Background and Purpose—Ischemic stroke has a strong familial component to risk. The Siblings With Ischemic Stroke Study (SWISS) is a genome-wide, family-based analysis that included use of imputed genotypes. The Siblings With Ischemic Stroke Study was conducted to examine the associations between single-nucleotide polymorphisms (SNPs) and risk of stroke and stroke subtypes within pairs.

Methods—The Siblings With Ischemic Stroke Study enrolled 312 probands with ischemic stroke from 70 US and Canadian centers. Affected siblings were ascertained by centers and confirmed by central record review; unaffected siblings were ascertained by telephone contact. Ischemic stroke was subtyped according to Trial of Org 10172 in Acute Stroke Treatment criteria. Genotyping was performed with an Illumina 610 quad array (proband) and an Illumina linkage V array (affected siblings). SNPs were imputed by using 1000 Genomes Project data and MACH software. Family-based association analyses were conducted by using the sibling transmission-disequilibrium test.

Results—For all pairs, the correlation of age at stroke within pairs of affected siblings was \( r = 0.83 \) (95% CI, 0.78–0.86; \( P<2.2\times10^{-16} \)). The correlation did not differ substantially by subtype. The concordance of stroke subtypes among affected pairs was 33.8\% (kappa = 0.13; \( P=5.06\times10^{-4} \)) and did not differ by age at stroke in the proband. Although no SNP achieved genome-wide significance for risk of ischemic stroke, there was clustering of the most associated SNPs on chromosomes 3p (neuronal nitric oxide synthase) and 6p.

Conclusions—Stroke subtype and age at stroke in affected sibling pairs exhibit significant clustering. No individual SNP reached genome-wide significance. However, 2 promising candidate loci were identified, including 1 that contains neuronal nitric oxide synthase, although these risk loci warrant further examination in larger sample collections. (Stroke. 2011; 42:2726-2732.)

Key Words: cerebral infarct • genetics

Ischemic stroke has long been recognized to cluster in families. This clustering has been attributed to both genetic factors and common environmental exposures. Gene mutations have been identified that lead to rare ischemic stroke syndromes like CADASIL and MELAS. However, the search for genetic loci associated with ischemic stroke risk has yielded meager results thus far, despite several candidate-gene\(^1\) and large-scale, genome-wide association\(^2,3\) studies. The Siblings With Ischemic Stroke Study (SWISS) provides a different and promising approach to discover novel risk factors for ischemic stroke through the study of unrelated families with affected and unaffected siblings. Here we present the results of the family-based, genome-wide scan of SWISS after achieving the original recruitment goal.

Methods

Study Subjects

Probands were recruited at 70 US and Canadian medical centers. Probands were adult (\( >18 \) years old) men and women presenting to a participating center with a study neurologist–confirmed ischemic stroke. Stroke was defined as rapidly developing signs of a focal or global disturbance of cerebral function, with symptoms lasting at

Received March 16, 2011; accepted April 1, 2011.
From the Department of Neurology (J.F.M., T.G.B.), Mayo Clinic, Jacksonville, FL; Department of Neurology (R.D.B. Jr), Mayo Clinic, Rochester, MN; Department of Molecular Neuroscience Institute of Neurology (J.H.), University College London, London, UK; Department of Neurology (B.K.), University of Cincinnati, Cincinnati, OH; National Institute on Aging (M.N., A.S., D.H., L.F.), Bethesda, MD; University College London Genomics and Institute of Child Health (K.P.), London, UK; Coriell Institute for Medical Research (M.K.), Camden, NJ; Center for Public Health Genomics (S.S.R.), University of Virginia, Charlottesville, VA; Departments of Neurology and Public Health Sciences (B.B.W.), University of Virginia, Charlottesville, VA; and Department of Clinical and Experimental Epilepsy (M.M.), UCL Institute of Neurology, London, UK.

The online-only Data Supplement is available at http://stroke.ahajournals.org/lookup/suppl/doi:10.1161/STROKEAHA.111.620484/-/DC1.

*See Appendix for a list of participating centers and investigators.

