Published in final edited form as:
Lancet Neurol. 2007 May ; 6(5): 414–420. doi:10.1016/S1474-4422(07)70081-9.

A genome-wide genotyping study in patients with ischaemic stroke:

initial analysis and data release

Mar Matarín, PhD[*], **W Mark Brown, MA**[*], **Sonja Scholz, MD**[*], **Javier Simón-Sánchez, BS**[*], **Hon-Chung Fung, MD**[*], **Dena Hernandez, MS**, **J Raphael Gibbs, BS**, **Fabienne Wavrant De Vrieze, PhD**, **Cynthia Crews**, **Angela Britton, MS**, **Carl D Langefeld, PhD**, **Thomas G Brott, MD**, **Robert D Brown Jr, MD**, **Bradford B Worrall, MD**, **Michael Frankel, MD**, **Scott Silliman, MD**, **L Douglas Case, PhD**, **Andrew Singleton, PhD**, **John A Hardy, PhD**, **Stephen S Rich, PhD**, and **James F Meschia, MD**

Molecular Genetics Unit

Division of Public Health Sciences, Wake Forest University School of Medicine, Winston-Salem, NC, USA

Department of Molecular Neuroscience, Institute of Neurology, Queen Square, London, UK

Molecular Genetics Unit, National Institute on Aging, National Institutes of Health, Bethesda, MD, USA; Unitat de Genética Molecular, Departamento de Genómica y Proteómica, Instituto de Biomedicina de Valencia-CSIC, Valencia, Spain

Laboratory of Neurogenetics, Department of Neurology, Chang Gung Memorial Hospital and College of Medicine, Chang Gung University, Taipei, Taiwan, Reta Lila Weston Institute of Neurological Studies, University College London, London, UK

Molecular Genetics Unit

Computational Biology Core, Department of Molecular Neuroscience, Institute of Neurology, Queen Square, London, UK

Correspondence to: Dr Andrew Singleton, Molecular Genetics Unit, National Institute on Aging, National Institutes of Health, Bethesda, MD 20892, USA singleta@mail.nih.gov.

[*]These authors contributed equally

Contributors MM, SS, JS-S, H-CF,DH, FWdV and AB did the genotyping. Data analysis was done by WMB, JRG and CDL, LDC, and AS. CC, TGB, RDB, BBW, MF, SS, and JFM collected and characterised samples. They decided also the study design, together with AS, JAH, and SSR. AS, JAH, SSR, JFM, MM, WMB, and JRG wrote the manuscript and helped in its critical revision.

Conflicts of interest We have no conflicts of interest.

ISGS Investigative Team

Executive Committee James F Meschia (Principal investigator), chair; Thomas G Brott, Robert D Brown Jr, Michael Frankel, John Hardy, Stephen S Rich, Scott Silliman, Bradford B Worrall. *Data Management Centre* Wake Forest University Medical Center: L Douglas Case (director), Laurie Russell, Carolyn Bell, Darrin Harris, Wes Roberson.*Clinical Coordinating Centre* Mayo Clinic, Jacksonville, FL, USA: James F Meschia, Alexa Richie, Dale Gamble, Sothear Luke. *DNA Repository* Coriell Institute for Medical Research, Camden, NJ, USA: Roderick A Corriveau. *Genetics Laboratory* National Institute on Aging, Bethesda MD, USA: John Hardy, Andrew Singleton. *Statistical Genetics* Wake Forest University Medical Center, Winston-Salem NC, USA: Stephen S Rich, W Mark Brown, Carl D Langefeld.

Sites and Investigators as of Sept 26, 2006

Mayo Clinic, Jacksonville, FL, USA (271 patients)—Principal investigator: James F Meschia. Coordinators: Alexa Richie, Dale Gamble, Sothear Luke. Subinvestigators: Thomas G Brott , Benjamin H Eidelman. University of Florida/Shands Hospital, Jacksonville, Florida (216 patients)—Principal investigator: Scott Silliman. Coordinators: Yvonne Douglas, Raam Sambandam. Subinvestigators: Nader Antonios. Emory University School of Medicine, Atlanta, Georgia (237 patients)—Principal investigator: Michael Frankel. Coordinator: Sharion Sailor-Smith. Mayo Clinic, Rochester, Minnesota (266 patients)—Principal investigator: Robert D Brown Jr. Coordinator: Colleen S Albers. University of Virginia, Charlottesville, Virginia (231 patients)—Principal investigator: Bradford Worrall. Coordinator: Daniel Chernavvsky.

Laboratory of Neurogenetics

Laboratory of Neurogenetics

Molecular Genetics Unit

Division of Public Health Sciences, Wake Forest University School of Medicine, Winston-Salem, NC, USA

Deparment of Neurology, Mayo Clinic, Jacksonville, FL, USA

Department of Neurology, Mayo Clinic, Rochester, MN, USA

Departments of Neurology and Public Health Sciences, University of Virginia, VA, USA

Department of Neurology, Emory University School of Medicine, GA, USA

Department of Neurology, University of Florida College of Medicine, Jacksonville, FL, USA

Division of Public Health Sciences, Wake Forest University School of Medicine, Winston-Salem, NC, USA

Molecular Genetics Unit

Laboratory of Neurogenetics, Department of Molecular Neuroscience, Institute of Neurology, Queen Square, London, UK

Division of Public Health Sciences, Wake Forest University School of Medicine, Winston-Salem, NC, USA

Deparment of Neurology, Mayo Clinic, Jacksonville, FL, USA

Summary

Background—Despite evidence of a genetic role in stroke, the identification of common genetic risk factors for this devastating disorder remains problematic. We aimed to identify any common genetic variability exerting a moderate to large effect on risk of ischaemic stroke, and to generate publicly available genome-wide genotype data to facilitate others doing the same.

