Genetics of common cerebral small vessel disease

Constance Bordes[1], Muralidharan Sargurupremraj[1,2], Aniket Mishra[1] and Stéphanie Debette[1,3†]

[1]Bordeaux Population Health Research Center, Inserm U1219, University of Bordeaux, Bordeaux, France

[2]Glenn Biggs Institute for Alzheimer's & Neurodegenerative Diseases, University of Texas Health Sciences Center, San Antonio, Texas, USA

[3]Department of Neurology, Bordeaux University Hospital, Institute for Neurodegenerative Diseases, Bordeaux, France

[†]email: stephanie.debette@u-bordeaux.fr

Abstract | Cerebral small vessel disease (cSVD) is a leading cause of ischaemic and haemorrhagic stroke and a major contributor to dementia. Covert cSVD, which is detectable with brain MRI but not manifest as clinical stroke, is highly prevalent in the general population with increasing age. Advances in technologies and collaborative work have led to substantial progress in the identification of common genetic variants that are associated with cSVD-related stroke (ischaemic and haemorrhagic) and MRI-defined covert cSVD. In this Review, we provide an overview of collaborative studies — mostly genome-wide association studies (GWAS) — that have identified >50 independent genetic loci associated with the risk of cSVD. We describe how these associations have provided novel insight into the biological mechanisms involved in cSVD, revealed patterns of shared genetic variation across cSVD traits, and shed new light on the continuum between rare, monogenic and common, multifactorial cSVD. We consider how GWAS summary statistics have been leveraged for Mendelian randomization studies to explore causal pathways in cSVD and provide genetic evidence for drug effects, and how the combination of findings from GWAS with gene expression resources and drug target databases has enabled identification of putative causal genes and provided proof-of-concept for drug repositioning potential. We also discuss opportunities for polygenic risk prediction, multiancestry approaches and integration with other omics data.

[H1] Introduction

As average lifespans have increased, stroke, cognitive impairment and dementia have become the most common age-related sources of disability and dependence and have a huge societal burden. One in three people who are healthy at 65 years of age will develop stroke, dementia or both during their remaining lifetime[1]. Cerebral small vessel disease (cSVD), which encompasses a group of pathological processes that affect small penetrating vessels (arterioles, capillaries and venules) of the brain[2,3], makes a major contribution to the occurrence of stroke and dementia. cSVD is a leading cause of ischaemic and haemorrhagic stroke, accounting for up to one third of strokes worldwide[4-6]. It is also the main pathological substrate of the vascular contribution to cognitive decline and dementia[7] - the vascular component of dementia is now recognized as a major driver of the pathophysiology[8,9].

Most often, vascular brain injury is covert (not associated with clinical stroke) and caused by underlying cSVD[10-12]. Covert cSVD is highly prevalent in elderly people[3] and encompasses several entities that are defined by brain MRI abnormalities[13]. The most studied of these abnormalities are white matter hyperintensities (WMHs) of presumed vascular origin (sometimes referred to as leukoaraiosis)[14], lacunes of presumed vascular origin, and cerebral microbleeds[15]. In the past 5 to 10 years, novel markers of covert cSVD have emerged that require more elaborate imaging acquisition and/or processing; these markers included perivascular spaces[16,17], microinfarcts[18,19] and subtle changes of white matter microstructure detectable with diffusion tensor imaging (DTI)[20-22]. Covert cSVD is a powerful determinant of stroke risk, unfavourable stroke outcome, cognitive decline and dementia[14,23,24], and therefore represents a major target for the prevention of these conditions and promotion of healthier brain ageing.

The development of cSVD is influenced by a complex mix of genetic and environmental risk factors, among which age and hypertension are currently deemed the most important[4]. However, the cascade of events and molecular pathways that lead to the alterations in the structure and function of small brain vessels that characterize cSVD are largely unknown. Molecular biomarkers for clinical use have not been identified and, besides risk factor (especially blood pressure) management, no mechanism-based treatments are available to prevent cSVD progression and complications. Determining the genetic underpinnings of common, multifactorial cSVD promises to reveal molecular mechanisms

that underlie cSVD and thereby accelerate the discovery of novel biomarkers, drug targets or drug repositioning opportunities[25,26]. Exploring the genomics of common cSVD also provides tools to indirectly examine the effects of known or novel drugs and test causal relationships of putative risk factors with cSVD. Furthermore, better knowledge of the genetic contribution to cSVD could help with identification and stratification of individuals who are at high risk of cSVD and could benefit from tailored preventive interventions (**Fig. 1**).

In the past 10 years, large, international collaborative work based on high-throughput genotyping has driven major progress in the identification of common genetic risk variants that underlie cSVD. In this Review, we provide a detailed overview of the current knowledge on genetic variants associated with common cSVD, encompassing cSVD-related stroke (ischaemic and haemorrhagic) and MRI-defined covert cSVD. We will not discuss single-gene disorders that cause cSVD, which have been reviewed in detail elsewhere[27-29]. We first focus on the discovery of genetic variants associated with cSVD-related stroke and MRI-defined covert cSVD before considering the shared genetic variation between cSVD phenotypes and vascular risk factors, and exploring the continuum between monogenic and multifactorial forms of cSVD. Finally, we discuss the rapidly accumulating Mendelian randomization studies of cSVD and reflect on the clinical and therapeutic implications of common cSVD genomics research.

[H1] Discovery of genetic risk loci

In this section, we summarize the discovery of genetic risk loci for common cSVD, focusing on robust results derived from large datasets and state-of-the-art methodology, mostly genome-wide association studies (GWAS, **Box 1**). We also highlight pilot studies of exome chip and next-generation sequencing approaches to explore rare genetic risk variants for cSVD.

[H2] cSVD-related stroke

cSVD-related stroke encompasses several distinct but correlated entities: small vessel stroke (SVS; also called lacunar stroke), which results from small subcortical brain infarcts with or without associated moderate to severe confluent leukoaraiosis[30]; lobar intracerebral haemorrhage (ICH), which originates at the cortex or cortico-subcortical junction and is

commonly associated with cerebral amyloid angiopathy; and non-lobar ICH, which originates in deep structures of the cerebral hemispheres, the brainstem and the cerebellum, and is typically associated with hypertensive vasculopathy[31]. Characterization of the genetic determinants of cSVD-related stroke has been challenging owing to the clinical heterogeneity of the condition and the small size of well-characterized datasets.

Work by large international consortia, such as the International Stroke Genetics Consortium (ISGC), has led to substantial progress in the discovery of genes that underlie cSVD-related stroke. To date, GWAS have identified two risk loci for the occurrence of ICH (one for lobar ICH and one for non-lobar ICH), two risk loci associated with ICH volume (one for lobar ICH and one for non-lobar ICH)[31-33] and seven risk loci for SVS[34-38] at the genome wide significance level ($P < 5 \times 10^{-8}$). The majority of these loci have been identified in populations of European ancestry (**Table 1**).

The discovery of risk loci for SVS has been substantially improved not only by an increase in sample size but also by MRI confirmation of SVS, which was shown to be more reliable than standard phenotyping[34]. MRI-confirmed SVS is defined as a clinical lacunar syndrome with an anatomically compatible infarct on MRI (subcortical and ≤15 mm in diameter)[39], either as a region of high intensity on diffusion-weighted imaging (acute) or as a region of low intensity on fluid-attenuated inversion recovery (FLAIR) or T1 imaging (non-acute)[13], and the absence of causes of stroke other than cSVD[34].

Genetic discoveries were further improved by the introduction of multivariate analyses based on the multi-trait analysis of genome-wide association summary statistics (MTAG)[40]; this approach involves joint analysis of multiple (genetically correlated) traits, which increases statistical power to detect genetic associations for each trait (**Box 1**). With the use of this method, two additional risk loci were revealed for ICH[33] in a joint analysis with SVS, and seven additional risk loci for SVD in a joint analysis with WMH[34] (**Table 1**). Most of the identified risk loci for cSVD-related stroke require definitive confirmation in independent follow-up samples, even though rigorous methodology was used. Particular caution is warranted for the additional loci derived from MTAG, as this approach relies strongly on the assumption that the associated loci act on both traits[40], so additional sensitivity analyses are required to test the relative contribution of the included traits.

When univariate and multivariate findings are combined, and the fact that three loci are shared between ICH and SVS (at chr1q22, chr2q33 and chr13q34) is taken into account, a

total of 17 independent risk loci for cSVD-related stroke have been identified to date. At nearly all of these loci, the lead risk variants are common variants (minor allele frequency ≥5%), except for one low-frequency variant (minor allele frequency 3%) at chr22q13 associated with haematoma volume of lobar ICH[32].

With the possible exception of risk loci in or near genes that are known to cause Mendelian forms of cSVD (for example, the *COL4A1–COL4A2* and *HTRA1* loci), which are discussed in detail below, the biology that underlies the genetic associations with cSVD-related stroke is still uncertain and requires further investigation. Indeed, although risk loci are often labelled with the name of the gene that is nearest to the lead risk variant, the gene through which risk variants exert their effects could be more distant and cannot be identified with certainty without extensive functional explorations. One such functional exploration is the combination of GWAS findings with those from gene expression quantitative loci eQTL studies (**Box 1**) in transcriptome-wide association studies (TWAS; **Box 1**), which can provide important information about putative causal genes. Use of this approach has suggested that the gene closest to the GWAS top hit is often not the causal gene[41].

A combination of transcriptome-wide association studies[42] and colocalization analyses (**Box 1**)[43] has been used to gain insight into specific genes associated with SVS. These studies rely on existing gene expression databases in brain and vascular tissues that were not specifically designed for the study of cSVD[44]. Nevertheless, they provide important information that enables loci identified in GWAS (some of which encompass a large number of genes) to be refined to a limited set of putative causal genes that can be prioritized for functional follow-up. This approach has indicated that the expression level of several genes is associated with SVS: *SLC25A44* at chr1q22, *ULK4* at chr3p22 and *ICA1L, FAM117B, CARF* and *NBEAL1* at chr2q33[34].

SLC25A44 encodes a mitochondrial solute transporter of branched-chain amino acids[34]. Interestingly, another locus associated with SVS also contains the solute carrier gene *SLC39A13* — this gene encodes a zinc transport transmembrane protein, mutation of which causes spondylocheirodysplastic Ehlers-Danlos syndrome, an inherited connective tissue disorder with rare vascular complications[45]. Solute carriers are known to play a major role in cerebrospinal fluid secretion and transport of various substances at the blood–cerebrospinal fluid barrier[46] and the blood–brain barrier[47]. Their expression is enriched in human blood–

brain barrier tissue and they are thought to have a role in xenobiotic transport[48,49]. Solute carriers could, therefore, be relevant in cSVD because dysfunction of the blood–brain barrier is thought to be an important feature of the condition and to play a deleterious role by promoting inflammation, inhibiting myelin formation and repair processes, and contributing to stiffening vessel walls[12].

ULK4 encodes a serine–threonine protein kinase[34] that has a role in neurogenesis in mice[50] and deficiency of which results in hypomyelination[51]. In UKL4 mutant mice, major transcription factors involved in oligodendrogenesis are downregulated and myelination is reduced by 50%, demonstrating that ULK4 is an important regulator of myelination[51]. Neuropathological and experimental studies suggest that myelin injury is an important and early feature of cSVD[52,53]. At the chr2q33 locus, TWAS and colocalization analyses suggested that expression levels of both ICA1L and NBEAL1 are associated with cSVD, in brain tissue for ICA1L and vascular tissue for NBEAL1[44]. ICA1L encodes islet cell autoantigen 1 like and mutations in this gene cause juvenile amyotrophic lateral sclerosis[54]. NBEAL1 encodes neurobeachin-like 1 protein, which regulates cholesterol metabolism by modulating expression of low-density lipoprotein-receptor expression[55]. Further studies are needed to explore the potential functional links of these genes with cSVD.

We have deliberately chosen not to discuss candidate gene association studies (**Box 1**) owing to methodological flaws in the majority of these studies. Nevertheless, two risk loci for cSVD-related stroke that were identified by GWAS had already been described in large candidate gene association studies. These were the association of the APOE ε2 and ε4 alleles (at chr19q13) with an increased risk of ICH[56] and the association of the COL4A1–COL4A2 locus (at chr13q34) with ICH[57,58] and SVS[58].

[H2] MRI markers of covert cSVD

Large collaborative projects conducted by the Cohorts for Heart and Aging Research in Genomic Epidemiology (CHARGE) consortium and the ISGC in combination with the large UK Biobank brain imaging dataset have led to substantial progress in the discovery of genes associated with MRI markers of covert cSVD. Genetic associations with WMHs are the most extensively studied. Some studies have also been done to explore the genetic underpinnings

of MRI-defined brain infarcts, lacunes, cerebral microbleeds and DTI markers of white matter integrity, such as fractional anisotropy and mean diffusivity. These DTI markers are less specific than the aforementioned, more traditional MRI markers of cSVD but can reveal alterations at a young age, and some evidence suggests that they precede the occurrence of WMHs in cSVD[59]. Overall, GWAS of quantitative MRI markers of cSVD traits (WMHs and DTI metrics) that largely rely on automated measurements have been better powered than similarly sized GWAS of binary traits (brain infarcts, lacunes and cerebral microbleeds) that are assessed with semi-quantitative visual rating.

To date, GWAS have identified a total of 44 independent genetic risk loci for MRI markers of covert cSVD at the genome wide significance level, in 35 regions (multiple independent signals have been identified at some loci). Twenty-nine genomic regions are associated with larger WMH volume[21,22,44,60-63]; two independent signals were identified in four of these regions and three independent signals were identified in one region, leading to 34 independent associations. Two loci, at *NOS3* (chr7q36) and *TSPAN14* (chr10q23) were specifically associated with periventricular WMH volume (not with overall or deep WMH volume) at genome-wide significance (**Table 2**). Seven genomic regions are associated with reduced fractional anisotropy, and two independent signals were identified in one of these regions, leading to eight independent loci[21]; six loci are associated with higher mean diffusivity[21], one with the risk of cerebral microbleeds[64] and two with the risk of brain infarcts[65] (**Table 2**). Again, associations were observed predominantly in populations of European ancestry and all lead single-nucleotide polymorphisms (SNPs) were common variants. According to gold standards, replication of these associations in independent samples of adequate size is required for definitive confirmation, but maximizing the size of discovery samples has so far been prioritized.

The biology underlying genetic associations with MRI markers of covert cSVD remains largely elusive. However, TWAS combined with colocalization analyses have provided preliminary evidence that expression of many genes is associated with WMH volume: *CALCRL* at chr2q32; *ICA1L, CARF, NBEAL1* and *KRT8P15* at chr2q33; *DEGS2* and *CCDC88C* at chr14q32; *DCAKD, NMT1, NT5C3B* and *STAT5B* at chr17q21; and *TRIM47, TRIM65, WBP2 and RP11-552F3.9* at chr17q25[21,44].