Correspondence to James F. Meschia, Department of Neurology, Mayo Clinic, 4500 San Pablo Rd, Jacksonville, FL 32224. E-mail meschia.james@mayo.edu

© 2011 American Heart Association, Inc.

Stroke is available at http://stroke.ahajournals.org

DOI: 10.1161/STROKEAHA.111.620484
least 24 hours or leading to death, with no apparent cause other than vascular origin (World Health Organization definition).4 Stroke was defined as ischemic when computed tomography or magnetic resonance imaging of the brain was performed within 7 days of onset of stroke symptoms and identified the symptomatic cerebral infarct or failed to identify an alternative cause of symptoms. Probands were required to have reported at least 1 living full sibling with a history of stroke. No probands were enrolled with iatrogenic vasospastic or vasculitic stroke or if the stroke occurred in the setting of a mechanical heart valve or in the setting of untreated or actively treated bacterial endocarditis. Probands were also excluded if they were known to have CADASIL, Fabry disease, homocysteinuria, MELAS, or sickle-cell anemia. Study neurologists at each center assigned to the qualifying ischemic stroke of each proband a Trial of Org 10172 in Acute Stroke Treatment (TOAST) subtype diagnosis.5

Strokes-affected siblings of the proband (concordant siblings) were recruited by using proband-initiated contact.6 Telephone interviews were performed to obtain demographic and clinical information and to gain permission for obtaining medical records pertaining to treatment for stroke. Medical records were compiled and adjudicated by a central committee (J.F.M., T.G.B.) to verify the diagnosis of ischemic stroke and to assign a TOAST subtype diagnosis. Assignment of TOAST subtype diagnoses to SWISS concordant siblings has moderate interrater reliability.7 Unaffected siblings were ascertained by telephone contact and interview.

Genotyping Considerations
The establishment of lymphoblastoid cell lines, quality control of genomic DNA, acquisition of genetic data, and genotyping quality control metrics were performed according to procedures. Please see the online-only Data Supplement for these details.

Consensus single-nucleotide polymorphisms (SNPs) that passed quality control in both phases (genome-wide association and family-based phases) were merged for all available sibships (2239 SNPs were imputed in the probands). Using all 5612 SNPs in the merged dataset, we verified reported relationships by using pi_hat estimates. Sibships were confirmed when pairwise pi_hat values were between 0.35 and 0.65; samples were removed from a sibship when the estimated pi_hat value was not in this range. This dataset of the combined genotyping phases represents the final dataset for all subsequently described analyses. The flow of patients in the study is shown in Figure 1.

Genetic Data Analysis
All family-based analyses were conducted with PLINK 1.07 software.8 The dfam utility within PLINK implements a siblings-based transmission-disequilibrium test and was used to conduct these analyses. The dfam option is a powerful test for sibling-only datasets, incorporating data across sibships as well as using data from estimated parental genotypes to calculate expected allele frequencies for comparison with observed allele frequencies. The association test is based on the Cochran-Mantel-Haenszel test. Bonferroni correction for the number of tested SNPs corresponds to a minimum probability value for a genome-wide significance of $P<8.91\times10^{-6}$.

Additional Statistical Analyses
Frequencies of stroke risk factors (hypertension, hyperlipidemia, and diabetes) between affected and unaffected participants were compared by using $\chi^2$ tests. The correlation between affected sibling age at stroke was estimated by using the Pearson test of correlation. These analyses were conducted across all TOAST subtypes as well as after stratification by concordant and discordant subtypes among
affected sibling pairs. Linear regression was used to determine the confidence intervals and linear fit of the age association, as shown in Figure 2. Kappa statistics were calculated to quantify concordance of phenotypes of interest within sibling pairs for all ages and stroke subtypes as well as models stratified by age (<65-year proband as defining age strata) and stroke subtype. All analyses that did not include genetic data were conducted by using scripts written in R (R Development Core Team, 2008).9

Results

A total of 312 affected sibling pairs (312 probands) were enrolled at 70 centers across the United States and Canada. After quality control filtering, the final study population consisted of 223 probands, 248 stroke-affected siblings, and 84 stroke-unaffected siblings (total sample size, 555). Ischemic stroke–affected individuals had expected high rates of conventional atherosclerotic risk factors (Table 1). Stroke-affected individuals (probands and affected siblings) were significantly more likely to have hypertension (P < 0.0001), hyperlipidemia (P = 0.002), and diabetes (P = 0.008) than were stroke-unaffected individuals. Stroke-affected siblings were somewhat older than the probands. This difference of 2 years (P = 0.057) is expected, as an older sibling of the proband would be more likely to have a stroke than a younger sibling.

