

REVIEW ARTICLE

Exploring the genetic basis of stroke. Spanish stroke genetics consortium[☆]

E. Giralt-Steinhauer^{a,*}, J. Jiménez-Conde^a, C. Soriano Tàrraga^a, M. Mola^a,
A. Rodríguez-Campello^a, E. Cuadrado-Godía^a, A. Ois^a, I. Fernández-Cádenas^b,
C. Carrera^b, J. Montaner^b, R.M. Díaz Navarro^c, C. Vives-Bauzá^c, J. Roquer^a

^a Grupo de Investigación Neurovascular, Departamento de Neurología, IMIM-Hospital del Mar, Universitat Autònoma de Barcelona, Barcelona, Spain

^b Laboratorio de Investigación Neurovascular, Institut de Recerca, Hospital Vall d'Hebron, Universitat Autònoma de Barcelona, Barcelona, Spain

^c Departamento de Neurología, Hospital Universitari Son Espases, Mallorca, Balears, Spain

Received 28 December 2012; accepted 21 April 2013

Available online 22 October 2014

KEYWORDS

Stroke;
Genetics;
Association;
Genetic risk factors;
Prediction;
Consortium

PALABRAS CLAVE

Ictus;
Genética;
Asociación;
Factores de riesgo
genético;
Predicción;
Consortio

Abstract This article provides an overview of stroke genetics studies ranging from the candidate gene approach to more recent studies by the genome wide association. It highlights the complexity of stroke owing to its different aetiopathogenic mechanisms, the difficulties in studying its genetic component, and the solutions provided to date. The study emphasises the importance of cooperation between the different centres, whether this takes place occasionally or through the creation of lasting consortiums. This strategy is currently essential to the completion of high-quality scientific studies that allow researchers to gain a better knowledge of the genetic component of stroke as it relates to aetiology, treatment, and prevention.

© 2012 Sociedad Española de Neurología. Published by Elsevier España, S.L.U. All rights reserved.

Aproximación al conocimiento de las bases genéticas del ictus. Consorcio español de genética del ictus

Resumen El presente artículo revisa la evolución de los estudios en genética del ictus desde la aproximación por gen candidato hasta los recientes estudios de *genome wide association*. Se destaca la complejidad de esta afección por sus muy variados mecanismos etiopatogénicos, las dificultades que comporta el estudio de su componente genético y las soluciones que se han aportado. Se subraya en especial el valor de las colaboraciones entre distintos centros, ya sea

[☆] Please cite this article as: Giralt-Steinhauer E, Jiménez-Conde J, Soriano Tàrraga C, Mola M, Rodríguez-Campello A, Cuadrado-Godía E, et al. Aproximación al conocimiento de las bases genéticas del ictus. Consorcio español de genética del ictus. Neurología. 2014;29:560–566.

* Corresponding author.

E-mail address: 96714@parcdesalutmar.cat (E. Giralt-Steinhauer).

de manera puntual o sobre todo a través de la creación de consorcios estables. Esta estrategia actualmente se hace imprescindible a la hora de realizar estudios de alta calidad científica que permitan seguir avanzando en el conocimiento de las bases genéticas del ictus tanto en etiología, como en tratamiento y prevención.

© 2012 Sociedad Española de Neurología. Publicado por Elsevier España, S.L.U. Todos los derechos reservados.

Introduction

In Spain, stroke is currently the second most frequent cause of death in the general population, as well as the leading cause of death in women (<http://www.ine.es/prensa/np703.pdf>). It is also the leading cause of disability in adults and the second most common cause of dementia.¹

However, despite important advances in stroke prevention, diagnosis, and treatment in the past 2 decades, the World Health Organisation estimates that the incidence of this entity will increase by 27% by the year 2025.² Researchers have described numerous risk factors associated with stroke, although in order to understand its causes it seems necessary to consider both external factors and the predisposition of each individual.³ The disease has been shown to have a hereditary component, and genetic load plays a very important role in its development.⁴ Technological advances in recent years have promoted genome-wide association studies (GWAS), which let researchers perform non-biased genetic analyses and discover new genes and metabolic pathways associated with stroke.

The purpose of this study is to review the current status and future perspectives of genetics in the field of cerebrovascular disease.

