest one to 100% (efficiency of 2 = 100%) and samples showed a very neat melting curve.

57ºC

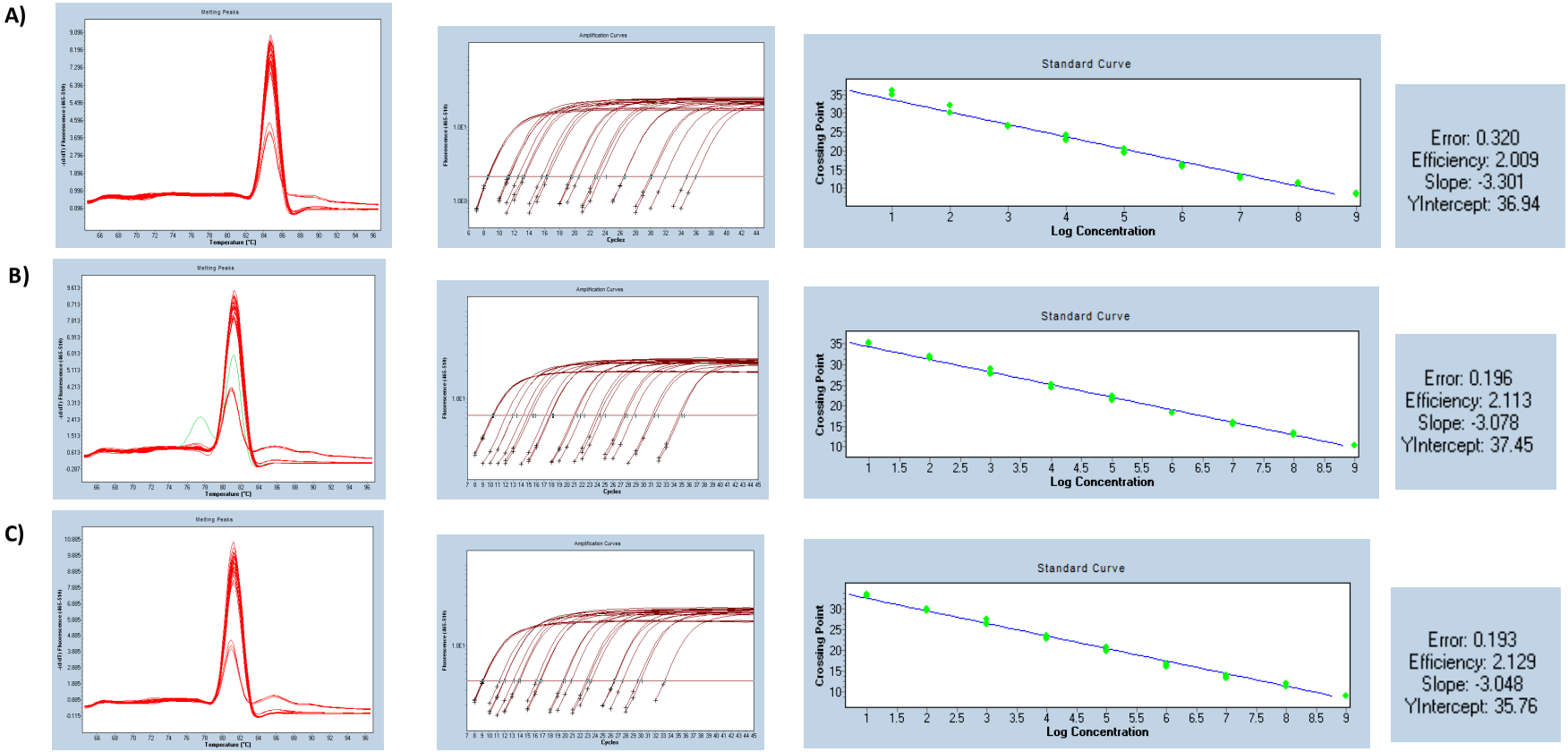


Figure 5.2. Primer pairs optimization for qPCR targeting *Sa*Cas9. Three primer pairs targeting *Sa*Cas9 sequence were designed and tested at 57ºC. Presented from left to right colum: Melting curve, amplification curve of standards and standard curve with error, efficiency, slope and Y-intersection indicated. Primer pairs: A) SaCas9.1 (FW 5´-CTGGAACGGCTGAAGAAAGA-3´, RV 5´-GTCGATGTAGGTGTCGATGAAG-3´), B) SaCas9.1 (FW 5´-CAAGTGCTATGAGGAAGCTAAGA-3´, RV 5´-GTTCACGCCGATCACTCTATAC-3´), C) SaCas9.3 (FW 5´-AACCGAGCAGGAGTACAAAG-3´RV 5´-GGAGTACAGGGTGTCGTTAATC-3´).

60ºC

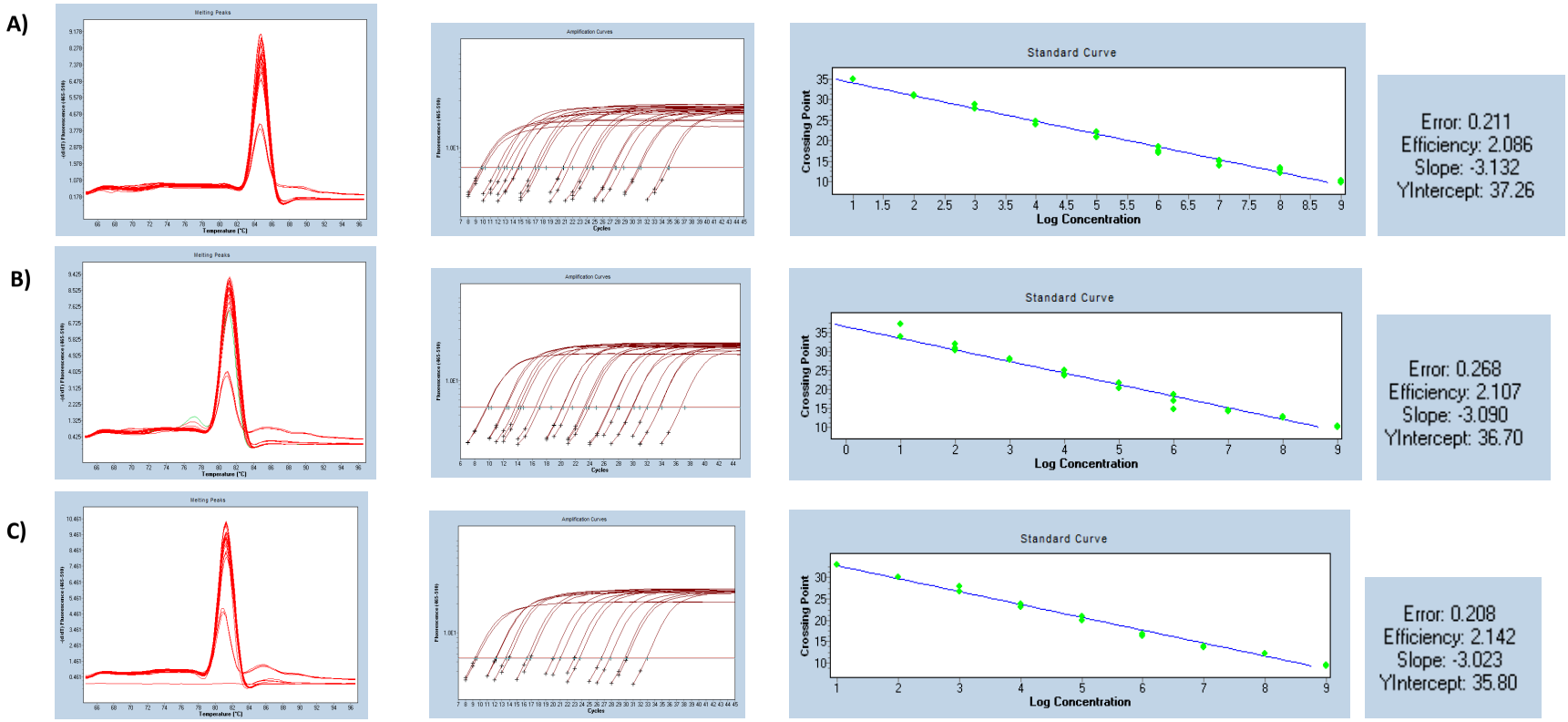


Figure 5.3. Primer pairs optimization for qPCR targeting *Sa*Cas9. Three primer pars targeting *Sa*Cas9 sequence were designed and tested at 60ºC. Presented from left to right colum: Melting curve, amplification curve of standards and standard curve with error, efficiency, slope and Y-intersection indicated. Primer pairs: A) SaCas9.1 (FW 5´-CTGGAACGGCTGAAGAAAGA-3´, RV 5´-GTCGATGTAGGTGTCGATGAAG-3´), B) SaCas9.1 (FW 5´-CAAGTGCTATGAGGAAGCTAAGA-3´, RV 5´-GTTCACGCCGATCACTCTATAC-3´), C) SaCas9.3 (FW 5´-AACCGAGCAGGAGTACAAAG-3´RV 5´-GGAGTACAGGGTGTCGTTAATC-3´).

To titre the AAV9 preps, samples from each prep were prepared as described in Materials and Methods Section 2.14.2 and qPCRs were run with primers binding the *Sa*Cas9 and with previously optimized primers targeting the GFP sequence (Forward primer: 5´- CAA GAT CCG CCA CAA CAT CG -3´and reverse primer 5´- GAC TGG GTG CTC AGG TAG TG -3´).

Melt curve, amplification curves and standard curve obtained from samples from all AAV preps with an *Sa*Cas9 can be seen on Fig. 5.4. Final titres are presented on Table 5.1.

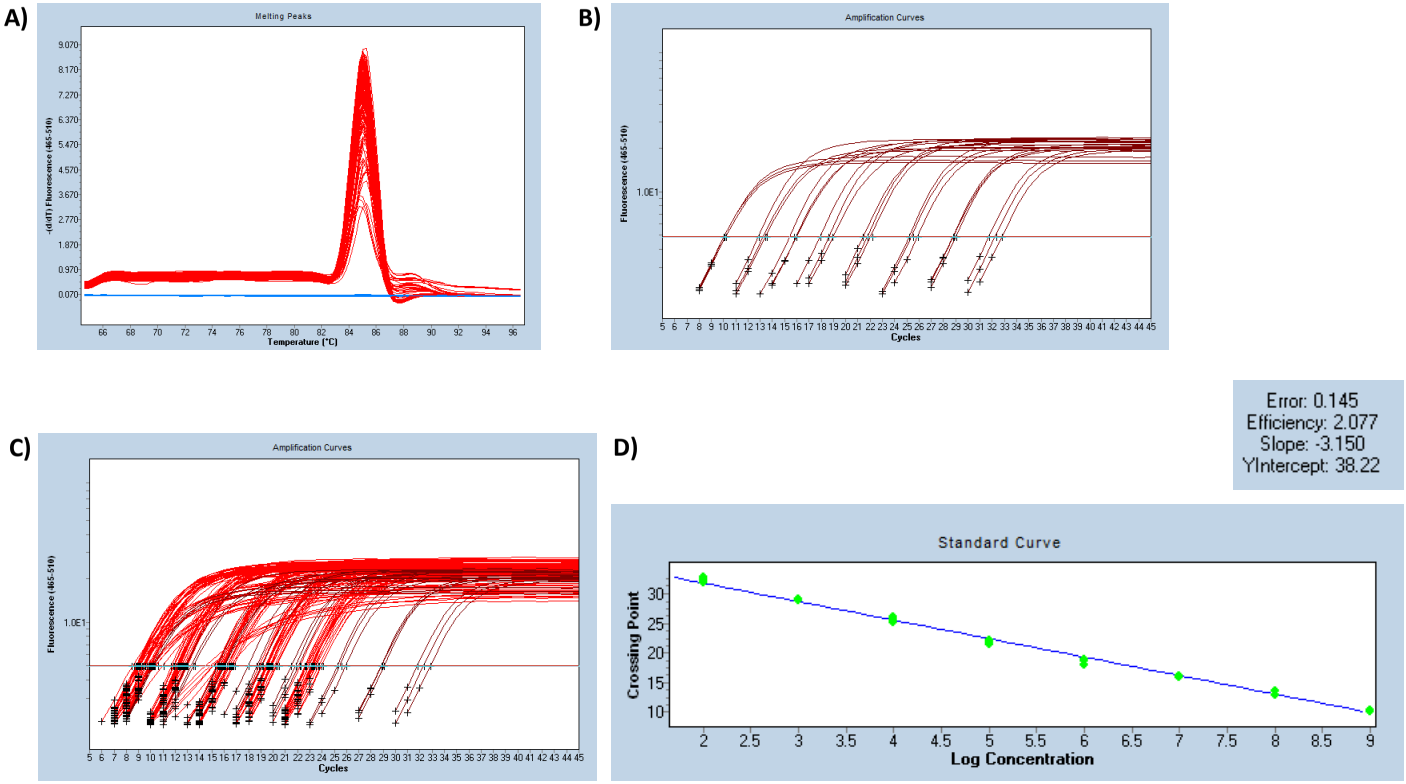


Figure 5.4. LightCycler480 Software analysis of qPCRs from AAV9 preps with *Sa*Cas9 (AAV9-Spc512-Multiplex-BbsI-BsaI (Empty), AAV9-Spc512-Multiplex-G14-G18, AAV9-Spc512-Multiplex-G14-BsaI and AAV9-Spc512-Multiplex-BbsI-G18). A) Melting curve, B) Amplification curve of stadard samples, C) Amplifications curve of standards and samples from all preps and D) Standard with error, efficiency, slope and Y-intersection indicated.

Table 5.1. AVV9 vector titre per prep obtained by qPCRs. Results presented in viral genomes per mL or µL.

AAV9 Prep	Titre (vg/mL)	Titre (vg/µL)	Obtained volume (µL)
AAV9-Spc512-GFP	1.35E+14	1.35E+11	350
AAV9-Spc512-Multiplex-BbsI-BsaI (Empty)	6.35E+13	6.35E+10	450
AAV9-Spc512-Multiplex-G14-G18	4.65E+13	4.65E+10	450
AAV9-Spc512-Multiplex-G14-BsaI	3.54E+13	3.54E+10	450
AAV9-Spc512-Multiplex-BbsI-G18	5.69E+13	5.69E+10	450

5.2. *IN-VIVO* TRANSDUCTION OF *MDX* MICE TIBIALIS ANTERIOR (TA) MUSCLES WITH AAV9 VECTORS.

5.2.1. EXPERIMENTAL DESIGN FOR IN-VIVO TRANSDUCTIONS OF MDX MICE.

A total of 24 *mdx* mice and 4 wild type (C57 black) mice were bred for this experiment. All mice used for this experiment were the same sex; female *mdx* mice were used due to mice availability.

It is known that *mdx* mice develop a progressive dystrophic muscle histopathology as they age (Chamberlain et al., 2007, Vohra et al., 2017) with an increase in muscle fibrosis (Hakim & Duan, 2012). An early intervention in young *mdx* mice could aid with prevention of muscle damage, hence it was decided to treat 2-months old *mdx* mice. An n = 4 mice per group was considered, therefore an n = 8 of TA muscles per group would be available for analysing. Mice were injected on both TA muscles with a dose of $1x10^{11}$ vp / 30 µL per TA of saline solution. Calculations to prepare injections were done for 10 TA muscles per group so there would be enough mix for all injections (presented on Table 5.2).

Table 5.2. Calculations for injections of 7 groups of female *mdx* mice, n = 4, TA muscles per group = 8, calculations for 10 TA muscles per group, Dose per TA = 10E+11 VP in 30 µL of saline solution.

Group	AAV prep	Titre (vp/ml)	Prep vol/group (µL)	Saline Vol/group (µL)	Final vol/ 10 legs (µL)
1	AAV9-Spc512-Multiplex-BbsI-BsaI (Empty)	6.35488E+13	15.7359507	284.264049	300
2	AAV9-Spc512-Multi-G14-G18	4.64521E+13	21.5275598	278.47244	300
3	AAV9-Spc512-Multiplex-G14-BsaI	3.53767E+13	28.2672194	271.732781	300
4	AAV9-Spc512-Multiplex-BbsI-G18	5.68563E+13	17.5882159	282.411784	300
5	AAV9-Spc512-Multiplex-G14-BsaI + AAV9-Spc512-Multiplex-BbsI-G18	4.61165E+13	45.8554353	254.144565	300
6	Saline	-	-	300	300
7	Wild Type	-	-	300	300

TA muscles were harvested 2 months after treatment, measurements of body weight, TA muscle weight and TA length were taken after performing muscle electrophysiology.

5.2.2. TRANSDUCED TA MUSCLES ELECTROPHYSIOLOGY ANALYSIS TO ASSESS POTENTIAL FUNCTIONALITY EFFECTS OF TREATMENTS.

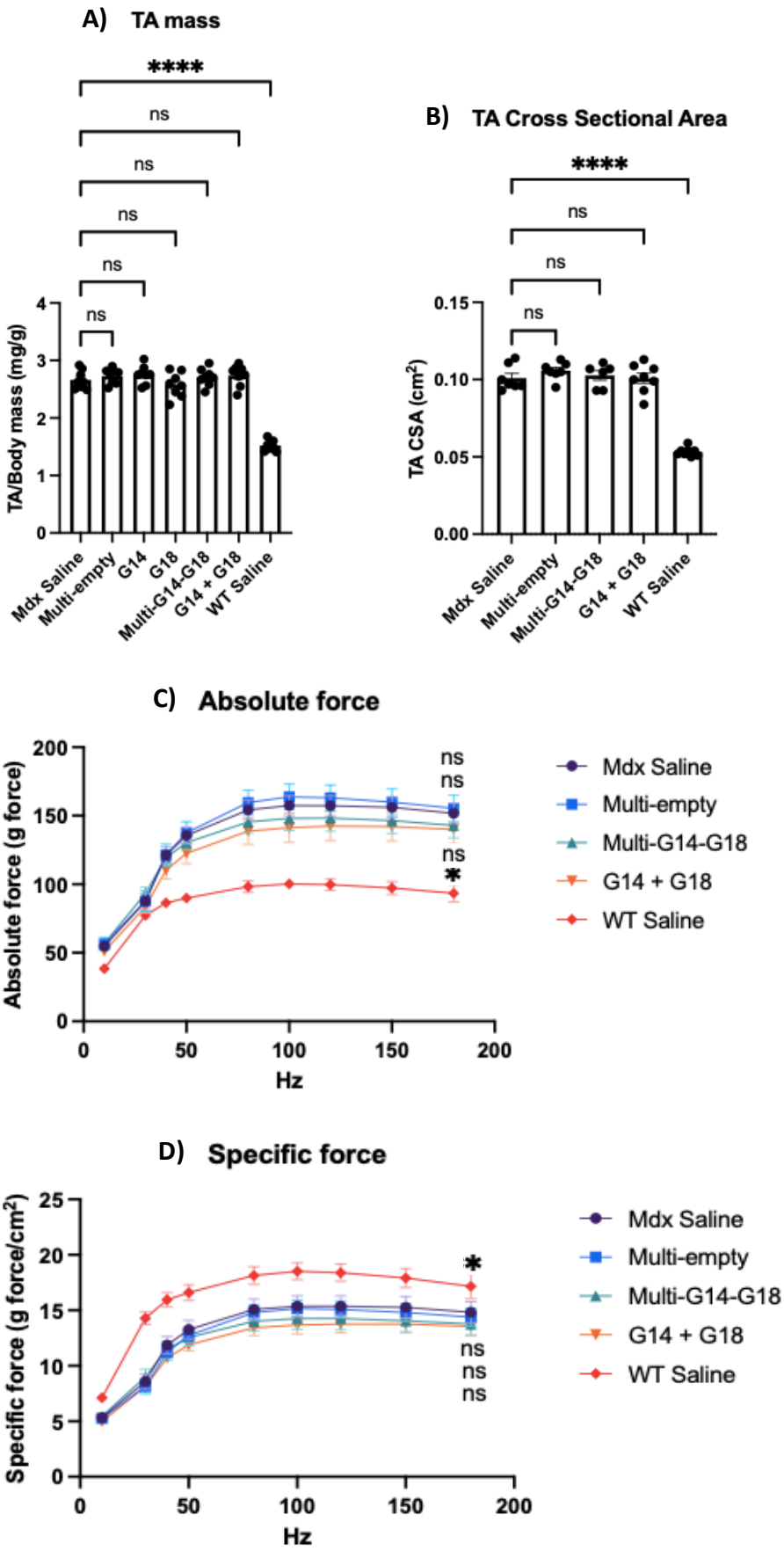
TA muscle length and weight measured after electrophysiology are reported on Table 5.3 alongside mice details. No physiology analysis was performed on groups treated with individual gRNAs as no effect nor deletion was expected from these treatments.

Table 5.3. ***In-vivo*** **experiment details. Contructs used per gorup, mice ID number, date of birth, sex, box, mark to identify (R-right ear), from = parental IDs, body weight in grams, lenght in mm of TAL (Tibialis anterior left) TAR (TA right), weight in mg of TAL and TAR, TA weight split with Body weight to obtain TA mass and harvesting date. No TA lenght measure for groups 3 & 4, as no**

CONSTRUCT	ID #	BORN	SEX	BOX	MARK	FROM	BODY WEIGHT (g)	TAL(mm)	TAR (mm)	TAL (mg)	TAR (mg)	TAL/BW	TAR/BW	HARVEST
Mdx + Saline	732	14/05/22	♀	221	-	683/693	33.5	x	12.97	85.0	83.4	2.54	2.49	29-sep
	733	14/05/22	♀	221	1R	683/693	33.8	13.28	13.90	96.7	98.7	2.86	2.92	29-sep
	741	14/05/22	♀	224	-	689/680	31.3	13.29	13.79	82.1	86.0	2.62	2.75	30-sep
	742	14/05/22	♀	224	1R	689/680	31.8	13.37	13.50	79.5	83.3	2.50	2.62	30-sep
AAV9-Spc512-Multiplex-BbsI-BsaI (Empty)	749	07/06/22	♀	227	-	647/665	32.0	13.42	x	90.1	88.3	2.82	2.76	29-sep
	750	07/06/22	♀	227	1R	647/665	31.3	13.19	13.42	88.1	81.4	2.81	2.60	29-sep
	759	11/06/22	♀	231	-	677/703	35.4	14.20	14.20	102.7	98.6	2.90	2.79	30-sep
	760	11/06/22	♀	231	1R	677/703	35.0	13.00	13.07	91.7	88.1	2.62	2.52	04-oct
AAV9-Spc512-Multiplex-G14-G18	753	11/06/22	♀	229	-	674/651	30.0	13.50	13.64	80.2	81.0	2.67	2.70	29-sep
	754	11/06/22	♀	229	1R	674/651	32.6	13.10	13.80	90.1	92.6	2.76	2.84	30-sep
	755	11/06/22	♀	229	1L	674/651	32.3	12.69	13.43	87.1	95.2	2.70	2.95	30-sep
	756	11/06/22	♀	229	2R	674/651	30.0	x	x	77.3	73.6	2.58	2.45	30-sep
AAV9-Spc512-Multiplex-G14-BsaI	765	09/06/22	♀	233	-	647/666	29.5	x	x	84.7	75.4	2.87	2.56	06-oct
	766	09/06/22	♀	233	1R	647/666	32.5	x	x	98.3	89.7	3.02	2.76	06-oct
	767	09/06/22	♀	233	1L	647/666	31.6	x	x	88.0	86.9	2.78	2.75	06-oct
	768	09/06/22	♀	223	2R	647/666	26.6	x	x	67.0	73.1	2.52	2.75	06-oct
AAV9-Spc512-Multiplex-BbsI-G18	771	09/06/22	♀	235	-	667/643	34.3	x	x	92.4	76.6	2.69	2.23	06-oct
	772	09/06/22	♀	235	1R	667/643	30.7	x	x	86.8	87.5	2.83	2.85	06-oct
	773	09/06/22	♀	235	1L	667/643	34.4	x	x	88.2	82.0	2.56	2.38	06-oct
	774	09/06/22	♀	235	2R	667/643	34.1	x	x	83.7	88.8	2.45	2.60	06-oct
AAV9-Spc512Multiplex-G14-BsaI + AAV9-Spc512-Multiplex-BbsI-G18	785	07/06/22	♀	239	-	689/701	32.8	14.55	14.05	78.6	83.8	2.40	2.55	05-oct
	786	07/06/22	♀	239	1R	689/701	30.8	13.07	13.88	83.9	88.4	2.72	2.87	05-oct
	787	07/06/22	♀	239	1L	689/701	31.4	13.60	13.50	88.7	92.6	2.82	2.95	05-oct
	799	09/06/22	♀	243	2R	690/702	33.4	13.17	13.09	92.0	95.1	2.75	2.85	04-oct
WT (C57BL/10) + Saline	529	19/03/22	♀	178	-	508/493	28.6	12.38	12.43	42.7	40.7	1.49	1.42	06-oct
	530	19/03/22	♀	178	1R	508/493	29.3	11.89	11.42	41.2	43.2	1.41	1.47	06-oct
	532	19/03/22	♀	180	-	486/483	25.9	12.16	12.44	40.6	39.7	1.57	1.53	06-oct
	533	19/03/22	♀	180	1R	486/483	27.0	12.47	13.67	43.3	45.4	1.60	1.68	06-oct

Muscle electrophysiology data analysis is presented on Fig. 5.5. No significant effects were observed with any of the treatments. There was no difference in TA mass or cross-sectional area, no improvements in absolute force, specific force nor eccentric force when compared to *mdx* mice treated with saline solution.

To draw any conclusions further analysis was needed. Therefore, each TA muscle was cut in half, one half was sectioned for immunohistochemistry analysis, DNA & RNA extractions and the other half was used for protein extraction, so protein expression of Del19-55 dystrophin could be analysed by Western Blots.


A) TA mass
****
ns
ns
ns
ns
ns
TA/Body mass (mg/g)
0
1
2
3
4
Mdx Saline
Multi-empty
G14
G18
Multi-G14-G18
G14 + G18
WT Saline
B) TA Cross Sectional Area
****
ns
ns
ns
TA CSA (cm²)
0.00
0.05
0.10
0.15
Mdx Saline
Multi-empty
Multi-G14-G18
G14 + G18
WT Saline
C) Absolute force
Absolute force (g force)
0
50
100
150
200
Hz
ns
ns
ns
*
Mdx Saline
Multi-empty
Multi-G14-G18
G14 + G18
WT Saline
D) Specific force
Specific force (g force/cm²)
0
5
10
15
20
25
Hz
*
ns
ns
ns
Mdx Saline
Multi-empty
Multi-G14-G18
G14 + G18
WT Saline

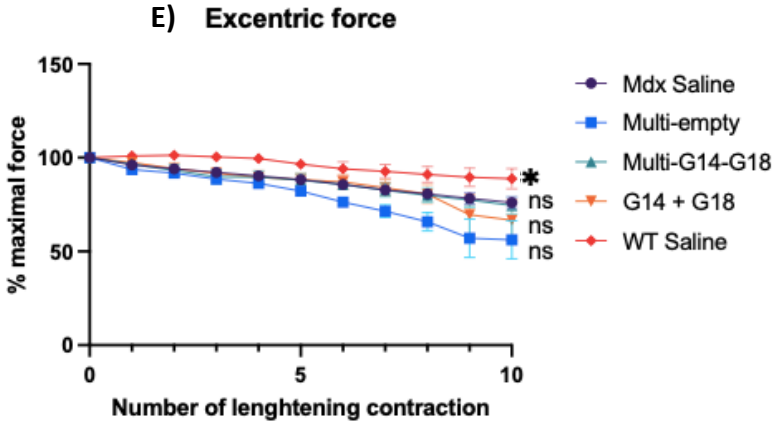


Figure 5.5. Analysis from data obtained from muscle electrophysiology of *mdx* mice harvested 2 months after treatment with AAV9 vectors. A) TA mass was obtained by dividing TA over body weight (mg/g). B) TA Cross Sectional area was calculated: CSA (mm^2) = TA weight / (TA length x 0.6 x 1.067), where 1.067 (mg/mm^3) is the density of mammalian muscle and 0.6 is the optimum muscle length/fibre length ratio for TA muscle. C) Absolute force was measured in a 9-protocol sequence with different frequency of stimulation at 10, 30, 40, 50, 80, 100, 120, 150 and 180 Hz. The entire sequence lasts ~ 7 minutes, nerve was maintained moist and at optimal tension (~1.232 g). D) Specific force (mN/mm^2) calculated as maximal force/CSA. E) Eccentric force calculated as percentage of force drop in Eccentric contraction (ECC) = (ECC_n x 100)/ECC_1. Per group: n = 8 biological repeats. Statistical analysis by mean comparison against *mdx* saline samples with a One-way ANOVA (95% confidence interval and p-value<0.05), followed by a Dunnett´s test. For A) and B), ****adjusted p-value<0.0001. For C) and D), means compared at 180 Hz. For E), means compared at the 10th lengthening contraction. *adjusted p-value<0.02.

5.2.3. Analysis of DNA extracted from transduced TA muscles.

5.2.3.1. Assessment of individual gRNA efficiency.

DNA was extracted from intersections from half of sectioned TA muscles. An end-point PCR with previously designed primer pairs (used for individual gRNA assessment in Section 4.2.2) flanking G14 cut site (5´- CCCAGGCAAACATGATACAATTAG -3´ and reverse 5´- AGCATGAGAGCAAAGGTGAG -3) and G18 cut site (5´- GCTAATCAAATCTGTGCATGGT -3´and reverse 5´- CTGGTCCATGCCTAACCATAT -3´), which produce a 1043 bp and a 548 bp product respectively were used. A single PCR product can be observed for all samples on Fig. 5.6. PCR reactions were cleaned and sent for Sanger sequencing with appropriate primers. Guide RNA efficiency was evaluated using TIDE analysis (as described in Materials & Methods section 2.9) and results can be observed on Fig. 5.7. Once aberrant samples were eliminated (Fig. 5.7.B), G14 showed an editing efficiency of ∼5% and G18 of ∼12%. A lot of background noise can be observed on the control sequence of G18 on representative Figure 5.9. As discussed previously, an AT rich region downstream of the cut site might be affecting the quality of the sequencing traces and interfering with the TIDE analysis.

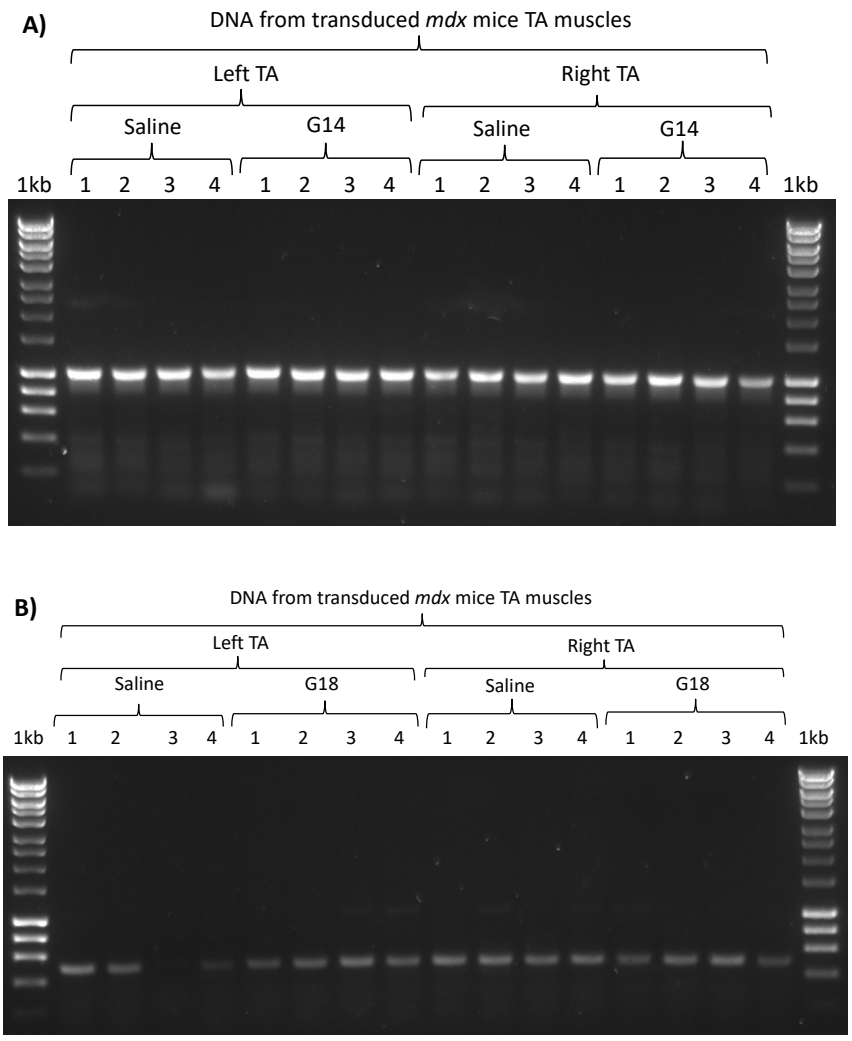


Figure 5.6. Gel images of PCR products from DNA extracted from *mdx* mice TA muscles transduced with individual gRNAs. A) PCR products of expected size: 1073 bp from samples treated with AAV9-sPC512-SaCas9-G14. PCR primers: forward 5´-CCCAGGCAAACATGATACAATTAG -3´ and reverse 5´- AGCATGAGAGCAAAGGTGAG -3´. B) PCR products of expected size (548 bp) from samples treated with AAV9-Spc512-SaCas9-G18. PCR primers: forward 5´- GCTAATCAAATCTGTGCATGGT -3´and reverse 5´-CTGGTCCATGCCTAACCATAT -3´. Both gels were 1% agarose (w/v) with 0.5X SYBR Safe in 1X TAE Buffer. Hyperladder I was used (1kb).

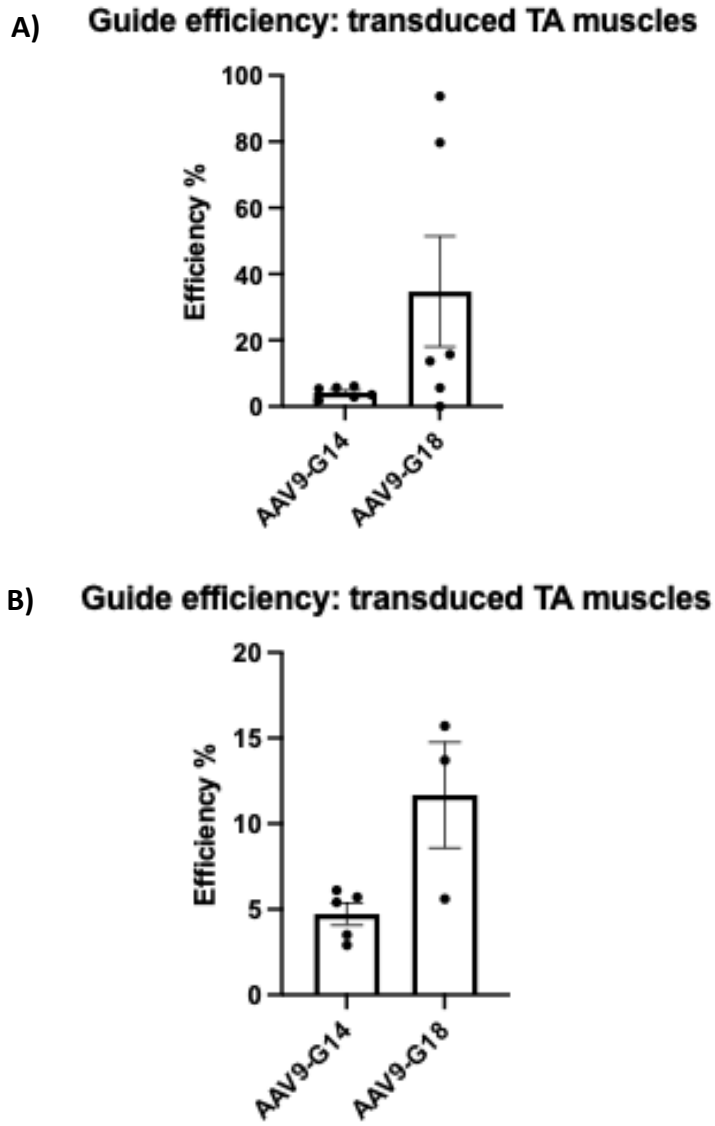


Figure 5.7. Graphical summary of *Sa*Cas9 gRNA cutting efficiency based on TIDE Analysis, bar charts show: A) Transduced TA muscles from *mdx* mice, n = 6 biological repeats. B) Transduced TA muscles from *mdx* mice, with aberrant samples eliminated (n = 5 biological repetas for G14 and n = 3 biological repeats for G18). DNA extracted from intersections of TA muscle. Appropriate PCR primers were designed targeting the sequence flanking the editing target site. PCRs were performed for each sample and ran on a 1% (w/v) agarose gel, PCR products were extracted, cleaned and sequenced (by Eurofins) with appropriate primers. Sequence traces were then analysed by TIDE Analysis. TIDE web tool algorithm reconstructs the spectrum of indels from two sequencing traces per gRNA (an edited vs. untreated trace). The output reports identity and frequency of detected indels, as a percentage, generated in a pool of cells (Brinkman et al., 2014). Data plotted on Prism9 Software. Error bars represent standard error of the mean.

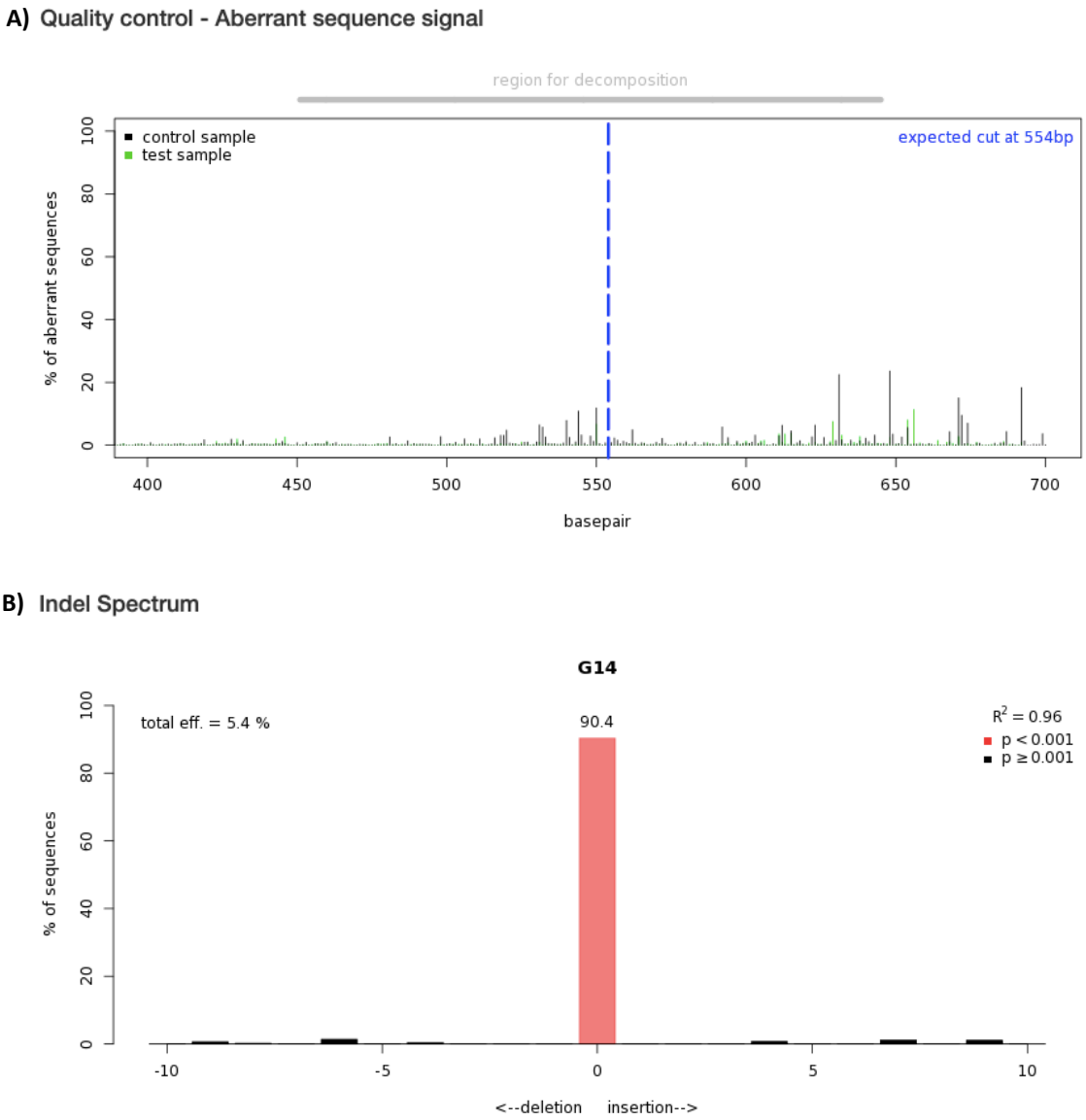


Figure 5.8. Representative outputs from TIDE Analysis Web Tool. Analysis of Guide 14 presented. A) Decomposition trace, aberrant sequence signal (green) compared to control trace (black). Dotted blue line indicates cut site. B) Bar chart indicating indel spectrum output. X-axis indicates small deletions of up to 10 base pairs on a negative scale (-10 to 0) and insertions on a positive scale (0 to 10). In this example, the red bar indicates 90.4% of traces had 0 deletions. Total efficiency of 5.4% from Guide 18, indicated at the top left of the graph. Numbers at the top right corner: coefficient of determination (R^2), to evaluate model accuracy (values from 0 to 1). P-values indicate significance cutoff, set up at $p<0.001$. Significant outputs in red, non-significant ($p\geq0.001$) in black.

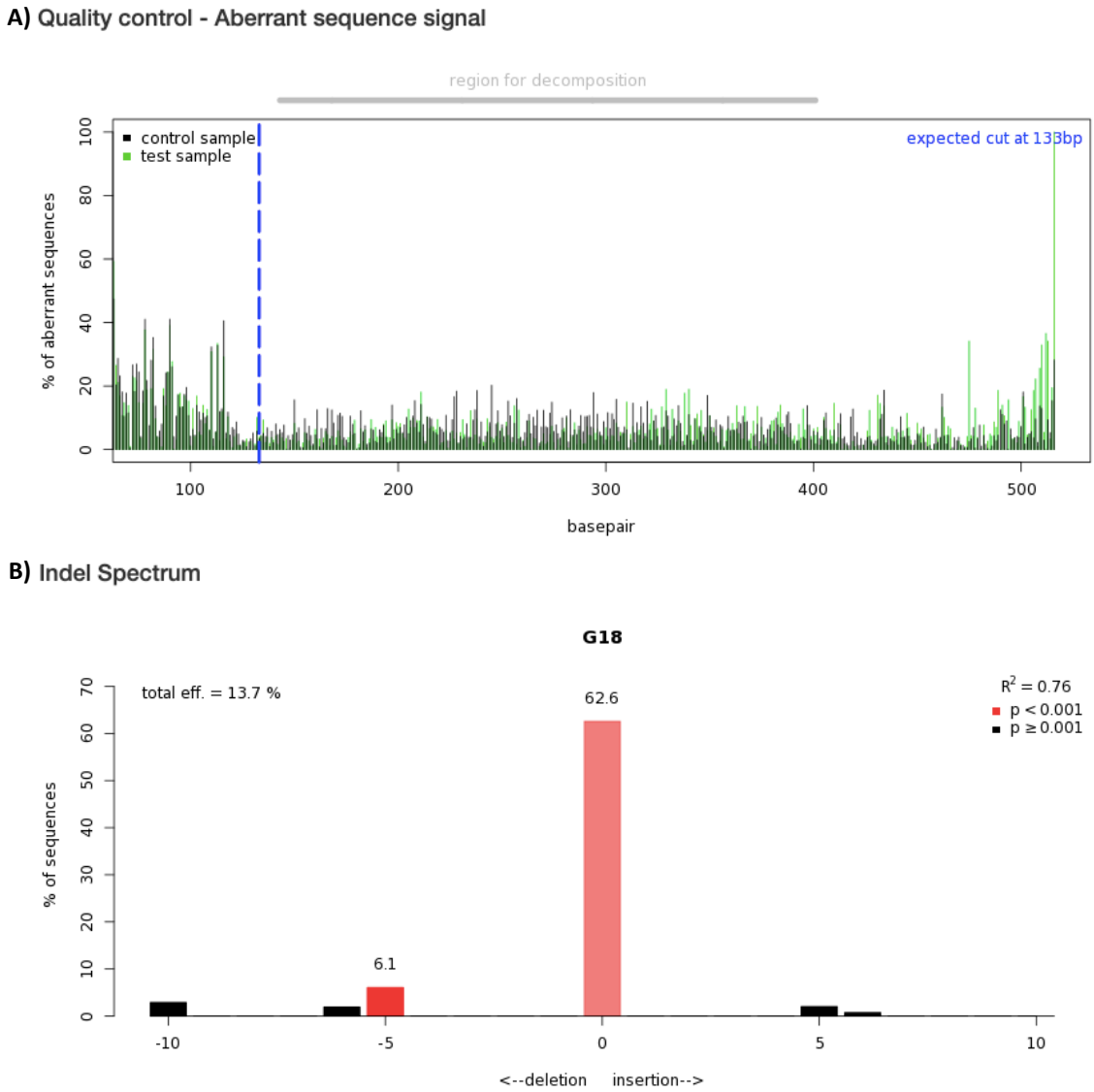


Figure 5.9. Representative outputs from TIDE Analysis Web Tool. Analysis of Guide 18 presented. A) Decomposition trace, aberrant sequence signal (green) compared to control trace (black). Dotted blue line indicates cut site. B) Bar chart indicating indel spectrum output. X-axis indicates small deletions of up to 10 base pairs on a negative scale (-10 to 0) and insertions on a positive scale (0 to 10). In this example, the red bars indicate 62.6% of traces had 0 deletions or insertions, 6.1% had -5 deleted bp and lower percentages of edited populations (black bars) indicate some -6 and -10 bp deletions and some +5 and +6 insertions. These percentages add up to a total efficiency of 13.7% from Guide 18, indicated at the top left of the graph. Numbers at the top right: coefficient of determination (R^2), to evaluate model accuracy (values from 0 to 1). Low R^2 can be due to poor sequence quality or non-optimal setting. P-values indicate significance cutoff, set up at $p<0.001$. Significant outputs in red, non-significant ($p \geq 0.001$) in black.

5.2.3.2. *Assessment of a deletion between introns 18 and 55 by PCR in DNA obtained from transduced TA muscles.*

From the same genomic DNA samples, previously extracted from intersections of half TA muscles, an end-point PCR to detect a deletion was performed with previously designed primers targeting intronic regions (described and used to detect deletion *in-vitro* in Section 4.4.2.1). Primers were designed to express a 970 bp product if there is a deletion, if there is no deletion PCR product would be too large to be amplified. None of the samples showed a clear unique product. There seemed to be multiple faint bands on all samples, including from wild type and saline-injected *mdx* mice. Two products close to the expected size were observed on one of the multiplex samples and one of the co-transduced samples (Fig. 5.19). These bands were extracted and sent for sequencing. However, it was not possible to obtain a clean trace from these samples and no further conclusions could be made at this point. There was a possibility that there was a deletion, but the levels were too low to be detected by end-point PCR. Furthermore, it was not possible to quantify deletion by genomic qPCR because it was not possible to synthetise a g-block containing the sequence of the *de novo* junction between introns 18 and 55 due to sequence complexity, including a low GC content, repeated GTTGT sequences and TGTTGTTGTT sequences constituting approximately 17% of the overall sequence (assessed with IDT online tool for g-block design: https://eu.idtdna.com/site/order/gblockentry). Therefore, it was decided to focus on attempting to detect and quantify the deletion by RT-qPCR from RNA samples.

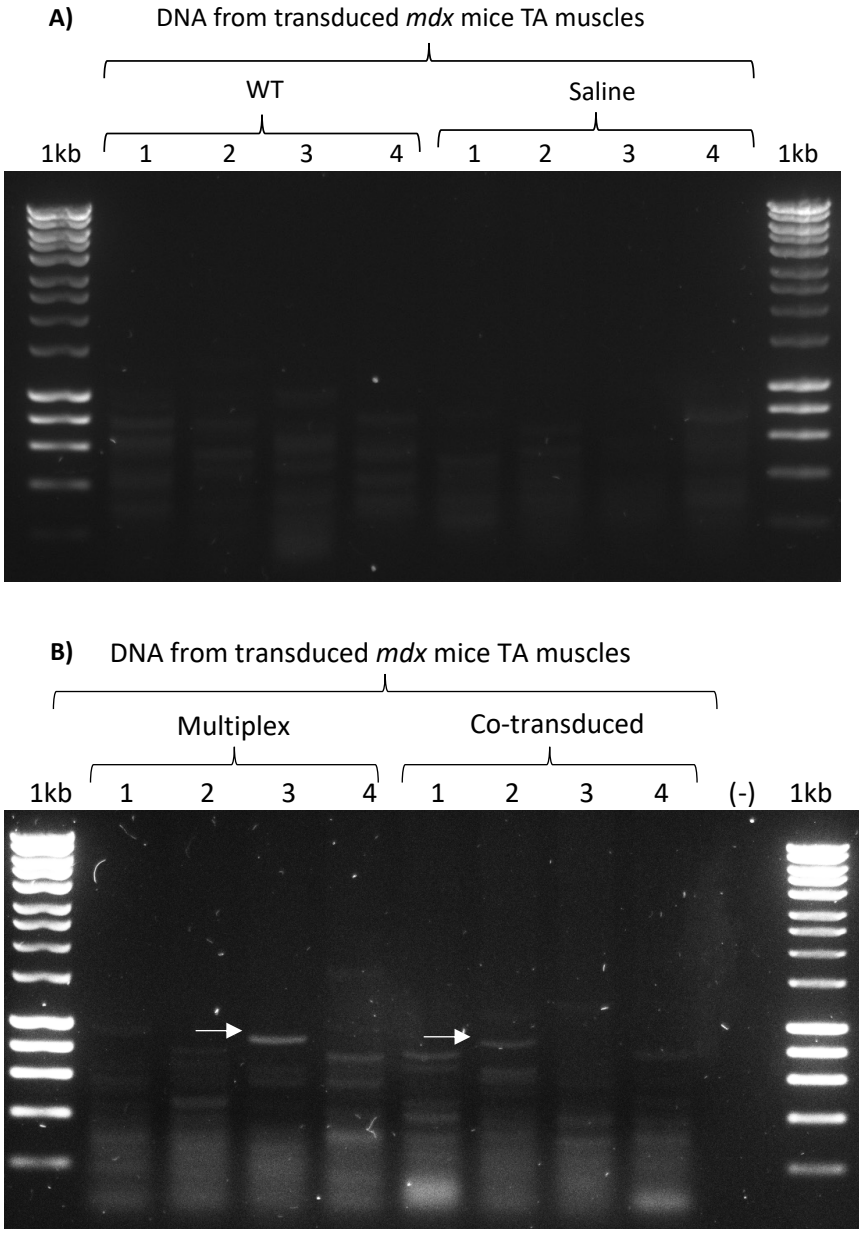


Figure 5.10. Gel images of PCR products from DNA samples extracted from TA muscles of treated *mdx* mice. A) Wild type and *mdx* mice samples treated with saline solution. B) "Multiplex" samples were treated with AAV9-Spc512-SaCas9-multiplex-G14-G18 and "Co-transduced" samples with AAV9-Spc512-SaCas9-G14 and AAV9-Spc512-SaCas9-G18. If Exons 19 to 55 were deleted, a PCR product of 970 bp was expected. Gel was 1% agarose (w/v) with 0.5X SYBR Safe in 1X TAE Buffer. Hyperladder I was used. White arrows indicate products with potential expected size. Indicated bands were extracted and sent for sequencing.

5.2.4. Assessment of SaCas9 expression and deletion of exons 19 to 55 in RNA from transduced TA muscles by RT-qPCR.

To assess expression of the transgene delivered by AAV9 vectors and quantify deletion of exons 19 to 55, two RT-qPCRs were performed. The first one with previously optimized primers binding to the *Sa*Cas9 and the second one with two primer pairs: first primer pair targeting *Dmd* Exons 6-7 and the second primer pair targeting *Dmd* Exons 20-21, which would be deleted if our *de novo* intron junction was created.

RNA was extracted from intersections of TA muscles and cDNA was obtained by reverse transcription with a QuantiTect reverse transcription kit from QIAGEN. Standard amplification curves were prepared by serial dilutions from 1E+10 to 1E+1 copy numbers of a plasmid expressing an *Sa*Cas9 (pAAV-Spc515-SaCas9-multiplex-G14-G18) or g-blocks expressing cDNA of exons 6-7 and exons 20-21 respectively. Samples were prepared as described in Materials and Methods Section 2.14.3. Plates were processed on a LightCycler480 Instrument II from Roche and data was analysed on the LightCycler480 Software to obtain the melting curve, amplification curves, standard curve and its efficiency, Cp values and concentration of each sample.

5.2.4.1. RT-qPCR TO DETECT SaCas9 EXPRESSION.

*Sa*Cas9 expression from AAV9 vectors normalized against reference gene Rplp0 can be seen on Fig. 5.11. From this experiment it can be concluded that the AAV9 vectors were expressing the constructs containing an *Sa*Cas9. Levels of expression seem to vary between samples and between constructs, however there was no significant difference between treated groups.

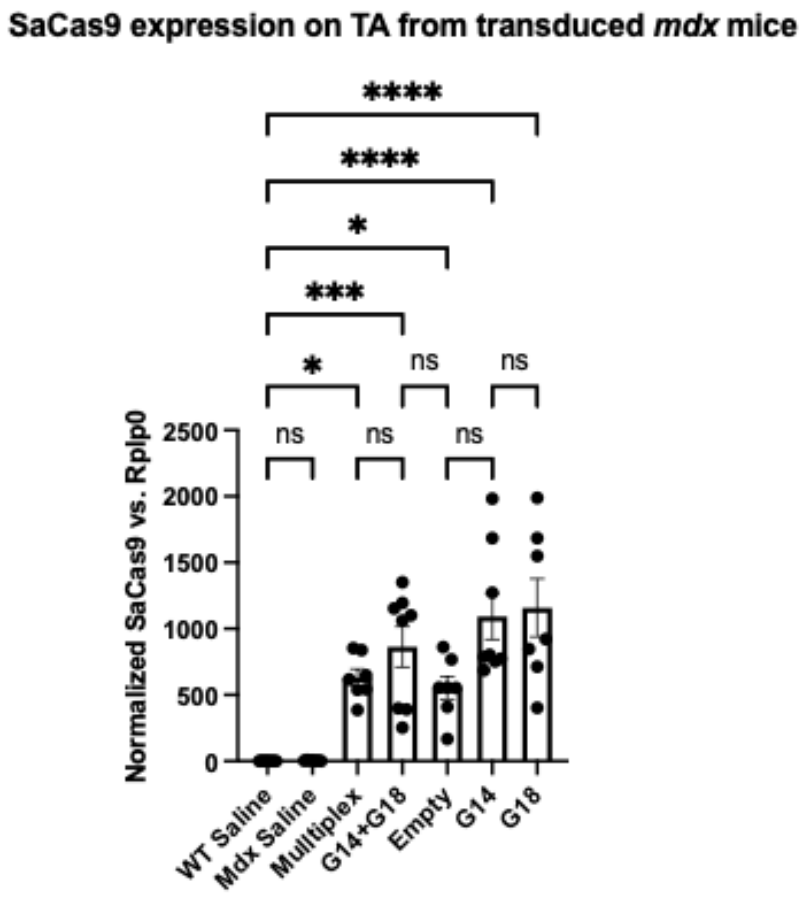


Figure 5.11. Normalized *Sa*Cas9 expression against reference gene Rplp0, from transduced TA muscles from *mdx* mice. N = 8 biological repeats per group. From left to right: wild type mice injected with saline solution, *mdx* mice injected with saline solution, *mdx* mice treated with multiplex construct (AAV9-Spc512-SaCas9-multiplex-G14-G14), co-transduced with G14 and G18 (AAV9-Spc512-SaCas9-G14-BsaI and AAV9-Spc512-SaCas9-BbsI-G18), transduced with empty construct (AAV9-Spc512-SaCas9-BbsI-BsaI) and individual gRNA constructs. *Sa*Cas9 expression was found significant by mean comparison against wild type samples treated with saline solution with a One-way ANOVA Analysis (95% confidence interval, p-value<0.05), followed by post-hoc Holm-Šídák´s multiple comparisons test. Adjusted p-values for: WT saline vs. multiplex, p-value = 0.0181 (*), WT saline vs. empty, p-value = 0.0490 (*), WT saline vs. G14+G18, p-value = 0.0002 (***) and WT saline vs G14 and G18, p-value<0.0001 (****). Non-significance = ns. Error bars represent standard error of the mean.

5.2.4.2. Assessment of exons 19 to 55 deletion on RNA from transduced TA muscles by RT-qPCR.

To assess deletion of exons 19 to 55, the qPCRs previously described (targeting *Rplp0* reference gene, *Dmd* exons 6-7 and *Dmd* exons 20-21) were performed. Absolute quantification was performed to detect if there was a decrease in exons 20-21 expression after G14/G18 treatments.

Normalised copy numbers per reaction were calculated as detailed in Figure 5.12. Results were plotted and analysed on Prism9 Software (Figure 5.12).

There was no significant change in expression of exons 6-7 nor exons 20-21 when analysed by mean comparison against "*mdx* saline" samples with a two-way ANOVA Analysis and a 95% confidence interval (p-value<0.05). From this it can be concluded that there were no detectable levels of deletion of exons 20-21 and therefore of exons 19 to 55, when analysing cDNA expression by RT-qPCRs.

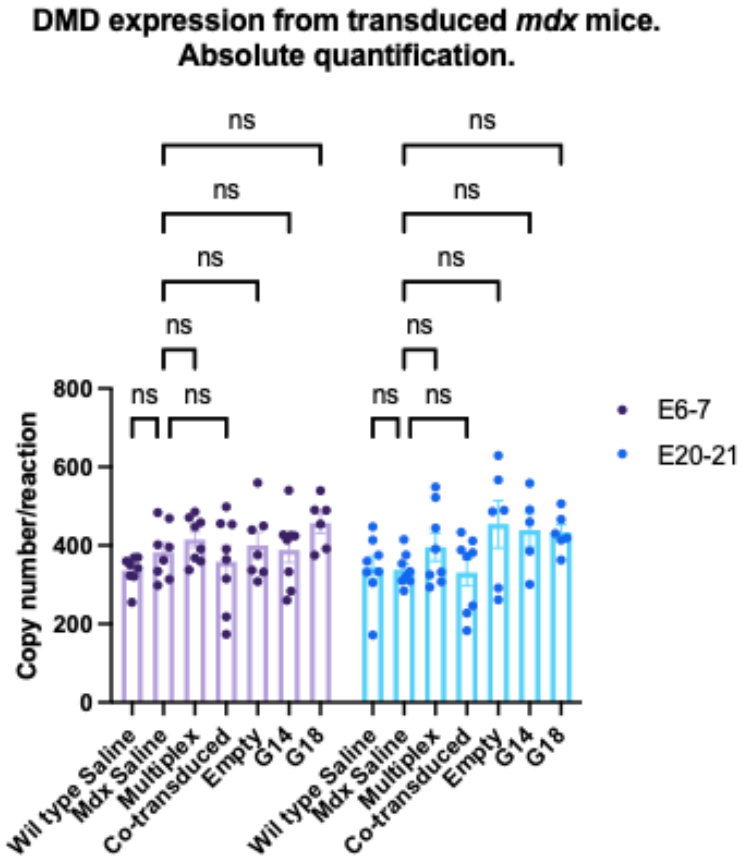


Figure 5.12. Absolute quantification of *Dmd* exons 6-7 and exons 20-21 expression. RNA extracted from intersections of TA muscles (n = 8 biological repeats); cDNA obtained by reverse transcription PCR with a QuantiTect reverse transcription kit from QIAGEN. Standard curves prepared by serial dilutions from 1E+10 to 1E+1 copy numbers of g-blocks expressing cDNA of exons 6-7 and exons 20-21 from *Dmd* mouse gene and *Rplp0* as a reference gene. Master mix of SYBR green (FastStart Universal SYBR Green Master mix 2X with FastStart Taq DNA Polymerase, Reaction Buffer, Nucleotides (dATP, dCTP, dGTP, dUTP), SYBR Green I and a reference dye from Roche) prepared to 1X for each reaction with 400 nM of each primer (forward and reverse). Then, 6 µL of mix and 4 µL of each sample, including standard curve samples, were loaded per well on a 96-well plate by triplicates, plates were processed on a LightCycler480 Instrument II from Roche and data was analysed on the LightCycler480 Software to obtain concentration of each sample, calculated by the Software based on standard curve from serial dilutions. To obtain normalised copy numbers per reaction: concentrations of samples were averaged, geometric mean of averaged concentrations was calculated for reference gene *Rplp0*, normalisation factor for each sample was obtained by dividing average *Rplp0* concentration by *Rplp0* geometric mean. Copy numbers per reaction were obtained for samples by dividing average gene of interest expression by normalisation factor. Results were graphed and analysed on Prism9 Software. There was non-significant change in expression of exons 6-7 nor exons 20-21 when analysed by mean comparison against "*mdx* saline" samples with a two-way ANOVA Analysis, followed by a Dunnett´s test (p-value<0.05).

5.2.5. *Assessment of Del19-55 dystrophin protein expression after AAV9 transduction of mdx mice.*

To assess potential protein expression of the truncated Del19-55 dystrophin after treatment of *mdx* mice, immunohistochemistry was performed, and dystrophin positive fibres were counted. Then, to further confirm results, protein was extracted from half TA muscles and analysed by Western Blotting.

5.2.5.1. *Immunohistochemistry & dystrophin positive fibre count.*

Sections of wild type and *mdx* mice TA muscles injected with saline solution were analysed as controls alongside with sections of *mdx* mice treated with the multiplex *Sa*Cas9 system (AAV9-Spc512-SaCas9-multiplex-G14-G18) and co-transduced with individual gRNAs (AAV9-Spc512-SaCas9-G14-BsaI and AAV9-Spc512-SaCas9-BbsI-G18). 10 µm sections were fixed and stained with Manex1011C (mouse monoclonal at 1:50), alpha laminin (rabbit polyclonal at 1:400) and DAPI (1:1000) to stain dystrophin, laminin and central nuclei respectively. Then, secondary antibodies anti-mouse-568 and anti-rabbit-488 were added at 1:200 and washed accordingly. Slides with sections were mounted on Mowiol and 6 fields per section were imaged with a Zeiss fluorescence microscope. Representative field images can be seen on Fig. 7.15.

Total fibres were counted using the FIJI Software and “MuscleJ plugin” and dystrophin positive fibres were counted manually on the FIJI Software. The percentage of dystrophin positive fibres was calculated and is presented on Fig. 5.13.

There was a significant difference (p-value=0.0029) between *mdx* samples treated with saline solution vs. treated with multiplex gRNAs, with 0.73% and 1.45% dystrophin positive fibres respectively. There was no significant difference between *mdx* samples treated with saline solution vs. co-transduced samples showing only 0.85% dystrophin positive fibres. To assess if these levels of dystrophin positive fibres were enough to express detectable levels of dystrophin, samples were further analysed by Western Blotting.

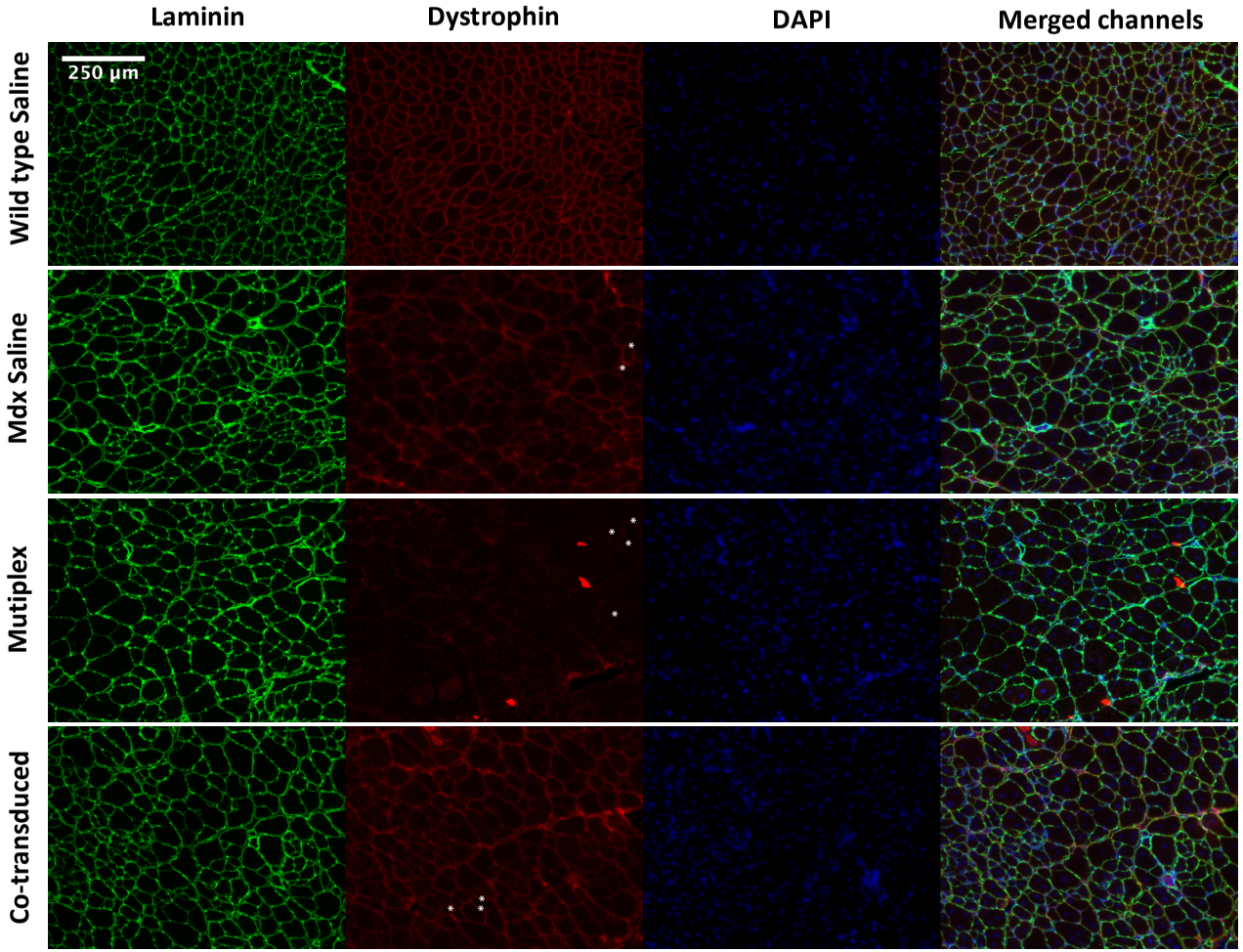


Figure 5.13. Representative immunohistochemistry field images from TA sections of treated *mdx* mice and controls (wild type and *mdx* mice injected with saline solution). 10 µm sections were fixed and stained with alpha laminin (1:400, green) for laminin, Manex1011C (1:50, red) for dystrophin and counterstained with DAPI (1:1000, blue) for central nuclei. Dystrophin positive fibres are indicated with a white * on field stained for dystrophin. Fields imaged and acquired with a Zeiss fluorescence microscope (Axio Vision D1 with AxioCam MRm) and Software ZEN 2012.

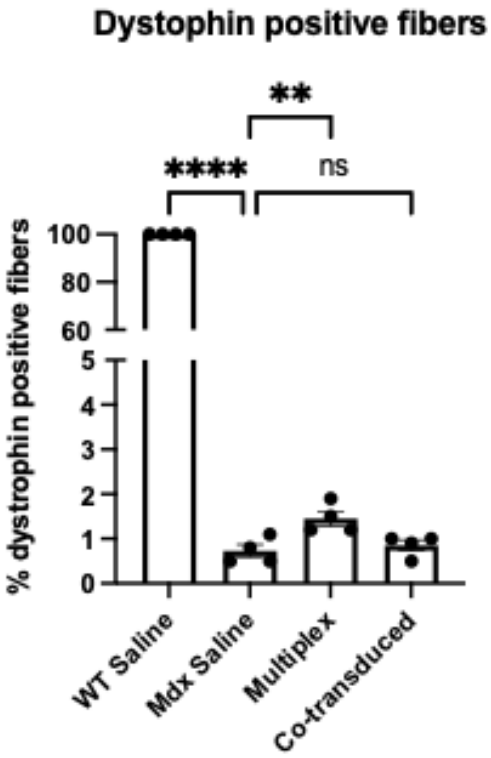


Figure 5.14. Percentage of dystrophin positive fibres in TA muscles. Samples analysed from wild type mice and *mdx* mice treated with saline solution, *mdx* mice transduced with AAV9-Spc512-SaCas9-multiplex-G14-G18 (Multiplex) and co-transduced with both gRNA constructs (G14 and G18) (Co-transduced) with an n = 4 biological repeats. Differences in positive fibres from multiplex samples were found significant by mean comparison against *mdx* saline samples, and no significance difference was found in co-transduced samples. Analysis done in Prism9 by a One-way ANOVA, with a 95% confidence interval (p-value<0.05), followed by a Dunnett´s test. Adjusted p-values for: WT salines vs. Mdx Saline, p-value<0.0001 (****); Mdx Saline vs. Multiplex, p-value=0.0039 (**) and Mdx Saline vs. Co-transduced, p-value=0.8210 (ns=non-significant). Error bars represent standard error of the mean.

5.2.5.2. Assessment of Del19-55 dystrophin expression in transduced TA muscles by Western blot.

Protein was extracted from one half of each treated TA muscle, 30 µL of protein lysate was loaded per well on 3-8% Tris-Acetate gel alongside with HiMark pre-stained ladder from ThermoFisher. Membrane was processed with Manex1011C (1:100) for dystrophin and alpha-tubulin (1:10,000) as loading control. Del19-55 dystrophin can be seen expressed from control sample (cells transfected with pAAV-CMV-hDys-Del19-55-GFP) at 224 kilodaltons, however none of the treated samples showed the expected band. Wild type sample expressed full length dystrophin as expected.

Based on these results, it can be concluded that no Del19-55 dystrophin was detected in any of the treated samples. This could mean that the protein levels were too low to detect by Western Blot or that there was not enough editing to express detectable levels of the Del19-55 truncated form of dystrophin.

To furthers assess potential causes of the lack of editing, it was decided to test AAV9 vectors *in-vitro*.

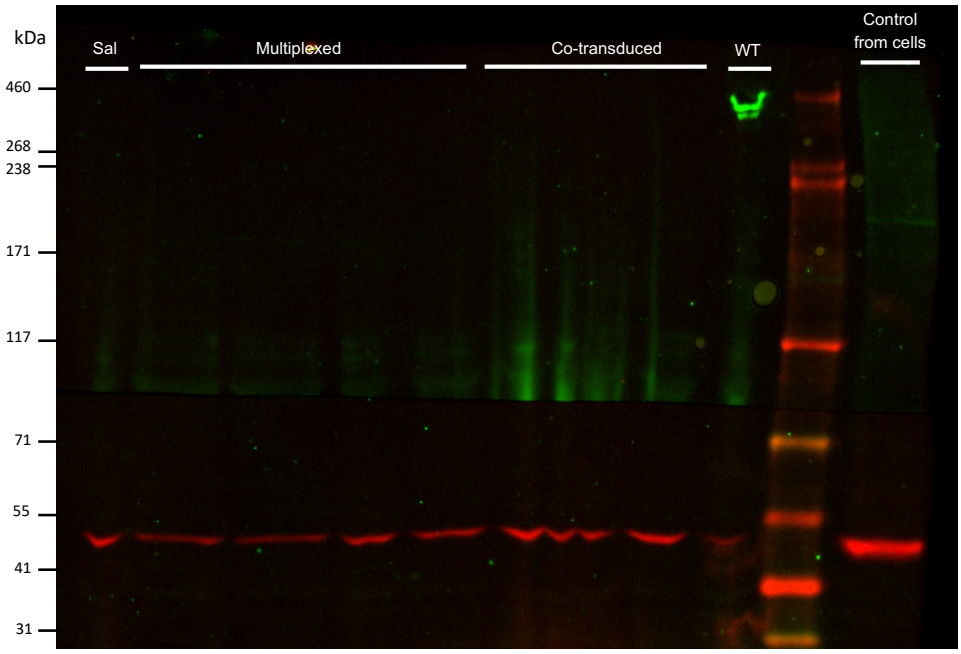


Figure 5.15. Western Blot to detect dystrophin from protein extracted from TA samples from treated *mdx* mice. From left to right, samples extracted from TA muscles from *mdx* injected with saline solution, treated with multiplex construct and co-transduces with both gRNA constructs. Wild type injected with saline expressing full length dystrophin, HiMark pe-stained ladder from ThermoFisher and control from protein sample extracted from transfected HEK293T cells with pAAV-Spc512-hDys-Del19-55-GFP. 30 µg of protein lysate per well were loaded on a 3-8 Tris-Acetate gel and analysed with antibodies: Manex1011C (1:100, green) for dystrophin and alpha-tubulin (1:10,000, red) as loading control.

5.3. *IN-VITRO* ASSESSMENT OF AAV9 VECTORS BY REVERSE TRANSDUCTION OF C2C12 AND H2KB-*MDX* CELLS.

To assess functionality of packaged AAV9 vectors, cells were transduced with the same batches of vectors used for the *in-vivo* experiments. Two cell lines were transduced, C2C12 (mouse myoblasts, ATCC, CRL-1772) and H2KB-*mdx*, a dystrophin deficient smooth muscle cell line derived from the production of a transgenic mouse model by crossing by the H-2k^{b}-tsA58 background and *mdx* mice (Morgan et al., 1994). Since C2C12 cells express full length dystrophin, if a deletion occurs (<100% efficiency), cells would express full length and Del19-55 dystrophin. On the other hand, since H2KB-*mdx* cells do not express dystrophin (due to a nonsense mutation in exon 23 that stops dystrophin expression) if there is a deletion, dystrophin expression would be recovered, and only Del19-55 dystrophin would be detected.

5.3.1. OPTIMIZATION OF C2C12 CELL DENSITY FOR REVERSE TRANSDUCTION & DIFFERENTIATION INTO MYOTUBES.

To select the optimal cell density for a transduction protocol, four different cell densities of C2C12 cells were seeded on 6-well ECM (extra cellular matrix) coated plates and reverse transduced with AAV9-Spc512-GFP. Reverse transduction, also referred to as substrate-mediated gene delivery, consists in coating a surface with viral vectors and then adding cells for seeding, which will uptake viral vectors. It has been shown that this

method improves delivery efficiency of AAV9 vectors in mammalian cells (E. J. Lee et al., 2018). Experimental conditions can be found on Table 5.4.

Table 5.4. Experimental conditions for four groups, each one with a different seeding cell density (indicated as cells seeded/well). The MOI (multiplicity of infection, in this case: ratio of viral particles to cells) was the same for all groups.

C2C12 cell transduction with AAV9-Spc512-GFP				
Cells seeded/well	**5.00E+04**	**1.60E+05**	**2.00E+05**	**4.40E+05**
Wells/group	3	3	3	3
MOI	1.00E+06	1.00E+06	1.00E+06	1.00E+06
Virus /well	5.00E+10	1.60E+11	2.00E+11	4.40E+11
Total virus	1.50E+11	4.80E+11	6.00E+11	1.32E+12

The next day after seeding and reverse transduction, media was changed to differentiation media and cells were imaged for the next 11 days (representative images shown on Fig. 5.16). Seeding cell density of $2x10^5$ was determined to be optimal since it allowed for differentiation within a week with minimal cell death (assessed by microscopy). With this seeding density, 5 days after transduction cells displayed GFP fluorescence when visualised microscopically from AAV9-Spc515-GFP and could be harvested for DNA and RNA analysis. On day 7, cells were assessed by microscopy and differentiation was detected, leading to cell harvesting for protein analysis. On day 8, cells started to detach, hence it was decided to harvest for protein on day 7 the latest.

A double coating approach, by coating the plate with ECM before seeding the cells and adding another ECM coat after reverse transduction and cell seeding, was also tested

but showed no improvement in cell differentiation nor avoided cell death or cell detachment (data not shown).

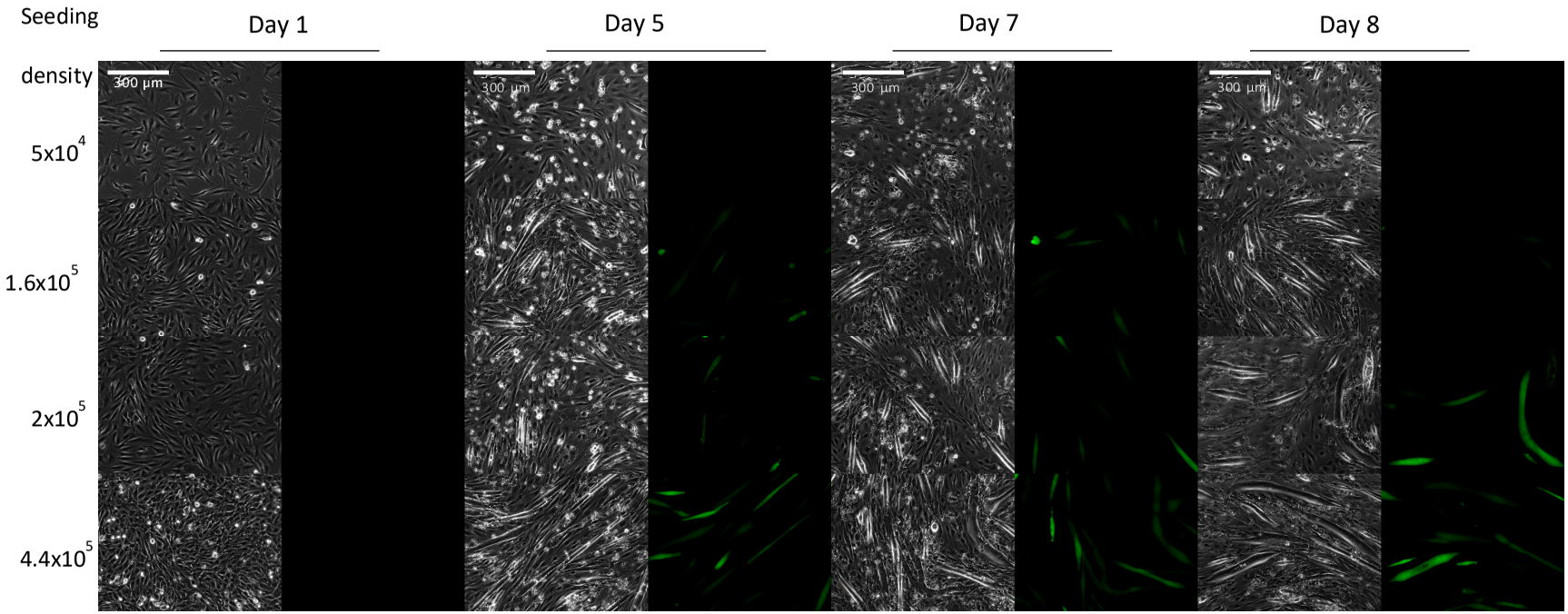


Figure 5.16. Representative images of C2C12 cells transduced with AAV9-Spc512-GFP, on day 1, 5, 7 and 8 after reverse transduction. Media changed to differentation media on day 1 after transduction. Brighfield and GFP channels can be observed for all groups. Cells imaged and acquired with a Zeiss fluorescence microscope (Axio Vision D1 with AxioCam MRm) and Software ZEN 2012. Magnification bar (300 µm) indicated in white.

5.3.2. *Reverse transduction of C2C12 cells with AAV9 vectors containing the SaCas9 multiplex constructs and individual gRNA constructs.*

Cells were seeded on 6-well plates with a cell density of $2x10^5$ cells/well for transduction with AAV9 with an MOI of $1x10^6$. Cells were harvested on day 5 after reverse transduction for DNA and RNA extraction and cells were harvested on day 7 for protein. Cell density and differentiation of cells transduced with: AAV9-Spc512-GFP (EGFP) as a control, AAV9-Spc512-SaCas9-BsaI-BbsI (Empty), AAV9-Spc512-SaCas9-multiplex-G14-G18 (Multiplex), AAV9-Spc512-SaCas9-G14 alone, AAV9-Spc512-SaCas9-G18 alone and co-transduced with both AAV9-Spc512-SaCas9-G14 and AAV9-Spc512-SaCas9-G18 (each at MOI of $0.5 x10^6$). Representative cell images can be observed in Fig. 5.17. There was no visible difference in cell death between treated and untreated groups. GFP was expressed from the positive control, confirming successful transduction.

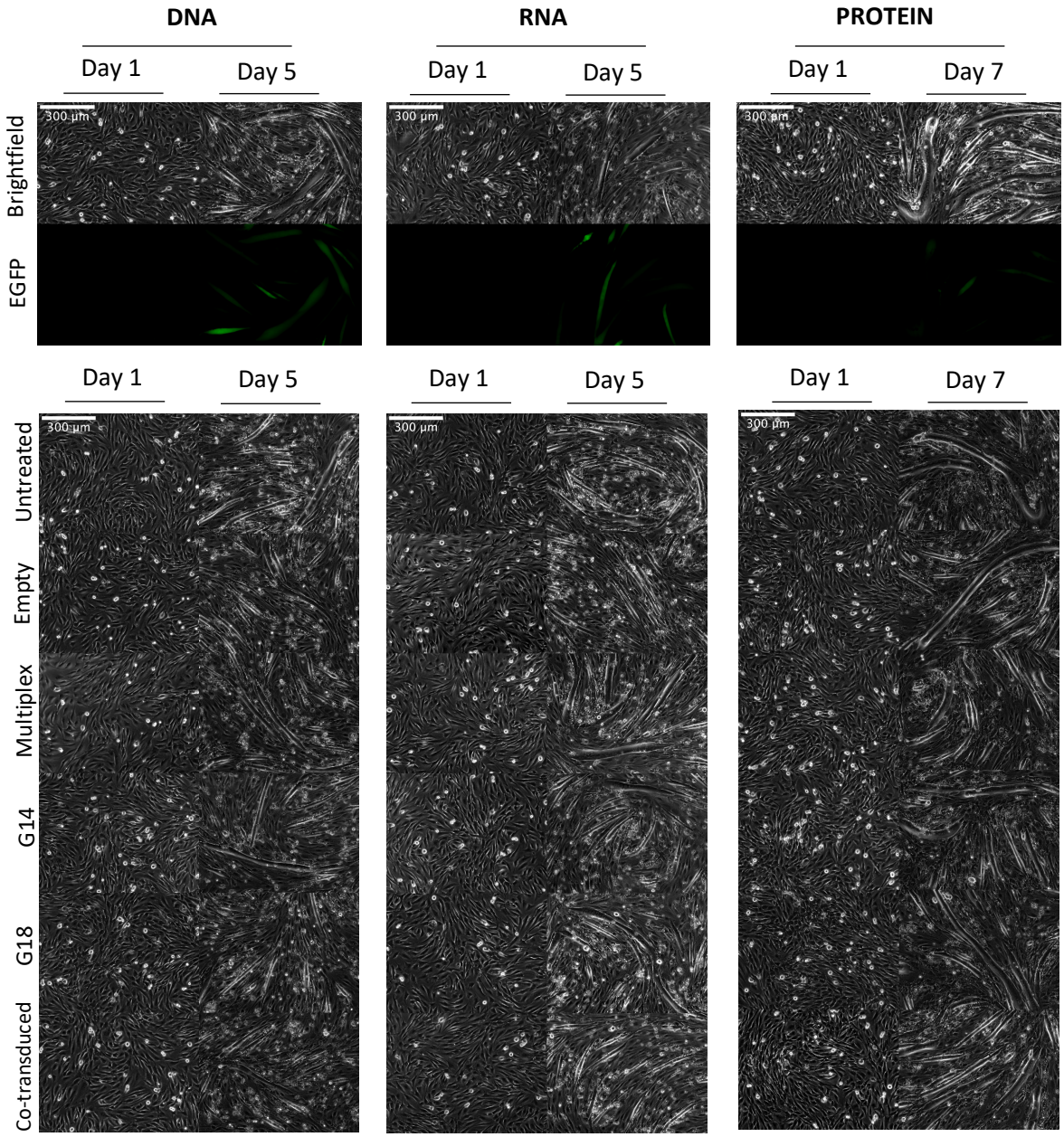


Figure 5.17. C2C12 cell images after reverse transduction. Cell transduced with: AAV9-Spc512-GFP (EGFP) as a control, AAV9-Spc512-SaCas9-BsaI-BbsI (Empty), AAV9-Spc512-SaCas9-multiplex-G14-G18 (Multiplex), AAV9-Spc512-SaCas9-G14, AAV9-Spc512-SaCas9-G18 and co-transduced with G14 and G18. Images of cells for DNA and RNA extraction taken on day 1 and 5; images for cells harvested for protein extraction taken on day 1 and 7. Cells imaged and acquired with a Zeiss fluorescence microscope (Axio Vision D1 with AxioCam MRm) and Software ZEN 2012.

5.3.2.1. *Analysis of DNA obtained from C2C12 cells transduced with AAV9 vectors.*

Genomic DNA was extracted from C2C12 cells harvested 5 days after reverse transduction. An end-point PCR with primer pairs designed to produce a band of 970 bp if there is a deletion between introns 18 and 55 were used (previously described on Section 4.4.2.1). No bands could be observed from any of the samples (Fig. 5.18), indicating that no deletion between introns 18 and 55 was detected from DNA samples.

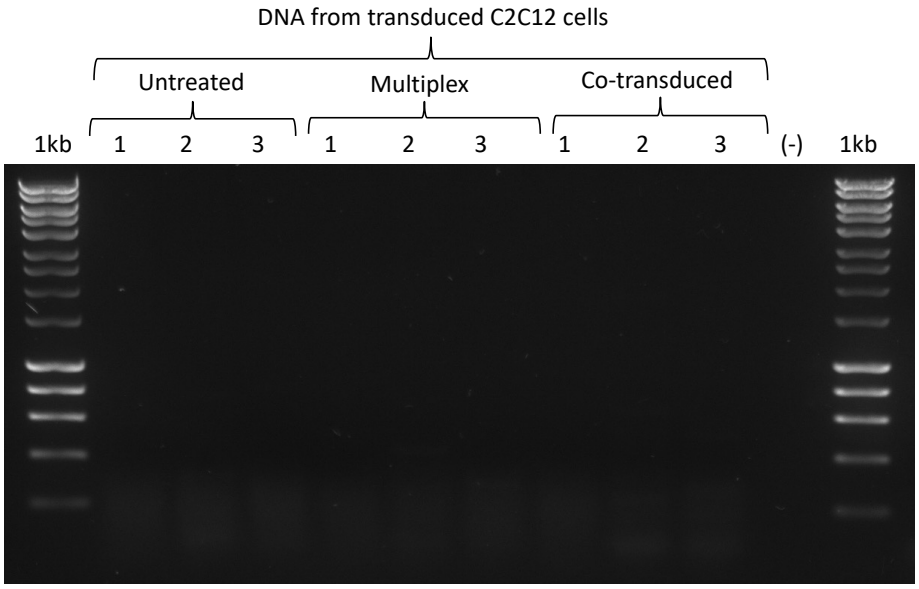


Figure 5.18. Gel image of PCR from DNA samples extracted from transduced C2C12 cells. From left to right in triplicates: untreated cells, transduction with AAV9-Spc512-SaCas9-multiplex-G14-G18 (Multiplex) and co-transduction with AAV9-Spc512-SaCas9-G14 and AAV9-Spc512-SaCas9-G18. If there is a deletion between introns 18 and 55, a PCR product of 970 bp would be amplified. Gel was 1% agarose (w/v) with 0.5X SYBR Safe in 1X TAE Buffer. Hyperladder I was used.

5.3.2.2. Assessment of SaCas9 expression on transduced C2C12 cells by RT-qPCR.

RNA was extracted from C2C12 cells harvested 5 days after reverse transduction. cDNA was obtained and processed the same way as cDNA from TA samples (described in Section 5.2.4). *Sa*Cas9 expression was normalised against the reference gene *Rplp0*. Results are shown in Fig. 5.19, *Sa*Cas9 seems to be expressed at the RNA level at least in all the treated groups with a significant difference when compared to untreated samples**:** untreated vs empty, p-value= 0.4681 (non-significant), untreated vs. multiplex, p-value=0.0209, untreated vs. co-transduced, p-value<0.0001, untreated vs G14, p-value=0.0024, untreated vs G18, p-value=0.0020. Co-transduced samples seem to be expressing the most *Sa*Cas9 with a significant difference (p-value=0.0220) when compared to samples treated with multiplex gRNAs, which is interesting considering the total MOI remained the same for al treatments.

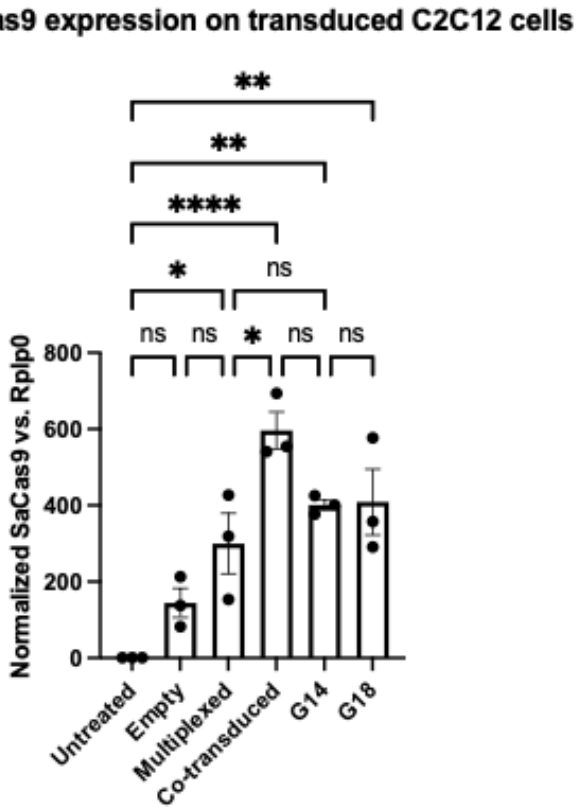


Figure 5.19. Normalized *Sa*Cas9 expression against reference gene *Rplp0*, from transduced C2C12 cells. From left to right: untreated cells, cells transduced with AAV9-Spc512-SaCas9-BbsI-BsaI (Empty), AAV9-Spc512-SaCas9-multiplex-G14-G14 (Multiplex), co-transduced with G14 and G18 (AAV9-Spc512-SaCas9-G14-BsaI and AAV9-Spc512-SaCas9-BbsI-G18) and transduced with G14 and G18 individually (n= 3 technical repeats per group). *Sa*Cas9 expression was found significant on all treated groups by mean comparison against untreated samples with a One-way ANOVA Analysis and a 95% confidence interval (p-value<0.05), followed by a Tukey´s test. Adjusted p-values for: untreated vs empty, p-value=0.4681 (non-significant), untreated vs. multiplex, p-value=0.0209 (*), untreated vs. co-transduced, p-value<0.0001 (****), untreated vs G14, p-value=0.0024 (**), untreated vs G18, p-value=0.0020(**). Comparison between groups: Empty vs. Multiplex = ns (non-significant), Multiplex vs. co-transduced, p-value=0.0220 (*), Co-transduced vs. G14 = ns, G14 vs. G18 = ns, Multiplex vs. G14 = ns. Error bars represent standard error of the mean. Graph and statistical analysis performed on Prism9 Software.

5.3.2.3. Assessment of Dmd expression by RT-qPCR on transduced C2C12 cells.

Standard curves and samples were prepared, and normalised copy numbers per reaction were obtained as described in Section 5.2.4.2. Results graphed and analysed on Prism9 Software can be seen in Figure 5.20.

There was no significant difference in expression of *Dmd e*xons 6-7 nor *Dmd* exons 20-21 when analysed by multiple mean comparison with a two-way ANOVA Analysis and a 95% confidence interval (p-value<0.05). From this it can be concluded that there were no detectable levels of deletion of exons 20-21 and therefore of exons 19 to 55, when analysing cDNA expression by qPCRs.

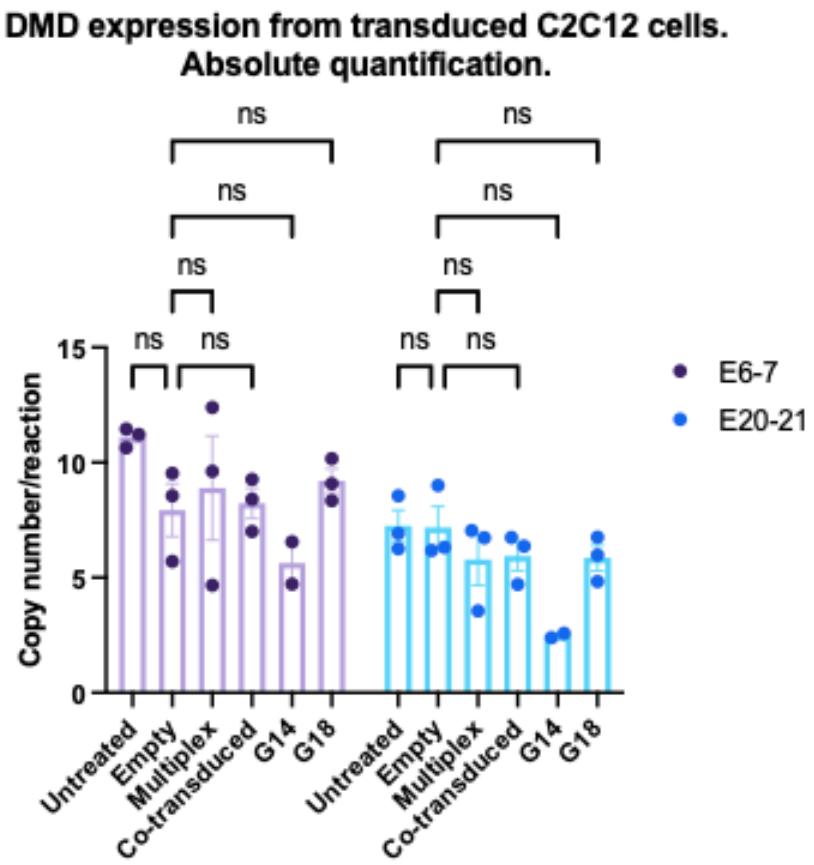


Figure 5.20. Absolute quantification of *Dmd* Exons 6-7 and Exons 20-21 expression. RNA extracted from transduced C2C12 cells (n = 3 technical repeats); cDNA obtained by reverse transcription PCR with a QuantiTect reverse transcription kit from QIAGEN. Standard curves prepared by serial dilutions from 1E+10 to 1E+1 copy numbers of g-blocks expressing cDNA of exons 6-7 and exons 20-21 from *Dmd* mouse gene and *Rplp0* as a reference gene. Master mix of SYBR green (FastStart Universal SYBR Green Master mix 2X with FastStart Taq DNA Polymerase, Reaction Buffer, Nucleotides (dATP, dCTP, dGTP, dUTP), SYBR Green I and a reference dye from Roche) prepared to 1X for each reaction with 400 nM of each primer (forward and reverse). 6 µL of mix and 4 µL of each sample, including standard curve samples, were loaded per well on a 96-well plate by triplicates, plates were processed on a LightCycler480 Instrument II from Roche and data was analysed on the LightCycler480 Software to obtain concentration of each sample, calculated by the Software based on standard curve from serial dilutions. To obtain normalised copy numbers per reaction: concentrations of samples were averaged, geometric mean of averaged concentrations was calculated for reference gene *Rplp0*, normalisation factor for each sample was obtained by dividing average *Rplp0* concentration by *Rplp0* geometric mean. Copy numbers per reaction were obtained for samples by dividing average gene of interest expression by normalisation factor. Results were graphed and analysed on Prism9 Software. There was non-significant difference in expression of exons 6-7 nor exons 20-21 between samples treated with the negative control (empty) and the rest of the groups, when analysed by multiple mean comparison with a two-way ANOVA Analysis, followed by a Tukey´s test (p-value<0.05).

5.3.2.4. Dystrophin protein expression assessment by Western Blot on transduced C2C12 cells.

Protein was extracted C2C12 cells harvested 7 days after reverse transduction, 30 µL of protein lysate were loaded per well on 3-8% Tris-Acetate gel alongside with HiMark pre-stained ladder from ThermoFisher. Membrane was processed with Manex1011C (1:100) for dystrophin and alpha-tubulin (1:10,000) as a reference gene. Del19-55 dystrophin can be seen expressed from control sample (cells transfected with pAAV-CMV-hDys-Del19-55-GFP) at 224 kilodaltons. All samples expressed full length dystrophin as expected. However, none of the samples showed a band for Del19-55 dystrophin.

From this, it can be concluded that there were no detectable levels of Del19-55 dystrophin expression in any of the treated samples. This could mean that protein levels were too low to detect by Western Blot or that there was no expression at all.

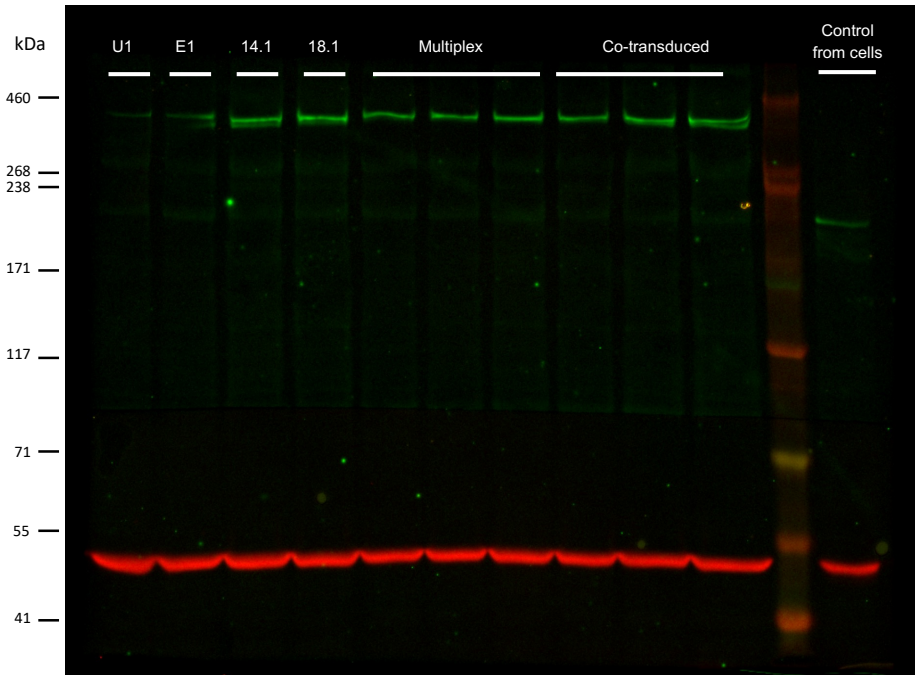


Figure 5.21. Western Blot to detect dystrophin from protein extracted from transduced C2C12 cells. From left to right: untreated, cells transduced with AAV9-Spc512-SaCas9-BsaI-BbsI (E1), AAV9-Spc512-SaCas9-G14-BbsI (14.1), AAV9-Spc512-SaCas9-BsaI-G18 (18.1), AAV9-Spc512-SaCas9-multiplex-G14-G18 (Multiplex by triplicates) and co-transduced with AAV9-Spc512-SaCas9-G14-BbsI and AAV9-Spc512-SaCas9-BsaI-G18 (by triplicates); HiMark pe-stained ladder from ThermoFisher and control from protein sample extracted from transfected HEK293T cells with pAAV-Spc512-hDys-Del19-55-GFP. 30 µg of protein lysate per well were loaded on a 3-8 Tris-Acetate gel and analysed with antibodies: Manex1011C (1:100, green) for dystrophin and alpha-tubulin (1:10,000, red) as a reference gene.

5.3.3. Transduction on H2KB-mdx cells: H2KB-mdx cell density optimization.

To select the optimal cell density for a transduction protocol, four different cell densities of H2KB-*mdx* cells were seeded on 6-well ECM (extra cellular matrix) coated plates and reverse transduced with AAV9-Spc512-GFP. Experimental conditions were the same as described ones on Section 5.3.1.

The next day after seeding and reverse transduction, media was changed to differentiation media and cells were imaged for the next 11 days (representative images shown on Fig. 5.22). It was decided that seeding cell density of $2x10^5$ was the optimal one to allow for differentiation within a week. With this seeding density, on day 5, cells showed fluorescence from AAV9-Spc515-GFP, although less when compared to GFP expression on C2C12s under the same experimental conditions. On day 7, cells showed differentiation. On day 8 some cells started to detach, hence it was decided to harvest for protein on day 7 the latest.

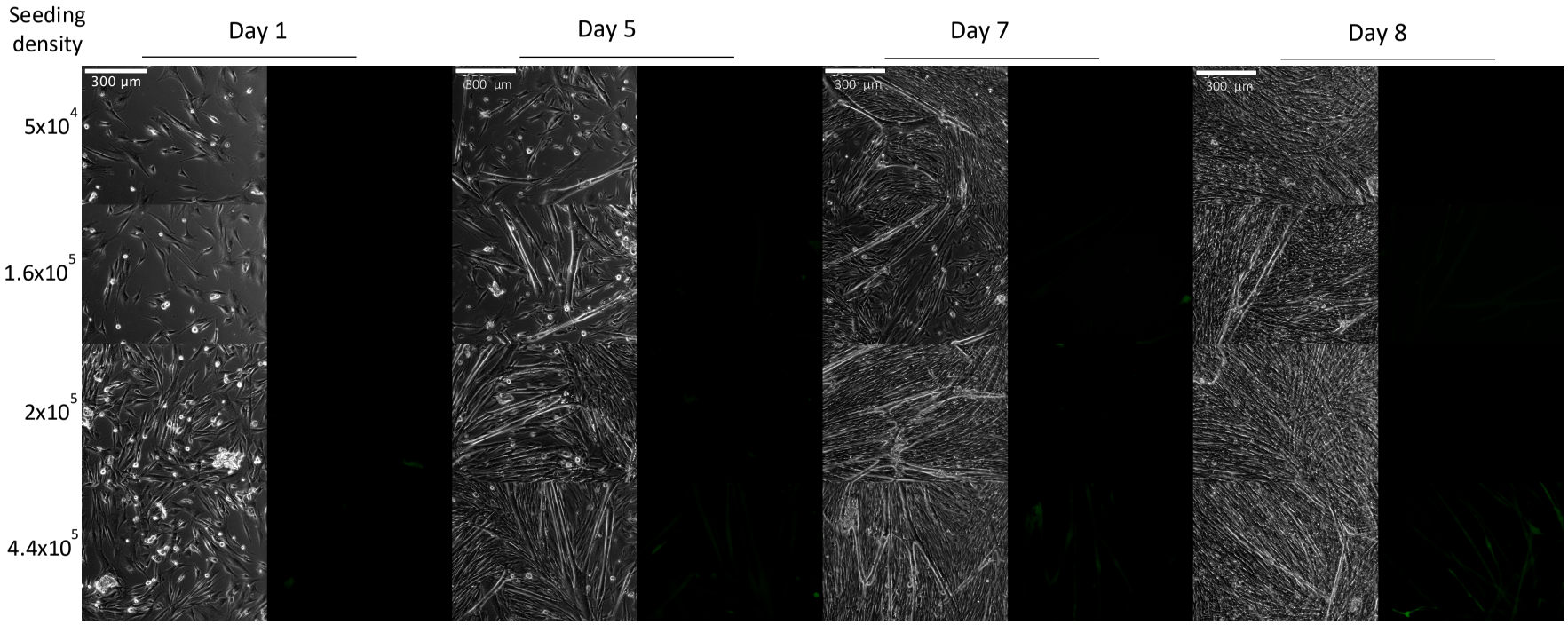


Figure 5.22. Representative images of H2KB-*mdx* cells transduced with AAV9-Spc512-GFP, on day 1, 5, 7 and 8 after reverse transduction. Media changed to differentation media on day 1 after transduction. Brighfield and GFP channels can be observed for all groups. Cells imaged and acquired with a Zeiss fluorescence microscope (Axio Vision D1 with AxioCam MRm) and Software ZEN 2012.

5.3.4. Transduction of H2KB-mdx cells with AAV9 vectors carrying SaCas9 multiplex constructs and individual gRNA constructs.

Cells were seeded on 6-well plates with a cell density of $2x10^5$ cells/well for transduction with AAV9 with an MOI of $1x10^6$. Cells were harvested on day 5 after reverse transduction for DNA and RNA extraction and on day 6 for protein extraction. Cell density and differentiation of cells transduced with: AAV9-Spc512-GFP (EGFP) as a control, AAV9-Spc512-SaCas9-BsaI-BbsI (Empty), AAV9-Spc512-SaCas9-multiplex-G14-G18 (Multiplex), AAV9-Spc512-SaCas9-G14, AAV9-Spc512-SaCas9-G18 and co-transduced with G14 and G14, can be observed on Fig. 5.23. There was no visible difference in cell death between treated and untreated groups. However, there was more cell death than expected in all samples on day 6, hence it was decided to harvest for protein on day 6. GFP expression from the positive control can be observed on day 6, confirming successful transduction.

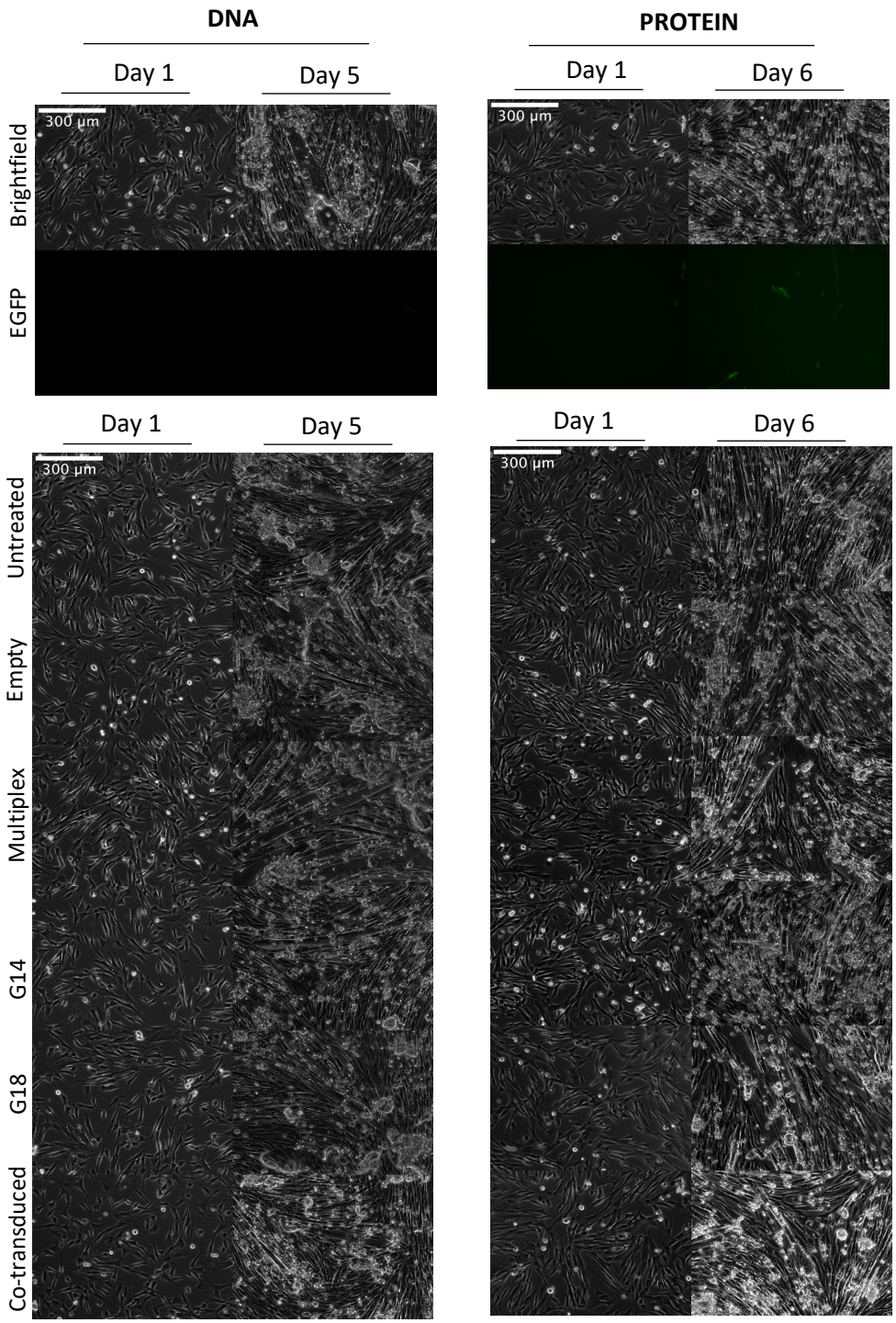


Figure 5.23. H2KB-*mdx* cell images after reverse transduction. Cell transduced with: AAV9-Spc512-GFP (EGFP) as a control, AAV9-Spc512-SaCas9-BsaI-BbsI (Empty), AAV9-Spc512-SaCas9-multiplex-G14-G18 (Multiplex), AAV9-Spc512-SaCas9-G14, AAV9-Spc512-SaCas9-G18 and co-transduced with G14 and G18. Images of cells for DNA extraction taken on day 1 and 5; images for cells harvested for protein extraction taken on day 1 and 6. Cells imaged and acquired with a Zeiss fluorescence microscope (Axio Vision D1 with AxioCam MRm) and Software ZEN 2012.

5.3.4.1. *Analysis of DNA obtained from H2kb-mdx cells transduced with AAV9 vectors.*

DNA was extracted from H2KB-*mdx* cells harvested 5 days after reverse transduction. An end-point PCR with primer pairs designed to produce a band of 970 bp if there is a deletion from intron 18 to 55 were used (previously described on Section 4.4.2.1). No bands could be observed from any of the samples (Fig. 5.24), indicating that no deletion was detected from DNA samples.

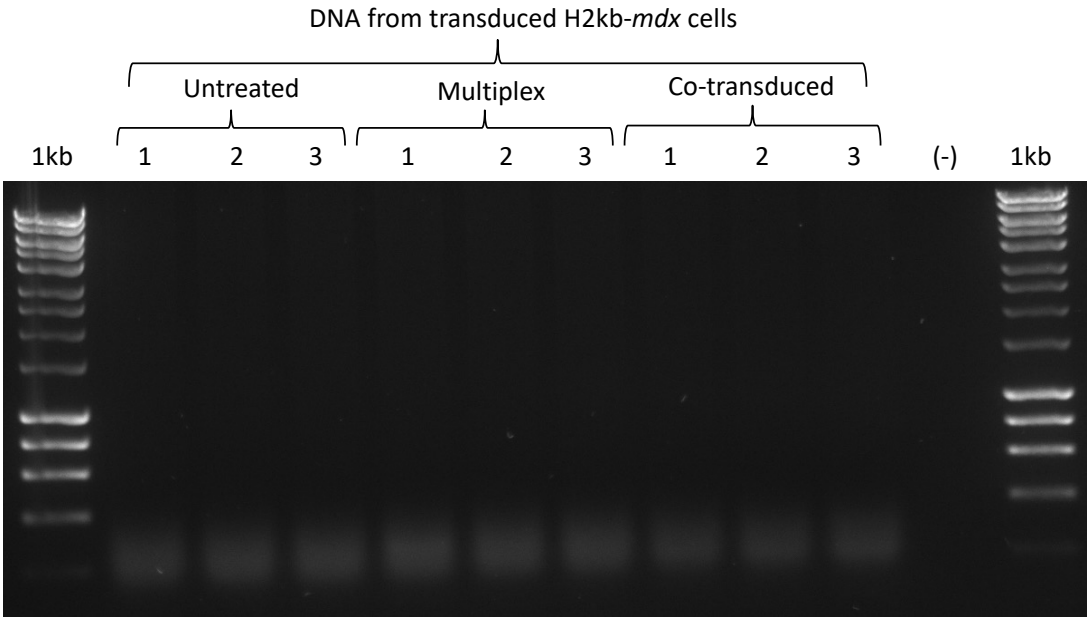


Figure 5.24. Gel image of PCR from DNA samples extracted from transduced H2kb-*mdx* cells. From left to right in triplicates: untreated cells, transduction with AAV9-Spc512-SaCas9-multiplex-G14-G18 (Multiplex) and co-transduction with AAV9-Spc512-SaCas9-G14 and AAV9-Spc512-SaCas9-G18. If Exons 19 to 55 were deleted, a PCR product of 970 bp was expected. Gel was 1% agarose (w/v) with 0.5X SYBR Safe in 1X TAE Buffer. Hyperladder I was used.