Sibling age at the time of stroke was strongly correlated with proband age at the time of stroke, despite the sibling’s being older. As shown in Figure 2 for all sibling pairs, the correlation coefficient was r = 0.83 (95% CI, 0.78–0.86; P < 2.2 × 10−16). For affected sibling pairs who had the same stroke subtype, the correlation coefficient was not different from all pairs, r = 0.83 (95% CI, 0.75–0.89; P < 2.2 × 10−16). This was the same for sibling pairs in which the affected siblings had different stroke subtypes, r = 0.83 (95% CI, 0.77–0.87; P < 2.2 × 10−16). More than 50% of the variance in age at stroke onset in siblings could be predicted by the age of the proband at the time of stroke. As shown in Table 2, there was significant concordance with affected siblings for TOAST subtype (kappa = 0.13, P = 5.06 × 10−4); this relation remained significant for sibling pairs in which the proband was <65 years old at the time of stroke and for sibling pairs in which the proband was 65 years or older.

Results of the genome-wide analyses are shown in Figure 3. Although no SNP association with ischemic stroke achieved a genome-wide level of significance, the 10 most associated SNPs exhibited clustering in several genomic regions. These 10 most significantly associated SNPs, their locations, frequencies, and effect estimates are shown in Table 3. The 10 SNPs represent 8 genomic loci, with minor allele frequencies ranging from 0.38 to 0.48 (common alleles). The effects for each are small, with odds ratios ranging from 0.96 to 1.04. The location of the most significantly associated SNPs (as well as others within 2.5 megabases) is shown in Figure 3 (indicated by blue shading). There are clusters of associated SNPs on chromosomes 3p and 6p. The SNPs on chromosome 3p lie in a strong candidate gene, neuronal nitric oxide synthase (NOS1).

Discussion

Our genome-wide scan for risk factors for ischemic stroke was performed in the largest collection of affected sibling pairs to date and showed potential loci of interest, includ-
ing a locus on NOS1. We are unaware of other genetic studies showing NOS1 variants to be associated with human ischemic stroke risk. However, knockout mouse models show that a deficiency in the NOS1 gene causes worsening neointimal formation and constrictive vascular remodeling. Genetic variants theoretically could impart elevated ischemic stroke risk by a vascular mechanism, by a thrombotic mechanism, or even by a parenchymal...
mechanism, whereby brain tissue is sensitized to focal ischemia. In the case of NOS1, it seems that variants would be more likely to impart risk by a vascular mechanism than by a parenchymal mechanism, since NOS1-knockout mice have smaller infarcts than wild-type mice in a middle cerebral artery occlusion model.11 It is important not to speculate beyond the strength of our observations, as no locus achieved genome-wide statistical significance.

Genome-wide studies have had mixed results in ischemic stroke. When SWISS was initiated, the human genome had only been sequenced in draft form.12 SWISS was predicated on the hypothesis that ischemic stroke obeyed the common disease, common variant hypothesis, which states that the genetic influences on many common diseases are attributable to a limited number of allelic variants present in >1% to 5% of the population.13 It has since become less clear that the hypothesis holds for ischemic stroke. No single locus has been identified in 2 genome-wide association studies at a genome-wide level of significance.14 Our study supports the idea that no single locus substantially contributes to ischemic stroke risk from the perspective of common variants contributing to disease risk, although future sequencing-based studies of rare variants may meet with substantially more success.

