Stroke genetics: discovery, biology, and clinical applications

Martin Dichgans, Sara L Pulit, Jonathan Rosand

Stroke, a leading cause of long-term disability and death worldwide, has a heritable component. Recent gene discovery efforts have expanded the number of known single-gene disorders associated with stroke and have linked common variants at approximately 35 genetic loci to stroke risk. These discoveries have highlighted novel mechanisms and pathways implicated in stroke related to large artery atherosclerosis, cardioembolism, and small vessel disease, and defined shared genetic influences with related vascular traits. Genetics has also successfully established causal relationships with risk factors and holds promise for prioritising targets for exploration in clinical trials. Genome-wide polygenic scores enable the identification of high-risk individuals before the emergence of vascular risk factors. Challenges ahead include a better understanding of rare variants and ancestral differences for integration of genetics into precision medicine, integration with other omics data, uncovering the genetic factors that govern stroke recurrence and stroke outcome, and the conversion of genetic discoveries to novel therapies.

Lancet Neurol 2019; 18: 587–99

Published Online
April 8, 2019
http://dx.doi.org/10.1016/S1474-4422(19)30043-2

Institute for Stroke and Dementia Research, Klinikum der Universität München, Ludwig-Maximilians-Universität, Munich, Germany (Prof M Dichgans MD); **German Center for Neurodegenerative Diseases, Munich, Germany** (Prof M Dichgans); **Munich Cluster for Systems Neurology (SyNergy), Munich, Germany** (Prof M Dichgans); **Department of Genetics, Centre for Molecular Medicine, University Medical Centre Utrecht, Utrecht, Netherlands** (S L Pulit PhD); **Big Data Institute, Li Ka Shing Centre for Health Information and Discovery, Oxford University, Oxford, UK** (S L Pulit); **Program in Medical Population and Genetics, Broad Institute, Cambridge, MA, USA** (S L Pulit, Prof J Rosand MD); **and Henry and Allison McCance Center for Brain Health, and Center for Genomic Medicine, Massachusetts General Hospital, Boston, MA, USA** (Prof J Rosand)

Correspondence to:
Prof Martin Dichgans, Institute for Stroke and Dementia Research, Klinikum der Universität München, D-81377 Munich, Germany
martin.dichgans@med.uni-muenchen.de

Introduction

Stroke is a leading cause of death and long-term disability worldwide.[1] Despite the discovery of modifiable and non-modifiable risk factors and effective treatments, novel therapeutic approaches are urgently needed to limit the growing burden of stroke. Uncovering the genetic contributions to stroke could lead to a better definition of causal pathways, the identification of novel therapeutic targets, and improved options for diagnosis and prognosis.[2–7]

The past 5 years have seen substantial advances in genomic technologies, sequencing costs, biobanking, and data sharing, which collectively have accelerated genetic discovery (panel 1).[2,8,9] Genetic studies in stroke are now interrogating both common and rare genetic variation for a causal role in disease. Genome-wide association studies (GWASs) in stroke and other vascular traits, such as blood pressure and atrial fibrillation, have tested more than a million samples and associated an ever-increasing number of loci with disease risk.[2,10] These discoveries, along with the expanding availability of other omics data and rare genetic variants, have begun to elucidate causal pathways, relevant cell and tissue types, and in some instances, have yielded novel drug targets (eg, proprotein convertase subtilisin/kexin type 9, PCSK9).[2,11]

In this Review, we summarise the latest discoveries in stroke genetics. In particular, we discuss the identification of novel mendelian causes of stroke,[4,5,12] the discovery of at least 35 stroke risk loci harbouring common genetic variants,[2,7] insights into subtype-specific mechanisms for stroke, genetic overlap with related traits, and efforts to understand the underlying biological mechanisms.[2,3,7] We also discuss how genetic discoveries could improve diagnosis, risk prediction, and treatment of stroke. Indeed, observations in case-control studies and population-based cohorts have elucidated how genetics could be leveraged to discover novel drug targets and identify high-risk individuals many years before the emergence of classic indicators of stroke risk (eg, by the application of polygenic risk scores).[2,6,13,14] Because of their substantially different pathophysiologies, we have not considered subarachnoid haemorrhage, cerebral aneurysms, cavernous malformations, dissections, and cerebral venous thrombosis in this Review.

Genetic discovery for stroke

Mendelian stroke

Advances in sequencing technology have facilitated the discovery of single-gene disorders associated with stroke beyond classic syndromes, such as cerebral autosomal dominant arteriopathy with subcortical infarcts and leukoencephalopathy (CADASIL) and sickle-cell disease (table). Most notably, there has been a substantial expansion of genetically defined types of ischaemic small vessel disease, which can manifest with ischaemic stroke (small vessel stroke), cognitive decline, and other manifestations. Small vessel stroke (also known as small artery stroke, small artery ischaemic stroke, and lacunar stroke) refers to an acute ischaemic stroke likely to be caused by small vessel disease based on diagnostic assessment. Among the latest discoveries are heterozygous mutations within the 3′ untranslated region of *COL4A1* (the gene encoding collagen 4A1) as a cause of pontine autosomal dominant microangiopathy with leukoencephalopathy (PADMAL), a severe form of small vessel disease that typically manifests with early-onset ischaemic stroke.[4] These mutations disrupt a binding site for the microRNA miR-29 and upregulate *COL4A1* mRNA expression (miRNA is a small non-coding RNA molecule that functions in RNA silencing and post-transcriptional regulation of gene expression). Heterozygous mutations (in particular, glycine substitutions) in the triple helical domains of *COL4A1* or *COL4A2* cause a different syndrome characterised by haemorrhagic stroke along with additional neurological and non-neurological manifestations (table).[23] Sequencing has further pinpointed heterozygous mutations in *HTRA1* (encoding high temperature requirement serine protease A1, HTRA1) in families with autosomal dominant small vessel disease.[5] The condition typically manifests with stroke and cognitive decline in mid-to-late adulthood (age >45 years) with more and more cases reported in the

Panel 1: **Glossary of key terms and concepts relevant to stroke genetics**

3′ untranslated region
The untranslated segment of mRNA between the stop codon and the 3′ end of the transcript.

Allele
Alternative forms or varieties of a gene, usually arising through mutations, that are responsible for hereditary variation.

Gene set enrichment analysis
A method to identify classes of genes or proteins that are over-represented in a large set of genes or proteins and might have an association with disease phenotypes.

Exome
Part of the genome composed of exons, the sequences which, when transcribed, remain within the mature RNA (after removal of introns) and contribute to the final protein product encoded by that gene.

Fine mapping
A process by which a trait-associated region from a genome-wide association study is analysed to identify the particular genetic variants that are likely to causally influence the examined trait.

Functional genomics
The study of genes and their resulting gene products, and their role in biological processes.

Genetic variant
An alteration in the most common DNA nucleotide sequence. The alteration may be benign, pathogenic, or of unknown significance.

Genome-wide association study
Study of a genome-wide set of genetic variants in different individuals to see if any variant is associated with a trait. Genome-wide association studies typically focus on associations between single nucleotide polymorphisms and specific traits including diseases but can equally be applied to any other genetic variants.

Intergenic region
A stretch of DNA sequences located between genes. Intergenic regions are a subset of non-coding DNA.

Lead single nucleotide polymorphism
A single nucleotide polymorphism is a DNA sequence variation occurring when a single nucleotide (adenine, thymine, cytosine, or guanine) in the genome differs between individuals. The lead single nucleotide polymorphism is the polymorphism with the most significant p value at a specific genetic locus.

Minor allele frequency
The frequency at which the second most common allele occurs in a given population.

Non-coding DNA
Components of DNA that do not encode protein sequences. Some non-coding DNA is transcribed into functional non-coding RNA molecules. Other functions of non-coding DNA include the transcriptional and translational regulation of protein-coding sequences.

Polygenic risk score
Summarises genome-wide genotype data into a single variable that measures genetic liability to a disorder or a trait.

Population attributable risk
Indicates the number of cases that would not occur in a population if the factor were eliminated.

Protein-coding sequence
DNA sequence that is transcribed into mRNA and for which the corresponding mRNA molecule is translated into a polypeptide chain.

past 4 years. By contrast, homozygous and compound heterozygous *HTRA1* mutations causing cerebral autosomal recessive arteriopathy with subcortical infarcts and leukoencephalopathy (CARASIL) seem very rare.[24] CARASIL manifests at a much younger age (10–30 years) than the syndrome of heterozygous mutation carriers and further features non-neurological symptoms such as alopecia and spondylosis.

Hereditary small vessel disease syndromes typically presenting during childhood include deficiency of adenosine deaminase 2, an autoinflammatory disease manifesting with small vessel vasculitis caused by *ADA2* (*CECR1)* mutations.[19] Mutations in *CTSA* (encoding cathepsin A)[12] and *FOXC1* (encoding forkhead box C1)[18] are associated with autosomal dominant small vessel disease typically manifesting after the age of 40 years and as young as 1 year of age, respectively.

Common sporadic stroke

Sporadic stroke arises through multiple risk factors and mechanisms. It is broadly classified as ischaemic (large artery stroke, cardioembolic stroke, and small vessel stroke) and haemorrhagic (deep and lobar intracerebral haemorrhage). The assignment to specific subtypes draws on the presence of established stroke risk factors and intermediate phenotypes such as carotid stenosis (for large artery stroke) or atrial fibrillation (for cardioembolic stroke). Yet, in a substantial proportion of cases, the underlying stroke mechanism remains uncertain either because diagnostic work-up fails to find an established cause of stroke or because there are multiple competing causes. This complexity, along with the existence of different algorithms for stroke classification (discussed elsewhere[25]), has posed challenges to unravelling the genetic underpinnings of sporadic multifactorial stroke.

