Mechanisms and treatment of ischaemic stroke—insights from genetic associations

Hugh S. Markus and Steve Bevan

Abstract | The precise pathophysiology of ischaemic stroke is unclear, and a greater understanding of the different mechanisms that underlie large-artery, cardioembolic and lacunar ischaemic stroke subtypes would enable the development of more-effective, subtype-specific therapies. Genome-wide association studies (GWASs) are identifying novel genetic variants that associate with the risk of stroke. These associations provide insight into the pathophysiological mechanisms, and present opportunities for novel therapeutic approaches. In this Review, we summarize the genetic variants that have been linked to ischaemic stroke in GWASs to date and discuss the implications of these associations for both our understanding and treatment of ischaemic stroke. The majority of genetic variants identified are associated with specific subtypes of ischaemic stroke, implying that these subtypes have distinct genetic architectures and pathophysiological mechanisms. The findings from the GWASs highlight the need to consider whether therapies should be subtype-specific. Further GWASs that include large cohorts are likely to provide further insights, and emerging technologies will complement and build on the GWAS findings.

Introduction

Though sometimes thought of as a uniform disease, stroke is the end result of many different pathophysiological processes, which result in multiple subtypes (Box 1). Identification of individual stroke subtypes requires a detailed clinical work-up, but even with appropriate investigations, an underlying pathology cannot be found in 25–40% of patients with ischaemic stroke.[1] This heterogeneity implies that different pathological mechanisms are responsible for different subtypes of stroke; identification of genetic variations that are associated with these subtypes might provide insight into these mechanisms and provide opportunities to develop new, more-effective therapeutic approaches.

Studies of twins and analyses of family history both suggest a substantial genetic predisposition to stroke, with implications for how molecular genetic studies of stroke are conducted and interpreted.[2–4] Family history data suggest that genetic predisposition differs by subtype of stroke, with greater familial associations for large-artery and small-artery (lacunar) stroke than for cardioembolic stroke.[3] These data also indicate that stroke and coronary artery disease share genetic factors, as a family history of coronary artery disease is associated with an increased risk of large-artery stroke,[5] and a family history of stroke is associated with acute coronary syndrome.[6] Furthermore, a family history of stroke is also strongly associated with risk factors for stroke, such as hypertension and diabetes mellitus, which makes it important to determine whether any identified genetic factors predispose patients to stroke indirectly via risk factors, or independently of the risk factors.[7] Sex differences have also been reported: women were found to be approximately 50% more likely to have a maternal history of stroke than a paternal history of stroke, whereas no similar excess exists in men.[8]

To date, the candidate gene approach to complex diseases, such as coronary artery disease and diabetes mellitus, has proven disappointing, and this has also been the case for stroke. Associations with stroke have been reported for a number of candidate genes, but few of these findings have been replicated.[9] Possible reasons for this lack of replication include small sample size, publication bias, and a failure to recognize the phenotypic heterogeneity of stroke.[10]

By contrast, the use of genome-wide association study (GWAS) technology, in combination with very large sample sizes and rigorous replication, has had a major impact on our current understanding of the pathogenesis of complex diseases, with well over 1,000 novel associations reported.[11] Unlike the candidate gene approach, GWASs allow detection of completely novel and unexpected associations owing to the fact that as many as 10 million genetic markers can be directly genotyped, and imputation allows estimates of over 15 million markers to provide coverage of the whole genome (Box 2).

GWASs in stroke initially lagged behind those in other complex diseases, but larger GWASs have begun to identify novel genetic associations with stroke. To date, the findings of these studies have had little immediate clinical impact; nevertheless, they are providing insights into the pathophysiology of stroke and could open up novel,

Stroke Research Group, Department of Neurology, R3, Box 83, University of Cambridge, Cambridge Biomedical Campus, Cambridge CB2 0QQ, UK (**H.S.M.**, **S.B.**).

Correspondence to: H.S.M. hsm32@ medschl.cam.ac.uk

Competing interests
The authors declare no competing interests.

Box 1 | Subtypes of ischaemic stroke

Large-artery stroke
Large-artery stroke results from atherosclerotic plaques in the carotid, verterbral or major intracerebral arteries. Plaques become unstable, leading to thrombus formation. The thrombus embolizes distally to occlude intracerebral vessels.

Cardioembolic stroke
Cardioembolic stroke results from a variety of intracardiac pathologies, including atrial fibrillation, cardiomyopathy, and the site of a previous myocardial infarction. The thrombus forms in the heart and embolizes to the brain.

Small-vessel disease (lacunar stroke)
Small-vessel stroke, also known as lacunar stroke, affects the small perforating arteries supplying subcortical structures and results in small lacunar infarcts that affect the white matter and deep grey matter nuclei. The main risk factor is hypertension.

previously unexpected treatment strategies. The GWAS data might also have important generic implications for the way that treatments for stroke are designed.

Here, we review the progression of GWAS use for identifying genetic associations with stroke, and summarize the specific genetic loci that have been associated with different subtypes of stroke to date. We explore the implications of these associations for the clinician in terms of predictive testing, therapeutic potential, and the impact on clinical management, as well as the insights into stroke pathophysiology, and consider aspects that need to be addressed in future studies to maximize the potential of GWASs. This Review focuses on findings in ischaemic stroke, although advances are also being made in intracerebral haemorrhage,[12] including the genetics of intracranial aneurysms.[13]

GWASs in stroke

The use of GWASs has already identified a number of novel variants associated with risk of stroke (Table 1), and has provided important insights into the genetic basis of this disease.

Early studies

Initial studies utilizing GWASs in stroke did not aim to identify completely novel stroke-related genes. Instead, they focused on genetic variants identified in a range of other cardiovascular diseases, such as atrial fibrillation, that increase an individual's risk of stroke. The goal of these studies was to determine whether such variants were also directly associated with ischaemic stroke.