Methods—We applied a genome-wide high-density single-nucleotide-polymorphism (SNP) genotyping approach to a cohort of samples with and without ischaemic stroke (n=278 and 275, respectively), and did an association analysis adjusted for known confounders in a final cohort of 249 cases and 268 controls. More than 400 000 unique SNPs were assayed.

Findings—We produced more than 200 million genotypes in 553 unique participants. The raw genotypes of all the controls have been posted publicly in a previous study of Parkinson's disease. From this effort, results of genotype and allele association tests have been publicly posted for 88% of stroke patients who provided proper consent for public release. Preliminary analysis of these data did not reveal any single locus conferring a large effect on risk for ischaemic stroke.

Interpretation—The data generated here comprise the first phase of a genome-wide association analysis in patients with stroke. Release of phase I results generated in these publicly available samples from each consenting individual makes this dataset a valuable resource for data-mining and augmentation.

Introduction

Ischaemic stroke is a common neurological disease and a leading cause of severe disability and death in developed countries.[1] About 85-90% of strokes are ischaemic.[2,3] In most cases, stroke is thought to be a multifactorial disorder or complex trait for which classic patterns of

inheritance cannot be shown. However, studies in family and animal models have consistently indicated a genetic influence on stroke risk and prognosis.[4-6]

Historically, the two most common approaches to finding genes involved in disease have been linkage and candidate gene association studies. For familial disorders with suitable family structures available for sampling, linkage has been a successful approach. The identification of genetic lesions underlying monogenic disorders has now become routine in many laboratories. In the context of stroke, this approach has provided some success, most notably with the identification of *NOTCH3* mutations underlying cerebral autosomal dominant arteriopathy with subcortical infacts and leukoencephalopathy (CADASIL).[5] Combined linkage and association approaches have led to the identification of putative risk factor loci for ischaemic stroke, namely *PDE4D* and *ALOX5AP* in the Icelandic population.[7,8]

In the context of diseases with an oligogenic or polygenic basis of genetic risk, population-based association studies provide more statistical power than family-based linkage studies.[9, 10] However, similar to other late-onset disorders, the identification of common alleles that affect stroke has been problematic. Until the advent of technology that made feasible the testing of thousands of genetic variants at once, the most practical and widely employed approach to identify disease-associated loci has been candidate gene association analysis. Many candidate genes for ischaemic stroke have been investigated with numerous statistically significant associations reported; however, few associations have been consistently replicated.[11]

The completion of the International Haplotype Map Project (HapMap), coupled with the availability of high-throughput genotyping methods, affords the opportunity to apply the powerful approach of association testing in a genome-wide manner.[12,13] A handful of genome-wide association studies have so far been published on diseases including age-related macular degeneration, Parkinson's disease, myocardial infarction, and inflammatory bowel disease, and at least two of these have identified novel loci that have been independently replicated.[14-18]

In an attempt to define whether there is a common genetic risk factor underlying risk for stroke, and to generate publicly available genome-wide genotypes that can be reanalysed or augmented by others, we did whole-genome genotyping using more than 400 000 unique SNPs from the Illumina Infinium Human-1 and HumanHap300 assays in a cohort of 278 patients with ischaemic stroke and 275 neurologically normal controls. Here we present these data and an initial analysis of this genotyping effort.

Methods

Participants

All control samples were from the National Institute of Neurological Disorders and Stroke Neurogenetics Repository. All individuals involved in this study gave written consent for the genetic analysis. As described previously,[16] the panels containing neurologically normal control samples were *NDPT002*, *NDPT006*, and *NDPT008*; these consist of DNA from 275 unique individuals and one replicate sample. Blood samples were drawn from white individuals who were unrelated and neurologically normal at many different sites within the USA. All individuals underwent a detailed medical history interview. None had a history of stroke, Alzheimer's disease, amyotrophic lateral sclerosis, ataxia, autism, bipolar disorder, brain aneurysm, dementia, dystonia, or Parkinson's disease. Sum scores on the mini-mental state examination[19] ranged from 26 to 30, and all were interviewed for detailed family history. None had any first-degree relative with a known primary neurological disorder including amyotrophic lateral sclerosis, ataxia, autism, brain aneurysm, dystonia, Parkinson's disease, or schizophrenia. The mean age at sample collection was 68 years (range 55-88 years).

All stroke samples used in the current study came from the Ischaemic Stroke Genetics Study (ISGS), which was a prospective five-centre case-control study in the USA. The protocol for ISGS has been reported previously.[20] For the stroke cohort, all cases had recent (within 30 days) first-ever ischaemic stroke confirmed by history, physical examination, and head imaging (CT or MRI). Stroke was defined according to the WHO definition.[21] Iatrogenic, septic embolic, vasospastic, and vasculitic stroke cases were excluded.

A genotype-blinded neurologist rater (RDB) classified ischaemic strokes according to the pre-specified Trial of Org 10172 in Acute Stroke Treatment (TOAST),[22] Oxfordshire,[23] and Baltimore[24] criteria on the basis of a detailed medical record review. Video-certified examiners assessed neurological impairment using the NIH stroke scale.[25] Functional status at baseline and at 90 days was assessed using the Barthel index,[26] Oxford handicap scale,[27] and the Glasgow outcome score.[28]

The work presented here represents the first phase of a multi-stage genetic association study. The first stage is a 275 case, 275 control, genome-wide association scan using more than 408000 SNPs. In the second stage, 1225 independent cases and 1225 independent controls will be genotyped for roughly 3000 of the most highly associated SNPs. The significance level for proceeding to genotype in the second stage is 0.0075 (3000 of 400000).