Evidence

Genetic factors in stroke

Unlike monogenic disorders, complex diseases are characterised by their considerable aetiological and pathological diversity. Family history is recognised as one of the most important risk factors for many complex disorders; this is also the case for cardiovascular disease, cancer, or autoimmune disorders.⁵ In conclusion, even though these disorders are not monogenic or oligogenic, inherited genetic variations are apparent. Scientists believe that co-presence of different low-risk allelic variants whose combined risk is additive or multiplicative, and interaction of those variants with the environment, may play a crucial role in the development of complex disorders.^{5,6}

Heritability rates between 30% and 50% have been described for ischaemic stroke.⁴ Because of its multiple causes, the numerous risk factors involved, and its array of different forms of presentation, stroke may be considered the final manifestation of any of a number of complex diseases. Apart from the simplest stroke classification scheme that distinguishes between ischaemic and haemorrhagic

events, we can also identify several distinct aetiopathogenic subtypes. Several recently published studies have shown differences between the genetic heritability of lacunar, cardioembolic, and atherosclerotic strokes; the latter are the most strongly associated with family history.^{4,7,8}

This also occurs in intracerebral haemorrhages (ICH), which account for 15% of all strokes. Pathogenic pathways involved in lobar haemorrhage, which is closely linked to cerebral amyloid angiopathy (CAA), may differ substantially from those involved in hypertensive haemorrhages.^{9,10}

Therefore, genetic studies having to do with stroke should be designed so as to examine these particularities and accommodate the difficulties inherent to investigating heterogeneous diseases.

Genetic association studies in complex disorders: techniques and approaches

Research on most genetic causes of stroke, excluding those that follow classic Mendelian inheritance patterns, are generally based on case-control genetic association studies. These studies analyse the different frequencies of genetic variants in patients and healthy controls. Studying common variants or single nucleotide polymorphisms (SNPs) largely involves candidate-gene and genome-wide association, whereas the recently developed technique of complete exome sequencing lets scientists examine uncommon variants. Other techniques, such as analysis of copy-number variation or epigenome studies, will also be discussed below.

Candidate-gene studies

For several decades, and until recently, the most common technique was the candidate-gene study. This approach consists of selecting genes that may be associated with a disease based on the pathophysiology of that disease and the pathways believed to be involved. Scientists identify previously described SNPs located near the candidate genes and select the most representative ones (tag SNPs) which will typically display linkage disequilibrium.¹¹ These candidate tag SNPs are genotyped within a population and statistical analysis is used to determine if there are associations between genetic variants and the clinical spectrum.

These studies present several problems. On the one hand, the fact that these genes are selected based on a priori hypotheses may mean that some researchers look for spurious differences based on preconceived notions. Meanwhile, they may fail to analyse genes that are potentially involved in the disease.

Multiple genes have been linked to increased risk of ischaemic stroke by means of candidate gene studies. However, a recent study showed that none of these potential associations were statistically significant once a statistical correction had been applied to adjust for the number of SNPs analysed.⁴ Furthermore, no results obtained using this technique have ever been clearly replicated, which shows that the approach is inconsistent and unreliable where ischaemic stroke is concerned.^{4,12}

Even so, this technique has shown that mutations in the *COL4A1* and *COL4A2* genes are associated with ICH. Heterotrimers formed by collagen IV alpha 1 (*COL4A1*) and alpha 2 (*COL4A2*) are one of the main components of nearly all basement membranes of the body, and they have been associated with numerous diseases that may be ocular, renal, or muscular as well as cerebrovascular.¹³ Initially, doctors using familial inheritance studies found that mutations in *COL4A1* caused porencephaly type I, and this condition was later shown to cause both sporadic and recurrent ICH, as well as lacunar stroke and leukoaraiosis.^{14–16}

A recent study, also related to ICH, showed that patients bearing haplotypes $\epsilon 2$ and $\epsilon 4$ of the *APOE* gene are at higher risk for lobar ICH (odds ratio [OR] 1.82 [95% CI, 1.50–2.23]; $P = 6.6 \times 10^{-10}$ for $\epsilon 2$ and OR 2.20 [95% CI, 1.85–2.63]; $P = 2.4 \times 10^{-11}$, for $\epsilon 4$). This is probably due to the effect of these variants on the risk of developing CAA.¹⁷ The same group later demonstrated that the *APOE*\epsilon 2* variant was associated with a greater volume of ICH, and consequently with poorer functional prognosis and higher mortality rate.¹⁸