Variants in two genes associated with atrial fibrillation, *PITX2* and *ZFHX3*, were found to be independent risk factors for ischaemic stroke;[14,15] further analysis determined that these associations were specific to the development of cardioembolic stroke.[16] *PITX2* encodes a transcription factor regulated by β-catenin, whereas *ZFHX3* encodes a zinc finger protein. Furthermore, a genetic variant at the 9p21 locus on the short arm of chromosome 9 that has a widely reported association with myocardial infarction and coronary artery disease[17] was associated with ischaemic stroke in a meta-analysis across multiple cohorts.[18] This association was specific to the large-artery subtype of stroke (OR 1.21), and was not observed among patients with cardioembolic or small-vessel stroke. This region of chromosome 9 contains genes that encode cyclin-dependent kinase inhibitors (*CDKN2A* and *CDKN2B*) and a noncoding RNA of unknown function (*CDKN2B-AS1*, also known as *ANRIL*).

Novel associations

The first widely replicated novel genetic association with ischaemic stroke was found in a GWAS conducted by the Wellcome Trust Case Control Consortium 2 (WTCCC2).[19] This study enrolled a discovery population of approximately 3,500 patients with stroke from four centres in Europe, and identified a genetic variant at the chromosome 7p21 locus that was associated specifically with the large-artery subtype of stroke (OR 1.42).[19] This association was replicated in the same study in a further 6,000 patients from Europe, America and Australia,[19] and in subsequent studies.[20,21]

The 7p21 locus resides between two recombination hotspots that encompass the 3' end of the histone deacetylase 9 (*HDAC9*) gene, making it likely that this gene underlies the observed association with large-artery stroke.[19] The variants identified in the WTCCC2 GWAS, and a nearby linked variant that was identified in a subsequent meta-analysis (METASTROKE),[20] are located within the introns of *HDAC9* rather than in the coding regions. However, the primary *HDAC9* single nucleotide polymorphism (SNP) that was identified in the METASTROKE study may have an effect on gene regulation; it lies within a DNAse hypersensitivity cluster and its sequence is conserved in mammals—observations that are consistent with this SNP being a functional variant. Whether or not this particular variant underlies the association of *HDAC9* with large-artery stroke, or is merely a marker for another nearby causal variant, remains to be determined.[22]

Stroke subtype-specific associations

Further GWASs have reported specific associations with different subtypes of stroke. For example, a genetic variant located at chromosome 6p21.1 was associated

Box 2 | Genome-wide association studies

Genome-wide association studies search the genome for small variations called single nucleotide polymorphisms (SNPs) that occur more frequently in people with a particular disease than in people without the disease. The method uses fluorescent labelling of DNA in microarrays to type as many as 5 million or more SNPs. The frequency of each SNP in individuals with the disease is compared with the frequency in healthy controls. Due to the very large number of statistical comparisons involved, a high level of statistical significance is used, and the P value threshold for significance is $<5\times10^{-8}$. Detection of a significant association does not mean that the identified SNP is causal; it is usually merely a marker for a causal functional SNP that is located on a nearby region of the genome.

with large-artery atherosclerotic stroke, but not other subtypes of stroke, among white populations.[23] An age-of-onset analysis that was weighted towards younger age at stroke onset found a specific association of large-artery atherosclerotic stroke with a variant in the *MMP12* gene, which encodes macrophage metalloelastase, also known as matrix metalloproteinase 12.[24] This observation was supported by a study that showed increased levels of *MMP12* mRNA in the atherosclerotic carotid plaque compared with normal arterial wall tissue.[24]

In contrast to the success of identifying genetic associations with large-artery and cardioembolic stroke, progress with small-artery stroke has been slower. A GWAS conducted in a group of Japanese patients with ischaemic stroke identified a genetic variant in *PRKCH*.[23] This gene encodes a serine–threonine kinase that regulates a variety of cellular functions including proliferation, differentiation and apoptosis. The variant identified was specifically associated with small-vessel stroke, a result that was replicated in both an independent Japanese cohort and a Chinese population.[25] The same variant was also found to be associated with MRI-determined silent brain infarction, which has a predominantly small-vessel

aetiology.[26] The identified SNP is monomorphic in white populations, so direct replication is not possible in these populations, but a large meta-analysis of GWAS data from white patients found no association between ischaemic stroke and a chromosomal region surrounding the *PRKCH* variant.[20]

Associations with all ischaemic stroke

Genetic variants that are associated with all types of ischaemic stroke have been reported, although in considerably smaller numbers than variants associated with specific subtypes of stroke. A meta-analysis of prospective cohort studies reported that an SNP in the chromosome 12p13 region was associated with all types of ischaemic stroke; *NINJ2* was suggested to be the underlying gene.[27] However, this association could not be replicated in large case–control meta-analyses.[20,28] The first association of a genetic variant with all types of ischaemic stroke to be replicated convincingly was reported with a locus at 12q24 (OR 1.1).[29] The same gene region has been associated with cardiovascular risk factors, including hypertension and type 1 diabetes mellitus,[29] although the SNP associated with these diseases was not the same as that found to be associated with stroke. The mechanism of action, and whether or not the association occurs independently of or via conventional risk factors, is not clear.