SWISS was designed to treat all types of ischemic stroke as a single phenotype. When SWISS was initiated, the human genome had only been sequenced in draft form.12 SWISS was predicated on the hypothesis that ischemic stroke obeyed the common disease, common variant hypothesis, which states that the genetic influences on many common diseases are attributable to a limited number of allelic variants present in >1% to 5% of the population.13 It has since become less clear that the hypothesis holds for ischemic stroke. No single locus has been identified in 2 genome-wide association studies at a genome-wide level of significance.14 Our study supports the idea that no single locus substantially contributes to ischemic stroke risk from the perspective of common variants contributing to disease risk, although future sequencing-based studies of rare variants may meet with substantially more success.

SWISS was designed to treat all types of ischemic stroke as a single phenotype. The phenotypic heterogeneity of ischemic stroke has long been appreciated, but categorizing subtypes of ischemic stroke categorically with little consistency in genetic research.15 Despite this methodological limitation, genetic risk factors have been identified that appear to be specific for certain ischemic stroke subtypes. For example, the chromosome 9p21 locus appears to impart risk for so-called large-vessel atherosclerotic stroke.16,17 The atrial fibrillation locus 4q25 appears to impart risk for cardioembolic stroke.18,19 Collaborating with investigators from Sweden, we previously assessed whether ischemic stroke subtypes clustered among affected sibships, showing low aggregation rates.20 We continue to see low aggregation rates, but the relation is significant. In addition to having larger numbers, the current analysis is restricted to those affected sibling pairs confirmed to be full siblings through genomic analysis.21

Age at onset of stroke may be a quantitative phenotype more tractable to genomic analysis. In an interim analysis, we had observed a significant association of proband age at stroke onset and sibling age at stroke onset.22 As with the subtype aggregation reanalysis, the current analysis involves a larger sample size and is restricted to those affected sibling pairs confirmed to be full siblings through genomic analysis. As a phenotype, age at stroke onset has the limitation that it

Table 3. Top 10 Most Significant SNPs Representing 8 Genomic Loci

<table>
<thead>
<tr>
<th>SNP</th>
<th>CHR</th>
<th>MB</th>
<th>cM</th>
<th>Minor Allele</th>
<th>MAF</th>
<th>Odds Ratio for Minor Allele</th>
<th>$\chi^2$ Statistic</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>rs1383407</td>
<td>3</td>
<td>78.991</td>
<td>107.18</td>
<td>C</td>
<td>0.4488</td>
<td>0.961</td>
<td>15.65</td>
<td>7.63×10⁻⁵</td>
</tr>
<tr>
<td>rs328049</td>
<td>3</td>
<td>79.068</td>
<td>107.21</td>
<td>A</td>
<td>0.4766</td>
<td>0.968</td>
<td>10.31</td>
<td>1.32×10⁻³</td>
</tr>
<tr>
<td>rs986692</td>
<td>3</td>
<td>107.766</td>
<td>116.1</td>
<td>T</td>
<td>0.4120</td>
<td>1.040</td>
<td>10.80</td>
<td>1.01×10⁻³</td>
</tr>
<tr>
<td>rs1053110</td>
<td>5</td>
<td>180.421</td>
<td>205.94</td>
<td>T</td>
<td>0.4704</td>
<td>1.034</td>
<td>9.92</td>
<td>1.64×10⁻³</td>
</tr>
<tr>
<td>rs1293457</td>
<td>6</td>
<td>44.866</td>
<td>68.46</td>
<td>T</td>
<td>0.4372</td>
<td>1.036</td>
<td>12.20</td>
<td>4.79×10⁻⁴</td>
</tr>
<tr>
<td>rs3778507</td>
<td>6</td>
<td>45.005</td>
<td>68.69</td>
<td>A</td>
<td>0.4075</td>
<td>1.043</td>
<td>14.53</td>
<td>1.38×10⁻⁴</td>
</tr>
<tr>
<td>rs179209</td>
<td>16</td>
<td>19.215</td>
<td>41.4</td>
<td>A</td>
<td>0.4749</td>
<td>0.967</td>
<td>9.773</td>
<td>1.77×10⁻³</td>
</tr>
<tr>
<td>rs750740</td>
<td>16</td>
<td>87.335</td>
<td>129.03</td>
<td>T</td>
<td>0.4695</td>
<td>1.037</td>
<td>12.26</td>
<td>4.63×10⁻⁴</td>
</tr>
<tr>
<td>rs897783</td>
<td>19</td>
<td>56.723</td>
<td>88.79</td>
<td>A</td>
<td>0.3779</td>
<td>1.036</td>
<td>10.36</td>
<td>1.29×10⁻³</td>
</tr>
<tr>
<td>rs976192</td>
<td>20</td>
<td>1.444</td>
<td>5.81</td>
<td>C</td>
<td>0.3797</td>
<td>1.041</td>
<td>10.35</td>
<td>1.30×10⁻³</td>
</tr>
</tbody>
</table>