CALCRL encodes a component of the calcitonin gene-related peptide receptor (CGRP), and low levels of the *CALCRL* transcript in arterial and nerve tissue and high levels in blood were associated with larger WMH volume. At the chr2q33 locus, three of the identified genes[44] are the same as genes identified in the SVS (*ICA1L, CARF* and *NBEAL1*)[34]. *DEGS2* encodes a bifunctional enzyme involved in the biosynthesis and metabolism of sphingolipids, which have been associated with white matter integrity[66] and stroke outcome [67], and a common missense variant in *DEGS2* is associated with cognition in schizophrenia[68]. *CCDC88C* encodes a coiled-coil domain-containing protein that acts as a negative regulator of the Wnt signalling pathway, mutations of which are known to cause autosomal recessive hydrocephalus[69]. The canonical Wnt–β-catenin pathway plays an important role in regulating blood–brain barrier integrity and function[70], is involved in response to vascular lesions or ischaemic events in vascular cells[71], and seems to be involved in oligodendrogenesis and myelination[72], all of which are factors in cSVD. More work is needed to fully understand the functional roles of these genes in cSVD.

The chr17q21 region is intriguing, as it lies within a region of high linkage disequilibrium that spans ~1.8 megabases and contains a 900 kb inversion polymorphism that encompasses *MAPT* (which encodes microtubule associated protein tau) and other genes, including *CRHR1* and *KANSL1*[73]. This inversion polymorphism has been associated with brain disorders[74-79], cognitive[80] and behavioural[81-83] phenotypes and brain MRI markers[84-86]. The lead WMH risk variants at chr17q21 are located upstream of the inversion polymorphism, in an intron of *NMT1* and near *DCAKD, NT5C3B* and *STAT5B*, whereas the lead mean diffusivity risk variant in this region lies in an intron of *MAPT*. Interestingly TWAS and colocalization analyses revealed transcriptome-wide associations of WMH volume not only with *NMT1, DCAKD, NT5C3B* and *STAT5B* but also with *MAPT, CRHR1* and *KANSL1*[21,44], highlighting the distant cis-effect conferred by an SNP–eQTL pair. *DCAKD* encodes dephospho-CoA kinase domain containing protein and has a putative role in neurodevelopment[87]. *NMT1* encodes N-myristoyltransferase 1, which is thought to affect astrocyte function[61] and is involved in early embryogenesis[87]. eQTL analyses have indicated that *DCAKD* and *NMT1* are also associated with fractional anisotropy and mean diffusivity in multiple white matter tracts[87]. Chromosomal rearrangements are common in this dynamic region[88,89], so the study of structural variants could provide insight into the risk architecture of this pleotropic and complex genetic locus.

The chr17q25 locus is the most significant risk locus for MRI markers of covert cSVD, specifically for WMH volume, and its association has been replicated in multiple independent European and Japanese samples[90-92]. In this gene-rich region (40 genes within 500 kb of the lead variant), *TRIM47, TRIM65, WBP2* and *RP11-552F3*.9 are the only significantly associated genes in TWAS and colocalization studies[21,44], which makes them prime candidates for causal genes **(Box 1)**. *TRIM47* and *TRIM65* encode E3 protein ligases, which are crucial regulatory proteins in multiple signalling pathways. *WBP2* is involved in the aforementioned Wnt–β-catenin signalling pathway. Extensive functional studies are warranted to identify the causal gene or genes at this locus. Use of exome chip data[93] has also revealed an association of two low-frequency missense variants in *MRPL38*, which is located at chr17q25.1, with WMH burden[93], but this association requires confirmation.

Finally, data from a study published in 2020 provide important insight into the lifetime effects of genetic risk factors for cSVD[44]. Risk variants associated with extensive WMH volume in old age (in aggregate and individually, most prominently a variant in *VCAN*) were already associated with subtle changes in white matter microstructure detectable with diffusion-weighted imaging in adults in their twenties. This finding suggests that biological pathways contributing to WMHs in old age already have an impact on brain microstructure at a very young age[44]. The changes could reflect a very early stage of cSVD or could indicate pleiotropy between genes associated with cSVD and those that influence white matter maturation, which peaks in early adulthood.

[H2] Biological insights

More work is required to precisely decipher the molecular pathways that underlie the observed genetic associations with cSVD, but the existing data provide some preliminary indications. However, several biological processes are implicated by genetic associations, including the structure and function of the extracellular matrix (ECM), myelination and membrane transport (**Table 3**). In addition, some genes associated with cSVD are also associated with other brain disorders, such as Alzheimer disease and psychiatric disorders, suggesting either mechanistic overlap or genetic pleiotropy.

[H3] Molecular pathways

A considerable number of genetic associations point to an important role for genes that are involved in the structure and function of the extracellular matrix (ECM; **Table 3**), including: genes that can harbour common variants that predispose to cSVD or rare mutations that cause monogenic cSVD (*COL4A2* and *HTRA1*); genes that TWAS and colocalization analyses suggest are causally implicated in cSVD (*EFEMP1, ADAMTSL4, VCAN*); and genes that are in the immediate vicinity of loci associated with cSVD and covert cSVD (*GPR126, LOX, SH3PXD2A, NID2, FBN2* and *HBEGF*). The roles of these genes in the ECM indicate overlap with mechanisms of monogenic SVD, even for genes that are not known to harbour mutations that cause monogenic cSVD. For example, in cerebral autosomal recessive arteriopathy with subcortical infarcts and leukoencephalopathy (CARASIL; a form of monogenic cSVD caused by *HTRA1* mutations), accumulation of versican core protein, encoded by *VCAN*, has been observed in the thickened arterial wall[94]. Similarly, application of heparin-binding EGF-like growth factor, which is encoded by *HBEGF*, in mouse a model of CADASIL (the most common monogenic cSVD caused by *NOTCH3* mutations) rescued reduced myogenic responses of pial and parenchymal arteries through functional upregulation of $K_V 1$ channels at the plasma membrane of arterial smooth muscle cells[95]. Thus, in line with converging evidence from monogenic SVD, perturbations of the ECM also seem to play a central role in common cSVD pathophysiology[96].

Specifically, perturbations of the cerebrovascular matrisome have been proposed as a convergent pathological pathway in monogenic cSVD and preliminary indirect evidence suggests that this also applies to common cSVD[96]. The term matrisome refers to the ensemble of proteins that make up the ECM together with proteins that are associated with the ECM[96,97], which seem to be expressed in an organ-specific pattern. Preliminary characterization of the matrisome of pial arteries and brain microvessels indicates the presence of several proteins encoded by genes associated with cSVD (*COL4A2, HTRA1, ADAMTSL4, NID2, FBN2* at pial arteries and *COL4A2, VCAN, NID2* at microvessels*)*, and others by genes (*EFEMP1* and *LOX*) that are involved in the cerebrovascular matrisome[96].

Risk loci for common cSVD also indicate involvement of genes that are known to be involved in myelination: *ULK4*, *VCAN* and *ADGRG6* (**Table 3**). As described above, *ULK4* is an important regulator of myelination[51]. Mutations of *adgrg6* in zebrafish and mice lead to

Schwann cell myelination dysfunction in the PNS[98,99] and *VCAN* encodes versican core protein, which can form complexes that inhibit oligodendrocyte maturation and affect myelination[100]. Combined TWAS and colocalization analyses have also identified significant associations of *ULK4* and *VCAN* expression with cSVD in relevant tissues[51,100], indicating causal involvement of these genes. Interestingly, in zebrafish, mutations of *adgrg6* lead to a failure to down-regulate *versican*[101,102], indicating that *ADGRG6* might be involved via downstream effects on *VCAN* expression.

Several cSVD lead risk variants also involve genes that encode membrane transport proteins, including ion channel proteins (*KCNK2*, encoding a potassium channel expressed most prominently in fetal brain) and the solute carrier proteins (*SLC25A44, SLC39A13*) discussed above[34,44] (**Table 3**). Interestingly, in a genetic model of cSVD (CADASIL), potassium channelopathy-like defects were shown to underlie early-stage cerebrovascular dysfunction, although these did not directly involve *KCNK2*[95]. Further research is needed to explore the possible involvement of these genes in cSVD.

Analysis of brain-specific, single-cell expression data in mice has revealed that genes within loci associated with WMHs are enriched in endothelial cells[44]. Interestingly, endothelial cell dysfunction has been described in the spontaneously hypertensive stroke-prone rat model of cSVD, — vascular tight junctions were compromised, and endothelial dysfunction impaired oligodendrocyte maturation and, consequently, myelination, and increased microglial activation[103]. Nominally significant enrichment of genes within WMH risk loci has also been observed in vascular leptomeningeal cells, oligodendrocytes, oligodendrocyte precursors and ependymal astrocytes[44], suggesting that these cell types are also involved in cSVD pathophysiology.

[H3] Associations with other brain disorders

Several risk loci associated with covert cSVD are near genes that have a role in the pathogenesis of Alzheimer disease (AD) or psychiatric disorders, which can be associated with features of cSVD. For AD, these loci include *APOE, ECHDC3, HAAO, MAPT* and *SH3PXD2A*. The *APOE* ε4 allele is the most important genetic risk factor for AD[104]. *ECHDC3* has been associated with AD in multi-ancestry GWAS and its expression is altered in the brains of

people with AD compared with controls without AD[105]. *HAAO* is involved in the kynurenine pathway, which has been implicated in the pathogenesis of neurodegenerative disorders, including AD[106]. *MAPT* encodes microtubule associated protein tau, which is a key protein in AD pathogenesis[107]. *SH3PXD2A* is involved in neurotoxicity of amyloid-β, which is widely accepted as the pathological protein that drives AD[108]. However, whether these associations reflect pleiotropy, whereby some genetic risk loci independently influence cSVD and AD, or whether they point to shared mechanisms or mediating effects between both diseases cannot be inferred from these data[23,44].

Some loci identified in TWAS analyses as being associated with cSVD have also been associated with psychiatric disorders[21]. Most of them (*PLEKHM1*, *CRHR1–IT1*, *DND1P1* and *KANSL1–AS1*) are located in the chr17q21 region described above and have been associated with depressive symptoms and neuroticism[109]. Others (*BTN3A2*[110], *MICA*[111] and *ZNF165*[111]) have been associated with schizophrenia. These genes were mainly associated with fractional anisotropy and mean diffusivity[21], suggesting that DTI markers could capture alterations in brain networks that are involved in the genesis of psychiatric disorders. How these genes may be involved in cSVD requires further investigation.

[H1] Shared genetic variation

Use of linkage disequilibrium score regression (**Box 1**) has revealed significant genetic correlation (reflecting shared genetic determinants at the genome-wide level) between WMH volume and all stroke, ischaemic stroke, SVS, and lower general cognitive function[44]. Strong evidence also indicates significant shared genetic risk factors between WMH, fractional anisotropy and mean diffusivity[21].

Considerable overlap also exists between the genome-wide significant risk loci associated with different cSVD-related phenotypes (**Fig. 2A**). From a total of 52 independent genome-wide significant risk loci for cSVD phenotypes, 34% were associated with at least two distinct cSVD traits at the genome-wide significance level. Some genomic regions are particularly pleiotropic: the chr16q24 locus at *ZCCHC14* is associated with WMH[21,44], fractional anisotropy[21] and SVS[36,37,112]; the chr19q13 locus at *APOE* is associated with WMHs[21,44] ICH[31] and cerebral microbleeds[64]; the chr2q33 locus at *ICA1L, CARF* and *NBEAL1* is

a risk locus for fractional anisotropy, WMH[21,44,61] and ICH[33]; and the chr5q14 locus at *VCAN* is associated with WMHs, fractional anisotropy and mean diffusivity[21].

We systematically explored the overlap between published genome-wide significant risk loci for cSVD (brain infarcts, cerebral microbleeds, fractional anisotropy, mean diffusivity, ICH and SVS) and variants associated with established vascular risk factors (systolic blood pressure (SBP), diastolic blood pressure (DBP), pulse pressure[113], type 2 diabetes mellitus[114], low-density lipoprotein [LDL] cholesterol, high-density lipoprotein [HDL] cholesterol, triglycerides[115], BMI[114] and lifetime smoking index)[116]. Briefly, variants located in close proximity (±250 kb) to the lead cSVD risk variants and variants in moderate to high linkage disequilibrium (r^2 >0.5) with these lead SNPs were explored for significant associations with the vascular risk factors, while accounting for the number of independent cSVD risk variants and vascular risk factors tested (**Fig. 2B**). Of the 29 WMH risk loci, 18 (62%) were significantly associated with at least one other vascular risk factor[21,44], especially blood pressure traits (59%), most often at genome-wide significance. A much smaller proportion (25–29%) of overlap was observed for fractional anisotropy and mean diffusivity (**Fig. 2B**), a finding that can at least partly be attributed to the smaller sample size available for these traits. The cSVD risk locus with the most prominent 'vascular pleiotropy' was at chr2q33 (*ICA1L, CARF* and *NBEAL1*) and was significantly associated with smoking, BMI, type 2 diabetes mellitus, lipid levels and blood pressure (**Fig. 2B**).

Importantly, the relationship of pleiotropic loci with cSVD and vascular risk factors is not necessarily driven by the same causal variant and does not imply mediation of the genetic association with cSVD by the vascular risk factors. Moreover, 25 of 52 cSVD risk loci are not shared with any vascular risk factor, suggesting novel mechanistic pathways. These loci are involved in organization of the ECM (*HTRA1, NID2, VCAN, LOX, FBN2, GPR126*), myelination (VCAN, GPR126), membrane transport (*KCNK2*), and loci that are known to be target genes for microRNA-mediated silencing (*DENND1B, KCNK2, ZDHHC20, ZCCHC14, FOXQ1, NID2, VCAN*)[117].

Of note, brain MRI features of cSVD with or without clinical symptoms are not entirely specific for underlying cSVD. For example, in rare instances, lacunes or SVS could be caused by emboli that originate from undetected cardiac sources or parent artery atheroma. GWAS have identified no overlap of genome-wide significant risk loci for SVS with cardioembolic

stroke, large artery stroke or atrial fibrillation[36], but significant genetic correlation was observed between large artery stroke and SVS, suggesting that some genetic contribution to these phenotypes is shared[118]. Further research is needed to differentiate causally distinct subtypes of lacunes and SVS, and use of more descriptive terminology that does not imply aetiology has been recommended; for example, 'recent small subcortical infarcts' to refer to neuroimaging evidence of recent infarction in the territory of one perforating arteriole[13]. Composite brain imaging phenotypes that combine lacunes and recent small subcortical infarcts with extensive WMH have also been proposed as a way to increase specificity for cSVD[30,119].