Implications for the clinician
Predictive genetic testing

Genetic variants identified by GWASs have been suggested as potential predictors of cardiovascular risk in individual patients. Consequently, tests have been developed that use saliva swabs to detect SNPs associated with a risk of cardiovascular disease. However, the current scientific basis for such approaches in stroke is weak. For example, calculations show that 100–300 different genetic variants would be required to explain the reported sibling relative risk of stroke (2–3),[30,31] so the few genetic variants that have been associated with stroke to date can only account for a fraction of the overall risk. Until additional associated genetic variants have been identified and we better understand the molecular basis of stroke, this kind of predictive testing can provide only limited information. This view was reflected in the decision by the FDA in 2013 to ban direct marketing of saliva-based DNA tests by the company 23andMe.[32]

There are other challenges in implementing personalized testing for patients with complex diseases such as stroke. Patient compliance with strategies to reduce established stroke risk factors, such as hypertension and smoking, is often poor, and might not improve even in those who have a high genetic risk. Furthermore, patients with a low genetic risk might take less care to minimize general risk factors, thereby increasing their risk of stroke. A recent review found conflicting results on whether smoking cessation—a major behavioural risk factor for stroke—was increased after genetic profiling for the glutathione *S*-transferase M1 (*GSTM1*) gene status, which is associated with risk of lung cancer.[33] The

Table 1 | Genetic variants associated with ischaemic stroke, identified by GWASs

Stroke subtype	Chromosomal locus	Candidate gene	Replicated in other studies
Large-artery stroke[18,19,23,24]	9p21	*CDKN2B-AS1* (*ANRIL*) or *CDKN2A* and/or *CDKN2B*	Yes[20]
	7p21	*HDAC9*	Yes[20,21]
	6p21.1	Unknown	Not yet attempted
	11q22	*MMP12*	Not yet attempted
Cardioembolic stroke[14,15]	4q25	*PITX2*	Yes[16,20]
	16q22	*ZFHX3*	Yes[20]
Small-vessel disease[25]	14q22	*PRKCH*	Yes (in Chinese and Japanese cohorts[26]); monomorphic in white populations, and no association between this gene region and stroke was found in the METASTROKE study[20]
All ischaemic stroke[27,29]	12p13	*NINJ2*	Not replicated in large case–control analyses[20,28]
	12q24	Unknown	Not yet attempted

Abbreviation: GWASs, genome-wide association studies.

review concluded that more data are needed to determine the effects of genetic information on the success of behavioural risk factor modification.

For these reasons, individual predictive genetic testing is not a current reality. In the immediate future, stroke variants identified using GWASs are more likely to provide novel insights into specific aspects of stroke pathogenesis, allowing new treatment pathways to be developed, and more-general insights into the pathophysiology of stroke, with broader clinical implications.

Novel variants with therapeutic potential

Although progress has been slower than initially hoped, information from GWASs is now being used to develop novel treatments for complex diseases, such as macular degeneration,[34] autoimmune disease[35] and hypercholesterolaemia.[36] Though still a long way from clinical translation, the *HDAC9* association identified in the WTCCC2 study highlights how GWASs might lead to new potential therapies for stroke.

HDAC9 belongs to a family of genes that encode proteins that deacetylate histones, thereby regulating chromatin structure and gene transcription.[37] *HDAC9* is ubiquitously expressed, with high levels of expression in cardiac muscle tissue and brain. The specific association of *HDAC9* with large-artery stroke suggests that the variant somehow increases the risk of atherosclerosis or plaque instability. One way of determining whether the *HDAC9* variant is associated with early atherosclerosis is to look for associations with asymptomatic carotid plaque and carotid intima–media thickness in community populations. One study that used ultrasonography to determine the presence of asymptomatic plaques in 25,179 individuals and carotid intima–media thickness in 31,210 individuals found statistically significant associations between the *HDAC9* variant and both of these phenotypes.[38]

HDAC9 is expressed in the endothelial and smooth muscle cells of both intracranial and systemic large arteries, including the carotid, middle cerebral and coronary arteries.[39] Levels of *HDAC9* mRNA expression were found to be elevated above normal levels in carotid atherosclerotic plaques compared with nondiseased arterial wall.[38] Interestingly, this elevation was seen to a greater degree in carotid plaques than in atherosclerotic plaques elsewhere in the body. The reasons for this difference are unknown, although the *HDAC9* variant was recently associated with coronary artery disease,[40] and it is of interest that this association was weaker than the association with large-artery stroke.

Taken together, these data suggest that the *HDAC9* variant increases atherosclerosis, particularly at the carotid bifurcation. This raises the intriguing possibility that modulation of this pathway might reduce atherosclerosis and the risk of large-artery stroke. The commonly used antiepileptic drug sodium valproate nonspecifically inhibits histone deacetylase 9, the protein product of *HDAC9*, and can protect against atherosclerosis in animal models.[41] Sodium valproate therapy was associated with lower stroke and myocardial risk rates compared with other antiepileptic drugs in a large community-based study,[42] although a replication study provided more-ambiguous results.[43] In this second study, treatment with sodium valproate was associated with a reduced risk of myocardial infarction, but not with a reduced risk of ischaemic stroke overall. However, ischaemic stroke risk correlated inversely with the duration of sodium valproate treatment.[43]

Further studies are now needed to map the 7p21.1 locus in detail and identify causal variants to confirm that these affect the *HDAC9* gene function. Functional studies and intervention studies using histone deacetylase inhibitors in animal models will provide further information as to whether inhibition of histone deacetylase 9 does indeed offer a feasible approach to stroke prevention.

Insights into stroke pathophysiology

The fact that the majority of associations reported in GWASs of ischaemic stroke have been specific to one subtype of stroke (Table 1) indicates that the genetic architecture of the various subtypes differs substantially. This is consistent with previous observations, such as differential associations between family history and specific stroke subtypes,[3] and the finding that most monogenic causes of stroke predispose to a single subtype.[44] GWASs have now provided convincing support for this hypothesis.

The confirmation that different stroke subtypes have different pathophysiological mechanisms raises the possibility that subtype-specific treatments might be necessary. However, with a few exceptions (such as carotid endarterectomy for symptomatic carotid stenosis, and anticoagulation therapy for secondary prevention of cardioembolic stroke due to atrial fibrillation), the vast majority of stroke treatments currently available have been tested in all subtypes of ischaemic stroke.

