Stroke Genetics Network (SiGN) Study

Design and Rationale for a Genome-Wide Association Study of Ischemic Stroke Subtypes

James F. Meschia, MD; Donna K. Arnett, PhD; Hakan Ay, MD; Robert D. Brown Jr, MD; Oscar R. Benavente, MD; John W. Cole, MD, MS; Paul I.W. de Bakker, PhD; Martin Dichgans, MD; Kimberly F. Doheny, PhD; Myriam Fornage, PhD; Raji P. Grewal, MD; Katrina Gwinn, MD; Christina Jern, MD; Jordi Jimenez Conde, MD, PhD; Julie A. Johnson, PharmD; Katarina Jood, MD; Cathy C. Laurie, PhD; Jin-Moo Lee, MD, PhD; Arne Lindgren, MD; Hugh S. Markus, FRCP; Patrick F. McArdle, PhD; Leslie A. McClure, PhD; Braxton D. Mitchell, PhD; Reinhold Schmidt, MD; Kathryn M. Rexrode, MD; Stephen S. Rich, PhD; Jonathan Rosand, MD, MSc; Peter M. Rothwell, MD; Tatjana Rundek, MD; Ralph L. Sacco, MD; Pankaj Sharma, MD; Alan R. Shuldiner, MD; Agnieszka Slowik, MD; Sylvia Wasserteil-Smoller, PhD; Cathie Sudlow, MD; Vincent N.S. Thijs, MD; Daniel Woo, MD; Bradford B. Worrall, MD, MSc; Ona Wu, PhD; Steven J. Kittner, MD; on behalf of the NINDS SiGN Study

Background and Purpose—Meta-analyses of extant genome-wide data illustrate the need to focus on subtypes of ischemic stroke for gene discovery. The National Institute of Neurological Disorders and Stroke SiGN (Stroke Genetics Network) contributes substantially to meta-analyses that focus on specific subtypes of stroke.

Methods—The National Institute of Neurological Disorders and Stroke SiGN includes ischemic stroke cases from 24 genetic research centers: 13 from the United States and 11 from Europe. Investigators harmonize ischemic stroke phenotyping using the Web-based causative classification of stroke system, with data entered by trained and certified adjudicators at participating genetic research centers. Through the Center for Inherited Diseases Research, the Network plans to genotype 10296 carefully phenotyped stroke cases using genome-wide single nucleotide polymorphism arrays and adds to these another 4253 previously genotyped cases, for a total of 14549 cases. To maximize power for subtype analyses, the study allocates genotyping resources almost exclusively to cases. Publicly available studies provide most of the control genotypes. Center for Inherited Diseases Research—generated genotypes and corresponding phenotypes will be shared with the scientific community through the US National Center for Biotechnology Information database of Genotypes and Phenotypes, and brain MRI studies will be centrally archived.

Conclusions—The Stroke Genetics Network, with its emphasis on careful and standardized phenotyping of ischemic stroke and stroke subtypes, provides an unprecedented opportunity to uncover genetic determinants of ischemic stroke. (Stroke. 2013;44:2694-2702.)

Key Words: cerebral infarct ▪ genetics ▪ genomics

Received May 2, 2013; accepted July 3, 2013.

From the Mayo Clinic Jacksonville, FL (J.F.M.); University of Alabama at Birmingham (D.K.A., L.A.M.); Mayo Clinic Rochester, MN (R.D.B.); Massachusetts General Hospital, Boston, MA (H.A., J.R., O.W.); University of British Columbia, Vancouver, British Columbia, Canada (O.R.B.); University of Maryland School of Medicine and Veterans Administration Medical Center, Baltimore, MD (J.W.C., P.F.M., B.D.M., A.R.S., J.K.); University Medical Center Utrecht, Utrecht, The Netherlands (P.I.W.d.B.); Brigham and Women’s Hospital, Harvard Medical School, Boston, MA (P.I.W.d.B.); Broad Institute of Harvard, MIT, Cambridge, MA (P.I.W.d.B.); Klinikum der Universität München, Ludwig-Maximilians-Universität, Munich, Germany (M.D.); Johns Hopkins University, Baltimore, MD (K.F.D.); University of Texas Health Science Center at Houston (M.F.D.); Saint Francis Medical Center, Lynwood, CA (R.P.G.); NINDS Neurogenetics Cluster, Bethesda, MD (K.G.); IMIM-Hospital Universitari del Mar, Barcelona, Spain (J.J.C.); University of Florida, College of Pharmacy, Gainesville, FL (J.A.J.); Institute of Neuroscience and Physiology, the Sahlgrenska Academy at University of Gothenburg, Sweden (K.J., C.J.); University of Washington, Seattle, WA (C.C.L.); Washington University School of Medicine, St. Louis, MO (J.-M.L.); Lund University, Lund, Sweden (A.L.); St. George’s University of London, London, United Kingdom (H.S.M.); Medical University Graz, Graz, Austria (R.S.); Brigham and Women’s Hospital, Boston, MA (K.M.R.); University of Virginia, Charlottesville, VA (S.R., B.B.W.); Radcliffe Infirmary, Oxford, England (P.M.R.); University of Miami. Coral Gables, FL (T.R., R.L.S.); Imperial College London, London, United Kingdom (P.S.); Jagiellonian University, Krakow, Poland (A.S.); Albert Einstein College of Medicine, Bronx, NY (S.W.—S.); University of Edinburgh, Edinburgh, United Kingdom (C.S.); Université Ziekenhuizen Leuven, Leuven, Belgium (V.N.S.T.); and University of Cincinnati, Cincinnati, OH (D.W.).

Guest Editor for this article was Steven C. Cramer, MD, MMSc.

The online-only Data Supplement is available with this article at http://stroke.ahajournals.orglookup/suppl/doi:10.1161/STROKEAHA.113.001857/-/DC1.

Correspondence to Steven Kittner, MD, Department of Neurology, University of Maryland School of Medicine, Baltimore, MD 21201. E-mail skittner@maryland.edu

© 2013 American Heart Association, Inc.

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

DOI: 10.1161/STROKEAHA.113.001857
See related articles, p 2679 and e120.

Genome-wide association studies have been remarkably successful in identifying loci contributing to the genetic basis of human disease and complex phenotypes. Just >5 years ago, knowledge of the genetic variants influencing disease risk was largely restricted to rare familial conditions that could be linked to rare mutations in single genes with high penetrance. Although heritability and family studies consistently pointed to a substantial genetic contribution to common complex conditions, such as type 2 diabetes mellitus, coronary heart disease, and ischemic stroke, the genetic loci accounting for a substantial component of risk for these disorders remained almost completely undiscovered. By applying technologies that allow genotyping of hundreds of thousands of variants across the genome in thousands of individuals at high accuracy, investigators have subsequently discovered many loci (and in some cases, the underlying causal genes within a locus) contributing to risk of common diseases (http://genome.gov/gwastudies/). The genotyping arrays were constructed using common variants for capturing regions of genomic variation, and variants in a locus that significantly associated with disease risk were common in the population and exerted relatively small effects. Thus, thousands of well-phenotyped individuals were required to identify these risk loci.

Founded in 2007, the International Stroke Genetics Consortium facilitates assembly of genome-wide data in thousands of cases, controls, and families with ischemic stroke for the purpose of collaborative meta-analyses. Although there have been no consistently replicated loci associated with ischemic stroke, the International Stroke Genetics Consortium has identified several loci associated with ischemic stroke subtypes. Variants on chromosome 9p21, 6p21.1, and near the HDAC9 gene have been related to large-vessel atherosclerotic stroke. Genetic variants associated with atrial fibrillation, a condition strongly predisposing to ischemic stroke, have also been identified by linkage and sequencing studies, including mutations in several ion channels a locus on chromosome 4 was identified through genome-wide association study (GWAS). It was quickly realized that much larger sample sizes would be required to detect further risk loci for ischemic stroke overall and stroke subtypes. In addition, the underlying heterogeneous cause of ischemic stroke (including small-vessel, cardioembolic, and large-vessel mechanisms) suggests that careful and systematic phenotyping and subtype-specific analyses are essential for successful gene discovery.

The success of subtype-specific genetic studies of ischemic stroke faces another substantial hurdle: the lack of agreement across centers on subtype assignment. To address these challenges, the US National Institute of Neurological Disorders and Stroke (NINDS) has established the Stroke Genetics Network (SiGN) with the goal of assembling the largest possible sample of individuals with ischemic stroke for genetic studies, where each individual has been uniformly and thoroughly characterized for stroke subtyping. Similar to the earlier Wellcome Trust Case Control Consortium-2 GWAS of ischemic stroke, SiGN has grown out of the International Stroke Genetics Consortium and is committed to the widest possible sharing of data among all investigators dedicated to discovering the role of genetic variation in risk of ischemic and hemorrhagic stroke and related phenotypes and exploiting this knowledge for the benefit of patients. This article describes and explains the rationale for key aspects of the design of SiGN.

Methods
SiGN responded to a request for applications issued by the NINDS that proposed establishing a GWAS consortium focused on identifying genes or genomic regions that affect either the susceptibility to or outcome of ischemic stroke. The request for applications specified that multiple genetic research centers (GRC) be established that have access to well-characterized ischemic stroke cases in whom extensive phenotype, covariate, and exposure data are available and high-quality DNA are banked or could be isolated from stored specimens and that standardized, validated, and easily replicated methods should be used to assign stroke subtypes. The request for applications further specified that investigators submit the harmonized phenotype data used for stroke subtyping and the newly generated genotype data to the National Center of Biotechnology Information supported database of Genotypes and Phenotypes (dbGaP) to create a national resource of high-quality information for data mining, replication studies, and future hypothesis generation.

Structure of SiGN
SiGN consists of 24 GRCs: 13 from the United States and 11 from Europe (Table I for summary and online-only Data Supplement I for descriptions). The GRCs represent centers that have existing collections of DNA samples from ischemic stroke cases and agree to characterize all cases for stroke subtype using a single standardized protocol requiring detailed imaging and clinical information. Informed consent for data sharing is a requirement for a GRC. The figure shows the administrative structure. The Scientific Steering Committee leads SiGN. Its members include co-principal investigators, the Analysis Committee, and NINDS staff. The Scientific Steering Committee is responsible for scientific direction and policy decisions. It also oversees the Publications and Data Access Committee, which develops guidelines for publication and authorship, prioritizes analysis resources for manuscript proposals, and recommends approval of proposals and manuscripts to the Scientific Steering Committee. The study has 4 cores: Administrative, Data Management, Imaging, and Genotyping. The Administrative Core and Data Management Core monitor study progress, maintain efficient interactions among the cores and the participating GRCs, ensure regulatory compliance, and are responsible for submitting the genotype and phenotype data to dbGaP. The Data Management Core also works closely with the Analysis Committee in the preparation of publications. The Analysis Committee, composed of genetic epidemiologists and statistical geneticists from 4 different institutions, advises the Scientific Steering Committee on design issues and is responsible for the genetic analyses (online-only Data Supplement II). The Phenotype Committee, detailed below, is responsible for training and quality assurance of ischemic stroke subtyping at the GRCs. The Imaging Core, detailed below, is the centralized repository for clinically obtained MRI data from the GRCs. The Genotyping Core is the NINDS-designated Center for Inherited Disease Research (CIDR, Baltimore, MD). The Genotyping Core performs quality control of the submitted DNA, as well as initial quality control of the GWAS and exome-enriched genotyping. The Center for Biomedical Statistics (CBS) at the University of Washington (Seattle) provides more extensive quality control of the genotype data through a subcontract with CIDR, CBS, and the Analysis Committee jointly decided on the design of the study, including choice of controls, and selection and use of within- and cross-study duplicates.

