Wellcome Trust Genome-Wide Association Study of Ischemic Stroke

Hugh S. Markus, DM

Twin and family history studies suggest that genetic risk factors are important for ischemic stroke. However, identifying the underlying genes has proved challenging. Until recently, the main technique used was the candidate gene method. In this, genetic variants, usually single nucleotide type polymorphisms (SNPs), are identified in a proposed candidate gene, which is thought to be involved in stroke risk. The frequency of the SNP in patients with stroke compared with controls is then determined. Many candidate gene studies have been published in stroke, but the results of these have been disappointing with few replicable associations. This situation is common with the genetics of many other complex diseases, and the reasons for lack of success include small sample sizes, a failure to adequately phenotype stroke subtypes, and a lack of primary replication of positive associations. An additional problem with candidate gene studies is that associations can only be identified in genes already known about and implicated in stroke risk; therefore, completely novel genes cannot be identified.

The genetics of complex diseases such as stroke, in which multiple genes interact with environmental risk factors to increase risk, has been revolutionized by the Genome-Wide Association Study (GWAS) approach. This uses microarray technology to genotype up to ≥1 million SNPs, which span the whole genome. A case–control methodology is used to compare the frequency of individual SNPs between cases and controls. This is combined with rigorous multiple comparison techniques to account for the many associations tested. A great advantage of the GWAS approach, in contrast to the candidate gene method, is that it allows associations between completely novel chromosomal loci and disease to be identified. This technology has been combined with much improved study design with both very large sample sizes and replication of positive associations before publication. GWAS had resulted in >1600 novel associations with many complex diseases reported by September 2011. Novel genetic associations have been reported with many cardiovascular diseases, including myocardial infarction, diabetes mellitus, hypertension, and hyperlipidemia. Most genetic variants discovered using GWAS account for only a small increase in disease risk with odds ratios commonly between 1.1 and 1.3. An implication of this is that large sample sizes are required to identify such variants, and this has resulted in the formation of disease consortiums, which combine data from multiple studies in meta-analysis, often in tens of thousands of cases.

GWAS studies in stroke have progressed less rapidly than in other cardiovascular diseases, but the approach is now beginning to identify novel genetic variants.

Initial studies attempted to replicate associations that had initially been found by applying the GWAS method to other diseases, which themselves are associated with increased stroke risk. Two variants (PITX2 and ZFHX3), which were initially associated with atrial fibrillation, have both been shown to be independent risk factors for ischemic stroke. These associations were only apparent with cardioembolic stroke and not with other stroke subtypes. An association between SNPs in the chromosome 9p21 region and coronary artery disease has been reported by many groups. A meta-analysis across multiple ischemic stroke populations demonstrated an association with large artery stroke independently of vascular risk factors and coronary artery disease. No association was found with other stroke subtypes. This locus has also been associated with abdominal aortic aneurysms and intracranial aneurysms.

Until recently, there have been few results available from GWAS studies identifying novel associations with ischemic stroke. In 2012, the Wellcome Trust Case Control Consortium 2 (WTCCC2) Ischemic Stroke GWAS published its initial results. WTCCC2 performed a GWAS for ischemic stroke and its 3 major subtypes (cardioembolic stroke, large artery stroke, and small-vessel disease) in 3548 affected individuals and 5972 controls, all of European ancestry. These discovery cohort cases were recruited from the United Kingdom and Germany. Replication of potential signals was performed in 5859 affected individuals and 6281 controls, also of European ancestry. Previous associations with cardioembolic stroke at PITX2 and ZFHX3, and for large artery stroke at the 9p21 locus, were confirmed. In addition, a novel association for large artery stroke was identified on chromosome 7p21.1. This association was further replicated in an additional 735 cases with large artery stroke and 28 583 controls. The top SNP (RS11984041) was associated with a combined P value of 1.87×10^{-11}. The odds ratio was 1.42 (95% confidence intervals, 1.28–1.57).
All variants with an association signal in the 7p21.1 locus resided within a peak between 2 combination hotspots that encompasses the tail end of a protein called histone deacetylase 9 (HDAC9). This is therefore, the most likely gene underlying the association. However, the downstream TWIST1 and FERD3L genes are relatively close to the identified peak and cannot be excluded as possible mechanisms through which genetic variants might exert cis effects.

HDAC9 is a member of a large family of genes that encode proteins that deacetylate histones, thereby regulating chromatin structure and gene transcription. HDAC9 is ubiquitously expressed with high levels of expression in cardiac tissue, muscle, and brain. Although called HDACs, these proteins also act on other substrates and can lead to both upregulation and downregulation of genes.