Stroke GWASs have emphasized the need for rigorous subtyping by means of a pathophysiological classification, such as the system developed for the Trial of Org 10172 in Acute Stroke Treatment (TOAST) study, which classifies stroke according to aetiology.[45] Only by classing stroke into these pathophysiological subtypes has it been possible to identify most genetic associations. However, classification with the TOAST system requires detailed investigation, and exclusion of pathology such as large-artery stenosis and cardioembolic sources; such information is frequently not available in large clinical trials.

Some clinical trials have attempted secondary analyses to determine treatment efficacy in different subtypes of stroke; however, the subtyping has usually been inadequate to obtain reliable data. A major limitation in previous stroke trials that attempted such classification is that they often used classifications that were developed before brain imaging was widely available, thereby relying on the clinical syndrome rather than test results. As an example of the problems that can arise with these classifications, studies have shown that small-vessel disease can be markedly over-diagnosed if the clinical Oxfordshire Community Stroke Project
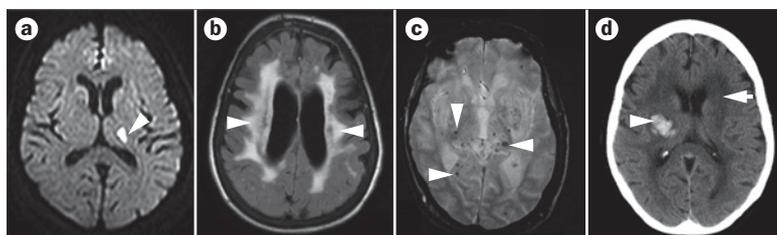
Figure 1 | Appearance of different consequences of cerebral small-vessel disease in brain images. **a** | Acute lacunar infarct, detected with diffusion-weighted MRI and visible as a small region of high signal in the left thalamus (arrowhead). **b** | Leukoaraiosis detected with a fluid-attenuated inversion recovery MRI sequence and visible as areas of high signal in the white matter (arrowheads). **c** | Microbleeds detected with gradient echo MRI and visible as small regions of low signal caused by haemosiderin deposition from breakdown of red blood cells (arrowheads). **d** | Right-sided subcortical haemorrhage detected with a CT scan and visible as an area of high signal (arrowhead) in the right posterior limb of the internal capsule and surrounding region. This patient also has extensive leukoaraiosis, which is visible as areas of low signal in CT scans (arrow).

(OSCP) classification is used to diagnose the lacunar syndrome category, owing to inclusion of a number of individuals with stroke due to larger nonlacunar infarcts in this group.[46,47]

Insights into specific pathogenic mechanisms

Genetic information can provide insight into specific pathogenic mechanisms, particularly where there are questions regarding whether a disease risk factor plays a causal role or whether the observed association is a result of confounding factors. An illustrative case is provided by determination of the appropriate use of antithrombotic drugs in secondary prevention of ischaemic stroke.

Considerable evidence—almost all from trials that involved all ischaemic stroke types—shows benefits of antiplatelet agents in secondary stroke prevention. A growing body of data suggests that intensive antiplatelet regimens, such as aspirin and clopidogrel, could be beneficial in large-artery stroke, particularly during the first few months after the event.[48,49] However, concern has been expressed that these regimens might be associated with an increased risk of intracerebral bleeding in small-vessel stroke.

The underlying pathophysiology of lacunar or small-vessel stroke is less well understood than that of large-artery and cardioembolic stroke. Lacunar stroke seems to belong to a spectrum of small-artery disease, other features of which are white matter hyperintensities—or leukoaraiosis—seen on T2-weighted MRI, asymptomatic microbleeds detected on gradient echo MRI, and symptomatic subcortical haemorrhage (Figure 1).[50] One study found overlap between the genetic basis of subcortical haemorrhage and ischaemic small-vessel stroke.[51]

Clinical studies have shown that warfarin treatment increases the risk of intracerebral haemorrhage among patients with lacunar stroke and leukoaraiosis.[52,53] This observation demonstrates that the potential overlap between lacunar stroke and intracerebral haemorrhage makes optimal antithrombotic treatment crucial in patients with lacunar stroke; genetic studies could help with the optimization.

The role of thrombotic factors in lacunar stroke could be evaluated by comparing blood levels of haemostatic markers among patients who have this condition with levels in healthy controls, but causality is difficult to determine because elevated levels of these markers could be secondary to the development of stroke rather than being a predisposing factor. A technique called Mendelian randomization is increasingly being used to determine causality in such situations.[54] In this technique, one selects a functional variant that is associated with an increased level of the marker in question, and determines whether it is also associated with an increased risk of the phenotype, in this case lacunar stroke. This approach has been used successfully in many cardiovascular diseases; for example, it has shown that elevated levels of C-reactive protein in patients with cardiovascular disease are likely to be secondary to atherosclerosis rather than having a causal role in disease pathogenesis.[55]

One study used Mendelian randomization to investigate the role of haemostatic factors in ischaemic stroke.[56] A GWAS identified a variant in the *ABO* gene that had a genome-wide association with haemostatic and thrombotic markers. The study then tested whether this variant was also associated with ischaemic stroke and its subtypes. A statistically significant association was found with large-artery and cardioembolic stroke, but not with small-artery stroke.[57] These two associations are not unexpected, as thromboembolism from an activated carotid plaque or from the heart are the primary disease mechanisms in these stroke subtypes. However, the lack of association with lacunar stroke is interesting, and suggests that thrombotic factors may be less important for this stroke subtype.

This finding emphasizes the need for clinical trials that examine the risk-to-benefit ratio of antiplatelet agents as a therapeutic option for patients with small-vessel stroke. The first large randomized controlled trial to specifically recruit patients with MRI-confirmed lacunar stroke was published in 2012, and demonstrated that clopidogrel alone was more effective than clopidogrel combined with aspirin for long-term prevention of recurrent stroke after a symptomatic lacunar stroke.[57] Adding aspirin to the treatment provided little benefit in terms of stroke reduction, and significantly increased the risk of bleeding.