5.3.4.2. Assessment of SaCas9 expression by RT-qPCRs from transduced H2kb-mdx cells.

RNA was extracted from H2kb-mdx cells harvested 5 days after reverse transduction. Then, cDNA was obtained and processed as described in Section 5.2.4. *Sa*Cas9 expression was normalised against reference gene *Rplp0*, results can be seen on Fig. 5.25. *Sa*Cas9 seems to be expressed by all the treated groups. However, when compared to untreated samples mean, only groups treated with individual guides showed a significant difference: untreated vs. G14 showed a p-value<0.0001 and untreated vs. G18 a p-value=0.0014.

SaCas9 expressed on transduced H2kb-*mdx* cells.

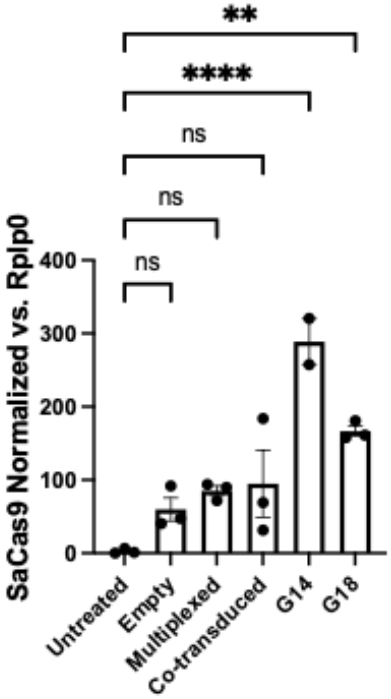


Figure 5.25. Normalized *Sa*Cas9 expression against reference gene *Rplp0*, from transduced H2kb-*mdx* cells. From left to right: untreated cells, cells transduced with AAV9-Spc512-SaCas9-BbsI-BsaI (Empty), AAV9-Spc512-SaCas9-multiplex-G14-G14 (Multiplex), co-transduced with G14 and G18 (AAV9-Spc512-SaCas9-G14-BsaI and AAV9-Spc512-SaCas9-BbsI-G18) and transduced with G14 and G18 individually (n = 3 technical repeats per group). *Sa*Cas9 expression was found significant on groups treated with individual guides by mean comparison against untreated samples with a One-way ANOVA Analysis and a 95% confidence interval (p-value<0.05), followed by a Dunnett´s test. Adjusted p-values for: untreated vs G14, p-value<0.0001 (****), untreated vs G18, p-value=0.0014 (**). Error bars represent standard error of the mean. Graph and statistical analysis performed on Prism9 Software.

5.3.4.3. Assessment of DMD expression by RT-qPCR on transduced H2kb-mdx Cells.

Standard curves and samples were prepared, and normalised copy numbers per reaction were obtained as described in Section 5.2.4.2. Results graphed and analysed on Prism9 Software can be seen on Figure 5.26.

There was no significant difference in expression of exons 6-7 nor exons 20-21 when analysed by multiple mean comparison with a two-way ANOVA Analysis and a 95% confidence interval (p-value<0.05). From this it can be concluded that there were no detectable levels of deletion of exons 20-21 and therefore of exons 19 to 55, when analysing cDNA expression by RT-qPCR.

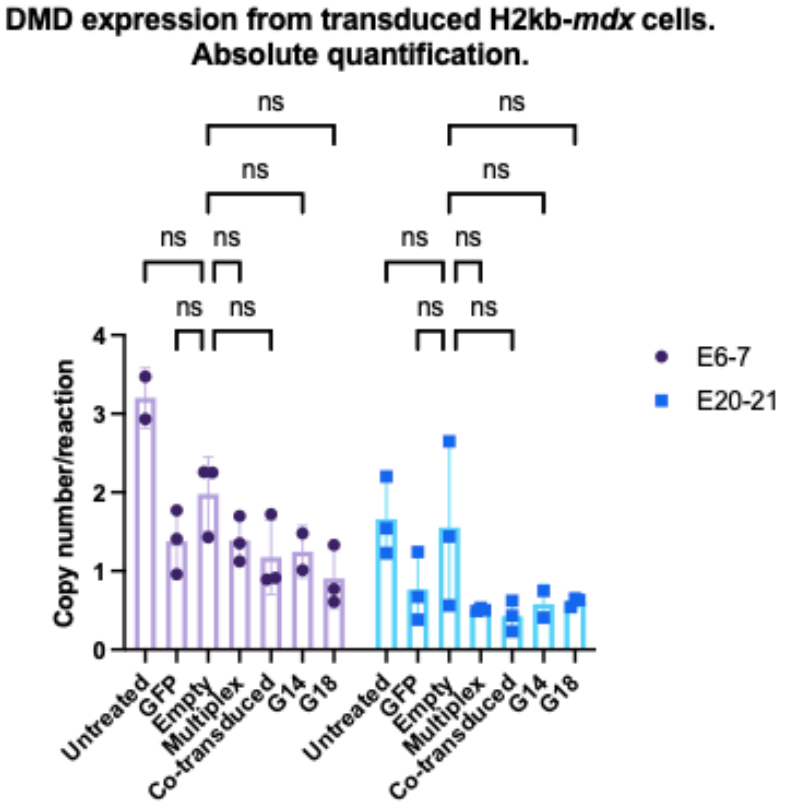


Figure 5.26. Absolute quantification of *Dmd* exons 6-7 and exons 20-21 expression. RNA extracted from transduced H2kb-*mdx* cells; cDNA obtained by reverse transcription PCR with a QuantiTect reverse transcription kit from QIAGEN (n = 3 technical repeats). Standard curves prepared by serial dilutions (1E+10 to 1E+1 copy numbers) of g-blocks expressing cDNA of exons 6-7 and exons 20-21 from *DMD* mouse gene and *Rplp0* as a reference gene. Master mix of SYBR green (FastStart Universal SYBR Green Master master mix 2X with FastStart Taq DNA Polymerase, Reaction Buffer, Nucleotides (dATP, dCTP, dGTP, dUTP), SYBR Green I and a reference dye from Roche) prepared to 1X for each reaction with 400 nM of each primer (forward and reverse). 6 µL of mix and 4 µL of each sample, including standard curve samples, were loaded per well on a 96-well plate by triplicates, plates were processed on a LightCycler480 Instrument II from Roche. Data was analysed on LightCycler480 Software to obtain concentration of each sample, calculated by the Software based on standard curve from serial dilutions. To obtain normalised copy numbers per reaction: concentrations of samples were averaged, geometric mean of averaged concentrations was calculated for reference gene *Rplp0*, normalisation factor for each sample was obtained by dividing average *Rplp0* concentration by *Rplp0* geometric mean. Copy numbers per reaction were obtained for samples by dividing average gene of interest expression by normalisation factor. Results were graphed and analysed on Prism9 Software. There was non-significant difference in expression of exons 6-7 nor exons 20-21 between samples treated with the negative control (empty) and the rest of the groups, when analysed by multiple mean comparison with a two-way ANOVA Analysis, followed by a Tukey´s test (p-value<0.05).

5.3.4.4. *Dystrophin protein expression assessment by Western blot from H2kb-mdx cells transduced with AAV9.*

Protein was extracted from H2kb-*mdx* cells harvested 6 days after reverse transduction and samples were processed the same as protein samples from transduced C2C12 cells (described in Section 5.3.2.4). None of the treated samples expressed a product band for Del19-55 dystrophin. It can be concluded that there were no detectable levels of Del19-55 dystrophin expression in any of the treated samples. This could mean that protein levels were too low to detect by Western Blot or that there was no expression at all.

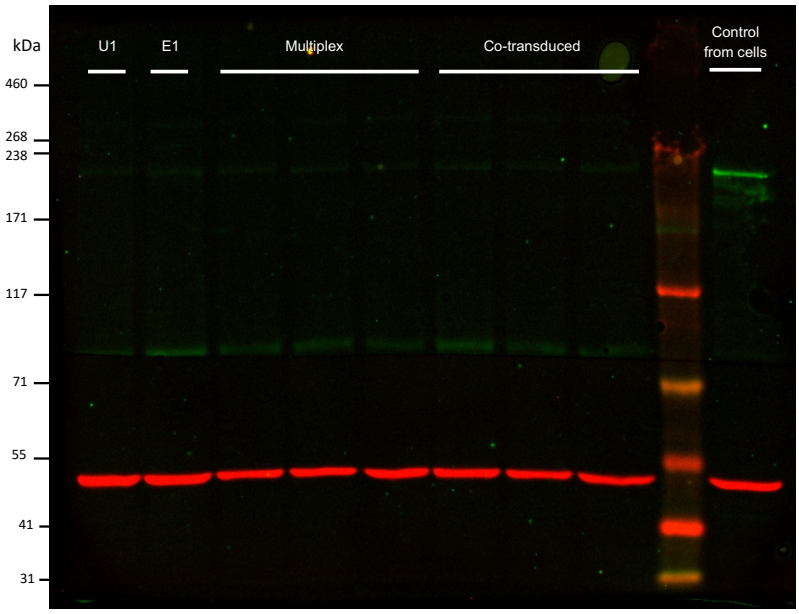


Figure 5.27. Western Blot to detect dystrophin from protein extracted from transduced H2kb-*mdx* cells. From left to right: untreated, cells transduced with AAV9-Spc512-SaCas9-BsaI-BbsI (E1), AAV9-Spc512-SaCas9-G14-BbsI (14.1), AAV9-Spc512-SaCas9-BsaI-G18 (18.1), AAV9-Spc512-SaCas9-multiplex-G14-G18 (Multiplex by triplicates) and co-transduced with AAV9-Spc512-SaCas9-G14-BbsI and AAV9-Spc512-SaCas9-BsaI-G18 (by triplicates); HiMark pe-stained ladder from ThermoFisher and control from protein sample extracted from transfected HEK293T cells with pAAV-Spc512-hDys-Del19-55-GFP. 30 µg of protein lysate per well were loaded on a 3-8 Tris-Acetate gel and analysed with antibodies: Manex1011C (1:100, green) for dystrophin and alpha-tubulin (1:10,000, red) as a reference gene.

5.4. Assessment of AAV vectors infectivity by Infectious Centre Assay (ICA).

Since no editing was detected from the AAV vectors with our CRISPR systems *in-vivo* nor *in-vitro*, it was hypothesized that there could be a problem with the vectors or that the editing efficiency from the system was too low and therefore deletion levels were too low be detected.

To rule out an issue with the vectors, it was decided to assess our AAV9 vectors infectivity. Samples from each prep were sent to the clinical vector core of the UMR1089 to perform an Infectious Center Assay (ICA). The ICA allows the quantification of infectious particles in each AAV prep (Salvetti et al., 1998). The assay was performed as described in Materials & Methods Section 2.17. Then, replication events from infectious particles were detected by chemiluminescence and quantified following hybridization with a transgene specific probe. Results are presented in Table 5.5.

Expected ratios for AAV9 vectors range between $1x10^4$ and $5x10^4$. All our vector preps showed a ratio within this range, alongside the internal controls from the vector core of the UMR1089 (expressing mouse micro-dystrophin MD1). Thus, confirming appropriate levels of infection achieved with our vectors. These results rule out any potential issues with the vector capsid or the ITRs and the packaging of the vectors.

Table 5.5. Infectious Centre Assay results for AAV vectors. Sample titres quantified by qPCRs targeting the *Sa*Cas9 (as described on Section 5.1.2) or targeting an ITR sequence (based on (D'Costa et al., 2016) and performed by the vector core of the UMR1089), reported in viral genomes per mL. Infectious titre reported in infectious particles per mL. Infectivity ratio reported in viral genomes (quantified by ITR qPCR) per infectious particles.

AAV batch	vg/mL titer (SaCas9 qPCR)	vg/mL titer (ITR qPCR)	Infectious titer (ip/mL)	Probe target	Ratio vg/ip
AAV9-Spc512-GFP	1.35E+14	9.60E+13	7.20E+09	Spc512 promoter	1.33E+04
AAV9-Spc512-Multi-G14-G18	4.65E+13	2.40E+13	2.20E+09	Spc512 promoter	1.09E+04
AAV9-Spc512-G18	5.69E+13	3.60E+13	2.10E+09	Spc512 promoter	1.71E+04
AAV9-Spc512-Empty	6.35E+13	3.20E+13	1.10E+09	Spc512 promoter	2.91E+04
AAV9-Spc512-mMD1 (internal control)	N/A	6.90E+12	5.30E+08	Spc512 promoter	1.30E+04
AAV9-Spc5.12-mMD1 (internal control)	N/A	6.90E+12	4.90E+08	mMD1	1.41E+04

5.5. DISCUSSION.

An *Sa*Cas9 system with two multiplex gRNAs, targeting intron 18 and 55, was established to generate the deletion of exons 19 to 55 of the *Dmd* gene. Previous *in-vitro* results confirmed achievement of this large deletion in mouse cell lines (N2A and C2C12 cells) with our multiplex system and by co-delivery of constructs with individual gRNAs.

To improve delivery efficiency, constructs were packaged into AAV9 vectors and tested *in-vivo* in *mdx* mice. To assess potential functionality generated by the treatments, muscle electrophysiology analysis was performed on groups co-transduced and treated with multiplex construct, alongside controls. No significant effects were observed with any of the treatments. DNA was then analysed; individual gRNAs showed an editing efficiency of ~5% (G14) and ~12% (G18). However, it was not possible to confirm a deletion at DNA level by PCR, most likely due to extremely low levels of deletions, nor by qPCR due to the complexity of the region flanking the junction of introns 18 and 55 (150 bp, 75 bp on each end of the junction), which has a low GC content and "TGTTGTTGTT" repeats constituting 20% of the overall sequence.

To assess expression from AAV9 vectors, an RT-qPCR was established to detect *Sa*Cas9 expression. When normalised against reference gene *Rplp0* and compared to wild type samples treated with saline solution, there was significant *Sa*Cas9 expression from all treated groups. Transduction efficiency could have been further investigated by running

a qPCR on genomic DNA to assess *Sa*Cas9 expression normalised against a reference DNA sequence.

To assess deletion of exons 19 to 55 at RNA level, RT-qPCRs detecting exons 6-7 and exons 20-21 expression were performed. No significant differences were detected in any of the groups, implicating that no deletion of exons 20-21 was detected. Since it is possible that deletion levels were too low to be detected by end-point PCR or RT-qPCR, dystrophin protein expression was assessed by immunohistochemistry and Western Blot. There was a significant increase in dystrophin positive fibres on samples treated with the multiplex construct when compared to negative control (*mdx* mice treated with saline solution), from 0.73% to 1.45% positive fibres. There was non-significant increase (to 0.85% positive fibres) on samples co-transduced with both guides. Even though there was in increase in samples treated with the multiplex construct, positive fibres levels were still below 2%.

At this stage it was not possible to confirm the exact reason for the lack of deletion. So, vectors were further tested *in-vitro* on two cell lines (C2C12 and H2kb-*mdx* cells). It was not possible to detect a deletion at DNA level by end-point PCR from either cell line. *Sa*Cas9 expression was detected on both cell lines by RT-qPCRs, although expression was lower on H2kb-*mdx* cells when normalised against reference gene *Rplp0*, ranging from 50-300X higher than *Rplp0*, compared to *Sa*Cas9 expression 100-600X higher than *Rplp0* on C2C12 cells. However, it was no possible to detect a significant difference in exons

20-21 expression from any of the groups, indicating that no deletion was detected at RNA level. To overcome detection of low levels of deletion, a nested end point PCR was attempted from cDNA obtained from treated samples. However, finding specific primer pairs was not possible. If time had permitted more primer pairs could have been screened.

At this stage the lack of deletion could be due to low levels of deletion, too low to be detected, or a lack of activity from our AAV vectors. A potential issue with the transgene expression was ruled out as the cassettes were previously tested *in-vitro* by plasmid transfection in N2A cells and by nucleofection in C2C12 cells and achieved a deletion between introns 18 and 55 in both cell lines. To confirm there was not an issue related to the AAV vectors, an ICA assay was performed. Results confirmed appropriate infectivity ratios from all our vectors. Based on this, it can be concluded that a lack of deletion in muscles was most likely caused by low editing efficiency from our systems and hence extremely low levels of deletion. Furthermore, C2C12 and H2KB-*mdx* myoblasts are not permissive cell lines, this could lead to very low transduction efficiencies (Zentilin et al., 2001, Cervelli et al., 2008, Lovric et al., 2012) and hence a lack of a detectable deletion in transduced cells.

The main limiting factor for the deletion efficiency is the gRNA editing efficiencies. When individually assessed G14 achieved 5% editing and G18 achieved 12% editing *in-vivo,* meaning that the maximum potential deletion efficiency would be 5%. However, there

are other factors that would decrease the efficiency of such a large deletion, such as re-insertions or inversions of the excised regions (Canver et al., 2014, Y. Wang et al., 2018) or AAV sequence integration (Nelson et al., 2019).

It is noteworthy that a deletion as large as the one achieved in this project (approximately 800 kbp), had not been attempted in the *DMD* gene and with an *Sa*Cas9 system. In previous studies using *Sp*Cas9 systems, it was shown that paired gRNAs can precisely generate large deletions up to 23 kbp (Han et al., 2014). Deletions from 1.3 kb to greater than 1 Mb have also been achieved (Canver et al., 2014). However, in the latter study, deletions of 1 Mb only achieved 0.58% of deletion frequency and it was concluded that the larger the deletion is, the less efficient the cutting made by the gRNAs is. To assess how the size of the deletion in the *DMD* gene affects the efficiency, including a system aiming for a smaller excision alongside, such as exons 45-55 or 47-58, which has been achieved in *mdx* mice with an *Sa*Cas9 system targeting exons 47 and 58 (Duchêne et al., 2018), would have been useful to compare with our system.

Furthermore, it was not possible to quantify the deletion efficiency achieved with our *Sa*Cas9 systems. This could be assessed with a modified TIDE analysis, in which a PCR product composed of the expected "edited sequence" would be used as the control sequence, as shown by (Duchêne et al., 2018). However, they generated the control PCR product from a purified PCR product with the expected deletion, that was previously cloned into a plasmid. In contrast, our "edited sequence", meaning the region flanking

the *de novo* junction of introns 18 and 55, is an AT rich region which complicates cloning of its PCR products or synthesis of a g-block to use as a control. This could have been circumvented by implementing cell sorting for the *in-vitro* work to enable the generation and enrichment of monoclonal populations.

In our *in-vivo* experiment, mice were injected with a dose of $1x10^{11}$ vp/TA muscle, which is comparable with doses used in other studies (Hanson et al., 2022, Long et al., 2016; Nelson et al., 2016, Tabebordbar et al., 2016). However, a strategy to increase efficacy of our system could be to increase the treatment dose, although there would be an increase in the risk of an immune reaction. Alternatively, an early intervention in neonate mice before replacement of muscle cells with fibrotic or adipose tissue could allow functional recovery and prevent abnormalities associated with the disease, as shown in other studies (Long et al., 2016), (Nelson et al., 2016), (Tabebordbar, Zhu, Cheng, Widrick, et al., 2016), (Bengtsson et al., 2017) and (Y. Zhang et al., 2022).

Additionally, it has been shown that dystrophin accumulation is progressive and maintained up to 6-months (Nelson et al., 2016), 12-months (Nelson et al., 2019) and 18-months (Hakim et al., 2018) in *mdx* mice. It would be interesting to assess a second time point after treatment with our constructs (only assessed 2 months after treatment).

Strategies that could be attempted to improve our systems include improving gRNA design and editing efficiency. This could be achieved by screening more gRNAs targeting different regions within exons 18 and 55 that present less complexity or assessing different Cas systems, such as *Sp*Cas9 (Long et al., 2014, Long et al., 2016, Nelson et al., 2016, Tabebordbar et al., 2016) and Cpf1 (Y. Zhang et al., 2017), which have been used in *mdx* mice to successfully skip exon 23.

6. General discussion.

6.1. Discussion.

This PhD thesis has provided a comprehensive exploration of the development of a novel genome editing strategy for a particular neuromuscular disease and highlights the complexities and challenges of translating research to *in-vivo* applications. In this research project, the generation of a large deletion that would eliminate approximately 81% of genetic mutations, that lead to Duchenne muscular dystrophy (DMD), was explored using a CRISPR/Cas9 system.

DMD is caused by mutations within the *DMD* gene that lead to the lack of dystrophin protein expression and therefore a lack of muscle stability (Hoffman et al., 1987). Analysis of patients' phenotypes and their correlation to their genotypes have shown that truncated dystrophin forms can be functional and lead to a less severe phenotype (England et al., 1990). It was hypothesised that by generating a deletion between introns 18 and 55 of the *DMD* gene, a *de novo* intronic junction would be generated through NHEJ repair and would lead to expression of a truncated but functional dystrophin protein, which was named Del19-55 dystrophin.

The genome editing strategy proposed in this work has some advantages over strategies like gene addition and exon skipping. Unlike exon skipping strategies, this genome

editing strategy would possess a high patient applicability, as it would remove ~81% of mutations in the *DMD* gene including the mutational hotspot in exons 45-55 (Tuffery-Giraud et al., 2009) and opposite to gene addition strategies, a genome editing strategy would most likely not require repeated administration as the genetic corrections are permanent. Reduced need of re-administration would decrease the chances of immune response. Furthermore, to increase safety of this strategy, it was decided to target intronic regions to avoid unintentional effects in expression from exons.

The main research questions were:

- Does the truncated Del19-55 dystrophin protein possess potential functionality?
- Can a large deletion, of approximately 800 kbp, between introns 18 and 55 of the *Dmd* gene be achieved with an *Sa*Cas9 system?
- Can the in-frame deletion of exons 19 to 55 be achieved *in-vivo* with our CRISPR/Cas9 systems and would this deletion lead to the expression of Del19-55 dystrophin? Would the treatments have a beneficial functional effect in dystrophic muscles?

In this research project, a new truncated dystrophin protein was generated, Del19-55 dystrophin. The protein functionality of Del19-55 dystrophin was assessed *in-silico* and results suggested that Del19-55 dystrophin should express as a functional protein. These results were validated by generating a cDNA construct expressing this truncated protein

and testing it *in-vitro* and *in-vivo*. Co-localisation of dystrophin, GFP (fused to Del19-55 cDNA construct) and proteins from the DAPC (α-sarcogylcan and β-dystroglycan) at the sarcolemma in TA muscle sections from treated mice, suggest that Del19-55 protein has potential to be functional. However, the effect of the lack of interaction with nNOS, due to the lack of its domain harboured in spectrin-like repeats 16-17 (encoded by deleted exons 41-45) needs further investigation. It is possible that Del19-55 would be partially functional, if so, further investigation would be needed to confirm if functionality levels are enough to produce beneficial effects in muscles. This could be explored by generating a mouse model with exons 19 to 55 deleted and assessing the phenotype.

Interestingly, a DMD mouse model was generated by deleting exons 8 to 34 (430 kbp) with CRISPR/Cas9, which disrupted the reading frame and led to the absence of functional dystrophin production. This was reflected in the mice phenotype, which presented similar characteristics to *mdx* mice, including age-related decrease in muscle strength, increase creatine kinase, muscle fibrosis and central nucleation (Egorova et al., 2019). Furthermore, a humanized mouse model was generated by replacing mouse exon 51 with human exon 51 and then deleting exon 50. Then, to assess a genome editing strategy targeting splice acceptor of exon 51 and restore the reading frame to rescue dystrophin expression, exon 51 was deleted with an *Sp*Cas9 system delivered in a dual AAV9 vector. Dystrophin was restored to 18-26% of wild type levels in multiple skeletal muscle and heart. This humanized model allowed for *in-vivo* assessment of human gRNAs, which would facilitate clinical translation of the system (Y. Zhang et al., 2022).

An alternative strategy to assess Del19-55 dystrophin functionality would be to use our cDNA construct as a mini-gene and attempt a gene addition strategy. However, our current cDNA construct expressing Del19-55 dystrophin is too large to be packaged into an AAV vector. To further analyse this construct *in-vivo*, it could be packaged into a lentiviral vector to improve delivery efficiency (replacing plasmid injection experiments). A recent study demonstrated successful delivery of a micro-dystrophin with a lentiviral vector into *mdx* mice and restored dystrophin expression in skeletal and cardiac muscles, leading to a statistically significant improvement in motor performance (Eren et al., 2023). Additionally, expression of the construct expressing Del19-55 dystrophin could be increased by codon optimising the cDNA sequence, as it has been done in micro-dystrophins (Athanasopoulos et al., 2011). Furthermore, if Del19-55 dystrophin was assessed as a mini- or micro-dystrophin, improvement of potential functionality could be achieved by including the nNOS domain expression.

One of the main differences between Del19-55 dystrophin and micro-dystrophin MD1 (ΔR4-23/ΔCT) is that our truncated protein retains the C-terminal domain, which is largely truncated or eliminated in most micro-dystrophins (Duan, 2018). Interestingly, a recent study evaluated for the first time the impact of the inclusion of a full-length C-terminal domain in MD1 micro-dystrophin (ΔR4-23/ΔCT). It was demonstrated that MD1 can restore normal levels of interaction with most DAPC partners in skeletal and cardiac muscles of DMD^{mdx} rats in the absence of the C-terminal domain, while inclusion of the

C-terminal domain resulted in a supra-physiological association with some of the DAPC. However, it was not possible to demonstrate if its inclusion led to added therapeutical benefits (Bourdon et al., 2022).

To investigate if a large deletion, of approximately 800 kbp, between introns 18 and 55 could be achieved with an *Sa*Cas9 system, a multiplex *Sa*Cas9 system expressing two gRNAs, one targeting *Dmd* intron 18 and the second one targeting *Dmd* intron 55 was established. The large genomic deletion was achieved *in-vitro* with the multiplex construct and by co-delivery of individual gRNAs in their respective constructs. However, various limitations were highlighted in the assessment of these systems. Quantification of such a large deletion was challenging. This was mainly due to the sequence complexity of the intronic region flanking the target site. It has been showed that complexity of introns is lower than that of coding regions. Low complexity reflects a biased nucleotide composition due to simple sequence repeats or imperfect direct and inverted repeats (Orlov et al., 2006), as particularly observed in intron 55. Alternative introns that could be targeted with this strategy to avoid intron 55, would be introns 17 and 54. However, intron 54 sequence also presents a low GC content. This could be circumvented by target exons, which is a common strategy in the context of DMD. Deletion of exons 47 and 58 was achieved with an *Sa*Cas9 system, forming a hybrid exon 47-58, which led to dystrophin restoration in del52h*DMD/mdx* mice (Duchêne et al., 2018). In a different study, an *Sa*Cas9 system achieved deletion of exon 23 by targeting the 5´and 3´ end of the exon. This system was delivered locally by dual AAV9 vectors injected into the TA

muscles of adult *mdx* mice. Deletion levels of exon 23 reached an average excision rate of ∼40% (quantified by TaqMan PCR) and led to a significant increase in specific force and attenuated force drop after eccentric damage (Tabebordbar, Zhu, Cheng, Widrick, et al., 2016).

In this project, predictive algorithms that calculate on- and off-target scores, adapted for *Sa*Cas9 (Hsu et al., 2013, Doench et al., 2014, Najm et al., 2018, Tycko et al., 2018) were used for gRNA design. However, these algorithms rely on gRNA sequence features and the PAM recognition at target site. Algorithms used to predict on-target activity of *Sp*Cas9 and *Sa*Cas9 sgRNA consider single and dinucleotide position-specific nucleotides and GC content within the sgRNA sequence: gRNAs with low GC content tend to be less active and certain nucleotides in particular position of the sgRNA sequence allow higher activity, such as a guanine at the nucleotide immediately adjacent to the PAM sequence (position 20 at 3´end), cytosine is unfavourable at this position but preferred at position 16 and adenine is preferred in the middle of the sgRNA sequence (Doench et al., 2014, Doench et al., 2016, Najm et al., 2018). However, it has been suggested that the parameters considered by these algorithms, do not consider changes in genomic context that could impact Cas9 activity independently from cleavage at the target site (Moreb & Lynch, 2021).

There is strong evidence suggesting that gRNA sequence activity is largely influence by the ability of the Cas9/gRNA complex to find the target site, rather than the activity at

the target site itself (Moreb & Lynch, 2021). Some contributing factors include non-target interactions, that depend on the number of potential PAM sites within the whole genome that could compete with interaction of the Cas/gRNA complex and the PAM sequence at the target site, limiting the ability of the Cas9/gRNA complex to find its target site by increasing the "search time" (Sternberg et al., 2014, Moreb et al., 2020); unwanted secondary structures that can form within the gRNA and lead to reduced functionality, although the degree to which a predicted structure can inhibit Cas9/gRNA activity is not well characterized (Moreb & Lynch, 2022), and Cas9 target accessibility, which is impeded by regions with high nucleosome occupancy (Horlbeck et al., 2016).

Furthermore, it should be considered that to generate a large deletion, gRNAs need to achieve high cleavage efficiencies. There are a couple of strategies that could be implemented to achieve this: improving gRNA design by improving current algorithms or improving the gRNA secondary structure. In a recent study, gRNAs were modified to avoid potential gRNA misfolding that interferes with CRISPR/Cas9 cleavage. To overcome this, gRNAs were engineered with a highly stable hairpin in their constant parts. This approach was named "genome-editing optimized locked design" or GOLD-gRNA and increased editing efficiency to around 1000-fold (Riesenberg et al., 2022). Authors suggest that this method would be particularly useful if target genomic sites are difficult to edit. This approach could be applied to the gRNAs that have been optimised in this project (G14 targeting intron 18 and G18 targeting intron 55) by modifying the

first hairpin on the tracrRNA and adding and extremely stable C(UUCG)G loop motif (referred to as GOLD tracrRNA).

Originally, a TTTT motif was used to connect the gRNA-tracrRNA (gRNA scaffold) (Jinek et al., 2013). The *Sa*Cas9 constructs used in this project still maintain this TTTT motif at the gRNA scaffold. This TTTT region can inhibit transcription from the U6 promoter by RNA polymerase III, which is why they should be avoided within the gRNA sequence (Wong et al., 2015). However, strategies to optimise gRNA structure and improve its expression levels include substituting on of the "Ts" from this TTTT motif at the scaffold with an A and extending the gRNA duplex region by five nucleotides (Chen et al., 2013, Dang et al., 2015). This strategy has been tested on *Sa*Cas9 systems producing promising results (Chen et al., 2016), including the excision of exon 23 of the *Dmd* gene with a modified *Sa*Cas9-gRNA to enhance activity (Tabebordbar, Zhu, Cheng, Chew, et al., 2016). Implementing these strategies, particularly the T to A bp change in our constructs could improve gRNA expression and activity.

To attempt the deletion of exons 19 to 55 *in-vivo*, AAV9 vectors packaging our CRISPR/Cas9 systems were produced to assess their efficiency and potential functionality in a dystrophic mouse model (*mdx* mice). Unfortunately, it was not possible to confirm the deletion of exons 19 to 55 *in-vivo*. Our system had a few limitations that could have affected the outcomes, such as low gRNA efficiency and the size of the ultra large deletion (800 kpb) attempted *in-vivo*.

Other studies have implemented a dual gRNA approach with a CRISPR system to excise exons, that could lead to dystrophin restoration. In a recent study, an *Sa*Cas9 system was used to target intronic regions and delete *Dmd* exon 23 (Hanson et al., 2022). In this study, the maximum expression of dystrophin was 5.7% and was insufficient to extend life span of treated dKO (double dystrophin and utrophin knockout) mice. The main differences between their strategy and the one used in this project are the delivery methods (dual vs single AAV vector) and the length of the deletion produced (437 bp vs. 800 kbp). For the deletion of exon 23, a dual vector delivery was attempted with one AAV vector expressing the *Sa*Cas9 and the second AAV vector expressing both gRNAs. In an attempt to enhance system efficacy and to make the strategy translationally relevant, we delivered both gRNAs and the *Sa*Cas9 in one multiplexed construct. Interestingly, Hanson et al., (2022) report a seamless repair of the up and downstream introns at the expected gRNA cut site with low evidence on indels, which was also observed in our results.

Deletion of exon 23 has been achieved with *Sa*Cas9 systems in neonatal and adult *mdx* mice, a mouse model harbouring a nonsense mutation in exon 23 with terminates dystrophin production, and results have shown enhancement in muscle force. In a study targeting introns 22 and 23 with a dual gRNAs approach to generate a 1,171 bp deletion, exon 23 was successfully deleted. The *Sa*Cas9 system was delivered with dual AAV8 vector intramuscularly injected into TA muscles of *mdx* mice, with one vector carrying

the *Sa*Cas9 and the second vector carrying both gRNAs (targeting introns 22 and 23). Exon 23 was deleted in 2% of alleles from the whole muscle lysate (quantified by ddPCR) and led to dystrophin restoration to levels of 67% dystrophin positive fibres and 8% dystrophin protein expression (quantified by Western Blot). Interestingly, gRNAs were screened in C2C12s by electroporation and both gRNAs showed around 12% editing activity (Nelson et al., 2016). The gRNAs designed in this project, were originally screened in N2A cells, in which they showed an editing activity of ~12% (G14) and ~18% (G18). However, when tested on C2C12 cells, they showed ~5% (G14) and ~10-30% (G18) editing activity, which reflects activity showed *in-vivo* of ~5% and ~12%. Furthermore, on-target predicted activity calculated by algorithms was of 20% (G14) and 45% (G18). These results can lead to conclude two main lessons, activity predicted by algorithms is not accurate and more importantly, N2A cells are not an optimal cell line to screen gRNAs targeting the *Dmd* gene. In contrast, C2C12 cells, a mouse muscle cell line, would be the optimal cell line to assess editing activity of gRNAs targeting *Dmd*. However, it must be noted that gRNA screening in C2C12 cells would require plasmid delivery by nucleofection rather than transient transfection (allowed by N2A cells) with the disadvantage that this method is more costly and time consuming. It also must be noted, that although it was not quantified, a deletion between intron 18 and 55 was achieved in C2C12 cells with our multiplex plasmid and by co-nucleofection of G14 and G18.

Furthermore, it has been shown that dystrophin restoration is sustained for at least a 1 year after a single administration of AAV-CRISPR in *mdx* mice. In a study, an *Sa*Cas9 delivered in a dual AAV8 systems was designed to excise exon 23. Adult and neonate mice were treated, with intramuscular injections ($5.6x10^{11}$ vg per vector per mouse) or intravenous facial-vein injection ($5.4x10^{11}$ vg per vector per mouse) respectively. *Dmd* mRNA transcripts were analysed by ddPCR and showed deletion levels of ~8% and 2% after 8 weeks and 6 months respectively, in adult mice. While in neonate mice edit levels showed a modest statistically significant increase in genome editing over a year, particularly in the heart where there was an increase from 5% at 8 weeks post treatment to 8% at one-year post treatment in the heart. Furthermore, serum creatine kinase levels were reduced (8 weeks post treatment) in neonate mice and muscles were protected from damage by the restored dystrophin. In addition, immune response against the Cas9 was assessed. It was also shown that humoral and cellular immune response occurred in treated adult mice (resolved without intervention), while no immune response was detected in treated neonate mice (Nelson et al., 2019). Results from comparing editing efficiencies in adult vs. neonate mice, showing an increase in editing in neonates, encourages to assess the system developed in this research project in neonate *mdx* mice. An increase in editing efficiency would increase the odds of detecting our large deletion.

Deletion of exons 52 to 53 to restore the reading frame was achieved by targeting introns 51 and 53 with in mdx^{4cv} mice, generating a ~45kb in-frame deletion. In this

study, an *Sp*Cas9 system delivered by a dual AAV6 system and a multiplex dual gRNA *Sa*Cas9 delivered by a single AAV6 vector were tested by TA injections. To quantify this large deletion, deep sequencing of PCR amplicons across individual target site was used to quantify instances where on-target DNA cleavage did not excise exons 52-53, this was used to calculate editing efficiency at each target. Editing efficiencies at introns 51 and 53 respectively were 8.6% and 8.2% with the *Sp*Cas9 system and 3.5% and 2.7% with the *Sa*Cas9 system, which led to dystrophin expression levels of 0.8-18.6% (with the *Sp*Cas9 system) and of 1.5-22.9% (with the *Sa*Cas9 system), which led to significant increase in specific force generating capacity and protection from contraction-induced injury (Bengtsson et al., 2017). It is interesting that the *Sp*Cas9 system showed higher editing (approx. 5% on average) but the *Sa*Cas9 system led to slightly higher dystrophin expression levels. It is also relevant to note that modest levels of editing led to dystrophin recovery. However, the deletion performed in this study spanned ~45kb, while deletion of exons 19 to 55 would span ~800kb.

Deletion of exons 50 to 54 has been achieved with a *Sp*Cas9 system in DMD patient myoblasts (with a deletion of exons 51-53) and in hDMD/*mdx* mice (that contain a full length human *DMD* gene). In this study gRNAs were design to target exons, generate a ~160kb deletion and form a hybrid 50-54 exon. Guide RNAs were screened in HEK293T cells and optimal gRNA pairs were tested in DMD myoblasts, in which the deletion was achieved and led to the generation of hybrid exons 50-54. PCR amplicons from edited genomic DNA from the myoblasts were cloned into a plasmid (pMiniT). 45 edited clones

were analysed and 56% of them contained the expected junction of exons 50 and 54. Then, plasmids were electroporated in TAs of hDMD/*mdx* mice to assess the deletion *in-vivo*. Deletion was confirmed by PCR and 11 clones were generated. Interestingly, 64% of them showed the same repaired sequence as the one obtained *in-vitro* (Iyombe-Engembe et al., 2016).

Based on strategies implemented in other studies, our *in-vivo* experimental design could be improved by increasing the dose administered in adult *mdx* mice from $1x10^{11}$ to $1x10^{12}$ vg per TA muscle and assess later time points (rather than 1 time point two months post treatment) to allow for dystrophin accumulation, which has shown to be progressive in *mdx* mice (Nelson et al., 2016, Hakim et al., 2018, Nelson et al., 2019). We could also assess systemic delivery of our AAV9 vectors and analyse editing levels in additional tissues, as dystrophin recovery seems to vary depending on the muscle.

Furthermore, it has been shown that gRNA:Cas9 ratio also affects genome editing success. In a study using an *Sp*Cas9 system, authors observed skeletal muscle restoration after excision of exons 52-53, only with the highest administered doses of $1x10^{13}$ vp of *Sp*Cas9-AAV and $4x10^{12}$ of sg-RNA-AAV and a 2.5:1 ratio of Cas to gRNA (Bengtsson et al., 2017). Assessing Cas to gRNA ratio with our systems could be an interesting approach to assess if editing efficiency can be increased. However, our constructs express the *Sa*Cas9 and the gRNAs (either an individual gRNA or two multiplex gRNAs) from the same

plasmid, which means that a third plasmid expressing an *Sa*Cas9 would be needed to increase the Cas9 to gRNA ratio, i.e to 2.5:1 Cas to gRNA rather than a 1:1 ratio.

It must be noted that strategies that could be complementary to correcting mutations in the *DMD* gene, aiming to correct damaged muscle characteristics, are also being investigated, such as activating muscle growth by downregulation of myostatin (Kang et al., 2011, Malerba et al., 2012), reducing inflammation with novel anti-inflammatory steroids, such as Vamorolone (Kourakis et al., 2021) and reducing fibrosis with small molecules that regulate pathways involved in fibrosis activation (Levi et al., 2015, Bettica et al., 2016).

Findings in this thesis pave the way for future research exploring new forms of truncated dystrophins. Exploring the use of Del19-55 dystrophin as a mini/micro-gene holds potential, considering that our current cDNA constructs is not packable in an AAV vector and some optimisation of the sequence or assessing other delivery methods would be required. Furthermore, the achievement of such a large deletion (~800kb) with an *Sa*Cas9 CRISPR system serves as a cornerstone for genome editing strategies. The fact that this deletion did not translate *in-vivo* should lead to re-establishing the criteria used for gRNA assessment and the design of future strategies. It is relevant to highlight that the treatment showed no detrimental effect either, thus encouraging further research into the strategy.

6.2. FUTURE WORK.

Further experimental milestones for this project would be to increase gRNA cleavage efficiency. This could be attempted by screening additional gRNAs, targeting different regions within introns 18 and 55, or exploring additional targets by targeting different introns, such as 17 and 54 (the structural domain would theoretically remain the same for Del17-54 dystrophin but intron 55 could be avoided) or by targeting exons. Strategies to increasing efficiency of our current gRNAs could be implemented by modifying their structure with a "GOLD tracr" (adding a stable hairpin to our tracrRNA), by substituting on of the "Ts" from the TTTT motif at the gRNA scaffold with an A or by enhancing deletion levels with strategies such as pharmacological delay of DNA-PKcs, which has shown to increase DNA deletion levels by delaying kinetics of NHEJ relative to DSB formation and increasing the likelihood of both DSB to co-occur at both ends of the deletion (Bosch-Guiteras et al., 2021). In addition to improvements suggested to improve our *in-vivo* experimental design.

The system potency could also be improved by using engineered AAV vectors, such AAVMYO2 and AAVMYO3. These vectors were generated by a semi-rational combinatorial bioengineering approach, which consisted in *de novo* screens of two shuffled AAV capsid libraries in murine musculature, then the top hits were combined with a myotropic peptide, assessed in a previous screen of capsid variants that led to the generation of AAVMYO (Weinmann et al., 2020), and vectors were validate *in-vivo*

in two mouse strains. Variants AAVMYO2 and AAVMYO3 displaying a myotropic peptide on the capsid surface, showed increased specificity in murine skeletal muscle, diaphragm and heart and de-targeting of the liver. Furthermore, AAVMYO3 was compared to AAV9, both vectors packaging a micro-dystrophin were delivered into *mdx* mice. Results showed significant higher expression of AAVMYO3 in skeletal muscle, a more robust expression in the heart and diaphragm and a trend towards better results in strength tests (longest hanging time and four-limb grip strength) (El Andari et al., 2022).

Lasty, alternative Cas proteins could be assessed to generate deletion of exons 19 to 55, such as Cpf1, which has shown to target up to four genes simultaneously with gRNAs multiplexed in the same construct (Zetsche et al., 2017) and was used to successfully skip exon 23 in *mdx* mice (Y. Zhang et al., 2017). Furthermore, prime editing systems have been optimised to generate large precise deletions. In a recent study, an active Cas9 nuclease was conjugated to a reverse transcriptase, to create PE-Cas9. This complex was combined with two prime editing gRNAs (pegRNA), rather than the usual one pegRNA, targeting complementary DNA strands. This system introduces DSBs and incorporates desired edits using the reverse transcriptase template at the 3´extension of the pegRNAs. The two complementary edits function as homologous sequences to direct ligation and repair of the deletion junction, referred to as “PE-Cas9-based deletion and repair”. Deletions of up to 10kbp and insertions of up to 60bp were demonstrated with this system *in-vitro.* Then the system was assessed *in-vivo* and a 1.38kb deletion

was achieved, to eliminate a pathogenic insertion within the *Fah* gene in a tyrosinemia mouse model, and led to precise repair of the junction (Jiang et al., 2022). It would be interesting to assess this system in the context of DMD. The precise repair that can be induced with this system at the junction would aid to generate edits that maintain the reading frame in the *DMD* gene and avoid random insertions that could potentially lead to a stop codon and therefore a lack of truncated dystrophin expression.

6.3. CONCLUSIONS.

Ongoing research in the field of gene and cell therapies to treat Duchenne muscular dystrophies is showing promising results. Clinical trials with micro-dystrophins by Solid Biosciences (NCT03368742), Sarepta Therapeutics (NCT03375164) and Pfizer (NCT03362502) hold great potential as a therapy. However, the recent death in Pfizer´s clinical trial, assessing delivery of a micro-dystrophin with a high-dose of AAV9 vectors (Philippidis, 2022a, Philippidis, 2022b), in addition to patient deaths in Audentes´ Therapeutics trial for X-linked myotubular myopathy, heightened safety concerns related to immune responses to high AAV doses in clinic. It must be noted that adverse effects differed in these trials, patients in Audentes´ trial delivered with AAV8 vectors showed liver dysfunction and gastrointestinal bleeding (Philippidis, 2020) most likely due to liver toxicity (Nature Biotechnology, 2020), while patient in Pfizer´s trial presented hypovolemia and cardiogenic shock related to an immune response to AAV vectors (Philippidis, 2022a, Philippidis, 2022b). Strategies to overcome liver toxicity and

immune responses include improving AAV vector efficiency so lower doses are needed to achieve benefits, modifying vector capsids to evade the immune system, immunosuppress patients while the vectors are active and explore nonviral delivery methods (Bessis et al., 2004, Nelson & Gersbach, 2016).

In a recent study, gene editing with an *Sa*Cas9 system driven by a muscle specific promoter (CK8) delivered by AAV6 vectors was compared to gene addition of micro-dystrophin driven by CK8 promoter and delivered in AAV6 vectors, in CXMD dogs of 3 and 8 years of age. 6 weeks post treatment, the gene editing strategy, aimed at deleting exons 6 to 8 (105kbp), restored dystrophin reading frame in 1.3% of genomes (assessed by digital PCR) and up to 4% dystrophin transcripts (assessed by RT-PCR). It was stated that asynchronous CRISPR activity (6% editing frequency at intron 8 and 1.25% at intron 5) likely contributed to low deletion frequency. Dystrophin positive fibres numbers were greater in micro-dystrophin injected 3-year-old dogs, while differences between micro-dystrophin injected and gene edited dogs were less obvious in 8-year-old CXMD dogs. However, dystrophin expression levels and effects on muscle pathology were greater with the micro-dystrophin strategy (Bengtsson et al., 2022). This study highlights that gene editing treatment efficacy is linked to the state of muscle pathology at the time of intervention and the need for methodological optimizations related to age and disease progression of DMD to achieve potential clinical translation (Bengtsson et al., 2022). Furthermore, gene editing strategies would have to rival benefits obtained by micro-dystrophins to achieve clinical translation.

An alternative strategy being explored to circumvent the need of repeated administration from gene addition strategies is the integration of a micro-dystrophin in a safe site. In a proof-of-concept study, this was achieved by targeting the ribosomal RNA gene (rDNA) locus with TALENickases in patient derived induces pluripotent stem cells (iPSC). Mini-dystrophin expression was achieved in edited iPSC and their derived cardiomyocytes (Zeng et al., 2021). Another strategy that could circumvent the need for repeated administration and the loss of CRISPR-edits in skeletal muscle is editing of satellite cells (muscle stem cells). This has been achieved *in-vivo* in *mdx* mice with an *Sa*Cas9 system and two gRNA, designed to excise exon 23, delivered by AAV9 vectors. However efficiencies were very low (Tabebordbar, Zhu, Cheng, Chew, et al., 2016). In a different study, satellite cells were also transduced in-vivo in *mdx* mice with AAV9 and AAV8 vectors. However, editing levels of satellite cells were significantly lower than in muscle cells (approximately 0.02% of deletion levels achieved in satellite cells vs. 3% deletion levels achieved in bulk muscle genomic DNA) (Kwon et al., 2020). Satellite cell editing has proven challenging and particularly low editing efficiencies in these cells are still a challenge.

As discussed in Section 1.3.3.3. and in this general discussion, gene editing strategies have shown promising results in pre-clinical studies. However, these strategies still need to overcome some challenges, such as achieved levels of editing and avoidance of off-target events (Happi Mbakam et al., 2022). In addition to avoiding immune responses to

viral vectors (Verdera et al., 2020, Weber, 2021) or Cas proteins (Simhadri et al., 2018, Charlesworth et al., 2018, Crudele & Chamberlain, 2018).

This research project explored a gene editing strategy, with an *Sa*Cas9 system, that would possess high patient applicability by eliminating approximately 81% of mutations that lead to DMD. Our hypothesis was that a deletion between introns 18 and 55 of the *DMD* gene would generate a *de novo* intronic junction and lead to expression of a truncated but functional dystrophin protein, which was named Del19-55 dystrophin. To test our hypothesis the following research questions were investigated: Does the truncated Del19-55 dystrophin protein possess potential functionality? Can a deletion, of approximately 800 kbp, between introns 18 and 55 be achieved with an *Sa*Cas9 system? Can the in-frame deletion of exons 19 to 55 be achieved *in-vivo* with our CRISPR/Cas9 systems? Would this deletion lead to the expression of Del19-55 dystrophin in sufficient levels to see a beneficial functional effect in dystrophic muscles?

In conclusion, a deletion between introns 18 and 55 was achieved *in-vitro* by two strategies: co-delivery of individual gRNAs in plasmids expressing an *Sa*Cas9 driven by an Spc512 muscle specific promoter and by delivery of a single construct with two multiplex gRNAs and an *Sa*Cas9 driven by muscle specific Spc512 promoter. These constructs were packaged into AAV9 vectors and delivered into *mdx* mice. However, gRNA efficiency was not sufficient to achieve this deletion *in-vivo* in detectable levels, therefore it was not possible to confirm expression of Del19-55 dystrophin after DNA

repair. Nonetheless, *in-vivo* studies with our positive control cDNA plasmid expressing Del19-55 dystrophin suggest that this truncated protein could be functional, which encourages pursuit of further optimisation of our CRISPR *Sa*Cas9 system to increase editing efficiency levels and re-assess potential beneficial effects.

7. REFERENCES.

Aartsma-Rus, A., & Corey, D. R. (2020). The 10th Oligonucleotide Therapy Approved: Golodirsen for Duchenne Muscular Dystrophy. *Nucleic Acid Therapeutics*, *30*(2), 67–70. https://doi.org/10.1089/nat.2020.0845

Aartsma-Rus, A., Fokkema, I., Verschuuren, J., Ginjaar, I., Deutekom, J. van, Ommen, G.-J. van, & Dunnen, J. T. den. (2009). Theoretic applicability of antisense-mediated exon skipping for Duchenne muscular dystrophy mutations. *Human Mutation*, *30*(3), 293–299. https://doi.org/10.1002/humu.20918

Aartsma-Rus, A., & Goemans, N. (2019). A Sequel to the Eteplirsen Saga: Eteplirsen Is Approved in the United States but Was Not Approved in Europe. *Nucleic Acid Therapeutics*, *29*(1), 13–15. https://doi.org/10.1089/nat.2018.0756

Acsadi, G., Dickson, G., Love, D. R., Jani, A., Walsh, F. S., Gurusinghe, A., Wolff, J. A., & Davies, K. E. (1991). Human dystrophin expression in mdx mice after intramuscular injection of DNA constructs. *Nature*, *352*(6338), Article 6338. https://doi.org/10.1038/352815a0

Agarwal, R., Chaturvedi, S., Chhillar, N., Pant, I., & Sharma, A. (2017). Duchenne muscular dystrophy: A immunohistochemical profile and deletion pattern in dystrophin gene in North Indian population. *Asian Journal of Medical Sciences*, *8*(6), Article 6. https://doi.org/10.3126/ajms.v8i6.18281

Aiuti, A., Roncarolo, M. G., & Naldini, L. (2017). Gene therapy for ADA-SCID, the first marketing approval of an ex vivo gene therapy in Europe: Paving the road for the next generation of advanced therapy medicinal products. *EMBO Molecular Medicine*, *9*(6), 737–740. https://doi.org/10.15252/emmm.201707573

Alberts, B., Johnson, A., Lewis, J., Raff, M., Roberts, K., & Walter, P. (2002). Genesis, Modulation, and Regeneration of Skeletal Muscle. In *Molecular Biology of the Cell. 4th edition*.

Garland Science. https://www.ncbi.nlm.nih.gov/books/NBK26853/

Alfano, L. N., Charleston, J. S., Connolly, A. M., Cripe, L., Donoghue, C., Dracker, R., Dworzak, J., Eliopoulos, H., Frank, D. E., Lewis, S., Lucas, K., Lynch, J., Milici, A. J., Flynt, A., Naughton, E., Rodino-Klapac, L. R., Sahenk, Z., Schnell, F. J., Young, G. D., … Lowes, L. P. (2019). Long-term treatment with eteplirsen in nonambulatory patients with Duchenne muscular dystrophy. *Medicine*, *98*(26), e15858. https://doi.org/10.1097/MD.0000000000015858

Allen, D., Rosenberg, M., & Hendel, A. (2021). Using Synthetically Engineered Guide RNAs to Enhance CRISPR Genome Editing Systems in Mammalian Cells. *Frontiers in Genome Editing*, *2*. https://www.frontiersin.org/articles/10.3389/fgeed.2020.617910

Amann, K. J., Renley, B. A., & Ervasti, J. M. (1998). A Cluster of Basic Repeats in the Dystrophin Rod Domain Binds F-actin through an Electrostatic Interaction. *Journal of Biological Chemistry*, *273*(43), 28419–28423. https://doi.org/10.1074/jbc.273.43.28419

Amoasii, L., Hildyard, J. C. W., Li, H., Sanchez-Ortiz, E., Mireault, A., Caballero, D., Harron, R., Stathopoulou, T.-R., Massey, C., Shelton, J. M., Bassel-Duby, R., Piercy, R. J., & Olson, E. N. (2018). Gene editing restores dystrophin expression in a canine model of Duchenne muscular dystrophy. *Science (New York, N.Y.)*, *362*(6410), 86–91. https://doi.org/10.1126/science.aau1549

Anderson, E. M., Haupt, A., Schiel, J. A., Chou, E., Machado, H. B., Strezoska, Ž., Lenger, S., McClelland, S., Birmingham, A., Vermeulen, A., & Smith, A. van B. (2015). Systematic analysis of CRISPR–Cas9 mismatch tolerance reveals low levels of off-target activity. *Journal of Biotechnology*, *211*, 56–65. https://doi.org/10.1016/j.jbiotec.2015.06.427

Andersone Pauline. (2016). *FDA Declines Approval for Drisapersen in DMD*. Medscape. https://www.medscape.com/viewarticle/857406

Andrews, J. G., Lamb, M. M., Conway, K., Street, N., Westfield, C., Ciafaloni, E., Matthews,

D., Cunniff, C., Pandya, S., Fox, D. J., & MD STARnet. (2018). Diagnostic Accuracy of Phenotype Classification in Duchenne and Becker Muscular Dystrophy Using Medical Record Data1. *Journal of Neuromuscular Diseases*, *5*(4), 481–495. https://doi.org/10.3233/JND-180306

Angellotti, M. C., Bhuiyan, S. B., Chen, G., & Wan, X.-F. (2007). CodonO: Codon usage bias analysis within and across genomes. *Nucleic Acids Research*, *35*(Web Server issue), W132-136. https://doi.org/10.1093/nar/gkm392

Anzalone, A. V., Randolph, P. B., Davis, J. R., Sousa, A. A., Koblan, L. W., Levy, J. M., Chen, P. J., Wilson, C., Newby, G. A., Raguram, A., & Liu, D. R. (2019). Search-and-replace genome editing without double-strand breaks or donor DNA. *Nature*, *576*(7785), Article 7785. https://doi.org/10.1038/s41586-019-1711-4

Apolonia, L., Waddington, S. N., Fernandes, C., Ward, N. J., Bouma, G., Blundell, M. P., Thrasher, A. J., Collins, M. K., & Philpott, N. J. (2007). Stable Gene Transfer to Muscle Using Non-integrating Lentiviral Vectors. *Molecular Therapy*, *15*(11), 1947–1954. https://doi.org/10.1038/sj.mt.6300281

Arabi, F., Mansouri, V., & Ahmadbeigi, N. (2022). Gene therapy clinical trials, where do we go? An overview. *Biomedicine & Pharmacotherapy*, *153*, 113324. https://doi.org/10.1016/j.biopha.2022.113324

Athanasopoulos, T., Foster, H., Foster, K., & Dickson, G. (2011). Codon Optimization of the Microdystrophin Gene for Duchenne Muscular Dystrophy Gene Therapy. In D. Duan (Ed.), *Muscle Gene Therapy: Methods and Protocols* (pp. 21–37). Humana Press. https://doi.org/10.1007/978-1-61737-982-6_2

Balaban, B., Matthews, D. J., Clayton, G. H., & Carry, T. (2005). Corticosteroid treatment and functional improvement in Duchenne muscular dystrophy: Long-term effect. *American Journal of Physical Medicine & Rehabilitation*, *84*(11), 843–850.

https://doi.org/10.1097/01.phm.0000184156.98671.d0

Balakumar, P., Rohilla, A., & Thangathirupathi, A. (2010). Gentamicin-induced nephrotoxicity: Do we have a promising therapeutic approach to blunt it? *Pharmacological Research*, *62*(3), 179–186. https://doi.org/10.1016/j.phrs.2010.04.004

Banks, G. B., Chamberlain, J. S., & Froehner, S. C. (2009). Truncated dystrophins can influence neuromuscular synapse structure. *Molecular and Cellular Neuroscience*, *40*(4), 433–441. https://doi.org/10.1016/j.mcn.2008.12.011

Banks, G. B., Combs, A. C., Chamberlain, J. R., & Chamberlain, J. S. (2008). Molecular and cellular adaptations to chronic myotendinous strain injury in mdx mice expressing a truncated dystrophin. *Human Molecular Genetics*, *17*(24), 3975–3986. https://doi.org/10.1093/hmg/ddn301

Banks, G. B., Judge, L. M., Allen, J. M., & Chamberlain, J. S. (2010). The Polyproline Site in Hinge 2 Influences the Functional Capacity of Truncated Dystrophins. *PLOS Genetics*, *6*(5), e1000958. https://doi.org/10.1371/journal.pgen.1000958

Baumbach, L. L., Chamberlain, J. S., Ward, P. A., Farwell, N. J., & Caskey, C. T. (1989). Molecular and clinical correlations of deletions leading to Duchenne and Becker muscular dystrophies. *Neurology*, *39*(4), 465–465. https://doi.org/10.1212/WNL.39.4.465

Becker, P. E. (1962). Two families of benign sex-linked recessive muscular dystrophy. *Revue Canadienne De Biologie*, *21*, 551–566.

Becker, P. E., & Kiener, F. (1955). A new x-chromosomal muscular dystrophy. *Archiv Fur Psychiatrie Und Nervenkrankheiten, Vereinigt Mit Zeitschrift Fur Die Gesamte Neurologie Und Psychiatrie*, *193*(4), 427–448. https://doi.org/10.1007/BF00343141

Beenakker, E. A. C., Fock, J. M., Van Tol, M. J., Maurits, N. M., Koopman, H. M., Brouwer, O. F., & Van der Hoeven, J. H. (2005). Intermittent prednisone therapy in Duchenne muscular

dystrophy: A randomized controlled trial. *Archives of Neurology*, *62*(1), 128–132. https://doi.org/10.1001/archneur.62.1.128

Belanto, J. J., Mader, T. L., Eckhoff, M. D., Strandjord, D. M., Banks, G. B., Gardner, M. K., Lowe, D. A., & Ervasti, J. M. (2014). Microtubule binding distinguishes dystrophin from utrophin. *Proceedings of the National Academy of Sciences of the United States of America*, *111*(15), 5723–5728. JSTOR.

Bell, C. C., Magor, G. W., Gillinder, K. R., & Perkins, A. C. (2014). A high-throughput screening strategy for detecting CRISPR-Cas9 induced mutations using next-generation sequencing. *BMC Genomics*, *15*(1), 1002. https://doi.org/10.1186/1471-2164-15-1002

Bengtsson, N. E., Crudele, J. M., Klaiman, J. M., Halbert, C. L., Hauschka, S. D., & Chamberlain, J. S. (2022). Comparison of dystrophin expression following gene editing and gene replacement in an aged preclinical DMD animal model. *Molecular Therapy: The Journal of the American Society of Gene Therapy*, *30*(6), 2176–2185. https://doi.org/10.1016/j.ymthe.2022.02.003

Bengtsson, N. E., Hall, J. K., Odom, G. L., Phelps, M. P., Andrus, C. R., Hawkins, R. D., Hauschka, S. D., Chamberlain, J. R., & Chamberlain, J. S. (2017). Muscle-specific CRISPR/Cas9 dystrophin gene editing ameliorates pathophysiology in a mouse model for Duchenne muscular dystrophy. *Nature Communications*, *8*. https://doi.org/10.1038/ncomms14454

Béroud, C., Tuffery-Giraud, S., Matsuo, M., Hamroun, D., Humbertclaude, V., Monnier, N., Moizard, M.-P., Voelckel, M.-A., Calemard, L. M., Boisseau, P., Blayau, M., Philippe, C., Cossée, M., Pagès, M., Rivier, F., Danos, O., Garcia, L., & Claustres, M. (2007). Multiexon skipping leading to an artificial DMD protein lacking amino acids from exons 45 through 55 could rescue up to 63% of patients with Duchenne muscular dystrophy. *Human Mutation*, *28*(2), 196–202. https://doi.org/10.1002/humu.20428

Bertoni, C., Morris, G. E., & Rando, T. A. (2005). Strand bias in oligonucleotide-mediated dystrophin gene editing. *Human Molecular Genetics*, *14*(2), 221–233. https://doi.org/10.1093/hmg/ddi020

Bessis, N., GarciaCozar, F. J., & Boissier, M.-C. (2004). Immune responses to gene therapy vectors: Influence on vector function and effector mechanisms. *Gene Therapy*, *11*(1), Article 1. https://doi.org/10.1038/sj.gt.3302364

Bettica, P., Petrini, S., D'Oria, V., D'Amico, A., Catteruccia, M., Pane, M., Sivo, S., Magri, F., Brajkovic, S., Messina, S., Vita, G. L., Gatti, B., Moggio, M., Puri, P. L., Rocchetti, M., De Nicolao, G., Vita, G., Comi, G. P., Bertini, E., & Mercuri, E. (2016). Histological effects of givinostat in boys with Duchenne muscular dystrophy. *Neuromuscular Disorders*, *26*(10), 643–649. https://doi.org/10.1016/j.nmd.2016.07.002

Bhat, K. P., & Cortez, D. (2018). RPA and RAD51: Fork reversal, fork protection, and genome stability. *Nature Structural & Molecular Biology*, *25*(6), 446–453. https://doi.org/10.1038/s41594-018-0075-z

Bianchi, M. L., Mazzanti, A., Galbiati, E., Saraifoger, S., Dubini, A., Cornelio, F., & Morandi, L. (2003). Bone mineral density and bone metabolism in Duchenne muscular dystrophy. *Osteoporosis International: A Journal Established as Result of Cooperation between the European Foundation for Osteoporosis and the National Osteoporosis Foundation of the USA*, *14*(9), 761–767. https://doi.org/10.1007/s00198-003-1443-y

Bione, S., Maestrini, E., Rivella, S., Mancini, M., Regis, S., Romeo, G., & Toniolo, D. (1994). Identification of a novel X-linked gene responsible for Emery-Dreifuss muscular dystrophy. *Nature Genetics*, *8*(4), 323–327. https://doi.org/10.1038/ng1294-323

Birch, S. M., Lawlor, M. W., Conlon, T. J., Guo, L.-J., Crudele, J. M., Hawkins, E. C., Nghiem, P. P., Ahn, M., Meng, H., Beatka, M. J., Fickau, B. A., Prieto, J. C., Styner, M. A., Struharik,

M. J., Shanks, C., Brown, K. J., Golebiowski, D., Bettis, A. K., Balog-Alvarez, C. J., … Kornegay, Joe. N. (2023). Assessment of systemic AAV-microdystrophin gene therapy in the GRMD model of Duchenne muscular dystrophy. *Science Translational Medicine*, *15*(677), eabo1815. https://doi.org/10.1126/scitranslmed.abo1815

Birnkrant, D. J., Bushby, K., Bann, C. M., Alman, B. A., Apkon, S. D., Blackwell, A., Case, L. E., Cripe, L., Hadjiyannakis, S., Olson, A. K., Sheehan, D. W., Bolen, J., Weber, D. R., Ward, L. M., & DMD Care Considerations Working Group. (2018). Diagnosis and management of Duchenne muscular dystrophy, part 2: Respiratory, cardiac, bone health, and orthopaedic management. *The Lancet. Neurology*, *17*(4), 347–361. https://doi.org/10.1016/S1474-4422(18)30025-5

Birnkrant, D. J., Bushby, K., Bann, C. M., Apkon, S. D., Blackwell, A., Brumbaugh, D., Case, L. E., Clemens, P. R., Hadjiyannakis, S., Pandya, S., Street, N., Tomezsko, J., Wagner, K. R., Ward, L. M., Weber, D. R., & DMD Care Considerations Working Group. (2018). Diagnosis and management of Duchenne muscular dystrophy, part 1: Diagnosis, and neuromuscular, rehabilitation, endocrine, and gastrointestinal and nutritional management. *The Lancet. Neurology*, *17*(3), 251–267. https://doi.org/10.1016/S1474-4422(18)30024-3

Birnkrant, D. J., Bushby, K., Bann, C. M., Apkon, S. D., Blackwell, A., Colvin, M. K., Cripe, L., Herron, A. R., Kennedy, A., Kinnett, K., Naprawa, J., Noritz, G., Poysky, J., Street, N., Trout, C. J., Weber, D. R., Ward, L. M., & DMD Care Considerations Working Group. (2018). Diagnosis and management of Duchenne muscular dystrophy, part 3: Primary care, emergency management, psychosocial care, and transitions of care across the lifespan. *The Lancet. Neurology*, *17*(5), 445–455. https://doi.org/10.1016/S1474-4422(18)30026-7

Bladen, C. L., Salgado, D., Monges, S., Foncuberta, M. E., Kekou, K., Kosma, K., Dawkins, H., Lamont, L., Roy, A. J., Chamova, T., Guergueltcheva, V., Chan, S., Korngut, L., Campbell, C., Dai, Y., Wang, J., Barišić, N., Brabec, P., Lahdetie, J., … Lochmüller, H. (2015). The TREAT-

NMD DMD Global Database: Analysis of More than 7,000 Duchenne Muscular Dystrophy Mutations. *Human Mutation*, *36*(4), 395–402. https://doi.org/10.1002/humu.22758

Bo, W., Y, L., Pa, M., Tj, D., P, L., & Ql, L. (2009). Octa-guanidine morpholino restores dystrophin expression in cardiac and skeletal muscles and ameliorates pathology in dystrophic mdx mice. *Molecular Therapy : The Journal of the American Society of Gene Therapy*, *17*(5). https://doi.org/10.1038/mt.2009.38

Boch, J., Scholze, H., Schornack, S., Landgraf, A., Hahn, S., Kay, S., Lahaye, T., Nickstadt, A., & Bonas, U. (2009). Breaking the Code of DNA Binding Specificity of TAL-Type III Effectors. *Science*, *326*(5959), 1509–1512. https://doi.org/10.1126/science.1178811

Boel, A., De Saffel, H., Steyaert, W., Callewaert, B., De Paepe, A., Coucke, P. J., & Willaert, A. (2018). CRISPR/Cas9-mediated homology-directed repair by ssODNs in zebrafish induces complex mutational patterns resulting from genomic integration of repair-template fragments. *Disease Models & Mechanisms*, *11*(10). https://doi.org/10.1242/dmm.035352

Bosch-Guiteras, N., Uroda, T., Guillen-Ramirez, H. A., Riedo, R., Gazdhar, A., Esposito, R., Pulido-Quetglas, C., Zimmer, Y., Medová, M., & Johnson, R. (2021). Enhancing CRISPR deletion via pharmacological delay of DNA-PKcs. *Genome Research*, *31*(3), 461–471. https://doi.org/10.1101/gr.265736.120

Bourdon, A., François, V., Zhang, L., Lafoux, A., Fraysse, B., Toumaniantz, G., Larcher, T., Girard, T., Ledevin, M., Lebreton, C., Hivonnait, A., Creismeas, A., Allais, M., Marie, B., Guguin, J., Blouin, V., Remy, S., Anegon, I., Huchet, C., … Le Guiner, C. (2022). Evaluation of the dystrophin carboxy-terminal domain for micro-dystrophin gene therapy in cardiac and skeletal muscles in the DMDmdx rat model. *Gene Therapy*, *29*(9), Article 9. https://doi.org/10.1038/s41434-022-00317-6

Bowles, D. E., McPhee, S. W. J., Li, C., Gray, S. J., Samulski, J. J., Camp, A. S., Li, J., Wang,

B., Monahan, P. E., Rabinowitz, J. E., Grieger, J. C., Govindasamy, L., Agbandje-McKenna, M., Xiao, X., & Samulski, R. J. (2012). Phase 1 gene therapy for Duchenne muscular dystrophy using a translational optimized AAV vector. *Molecular Therapy: The Journal of the American Society of Gene Therapy*, *20*(2), 443–455. https://doi.org/10.1038/mt.2011.237

Brais, B., Bouchard, J. P., Xie, Y. G., Rochefort, D. L., Chrétien, N., Tomé, F. M., Lafrenière, R. G., Rommens, J. M., Uyama, E., Nohira, O., Blumen, S., Korczyn, A. D., Heutink, P., Mathieu, J., Duranceau, A., Codère, F., Fardeau, M., & Rouleau, G. A. (1998). Short GCG expansions in the PABP2 gene cause oculopharyngeal muscular dystrophy. *Nature Genetics*, *18*(2), 164–167. https://doi.org/10.1038/ng0298-164

Brais, B., Rouleau, G. A., Bouchard, J. P., Fardeau, M., & Tomé, F. M. (1999). Oculopharyngeal muscular dystrophy. *Seminars in Neurology*, *19*(1), 59–66. https://doi.org/10.1055/s-2008-1040826

Bravo, J. P. K., Liu, M.-S., Hibshman, G. N., Dangerfield, T. L., Jung, K., McCool, R. S., Johnson, K. A., & Taylor, D. W. (2022). Structural basis for mismatch surveillance by CRISPR–Cas9. *Nature*, *603*(7900), 343–347. https://doi.org/10.1038/s41586-022-04470-1

Brennan, T. A., & Wilson, J. M. (2014). The special case of gene therapy pricing. *Nature Biotechnology*, *32*(9), 874–876. https://doi.org/10.1038/nbt.3003

Brinkman, E. K., Chen, T., Amendola, M., & van Steensel, B. (2014). Easy quantitative assessment of genome editing by sequence trace decomposition. *Nucleic Acids Research*, *42*(22), e168. https://doi.org/10.1093/nar/gku936

Broglio, L., Tentorio, M., Cotelli, M. S., Mancuso, M., Vielmi, V., Gregorelli, V., Padovani, A., & Filosto, M. (2010). Limb-girdle muscular dystrophy-associated protein diseases. *The Neurologist*, *16*(6), 340–352. https://doi.org/10.1097/NRL.0b013e3181d35b39

Brolin, C., & Shiraishi, T. (2011). Antisense mediated exon skipping therapy for duchenne

muscular dystrophy (DMD). *Artificial DNA, PNA & XNA*, *2*(1), 6–15. https://doi.org/10.4161/adna.2.1.15425

Bushby, K., Finkel, R., Birnkrant, D. J., Case, L. E., Clemens, P. R., Cripe, L., Kaul, A., Kinnett, K., McDonald, C., Pandya, S., Poysky, J., Shapiro, F., Tomezsko, J., Constantin, C., & DMD Care Considerations Working Group. (2010). Diagnosis and management of Duchenne muscular dystrophy, part 2: Implementation of multidisciplinary care. *The Lancet. Neurology*, *9*(2), 177–189. https://doi.org/10.1016/S1474-4422(09)70272-8

Bushby, K., Finkel, R., Wong, B., Barohn, R., Campbell, C., Comi, G. P., Connolly, A. M., Day, J. W., Flanigan, K. M., Goemans, N., Jones, K. J., Mercuri, E., Quinlivan, R., Renfroe, J. B., Russman, B., Ryan, M. M., Tulinius, M., Voit, T., Moore, S. A., … Group, F. the P.-G.-007-D. S. (2014). Ataluren treatment of patients with nonsense mutation dystrophinopathy. *Muscle & Nerve*, *50*(4), 477–487. https://doi.org/10.1002/mus.24332

Bushby, K. M. (1999). Making sense of the limb-girdle muscular dystrophies. *Brain: A Journal of Neurology*, *122 (Pt 8)*, 1403–1420. https://doi.org/10.1093/brain/122.8.1403

Bushby, K. M. D. (1999). The Limb-Girdle Muscular Dystrophies—Multiple Genes, Multiple Mechanisms. *Human Molecular Genetics*, *8*(10), 1875–1882. https://doi.org/10.1093/hmg/8.10.1875

Byers, T. J., Lidov, H. G., & Kunkel, L. M. (1993). An alternative dystrophin transcript specific to peripheral nerve. *Nature Genetics*, *4*(1), 77–81. https://doi.org/10.1038/ng0593-77

Canver, M. C., Bauer, D. E., Dass, A., Yien, Y. Y., Chung, J., Masuda, T., Maeda, T., Paw, B. H., & Orkin, S. H. (2014). Characterization of genomic deletion efficiency mediated by clustered regularly interspaced short palindromic repeats (CRISPR)/Cas9 nuclease system in mammalian cells. *The Journal of Biological Chemistry*, *289*(31), 21312–21324. https://doi.org/10.1074/jbc.M114.564625

Capecchi, M. R. (1989). Altering the Genome by Homologous Recombination. *Science*, *244*(4910), 1288–1292. https://doi.org/10.1126/science.2660260

Carter, B. J. (1992). Adeno-associated virus vectors. *Current Opinion in Biotechnology*, *3*(5), 533–539. https://doi.org/10.1016/0958-1669(92)90082-t

Castanotto, D., & Rossi, J. J. (2009). The promises and pitfalls of RNA-interference-based therapeutics. *Nature*, *457*(7228), Article 7228. https://doi.org/10.1038/nature07758

Cervelli, T., Palacios, J. A., Zentilin, L., Mano, M., Schwartz, R. A., Weitzman, M. D., & Giacca, M. (2008). Processing of recombinant AAV genomes occurs in specific nuclear structures that overlap with foci of DNA-damage-response proteins. *Journal of Cell Science*, *121*(Pt 3), 349–357. https://doi.org/10.1242/jcs.003632

Chamberlain, J. S., Metzger, J., Reyes, M., Townsend, D., & Faulkner, J. A. (2007). Dystrophin-deficient mdx mice display a reduced life span and are susceptible to spontaneous rhabdomyosarcoma. *The FASEB Journal*, *21*(9), 2195–2204. https://doi.org/10.1096/fj.06-7353com

Chapdelaine, P., Pichavant, C., Rousseau, J., Pâques, F., & Tremblay, J. P. (2010). Meganucleases can restore the reading frame of a mutated dystrophin. *Gene Therapy*, *17*(7), 846–858. https://doi.org/10.1038/gt.2010.26

Chapman, M. S., & Agbandje-McKenna, M. (2005). *Atomic structure of viral particles. In: Parvoviruses.* CRC Press.

Charlesworth, C. T., Deshpande, P. S., Dever, D. P., Dejene, B., Gomez-Ospina, N., Mantri, S., Pavel-Dinu, M., Camarena, J., Weinberg, K. I., & Porteus, M. H. (2018). *Identification of Pre-Existing Adaptive Immunity to Cas9 Proteins in Humans* (p. 243345). bioRxiv. https://doi.org/10.1101/243345

Chemello, F., Chai, A. C., Li, H., Rodriguez-Caycedo, C., Sanchez-Ortiz, E., Atmanli, A.,

Mireault, A. A., Liu, N., Bassel-Duby, R., & Olson, E. N. (2021). Precise correction of Duchenne muscular dystrophy exon deletion mutations by base and prime editing. *Science Advances*, *7*(18), eabg4910. https://doi.org/10.1126/sciadv.abg4910

Chen, B., Gilbert, L. A., Cimini, B. A., Schnitzbauer, J., Zhang, W., Li, G.-W., Park, J., Blackburn, E. H., Weissman, J. S., Qi, L. S., & Huang, B. (2013). Dynamic imaging of genomic loci in living human cells by an optimized CRISPR/Cas system. *Cell*, *155*(7), 1479–1491. https://doi.org/10.1016/j.cell.2013.12.001

Chen, B., Hu, J., Almeida, R., Liu, H., Balakrishnan, S., Covill-Cooke, C., Lim, W. A., & Huang, B. (2016). Expanding the CRISPR imaging toolset with Staphylococcus aureus Cas9 for simultaneous imaging of multiple genomic loci. *Nucleic Acids Research*, *44*(8), e75. https://doi.org/10.1093/nar/gkv1533

Chew, W. L. (2018). Immunity to CRISPR Cas9 and Cas12a therapeutics. *Wiley Interdisciplinary Reviews: Systems Biology and Medicine*, *10*(1). Scopus. https://doi.org/10.1002/wsbm.1408

Choi, I.-K., & Yun, C.-O. (2013). Recent developments in oncolytic adenovirus-based immunotherapeutic agents for use against metastatic cancers. *Cancer Gene Therapy*, *20*(2), Article 2. https://doi.org/10.1038/cgt.2012.95

Cirak, S., Arechavala-Gomeza, V., Guglieri, M., Feng, L., Torelli, S., Anthony, K., Abbs, S., Garralda, M. E., Bourke, J., Wells, D. J., Dickson, G., Wood, M. J. A., Wilton, S. D., Straub, V., Kole, R., Shrewsbury, S. B., Sewry, C., Morgan, J. E., Bushby, K., & Muntoni, F. (2011). Exon skipping and dystrophin restoration in patients with Duchenne muscular dystrophy after systemic phosphorodiamidate morpholino oligomer treatment: An open-label, phase 2, dose-escalation study. *Lancet (London, England)*, *378*(9791), 595–605. https://doi.org/10.1016/S0140-6736(11)60756-3

Clemens, P. R., Rao, V. K., Connolly, A. M., Harper, A. D., Mah, J. K., McDonald, C. M., Smith, E. C., Zaidman, C. M., Nakagawa, T., CINRG DNHS Investigators, & Hoffman, E. P. (2022). Long-Term Functional Efficacy and Safety of Viltolarsen in Patients with Duchenne Muscular Dystrophy. *Journal of Neuromuscular Diseases*, *9*(4), 493–501. https://doi.org/10.3233/JND-220811

Clemens, P. R., Rao, V. K., Connolly, A. M., Harper, A. D., Mah, J. K., Smith, E. C., McDonald, C. M., Zaidman, C. M., Morgenroth, L. P., Osaki, H., Satou, Y., Yamashita, T., & Hoffman, E. P. (2020). Safety, Tolerability, and Efficacy of Viltolarsen in Boys With Duchenne Muscular Dystrophy Amenable to Exon 53 Skipping. *JAMA Neurology*, *77*(8), 1–10. https://doi.org/10.1001/jamaneurol.2020.1264

Clément, N., & Grieger, J. C. (2016). Manufacturing of recombinant adeno-associated viral vectors for clinical trials. *Molecular Therapy. Methods & Clinical Development*, *3*, 16002. https://doi.org/10.1038/mtm.2016.2

Cole-Strauss, A., Yoon, K., Xiang, Y., Byrne, B. C., Rice, M. C., Gryn, J., Holloman, W. K., & Kmiec, E. B. (1996). Correction of the mutation responsible for sickle cell anemia by an RNA-DNA oligonucleotide. *Science (New York, N.Y.)*, *273*(5280), 1386–1389. https://doi.org/10.1126/science.273.5280.1386

Collins, C. A., & Partridge, T. A. (2005). Self-Renewal of the Adult Skeletal Muscle Satellite Cell. *Cell Cycle*, *4*(10), 1338–1341. https://doi.org/10.4161/cc.4.10.2114

Concordet, J.-P., & Haeussler, M. (2018). CRISPOR: Intuitive guide selection for CRISPR/Cas9 genome editing experiments and screens. *Nucleic Acids Research*, *46*(W1), W242–W245. https://doi.org/10.1093/nar/gky354

Conlin, M. P., Reid, D. A., Small, G. W., Chang, H. H., Watanabe, G., Lieber, M. R., Ramsden, D. A., & Rothenberg, E. (2017). DNA Ligase IV Guides End-Processing Choice during

Nonhomologous End Joining. *Cell Reports*, *20*(12), 2810–2819. https://doi.org/10.1016/j.celrep.2017.08.091

Cossu, G., Previtali, S. C., Napolitano, S., Cicalese, M. P., Tedesco, F. S., Nicastro, F., Noviello, M., Roostalu, U., Natali Sora, M. G., Scarlato, M., De Pellegrin, M., Godi, C., Giuliani, S., Ciotti, F., Tonlorenzi, R., Lorenzetti, I., Rivellini, C., Benedetti, S., Gatti, R., … Ciceri, F. (2015). Intra-arterial transplantation of HLA-matched donor mesoangioblasts in Duchenne muscular dystrophy. *EMBO Molecular Medicine*, *7*(12), 1513–1528. https://doi.org/10.15252/emmm.201505636

Cox, D. B. T., Platt, R. J., & Zhang, F. (2015). Therapeutic genome editing: Prospects and challenges. *Nature Medicine*, *21*(2), Article 2. https://doi.org/10.1038/nm.3793

CRISPR gRNA Design Tool | Benchling. (n.d.). Retrieved 24 February 2023, from https://www.benchling.com/crispr

Crudele, J. M., & Chamberlain, J. S. (2018). Cas9 immunity creates challenges for CRISPR gene editing therapies. *Nature Communications*, *9*(1), 3497. https://doi.org/10.1038/s41467-018-05843-9

Daley, J. M., Jimenez-Sainz, J., Wang, W., Miller, A. S., Xue, X., Nguyen, K. A., Jensen, R. B., & Sung, P. (2017). Enhancement of BLM-DNA2-Mediated Long-Range DNA End Resection by CtIP. *Cell Reports*, *21*(2), 324–332. https://doi.org/10.1016/j.celrep.2017.09.048

Dang, Y., Jia, G., Choi, J., Ma, H., Anaya, E., Ye, C., Shankar, P., & Wu, H. (2015). Optimizing sgRNA structure to improve CRISPR-Cas9 knockout efficiency. *Genome Biology*, *16*(1), 280. https://doi.org/10.1186/s13059-015-0846-3

Darras, B. T., Menache-Starobinski, C. C., Hinton, V., & Kunkel, L. M. (2015). Chapter 30—Dystrophinopathies. In B. T. Darras, H. R. Jones, M. M. Ryan, & D. C. De Vivo (Eds.), *Neuromuscular Disorders of Infancy, Childhood, and Adolescence (Second Edition)* (pp. 551–

592). Academic Press. https://doi.org/10.1016/B978-0-12-417044-5.00030-5

Dastur, R. S., Gaitonde, P. S., Khadilkar, S. V., & Nadkarni, J. J. (2008). Becker muscular dystrophy in Indian patients: Analysis of dystrophin gene deletion patterns. *Neurology India*, *56*(3), 374–378. https://doi.org/10.4103/0028-3886.40961

Davis, A. J., Chen, B. P. C., & Chen, D. J. (2014). DNA-PK: A dynamic enzyme in a versatile DSB repair pathway. *DNA Repair*, *17*, 21–29. https://doi.org/10.1016/j.dnarep.2014.02.020

Davison, A. J., Benkő, M., & Harrach, B. (2003). Genetic content and evolution of adenoviruses. *The Journal of General Virology*, *84*(Pt 11), 2895–2908. https://doi.org/10.1099/vir.0.19497-0

D'Costa, S., Blouin, V., Broucque, F., Penaud-Budloo, M., François, A., Perez, I. C., Le Bec, C., Moullier, P., Snyder, R. O., & Ayuso, E. (2016). Practical utilization of recombinant AAV vector reference standards: Focus on vector genomes titration by free ITR qPCR. *Molecular Therapy. Methods & Clinical Development*, *5*, 16019. https://doi.org/10.1038/mtm.2016.19

Dellavalle, A., Maroli, G., Covarello, D., Azzoni, E., Innocenzi, A., Perani, L., Antonini, S., Sambasivan, R., Brunelli, S., Tajbakhsh, S., & Cossu, G. (2011). Pericytes resident in postnatal skeletal muscle differentiate into muscle fibres and generate satellite cells. *Nature Communications*, *2*(1), Article 1. https://doi.org/10.1038/ncomms1508

Devkota, S. (2018). The road less traveled: Strategies to enhance the frequency of homology-directed repair (HDR) for increased efficiency of CRISPR/Cas-mediated transgenesis. *BMB Reports*, *51*(9), 437–443. https://doi.org/10.5483/BMBRep.2018.51.9.187

Dhatterwal, P., Mehrotra, S., & Mehrotra, R. (2017). Optimization of PCR conditions for amplifying an AT-rich amino acid transporter promoter sequence with high number of tandem repeats from Arabidopsis thaliana. *BMC Research Notes*, *10*. https://doi.org/10.1186/s13104-017-2982-1

Doench, J. G., Fusi, N., Sullender, M., Hegde, M., Vaimberg, E. W., Donovan, K. F., Smith, I., Tothova, Z., Wilen, C., Orchard, R., Virgin, H. W., Listgarten, J., & Root, D. E. (2016). Optimized sgRNA design to maximize activity and minimize off-target effects of CRISPR-Cas9. *Nature Biotechnology*, *34*(2), Article 2. https://doi.org/10.1038/nbt.3437

Doench, J. G., Hartenian, E., Graham, D. B., Tothova, Z., Hegde, M., Smith, I., Sullender, M., Ebert, B. L., Xavier, R. J., & Root, D. E. (2014). Rational design of highly active sgRNAs for CRISPR-Cas9–mediated gene inactivation. *Nature Biotechnology*, *32*(12), 1262–1267. https://doi.org/10.1038/nbt.3026

Doetschman, T., Gregg, R. G., Maeda, N., Hooper, M. L., Melton, D. W., Thompson, S., & Smithies, O. (1987). Targetted correction of a mutant HPRT gene in mouse embryonic stem cells. *Nature*, *330*(6148), Article 6148. https://doi.org/10.1038/330576a0

Dong, J.-Y., Fan, P.-D., & Frizzell, R. A. (1996). Quantitative Analysis of the Packaging Capacity of Recombinant Adeno-Associated Virus. *Human Gene Therapy*, *7*(17), 2101–2112. https://doi.org/10.1089/hum.1996.7.17-2101

Dostert, A., & Heinzel, T. (2004). Negative glucocorticoid receptor response elements and their role in glucocorticoid action. *Current Pharmaceutical Design*, *10*(23), 2807–2816. https://doi.org/10.2174/1381612043383601

Dreghici, R. D., Redican, S., Lawrence, J., Brown, K., Wang, F., Gonzalez, J., Schneider, J., Morris, C., Shieh, P., & Byrne, B. (2022). FP.28 IGNITE DMD phase I/II study of SGT-001 microdystrophin gene therapy for DMD: Long-term outcomes and expression update. *Neuromuscular Disorders*, *32*, S98. https://doi.org/10.1016/j.nmd.2022.07.234

Duan, D. (2018). Systemic AAV Micro-dystrophin Gene Therapy for Duchenne Muscular Dystrophy. *Molecular Therapy*, *26*(10), 2337–2356. https://doi.org/10.1016/j.ymthe.2018.07.011

Duan, D., Goemans, N., Takeda, S., Mercuri, E., & Aartsma-Rus, A. (2021). Duchenne muscular dystrophy. *Nature Reviews. Disease Primers*, *7*(1), 13. https://doi.org/10.1038/s41572-021-00248-3

Duchêne, B. L., Cherif, K., Iyombe-Engembe, J.-P., Guyon, A., Rousseau, J., Ouellet, D. L., Barbeau, X., Lague, P., & Tremblay, J. P. (2018). CRISPR-Induced Deletion with SaCas9 Restores Dystrophin Expression in Dystrophic Models In Vitro and In Vivo. *Molecular Therapy*, *26*(11), 2604–2616. https://doi.org/10.1016/j.ymthe.2018.08.010

Dull, T., Zufferey, R., Kelly, M., Mandel, R. J., Nguyen, M., Trono, D., & Naldini, L. (1998). A Third-Generation Lentivirus Vector with a Conditional Packaging System. *Journal of Virology*, *72*(11), 8463–8471. https://doi.org/10.1128/JVI.72.11.8463-8471.1998

Egorova, T. V., Zotova, E. D., Reshetov, D. A., Polikarpova, A. V., Vassilieva, S. G., Vlodavets, D. V., Gavrilov, A. A., Ulianov, S. V., Buchman, V. L., & Deykin, A. V. (2019). CRISPR/Cas9-generated mouse model of Duchenne muscular dystrophy recapitulating a newly identified large 430 kb deletion in the human DMD gene. *Disease Models & Mechanisms*, *12*(4), dmm037655. https://doi.org/10.1242/dmm.037655

El Andari, J., Renaud-Gabardos, E., Tulalamba, W., Weinmann, J., Mangin, L., Pham, Q. H., Hille, S., Bennett, A., Attebi, E., Bourges, E., Leborgne, C., Guerchet, N., Fakhiri, J., Krämer, C., Wiedtke, E., McKenna, R., Guianvarc'h, L., Toueille, M., Ronzitti, G., … Grimm, D. (2022). Semirational bioengineering of AAV vectors with increased potency and specificity for systemic gene therapy of muscle disorders. *Science Advances*, *8*(38), eabn4704. https://doi.org/10.1126/sciadv.abn4704

El Refaey, M., Xu, L., Gao, Y., Canan, B. D., Adesanya, T. M. A., Warner, S. C., Akagi, K., Symer, D. E., Mohler, P. J., Ma, J., Janssen, P. M. L., & Han, R. (2017). In Vivo Genome Editing Restores Dystrophin Expression and Cardiac Function in Dystrophic Mice. *Circulation Research*, *121*(8), 923–929. https://doi.org/10.1161/CIRCRESAHA.117.310996

EMA. (2018, September 17). *Advanced therapy medicinal products: Overview* [Text]. European Medicines Agency. https://www.ema.europa.eu/en/human-regulatory/overview/advanced-therapy-medicinal-products-overview

EMA. (2020a, October 13). *Libmeldy* [Text]. European Medicines Agency. https://www.ema.europa.eu/en/medicines/human/EPAR/libmeldy

EMA. (2020b, October 13). *Tecartus* [Text]. European Medicines Agency. https://www.ema.europa.eu/en/medicines/human/EPAR/tecartus

EMA. (2021a, January 4). *Spikevax (previously COVID-19 Vaccine Moderna)* [Text]. European Medicines Agency. https://www.ema.europa.eu/en/medicines/human/EPAR/spikevax

EMA. (2021b, June 23). *Abecma* [Text]. European Medicines Agency. https://www.ema.europa.eu/en/medicines/human/EPAR/abecma

Emery, A. E. (1989). Emery-Dreifuss syndrome. *Journal of Medical Genetics*, *26*(10), 637–641. https://doi.org/10.1136/jmg.26.10.637

Emery, A. E. H. (2002). *The muscular dystrophies*. *359*(9307), 687–695. https://doi.org/10.1016/S0140-6736(02)07815-7

Emery, A. E. H., Muntoni, F., & Quinlivan, R. (2015). *Duchenne Muscular Dystrophy*. Oxford University Press.

England, S. B., Nicholson, L. V., Johnson, M. A., Forrest, S. M., Love, D. R., Zubrzycka-Gaarn, E. E., Bulman, D. E., Harris, J. B., & Davies, K. E. (1990). Very mild muscular dystrophy associated with the deletion of 46% of dystrophin. *Nature*, *343*(6254), 180–182. https://doi.org/10.1038/343180a0

Eren, S. A., Tastan, C., Karadeniz, K. B., Turan, R. D., Cakirsoy, D., Kancagi, D. D., Yilmaz, S. U., Oztatlici, M., Oztatlici, H., Ozer, S., Tumentemur, G., Baykal, A. T., & Ovali, E. (2023).

Lentiviral micro-dystrophin gene treatment into late-stage mdx mice for Duchene Muscular Dystrophy disease. *Current Gene Therapy*. https://doi.org/10.2174/1566523223666230407091317

ExPASy—Translate tool. (n.d.). Retrieved 25 February 2023, from https://web.expasy.org/translate/

Fanin, M., Freda, M. P., Vitiello, L., Danieli, G. A., Pegoraro, E., & Angelini, C. (1996). Duchenne phenotype with in-frame deletion removing major portion of dystrophin rod: Threshold effect for deletion size? *Muscle & Nerve*, *19*(9), 1154–1160. https://doi.org/10.1002/mus.880190902

FDA. (1993). *Application of current statutory authorities to human somatic cell therapy products and gene therapy products*. Vol. 58. Federal Register 53248–53251. https://fda.report/media/76647/Application-of-Current-Statuatory-Authorities-to-Human-Somatic-Cell-Therapy-Products-and-Gene-Therapy-Products.pdf

FDA. (2018). What is Gene Therapy? *FDA*. https://www.fda.gov/vaccines-blood-biologics/cellular-gene-therapy-products/what-gene-therapy

FDA. (2021a). FDA approves brexucabtagene autoleucel for relapsed or refractory B-cell precursor acute lymphoblastic leukemia. *FDA*. https://www.fda.gov/drugs/resources-information-approved-drugs/fda-approves-brexucabtagene-autoleucel-relapsed-or-refractory-b-cell-precursor-acute-lymphoblastic

FDA. (2021b). FDA approves idecabtagene vicleucel for multiple myeloma. *FDA*. https://www.fda.gov/drugs/resources-information-approved-drugs/fda-approves-idecabtagene-vicleucel-multiple-myeloma

FDA. (2021c). FDA approves lisocabtagene maraleucel for relapsed or refractory large B-cell lymphoma. *FDA*. https://www.fda.gov/drugs/resources-information-approved-drugs/fda-

approves-lisocabtagene-maraleucel-relapsed-or-refractory-large-b-cell-lymphoma

FDA. (2023a). CARVYKTI. *FDA*. https://www.fda.gov/vaccines-blood-biologics/carvykti

FDA. (2023b). SPIKEVAX. *FDA*. https://www.fda.gov/vaccines-blood-biologics/spikevax

FDA. (2019). *FDA grants accelerated approval to first targeted treatment for rare Duchenne muscular dystrophy mutation*. FDA; FDA. https://www.fda.gov/news-events/press-announcements/fda-grants-accelerated-approval-first-targeted-treatment-rare-duchenne-muscular-dystrophy-mutation

FDA. (2020). *FDA Approves Targeted Treatment for Rare Duchenne Muscular Dystrophy Mutation*. FDA; FDA. https://www.fda.gov/news-events/press-announcements/fda-approves-targeted-treatment-rare-duchenne-muscular-dystrophy-mutation

FDA. (2021d, February 25). *FDA Approves Targeted Treatment for Rare Duchenne Muscular Dystrophy Mutation*. FDA; FDA. https://www.fda.gov/news-events/press-announcements/fda-approves-targeted-treatment-rare-duchenne-muscular-dystrophy-mutation-0

Fell, V. L., & Schild-Poulter, C. (2015). The Ku heterodimer: Function in DNA repair and beyond. *Mutation Research. Reviews in Mutation Research*, *763*, 15–29. https://doi.org/10.1016/j.mrrev.2014.06.002

Fishman-Lobell, J., Rudin, N., & Haber, J. E. (1992). Two alternative pathways of double-strand break repair that are kinetically separable and independently modulated. *Molecular and Cellular Biology*, *12*(3), 1292–1303.

Fletcher, S., Honeyman, K., Fall, A. M., Harding, P. L., Johnsen, R. D., Steinhaus, J. P., Moulton, H. M., Iversen, P. L., & Wilton, S. D. (2007). Morpholino Oligomer–Mediated Exon Skipping Averts the Onset of Dystrophic Pathology in the mdx Mouse. *Molecular Therapy*, *15*(9), 1587–1592. https://doi.org/10.1038/sj.mt.6300245

Folger, K. R., Thomas, K., & Capecchi, M. R. (1985). Nonreciprocal exchanges of information between DNA duplexes coinjected into mammalian cell nuclei. *Molecular and Cellular Biology*, *5*(1), 59–69. https://doi.org/10.1128/mcb.5.1.59-69.1985

Foster, H., Sharp, P. S., Athanasopoulos, T., Trollet, C., Graham, I. R., Foster, K., Wells, D. J., & Dickson, G. (2008). Codon and mRNA sequence optimization of microdystrophin transgenes improves expression and physiological outcome in dystrophic mdx mice following AAV2/8 gene transfer. *Molecular Therapy: The Journal of the American Society of Gene Therapy*, *16*(11), 1825–1832. https://doi.org/10.1038/mt.2008.186

Frangoul, H., Altshuler, D., Cappellini, M. D., Chen, Y.-S., Domm, J., Eustace, B. K., Foell, J., de la Fuente, J., Grupp, S., Handgretinger, R., Ho, T. W., Kattamis, A., Kernytsky, A., Lekstrom-Himes, J., Li, A. M., Locatelli, F., Mapara, M. Y., de Montalembert, M., Rondelli, D., … Corbacioglu, S. (2021). CRISPR-Cas9 Gene Editing for Sickle Cell Disease and β-Thalassemia. *New England Journal of Medicine*, *384*(3), 252–260. https://doi.org/10.1056/NEJMoa2031054

Frank, D. E., Schnell, F. J., Akana, C., El-Husayni, S. H., Desjardins, C. A., Morgan, J., Charleston, J. S., Sardone, V., Domingos, J., Dickson, G., Straub, V., Guglieri, M., Mercuri, E., Servais, L., Muntoni, F., & SKIP-NMD Study Group. (2020). Increased dystrophin production with golodirsen in patients with Duchenne muscular dystrophy. *Neurology*, *94*(21), e2270–e2282. https://doi.org/10.1212/WNL.0000000000009233

Frels, W. I., Bluestone, J. A., Hodes, R. J., Capecchi, M. R., & Singer, D. S. (1985). Expression of a microinjected porcine class I major histocompatibility complex gene in transgenic mice. *Science (New York, N.Y.)*, *228*(4699), 577–580. https://doi.org/10.1126/science.3885396

Fridovich-Keil, Judith L. (2019). *Gene editing | Definition, History, & CRISPR-Cas9*. Encyclopedia Britannica. https://www.britannica.com/science/gene-editing

Friedmann, T., & Roblin, R. (1972). Gene Therapy for Human Genetic Disease? *Science*, *175*(4025), 949–955. https://doi.org/10.1126/science.175.4025.949

Fu, B. X. H., Hansen, L. L., Artiles, K. L., Nonet, M. L., & Fire, A. Z. (2014). Landscape of target:guide homology effects on Cas9-mediated cleavage. *Nucleic Acids Research*, *42*(22), 13778–13787. https://doi.org/10.1093/nar/gku1102

Gaj, T., Gersbach, C. A., & Barbas, C. F. (2013). ZFN, TALEN, and CRISPR/Cas-based methods for genome engineering. *Trends in Biotechnology*, *31*(7), 397–405. https://doi.org/10.1016/j.tibtech.2013.04.004

Galbiati, F., Razani, B., & Lisanti, M. P. (2001). Caveolae and caveolin-3 in muscular dystrophy. *Trends in Molecular Medicine*, *7*(10), 435–441. https://doi.org/10.1016/s1471-4914(01)02105-0

Gales, L. (2019). Tegsedi (Inotersen): An Antisense Oligonucleotide Approved for the Treatment of Adult Patients with Hereditary Transthyretin Amyloidosis. *Pharmaceuticals (Basel, Switzerland)*, *12*(2), 78. https://doi.org/10.3390/ph12020078

Gao, G., Vandenberghe, L. H., & Wilson, J. M. (2005). New recombinant serotypes of AAV vectors. *Current Gene Therapy*, *5*(3), 285–297. https://doi.org/10.2174/1566523054065057

Gao, Q., & McNally, E. M. (2015). The Dystrophin Complex: Structure, function and implications for therapy. *Comprehensive Physiology*, *5*(3), 1223–1239. https://doi.org/10.1002/cphy.c140048

Garcia, V., Phelps, S. E. L., Gray, S., & Neale, M. J. (2011). Bidirectional resection of DNA double-strand breaks by Mre11 and Exo1. *Nature*, *479*(7372), 241–244. https://doi.org/10.1038/nature10515

Gaudelli, N. M., Komor, A. C., Rees, H. A., Packer, M. S., Badran, A. H., Bryson, D. I., & Liu, D. R. (2017). Programmable base editing of A•T to G•C in genomic DNA without DNA

cleavage. *Nature*, *551*(7681), Article 7681. https://doi.org/10.1038/nature24644

Gillmore, J. D., Gane, E., Taubel, J., Kao, J., Fontana, M., Maitland, M. L., Seitzer, J., O'Connell, D., Walsh, K. R., Wood, K., Phillips, J., Xu, Y., Amaral, A., Boyd, A. P., Cehelsky, J. E., McKee, M. D., Schiermeier, A., Harari, O., Murphy, A., … Lebwohl, D. (2021). CRISPR-Cas9 In Vivo Gene Editing for Transthyretin Amyloidosis. *New England Journal of Medicine*, *385*(6), 493–502. https://doi.org/10.1056/NEJMoa2107454

Godfrey, C., Muses, S., McClorey, G., Wells, K. E., Coursindel, T., Terry, R. L., Betts, C., Hammond, S., O'Donovan, L., Hildyard, J., El Andaloussi, S., Gait, M. J., Wood, M. J., & Wells, D. J. (2015). How much dystrophin is enough: The physiological consequences of different levels of dystrophin in the mdx mouse. *Human Molecular Genetics*, *24*(15), 4225–4237. https://doi.org/10.1093/hmg/ddv155

Goemans, N., Mercuri, E., Belousova, E., Komaki, H., Dubrovsky, A., McDonald, C. M., Kraus, J. E., Lourbakos, A., Lin, Z., Campion, G., Wang, S. X., Campbell, C., Araujo, A., Bertini, E., Born, P., Cances, C., Chabrol, B., Chae, J.-H., Colomer Oferil, J., … Wilichowski, E. (2018). A randomized placebo-controlled phase 3 trial of an antisense oligonucleotide, drisapersen, in Duchenne muscular dystrophy. *Neuromuscular Disorders*, *28*(1), 4–15. https://doi.org/10.1016/j.nmd.2017.10.004

Gollins, H., McMahon, J., Wells, K. E., & Wells, D. J. (2003). High-efficiency plasmid gene transfer into dystrophic muscle. *Gene Therapy*, *10*(6), Article 6. https://doi.org/10.1038/sj.gt.3301927

Gonçalves, M. A. F. V., & de Vries, A. A. F. (2006). Adenovirus: From foe to friend. *Reviews in Medical Virology*, *16*(3), 167–186. https://doi.org/10.1002/rmv.494

Gong, B., Shin, M., Sun, J., Jung, C.-H., Bolt, E. L., van der Oost, J., & Kim, J.-S. (2014). Molecular insights into DNA interference by CRISPR-associated nuclease-helicase Cas3.

Proceedings of the National Academy of Sciences of the United States of America, *111*(46), 16359–16364. https://doi.org/10.1073/pnas.1410806111

Goyenvalle, A., Griffith, G., Babbs, A., El Andaloussi, S., Ezzat, K., Avril, A., Dugovic, B., Chaussenot, R., Ferry, A., Voit, T., Amthor, H., Bühr, C., Schürch, S., Wood, M. J. A., Davies, K. E., Vaillend, C., Leumann, C., & Garcia, L. (2015). Functional correction in mouse models of muscular dystrophy using exon-skipping tricyclo-DNA oligomers. *Nature Medicine*, *21*(3), 270–275. https://doi.org/10.1038/nm.3765

Goyenvalle, A., Leumann, C., & Garcia, L. (2016). Therapeutic Potential of Tricyclo-DNA antisense oligonucleotides. *Journal of Neuromuscular Diseases*, *3*(2), 157–167. https://doi.org/10.3233/JND-160146

Goyenvalle, A., Vulin, A., Fougerousse, F., Leturcq, F., Kaplan, J.-C., Garcia, L., & Danos, O. (2004). Rescue of dystrophic muscle through U7 snRNA-mediated exon skipping. *Science (New York, N.Y.)*, *306*(5702), 1796–1799. https://doi.org/10.1126/science.1104297

Grain, L., Cortina-Borja, M., Forfar, C., Hilton-Jones, D., Hopkin, J., & Burch, M. (2001). Cardiac abnormalities and skeletal muscle weakness in carriers of Duchenne and Becker muscular dystrophies and controls. *Neuromuscular Disorders: NMD*, *11*(2), 186–191. https://doi.org/10.1016/s0960-8966(00)00185-1

Gregorevic, P., Allen, J. M., Minami, E., Blankinship, M. J., Haraguchi, M., Meuse, L., Finn, E., Adams, M. E., Froehner, S. C., Murry, C. E., & Chamberlain, J. S. (2006). RAAV6-microdystrophin preserves muscle function and extends lifespan in severely dystrophic mice. *Nature Medicine*, *12*(7), 787–789. https://doi.org/10.1038/nm1439

Gregorevic, P., Blankinship, M. J., Allen, J. M., & Chamberlain, J. S. (2008). Systemic Microdystrophin Gene Delivery Improves Skeletal Muscle Structure and Function in Old Dystrophic mdx Mice. *Molecular Therapy*, *16*(4), 657–664. https://doi.org/10.1038/mt.2008.28

Griggs, R. C., Moxley, R. T., Mendell, J. R., Fenichel, G. M., Brooke, M. H., Pestronk, A., Miller, J. P., Cwik, V. A., Pandya, S., & Robison, J. (1993). Duchenne dystrophy: Randomized, controlled trial of prednisone (18 months) and azathioprine (12 months). *Neurology*, *43*(3 Pt 1), 520–527. https://doi.org/10.1212/wnl.43.3_part_1.520

Grimm, D., Lee, J. S., Wang, L., Desai, T., Akache, B., Storm, T. A., & Kay, M. A. (2008). In vitro and in vivo gene therapy vector evolution via multispecies interbreeding and retargeting of adeno-associated viruses. *Journal of Virology*, *82*(12), 5887–5911. https://doi.org/10.1128/JVI.00254-08

Gruber, K. (2012). Europe gives gene therapy the green light. *Lancet (London, England)*, *380*(9855), e10. https://doi.org/10.1016/s0140-6736(12)61992-8

Gruntman, A. M., Bish, L. T., Mueller, C., Sweeney, H. L., Flotte, T. R., & Gao, G. (2013). Gene Transfer in Skeletal and Cardiac Muscle Using Recombinant Adeno-Associated Virus. *Current Protocols in Microbiology*, *28*(1), 14D.3.1-14D.3.19. https://doi.org/10.1002/9780471729259.mc14d03s28

Guiraud, S., Squire, S. E., Edwards, B., Chen, H., Burns, D. T., Shah, N., Babbs, A., Davies, S. G., Wynne, G. M., Russell, A. J., Elsey, D., Wilson, F. X., Tinsley, J. M., & Davies, K. E. (2015). Second-generation compound for the modulation of utrophin in the therapy of DMD. *Human Molecular Genetics*, *24*(15), 4212–4224. https://doi.org/10.1093/hmg/ddv154

Guo, R., Zhu, G., Zhu, H., Ma, R., Peng, Y., Liang, D., & Wu, L. (2015). DMD mutation spectrum analysis in 613 Chinese patients with dystrophinopathy. *Journal of Human Genetics*, *60*(8), 435–442. https://doi.org/10.1038/jhg.2015.43

Guo, W., & Song, H. (2018). Development of Gene Therapeutics for Head and Neck Cancer in China: From Bench to Bedside. *Human Gene Therapy*, *29*(2), 180–187. https://doi.org/10.1089/hum.2017.230

Gupta, R. M., & Musunuru, K. (2014). Expanding the genetic editing tool kit: ZFNs, TALENs, and CRISPR-Cas9. *Journal of Clinical Investigation*, *124*(10), 4154–4161. https://doi.org/10.1172/JCI72992

Gussoni, E., Bennett, R. R., Muskiewicz, K. R., Meyerrose, T., Nolta, J. A., Gilgoff, I., Stein, J., Chan, Y., Lidov, H. G., Bönnemann, C. G., Moers, A. von, Morris, G. E., Dunnen, J. T. den, Chamberlain, J. S., Kunkel, L. M., & Weinberg, K. (2002). Long-term persistence of donor nuclei in a Duchenne muscular dystrophy patient receiving bone marrow transplantation. *The Journal of Clinical Investigation*, *110*(6), 807–814. https://doi.org/10.1172/JCI16098

Haas, M., Vlcek, V., Balabanov, P., Salmonson, T., Bakchine, S., Markey, G., Weise, M., Schlosser-Weber, G., Brohmann, H., Yerro, C. P., Mendizabal, M. R., Stoyanova-Beninska, V., & Hillege, H. L. (2015). European Medicines Agency review of ataluren for the treatment of ambulant patients aged 5 years and older with Duchenne muscular dystrophy resulting from a nonsense mutation in the dystrophin gene. *Neuromuscular Disorders: NMD*, *25*(1), 5–13. https://doi.org/10.1016/j.nmd.2014.11.011

Hakim, C. H., & Duan, D. (2012). Gender differences in contractile and passive properties of mdx extensor digitorum longus muscle. *Muscle & Nerve*, *45*(2), 250–256. https://doi.org/10.1002/mus.22275

Hakim, C. H., Wasala, N. B., Nelson, C. E., Wasala, L. P., Yue, Y., Louderman, J. A., Lessa, T. B., Dai, A., Zhang, K., Jenkins, G. J., Nance, M. E., Pan, X., Kodippili, K., Yang, N. N., Chen, S.-J., Gersbach, C. A., & Duan, D. (2018). AAV CRISPR editing rescues cardiac and muscle function for 18 months in dystrophic mice. *JCI Insight*, *3*(23), e124297, 124297. https://doi.org/10.1172/jci.insight.124297

Hakim, C. H., Wasala, N. B., Pan, X., Kodippili, K., Yue, Y., Zhang, K., Yao, G., Haffner, B., Duan, S. X., Ramos, J., Schneider, J. S., Yang, N. N., Chamberlain, J. S., & Duan, D. (2017). A Five-Repeat Micro-Dystrophin Gene Ameliorated Dystrophic Phenotype in the Severe DBA/2J-

mdx Model of Duchenne Muscular Dystrophy. *Molecular Therapy - Methods & Clinical Development*, *6*, 216–230. https://doi.org/10.1016/j.omtm.2017.06.006

Han, J., Zhang, J., Chen, L., Shen, B., Zhou, J., Hu, B., Du, Y., Tate, P. H., Huang, X., & Zhang, W. (2014). Efficient in vivo deletion of a large imprinted lncRNA by CRISPR/Cas9. *RNA Biology*, *11*(7), 829–835. https://doi.org/10.4161/rna.29624

Hanson, B., Stenler, S., Ahlskog, N., Chwalenia, K., Svrzikapa, N., Coenen-Stass, A. M. L., Weinberg, M. S., Wood, M. J. A., & Roberts, T. C. (2022). Non-uniform dystrophin re-expression after CRISPR-mediated exon excision in the dystrophin/utrophin double-knockout mouse model of DMD. *Molecular Therapy - Nucleic Acids*, *30*, 379–397. https://doi.org/10.1016/j.omtn.2022.10.010

Happi Mbakam, C., Lamothe, G., Tremblay, G., & Tremblay, J. P. (2022). CRISPR-Cas9 Gene Therapy for Duchenne Muscular Dystrophy. *Neurotherapeutics*, *19*(3), 931–941. https://doi.org/10.1007/s13311-022-01197-9

Harper, S. Q., Hauser, M. A., DelloRusso, C., Duan, D., Crawford, R. W., Phelps, S. F., Harper, H. A., Robinson, A. S., Engelhardt, J. F., Brooks, S. V., & Chamberlain, J. S. (2002). Modular flexibility of dystrophin: Implications for gene therapy of Duchenne muscular dystrophy. *Nature Medicine*, *8*(3), 253. https://doi.org/10.1038/nm0302-253

Hayashi, Y. K., Chou, F. L., Engvall, E., Ogawa, M., Matsuda, C., Hirabayashi, S., Yokochi, K., Ziober, B. L., Kramer, R. H., Kaufman, S. J., Ozawa, E., Goto, Y., Nonaka, I., Tsukahara, T., Wang, J. Z., Hoffman, E. P., & Arahata, K. (1998). Mutations in the integrin alpha7 gene cause congenital myopathy. *Nature Genetics*, *19*(1), 94–97. https://doi.org/10.1038/ng0598-94

Heier, C. R., Damsker, J. M., Yu, Q., Dillingham, B. C., Huynh, T., Van der Meulen, J. H., Sali, A., Miller, B. K., Phadke, A., Scheffer, L., Quinn, J., Tatem, K., Jordan, S., Dadgar, S., Rodriguez, O. C., Albanese, C., Calhoun, M., Gordish-Dressman, H., Jaiswal, J. K., …

Nagaraju, K. (2013). VBP15, a novel anti-inflammatory and membrane-stabilizer, improves muscular dystrophy without side effects. *EMBO Molecular Medicine*, *5*(10), 1569–1585. https://doi.org/10.1002/emmm.201302621

Hendriksen, J. G. M., & Vles, J. S. H. (2008). Neuropsychiatric disorders in males with duchenne muscular dystrophy: Frequency rate of attention-deficit hyperactivity disorder (ADHD), autism spectrum disorder, and obsessive--compulsive disorder. *Journal of Child Neurology*, *23*(5), 477–481. https://doi.org/10.1177/0883073807309775

Henry, S. P., Kim, T.-W., Kramer-Stickland, K., Zanardi, T. A., Fey, R. A., & Levin, A. A. (2007). Toxicologic Properties of 2-O-Methoxyethyl Chimeric Antisense Inhibitors in Animals and Man. In *Antisense Drug Technology* (2nd ed.). CRC Press.

Heo, Y.-A. (2020). Golodirsen: First Approval. *Drugs*, *80*(3), 329–333. https://doi.org/10.1007/s40265-020-01267-2

Heyer, W.-D., Ehmsen, K. T., & Liu, J. (2010). Regulation of homologous recombination in eukaryotes. *Annual Review of Genetics*, *44*, 113–139. https://doi.org/10.1146/annurev-genet-051710-150955

Hoe, N., Nakashima, K., Grigsby, D., Pan, X., Dou, S. J., Naidich, S., Garcia, M., Kahn, E., Bergmire-Sweat, D., & Musser, J. M. (1999). Rapid molecular genetic subtyping of serotype M1 group A Streptococcus strains. *Emerging Infectious Diseases*, *5*(2), 254–263.