How variants in the HDAC9 region increase large artery stroke is not yet clear. The specific association with this stroke subtype would be consistent with the variant increasing the risk of atherosclerosis. The HDAC9 protein inhibits myogenesis and is involved in heart development, but pathological effects on systemic arteries have not yet been reported. Alternatively, it could increase risk by altering brain ischemic responses and might therefore have effects on neuronal survival. The HDAC9 protein has been shown to protect apoptosis, both by inhibiting JUN phosphorylation by MAPK10 and by repressing JUN transcription. HDAC inhibitors have been suggested as a treatment for stroke, and evidence in experimental models supports such an effect. However, an effect via increasing neuronal survival might also be expected to increase risk for stroke caused by cardiac embolism as well as large artery embolism, and there was no evidence of any association between the 7p21.1 locus and cardioembolic stroke. Initial studies in our laboratory have shown that HDAC9 is expressed in both intracranial and systemic large arteries, including carotid and coronary arteries. Abundant staining was found both in the endothelium and smooth muscle cells. Staining was present in both nuclei and cytoplasmic locations. We also assessed the effect of pulsatile stretch on HDAC9 expression in vascular smooth muscle cells, which play a central role in atherogenesis. HDAC was detected in aortic vascular smooth muscle cell in vitro by both Western blot and immunofluorescence. In cultures subjected to pulsatile stretching for 19 hours, HDAC9 abundance rose by a factor of 4 and then declined during the succeeding static period returning to near resting levels.

The commonly used antiepileptic drug sodium valproate has HDAC inhibitory properties and has been shown to inhibit atherosclerosis in animal models. Intriguingly, sodium valproate therapy in man has been associated with lower stroke and myocardial infarction rates compared with other antiepileptic drugs. Further studies are now required to fine map the 7p21.1 locus and confirm that the at risk variants do indeed lie within the HDAC9 gene. Functional studies are required to understand how these variants result in an increased stroke risk.

The WTCCC2 Ischemic Stroke Study shows that novel associations can be identified with ischemic stroke using the GWAS approach. Analyses of associations of PITX2 and ZFHX3 and HDAC9 across both discovery and replication populations, using a Bayesian modeling approach, suggested these loci were associated with specific stroke subtypes, consistent with different genetic architecture for different stroke subtypes. This finding has important implications. First, if studies identifying novel genetic risk factors are to be successful, careful stroke subtyping is important. Second, it suggests different stroke subtypes have different underlying pathophysiological mechanisms, and therefore, by implication may have different treatment responses.

Although the WTCCC2 Ischemic Stroke Study was the largest stroke study to date, the number of cases of each individual stroke subtype in a discovery population was modest, with 844 large artery stroke cases, 790 cardioembolic cases, and 580 small-vessel disease cases. In other complex diseases, studying much larger sample sizes has identified additional novel associations. The same situation may well apply to ischemic stroke. There have been a number of smaller GWASs in ischemic stroke. These have been brought together with WTCCC2 to form the Meta-Stroke Collaboration. This has provided further evidence for the association with HDAC9 and reinforced the subtyped specificity of associations reported to date. In addition, further large GWASs in ischemic stroke are underway. In particular, the National Institute of Neurological Disorders and Stroke–funded Stroke Genetics Network collaborative study will include genotyping on many thousands more cases.

The current genetic variants that have been identified for ischemic stroke explain only a small proportion of attributable risk. Genetic epidemiological studies have suggested that the sibling relative risk for stroke is ≈2 to 3. Assuming that genetic variants confer odds ratios of between 1.1 and 1.2, it has been calculated that 100 to 300 different variants will explain this order of relative risk. Therefore, we need to identify many more of the genetic variants accounting for stroke risk before such information is likely to have any useful predictive value. In the more immediate future, identifying novel variants may contribute to treating disease by identifying new pathways involved in disease pathogenesis. One criticism levelled at GWAS has been that the genetic associations identified are associations with only a small increase in risk (odds ratio) and, therefore, are likely to be unimportant. However, this does not mean that they will not tell us important things about disease processes. For example, the total variance in disease risk explained by genes involved in lipid pathways targeted by one of the most successful cardiovascular drugs we have, the statins, is small. Therefore, although only associated with moderate odds ratios, associations, such as HDAC9 with large artery stroke, may highlight disease pathways that can be treated with novel therapeutic approaches to reduce ischemic stroke risk.

Acknowledgments
The author thanks Gill Cockerill, PhD for providing the functional histone deacetylase 9 data.

Sources of Funding
The Wellcome Trust funded the Wellcome Trust Case Control Consortium 2 Ischemic Stroke Study.
Disclosures

None.

References