GWAS

GWAS, which let researchers measure the allele frequency of hundreds of thousands of SNPs in cases versus controls, were developed for 2 main reasons. First came the creation of the International HapMap Project which determined the characteristics, linkage equilibrium, and allele frequencies of SNPs found in the genome of 270 individuals representing 4 different ancestral origins. Its purpose was to create a public genome-wide database that could be freely accessed by all researchers.^{5,19} Secondly, significant technological advances and platforms were developed that permitted genotyping 500 000 to 2.7 million SNPs per individual at consistently lower costs. GWAS studies constitute a much more exhaustive means of studying the genome than candidate studies.²⁰ They are not affected by preconceived hypothesis biases and may also let scientists identify new associations even when the gene has not yet been linked to the disease, unlike earlier approaches. Nevertheless, GWAS has other types of limitations. Its main 2 limitations are the large sample sizes that are needed and the difficulty of interpreting potential positive results.²¹ These studies also require strict quality control processes to detect genotyping errors and assess possible stratification in populations.²² The latter limitation arises because the allele frequencies of an SNP vary according to each individual's ethnic makeup, even between individuals sharing an ethnic phenotype. Since the above situation may give rise to false results if the case and control populations do not have comparable genetic origins, the pertinent adjustments have to be made. In fact, the best way of eliminating these inconsistencies is by replicating results obtained from very large independent samples.²³

Considering the large numbers of analyses that must be performed (at least one per genotyped SNP), statistical penalties should be applied for multiple comparisons such that the significance level (P -value) may be set at 10^{-8} (in cases of analysing one million SNPs).⁶ If studies are to reach the necessary statistical power, their sample sizes must be increased by a significant amount.

The first results having to do with complex diseases were published in 2005, when complement factor H was found to be involved in age-related macular degeneration.²⁴ Interesting discoveries have since been made in the areas of type 1 diabetes mellitus (*PPARG*²⁵, *KCNJ11*,²⁶ and *TCF7L2*²⁷); rheumatoid arthritis and type 2 diabetes mellitus (*PTPN22*^{28,29}) and Crohn disease (*NOD2*)³⁰ and many others. Large numbers of loci have been found to be implicated in the predisposition or risk of developing the above diseases.

In contrast with the discouraging results that have been delivered by other approaches, several associations with ischaemic stroke have been replicated consistently. Such is the case of associations discovered through GWAS for other diseases associated with ischaemic stroke and which have also been determined to be stroke risk factors.³¹ As such, 2 variants, rs1906591 (OR 1.48 [95% CI, 1.28–1.71]; $P = 4.7 \times 10^{-6}$) and rs7193343 (OR 1.22 [95% CI, 1.10–1.35]; $P = .00021$), in chromosomes 4q25 and 16q22 in regions near genes *PITX2* and *ZFX3* respectively, were initially identified as contributing to the risk of developing atrial fibrillation. Since then, they have also been associated with higher risk of presenting a cardioembolic stroke.^{32,33}

A region in chromosome 9p21 near the *CDKN2A/B* genes has been linked to myocardial infarction and coronary artery disease. A later candidate gene study of these loci using a candidate gene found an association between the SNP rs1537378 (OR 1.21; [95% CI, 1.07–1.37]; $P = .002$) and atherothrombotic stroke.³⁴ GWAS has also identified a new association with atherothrombotic stroke: polymorphism rs11984041 (OR 1.42 [95% CI, 1.28–1.57]; $P = 1.87 \times 10^{-11}$). This variant is located in the region of gene *HDAC9* on chromosome 7p21.1, but its pathogenic mechanism is unknown (Fig. 1).³⁵

Regarding new associations, a Japanese team has identified a significant link between a polymorphism of the *PRKCH* gene, rs2230500 on chromosome 14 (OR 1.66 [95% CI, 1.33–2.09]; $P = 9.84 \times 10^{-6}$), with lacunar stroke. Replication results have been dissimilar in other ethnic groups.^{4,36}

Exome sequencing

One of the weak points of GWAS is that it only analyses common variants (population frequency > 1%) without considering rare variants. There may in fact be several rare variants that contribute to genetic risk of complex diseases. In order to identify them, scientists are developing new exome sequencing techniques. These new techniques identify all polymorphisms or mutations found in protein-coding regions of the genome. As such, they are able to detect the rare variants that cannot be identified by GWAS. While this approach is promising, its methodology is still being developed and doctors have yet to agree on the best way to manage the huge amounts of data generated by these analyses.³⁷