Using an estimated cumulative incidence of ischaemic stroke of 10% for adults over age 55 years, we have calculated the power estimates to detect an associated SNP with given minor allele frequency and odds ratio in a series of 250 cases and 250 controls (webtable 1). Thus, for the ultimate two-stage design, the power to detect a SNP with odds ratio of 1.50, assuming a minor allele frequency of 0.20 at $p=5\times10^{-7}$, is 89%. Using this two-stage strategy, there is excellent power (greater than 80%) to detect stroke susceptibility loci with realistically modest effect (odds ratio around 1.5) and low (but still common) disease allele frequency (around 0.15). Although this calculation does not take into account incomplete coverage, we estimate that the combination of data from the gene-centric Human-1 and haplotype tagging HumanHap300 chips provides excellent coverage in our population, and thus feel that this would have minimum effect on these power calculations.

In comparison to a two-stage design, the power of a single-stage design with 250 cases and 250 controls is substantially lower (webtable 2). If this collection of cases and controls was the only component of the genetic study, only susceptibility loci with large effect (odds ratio >2) and common frequency (>0.30) could be detected. Such loci would probably be detected by linkage, since they would have an effect equivalent to that of HLA and in patients with type 1 diabetes (originally detected with 100 cases and 100 controls).

All the control samples and 88% (219 of 249) of the stroke patients included in the association analysis gave consent for public release of their genotype data. 24 patients with stroke, whose data predated the inclusion within the NINDS Neurogenetics Repository, had given consent but did not explicitly agree to public sample release or data sharing, so we have not released raw genotype data on these individuals.

Procedures

Epstein-Barr virus immortalisation of peripheral blood lymphocytes was done as previously described[29,30] and DNA was extracted using a modified salting out procedure.[31] DNA was also extracted from 0.5 mL of blood for subsequent quality control steps in the cell banking process. DNA for the analyses was extracted from the Epstein-Barr virus immortalised lymphocyte cell lines; these cell lines remain largely faithful to the genotype of the source tissue when examined by high-density SNP genotyping assays.[32]

All samples were assayed with the Illumina Infinium Human-1 and HumanHap300 SNP chips (Illumina Inc, San Diego, CA, USA). The Human-1 product assays 109 365 gene-centric SNPs and the HumanHap300 product assays 317 511 haplotype tagging SNPs derived from phase I of the international HapMap project. There are 18 073 SNPs in common between the two arrays; thus the assays combined provide data on 408 803 unique SNPs. Any assay with a call rate below 95% was repeated on a fresh DNA aliquot; if the call rate persisted below this level the sample was excluded from the analysis.

All chips were scanned using the Illumina BeadStation system. Human-1 chips were all scanned with settings standard for that product; HumanHap300 chips were scanned with one of two settings; a slow scan setting (about 90 min scan time) or a fast scan setting (around 40 min scan time), which was made available during this analysis. Genotype concordance rates between these two analyses are extremely high (0.9999 [SD 0.001]).

Data were analysed with BeadStudio v2.1.10.0 (Illumina Inc, San Diego CA, USA). All genotypes were stored within an open source in-house genotype database GERON genotyping; this database was also used for data manipulation and export for analysis using the programs STRUCTURE[33] version 2.1 and SNPGWA version 2.2.

Statistical analysis

Population substructure was examined with STRUCTURE version 2.1 in the entire sample of cases and controls.[33] Specifically, 267 SNPs were selected from across the autosomes ensuring that no two of the selected SNPs were in linkage disequilibrium and that each had a minor allele frequency of more than 10%. Global tests for substructure were computed and individual observations that showed departure from the general population were identified.

Statistical analysis of the raw genotype data was done with the software SNPGWA. Each SNP was tested for departures from Hardy-Weinberg equilibrium (HWE) expectations in the case, control, and combined samples using the exact test.[34] Case departures from HWE expectations were compared with control proportion patterns for insight into possible genetic models. Linkage disequilibrium statistics, D' and r^2, were computed for each tandem pair of SNPs. To identify any association between the individual polymorphism and stroke status, several tests were done with SNPGWA. These included the overall 2-degree of freedom test (genotype), tests of the additive genetic model (Cochran-Armitage trend test), and the corresponding test for lack of fit to additivity. Tests of allelic association and association under dominant and recessive models are also reported. For SNPs on chromosome X not within the PAR1 (pseudoautosomal region 1) and PAR2 (pseudoautosomal region 2) regions, only the dominant genetic model was considered. The human X and Y chromosomes are morphologically and genetically distinct; however, there are X-Y homologous regions, PAR1 and PAR2, which pair and recombine at meiosis.

For sets of tandem SNPs, allelic and two-marker and three-marker moving window haplotype tests were computed using the expectation-maximisation algorithm implemented in SNPGWA. For all analyses, odds ratios and 95% CIs were computed. The above analyses were repeated for the separate TOAST subtypes.

Using the case-control data, we computed a series of generalised estimating equations[35] that included relevant covariates (age, sex, hypertension, smoking status, diabetes mellitus, and heart disease). All modelling was done in a hierarchical manner, with a baseline model that included only the single nucleotide polymorphism as the predictor. Additional models were tested, with age and sex; further models were then tested by adding an individual stroke risk factor variable as a covariate. A final, fully saturated model that included all relevant covariates was analysed. P values were computed using the 2 degree-of-freedom generalised test of

association. A series of genetic models were tested (dominant, additive, recessive) for estimation of best fit for risk. Results were examined only for SNPs that were in HWE in controls and had less than a 5% missing genotype rate. Only the additive model was considered with the exception of SNPs on chromosome X, where the dominant model was examined.

Role of the funding source

The study sponsors had no role in the study design, data collection, data analysis, data interpretation, or writing of the report. The corresponding author had full access to all the data in the study and had final responsibility for the decision to submit for publication.