[H1] Continuum with monogenic cSVD

Mounting evidence suggests a continuum between the genetic causes of monogenic cSVD and of common cSVD[27]. In particular, genetic variants that affect the *HTRA1* and *COL4A1–COL4A2* loci have been associated with monogenic and common forms of cSVD[120,121].

Mutations in *HTRA1* cause CARASIL, a rare autosomal recessive form of cSVD that is caused by loss of function of *HTRA1* and has been reported mainly in the Japanese population[94], and *HTRA1*-autosomal dominant cSVD, caused by dominant negative mutations that lead to decreased serine protease HTRA1 activity[122]. Mutations in the *COL4A1* and *COL4A2* genes lead to impaired deposition of collagen in vessel walls and can cause collagen 4A1/A2 microangiopathy, an autosomal dominant cSVD that was initially thought to primarily cause ICH. However, the spectrum of phenotypes caused by *COL4A1* and *COL4A2* mutations seems to be very broad, including specific entities such as hereditary angiopathy, nephropathy, aneurysms and cramps (HANAC), pontine autosomal dominant microangiopathy and leukoencephalopathy (PADMAL), and multi-infarct dementia of the Swedish type[27,123,124].

Common variants in or near *HTRA1* and the *COL4A1–COL4A2* locus have been associated at the genome-wide significance level with WMH burden (*COL4A1–COL4A2*)[21,44,60,61], with ICH in multi-trait analyses with SVS (*COL4A1–COL4A2*)[33] and with SVS in multivariate analyses accounting for WMH (*COL4A1–COL4A2* and *HTRA1*) (**Tables 1 and 2**)[34]. In addition, well-powered candidate gene studies had previously identified associations of common variants at the *COL4A1–COL4A2* locus with ICH[57,58], SVS[58] and WMH burden[125],

and of a common variant near HTRA1 (also an eQTL of HTRA1) with SVS and a composite MRI-defined extreme cSVD phenotype with extensive WMH burden and/or presence of lacunes[119]. In the largest published GWAS for stroke, which involved >67,000 patients, common variants at the HTRA1 and COL4A1–COL4A2 loci were also associated with any stroke[36,126], probably reflecting an underlying contribution of cSVD.

Another locus that causes monogenic cSVD and is associated with common cSVD is the SH3PXD2A–STN1 locus at chr10q24. Mutations in STN1 cause cerebroretinal microangiopathy known as Coats-plus syndrome[127]. Common variants at the SH3PXD2A–STN1 locus have been associated at the genome-wide significance level with WMH burden, mean diffusivity, and SVS in a multitrait GWAS with WMH burden (**Tables 1 and 2**).

Published GWAS of cSVD phenotypes have not revealed any genome-wide significant associations with common variants in or near NOTCH3, mutation of which causes cerebral autosomal dominant arteriopathy with subcortical infarcts and leukoencephalopathy (CADASIL), the most prevalent monogenic form of cSVD[27]. However, a population-based candidate gene studyidentified a nominally significant association of common variants in or near the NOTCH3 locus with WMH volume and WMH progression in participants with hypertension[128], and a candidate whole-exome sequencing study demonstrated a significant association between the burden of protein-modifying rare or low-frequency variants in NOTCH3 and extreme cSVD in people aged $\geq$65 years[119]. In addition, cysteine-altering rare variants in NOTCH3 that are typical in CADASIL are much more common in community exome biobanks (1 in 300) and in elderly people in the community (1 in 300–500)[119,129-131] than expected on the basis of the reported prevalence of CADASIL (~4 in 100,000[120,121]), especially variants in EGFr domains 7–34 of NOTCH3, which are associated with milder CADASIL phenotypes than mutations in EGFr domains 1–6, which cause severe CADASIL phenotypes[130]. Interestingly, in over 200,000 participants of the population-based UK Biobank, these 'asymptomatic' cysteine-altering NOTCH3 variants were associated with an increased risk of stroke and vascular dementia, increased WMH volume and an increased prevalence of lacunes and cerebral microbleeds[131], accounting for an increased burden of apparently 'sporadic' stroke and dementia[131].

Another important locus that has been associated with both monogenic and multifactorial cSVD is FOXF2–FOXC1. Common variants near FOXF2 have been associated with

all stroke, with a strong indication that the signal was driven by underlying cSVD[132], and with SVS at genome-wide significance in a multitrait analysis with WMH[34]. *FOXF2* is located near *FOXC1*, mutations or copy number variations in which cause Axenfeld–Rieger syndrome, a rare monogenic ophthalmological condition co-occurring with cerebral lesions seen in cSVD. The WMH burden is nearly tenfold greater among patients with segmental deletions that encompass both *FOXC1* and *FOXF2* than in patients with deletions of *FOXC1* only[132], suggesting involvement of *FOXF2* in cSVD pathogenesis. Further supporting this conclusion, deletion of *Foxf2* in adult mice leads to cerebral infarction, reactive gliosis and microhaemorrhage, and cerebral vessels of mutant *foxf2b*$^{-/-}$ zebrafish have decreased smooth muscle cell and pericyte coverage, suggesting that the mutation leads to defects in differentiation of cerebral vascular mural cells[132]. Interestingly, although *FOXF2* does not directly encode an ECM protein, its protein product, forkhead box protein F2 (FOXF2) is also involved in vascular stability[132-135] and blood–brain barrier formation[136]. In the context of cancer, FOXF2 modulates the degradation of the ECM by activating metalloproteinase inhibitor 3 (TIMP3)[137], which has an important role in CADASIL pathophysiology[96].

[H1] Mendelian randomization studies

Mendelian randomization is an instrumental variable analysis that can be used to investigate the likelihood that an exposure contributes to the causal pathway of a specified outcome (**Box 2**)[138]. Numerous Mendelian randomization studies have been done to explore the causal relationships of putative risk factors and biomarkers with cSVD, and the list is continuously growing (**Fig. 3**). We acknowledge that the following summary of Mendelian randomization studies might soon be outdated, given the rapid growth of this body of literature. Moreover, we have not systematically reported how potential biases in estimating the causal inference have been addressed, as methods used differ across studies and a comprehensive description of these is beyond the scope of this Review.

Mendelian randomization has been used in a number of studies to explore the causal relationship between vascular and lifestyle risk factors and cSVD phenotypes. Although these factors are established risk factors for other common types of ischaemic stroke, especially large artery stroke, their relationship to cSVD-related stroke and MRI-defined covert cSVD has been controversial. The most undisputed association, which has been confirmed by

Mendelian randomization approaches[44,139], is the strong relationship between high blood pressure and most cSVD phenotypes. Importantly, genetically predicted systolic and diastolic blood pressure were associated with WMH volume even in participants without clinically defined hypertension (systolic blood pressure ≥140 mmHg, diastolic blood pressure ≥90 mmHg or antihypertensive drug intake), highlighting that higher blood pressure is likely to be causally associated with covert cSVD even at blood pressures below the thresholds that typically define hypertension[44].

cSVD phenotypes also seem to be associated with genetically determined lipid levels, but these associations are more complex. Mendelian randomization studies have shown that genetically predicted higher levels of HDL-cholesterol are associated with a lower risk of SVS[36,140,141] and, in some studies, lower WMH burden[141]. However, genetically predicted higher levels of triglycerides were associated with a higher risk of SVS and cerebral microbleeds[64,140,141]. Genetically determined high levels of LDL-cholesterol have been associated with an increased risk of SVS and a lower risk of ICH in some studies[140,142] but not all[34,141,143]. Single studies have identified significant associations of genetically predicted higher levels of lipoprotein (a)[144] and apolipoprotein A-I[140] with a lower risk of SVS and higher levels of apolipoprotein B with an increased risk of SVS[140].

Mendelian randomization studies have also been done to explore the relationship between type 2 diabetes mellitus and cSVD. Several studies of partly overlapping cohorts of patients with SVS have shown associations between genetically predicted type 2 diabetes mellitus and SVS[34,36,145-147] and no significant association of diabetes with ICH[145,147]. The largest of three studies[44,145,147] showed a significant association of genetically predicted type 2 diabetes mellitus with larger WMH volume, although residual pleiotropy was observed after removal of outlier variants. In the UK Biobank, genetically predicted type 2 diabetes mellitus has been associated with lower mean fractional anisotropy[145,147].

Mendelian randomization studies have also identified at least one significant association of several other factors with cSVD, including other vascular risk factors[36,44,64,65,125,145-149], lifestyle factors such as diet[36,44,150-154], comorbidities[36,44,155,156] and some circulating biomarkers[157-161] (**Fig. 3**). In most of these Mendelian randomization studies, two-sample Mendelian randomization has been used (**Box 2**), and where several studies have been published on the same exposure–outcome relationship, they were often based on the same or overlapping datasets. Discrepancies in results between studies are mostly related to

variation in the power of the summary statistics used for the outcome (for example, use of results from the earlier METASTROKE GWAS or the more recent MEGASTROKE SVS GWAS), in the definition of genetic instruments, or in the extent to which residual pleiotropy or reverse causation was addressed.

[H1] Clinical and therapeutic implications

Clinical interpretation of the aforementioned genetic risk loci for cSVD requires caution, as most of them need independent replication and/or functional validation in future studies. Transitioning from discovery of genetic risk loci to robust identification of affected genes and pathways that underlie cSVD pathophysiology is one of the major challenges ahead. This process is likely to require various complementary approaches, such as next-generation sequencing, multiancestry fine-mapping, systems genetics, multiomics, single-cell genomics, bioinformatics approaches and, importantly, experimental follow-up. Such experimental follow-up could range from high-throughput assays that provide functional readouts for putative regulatory elements and variants, such as Massively Parallel Reporter Assays (MPRA)[162], to elaborate, customized experiments in various model organisms, such as zebrafish or mice[132]. Below, we consider the potential of common cSVD genomics for eventual clinical and therapeutic application.

Although still in its infancy, bioinformatics and experimental follow-up of loci identified in GWAS can inform drug development, with the potential to accelerate drug discovery through the identification of novel drug targets, indirect measurement of drug effects using genetic instruments, and by providing evidence for drug repositioning (**Fig. 1**)[26,163]. Strong bioinformatics and preliminary experimental evidence support the involvement of several genes in the pathophysiology of cSVD that are currently being followed up experimentally to confirm their involvement and, when relevant, explore their potential as drug targets for cSVD. Moreover, increasingly rich molecular and bioinformatics resources are being used to generate additional evidence on the cSVD risk loci identified to discover further candidates for experimental follow-up.

With the increasing popularity of Mendelian randomization, there is also growing interest in the use of genetic instruments to mimic the effect of drugs of interest[25]. In this approach, genetic instruments from genetic regions involved in the production of a drug

target are prioritized when constructing instruments for a given exposure (for example *ACE,* which encodes a target for ACE inhibitors which lower blood pressure, would be prioritized when constructing instruments for high blood pressure)[164]. It has been estimated that genetic support of drug effects through such Mendelian randomization approaches could double the success rate of clinical trials[25]. Mendelian randomization can also be used as an indication for preferential effects of certain drug classes on a given phenotype, keeping in mind methodological limitations and caveats and the need for validation in a trial setting. For instance, genetic proxies for calcium channel blockers — that is, in genetic instruments that mimic the effects of calcium channel blockers — but not β-blockers were associated with a lower risk of SVS and WMH[139], though some evidence indicated an association with diverticulosis[164]. Similarly, genetic proxies for cholesterylester transfer protein (CETP) inhibitors (which raise levels of HDL) were shown to be associated with a lower risk of SVS and a lower WMH burden, whereas variants at the *CETP* locus were associated with a higher risk of ICH (both deep and lobar)[141]. Confirmation of these findings in a clinical trial setting is required.

Mendelian randomization can also be used to explore pathways of therapeutic relevance. For example, Mendelian randomization studies have shown that a predisposition to β-cell dysfunction was associated with SVS and ICH, which might have implications for use of anti-diabetic treatments to target these mechanisms[145]. A Mendelian randomization approach based on data from the MEGASTROKE GWAS identified a significant association between a weighted genetic risk score for venous thromboembolism and all ischaemic stroke subtypes except SVS, despite similar or larger power across subtypes, suggesting that thrombotic processes play a less important role in SVS. This insight has potential implications for therapeutic research on antithrombotic treatment in cSVD[36,165].

Drug target enrichment analyses can be used to seek evidence for drug repositioning opportunities. For instance, this approach has been applied to the largest WMH GWAS to quantify the enrichment of the 39 associated genes that emerged from the TWAS–colocalization analysis. Use of curated drug-target lists classified on the basis of the International Classification of Diseases 10 (ICD10)[44] showed that of the 39 genes, four (*MAPT, CRHR1, CALCRL* and *NOTCH4*) were registered as targets of approved drugs in the DrugBank database and the Therapeutic Target Database. *MAPT* is a drug target under investigation for neurodegenerative disorders and TWAS suggested that larger WMH volume is associated

with upregulated *MAPT* expression. CGRP, the product of *CALCRL*, is targeted by monoclonal antibodies in the treatment of migraine[44].

Findings from GWAS also have the potential to enable risk stratification and risk prediction of cSVD and its major complications (stroke and dementia)[166]. In epidemiological studies, MRI markers of covert cSVD have repeatedly been associated with an increased risk of developing stroke (ischaemic and haemorrhagic) and dementia, including AD[23]. Extensive WMH burden has also been shown to predict amyloid-β positivity after the age of 65 years, although the causal relationship between cSVD and AD was uncertain[167]. Mendelian randomization has provided evidence for an association of larger genetically determined WMH volume with an increased risk of ischaemic stroke, intracerebral haemorrhage and Alzheimer type dementia, supporting causality[44]. Covert cSVD, therefore, represents a major opportunity to detect individuals who are at risk of developing major neurological events and could benefit from targeted preventive interventions. However, risk stratification in people with cSVD is lacking, as is accurate, quantitative estimates of the ability of the full covert cSVD spectrum to predict the risk of stroke and dementia. This estimate needs to account for several methodological challenges, such as time-varying exposure, competing risk of death, and modifying factors such as coexisting neurodegenerative disease and vascular comorbidity. Integration of cSVD genetic risk data into risk stratification and risk prediction algorithms is becoming more and more relevant, as increasingly powerful genetic instruments are available that at least partly address the aforementioned challenges.

