FOXF2, a novel risk locus for stroke and small artery disease: a genome-wide association study

The Neurology Working Group of the Cohorts for Heart and Aging Research in Genomic Epidemiology (CHARGE) Consortium, the Stroke Genetics Network (SiGN), and the International Stroke Genetics Consortium (ISGC)[†]

[†]Authors listed at end of the paper

Corresponding author:

Stéphanie Debette, MD

Inserm U1219, University of Bordeaux, 146 rue Leo Saignat, 33076 Bordeaux, France; Phone: +33-5-57-57-16-59; Fax: +33-5-47-30-42-09; E-mail: stephanie.debette@u-bordeaux.fr &

Department of Neurology, Boston University School of Medicine, 72 East Concord Street, B602; Boston, MA 02118, USA; Phone: 617-638-2744; Fax: 617-638-8086; E-mail: sdebette@bu.edu

For details on participating studies please see http://www.chargeconsortium.com/ and http://strokegenetics.org/

Summary

Background Genetic determinants of stroke, the leading neurological cause of death and disability, are poorly understood and have seldom been explored in the general population. Our aim was to identify additional loci for stroke by conducting a meta-analysis of genome wide association studies.

Methods We performed a genome-wide screen for common genetic variants associated with incident stroke risk in 18 prospective population-based cohorts comprising 84,961 participants, of whom 4,348 experienced stroke. Stroke (as per WHO definition) was ascertained and validated prospectively by study investigators. Mean age at stroke ranged between 45·8 and 76·4 years in the cohorts, and data collection took place between 1948 and 2013. We followed-up variants yielding an association at $p<5\times10^{-6}$ with all stroke, ischemic stroke, cardioembolic, or non-cardioembolic ischemic stroke in the largest available cross-sectional studies (70,804 participants of whom 19,816 experienced stroke). Summary-level results of discovery and follow-up stages were combined using inverse variance-weighted fixed-effects meta-analysis and look-up was performed in stroke sub-types. For genome-wide significant findings ($p<5\times10^{-8}$), we explored associations with additional cerebrovascular phenotypes and undertook functional analyses by conditional (inducible) deletion of the likely causal gene in mouse and also studied the expression of the latter and effects on cerebral vasculature in zebrafish mutants.

Findings We replicated seven of eight known loci for ischemic stroke and identified a novel locus at chr6p25 (rs12204590, near *FOXF2*) associated with risk of all stroke (odds ratio [OR] = 1·08, 95% CI 1·05-1·12, $p=1\cdot48\times10^{-8}$ [minor allele frequency 21%]). The rs12204590 stroke risk allele also increased MRI-defined white matter hyperintensity (WMH) burden, a marker of cerebral small artery disease, in stroke-free adults (N=21,079; p=0·0025). Consistently, young patients **(**age range 2-32 years**)** with segmental deletions of *FOXF2* exhibited extensive WMH burden. Deletion of *Foxf2* in adult mice resulted in cerebral infarction, reactive gliosis, and microhemorrhage. The zebrafish equivalents of *FOXF2* (orthologs) foxf2b/foxf2a were expressed in brain pericytes and mutant foxf2b-/- cerebral vessels showed decreased smooth muscle cell and pericyte coverage.

Interpretation In our study of 155,765 persons in total (24,164 with stroke), we identified common variants near *FOXF2* associated with increased stroke susceptibility. Extensive epidemiological and experimental data suggest that *FOXF2* mediates this association, potentially via differentiation defects of cerebral vascular mural cells. Further expression studies in appropriate human (including fetal) tissues and further functional experiments with longer follow-up periods are required to fully understand the underlying mechanisms.

Funding NIH, NHMRC, CIHR, European national research institutions, Fondation Leducq

Research in context

Evidence before this study

Stroke is the leading neurological cause of death and disability worldwide. A substantial proportion of stroke risk remains unexplained, and a contribution of genetic factors is supported by recent discoveries of common genetic variants associated with stroke risk, identified through large, collaborative, genome-wide association studies (GWAS). However, most of these showed association with cardioembolic and large artery ischemic stroke and no robust genetic association has been reported for other subtypes, especially the very common but poorly understood small artery ischemic stroke. Genetic associations with all stroke are also scarce, with only few genetic studies looking comprehensively at incident stroke in a population-based longitudinal setting.

Added value of this study

This study is novel by several aspects. First, leveraging on extensive phenotypic and genotypic information from very large population-based and hospital-based studies we identified a novel risk locus for stroke that appears to be mediated by small artery disease. Although small artery disease is one of the major subtypes of stroke, GWAS have failed so far to discover risk loci for small artery ischemic stroke (except for an association with the *PRKCH* locus identified in a Japanese study, which was not found in European populations). Second, while GWAS have successfully identified numerous genetic associations with complex diseases, including stroke, biological mechanisms underlying these associations still remain elusive for most of the variants, precluding any clinical applications beyond risk prediction. Here we provide important preliminary experimental evidence from zebrafish and mouse models that the observed statistical association reflects an effect of the nearby transcription factor *FOXF2* (a gene predominantly expressed in fetal tissue), on the development of cerebral vasculature. Conditional deletion of *Foxf2* in adult mice led to cerebral infarction, reactive gliosis and, to a lesser extent, microhemorrhage. In zebrafish *foxf2b*$^{-/-}$ zebrafish mutants showed decreased smooth muscle cell and pericyte coverage. Third, we show that patients with a rare monogenic ophthalmologic condition due to segmental deletions encompassing *FOXF2* also exhibit features of cerebral small artery disease, providing an example of how monogenic conditions can inform the mechanisms of complex diseases.

Implications of all the available evidence

The present findings provide important novel insight into the genetic underpinnings of stroke, especially of the small artery subtype, with strong evidence from multiple approaches for a pivotal role of *FOXF2*, a neural crest expressed transcription factor involved in cerebral vessel development. Cerebral small artery disease is a major, but poorly understood, cause of stroke in all ethnic groups, and "subclinical" small artery disease (here also showing association with the stroke risk variants near *FOXF2*) has been associated with progressive functional and cognitive decline and increased risk of dementia. Currently no mechanism-based treatment is available for small artery disease, other than management of risk factors. The present findings provide promising grounds for follow-up, pointing to a possible novel mechanism of stroke and small artery disease. Further research is warranted to explore whether this can translate into clinical applications.

Introduction

Stroke represents the leading neurological cause of death and disability worldwide.[1] A substantial proportion of stroke risk remains unexplained, and a contribution of genetic factors is supported by recent discoveries of common genetic variation associated with stroke risk, identified through large, collaborative, genome-wide association studies (GWAS).[2] These studies have estimated the proportion of phenotype variance explained by the genome-wide genotypes to range between 16-40% for ischemic stroke and between 34-73% for intracerebral hemorrhage.[3] Most associations to date have been specific to certain ischemic or hemorrhagic stroke types, although a few risk loci for overall ischemic stroke (IS) have also been reported.[2] Overall, the search for stroke loci has been less successful than for other complex phenotypes.[4] Potential explanations include heterogeneity of stroke and also limited ability to detect genetic variants increasing both stroke risk and severity, due to the cross-sectional design of most studies, with hospital-based case ascertainment and non-inclusion of severe strokes with early mortality. Population-based cohort studies, with blood samples drawn at recruitment and prospective incident stroke ascertainment, offer the advantage of including severe strokes leading to early death.

We performed a genome-wide screen for common genetic variants associated with an increased risk of incident stroke in prospective population-based cohort studies and followed up these results in the largest available cross-sectional studies. Detailed functional exploration of novel genome-wide significant association was conducted in zebrafish and mice.

Methods

Study population for discovery GWAS

The GWAS of incident stroke comprised 84,961 participants of European origin from 18 community-based prospective cohort studies participating in the Cohorts of Heart and Aging Research in Genomic Epidemiology (CHARGE) consortium. All participants were free of stroke at baseline and 4,348 developed incident stroke during 10 years of average follow-up (table 1, appendix). Some but not all of the cohorts included in the present analysis (representing <1,544 incident stroke cases) have participated in published HapMap-based stroke GWAS.[5,6]

The study was approved by the ethics committee of the participating studies and written informed consent was obtained by all study participants.

Stroke definition and classification of subtypes

Stroke was defined as a focal neurologic deficit of presumed vascular origin with sudden onset and lasting for ≥24 hours, or until death if the participant died <24 hours after onset of symptoms. Strokes were classified as IS (N=3,028), intracerebral hemorrhage (ICH, N=277), or unknown type based on clinical and imaging criteria; for cohorts where IS subtypes were available (Table 1), IS was subdivided into cardioembolic (N=602) and non-cardioembolic (N=1,770) subtypes. Numbers were too small to analyze large artery IS (N=117) and small artery IS (N=87) separately in this discovery dataset. Subarachnoid hemorrhage was not considered due to distinct mechanisms and very small numbers. Detailed definitions of stroke types and subtypes are given in the appendix.

Genotyping and imputation

Genotyping platforms and quality control filters are described in supplementary tables 1-2. All but one study used imputed genotypes based on the 1000GpIv3 "All" reference panel (supplementary tables 3-4).

Genome-wide association analyses

Using genome-wide multivariable Cox regression, we tested associations of genetic variants with incident stroke (all stroke, IS, cardioembolic IS, non-cardioembolic IS, and in secondary analyses ICH) under an additive genetic model, adjusting for sex, age, and when relevant principal components of population stratification, study site, or familial structure (appendix, supplementary table 5). Meta-analysis of study-specific association statistics was performed at two sites (G.C. and A.Y.C.) using inverse-variance weighted meta-analysis with METAL (http://csg.sph.umich.edu/abecasis/Metal/). The QQ plots and values of the genomic inflation factor lambda suggest no systematic inflation of test statistics due to population stratification, cryptic relatedness, or technical artefacts (supplementary table 6 and supplementary figure 1). Power of the discovery stage to detect association with various stroke subtypes is presented in supplementary figure 2.

Follow-up of most significant genetic associations

We selected variants with high imputation accuracy (mean R^2>0·80) yielding an association $p<5\times10^{-6}$ with all stroke, IS, cardioembolic, or non-cardioembolic IS. In total 177 variants belonging to 21 loci (linkage disequilibrium [LD] r^2>0·7 within each locus) were selected. In silico follow-up association analyses were performed in four independent cross-sectional studies, with mostly hospital-based stroke ascertainment, totaling 19,816 stroke patients (table 1) and 50,988 controls, from the Stroke genetics network (SiGN),[7,8] METASTROKE,[6] Heart and Vascular Health 1 (HVH1),[9] and Cervical Artery Dissections and IS Patients (CADISP) studies.[10] Except for 4,963 African-American and 3,371 Hispanic participants (cases and controls) in SiGN, all follow-up samples were of European ancestry. Follow-up analyses were performed with the same, or most similar, phenotype as in discovery (table 1), using logistic regression under an additive genetic model, followed by inverse-variance weighted meta-analysis of study-specific association statistics (appendix).

After Bonferroni correction for the number of independent loci (r^2<0.01 reflecting absence of linkage disequilibrium), $p<2\cdot38\times10^{-3}$ was considered significant evidence of replication. We did not correct for the number of stroke phenotypes (all stroke, IS, cardioembolic, and non-cardioembolic IS) as they are not independent (the latter being subtypes of the former). Only loci reaching genome-wide significance, at $p<5\times10^{-8}$, in the combined meta-analysis of discovery and follow-up samples were given further consideration.