Results

DNA samples from 278 patients with stroke and 275 control subjects were genotyped using the Human-1 beadchip; 2 stroke cases and 1 control subject were genotyped in replicate to assess reproducibility of the genotyping platform. Ten stroke samples were dropped because of insufficient DNA and nine stroke samples did not reach the quality threshold of 95% call rate, or showed discordance between expected sex and genotype. Thus the ischaemic stroke cohort taken through to the analysis by STRUCTURE consisted of 259 unique individuals. In the control cohort, four samples were dropped because of low-quality genotyping, and two samples showed a discrepancy between expected sex and genotype. The control cohort used for STRUCTURE analysis therefore consisted of 269 unique individuals. The median genotype success rate for the 408 803 SNPs studied in our population was 99.72% (mean 99.44%; range 96.14% to 99.97%) for samples that passed the initial quality check (call rate ≥95%). Genotyping of three replicate samples showed concordance rates of 99.55% (99.81% for the control replicate with Human-1, and 99.12% and 99.71% for the two cases replicated with the HumanHap300 assays). Analysis of the 18 073 SNPs that overlap between the Human-1 and HumanHap300 products revealed genotype concordance rates of 99.97% (range 98.53% to 99.99%) between the assays across 528 samples (259 samples from patients with ischaemic stroke and 269 from unrelated population controls) including both samples scanned with fast scan and slow scan settings.

STRUCTURE failed to detect overall differences in the population substructure between cases and controls. However, 2% (11 of 528) of samples, which consisted of the ten stroke cases and one control reported previously,[16] seem to have a substantially different genetic background (figure). Assessment of these samples revealed that all were from self-identified African Americans. On the basis of the STRUCTURE results, these 11 patients were not included in subsequent analyses. Statistical comparison for association was therefore done with 249 white patients with stroke and 268 white neurologically normal controls.

Table 1 shows the clinical and demographic characteristics of the study population by case status. As expected, traditional vascular risk factors are more common in cases with ischaemic stroke than in stroke-free controls. Table 2 shows stroke severity, stroke subtype, and functional status among cases.

Hardy-Weinberg equilibrium (HWE) was tested in the controls for all loci with a call rate of 95% or more. A total of 394 513 SNPs and 374 544 SNPs had a HWE p value higher than 0.001 and 0.05, respectively. Statistical analysis of association was done on all genotypes irrespective of Hardy-Weinberg disequilibrium or minor allele frequency. However, in table 3 we present the SNPs with a p value of $<1\times10^{-5}$ after adjusting for demographic and stroke risk factors (age, sex, hypertension, smoking status, diabetes mellitus, and heart disease) with call rates more than 95% and HWE of $p<0.001$.

Analyses according to TOAST stroke subtypes were also done. Webtables 3, 4, and 5 summarise the most significant SNPs for each of cardioembolic, large vessel and small vessel stroke subtypes with call rates more than 95% and HWE of $p<0.001$ in the control cohort.

Raw p values and genotype and allele frequencies for all loci and under each model are available at the National Center for Biotechnology Information dbGAP website. Individual genotype data for all the controls and 88 % of cases are also available at this site. Within 9 months of publication we will also release raw scan data to enable the analysis of structural genomic variation.

Discussion

We present here an initial genome-wide SNP association study in ischaemic stroke, which compared 408 803 unique SNPs in 249 white patients with ischaemic stroke and 268 white neurologically normal controls. The ischaemic stroke cohort, prospectively ascertained at five US stroke centres, is comparable to population-based cohorts in the United States in terms of its conventional atherosclerotic risk factor profile.[36] The control cohort showed a paucity of conventional risk factors relative to what would be expected from an age-related population-based sample, probably due to the restrictive control eligibility criteria and possible volunteer bias.

As expected, our screening approach yields hundreds of nominally statistically significant associated markers, leaving the challenge to distinguish true associations from those that are false positives. None of our results are significant after Bonferroni correction. However, in view of the correlation between tests in most SNP settings, this correction is probably overly conservative. The individual risk provided by SNPs (table 3) is moderate-high with odds ratios ranging from 0.40 (0.26-0.62) to 0.54 (0.41-0.74) and 1.9 (1.42-2.65) to 8 (2.85-22.33). In a recent meta-analysis[37] from 120 case-control studies, 32 genes, and around 18 000 cases of stroke and 58 000 controls, the summary odds ratio for those candidate genes with significant association varied from 1.21 (95% CI 1.08- 1.35) for the angiotensin-converting enzyme (ACE) insertion or deletion polymorphism to 1.88 (1.28-2.76) for a polymorphism in the Kozak sequence of GPIBA (encoding glycoprotein Ib-α). Most published candidate genes studies report significant association with ischaemic stroke with an odds ratio of <3.11 Clearly the current study in isolation is underpowered to detect genetic loci that exert a moderate risk for stroke; however, as discussed, as a part of our two stage design, we have excellent power to detect effects of odds ratios higher than 1.5 conferred by common risk alleles. The posting of the raw genotype data will allow other interested parties to also pursue independent follow-up work.