Hoffman, E. P., Brown, R. H., & Kunkel, L. M. (1987). Dystrophin: The protein product of the duchenne muscular dystrophy locus. *Cell*, *51*(6), 919–928. https://doi.org/10.1016/0092-8674(87)90579-4

Holkers, M., Maggio, I., Liu, J., Janssen, J. M., Miselli, F., Mussolino, C., Recchia, A., Cathomen, T., & Gonçalves, M. A. F. V. (2013). Differential integrity of TALE nuclease genes following adenoviral and lentiviral vector gene transfer into human cells. *Nucleic Acids*

Research, *41*(5), e63. https://doi.org/10.1093/nar/gks1446

Horlbeck, M. A., Witkowsky, L. B., Guglielmi, B., Replogle, J. M., Gilbert, L. A., Villalta, J. E., Torigoe, S. E., Tjian, R., & Weissman, J. S. (2016). Nucleosomes impede Cas9 access to DNA in vivo and in vitro. *ELife*, *5*, e12677. https://doi.org/10.7554/eLife.12677

Hsu, P. D., Scott, D. A., Weinstein, J. A., Ran, F. A., Konermann, S., Agarwala, V., Li, Y., Fine, E. J., Wu, X., Shalem, O., Cradick, T. J., Marraffini, L. A., Bao, G., & Zhang, F. (2013). DNA targeting specificity of RNA-guided Cas9 nucleases. *Nature Biotechnology*, *31*(9), 827–832. https://doi.org/10.1038/nbt.2647

Huertas, P., & Jackson, S. P. (2009). Human CtIP mediates cell cycle control of DNA end resection and double strand break repair. *The Journal of Biological Chemistry*, *284*(14), 9558–9565. https://doi.org/10.1074/jbc.M808906200

Huo, Y., Nam, K. H., Ding, F., Lee, H., Wu, L., Xiao, Y., Farchione, M. D., Zhou, S., Rajashankar, K., Kurinov, I., Zhang, R., & Ke, A. (2014). Structures of CRISPR Cas3 offer mechanistic insights into Cascade-activated DNA unwinding and degradation. *Nature Structural & Molecular Biology*, *21*(9), Article 9. https://doi.org/10.1038/nsmb.2875

Ibrahim, Y. F., Hammady, S. H., Rifaai, R. A., Waz, S., Ibrahim, M. A., & Hafez, H. M. (2022). Dose-dependent ameliorating effect of lipoxin A4 on gentamicin-induced nephrotoxicity in rats: The role of TNFα, TGF-β, ICAM-1, and JNK signaling. *Chemico-Biological Interactions*, *366*, 110139. https://doi.org/10.1016/j.cbi.2022.110139

Ilsley, J. L., Sudol, M., & Winder, S. J. (2002). The WW domain: Linking cell signalling to the membrane cytoskeleton. *Cellular Signalling*, *14*(3), 183–189. https://doi.org/10.1016/S0898-6568(01)00236-4

Ishino, Y., Shinagawa, H., Makino, K., Amemura, M., & Nakata, A. (1987). Nucleotide sequence of the iap gene, responsible for alkaline phosphatase isozyme conversion in

Escherichia coli, and identification of the gene product. *Journal of Bacteriology*, *169*(12), 5429–5433. https://doi.org/10.1128/jb.169.12.5429-5433.1987

Iyombe-Engembe, J.-P., Ouellet, D. L., Barbeau, X., Rousseau, J., Chapdelaine, P., Lagüe, P., & Tremblay, J. P. (2016). Efficient Restoration of the Dystrophin Gene Reading Frame and Protein Structure in DMD Myoblasts Using the CinDel Method. *Molecular Therapy. Nucleic Acids*, *5*, e283. https://doi.org/10.1038/mtna.2015.58

Jansen, Ruud., Embden, Jan. D. A. van, Gaastra, Wim., & Schouls, Leo. M. (2002). Identification of genes that are associated with DNA repeats in prokaryotes. *Molecular Microbiology*, *43*(6), 1565–1575. https://doi.org/10.1046/j.1365-2958.2002.02839.x

Jasin, M., & Berg, P. (1988). Homologous integration in mammalian cells without target gene selection. *Genes & Development*, *2*(11), 1353–1363. https://doi.org/10.1101/gad.2.11.1353

Jiang, T., Zhang, X.-O., Weng, Z., & Xue, W. (2022). Deletion and replacement of long genomic sequences using prime editing. *Nature Biotechnology*, *40*(2), 227–234. https://doi.org/10.1038/s41587-021-01026-y

Jinek, M., Chylinski, K., Fonfara, I., Hauer, M., Doudna, J. A., & Charpentier, E. (2012). A programmable dual RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science (New York, N.Y.)*, *337*(6096), 816–821. https://doi.org/10.1126/science.1225829

Jinek, M., East, A., Cheng, A., Lin, S., Ma, E., & Doudna, J. (2013). RNA-programmed genome editing in human cells. *ELife*, *2*, e00471. https://doi.org/10.7554/eLife.00471

Johnson, N. E., & Statland, J. M. (2022). The Limb-Girdle Muscular Dystrophies. *Continuum (Minneapolis, Minn.)*, *28*(6), 1698–1714. https://doi.org/10.1212/CON.0000000000001178

Juan-Mateu, J., Gonzalez-Quereda, L., Rodriguez, M. J., Baena, M., Verdura, E., Nascimento, A., Ortez, C., Baiget, M., & Gallano, P. (2015). DMD Mutations in 576 Dystrophinopathy Families: A Step Forward in Genotype-Phenotype Correlations. *PLoS ONE*, *10*(8), e0135189.

https://doi.org/10.1371/journal.pone.0135189

Judge, L. M., Arnett, A. L. H., Banks, G. B., & Chamberlain, J. S. (2011). Expression of the dystrophin isoform Dp116 preserves functional muscle mass and extends lifespan without preventing dystrophy in severely dystrophic mice. *Human Molecular Genetics*, *20*(24), 4978–4990. https://doi.org/10.1093/hmg/ddr433

Kang, J. K., Malerba, A., Popplewell, L., Foster, K., & Dickson, G. (2011). Antisense-induced myostatin exon skipping leads to muscle hypertrophy in mice following octa-guanidine morpholino oligomer treatment. *Molecular Therapy: The Journal of the American Society of Gene Therapy*, *19*(1), 159–164. https://doi.org/10.1038/mt.2010.212

Kantor, A., McClements, M. E., & MacLaren, R. E. (2020). CRISPR-Cas9 DNA Base-Editing and Prime-Editing. *International Journal of Molecular Sciences*, *21*(17), 6240. https://doi.org/10.3390/ijms21176240

Karpati, G., Ajdukovic, D., Arnold, D., Gledhill, R. B., Guttmann, R., Holland, P., Koch, P. A., Shoubridge, E., Spence, D., Vanasse, M., Watters, G. V., Abrahamowicz, M., Duff, C., & Worton, R. G. (1993). Myoblast transfer in duchenne muscular dystrophy. *Annals of Neurology*, *34*(1), 8–17. https://doi.org/10.1002/ana.410340105

Katrekar, D., Moreno, A. M., Chen, G., Worlikar, A., & Mali, P. (2018). Oligonucleotide conjugated multi-functional adeno-associated viruses. *Scientific Reports*, *8*(1), 3589. https://doi.org/10.1038/s41598-018-21742-x

Kay, M. A. (2011). State-of-the-art gene-based therapies: The road ahead. *Nature Reviews Genetics*, *12*(5), Article 5. https://doi.org/10.1038/nrg2971

Kelley, L. A., Mezulis, S., Yates, C. M., Wass, M. N., & Sternberg, M. J. E. (2015). The Phyre2 web portal for protein modeling, prediction and analysis. *Nature Protocols*, *10*(6), Article 6. https://doi.org/10.1038/nprot.2015.053

Kesselheim, A. S., & Avorn, J. (2016). Approving a Problematic Muscular Dystrophy Drug: Implications for FDA Policy. *JAMA*, *316*(22), 2357–2358. https://doi.org/10.1001/jama.2016.16437

Kher, G., Trehan, S., & Misra, A. (2011). Antisense Oligonucleotides and RNA Interference. In A. Misra (Ed.), *Challenges in Delivery of Therapeutic Genomics and Proteomics: Vol. Chapter 7* (pp. 325–386). Elsevier. https://doi.org/10.1016/B978-0-12-384964-9.00007-4

Khirani, S., Ramirez, A., Aubertin, G., Boulé, M., Chemouny, C., Forin, V., & Fauroux, B. (2014). Respiratory muscle decline in Duchenne muscular dystrophy. *Pediatric Pulmonology*, *49*(5), 473–481. https://doi.org/10.1002/ppul.22847

Khurana, T. S., Watkins, S. C., Chafey, P., Chelly, J., Tomé, F. M., Fardeau, M., Kaplan, J. C., & Kunkel, L. M. (1991). Immunolocalization and developmental expression of dystrophin related protein in skeletal muscle. *Neuromuscular Disorders: NMD*, *1*(3), 185–194. https://doi.org/10.1016/0960-8966(91)90023-l

Kim, E., Koo, T., Park, S. W., Kim, D., Kim, K., Cho, H.-Y., Song, D. W., Lee, K. J., Jung, M. H., Kim, S., Kim, J. H., Kim, J. H., & Kim, J.-S. (2017). In vivo genome editing with a small Cas9 orthologue derived from Campylobacter jejuni. *Nature Communications*, *8*, 14500. https://doi.org/10.1038/ncomms14500

Kim, S., Campbell, K. A., Fox, D. J., Matthews, D. J., & Valdez, R. (2015). Corticosteroid Treatments in Males With Duchenne Muscular Dystrophy: Treatment Duration and Time to Loss of Ambulation. *Journal of Child Neurology*, *30*(10), 1275–1280. https://doi.org/10.1177/0883073814558120

Kim, S., Kim, D., Cho, S. W., Kim, J., & Kim, J.-S. (2014). Highly efficient RNA-guided genome editing in human cells via delivery of purified Cas9 ribonucleoproteins. *Genome Research*, *24*(6), 1012–1019. https://doi.org/10.1101/gr.171322.113

Kim, V. N. (2005). MicroRNA biogenesis: Coordinated cropping and dicing. *Nature Reviews Molecular Cell Biology*, *6*(5), Article 5. https://doi.org/10.1038/nrm1644

Kim, Y. G., Cha, J., & Chandrasegaran, S. (1996). Hybrid restriction enzymes: Zinc finger fusions to Fok I cleavage domain. *Proceedings of the National Academy of Sciences of the United States of America*, *93*(3), 1156–1160.

Knopp, Y., Geis, F. K., Heckl, D., Horn, S., Neumann, T., Kuehle, J., Meyer, J., Fehse, B., Baum, C., Morgan, M., Meyer, J., Schambach, A., & Galla, M. (2018). Transient Retrovirus-Based CRISPR/Cas9 All-in-One Particles for Efficient, Targeted Gene Knockout. *Molecular Therapy - Nucleic Acids*, *13*, 256–274. https://doi.org/10.1016/j.omtn.2018.09.006

Kobayashi, K., Nakahori, Y., Miyake, M., Matsumura, K., Kondo-Iida, E., Nomura, Y., Segawa, M., Yoshioka, M., Saito, K., Osawa, M., Hamano, K., Sakakihara, Y., Nonaka, I., Nakagome, Y., Kanazawa, I., Nakamura, Y., Tokunaga, K., & Toda, T. (1998). An ancient retrotransposal insertion causes Fukuyama-type congenital muscular dystrophy. *Nature*, *394*(6691), 388–392. https://doi.org/10.1038/28653

Koenig, M., Hoffman, E. P., Bertelson, C. J., Monaco, A. P., Feener, C., & Kunkel, L. M. (1987). Complete cloning of the duchenne muscular dystrophy (DMD) cDNA and preliminary genomic organization of the DMD gene in normal and affected individuals. *Cell*, *50*(3), 509–517. https://doi.org/10.1016/0092-8674(87)90504-6

Koenig, M., & Kunkel, L. M. (1990). Detailed analysis of the repeat domain of dystrophin reveals four potential hinge segments that may confer flexibility. *Journal of Biological Chemistry*, *265*(8), 4560–4566.

Koenig, M., Monaco, A. P., & Kunkel, L. M. (1988). The complete sequence of dystrophin predicts a rod-shaped cytoskeletal protein. *Cell*, *53*(2), 219–228.

Kole, R., & Krieg, A. M. (2015). Exon skipping therapy for Duchenne muscular dystrophy.

Advanced Drug Delivery Reviews, *87*, 104–107. https://doi.org/10.1016/j.addr.2015.05.008

Koller, B. H., Hagemann, L. J., Doetschman, T., Hagaman, J. R., Huang, S., Williams, P. J., First, N. L., Maeda, N., & Smithies, O. (1989). Germ-line transmission of a planned alteration made in a hypoxanthine phosphoribosyltransferase gene by homologous recombination in embryonic stem cells. *Proceedings of the National Academy of Sciences*, *86*(22), 8927–8931. https://doi.org/10.1073/pnas.86.22.8927

Komaki, H., Nagata, T., Saito, T., Masuda, S., Takeshita, E., Sasaki, M., Tachimori, H., Nakamura, H., Aoki, Y., & Takeda, S. (2018). Systemic administration of the antisense oligonucleotide NS-065/NCNP-01 for skipping of exon 53 in patients with Duchenne muscular dystrophy. *Science Translational Medicine*, *10*(437), eaan0713. https://doi.org/10.1126/scitranslmed.aan0713

Komor, A. C., Kim, Y. B., Packer, M. S., Zuris, J. A., & Liu, D. R. (2016). Programmable editing of a target base in genomic DNA without double-stranded DNA cleavage. *Nature*, *533*(7603), Article 7603. https://doi.org/10.1038/nature17946

Koo, T., Popplewell, L., Athanasopoulos, T., & Dickson, G. (2014). Triple trans-splicing adeno-associated virus vectors capable of transferring the coding sequence for full-length dystrophin protein into dystrophic mice. *Human Gene Therapy*, *25*(2), 98–108. https://doi.org/10.1089/hum.2013.164

Kornegay, J. N., Li, J., Bogan, J. R., Bogan, D. J., Chen, C., Zheng, H., Wang, B., Qiao, C., Howard, J. F., & Xiao, X. (2010). Widespread Muscle Expression of an AAV9 Human Mini-dystrophin Vector After Intravenous Injection in Neonatal Dystrophin-deficient Dogs. *Molecular Therapy*, *18*(8), 1501–1508. https://doi.org/10.1038/mt.2010.94

Kosicki, M., Tomberg, K., & Bradley, A. (2018). Repair of double-strand breaks induced by CRISPR-Cas9 leads to large deletions and complex rearrangements. *Nature Biotechnology*,

36(8), 765–771. https://doi.org/10.1038/nbt.4192

Kotin, R. M., & Snyder, R. O. (2017). Manufacturing Clinical Grade Recombinant Adeno-Associated Virus Using Invertebrate Cell Lines. *Human Gene Therapy*, *28*(4), 350–360. https://doi.org/10.1089/hum.2017.042

Kotterman, M. A., & Schaffer, D. V. (2014). Engineering adeno-associated viruses for clinical gene therapy. *Nature Reviews Genetics*, *15*(7), Article 7. https://doi.org/10.1038/nrg3742

Kourakis, S., Timpani, C. A., Campelj, D. G., Hafner, P., Gueven, N., Fischer, D., & Rybalka, E. (2021). Standard of care versus new-wave corticosteroids in the treatment of Duchenne muscular dystrophy: Can we do better? *Orphanet Journal of Rare Diseases*, *16*(1), 117. https://doi.org/10.1186/s13023-021-01758-9

Kozak, M. (2005). Regulation of translation via mRNA structure in prokaryotes and eukaryotes. *Gene*, *361*, 13–37. https://doi.org/10.1016/j.gene.2005.06.037

Krag, T. O. B., Bogdanovich, S., Jensen, C. J., Fischer, M. D., Hansen-Schwartz, J., Javazon, E. H., Flake, A. W., Edvinsson, L., & Khurana, T. S. (2004). Heregulin ameliorates the dystrophic phenotype in mdx mice. *Proceedings of the National Academy of Sciences of the United States of America*, *101*(38), 13856–13860. https://doi.org/10.1073/pnas.0405972101

Kragelund, B. B., Weterings, E., Hartmann-Petersen, R., & Keijzers, G. (2016). The Ku70/80 ring in Non-Homologous End-Joining: Easy to slip on, hard to remove. *Frontiers in Bioscience (Landmark Edition)*, *21*(3), 514–527. https://doi.org/10.2741/4406

Kuehn, M. R., Bradley, A., Robertson, E. J., & Evans, M. J. (1987). A potential animal model for Lesch–Nyhan syndrome through introduction of HPRT mutations into mice. *Nature*, *326*(6110), Article 6110. https://doi.org/10.1038/326295a0

Kumar, N., Stanford, W., de Solis, C., Aradhana, Abraham, N. D., Dao, T.-M. J., Thaseen, S., Sairavi, A., Gonzalez, C. U., & Ploski, J. E. (2018). The Development of an AAV-Based

CRISPR SaCas9 Genome Editing System That Can Be Delivered to Neurons in vivo and Regulated via Doxycycline and Cre-Recombinase. *Frontiers in Molecular Neuroscience*, *11*, 413. https://doi.org/10.3389/fnmol.2018.00413

Kwon, J. B., Ettyreddy, A. R., Vankara, A., Bohning, J. D., Devlin, G., Hauschka, S. D., Asokan, A., & Gersbach, C. A. (2020). In Vivo Gene Editing of Muscle Stem Cells with Adeno-Associated Viral Vectors in a Mouse Model of Duchenne Muscular Dystrophy. *Molecular Therapy. Methods & Clinical Development*, *19*, 320–329. https://doi.org/10.1016/j.omtm.2020.09.016

Lai, Y., Thomas, G. D., Yue, Y., Yang, H. T., Li, D., Long, C., Judge, L., Bostick, B., Chamberlain, J. S., Terjung, R. L., & Duan, D. (2009). Dystrophins carrying spectrin-like repeats 16 and 17 anchor nNOS to the sarcolemma and enhance exercise performance in a mouse model of muscular dystrophy. *The Journal of Clinical Investigation*, *119*(3), 624–635. https://doi.org/10.1172/JCI36612

Lam, J. K. W., Chow, M. Y. T., Zhang, Y., & Leung, S. W. S. (2015). SiRNA Versus miRNA as Therapeutics for Gene Silencing. *Molecular Therapy. Nucleic Acids*, *4*(9), e252. https://doi.org/10.1038/mtna.2015.23

Lamb, Y. N. (2021). BNT162b2 mRNA COVID-19 Vaccine: First Approval. *Drugs*, *81*(4), 495–501. https://doi.org/10.1007/s40265-021-01480-7

Landfeldt, E., Lindgren, P., Bell, C. F., Schmitt, C., Guglieri, M., Straub, V., Lochmüller, H., & Bushby, K. (2014). The burden of Duchenne muscular dystrophy: An international, cross-sectional study. *Neurology*, *83*(6), 529–536. https://doi.org/10.1212/WNL.0000000000000669

Larkindale, J., Yang, W., Hogan, P. F., Simon, C. J., Zhang, Y., Jain, A., Habeeb-Louks, E. M., Kennedy, A., & Cwik, V. A. (2014). Cost of illness for neuromuscular diseases in the United States. *Muscle & Nerve*, *49*(3), 431–438. https://doi.org/10.1002/mus.23942

Lattanzi, A., Duguez, S., Moiani, A., Izmiryan, A., Barbon, E., Martin, S., Mamchaoui, K., Mouly, V., Bernardi, F., Mavilio, F., & Bovolenta, M. (2017). Correction of the Exon 2 Duplication in DMD Myoblasts by a Single CRISPR/Cas9 System. *Molecular Therapy. Nucleic Acids*, *7*, 11–19. https://doi.org/10.1016/j.omtn.2017.02.004

Le Guiner, C., Servais, L., Montus, M., Larcher, T., Fraysse, B., Moullec, S., Allais, M., François, V., Dutilleul, M., Malerba, A., Koo, T., Thibaut, J.-L., Matot, B., Devaux, M., Le Duff, J., Deschamps, J.-Y., Barthelemy, I., Blot, S., Testault, I., … Dickson, G. (2017). Long-term microdystrophin gene therapy is effective in a canine model of Duchenne muscular dystrophy. *Nature Communications*, *8*(1), Article 1. https://doi.org/10.1038/ncomms16105

Lee, B. L., Nam, S. H., Lee, J. H., Ki, C. S., Lee, M., & Lee, J. (2012). Genetic analysis of dystrophin gene for affected male and female carriers with Duchenne/Becker muscular dystrophy in Korea. *Journal of Korean Medical Science*, *27*(3), 274–280. https://doi.org/10.3346/jkms.2012.27.3.274

Lee, C. S., Bishop, E. S., Zhang, R., Yu, X., Farina, E. M., Yan, S., Zhao, C., Zeng, Z., Shu, Y., Wu, X., Lei, J., Li, Y., Zhang, W., Yang, C., Wu, K., Wu, Y., Ho, S., Athiviraham, A., Lee, M. J., … He, T.-C. (2017). Adenovirus-mediated gene delivery: Potential applications for gene and cell-based therapies in the new era of personalized medicine. *Genes & Diseases*, *4*(2), 43–63. https://doi.org/10.1016/j.gendis.2017.04.001

Lee, E. J., Robinson, T. M., Tabor, J. J., Mikos, A. G., & Suh, J. (2018). Reverse Transduction Can Improve Efficiency of AAV Vectors in Transduction-Resistant Cells. *Biotechnology and Bioengineering*, *115*(12), 3042–3049. https://doi.org/10.1002/bit.26830

Lee, J., Echigoya, Y., Duddy, W., Saito, T., Aoki, Y., Takeda, S., & Yokota, T. (2018). Antisense PMO cocktails effectively skip dystrophin exons 45-55 in myotubes transdifferentiated from DMD patient fibroblasts. *PLoS ONE*, *13*(5), e0197084. https://doi.org/10.1371/journal.pone.0197084

Lee, K.-W., Kim, D.-S., & Kwon, H.-J. (2004). CG sequence- and phosphorothioate backbone modification-dependent activation of the NF-κB-responsive gene expression by CpG-oligodeoxynucleotides in human RPMI 8226 B cells. *Molecular Immunology*, *41*(10), 955–964. https://doi.org/10.1016/j.molimm.2004.06.022

Leibowitz, D., & Dubowitz, V. (1981). Intellect and behaviour in Duchenne muscular dystrophy. *Developmental Medicine and Child Neurology*, *23*(5), 577–590. https://doi.org/10.1111/j.1469-8749.1981.tb02039.x

Lemmers, R. J. L. F., van der Vliet, P. J., Klooster, R., Sacconi, S., Camaño, P., Dauwerse, J. G., Snider, L., Straasheijm, K. R., van Ommen, G. J., Padberg, G. W., Miller, D. G., Tapscott, S. J., Tawil, R., Frants, R. R., & van der Maarel, S. M. (2010). A unifying genetic model for facioscapulohumeral muscular dystrophy. *Science (New York, N.Y.)*, *329*(5999), 1650–1653. https://doi.org/10.1126/science.1189044

Levi, O., Genin, O., Angelini, C., Halevy, O., & Pines, M. (2015). Inhibition of muscle fibrosis results in increases in both utrophin levels and the number of revertant myofibres in Duchenne muscular dystrophy. *Oncotarget*, *6*(27), 23249–23260.

Li, C., & Samulski, R. J. (2020). Engineering adeno-associated virus vectors for gene therapy. *Nature Reviews Genetics*, *21*(4), Article 4. https://doi.org/10.1038/s41576-019-0205-4

Li, H. L., Fujimoto, N., Sasakawa, N., Shirai, S., Ohkame, T., Sakuma, T., Tanaka, M., Amano, N., Watanabe, A., Sakurai, H., Yamamoto, T., Yamanaka, S., & Hotta, A. (2014). Precise Correction of the Dystrophin Gene in Duchenne Muscular Dystrophy Patient Induced Pluripotent Stem Cells by TALEN and CRISPR-Cas9. *Stem Cell Reports*, *4*(1), 143–154. https://doi.org/10.1016/j.stemcr.2014.10.013

Li, J., Sun, W., Wang, B., Xiao, X., & Liu, X.-Q. (2008). Protein trans-splicing as a means for viral vector-mediated in vivo gene therapy. *Human Gene Therapy*, *19*(9), 958–964.

https://doi.org/10.1089/hum.2008.009

Li, T., Huang, S., Jiang, W. Z., Wright, D., Spalding, M. H., Weeks, D. P., & Yang, B. (2011). TAL nucleases (TALNs): Hybrid proteins composed of TAL effectors and FokI DNA-cleavage domain. *Nucleic Acids Research*, *39*(1), 359–372. https://doi.org/10.1093/nar/gkq704

Li, X., Eastman, E. M., Schwartz, R. J., & Draghia-Akli, R. (1999). Synthetic muscle promoters: Activities exceeding naturally occurring regulatory sequences. *Nature Biotechnology*, *17*(3), 241–245. https://doi.org/10.1038/6981

Liang, M. (2018). Oncorine, the World First Oncolytic Virus Medicine and its Update in China. *Current Cancer Drug Targets*, *18*(2), 171–176. https://doi.org/10.2174/1568009618666171129221503

Lim, K. R. Q., Echigoya, Y., Nagata, T., Kuraoka, M., Kobayashi, M., Aoki, Y., Partridge, T., Maruyama, R., Takeda, S., & Yokota, T. (2019). Efficacy of Multi-exon Skipping Treatment in Duchenne Muscular Dystrophy Dog Model Neonates. *Molecular Therapy*, *27*(1), 76–86. https://doi.org/10.1016/j.ymthe.2018.10.011

Lim, K. R. Q., Maruyama, R., & Yokota, T. (2017). Eteplirsen in the treatment of Duchenne muscular dystrophy. *Drug Design, Development and Therapy*, *11*, 533–545. https://doi.org/10.2147/DDDT.S97635

Lim, K. R. Q., Nguyen, Q., & Yokota, T. (2020). Genotype-Phenotype Correlations in Duchenne and Becker Muscular Dystrophy Patients from the Canadian Neuromuscular Disease Registry. *Journal of Personalized Medicine*, *10*(4), 241. https://doi.org/10.3390/jpm10040241

Lim, K. R. Q., Woo, S., Melo, D., Huang, Y., Dzierlega, K., Shah, M. N. A., Aslesh, T., Roshmi, R. R., Echigoya, Y., Maruyama, R., Moulton, H. M., & Yokota, T. (2022). Development of DG9 peptide-conjugated single- and multi-exon skipping therapies for the treatment of Duchenne muscular dystrophy. *Proceedings of the National Academy of Sciences*,

119(9), e2112546119. https://doi.org/10.1073/pnas.2112546119

Ling, S., Yang, S., Hu, X., Yin, D., Dai, Y., Qian, X., Wang, D., Pan, X., Hong, J., Sun, X., Yang, H., Paludan, S. R., & Cai, Y. (2021). Lentiviral delivery of co-packaged Cas9 mRNA and a Vegfa-targeting guide RNA prevents wet age-related macular degeneration in mice. *Nature Biomedical Engineering*, *5*(2), Article 2. https://doi.org/10.1038/s41551-020-00656-y

Liu, P., Chen, M., Liu, Y., Qi, L. S., & Ding, S. (2018). CRISPR-Based Chromatin Remodeling of the Endogenous Oct4 or Sox2 Locus Enables Reprogramming to Pluripotency. *Cell Stem Cell*, *22*(2), 252-261.e4. https://doi.org/10.1016/j.stem.2017.12.001

Long, C., Amoasii, L., Mireault, A. A., McAnally, J. R., Li, H., Sanchez-Ortiz, E., Bhattacharyya, S., Shelton, J. M., Bassel-Duby, R., & Olson, E. N. (2016). Postnatal genome editing partially restores dystrophin expression in a mouse model of muscular dystrophy. *Science*, *351*(6271), 400–403. https://doi.org/10.1126/science.aad5725

Long, C., McAnally, J. R., Shelton, J. M., Mireault, A. A., Bassel-Duby, R., & Olson, E. N. (2014). Prevention of muscular dystrophy in mice by CRISPR/Cas9–mediated editing of germline DNA. *Science*, *345*(6201), 1184–1188. https://doi.org/10.1126/science.1254445

Lonowski, L. A., Narimatsu, Y., Riaz, A., Delay, C. E., Yang, Z., Niola, F., Duda, K., Ober, E. A., Clausen, H., Wandall, H. H., Hansen, S. H., Bennett, E. P., & Frödin, M. (2017). Genome editing using FACS enrichment of nuclease-expressing cells and indel detection by amplicon analysis. *Nature Protocols*, *12*(3), Article 3. https://doi.org/10.1038/nprot.2016.165

Lovric, J., Mano, M., Zentilin, L., Eulalio, A., Zacchigna, S., & Giacca, M. (2012). Terminal Differentiation of Cardiac and Skeletal Myocytes Induces Permissivity to AAV Transduction by Relieving Inhibition Imposed by DNA Damage Response Proteins. *Molecular Therapy*, *20*(11), 2087–2097. https://doi.org/10.1038/mt.2012.144

Lu, B., Javidi-Parsijani, P., Makani, V., Mehraein-Ghomi, F., Sarhan, W. M., Sun, D., Yoo, K.

W., Atala, Z. P., Lyu, P., & Atala, A. (2019). Delivering SaCas9 mRNA by lentivirus-like bionanoparticles for transient expression and efficient genome editing. *Nucleic Acids Research*, *47*(8), e44. https://doi.org/10.1093/nar/gkz093

Lv, H., Zhang, S., Wang, B., Cui, S., & Yan, J. (2006). Toxicity of cationic lipids and cationic polymers in gene delivery. *Journal of Controlled Release*, *114*(1), 100–109. https://doi.org/10.1016/j.jconrel.2006.04.014

Lyu, P., Wang, L., & Lu, B. (2020). Virus-Like Particle Mediated CRISPR/Cas9 Delivery for Efficient and Safe Genome Editing. *Life*, *10*(12), 366. https://doi.org/10.3390/life10120366

Macdonald, G. (2017, October 26). *US FDA rejects Translarna and says additional trial and CMC data is needed*. Outsourcing-Pharma.Com. https://www.outsourcing-pharma.com/Article/2017/10/26/US-FDA-rejects-Translarna-and-says-additional-trial-and-CMC-data-is-needed

Madeira, F., Pearce, M., Tivey, A. R. N., Basutkar, P., Lee, J., Edbali, O., Madhusoodanan, N., Kolesnikov, A., & Lopez, R. (2022). Search and sequence analysis tools services from EMBL-EBI in 2022. *Nucleic Acids Research*, gkac240. https://doi.org/10.1093/nar/gkac240

Maeder, M. L., & Gersbach, C. A. (2016). Genome-editing Technologies for Gene and Cell Therapy. *Molecular Therapy*, *24*(3), 430–446. https://doi.org/10.1038/mt.2016.10

Maeder, M. L., Stefanidakis, M., Wilson, C. J., Baral, R., Barrera, L. A., Bounoutas, G. S., Bumcrot, D., Chao, H., Ciulla, D. M., DaSilva, J. A., Dass, A., Dhanapal, V., Fennell, T. J., Friedland, A. E., Giannoukos, G., Gloskowski, S. W., Glucksmann, A., Gotta, G. M., Jayaram, H., … Jiang, H. (2019). Development of a gene-editing approach to restore vision loss in Leber congenital amaurosis type 10. *Nature Medicine*, *25*(2), 229–233. https://doi.org/10.1038/s41591-018-0327-9

Maggio, I., Stefanucci, L., Janssen, J. M., Liu, J., Chen, X., Mouly, V., & Gonçalves, M. A. F.

V. (2016). Selection-free gene repair after adenoviral vector transduction of designer nucleases: Rescue of dystrophin synthesis in DMD muscle cell populations. *Nucleic Acids Research*, *44*(3), 1449–1470. https://doi.org/10.1093/nar/gkv1540

Mahajan, R. (2019). Onasemnogene Abeparvovec for Spinal Muscular Atrophy: The Costlier Drug Ever. *International Journal of Applied & Basic Medical Research*, *9*(3), 127–128. https://doi.org/10.4103/ijabmr.IJABMR_190_19

Maheshri, N., Koerber, J. T., Kaspar, B. K., & Schaffer, D. V. (2006). Directed evolution of adeno-associated virus yields enhanced gene delivery vectors. *Nature Biotechnology*, *24*(2), 198–204. https://doi.org/10.1038/nbt1182

Makarova, K. S., Aravind, L., Grishin, N. V., Rogozin, I. B., & Koonin, E. V. (2002). A DNA repair system specific for thermophilic Archaea and bacteria predicted by genomic context analysis. *Nucleic Acids Research*, *30*(2), 482–496.

Makarova, K. S., Aravind, L., Wolf, Y. I., & Koonin, E. V. (2011). Unification of Cas protein families and a simple scenario for the origin and evolution of CRISPR-Cas systems. *Biology Direct*, *6*(1), 38. https://doi.org/10.1186/1745-6150-6-38

Makarova, K. S., Grishin, N. V., Shabalina, S. A., Wolf, Y. I., & Koonin, E. V. (2006). A putative RNA-interference-based immune system in prokaryotes: Computational analysis of the predicted enzymatic machinery, functional analogies with eukaryotic RNAi, and hypothetical mechanisms of action. *Biology Direct*, *1*, 7. https://doi.org/10.1186/1745-6150-1-7

Makarova, K. S., Wolf, Y. I., Alkhnbashi, O. S., Costa, F., Shah, S. A., Saunders, S. J., Barrangou, R., Brouns, S. J. J., Charpentier, E., Haft, D. H., Horvath, P., Moineau, S., Mojica, F. J. M., Terns, R. M., Terns, M. P., White, M. F., Yakunin, A. F., Garrett, R. A., van der Oost, J., … Koonin, E. V. (2015). An updated evolutionary classification of CRISPR–Cas systems. *Nature Reviews Microbiology*, *13*(11), 722–736. https://doi.org/10.1038/nrmicro3569

Malerba, A., Kang, J. K., McClorey, G., Saleh, A. F., Popplewell, L., Gait, M. J., Wood, M. J., & Dickson, G. (2012). Dual Myostatin and Dystrophin Exon Skipping by Morpholino Nucleic Acid Oligomers Conjugated to a Cell-penetrating Peptide Is a Promising Therapeutic Strategy for the Treatment of Duchenne Muscular Dystrophy. *Molecular Therapy. Nucleic Acids*, *1*(12), e62. https://doi.org/10.1038/mtna.2012.54

Malik, V., Rodino-Klapac, L. R., Viollet, L., Wall, C., King, W., Al-Dahhak, R., Lewis, S., Shilling, C. J., Kota, J., Serrano-Munuera, C., Hayes, J., Mahan, J. D., Campbell, K. J., Banwell, B., Dasouki, M., Watts, V., Sivakumar, K., Bien-Willner, R., Flanigan, K. M., … Mendell, J. R. (2010). Gentamicin-induced readthrough of stop codons in Duchenne muscular dystrophy. *Annals of Neurology*, *67*(6), 771–780. https://doi.org/10.1002/ana.22024

Manuvakhova, M., Keeling, K., & Bedwell, D. M. (2000). Aminoglycoside antibiotics mediate context-dependent suppression of termination codons in a mammalian translation system. *RNA*, *6*(7), 1044–1055.

Manzur, A. Y., Kuntzer, T., Pike, M., & Swan, A. V. (2008). Glucocorticoid corticosteroids for Duchenne muscular dystrophy. *Cochrane Database of Systematic Reviews*, *1*. https://doi.org/10.1002/14651858.CD003725.pub3

Masepohl, B., Görlitz, K., & Böhme, H. (1996). Long tandemly repeated repetitive (LTRR) sequences in the filamentous cyanobacterium Anabaena sp. PCC 7120. *Biochimica et Biophysica Acta (BBA) - Gene Structure and Expression*, *1307*(1), 26–30. https://doi.org/10.1016/0167-4781(96)00040-1

Mashal, R. D., Koontz, J., & Sklar, J. (1995). Detection of mutations by cleavage of DNA heteroduplexes with bacteriophage resolvases. *Nature Genetics*, *9*(2), Article 2. https://doi.org/10.1038/ng0295-177

Mashimo, T., Takizawa, A., Voigt, B., Yoshimi, K., Hiai, H., Kuramoto, T., & Serikawa, T.

(2010). Generation of knockout rats with X-linked severe combined immunodeficiency (X-SCID) using zinc-finger nucleases. *PloS One*, *5*(1), e8870. https://doi.org/10.1371/journal.pone.0008870

Matthews, E., Brassington, R., Kuntzer, T., Jichi, F., & Manzur, A. Y. (2016). Corticosteroids for the treatment of Duchenne muscular dystrophy. *The Cochrane Database of Systematic Reviews*, *2016*(5), CD003725. https://doi.org/10.1002/14651858.CD003725.pub4

Maurer, M. S., Schwartz, J. H., Gundapaneni, B., Elliott, P. M., Merlini, G., Waddington-Cruz, M., Kristen, A. V., Grogan, M., Witteles, R., Damy, T., Drachman, B. M., Shah, S. J., Hanna, M., Judge, D. P., Barsdorf, A. I., Huber, P., Patterson, T. A., Riley, S., Schumacher, J., … ATTR-ACT Study Investigators. (2018). Tafamidis Treatment for Patients with Transthyretin Amyloid Cardiomyopathy. *The New England Journal of Medicine*, *379*(11), 1007–1016. https://doi.org/10.1056/NEJMoa1805689

McCarty, N. S., Graham, A. E., Studená, L., & Ledesma-Amaro, R. (2020). Multiplexed CRISPR technologies for gene editing and transcriptional regulation. *Nature Communications*, *11*(1), Article 1. https://doi.org/10.1038/s41467-020-15053-x

McClorey, G., Fall, A. M., Moulton, H. M., Iversen, P. L., Rasko, J. E., Ryan, M., Fletcher, S., & Wilton, S. D. (2006). Induced dystrophin exon skipping in human muscle explants. *Neuromuscular Disorders*, *16*(9), 583–590. https://doi.org/10.1016/j.nmd.2006.05.017

McClorey, G., Moulton, H. M., Iversen, P. L., Fletcher, S., & Wilton, S. D. (2006). Antisense oligonucleotide-induced exon skipping restores dystrophin expression in vitro in a canine model of DMD. *Gene Therapy*, *13*(19), Article 19. https://doi.org/10.1038/sj.gt.3302800

McDonald, C. M., Campbell, C., Torricelli, R. E., Finkel, R. S., Flanigan, K. M., Goemans, N., Heydemann, P., Kaminska, A., Kirschner, J., Muntoni, F., Osorio, A. N., Schara, U., Sejersen, T., Shieh, P. B., Sweeney, H. L., Topaloglu, H., Tulinius, M., Vilchez, J. J., Voit, T., … Vita, G.

(2017). Ataluren in patients with nonsense mutation Duchenne muscular dystrophy (ACT DMD): A multicentre, randomised, double-blind, placebo-controlled, phase 3 trial. *The Lancet*, *390*(10101), 1489–1498. https://doi.org/10.1016/S0140-6736(17)31611-2

McDonald, C. M., Shieh, P. B., Abdel-Hamid, H. Z., Connolly, A. M., Ciafaloni, E., Wagner, K. R., Goemans, N., Mercuri, E., Khan, N., Koenig, E., Malhotra, J., Zhang, W., Han, B., Mendell, J. R., & the Italian DMD Telethon Registry Study Group, Leuven NMRC Registry Investigators, CINRG Duchenne Natural History Investigators, and PROMOVI Trial Clinical Investigators. (2021). Open-Label Evaluation of Eteplirsen in Patients with Duchenne Muscular Dystrophy Amenable to Exon 51 Skipping: PROMOVI Trial. *Journal of Neuromuscular Diseases*, *8*(6), 989–1001. https://doi.org/10.3233/JND-210643

Meliani, A., Boisgerault, F., Fitzpatrick, Z., Marmier, S., Leborgne, C., Collaud, F., Simon Sola, M., Charles, S., Ronzitti, G., Vignaud, A., van Wittenberghe, L., Marolleau, B., Jouen, F., Tan, S., Boyer, O., Christophe, O., Brisson, A. R., Maguire, C. A., & Mingozzi, F. (2017). Enhanced liver gene transfer and evasion of preexisting humoral immunity with exosome-enveloped AAV vectors. *Blood Advances*, *1*(23), 2019–2031. https://doi.org/10.1182/bloodadvances.2017010181

Meliani, A., Boisgerault, F., Hardet, R., Marmier, S., Collaud, F., Ronzitti, G., Leborgne, C., Costa Verdera, H., Simon Sola, M., Charles, S., Vignaud, A., van Wittenberghe, L., Manni, G., Christophe, O., Fallarino, F., Roy, C., Michaud, A., Ilyinskii, P., Kishimoto, T. K., & Mingozzi, F. (2018). Antigen-selective modulation of AAV immunogenicity with tolerogenic rapamycin nanoparticles enables successful vector re-administration. *Nature Communications*, *9*(1), 4098. https://doi.org/10.1038/s41467-018-06621-3

Mendell, J. R., Al-Zaidy, S. A., Rodino-Klapac, L. R., Goodspeed, K., Gray, S. J., Kay, C. N., Boye, S. L., Boye, S. E., George, L. A., Salabarria, S., Corti, M., Byrne, B. J., & Tremblay, J. P. (2021). Current Clinical Applications of In Vivo Gene Therapy with AAVs. *Molecular*

Therapy, *29*(2), 464–488. https://doi.org/10.1016/j.ymthe.2020.12.007

Mendell, J. R., Al-Zaidy, S., Shell, R., Arnold, W. D., Rodino-Klapac, L. R., Prior, T. W., Lowes, L., Alfano, L., Berry, K., Church, K., Kissel, J. T., Nagendran, S., L'Italien, J., Sproule, D. M., Wells, C., Cardenas, J. A., Heitzer, M. D., Kaspar, A., Corcoran, S., … Kaspar, B. K. (2017). Single-Dose Gene-Replacement Therapy for Spinal Muscular Atrophy. *New England Journal of Medicine*, *377*(18), 1713–1722. https://doi.org/10.1056/NEJMoa1706198

Mendell, J. R., Goemans, N., Lowes, L. P., Alfano, L. N., Berry, K., Shao, J., Kaye, E. M., Mercuri, E., & Eteplirsen Study Group and Telethon Foundation DMD Italian Network. (2016). Longitudinal effect of eteplirsen versus historical control on ambulation in Duchenne muscular dystrophy. *Annals of Neurology*, *79*(2), 257–271. https://doi.org/10.1002/ana.24555

Mendell, J., Sahenk, Z., Lehman, K., Nease, C., Lowes, L., Reash, N., Iammarino, M., Alfano, L., Vaiea, J., Lewis, S., Church, K., Shell, R., Potter, R., Griffin, D., Pozsgai, E., Hogan, M., Hu, L., Giblin, K., & Rodino-Klapac, L. (2022). *Phase 1/2a trial of delandistrogene moxeparvovec (SRP-9001) in patients with Duchenne muscular dystrophy: 3-year safety and functional outcomes*.

Meng, J., Moore, M., Counsell, J., Muntoni, F., Popplewell, L., & Morgan, J. (2022). Optimized lentiviral vector to restore full-length dystrophin via a cell-mediated approach in a mouse model of Duchenne muscular dystrophy. *Molecular Therapy - Methods & Clinical Development*, *25*, 491–507. https://doi.org/10.1016/j.omtm.2022.04.015

Mercuri, E., Bönnemann, C. G., & Muntoni, F. (2019). Muscular dystrophies. *Lancet (London, England)*, *394*(10213), 2025–2038. https://doi.org/10.1016/S0140-6736(19)32910-1

Meselson, M., & Yuan, R. (1968). DNA restriction enzyme from E. coli. *Nature*, *217*(5134), 1110–1114. https://doi.org/10.1038/2171110a0

Messina, S., Vita, G. L., Aguennouz, M., Sframeli, M., Romeo, S., Rodolico, C., & Vita, G.

(2011). Activation of NF-kappaB pathway in Duchenne muscular dystrophy: Relation to age. *Acta Myologica: Myopathies and Cardiomyopathies: Official Journal of the Mediterranean Society of Myology*, *30*(1), 16–23.

Migliorati, J. M., Jin, J., & Zhong, X.-B. (2022). SiRNA drug Leqvio (inclisiran) to lower cholesterol. *Trends in Pharmacological Sciences*, *43*(5), 455–456. https://doi.org/10.1016/j.tips.2022.02.003

Milone, M. C., & O'Doherty, U. (2018). Clinical use of lentiviral vectors. *Leukemia*, *32*(7), Article 7. https://doi.org/10.1038/s41375-018-0106-0

Mingozzi, F., Chen, Y., Murphy, S. L., Edmonson, S. C., Tai, A., Price, S. D., Metzger, M. E., Zhou, S., Wright, J. F., Donahue, R. E., Dunbar, C. E., & High, K. A. (2012). Pharmacological modulation of humoral immunity in a nonhuman primate model of AAV gene transfer for hemophilia B. *Molecular Therapy: The Journal of the American Society of Gene Therapy*, *20*(7), 1410–1416. https://doi.org/10.1038/mt.2012.84

Miraldi Utz, V., Coussa, R. G., Antaki, F., & Traboulsi, E. I. (2018). Gene therapy for RPE65-related retinal disease. *Ophthalmic Genetics*, *39*(6), 671–677. https://doi.org/10.1080/13816810.2018.1533027

Moat, S. J., Bradley, D. M., Salmon, R., Clarke, A., & Hartley, L. (2013). Newborn bloodspot screening for Duchenne muscular dystrophy: 21 years experience in Wales (UK). *European Journal of Human Genetics: EJHG*, *21*(10), 1049–1053. https://doi.org/10.1038/ejhg.2012.301

Mohammed, F., Elshafey, A., Al-balool, H., Alaboud, H., Ali, M. A. B., Baqer, A., & Bastaki, L. (2018). Mutation spectrum analysis of Duchenne/Becker muscular dystrophy in 68 families in Kuwait: The era of personalized medicine. *PLOS ONE*, *13*(5), e0197205. https://doi.org/10.1371/journal.pone.0197205

Mojica, F. J. M., Díez-Villaseñor, C., García-Martínez, J., & Almendros, C. (2009). Short motif

sequences determine the targets of the prokaryotic CRISPR defence system. *Microbiology*, *155*(3), 733–740. https://doi.org/10.1099/mic.0.023960-0

Mojica, F. J. M., Díez-Villaseñor, C., Soria, E., & Juez, G. (2000). Biological significance of a family of regularly spaced repeats in the genomes of Archaea, Bacteria and mitochondria. *Molecular Microbiology*, *36*(1), 244–246. https://doi.org/10.1046/j.1365-2958.2000.01838.x

Mojica, F. j. m., Ferrer, C., Juez, G., & Rodríguez-Valera, F. (1995). Long stretches of short tandem repeats are present in the largest replicons of the Archaea Haloferax mediterranei and Haloferax volcanii and could be involved in replicon partitioning. *Molecular Microbiology*, *17*(1), 85–93. https://doi.org/10.1111/j.1365-2958.1995.mmi_17010085.x

Monaco, A. P., Bertelson, C. J., Liechti-Gallati, S., Moser, H., & Kunkel, L. M. (1988). An explanation for the phenotypic differences between patients bearing partial deletions of the DMD locus. *Genomics*, *2*(1), 90–95. https://doi.org/10.1016/0888-7543(88)90113-9

Moore, J. K., & Haber, J. E. (1996). Cell cycle and genetic requirements of two pathways of nonhomologous end-joining repair of double-strand breaks in Saccharomyces cerevisiae. *Molecular and Cellular Biology*, *16*(5), 2164–2173.

Moreb, E. A., Hutmacher, M., & Lynch, M. D. (2020). CRISPR-Cas 'Non-Target' Sites Inhibit On-Target Cutting Rates. *The CRISPR Journal*, *3*(6), 550–561. https://doi.org/10.1089/crispr.2020.0065

Moreb, E. A., & Lynch, M. D. (2021). Genome dependent Cas9/gRNA search time underlies sequence dependent gRNA activity. *Nature Communications*, *12*(1), 5034. https://doi.org/10.1038/s41467-021-25339-3

Moreb, E. A., & Lynch, M. D. (2022). A Meta-Analysis of gRNA Library Screens Enables an Improved Understanding of the Impact of gRNA Folding and Structural Stability on CRISPR-Cas9 Activity. *The CRISPR Journal*, *5*(1), 146–154. https://doi.org/10.1089/crispr.2021.0084

Morgan, J. E., Beauchamp, J. R., Pagel, C. N., Peckham, M., Ataliotis, P., Jat, P. S., Noble, M. D., Farmer, K., & Partridge, T. A. (1994). Myogenic cell lines derived from transgenic mice carrying a thermolabile T antigen: A model system for the derivation of tissue-specific and mutation-specific cell lines. *Developmental Biology*, *162*(2), 486–498. https://doi.org/10.1006/dbio.1994.1103

Mulepati, S., & Bailey, S. (2011). Structural and biochemical analysis of nuclease domain of clustered regularly interspaced short palindromic repeat (CRISPR)-associated protein 3 (Cas3). *The Journal of Biological Chemistry*, *286*(36), 31896–31903. https://doi.org/10.1074/jbc.M111.270017

Muntoni, F., Tejura, B., Spinty, S., Roper, H., Hughes, I., Layton, G., Davies, K. E., Harriman, S., & Tinsley, J. (2019). A Phase 1b Trial to Assess the Pharmacokinetics of Ezutromid in Pediatric Duchenne Muscular Dystrophy Patients on a Balanced Diet. *Clinical Pharmacology in Drug Development*, *8*(7), 922–933. https://doi.org/10.1002/cpdd.642

Najm, F. J., Strand, C., Donovan, K. F., Hegde, M., Sanson, K. R., Vaimberg, E. W., Sullender, M. E., Hartenian, E., Kalani, Z., Fusi, N., Listgarten, J., Younger, S. T., Bernstein, B. E., Root, D. E., & Doench, J. G. (2018). Orthologous CRISPR–Cas9 enzymes for combinatorial genetic screens. *Nature Biotechnology*, *36*(2), Article 2. https://doi.org/10.1038/nbt.4048

Nakamura, A., & Takeda, S. (2009). Exon-skipping therapy for Duchenne muscular dystrophy. *Neuropathology: Official Journal of the Japanese Society of Neuropathology*, *29*(4), 494–501. https://doi.org/10.1111/j.1440-1789.2009.01028.x

Naldini, L. (2015). Gene therapy returns to centre stage. *Nature*, *526*(7573), Article 7573. https://doi.org/10.1038/nature15818

Naldini, L., Blömer, U., Gallay, P., Ory, D., Mulligan, R., Gage, F. H., Verma, I. M., & Trono, D. (1996). In Vivo Gene Delivery and Stable Transduction of Nondividing Cells by a Lentiviral

Vector. *Science*, *272*(5259), 263–267. https://doi.org/10.1126/science.272.5259.263

Namgoong, J. H., & Bertoni, C. (2016). Clinical potential of ataluren in the treatment of Duchenne muscular dystrophy. *Degenerative Neurological and Neuromuscular Disease*, *6*, 37–48. https://doi.org/10.2147/DNND.S71808

Nature Biotechnology. (2020). High-dose AAV gene therapy deaths. *Nature Biotechnology*, *38*(8), Article 8. https://doi.org/10.1038/s41587-020-0642-9

Nelson, C. E., & Gersbach, C. A. (2016). Engineering Delivery Vehicles for Genome Editing. *Annual Review of Chemical and Biomolecular Engineering*, *7*, 637–662. https://doi.org/10.1146/annurev-chembioeng-080615-034711

Nelson, C. E., Hakim, C. H., Ousterout, D. G., Thakore, P. I., Moreb, E. A., Rivera, R. M. C., Madhavan, S., Pan, X., Ran, F. A., Yan, W. X., Asokan, A., Zhang, F., Duan, D., & Gersbach, C. A. (2016). In vivo genome editing improves muscle function in a mouse model of Duchenne muscular dystrophy. *Science*, *351*(6271), 403–407. https://doi.org/10.1126/science.aad5143

Nelson, C. E., Wu, Y., Gemberling, M. P., Oliver, M. L., Waller, M. A., Bohning, J. D., Robinson-Hamm, J. N., Bulaklak, K., Castellanos Rivera, R. M., Collier, J. H., Asokan, A., & Gersbach, C. A. (2019). Long-term evaluation of AAV-CRISPR genome editing for Duchenne muscular dystrophy. *Nature Medicine*, *25*(3), Article 3. https://doi.org/10.1038/s41591-019-0344-3

Nigro, G., Comi, L. I., Politano, L., & Bain, R. J. (1990). The incidence and evolution of cardiomyopathy in Duchenne muscular dystrophy. *International Journal of Cardiology*, *26*(3), 271–277. https://doi.org/10.1016/0167-5273(90)90082-g

Nonaka, I. (1999). Distal myopathies. *Current Opinion in Neurology*, *12*(5), 493–499. https://doi.org/10.1097/00019052-199910000-00002

Norwood, F. L. M., Harling, C., Chinnery, P. F., Eagle, M., Bushby, K., & Straub, V. (2009).

Prevalence of genetic muscle disease in Northern England: In-depth analysis of a muscle clinic population. *Brain: A Journal of Neurology*, *132*(Pt 11), 3175–3186. https://doi.org/10.1093/brain/awp236

Nuñez, J. K., Kranzusch, P. J., Noeske, J., Wright, A. V., Davies, C. W., & Doudna, J. A. (2014). Cas1-Cas2 complex formation mediates spacer acquisition during CRISPR-Cas adaptive immunity. *Nature Structural & Molecular Biology*, *21*(6), 528–534. https://doi.org/10.1038/nsmb.2820

Oakley, R. H., & Cidlowski, J. A. (2013). The biology of the glucocorticoid receptor: New signaling mechanisms in health and disease. *The Journal of Allergy and Clinical Immunology*, *132*(5), 1033–1044. https://doi.org/10.1016/j.jaci.2013.09.007

Odom, G. L., Gregorevic, P., Allen, J. M., & Chamberlain, J. S. (2011). Gene therapy of mdx mice with large truncated dystrophins generated by recombination using rAAV6. *Molecular Therapy: The Journal of the American Society of Gene Therapy*, *19*(1), 36–45. https://doi.org/10.1038/mt.2010.205

Oleykowski, C. A., Bronson Mullins, C. R., Godwin, A. K., & Yeung, A. T. (1998). Mutation detection using a novel plant endonuclease. *Nucleic Acids Research*, *26*(20), 4597–4602. https://doi.org/10.1093/nar/26.20.4597

Orlov, Y. L., Te Boekhorst, R., & Abnizova, I. I. (2006). Statistical measures of the structure of genomic sequences: Entropy, complexity, and position information. *Journal of Bioinformatics and Computational Biology*, *4*(2), 523–536. https://doi.org/10.1142/s0219720006001801

Ott, P. A., & Hodi, F. S. (2016). Talimogene Laherparepvec for the Treatment of Advanced Melanoma. *Clinical Cancer Research: An Official Journal of the American Association for Cancer Research*, *22*(13), 3127–3131. https://doi.org/10.1158/1078-0432.CCR-15-2709

Ottesen, E. W. (2017). ISS-N1 makes the First FDA-approved Drug for Spinal Muscular

Atrophy. *Translational Neuroscience*, *8*, 1–6. https://doi.org/10.1515/tnsci-2017-0001

Ousterout, D. G., Kabadi, A. M., Thakore, P. I., Majoros, W. H., Reddy, T. E., & Gersbach, C. A. (2015). Multiplex CRISPR/Cas9-based genome editing for correction of dystrophin mutations that cause Duchenne muscular dystrophy. *Nature Communications*, *6*(1), Article 1. https://doi.org/10.1038/ncomms7244

Ousterout, D. G., Kabadi, A. M., Thakore, P. I., Perez-Pinera, P., Brown, M. T., Majoros, W. H., Reddy, T. E., & Gersbach, C. A. (2015). Correction of dystrophin expression in cells from Duchenne muscular dystrophy patients through genomic excision of exon 51 by zinc finger nucleases. *Molecular Therapy: The Journal of the American Society of Gene Therapy*, *23*(3), 523–532. https://doi.org/10.1038/mt.2014.234

Ouyang, L., Grosse, S. D., & Kenneson, A. (2008). Health care utilization and expenditures for children and young adults with muscular dystrophy in a privately insured population. *Journal of Child Neurology*, *23*(8), 883–888. https://doi.org/10.1177/0883073808314962

Padhy, S. K., Takkar, B., Narayanan, R., Venkatesh, P., & Jalali, S. (2020). Voretigene Neparvovec and Gene Therapy for Leber's Congenital Amaurosis: Review of Evidence to Date. *The Application of Clinical Genetics*, *13*, 179–208. https://doi.org/10.2147/TACG.S230720

Paik, J., & Duggan, S. (2019). Volanesorsen: First Global Approval. *Drugs*, *79*(12), 1349–1354. https://doi.org/10.1007/s40265-019-01168-z

Partridge, T. A., Morgan, J. E., Coulton, G. R., Hoffman, E. P., & Kunkel, L. M. (1989). Conversion of mdx myofibres from dystrophin-negative to -positive by injection of normal myoblasts. *Nature*, *337*(6203), Article 6203. https://doi.org/10.1038/337176a0

Paunovska, K., Loughrey, D., & Dahlman, J. E. (2022). Drug delivery systems for RNA therapeutics. *Nature Reviews Genetics*, *23*(5), Article 5. https://doi.org/10.1038/s41576-021-00439-4

Pearce, M., Blake, D. J., Tinsley, J. M., Byth, B. C., Campbell, L., Monaco, A. P., & Davies, K. E. (1993). The utrophin and dystrophin genes share similarities in genomic structure. *Human Molecular Genetics*, *2*(11), 1765–1772. https://doi.org/10.1093/hmg/2.11.1765

Pearson, S., Jia, H., & Kandachi, K. (2004). China approves first gene therapy. *Nature Biotechnology*, *22*(1), Article 1. https://doi.org/10.1038/nbt0104-3

Péault, B., Rudnicki, M., Torrente, Y., Cossu, G., Tremblay, J. P., Partridge, T., Gussoni, E., Kunkel, L. M., & Huard, J. (2007). Stem and progenitor cells in skeletal muscle development, maintenance, and therapy. *Molecular Therapy: The Journal of the American Society of Gene Therapy*, *15*(5), 867–877. https://doi.org/10.1038/mt.sj.6300145

Péladeau, C., Adam, N., Bronicki, L. M., Coriati, A., Thabet, M., Al-Rewashdy, H., Vanstone, J., Mears, A., Renaud, J.-M., Holcik, M., & Jasmin, B. J. (2020). Identification of therapeutics that target eEF1A2 and upregulate utrophin A translation in dystrophic muscles. *Nature Communications*, *11*(1), Article 1. https://doi.org/10.1038/s41467-020-15971-w

Péladeau, C., Ahmed, A., Amirouche, A., Crawford Parks, T. E., Bronicki, L. M., Ljubicic, V., Renaud, J.-M., & Jasmin, B. J. (2016). Combinatorial therapeutic activation with heparin and AICAR stimulates additive effects on utrophin A expression in dystrophic muscles. *Human Molecular Genetics*, *25*(1), 24–43. https://doi.org/10.1093/hmg/ddv444

Phelps, S. F., Hauser, M. A., Cole, N. M., Rafael, J. A., Hinkle, R. T., Faulkner, J. A., & Chamberlain, J. S. (1995). Expression of full-length and truncated dystrophin mini-genes in transgenic mdx mice. *Human Molecular Genetics*, *4*(8), 1251–1258. https://doi.org/10.1093/hmg/4.8.1251

Philippidis, A. (2020). After Third Death, Audentes' AT132 Remains on Clinical Hold. *Human Gene Therapy*, *31*(17–18), 908–910. https://doi.org/10.1089/hum.2020.29133.bfs

Philippidis, A. (2022a). After Patient Death, FDA Places Hold on Pfizer Duchenne Muscular

Dystrophy Gene Therapy Trial. *Human Gene Therapy*, *33*(3–4), 111–115. https://doi.org/10.1089/hum.2022.29198.bfs

Philippidis, A. (2022b). Pfizer Eyes Resuming Phase III Enrollment, Investigates Phase Ib Death Tied to Duchenne Muscular Dystrophy Candidate. *Human Gene Therapy*, *33*(5–6), 215–217. https://doi.org/10.1089/hum.2022.29203.bfs

Phillips, J. W., & Morgan, W. F. (1994). Illegitimate recombination induced by DNA double-strand breaks in a mammalian chromosome. *Molecular and Cellular Biology*, *14*(9), 5794–5803.

Pisani, C., Strimpakos, G., Gabanella, F., Di Certo, M. G., Onori, A., Severini, C., Luvisetto, S., Farioli-Vecchioli, S., Carrozzo, I., Esposito, A., Canu, T., Mattei, E., Corbi, N., & Passananti, C. (2018). Utrophin up-regulation by artificial transcription factors induces muscle rescue and impacts the neuromuscular junction in mdx mice. *Biochimica Et Biophysica Acta. Molecular Basis of Disease*, *1864*(4 Pt A), 1172–1182. https://doi.org/10.1016/j.bbadis.2018.01.030

Poh, A. (2016). First Oncolytic Viral Therapy for Melanoma. *Cancer Discovery*, *6*(1), 6. https://doi.org/10.1158/2159-8290.CD-NB2015-158

Popplewell, L., Koo, T., Leclerc, X., Duclert, A., Mamchaoui, K., Gouble, A., Mouly, V., Voit, T., Pâques, F., Cédrone, F., Isman, O., Yáñez-Muñoz, R. J., & Dickson, G. (2013). Gene Correction of a Duchenne Muscular Dystrophy Mutation by Meganuclease-Enhanced Exon Knock-In. *Human Gene Therapy*, *24*(7), 692–701. https://doi.org/10.1089/hum.2013.081

Potaczek, D. P., Garn, H., Unger, S. D., & Renz, H. (2016). Antisense molecules: A new class of drugs. *Journal of Allergy and Clinical Immunology*, *137*(5), 1334–1346. https://doi.org/10.1016/j.jaci.2015.12.1344

Protein Molecular Weight. (n.d.). Retrieved 25 February 2023, from https://www.bioinformatics.org/sms/prot_mw.html

Provasi, E., Genovese, P., Lombardo, A., Magnani, Z., Liu, P.-Q., Reik, A., Chu, V., Paschon, D. E., Zhang, L., Kuball, J., Camisa, B., Bondanza, A., Casorati, G., Ponzoni, M., Ciceri, F., Bordignon, C., Greenberg, P. D., Holmes, M. C., Gregory, P. D., … Bonini, C. (2012). Editing T cell specificity towards leukemia by zinc finger nucleases and lentiviral gene transfer. *Nature Medicine*, *18*(5), Article 5. https://doi.org/10.1038/nm.2700

Qasim, W., Zhan, H., Samarasinghe, S., Adams, S., Amrolia, P., Stafford, S., Butler, K., Rivat, C., Wright, G., Somana, K., Ghorashian, S., Pinner, D., Ahsan, G., Gilmour, K., Lucchini, G., Inglott, S., Mifsud, W., Chiesa, R., Peggs, K. S., … Veys, P. (2017). Molecular remission of infant B-ALL after infusion of universal TALEN gene-edited CAR T cells. *Science Translational Medicine*, *9*(374), eaaj2013. https://doi.org/10.1126/scitranslmed.aaj2013

Qiao, C., Koo, T., Li, J., Xiao, X., & Dickson, J. G. (2011). Gene Therapy in Skeletal Muscle Mediated by Adeno-Associated Virus Vectors. In R. O. Snyder & P. Moullier (Eds.), *Adeno-Associated Virus: Methods and Protocols* (pp. 119–140). Humana Press. https://doi.org/10.1007/978-1-61779-370-7_5

Raguram, A., Banskota, S., & Liu, D. R. (2022). Therapeutic in vivo delivery of gene editing agents. *Cell*, *185*(15), 2806–2827. https://doi.org/10.1016/j.cell.2022.03.045

Ran, F. A., Cong, L., Yan, W. X., Scott, D. A., Gootenberg, J. S., Kriz, A. J., Zetsche, B., Shalem, O., Wu, X., Makarova, K. S., Koonin, E. V., Sharp, P. A., & Zhang, F. (2015). In vivo genome editing using Staphylococcus aureus Cas9. *Nature*, *520*(7546), Article 7546. https://doi.org/10.1038/nature14299

Rando, T. A. (2001). The dystrophin-glycoprotein complex, cellular signaling, and the regulation of cell survival in the muscular dystrophies. *Muscle & Nerve*, *24*(12), 1575–1594. https://doi.org/10.1002/mus.1192

Rando, T. A., Disatnik, M.-H., & Zhou, L. Z.-H. (2000). Rescue of dystrophin expression in

mdx mouse muscle by RNA/DNA oligonucleotides. *Proceedings of the National Academy of Sciences*, *97*(10), 5363–5368. https://doi.org/10.1073/pnas.97.10.5363

Ratcliff, R., Evans, M. J., Cuthbert, A. W., MacVinish, L. J., Foster, D., Anderson, J. R., & Colledge, W. H. (1993). Production of a severe cystic fibrosis mutation in mice by gene targeting. *Nature Genetics*, *4*(1), Article 1. https://doi.org/10.1038/ng0593-35

RCSB PDB. (n.d.). Retrieved 9 May 2023, from https://www.rcsb.org/

Reeves, E. K. M., Hoffman, E. P., Nagaraju, K., Damsker, J. M., & McCall, J. M. (2013). VBP15: Preclinical characterization of a novel anti-inflammatory delta 9,11 steroid. *Bioorganic & Medicinal Chemistry*, *21*(8), 2241–2249. https://doi.org/10.1016/j.bmc.2013.02.009

Relizani, K., Griffith, G., Echevarría, L., Zarrouki, F., Facchinetti, P., Vaillend, C., Leumann, C., Garcia, L., & Goyenvalle, A. (2017). Efficacy and Safety Profile of Tricyclo-DNA Antisense Oligonucleotides in Duchenne Muscular Dystrophy Mouse Model. *Molecular Therapy. Nucleic Acids*, *8*, 144–157. https://doi.org/10.1016/j.omtn.2017.06.013

Renkawitz, J., Lademann, C. A., & Jentsch, S. (2014). Mechanisms and principles of homology search during recombination. *Nature Reviews. Molecular Cell Biology*, *15*(6), 369–383. https://doi.org/10.1038/nrm3805

Renneberg, D., Bouliong, E., Reber, U., Schümperli, D., & Leumann, C. J. (2002). Antisense properties of tricyclo-DNA. *Nucleic Acids Research*, *30*(13), 2751–2757.

Ricotti, V., Spinty, S., Roper, H., Hughes, I., Tejura, B., Robinson, N., Layton, G., Davies, K., Muntoni, F., & Tinsley, J. (2016). Safety, Tolerability, and Pharmacokinetics of SMT C1100, a 2-Arylbenzoxazole Utrophin Modulator, following Single- and Multiple-Dose Administration to Pediatric Patients with Duchenne Muscular Dystrophy. *PLoS ONE*, *11*(4). https://doi.org/10.1371/journal.pone.0152840

Riesenberg, S., Helmbrecht, N., Kanis, P., Maricic, T., & Pääbo, S. (2022). Improved gRNA

secondary structures allow editing of target sites resistant to CRISPR-Cas9 cleavage. *Nature Communications*, *13*(1), 489. https://doi.org/10.1038/s41467-022-28137-7

Roberts, R. G., Coffey, A. J., Bobrow, M., & Bentley, D. R. (1992). Determination of the exon structure of the distal portion of the dystrophin gene by vectorette PCR. *Genomics*, *13*(4), 942–950. https://doi.org/10.1016/0888-7543(92)90005-D

Roshmi, R. R., & Yokota, T. (2019). Viltolarsen for the treatment of Duchenne muscular dystrophy. *Drugs of Today (Barcelona, Spain: 1998)*, *55*(10), 627–639. https://doi.org/10.1358/dot.2019.55.10.3045038

Roshmi, R. R., & Yokota, T. (2021). Pharmacological Profile of Viltolarsen for the Treatment of Duchenne Muscular Dystrophy: A Japanese Experience. *Clinical Pharmacology: Advances and Applications*, *13*, 235–242. https://doi.org/10.2147/CPAA.S288842

Rouet, P., Smih, F., & Jasin, M. (1994). Introduction of double-strand breaks into the genome of mouse cells by expression of a rare-cutting endonuclease. *Molecular and Cellular Biology*, *14*(12), 8096–8106.

Russell, S., Bennett, J., Wellman, J. A., Chung, D. C., Yu, Z.-F., Tillman, A., Wittes, J., Pappas, J., Elci, O., McCague, S., Cross, D., Marshall, K. A., Walshire, J., Kehoe, T. L., Reichert, H., Davis, M., Raffini, L., George, L. A., Hudson, F. P., … Maguire, A. M. (2017). Efficacy and safety of voretigene neparvovec (AAV2-hRPE65v2) in patients with RPE65-mediated inherited retinal dystrophy: A randomised, controlled, open-label, phase 3 trial. *The Lancet*, *390*(10097), 849–860. https://doi.org/10.1016/S0140-6736(17)31868-8

Ryu, S.-M., Koo, T., Kim, K., Lim, K., Baek, G., Kim, S.-T., Kim, H. S., Kim, D.-E., Lee, H., Chung, E., & Kim, J.-S. (2018). Adenine base editing in mouse embryos and an adult mouse model of Duchenne muscular dystrophy. *Nature Biotechnology*, *36*(6), 536–539. https://doi.org/10.1038/nbt.4148

Sadoulet-Puccio, H. M., Rajala, M., & Kunkel, L. M. (1997). Dystrobrevin and dystrophin: An interaction through coiled-coil motifs. *Proceedings of the National Academy of Sciences of the United States of America*, *94*(23), 12413–12418.

Sakuma, T., Nishikawa, A., Kume, S., Chayama, K., & Yamamoto, T. (2014). Multiplex genome engineering in human cells using all-in-one CRISPR/Cas9 vector system. *Scientific Reports*, *4*(1), Article 1. https://doi.org/10.1038/srep05400

Salmaninejad, A., Jafari Abarghan, Y., Bozorg Qomi, S., Bayat, H., Yousefi, M., Azhdari, S., Talebi, S., & Mojarrad, M. (2021). Common therapeutic advances for Duchenne muscular dystrophy (DMD). *The International Journal of Neuroscience*, *131*(4), 370–389. https://doi.org/10.1080/00207454.2020.1740218

Salvetti, A., Orève, S., Chadeuf, G., Favre, D., Cherel, Y., Champion-Arnaud, P., David-Ameline, J., & Moullier, P. (1998). Factors influencing recombinant adeno-associated virus production. *Human Gene Therapy*, *9*(5), 695–706. https://doi.org/10.1089/hum.1998.9.5-695

Sampaolesi, M., Blot, S., D'Antona, G., Granger, N., Tonlorenzi, R., Innocenzi, A., Mognol, P., Thibaud, J.-L., Galvez, B. G., Barthélémy, I., Perani, L., Mantero, S., Guttinger, M., Pansarasa, O., Rinaldi, C., Cusella De Angelis, M. G., Torrente, Y., Bordignon, C., Bottinelli, R., & Cossu, G. (2006). Mesoangioblast stem cells ameliorate muscle function in dystrophic dogs. *Nature*, *444*(7119), Article 7119. https://doi.org/10.1038/nature05282

San Filippo, J., Sung, P., & Klein, H. (2008). Mechanism of eukaryotic homologous recombination. *Annual Review of Biochemistry*, *77*, 229–257. https://doi.org/10.1146/annurev.biochem.77.061306.125255

Sardone, V., Zhou, H., Muntoni, F., Ferlini, A., & Falzarano, M. S. (2017). Antisense Oligonucleotide-Based Therapy for Neuromuscular Disease. *Molecules : A Journal of Synthetic Chemistry and Natural Product Chemistry*, *22*(4), 563.

https://doi.org/10.3390/molecules22040563

Sarepta Therapeutics. (2023). *A Phase 2, Two-Part, Multiple-Ascending-Dose Study of SRP-5051 for Dose Determination, Then Dose Expansion, in Patients With Duchenne Muscular Dystrophy Amenable to Exon 51-Skipping Treatment* (Clinical Trial Registration No. NCT04004065). clinicaltrials.gov. https://clinicaltrials.gov/ct2/show/NCT04004065

Sarepta Therapeutics. (2022a). *Community Letter: Momentum Trial | Sarepta Therapeutics*. https://www.sarepta.com/community-letter-momentum-trial

Sarepta Therapeutics. (2022b). *Sarepta Therapeutics Announces That FDA has Lifted its Clinical Hold on SRP-5051 for the Treatment of Duchenne Muscular Dystrophy | Sarepta Therapeutics, Inc.* https://investorrelations.sarepta.com/news-releases/news-release-details/sarepta-therapeutics-announces-fda-has-lifted-its-clinical-hold

Schimmer, J., & Breazzano, S. (2016). Investor Outlook: Rising from the Ashes; GSK's European Approval of Strimvelis for ADA-SCID. *Human Gene Therapy. Clinical Development*, *27*(2), 57–61. https://doi.org/10.1089/humc.2016.29010.ind

Schunder, E., Rydzewski, K., Grunow, R., & Heuner, K. (2013). First indication for a functional CRISPR/Cas system in Francisella tularensis. *International Journal of Medical Microbiology: IJMM*, *303*(2), 51–60. https://doi.org/10.1016/j.ijmm.2012.11.004

Scott, L. J. (2020). Givosiran: First Approval. *Drugs*, *80*(3), 335–339. https://doi.org/10.1007/s40265-020-01269-0

Scott, L. J., & Keam, S. J. (2021). Lumasiran: First Approval. *Drugs*, *81*(2), 277–282. https://doi.org/10.1007/s40265-020-01463-0

Seimetz, D., Heller, K., & Richter, J. (2019). Approval of First CAR-Ts: Have we Solved all Hurdles for ATMPs? *Cell Medicine*, *11*, 2155179018822781. https://doi.org/10.1177/2155179018822781

Sentmanat, M. F., Peters, S. T., Florian, C. P., Connelly, J. P., & Pruett-Miller, S. M. (2018). A Survey of Validation Strategies for CRISPR-Cas9 Editing. *Scientific Reports*, *8*(1), Article 1. https://doi.org/10.1038/s41598-018-19441-8

Servais, L., Mercuri, E., Straub, V., Guglieri, M., Seferian, A. M., Scoto, M., Leone, D., Koenig, E., Khan, N., Dugar, A., Wang, X., Han, B., Wang, D., Muntoni, F., & SKIP-NMD Study Group. (2022). Long-Term Safety and Efficacy Data of Golodirsen in Ambulatory Patients with Duchenne Muscular Dystrophy Amenable to Exon 53 Skipping: A First-in-human, Multicenter, Two-Part, Open-Label, Phase 1/2 Trial. *Nucleic Acid Therapeutics*, *32*(1), 29–39. https://doi.org/10.1089/nat.2021.0043

Servais, L., Montus, M., Guiner, C. L., Ben Yaou, R., Annoussamy, M., Moraux, A., Hogrel, J.-Y., Seferian, A. M., Zehrouni, K., Le Moing, A.-G., Gidaro, T., Vanhulle, C., Laugel, V., Butoianu, N., Cuisset, J.-M., Sabouraud, P., Cances, C., Klein, A., Leturcq, F., … Voit, T. (2015). Non-Ambulant Duchenne Patients Theoretically Treatable by Exon 53 Skipping have Severe Phenotype. *Journal of Neuromuscular Diseases*, *2*(3), 269–279. https://doi.org/10.3233/JND-150100

Seto, J. T., Bengtsson, N. E., & Chamberlain, J. S. (2014). Therapy of Genetic Disorders: Novel Therapies for Duchenne Muscular Dystrophy. *Current Pediatrics Reports*, *2*(2), 102–112. https://doi.org/10.1007/s40124-014-0044-x

sgRNA Designer: CRISPRko. (n.d.). Retrieved 24 February 2023, from https://portals.broadinstitute.org/gpp/public/analysis-tools/sgrna-design

Sherkow, J. S., Zettler, P. J., & Greely, H. T. (2018). Is it 'gene therapy'? *Journal of Law and the Biosciences*, *5*(3), 786–793. https://doi.org/10.1093/jlb/lsy020

Shin, J.-H., Pan, X., Hakim, C. H., Yang, H. T., Yue, Y., Zhang, K., Terjung, R. L., & Duan, D. (2013). Microdystrophin ameliorates muscular dystrophy in the canine model of duchenne

muscular dystrophy. *Molecular Therapy: The Journal of the American Society of Gene Therapy*, *21*(4), 750–757. https://doi.org/10.1038/mt.2012.283

Shirley, M. (2021). Casimersen: First Approval. *Drugs*, *81*(7), 875–879. https://doi.org/10.1007/s40265-021-01512-2

Sienko, S., Buckon, C., Fowler, E., Bagley, A., Staudt, L., Sison-Williamson, M., Zebracki, K., McDonald, C. M., & Sussman, M. (2016). Prednisone and Deflazacort in Duchenne Muscular Dystrophy: Do They Play a Different Role in Child Behavior and Perceived Quality of Life? *PLoS Currents*, *8*, ecurrents.md.7628d9c014bfa29f821a5cd19723bbaa. https://doi.org/10.1371/currents.md.7628d9c014bfa29f821a5cd19723bbaa

Simhadri, V. L., McGill, J., McMahon, S., Wang, J., Jiang, H., & Sauna, Z. E. (2018). Prevalence of Pre-existing Antibodies to CRISPR-Associated Nuclease Cas9 in the USA Population. *Molecular Therapy - Methods & Clinical Development*, *10*, 105–112. https://doi.org/10.1016/j.omtm.2018.06.006

Sinkunas, T., Gasiunas, G., Fremaux, C., Barrangou, R., Horvath, P., & Siksnys, V. (2011). Cas3 is a single-stranded DNA nuclease and ATP-dependent helicase in the CRISPR/Cas immune system. *The EMBO Journal*, *30*(7), 1335–1342. https://doi.org/10.1038/emboj.2011.41

Skuk, D., & Tremblay, J. P. (2003). Myoblast transplantation: The current status of a potential therapeutic tool for myopathies. *Journal of Muscle Research & Cell Motility*, *24*(4), 287–302. https://doi.org/10.1023/A:1025425823322

Smith, A. D., Koreska, J., & Moseley, C. F. (1989). Progression of scoliosis in Duchenne muscular dystrophy. *The Journal of Bone and Joint Surgery. American Volume*, *71*(7), 1066–1074.

Smith, E. C., Conklin, L. S., Hoffman, E. P., Clemens, P. R., Mah, J. K., Finkel, R. S., Guglieri, M., Tulinius, M., Nevo, Y., Ryan, M. M., Webster, R., Castro, D., Kuntz, N. L., Kerchner, L.,

Morgenroth, L. P., Arrieta, A., Shimony, M., Jaros, M., Shale, P., … CINRG VBP15 and DNHS Investigators. (2020). Efficacy and safety of vamorolone in Duchenne muscular dystrophy: An 18-month interim analysis of a non-randomized open-label extension study. *PLoS Medicine*, *17*(9), e1003222. https://doi.org/10.1371/journal.pmed.1003222

Smith, H. O., & Welcox, K. W. (1970). A Restriction enzyme from Hemophilus influenzae: I. Purification and general properties. *Journal of Molecular Biology*, *51*(2), 379–391. https://doi.org/10.1016/0022-2836(70)90149-X

Soblechero-Martín, P., López-Martínez, A., de la Puente-Ovejero, L., Vallejo-Illarramendi, A., & Arechavala-Gomeza, V. (2021). Utrophin modulator drugs as potential therapies for Duchenne and Becker muscular dystrophies. *Neuropathology and Applied Neurobiology*, *47*(6), 711–723. https://doi.org/10.1111/nan.12735

Solid Biosciences. (2022). *Solid Biosciences Presents New SGT-001 IGNITE DMD Study Results at World Muscle Society 2022 Congress Demonstrating Improvements in Ambulatory Function* [Text]. Solid Biosciences. https://www.solidbio.com/about/media/press-releases/solid-biosciences-presents-new-sgt-001-ignite-dmd-study-results-at-world-muscle-society-2022-congress-demonstrating-improvements-in-ambulatory-function

Stadtmauer, E. A., Fraietta, J. A., Davis, M. M., Cohen, A. D., Weber, K. L., Lancaster, E., Mangan, P. A., Kulikovskaya, I., Gupta, M., Chen, F., Tian, L., Gonzalez, V. E., Xu, J., Jung, I., Melenhorst, J. J., Plesa, G., Shea, J., Matlawski, T., Cervini, A., … June, C. H. (2020). CRISPR-engineered T cells in patients with refractory cancer. *Science*, *367*(6481), eaba7365. https://doi.org/10.1126/science.aba7365

Sternberg, S. H., Redding, S., Jinek, M., Greene, E. C., & Doudna, J. A. (2014). DNA interrogation by the CRISPR RNA-guided endonuclease Cas9. *Nature*, *507*(7490), 62–67. https://doi.org/10.1038/nature13011

Stinson, B. M., Moreno, A. T., Walter, J. C., & Loparo, J. J. (2020). A Mechanism to Minimize Errors during Non-homologous End Joining. *Molecular Cell*, *77*(5), 1080-1091.e8. https://doi.org/10.1016/j.molcel.2019.11.018

Summit Therapeutics. (2019). *Phaseout DMD: A Phase 2 Clinical Study to Assess the Activity and Safety of Utrophin Modulation With Ezutromid in Ambulatory Paediatric Male Subjects With Duchenne Muscular Dystrophy (SMT C11005)* (Clinical Trial Registration No. NCT02858362). clinicaltrials.gov. https://clinicaltrials.gov/ct2/show/NCT02858362

Sun, C., Serra, C., Lee, G., & Wagner, K. R. (2020). Stem cell-based therapies for Duchenne muscular dystrophy. *Experimental Neurology*, *323*, 113086. https://doi.org/10.1016/j.expneurol.2019.113086

Sun, N., & Zhao, H. (2014). Seamless correction of the sickle cell disease mutation of the HBB gene in human induced pluripotent stem cells using TALENs. *Biotechnology and Bioengineering*, *111*(5), 1048–1053. https://doi.org/10.1002/bit.25018

Sweeney, N. P., & Vink, C. A. (2021). The impact of lentiviral vector genome size and producer cell genomic to gag-pol mRNA ratios on packaging efficiency and titre. *Molecular Therapy - Methods & Clinical Development*, *21*, 574–584. https://doi.org/10.1016/j.omtm.2021.04.007

Symington, L. S. (2016). Mechanism and regulation of DNA end resection in eukaryotes. *Critical Reviews in Biochemistry and Molecular Biology*, *51*(3), 195–212. https://doi.org/10.3109/10409238.2016.1172552

Symington, L. S., & Gautier, J. (2011). Double-strand break end resection and repair pathway choice. *Annual Review of Genetics*, *45*, 247–271. https://doi.org/10.1146/annurev-genet-110410-132435

Tabebordbar, M., Lagerborg, K. A., Stanton, A., King, E. M., Ye, S., Tellez, L., Krunnfusz, A.,

Tavakoli, S., Widrick, J. J., Messemer, K. A., Troiano, E. C., Moghadaszadeh, B., Peacker, B. L., Leacock, K. A., Horwitz, N., Beggs, A. H., Wagers, A. J., & Sabeti, P. C. (2021). Directed evolution of a family of AAV capsid variants enabling potent muscle-directed gene delivery across species. *Cell*, *184*(19), 4919-4938.e22. https://doi.org/10.1016/j.cell.2021.08.028

Tabebordbar, M., Zhu, K., Cheng, J. K. W., Chew, W. L., Widrick, J. J., Yan, W. X., Maesner, C., Wu, E. Y., Xiao, R., Ran, F. A., Cong, L., Zhang, F., Vandenberghe, L. H., Church, G. M., & Wagers, A. J. (2016). In vivo gene editing in dystrophic mouse muscle and muscle stem cells. *Science (New York, N.Y.)*, *351*(6271), 407–411. https://doi.org/10.1126/science.aad5177

Tabebordbar, M., Zhu, K., Cheng, J., Widrick, J., Yan, W., Xiao, R., Vandenberghe, L., Zhang, F., & Wagers, A. (2016). 483. In Vivo DMD Gene Editing in Muscles and Muscle Stem Cells of Dystrophic Mice. *Molecular Therapy*, *24*, S191–S192. https://doi.org/10.1016/S1525-0016(16)33292-0

Takeda, S. (2001). [Development of new therapy on muscular dystrophy]. *Rinsho Shinkeigaku = Clinical Neurology*, *41*(12), 1154–1156.

Takemitsu, M., Ishiura, S., Koga, R., Kamakura, K., Arahata, K., Nonaka, I., & Sugita, H. (1991). Dystrophin-related protein in the fetal and denervated skeletal muscles of normal and mdx mice. *Biochemical and Biophysical Research Communications*, *180*(3), 1179–1186. https://doi.org/10.1016/s0006-291x(05)81320-8

Takeshima, Y., Yagi, M., Okizuka, Y., Awano, H., Zhang, Z., Yamauchi, Y., Nishio, H., & Matsuo, M. (2010). Mutation spectrum of the dystrophin gene in 442 Duchenne/Becker muscular dystrophy cases from one Japanese referral center. *Journal of Human Genetics*, *55*(6), Article 6. https://doi.org/10.1038/jhg.2010.49

Takeshima, Y., Yagi, M., Wada, H., Ishibashi, K., Nishiyama, A., Kakumoto, M., Sakaeda, T., Saura, R., Okumura, K., & Matsuo, M. (2006). Intravenous infusion of an antisense

oligonucleotide results in exon skipping in muscle dystrophin mRNA of Duchenne muscular dystrophy. *Pediatric Research*, *59*(5), 690–694. https://doi.org/10.1203/01.pdr.0000215047.51278.7c

Takeshima, Y., Yagi, M., Wada, H., & Matsuo, M. (2005). Intraperitoneal administration of phosphorothioate antisense oligodeoxynucleotide against splicing enhancer sequence induced exon skipping in dystrophin mRNA expressed in mdx skeletal muscle. *Brain & Development*, *27*(7), 488–493. https://doi.org/10.1016/j.braindev.2004.12.006

Tasca, F., Wang, Q., & Gonçalves, M. A. F. V. (2020). Adenoviral Vectors Meet Gene Editing: A Rising Partnership for the Genomic Engineering of Human Stem Cells and Their Progeny. *Cells*, *9*(4), Article 4. https://doi.org/10.3390/cells9040953

Tawil, R., & Van Der Maarel, S. M. (2006). Facioscapulohumeral muscular dystrophy. *Muscle & Nerve*, *34*(1), 1–15. https://doi.org/10.1002/mus.20522

Tebas, P., Stein, D., Tang, W. W., Frank, I., Wang, S. Q., Lee, G., Spratt, S. K., Surosky, R. T., Giedlin, M. A., Nichol, G., Holmes, M. C., Gregory, P. D., Ando, D. G., Kalos, M., Collman, R. G., Binder-Scholl, G., Plesa, G., Hwang, W.-T., Levine, B. L., & June, C. H. (2014). Gene Editing of CCR5 in Autologous CD4 T Cells of Persons Infected with HIV. *New England Journal of Medicine*, *370*(10), 901–910. https://doi.org/10.1056/NEJMoa1300662

The DMD mutations database. (n.d.). Retrieved 22 March 2019, from http://www.umd.be/DMD/W_DMD/index.html

Thompson, A. A., Walters, M. C., Kwiatkowski, J., Rasko, J. E. J., Ribeil, J.-A., Hongeng, S., Magrin, E., Schiller, G. J., Payen, E., Semeraro, M., Moshous, D., Lefrere, F., Puy, H., Bourget, P., Magnani, A., Caccavelli, L., Diana, J.-S., Suarez, F., Monpoux, F., … Cavazzana, M. (2018). Gene Therapy in Patients with Transfusion-Dependent β-Thalassemia. *New England Journal of Medicine*, *378*(16), 1479–1493. https://doi.org/10.1056/NEJMoa1705342

Tinsley, J. M., Fairclough, R. J., Storer, R., Wilkes, F. J., Potter, A. C., Squire, S. E., Powell, D. S., Cozzoli, A., Capogrosso, R. F., Lambert, A., Wilson, F. X., Wren, S. P., De Luca, A., & Davies, K. E. (2011). Daily treatment with SMTC1100, a novel small molecule utrophin upregulator, dramatically reduces the dystrophic symptoms in the mdx mouse. *PloS One*, *6*(5), e19189. https://doi.org/10.1371/journal.pone.0019189

Tomé FM, Evangelista T, Leclerc A, Sundada Y, Manole E, Estournet B, Barois A, Campbell K, & Fardeau M. (1994). Congenital muscular dystrophy with merosin deficiency. *C R Acad Sci*, *III*, 351–357.

Tornabene, P., & Trapani, I. (2020). Can Adeno-Associated Viral Vectors Deliver Effectively Large Genes? *Human Gene Therapy*, *31*(1–2), 47–56. https://doi.org/10.1089/hum.2019.220

Torrente, Y., Belicchi, M., Marchesi, C., D'antona, G., Cogiamanian, F., Pisati, F., Gavina, M., Giordano, R., Tonlorenzi, R., Fagiolari, G., Lamperti, C., Porretti, L., Lopa, R., Sampaolesi, M., Vicentini, L., Grimoldi, N., Tiberio, F., Songa, V., Baratta, P., … Bresolin, N. (2007). Autologous Transplantation of Muscle-Derived CD133+ Stem Cells in Duchenne Muscle Patients. *Cell Transplantation*, *16*(6), 563–577. https://doi.org/10.3727/000000007783465064

Tuffery-Giraud, S., Béroud, C., Leturcq, F., Yaou, R. B., Hamroun, D., Michel-Calemard, L., Moizard, M.-P., Bernard, R., Cossée, M., Boisseau, P., Blayau, M., Creveaux, I., Guiochon-Mantel, A., de Martinville, B., Philippe, C., Monnier, N., Bieth, E., Khau Van Kien, P., Desmet, F.-O., … Claustres, M. (2009). Genotype-phenotype analysis in 2,405 patients with a dystrophinopathy using the UMD-DMD database: A model of nationwide knowledgebase. *Human Mutation*, *30*(6), 934–945. https://doi.org/10.1002/humu.20976

Tuggle, K. L., Birket, S. E., Cui, X., Hong, J., Warren, J., Reid, L., Chambers, A., Ji, D., Gamber, K., Chu, K. K., Tearney, G., Tang, L. P., Fortenberry, J. A., Du, M., Cadillac, J. M., Bedwell, D. M., Rowe, S. M., Sorscher, E. J., & Fanucchi, M. V. (2014). Characterization of Defects in Ion Transport and Tissue Development in Cystic Fibrosis Transmembrane

Conductance Regulator (CFTR)-Knockout Rats. *PLOS ONE*, *9*(3), e91253. https://doi.org/10.1371/journal.pone.0091253

Tycko, J., Barrera, L. A., Huston, N. C., Friedland, A. E., Wu, X., Gootenberg, J. S., Abudayyeh, O. O., Myer, V. E., Wilson, C. J., & Hsu, P. D. (2018a). Pairwise library screen systematically interrogates Staphylococcus aureus Cas9 specificity in human cells. *Nature Communications*, *9*, 2962. https://doi.org/10.1038/s41467-018-05391-2

Tycko, J., Barrera, L. A., Huston, N. C., Friedland, A. E., Wu, X., Gootenberg, J. S., Abudayyeh, O. O., Myer, V. E., Wilson, C. J., & Hsu, P. D. (2018b). Publisher Correction: Pairwise library screen systematically interrogates Staphylococcus aureus Cas9 specificity in human cells. *Nature Communications*, *9*, 3542. https://doi.org/10.1038/s41467-018-06029-z

UCSC Genome Browser Home. (n.d.). Retrieved 23 February 2023, from http://genome.ucsc.edu/index.html

van der Pijl, E. M., van Putten, M., Niks, E. H., Verschuuren, J. J. G. M., Aartsma-Rus, A., & Plomp, J. J. (2018). Low dystrophin levels are insufficient to normalize the neuromuscular synaptic abnormalities of mdx mice. *Neuromuscular Disorders*, *28*(5), 427–442. https://doi.org/10.1016/j.nmd.2018.02.013

Vannucci, L., Lai, M., Chiuppesi, F., Ceccherini-Nelli, L., & Pistello, M. (2013). Viral vectors: A look back and ahead on gene transfer technology. *The New Microbiologica*, *36*(1), 1–22.

Vengalil, S., Preethish-Kumar, V., Polavarapu, K., Mahadevappa, M., Sekar, D., Purushottam, M., Thomas, P. T., Nashi, S., & Nalini, A. (2017). Duchenne Muscular Dystrophy and Becker Muscular Dystrophy Confirmed by Multiplex Ligation-Dependent Probe Amplification: Genotype-Phenotype Correlation in a Large Cohort. *Journal of Clinical Neurology (Seoul, Korea)*, *13*(1), 91–97. https://doi.org/10.3988/jcn.2017.13.1.91

Verdera, H. C., Kuranda, K., & Mingozzi, F. (2020). AAV Vector Immunogenicity in Humans:

A Long Journey to Successful Gene Transfer. *Molecular Therapy*, *28*(3), 723–746. https://doi.org/10.1016/j.ymthe.2019.12.010

Vieitez, I., Gallano, P., González-Quereda, L., Borrego, S., Marcos, I., Millán, J. M., Jairo, T., Prior, C., Molano, J., Trujillo-Tiebas, M. J., Gallego-Merlo, J., García-Barcina, M., Fenollar, M., & Navarro, C. (2017). Mutational spectrum of Duchenne muscular dystrophy in Spain: Study of 284 cases. *Neurologia (Barcelona, Spain)*, *32*(6), 377–385. https://doi.org/10.1016/j.nrl.2015.12.009

Vincent, N., Ragot, T., Gilgenkrantz, H., Couton, D., Chafey, P., Grégoire, A., Briand, P., Kaplan, J.-C., Kahn, A., & Perricaudet, M. (1993). Long–term correction of mouse dystrophic degeneration by adenovirus–mediated transfer of a minidystrophin gene. *Nature Genetics*, *5*(2), Article 2. https://doi.org/10.1038/ng1093-130

Vohra, R., Batra, A., Forbes, S. C., Vandenborne, K., & Walter, G. A. (2017). Magnetic Resonance Monitoring of Disease Progression in mdx Mice on Different Genetic Backgrounds. *The American Journal of Pathology*, *187*(9), 2060–2070. https://doi.org/10.1016/j.ajpath.2017.05.010

Voit, T., Topaloglu, H., Straub, V., Muntoni, F., Deconinck, N., Campion, G., De Kimpe, S. J., Eagle, M., Guglieri, M., Hood, S., Liefaard, L., Lourbakos, A., Morgan, A., Nakielny, J., Quarcoo, N., Ricotti, V., Rolfe, K., Servais, L., Wardell, C., … Kraus, J. E. (2014). Safety and efficacy of drisapersen for the treatment of Duchenne muscular dystrophy (DEMAND II): An exploratory, randomised, placebo-controlled phase 2 study. *The Lancet. Neurology*, *13*(10), 987–996. https://doi.org/10.1016/S1474-4422(14)70195-4

Volpers, C., & Kochanek, S. (2004). Adenoviral vectors for gene transfer and therapy. *The Journal of Gene Medicine*, *6*(S1), S164–S171. https://doi.org/10.1002/jgm.496

Vouillot, L., Thélie, A., & Pollet, N. (2015). Comparison of T7E1 and Surveyor Mismatch

Cleavage Assays to Detect Mutations Triggered by Engineered Nucleases. *G3 Genes|Genomes|Genetics*, *5*(3), 407–415. https://doi.org/10.1534/g3.114.015834

Vulin, A., Barthélémy, I., Goyenvalle, A., Thibaud, J.-L., Beley, C., Griffith, G., Benchaouir, R., le Hir, M., Unterfinger, Y., Lorain, S., Dreyfus, P., Voit, T., Carlier, P., Blot, S., & Garcia, L. (2012). Muscle function recovery in golden retriever muscular dystrophy after AAV1-U7 exon skipping. *Molecular Therapy: The Journal of the American Society of Gene Therapy*, *20*(11), 2120–2133. https://doi.org/10.1038/mt.2012.181

Wagner, K. R., Kuntz, N. L., Koenig, E., East, L., Upadhyay, S., Han, B., & Shieh, P. B. (2021). Safety, tolerability, and pharmacokinetics of casimersen in patients with Duchenne muscular dystrophy amenable to exon 45 skipping: A randomized, double-blind, placebo-controlled, dose-titration trial. *Muscle & Nerve*, *64*(3), 285–292. https://doi.org/10.1002/mus.27347

Wang, B., Li, J., & Xiao, X. (2000). Adeno-associated virus vector carrying human minidystrophin genes effectively ameliorates muscular dystrophy in mdx mouse model. *Proceedings of the National Academy of Sciences*, *97*(25), 13714–13719. https://doi.org/10.1073/pnas.240335297

Wang, D., Li, S., Gessler, D. J., Xie, J., Zhong, L., Li, J., Tran, K., Van Vliet, K., Ren, L., Su, Q., He, R., Goetzmann, J. E., Flotte, T. R., Agbandje-McKenna, M., & Gao, G. (2018). A Rationally Engineered Capsid Variant of AAV9 for Systemic CNS-Directed and Peripheral Tissue-Detargeted Gene Delivery in Neonates. *Molecular Therapy. Methods & Clinical Development*, *9*, 234–246. https://doi.org/10.1016/j.omtm.2018.03.004

Wang, H.-X., Li, M., Lee, C. M., Chakraborty, S., Kim, H.-W., Bao, G., & Leong, K. W. (2017). CRISPR/Cas9-Based Genome Editing for Disease Modeling and Therapy: Challenges and Opportunities for Nonviral Delivery. *Chemical Reviews*, *117*(15), 9874–9906. https://doi.org/10.1021/acs.chemrev.6b00799

Wang, P., Li, H., Zhu, M., Han, R. Y., Guo, S., & Han, R. (2023). Correction of DMD in human iPSC-derived cardiomyocytes by base-editing-induced exon skipping. *Molecular Therapy - Methods & Clinical Development*, *28*, 40–50. https://doi.org/10.1016/j.omtm.2022.11.010

Wang, X., Tokheim, C., Gu, S. S., Wang, B., Tang, Q., Li, Y., Traugh, N., Zeng, Z., Zhang, Y., Li, Z., Zhang, B., Fu, J., Xiao, T., Li, W., Meyer, C. A., Chu, J., Jiang, P., Cejas, P., Lim, K., … Liu, X. S. (2021). In vivo CRISPR screens identify the E3 ligase Cop1 as a modulator of macrophage infiltration and cancer immunotherapy target. *Cell*, *184*(21), 5357-5374.e22. https://doi.org/10.1016/j.cell.2021.09.006

Wang, Y., Hao, L., Wang, H., Santostefano, K., Thapa, A., Cleary, J., Li, H., Guo, X., Terada, N., Ashizawa, T., & Xia, G. (2018). Therapeutic Genome Editing for Myotonic Dystrophy Type 1 Using CRISPR/Cas9. *Molecular Therapy*, *26*(11), 2617–2630. https://doi.org/10.1016/j.ymthe.2018.09.003

Wanisch, K., & Yáñez-Muñoz, R. J. (2009). Integration-deficient Lentiviral Vectors: A Slow Coming of Age. *Molecular Therapy*, *17*(8), 1316–1332. https://doi.org/10.1038/mt.2009.122

Waterkamp, D. A., Müller, O. J., Ying, Y., Trepel, M., & Kleinschmidt, J. A. (2006). Isolation of targeted AAV2 vectors from novel virus display libraries. *The Journal of Gene Medicine*, *8*(11), 1307–1319. https://doi.org/10.1002/jgm.967

Watts, G. (2007). Nobel prize is awarded for work leading to “knockout mouse”. *BMJ : British Medical Journal*, *335*(7623), 740. https://doi.org/10.1136/bmj.39364.367361.DB

Way, M., Pope, B., Cross, R. A., Kendrick-Jones, J., & Weeds, A. G. (1992). Expression of the N-terminal domain of dystrophin in E. coli and demonstration of binding to F-actin. *FEBS Letters*, *301*(3), 243–245. https://doi.org/10.1016/0014-5793(92)80249-G

Weber, T. (2021). Anti-AAV Antibodies in AAV Gene Therapy: Current Challenges and Possible Solutions. *Frontiers in Immunology*, *12*.

https://www.frontiersin.org/articles/10.3389/fimmu.2021.658399

Weinmann, J., Weis, S., Sippel, J., Tulalamba, W., Remes, A., El Andari, J., Herrmann, A.-K., Pham, Q. H., Borowski, C., Hille, S., Schönberger, T., Frey, N., Lenter, M., VandenDriessche, T., Müller, O. J., Chuah, M. K., Lamla, T., & Grimm, D. (2020). Identification of a myotropic AAV by massively parallel in vivo evaluation of barcoded capsid variants. *Nature Communications*, *11*(1), 5432. https://doi.org/10.1038/s41467-020-19230-w

Welch, E. M., Barton, E. R., Zhuo, J., Tomizawa, Y., Friesen, W. J., Trifillis, P., Paushkin, S., Patel, M., Trotta, C. R., Hwang, S., Wilde, R. G., Karp, G., Takasugi, J., Chen, G., Jones, S., Ren, H., Moon, Y.-C., Corson, D., Turpoff, A. A., … Sweeney, H. L. (2007). PTC124 targets genetic disorders caused by nonsense mutations. *Nature*, *447*(7140), 87–91. https://doi.org/10.1038/nature05756

Wells, D. J. (2019). What is the level of dystrophin expression required for effective therapy of Duchenne muscular dystrophy? *Journal of Muscle Research and Cell Motility*, *40*(2), 141–150. https://doi.org/10.1007/s10974-019-09535-9

Wojtal, D., Kemaladewi, D. U., Malam, Z., Abdullah, S., Wong, T. W. Y., Hyatt, E., Baghestani, Z., Pereira, S., Stavropoulos, J., Mouly, V., Mamchaoui, K., Muntoni, F., Voit, T., Gonorazky, H. D., Dowling, J. J., Wilson, M. D., Mendoza-Londono, R., Ivakine, E. A., & Cohn, R. D. (2016). Spell Checking Nature: Versatility of CRISPR/Cas9 for Developing Treatments for Inherited Disorders. *American Journal of Human Genetics*, *98*(1), 90–101. https://doi.org/10.1016/j.ajhg.2015.11.012

Wong, N., Liu, W., & Wang, X. (2015). WU-CRISPR: Characteristics of functional guide RNAs for the CRISPR/Cas9 system. *Genome Biology*, *16*(1), 218. https://doi.org/10.1186/s13059-015-0784-0

Wright, J. F. (2008). Manufacturing and characterizing AAV-based vectors for use in clinical

studies. *Gene Therapy*, *15*(11), 840–848. https://doi.org/10.1038/gt.2008.65

Xu, L., Lau, Y. S., Gao, Y., Li, H., & Han, R. (2019). Life-Long AAV-Mediated CRISPR Genome Editing in Dystrophic Heart Improves Cardiomyopathy without Causing Serious Lesions in mdx Mice. *Molecular Therapy*, *27*(8), 1407–1414. https://doi.org/10.1016/j.ymthe.2019.05.001

Xu, L., Park, K. H., Zhao, L., Xu, J., El Refaey, M., Gao, Y., Zhu, H., Ma, J., & Han, R. (2016). CRISPR-mediated Genome Editing Restores Dystrophin Expression and Function in mdx Mice. *Molecular Therapy: The Journal of the American Society of Gene Therapy*, *24*(3), 564–569. https://doi.org/10.1038/mt.2015.192

Yang, H., Ren, S., Yu, S., Pan, H., Li, T., Ge, S., Zhang, J., & Xia, N. (2020). Methods Favoring Homology-Directed Repair Choice in Response to CRISPR/Cas9 Induced-Double Strand Breaks. *International Journal of Molecular Sciences*, *21*(18), 6461. https://doi.org/10.3390/ijms21186461

Yang, Z., Steentoft, C., Hauge, C., Hansen, L., Thomsen, A. L., Niola, F., Vester-Christensen, M. B., Frödin, M., Clausen, H., Wandall, H. H., & Bennett, E. P. (2015). Fast and sensitive detection of indels induced by precise gene targeting. *Nucleic Acids Research*, *43*(9), e59. https://doi.org/10.1093/nar/gkv126

Ye, L., Wang, J., Tan, Y., Beyer, A. I., Xie, F., Muench, M. O., & Kan, Y. W. (2016). Genome editing using CRISPR-Cas9 to create the HPFH genotype in HSPCs: An approach for treating sickle cell disease and β-thalassemia. *Proceedings of the National Academy of Sciences of the United States of America*, *113*(38), 10661–10665. https://doi.org/10.1073/pnas.1612075113

Yin, C., Zhang, T., Qu, X., Zhang, Y., Putatunda, R., Xiao, X., Li, F., Xiao, W., Zhao, H., Dai, S., Qin, X., Mo, X., Young, W.-B., Khalili, K., & Hu, W. (2017). In Vivo Excision of HIV-1 Provirus by saCas9 and Multiplex Single-Guide RNAs in Animal Models. *Molecular Therapy:*

The Journal of the American Society of Gene Therapy, *25*(5), 1168–1186. https://doi.org/10.1016/j.ymthe.2017.03.012

Yin, D., Ling, S., Wang, D., Dai, Y., Jiang, H., Zhou, X., Paludan, S. R., Hong, J., & Cai, Y. (2021). Targeting herpes simplex virus with CRISPR–Cas9 cures herpetic stromal keratitis in mice. *Nature Biotechnology*, *39*(5), Article 5. https://doi.org/10.1038/s41587-020-00781-8

Yin, H., Kauffman, K. J., & Anderson, D. G. (2017). Delivery technologies for genome editing. *Nature Reviews Drug Discovery*, *16*(6), Article 6. https://doi.org/10.1038/nrd.2016.280

Yin, H., Song, C.-Q., Dorkin, J. R., Zhu, L. J., Li, Y., Wu, Q., Park, A., Yang, J., Suresh, S., Bizhanova, A., Gupta, A., Bolukbasi, M. F., Walsh, S., Bogorad, R. L., Gao, G., Weng, Z., Dong, Y., Koteliansky, V., Wolfe, S. A., … Anderson, D. G. (2016). Therapeutic genome editing by combined viral and non-viral delivery of CRISPR system components in vivo. *Nature Biotechnology*, *34*(3), 328–333. https://doi.org/10.1038/nbt.3471

Yin, H., Song, C.-Q., Suresh, S., Wu, Q., Walsh, S., Rhym, L. H., Mintzer, E., Bolukbasi, M. F., Zhu, L. J., Kauffman, K., Mou, H., Oberholzer, A., Ding, J., Kwan, S.-Y., Bogorad, R. L., Zatsepin, T., Koteliansky, V., Wolfe, S. A., Xue, W., … Anderson, D. G. (2017). Structure-guided chemical modification of guide RNA enables potent non-viral in vivo genome editing. *Nature Biotechnology*, *35*(12), Article 12. https://doi.org/10.1038/nbt.4005

Yiu, E. M., & Kornberg, A. J. (2015). Duchenne muscular dystrophy. *Journal of Paediatrics and Child Health*, *51*(8), 759–764. https://doi.org/10.1111/jpc.12868

Young, C. S., Hicks, M. R., Ermolova, N. V., Nakano, H., Jan, M., Younesi, S., Karumbayaram, S., Kumagai-Cresse, C., Wang, D., Zack, J. A., Kohn, D. B., Nakano, A., Nelson, S. F., Miceli, M. C., Spencer, M. J., & Pyle, A. D. (2016). A Single CRISPR-Cas9 Deletion Strategy that Targets the Majority of DMD Patients Restores Dystrophin Function in hiPSC-Derived Muscle Cells. *Cell Stem Cell*, *18*(4), 533–540. https://doi.org/10.1016/j.stem.2016.01.021

Young, C. S., Mokhonova, E., Quinonez, M., Pyle, A. D., & Spencer, M. J. (2017). Creation of a Novel Humanized Dystrophic Mouse Model of Duchenne Muscular Dystrophy and Application of a CRISPR/Cas9 Gene Editing Therapy. *Journal of Neuromuscular Diseases*, *4*(2), 139–145. https://doi.org/10.3233/JND-170218

Yuasa, K., Ishii, A., Miyagoe, Y., & Takeda, S. (1997). Introduction of rod-deleted dystrophin cDNA, delta DysM3, into mdx skeletal muscle using adenovirus vector. *Nihon rinsho Japanese journal of clinical medicine*, *55*(12), 3148–3153.