SNP indicates single nucleotide polymorphism; CHR, chromosome; MB, megabase; cM, centiMorgan; MAF, minor allele frequency.
does not necessarily reflect the burden of ischemic disease at any given moment in the lifespan of a patient. Some cerebral infarcts are asymptomatic, whereas other cerebral infarcts may be symptomatic but undiagnosed.

In summary, we have described an affected relative–based genetic analysis of ischemic stroke. This work provides preliminary evidence for the involvement of several loci in risk for this disease, and these loci certainly warrant follow-up. This work also suggests that any individual risk variants involved in ischemic stroke are likely to have a low population-attributable risk. Attributable risk could be low if the risk conferred is relatively low; it could also be low if there is extensive allelic and/or genetic heterogeneity in stroke, with no single locus being a common, high risk–confling locus. Clearly we can hope that the future application of now- and next-generation technologies in large and extremely well-characterized cohorts will enable identifying genetic risks for ischemic stroke.

Appendix: SWISS Investigative Team

Executive Committee: James F. Meschia, chair; Thomas G. Brott, Robert D. Brown Jr, Brett Kissela, John Hardy, Stephen S. Rich, Andrew Singleton, Bradford Worrell. Statistical Center: University of Virginia, Charlottesville, VA; Stephen S. Rich, Director of the Center for Public Health Genomics. DNA Repository: Cor nell Institute for Medical Research, Camden, NJ; Margaret Keller, PhD, Director. Data Management: Mayo Clinic, Rochester, MN; Barry Bisbee, Data Management Specialist.


Acknowledgments
This study used the high-performance computational capabilities of the Biowulf Linux cluster at the National Institutes of Health, Bethesda, MD. This work used samples and clinical data from the National Institutes of Neurological Disorders and Stroke Human Genetics Resource Center DNA and Cell Line Repository (http://crr.coriell.org/ninds), human subjects protocol numbers 2003-081 and 2004-147.

Sources of Funding
SWISS is funded by the National Institute for Neurological Disorders and Stroke (NINDS) through grant number R01 NS39987 (J.F.M., PI), IRB 1082-99 (Mayo Clinic). The inclusion of BLSA samples was supported in part by the Intramural Research Program of the National Institute on Aging, National Institutes of Health, Department of Health and Human Services; project number Z01 AG000015-50, human subjects protocol number 2003-078. Dr Hardy is supported by an MRC Returning Scientist award. The AG000015-50, human subjects protocol number 2003-078. Drs Adams, NHLBI, and the Department of Health and Human Services; project number Z01 AG000015-50, human subjects protocol number 2003-081.

Disclosures
None.

References
Siblings With Ischemic Stroke Study: Results of a Genome-Wide Scan for Stroke Loci
James F. Meschia, Michael Nalls, Mar Matarin, Thomas G. Brott, Robert D. Brown, Jr, John Hardy, Brett Kissela, Stephen S. Rich, Andrew Singleton, Dena Hernandez, Luigi Ferrucci, Kerra Pearce, Margaret Keller, Bradford B. Worrall and for the Siblings With Ischemic Stroke Study Investigators

Stroke. published online September 22, 2011;
Stroke is published by the American Heart Association, 7272 Greenville Avenue, Dallas, TX 75231
Copyright © 2011 American Heart Association, Inc. All rights reserved.
Print ISSN: 0039-2499. Online ISSN: 1524-4628