Notably, some of the most significant SNPs listed in table 3 are within or near interesting candidate loci; for example, two genes involved in potassium transport, *KCNIP4* and *KCNK17*. Two different functions have been suggested for *KCNIP4*. Firstly, all KCNIP family proteins modulate the activity of Kv4 A-type potassium channels, which contribute to the frequency of slow repetitive firing and back-propagation of action potentials in neurons and shape the action potential in the heart.[38] Secondly, *KCNIP4* has been shown to have a role in presenilin function.[39,40] KCNK17 is a member of the acid-sensitive subfamily of tandem pore K^+ channels, which are open at all membrane potentials and contribute to cellular resting membrane potential. KCNK17 transcripts are widely expressed in humans, with highest levels in liver, lung, pancreas, placenta, aorta, and heart.[41] In the heart, background K^+ currents are thought to modulate the cardiac action potential.[42,43]

In the stroke cohort, ischaemic strokes were classified according to TOAST criteria. Subtype analysis greatly reduces statistical power, making all conclusions preliminary. Nonetheless, it

is of interest to note that the cardioembolic subtype of ischaemic stroke showed an association with common variation in *APEG-1* (aortic preferentially expressed gene 1). Expression of *APEG-1* gene is thought to serve as a marker for differentiated vascular smooth muscle cells; given that alterations in arterial smooth muscle cells phenotypes have an important role in the pathogenesis of vascular diseases and angiogenesis, *APEG-1* could be a good candidate.[44]

Although there is a temptation to speculate on the potential pathogenic roles that the most statistically significantly associated genes might have in disease, we should emphasise that this first-stage association work is not designed to unequivocally link genetic variability, conferring a risk of the size identified here, with stroke. As described, the approach presented here is taken as an intermediate step where putative associations will be re-tested in separate cohorts. As such, the loci listed in table 3 should not be regarded as the only regions warranting follow-up. More appropriately, several thousand of the most significantly associated loci will be re-tested in independent series. Some studies point out that analysing the data from both stages jointly can be more powerful than treating the second stage as a stand-alone replication study.[45] Public release of genotype data makes the comparison and combination of experiments easier thus increasing power.

Our data strongly suggest that there is no single common genetic variant exerting a major risk on stroke. This finding contrasts with our recent study of Alzheimer's disease, which showed that APOE ε4 was the only allele in the entire genome conferring a risk of an odds ratio higher than 2.[46] Although the absence of evidence does not necessarily imply evidence of absence, the genomic coverage of the methods applied here for white populations is in the order of 90% (at r^2 of 0.8). Thus, we have a fair degree of certainty to believe in the absence of one common variant that exerts a strong effect on risk for ischaemic stroke in this population. However, subsequent genome-wide association studies in larger cohorts, and focused follow-up of candidate loci, will be key steps in delineating the role that common genetic variability has in risk for ischaemic stroke. We believe this study is a necessary first step in the elucidation of genetic risk factors underpinning the third most common cause of mortality in the developed world.

Acknowledgments

The Ischaemic Stroke Genetics study (ISGS) is funded by a grant from the National Institute of Neurological Disorders and Stroke (R01 NS42733). We thank the participants and the submitters for depositing samples at the NINDS neurogenetics repository. Many samples for this study are derived from the NINDS neurogenetics repository at Coriell Cell Repositories, and the data are available from the website. This study in part used the high-performance computational capabilities of the Biowulf PC/Linux cluster at the NIH, Bethesda, MD, USA. This work was supported by the intramural programmes of the National Institute on Aging and NINDS and by an extramural NINDS contract funding the Coriell Repository.

References

1. Bonita R. Epidemiology of stroke. Lancet 1992;339:342–44. [PubMed: 1346420]
2. Rothwell PM, Coull AJ, Giles MF, et al. Change in stroke incidence, mortality, case-fatality, severity, and risk factors in Oxfordshire, UK from 1981 to 2004 (Oxford Vascular Study). Lancet 2004;363:1925–33. [PubMed: 15194251]
3. Brown RD, Whisnant JP, Sicks JD, O'Fallon WM, Wiebers DO. Stroke incidence, prevalence, and survival: secular trends in Rochester, Minnesota, through 1989. Stroke 1996;27:373–80. [PubMed: 8610298]
4. Hassan A, Markus HS. Genetics and ischaemic stroke. Brain 2000;123:1784–812. [PubMed: 10960044]
5. Joutel A, Corpechot C, Ducros A, et al. Notch3 mutations in CADASIL, a hereditary adult-onset condition causing stroke and dementia. Nature 1996;383:707–10. [PubMed: 8878478]