Yue, Y., Pan, X., Hakim, C. H., Kodippili, K., Zhang, K., Shin, J.-H., Yang, H. T., McDonald, T., & Duan, D. (2015). Safe and bodywide muscle transduction in young adult Duchenne muscular dystrophy dogs with adeno-associated virus. *Human Molecular Genetics*, *24*(20), 5880–5890. https://doi.org/10.1093/hmg/ddv310

Zatz, M., Rapaport, D., Vainzof, M., Passos-Bueno, M. R., Bortolini, E. R., Pavanello, R. de C., & Peres, C. A. (1991). Serum creatine-kinase (CK) and pyruvate-kinase (PK) activities in Duchenne (DMD) as compared with Becker (BMD) muscular dystrophy. *Journal of the Neurological Sciences*, *102*(2), 190–196. https://doi.org/10.1016/0022-510x(91)90068-i

Zeng, B., Zhou, M., Liu, B., Shen, F., Xiao, R., Su, J., Hu, Z., Zhang, Y., Gu, A., Wu, L., Liu, X., & Liang, D. (2021). Targeted addition of mini-dystrophin into rDNA locus of Duchenne muscular dystrophy patient-derived iPSCs. *Biochemical and Biophysical Research Communications*, *545*, 40–45. https://doi.org/10.1016/j.bbrc.2021.01.056

Zentilin, L., Marcello, A., & Giacca, M. (2001). Involvement of cellular double-stranded DNA break binding proteins in processing of the recombinant adeno-associated virus genome. *Journal of Virology*, *75*(24), 12279–12287. https://doi.org/10.1128/JVI.75.24.12279-12287.2001

Zetsche, B., Heidenreich, M., Mohanraju, P., Fedorova, I., Kneppers, J., DeGennaro, E. M.,

Winblad, N., Choudhury, S. R., Abudayyeh, O. O., Gootenberg, J. S., Wu, W. Y., Scott, D. A., Severinov, K., van der Oost, J., & Zhang, F. (2017). Multiplex gene editing by CRISPR–Cpf1 using a single crRNA array. *Nature Biotechnology*, *35*(1), 31–34. https://doi.org/10.1038/nbt.3737

Zhang, F., Cong, L., Lodato, S., Kosuri, S., Church, G. M., & Arlotta, P. (2011). Efficient construction of sequence-specific TAL effectors for modulating mammalian transcription. *Nature Biotechnology*, *29*(2), 149–153. https://doi.org/10.1038/nbt.1775

Zhang, H.-X., Zhang, Y., & Yin, H. (2019). Genome Editing with mRNA Encoding ZFN, TALEN, and Cas9. *Molecular Therapy: The Journal of the American Society of Gene Therapy*, *27*(4), 735–746. https://doi.org/10.1016/j.ymthe.2019.01.014

Zhang, W., Cao, S., Martin, J. L., Mueller, J. D., Mansky, L. M., Zhang, W., Cao, S., Martin, J. L., Mueller, J. D., & Mansky, L. M. (2015). Morphology and ultrastructure of retrovirus particles. *AIMS Biophysics*, *2*(3), 343–369. https://doi.org/10.3934/biophy.2015.3.343

Zhang, Y., Li, H., Nishiyama, T., McAnally, J. R., Sanchez-Ortiz, E., Huang, J., Mammen, P. P. A., Bassel-Duby, R., & Olson, E. N. (2022). A humanized knockin mouse model of Duchenne muscular dystrophy and its correction by CRISPR-Cas9 therapeutic gene editing. *Molecular Therapy. Nucleic Acids*, *29*, 525–537. https://doi.org/10.1016/j.omtn.2022.07.024

Zhang, Y., Long, C., Li, H., McAnally, J. R., Baskin, K. K., Shelton, J. M., Bassel-Duby, R., & Olson, E. N. (2017). CRISPR-Cpf1 correction of muscular dystrophy mutations in human cardiomyocytes and mice. *Science Advances*, *3*(4). https://doi.org/10.1126/sciadv.1602814

Zimowski, J. G., Massalska, D., Holding, M., Jadczak, S., Fidziańska, E., Lusakowska, A., Kostera-Pruszczyk, A., Kamińska, A., & Zaremba, J. (2014). MLPA based detection of mutations in the dystrophin gene of 180 Polish families with Duchenne/Becker muscular dystrophy. *Neurologia I Neurochirurgia Polska*, *48*(6), 416–422.

https://doi.org/10.1016/j.pjnns.2014.10.004

Zincarelli, C., Soltys, S., Rengo, G., & Rabinowitz, J. E. (2008). Analysis of AAV Serotypes 1–9 Mediated Gene Expression and Tropism in Mice After Systemic Injection. *Molecular Therapy*, *16*(6), 1073–1080. https://doi.org/10.1038/mt.2008.76

8. APPENDICES.

8.1. APPENDIX A: ALIGNMENT OF INTRONS 18 AND 55 FROM *DMD/DMD* GENES ON EMBOSS.

8.1.1. ALIGNMENT OF HUMAN AND MOUSE INTRON 18 OF DMD/DMD GENE.

```
#######################################
# Rundate: Wed  3 Jul 2019 11:37:35
# Commandline: needle
# Report_file: stdout
#######################################

#=======================================
#
# Aligned_sequences: 2
# 1: EMBOSS_001 = human
# 2: EMBOSS_002 = mouse
# Matrix: EDNAFULL
# Gap_penalty: 10.0
# Extend_penalty: 0.5
#
# Length: 21103
# Identity:    9379/21103 (44.4%)
# Similarity:  9379/21103 (44.4%)
# Gaps:        9007/21103 (42.7%)
# Score: 16789.0
#
#
#=======================================

EMBOSS_001         1 GTAGGTTATGCATTAAT-TTTTATATCTGTACTCATTTTGTGCTGCTTGT     49
                     |||||||.||||.|||| ||.|||.|||      ||||||||||.|||..
EMBOSS_002         1 GTAGGTTCTGCACTAATCTTATATTTCT------ATTTTGTGCTACTTTC     44

EMBOSS_001        50 AAACTCCGTGCTTTGT------TATCTGTGATTCTACTAG---TTGATAG     90
                     |||.|   |..|||||      |||||.||   ||| |||   |||  |.
EMBOSS_002        45 AAATT---TCTTTTGTTTCAAATATCTCTG---CTA-TAGTCATTG--AA     85
```

```
EMBOSS_001        91 ATTAATGC--GATATGATAAAATGCTGTTTAATATGTCTTAAAAATATAT    138
                     |||||.||  ||.||  |||||||.||||.|||||   .||||||   ..
EMBOSS_002        86 ATTAAAGCAAGAAAT--TAAAATGATGTTCAATAT---ATAAAAA---CG    127

EMBOSS_001       139 TTTAGATAGCAAGCAAAATCAAAGT-TTAGCCTAGAGAAATCT-TTCTGG    186
                     ||||||||.||||||||| ||||.| ||.||.|||||..||||| ||| .|
EMBOSS_002       128 TTTAGATAACAAGCAAAA-CAAAATATTTGCTTAGAACAATCTATTC-AG    175

EMBOSS_001       187 ATTCT--TCTGTGACTGACTAGGGATGCCTTTCATAAACTTGCCTTAACT    234
                     |.|||  ||||.|       |.||||   |||||||    |.|.|||||.|
EMBOSS_002       176 AATCTGCTCTGGG-------AAGGAT---TTTCATA----TTCTTTAAGT    211

EMBOSS_001       235 TACATATTATTTAACTA--TAAA---TTA-ACAATTTTATGACATTAATT    278
                     ||...|.||||||  ||  ||||   ||| |||||.||          ||
EMBOSS_002       212 TAAGCAGTATTTA--TAAGTAAAACTTTAGACAATCTT----------TT    249

EMBOSS_001       279 CATGTTGATTCCT-TTCATATATTGTTCAGTTACCTCAAGGCTAGGCCTT    327
                     .|||||||..|.| |||||||.|||||              ||.||||||
EMBOSS_002       250 TATGTTGACACGTCTTCATATCTTGTT--------------CTTGGCCTT    285

EMBOSS_001       328 AGCAA--------GCTTGCACAATTGTGTAGAGTAGAATAGAGTAAAATA    369
                     ||.||        .||.|.|||   |||.|.|.|||||.||.||||||||
EMBOSS_002       286 AGGAAATTTATTGACTCGAACA---GTGCATAATAGAGTACAGTAAAAT-    331

EMBOSS_001       370 GAATAGAATAATATAATAGATACCATACTTAGGAC----TATGTCTATAT    415
                        ||| .||..|||||||           |||||    ||||
EMBOSS_002       332 ---TAG-GTAGGATAATAG-----------AGGACACTTTATG-------    359

EMBOSS_001       416 TCTCACTGTTATCAAAA---AGTTATTCTTTATAGTCATCAATCCCTGTC    462
                        ||   ..|||||||   ||     .||||||||.|||||||.||||||
EMBOSS_002       360 ---CA---GGATCAAAATGCAG-----GTTTATAGTAATCAATACCTGTC    398

EMBOSS_001       463 ATATATTATAAGATATTTAAATTTCCCAGATATTTGTGTATT-----ACC    507
                     ||| ||.|||||.|||||||.||.||||.|.|||||.|.|.||||     |.|
EMBOSS_002       399 ATA-ATGATAACATATTTAGATATCCAATATATGTATATATTGCATCAGC    447

EMBOSS_001       508 TTATCTAAAAGGTGTCCAGCCTTTCTCAGTATTTACATGTGGTAGCAGAT    557
                     |.|.|||||||.||  ||..||.|.||||||||||..|.|.|||||.||||
```

```
EMBOSS_002       448 TCACCTAAAAGATG--CATACTCTTTCAGTATTTATTTTTAGTAGTAGAT    495

EMBOSS_001       558 GCTTTG-----TTGTTGTTATTCCCTTCTGAGATGGTATA-ATTTTGACC    601
                     ..||||     |     |||.||||.|.||||||  |||| |.|||||.||
EMBOSS_002       496 AGTTTGCAACAT-----TTACTCCCGTATGAGAT--TATATAGTTTGTCC    538

EMBOSS_001       602 TTCACAAATAGGAAGTTTTATTAATTTAATATAAAG---TGTTTCATCTT    648
                     ||||| ||.||||..|||||||.|..|  |||||.|   ||||..|||||
EMBOSS_002       539 TTCAC-AACAGGATTTTTTATTGACCT--TATAAGGAGCTGTTATATCTT    585

EMBOSS_001       649 TAGGGAAATTGTCAAGGCTTTAA-TTTTTCA-------ACTTCAACAATT    690
                     ..||.||||||.|||..|||.||| |||.|||       ||||.||||||.
EMBOSS_002       586 CTGGAAAATTATCAGTGCTATAACTTTATCACTTTATCACTTTAACAATG    635

EMBOSS_001       691 ATAAAACATGCATCTTT-GTATATGATTGGCAACAGTTTGTAAAATAATT    739
                     |.|||||||.|||.||| |.||.|.||||.|||.|.|||.|||.|.|.||
EMBOSS_002       636 ACAAAACATTCATTTTTAGAATGTAATTGACAATATTTTCTAAGAGACTT    685

EMBOSS_001       740 TCTAACAGGCAGAGTAAATAAAATGGGCAAGCT--AA----------GGA    777
                     ||||..|..||..||||||||||||..||.|||  ||          .||
EMBOSS_002       686 TCTACAAATCAAGGTAAATAAAATGAACAGGCTCCAAATACTGAATGTGA    735

EMBOSS_001       778 AGCAGT-----------GATTATGAATGAAAAGTAATT-----TTA--TC    809
                     |.||.|           .|||||.|||...||||.|||     |||  ||
EMBOSS_002       736 AACATTTTTAGTCGTAAAATTATCAATTGCAAGTGATTAGCAATTAAGTC    785

EMBOSS_001       810 CATG----GGAAGATCACATCTAAATGTGT---------ATATTAGCT--    844
                     | ||    |.||.||| .||.||.||||..|         |||.||.||
EMBOSS_002       786 C-TGACCTGTAAAATC-TATTTACATGATTTCACATATAATACTACCTCA    833

EMBOSS_001       845 ---CTG----------------CTTTTTCTCTCCAGGATTCA--------    867
                        |||                |||||..| .||||.|.|||
EMBOSS_002       834 CAGCTGGGCATGGTGGCCCACCCTTTTAAT-GCCAGCACTCAGGAGGCAG    882

EMBOSS_001       868 ------------AATCA---------------------------TGAAGT    878
                                 ||.||                           |||..|
EMBOSS_002       883 AGGCAGAGGCAGAAGCAGGCAGATTTCTGAGCTCGAGGCCAGCCTGATCT    932

EMBOSS_001       879 TCAAA--------------ATCCATTG-TATTTTTA--------------    899
```

```
                     .||||              |.|||..| ||    ||
EMBOSS_002       933 ACAAAGTGAGTCCCAGGACAGCCAGGGCTA----TACAGAGAAACCCTGT    978

EMBOSS_001       900 ----------CAAATACGAC------------------------------    909
                               ||||.||.||
EMBOSS_002       979 CTCGAAAAACCAAAAACAACAACAACAACAACAACAAAAACAAACAAACA   1028

EMBOSS_001       910 --------------------------------------------------    909

EMBOSS_002      1029 AACAAAAAAACAAAAAAACCCAACCAACCAACCAAACAAACAAAAAACCA   1078

EMBOSS_001       910 --ATTACCTCAGTTACAAGCTAATGT----TTGCT---GTTGGG-GTTGG    949
                       |.|||||||        |.|||||    |||||   .||||| ..|.|
EMBOSS_002      1079 ATACTACCTCA--------CAAATGTGCTCTTGCTTTACTTGGGACATAG   1120

EMBOSS_001       950 AACTTTTGGAGA---TCAAC----AAAAGATATATAT------ATATTCT    986
                     ||.|.||..|||   .||||    ||||||||| |.|      |.|||.|
EMBOSS_002      1121 AAATGTTAAAGAGTTACAACTCCAAAAAGATAT-TCTGGTGTAACATTGT   1169

EMBOSS_001       987 GGAAAAAAAATCTATTTTT------AGGCTGCTTGAAAAAGG---GAAG   1027
                     ||        |.|.||||||      |||    .|||..||||   .|||
EMBOSS_002      1170 GG--------TTTTTTTTTTTTTTTAAGG----ATGACCAAGGAGCAAAG   1207

EMBOSS_001      1028 ACAATTTTGTCACCAGTTCTTCTTAG-AGTTAAC--TACTTATAAATTGG   1074
                     |.|| |||||   |||||    |||| |.|.|||  ||.|||| |.||..
EMBOSS_002      1208 ATAA-TTTGT---CAGTT----TTAGAAATCAACATTATTTAT-ATTTAC   1248

EMBOSS_001      1075 AA---AGCTATTTTAAATTACTTATTAGCTTTATA--AGACATGCTGTTG   1119
                     ||   ||.||||||.||||.|.||||||||.||||  |||.|||.|.||.
EMBOSS_002      1249 AACATAGTTATTTTGAATTTCATATTAGCTGTATATTAGATATGATTTTA   1298

EMBOSS_001      1120 TCAGCATTATAATAGACTA-------TTC---TA--------AATTGTTT   1151
                     |.|.|||.||||||||||       |||   ||        ||||.||.
EMBOSS_002      1299 TTATCATCATAATAGACTAGGATAGTTTCACATAATAAGTATAATTTTTA   1348

EMBOSS_001      1152 CAAGAAATGGGAAATATGAAAACTGAAA------GATAATATATAATTGT   1195
                     ..|..||||..|.||.|.|||| |.|||      |||||||.|.||
EMBOSS_002      1349 TTATTAATGTTACATTTTAAAA-TTAAATTAAATGATAATACACAA----   1393
```

```
EMBOSS_001      1196 AGAAAATTAGCTAAA------TGTCTTTTTTCAAGTATACTTCTTTGAAG   1239
                       |||||.|.||.||      |||||...||.|||||||..|.|||.||.
EMBOSS_002      1394 --AAAATCAACTTAACAAGATTGTCTAGGTTTAAGTATATGTATTTTAAA   1441

EMBOSS_001      1240 GTAAATTGTTTGCTGTGATTTTTGAGGGA-AATTC---------------   1273
                     .||.| |||||  ..||||||||.|   | |||||
EMBOSS_002      1442 ATATA-TGTTT--CATGATTTTTAA---ATAATTCTTTTTAAATAAGATT   1485

EMBOSS_001      1274 --TCTTAGTAGTAGAAAAT----TAATT----AGAGAACATTAAACTCTA   1313
                       |.|||.|..|.||||||    |||||    ||||..|||| ..|||.|
EMBOSS_002      1486 CTTTTTATTCTTTGAAAATTTCCTAATTTTAGAGAGTTCATT-TCCTCAA   1534

EMBOSS_001      1314 ATT-TTCAAACATCATGTTAC-ATATTTGACCA----CTGAAAGTA-TGA   1356
                     ||| ||  |.|||| |.|||| .|||     ||    |||||..|| ||
EMBOSS_002      1535 ATTATT--ATCATC-TCTTACTCTAT-----CACTTTCTGAACTTAGTG-   1575

EMBOSS_001      1357 ATCACCTTTAGCATTTTATTAGCTAATTAAA-----------AAT-----   1390
                      ||..||||.|   .||||||  |||.||.|           |||
EMBOSS_002      1576 -TCCTCTTTTG---ATTATTA--TAAATACACTCCCAAGTCCAATTTGTG   1619

EMBOSS_001      1391 ----GATTATTGATCT---------GATTC--------------------   1407
                         .|.|||||||||         ||.||
EMBOSS_002      1620 CTACCAGTATTGATCTTGTGTTGAGGAATCAAAAAGAGCATGGGCAATCT   1669

EMBOSS_001      1408 -TCA-----CAGGTTTCTGTCTGATA------CTCTGTT------TCTTT   1439
                      |||     ||.|||.||...|||.|      |||..||      ||.||
EMBOSS_002      1670 ATCAGCAGCCATGTTCCTAAATGACAGTGACTCTCCATTTAGCAATCATT   1719

EMBOSS_001      1440 ATCT-----------------TAA-------GAC----------------   1449
                     |.||                 |||       |||
EMBOSS_002      1720 AACTGCTAGTAGCTGCACCTCTAAGTGTAAGGACTTGGTAGCCCCACATC   1769

EMBOSS_001      1450 ------------AGCTTTGA---------------TCAGG------CAG-   1465
                                 ||||||||               |||||      |||
EMBOSS_002      1770 TCTTCATGGTGTAGCTTTGAATTGCTTAATCTGATTCAGGTAACCACAGA   1819

EMBOSS_001      1466 ----GAAAATTGAGTTTCAGTAT-----------------TTCAGGAATA   1494
                         |.||||||.|   ||  |||                 |..||||.|.
EMBOSS_002      1820 TGCTGTAAATTCA---TC--TATGTGAAAGCCATGTGGGGTCAAGGATTC   1864
```

```
EMBOSS_001      1495 TAAATTTTAAAAGT-TCCTGTTA--TCATGTGGTTACCTATGAACGAT--   1539
                     |.|||.|.|.|..| ||.||.||  ||||..|.|  |.|||||..|..|
EMBOSS_002      1865 TGAATCTCACAGCTCTCTTGCTAATTCATTAGAT--CTTATGGTCTGTCC   1912

EMBOSS_001      1540 TTGCATT-TTATCAATAGCGGGAG----CATGAC----------------   1568
                     ||.|||| ||.|       |||||    ||||||
EMBOSS_002      1913 TTTCATTCTTCT-------GGGAGGATCCATGACCCTTGCATGAGGGAAT   1955

EMBOSS_001      1569 -----------AGACTAAAAGTGCTCA---AAAATGTACAGTGT------   1598
                                .||.|.||| ||.|.|   |.|||| |||||.|
EMBOSS_002      1956 GTTGATATAGATGATTCAAA-TGGTTACTGACAATG-ACAGTTTAATCAG   2003

EMBOSS_001      1599 ---TTTATAACTGGT-------------------TTTCTGATACC----T   1622
                        |.|||..|||.|                  ||||||..|||    |
EMBOSS_002      2004 ATATATATTCCTGCTTACTCACTACATAAAGAAGTTTCTGTGACCAAAGT   2053

EMBOSS_001      1623 TGTGAGTCCC------CTTTATCAA------------------CCTGAAC   1648
                     ||.|.||..|      ||.|||||.|                  ||.|||.
EMBOSS_002      2054 TGAGTGTAGCATAAATCTATATCTAAAAAATAGAGGTCAATTTCCAGAAT   2103

EMBOSS_001      1649 TCTCATTTTCTGAA--ATAACAACATCA-----TTCT-------------   1678
                     ..||||||     |  ||||||..||.|     ||||
EMBOSS_002      2104 GATCATTT-----AGCATAACACTATTAGTAGGTTCTACCACAGAGAGCA   2148

EMBOSS_001      1679 TAATTGCC-------------TACTAAAATGGGACATGAAAT--------   1707
                     |.|.|.||             ||| |.|||||.|||||||||||
EMBOSS_002      2149 TGAGTTCCCAGGCTTTGATTTTAC-AGAATGAGACATGAAATACCTCCAG   2197

EMBOSS_001      1708 --GATTTAC----------------CTTGA-TGTTTCCCTC-AGGCTAAA   1737
                       ||   ||                |||||||    |||| ||  ||||
EMBOSS_002      2198 TGGA---ACAGCTAATCAGGGAGCACTTGACTG----CCTCTAG--TAAA   2238

EMBOSS_001      1738 TACAATGTATATGTAT---TAAAAATCCTTGAGAGAA-----ATACATAA   1779
                        ||| ||.|||.||   ||||.|||      ||.|     ||.|| ||
EMBOSS_002      2239 ---AAT-TACATGAATGAGTAAATATC------AGCAACCCTATTCA-AA   2277

EMBOSS_001      1780 TTACGATGATGGATATTCAGT--------------TT--------AGGTA   1807
                     |||.||..||.| |..|.|||              ||        .||||
```

```
EMBOSS_002      2278 TTATGAGTATAG-TGCTAAGTGAGACTGCTGTGCCTTCTCTCTCCTGGTA   2326

EMBOSS_001      1808 TCCATT----AATTTTCTTACTAC--ATCA-------CCTGTCGG-----   1839
                     .||..|    |||.||.|.| |||  |.||       |||.|.||
EMBOSS_002      2327 GCCTGTGTAGAATCTTATGA-TACCAAGCAAGCAAAGCCTCTGGGAACTA   2375

EMBOSS_001      1840 --TTTAT----AT---AGG-TTATTTTAAACTCATGGCAGGCTCGTGTCA   1879
                       |||.|    ||   ||| |.|||||  .|| |||   .||| ||.| |
EMBOSS_002      2376 GTTTTCTGGTCATTCAAGGATGATTTT--GCT-ATG---TGCT-GTAT-A   2417

EMBOSS_001      1880 TGAACTTGAGTGGTATT------AAT--CCATTT---------TTAGGCT   1912
                     | ||..|||||||.|||      |||  ..||||         |||  ||
EMBOSS_002      2418 T-AAAGTGAGTGGAATTTTCAGCAATAGGAATTTAACACCTACTTA--CT   2464

EMBOSS_001      1913 TT---CATTAAAAATG-GCTAAAT----ATTACCAACTCTTTTTTTTT-   1953
                     .|   |||   ||||| |.|||||    ||     |.|||.|.||.|||
EMBOSS_002      2465 ATGGACAT---AAATGAGATAAATGCCGAT-----AGTCTGTGTTGTTTA   2506

EMBOSS_001      1954 -----CACTTGATTTCTTC---------ATTTACTCCTTTCAAATGCTTT   1989
                          |.||.||   ||||         |||||     ||||.|||
EMBOSS_002      2507 GATGCCTCTGGA---CTTCCCTGACCATATTTA-----TTCATATG----   2544

EMBOSS_001      1990 TAGCCA--GTTAGCATATCATAGGCCATTATCAGTTCA------------   2025
                      ||.||  .||.|||| |.|||    ||| |||||| |
EMBOSS_002      2545 -AGACATTCTTGGCAT-TGATA----ATT-TCAGTT-AGTAACCCTGTGT   2586

EMBOSS_001      2026 ---------------ATT--TCACATATTTTAAAAAATGCAACTTTT---   2055
                                        |||  ||||..|||     |||.|| ||||||
EMBOSS_002      2587 TGTTTGGTAAAAAGCATTGATCACCCATT-----AAAGGC-ACTTTTTGA   2630

EMBOSS_001      2056 -------------AAACCTAAGCATAATTTTTTC-----CTTGTATATTC   2087
                                  ||||||.||    |.|||.||     |||.|.|||.|
EMBOSS_002      2631 TTTTTGTAGCTAAAAACCTGAG----AATTTCTCAAAGACTTCTTTATAC   2676

EMBOSS_001      2088 ACCATGTGTTTTATTCATATAATTAACAAAATGCACT------TATTTAA   2131
                     |     |.||..|||.|||.|.|.|..||..|.||||      |||...|
EMBOSS_002      2677 A-----TCTTGAATTGATACATTCAGTAATGTCCACTGACCTCTATCCCA   2721

EMBOSS_001      2132 AAGCTGACTTCTGGGGGACCTTTCTAATCAAA--------AT---GTCCC   2170
```

```
                        |..|||.|| |.|.||.|.||.||.||||        ||   .||||
EMBOSS_002      2722 ---CCCACTCCT-GTGCACTTCTCCAAACAAACCTTTAGCATATCCTCCC   2767

EMBOSS_001      2171 TGGACCACCAATAT----TGACATTATC----------TGGGAGCTTG--   2204
                     |...||.|| ||.|    ||.|||||.|          |||||  |||
EMBOSS_002      2768 TTTGCCTCC-ATGTAACCTGTCATTAACTTTCCTAACTTGGGA--TTGAA   2814

EMBOSS_001      2205 --TCAGA-----AA-----TATAGAATCTCAGAACTAGTAAATC----AA   2238
                       |||||     ||     |||.||   .||.||.||.|  |||    ||
EMBOSS_002      2815 AATCAGAGGTGGAAAAATTTATTGA---GCAAAAATACT--ATCTAGAAA   2859

EMBOSS_001      2239 AATTTG-CCTTTGAACATTTTCCAGGAGTGATTCCTATGCACATA---AA   2284
                     ||.||| .||.|    ||.||.|| |||   ||..|||| |.|||   ||
EMBOSS_002      2860 AAGTTGTTCTCT----ATGTTTCA-GAG---TTTTTATG-ATATATTCAA   2900

EMBOSS_001      2285 GGCTCGAGAATCAATATTCT--AGGACACAAT--CTT--GGGCTTTA---   2325
                     ..|.|.||.||.|||.|.||  |.||...|||  .||  |||.||||
EMBOSS_002      2901 TACACAAGTATAAATTTACTTAATGATTGAATTAATTGAGGGTTTTATTG   2950

EMBOSS_001      2326 --------CGTAGCAGTTACCTCTCACATTCTTTTAGTATCACTGCAGTT   2367
                             .|||| ||||          ||||.|  ||||     ||.||
EMBOSS_002      2951 ATTTGATTTGTAG-AGTT----------TTCTGT--GTAT-----CATTT   2982

EMBOSS_001      2368 CAGACT-TTTTA-----GCTGAATTAACAGATTCTAA----CAACTCGAA   2407
                     .||.|| |||||     |||   ||||     ||.||    .||.|.|||
EMBOSS_002      2983 TAGTCTGTTTTAAAATGGCT---TTAA-----TCAAATATGAAAATGGAA   3024

EMBOSS_001      2408 TTTTTTCCGTGAGCACATGGTA-ACGCAAA---------------TTGG-   2440
                     ||.|.|.|.|.||.|.|| .|| |..||||               ||||
EMBOSS_002      3025 TTCTATACTTTAGGAAAT-ATATATTCAAAATTTCCTCTTATCCTTTGGC   3073

EMBOSS_001      2441 TA-----------------------------------------------   2442
                     ||
EMBOSS_002      3074 TACTTAAAAGCATATATATATATATATATATATATATATATATATATATA   3123

EMBOSS_001      2443 -----ATGAAACAA-----------------------ACAAAAAA-----   2459
                          ||.||||||                      ||.|||||
EMBOSS_002      3124 TATATATAAAACAATGCCAATATGATAGACCAAAGGTACTAAAAATGTAC   3173
```

```
EMBOSS_001      2460 AATA--------GCCTCCTTTGATCACAAGGGGA----------------   2485
                     ||||        |||||..|     || ||||||
EMBOSS_002      3174 AATATGTTATATGCCTCGGT-----AC-AGGGGAACGCCGGGGCCAAAAA   3217

EMBOSS_001      2486 -TTTGATAAGGTGTCT-------------GAAA----ATCTGGAGCTTCT   2517
                      |..||..|||||..|             ||.|    ||..||..|||.|
EMBOSS_002      3218 GTGGGAGTAGGTGGGTAGGGGAGTCGGGGGAGAGGGTATGGGGGACTTTT   3267

EMBOSS_001      2518 GTGAGGTT--CTCTTGAAACC-----------------------------   2536
                     |   ||.|  |..|||||||.|
EMBOSS_002      3268 G---GGATAGCATTTGAAATCTAAATGAAGAAAATACCGAATAAAAAAGA   3314

EMBOSS_001      2537 ----------------------CTCC-CTATCAA-GTCGTTATGGTCATC   2562
                                           |.|| ||.|||| .|||  |...|||||
EMBOSS_002      3315 TATTTTTCAATACAATTTCAGTCACCACTCTCAACTTCG--AATCTCATC   3362

EMBOSS_001      2563 -----------TGGCTTTGGGCTCTT-----CC--------TCGGATCTG   2588
                                |.||.||  |.||||     ||        |.|||..||
EMBOSS_002      3363 TTCAGAAACAATAGCATT--GTTCTTAATTGCCCAGTAAAATAGGACATG   3410

EMBOSS_001      2589 CACTTGAT-TACCC--------TCTTTCAGAATCAACTTT-------TTA   2622
                     .|| |||| |.|||        .||||.||  ||||.|.|       |||
EMBOSS_002      3411 AAC-TGATGTGCCCAGAAATTGCCTTTAAG--TCAATTATATATGCATTA   3457

EMBOSS_001      2623 CCAACTCCCG--------------------CTGGAAATCATCA--AGCCA   2650
                     ...||||..|                    ||.||..||.|||  .||||
EMBOSS_002      3458 AACACTCTAGAGAGGAATAATGATGATTTTCTTGATTTCCTCATTTGCCA   3507

EMBOSS_001      2651 GTTAATG-------CTTCCAATT-------------TCTCCT----ATCT   2676
                     |||.|||       .||..||||             ||||.|    |...
EMBOSS_002      3508 GTTTATGTAAGTCATTTTGAATTCAGGGGAGGGTCGTCTCATGAAAAGAC   3557

EMBOSS_001      2677 CTGTTATCTGTGTTTACTACCCAGAG-TAT------------AGGTACCT   2713
                     |||.||  .|..||||||   ||.|| |||            |.|.|.|.
EMBOSS_002      3558 CTGGTA--GGAATTTACT---CATAGATATCCATTTAAAATGATGAAGCA   3602

EMBOSS_001      2714 CGAGTCTATATTTTCTTCTAT------CTTTCTCTCTCTGCTCTATTGCA   2757
                     ..|.|..||.|||||..||||      .|||||| ||.||.|.|.|||.|
EMBOSS_002      3603 TTATTTGATTTTTTCACCTATTGCTGAATTTCTC-CTATGTTATTTTGTA   3651
```

```
EMBOSS_001      2758 C------TTTCTTTTCCTTG------------CTTCACTTCCTCTGCTTT   2789
                     |      |...||.|..|||            |||.||||      .|||
EMBOSS_002      3652 CCTAAAGTGAGTTGTTGTTGAAATAAGAATATCTTAACTT------ATTT   3695

EMBOSS_001      2790 CCCACACTCTCACTCGGCAGGCTCAAAC-----ACTTTTTCAAGACAAGC   2834
                     |.||.|.|.||              |||     |.||||||.||....|||
EMBOSS_002      3696 CACATATTTTC--------------AACTGTGTAATTTTTAAACTTTAGC   3731

EMBOSS_001      2835 ------------CTTCTAT-----ATACTTATTCTACTCATACGTGGAAA   2867
                                 .||||||     ||| |..||.||.|||||  |||..|
EMBOSS_002      3732 ATAAGTTTATCTTTTCTATTGATGATA-TGTTTTTATTCATA--TGGTTA   3778

EMBOSS_001      2868 CCAA--------TTTTACCCTCGAAAACAGCACTATGACTCATCCCTCTT   2909
                     .|||        |.|||                ||||
EMBOSS_002      3779 ACAAAATGTTCTTATTA----------------TATG-------------   3799

EMBOSS_001      2910 TTGTACCCCCACTACTTTAGAATGACTGACTCCCTCACCTAA--------   2951
                      |||||..|        .|||.||.|         |.||.|||
EMBOSS_002      3800 -TGTAGAC--------ATAGGATAA---------TAACATAATTTTGGAG   3831

EMBOSS_001      2952 --------CAGAGACAATTC-------------CAAAGATACTGC-----   2975
                              ||||||.|||.|             ||.|.|||.|||
EMBOSS_002      3832 CTGCTAGTCAGAGAGAATGCAAATTTTTTTTCTCACACATAATGCCCAGT   3881

EMBOSS_001      2976 ------TCTCTTCATGTAAAAATACCT-----CTT-----TCCTT--GCA   3007
                           ||.|||.||       |||||     |||     |.|||  ||
EMBOSS_002      3882 ACCCAATCCCTTTAT-------TACCTGAGAACTTGTTTCTACTTAAGC-   3923

EMBOSS_001      3008 GTTA------ATGAC-TTCTGGCCATCTCTTA-TCTACCATTTTATCATA   3049
                      |||      ||.|| |.||.|.||  |.||| |.||.|||.|  ||.||
EMBOSS_002      3924 -TTAGAATATATCACATACTTGGCA--TTTTACTGTATCATCT--TCTTA   3968

EMBOSS_001      3050 TTTTCTCATTGAAGACTTTGGCACTAGGTTAGT--AGATTTTTTC-AACC   3096
                     |.|||||.|||||||||.|||..||.||||   ||  |..||||||| |...
EMBOSS_002      3969 TCTTCTTATTGAAGAATTTCTCAGTAGG---GTCAAACTTTTTTCTATTT   4015

EMBOSS_001      3097 AATTTGTTCTTTCATTCTTAGTTATTTCA----GCATCCCTGTGGGCTCT   3142
                     ||||.|.|.||| |.|.||||||||.|||    .||||    |||
```

```
EMBOSS_002      4016 AATTGGATATTT-AATATTAGTTATCTCAATACACATC----TGG-----   4055

EMBOSS_001      3143 CCTGCTATCACTAAAAGCTGTTGTTTC-----TCTC-ATCTC--------   3178
                                ||    |||| ||||     |||| |||||
EMBOSS_002      4056 ------------AA----TGTT-TTTCTGTATTCTCAATCTCATATTTGT   4088

EMBOSS_001      3179 CTCAATTCAGCTAC------CTGCT---------GTCATGATTGTATGTA   3213
                     |||||||||..|||      .||.|         |.|||.|||||   |.
EMBOSS_002      4089 CTCAATTCATTTACTTATATATGATCACAGTATAGGCATTATTGT---TG   4135

EMBOSS_001      3214 TGTAGA---------TGTTGTCATTGTCAGG----CA---TTAC------   3241
                     |.||.|         |.||.|.|||||||..|    ||   ||||
EMBOSS_002      4136 TTTAAAATCAACTATTATTATTATTGTCTTGACTACAATTTTACCCCCTC   4185

EMBOSS_001      3242 -------TCCATT-------------------------------------   3247
                            ||||||
EMBOSS_002      4186 CTCGCCTTCCATTTGCTTCTCTTCCCTTCCCGTTGCCATCTTCCCTCTGC   4235

EMBOSS_001      3248 ---------------------------------TCTTTAATAAGA-----   3259
                                                      |.|||||||.|||
EMBOSS_002      4236 CCTTTCTTTGTCTAGTCCTCCTCCACTCCTCAGTTTTTAATTAGAAAAGG   4285

EMBOSS_001      3260 ---------------------------------TTATTTTTTTG-----A   3271
                                                      .||||...|||     |
EMBOSS_002      4286 GCCAGCCTCCCATAAATATAAACCAACCCTGGCATATTAAGTTGCAGTAA   4335

EMBOSS_001      3272 GACGGAGTCTCGCT---------CTGTCCCAGGC-------------GG-   3298
                     |.|..|||.||.||         .||..|.||||             ||
EMBOSS_002      4336 GGCTTAGTATCTCTAGTATTAGAATGGGCAAGGCAACACATTATAATGGA   4385

EMBOSS_001      3299 ------------------------GAGTGCAGTGGC-GCGATCTCGGCT--   3322
                                             |||| |||.|.| ||   |.|.|||
EMBOSS_002      4386 AATGTTCCCAAAAGCCTGCAAAAGAGT-CAGAGACAGC---CACTGCTCC   4431

EMBOSS_001      3323 CACTGCAAGCTCTGCC-------TCCCGGGTT-CAC--------------   3350
                     |||||..||...|.||       |.||..||| |||
EMBOSS_002      4432 CACTGTTAGTAGTCCCACAAGAATACCATGTTACACAACTGTAACATATA   4481

EMBOSS_001      3351 --------GCC---GT---TCTCCTG-------CCT--CAGCCTCCTGAG   3377
```

```
                            |||   ||   ||||.||       |||  ..|.||.|||||
EMBOSS_002      4482 TGCAGAGGGCCTAGGTCAGTCTCATGCAGGCCCCCTTGTTGTCTGCTGAG   4531

EMBOSS_001      3378 TAGCTGGGACTACAGGCGCCCGCCACCACACCCGGCTAAGTTT-------   3420
                     |..|||.||            |||...|.|   .|||.|.|||
EMBOSS_002      4532 TCTCTGTGA------------GCCCTTATA---AGCTGACTTTTGTTTGT   4566

EMBOSS_001      3421 -------TTGTATTTTT--AGTAGAGACGGGGTTTCACCGTGTTAGCCAG   3461
                            ||.|||||||  ||    |||.||.||||.|.||| ||||
EMBOSS_002      4567 TTGTTTGTTTTATTTTTCCAG----GACAGGATTTCTCAGTG-TAGC---   4608

EMBOSS_001      3462 GATGGTCTCGATCTC-CT--------------------GACCTCGT---G   3487
                          |||.|||||| ||                    ||.|||.|   |
EMBOSS_002      4609 -----TCTGGATCTCACTCTATAGACCAGATTAACTTAGAACTCATAAAG   4653

EMBOSS_001      3488 ATCCGCCCGCCTCGGCCTCCCAAAGTGCTGGGATTACAGGCGTGAGCC--   3535
                     ||...||.|.|||.|||| .|.|||||||||.||.|.|.||.|..|||
EMBOSS_002      4654 ATAGACCTGACTCTGCCT-TCTAAGTGCTGGAATAAAAAGCATATGCCAA   4702

EMBOSS_001      3536 -ACCGCGCCCGGCACTAACA-----------------ACATATTTTAACA   3567
                      |||.|        |||.||                 |||.|.|||||.|
EMBOSS_002      4703 AACCAC--------CTAGCAAACTCTGTTTCTCAAGTACACACTTTAAGA   4744

EMBOSS_001      3568 AAACCAG-----CAGTTCTTCAT-AAGCTTTCCGTGCATCCTAAGTATAT   3611
                     ||..|||     ||  ||||||| .|| |.|||.|.||||        ||
EMBOSS_002      4745 AAGTCAGGGTTACA--TCTTCATGGAG-TATCCATACATC--------AT   4783

EMBOSS_001      3612 TTTCAACTTAGTT---------TCTCTCAACAATTATTTCAAGCTTTTGT   3652
                     |    |||||.||         ||||| ...||||||||..||..|||.|
EMBOSS_002      4784 T----ACTTAATTATACCTCGATCTCT-TTTAATTATTTTCAGGATTTAT   4828

EMBOSS_001      3653 GGGGTACCT-AAACTTTCCCCATCATTC--ACAGTAAATGATTTCATCTC   3699
                     ....|.||| ||||||||.||.| ||.|  |||||..||..||||  ||.
EMBOSS_002      4829 AATTTCCCTAAAACTTTCTCCCT-ATACCTACAGTGTATAGTTTC--CTA   4875

EMBOSS_001      3700 CTACTTTAG-TCGAAATAATATTTCTCAAATTGTGGGTATTTGTGCTCAT   3748
                     .|| ||||| ||.|||.|.||.||||.|||.|.|...||||.||.|.|||
EMBOSS_002      4876 TTA-TTTAGATCAAAACAGTACTTCTAAAAGTATATATATTGGTACCCAT   4924
```

```
EMBOSS_001      3749 AGTATCTTTTTTTTTTTTTTTTTTCCACCTGCC--TTAGTGGAAGAG---   3793
                     |||||||.|.|                 |||||  |.|.|.||.|||
EMBOSS_002      4925 AGTATCTGTGT-----------------CTGCCTGTCATTTGATGAGGGA   4957

EMBOSS_001      3794 -GCTATCCTTCACCTGGTTGAG----GCTCATCTCTGGGTGTGTGTTCTC   3838
                      |||||     ||.|..|||||    |.| |||.||            ..
EMBOSS_002      4958 TGCTAT-----ACATTTTTGAGATAAGGT-ATCACT------------AT   4989

EMBOSS_001      3839 AG---CAGCATCACTGAC----TATGTA-TTAAGCCACCTGGTTCCATTC   3880
                     ||   |||.||||||| |    |.|||| |.|||  |.|.|.|||||||.
EMBOSS_002      4990 AGACTCAGTATCACTG-CAGTGTCTGTACTAAAG--AACAGCTTCCATTT   5036

EMBOSS_001      3881 AGCT--------GTATATCCAGATTGTCAAAAATCTACATCC-CAGGTCT   3921
                     ||||        .|||.|.|||.||       .|||..|||| .||.|.|
EMBOSS_002      5037 AGCTACACACCCATATCTACAGTTT-------TTCTTTATCCTGAGATTT   5079

EMBOSS_001      3922 TTATCATTAGCTTTTAAAC--CGGTTGTGTTTTC-ATCTTTTAGAACGTG   3968
                     |||||||.|...|.|||||  |..|| ||||||| .|.|||||.||....
EMBOSS_002      5080 TTATCATCATAATCTAAACAACTCTT-TGTTTTCTGTGTTTTAAAAATCA   5128

EMBOSS_001      3969 TCCTCTCCTATAAACATGCATGTGAATAC--TTAACGTGCCTTATCTT--   4014
                     ||.|.||||..|||||  |||.|||||||  ||   ||.|..|| |||
EMBOSS_002      5129 TCTTGTCCTTGAAACA--CATTTGAATACCTTT---GTACACTA-CTTCA   5172

EMBOSS_001      4015 ACTCAATCCCTCTGTAGAACTAGAACCTATTAACCTCTTTTCCCTCATGA   4064
                     ||...||     ||.||.||.|.|..|||.|.|||..|||||||   |||
EMBOSS_002      5173 ACGAGAT-----TGAAGGACAACATTCTAGTCACCATTTTTCCC---TGA   5214

EMBOSS_001      4065 AATTCTTTCCTC-----GCTTGGCT--TA---------------------   4086
                     |.|||||..|||     |||| |||  ||
EMBOSS_002      5215 AGTTCTTCTCTCCATTAGCTT-GCTAGTATGTATTATGAATTTTTACATA   5263

EMBOSS_001      4087 -TGGGATAC-TGCTTTCTCCTG------------ATTAATTCTTTCACA-   4121
                      ||||..|| ||.|||.|  ||            ||||||| |..||||
EMBOSS_002      5264 TTGGGTCACTTGTTTTAT--TGAACAGTTTGAGAATTAATT-TCCCACAC   5310

EMBOSS_001      4122 TTCTTGGCCATT--------------TGTTCCAAT---------------   4142
                     |||||  ||.||              ||||.|.||
EMBOSS_002      5311 TTCTT--CCTTTATTTTTGTTTTAGATGTTTCTATATTCATTCTTATTAA   5358
```

```
EMBOSS_001      4143 ------------------------------TCAGT------------TTT-   4150
                                                     |||||            |||
EMBOSS_002      5359 ATTAGCATCTAAAGGTCTAAAGACTTGGCTCAGTGGTTAAGAACACTTTT   5408

EMBOSS_001      4151 --TCTT-TA------ATAT--------------------TGAT-------   4164
                       |||| ||      ||||                    ||||
EMBOSS_002      5409 GCTCTTGTAGAGAGCATATTACAGTTCCCAGAACCAATATGATAACTCAA   5458

EMBOSS_001      4165 -ATTCTTCAG----------AGTCCTTG-TG--------TTTC---TTCT   4191
                      ||..|.|||          |||.||.| ||        ||||   ||.|
EMBOSS_002      5459 AATCATCCAGAAATCTATCTAGTACTAGATGACAAAGACTTTCTGATTTT   5508

EMBOSS_001      4192 C-TGG---------ATAT----------CCTCATTCATTCCC--------   4213
                     | |||         ||||          |.| |||||..|||
EMBOSS_002      5509 CATGGGCAGCCAGTATATAGGTAGTACACAT-ATTCACACCCAGGTAAAA   5557

EMBOSS_001      4214 ----------------ATTCAAATAGTGTCTAT--------TTTCAAATG   4239
                                        |||.|||.|...|||||        |.|||.||.
EMBOSS_002      5558 CACTGAAACACATAAAATTTAAAAAAAATCTATAAGTATTATATCATATA   5607

EMBOSS_001      4240 AACCCTTC--------------TCC-------------CATATGTATCT-   4261
                     .|..||||              |||             ||||||||||
EMBOSS_002      5608 TATTCTTCCTCTAATATCTTTTTCCTAAATATCTTTAACATATGTATCTA   5657

EMBOSS_001      4262 -TTAGCTCATCTCTTTCTTAAAACCCAGGGTGATATATCTGACTGCCT--   4308
                      ||||.| ||.|     ||||       |.||||||.|.||..||..|
EMBOSS_002      5658 ATTAGGT-ATAT-----TTAA-------GCTGATATGTGTGTGTGTGTGT   5694

EMBOSS_001      4309 -----------------AATAGACAAT-----------------------   4318
                                      ..||||||||
EMBOSS_002      5695 GTGTGTGTGTGTGTGTGTGTAGACAATGGGATGTCATATAATGTCCATAA   5744

EMBOSS_001      4319 -----TTTAATT----------------GATGTTT--GCCTGAT------   4339
                          |.|||||                .|||.||  .||||||
EMBOSS_002      5745 TTATATGTAATTTCTATTATTTTAGGGATATGATTAAACCTGATTTACAA   5794

EMBOSS_001      4340 ----------------------ATCTT-AAAT----CTCATTATTTC----   4359
                                             ||||| ||||    ||.||||..||
```

```
EMBOSS_002      5795 AGAAAAATAATAGACAACAGAATCTTGAAATATACCTAATTAACTCTGTA   5844

EMBOSS_001      4360 -CCCAGACCTCCTTTATTCTTTCTCCTTCCTTTCACCCTGCAAACAGCCT   4408
                      ||.|     |.|||||||.||||||||..|||  .|.||        .|
EMBOSS_002      5845 TCCTA-----CTTTTATTCATTCTCCTTTATTT--TCTTG--------TT   5879

EMBOSS_001      4409 AATCCAGAAACCTG--AAAGTGAACTGGGTCACCTAGATTTTTCT-TCTA   4455
                     ||.|.||.||  ||  |.||..||.|.|||||.|||.|..|||.| |.||
EMBOSS_002      5880 AAACTAGGAA--TGGAACAGAAAAATAGGTCATCTACAAGTTTGTATATA   5927

EMBOSS_001      4456 C--CATTCACAAG----TCCAATCTTTGACAGGTTCAATTAATCCTGCTC   4499
                     |  .||| ||.||    |..||  ||||| |..|..|||||
EMBOSS_002      5928 CTTTATT-ACTAGATAATTAAA--TTTGA-ATATAGAATTA---------   5964

EMBOSS_001      4500 CCTTAACAGCTCTTGAAAATGTCCAGTTCTCTCTATTCCCATCACCAAAA   4549
                        |||      |..|||||.|..|.||   |||.|.||          |
EMBOSS_002      5965 ---TAA------TATAAAATATAGAATT---TCTCTGCC----------A   5992

EMBOSS_001      4550 CCCT----GA------TTAAAGCTAA----TATCATGGCTTACTTCTATT   4585
                     ||||    ||      |||||.||||    ||||.||..|..|.|.||||
EMBOSS_002      5993 CCCTACAGGATGAGAGTTAAACCTAACAATTATCTTGAATGTCATTTATT   6042

EMBOSS_001      4586 TAAGATGAC----AATCCTGTAGCTGGTCTTC-CCTGCTACTAATGTTAT   4630
                     ||||||.||    ||||.|.|||.||  |||| |||..||..|.|..|.|
EMBOSS_002      6043 TAAGATAACAATGAATCTTTTAGTTG--CTTCACCTTTTATCACTTCTCT   6090

EMBOSS_001      4631 ATTCCTTTAGA-AGTTTCCACATCACCTCTGC---CAAAGCGA-ACATAT   4675
                     ||||||||.|| |.|||    ||  |.|.|||   |.||||.| ||.||.
EMBOSS_002      6091 ATTCCTTTTGACATTTT----AT--CTTATGCCTGCTAAGCCAGACTTAC   6134

EMBOSS_001      4676 TAAAACCTCAGTATGATCTTGTTATGTCTGTGTTTATAATCATTTGAGAA   4725
                     ||.||..|.|.||.|||.|.|...|||||...|.||||.||| ||.|.||
EMBOSS_002      6135 TATAATTTAAATAGGATATAGAGTTGTCTCCATGTATAGTCA-TTAATAA   6183

EMBOSS_001      4726 AGATAAAATCAGCTACTGATAGCCAGGAGCTGGCTTGGCACTTCAGTTAG   4775
                     |.|.||||               |||.|||.|            ||||
EMBOSS_002      6184 ATACAAAA---------------CAGCAGCAG------------AGTT--   6204

EMBOSS_001      4776 GCTTTGGAACTGCTAGATGTTGCCTTGGTACTTGCAGCCAGTCTGATGTT   4825
```

```
                     |||                                     .||.|||
EMBOSS_002      6205 -CTT----------------------------------------AGAGGTT   6214

EMBOSS_001      4826 CTCTGGTTGAATATAAATAATTGCTAGGGATGCCAGCATCATACAGGGAC   4875
                     ||.||.|.|||       ||||
EMBOSS_002      6215 CTATGCTGGAA-------AATT----------------------------   6229

EMBOSS_001      4876 ACGCTATGGTTGTGATGAATCAAGATAAAAGCAAGACCACTTTGTA--AT   4923
                                            ||||      |||||||||.||||  ||
EMBOSS_002      6230 -----------------------AGAT------AAGACCACTCTGTATGAT   6251

EMBOSS_001      4924 CATATCTTAATACAG----TAAAAAAGAATAGCATTGTCCAAACCACGCC   4969
                     ||..|||||.|.|||    ||.|||              .|.||.|.||
EMBOSS_002      6252 CACTTCTTAGTGCAGGTACTACAAA--------------TACACTAAGC-   6286

EMBOSS_001      4970 TAATGACCAAATATCCCTCATCTTCATTAATGTGACACATTA--TGCTTC   5017
                      |||   ||..||||..|||||.|||.||.||   ||| |.|  || |||
EMBOSS_002      6287 -AAT---CAGGTATCTTTCATCATCACTATTG---CAC-TGAAGTG-TTC   6327

EMBOSS_001      5018 -TTTACTAAA--CACAGCTTTAGCCTCTCTGCATTCCTCCTGCCATCTAG   5064
                      ||||..|||  ||.|..|||||.|.||.|.|||.|.|  ||    ||.|
EMBOSS_002      6328 GTTTAAAAAAACCATAAATTTAGTCACTTTACATGCTT--TG----CTTG   6371

EMBOSS_001      5065 A----TAAAAATTTTAGGACACTCAACATAGAATTACCCT-----ACTTC   5105
                     |    ||||||.|..|..|.|||.||.||| ||||||.||     |.||
EMBOSS_002      6372 ATTTTTAAAAACTAAAAAAAACTGAATATA-AATTACTCTTAATGATTT-   6419

EMBOSS_001      5106 CTTACAACATCCAATCAAAACAAAGCCCCACTTCC------TTGATGCCC   5149
                      ||.||    ..|||.||||.|||....||.||||      .||.|.||
EMBOSS_002      6420 -TTGCA----TGAATTAAAAAAAAAAAAACAATTCCAACTCACTGGTTCC-   6463

EMBOSS_001      5150 CCTATCATGTAACACAAGCCCACATCTTAA-AATAAGCTTTTTCTAACAT   5198
                         ||||.||.|.|.||..| |||.|||| |||||||        |.||
EMBOSS_002      6464 ----TCATCTACCTCTAGTTC-CATTTTAATAATAAGC--------AAAT   6500

EMBOSS_001      5199 ACTCTTAGAGACAGCCC----CATAGTTCCCTGTGATGCGCTGTCTCTCT   5244
                         |||     |.|||    ||| ||||              ||.||||
EMBOSS_002      6501 ----TTA-----AACCCTTTTCAT-GTTC--------------TCACTCT   6526
```

```
EMBOSS_001      5245 CGTTGCAGCGTGTAAATCAGGTTAACTTTGATAACCTACAAATATGTTCC   5294
                           ||.||                  ||||.|||.|
EMBOSS_002      6527 -------GCCTG-------------------TAACTTACCA---------   6541

EMBOSS_001      5295 AGTGGTCTTTGTCTGA--AGAGTGTTGAC---AAAGTCCATGTCTTCATA   5339
                             |||||||  |||    ||.|   |||       |..|||.|
EMBOSS_002      6542 ---------TGTCTGAGTAGA----TGGCAGGAAA-------TAATCAAA   6571

EMBOSS_001      5340 TTGCCCTC-----TTAAAGCCTACAATGATCTAATTTTTGTTTACTTGCA   5384
                         |||     |.||| .|.||.|.||...|                |
EMBOSS_002      6572 -----CTCAGGGGTGAAA-TCAACCAAGAGGAA----------------A   6599

EMBOSS_001      5385 CA-----AACTAATTTCTAGCCATTTTCATTTCAATTCCATTGTTGGTTT   5429
                     ||     |||||                  |||||
EMBOSS_002      6600 CAAGAAGAACTA------------------TTCAA---------------   6616

EMBOSS_001      5430 CTAGAAATCAGGAC--TTGAACGTGAAGTGGCTTTGGTTTTATGTGTGTG   5477
                       || |||||  ||  |   |||.|.||     .|||||.|.||.|
EMBOSS_002      6617 --AG-AATCA--ACCAT---ACGAGGAG-----CTGGTTCTTTGAG----   6649

EMBOSS_001      5478 TGTTTTGTTTTTGGTGTCCCCATGCCTAA---CACACAAACCA-ATAAAT   5523
                                                ||   || |||||..| |||||.
EMBOSS_002      6650 ---------------------------AAAATCA-ACAAAATAGATAAAC   6671

EMBOSS_001      5524 GTTTA---------ACTATATGGATATGTGATTTTTTTCAGTTCTGCAGT   5564
                     ..|||         ||||.|.||               ||      |||
EMBOSS_002      6672 CCTTAGCTAGACTCACTAGAGGG---------------CA------CAG-   6699

EMBOSS_001      5565 TGCATTTAAGATAGGATC--TATTACCAATAT--GAA----------GGA   5600
                             .||.||.|||  .||||.|||.||  |||          .||
EMBOSS_002      6700 --------GGACAGCATCCTAATTAACAAAATCAGAAATGAAAAGGGAGA   6741

EMBOSS_001      5601 TAGAACAAC---TTCTGAATAAA----------CAT-------TTGAGTG   5630
                     .|.||||||   |.|||||.|||          |||       ||  .|.
EMBOSS_002      6742 CATAACAACAGATCCTGAAGAAATCCAAAACACCATCAAATCCTT--ATA   6789

EMBOSS_001      5631 AATAAG---TTACTCTTCAGTCAAAGGGTAAAACTTCATTATAGAATGTT   5677
                     .|||||   .||||||..||   |||..|.||||||.|       |.|||
EMBOSS_002      6790 CATAAGGATATACTCAACA---AAACTGGAAAACCT-------GGATG--   6827
```

```
EMBOSS_001      5678 AAATGAGTAGATTCTTAATTACTTCTAATGTCGGAGTAGGACTGATA-GA   5726
                     |||||...|.||          |||||            |||.|||| .|
EMBOSS_002      6828 AAATGGACAAAT----------TTCTA------------GACAGATACCA   6855

EMBOSS_001      5727 GAT-CCATAACTTATTT-----TCA---TAATGTTAGAGAAGA-------   5760
                     |.| ||| ||.|||..|     |||   |||||.|..| ||.|
EMBOSS_002      6856 GGTACCA-AAGTTAAATCGCGATCAGGTTAATGATCTA-AACAGTCCCAT   6903

EMBOSS_001      5761 ------TACAGAAAGAGGAG--GCTTTTAAGAG-GTGCTAA-TAAAGAAG   5800
                           ||.|.||||||.||  |...||||.|| .|.|.|| .|||.||.
EMBOSS_002      6904 ATCCCCTAAATAAAGAGAAGCAGTCATTAATAGTCTCCCAACCAAAAAAC   6953

EMBOSS_001      5801 G---AGGAACAGA--GGTCTGAAAGAG----GTTTTAATCAACTTATTGA   5841
                     |   ||||.||||  |||.|   ||||    |||.| ||||..|..|..|
EMBOSS_002      6954 GCCCAGGACCAGATGGGTTT---AGAGCAGAGTTCT-ATCAGATCTTCAA   6999

EMBOSS_001      5842 GGAATTTGTGAACT--TCCCTGTGTTGTATTCTTTGCTTCAAATTATTTT   5889
                     .|||     ||.||  |.||.||..||.|          ||||.|||...
EMBOSS_002      7000 AGAA-----GATCTAATTCCAGTTCTGCA----------CAAACTATGCC   7034

EMBOSS_001      5890 ACTTTTCTTCCTTTTATTGAATATTATATAACTTAGTTGCCTTTTATTTT   5939
                     ||                   |..|||.||
EMBOSS_002      7035 AC--------------------AAAATAGAA-------------------   7045

EMBOSS_001      5940 GCATTTAGGT----TACCCATTTTAGTATT-TAT---------AATTCTT   5975
                     ||| ..||||    ||||||.||   .||| |||         |||.||
EMBOSS_002      7046 GCA-GAAGGTACTCTACCCAATT---CATTCTATGAAGCCACAAATACT-   7090

EMBOSS_001      5976 CATGATCTCTA------AGAATTTTCATGAGCCCACAATAAT--------   6011
                      .|||||..|||      |||      |.||.||.||||..||
EMBOSS_002      7091 -CTGATACCTAAACCACAGA------AAGATCCAACAAAGATAGAGAACG   7133

EMBOSS_001      6012 ------CACTCAT---TTCTGATACATTGATTTACTGATGCCTTTTTTAG   6052
                           ||.| ||   ||.||| |.||.|||..|...||.|
EMBOSS_002      7134 TCAGACCAAT-ATCCCTTATGA-ATATCGATGCAAAAATAC---------   7172

EMBOSS_001      6053 CTTCTGTATACTTTACTTGAAATTATGTGTACTAAATAATTTAGAATCCG   6102
                       ||.|||           |||||.|   .||||||.|      ||||||.
```

```
EMBOSS_002      7173 --TCAGTA-----------AAATTCT---CACTAACT------GAATCCA   7200

EMBOSS_001      6103 AG-------TAGTGGCTTCTAAACATGTGATCAT--ATATTAAGCATATA   6143
                     ||       ||         |||||    |||||  .|..|||.|
EMBOSS_002      7201 AGAACACATTA---------AAACA----ATCATCCTTCCTAACC-----   7232

EMBOSS_001      6144 TTGAGTAAAGTA--TTTAGTATTCTA--TATG----------TTAATA-G   6178
                           |||||  |||  |||||.|  .|||          |||||| .
EMBOSS_002      7233 -------AAGTAGGTTT--TATTCCAGGGATGCAGGGATGGTTTAATATA   7273

EMBOSS_001      6179 CA-TAATTTAAAAACTGTATGTTATGAAATATAGATCCTTTCAAGATTAA   6227
                     || .|||..|..||| |||               ||||.||   |.||||
EMBOSS_002      7274 CAGAAATCCATCAAC-GTA---------------ATCCATT---GTTTAA   7304

EMBOSS_001      6228 A-------AATGTATATGCATGTAGACACA---TATATGATAAACAGCTC   6267
                     |       ||             |||||.|   .|.|||||   ||.|||
EMBOSS_002      7305 ACAAACTCAA-------------AGACAAAAACCACATGAT---CATCTC   7338

EMBOSS_001      6268 TATAGA-----AAAACACACATACAGAAAAGCTGGATGAC---TTCATGA   6309
                     ..||||     |.||..||..|   |||||..|..|..||   ||||||
EMBOSS_002      7339 GTTAGAGGCAGAGAAAGCATTT---GAAAAAATCCAACACCCATTCATG-   7384

EMBOSS_001      6310 GGACGAAATTTAAGA----GAAAGAT----AATATTGTCAAGGGCAGAGG   6351
                           ||..||||    |||||||    |||    ||||||.|.|
EMBOSS_002      7385 -------ATAAAAGACTTGGAAAGATCAGGAAT----TCAAGGCCCG---   7420

EMBOSS_001      6352 TATTTTTA--TGAACATG------GCAATTT--------CGAGGAGGTAA   6385
                          ||  |.||||||      |||||.|        |.||.||..||
EMBOSS_002      7421 ------TACCTAAACATGATAAAAGCAATCTACAGCAAACCAGTAGCCAA   7464

EMBOSS_001      6386 TTGTAGTTTTTATC--AGTACTGAATGTTGTAAGAAGCATTTTTCCTAAC   6433
                              .|||  ||||   ||||  ||.||||| ||
EMBOSS_002      7465 ----------CATCAAAGTA---AATG--GTGAGAAG-AT----------   7488

EMBOSS_001      6434 AAGAAG----------AAAATTAATGACTTTCTTTTCTTTTTTTTTTTTT   6473
                      .||||          |||||.|..||||
EMBOSS_002      7489 -GGAAGCAATCCCACTAAAATCAGGGACT---------------------   7516

EMBOSS_001      6474 TTTTTTTTTTCCCGAGACGGAGCCTCGCTCTGTCACCCAGGCTGGAGTGC   6523
```

```
                                  |||                  |.|||||
EMBOSS_002      7517 --------------AGA-------------------CAAGGCT-------   7526

EMBOSS_001      6524 AGTGGCGCAATCTTGGCTCACTGCAACTTCTGTCTCCCGGGTTCAAGCGA   6573
                                    ||.||||     ||||..||.|| ..|||||  .|
EMBOSS_002      7527 ---------------GCCCACT-----TTCTCCCTACC-TATTCAA--CA   7553

EMBOSS_001      6574 TTCTCCT----GCCTCAGCC--------TCCCG--------AGTAGCT--   6601
                     ||.|.||    |.||.||||        |  ||        ||.||.|
EMBOSS_002      7554 TTGTACTTGAAGTCTTAGCCAGAACAATT--CGACAACAAAAGGAGATCA   7601

EMBOSS_001      6602 ---GGGACTACA----GGCACGTGCCACCACGTCCAGCTAAT-----TTT   6639
                        |||| ||||    ||.|.|   .|..|.|||.|   |||     |||
EMBOSS_002      7602 AGGGGGA-TACAAATTGGAAAG---GAAGAAGTCAA---AATATCGCTTT   7644

EMBOSS_001      6640 TTG-------TATTTTTAGTAGAGACGGGGTTTCAACTGTTGGCC---AG   6679
                     |||       || |..|||||.|.|.|             ||.||   |.
EMBOSS_002      7645 TTGCAGATGATA-TGATAGTATATATG-------------TGACCCTAAA   7680

EMBOSS_001      6680 GATGGTCTCC-------ATCTCCT-GACCTTGTGATCCACCTGCCTTGGC   6721
                     .||    |||       |.||||| .|||   |||| .|.|.||.||||
EMBOSS_002      7681 AAT----TCCACCAAAGAACTCCTAAACC---TGAT-AAACAGCTTTGG-   7721

EMBOSS_001      6722 CTCCCAAAGT-GCTGG------GATTACAGGCGTGAGCCCCTGCGCCCGG   6764
                        ...|||| |||||      .||||                     .|
EMBOSS_002      7722 ---GTGAAGTGGCTGGATATAAAATTA---------------------AG   7747

EMBOSS_001      6765 CCAAA----TTAATGACTTT------CACAA--AAT-AACTTTTTG-TAC   6800
                     .||||    |.||||.|.||      |||||  ||| ||||..||| .|.
EMBOSS_002      7748 TCAAACAAGTCAATGGCCTTTCTCTACACAAAGAATAAACTGGTTGAGAA   7797

EMBOSS_001      6801 AGTAGTGAAGG-------------CTGA--AGTCTGACTGTAGGTATTTG   6835
                     ||.|.|.|.||             ||.|  ||||..|
EMBOSS_002      7798 AGAAATTAGGGATCCAACACCCTTCTCAATAGTCACA-------------   7834

EMBOSS_001      6836 GAGTAATATTGAAATACTGGAATTTGG----------ACTAA--------   6867
                      |.||||| |.||||||     .||||          |||||
EMBOSS_002      7835 -AATAATA-TAAAATAC-----CTTGGCCTGACTCTAACTAAGGAAGTGA   7877
```

```
EMBOSS_001      6868 --------TATGGTAATATTAGTCCAAATTTCATTGAAGATGAAATATAT   6909
                             ||||  ||.|..|.|.|||.|.||  ||||   ||||.|.||
EMBOSS_002      7878 AAGATCTCTATG--AAAAGAATTTCAAGTCTC--TGAA---GAAAGAAAT   7920

EMBOSS_001      6910 AAAAGAAAAT----GAAGACTGTGGTTGAAGAGTATTCAGATGATACCCT   6955
                     .||||||.||    |||      |||.|||        ||||  ..|||.
EMBOSS_002      7921 TAAAGAAGATCTCAGAA------GGTGGAA--------AGAT--CTCCCA   7954

EMBOSS_001      6956 TGGTCTTGAA-----------AACATGG--GAAAGGGGTAT---------   6983
                     ||.||.||.|           ||.||.|  .|||.||.|||
EMBOSS_002      7955 TGCTCATGGATTGGCAGGATCAATATAGTAAAAATGGCTATCTTGCCAAA   8004

EMBOSS_001      6984 ----ATATACAG-------------------------------------   6991
                         ||.|||||
EMBOSS_002      8005 AGCAATCTACAGATTCAATGCAATCCCCATCAAAATTTCAACTCAATTCT   8054

EMBOSS_001      6992 --TATGTATGTGTATGTACATATATGTACATACA--------CACACATA   7031
                       .|||.||..|.|.|.||| ||.||.|.||.||        .|||.|||
EMBOSS_002      8055 TCAATGAATTAGAAAGAACA-ATCTGCAAATTCATCTGGAATAACAAATA   8103

EMBOSS_001      7032 ------GAGAG---------------AGGA-------------GGTGGAA   7047
                           ||.||               ||||             ||||||
EMBOSS_002      8104 ACCTAGGATAGCAAAAACTCTTCTCAAGGATAAAAGAATCTCTGGTGGA-   8152

EMBOSS_001      7048 GATCAGAGCCA----TGAAATAAAAACAGAGAGGTAAGATTCTAC---CA   7090
                      ||||   |||    |||..|||      ||..|||     ||||   .|
EMBOSS_002      8153 -ATCA---CCATGCCTGACCTAA------AGCTGTA-----CTACAGAAA   8187

EMBOSS_001      7091 AGTTTTGACTCA---TGGATTGTATATGGTACTCTTTCTTTTCTACTTAG   7137
                     |.||.|||.|.|   ||       .||||||||                |
EMBOSS_002      8188 AATTGTGATTAAAACTG-------CATGGTACT----------------G   8214

EMBOSS_001      7138 GTATTG----------GTTTATTACCAGTATCAGGACTTGAGCTGAAGA-   7176
                     ||||.|          ||   ..||||||    |||.|.||..||||||
EMBOSS_002      8215 GTATAGCAGCAGACAAGT---AGACCAGT----GGAATAGAATTGAAGAC   8257

EMBOSS_001      7177 --AAAGATGGATGTGCCCA-ACTCCCTAACTAATATGCAGTTTCTAGAAT   7223
                       |.|.|||.|    |||| ||.||        ||||  ||..||.||.|
EMBOSS_002      8258 CCAGAAATGAA----CCCACACACC--------TATG--GTCACTTGATT   8293
```

```
EMBOSS_001      7224 ACTTTGA--------TTTAATCCATCC--------------------TTC   7245
                       |||||        .|.||.||||||                    |||
EMBOSS_002      8294 --TTTGACAAGGGAGCTAAAACCATCCAGTGGAAGAAAGACAGCATTTTC   8341

EMBOSS_001      7246 ---------------CTTTACTG------AGCTTGCAGCA----------   7264
                                    |...||||      |.|.||.||.|
EMBOSS_002      8342 AACAATTGGTGCTGGCACAACTGGTGGTTAACATGTAGAAGAATGCGAAT   8391

EMBOSS_001      7265 ---------CCTGTCCCCATGTACT--GGT----TTTATG--------GC   7291
                             |||.||.||.||||||  .||    |.||.|        |.
EMBOSS_002      8392 TGATCCATTCCTATCTCCTTGTACTAACGTCAAATCTAAGTAGATCAAGG   8441

EMBOSS_001      7292 CACTTTATATAAAGCTATTAGAGTTCGTTATGAATTTGCACTTTAAGAG-   7340
                     .|||..|.|||||   |..||||...|   ||||     |||||..||||
EMBOSS_002      8442 AACTCCACATAAA---ACCAGAGACAG---TGAA-----ACTTATAGAGG   8480

EMBOSS_001      7341 --------------AGCAT------TGTGGGTAAGGGGAGAGA---CCAT   7367
                                   |||.|      ||||||.|..|||||.|.|   ||..
EMBOSS_002      8481 AGAAAGTGGGGAAAAGCCTCGAAGATGTGGGCACAGGGAAAAAAATCCTG   8530

EMBOSS_001      7368 AATA-------------TTGTATATGGACGCGTTACGCTAAAATAGAGC-   7403
                     ||||             |||| |.||.|.| .|||    |.|||.|| |
EMBOSS_002      8531 AATAGAACAGCAATGGCTTGT-TCTGTAAG-ATTA----AGAATTGA-CA   8573

EMBOSS_001      7404 --TCG--CCTTATTA---TGTA-----CCTATAATG-------TACCA--   7432
                       |||  |||.||.|   ||||     .||.||||.|       .||||
EMBOSS_002      8574 AATCGGACCTCATAAAATTGTAAAGCTTCTGTAAGGCAAAGGACACCATC   8623

EMBOSS_001      7433 --------------GCCACTTTCATGGAGACTTGGGATAGAGTAGAGAAG   7468
                                   ||||   ||| ..||||| |||||          |||
EMBOSS_002      8624 AATAAGACAAAAAGGCCA---TCA-ACAGAC-TGGGA----------AAG   8658

EMBOSS_001      7469 GATAT----------------AGATAAGG-----ATACCTGACCTTATAT   7497
                     |||.|                |||||.||     |||.||.|   |||||
EMBOSS_002      8659 GATCTTTACCTATCCTAAATCAGATAGGGGACTAATATCTAA---TATAT   8705

EMBOSS_001      7498 AGTTTACATTCTAGTGGGAAAGATGGAAATGAAGCAAGACTAATATGTAA   7547
                     |   ||.|||        |.|.||..||.||||| |||
```

```
EMBOSS_002      8706 A---TATATT--------ATATATATAACTGAAG-AAG------------   8731

EMBOSS_001      7548 AGCATGTGGAATGTTAGTGATAAGTTCCA-AGGAGCAGA-------AT--   7587
                          |||||              .|||| |..|.||||       ||
EMBOSS_002      8732 -----GTGGA--------------CTCCAGAAAATCAGATAACCCCATTA   8762

EMBOSS_001      7588 AAAATGAGAGGGTATGCATGTGCTTGTAGACAGAGTGGCACTAAA-----   7632
                     ||||||   ||||                ||||||     |||||
EMBOSS_002      8763 AAAATG---GGGT----------------ACAGAG-----CTAAACAATG   8788

EMBOSS_001      7633 ------------AGGCCTA-------GCTGAAAAAGTGGCATTTGATGGG   7663
                                 |||..||       |||||.|||    ||..||||..|
EMBOSS_002      8789 AATTCTCACCCGAGGAATACCGAATGGCTGAGAAA----CACCTGAAAG-   8833

EMBOSS_001      7664 GGAAAATGTCCAAGTAAGTAATTTAGGGCCTTTTTAGGCAT-GGGGAA--   7710
                        ||||||.| ||||              |..|||..||| .|||||
EMBOSS_002      8834 ---AAATGTTC-AGTA--------------TCCTTAATCATCAGGGAAAT   8865

EMBOSS_001      7711 --AAAT-AAATGCACAGCCGTG----------TAACAGCAGCATGCCTGT   7747
                       |||| |||   ||||||.||          |.|||.|||
EMBOSS_002      8866 GCAAATCAAA---ACAGCCCTGAGATTCCATCTCACACCAG---------   8903

EMBOSS_001      7748 AGCATCAGGGAAAAGG-TAAGCGTGTGGCTGCAGCAGAAAAATCACGATG   7796
                         ||||   ||.|| ||||           |.|| ||||.|||.|
EMBOSS_002      8904 ----TCAG---AATGGCTAAG-----------ATCA-AAAATTCAGG---   8931

EMBOSS_001      7797 TTTAGCAGCAGAAG-----AA-----TTCAGAAAGGTTAAGGAAGTTAAT   7836
                       |..|||||||.|     ||     |..||||||    |||||
EMBOSS_002      8932 --TGACAGCAGATGCTGGCAAGGATGTGGAGAAAG----AGGAA------   8969

EMBOSS_001      7837 TAAGATCA---GACCAT----GGTAAGGACTTG--AGCTTTT--------   7869
                           ||   ..||||    |||  ||..|||  |||||.|
EMBOSS_002      8970 ------CACTCCTCCATTTTGGGT--GGGATTGCAAGCTTGTACAACCAC   9011

EMBOSS_001      7870 ------ATTCAG-CTG--------TCAGCGAGTTG-----AG--CT----   7893
                           |.|||| |||        ||||..|.|||     ||  ||
EMBOSS_002      9012 TCTGGAAATCAGTCTGGCAGTTCCTCAGAAAATTGGACATAGTACTACTG   9061

EMBOSS_001      7894 --GAATCCTGTTGTCTCTGTCTGCTGTGTTGATAATATATAATCAGAGAA   7941
```

```
                     |.||||||...|..|  |||.|||.|       .|||||..||||.||
EMBOSS_002      9062 GAGGATCCTGCAATACC--TCTCCTGGG-------CATATATCCAGAAAA   9102

EMBOSS_001      7942 -----CA------AAGGGTGAAATAT----AAAGACCTT--TAGTATCCT   7974
                          ||      |||...||||.||    .|..|..||  |||.|.|||
EMBOSS_002      9103 TGTTCCAACCGATAAGAAGGAAACATGCTCCACTATGTTCATAGCAGCCT   9152

EMBOSS_001      7975 ATTTT-----------------GAAACATTC---ATACCCCT---CAGAG   8001
                     ..|||                 ||||.|..|   ||.|||||   |||||
EMBOSS_002      9153 TATTTATAATAGCCGGAAGCTGGAAAGAACCCAGATGCCCCTCAACAGAG   9202

EMBOSS_001      8002 --AT-GATGGGGTACTTTGTCTAATGTGGTA---GTAGTGAGCACCAT-G   8044
                       || |||.|.|.|        |||||||||   .||     |||.|| |
EMBOSS_002      9203 GAATGGATAGAGAA--------AATGTGGTACATTTA-----CACAATGG   9239

EMBOSS_001      8045 ATAACTCACTTTCTGGCTATGTTTTGAGGTTAGAGCCAACAG---GACTT   8091
                     |..||| |||   |.||||          |||     ||.||   ||.||
EMBOSS_002      9240 AGTACT-ACT---TAGCTA----------TTA-----AAAAGAATGAATT   9270

EMBOSS_001      8092 TGT--------------GATGGATTGAATGTGAGGCTTGAGAGAGAGAGG   8127
                     |.|              .||||.||||        |.||       ||||
EMBOSS_002      9271 TATAAAATTCCTAGGCAAATGGGTTGA--------CCTG-------GAGG   9305

EMBOSS_001      8128 GGA-------GTG-GAGGATTCTGTTTTTGGCCTAA--ACAAAGGAAAAC   8167
                     |.|       ||| ||||          ||.||.||  |||||.|||..|
EMBOSS_002      9306 GCATCATCCTGTGTGAGG----------TGTCCCAATCACAAAAGAACTC   9345

EMBOSS_001      8168 ACAGATTGAGGCCTAGG---ACACTCCAAAGTGCG-----AGCTCAGGGA   8209
                     |||   |||    ||.|   .||||...||||| |     |||.|||..|
EMBOSS_002      9346 ACA---TGA----TATGTACTCACTGATAAGTG-GATATTAGCCCAGAAA   9387

EMBOSS_001      8210 CATGATTAGGA-ACCTGCAAAGGAGATGGAGGAAAGGTGACCTCTGAAGT   8258
                     |    ||||.| ||||      .||||..|.||
EMBOSS_002      9388 C----TTAGAATACCT------AAGATACAAGA-----------------   9410

EMBOSS_001      8259 TGGAGGAAAACTGTGTTCAAGTAGTACGTGAAGCCAAATTAGCAAATACT   8308
                                        |.|||  ||.||.||||               |||
EMBOSS_002      9411 ---------------TACAA--AGCACATGAA---------------ACT   9428
```

```
EMBOSS_001      8309 TCTAAGAAGAAA-AAGGAATCCTCTCAATCAAATCATGG-----------   8346
                       .||||||||| |||           |.||||...|||
EMBOSS_002      9429 --CAAGAAGAAAGAAG-----------ACCAAAGTGTGGACACTTTGCCC   9465

EMBOSS_001      8347 --TGATAGGTAAAGTGAGAACTGAACATCAGTCATTGAATGCAGTAGCTT   8394
                       |..|||   ||.||.||||..||||.|   |||.||    ||||| ||
EMBOSS_002      9466 ATTCTTAG---AATTGGGAACAAAACACC---CATGGA----AGTAG-TT   9504

EMBOSS_001      8395 GGGGGCCAATGGTGAGACTGATAAGAGCAGTTTTGA--TGGGACTGAGTA   8442
                     |     ||     |||||    ||    |||||.||  ||.|||     |
EMBOSS_002      9505 G-----CA-----GAGAC----AA----AGTTTGGAGCTGAGAC-----A   9531

EMBOSS_001      8443 AGAGAATAGATGTGAGAGCCAGTTTTTGTTTTTGTTTTGA----ACCA--   8486
                     |.||.||.||        |||             |.|.||    ||||
EMBOSS_002      9532 AAAGGATGGA--------CCA-------------TCTAGAGACTACCATA   9560

EMBOSS_001      8487 ---GTTTTTACTTTGCAAGAT-AG--ACCTGTGGAAGGCCTATGTTCCTC   8530
                        |..|.|.|||..||..|| ||  |||
EMBOSS_002     9561 TCCGGGTATCCTTCCCATAATCAGCCACC---------------------   9589

EMBOSS_001      8531 TTTGTTGATACACAGTCACTTAAAATGAGGACAGTTATCTTCCTTTTTC-   8579
                            |.||||||.|||            ||      ||||.|..||
EMBOSS_002      9590 -------AAACACAGACAC------------CA------TTCCATACTCT   9614

EMBOSS_001      8580 AGCAAAATATATCTCTTT-------AGACTTTGTTTTTGGTTTTTGTGCT   8622
                     |||||.||       |||       .||||.||.|.|.|
EMBOSS_002      9615 AGCAAGAT-------TTTGCTGAGAGGACTCTGATATAG-----------   9646

EMBOSS_001      8623 TTTCTGTGCTTTTTAATGTATATTTCTGTGAGTTAATGGGAGACGGTAGA   8672
                        |||| ||.|              ||||||...||    |.|||
EMBOSS_002      9647 ---CTGT-CTCT--------------TGTGAGACTAT----GCCGG----   9670

EMBOSS_001      8673 AATTTTAGGAACTACTGATCCAACCTACCTGAAA-ACAGAAGT-------   8714
                             ||              ||||.|  ||| ||||||||
EMBOSS_002      9671 -------GG--------------CCTAGC--AAACACAGAAGTGGATGCT   9697

EMBOSS_001      8715 -----TCATATATTTCATTTGAGGACGAA-----------CGTATGG-GT   8747
                          |||..||||        |||.|.|           ||.|||| |.
EMBOSS_002      9698 CACAATCAGCTATT--------GGATGGATCACAGGGCCCCGAATGGAGG   9739
```

```
EMBOSS_001      8748 GGCCAGA-----TCCCAGGAG--AAATGGCGATTTTCAGTGATGTTTATT   8790
                     .||.|||     |||.|||||  |||.||                  ||.
EMBOSS_002      9740 AGCTAGAGAAAGTCCAAGGAGCTAAAGGG------------------ATC   9771

EMBOSS_001      8791 TACTTAACCTTTTTTTTTTTTTTTTTTTAACACTTTTTATATGTGG----   8836
                     |||  ||||                          .||||.||.|
EMBOSS_002      9772 TAC--AACC---------------------------CTATACGTAGAACA   9792

EMBOSS_001      8837 -TTATTTGGACATTTCCACGAACAAGGTAC-------TCTCATCATCTCT   8878
                      |.||.||.|| |..|||        ||||       .|||.| .|||||
EMBOSS_002      9793 ATAATCTGAAC-TAACCA--------GTACCCCTGGAGCTCCT-TTCTCT   9832

EMBOSS_001      8879 TG-TGTCAAAAATGTTTCCG----TGGCC---------------------   8902
                     .| |||   |.||||.|..|    |||||
EMBOSS_002      9833 AGCTGT---ATATGTATTAGAAGATGGCCTAGTTGGCCATCAGTGGAAAG   9879

EMBOSS_001      8903 --AGGCACGGTGGCTCACGC---CTGTA-AT--CTCAGCACTTTGGG---   8941
                       ||||.|..||| .|||.|   ||.|| ||  |||||.||  .|||
EMBOSS_002      9880 AGAGGCCCATTGG-ACACACAAACTTTATATGCCTCAGTAC--AGGGGAA   9926

EMBOSS_001      8942 ----AGGTC--------GAG-GCGGGTGGATCA------TGAAGTCAGGA   8972
                         |||.|        ||| |.||||||.| |      ||.||..||||
EMBOSS_002      9927 TGCCAGGGCCAAGAACTGAGAGTGGGTGGGT-AGGGGAGTGGAGGGAGGA   9975

EMBOSS_001      8973 GATCA-AGACCATCATGGTTGACACCAT--GAA--GCCTCGTCTCTACTA   9017
                     .||.| .|||  |..|||  ||.|.|||  |||  |            ||
EMBOSS_002      9976 TATGAGGGAC--TTTTGG--GATAGCATTGGAAATG------------TA  10009

EMBOSS_001      9018 -------AAAATAC--------AAAAAATTAGCCGGGCATGGTTGCGGGC   9052
                            |||||||        |.||||..|.||...|||
EMBOSS_002     10010 AATGAGGAAAATACCTAATAAAATAAAACCATCCTAACAT----------  10049

EMBOSS_001      9053 GCCTGTA-------GTCCCAGCTACTC--AGGAGGCTGAGGAAGGAGAA-   9092
                      .||.||       ||     ||||||  ||..|.||..|| |.||.||
EMBOSS_002     10050 -ACTTTATTTGTTCGT-----CTACTCCTAGTGGTCTTTGG-ATGAAAAT  10092

EMBOSS_001      9093 ---TGGC---GTGAACCCGGGAGGCGGAGCTTGCAGTGAGC---CAAGAT   9133
                        ||||   ||.|||           |.||..||| |.||   ||.|||
```

```
EMBOSS_002     10093 TTTTGGCAATGTCAAC-----------ATCTCACAG-GTGCTTTCATGAT  10130

EMBOSS_001      9134 TGCGCCACTGCACTCCAGCCTGGGCGAC-AGAGCGA---GACTCCGTCTC   9179
                                              ||| ||||.||   |..|||.|.|.
EMBOSS_002     10131 -------------------------GACTAGAGTGACCTGTTTCCCTTTT  10155

EMBOSS_001      9180 AA----------------AAAGAAAAAAAAAAAAAGTTTCCTAAATTTCT   9213
                     ||                |.||..|.||.||.|.|||||  |||
EMBOSS_002     10156 AATTGTTCATGCTGATTTAGAGGCATAATAATATAGTTT--TAA------  10197

EMBOSS_001      9214 TTTTAAGATTATTGAAGTTACCTTTAAGTAGAATGAGCACATATTTCAGT   9263
                     ||.|..||||||.|||.|||.|   |.||.|||    |..||.|||||.||
EMBOSS_002     10198 TTGTTGGATTATAGAAATTAGC---AGGTGGAA----CTTATTTTTCTGT  10240

EMBOSS_001      9264 TTGTCGGGTATAGTCTTAGCTTATGTCAAACCATT-------TTTCTCAG   9306
                     |         ||.||||.|||.||      |||||       |||||
EMBOSS_002     10241 T---------TACTCTTTGCTGAT------CCATTTATTGAATTTCT---  10272

EMBOSS_001      9307 CTTAAT----TATTCCTAGCTCCCCAT----------TTT---ACTATAA   9339
                     .||.||    ||.|||.||     .||          |||   ||.||||
EMBOSS_002     10273 TTTGATTTCATAATCCAAG-----AATGAAATTGGTATTTAACACAATAA  10317

EMBOSS_001      9340 AAACTATCTCATTTTGGGAAATAAA-TAGCGAGCTTCCTTTAATCCCCTC   9388
                     |..||   |.||.|....||||||| ||      ||  |||||
EMBOSS_002     10318 ATGCT---TAATATATACAAATAAATTA------TT--TTTAA-------  10349

EMBOSS_001      9389 AGAAGTCCCACATGG--TAGGTATCTTTATT-GCTATTTTACATAGG---   9432
                     ||     |.|||.||  ||.|||..|..||| |..||.|.|   |||
EMBOSS_002     10350 AG-----CTACACGGGTTATGTAAATACATTAGAGATGTGA---AGGATG  10391

EMBOSS_001      9433 ----AAATACTGAGGT-TCAGAAAG----------------GTCGAAGT-   9460
                         ||.|.||||..| |..|||||                |||.||||
EMBOSS_002     10392 CCAAAACTTCTGAATTATGTGAAAGAATAAATTAGCACTCTGTCAAAGTA  10441

EMBOSS_001      9461 ----ACTTTATTAGGGACAGATAGTAAGCAGTAGAGACAGAACTGATAAC   9506
                         |||.||||||||||..|.||      |||              |||.
EMBOSS_002     10442 CAAAACTATATTAGGGAACGCTA------AGT--------------TAAT  10471

EMBOSS_001      9507 AGCAGAGCAAAAAGGGTGACTGCAGGCCCTTTTGTTGTAGTATCTTCTTT   9556
```

```
                    ||      ||||                  ||..|||      |||||..
EMBOSS_002     10472 AG------AAAA------------------TTAATTG------CTTCTAA  10491

EMBOSS_001      9557 AA--CATCT-----CTTTGTTATACTAAATTTTATTATTTTTAATCAACT   9599
                    ||  .||.|     |.|||..|.|.|.||.||||.|       ||||  |
EMBOSS_002     10492 AAACAATATAGGGCCATTGGGAAAATCAACTTTACT-------ATCA--T  10532

EMBOSS_001      9600 ATC--TAC---AGCTGTGTCTGAAATG------AAATG------TTTTCT   9632
                    |||  |||   ||.|| |..||||||.|      |||||      ||||..
EMBOSS_002     10533 ATCATTACAAAAGATG-GATTGAAAGGAGAAATAAATGTCATGATTTTAA  10581

EMBOSS_001      9633 GGGCTCTTCACTGTCT----CATTCACACT---CTTCAAGGCATACTTGT   9675
                    .|     ||||| |.|    || |||.|||   ||| ||.|.|||.||.|
EMBOSS_002     10582 AG-----TCACT-TATTAACCA-TCAAACTTCCCTT-AATGTATAATTTT  10623

EMBOSS_001      9676 ATG----CATATTCCTCTTCATTTTGCAATGTTCCTCTTGAAAGTAATGT   9721
                    .||    .||.||.|||||| |.||.||.|.||  ..||.||   ||||
EMBOSS_002     10624 TTGGAATTATTTTACTCTTC-TATTACATTTTT--ATTTCAA---AATG-  10666

EMBOSS_001      9722 GCCCTAACAG-AACC---CAGATATCGTCTGAACATTG-----ACAAATA   9762
                           ||| .|||   ||.||    |||  ||||||     |||..|.
EMBOSS_002     10667 -------CAGTTACCTTTCATAT----TCT--ACATTGAGGGTACATTTT  10703

EMBOSS_001      9763 TGCA-GATACCATTGACTGCTTCTTTCTGGGGTCTATGTGTCTCTTTATG   9811
                    ||.| .|||  |||..|||.||.|.|||      ||..|     ||||||
EMBOSS_002     10704 TGAATTATA--ATTACCTGATTATCTCT------TAACT-----TTTATG  10740

EMBOSS_001      9812 CACC-AGTTATTCAGTTTATTTGAGCTGTATA----CAGAG---TTGGTA   9853
                    .||| |..|||| |.|||||.|      ||||    .||||   ||.||
EMBOSS_002     10741 AACCTATATATT-ATTTTATAT------TATATGTGAAGAGCCCTTAGT-  10782

EMBOSS_001      9854 ATCAGCCTATAG---ACTGTCAGAGACCATTCTGAGTGACCCGGCATTCC   9900
                     ||||  |.|.|   |||.|        |||||   ||||..|..|||
EMBOSS_002     10783 -TCAG--TTTGGATTACTAT--------ATTCT---TGACTTGAAATT--  10816

EMBOSS_001      9901 TTGTGATATCTTGTTGGCCCTGCTGAATTA-------------------A   9931
                         ||||||..|     .|.||.|||||                   |
EMBOSS_002     10817 -----ATATCTATT-----ATACTAAATTATTTAGAATCTAGTTTAACTA  10856
```

```
EMBOSS_001      9932 TAG-------TTAAA---------------TATGT-TGAACATCCCTAAT   9958
                     |||       |||||               ||||| |.||.|     |||
EMBOSS_002     10857 TAGGATGATATTAAACAAAAACTGCTGTTCTATGTGTTAAGA-----AAT  10901

EMBOSS_001      9959 ACATCCAAACTATAAGCTCT-----------------GTTTGCTCT----   9987
                     .|.||.|||.||    ||||                 .|||..|.|
EMBOSS_002     10902 GCTTCAAAATTA----CTCTACTGAAATATAAAATAATTTTATTATTAAA  10947

EMBOSS_001      9988 -------------GATATATATTCTGATACGCTCACACCTTTTTGA----  10020
                                  .||||||||    ||||.|.|||||       |
EMBOSS_002     10948 ATAATTATATATATATATATAT----ATACACACACAC-------ACAGA  10986

EMBOSS_001     10021 -ATAATATCCATTGAAATACCCCTAGC--------------AACAGCAGA  10055
                      ||.|||...|||.|||     .||||              ||.|.||||
EMBOSS_002     10987 GATCATACATATTAAAA-----GTAGCATACAATCATACATAAAAACAGA  11031

EMBOSS_001     10056 GC-ATGTCT----GAAAGAAAGCT---GAAATTATAAGTAGAAGA-GCAG  10096
                     .| ||.|.|    ||||.|.||||   |||| .||||  |.|.|| |.||
EMBOSS_002     11032 ACTATTTTTTAGAGAAAAACAGCTAGAGAAA-AATAA--AAATGAGGGAG  11078

EMBOSS_001     10097 GGTTCTAATATTGAATCTAAAACTTCCT-AATATGATGACC------TTT  10139
                     |.|    |||||           |||.| ||.|..||||||      .|.
EMBOSS_002     11079 GAT----ATATT-----------TTCATAAACAGTATGACCCAAAGAATG  11113

EMBOSS_001     10140 GATATACGAT-TTAAT-CTTACT-----------GATGCTCAGTTTTCTC  10176
                     ||..||..|| |||.| ||||||           .|||.|.|||
EMBOSS_002     11114 GAGTTATCATATTAGTACTTACTAGAAGAATCACAATGGTAAGT------  11157

EMBOSS_001     10177 TTGTATAAAATGGTATAC--------AACAATGCTGATTTCT-----TCT  10213
                         .||||||.|||||        ||.|||.|| .|||.|     |.|
EMBOSS_002     11158 -----AAAAATGTTATACATTTTCTGAAAAATACT-CTTTATGGTAGTTT  11201

EMBOSS_001     10214 TTCTTACA----------ATTATT-----------------------ATG  10230
                     ||.|||||          |.||||                       |||
EMBOSS_002     11202 TTATTACATATATGTAGTAGTATTTCAGAGATTGTATAAGAAAAGAGATG  11251

EMBOSS_001     10231 A-CATA--------CATATGG-----AAATTAT--GT----ACATAAA-C  10259
                     | .|||        |||||||     ||.||||  ||    |.||||| .
EMBOSS_002     11252 AGGATACCAGGAATCATATGGCTCATAACTTATCAGTGATGATATAAATG  11301
```

```
EMBOSS_001     10260 TATAA-------TTTGG------------TAGTATAAGGAATCTTA----  10286
                     |||||       .||||            |||||     |||.|||
EMBOSS_002     11302 TATAAAATACATGTTGGTGATAAATGTCATAGTA-----AATATTAGAAG  11346

EMBOSS_001     10287 ------TTT----------------------------------------  10289
                           |||
EMBOSS_002     11347 AAGAGGTTTAGGTGTGTGTGTGTGTTTGTGTGTGTGTGTGTGTGTGTGTG  11396

EMBOSS_001     10290 --TCAATGTGGCTAT---TGCAT---TGTTTTCAAAATTTTCACAAAC--  10329
                       |...||||..|||   .||||   ||..|.||       ||||..|
EMBOSS_002     11397 TGTGTGTGTGTGTATGCGGGCATAGGTGCGTGCA-------CACAGTCAT  11439

EMBOSS_001     10330 -----CTCTCTGTGCCTCACTTCCCTCCTCTGTCAAA--ATGGGAAAGTA  10372
                          ||| ||||       ||   ||||.||||.||  |.||.|.|..|
EMBOSS_002     11440 GTGTACTC-CTGT-------TT---TCCTGTGTCCAAGCAGGGCAGAACA  11478

EMBOSS_001     10373 AT-AGTACCTA-CTTTA----TAGATTATTGTTTTTGT------------  10404
                     .| |||||.|| |||||    .|||..||.|.|.||||
EMBOSS_002     11479 CTCAGTACATATCTTTAAAGTAAGAAAATGGCTGTTGTATCGGGTAACAA  11528

EMBOSS_001     10405 -------TGTAGTTGTGGTGGTTGTTACTG-AGACTATTTATA-------  10439
                            |||.|.||.|||..|.||||..| |.||.|.|||||
EMBOSS_002     11529 GAAAGCATGTGGATGGGGTTTTAGTTATAGCAAACAAATTATAAAGTGTA  11578

EMBOSS_001     10440 ------TTAGAATAA---------------------------------CT  10450
                           ||||.||.|                                 .|
EMBOSS_002     11579 AAGTGTTTAGGATTAGGCAGAGCTTGAGAATTTACTGAGCTGTAGAGTGT  11628

EMBOSS_001     10451 GGAACA---------ATGTGGTTAACAAGATATTAAAGGTCAAATATGAT  10491
                     ||||||         |||||    ||.||.|.||.|||.|   ||.|||
EMBOSS_002     11629 GGAACATAGTCCTTGATGTG----ACTAGCTTTTTAAGCT---ATCTGA-  11670

EMBOSS_001     10492 TCCTGAGATATTCAAATATAAG-------TTTTGTTCTCTG----TTAGT  10530
                                 .|||.|||||       |    |||..||    |||||
EMBOSS_002     11671 ------------AAAACATAAGAAAGCAAT----TTCCATGATGCTTAGT  11704

EMBOSS_001     10531 AATATTAATTGA-----------TTCTAGTAAG---CTTAGAATAAAGAG  10566
                     ||..||||||.|           |||||..|..|   ||||..|||.| |.
```

```
EMBOSS_002     11705 AAGCTTAATTAACATGTTTTTCTTTCTTTTTTGCTACTTATCATACA-AT  11753


EMBOSS_001     10567 ATTA-TGATGAT---------AACAT---TTAGAGGGCTATTGATA----  10599
                     |||| |.||.||         |||||   .||.|.|.||||| |||
EMBOSS_002     11754 ATTACTAATAATGCAATATTCAACATATAATATATGTCTATT-ATATATA  11802

EMBOSS_001     10600 -------ATTTATTTTCAATATTTATAAAGACCACATAGTGTTCTCATTC  10642
                             ||||||     ||||.||||.|.|              |||||
EMBOSS_002     11803 CACATATATTTAT-----ATATGTATATATA--------------CATTC  11833

EMBOSS_001     10643 TACTTTTTATATCATTTCACAAAAACTAAAG---CAAAAGCTCAAT----  10685
                     ||...|.||||||.||          |||.|   .|||.|.|..||
EMBOSS_002     11834 TATAATATATATCCTT----------TAATGGTATAAATGTTTTATTTTT  11873

EMBOSS_001     10686 --------GTA-----GATAAAATAGAAACCCCTCTGATAGGAGAAGATA  10722
                              |||     |.|.|||||.|..|                ||.
EMBOSS_002     11874 ATTTTGCAGTACTTTTGTTGAAATACAGGC-----------------ATT  11906

EMBOSS_001     10723 ACTCTAAAAGTAACTCAAAGAACATCA--------CTTTACT--------  10756
                     ||      |.|.|| |||.||||..||        |||||.|
EMBOSS_002     11907 AC------ATTTAC-CAATGAACTACATATCTGTCCTTTAATATTATATA  11949

EMBOSS_001     10757 ----------CAGCT--------TGTTGTCATTGTTCCTTTATGTGTTGT  10788
                                ||.||        |.|||| |||.|..||.|.|..||..|
EMBOSS_002     11950 TCGAAATAGGCATCTCTCATAAATATTGT-ATTTTGACTATTTTAGTAAT  11998

EMBOSS_001     10789 -------ATTGTGGTATATATTTTTTCTTTCATCTGTATATTGTTTTCTG  10831
                             ||.|||.| ||||    .|||.|.||.||..|| ||||||
EMBOSS_002     11999 AATGAAAATAGTGAT-TATA----ATCTATTATTTGGTTA-TGTTTT---  12039

EMBOSS_001     10832 GATA--AAGTTCAAATTCTGGACACCCCAAATCCTCCTGTCTTTCT----  10875
                     ||||  |||||||||   .|||            ...|||.|||.|
EMBOSS_002     12040 GATATTAAGTTCAAA---AGGA------------GTGTGTGTTTGTGTAT  12074

EMBOSS_001     10876 ---TGATATTTACATT---ACATATTTA-----CTTAAATGTGCTTATTT  10914
                        || |.|.|.||||   ||||.|.||     .|||||..||||||||||
EMBOSS_002     12075 GTGTG-TGTGTGCATTCACACATGTGTATTTGTGTTAATAGTGCTTATTT  12123
```

```
EMBOSS_001     10915 ---TTTCT----------CC-----TCTAAGA--TGCTCAGG---GTGCT  10941
                        |||.|          ||     |.|||.|  |||| |||   .|..|
EMBOSS_002     12124 ATGTTTATAAGAAATCAACCCATAATGTAACACTTGCT-AGGCAAATATT  12172

EMBOSS_001     10942 TT------------------------CCATATAACATTT--ACCTTTT--  10963
                     ||                        |||| |||.||||  |.|||||
EMBOSS_002     12173 TTTGGGATAAGTTTCCTGAAACAAAACCAT-TAAGATTTTAAACTTTTTT  12221

EMBOSS_001     10964 -AA--TGGAG-------------AATAAATGTAAGTAGTGA-AAAACAAG  10996
                      ||  |||||             |||||||.||||.|.||| ||||.||.
EMBOSS_002     12222 GAATGTGGAGTTTGTATGGAAAAAATAAATATAAGAAATGAGAAAATAAT  12271

EMBOSS_001     10997 TATGCCATAAGCTGCTAAGGACAGATAACTTAGAT----CACAAAAACT-  11041
                     |      |.||||..|||  |.||||...||||||    |       ||
EMBOSS_002     12272 T------TTAGCTCATAA--AAAGATTTTTTAGATTTGGC-------CTG  12306

EMBOSS_001     11042 AATAAGTTGAGAAAATAGA--AATATTGG--AGCTTCATCT---------  11078
                     |.||||.|.||||||..||  |||..|.|  ||||.||.||
EMBOSS_002     12307 ATTAAGATAAGAAAAATGATTAATGCTTGAAAGCTACAGCTCAAGGAGAT  12356

EMBOSS_001     11079 -----------AGGATCTTACTT---TTAACTAGATGTTGTAGCTACTCT  11114
                                 ||||..||||.|   .|||  |||||         ||||
EMBOSS_002     12357 AGACAGGAAACAGGAAATTACATCAGGTAA--AGATG---------CTCT  12395

EMBOSS_001     11115 --ATTT----------TCATGGACTTTTCAGTTCAACCAA-CAAATTAGA  11151
                       ||.|          |||||.|..||.|...||.||.|| ||||  |||
EMBOSS_002     12396 TGATGTCAGAGGAAAGTCATGCATGTTACTACTCTACAAAGCAAA--AGA  12443

EMBOSS_001     11152 ACCATCCATGATTGTGCACGTTCCTG--GAAA---CAATTCAAT------  11190
                            ||.|||    ||.|||.||  ||||   .||||||||
EMBOSS_002     12444 -------ATAATT----ACTTTCTTGAAGAAAAGGGAATTCAATGGATTA  12482

EMBOSS_001     11191 ACTGTCTTTGT----GAAACTGAATTTCAAGACAACTCAGAAGT----TT  11232
                     |     |||.|    ||||.|.||.|    ||.||||.| ||||    |.
EMBOSS_002     12483 A-----TTTATACTCGAAAGTAAACT----GAGAACTGA-AAGTCAGATA  12522

EMBOSS_001     11233 TA-CCTGC--TAATTGGAAAGGGAT------CCTGGGAACATAAAATCAC  11273
                     || |||||  |||.|.|..||..||      |||.|  ||||    |.||
EMBOSS_002     12523 TAGCCTGCAATAAATAGGCAGACATTGCTGACCTAG--ACAT----TGAC  12566
```

```
EMBOSS_001     11274 AA----------------------AATACCACAAACTGGGTGCTTTG---  11298
                     ||                      ||||         ||.|.|||||
EMBOSS_002     12567 AATTGTGTTAGGATTGAGTAAGGTAATA---------GGTTACTTTGCTT  12607

EMBOSS_001     11299 ----------------------AAAT----TTTCAAGTGGAATGTAAGAT  11322
                                            |.||    |.|||..|..||..|||||..
EMBOSS_002     12608 TTGCTCTTGGTTTCATTTATTTATATTCTATATCATTTTTAAATTAAGCA  12657

EMBOSS_001     11323 ATTAAAAAGTGTCTGTCACTTTCAGAGT-----TTCTT-AACT--AA---  11361
                     |.||||||.|            ||||.||     ||||| ||.|  ||
EMBOSS_002     12658 AATAAAAGG------------TCAGTGTGTTCCTTCTTGAAATGCAATAC  12695

EMBOSS_001     11362 -TTAGACAAAA--CTGAGTTTTCAT-------TTGTTGCCCTTC------  11395
                      |||.|.|||||  |||.|.||.|.|       |||||  .|||||
EMBOSS_002     12696 TTTAAAAAAAAATCTGTGATTCCGTTCCAGGATTGTT--ACTTCTAGGCA  12743

EMBOSS_001     11396 ----------AGAGCACCATGTAAAATGTGT-------ATCTTTCAG--C  11426
                                .|||    .|||.|.|||.||       ||||||..|  |
EMBOSS_002     12744 TTGTTTTGGTTGAG----GTGTGATATGGGTTTTTCCCATCTTTGTGTTC  12789

EMBOSS_001     11427 TAAGAGATAGGTATTTAAGAAAATAAAACAAAA----TTACTACCTTATA  11472
                     |..|..|.|.||.|.|..|||   |||||||||    .||..||.|.|.|
EMBOSS_002     12790 TTTGTCACATGTCTCTCTGAA---AAAACAAAAGGTGGTAAGACATAAAA  12836

EMBOSS_001     11473 GATT----TTTTTTAAATTTTACTCTTAATTTATTCTCAG---TTTTAAT  11515
                     .|||    .|||.|||    |||||  ||||..||..|||   |.||.||
EMBOSS_002     12837 TATTAGGAATTTCTAA----TACTC--AATTCTTTAACAGAACTGTTCAT  12880

EMBOSS_001     11516 TAGTATTTTATATATTACAGCTAT-TCCTGATG--------------GAG  11550
                       .||||||.|.|.||||    |.| |||| |||              |||
EMBOSS_002     12881 --ATATTTGAAAAATTA----TGTATCCT-ATGAGAAAGCAGTTTCAGAG  12923

EMBOSS_001     11551 TTT-----ATTCT--GTGAA--TGATTGGCTTTTTTTTTTTTTTTAGCA  11591
                     |||     ||.||  ||.||  |.|||..|.|.|||.||||.||||  ||
EMBOSS_002     12924 TTTCAGTGATGCTTAGTAAAGTTTATTAACATATTTCTTTTCTTTT--CA  12971

EMBOSS_001     11592 TTTTGTACT-CTTGATATTGAATCCAGTATTCCTAATTCATGAAACCTTT  11640
                     |||..||.| |||    ||..||.||.||    ||.||||
```

```
EMBOSS_002     12972 TTTGTTATTACTT----TTCCATACAATA----TATTTCA----------  13003

EMBOSS_001     11641 TTTTGTG-ATTTTAATAGACTCC----ATTTTTAGAGT-AAT--------  11676
                      .||||| |.|||..|   ||||    |.||.||..|| |||
EMBOSS_002     13004 -ATTGTGCAGTTTCTT---CTCCCTAAACTTCTAATGTGAATATTGTTGA  13049

EMBOSS_001     11677 TATAGGTTCACAGCACAATTGAACAGAAGGTACAAAAATTTCCCATATAA  11726
                     ||||||..|||         |||||  ||||||            ||.||
EMBOSS_002     13050 TATAGTCTCA---------TGAAC--AAGGTA------------ATCTA-  13075

EMBOSS_001     11727 CCCGGCTCCCACAGATACATAGCCTGCCCCACTGTCAACATTTCCCACTA  11776
                               |.|||||  ||.|          ||||||..||  |.||||
EMBOSS_002     13076 ---------CTCAGAT--ATTG----------TGTCAAGGTT--CTACTA  13102

EMBOSS_001     11777 GAGTGGTATATTTGTTTCAACAGATG--------------AACTTACATT  11812
                                    |||.|  |||||              ||.|||.|.|
EMBOSS_002     13103 --------------TTTTA--AGATGATTGGGGTCATCCTAAATTAAAAT  13136

EMBOSS_001     11813 G---ACACATCATTATCTCCCAAAGTTCATAGTTTATACA-TTAGCTTTC  11858
                     |   |||.|.|||           |||  |.|||||   | ||||
EMBOSS_002     13137 GAGTACATAACAT-----------GTT--TTGTTTA---ATTTAG-----  13165

EMBOSS_001     11859 CCTATTGGTGCCACACATTCTGTGAGTTTAAACAAATTTACAATGACATG  11908
                       |.||.|||.|..||      |.||.||||  |||||    ||.|| ||
EMBOSS_002     13166 --TCTTAGTGTCTTAC------TCAGATTAA--AAATT----ATCAC-TG  13200

EMBOSS_001     11909 TATTTATCTTTAGAGTGTTATGCGGTGTAATTTCACTGCCATAAAA---A  11955
                     ||.|                        ||||.||||..|..|||   |
EMBOSS_002     13201 TACT-------------------------ATTTTACTGGGAAGAAAATGA  13225

EMBOSS_001     11956 TCCTCAATACTTTTCCCGTTCATCT---ACACT----TC-------TCTC  11991
                     |||||              ||||.|   |.|.|    ||       ||||
EMBOSS_002     13226 TCCTC--------------TCATTTTGGAGATTGTTATCTGATGAGTCTC  13261

EMBOSS_001     11992 C--TAATCT-CTAGAAACCACTGATTGTT-TTG-TAGTT--TACATAGTT  12034
                     |  ||.||| |||..||.|.||  ||.|| ||| |||||  ||||
EMBOSS_002     13262 CTTTAGTCTGCTAAGAAGCTCT--TTATTATTGTTAGTTAATACA-----  13304

EMBOSS_001     12035 TTACCTTTTCCAGACTATCATATATTCGGAATTGGAATCTAGGGTGAAAT  12084
```

```
                                |.||.|.||||.|     |||...|||        ||||
EMBOSS_002     13305 -----------AAACAAACATACA-----AATAAAAAT--------AAAT  13330

EMBOSS_001     12085 -CCTGCGGTAAAATATATAGCCTTTTCAGATTATCTTCTTTAACTTAGTA  12133
                      ||      |||..|||                     .||||||.|  |
EMBOSS_002     13331 ACC------AAACAATA----------------------ATAACTCA--A  13350

EMBOSS_001     12134 ACATGCATTTA--------AGT---TCTTCCATGTCTT--------TTTA  12164
                     |||  ||||||        |||   ||.|||.|.||||        ||||
EMBOSS_002     13351 ACA--CATTTATATTTGTTAGTTTATCATCCCTTTCTTATCCTTTATTTA  13398

EMBOSS_001     12165 TGGCT-TGAT-----AGCT-----CATTTTTAA---------GCACTGAA  12194
                     |  || ||||     |.||     |||.|||||         ..||||
EMBOSS_002     13399 T--CTGTGATCAAACACCTACAACCATCTTTAAATTCAAAGTTTACTG--  13444

EMBOSS_001     12195 TAATATTCCATTGTCTGG--ATGTCTCACAGTTTAT--TTA----TCCAT  12236
                      ||.|.|.||||||||.|  || |.||    |||||  |||    |||||
EMBOSS_002     13445 -AACACTTCATTGTCTAGTCAT-TTTC----TTTATGCTTAACTTTCCAT  13488

EMBOSS_001     12237 TCATCTACTGAAGAACATCTTGATTGTTTCCAAGTCTTGGCAAAT-----  12281
                     |||| ||.|.||.|             ||||    ||||.||.||
EMBOSS_002     13489 TCAT-TATTTAATA-------------TTCC----CTTGTCATATGGTTT  13520

EMBOSS_001     12282 --ACAAACAAAGCTGCTGCAGCATTCATGTGCAGGAATTTGAGTGGACAT  12329
                       .|.|||||           .|.||||||    |||||        ||.
EMBOSS_002     13521 GACCTAACAA-----------TACTCATGT----GAATT--------CAA  13547

EMBOSS_001     12330 A--TGACTCCTTTGTGTAAAT-------GC--CTAAGTGCATGATTGCTG  12368
                     |  ||||...||  |||||||       ||  ||||||      |||..|
EMBOSS_002     13548 AATTGACAAATT--TGTAAATAGTGTTAGCAACTAAGT------TTGTAG  13589

EMBOSS_001     12369 G------ATC---ATATA---GTA---------------GAAT-GTTTAC  12390
                     |      |||   |||||   |||               |.|| ||.|||
EMBOSS_002     13590 GGCTTCTATCTCTATATATGTGTAAGTTGCTCCTTGGCTGTATAGTGTAC  13639

EMBOSS_001     12391 -TTTTGTAA--GAAACCTCCA----AACTG-----------TCTT-CCAA  12421
                      |.|||.||  |.||||..||    ||.||           |||| ||||
EMBOSS_002     13640 ATCTTGCAAATGGAACCAACAATTGAATTGCATCTGATGTTTCTTACCAA  13689
```

```
EMBOSS_001     12422 AGTGGCT---GTGCCATTTTGCATT------------------GCCA-CC  12449
                       ||.||   |.||||   ||||||                  .||| .|
EMBOSS_002     13690 --TGACTCCAGGGCCA---TGCATTTATTCTAATATCTTAGTACCCATGC  13734

EMBOSS_001     12450 AGAAATGAAT----GAGAGTTCCCAT----------TGCCCCAT------  12479
                     .| |||.|||    ||.||||   ||          |.|.|.||
EMBOSS_002     13735 TG-AATTAATAGCAGATAGTT---ATTTAGAATACCTACACTATCTAGAA  13780

EMBOSS_001     12480 --------TTCTCTGTCGGCTTT----TG-GTTTTGTCAGTGTTCTTGGA  12516
                             |||||      ||||    || |||      |||||||.|.|
EMBOSS_002     13781 ACTAAATATTCTC------CTTTACTATGAGTT------GTGTTCTGGCA  13818

EMBOSS_001     12517 TTTTGGCCAT-TCTAATAGGTGTGTAGTGGTATCTCATTGTTTAATTTGT  12565
                     |..|   ||| .|||.||||||..|       |.||||||  .|||
EMBOSS_002     13819 TACT---CATGCCTATTAGGTGAAT-------TTTCATTG--AAAT----  13852

EMBOSS_001     12566 AATTCCCTG--GTAACATATGATGTGAAACA-TATTTTCATTTGCTTATT  12612
                      |||||.||  |.|.||.||  |||||||.| ||          ||.|
EMBOSS_002     13853 -ATTCCTTGAAGAAGCAAAT--TGTGAAAGAGTA----------CTGA--  13887

EMBOSS_001     12613 TGCCATCTGTATACTTCCTTTGGTGA------GGTGTCTGTTAAGG----  12652
                         |..|||||.||         .||      |||     ||.|||
EMBOSS_002     13888 ----ACTTGTAAAC---------AGAAAAAATGGT-----TTTAGGAAAC  13919

EMBOSS_001     12653 ATTTTTGCCCATGT------TTTAGTTTG-TTTGT--TTT--TGTTACTG  12691
                     |||||    ||.||      |||  |||| |||||  |||  |.|.|.||
EMBOSS_002     13920 ATTTT----CAAGTCCTGTCTTT--TTTGATTTGTCATTTAGTCTCATTG  13963

EMBOSS_001     12692 AGTTTTAAGAGTTCTTTGTGTATTTCGGATAAGAATTCTCTACCAG-ATA  12740
                     |..||||  |.||||             ||||         |.||| |||
EMBOSS_002     13964 ATGTTTA--ACTTCT-------------ATAA---------AACAGCATA  13989

EMBOSS_001     12741 TGTCTTTCGTAAAT-ATTTTCTTTGAGCCTGTGGCTTTTCTTTTC-----  12784
                        ||       || ||.|.|||||            |||||.||
EMBOSS_002     13990 ---CT-------ATAATGTACTTTG------------TTCTTGTCCACAA  14017

EMBOSS_001     12785 --GTTCTTTTGGCAGCATCTTTCACAGAGCAGAACCTTTTA-----ACTT  12827
                       |||.|    |.||.||.||    |||    ||..||.||     ||.|
EMBOSS_002     14018 ATGTTGT----GAAGTATGTT----AGA----AATATTATATATATACAT  14055
```

```
EMBOSS_001     12828 AGTAC-TAT----------TCAACTTAT--------TAA--TTCTT-TCT  12855
                     | ||| |||          | |||||||        |||  |.||| ..|
EMBOSS_002     14056 A-TACATATATATATAACAT-AACTTATAGGCAGTATAAGGTACTTGAGT  14103

EMBOSS_001     12856 TTCATGGATCATCCCTTTGGTGGTGTATCTAAAAAACTATCAACAAACTC  12905
                     ||.|..|||      .|||     ||||.|..|||..|.|       |||
EMBOSS_002     14104 TTGAAAGAT------ATTG-----GTATTTTCAAATGTTT-------CTC  14135

EMBOSS_001     12906 AAAGTCACCTCTG--------TTTTCTCCACTGCTATCTGCT------AG  12941
                     ||| ||| ||.||        ||||.|||.|||        |      |.
EMBOSS_002     14136 AAA-TCA-CTTTGTGTCTTAATTTTATCCCCTG--------TCAAAACAA  14175

EMBOSS_001     12942 GAGTTTTATAGTTTT------ACA----TTTTACTCTAGAGTCTGTGATT  12981
                     |||....|||.||||      |||    .||||.|.|   || ||||.||
EMBOSS_002     14176 GAGAGAAATAATTTTCTTTCAACAGATTATTTATTTT---GT-TGTGGTT  14221

EMBOSS_001     12982 GATTCATTATT-AATTAACTT---TATGAA-----------------TTA  13010
                      ||  |||||| ..||.||||   ||.|||                 |||
EMBOSS_002     14222 -AT--ATTATTGTCTTTACTTATGTAAGAATAACTAGAGTAAAGTGGTTA  14268

EMBOSS_001     13011 TATGAACACCACAATTACTTTCAA--GCAG---TTACAC----AAATTTC  13051
                     |...|||||||       |||.||  ||||   |||..|    ||.||||
EMBOSS_002     14269 TCAAAACACCA-------TTTGAAATGCAGATATTATGCCTGAAATTTTC  14311

EMBOSS_001     13052 TAAATTCA-GTGGA-----------------------AATGTATTTATT-  13076
                     |||..|.| ||.||                       |||.|..|||||
EMBOSS_002     14312 TAATGTAAGGTTGACCTTCTAGACAGAATAGCTAATTAATCTTCTTATTA  14361

EMBOSS_001     13077 ----TTTGG-----------TATGTTTATTTTTCATTTTTATTTTATTTT  13111
                         |||||           ||||||.|  |..|||||.||....|.||.
EMBOSS_002     14362 GACATTTGGAATAAAGAGACTATGTTAA--TGGCATTTATAAGGGAATTA  14409

EMBOSS_001     13112 T---TATAATTTCAACT--TTTG--GATTCAAGAGTAAATGTGCAGGTT-  13153
                     |   ||||.|||  |||  ||||  .||||       .||||   ||||
EMBOSS_002     14410 TACATATAGTTT--ACTTATTTGTCTATTC-------TATGT---GGTTC  14447

EMBOSS_001     13154 --TGTTACGTGTGTATATTGTGTGACACTGAGGTTTGGTGCTGAGCATAG  13201
                       ||.|.|.|.|.|.|||      |.|.|    |||    .|.||.|.|.
```

```
EMBOSS_002     14448 TCTGATTCTTTTCTTTAT------ATAAT----TTT----ATAAGTAAAA  14483

EMBOSS_001     13202 TATCCAAT--------------AGCTACT---------------------  13216
                     || |||||              |..||||
EMBOSS_002     14484 TA-CCAATCTCATCAAATATAAAAATACTATGACAATAAAACATAACAGG  14532

EMBOSS_001     13217 --------ATTTTG-----------GACCTTTTCCCCTCACTCCCTCCC-  13246
                             |||.||           |||   ||.|.|||.||.||||||
EMBOSS_002     14533 AAAAATTCATTATGAATGCATGCATGAC---TTTCTCTCTCTTCCTCCCT  14579

EMBOSS_001     13247 CAC------------------TCTA-GTAGTCCGC---AGTGTTTCTTGT  13274
                     |||                  |||| |||||.|||   ||||      |.
EMBOSS_002     14580 CACTTTGTGTGTGTGTGTGTGTCTATGTAGTGCGCATAAGTG------GG  14623

EMBOSS_001     13275 TCCCATCTTTATATCCATGTGTATCCAAGA--AAA------CATGGTTTT  13316
                     .|.|||          |.|||.||..||||  |||      .|||..|||
EMBOSS_002     14624 ACACAT----------AAGTGAATATAAGAGTAAAAGTTTGGATGAATTT  14663

EMBOSS_001     13317 CA------ACCATAATGCTGTGTTCCACAAAAATATATGTTCATGTTATT  13360
                     ||      |||||.|..||.|.||                    ||.||
EMBOSS_002     14664 CATTCTCCACCATTACTCTTTTTT---------------------TTTTT  14692

EMBOSS_001     13361 TTATTTTATTTATTT--ATTTTTGAGACAGAGTCTTACTCTGTCA---CC  13405
                     ||.||||.|||.|||  .||||.|||||||.|  ||.|||||| |   |.
EMBOSS_002     14693 TTTTTTTTTTTTTTTGGGTTTTCGAGACAGGG--TTTCTCTGT-ATAGCT  14739

EMBOSS_001     13406 CAGGCTGGAGTGCAATGGTGTGATCTCGGCTCAC--TGCA-ACC--TCT-  13449
                     |.||||||...||.||              |||||  ||.| |||  .||
EMBOSS_002     14740 CTGGCTGTCCTGGAA--------------CTCACTTTGTAGACCAGGCTG  14775

EMBOSS_001     13450 GCCTCCCGGGTTCAAGGGATTTTCCTGCCTCAGCCTCCAGAGTAGCTGGG  13499
                     ||||  ||...|| ||..||...||||||||.||.|||.|||| ||||||
EMBOSS_002     14776 GCCT--CGAACTC-AGAAATCCGCCTGCCTCTGCTTCCCGAGT-GCTGGG  14821

EMBOSS_001     13500 ATTATAGGTGCCTGCCACCATGCCCAGATAATTATGTTTGTA-TTTTTAG  13548
                     ||||.|||.|..||||||||.||||.|.|.|..||     || |.||.|.
EMBOSS_002     14822 ATTAAAGGCGTGTGCCACCACGCCCGGCTCACCAT-----TACTCTTCAA  14866

EMBOSS_001     13549 T---AGAGACAGGGTTTCACC---ATGTTG-GCCAGGCTGGTCTCGAACT  13591
```

```
                    |   |||..|     ||..||   ||||.| ||..||.|..|||| ||||
EMBOSS_002     14867 TGCAAGAATC-----TTTTCCTGAATGTAGAGCTTGGATATTCTC-AACT  14910

EMBOSS_001     13592 CCCGACCTCAGGTGATCCACCCGCGTTGGCCCCCCAAAGTGCTGGGATTA  13641
                               ||.|||...||..||  ..||||               |.|
EMBOSS_002     14911 ---------AGATGAGAAACTAGC--AAGCCC---------------TAA  14934

EMBOSS_001     13642 CAG-GTGTGAGCCACTGCACCCAGCCCATGTTATCA----TTTGTACAGA  13686
                     ||| ||.|||..|.||||..||           |||    ||||    |.
EMBOSS_002     14935 CAGTGTATGAATCCCTGCTGCC-----------TCAGAGCTTTG----GT  14969

EMBOSS_001     13687 TATA----TTCAATG------TTGAAATTTTAAAGTGAATTT--GTGATA  13724
                     ||||    |||  ||      .|.||||||     ||.||.|  |||.||
EMBOSS_002     14970 TATAGTTTTTC--TGAGGGAAGTCAAATTT-----TGTATGTAGGTGCTA  15012

EMBOSS_001     13725 AAA------------TTCACAATAATGTCAGATATTTCACCATCAAAAGT  13762
                     |.|            ||||.||.||.|||   ||.||.|||||..|||
EMBOSS_002     15013 AGATCCAGAGTCTGGTTCAAAAGAAGGTC---TACTTAACCATGGAAA--  15057

EMBOSS_001     13763 ATAAACTTCCA-----ATGAAGTTGATCCAGTGAATAACATGAAAGTAAA  13807
                     |.|.|||.|||     || ||| ||||                |.|||||
EMBOSS_002     15058 AAATACTCCCAGCCCCAT-AAG-TGAT----------------ATGTAAA  15089

EMBOSS_001     13808 TAAATTGGAATTAAATAATTGTAATTAAAACAAAAGAAGGGTGTAAGGTT  13857
                     |.|||..|.||.||
EMBOSS_002     15090 TTAATGAGCATCAA-----------------------------------  15103

EMBOSS_001     13858 TATGCCTAGATAAACTTTTTTGCATGTAGATGTTGTGTTGTTCAGGACCA  13907
                         |.|||     ||||||   || ||.||      |||||||
EMBOSS_002     15104 ----CTTAG-----CTTTTT---AT-TATAT------TTGTTCA------  15128

EMBOSS_001     13908 TTTGTTTAAAGACTTTTTCTCCACTGTATT---GCCTTGG---TTCCTTT  13951
                                 ||||     ||||||||   .|.|.||   ||.||||
EMBOSS_002     15129 -------------TTTT-----ACTGTATTAAAACATAGGATTTTTCTTT  15160

EMBOSS_001     13952 GCCAAAAATCAATTGATTATATTTATGTGGATCTGTTTCTGAGCTTTCTA  14001
                     |           ||.|.|||.||.||.|       .||.|.|   |||.|
EMBOSS_002     15161 G-----------TTTAATATTTTCATAT-------ATTATAA---TTCAA  15189
```

```
EMBOSS_001     14002 TTCTGTTCCATTGATTTA------CCTGTCTGTTCT-----TTTGCCAAT  14040
                     .|..|    |.|.|.|||      ||||| ||.|||     |||   |||
EMBOSS_002     15190 ATAAG----AATAACTTAAACCCTCCTGT-TGCTCTTGGTATTT---AAT  15231

EMBOSS_001     14041 ATTACTCTGTCTTGATTACTGCAGTTTTATATCAAGCCT----TGAAGTC  14086
                     |.||   ||   |.|||||| .||..||||..|...|||    ||.|  |
EMBOSS_002     15232 AGTA---TG---TAATTACT-TAGAATTATCCCCTTCCTACTGTGCA--C  15272

EMBOSS_001     14087 AGA------TATCGTCAGGTCTCCCACTTAGGAAACTCTTTAAAATAAAT  14130
                     |||      ||||             .|||.|||      |||.||...|
EMBOSS_002     15273 AGAGCACATTATC-------------ATTATGAA------TAACATTTCT  15303

EMBOSS_001     14131 TTTT----AAGAA-------AACCTTTATTCAGGCCAGGCACGGTGGCTC  14169
                     ||||    |||||       ||   |||||             |.|..||
EMBOSS_002     15304 TTTTAGTGAAGAAGGTAGTGAA---TTATT-------------GGGAATC  15337

EMBOSS_001     14170 ACACCTGTAATCCCCCCGCCCCACCCCAGCACTGAGCTGGGCGGATGATG  14219
                     |.|                       ||.||.|.|||||
EMBOSS_002     15338 ATA-----------------------CATCAGTAAGCTG-----------  15353

EMBOSS_001     14220 AGGTCAGGAGTTTGAGACCAGCCTGGCCAACATGGTG----AAACCCTGT  14265
                               .|.||..||||.| .|.||.|.|.||    ||||..||
EMBOSS_002     15354 -----------CTAAGCACAGCAT-ACGAATAAGTTGTTACAAACTATG-  15390

EMBOSS_001     14266 CTCTACTAAAAATACAAAAATT----AGCTGGGCGTGGTGGTGCACGCCT  14311
                             .||.|||||||||    ||.||.|
EMBOSS_002     15391 ---------GAACACAAAAATTGAACAGATGAG-----------------  15414

EMBOSS_001     14312 GTAATCCCGGCTACTCGGGAGGCTGAA-GCAGGAGAGTTGC-------TT  14353
                       ||               |||.|||| ||..|||..||.|       ||
EMBOSS_002     15415 --AA---------------AGGGTGAATGCTAGAGCTTTACTTTTGATTT  15447

EMBOSS_001     14354 GAACC--CGGGAGGCAGAGGTTGCAGTG-AGCTG------------AGA-  14387
                     .||||  ..|||||.|      |..|||| |.|||            |||
EMBOSS_002     15448 TAACCTTATGGAGAC------TCTAGTGTAACTGTTTTATTTTTACAGAC  15491

EMBOSS_001     14388 ---TCATGCCACTGCACTCCAGCCTGGGCGACAGAGTGAGACTCTGTCTC  14434
                        |||.|.||.|      ||||           |||     ||
EMBOSS_002     15492 TTCTCAGGTCAAT------CAGC-----------AGT-----TC------  15513
```

```
EMBOSS_001     14435 AAAGAGAAACAAAAACAAAAAGCTTTACATTTGTAGCTATTTCT------  14478
                         |||||.|            |||.|..|||||  |.|.||||
EMBOSS_002     15514 ----AGAACC------------CTTCATGTTTGT--CCAATTCTCTCCTT  15545

EMBOSS_001     14479 -TTAAATTAGATAGTCTTTTCA---GATACATAAA------------AAT  14512
                      |.||||||..|      |||||   .|||| ||.|            |||
EMBOSS_002     15546 ATGAAATTGAA------TTTCAAGGCATAC-TAGAGCTTTTCCTGTTAAT  15588

EMBOSS_001     14513 ATGTGGATTATTGTAT--------------------AAGCATTTACAT--  14540
                        ||||  ||.|.||                    ||.||     ||
EMBOSS_002     15589 ---TGGA--ATGGGATCCTGGGAACACAGAATCACAAAACA-----ATGC  15628

EMBOSS_001     14541 AATTTAATGTGCTTAGTAAAT-----AGTG-----TTGTCTAGATGATTA  14580
                     |||.| ||||| ||.|.||||     ||||     |  .|.|||| ||||
EMBOSS_002     15629 AATCT-ATGTG-TTGGGAAATTTTCAAGTGAAAAAT--CCAAGAT-ATTA  15673

EMBOSS_001     14581 --------CTTTTTCTC-TGCAACTATTTCTGATATTGCCATATTATATG  14621
                             | |||||.| |.|||.|.||||   .|||  ||..||||||..
EMBOSS_002     15674 AAACATGTC-TTTTCACTTTCAAATCTTTC---AATT--CAGGTTATAAA  15717

EMBOSS_001     14622 AAAGAGTT-TCCCTCTGTCGCCCAGGCTGG------AGTACAGTGGCGCA  14664
                     |.||.||| ||.||..|..|||   ..|||      ||||.|     ..|
EMBOSS_002     15718 ACAGGGTTATCTCTTGGATGCC---TTTGGCAACATAGTAAA-----AAA  15759

EMBOSS_001     14665 ATG-ATAGTTCGTT-------GCA-GCCTACTCAGACTACTGTGCTCAAG  14705
                     ||| |||.||  ||       ||| |||          |.|||.|.||
EMBOSS_002     15760 ATGTATATTT--TTCATGTAGGCATGCC----------AATGTACCCA--  15795

EMBOSS_001     14706 CAGTCCTTCTGCCTCAGATTCCTGAGTAGCTGGGACTACAGGCTCCTGTC  14755
                          ||  ||..|.|..||      ||||                 ||
EMBOSS_002     15796 -----CT--TGGTTGAATTT------TAGC------------------TC  15814

EMBOSS_001     14756 ACCACGCCT---GACTAATTTTTTAAAATTGTTTGTTTAGAGATGGAGGT  14802
                     ||.||.|.|   .|.|||.||||.||||||..||||||.||..||
EMBOSS_002     15815 ACTACTCATTTATAGTAAATTTTAAAAATACTTTGTATATTGA-------  15857

EMBOSS_001     14803 CTCTCGAT--TTTGCCTAGG------CTGATCTCAAATCCCTGGACTGAA  14844
                           ||  |||| |||.|      |||||         ||||..|||
```

```
EMBOSS_002     15858 ------ATCATTTG-CTATGAAAAATCTGAT---------CTGGTTTGA-  15890

EMBOSS_001     14845 GCAGTCTTCCTGCCCAGGCCTTCTGAGTAGCTCGGATTACAGGCCTATCT  14894
                         |.||         ...||.|.|.||  |..|.|||.|....|||||
EMBOSS_002     15891 ----TGTT---------AAATTATTATTA--TTTGTTTATATTTATATCT  15925

EMBOSS_001     14895 --------CTTAACTATTTGCTTTTAA-AGGAATGTATTAGATCTGCGCT  14935
                             |||.||   |||.|||.|| |||||           .||.|.
EMBOSS_002     15926 TCATCTGCCTTCAC---TTGTTTTCAACAGGAA-----------GGCCCA  15961

EMBOSS_001     14936 G----------TACAATCACAGTTAACACTGGTCACATGTAACCTTTAAA  14975
                     |          ||||||.|.||.|||||.||||.|||||||      |||
EMBOSS_002     15962 GGCAAACATGATACAATTAGAGATAACAGTGGTAACATGTA------AAA  16005

EMBOSS_001     14976 TTTTAAATTATTTAAAATAAATAAAATTAGATTAAATTAAATATTCTTT-  15024
                     |||              |||||.|.|||||||||||||||     |||
EMBOSS_002     16006 TTT--------------TAAATCACATTAGATTAAATTAAA-----TTTA  16036

EMBOSS_001     15025 ----TATTTTC--AAGTCATGCTGACCACCTTTCAAGTGCTGAATACCCA  15068
                         ||.||||  ..||.||  ||.|.||||||.|..|.||.|.||  |.
EMBOSS_002     16037 AAAATAATTTCCTGTGTAAT--TGGCAACCTTTTAGCTACTCAGTA--CT  16082

EMBOSS_001     15069 CATGTGGTTATTGGCTATCATATTG-----GATGACACAGATA-CAGATC  15112
                     ||||||  ||.||.||| .|||.||     ||..||||||.|| .|.||
EMBOSS_002     16083 CATGTG--TAGTGACTA-GATACTGTAATAGAAAACACAGCTATTAAAT-  16128

EMBOSS_001     15113 ATTTCCATCCCTGC-ACAAATTTATACTCAACAGTGGTGCAGACTGTCCA  15161
                     ||||.|||  .||| |||.||.|.||.|.||.||...||.|||.|.||||
EMBOSS_002     16129 ATTTACAT--TTGCAACACATGTCTAATAAAAAGCATTGAAGAATATCCA  16176

EMBOSS_001     15162 TGTATCTGTTATTTGCCCTCCCAT------CTG----TACTCCAGTATTT  15201
                     .|                |.||||      |||    |||||.|.||||
EMBOSS_002     16177 AG----------------TTCCATAAAATGCTGATTATACTCTAATATT-  16209

EMBOSS_001     15202 TACCTGGA-------AATATTTTTTTAAATGCCTGTGGATGTAAATGAAG  15244
                           ||       |||..||.|||||||.|..||.||  .|||||.||.
EMBOSS_002     16210 ------GAATCATAGAATGCTTCTTTAAATTCAAGTAGA--AAAATTAAA  16251

EMBOSS_001     15245 TCTTAA-----AAAATAGTCTCATGTAGAATAACCTTCTGCCTCAAAT-T  15288
```

```
                    |.||.|     ||.|||    .||..|||.||||.|         ||| |
EMBOSS_002     16252 TTTTGATTTGGAATATA----AATCCAGACTAACTT---------AATCT  16288

EMBOSS_001     15289 CAAGAGAAATCTTTGTT----------ATAATATGATTGTTAAGTCACAG  15328
                     ||.|..||||||.|.||          ||||.||||||        |.||
EMBOSS_002     16289 CATGTAAAATCTATCTTTCTATAGCAGATAAAATGATT--------ATAG  16330

EMBOSS_001     15329 GTGT-GTTGT-----TC----TTACTATATT-TTAATTGCGT------TA  15361
                     .||| |.|.|     ||    ||| |..||| ||||||||.|      |.
EMBOSS_002     16331 ATGTAGATATAGACATCAGAATTA-TGGATTGTTAATTGCATTTTCACTC  16379

EMBOSS_001     15362 TAGTTATTT------GAATAGTTATGTTTTAGGACATACACATTGTTATA  15405
                     ||.||.|||      .||||     |.||.||||||||.|       |||
EMBOSS_002     16380 TATTTCTTTTGTAAGAAATA-----GGTTGAGGACATAAA-------ATA  16417

EMBOSS_001     15406 GTCAACACTTCTTGATGATTATAGTAATGTAGTTCTAC---------TCT  15446
                      |||               |||| ||| .|||..|.||         |..
EMBOSS_002     16418 -TCA---------------TATA-TAA-CTAGAGCCACATTAAAAAGTTA  16449

EMBOSS_001     15447 GTCATAGTTTAAAATTTATGTCTCCATAGG----AATAT---TATTAGCT  15489
                     |||||| ||||..|.||||.|.|.|||..|    .||||   |||.|| |
EMBOSS_002     16450 GTCATA-TTTATTAATTATTTGTGCATGTGTTTCTATATGGGTATGAG-T  16497

EMBOSS_001     15490 AAACT-CA----ATAGTATGAAATTCTGGCTATTGTTAACTGAAATAATT  15534
                     |.||| ||    ||| |||||.|     ||.|  |..|||
EMBOSS_002     16498 ATACTACAGAATATA-TATGACA-----GCCA--GAGAAC----------  16529

EMBOSS_001     15535 TTGCATTTGTGGTTCAAAATGTAACTCTTAAGGGAAATAA-----AGGAA  15579
                                            |||| ||.|||| |||||     |||||
EMBOSS_002     16530 ----------------------AACT-TTCAGGG-AATAACGTACAGGAA  16555

EMBOSS_001     15580 TTC-AGAAAGT---TTTTTGA-----AACTCAATCT-------TGTC---  15610
                     ||| |..|.||   ||.|.||     ||||||..||       ||.|
EMBOSS_002     16556 TTCTACCATGTGGGTTCTGGAAAATGAACTCAGGCTGTTAGTATGGCAGT  16605

EMBOSS_001     15611 ------TTC--CTTGGAAATAGGAGTACTTTGATTATTTTTGAAAGAAAA  15652
                           |||  ||||     .|||||..|.|.|.|
EMBOSS_002     16606 GATCAGTTCCACTTG-----CGGAGTCATCTCACT---------------  16635
```

```
EMBOSS_001     15653 ATTAGCCTACAGC----CTGGTATTTAAATATATATGTACATATATTATA  15698
                         .|||||||    .|.||.||.|||...||     ||||    |||
EMBOSS_002     16636 -----GCTACAGCTAAGATAGTCTTAAAAGCGAT-----CATA----ATA  16671

EMBOSS_001     15699 TATTTCTGCTAAATTATTTATGGTAGTTT--ATTTTTTCCATCTTATATA  15746
                             .|||||          ||||  |||       ||||||
EMBOSS_002     16672 ---------AAAATT----------GTTTGAATT-------TCTTAT---  16692

EMBOSS_001     15747 CATACTGGATTCTCAATTTGATTTTTAATACCGCCTATATACTTATTAGT  15796
                             .|||.|   .||||||||    .|.|||||||...|||.|.
EMBOSS_002     16693 ---------GTCTAA---CGATTTTTA----TGGCTATATAGCCATTGGG  16726

EMBOSS_001     15797 AATTTCAATGGTGTATCTTTAAAAGATAAATTTCATTTTAGTTATGTGAC  15846
                     |||                   |||||||                .||.|
EMBOSS_002     16727 AAT-------------------AAGATAA----------------TTGGC  16741

EMBOSS_001     15847 ACTTTATCTTTCATTGTTA---TGAATTGCCTTTTTACTTTTTGCAGTCT  15893
                     |..|.||.|||  ||||||   ||||.||  ||||.|.|||
EMBOSS_002     16742 AGGTAATATTT--TTGTTAATTTGAACTG--TTTTAAATTT---------  16778

EMBOSS_001     15894 TGCGTTGAAATGTATCAGAAACTATAATGTAAAAAAAAGCTGAGTAGAAA  15943
                             |.||| ||| |||...||||          .||||.|||.||
EMBOSS_002     16779 --------ATTGT-TCA-AAAGGTTAAT----------TCTGATTAGGAA  16808

EMBOSS_001     15944 TCTT----ATAATTAAAAGTTGTAGCAAG-TC-------ATGAAAATGGC  15981
                     |.||    |     ||.|||| |||.||| ||       |||||.||
EMBOSS_002     16809 TTTTGAGCA-----AACAGTT-TAGAAAGCTCTTAGAATATGAACAT---  16849

EMBOSS_001     15982 TCATG-CTTTTATTGCCATTTTG-----ATG-----TT---TTTGATGGC  16017
                        || |||||       .||||     |||     ||   .|||||
EMBOSS_002     16850 ---TGACTTTT-------GTTTGAAAATATGAGAACTTCGAATTGAT---  16886

EMBOSS_001     16018 AAAAGTGTTGAGAAAA--AGTCTT-TAGATTCACGTGATAAGCTGACAGA  16064
                       .|.|||||| ||||  |||.|| .|||||.||.||||.||.|||.|.|
EMBOSS_002     16887 --TATTGTTGA-AAAATCAGTATTAAAGATTTACATGATGAGTTGATAAA  16933

EMBOSS_001     16065 GTGAAACATCTTAAGGCTTGAAAGGGCAAGTAGAAGTTATAATTATTGTG  16114
                     .|||||..|||..||||..||||||...||.||...||||||.||||||.|||.
EMBOSS_002     16934 ATGAAGTATCAGAAGAATTGAAAAATCAGGTTACAGTTACAATTACTGTT  16983
```

```
EMBOSS_001     16115 TAGATTCAC--AGTCCTTGTATTGAATTACTCATCTTTGCTCTCATGCTG  16162
                     ||||  |||  |||..|||.|.||||.|.||||.|||||||||||||||
EMBOSS_002     16984 TAGA--CACCAAGTTGTTGCACTGAAATGCTCACCTTTGCTCTCATGCTG  17031


EMBOSS_001     16163 CAG  16165
                     |||
EMBOSS_002     17032 CAG  17034


#---------------------------------------

#---------------------------------------
```