Secondary analyses

We examined the association of novel genome-wide significant all-stroke risk variants with stroke subtypes in CHARGE and follow-up samples (using TOAST subtyping [11] and in sensitivity analyses the Causative Classification System (CCS) implemented in SiGN[7,8]). We examined whether the same variants were associated with white matter hyperintensity (WMH) burden, a quantitative MRI-marker of cerebral small artery disease, in 21,079 community-dwelling participants,[12] and with ICH in an independent study comprising 1,576 patients (682 lobar, 894 deep ICH patients) and 1,303 controls of European ancestry.[13] We tested whether risk variants with genome-wide or suggestive association for all-stroke or IS in the population-based discovery stage were associated with incident fatal and non-fatal stroke. We also looked for a significant association of these risk variants comparing fatal and non-fatal stroke, which

would suggest different genetic influences in the two groups of cases, assuming systematic bias in follow-up does not influence the frequency of the candidate variant in either group.

Functional exploration of novel stroke risk locus

Based on in silico functional annotation (appendix) and literature review of the novel genome-wide significant stroke risk locus identified by the aforementioned approach, we examined the effect of loss of function of the putative causal gene on brain vasculature and stroke-related phenotypes, in humans, mice, and zebrafish (appendix).

To complement the genome wide analysis, we utilized a rare cohort of Axenfeld-Rieger syndrome patients with deletions of the novel stroke risk locus to directly determine if loss of this locus resulted in more severe cerebro-vascular MRI phenotypes. By extracting data from individual MRI slices, we calculated the volume of WMH in two patients with large segmental deletions encompassing the putative causal gene (aged 2 and 32 years) and two patients with smaller deletions in whom this gene was intact (aged 15 and 17 years).

The putative causal gene was deleted in adult (12 weeks) conditional (floxed) mouse mutants by Cre-ERT2, an inducible Cre recombinase, as previously described,[14] and sacrificed six weeks later. Brains of conditional knockouts and controls were examined by histology (using hematoxylin/eosin or Richardson's methylene blue/Azure II) and glial fibrillary acidic protein (GFAP) immunofluorescence (DAKO Z0334), to search for features of vascular brain injury.

Zebrafish permit live imaging of blood vessel and mural cell interactions with exceptional clarity, using transgenic lines which permit in vivo visualization of endothelium and smooth muscle. Expression of the putative causal gene in the brain of zebrafish larvae was assessed by in situ hybridization, and compared to that of established pericyte markers, *notch3*, and *pdgfrβ*.[15] Function of *FoxF2* was assessed by knockout using transcription activator-like effector nucleases (TALENs) to create targeted nonsense mutations in the DNA binding domain (supplementary figure 3). Smooth muscle cell coverage of branches of brain vessels was examined in live transgenic mutant and wild type zebrafish embryos; these cells were modified to express green fluorescent protein using the α-smooth muscle actin promoter (acta2:GFP) at 4-6 days post-fertilization. Pericyte density was assessed by *pdgfrβ* in situ hybridization.

Role of the funding source

The funder was involved in the study design but had no role in data collection.

Results

In the population-based discovery stage (4,348 stroke cases versus 80,613 controls) 177 genetic variants in 21 independent loci reached $p<5\times10^{-6}$ in association with incident all-stroke, IS, cardioembolic, or non-cardioembolic IS (supplementary table 7). Eleven loci showed suggestive association at $p<5\times10^{-6}$ with incident all-stroke or IS. Ten additional loci showed association at $p<5\times10^{-6}$ with incident cardioembolic IS, one locus (lead-SNP rs72794386, in *SLC12A2*) showing genome-wide significance: HR=1·67 (95%CI:1·39-2·00), $p=4\cdot37\times10^{-8}$ (minor allele frequency [MAF]= 10%) (table 2, supplementary figure 4).

In the cross-sectional follow-up samples (19,816 stroke cases versus 50,988 controls) two loci replicated at $p<2\cdot38\times10^{-3}$ and reached genome-wide significance ($p<5\times10^{-8}$) in the combined analysis (table 3). The first one is a novel locus (chr6p25·3, lead SNP rs12204590),

lying between *FOXQ1* and *FOXF2*, showing association with incident all-stroke: combined OR=1·08(1·05-1·12), $p=1{\cdot}48\times10^{-8}$ (MAF=21%, figure 1). Associations in each study are shown in supplementary figure 5. The second (chr4q25, near *PITX2*) is a known risk locus for cardioembolic IS: combined OR=1·37(1·29-1·46), $p=4{\cdot}72\times10^{-24}$ for incident cardioembolic IS (MAF=12%). The *SLC12A2* locus (genome-wide significant in the small cardioembolic IS discovery sample) did not show evidence of replication (p=0·27).

We also explored association of the chr6p25·3 locus with stroke subtypes. In the discovery sample, lead SNP rs12204590 was associated with incident IS: HR=1·13[1·06-1·20], $p=1{\cdot}64\times10^{-4}$ (non-cardioembolic IS: HR=1·12[1·04-1·22], $p=4{\cdot}35\times10^{-3}$; cardioembolic IS: HR=1·10[0·95-1·27], p=0·21). In follow-up samples, we observed association with small artery IS: OR=1·08[1·02-1·14], p=0·0094 for rs12200309 (in complete LD with rs12204590), using TOAST subtypes, while association with large artery and cardioembolic IS was non-significant (p>0·35). The association with small artery IS was even more marked when using CCS-causative subtyping where available (SiGN): OR=1·11[1·05-1·18], p=0·00029 (supplementary table 8). We also observed significant association of chr6p25·3 with increasing WMH burden in the general population (lowest p=0·0025, supplementary table 9). The HR for association of rs12204590 with incident fatal IS (n=271, HR=1·21[0·99-1·50], p=0·0684) was higher than with non-fatal IS (n=2300, HR=1·14[1·06-1·22], $p=4{\cdot}93\times10^{-4}$), but the difference was not statistically significant (supplementary tables 10-11). We did not observe any heterogeneity by ethnicity for associations between the chr6p25 locus and stroke risk (supplementary table 12).

The genomic region where the variant associated with increased stroke risk and adjacent linked variants (linkage disequilibrium $r^2>0.50$) are located, appears to include enhancers (regions of DNA that activate transcription of nearby genes). This genomic region also includes DNAaseI hypersensitive regions, a marker of open chromatin associated with active cis-regulatory elements important for transcription of nearby genes (figure 1, appendix, supplementary tables 13-14). Two SNPs (rs7750826 and rs2006798, r2>0•75 with rs12204590) had RegulomeDB scores of "2b" suggesting a likely role in regulating gene expression (combination of transcription factor binding site, and DNAse peak and footprint, figure 1 and supplementary table 15). The genomic region that includes the lead variant and variants in LD with it includes two protein coding genes *FOXQ1* and *FOXF2*. The same region also includes a microRNA (*MIR6720*). However if we expand the regional plot to a 1 MB region around the lead variant, the region also includes two other protein coding genes (*FOXC1* and *GMDS*) and a long non-coding RNA (LINC01622) (figure 1). We performed an extensive search of publically available eQTL[16] and miRNA databases[17], examined mRNA expression of *FOXF2* and adjacent genes in dorsolateral prefrontal cortex of 508 persons enrolled in the Religious Orders Study and the Rush Memory and Aging Project,[18] and mined large sets of epigenomic data from the International Human Epigenome Consortium (www.ihec.org) (supplementary tables 16-17). Expression quantitative trait loci or methylation QTLs in this region were lacking, the only eQTL observed was for a long non-coding RNA (LOC285768) in the human brain ($p=5{\cdot}25\times10^{-7}$ for rs7746700, average in ten brain regions, http://www.braineac.org/). However, the different distribution of histone modifications associated with active genes in cells expressing *FOXF2* or *FOXQ1* suggested that these variants are likely to lie within the regulatory region of *FOXF2* (supplementary figure 6). Moreover, we observed lowest CpG methylation levels (indicating highest activity) at this locus in fetal brain as compared to any other tissue.[19]

We have previously described that patients with Axenfeld-Rieger syndrome, a rare heterogeneous condition with maldevelopment of the ocular anterior segment, attributable to mutation or copy number variation of *FOXC1*, adjacent to *FOXF2* on chr6p25, have increased burden of MRI-markers of cerebral small artery disease.[20] Within this cohort of *FOXC1*-

attributable Axenfeld-Rieger syndrome we identified two patients with 300kb segmental deletions encompassing both *FOXC1* and *FOXF2* and found that they had extensive, confluent WMH, with more than ten-fold larger WMH volume than two patients with segmental deletions of *FOXC1* only (30kb), although small numbers do not allow a formal statistical comparison. All patients were younger than 35 years (age range 2-32 years) and lacked vascular risk factors (figure 2A-E, supplementary table 18). White matter hyperintensities are normally absent or negligible in this age-range in the general population.[17,21]

We next used mice and zebrafish, two common model organisms for brain and vascular development, to show that disturbances of FoxF2 expression are associated with neurovascular pathology and vascular mural cell defects. We recently showed that murine *Foxf2* is expressed in vascular mural cells, specifically in the central nervous system, and is essential for pericyte differentiation, vascular maturation and formation of the blood-brain barrier (BBB).[14] Moreover, we demonstrated that inactivation of *Foxf2* in adult mice leads to gradual breakdown of the BBB and increased mortality.[14] To understand the mechanism at the tissue level we used histology of the brains from six mice six weeks after *Foxf2* inactivation and found areas of neurons with pyknotic nuclei and eosinophilic cytoplasm (figure 2F-F''), suggestive of ischemic infarction, in five of the brains. Patches with elevated levels of glial fibrillary acidic protein (GFAP) in astrocytes (figure 2G) indicated reactive gliosis. Higher magnification revealed a few instances of microhemorrhage with extravascular ertythrocytes (figure 2I). In contrast, brains from control mice contained only occasional and scattered neurons that showed signs of apoptosis, or isolated astrocytes with increased GFAP immuonoreactivity, and no visible hemorrhagic lesions (figure 2H). Mice in which *Foxf2* had been deleted had significantly higher mortality than control mice. Usually the animals were found dead, but some had to be euthanized after exhibiting behaviour indicative of brain damage, such as circling or lopsided gait.

In zebrafish we show that *foxf2a* and *foxf2b* (*FOXF2* orthologs) are expressed in pericytes closely wrapping the cerebral endothelium, similar to mice (figure 2J-K). Expression occurs in a pattern similar to established pericyte markers, *platelet derived growth factor receptor β* (*pdgfrβ)*, and *notch3* (figure 2L-M). We made two zebrafish knockout lines *foxf2b*^{ca22} *and foxf2b*^{ca23} with nonsense mutations in the first exon of *foxf2b* that would result in a translation block prior to the essential DNA binding domain (supplementary figure 3). Both alleles have identical phenotypes. *foxf2b* mutants had decreased expression of the brain pericyte marker *pdgfrβ* (Fig 2N-Q) suggesting pericyte maturation defects, and decreased acta2-positive smooth muscle cell coverage on large cerebral vessels suggestive of smooth muscle defects (figure 2R-U, supplementary figure 7). In homozygous mutants, acta2 was visible up to second order of cerebral vessel branching or less versus fourth or higher order branch in wildtype or heterozygous embryos.