6. Tournier-Lasserve E. New players in the genetics of stroke. N Engl J Med 2002;347:1711–12. [PubMed: 12444190]
7. Helgadottir A, Manolescu A, Thorleifsson G, et al. The gene encoding 5-lipoxygenase activating protein confers risk of myocardial infarction and stroke. Nat Genet 2004;36:233–39. [PubMed: 14770184]
8. Gretarsdottir S, Thorleifsson G, Reynisdottir ST, et al. The gene encoding phosphodiesterase 4D confers risk of ischaemic stroke. Nat Genet 2003;35:131–38. [PubMed: 14517540]
9. Gershon ES, Goldin LR. Clinical methods in psychiatric genetics. I. Robustness of genetic marker investigative strategies. Acta Psychiatr Scand 1986;74:113–18. [PubMed: 3465198]
10. Risch N, Merikangas K. The future of genetic studies of complex human diseases. Science 1996;273:1516–17. [PubMed: 8801636]
11. Casas JP, Hingorani AD, Bautista LE, Sharma P. Meta-analysis of genetic studies in ischaemic stroke: thirty-two genes involving approximately 18,000 cases and 58,000 controls. Arch Neurol 2004;61:1652–61. [PubMed: 15534175]
12. The International HapMap Project. Nature 2003;426:789–96. [PubMed: 14685227]
13. A haplotype map of the human genome. Nature 2005;437:1299–320. [PubMed: 16255080]
14. Maraganore DM, de Andrade M, Lesnick TG, et al. High-resolution whole-genome association study of Parkinson disease. Am J Hum Genet 2005;77:685–93. [PubMed: 16252231]
15. Klein RJ, Zeiss C, Chew EY, et al. Complement factor H polymorphism in age-related macular degeneration. Science 2005;308:385–89. [PubMed: 15761122]
16. Fung HC, Scholz S, Matarin M, et al. Genome-wide genotyping in Parkinson's disease and neurologically normal controls: first stage analysis and public release of data. Lancet Neurol 2006;5:911–16. [PubMed: 17052657]
17. Ozaki K, Ohnishi Y, Iida A, et al. Functional SNPs in the lymphotoxin-alpha gene that are associated with susceptibility to myocardial infarction. Nat Genet 2002;32:650–54. [PubMed: 12426569]
18. Duerr RH, Taylor KD, Brant SR, et al. A genome-wide association study identifies IL23R as an inflammatory bowel disease gene. Science 2006;314:1461–63. [PubMed: 17068223]
19. Folstein MF, Folstein SE, McHugh PR. Mini-mental state: a practical method for grading the cognitive state of patients for the clinician. J Psychiatr Res 1975;12:189–98. [PubMed: 1202204]
20. Meschia JF, Brott TG, Brown RD Jr, et al. The Ischaemic Stroke Genetics Study (ISGS) Protocol. BMC Neurol 2003;3:4. [PubMed: 12848902]
21. Investigators WMPP. The World Health Organization MONICA Project (monitoring trends and determinants in cardiovascular disease): a major international collaboration. J Clin Epidemiol 1998;41:105–14.
22. Adams HP Jr. Bendixen BH, Kappelle LJ, et al. Classification of subtype of acute ischaemic stroke. Definitions for use in a multicenter clinical trial. TOAST. Trial of Org 10172 in Acute Stroke Treatment. Stroke 1993;24:35–41. [PubMed: 7678184]
23. Bamford J, Sandercock P, Dennis M, Burn J, Warlow C. Classification and natural history of clinically identifiable subtypes of cerebral infarction. Lancet 1991;337:1521–26. [PubMed: 1675378]
24. Johnson CJ, Kittner SJ, McCarter RJ, et al. Interrater reliability of an etiologic classification of ischemic stroke. Stroke 1995;26:46–51. [PubMed: 7839396]
25. Lyden P, Brott T, Tilley B, et al. Improved reliability of the NIH Stroke Scale using video training. NINDS TPA Stroke Study Group. Stroke 1994;25:2220–26. [PubMed: 7974549]
26. Collin C, Wade DT, Davies S, Horne V. The Barthel ADL Index: a reliability study. Int Disabil Stud 1988;10:61–63. [PubMed: 3403500]
27. Bamford JM, Sandercock PA, Warlow CP, Slattery J. Interobserver agreement for the assessment of handicap in stroke patients. Stroke 1989;30:828. [PubMed: 2728057]
28. Jennett B, Bond M. Assessment of outcome after severe brain damage. Lancet 1975;305:480–84. [PubMed: 46957]
29. Miller G, Shope T, Lisco H, Stitt D, Lipman M. Epstein-Barr virus: transformation, cytopathic changes, and viral antigens in squirrel monkey and marmoset leukocytes. Proc Natl Acad Sci USA 1972;69:383–87. [PubMed: 4333982]

30. Tumilowicz JJ, Gallick GE, East JL, Pathak S, Trentin JJ, Arlinghaus RB. Presence of retrovirus in the B95-8 Epstein-Barr virus-producing cell line from different sources. In Vitro 1984;20:486–92. [PubMed: 6086497]

31. Miller SA, Dykes DD, Polesky HF. A simple salting out procedure for extracting DNA from human nucleated cells. Nucleic Acids Res 1988;16:1215. [PubMed: 3344216]

32. Simon-Sanchez J, Scholz S, Fung HC, et al. Genome-wide SNP assay reveals structural genomic variation, extended homozygosity and cell-line induced alterations in normal individuals. Hum Mol Genet 2007;16:1–14. [PubMed: 17116639]

33. Falush D, Stephens M, Pritchard JK. Inference of population structure using multilocus genotype data: linked loci and correlated allele frequencies. Genetics 2003;164:1567–87. [PubMed: 12930761]

34. Wigginton JE, Cutler DJ, Abecasis GR. A note on exact tests of Hardy-Weinberg equilibrium. Am J Hum Genet 2005;76:887–93. [PubMed: 15789306]

35. Zeger SL, Liang KY. Longitudinal data analysis for discrete and continuous outcomes. Biometrics 1986;42:121–30. [PubMed: 3719049]

36. White H, Boden-Albala B, Wang C, et al. Ischaemic stroke subtype incidence among whites, blacks, and Hispanics: the Northern Manhattan Study. Circulation 2005;111:1327–31. [PubMed: 15769776]

37. Casas JP, Hingorani AD, Bautista LE, Sharma P. Arch Neurol 2004;61:1652–61. [PubMed: 15534175]

38. Holmqvist MH, Cao J, Hernandez-Pineda R, et al. Elimination of fast inactivation in Kv4 A-type potassium channels by an auxiliary subunit domain. Proc Natl Acad Sci USA 2002;99:1035–40. [PubMed: 11805342]

39. Pruunsild P, Timmusk T. Structure, alternative splicing, and expression of the human and mouse KCNIP gene family. Genomics 2005;86:581–93. [PubMed: 16112838]

40. Morohashi Y, Hatano N, Ohya S, et al. Molecular cloning and characterization of CALP/KChIP4, a novel EF-hand protein interacting with presenilin 2 and voltage-gated potassium channel subunit Kv4. J Biol Chem 2002;277:14965–75. [PubMed: 11847232]

41. Decher N, Maier M, Dittrich W, et al. Characterization of TASK-4, a novel member of the pH-sensitive, two-pore domain potassium channel family. FEBS Lett 2001;492:84–89. [PubMed: 11248242]

42. Lesage F, Reyes R, Fink M, Duprat F, Guillemare E, Lazdunski M. Dimerization of TWIK-1 K+ channel subunits via a disulfide bridge. Embo J 1996;15:6400–07. [PubMed: 8978667]