8.1.2. Alignment of human and mouse intron 55 of DMD/Dmd gene.

```
#######################################
# Rundate: Thu 23 Feb 2023 15:42:50
# Commandline: matcher
#######################################

#=======================================
#
# Aligned_sequences: 2
# 1: EMBOSS_001 - HUMAN DMD INTRON 55
# 2: EMBOSS_002 - MOUSE DMD INTRON 55
# Matrix: EDNAFULL
# Gap_penalty: 16
# Extend_penalty: 4
#
# Length: 34697
# Identity:   18393/34697 (53.0%)
# Similarity: 18393/34697 (53.0%)
# Gaps:        5254/34697 (15.1%)
# Score: 9241
#
#
#=======================================

EMBOSS_001         1 GTAAGTCAGGCATTTCCGCTTTAGCACTCTTGTGGATCCAATTGAACAAT     50
                     ||||||...|..|||||.||||||.||...|..|||.|||||.|.|||||..|.
EMBOSS_002         1 GTAAGTTGAGTGTTTCAGCTTTGGCTGGCAAGTGAATCCCACTGAAGCAG     50

EMBOSS_001        51 TCTCAGCATTTGTACTTGTAACTGACAAGCCAGGGACAAAACAAAATAGT    100
                     |||.|||||||||||||||..||||||||.|..||||||||||..||  |||
EMBOSS_002        51 TCTAAGCATTTGTACTTGATACTGACAAACTGGGGACAAAAATAA--AGT     98

EMBOSS_001       101 TGCTTTTATACAGCCTGATGTATTTCGGTATTTGGACAAGGAGGAGAGAG    150
                     || |||.|..||||.||||.||.||..|.||| ||.||||.||.|.||||
EMBOSS_002        99 TG-TTTCACTCAGCTTGATATAGTTTAGCATT-GGGCAAGAAGAAAAGAG    146

EMBOSS_001       151 GCAGAGGGAGAAGGAAACATCATTTATAATTCCACTTAACACCCTCGTCT    200
                     |||||||.||.|..||...||||||.|.|||.||.|.||.|||.|...||
EMBOSS_002       147 GCAGAGGCAGTAAAAAGACTCATTTGTGATTTCATTAAATACCATGAGCT    196

EMBOSS_001       201 TAGAAAAAGTACA----TGCTCTGACCAGGAAAACATTTGCATATAAAAC    246
                     |..|.||||||||    |.|| |.|||||.||||.|.||||.|| ||.|.
EMBOSS_002       197 TCAAGAAAGTACACATCTACT-TCACCAGAAAAATACTTGCCTA-AAGAG    244

EMBOSS_001       247 CAGAGCTTCGGTCAAGGAGAAACTTTGCTCAGAGAAATAACTT-AGGGAT    295
                     ||.||.||..||||||.||||||.|..||||||||||.|||.|| ||.|.|
EMBOSS_002       245 CATAGGTTTTGTCAAGCAGAAACGTAACTCAGAGAAGTAAGTTCAGCGTT    294
```

```
EMBOSS_001       296 TGGTTTATTAAATTTTAAAAGTTGACATTTTTGAGTGTTTATTTAATATT    345
                     |..|||||||||||..||..||||.||||||||.|.||||||||||||.|.
EMBOSS_002       295 TAATTTATTAAATTAAAATGGTTGCCATTTTTCACTGTTTATTTAATGTC    344

EMBOSS_001       346 TTACAGGGAAAGCATCTGTATGAATTGTCTGTTTTATTTAGCGTTGCTAA    395
                     ||||||||||||||||||||.|||.||||||||||||||||||||||||||
EMBOSS_002       345 TTACAGGGAAAGCATCTGTAGGAACTGTCTGTTTTATTTAGCGTTGCTAA    394

EMBOSS_001       396 CTGAATCAGTTTCCCTTCATTACTTTCAAATATGTTTTGAAATGTTAATC    445
                     ||.||||||||||||||||.||||||||.||||..||.||||||||||||||
EMBOSS_002       395 CTAAATCAGTTTCCCTTCGTTACTTTCCAATACATTCTGAAATGTTAATC    444

EMBOSS_001       446 TGGCATTTTGTAGCTTTCTTCCTAACATGATCTGTGAAAATAAGAATGAG    495
                     ||.||||||||||| |||||||||||||||||||||||||||||||||||||
EMBOSS_002       445 TGACATTTTGTAG-TTTCTTCCTAACATGATCTGTGAAAATAAGAATGAG    493

EMBOSS_001       496 ATGGCTGAATTTGTCGTAGTTAATGATCAAACAATTTTCAGACAATTGTT    545
                     ||.|||.|||||||..||||||.||.|....||||||||||||||||||||
EMBOSS_002       494 ATTGCTAAATTTGTTATAGTTAGTGGTTGTGCAATTTTCAGACAATTGTT    543

EMBOSS_001       546 TTT-CCTAGAAACAAAAATTATTTCCATAAAGTTCCATATGCATAAACAG    594
                     ||| |||.||||||||||||||.|       |.|||..|.||||||.|||||
EMBOSS_002       544 TTTTCCTGGAAACAAAAATTAGT-------AATTCTGTGTGCATGAACAG    586

EMBOSS_001       595 TGAAAACAGAA-CGTGGGGTAGTTTTGTTTAAATGAAGTCTTGGTGAGAA    643
                     ..||||||||| |.|.|.|||.|.|||||||||||            |||
EMBOSS_002       587 CAAAAACAGAAGCCTAGAGTAATATTGTTTAAATG------------GAA    624

EMBOSS_001       644 TCATATTCTGTAGTACAAGGAGGCTCTTAAAGTTTA-TTCTCAATACCTG    692
                     ||||.|||||||||.|. |.||| |||.||||.||| ||||.|||..  |
EMBOSS_002       625 TCATGTTCTGTAGTTCC-GTAGG-TCTCAAAGGTTAATTCTTAATGG--G    670

EMBOSS_001       693 ATATAATTTTCCTGAACTATTATGGAGTTTTGTTATGTATAGTTGGTTTT    742
                     ||...|||||.|||||.||||.||.|...|||..||.|||
EMBOSS_002       671 ATGGGATTTTTCTGAATTATTTTGCAACATTGCCATATAT----------    710

EMBOSS_001       743 TCTGACTTGATATAATAACTTTACTAGTCTCTCAAATACAATTTGGATAT    792
                                             |.|.|||           ||||.
EMBOSS_002       711 -----------------------CCATTCT-----------TTTGC----    722

EMBOSS_001       793 AAATCATTATAATAAGATGATTGATTTTTTAGACTAACTTTATTTTTTGA    842
                                         |..|||.||||||.|      |||||
EMBOSS_002       723 ------------------GCCTGAGTTTTTGG------TTTAT-------    741

EMBOSS_001       843 TATTTTTAAACTATTATGAAAAACTATTATGAAACTATTATGATATTTTT    892
                                     |||.|||..||. |||.|..|..| ||||||.|..
EMBOSS_002       742 ---------------ATGTAAATGTAA-ATGCACATCAT-TGATATCTAG    774

EMBOSS_001       893 AAACTATTATGAAAAGTATATTCTAGTTTGAATAATTCCAGAATCAAATC    942
                     ||..||.|.||    |.||||.|  |.|||.||.||||
EMBOSS_002       775 AATGTAGTTTG----GCATAGT--AATTTAAAGAATT-------------    805

EMBOSS_001       943 ATAATAAGCAGAAGTTCTTCTCCTCTCCCTCCTATCGTTCTCCTTCTCCT    992
                                                 ||||||..
EMBOSS_002       806 ------------------------------TCCTAGA--------------    812

EMBOSS_001       993 GTTTTTCTTTTTTGATATGATAGTTGATCTACTTTGCTGCTCTGTTGCAT   1042
                           ||.||||||.|.|||.|.|||.| ||.||||.||||.||.||||.
EMBOSS_002       813 ------CTATTTTGTTGTGACAATTGGT-TAATTTGTTGCTTTGGTGCAG    855

EMBOSS_001      1043 AGAGTACGTAACAGTGGCAAT-GTATGGC-TCCTGAATTTATCGTTCTTG   1090
                     |.||||.||||||.....||| ||||..| ||.||||.||||...|||||.
EMBOSS_002       856 ACAGTAGGTAACAACAATAATTGTATTACATCTTGAACTTATTACTCTTT    905

EMBOSS_001      1091 CTTCATCATCCTGCTTTGACCCCACTTTCTCCTCCAAAATGCGTGTTGAG   1140
                     .||    |||.|.||  |.||||||||.|...||||.||.|. ||.|...
EMBOSS_002       906 TTT----ATCTTTCT--GGCCCCACTTGCCTATCCACAACGA-TGCTAGT    948

EMBOSS_001      1141 TTAGTTTGATCATTTGGAGGTAATTTGTTTGGAACAGTATCAGACTTTAT   1190
                     ||.|||||.||.|.|||||||||||||..|..||.|||.||.||||||||
EMBOSS_002       949 TTTGTTTGCTCCTGTGGAGGTAATTTGAATTAAATAGTTTCTGACTTTAT    998
```

```
EMBOSS_001      1191 ------AGATATCTCCCATGGCTTGTGATAGAATATAAGGGCAATGCAAA   1234
                           |||.||||||||.             |..|||..|..|||.|||
EMBOSS_002       999 CTCTGCAGAGATCTCCCAC-------------ACGTAAAAGTCATGAAAA   1035

EMBOSS_001      1235 TGTAGAGTTTTTTGCTCACTCTTCGATGTATGGTTAGAC--AATGTACCA   1282
                     |..|||.|||.||        ||.|.||.|....||.|.  |.||||...
EMBOSS_002      1036 TCCAGATTTTATT--------TTGGCTGAACCACTACAGTGATTGTAGTG   1077

EMBOSS_001      1283 CTGTAATATATTTGGCTTAGGCTATTTCATAAATAAAATTTTATTATAAA   1332
                     .|.||....||||||||||.||.|||.||||||||||.|.|||||..|..
EMBOSS_002      1078 ATATAGCTCATTTGGCTTATGCCATTCCATAAATAAAAATGTATTTCACC   1127

EMBOSS_001      1333 ATATTATAAATGCTGATAAAGCTACTCCAGAATTTTAATAGATATGTGGG   1382
                        |||||||||.....|||||.|....||||.||||||.|.||.|||..
EMBOSS_002      1128 ---TTATAAATGGCTGCAAAGCCATAATAGAAATTTAATTGGTACGTGAA   1174

EMBOSS_001      1383 TTTCCCGGCCAGATGCGGTGGCTCATGCCTGTAACCCCAGCACTTTGGGA   1432
                     ||       ||.||.       |..|.||||.|||  .|..|  |||..|
EMBOSS_002      1175 TT-------CACATT-------TTTTTCCTGCAAC--TATGA--TTGATA   1206

EMBOSS_001      1433 GGCCGAGGTGGGTGGATCACCTGAAGTCAGGAGTTCGAGACCAGCCTGGC   1482
                     ..|.|....|..|.|||.||.|.       |..||..|.||.|
EMBOSS_002      1207 AACAGTATAGCTTTGATTACATA-------GCTTTTTAAACAA-------   1242

EMBOSS_001      1483 CAACATGGCGAAACCCCATCTCTACTAAAAATACAAAAATTAGCTGGGTA   1532
                      |..|||..||||          |.|||||.|....||||||.||...||
EMBOSS_002      1243 -ATTATGCTGAAA----------ATTAAAACTGATCAAATTACCTAAATA   1281

EMBOSS_001      1533 TGGTGACCTGCGCCTGTAATCCTAGCTACTTGGGAGGCTGAGGTGGGAGA   1582
                           |||.        ||...||.||.||...||.||
EMBOSS_002      1282 ------CCTT--------ATGTCAGATATTTAATAGACT-----------   1306

EMBOSS_001      1583 ATCGCTTGAACCCAGGAGGCAGAGGTTGCAGTGAGCCGAGGTGGCGCCAC   1632
                      ||...|.||...|          ||.|.|||.||     ||.|.|
EMBOSS_002      1307 -TCCTCTCAAAATA----------GTGGAAGTCAG-----GTAGGG----   1336

EMBOSS_001      1633 TGCACTCCAGCCTGGGTGACAAAGTGAGACTTCATCTCAAAACAAATAAA   1682
                                  |||..|.||||   |...|||||.    |.||..||.|
EMBOSS_002      1337 ------------TGGAAGTCAAA---AATATTCATG----ATCATTTATA   1367

EMBOSS_001      1683 TAAATAAATAAAAATACATGGGTTTACATTTTACCCATCAGCTATGGTAG   1732
                     |...||.|||....|   |..||.||..||.|||.|.|.||..||..||
EMBOSS_002      1368 TTTGTATATATGGTT---TATGTGTAAGTTATACTCTTTAGAAATTATA-   1413

EMBOSS_001      1733 GTAAATAATAAGCTTTGATTAAGTCTATTTTAGTCTATTTTTAGCAGATT   1782
                       ||||..|.|   .|.|||||.|..|...|.|.||||         |||
EMBOSS_002      1414 --AAATTCTGA---ATAATTAAATTAAAAATTGGCTAT---------ATT   1449

EMBOSS_001      1783 ACTTTGAAAAATAAAGAATAACCCAATGACTAAAAAATTATTTTATGTCA   1832
                     .|||||..|||.||. |||||       ||..|.||......|||.|||..
EMBOSS_002      1450 TCTTTTTAAAGTAT-GAATA-------GAGAAGAAGGAACATTTGTGTTT   1491

EMBOSS_001      1833 GGGATTTAATAAAACATATCTTTAAATCTAGTTGAGGGCAAAAATACGTC   1882
                     .|.|||||.        |||||||    ||  ||     ||||||.|..|.
EMBOSS_002      1492 TGAATTTT--------TATCTTT----CT--TT-----CAAAAACAATTG   1522

EMBOSS_001      1883 TATTTTCTACTATACAATTTGTATTTATATCTGCTGTATTATATAATGAA   1932
                     .|..|  |||..|..||..||...||||..|.|..|||||    ||.|.||
EMBOSS_002      1523 GAAAT--TACCCTGAAAAGTGCTATTAGCTTTCTTGTAT----TACTTAA   1566

EMBOSS_001      1933 AATTTATCTCTATTTCTAATCTCAAGAAACTGCAAGCTTCTGAATCATTA   1982
                     |||.|||.|.|.|.|.||            ||||.||.|.|...| |||
EMBOSS_002      1567 AATGTATGTATGTATGTA------------TGCATGCATGTATGT-ATT-   1602

EMBOSS_001      1983 AAGGGAAGATTCACCATGTGTCC-TAACTATATTTAC-TATGGAAGCATG   2030
                         |...||..|..||||||.. ||..|||||.||. |||..|...|||
EMBOSS_002      1603 ----GTGTATGTATTATGTGTATATATATATATATATATATATATATATG   1648

EMBOSS_001      2031 GAAAATAAATATTTT-ATG-TTTAGATTTCTGATCTCTCTTTCAAAAGCA   2078
                     ..||..||||.|.|. ||| |||..||||||.||     ||.|..||.||
EMBOSS_002      1649 CTAATCAAATCTGTGCATGGTTTTTATTTCTTAT-----TTGCTCAATCA   1693

EMBOSS_001      2079 GTTGGAAATTATGCTGAGAAAATGTCTTAGCTTATCCCATGTTACTCAAG   2128
```

```
                          ||||.||.|||.||..||..|...||..||.     |||.|.
EMBOSS_002      1694 --------TTATTCTCAGATAAACTCCAATGATAAACCG-----CTCGAC   1730

EMBOSS_001      2129 AAAATGTATTTATTCGTTTTTGTCCAGTGGCTTAACCAAACCACAGTTTA   2178
                     |     |.||||..|..|..||.    ||||. |...|||.||......|
EMBOSS_002      1731 A-----TCTTTAAGCTCTAATGC----TGGCC-ACGTAAAACATTCAAAA   1770

EMBOSS_001      2179 TTTGTTGCTCACATAAAGTCCAGTGTCGA-TCAGGCTACTCTTT-TCCAT   2226
                     |||...|.|||.|.||.||..|.|.|.|| |.|||.....|||. |..|.
EMBOSS_002      1771 TTTTGGGGTCATAAAATGTTAATTTTAGAATAAGGAAGGACTTCCTAAAA   1820

EMBOSS_001      2227 CTTTGAGCTAAGGCACATATTA----CACATAACTTTCAGTGTACCCGAG   2272
                     .|.|.||...||.|...|.||.    |.||...||||.|   |.||.|..
EMBOSS_002      1821 GTCTTAGTGTAGCCGGGTGTTGGTGGCGCACTCCTTTAA---TCCCAGCA   1867

EMBOSS_001      2273 GTAGAAAAAGAGAGAGCTTGGGAATAAGGCAGGGGCTTTTTACTGTCTCA   2322
                     .|    .|.|||..||          |||||||.|..|||....|| |||
EMBOSS_002      1868 CT----CAGGAGGCAG----------AGGCAGGCGGATTTCTGAGT-TCA   1902

EMBOSS_001      2323 ACCCCAAAGTGATAAACTACATTTATTCTCA--AAATCCAGATAAAACTC   2370
                     |..||....||.|..||.| |.|.|.|..||  |.|.||||..|.| |.|
EMBOSS_002      1903 AGGCCTGCCTGGTCTACAA-AGTGAGTTCCAGGACAGCCAGGGATA-CAC   1950

EMBOSS_001      2371 CCATAGAGCCTCTGAAAACCTCAACATTTGCGTCTTAACTATAATAAGGT   2420
                     ..|.|.|.|||.|...||....||||...    |..|||.|.||.||...
EMBOSS_002      1951 TGAGAAACCCTGTCTCAAAAAAAACAAAA----CAAAACAAAAACAAAAA   1996

EMBOSS_001      2421 TAACTAAGATTCCAAAATTATTTTAAAACAGAGAC-AGTTTCCCTCTTCC   2469
                     .||  ||.|...|||||.|.||...|.|.||.||| |||.....|.||..
EMBOSS_002      1997 CAA--AACAAAACAAAAGTCTTAGTATAAAGTGACGAGTGGATATGTTGT   2044

EMBOSS_001      2470 C-TGGCAGCTAATATTGTATTTTCTATAAATCCACTTGCCCAAGGTTTAA   2518
                     . |.|..|.|..||||||..||..|.|.||.....|||..||     |||
EMBOSS_002      2045 TGTTGTTGTTGTTATTGTTGTTGTTGTTAAGGGTGTTGTTCA-----TAA   2089

EMBOSS_001      2519 ACTACATTTTATGGATTGAAATGACATTTATAGCCAACTCCTGATTTTTA   2568
                     || |.|||.|.|.            |||..||||||..||.|.||||.||
EMBOSS_002      2090 AC-AGATTCTTTA------------ATTGTTAGCCAGTTCATTATTTGTA   2126

EMBOSS_001      2569 GTTAGATGGTTGGATAATGATCTTTTGATGAAAGACTCGGAGATGTCATG   2618
                     |.     ||....|||.....|.|.||..||.|.|...||||  |.|.|.
EMBOSS_002      2127 GG-----GGCCAAATACAACACATGTGTGGAGAAAAAAGGAG--GACCTT   2169

EMBOSS_001      2619 GTAAAACGGTGAACTACTGAAACTATTGATTATTGTTAATGGCACATTT-   2667
                     |......|||..|...||.|...||||..||||.||  |.|..|..|.|
EMBOSS_002      2170 GGGTGCTGGTCCATGCCTAACCATATTACTTATGGT--AGGAAATCTCTG   2217

EMBOSS_001      2668 CAGCTGATTGAATTGAGTCAAGAAACTGGTGTTGAAGAG-CAACAAATGG   2716
                     |.||||.|||...| || |..|.|.||..||..||..|. |....|.|
EMBOSS_002      2218 CTGCTGCTTGCCCT-AG-CTGGGAGCTTCTGGAGATTATTCTGTTACT--   2263

EMBOSS_001      2717 AAATGCCGAG-CTTGAAAATAAATAAAGCAGCATACCTTAAGAGATTACA   2765
                     |..|.|.|.. ||||.|...  ||...|.|.. |||.|.|||...|||||
EMBOSS_002      2264 ACTTTCTGTTTCTTGCAGGC--ATCTTGGATT-TACATCAAGTATTTACA   2310

EMBOSS_001      2766 TGCAATTTCAGTATTTCAGCTAAA-TGGAAGTGTTTGCTTTTTTTCCTCT   2814
                     ||.|.|.|.||..|.|.|.||.|. |.|.|..||.|||.     ..|.|.
EMBOSS_002      2311 TGGATTCTGAGGCTCTGATCTCAGGTTGTAAAGTGTGCA-----GGCACC   2355

EMBOSS_001      2815 ATGAATTTTTATTTTGAACAAAAGGAATTTTCTATAATATGTAGGTAGGA   2864
                     .||.|.|||.|...||.||....|   ||.|.|.|||||..|||..|...|
EMBOSS_002      2356 CTGTACTTTCACCCTGTACTCTGG---TTCTTTCTAATTAGTAATTCTCA   2402

EMBOSS_001      2865 GAAAAGTGAAATGGCATGCTTTTTCACTTCATTTGAAGAAGCTGGTAGCA   2914
                     |.          |.||..|||..|.|  .|||||.|||..| |..|||
EMBOSS_002      2403 GC----------GTCAGACTTCCTGA--ACATTAGAATTA-CAAGTA---   2436

EMBOSS_001      2915 TTGTATTCATAGATTCATGCTGTATAGCAATCATAGTTCTCATATATTAA   2964
                      |.|||||...||....||.|||.|..|.|.|..||.||.|||.|....||
EMBOSS_002      2437 -TATATTGCCAGGCATATACTGAACTGTATTATTAATTATCAAAAGAAAA   2485

EMBOSS_001      2965 AAAAAAAGGAAATTTGAAATGCCTAGCCAAAGCAACAGCTCTGCCAACAG   3014
                     ||..||.|||.|.   |||||..|||.||||||.||.||||||..||||||
```

```
EMBOSS_002      2486 AAGGAAGGGACAC---AAATGTGTAGTCAAAGTAATAGCTCTCTCAACAG   2532

EMBOSS_001      3015 ATTTTGATATATCTGTCTACCCCAAAAGTAGTGATGATTTACTTCATACA   3064
                     .|.|||.|...||||||..||.|.||||.|......||||||..||...|||
EMBOSS_002      2533 TTCTTGTTTCCTCTGTTCACTCAAAAATTGAGCCAGATTTGTTTTCCACA   2582

EMBOSS_001      3065 AATGCTAGTGAATGAAGAGAGAGGGTGAAAACCTTCACAAAATGTGTTTT   3114
                     ||.||||.||.|||.|||.|.|.|.|.|.||.||.||   |||||    |.|
EMBOSS_002      2583 AAAGCTATTGCATGGAGAAAAATGATAAATACTTT---AAAAT----TAT   2625

EMBOSS_001      3115 TCTCTAAGACTGTCAATCCGTTTTTCTATATATGGAGA------CTCCAG   3158
                     |.|||.|.|.||.||||    |.|||.||||||.||||      |||...
EMBOSS_002      2626 TTTCTGAAATTGCCAAT----TGTTCCATATATTGAGAAAAAGACTCTTA   2671

EMBOSS_001      3159 CTCTTGCTAGACTACCTATCACTTTCGTCTATCAGCCACTTCGTAAGATA   3208
                     .|.||.|||||.|..||||..|.||.|..|||.|||.|||||...|..||
EMBOSS_002      2672 ATATTCCTAGATTGACTATAGCCTTTGGTTATTAGCTACTTCAATATCTA   2721

EMBOSS_001      3209 TTTA-TTCTCTCAGCAATAATCATAATTCATAGATTCTTTAAA-CATACA   3256
                     |||| ||||||||.|||.||||||||||.||.|..|||.|.|. |.|.||
EMBOSS_002      2722 TTTAATTCTCTCATCAAGAATCATAATTTATGGGGTCTATGAGTCCTGCA   2771

EMBOSS_001      3257 TGTAATATAAAGCATATACATTCTGAATGGAATTAACATGATTAATTCTT   3306
                     |.||.||||||.|||.....|.||.|||..||||||..|. ||..|||...
EMBOSS_002      2772 TATATTATAAATCATGAGATTCCTTAATAAAATTAGTAA-ATACATTTCA   2820

EMBOSS_001      3307 CTCTGAAAGACATTAGAATTTCCTCCCGTATTA-TAAAAAGGTGTAACTC   3355
                     |.||||||.|.||||.||||||||...|||||| ||||||||..|..
EMBOSS_002      2821 CCCTGAAACATATTAAAATTTCCTAGTGTATTAGTAAAAAGGGTTGT---   2867

EMBOSS_001      3356 ACTTTCCTTACTAAAATCAAGAACTTTACCGT-CGTCCTTGTACTTCAGG   3404
                       ||||.|.||||||.|||||||...||...| |.|.||...||...|.|
EMBOSS_002      2868 --TTTCTTAACTAAAGTCAAGAATAATATTTTGCTTGCTCACACACTATG   2915

EMBOSS_001      3405 ATAAGGGGGTGTTTCTTATAAATATTGTTATTTCTGATATGCTAACTGGA   3454
                     .|.|.|...|.|||||||.|      ||||.||.||||||..||||..||||
EMBOSS_002      2916 GTGAAGACATCTTTCTTACA------GTTACTTGTGATAGTCTAGATGGA   2959

EMBOSS_001      3455 ATTTTTAAGCAAATGTATTTTTATAGAACGCCATACAAAGCCTTTAGGGG   3504
                     .|.|||||||.|......|||||||.||....|||.|.|.|||. ||..|
EMBOSS_002      2960 CTATTTAAGCCACCAAGATTTTATATAATATAATATAGATCCTG-AGCAG   3008

EMBOSS_001      3505 TGAAAGTTTCAGGATTTTTAAATTGCAGATTTATCCTTTAAATAAAAA-A   3553
                     |||...|||.||.||..|||| |||||.||||||  ||.|||.|.||| |
EMBOSS_002      3009 TGAGTATTTAAGAATGCTTAA-TTGCACATTTAT--TTCAAAAATAAAGA   3055

EMBOSS_001      3554 ACTATATTCGTAATTGAA---------------TCGGATTATTTCTCTAT   3588
                     |||..|||..||||||||.               ||.|||||.|||.||.||
EMBOSS_002      3056 ACTGAATTAATAATTGATCTTGTCAAGATCTAGTCAGATTCTTTTTCCAT   3105

EMBOSS_001      3589 CCAAAACATTTTCTGCTTTGGGCCTAAGAAGAGTTGACAAAGCTGTTCAT   3638
                     |.||| ||||||||  |||.|.|||.|..||||||.|.|||..|..|.||
EMBOSS_002      3106 CTAAA-CATTTTCT--TTTAGACCTTAAGAGAGTTAATAAAAATAATAAT   3152

EMBOSS_001      3639 GGTTCA------AAGTACTACCATAAAACCCTGGGT------AACTAACT   3676
                     ||||||      ||.||.....|||||....|...|      ||||.|.|
EMBOSS_002      3153 GGTTCATCTTGTAAATAAATAAATAAATAAATATTTTTTAAAAACTTAGT   3202

EMBOSS_001      3677 GAAA--ATGGAAAGACTCTGT--CTTTCTGAATATTTCACAAGAGTTTCA   3722
                     .|||  |..||||||||.|||  ||||...|||||..|||||  ||||..
EMBOSS_002      3203 TAAATAAATGAAAGACTTTGTAACTTTGCAAATATACCACAA--GTTTTG   3250

EMBOSS_001      3723 CAAATATTAAGTGGTTCTCTAAGTACCCCTGAGAGATC-ATTGTAATATT   3771
                     .||.|.|||.|||.||||.|  ||.....|.||.|.|. ||||....|..
EMBOSS_002      3251 TAATTCTTATGTGATTCTTT--GTCTTAGTTAGGGTTTTATTGCTGCAAA   3298

EMBOSS_001      3772 AGCTTGTAAAGACAATGTGGGGGTGTGG---GTATGTGGTGACCTTTATG   3818
                     ...|....|||||.|.| |....|..||   |.||.|...||..||.|..
EMBOSS_002      3299 CAGTCACGAAGACCAAG-GCTTATAAGGACAGCATTTAAAGAGGTTCAGT   3347

EMBOSS_001      3819 ATGTT--CATAAAGGTGGTGTAAT-TAACATATTTTTCTCAG-CAAG---   3861
                     .|.||  |||.||||  .||.||| |..||...|.....||| ||.|
EMBOSS_002      3348 CTATTATCATCAAGG--CTGGAATATGGCAGGATCCAGGCAGGCATGGTG   3395
```

```
EMBOSS_001      3862 -ACAAACTAAGGAG------CAATAAATATATGAGATACCTTCATCTGTG   3904
                      ||||...|..|||      |.||....||.|||....||...|...|..
EMBOSS_002      3396 CACAAGGAACTGAGAGTTTTCCATTCTCATTTGACTGGCCACTAAGAGAA   3445

EMBOSS_001      3905 ATCTGGGTCATGTCTCAGGC--CATATCTTTCAAATCACTCCCTTCCCTA   3952
                     ..||...||.| ||||||.|  ||.||||||.||..|...|||..|.||.
EMBOSS_002      3446 GACTAATTCTT-TCTCAGTCAGCAGATCTTTAAAGCCCACCCCCACACTG   3494

EMBOSS_001      3953 AT-CTCGTGTTTTACCTACGTCTCCTCTC--AATCCCCCCATTATAAA--   3997
                     |. |.|.|.||..|.|.|.|.|.|..||.  |||....|||||.....
EMBOSS_002      3495 ACACACATCTTCCAACAAGGCCACACCTTCAAATAGTGCCATTCCTTGGG   3544

EMBOSS_001      3998 --AATTGTCTTCTGATGAATAAAACATTTCCAGAGAGACA-AGTTTCATA   4044
                       ||...|.||...|..|..||||||....||.|.|.||| |....||.|
EMBOSS_002      3545 CCAAGCATTTTAAAACCACCAAAACACACACACACACACACACACACACA   3594

EMBOSS_001      4045 AAGTTTGAATTGTACATCTGAGTACACCTATGAATTAAGATATCTTTGAT   4094
                     .|.....|.....|||.......||||..|...|...|.|||......|.
EMBOSS_002      3595 CACACACACACACACACACACTCACACACACACACACAAATACAGAGTAG   3644

EMBOSS_001      4095 TTCTAATATGTTATTAAAATTGGG-------TGTGGTGGCTCACGCCTGT   4137
                     |..|||.|.....||.|||.||||       ||||.|...||....||.|
EMBOSS_002      3645 TCGTAAAAGAAGTTTGAAAGTGGGAGATAAATGTGATACATCTTTTCTTT   3694

EMBOSS_001      4138 AATCCCAGCACTTTGGGAGG--CAGAGGCGGGCGGATCACGAGGTCAAGA   4185
                     ....|...|||||.|.|...  |..|||...||..| |||.|..||..|.
EMBOSS_002      3695 GGATCTGCCACTTAGAGCTTTTCCTAGGAAAGCATA-CACTATTTCTTGG   3743

EMBOSS_001      4186 GATCGAGACCATCC---TGGCCACAAGGTGAAACCCCCGTCTCTACTAAA   4232
                     .||...|..|||.|   |....|.||.||...||.......|..|.|.|.
EMBOSS_002      3744 AATATTGCACATGCAAATTTTAATAATGTACCACTTAAAATTTAAGTGAT   3793

EMBOSS_001      4233 AATACAAAAATTA--GCCGGGTGTGGTGGAGTACGCCTGTAGTCCCAGCT   4280
                     |..|...|||.|.  |...||.|....|.|.||.....|.||||.. ||
EMBOSS_002      3794 ATAATTGAAAGTGATGATTGGAGATTAGAAATAAAAGGATTGTCCTT-CT   3842

EMBOSS_001      4281 ACTCAGGAGGCTGAGGCAGGAGAATT---GCTTGAACCCAGGAGGTGG--   4325
                     ||...|.|.|.|..||||...|..||   |.|.|.|||....|..|..
EMBOSS_002      3843 ACAGTGAAAGTTTTGGCATCTGCCTTAAGGGTAGCACCATTTATTTTCTA   3892

EMBOSS_001      4326 ---------AGGTTGCAGTGAGCCGAGATGGCGCCACTGCACTCCAGCCT   4366
                              |.||.|||...|.....|||.|.|..|||.||...|.|.|
EMBOSS_002      3893 AATAAAAATATGTAGCAACCATTGTTGATAGGGTTACTCCATATCTGGC-   3941

EMBOSS_001      4367 GGTGAAAGAGCAGACTCTGTCTCAAAAAAATAAATTAAAATAAAATAAAA   4416
                         ||||.. |.|.| ||||.||.|..||....||.|..|..|.|| ||.|
EMBOSS_002      3942 ----AAAGCT-ACAAT-TGTCACACATTAACCTTTTCACTTTGAA-AATA   3984

EMBOSS_001      4417 TAAAATTGG-AGAAGTTTCTCACCAAAATTT--TGGCGCACGGATTAATT   4463
                     |||.|||.. |..||.....||.|||.||.|  |..|..| ||.||.|.|
EMBOSS_002      3985 TAACATTCCCATTAGCAGTACAACAAGATATGATACCTGA-GGTTTCACT   4033

EMBOSS_001      4464 CTGAAGAAAGAAGAAAGAATGCAATCTTAGTAGCACAATTAG---TACCT   4510
                     ..||||....|| |.|||....|..|..|||....|||||..|   ||...
EMBOSS_002      4034 TGGAAGTTCCAA-ATAGACGTTATACAAAGTTAACCAATGGGGTGTAGTG   4082

EMBOSS_001      4511 TGAATAAATTGGAGTATCG-TATTTCTTGGACTATCTGAGAATGCAGAGG   4559
                     |..|.|...||.|.|.|.| |||...|..||...|...|...|||...|.
EMBOSS_002      4083 TCCAAAGTGTGCACTGTGGCTATAAATATGATGCTGAAACTTTGCTCTGA   4132

EMBOSS_001      4560 CAATTTAAGGATCCCTAATTCTA-AGGAGAAGAAACCTTTAGTGTATTCC   4608
                     .||||.||.|.......|||.|| |..|.|..|.|...|||||.|.|||||.
EMBOSS_002      4133 AAATTAAATGTAAGACCATTATATACAACATCACAAAATTAGAGCATTCA   4182

EMBOSS_001      4609 TTCCTGTTGCTTTAGTTTGAATTGAGTTTTATATGTATTTTTTAATCTTT   4658
                     |.   |.||.|.|.|||...|||..||...|| |..||.||..||.|.|.
EMBOSS_002      4183 TA---GATGATATGGTTAATATTCTGTGGCAT-TAAATATTAGAAACATA   4228

EMBOSS_001      4659 CTATTTTGATTGTTGTCTAA---AGAGTGTG---AAAGTGAATTTTGA-T   4701
                     ..|.|.|.|.|...|..||.   |||||.||   |.|...||...||. |
EMBOSS_002      4229 ACAGTCTCACTTAAGGATATGGTAGAGTTTGGTCAGATAAAAAGGTGTCT   4278
```

```
EMBOSS_001      4702 ATTTTTATTTTGCCTGGCGATGAATGCCTTCTG------CTCTGGATATT   4745
                     ||.||.|..||..||.|.|....|...|||.||      ||.|...||||
EMBOSS_002      4279 ATCTTAAGGTTCTCTAGAGTCACAGAACTTATGGAATATCTTTATTTATT   4328

EMBOSS_001      4746 TAAAAATTATATACACATATATGTGTGTGTGT-GTGTGTGTGTGTG----   4790
                     .|.|||.|.|||....||...|....|..||| ||.....|...|.
EMBOSS_002      4329 AAGAAAATTTATTGTAATGACTTATAGACTGTAGTCCAACTTACTCAACC   4378

EMBOSS_001      4791 -TGTGTATATATATATATATATATATAT-ATATA-TATAAAATTTTTCTG   4837
                      ||.|.|..|||..||....|.....|| ||.|| ||.....|...||..
EMBOSS_002      4379 ATGGGCAGCTATGAATGGGAAGTCCAATGATCTAGTAGCTGCTCAGTCCC   4428

EMBOSS_001      4838 AGAACTTTTATTAATTCAGCGTATCTTTGCT--AAACACCTGCCATGTGT   4885
                     |..|...|..||..||..||.|.|||||...|  ||..|..|.|||...|.
EMBOSS_002      4429 ACGAAGCTAGTTGTTTATGCTTGTCTTCTGTGGAAGTAGGTTCCAACAGA   4478

EMBOSS_001      4886 CGTGGTGTTAGGTCTGGTGATACAAACATGTTCAGAGAGATGATTTTCTT   4935
                     .||..||..|.||..| ||..|.||........||||.||  ||.|||.|
EMBOSS_002      4479 TGTTCTGGCAAGTAAG-TGCAAGAAGTCAAAGAAGAGTGA--ATCTTCCT   4525

EMBOSS_001      4936 TCTT---------TTTTGGGGG---------GTGGGTAAGG----GAAAG   4963
                     ||||         ||.||..||         |..|.||.||    ||..|
EMBOSS_002      4526 TCTTCTAATGTCCTTATGTAGGCCTCCAGCAGAAGATATGGCCCAGATTG   4575

EMBOSS_001      4964 AAGGCTTATACAACAGAATCT-TATTTCTCACA-GTTCTGG---AGGCTG   5008
                     ||||..|.|.|.||...|.|| |||.|...||. |.|.||.   ||||||
EMBOSS_002      4576 AAGGTGTGTGCCACCACACCTGTATCTGGAACTTGCTTTGTCCCAGGCTG   4625

EMBOSS_001      5009 GGATTCCAAGATCAGGGC-CTGGTGAGGGCCCCTCTTCCTGGT-TTGCAG   5056
                     ...||..|.  ||||.|. |||.|  .|.|..|..|.|||||. ||..||
EMBOSS_002      4626 TCTTTGAAC--TCAGAGATCTGCT--TGCCTGCGTTGCCTGGAATTAGAG   4671

EMBOSS_001      5057 ATGGCTTCCTTCTC-TCTGTGTCCTAACATAGCAAAGAGAGACAGAGCTC   5105
                     |....|.|..|||. ||.|.| |||||..|....|.| |..||...|| |
EMBOSS_002      4672 ACATGTACTATCTTGTCAGAG-CCTAAGCTTTTCATG-GCCACTATGC-C   4718

EMBOSS_001      5106 TGATGACACTTCC---TCTTGTT----ATAAGGGAACTAA-TTCCA-TCA   5146
                     |.|.||.  ||||   ||..|.|    ||.||..|.||.. ||||| ||.
EMBOSS_002      4719 TCAAGAT--TTCCATGTCAAGATCCAGATCAGAAACCTTTCTTCCAGTCT   4766

EMBOSS_001      5147 TAAGGGCCCCAAGAAAG--GTGCTTTTCA-------AAAACAGTTCAGT-   5186
                     .|||..|.|.|..|.||  ||||..|.||       ||...||||||.|
EMBOSS_002      4767 CAAGATCTCGATCACAGATGTGCCCTCCATTTCTGGAATGTAGTTCATTC   4816

EMBOSS_001      5187 -AAAAGTACTGGGTTGTATAATCACTTTAATGAGTATCAATCCATATTTT   5235
                      |.|.|||.|....|..|.||.||.....|..|.|..||.||.||.||||
EMBOSS_002      4817 CAGATGTAATCAAGTTGACAACCAGGAATAGCAATCACATTCTATTTTTT   4866

EMBOSS_001      5236 TAAGATAGAAATGAATGAAATTAG---TAAAATAGAATAGAAATAAGGAG   5282
                     |||....|.|.|.|..|.|||.   |..|||...||...||.|....|
EMBOSS_002      4867 TAATTAGGTATTTATTTCATTTACATTTCCAATGCTATCCCAAAAGTTCG   4916

EMBOSS_001      5283 TCCAT-CACTTTTA-AGTAAGTT-TCAATATTG---TTCGTAAAACTTTG   5326
                     .|||. |.||...| |...|.|. ||.|.....   |||..|...||| |
EMBOSS_002      4917 CCCACACCCTCCCACACCCACTCCTCCACCCACCCATTCCCACCTCTT-G   4965

EMBOSS_001      5327 GTTCGGTGGTTTGTGTGTG-TGTG---TATTTGTGTGTG---------TG   5363
                     |..|.|...||....|||. ||.|   |||..| ||.||         ||
EMBOSS_002      4966 GCCCTGGCATTCCCCTGTACTGAGGCATATAAG-GTCTGCACAACCAATG   5014

EMBOSS_001      5364 TGTGTGTGTGTCTGTCGGTGTGGAAATACTGGATCACTTTGTAACATATA   5413
                     .|..|.|.|.||....|.|| |..||||......||..|.|..||.|...
EMBOSS_002      5015 GGCCTCTCTTTCCACTGATG-GCCAATATATATGCAGCTAGAGACGTGAG   5063

EMBOSS_001      5414 TTCAAA---------AGCCTCTGTATT-TTAACATTATT--TCTGCCTTT   5451
                     .||...         ||...|||..|. ||.|.|||.||  |.|.|||.|
EMBOSS_002      5064 CTCTGGGGGGGTGGGAGGTACTGGTTAGTTCATATTGTTGTTTTACCTAT   5113

EMBOSS_001      5452 GAGAGGTTCACATTCCAG-AGGTGAAGACATACATCCTA-AGA--CAAAA   5497
                      ||.|.|.||.||.||.. ||.|.......|||.|||||. ||.  |...|
EMBOSS_002      5114 -AGGGTTGCAGATCCCCCCAGCTCCCTGAGTACTTCCTCCAGCTCCTCCA   5162

EMBOSS_001      5498 TTAT---------AATAGCATTATGAGA--ATTACAGTAGAGAGCTGGAC   5536
```

```
                     ||..         |.||..|||.|||.|  |||..|||||....|||...
EMBOSS_002      5163 TTGGGGGCCCTGAAGTAAAATTGTGATATCATTGTAGTAGTATCCTGTGA   5212

EMBOSS_001      5537 AGG----GTCTAGCAAAAA-CAGAAGACTAGGC-TAAACCTTCCAAAGAG   5580
                     ||.    |.||.||.|.|. ||...|..|.||. |.|..||...||..|.
EMBOSS_002      5213 AGAAGCTGCCTGGCTACAGGCATGTGGTTGGGAATTAGTCTAAGAACTAA   5262

EMBOSS_001      5581 GCCAGGAAACTCAC-CTAGAACGGTGGATTTTAACCTTGCTTATGCACTG   5629
                     |..|||||.|||.| ||    |..|...|.|....||..||....|.||.
EMBOSS_002      5263 GTAAGGAATCTCTCTCT----CTCTCTCTCTCTCTCTCTCTCTCTCTCTC   5308

EMBOSS_001      5630 GGGGAGATTTTAAAAATATCTCTGCCCACAATAGATACCAACTGAATTGA   5679
                     ........|||.....|.||||| |.|.|..|...|.||..||....|..
EMBOSS_002      5309 TCTCTCTCTTTCTCTCTCTCTCT-CTCTCTCTTCCTCCCCTCTTCCCTTT   5357

EMBOSS_001      5680 GCATAGCATGTCCTACCCAT---GAATCTAT-TGTCCAGTGAGAACCTCT   5725
                     ...|..|.|..|...|||.|   |..|.||| |||   ||..|||||.|..
EMBOSS_002      5358 ATCTCCCCTCCCTCCCCCCTCTGGTGTGTATGTGT---GTATGAACATTA   5404

EMBOSS_001      5726 GTTTAGAGAA---AGTCACCTTAGAAGAATTGTTAGGAGTT-ATTTAGGT   5771
                     |||.....||   |.|.||||..  |.|||||.|.|..|.|| |||||...
EMBOSS_002      5405 GTTACTTCAATTTATTTACCTAT--ATAATTTTCAACATTTCATTTACCA   5452

EMBOSS_001      5772 TCATGGGGTTGAAAAGAGCATTCGTGATAGAGGAAACACCATAT--CCAA   5819
                     ..||..   |||||||    ||| |.|||.|..|||.|..||.|  |||.
EMBOSS_002      5453 ATATCA---TGAAAAG----TTC-TCATATATTAAATAAAATGTTGCCAG   5494

EMBOSS_001      5820 AGGCTTAGTCAGTGTGGTAGTGTGAGAAT-------CTGAAGGAAC----   5858
                     ||.   |..|||||.|||||...|..||||       ||||||...||
EMBOSS_002      5495 AGA---AACCAGTTTGGTAAACTTTGAATTATACATCTGAATATACCTAT   5541

EMBOSS_001      5859 ---TTGGCTGGGGTATGGTTGCTACAAGAA-ATGAAATTAGATCAACTGG   5904
                        ||.|....|.|.||..|.||...|.|. |||.|..|.||..||.|.|
EMBOSS_002      5542 AAATTAGGACAGTTTTGACTTCTGATACATTATGGAGATTGAAAAATTAG   5591

EMBOSS_001      5905 GG-CTAAATTATGTGGAA-AGACAGCAT-GA------TGTAGCAGCTAG-   5944
                     .. ||..|.|||.|..|| |.|.|..|| ||      |||.....|.||
EMBOSS_002      5592 ACTCTTCAGTATTTCTAATACAAAATATTGACCCATGTGTTCATTCAAGG   5641

EMBOSS_001      5945 -AGTAT-----GGACCTTGTAAGCAGGAAGACCCCTTATTTA---GCACT   5985
                      |||||     |.||||||..|...|...|.|||.|..|||||.   |.|..
EMBOSS_002      5642 TAGTATAATCAGCACCTTAAATAAAATGAAACCTCACATTTCATGGGATA   5691

EMBOSS_001      5986 TACTAGCTTA----TTGTCTG---ACCTCTGAGTCCCAATTTTACTCTTC   6028
                     |..||...||    |.|||||   |.||||.|.||. ||....|| ||||
EMBOSS_002      5692 TTGTACAATAGAAATGGTCTGTGGATCTCTAATTCT-AAGGAAAC-CTTC   5739

EMBOSS_001      6029 TATACAATGAGTACATCACA---GGATTTTATCAGGTTTAAATGATAAGA   6075
                     .|||.|.|.|.||. ||.||   ||.|||.||....|.||..|..|.|..
EMBOSS_002      5740 AATATATTTACTAT-TCCCAATTGGTTTTCATTGTCTGTATTTTTTTATT   5788

EMBOSS_001      6076 TA--TATGTAAAATGCATACCAGAGAGGCAGACTATTGGACTCGAAGGGC   6123
                     |.  |||.|..|.||. |.|..|||||....|...|....|||.|
EMBOSS_002      5789 TGCCTATTTTGATTGT-TTCTTGAGAGTATCAGGGTGATTCTCCA-----   5832

EMBOSS_001      6124 TCAGTAAGTGTAAGCTGGCTCTCTCTGCCCCTTGCCACCTATTTTTCAG-   6172
                     |...||..|.|...||.|...|||||..|..|| |.||.|.|.|...|.
EMBOSS_002      5833 TATATACCTTTTGTCTTGTGTTCTCTTACTGTT-CTACATGTGTAAAAAT   5881

EMBOSS_001      6173 ACTCTGGACTTTTATCACTTTAAGTCATAG-CCTAGT-TCTAAGCA---A   6217
                     |.|.|...|.|.|.|...|..||||.|||. ||||.. |.||||||.   |
EMBOSS_002      5882 ATTATCATCATATTTACATGCAAGTTATAATCCTACAATTTAAGCTTTCA   5931

EMBOSS_001      6218 GGAAATGGACTAATCAG-AC-ATGTTTTTAAAAGATCATTCTGGTAGTGG   6265
                     ..||||.|...|.||.. || |....|.||..||.|..||..|||| |.|
EMBOSS_002      5932 TTAAATTGGAAAGTCTTTACTAAACATCTACCAGGTGCTTTAGGTA-TAG   5980

EMBOSS_001      6266 TTAGGAGAATGAAT-TGGAAAGATATGAGACCCATG---CAGGGACAACA   6311
                     |.|.||.|||..|. |..|.|||..|| ||.|..||   |||.||..|||
EMBOSS_002      5981 TGATGACAATATACATACAGAGAGGTG-GAACTTTGTGTCAGAGAACACA   6029

EMBOSS_001      6312 GTTAGGACATTATTTCTGTAATAA-GCC--AAGCAAGAATTGATGATCAA   6358
                     ...||||||| |..|.|||.|||. |||  ||||.|||...|..|||||.|
```

```
EMBOSS_002      6030 CCAAGGACAT-AGCTATGTCATAGTGCCCAAAGCTAGACAGGTAGATCTA   6078

EMBOSS_001      6359 -AGTGGT--GAGGTTGAACAAA----CAA-AACAGAT----ACGTG----   6392
                      |.|.||  ||.|...|..|||    ||| ||||.||    ||.|.
EMBOSS_002      6079 TACTTGTCTGAAGGCCAGAAAATATTCAAGAACATATTTTAACCTTGCTT   6128

EMBOSS_001      6393 AGCTATTTGGAGATAAAATCAACACTGTCATATGTTTTGTGGGAGGTGGA   6442
                     |..||||.||..|.|...|..|.| ||.||.|..||..||.|||..|.||
EMBOSS_002      6129 ATGTATTGGGTAAGATTTTTTATA-TGCCAAACCTTAAGTTGGACCTTGA   6177

EMBOSS_001      6443 GGTGAGCAGAAAATGTGAGGTAAAATGAGAAATCAGTGCC-TGCTTACCA   6491
                     .  ||.....||.|..||...|.|||...|.|||...||| ||.||...|
EMBOSS_002      6178 A--GAAATTCAAGTTAGATTGATAATACCATATCCTGGCCATGATTCTTA   6225

EMBOSS_001      6492 CTTGGCATGATTGACTGA-AGGTAGTGTCTTCACTCAATCATG----AG-   6535
                     ..|...||.|..|..||| |..|..|||.|..|...|.|.|||    ||
EMBOSS_002      6226 TCTACAATCAAAGGATGAGAAATGCTGTTTAGAGAAAGTTATGTTTAAGA   6275

EMBOSS_001      6536 ---TTGCAGAA--TTCAAGATGGCA---------AACAGTTG----TGAG   6567
                        ||.|||||  |.|.||.|||.|         |||...||    ||||
EMBOSS_002      6276 ATATTCCAGAAGGTCCCAGGTGGTACTTGTGGAGAACTCCTGCATTTGAG   6325

EMBOSS_001      6568 GAGCAAAGTCAAGAACGTGTTTGATTTTGAGGTATCTGTAAGTGAAAAAT   6617
                     ...|.|||.|| |||.|.   |.||.|||||||.....||...||.|.|.
EMBOSS_002      6326 AGACTAAGCCA-GAAAGA---TAATATTGAGGTCAACCTAGTCGACATAG   6371

EMBOSS_001      6618 CA-GAG---GTGAAAACCTT---ACCTCTCTTGAAGCAG----TTGTGA-   6655
                     || ||.   ||.|.||..||   |......|.|.||..|    ||.|||
EMBOSS_002      6372 CAAGATCGTGTCACAATTTTGAAAAGGGCATAGTAGTGGCTTATTTTGAG   6421

EMBOSS_001      6656 ------ATGTAAATC-TA-AGGTTTGGAAAAAGATCTGGGTTAAAGATTT   6697
                           |.|.||.|| || |.||.|.||.|||||.|...|.|..|||||..|
EMBOSS_002      6422 ATCCTGAAGAAACTCCTATATGTGTAGATAAAGTTTCTGCTACAAGAAAT   6471

EMBOSS_001      6698 AAAATTGAAGGACATCA---ACATGGAAGCCATAGAAATAA-ATTATATT   6743
                     .|..|||. |.|.||||   |...|.|||.|||..|||.|. |..|||..
EMBOSS_002      6472 GAGGTTGG-GTAGATCAGATAAGAGTAAGTCATGTAAACACTAGGATAAA   6520

EMBOSS_001      6744 ACACACAAAT---TTATGTCGTTATTTGA-AT-----TTCTCCATGGTCC   6784
                     |.|.|.||.|   ||..||.||.|...|| ||     ||||.|||.  ||
EMBOSS_002      6521 ATAGAGAATTGGTTTTAGTGGTAAAGAGAGATCCCCATTCTGCATA--CC   6568

EMBOSS_001      6785 ACTCAGAAATATATCTAAAT-GTCACCAAAATGTTACTTACTGTAGTACA   6833
                     |.||..||...|.|||..|| |||.|.|.....||.||...||.|.|| |
EMBOSS_002      6569 ATTCTTAACCCTTTCTCCATTGTCCCAAGCCCCTTTCTGTTTGCATTA-A   6617

EMBOSS_001      6834 GAATTGGTATTAAGTGATACT----ATTGTCCATGTTATTCAAAAAGACA   6879
                     .|.||..|......|..||||    |....|||.||.......|.|.|||
EMBOSS_002      6618 CACTTCCTTCCTTATTTTACTGCCAAACCACCAAGTCTCAGTGACACACA   6667

EMBOSS_001      6880 GTTATAGGGACCCTCTTAATAAACTAATTGTGAAAAAGGCAAAGAATTAG   6929
                     ..||.||||||..|..||.|.|||..|.|.....|.|...||||||.||
EMBOSS_002      6668 TATACAGGGACTATAATACTTAACAGAATACAGTCAGGATTAAGAATAAG   6717

EMBOSS_001      6930 CAA-AGCTT---TGGCATAAAATTCATATCA-TGGGCCA----GGC---G   6967
                     ..| |..||   |..||..|||.....|.|| ||.||.|    |||   |
EMBOSS_002      6718 ATACATGTTACTTTCCACCAAAGAGGCAACACTGTGCTAAATAGGCTCAG   6767

EMBOSS_001      6968 TGG--------TGGCT-----------CATGCATATAATCCCAGCACTTT   6998
                     ||.        |||||           ||..||.|.|..|||..|||...
EMBOSS_002      6768 TGAATAAAAACTGGCTTTTTCTCCACACACACACACACACCCCACACACA   6817

EMBOSS_001      6999 GGGAGGCTG--------AGGTGGGCAGATCACCTGAGGTCG----GGAGT   7036
                     ..||..|||        |.||..| ||.|||.|||..|||.    |..||
EMBOSS_002      6818 CTGACACTGTTTTTTAAAAGTTTG-AGTTCATCTGTAGTCTTTCAGCTGT   6866

EMBOSS_001      7037 TCGAGACCAGCCTGACCAACATGGCGAAACCCCGTCTCTACTAAAAAT--   7084
                     ...|....||..|.....|.||...|.|||||..|....||.|.|.|.|
EMBOSS_002      6867 CATATCTTAGTATTCAGCAAATATTGCAACCAAGAAATTAATTATAGTTT   6916

EMBOSS_001      7085 --ACAAAAATTAGCCAGGTG-TG-GTGGCAC---ACGCCTGTAATCCCAA   7127
                       |..|||||.|...|||.| || |||.||.   ||.|..|.|.||...|
EMBOSS_002      6917 TTAATAAAATGATTAAGGCGGTGAGTGACATTGGACACAGGAACTCTATA   6966
```

```
EMBOSS_001      7128 CTAC--TCGGGAGGCAGAGGCAGGAGAATC--GCTTGAAC---GTAGGAG   7170
                     ..|.  |.|...||.||...||.|.|.||.  ||| |.||   |||.|||
EMBOSS_002      6967 AGAAAGTTGTTTGGAAGTTACAAGGGTATTTTGCT-GCACTCTGTAAGAG   7015

EMBOSS_001      7171 GCAGAG--GATGCAGTGAGCTGAGATCGTGCCATTGCACTC--CAGCCTG   7216
                     ||||.|  ||.|...|.|.|||||.|..|   ||.|.|.||  ||.|.|.
EMBOSS_002      7016 GCAGTGATGAGGAGATAATCTGAGGTAAT---ATGGTAGTCAGCATCTTC   7062

EMBOSS_001      7217 GGTGACACAGTGAGACTCCATCTCAAAAAAAAAAAAAAAAAAATTATGTCA   7266
                     ..||||.||.|||.|.|..... |.|||.|.........|...|||.||
EMBOSS_002      7063 ATTGACTCAATGAAATTTTGAGG-ACAAACAGTTGTGGGGAGGGATGCCA   7111

EMBOSS_001      7267 TGGAAAAAGTAAAAGTCTTTGCATAA---TGTATCCAAGATCATGAAA--   7311
                        |.|..||....||.||....||.   ||.||...||||.|..|||
EMBOSS_002      7112 ---AGACTGTGGTTGTTTTGAGGTACCCGTGCATAATAGATGAGAAAATC   7158

EMBOSS_001      7312 AACTCTTTTCAATAAGATAATTAGTTCCTTTTC---TTAT--ATAAACAT   7356
                     |.|||..||.||..||.|. |.|.|.|.|.||.   ||||  ||.|||.|
EMBOSS_002      7159 ACCTCCGTTGAAGCAGTTG-TGAATGCATATTTAAATTATGGATTAACTT   7207

EMBOSS_001      7357 GGAAATTTTCATTTT-----TCCTT-TTATTCTCATATTGATACTATAAA   7400
                     |||||...|.||...     ||||. |||....||.||. |||.||.|.|
EMBOSS_002      7208 GGAAAAAATTATGGAGAAAATCCTAGTTAAAGGCAGATC-ATAATAAATA   7256

EMBOSS_001      7401 AACCCCATCCTCATTCACAATACTACTGTCTCTACCCTCG-ATAGATACC   7449
                     |.....|.....||||...|.| ||||..||.|..||..| |||| |.|.
EMBOSS_002      7257 ATGATAACTAATATTCTTTAAA-TACTCACTATGACCAAGCATAG-TGCT   7304

EMBOSS_001      7450 AGTTCAATTGAACGTA-GCATGTTCTACCCATGAATCTATTGTTCAGTGA   7498
                     |..|..||||||..||| |.||.||.   |||| |||||..|.|||||.
EMBOSS_002      7305 ATATGTATTGATTGTACGAATATTA---CCAT-AATCTTATTTTCAT---   7347

EMBOSS_001      7499 GAACCTCTGACTAT----AATGCTCAGGAAT-----ACTCAAGACTCACA   7539
                     |||...||.|.|||    |..|.|.|..||.     |||..|.|...|.|
EMBOSS_002      7348 GAAATGCTAAATATTTGGAGGGTTGACCAACTGCAAACTATATAGGGAAA   7397

EMBOSS_001      7540 TGATTGT--CTTCTTGCTATATTTAGTTACTTTATTATTTTC---CATTT   7584
                     ..|.||.  ||.|||.| |.|..||.|.|..|||||||||...   |.|..
EMBOSS_002      7398 CCAATGATACTACTTAC-ACACATATTGATATTATTATTGAAACGCCTCC   7446

EMBOSS_001      7585 TGG--GACCCTGAATTCCTGTAGATCTCAGAGAAAATC------CGAAAT   7626
                     |||  ||.||.|||.||||..|..|...|....||||||      |....|
EMBOSS_002      7447 TGGTTGAACCAGAAATCCACTTAAATGCTACTAAAATCTTCCTGCATGTT   7496

EMBOSS_001      7627 GAAAT--AATGAAAA-TAATTAAAAGTT-TAGAAAAGGGAGTCAATGGGG   7672
                     ||.||  |.|||... |...||.|.|.| |...|||.|.|.|||.|....
EMBOSS_002      7497 GATATGCACTGATCCCTTCCTATATGCTATTTGAAATGAATTCAGTATAA   7546

EMBOSS_001      7673 ACAAATGTT--CAGGACTGGTCTTTTATCTC-----CTGC-AGGAAGAAA   7714
                     |....|.||  |||.....|.||.||||...     |||| |.|||.|.|
EMBOSS_002      7547 AACGTTCTTTCCAGCTTCTGGCTATTATGAAGAAGGCTGCTATGAACATA   7596

EMBOSS_001      7715 GACTGAATGCAGAAAAT-TAGAATC------CATTTTTCATCCAG-TCAC   7756
                     |....|||...|....| |||..|.      |||.|||.....|. |.||
EMBOSS_002      7597 GTGGAAATTGTGCCCCTGTAGCCTGGTGGGGCATATTTTGGGTATATGAC   7646

EMBOSS_001      7757 CCCAATT--TA-ATGCAA-TATGAGTTTAGCTA-TTTGATTTTAAGTGTT   7801
                     |...|.|  || ||.||| |||..||..|.||| ||||||||.|..||.|
EMBOSS_002      7647 CAAGAGTGGTATATACAAGTATTTGTGAAACTACTTTGATTTAATATGAT   7696

EMBOSS_001      7802 GTACCGTTTTGGACCATGTTACCATGGT------AACATGAACC----AT   7841
                     .|....||...|||....|||..|.|||      |...||.||.    ||
EMBOSS_002      7697 TTGAATTTCAAGACATCATTATAAAGGTTGGGCTACTTTGGACATATTAT   7746

EMBOSS_001      7842 GTCTCATTCATACGTAAACATGTTAATTG----TATTAAAACCTTTAAAA   7887
                     .|||..||||||...||||||||.|.|.|    ||||.|||..||.|.||
EMBOSS_002      7747 ATCTTGTTCATATTGAAACATGTGAGTAGAAAGTATTGAAAATTTCAGAA   7796

EMBOSS_001      7888 CCTACTT--CTGGATGTTGCCATTACATTAAACAATTATCTAGAATGATA   7935
                     ||.||||  |||.||.||||.||..||.||...|||.|...|.||||.||
EMBOSS_002      7797 CCAACTTGTCTGAATATTGCTATGTCACTAGGAAATAACACAAAATGGTA   7846
```

```
EMBOSS_001      7936 CAAAGTAATGACTAAATTGAATAACTTTGTAAATTAACTATTGGATTTTG   7985
                     ||.|..||..|||||.|||||||||.||...|..|||||||...|||||.|.
EMBOSS_002      7847 CACATGAAATACTACATTGAATAATTTACAACTTTAACTACCAGATTATA   7896

EMBOSS_001      7986 TAATTTTATATCTATAAACCAAAAGAAAAGCCCACATTGGTAAGAAGACA   8035
                     || |||||||||||.|||..||||||||||||||..|.|..|||||   ||
EMBOSS_002      7897 TA-TTTTATATCTACAAATAAAAAGAAAAGCCCCTAATCATAAGA---CA   7942

EMBOSS_001      8036 CTGTGCATACTGAAAAGTCAATTTTGTTAGCCTCCAATAACCATTGTGTT   8085
                     ||||||||||||.|||||.||||||||.|||||.|..|||.|.|||||||.||
EMBOSS_002      7943 CTGTGCATACTGCAAAGGCAATTTTGATAGCTTATAATGATCATTGTCTT   7992

EMBOSS_001      8086 TTATTCCTCGCAGAGCTTTTGTGAGGATCTTATAAGGGAATAAATATGAA   8135
                     |||||.|.||.|||||.||||||||..|.|||||.|||||.|||||||||.||
EMBOSS_002      7993 TTATACTTCACAGAGATTTTGTGATAAACTTAGAAGGTAATAAATATAAA   8042

EMBOSS_001      8136 AGCACTTTGAAAAAGCTTTC---AAGTGAAAGGTCCTTATTAATTTTATG   8182
                     ||||||||||||| |||||.   |||||||||||||||.|||||.|||||.
EMBOSS_002      8043 AGCACTTTGAAAA-GCTTTTTTTAAGTGAAAGGTCCTTGTTAATATTATT   8091

EMBOSS_001      8183 AATTACCATTAAACAAAAGTCAAACTGAAGATGTAAATCTAATAGGATGC   8232
                     ||||||.|||||.||||||||||||||||||.|.|||||||||.|.....|
EMBOSS_002      8092 AATTACTATTAAGCAAAAGTCAAACTGAAGCTATAAATCTAAAATTGCAC   8141

EMBOSS_001      8233 TCTTAAAAGTCAATGGATCAAAGTTATATTAATTAATAAAGAATAATAAC   8282
                     ||||.|.|..|||||.||||||||||.|||||.|||||       |||||
EMBOSS_002      8142 TCTTTACAACCAATGAATCAAAGTTACATTAAGTAATA-------ATAAC   8184

EMBOSS_001      8283 TAAATATTTTATGTTTCATAATTGGCAAAGTATCTTTACTGTCATTTTCT   8332
                     ||.|..|||||..|||.|||||...|.||||||..|.|||||.|||||||||
EMBOSS_002      8185 TACAGTTTTTACATTTTATAATCACCCAAGTAATTATACTTTCATTTTCT   8234

EMBOSS_001      8333 AATTTGATCCTTAGTGAAAACCTGTGATGTTGGTACTCCTATTATTTCCA   8382
                     |||.||.|  ||..|.|||||   .|||||.|| ||||.....||||||||
EMBOSS_002      8235 AATCTGGT--TTTCTAAAAAC---GGATGCTG-TACTAGGGCTATTTCCA   8278

EMBOSS_001      8383 TTTTCATTTGAGAAGAATAAAATTGGAGAGGTTAAGTAATTTATCTATTG   8432
                     ||||.|.||....||||||||||||||||...||.|.||||||||||||.|
EMBOSS_002      8279 TTTTTAATTACAGAGAATAAAATTGGAGAAAGTATGCAATTTATCTATGG   8328

EMBOSS_001      8433 CTACTTGTTAAAATAACTACTAAATTTTATTACTC----CCAG-TTAGGA   8477
                     .|||.|..|.|.|.|...|||||.||.|.||.|.|    |||| ||||||
EMBOSS_002      8329 TTACATCATTAGAAATTAACTAATTTGTGTTTCACTATTCCAGATTAGGA   8378

EMBOSS_001      8478 GGGCAATTATATAAACTAAAAGCTTGTCACAATAAATGTTTACTTTTCTG   8527
                     |||.||||||.|||||.|.|||...|.|||..|.|.|||||.|    |.|||
EMBOSS_002      8379 GGGAAATTATGTAAAATCAAATGCTATCAGGAAAGATGTGT----TCCTG   8424

EMBOSS_001      8528 GGATTAAAGTCATCATGTATTTTTCAATTATTAAGGGGGGTAATAA-TAA   8576
                     |||||.||...|.|||.|.||...|||.||..|       |||||| |||
EMBOSS_002      8425 GGATTTAAAATAGCATATTTTGCCCAACTAACA-------TAATAAATAA   8467

EMBOSS_001      8577 TAATAGCTACCTTT-TTAAAATAGTTACTATGTGCCAAGGTGTGTACTAA   8625
                     |.|||||||...|| ||.||||||.||||||||. |||||
EMBOSS_002      8468 TGATAGCTAAAATTCTTTAAATACTTACTATGA-CCAAG-----------   8505

EMBOSS_001      8626 GTGCTTTGCTTGCATGATGTAATACCATCGTATATTTAGTACAGAGGAAA   8675
                                 |||||.||..|||        |||||
EMBOSS_002      8506 ------------CATGGTGCTATA--------TATTT-------------   8522

EMBOSS_001      8676 AACTGAGAGGCTGGGTAACTTCTACTAAGGTAACACACAAGTACTGGTTG   8725
                        |||.    ||.||.|||..|||          |||
EMBOSS_002      8523 ---TGAT----TGTGTGACTATTAC----------CAC------------   8543

EMBOSS_001      8726 AGTATCCCTTATCCAAAACACTTGGGACCACAAGTGTTATGGATATCAAT   8775
                                |||              |||||      |||||..|||.|
EMBOSS_002      8544 ---------TAT--------------ACCAC------TATGACTATTA--   8562

EMBOSS_001      8776 TTTTTTCTGATTCTTTTTTTGGATTTCAGATTTTTTCAGATTTTGGATTA   8825
                                  |||||.||||
EMBOSS_002      8563 -----------TCTTATTTT------------------------------   8571

EMBOSS_001      8826 CTTGCTTTATAATTATGGGTTAAGCATCCCAAACCCCAAAATTCAAAATT   8875
```

```
                                          |||
EMBOSS_002      8572 ------------------------CAT-----------------------   8574

EMBOSS_001      8876 GGAAATACTCCAATGAGCATTTACTTTGAGAATCATGTCGGCGCTCAAAA   8925
                      ||||||||..|         ||.||.|||            |.||||..
EMBOSS_002      8575 -GAAATACTAAA---------TATTTGGAG------------GGTCAACC   8602

EMBOSS_001      8926 ATTTTCAGCTTTTAGAGTTTTTTGGATTTTGGATTTTCAGATTTGGGATG   8975
                     |..|.||
EMBOSS_002      8603 AACTGCA-------------------------------------------   8609

EMBOSS_001      8976 CTCAACCCGAATATATAGAAAAGTCAGCATTTGAACCTAAGTTTGACTTT   9025
                        |||.   |||||..||.||.|.||||||||||..|..||||||||||
EMBOSS_002      8610 ---AACT---ATATAGGGAGAAATGAGCATTTGAAATTCTGTTTGACTTT   8653

EMBOSS_001      9026 CTGATCTTCTACCAACTCTACTGTCCTACCCATTACTCTACATTGACTCA   9075
                     ||||.|||||...||..||..|.|||||.|..|||||.|||....||.||||
EMBOSS_002      8654 CTGACCTTCTGAGAATCCTGATATCCTGCTTATTATTCTGAGCTGGCTCA   8703

EMBOSS_001      9076 GCATTACAGGGAAAGACCCAAGATCACCAAAAGCAAGCTTCAAATCACTC   9125
                     |||.|.||.|..|| ||||.|.||..|.|||||..||..|||||||||.||
EMBOSS_002      8704 GCAATTCAAGCCAA-ACCCTAAATATCAAAAAAGAAATTTCAAATCATTC   8752

EMBOSS_001      9126 ATCTAATAGAAATTAGTG--------GAAATATTTCTACTTCCTAAACAT   9167
                     |..|.||.|.|||..||.        ||||.|||... ||.|||||..|||
EMBOSS_002      8753 ACTTCATTGGAATCTGTCAGTCCATAGAAACATTAGG-CTCCCTAGGCAT   8801

EMBOSS_001      9168 CCATCTTTCCTTTACATTTTAAAGTCAAGTTTCTACATCTGCCTCCCAAC   9217
                     .||| |||..|||||.|||||....||.|||||||.|||.|..||  ||.|
EMBOSS_002      8802 ACAT-TTTGTTTTATATTTTTCTTTCTAGTTTCTGCATGTTTCT--CAGC   8848

EMBOSS_001      9218 TGAAACACTTCTCTATGAAATCACCATAACTACCAAATGCAAATATTTTT   9267
                     |||||..|. ||||.|||||.||.. |.|..|.||.|||.|||||.||.||
EMBOSS_002      8849 TGAAATGCA-CTCTGTGAAACCAGG-TCATGATCAGATGGAAATGTTCTT   8896

EMBOSS_001      9268 ATCAAGTCCTCATTGCCCTAGAAA--TCTACTCATATTTTGTTATTACTG   9315
                     .|.|||.||||.||...|.||||.  |.|...|.|.||.||||||
EMBOSS_002      8897 GTTAAGGCCTCCTTTTGCCAGAATGGTGTGTGCTTGTTCTGTTAT-----   8941

EMBOSS_001      9316 CTCACTACAGCCTACTGAAAAATGTCTCACCTTTTGACTTGCCAGGGTGA   9365
                        |||..|.|  ||||||.|||.||.|.|||.|||||||.|..|..|..
EMBOSS_002      8942 ---ACTTGACC--ACTGAACAATTTCACTCCTCTTGACTTACTGGCCTAG   8986

EMBOSS_001      9366 TATATTATACTAATTGTCTCCTTGTCT-CTCTAAGCACTCAT--TCCTTC   9412
                     .|.|.|.||.|||.|..|||....||| |||.|.|...|..|  ||.|.|
EMBOSS_002      8987 CAGAATGTAGTAAATCCCTCTCCATCTGCTCAATGACATGGTCTTCATGC   9036

EMBOSS_001      9413 CTCTTTCTTTCT-----TCTTTTTTTTTTTTTCACTT---TTATTTTAAG   9454
                     .||.|| ||.||     |||.||...|.|.|||.|||   ||.|.||.|.
EMBOSS_002      9037 ATCATT-TTCCTGAGGATCTATTCCATGTATTCCCTTGATTTTTCTTCAT   9085

EMBOSS_001      9455 CTCTAGGGGCACATGTGCAGGTTTGTTACATGGGTAAATTGCA--TGTCA   9502
                     ||| |....|.|.|.|..|   |...|||.|.. |||||..||  |.||
EMBOSS_002      9086 CTC-ATTATCTCCTCTCAA---TCAATACTTCC-TAAATGTCAGTTATC-   9129

EMBOSS_001      9503 TGGGAGTTTGGTGAACAGATTATTTTGTCACCCAGATAATAAGCATGGTA   9552
                     ||..||.||..|....|.|||...|..|||.|...||.|.|...|||.|.
EMBOSS_002      9130 TGTTAGATTTCTTTCTAAATTGACTGCTCATCTCCATTAAACCAATGATT   9179

EMBOSS_001      9553 CCTGATAGGTAGTTTCTCAGTCTTCACCATCCTCCCACCCTCCACCCTAG   9602
                     |     |.|||  ||||    .|||    ||.|..||...|.||.||||.
EMBOSS_002      9180 C-----AAGTA--TTCT----GTTC----TCTTTGCAAGATTCATCCTAC   9214

EMBOSS_001      9603 AGTAGATCCTGGTTTCTGTTGTTCCCTTCTTTGTGTTCATATGT-ACTCA   9651
                     |.|...|..||..|.||||..|||    ||||......|||||.| |...|
EMBOSS_002      9215 ATTTCTTATTGAATGCTGGAGTT----TCTTAACTCACATATCTCAGAAA   9260

EMBOSS_001      9652 GTGTTTAGCTCCACTTATAAGTGAG-AATATATGGTATTTGGTTTTCTGT   9700
                     ..||..|....||.|.|.|.|..|| ||.|....||.||   ||.||
EMBOSS_002      9261 AAGTGCAAACTCAGTAAGATGCAAGCAAGACCATGTCTT---TTATC---   9304

EMBOSS_001      9701 TCCTATGTTATTTCACCTAGGATAATGGCCTCCAGCTCCATCCATGTTGC   9750
                     ||||  |.|||||..||..||.||..   |||||.       ||...|...|
```

```
EMBOSS_002      9305 TCCT--GCTATTCAACACAGTATCC---CCTCT-------TCTGGGAGAC   9342

EMBOSS_001      9751 TGCAAAGAACATAATCTCATTCTTTTTTCTGGCTGCACAGTATTCCCTGG   9800
                     |.||  ||.|||.|    |||..|||.|.|..||.|    |.||   |..
EMBOSS_002      9343 TTCA--GATCATGA----ATTTATTTATATATCTTC----TTTT---TTA   9379

EMBOSS_001      9801 TGTATATGTACCACATTTTCTATATCTGATCTACCATTGATGGGCATTTA   9850
                     |||..||.|  .|||||||||.|.| |||.||.       ||||  ||
EMBOSS_002      9380 TGTCAATTT--TACATTTTCTTTTT-TGAACTC-------TGGG--TT--   9415

EMBOSS_001      9851 GGTTGATTCCATGTCTTTGGTATTGGGAATAGTGCAGCAATGAACATACA   9900
                       |||..|.||||.   |.||||..|   ||||.|   ..|..||||||.
EMBOSS_002      9416 --TTGTCTACATGA---TTGTATATG---TAGTTC---CCTTTACATACT   9454

EMBOSS_001      9901 GCTGCATGTGTC-TTTATGGTAGAATGATTTATATTCCTTTGGGTATATA   9949
                     .||.|...|.|. |||.||.||.     |.|.|..||.|.|..||.|..|
EMBOSS_002      9455 ACTACTACTATAATTTGTGTTAT-----TCTTTGCTCGTATCTGTTTCAA   9499

EMBOSS_001      9950 CCCAGTAATGGCAT-TGCTGGGTTGAACGGTAGTTCAGTTTTGAGTTCTT   9998
                     |.||.|..|||||. |||||.| ||..|....|...||..|.||....|||
EMBOSS_002      9500 CACATTTCTGGCAACTGCTTG-TTCCAATCCAACCCACATATGTAAACTT   9548

EMBOSS_001      9999 AGAGGTATTTCCAAACTGCTTTCCACAGTGGCTGAACTAATT--TACATT  10046
                     ..   ||...|..||.|....|....|.|.| |.||..||||  |.||.|
EMBOSS_002      9549 CC---TAAAACATAATTTAGATGGTAACTAG-TCAAGCAATTGTTCCAAT   9594

EMBOSS_001     10047 CCCACCAACAGGGTATAAGCATTCC--CCTTTCTTCACAACCTCACCAGC  10094
                     .|..||..|||..||.|..|..|..  |.|..|.|.| |..||.||.||.
EMBOSS_002      9595 TCACCCTGCAGTTTAAACTCCCTAAAGCATAACCTAA-ATGCTAACTAGA   9643

EMBOSS_001     10095 ATCTGGTATTTTTTGACTTTTTTTTTTTTTTTTTTTTTTTTTTTGAGACG  10144
                     ||.|.| |||||.||..........||...|.|...||..|||....|.|
EMBOSS_002      9644 ATGTAG-ATTTTATGCAAGAAGAGATTAAATGTGAATTCATTTACTTATG   9692

EMBOSS_001     10145 AAGTCTCGCTCTTGTCCCCCAGGCTGGAGTGCAATGGCGCAATCTTGGCT  10194
                     .|.||   ||.||||||... |..|||.|.||   ||....|....||.||
EMBOSS_002      9693 TATTC---CTATTGTCAAG-ACACTGTATTG---TGTATGATGGGTGTCT   9735

EMBOSS_001     10195 CACTGCAACCTCCACCTCCCGGGTTCAAGTGATTCTCCTGCCTCAGCCTC  10244
                     .|.|. ||..|..|.....|...|   ||.||..|...||.|||    ||
EMBOSS_002      9736 TAATA-AATATTTATTAGTCAATT---AGAGAAACATATGACTC----TC   9777

EMBOSS_001     10245 CCAAGTAGCTGGGATTAGAGGCGCCTTCCACCATGCCTGGCTAATTTTTT  10294
                     ||||.||..|    |||       ||||.            |||..|.||
EMBOSS_002      9778 CCAAATATTT----TTA-------CTTCT------------TAAAATATT   9804

EMBOSS_001     10295 ATTTTTAGTACAGACAGGGTTTCACCAGGTTGGCCAGGCTGGTCGCAAAC  10344
                     |.       |.|.|.||..||..|.||||  .|...|.||.|||    |.
EMBOSS_002      9805 AC-------AAAAAAAGCCTTGTAGCAGG--AGATGGCCTAGTC----AG   9841

EMBOSS_001     10345 TCCTGACCTCAGGTGATGCGCCCGCCCCGGCCTCCCAAAACGCTGAGATT  10394
                     ||.|.|    |.|.||.|.|.  |.|||..|.||....||     |..||
EMBOSS_002      9842 TCATCA----ATGGGAGGGGA--GGCCCTTCGTCTTGTAA-----ACTTT   9880

EMBOSS_001     10395 ACAGGTGTGAGCCACCACACCAAGCCCACAGTATCAATTCTATGCATTCT  10444
                     |.|    ||..|||..|||......|..|||...|||......|.|.|..
EMBOSS_002      9881 ATA----TGCCCCAGTACAGGGGAACGCCAGGGCCAAGAAGTGGGAGTGG   9926

EMBOSS_001     10445 TTTCTGATTTCATTAATCTCATTATCTTCATTTGATATTTAGTCAATAGT  10494
                     .|....|....|.      ||...|........|.||  |||      |.
EMBOSS_002      9927 GTGGGTAGGGGAG------CAGGGTGGAGGGACGGTA--TAG------GG   9962

EMBOSS_001     10495 TACTGTCAGTTATGTGTTAGTTATTATACTAGAAACAGTCTTTTCTCCAT  10544
                     .|||.||||....|..||.|..||...|.||.|.|.|.|.|.|..|..|.
EMBOSS_002      9963 GACTTTCAGGGTAGCATTTGAAATGTAAATAAAGAAAATATCTAATAAAC  10012

EMBOSS_001     10545 CTCCTTTAATCCAATGATTTGAACATTTTTATTCCTTTCCAATGTCTGTC  10594
                     ..|||..|.|.|.||.|..|||||||             |||
EMBOSS_002     10013 ACCCTCGATTTCTATAAAATGAACAT-------------CAA--------  10041

EMBOSS_001     10595 CCACATTTCTTACTGTATGTAGGACATTTCTTACTCAAATGTCTCACAAA  10644
                       |||.    .|||..| ||.|.|..|||..||.|.|.||   |||
EMBOSS_002     10042 --ACAA----GACTAGA-GTTGTATTTTTAATAATTATAT---TCA----  10077
```

```
EMBOSS_001     10645 TGACATAAATTCAGTATGACCCAAATAGGCCATTTTTTATACCAAGTCTT  10694
                     ||.| ||.|  ||||.|.|..|...||.||.||.|.|        |.|||
EMBOSS_002     10078 TGCC-TAGA--CAGTTTTATACTTGTATGCAATATAT--------GCCTT  10116

EMBOSS_001     10695 ATTTCCTATCCTGCTGTTCATCCCGGTACCATCTTTTCAGTCAGAGAGTT  10744
                     ||.||| ||| |.|..|||.|..|   ||.|.||||.|..|    .|.|.
EMBOSS_002     10117 ATCTCC-ATC-TCCCTTTCCTAAC---ACAAACTTTCCTAT----AAATC  10157

EMBOSS_001     10745 CAGATCATATAGTCATTTCTAAATCTCCCACTTACTTGCCTCACTTTCAA  10794
                     |...||..|.|.||||.|...||      |||.|...|..|...||
EMBOSS_002     10158 CCTCTCCCACATTCATGTGGGAA------ACTAAAAGGGGTTGTTT----  10197

EMBOSS_001     10795 GTTCATTTTTAAGGTCTGTAGATTCTGCCTCCCTAATTCTTTATGACCAT  10844
                     |||.||||.|..|.| |||...||.||.|.|.||.|.|.|..||.|
EMBOSS_002     10198 GTTTATTTGTTTGTT-TGTTTGTTTTGGCACACTTAGTTTAAATAA----  10242

EMBOSS_001     10845 TCCTTTCTCACTAGCCCCTTACCTCCACTCTCATTCACACTCTTACTATT  10894
                     ||....|...|||..   |.|||.....|.|.|...|.||..|||..|.|
EMBOSS_002     10243 TCGGAACCAGCTATG---TGACCATTGGTTTGAAATAAACAATTAGAACT  10289

EMBOSS_001     10895 TTTTACCCTCCTCCACTCATTCCTGCCCACCAGTGGCTCCAATCCAACTT  10944
                     |..|.                  |||.|||||...|||.||   |||||
EMBOSS_002     10290 TGGTG------------------TGCTCACCAACAGCTACA---CAACT-  10317

EMBOSS_001     10945 GCAGATTTCCATTTAAATTAAGCTTCCTAAAACATAGCTTAGGTTGTAAC  10994
                        ||...|.||.   |.||..|||.||..|..|....||..||.|..|.
EMBOSS_002     10318 ---GAAAACAATG---ACTACTCTTACTCCAGGAATTGTTCAGTAGCCAG  10361

EMBOSS_001     10995 TACAATGCAAATTCCATGAGAGCAAAGATTTCATCTGCTTTATTCACTTG  11044
                     ||...|    |||.      |.|||.||.. |||.||||.      |||.
EMBOSS_002     10362 TAGTTT----ATTA------AACAAGGACC-CATGTGCTC------CTTC  10394

EMBOSS_001     11045 TATATATCCATTGTCCAAGACTGTGTGTGTCACATGAAAAGTGTTCAATA  11094
                     |      ||||   |||.||.||..|       |||.|.||||.....|.
EMBOSS_002     10395 T------CCAT---CCATGATTGACT-------ATGGACAGTGCCAGCTT  10428

EMBOSS_001     11095 AGTATTTGTCAGTGAACGAAAATAATATATGACTCCCCTCTTCAA-ACAC  11143
                     |||    |..|||.|.|..||.|..|.|  ||.|    |||||| |||.
EMBOSS_002     10429 AGT----GCTAGTAATCACAACTGCTCT--GAGT----TCTTCATTACAA  10468

EMBOSS_001     11144 CTTTTTTGACTTCAAAGCCCTTCAGAATATTCTACAGACTCCTTCACCTG  11193
                     ||..||.  .|||.|  |.||||       ||      ||.|||||.||.
EMBOSS_002     10469 CTGGTTC--ATTCTA--CACTTC-------TC------CTTCTTCATCTC  10501

EMBOSS_001     11194 GCTCTCCACAATTGCCCCTGAGTCTCGTTTCCAATCTTATTTCTTATTTT  11243
                     .|..|.|.|||.|||..||.|  ||.||.   ||....||..|.|||.||
EMBOSS_002     10502 ACCTTACCCAAATGCTACTTA--CTTGTG---AAAAAAATACCATATATT  10546

EMBOSS_001     11244 ACCTCTCAATGCACCTTCAACTCCTACTAAAATGAACAGCTAGCCAGCTT  11293
                     ..|. |||||.|...|||.       ||.|..||||...||  ||...||
EMBOSS_002     10547 GACA-TCAATACTTGTTCC-------CTTAGCTGAAAGTCT--CCTATTT  10586

EMBOSS_001     11294 ACTTCTGTGTCTTTCGATGATCTTGTTTTTTGTCTTGAGATTCCTTTTTT  11343
                     .|||||     |||..|..|..||.||.||       ||||  .||||.|
EMBOSS_002     10587 CCTTCT-----TTTAAAAAAAATTCTTATT-------AGAT--ATTTTCT  10622

EMBOSS_001     11344 TCATCTAAGCTTACCCAAACATTACCTACTTTTCAAGGAAAGCCATTTTC  11393
                     |.||.|||..||   ||||..|||.|  |||||....|....||..|.|.
EMBOSS_002     10623 TTATTTAAATTT---CAAATGTTATC--CTTTTTCCTGGTTTCCGCTCTG  10667

EMBOSS_001     11394 GAATCTTCCCTTTTTCCCTGAGCCCCCAAGCTGGAAGACATCTTGTCTCC  11443
                     .||.||  ||.|.|.|.||    ||||...||        ||.|| ||||
EMBOSS_002     10668 AAAACT--CCCTATCCACT----CCCCTCTCT--------TCCTG-CTCC  10702

EMBOSS_001     11444 ATCTCAATTCCTATAGGCATTTCTCTGCACTTTAAATGACG--TTTAGTA  11491
                     .|   ||||||...|||||||||||||.|.|.|.|||.|.|  ||...|.
EMBOSS_002     10703 TT---AATTCCCGAAGGCATTTCTCTGAATTATCAATAATGGGTTATTTT  10749

EMBOSS_001     11492 CTTCTGACATTGCATTAGAGAGAGGCTGGGGTGGATAGTGTTTCATAGTG  11541
                     .||||.....|.|.||..|||.|..||.|.|.|||.|||.|||.|||.|.
EMBOSS_002     10750 TTTCTTTTGCTTCCTTTTAGAAAAACTAGAGAGGACAGTATTTAATACTT  10799
```

```
EMBOSS_001     11542 TGAACTTTGAAGCCCGACTGCCTGAGTTTAAATCGTGATTCTGGGGCTTA  11591
                     ||||.|||  |||     ||..|..|||||||.|.||.|||||...|.|.
EMBOSS_002     10800 TGAAGTTT--AGC-----TGTGTAGGTTTAAACCATGGTTCTGATTCCTG  10842

EMBOSS_001     11592 CTGACCATAGACGCATTTCTGAATTGCTCTCAGATTATGGAGC-ATAAAT  11640
                     .|||||||..||..|||||||.|||||||.|...|||||...|| ||||||
EMBOSS_002     10843 ATGACCATGAACTTATTTCTTAATTGCTTTTGTATTATATTGCTATAAAT  10892

EMBOSS_001     11641 CAAAAGTAATGACAGCTACCTCTTCAGGTTGTTG-TGAGGGTGATGCGAA  11689
                     .||.|..||..|.|.|||.. |||||.||..||. |.|...|||||..|||
EMBOSS_002     10893 AAACAACAACCATATCTATA-CTTCATGTACTTTATAAAACTGATATGAA  10941

EMBOSS_001     11690 TTAATGTAC-TGAAGTGCATGGAACAGTTTCTGGCACACGGTAAGCACCC  11738
                     .|.||.|.. ||||.|.|.|..|||.||..|||.|. |..||||||||||
EMBOSS_002     10942 GTCATATCTATGAAATTCTTAAAACTGTACCTGACT-ATAGTAAGCACCC  10990

EMBOSS_001     11739 AATAAACATAGCTAATATTATGTTATTACTATTTTCAGGCTTATTTTTAT  11788
                     |||||.||||   |.||..|||..||||| |||......|..|||||.|||
EMBOSS_002     10991 AATAAGCATA---ATTAGCATGCCATTAC-ATTCATTCTCACATTTATAT  11036

EMBOSS_001     11789 GTATACATATAGTATGTAATTTTATGTCAATATGTATAAATAGACTTTGG  11838
                     ||..|..|||||| |.|| ||||.||...|||||.|....|||.|.||...
EMBOSS_002     11037 GTGCAAGTATAGT-TATA-TTTTTTGCTCATATTTGGTGATACATTTCCA  11084

EMBOSS_001     11839 TATTGTTTATT-TCA-CTATCACCTTGAGAGCACAATTCTCATTTGATTT  11886
                     ||.|.||.||| ||| ||....||......|.|.|||||.||||||||||
EMBOSS_002     11085 TAGTATTCATTGTCATCTGGGCCCACTGTGGAATAATTCCCATTTGATTT  11134

EMBOSS_001     11887 GTGTGAGAAACTACTTAGAAAGAAATAGACGTGTGAATGAAACTATGCTT  11936
                     |..||||||...||     |.|||||.||.||   |..||.||.|| |.|
EMBOSS_002     11135 GAATGAGAATGGAC-----ATGAAATGGATGT---ACAGATACAAT-CCT  11175

EMBOSS_001     11937 GAAATATTGGTTACTGTGAGTGTTGAAAATCCATTTTGTTTAAAGAAAGC  11986
                     .||..|||||.|.|..||||||.|||.                 ||||..
EMBOSS_002     11176 AAAGCATTGGCTTCCATGAGTGCTGAG-----------------GAAAAA  11208

EMBOSS_001     11987 TTCAATTGTTAATCTTCCATAAATTTTAGTTCTTAAGCGTTCATATTGAC  12036
                     |  |.|||||||||.||..||.||||||.|||||||..||..|||||||
EMBOSS_002     11209 T--ATTTGTTAATCGTCAGTATATTTTACTTCTTAAATGTCAATATTGA-  11255

EMBOSS_001     12037 TCGTTTTGGAAAAGCTCTTTAAAGTCTTGGGATATAAACAAGGCTGAATA  12086
                         ||....|||.|||||||||.|||...|||| |||.||..||.||||
EMBOSS_002     11256 ----TTAAACAAAACTCTTTAAACTCTCCAGATA-AAATAAACCTTAATA  11300

EMBOSS_001     12087 CCCTCATTCATGATAACAAACATATTATACTGAAAATTGTAAGAGAGATA  12136
                     .||| |||..|||.||.|||.||||.|||.|||||.|...||.|.|.|.|
EMBOSS_002     11301 ACCT-ATTTTTGAGAATAAATATATCATATTGAAAGTAAAAAAAAATAAA  11349

EMBOSS_001     12137 -------------------TTTTATCTTTCATAATGCCCTCCTTGGGAAA  12167
                                         ||||.|.||.||||..|.||||||||||.||
EMBOSS_002     11350 AAAAAACAAACAAACAGATTTTTGTTTTGCATAGCGACCTCCTTGGGGAA  11399

EMBOSS_001     12168 ATACATTGACTTGGC--CCTTCTCTTTCAATCAGACACCAAAGTTGAGAT  12215
                     .|| |.|||||||..  ||||..|||||||..||||||.|     .|||||
EMBOSS_002     11400 GTA-ACTGACTTGTTATCCTTTGCTTTCAACTAGACATC-----CGAGAT  11443

EMBOSS_001     12216 TGCCTGAAACACAGTTTGGTAAAAGGAGTTTCTTTTTCCCAAACATCCTG  12265
                     | |..||||||..|.|||||..||||||..|.||.||||||.||..   |.|||
EMBOSS_002     11444 T-CTGGAAACCTACTTTGTCAAAAGTGGCTTATTTTTTCCTG---TTCTG  11489

EMBOSS_001     12266 AGTAACACAGGAAATCACACCAATGACTGATAGATAACGTTAATAAAATT  12315
                     ||.||||.|||||||.|||.|||.|.||.|.|||.|.|.|.|.||||||..
EMBOSS_002     11490 AGCAACAGAGGAAATGACATCAAGGTCTTACAGAAATCTTAATTAAAAAA  11539

EMBOSS_001     12316 AATAAAGTTGTTTTAAATG--------CATACCATGGG------------  12345
                     ||.||..|||||||...|.        |.|||.||...
EMBOSS_002     11540 AAAAACTTTGTTTTGTTTTAGTTTTCCCTTACTATATTTTGTTTGTTTTT  11589

EMBOSS_001     12346 -----GCAGTGGCAATGAAAACATTG---------------AGAAGGCT-  12374
                           |...||.|..|..|||..||.               ||||.|.|
EMBOSS_002     11590 TCTTTGTTTTGACTTTTTAAATGTTTCTGTTTTTGAAAGACAGAATGGTA  11639

EMBOSS_001     12375 ---GGGACTATTTGCC-----AACTTTCTTTGATCTCCATTAG-------  12409
```

```
                        |..|||.|.||.|     |..|.....||.||||.|..||
EMBOSS_002     11640 ATTGATACTCTGTGGCTGGGGAGGTGGGGATGGTCTCAAGGAGCTGGGAG  11689

EMBOSS_001     12410 AACCTGGACAA----GATCCACATA-ATTTCAGAACTTCTTCTCCAAAC-  12453
                     ||..|..|.||    ||||.|.||| ||||.|.|...|.|..|..|||.
EMBOSS_002     11690 AAGTTAAAAAAATATGATCAAAATATATTTTAAAGAATTTAATTAAAATT  11739

EMBOSS_001     12454 -AAGAATTGAAAAGGTCAGGAAAAGTTTGACCACAGAAA--AATGTCAAA  12500
                      ||.||...||||...|...|||....||.|.|.||.||  |.||...||
EMBOSS_002     11740 GAATAAAATAAAATTGCTTTAAATAAATGCCAAGAGGAATCACTGGAGAA  11789

EMBOSS_001     12501 GAATTTTGTG-TCA-----CTTTCTCCTCCTC----CCTTCCTCTAACCT  12540
                     .|.|..||.| |||     |..||.|.|.|||    |||..|.|..||||
EMBOSS_002     11790 AATTAATGAGGTCAACAAACACTCACTTTCTCTGAACCTGGCCCAGACCT  11839

EMBOSS_001     12541 TGAATAATTTTTTAGGGTTATTGGTCTTTGGGAGC---AGACTTTCTAGA  12587
                     ..| |||||||||.....|...||.||....||.   |||.....||||
EMBOSS_002     11840 ACA-TAATTTTTTTTCCAAACAAGTATTAAAAAGGTCAAGAAAAATTAGA  11888

EMBOSS_001     12588 CCAAAACAAAAAAAAATGATATTCCTCTATGTGATAGGTAACAA-TCACTA  12636
                     ..|.|..|||||..|.|..|||..|...||||.|   ||..|.. ||..|.
EMBOSS_002     11889 GTACAGAAAAAATCAAGTAATTAATTGCTGTCA---GTCTCTTGTCTTTT  11935

EMBOSS_001     12637 CCCATCCTACTGGAAAATTCTCAAAGTGT-------AAATTGA---GGG-  12675
                     |...|..|.||||||.|.|.|...|||||       |.||||.|   |||
EMBOSS_002     11936 CTTTTTTTCCTGGAAGAGTTTTTGAGTGTTTGGAGCAGATTTACCAGGGC  11985

EMBOSS_001     12676 --------GATAAA------------------------------AAAAGA  12687
                             ||||||                              ||||||
EMBOSS_002     11986 TGGTATATGATAAATGGTAGACCTTCCATTTTGCGTTGAATTGGAAAAGA  12035

EMBOSS_001     12688 ATCTTAAGTCCTTTA----------AATTATTTTTAAGATGAACTACAT-  12726
                     .|..|.|||.||.||          ||..|..|.||...|.|..|.|||
EMBOSS_002     12036 CTAATCAGTACTGTATTTAAAAAAAAAAAAAACTATACTTTAATTTCCATC  12085

EMBOSS_001     12727 -TAGTGCCTCTCTTGTGCCTTTCATAATTCTGATAATAAAACATTCCAGG  12775
                      ||.|...||||||.|.|||||..|||||||.|.|||||||..||..|..
EMBOSS_002     12086 TTAATATTTCTCTTCTACCTTTGTTAATTCTTAAAATAAAAATTTTTAAT  12135

EMBOSS_001     12776 TATTAGTCAAAGATTAATGGTATTGAAAATAATTTAGGTTATCAGCATGT  12825
                     |||.|||.|||.|.|.|...|.||.||||||.|||||.||.|||||||.|
EMBOSS_002     12136 TATCAGTTAAATAGTGAACATGTTCAAAATATTTTAGTTTGTCAGCATAT  12185

EMBOSS_001     12826 GATTTTCATTCCACATGAGGTCCTTTTGCAGTTTACATGGTTTTCTAAAT  12875
                     |||.|||||||||||||||.|||.||||.||||||..||||||||||||
EMBOSS_002     12186 GATATTCATTCCACATGAGCTCCATTTGTAGTTTATGTGGTTTTCTAAAT  12235

EMBOSS_001     12876 TATATTAAAATAAAATGTCAGAAAGTTCACATTTTTTT------------  12913
                     ||.|||||| |.||||||||||||.|||.||.|||||.|
EMBOSS_002     12236 TACATTAAA-TGAAATGTCAGAATGTTGACCTTTTTATTTACCTTTTTAT  12284

EMBOSS_001     12914 -----------------------CATGT-----------------------  12918
                                             |||||
EMBOSS_002     12285 TGATTCTTTTTGAGTTTCACATCATGTAACCCAAACCCACTTATCTTCCT  12334

EMBOSS_001     12919 --------------------------------------------------  12918

EMBOSS_002     12335 GTCCCTTAATACACACCCTCTGCCCATGCAACTTCCCCCATGATAAGAAA  12384

EMBOSS_001     12919 --------TTAACAGCATCAATCTTTAAAGAAAAGTTATTGCACAAAGGT  12960
                             |||.|||.|||||..|||||||||||||||.|||.||||||||
EMBOSS_002     12385 GTAGACTTTTATCAGTATCAAATTTTAAAGAAAAGTTCTTGAACAAAGGT  12434

EMBOSS_001     12961 CTGTGCATAAATCAGCCATTCTCCGAAGAGGTAAAAGAAGTCATTACGCC  13010
                     .|||||||||.|..|..||||..||..|.||.||.||.||.||||   |||
EMBOSS_002     12435 GTGTGCATAAGTTGGGTATTTCCCATATAGATAGAATAAATCAT---GCC  12481

EMBOSS_001     13011 TGGTTATGAGAGAGAGTTTCATGAATGTAAGAGACATAAATCATTTCCCA  13060
                     |||||.|||.||||.|.|.|||.|||.||||||||.||||||..||||||..|
EMBOSS_002     12482 TGGTTGTGAAAGAGTGATGCATTAATATAAGAGATATAAACTATTTCTTA  12531

EMBOSS_001     13061 CTGGAGATCATATTAGTCTAGATGGAAGAATGTCTGTTTCTTGATAGTGA  13110
                     .||||||||..||||..|.||||||||||.||.||||.||||..|||||||||
```

```
EMBOSS_002     12532 TTGGAGATCCCATTCATATAGATGGAAGGATTTCTATTTTCTGATAGTGA  12581

EMBOSS_001     13111 GAAAGCAACAAATTACTTTTGTTTGCTCCTGAGTCTGTGGTTGTCCTTGA  13160
                     .|||||||||||||||.|....||||||.||...||.||||||||.|||||
EMBOSS_002     12582 AAAAGCAACAAATTATTACCATTTGCTTCTACATCCGTGGTTGTTCTTGA  12631

EMBOSS_001     13161 GAGGTCTGTTAGCATGTTGACTATTGACTATTCAATATTAGCATTATAAT  13210
                     |.|||||||||..||||||||       |.||||.||||||..||||.||.
EMBOSS_002     12632 GGGGTCTGTTAATATGTTGAC-------TTTTCAGTATTATAATTACAAC  12674

EMBOSS_001     13211 AACTTACAATGATCTGAGTCACATAAATATAATCTTTCAGTTCTCTAAAG  13260
                     |.|..| |.|.||.    ||||    ||||.||||||...|||||||||.
EMBOSS_002     12675 ATCCCA-AGTCATT----TCAC----ATATCATCTTTGCATTCTCTAAAT  12715

EMBOSS_001     13261 ATTTTACTTTTT-CCTCTCTAATATCTATT-CACCTCCAACACCTT----  13304
                     |||||..||||| |||.||||||.|.|.|| ||||||  .||||||
EMBOSS_002     12716 ATTTTCTTTTTTTCCTGTCTAATGTTTTTTTCACCTC--TCACCTTGTGT  12763

EMBOSS_001     13305 -------------------------TGCAAATATATT-----ATTCT---  13321
                                             ||....|.||||     ||||.
EMBOSS_002     12764 GTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTATTTGTGTATTCGTGT  12813

EMBOSS_001     13322 ----CTGGGAGTTACAAAGAAAGTTATTCTCTGCAGGAAGCAGCATTTCA  13367
                         ||.|||.||..|.|.|.|.|||||||||||||.|..||.||.||..
EMBOSS_002     12814 ATTCCTAGGACTTGGAGAAATAATTATTCTCTGCAGAATCCAACAATTTT  12863

EMBOSS_001     13368 GTTGCTCTCAGGAGCCAACCACATTTCACCTCAATTCTTTGCTCCCAATT  13417
                      |.||||....|.||||.|||..||.||||.||||.||||||||||.||
EMBOSS_002     12864 -TGGCTCCTTTGTGCCAGCCATTTTCCACCCCAATACTTTGCTCCCACTT  12912

EMBOSS_001     13418 CAACAATTCAATATTGGATTAAATTCAAGGCTGTGACCCCAAATAGAATG  13467
                     ||.||.|||||||.|.||||.||||.||.||||||||.|||.||||||||
EMBOSS_002     12913 CAGCATTTCAATACTAGATTCAATTAAATGCTGTGACTCCAGATAGAATG  12962

EMBOSS_001     13468 AGACCTGGATATTTATGAACCACTTGACCAGGCATTCTTCCCATGATTTA  13517
                     ||.|.||||||.||||..||..|||||...||.|||.|| |.||.|||||.
EMBOSS_002     12963 AGTCTTGGATACTTACTAAATACTTGGTAAGACATCCT-CTCAGGATTTT  13011

EMBOSS_001     13518 --CTCCATAAATCCT--TTTTAGTTTTTGCAGTAGCTTTACAAATATTTG  13563
                       ||||||||||.||  ||||||||||. ||.|||.||.|..|||..|.|
EMBOSS_002     13012 ACCTCCATAAATGCTATTTTTAGTTTTA-CAATAGTTTCATGAATGGTGG  13060

EMBOSS_001     13564 GAAAATGGCTGTGCAATGCAGTTTTAAAAAGTGCAATGAGTAGAGGTAGC  13613
                     .|||||..|.|.|.||.||.|.||. |||.|..||||||||...||.||||
EMBOSS_002     13061 TAAAACTGTTATACACTGTACTTG-AAAGATGGCAATGAGCTCAGTTAGC  13109

EMBOSS_001     13614 TTCTTCACCTGGTATGGTAAATTGTTGATTCTCTT--TTGGAG-TGGAAA  13660
                     ...||.||.|...||||.|||......||..||||  ||||.| |.|..|
EMBOSS_002     13110 AAGTTTACATACCATGGGAAAGAAGCAATATTCTTGTTTGGGGATAGGGA  13159

EMBOSS_001     13661 ACAAGTGTTCTTATTTGGATGCAACCATTGCATTGATTAGACAACCCTAA  13710
                     |.|  ||||.||||||..|.|||||    ||| ||||.|||..||....||
EMBOSS_002     13160 ATA--TGTTATTATTGAGGTGCAA----TGC-TTGACTAGGTAATTGCAA  13202

EMBOSS_001     13711 ATTCATCTTTCATCCATGACCTGAAAGAAATTTTGAAATTCATGCAATAT  13760
                     ||..||.|..|.||||||||.  |||||||.....|.||||.|.|||||.|
EMBOSS_002     13203 ATGTATTTGCCTTCCATGACA--AAAGAAACAGGTATATTCTTACAATGT  13250

EMBOSS_001     13761 ATACCCGTAGTGGAA-AATGTACTTTTTGAATGGATTCCTGAATGTGACT  13809
                     |||....||||.||| ||||...||.||||.||.||||.||||||||.|.
EMBOSS_002     13251 ATATATCTAGTAGAATAATGGTTTTGTTGACTGAATTCTTGAATGTGTCC  13300

EMBOSS_001     13810 TTTAAGAAGAGCTATTA---AGAAGTGGGATCTTCTACAGAACAGTAAAC  13856
                     |||.||..||.||||||   |.||||||.||.|.||.|||.|.|.||||
EMBOSS_002     13301 TTTCAGGTGAACTATTATTGACAAGTGGAATTTCCTGCAGCAGAATAAA-  13349

EMBOSS_001     13857 AGGCATGAAAATATACAAGTTGATAAGATATGGAACTACCCCAAAAGAGG  13906
                           ||.|||..|.| |||..||||||.|  |||.|..|||..|.||||
EMBOSS_002     13350 ------GAGAATCAATA-GTTTTTAAGACA--GAAATGTCCCTGAGGAGG  13390

EMBOSS_001     13907 AATTAATAGTGGTGGGGCTTGGGGCAGGAGGACAGAGA-GACCTAGCCAA  13955
                     ||.|.|.|||.||||||..||||...|.||||..||... ||||||||.|.
EMBOSS_002     13391 AACTGAGAGTAGTGGGAATTGGCAGAAGAGGGTAGGTTTGACCTAGCTAT  13440
```

```
EMBOSS_001     13956 GGAAGGAAGGGCT-ATATTATAATAGAGTACAAAGTCCTTTAGTCATCCA  14004
                     |.|...||.|..| |.||.|.|||| |||||||...|. ||||.|.|.|||
EMBOSS_002     13441 GAAGAAAAAGAATGAAATGAAAATA-AGTACAGTATT-TTTAATGACCCA  13488

EMBOSS_001     14005 AGAGAAGGGGC-ACCTTCTGCATCCCTTATGAGTAAGATCAGAGAAGGTA  14053
                      ||||||..|. |..||||.      |||||....| ||||.|.|||||.|
EMBOSS_002     13489 -GAGAAGAAGATATATTCTA------TTATGCCCCA-ATCAAACAAGGGA  13530

EMBOSS_001     14054 TTCTAGTTAACTTTTGCTACATAACAAGCCAGCCCAAAACTTCATGGCTT  14103
                     |||||||||.|||||||||.|||   |.||......|||.|||||||.||
EMBOSS_002     13531 TTCTAGTTAGCTTTTGCTATATA---ATCCTCAATGAAATTTCATGGTTT  13577

EMBOSS_001     14104 ----CAGTAAAAATTACTTGTTTTGTTCATGAATCTACAGTTTGCTCAAG  14149
                         ||.|...|.|.|||..|||...|.||..|||||.|.||||...||.
EMBOSS_002     13578 ACTACATTTTTAGTCACTAATTTCACTTATTGATCTATAATTTGAGTAAA  13627

EMBOSS_001     14150 GTTCAATGGGGCTTGCTTATCCCTGTTTCAGTTGATATCAGTTGGGGTAG  14199
                     |..||....|.||.||  ||.|.||||....||.|.||......|.|.||
EMBOSS_002     13628 GGCCAGCATGTCTCGC--ATGCATGTTCTTCTTTAAATACTCCAGAGCAG  13675

EMBOSS_001     14200 ATTGCCTGATGCTGGAGGATTCACTTCCAAGAGGGCTCACTCACATGCCT  14249
                     .|...||...|||||.||||..||||||||||||....|||....|||||||
EMBOSS_002     13676 GTCAGCTATCGCTGCAGGAACCACTTCCAAGAACAGTCATATGCATGCCT  13725

EMBOSS_001     14250 GGAAAATAGGTGCTGACT-GTCAGTTTTTCTTCATGTGGACCTCTCCATG  14298
                     .|...|||    |..|.| ||||..||.|.|..||......|.||....|
EMBOSS_002     13726 TGTGCATA----CCCATTAGTCACCTTATTTGGATACTTGTCACTAATCG  13771

EMBOSS_001     14299 GAGCAGTTTGGGCTTTTTCACAGTGTAAGAGTTGGGTCCCAAGAGCAATT  14348
                     |||||       |.|..|..|  |.||||||..|..||.          ||
EMBOSS_002     13772 GAGCA-------CATAGTAGC--TCTAAGATCTTAGTT----------TT  13802

EMBOSS_001     14349 ATCCTAAGGGACAAGAAATTAAAGCTGCAAGCTTCTCAAGGCCTGCCCTA  14398
                     .||.|.|.           ||.|.|| |...|..|||....|.|.||||.
EMBOSS_002     13803 CTCTTTAA-----------TATATCT-CTTTCCACTCTCTTCTTTCCCTT  13840

EMBOSS_001     14399 AAAGCAAGAATGGTTT--TGCTTCTCCCATATTCTATTTGTCAATCAGTG  14446
                     ....||...||..|||  ||.||.|....| ||.|.||| |||||   ||
EMBOSS_002     13841 TCTACATATATCCTTTGATGTTTGTTGGTT-TTATCTTT-TCAAT---TG  13885

EMBOSS_001     14447 ACAGAGCTCTGATTCAAGGGGATGAGAACATAAACTCCACCTTTCCATGG  14496
                     ||.         ||||.....|| ||||.||.||.|.||.      ||.|
EMBOSS_002     13886 ACT---------TTCATCTATAT-AGAAAATGAATTTCAA------ATCG  13919

EMBOSS_001     14497 AGAAGTATCAAAAAGTT-TTGATGCCATTTAATTAAAGCTGCCATACAAA  14545
                     |...||.||.|...|.| |||||....||.|...|||....|..|...||
EMBOSS_002     13920 ACCTGTCTCTAGGCGCTATTGATCTTGTTCACCCAAATGCTCTTTCTGAA  13969

EMBOSS_001     14546 GTTT--CTTATAAATGACACTGAGCTGAATGAATACTAAACAGCAAGTAG  14593
                     ||||  |.|.|...|..|| |.|.|..|.|||||.. |.||||.....|||
EMBOSS_002     13970 GTTTTCCATTTGTTTTCCA-TCAACCCATTGAAGT-TGAACAAATGATAG  14017

EMBOSS_001     14594 TCATTATCCCAGTCAAGAGAAGTTATCTTTGCTCAGAATACCCTTTCTCT  14643
                     ..|||||             |..|..|...||.||||||.|.|.|
EMBOSS_002     14018 AAATTA-------------ATTTCTTAAATGATCAGAACAGCAT------  14048

EMBOSS_001     14644 CCTTGTCTACCTGGAAAATTCAACTCTTGGCCAAAGCCCTACCTCTTCTC  14693
                      |.||..||   ||||   ||||.|.|....|||    |.|.||......
EMBOSS_002     14049 -CCTGAATA---GGAA---TCAAATATACTTCAA----CAAACTGAAAAA  14087

EMBOSS_001     14694 GA-AAGCATTACCAGGCCTTGCCTCTAAGTGTACAATTGGAGATACACCA  14742
                     |. |||.|.||.|||.|   ..|| |||||| |||.|..|.....||..|
EMBOSS_002     14088 GTGAAGTAATATCAGCC---AACT-TAAGTG-ACATTCTGCCTAGCATAA  14132

EMBOSS_001     14743 GTATACTGATGTTTTTAAAACTTTAAACTTTTTTCTACAATAAAACATAA  14792
                     ..|..|..|....|||.|||||.|..  |||....||....|...|||||.
EMBOSS_002     14133 CAAATCAAACAAATTTTAAACTGTGG--TTTCACTTATGGCAGGCCATAG  14180

EMBOSS_001     14793 ATTAAATAACTTCCCTTCTGACTTAAAAGCTGCAAAATGCTCATGACAGT  14842
                     |.||...||.|||  .|.|..||..|||..||.||.||||.|  |..||||
EMBOSS_002     14181 AATAGCCAAGTTC--ATATTGCTAGAAAATTGAAAGATGAT--TTCCAGT  14226
```

```
EMBOSS_001     14843 AACTATAT-----AAATTAAAATTAAATCTTAAGCACGATAAATACCT--  14885
                     .. |||.|     |.|..|.|...|..|||||...||  |..||.|||
EMBOSS_002     14227 GG-TATGTGGAGCAGAGGATAGGAATTTCTTATTTAC--TGGATTCCTGA  14273

EMBOSS_001     14886 CTCGAATAGCAACATAGATGCTTACTTCT--TTATTTCACTTCTTTATTT  14933
                     ||...|..|..|.||| |||...|.||||  |.||......|..||||.|
EMBOSS_002     14274 CTTTCAGTGAGAGATA-ATGAAAAATTCTAATGATGGGTAGTAATTATAT  14322

EMBOSS_001     14934 GCTTTTCTTTGTCTATAGTTTGCCCCAAAGGTATTTTAATAATATCGGGT  14983
                     ...|||.|||||.| ||.|||.....||  .|.|||.|.|.|| ||...|
EMBOSS_002     14323 TTCTTTGTTTGTTT-TAATTTAATTTAA--TTGTTTAACTTAT-TCACTT  14368

EMBOSS_001     14984 TCCATGTATACCAGTGTGTACCAATTAATATTTAGAATATACCTGTTAAT  15033
                     |.|||......|||||    ||..|...|   .||....|.|||...|..
EMBOSS_002     14369 TACATCCTGCTCAGTG----CCCCTCTCT---CAGTCACTCCCTCCCACA  14411

EMBOSS_001     15034 AACCTCATTTGCATAGCCCTACTAATCTGAGCACAGCGCAGCCT-TAAGA  15082
                     |.|||  ||..|.|..|||||.||..|||             ||| |||||
EMBOSS_002     14412 ATCCT--TTCCCCTTTCCCTCCTCTTCT-------------CCTCTAAGA  14446

EMBOSS_001     15083 AAGTCTTAGTTTTTCTCAGTTTAGTTCATCTCTCTTCTCTT-CTCCTCCT  15131
                     ..||....|.||.|||..|..||...|.........|.||. |.|.||..
EMBOSS_002     14447 GGGTGGGGGGTTCTCTGGGAATACCCCCCACACACACACTGACACTTCAA  14496

EMBOSS_001     15132 GTCTCTCTTATTTCCTATTTCTTTTTCTTTTCAAGTGACTTTCAACTAAG  15181
                     ||.|||.|.|.  .|||.. |...|.||.|.|.|.|||......||.|.|
EMBOSS_002     14497 GTATCTGTGAG--ACTAGG-CACATCCTCTCCCACTGAGGCCACACAAGG  14543

EMBOSS_001     15182 TAGAAAATGCATTTCACATCACTATGCCGGCCTCCAGGCTCTGTCTATTT  15231
                     .||...| || |...|.|.|| |||.||.....| ||||.....||.||.
EMBOSS_002     14544 CAGCCCA-GC-TAGAAGAACA-TATTCCATAGAC-AGGCAACAGCTTTTG  14589

EMBOSS_001     15232 CATTCACCCAGGAATGCCCTTTCTGAATGCTTTCTCTCATTTAGCAGCTA  15281
                     ...|..|||..|  |.||..|  ||..||...|| |.|||..|| |.|.|
EMBOSS_002     14590 GGATAGCCCTTG--TTCCAGT--TGTTTGGAATC-CACATGAAG-ACCAA  14633

EMBOSS_001     15282 TCTATTGAAGTTGGACAAATGATAGAAATTCATTTCTTAAAGAGCCAGAA  15331
                     .|||...|..|...|.|.|||....|......||...||.|||.|..|.|
EMBOSS_002     14634 GCTACACATTTGCTATATATGTGCTAGGAGGCTTAGGTAGAGAACATGTA  14683

EMBOSS_001     15332 CATCATCTTGAACAAGAAGTTAAAA----GAATTCAGCAAATCAAAAGAT  15377
                       |..||||.....||..|||.|.|    ||...|..|||......||.|
EMBOSS_002     14684 --TGTTCTTTGGTTAGTGGTTCAGACTCTGAGAGCCCCAAGGGTCTAGGT  14731

EMBOSS_001     15378 GAGCTAA---TATGGGTGAATCT-TAGAGGCATTATGC---TAAGTGAAA  15420
                     .||.|.|   |.|.|||....|| |.|||...||||.|   |.||.|..|
EMBOSS_002     14732 TAGTTGACTCTCTTGGTCTTCCTGTGGAGTTCTTATCCCCTTCAGGGCCA  14781

EMBOSS_001     15421 TAAACCAGA-CACAAAATGAAA-AATATTGTATGAT-----TCCACTGGT  15463
                     ..||..|.. |.|..|.|.... ||.|.|.....||     |||||||.|
EMBOSS_002     14782 GCAATTACTTCTCCTATTCTTCTAAGAGTCCCCAATCTCCATCCACTGTT  14831

EMBOSS_001     15464 ATGA----GCTACCTACAACAGTCAAATTTA--TACAGACGTAAAGTTGA  15507
                     ..||    |.|..||..|.|.|||..|.|.|  |||.|| ||..||....
EMBOSS_002     14832 TGGATGTGGGTGTCTGTATCTGTCTGAGTCAGCTACTGA-GTGGAGCCTC  14880

EMBOSS_001     15508 AGGATGTTACCAGGAGCTGGA---------GGAAGAAGAGAATGAGGGCT  15548
                     .  .||..||.|...||||||         |.|||.|.||.|.  .|..|
EMBOSS_002     14881 T--CTGGGACAACATGCTGGACTCATGTCTGCAAGCATAGCAG--TGTAT  14926

EMBOSS_001     15549 TATTGTTTAATGAGTACCTGAGTT-TCAGTTTGGGATGATGAAAACATTC  15597
                     .|||..|||||.|||  .||||.| |..||.|.....|||.||....||
EMBOSS_002     14927 CATTAATTAATTAGT--GTGAGGTATTGGTGTTTACCCATGGAATGGGTC  14974

EMBOSS_001     15598 TAGAGATGGA-----TAGTGGTGA-----TGGTTCAACGATAATAATAAT  15637
                     |..||.|||.     ||.||||.|     |...|||...||||.|.|||||.
EMBOSS_002     14975 TCAAGCTGGGCCAGTTATTGGTTAGCCACTCCCTCATTAATATTTATAAA  15024

EMBOSS_001     15638 ATAATATTAATGTACTTAATAGTACTCAACTGTATACTTAAAAATGGTCA  15687
                     |.|.||..|||||  ||.|||..|...||.||||.|.||||||||.|.
EMBOSS_002     15025 ACACTAGGAATGT--TTTATACAATCAAAATGTACATTTAAAAATTGC--  15070

EMBOSS_001     15688 AGAAAATGGTACCCCGTTATCCTGATGTGATTATTACACATTGTAGGCCT  15737
```

```
                    ||||.||..|      ||||    .|.| |||.|||.||||.|||.|
EMBOSS_002     15071 AGAACATATT------TTAT----GTAT-ATTTTTAAACATAGTATG---  15106

EMBOSS_001     15738 ATATCAAAATATCTCATGTACCCCGTAAATATATGCACCTACTATGTACC  15787
                     |.||.|.||....|.|..||.|....||||...||.|   |..||.|...
EMBOSS_002     15107 AAATAAGAAGTAATGAGTTAACAGACAAATCCTTGAA---AGAATTTTTA  15153

EMBOSS_001     15788 CATAAAAAAAATTTAAAGGCTAAATGGCCAGGCATTGTGGGTCACTTCTG  15837
                     ...|||.||.|||...||...|.||.   |..||||.|...|...||||.
EMBOSS_002     15154 TGGAAATAATATTACCAGAGAATATA---AAACATTTTATATTGTTTCTT  15200

EMBOSS_001     15838 TAATCCCA---GAACTGTGGGAGGC--TAAAGCAGGAGGA-TCACTTGAG  15881
                     |.|.|..|   ||..||.||.|..|  |||....||||.. ||||.|...
EMBOSS_002     15201 TTAGCTAATGTGATTTGAGGAATACAATAATTTTGGAGCTCTCACATTTA  15250

EMBOSS_001     15882 CTCAGGAGTTCAAGACCAGCCTGG-GC-AACATGGCAAGGCCCCATCTCT  15929
                     .|      ||..||||..|....|| || ||..||||||..    ||.||.|
EMBOSS_002     15251 TT------TTATAGAAAATGTGGGAGCTAATTTGGCAGA----CACCTTT  15290

EMBOSS_001     15930 ACAAAGAATTCAAAAATTAACTGGGTGTGGGAGCTCATGCTTGTAGTCCC  15979
                     |..|||| |.|.||..|.||     ||..||...|..||.||||..|.||
EMBOSS_002     15291 AGTAAGA-TGCCAATGTAAA-----TGCTGGTATTTTTGTTTGTTTTACC  15334

EMBOSS_001     15980 AGACACACTGGAGGCTGAGGCAGGAGGATTCCTTGAACCCAGGAACTGGA  16029
                     |.   ||.|||....|.|.        |||.||   ||.||.||..|..|
EMBOSS_002     15335 AT---CATTGGTAATTTAA--------ATTTCT---ACTCATGATTTTCA  15370

EMBOSS_001     16030 GGAAGCAGTGAATGACACTGTACCCCAGCATGGTCAAGATCCCA-AATCA  16078
                     |.|.....|.|||..  .|..|.|..|..|...|.|..||.|.| ||..|
EMBOSS_002     15371 GAATTTTCTAAATTT--TTCAAGCATATAAATATGAGCATGCTATAAAGA  15418

EMBOSS_001     16079 AAAAGAAATGATTAAAATGATCAATTTTATGTTGTGTATATTTTGC-CAC  16127
                     ..||.|.|..||.|||.|...|...|.|||||....|..|.|..|. |||
EMBOSS_002     15419 TGAACATAGCATGAAATTATACCTCTCTATGTACCTTCAAATAGGTACAC  15468

EMBOSS_001     16128 AATACAAAAATGG----GGAAAAGCCTATTCGCTTTTAAGTATCCTTAAA  16173
                     .|..||.|.||.|    |.||||..|..|||...||.||.|||.|..|..
EMBOSS_002     15469 TAGTCAGATATAGCTGTGCAAAATACACTTCCTCTTCAATTATTCACAGC  15518

EMBOSS_001     16174 AAGGCACAGCTTCTTCAGCTAACAGAC-TCTAAAACTTTTTTAATAGAA  16222
                     |.|...|||.....|.|..|.|...|. |.||..|||.|.|.|.|||.|.
EMBOSS_002     15519 ATGTGCCAGTCAGATTACTTCATTAATGTTTATGACTCTCTCTCATATAC  15568

EMBOSS_001     16223 GTATTAAGGTATTTAGAGAGTGCAAAATATCTTATTT---TAAGTCAA-G  16268
                      |||.    ||||.|.| |.|.||.||.||.||..||   ||.||..| |
EMBOSS_002     15569 -TATA----TATTCATA-ATTACACAAGATTTTTCTTGCATATGTGTATG  15612

EMBOSS_001     16269 AAGTTAGGGTCCTGTTCCTAAACACTAG---CCTCTGTAATCCTGGGGAA  16315