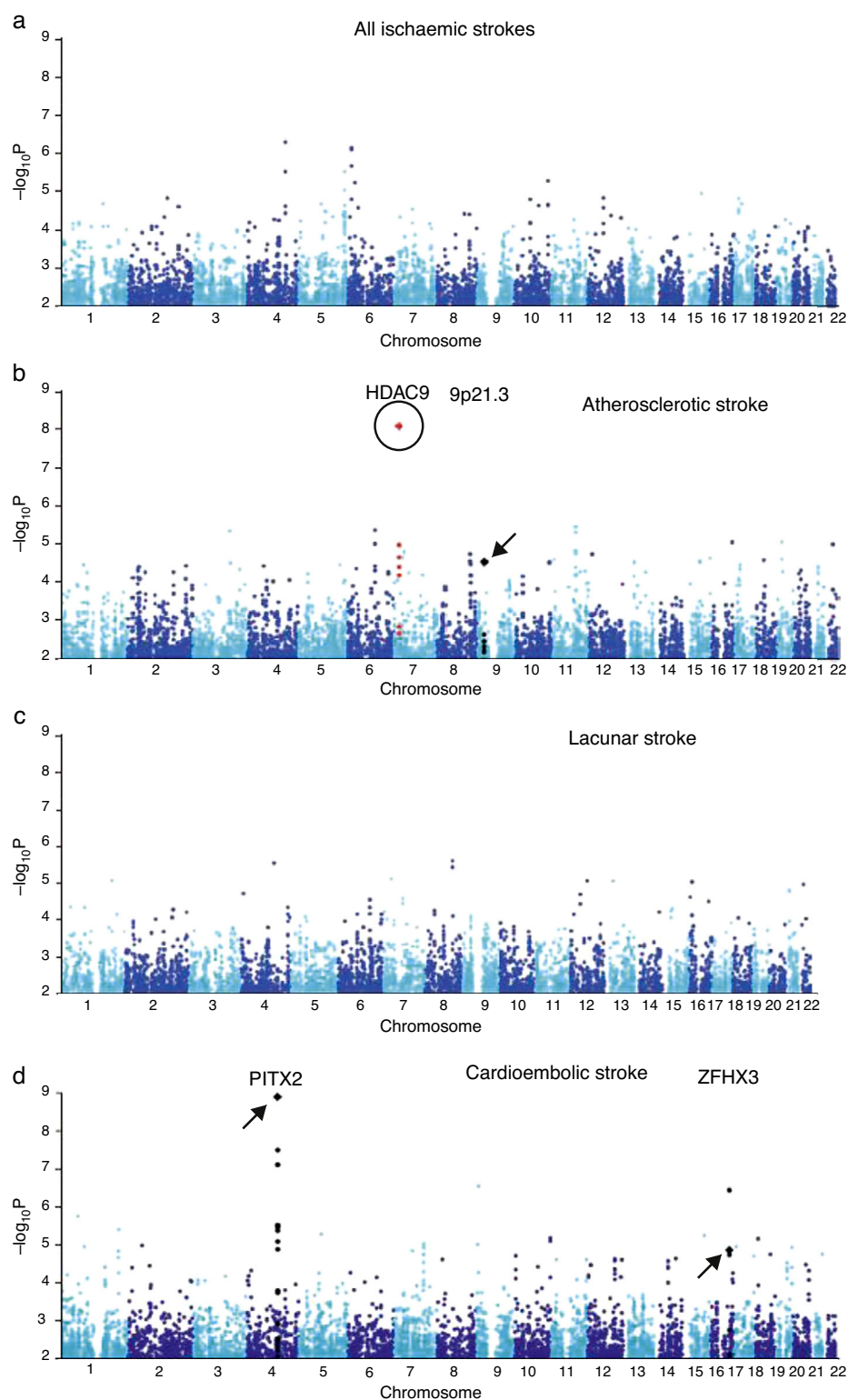


Figure 1 GWAS results from samples from the UK and Germany. (a) Results from all ischaemic strokes. (b) Atherosclerotic stroke. (c) Lacunar stroke. (d) Cardioembolic stroke. The circle indicates the new locus in *HDAC9*. Loci described in earlier literature are indicated with arrows.

