Many hypotheses but no replication for the association between *PDE4D* and stroke

To the Editor:

Stroke is the second-leading cause of death and a major cause of long-term disability throughout the world. Accumulated evidence suggests that inherited genetic variation plays a role in its pathogenesis. The need for identifying culprit genetic variants is great, since they may provide novel targets for the preventative therapeutics that are desperately needed.

A heterogeneous disorder, stroke is either ischemic, resulting from blockage of a blood vessel in the brain by a blood clot, or hemorrhagic, the result of rupture of one of these vessels. Roughly 85% of strokes are ischemic in nature, and these, in turn, are routinely divided into subtypes, each thought to represent a different pathophysiological process.

The candidate gene for stroke most intensively investigated thus far is *PDE4D*, following an initial report in *Nature Genetics*[1]. This publication identified a risk haplotype of a SNP (called SNP45) and microsatellite AC008818-1 that was associated with two of the major subtypes of stroke (cardiogenic and large-vessel) to the exclusion of the other subtypes. With accompanying *P* values on the order of 10^{-4} to 10^{-8}, it generated enormous enthusiasm and expectation in the stroke genetics community.

Numerous follow-up studies have since been published. Unfortunately, the validity of the original association remains unresolved, as these follow-up studies are roughly divided between those that claim replication[2–6] and those that do not[7–10]. Importantly, meta-analysis of the accumulated evidence has proven difficult, mainly because different groups examined different SNPs in this ~1.5-Mb gene (**Supplementary Table 1** online).

Heterogeneity among studies also extends to the phenotypes examined. Indeed, among those studies claiming association for *PDE4D*, implicated phenotypes have included all ischemic stroke, small-vessel stroke, large-vessel stroke, cardioembolic stroke and, emulating the approach of the original study[1], the combination of large-vessel and cardioembolic stroke. Of the three publications that included genotypes of microsatellite as well as SNP variants and conducted an analysis identical to that of the initial study (identical markers in the identical stroke subtypes), none replicated the original finding (**Supplementary Fig. 1** online)[2,7,8]. In eight studies that examined SNP45 without microsatellite AC008818-1 (refs. 2,4–10), only one found a nominal association with combined large-vessel and cardiogenic stroke ($P = 0.01$)[2]. Many of the studies reported 'novel' associated alleles with a nominal *P* value ≤ 0.05. In total, 11 SNPs have been reported as associated, some for different subtypes of stroke not originally implicated by the original report (**Supplementary Table 1**).

To determine whether the results of these follow-up studies are consistent with replication, we undertook a systematic evaluation of the 11 SNPs and the microsatellite within *PDE4D* that have been reported as associated with stroke or subtypes of stroke[2–10]. If any of these associations represent evidence of a true association between a *PDE4D* variant and stroke, it is expected that these 'positively' associated alleles should be correlated to one another. Such correlation would reflect that these alleles act as proxies for one another—all in linkage disequilibrium with the true causal variant. To test this prediction, we used data from the International HapMap Project to estimate the correlation (r^2) among the 11 associated SNPs and the original risk haplotype (**Supplementary Table 2** and **Supplementary Methods** online)[1].

We find that none of these associated SNPs is significantly correlated to the G0 risk haplotype in HapMap CEU samples: the