43. Kim Y, Bang H, Kim D. TASK-3, a new member of the tandem pore K(+) channel family. J Biol Chem 2000;275:9340–47. [PubMed: 10734076]

44. Hsieh CM, Fukumoto S, Layne MD, et al. Striated muscle preferentially expressed genes alpha and beta are two serine/threonine protein kinases derived from the same gene as the aortic preferentially expressed gene-1. J Biol Chem 2000;275:36966–73. [PubMed: 10973969]

45. Skol AD, Scott LJ, Abecasis GR, Boehnke M. Joint analysis is more efficient than replication-based analysis for two-stage genome-wide association studies. Nat Genet 2006;38:209–13. [PubMed: 16415888]

46. CoonKMyersAJCraigDWA high-density whole-genome association study reveals that apoe is the major susceptibility gene for sporadic late-onset Alzheimer’s disease. J Clin Psychiatry2007 (in press).

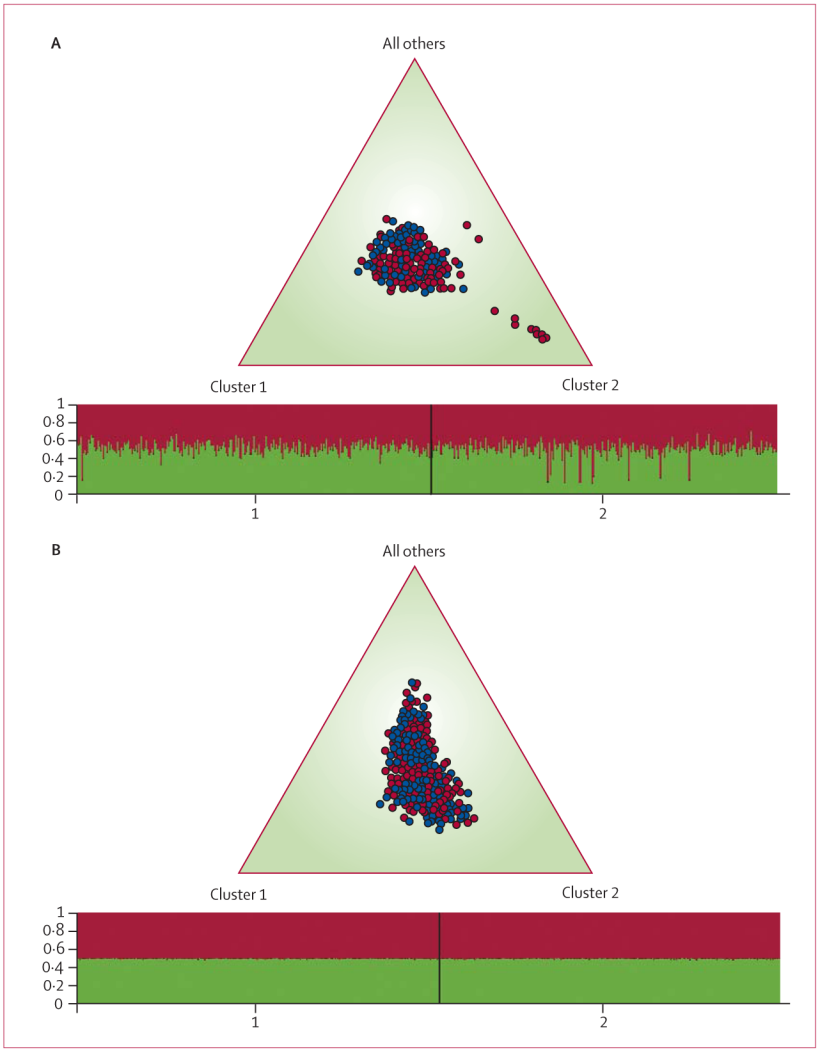

Figure. STRUCTURE results on cases and controls

276 autosomal markers were analysed using STRUCTURE (version 2.1) to determine if there was substantial population substructure within or between the cases and controls. Shown in both the upper and lower panels are triangle and bar plots with the number of assumed populations set to 3 and 2 respectively. In triangle plots, cases are red and controls are blue, and in the bar plot cases are group 2 and controls group 1. Analysis of all 259 cases and 269 controls (panel A) shows 11 samples clustering outside of the main group. These 11 outlier samples are self-reported African American patients (ten cases and one control). Panel B shows the results of analysis using STRUCTURE following removal of these 11 samples; these results show that there is a lack of substantial population substructure in this cohort and this was the cohort used for the final analysis.

Table 1

Demographic and clinical characteristics of cases and controls in ISGS

	Controls (%)	Cases (%)	p
Total (number of patients)	268	249	
Age (years)			
Mean	69.5	71.8	0.0096
Sex			
Men	128 (47.8)	134 (53.8)	0.1689
Women	140 (52.2)	115 (46.2)	
History			
Heart disease	35/268 (13.1)	88/249 (35.3)	<0.0001
Hypertension	66/268 (24.6)	173/247 (70.0)	<0.0001
Diabetes mellitus	19/268 (7.1)	61/249 (24.5)	<0.0001
Smoking			
Never	146/256 (57)	82/249 (32.9)	<0.0001
Former	91/256 (35.6)	122/249 (49.0)	0.0015
Active smoker	19/256 (7.4)	45/249 (18.1)	<0.0006

