**Whole-Genome sequencing of sporadic primary immunodeficiency cohort**

Primary immunodeficiency (PID) is characterized by recurrent and often life-threatening infections, autoimmunity and cancer, and it poses major diagnostic and therapeutic challenges. Although the most severe forms of PID are identified in

early childhood, most patients present in adulthood, typically with no apparent family history and a variable clinical phenotype of widespread immune dysregulation:

25% of patients have autoimmune disease, allergy is prevalent and up to 10% develop lymphoid malignancies1–3. Consequently, in sporadic (or non-familial) PID genetic diagnosis is difficult and the role of genetics is not well defined. Here we address these challenges by performing whole-genome sequencing in a large PID cohort of 1,318 participants. An analysis of the coding regions of the genome in 886 index cases of PID found that disease-causing mutations in known genes that are implicated in monogenic PID occurred in 10.3% of these patients, and a Bayesian approach (BeviMed4) identified multiple new candidate PID-associated genes, including *IVNS1ABP*. We also examined the noncoding genome, and found deletions in regulatory regions that contribute to disease causation. In addition, we used a genome-wide association study to identify loci that are associated with PID, and found evidence for the colocalization of—and interplay between—novel high-penetrance monogenic variants and common variants (at the *PTPN2* and *SOCS1* loci). This begins to explain the contribution of common variants to the variable penetrance and phenotypic complexity that are observed in PID. Thus, using a c oh or t- based whole-genome-sequencing approach in the diagnosis of PID can increase diagnostic yield and further our understanding of the key pathways that influence immune responsiveness in humans.

The phenotypic heterogeneity of PID leads to diagnostic difficulty, and almost certainly to an underestimation of its true incidence. Our cohort reflects this heterogeneity, although it is dominated by adult-onset, sporadic antibody-deficiency-associated PID (AD-PID)—comprising common variable immunodeficiency (CVID), combined immunodeficiency (CID) and isolated antibody deficiency. Identifying a specific genetic cause of PID can facilitate definitive treatment including haematopoietic stem cell transplantation, genetic counselling and the possibility of gene-specific therapy2, and can also contribute to our understanding of the human immune system5. Unfortunately, a genetic

cause of disease has been identified in only 29% of patients with PID6, with the lowest rate in patients who present as adults and have no apparent family history. Although variants in over 300 genes have been described as monogenic causes of PID3, it is frequently difficult to match the clinical phenotype to a known genetic cause, because phenotypes are heterogeneous and disease penetrance is often low2,7. Furthermore, a common-variant analysis of CVID identified new disease-associated loci, and raised the possibility that common variants may influence clinical presentation8. We therefore investigated whether applying whole-genome sequencing (WGS) across an unselected PID cohort might reveal the complex genetics of the range of conditions collectively termed PID. Our approach is summarized in Extended Data Fig. 1.

# Patient cohort

We sequenced the genomes of 1,318 individuals who were recruited as part of the PID domain of the UK National Institute for Health Research (NIHR) BioResource–Rare Diseases programme (NBR-RD) (Extended Data Fig. 2, Supplementary Methods). The cohort comprised patients with both sporadic and familial PID (*n* = 974) and family members. Of the patients, 886 were index cases who could be assigned to one of the diagnostic categories of the European Society for Immunodeficiencies (ESID) Registry ([https://esid.org/](https://esid.org/Working-Parties/Registry-Working-Party/Diagnosis-criteria)

[Working-Parties/Registry-Working-Party/Diagnosis-criteria)](https://esid.org/Working-Parties/Registry-Working-Party/Diagnosis-criteria) (Fig. 1a, Extended Data Table 1). These 974 patients represent a third of all the UK-registered patients with CVID and half of all those with CID9. Clinical phenotypes were dominated by adult-onset sporadic AD-PID: all patients had recurrent infections, 28% had autoimmunity and 8% had malignancy (Fig. 1a, b, Extended Data Table 2), mirroring the UK national PID registry6.

