In search of genes for stroke

See **Articles** page 414

Stroke is the third most common cause of death and the leading cause of long-term disability worldwide, but it has proven resistant to genetic investigation. Strong evidence suggests that genetic variation has a role in its pathogenesis, and although there has been no shortage of candidate gene studies, few (if any) genes have been irrefutably associated with stroke[1,2]—a story that has been repeated for a host of other common diseases.[3]

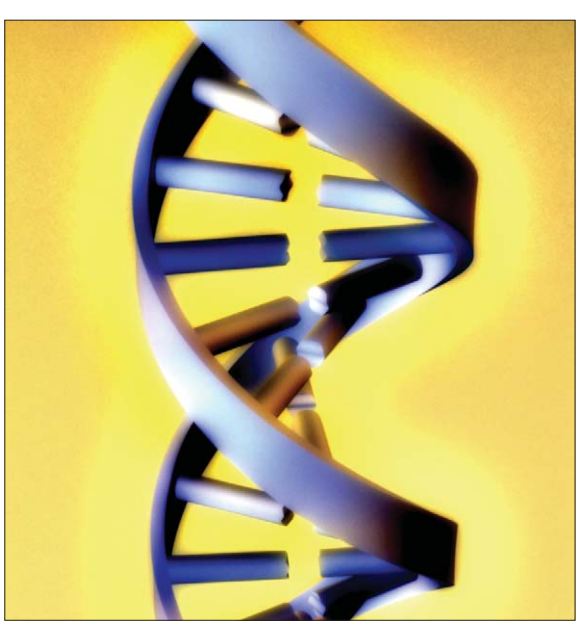
Relations between gene variants and the traits or diseases they influence generally fall into two models. For diseases that have a predictable familial pattern of inheritance, the gene variants or alleles responsible tend to be rare, and possession of the mutated gene is both necessary and sufficient to cause disease. Cerebral autosomal dominant arteriopathy with subcortical infarcts and leucoencephalopathy (CADASIL) is a notable example of a monogenic disease in which stroke is a cardinal manifestation. The vast majority of symptomatic strokes, however, do not cluster in families in any appreciable pattern. In late-onset complex traits, such as stroke, it is widely believed that multiple variants involving multiple genes each have a relatively modest effect on disease risk. Unlike the mutations in *NOTCH3* that cause CADASIL, these alleles are thought to be common among the general population and, by virtue of their prevalence, contribute substantially to disease burden despite their small effect size.

Whole-genome association studies, such as that described by Matarin and colleagues[4] in this issue, allow investigators, for the first time, to search for these common alleles of small-to-moderate effect with an unbiased approach, by testing known single-nucleotide polymorphisms (SNPs) across the genome for association with disease. Because these studies involve simultaneously testing hundreds of thousands of SNPs, the possibility of discovering a chance (ie, false positive) association becomes a certainty unless the p value threshold is set sufficiently low ($p<5\times10^{-7}$ in Matarin and colleagues' study, for example). Associations that are declared significant at such a low threshold, however, can still represent false-positive results. Ultimately, only replication of the finding in independent datasets can provide unequivocal evidence for a bona fide genotype–phenotype correlation.

Although conservative thresholds of significance are essential for such studies, they raise the possibility that SNPs that are truly associated with disease will be inappropriately rejected if their corresponding p values do not pass the predefined threshold. Only by increasing the number of study participants can investigators reduce both the possibility of a false-positive finding and the likelihood of rejecting SNPs that are truly associated with disease.

Indeed, lessons from recently released whole-genome studies underscore the importance of large sample sizes to improve statistical power to discover an association. Sladek and colleagues undertook a genome-wide association study of 694 type 2 diabetes patients and 669 controls, and fast-tracked the 59 most "promising" SNPs for replication in another 5511 samples.[5] They reported four novel susceptibility loci containing common variants (27–40% allele frequency) but with small effects on disease risk (odds ratios in the range of 1·14 to 1·27).

In this issue of *The Lancet Neurology*, Matarin and colleagues describe a genome-wide association study in which approximately 400 000 SNPs were genotyped in 249 patients with ischaemic stroke. Allele frequencies for each SNP were compared with those in 268 controls recruited as part of a prior study. This publication—the first genome-wide association study for stroke—is an important step forward for stroke genetics. The group responsible for ascertaining and phenotyping the cases have been pioneers in the development of



tools for multicentre genetic studies. Investigators in the International Stroke Genetics Study (ISGS)[6] and the Siblings With Ischemic Stroke Study (SWISS)[7] recognised that only through collaboration could large numbers of carefully phenotyped subjects be assembled; thus, these researchers developed and validated tools to ensure uniform characterisation of patients across multiple sites, and the studies have become models for development of subsequent multicentre genetic stroke studies.