The future of GWASs

GWASs have started to provide novel insights into the molecular mechanisms that underlie genetic predisposition to stroke, and further studies in large, well-phenotyped populations will probably yield additional insights. It is important that a number of considerations are addressed to maximize the benefits of GWASs in identifying associations that have clinical potential.

Bigger sample sizes

Sample sizes analysed in GWASs of stroke are still some way behind those for other complex diseases. The METASTROKE meta-analysis of ischaemic stroke GWASs brought together over 10,000 cases, but this

number only included approximately 2,000 cases of each subtype of ischaemic stroke.[20] Meta-analyses of GWAS data in other complex diseases, such as hypertension, diabetes mellitus and coronary artery disease, with cohorts that range from tens of thousands to hundreds of thousands of patients, consistently show that the number of variants identified is highly dependent on the sample size.[58] One challenge in stroke is that studies must be adequately powered to detect associations with individual stroke subtypes, which requires larger overall numbers of participants than for many other complex diseases. Therefore, as stroke sample sizes increase further, it is highly likely that many additional associations will be determined. Further large GWASs of stroke, including the National Institute of Neurological Disorders and Stroke SiGN study, which will contribute another 10,000 stroke cases, are nearing completion.[59]

Accurate classification of subtypes

The subtype specificity of associations identified to date highlights the importance of accurate phenotyping with a system that classifies according to pathophysiology. The fact that subtype-specific associations have been found by using the TOAST classification system is reassuring, but improved subtyping classifications may increase the yield further. A computerized algorithm may reduce inter-observer differences in classification, and this approach has been successfully applied to the TOAST system.[60]

Difficulties with the TOAST system include identifying the most likely cause where multiple potential mechanisms are found, and uncertainty regarding how to deal with cases with incomplete diagnostic evaluation. To address these difficulties, modifications to the TOAST system have been developed, including the ASCO (Atherosclerosis, Small-vessel disease, Cardiac source, Other cause) system,[61] and the web-based Causative Classification of Stroke (CCS), which is specifically recommended for genetic studies.[62] Comparisons of these systems have generally shown good agreement,[63,64] but future studies will reveal any differences in their ability to detect genome-wide associations.

Sex and ethnic differences

GWASs have confirmed many findings from twin and family history studies, including different genetic architectures for different stroke subtypes;[19] greater importance of some variants for younger-onset stroke;[24] and shared genetic risk factors for coronary artery disease and stroke, particularly large-artery stroke.[65] Whether or not sex differences are important, as suggested by family history data,[30] remains to be determined by GWASs. Further work will also determine the extent to which identified genetic risk factors contribute to stroke via conventional stroke risk factors and independent effects. A genetic risk score based on risk factors already identified in GWASs did correlate with stroke, but did not add predictive information.[66]

Extensive epidemiological data demonstrate marked ethnic differences in stroke incidence and the distribution of subtypes. For example, intracranial stenosis and intracerebral haemorrhage are more frequent in Far East Asian populations compared with white populations,[67,68] and small-vessel disease is more common among individuals of African ancestry living in London, UK or in the USA than in white individuals living in the same countries.[47,69] However, few studies have been conducted in non-white populations, and more are needed. Whether or not the variants associated with increased stroke risk in white populations also increase risk in other ethnicities remains to be determined, although findings from other complex diseases suggest that most associations are likely to be shared by multiple ethnicities. Admixture mapping using GWAS data, which relies on genetic variants that differ in frequency across populations, might reveal the contribution of genetic factors to ethnic differences in stroke, and identify novel genetic variants; this approach has been successful in other complex traits.[70]

Emerging techniques in stroke genetics

While GWASs have provided most of the genetic associations with stroke that have been identified to date, several new technologies could complement GWAS findings and provide further insights into the genetics of stroke.

GWASs are designed to identify common variants that contribute to disease risk, whereas whole-genome sequencing can identify rare variants that might contribute. The cost of whole-genome sequencing has fallen dramatically in recent years, and is now just US$1,000 per genome,[71] so could be a more cost-effective approach to screening for multiple monogenic causes of stroke in a single assay than is the current gene-by-gene testing approach.[72] The clinical benefits of next-generation sequencing are currently being assessed in projects such as the Genomics England 100,000 Genomes Project.[73]

Whole-exome and whole-genome sequencing are also being used to identify novel rare variants that could contribute to common polygenic diseases, including stroke.[74] Studies of multiple family members in combination with these techniques are likely to be particularly useful for determining which variants are causal and which are non-causal. Identification of rare variants that contribute to complex polygenic disease such as apparently sporadic stroke will present challenges because larger sample sizes are required to detect the low-frequency variants, unless the odds ratios associated with such variants are much greater than those associated with GWAS-derived variants.

Epigenetic changes such as DNA methylation and histone modification do not alter the underlying DNA sequence, but are heritable and can affect gene expression and cellular phenotype, and could play a part in stroke pathogenesis. Array-based methods can analyse patterns of DNA methylation in large populations, akin to GWASs at the DNA level. With small patient cohorts, this approach has demonstrated differential methylation in stroke-relevant risk factors such as BMI,[75] lipid levels[76] and smoking status.[77] In view of the probable importance of gene–environment interactions in cardiovascular

disease and stroke, this approach may be particularly interesting in stroke.

In light of GWAS findings, there is also a move towards investigating the interactions of the transcriptome and proteome with each other and with the genome in relation to disease phenotypes. Identification of motifs such as transcription factor binding sites on a genome-wide basis is already possible with techniques such as chromatin immunoprecipitation followed by next-generation sequencing,[78] although interpretation is limited by availability and ongoing development of tools for analysis.[79] However, discovering how the interactions between the genome, transcriptome and proteome lead to the final phenotype is becoming a major focus in unravelling the molecular pathology of disease.