# Identifying pathogenic variants in known genes

We analysed the coding regions of genes in which disease-causing variants in PID have been previously reported10 (Methods). On the basis of the filtering criteria for diagnostic reporting that are provided in the guidelines of the American College of Medical Genetics11 and described in the Methods, we identified and reported to the referring clinicians 104 known or likely pathogenic variants in 91 index cases (10.3%) across 41 genes that are implicated in monogenic PID (Fig. 1c, Supplementary Table 1). Sixty patients (6.8%) had a previously reported pathogenic variant in the disease modifier *TNFRSF13B* (also known as *TACI*), increasing the proportion of cases with a reportable variant to 17.0% (151 patients). Notably, five patients with a monogenic diagnosis (in *BTK*, *LRBA*, *MAGT1*, *RAG2* or *SMARCAL1*) also had a pathogenic *TNFRSF13B* variant. Of the 103 monogenic variants that we report here, 69 (67.0%) had not been previously described (Supplementary Table 1) and 8 were structural variants—including deletions of single exons and noncoding regions of promoters, which are unlikely to have been detected by whole-exome sequencing12.

In 22 patients with variants in 14 genes (34% of 41 identified genes) that were reported as pathogenic, the clinical presentation deviated from the phenotypes that are typically associated with those genes. One example was chronic mucocutaneous candidiasis (CMC), which is the trigger for clinical genetic testing for *STAT1* gain-of-function variants as it was present in 98% of reported cases13,14. Our findings, together with single case reports15,16, now indicate that patients with *STAT1* gain-of-function variants may present with phenotypes as diverse as CVID or primary antibody deficiency. As many PID-associated genes were first discovered in a small number of familial cases, it is not surprising that the phenotypes described in the literature do not reflect the true clinical diversity. Thus, a cohort-based WGS approach to PID provides a diagnostic yield even in a predominantly sporadic cohort, allows diagnoses that are not constrained by pre-existing assumptions about genotype–phenotype relationships and suggests caution in the use of clinical phenotype in targeted gene screening and the interpretation of PID genetic data.

# Prioritizing candidate PID genes in a WGS cohort

We next investigated whether the cohort-based WGS approach could identify new genetic associations with PID. We included all 886 index cases in a single cohort, to optimize statistical power and because the correlation between genotype and phenotype in PID is incompletely understood. We used a Bayesian inference procedure, BeviMed4, to determine the posterior probability of association (PPA) between each gene and the case or control status of the 886 index cases and 9,284 unrelated controls (Methods). We obtained a BeviMed PPA for 31,350 genes in the human genome; the 25 highest-ranked genes are shown in Fig. 2a (see also Supplementary Table 2, Supplementary Note 2). Overall, genes with a BeviMed PPA greater than 0.1 were strongly enriched for known PID-associated genes (odds ratio = 15.1, *P* = 3.1 × 10−8, Fisher’s exact test), demonstrating that a statistical genetic association approach can identify genes that cause PID.

This method produces a posterior probability of association; it is therefore inevitable that, in cases in which the PPA value is less than 1, some of the genes identified will not be found to be causal. Such false positives are an integral feature of a method that does not provide statistical proof of causality, but rather ranks and prioritizes genes for subsequent functional assessment. They can be minimized by ensuring reasonable assumptions in the Bayesian algorithm4, and by taking care to detect and minimize relatedness and population stratification (see Methods, Supplementary Note 2, Supplementary Table 2).

*NFKB1* and *ARPC1B* were first associated with PID in the literature as a result of familial co-segregation studies17,18, and were highly ranked in the BeviMed analysis, validating it as a gene-discovery tool in PID. *NFKB1* had the strongest probability of association (PPA = 1 − (1.25 × 10−8))— driven by the presence of truncating heterozygous variants in 13 patients, which led to our previous report in which we identified *NFKB1* haploinsufficiency as the most common monogenic cause of CVID19. The association of *ARPC1B* with PID (PPA = 0.18) was identified by BeviMed on the basis of two recessive cases; the first of these has been described in a previous report18 and the other is described below.

To further demonstrate the effectiveness of BeviMed at prioritizing PID-related genetic variants in the cohort, we selected *IVNS1ABP* for validation. BeviMed enrichment (PPA = 0.33) of *IVNS1ABP* was driven by three independent heterozygous protein-truncating variants, suggesting haploinsufficiency; by contrast, no such variants were observed in control individuals (Fig. 2b). A pathogenic role for *IVNS1ABP* was supported by its intolerance to LOF mutations (probability of being LOF-intolerant (pLI) = 0.994), and by a distinctive clinical similarity between affected patients—all had severe warts (Supplementary Note 1). Expression of IVNS1ABP protein in heterozygous patients was around 50% of that in control individuals, consistent with haploinsufficiency (Fig. 2c). The patients also shared a previously undescribed peripheral leukocyte phenotype, with low or normal numbers of CD4+ T cells and B cells and aberrant increased expression of CD127 and PD-1 on naive T cells (Fig. 2d, e). Together, these data suggest that *IVNS1ABP* haploinsufficiency is a monogenic cause of PID (Supplementary Note 1).