                     .|||.|....|.||||...||....||.   |.|.|.||....||.|.|.
EMBOSS_002     15613 TAGTAAATTGCATGTTTTGAATATATATGTACATATTTACAGATGTGTAT  15662

EMBOSS_001     16316 GTCAGTGCTGTTGGAG--ATCTCAGGTT------------GATCTTCTGA  16351
                     .|.. |...||||.||  |..|.|.|||            ||....|.||
EMBOSS_002     15663 ATGT-TTAAGTTGCAGCCAAATAATGTTTTATAAAGGAAAGAATGCCAGA  15711

EMBOSS_001     16352 AAAATGATGGATCTAGGTAAAAGATATGTTTCTC-CAGGTTTACATACCA  16400
                     .|.|..|.|.|...|||..|||||.|.|....|. |||...||.|.|.||
EMBOSS_002     15712 GAGAGAAAGAAAGAAGGGGAAAGAAAAGAAAATGTCAGAGCTATACAACA  15761

EMBOSS_001     16401 CGGACA-----CCATCTTTACTTGGAAACTTTATTAAAAATG--CATTGT  16443
                     .|.|.|     |||..||||..|.|    |||||..|...||  |.||||
EMBOSS_002     15762 AGAATATGTCTCCAAATTTATGTTG----TTTATCCATTCTGGTCCTTGT  15807

EMBOSS_001     16444 GT-CAGAAGCTCTCTGGGG------ATGGGTCGTGGAATCTGCATATGTA  16486
                     |. |.|||..|.|......      ||...|.|..||..|.|.||||.|.
EMBOSS_002     15808 GCGCCGAAAGTATAATATATCAACTATCTATAGATGAGCCAGAATATCTG  15857

EMBOSS_001     16487 AAGAGCCCCTAGGTAGTTCTTGTGCCCACTTAAATTTGAGAACCACTAGA  16536
                     ...||.||.|||.|   ||||.|.||||||||||||.|||.|||.|||..|
EMBOSS_002     15858 CCTAGGCCTTAGCT---TCTTCTACCCACTTAAATCTGAAAACAACTGAA  15904

EMBOSS_001     16537 CCAGATGTTTTGCTTATGGCCCTTTCAGCTCTGAAATTTGAAAAAAAAAA  16586
                     |.||||..|| |..|.||.||.||| |||||||....||||.|||...|
```

```
EMBOSS_002     15905 CTAGATACTT-GAATGTGTCCATTT-AGCTCTGGCGGTTGAGAAATTTA-  15951

EMBOSS_001     16587 AAATGATTCTG-CAAGACAGAGTCTCTGTGCTTTTGCAGGATAAAGAAAT  16635
                         ||.|||. |||.|||||.|.|              ||.||.|||||
EMBOSS_002     15952 ----GAGTCTTACAAAACAGATTGT--------------GAGAAGGAAAT  15983

EMBOSS_001     16636 GAAGAAAATAATAC--TTCCTGCTTGTGTTGGAGCATTTTTTTCATTTGG  16683
                     |.|||||||||.|.  ||.||||||.|.|||.||.||||||  ||||||.
EMBOSS_002     15984 GGAGAAAATAACAGCTTTTCTGCTTATATTGAAGTATTTTT--CATTTGC  16031

EMBOSS_001     16684 TATCCCCATCTCCAGTGGCTAGCCAATCAAGAATAGTATTGTTTATTCTT  16733
                     |.||.|||||||||..|..||.||||||||.||||||||||||||||||.|
EMBOSS_002     16032 TGTCTCCATCTCCAAAGAATATCCAATCATGAATAGTATTGTTTATTCCT  16081

EMBOSS_001     16734 CCCACTGTTTTGA-AGATACAAAAGGAAAAGCTAAGCCAGATGACACCTA  16782
                     .||||.||||.|. |||..||..|...|||||||||||||||..|.|||
EMBOSS_002     16082 ACCACAGTTTGGGGAGAAGCATGAATGAAAGCTAAGCCAGATGGTATCTA  16131

EMBOSS_001     16783 -AAGGCTTCCATTACCATTTTCATG-TTTTTCCCTTTGCATA------AA  16824
                      |||..||.||||.||||||||||. |||||||.||||||||      ||
EMBOSS_002     16132 GAAGATTTACATTGCCATTTTCATTCTTTTTCCATTTGCATATTATATAA  16181

EMBOSS_001     16825 AACTGTCCATGCCTCCATCAGAGCCATGATCACTAGTACAATGTTACACT  16874
                     |||||.||||||.||||||.||..||||||||.||.|..|||.|||...|
EMBOSS_002     16182 AACTGCCCATGCTTCCATCTGACACATGATCATTAATTTAATATTATGAT  16231

EMBOSS_001     16875 CTAATGACTCATGACATTAAATTATATC-TTAGCCTAATATGACCAAATT  16923
                     ||||||.||||||||||||||||||||.| ||||.|||..||||||.||||
EMBOSS_002     16232 CTAATGGCTCATGACATTAAATTATAACATTAGTCTAGCATGACCGAATT  16281

EMBOSS_001     16924 ACAATA-TCAGAATAAAAATTTCTTTTTTCAGGTTGAATCCCATAACTTA  16972
                     |.|.|. ||..||..||||   ||||||| ||.||.||||||...||.|||
EMBOSS_002     16282 ATAGTTCTCTAAAGGAAAA---CTTTTTT-AGATTTAATCCTTCAATTTA  16327

EMBOSS_001     16973 ATCCAATTATAATACTGGCTGAATTTTTCACAATTATGTCTCAGTCTTGA  17022
                     |.|.|||.||||.|||...|| |||||.|||.||||||||.|..|..||||
EMBOSS_002     16328 AGCAAATAATACTACCAACT-AATTTGTCATAATTATGTTTTGGCTTTGA  16376

EMBOSS_001     17023 TTTAGGGAATCTTCTCTTTATCATAAAAATGCATTTTGTTAAACATGTTT  17072
                     |||||..|.|.| |||.|  |||.||||||.||.||||||||||.|.|.
EMBOSS_002     16377 TTTAGACAGTAT-CTCCT--TCACAAAAATACACTTTGTTAAACAAGATC  16423

EMBOSS_001     17073 CA--TTATAATCAATTTCTCAAAAGTAAAGTTAATCAAGAGAAGGAAAAA  17120
                     .|  ||||||||||||||||||||.||.|||||    ||| |||||||||
EMBOSS_002     16424 AAGCTTATAATCAATTTCTCAAAATTATAGTTA----AGA-AAGGAAAAA  16468

EMBOSS_001     17121 AGGTTTTGTTTTGATTTGATTTGGAATGTGTATGTGTGTTTACTGTATTG  17170
                     ....|||||.||   |.||||....|.||| ||.|||||||...|||.|.
EMBOSS_002     16469 TATATTTGTATT---TGGATTATAGAAGTG-ATATGTGTTTGTAGTAATA  16514

EMBOSS_001     17171 AAATAGATTCTGTCTGAAAGACTGTATATAAGATAAAAAGTACAGAAGAG  17220
                     ||||...||..||  .||.|||..|||.|||.||..||..|..||.|.|.
EMBOSS_002     16515 AAATCAGTTTGGT--TAAGGACAATATTTAAAATTTAATCTGAAGGAAAA  16562

EMBOSS_001     17221 TAGTCAGAGAGTTATTACCCACCCCTGACTGATGGTGAATAGATTATCTA  17270
                     |..|.|.|.|.||..||.||||...|||.||..||||..|.|||||||.|
EMBOSS_002     16563 TCATAAAATATTTGCTATCCACATTTGATTGGAGGTGGGTGGATTATCAA  16612

EMBOSS_001     17271 AGTATCCCGTAAAAGGCACAACTCCTTCAGGTATATTTTACAAATTAATT  17320
                     ...|    ||...|..||.|..|...||..|.| ||.||||  ||...|.
EMBOSS_002     16613 TAAA----GTTCTAACCAAATTTGTGTCCTGAA-ATATTAC--ATCGGTA  16655

EMBOSS_001     17321 AGTAACTTTCTAGCCAAATTTGTGTCTTAAAGACACCAGCTAGAAC-TTG  17369
                     |.|.|.|.|.   ||..|||..|...|..|..|.|.|||.||..|| |||
EMBOSS_002     16656 ATTCAGTGTT---CCTTATTCTTACATATATAAGATCAGGTAATACATTG  16702

EMBOSS_001     17370 GTTAGTTCTAGCAAAG---AAGATTATTTTATTCTGAAACAGGTTTTTGT  17416
                     ||.|....|||..||.   ||.|....|||.|..||||..|||||||...|
EMBOSS_002     16703 GTCATCCTTAGGTAATTCAAACAAAGATTTGTCGTGAGGCAGGTTTGAAT  16752

EMBOSS_001     17417 T-GTCGTTTTACTTATTTGAACTTTTTTCTTGAATATGTATTTCTTTGCA  17465
                     | |||.|||||.||||||||   .||||.|||.|.|.. |.||.||...|
EMBOSS_002     16753 TTGTCATTTTAATTATTTGA---GTTTTATTGCACAAA-ACTTTTTCTAA  16798
```

```
EMBOSS_001     17466 CATAAAATATATTGACTTATGAATGTGATTAAAATGGAAAATAATTAGTT  17515
                     |||.||||||..||...||     ||||||||||  .||||||||||||
EMBOSS_002     16799 CATGAAATATTGTGTTCTA-----GTGATTAAAA--CAAAATAATTAGTT  16841

EMBOSS_001     17516 GATTTTAGAGAGACAGAGAGAGGAGAAGAGAAGTGTGAAGGAGAGAGGGA  17565
                     ||.||||              |||||||.|         ||.|||..
EMBOSS_002     16842 GACTTTA-------------AGGAGAATA---------AGCAGATG----  16865

EMBOSS_001     17566 GGATAGAAAGGAGAGAGGGAGAACAGGAAGGACAGAGGGAGAATGGGAAG  17615
                        |||.|||||          |||||||.| |||    |||||.
EMBOSS_002     16866 ---TAGGAAGGA----------ACAGGAAAG-CAG----AGAATA-----  16892

EMBOSS_001     17616 GAGAGGGAGAGAGAGAGACAGAGAGAGAGGAATGGAGTGGGTAATAAGCA  17665
                                           .||||||              ||||||.|
EMBOSS_002     16893 ----------------------TGAGAGA--------------AATAAGTA  16907

EMBOSS_001     17666 AGAGAAAAATGCCAATCATATGCTTTGCTAGTGTGTAAAGTCTGATAACC  17715
                     |   |||||    |||||.||....|..||.|...||.||||.|||||..
EMBOSS_002     16908 A---AAAAA----AATCACATAACATCTTATTCAATAGAGTCGGATAAAT  16950

EMBOSS_001     17716 CAAGGGAGAGAGGACTACTCTGGCCTAGTGAAACAAAGGAAAGAGAAATA  17765
                     |||||.|.||||.|.|..|.||||.|..|..|||    ||||.||.|.||
EMBOSS_002     16951 CAAGGCAAAGAGAAATGATTTGGCTTTCTAGAAC----GAAAAAGTAGTA  16996

EMBOSS_001     17766 TGGTAGAATATTCTCCTGGTGCTTCACCAAATGTGACACCAGAAGTCTGA  17815
                     .|||||||.|.|||.|||   |||
EMBOSS_002     16997 AGGTAGAAAAGTCTGCTG---CTT--------------------------  17017

EMBOSS_001     17816 CAGAAGTCATGTCAGCATTTGAGCTCCATAAAACTCAGGCTATCGACCTA  17865
                                 |||.|||||.|||||...||||.|..|..|.|..||.|||
EMBOSS_002     17018 -----------TCAACATTTCAGCTCATGAAACCATAAACAACTGAACTA  17056

EMBOSS_001     17866 CCATGTGAGAGTCTCAAAATGAGTTTAGGTAGGGGCAGAGGAGTTGAAAT  17915
                     ...|.|||||. ||.|||||.|..|||||||..|.|.|.|| |..|||||.
EMBOSS_002     17057 TTCTATGAGAA-CTTAAAATAATCTTAGGTGAGAGGATAG-ACATGAAAC  17104

EMBOSS_001     17916 CCAGTAACATATGCAACAGTGATCACACCAGGATTGCACATAGAAAGCAA  17965
                     |||.|||...||||||.|..|.|||||||..|...||||||.|.||||||.||
EMBOSS_002     17105 CCAATAATGCATGCACCCATCATCACAATAATGTTGCACCTGGAAAGGAA  17154

EMBOSS_001     17966 ATTAGTCCTCTAATAGAGACGCCAATTTGAAATTCACCCTCTGAGCAGGT  18015
                     ||.|  .||||.|||.|||.||||.|||||||.|.|..|||.|   |.||
EMBOSS_002     17155 ATGA--TCTCTGATAAAGATGCCAGTTTGAAAATTATTCTCAG---AAGT  17199

EMBOSS_001     18016 TTTTAAGCACAC---TCTTCTTTTACTTTTCT--ATTTACAAAAATGGAA  18060
                     .||.|||..||.   |||.||.||.||..||.  |||.|.|||||.|..||
EMBOSS_002     17200 CTTCAAGTTCAGGATTCTCCTCTTCCTCCTCCTCATTCATAAAAGTAAAA  17249

EMBOSS_001     18061 CACCACCAGAAAAACAAGAATTTGAAAG---ACGAGATGAG--AAAAGTA  18105
                     .|..|.|||.|||||||.|||...||.|   |..|||||.|.  ||||.||
EMBOSS_002     17250 GAAAATCAGCAAAACAAAAATCAAAACGTGTATAAGATAAACAAAAACTA  17299

EMBOSS_001     18106 AGTTGTAATTGGAAACAGACAGAATGTGTA-CACAAACACACACACAC-A  18153
                     ..|.|.||||..||..|||.|||   |||| ||.|.|.|.|...|.|. |
EMBOSS_002     17300 GTTAGGAATTCTAAGGAGAGAGA---TGTAGCAGAGATAGAGGAAGATGA  17346

EMBOSS_001     18154 CGCACACACACGT----GCATGCACAGGTGATGAGAG-AGTAGTTTGCC-  18197
                     .|.|...|.|.|.    |.|||...||..||.|||.| |||||...|..
EMBOSS_002     17347 AGAATGGAAAGGGTGAAGAATGGGAAGAGGAAGAGCGGAGTAGGGAGTAA  17396

EMBOSS_001     18198 -TACATGGTGTATCTGACTAAGAAGACTTTTTGCTCTGGTTGTCTT-ACA  18245
                      ||||..|||.||||||||.||||.....||||||...|.|..||| |..
EMBOSS_002     17397 TTACACAGTGAATCTGACTGAGAAATTCCTTTGCTACAGATCACTTCATC  17446

EMBOSS_001     18246 GGAAGTGACTAAATCTCATGATGTGAAATATTTTCTTGCATATTGTATTG  18295
                     .|||.||| |||.|     |.||||  |||.||||.|  .||.|.|||..
EMBOSS_002     17447 AGAAATGA-TAACT-----GCTGTG--ATACTTTCAT--TTACTATATGA  17486

EMBOSS_001     18296 GAAAAGAAAATAATTTTCCCAAACTCCTTAGGGGCAGTGTTGTCTTATAA  18345
                     .|...||||..|||.|..||||||||.|| |..|||.||||||.||.|..
EMBOSS_002     17487 CACTGGAAATAAATGTGTCCAAACTCATT-GATGCAATGTTGTATTCTTG  17535
```

```
EMBOSS_001     18346 TTCCCATATAGTATA-----TGCT---CTTCAAGTAAGTAACTCCAGAGT  18387
                     ..|||.|.||.||||     ||.|   |||.||| ||||||..|.|..||
EMBOSS_002     17536 CCCCCGTTTAATATAACCACTGATAACCTTTAAG-AAGTAATGCTAACGT  17584

EMBOSS_001     18388 TGAGTAAGACAAGACTCGTGACTCAGATGGCATGCTCTGCTCCCTAGACT  18437
                     ||.|||  |||.|||.|.|||.||......||..|||||..||||||..|||
EMBOSS_002     17585 TGGGTA--ACATGACACTTGAATCCAGGAACAATCTCTTGTCCCTCTACT  17632

EMBOSS_001     18438 AGACATTGCATCAGTCTGCCTATACTCACATCCGCTGTTAAAGGATTGCC  18487
                     .||.|.|.| |.||.|            |||..|...|||||.|||||..|
EMBOSS_002     17633 CGAAAGTAC-TTAGAC------------CATTTGTAATTAAAAGATTTTC  17669

EMBOSS_001     18488 TCCAGTAAAATATGTCTTTTAATTCCTTATACAAGAATCTGGAAAAAAAA  18537
                     ||||||||||||||..|||||||.||||||||.|.||.|.|.....|.|
EMBOSS_002     17670 TCCAGTAAAATATGTGCTTTAATTGCTTATACAGGTATGTAGGGTGGAGA  17719

EMBOSS_001     18538 AGTAAGATTCTCTATTTCTTAAATTTAGCAGCAGGTTAATCACTGATAAC  18587
                     .|..||      ||||            ||.|.||| |||
EMBOSS_002     17720 GGGGAG------TATT------------CACCCGGT-AAT----------  17740

EMBOSS_001     18588 AATAAAAATACATAACAATCATCTAGCACGGGTAAATATTGTGGCAAAAA  18637
                      |||||..|||        ||.|.||||.|| ||||||||.||.|||.||
EMBOSS_002     17741 -ATAAACCTAC--------CACCAAGCATGG-TAAATATTATGCCAATAA  17780

EMBOSS_001     18638 TTACACCCTGAAGAATTCAGTCAAAGATATAAGTAAGTACACATCATTGT  18687
                     .|..||    |||  |.|||                  |||..||||..|
EMBOSS_002     17781 ATGAAC----AAG--TGCAG------------------ACATGTCATAAT  17806

EMBOSS_001     18688 CATGTTCCACAATATATCATCTGCTTTAAAGAAACTGTTATGTAGCTGTA  18737
                     |.||||.||.|||||||.||.|||||||.||||||||.||..||..|||.
EMBOSS_002     17807 CCTGTTTCATAATATATTATTTGCTTTAGAGAAACTGCTACATACTTGTG  17856

EMBOSS_001     18738 GTAGATTTAATCATTAATCCCATTTCTTCTCCACCTTCTGCAATCACAAC  18787
                     ..||.||.|.|.|.|||.||.||||||||.|....|.||.||.||||||.
EMBOSS_002     17857 AAAGCTTCAGTTACTAACCCTATTTCTTCCCTCATTCCTACAGTCACAAT  17906

EMBOSS_001     18788 CTTAACAATGCCT-CCTTATGAGTGGAATGTACTTCCCAACCCCTAGTCT  18836
                     .||||||.|.|.| |||.||.|||.||.||.|.||.||||.||||.|.||
EMBOSS_002     17907 TTTAACATTACTTTCCTCATCAGTAGAGTGAATTTTCCAAACCCTGGCCT  17956

EMBOSS_001     18837 TAGGGGTTGGCCATGTGATTTGCTTTAGCAAATGGTAAATGAGCAGGAGT  18886
                     |.|...|||..|..||..||||||||||||.|||  |.|||.|.|. ||.
EMBOSS_002     17957 TGGCCATTGATCTGGTATTTTGCTTTAGCATATG--ACATGTGAAT-AGC  18003

EMBOSS_001     18887 GAGAGGTGACAGTTTTCAGCCTAGGCCTTAAGAGATCTATACATTCCTGT  18936
                     .|||.||||||||||   |||.  |..|||||.   ||.|    ||.|.|
EMBOSS_002     18004 TAGAAGTGACAGTTT---GCCA--GTTTTAAGC---CTCT----TCTTCT  18041

EMBOSS_001     18937 TTGTGCTTCTGCTATCATTCTGAGAACACGTCCATCTAGGCTGCTGGTCT  18986
                     |||.|||||||..||...|.||||||.|.|....||...|..||||.|.|
EMBOSS_002     18042 TTGGGCTTCTGGCATATGTATGAGAAGAAGATGCTCACTGAGGCTGTTTT  18091

EMBOSS_001     18987 CAGGAAAACGATAAAAGACATGAACAGCAGGGCTGCACTAGCCATTCACA  19036
                     ||||||||.|||.|.|||.|||.|..|||..|||..|.|||...|.||||
EMBOSS_002     18092 CAGGAAAAGGATGAGAGAGATGTAATGCAAAGCTATAGTAGATGTGCACA  18141

EMBOSS_001     19037 TCCAGGAAAAGAAATGATTGTTGCATAAAGCCATTGAGCTTTATTCTACA  19086
                     |..||||||.|   |.|..||       |||.|||..|.|.|.|||...||
EMBOSS_002     18142 TTTAGGAATA---AGGTCTG-------AAGTCATCCAACCTCATTTCTCA  18181

EMBOSS_001     19087 TTACTGTGACAATAGCTAATTGAAATAGTAAATATACTTTGGTTTTTCCT  19136
                     |||...|||.||.||..|.||||||..||.||||||.|       ||.||
EMBOSS_002     18182 TTATGTTGAGAACAGTGACTTGAAGCAGAAAATAAA-------TTGTC--  18222

EMBOSS_001     19137 AAATGCATATTGAAAATTAATAATATTAGCCATCTGTATGATAAAAATAT  19186
                     ||.|||                           ||.|.|..|||||.|.|
EMBOSS_002     18223 AACTGC---------------------------CTTTGTCCTAAAATTCT  18245

EMBOSS_001     19187 AAAGCCTATGTTTTATTTT-----TTAATGGTTCACTGCCCTAAATAAAT  19231
                     .||||||||.|||||||||     |||| ||.||| |||.||   |||||
EMBOSS_002     18246 GAAGCCTATATTTTATTTTAGTTGTTAA-GGATCA-TGCACT---TAAAT  18290

EMBOSS_001     19232 TTCCAAAAAGTAGATGTTCCCTTGTCTAGTGATGTCATTATATTTTATTT  19281
```

```
                        |.||.|||||||| |||||   .||||||   ||||.|||.|.|||
EMBOSS_002     18291 -----ATAACTAGATGT-CCCTT---CAGTGAT---ATTACATTGTGTTT  18328

EMBOSS_001     19282 ATACATCATAAACACACTGTTTATTTCTGCTCATTTTTTTGTAAGTAACA  19331
                     .||..|. |.||...|| |||.||.||||.|.|.|..||     |||.||
EMBOSS_002     18329 GTATGTT-TGAATGTAC-GTTAATCTCTGTTAAATCATT-----GTAGCA  18371

EMBOSS_001     19332 TGTG--TTACCGCCAATCTTGAGATGATACA----------CACACTTCT  19369
                     |.|.  ||||.|||||.||||...|||||||          ||||.||||
EMBOSS_002     18372 TATTACTTACTGCCAAACTTGCATTGATACATAGGTTTTCACACATTTCT  18421

EMBOSS_001     19370 GTACTAAATTTT--GGAAAACA-TATTAGCTACCCACTCCTTATATCAAA  19416
                     |.|||||.|.||  |||||||| |.|.||||...|||||||||||||.|
EMBOSS_002     18422 GGACTAAGTGTTAGGGAAAACAATCTCAGCTCTTCACTCCTTATATCACA  18471

EMBOSS_001     19417 ATATTGCCTAATAATGTGTTTTGTTTTAATCCTTCATGAATTTCCAGGAG  19466
                     ..|.|||...|.||||||.||..|||||||.|||||||..|||.|||||||
EMBOSS_002     18472 GCACTGCACGAGAATGTGCTTCATTTTAATGCTTCATACATTGCCAGGAG  18521

EMBOSS_001     19467 AACTGAACTGATACTTGGGTT-TGT-GAGATATATGAAAATAGTGAACAT  19514
                     .|....||||.||||||..|| ||| |.||||||  |||..||||.|.|.
EMBOSS_002     18522 CATCTGACTGGTACTTGTTTTATGTTGCGATATA--AAATCAGTGCATAC  18569

EMBOSS_001     19515 GAACTTCTGGTTTAACCCTTGTGATGATAATGGAATCATAGCTCTGTTAA  19564
                     ..||...|||.||||.|.|..||||| |.||..||.|..|      ||||
EMBOSS_002     18570 TGACACATGGCTTAAACTTCATGATG-TCATATAACCTCA------TTAA  18612

EMBOSS_001     19565 TTACTCTTGTGGTTTGTCTTCCTAGAGATAATCATGTACAAAATTCCTTT  19614
                     ||.|.||||||..|||||||||.|||||||                ||||||
EMBOSS_002     18613 TTGCCCTTGTAATTTGTCTTCATAGAGAT---------------TCCTTT  18647

EMBOSS_001     19615 CCAATTTGTTATATAATATTAGAAATACTTCCAAAATTGGCATGGATTTA  19664
                     .||.|.|||.||||||||||||||||..|.....||||||||||.|||..
EMBOSS_002     18648 TCATTCTGTAATATAATATTAGAAATGATATTTGAATTGGCATGCATTAG  18697

EMBOSS_001     19665 TTGTTATCATTTGTTGGCACAATCATTAAAACGAAACCCATAAAGCTAGA  19714
                     ||||.|..||||  ||.||.||    |||||..|||..||||.|.|.|.|
EMBOSS_002     18698 TTGTAACTATTT--TGACAAAA----TAAAATTAAATGCATAGATCCAAA  18741

EMBOSS_001     19715 TAATTAAATGTTTACAAAGCTATAGTACTCAAAACAAAAACACTGTGAAA  19764
                     ||.||||||..|||.||||..|||.|...||.|  |||||.|||||||..
EMBOSS_002     18742 TACTTAAATCGTTATAAAGTAATATTGTGCAGA--AAAAATACTGTGATC  18789

EMBOSS_001     19765 AGAGATTTTTTAAATAATAGTTTTTGCATGCCTTTTGAATAATTGGATTA  19814
                     |          |.|.||||..|||..|||||||||.|||||||||||||
EMBOSS_002     18790 A----------AGACAATATATTTCACATGCCTTTAGAATAATTGGATTA  18829

EMBOSS_001     19815 TTCTGAATTTCTTCATGTTTAGTCCCTGAATCTAAGTCATACCGTCTACA  19864
                     |||.|.||||..||.|.|||.||...|||||.||.|        ||||..
EMBOSS_002     18830 TTCGGTATTTTCTCCTCTTTGGTTTTTGAATATATG--------TCTATG  18871

EMBOSS_001     19865 TAAAAATAGATGTCAGCTGAAGAAAACCAGGCAA-TGGATTTGTCTTGAC  19913
                     ||||||||.|||||..||||||||.||..||..|| |.||||||.||||||.
EMBOSS_002     18872 TAAAAATACATGTTGGCTGAAGATAAGTAGTAAAGTAGATTTGCCTTGAA  18921

EMBOSS_001     19914 GACAATCTTTT--TATATGTTCAGACTTCATTTAACATTAGACTTGTCTG  19961
                     .|||||..||||  |||||.||..||||||||.|||.|||||.|  |||||.
EMBOSS_002     18922 AACAACATTTTGATATATATTTTGACTTCATCTAATATTATA--TGTCTT  18969

EMBOSS_001     19962 TA--TTTGAAATTGGTA-TTTCTTTACATTTCTGAATTTAGGGAAATGGC  20008
                     .|  |||||||..||.| ||||||.     ||||||||||.||||||||.|
EMBOSS_002     18970 CAGTTTTGAAAGGGGGAATTTCTTC-----TCTGAATTTATGGAAATGAC  19014

EMBOSS_001     20009 ACAAGAGAATAACATTAATTTCCTCTG-CATTTTGGCCTAATCAAATTTG  20057
                     ||||||||||||.||||.||||.||||| ||..||||.|.|||.||||...
EMBOSS_002     19015 ACAAGAGAATAGCATTGATTTGCTCTGTCACATTGGTCAAATGAAATAAA  19064

EMBOSS_001     20058 AGCCTTTCAAGAGACACAGCCAAGTCAATTCAAAGAGACATATGAAAAGA  20107
                     |     |.||.|||||....|...|.||||||  ||.||||.||.||.|.
EMBOSS_002     19065 A-----TAAAAAGACATTATCTCATTAATTCA--GACACATTTGTAATGG  19107

EMBOSS_001     20108 CTACTGTTAATGTATCTTTAAAATGAATTAGCGGCATGAACTGTTGCTAG  20157
                     .|..||||||..||.|.||||..|.||.||..|.|||.|||.||||.|.|.
```

```
EMBOSS_002     19108 TTCTTGTTAACATAGCCTTAAGGTTAACTAATGACATAAACCGTTACCAA  19157

EMBOSS_001     20158 GTGAGTTAGGTATAGTTGTAGTTTTTAGTAACCCTAAGAGAAGATGCAGT  20207
                     ||||..||.||..|..|.||.||||.|.|||....||.||||||.|..|..
EMBOSS_002     19158 GTGAACTAAGTCCAAATATAATTTTCACTAATTGCAAAAGAAGGTAAAAA  19207

EMBOSS_001     20208 GCATTCTAAAATGTCACAAGGAGTTTGATTGCTCAAAATTCTGGGAGATT  20257
                     |||||..|.|||.|||.|   |||.|||||...||||||..... |||||
EMBOSS_002     19208 GCATTTGAGAATATCATA---AGTATGATTTAACAAAATGACAA-AGATT  19253

EMBOSS_001     20258 GGCTCTCTGCAAGGCTTCTTGATGTCATTGTTCCTAGAGGAATGTTGTTC  20307
                     |....||||||..|..|.|||||.|||.|.       |.|||..||..||
EMBOSS_002     19254 GATCATCTGCACAGTGTGTTGATTTCAATA-------AAGAAGCTTTCTC  19296

EMBOSS_001     20308 CAGTACCTATAGCGATTGCAGCCATAAC-----------------TATTT  20340
                     ||..|.|||||||.||.|.|||..|.||                 ||.||
EMBOSS_002     19297 CAAAATCTATAGCAATAGTAGCACTGACATTTTCTTTTCCTAGTTTAATT  19346

EMBOSS_001     20341 ATGTGTCATTGTAGCCATTGTTATTACTACATGCT---------TCA---  20378
                     .|.||...||.|.|..|.|.||||.|.|...||||         |||
EMBOSS_002     19347 TTTTGGTTTTCTTGAGAATTTTATAAGTGAGTGCTGAATTTACATCATTA  19396

EMBOSS_001     20379 ------------------------------CATACCT-CTACTGAGGTCT  20397
                                                   |||.||| ||.||..|..||
EMBOSS_002     19397 CTAGTACTCTCTCTCCCCTTCCAACTCCCCCATCCCTGCTTCTATGCCCT  19446

EMBOSS_001     20398 AAAGAATTAGTGGACT-------------TCATATTCTGGAGAGA-----  20429
                     .||.|||| |..||||             |.||||.||..|.|||
EMBOSS_002     19447 TAAAAATT-GCTGACTGTCGGTATTTTTGTTATATACTACACAGACACTT  19495

EMBOSS_001     20430 --ACACTTGAAGA---AC-----------------------CAAACAGA-  20450
                       ||||.|||||.|   ||                       ||.|||||
EMBOSS_002     19496 ACACACCTGAACACAAACACACATACACACACACACACACACACACAGAG  19545

EMBOSS_001     20451 AGTTTGAT-GTGAATCTGC-----------------ATATCCACCATTAT  20482
                     ||...||. |.||....|.                 |.|...|.....|.
EMBOSS_002     19546 AGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG  19595

EMBOSS_001     20483 TGTTCATAG----GTTCTCAGGATTAGTTGAGTG---ATGCC--------  20517
                     |.|...|||    |.|||.|||..||..||..||   |||.|
EMBOSS_002     19596 TCTCTTTAGTGTTGCTCTTAGGTATATGTGCTTGGGGATGACCACTTGAA  19645

EMBOSS_001     20518 -TTAAAGAAA--------GAA--AGTCAGATGA------TAGGTCTT---  20547
                      ||..|.|||        |||  ||||....||      ||..||||
EMBOSS_002     19646 ATTGGATAAACTCTCAGGGAACTAGTCCCTGGACAAAACTAATTCTTATT  19695

EMBOSS_001     20548 -----CCTGCTGC-CCGCACCACATCATGAG------TGTTATTCCTATA  20585
                           |||||||| ||.||.|...|..||||      ||..||||||||.
EMBOSS_002     19696 GATTTCCTGCTGCTCCTCATCTAGTGGTGAGGCCTTGTGAAATTCCTATT  19745

EMBOSS_001     20586 GAGGAG--GAGTAAAGAG-TGGGAAG--------AAAATGAAATCT----  20620
                     ...|..  |.||.|..|| |.||..|        |..|||.|.|||
EMBOSS_002     19746 TGTGTTTTGTGTTAGCAGGTTGGCTGCTGTTGTCATTATGCATTCTTGCT  19795

EMBOSS_001     20621 -----GTCAATACTGTGAATATATAAA-------------------TAAT  20646
                           |.||||||||||..|.||.|.|.                   ||||
EMBOSS_002     19796 TAAGTGACAATACTGTTGAGATTTCATGTGTGAGTTTCAATGCCTTTAAT  19845

EMBOSS_001     20647 AAAAGTAGCAGTAGGACTGATTAATTCTGAATC--ATCTTTATGAAATGA  20694
                     ||||..|..|....||..||...||.|||..||  |...||.||.|||..
EMBOSS_002     19846 AAAAACACTACATAGAAGGAGGTATCCTGGTTCTCAGGCTTTTGTAATCT  19895

EMBOSS_001     20695 CTGGAG--CCGTGAAAATGCTCAGTCTGCACAGCTGATTGAGAAATGTAT  20742
                     .|..|.  ||.||...|||.|  ||.|.|..||||  ||   ||||||..
EMBOSS_002     19896 TTCCATTCCCTTGTTCATGGT--GTTTCCTGAGCT--TT---AAATGTGA  19938

EMBOSS_001     20743 GCAATCTGTTGATCGGAATTTATTTG--TGAATGCTCTCTTCCAGAGATT  20790
                     |.|.|.|||||.....|..|.|||||  |||.|.|.|..| |||..|.|.
EMBOSS_002     19939 GTATTGTGTTGTAGATATATCATTTGGCTGAGTACCCAGT-CCACTGTTC  19987

EMBOSS_001     20791 TATA----TACCAGAGTTCTTAAAACGAATTTTGTCCCCATGAAAAGAAA  20836
                     |.|.    |.|...|||||.||......|.|..|.|||...||...|||||
EMBOSS_002     19988 TCTGGTTTTTCACTAGTTGTTTTTTGTATTGGTCTCCATCTGCTGAGAAA  20037
```

```
EMBOSS_001     20837 ACTACAGATCTGTAAGACTGCAATTTAAAATGGAAGAAAACATGTTCCCA  20886
                     | ||.|.|||||.|.   |||..||.|.|    |.|..|.||||.||.|..
EMBOSS_002     20038 A-TAAACATCTTTG---CTGAGATGTGA----GTACTATACATTTTTCTT  20079

EMBOSS_001     20887 CTTGAAGAACAACTTTCAAACAAACAACTGATACAAAAAAGT-CAAAAGC  20935
                     .|..|......|.|||..|....|||....|.|.||||.|.| ||....|
EMBOSS_002     20080 TTACAGTTTTTATTTTTGAGTTTACACTATAAATAAAACATTTCATCTTC  20129

EMBOSS_001     20936 TGT---TTTGTTTTATATAATAGTTTCAGAATA------CTTCCAGTCAA  20976
                     |..   |||..||..|..||...||.|..||||      |||.|..|||.
EMBOSS_002     20130 TTCCCCTTTCCTTCTTCCAAGCCTTCCCCAATAACTCTCCTTGCTCTCAT  20179

EMBOSS_001     20977 T-ATATACCT----TGGTTTGGTGAAAAAATAAAAAGCTAAATCCTTAGA  21021
                     | |.||.|.|    ||.|.|..||.||..|||.|....|..|||..|..|
EMBOSS_002     20180 TCAAATTCTTGCAATGATATCTTGCAATGATATATCTATCTATCTATCTA  20229

EMBOSS_001     21022 TC-AT-TAACTAGAAATTTTTGTAAA-ATAAATAAAAGCCGTGGGTTTTA  21068
                     || || ||.|||...||.|.|.||.. ||..||..|. ||.|  .|.|..
EMBOSS_002     20230 TCTATCTATCTATCTATCTATCTATCTATCTATCTAT-CCAT--CTATCC  20276

EMBOSS_001     21069 GTGCAGTGATCC---CATGAAG--AGGAATATATTCACC-ATTGGTCTCT  21112
                     .|.||...||||   |||..|.  |...||.|||..|.| ||...|||.|
EMBOSS_002     20277 ATCCATCCATCCATCCATCCATCCATCCATCTATCTATCTATCTATCTAT  20326

EMBOSS_001     21113 TAATCTCAGATAGAATGTACATGTTACTTTATTTTATAACGAA--AGCAA  21160
                     ..||||........||.||..|.|..|.. ||.||||||.||.  |||||
EMBOSS_002     20327 CTATCTATCTATCTATCTATCTATCCCAA-ATGTTATAAAGACTGAGCAA  20375

EMBOSS_001     21161 CTGTGTTGTG---ATAT---TATGTATA-ATAT---TATAACAGG---AG  21197
                     |....||.||   ||||   ||..|||| |.||   ||.|||||.   ||
EMBOSS_001     20376 CACATTTCTGTGGATATAAGTACATATATAAATACATAGAACAGAGATAG  20425

EMBOSS_001     21198 AAGTCCT-----CTTAGCTAACTC--AGTAATCAATAACATTGTACGTTG  21240
                     ||....|     |||||..||.|.  ||||.|.........|.||..||.
EMBOSS_002     20426 AAAGTATAATGACTTAGGAAAATGGTAGTAGTAGGCTCTCCTCTAGATTC  20475

EMBOSS_001     21241 TGTGTTATTGTAA-CCAAA-AACTATGACAGAACCCCATTTCATAAGATC  21288
                     ..||...|..|.| |||.| ||...||..|||....||.|...|...||.
EMBOSS_002     20476 CATGGCCTCATCAGCCACAGAAAGTTGCTAGATTTACAGTAACTGGCATA  20525

EMBOSS_001     21289 AGTTTATCCACCTATATGATT-TATATTTGAATATTCATTTCAGTACTTA  21337
                     ..|||.|.|  |||. ||||| ||..||    |..||.|.||| ||...|
EMBOSS_002     20526 CATTTCTTC--CTAC-TGATTGTAGCTT----TTGTCCTATCA-TATGGA  20567

EMBOSS_001     21338 TGTTGCTTAAACA-AAGCTACTG-----TATTAGTCCAT--TTTCA--TA  21377
                     |||||.|||..|. |||.||.|.     ||.||.|.||.  ||.||  ||
EMBOSS_002     20568 TGTTGGTTACTCCCAAGATAATAGTGACTACTACTACAGCCTTGCAAATA  20617

EMBOSS_001     21378 C--TGCTATAAAGAACTGCCCGAGACTGGGTAATTTCTAAAGGAAAGAGG  21425
                     |  |||.||.|.|..|||....||..|  |||..||.||.|...|||.|.
EMBOSS_002     20618 CCTTGCCATGATGTTCTGTTATAGTTT--GTAGGTTTTACATCTAAGTGA  20665

EMBOSS_001     21426 TTTAATTGACTCACAGTTCCACATGGC----TGGGTAGGCCT--CAG---  21466
                     ...|||||..|.....|..||..||||    ||.||||..||  |||
EMBOSS_002     20666 GACAATTGGGTGTTTTTCTCATTTGGCAGTTTGAGTAGTACTTACAGATA  20715

EMBOSS_001     21467 ----GAAA---CTTACAA-TCATGGCAGAAGGTGAAGGGGAAGCAAGCAT  21508
                         ||||   |||.||. ||||..|  .||.|.||||.....|..|.|
EMBOSS_002     20716 CTATGAAAAGGCTTCCAGGTCATTTC--CAGCTTAAGGTTTCCAAGTCCT  20763

EMBOSS_001     21509 CT-TCTTCA-CA-AGGCCGCAGGAAGGAGAAGCGCCCAGCG---AAGTA-  21551
                     || |||..| || |.||.|.....||.|..||.|.|...|.   |||||
EMBOSS_002     20764 CTGTCTGAAGCATATGCTGTTTTCAGCAATAGTGACTTACTTTCAAGTAC  20813

EMBOSS_001     21552 --GGAAGAGCCCCTTATAAAACCATCAGATCCCGCT----ATCAT----G  21591
                       |||.|...||...|..||....|..||...|.||    |||..    |
EMBOSS_002     20814 TGGGAGGCAACCAAGAGCAATGTTTGTGACATCTCTTGGAATCCCACTGG  20863

EMBOSS_001     21592 AGAACAGCATGGGAGAAACTGCCCTTATGATTC---CATTACCTCCACCT  21638
                     ..||||.|.||...||||.|..| |.|||.||.   |...|..|.....|
EMBOSS_002     20864 CCAACAACTTGAAGGAAAGTTTC-TCATGCTTTGTACTAGAGATTTTATT  20912
```

```
EMBOSS_001     21639 GGTCTCTCCCT--TGACACGTGG-GGATTATGGA-----GGTTAT-----  21675
                     .||||.|.|.|  |..||..|.. |||||.||.|     |.||||
EMBOSS_002     20913 AGTCTATGCATAATTCCATTTAAAGGATTTTGTAACTAAGCTTATACAAT  20962

EMBOSS_001     21676 ----------GGGGATTACAATTTAAGATG-----AGATT--------GT  21702
                               |.|.|||..||.|....|||     |||||        |.
EMBOSS_002     20963 AATGTCTTGAGAGCATTCTAACTGGGAATGGGGAGAGATTACAATACTGA  21012

EMBOSS_001     21703 GGGGTGGGGACACAGCCAAGCCATACCAAAAACTCTGTTTT---------  21743
                     |||...||.|.|..|..||...|||.|||.||...|.|||.
EMBOSS_002     21013 GGGAAAGGAAAAGGGAAAATAAATAACAATAAGGATTTTTGAAAAAGCTC  21062

EMBOSS_001     21744 -----TTGTTTTTGTT--TAATGGA---AATGATTTAGAACTTTATTT--  21781
                          ||.|||||.||  |..||.|   ||.|.|..|...||.|||||
EMBOSS_002     21063 AGACCTTATTTTTATTACTGCTGTAGTCAAAGCTACAACTCTGTATTTCA  21112

EMBOSS_001     21782 TCTGATGTTTCTTTTTCATAAAACCACGACACCAAAATCTACTT----TT  21827
                     |.|.|...||||||.|.||..|......|..|.||||..||..|    ||
EMBOSS_002     21113 TATCAATATTCTTTGTGATCTACTATATAGTCTAAAACGTAAATAAACTT  21162

EMBOSS_001     21828 CACTGCTCCATTCAACTAGT---AGAGAATA---------TCTAATCTCT  21865
                     .|.|..||.|...||..|..   ||||||.|         ||...|...|
EMBOSS_002     21163 TAATTTTCTAGAGAAGGACCTTAAGAGAACAAAGGAAGGTTCGTGTGAAT  21212

EMBOSS_001     21866 TCTCAAGTATTTCT---TTCTCAATTATGGTGGTTT----TAGCTAAGAA  21908
                     |..||.||..|.||   ||.||..||.||||..|.|    |||.||||.|
EMBOSS_002     21213 TTGCATGTTATGCTACCTTATCCGTTTTGGTTATCTGGATTAGTTAAGTA  21262

EMBOSS_001     21909 CAG-CTTATGGCATGCTTTTC---TAAATAATATTAGAACACATAAATTA  21954
                     ..| |.||..|.|.....|||   |.||.|.|.||....|.|.....|..
EMBOSS_002     21263 ATGTCGTAAAGGAAAGGGTTCACATGAAAAGTCTTGAGTCTCCCCTGTCC  21312

EMBOSS_001     21955 TCTGTACCTGGTATTACCACATTC--------------ATTGCTCATTTT  21990
                     ..|.|....|||.|||..||..|.              |.||...|....
EMBOSS_002     21313 CATATCATGGGTTTTAGTACTATAGAAGAGAAGAAGAGAGTGGGAAGAAA  21362

EMBOSS_001     21991 AAGATCTCAATTGATACATTCAATTCATATATATTTAAAAT---------  22031
                     |.||..||.....||||..|.|||..|||.|||.|.|||.|
EMBOSS_002     21363 ATGAAATCTGCCAATACTATAAATATATAAATAATAAAAGTGGCAAGGGG  21412

EMBOSS_001     22032 --TGATTCATTTAGAGCAAGAGA-TACAGGCATTTTAATGTATTACACTG  22078
                       |||.||||.|.||..|||.|| |..||.|.|...||||. |.|..|||
EMBOSS_002     21413 ACTGAATCATCTTGATAAAGTGACTGGAGCCCTGAAAATGC-TCAATCTG  21461

EMBOSS_001     22079 CTACTAAAGCTTAGCAAAT-TATTCTTTTTTGTGCCCACAAATTA-TCAT  22126
                     | ||....|.||.|.|.|| |||.|..|.|..||..||.||.|.| |.||
EMBOSS_002     21462 C-ACAGCTGATTGGGAGATATATGCAATCTGTTGATCAGAATTCACTTAT  21510

EMBOSS_001     22127 CCATTCA-TGTCCTAAAAA----TAAAATTGAATTTATTATACTTTCCCA  22171
                     ..||.|. |.|.|.|.|||    |..|||.||||||...|.|..||
EMBOSS_002     21511 GAATGCTCTCTTCCAGAAACTGGTTTAATGGAATTTTCAAAATATT----  21556

EMBOSS_001     22172 TTTATCCA---AAAAAAAGGTTTTTTTTAACAATTGATGCAGATACAC--  22216
                     |||.||||   ||||||||         |||.|..|| |||||.||||
EMBOSS_002     21557 TTTCTCCACGTAAAAAAAG---------AACTACAGA-GCAGAGACACCA  21596

EMBOSS_001     22217 ATTTTCAAGCTAAAAA---TATGTGTGAAAGT---GGCCTCTTTC--TCA  22258
                     |||||.|||..|||||   |.|.|....|||.   |..|..||||  |||
EMBOSS_002     21597 ATTTTAAAGGGAAAAAAGCTTTTTCCCTAAGAAAAGAACAGTTTCCATCA  21646

EMBOSS_001     22259 TAGTATTTAT--------TTTAGGAGTCTAGCAATAATTTTTCTTAGGTT  22300
                     .|.|||||||        |||..|.|.||..|...||..|. |.|..|.|
EMBOSS_002     21647 CATTATTTATCTATTTATTTTGTGTGCCTGTCTGGAAGGTG-CATGTGCT  21695

EMBOSS_001     22301 ATCA-GCACATGTCTTAGCCTGA--ATTATTTGAATTCAGTCTGTGTCTT  22347
                     ||.| |.||||||.....|.||  |....|.|.|...|||.| |.||||
EMBOSS_002     21696 ATAATGTACATGTGGAGAGCAGAGGACATCTAGCAGCAAGTGT-TCTCTT  21744

EMBOSS_001     22348 CAAGTTCAGATGGTTATGTGATCTTGTTAAGATCTCAAAGTAGTGGGAAT  22397
                     |.|  .||..|||.||...    |.|.|.|.|| |||..||.||.||..|
EMBOSS_002     21745 CTA--CCATTTGGGTACCA----TGGATTAAAT-TCATGGTGGTAGGCTT  21787

EMBOSS_001     22398 GATGGAGTATACAACAACCTCATTGTTTTTTATGGCAACTGTCATTTACT  22447
```

```
                     ..|||      |||..||||  |....|..|..|.||||.  ||.||...
EMBOSS_002     21788 TGTGG------CAAGCACCT--TCACCTGCTGAGCCAACC--CACTTGGC  21827

EMBOSS_001     22448 GAAGGACATAAGGCTAGCAGAACATGGTCAGAGAAGGAATCAAAG-----  22492
                     .||||||.|.....|.|.||.|..|..|..|.| ||.|.|||||||
EMBOSS_002     21828 CAAGGAAAACTTTCAAACATATGACTGATACA-AATGTATCAAAGGTGCT  21876

EMBOSS_001     22493 TTTGGT-CAGCCAACTCTGCTCCACAGCTACAAGCTGCTA--------GA  22533
                     |||||| ||...||...|.||...|....|.||.|||.||        ||
EMBOSS_002     21877 TTTGGTTCATAGAATAGTTCTGACCTATAAAAACCTGTTATTTGTTGAGA  21926

EMBOSS_001     22534 CAGGCATAAATTTTTCCAAACCTACACAA-AGGGACTTAGGGCCCTTGGC  22582
                     .|..|.||||.|.|.....|.|||..||. |...|.||||.   .|....
EMBOSS_002     21927 AACACTTAAAATATAAAGTATCTAATCATTAACTATTTAGT---ATAAAA  21973

EMBOSS_001     22583 TGAGAGCGACATTCTAACCACTTCCTTATTTATGGCTGGTGGGGTTTGTA  22632
                     |||..||..||||   |....||..|.|..|||..|||....|.....||
EMBOSS_002     21974 TGATGGCATCATT---AAGTTTTAGTAAAGTATTTCTGTGAAGAAAAATA  22020

EMBOSS_001     22633 CATTTTCT-CATTTCTGTATA--ACATTTCTTGACTGTAATAAGCA---A  22676
                     .|||..|. ||...|.|||||  |||.......||....|.|..||   |
EMBOSS_002     22021 TATTCCCCACAAAACAGTATATTACACACACACACACACACACACACA  22070

EMBOSS_001     22677 TGTATTCATTCTGCTTTACCACTTTCACTA-ACCT-TAACCTCAATATAT  22724
                     ...|..|||.|...|.||....|.|...|| |..| ||.|||..||.|.|
EMBOSS_002     22071 CACACACATACATATATATATATATATATATATATATATCCTGTATTTTT  22120

EMBOSS_001     22725 ACTCAATTAAGCAATTGAAAACAGCAGTTTTAATCTTTTGACATAAATGA  22774
                     ..|.|...||..|||||..||..|.. ||.|.|.|.|.|...|||||..||
EMBOSS_002     22121 CATAATGAAATGAATTTGAATTATG-GTATGATTTTATATCCATAAGAGA  22169

EMBOSS_001     22775 TTTCCTCCGAAGCAAAATGCTGGAAATCCCCTCAAATGCACCTTTTATTG  22824
                     ..||||...|.|.||...|.|....|.....|....||||...|.||||.
EMBOSS_002     22170 AGTCCTTTTAGGTAACTAGTTAATCAGTAATTGTTTTGCATTGTATATTT  22219

EMBOSS_001     22825 ATGAATACCTATAAGCACCACCTACAGTCGCTGGAGGCTGACAGGAACCA  22874
                     .|...|...|...||.|.| ||.||.|.|...|.||..|..||..|...|
EMBOSS_002     22220 GTTCTTGTTTTACAGAAAC-CCAACTGACAAAGTAGTTTATCAATATTAA  22268

EMBOSS_001     22875 AACTTGATGATAACCACTGAG--CTGAGAATTTTCAACTC----------  22912
                     ||.|...||.||..|..|.|.  |||. ||||||||..||
EMBOSS_002     22269 AAGTCATTGTTATTCCATAAATACTGT-AATTTTCAGGTCCTTTAACATA  22317

EMBOSS_001     22913 ACTCTTTTTCCCTGTATGGTTCTTCTAGCTGCATTATTTCCCACTA-TTT  22961
                     |.|...|.|.|.||..|.|||.||..|..||||| |||||..|..| |||
EMBOSS_002     22318 AATTAATGTGCATGCTTTGTTTTTTAAAATGCAT-ATTTCTTAAAACTTT  22366

EMBOSS_001     22962 A---AAGCTACA-GCTGGTGAACT-ATTCA-AATATTTAAACTTTGGAGA  23005
                     |   |||.|.|. |.|..|.||.| |.||. ||.|.|.|.|||||  |..
EMBOSS_002     22367 ACTCAAGTTTCTTGTTTATAAAATCACTCTCAACACTAAGACTTT--ACT  22414

EMBOSS_001     23006 AGAAAATATCAACTTATCACAACCCTCTTTTTATATTCTAAATTCATATA  23055
                     ...|.||.| |||||...|.||...|...|.|||.|||...|.|..|.|.
EMBOSS_002     22415 TTTACATGT-AACTTGGTAGAATATTTAATCTATTTTCAGTACTGTTTTC  22463

EMBOSS_001     23056 CCTGTTTGGTACTTAAAGGAAAAA---TATGCTGAGGAA-CAGGCTGG-T  23100
                     ||..|||.|...||..||.|||.|   ||.....|..|| |.|..||. |
EMBOSS_002     22464 CCGCTTTAGAGTTTTTAGTAAAGAACCTAAATATATCAATCTGTATGTAT  22513

EMBOSS_001     23101 CATAAGACTG-------TATAGAACGTGCATCTT------CCATCCTATT  23137
                     |.||..|||.       ||.|..|..|..|||||.      |..|||.|||
EMBOSS_002     22514 CTTAGTACTAGGATACCTAAATTATCTATATCTGTGGTACCGTTCCCATT  22563

EMBOSS_001     23138 GAGGTGACTCCTAGACAATGGGAAAAATGCCTTCACTCGACTTGCTCATT  23187
                     ....|..||...|.||.|.|.    |||||.||||||..||.|..|  ||
EMBOSS_002     22564 ATTTTTTCTAGGATACCAAGT----AATGCATTCACTTCACATTGT--TT  22607

EMBOSS_001     23188 AAATG--TGACCGTAGCTGCTAATCTTTT--------GGCGCTGTCTCGA  23227
                     |||||  |.|...||||.|..|||.|.||        |||.||...|..|
EMBOSS_002     22608 AAATGGATTAAATTAGCAGAAAATTTATTTGGAAAATGGCACTTATTAAA  22657

EMBOSS_001     23228 ACTTTAATTAGATG-TGCTCTTC--TCTTG-AAGGTTGGAACTAC-AGTA  23272
                     |.|.||.|.||..| |.|..||.  ||||. ||..||..|.|||| |.||
```

```
EMBOSS_002     22658 AGTGTAGTAAGTGGGTACATTTTAATCTTACAAAATTTCATCTACTAATA  22707

EMBOSS_001     23273 TCCAGAGAC-CATAGAATCACA-GA----GTTGAAAACAAAATCTTGGAA  23316
                     ||...|.|. ||||.|||.|.. ||    |.|.||||.|..|..||.|..
EMBOSS_002     22708 TCTTCAAATTCATATAATAATTTGATCTTGCTTAAAATATCAAATTAGTT  22757

EMBOSS_001     23317 ATCATTGA---ATCCACTTATCAGATGAGAAAAAAAAAATAAGCCCATGG  23363
                     |.||.|||   ||..|.||||.|  |.||..|..|..|||||... |||.
EMBOSS_002     22758 ACCACTGATTTATGAAATTATTA--TCAGTTACCACTAATAATTT-ATGA  22804

EMBOSS_001     23364 AGATAGCCATTTTAAAACATATCATTCTATTTAGCCTCC-AATGTAAAAC  23412
                     |..||    |.|.||.|||||.|||.|||..|...|.. ||.|||.||.
EMBOSS_002     22805 AATTA----TATAAATATATATAATTTTATGAAATTTATGAAAGTATAAT  22850

EMBOSS_001     23413 AATGAGTTACTATGTTTCAATAATGTTGATGTTAAGA--AATTATTTGAT  23460
                     |||.|.|||..|      |||.||.|||.|.|.||.|  ||.|||.|.||
EMBOSS_002     22851 AATAATTTATGA------AATTATCTTGCTTTAAATATCAAATATATAAT  22894

EMBOSS_001     23461 AGCTTCCTC---ACTTGGTCTCCTA-TATTCCTCCA-AGGTTAC-TAGTT  23504
                     |...|..|.   |||||..|.|..|| ||||..|..| |..|||. |||||
EMBOSS_002     22895 ATAATTTTTTATACTTTATATTTTAATATTTATATATATTTTATATAGTT  22944

EMBOSS_001     23505 AGGAAGACTGTCATTCAAATTTGGAGACTACATAAGAA-GCAGAAAAAGC  23553
                     |..|.|..|.| ||..||.|.|..|...||.|||...| |.|.|||.|..
EMBOSS_002     22945 ATTATGTTTAT-ATATAACTATATATTTTATATATTTATGTATAAATATA  22993

EMBOSS_001     23554 ATATAAAGAG--GCACATGAAATTG-GAACT-----TTTCTGGTAAAATC  23595
                     .||.|.|...  |.|.|||||.||. |||.|     |||.||..|..||.
EMBOSS_002     22994 TTAAATATGTTTGTATATGAAGTTATGAAATAATAATTTATGAAATTATT  23043

EMBOSS_001     23596 TTCTTTCTTAAACTCTCCTCAAATAAGCTGTTGGTGGC--AGGAGGTGAA  23643
                     .||.|.||||||.|.||   ||||.||.|.....||..  |.||..|.|.
EMBOSS_002     23044 ATCCTGCTTAAAATATC---AAATTAGTTACCACTGATTTATGAAATTAT  23090

EMBOSS_001     23644 AGACAGCCTCCACCC----TTTAGCACAGTCCGTACTTGTCAGCATTTCC  23689
                     ....||...|.||..    ||.|..|...|...||..|.|.|...|.|..
EMBOSS_002     23091 TATTAGTTACAACTAATAATTCATGAAGTTATATAAATATAAATTTATGA  23140

EMBOSS_001     23690 CAGGA--AGGGTGATGTCTGGAAATGATAG-AGATTGTGGAAGCACATTG  23736
                     .|..|  |...|.||.|.||.||||..||. ||.|......|....||..
EMBOSS_002     23141 AATAATAATAATAATTTATGAAAATATTATTAGTTACCACTAATTTATGA  23190

EMBOSS_001     23737 CATT-ATGGGTCAAGAATGCGAAGGTCAAGGAGTGGAGTCT----TCC--  23779
                     .||| ||...|||||.. |..|||..||..||..|....||.    |||
EMBOSS_002     23191 AATTTATAATTCAATT-TATGAAAATCTTGGTATCTCTTCCAGAGTCCAA  23239

EMBOSS_001     23780 --TTTACGAAGTAGTGTTA-ACTGCTTGGCG-TGGCATTGTTGTAAACAG  23825
                       |||||.||| |||.|.| ||.....|.|. ||||.......|.||.|.
EMBOSS_002     23240 TATTTACTAAG-AGTCTGAGACCAACAGACTATGGCTAGAGAATGAAAAT  23288

EMBOSS_001     23826 A-AGCCACCAGGAAGGATCATCCTTAGGAGGGAACCTGTAGATATGACTG  23874
                     | ||.|||  |.||...|..|||.||..... ||.|||..||..||.||.
EMBOSS_002     23289 ATAGTCAC--GCAACCCTTTTCCGTATCTAT-AAGCTGCTGAGCTGTCTT  23335

EMBOSS_001     23875 AAAACAAGAGAGATCCAGT---TTTACCACTCTGGAAACATAGGTAATAG  23921
                     .|||.|.|......||..|   ||.||.||| |||...|.|.|.|||.||
EMBOSS_002     23336 GAAAAATGCCCATGCCTATACCTTGACAACT-TGGGCCCTTGGCTAAGAG  23384

EMBOSS_001     23922 AA--AGCCCAAAAGGTACCTTATCACTTGTTTGTTCCTTTCTGTACAAAA  23969
                     ..  |||||||      ||.|.||. |||||||.... |||...||||...
EMBOSS_002     23385 GGGCAGCCCAA------CCATTTCT-TTGTTTAGGG-TTTGCCTACATTT  23426

EMBOSS_001     23970 GGACTTAAATCCTTTCTGAGCAAGAAAGATATTTGAGAAT---CCAATTT  24016
                     |.||||  .||.|||||   |||.|||....|||||..||   ||||...
EMBOSS_002     23427 GAACTT--CTCATTTCT---CAACAAAATGTTTTGAATATTCGCCAAGCC  23471

EMBOSS_001     24017 TGTTTTAAAC-TTGAGCTTAGC-ATTTTGGAACTAT---TCCAAAGACCA  24061
                     ||.|.|.|.| ||||.|||..| |.||...|||..|   |.|||.|...|
EMBOSS_002     23472 TGGTATTATCATTGATCTTTTCCACTTAATAACCTTAACTTCAATGTATA  23521

EMBOSS_001     24062 CAGAATTCAC--AGTCATTAGCATACCACAGCAGACTCTTTT---CAAAT  24106
                     ||.||.|.||  ||.||.|.|.|.||.|.||     |||||.   ||.||
EMBOSS_002     23522 CACAAATTACTAAGGCAATTGTAAACAATAG-----TCTTTATCCCACAT  23566
```

```
EMBOSS_001     24107 AT-TGCAAACC---AGAACAGTCTGCTTGAAA-ACCTGGAAATACGACCT  24151
                     |. ||..|.||   ||||||...||||.|||| ||....||||.|.||.|
EMBOSS_002     23567 AAATGATATCCTTAAGAACAACATGCTGGAAACACACTAAAATGCAACTT  23616

EMBOSS_001     24152 AGTGGGTTCAACTTGACTTTTTTTATTTCTAACCCTTACCCCTAGGCAAT  24201
                           ||.....||.||.|.|.|...|...|||..||.|...|||....
EMBOSS_002     23617 ------TTATTAATGGCTATATATTAGTAGCACCTGTATCTACAGGATCA  23660

EMBOSS_001     24202 TATTG-ATAACTCATTCTGGTACCTGGTATGTATATGGACTTTGTTAGAA  24250
                     |||||. ||.|..|.|......||.|||||.|.||.|...|.|||.|.....
EMBOSS_002     23661 TATTTCATGATGCCTAGGAAGACTTGGTCTTTAAACTAATTTTTTCTCCC  23710

EMBOSS_001     24251 GAATTTGACAACTTTCTAATCATCTGTTTTTTTTCTTTTGCTTGATAGAC  24300
                     ..||.|||  .|||||..|||    ||..||.|||.|.....||.|.||
EMBOSS_002     23711 TGATATGA--TCTTTTAAAT----TGCATTATTTATAACTATTAAGAG--  23752

EMBOSS_001     24301 ATACATTTAGTAGAACTTTACTGGATTGTATTGAT-TATAAACCACATTT  24349
                        |..|.||||.....|||.|..|...|.|..|| ||.|||....|.|.
EMBOSS_002     23753 ---CTCTGAGTAATGAATTATTCAAGCCTTTAAATCTAGAAAGATGAATG  23799

EMBOSS_001     24350 CAGTTCATATCAGTCCATTTTGCTGCACAATAAACAACCAAAAAAATTTA  24399
                     .||. ||..||||..|.|.||.....|.|.|...|..|||.||||..|.||
EMBOSS_002     23800 AAGC-CAGTTCAACCAAGTTCTAATCTCCAAGTAATACCTAAAATTTATA  23848

EMBOSS_001     24400 A-TTCAGTGGCTAAT--AACAACAATATTGATTTATTCA----TGGAGCT  24442
                     . ||.|.|||.||.|  ||....|||.|||.....|..|    |||...|
EMBOSS_002     23849 CGTTTATTGGATACTTCAAGGGAAATTTTGCCCAGTAGAAGTTTGGTCAT  23898

EMBOSS_001     24443 GCAGTTTGGTAGGGTTTGGCCAATCA-TGGCT-----GGAAATGG-----  24481
                     ...|||...|||..|.||...|||.| ||.||     |.|||.||
EMBOSS_002     23899 AGGGTTATATAGAATATGCATAATTAATGACTTATTGGTAAAAGGGGTCT  23948

EMBOSS_001     24482 TTTAGCTATGCTTATCTCTA--GGCCGTCGGTTCTGT-TCGGGTCTATAC  24528
                     ||||...||...||..|.||  |..|.||..||...| |.||..|...|.
EMBOSS_002     23949 TTTAAAAATAAATAATTATATTGTTCTTCTCTTGAATCTTGGAACAGCAG  23998

EMBOSS_001     24529 CACATATTTTCTTCTGAGACTCAA---GCTGAAGGGACA---------TC  24566
                     .||.|||... ||.| ||||||.|   ||||||...|.|         ||
EMBOSS_002     23999 TACCTATAGA-TTAT-AGACTCTAAGGGCTGAAAACAAAATACTGCAATC  24046

EMBOSS_001     24567 AGCTACTCGGGGTATGACAGAGTAGCACAAGGCAATGACAGAAGCACAAA  24616
                     |...|.||..|.|||.  |||..||.|.||...|.|.|.|||||.|||..
EMBOSS_002     24047 ATTGAATCTAGTTATT--AGATAAGAAAAATAAAGTAATAGAAGTACACC  24094

EMBOSS_001     24617 CAACACTTTTCAAAATCTCTC---CTCT--TGTCACATTTGT--TTA-TA  24658
                     ..|.|.|.|.|.||.|.|.||   |||.  ||..|.|...||  ||| ||
EMBOSS_002     24095 TTAAAATATACCAACTTTATCAACCTCCAATGCAAGAGAAGTAATTACTA  24144

EMBOSS_001     24659 GC---CCATTAG----------ACAAAACATGTC-TTGTGGCCAAGCCCA  24694
                     ||   |.|||||          |..||.|.|.|| |||...||.....||
EMBOSS_002     24145 GCTTTCAATTAGTTTGATATTAAGGAATCCTTTCATTGCTTCCTTAGTCA  24194

EMBOSS_001     24695 AAGTCAAGGGGTAGGAAAATACTTTCCACCTATGTGAGGC---CATGGCT  24741
                     ...||.....|.....|||.||||||.||..|..|.||   |||...|
EMBOSS_002     24195 GTCTCCCATACTTTCTCCATATTTTCCAGCTTAGGAAAGCTGTCATTTGT  24244

EMBOSS_001     24742 GGAGCGTGAATGTATGATACTA-------------CTAGGGATGTGAAA-  24777
                     .....|...|..||.||.||.|             |||.| |..|||||.
EMBOSS_002     24245 ATTCAGACTACCTAAGACACAAAGAAAGCGTTTGACTAAG-ACATGAACT  24293

EMBOSS_001     24778 -GGATTGAGGCCAATAATTCAATCTTCT-ATTGGAGACAAGCTCAACGAG  24825
                      |||.|...||||..||||..|||||.| |.|....|.||...|....||
EMBOSS_002     24294 TGGAATTCTGCCAGCAATTTTATCTTTTCACTTCTCAAAATAGCTCTCAG  24343

EMBOSS_001     24826 TTAGTTAAAATGGAAGGCTAATATTTACTAACTTTGCAACCCAAGGAAGA  24875
                     ||.||...||||.|||..|||..||.|||||..|. ||||.      |||.
EMBOSS_002     24344 TTGGTAGGAATGAAAGAATAACTTTAACTATGTA-GCAAT------AAGC  24386

EMBOSS_001     24876 GAAAGCAGGATCTCTCTGACGATGACGGAATTTCATA---CCCTCATCTT  24922
                     ..|||.||||..|.|..||| ||||.||||....|.|   |.|.||| ||
EMBOSS_002     24387 ATAAGAAGGACTTTTTAGAC-ATGATGGAAAACAAGAGAGCACCCAT-TT  24434
```

```
EMBOSS_001     24923 TGAAGTTATACTA-AAGCTTAGGAACAACCGTCAGATAGGACTGAA-TTG  24970
                     |......|..||| ||...|||.|..|||.|  |.|| |...|||. ||.
EMBOSS_002     24435 TACTACCACCCTACAAAGATAGAAGTAACAG--AAAT-GCTGTGAGCTTC  24481

EMBOSS_001     24971 CTC--CCCCTTCCAGATTCAGCATGTGAAGTATGCAGCATCTTATTATAG  25018
                     |||  |.|.|||....|||....|.|..||...||..|.|||..||.|.
EMBOSS_002     24482 CTCATCACTTTCTTCTTTCTTTCTATACAGAGGGCTCCTTCTGTTTCTG-  24530

EMBOSS_001     25019 CAGTAGCCAAAACAGCCGTTTTCTTCAATTTGGGAATACAATGTAGGTGT  25068
                      ||||      |.|          ||.||.||.|.||.|||||.|   |||
EMBOSS_002     24531 -AGTA------AGA----------TCTATCTGAGGATTCAATTT---TGT  24560

EMBOSS_001     25069 GTTAATTTTCAATTAAGAGTTCTAAACTTATT---ATCTGCTTGG---TA  25112
                     .|||||..||.|||||.||..||.|.||..||||   |....|.|.|   ||
EMBOSS_002     24561 TTTAAACTTGAATTTAGCATTTTGAAACTATTCCAAAAACCATTGATTTA  24610

EMBOSS_001     25113 GCTCTTC--CATGTGACAGTCATTCCATCTG----ACTCTTCATGTTGGC  25156
                     |.||.|.  |||.|.|||| ||..||.|.|.    |...|..|||.|.|.
EMBOSS_002     24611 GGTCATTAGCATATCACAG-CAGGCCGTTTTTTAAAGATTGTATGCTTGA  24659

EMBOSS_001     25157 TTTTGAACTAAATTTTAAAGGAACCGCCAA--AATTTAAGG-GCCATGTA  25203
                     .||||....|||..|.|||..|.|..|.||  ||.|.||.. |.|..|||
EMBOSS_002     24660 GTTTGCTTGAAACATGAAACTATCACCTAATGAAGTAAATTTGACTAGTA  24709

EMBOSS_001     25204 CTTTTTATAACCTGTTTGTGGTCTGGGTAAGAAAATAAAAATTATACAAC  25253
                     .|||||.     ||.|||.||.|.|...|....||.|   ..|....|||.
EMBOSS_002     24710 ATTTTG-----CTTTTTTTGCTTTTTATTTCTAACT---TTTGCCCCAAG  24751

EMBOSS_001     25254 TGT-TCTTTTTGACCAG--CCACAAGCATGTAATGAAAATGA-CTGTTTT  25299
                     |.. ||.||..|||||.|  ||..|....||..||.||.|||| ||.||||
EMBOSS_002     24752 TCACTCATTGGGACCCGATCCTGATATCTGGCATAAATATGAGCTTTTTT  24801

EMBOSS_001     25300 GGCTAGCAGATGTATTAGAAGCTTTCAA-GGTGTTTAAAAAAAAAAAAAA  25348
                     ....||...||.|..|.|||  |||.|. |||||              ||
EMBOSS_002     24802 TATGAGACTATTTTCTTGAA--TTTGATTGGTGT--------------AA  24835

EMBOSS_001     25349 AAACTGGAGAAAGGAGCCAGT-GAATTGACCTCA-AACAAAAC--AAGAA  25394
                     |..||..|||.......|.|| |.|||| ||||| |||||.|..  ||...
EMBOSS_002     24836 ATTCTTCAGAGTCTTTTCTGTTGTATTG-CCTCAGAACACATTTTAATTT  24884

EMBOSS_001     25395 CAAATAAACAAAACACTTGTCTGCACTTCCAAGGAAGGGTGATATCTAGA  25444
                     .|.||.|...|..||.|..|.||    ||||...||.||..| |||   |
EMBOSS_002     24885 TATATTATGTATCCATTGATTTG----TCCACATAATGGACA-ATC---A  24926

EMBOSS_001     25445 AAAGATAGAGATGATGGAAGCACCTTGCATTATGGGTCACAAACGTGAAG  25494
                     |||.|||   ||..|.|| ||    ||||.|||...|   |.|..|.|.|
EMBOSS_002     24927 AAACATA---ATTTTAGA-GC----TGCAATATAAAT---ATATCTTACG  24965

EMBOSS_001     25495 GTCAAGGGGTGGCGTCTTCCTTTATGAAGTAGTATTAACTGCTTGGC--A  25542
                     |.......|||| ||.|.|...|.|.||||..|..|.....|||.||  |
EMBOSS_002     24966 GGATTTTTGTGG-GTATACACCTTTTAAGTGTTTATCCAGTCTTAGCTCA  25014

EMBOSS_001     25543 GGGCATT---GTTGTAAAAAGAATCCACCAGAAG-TGAAACAAG--CAGC  25586
                     |.|..||   ||||||.|.|...||..|....|| |..|.|..|  ||..
EMBOSS_002     25015 GTGTTTTCTGGTTGTACATACTTTCTCCATACAGATTCAGCTTGTACATA  25064

EMBOSS_001     25587 ACTAAAAGTTAAAAGATTTATGTGTAAACCTCATCTAAG--GCAACAGAA  25634
                     ..|...|..|.|.||||||||   |||...||||..|||.  |..|.|..|
EMBOSS_002     25065 CTTTCTACATTACAGATTTAT---TAAGAATCATGCAAACAGGGATATCA  25111

EMBOSS_001     25635 GCCATTTCTATAAAATAGTATAGGACCTTTTATTATATATGGTCCTAGAG  25684
                     ||.||||||.|    .||.|||    |.|||||.|.|.|.|    |..|||.
EMBOSS_002     25112 GCTATTTCCA----GTATTAT----CGTTTTGTGACAAA----CAAAGAA  25149

EMBOSS_001     25685 T-ATATTAAAATAAGTCTGTTTGGGTCCAT---TTGCAGCTCAT--TTG-  25727
                     . |.|||.|||.....||.|||..||..||   ||.|||...||  |||
EMBOSS_002     25150 ACACATTCAAACGCTGCTTTTTTTGTTTATATATTACAGACAATACTTGT  25199

EMBOSS_001     25728 -AAG---ATTTTTATAGGAAAAACATC-CTCAAAAATATCATACTACAGT  25772
                      ||.   |..|..|..||.|||..|.| ||....|.||||||..|.| |||
EMBOSS_002     25200 CAAACCCAAATCCAAGGGTAAAGGAACACTTTCCAGTATCAAGCCA-AGT  25248

EMBOSS_001     25773 --GCCTTGATGCTTTTTTCTTTTTATAAGGTACTGCCAGCCCAAATAGTA  25820
```

```
                        |.|.|| |...|.|.|..|.||||.|||.|.....|......|..||.
EMBOSS_002     25249 CAGGCATG-TAAATGTGTAGTATTATTAGGGAAGTAAAATAATCAAGGTC  25297

EMBOSS_001     25821 AGAAACCGATATGATTTTTGTCCATGTGAGGTGTTTAATTGCTTCCCAAA  25870
                     |||||...|....|.|||...|  |||...|....|||.|  ||....||
EMBOSS_002     25298 AGAAATTTAACCTACTTTCTAC--TGTATTGGAGATAACT--TTAGTGAA  25343

EMBOSS_001     25871 ATATGGTTATTGTGTAGA-TG-TCACTAACGAAATATATAAAGAGCAGTA  25918
                     .|||  |||  |..|||| || |.|||.|.||. ||..||....||||..
EMBOSS_002     25344 TTAT--TTA--GAATAGAATGCTAACTGATGAG-TAGCTATTATGCAGCT  25388

EMBOSS_001     25919 TTTGGGAAAATTTATTTTAATACCACCTTTTTCCTTTTTTACCCTAAAAG  25968
                     .|||.|||.||...|...|.||..|    ||||.||.|.|||||||||..
EMBOSS_002     25389 ATTGCGAAGATCCCTCGGATTATTA----TTTCATTGTCTACCCTAAACC  25434

EMBOSS_001     25969 TATTTATTTTTTTCGTAGCATACACTC-TGTGTCTCAGTA--TCATTGTT  26015
                     . ||.|.||.|.|.|||||.||.|..| |...||..|.||  |.|.||.|
EMBOSS_002     25435 C-TTAAGTTATATTGTAGCCTAGAAGCATCCATCAAATTAAATGAATGGT  25483

EMBOSS_001     26016 TTTCATAAAAACATAAATTCTTAACA-GAAAATTTCCTGCAAGCTCCCCT  26064
                     ||.|..     |.||.|.|.||..|| |..||.|.  |||||..||   |
EMBOSS_002     25484 TTCCTC-----CTTAGAGTTTTCGCATGTGAAATA--TGCAACATC---T  25523

EMBOSS_001     26065 AAGCTTGAAGAGACAAAGGAGAT---TTGTAATGTAGCTCAGCCCCAATC  26111
                     ||...||.|||.|..|||...||   ||.||.|..|..|.||    |||
EMBOSS_002     25524 AATAATGTAGATAGTAAGACAATCCATTTTATTCAATTTAAG----AAT-  25568

EMBOSS_001     26112 AGGGTAAAAGAATGCAGGGCTGACTTTATACTTATA-ACTCAGAAAAAGG  26160
                        |.|||.|..|..  |..|||.|||..|.|||.. |||||.  |||...
EMBOSS_002     25569 ---GCAAATGTTTTT--GTGTGAATTTTCAATTAAGGACTCT--AAACTT  25611

EMBOSS_001     26161 TTATGCTTCCCGTCTCTTCACAGAGCTAGTCTCT-TAATTGAT--TCCGA  26207
                     .|.|||||.....|.|||||.|.||...|.|||.| |.|.|.||  |.|..
EMBOSS_002     25612 ATCTGCTTGGAAACCCTTCCCTGAAACACTCTATGTGACTAATAATGCTG  25661

EMBOSS_001     26208 ACTA--GGA--ACATGTACAAGTGGCCCACGATCTGGAACAGACTGGCGG  26253
                     |||.  ||.  |.||.|.||||...||||||.|.....||.||.||.
EMBOSS_002     25662 ACTTTTGGCCTAAATTTGCAAGAAACCCACAAAGCATAAGAGCCTT----  25707

EMBOSS_001     26254 ATAATGG-AATATTGAGACCTTGTCTATGGTCAGCCATATTAACACTGGA  26302
                     |||.|.| |||.||.|.|.|.|.|.|||||.     |||||    |.|.||
EMBOSS_002     25708 ATAGTAGTAATTTTTACAACCTTTTTATGT-----CATAT----AATTGA  25748

EMBOSS_001     26303 TAAGTCTGATAACACTGTGATTACATATGTAT--CAATATAGTATGCTGT  26350
                     .||...|.| ||||...|.|||...|.||..|  ||..|..||.|..||.
EMBOSS_002     25749 GAAAATTCA-AACAGGCTCATTGTTTCTGAGTGGCACAAGTGTGTAATGA  25797

EMBOSS_001     26351 TAATATATTAAA----AACTTATTTACAA-----CATGATTA--TTGGAC  26389
                     .|||..|||.||    |.|...|.||.||     ||..|.||  |..||.
EMBOSS_002     25798 AAATTAATTCAAGTGGAGCAAGTATAAAAGGCTTCAAAAGTACCTCAGAA  25847

EMBOSS_001     26390 AACTGTTACAGTACAGCCA-CATCAATCCTATATCAAGTTAG-ACCATGT  26437
                     .|..||||.||.||||..|| ||.|||..|...|.||||..|. |.|||.|
EMBOSS_002     25848 GAAAGTTATAGAACAAGCATCAACAAAACGCAAACAAGAAAACAACATAT  25897

EMBOSS_001     26438 CAACTGGTTTTGTGTTGAGACACCTGTGTATGGACATAGTCTGAACTTTT  26487
                     ...|..||.||    |.|.|||       |.||...|..||||..|..|..
EMBOSS_002     25898 TTTCCAGTATT----TCACACA-------AAGGGTGTTATCTAGAAATGA  25936

EMBOSS_001     26488 CATAGTTTGTGCTAAATGATAGCAATCAACATCGGTATGGCACTTACAGT  26537
                     ||.||.|||||  |||..|.|..||  |...|.|||...|.|..||.||.
EMBOSS_002     25937 CACAGATTGTG--AAAGCACATTAA--ATTGTGGGTCATGAATGTAAAGG  25982

EMBOSS_001     26538 TTACTGATAACTTTCATGCCCATTAACATAGTACCGCAATAACTCTGTGA  26587
                     |.|..||.....|||.|.|.||.|.|   ||.||.  |.|||||...||
EMBOSS_002     25983 TCAAGGAATTAATTCCTTCTCAATGA---AGAACT--ATTAACTGCTTG-  26026

EMBOSS_001     26588 AGTGCTGAATTTGTGTCCTGTTTCATGATTGTATTTGTGTTGATATCTCA  26637
                        ||.|||..| |||  |||...|.||.|.||.|.|...|||.|.....
EMBOSS_002     26027 ---GCAGAACAT-TGT--TGTAAAAAGAATCTACTAGAAATGAAACAAGG  26070

EMBOSS_001     26638 GTCAGTCAGAGTCCCAACAAGAAACAGATGGC-ACATTCAGATTAGGGTA  26686
                     ..||.|.|..||      ||.||.|..||||. |.|.|.|..|.||....
```

```
EMBOSS_002     26071 AGCACTAAAGGT------AAAAAGCTTATGGGTAAACTGACCTAAGTCAG  26114

EMBOSS_001     26687 AGTTGAGGAGTCTTTATTTA-CAAGGCACTACATAC--TCAGGATTGGGC  26733
                     |.|.....||| |.|||..| ||.|||..||.||.|  |.|....||..|
EMBOSS_002     26115 AATAAGTTAGT-TCTATACAGCAGGGCTGTATATTCCTTTATATGTGATC  26163

EMBOSS_001     26734 AGGGTGTAGGGAAATCTCACAAGATA---GCACA-----AGACTCTAGGA  26775
                     ..||.|||...||.|.|..|..|.|.   |..||     |||.|.|..||
EMBOSS_002     26164 CTGGAGTATTAAACTGTGTCCTGTTGTGTGTCCATTGACAGATTTTTGA  26213

EMBOSS_001     26776 CTAGCA--GCAGCAGAGCTGTCACCTCTCCTAGACCTGAAGCCG-TTGTT  26822
                     ..|.|.  |.|..|.|....||.|.||...|||...|..||... |||||
EMBOSS_002     26214 AGAACTTTGTAAGAAAATCATCCCGTCAAGTAGTATTACAGAATATTGTT  26263

EMBOSS_001     26823 GGGGAGAGAGGTTTCTCA-GAG------CCCAGAAAAAAAGAAAAAAAAA  26865
                     ...     |.||||.||| |||      ||||....||.||.|..|||..
EMBOSS_002     26264 TTA-----ATGTTTTTCATGAGGCATTGCCCACCTCAATAGTAGGAAAGT  26308

EMBOSS_001     26866 AAAAACATCATGCAGATTTTAATGCCTTGGGAGGAGCAGTGGCTTTCTCT  26915
                     .||.|..    |||||||  |||...||...||.|        |.|.|.|
EMBOSS_002     26309 TAATATG----GCAGATT--AATTAGTTCCTAGAA--------TATATTT  26344

EMBOSS_001     26916 TAAGGACAGAATTTGCCTCGAAATGATACTCAGGGAAAAAGA-GATGAAG  26964
                     |||.. |||...|..|.||..||..||....||.|.|||||.. |...||.
EMBOSS_002     26345 TAACT-CAGGTGTCCCTTCTGAAGTATTTAAAGAGCAAAATTTGCCCAAT  26393

EMBOSS_001     26965 GGAATCAA-TACTCTGACCCAAGACTCTCCCTTCTCTGCAGTGGTTTGCT  27013
                     |.|.|.|| |||    |.|..| |.|.|...|.||.||..|...|||..|
EMBOSS_002     26394 GTATTTAAATAC----AACTTA-AATATGTTTGCTATGATGACATTTATT  26438

EMBOSS_001     27014 AGTCCTCTCC-TTGGTCAAACCCAAACAGAAAACCATAGGGCATAGGAGT  27062
                     ..|..|.|.| ||..|||..|.....|...|.|..||....|||||..|||
EMBOSS_002     26439 CATTTTGTGCATTACTCACTCTGTGTCTTGATATTATTTTCCATAAAAGT  26488

EMBOSS_001     27063 CTAATGATGTAATCCAAGTCAGCCCCCTGGAAGGTGGAAAAAGAAGGGAA  27112
                     .||| ||...|.|.||||   ||....|..|.....|.||...|   |||
EMBOSS_002     26489 GTAA-GAGTCAGTGCAAG---GCTATTTTTATTTCTGCAATCCA---GAA  26531

EMBOSS_001     27113 AATGGATCTGGAT--CTGGAGGGATACCAAAAAAAAAAAAAAAAAAAAAA  27160
                     ||  |||...|||  |||......|||...|..|..........||..|.
EMBOSS_002     26532 AA--GATTGTGATTTCTGACCATTTACAGGACTAGTCTCTTTTTAATTAT  26579

EMBOSS_001     27161 AACCATAGTTGGCATGCTTGTTTATTGATATTTTCTTG-CATGATATAAG  27209
                     ..|||.|.|.||.||..||..  |.||||.||...|.| ||...|||.||
EMBOSS_002     26580 TTCCAAATTGGGAATATTTAC--AGTGATCTTGAATAGACACTGTATTAG  26627

EMBOSS_001     27210 AATCCAGATAAATATAGTAAGAG-GTCTATTTTACTAACAATT-------  27251
                     |||...||.|.||.|.....||. |||..||.||..||||.|.
EMBOSS_002     26628 AATATTGAGACATGTCTATGGATAGTCAGTTATATAAACAGTGGTTAAAC  26677

EMBOSS_001     27252 TTAGGCACCTAATAATAATACTCCTTCTTTGAATGTATAACCTCTAGAAT  27301
                     |||.|.|||.|||||..|.||||     |..|||....|.|.||.||..
EMBOSS_002     26678 TTAAGGACCCAATATAATTACT-----TACGAAACATTGAACTATACT--  26720

EMBOSS_001     27302 TGGTTCAGAAAT-GTAACTGTGCCGTTACAATTTCTATTAGTATTCAACA  27350
                       ||.||.|.|| |.|||||||..|. |.||||..|..|....|...|.|
EMBOSS_002     26721 --GTACATACATTGAAACTGTGTAGC-ATAATTAATGCTCAGCTGACATA  26767

EMBOSS_001     27351 GTAGATTCATATCCATTCATCTATGACTGGAGTATC--TGCCATTTGCTG  27398
                     .||.|..|||.|||||.| |.||.. |.|||..|..  |||||.|||
EMBOSS_002     26768 ATACAGACATCTCCATCC-TATACT-CAGGAAAAGGCATGCCAATTG---  26812

EMBOSS_001     27399 GTTAGTTACTGTGTAAGGTACTTTGTAAGGTATAGAAATACACTTGGGGT  27448
                      ||.|.|.|...|..|..|.|.|....|....|.||||||...||.|..|
EMBOSS_002     26813 -TTTGCTGCCAGGACATCTGCATGAACACAGTTTGAAATATTTTTAGTTT  26861

EMBOSS_001     27449 GCGATGGC---TCATGCCTGTAATCCCA----AGGATTTGGGAAGCTGAG  27491
                     ....||..   ||||.|||.||.|..||    |.|||.|...|| ||||.
EMBOSS_002     26862 ATACTGAACAGTCATACCTATATTATCATAGTATGATATAATAA-CTGAT  26910

EMBOSS_001     27492 GCAGGCAGATCACTTGAG----TCCAGGAGTTTGAGATCAGCC----TGG  27533
                     .|..|.| |||. ||||.    ||||.||..||.....|.|||    ||.
EMBOSS_002     26911 ACCAGGA-ATCT-TTGAACTATTCCAAGATCTTATTTACTGCCAATCTGT  26958
```

```
EMBOSS_001     27534 GCAACATGGTGAAACCCCATCTCTACAAAAAA---TGCAAAAAGAGTACC  27580
                     | |..||..|..|...|.|....||.|.|..|   ||.||..|||.||.|
EMBOSS_002     26959 G-ATGATATTTCATTACAAGTAATAGATATCACATTGAAATTAGACTAAC  27007

EMBOSS_001     27581 ---TGCGCATGGTGGCATGTGCCTGTAGTCCCAGCTAC-TCGGGAGCCTG  27626
                        |...||||.|....||| |.|.||.|.||..|.|. |....|..|..
EMBOSS_002     27008 ATTTTTACATGCTACAGTGT-CTTTTATTGCCTTCAATATTTCCACACAT  27056

EMBOSS_001     27627 AGGTAGAAGGATCACGTGAACCCAGGAAGTCGAGGCTGCAGTGAGCCATA  27676
                     |||.|.||..|.||.|     |||...|||..|.  ||.|   ||...||
EMBOSS_002     27057 AGGCAAAATAAGCATG-----CCATTTAGTGTAT--TGGA---AGTTGTA  27096

EMBOSS_001     27677 ATGGCACAACTGCACTCCAGCCTGGATGACAGAGTGAGACCCTATCAAAA  27726
                     |.  ||||||   |.|   ||||.|||.|.|||   |.|||||..  |||.
EMBOSS_002     27097 AA--CACAAC---ATT---GCCTAGATAAGAGA---AAACCCAG--AAAC  27133

EMBOSS_001     27727 AAAAATAAGAAATAAATTTGAGCTCAGTGACCTACATTCTAGTGCAGAAA  27776
                     |.|..||||.||.|      |||||| ||..||.||  ||..|....||.
EMBOSS_002     27134 ATACTTAAGCAAAA------AGCTCA-TGTACTCCA--CTGATATCCAAG  27174

EMBOSS_001     27777 AAAATGACCATAGTTGATTATGAGATTTTAAAGCAATAAACCA-CATGAG  27825
                     ||..|||. ||.||...||.|....|||......|.||.|||. |||.|.
EMBOSS_002     27175 AATGTGAA-ATTGTATGTTCTAGACTTTGCTCTGATTAGACCTTCATTAA  27223

EMBOSS_001     27826 ACATACTAATGAGCTCATAAGATCATTCAGAAATTGTTTATTAT--GAAC  27873
                     .|  ||.|.|||.......|..||.|..|..||..||||.||.|  |||.
EMBOSS_002     27224 CC--ACCATTGAATAACCCAACTCCTATATCAACAGTTTCTTGTCAGAAA  27271

EMBOSS_001     27874 ACATAGTACTTTCAGTGTGGCATTAAACAGAGATCACTGTCCTTAAACAA  27923
                     .|||.|.|.|..|||....|...|   |..||||.|.|.|..|...||
EMBOSS_002     27272 TCATTGAAATCCCAGAAAAGTTCT---CCTAGATAAATCTGGTATCAC--  27316

EMBOSS_001     27924 GTTAAAAGCAGAATCAA---ATCATCTGCAAATTAACACACCACTAAACT  27970
                      |.|||.||.|.||.||   |..||.||.|||..||.||.|.....||||
EMBOSS_002     27317 -TAAAATGCTGTATTAAGGGAGAATATGTAAAGCAAGACCCTGGCTAACT  27365

EMBOSS_001     27971 TTAAGCTTCTTGAGTGATTCTGTAATTTTTA---AAATGTC--TTCAGCA  28015
                     |.||...|   .|||....||||.||||.||   |.|||.|  ||..|.|
EMBOSS_002     27366 TCAAATGT---AAGTTTGGCTGTCATTTCTATGAAGATGACATTTGTGAA  27412

EMBOSS_001     28016 TTTCAG-----TGTCA---AGAT------AGT-----GCAAACT--CAGT  28044
                     |.||||     |.|||   ||||      |||     |..||.|  |||.
EMBOSS_002     27413 TATCAGGTTTCTATCAGGCAGATTACCTAAGTTTTGAGAGAAGTTCCAGG  27462

EMBOSS_001     28045 AAAAGCTTGTGGAATTGC-ATTAAACAA-AACCA---AAATAAATAGATT  28089
                     ||||.......|..|||. |..|.|.|| ||||.   ||||.|||.||..
EMBOSS_002     27463 AAAATGAAAGAGCCTTGAGAGCAGAGAAGAACCCTGGAAATCAATGGAGC  27512

EMBOSS_001     28090 TTAT----TAAAACTATATACA-----ATTGTCTTTCTAATCATATCCTC  28130
                     ....    ||||||||.|.||.     |||.|||||.||....|| |||.
EMBOSS_002     27513 CAGGAGCCTAAAACTAGAGACGTTGCTATTTTCTTTTTAGGTCTA-CCTT  27561

EMBOSS_001     28131 TCCATGAATAGGGAAGAAATAATT-----TTAGGAATTTAAATATCTTCT  28175
                     ||.....||..|||||||||||...     |||..|||||||||||.|||||
EMBOSS_002     27562 TCTTATCATTAGGAAGAAATACAGGTCCCTTACAAATTTAAATACCTTCT  27611

EMBOSS_001     28176 ATCTTAATAGTTCCTCTTATTTCCCTCTTAAGCAATGTTCACTCCTTCAA  28225
                     ..||.||||.||.||.|||||.||.||||..||.|||.|.||.||..||||
EMBOSS_002     27612 TCCTAAATAATTTCTTTTATTGCCTTCTCTAGAAATATCCAGTCTCTCAA  27661

EMBOSS_001     28226 AAATATTTATTGAGCATCTAATATGTACTTAACACTGTGCCAGGTGCTGT  28275
                     .. ||||||||||..||||.|||||||    ||||...|.|||.|||..|
EMBOSS_002     27662 GG-TATTTATTGATTATCTCATATGTA----ACACAAAGTCAGATGCCAT  27706

EMBOSS_001     28276 GAAGAATGCCAAGGAAATAGAATGAACTTCTAATTCTTTGGAGTTCCAAT  28325
                     ||    |.|.|||||||.||||.||||||.||||||||..||..||||||
EMBOSS_002     27707 GA----TTCTAAGGAAAGAGAAGGAACTTGTAATTCTTCAGATCTCCAAT  27752

EMBOSS_001     28326 TAAATAACCTAAAGTTAAATTGGTTTCGGAGAGAACATTATGCCTTCGAG  28375
                     |.||||||||.||||||||....|||..|| |.|||||||||||||.|||
EMBOSS_002     27753 TGAATAACCTGAAGTTAAACCAATTTGAGA-ACAACATTATGCCTTTGAG  27801
```

```
EMBOSS_001     28376 ACTGTAGGCTTCTCTTGATTAGAAAGTCTTAAACATTTTAAGTAACTAAA  28425
                     |||.||||  |||||||||||||.||.||||||||||.||.|||||||
EMBOSS_002     27802 ACTCTAGG--TCTCTTGATTAGAATGTATTAAACATTTTCAGCAACTAAA  27849

EMBOSS_001     28426 CAGATTAAGGAGAATTCAAGGATGCCTCTCACTAGTAAATTTGGATTAGT  28475
                     .|||||||.|.|||||||||.||||||||||.|||||||||||||||.||
EMBOSS_002     27850 TAGATTAAAGGGAATTCAAGAATGCCTCTCACCAGTAAATTTGGATTCGT  27899

EMBOSS_001     28476 CTGGCAAACTTCAGACCTTAAATGCAAGATTTTTAATAATTAAAAGAAGA  28525
                     ||||||||.|||.||||||||||||.|||.  |||||||.||||||||||
EMBOSS_002     27900 CTGGCAAATTTAAGACCTTAAATGGAAC--TTTTAATGATTAAAAGAAG-  27946

EMBOSS_001     28526 GAGAAAATGATAATTACA---TTTCTAGAGTCTATGTTTACCATTCAGCC  28572
                     ||||||||||||||||||   ||.||||||.||||||||||||||.|.|..|
EMBOSS_002     27947 GAGAAAATGATAATTGCACCATTCCTAGAATCTATGTTTACCACTAACAC  27996

EMBOSS_001     28573 TTCTTAATCATTTCCTAAGTATATCTGGTGATCAGGATTTTATAACTCCA  28622
                     ||||||||||..||...|||||||.|.|||.||||||||||||||.|||.|
EMBOSS_002     27997 TTCTTAATCACCTCTGGAGTATATTTAGTGTTCAGGATTTTATAGCTCAA  28046

EMBOSS_001     28623 GAAAATCTTTCTATACATCGCATAAATCTCTTCTTTTAAAAAGCTCTTCA  28672
                     |.|||.||||..|||||||...||||||.||||||||.|||.|||.|||.|
EMBOSS_002     28047 GGAAACCTTTGCATACATCCAGTAAATTTCTTCTTTCAAAGAGCACTTTA  28096

EMBOSS_001     28673 ATTTTGTATTTTGTTAAAACT-TAAAAGCCTCCATGAAAAATGAGACAAA  28721
                     ||||||||||||||||||.||. |.||||||..|||||||
EMBOSS_002     28097 ATTTTGTATTTTGTTAACACACTGAAAGCTGCCATGAA------------  28134

EMBOSS_001     28722 AGTCAGTGAGAGGCTGTAGCAATAAAAATCAGATGTGATTTTCTTTTGAA  28771
                                   |||..||||||||  |||||..|  |||||||| |||
EMBOSS_002     28135 -------------CTGGGGCAATAAA--TCAGACTT--TTTTCTTT-GAA  28166

EMBOSS_001     28772 TAACATCTGTTTTTACAGTCCTTTCATGTTAAACTTTATAAGAATTTATT  28821
                     |||.||.|..||||||...||..|||||||.||||||||||||....|.|||
EMBOSS_002     28167 TAATATTTTGTTTTAACATCTCTTCATGCTAAACTTTATAAATGGTCATT  28216

EMBOSS_001     28822 ATAAACA--GCTTTATTGACAGTTCAATCCTATTTCTAAAAGGATTTATT  28869
                     |.|||.|  ||||||||||.||||.|.|.|||||.  |..|.|||||||.
EMBOSS_002     28217 ACAAAAAATGCTTTATTGAGAGTTAACTTCTATTA--ACTATGATTTATA  28264

EMBOSS_001     28870 T-TCCCCCAATGGTAAGAGTTTTCTTTTCTTAAACCTAACTAGTTGCAGA  28918
                     | ||.|.||.||||||...|||||||.|.||..|.|.|  |.||  ||||
EMBOSS_002     28265 TGTCTCTCAGTGGTAAAGATTTTCTTCTATTTCATCCA--TTGT--CAGA  28310

EMBOSS_001     28919 TATTTCAGATACTACATTTCTCATTGTGTAAGGTAAAGTTTCTGACCACC  28968
                     ||.|||..|. |||.||||||||.|.||||||...|....|.|||..||||
EMBOSS_002     28311 TACTTCTTAG-CTATATTTCTCACTATGTAAATGACTACTCCTGGTCACC  28359

EMBOSS_001     28969 TGAATATGACTTGTAGCTCCTGAGAA-CAATTTGTTTAGTACCGATATCA  29017
                     |..|||.||||..||..|..|..||| ||.||||..||.||||.||.|||
EMBOSS_002     28360 TTCATAAGACTCTTAATTTTTATGAAGCATTTTGCATAATACCAATGTCA  28409

EMBOSS_001     29018 TGCAGTGACATTGGTACAAAGGAATTTTCTTTATTTCACTGTACTGTTTT  29067
                     ||||.||||....|.|||| ||.||||||.|||||.||||.||.|..|.|
EMBOSS_002     28410 TGCAATGACCACAGCACAA-GGCATTTTCCTTATTCCACTTTATTACTAT  28458

EMBOSS_001     29068 CAGTTTTATTCTATAGTTGTTAAATAAGACCATTAAATATTTTTATTAGT  29117
                       ||||||||||.|.||||||||.||||||.||||||..||||.. |||||.
EMBOSS_002     28459 --GTTTTATTCTGTTGTTGTTAATTAAGATCATTAATCATTCC-ATTAGC  28505

EMBOSS_001     29118 CTTATTTCCTGTTTAACTAGGTGGGTTTTTGATCTCTGTTCAGTAAAGCA  29167
                     ..|.||||||||||.|.|||.||||||.|.|||||..|||.||||||.|||
EMBOSS_002     28506 ACTGTTTCCTGTTCACCTAAGTGGGTGTGTGATTGCTATTCAGCAAGGCA  28555

EMBOSS_001     29168 TTGTGCTCTTCAGAGCAAGCAATTGAAAAGCAAATAGTGAGTATTTCTAC  29217
                     .||||||..|||.||.||.|||||||||..    ||.||||||..||...|
EMBOSS_002     28556 CTGTGCTAATCAAAGGAAACAATTGAAATC----TACTGAGTTGTTTCTC  28601

EMBOSS_001     29218 TGTAAAAGTTTAACATTAAAAGATATACACACAGCCAGGCAAGGTGGCTC  29267
                     ..||.||||||||||.|.|||.|..|.|.|.|||  ||....||..|..|.
EMBOSS_002     28602 CATAGAAGTTTAACCTAAAAGGGCACATATACA--CATTTGAGCAGCTTA  28649

EMBOSS_001     29268 ACGACTGTAATCCCAGCAATTTGGGAGGCTAAGGCAGGAGAATCGCTTGA  29317
```

```
                    |.|.||.||...|||......|.......|||.|||....|.|..||..|
EMBOSS_002     28650 ATGTCTCTACCTCCATTTTAATTATTTTATAATGCAACTAATTTTCTATA  28699

EMBOSS_001     29318 GCCCAGGAGTTCGAGACCAGTCTGGGAACCATAGCAAGACTCCGTCTCTA  29367
                     ...    |.|||.|...||.|.||..||   ||.|||   ||.||.|||
EMBOSS_002     28700 CTT----AATTCCATGTCAATATGTAAA---TACCAA---TCTGTTTCT-  28738

EMBOSS_001     29368 CCAAAAAAATTTTTTAAAAAATAGTTGGATGTGGTGGAACACCTCTGTAA  29417
                               |||||.|.|||..|   ||.|.         ||.|.|.|
EMBOSS_002     28739 -----------TTTTAGACAATTAT---ATTTT---------CTTTTTTA  28765

EMBOSS_001     29418 TCCCAG-CTACTCAGGACGCTGAGGCAGGAGGATTGCTTGAGCCTGGGAG  29466
                     |...|| ||..|..||.|.||     ||.||||.|||||..|       |.
EMBOSS_002     28766 TTAGAGTCTCTTTTGGTCACT-----AGCAGGAATGCTAAA-------AA  28803

EMBOSS_001     29467 GTCAAGGCTGCAAGGCTGCAGGGAGCTGTGACTATGCTACTGTACTCCAG  29516
                     ||.||     ||||..|.          |.|..||.|||..||.....|.
EMBOSS_002     28804 GTGAA-----CAAGTTTT----------TCAACATACTAGGGTCAAAGAA  28838

EMBOSS_001     29517 TCTAGGTGACAGAATGAGACCCTCTCTCTCTCAATTAAAAAAAAAAAAAC  29566
                     .||..|..|.|.|.||. |||.|.|||   |.|||||.....|.|....|
EMBOSS_002     28839 ACTCAGAAAGATAGTGT-ACCATATCT---TGAATTAGGCCCACATCCTC  28884

EMBOSS_001     29567 AAGATACACACACATATATTTGCGTAGGTAACTCTAATTTCATTTCAAGT  29616
                     ..||||.|.|..||.|.||..||.|...|.||   |...|||.||.   |
EMBOSS_002     28885 CTGATATATATTCACAAATCAGCTTCCATGAC---AGCATCACTTA---T  28928

EMBOSS_001     29617 ATGTTATGTAACAACCATTTGTGTAGTGCTTGTAACAGT-CAATATGTAA  29665
                     ..|..||..||.||.|.|.||    |.|.|..||..||| |.||.|.|||
EMBOSS_002     28929 CAGAAATAAAATAATCCTATG----GAGGTATTATTAGTGCTATTTTTAA  28974

EMBOSS_001     29666 ATACTGACTCATCTTCTTTGACAATTCTACCTAGATACTTATTAGAGTCC  29715
                          ||   ||.|||||..|..||..||..||..||.|||||....||.
EMBOSS_002     28975 -----GA---ATGTTCTTAAAATATAGTATTTATTTATTTATTTATCTCA  29016

EMBOSS_001     29716 CCCTTAGTCATTGAAAGGAAGGTTAAAATCAAAAGACGTTGTTTGCCAAA  29765
                     .|.|.|  |.||.|               ||.|....|||.|....||||
EMBOSS_002     29017 TCATCA--CCTTCA---------------CATATTCTGTTCTCCTTCAAA  29049

EMBOSS_001     29766 GTAATGAAAGAAAAC-TTATAAACACAATGTATCATGTCTGGGGCTGAAC  29814
                     |....|....||..| ||.|.||.|.||  |||..||..|     |.||.
EMBOSS_002     29050 GCCCAGCTCCAATTCCTTTTTAATAAAA--TATTCTGCAT-----TAAAT  29092

EMBOSS_001     29815 TAAAACCCTTCTGATATGTGGTATTAACAGATCATCTTTCATGACAGTAC  29864
                     |.|.||....|||.|.|.|.....|.||   ||.||.|||...||.||..
EMBOSS_002     29093 TGAGACAACACTGGTCTCTTTCTCTCAC---TCTTCATTCTGCACTGTTG  29139

EMBOSS_001     29865 CAGTTATTAGAAATAAAATGATTGGAGTTATTATTAATACTAACAATAGT  29914
                     .|    |||||   |||||.|   ||      |.|.|.||||.||
EMBOSS_002     29140 AA----TTAGA---AAAATCA---GA------ACTCAGACTACCA-----  29168

EMBOSS_001     29915 GGTATTCTTAAAATGACTTCCTTATTTATCTTCACCTTTATACATTCTAC  29964
                     ||.|.|||.|    |.|                     |.|.||..|..
EMBOSS_002     29169 GGAAATCTCA----GGC----------------------AAAAATGGTGG  29192

EMBOSS_001     29965 TACTGCTTCAAGACCCATCTTGAATTCTTCTTCCACAGAACATTCTGCAT  30014
                     .||       |.||||||.|    ||||...|...|.|||..|. |||..|
EMBOSS_002     29193 GAC-------ATACCCAAC----ATTCAGGTGGTAGAGATGAC-CTGGCT  29230

EMBOSS_001     30015 TAATTTCAGCCAACATTGATTTCTCTTTTTAAAATTTGTCTTGCACAGTG  30064
                     |   .||||...|..|.|...|.|...|..|.|...|||       ||.|
EMBOSS_002     29231 T---GTCAGGAGAGTTAGGGCTATACATAGAGACCCTGT-------AGAG  29270

EMBOSS_001     30065 AATTAGAAAACCAGGAATTGGAAAACCAGAAAAGCTTATTAAGTAAGAAG  30114
                     |...||||||   |..|.||||.|..  |.|.||       ||.|||||..
EMBOSS_002     29271 AGGGAGAAAA---GAGAGTGGAGAGG--GGAGAG-------AGGAAGAGA  29308

EMBOSS_001     30115 CAGAGAG-GAGAGAGTTTCAACAAAGGGCCATTCTAAAGTGGTCTACTGC  30163
                     ||||||| ||.||||     |||.||.|..|..|   ||      ||. |
EMBOSS_002     29309 CAGAGAGAGACAGAG-----ACAGAGAGAGACAC---AG------ACA-C  29343

EMBOSS_001     30164 GGACACCATACTGATTATAGTTGGTGATTAAATCTTATCTTTCCAACTGA  30213
                     .||||| |.||.||....|..      |||.|| ||.||||||..||||..|
```

```
EMBOSS_002     29344 AGACAC-AGACAGACACAAAC------TTAGAT-TTGTCTTTTTAACATA  29385

EMBOSS_001     30214 TTATAAACTCCTCCAGGGCATACTCTTATATTCCACAAGATGCTTATCTG  30263
                     .|..|||.|||||.|.||.|||..|||...|  |.|.||.||..||.||.
EMBOSS_002     29386 GTGCAAAGTCCTCTATGGAATATGCTTGGTT--CTCCAGTTGACTAGCTC  29433

EMBOSS_001     30264 GGTGCAGAGCATGCATGCAGTTGGTATTTGCTGATTTATCAACTAACTAA  30313
                     .|||||.|...|..||....||.|.|||||..|||||||..||.|||||||
EMBOSS_002     29434 AGTGCAAACAGTATATATGTTTAGCATTTTTTGATTTAGAAATTAACTAA  29483

EMBOSS_001     30314 ATCTTAACATATTATTATTAACAATTTAAAATAAAGTTAAATGTATCACT  30363
                     |..|||||||.||||||.|...|.|.|..||     |||||||.|.||.|
EMBOSS_002     29484 AATTTAACATGTTATTAATCTAACTCTCTAA-----TTAAATGCAACATT  29528

EMBOSS_001     30364 CTCCAC--CCCTCAAAGCCATTTCTGT-TCTTTGTTTTCATAGCACCATT  30410
                     ...||.  ||.||..||.||   |.|| |.|||.||||||.|||||. |.
EMBOSS_002     29529 AAACATGTCCTTCTCAGTCA---CAGTATTTTTTTTTTCACAGCACT-TG  29574

EMBOSS_001     30411 ATTATTTCCTGCATAGTATTTTTTAAAAACCGTATTTTTAAAATTTATAT  30460
                     |||||||..|.||.||.|.|||.|..|.          ||.|  |.||.|
EMBOSS_002     29575 ATTATTTTTTTCACAGAAATTTCTTGAC----------TAGA--TAATCT  29612

EMBOSS_001     30461 ATTTGTT-TATTTGGGTATACTTCACTAGATTGTAAGCGTCACAAAAGCA  30509
                     |||.||. ||.||.||.|.|  |.|.|..|.|.|..|| |....|..|.|
EMBOSS_002     29613 ATTGGTACTACTTAGGAAAA--TAATTCCACTCTGCGC-TGTGCACTGTA  29659

EMBOSS_001     30510 GAACTATTATAACCCCAGCCACTAACACAATGCCTAACAAATAGTAGGTT  30559
                     |  ||..|||    |.||...||.|.||| |||..|.|||.|     ||.
EMBOSS_002     29660 G--CTCATAT----CTAGATTCTCAGACA-TGCAAAGCAAGT-----GTC  29697

EMBOSS_001     30560 CTCAATATTTGTTGAATGAATGACCTACAGATATTACTTCATTATGAAAG  30609
                     ||            |||||. |.|.|||.|....|.|||.|    ||.|.|
EMBOSS_002     29698 CT------------AATGC-TAATCTATACCCCTGACTGC----TGGATG  29730

EMBOSS_001     30610 ATTTTGCTAAGTTGTTTTA-CATCTATTTTATCCAAAACTAAAGTTCTTG  30658
                     .|.||..||.....||||| |..|||||||.||||||.. ||.|.|.|| |
EMBOSS_002     29731 TTCTTTTTATTGCCTTTTAACCGCTATTTTCTCCAAGG-TACACTGCT-G  29778

EMBOSS_001     30659 AGGCAAAGCCTAGAATAT-CTTCTATGTTCTCACAATGCTCTGAATCAGT  30707
                     ||...||  |.|||.||| |.|.|.||| |||.||.|.|..|||  ||..
EMBOSS_002     29779 AGATTAA--CAAGAGTATGCATGTTTGT-CTCTCAGTTCCATGA--CATA  29823

EMBOSS_001     30708 GCTTCTCTTAATATGCATAGCAATTGCCTGGAGAGCTTGTTAAAACATAG  30757
                     |||||..||       ||||..||.|.|||..||   .|||     |||.|
EMBOSS_002     29824 GCTTAGCT-------CATAAGAACTACCTTCAG---GTGT-----CATGG  29858

EMBOSS_001     30758 ATTACTTAGCCCCAACCCCAGAGATGCTGATTCAGTAGGTCCCAGGTGAT  30807
                     ||     .|.||||...|.|||.|..|||||.|  ||.||.||.....||
EMBOSS_002     29859 AT-----GGTCCCACATCTAGACAGTCTGATCC--TATGTTCCCAAAAAT  29901

EMBOSS_001     30808 GCTGCTGC-----TGTCAGTCTCTGGCGCACACTTTGAGTAGTAGGGCTC  30852
                     .|||....     ||.|||||||.|.|.|||.|||||||..||||||.||.
EMBOSS_002     29902 ACTGAAATAAAATTGCCAGTCTGTTGAGCATACTTTGATGAGTAGGTCTG  29951

EMBOSS_001     30853 TAGGATGTTATATGTACAGACACATGCTGAATAGTGGGCTATGTGCTTAC  30902
                     ||||.||||.||||||.|||.|.||||||.||....|...||||||||..
EMBOSS_002     29952 TAGGGTGTTTTATGTAAAGAAATATGCTGTATGAAAGAGAATGTGCTTGT  30001

EMBOSS_001     30903 TTGCTGGCTAAATAATAAATGTTCTCACT-GAGTCATA----GAAC----  30943
                     ||....||.|||||..||...|||....| ||..||.|    ||||
EMBOSS_002     30002 TTCTCAGCAAAATAGCAATATTTCAAGTTTGAAACACATAAGGAACCGTG  30051

EMBOSS_001     30944 -------------TTTGAA--------ATTTGCAAGGACTTTTGCTATTA  30972
                                 ||||.|        |||||..|..||.|.|..|||||
EMBOSS_002     30052 CTAACATGTAGTGTTTGTATGACAAATATTTGATATTACATATAGTATTA  30101

EMBOSS_001     30973 TCTA--GTCTATGGATAGCAA---ATAACCTG--ATACCGTGCTATAGTG  31015
                     |.||  |.||||...|..|.|   ||||..|.  ||...|...|||.|.|
EMBOSS_002     30102 TATATGGACTATACTTTACTATACATAATATAGCATTTAGAAATATGGAG  30151

EMBOSS_001     31016 CTTGACTGCATTTAACCTGCAGAATCCTCATGAGCAG--CCCAGCACCAT  31063
                     ||.|..|.....|.|||| ||||||.||.| |||..|  ||||..||.||
EMBOSS_002     30152 CTAGCTTTTTAGTGACCT-CAGAATTCTTA-GAGGTGTTCCCATAACTAT  30199
```

```
EMBOSS_001     31064 CACTCCAAGTGAAACTA--CTCTCTTCTTGAGGTTGTCCAATTCTATCAA  31111
                     ..||    |.|||.|..  ||.|||| ||.....|||...|.|||||||.|
EMBOSS_002     30200 GTCT----GGGAAGCATGGCTGTCTT-TTTTTCCTGTTACAGTCTATCTA  30244

EMBOSS_001     31112 TTAAAGATGAAAACCAGGTTCTGAGAGTTGAAATCTCTGGACTTCAAAGG  31161
                     |||. |||||..|  ||...|||| ||||.|.|           |||||.
EMBOSS_002     30245 TTAT-GATGATGA--AGAAACTGA-AGTTCAGA-----------CAAAGA  30279

EMBOSS_001     31162 TCCAACAGCCCAGGTCTTCTCAATTCTCGTTAGTGTTTCAGCAGCTGAAT  31211
                     |  |||||||.||.|.||.|          ||.||||    ||.||||||
EMBOSS_002     30280 T--AACAGCCTAGATGTTTT----------TATTGTT----CATCTGAAT  30313

EMBOSS_001     31212 ACAAATTTATTAAGCTGTATCAGAGTAGTATCTGTCAAATTGGAGTGTCC  31261
                     ..||.|||.|||||.|..||...|||||..|||.||||||..||.|||||
EMBOSS_002     30314 CAAATTTTCTTAAGATAAATTTTAGTAGCCTCTATCAAATGTGAATGTCC  30363

EMBOSS_001     31262 ATAATATGCTTAAACAGAGAACTCCATTCCAATAACATGAACTTTCCTTA  31311
                     ||.||||||| ||.||||
EMBOSS_002     30364 ATGATATGCT-AAGCAGA--------------------------------  30380

EMBOSS_001     31312 TGCTTTATTCATCATCGCTTGAAATTTTGAATTTTGCCCAAAGAAGTTTA  31361
                                             |.|||..|||||     ||.||
EMBOSS_002     30381 -----------------------AGTTTTCATTTT-----AATAA-----  30397

EMBOSS_001     31362 TACCAGTACATGTTAAATTACATCATAGCCTTCTTTGTATAAATCTTAGA  31411
                           ||.|.||||||||||...||||.|.|.|||||.|||..|.||||
EMBOSS_002     30398 ------TAAAAGTTAAATTACTCTATAGACATTTTTGTGTAAGCCATAGA  30441

EMBOSS_001     31412 GTAGTTTACTGAAGTACATCGCAAAGTTTTGTTGTTTCTTAGGTGATTTT  31461
                     .||..|||||.||.||..||..|.| ||||||||||.. |..||||||
EMBOSS_002     30442 ATAACTTACTAAAATATGTCATATA-TTTTGTTGTTTCAC-GAGGATTTT  30489

EMBOSS_001     31462 AATTATGTATGTTTACTTTCAGTAATGCATCTTTTCTCCTTCATCAATAT  31511
                     ||.|||||||.|||.|.|||||||||.|||.|||.|..||||||||..||
EMBOSS_002     30490 AACTATGTATATTTTCATTCAGTAATACATATTTGCCACTTCATCAGCAT  30539

EMBOSS_001     31512 TATGTTATGCTAGCTGTAAGTACAAAATAATTGAGAACAAATTATGACAA  31561
                     |..|| ||..|||.||||||.|||.|||||||||||||||  |.||||||
EMBOSS_002     30540 TGCGT-ATATTAGTTGTAAGAACAGAATAATTGAGAACAA--TCTGACAA  30586

EMBOSS_001     31562 ATTGAACCAAGCCACAAAAAAAGGAGAAACCAAATACTTTTGTGATTTGA  31611
                     |||||||.||.|||.|||.|.|  |||||..||||||...||||||||||
EMBOSS_002     30587 ATTGAACTAAACCAGAAACATA--AGAAAGAAAATAAACTTGTGATTTGA  30634

EMBOSS_001     31612 GCTTTTTTCAGTCCTTGAAACTTTAAGAATATCTGTCTTTATTAACTTTT  31661
                     ||||||.|......|||.||||||.|||||||||||||||.||||||||||
EMBOSS_002     30635 GCTTTTGTGTTCAGTTGTAACTTTGAGAATATCTGTCTTCATTAACTTTT  30684

EMBOSS_001     31662 GCTTTTTGCTGATGGTTTCTCTCATTTTATTATAGCTTATAGCATTGTAA  31711
                     ||.||.|||||.||||||||||||||||||||||||||.||||||.|||||
EMBOSS_002     30685 GCCTTCTGCTGGTGGTTTCTCTCATTTTATTATAGCTTGTAGCACTGTAA  30734

EMBOSS_001     31712 ATTAATTTAACATGAAAGGATAAAAACGTTGCTTTTGAAATGTTTCTCAT  31761
                     ||||||||||||||||||||||||||.||||||||||||||||||||||||
EMBOSS_002     30735 ATTAATTTAACATGAAAGGATAAAAATGTTGCTTTTGAAATGTTTCTCAT  30784

EMBOSS_001     31762 TAAATTATGAAAAAATATTACACTAAATAAAAGAAAGGAATGCCTCTGGT  31811
                     |||||||||.||||||||||.|.||.||||||||||||||||||||||.|
EMBOSS_002     30785 TAAATTATGGAAAAATATTATAATAGATAAAAGAAAGGAATGCCTCTGCT  30834

EMBOSS_001     31812 ACCAGCTTCTGTTTGCTCAATTATTGCAGTACCCAAAGTGAATTATTACA  31861
                     |||||||||||||||||||||||.|.|.|..|   ||.||.|||||||.||
EMBOSS_002     30835 ACCAGCTTCTGTTTGCTCAATTGTGGAATGA---AATGTAAATTATTTCA  30881

EMBOSS_001     31862 CAGTTAACTCAGAGGCAATATTATTGTCATTATATTATAAAATAGATGAG  31911
                     ..||||||.||.|||||||||||||.||||.||     ||.||.||||.
EMBOSS_002     30882 TGGTTAACACAAAGGCAATATTATTGCCATTGTA-----AAGTACATGAA  30926

EMBOSS_001     31912 TTGCAATCTTCAAAAAAAAAAAACAGCATAGGTCCTTTGAAAGTGAAATA  31961
                     ||.|.||||..||||.|||.|.|..|||.||||||||||||||||.|.|.
EMBOSS_002     30927 TTCCTATCTCTAAAAGAAATATATGGCACAGGTCCTTTGAAAGTGGACTT  30976
```

```
EMBOSS_001     31962 CCTTTTTTCCTTGTGCTTCATTTAAATATATACTGACCCCAGTTTTGTTT  32011
                     . ||||||.|||.|...|||.||||||..|..||||.|.||.    ||.|
EMBOSS_002     30977 T-TTTTTTCTTATATATCACTTAAATCCAAGCTGAACTCAA----GTCT  31021

EMBOSS_001     32012 TTGTTTTTCCTTTTTAGAGTTCTTGCTAATGATGGGCCCAAAGTTATATT  32061
                     |.||.||..|||.|||.|  ||...||||||||.|||.||.|||||||.|
EMBOSS_002     31022 TAGTGTTGTCTTCTTATA--TCCGCCTAATGATTGGCTCAGAGTTATACT  31069

EMBOSS_001     32062 AAGAACTGC--AAAGTAAATTTCAACCAATTACTTTATTCAGGGGAGTCA  32109
                     .|.||.||.  ||||.|||||...||||||||.|.|||||||..||.|.|
EMBOSS_002     31070 TAAAAGTGGTAAAAGAAAATTCTGACCAATTATTATATTCAGAAGAATAA  31119

EMBOSS_001     32110 TTAAATTGAGGTACCTCTGAAATTTTGGAAGGAATGTACTGCCAATTAGC  32159
                     |||.||.||.||||||.||.||||||||||.|||||.||.|||.||||||.|
EMBOSS_002     31120 TTATATGGAAGTACCTATGGAATTTTGGAATGAATATATTGCTAATTATC  31169

EMBOSS_001     32160 CGAAAGCACTACTCAATGTCCTTTCTATGGTTATAATCTCTCTAGTGTAT  32209
                     .||||||||||.||||.......||||||||||...|||||.|.||.||.
EMBOSS_002     31170 AGAAAGCACTATTCAACACTAGCTCTATGGTTATGCCCTCTTTTGTATAG  31219

EMBOSS_001     32210 TTTTAATTGAAGACAACCTCTATAGAGGAGGTGAGAAGTTGCTATTTATT  32259
                     |||||||.|||.|||  .||||.||||||..||||||.||||||.|||||.|
EMBOSS_002     31220 TTTTAATAGAAAACA--GTCTACAGAGGAACTGAGAAATTGCTGTTTAGT  31267

EMBOSS_001     32260 GGTACTTGTTAGGATGGAATCAAGGGTGTGGAAGATATTCATCTATTTCT  32309
                     .||||||||||..||||||.|||..|||...|||.|||||||.|||||||
EMBOSS_002     31268 AGTACTTGTTAAAATGGAACCAAAAGTGGAAAAGGTATTCATTTATTTCT  31317

EMBOSS_001     32310 CTCTCCAGCTCCCCCACACAAAAAGAATGGTGCTTAATCCATCTGAAGCA  32359
                                                   |..||..||||.||.|
EMBOSS_002     31318 ------------------------------TATTTTTTCCAACTCA----  31333

EMBOSS_001     32360 TTTGGGGAGCGAGGGTAAAGATGTAATATTTACCATGAGCCGAAACAGAT  32409
                         |..||.||..||.|..|||||   ||||..||.|.|..||||||..|
EMBOSS_002     31334 ----GAAAGAGACAGTTACAATGTA---TTTATTATCAACTAAAACAATT  31376

EMBOSS_001     32410 CTTCAGAAGTGGA-AAATGGAAGCATATTGAAGTCCCTCAACTAAACAGA  32458
                     ..|.|.||.||.| ||||..||||...||||||.||||||||||.|.. |
EMBOSS_002     31377 GCTGAAAATTGAAGAAATACAAGCTGTTTGAAGCCCCTCAACTAGATC-A  31425

EMBOSS_001     32459 CTTTCTTCCATATGGAATTCAATGCATTAATGTTTTCAAATTCTATAGCT  32508
                     .|||.|||||.|||.||| |||..|||...|||.|.||.|.|..||||.
EMBOSS_002     31426 ATTTTTTCCACATGAAAT-CAACACATAGGTGTATACATAATGCATAGA-  31473

EMBOSS_001     32509 TCAAATTCTTAATATTTTCAAATTATGTGAGCTTATGTCAAAACATTTAA  32558
                       |||                        |||||||.|       |.|||
EMBOSS_002     31474 --AAA------------------------AGCTTATAT-------TGTAA  31490

EMBOSS_001     32559 GTGAGCTTTTAACAATGAGGCAAATATTTGAAT--CATTTGTCTACA  32603
                     ||||||||||||.||||..|||||.||||||.|||  |||.|.||||||
EMBOSS_002     31491 GTGAGCTTTTAATAATATGGCAGATATTTAAATAGCATATCTCTACA  31537


#---------------------------------------
#---------------------------------------
```

