Elsevier Editorial System(tm) for The Lancet Neurology

Manuscript Draft

Manuscript Number: THELANCETNEUROLOGY-D-18-01037R2

Title: Stroke genetics: discovery, biology, and clinical applications

Article Type: Review

Corresponding Author: Professor Martin Dichgans,

Corresponding Author's Institution: University of Munich

First Author: Martin Dichgans

Order of Authors: Martin Dichgans; Sara Pulit; Jonathan Rosand

Abstract: Abstract
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 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. Further, genetics has successfully established causal relationships with risk factors and holds promise for prioritizing 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, ultimately, the conversion of genetic discoveries to novel therapies.

Stroke genetics: discovery, biology, and clinical applications

Martin Dichgans, MD[1,2,3] Sara L Pulit, PhD4[,5,6] Jonathan Rosand, MD, MSc[6,7,8]

[1]Institute for Stroke and Dementia Research, Klinikum der Universität München, Ludwig-Maximilians-Universität LMU, Munich, Germany

[2]German Center for Neurodegenerative Diseases (DZNE, Munich), Munich, Germany

[3]Munich Cluster for Systems Neurology (SyNergy), Munich, Germany

[4]Department of Genetics, Center for Molecular Medicine, University Medical Center Utrecht, Utrecht, The Netherlands

[5]Big Data Institute, Li Ka Shing Center for Health Information and Discovery, Oxford University, Oxford, UK

[6]Program in Medical Population and Genetics, Broad Institute, Cambridge, MA, USA

[7]Henry and Allison McCance Center for Brain Health, Massachusetts General Hospital, Boston, MA, USA

[8]Center for Genomic Medicine, Massachusetts General Hospital, Boston, MA, USA

Correspondence to:

Martin Dichgans, Institute for Stroke and Dementia Research; Klinikum der Universität München; Feodor-Lynen-Straße 17, D-81377 München, Germany

Key words: stroke; hereditary; common, rare, risk prediction, precision medicine

Abstract

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 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. Further, genetics has successfully established causal relationships with risk factors and holds promise for prioritizing 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, ultimately, the conversion of genetic discoveries to novel therapies.

Introduction

Stroke remains a leading cause of death and long-term disability across the globe.[1] Despite the discovery of modifiable and non-modifiable risk factors as well as effective treatments novel therapeutic approaches are urgently needed. Uncovering the genetic contributions to stroke promises a better definition of causal pathways, the identification of novel therapeutic targets, and improved options for diagnosis and prognostication.[2-7]

The last five years have seen dramatic advances in genomic technologies, sequencing costs, biobanking, and data sharing, which collectively have accelerated genetic discovery.[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 (GWAS) in stroke and other vascular traits such as blood pressure and atrial fibrillation have tested over 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, have begun to elucidate causal pathways, relevant cell and tissue types, and, in some instances, yielded novel drug targets.[2]

Here, we review the latest discoveries in stroke genetics. In particular, we discuss the identification of novel Mendelian causes of stroke,[4,5,11] the discovery of at least 35 stroke risk loci harboring 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 further discuss how genetic discoveries could improve diagnosis, risk prediction, and the treatment of stroke. Indeed, recent observations have elucidated how genetics could be leveraged to discover novel drug targets and identify high-risk individuals many years before the emergence of classical indicators of stroke risk. [2,6,12,13] Because of their substantially different pathophysiologies, the following conditions were not considered: subarachnoid haemorrhage, cerebral aneurysms, cavernous malformations, dissections, and cerebral venous thrombosis.

Genetic discovery for stroke

Family-based studies and Mendelian stroke

Advances in sequencing technology have facilitated the discovery of single-gene disorders associated with stroke beyond classical syndromes such as CADASIL and sickle-cell disease (**Table 1**). Most notably, there has been a substantial expansion of the spectrum of ischemic

small vessel disease (SVD), which may manifest with ischemic stroke ('small vessel stroke', **Panel**), cognitive decline, and other manifestations. Among the most recent 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 SVD that typically manifests with early onset ischemic stroke.[4] These mutations disrupt a binding site for miR-29 and were shown to upregulate *COL4A1* mRNA expression. Heterozygous mutations (in particular, glycine substitutions) in the triple helical domains of COL4A1 or COL4A2 cause a different syndrome characterized by hemorrhagic stroke[14] along with additional neurological and non-neurological manifestations (**Table 1**).[4,5,11,14-23] Sequencing has further pinpointed heterozygous mutations in *HTRA1* (encoding high temperature requirement serine protease A1, HTRA1) in families with autosomal dominant SVD.[5] The condition typically manifests with stroke and cognitive decline in mid to late adulthood with more and more cases reported over the last years. In 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 than the syndrome of heterozygous mutation carriers and further features non-neurological symptoms.

Hereditary SVD syndromes typically presenting during childhood include deficiency of adenosine deaminase 2 (DADA2), an autoinflammatory disease manifesting with small vessel vasculitis caused by *CECR1* mutations.[20] Recently described mutations in *CTSA* (encoding cathepsin A)[11] and *FOXC1* (encoding forkhead box C1)[19] are associated with young onset autosomal dominant SVD.

Common sporadic stroke

Sporadic stroke arises through multiple risk factors and mechanisms. It is broadly classified as ischemic (large artery stroke, cardioembolic stroke, and small vessel stroke) and hemorrhagic (deep and lobar intracerebral haemorrhage [ICH]). 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 responsible stroke mechanism remains uncertain either because diagnostic work-up fails to find an established stroke etiology or because there are multiple competing etiologies. This complexity, along with the existence of different algorithms for stroke classification (discussed in [25]), has posed challenges to unravelling the genetic underpinnings of sporadic multifactorial stroke.

Common genetic variants associated with stroke

Common genetic variants, defined here as an allele frequency of ≥0.5% (one carrier per 100 individuals), have been associated with stroke using genome-wide association studies (GWAS), which compare the frequency of variants (mostly single nucleotide variants, SNPs) between groups of individuals. The first successful stroke GWAS included 3,548 patients and 5,972 controls and identified common variants at *HDAC9* (encoding histone deacetylase 9), that conferred an ~40% increased risk for large artery stroke per copy of the risk allele.[26] Since then, GWAS in progressively larger sample sizes have identified at least 35 loci with robust links to stroke risk.[2,27,28] Collectively, these studies enable the following conclusions: First, the vast majority of associated SNPs have a minor allele frequency of >5% and are associated with a modest increase in stroke risk (typically < 30% increase per allele) (**Table 2**). Second, most (~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 (AS), any ischemic stroke (AIS), large artery stroke, cardioembolic stroke, small vessel stroke, and ICH, both lobar and deep ICH (**Table 2,** and [27,28]). Fourth, at several loci, the association is restricted to a specific etiological stroke subtype. For instance, 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 ($p<5 \times 10^{-8}$) for large artery stroke but showed no association (p>0.05) with other stroke subtypes, implying mechanisms limited to atherosclerosis. Fifth, many genetic variants that confer risk of stroke also influence risk of related traits (**Figure 1**). Notably, about a quarter of the 32 genome-wide significant loci identified by MEGASTROKE are established risk loci for high blood pressure. Other loci (e.g., near *SMARCA4-LDLR*) are established risk loci for lipid levels.[2] Also, there is genetic overlap between large artery stroke and the related trait of coronary artery disease (CAD) (e.g. near *HDAC9-TWIST1*, *EDNRA*, and *9p21*)[2,29,30] and the intermediate phenotypes of carotid intima media thickness (IMT) and carotid plaque (e.g., at EDNRA)[2,31]. Sixth, several stroke risk loci contain genes previously implicated in monogenic SVD (e.g. *COL4A1*, *COL4A2*, and *HTRA1*)[2,7] (**Table 3**) suggesting a biological overlap between common stroke and rare familial stroke syndromes. And finally, while several risk loci integrate into previously suspected biological pathways for stroke, roughly one third show no obvious relationship with known pathways, pointing to mechanisms not previously implicated in stroke pathophysiology.

The heritability of stroke (i.e. the proportion of variation in risk attributable to inherited genetic variation) calculated from genome-wide data has been estimated to be 30%-40%[32,33] 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 to other common conditions including CAD. This is partly due to the complexity of stroke as a phenotype as well as the smaller sample sizes that are as yet available for etiologically-defined subtypes in recent GWAS. Not surprisingly, the lead variants at stroke risk loci identified to date explain roughly 1 to 2% of stroke heritability,[2] which is markedly lower than published figures for other traits.[30] Given the impact 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 this 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.[34] The two largest stroke studies[32,35] to date focused on coding regions (the exome) rather than the whole genome. While 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 (see below), 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, which often vary from locus to locus. These challenges notwithstanding, recent functional genomics studies have provided initial insights into the mechanisms linking common variants with stroke risk. In the following, we highlight key observations on biological mechanisms derived from recent genetic discoveries starting with all stroke and then moving to etiological 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 (P262W) in the lymphocyte adaptor protein LNK (the protein

encoded by *SH2B3)* and also associates with hypertension, type 1 diabetes, CAD, platelet counts, and leukocytosis, collectively suggesting an involvement of this locus in multiple mechanisms relevant to stroke pathophysiology. Genetic variation at *SH2B3* associates with both the broader phenotypes of AS, AIS, and with the etiologically defined stroke subtypes large artery stroke and small vessel stroke, providing an example of shared genetic influences between etiological subtypes.[2] *LNK* is predominantly expressed in hematopoietic and endothelial cells and negatively regulates cytokine signaling and cell proliferation.[36] *LNK* deficiency is associated with increased platelet production and activation, accelerated arterial thrombosis and atherosclerosis in hypercholesterolemic mice,[37] and other mechanisms under investigation. Collectively, these findings highlight the potential of genetics to uncover disease mechanisms while also illustrating challenges of genetic epidemiology as *SH2B3* has also been associated with traits whose link to stroke is more difficult to explain.

A substantial proportion of stroke risk variants associate with the mRNA expression, methylation, or protein levels of nearby genes primarily in stroke-relevant tissues (vascular, brain) and cell types (e.g., endothelial cells, blood, and immune cells) thus emphasizing 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 (e.g., 'enlarged heart' and 'cardiomyocyte differentiation via bone-morphogenetic-protein receptors') the coagulation system, and nitric oxide (NO) metabolism but also other pathways.[2]

Large artery stroke

The strongest association signal for large artery stroke identified to date is near *HDAC9*[2,26,32] a locus that also associates with CAD (**Figure 1**), peripheral artery disease, and moyamoya disease.[2,30,38] rs2109595, the lead SNP for stroke and likely causal variant[2] resides in regulatory DNA 3' to *HDAC9*. Risk variants associate with elevated *HDAC9* expression levels in blood cells with a gene dosage effect,[39] and deficiency of Hdac9 attenuates atherosclerosis in experimental mice.[39,40] Studies in $Ldlr^{-/-}$ mice suggest a pro-atherogenic effect of *Hdac9* in macrophages via changes in their activation status and cholesterol efflux[40] although *HDAC9* is also expressed in other cell types relevant to atherosclerosis.[41] 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 CAD but subsequently shown to also associate with peripheral

artery disease, intracranial and abdominal aortic aneurysms, carotid plaque,[31] and stroke, in particular large artery stroke.[2,42] Risk variants at 9p21 (i) are responsible for up to 20% of the population attributable risk of large artery stroke,[2,42] (ii) are considered to mediate their effects through mechanisms that are largely independent from established risk factors for atherosclerosis, and (iii) likely act through several genes at this locus. Specifically, these variants are associated with reduced expression of the cell cycle suppressor genes *CDKN2A* and *CDKN2B*[43,44] and with vascular smoot muscle cell (SMC) proliferation and SMC content in human atherosclerotic plaques.[43,44] *CDKN2B* has been implicated in SMC physiology and clearance of apoptotic debris, and mice deficient of *Cdkn2b* develop advanced atherosclerotic lesions composed of large, lipid-laden necrotic cores[45]. Altogether, the available evidence suggests a role of 9p21 in vascular remodeling, a key process in atherosclerosis. Importantly, CDKN2A/B and a circular non-coding antisense RNA (circANRIL) that is also transcribed at 9p21[46] have been suggested as possible targets for preventive strategies.

Recent GWAS further found variants near *TSPAN2*,[2,25] *LINC01492*,[2] *EDNRA*,[2] and *MMP12*[2,47] to be associated with large artery stroke. *EDNRA* and *MMP12* offer potential mechanistic insights: *ENDRA* (encoding endothelin receptor A, ET_A) is expressed in SMCs, endothelial cells, and macrophages. Variants at *EDNRA* also associate with carotid IMT (reflecting early atherosclerosis),[31] CAD[30] (**Figure 1**), and intracranial aneurysms, hence suggesting a broader role of this gene and Endothelin-1 signaling in vascular disease.[48] Activation of ET_A has effects on vasoconstriction, SMC proliferation, extracellular matrix (ECM) production, and fibrosis,[48] i.e., processes relevant to atherosclerosis. Moreover, pharmacological inhibition of ET_A normalized NO-mediated endothelium-dependent dysfunction[49] and attenuated atheroma formation in atherosclerotic mice[49] thus offering a perspective for alternative preventive strategies. *MMP12* (encoding matrix metalloproteinase 12, *MMP12*) was originally identified by an age-at-onset informed approach.[47] MMPs degrade ECM proteins, show increased activity in atherosclerotic plaques and have suspected roles in the growth, destabilization, and rupture of atherosclerotic lesions. Collectively, these findings suggest a causal role of the MMP12 protease in atherosclerosis (reviewed in [47]). Additional mechanism supported by recent stroke GWAS include changes in lipoprotein(a) metabolism (evidenced by *LPA*, a subthreshold locus for large artery stroke [p=1.3E-7][2] known to be implicated in atherosclerosis[50]) and thromboembolism (evidenced by *FGA*, *ILF3-SLC44A2*, and *ABO*, established risk loci for venous thromboembolism[51]). Hence, risk loci for large artery stroke highlight different aspects of the biology underlying large artery stroke (**Figure 1**).