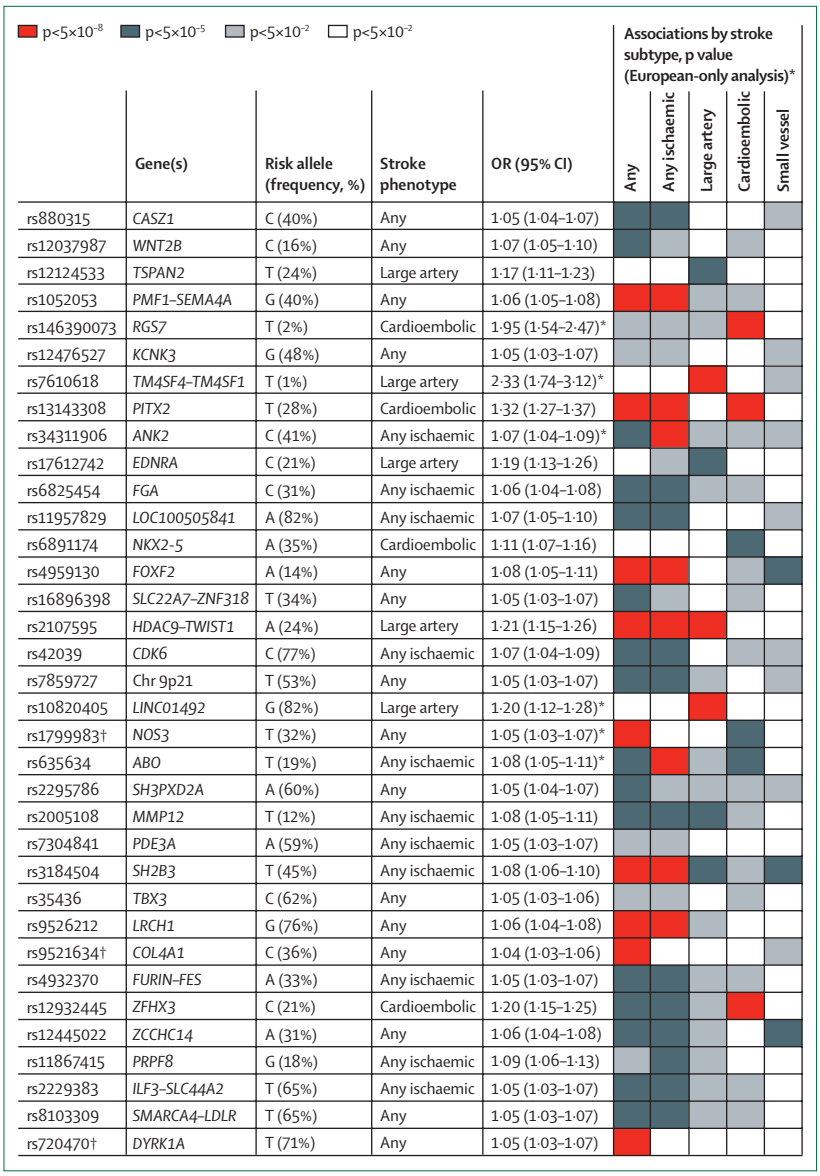
	Mode of inheritance	Underlying gene(s)	Stroke mechanism	Comment and selected key references*
Mendelian conditions mostly manifesting with ischaemic stroke				
CADASIL	Autosomal dominant	*NOTCH3*	Small vessel disease	Most common hereditary stroke syndrome[15]
CARASIL	Autosomal recessive	*HTRA1*	Small vessel disease	Heterozygous mutations in *HTRA1* might cause late-onset small vessel disease[5]
CARASAL	Autosomal dominant	*CTSA*	Small vessel disease	Can manifest with both ischaemic and haemorrhagic stroke, and hypertension[12]
Fabry's disease	X-linked	*GLA*	Small vessel disease, large artery disease, cardioembolism	Multi-organ disease; enzyme-replacement therapy available
PADMAL	Autosomal dominant	*COL4A1*	Small vessel disease	Ischaemic lacunar infarctions in the pons as a common presentation[4]
RVCL-S	Autosomal dominant	*TREX1*	Small vessel disease	Retinopathy and rim-enhancing mass lesions on brain MRI[16]
Sickle-cell disease	Autosomal recessive	*HBB*	Prothrombotic state, large artery disease	Most common cause of stroke in children; haemorrhagic strokes in adult patients[17]
FOXC1 deletion-related angiopathy	Autosomal dominant	*FOXC1*	Small vessel disease	Common manifestations are white matter hyperintensities on brain MRI[18]
DADA2	Autosomal recessive	*ADA2* (*CECR1*)	Small vessel vasculitis	Typically manifests in early childhood (before age 6 years); fever, skin changes, polyarteritis nodosa[19]
Pseudoxanthoma elasticum	Autosomal recessive	*ABCC6*	Large artery disease, small vessel disease	Common manifestations are skin and retinal changes; calcified elastic fibres[20]
Homocystinuria	Autosomal recessive	*CBS* and others (eg, *MTHFR*)	Large artery disease, cardioembolism, small vessel disease, arterial dissection	Common manifestations are thromboembolism, premature atherosclerosis, mental retardation, Marfan-like skeletal abnormalities
Marfan's syndrome	Autosomal dominant	*FBN1*	Cardioembolism, arterial dissection	Clinical diagnosis based on skeletal abnormalities, aortic root aneurysm, ectopia lentis, and other features
Vascular Ehlers-Danlos syndrome	Autosomal dominant	*COL3A1*	Arterial dissection	Stroke typically before 40 years of age[21]
MELAS	Maternal	Mitochondrial DNA	Microvascular and neuronal factors	Stroke-like episodes typically before 40 years of age; seizures, encephalopathy
Hereditary haemorrhagic telangiectasia	Autosomal dominant	*ENG* or *ALK1* in about 85% of cases	Arteriovenous malformations	Pulmonary arteriovenous malformations as a cause of ischaemic stroke; cerebral arteriovenous malformations as a cause of intracerebral haemorrhage[22]
Mendelian conditions mostly manifesting with haemorrhagic stroke				
COL4A1 or *COL4A2*-related angiopathies	Autosomal dominant and de novo	*COL4A1*, *COL4A2*	Small vessel disease	About 50% are sporadic cases; haemorrhages manifest perinatally, in childhood, or in adulthood[23]
Cerebral amyloid angiopathy	Autosomal dominant	*APP*, *CST3*	Cerebral amyloid angiopathy	Manifests with stroke (mostly haemorrhagic) and dementia
Cerebral cavernous malformations	Autosomal dominant	*KRIT1 (CCM1)*, *MGC4607 (CCM2)*, *PDCD10 (CCM3)*	Cerebral cavernous malformations	Multiple cavernomas; proportion of familial cases is up to 50% in Hispanic-American patients

CADASIL=cerebral autosomal dominant arteriopathy with subcortical infarcts and leukoencephalopthy. CARASIL=cerebral autosomal recessive arteriopathy with subcortical infarcts and leukoencephalopthy. CARASAL=cathepsin A-related arteriopathy with strokes and leukoencephalopathy. PADMAL=pontine autosomal dominant microangiopathy with leukoencephalopathy. RVCL-S=retinal vasculopathy with cerebral leukoencephalopathy (and systemic manifestations). DADA2=deficiency of adenosine deaminase 2. MELAS=mitochondrial myopathy, encephalopathy, lactic acidosis, and stroke-like episodes. *Selected key references mostly from recent literature are provided.

Table: **Mendelian causes of stroke**

Common genetic variants associated with stroke

Common genetic variants, defined here as an allele frequency of at least 0·5% (one carrier per 100 individuals), have been associated with stroke using GWASs, which compared the frequency of variants (mostly single nucleotide polymorphisms, SNPs) between groups of individuals. The first stroke GWAS included 3548 cases and 5972 controls and identified common variants at *HDAC9* (encoding histone deacetylase 9), which conferred a 40% increased risk for large artery stroke per copy of the risk allele.[26] Since then, GWASs in progressively larger sample sizes (eg, >70 000 cases) have identified at least 35 loci with robust links to stroke risk.[2,27,28] Collectively, these studies enable the following conclusions:[7] first, the majority of associated SNPs have a minor allele frequency of more than 5% and are associated with a modest increase in stroke risk (typically <30% increase per allele; figure 1). Second, most (about 90%) associated variants reside outside protein-coding sequences and approximately half are intergenic. Third, there are risk loci for all major diagnostic categories including any stroke, any ischaemic stroke, large artery stroke, cardioembolic stroke, small vessel stroke, and both lobar and deep intracerebral haemorrhage (figure 1, and discussed elsewhere[27,28]). Fourth, at several loci, the association is restricted to a specific aetiological stroke subtype. For example, in the MEGASTROKE GWAS (>67 000 stroke cases and >450 000 controls),[2] the lead variants near *EDNRA*, *TSPAN2*, and *LINC01492* reached genome-wide significance ($p<5\times10^{-8}$) for large artery stroke but showed no association with other stroke subtypes, implying mechanisms limited to atherosclerosis.



	Gene(s)	Risk allele (frequency, %)	Stroke phenotype	OR (95% CI)
rs880315	*CASZ1*	C (40%)	Any	1·05 (1·04–1·07)
rs12037987	*WNT2B*	C (16%)	Any	1·07 (1·05–1·10)
rs12124533	*TSPAN2*	T (24%)	Large artery	1·17 (1·11–1·23)
rs1052053	*PMF1–SEMA4A*	G (40%)	Any	1·06 (1·05–1·08)
rs146390073	*RGS7*	T (2%)	Cardioembolic	1·95 (1·54–2·47)*
rs12476527	*KCNK3*	G (48%)	Any	1·05 (1·03–1·07)
rs7610618	*TM4SF4–TM4SF1*	T (1%)	Large artery	2·33 (1·74–3·12)*
rs13143308	*PITX2*	T (28%)	Cardioembolic	1·32 (1·27–1·37)
rs34311906	*ANK2*	C (41%)	Any ischaemic	1·07 (1·04–1·09)*
rs17612742	*EDNRA*	C (21%)	Large artery	1·19 (1·13–1·26)
rs6825454	*FGA*	C (31%)	Any ischaemic	1·06 (1·04–1·08)
rs11957829	*LOC100505841*	A (82%)	Any ischaemic	1·07 (1·05–1·10)
rs6891174	*NKX2-5*	A (35%)	Cardioembolic	1·11 (1·07–1·16)
rs4959130	*FOXF2*	A (14%)	Any	1·08 (1·05–1·11)
rs16896398	*SLC22A7–ZNF318*	T (34%)	Any	1·05 (1·03–1·07)
rs2107595	*HDAC9–TWIST1*	A (24%)	Large artery	1·21 (1·15–1·26)
rs42039	*CDK6*	C (77%)	Any ischaemic	1·07 (1·04–1·09)
rs7859727	Chr 9p21	T (53%)	Any	1·05 (1·03–1·07)
rs10820405	*LINC01492*	G (82%)	Large artery	1·20 (1·12–1·28)*
rs1799983†	*NOS3*	T (32%)	Any	1·05 (1·03–1·07)*
rs635634	*ABO*	T (19%)	Any ischaemic	1·08 (1·05–1·11)*
rs2295786	*SH3PXD2A*	A (60%)	Any	1·05 (1·04–1·07)
rs2005108	*MMP12*	T (12%)	Any ischaemic	1·08 (1·05–1·11)
rs7304841	*PDE3A*	A (59%)	Any ischaemic	1·05 (1·03–1·07)
rs3184504	*SH2B3*	T (45%)	Any ischaemic	1·08 (1·06–1·10)
rs35436	*TBX3*	C (62%)	Any	1·05 (1·03–1·06)
rs9526212	*LRCH1*	G (76%)	Any	1·06 (1·04–1·08)
rs9521634†	*COL4A1*	C (36%)	Any	1·04 (1·03–1·06)
rs4932370	*FURIN–FES*	A (33%)	Any ischaemic	1·05 (1·03–1·07)
rs12932445	*ZFHX3*	C (21%)	Cardioembolic	1·20 (1·15–1·25)
rs12445022	*ZCCHC14*	A (31%)	Any	1·06 (1·04–1·08)
rs11867415	*PRPF8*	G (18%)	Any ischaemic	1·09 (1·06–1·13)
rs2229383	*ILF3–SLC44A2*	T (65%)	Any ischaemic	1·05 (1·03–1·07)
rs8103309	*SMARCA4–LDLR*	T (65%)	Any	1·05 (1·03–1·07)
rs720470†	*DYRK1A*	T (71%)	Any	1·05 (1·03–1·07)