Phenotyping Methods
SiGN uses the causative classification of stroke (CCS) system for phenotyping of ischemic stroke cases. CCS incorporates multiple aspects of present-day diagnostic stroke evaluation (diffusion-weighted imaging, perfusion-weighted imaging, computed tomography and
Table 1. Cases of Ischemic Stroke (N=17,298) Classified Using the Causative Classification of Stroke (CCS) System as of April 3, 2013, and Their Demographic Characteristics for Each Genetic Research Center (GRC)

<table>
<thead>
<tr>
<th>GRC</th>
<th>Location</th>
<th>Recruitment Source</th>
<th>Recruitment Years</th>
<th>Cases (n)</th>
<th>Age Range, y</th>
<th>Female, %</th>
<th>European Descent, %</th>
<th>African Descent, %</th>
</tr>
</thead>
<tbody>
<tr>
<td>BASICMAR</td>
<td>Barcelona, Spain</td>
<td>Hospital based</td>
<td>2005–2012</td>
<td>1088</td>
<td>30–101</td>
<td>47</td>
<td>97</td>
<td>0</td>
</tr>
<tr>
<td>EDINBURGH</td>
<td>Edinburgh, Scotland</td>
<td>Hospital based</td>
<td>2002–2005</td>
<td>626</td>
<td>29–97</td>
<td>45</td>
<td>100</td>
<td>0</td>
</tr>
<tr>
<td>GASROS</td>
<td>Boston, United States</td>
<td>Hospital based</td>
<td>2003–2009</td>
<td>686</td>
<td>18–100</td>
<td>36</td>
<td>90</td>
<td>4</td>
</tr>
<tr>
<td>GCNKSS</td>
<td>Greater Cincinnati region, United States</td>
<td>Population based</td>
<td>1999–2006</td>
<td>642</td>
<td>20–104</td>
<td>50</td>
<td>75</td>
<td>24</td>
</tr>
<tr>
<td>GRAZ</td>
<td>Graz, Austria</td>
<td>Hospital based</td>
<td>1992–2011</td>
<td>685</td>
<td>19–101</td>
<td>41</td>
<td>100</td>
<td>0</td>
</tr>
<tr>
<td>ISGS</td>
<td>Multicenter, United States</td>
<td>Hospital based</td>
<td>2002–2008</td>
<td>675</td>
<td>19–94</td>
<td>43</td>
<td>71</td>
<td>26</td>
</tr>
<tr>
<td>KRAKOW</td>
<td>Krakow, Poland</td>
<td>Hospital based</td>
<td>2001–2011</td>
<td>1487</td>
<td>19–100</td>
<td>48</td>
<td>100</td>
<td>0</td>
</tr>
<tr>
<td>LEUVEN</td>
<td>Leuven, Belgium</td>
<td>Hospital based</td>
<td>2005–2009</td>
<td>524</td>
<td>18–97</td>
<td>42</td>
<td>100</td>
<td>0</td>
</tr>
<tr>
<td>LUND</td>
<td>Lund, Sweden</td>
<td>Hospital based</td>
<td>2006–2010</td>
<td>818</td>
<td>22–99</td>
<td>49</td>
<td>100</td>
<td>0</td>
</tr>
<tr>
<td>MCISS</td>
<td>New Jersey, United States</td>
<td>Hospital based</td>
<td>1999–2009</td>
<td>876</td>
<td>19–98</td>
<td>49</td>
<td>68</td>
<td>12</td>
</tr>
<tr>
<td>MIAMISR</td>
<td>Miami, United States</td>
<td>Hospital based</td>
<td>2008–2011</td>
<td>331</td>
<td>18–92</td>
<td>36</td>
<td>15</td>
<td>29</td>
</tr>
<tr>
<td>MUNICH*</td>
<td>Munich, Germany</td>
<td>Hospital based and cohort (2 sources)</td>
<td>2002–2009</td>
<td>524</td>
<td>17–97</td>
<td>41</td>
<td>100</td>
<td>0</td>
</tr>
<tr>
<td>NHS</td>
<td>National sample, United States</td>
<td>Hospital based and cohort study</td>
<td>1989–1992</td>
<td>470</td>
<td>45–85</td>
<td>100</td>
<td>93</td>
<td>1</td>
</tr>
<tr>
<td>NOMAS(S)</td>
<td>Manhattan, United States</td>
<td>Population based and cohort (2 sources)</td>
<td>1993–2001</td>
<td>578</td>
<td>33–104</td>
<td>55</td>
<td>20</td>
<td>25</td>
</tr>
<tr>
<td>OXVASC</td>
<td>Oxfordshire, England</td>
<td>Population based</td>
<td>2002–2010</td>
<td>554</td>
<td>33–96</td>
<td>51</td>
<td>100</td>
<td>0</td>
</tr>
<tr>
<td>REGARDS</td>
<td>National sample, United States</td>
<td>Cohort study</td>
<td>2003–2007</td>
<td>555</td>
<td>46–93</td>
<td>46</td>
<td>57</td>
<td>43</td>
</tr>
<tr>
<td>SAHLSIS</td>
<td>Gothenburg, Sweden</td>
<td>Hospital based</td>
<td>1998–2012</td>
<td>1085</td>
<td>16–69</td>
<td>36</td>
<td>100</td>
<td>0</td>
</tr>
<tr>
<td>SPS3*</td>
<td>Multicenter; United States, Latin America, Spain</td>
<td>Hospital based</td>
<td>2003–2011</td>
<td>1139</td>
<td>32–89</td>
<td>37</td>
<td>47</td>
<td>14</td>
</tr>
<tr>
<td>ST GEORGE’S</td>
<td>London, England</td>
<td>Hospital based</td>
<td>1995–2008</td>
<td>684</td>
<td>18–102</td>
<td>47</td>
<td>100</td>
<td>0</td>
</tr>
<tr>
<td>SWISS</td>
<td>Multicenter, United States</td>
<td>Hospital based and cohort study (2 sources)</td>
<td>1999–2011</td>
<td>407</td>
<td>21–93</td>
<td>47</td>
<td>91</td>
<td>2</td>
</tr>
<tr>
<td>WHI</td>
<td>National sample, United States</td>
<td>Hospital based</td>
<td>1993–1998</td>
<td>840</td>
<td>54–87</td>
<td>100</td>
<td>86</td>
<td>8</td>
</tr>
<tr>
<td>WUSTL*</td>
<td>St. Louis, United States</td>
<td>Hospital based</td>
<td>2008–2012</td>
<td>535</td>
<td>20–90</td>
<td>42</td>
<td>47</td>
<td>28</td>
</tr>
</tbody>
</table>

The proportion Hispanics by self-identification was as follows: BASICMAR (0%), BRAINS (0%), EDINBURGH (0%), GASROS (4.1%), GCNKSS (0.3%), GEOS (2.4%), SAHLSIS (0%), GRAZ (0.2%), ISGS (1.2%), KRAKOW (0%), LEUVEN (0%), LUND (0%), MC ISS (6.5%), MIAMISR (54.1%), MUNICH (0%), NHS (1.3%), NOMAS(S) (52.9%), OX VASC (0%), REGARDS (0%), SPS3 (45%), ST. GEORGE’S (0%), SWISS (1.5%), WHI (2.0%), WUSTL (0%). GASROS indicates Genes Affecting Stroke Risks and Outcomes Study; GRC, Genetic Research Center; ISGS, Ischemic Stroke Genetics Study; NHS, Nurses’ Health Study; NOMAS, Northern Manhattan Study; OXVASC, Oxford Vascular Study; REGARDS, Reasons for Geographic and Racial Differences in Stroke Study; SWISS, Sibling with Ischemic Stroke Study; and WHI, Women’s Health Initiative. *Cases being phenotyped.
Figure. Organizational structure of Stroke Genetics Network (SiGN) study. CIDR indicates Center for Inherited Diseases Research; and GRC, genetic research center.

MR angiography of extracranial and intracranial arteries, transthoracic and transesophageal echocardiography, and ambulatory electrocardiography) in a standardized manner to identify both likely causative and phenotypic subtypes. There is Web-based, semiautomated CCS software to assign the most likely causative mechanism. The CCS divides ischemic stroke into 5 causative subtypes based on a framework that is well defined, easily replicable, and evidence based: supra-aortic large artery atherosclerosis, cardioaortic embolism, small artery occlusion, other uncommon causes, and undetermined causes. The system permits distinguishing patients with symptomatic intracranial atherosclerosis from patients with symptomatic extracranial atherosclerosis. The Web-based CCS allows for remote data entry, as well as structuring and archiving of individual data elements, such as diagnostic test findings. In an international multicenter study, a high degree of reliability (k statistic, 0.80) was found among 20 raters from 13 centers in 8 countries when applying the Web-based CCS to the same set of 50 consecutive abstracted case summaries.

With the exception of ST GEORGE’S, BASICMAR (early cases), and the Secondary Prevention of Small Subcortical Strokes (SPS3) trial, physician adjudicators from each GRC adjudicate clinical histories, physical examination findings, and the results of diagnostic testing and enter the information into the Web-based CCS system. Every adjudicator is required to have undergone formal training and certification in use of the CCS. The CCS Web site contains an interactive training module that has 10 training cases. A Phenotype Committee trainer provides training lectures to adjudicators at scheduled study meetings. Finally, the Committee presents a 90-minute webinar for additional training. The webinar reviews case numbering conventions, data entry, data submission, and archiving, as well as standardized consensus responses to frequently asked questions about specific CCS items. All GRC adjudicators and members of the Phenotype Committee are required to take and pass an on-line CCS certification examination. The CCS certification examination consists of 5 clinical vignettes (randomly selected from a pool of 15 vignettes) from which the test taker abstracts and enters data into Web forms. The CCS assigns weights to test items based on their importance in determining subtype diagnosis. The total score is on a 40 to 100 point scale. Points are deducted when critical data elements are missed or non-existing data elements are substituted. The minimum passing score is 80 points. The Phenotype Committee allows ≤5 attempts by test takers to pass. Individuals who achieve certification in CCS receive online confirmation of having passed. The GRC and the Phenotype Committee administrators retain copies of the certificate.