These observations highlight different biological pathways underpinning large artery stroke. 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 exhibits close links with the genetics of atrial fibrillation (AF): (i) all loci reaching genome-wide significance for common variant association with cardioembolic stroke (nearest genes *PITX2*, *ZFHX2*, *NKX2-5*) are established risk loci for AF[10]; (ii) at all loci the association signal for stroke is confined to cardioembolic stroke; (iii) the association signals here and at subthreshold loci for cardioembolic stroke (near *CAV1/2* and *PRRX1*)[2] overlap with those for AF, and (iv) show a similar architecture (e.g. near *PITX2).* In addition, there are strong genetic links with venous thromboembolism: loci near *ABO*, *FGA* (encoding fibrinogen alpha), and *F11* (a subthreshold locus for cardioembolic stroke [$p=5.2 \times 10^{-8}$], encoding factor XI) are established risk loci for venous thromboembolism[51] and are known to modulate haemostatic traits.[51] Interestingly, genetic risk for the broader phenotype of AS or AIS further shows links with cardiac mechanisms beyond those implicated in AF as exemplified by the associations near *LRCH1* and *ANK2*, found in cardiac pacing and familial forms of cardiac disease (discussed in [2]).

Small vessel stroke

The identification of genes implicated in monogenic SVD has been instrumental to understanding the biology of small vessel stroke and SVD. Experimental studies in mutant mice and cultured cells highlight key physiological and pathological processes linked to genes implicated in monogenic SVD (**Table 3**).[11,14,15,24,52-63] Recurrent themes include TGF-ß signaling,[24,52] ECM formation/biology,[11,14,15,53-56] blood-brain barrier (BBB) function,[52,57,58] and specific cellular constituents of brain microvessels, in particular endothelial cells, pericytes, and SMCs (**Table 3**).[56,59,60] Proteomics studies of isolated brain microvessels from mutation carriers and mouse models further revealed an unexpected molecular link between CADASIL and CARASIL that involves accumulation of the HTRA1 protease and its substrates in microvessels from NOTCH3 mutation carriers.[54]

Common variant association studies identified various risk loci that show predominant association with small vessel stroke (**Table 2**). *ZCCHC14*, *FOXF2*, and *CASZ1* encode transcription factors expressed in the vasculature.[52,61,62] Foxf2 (expressed in brain endothelial cells and pericytes) is required for brain pericyte differentiation and development and maintenance of the BBB (**Table 3**).[52] Most notably, Foxf2 deficient mice develop microhemorrhages and neuronal loss thus recapitulating features of human SVD.[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]) also associate with radiological white matter hyperintensities (WMH), thus highlighting shared biology of small vessel stroke and WMH. *PMF1–SEMA4A* further reached genome-wide significance for association with deep (nonlobar) ICH,[28] consistent with nonlobar ICH being part of the phenotypic spectrum of SVD. Interestingly, 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 SVD-related phenotypes.[3,19]

Hemorrhagic stroke

ICH 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 sporadic ICH identified coding variants with *in vitro* pathological consequences resembling familial mutations in the same genes.[66] GWAS of ICH, on the other hand, have identified associations in the region of *COL4A2* in nonlobar ICH, as well as small vessel stroke and WMH, highlighting the shared pathways underlying the various manifestations of cerebral SVD. The shared biology between ICH and small vessel stroke is also underscored by associations at *PMF1/SLC25A44*.[2,28] An additional important observation has been that variation associated with risk of ICH can also affect the extent of bleeding, as measured by haematoma volume,[67] providing support for hypotheses generated from histopathologic data that ICH 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 (PRS) enables ascertainment of high-risk individuals at young age thus offering opportunities for early prevention.[68,69] For instance, a recent study found that applying a PRS consisting of 90 SNPs

to data from UK Biobank (>300,000 subjects) could identify individuals with a 35% increased risk of incident stroke (hazard ratio: 1.35). This cutoff included a third of the population.[6] Compared to individuals in the bottom tertile of the PRS who had a favourable lifestyle (defined as 3 or 4 healthy lifestyle factors) individuals in the upper tertile of the PRS had a relative risk (hazard ratio) of 1.44, 1.70 and 2.30 depending on whether they had a favourable, intermediate (2 healthy lifestyle factors) or unfavorable (0 or 1 healthy lifestyle factors) lifestyle, respectively. Indeed, lifestyle risk was similar across all PRS strata, thus highlighting the potential for early risk stratification and prevention via genetics.[6] 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 unfavorable consequences including costs and psychological distress.[69]

Exploration of potential therapeutic targets by Mendelian Randomisation

An increasingly recognized opportunity for genetics is the exploration of causal relationships between risk factors (exposures) and disease outcomes through Mendelian Randomisation (MR).[70,71] MR exploits genetic variants causally related to an exposure as instruments, and investigates their associations with a disease (**Panel** and **Figure 2**).[72-89] Because individual alleles are allocated at random during meiotic segregation, MR shares many features with randomized controlled trials (RCTs). Like RCTs, MR overcomes limitations of observational studies, in particular bias resulting from conventional confounding or reverse causality.

An important requirement for MR studies is a robust association between the genetic instrument (single SNP or multiple variants) and the exposure of interest. For instance, an intronic variant (rs6511720) in the low-density lipoprotein (LDL) receptor (*LDLR*) gene has consistently been shown to associate with LDL cholesterol (LDLC) levels. Like other variants known to influence LDLC, rs6511720 further associates with risk of myocardial infarction in a concordant fashion.[90] MR studies have demonstrated a dose-response relationship between genetically elevated LDLC levels and CAD[91] consistent with trials involving statins and other cholesterol lowering interventions. Conversely, MR studies examining HDL as an exposure found no causal effect of genetically determined HDL levels on the risk of CAD again in accordance with RCTs (reviewed in [71]).

MR studies may predict the success or failure of RCTs, thus saving on risks for study participants and on costs.[71] Examples, in which MR studies predicted the success or failure of RCTs include studies on inactivating mutations in *NPC1L1* (encoding the drug target for ezetimibe)[92] and variants in *PLA2G7* (encoding lipoprotein-associated phospholipase A2, the drug target for darapladib, reviewed in [71]), respectively.

Only recently, investigators have started to apply MR to stroke.[93] A MR study on blood lipids found genetically elevated LDLC levels associate with risk of large artery stroke but not small vessel stroke and cardioembolic stroke.[73] In contrast, genetically elevated HDL cholesterol associated with a reduced risk of small vessel stroke and no effects on large artery stroke and cardioembolic stroke. Differential effects on etiological stroke subtypes were further seen for T2D in that genetically defined T2D was found to associate with both small vessel stroke and large artery stroke but not cardioembolic stroke and ICH.[76,77] A MR study on the waist-to-hip ratio adjusted for body mass index (BMI) found waist-to-hip ratio to be causally related to a higher risk of ischemic stroke. This finding adds to observational data emphasizing the need to consider measures of adiposity beyond BMI for risk prediction and patient management.[94] MR may further identify novel risk factors and potential drug targets for stroke as illustrated by a recent study that found genetically elevated levels of the inflammatory cytokine MCP-1/CCL2 to be associated with risk of stroke.[83] **Table 4** provides and overview on notable MR studies in stroke.

While potentially highly informative, MR analyses also pose specific challenges that require consideration.[71] MR 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 2**). 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 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 MR results (**Table 4**). With increasing availability of large-scale genetic data and improved databases, MR studies will become even more informative and relevant for clinical practice.[50,95]

Exploiting genetics for drug discovery

Genetics holds great promise for catalyzing drug development and prioritizing targets for exploration in RCTs. Aside from identifying causal pathways and novel drug targets, genetics may help in anticipating the full range of efficacy and safety consequences of pharmacological interventions.[12,50] Key approaches aside from MR[71] include the exploitation of protective variants,[96-99] the examination of naturally occurring human knockouts,[100] and phenome-wide association studies (PheWAS)[75,101] (**Figure 3**)**.** PheWAS studies benefit from large datasets with detailed genotyping and extensive phenotyping of multiple traits. Such

datasets are currently generated through commercial (e.g. DECODE genetics), government (e.g. UK Biobank), and institutional (e.g. Kaiser Permanente research bank) funding. By leveraging data from >100.000 UK Biobank participants a recent PheWAS study provided a comprehensive account of the phenotypic consequences of genetically lowered Lp(a) levels on various disease states including a reduced risk of stroke.[75] 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 recent observation that risk loci for stroke are significantly enriched in drug-target genes for antithrombotic therapy. Specifically, *FGA* (**Table 2**) is a target for alteplase and other thrombolytic agents, and *PDE3A* is a target for cilostazol an antiplatelet agent approved for stroke prevention in Asia.[2] Utilizing genetics for drug discovery and genomic medicine represents a major area of future research.

Monogenic stroke

Advances in Mendelian stroke genetics have improved options for molecular diagnosis, prognostication, counseling, 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 (e.g., involving the skin, eye, and skeletal system) (**Table 1**). In most instances, 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 Syndrome) may suffice. In light 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 their 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 pathology holds the promise of identifying novel targets for intervention.[2,12,13] Genetic information can further be used to improve stroke diagnosis and prognostication.[5,6,14,68] Recent discoveries have expanded the number of genes and mutations proven to cause familial stroke, while GWAS have yielded at least 35 independent loci across the genome with an impact on stroke risk.[4,5,7,11,14,17,19,20,68] Functional exploration of these loci will be key to generating drug targets and novel therapies. The application of analytic techniques such as MR facilitates the exploration of causal relationships with

exposures and prioritize potential therapeutic targets for stroke.[71,83,92] Aggregation of multiple SNPs from across the genome into PRS enables the identification of individuals at high risk for stroke at young age.[6]

Several lines of research promise to 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.[2,7] Biobanks, such as those of the UK,[9] and studies such as China-Kadoorie, will vastly expand the opportunity for gene discovery in the near future and the broad sharing of data will undoubtedly accelerate the process (see www.cerebrovascularportal.org). Genetic discovery will benefit from further refinement of diagnostic categories (e.g. cardioembolic stroke of specific origin). However, this will require resources to ensure a minimum common level of investigation in large patient samples. A crucial gap that must be filled, however, is the relative absence of cases of non-European ancestry.[2] Along these same lines, ancestry-specific studies will be vital 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 as a result of the GWAS revolution.[50] 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.[13] Further progress will require the development of novel cell and tissue models (e.g., employing genome editing in human inducible pluripotent stem cells) and advances in functional genomics and multi-level 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 ultimately 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 focusing on outcome and recovery after stroke are just at the beginning but expected to further improve options for clinical applications.[102,103]

Search strategy and selection criteria

We identified articles published in English by searches of PubMed between Jan 1, 2012 and December 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 generated based on relevance to the topics covered in this review.

Declaration of interest

MD has received grants from the European Union's Horizon 2020 research and innovation programme under grant agreement No 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) through 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 grants R01NS036695, UM1HG008895, R01NS093870, R24NS092983, and has consulted for New Beta Innovations, Boehringer Ingelheim, and Pfizer Inc; SP received grants from the Dutch Organization for Scientific Research (Nederlandse Organisatie voor Wetenschappelijk Onderzoek, NOW; the Veni Fellowship 016.186.071, ZonMW) and NIH grant R01NS100178.

Authors' contributions

MD and JR generated the outline of the Review. MD, SP, and JR drafted 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.

Figure legends

Figure 1. Risk loci for large artery stroke and their association with related vascular traits. Shown are genetic loci that either reached genome-wide significance for association with large artery stroke (bold) or reached genome-wide significance for any ischemic stroke and in addition showed a strong association signal ($p<5 \times 10^{-4}$) for large artery stroke in MEGASTROKE.[2] Several of the risk loci for large artery stroke are established risk loci for related vascular traits.[2,31] IMT, intima media thickness.

Figure 2. Exploration of potential therapeutic targets by Mendelian Randomisation. Schematic representation of the principles and requirements of instrumental variable analysis to generate causal estimates through MR: the genetic instrument (single variant or multiple variants) associates with the exposure; the genetic instrument must not associate with known or unknown confounders; any influence of the genetic instrument on the outcome phenotype is through the intermediate phenotype

Figure 3. Exploiting genetics for drug discovery. Shown are approaches that facilitate the discovery, development, and prioritization of drug targets.

Tables

Table 1. Mendelian causes of stroke

Table 2. Common variants associated with stroke. Shown are the top signals from previous GWAS meta-analyses in MEGASTROKE[2] and a combined meta-analysis of MEGASTROKE with data from UK biobank.[7] For each locus, the variant showing the lowest p-value in the fixed effects trans-ancestral or European-only meta-analysis is shown**. ***results obtained in the European-only meta-analysis. For consistency, associations in stroke subtypes are taken from the respective European-only analyses. OR, odds ratio; CI, confidence interval.