Figure 1: **Common variants associated with stroke**
Top signals from previous genome-wide association study meta-analyses in MEGASTROKE (all rows apart from *NOS3*, *COL4A1*, and *DYRK1A*)[2] and a combined meta-analysis of MEGASTROKE with data from the UK Biobank (rows for *NOS3*, *COL4A1*, and *DYRK1A* only).[7] The order of loci is by position within the genome, starting with the short arm of human chromosome 1 (*CASZ1*) and ending with human chromosome 21 (*DYRK1A*). Shown are loci reaching genome-wide significance ($p<5\times10^{-8}$). For each locus the variant showing the lowest p value in the fixed-effects transancestral or European-only meta-analysis is shown. OR=odds ratio. C=cytosine. T=thymine. G=guanine. A=adenine. Chr=chromosome. *Results obtained in the European-only analysis. For consistency, p values are taken from the respective European-only analyses. †Variants from a combined meta-analysis of MEGASTROKE with data from the UK Biobank.[7]

Fifth, many genetic variants that confer risk of stroke also influence risk of related traits (figure 2). Notably, about a quarter of the 32 genome-wide significant loci identified by MEGASTROKE[2] are established risk loci for high blood pressure. Other loci (eg, near *SMARCA4–LDLR*) are established risk loci for lipid concentrations.[2] Also, there is genetic overlap between large artery stroke and the related trait of coronary artery disease (eg, near *HDAC9–TWIST1, EDNRA*, and chromosome 9p21)[2,30,31] and the intermediate phenotypes of carotid intima-media thickness and carotid plaque (eg, at *EDNRA*).[2,29] Sixth, several stroke risk loci contain genes previously implicated in monogenic small vessel disease (eg, *COL4A1, COL4A2*, and *HTRA1*;[2,7] figure 3), suggesting a biological overlap between common stroke and rare familial stroke syndromes. And finally, although several risk loci integrate into previously suspected biological pathways for stroke, roughly a third show no obvious relationship with known pathways, pointing to mechanisms not previously implicated in stroke pathophysiology.

The heritability of stroke (ie, the proportion of variation in risk attributable to inherited genetic variation) calculated from genome-wide data has been estimated to be 30–40%,[45,46] although a substantial proportion of this variation is likely to be mediated by known risk factors for stroke. Overall, the number of risk loci identified to date remains relatively small when compared with other common conditions including coronary artery disease. This difference is partly because of the complexity of stroke as a phenotype as well as the smaller sample sizes that are as yet available for aetiologically defined subtypes in GWASs. Not surprisingly, the lead variants at stroke risk loci identified to date explain roughly 1–2% of stroke heritability,[2] which is markedly lower than the percentage of heritability for other traits (eg, coronary artery disease).[31] Given the effect of larger sample sizes on genetic discovery in other diseases, it is reasonable to assume that analyses of larger numbers of stroke cases will yield substantial numbers of new loci.

Rare variants in sporadic stroke

The rapid development of low-cost sequencing technologies now enables genome-wide genotyping of all variants across the genome. Rare variant detection requires large-scale sequencing and benefits from analytical strategies that aggregate rare variants in a given gene into variant sets, enabling a comparison of the aggregate frequency across groups.[47] The two largest stroke studies to date (3127 cases and 2088 cases, respectively)[45,48] focused on coding regions (the exome) rather than the whole genome. Although too small to detect robust associations with rare (allele frequency <0·5%) variants, these studies provide a first step towards future discovery. With the rapid decline in sequencing costs and expected gain of information on rare variants for precision medicine approaches, much larger datasets are expected in the near future.

From genetic discovery to biological mechanisms

Because of the complex nature of GWAS signals, the often large number of genes within risk loci, and the complexity of gene regulatory mechanisms occasionally involving multiple or distant genes, mapping GWAS signals to their causal mechanisms is rarely straightforward. Elucidating these mechanisms requires additional genetic data and

work in animal and cellular models, with the most appropriate cellular models varying from locus to locus. These challenges notwithstanding, functional genomics studies have provided initial insights into the mechanisms linking common variants with stroke risk. We highlight key observations on biological mechanisms derived from recent genetic discoveries starting with all stroke and then moving to aetiological stroke subtypes. We focus on genes and gene loci that have received particular attention.

All stroke

Fine mapping at the *SH2B3* locus identified rs3184504 as the most likely causal variant.[2] Further functional annotation revealed *SH2B3* as the most likely causal gene.[2] rs3184504 causes an amino acid exchange (Pro262Trp) in the lymphocyte adaptor protein LNK (also known as SH2B3, the protein encoded by *SH2B3)* and is also associated with hypertension, type 1 diabetes, coronary artery disease, platelet counts, and leukocytosis, suggesting an involvement of this locus in multiple mechanisms relevant to stroke pathophysiology. Genetic variation at *SH2B3* is associated with both the broader phenotypes of any stroke, any ischaemic stroke, and with the aetiologically defined stroke subtypes large artery stroke and small vessel stroke, providing an example of shared genetic influences between aetiological subtypes.[2] LNK is predominantly expressed in haematopoietic and endothelial cells and negatively regulates cytokine signalling and cell proliferation.[49] LNK deficiency is associated with increased platelet production and activation, accelerated arterial thrombosis, and atherosclerosis in hypercholesterolaemic mice,[50] and other mechanisms under investigation. These findings highlight the potential of genetics to uncover disease mechanisms while also illustrating challenges of genetic epidemiology, because *SH2B3* has also been associated with traits for which the link to stroke is more difficult to explain (eg, cancer).

A substantial proportion of stroke risk variants are associated with the mRNA expression, methylation, or protein concentrations of nearby genes primarily in stroke-relevant tissues (vascular, brain) and cell types (eg, endothelial cells, blood, and immune cells), emphasising their role in stroke pathophysiology and providing an entry point for functional exploration. Bioinformatic analyses further highlight a role of specific pathways, most notably cardiac pathways (eg, "enlarged heart" and "cardiomyocyte differentiation via bone morphogenetic protein receptors"), the coagulation system, and nitric oxide metabolism, but also other pathways.[2]

Large artery stroke

The strongest association signal for large artery stroke identified to date is near *HDAC9*,[2,26,45] a locus that is also associated with coronary artery disease (figure 2), peripheral artery disease, and moyamoya disease.[2,31,51] rs2107595, the lead SNP for stroke and likely causal

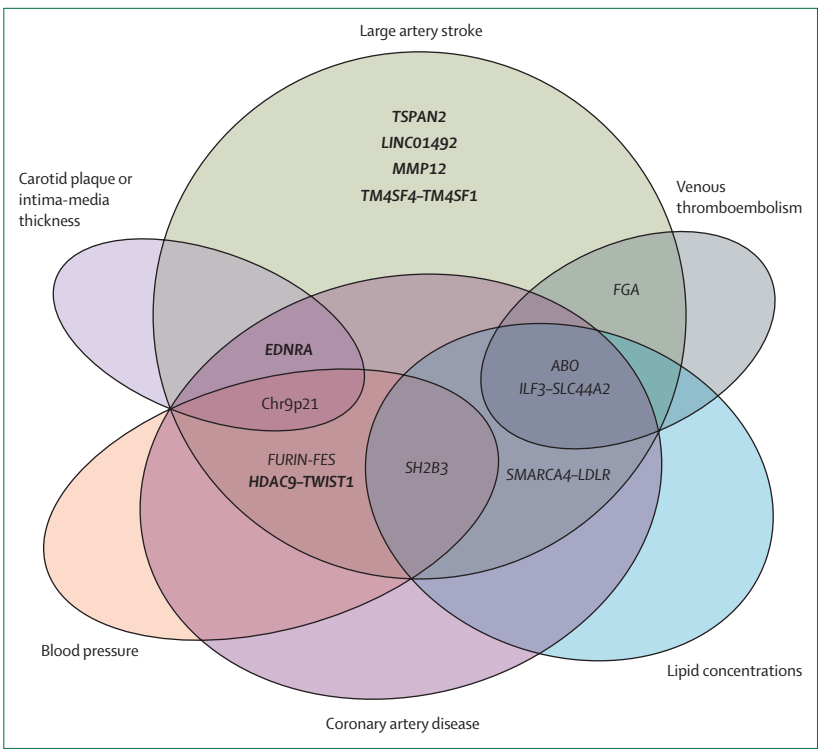


Figure 2: **Risk loci for large artery stroke and their association with related vascular traits**
Genetic loci that either reached genome-wide significance for association with large artery stroke (bold) or reached genome-wide significance for any ischaemic stroke and showed a strong association signal ($p<5\times10^{-4}$) for large artery stroke in MEGASTROKE.[2] Several of the risk loci for large artery stroke are established risk loci for related vascular traits.[2,29]

variant[2] resides in regulatory DNA 3′ to *HDAC9*. Risk variants are associated with elevated *HDAC9* expression levels in blood cells with a gene dosage effect,[52] and deficiency of Hdac9 attenuates atherosclerosis in experimental mice.[52,53] Studies in *Ldlr*$^{-/-}$ mice suggest a proatherogenic effect of *Hdac9* in macrophages via changes in their activation status and cholesterol efflux,[53] although *HDAC9* is also expressed in other cell types relevant to atherosclerosis.[54] Irrespective of the specific mechanisms and cell types involved, pharmacological inhibition with subclass-specific HDAC inhibitors seems a potential strategy for atheroprotection that deserves further study.