Recent work has highlighted the potential of polygenic risk scores (PRS) for prediction of common diseases[168,169]. PRS aggregate small effects of independent genetic variants associated with a disease (at varying significance thresholds) into a single score, which reflects the cumulative genetic risk of a disease[170]. For some common diseases, use of PRS has identified individuals with a risk equivalent to that for people with a monogenic mutation that causes the diseases; for example, PRS showed that 8% of the population with the largest number of risk alleles for coronary artery disease had a threefold higher risk of developing this condition than the remainder of the population[171]. A novel approach called metaGRS combines several PRS into the same model and enables greater risk discrimination, including for stroke[172]. UK biobank participants (from the general population) who had metaGRS scores in the top 0.25% for stroke and vascular risk factors also had a threefold higher risk of

developing ischaemic stroke during follow-up than the rest of the population[172]. Similar approaches could be efficient ways to model cSVD risk and could guide prevention approaches by identifying individuals who are at the highest risk of developing specific cSVD complications and could benefit from lifestyle modifications in a precision medicine setting[173] or could be enrolled in clinical trials to assess preventive interventions. The added value of genetic risk variants in the prediction of stroke or dementia above and beyond covert cSVD MRI markers still needs to be demonstrated, but genetic variants have the advantage of being measurable at increasingly low cost and throughout a lifetime, including before visible covert cSVD develops, so could enable very early detection of high-risk populations.

[H1] Conclusions and perspectives

In summary, through large international collaborative efforts, major progress has been made in deciphering the genetic underpinnings of cSVD. Fifty-two independent genetic risk loci for various cSVD-related phenotypes have been identified by combining large population-based and hospital-based cohorts with cutting-edge analytical approaches. Moreover, association statistics from published GWAS have been leveraged for a wide array of secondary analyses, such as assessment of shared genetic variation across cSVD phenotypes and exploration of causal pathways using Mendelian randomization. Although more research is required to derive clinical and therapeutic applications, these findings already provide extensive novel insight into the biological pathways that underlie cSVD, the roles of putative risk factors, the clinical significance of cSVD across the adult lifespan, and the potential implications for therapy. They also highlight the complexity of the cSVD spectrum by revealing that partly distinct genetic profiles underlie the different phenotypic expressions of the disease, and that a continuum exists between common variants associated with common cSVD and genes that harbour causal mutations for rare monogenic cSVD or other genes involved in monogenic cSVD pathophysiology.

Additional progress is to be expected in the near future as a result of multiple developments. First, to enhance genomic discovery, more cSVD genomic studies are needed that encompass exploration of emerging MRI markers of covert cSVD, such as dilated perivascular spaces and novel diffusion-based metrics. These markers also provide an opportunity to capture early mechanisms, as they can be detected and quantified much

earlier in life than traditional MRI phenotypes of cSVD. Preliminary results discussed above support such a lifespan approach. Second, the availability of next-generation sequencing (whole-exome and whole-genome sequencing) of increasingly large samples provides opportunities to discover relevant genetic variation other than single nucleotide variants, such as structural variants or short tandem repeats, and to maximize the fine-mapping potential of GWAS loci to identify causal variants. Third, cSVD genomics have been explored predominantly in populations of European ancestry to date, so expanding analyses to other ethnic groups, some of which have a considerably higher prevalence of cSVD than Europeans (for example, populations of African and East-Asian ancestry)[174-176], will be crucial to establish the generalizability of associations and maximize the potential for fine-mapping and risk prediction[177]. Finally, combining cSVD genomics with other molecular approaches (such as epigenomics, transcriptomics, proteomics and metabolomics) will be essential to capture the complex interplay of factors beyond DNA sequence that lead to disease[178]. These combined approaches are currently being explored in the CHARGE and ISGC consortia[179], as well as in several funded initiatives, including COSTREAM, MarkVCID, Discovery or SHIVA.

1   Seshadri, S. & Wolf, P. A. Lifetime risk of stroke and dementia: current concepts, and estimates from the Framingham Study. *The Lancet Neurology* **6**, 1106-1114 (2007).
2   Cuadrado-Godia, E. *et al.* Cerebral Small Vessel Disease: A Review Focusing on Pathophysiology, Biomarkers, and Machine Learning Strategies. *Journal of Stroke* **20**, 302-320 (2018).
3   Wardlaw, J. M., Smith, C. & Dichgans, M. Mechanisms of sporadic cerebral small vessel disease: insights from neuroimaging. *The Lancet Neurology* **12**, 483-497 (2013).
4   Greenberg, S. M. Small Vessels, Big Problems. *New England Journal of Medicine* **354**, 1451-1453 (2006).
5   Feigin, V. L., Lawes, C. M. M., Bennett, D. A., Barker-Collo, S. L. & Parag, V. Worldwide stroke incidence and early case fatality reported in 56 population-based studies: a systematic review. *The Lancet Neurology* **8**, 355-369 (2009).
6   Mozaffarian, D. *et al.* Heart disease and stroke statistics--2015 update: a report from the American Heart Association. *Circulation* **131**, e29-322 (2015).
7   Alber, J. *et al.* White matter hyperintensities in vascular contributions to cognitive impairment and dementia (VCID): Knowledge gaps and opportunities. *Alzheimer's & Dementia (New York, N. Y.)* **5**, 107-117 (2019).
8   Gladman, J. T. *et al.* Vascular contributions to cognitive impairment and dementia: Research consortia that focus on etiology and treatable targets to lessen the burden of dementia worldwide. *Alzheimer's & Dementia (New York, N. Y.)* **5**, 789-796 (2019).
9   Viswanathan, A., Rocca, W. A. & Tzourio, C. Vascular risk factors and dementia: how to move forward? *Neurology* **72**, 368-374 (2009).
10   Kapasi, A., DeCarli, C. & Schneider, J. A. Impact of multiple pathologies on the threshold for clinically overt dementia. *Acta Neuropathologica* **134**, 171-186 (2017).

11   Bos, D. *et al.* Cerebral small vessel disease and the risk of dementia: A systematic review and meta-analysis of population-based evidence. *Alzheimer's & Dementia: The Journal of the Alzheimer's Association* **14**, 1482-1492 (2018).

12   Wardlaw, J. M., Smith, C. & Dichgans, M. Small vessel disease: mechanisms and clinical implications. *The Lancet Neurology* **18**, 684-696 (2019). **An in-depth overview of the current knowledge of mechanisms of small vessel disease.**

13   Wardlaw, J. M. *et al.* Neuroimaging standards for research into small vessel disease and its contribution to ageing and neurodegeneration. *Lancet Neurol* **12**, 822-838 (2013).

14   Debette, S. & Markus, H. S. The clinical importance of white matter hyperintensities on brain magnetic resonance imaging: systematic review and meta-analysis. *BMJ (Clinical research ed.)* **341**, c3666 (2010).

15   Maxwell, S. S. *et al.* Genetic associations with brain microbleeds: Systematic review and meta-analyses. *Neurology* **77**, 158-167 (2011).

16   Duperron, M.-G. *et al.* Burden of Dilated Perivascular Spaces, an Emerging Marker of Cerebral Small Vessel Disease, Is Highly Heritable. *Stroke* **49**, 282-287 (2018).

17   Dubost, F. *et al.* Enlarged perivascular spaces in brain MRI: Automated quantification in four regions. *NeuroImage* **185**, 534-544 (2019).

18   van Veluw, S. J. *et al.* Detection, risk factors, and functional consequences of cerebral microinfarcts. *The Lancet. Neurology* **16**, 730-740 (2017).

19   van Veluw, S. J. *et al.* Different microvascular alterations underlie microbleeds and microinfarcts. *Annals of Neurology* **86**, 279-292 (2019).

20   Baykara, E. *et al.* A Novel Imaging Marker for Small Vessel Disease Based on Skeletonization of White Matter Tracts and Diffusion Histograms: Novel SVD Imaging Marker. *Annals of Neurology* **80**, 581-592 (2016).

21   Persyn, E. *et al.* Genome-wide association study of MRI markers of cerebral small vessel disease in 42,310 participants. *Nature Communications* **11**, 2175 (2020).

22   Rutten-Jacobs, L. C. A. *et al.* Genetic Study of White Matter Integrity in UK Biobank (N=8448) and the Overlap With Stroke, Depression, and Dementia. *Stroke* **49**, 1340-1347 (2018).

23   Debette, S., Schilling, S., Duperron, M.-G., Larsson, S. C. & Markus, H. S. Clinical Significance of Magnetic Resonance Imaging Markers of Vascular Brain Injury: A Systematic Review and Meta-analysis. *JAMA neurology* **76**, 81-94 (2019).

24   Georgakis, M. K., Duering, M., Wardlaw, J. M. & Dichgans, M. WMH and long-term outcomes in ischemic stroke: A systematic review and meta-analysis. *Neurology* **92**, e1298-e1308 (2019).

25   Nelson, M. R. *et al.* The support of human genetic evidence for approved drug indications. *Nature Genetics* **47**, 856-860 (2015).

26   Okada, Y. *et al.* Genetics of rheumatoid arthritis contributes to biology and drug discovery. *Nature* **506**, 376-381 (2014).

27   Mancuso, M. *et al.* Monogenic cerebral small-vessel diseases: diagnosis and therapy. Consensus recommendations of the European Academy of Neurology. *Eur J Neurol* **27**, 909-927 (2020). **Updated recommendations for the management of monogenic cSVD.**

28   Marini, S., Anderson, C. D. & Rosand, J. Genetics of Cerebral Small Vessel Disease. *Stroke* **51**, 12-20 (2020).

29   Falcone, G. J., Malik, R., Dichgans, M. & Rosand, J. Current concepts and clinical applications of stroke genetics. *The Lancet Neurology* **13**, 405-418 (2014).

30   Traylor, M. *et al.* Genetic Architecture of Lacunar Stroke. *Stroke* **46**, 2407-2412 (2015).

31   Woo, D. *et al.* Meta-analysis of genome-wide association studies identifies 1q22 as a susceptibility locus for intracerebral hemorrhage. *American Journal of Human Genetics* **94**, 511-521 (2014).

32   Marini, S. *et al.* 17p12 Influences Hematoma Volume and Outcome in Spontaneous Intracerebral Hemorrhage. *Stroke* **49**, 1618-1625 (2018).

33   Chung, J. *et al.* Genome-wide association study of cerebral small vessel disease reveals established and novel loci. *Brain: A Journal of Neurology* **142**, 3176-3189 (2019).

34   Traylor, M. *et al.* Genetic basis of lacunar stroke: a pooled analysis of individual patient data and genome-wide association studies. *Lancet Neurol* **20**, 351–361 (2021).
The largest GWAS of lacunar stroke in which a multi-trait approach is used with white matter hyperintensities to discover additional lacunar stroke loci.

35   von Berg, J. *et al.* Alternate approach to stroke phenotyping identifies a genetic risk locus for small vessel stroke. *European journal of human genetics: EJHG* **28**, 963-972 (2020).

36   Malik, R. *et al.* Multiancestry genome-wide association study of 520,000 subjects identifies 32 loci associated with stroke and stroke subtypes. *Nature Genetics* **50**, 524-537 (2018).
The largest multi-ancestry GWAS meta-analysis of stroke and its subtypes; few loci were associated with small vessel stroke but follow-up bioinformatics analyses provided insight into the genetics of cSVD.

37   Traylor, M. *et al.* Genetic variation at 16q24.2 is associated with small vessel stroke. *Annals of Neurology* **81**, 383-394 (2017).

38   NINDS Stroke Genetics Network & International Stroke Genetics Consortium. Loci associated with ischaemic stroke and its subtypes (SiGN): a genome-wide association study. *Lancet Neurol* **15**, 174-184 (2016).

39   Rajajee, V. *et al.* Diagnosis of lacunar infarcts within 6 hours of onset by clinical and CT criteria versus MRI. *J Neuroimaging* **18**, 66-72 (2008).

40   Turley, P. *et al.* Multi-trait analysis of genome-wide association summary statistics using MTAG. *Nat Genet* **50**, 229-237 (2018).

41   Zhu, Z. *et al.* Integration of summary data from GWAS and eQTL studies predicts complex trait gene targets. *Nat Genet* **48**, 481-487 (2016).

42   Gusev, A. *et al.* Integrative approaches for large-scale transcriptome-wide association studies. *Nat Genet* **48**, 245-252 (2016).

43   Giambartolomei, C. *et al.* Bayesian test for colocalisation between pairs of genetic association studies using summary statistics. *PLoS Genet* **10**, e1004383 (2014).

44   Sargurupremraj, M., Suzuki, H., Jian, X., Fornage, M. & Debette, S. Cerebral small vessel disease genomics and its implications across the lifespan. *Nature Communications* (2020).
The largest GWAS of WMHs, accounting for interaction with hypertension, providing important insight into the lifetime impact of genetic risk for cSVD.

45   Bin, B. H. *et al.* Molecular pathogenesis of spondylocheirodysplastic Ehlers-Danlos syndrome caused by mutant ZIP13 proteins. *EMBO Mol Med* **6**, 1028-1042 (2014).

46   Ho, H. T., Dahlin, A. & Wang, J. Expression Profiling of Solute Carrier Gene Families at the Blood-CSF Barrier. *Front Pharmacol* **3**, 154 (2012).

47   Kusuhara, H. & Sugiyama, Y. Active efflux across the blood-brain barrier: role of the solute carrier family. *NeuroRx: The Journal of the American Society for Experimental NeuroTherapeutics* **2**, 73-85 (2005).

48   Geier, E. G. *et al.* Profiling solute carrier transporters in the human blood-brain barrier. *Clinical Pharmacology and Therapeutics* **94**, 636-639 (2013).

49   Wang, W. W., Gallo, L., Jadhav, A., Hawkins, R. & Parker, C. G. The Druggability of Solute Carriers. *Journal of Medicinal Chemistry* **63**, 3834-3867 (2020).

50   Liu, M., Xu, P., O'Brien, T. & Shen, S. Multiple roles of Ulk4 in neurogenesis and brain function. *Neurogenesis (Austin)* **4**, e1313646 (2017).

51   Liu, M. *et al.* Ulk4 deficiency leads to hypomyelination in mice. *Glia* **66**, 175-190 (2018).

52   Humphreys, C. A., Smith, C. & Wardlaw, J. M. Correlations in post-mortem imaging-histopathology studies of sporadic human cerebral small vessel disease: A systematic review. *Neuropathology and Applied Neurobiology*, https://doi.org/10.1111/nan.12737 (2021).

53   Rajani, R. M. *et al.* Characterisation of early ultrastructural changes in the cerebral white matter of CADASIL small vessel disease using high-pressure freezing/freeze-substitution. *Neuropathology and Applied Neurobiology* **47**, 694-704 (2021).

54   Hadano, S. *et al.* A gene encoding a putative GTPase regulator is mutated in familial amyotrophic lateral sclerosis 2. *Nature Genetics* **29**, 166-173 (2001).

55   Bindesbøll, C. *et al.* NBEAL1 controls SREBP2 processing and cholesterol metabolism and is a susceptibility locus for coronary artery disease. *Scientific Reports* **10**, 4528 (2020).

56   Biffi, A. *et al.* Variants at APOE influence risk of deep and lobar intracerebral hemorrhage. *Annals of Neurology* **68**, 934-943 (2010).