Table 3. Pathological and physiological processes related to genes and gene loci associated with small vessel stroke. Shown are genes implicated in Mendelian stroke (top) and gene loci associated with sporadic stroke (bottom) along with the presumed relationship to pathological (black symbols) and physiological (grey symbols) processes as supported by experimental studies or observations in humans (references on the right). For instance, *NOTCH3* has a physiological role in cell-cell signaling while disease-associated mutations cause abnormal folding, and accumulation of the NOTCH3 protein. Genes primarily implicated in intracerebral haemorrhage are underlined. **HTRA1*, *COL4A1/A2*, and *FOXF2* have been shown to affect Tgf-ß signaling.

Table 4. Notable Mendelian Randomisation studies in stroke. Note that absence of support for a causal relationship may also relate to methodological aspects such as limited statistical power. For explanation see text.

Panels

Panel 1. Glossary of terms and concepts relevant to stroke genetics

References

1. Collaborators GBDCoD. 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**(10159): 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. *Nature genetics* 2018; **50**(4): 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 neurology* 2016; **15**(7): 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. *Annals of neurology* 2016; **80**(5): 741-53.
5. Verdura E, Herve D, Scharrer E, et al. Heterozygous HTRA1 mutations are associated with autosomal dominant cerebral small vessel disease. *Brain : a journal of neurology* 2015; **138**(Pt 8): 2347-58.
6. Rutten-Jacobs L, 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 three novel loci associated with stroke. *Annals of neurology* 2018; **84**(6): 934-9.
8. Boycott KM, Vanstone MR, Bulman DE, MacKenzie AE. Rare-disease genetics in the era of next-generation sequencing: discovery to translation. *Nature reviews Genetics* 2013; **14**(10): 681-91.
9. Bycroft C, Freeman C, Petkova D, et al. The UK Biobank resource with deep phenotyping and genomic data. *Nature* 2018; **562**(7726): 203-9.
10. Roselli C, Chaffin MD, Weng LC, et al. Multi-ethnic genome-wide association study for atrial fibrillation. *Nature genetics* 2018.
11. Bugiani M, Kevelam SH, Bakels HS, et al. Cathepsin A-related arteriopathy with strokes and leukoencephalopathy (CARASAL). *Neurology* 2016; **87**(17): 1777-86.
12. Plenge RM, Scolnick EM, Altshuler D. Validating therapeutic targets through human genetics. *Nature reviews Drug discovery* 2013; **12**(8): 581-94.
13. Finan C, Gaulton A, Kruger FA, et al. The druggable genome and support for target identification and validation in drug development. *Science translational medicine* 2017; **9**(383).
14. 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.
15. 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**(18): 1555-65.
16. Chabriat H, Joutel A, Dichgans M, Tournier-Lasserve E, Bousser MG. Cadasil. *Lancet neurology* 2009; **8**(7): 643-53.
17. Stam AH, Kothari PH, Shaikh A, et al. Retinal vasculopathy with cerebral leukoencephalopathy and systemic manifestations. *Brain : a journal of neurology* 2016; **139**(11): 2909-22.
18. Switzer JA, Hess DC, Nichols FT, Adams RJ. Pathophysiology and treatment of stroke in sickle-cell disease: present and future. *Lancet neurology* 2006; **5**(6): 501-12.
19. French CR, Seshadri S, Destefano AL, et al. Mutation of FOXC1 and PITX2 induces cerebral small-vessel disease. *The Journal of clinical investigation* 2014; **124**(11): 4877-81.
20. Zhou Q, Yang D, Ombrello AK, et al. Early-onset stroke and vasculopathy associated with mutations in ADA2. *The New England journal of medicine* 2014; **370**(10): 911-20.
21. Roach ES, Islam MP. Pseudoxanthoma elasticum. *Handb Clin Neurol* 2015; **132**: 215-21.
22. Debette S, Goeggel Simonetti B, Schilling S, et al. Familial occurrence and heritable connective tissue disorders in cervical artery dissection. *Neurology* 2014; **83**(22): 2023-31.
23. Chung MG. Hereditary hemorrhagic telangiectasia. *Handb Clin Neurol* 2015; **132**: 185-97.
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. *Proceedings of the National Academy of Sciences of the United States of America* 2014; **111**(46): 16496-501.
25. Network NSG, International Stroke Genetics C. Loci associated with ischaemic stroke and its subtypes (SiGN): a genome-wide association study. *Lancet neurology* 2016; **15**(2): 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. *Nature genetics* 2012; **44**(3): 328-33.
27. Biffi A, Sonni A, Anderson CD, et al. Variants at APOE influence risk of deep and lobar intracerebral hemorrhage. *Annals of neurology* 2010; **68**(6): 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. *American journal of human genetics* 2014; **94**(4): 511-21.
29. Dichgans M, Malik R, König I, et al. Shared genetic susceptibility to ischemic stroke and coronary artery disease – a genome-wide analysis of common variants. *Stroke; a journal of cerebral circulation* 2014; **45**(1): 24-36.
30. Nikpay M, Goel A, Won HH, et al. A comprehensive 1,000 Genomes-based genome-wide association meta-analysis of coronary artery disease. *Nature genetics* 2015; **47**(10): 1121-30.
31. 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**(9): 5141.
32. Malik R, Dau T, Gonik M, et al. Common coding variant in SERPINA1 increases the risk for large artery stroke. *Proceedings of the National Academy of Sciences of the United States of America* 2017; **114**(14): 3613-8.
33. Falcone GJ, Woo D. Genetics of Spontaneous Intracerebral Hemorrhage. *Stroke; a journal of cerebral circulation* 2017; **48**(12): 3420-4.
34. Lee S, Abecasis GR, Boehnke M, Lin X. Rare-variant association analysis: study designs and statistical tests. *American journal of human genetics* 2014; **95**(1): 5-23.
35. 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**(7): 781-8.
36. Bersenev A, Wu C, Balcerek J, Tong W. Lnk controls mouse hematopoietic stem cell self-renewal and quiescence through direct interactions with JAK2. *The Journal of clinical investigation* 2008; **118**(8): 2832-44.
37. Wang W, Tang Y, Wang Y, et al. LNK/SH2B3 Loss of Function Promotes Atherosclerosis and Thrombosis. *Circulation research* 2016; **119**(6): e91-e103.
38. Duan L, Wei L, Tian Y, et al. Novel Susceptibility Loci for Moyamoya Disease Revealed by a Genome-Wide Association Study. *Stroke; a journal of cerebral circulation* 2018; **49**(1): 11-8.
39. 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; a journal of cerebral circulation* 2015; **46**(1): 197-202.
40. 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. *Arteriosclerosis, thrombosis, and vascular biology* 2014; **34**(9): 1871-9.
41. 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**(1): 1009.
42. Gschwendtner A, Bevan S, Cole JW, et al. Sequence variants on chromosome 9p21.3 confer risk for atherosclerotic stroke. *Annals of neurology* 2009; **65**(5): 531-9.
43. Motterle A, Pu X, Wood H, et al. Functional analyses of coronary artery disease associated variation on chromosome 9p21 in vascular smooth muscle cells. *Human molecular genetics* 2012; **21**(18): 4021-9.
44. 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**(21): 1969-78.
45. Kojima Y, Downing K, Kundu R, et al. Cyclin-dependent kinase inhibitor 2B regulates efferocytosis and atherosclerosis. *The Journal of clinical investigation* 2014; **124**(3): 1083-97.
46. 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.
47. 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 genetics* 2014; **10**(7): e1004469.
48. 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**(3): 522-33 e15.
49. 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. *Proceedings of the National Academy of Sciences of the United States of America* 1998; **95**(24): 14367-72.
50. Khera AV, Kathiresan S. Genetics of coronary artery disease: discovery, biology and clinical translation. *Nature reviews Genetics* 2017; **18**(6): 331-44.

51. Germain M, Chasman DI, de Haan H, et al. Meta-analysis of 65,734 individuals identifies TSPAN15 and SLC44A2 as two susceptibility loci for venous thromboembolism. *American journal of human genetics* 2015; **96**(4): 532-42.
52. 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**(1): 19-32.
53. 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**.
54. Zellner A, Scharrer E, Arzberger T, et al. CADASIL brain vessels show a HTRA1 loss-of-function profile. *Acta neuropathologica* 2018.
55. 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.
56. 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.
57. 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. *Annals of neurology* 2015; **78**(6): 887-900.
58. 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**(4): 408-20.
59. Schiffmann R. Fabry disease. *Pharmacol Ther* 2009; **122**(1): 65-77.
60. 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**(1): 60-70.
61. Traylor M, Malik R, Nalls MA, et al. Genetic variation at 16q24.2 is associated with small vessel stroke. *Annals of neurology* 2017; **81**(3): 383-94.
62. 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**(2): 132-43.
63. 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**(9): e107674.
64. Rannikmae K, Davies G, Thomson PA, et al. Common variation in COL4A1/COL4A2 is associated with sporadic cerebral small vessel disease. *Neurology* 2015.
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**(2): 146-53.
66. Weng YC, Sonni A, Labelle-Dumais C, et al. COL4A1 mutations in patients with sporadic late-onset intracerebral hemorrhage. *Annals of neurology* 2012; **71**(4): 470-7.
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 neurology* 2011; **10**(8): 702-9.
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; a journal of cerebral circulation* 2014; **45**(2): 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. *Nature genetics* 2018; **50**(9): 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): 1-22.
71. Holmes MV, Ala-Korpela M, Smith GD. Mendelian randomization in cardiometabolic disease: challenges in evaluating causality. *Nat Rev Cardiol* 2017; **14**(10): 577-90.
72. Beheshti S, Madsen CM, Varbo A, Benn M, Nordestgaard BG. Relationship of Familial Hypercholesterolemia and High LDL Cholesterol to Ischemic Stroke: The Copenhagen General Population Study. *Circulation* 2018.
73. 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; a journal of cerebral circulation* 2018; **49**(4): 820-7.
74. Hopewell JC, Malik R, Valdes-Marquez E, Worrall BB, Collins R, ISGC MCot. Differential effects of PCSK9 variants on risk of coronary disease and ischaemic stroke. *European heart journal* 2018; **39**(5): 354-9.
75. Emdin CA, Khera AV, Natarajan P, et al. Phenotypic Characterization of Genetically Lowered Human Lipoprotein(a) Levels. *Journal of the American College of Cardiology* 2016; **68**(25): 2761-72.
76. Larsson SC, Scott RA, Traylor M, et al. Type 2 diabetes, glucose, insulin, BMI, and ischemic stroke subtypes: Mendelian randomization study. *Neurology* 2017; **89**(5): 454-60.

77. 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; a journal of cerebral circulation* 2018; **49**(6): 1325-31.
78. 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**(24): 2373-88.
79. 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.
80. Casas JP, Bautista LE, Smeeth L, Sharma P, Hingorani AD. Homocysteine and stroke: evidence on a causal link from mendelian randomisation. *Lancet* 2005; **365**(9455): 224-32.
81. 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**(9791): 584-94.
82. 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**(6): e1001976.
83. Georgakis M, Gill D, Rannikmäe K, et al. Genetically Determined Levels of Circulating Cytokines and Risk of Stroke: Role of Monocyte Chemoattractant Protein-1. *Circulation* 2018: epub ahead of print.
84. Interleukin 1 Genetics C. Cardiometabolic effects of genetic upregulation of the interleukin 1 receptor antagonist: a Mendelian randomisation analysis. *Lancet Diabetes Endocrinol* 2015; **3**(4): 243-53.
85. Interleukin-6 Receptor Mendelian Randomisation Analysis C, 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**(9822): 1214-24.
86. 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**(10): e1001751.
87. Keenan T, Zhao W, Rasheed A, et al. Causal Assessment of Serum Urate Levels in Cardiometabolic Diseases Through a Mendelian Randomization Study. *Journal of the American College of Cardiology* 2016; **67**(4): 407-16.
88. van der Laan SW, Fall T, Soumare A, et al. Cystatin C and Cardiovascular Disease: A Mendelian Randomization Study. *Journal of the American College of Cardiology* 2016; **68**(9): 934-45.
89. Lee SJ, Jee YH, Jung KJ, Hong S, Shin ES, Jee SH. Bilirubin and Stroke Risk Using a Mendelian Randomization Design. *Stroke; a journal of cerebral circulation* 2017; **48**(5): 1154-60.
90. Voight BF, Peloso GM, Orho-Melander M, et al. Plasma HDL cholesterol and risk of myocardial infarction: a mendelian randomisation study. *Lancet* 2012; **380**(9841): 572-80.
91. 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. *Journal of the American College of Cardiology* 2012; **60**(25): 2631-9.
92. 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 2 x 2 factorial Mendelian randomization study. *Journal of the American College of Cardiology* 2015; **65**(15): 1552-61.
93. Hopewell JC, Clarke R. Emerging Risk Factors for Stroke: What Have We Learned From Mendelian Randomization Studies? *Stroke; a journal of cerebral circulation* 2016; **47**(6): 1673-8.
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**(10046): 761-75.
95. Davey Smith G, Paternoster L, Relton C. When Will Mendelian Randomization Become Relevant for Clinical Practice and Public Health? *JAMA : the journal of the American Medical Association* 2017; **317**(6): 589-91.
96. Cohen JC, Boerwinkle E, Mosley TH, Jr., Hobbs HH. Sequence variations in PCSK9, low LDL, and protection against coronary heart disease. *The New England journal of medicine* 2006; **354**(12): 1264-72.
97. Sabatine MS, Giugliano RP, Keech AC, et al. Evolocumab and Clinical Outcomes in Patients with Cardiovascular Disease. *The New England journal of medicine* 2017; **376**(18): 1713-22.
98. Gaudet D, Alexander VJ, Baker BF, et al. Antisense Inhibition of Apolipoprotein C-III in Patients with Hypertriglyceridemia. *The New England journal of medicine* 2015; **373**(5): 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**(10002): 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**(7649): 235-9.