8.2. APPENDIX B: ATTEMPT TO ESTABLISH A CPF1 SYSTEM.

A Cpf1 system was assessed to attempt deletion of exons 19 to 55 of the *DMD/Dmd* genes. Cpf1 was considered an alternative candidate (to our *Sa*Cas9 system) as its size would also allow for packaging in an AAV vectors. Furthermore, the sticky end generated by this system after a double strand DNA break, would allow for insertion of a repair template through micro mediated homology end joining, if needed. Guide RNAs targeting introns 18 and 55 of human and mouse *DMD/Dmd* genes were designed with online tools (as described in Section 4.1.1). Cloning of selected gRNAs was attempted. Once gRNAs were cloned in a suitable plasmid, they were screened *in-vitro* in human and mouse cell lines respectively (HEK293T and N2A cells). Unfortunately, none of the screened gRNA showed any activity so it was decided to proceed further experiments in this project only with the *Sa*Cas9 system.

In this section cloning strategies attempted to establish a Cpf1 systems are presented, alongside results from Cpf1 gRNA screening *in-vitro*.

8.2.1. ATTEMPT TO ESTABLISH A CPF1 SYSTEM BY CLONING.

To screen Cpf1 gRNAs *in-vitro* two plasmids were needed. One of them expressing the gRNAs (Fig. 8.1) and the second one expressing the Cpf1 protein (Fig. 8.2).

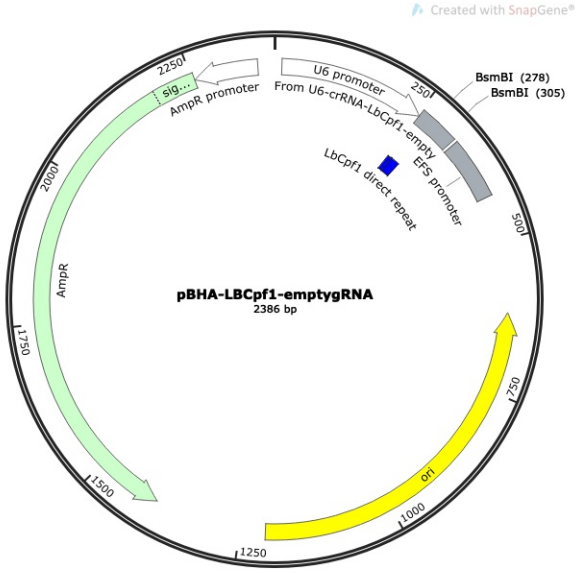


Figure 8.1. pBHA-LbCpf1-emptyRNA (also called p-empty-Cpf1). Plasmid expressing an ampicillin resistance cassette and a U6 promoter driving gRNA expression. Guide RNAs were cloned in with BsmBI restriction enzyme.

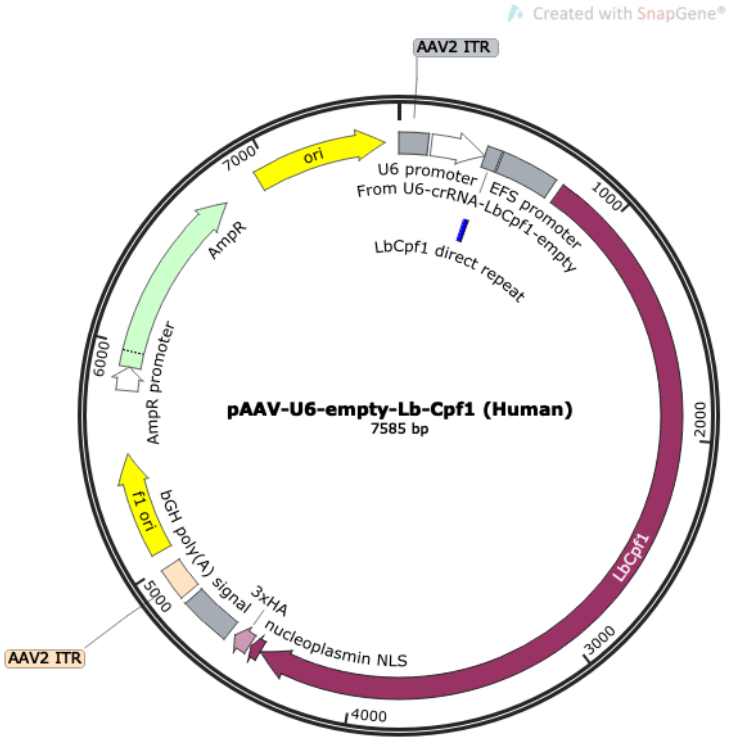
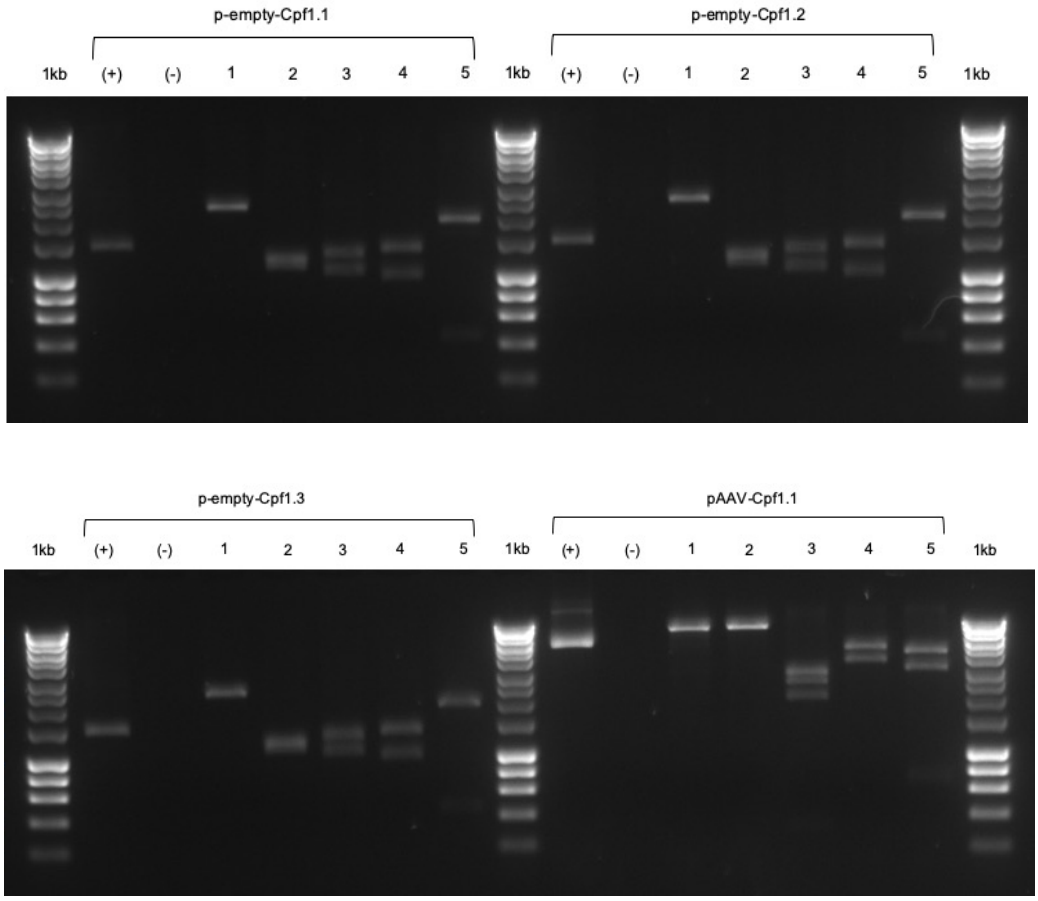


Figure 8.2. Plasmid expressing an *Lb*Cpf1 protein under an EFS promoter in an AAV backbone.

Integrity of both Cpf1 plasmids was confirmed with the respective restriction enzyme digests before starting the cloning experiments. Results from the restriction digestions from p-empty-Cpf1 and pAAV-Cpf1, showing the expected fragment sizes, can be seen in Figure 8.3.



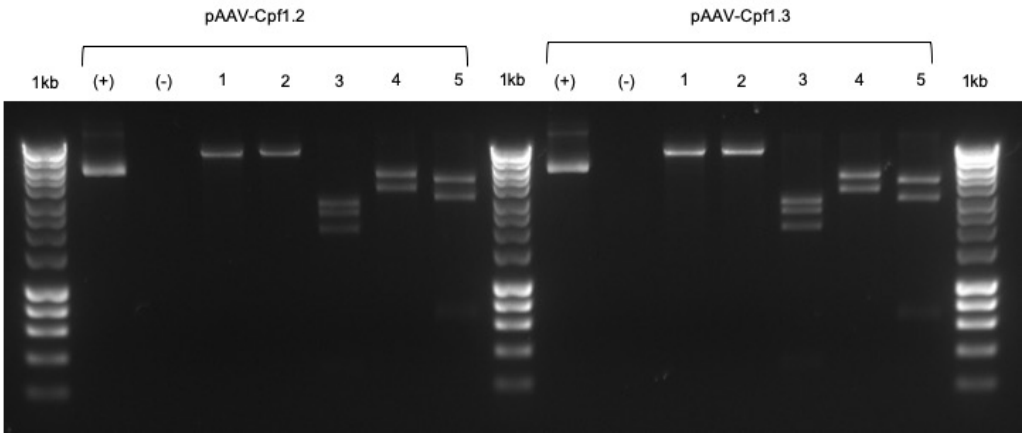


Figure 8.3. Gel Images from p-empty-Cpf1 and pAAV-Cpf1 restriction digests. 1% (w/v) agarose gel with 0.5X SYBR Safe in 1X TAE (Tris-Acetate–EDTA) Buffer. From left to right for p-empty-Cpf1: Hyperladder I from Bioline, positive control (undigested plasmid), negative control (enzyme only). Lane 1 - BamHI: 1. 2386 bp. Lane 2- ApaLI: 1. 1261 bp, 2. 1125 bp. Lane 3- BanI: 1. 1327 bp, 2. 1059 bp. Lane 4 - BspHI: 1. 1363 bp, 2. 1023 bp. Lane 5 - EcoRI: 1. 1960 bp, 2. 426 bp. (Obtained from SnapGene). Expected band sizes for For pAAV-Cpf1, from left to right: Lane 1 - EcoRI: 1. 7585 bp. Lane 2 - BamHI: 1. 4059 bp, 2. 3526 bp. Lane 3 - MscI: 1. 2820 bp, 2. 2457 bp, 3. 2002 bp, 4. 306 bp. Lane 4 - NdeI: 1. 4223 bp, 2. 3362 bp. Lane 5 - ScaI: 1. 3806 bp, 2. 31136 bp, 3. 666 bp. (Obtained from SnapGene).

All the mini-preps of p-empty-Cpf1 showed the expected band pattern, confirming plasmid integrity. However, the bands obtained from pAAV-Cpf1 restriction digest with BamHI did not match the expected band pattern, showing one large band with the same size as the band digested with a single cutter enzyme. The whole plasmid was sent for sequencing with staggered primers covering the whole plasmid sequence. Sequencing results are shown in Figures 8.4 and 8.5.

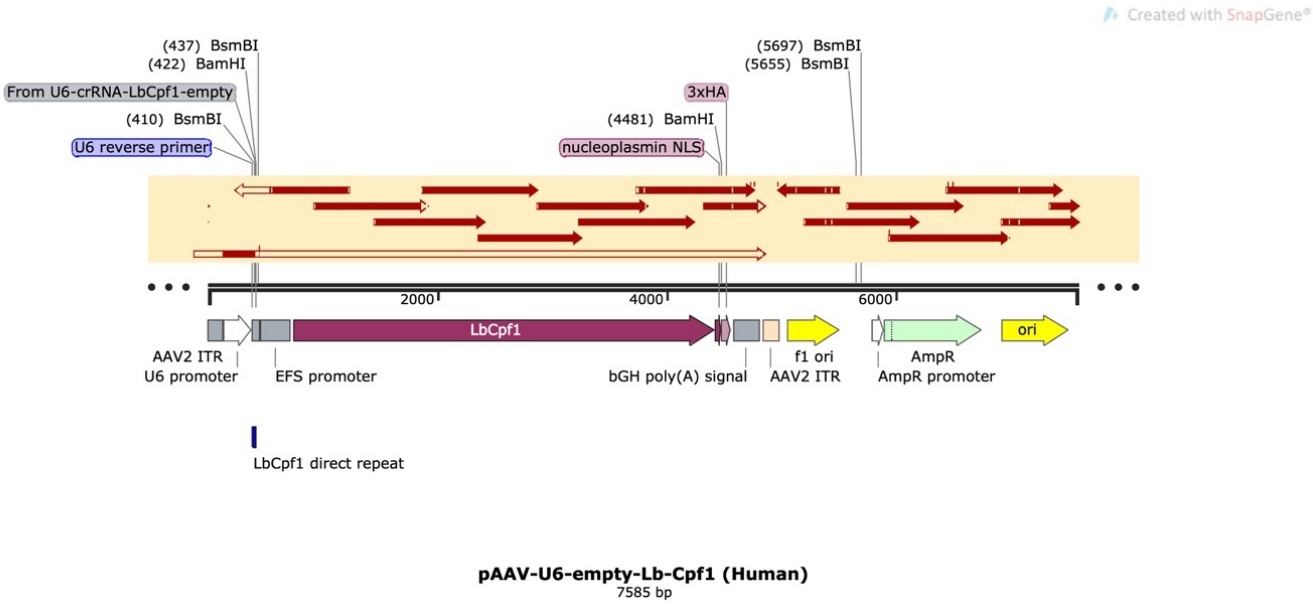


Figure 8.4. Sequencing results from pAAV-Cpf1 plasmid samples aligned against pAAV-Cpf1 plasmid map. The whole plasmid was sequenced with 17 staggered sequencing primers to confirm plasmid integrity. Alignments were performed on SnapGene.

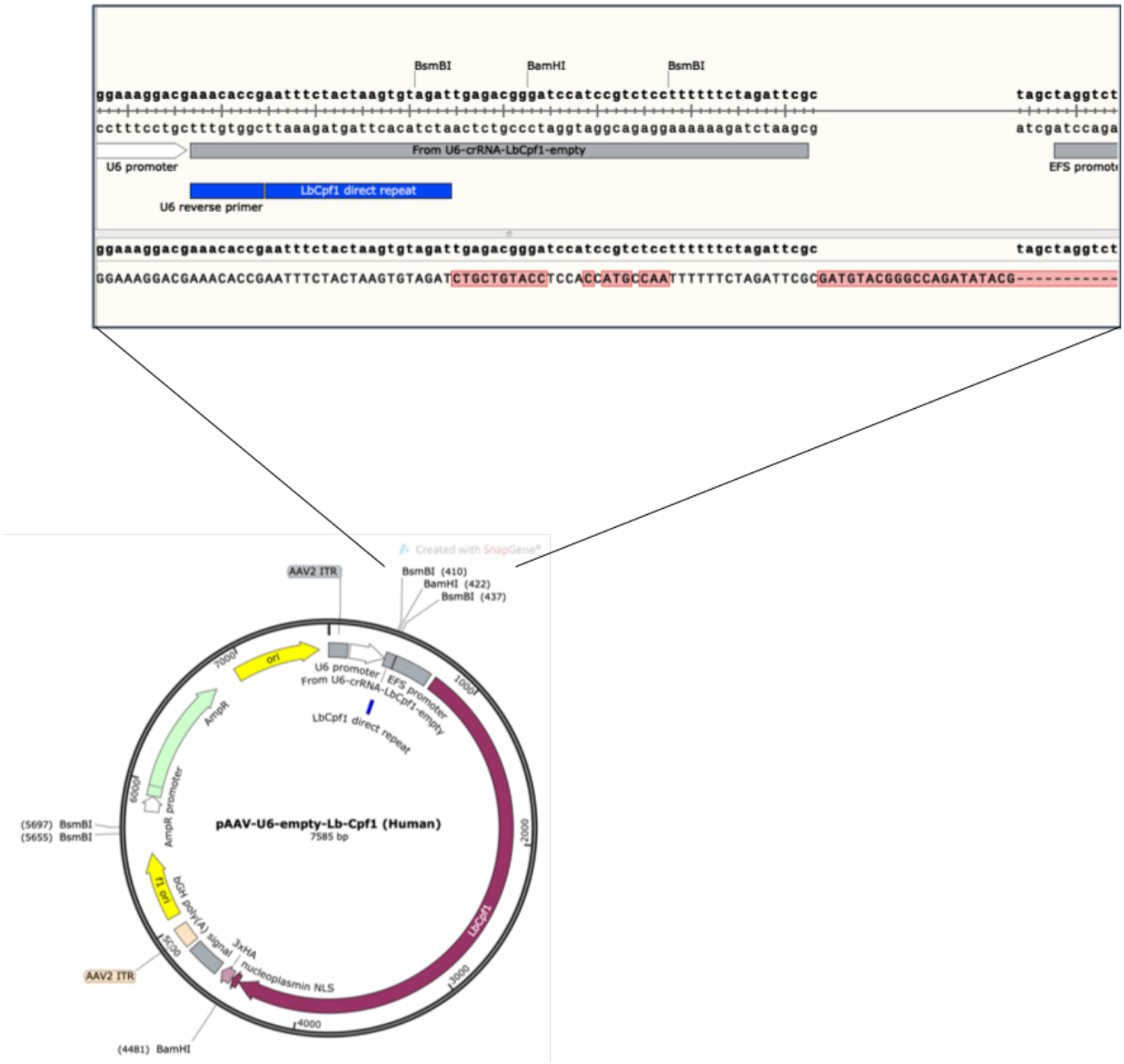


Figure 8.5. Zoom in of pAAV-Cpf1 plasmid and sequence alignment of plasmid map and sequenced samples at the region across the BsmBI restriction sites. Sequencing trace from sample showing BamHI site missing in the plasmid, which is located in the region that would be cut out to clone in the CRISPR gRNAs. Alignment performed on SnapGene.

Based on the sequencing results, it was concluded that one of the BamHI restriction sites was not present in the pAAV-Cpf1 plasmid. Nevertheless, the missing site was within the

region that would be cut out to clone the CRISPR gRNAs in and therefore would not affect the plasmid appropriateness for gRNA cloning.

Once the integrity of the plasmids was confirmed, the goal was to construct a plasmid expressing Cpf1 where the gRNAs could be cloned in directly, therefore the cloning strategy depicted in figure 8.6 was attempted. The goal was to construct a plasmid from pAAV-Cpf1 without a BsaI site in the Ampicillin resistance gene and to replace the BsmBI sites with BsaI sites, so BsaI could be used for direct gRNA cloning into this new plasmid instead of having to clone the gRNAs into p-empty-Cpf1 and then subclone them into pAAV-Cpf1 or co-deliver them.

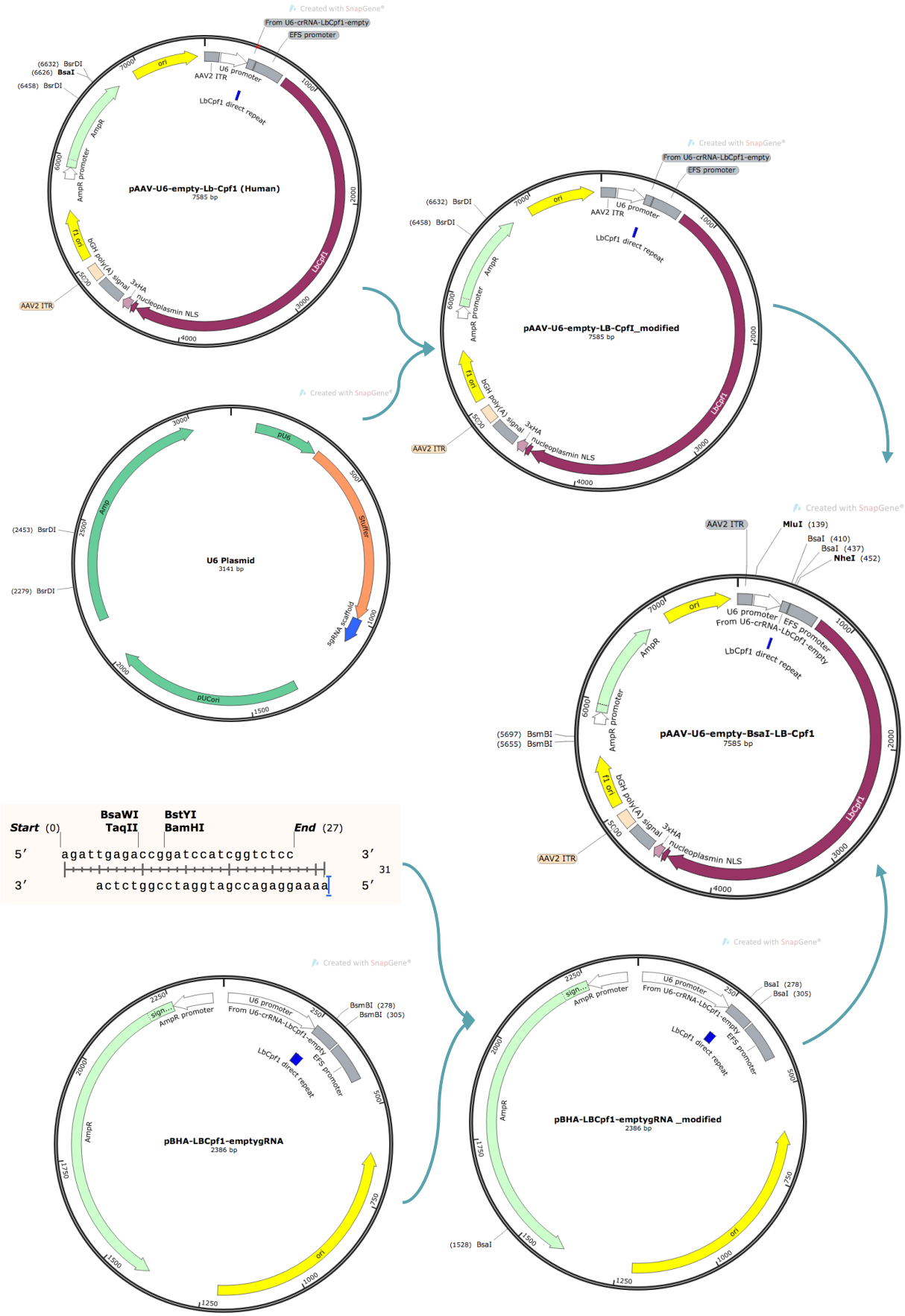

Created with SnapGene®
From U6-crRNA-LbCpf1-empty
EFS promoter
(6632) BsrDI
(6626) BsaI
(6458) BsrDI
ori
AAV2 ITR
U6 promoter
LbCpf1 direct repeat
AmpR
AmpR promoter
pAAV-U6-empty-Lb-Cpf1 (Human)
7585 bp
f1 ori
bGH poly(A) signal
3xHA
nucleoplasmin NLS
LbCpf1
AAV2 ITR
pAAV-U6-empty-LB-CpfI_modified
7585 bp
U6 Plasmid
3141 bp
pU6
Stuffer
sgRNA scaffold
Amp
pUCori
(2453) BsrDI
(2279) BsrDI
MluI (139)
BsaI (410)
BsaI (437)
NheI (452)
pAAV-U6-empty-BsaI-LB-Cpf1
7585 bp
(5697) BsmBI
(5655) BsmBI
BsaWI
BstYI
TaqII
BamHI
Start (0)
End (27)
5′ agattgagaccggatccatcggtctcc 3′
3′ actctggcctaggtagccagaggaaaa 5′
31
pBHA-LBCpf1-emptygRNA
2386 bp
BsmBI (278)
BsmBI (305)
pBHA-LBCpf1-emptygRNA _modified
2386 bp
BsaI (278)
BsaI (305)
(1528) BsaI

Figure 8.6. Cloning strategy to generate pAAV-Cpf1-modified. This strategy was aiming to clone a plasmid expressing Cpf1 where CRISPR gRNAs could be cloned in directly in order to avoid a two-step cloning strategy for each gRNA. To achieve this, a region from the ampicillin resistance gene not containing a BsaI site would be recovered from plasmid U6 (pU6) in order to replace the ampicillin region containing a BsaI site on pAAV-Cpf1. Then, the annealed oligonucleotides indicated in the figure would be cloned into p-empty-Cpf1, so this construct could then be sub-cloned into the modified pAAV-Cpf1. Guide RNAs could then be directly cloned into this final construct.

The first part of the cloning strategy aimed to modify p-empty-Cpf1 to switch BsmBI sites to BsaI sites for gRNA cloning. In order to achieve this, the following oligonucleotides containing BsaI sites with the appropriate overhangs to be cloned with BsmBI were designed and ordered from IDT: 5'-agattgagaccggatccatcggtctcc-3' 5'-aaaaggagaccgatggatccggtctca-3'. Then, p-empty-Cpf1 was digested with BsmBI to recover the backbone (Fig. 8.7) and clone the annealed oligonucleotides in. The goal was to sub-clone this modified plasmid into a modified pAAV-Cpf1 in which BsaI sites located in the ampicillin resistance gene were removed in order to avoid cutting the plasmid when the gRNAs were cloned in.

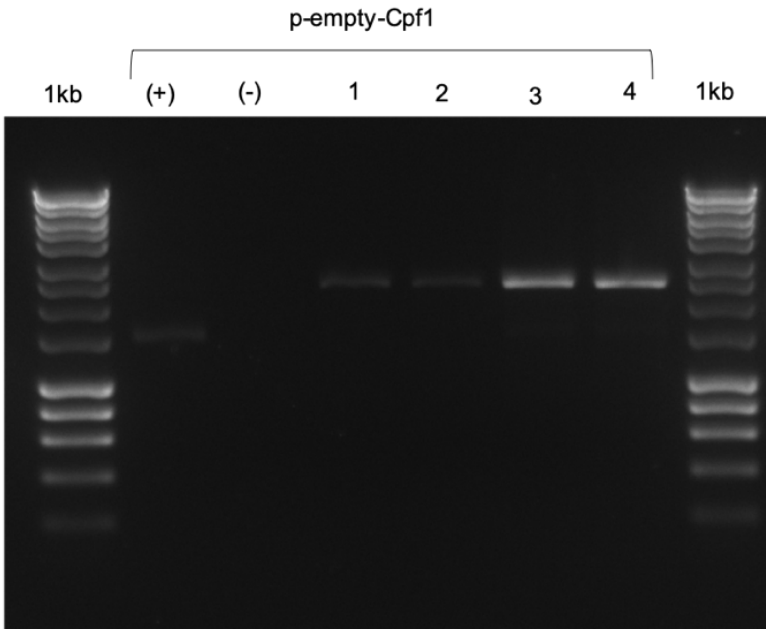


Figure 8.7. Gel Image from p-empty-Cpf1 restriction digestion. 1% (w/v) agarose gel with 0.5X SYBR Safe in 1X TAE (Tris-Acetate–EDTA) Buffer. From left to right: Hyperladder I from Bioline, positive control (undigested plasmid), negative control (enzyme only), Lane 1: BamHI, Lanes 2, 3 and 4: BsmBI

After the backbone was recovered, it was ligated with the annealed oligos containing BsaI sites. Clones 1, 3 and 4 were digested (Fig. 8.8) to confirm a successful cloning of the fragment. Clone pCpf1.2 showed an unexpected fragment size when digested with BsaI (Fig. 8.8, pCpf1.2, band 3).

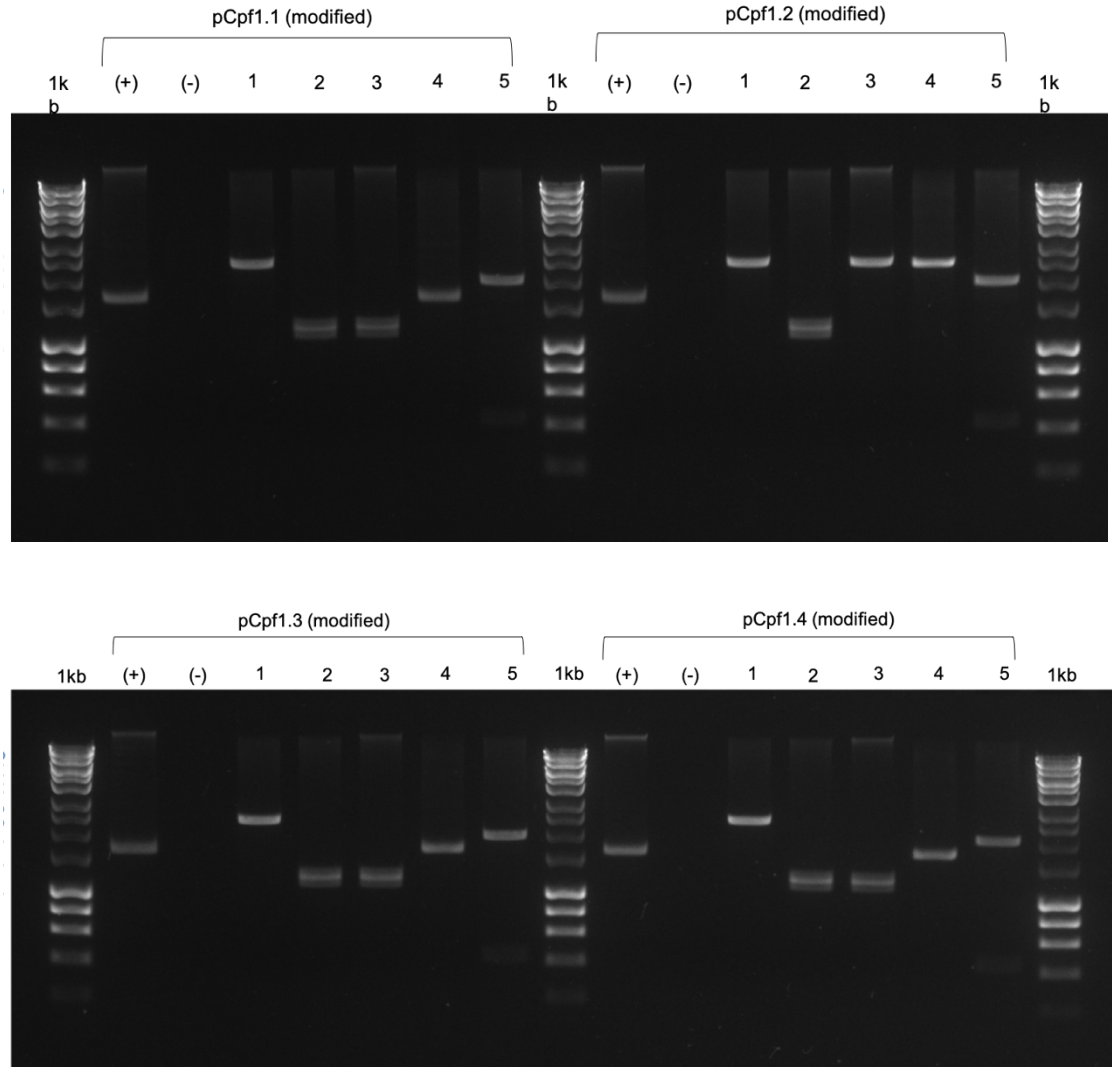


Figure 8.8. Gel Images from four clones of p-empty-Cpf1-modified restriction digestions. 1% (w/v) agarose gel with 0.5X SYBR Safe in 1X TAE Buffer. From left to right: Hyperladder I from Bioline, positive control (undigested plasmid), negative control (enzyme only), Lane 1: NdeI, Lane 2: ApaLI, Lane 3: BsaI, Lane 4: BsmBI, Lane 5: EcoRI. Clones 1, 3 and 4 match the following expected band pattern: Lane 1. NdeI: 2,386 bp, Lane 2. ApaLI: 1,262 bp, 1,125 bp, Lane 3. BsaI: 1,223 bp, 1,136 bp, 27 bp, Lane 4. BsmBI: noncutter, Lane 5. EcoRI: 1,960 bp, 426 bp.

Once p-empty-Cpf1 was successfully modified, the other half of the cloning strategy from Fig. 8.6 was attempted. To achieve this, a region from the ampicillin resistance gene not containing a BsaI site was recovered from plasmid U6 (pU6) in order to replace the ampicillin region containing a BsaI site on pAAV-Cpf1. The backbone from pAAV-Cpf1

was successfully recovered digesting the plasmid with BsrDI (Figs. 8.9). The ampicillin region from U6 was successfully recovered (Fig. 8.19). Nevertheless, it was not possible to ligate the ampicillin region recovered from pU6 to the backbone even though the cloning was attempted twice. In total 16 clones were tested and none of them showed a proper integration and plasmid integrity (Figs. 8.11 and 8.12). Therefore, a different strategy was attempted afterwards.

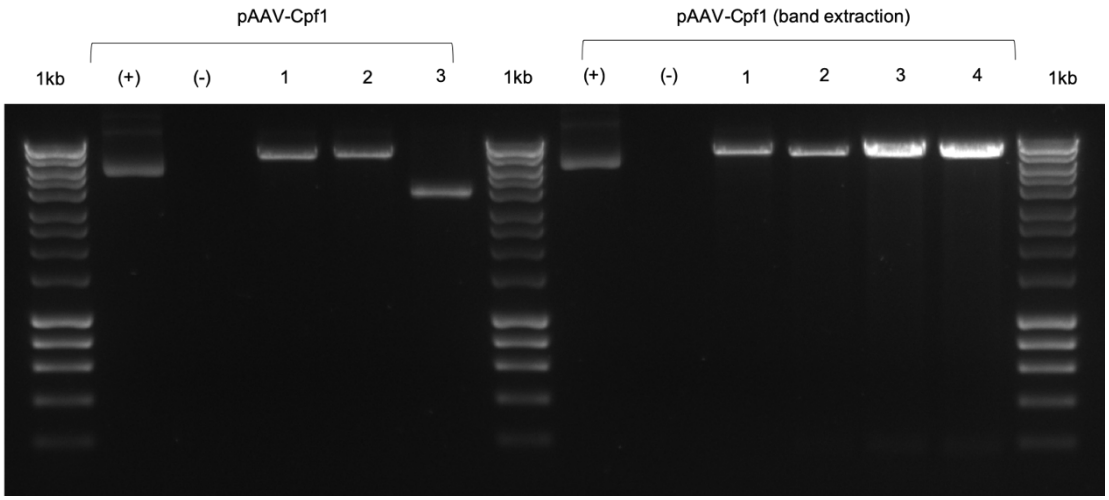


Figure 8.9. Gel Image from pAAV-Cpf1 restriction digestion. 1% (w/v) agarose gel with 0.5X SYBR Safe in 1X TAE Buffer. From left to right: Hyperladder I from Bioline, positive control (undigested plasmid), negative control (enzyme only) and the following enzymes: 1. BamHI, 2. KpnI, 3. BamHI + KpnI, confirming plasmid integrity. For band extraction: 1. EcoRI, 2, 3 and 4. BsrDI.

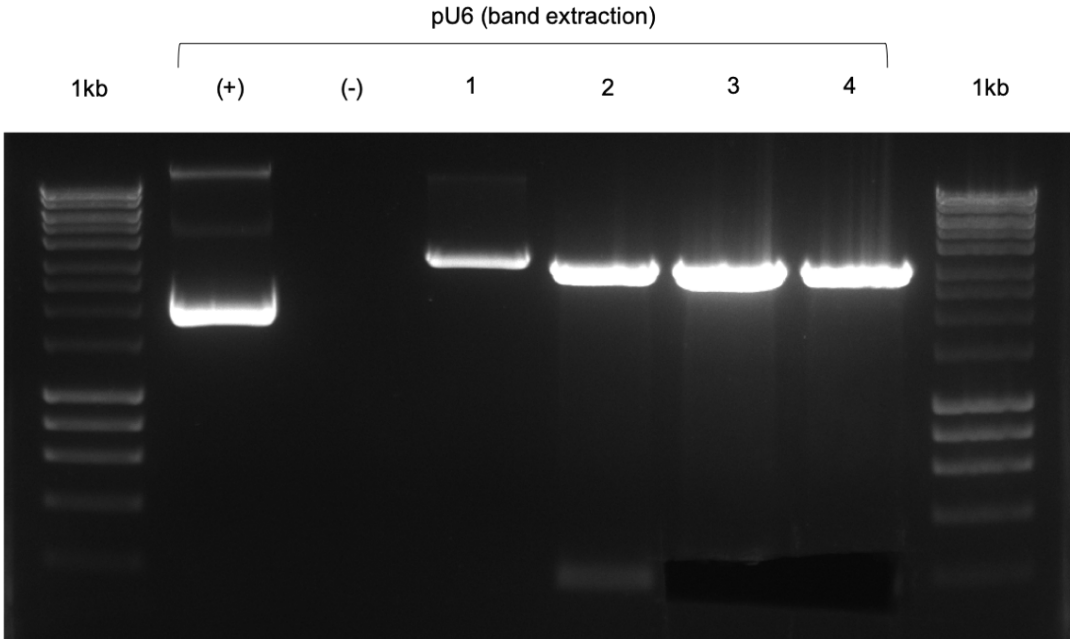


Figure 8.10. Gel Image from pU6 after band extraction. 1% (w/v) agarose gel with 0.5X SYBR Safe in 1X TAE Buffer. From left to right: Hyperladder I from Bioline, positive control (undigested plasmid), negative control (enzyme only) and the following enzymes from left to right: 1. HindIII, 2. BsrDI, 3. BsrDI and 4. BsrDI digestion of 2000 ng of DNA for band extraction.

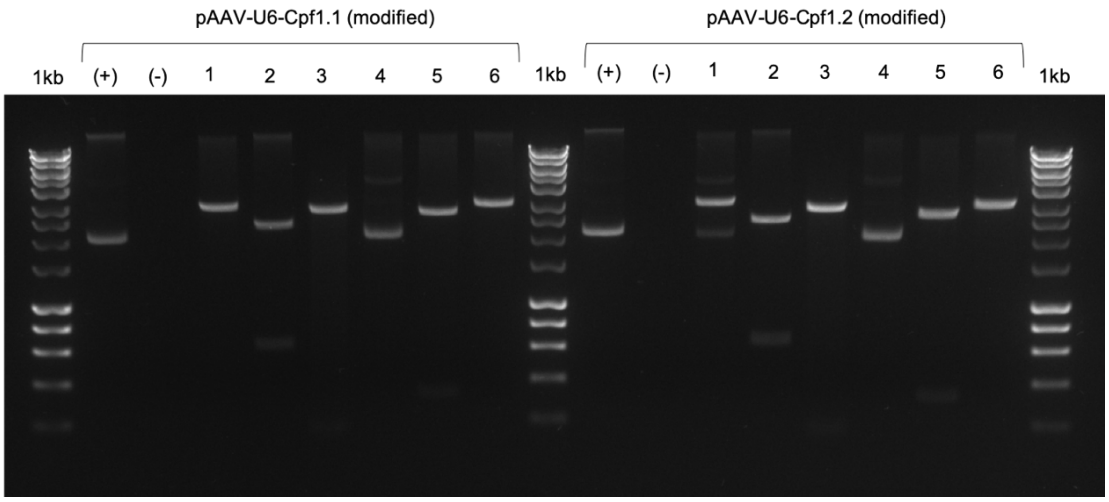


Figure 8.11. Restriction digestion of two clones of pAAV-U6-Cpf1-modified. 1% (w/v) agarose gel with 0.5X SYBR Safe in 1X TAE Buffer. From left to right: Hyperladder I from Bioline, positive control (undigested plasmid), negative control (enzyme only) and the following enzymes from left to right: 1. SphI, 2. BsaI, 3. BsrDI, 4. MscI, 5. NdeI and 6. ScaI. The obtained fragment did not match the expected fragment sizes: 1. SphI: 7,585 bp, 2. BsaI: non-cutter, 3. BsrDI: 7,411 bp, 174 bp, 4. MscI: 2,820 bp, 2,457 bp, 2,002 bp, 306 bp, 5. NdeI: 4,223 bp, 3,362 bp and 6. ScaI: 3,806 bp, 3,313 bp, 666 bp (Obtained from SnapGene).

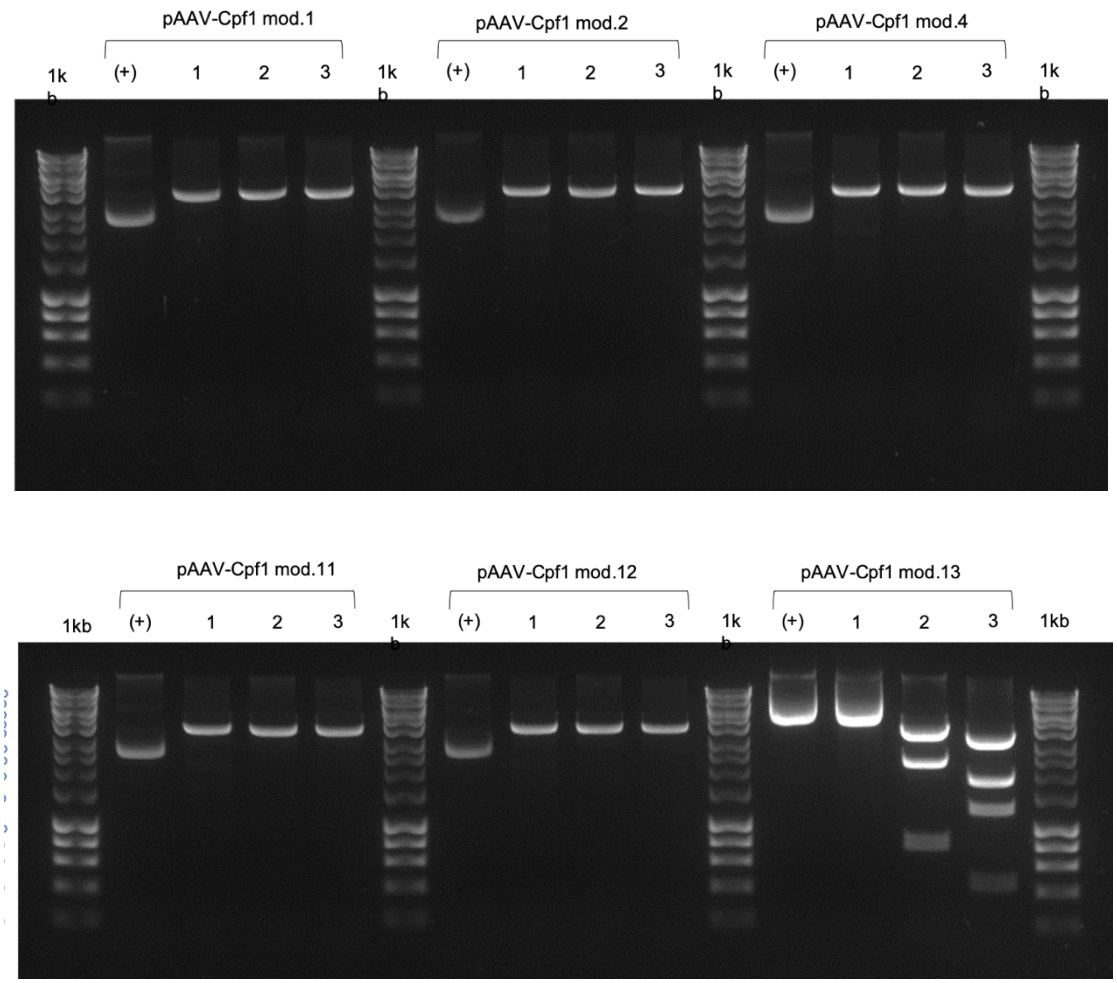


Figure 8.12. Restriction digestions of some of the 16 clones of pAAV-U6-Cpf1-modified. 1% (w/v) agarose gel with 0.5X SYBR Safe in 1X TAE Buffer. Using the following enzymes for a quick scan, from left to right: Hyperladder I from Bioline, positive control (undigested plasmid), negative control (enzyme only), 1. EcoRI-HF, 2. BsaI, 3. MscI. The obtained fragments did not match the expected fragment sizes: 1. EcoRI-HF: 7,585 bp, 2. BsaI: non-cutter, 3. MscI: 2,820 bp, 2,457 bp, 2,002 bp, 306 bp (Obtained from SnapGene).

In order to avoid potential issues while ligating the fragment recovered from the U6 plasmid, a strategy using a g-block to replace the ampicillin region containing the BsaI site was attempted. Instead of recovering a region of the ampicillin sequence from U6 plasmid, a double stranded block of DNA containing the desired sequence was ordered from IDT (Fig. 8.13) and a cloning strategy with NEB HiFi builder kit was attempted.

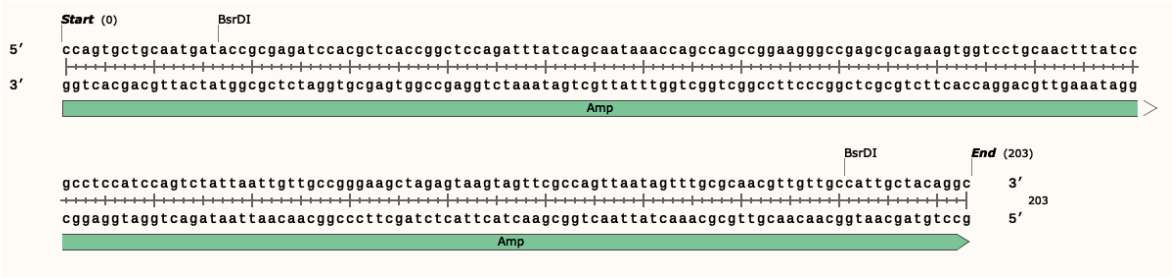


Figure 8.13. G-block design of a double stranded DNA Ampicillin fragment. Designed using SnapGene and ordered from IDT.

The cloning of the g-block with the NEB HiFi builder kit did not work, as shown on Fig. 8.14. Therefore, a different approach was tested.

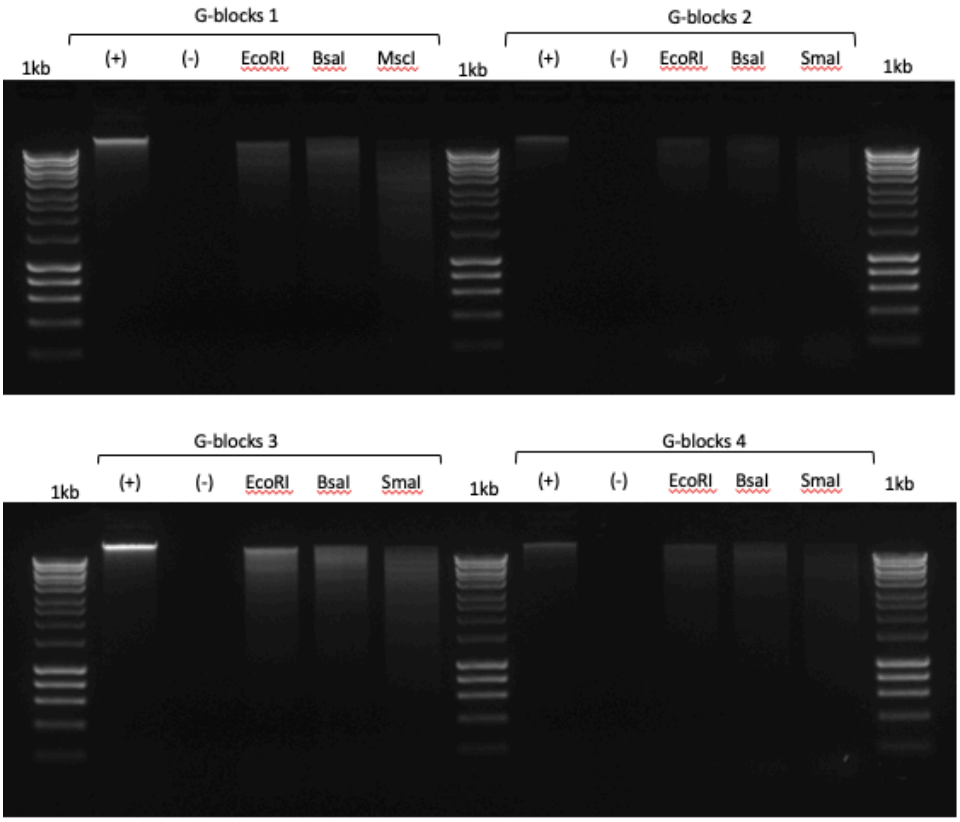


Figure 8.14. Restriction digestions of some of the clones obtained from the g-blocks assembly. 1% (w/v) agarose gel with 0.5X SYBR Safe in 1X TAE. None of the clones seemed to have worked.

A two-step cloning strategy using the original Cpf1 plasmids was designed (Fig. 8.15), in which the gRNAs were cloned into p-empty-Cpf1 with BsmBI and a fragment containing the gRNAs would be sub-cloned into pAAV-Cpf1 using NdeI and MluI once the guides were tested *in-vitro*.

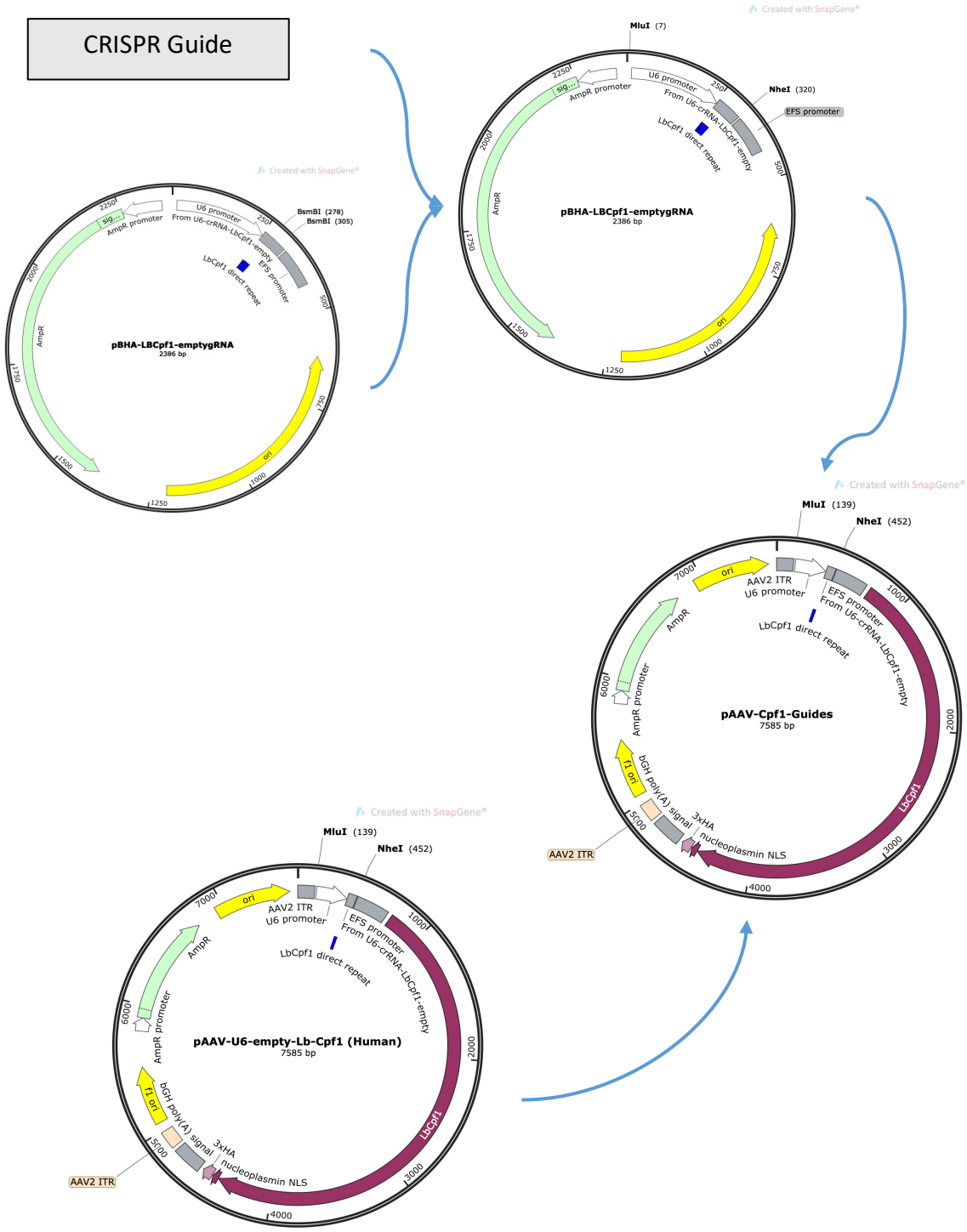


Figure 8.15. Two-step cloning strategy to clone Cpf1 CRIPSR gRNAs. This strategy was aiming to clone Cpf1 CRISPR gRNAs into p-empty Cpf1 with BsmBI and then subclone the region with the gRNAs into pAAV-Cpf1 using NdeI and MluI.

Before cloning the gRNAs into p-empty-Cpf1, the plasmid was maxi-prepped and its integrity was confirmed (Fig. 8.16). Then, a preparative restriction digestion was performed, and the plasmid backbone extraction was confirmed by gel imaging (Figs. 8.17 and 8.18). Plasmid pAAV-Cpf1 was also maxi-prepped and plasmid integrity was confirmed by restriction digestion (Fig. 38).

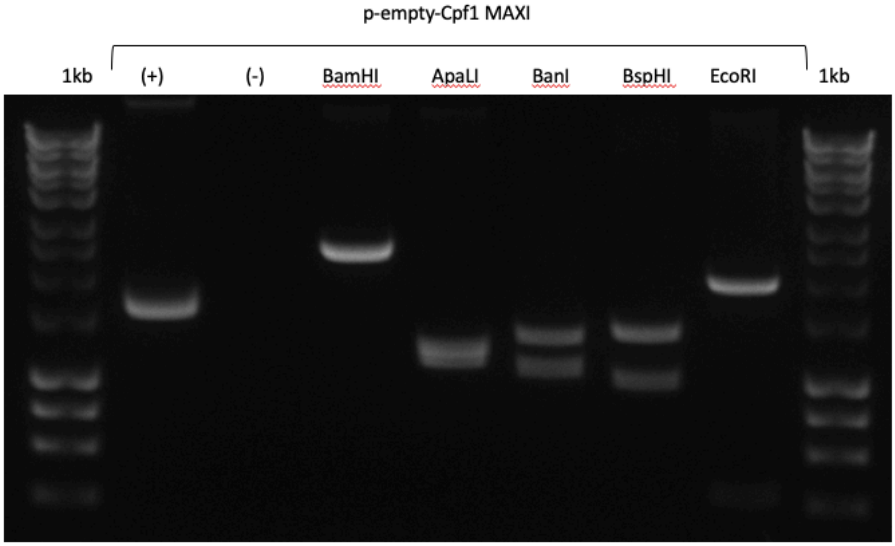


Figure 8.16. Gel Image from p-empty-Cpf1 maxi-prep restriction digestion. 1% (w/v) agarose gel with 0.5X SYBR Safe in 1X TAE Buffer. Plasmid integrity was confirmed. From left to right: Hyperladder I from Bioline, positive control (undigested plasmid), negative control (enzyme only) and the enzymes labelled on the image. Fragments show the following expected fragments sizes: BamHI: 1. 2,386 bp. ApaLI: 1. 1,261 bp, 2. 1,125 bp. BanI: 1.1,327 bp, 2. 1,059 bp. BspHI: 1. 1,363, 2. 1023. EcoRI: 1. 1,960, 2. 426 bp obtained from SnapGene.

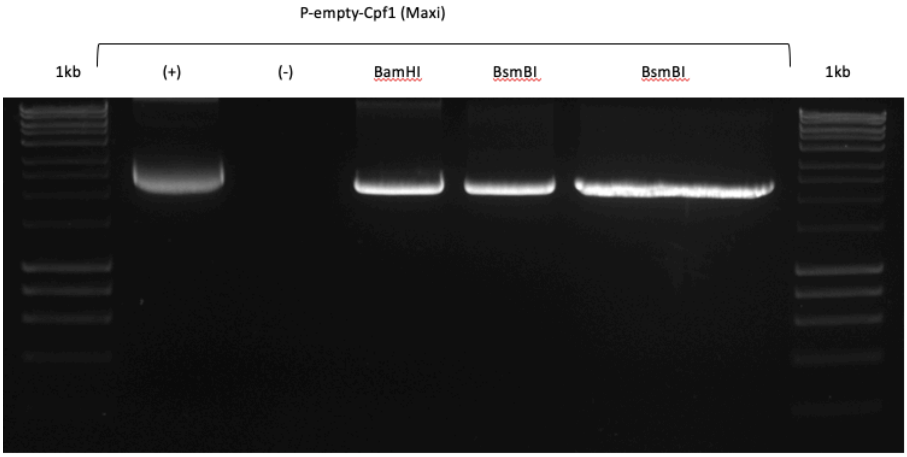


Figure 8.17. Gel Image from pAAV-Cpf1 (maxi-prepped) preparative restriction digestion. 1% (w/v) agarose gel with 0.5X SYBR Safe in 1X TAE Buffer. Plasmid digested with BamHI as a control and with BsmBI to recover the backbone.

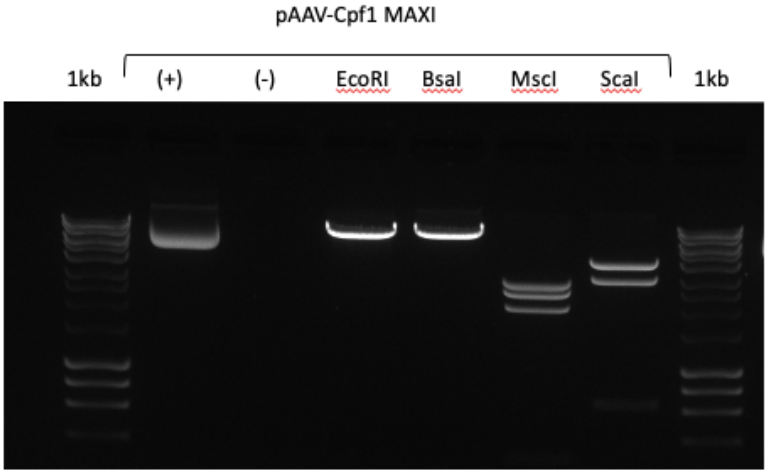
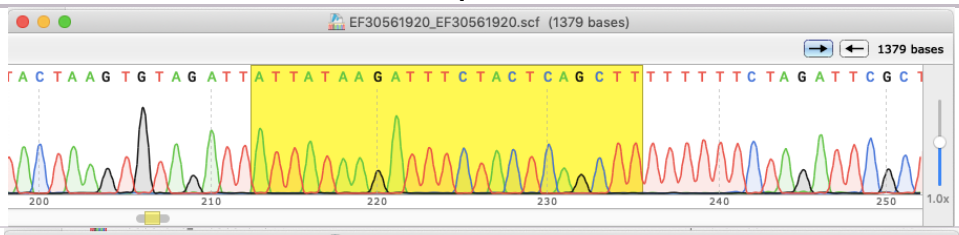
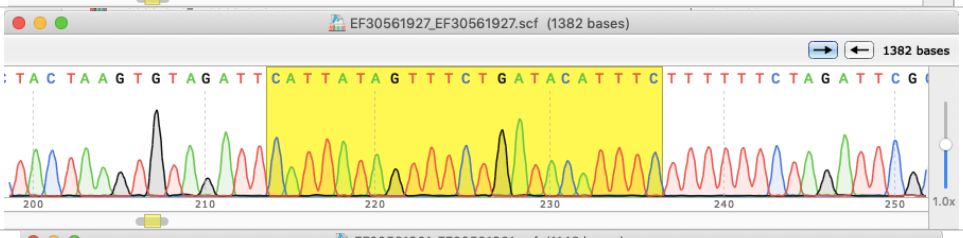
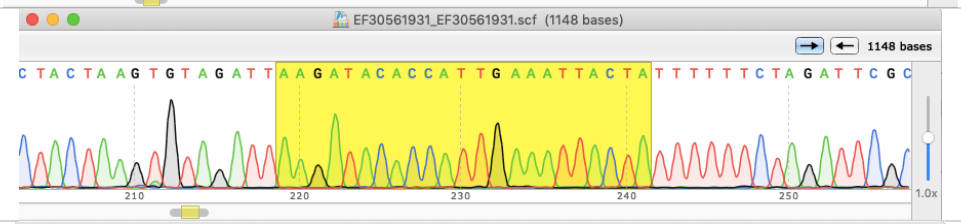
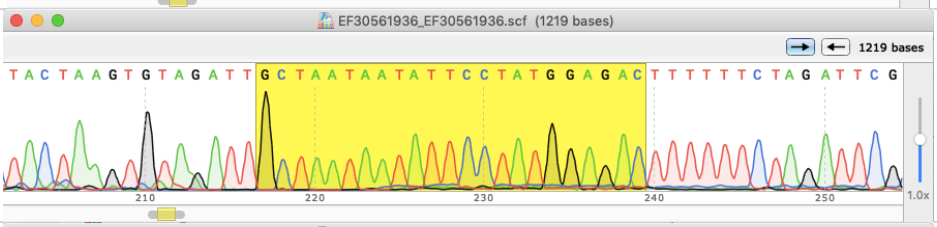
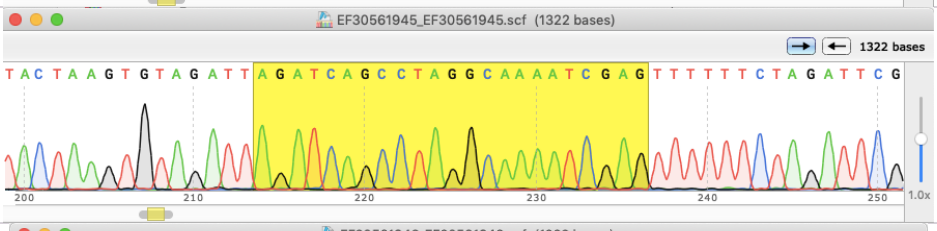
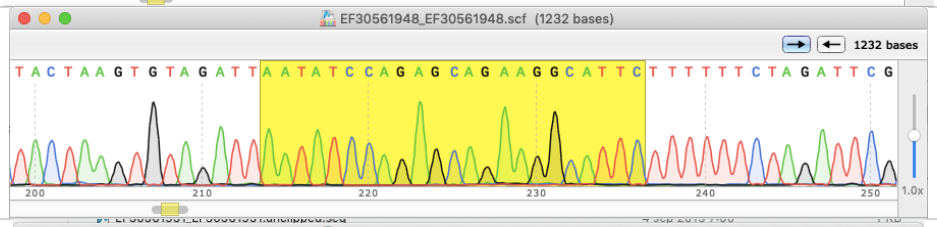
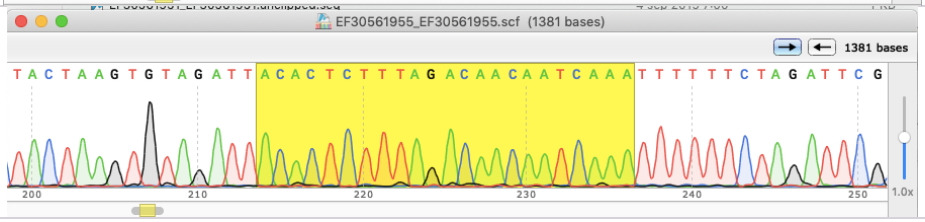
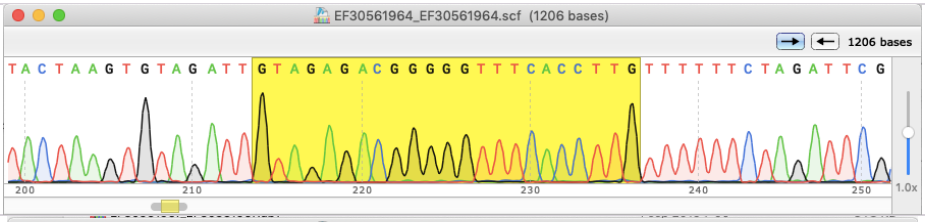
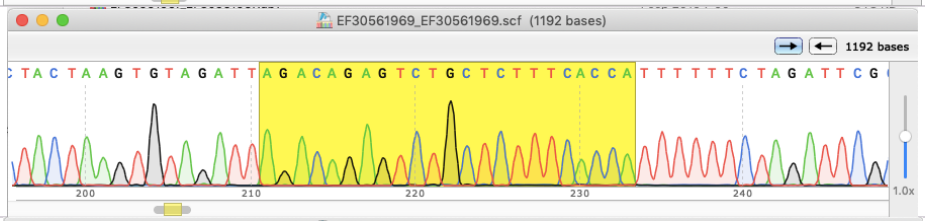
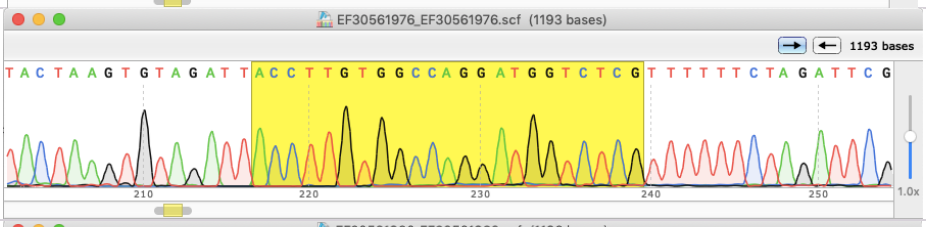
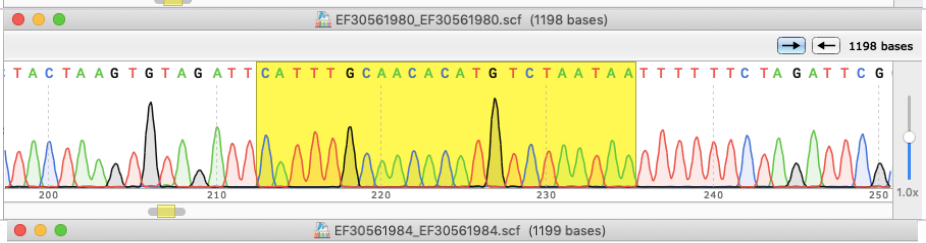
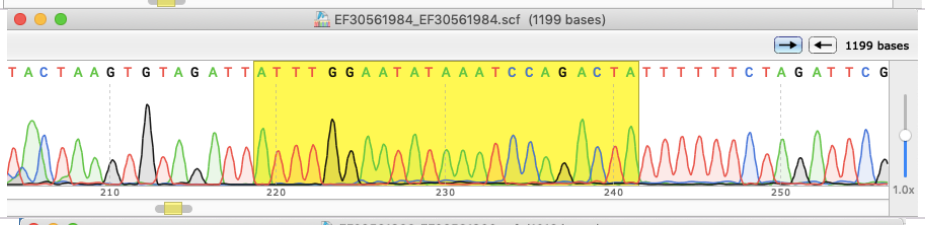
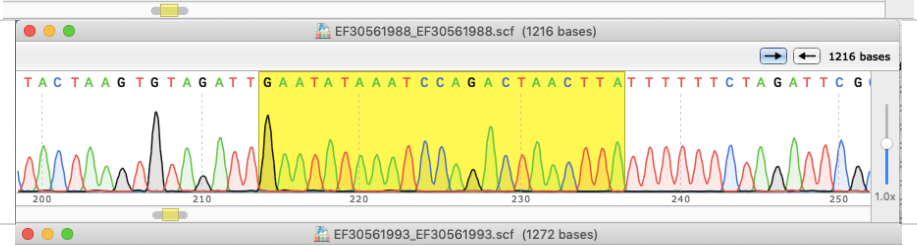
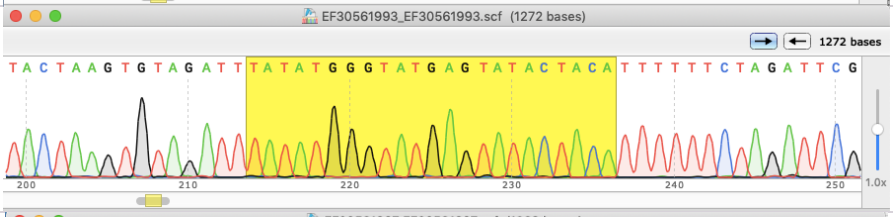
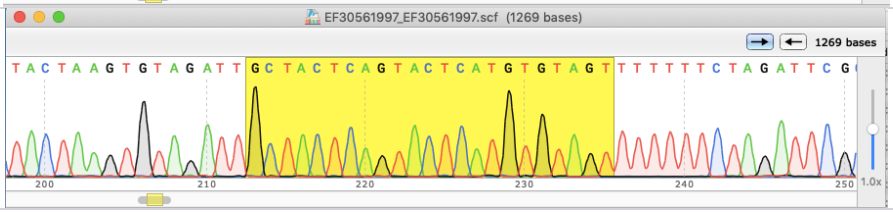


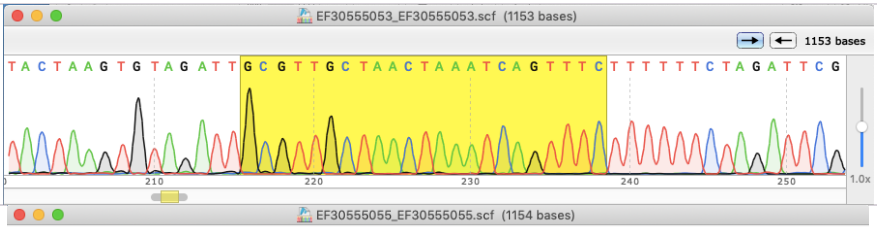
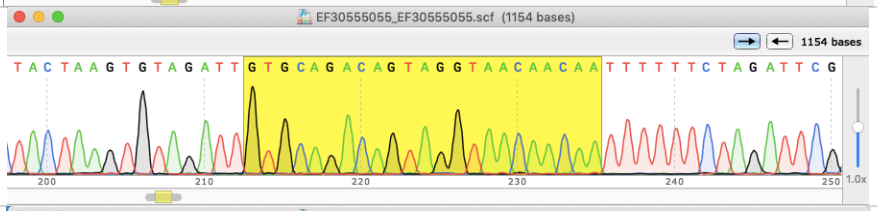
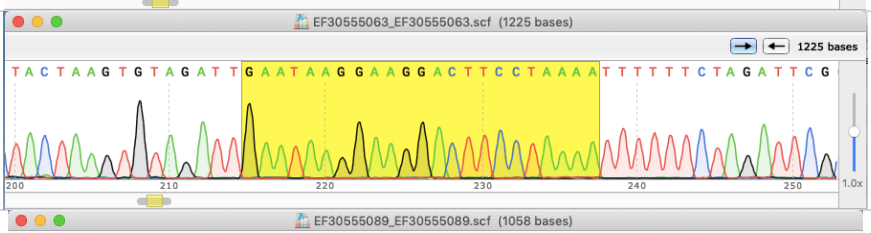
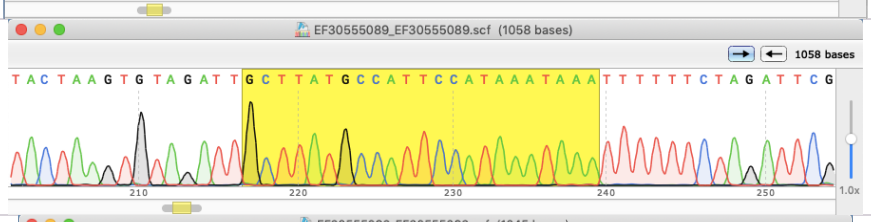
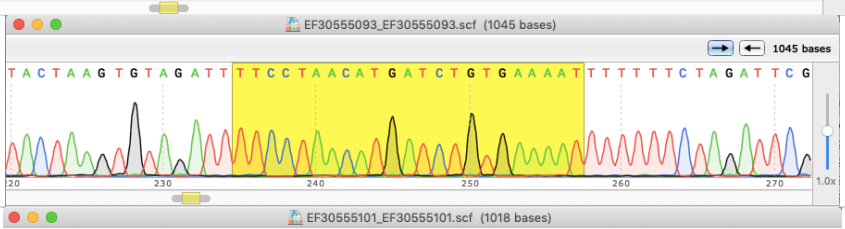
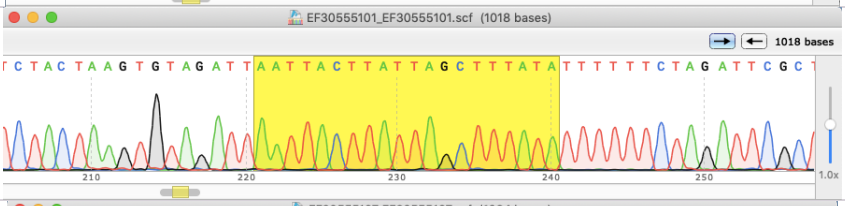
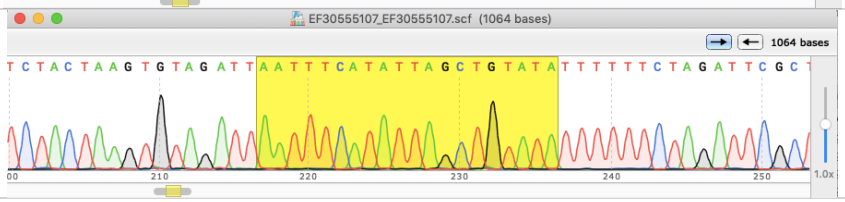
Figure 8.18. Gel Image from pAAV-Cpf1 (maxiprepped) restriction digestion. 1% (w/v) agarose gel with 0.5X SYBR Safe in 1X TAE Buffer. Plasmid integrity was confirmed. Fragments show the following expected fragments sizes: EcoRI: 1. 7,585 bp. BsaI: 1. 7,585 bp. MscI: 1. 2,820 bp, 2. 2,457 bp, 3. 2,002 bp, 4. 306 bp. ScaI: 1. 3,806 bp, 2. 3,113 bp, 3. 666 bp.