ST GEORGE’S entered data into the publically available version of CCS. For the first set of cases classified by BASICMAR, investigators at BASICMAR mapped precollected stroke research data that had been stored in an electronic database to the study-specific CCS. From the SPS3 trial, precollected clinical trial data captured on case report forms and stored electronically are being mapped to the CCS using decision rules authored by the Phenotype Committee in collaboration with the principal investigator of the trial.

The Phenotype Committee tracks progress in CCS adjudication center by center in a weekly conference call. The Committee also monitors data quality by assessing inter-rater reliability of case adjudication. An independent 10% random sample of cases is rejudicated for each GRC. For US centers, vascular neurologist members of the Phenotype Committee rejudicate cases. For non-US centers, a CCS-certified member of the local investigative team rejudicates cases. Raters perform all rejudications blinded to the results of the initial adjudication. For any given case, the adjudicator and the re adjudicator are different individuals. When CCS reliability results fall <50% complete agreement, the Phenotype Committee reviews which aspects of the CCS adjudication seem to be most problematic for the adjudicator, engages in retraining, and requires rejudication at a center.

The SiGN Imaging Platform

A distinctive feature of SiGN is that the imaging platform assembles in a central location all available brain MR images obtained at the time of or during follow-up after stroke for genotyped subjects. Deidentified images have been stored on a central server. The goal is to have this resource used in future investigations by members of SiGN, the International Stroke Genetics Consortium, or other investigators to advance understanding of the role of genetic variation in stroke and MRI-derived phenotypes. The imaging platform is based on the Extensible Neuroimaging Archive Toolkit, an open-source software specifically designed to facilitate common management and productivity tasks for neuroimaging and associated metadata, such as image capture, quality control, automation, local use, collaborative use, and public access. The Imaging Core has integrated Extensible Neuroimaging Archive Toolkit with a production-ready, open-source content management framework called Plone (http://www.plone.org), which provides an easily customizable front end and a streamlined interface for imaging and clinical data management. Data dictionaries from each of the sites, consisting of SiGN ID, sex, race, ethnicity, age, and infarct location, are uploaded along with imaging data. Images can be viewed on Web browsers using a Java-based image viewer. Search query capabilities are provided with a concise interface similar to the
advanced search feature on PubMed. Users are able to query the data stored in the imaging repository, including metadata in the data dictionary, using field-based key words. For example, a site can find all female patients aged <65 years for which imaging exists in the repository. Search results are available for export into an Excel-compatible file. The file can then be used to request detailed genetic information and raw images.

**Overall Genotyping and Analysis Strategy**

The support from NINDS to the SiGN study allowed 11,644 samples to be genotyped at CIDR using the Illumina Infinium Omni5 genotype array with exome content. To maximize power to detect associations with stroke subtypes, a strategic decision was made to genotype primarily ischemic stroke cases for comparison with publicly available previously genotyped controls where possible. Cases from participating GRCs were prioritized for genotyping based on CCS subtyping: (1) cases with a determined CCS subtype excluding certain known rare causes (migraine-related stroke, acute arterial dissection, dilated cardiomyopathy, infective endocarditis, papillary fibroelastoma, left atrial myxoma, cerebral venous thrombosis, acute disseminated intravascular coagulation, drug-induced, heparin-induced thrombocytopenia type II, cerebral autosomal-dominant arteriopathy with subcortical infarcts and leukoencephalopathy, iatrogenic causes, mitochondrial encephalopathy with lactic acidosis and stroke-like episodes, meningitis, primary infection of the arterial wall, and sickle cell disease), and (2) cryptogenic CCS subtype, despite adequate evaluation. Additional cases were prioritized independent of CCS subtyping based on the availability of specific other desirable phenotypic information, including digital MRI data and longitudinal outcome data. The total number of CCS-phenotyped cases in SiGN is 16,411, which were contributed by 24 sites in the United States and Europe. A total of 10,236 cases have been prioritized for genotyping. A total of 4,253 additional CCS-phenotyped cases from 10 GRCs across Europe and the United States have been previously genotyped before initiating SiGN (BRAINS, EDINBURGH, GASROS, GEOS, ISGS, Nurses’ Health Study [NHS], MUNICH, OXVASC, ST GEORGE’S, and SWISS). A total of 14,549 cases will have genotypes from an Illumina platform, with ≥610,000 single nucleotide polymorphism (SNP) genome-wide coverage available for analysis.

**Selection of Control Subjects for SiGN**

Where possible, controls with publicly available genotype data were selected to ancestry-match cases at each GRC. A key criterion for selection of these control groups was that they had been genotyped on an Illumina Omni series GWAS array to minimize technical artifacts between cases and controls. Table 2 summarizes the key features of the control groups selected for each GRC. For the US samples, we identified 2 large population studies that had been genotyped previously to serve as control groups: (1) the Health and Retirement Study (HRS), a nationally representative sample of ≥2200 adults ≥50 years of age launched in 1992 to provide information about health and social issues relating to retirement (n=12,507 genotyped subjects); (2) the Osteoarthritis Initiative (OAI), a prospective study of ≥5000 adults, with the primary objective of identifying risk factors for incidence and progression of tibiofemoral knee osteoarthritis (http://oai.epi-ucsf.org/dataload/ default.asp; n=4130 genotyped subjects). HRS and OAI used the Illumina Omni Quad 2.5M array. These studies, which include European white, black, and Hispanic ethnicities, were selected to provide controls for all stroke cases to be genotyped through the 13 US GRCs. The GEOS GRC had previously genotyped patients and controls.

SiGN has identified separate control groups for the 11 European GRCs. These include previously genotyped control groups from Sweden, Spain, the UK, and Austria to be paired with the cases to be genotyped from the 2 GRCs from Sweden (LUND and SAHLSIS) and 2 from the GRCs from Spain (BASICMAR), the UK (BRAMS), and Austria (GRAZ). Because suitable control groups with available GWAS genotyping could not be identified from Poland and Belgium, controls from the KRAKOW and LEUVEN GRCs will also be genotyped as part of SiGN.

**Genotyping Methods**

The Illumina Infinium Omni5 genotype array with exome content has been selected as the genotyping platform in consultation with NINDS and CIDR. This array includes ≥4.3 million SNPs across the genome, with excellent coverage of common and infrequent variants (minor allele frequency, >1%). This array also includes ≥240,000 rare but polymorphic variants selected from >1200 individually sequenced exomes and 475 mitochondrial markers. CIDR will perform all genotyping. The 11,644 samples allocated for genotyping will include ischemic stroke cases (n=10,296), controls from the KRAKOW and LEUVEN sites (n=1282), and 66 additional samples selected for quality control (see below). In the total sample of cases, there are slightly more women, because 2 GRCs enrolled only women (Women’s Health Initiative and NHS). All GRCs provide subjects of European ancestry; in addition, 9 of the US sites provide black subjects and 5 provide Hispanic subjects. There are no whole-genome amplified DNA samples used in the genotyping process.

Three types of samples will be genotyped for quality control purposes: (1) cross-study duplicates (previously genotyped control samples regenotyped by CIDR to assess genotype concordance rates and identify SNPs that perform differently); (2) a 2% sample of duplicates from each GRC; and (3) HapMap controls. HRS, OAI, GRAZ, BASICMAR, and LUND will each provide 30 duplicates from their control groups for replicate genotyping.

For allocation to genotyping plates, DNA samples sent for genotyping at CIDR will be separated into 21 groups, based on GRC (19 sites that contribute ischemic stroke DNA samples) and disease status (2 sites contributing control DNA samples). DNA samples will be randomized within these levels and genotyped in 48-sample batches. Each batch will contain 1 HapMap control and 1 study duplicate. The duplicates will be randomly selected and be representative of the overall sample distribution. The 30 cross-study duplicates contributed by the 5 control groups will be distributed evenly across plates.

Before genotyping with the Illumina Infinium Omni5 genotype array with exome content, all samples will be genotyped with a 96-SNP barcode panel composed of autosomal, X- and Y-chromosome markers. The pretesting process allows for assurance of proper sample tracking and file tracking throughout data generation and release processes (concordance between pretesting genotypes and genotypes generated from 5M plus exome array data; confirmation of expected relationships and duplicates; and identification of file creation and aliquoting errors, primarily sex discrepancies and unexpected first-degree relatives among subjects). DNA aliquots that perform unexpectedly or result in poor data quality with the pretesting assay will be flagged for possible replacement or removal from the study. After the final sample set is determined, the GWAS assay will be performed. Poorly performing samples, usually those with a call rate <98%, will be genotyped a second time. In the GWAS processing of data, genotype cluster definitions will be determined using the Illumina Gentrain algorithm version 1.0 contained in Illumina GenomewStudio software (Illumina, Inc, San Diego, CA). We initially use this software to determine cluster boundaries, including all samples for a project. Sample call rates and quality metrics will be evaluated. From previous CIDR experience, it is anticipated that a small portion of samples will be marked for exclusion from project release as a result of poor data quality (call rate generally <97%–98% for genomic DNAs). After exclusion of poor quality experiments, the clustering algorithm will be run again to determine final cluster positions, because it is important to include only high-quality raw data for accurate clustering. Any genotype with a quality value <0.15 will not be provided for analysis. Genotype cluster boundaries will be manually reviewed for all XY, Y, and mitochondrial SNPs and adjusted as necessary. Additional SNP filtering will be performed, with the goal being to remove genotypes only for markers that are complete assay failure.

Phenotype data and corresponding genotypes generated through SiGN will be made available on dbGaP. For dbGaP posting purposes,
data for all SNPs will be provided for all samples that pass quality control at CIDR and for which no sample identity issues arise during quality control. The released data sets will include the raw data files (.idat files); genotypes for forward, A/B, design, and top alleles; quality scores and intensity values (raw and normalized); SNP and sample summary tables, including quality flags and comments; SNP cluster definition files; and project summary and quality statistics. Reported quality statistics will include sample success rates, missing data rates, Mendelian consistency rates, investigator duplicate reproducibility rates, and HapMap concordance rates.