101. Bush WS, Oetjens MT, Crawford DC. Unravelling the human genome-phenome relationship using phenome-wide association studies. *Nature reviews Genetics* 2016; **17**(3): 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**(3): 9.

103. Mola-Caminal M, Carrera C, Soriano-Tarraga C, et al. PATJ Low Frequency Variants Are Associated with Worse Ischemic Stroke Functional Outcome: A Genome-Wide Meta-Analysis. *Circulation research* 2018.

Panel. Glossary of terms and concepts relevant to stroke genetics

Key genetic terms and concepts

- **3' untranslated region** - the untranslated segment of mRNA between the stop codon and the 3′ end of the transcript
- **Allele** - alternate 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 may 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** - process by which a trait-associated region from a GWAS is analysed to identify the particular genetic variants that are likely to causally influence the examined trait
- **Functional genomics** - 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 (GWAS)** - study of a genome-wide set of genetic variants in different individuals to see if any variant is associated with a trait. GWASs typically focus on associations between SNPs and traits like major human diseases but can equally be applied to any other genetic variants
- **Intergenic region** – refers to a stretch of DNA sequences located between genes. Intergenic regions are a subset of noncoding DNA
- **Lead SNP** – refers to the SNP with the most significant p-value at a specific genetic locus
- **Minor allele frequency** - refers to the frequency at which the second most common allele occurs in a given population
- **miR** – microRNA is a small non-coding RNA molecule that functions in RNA silencing and post-transcriptional regulation of gene expression
- **Non-coding DNA** - components of DNA that do not encode protein sequences. Some noncoding DNA is transcribed into functional non-coding RNA molecules. Other functions of noncoding 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 in which the corresponding mRNA molecule is translated into a polypeptide chain
- **Small vessel stroke** – refers to an acute ischemic stroke likely to be caused by cerebral small vessel disease on the basis of diagnostic evaluation. Synonymous terms include "small artery stroke" "small artery ischemic stroke", and "lacunar stroke"

Key features of Mendelian Randomisation studies

- use genetic information to confirm a causal relationship between an exposure and an outcome
- 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 GWAS datasets become available
- have successfully been applied to vascular traits
- may identify potential drug targets
- may identify potential risks associated with pharmacological interventions
- may replace RCTs in settings in which RCTs are not feasible or ethical to conduct

Stroke genetics: discovery, biology, and clinical applications

Martin Dichgans, MD[1,2,3] Sara L Pulit, PhD[4,5,6] Jonathan Rosand, MD, MSc[6,7,8]

[1]Institute for Stroke and Dementia Research, Klinikum der Universität München, Ludwig-Maximilians-Universität LMU, Munich, Germany

[2]German Center for Neurodegenerative Diseases (DZNE, Munich), Munich, Germany

[3]Munich Cluster for Systems Neurology (SyNergy), Munich, Germany

[4]Department of Genetics, Center for Molecular Medicine, University Medical Center Utrecht, Utrecht, The Netherlands

[5]Big Data Institute, Li Ka Shing Center for Health Information and Discovery, Oxford University, Oxford, UK

[6]Program in Medical Population and Genetics, Broad Institute, Cambridge, MA, USA

[7]Henry and Allison McCance Center for Brain Health, Massachusetts General Hospital, Boston, MA, USA

[8]Center for Genomic Medicine, Massachusetts General Hospital, Boston, MA, USA

Correspondence to:

Martin Dichgans, Institute for Stroke and Dementia Research; Klinikum der Universität München; Feodor-Lynen-Straße 17, D-81377 München, Germany

Key words: stroke; hereditary; common, rare, risk prediction, precision medicine

Abstract

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 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. Further, genetics has successfully established causal relationships with risk factors and holds promise for prioritizing 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, ultimately, the conversion of genetic discoveries to novel therapies.

Introduction

Stroke remains a leading cause of death and long-term disability across the globe.[1] Despite the discovery of modifiable and non-modifiable risk factors as well as effective treatments novel therapeutic approaches are urgently needed. Uncovering the genetic contributions to stroke promises a better definition of causal pathways, the identification of novel therapeutic targets, and improved options for diagnosis and prognostication.[2-7]

The last five years have seen dramatic advances in genomic technologies, sequencing costs, biobanking, and data sharing, which collectively have accelerated genetic discovery.[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 (GWAS) in stroke and other vascular traits such as blood pressure and atrial fibrillation have tested over 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, have begun to elucidate causal pathways, relevant cell and tissue types, and, in some instances, yielded novel drug targets.[2]

Here, we review the latest discoveries in stroke genetics. In particular, we discuss the identification of novel Mendelian causes of stroke,[4,5,11] the discovery of at least 35 stroke risk loci harboring 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 further discuss how genetic discoveries could improve diagnosis, risk prediction, and the treatment of stroke. Indeed, recent observations have elucidated how genetics could be leveraged to discover novel drug targets and identify high-risk individuals many years before the emergence of classical indicators of stroke risk. [2,6,12,13] Because of their substantially different pathophysiologies, the following conditions were not considered: subarachnoid haemorrhage, cerebral aneurysms, cavernous malformations, dissections, and cerebral venous thrombosis.

Genetic discovery for stroke

Family-based studies and Mendelian stroke

Advances in sequencing technology have facilitated the discovery of single-gene disorders associated with stroke beyond classical syndromes such as CADASIL and sickle-cell disease (**Table 1**). Most notably, there has been a substantial expansion of the spectrum of ischemic

small vessel disease (SVD), which may manifest with ischemic stroke ('small vessel stroke', **Panel**), cognitive decline, and other manifestations. Among the most recent 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 SVD that typically manifests with early onset ischemic stroke.[4] These mutations disrupt a binding site for miR-29 and were shown to upregulate *COL4A1* mRNA expression. Heterozygous mutations (in particular, glycine substitutions) in the triple helical domains of COL4A1 or COL4A2 cause a different syndrome characterized by hemorrhagic stroke[14] along with additional neurological and non-neurological manifestations (**Table 1**).[4,5,11,14-23] Sequencing has further pinpointed heterozygous mutations in *HTRA1* (encoding high temperature requirement serine protease A1, HTRA1) in families with autosomal dominant SVD.[5] The condition typically manifests with stroke and cognitive decline in mid to late adulthood with more and more cases reported over the last years. In 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 than the syndrome of heterozygous mutation carriers and further features non-neurological symptoms.

Hereditary SVD syndromes typically presenting during childhood include deficiency of adenosine deaminase 2 (DADA2), an autoinflammatory disease manifesting with small vessel vasculitis caused by *CECR1* mutations.[20] Recently described mutations in *CTSA* (encoding cathepsin A)[11] and *FOXC1* (encoding forkhead box C1)[19] are associated with young onset autosomal dominant SVD.

Common sporadic stroke

Sporadic stroke arises through multiple risk factors and mechanisms. It is broadly classified as ischemic (large artery stroke, cardioembolic stroke, and small vessel stroke) and hemorrhagic (deep and lobar intracerebral haemorrhage [ICH]). 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 responsible stroke mechanism remains uncertain either because diagnostic work-up fails to find an established stroke etiology or because there are multiple competing etiologies. This complexity, along with the existence of different algorithms for stroke classification (discussed in [25]), has posed challenges to unravelling the genetic underpinnings of sporadic multifactorial stroke.

Common genetic variants associated with stroke

Common genetic variants, defined here as an allele frequency of ≥0.5% (one carrier per 100 individuals), have been associated with stroke using genome-wide association studies (GWAS), which compare the frequency of variants (mostly single nucleotide variants, SNPs) between groups of individuals. The first successful stroke GWAS included 3,548 patients and 5,972 controls and identified common variants at *HDAC9* (encoding histone deacetylase 9), that conferred an ~40% increased risk for large artery stroke per copy of the risk allele.[26] Since then, GWAS in progressively larger sample sizes have identified at least 35 loci with robust links to stroke risk.[2,27,28] Collectively, these studies enable the following conclusions: First, the vast majority of associated SNPs have a minor allele frequency of >5% and are associated with a modest increase in stroke risk (typically < 30% increase per allele) (**Table 2**). Second, most (~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 (AS), any ischemic stroke (AIS), large artery stroke, cardioembolic stroke, small vessel stroke, and ICH, both lobar and deep ICH (**Table 2,** and [27,28]). Fourth, at several loci, the association is restricted to a specific etiological stroke subtype. For instance, 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 ($p<5 \times 10^{-8}$) for large artery stroke but showed no association (p>0.05) with other stroke subtypes, implying mechanisms limited to atherosclerosis. Fifth, many genetic variants that confer risk of stroke also influence risk of related traits (**Figure 1**). Notably, about a quarter of the 32 genome-wide significant loci identified by MEGASTROKE are established risk loci for high blood pressure. Other loci (e.g., near *SMARCA4-LDLR*) are established risk loci for lipid levels.[2] Also, there is genetic overlap between large artery stroke and the related trait of coronary artery disease (CAD) (e.g. near *HDAC9-TWIST1*, *EDNRA*, and *9p21*)[2,29,30] and the intermediate phenotypes of carotid intima media thickness (IMT) and carotid plaque (e.g., at EDNRA)[2,31]. Sixth, several stroke risk loci contain genes previously implicated in monogenic SVD (e.g. *COL4A1*, *COL4A2*, and *HTRA1*)[2,7] (**Table 3**) suggesting a biological overlap between common stroke and rare familial stroke syndromes. And finally, while several risk loci integrate into previously suspected biological pathways for stroke, roughly one third show no obvious relationship with known pathways, pointing to mechanisms not previously implicated in stroke pathophysiology.

The heritability of stroke (i.e. the proportion of variation in risk attributable to inherited genetic variation) calculated from genome-wide data has been estimated to be 30%-40%[32,33] 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 to other common conditions including CAD. This is partly due to the complexity of stroke as a phenotype as well as the smaller sample sizes that are as yet available for etiologically-defined subtypes in recent GWAS. Not surprisingly, the lead variants at stroke risk loci identified to date explain roughly 1 to 2% of stroke heritability,[2] which is markedly lower than published figures for other traits.[30] Given the impact 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 this 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.[34] The two largest stroke studies[32,35] to date focused on coding regions (the exome) rather than the whole genome. While 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 (see below), 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, which often vary from locus to locus. These challenges notwithstanding, recent functional genomics studies have provided initial insights into the mechanisms linking common variants with stroke risk. In the following, we highlight key observations on biological mechanisms derived from recent genetic discoveries starting with all stroke and then moving to etiological 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 (P262W) in the lymphocyte adaptor protein LNK (the protein

encoded by *SH2B3)* and also associates with hypertension, type 1 diabetes, CAD, platelet counts, and leukocytosis, collectively suggesting an involvement of this locus in multiple mechanisms relevant to stroke pathophysiology. Genetic variation at *SH2B3* associates with both the broader phenotypes of AS, AIS, and with the etiologically defined stroke subtypes large artery stroke and small vessel stroke, providing an example of shared genetic influences between etiological subtypes.[2] *LNK* is predominantly expressed in hematopoietic and endothelial cells and negatively regulates cytokine signaling and cell proliferation.[36] *LNK* deficiency is associated with increased platelet production and activation, accelerated arterial thrombosis and atherosclerosis in hypercholesterolemic mice,[37] and other mechanisms under investigation. Collectively, these findings highlight the potential of genetics to uncover disease mechanisms while also illustrating challenges of genetic epidemiology as *SH2B3* has also been associated with traits whose link to stroke is more difficult to explain.

A substantial proportion of stroke risk variants associate with the mRNA expression, methylation, or protein levels of nearby genes primarily in stroke-relevant tissues (vascular, brain) and cell types (e.g., endothelial cells, blood, and immune cells) thus emphasizing 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 (e.g., 'enlarged heart' and 'cardiomyocyte differentiation via bone-morphogenetic-protein receptors') the coagulation system, and nitric oxide (NO) metabolism but also other pathways.[2]

Large artery stroke

The strongest association signal for large artery stroke identified to date is near *HDAC9*[2,26,32] a locus that also associates with CAD (**Figure 1**), peripheral artery disease, and moyamoya disease.[2,30,38] rs2109595, the lead SNP for stroke and likely causal variant[2] resides in regulatory DNA 3' to *HDAC9*. Risk variants associate with elevated *HDAC9* expression levels in blood cells with a gene dosage effect,[39] and deficiency of Hdac9 attenuates atherosclerosis in experimental mice.[39,40] Studies in Ldlr$^{-/-}$ mice suggest a pro-atherogenic effect of *Hdac9* in macrophages via changes in their activation status and cholesterol efflux[40] although *HDAC9* is also expressed in other cell types relevant to atherosclerosis.[41] 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 CAD but subsequently shown to also associate with peripheral

artery disease, intracranial and abdominal aortic aneurysms, carotid plaque,[31] and stroke, in particular large artery stroke.[2,42] Risk variants at 9p21 (i) are responsible for up to 20% of the population attributable risk of large artery stroke,[2,42] (ii) are considered to mediate their effects through mechanisms that are largely independent from established risk factors for atherosclerosis, and (iii) likely act through several genes at this locus. Specifically, these variants are associated with reduced expression of the cell cycle suppressor genes *CDKN2A* and *CDKN2B*[43,44] and with vascular smoot muscle cell (SMC) proliferation and SMC content in human atherosclerotic plaques.[43,44] *CDKN2B* has been implicated in SMC physiology and clearance of apoptotic debris, and mice deficient of *Cdkn2b* develop advanced atherosclerotic lesions composed of large, lipid-laden necrotic cores[45]. Altogether, the available evidence suggests a role of 9p21 in vascular remodeling, a key process in atherosclerosis. Importantly, CDKN2A/B and a circular non-coding antisense RNA (circANRIL) that is also transcribed at 9p21[46] have been suggested as possible targets for preventive strategies.