Guide RNAs were cloned into p-empty-Cpf1 and correct guide insertion was confirmed by sequencing with “pCpf1 Guides” primer (5´- TTG CAT ATA CGA TAC AAG GCT G -3´) as shown in Table 8.1.

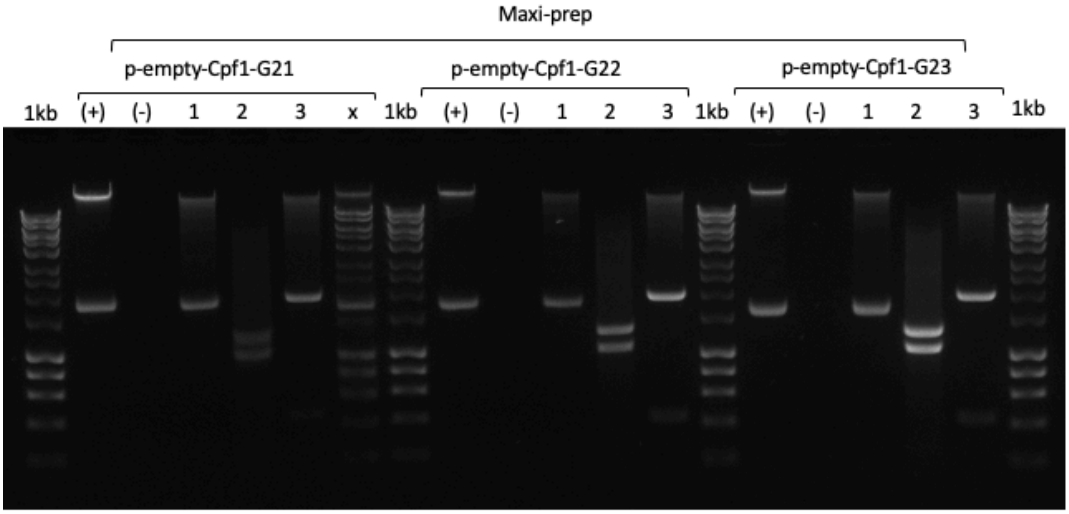
Table 8.1. Representative Sanger sequencing results of p-empty-Cpf1 mini-preps with gRNAs cloned in and highlighted in yellow. Only one of the four results per gRNA shown.

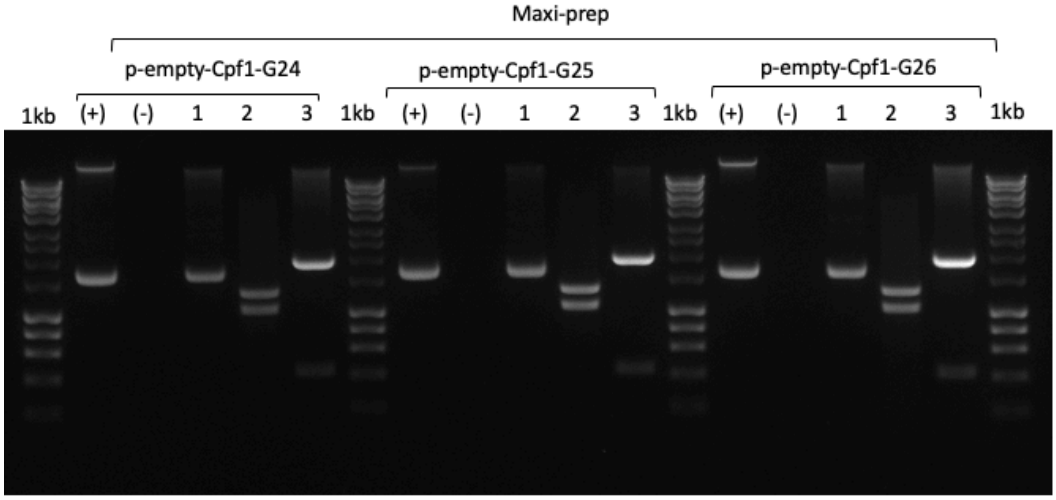
Guide	Sequence
H-I-18-G21.1	
H-I-18-G22.2	
H-I-18-G23.1	
H-I-18-G24.1	
H-I-18-G25.3	
H-I-55-G26.1	
H-I-55-G27.1	

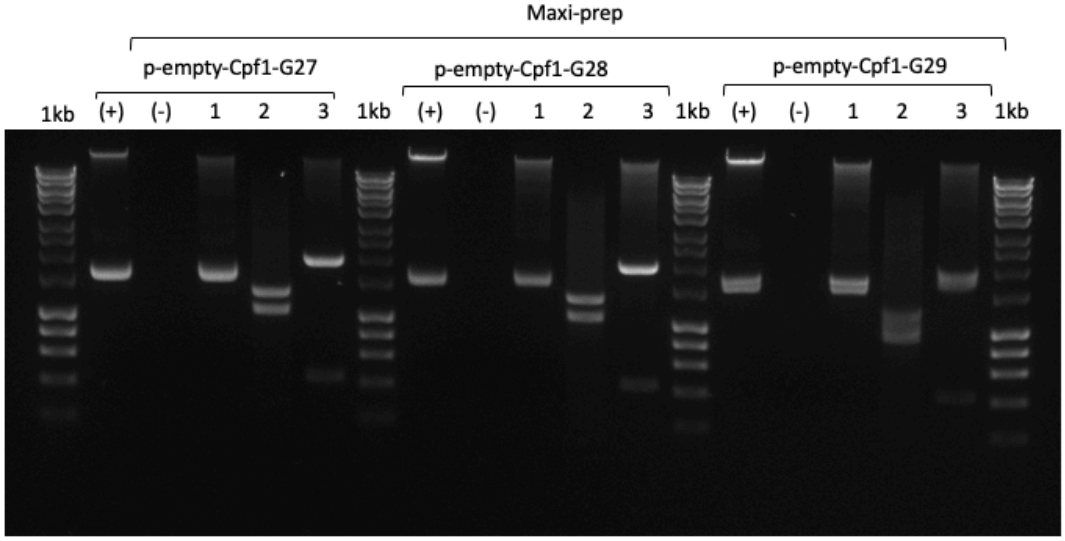
H-I-55-G28.1	
H-I-55-G29.1	
H-I-55-G30.2	
M-I-18-G31.1	
M-I-18-G32.2	
M-I-18-G33.2	
M-I-18-G34.1	
M-I-55-G35.1	

M-I-55-G36.3	
M-I-55-G37.1	
M-I-55-G38.3	
M-I-55-G39.1	
M-I-55-G40.1	
H-I-18-G43.2	
M-I-18-G44.2	

Then, each construct was maxi-prepped and plasmid integrity was confirmed by sequencing and restriction digestions, as seen on Fig. 8.19 and Table 8.2.


Maxi-prep
p-empty-Cpf1-G21
p-empty-Cpf1-G22
p-empty-Cpf1-G23
1kb (+) (-) 1 2 3 x 1kb (+) (-) 1 2 3 1kb (+) (-) 1 2 3 1kb


Maxi-prep
p-empty-Cpf1-G24
p-empty-Cpf1-G25
p-empty-Cpf1-G26
1kb (+) (-) 1 2 3 1kb (+) (-) 1 2 3 1kb (+) (-) 1 2 3 1kb


Maxi-prep
p-empty-Cpf1-G27
p-empty-Cpf1-G28
p-empty-Cpf1-G29
1kb (+) (-) 1 2 3 1kb (+) (-) 1 2 3 1kb (+) (-) 1 2 3 1kb

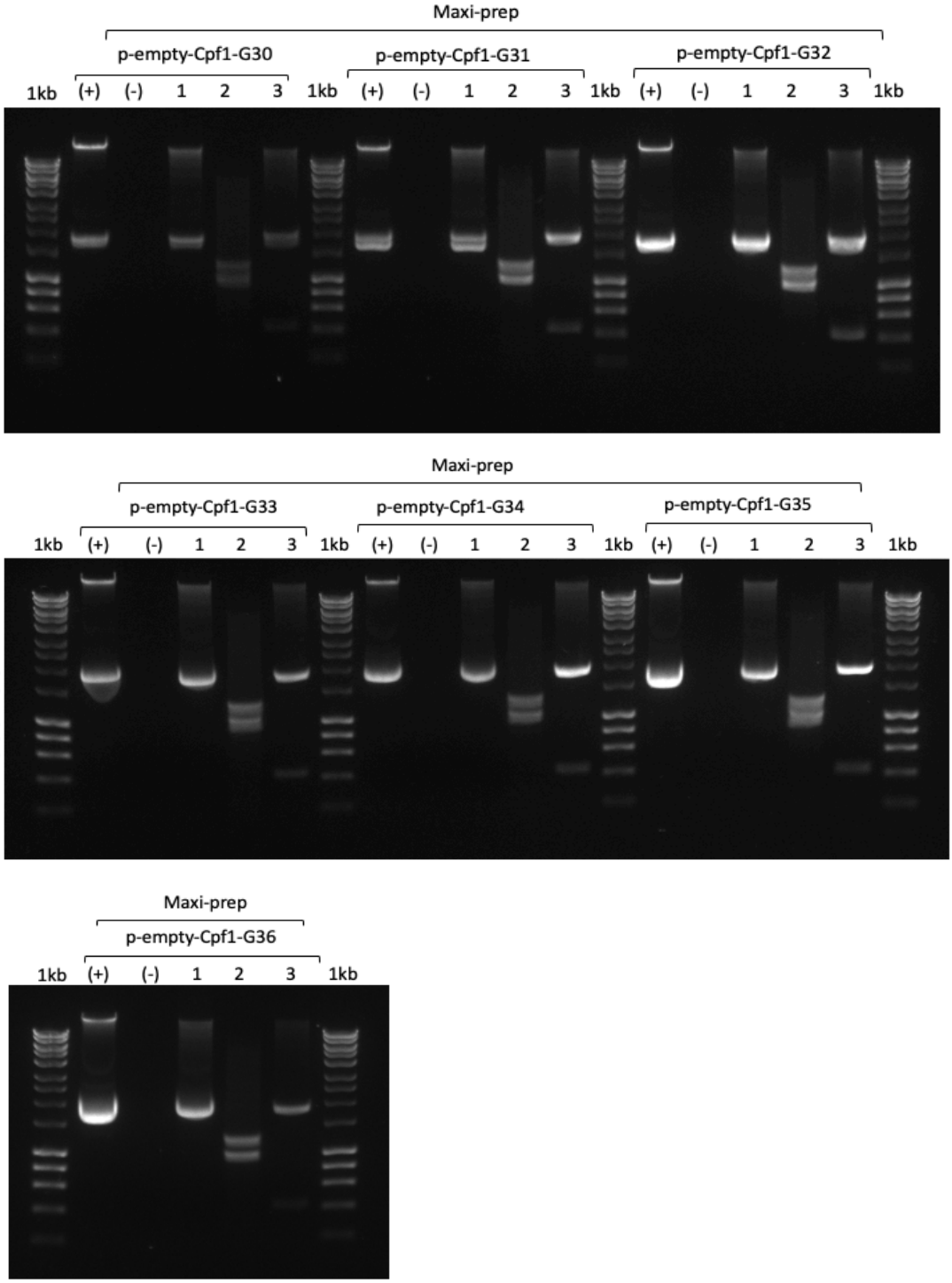


Figure 8.19. Gel Image from restrictions digestion of p-empty-Cpf1 with guides cloned and (maxiprepped). 1% (w/v) agarose gel with 0.5X SYBR Safe in 1X TAE Buffer. Plasmids digested with 1=BamHI, 2=BanI and 3=EcoRI. Plasmid integrity was confirmed. Fragments show the following expected fragments sizes: BamHI: 1. 2,386 bp. BanI: 1. 1,327 bp, 2. 1,059 and for EcoRI: 1. 1,969 bp, 2. 426 bp.

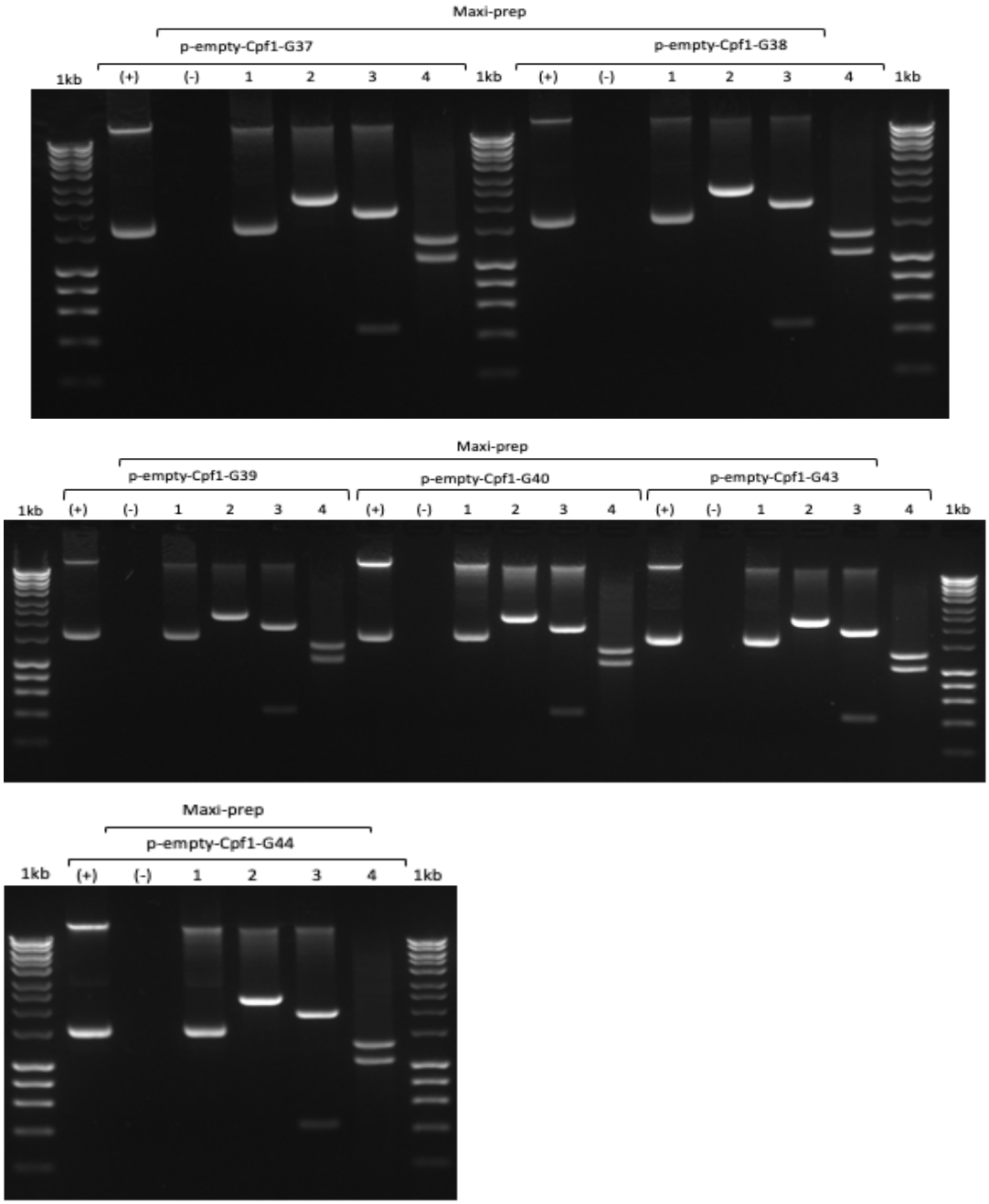
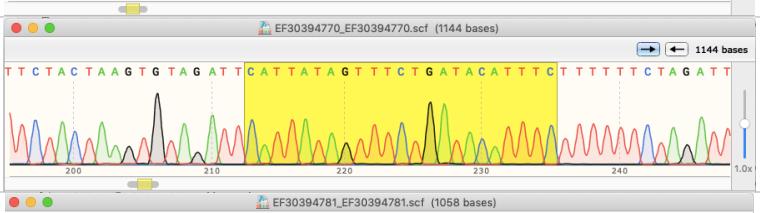
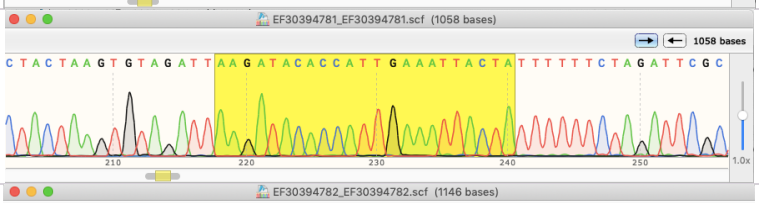
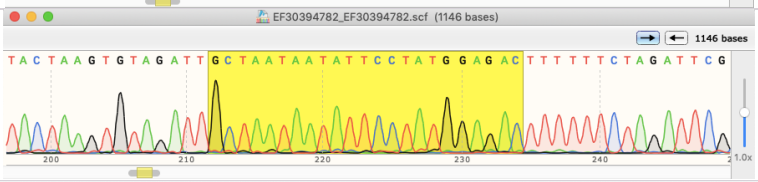
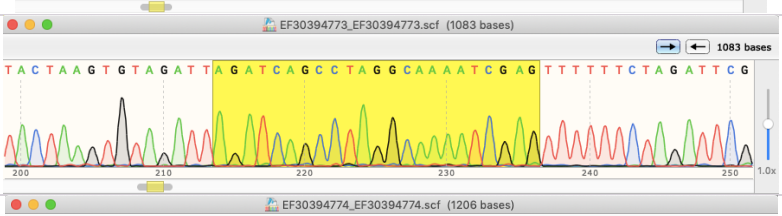
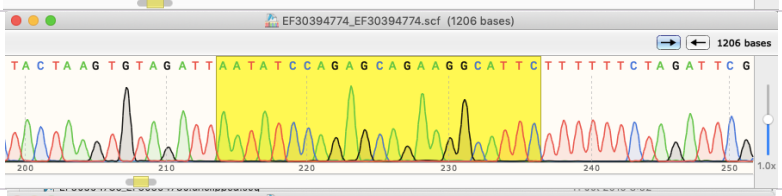
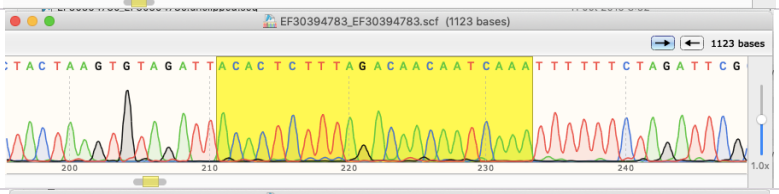
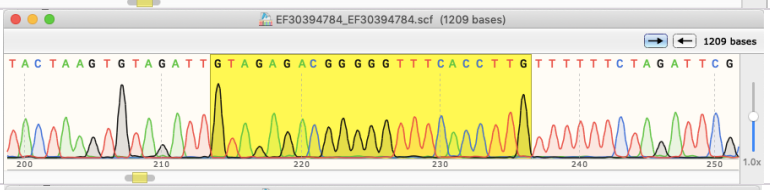
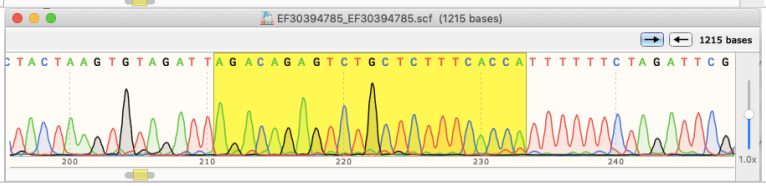
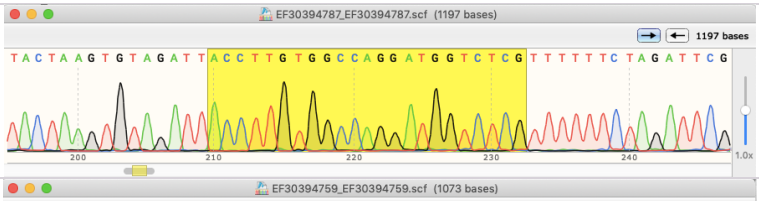
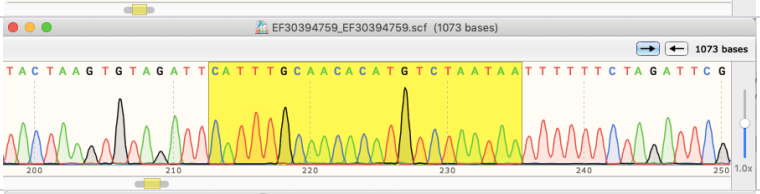
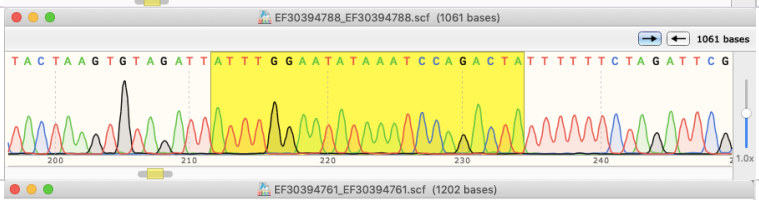
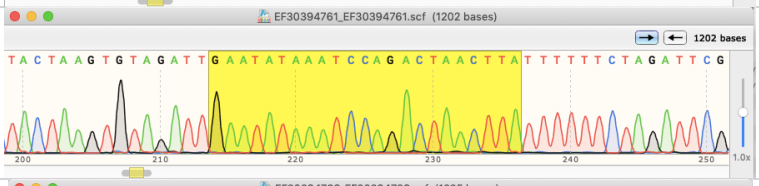
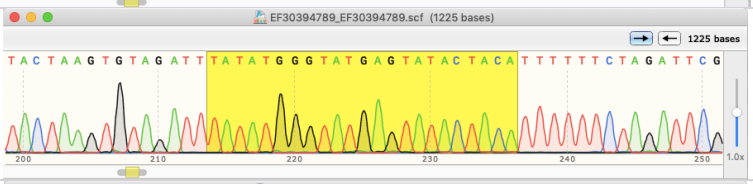
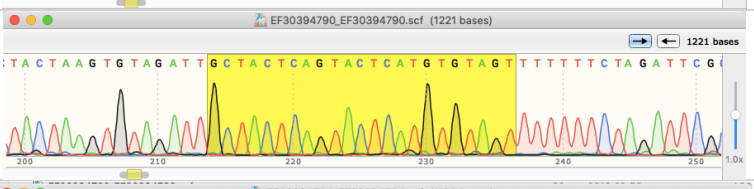
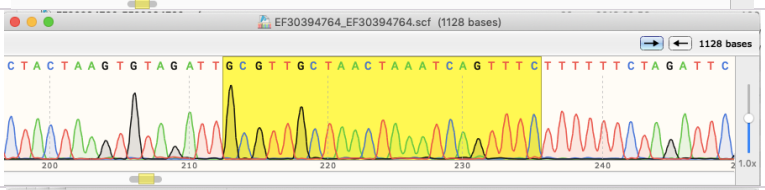
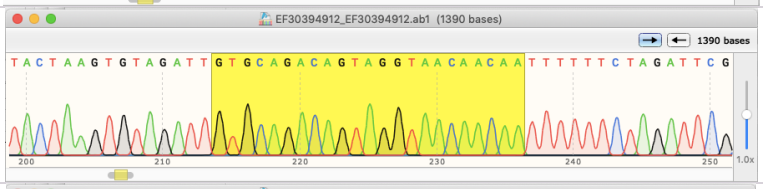
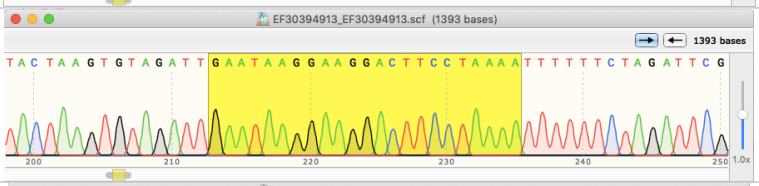
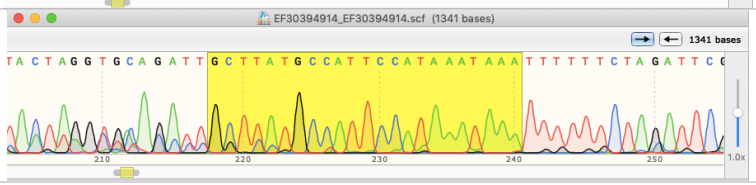
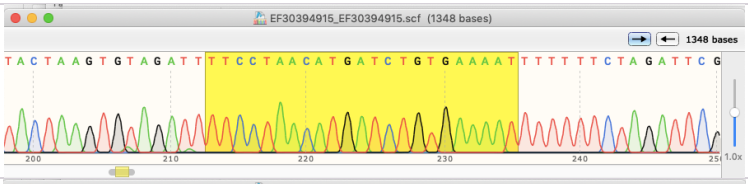
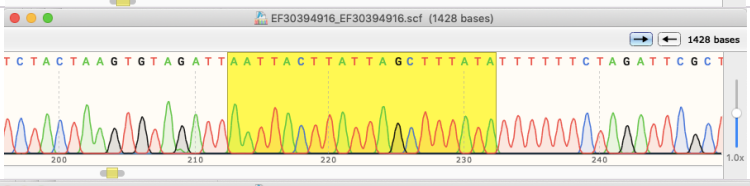
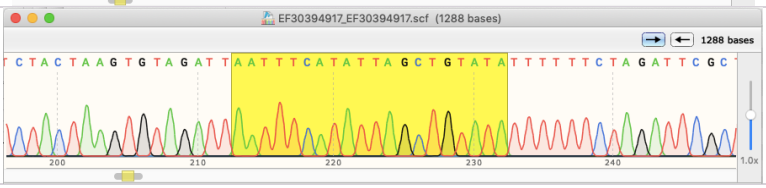


Figure 8.20. Gel Image from restrictions digestion of p-empty-Cpf1 with guides cloned and (maxiprepped). 1% (w/v) agarose gel with 0.5X SYBR Safe in 1X TAE. Plasmids digested with 1=BamHI, 2=NdeI and 3=EcoRI. 4=BanI Plasmid integrity was confirmed. Fragments show the following expected fragments sizes: BamHI: 1. 2,386 bp. (in theory BamHI site was removed when the guides were cloned in, hence this band should match the positive undigested control) NdeI: 1. 2,386 bp. EcoRI: 1. 1,969 bp, 2. 426 bp. and BanI: 1. 1,327 bp, 2. 1,059.

Table 8.2. Representative Sanger sequencing results of maxi-preps from p-empty-Cpf1 with gRNAs cloned (gRNA highlighted in yellow). Only one of the four results per gRNA shown.

Guide	Sequence
H-I-18-G21	
H-I-18-G22	
H-I-18-G23	
H-I-18-G24	
H-I-18-G25	
H-I-55-G26	
H-I-55-G27	
H-I-55-G28	
H-I-55-G29	

H-I-55-G30	
M-I-18-G31	
M-I-18-G32	
M-I-18-G33	
M-I-18-G34	
M-I-55-G35	
M-I-55-G36	
M-I-55-G37	
M-I-55-G38	
M-I-55-G39	

M-I-55-G40	
H-I-18-G43	
M-I-18-G44	

Co-transfection of p-empty-Cpf1 with each gRNA and pAAV-Cpf1 expressing Cpf1 for testing guide efficiency, would have likely resulted in low efficiencies of co-delivery, so the two-step cloning strategy was explored.

Nevertheless, it was found that the NheI site on pAAV-Cpf1 needed for the second step of the cloning strategy was inexistent and made it impossible to clone the gRNAs from p-empty-Cpf1 into pAAV-Cpf1. It was then decided to clone all gRNAs into pY095 to perform the screening *in-vitro*.

8.2.2. Cpf1 gRNA in-vitro screening & TIDE Analysis.

To find the optimal DNA amount for an efficient transfection, a dose response was performed with pY095-GFP on HEK293T and Neuro2A cells. Once the optimal dose was

confirmed by fluorescence microscopy and FACS analysis, it was proceeded to gRNA cloning on an AAV plasmid.

Plasmid integrity of pY095 was confirmed by restriction digests (Data not shown). Then an experiment was set up testing different DNA dose responses using Vifect transfection reagent (4:1 to DNA), with 4, 6 and 8 μg of DNA. Microscopy of transfected HEK293T cells can be observed in Fig. 8.21. FACS Analysis results are presented on Figure 8.22.

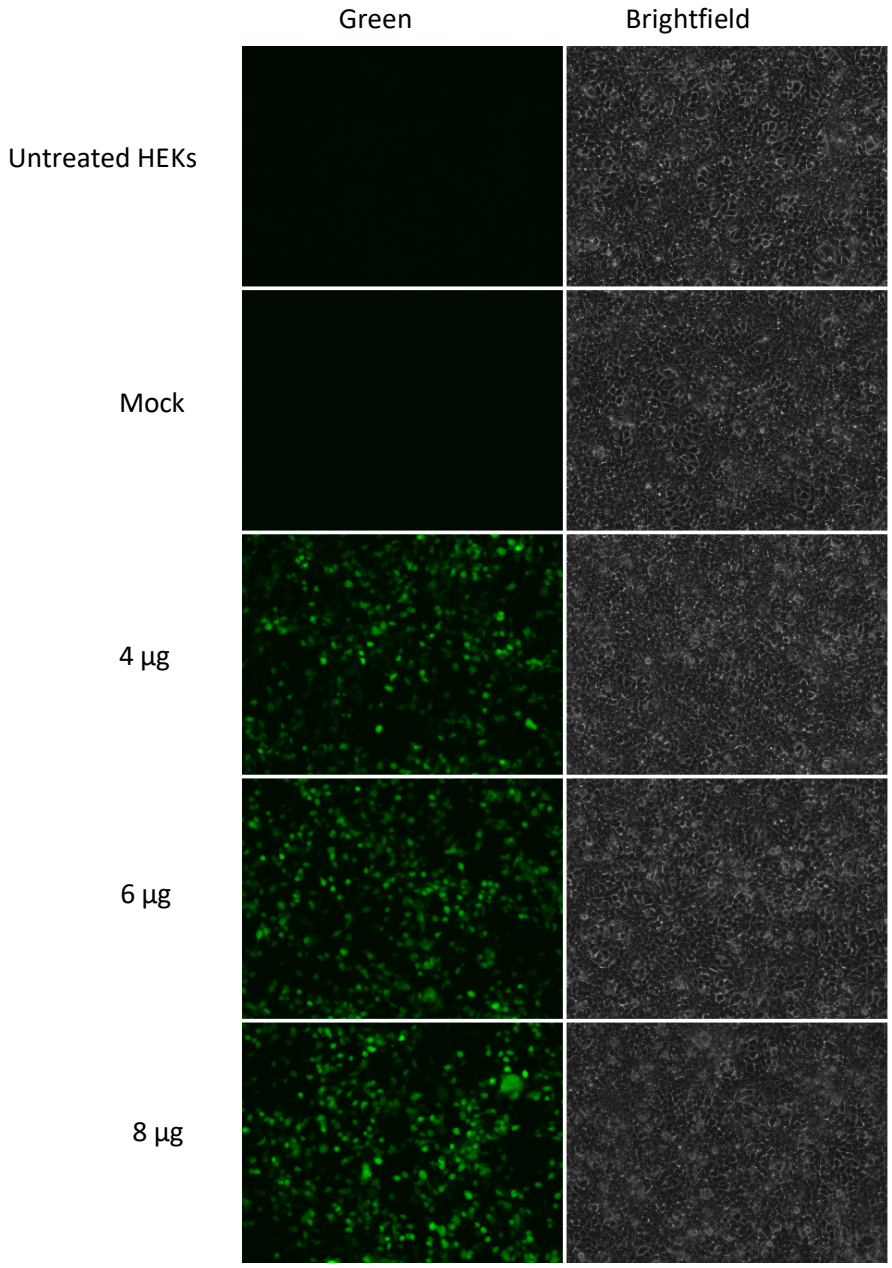


Figure 8.21. Microscopy of HEK293T cells transfected with pY095. Images taken 48 hrs. after transfection with Viafect 4:1 to DNA.

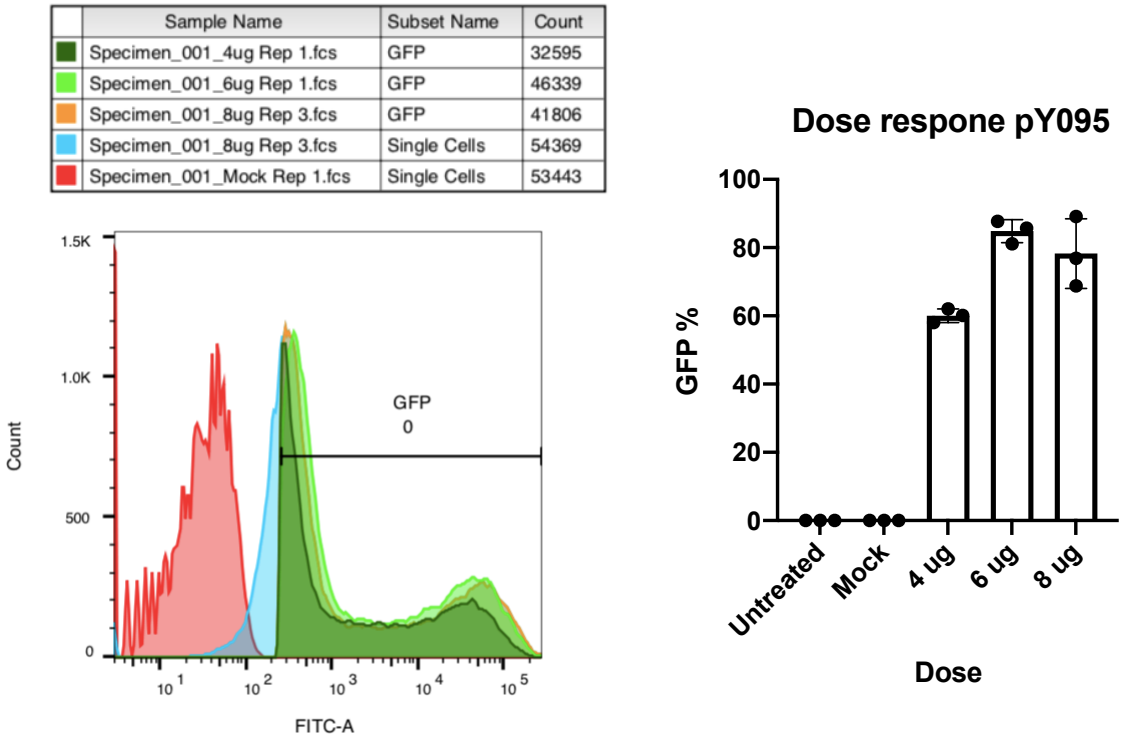


Figure 8.22. FACS Analysis of pY095 transfected on HEKs (dose response from 4-8 µg). From left to right: Histogram showing cell counts for mock, positive control and all different doses. The bar graph shows the percentage of positive cells according to different doses.

It was concluded that a transfection efficiency between 60-80% should be enough to detect gRNA cutting. In order to save reagents, all gRNA screening was performed at a 4 µg dose with a 1:4 DNA to Viafect ratio.

After transfection, DNA harvesting, PCR with appropriate primers (as demonstrated on Fig. 8.23), PCR product clean-up and Sanger sequencing, samples were analysed by TIDE, results of gRNA cutting efficiency can be seen on Fig. 8.24.

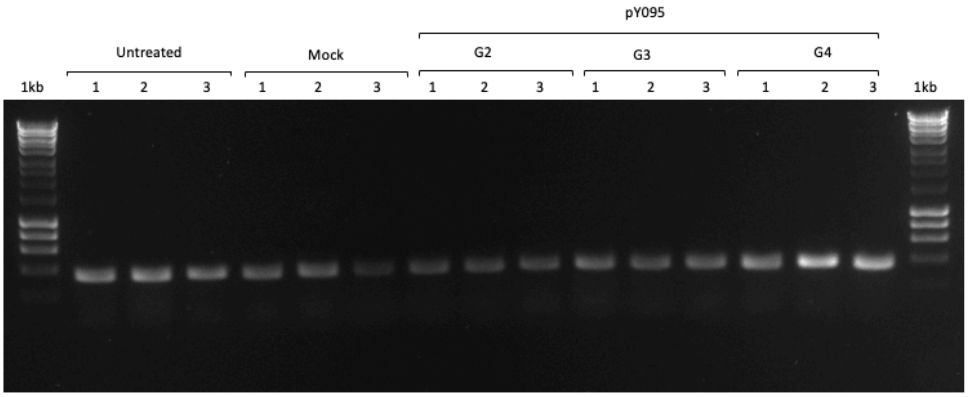


Figure 8.23. Representative image of PCRs from Cpf1 gRNA screening showing PCR products of Guides 2, 3 and 4. Samples were run on a 1% (w/v) agarose gel with 0.5X SYBR Safe in 1X TAE Buffer. Primer Set #17 was used for sequencing the PCR product. Expected band size around 400 bp.

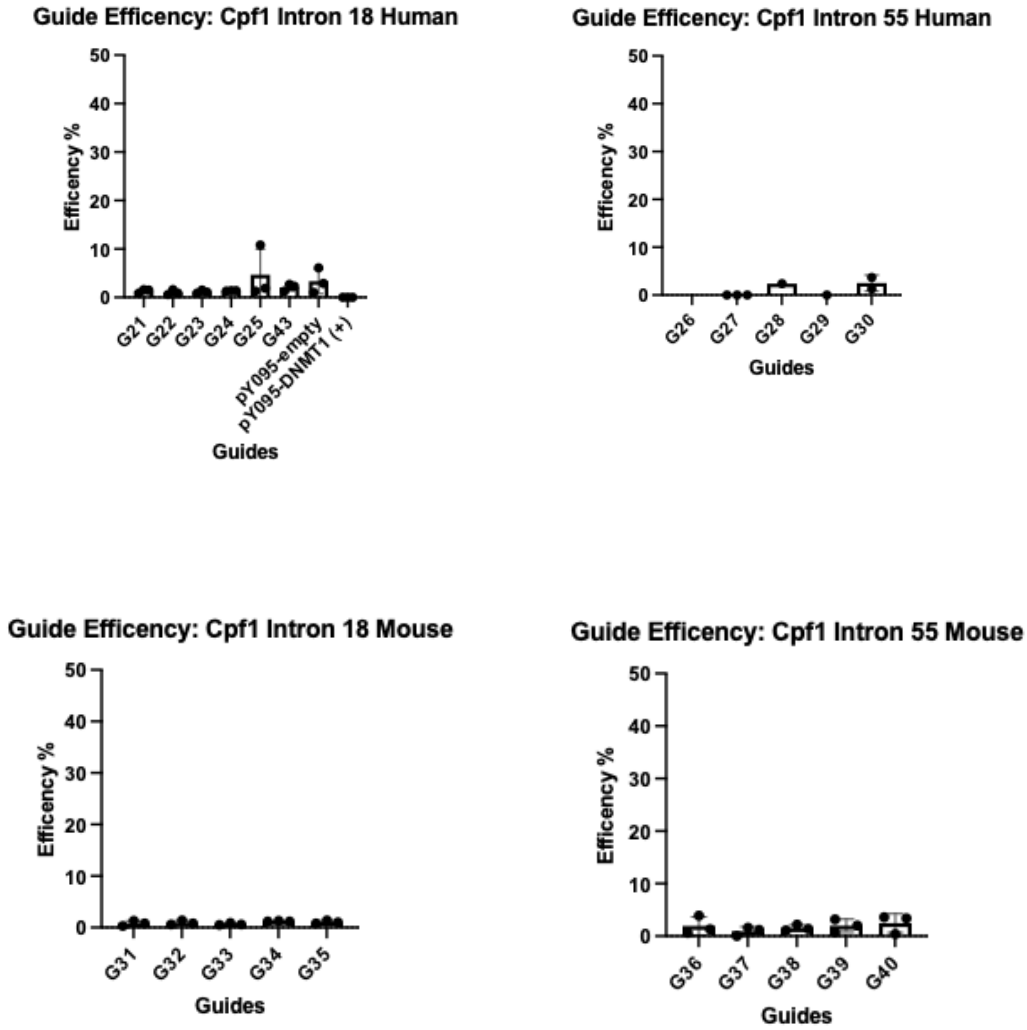


Figure 8.24. Graphical summary of Cpf1 gRNA cutting efficiency assessed by TIDE. Guide RNAs targeting intron 18 and 55 of the *DMD/Dmd* human and mouse gene. Sequencing analysed with TIDE.

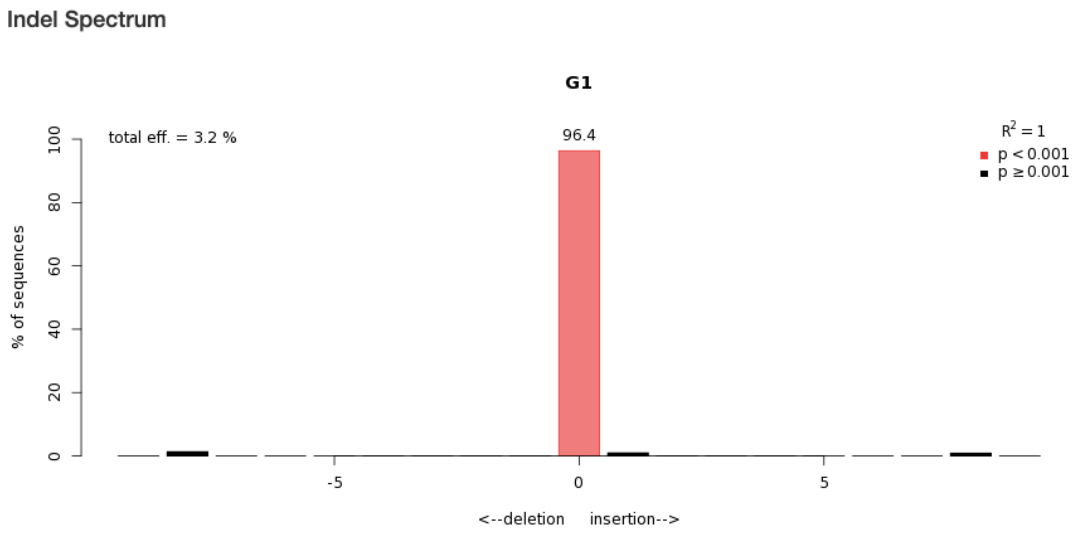
8.3. APPENDIX C: TIDE ANALYSIS RESULTS FROM SACAS9 GRNAS TARGETING INTRONS 18 AND 55 OF THE HUMAN AND MOUSE *DMD/DMD* GENES.

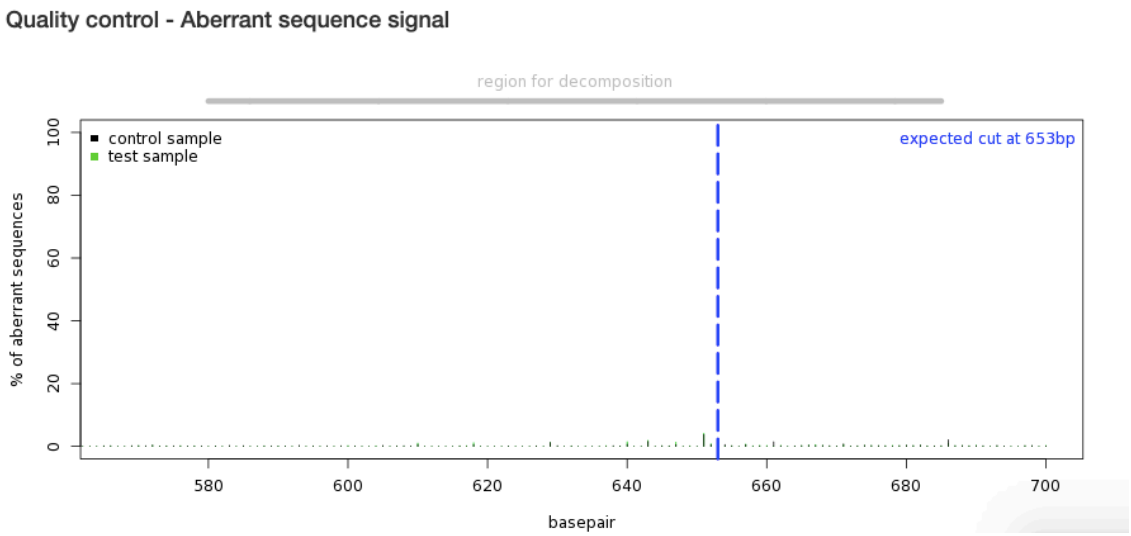
Representative images of outputs from TIDE analysis Software for each *Sa*Cas9 gRNA are presented in this section.

Outputs per gRNA show:

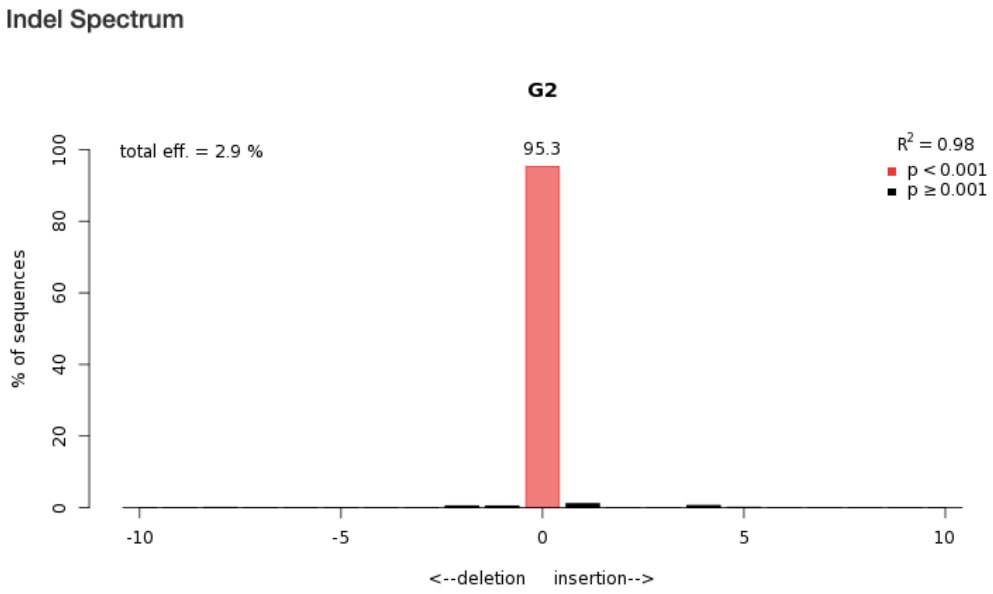
- A bar chart indicating indel spectrum output. X-axis indicates small deletions of up to 10 base pairs on a negative scale (-10 to 0) and insertions on a positive scale (0 to 10). Numbers at the top right corner denote the coefficient of determination (R^2), a statistical measure to evaluate model accuracy with values from 0 to 1. A low R^2 can be caused by poor sequence quality or non-optimal setting. P-values indicate significance cutoff, set up at $p<0.001$. Significant outputs indicated in red, non-significant ($p\geq0.001$) indicated in black.
- A decomposition trace, aberrant sequence signal (green) compared to control trace (black). Dotted blue line indicates cut site.

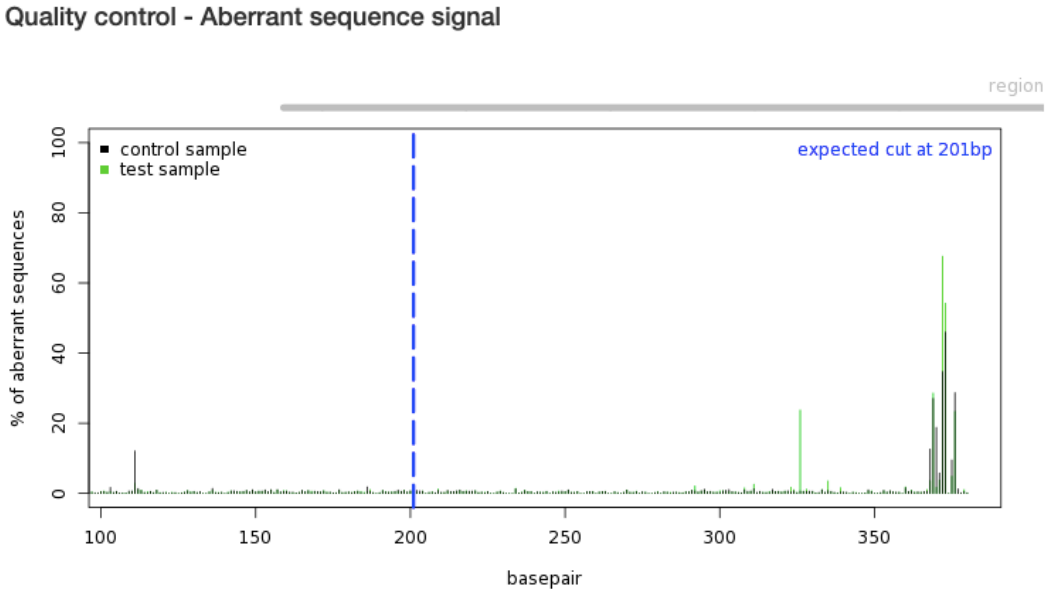
TIDE analysis: H-I18-G1.



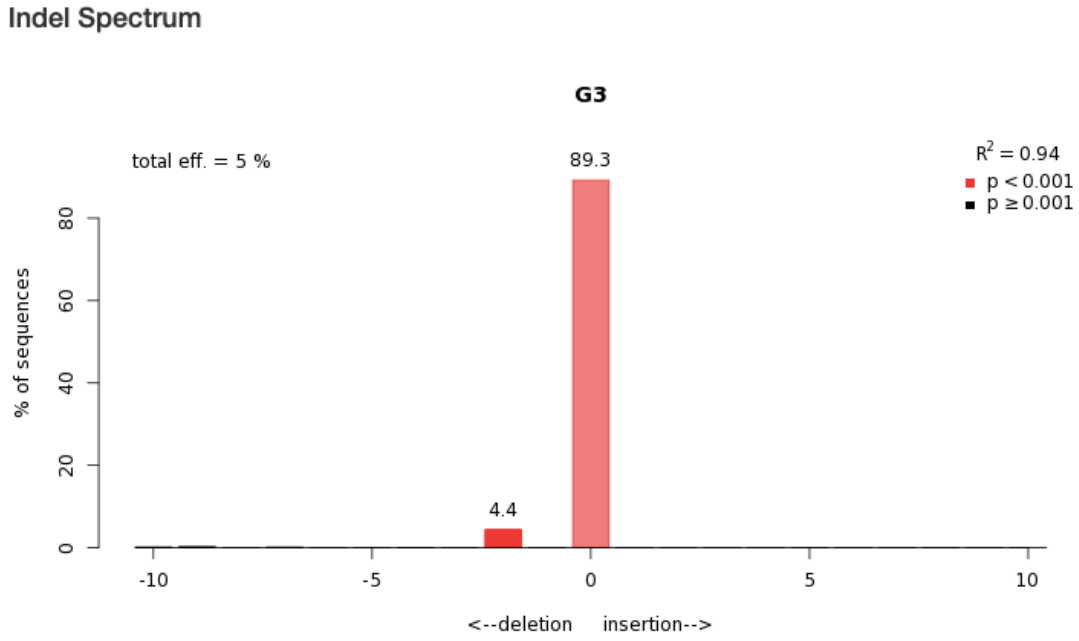


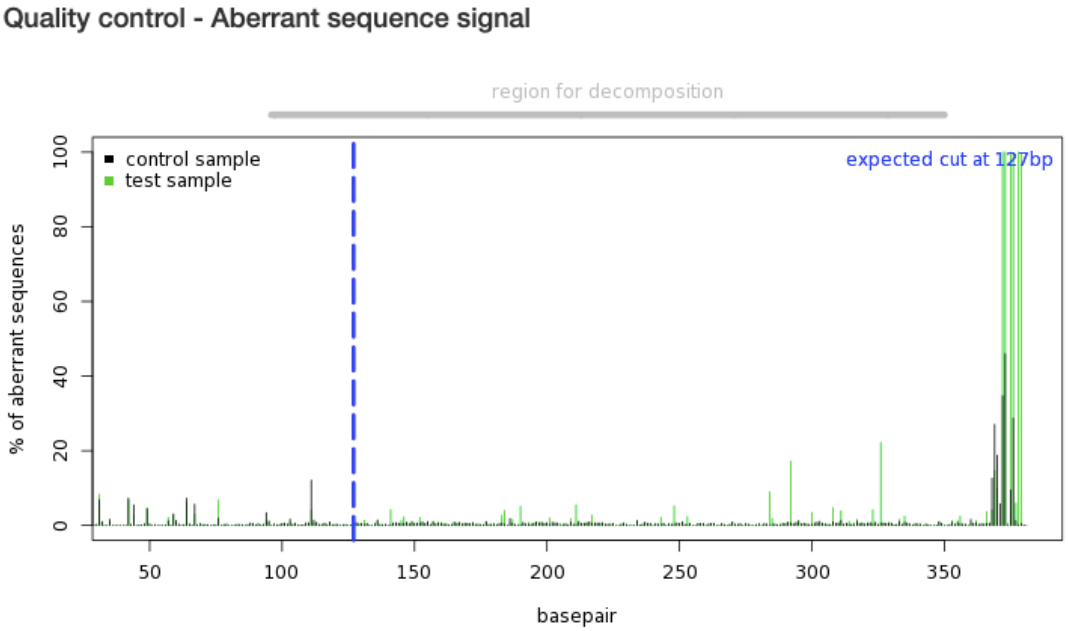
TIDE analysis: H-I18-G2.



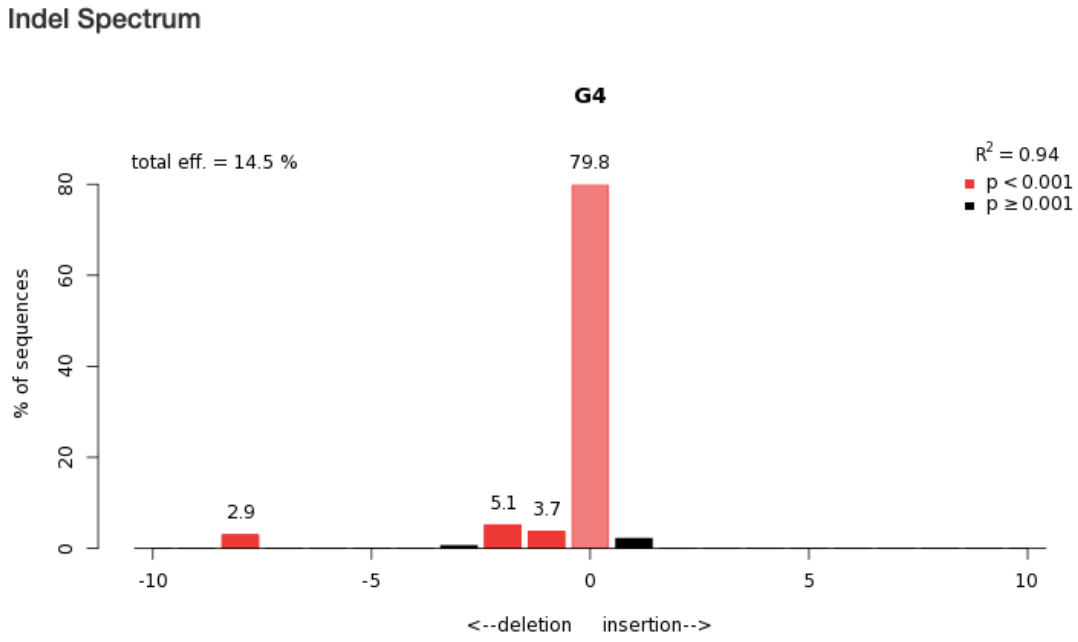


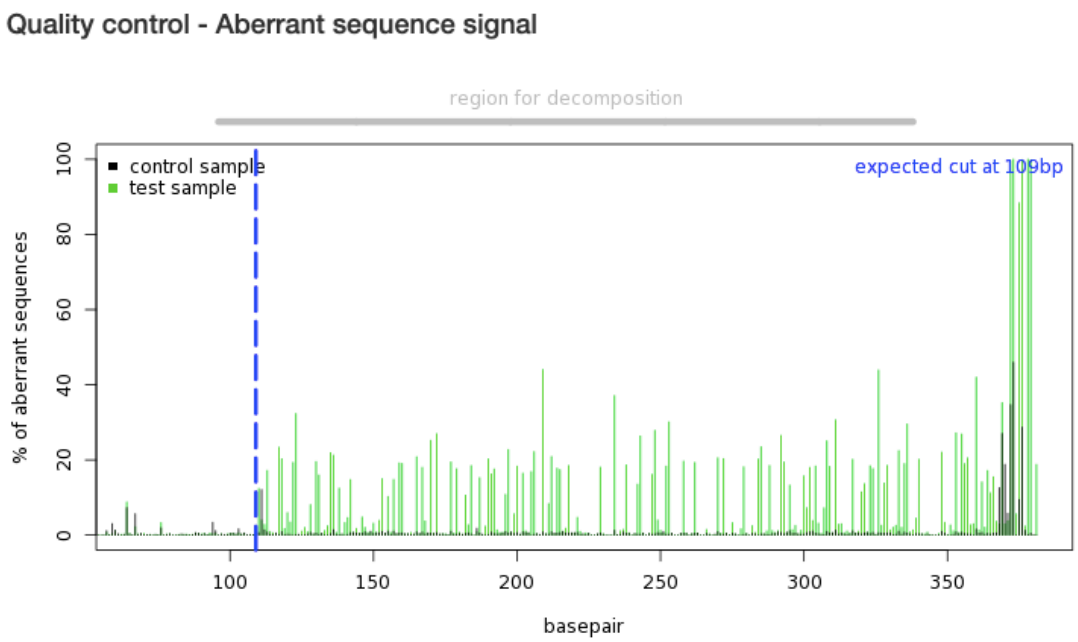
TIDE analysis: H-I18-G3.



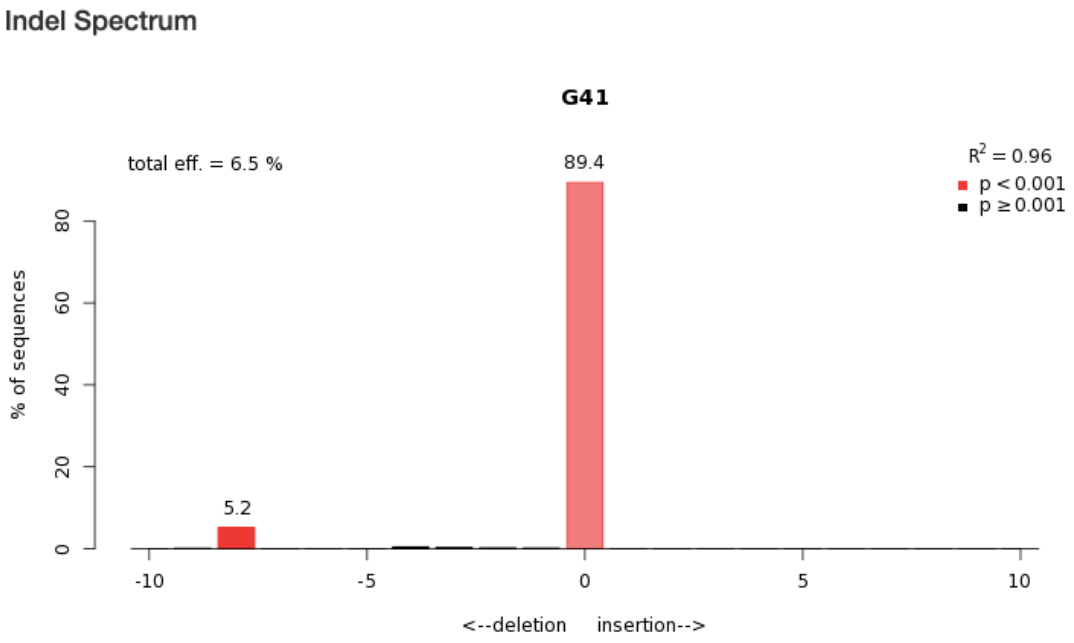


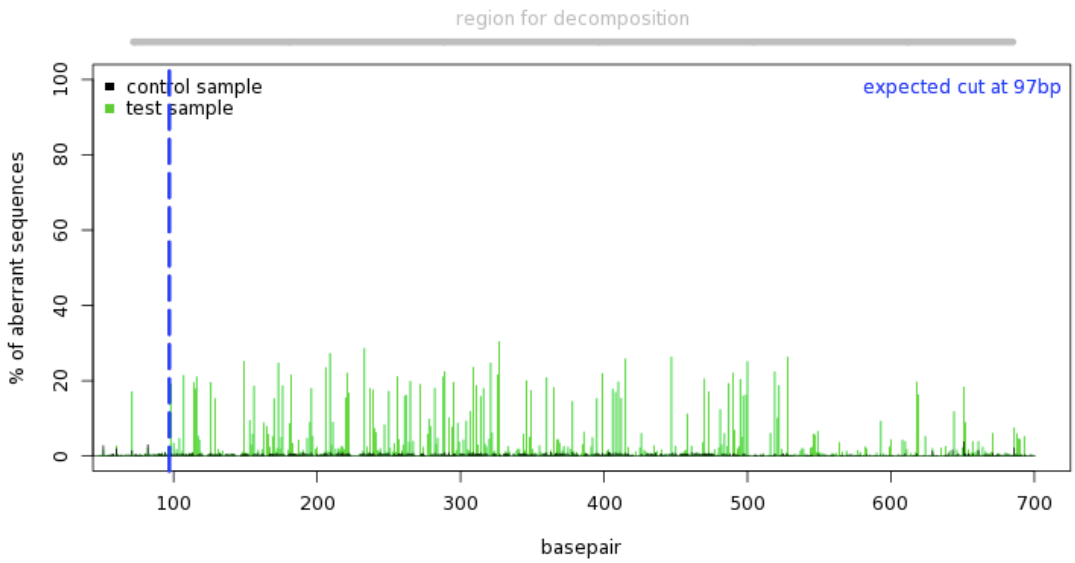
TIDE analysis: H-I18-G4.



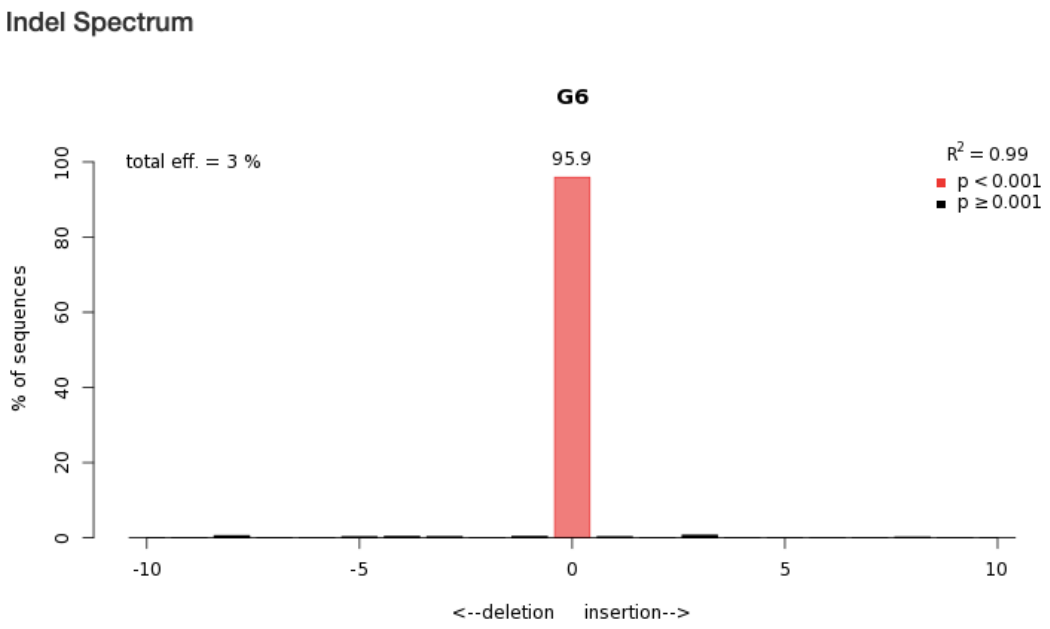


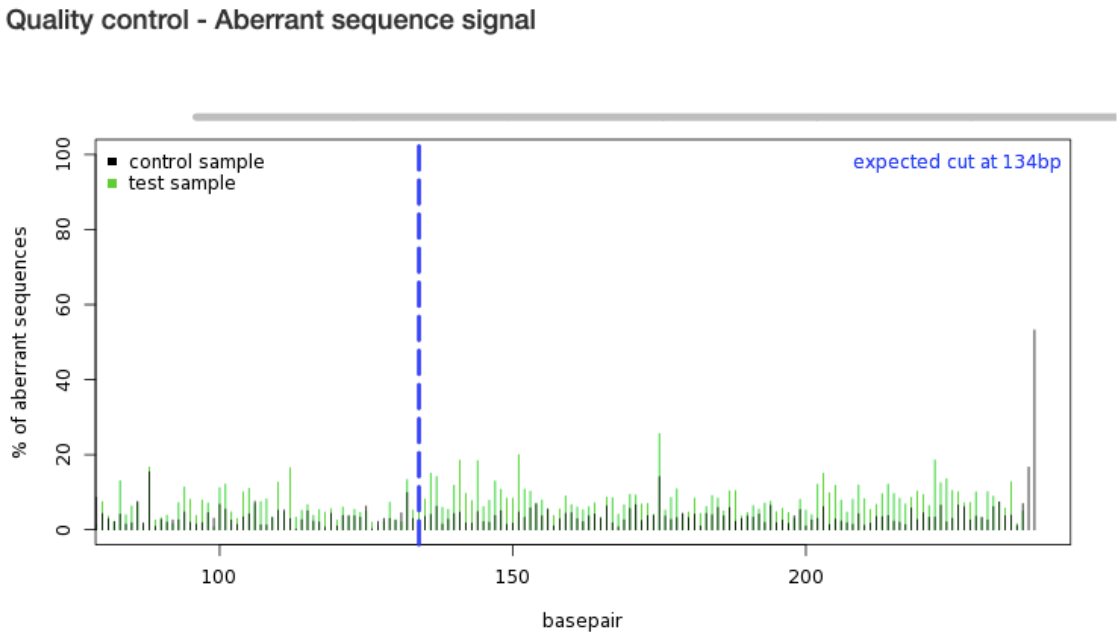
TIDE analysis: H-I18-G41.



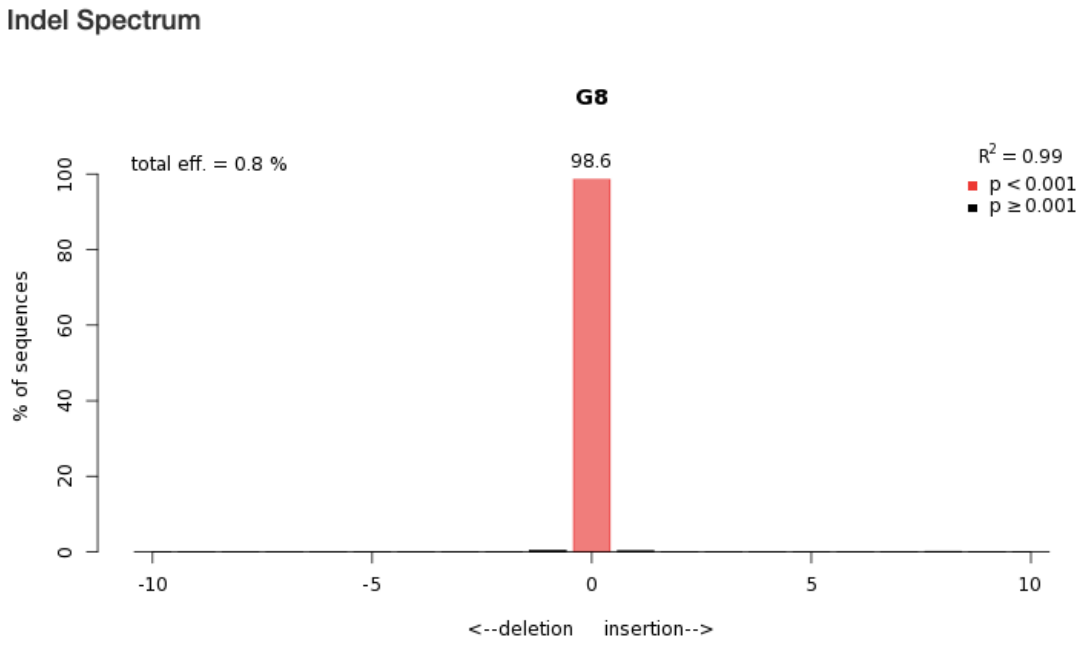


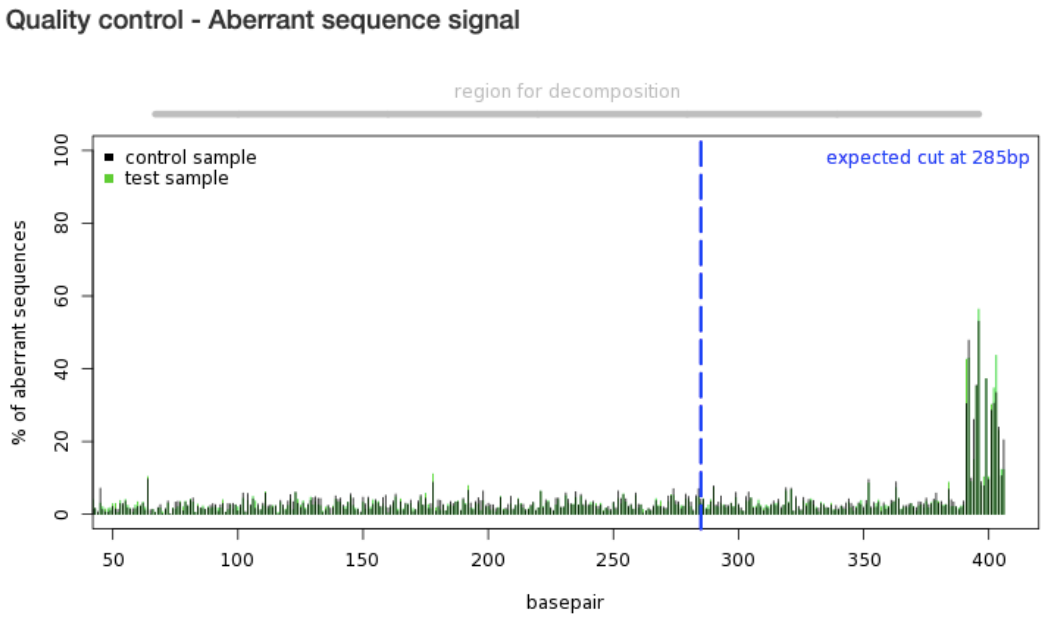
TIDE analysis: H-I55-G6.



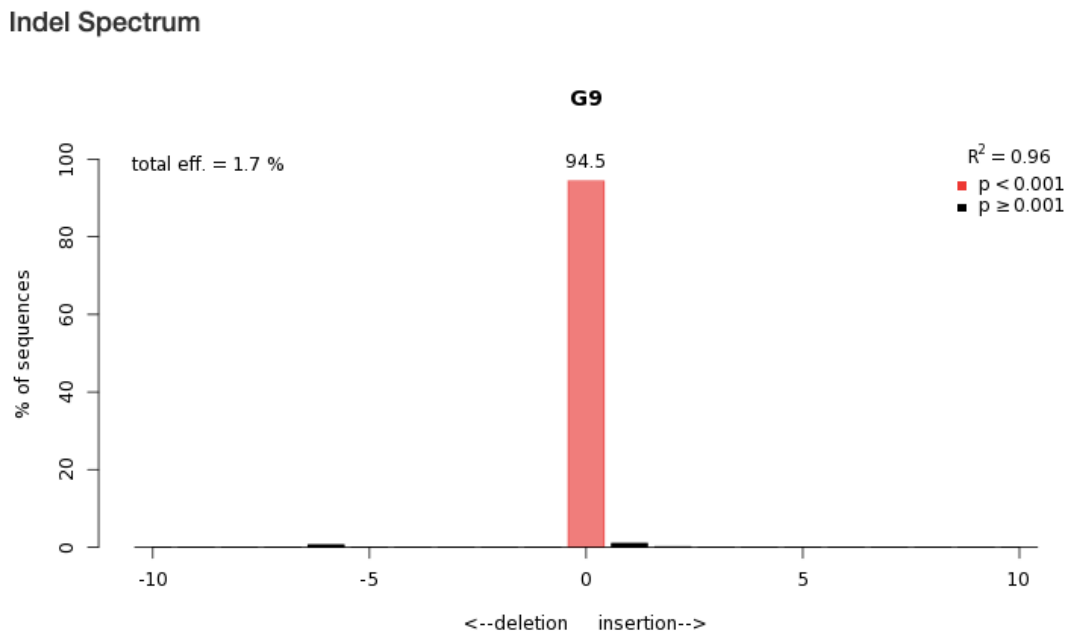


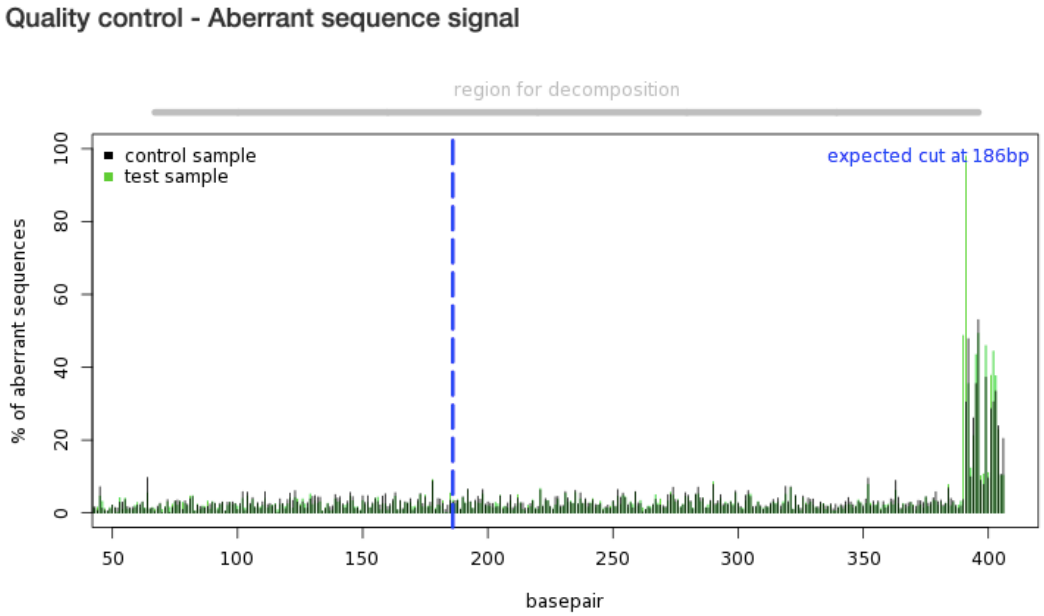
TIDE analysis: H-I55-G8.



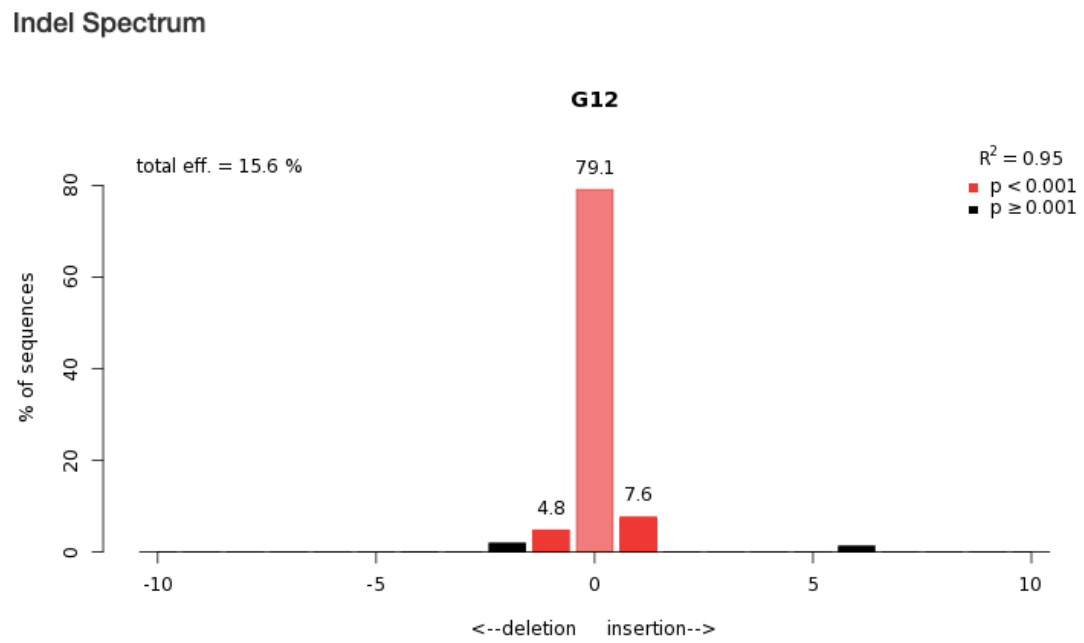


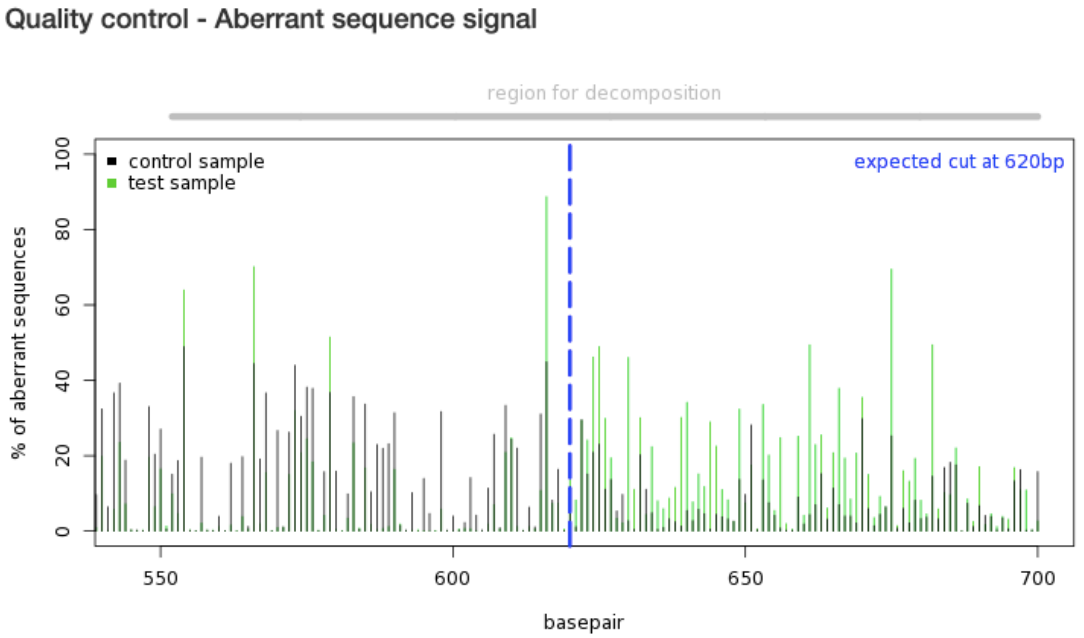
TIDE analysis: H-I55-G9.



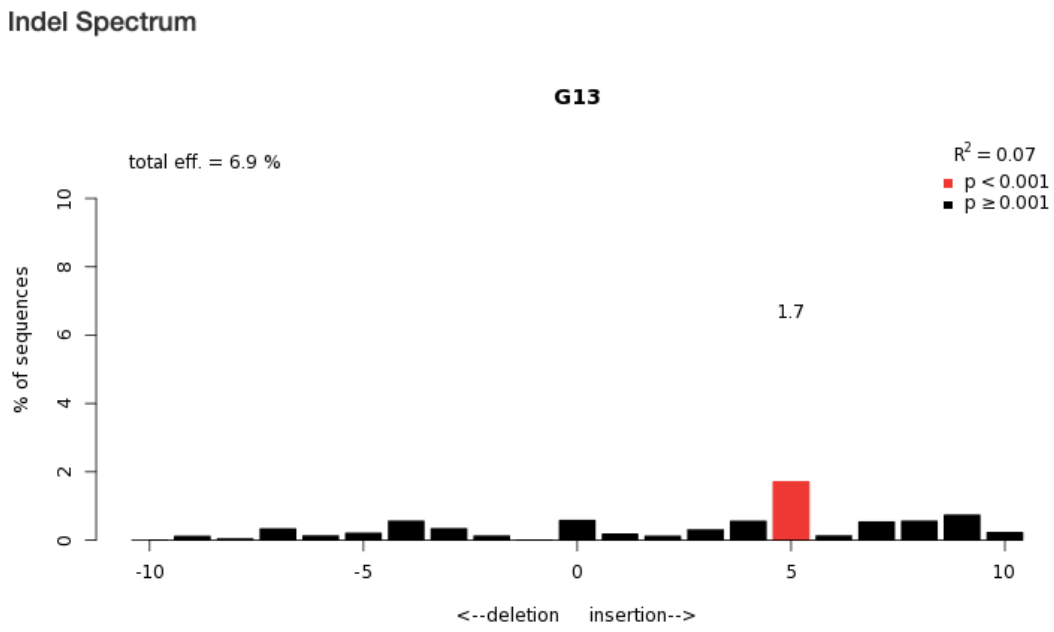


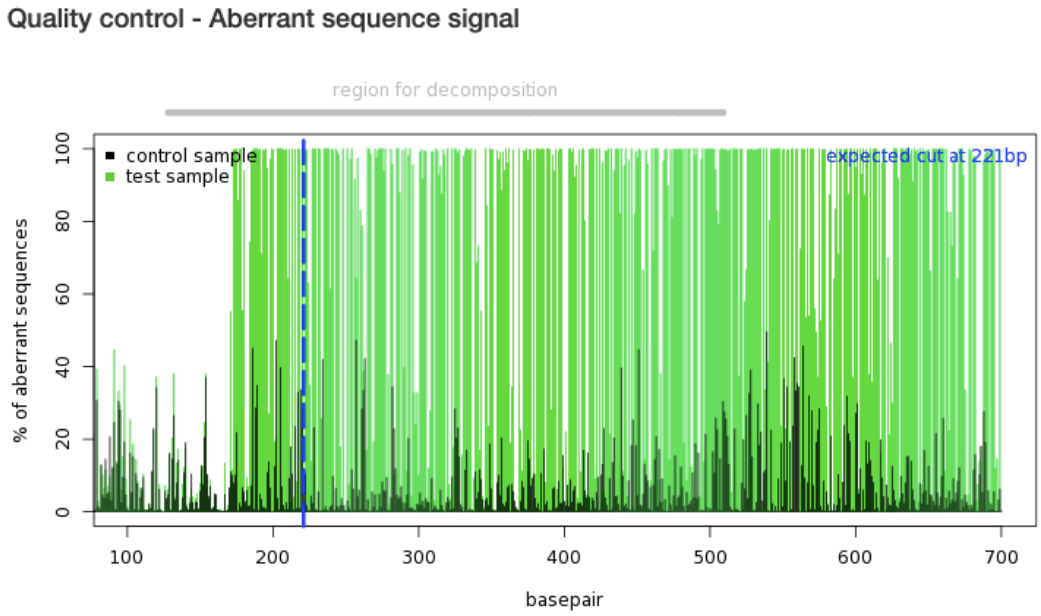
TIDE analysis: M-I18-G12.



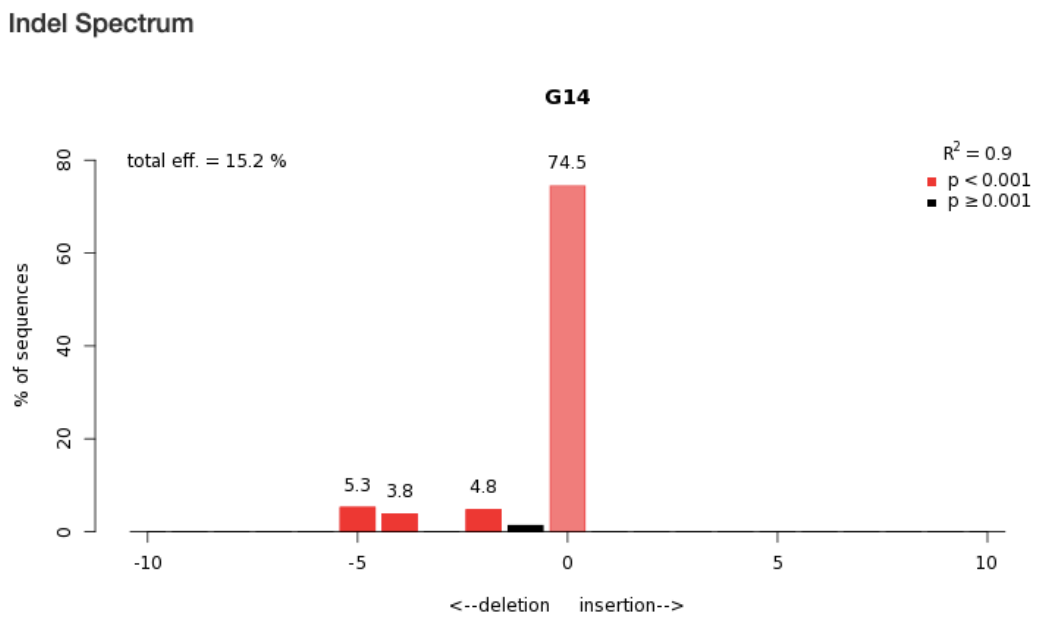


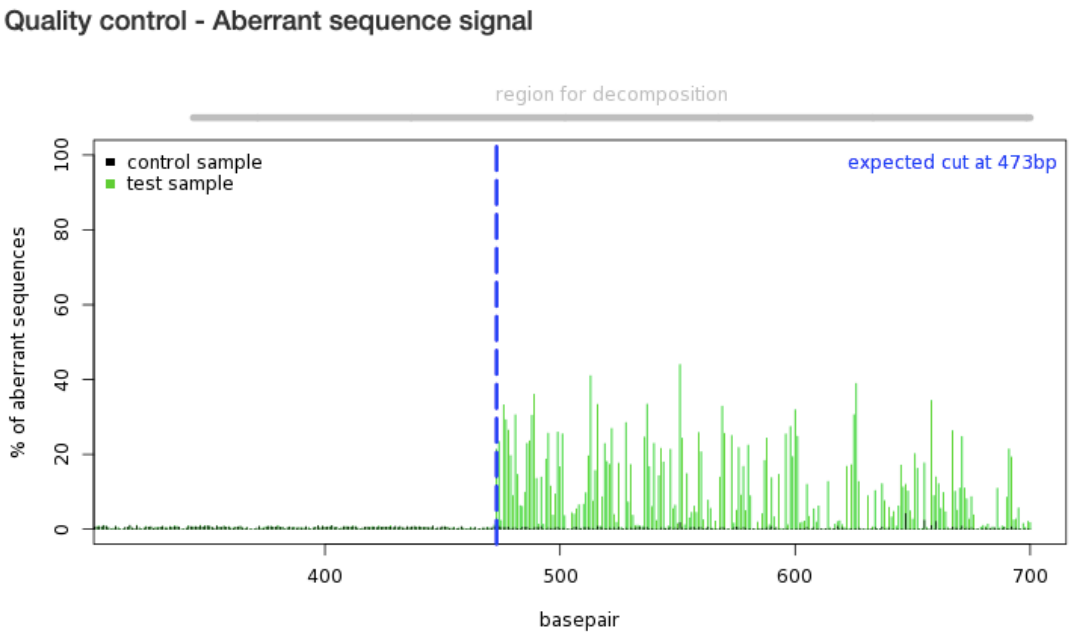
TIDE analysis: M-I18-G13.



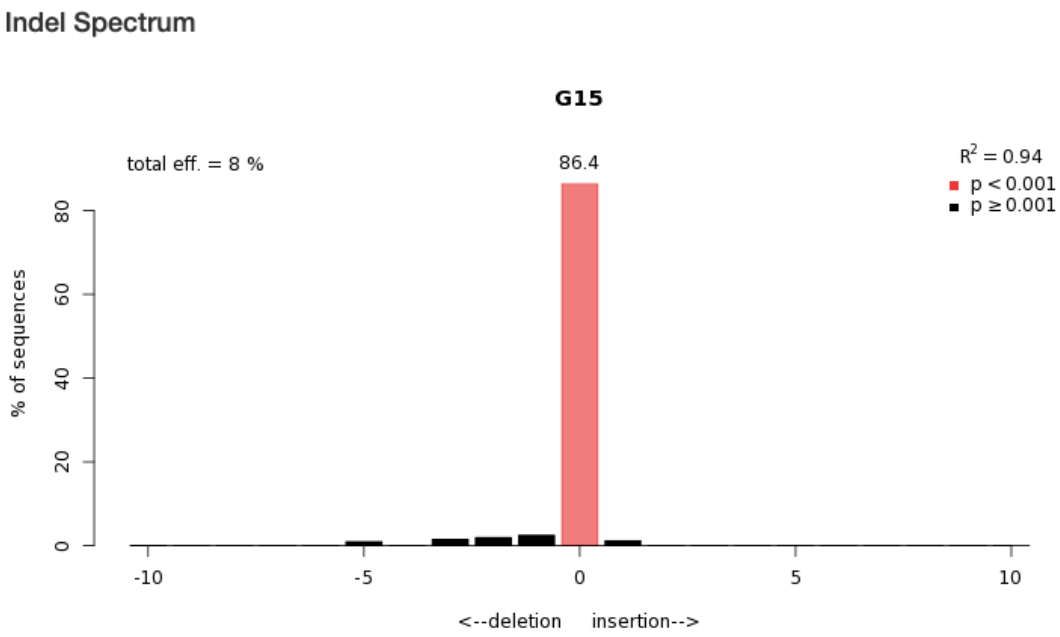


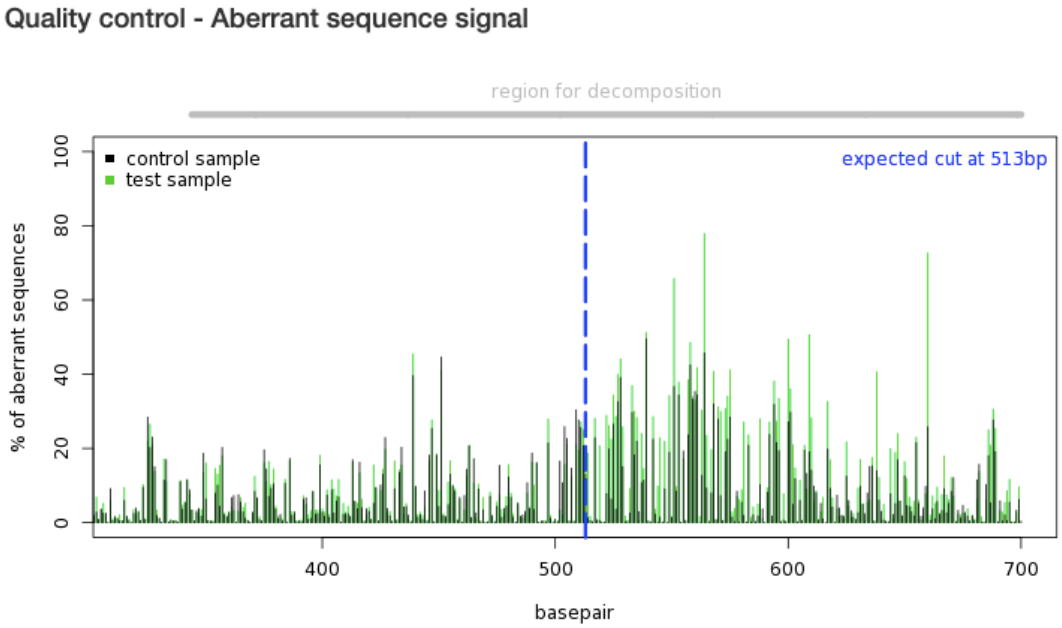
TIDE analysis: M-I18-G14.



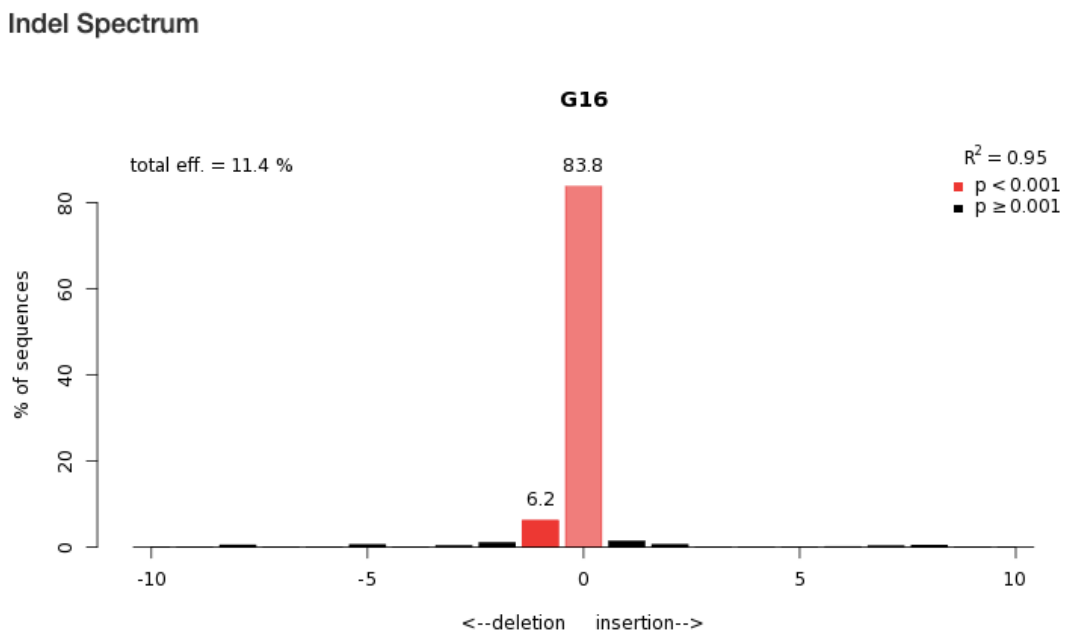


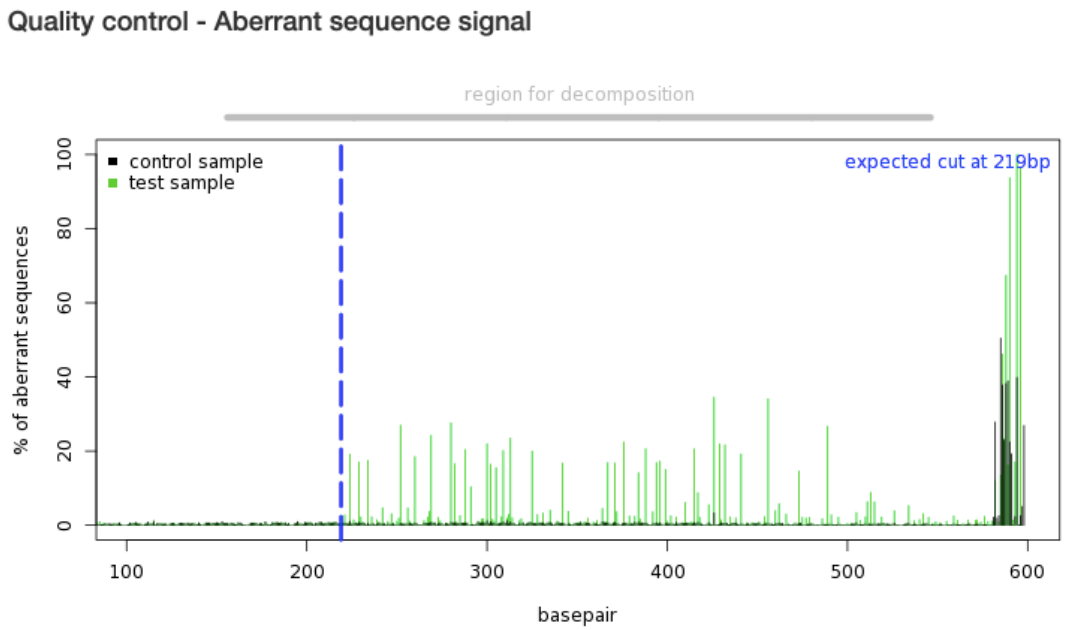
TIDE analysis: M-I18-G15.



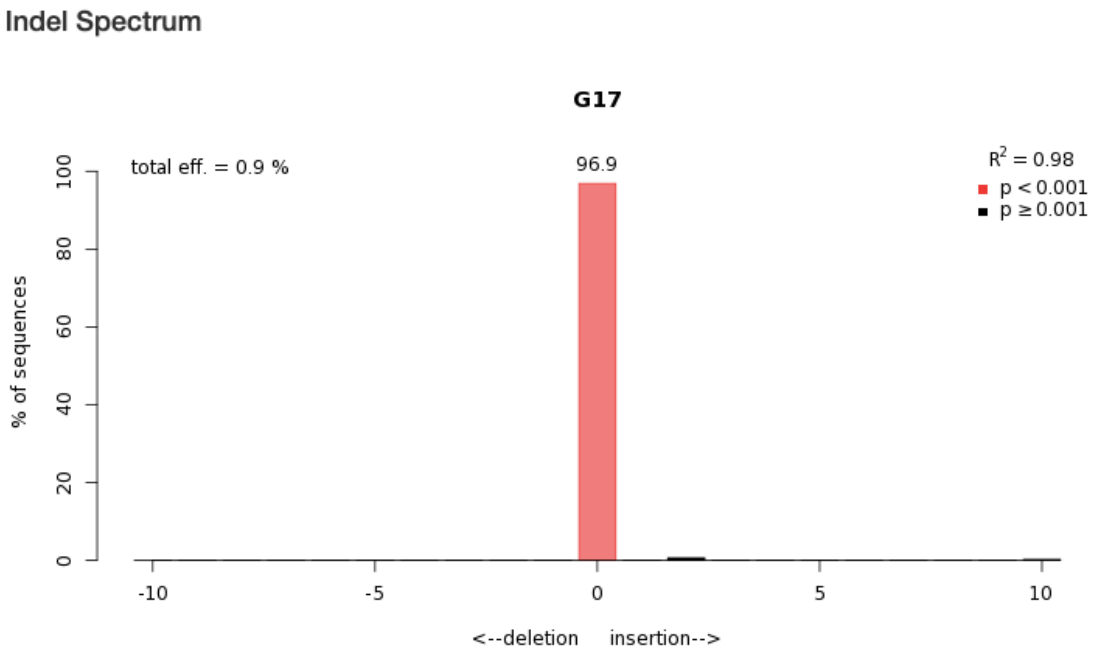


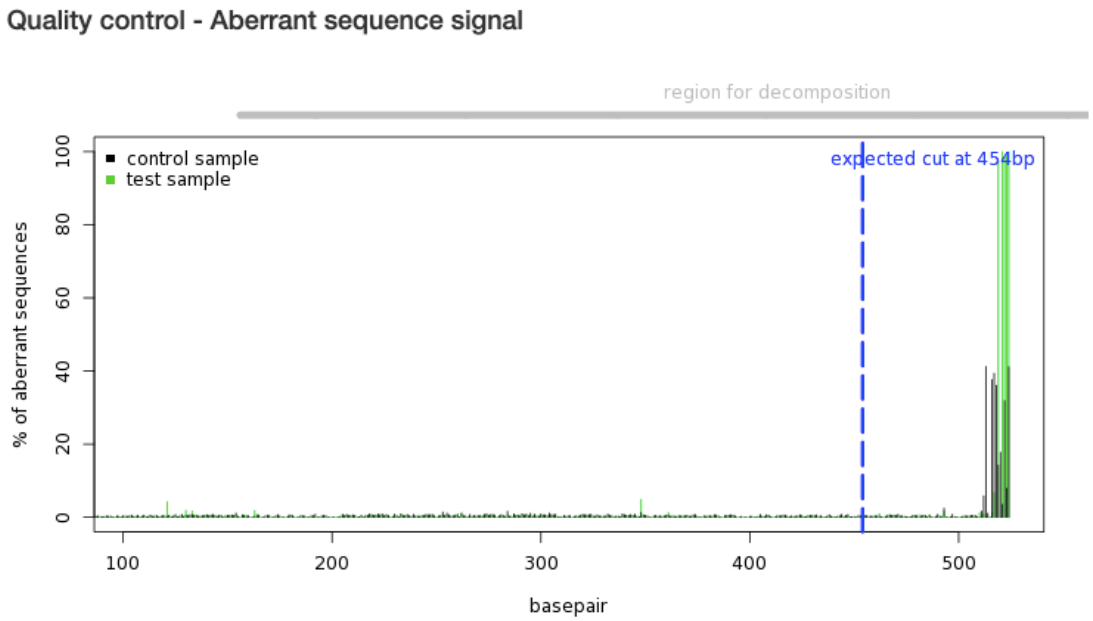
TIDE analysis: M-I55-G12.



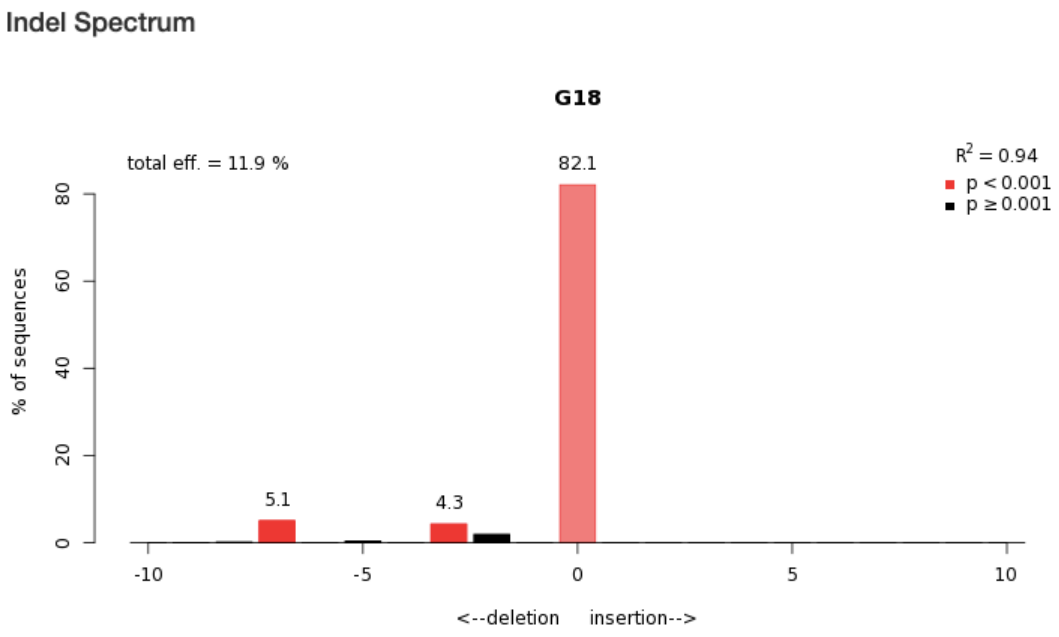


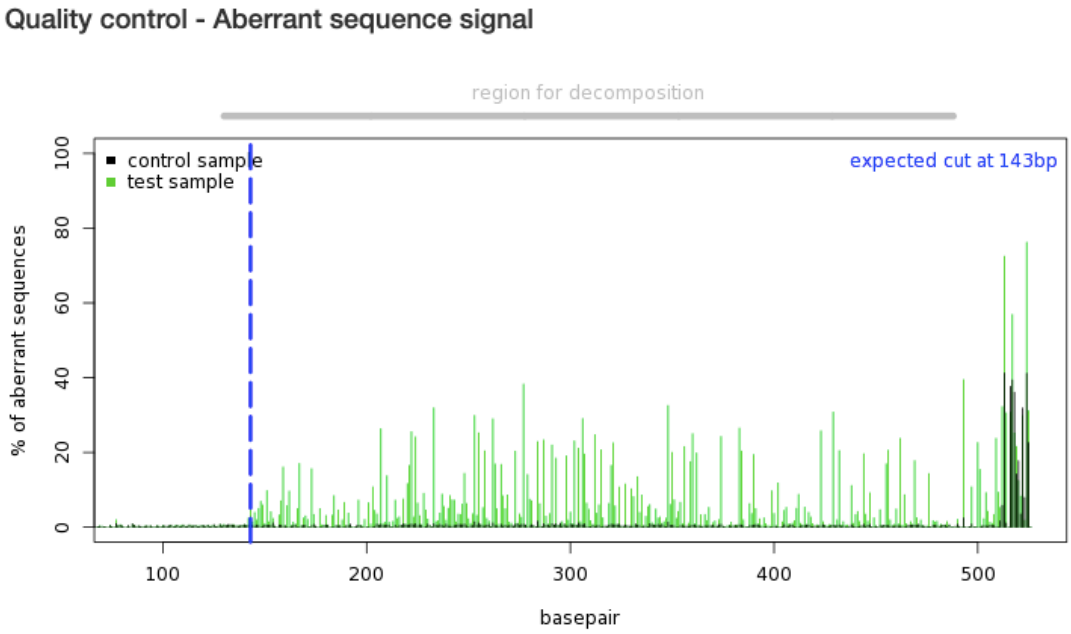
TIDE analysis: M-I55-G17.





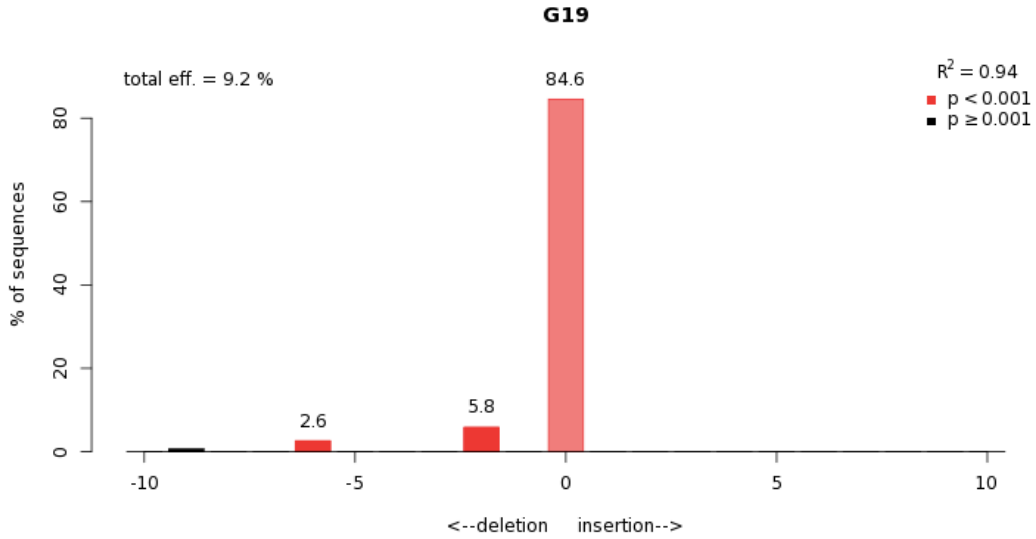
TIDE analysis: M-I55-G18.





TIDE analysis: M-I55-G19.

Indel Spectrum



Quality control - Aberrant sequence signal