The CBS at the University of Washington will perform additional postrelease data processing as described previously. This group

### Table 2. Genotyping Platforms Used for Cases for Each Genetic Research Center and Its Associated Control Populations

<table>
<thead>
<tr>
<th>Cases</th>
<th>Platforms for Cases</th>
<th>Controls</th>
<th>Platform for Controls</th>
</tr>
</thead>
<tbody>
<tr>
<td>BASICMAR</td>
<td>Illumina Infinium Omni5</td>
<td>INMA Project</td>
<td>Illumina Omni1 array*</td>
</tr>
<tr>
<td>BRAINS</td>
<td>Illumina Infinium Omni5</td>
<td>ADHD Study</td>
<td>Illumina Omni1 array*</td>
</tr>
<tr>
<td>EDINBURGH</td>
<td>Illumina Human 660W-Quad</td>
<td>Wellcome Trust</td>
<td>Illumina Human1M-Duo*</td>
</tr>
<tr>
<td>GASROS</td>
<td>Illumina Infinium Omni5</td>
<td>HRS</td>
<td>Illumina Omni Quad 2.5M*</td>
</tr>
<tr>
<td></td>
<td>Illumina 610*</td>
<td>OAI</td>
<td>Illumina Omni Quad 2.5M*</td>
</tr>
<tr>
<td>KRAKOW</td>
<td>Illumina Infinium Omni5</td>
<td>KRAKOW</td>
<td>Illumina Infinium Omni5</td>
</tr>
<tr>
<td>LEUVEN</td>
<td>Illumina Infinium Omni5</td>
<td>LEUVEN</td>
<td>Illumina Infinium Omni5</td>
</tr>
<tr>
<td>LUND</td>
<td>Illumina Infinium Omni5</td>
<td>South Sweden Study</td>
<td>Pending</td>
</tr>
<tr>
<td>MCISS</td>
<td>Illumina Infinium Omni5</td>
<td>HRS</td>
<td>Illumina Omni Quad 2.5M*</td>
</tr>
<tr>
<td></td>
<td>OAI</td>
<td>Illumina Omni Quad 2.5M*</td>
<td></td>
</tr>
<tr>
<td>MIAMISR</td>
<td>Illumina Infinium Omni5</td>
<td>Study of Latinos</td>
<td>Pending</td>
</tr>
<tr>
<td>MUNICH</td>
<td>Illumina Human660W-Quad*</td>
<td>Wellcome Trust</td>
<td>IlluminaHuman1M-Duo*</td>
</tr>
<tr>
<td>NS</td>
<td>Illumina Infinium Omni5</td>
<td>HRS</td>
<td>Illumina Omni Quad 2.5M*</td>
</tr>
<tr>
<td></td>
<td>OAI</td>
<td>Illumina Omni Quad 2.5M*</td>
<td></td>
</tr>
<tr>
<td>NOMAS(S)</td>
<td>Illumina Infinium Omni5</td>
<td>HRS</td>
<td>Illumina Omni Quad 2.5M*</td>
</tr>
<tr>
<td></td>
<td>OAI</td>
<td>Illumina Omni Quad 2.5M*</td>
<td></td>
</tr>
<tr>
<td>OXVASC</td>
<td>Illumina Human660W-Quad</td>
<td>Wellcome Trust</td>
<td>IlluminaHuman1M-Duo*</td>
</tr>
<tr>
<td>REGARDS</td>
<td>Illumina Infinium Omni5</td>
<td>HRS</td>
<td>Illumina Omni Quad 2.5M*</td>
</tr>
<tr>
<td></td>
<td>OAI</td>
<td>Illumina Omni Quad 2.5M*</td>
<td></td>
</tr>
<tr>
<td>SAHLSIS</td>
<td>Illumina Infinium Omni5</td>
<td>South Sweden Study</td>
<td>Pending</td>
</tr>
<tr>
<td>SPS3</td>
<td>Illumina Infinium Omni5</td>
<td>HRS</td>
<td>Illumina Omni Quad 2.5M*</td>
</tr>
<tr>
<td></td>
<td>OAI</td>
<td>Illumina Omni Quad 2.5M*</td>
<td></td>
</tr>
<tr>
<td>ST GEORGE'S</td>
<td>Illumina Human660W-Quad*</td>
<td>Wellcome Trust</td>
<td>IlluminaHuman1M-Duo*</td>
</tr>
<tr>
<td>SWISS</td>
<td>Illumina Infinium Omni5</td>
<td>HRS</td>
<td>Illumina Omni Quad 2.5M*</td>
</tr>
<tr>
<td></td>
<td>Illumina 610/660*</td>
<td>OAI</td>
<td>Illumina Omni Quad 2.5M*</td>
</tr>
<tr>
<td>WHI</td>
<td>Illumina Infinium Omni5</td>
<td>HRS</td>
<td>Illumina Omni Quad 2.5M*</td>
</tr>
<tr>
<td></td>
<td>OAI</td>
<td>Illumina Omni Quad 2.5M*</td>
<td></td>
</tr>
<tr>
<td>WUSTL</td>
<td>Illumina Infinium Omni5</td>
<td>HRS</td>
<td>Illumina Omni Quad 2.5M*</td>
</tr>
<tr>
<td></td>
<td>OAI</td>
<td>Illumina Omni Quad 2.5M*</td>
<td></td>
</tr>
</tbody>
</table>

ADHD indicates Attention Deficit Hyperactivity Disorder; HRS, Health and Retirement Study; INMA, INFancia y Medio Ambiente; NOMAS, Northern Manhattan Study; OAI, Osteoarthritis Initiative; OXVASC, Oxford Vascular Study; REGARDS, Reasons for Geographic and Racial Differences in Stroke Study; and WHI, Women’s Health Initiative.

*Platforms used to genotype cases and controls before Stroke Genetics Network (SiGN) that will be used as part of the analysis.
will assist the Analysis Committee with data cleaning and, if requested, posting of data sets to dbGaP, as well as use imputation using reference data from the 1000 Genomes Project. The GWAS data cleaning process typically focuses first on resolving any sample identity problems identified at release (e.g., sex mismatch, unexpected sample duplicates, and cryptic relatedness). Samples will be identified that should be removed for some analyses but may be retained as part of the posting to dbGaP, such as unexpected relatedness. Chromosome anomalies will be identified, and genotypes will be filtered from an anomalous region. Batch effects (samples processed together, DNA source or extraction method, study) will be checked, and the analysis will control for differences in ethnicity. Principal component analysis will be used to identify ethnic outliers and to adjust for population stratification in association analyses. SNP filters will be developed, including missing data filters, duplicate errors, minor allele frequency, and Hardy-Weinberg equilibrium. The CBS typically performs a relatively simple association (precompute) analysis to determine whether there is a problematic level of genomic inflation, suggesting false-positives. Given the complexity of the SiGN data set with its multiple GRCs and control groups, this precompute will be performed within multiple strata to accommodate proper matching of cases and controls (e.g., cases from US GRCs versus US control groups; Swedish cases versus Swedish controls). The precompute will also allow investigators who access data to verify that they were able to download data, merge the genotype and phenotype data sets, and apply the filters correctly by repeating the precompute results. A quality control report describing the data set and results of data cleaning will be posted on dbGaP. In addition, the CBS will impute untyped variants across the genome using 1000 Genomes Project data as a reference and post the results on dbGaP.

Analysis Strategy

The Data Management Core will store cleaned genotype data for distribution to the Analysis Committee that will conduct the primary GWAS analysis. The analytic strategy will initially involve logistic regression models adjusting for GRC, country, or principal components to test the overall behavior of the test statistic. If the data show unacceptable levels of statistical inflation, the Analysis Committee will likely adopt linear mixed models to account for hidden structure in the case-control data. This approach has worked successfully for a genome-wide study with a similarly heterogeneous source of case and control samples. The Analysis Committee will adjust for age and sex in the final association analysis. Analyses will be performed for total ischemic stroke and for each subtype. Because most of the available control groups have not been currently genotyped for exome content, the exome analysis will be a secondary project that will proceed as ancestry-matched exome data become available in large control populations.

Power Estimates

Most cases are of European white ancestry, although some US sites also contributed black and Hispanic cases. Power estimates indicate that the available number of European white cases (n=10633) and equivalent number of controls would provide 80% power to detect stroke-associated SNPs, with odds ratios of 1.05 to 1.09 across allele frequencies ranging from 0.10 to 0.50. For the 2 most common CCS-defined stroke subtypes, lacunar and cardioembolic stroke, power would be 80% to detect odds ratios ranging from 1.10 to 1.17.

Conclusions

The NINDS-supported SiGN is a large-scale international collaboration aimed to discover genetic determinants of ischemic stroke and its subtypes. SiGN is uniquely positioned to successfully accomplish this objective because of sufficient power to detect the genetic associations, a standardized approach to classification of ischemic stroke subtypes, a centralized approach to genotyping, and a large collection of clinical-phenotypic and imaging data, as well as genotypes for future discoveries. SiGN investigators have emphasized quality control of the phenotype data, including blinded readjudication of ischemic stroke subtypes, as well as other quality control checks. The NINDS provides considerable management and scientific input to the Scientific Steering Committee and the SiGN investigators, consistent with U01 Cooperative Agreement funding. The SiGN organizational structure with the management leadership team represented through the Scientific Steering Committee is open, flexible, and transparent with a collaborative spirit, which has been essential to resolve scientific issues and accomplish project tasks in a timely manner.

The next challenge for SiGN will be to develop open collaborations with other studies and consortia. An ongoing collaboration with Cohorts for Heart and Aging Research in Genomic Epidemiology (CHARGE) has already been established to coordinate future proposals and analyses. It would be scientifically advantageous to collaborate with groups that have access to large cohorts of Asian descent. Genetic associations that are validated across diverse populations are more likely related to functional variants. The SiGN Publications and Data Access Committee has adopted policies that are open to collaboration with all researchers, with the goal of maximizing progress toward understanding the role of genetic variation in risk of stroke. The SiGN Publications and Data Access Committee helps the Scientific Steering Committee to prioritize analyses and publications and assures recognition of the scientific efforts of all investigators involved in SiGN.

Appendix

Collaborators

The SiGN study also includes the following collaborators: Christopher D. Anderson, MD; Kerstin Andreu, MD; Gunnar Andsberg, MD; Ethem Arsava, MD; Kevin M. Barrett, MD, MSc; Thomas Benner, PhD; Paul Bentley, MD; Steven Bevan, PhD; Alessandro Biffi, MD; David Brenner, MD; Sherita Chapman, MD; Yu-Ching Cheng, PhD; Hossein Delavaran, MD; Christopher Deline, MD; Tomasz Dziedzic, MD; Dale Gamble, CCRP, MHSc; Eva Giralt, MD; Anja Grazer, MD; Andreas Gschwendtner, MD; Caityln Hegge, BS; Laura Heitsch, MD; Johanna Helenius, MD; Lukas Holmegaard, MD; Mohammed Huq, MD; Robert E. Irie, PhD; Rebecca Jackson, MD; Petra Katschnig, MD; Michael Katsnelson, MD; Naim Khoury, MD, MS; Selma Kilinc, MD; Gyanendra Kumar, MD; Daniel Labovitz, MD; Silvia Lanfranconi, MD; Carl D. Langefeld, PhD; Robin Lemmens, MD; Linxin Li, MD; Shaneela Malik, MD; Olle Melander, PhD; Sara Nordstrom, MD; Bo Norrving, MD; Angel Ois, MD PhD; Raid Ossi, MD; Leema Reddy Peddareddygari, MD; Annie Pedersén, MD; Joanna Pera, PhD; Mateusz Pucek, MD; Kristiina Rannikmae, MD; Petra Redfors, MD; J. David Rhodes, BSN, MPH; Marta Ribas, PhD; Neha Saraf, MD; Markus Schuerks, MD; Stephan Seiler, MD; Huma Sheikh, MD; Andrew M Southerland, MD; Mary J. Sparks, RN, BSN; Eva Stoegerer, MD; Jordi Sunyer, MD, PhD; Ella Temple, PhD; Raffaeilla Valenti, MD; David Weir, MD; Darren Weissman, MD; Rebecca Woodfield, MD; Gabriel S.C. Yin, MRCP.