Recent GWAS further found variants near *TSPAN2*,[2,25] *LINC01492*,[2] *EDNRA*,[2] and *MMP12*[2,47] to be associated with large artery stroke. *EDNRA* and *MMP12* offer potential mechanistic insights: *ENDRA* (encoding endothelin receptor A, ET_A) is expressed in SMCs, endothelial cells, and macrophages. Variants at *EDNRA* also associate with carotid IMT (reflecting early atherosclerosis),[31] CAD[30] (**Figure 1**), and intracranial aneurysms, hence suggesting a broader role of this gene and Endothelin-1 signaling in vascular disease.[48] Activation of ET_A has effects on vasoconstriction, SMC proliferation, extracellular matrix (ECM) production, and fibrosis,[48] i.e., processes relevant to atherosclerosis. Moreover, pharmacological inhibition of ET_A normalized NO-mediated endothelium-dependent dysfunction[49] and attenuated atheroma formation in atherosclerotic mice[49] thus offering a perspective for alternative preventive strategies. *MMP12* (encoding matrix metalloproteinase 12, *MMP12*) was originally identified by an age-at-onset informed approach.[47] MMPs degrade ECM proteins, show increased activity in atherosclerotic plaques and have suspected roles in the growth, destabilization, and rupture of atherosclerotic lesions. Collectively, these findings suggest a causal role of the MMP12 protease in atherosclerosis (reviewed in [47]). Additional mechanism supported by recent stroke GWAS include changes in lipoprotein(a) metabolism (evidenced by *LPA*, a subthreshold locus for large artery stroke [p=1.3E-7][2] known to be implicated in atherosclerosis[50]) and thromboembolism (evidenced by *FGA*, *ILF3-SLC44A2*, and *ABO*, established risk loci for venous thromboembolism[51]). Hence, risk loci for large artery stroke highlight different aspects of the biology underlying large artery stroke (**Figure 1**).

These observations highlight different biological pathways underpinning large artery stroke. 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 exhibits close links with the genetics of atrial fibrillation (AF): (i) all loci reaching genome-wide significance for common variant association with cardioembolic stroke (nearest genes *PITX2*, *ZFHX2*, *NKX2-5*) are established risk loci for AF[10]; (ii) at all loci the association signal for stroke is confined to cardioembolic stroke; (iii) the association signals here and at subthreshold loci for cardioembolic stroke (near *CAV1/2* and *PRRX1*)[2] overlap with those for AF, and (iv) show a similar architecture (e.g. near *PITX2).* In addition, there are strong genetic links with venous thromboembolism: loci near *ABO*, *FGA* (encoding fibrinogen alpha), and *F11* (a subthreshold locus for cardioembolic stroke [$p=5.2 \times 10^{-8}$], encoding factor XI) are established risk loci for venous thromboembolism[51] and are known to modulate haemostatic traits.[51] Interestingly, genetic risk for the broader phenotype of AS or AIS further shows links with cardiac mechanisms beyond those implicated in AF as exemplified by the associations near *LRCH1* and *ANK2*, found in cardiac pacing and familial forms of cardiac disease (discussed in [2]).

Small vessel stroke

The identification of genes implicated in monogenic SVD has been instrumental to understanding the biology of small vessel stroke and SVD. Experimental studies in mutant mice and cultured cells highlight key physiological and pathological processes linked to genes implicated in monogenic SVD (**Table 3**).[11,14,15,24,52-63] Recurrent themes include TGF-ß signaling,[24,52] ECM formation/biology,[11,14,15,53-56] blood-brain barrier (BBB) function,[52,57,58] and specific cellular constituents of brain microvessels, in particular endothelial cells, pericytes, and SMCs (**Table 3**).[56,59,60] Proteomics studies of isolated brain microvessels from mutation carriers and mouse models further revealed an unexpected molecular link between CADASIL and CARASIL that involves accumulation of the HTRA1 protease and its substrates in microvessels from NOTCH3 mutation carriers.[54]

Common variant association studies identified various risk loci that show predominant association with small vessel stroke (**Table 2**). *ZCCHC14*, *FOXF2*, and *CASZ1* encode transcription factors expressed in the vasculature.[52,61,62] Foxf2 (expressed in brain endothelial cells and pericytes) is required for brain pericyte differentiation and development and maintenance of the BBB (**Table 3**).[52] Most notably, Foxf2 deficient mice develop microhemorrhages and neuronal loss thus recapitulating features of human SVD.[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]) also associate with radiological white matter hyperintensities (WMH), thus highlighting shared biology of small vessel stroke and WMH. *PMF1–SEMA4A* further reached genome-wide significance for association with deep (nonlobar) ICH,[28] consistent with nonlobar ICH being part of the phenotypic spectrum of SVD. Interestingly, 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 SVD-related phenotypes.[3,19]

Hemorrhagic stroke

ICH 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 sporadic ICH identified coding variants with *in vitro* pathological consequences resembling familial mutations in the same genes.[66] GWAS of ICH, on the other hand, have identified associations in the region of *COL4A2* in nonlobar ICH, as well as small vessel stroke and WMH, highlighting the shared pathways underlying the various manifestations of cerebral SVD. The shared biology between ICH and small vessel stroke is also underscored by associations at *PMF1/SLC25A44*.[2,28] An additional important observation has been that variation associated with risk of ICH can also affect the extent of bleeding, as measured by haematoma volume,[67] providing support for hypotheses generated from histopathologic data that ICH 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 (PRS) enables ascertainment of high-risk individuals at young age thus offering opportunities for early prevention.[68,69] For instance, a recent study found that applying a PRS consisting of 90 SNPs

to data from UK Biobank (>300,000 subjects) could identify individuals with a 35% increased risk of incident stroke (hazard ratio: 1.35). This cutoff included a third of the population.[6] Compared to individuals in the bottom tertile of the PRS who had a favourable lifestyle (defined as 3 or 4 healthy lifestyle factors) individuals in the upper tertile of the PRS had a relative risk (hazard ratio) of 1.44, 1.70 and 2.30 depending on whether they had a favourable, intermediate (2 healthy lifestyle factors) or unfavorable (0 or 1 healthy lifestyle factors) lifestyle, respectively. Indeed, lifestyle risk was similar across all PRS strata, thus highlighting the potential for early risk stratification and prevention via genetics.[6] 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 unfavorable consequences including costs and psychological distress.[69]

Exploration of potential therapeutic targets by Mendelian Randomisation

An increasingly recognized opportunity for genetics is the exploration of causal relationships between risk factors (exposures) and disease outcomes through Mendelian Randomisation (MR).[70,71] MR exploits genetic variants causally related to an exposure as instruments, and investigates their associations with a disease (**Panel** and **Figure 2**).[72-89] Because individual alleles are allocated at random during meiotic segregation, MR shares many features with randomized controlled trials (RCTs). Like RCTs, MR overcomes limitations of observational studies, in particular bias resulting from conventional confounding or reverse causality.

An important requirement for MR studies is a robust association between the genetic instrument (single SNP or multiple variants) and the exposure of interest. For instance, an intronic variant (rs6511720) in the low-density lipoprotein (LDL) receptor (*LDLR*) gene has consistently been shown to associate with LDL cholesterol (LDLC) levels. Like other variants known to influence LDLC, rs6511720 further associates with risk of myocardial infarction in a concordant fashion.[90] MR studies have demonstrated a dose-response relationship between genetically elevated LDLC levels and CAD[91] consistent with trials involving statins and other cholesterol lowering interventions. Conversely, MR studies examining HDL as an exposure found no causal effect of genetically determined HDL levels on the risk of CAD again in accordance with RCTs (reviewed in [71]).

MR studies may predict the success or failure of RCTs, thus saving on risks for study participants and on costs.[71] Examples, in which MR studies predicted the success or failure of RCTs include studies on inactivating mutations in *NPC1L1* (encoding the drug target for ezetimibe)[92] and variants in *PLA2G7* (encoding lipoprotein-associated phospholipase A2, the drug target for darapladib, reviewed in [71]), respectively.

Only recently, investigators have started to apply MR to stroke.[93] A MR study on blood lipids found genetically elevated LDLC levels associate with risk of large artery stroke but not small vessel stroke and cardioembolic stroke.[73] In contrast, genetically elevated HDL cholesterol associated with a reduced risk of small vessel stroke and no effects on large artery stroke and cardioembolic stroke. Differential effects on etiological stroke subtypes were further seen for T2D in that genetically defined T2D was found to associate with both small vessel stroke and large artery stroke but not cardioembolic stroke and ICH.[76,77] A MR study on the waist-to-hip ratio adjusted for body mass index (BMI) found waist-to-hip ratio to be causally related to a higher risk of ischemic stroke. This finding adds to observational data emphasizing the need to consider measures of adiposity beyond BMI for risk prediction and patient management.[94] MR may further identify novel risk factors and potential drug targets for stroke as illustrated by a recent study that found genetically elevated levels of the inflammatory cytokine MCP-1/CCL2 to be associated with risk of stroke.[83] **Table 4** provides and overview on notable MR studies in stroke.

While potentially highly informative, MR analyses also pose specific challenges that require consideration.[71] MR 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 2**). 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 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 MR results (**Table 4**). With increasing availability of large-scale genetic data and improved databases, MR studies will become even more informative and relevant for clinical practice.[50,95]

Exploiting genetics for drug discovery

Genetics holds great promise for catalyzing drug development and prioritizing targets for exploration in RCTs. Aside from identifying causal pathways and novel drug targets, genetics may help in anticipating the full range of efficacy and safety consequences of pharmacological interventions.[12,50] Key approaches aside from MR[71] include the exploitation of protective variants,[96-99] the examination of naturally occurring human knockouts,[100] and phenome-wide association studies (PheWAS)[75,101] (**Figure 3**)**.** PheWAS studies benefit from large datasets with detailed genotyping and extensive phenotyping of multiple traits. Such

datasets are currently generated through commercial (e.g. DECODE genetics), government (e.g. UK Biobank), and institutional (e.g. Kaiser Permanente research bank) funding. By leveraging data from >100.000 UK Biobank participants a recent PheWAS study provided a comprehensive account of the phenotypic consequences of genetically lowered Lp(a) levels on various disease states including a reduced risk of stroke.[75] 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 recent observation that risk loci for stroke are significantly enriched in drug-target genes for antithrombotic therapy. Specifically, *FGA* (**Table 2**) is a target for alteplase and other thrombolytic agents, and *PDE3A* is a target for cilostazol an antiplatelet agent approved for stroke prevention in Asia.[2] Utilizing genetics for drug discovery and genomic medicine represents a major area of future research.

Monogenic stroke

Advances in Mendelian stroke genetics have improved options for molecular diagnosis, prognostication, counseling, 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 (e.g., involving the skin, eye, and skeletal system) (**Table 1**). In most instances, 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 Syndrome) may suffice. In light 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 their 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 pathology holds the promise of identifying novel targets for intervention.[2,12,13] Genetic information can further be used to improve stroke diagnosis and prognostication.[5,6,14,68] Recent discoveries have expanded the number of genes and mutations proven to cause familial stroke, while GWAS have yielded at least 35 independent loci across the genome with an impact on stroke risk.[4,5,7,11,14,17,19,20,68] Functional exploration of these loci will be key to generating drug targets and novel therapies. The application of analytic techniques such as MR facilitates the exploration of causal relationships with

exposures and prioritize potential therapeutic targets for stroke.[71,83,92] Aggregation of multiple SNPs from across the genome into PRS enables the identification of individuals at high risk for stroke at young age.[6]

Several lines of research promise to 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.[2,7] Biobanks, such as those of the UK,[9] and studies such as China-Kadoorie, will vastly expand the opportunity for gene discovery in the near future and the broad sharing of data will undoubtedly accelerate the process (see www.cerebrovascularportal.org). Genetic discovery will benefit from further refinement of diagnostic categories (e.g. cardioembolic stroke of specific origin). However, this will require resources to ensure a minimum common level of investigation in large patient samples. A crucial gap that must be filled, however, is the relative absence of cases of non-European ancestry.[2] Along these same lines, ancestry-specific studies will be vital 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 as a result of the GWAS revolution.[50] 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.[13] Further progress will require the development of novel cell and tissue models (e.g., employing genome editing in human inducible pluripotent stem cells) and advances in functional genomics and multi-level 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 ultimately 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 focusing on outcome and recovery after stroke are just at the beginning but expected to further improve options for clinical applications.[102,103]

Search strategy and selection criteria

We identified articles published in English by searches of PubMed between Jan 1, 2012 and December 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 generated based on relevance to the topics covered in this review.

Declaration of interest

MD has received grants from the European Union's Horizon 2020 research and innovation programme under grant agreement No 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) through 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 grants R01NS036695, UM1HG008895, R01NS093870, R24NS092983, and has consulted for New Beta Innovations, Boehringer Ingelheim, and Pfizer Inc; SP received grants from the Dutch Organization for Scientific Research (Nederlandse Organisatie voor Wetenschappelijk Onderzoek, NOW; the Veni Fellowship 016.186.071, ZonMW) and NIH grant R01NS100178.