57   Rannikmäe, K. *et al.* Common variation in COL4A1/COL4A2 is associated with sporadic cerebral small vessel disease. *Neurology* **84**, 918-926 (2015).

58   Rannikmäe, K. *et al.* COL4A2 is associated with lacunar ischemic stroke and deep ICH: Meta-analyses among 21,500 cases and 40,600 controls. *Neurology* **89**, 1829-1839 (2017).

59   Pelletier, A. *et al.* Age-Related Modifications of Diffusion Tensor Imaging Parameters and White Matter Hyperintensities as Inter-Dependent Processes. *Front Aging Neurosci* **7**, 255 (2015).

60   Traylor, M. *et al.* Genome-wide meta-analysis of cerebral white matter hyperintensities in patients with stroke. *Neurology* **86**, 146-153 (2016).

61   Armstrong, N. J. *et al.* Common Genetic Variation Indicates Separate Causes for Periventricular and Deep White Matter Hyperintensities. *Stroke* **51**, 2111-2121 (2020).

62   Fornage, M. *et al.* Genome-wide association studies of cerebral white matter lesion burden: the CHARGE consortium. *Annals of Neurology* **69**, 928-939 (2011).

63   Verhaaren, B. F. J. *et al.* Multiethnic genome-wide association study of cerebral white matter hyperintensities on MRI. *Circulation. Cardiovascular Genetics* **8**, 398-409 (2015).

64   Knol, M. J. *et al.* Association of common genetic variants with brain microbleeds: A Genome-wide Association Study. *Neurology* (2020).

65   Chauhan, G. *et al.* Genetic and lifestyle risk factors for MRI-defined brain infarcts in a population-based setting. *Neurology* **95**, e3331–e3343 (2019).

66   Gonzalez, C. E. *et al.* Peripheral sphingolipids are associated with variation in white matter microstructure in older adults. *Neurobiology of Aging* **43**, 156-163 (2016).

67   Sun, N., Keep, R. F., Hua, Y. & Xi, G. Critical Role of the Sphingolipid Pathway in Stroke: a Review of Current Utility and Potential Therapeutic Targets. *Translational Stroke Research* **7**, 420-438 (2016).

68   Ohi, K. *et al.* DEGS2 polymorphism associated with cognition in schizophrenia is associated with gene expression in brain. *Transl Psychiatry* **5**, e550 (2015).

69   Wallis, M. *et al.* Surprisingly good outcome in antenatal diagnosis of severe hydrocephalus related to CCDC88C deficiency. *Eur J Med Genet* **61**, 189-196 (2018).

70   Jia, L., Piña-Crespo, J. & Li, Y. Restoring Wnt/β-catenin signaling is a promising therapeutic strategy for Alzheimer's disease. *Molecular Brain* **12**, 104 (2019).

71   Couffinhal, T., Dufourcq, P. & Duplàa, C. β-Catenin Nuclear Activation. *Circulation Research* **99**, 1287-1289 (2006).

72   Guo, F. *et al.* Canonical Wnt signaling in the oligodendroglial lineage--puzzles remain. *Glia* **63**, 1671-1693 (2015).

73   Zody, M. C. *et al.* Evolutionary toggling of the MAPT 17q21.31 inversion region. *Nature Genetics* **40**, 1076-1083 (2008).

74   Hyde, C. L. *et al.* Identification of 15 genetic loci associated with risk of major depression in individuals of European descent. *Nature Genetics* **48**, 1031-1036 (2016).

75   Ikeda, M. *et al.* Genome-Wide Association Study Detected Novel Susceptibility Genes for Schizophrenia and Shared Trans-Populations/Diseases Genetic Effect. *Schizophrenia Bulletin* **45**, 824-834 (2019).

76   Kouri, N. *et al.* Genome-wide association study of corticobasal degeneration identifies risk variants shared with progressive supranuclear palsy. *Nature Communications* **6**, 7247 (2015).

77   Kunkle, B. W. *et al.* Genetic meta-analysis of diagnosed Alzheimer's disease identifies new risk loci and implicates Aβ, tau, immunity and lipid processing. *Nature Genetics* **51**, 414-430 (2019).

78   Pain, O. *et al.* Novel Insight Into the Etiology of Autism Spectrum Disorder Gained by Integrating Expression Data With Genome-wide Association Statistics. *Biological Psychiatry* **86**, 265-273 (2019).

79   Simón-Sánchez, J. *et al.* Genome-wide association study reveals genetic risk underlying Parkinson's disease. *Nature Genetics* **41**, 1308-1312 (2009).

80   Davies, G. *et al.* Study of 300,486 individuals identifies 148 independent genetic loci influencing general cognitive function. *Nature Communications* **9**, 2098 (2018).

81   Koolen, D. A. *et al.* A new chromosome 17q21.31 microdeletion syndrome associated with a common inversion polymorphism. *Nature Genetics* **38**, 999-1001 (2006).

82   Lee, J. J. *et al.* Gene discovery and polygenic prediction from a genome-wide association study of educational attainment in 1.1 million individuals. *Nature Genetics* **50**, 1112-1121 (2018).

83   Sanchez-Roige, S. *et al.* Genome-Wide Association Study Meta-Analysis of the Alcohol Use Disorders Identification Test (AUDIT) in Two Population-Based Cohorts. *The American Journal of Psychiatry* **176**, 107-118 (2019).

84   Ikram, M. A. *et al.* Common variants at 6q22 and 17q21 are associated with intracranial volume. *Nature Genetics* **44**, 539-544 (2012).

85   Jonsson, B. A. *et al.* Brain age prediction using deep learning uncovers associated sequence variants. *Nature Communications* **10**, 5409 (2019).

86   Zhao, B. *et al.* Common genetic variation influencing human white matter microstructure. *Science (New York, N.Y.)* **372**, eabf3736 (2021).

87   Barbu, M. C. *et al.* Expression quantitative trait loci-derived scores and white matter microstructure in UK Biobank: a novel approach to integrating genetics and neuroimaging. *Transl Psychiatry* **10**, 55 (2020).
      Analysis of unexplored relationships between gene expression and heritable in vivo measures of human brain structural connectivity by combining expression quantitative trait loci and genome-wide genotype data.

88   Boettger, L. M., Handsaker, R. E., Zody, M. C. & McCarroll, S. A. Structural haplotypes and recent evolution of the human 17q21.31 region. *Nature Genetics* **44**, 881-885 (2012).

89   Steinberg, K. M. *et al.* Structural diversity and African origin of the 17q21.31 inversion polymorphism. *Nature Genetics* **44**, 872-880 (2012).

90   Tabara, Y. *et al.* Association of Chr17q25 with cerebral white matter hyperintensities and cognitive impairment: the J-SHIPP study. *European Journal of Neurology* **20**, 860-862 (2013).

91   Verhaaren, B. F. J. *et al.* Replication study of chr17q25 with cerebral white matter lesion volume. *Stroke* **42**, 3297-3299 (2011).

92   Adib-Samii, P. *et al.* 17q25 Locus is associated with white matter hyperintensity volume in ischemic stroke, but not with lacunar stroke status. *Stroke* **44**, 1609-1615 (2013).

93   Jian, X. *et al.* Exome Chip Analysis Identifies Low-Frequency and Rare Variants in MRPL38 for White Matter Hyperintensities on Brain Magnetic Resonance Imaging. *Stroke* **49**, 1812-1819 (2018).

94   Hara, K. *et al.* Association of HTRA1 mutations and familial ischemic cerebral small-vessel disease. *N Engl J Med* **360**, 1729-1739 (2009).

95   Dabertrand, F. *et al.* Potassium channelopathy-like defect underlies early-stage cerebrovascular dysfunction in a genetic model of small vessel disease. *Proceedings of the National Academy of Sciences of the United States of America* **112**, E796-805 (2015).

96   Joutel, A., Haddad, I., Ratelade, J. & Nelson, M. T. Perturbations of the cerebrovascular matrisome: A convergent mechanism in small vessel disease of the brain? *J Cereb Blood Flow Metab* **36**, 143-157 (2016).
      A description of the concept that cerebrovascular matrisome perturbations are a convergent pathological pathway in monogenic SVD and are likely to be relevant in cSVD.

97   Naba, A. *et al.* The matrisome: in silico definition and in vivo characterization by proteomics of normal and tumor extracellular matrices. *Molecular & cellular proteomics: MCP* **11**, M111.014647 (2012).

98   Mehta, P. & Piao, X. Adhesion G-protein coupled receptors and extracellular matrix proteins: Roles in myelination and glial cell development. *Developmental Dynamics: An Official Publication of the American Association of Anatomists* **246**, 275-284 (2017).

99   Monk, K. R., Oshima, K., Jörs, S., Heller, S. & Talbot, W. S. Gpr126 is essential for peripheral nerve development and myelination in mammals. *Development (Cambridge, England)* **138**, 2673-2680 (2011).

100   Chang, A. *et al.* Cortical remyelination: a new target for repair therapies in multiple sclerosis. *Annals of Neurology* **72**, 918-926 (2012).

101   Diamantopoulou, E. *et al.* Identification of compounds that rescue otic and myelination defects in the zebrafish adgrg6 (gpr126) mutant. *eLife* **8**, e44889 (2019).

102   Paavola, K. J., Sidik, H., Zuchero, J. B., Eckart, M. & Talbot, W. S. Type IV collagen is an activating ligand for the adhesion G protein-coupled receptor GPR126. *Science Signaling* **7**, ra76 (2014).

103   Rajani, R. M. *et al.* Reversal of endothelial dysfunction reduces white matter vulnerability in cerebral small vessel disease in rats. *Science Translational Medicine* **10**, eaam9507 (2018).

104   Corder, E. H. *et al.* Protective effect of apolipoprotein E type 2 allele for late onset Alzheimer disease. *Nat Genet* **7**, 180-184 (1994).

105   Jun, G. R. *et al.* Transethnic genome-wide scan identifies novel Alzheimer's disease loci. *Alzheimers Dement* **13**, 727-738 (2017).

106   Maddison, D. C. & Giorgini, F. The kynurenine pathway and neurodegenerative disease. *Semin Cell Dev Biol* **40**, 134-141 (2015).

107   Strang, K. H., Golde, T. E. & Giasson, B. I. MAPT mutations, tauopathy, and mechanisms of neurodegeneration. *Laboratory Investigation; a Journal of Technical Methods and Pathology* **99**, 912-928 (2019).

108   Harold, D. *et al.* Interaction between the ADAM12 and SH3MD1 genes may confer susceptibility to late-onset Alzheimer's disease. *Am J Med Genet B Neuropsychiatr Genet* **144B**, 448-452 (2007).

109   Okbay, A. *et al.* Genetic variants associated with subjective well-being, depressive symptoms, and neuroticism identified through genome-wide analyses. *Nature Genetics* **48**, 624-633 (2016).

110   Wu, Y. *et al.* Identification of the primate-specific gene BTN3A2 as an additional schizophrenia risk gene in the MHC loci. *EBioMedicine* **44**, 530-541 (2019).

111   Schizophrenia Working Group of the Psychiatric Genomics Consortium. Biological insights from 108 schizophrenia-associated genetic loci. *Nature* **511**, 421-427 (2014).

112   Larsson, S. C. *et al.* Serum magnesium and calcium levels in relation to ischemic stroke: Mendelian randomization study. *Neurology* **92**, e944-e950 (2019).

113   Evangelou, E. *et al.* Genetic analysis of over 1 million people identifies 535 new loci associated with blood pressure traits. *Nat Genet* **50**, 1412-1425 (2018).

114   Xue, A. *et al.* Genome-wide association analyses identify 143 risk variants and putative regulatory mechanisms for type 2 diabetes. *Nat Commun* **9**, 2941 (2018).

115   Willer, C. J. *et al.* Discovery and refinement of loci associated with lipid levels. *Nat Genet* **45**, 1274-1283 (2013).

116   Wootton, R. E. *et al.* Evidence for causal effects of lifetime smoking on risk for depression and schizophrenia: a Mendelian randomisation study. *Psychol Med* **50**, 2435-2443 (2020).

117   Liberzon, A. *et al.* Molecular signatures database (MSigDB) 3.0. *Bioinformatics (Oxford, England)* **27**, 1739-1740 (2011).

118   Holliday, E. G. *et al.* Genetic overlap between diagnostic subtypes of ischemic stroke. *Stroke* **46**, 615-619 (2015).

119 Mishra, A. *et al.* Association of variants in HTRA1 and NOTCH3 with MRI-defined extremes of cerebral small vessel disease in older subjects. *Brain: A Journal of Neurology* **142**, 1009-1023 (2019).

In this study , whole-exome sequencing in a population-based setting provided evidence that a composite, extreme MRI-derived phenotype for cSVD is useful for identification of rare variants, adding to evidence that *NOTCH3* mutations that cause CADASIL are more common than previously suspected.

120 Narayan, S. K., Gorman, G., Kalaria, R. N., Ford, G. A. & Chinnery, P. F. The minimum prevalence of CADASIL in northeast England. *Neurology* **78**, 1025-1027 (2012).

121 Razvi, S. S. M., Davidson, R., Bone, I. & Muir, K. W. The prevalence of cerebral autosomal dominant arteriopathy with subcortical infarcts and leucoencephalopathy (CADASIL) in the west of Scotland. *Journal of Neurology, Neurosurgery, and Psychiatry* **76**, 739-741 (2005).

122 Verdura, E. *et al.* Heterozygous HTRA1 mutations are associated with autosomal dominant cerebral small vessel disease. *Brain* **138**, 2347-2358 (2015).

123 Siitonen, M. *et al.* Multi-infarct dementia of Swedish type is caused by a 3'UTR mutation of COL4A1. *Brain: A Journal of Neurology* **140**, e29 (2017).

124 Lemmens, R. *et al.* Novel COL4A1 mutations cause cerebral small vessel disease by haploinsufficiency. *Human Molecular Genetics* **22**, 391-397 (2013).

125 Li, J. et al. Replication of top loci from COL4A1/2 associated with white matter hyperintensity burden in patients with ischemic stroke. *Stroke* **51**, 3751–3755(2020).

126 Malik, R. *et al.* Genome-wide meta-analysis identifies 3 novel loci associated with stroke: MEGASTROKE and UK Biobank GWAS. *Annals of Neurology* **84**, 934-939 (2018).

127 Simon, A. J. *et al.* Mutations in STN1 cause Coats plus syndrome and are associated with genomic and telomere defects. *J Exp Med* **213**, 1429-1440 (2016).

128 Schmidt, H. *et al.* Genetic variants of the NOTCH3 gene in the elderly and magnetic resonance imaging correlates of age-related cerebral small vessel disease. *Brain: A Journal of Neurology* **134**, 3384-3397 (2011).

