In Hans Christian Andersen’s beloved tale, it is the innocent child who finally reveals what others had been unable to admit. Deference to the King’s court and perceived wisdom discomforts other subjects in the kingdom from questioning the sovereign’s taste in garments. Since the original article identifying PDE4D as the putative stroke 1 (STRK1) locus, many groups from around the globe have attempted to validate the association. Stroke is a syndrome not a disease, with numerous interrelated phenotypes and subphenotypes. The challenges of sorting out the genetic contributions to complex diseases such as stroke are substantial, but the potential rewards in the forms of new treatments and improved understanding of pathophysiology are great. In the face of the great promise of the PDE4D tale, we need to assess the state of our knowledge clearly and honestly.

Before the report by deCODE Genetics, the PDE4 family of genes had not been tested as candidate genes for any disease or phenotype, although the biochemical role of PDE4D as a regulator of cAMP signal transduction was recognized. The PDE4D gene product (cAMP specific 3',5'-cyclic phosphodiesterase 4D) appears to be a secondary signal pathway regulator of phenotype, with indirect effects on cardiovascular or stroke biomarkers. Nonetheless, because PDE4 enzymes predominate cAMP metabolism in inflammatory cells, and PDE4D accounts for at least 80% of PDE activity in inflammatory cells, the PDE4D is a plausible candidate gene. PDE4D enzyme activity could play an important role in stroke risk through its effects on inflammation, plaque stability, response to injury, angiogenesis, and susceptibility to low grade infections. The deCODE investigators and others have argued that isoform expression may regulate PDE4D activity based on their relative expression profile. PDE4D clearly has a complex role in regulating biological activity of tissues in which it is expressed.

A fundamental flaw in the original deCODE article was the lack of independent replication. The deCODE team did survey PDE4D transcript expression in transformed B-cells, which appeared to corroborate the hypothesis that noncoding PDE4D variants are associated both...
with measurable PDE4D biological activity and ischemic stroke as a major outcome. The limited assay scope (only B-cell lines), the lack of clear relationship to genotype/haplotype, and the known complexity of the autoregulation of PDE4D via the cAMP feedback loop\(^9\) argues against drawing general conclusions.

Since this initial report, the stroke genetics community has properly attempted to replicate these findings in other cohorts and populations. There are 10 published follow-up studies by independent groups attempting to reproduce and inform the original findings.\(^8\) Association studies have tested >55 distinct single-nucleotide proteins (SNPs) in the PDE4D gene although the majority has only been reported in 1 or 2 articles. These studies tested in cohorts of differing ethnicity and ascertainment criteria with nonstandardized adjustments for modifiable and nonmodifiable stroke risk factors. To date, none of these studies have found the same association of the high-risk haplotype and ischemic stroke or any subphenotype.\(^10\)\(^-\)\(^15\) Several SNPs have shown positive associations at the 0.05 level in some studies, although the effect size has generally been weaker than in the original report.

Adding to the recently burgeoning list of replication studies, in this issue of Stroke Zee et al report the results of a nested case-control study of 259 white men with incident adjudicated stroke drawn from the Physician’s Health Study, matched to an equal number of participants free of stroke or vascular disease, at the time of the primary qualifying event in the cases.\(^16\) They included adjustments for main study randomization arm (aspirin or beta carotene), diabetes, and BMI, and tested 9 SNP markers with prior evidence of association in either the deCODE cohort or in the recently published nested case-control analysis of women in the Study of Osteoporotic Fractures (SOF).\(^2\)\(^,\)\(^15\) After adjusting for traditional stroke risk factors, SNP56 showed nominal significance with ischemic stroke risk (recessive: odds ratio=2.26, 95% CI=1.11 to 4.61; \(P=0.03\)). In a stratified analysis of nonhypertensive individuals, SNP45 (dominant: odds ratio=2.24, 95% CI=1.00 to 5.00; \(P=0.05\)) and SNP56 (additive: odds ratio=1.77, 95% CI=1.02 to 3.10; \(P=0.04\)) showed modest association with increased risk of ischemic stroke. As Zee et al acknowledge, their SNP45 findings of allele A association with ischemic stroke represent a reversal of risk allele from that reported in the original\(^2\) and some\(^12\)\(^,\)\(^17\) but not all follow-up\(^10\)\(^,\)\(^11\)\(^,\)\(^15\)\(^,\)\(^18\)\(^,\)\(^19\) studies. Interpretation of these results is made more difficult by the observation that 2 groups found the SNP45 locus to be monomorphic in a Japanese and a biracial US population.\(^13\)\(^,\)\(^14\)

The observation that variants in PDE4D manifest reversed direction of effects in differing studies suggests the possibility of spurious associations (Cardon LR, personal communication, 2006). Although hard to quantify, discrete variants that show a consistent effect direction across multiple independent studies\(^20\)\(^,\)\(^21\) are more likely to be ‘real’ as judged by meta-analysis. Race-specific interactions are sometimes advanced as a possible reason for inconsistency of results across studies, but experience suggests that true associated variants are robust to cross-race effects.\(^22\) In other words, race-dependent biological effects in common complex diseases arise from factors that affect the allelic frequencies in the population\(^22\) rather than through modulation or reversal of the effect itself.