Authors' contributions

MD and JR generated the outline of the Review. MD, SP, and JR drafted 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.

Figure legends

Figure 1. Risk loci for large artery stroke and their association with related vascular traits. Shown are genetic loci that either reached genome-wide significance for association with large artery stroke (bold) or reached genome-wide significance for any ischemic stroke and in addition showed a strong association signal ($p<5 \times 10^{-4}$) for large artery stroke in MEGASTROKE.[2] Several of the risk loci for large artery stroke are established risk loci for related vascular traits.[2,31] IMT, intima media thickness.

Figure 2. Exploration of potential therapeutic targets by Mendelian Randomisation. Schematic representation of the principles and requirements of instrumental variable analysis to generate causal estimates through MR: the genetic instrument (single variant or multiple variants) associates with the exposure; the genetic instrument must not associate with known or unknown confounders; any influence of the genetic instrument on the outcome phenotype is through the intermediate phenotype

Figure 3. Exploiting genetics for drug discovery. Shown are approaches that facilitate the discovery, development, and prioritization of drug targets.

Tables

Table 1. Mendelian causes of stroke

Table 2. Common variants associated with stroke. Shown are the top signals from previous GWAS meta-analyses in MEGASTROKE[2] and a combined meta-analysis of MEGASTROKE with data from UK biobank.[7] For each locus, the variant showing the lowest p-value in the fixed effects trans-ancestral or European-only meta-analysis is shown**. ***results obtained in the European-only meta-analysis. For consistency, associations in stroke subtypes are taken from the respective European-only analyses. OR, odds ratio; CI, confidence interval.

Table 3. Pathological and physiological processes related to genes and gene loci associated with small vessel stroke. Shown are genes implicated in Mendelian stroke (top) and gene loci associated with sporadic stroke (bottom) along with the presumed relationship to pathological (black symbols) and physiological (grey symbols) processes as supported by experimental studies or observations in humans (references on the right). For instance, *NOTCH3* has a physiological role in cell-cell signaling while disease-associated mutations cause abnormal folding, and accumulation of the NOTCH3 protein. Genes primarily implicated in intracerebral haemorrhage are underlined. **HTRA1*, *COL4A1/A2*, and *FOXF2* have been shown to affect Tgf-ß signaling.

Table 4. Notable Mendelian Randomisation studies in stroke. Note that absence of support for a causal relationship may also relate to methodological aspects such as limited statistical power. For explanation see text.

Panels

Panel 1. Glossary of terms and concepts relevant to stroke genetics

References

1. Collaborators GBDCoD. 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**(10159): 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. *Nature genetics* 2018; **50**(4): 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 neurology* 2016; **15**(7): 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. *Annals of neurology* 2016; **80**(5): 741-53.
5. Verdura E, Herve D, Scharrer E, et al. Heterozygous HTRA1 mutations are associated with autosomal dominant cerebral small vessel disease. *Brain : a journal of neurology* 2015; **138**(Pt 8): 2347-58.
6. Rutten-Jacobs L, 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 three novel loci associated with stroke. *Annals of neurology* 2018; **84**(6): 934-9.
8. Boycott KM, Vanstone MR, Bulman DE, MacKenzie AE. Rare-disease genetics in the era of next-generation sequencing: discovery to translation. *Nature reviews Genetics* 2013; **14**(10): 681-91.
9. Bycroft C, Freeman C, Petkova D, et al. The UK Biobank resource with deep phenotyping and genomic data. *Nature* 2018; **562**(7726): 203-9.
10. Roselli C, Chaffin MD, Weng LC, et al. Multi-ethnic genome-wide association study for atrial fibrillation. *Nature genetics* 2018.
11. Bugiani M, Kevelam SH, Bakels HS, et al. Cathepsin A-related arteriopathy with strokes and leukoencephalopathy (CARASAL). *Neurology* 2016; **87**(17): 1777-86.
12. Plenge RM, Scolnick EM, Altshuler D. Validating therapeutic targets through human genetics. *Nature reviews Drug discovery* 2013; **12**(8): 581-94.
13. Finan C, Gaulton A, Kruger FA, et al. The druggable genome and support for target identification and validation in drug development. *Science translational medicine* 2017; **9**(383).
14. 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.
15. 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**(18): 1555-65.
16. Chabriat H, Joutel A, Dichgans M, Tournier-Lasserve E, Bousser MG. Cadasil. *Lancet neurology* 2009; **8**(7): 643-53.
17. Stam AH, Kothari PH, Shaikh A, et al. Retinal vasculopathy with cerebral leukoencephalopathy and systemic manifestations. *Brain : a journal of neurology* 2016; **139**(11): 2909-22.
18. Switzer JA, Hess DC, Nichols FT, Adams RJ. Pathophysiology and treatment of stroke in sickle-cell disease: present and future. *Lancet neurology* 2006; **5**(6): 501-12.
19. French CR, Seshadri S, Destefano AL, et al. Mutation of FOXC1 and PITX2 induces cerebral small-vessel disease. *The Journal of clinical investigation* 2014; **124**(11): 4877-81.
20. Zhou Q, Yang D, Ombrello AK, et al. Early-onset stroke and vasculopathy associated with mutations in ADA2. *The New England journal of medicine* 2014; **370**(10): 911-20.
21. Roach ES, Islam MP. Pseudoxanthoma elasticum. *Handb Clin Neurol* 2015; **132**: 215-21.
22. Debette S, Goeggel Simonetti B, Schilling S, et al. Familial occurrence and heritable connective tissue disorders in cervical artery dissection. *Neurology* 2014; **83**(22): 2023-31.
23. Chung MG. Hereditary hemorrhagic telangiectasia. *Handb Clin Neurol* 2015; **132**: 185-97.
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. *Proceedings of the National Academy of Sciences of the United States of America* 2014; **111**(46): 16496-501.
25. Network NSG, International Stroke Genetics C. Loci associated with ischaemic stroke and its subtypes (SiGN): a genome-wide association study. *Lancet neurology* 2016; **15**(2): 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. *Nature genetics* 2012; **44**(3): 328-33.
27. Biffi A, Sonni A, Anderson CD, et al. Variants at APOE influence risk of deep and lobar intracerebral hemorrhage. *Annals of neurology* 2010; **68**(6): 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. *American journal of human genetics* 2014; **94**(4): 511-21.
29. Dichgans M, Malik R, König I, et al. Shared genetic susceptibility to ischemic stroke and coronary artery disease – a genome-wide analysis of common variants. *Stroke; a journal of cerebral circulation* 2014; **45**(1): 24-36.
30. Nikpay M, Goel A, Won HH, et al. A comprehensive 1,000 Genomes-based genome-wide association meta-analysis of coronary artery disease. *Nature genetics* 2015; **47**(10): 1121-30.
31. 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**(9): 5141.
32. Malik R, Dau T, Gonik M, et al. Common coding variant in SERPINA1 increases the risk for large artery stroke. *Proceedings of the National Academy of Sciences of the United States of America* 2017; **114**(14): 3613-8.
33. Falcone GJ, Woo D. Genetics of Spontaneous Intracerebral Hemorrhage. *Stroke; a journal of cerebral circulation* 2017; **48**(12): 3420-4.
34. Lee S, Abecasis GR, Boehnke M, Lin X. Rare-variant association analysis: study designs and statistical tests. *American journal of human genetics* 2014; **95**(1): 5-23.
35. 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**(7): 781-8.
36. Bersenev A, Wu C, Balcerek J, Tong W. Lnk controls mouse hematopoietic stem cell self-renewal and quiescence through direct interactions with JAK2. *The Journal of clinical investigation* 2008; **118**(8): 2832-44.
37. Wang W, Tang Y, Wang Y, et al. LNK/SH2B3 Loss of Function Promotes Atherosclerosis and Thrombosis. *Circulation research* 2016; **119**(6): e91-e103.
38. Duan L, Wei L, Tian Y, et al. Novel Susceptibility Loci for Moyamoya Disease Revealed by a Genome-Wide Association Study. *Stroke; a journal of cerebral circulation* 2018; **49**(1): 11-8.
39. 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; a journal of cerebral circulation* 2015; **46**(1): 197-202.
40. 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. *Arteriosclerosis, thrombosis, and vascular biology* 2014; **34**(9): 1871-9.
41. 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**(1): 1009.
42. Gschwendtner A, Bevan S, Cole JW, et al. Sequence variants on chromosome 9p21.3 confer risk for atherosclerotic stroke. *Annals of neurology* 2009; **65**(5): 531-9.
43. Motterle A, Pu X, Wood H, et al. Functional analyses of coronary artery disease associated variation on chromosome 9p21 in vascular smooth muscle cells. *Human molecular genetics* 2012; **21**(18): 4021-9.
44. 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**(21): 1969-78.
45. Kojima Y, Downing K, Kundu R, et al. Cyclin-dependent kinase inhibitor 2B regulates efferocytosis and atherosclerosis. *The Journal of clinical investigation* 2014; **124**(3): 1083-97.
46. 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.
47. 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 genetics* 2014; **10**(7): e1004469.
48. 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**(3): 522-33 e15.
49. 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. *Proceedings of the National Academy of Sciences of the United States of America* 1998; **95**(24): 14367-72.
50. Khera AV, Kathiresan S. Genetics of coronary artery disease: discovery, biology and clinical translation. *Nature reviews Genetics* 2017; **18**(6): 331-44.

51. Germain M, Chasman DI, de Haan H, et al. Meta-analysis of 65,734 individuals identifies TSPAN15 and SLC44A2 as two susceptibility loci for venous thromboembolism. *American journal of human genetics* 2015; **96**(4): 532-42.
52. 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**(1): 19-32.
53. 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**.
54. Zellner A, Scharrer E, Arzberger T, et al. CADASIL brain vessels show a HTRA1 loss-of-function profile. *Acta neuropathologica* 2018.
55. 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.
56. 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.
57. 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. *Annals of neurology* 2015; **78**(6): 887-900.
58. 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**(4): 408-20.
59. Schiffmann R. Fabry disease. *Pharmacol Ther* 2009; **122**(1): 65-77.
60. 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**(1): 60-70.
61. Traylor M, Malik R, Nalls MA, et al. Genetic variation at 16q24.2 is associated with small vessel stroke. *Annals of neurology* 2017; **81**(3): 383-94.
62. 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**(2): 132-43.
63. 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**(9): e107674.
64. Rannikmae K, Davies G, Thomson PA, et al. Common variation in COL4A1/COL4A2 is associated with sporadic cerebral small vessel disease. *Neurology* 2015.
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**(2): 146-53.
66. Weng YC, Sonni A, Labelle-Dumais C, et al. COL4A1 mutations in patients with sporadic late-onset intracerebral hemorrhage. *Annals of neurology* 2012; **71**(4): 470-7.
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 neurology* 2011; **10**(8): 702-9.
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; a journal of cerebral circulation* 2014; **45**(2): 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. *Nature genetics* 2018; **50**(9): 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): 1-22.
71. Holmes MV, Ala-Korpela M, Smith GD. Mendelian randomization in cardiometabolic disease: challenges in evaluating causality. *Nat Rev Cardiol* 2017; **14**(10): 577-90.
72. Beheshti S, Madsen CM, Varbo A, Benn M, Nordestgaard BG. Relationship of Familial Hypercholesterolemia and High LDL Cholesterol to Ischemic Stroke: The Copenhagen General Population Study. *Circulation* 2018.
73. 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; a journal of cerebral circulation* 2018; **49**(4): 820-7.
74. Hopewell JC, Malik R, Valdes-Marquez E, Worrall BB, Collins R, ISGC MCot. Differential effects of PCSK9 variants on risk of coronary disease and ischaemic stroke. *European heart journal* 2018; **39**(5): 354-9.
75. Emdin CA, Khera AV, Natarajan P, et al. Phenotypic Characterization of Genetically Lowered Human Lipoprotein(a) Levels. *Journal of the American College of Cardiology* 2016; **68**(25): 2761-72.
76. Larsson SC, Scott RA, Traylor M, et al. Type 2 diabetes, glucose, insulin, BMI, and ischemic stroke subtypes: Mendelian randomization study. *Neurology* 2017; **89**(5): 454-60.