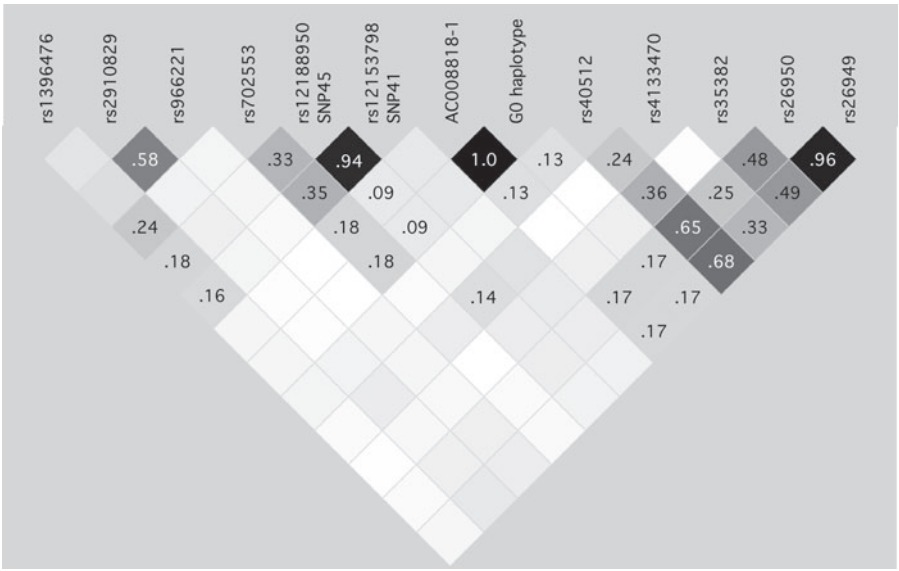

Figure 1 Pairwise r^2 plot for 11 associated SNPs and allele 0 of microsatellite AC008818-1 in the HapMap CEU panel. G0 denotes the risk haplotype formed by rs12188950 (SNP45) and AC008818-1 (ref. 1). Both the haplotype frequencies and the extent of linkage disequilibrium between the two markers found in HapMap CEU are similar to those found in Icelandic population data from the deCODE study (**Supplementary Table 2**). None of the associated SNPs in follow-up studies demonstrates a correlation of $r^2 > 0.18$ to the G0 haplotype. Blank squares indicate $r^2 < 0.15$. Figure was created using Haploview 3.32.

maximum r^2 of the 11 SNPs to G0 is ~0.2 (**Fig. 1**). As was observed in the data of the deCODE study[1], the 0 allele of AC008818-1 determines the G0 haplotype perfectly and is in complete linkage disequilibrium ($D' = 1$) with SNP45 (rs12188950). Further, the r^2 (<0.1) between SNP45 and the 0 allele of AC008818-1 as well as the frequencies of the two-marker haplotype in both the deCODE study and the present HapMap data are very similar (**Supplementary Table 2**). These results make it difficult to reconcile follow-up association findings in European-derived populations under the assumption that the original association[1] is a true positive. Consequently, many of the follow-up publications raise novel hypotheses (**Supplementary Table 3** online) that may themselves require replication in future studies, but none of them constitutes a replication of the original finding.

The search for genetic determinants of complex diseases has been notably successful in the past year. In age-related macular degeneration, studies of different populations have repeatedly identified the identical SNP in complement factor H as associated with disease. Similarly, in type 2 diabetes, recent publications have identified a single microsatellite in the transcription factor 7–like 2 gene as associated with disease in multiple population samples from different countries. In both of these examples, replication involved an identical analysis of data drawn from different populations, and the results of each analysis confirmed the identical hypothesis: a single allele is associated with disease.

The stroke phenotype presents numerous challenges to investigators in search of genes. Debate about the genetic architecture of ischemic stroke—whether it represents one or multiple phenotypes—continues, and the heterogeneity of phenotypes associated with *PDE4D* variants in replication studies has not brought clarification. The originally published association between *PDE4D* and stroke demonstrated an association only with large-vessel and cardiogenic stroke subtypes[1]. Follow-up studies, however, identified nominal associations with other stroke subtypes as well as with all subtypes combined.

The possibility remains that a single disease phenotype can arise from multiple variants in the same gene, as is often the case in monogenic or mendelian disorders. However, the challenge for the study of complex disorders such as stroke is that, given the large number of presumably non-causal SNPs in the genome, the prior probability that any one of them causes disease is very small. Beyond functional confirmation in a model system, where introduction of an altered gene produces a predicted change in phenotype, replication in additional samples is therefore essential. Indeed, we emphasize that reproducible associations can be obtained, as has been demonstrated in other complex diseases. Our results do not rule out a true association of any of the reported variants. Nonetheless, we strongly argue that the original *PDE4D* association be viewed with skepticism, as we have learned from past experience with multiple irreproducible associations.