Published with permission from Macmillan Publishers Ltd: *Nature Genetics*, 2012.³⁵

Adapted with permission from Macmillan Publishers Ltd: *Nature Genetics*, 2012.³⁵

Table 1 Main articles completed within the framework of the International Stroke Genetics Consortium and published in major journals

Authors, year of publication, journal	Stroke type and subtype	No. cases/No. controls	Affected chromosome	Conclusion
Gschwendter A. <i>Annals of Neurology</i> , 2009. ³⁴	Ischaemic stroke TOAST: atherosclerotic	1090/1244, southern Germany 758/872, United Kingdom Replica study: 2528/2189	Chromosome 9p region 21.3: rs1537378	This locus, also implicated in IHD, is related to atherothrombotic stroke
Biffi A. <i>Annals of Neurology</i> , 2010. ¹⁷	Hypertensive and lobar haemorrhages	Lobar: 931/3744 Deep haemorrhage: 1085/3657	2 variants in <i>APOE</i> chromosome 19: rs7412 and rs429358	Strong association between <i>APOE</i> variants and lobar ICH, especially ICH linked to CAA
Lemmens R. <i>Stroke</i> , 2010. ³²	Ischaemic stroke TOAST: cardioembolic	4199/3750	Variant of chromosome 4q25: rs1906591	Replica study supporting link between variant rs1906591 of chromosome 4q25 and FA and CE stroke. Not associated with other stroke subtypes
Biffi A. <i>Lancet Neurology</i> , 2011. ¹⁸	Lobar ICH	Cases: 865 subjects. Replica study 1: 946; replica study 2: 214	2 variants in <i>APOE</i> chromosome 19: rs7412 and rs429358	The presence of <i>APOE</i> * ϵ 2 alleles is associated with a higher ICH volume. This is probably due to the role played by ϵ 2 in the severity of vasculopathy due to CAA.
ISGC & WTCCC2. <i>Nature Genetics</i> , 2012. ³⁵	Ischaemic stroke TOAST: atherosclerotic	3548/5972 Replica study: 5859/6281	Chromosome 7p21.1 (rs11984041). Gene <i>HDAC9</i>	Replicates previously signalled locus 9p21 Identification of new association in chromosome 7p21.1; unknown mechanism

CAA: cerebral amyloid angiopathy; CE: cardioembolic; IHD: ischaemic heart disease; AF: atrial fibrillation; ICH: intracerebral haemorrhage; TOAST: Trial of Org 10172 in Acute Stroke Treatment.

Analysis of copy-number variation

Copy-number variation (CNV) is a structural variation of the genome that may occur as (micro) deletions, duplications, or inversions, creating an anomalous number of copies of a DNA fragment. This variation may affect both the expression and the makeup of proteins that are coded in this region.^{38–40} At present, very little has been published regarding stroke. While one study of patients with ischaemic stroke did not detect associated CNV, the sample size was small and the different stroke subtypes were not taken into account.⁴¹ However, there are genome-wide studies currently underway that include analysis of multiple CNVs in their analysis platforms.

Epigenetics

Another emerging field is epigenetics, a discipline that studies heritable changes in gene regulation that do not affect the DNA sequence. Although all the cells in our body display identical genomes, genetic transcription in each tissue type is regulated by different epigenetic mechanisms. Chief among these mechanisms are DNA methylation, and modifications to histones and non-coding RNA.⁴² Of these mechanisms, DNA methylation is the most studied and best understood to date. The most important findings were obtained from epigenetic studies in tumour tissue. For example, methylation of the *MGMT* gene inactivates enzyme

O (6)-methylguanine-DNA methyltransferase in cases of glioblastoma multiforme, which is associated with a better response to chemotherapy with temozolamide. This finding is already being applied in clinical practice.⁴³ Epigenetic studies are currently being expanded to contemplate other diseases, including stroke. One describes an association between stroke and overall DNA hypomethylation, and that study is currently being replicated.⁴⁴ Within the framework of the Spanish Consortium on Stroke Genetics, several studies in this field are currently underway and results are not yet available.

Risk scales and genetic networks

Designing genetic scores and genetic networks is becoming increasingly relevant for purposes of integrating the information and predictive ability of each SNP associated with a phenotype. The most commonly used methods for creating genetic scores are summing risk alleles (1 point per allele) or calculating weighted risk (the score for each allele depends on the OR reported for that allele). A very recent study showed an association between the score calculated based on risk alleles for arterial hypertension and the risk of experiencing deep ICH.⁴⁵

One way of organising knowledge is by creating networks or algorithms.⁴⁶ One example is the Bayesian network, a tool in statistics that represents an array of associated

variables set out according to the conditional dependencies that have been established between them. These models let researchers estimate the probability of the appearance of latent variables based on known probabilities. The main advantages of Bayesian networks are that they permit bidirectional interference (that is, effect-to-cause and cause-to-effect) as well as rapid updating of knowledge.⁴⁷