Among the most extensively studied risk loci for atherosclerosis is a region on chromosome 9p21 originally identified in coronary artery disease but subsequently shown to also be associated with peripheral artery disease, intracranial and abdominal aortic aneurysms, carotid plaque,[29] and stroke, in particular large artery stroke.[2,55] Risk variants at 9p21 are responsible for up to 20% of the population attributable risk of large artery stroke,[2,55] considered to mediate their effects through mechanisms that are largely independent from established risk factors for atherosclerosis, and likely to act through several genes at this locus. Specifically, these variants are associated with reduced expression of the

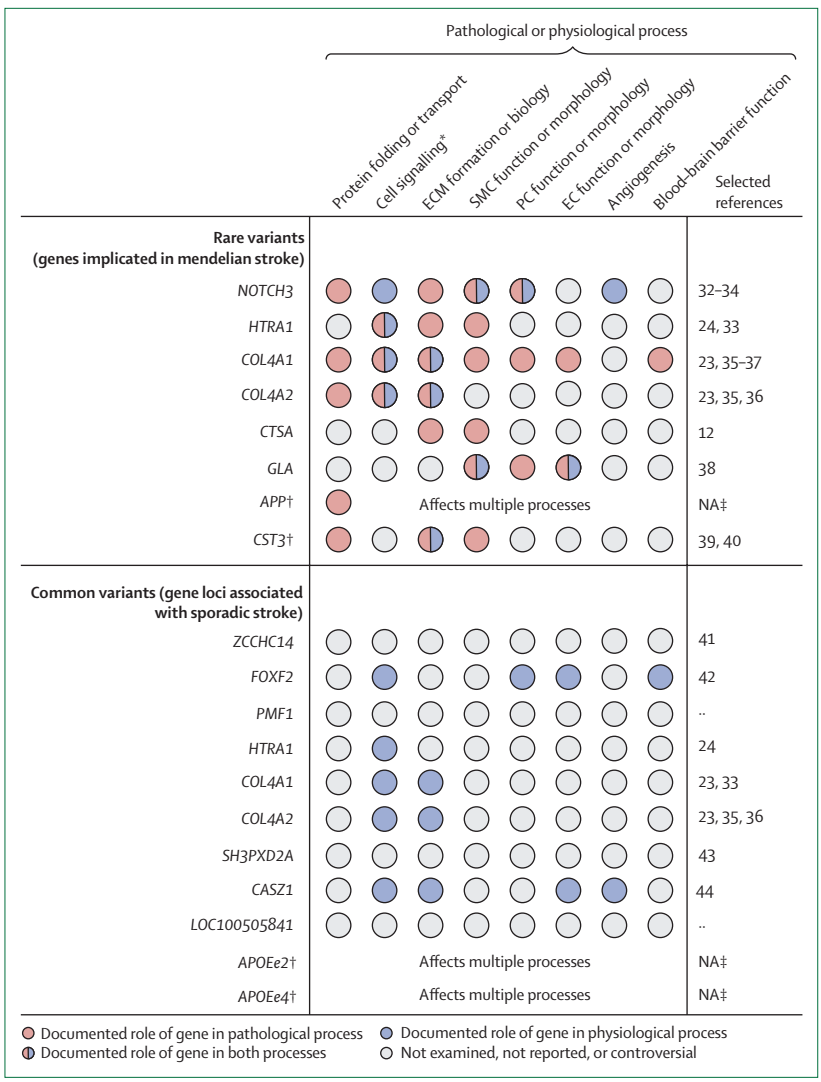


Figure 3: **Pathological and physiological processes related to genes and gene loci associated with small vessel stroke**

Genes implicated in mendelian stroke and gene loci associated with sporadic stroke with the presumed relationship to pathological and physiological processes as supported by experimental or human studies. For example, *NOTCH3* has a physiological role in cell–cell signalling while disease-associated mutations cause abnormal folding and accumulation of the NOTCH3 protein. References were selected following a detailed review of the literature. ECM=extracellular matrix. SMC=smooth muscle cell. PC=pericyte. EC=endothelial cell. NA=not applicable. **HTRA1*, *COL4A1*/*A2*, and *FOXF2* have been shown to affect transforming growth factor (TGF)-β signalling. †Genes primarily implicated in intracerebral haemorrhage. ‡Most of the evidence of the biological effects of *APP* and *APOE* is based on research in Alzheimer's disease (for *APP* and *APOE*) and in atherosclerosis (for *APOE*) rather than in small vessel stroke. Disentangling the contribution of the individual processes in the context of small vessel stroke or cerebral amyloid angiopathy is challenging and beyond the scope of this Review.

cell cycle suppressor genes *CDKN2A* and *CDKN2B*,[56,57] and with vascular smooth muscle cell proliferation and smooth muscle cell content in human atherosclerotic plaques.[56,57] *CDKN2B* has been implicated in smooth muscle cell physiology and clearance of apoptotic debris, and mice deficient of *Cdkn2b* develop advanced atherosclerotic lesions composed of large, lipid-laden necrotic cores.[58] Altogether, evidence suggests a role of 9p21 in vascular remodelling, a key process in atherosclerosis. Importantly, *CDKN2A* and *CDKN2B* and a circular non-coding antisense RNA (*circANRIL*) that is also transcribed at 9p21[59] have been suggested as possible targets for stroke prevention strategies.

GWASs have also found variants near *TSPAN2*,[2,25] *LINC01492*,[2] *EDNRA*,[2] and *MMP12*[2,60] to be associated with large artery stroke. *EDNRA* and *MMP12* offer potential mechanistic insights: *EDNRA* (encoding endothelin receptor type A, ET_A) is expressed in smooth muscle cells, endothelial cells, and macrophages. Variants at *EDNRA* are also associated with carotid intima-media thickness (reflecting early atherosclerosis),[29] coronary artery disease[31] (figure 2), and intracranial aneurysms, suggesting a broader role of this gene and endothelin-1 signalling in vascular disease.[61] Activation of ET_A has effects on vasoconstriction, smooth muscle cell proliferation, extracellular matrix production, and fibrosis[61] (ie, processes relevant to atherosclerosis). Moreover, pharmacological inhibition of ET_A normalised nitric oxide-mediated endothelium-dependent dysfunction and attenuated atheroma formation in atherosclerotic mice,[62] offering a perspective for alternative stroke prevention strategies. The association between *MMP12* (encoding matrix metalloproteinase 12, MMP12) and large artery stroke was originally identified by an age-at-onset informed approach (a covariate-informed approach that also adjusts for age-at-onset).[60] MMPs degrade extracellular matrix proteins, show increased activity in atherosclerotic plaques, and have suspected roles in the growth, destabilisation, and rupture of atherosclerotic lesions. Together, these genetic and experimental findings suggest a causal role of the MMP12 protease in atherosclerosis (reviewed elsewhere[60]). Additional mechanisms supported by stroke GWASs include changes in lipoprotein(a) metabolism (evidenced by *LPA*, a subthreshold locus for large artery stroke [$p=1\cdot3\times10^{-7}$][2] known to be implicated in atherosclerosis[11]) and thromboembolism (evidenced by *FGA*, *ILF3–SLC44A2*, and *ABO*, established risk loci for venous thromboembolism[63]). Hence, risk loci for large artery stroke highlight different aspects of the biology underlying this stroke subtype (figure 2). However, additional work, including fine mapping, gene set enrichment analyses, and work in experimental models, is needed to robustly establish relationships between risk-associated variants and biological mechanisms that result in stroke or stroke-related phenotypes.

Cardioembolic stroke

Genetic risk for cardioembolic stroke has close links with the genetics of atrial fibrillation: all loci reaching genome-wide significance for common variant association with cardioembolic stroke (nearest genes *PITX2*, *ZFHX2*, *NKX2-5*) are established risk loci for atrial fibrillation;[10] at all loci the association signal for stroke is confined to cardioembolic stroke; and the association signals here and at subthreshold loci for cardioembolic stroke (near *CAV1* and *CAV2* and *PRRX1*)[2] overlap with those for atrial

fibrillation, and show a similar architecture (eg, near *PITX2*). Additionally, there are strong genetic links with venous thromboembolism: loci near *ABO*, *FGA* (encoding fibrinogen alpha), and *F11* (a subthreshold locus for cardioembolic stroke [$p=5\cdot2\times10^{-8}$], encoding factor XI) are established risk loci for venous thromboembolism and are known to modulate haemostatic traits.[63] Interestingly, genetic risk for the broader phenotype of any stroke or any ischaemic stroke further shows links with cardiac mechanisms beyond those implicated in atrial fibrillation as exemplified by the associations near *LRCH1* and *ANK2*, found in cardiac pacing and familial forms of cardiac disease (discussed elsewhere[2]).

Small vessel stroke

The identification of genes implicated in monogenic small vessel disease has been instrumental to understanding the biology of small vessel stroke and small vessel disease. Experimental studies in mutant mice and cultured cells highlight key physiological and pathological processes linked to genes implicated in monogenic small vessel disease (figure 3).[12,23,24,32–44] Recurrent themes include transforming growth factor (TGF)-β signalling,[24,42] extracellular matrix formation or biology,[12,23,32,33,35,36,39] blood–brain barrier function,[34,37,42] and specific cellular constituents of brain microvessels, in particular endothelial cells, pericytes, and smooth muscle cells (figure 3).[38–40] A proteomics study of isolated brain microvessels from mutation carriers and mouse models showed an unexpected molecular link between CADASIL and CARASIL that involves accumulation of the HTRA1 protease and its substrates in microvessels from *NOTCH3* mutation carriers.[33]

Common variant association studies identified various risk loci that show predominant association with small vessel stroke (figure 1). *ZCCHC14*, *FOXF2*, and *CASZ1* encode transcription factors expressed in the vasculature.[41,42,44] *Foxf2* (expressed in brain endothelial cells and pericytes) is required for brain pericyte differentiation and development and maintenance of the blood–brain barrier in mice (figure 3).[42] Most notably, *Foxf2*-deficient mice develop microhaemorrhages and neuronal loss, common features of human small vessel disease.[3] The specific mechanisms by which common variants at *HTRA1*, *COL4A1*, and *COL4A2* confer risk of small vessel stroke are still elusive but possibly include altered expression levels of these genes, a finding consistent with the mechanisms seen for rare mutations in these genes. Importantly, several of the risk loci for small vessel stroke (in particular *PMF1–SEMA4A*, *LOC100505841*, *SH3PXD2A*, and *COL4A2*[2,64,65]) are also associated with white matter hyperintensities, highlighting shared biology of small vessel stroke and white matter hyperintensities. *PMF1–SEMA4A* also reached genome-wide significance for association with deep (non-lobar) intracerebral haemorrhage,[28] consistent with non-lobar intracerebral haemorrhage being a manifestation of small vessel disease. Genetic studies in humans in combination with experimental data suggest a broader role of the neural crest genes *FOXF2*, *FOXC1* (near *FOXF2*), and *PITX2* in small vessel disease-related phenotypes.[3,18]

Haemorrhagic stroke

Intracerebral haemorrhage is one subtype of stroke for which genetic evidence points to shared pathways between familial and sporadic disease. Sequencing of *COL4A1* and *COL4A2* in a cohort of patients with sporadic intracerebral haemorrhage identified coding variants with in-vitro pathological consequences resembling familial mutations in the same genes.[66] GWASs of intracerebral haemorrhage have identified associations in the region of *COL4A2* in non-lobar intracerebral haemorrhage, small vessel stroke, and white matter hyperintensities, highlighting the shared pathways underlying the various manifestations of cerebral small vessel disease. The shared biology between intracerebral haemorrhage and small vessel stroke is also underscored by associations at *PMF1–SLC25A44–SEMA4A*.[2,28] Note that variation associated with risk of intracerebral haemorrhage can also affect the extent of bleeding, as measured by haematoma volume,[67] providing support for hypotheses generated from histopathological data that intracerebral haemorrhage expands through the disruption of diseased vessels in the periphery of the original vessel rupture.