Seven of eight risk loci for IS[2] identified in cross-sectional studies were associated with incident stroke in the population-based GWAS (discovery), predominantly in the same subtype as the original study (p-range=0·047-7·82×10^{-5}, supplementary table 19). One risk locus for IS (*PITX2*) also showed association with incident ICH (p=0·0031), and one risk locus for ICH (*PMF1-BGLAP*), also a risk locus for increasing WMH burden,[12] was associated with incident IS (p=0·00064), both in the same direction (please see appendix for more details).

Discussion

In a large population-based GWAS meta-analysis of incident stroke, with follow-up in the largest available cross-sectional stroke GWAS, we identified a novel genome-wide significant association of common variants in the chr6p25·3 region, near *FOXF2*, with risk of stroke. Associations predominated with small artery IS, and significant association was observed with WMH burden; in addition, patients with rare segmental deletions of *FOXF2* also showed extensive WMH burden. These findings suggest an effect of this locus on cerebral small artery disease, however the mechanism by which this transcription factor results in cerebral small artery disease and stroke is unclear. To support plausibility of *FOXF2* we undertook experimental studies in two animal species to examine its role in cerebral vessel development and stability. We demonstrate areas of infarction and to a lesser extent microhemorrhages in brains of conditional *Foxf2* mutant mice. We show in zebrafish that *foxf2b* is expressed in brain pericytes (as in mice[14]) and that reduction in *foxf2b* function leads to differentiation defects of both pericytes and smooth muscle cells in the developing cerebral vasculature.

Converging evidence, both from the present work and previous publications, suggests an important role of *Forkhead transcription factor 2* (*FOXF2*) in cerebrovascular disease. In mice, we recently showed that *Foxf2* is required for brain pericyte differentiation and BBB development, with *Foxf2*$^{-/-}$ embryos exhibiting thickened endothelium, perivascular edema, thinning of the vascular basal lamina, and leaky BBB.[14] We had also described that *Foxf2* inactivation in adult mice results in endothelial thickening and BBB breakdown,[14] an important mediator of cerebral small artery disease, and increased mortality.[22] In this conditional *Foxf2* mutant, we now analyzed mouse brains six weeks after *Foxf2* inactivation and interestingly we observed signs of brain infarction, with reactive astrogliosis, and to a lesser extent microhemorrhages. Of note, in prior experiments *Foxf2-/-* mouse embryos were found to develop intracranial hemorrhage,[14] while areas of microhemorrhage were scarce in conditional *Foxf2* mutant brains. We were also not able to demonstrate an association of the chr6p25.3 locus with intracerebral hemorrhage in the largest available collaborative genetic association study, although power may have been limited (supplementary table 20). In zebrafish, *foxf2b* knockout led to disruption of cerebral vasculature with decreased pericyte density and smooth muscle cell coverage (figure 2). *Foxf2* is first expressed in the neural crest and in mice regulates pathways involved in mural cell (pericyte and vascular smooth muscle cell) differentiation including the pdgfβ and serum response factor pathways.[14,23] We did not observe hemorrhage in the zebrafish *foxf2* mutant model during embryonic stages. We note that we have knocked out only one of two *foxf2* genes in fish and even though mural cell markers have changed expression in *foxf2b* mutants, there may be genetic compensation from the *foxf2a* gene that may make the phenotype less severe. However we have not found expression changes in foxf2a in foxf2b mutants. We cannot exclude haemorrhage occurring at juvenile or adult stages that we have not been able to examine. In summary our data suggest that association of *FOXF2* and stroke may arise from differentiation defects of cerebral vascular mural cells. It is beyond the scope of this study, but demonstrating the cellular requirement for *FoxF2* by expression under a vascular mural cell promoter for prevention of stroke and mural cell phenotypes in mutants is an important future experiment. No current promoters available in fish express in both smooth muscle and pericytes where we find *foxf2* expression.

Forkhead transcription factors are involved in various developmental and biological processes, and tend to be distributed in clusters on the genome.[24,25] The evolutionarily conserved chr6p25 cluster comprises *FOXQ1, FOXF2,* and *FOXC1*. Our lead stroke risk variants lie between *FOXQ1* (22·4kb) and *FOXF2* (52·7kb). In contrast with the compelling experimental evidence for a central role of *FOXF2* in cerebrovascular disease, *FOXQ1* has not been implicated in cerebrovascular phenotypes; mutant mice for this gene show mainly altered hair differentiation and gastric mucin secretion.[26,27] The third gene in the cluster, *FOXC1* (225kb downstream of

FOXF2 and 273.3kb downstream of stroke risk variants), was also recently shown to be expressed in the brain vasculature and influence vessel morphogenesis,[28] and arteriovenous specification.[29] MRI analysis of patients with *FOXC1*-attributable Axenfeld-Rieger syndrome revealed MRI-features of cerebral small artery disease (including increased WMH burden), and genetic variants downstream of *FOXC1* were associated with WMH burden in the general population.[20] These previously described variants are independent (r^2<0·017) from the stroke risk variants near *FOXF2* described here.[20,30] This, together with the tenfold WMH burden in patients with segmental deletion of both *FOXC1* and *FOXF2* versus *FOXC1* alone, suggests an independent role of *FOXF2* in cerebral small artery disease. Differences in the roles of *foxf2* and *foxc1a/b* are also seen in the zebrafish model. Knockdown of zebrafish *foxc1a/b* leads to embryonic cerebral hemorrhage in embryos,[31] while knockout of *foxf2b* at the same developmental stage, does not. *foxc1a/b* morphants showed unchanged *pdgfrβ* expression while *foxf2b* mutants show decreased *pdgfrβ* expression. We show that *foxf2b* mutants have cerebral smooth muscle defects while to date we have only examined ventral head smooth muscle defects in *foxc1a/b* morphants. Thus although the two genes are closely related, there are indications that their roles in vascular mural cells may be distinct.

Intriguingly, *PITX2* (chr4q25), a known risk locus for cardioembolic IS and atrial fibrillation,[32,33] also genome-wide significant for cardioembolic IS in the present sample, encodes a neural crest–expressed transcription factor that physically interacts with FOXC1 and harbors causal mutations for Axenfeld-Rieger syndrome.[20] Variants near *PITX2* were associated with WMH burden.[20,30] In the present study, common variants near *PITX2* were also associated with ICH risk, of which the main mechanism in the general population is small artery disease.[34] These observations suggest that *FOXF2*, *FOXC1* and *PITX2* could perhaps contribute to cerebrovascular disease via partly shared pathways, involving cerebral small arteries.

The relatively limited number of incident strokes (N=4,300), particularly stroke subtypes (N=602 for cardioembolic IS while small artery occlusion, large artery IS and other stroke subtypes had to be merged into a single category) may have hampered our ability to detect additional associations. We were also underpowered (<80% power) to detect associations with smaller effect sizes and lower allele frequencies (supplementary figure 2). While there are strong arguments from the described animal experiments suggesting that *FOXF2* is the causal gene underlying the observed genetic association with stroke and small artery disease at chr6p25, functional annotation of the identified risk variants is limited, with a lack of expression quantitative trait loci despite an extensive search of publicly available and other databases, possibly reflecting tissue specificity, or primarily developmental effects, as supported by higher expression of *FOXF2* in fetal than adult brain (www.ihec.org) and lower methylation levels of the stroke risk locus in fetal versus adult tissues. Another limitation is that we only explored common variants and not rare variants in this region. In addition, we have used an additive genetic model only, the most powerful approach when the underlying genetic model is unknown, but we cannot exclude that associations with genetic risk loci following a recessive or dominant model may have been missed. Nevertheless, we confirmed and extended the range of associations for previously discovered stroke risk loci, in a population-based sample. For the first time we describe shared genetic variation underlying both IS and ICH (chr4q25, chr1q22), in agreement with some monogenic strokes having both ischemic and hemorrhagic phenotypes, mostly with underlying small artery disease.[35-37] The association we previously described between chr12p13 and incident stroke in a smaller, overlapping sample was also the most significant association with IS in the present population-based GWAS,[5] showing a stronger association with incident fatal versus non-fatal stroke, suggesting an effect on stroke survival (supplementary table 12).

In summary, we identified common variants near *FOXF2* associated with increased stroke susceptibility, especially of the small artery subtype, and with extensive "subclinical" small artery disease. This is particularly interesting, as GWAS have failed so far in discovering risk loci for small artery IS (except for an association with the *PRKCH* locus identified in a Japanese study, which was not found in European populations).[6,38] Brain imaging data from patients with rare segmental deletions encompassing *FOXF2*, and extensive functional experiments across evolutionarily separated species, suggest an important role of *FOXF2* in cerebrovascular disease, especially cerebral small artery disease, possibly by affecting differentiation of cerebral vascular mural cells. Cerebral small artery disease is a major, but poorly understood, cause of stroke in all ethnic groups, and "subclinical" small artery disease has been associated with progressive functional and cognitive decline and increased risk of dementia. Currently no mechanism-based treatment is available, other than management of risk factors. The present findings provide promising grounds for follow-up, pointing to a possible novel mechanism of stroke and small artery disease. Further research is warranted to explore whether this can translate into clinical applications.

[†]The Neurology Working Group of the Cohorts for Heart and Aging Research in Genomic Epidemiology (CHARGE) Consortium, the Stroke Genetics Network (SiGN)[‡], and the International Stroke Genetics Consortium (ISGC):

Ganesh Chauhan* (PhD), Corey R. Arnold* (BSc), Audrey Y. Chu* (PhD), Myriam Fornage* (PhD), Azadeh Reyahi (PhD)*, Joshua C. Bis* (PhD), Aki S. Havulinna* (DSc), Muralidharan Sargurupremraj (PhD), Albert Vernon Smith (PhD), Hieab H.H. Adams (MSc), Seung Hoan Choi (MA), Sara L. Pulit (BA), Stella Trompet (PhD), Melissa E. Garcia (MPH), Ani Manichaikul (PhD), Alexander Teumer (PhD), Stefan Gustafsson (PhD), Traci M. Bartz (MS), Céline Bellenguez (PhD), Jean Sebastien Vidal (MD), Xueqiu Jian (PhD), Olafur Kjartansson (MD), Kerri L. Wiggins (MS), Claudia L. Satizabal (PhD), Flora Xue (MS), Samuli Ripatti (PhD), Yongmei Liu (PhD), Joris Deelen (PhD), Marcel den Hoed (PhD), Steve Bevan (PhD), Jemma C. Hopewell (PhD), Rainer Malik (PhD), Susan R. Heckbert (MD), Kenneth Rice (PhD), Nicholas L. Smith (PhD), Christopher Levi (MBBS), Pankaj Sharma (MD), Cathie LM Sudlow (BMBCh), Ali Moussavi Nik (DVM, PhD), John W. Cole (MD), Reinhold Schmidt (MD), James Meschia (MD), Vincent Thijs (MD), Arne Lindgren (MD), Olle Melander (MD), Raji P. Grewal (MD), Ralph L. Sacco (MD), Tatjana Rundek (MD), Peter M Rothwell (MD), Donna K. Arnett (PhD), Christina Jern (MD), Julie A. Johnson (PharmD), Oscar R. Benavente (MD), Sylvia Wassertheil-Smoller (PhD), Jin-Moo Lee (MD, PhD), Quenna Wong (MS), Hugo J. Aparicio (MD), Stefan T Engelter (MD), Manja Kloss (MD), Didier Leys (MD), Alessandro Pezzini (MD), Julie E. Buring (ScD), Paul M Ridker (MD), Claudine Berr (MD), Jean-François Dartigues (MD), Anders Hamsten (MD), Patrik K. Magnusson (PhD), Matthew Traylor (PhD), Nancy L. Pedersen (PhD), Lars Lannfelt (MD), Lars Lind (MD), Cecilia M. Lindgren (PhD), Andrew P. Morris (PhD), Jordi Jimenez-Conde (MD), Joan Montaner (MD), Farid Radmanesh (MD), Agnieszka Slowik (MD), Daniel Woo (MD), Albert Hofman (MD), Peter J. Koudstaal (MD), Marileen LP. Portegies (MD), André G. Uitterlinden (PhD), Anton JM de Craen (PhD), Ian Ford (MD), J. Wouter Jukema (MD), David J Stott (MD), Norrina B. Allen (PhD), Michele M. Sale (PhD), Andrew D Johnson (PhD), David A. Bennett (MD), Philip L. De Jager (MD, PhD), Charles C. White (PhD), Hans Jörgen Grabe (MD), Marcello Ricardo Paulista Markus (MD), Ulf Schminke (MD), Giorgio B Boncoraglio (MD), Robert Clarke (MD), Yoichiro Kamatani (MD), Jean Dallongeville (MD), for the Cervical Artery Dissection and Ischemic