Acknowledgments

National Institute of Neurological Disorders and Stroke program officials are Drs Gwinn and Corriveau.

Sources of Funding

SiGN: The study was funded by a cooperative agreement grant from the National Institute of Neurological Disorders and Stroke (NINDS)
supports the databases of the Graz Stroke Study and the Austrian Stroke Prevention Study.

ISGS and SWISS: The Ischemic Stroke Genetics Study (ISGS) was supported by the NINDS (R01 NS42733; PI Dr Meschia). The Sibling with Ischemic Stroke Study (SWISS) was supported by the NINDS (R01 NS39987; PI Dr Meschia). Both SWISS and ISGS received additional support, in part, from the Intramural Research Program of the National Institute on Aging (Z01 AG000054-06; PI Andrew Singleton). SWISS and ISGS used samples and clinical data from the NIH-NINDS Human Genetics Resource Center DNA and Cell Line Repository (http://ccr.coriell.org/ninds), human subjects protocol numbers 2003-081 and 2004-147. SWISS and ISGS used stroke-free participants from the Baltimore Longitudinal Study of Aging (BLSA) as controls with the permission of Dr Luigi Ferrucci. The inclusion of BLSA samples was supported, in part, by the Intramural Research Program of the National Institute on Aging (Z01 AG000015-50), human subjects protocol number 2003-078. This study used the high-performance computational capabilities of the Biowulf/Linux cluster at the NIH (http://biowulf.nih.gov/).

KRAKOW: Phenotypic data and genetic specimens collection were funded by the grant from the Polish Ministry of Science and Higher Education for Leading National Research Centers (KNOW) and by the grant from the Medical College, Jagiellonian University in Krakow, Poland: K/ZDS/002848.

LEUVEN: The Leuven Stroke genetics study was supported by personal research funds from the Department of Neurology of the University Hospitals Leuven. Vincent Thijs is supported by a Fundamental Clinical Research grant from FWO Flanders (numbers 1.8.09.01-00 and 1.009.05.000).

LUND: The Lund Stroke Register was supported by the Swedish Research Council (K2010-61X-20378-04-3), Region Skåne, the Freemasons Lodge of Instruction EOS in Lund, King Gustaf V’s and Queen Victoria’s Foundation, Lund University, and the Swedish Stroke Association. Biobank services were provided by Region Skåne Competence Centre (RSKC Malmö), Skåne University Hospital, Malmö, Sweden, and Biobank, LabmedicinSkåne, University and Regional Laboratories Region Skåne, Sweden.

MCISS: The Middlesex County Ischemic Stroke Study (MCISS) was supported by intramural funding from the New Jersey Neuroscience Institute/IFK Medical Center, Edison, NJ, and The Neurogenetics Foundation, Cranbury, NJ. We acknowledge Dr Souvik Sen for his advice and encouragement in the initiation and design of this study.

MIAMISR and NOMAS(S): The Northern Manhattan Study (NOMAS) was supported by grants from the NINDS (R37 NS029993, R01 NS27517). The Cerebrovascular Biorepository at University of Miami/Jackson Memorial Hospital (The Miami Stroke Registry, Institutional Review Board No. 20070386) was supported by the Department of Neurology at University of Miami Miller School of Medicine and Evelyn McKnight Brain Institute. Biorepository and DNA extraction services were provided by the Hussmann Institute for Human Genomics at the Miller School of Medicine.

MUNCH: The MUNCH study was supported by the Vascular Dementia Research Foundation and the Jackstaedt Stiftung.

NHS: The Nurses’ Health Study work on stroke is supported by grants from the National Institutes of Health, including HL088521 and HL34594 from the National Heart, Lung, and Blood Institute, as well as grants from the National Cancer Institute funding the questionnaire follow-up and blood collection: CA87969 and CA49449.

OXVASC: The Oxford Vascular Study was supported by the Stroke Association, Medical Research Council, Wellcome Trust, Dunhill Medical Trust, National Institutes of Health Research (NIHR), and NIHR Oxford Biomedical Research Centre based at Oxford University Hospitals NHS Trust and University of Oxford. Rothwell is in receipt of Senior Investigator Awards from the Wellcome Trust and the NIHR.

REGARDS: The Reasons for Geographic and Racial Differences in Stroke Study was supported by a cooperative agreement U01 NS041588 from the NINDS, National Institutes of Health, and Department of Health and Human Service. A full list of participating REGARDS investigators and institutions can be found at http://www.regardsstudy.org.

SAHLISIS: SAHLISIS was supported by the Swedish Research Council (K2011-65X-14605-09-6), the Swedish Heart and Lung
None.

References
Stroke Genetics Network (SiGN) Study: Design and Rationale for a Genome-Wide Association Study of Ischemic Stroke Subtypes


on behalf of the NINDS SiGN Study

*Stroke.* 2013;44:2694-2702; originally published online September 10, 2013;
doi: 10.1161/STROKEAHA.113.001857

*Stroke* is published by the American Heart Association, 7272 Greenville Avenue, Dallas, TX 75231
Copyright © 2013 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/44/10/2694

Data Supplement (unedited) at:
http://stroke.ahajournals.org/content/suppl/2013/09/10/STROKEAHA.113.001857.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

ONLINE SUPPLEMENT 1: Description of case ascertainment among the genetic research centers.

BASICMAR

BASe de datos de ICTus del hospital del MAR (BASICMAR) is an ongoing prospective study of all acute strokes assessed since 2005 at the IMIM-Hospital Universitari del Mar (Barcelona, Spain). It includes both first ever and recurrent strokes. There were no exclusion criteria regarding age or race-ethnicity of the individuals. All patients had an ECG, a blood analysis and neuroimaging at the acute stage. Additional diagnostic work-up was performed when clinically indicated. A follow-up of 3 months after stroke was completed for all survivor patients. Ischemic stroke etiologic subtypes were classified according to Trial of Org 10172 in Acute Stroke Treatment (TOAST) criteria.¹ For this study, only individuals of European origin with ischemic stroke were selected from BASICMAR, defining the eligible events as a clinical syndrome of any duration associated with a radiographically proven acute infarct, without radiographic evidence of a demyelinating or neoplastic disease or other structural disease including primary intracerebral hemorrhage.

BRAINS

The Bio-Repository of DNA in Stroke (BRAINS) is an on-going, hospital-based study that seeks to establish a high quality biobank resource. Extensive phenotype information is collected including subtype of stroke, past and family cardiovascular history, blood pressure data, MRI or CT brain imaging, carotid anatomy and blood tests (including cholesterol). All hospital admitted patients over the age of 18 years with first-ever or recurrent stroke that provided informed consent (or caregivers on their behalf) were recruited. Patients must have image positive lesions. Exclusion criteria are mainly for brain image negative patients even if the clinical presentation is that of stroke. There are no eligibility criteria based on stroke severity or participation in a treatment trial. Inability to obtain consent results in mandatory exclusion. Additional information about BRAINS can be found in prior publications.²³

EDINBURGH

Edinburgh Stroke Study, Scotland, UK: Between 2002 and 2005 consecutive consenting patients with stroke who were admitted to or seen as outpatients at the Western General Hospital, Edinburgh were prospectively recruited. There was no selection of cases based on
age, stroke severity, or inclusion in other clinical research studies. Cases in this study were of European origin, with a clinically evident stroke, demonstrated by brain imaging (CT or MRI) to be ischemic. An experienced stroke physician assessed each patient as soon as possible after the stroke, prospectively recording demographic and clinical details, including vascular risk factors and results of brain imaging and other investigations. Ischemic subtypes were assigned according to the TOAST criteria and, subsequently, using the Computerized Classification System (CCS), specifically for the purposes of the SIGN study.

**GASROS**

The Genes Affecting Stroke Risk and Outcome Study (GASROS) enrolled all ischemic stroke subjects as part of an ongoing single-center prospective cohort study of consecutive patients with ischemic stroke aged ≥18 years admitted to the Massachusetts General Hospital Stroke Unit after presenting to the emergency department within 24 hours of symptom onset between 2003 and 2011. Ischemic stroke was defined as a clinical syndrome of any duration associated with a radiographically proven acute infarct consistent with a vascular pattern of involvement and without radiographic evidence of a demyelinating or neoplastic disease or other structural disease, including vasculitis, subacute bacterial endocarditis, vasospasm due to subarachnoid hemorrhage or cocaine abuse, or primary intracerebral hemorrhage. Diagnosis of acute cerebral ischemia was confirmed for all subjects in the present study by admission diffusion-weighted MRI completed within 48 hours after symptom onset. Vascular and critical care neurologists subtyped ischemic strokes by systematic medical record review using the TOAST system as well as the Causative Classification of Stroke System. Controls were matched to cases on age, sex and race/ethnicity and drawn from stroke-free individuals who received care at primary care practices within Massachusetts General Hospital.

**GCNKSS**

The Greater Cincinnati/Northern Kentucky Stroke Study (GCNKSS) is a population-based epidemiological study of stroke in blacks and whites that is designed to measure temporal trends and racial differences in incidence of stroke. The GCNK region includes two southwestern Ohio counties (Hamilton, which includes the city of Cincinnati, and Clermont to the east) and three Northern Kentucky counties (Boone, Kenton, and Campbell) to the south of Cincinnati across the Ohio River. As part of the GCNKSS for calendar years 1999 and 2005, prospective cohorts of first-ever and recurrent ischemic stroke cases were assembled using “hot pursuit” methodology at all local hospitals in the region (18 in 1999, and 17 in 2005), except for
one hospital that is solely devoted to pediatric cases. Participants remained eligible if they were in a treatment study, but participation in a treatment study was not a requirement of enrollment. Patients with all degrees of severity of stroke were eligible, and no particular racial group was oversampled (approximately 80% white, 20% black). Study research nurses prospectively screened inpatient admission and emergency department logs to identify acute ischemic stroke patients. When a case was identified and the treating physician had given permission to approach the patient, a study nurse asked the patient or a proxy (the most closely related competent individual, preferably a person living with the patient prior to the stroke) to consent to participate in the cohort. If consent was granted, an extensive interview was performed by a study nurse, and a blood draw was performed for genetic analysis. In addition, a study nurse abstracted information about the patient, the patient’s medical history, the stroke event, and imaging studies from the hospital chart. A study physician reviewed every abstract, along with the imaging studies, to verify that an acute stroke had occurred, and to classify the event according to TOAST criteria.\(^1\)