Clinical applications

Substantial progress in understanding the genetic underpinnings of stroke has begun to lay the groundwork for future integration of genetic data into routine clinical practice (genomic medicine).

Risk prediction

Aggregation of multiple common variants into polygenic risk scores enables ascertainment of high-risk individuals at a young age (in principle even in childhood), offering opportunities for early prevention.[68,69] For example, one study showed that application of a polygenic risk score consisting of 90 SNPs to data from the UK Biobank (>300 000 participants) could identify individuals with a 35% increased risk of incident stroke (hazard ratio 1·35, 95% CI 1·21–1·50).[6] This cutoff included a third of the population. Compared with individuals in the bottom tertile of the polygenic risk score who had a favourable lifestyle (defined as three or four of the following healthy lifestyle factors: non-smoker, healthy diet, body-mass index <30 kg/m², and regular physical exercise), individuals in the upper tertile of the polygenic risk score had a relative risk (hazard ratio) of 1·44, 1·70, and 2·30, depending on whether they had a favourable, intermediate (two healthy lifestyle factors), or unfavourable (none or one healthy lifestyle factors) lifestyle, respectively. Indeed, lifestyle risk was similar across all polygenic risk score strata, highlighting the potential for early risk stratification and prevention via genetics.[6]

***Panel 2:* Key features of mendelian randomisation studies**

- Use genetic information to confirm a causal relationship between an exposure and an outcome
- Are unaffected by conventional confounding (random assortment of alleles at meiotic segregation)
- Benefit from absence of reverse causality (non-modifiable nature of transmitted germline genome)
- Require a robust association between the genetic instrument and the exposure
- Are increasingly powerful as larger genome-wide association study datasets become available
- Have successfully been applied to vascular traits
- Might identify potential drug targets
- Might identify potential risks associated with pharmacological interventions
- Might replace randomised controlled trials in settings in which such trials are not feasible or ethical to conduct

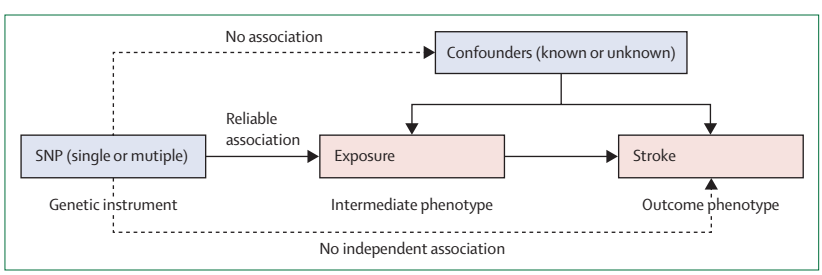


Figure 4: **Exploration of potential therapeutic targets by mendelian randomisation**
Schematic representation of the principles and requirements of instrumental variable analysis to generate causal estimates through mendelian randomisation: the genetic instrument (single variant or multiple variants) is associated with the exposure; the genetic instrument must not be associated with known or unknown confounders; any influence of the genetic instrument on the outcome phenotype is through the intermediate phenotype. SNP=single nucleotide polymorphism.

Additional work is needed to further improve genetic risk prediction, to extend this approach across other ancestral backgrounds, and to weigh its benefits against possible unfavourable consequences including costs and psychological distress.[69]

Exploration of potential therapeutic targets by mendelian randomisation

An increasingly recognised opportunity for genetics is the exploration of causal relationships between risk factors (exposures) and disease outcomes through mendelian randomisation.[70,71] Mendelian randomisation uses genetic variants that are causally related to an exposure as instrumental variables to determine their associations with a disease (panel 2, figure 4). Because individual alleles are allocated at random during meiotic segregation, mendelian randomisation shares many features with randomised controlled trials. Like randomised controlled trials, mendelian randomisation overcomes limitations of observational studies, in particular bias resulting from conventional confounding or reverse causality.

An important requirement for mendelian randomisation studies is a robust association between the genetic instrument (single SNP or multiple variants) and the exposure of interest. For example, an intronic variant (rs6511720) in the LDL receptor (*LDLR*) gene has consistently been shown to be associated with LDL cholesterol concentrations. Like other variants known to influence LDL cholesterol, rs6511720 also associates with risk of myocardial infarction in a concordant fashion.[72] Mendelian randomisation studies have demonstrated a dose-response relationship between genetically elevated LDL cholesterol concentrations and risk of coronary artery disease,[73] consistent with trials involving statins and other cholesterol-lowering interventions. Conversely, mendelian randomisation studies examining HDL as an exposure found no causal effect of genetically determined HDL concentrations on the risk of coronary artery disease, again in accordance with findings from randomised controlled trials (reviewed elsewhere[71]).

Mendelian randomisation studies can predict the success or failure of randomised controlled trials, reducing risks for study participants and reducing risks on costs.[71] Examples in which mendelian randomisation studies predicted the success or failure of randomised controlled trials are studies on inactivating mutations in *NPC1L1* (encoding the drug target for ezetimibe, a lipd-lowering drug)[74] and variants in *PLA2G7* (encoding lipoprotein-associated phospholipase A2, the drug target for darapladib, a drug developed for the prevention of atherosclerosis, reviewed elsewhere[71]), respectively.

Investigators have now started to apply mendelian randomisation to stroke.[75–93] For example, a mendelian randomisation study on blood lipids found that genetically elevated LDL cholesterol concentrations were associated with risk of large artery stroke but not small vessel stroke and cardioembolic stroke.[76] By contrast, genetically elevated HDL cholesterol concentrations were associated with a reduced risk of small vessel stroke but had no effect on large artery stroke and cardioembolic stroke.[76] Differential effects on aetiological stroke subtypes were further reported for type 2 diabetes in that genetically defined type 2 diabetes was found to be associated with both small vessel stroke and large artery stroke but not cardioembolic stroke and intracerebral haemorrhage.[79,80] A mendelian randomisation study on waist-to-hip ratio adjusted for body-mass index found waist-to-hip ratio to be causally related to a higher risk of ischaemic stroke. This finding adds to observational data emphasising the need to consider measures of adiposity beyond body-mass index for risk prediction and patient management.[94] Mendelian randomisation might further identify novel risk factors and potential drug targets for stroke as illustrated by a study showing that genetically elevated concentrations of the inflammatory cytokine CCL2 (MCP-1) were associated with risk of stroke.[86] Figure 5 provides an overview of notable mendelian randomisation studies in stroke.[75–92]

Although potentially highly informative, mendelian randomisation analyses also pose specific challenges that require consideration.[71] Mendelian randomisation is based on the crucial assumptions that the genetic instruments are not associated with potential confounders and that they influence the risk of the disease under study only through the risk factor of interest (figure 4). One example of a violation of these assumptions is horizontal pleiotropy, which refers to a genetic variant being associated with traits on discrete pathways that are also causal in disease. Horizontal pleiotropy can be formally assessed using specific algorithms.[71] Also, there are several methodological requirements (eg, robust and sufficiently strong associations between the variants selected and the exposure of interest) that must be met to confidently exclude a causal relationship between an exposure and disease. These requirements must be kept in mind when interpreting negative mendelian randomisation results (figure 5). With increasing availability of large-scale genetic data and improved databases (eg, MR-Base or PhenoScanner), mendelian randomisation studies will become even more informative and relevant for clinical practice.[11,95]

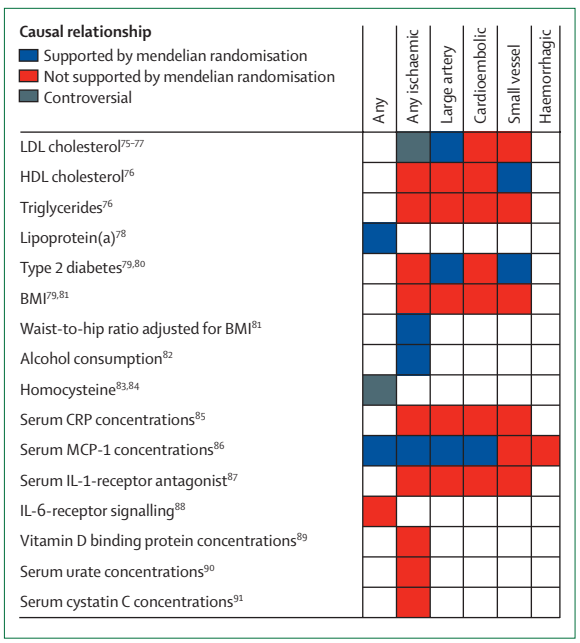


Figure 5: **Mendelian randomisation studies in stroke, by exposure (risk factor)**
Note that absence of support for a causal relationship might also relate to methodological aspects such as limited statistical power. White boxes indicate a lack of available data. BMI=body-mass index. CRP=C-reactive protein. MCP-1=monocyte chemoattractant potein-1. IL-1=interleukin 1. IL-6=interleukin 6.