Stroke Patients (CADISP) study, Oscar L Lopez (MD), Jerome I. Rotter (MD), Michael A. Nalls (PhD), Rebecca F. Gottesman (MD), Michael E. Griswold (PhD), David S. Knopman (MD), B. Gwen Windham (MD), Alexa Beiser (PhD), Hugh S Markus (DM), Erkki Vartiainen (MD), Curtis R. French (PhD), Martin Dichgans (MD), for the METASTROKE consortium, Tomi Pastinen (PhD), Mark Lathrop (PhD), Vilmundur Gudnason (MD), Tobias Kurth (MD), Bruce M. Psaty (MD), Tamara B. Harris (MD), Stephen S Rich (PhD), Anita L. deStefano (PhD), Carsten Oliver Schmidt (PhD), Bradford B Worrall (MD), Jonathan Rosand (MD), Veikko Salomaa (MD), Thomas H. Mosley (PhD), Erik Ingelsson (MD, PhD), Cornelia M. van Duijn (PhD), Christophe Tzourio (MD), Kathryn M. Rexrode (MD), Ordan J. Lehmann** (MD, PhD), Lenore J Launer** (PhD), M. Arfan Ikram** (MD), Peter Carlsson** (PhD), Daniel I. Chasman** (PhD), Sarah J Childs** (PhD), William T. Longstreth, Jr.** (MD), Sudha Seshadri** (MD), Stéphanie Debette** (MD)

* Equal contributions; ** Co-directed this work

‡ See appendix for more details on study organization

Affiliations

University Medicine Greifswald, Institute for Community Medicine, SHIP-KEF, Greifswald, Germany (AT, COS); University Medicine Greifswald, Department of Neurology, Greifswald, Germany (US); University Medicine Greifswald, Department of Psychiatry and Psychotherapy, Greifswald, Germany (HJG); University Medicine Greifswald, Department of Internal Medicine B, DZHK (German Centre for Cardiovascular Research), partner site Greifswald, Greifswald, Germany (MRPM); Division of Preventive Medicine, Brigham and Women's Hospital, Boston, MA, USA (JEB, PMRi, DIC, AYC, TK); Harvard Medical School, Boston, MA, USA (JEB, PMRi, DIC);; Inserm Research Center for Epidemiology and Biostatistics (U897) - Team Neuroepidemiology, Bordeaux, France (MS, CT, JFD, GC, SD); Institute of Public Health, Charité - Universitätzmedizin, Berlin, Germany (TK); University of Virginia, Charlottesville, VA USA (MMS, AM, SSR); Northwestern University, Chicago, IL, USA (NBA); National institute for Health and Welfare, Helsinki, Finland (ASH, EV, VS); Public Health, University of Helsinki, Finland (SR); Institute for Molecular Medicine Finland (FIMM), University of Helsinki (SR); Wellcome Trust Sanger Institute, UK (SR); Department of Epidemiology, Erasmus MC, Rotterdam, The Netherlands (HHHA, MLP, AHo, CMvD, MAI); Department of Neurology, Erasmus MC, Rotterdam, The Netherlands (MLP, PJK); Department of Internal Medicine, Erasmus MC , Rotterdam, The Netherlands (AGU); Cardiovascular Health Research Unit, Department of Medicine, University of Washington, Seattle, WA, USA (JCB, BMP, KLW); Department of Epidemiology, University of Washington, Seattle, WA, USA (BMP, WTL, NLS, SRH); Group Health Research Institute, Group Health Cooperative, Seattle, WA, USA (SRH); Department of Health Services, University of Washington, Seattle, WA, USA (BMP); Group Health Research Institute, Group Health Cooperative, Seattle, WA, USA (BMP); Institute for Translational Genomics and Population Sciences, Los Angeles Biomedical Research Institute at Harbor-UCLA Medical Center, Torrance, California, USA (JIR); Division of Genomic Outcomes, Department of Pediatrics, Harbor-UCLA Medical Center, California, USA (JIR); UCLA, California, USA (JIR); Department of Biostatistics, University of Washington, Seattle, WA, USA (TMB, KR, QW); Department of Neurology; University of Washington, Seattle, WA, USA (WTL); Seattle Epidemiologic Research and Information Center of the Department of Veterans Affairs Office

of Research and Development, Seattle, WA, USA (NLS); Department of Cardiology, Leiden University Medical Center, Leiden The Netherlands (ST, JWJ); Department of Gerontology and Geriatrics, Leiden University Medical Center, Leiden, The Netherlands (ST, AJMdC); Department of Molecular Epidemiology, Leiden University Medical Center, Leiden, The Netherlands (JDe); Robertson Center for Biostatistics, University of Glasgow, United Kingdom (IF); Institute of Cardiovascular and Medical Sciences, Faculty of Medicine, University of Glasgow, United Kingdom (DJS); Center for Human Genetics, Division of Public Health Sciences, Wake Forest School of Medicine, Winston Salem, NC, USA (YL); Laboratory of Epidemiology and Population Sciences, National Institute on Aging, Bethesda, MD, USA (TBH, MEGa); Laboratory of Neurogenetics, National Institute on Aging, Bethesda, MD, USA (MAN); Department of Medical Sciences, Molecular Epidemiology and Science for Life Laboratory, Uppsala University, Uppsala, Sweden (SG, MdH, EI); Department of Medicine, Division of Cardiovascular Medicine, Stanford University School of Medicine, Stanford, CA 94305, USA (EI); Department of Medical Sciences, Uppsala University, Uppsala, Sweden (LLi); Department of Public Health and Caring Sciences, Uppsala University, Uppsala, Sweden (LLa); Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, Stockholm, Sweden (NLP, PKM); Genetic and Genomic Epidemiology Unit, Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford, UK (CML, APM); Department of Biostatistics, University of Liverpool, Liverpool, UK (APM); Atherosclerosis Research Unit, Department of Medicine Solna, Karolinska Institutet, Stockholm, Sweden (AHa); University of Bordeaux, Bordeaux, France (MS, CT, JFD, GC, SD); Inserm, U1167, F-59000 Lille, France (CeB), Institut Pasteur de Lille, F-59000 Lille, France (CeB), Univ. Lille, U1167, F-59000 Lille, France (CeB); INSERM U1167, Université Lille, Lille, Lille, France (JDa, DL); Institut Pasteur de Lille, Lille, France (JDa); Department of Neurology, Lille University Hospital, Lille, France (DL); INSERM U1171, Lille, France (DL); INSERM U1061, Montpellier, France (ClB); University Montpellier 1, Montpellier, France (ClB); Department of Neurology, Bordeaux University Hospital, Bordeaux, France (SD, JFD); Boston University School of Medicine, Boston, MA, USA (SD); IRP/NIA/NIH, Bethesda, MD, USA (LJL); Department of Neurology, University of Pittsburgh, PA, USA (OLL); Landspitali National University Hospital, Department of Radiology, Reykjavik, Iceland (OK); Icelandic Heart Association, Kópavogur, Iceland (VG, AVS); University of Iceland, Faculty of Medicine, Reykjavik, Iceland (VG); Institute of Molecular Medicine and Human Genetics Center; University of Texas Health Science Center at Houston, Houston, TX, USA (MF, XJ); Department of Neurology; Johns Hopkins University, Baltimore, MD, USA (RFG); Center of Biostatistics and Bioinformatics; University of Mississippi Medical Center, Jackson, MS, USA (MEGr); Department of Neurology; Mayo Clinic, Rochester, MN, USA (DSK); Department of Medicine, Division of Geriatrics; University of Mississippi Medical Center, Jackson, MS, USA (THM, BGW); CTSU, Nuffield Department of Population Health, University of Oxford, UK (JCH, RC); Institute for Stroke and Dementia Research, Klinikum der Universität München, Ludwig-Maximilians University, Munich, Germany (MD, RM); Munich Cluster for Systems Neurology (SyNergy), Munich, Germany (MD); Department of Clinical Neurosciences, University of Cambridge, Cambridge, UK (SB); Departments of Neurology and Public Health Sciences, University of Virginia, Charlottesville, Virginia, USA (BBW); Department of Cerebrovascular Diseases, Fondazione IRCCS Istituto Neurologico Carlo Besta, Milano, Italy (GBB); Department of Neurology and Stroke Center, Basel University Hospital, Switzerland (STE); Neurorehabilitation Unit, University and University Center for Medicine of Aging and Rehabilitation Basel, Felix Platter Hospital, Basel, Switzerland (STE); Department of Neurology, Heidelberg University Hospital, Germany (MK); Department of Clinical and Experimental Sciences, Neurology Clinic, University of Brescia, Italy (AP); McGill University and Genome Quebec Innovation Center (ML, TP); Neurovascular Research Group (NEUVAS),