Matarin and colleagues have set an important standard for those who follow, by committing to full release of the raw genotype and phenotype data (for almost all participants) and results of the association analysis into the public domain. This approach is being widely adopted throughout the genetics community, and assures maximum utility by enabling investigators worldwide to examine and analyse the data with fresh ideas.

A crucial limitation of the present study is its sample size. When completed, the researchers' study will include a total of 1500 stroke cases and 1500 controls. Power will exceed 80% only to detect SNPs with odds ratios greater than 1·5 and frequencies in the population that exceed 20% (at $p<5\times10^{-7}$). In fact, a study the size of the ISGS might not be large enough to identify alleles of smaller effect and lower frequency. Additional samples will also be needed for subsequent replication.

The stroke genetics community must now face the challenge of how best to assemble well-phenotyped cohorts that greatly exceed the size of the ISGS. Lessons from other diseases abound. For example, for multiple sclerosis, a disease in which genes appear to have an even larger role in susceptibility than in stroke, the International Multiple Sclerosis Genetics Consortium has assembled over 4000 cases.[8] Similar consortia exist for diabetes, coronary artery disease, psychiatric disease, cancer, inflammatory disease, and many others—some with as many as 10 000 cases assembled. Any large consortium devoted to finding stroke genes will have to decide on a consistent definition of disease phenotype. Matarin and colleagues have taken the approach of "lumping" rather than "splitting"—a reasonable approach because there are few data at present to suggest that the standard clinical stroke subtypes are genetically determined.[9,10] Given the heterogeneous pathophysiology of stroke, however, it is likely that novel, more biologically based phenotyping approaches will be necessary to find the range of genetic variants likely to contribute to cerebrovascular disease.

The discovery of genes that contribute to stroke risk will revolutionise the prevention of the leading cause of combined death and disability in the developed world and will help to identify novel pathways for pharmacological intervention. Matarin and colleagues have taken a first crucial step towards that goal.

Paul IW de Bakker, Jonathan Rosand
Center for Human Genetic Research, Massachusetts General Hospital, 185 Cambridge Street, Boston MA 02114, USA
jrosand@partners.org

We have no conflicts of interest.

1 Casas JP, Hingorani AD, Bautista LE, Sharma P. Meta-analysis of genetic studies in ischemic stroke: Thirty-two genes involving approximately 18 000 cases and 58 000 controls. *Arch Neurol* 2004; **61:** 1652–61.
2 Dichgans M. Genetics of ischaemic stroke. *Lancet Neurol* 2007; **6:** 149–61.
3 Hirschhorn JN, Lohmueller K, Byrne E, Hirschhorn K. A comprehensive review of genetic association studies. *Genet Med* 2002; **4:** 45-61.
4 Matarin M, Brown WM, Scholz, et al. A genome-wide SNP association study in patients with ischaemic stroke: initial analysis and data release. *Lancet Neurol* 2007; **6:** 414–20.
5 Sladek R, Rocheleau G, Rung J, et al. A genome-wide association study identifies novel risk loci for type 2 diabetes. *Nature* 2007; **445:** 881–85.
6 Meschia JF, Brott TG, Brown RD, et al. The Ischemic Stroke Genetics Study (ISGS) protocol. *BMC Neurol* 2003; **3:** 4.
7 Meschia JF, Brown RD, Jr, Brott TG, Chukwudelunzu FE, Hardy J, Rich SS. The Siblings With Ischemic Stroke Study (SWISS) protocol. *BMC Med Genet* 2002; **3:** 1.
8 International Multiple Sclerosis Genetics Consortium. A high-density screen for linkage in multiple sclerosis. *Am J Hum Genet* 2005; **77:** 454–67.
9 Schulz UG, Flossmann E, Rothwell PM. Heritability of ischemic stroke in relation to age, vascular risk factors, and subtypes of incident stroke in population-based studies. *Stroke* 2004; **35:** 819–24.
10 Wiklund PG, Brown WM, Brott TG, et al. Lack of aggregation of ischemic stroke subtypes within affected sibling pairs. *Neurology* 2007; **68:** 427–31.

Pathological gambling in Parkinson's disease

The rapid proliferation of opportunities to gamble money on the internet, in casinos, and even on some budget airlines has raised concern that pathological gambling could reach epidemic proportions, especially among young people in the near future. Pathological gambling was first reported with anti-Parkinsonian medication 7 years ago,[1] but raised few initial concerns. Subsequently, a number of tragic cases have been reported in the international press where fortunes have been recklessly dissipated and lives ruined. High-profile legal cases have followed, with pharmaceutical companies being sued for large settlements.[2] This attention has, in turn, fuelled