Genomic convergence

The concept of genomic convergence is becoming increasingly relevant to genetic studies of complex diseases. It is based on combining and unifying the information and data provided by each different type of study to achieve a greater capacity for discovery and more robust results.⁴⁸ Several polymorphisms of gene *TTC7B* were recently shown to confer susceptibility to ischaemic stroke.⁴⁹

The creation of consortia: GeneStroke, the Spanish consortium on stroke genetics

Following the creation of large international consortia, such as the UK's Wellcome Trust Case Control Consortium, the International Stroke Genetics Consortium (ISGC) was formed in 2007 (<http://www.strokegenetics.org>). The ISGC's activities have included identifying and fostering numerous associations (Table 1) and many more ambitious international projects in this field are currently underway.

GeneStroke, the Spanish Stroke Genetics Consortium (<http://www.genestroke.com>), was formed in 2009 for the main purpose of studying the genetic basis of stroke and promoting interaction between groups with common interests in this field. GeneStroke has launched a number of multi-centre cooperative projects related to stroke aetiology, progression, and prognosis. Further information about these projects and participants is available on the website (www.genestroke.com).

The Spanish Stroke Genetics Consortium promotes a model of large collaborative studies that are able to meet the challenges faced by research in the area of cerebrovascular disease genetics.

Conclusions

Stroke is one of the most prevalent diseases in our setting and it has a major impact in terms of morbidity and mortality. Its heritability is high, and a better knowledge of its genetic basis may contribute to identifying at-risk individuals and designing more efficient treatments and prevention strategies. The study of the genetic substrate should be a priority area, along with cooperation between different groups and professionals striving to complete high-quality scientific projects that may benefit society.

Conflicts of interest

The authors have no conflicts of interest to declare.

Acknowledgements

We would first like to thank all permanent members of the Spanish Stroke Genetics Consortium (GeneStroke) and any other members who have made sporadic contributions to these projects. We would also like to express our particular gratitude to the members of the International Stroke Genetics Consortium.

References

1. Alvarez Sabin J. In-hospital mortality in stroke patients. *Rev Esp Cardiol.* 2008;61:1007–9.
2. Truelsen T, Piechowski-Jozwiak B, Bonita R, Mathers C, Bogouslavsky J, Boysen G. Stroke incidence and prevalence in Europe: a review of available data. *Eur J Neurol.* 2006;13:581–98.
3. Grau AJ, Weimar C, Buggle F, Heinrich A, Goertler M, Neumaier S, et al. Risk factors, outcome, and treatment in subtypes of ischemic stroke: the German stroke data bank. *Stroke.* 2001;32:2559–66.
4. Bevan S, Traylor M, Adib-Samii P, Malik R, Paul NL, Jackson C, et al. Genetic heritability of ischemic stroke and the contribution of previously reported candidate gene and genome-wide associations. *Stroke.* 2012;43:3161–7.
5. International HapMap Consortium. A haplotype map of the human genome. *Nature.* 2005;437:1299–320.
6. Risch N, Merikangas K. The future of genetic studies of complex human diseases. *Science.* 1996;273:1516–7.
7. Jerrard-Dunne P, Cloud G, Hassan A, Markus HS. Evaluating the genetic component of ischemic stroke subtypes: a family history study. *Stroke.* 2003;34:1364–9.
8. Traylor M, Farrall M, Holliday EG, Sudlow C, Hopewell JC, Cheng YC, et al. Genetic risk factors for ischaemic stroke and its subtypes (the METASTROKE collaboration): a meta-analysis of genome-wide association studies. *Lancet Neurol.* 2012;11:951–62.
9. Caplan LR. Intracerebral haemorrhage. *Lancet.* 1992;339:656–8.
10. Flaherty ML, Woo D, Broderick J. The incidence of deep and lobar intracerebral hemorrhage in whites, blacks, and hispanics. *Neurology.* 2006;66:956–7, author reply 956–7.
11. Tabor HK, Risch NJ, Myers RM. Candidate-gene approaches for studying complex genetic traits: practical considerations. *Nat Rev Genet.* 2002;3:391–7.
12. Dichgans M, Markus HS. Genetic association studies in stroke: methodological issues and proposed standard criteria. *Stroke.* 2005;36:2027–31.
13. Kuo DS, Labelle-Dumais C, Gould DB. COL4A1 and COL4A2 mutations and disease: insights into pathogenic mechanisms and potential therapeutic targets. *Hum Mol Genet.* 2012;21(R1):R97–110.
14. Breedveld G, de Coe IF, Lequin MH, Arts WF, Heutink P, Gould DB, et al. Novel mutations in three families confirm a major role of COL4A1 in hereditary porencephaly. *J Med Genet.* 2006;43:490–5.
15. Gould DB, Phalan FC, van Mil SE, Sundberg JP, Vahedi K, Massin P, et al. Role of COL4A1 in small-vessel disease and hemorrhagic stroke. *N Engl J Med.* 2006;354:1489–96.
16. Weng YC, Sonni A, Labelle-Dumais C, de Leau M, Kauffman WB, Jeanne M, et al. COL4A1 mutations in patients with sporadic late-onset intracerebral hemorrhage. *Ann Neurol.* 2012;71:470–7.
17. Biffi A, Sonni A, Anderson CD, Kissela B, Jagiella JM, Schmidt H, et al. Variants at APOE influence risk of deep and lobar intracerebral hemorrhage. *Ann Neurol.* 2010;68:934–43.