Conclusions

The genetic variants that have been associated with ischaemic stroke in GWASs to date account for only a small proportion of the overall risk of stroke, so the clinical impact of the findings has so far been limited. However, the novel insights into the pathogenesis of ischaemic stroke and its subtypes that are provided by these associations can help to direct clinical practice to achieve the best outcomes.

Treatment approaches that take into account the different pathophysiologies of the subtypes of stroke are likely to be more successful than approaches that are used for all subtypes of stroke. However, randomized clinical trials will be required to determine whether responses to treatments are indeed subtype-specific, and these trials will require the use of a detailed pathophysiological system such as the TOAST classification system to ensure that subtypes are classified correctly and consistently.

For many of the genetic variants that have been associated with ischaemic stroke in GWASs, the exact causal relationship is yet to be determined. Where individual genes are suspected to be responsible for the associations with specific loci, such as *HDAC9*, functional studies are needed to determine the precise mechanisms of stroke pathogenesis and to enable the development of new therapeutics that target these pathways. Focused genetic studies that use techniques such as Mendelian randomization might provide further insights into pathogenesis, particularly the role of specific molecules or pathways in disease and whether their associations with stroke are causal.

The fact that GWASs of stroke have until now included relatively small numbers of patients means that applying GWASs to much larger stroke cohorts is likely to identify further genetic associations. Furthermore, novel techniques, particularly next-generation sequencing, are likely to complement and add to GWAS findings, for example, by identifying rare genetic associations that GWASs could miss. Such future work is likely to further improve both disease understanding and clinical care.

Review criteria

PubMed was searched on 16th August 2014 using the term "stroke AND genetics". All articles identified were English-language, full-text papers. We also searched the reference lists of important articles and reviews to find further papers.

1. Lanfranconi, S. & Markus, H. S. Stroke subtyping for genetic association studies? A comparison of the CCS and TOAST classifications. *Int. J. Stroke* **8**, 626–631 (2013).
2. Flossmann, E., Schulz, U. G. & Rothwell, P. M. Systematic review of methods and results of studies of the genetic epidemiology of ischemic stroke. *Stroke* **35**, 212–227 (2004).
3. Jerrard-Dunne, P., Cloud, G., Hassan, A. & Markus, H. S. Evaluating the genetic component of ischemic stroke subtypes: a family history study. *Stroke* **34**, 1364–1369 (2003).
4. Seshadri, S. *et al.* Parental occurrence of stroke and risk of stroke in their children: the Framingham study. *Circulation* **121**, 1304–1312 (2010).
5. Khaleghi, M. *et al.* Family history as a risk factor for carotid artery stenosis. *Stroke* **45**, 2252–2256 (2014).
6. Banerjee, A. *et al.* Familial history of stroke is associated with acute coronary syndromes in women. *Circ. Cardiovasc. Genet.* **4**, 9–15 (2011).
7. Kennedy, R. E. *et al.* Association between family risk of stroke and myocardial infarction with prevalent risk factors and coexisting diseases. *Stroke* **43**, 974–979 (2012).
8. Touzé, E. & Rothwell, P. M. Sex differences in heritability of ischemic stroke: a systematic review and meta-analysis. *Stroke* **39**, 16–23 (2008).
9. Bevan, S. *et al.* Genetic heritability of ischemic stroke and the contribution of previously reported candidate gene and genomewide associations. *Stroke* **43**, 3161–3167 (2012).
10. Dichgans, M. & Markus, H. S. Genetic association studies in stroke: methodological issues and proposed standard criteria. *Stroke* **36**, 2027–2031 (2005).
11. Hindorff, L. A. *et al.* A catalog of published genome-wide association studies. *National Human Genome Research Institute* [online], http://www.genome.gov/gwastudies/ (2014).
12. Woo, D. *et al.* Meta-analysis of genome-wide association studies identifies 1q22 as a susceptibility locus for intracerebral hemorrhage. *Am. J. Hum. Genet.* **94**, 511–521 (2014).
13. Alg, V. S., Sofat, R., Houlden, H. & Werring, D. J. Genetic risk factors for intracranial aneurysms: a meta-analysis in more than 116,000 individuals. *Neurology* **80**, 2154–2165 (2013).
14. Gretarsdottir, S. *et al.* Risk variants for atrial fibrillation on chromosome 4q25 associate with ischemic stroke. *Ann. Neurol.* **64**, 402–409 (2008).
15. Gudbjartsson, D. F. *et al.* A sequence variant in *ZFHX3* on 16q22 associates with atrial fibrillation and ischemic stroke. *Nat. Genet.* **41**, 876–878 (2009).
16. Lemmens, R. *et al.* The association of the 4q25 susceptibility variant for atrial fibrillation with stroke is limited to stroke of cardioembolic etiology. *Stroke* **41**, 1850–1857 (2010).
17. Helgadottir, A. *et al.* A common variant on chromosome 9p21 affects the risk of myocardial infarction. *Science* **316**, 1491–1493 (2007).
18. Gschwendtner, A. *et al.* Sequence variants on chromosome 9p21.3 confer risk for atherosclerotic stroke. *Ann. Neurol.* **65**, 531–539 (2009).
19. International Stroke Genetics Consortium (ISGC) *et al.* Genome-wide association study identifies a variant in *HDAC9* associated with large vessel ischemic stroke. *Nat. Genet.* **44**, 328–333 (2012).
20. Traylor, M. *et al.* Genetic risk factors for ischaemic stroke and its subtypes (the METASTROKE collaboration): a meta-analysis of genome-wide association studies. *Lancet Neurol.* **11**, 951–962 (2012).
21. Han, Y. *et al.* HDAC9 gene is associated with stroke risk in a Chinese population. *Exp. Biol. Med. (Maywood)* **238**, 842–847 (2013).
22. ENCODE Project Consortium. A user's guide to the encyclopedia of DNA elements (ENCODE). *PLoS Biol.* **9**, e1001046 (2011).
23. Holliday, E. G. *et al.* Common variants at 6p21.1 are associated with large artery atherosclerotic stroke. *Nat. Genet.* **44**, 1147–1151 (2012).
24. Traylor, M. *et al.* A novel *MMP12* locus is associated with large artery atherosclerotic stroke using a genome-wide age-at-onset informed approach. *PLoS Genet.* **10**, e1004469 (2014).
25. Kubo, M. *et al.* A nonsynonymous SNP in *PRKCH* (protein kinase C η) increases the risk of cerebral infarction. *Nat. Genet.* **39**, 212–217 (2007).
26. Serizawa, M. *et al.* Association between *PRKCH* gene polymorphisms and subcortical silent brain infarction. *Atherosclerosis* **199**, 340–345 (2008).