GEOS

The Genetics of Early Onset Stroke (GEOS) Study is a population-based case-control study designed to identify the genetic determinants of early-onset ischemic stroke and to characterize interactions of stroke-associated genes with environmental risk factors. Cases with a first-ever ischemic stroke were identified by discharge surveillance from one of 59 hospitals in the greater Baltimore-Washington area and by direct referral from regional neurologists. Cases and controls were recruited in three different time periods: Stroke Prevention in Young Women-1 (SPYW-1) conducted from 1992-1996, Stroke Prevention in Young Women-2 (SPYW-2) conducted from 2001-2003, and Stroke Prevention in Young Men (SPYM) conducted from 2003-2007. SPYW-1 included cases aged 15-44 years recruited within one year of stroke and was designed with a 1:2 case-to-control ratio. SPYW-2 and SPYM included cases aged 15-49 recruited within three years of stroke and was designed with a 1:1 case-to-control ratio. Control participants without a history of stroke were identified by random-digit dialing. Controls were balanced to cases by age and region of residence in each study and were additionally balanced for ethnicity in SPYW-2 and SPYM. The number of cases and controls recruited in each study is as follows: 115 cases and 198 controls from SPYW-1, 234 cases and 209 controls from SPYW-2, and 478 cases and 500 controls from SPYM. The abstracted hospital records of potential cases were reviewed and adjudicated for ischemic stroke, ischemic stroke subtype, and modified Rankin Scale at discharge by a pair of vascular neurologists according to previously published procedures\(^4,5\) with
disagreements resolved by a third vascular neurologist. Stroke was defined according to the criteria of the World Health Organization and ischemic stroke was defined based on the criteria of the NINDS Stroke Data Bank. Cases had a head CT and/or brain MRI that was consistent with cerebral infarction. Visualization of the infarct was not required, only that no alternative etiology was identified. The ischemic stroke subtype classification system retains information on all probable and possible causes, and is reducible to the more widely used TOAST system that assigns each case to a single category. Ischemic strokes with the following characteristics were excluded from participation: stroke occurring as an immediate consequence of trauma, stroke within 48 hours after a hospital procedure, stroke within 60 days after the onset of a non-traumatic subarachnoid hemorrhage, and cerebral venous thrombosis. Additional exclusions for genetic analyses modified from were as follows: known single-gene or mitochondrial disorders recognized by a distinctive phenotype (e.g. cerebral autosomal dominant arteriopathy with subcortical infarcts and leukoencephalopathy (CADASIL), mitochondrial encephalopathy with lactic acidosis and stroke-like episodes (MELAS), homocystinuria, Fabry disease, or sickle cell anemia); mechanical aortic or mitral valve at the time of index stroke; untreated or actively treated bacterial endocarditis at the time of the index stroke; neurosyphilis or other central nervous system infections; severe sepsis with hypotension at the time of the index stroke; cerebral vasculitis by angiogram and clinical criteria; post-radiation arteriopathy; left atrial myxoma; major congenital heart disease; and cocaine use in the 48 hours prior to their stroke. There were no exclusions based on race or ethnicity, stroke severity, or participation in clinical trial research. Demographic variables, including age, ancestry, ethnicity, and established stroke risk factors were collected during a standardized interview. Risk factors included history of hypertension, diabetes, myocardial infarction and, defined as use within one month prior to event for cases and at a comparable reference time for controls, current smoking status and current oral contraceptive use.

GRAZ

Between 1994 and 2003 patients with first-ever and recurrent ischemic strokes admitted to the stroke unit of the Medical University Graz Department of Neurology were included. All race-ethnic groups were eligible and we did not target specific groups according to ethnicity. All age groups were allowed; however to our department only patients above the age of 18 are admitted. Ischemic stroke was defined as an episode of focal neurological deficits with acute onset and lasting > 24 hours. There were no selection criteria based on stroke severity. Patients in treatment trials were excluded. 685 patients were eligible to participate in this study (278
women, 407 men). All patients were Caucasian. Mean age was 68.9 ± 13.8 years with an age range from 19 to 101 years. In addition to a standardized protocol including a laboratory examination and carotid ultrasound or magnetic resonance angiography and ECG, 304 patients underwent neuroimaging by CT and 381 by MRI. More extensive cardiac examination, including transesophageal echocardiography or transthoracic echocardiography and Holter, was done in subjects with suspected cardiac embolism. The Stroke subtypes were assessed according to modified Trial of Org 10172 in Acute Stroke Treatment (TOAST) criteria¹ and were done by trained stroke neurologists.

ISGS

The Ischemic Stroke Genetics Study (ISGS) was a 5-center prospective hospital-based inception cohort study of first-ever ischemic stroke. Enrollment for ISGS began in December 2002 and was completed in July 2007. During this time 656 cases of first-ever ischemic stroke and 648 stroke-free controls were enrolled across the 5 centers (Mayo Clinic, Jacksonville, FL; Mayo Clinic Rochester, MN; University of Virginia, Charlottesville, VA; Shands Hospital, Jacksonville, FL; Grady Hospital, Atlanta, GA). All cases required meeting the World Health Organization definition for stroke and head imaging, by either head MRI or CT, confirmed no alternative cause for the stroke symptoms other than focal cerebral ischemia. All participants had to be over age 18 years. There were no eligibility criteria based on stroke severity or enrollment status in a treatment trial. Cases were excluded if they had CADASIL, MELAS, homocystinuria, or sickle cell anemia or if their stroke was due to vasculitis, vasospasm due to subarachnoid hemorrhage, mechanical aortic valve or mechanical mitral valve, or occurred within 30 days of a vascular surgical procedure: Baseline assessment of patients included standardized assessment of demographics, medical history, vital signs, results of baseline blood tests, pre-stroke functional status per modified Rankin Scale, and NIH Stroke Scale. Functional outcomes at 90 days post stroke onset were assessed using telephonic structured interview to obtain Oxford Handicap Scale, Glasgow Outcome Scale and Barthel Index. To minimize center-to-center variability, a single vascular neurologist (Robert D. Brown, Jr., MD) reviewed all available records of every ischemic stroke case for purposes of classification by etiology and syndrome. He used the TOAST criteria,¹ along with the Baltimore-Washington criteria and the Oxford Community Stroke Project criteria. He received medical records from the 5 centers that were stripped of personal identifiers, coded with study ID, and compiled in standard fashion.
KRAKOW

All consecutive patients with ischemic stroke (fulfilling WHO criteria) who were admitted to the Stroke Unit at the Jagiellonian University and who provided the informed consent were included into the study. The Stroke Unit serves as a stroke emergency center for one district of Krakow (200,000 inhabitants) and as a referral center for South East Poland (up to 15% of all admissions). For this on-going, prospective single center, hospital-based study participants with first ever or recurrent strokes were recruited from January 22, 2002 to September 9, 2010. The study was approved by local research ethics committee. Participants of the treatment trials were excluded. All patients were White. Stroke severity did not affect the inclusion or exclusion criteria. All patients had performed clinically relevant diagnostic workup, including brain imaging with computed tomography (CT) (100%) and/or magnetic resonance imaging (MRI) (up to 20%) as well as ancillary diagnostic investigations including duplex ultrasonography of the carotid and vertebral arteries (ca. 90%), transthoracic echocardiography (ca. 70%). MR-angiography, CT-angiography ECG-Holter monitoring, transesophageal echocardiography and blood tests for hypercoagulability were performed were indicated. Patients were classified into etiologic subtypes according to the Trial of Org 10172 in Acute Stroke Treatment (TOAST). All cases were phenotyped independently by two experienced stroke neurologists with review of original imaging.

LEUVEN

Patients of European descent with cerebral ischaemia, defined as a clinical stroke with imaging confirmation or a TIA with a new ischemic lesion on diffusion-weighted MRI, who were admitted to the Stroke Unit of the University Hospitals in Leuven were enrolled in the Leuven Stroke Genetics Study (LSGS) between 2005 and 2009. All patients from the LSGS study underwent brain imaging (MRI in 91% of patients, CT in the remainder) and a standardized protocol including lab examination, carotid ultrasound or CT angiography and cardiac examination (echocardiography and Holter monitoring) in all patients. Based on clinical presentation and results from the diagnostic work-up, case were classified into ischemic stroke etiologic subtypes according to modified Trial of Org 10172 in Acute Stroke Treatment (TOAST) criteria by a single reviewer who classified all cases. The reviewer had access to all information and imaging. Large-vessel disease was defined as either occlusive or significant stenosis (corresponding to >50% diameter reduction according to NASCET criteria) of a clinically relevant precerebral or cerebral artery, presumably due to atherosclerosis. Carotid ultrasound was used as a screening tool, and in principle, additional imaging with CTA or MRA was
performed when a high-grade stenosis was identified. In case CTA was used as the primary imaging modality, stenosis was confirmed by carotid ultrasound. In case of posterior circulation infarcts on imaging, CTA or MRA was used as the primary imaging modality to determine the degree of stenosis. Probable causes of cardiac embolism should be excluded. The presence of a patent foramen ovale was not considered a cardiac source in this context. Intracranial atherosclerosis was considered present only if repeat imaging after at least one week revealed a similar degree of stenosis or persistent occlusion. If not, the findings were interpreted as an embolism from a proximal source. Small-vessel disease was defined as a relevant infarct of <20 mm on DWI in areas supplied by single, small penetrating branches from middle cerebral artery, posterior cerebral artery or basilar artery in the absence of both a cardioembolic source and significant stenosis/occlusion due to atherosclerosis of an appropriate major brain artery. Cardioembolic stroke was defined as the presence of atrial fibrillation, sick sinus syndrome, myocardial infarction in the past four weeks, cardiac thrombus, infective endocarditis, atrial myxoma, prosthetic mitral or aortic valve, valvular vegetations, left ventricular akinetic segment, dilated cardiomyopathy, or patent foramen ovale or atrial septal aneurysm. Significant stenosis/occlusion due to atherosclerosis of an appropriate precerebral or cerebral artery should be excluded. Other determined cause of stroke included those with arterial dissection, vasculitis, hematologic disorders, monogenic syndromes and complications of cardiovascular procedures. Dissection was diagnosed by typical findings on contrast enhanced MRA and T1-fat suppressed MRI. Cryptogenic stroke was defined when no cause was identified despite an extensive evaluation. Strokes associated with significant aortic arch atheroma with plaques of ≥4 mm were also considered cryptogenic strokes.