The identification of both known and new PID-associated genes using BeviMed underlines the effectiveness of this method in cohorts of unrelated patients with sporadic disease. As the PID cohort grows, even causes of PID that are very rare should be detectable with a high positive predictive value (Extended Data Fig. 3).

# Regulatory elements that contribute to PID

Sequence variation within noncoding regions of the genome can have profound effects on gene expression and would be expected to contribute to susceptibility to PID. We combined rare-variant and large-deletion (more than 50 bp) events with a tissue-specific catalogue of *cis*-regulatory elements (CREs)20, generated using promoter capture Hi-C (pcHi-C)21, to prioritize putative causal genes for PID (Methods). We limited our initial analysis to rare large deletions that overlapped exon, promoter or ‘super-enhancer’ CREs of known PID genes. No homozygous deletions that affected CREs were identified, so we looked for individuals with two or more heterozygous variants that comprised a CRE deletion, together with either a rare coding variant or another large deletion in a pcHi-C-linked gene. Such compound heterozygote (cHET) variants have the potential to cause recessive disease. Out of 22,296 candidate cHET deletion events, after filtering by minor allele frequency (MAF), functional score and known PID gene status, we obtained ten events (Supplementary Table 3, Extended Data Fig. 4), and we describe the confirmation of three of these below.

The *LRBA* and *DOCK8* cHET variants were functionally validated (Extended Data Figs. 4, 5). In these two cases, the deletions encompassed both noncoding CREs and coding exons. However, using WGS PID cohorts to detect a contribution of CREs confined to the noncoding genome would represent a major advance in the pathogenesis and diagnosis of PID. *ARPC1B* fulfilled this criterion; its BeviMed probability of association was partially driven by a novel cHET variant (p.Leu247Glyfs\*25) in a patient, which results in a premature stop, and by a 9-kb deletion that spans the promoter region and includes an untranslated first exon (Fig. 3a). This deletion has no coverage in the ExAC database. Two unaffected first-degree relatives of this patient were heterozygous for the frameshift variant, and two for the promoter deletion (Fig. 3b), confirming compound heterozygosity in the patient. Western blotting demonstrated a complete absence of ARPC1B and an increase in the levels of ARPC1A in the platelets of the patient22 (Fig. 3c). *ARPC1B* mRNA was almost absent from mononuclear cells in the patient, and its levels were reduced in a clinically unaffected sister carrying the frameshift mutation (Supplementary Note 1). An allele-specific expression assay demonstrated that the promoter deletion essentially abolished mRNA expression (Supplementary Note 1). ARPC1B is part of the ARP2/3 complex that is necessary for the normal assembly of actin in immune cells23, and monocyte-derived macrophages from the patient had an absence of podosomes—phenocopying deficiency of the ARP2/3 regulator WASP (Fig. 3d).

Although examples of bi-allelic coding variants have been described as causing PID24,25, here we demonstrate the utility of WGS for detecting compound heterozygosity for a coding variant and a noncoding CRE deletion—a further advantage of a WGS approach to PID diagnosis. Improvements in analysis methodology and cohort size and better annotation of regulatory regions will be required to explore the noncoding genome more fully and discover additional disease-causing genetic variants.

# GWAS reveals PID-associated loci

The diverse clinical phenotypes and variable within-family disease penetrance of PID may be in part due to stochastic events (for example, unpredictable pathogen transmission), but might also have a genetic basis. We therefore performed a genome-wide association study (GWAS) of common single-nucleotide polymorphisms (SNPs) (MAF > 0.05) in 733 cases of AD-PID (Fig. 1a), restricted to reduce phenotypic heterogeneity (see Methods), and 9,225 unrelated NBR-RD controls, and then performed a fixed-effect meta-analysis of our AD-PID GWAS with a previous ImmunoChip study of CVID (778 cases, 10,999 controls)8. This analysis strengthened known associations of the major histocompatibility complex (MHC) locus and the 16p13.13 region with PID8, and also found suggestive associations, including at 3p24.1 within the promoter region of *EOMES* and at 18p11.21 proximal to *PTPN2*. We also examined SNPs of intermediate frequency (0.005 < MAF < 0.05) in AD-PID, and identified a *TNFRSF13B* p.Cys104Arg variant26 (odds ratio = 4.04, *P* = 1.37 × 10−12) (Fig. 4a, Extended Data Table 3, Extended Data Fig. 6, Supplementary Note 3). Conditional analysis of the MHC locus revealed independent signals at the class I and class II regions, driven by amino acid changes in the *HLA-B* and *HLA-DRB1* genes that are known to affect peptide binding (Extended Data Fig. 7). We next examined the enrichment of non-MHC-related AD-PID associations in nine other diseases, and found enrichment for allergic and immune-mediated diseases—suggesting that dysregulation of common pathways contributes to susceptibility to both these diseases and PID (Supplementary Note 4).