18. Biffi A, Anderson CD, Jagiella JM, Schmidt H, Kissela B, Hansen BM, et al. APOE genotype and extent of bleeding and outcome in lobar intracerebral haemorrhage: a genetic association study. *Lancet Neurol.* 2011;10:702–9.
19. Frazer KA, Ballinger DG, Cox DR, Hinds DA, Stuve LL, Gibbs RA, et al., International HapMap Consortium. A second generation human haplotype map of over 3.1 million SNPs. *Nature.* 2007;449:851–61.
20. Visscher PM, Hill WG, Wray NR. Heritability in the genomics era – concepts and misconceptions. *Nat Rev Genet.* 2008;9:255–66.
21. McCarthy MI, Abecasis GR, Cardon LR, Goldstein DB, Little J, Ioannidis JP, et al. Genome-wide association studies for complex traits: consensus, uncertainty and challenges. *Nat Rev Genet.* 2008;9:356–69.
22. Marchini J, Cardon LR, Phillips MS, Donnelly P. The effects of human population structure on large genetic association studies. *Nat Genet.* 2004;36:512–7.
23. Neale BM, Purcell S. The positives, protocols, and perils of genome-wide association. *Am J Med Genet B Neuropsychiatr Genet.* 2008;147B:1288–94.
24. Klein RJ, Zeiss C, Chew EY, Tsai JY, Sackler RS, Haynes C, et al. Complement factor H polymorphism in age-related macular degeneration. *Science.* 2005;308:385–9.
25. Altshuler D, Hirschhorn JN, Klannemark M, Lindgren CM, Vohl MC, Nemesh J, et al. The common PPARgamma Pro12Ala polymorphism is associated with decreased risk of type 2 diabetes. *Nat Genet.* 2000;26:76–80.
26. Florez JC, Hirschhorn J, Altshuler D. The inherited basis of diabetes mellitus: implications for the genetic analysis of complex traits. *Annu Rev Genomics Hum Genet.* 2003;4:257–91.
27. Zeggini E, Weedon MN, Lindgren CM, Frayling TM, Elliott KS, Lango H, et al. Replication of genome-wide association signals in UK samples reveals risk loci for type 2 diabetes. *Science.* 2007;316:1336–41.
28. Begovich AB, Carlton VE, Honigberg LA, Schrodi SJ, Chokkalingam AP, Alexander HC, et al. A missense single-nucleotide polymorphism in a gene encoding a protein tyrosine phosphatase (PTPN22) is associated with rheumatoid arthritis. *Am J Hum Genet.* 2004;75:330–7.
29. Bottini N, Musumeci L, Alonso A, Rahmouni S, Nika K, Rostamkhani M, et al. A functional variant of lymphoid tyrosine phosphatase is associated with type I diabetes. *Nat Genet.* 2004;36:337–8.
30. Ogura Y, Bonen DK, Inohara N, Nicolae DL, Chen FF, Ramos R, et al. A frameshift mutation in NOD2 associated with susceptibility to Crohn's disease. *Nature.* 2001;411:603–6.
31. Markus HS. Stroke genetics. *Hum Mol Genet.* 2011;20(R2):R124–31.
32. Lemmens R, Buysschaert I, Geelen V, Fernandez-Cadenas I, Montaner J, Schmidt H, et al. The association of the 4q25 susceptibility variant for atrial fibrillation with stroke is limited to stroke of cardioembolic etiology. *Stroke.* 2010;41:1850–7.
33. Gudbjartsson DF, Holm H, Gretarsdottir S, Thorleifsson G, Walters GB, Thorgeirsson G, et al. A sequence variant in ZFX3 on 16q22 associates with atrial fibrillation and ischemic stroke. *Nat Genet.* 2009;41:876–8.