Department of Neurology, IMIM (Institut Hospital del Mar d'Investigacions Mèdiques), Universitat Autonoma de Barcelona, Spain (JJC), Neurovascular Research Laboratory and Neurovascular Unit, Institut de Recerca, Hospital Vall d'Hebron, Universitat Autonoma de Barcelona, Spain (JMo), Center for Human Genetic Research, Massachusetts General Hospital, Boston, MA, USA (FR, JR), Division of Neurocritical Care and Emergency Neurology, Department of Neurology, Massachusetts General Hospital, Boston, MA, USA (FR, JR), J. Philip Kistler Stroke Research Center, Massachusetts General Hospital, Boston (FR, JR), MA, Program in Medical and Population Genetics, Broad Institute, Cambridge, MA (FR, JR), Department of Neurology, Jagiellonian University Medical College, Krakow, Poland (AS), Department of Neurology and Rehabilitation Medicine, University of Cincinnati, OH (DW), John Hunter Hospital, Hunter Medical Research Institute and University of Newcastle (CL), Institute of Cardiovascular Research, Royal Holloway college, University of London (ICR2UL) & Ashford and St Peters Hospital, Surrey (PS),Centre for Clinical Brain Sciences & Institute of Genetics and Molecular Medicine, University of Edinburgh (CLMS), University of Maryland School of Medicine and Baltimore VAMC (JWC), Department of Neurology, Clinical Division of Neurogeriatrics, Medical University Graz, Austria (RS),Department of Neurology, Mayo Clinic Jacksonville, Jacksonville, FL, USA (JMe), University of Leuven, Department of Neurosciences, Experimental Neurology and Leuven Research Institute for Neuroscience and Disease, Leuven, Belgium (VT), Vesalius Research Center, Laboratory of Neurobiology, Leuven, Belgium (VT), University Hospitals Leuven, Belgium (VT), Department of Neurology, Austin Health, 145 Studley Rd, Heidelberg, Victoria 3084, Australia (VT), Department of Clinical Sciences Lund, Neurology, Lund University, Lund, Sweden (AL), Department of Neurology and Rehabilitation Medicine, Skåne University Hospital, Lund, Sweden (AL), Lund University Department of Clinical Sciences Malmö (OM), Neuroscience Institute, Saint Francis Medical Center, School of Health and Medical Sciences, Seton Hall University, South Orange, New Jersey, USA (RPG), Department of Neurology, Miller School of Medicine, University of Miami (RLS), Department of Neurology, Miller School of Medicine, University of Miami (TR), Brigham and Women's Hospital (KMR), Nuffield Department of Clinical Neurosciences, University of Oxford (PMRo), Dean's Office, College of Public Health, University of Kentucky, Kentucky, USA (DKA), Department of Medical and Clinical genetics, Institute of Biomedicine, the Sahlgrenska Academy at University of Gothenburg, Gothenburg, Sweden (CJ), Department of Pharmacotherapy and Translational Research and Center for Pharmacogenomics, University of Florida, College of Pharmacy, Gainesville, FL, USA (JAJ), Division of Cardiovascular Medicine, College of Medicine, University of Florida, Gainesville, FL, USA (JAJ), Department of Neurology, University of British Columbia, Vancouver, British Columbia, Canada (ORB), Clinical Neurosciences, University of Cambridge, Cambridge, UK (HSM, MT), Department of Epidemiology and Population Health, Albert Einstein College of Medicine (SWS), Stroke Center, Department of Neurology, Washington University School of Medicine, St Louis, MO, USA (JML),Department of Medical Genetics, University Medical Center Utrecht, Utrecht, Netherlands (SLP), Ann Romney Center for Neurologic Diseases, Brigham and Women's Hospital, Boston, Massachusetts, USA (PLDJ, CCW), Rush Alzheimer's Disease Center, Rush University Medical Center, 600 S Paulina St, Chicago, IL 60612 (DAB), Hôpital Broca, Service de Gérontologie I, and Université Paris Descartes, Sorbonne Paris Cité, Paris, France (JSV), National Heart, Lung, and Blood Institute's Framingham Heart Study Population Sciences Branch, Framingham, Massachusetts, USA (ADJ), Fondation Jean Dausset, Centre d'Etude du Polymorphisme Humain (CEPH), Paris, France (YK), Center for Integrative Medical Sciences, RIKEN, Yokohama, Japan (YK), Department of Neurology, Boston University School of Medicine, Boston, Massachusetts, USA (AB, ALD, HJA, SS, CLS, SHC, FX), Department of Biostatistics, Boston University School of Public Health, Boston, Massachusetts, USA (AB, ALD, SS, SHC, FX), The National Heart,

Lung, and Blood Institute's Framingham Heart Study, Framingham, Massachusetts, USA (AB, ADJ, ALD, AYC, HJA, SS, CLS, SHC, FX), Department of Ophthalmology, University of Alberta, Edmonton, Alberta, Canada (CRF, OJL), Alberta Children's Hospital Research Institute and Department of Biochemistry and Molecular Biology, University of Calgary, Calgary, Alberta, Canada (CRA, SJC), Department of Medical Genetics, University of Alberta, Edmonton, Alberta, Canada (OJL), Department of Chemistry and Molecular Biology, University of Gothenburg, Gothenburg, Sweden (AR, AMN, PC).

Contributors

S.D., S.S., W.T.L., S.J.C., D.I.C., P.C., M.A.I. and L.J.L. jointly supervised research. G.C., C.R.A., A.Y.C., M.F., A.R., J.C.B. and A.S.H. contributed equally. S.D., S.S., W.T.L., S.J.C., P.C., J.R. and B.B.W. conceived and designed the experiments. G.C., A.Y.C., M.F., J.C.B., A.S.H., M.S., A.V.S., H.H.H.A., S.H.C., S.L.P., S.T., M.E.Ga., A.M., A.T., S.G., T.M.B., Ce.B., X.J., F.X., Y.L., M.d.H., T.P., C.C.W., S.B., R.M., Q.W., M.T. and Y.K. performed statistical analysis and other data analysis. M.A.I., L.J.L., K.M.R., C.T., O.J.L., C.M.v.D., E.I., T.H.M., V.S., J.R., B.B.W., C.O.S., A.L.D., S.S.R., T.B.H., B.M.P., T.K., V.G., M.L., T.P., C.C.W, P.L.D.J., D.A.B., M.D., C.R.F., E.V., H.S.M., A.B., B.G.W., D.S.K., M.E.Gr., R.F.G., M.A.N., J.I.R., O.L.L., Y.K., J.Da., R.C., G.B.B., U.S., M.R.P.M., H.J.G., A.D.J., M.M.S., N.B.A., D.J.S., J.W.J., I.F., A.J.M.d.C., A.G.U., M.L.P., P.J.K., A.Ho., D.W., A.S., F.R., J.Mo., J.J.C., A.P.M., C.M.L., L.Li., L.La., N.L.P., M.T., P.K.M., A.Ha., J.F.D., Cl.B., P.M.Ri., J.E.B., A.P., D.L., M.K., S.T.E., H.J.A., Q.W., J.M.L., S.W.S., O.R.B., J.A.J., C.J., D.K.A., P.M.Ro., T.R., R.L.S., R.P.G., O.M., A.L., V.T., J.Me., R.S., J.W.C., A.M.N., C.L.M.S., P.S., C.L., N.L.S., K.R., S.R.H., R.M., J.C.H., S.B., M.d.H., J.De., Y.L., S.R., F.X., C.L.S., K.L.W., O.K., X.J., J.S.V., Ce.B., T.M.B., S.G., A.T., A.M., M.E.Ga., S.T., S.L.P., S.H.C., H.H.H.A., A.V.S., M.S., A.S.H., J.C.B., A.R., M.F., A.Y.C. and C.R.A. revised the draft critically for important intellectual content. S.J.C., O.J.L., A.M.N., A.R. and C.R.A. performed the experiments. S.D., S.S., W.T.L., S.J.C., D.I.C., P.C. and G.C. wrote the paper.

DECLARATION OF INTERESTS

The authors declare no competing financial interests.

ACKNOWLEDGMENTS

The information about funding for each section is reported in the online appendix (page 63-73)

References

1. Johnston SC, Mendis S, Mathers CD. Global variation in stroke burden and mortality: estimates from monitoring, surveillance, and modelling. Lancet Neurol. 2009; **8**(4): 345-54.
2. Falcone GJ, Malik R, Dichgans M, Rosand J. Current concepts and clinical applications of stroke genetics. Lancet Neurol. 2014; **13**(4): 405-18.
3. Bevan S, Traylor M, Adib-Samii P, Malik R, Paul NL, Jackson C, et al. Genetic heritability of ischemic stroke and the contribution of previously reported candidate gene and genomewide associations. Stroke; a journal of cerebral circulation. 2012; **43**(12): 3161-7.
4. Manolio TA. Bringing genome-wide association findings into clinical use. Nat Rev Genet. 2013; **14**(8): 549-58.
5. Ikram MA, Seshadri S, Bis JC, Fornage M, DeStefano AL, Aulchenko YS, et al. Genomewide association studies of stroke. N Engl J Med. 2009; **360**(17): 1718-28.
6. Traylor M, Farrall M, Holliday EG, Sudlow C, Hopewell JC, Cheng YC, et al. Genetic risk factors for ischaemic stroke and its subtypes (the METASTROKE collaboration): a meta-analysis of genome-wide association studies. Lancet Neurol. 2012; **11**(11): 951-62.
7. Meschia JF, Arnett DK, Ay H, Brown RD, Jr., Benavente OR, Cole JW, et al. Stroke Genetics Network (SiGN) study: design and rationale for a genome-wide association study of ischemic stroke subtypes. Stroke. 2013; **44**(10): 2694-702.
8. Hofer E, Cavalieri M, Bis JC, DeCarli C, Fornage M, Sigurdsson S, et al. White Matter Lesion Progression: Genome-Wide Search for Genetic Influences. Stroke. 2015.
9. Klungel OH, Heckbert SR, Longstreth WT, Jr., Furberg CD, Kaplan RC, Smith NL, et al. Antihypertensive drug therapies and the risk of ischemic stroke. Arch Intern Med. 2001; **161**(1): 37-43.
10. Debette S, Kamatani Y, Metso TM, Kloss M, Chauhan G, Engelter ST, et al. Common variation in PHACTR1 is associated with susceptibility to cervical artery dissection. Nat Genet. 2015; **47**(1): 78-83.
11. Adams HP, Jr., Bendixen BH, Kappelle LJ, Biller J, Love BB, Gordon DL, et al. Classification of subtype of acute ischemic stroke. Definitions for use in a multicenter clinical trial. TOAST. Trial of Org 10172 in Acute Stroke Treatment. Stroke. 1993; **24**(1): 35-41.
12. Verhaaren BF, Debette S, Bis JC, Smith JA, Ikram MK, Adams HH, et al. Multi-Ethnic Genome-Wide Association Study of Cerebral White Matter Hyperintensities on MRI. Circ Cardiovasc Genet. 2015.
13. Woo D, Falcone GJ, Devan WJ, Brown WM, Biffi A, Howard TD, et al. Meta-analysis of Genome-wide Association Studies Identifies 1q22 as a Susceptibility Locus for Intracerebral Hemorrhage. Am J Hum Genet. 2014; **94**(4): 511-21.
14. Reyahi A, Nik AM, Ghiami M, Gritli-Linde A, Ponten F, Johansson BR, et al. Foxf2 Is Required for Brain Pericyte Differentiation and Development and Maintenance of the Blood-Brain Barrier. Dev Cell. 2015; **34**(1): 19-32.
15. Arnold CR, Lamont RE, Walker JT, Spice PJ, Chan CK, Ho CY, et al. Comparative analysis of genes regulated by Dzip1/iguana and hedgehog in zebrafish. Dev Dyn. 2015; **244**(2): 211-23.
16. Gorlov IP, Gorlova OY, Amos CI. Allelic Spectra of Risk SNPs Are Different for Environment/Lifestyle Dependent versus Independent Diseases. PLoS Genet. 2015; **11**(7): e1005371.
17. Maillard P, Seshadri S, Beiser A, Himali JJ, Au R, Fletcher E, et al. Effects of systolic blood pressure on white-matter integrity in young adults in the Framingham Heart Study: a cross-sectional study. Lancet Neurol. 2012; **11**(12): 1039-47.
18. de Bakker PI, Ferreira MA, Jia X, Neale BM, Raychaudhuri S, Voight BF. Practical aspects of imputation-driven meta-analysis of genome-wide association studies. Human molecular genetics. 2008; **17**(R2): R122-8.
19. Ziller MJ, Gu H, Muller F, Donaghey J, Tsai LT, Kohlbacher O, et al. Charting a dynamic DNA methylation landscape of the human genome. Nature. 2013; **500**(7463): 477-81.