27. Ikram, M. A. *et al.* Genomewide association studies of stroke. *N. Engl. J. Med.* **360**, 1718–1728 (2009).

28. International Stroke Genetics Consortium & Wellcome Trust Case–Control Consortium 2. Failure to validate association between 12p13 variants and ischemic stroke. *N. Engl. J. Med.* **362**, 1547–1550 (2010).

29. Kilarski, L. *et al.* Meta-analysis in more than 17,900 cases of ischemic stroke reveals a novel association at 12q24.12. *Neurology* **19**, 678–685 (2014).

30. Hassan, A., Sham, P. & Markus, H. S. Planning genetic studies in human stroke: sample size estimates based on family history data. *Neurology* **58**, 1483–1488 (2002).

31. Kraft, P. & Hunter, D. J. Genetic risk prediction —are we there yet? *N. Engl. J. Med.* **360**, 1701–1703 (2009).

32. FDA bans 23andme personal genetic tests. *BBC News* [online], http://www.bbc.co.uk/news/technology-25100878 (2013).

33. Schneider, K. I. & Schmidtke, J. Patient compliance based on genetic medicine: a literature review. *J. Community Genet.* **5**, 31–48 (2014).

34. Rohrer, B. *et al.* A targeted inhibitor of the alternative complement pathway reduces angiogenesis in a mouse model of age-related macular degeneration. *Invest. Ophthalmol. Vis. Sci.* **50**, 3056–3064 (2009).

35. Visscher, P. M., Brown, M. A., McCarthy, M. I. & Yang, J. Five years of GWAS discovery. *Am. J. Hum. Genet.* **90**, 7–24 (2012).

36. Dadu, R. T. & Ballantyne, C. M. Lipid lowering with PCSK9 inhibitors. *Nat. Rev. Cardiol.* **11**, 563–575 (2014).

37. Haberland, M., Montgomery, R. L. & Olson, E. N. The many roles of histone deacetylases in development and physiology: implications for disease and therapy. *Nat. Rev. Genet.* **10**, 32–42 (2009).

38. Markus, H. S. *et al.* Evidence HDAC9 genetic variant associated with ischemic stroke increases risk via promoting carotid atherosclerosis. *Stroke* **44**, 1220–1225 (2013).

39. Langley, B., Brochier, C. & Rivieccio, M. A. Targeting histone deacetylases as a multifaceted approach to treat the diverse outcomes of stroke. *Stroke* **40**, 2899–2905 (2009).

40. Schunkert, H. *et al.* Large-scale association analysis identifies 13 new susceptibility loci for coronary artery disease. *Nat. Genet.* **43**, 333–338 (2011).

41. Bowes, A. J., Khan, M. I., Shi, Y., Robertson, L. & Werstuck, G. H. Valproate attenuates accelerated atherosclerosis in hyperglycemic apoE-deficient mice: evidence in support of a role for endoplasmic reticulum stress and glycogen synthase kinase-3 in lesion development and hepatic steatosis. *Am. J. Pathol.* **174**, 330–342 (2009).

42. Olesen, J. B. *et al.* Effects of epilepsy and selected antiepileptic drugs on risk of myocardial infarction, stroke, and death in patients with or without previous stroke: a nationwide cohort study. *Pharmacoepidemiol. Drug Saf.* **20**, 964–971 (2011).

43. Dregan, A., Charlton, J. Wolfe, C. D., Gulliford, M. C. & Markus, H. S. Is sodium valproate, an HDAC inhibitor, associated with reduced risk of stroke and myocardial infarction? A nested case–control study. *Pharmacoepidemiol. Drug Saf.* **23**, 759–767 (2014).

44. Dichgans, M. Genetics of ischaemic stroke. *Lancet Neurol.* **6**, 149–161 (2007).

45. Adams, H. P. Jr *et al.* Classification of subtype of acute ischemic stroke. Definitions for use in a multicenter clinical trial. TOAST. Trial of Org 10172 in Acute Stroke Treatment. *Stroke* **24**, 35–41 (1993).

46. Bamford, J., Sandercock, P., Jones, L. & Warlow, C. The natural history of lacunar infarction: the Oxfordshire Community Stroke Project. *Stroke* **18**, 545–551 (1987).

47. Markus, H. S. *et al.* Differences in stroke subtypes between black and white patients with stroke: the South London Ethnicity and Stroke Study. *Circulation* **116**, 2157–2164 (2007).

48. Markus, H. S. *et al.* Dual antiplatelet therapy with clopidogrel and aspirin in symptomatic carotid stenosis evaluated using Doppler embolic signal detection: the Clopidogrel and Aspirin for Reduction of Emboli in Symptomatic Carotid Stenosis (CARESS) trial. *Circulation* **111**, 2233–2240 (2005).

49. Wong, K. S. *et al.* Early dual versus mono antiplatelet therapy for acute non-cardioembolic ischemic stroke or transient ischemic attack: an updated systematic review and meta-analysis. *Circulation* **128**, 1656–1666 (2103).