# GWAS enables identification of monogenic PID genes

To investigate whether loci identified by GWASs of AD-PID and other immune-mediated diseases might be used to prioritize candidate monogenic PID genes, we used the data-driven pcHiC omnibus gene score (COGS) approach21 (Methods, Supplementary Table 4). We selected six protein-coding genes with above-average prioritization scores in one or more diseases (Fig. 4b) and identified a single protein-truncating variant in each of *ETS1*, *SOCS1* and *PTPN2*, all of which occurred exclusively in patients with PID. We analysed the *SOCS1* and *PTPN2* variants in more detail.

SOCS1 limits the phosphorylation of target proteins that include STAT1, and is a key regulator of interferon γ (IFNγ) signalling27. The patient with a heterozygous de novo protein-truncating *SOCS1* variant (p.Met161Alafs\*46) presented with CVID complicated by lung and liver inflammation. GeneMatcher28 identified an independent pedigree with a protein-truncating variant p.Tyr64\* in *SOCS1*. All patients with *SOCS1* variants showed low or normal numbers of B cells, a population of CD4+ memory cells that was skewed towards T helper 1(TH1) cells, and a reduced number of CD4+CD25+ T regulatory (Treg) cells (Supplementary Note 1). *Socs1* haploinsufficient mice also exhibit B cell lymphopaenia27,29, a skew towards TH1 cells, decreased numbers of Treg cells30, and immune-mediated liver inflammation31. In T cell blasts from patients with *SOCS1* variants, the levels of SOCS1 protein were reduced and IFNγ-induced phosphorylation of STAT1 was increased (Fig. 4c). Together, this is consistent with *SOCS1* haploinsufficiency causing PID. The initial patient also carried the *SOCS1* pcHiC-linked 16p13.13 risk allele that was identified in the AD-PID GWAS (Supplementary Note 3) in *trans* with the novel *SOCS1*-truncating variant (Supplementary Note 1). Compound heterozygosity of this kind suggests that common and rare variants might combine to affect disease phenotype—a possibility that we explore further below.

A more detailed example of an interplay between rare and common variants is provided by a family of patients with variants in the *PTPN2* gene (Fig. 4d). *PTPN2* encodes the non-receptor T cell protein tyrosine phosphatase (TC-PTP), which negatively regulates immune responses by the dephosphorylation of proteins that mediate cytokine signalling. *Ptpn2*-deficient mice are B cell lymphopaenic32,33 and haematopoietic deletion of *Ptpn2* leads to the proliferation of B and T cells and autoimmunity34. A novel premature stop-gain mutation at p.Glu291 was identified in an individual with ‘sporadic’ PID. This patient presented with CVID at age 20, with B cell lymphopaenia, low levels of immunoglobulin G (IgG), rheumatoid-like polyarthropathy, severe recurrent bacterial infections, splenomegaly and inflammatory lung disease. His mother—also heterozygous for the *PTPN2* truncating variant—had systemic lupus erythematosus, insulin-dependent diabetes mellitus, hypothyroidism and autoimmune neutropaenia (Supplementary Note 1). Gain-of-function variants in *STAT1* can present as CVID (Supplementary Table 1), and TC-PTP—like SOCS1—reduces the levels of phosphorylated STAT1(Fig. 4e). Both mother and son showed a reduction in the expression of TC-PTP and hyperphosphorylation of STAT1—more pronounced in the son and similar to both *SOCS1* haploinsufficient and *STAT1* gain-of-function patients (Fig. 4f). Thus, *PTPN2* haploinsufficiency represents a new cause of PID that acts, at least in part, through increased phosphorylation of STAT1. The JAK1 and JAK2 inhibitor ruxolitinib has been used to control autoimmunity in patients with *STAT1* gain-of-function mutations35, suggesting that it could also be effective in treating patients who lack *SOCS1* or *PTPN2*.