The results of Zee et al corroborate an emerging view that the overall effects of genetic variation in PDE4D appear to be very modest at best relative to other risk factors in stroke populations defined by the broadest ischemic stroke inclusionary criteria. The difficulty in replicating the primary report of association of genetic variants with a complex disease has sadly been rehearsed many times over, and the possible statistical causes and epidemiological confounders recounted at some length. The initial report of any novel gene association must be treated with measured enthusiasm because the significance-based ranking of association test results across genes and markers inherently leads to bias in the estimated effect size, and concomitant inference of association.

\(\text{Stroke. Author manuscript; available in PMC 2006 October 19.}\)
One encouraging aspect of both the Zee study and a recently published analysis by the same authors of the SOF cohort is the suggestive increase in significance of test results for the nonhypertensive stratum, after stratifying the samples by history of hypertension, reflecting the relative importance of this factor in risk for ischemic stroke. There are 2 broad explanations for this gene-stratum interaction. Qualitatively, the effective removal of a major risk factor may help to reveal the influence of the underlying genetic variation with reduced confounding of the results (“swamping out” hypothesis). Alternatively, the physiological milieu of patients in this stratum may result in differential importance of PDE4D activity relative to nonstratum patients. In this sense, the analysis parallels approaches used in other complex diseases to disentangle genetic factors from modifiable behavioral or environmental effects.

Although disappointing overall, the Zee et al article offers some useful lessons and suggestions about how to design and perform these studies in the future. We recognize that investigators may have little control over the phenotypes and data that are available in existing stroke cohorts. However, we urge investigators that are planning new studies to include collection of data that will allow for satisfactory adjustment of nongenetic confounding and risk factors. Because common sporadic stroke cases occur in late adulthood, there will be a lifetime of accumulated background risk, and testing selective secondary hypotheses requires correspondingly larger total sample sizes. The stratified analysis of Zee et al was probably underpowered to achieve more compelling significance, given the reasonable estimated odds ratios range 1.6 to 2.2 (at least by complex disease standards) and allele frequencies for the associated SNPs. We remind the community of the following basic principles for conducting genetics studies:

1. Standardize epidemiological and statistical models to incorporate minimal set of recognized risk factors, both modifiable and not.
2. Perform power analyses for all studies, especially negative reports of association, preferably using shrunken estimates of effect size. At a minimum, studies should be powered to allow analysis of the subtype that is most relevant to the study population.
3. Continue to report negative results, particularly well-powered studies, to avoid publication bias.
4. Evaluate study design and results honestly in articles and discussions with reviewers.
5. Characterize study-sample phenotypes, and perform analyses based on rational mechanistic principles to minimize type 1 error, rather than exhaustively testing all possibilities.

Although replication of putative association results is necessary to establish the robustness and generality of a result, and to estimate effect sizes less tainted by biases, we should remember that PDE4D is only the first candidate of a new era in stroke genetics. From a public health perspective, variation in the PDE4D locus seems to contribute only modestly to overall stroke risk in the North American populations tested so far, but with national and international studies reaching maturity, and the availability of very large data sets from whole genome association studies within the next few years, the community will have other candidates and pathways to pursue. Pathways of inflammation, coagulation, cysteine metabolism, atherogenesis, and lipid metabolism may yet yield genes that contribute to a broad stroke outcome or stroke sub(phen) type. From the existing literature, there are already suggestions that the single extended STRK1 locus may contain multiple loci that contribute to discrete linkage peaks.

So after all of the admiring peeks and compliments, is the King really bare? As the first gene linked to ischemic stroke, the identification of PDE4D has been rightly heralded as a major milestone in the nascent field of stroke genetics. This gene and its associated molecular pathways may still prove to be important in cerebrovascular disease, but will require a...
concerted systematic and rigorous approach to collecting and analyzing data to disentangle the contributing factors.

Stroke is a youthful member of a widening family of diseases with a complex genetic epidemiology, a family which includes diabetes, obesity, cardiovascular disease, hypertension, asthma, schizophrenia, and autoimmune diseases. If the genetics research on more established complex diseases is any guide, there will be many new suits that are discarded as fashions change, hopefully interspersed with classic and more enduring attire. Realizing these successes may require more finely nuanced hypotheses based on the pathways involved and the physiological milieu in which the genetic variation is exercised. The court and the public await the tale to play out.

References


*Stroke*. Author manuscript; available in PMC 2006 October 19.