Table 2

Stroke characteristics in ISGS

Characteristic	Cases (%)
Total	249
TOAST criteria	
Large artery	
Probable	53 (21.3)
Possible	0
Lacunar	
Probable	38 (15.3)
Possible	4 (1.6)
Cardioembolic	
Probable	54 (21.7)
Possible	15 (6)
Other causes	
Probable	5 (2)
Possible	1 (0.4)
Undetermined cause	
>1 possible cause identifi ed	18 (72)
Negative work-up	52 (20.9)
Incomplete work-up	9 (3.6)
Oxfordshire criteria	
LACI	66 (26.5)
TACI	23 (9.2)
PACI	112 (45)
POCI	48 (19.3)
NIHSS score	
0–1	91 (36.6)
2–4	88 (35.3)
≥5	70 (28.1)
Oxford handicap scale at enrolment	
0	34 (13.7)
1	63 (25.3)
2	69 (27.7)
3–5	83 (33.3)
Barthel index at enrolment	
100–95	113 (45.4)
90–80	38 (15.3)
75–0	98 (39.4)

LACI=lacunar cerebral infarct; TACI=total anterior circulation cerebral infarct PACI=partial anterior circulation cerebral infarct; POCI=posterior circulation cerebral infarct

Table 3

SNPs with a p value less than 1×10^{-5} using the additive model adjusted for known stroke risk factors

Chromosome location	dbSNP ID	Location (Build 36.1)	Gene	HWE p value	p	MAF cases	MAF controls	Pct_ missing	Odds ratios*	95% CI*
1q31	rs11360	153442110	BCAN	0.07	3.57E-05	0.39	0.50	0	1.94	(1.42-2.65)
1q31	rs10494737	192594516	Intergenic	0.38	4.69E-05	0.30	0.22	0	0.50	(0.36-0.70)
2q21	rs2118844	135001332	MGAT5	0.79	4.35E-05	0.24	0.38	0	2.01	(1.44-2.82)
2q24.2	rs10497212	160790207	ITGB6	0.11	8.49E-05	0.12	0.19	0.0232	2.64	(1.65-4.23)
2q31.1	rs10204475	170758029	ZNF650	0.83	5.36E-05	0.09	0.18	0	2.50	(1.58-3.95)
4p15.31	rs4697177	20436358	KCNIP4	0.58	6.8E-05	0.25	0.33	0.0484	2.73	(1.74-4.29)
4q13	rs13126803	63207032	Intergenic	0.13	1.2E-05	0.10	0.18	0.0484	1.99	(1.42-2.78)
5q31	rs246341	134061676	Intergenic	0.42	7.72E-05	0.01	0.04	0.0019	2.00	(1.43-2.78)
5q34	rs32720	165225899	Intergenic	0.79	4.55E-05	0.26	0.39	0	7.98	(2.85-22.3)
6p21.1	rs2395721	39378456	KCNK17	1.00	5.53E-05	0.29	0.18	0	5.79	(2.66-12.6)
6p21.1	rs10947803	39378588	KCNK17	1.00	5.65E-05	0.28	0.18	0.0019	0.47	(0.32-0.68)
6p21.1	rs10807204	39381404	KCNK17	0.68	6.7E-05	0.27	0.18	0	0.47	(0.32-0.68)
6q21	rs783396	107094063	AIM1	0.49	9.24E-06	0.03	0.10	0.0348	0.46	(0.32-0.68)
7p21	rs10486776	15516026	Intergenic	1.00	6.16E-06	0.03	0.10	0.0251	5.62	(2.66-11.9)
7p21	rs2192476	19385545	Intergenic	1.00	4.92E-05	0.30	0.39	0	1.99	(1.43-2.78)
9q21	rs7043482	82365469	Intergenic	0.20	1.16E-05	0.29	0.39	0.0019	0.50	(0.37-0.68)
9q33.1	rs3761845	116850034	ASTN2	0.80	8.63E-05	0.50	0.38	0.0019	2.11	(1.51-2.95)
9q33.1	rs10817974	116855160	ASTN2	0.79	1.11E-05	0.50	0.37	0.0019	0.55	(0.41-0.74)
11p15	rs12146588	19191730	Intergenic	0.09	8.75E-05	0.05	0.12	0.0484	0.40	(0.26-0.62)
11q24	rs588407	128201807	Intergenic	1.00	3.45E-05	0.19	0.11	0.0155	3.31	(1.82-6.02)
p11,21	rs11052413	33097905	Intergenic	0.45	1.18E-05	0.45	0.42	0.0174	0.49	(0.35-0.67)
13q12.12	rs2793483	23657688	SPATA13	0.37	3.78E-05	0.41	0.29	0.0039	0.48	(0.35-0.66)
13q21	rs9536591	53479088	Intergenic	0.37	5.6E-06	0.49	0.39	0.0039	0.52	(0.38-0.71)
14q23-q24.2	rs229673	64351697	SPTB	0.88	6.7E-05	0.20	0.29	0.0097	2.10	(1.46-3.03)
18p11.2	rs7506045	11977272	IMPA2	0.15	7.07E-07	0.05	0.10	0.0039	5.39	(2.77-10.5)
18q11.2	rs1539829	20811162	Intergenic	1.00	3.78E-05	0.3394	0.2444	0	0.49	(0.35-0.69)
22q12	rs5753271	29395047	Intergenic	1.00	4.35E-05	0.0586	0.1311	0.0213	3.23	(1.84-5.67)

Ischaemic stroke risk factors: age, sex, hypertension, smoking status, diabetes mellitus, and heart disease. SNPs where HWE in controls was p<0.001 or genotypes were successful in less than 95% of samples were excluded. Although SNPs outlined here are candidates, an appropriate replication or joint-analysis follow-up is needed and would include genotyping of loci that are significant down to a less stringent p value. HWE=Hardy-Weinberg equilibrium; MAF=minor allele frequency; Pct missing=percentage of genotypes missing

* listed p values, odds ratios, and 95% confidence intervals were calculated using the additive model of genetic association.

Lancet Neurol. Author manuscript; available in PMC 2009 January 5.