Jonathan Rosand[1–3], Nick Bayley[2,3], Natalia Rost[1–3] & Paul I W de Bakker[2,3]

[1]Vascular and Critical Care Neurology and [2]Center for Human Genetic Research, Massachusetts General Hospital, Boston, Massachusetts 02114, USA. [3]Program in Medical and Population Genetics, Broad Institute of Harvard and MIT, Cambridge, Massachusetts 02142, USA.
e-mail: jrosand@partners.org

Note: Supplementary information is available on the Nature Genetics website.

1. Gretarsdottir, S. *et al. Nat. Genet.* **35**, 131–138 (2003).
2. Meschia, J.F. *et al. Ann. Neurol.* **58**, 351–361 (2005).
3. Saleheen, D. *et al. Stroke* **36**, 2275–2277 (2005).
4. Woo, D. *et al. Stroke* **37**, 371–376 (2006).
5. van Rijn, M.J.E. *et al. Neurology* **65**, 1203–1209 (2005).
6. Nakayama, T., Asai, S., Sato, N. & Soma, M. *Stroke* **37**, 69–76 (2006).
7. Lohmussaar, E. *et al. Stroke* **36**, 731–736 (2005).
8. Bevan, S., Porteous, L., Sitzer, M. & Markus, H.S. *Stroke* **36**, 949–953 (2005).
9. Nilsson-Ardnor, S. *et al. Stroke* **36**, 1666–1671 (2005).
10. Kuhlenbaumer, G. *et al. J. Neurol. Neurosurg. Psychiatry* **77**, 521–524 (2006).

Gulcher *et al.* reply:

We do not disagree with Rosand *et al.* that, in general, the published replication studies do not provide a clear and simple picture for the role of *PDE4D* in stroke, even though five out of the nine studies claim replication. We agree that the quality and rigor of many of the studies published so far as replication cohorts leave a little to be desired. Our initial report of association of *PDE4D* to stroke phenotypes was one of our first on genes contributing to the risk of common disease. We held ourselves to the highest standards of the time, but since then, we have adopted a policy to withhold publication of most of our discoveries until we have replication in sample sets powered for the phenotype of interest. That way we can have some control along with our collaborators over the quality and power of the replication cohorts used. We agree that claims of replication are premature, because none of the studies shows significant association to our original markers if fully corrected for the number of markers and phenotypes tested. In most cases, the effects of the associated variants are rather weak for ischemic stroke as a whole, but this is not surprising, because even in Iceland we have very modest effects for stroke in general. In six out of the nine studies, the largest stroke cohorts of single ethnicity range from only 88 to 279 patients and therefore are underpowered to properly test for the two subtypes of stroke in which we found stronger effect sizes—cardiogenic and large-vessel stroke[1–6]. This is especially true if the effects are somewhat smaller than originally estimated. Furthermore, there are clear differences in criteria used for subtyping ischemic stroke even among the groups using the TOAST research criteria. For example, our cutoff for large-vessel stroke was 70% stenosis, while most other groups use 50%.

Regardless, the strongest signal we reported was the G0 haplotype, and none of the studies has replicated that signal. We ourselves have conducted replication studies with two foreign cohorts, one from Sweden and one from Scotland, and we are now preparing the results for publication. In brief, the G0 haplotype also does not confer a significant increase in risk. However, the single allele G of SNP45 that is a part of the haplotype looks more promising in our hands and others. In Gretarsdottir *et al.*[7], haplotypes were partitioned into three groups. While haplotype G0 was found to confer an increase in risk, haplotypes that contained allele A of SNP45 showed reduced risk compared to the wild-type, the latter defined as haplotypes that have allele G for SNP45 but not allele 0 for the microsatellite marker AC008818-1. Hence, allele G of SNP45 is the complement of what we called a 'protective' variant. As we commented earlier[8], allele G of SNP45 also showed more promising results than the haplotype in the Bevan *et al.* study.

One major critique by Rosand *et al.* of some of the published replication studies seems to be that in certain cases, replication was claimed, but the associations were found to be with SNPs that are essentially uncorrelated with the original major finding. This is no doubt a legitimate criticism in many cases, but it is also