For **more on MR-Base** see http://www.mrbase.org/

For more on **PhenoScanner** see http://www.phenoscanner.medschl.cam.ac.uk/

Exploiting genetics for drug discovery

Genetics offers great potential for catalysing drug development and prioritising targets for exploration in randomised controlled trials. Aside from identifying causal pathways and novel drug targets, genetics might help in anticipating the full range of safety and efficacy consequences of pharmacological interventions.[11,13] Key approaches aside from mendelian randomisation[71] include the exploitation of protective variants,[96–99] the examination of naturally occurring human knockouts,[100] and phenome-wide association studies (PheWASs; figure 6).[78,101] PheWASs benefit from large datasets (from >200 000 individuals) with detailed genotyping and extensive phenotyping of multiple traits. Such datasets are currently generated through commercial (eg, DeCode genetics), government (eg, UK Biobank), and institutional (eg, Kaiser Permanente Research Bank) funding. By leveraging data from more than 100 000 UK Biobank participants, a PheWAS provided a comprehensive account of the phenotypic consequences of genetically lowered lipoprotein(a) concentrations on various disease states including a reduced risk of stroke.[78] A specific advantage of exquisitely phenotyped longitudinal cohort studies such as the Rotterdam, Framingham, 3C, and Whitehall studies is the quality of phenotyping and completeness of clinical data and the possibility to study association with life course and pre-event phenotype status. The prospects of genetics for drug discovery are highlighted by the observation that risk loci for stroke are substantially enriched in drug-target genes for antithrombotic therapy. Specifically, *FGA* (figure 1) is a target for alteplase and other thrombolytic agents, and *PDE3A* is a target for cilostazol, an antiplatelet drug approved for stroke prevention in Asia.[2] Using genetics for drug discovery and genomic medicine represents a major area of future research.

Monogenic stroke

Advances in mendelian stroke genetics have improved molecular diagnosis, prognosis, counselling, and in some instances prevention or treatment.[8] Diagnostic algorithms should consider the predominant stroke mechanism, mode of inheritance, and presence or absence of systemic manifestations (eg, involving the skin, eye, and skeletal system; table). In most cases, molecular genetic testing remains key to establishing a diagnosis. However, in some conditions, a positive skin biopsy (for CADASIL), laboratory test (for homocystinuria), or detailed clinical examination (for Marfan's syndrome) can be sufficient to establish a diagnosis. In view of falling costs for whole exome and whole genome sequencing, information on rare variants associated with stroke is quickly accumulating. Challenges arising from this information include its interpretation in terms of phenotypic consequences and relevance for disease as well as ethical and medicolegal aspects of handling genetic information.

Conclusions and future directions

The global burden of stroke remains high.[1] Uncovering the biological pathways from genetic variants to stroke

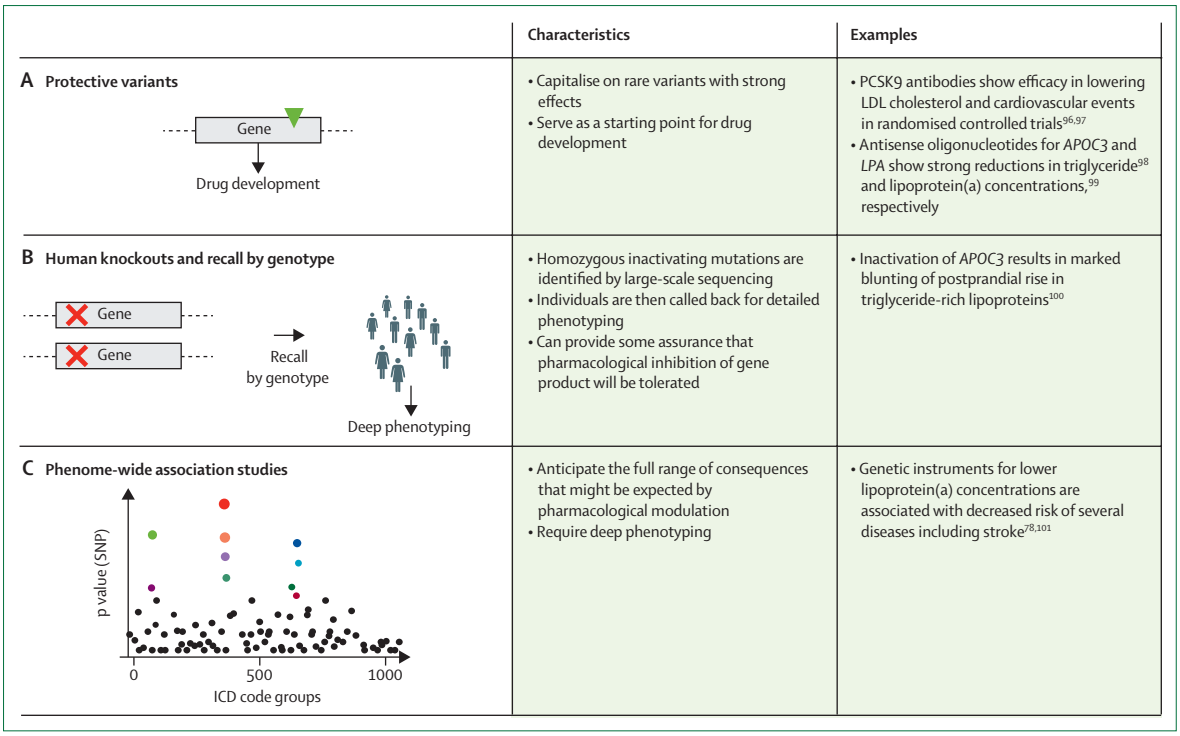


Figure 6: **Genetic approaches that facilitate the discovery, development, and prioritisation of drug targets**
The plot in (C) illustrates the principle of analysing the association between a given SNP and multiple diseases (represented by ICD codes); the coloured dots highlight the top signals for illustrative purposes. SNP=single nucleotide polymorphism. ICD=International Classification of Diseases.

pathology holds the promise of identifying novel targets for intervention.[2,13,14] Genetic information can further be used to improve stroke diagnosis and prognosis.[5,6,23,68] Recent genetic discoveries have expanded the number of genes and mutations proven to cause familial stroke, while GWASs have yielded at least 35 independent loci across the genome with an effect on stroke risk.[2,4,5,7,12,16,18,19,23,67] Functional exploration of these loci will be key to generating drug targets and novel therapies. The application of analytical techniques such as mendelian randomisation facilitates the exploration of causal relationships with exposures and prioritises potential therapeutic targets for stroke.[71,74,86] Aggregation of multiple SNPs from across the genome into polygenic risk scores enables the identification of individuals at high risk for stroke at young age.[6]

Several lines of research could accelerate the discovery of novel biological pathways for stroke. Investigators are working to identify the many undiscovered loci for stroke through the study of substantially larger sample sizes (>100 000 cases).[2,7] Biobanks, such as those in the UK,[9] and studies such as the China Kadoorie Biobank, will vastly expand the opportunity for gene discovery within the next 5 years and the broad sharing of data (eg, the Cerebrovascular Disease Knowledge Portal) will undoubtedly accelerate this process. Genetic discovery will benefit from further refinement of diagnostic categories (eg, cardioembolic stroke of specific origin). However, this will require substantial efforts and resources to ensure a minimum common level of investigation (eg, clinical and radiological characterisation) in large patient samples. A crucial gap that must be filled, however, is the relative absence of cases of non-European ancestry.[2] Thus, ancestry-specific studies will be essential for the development of genetics-derived strategies and tools that are effective across populations.

The development of novel effective drugs, the holy grail of the application of human genetics, is rapidly accelerating in a range of common diseases (eg, type 2 diabetes and coronary artery disease) as a result of the GWAS revolution.[11] Indeed, it has become clear that drugs supported by human genetic data are much more likely to advance to approval by regulatory agencies than those lacking such data.[14] Further progress will require the development of novel cell and tissue models (eg, employing genome editing in human inducible pluripotent stem cells) and advances in functional genomics and multilevel omics to uncover the flow of information in stroke pathophysiology.

Nearly all efforts in human stroke genetics completed thus far have targeted risk of stroke and can be expected to catalyse improvements in stroke prevention. The search for variants implicated in recurrent versus first-ever events provides another avenue for discovery. Equally needed, however, is progress in the understanding and treatment of stroke outcome. Genetic studies

For more on the **China Kadoorie Biobank** see http://www.ckbiobank.org/site/

For more on the **Cerebrovascular Disease Knowledge Portal** see http://www.cerebrovascularportal.org

Search strategy and selection criteria

We identified relevant articles in English for this Review by searching PubMed between Jan 1, 2012, and Dec 31, 2018, and from references cited in relevant articles. We used the search terms "stroke", "intracerebral AND haemorrhage", "genetics", "gene", "variant", "association", "mendelian", "drug", and "personalised medicine". We further checked reference lists of reviews and searched for articles describing the function of genes associated with stroke and of proteins encoded by these genes. The final reference list was made on the basis of relevance to the topics covered in this Review.

focusing on outcome and recovery after stroke are just beginning but are expected to further improve options for clinical applications.[102,103]

Contributors

MD and JR generated the outline of the Review. MD, SLP, and JR drafted the text, provided information, and weighed evidence and information through various discussions. MD prepared the figures. All authors thoroughly revised the manuscript and approved the final version.

Declaration of interests

MD has received grants from the European Union's Horizon 2020 research and innovation programme (grant agreement number 666881; SVDs@target); the Fondation Leducq (Transatlantic Networks of Excellence on the Pathogenesis of Small Vessel Disease of the Brain); the German Research Foundation (DFG; DI 722/13-1; CRC 1123 [B3]); the Munich Cluster for Systems Neurology (SyNergy), Munich, Germany; and the Vascular Dementia Research Foundation. JR has received grants from the US National Institutes of Health (R01NS036695, UM1HG008895, R01NS093870, R24NS092983), and has consulted for New β Innovation, Boehringer Ingelheim, and Pfizer. SLP has received grants from the Dutch Organization for Scientific Research (Nederlandse Organisatie voor Wetenschappelijk Onderzoek, NWO; the Veni Fellowship 016.186.071, ZonMW) and the National Institutes of Health (R01NS100178).

References

1 GBD 2017 Causes of Death Collaborators. Global, regional, and national age-sex-specific mortality for 282 causes of death in 195 countries and territories, 1980–2017: a systematic analysis for the Global Burden of Disease Study 2017. *Lancet* 2018; **392:** 1736–88.

2 Malik R, Chauhan G, Traylor M, et al. Multiancestry genome-wide association study of 520 000 subjects identifies 32 loci associated with stroke and stroke subtypes. *Nat Genet* 2018; **50:** 524–37.

3 Neurology Working Group of the Cohorts for Heart and Aging Research in Genomic Epidemiology (CHARGE) Consortium, the Stroke Genetics Network (SIGN), and the International Stroke Genetics Consortium (ISGC). Identification of additional risk loci for stroke and small vessel disease: a meta-analysis of genome-wide association studies. *Lancet Neurol* 2016; **15:** 695–707.

4 Verdura E, Herve D, Bergametti F, et al. Disruption of a miR-29 binding site leading to *COL4A1* upregulation causes pontine autosomal dominant microangiopathy with leukoencephalopathy. *Ann Neurol* 2016; **80:** 741–53.