LUND

Lund Stroke Register (LSR) is an ongoing study including consecutive patients with first-ever stroke since 1 March 2001 from the local uptake area of Skåne University Hospital, Lund Sweden. The WHO criteria for stroke diagnosis are used. Patients aged 18 years or older with stroke caused by cerebral infarct, intracerebral hemorrhage or subarachnoid hemorrhage are included. Patients are included regardless of stroke severity, race-ethnic group belonging, or participation in any treatment trial. Patients with iatrogenic or traumatic stroke are not included. In the SiGN study, patients from LSR with first-ever ischemic stroke between 1 March 2006 and 28 Feb 2010 were included if they or their next of kin provided informed consent. Age over 90 years is set to 90 years to maintain anonymous situation. All these patients underwent CT/MR/autopsy of the brain. EKG was performed on all patients. Echocardiography, ultrasound,
CT angiography or MR angiography of cerebral arteries was performed when judged clinically relevant. The subtype of ischemic stroke was determined by using the TOAST, the SSS-TOAST, or the CCS classification.¹,¹⁰

**MCISS**

The Middlesex County Ischemic Stroke Study (MCISS) was initiated as a prospective hospital based stroke registry at the New Jersey Neuroscience Institute, Edison, New Jersey. All patients over age 18 years were included and no specific ethnic/racial group was targeted or excluded. From 2000 to 2009, 1139 patients with ischemic strokes were enrolled in this registry. There was no selection criteria based upon stroke severity and both first ever and recurrent strokes were included in the study. Patients that were participants in treatment trials were not excluded. The major race/ethnic groups are Caucasians (67.2%), African Americans (14.3%), Asian Indians (8.2%), Hispanic (5.5%) and others (4.8%, Chinese and other Asians). All patients with clinical suspicion of a stroke were admitted through the emergency room to a dedicated stroke unit supervised by a vascular neurologist. After a history and neurological examination, a standardized series of investigations were performed: CBC and differential, comprehensive metabolic panel, electrolytes, BUN, creatinine, lipid panel (total cholesterol, LDL, HD) triglyceride levels, homocysteine levels, a cerebral MRI/MRA (if the MRI could not be performed, a cranial CT scan was done), carotid duplex ultrasound, electrocardiogram and an echocardiogram. The diagnosis of cerebral infarct was confirmed by the imaging studies. The epidemiological and clinical data on these patients was collected prospectively. Two independent investigators (one of which was a board certified neurologist with expertise in vascular neurology) reviewed the data, and all strokes were classified into etiological subtypes by applying the TOAST criteria.¹ In addition, the Oxfordshire stroke classification was applied and the vascular distribution of stroke was tabulated. All procedures, including the generation of the databases and recruitment of the stroke patients was conducted following Institutional Review Board policies and procedures at the New Jersey Neuroscience Institute/JFK Hospital.

**MIAMISR**

The Miami Stroke Registry and Biorepository (MIAMISR) at the University of Miami/Jackson Memorial Hospital is an ongoing prospective hospital registry of consecutive patients with prevalent stroke (ischemic and hemorrhagic) and TIA with available neuroimaging (CT or MRI) who provide informed consent. There is no specific exclusion criteria with the respect to patient age, stroke severity, disability or participation in treatment trials. It was established in November
of 2008 in order to investigate stroke type, ischemic stroke subtypes, stroke genetics and stroke outcomes in diverse ethnic population of Miami. The stroke population in Miami Stroke Registry is predominately Hispanic (63%) with Cuba (32%), Nicaragua (4.8%), Colombia (4.8%), and Puerto Rico (4.1)% being top countries of origin. Jackson Memorial Hospital is a 1550-bed county hospital affiliated with the University of Miami with 900 annual stroke and transient ischemic attack (TIA) admissions. Demographic, clinical data, and blood samples for genetic and other research have been collected prospectively during the hospitalizations and at 90 days by phone or in person. Trained research staff obtained written informed consent from the stroke patients or the health care proxy when available for participation in the prospective Miami Stroke Registry and Biorepository at the University of Miami/Jackson Memorial Hospital.11

MUNICH

Patients with first-ever or recurrent ischemic stroke were recruited consecutively from a single dedicated stroke unit (KlinikumGroßhadern, Ludwig-Maximilians-University of Munich) from 2002 onward. All patients were over the age of 18 years and of European descent. Brain imaging was performed in all patients, with most patients (>80%) undergoing MRI, including diffusion weighted imaging. Diagnosis of ischemic stroke was based on neurological symptoms in combination with a documented acute infarct on neuroimaging. Patients were not excluded based on stroke severity or whether they were enrolled in a treatment trial. Diagnostic workup included electrocardiography and duplex ultrasonography of the extracranial carotid arteries in all cases. Transcranial ultrasonography, CT angiography and/or magnetic resonance angiography, transthoracic and transesophageal echocardiography, and ambulatory electrocardiography were performed if clinically indicated.

NHS

The NHS cohort consists of 121,700 female registered nurses aged 30 to 55 years who were residing in 11 U.S. states and who were enrolled in 1976 through responding to a mailed questionnaire on their medical history and lifestyle practices. They have been followed with biennial mailed questionnaires collecting information on disease risk factors and health status. In 1989-1990, blood samples were collected from 32,826 participants. Among these participants, we prospectively identified incident strokes and confirmed ischemic stroke cases by medical record review. Clinical symptoms consistent with stroke and exclusion of alternate etiologies were required for classification of stroke. Virtually all cases had imaging, but confirmation on CT or MRI was not required. No participants were excluded based on
race/ethnicity. Neither stroke severity nor enrollment in a treatment trial was part of the eligibility criteria.

NOMAS

The Northern Manhattan Study (NOMAS) is an ongoing population-based study designed to determine stroke incidence, risk factors and outcome in an urban multiethnic population. The NOMAS started as a case-control study of index ischemic stroke patients admitted to the Columbia University Presbyterian Medical Center and affiliated hospitals and matching community controls in the 1993 (Northern Manhattan Stroke Study-NOMASS) and continued as a prospective stroke incidence study by following up controls in 1997 (NOMAS). Demographic and clinical data were collected prospectively during the hospitalizations and annually by phone or in person. Genetic samples were derived from these two sources: (1) the population-based case-control study conducted from 1993-98 (NOMASS) and (2) the ongoing prospective cohort study (NOMAS). First ever ischemic stroke cases were identified for the case-control study by screening of patient admissions, discharge codes, and referrals for neuroimaging at 15 acute care hospitals in the defined study area and multiple approaches to monitor for non-hospitalized cases. Incident ischemic stroke cases were identified from the prospective cohort study through follow-up visits and scheduled telephone contacts. Ischemic stroke cases from both sources were followed at 6 months by telephone and then annually afterwards in order to assess functional status and other outcomes. The administrative coordinating center of NOMAS has moved from NY to Miami in 2007, and the Institutional Review Boards of both institutions, Columbia University and the University of Miami approved the study.

OXVASC

OXVASC (Oxford Vascular Study) is an ongoing population-based study of the incidence and outcome of cerebrovascular, cardiovascular and peripheral vascular events since April 1, 2002. The OXVASC study population comprises all 91,105 individuals, irrespective of age, registered with 101 general practitioners in nine general practices in Oxfordshire, UK. Multiple overlapping methods of “hot” and “cold” pursuit are used to achieve near complete ascertainment of all cases as possible. All patients are consented and seen by study physicians as soon as possible after their initial presentation. In the SiGN study, patients of all ethnic groups from OXVASC with any ischemic stroke between April 1, 2002 and August 31, 2010 were included if they consented to have research DNA samples extracted. Ischemic stroke was defined as an episode of focal neurological deficits with acute onset lasting >24 hours or until death, with no apparent non-
vascular cause, and no signs of primary haemorrhage on brain imaging. An infarct did not need to be seen on CT or MRI to be included in this study. Patients were not excluded if they were of a treatment trial or for their stroke severity. Demographic data, major vascular risk factors (hypertension, diabetes, smoking, hyperlipidemia, prior TIA and history of coronary disease or peripheral vascular disease), and symptomatology were recorded in all patients. Patients routinely had brain imaging (CT or MRI), vascular imaging (Carotid Doppler or CTA/MRA or DSA), and 12-lead electrocardiography (EKG). Echocardiography and 24-hour EKG (HOLTER) monitoring were done in selected patients. All cases were subsequently reviewed by a senior neurologist and stroke etiology was classified according to modified Trial of Org 10172 in Acute Stroke Treatment (TOAST) criteria.\(^1\) Risk factors such as hypertension and diabetes were not included in the criteria. The patients were classified as undetermined stroke only if the diagnostic workup was complete (any form of brain imaging plus ECG and any form of vascular imaging), but no clear etiology was found. Patients of incomplete investigation were classified as unknown stroke while stroke of multiple causes was classified separately.

**REGARDS**

The REasons for Geographic and Racial Differences in Stroke (REGARDS) study is a US national, population-based, longitudinal cohort of 30,239 African-American and white adults aged ≥ 45 years, recruited January 2003-October 2007 with ongoing follow-up. Suspected stroke is queried every six months and triggered by participant self-report of stroke, stroke symptom(s), hospitalization or proxy report of death. Stroke severity and participation in a treatment trial did not limit inclusion in this study. Medical records for these reported events are retrieved and reviewed by at least two members of a committee of stroke experts with disagreements resolved by a third adjudicator. A symptom based approach, independent of neuroimaging outcome, is used to confirm events using the World Health Organization (WHO) definition of stroke. An infarct did not need to be seen on brain imaging to be included in this study. Ischemic stroke subtype classification is conducted using the TOAST system.\(^1,13\)

**SAHLSIS**

The Sahlgrenska Academy Study on Ischemic Stroke (SAHLSIS) is a case control study of ischemic stroke based in Gothenburg, Sweden.\(^14\) Adult patients who presented with first-ever or recurrent acute ischemic stroke before 70 years of age were recruited consecutively at stroke units in western Sweden 1998-2012. All participants were Caucasian. Patients were not excluded based on stroke severity or whether they were enrolled in a treatment trial. All patients
underwent ECG and neuroimaging at the acute stage (all by CT and 58% also by MRI). Additional diagnostic work-up was performed when clinically indicated. Inclusion criteria was ischemic stroke which was defined as an episode of focal neurological deficits with acute onset and lasting >24 hours or until death, with no apparent non-vascular cause, and no signs of primary hemorrhage on brain imaging. Patients were excluded if they had a diagnosis of cancer at advanced stage, infectious hepatitis or HIV. Stroke etiology was classified according to modified TOAST criteria, i.e. risk factors such as hypertension and diabetes were not included in the criteria.

SPS3

The Secondary Prevention of Small Subcortical Strokes (SPS3) trial (clinicaltrials.gov NCT00059306) is a randomized multicenter phase 3 trial of antiplatelet therapy and antihypertensive therapy. Participants are randomized to aspirin alone or the combination of aspirin and clopidogrel. Participants are also randomized to two groups of blood pressure control: either to a target systolic blood pressure (SBP) of 130 to 149 mmHg or < 130 mmHg. Principal eligibility criteria include man or woman at least 30 years of age with clinical evidence of small subcortical stroke and brain MRI evidence of small subcortical infarct. Patients were required to not have evidence of ipsilateral symptomatic cervical carotid stenosis or high risk cardioembolic sources for embolism. Further details of eligibility criteria have been published. Primary outcomes include ischemic and hemorrhagic stroke. DNA samples were collected from 38% (1139/3020) of participants in the trial. These samples were obtained from 46% (37/81) participating centers across the US, Spain, Mexico, Chile, Ecuador and Peru). No additional eligibility criteria were necessary beyond informed consent for participating in the DNA substudy. A total of 0.9% (10/1139) of DNA donors gave sample at time of randomization, with remainder donating at a later time point in follow-up.