20. French CR, Seshadri S, Destefano AL, Fornage M, Arnold CR, Gage PJ, et al. Mutation of FOXC1 and PITX2 induces cerebral small-vessel disease. J Clin Invest. 2014; **124**(11): 4877-81.
21. Orozco G, Barrett JC, Zeggini E. Synthetic associations in the context of genome-wide association scan signals. Hum Mol Genet. 2010; **19**(R2): R137-44.
22. Wardlaw JM, Smith C, Dichgans M. Mechanisms of sporadic cerebral small vessel disease: insights from neuroimaging. Lancet Neurol. 2013; **12**(5): 483-97.
23. Bolte C, Ren X, Tomley T, Ustiyan V, Pradhan A, Hoggatt A, et al. Forkhead box F2 regulation of platelet-derived growth factor and myocardin/serum response factor signaling is essential for intestinal development. J Biol Chem. 2015; **290**(12): 7563-75.
24. Lehmann OJ, Sowden JC, Carlsson P, Jordan T, Bhattacharya SS. Fox's in development and disease. Trends Genet. 2003; **19**(6): 339-44.
25. Carlsson P, Mahlapuu M. Forkhead transcription factors: key players in development and metabolism. Dev Biol. 2002; **250**(1): 1-23.
26. Hong HK, Noveroske JK, Headon DJ, Liu T, Sy MS, Justice MJ, et al. The winged helix/forkhead transcription factor Foxq1 regulates differentiation of hair in satin mice. Genesis. 2001; **29**(4): 163-71.
27. Verzi MP, Khan AH, Ito S, Shivdasani RA. Transcription factor foxq1 controls mucin gene expression and granule content in mouse stomach surface mucous cells. Gastroenterology. 2008; **135**(2): 591-600.
28. Siegenthaler JA, Choe Y, Patterson KP, Hsieh I, Li D, Jaminet SC, et al. Foxc1 is required by pericytes during fetal brain angiogenesis. Biol Open. 2013; **2**(7): 647-59.
29. Fish JE, Wythe JD. The molecular regulation of arteriovenous specification and maintenance. Dev Dyn. 2015; **244**(3): 391-409.
30. Fornage M, Debette S, Bis JC, Schmidt H, Ikram MA, Dufouil C, et al. Genome-wide association studies of cerebral white matter lesion burden: the CHARGE consortium. Ann Neurol. 2011; **69**(6): 928-39.
31. French CR, Seshadri S, Destefano AL, Fornage M, Arnold CR, Gage PJ, et al. Mutation of FOXC1 and PITX2 induces cerebral small-vessel disease. J Clin Invest. 2014.
32. Bellenguez C, Bevan S, Gschwendtner A, Spencer CC, Burgess AI, Pirinen M, et al. Genome-wide association study identifies a variant in HDAC9 associated with large vessel ischemic stroke. Nat Genet. 2012; **44**(3): 328-33.
33. Gudbjartsson DF, Arnar DO, Helgadottir A, Gretarsdottir S, Holm H, Sigurdsson A, et al. Variants conferring risk of atrial fibrillation on chromosome 4q25. Nature. 2007; **448**(7151): 353-7.
34. Greenberg SM. Small vessels, big problems. N Engl J Med. 2006; **354**(14): 1451-3.
35. Zhou Q, Yang D, Ombrello AK, Zavialov AV, Toro C, Stone DL, et al. Early-onset stroke and vasculopathy associated with mutations in ADA2. N Engl J Med. 2014; **370**(10): 911-20.
36. Sibon I, Coupry I, Menegon P, Bouchet JP, Gorry P, Burgelin I, et al. COL4A1 mutation in Axenfeld-Rieger anomaly with leukoencephalopathy and stroke. Ann Neurol. 2007; **62**(2): 177-84.
37. Lanfranconi S, Markus HS. COL4A1 mutations as a monogenic cause of cerebral small vessel disease: a systematic review. Stroke. 2010; **41**(8): e513-8.
38. Kubo M, Hata J, Ninomiya T, Matsuda K, Yonemoto K, Nakano T, et al. A nonsynonymous SNP in PRKCH (protein kinase C eta) increases the risk of cerebral infarction. Nat Genet. 2007; **39**(2): 212-7.

Tables and Figures

Table 1: Population characteristics

Studies	Country	Number of participants					Percentage of Women	Mean age of controls (±SD)	Mean age of stroke cases (±SD)	Mean years of follow up
		Total	All stroke	IS	CE-IS	Non-CE-IS †				
Discovery (prospective longitudinal population-based cohort studies)										
AGES	USA	2,996	114	99	n.a.	n.a.	58%	79·9 (5·5)	76·4 (5·5)	3·6 (1·1)
ARIC	USA	8,939	473	416	108	305	53%	69·3 (7·4)	54·2 (5·7)	19·1 (4·6)
CHS	USA	3,268	563	447	139	308	61%	82·7 (6·2)	72·3 (5·4)	13·5 (6·3)
FHS	USA	4,369	235	198	57	127	55%	73·4 (11·7)	65·5 (12·7)	8·0 (3·2)
FINRISK CoreExome	Finland	5,202	94	60	n.a.	n.a.	55%	69·9 (9·4)	45·8 (12·8)	13·3 (2·7)
FINRISK Corogene	Finland	1,887	60	46	n.a.	n.a.	49%	73·7 (9·5)	55·2 (12·2)	9·0 (4·0)
FINRISK PredictCVD	Finland	1,309	352	294	n.a.	n.a.	53%	66·5 (9·8)	46·5 (13·3)	10·0 (5·2)
Health-ABC	USA	1,661	124	n.a.	n.a.	n.a.	47%	80·1 (4·2)	73·8 (2·8)	9·3 (2·9)
MESA	USA	2,364	49	43	n.a.	35	52%	75·1 (8·9)	62·7 (10·2)	7·2 (1·4)
PROSPER	Netherlands	4,658	193	n.a.	n.a.	n.a.	53%	77·8 (3·6)	75·2 (3·3)	3·1 (0·7)
Rotterdam Study I	Netherlands	6,066	821	448	95	353	60%	80·6 (8·1)	69·2 (9·0)	13·0 (6·2)
Rotterdam Study II	Netherlands	2,080	125	88	17	71	54%	75·9 (9·2)	64·6 (7·9)	9·7 (2·5)
SHIP	Germany	3,112	75	37	n.a.	n.a.	52%	71·8 (10·6)	48·7 (15·2)	12·1 (2·5)
TWINGENE	Sweden	6,702	116	95	n.a.	n.a.	52%	75·6 (8·8)	64·9 (8·1)	3·2 (1·0)
ULSAM	Sweden	1,139	216	171	56	115	0%	79·9 (4·8)	71·0 (0·6)	12·8 (5·2)
WGHS	USA	23,294	499	402	82	320	100%	69·6 (9·3)	54·7 (7·1)	16·0 (3·2)
3C-Dijon	France	3,762	157	125	33	92	62%	81·5 (6·1)	72·4 (5·6)	8·7 (3·1)
3C-Bordeaux-	France	2,153	82	59	15	44	60%	81·7 (5·2)	73·9 (5·1)	7·8 (2·5)
TOTAL		*84,961*	*4,348*	*3,028*	*602*	*1,770*	*67%*	*75·8 (8·0)*	*63·7 (8·4)*	*10·0 (3·6)*
Follow-up (cross-sectional case-control studies) *										
SiGN	USA + Europe	49,324	16,851	16,851	3,427	2,346 3,150	46%	n.a.‡	66·5 (14·8)	n.a.
METASTROKE	USA + Europe	9,654	1,729	1,729	276	206 159	36%	60.6 (11.9)	67.0 (10.1)	n.a.
HVH1	USA	2,012	681	577	92	62 175	57%	66.7 (9.1)	68.8 (8.9)	n.a.
CADISP	Europe	9,814	555	555	211	67 31	61%	n.a.‡	43·7 (9·9)	n.a.
TOTAL		*70,804*	*19,816*	*19,712*	*4,006*	*2,681* *3,515*	*47%*	*n.a.*	*n.a.*	*n.a.*

N= number of participants; n.a.= not available; IS= Ischemic stroke; CE-IS= Cardioembolic ischemic stroke; SD= standard deviation; * More detailed descriptions of the composition of replication studies can be found in the appendix; † follow-up samples are from large artery ischemic stroke (first line) and small artery ischemic stroke (second line) for SiGN, METASTROKE, HVH1, and CADISP (TOAST subtyping) ; ǂ mean age of controls in the SiGN and CADISP study is not available as they were obtained from anonymous genotype databases

Table 2: Main association results ($p<5\times10^{-6}$) for all stroke, ischemic stroke, cardioembolic ischemic stroke, and non cardioembolic ischemic stroke

Marker Name	chr:position[ł]	Function	Gene	# var[ǂ]	MA	MAF	HR	P-value [†]	Direction[‡]	HetISq[‖]	HetPVal[§]	Imputation quality[*]
All stroke												
rs6433905	2:182138150	intergenic	*UBE2E3*	2	C	0·08	1·21 (1·12-1·31)	$2·54\times10^{-6}$	+++++++++++++++	0	0·54	0·97
rs12204590	6:1337393	intergenic	*FOXF2*	6	A	0·21	1·14 (1·08-1·20)	$2·17\times10^{-6}$	++++++++++++-+-++	0	0·60	1·00
rs790919	6:154298875	intergenic	*OPRM1*	2	A	0·44	1·12 (1·07-1·17)	$2·44\times10^{-6}$	+++++++++++++--++	16·2	0·26	0·96
rs11788316	9:13445687	intergenic	*MPDZ*	4	T	0·28	1·13 (1·07-1·19)	$2·49\times10^{-6}$	++-++++--++++++--+	0	0·84	0·96
rs11627959	14:35160471	intergenic	*CFL2*	4	A	0·44	0·89 (0·85-0·93)	$2·23\times10^{-6}$	--+------+--+----	0	0·87	0·93
rs4899120	14:64335447	intronic	*SYNE2*	1	T	0·09	1·19 (1·11-1·29)	$4·71\times10^{-6}$	+++-++++++++--++	24	0·18	0·98
Ischemic stroke												
rs62262077	3:105014929	intergenic	*ALCAM*	5	A	0·27	1·17 (1·10-1·24)	$6·04\times10^{-7}$	-++++++++++-+++	19·2	0·23	0·94
rs10037362	5:31110857	intergenic	*CDH6*	2	A	0·07	1·27 (1·15-1·40)	$4·41\times10^{-6}$	-+-++++++++	25·7	0·20	0·97
rs4448595	10:21666138	intergenic	*C10orf114*	8	G	0·16	0·83 (0·77-0·90)	$2·50\times10^{-6}$	--------+-------	0	0·80	0·98
rs11833579	12:775199	intergenic	*NINJ2*	2	A	0·24	1·19 (1·12-1·27)	$5·74\times10^{-8}$	-+++-++++++-++++	18·5	0·24	0·92
rs77858481	13:81142325	intergenic	*SPRY2*	1	G	0·06	1·38 (1·22-1·55)	$2·32\times10^{-7}$	++++++++	0	0·97	0·83
Cardioembolic ischemic stroke												
rs4284256	1:157675273	intergenic	*FCRL3*	1	T	0·18	1·41 (1·22-1·64)	$3·13\times10^{-6}$	-+++++	66·5	0·01	0·96
rs12646447	4:111699326	intergenic	*PITX2*	102	C	0·12	1·53 (1·31-1·80)	$1·92\times10^{-7}$	++++++	0	0·44	0·99
rs72184	5:123754837	intergenic	*ZNF608*	1	G	0·43	1·30 (1·17-1·46)	$2·29\times10^{-6}$	+++++++++	9·1	0·6	0·90
rs72794386	5:127479278	intronic	*SLC12A2*	22	T	0·10	1·67 (1·39-2·00)	$4·37\times10^{-8}$	+++++	0	0·87	0·97
rs1428155	5:151281633	intronic	*GLRA1*	2	C	0·38	1·28 (1·161·43)	$3·10\times10^{-6}$	+++++++++	36·2	0·13	1·00
rs7771564	6:22504092	intergenic	*HDGFL1*	4	G	0·10	1·53 (1·28-1·82)	$2·10\times10^{-6}$	++++++	0	0·67	0·99
rs1495081	8:15314955	intergenic	*TUSC3*	1	C	0·14	1·48 (1·25-1·74)	$3·09\times10^{-6}$	++-+++	49·9	0·08	0·88
rs2393938	10:44063812	UTR5	*ZNF239*	1	C	0·12	1·45 (1·24-1·70)	$3·47\times10^{-6}$	++++++	0	0·51	0·99
rs11021485	11:95968208	intronic	*MAML2*	1	A	0·12	1·60 (1·32-1·94)	$1·24\times10^{-6}$	+++++	30·5	0·22	0·82
rs710009	14:59184500	intergenic	*DACT1*	4	G	0·16	1·41 (1·22-1·64)	$3·62\times10^{-6}$	+++++++	0	0·84	0·98
Non-cardioembolic ischemic stroke												
rs77744591	13:81142325	intergenic	*SPRY2*	1	T	0·08	1·34 (1·18-1·51)	$3·44\times10^{-6}$	+++++-+	0	0·97	0·93