5 Verdura E, Herve D, Scharrer E, et al. Heterozygous *HTRA1* mutations are associated with autosomal dominant cerebral small vessel disease. *Brain* 2015; **138:** 2347–58.

6 Rutten-Jacobs LC, Larsson SC, Malik R, et al. Genetic risk, incident stroke, and the benefits of adhering to a healthy lifestyle: follow-up study of 306 473 UK Biobank participants. *BMJ* 2018; **363:** k4168.

7 Malik R, Rannikmae K, Traylor M, et al. Genome-wide meta-analysis identifies 3 novel loci associated with stroke. *Ann Neurol* 2018; **84:** 934–39.

8 Boycott KM, Vanstone MR, Bulman DE, MacKenzie AE. Rare-disease genetics in the era of next-generation sequencing: discovery to translation. *Nat Rev Genet* 2013; **14:** 681–91.

9 Bycroft C, Freeman C, Petkova D, et al. The UK Biobank resource with deep phenotyping and genomic data. *Nature* 2018; **562:** 203–09.

10 Roselli C, Chaffin MD, Weng LC, et al. Multi-ethnic genome-wide association study for atrial fibrillation. *Nat Genet* 2018; **50:** 1225–33.

11 Khera AV, Kathiresan S. Genetics of coronary artery disease: discovery, biology and clinical translation. *Nat Rev Genet* 2017; **18:** 331–44.

12 Bugiani M, Kevelam SH, Bakels HS, et al. Cathepsin A-related arteriopathy with strokes and leukoencephalopathy (CARASAL). *Neurology* 2016; **87:** 1777–86.

13 Plenge RM, Scolnick EM, Altshuler D. Validating therapeutic targets through human genetics. *Nat Rev Drug Discov* 2013; **12:** 581–94.

14 Finan C, Gaulton A, Kruger FA, et al. The druggable genome and support for target identification and validation in drug development. *Sci Transl Med* 2017; **9:** eaag1166.

15 Chabriat H, Joutel A, Dichgans M, Tournier-Lasserve E, Bousser MG. CADASIL. *Lancet Neurol* 2009; **8:** 643–53.

16 Stam AH, Kothari PH, Shaikh A, et al. Retinal vasculopathy with cerebral leukoencephalopathy and systemic manifestations. *Brain* 2016; **139:** 2909–22.

17 Switzer JA, Hess DC, Nichols FT, Adams RJ. Pathophysiology and treatment of stroke in sickle-cell disease: present and future. *Lancet Neurol* 2006; **5:** 501–12.

18 French CR, Seshadri S, Destefano AL, et al. Mutation of *FOXC1* and *PITX2* induces cerebral small-vessel disease. *J Clin Invest* 2014; **124:** 4877–81.

19 Zhou Q, Yang D, Ombrello AK, et al. Early-onset stroke and vasculopathy associated with mutations in *ADA2*. *N Engl J Med* 2014; **370:** 911–20.

20 Roach ES, Islam MP. Pseudoxanthoma elasticum. *Handb Clin Neurol* 2015; **132:** 215–21.

21 Debette S, Goeggel Simonetti B, Schilling S, et al. Familial occurrence and heritable connective tissue disorders in cervical artery dissection. *Neurology* 2014; **83:** 2023–31.

22 Chung MG. Hereditary hemorrhagic telangiectasia. *Handb Clin Neurol* 2015; **132:** 185–97.

23 Jeanne M, Gould DB. Genotype-phenotype correlations in pathology caused by collagen type IV alpha 1 and 2 mutations. *Matrix Biol* 2017; **57–58:** 29–44.

24 Beaufort N, Scharrer E, Kremmer E, et al. Cerebral small vessel disease-related protease HtrA1 processes latent TGF-beta binding protein 1 and facilitates TGF-beta signaling. *Proc Natl Acad Sci USA* 2014; **111:** 16496–501.

25 NINDS Stroke Genetics Network (SiGN), International Stroke Genetics Consortium (ISGC). Loci associated with ischaemic stroke and its subtypes (SiGN): a genome-wide association study. *Lancet Neurol* 2016; **15:** 174–84.

26 Bellenguez C, Bevan S, Gschwendtner A, et al. Genome-wide association study identifies a variant in *HDAC9* associated with large vessel ischemic stroke. *Nat Genet* 2012; **44:** 328–33.

27 Biffi A, Sonni A, Anderson CD, et al. Variants at *APOE* influence risk of deep and lobar intracerebral hemorrhage. *Ann Neurol* 2010; **68:** 934–43.

28 Woo D, Falcone GJ, Devan WJ, et al. Meta-analysis of genome-wide association studies identifies 1q22 as a susceptibility locus for intracerebral hemorrhage. *Am J Hum Genet* 2014; **94:** 511–21.

29 Franceschini N, Giambartolomei C, de Vries PS, et al. GWAS and colocalization analyses implicate carotid intima-media thickness and carotid plaque loci in cardiovascular outcomes. *Nat Commun* 2018; **9:** 5141.

30 Dichgans M, Malik R, König IR, et al. Shared genetic susceptibility to ischemic stroke and coronary artery disease: a genome-wide analysis of common variants. *Stroke* 2014; **45:** 24–36.

31 Nikpay M, Goel A, Won HH, et al. A comprehensive 1,000 genomes-based genome-wide association meta-analysis of coronary artery disease. *Nat Genet* 2015; **47:** 1121–30.

32 Capone C, Dabertrand F, Baron-Menguy C, et al. Mechanistic insights into a TIMP3-sensitive pathway constitutively engaged in the regulation of cerebral hemodynamics. *Elife* 2016; **5:** e17536.

33 Zellner A, Scharrer E, Arzberger T, et al. CADASIL brain vessels show a HTRA1 loss-of-function profile. *Acta Neuropathol* 2018; **136:** 111–25.

34 Ghosh M, Balbi M, Hellal F, Dichgans M, Lindauer U, Plesnila N. Pericytes are involved in the pathogenesis of cerebral autosomal dominant arteriopathy with subcortical infarcts and leukoencephalopathy. *Ann Neurol* 2015; **78:** 887–900.

35 Jeanne M, Jorgensen J, Gould DB. Molecular and genetic analyses of collagen type IV mutant mouse models of spontaneous intracerebral hemorrhage identify mechanisms for stroke prevention. *Circulation* 2015; **131:** 1555–65.

36 Mao M, Alavi MV, Labelle-Dumais C, Gould DB. Type IV collagens and basement membrane diseases: cell biology and pathogenic mechanisms. *Curr Top Membr* 2015; **76:** 61–116.

37 Ratelade J, Mezouar N, Domenga-Denier V, Rochey A, Plaisier E, Joutel A. Severity of arterial defects in the retina correlates with the burden of intracerebral haemorrhage in *COL4A1*-related stroke. *J Pathol* 2018; **244:** 408–20.

38 Schiffmann R. Fabry disease. *Pharmacol Ther* 2009; **122:** 65–77.

39 Osk Snorradottir A, Isaksson HJ, Kaeser SA, et al. Parenchymal cystatin C focal deposits and glial scar formation around brain arteries in hereditary cystatin C amyloid angiopathy. *Brain Res* 2015; **1622:** 149–62.

40 Levy E, Jaskolski M, Grubb A. The role of cystatin C in cerebral amyloid angiopathy and stroke: cell biology and animal models. *Brain Pathol* 2006; **16:** 60–70.

41 Traylor M, Malik R, Nalls MA, et al. Genetic variation at 16q24.2 is associated with small vessel stroke. *Ann Neurol* 2017; **81:** 383–94.

42 Reyahi A, Nik AM, Ghiami M, et al. Foxf2 is required for brain pericyte differentiation and development and maintenance of the blood-brain barrier. *Dev Cell* 2015; **34:** 19–32.

43 Cejudo-Martin P, Yuen A, Vlahovich N, Lock P, Courtneidge SA, Diaz B. Genetic disruption of the *Sh3pxd2a* gene reveals an essential role in mouse development and the existence of a novel isoform of tks5. *PloS One* 2014; **9:** e107674.

44 Charpentier MS, Christine KS, Amin NM, et al. CASZ1 promotes vascular assembly and morphogenesis through the direct regulation of an EGFL7/RhoA-mediated pathway. *Dev Cell* 2013; **25:** 132–43.

45 Malik R, Dau T, Gonik M, et al. Common coding variant in *SERPINA1* increases the risk for large artery stroke. *Proc Natl Acad Sci USA* 2017; **114:** 3613–18.

46 Falcone GJ, Woo D. Genetics of spontaneous intracerebral hemorrhage. *Stroke* 2017; **48:** 3420–24.

47 Lee S, Abecasis GR, Boehnke M, Lin X. Rare-variant association analysis: study designs and statistical tests. *Am J Hum Genet* 2014; **95:** 5–23.

48 Auer PL, Nalls M, Meschia JF, et al. Rare and coding region genetic variants associated with risk of ischemic stroke: the NHLBI Exome Sequence Project. *JAMA Neurol* 2015; **72:** 781–88.

49 Bersenev A, Wu C, Balcerek J, Tong W. Lnk controls mouse hematopoietic stem cell self-renewal and quiescence through direct interactions with JAK2. *J Clin Invest* 2008; **118:** 2832–44.

50 Wang W, Tang Y, Wang Y, et al. LNK/SH2B3 loss of function promotes atherosclerosis and thrombosis. *Circ Res* 2016; **119:** e91–103.

51 Duan L, Wei L, Tian Y, et al. Novel susceptibility loci for moyamoya disease revealed by a genome-wide association study. *Stroke* 2018; **49:** 11–18.

52 Azghandi S, Prell C, van der Laan SW, et al. Deficiency of the stroke relevant *HDAC9* gene attenuates atherosclerosis in accord with allele-specific effects at 7p21.1. *Stroke* 2015; **46:** 197–202.

53 Cao Q, Rong S, Repa JJ, St Clair R, Parks JS, Mishra N. Histone deacetylase 9 represses cholesterol efflux and alternatively activated macrophages in atherosclerosis development. *Arterioscler Thromb Vasc Biol* 2014; **34:** 1871–79.

54 Lino Cardenas CL, Kessinger CW, Cheng Y, et al. An HDAC9-MALAT1-BRG1 complex mediates smooth muscle dysfunction in thoracic aortic aneurysm. *Nat Commun* 2018; **9:** 1009.

55 Gschwendtner A, Bevan S, Cole JW, et al. Sequence variants on chromosome 9p21.3 confer risk for atherosclerotic stroke. *Ann Neurol* 2009; **65:** 531–39.