The index case—but not his mother—carried the G allele of variant rs2847297 at the *PTPN2* locus—an expression quantitative trait locus (eQTL)36 that is associated with rheumatoid arthritis37. His brother, who was healthy apart from severe allergic nasal polyposis, was heterozygous at rs2847297 (thus also carried the G allele) and did not inherit the rare variant (Fig. 4d). Allele-specific expression analysis in the brother showed a reduction in *PTPN2* transcription from the rs2847297 G allele, which explains the lower expression of TC-PTP and greater persistence of phosphorylated STAT1 in the index case compared to his mother (Fig. 4g). This could explain the variable disease penetrance in this family, with *PTPN2* haploinsufficiency alone driving autoimmunity in the mother, but the additional effects of the common variant causing immunodeficiency in the index case. This family illustrates the strength of a cohort-wide WGS approach to PID diagnosis, by revealing both a new monogenic cause of disease, and how the interplay between common and rare genetic variants may contribute to the variable clinical phenotypes of PID.

In summary, we show that cohort-based WGS is a powerful approach in PID for diagnosing known genetic defects and discovering new coding and noncoding variants associated with disease (for a comparison of WGS with other methodologies, see Supplementary Note 5). Improved analysis methodology and better integration of parallel datasets, such as GWAS and cell-surface or metabolic immunophenotyping, will allow further exploration of the noncoding space and thereby enhance diagnostic yield. Such approaches promise to transform our understanding of genotype–phenotype relationships in PID and related immune-mediated conditions, and could redefine the clinical boundaries of immunodeficiency, add to our understanding of human immunology and ultimately improve patient outcomes.