The online version of this article, along with updated information and services, is located on the World Wide Web at:
http://stroke.ahajournals.org/content/early/2011/09/22/STROKEAHA.111.620484

Data Supplement (unedited) at:
http://stroke.ahajournals.org/content/suppl/2011/09/22/STROKEAHA.111.620484.DC1

Permissions: Requests for permissions to reproduce figures, tables, or portions of articles originally published in Stroke can be obtained via RightsLink, a service of the Copyright Clearance Center, not the Editorial Office. Once the online version of the published article for which permission is being requested is located, click Request Permissions in the middle column of the Web page under Services. Further information about this process is available in the Permissions and Rights Question and Answer document.

Reprints: Information about reprints can be found online at:
http://www.lww.com/reprints

Subscriptions: Information about subscribing to Stroke is online at:
http://stroke.ahajournals.org/subscriptions/
SUPPLEMENTAL MATERIAL

Supplemental description of the methods.

_Establishing Lymphoblastoid Cell Lines from Peripheral Blood_

DNA samples were obtained by peripheral vein phlebotomy, often by a home health agency. When this was not practical because of geographic or linguistic barriers (French Canadian centers), phlebotomy was arranged through the respective local centers.

Two tubes of peripheral blood were shipped to Coriell Institute for Medical Research (Camden, NJ) for processing. Blood from one Vacutainer was used to isolate mononuclear cells for Epstein-Barr virus (EBV) transformation by standard methods. A small volume (0.5 ml) of blood was retained for identity tracking of resulting cell cultures and DNA by microsatellite analysis. Peripheral blood from a second Vacutainer was processed to generate mononuclear cells. The cells were re-suspended in 4 ml of freeze medium (RPMI 1640 with HEPES and 30% (v/v) fetal bovine serum and 6% dimethylsulfoxide) and distributed among four glass ampules. In rare cases where the establishment of a lymphoblastoid line from freshly-prepared peripheral blood mononuclear cells was unsuccessful, the cryopreserved mononuclear cells were retrieved from liquid nitrogen. The cells were thawed, washed, re-suspended in 2 ml of cell culture medium and added to a flask with 0.3 ml of PHA and 1 ml of EBV stock. Further handling was the same as for EBV-transformation from freshly prepared peripheral blood mononuclear cells.

_DNA Isolation from Lymphoblastoid Cell Lines_

An ampule of each lymphoblastoid cell line was recovered and the cells cultured to yield approximately 40 million cells. Cells were harvested and processed for isolation of genomic DNA using the Gentra Autopure robotic system.

_Quality Control of Genomic DNA_

DNA was allowed to solubilize over several days. UV spectrophotometry of each DNA sample was required to show a 260/280 nm absorbance ratio between 1.65 and 1.95 and a
concentration of at least 0.1 mg/ml with less than 0.1 µg protein per µg of DNA to be included in the study. DNA sample identity was confirmed by comparing a DNA fingerprint between DNA isolated directly from the peripheral blood to the DNA isolated from the lymphoblastoid cell line. DNA fingerprint analysis was performed using a multiplex PCR assay for 6 microsatellite markers: THO-1, D5S592, D10S526, vWA31, D22S417 and FES/FPS [S1]. Gender was determined in the same reaction using an additional primer pair designed to amplify a region of allelic difference between the X and Y chromosome amelogenin genes [S2]. Reactions were analyzed using an ABI 3730 DNA Analyzer. Each of these markers has sufficiently high heterozygosity (0.77, 0.83, 0.84, 0.82, 0.85 and 0.67, respectively) such that the probability of obtaining identical “profiles” from unrelated individuals is approximately 1 in 33 million. The gender determination was compared to the gender reported by the submitter of the specimen.