129 Rutten, J. W. *et al.* Archetypal NOTCH3 mutations frequent in public exome: implications for CADASIL. *Ann Clin Transl Neurol* **3**, 844-853 (2016).

130 Rutten, J. W. *et al.* Broad phenotype of cysteine-altering NOTCH3 variants in UK Biobank: CADASIL to nonpenetrance. *Neurology* **95**, e1835-e1843 (2020).

131 Cho, B. P. H. *et al.* NOTCH3 variants are more common than expected in the general population and associated with stroke and vascular dementia: an analysis of 200 000 participants. *Journal of Neurology, Neurosurgery, and Psychiatry* **92**, 694-701 (2021).).

Using UK Biobank data, this study provides evidence that NOTCH3 mutations that cause CADASIL are more common than expected in the general population and describes their phenotypic spectrum.

132 Chauhan, G. *et al.* Identification of additional risk loci for stroke and small vessel disease: a meta-analysis of genome-wide association studies. *The Lancet Neurology* **15**, 695-707 (2016).

This study identified a novel risk locus for any stroke and subclinical cerebral small vessel disease in a population-based setting, with experimental follow-up in mice and zebrafish implicating *FOXF2*.

133 French, C. R. *et al.* Mutation of FOXC1 and PITX2 induces cerebral small-vessel disease. *The Journal of Clinical Investigation* **124**, 4877-4881 (2014).

134 Whitesell, T. R. *et al.* foxc1 is required for embryonic head vascular smooth muscle differentiation in zebrafish. *Developmental Biology* **453**, 34-47 (2019).

135 Siegenthaler, J. A. *et al.* Foxc1 is required by pericytes during fetal brain angiogenesis. *Biology Open* **2**, 647-659 (2013).

136 Reyahi, A. *et al.* Foxf2 Is Required for Brain Pericyte Differentiation and Development and Maintenance of the Blood-Brain Barrier. *Developmental Cell* **34**, 19-32 (2015).

137 He, W. *et al.* FOXF2 acts as a crucial molecule in tumours and embryonic development. *Cell Death & Disease* **11**, 424 (2020).

138 Smith, G. D. & Ebrahim, S. 'Mendelian randomization': can genetic epidemiology contribute to understanding environmental determinants of disease? *Int J Epidemiol* **32**, 1-22 (2003).

139 Georgakis, M. K. *et al.* Genetically determined blood pressure, antihypertensive drug classes, and risk of stroke subtypes. *Neurology* **95**, e353-e361 (2020).

140 Yuan, S., Tang, B., Zheng, J. & Larsson, S. C. Circulating Lipoprotein Lipids, Apolipoproteins and Ischemic Stroke. *Ann Neurol* **88**, 1229-1236 (2020).

141 Georgakis, M. K. *et al.* Genetic determinants of blood lipids and cerebral small vessel disease: role of high-density lipoprotein cholesterol. *Brain* **143**, 597-610 (2020).

142 Falcone, G. J. *et al.* Genetically Elevated LDL Associates with Lower Risk of Intracerebral Hemorrhage. *Ann Neurol* **88**, 56-66 (2020).

143 Valdes-Marquez, E. *et al.* Relative effects of LDL-C on ischemic stroke and coronary disease: A Mendelian randomization study. *Neurology* **92**, e1176-e1187 (2019).

144 Pan, Y., Li, H., Wang, Y. & Meng, X. Causal Effect of Lp(a) [Lipoprotein(a)] Level on Ischemic Stroke and Alzheimer Disease: A Mendelian Randomization Study. *Stroke* **50**, 3532-3539 (2019).

145 Georgakis, M. K. *et al.* Diabetes Mellitus, Glycemic Traits, and Cerebrovascular Disease: A Mendelian Randomization Study. *Neurology* **96**, e1732–e1742 (2021).

146 Larsson, S. C. *et al.* Type 2 diabetes, glucose, insulin, BMI, and ischemic stroke subtypes: Mendelian randomization study. *Neurology* **89**, 454-460 (2017).

147 Liu, J., Rutten-Jacobs, L., Liu, M., Markus, H. S. & Traylor, M. Causal Impact of Type 2 Diabetes Mellitus on Cerebral Small Vessel Disease: A Mendelian Randomization Analysis. *Stroke* **49**, 1325-1331 (2018).

148 Dale, C. E. *et al.* Causal Associations of Adiposity and Body Fat Distribution With Coronary Heart Disease, Stroke Subtypes, and Type 2 Diabetes Mellitus: A Mendelian Randomization Analysis. *Circulation* **135**, 2373-2388 (2017).

149 Marini, S. *et al.* Mendelian Randomization Study of Obesity and Cerebrovascular Disease. *Ann Neurol* **87**, 516-524 (2020).

150 Qian, Y. *et al.* Coffee Consumption and Risk of Stroke: A Mendelian Randomization Study. *Ann Neurol* **87**, 525-532 (2020).

151 Wang, M. *et al.* Higher tea consumption is associated with decreased risk of small vessel stroke. *Clin Nutr* **40**, 1430-1435 (2021).

152 Qian, Y. *et al.* Role of cigarette smoking in the development of ischemic stroke and its subtypes: a Mendelian randomization study. *Clin Epidemiol* **11**, 725-731 (2019).

153 Larsson, S. C., Burgess, S. & Michaëlsson, K. Smoking and stroke: A mendelian randomization study. *Ann Neurol* **86**, 468-471 (2019).

154 Harshfield, E. L., Georgakis, M. K., Malik, R., Dichgans, M. & Markus, H. S. Modifiable Lifestyle Factors and Risk of Stroke: A Mendelian Randomization Analysis. *Stroke* **52**, 931-936 (2021).

155 Wang, T. *et al.* Birth Weight and Stroke in Adult Life: Genetic Correlation and Causal Inference With Genome-Wide Association Data Sets. *Front Neurosci* **14**, 479 (2020).

156 Cai, H. *et al.* Major depression and small vessel stroke: a Mendelian randomization analysis. *J Neurol* **266**, 2859-2866 (2019).

157 Maners, J. *et al.* A Mendelian randomization of γ' and total fibrinogen levels in relation to venous thromboembolism and ischemic stroke. *Blood* **136**, 3062-3069 (2020).

158 Georgakis, M. K. *et al.* Genetically Determined Levels of Circulating Cytokines and Risk of Stroke. *Circulation* **139**, 256-268 (2019).
This study demonstrates how Mendelian randomization can be used to provide evidence for causal involvement of circulating biomarkers for stroke (MCP-1), leveraging summary data from GWAS.

159 Lin, J., Wang, Y. & Pan, Y. Inflammatory biomarkers and risk of ischemic stroke and subtypes: A 2-sample Mendelian randomization study. *Neurol Res* **42**, 118-125 (2020).

160   Yuan, S., Lin, A., He, Q. Q., Burgess, S. & Larsson, S. C. Circulating interleukins in relation to coronary artery disease, atrial fibrillation and ischemic stroke and its subtypes: A two-sample Mendelian randomization study. *Int J Cardiol* **313**, 99-104 (2020).

161   Larsson, S. C., Traylor, M. & Markus, H. S. Homocysteine and small vessel stroke: A mendelian randomization analysis. *Ann Neurol* **85**, 495-501 (2019).

162   Mulvey, B., Lagunas, T. & Dougherty, J. D. Massively Parallel Reporter Assays: Defining Functional Psychiatric Genetic Variants Across Biological Contexts. *Biological Psychiatry* **89**, 76-89 (2021).

163   Sanseau, P. *et al.* Use of genome-wide association studies for drug repositioning. *Nature Biotechnology* **30**, 317-320 (2012).

164   Gill, D. *et al.* Use of Genetic Variants Related to Antihypertensive Drugs to Inform on Efficacy and Side Effects. *Circulation* **140**, 270-279 (2019).
In this study, genetic variants were used to explore the efficacy and side effects of different antihypertensive drug classes for stroke prevention, providing proof-of-concept that genomic data can provide genetic support for drug effects.

165   Wardlaw, J. M. *et al.* ESO Guideline on covert cerebral small vessel disease. *European Stroke Journal* **6**, CXI-CLXII (2021).

166   Abraham, G. & Inouye, M. Genomic risk prediction of complex human disease and its clinical application. *Curr Opin Genet Dev* **33**, 10-16 (2015).

167   Habes, M. *et al.* The Brain Chart of Aging: Machine-learning analytics reveals links between brain aging, white matter disease, amyloid burden, and cognition in the iSTAGING consortium of 10,216 harmonized MR scans. *Alzheimers Dement* **17**, 89-102 (2021).

168   Abraham, G. *et al.* Genomic prediction of coronary heart disease. *Eur Heart J* **37**, 3267-3278 (2016).

169   Escott-Price, V., Shoai, M., Pither, R., Williams, J. & Hardy, J. Polygenic score prediction captures nearly all common genetic risk for Alzheimer's disease. *Neurobiol Aging* **49**, 214 e217-214 e211 (2017).

170   Lambert, S. A., Abraham, G. & Inouye, M. Towards clinical utility of polygenic risk scores. *Hum Mol Genet* **28**, R133-R142 (2019).

171   Khera, A. V. *et al.* Genome-wide polygenic scores for common diseases identify individuals with risk equivalent to monogenic mutations. *Nat Genet* **50**, 1219-1224 (2018).

172   Abraham, G. *et al.* Genomic risk score offers predictive performance comparable to clinical risk factors for ischaemic stroke. *Nature Communications* **10**, 5819 (2019).
Abraham et al. developed a new polygenic risk score (PRS) approach based on the combination of PRS from various traits (metaGRS) and efficiently applied it to predict ischaemic stroke in the UK biobank.

173   Khera, A. V. *et al.* Genetic Risk, Adherence to a Healthy Lifestyle, and Coronary Disease. *N Engl J Med* **375**, 2349-2358 (2016).

174   Mok, V. *et al.* Race-ethnicity and cerebral small vessel disease--comparison between Chinese and White populations. *International journal of stroke : official journal of the International Stroke Society* **9 Suppl A100**, 36-42 (2014).

175   Keene, K. L. *et al.* Genome-Wide Association Study Meta-Analysis of Stroke in 22 000 Individuals of African Descent Identifies Novel Associations With Stroke. *Stroke* **51**, 2454-2463 (2020).

176   Carty, C. L. *et al.* Meta-Analysis of Genome-Wide Association Studies Identifies Genetic Risk Factors for Stroke in African Americans. *Stroke* **46**, 2063-2068 (2015).

177   Márquez-Luna, C., Loh, P. R. & Price, A. L. Multiethnic polygenic risk scores improve risk prediction in diverse populations. *Genet Epidemiol* **41**, 811-823 (2017).

178   Montaner, J. *et al.* Multilevel omics for the discovery of biomarkers and therapeutic targets for stroke. *Nature Reviews. Neurology* **16**, 247-264 (2020).

179   Woo, D. *et al.* Top research priorities for stroke genetics. *The Lancet. Neurology* **17**, 663-665 (2018).

180   Adams, H. P., Jr. *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* **24**, 35-41 (1993).

181   Ay, H. *et al.* A computerized algorithm for etiologic classification of ischemic stroke: the Causative Classification of Stroke System. *Stroke* **38**, 2979-2984 (2007).

182   Lin, M. K. *et al.* HTRA1, an age-related macular degeneration protease, processes extracellular matrix proteins EFEMP1 and TSP1. *Aging Cell* **17**, e12710 (2018).

183   Zagryazhskaya-Masson, A. *et al.* Intersection of TKS5 and FGD1/CDC42 signaling cascades directs the formation of invadopodia. *The Journal of Cell Biology* **219**, e201910132 (2020).

184   Leonardo, C. C., Eakin, A. K., Ajmo, J. M. & Gottschall, P. E. Versican and brevican are expressed with distinct pathology in neonatal hypoxic-ischemic injury. *J Neurosci Res* **86**, 1106-1114 (2008).

185   Menezes, M. J. *et al.* The extracellular matrix protein laminin α2 regulates the maturation and function of the blood-brain barrier. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience* **34**, 15260-15280 (2014).

186   Fukada, T. *et al.* The zinc transporter SLC39A13/ZIP13 is required for connective tissue development; its involvement in BMP/TGF-beta signaling pathways. *PloS One* **3**, e3642 (2008).

187   Yoneshiro, T. *et al.* BCAA catabolism in brown fat controls energy homeostasis through SLC25A44. *Nature* **572**, 614-619 (2019).

188   Bowden, J. & Holmes, M. V. Meta-analysis and Mendelian randomization: A review. *Res Synth Methods* **10**, 486-496 (2019).

189   Burgess, S. & Thompson, S. G. Avoiding bias from weak instruments in Mendelian randomization studies. *Int J Epidemiol* **40**, 755-764 (2011).

190   Hemani, G., Bowden, J. & Davey Smith, G. Evaluating the potential role of pleiotropy in Mendelian randomization studies. *Hum Mol Genet* **27**, R195-R208 (2018).

Acknowledgements

Author contributions

C.B. researched data for the article. All authors made substantial contributions to discussion of the content, wrote the article and edited and/or reviewed the manuscript before submission.

Competing interests

Peer review information

Review criteria

For our literature review, we searched PubMed for papers published between January 2007 and March 2021. We focused on this period because we chose to focus on genome-wide association studies (GWAS), which were not available at earlier timepoints in the field of cerebral small vessel disease (cSVD). We considered only genetic associations with common cSVD phenotypes that reached genome-wide significance ($P < 5 \times 10^{-8}$). When several genome-wide significant associations on the same locus were reported in several GWAS, we present the lead single-nucleotide polymorphism (SNP) and association statistics of the GWAS performed on the largest sample size (largest number of cases for binary traits) but still report references of the other publications. Where several genome-wide significant associations with different SNPs have been identified in the same locus, we report only those that correspond to independent signals (linkage disequilibrium $r^2 < 0.1$); for correlated SNPs ($r^2 \geq 0.1$), we selected the SNP with the lowest P value (if reported in the same study) or that from the GWAS performed on the largest sample size (if reported in two different studies). Results of candidate gene association studies were included if the studies were large ($n > 500$), robust methodology was used and the findings complement or support GWAS findings. Only peer-reviewed papers in English were considered. For Mendelian randomization studies, we searched the literature for publications that relate to cSVD phenotypes and include Mendelian randomization in the title or abstract. We also systematically screened all published GWAS of cSVD-related phenotypes for Mendelian randomization analyses.

Related links

COSTREAM http://www.costream.eu/

MarkVCID https://markvcid.partners.org/

Discovery https://discoverystudy.org/

SHIVA https://rhu-shiva.com/

Key points

- Fifty-two independent genetic loci have been associated with cerebral small vessel disease (cSVD) at the genome-wide significance level, including loci associated with cSVD-related stroke and loci associated with covert, MRI-defined cSVD.

- In silico functional explorations of the observed genetic associations point to a major role of blood pressure-related pathways and mechanisms independent of vascular risk factors, such as extracellular matrix structure and function.