56 Motterle A, Pu X, Wood H, et al. Functional analyses of coronary artery disease associated variation on chromosome 9p21 in vascular smooth muscle cells. *Hum Mol Genet* 2012; **21:** 4021–29.

57 Almontashiri NA, Antoine D, Zhou X, et al. 9p21.3 coronary artery disease risk variants disrupt TEAD transcription factor-dependent transforming growth factor beta regulation of p16 expression in human aortic smooth muscle cells. *Circulation* 2015; **132:** 1969–78.

58 Kojima Y, Downing K, Kundu R, et al. Cyclin-dependent kinase inhibitor 2B regulates efferocytosis and atherosclerosis. *J Clin Invest* 2014; **124:** 1083–97.

59 Holdt LM, Stahringer A, Sass K, et al. Circular non-coding RNA ANRIL modulates ribosomal RNA maturation and atherosclerosis in humans. *Nat Commun* 2016; **7:** 12429.

60 Traylor M, Makela KM, Kilarski LL, et al. A novel MMP12 locus is associated with large artery atherosclerotic stroke using a genome-wide age-at-onset informed approach. *PLoS Genet* 2014; **10:** e1004469.

61 Gupta RM, Hadaya J, Trehan A, et al. A genetic variant associated with five vascular diseases is a distal regulator of endothelin-1 gene expression. *Cell* 2017; **170:** 522–33.e15.

62 Barton M, Haudenschild CC, d'Uscio LV, Shaw S, Munter K, Luscher TF. Endothelin ETA receptor blockade restores NO-mediated endothelial function and inhibits atherosclerosis in apolipoprotein E-deficient mice. *Proc Natl Acad Sci USA* 1998; **95:** 14367–72.

63 Germain M, Chasman DI, de Haan H, et al. Meta-analysis of 65734 individuals identifies *TSPAN15* and *SLC44A2* as two susceptibility loci for venous thromboembolism. *Am J Hum Genet* 2015; **96:** 532–42.

64 Rannikmae K, Davies G, Thomson PA, et al. Common variation in *COL4A1/COL4A2* is associated with sporadic cerebral small vessel disease. *Neurology* 2015; **84:** 918–26.

65 Traylor M, Zhang CR, Adib-Samii P, et al. Genome-wide meta-analysis of cerebral white matter hyperintensities in patients with stroke. *Neurology* 2016; **86:** 146–53.

66 Weng YC, Sonni A, Labelle-Dumais C, et al. *COL4A1* mutations in patients with sporadic late-onset intracerebral hemorrhage. *Ann Neurol* 2012; **71:** 470–77.

67 Biffi A, Anderson CD, Jagiella JM, et al. *APOE* genotype and extent of bleeding and outcome in lobar intracerebral haemorrhage: a genetic association study. *Lancet Neurol* 2011; **10:** 702–09.

68 Malik R, Bevan S, Nalls MA, et al. Multilocus genetic risk score associates with ischemic stroke in case-control and prospective cohort studies. *Stroke* 2014; **45:** 394–402.

69 Khera AV, Chaffin M, Aragam KG, et al. Genome-wide polygenic scores for common diseases identify individuals with risk equivalent to monogenic mutations. *Nat Genet* 2018; **50:** 1219–24.

70 Smith GD, Ebrahim S. 'Mendelian randomization': can genetic epidemiology contribute to understanding environmental determinants of disease? *Int J Epidemiol* 2003; **32:** 1–22.

71 Holmes MV, Ala-Korpela M, Smith GD. Mendelian randomization in cardiometabolic disease: challenges in evaluating causality. *Nat Rev Cardiol* 2017; **14:** 577–90.

72 Voight BF, Peloso GM, Orho-Melander M, et al. Plasma HDL cholesterol and risk of myocardial infarction: a mendelian randomisation study. *Lancet* 2012; **380:** 572–80.

73 Ference BA, Yoo W, Alesh I, et al. Effect of long-term exposure to lower low-density lipoprotein cholesterol beginning early in life on the risk of coronary heart disease: a mendelian randomization analysis. *J Am Coll Cardiol* 2012; **60:** 2631–39.

74 Ference BA, Majeed F, Penumetcha R, Flack JM, Brook RD. Effect of naturally random allocation to lower low-density lipoprotein cholesterol on the risk of coronary heart disease mediated by polymorphisms in *NPC1L1, HMGCR*, or both: a 2x2 factorial mendelian randomization study. *J Am Coll Cardiol* 2015; **65:** 1552–61.

75 Beheshti S, Madsen CM, Varbo A, Benn M, Nordestgaard BG. Relationship of familial hypercholesterolemia and high low-density lipoprotein cholesterol to ischemic stroke. *Circulation* 2018; **138:** 578–89.

76 Hindy G, Engstrom G, Larsson SC, et al. Role of blood lipids in the development of ischemic stroke and its subtypes: a mendelian randomization study. *Stroke* 2018; **49:** 820–27.

77 Hopewell JC, Malik R, Valdes-Marquez E, Worrall BB, Collins R, METASTROKE Collaboration of the ISGC. Differential effects of *PCSK9* variants on risk of coronary disease and ischaemic stroke. *Eur Heart J* 2018; **39:** 354–59.

78 Emdin CA, Khera AV, Natarajan P, et al. Phenotypic characterization of genetically lowered human lipoprotein(a) levels. *J Am Coll Cardiol* 2016; **68:** 2761–72.

79 Larsson SC, Scott RA, Traylor M, et al. Type 2 diabetes, glucose, insulin, BMI, and ischemic stroke subtypes: mendelian randomization study. *Neurology* 2017; **89:** 454–60.

80 Liu J, Rutten-Jacobs L, Liu M, Markus HS, Traylor M. Causal impact of type 2 diabetes mellitus on cerebral small vessel disease: a mendelian randomization analysis. *Stroke* 2018; **49:** 1325–31.

81 Dale CE, Fatemifar G, Palmer TM, 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* 2017; **135:** 2373–88.

82 Holmes MV, Dale CE, Zuccolo L, et al. Association between alcohol and cardiovascular disease: mendelian randomisation analysis based on individual participant data. *BMJ* 2014; **349:** g4164.

83 Casas JP, Bautista LE, Smeeth L, Sharma P, Hingorani AD. Homocysteine and stroke: evidence on a causal link from mendelian randomisation. *Lancet* 2005; **365:** 224–32.

84 Holmes MV, Newcombe P, Hubacek JA, et al. Effect modification by population dietary folate on the association between *MTHFR* genotype, homocysteine, and stroke risk: a meta-analysis of genetic studies and randomised trials. *Lancet* 2011; **378:** 584–94.

85 Prins BP, Abbasi A, Wong A, et al. Investigating the causal relationship of C-reactive protein with 32 complex somatic and psychiatric outcomes: a large-scale cross-consortium mendelian randomization study. *PLoS Med* 2016; **13:** e1001976.

86 Georgakis MK, Gill D, Rannikmäe K, et al. Genetically determined levels of circulating cytokines and risk of stroke: role of monocyte chemoattractant protein-1. *Circulation* 2018; **139:** 256–68.

87 Interleukin 1 Genetics Consortium. Cardiometabolic effects of genetic upregulation of the interleukin 1 receptor antagonist: a mendelian randomisation analysis. *Lancet Diabetes Endocrinol* 2015; **3:** 243–53.

88 Interleukin-6 Receptor Mendelian Randomisation Analysis Consortium, Swerdlow DI, Holmes MV, et al. The interleukin-6 receptor as a target for prevention of coronary heart disease: a mendelian randomisation analysis. *Lancet* 2012; **379:** 1214–24.

89 Leong A, Rehman W, Dastani Z, et al. The causal effect of vitamin D binding protein (DBP) levels on calcemic and cardiometabolic diseases: a mendelian randomization study. *PLoS Med* 2014; **11:** e1001751.

90 Keenan T, Zhao W, Rasheed A, et al. Causal assessment of serum urate levels in cardiometabolic diseases through a mendelian randomization study. *J Am Coll Cardiol* 2016; **67:** 407–16.

91 van der Laan SW, Fall T, Soumare A, et al. Cystatin C and cardiovascular disease: a mendelian randomization study. *J Am Coll Cardiol* 2016; **68:** 934–45.

92 Lee SJ, Jee YH, Jung KJ, Hong S, Shin ES, Jee SH. Bilirubin and stroke risk using a mendelian randomization design. *Stroke* 2017; **48:** 1154–60.

93 Hopewell JC, Clarke R. Emerging risk factors for stroke: what have we learned from mendelian randomization studies? *Stroke* 2016; **47:** 1673–78.

94 O'Donnell MJ, Chin SL, Rangarajan S, et al. Global and regional effects of potentially modifiable risk factors associated with acute stroke in 32 countries (INTERSTROKE): a case-control study. *Lancet* 2016; **388:** 761–75.

95 Davey Smith G, Paternoster L, Relton C. When will mendelian randomization become relevant for clinical practice and public health? *JAMA* 2017; **317:** 589–91.

96 Cohen JC, Boerwinkle E, Mosley TH Jr, Hobbs HH. Sequence variations in *PCSK9*, low LDL, and protection against coronary heart disease. *N Engl J Med* 2006; **354:** 1264–72.

97 Sabatine MS, Giugliano RP, Keech AC, et al. Evolocumab and clinical outcomes in patients with cardiovascular disease. *N Engl J Med* 2017; **376:** 1713–22.

98 Gaudet D, Alexander VJ, Baker BF, et al. Antisense inhibition of apolipoprotein C-III in patients with hypertriglyceridemia. *N Engl J Med* 2015; **373:** 438–47.

99 Tsimikas S, Viney NJ, Hughes SG, et al. Antisense therapy targeting apolipoprotein(a): a randomised, double-blind, placebo-controlled phase 1 study. *Lancet* 2015; **386:** 1472–83.

100 Saleheen D, Natarajan P, Armean IM, et al. Human knockouts and phenotypic analysis in a cohort with a high rate of consanguinity. *Nature* 2017; **544:** 235–39.

101 Bush WS, Oetjens MT, Crawford DC. Unravelling the human genome-phenome relationship using phenome-wide association studies. *Nat Rev Genet* 2016; **17:** 129–45.

102 Maguire JM, Bevan S, Stanne TM, et al. GISCOME – Genetics of Ischaemic Stroke Functional Outcome network: a protocol for an international multicentre genetic association study. *Eur Stroke J* 2017; **2:** 9.

103 Mola-Caminal M, Carrera C, Soriano-Tarraga C, et al. *PATJ* low frequency variants are associated with worse ischemic stroke functional outcome. *Circ Res* 2018; **124:** 114–20.