Only associations with the lead SNP in each locus are shown in this table, the full set of genetic associations at $p < 5\times10^{-6}$ is presented in supplementary table 7; All results are presented with respect to the minor allele as coded allele; MA= Minor allele; MAF= Minor allele frequency; HR= Hazards ratio; * mean value of imputation quality across studies; ł Chromosome positions with respect to NCBI built 37; ǂ number of variants reaching $p < 5\times10^{-6}$ in the locus; † p-value after genomic control; ‡ Direction refers to direction of effect size with respect to the minor allele; ‖ Heterogeneity I^2, ranges between 0 to 100, higher values suggesting more heterogeneity; § P-value for heterogeneity, calculated using the Cochran's Q test

Table 3: Follow-up of top loci in independent cross-sectional case-control association studies

		Discovery		Follow-up										Combined all	
		CHARGE		SiGN		METASTROKE		HVH1		CADISP		Follow-up Meta-analysis		Meta-analysis	
Marker Name	Gene	HR (95%CI)	P	OR (95%CI)	P	OR (95%CI)	P	OR (95%CI)	P	OR (95%CI)	P	OR (95%CI)	P	OR (95%CI)	P
*All stroke**															
rs6433905	*UBE2E3*	1·21 (1·12-1·31)	$2{\cdot}54\times10^{-6}$	1·02 (0·97-1·08)	0·41	1·11 (0·98-1·27)	0·11	1·09 (0·83-1·44)	0·54	0·72 (0·54-0·95)	0·018	1·03 (0·98-1·07)	0·22	1·07 (1·03-1·12)	$8{\cdot}72\times10^{-4}$
rs12204590	***FOXF2***	**1·14 (1·08-1·20)**	$\mathbf{2{\cdot}17\times10^{-6}}$	**1·07 (1·03-1·11)**	$\mathbf{1{\cdot}02\times10^{-3}}$	**1·07 (0·98-1·16)**	**0·13**	**1·03 (0·86-1·24)**	**0·73**	**1·08 (0·92-1·26)**	**0·36**	**1·06 (1·03-1·09)**	$\mathbf{2{\cdot}15\times10^{-4}}$	**1·08 (1·05-1·12)**	$\mathbf{1{\cdot}48\times10^{-8}}$
rs790919	*OPRM1*	1·12 (1·07-1·17)	$2{\cdot}44\times10^{-6}$	1·00 (0·97-1·03)	0·88	1·01 (0·95-1·09)	0·70	0·88 (0·76-1·02)	0·10	1·06 (0·93-1·21)	0·37	1·00 (0·98-1·03)	0·80	1·03 (1·01-1·05)	0·013
rs11788316	*FLJ41200*	1·13 (1·07-1·19)	$2{\cdot}49\times10^{-6}$	0·99 (0·96-1·03)	0·73	1·07 (0·99-1·16)	0·074	1·01 (0·84-1·21)	0·93	0·99 (0·85-1·15)	0·89	1·01 (0·99-1·04)	0·38	1·03 (1·01-1·06)	$7{\cdot}53\times10^{-3}$
rs11627959	*CFL2*	0·89 (0·85-0·93)	$2{\cdot}23\times10^{-6}$	0·99 (0·96-1·02)	0·70	1·00 (0·93-1·08)	0·93	1·00 (0·85-1·17)	0·96	1·03 (0·90-1·18)	0·62	1·01 (0·98-1·03)	0·53	0·97 (0·95-0·99)	0·011
rs4899120	*SYNE2*	1·19 (1·11-1·29)	$4{\cdot}71\times10^{-6}$	1·02 (0·97-1·07)	0·50	1·00 (0·88-1·14)	0·98	1·05 (0·80-1·37)	0·73	1·16 (0·92-1·46)	0·21	1·02 (0·99-1·07)	0·23	1·06 (1·02-1·10)	$2{\cdot}00\times10^{-3}$
Ischemic stroke†															
rs62262077	*ALCAM*	1·17 (1·10-1·24)	$6{\cdot}04\times10^{-7}$	0·99 (0·96-1·02)	0·52	1·03 (0·95-1·12)	0·46	1·03 (0·84-1·26)	0·78	1·02 (0·88-1·18)	0·82	1·01 (0·98-1·03)	0·63	1·03 (1·01-1·06)	0·015
rs10037362	*CDH6*	1·27 (1·15-1·40)	$4{\cdot}41\times10^{-6}$	0·98 (0·93-1·03)	0·40	0·99 (0·87-1·12)	0·84	0·93 (0·69-1·26)	0·65	0·86 (0·66-1·13)	0·27	0·97 (0·93-1·02)	0·22	1·01 (0·97-1·05)	0·55
rs4448595	*NEBL-AS1*	0·83 (0·77-0·90)	$2{\cdot}50\times10^{-6}$	1·01 (0·97-1·05)	0·72	1·02 (0·93-1·12)	0·69	0·89 (0·72-1·10)	0·28	0·96 (0·81-1·14)	0·65	1·00 (0·97-1·03)	0·97	0·98 (0·95-1·00)	0·09
rs11833579	*NINJ2*	1·19 (1·12-1·27)	$5{\cdot}74\times10^{-8}$	0·98 (0·95-1·01)	0·21	0·96 (0·88-1·04)	0·28	1·04 (0·85-1·29)	0·69	0·96 (0·82-1·12)	0·57	0·98 (0·95-1·01)	0·14	1·01 (0·98-1·03)	0·47
rs77858481	*SPRY2*	1·38 (1·22-1·55)	2·32×10-7	0·99 (0·93-1·06)	0·75	0·90 (0·77-1·05)	0·17	1·02 (0·72-1·44)	0·93	0·91 (0·69-1·20)	0·48	0·98 (0·93-1·03)	0·50	1·03 (0·99-1·08)	0·17
Cardioembolic ischemic stroke‡															
rs4284256	*FCRL3*	1·41 (1·22-1·64)	$3{\cdot}13\times10^{-6}$	0·96 (0·90-1·04)	0·31	0·93 (0·78-1·12)	0·45	1·03 (0·69-1·55)	0·88	0·98 (0·75-1·28)	0·90	0·96 (0·90-1·01)	0·13	1·02 (0·97-1·09)	0·41
rs12646447	***PITX2***	**1·53 (1·31-1·80)**	$\mathbf{1{\cdot}92\times10^{-7}}$	**1·39 (1·29-1·50)**	$\mathbf{3{\cdot}15\times10^{-18}}$	**1·17 (0·98-1·41)**	**0·083**	**1·62 (1·09-2·42)**	**0·018**	**1·04 (0·78-1·37)**	**0·80**	**1·36 (1·28-1·44)**	**1·89E-23**	**1·37 (1·29-1·46)**	$\mathbf{4{\cdot}72\times10^{-24}}$
rs72184	*ZNF608*	1·30 (1·17-1·46)	$2{\cdot}29\times10^{-6}$	1·02 (0·96-1·08)	0·53	0·99 (0·87-1·13)	0·90	0·92 (0·67-1·28)	0·63	1·03 (0·85-1·26)	0·74	1·02 (0·97-1·06)	0·49	1·06 (1·01-1·10)	0·017
rs72794386	*SLC12A2*	1·67 (1·39-2·00)	$4{\cdot}37\times10^{-8}$	0·97 (0·89-1·06)	0·51	1·09 (0·87-1·37)	0·46	0·95 (0·56-1·62)	0·85	0·78 (0·54-1·14)	0·18	0·96 (0·90-1·03)	0·27	1·06 (0·99-1·14)	0·12
rs1428155	*GLRA1*	1·28 (1·16-1·43)	$3{\cdot}10\times10^{-6}$	0·97 (0·92-1·03)	0·33	0·95 (0·83-1·08)	0·44	0·90 (0·66-1·24)	0·51	0·99 (0·81-1·21)	0·94	0·99 (0·95-1·03)	0·64	1·02 (0·98-1·07)	0·38
rs7771564	*HDGFL1*	1·53 (1·28-1·82)	$2{\cdot}10\times10^{-6}$	1·01 (0·92-1·10)	0·87	0·88 (0·70-1·10)	0·25	1·23 (0·77-1·96)	0·39	1·20 (0·89-1·62)	0·24	1·00 (0·93-1·07)	0·97	1·08 (1·01-1·17)	0·031
rs1495081	*TUSC3*	1·48 (1·25-1·74)	$3{\cdot}09\times10^{-6}$	1·05 (0·98-1·14)	0·18	0·89 (0·72-1·09)	0·25	1·35 (0·84-2·16)	0·21	1·05 (0·79-1·40)	0·73	1·04 (0·98-1·10)	0·24	1·10 (1·03-1·17)	$5{\cdot}07\times10^{-3}$
rs2393938	*ZNF239*	1·45 (1·24-1·70)	$3{\cdot}47\times10^{-6}$	1·02 (0·94-1·10)	0·68	1·01 (0·84-1·22)	0·88	0·95 (0·60-1·52)	0·84	0·96 (0·71-1·29)	0·76	1·00 (0·94-1·07)	0·95	1·07 (1·01-1·15)	0·029
rs11021485	*MAML2*	1·60 (1·32-1·94)	$1{\cdot}24\times10^{-6}$	0·94 (0·86-1·03)	0·17	0·77 (0·63-0·95)	0·015	0·51 (0·25-1·04)	0·063	1·06 (0·78-1·43)	0·73	0·92 (0·86-0·99)	0·017	0·99 (0·92-1·07)	0·82
rs710009	*DACT1*	1·41 (1·22-1·64)	$3{\cdot}62\times10^{-6}$	1·00 (0·92-1·07)	0·93	0·96 (0·80-1·15)	0·68	1·21 (0·80-1·83)	0·37	1·10 (0·84-1·43)	0·50	1·00 (0·94-1·06)	0·96	1·06 (1·00-1·13)	0·048
Non-cardioembolic ischemic stroke§															
rs77744591	*SPRY2*	1·34 (1·18-1·51)	$3{\cdot}44\times10^{-6}$	1·08 (0·96-1·21)	0·19	1·08 (0·84-1·40)	0·53	1·39 (0·74-2·59)	0·30	0·91 (0·47-1·77)	0·78	1·08 (0·99-1·18)	0·80	1·18 (1·09-1·28)	$3{\cdot}22\times10^{-5}$
rs77744591	*SPRY2*	1·34 (1·18-1·51)	$3{\cdot}44\times10^{-6}$	1·12 (1·02-1·24)	0·023	1·06 (0·80-1·40)	0·68	0·83 (0·52-1·31)	0·42	0·73 (0·24-2·22)	0·56	1·11 (1·02-1·20)	0·16	1·18 (1·10-1·27)	$1{\cdot}08\times10^{-5}$