34. Gschwendtner A, Bevan S, Cole JW, Plourde A, Matarin M, Ross-Adams H, et al. Sequence variants on chromosome 9p21.3 confer risk for atherosclerotic stroke. *Ann Neurol.* 2009;65:531–9.
35. Bellenguez C, Bevan S, Gschwendtner A, Spencer CC, Burgess AI, Pirinen M, et al., International Stroke Genetics Consortium (ISGC), Wellcome Trust Case Control Consortium 2 (WTCCC2). Genome-wide association study identifies a variant in HDAC9 associated with large vessel ischemic stroke. *Nat Genet.* 2012;44:328–33.
36. Kubo M, Hata J, Ninomiya T, Matsuda K, Yonemoto K, Nakano T, et al. A nonsynonymous SNP in PRKCH (protein kinase C eta) increases the risk of cerebral infarction. *Nat Genet.* 2007;39:212–7.
37. Cole JW, Stine OC, Liu X, Pratap A, Cheng Y, Tallon LJ, et al. Rare variants in ischemic stroke: an exome pilot study. *PLoS One.* 2012;7:e35591.
38. Iafrate AJ, Feuk L, Rivera MN, Listewnik ML, Donahoe PK, Qi Y, et al. Detection of large-scale variation in the human genome. *Nat Genet.* 2004;36:949–51.
39. Sebat J, Lakshmi B, Troge J, Alexander J, Young J, Lundin P, et al. Large-scale copy number polymorphism in the human genome. *Science.* 2004;305:525–8.
40. Pollex RL, Hegele RA. Copy number variation in the human genome and its implications for cardiovascular disease. *Circulation.* 2007;115:3130–8.
41. Matarin M, Simon-Sanchez J, Fung HC, Scholz S, Gibbs JR, Hernandez DG, et al. Structural genomic variation in ischemic stroke. *Neurogenetics.* 2008;9:101–8.
42. Qureshi IA, Mehler MF. Emerging role of epigenetics in stroke: Part 1: DNA methylation and chromatin modifications. *Arch Neurol.* 2010;67:1316–22.
43. Skiriute D, Vaitkiene P, Saferis V, Asmoniene V, Skauminas K, Deltuva VP, et al. MGMT, GATA6, CD81, DR4, and CASP8 gene promoter methylation in glioblastoma. *BMC Cancer.* 2012;12:218.
44. Baccarelli A, Wright R, Bollati V, Litonjua A, Zanobetti A, Tarantini L, et al. Ischemic heart disease and stroke in relation to blood DNA methylation. *Epidemiology.* 2010;21:819–28.
45. Falcone GJ, Biffi A, Devan WJ, Jagiella JM, Schmidt H, Kissela B, et al. Burden of risk alleles for hypertension increases risk of intracerebral hemorrhage. *Stroke.* 2012.
46. Park YK, Bang OS, Cha MH, Kim J, Cole JW, Lee D, et al. SigCS base: an integrated genetic information resource for human cerebral stroke. *BMC Syst Biol.* 2011;14 5 (Suppl. 2). S10.0509-5-S2-S10. Epub 2011 Dec 14.
47. Ramoni RB, Himes BE, Sale MM, Furie KL, Ramoni MF. Predictive genomics of cardioembolic stroke. *Stroke.* 2009;40 3 Suppl.:S67–70.
48. Bell CG, Finer S, Lindgren CM, Wilson GA, Rakyan VK, Teschendorff AE, et al. Integrated genetic and epigenetic analysis identifies haplotype-specific methylation in the FTO type 2 diabetes and obesity susceptibility locus. *PLoS One.* 2010;5:e14040.
49. Krug T, Gabriel JP, Taipa R, Fonseca BV, Domingues-Montanari S, Fernandez-Cadenas I, et al. TTC7B emerges as a novel risk factor for ischemic stroke through the convergence of several genome-wide approaches. *J Cereb Blood Flow Metab.* 2012;32:1061–72.