1. Gathmann, B. et al. Clinical picture and treatment of 2212 patients with common variable immunodeficiency. *J. Allergy Clin. Immunol*. **134**, 116–126 (2014).
2. Lenardo, M., Lo, B. & Lucas, C. L. Genomics of immune diseases and new therapies. *Annu. Rev. Immunol*. **34**, 121–149 (2016).
3. Bousfiha, A. et al. The 2017 IUIS phenotypic classification for primary immunodeficiencies. *J. Clin. Immunol*. **38**, 129–143 (2018).
4. Greene, D., Richardson, S. & Turro, E. A fast association test for identifying pathogenic variants involved in rare diseases. *Am. J. Hum. Genet*. **101**, 104–114 (2017).
5. Casanova, J.-L. Human genetic basis of interindividual variability in the course of infection. *Proc. Natl Acad. Sci. USA* **112**, E7118–E7127 (2015).
6. Edgar, J. D. M. et al. The United Kingdom Primary Immune Deficiency (UKPID) registry: report of the first 4 years’ activity 2008–2012. *Clin. Exp. Immunol*. **175**, 68–78 (2014).
7. Pan-Hammarström, Q. et al. Reexamining the role of TACI coding variants in common variable immunodeficiency and selective IgA deficiency. *Nat. Genet*. **39**, 429–430 (2007).
8. Li, J. et al. Association of *CLEC16A* with human common variable immunodeficiency disorder and role in murine B cells. *Nat. Commun*. **6**, 6804 (2015).
9. Shillitoe, B. et al. The United Kingdom Primary Immune Deficiency (UKPID) registry 2012 to 2017. *Clin. Exp. Immunol*. **192**, 284–291 (2018).
10. Bousfiha, A. et al. The 2015 IUIS phenotypic classification for primary immunodeficiencies. *J. Clin. Immunol*. **35**, 727–738 (2015).
11. Richards, S. et al. Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. *Genet. Med*. **17**, 405–424 (2015).
12. Fromer, M. et al. Discovery and statistical genotyping of copy-number variation from whole-exome sequencing depth. *Am. J. Hum. Genet*. **91**, 597–607 (2012).
13. van de Veerdonk, F. L. et al. *STAT1* mutations in autosomal dominant chronic mucocutaneous candidiasis. *N. Engl. J. Med*. **365**, 54–61 (2011).
14. Liu, L. et al. Gain-of-function human *STAT1* mutations impair IL-17 immunity and underlie chronic mucocutaneous candidiasis. *J. Exp. Med*. **208**, 1635–1648 (2011).
15. Breuer, O. et al. Autosomal dominant gain of function STAT1 mutation and severe bronchiectasis. *Respir. Med*. **126**, 39–45 (2017).
16. Toubiana, J. et al. Heterozygous *STAT1* gain-of-function mutations underlie an unexpectedly broad clinical phenotype. *Blood* **127**, 3154–3164 (2016).
17. Fliegauf, M. et al. Haploinsufficiency of the NF-κB1 subunit p50 in common variable immunodeficiency. Am. *J. Hum. Genet*. **97**, 389–403 (2015).
18. Kuijpers, T. W. et al. Combined immunodeficiency with severe inflammation and allergy caused by *ARPC1B* deficiency. *J. Allergy Clin. Immunol*. **140**, 273–277.e10 (2017).
19. Tuijnenburg, P. et al. Loss-of-function nuclear factor κB subunit 1 (*NFKB1*) variants are the most common monogenic cause of common variable immunodeficiency in Europeans. *J. Allergy Clin. Immunol*. **142**, 1285–1296 (2018).
20. Hnisz, D. et al. Super-enhancers in the control of cell identity and disease. *Cell* **155**, 934–947 (2013).
21. Javierre, B. M. et al. Lineage-specific genome architecture links enhancers and non-coding disease variants to target gene promoters. *Cell* **167**, 1369–1384 (2016).
22. Kahr, W. H. A. et al. Loss of the Arp2/3 complex component ARPC1B causes platelet abnormalities and predisposes to inflammatory disease. *Nat. Commun*. **8**, 14816 (2017).
23. Burns, S., Cory, G. O., Vainchenker, W. & Thrasher, A. J. Mechanisms of WASp-mediated hematologic and immunologic disease. *Blood* **104**, 3454–3462 (2004).
24. Engelhardt, K. R. et al. Identification of heterozygous single- and multi-exon deletions in *IL7R* by whole exome sequencing. *J. Clin. Immunol*. **37**, 42–50 (2017).
25. Schepp, J. et al. Deficiency of adenosine deaminase 2 causes antibody deficiency. *J. Clin. Immunol*. **36**, 179–186 (2016).
26. Salzer, U. et al. Mutations in *TNFRSF13B* encoding TACI are associated with common variable immunodeficiency in humans. *Nat. Genet*. **37**, 820–828 (2005).
27. Alexander, W. S. et al. SOCS1 is a critical inhibitor of interferon γ signaling and prevents the potentially fatal neonatal actions of this cytokine. *Cell* **98**, 597–608 (1999).
28. Sobreira, N., Schiettecatte, F., Valle, D. & Hamosh, A. GeneMatcher: a matching tool for connecting investigators with an interest in the same gene. *Hum. Mutat*. **36**, 928–930 (2015).
29. Starr, R. et al. Liver degeneration and lymphoid deficiencies in mice lacking suppressor of cytokine signaling-1. *Proc. Natl Acad. Sci. USA* **95**, 14395–14399 (1998).
30. Horino, J. et al. Suppressor of cytokine signaling-1 ameliorates dextran sulfate sodium-induced colitis in mice. *Int. Immunol*. **20**, 753–762 (2008).
31. Yoshida, T. et al. SOCS1 is a suppressor of liver fibrosis and hepatitis-induced carcinogenesis. *J. Exp. Med*. **199**, 1701–1707 (2004).
32. Bourdeau, A. et al. TC-PTP-deficient bone marrow stromal cells fail to support normal B lymphopoiesis due to abnormal secretion of interferon-γ. *Blood* **109**, 4220–4228 (2007).
33. You-Ten, K. E. et al. Impaired bone marrow microenvironment and immune function in T cell protein tyrosine phosphatase-deficient mice. *J. Exp. Med*. **186**, 683–693 (1997).
34. Wiede, F., Sacirbegovic, F., Leong, Y. A., Yu, D. & Tiganis, T. PTPN2-deficiency exacerbates T follicular helper cell and B cell responses and promotes the development of autoimmunity. *J. Autoimmun*. **76**, 85–100 (2017).
35. Forbes, L. R. et al. Jakinibs for the treatment of immune dysregulation in patients with gain-of-function signal transducer and activator of transcription 1 (*STAT1*) or *STAT3* mutations. *J. Allergy Clin. Immunol*. **142**, 1665–1669 (2018).
36. Kilpinen, H. et al. Common genetic variation drives molecular heterogeneity in human iPSCs. *Nature* **546**, 370–375 (2017).

Okada, Y. et al. Meta-analysis identifies nine new loci associated with rheumatoid arthritis in the Japanese population. *Nat. Genet*. **44**, 511–516 (201