In bold are loci which reach genome wide significance ($P<5\times10^{-8}$) in the combined meta-analysis; HR= Hazards ratio; OR= Odds ratio; CI= Confidence interval; *Follow-up results are from association analyses of ischemic stroke for SiGN, METASTROKE, and CADISP, and of all stroke for HVH1; †follow-up results are from association analyses of ischemic stroke for SiGN, METASTROKE, HVH1, and CADISP; ‡follow-up results are from association analyses of cardioembolic ischemic stroke for SiGN, METASTROKE, HVH1, and CADISP (TOAST subtyping);

§follow-up results are from association analyses of large artery ischemic stroke (first line) and small artery ischemic stroke (second line) for SiGN, METASTROKE, HVH1, and CADISP (TOAST subtyping)

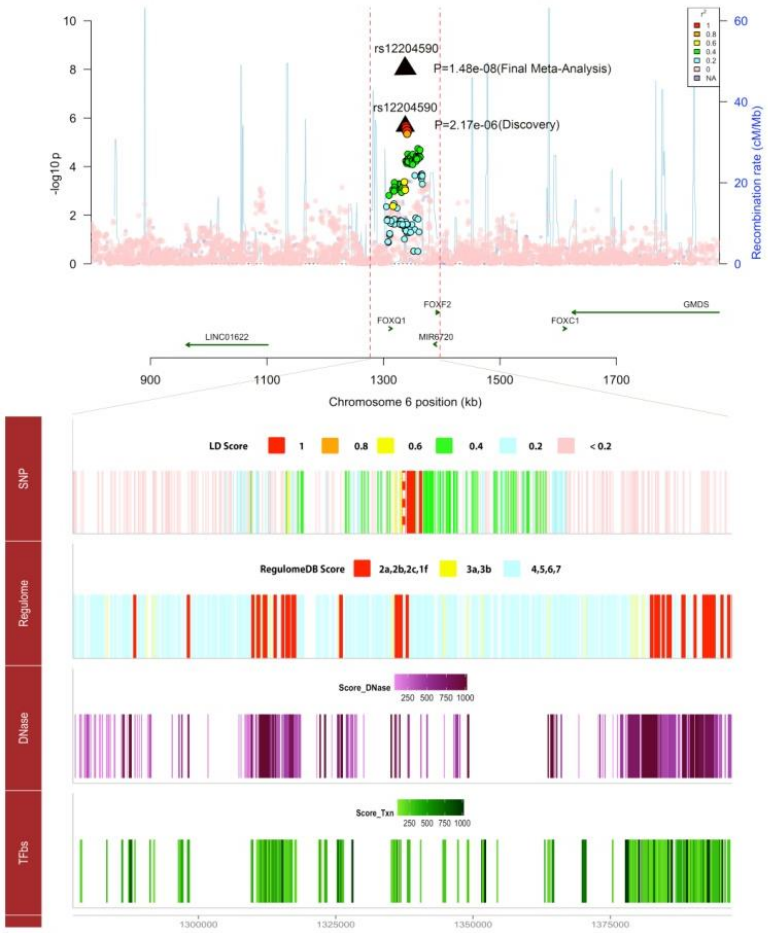

Figure 1: Regional association plot centered around rs12204590 in association with incident all stroke in the population-based discovery GWAS

All SNPs in the discovery stage (circles) are plotted with the negative log of their p-values against their genomic positions. The final meta-analysis P-value of rs12204590 is also plotted. The color of the circles represents the linkage disequilibrium between SNPs. The blue peaks represent estimated recombination rates. Genes are shown as green arrows with direction of arrow representing direction of transcription. Tracks in the bottom were added using UCSC genome browser and the RegulomeDB database: "SNP": SNP track showing the SNPs encompassing the selected region, red dotted line in the SNP track shows the position of top SNP (rs12204590); "Regulome": shows RegulomeDB scores, variants with lower scores having higher probability to act as regulatory variants (http://regulomedb.org/); "DNase": shows DNase hypersensitive regions assayed in a large collection of cell types (125 cell types), ENCODE project, Release 3 (2014); "TFbs": shows regions where transcription factors, proteins responsible for modulating gene transcription, bind to DNA as assayed by ChIP-seq assay, ENCODE project Release 3 (August 2013) https://www.encodeproject.org/

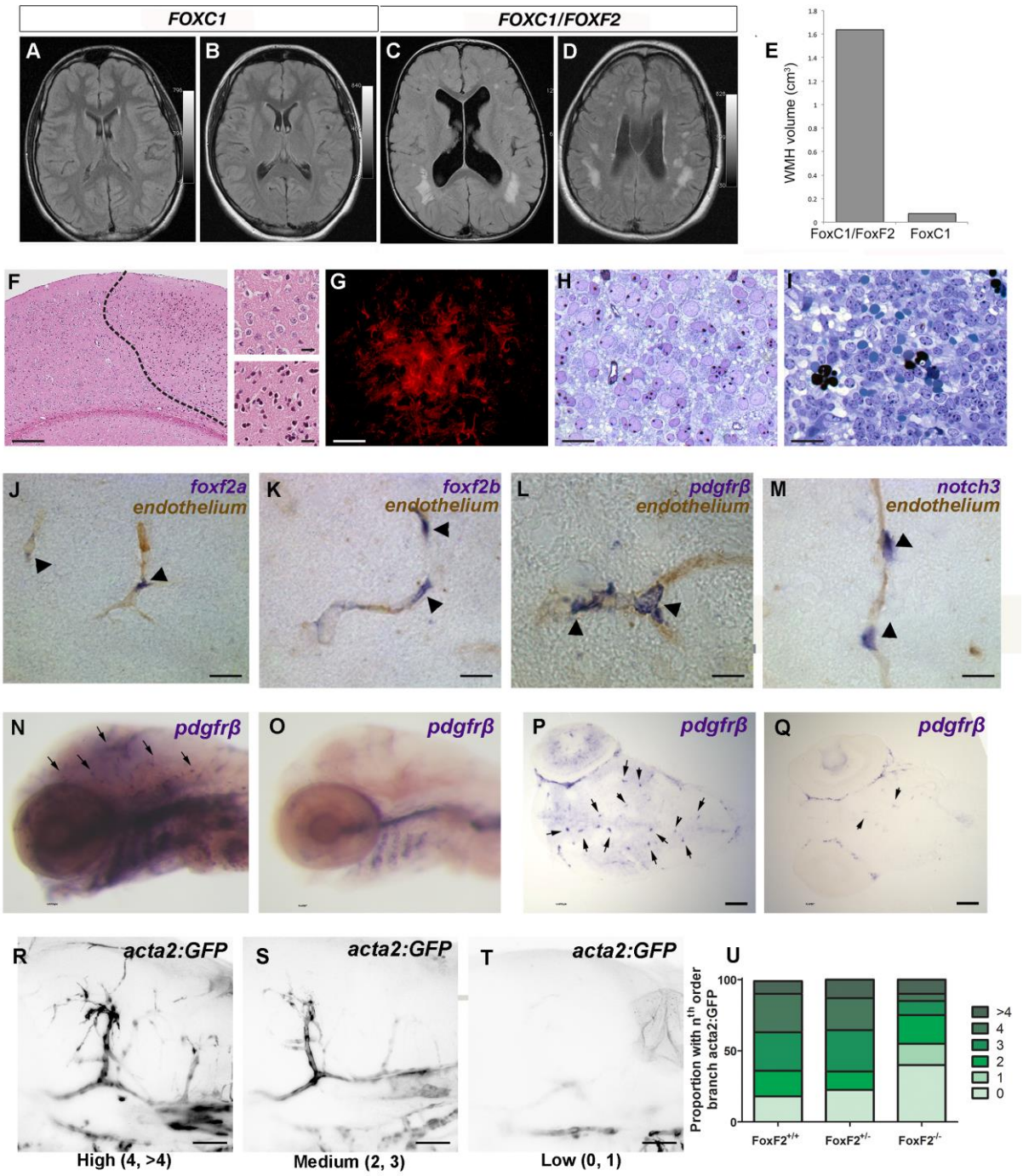

Figure 2: Cerebrovascular phenotype of *FOXF2* deletions in humans, *foxf2* expression in brain pericytes in zebrafish and loss of *Foxf2* leading to hallmarks of cerebrovascular disease in zebrafish and mice

(A-D) In patients with a segmental deletion encompassing *FOXC1* (n=2), white matter hyperintensities (WMHs) are observed in the periventricular region (A-B) and subcortical regions (B). In patients with a segmental deletion of both *FOXC1* and *FOXF2* (C,D; n=2), the mean WMH volume is increased, by more than tenfold (E), in both the subcortical and periventricular regions (see Supplementary table 18 for WMH volumes in each of the four patients). (F-I) Cerebral cortex of conditional *Foxf2* knockout mouse showing ischemic

infarction and hemorrhagic tissue. (F) An area with condensed eosinophilic cytoplasm and pyknotic nuclei (to the right of the dashed line) which indicates a recent ischemic infarction. (F') normal tissue and (F'') tissue with ischemic infarction at higher magnification. (G) GFAP immunofluorescence of an area with reactive astrogliosis in the cerebral cortex of *Foxf2* conditional knockout mouse. (H) Cerebral cortex from control mouse showing normal neuronal tissue and intact capillaries. (I) A hemorrhagic area of the cerebral cortex from a Foxf2 conditional knockout mouse. Extravascular erythrocytes are seen both as intact cells (homogenous greenish blue) and lysed cells (black). (J-M) RNA in situ hybridization (purple) of larval zebrafish brains shows expression of *foxf2a* (J) and *foxf2b* (K) with identical morphology and perivascular location as pericyte markers *pdgfrβ* (L) and *notch3* (M) (purple) around capillaries in 1 month old larval zebrafish (brown). (N-Q) *foxf2*$^{-/-}$ mutants have reduced expression of the pericyte marker *pdgfrβ* in 4 day postfertilization embryonic cerebral vasculature. (R-U) Loss of *foxf2b* results in a reduction of the smooth muscle marker *acta2:GFP* coverage of blood vessels in wildtype (n=11), *foxf2*$^{+/-}$ (n=31) and FoxF2$^{-/-}$ (n=20) embryonic cerebral vasculature. (R-T) show examples of high, medium, and low branch order coverage that were scored from 0 to the 4th order branch in vessel coverage presented as percentages of total embryo counts (U). Scale bars: A-D, G, P-T=50µm; F = 200 µm; F', F'', H and I = 20 µm; J-M = 10µm