77. 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; a journal of cerebral circulation* 2018; **49**(6): 1325-31.
78. 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**(24): 2373-88.
79. 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.
80. Casas JP, Bautista LE, Smeeth L, Sharma P, Hingorani AD. Homocysteine and stroke: evidence on a causal link from mendelian randomisation. *Lancet* 2005; **365**(9455): 224-32.
81. 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**(9791): 584-94.
82. 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**(6): e1001976.
83. Georgakis M, Gill D, Rannikmäe K, et al. Genetically Determined Levels of Circulating Cytokines and Risk of Stroke: Role of Monocyte Chemoattractant Protein-1. *Circulation* 2018: epub ahead of print.
84. Interleukin 1 Genetics C. Cardiometabolic effects of genetic upregulation of the interleukin 1 receptor antagonist: a Mendelian randomisation analysis. *Lancet Diabetes Endocrinol* 2015; **3**(4): 243-53.
85. Interleukin-6 Receptor Mendelian Randomisation Analysis C, 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**(9822): 1214-24.
86. 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**(10): e1001751.
87. Keenan T, Zhao W, Rasheed A, et al. Causal Assessment of Serum Urate Levels in Cardiometabolic Diseases Through a Mendelian Randomization Study. *Journal of the American College of Cardiology* 2016; **67**(4): 407-16.
88. van der Laan SW, Fall T, Soumare A, et al. Cystatin C and Cardiovascular Disease: A Mendelian Randomization Study. *Journal of the American College of Cardiology* 2016; **68**(9): 934-45.
89. Lee SJ, Jee YH, Jung KJ, Hong S, Shin ES, Jee SH. Bilirubin and Stroke Risk Using a Mendelian Randomization Design. *Stroke; a journal of cerebral circulation* 2017; **48**(5): 1154-60.
90. Voight BF, Peloso GM, Orho-Melander M, et al. Plasma HDL cholesterol and risk of myocardial infarction: a mendelian randomisation study. *Lancet* 2012; **380**(9841): 572-80.
91. 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. *Journal of the American College of Cardiology* 2012; **60**(25): 2631-9.
92. 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 2 x 2 factorial Mendelian randomization study. *Journal of the American College of Cardiology* 2015; **65**(15): 1552-61.
93. Hopewell JC, Clarke R. Emerging Risk Factors for Stroke: What Have We Learned From Mendelian Randomization Studies? *Stroke; a journal of cerebral circulation* 2016; **47**(6): 1673-8.
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**(10046): 761-75.
95. Davey Smith G, Paternoster L, Relton C. When Will Mendelian Randomization Become Relevant for Clinical Practice and Public Health? *JAMA : the journal of the American Medical Association* 2017; **317**(6): 589-91.
96. Cohen JC, Boerwinkle E, Mosley TH, Jr., Hobbs HH. Sequence variations in PCSK9, low LDL, and protection against coronary heart disease. *The New England journal of medicine* 2006; **354**(12): 1264-72.
97. Sabatine MS, Giugliano RP, Keech AC, et al. Evolocumab and Clinical Outcomes in Patients with Cardiovascular Disease. *The New England journal of medicine* 2017; **376**(18): 1713-22.
98. Gaudet D, Alexander VJ, Baker BF, et al. Antisense Inhibition of Apolipoprotein C-III in Patients with Hypertriglyceridemia. *The New England journal of medicine* 2015; **373**(5): 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**(10002): 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**(7649): 235-9.

101. Bush WS, Oetjens MT, Crawford DC. Unravelling the human genome-phenome relationship using phenome-wide association studies. *Nature reviews Genetics* 2016; **17**(3): 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**(3): 9.

103. Mola-Caminal M, Carrera C, Soriano-Tarraga C, et al. PATJ Low Frequency Variants Are Associated with Worse Ischemic Stroke Functional Outcome: A Genome-Wide Meta-Analysis. *Circulation research* 2018.

Panel. Glossary of terms and concepts relevant to stroke genetics

Key genetic terms and concepts

- **3' untranslated region** - the untranslated segment of mRNA between the stop codon and the 3′ end of the transcript
- **Allele** - alternate 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 may 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** - process by which a trait-associated region from a GWAS is analysed to identify the particular genetic variants that are likely to causally influence the examined trait
- **Functional genomics** - 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 (GWAS)** - study of a genome-wide set of genetic variants in different individuals to see if any variant is associated with a trait. GWASs typically focus on associations between SNPs and traits like major human diseases but can equally be applied to any other genetic variants
- **Intergenic region** – refers to a stretch of DNA sequences located between genes. Intergenic regions are a subset of noncoding DNA
- **Lead SNP** – refers to the SNP with the most significant p-value at a specific genetic locus
- **Minor allele frequency** - refers to the frequency at which the second most common allele occurs in a given population
- **miR** – microRNA is a small non-coding RNA molecule that functions in RNA silencing and post-transcriptional regulation of gene expression
- **Non-coding DNA** - components of DNA that do not encode protein sequences. Some noncoding DNA is transcribed into functional non-coding RNA molecules. Other functions of noncoding 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 in which the corresponding mRNA molecule is translated into a polypeptide chain
- **Small vessel stroke** – refers to an acute ischemic stroke likely to be caused by cerebral small vessel disease on the basis of diagnostic evaluation. Synonymous terms include "small artery stroke" "small artery ischemic stroke", and "lacunar stroke"

Key features of Mendelian Randomisation studies

- use genetic information to confirm a causal relationship between an exposure and an outcome
- 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 GWAS datasets become available
- have successfully been applied to vascular traits
- may identify potential drug targets
- may identify potential risks associated with pharmacological interventions
- may replace RCTs in settings in which RCTs are not feasible or ethical to conduct

COMMENTS BY THE EDITORS:

AUTHORS: We thank the Editors for their comments and believe that the additional changes further improved our manuscript. Below, we reply to all comments point-by-point providing the exact page of all changes to the revised "clean" version of the manuscript.

Editorial points:

1. **Please provide a clean manuscript version without any line numbers.**

   AUTHORS: done

2. **Please remove all tables and figures from the manuscript.**

   AUTHORS: done. However, we left the Panel in the manuscript as we understood the panel should not be removed.

3. **Please submit tables as separate word documents.**

   AUTHORS: done. These files have been uploaded as separate files to the submission system.

4. **Author signature form: please remove the role of the finding statement from here (ie, leave this section empty or add some text along the lines that this paper hasn't received any funding).**

   AUTHORS: done.

5. **Declaration of interest: Is "SVDs@taget" a funding source? If yes, this needs to be added to the corresponding ICMJE form.**

   AUTHORS: "SVDs@target" is the project name not a funding source. The funding source is: "European Union's Horizon 2020 research and innovation programme" the grant has the agreement No 666881 and is for the project entitled "SVDs@target".

6. **Figure 1: you mentioned that "modified from Malik et al.2 with additional data from Franceschini et al. 31) in the legend. In your point-by-point response, "We believe there is no need to obtain a copy of permission letter since the figure is sufficiently distinct from published figures". We need a copy of permission letter even if it's a modified version of the original figure. If the figure hasn't been used as a baseline for this one, please remove the text from the legend. Also, "with additional data from Franceschini et al. 31" needs further clarification or can it be omitted?**
   AUTHORS: done.

   The respective text has been modified as follows (page 16):

   Previous text:
   "Shown are genetic loci that 1) reached genome-wide significance for association with large artery stroke (bold) or 2) reached genome-wide significance for any ischemic stroke and in addition showed a strong association signal (p<5 x 10-4)

for large artery stroke (modified from Malik et al.2 with additional data from Franceschini et al. 31) IMT, intima media thickness."

New version:
"Shown are genetic loci that either reached genome-wide significance for association with large artery stroke (bold) or reached genome-wide significance for any ischemic stroke and in addition showed a strong association signal (p<5 x 10-4) for large artery stroke in MEGASTROKE.[2] Several of the risk loci for large artery stroke are established risk loci for related vascular traits.[2,31] IMT, intima media thickness."

We further slightly edited figure 1.

7. **Reviewer #4 concerns regarding the term "small vessel stroke": Could you please add something along the lines to your manuscript based on your point-by-point response to guide non-experts "While we agree that the term "small vessel stroke" is not ideal terminology we decided to stick to this term because this was the term used in the most recent and largest genome-wide association studies (GWAS) referenced both in Table 2 and throughout the manuscript…"**

AUTHORS: We now added a reference to the Panel when first mentioning on "small vessel stroke" (page 4) and added the following explanation to the Panel (page 24):

"• Small vessel stroke – refers to an acute ischemic stroke likely to be caused by cerebral small vessel disease on the basis of diagnostic evaluation. Synonymous terms include "small artery stroke" "small artery ischemic stroke", and "lacunar stroke"."

This explanation accounts for the comment by the reviewer who pointed out that "stroke is a clinical syndrome and the various radiological and pathological features of cerebral small vessel disease do not necessarily produce stroke as a phenotype" This had already been acknowledged by our previous revision in which we expanded on the text when first mentioning on "small vessel stroke" (page 3, bottom).

8. **Reviewer #3 concern regarding "SH2B3 locus": please add a sentence to the end of the paragraph to illustrate challenges of genetic epidemiology as requested by the reviewer.**

AUTHORS: We have now expanded on the paragraph adopting terminology suggested by Reviewer #3 as follows (page 7):

"Collectively, these findings highlight the potential of genetics to uncover disease mechanisms while also illustrating challenges of genetic epidemiology as SH2B3 has also been associated with traits whose link to stroke is more difficult to explain."

To enable readers to better appreciate research findings and to encourage full and transparent reporting of outcomes, The Lancet family journals offer to publish a webaddress in accepted paper that links to the study's protocol on the author's institutional website (see Lancet 2009; 373: 992). This is particularly encouraged for randomised controlled trials, but is welcome for all types of research.

AUTHORS: Since this is a review rather than original research we understand that there is no point in publishing a webaddress that links to the author’s institutional website.

We provide one "clean" copy and one copy where our changes are highlighted.

Table 1. Mendelian Causes of Stroke

Condition	Mode of inheritance	Underlying Gene(s)	Stroke mechanism	Comment and key references (selection)
Mendelian conditions mostly manifesting with ischemic stroke				
CADASIL	AD	*NOTCH3*	SVD	Most common hereditary stroke syndrome [16]
CARASIL	AR	*HTRA1*	SVD	Heterozygous mutations in HTRA1 may cause late onset SVD [5]
CARASAL	AD	*CTSA*	SVD	May manifest with both ischemic and hemorrhagic stroke; hypertension [11]
Fabry disease	X-linked	*GLA*	SVD, LAD, CE	Multi-organ disease; enzyme-replacement therapy available
PADMAL	AD	*COL4A1*	SVD	Ischemic lacunar infarctions in the pons as a common presentation [4]
RVCL-S	AD	*TREX1*	SVD	Retinopathy and Rim-enhancing mass lesions on neuroimaging [17]
Sickle-cell disease	AR	*HBB*	Prothrombotic state, LAD	Most common cause of stroke in children; hemorrhagic strokes in adult patients [18]
FOXC1 deletion related angiopathy	AD	*FOXC1*	SVD	White matter hyperintensities on brain MRI [19]
DADA2	AR	*CECR1*	Small vessel vasculitis	Typically manifests in early childhood; fewer, skin changes, polyarteritis nodosa [20]
Pseudoxanthoma elasticum	AR	*ABCC6*	LAD, SVD	Skin and retinal changes; calcified elastic fibers [21]
Homocystinuria	AR	*CBS and others*	LAD, CE, SVD, arterial dissection	Thromboembolism, premature atherosclerosis, mental retardation, Marfan-like skeletal abnormalities
Marfan Syndrome	AD	*FBN1*	CE, arterial dissection	Clinical diagnosis based on skeletal abnormalities, aortic root aneurysm, ectopia lentis, and other features
Vascular Ehlers-Danlos Syndrome	AD	*Col3A1*	Arterial dissection	Stroke typically before age 40 years [22]
MELAS	Maternal	*Mitochondrial DNA*	Microvascular & neuronal factors	Strokelike episodes typically before age 40 years; seizures, encephalopathy
Hereditary hemorrhagic telangiectasia	AD	*ENG* or *ALK1* in ~85% of cases	Arteriovenous malformations	Pulmonary AVMs as a cause of ischemic stroke; Cerebral AVMs as cause of intracerebral hemorrhage [23]
Mendelian conditions mostly manifesting with hemorrhagic stroke				
COL4A1/2-related angiopathies	AD and *de novo*	*COL4A1, COL4A2*	SVD	About 50% are sporadic cases; Hemorrhages manifest perinatal, in childhood, or in adulthood [14]
Cerebral Amyloid Angiopathy	AD	*APP, CST3*	Cerebral Amyloid Angiopathy	Manifests with stroke (mostly hemorrhagic) and dementia
Cerebral cavernous malformations	AD	KRIT1, MGC4607, PDCD10	Cerebral cavernous malformations	Multiple cavernomas; proportion of familial cases is up to 50% in Hispanic-American patients

Abbreviations: 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; SVD = small vessel disease; LAD = large artery disease; CE = cardioembolism.

Table 2. Common variants associated with stroke: Shown are the top signals from previous GWAS meta-analyses in MEGASTROKE[2] and a combined meta-analysis of MEGASTROKE with data from UK biobank.[7] For each locus the variant showing the lowest p-value in the fixed effects trans-ancestral or European-only meta-analysis is shown. *results obtained in the European-only meta-analysis. For consistency, associations in stroke subtypes are taken from the respective European-only analyses. OR, odds ratio; CI, confidence interval.