50. Pantoni, L. Cerebral small vessel disease: from pathogenesis and clinical characteristics to therapeutic challenges. *Lancet Neurol.* **9**, 689–701 (2010).

51. Anderson, C. D. *et al.* Common variants within oxidative phosphorylation genes influence risk of ischemic stroke and intracerebral hemorrhage. *Stroke* **44**, 612–619 (2013).

52. Smith, E. E., Rosand, J., Knudsen, K. A., Hylek, E. M. & Greenberg, S. M. Leukoaraiosis is associated with warfarin-related hemorrhage following ischemic stroke. *Neurology* **59**, 193–197 (2002).

53. [No authors listed] A randomized trial of anticoagulants versus aspirin after cerebral ischemia of presumed arterial origin. The Stroke Prevention in Reversible Ischemia Trial (SPIRIT) Study Group. *Ann. Neurol.* **42**, 857–865 (1997).

54. Ebrahim, S. & Davey Smith, G. Mendelian randomization: can genetic epidemiology help redress the failures of observational epidemiology? *Hum. Genet.* **123**, 15–33 (2008).

55. Kuper, H. *et al.* Evaluating the causal relevance of diverse risk markers: horizontal systematic review. *BMJ* **339**, b4265 (2009).

56. Williams, F. M. Ischemic stroke is associated with the ABO locus: the EuroCLOT study. *Ann. Neurol.* **73**, 16–31 (2013).

57. SPS3 Investigators *et al.* Effects of clopidogrel added to aspirin in patients with recent lacunar stroke. *N. Engl. J. Med.* **367**, 817–825 (2012).

58. Visscher, P. M., Brown, M. A., McCarthy, M. I. & Yang, J. Five years of GWAS discovery. *Am. J. Hum. Genet.* **90**, 7–24 (2012).

59. Meschia, J. F. *et al.* Stroke Genetics Network (SiGN) study: design and rationale for a genome-wide association study of ischemic stroke subtypes. *Stroke* **44**, 2694–2702 (2013).

60. Goldstein, L. B. *et al.* Improving the reliability of stroke subgroup classification using the Trial of ORG 10172 in Acute Stroke Treatment (TOAST) criteria. *Stroke* **32**, 1091–1098 (2001).

61. Amarenco, P., Bogousslavsky, J., Caplan, L. R., Donnan, G. A. & Hennerici, M. G. New approach to stroke subtyping: the A-S-C-O (phenotypic) classification of stroke. *Cerebrovasc. Dis.* **27**, 502–508 (2009).

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

63. Marnane, M. *et al.* Stroke subtype classification to mechanism-specific and undetermined categories by TOAST, A-S-C-O, and causative classification system: direct comparison in the North Dublin population stroke study. *Stroke* **41**, 1579–1586 (2010).

64. Lanfranconi, S. & Markus, H. S. Stroke subtyping for genetic association studies? A comparison of the CCS and TOAST classifications. *Int. J. Stroke* **8**, 626–631 (2013).

65. Dichgans, M. *et al.* Shared genetic susceptibility to ischemic stroke and coronary artery disease: a genome-wide analysis of common variants. *Stroke* **45**, 24–36 (2014).

66. Malik, R. *et al.* Multilocus genetic risk score associates with ischemic stroke in case–control and prospective cohort studies. *Stroke* **45**, 394–402 (2104).

67. Tsai, C. F., Thomas, B. & Sudlow, C. L. Epidemiology of stroke and its subtypes in Chinese vs white populations: a systematic review. *Neurology* **81**, 264–272 (2103).

68. Holmstedt, C. A., Turan, T. N. & Chimowitz, M. I. Atherosclerotic intracranial arterial stenosis: risk factors, diagnosis, and treatment. *Lancet Neurol.* **12**, 1106–1114 (2103).

69. Woo, D. *et al.* Incidence rates of first-ever ischemic stroke subtypes among blacks: a population-based study. *Stroke* **30**, 2517–2522 (1999).

70. Divers, J. *et al.* Admixture mapping of coronary artery calcified plaque in African Americans with type 2 diabetes mellitus. *Circ. Cardiovasc. Genet.* **6**, 97–105 (2013).

71. Sheridan, C. Illumina claims $1,000 genome win. *Nat. Biotechnol.* **32**, 115 (2014).

72. Singleton, A. B. Exome sequencing: a transformative technology. *Lancet Neurol.* **10**, 942–946 (2011).

73. The 100,000 Genomes Project. *Genomics England* [online], http://www.genomicsengland.co.uk/the-100000-genomes-project/ (2014).

74. Shirodkar, A. V. & Marsden, P. A. Epigenetics in cardiovascular disease. *Curr. Opin. Cardiol.* **26**, 209–215 (2011).

75. Dick, K. J. *et al.* DNA methylation and body-mass index: a genome-wide analysis. *Lancet* **383**, 1990–1998 (2104).

76. Irvin, M. R. *et al.* Epigenome-wide association study of fasting blood lipids in the genetics of lipid-lowering drugs and diet network study. *Circulation* **130**, 565–572 (2014).

77. Breitling, L. P., Yang, R., Korn, B., Burwinkel, B. & Brenner, H. Tobacco-smoking-related differential DNA methylation: 27K discovery and replication. *Am. J. Hum. Genet.* **88**, 450–457 (2011).

78. Furey, T. S. ChIP-seq and beyond: new and improved methodologies to detect and characterize protein-DNA interactions. *Nat. Rev. Genet.* **13**, 840–852 (2012).

79. Yan, H. *et al.* HiChIP: a high-throughput pipeline for integrative analysis of ChIP-Seq data. *BMC Bioinformatics* **15**, 280 (2014).

Author contributions
H.S.M. researched data for the article, and wrote the article. S.B. contributed to review and editing of the manuscript before submission.