- Transcriptome-wide association studies have provided evidence for associations between one or several genes at a cSVD risk locus and the corresponding cSVD traits, enabling prioritization of putative causal genes for functional follow-up.

- Mendelian randomization studies have been conducted to investigate the causal link between various factors and cSVD-related phenotypes; a dedicated systematic review and meta-analysis is required to confirm some causal relationships.

- Preliminary results provide proof-of-concept that cSVD genomics can inform therapeutic strategies by providing genetic evidence for drug effects, indicating pathways with therapeutic relevance and revealing potential for drug repositioning.

- Use of high-throughput molecular approaches, such as epigenomics, transcriptomics, proteomics and metabolomics, will enable integration of genetic associations with functional data to decipher the biological roles of genetic risk loci in cSVD.

Table 1 | Risk loci associated with cerebral small vessel disease-related stroke at the genome-wide significance level

Lead variant	Chromosome position	Risk allele	RAF	Nearest gene(s)	Population	Phenotype	Main ref[a]	Other ref[a]
ICH								
rs2984613	1q22	C	0.68	*SLC25A44* *PMF1* *BGLAP*	EUR	Non-lobar ICH	[31]	[33]
rs11655160	17p12[b]	G	0.76	*PIRT (intergenic)*	EUR	Haematoma volume of non-lobar ICH	[32]	NA
rs5117	19q13	C	NA	*APOE*	EUR	Lobar ICH	[33]	NA
rs9614326	22q13[b]	A	0.97	*PARVB*	EUR	Haematoma volume of lobar ICH	[32]	NA
ICH accounting for SVS, multivariate analysis								
rs72932727[c]	2q33	G	0.87	*ICA1L* *WDR12* *CARF* *NBEAL1*	EUR	Non-lobar ICH accounting for SVS[d]	[33]	NA
rs9515201	13q34	A	0.33	*COL4A1* *COL4A2*	EUR	Non-lobar ICH accounting for SVS[d]	[33]	NA
SVS								
rs72934535[c]	2q33	T	0.89	*ICA1L* *FAM117B* *CARF* *NBEAL1*	EUR	SVS[e,f]	[34]	NA
rs4621303	3p22	T	0.83	*ULK4*	TRANS	SVS[e,f]	[34]	NA
rs10029218	4q26[b]	A	0.12	*CAMK2D*	EUR	SVS[e,g]	[35]	NA
rs2293576	11p11	G	0.67	*SPI1* *SLC39A13* *PSMC3* *RAPSN*	EUR, TRANS	SVS[e,f]	[34]	NA
rs10744777	12q24	T	0.66	*SH2B3*	EUR	SVS[g]	[38]	[35]
rs12445022	16q24[b]	A	0.37	*ZCCHC14*	EUR	SVS[e,f]	[34]	[36,37]
rs9958650	18p11[b]	G	0.10	*ZBTB14* *EPB41L3*	TRANS	SVS[e,f]	[34]	NA
SVS accounting for WMH burden, multivariate analysis								
rs2984613	1q22	C	0.64	*SLC25A44* *PMF1* *BGLAP*	EUR	SVS[e,f]	[34]	NA
rs2303655	5q23[b]	T	0.81	*LOX* *ZNF474* *LOC100505841*	EUR	SVS[e,f]	[34]	NA
rs7766042	6p25[b]	C	0.11	*FOXF2* *FOXQ1*	EUR	SVS[e,f]	[34]	NA
rs225744	6q24[b]	C	0.77	*VTA1-GPR126*	EUR	SVS[e,f]	[34]	NA
rs61000833	10q24	T	0.60	*SH3PXD2A* *STN1*	EUR	SVS[e,f]	[34]	NA
rs79043147	10q26[b]	T	0.07	*HTRA1*	EUR	SVS[e,f]	[34]	NA
rs11838776	13q34	A	0.29	*COL4A1* *COL4A2*	EUR	SVS[e,f]	[34]	NA

[a]When several GWAS reported a genome-wide significant association with the same locus, we present the results of the GWAS performed on the largest sample size (largest number of cases), referred to as "main ref"; other studies reporting a genome-wide significant association with the same locus are listed under "other ref". [b]Not associated with vascular risk

factors. [c]rs72932727 (ICH) and rs72934535 (SVS) are in linkage disequilibrium (r^2 = 0.8317). [d]Multitrait analysis of non-lobar ICH and SVS. [e]SVS was defined according to Trial of Org 10172 Acute Stroke Treatment (TOAST)[180], which is based on a clinical lacunar syndrome (lacunar infarct ≤15 mm in diameter) and the absence of other causes of stroke, or non-lacunar infarction on CT. [f]SVS was defined as a clinical lacunar syndrome with an anatomically compatible lesion on MRI (subcortical infarct ≤15 mm in diameter). [g]SVS was defined according to Causative Classification of Stroke[181], which is based on imaging evidence of a clinically relevant acute infarction (≤20 mm in diameter) and the absence of other causes of stroke. Von Berg et al.[35] used an intersect phenotype (an individual must be assigned the same subtype across all subtyping systems). Traylor et al.[34] re-analyzed cases of lacunar stroke, identified with the TOAST system, with MRI confirmation. EUR, European; RAF, relative allele frequency; TRANS, transethnic; ICH, intracerebral haemorrhage; SVS, small vessel stroke; WMH, white matter hyperintensity.

Table 2 | Loci associated with MRI markers of covert cerebral small vessel disease at the genome-wide significance level

Lead variant[a]	Chromosome position	Risk allele	RAF	Nearest gene(s)	Population	Main ref[b]	Other ref[b]
Brain infarcts							
rs39938[c]	5q23[d]	T	0.21	*FBN2*	EUR	65	NA
rs12583648	13q12[d]	C	0.33	*ZDHHC20*	EUR	65	NA
Cerebral microbleeds							
rs769449	19q13	A	0.13	*APOE*	EUR	64	NA
Fractional anisotropy							
rs76122535	2q33	G	0.13	*ICA1L, CARF, NBEAL1*	EUR	21	NA
rs34380167	2q35[d]	ID	0.27	*SMARCAL1*	EUR	21	NA
rs35544841	5q14[d]	ID	0.20	*VCAN*	EUR	21	NA
rs4150221	5q31	C	0.26	*HBEGF*	EUR	21	NA
rs3129171[e]	6p22[d]	A	0.24	*ZFP57, OR2J2, OR2H4P*	EUR	21	NA
rs374598428[e]	6p22[d]	ID	0.14	*VN1R12P*	EUR	21	NA
rs112730611	16q24[d]	T	0.17	*SPIRE2, TCF25*	EUR	21	NA
rs6062264	20q13	T	0.28	*MIR1, 1HG*	EUR	21	NA
Mean diffusivity							
rs35544841	5q14[d]	ID	0.20	*VCAN*	EUR	21	NA
rs4150221	5q31	C	0.26	*HBEGF*	EUR	21	NA
rs7772614	6p21[d]	A	0.38	*HLA-B, HLA-S*	EUR	21	NA
rs1233587	6p22[d]	T	0.30	*ZFP57*	EUR	21	NA
rs11813268	10q24	T	0.15	*SH3PXD2A, STN1, PDCD11*	EUR	21	NA
rs55939347	17q21	ID	0.22	*MAPT, DCAKD, NMT1, NT5C3B, STAT5B*	EUR	21	NA
White matter hyperintensities							
rs786921	1p22	A	0.60	*PKN2*	EUR	44	NA
rs2984613	1q22	C	0.65	*SLC25A44, PMF1, BGLAP*	TRANS	63	NA
rs12120143	1q31[d]	T	0.03	*DENND1B*	EUR	21	NA
rs6540873	1q41[d]	A	0.61	*KCNK2*	TRANS	44	NA
rs7596872	2p16	A	0.10	*EFEMP1*	EUR	44	22,63,44,60,21,61
rs73923006	2p21	G	0.81	*HAAO*	EUR/TRANS	44	21,61
rs62172472	2q32	G	0.79	*CALCRL*[f]	EUR/TRANS	44	21
rs7603972	2q33	A	0.87	*ICA1L*[f]*, CARF*[f]*, NBEAL1*[f]*, KRT8P15*[f]*,*	EUR/TRANS	44	21,61,60
rs6797002	3q27	C	0.73	*KLHL24*[f]	EUR/TRANS	44	21
rs17205972	5q14[d]	T	0.20	*VCAN*	EUR/TRANS	44	NA
rs2303655[c]	5q23[d]	T	0.78	*FBN2*	EUR/TRANS	44	21
rs6940540	6q25	G	0.41	*PLEKHG1*	EUR/TRANS	44	61
rs3215395	7q22[d]	ID	0.29	*ZBN*	EUR	21	NA
rs1799983	7q36	T	0.32	*NOS3*	EUR	61	NA
rs73184312[g]	8p23	G	0.72	*TNKS, XKR6, SGK223*	EUR/TRANS	44	NA
rs11249945[g]	8p23	A	0.35	*TNKS, XKR6, SGK223*	EUR/TRANS	44	NA
rs11257311	10p14[d]	G	0.70	*ECHDC3*	AFR/TRANS	44	NA
rs1993484	10q23	T	0.23	*TSPAN14*	EUR	61	NA
rs10786772[h]	10q24	G	0.64	*SH3PXD2A, STN1, PDCD11*	EUR/TRANS	44	63,60
rs71471298[h]	10q24	T	0.11	*SH3PXD2A, STN1, PDCD11*	EUR/TRANS	44	
rs4630220[h]	10q24	G	0.71	*SH3PXD2A, STN1, PDCD11*	EUR/TRANS	44	
rs55940034	13q34	G	0.29	*COL4A1, COL4A2*	EUR/TRANS	44	21,61,60
rs72680374	14q22[d]	T	0.63	*NID2*	EUR	44	NA
rs1285847[i]	14q32	T	0.55	*EVL, DEGS2*[f]*, CCDC88C*[f]	EUR/TRANS	44	21,60
rs7157599[i]	14q32	C	0.29	*EVL, DEGS2*[f]*, CCDC88C*[f]	EUR/TRANS	44	
rs12443113	15q22	G	0.55	*MTFMT, SLC51B, RASL12*	EUR	44	21
rs1948948	16q12[d]	C	0.56	*SALL1*	EUR/TRANS	44	21
rs12921170	16q24[d]	A	0.58	*SPIRE2, TCF25*	EUR/TRANS	44	21,61
rs6587216	17p11	G	0.19	*EPN2*	EUR	21	NA
rs6503417[j]	17q21	C	0.63	*MAPT, DCAKD*[f]*, NMT1*[f]*, NT5C3B*[f]*, STAT5B*[f]	EUR/TRANS	44	21,61
rs962888[j]	17q21	G	0.71	*MAPT, DCAKD*[f]*, NMT1*[f]*, NT5C3B*[f]*, STAT5B*[f]	EUR	60	
rs34974290	17q25	A	0.19	*TRIM47*[f]*, TRIM65*[f]*, NEURL, WBP2*[f]*, RP11–552F3.9*[f]	EUR/TRANS	44	62,61,21,63,60,22

37

| rs429358 | 19q13 | C | 0.15 | *APOE* | EUR | [21] | NA |
| rs5762197 | 22q12[c] | C | 0.71 | *MN1* | EUR | [44] | NA |

[a]If several independent single nucleotide polymorphisms have been associated with cerebral small vessel disease at the genome-wide significance level in the same region (linkage disequilibrium, r^2 >0.1), all of them are reported in the table. [b]When several GWAS reported a genome-wide significant association with the same locus, we present the results of the GWAS performed on the largest sample size (largest number of cases), referred to as "main ref"; other studies reporting a genome-wide significant association with the same locus are listed under "other ref". [c]r^2 = 0.002. [d]not associated with vascular risk factors. [e]rs3129171 and rs374598428 have been reported as independent[21]. [f]Expression of these genes was associated with MRI-defined covert cerebral small vessel disease in TWAS and colocalization analyses[44]. [g]r^2 = 0.073. [h]all r^2 <0.01. [i]r^2 = 0.002. [j]r^2 = 0.013. ID, insertion–deletion; EUR, European; RAF, relative allele frequency; TRANS, transethnic.

Table 3 | Involvement of genes associated with cerebral small vessel disease in biological processes.

Biological process[a]	Gene	Protein	Involvement of protein in biological process	Ref
Extracellular matrix structure and function	EFEMP1	Fibulin 3	ECM glycoprotein localized in the basement membrane. Proteolytic target of serine protease HTRA1.	[182]
	LOX	Lysyl oxidase	Copper-dependent enzyme that cross-links collagens and elastin.	[96]
	SH3PXD2A	SH3 and PX domain-containing protein 2A	Adapter protein involved in ECM degradation via invadopodia and podosome formation	[183]
	NID2	Nidogen 2	ECM glycoprotein and a major component of basement membranes.	[96]
	ADAMTSL4	ADAMTS-like protein 4	Involved in cellular adhesion, angiogenesis and patterning of the developing nervous system.	[96]
	FBN2	Fibrillin 2	Component of connective tissue microfibrils that might be involved in elastic fibre assembly.	[65]
	VCAN	Versican core protein	Proteoglycan involved in cell adhesion and ECM assembly.	[184]
	AGRG6	Adhesion G-protein coupled receptor 6	ECM protein produced by astrocytes and involved in pericyte differentiation and blood–brain barrier development and function; activated by type IV collagen and binds to laminin211.	[102,185]
	HBEGF	Heparin-binding EGF-like growth factor	EGF-like is one of the characteristic domains of ECM proteins.	[96]
Myelination	ULK4[b]	Serine–threonine protein kinase ULK4	Regulates a core set of factors that are essential in the development of oligodendrocytes .	[51]
	VCAN	Versican core protein	Forms complexes that inhibit oligodendrocyte maturation	[100]
	AGRG6	Adhesion G-protein coupled receptor 6	Involved in Schwann cell myelination in the PNS	[98,99]
Membrane transport	KCNK2	Potassium channel subfamily K member 2	Enables potassium transport across the cytoplasmic membrane; defects underlie early-stage cerebrovascular dysfunction in CADASIL.	[95]
	SLC39A13	Solute carrier family 39 member 13	Transmembrane zinc transporter; mutations cause a form of Ehlers–Danlos syndrome, which can cause stroke	[186]
	SLC25A44[b]	Solute carrier family 25 member 44	Mitochondrial carrier protein with a role in catabolism of branched-chain amino acids in brown adipose tissue	[187]

[a]Selected biological processes associated with genes that have been associated with common small vessel disease. [b]Genes have been associated with common small vessel disease in transcriptome-wide analysis studies and colocalization analyses, providing stronger evidence for their causal involvement.