Top signal					Associations in stroke subtypes p-value*					Reference
rsID	Gene(s)	Risk allele (frequency in %)	Stroke phenotype	OR [95% CI]	Any	Any ischemic	Large artery	Cardioembolic	Small vessel	
rs880315	*CASZ1*	C (40)	any	1.05 [1.04-1.07]						2
rs12037987	*WNT2B*	C (16)	any	1.07 [1.05-1.10]						2
rs12124533	*TSPAN2*	T (24)	large artery	1.17 [1.11-1.23]						2
rs1052053	*PMF1-SEMA4A*	G (40)	any	1.06 [1.05-1.08]						2
rs146390073	*RGS7*	T (2)	cardioembolic	1.95 [1.54-2.47]*						2
rs12476527	*KCNK3*	G (48)	any	1.05 [1.03-1.07]						2
rs7610618	*TM4SF4-TM4SF1*	T (1)	large artery	2.33 [1.74-3.12]*						2
rs13143308	*PITX2*	T (28)	cardioembolic	1.32 [1.27-1.37]						2
rs34311906	*ANK2*	C (41)	any ischemic	1.07 [1.04-1.09]*						2
rs17612742	*EDNRA*	C (21)	large artery	1.19 [1.13-1.26]						2
rs6825454	*FGA*	C (31)	any ischemic	1.06 [1.04-1.08]						2
rs11957829	*LOC100505841*	A (82)	any ischemic	1.07 [1.05-1.10]						2
rs6891174	*NKX2-5*	A (35)	cardioembolic	1.11 [1.07-1.16]						2
rs4959130	*FOXF2*	A (14)	any	1.08 [1.05-1.11]						2
rs16896398	*SLC22A7-ZNF318*	T (34	any	1.05 [1.03-1.07]						2
rs2107595	*HDAC9-TWIST1*	A (24)	large artery	1.21 [1.15-1.26]						2
rs42039	*CDK6*	C (77)	any ischemic	1.07 [1.04-1.09]						2
rs7859727	*Chr9p21*	T (53	any	1.05 [1.03-1.07]						2
rs10820405	*LINC01492*	G (82)	large artery	1.20 [1.12-1.28]*						2
rs1799983	*NOS3*	T (32)	any	1.05 [1.03-1.07]*						7
rs635634	*ABO*	T (19)	any ischemic	1.08 [1.05-1.11]*						2
rs2295786	*SH3PXD2A*	A (60)	any	1.05 [1.04-1.07]						2
rs2005108	*MMP12*	T (12)	any ischemic	1.08 [1.05-1.11]						2
rs7304841	*PDE3A*	A (59)	any ischemic	1.05 [1.03-1.07]						2
rs3184504	*SH2B3*	T (45)	any ischemic	1.08 [1.06-1.10]						2
rs35436	*TBX3*	C (62)	any	1.05 [1.03-1.06]						2
rs9526212	*LRCH1*	G (76)	any	1.06 [1.04-1.08]						2
rs9521634	*COL4A1*	C (36)	any	1.04 [1.03-1.06]						7
rs4932370	*FURIN-FES*	A (33)	any ischemic	1.05 [1.03-1.07]						2
rs12932445	*ZFHX3*	C (21)	cardioembolic	1.20 [1.15-1.25]						2
rs12445022	*ZCCHC14*	A (31)	any	1.06 [1.04-1.08]						2
rs11867415	*PRPF8*	G (18)	any ischemic	1.09 [1.06-1.13]						2
rs2229383	*ILF3-SLC44A2*	T (65)	any ischemic	1.05 [1.03-1.07]						2
rs8103309	*SMARCA4-LDLR*	T (65)	any	1.05 [1.03-1.07]						2
rs720470	*DYRK1A*	T (71)	any	1.05 [1.03-1.07]						7

p-value: <5E-8 <5E-5 <5E-2 >5E-2

Table

		Pathological / Physiological Process								
		Protein folding/transport	Cell-signaling*	ECM formation/biology	SMC function/morphology	PC function/morphology	EC function/morphology	Angiogenesis	BBB function	Selected references
Rare variants (genes implicated in Mendelian stroke)	*NOTCH3*	●	○	●	◐	◐	○	○	○	53; 54; 57
	HTRA1	○	◐	●	●	○	○	○	○	24: 54
	COL4A1	●	◐	◐	●	●	●	○	●	14;15; 55; 58
	COL4A2	●	◐	◐	○	○	○	○	○	14;15; 55
	CTSA	○	○	●	●	○	○	○	○	11
	GLA	○	○	○	◐	●	◐	○	○	59
	APP	●	*affects multiple processes*							n.a.
	CST3	●	○	◐	●	○	○	○	○	56; 60
Common variants (gene loci associated with sporadic stroke)	*ZCCHC14*	○	○	○	○	○	○	○	○	61
	FOXF2	○	○	○	○	○	○	○	○	52
	PMF1	○	○	○	○	○	○	○	○	-
	HTRA1	○	○	○	○	○	○	○	○	24
	COL4A1	○	○	○	○	○	○	○	○	14; 54
	COL4A2	○	○	○	○	○	○	○	○	14; 15; 55
	SH3PXD2A	○	○	○	○	○	○	○	○	63
	CASZ1	○	○	○	○	○	○	○	○	62
	LOC100505841	○	○	○	○	○	○	○	○	-
	ApoEe2	affects multiple processes								n.a.
	ApoEe4	affects multiple processes								n.a.

● documented role of gene in pathological process
◐ documented role of gene in both processes
○ documented role of gene in physiological process
○ not examined, not reported, or controversial

Table

Exposure (risk factor)	Any stroke	Any ischemic	Large artery	Cardioembolic	Small vessel	Hemorrhagic	References
LDL Cholesterol							72-74
HDL Cholesterol							73
Triglycerides							73
Lp(a)							75
T2D							76; 77
BMI							76; 78
Waist-to-hip ratio adjusted for BMI							78
Alcohol consumption							79
Homocysteine							80; 81
Serum CRP levels							82
Serum MCP-1 levels							83
Serum IL1-Receptor antagonist							84
IL 6 receptor signaling							85
Vitamin D binding protein levels							86
Serum Urate levels							87
Serum Cystatin C levels							88

Causal relationship:

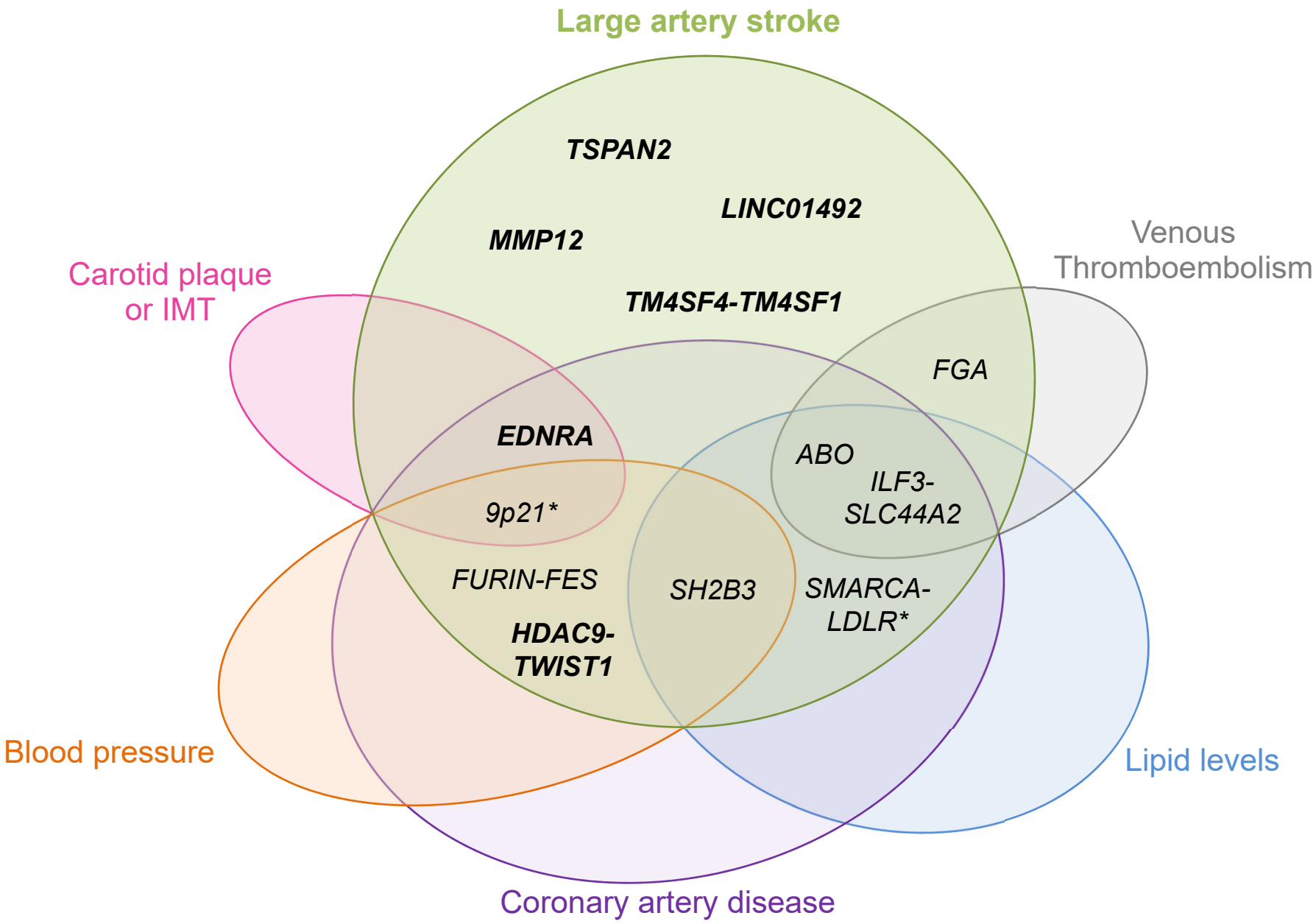
supported by MR

not supported by MR

controversial

Figure

Click here to download Figure: FIGURE 1R FINAL.pdf



Figure

Click here to download Figure: FIGURE 2R FINAL.pdf

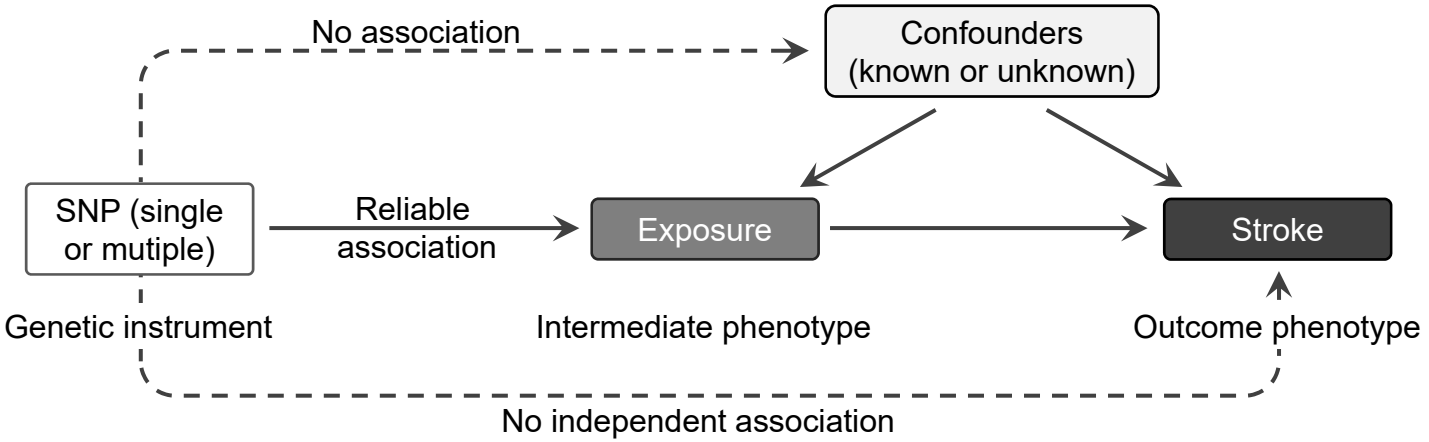
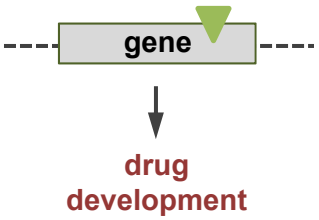
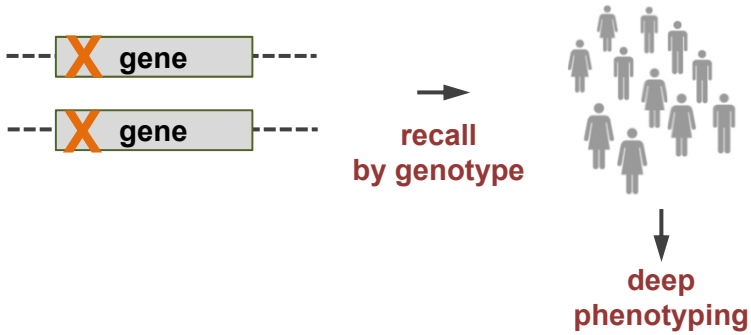
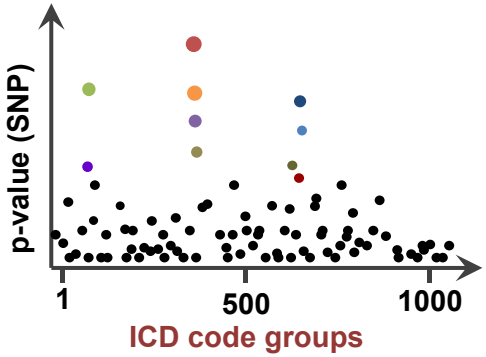


Figure
Click here to download Figure: FIGURE 3R.pdf

	Characteristics	Examples
(**A**) Protective variants 	• Capitalize on rare variants with strong effects • Serve as a starting point for drug development	• PCSK9 antibodies show efficacy in lowering LDLC and cardiovascular events in clinical endpoint trials[96;97] • Antisense oligonucleotides for *APOC3* and LPA show strong reductions in triglyceride[98] and Lp(a) levels,[99] respectively
(**B**) Human knockouts & recall by genotype 	• Homozygous inactivating mutations are identified by large-scale sequencing • Individuals are then called back for detailed phenotyping • Can provide some assurance that pharmacological inhibition of gene product will be tolerated	• Inactivation of *APOC3* results in marked blunting of post-prandial rise in triglyceride-rich lipoproteins[100]
(**C**) Phenome-wide association studies 	• Anticipate the full range of consequences that might be expected by pharmacological modulation • Require deep phenotyping	• Genetic instruments for lower Lp(a) levels associate with decreased risk of multiple diseases including stroke[75,101]