Study design for genotyping and genetic data acquisition

The two separate genotyping phases of SWISS were designed to conduct a genome-wide analysis of affected sibling pairs and, at the same time, identify unrelated participants for inclusion into a genome-wide association study. This division is due to the use and availability of data from separate genotyping arrays, first a large-scale dense coverage array utilized in genome-wide association studies and second, a less dense but compatible array (coverage of a majority of SNPs overlap across platforms) designed for use in genome-wide linkage mapping studies within families. The first phase of the genotyping was performed on DNA samples from probands using Human 610-Quadv1.0 DNA analysis beadchip (Illumina Inc, San Diego, CA). This density of SNPs across the genome also permitted incorporation of these data into analyses of family-based genome-wide association (the “genome-wide association phase”). The second phase of genotyping was performed on the DNA samples from the siblings of the probands using Illumina linkage V chip (the “genome-wide family-based phase”). After independent quality control of the two phases, the two datasets (genome-wide association and family-based) were merged to complete the final SWISS dataset. Imputation was used to
incorporate SNPs into the final dataset, allowing all SNPs passing quality control on the linkage array to be utilized.

Genotyping methods and quality control metrics – genome-wide association phase

DNA samples were assayed on the Human 610-Quadv1.0 DNA analysis beadchip (Illumina Inc, San Diego, CA) according to the Infinium HD Super assay protocol (Rev B, 2009, Illumina). Following the genotyping of probands on genome-wide arrays, standard quality control was applied. Illumina Beadstudio was used for genotype calling and clustering. After the initial genotype calling, inclusion thresholds for samples included call rate per sample > 0.95, call rate per SNP > 0.95, MAF > 0.05, and HWE p-value > 1x10^-7. Sex concordance based on X chromosome heterogeneity estimates were compared to self-reported data. Case data from this phase were then combined with a control set obtained from neurologically normal controls from the Coriell/NINDS repository and stroke-free participants in the Baltimore Longitudinal Study of Aging (BLSA). [S3]

After merging the probands genotyped in this phase with the control sets, the combined dataset was re-filtered on consensus SNPs with the following criteria: call rate per sample > 0.95, call rate per SNP > 0.95, MAF > 0.05, HWE P > 1x10^-5 in controls, and nonrandom missingness by haplotype and phenotype (P > 1x10^-5). Identity-by-state clustering was performed to remove cryptic related samples, removing all samples with pairwise pi_hat > 0.125; this effectively removes up to cousin-level related individuals. Study data were then merged with HapMap3 data and consensus SNPs were pruned based upon estimated linkage disequilibrium. Multidimensional scaling was conducted to remove samples > ± 6 SD from mean estimates for the first two component vectors from the mean of the combined CEU/TSI estimates.

Participants passing quality control from this phase had additional SNPs imputed using the 1000 Genomes Project reference haplotypes (released in June 2010). These data were obtained from the Abecasis MACH website (http://www.sph.umich.edu/csg/abecasis/MaCH/download/). A 2-stage imputation was performed using MACH v 1.0.16. [S4] [S5]. SNPs were included with a
minimum imputation quality indicated by MACH of RSQR ≥ 0.50. Imputed SNPs and genotyped SNPs from the probands were merged with the family-based phase (sibling) genotypes to create the final dataset of 555 individuals (across 223 sibships).

**Genotyping methods and quality control metrics – genome-wide family-based phase**

Illumina Beadstudio was used for genotype calling and clustering following the manufacturer’s protocol for the Linkage V array. The following filters were initially applied: call rate per sample > 0.95, call rate per SNP > 0.95, MAF > 0.05, HWE P > 1x10^{-7}, nonrandom missingness by haplotype and phenotype (P > 1x10^{-5}), and sex concordance based on X chromosome heterogeneity estimates compared to self reported data. Multi-dimensional scaling and identity-by-state clustering were utilized to remove ancestry outliers and cryptically related samples as previously described.
SUPPLEMENTAL REFERENCES


Development and applications of human diversity panels. American Society for 

S.2. Bender, P, Hansen, M.A., and Beck, J.C., Gender determination by amelogenin 
34: 29A.

S.3. Failure to validate association between 12p13 variants and ischemic stroke. N 


S.5. Li, YaA, G.R., Mach 1.0: Rapid Haplotype Reconstruction and Missing Genotype 