Fig. 1 | Potential clinical applications of cerebral small vessel disease genomics.
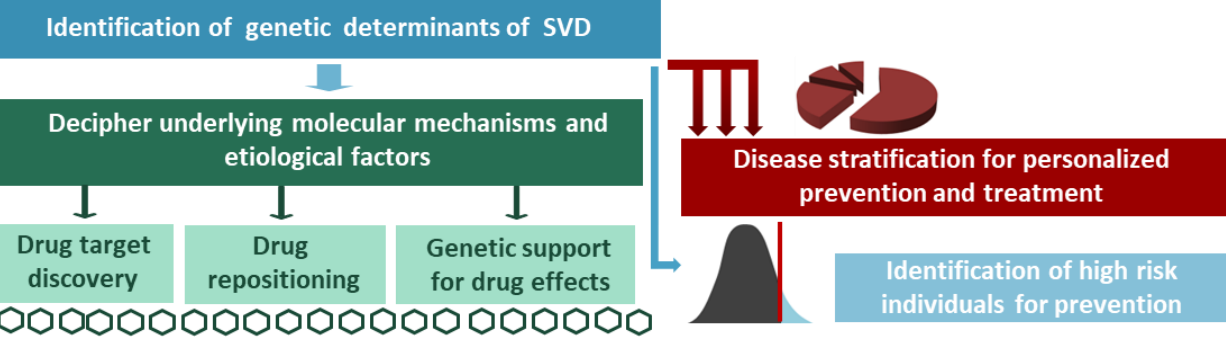
Knowledge of the genetic determinants of small vessel disease has several implications for drug development and clinical management.

Fig. 2 | Shared risk loci for cSVD phenotypes and associations with vascular risk factors.

A

B
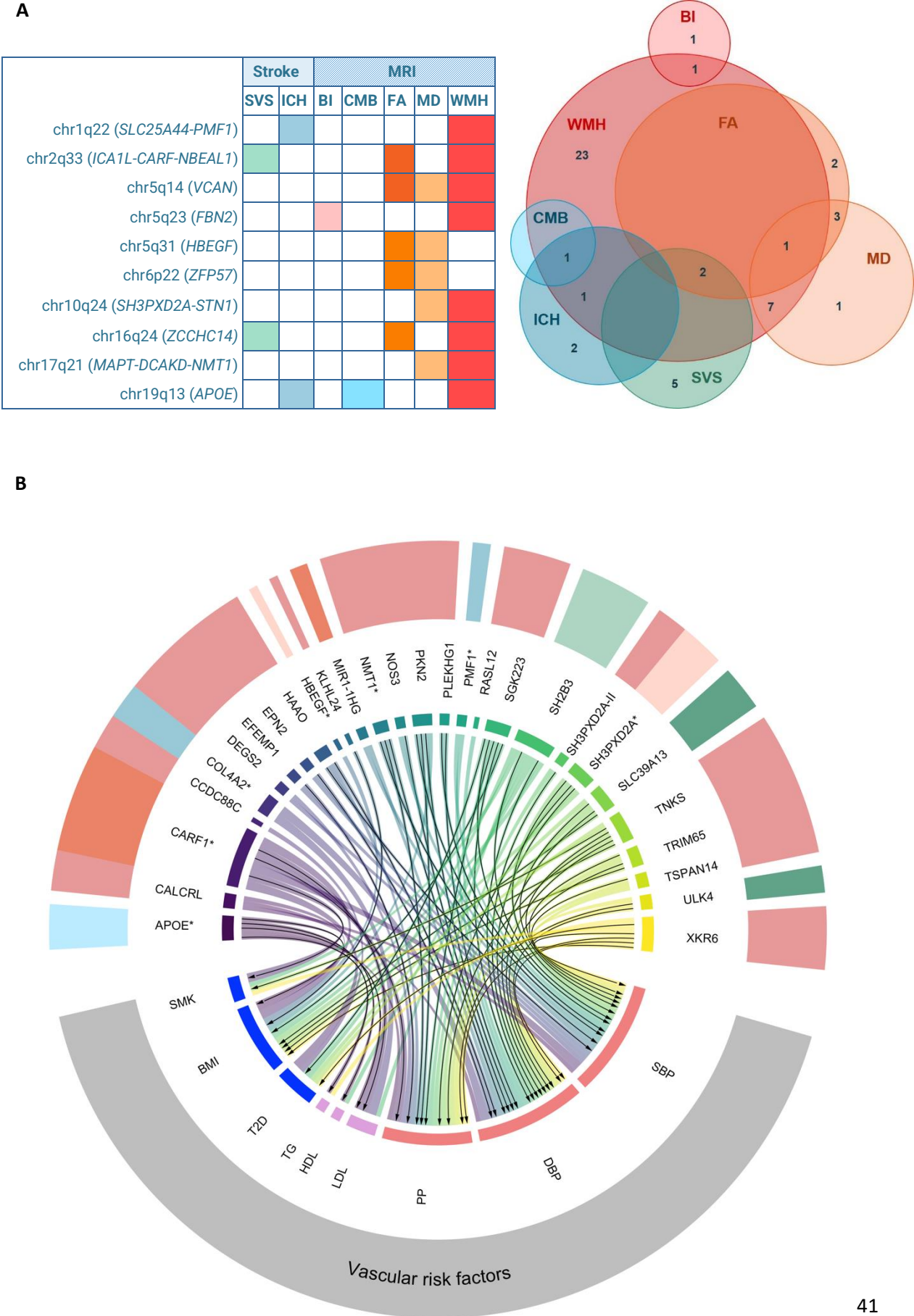
Loci reaching genome-wide significance in multivariate analyses based on two cSVD phenotypes are not included in this figure a | Venn diagram that shows the extent to which different cSVD phenotypes share genome-wide significant risk loci. The cSVD phenotypes shown include overt phenotypes (small vessel stroke (SVS) and intracerebral haemorrhage (ICH)) and covert, MRI-defined markers of ccSVD (white matter hyperintensities (WMH), brain infarcts (BI), cerebral microbleeds (CMB), fractional anisotropy (FA) and mean diffusivity (MD)). Where risk loci are shared between phenotypes, the chromosome coordinates and the nearest genes to these co-ordinates are specified in the surrounding boxes. b| Chord diagram that summarizes the association of genome-wide significant risk variants for cSVD traits (top half) with vascular risk factors (bottom half). The width of each of the inner stems reflects the number of traits associated with a given locus (top half) or the number of loci associated with a given vascular risk factor (bottom half). Black arrows indicate genome-wide significant associations. Outer coloured segments in the top half match colours in part a, and dark green represents lacunar stroke loci. [a]Locus overlaps with other cSVD traits — see part a for details.

	SVS	ICH	Lobar ICH	Non-lobar ICH	Lacunes**	CMB	WMH	FA*	MD
SBP	3/3	1/1	1/1	1/1	1/1	1/1[a]	2/2		
DBP	3/3	1/1	1/1	1/1	1/1	1/1[a]	2/2		
LDL	2/5	1/2	1/1	1/1	1/1	1/1	2/2		
Lp(A)	1/1								
HDL	3/4	2/2			1/1	1/1	1/3		
APO B	1/1								
APO A-I	1/1								
TG	2/3	2/2			1/1	1/1	1/3		
BMI	4/5	3/3	1/1	1/1	1/1	1/1	2/2		
WHR	2/3	1/3	1/1	1/1			1/1		
Type 2 diabetes	5/5	2/2	2/2	2/2	1/1	1/1	1/3	2/2	2/2
HbA1c	1/1	1/1	1/1	1/1			1/1	1/1	1/1
Smoking	3/4				1/1		1/1		
High education	1/1	1/1							
Tea intake	1/1 ‖								
Low birth weight (maternal effects)	1/1								
Major depressive disorder	1/1								
Migraine							1/1[b]		
Fibrinogen	1/1								
Homocysteine	1/1								
Folate	1/1 ‡								
B6	1/1 ‡								
Interleukin-6	1/1 §								

Blue boxes indicate that at least one Mendelian randomization study has shown a significant association between genetically predicted higher exposure to the risk factor and an increased risk or burden of the cerebral small vessel disease (cSVD). Pink boxes indicate that at least one Mendelian randomization study has shown a significant association between genetically predicted higher exposure to the risk factor and a reduced risk or burden of the cSVD. Numbers indicate the fraction of Mendelian randomization studies conducted that have shown a significant association. Dotted blue boxes indicate consistently non-significant associations. Empty boxes indicate no published studies. **For lacunes, the association of genetically predicted risk factors with lacunes was tested but pleiotropy was not formally ruled out. *For FA, association of genetically predicted risk factors with lower FA is being tested. [a]Only nominally significant after accounting for the number of risk factors tested. ‖Not significant when the weighted median or the Mendelian randomization Egger method were used. ‡Only

nominally significant after removing outlier single-nucleotide polymorphisms. [§]Weak instruments (pleiotropy not tested). [g]Borderline significant.

Box 1 | Methods to explore the genomics of common (multifactorial) diseases or traits

Candidate gene association studies

Candidate gene association studies test associations between genetic variants within a gene or region of interest and a relevant phenotype (disease or trait). The genes of interest are usually selected on the basis of *a priori* hypotheses on their relevance to the phenotype under investigation.

Genome-wide association studies

Genome-wide association studies (GWAS) offer an 'agnostic' approach by testing associations of millions of common genetic variants (minor allele frequency $\geq$5%) or low-frequency variants (1%$\leq$ minor allele frequency <5%) with a phenotype of interest at the genome-wide scale. To avoid false positive findings when multiple association tests are performed for millions of variants, a *P* value threshold of <5×10^{-8} is usually used to define genome-wide significance. Most genetic variants studied in GWAS are single-nucleotide polymorphisms (SNPs).

Expression quantitative loci studies

Expression quantitative loci (eQTL) refer to genetic variants associated with mRNA expression levels of a given gene within a given tissue. eQTL studies can be useful for in silico functional exploration of genetic variants identified in GWAS, by linking SNPs to tissue-specific changes in gene expression.

Transcriptome-wide association studies

Transcriptome-wide association studies enable integration of data from GWAS and eQTL studies to derive associations between tissue-specific gene expression levels and a given phenotype.

Colocalization analyses

Colocalization analyses test whether the same causal variant is responsible for the association of a genetic variant with a molecular trait, such as expression levels of a gene in a given tissue, and with a disease of interest.

Linkage disequilibrium score regression

Linkage disequilibrium score regression (LDSR) examines the relationships of linkage disequilibrium score and SNP association test statistics with a phenotype. LDSR is used to estimate SNP-based heritability of a given phenotype considering genome-wide significant SNPs and enables testing the genetic correlation between two phenotypes.

Multi-trait analysis of genome-wide association summary statistics

Multi-trait analysis of genome-wide association summary statistics (MTAG) enables a joint association analysis of multiple traits on the basis of individual trait GWAS summary statistics while accounting for sample overlap. This approach increases statistical power for detection of associations with genetically correlated traits.

Mendelian randomization

Mendelian randomization studies enable use of genetic variants to identify causal associations between specific factors or exposures and a specific trait or disease (see **Box 2**)

Box 2 | Mendelian randomization

Mendelian randomization is an analytical approach in which genetic variants are used as instrumental variables that approximate the effect of a given exposure or factor (X) to seek evidence for a causal association between this exposure and a disease or trait (Y). The method is based on the fact that genetic alleles are assorted randomly during meiosis, meaning that the analysis is less prone to confounding and reverse causation than are observational epidemiological studies. As a result, Mendelian randomization can generate more reliable evidence regarding the association of the exposure with disease and therefore provide indirect evidence for the development of interventions that could benefit health.

A set of genetic variants — referred to as a genetic instrument — can be used as a proxy for a given modifiable exposure if they are robustly associated with the risk of exposure. For example, a set of genetic variants associated with high blood pressure can be used as a proxy for exposure to high blood pressure. The instrumental variable must not be associated with Y independently of X (Figure).

Mendelian randomization can be performed at the individual data level (one-sample Mendelian randomization)[133] or at the summary statistics level (two-sample Mendelian randomization)[134]. In two-sample Mendelian randomization, data from two study samples are used to compare the association between the instrument and the exposure with the association between the instrument and the outcome across multiple single-nucleotide polymorphisms to estimate the causal role of the exposure in the outcome. Given that GWAS summary statistics are largely made publicly available, two-sample Mendelian randomization has become extremely popular.

Several notes of caution are required in relation to Mendelian randomization studies. Weak genetic instruments — for example, those that are pleiotropic and/or are directly associated with the outcome — will lead to regression dilution bias towards the null[188], so that the estimated causal effects are underestimations. Construction of instrumental variables from well-powered GWAS helps to generate stronger genetic instruments[189], but this approach also increases the likelihood of adding pleiotropic SNPs. Avoidance of this risk requires application of various methods that avoid pleiotropy[190].

Glossary

Genetic variant

A specific region of the genome that differs between individuals in the population.

Single nucleotide polymorphism

A single nucleotide variation in the DNA sequence.

Alleles

Two or more versions of a polymorphic genetic site; for single nucleotide polymorphisms, alleles correspond to two alternative nucleotides at the given position.

Linkage disequilibrium

The non-random association of alleles at two nearby genetic loci, reflecting haplotypes that descend from a common ancestor.

White matter hyperintensity of presumed vascular origin (WMH)

An abnormality of variable size in the white matter seen on MRI as a hyperintensity on T2-weighted images such as fluid-attenuated inversion recovery without cavitation (with a signal different from the cerebrospinal fluid).

Fractional anisotropy

A scalar measure derived from diffusion tensor imaging that quantifies the overall directionality of water diffusion in brain tissue. The measure is greater in organized white matter tracts and lower in disorganized tissues, such as CSF.

Mean diffusivity

A scalar measure obtained from diffusion tensor imaging that quantifies the average mobility of water molecules regardless of its direction within the brain tissue. It is used to study microstructural properties and the structural integrity of brain tissue.

Cerebral microbleed

A small area (usually 2–5 mm in diameter) of blood leakage that appears as a signal void with associated blooming on T2*-weighted MRI or susceptibility-weighted imaging.

Perivascular spaces

Fluid-filled spaces that surround perforating vessels in the brain, with a signal intensity similar to that of cerebrospinal fluid on all MRI sequences (diameter generally smaller than 3 mm).

Lacune of presumed vascular origin

A subcortical, round, fluid-filled cavity (3–15 mm diameter) consistent with a previous acute small subcortical infarct or haemorrhage in the territory of one perforating arteriole.

Pleiotropy

A phenomenon in which genes or genetic variants affect multiple, apparently unrelated, phenotypes.

Inversion polymorphism

A type of DNA structural variant that changes the orientation of a genomic segment.

In this Review, the authors provide an overview of the genetics of common small vessel disease, the insight into causal genes and the biological pathways involved, the overlap with monogenic small vessel disease, and the therapeutic implications.