ST GEORGE’S

First-ever and recurrent ischemic stroke patients of European descent attending a cerebrovascular service were recruited 1995-2008. All cases were phenotyped by one experienced stroke neurologist with review of original imaging. All patients had clinically relevant diagnostic workup performed, including brain imaging with computed tomography (CT) and/or magnetic resonance imaging (MRI) as well as ancillary diagnostic investigations including duplex ultrasonography of the carotid and vertebral arteries or magnetic resonance angiography.
MRA)/ CT-angiography (CTA), blood tests, and ECG, and where clinically indicated echocardiography and ambulatory electrocardiographic monitoring was performed. Patients were enrolled only if a symptomatic acute infarct was detected on head imaging. Participants had to be over the age of 18 years and have provided informed consent. No patient was excluded for participation in a treatment trial or because of stroke severity.

**SWISS**

The Siblings with Ischemic Stroke Study (SWISS) was a prospective, hospital-based affected sibling pair study of ischemic stroke. Enrollment for SWISS began in December 2000 and was completed in February 2011. DNA samples were collected from 312 ischemic stroke-affected sibling pairs. During this time, 1,026 probands with first-ever or recurrent ischemic stroke were enrolled across 70 centers in North America (66 US and 4 Canada). All probands required at least one living sibling with a history of stroke and required meeting the World Health Organization definition for stroke with head imaging, by either head MRI or CT, confirming no alternative cause for the stroke symptoms other than focal cerebral ischemia. Probands were excluded if they had CADASIL, MELAS, homocystinuria, or sickle cell anemia or if their stroke was due to vasculitis, vasospasm due to subarachoid hemorrhage, mechanical aortic valve or mechanical mitral valve, or occurred within 30 days of a vascular surgical procedure. Baseline assessment of patients included standardized assessment of demographic and medical history. Siblings were recruited primarily using proband-initiated contact. Stroke-affected siblings were screened using the Questionnaire for Verifying Stroke-free Status (QVSS). Eligibility criteria for affected siblings were the same as for probands. The Stroke Verification Committee, composed of two vascular neurologists, confirmed ischemic stroke status in affected siblings by medical record review. For probands, the center principal investigator classified ischemic stroke using the original TOAST classification system. Center principal investigators were neurologists certified in TOAST classification using stroke vignette training and certification process. The Stroke Verification Committee classified all affected siblings using TOAST based on medical record review. The Committee received medical records stripped of personal identifiers, coded with study ID, and compiled in standard fashion.

**WHI**

The Women’s Health Initiative Observational Study (WHI-OS) is a long-term follow-up study of post-menopausal women to identify and assess the impact of biological, genetic and lifestyle risk factors for cancer, cardiovascular disease, osteoporosis and other diseases of older
women. The National Institutes of Health Heart Lung and Blood Institute (NHLBI) sponsors the
WHI. The cases submitted here came from a case-control ancillary study nested within the
WHI-OS of the first 972 strokes occurring after WHI-OS baseline. This case-control study was
the Hormones and Biomarkers Predicting Stroke Study (HaBPS), conducted to examine blood
biomarkers in relation to stroke. HaBPS was sponsored by the National Institutes of
Neurological Disorders and Stroke (NINDS). Forty clinical centers throughout the United States
enrolled 93,676 women ages 50-79 at baseline into the parent study, the WHI-OS, between
September 1993 and February 28, 1997. Follow-up for clinical events and exposures is
ongoing. Recruitment into WHI-OS was mostly through mass mailings to age-eligible women
from large mailing lists such as voter registration, driver’s license, HCFA or other insurance
lists. Recruitment of minorities and older women was a particular study objective. Women
were either specifically recruited for the Observational Study or entered it because they were
ineligible or unwilling to be randomized into the Women’s Health Initiative Clinical Trials of
hormone therapy and/or dietary modification. Exclusions from WHI-OS were participation in
other randomized trials, predicted survival of less than 3 years, alcoholism, drug dependency,
mental illness, dementia, or other conditions making them unable to participate in the study.
Exclusions for the HaBPS case-control study of biomarkers of stroke were women with prior
history of myocardial infarction or stroke or those who did not have adequate blood sample for
biomarker assays. Strokes were first identified through annual mail and/or telephone follow-up,
and participant or third-party reports of overnight hospitalizations which were further
investigated by obtaining laboratory results, medical records and available imaging study
reports. Trained local physician adjudicators assigned a diagnosis according to standard
criteria. Locally adjudicated strokes were sent for central adjudication by three highly trained
neurologists. Each potential case was adjudicated by two neurologists and disagreements
were resolved by conference call consensus of the three neurologists. Only centrally
confirmed ischemic strokes that required hospitalization were used in this study. Transient
ischemic attacks (TIA’s) or hemorrhagic strokes (determined on review of reports of brain
imaging studies) were not included. Ischemic stroke was defined as the rapid onset of a
persistent neurologic deficit attributed to an obstruction lasting more than 24 hours and without
evidence for other causes. The deficit must have lasted more than 24 hours unless death
supervened or there was a lesion compatible with acute stroke demonstrated on computed
tomography (CT) or magnetic resonance imaging (MRI) scan. Ischemic strokes were also
centrally classified by TOAST criteria.
The Washington University St. Louis (WUSTL) patient collection included ischemic stroke patients admitted to Barnes-Jewish Hospital/Washington University Medical Center for genetic studies starting from August 1, 2008. Patients were identified for the genetic studies by screening admissions at our tertiary care hospital (both in the Emergency Department and on the Inpatient Stroke Service) without regard to age, race or ethnicity, including both first-ever and recurrent strokes. Patients were retained in the study if their discharge diagnosis was ischemic stroke (without requirement for the stroke to be visualized on CT or MRI). Demographic and clinical data were collected prospectively during the hospitalization and at 90 days, by phone or in person. Genetic samples were derived from subjects enrolled in 3 different studies: (1) Acute tPA pharmacogenomics study (Ischemic stroke patients who received tPA and were admitted to BJH/Washington University; serial NIHSS scores, and data on hemorrhagic transformation was collected), (2) Recovery Genomics after Ischemic Stroke Study (ReGenesIS, Ischemic stroke patients with NIHSS >3 without underlying chronic neurological disease, and expected survival up to 3 months after stroke), and (3) the Cognitive Recovery and Rehabilitation Group (CRRG) Registry (All ischemic stroke patients admitted to BJH/Washington University who consent to entering their clinical data into a stroke registry, and the collection of blood for genetic analysis). Patients that were part of a treatment trial were excluded from the tPA pharmacogenomics and ReGenesIS study, but not the CRRG registry.
ONLINE SUPPLEMENT 2: Memberships in SiGN committees, cores and centers.

**Scientific Steering Committee:** Steven Kittner, MD, James Meschia, MD, Donna Arnett, PhD, Kathryn Rexrode, MD, Jonathan Rosand, MD, Ralph Sacco, MD, Daniel Woo, MD, Raji Grewal, MD, Sylvia Smoller, MD, Bradford Worrall, MD, MSc, Patrick McArdle, PhD, Paul de Bakker, PhD, Miriam Fornage, PhD, Braxton D. Mitchell, PhD, Stephen Rich, PhD, Jin Moo Lee, MD, Jordi Jimenez-Conde, MD, Peter Rothwell, MD, Martin Dichgans, MD, Pankaj Sharma, MD, Cathie Sudlow, MD, Reinhold Schmidt, MD, Christina Jern, MD, Katarina Jood, MD, Agnieszka Slowik, MD, Arne Lindgren, MD, Hugh S Markus, MD, Vincent Thijs, MD

**Phenotype Committee:** James F. Meschia, MD, Chair, Robert D. Brown, Jr., MD, Bradford Worrall, MD, MSc, Hakan Ay, MD, Steven J. Kittner, MD, Patrick McArdle, PhD, Dale Gamble

**Analysis Committee:** Stephen Rich, PhD, Braxton D. Mitchell, PhD, Patrick McArdle, PhD, Paul de Bakker, PhD, Miriam Fornage, PhD

**Administrative Core:** Steven J. Kittner, MD, Mary J. Sparks.

**Genotyping Core:** Center for Inherited Disease Research (CIDR).

**Data Management Core:** Patrick McArdle, PhD.

**Imaging Core:** Jonathan Rosand, MD, MSc; Ona Wu, PhD

**Publications and Data Access Committee:** Tatjana Rundek, MD, Agnieszka Slowik, MD, James Meschia, MD, Steven J. Kittner, MD, Sylvia Smoller, MD, Bradford Worrall, MD, MSc, Daniel Woo, MD, Hakan Ay, MD, Jonathan Rosand, MD, Kathryn Rexrode, MD, Katrina Gwinn, MD, Braxton D. Mitchell, PhD, Patrick McArdle, PhD, Oscar Benavente, MD.

**Genetic Research Centers (GRC) (*certified adjudicator or certified re-adjudicator):**

MIAMISR: Ralph Sacco, MD; PI, Michael Katsnelson*, MD and Tatjana Rundek*, MD. Munich: Martin Dichgans, MD; PI, Andreas Gschwendtner*, MD, Selma Kilinc*, MD. NHS: Kathryn Rexrode, MD; PI, Markus Schuerks*, MD. NOMAS: Ralph Sacco, MD; PI, Michael Katsnelson*, MD and Tatjana Rundek*, MD. OXVASC: Peter Rothwell, MD; PI, Linxin Li*, Gabriel Yiin*, MD. REGARDS: Donna Arnett, PhD; PI, David Rhodes*, Neha Saraf*, MD, David Brenner*, MD. St. George: Hugh S Markus, MD; PI, Silvia Lanfranconi*, MD, Steve Bevan, PhD, Raffaella Valenti*, MD. SAHLSIS: Christina Jern, MD; PI, Katarina Jood*, MD; PI, Lukas Holmegaard*, MD, Annie Pedersen*, MD, Sara Nordstrom*, MD, Kerstin Andren*, MD, Petra Redfors*, MD. SPS3: Oscar Benavente, MD; PI, Julie A. Johnson, PharmD, Alan R. Shuldiner, MD, Leslie A. McClure, PhD., Sherita Chapman, MD SWISS: James F. Meschia, MD; PI, Raid Ossi*, MD, Shaneela Malik*, MD, and Andrew Southerland*, MD. WHI: Sylvia Smoller, MD; PI, Daniel Labovitz*, MD, Darren Weissman*, MD, Huma Sheikh*, MD. WUSTL: Jin-Moo Lee, MD; PI, Naim Khoury*, MD, Gyanendra Kumar*, MD and Laura Heitsch*, MD.
References for ONLINE SUPPLEMENT 1

References:


