Whole Genome Approaches in Ischemic Stroke

James F. Meschia, MD

Background and Purpose—The field of ischemic stroke genetics is moving beyond candidate gene studies into the realm of genomewide association studies. Such studies have resulted in discoveries in diverse, complex disorders.

Methods—The author conducted an informal qualitative review of peer-reviewed medical literature.

Results—The power of genomewide association studies to confirm prior associations and establish new ones is illustrated by recent work focusing on type 2 diabetes mellitus. A pilot genomewide association study of ischemic stroke failed to identify a single gene of major effect.

Conclusions—Follow-up studies with substantially greater statistical power are essential and are being planned by the Wellcome Trust and others. (Stroke. 2009;40[suppl 1]:S61-S63.)

Key Words: ischemia ■ whole genome ■ consesus data elements

Genomewide association studies have recently led to pivotal discoveries in a host of common disorders, including age-related macular degeneration, inflammatory bowel disease, and type 2 diabetes mellitus (T2D). Gene chip technology makes possible the genotyping of up to 500,000 to one million single-nucleotide polymorphisms (SNPs) in large numbers of disease-affected patients and unaffected control subjects. Investigators can now screen for regions within the human genome that associate with disease without having to hypothesize which variants of which genes are related to disease. This testing for genetic associations in a hypothesis-neutral way represents a striking departure from the common approaches of the past in which candidate genes were chosen a priori on the basis of assumed biological relevance. Hundreds of such candidate gene association studies have been done in ischemic stroke for various genes, including factor V, prothrombin, angiotensin-converting enzyme, and methylene tetrahydrofolate reductase. Studies generally lack the statistical power to exclude modest effects. Several meta-analyses have been done to increase statistical power.1,2 However, such analyses may magnify publication bias and are incapable of identifying novel genes.

Definitive genomewide association studies in ischemic and hemorrhagic stroke have yet to be completed. However, recent successes in other complex disorders illustrate the potential of the technique to yield important discoveries. Several large, well-executed genomewide studies have been done in T2D. In the interest of brevity, I focus on the design and findings of one study in particular, the Diabetes Genetics Initiative.3 The study was conducted in 2 stages. The first, so-called discovery stage involved 1464 patients with T2D and 1467 control subjects from Finland and Sweden. Investigators attempted to genotype a total of 500,568 SNPs, but only 386,731 SNPs survived rigorous quality control. The second, confirmatory stage did not require genomewide genotyping. A total of 10,850 additional subjects, both with and without T2D, from Sweden, Poland, and the United States, were genotyped for 107 SNPs. In addition to T2D, a series of secondary phenotypes were constructed, most of which were biometric data like body mass index and laboratory measures like insulinogenic index.

The Diabetes Genetics Initiative combined data for analysis from the Wellcome Trust Case Control Consortium (WTCCC) and the Finland–United States Investigation of NIDDM (Non-Insulin-Dependent Diabetes Mellitus) genetics studies, yielding a pooled sample size of more than 32,000 subjects. The result was confident identification of an association between several SNPs and T2D. Some findings were confirmatory; others novel. Investigators modeled sample sizes that would have been required, in retrospect, to discover associations between SNPs and T2D. They assumed 80% power, α of 0.05, observed pooled ORs, equal numbers of cases and control subjects, and 10% prevalence of T2D. Of the 8 SNPs that were modeled, sample size calculations ranged from approximately 800 to 7900 (median sample size estimate, 4250).

The first foray into genomewide association studies in ischemic stroke was a case–control study of approximately 500 subjects.4 Cases were recruited in the 5-center prospective inception cohort study known as the Ischemic Stroke Genetics Study (ISGS).5 All had neurologist-diagnosed recent ischemic stroke. All cases were centrally adjudicated for subtype of stroke using the Trial of Org 10172 in Acute Stroke Treatment (TOAST) criteria. Control subjects were

Received and accepted July 30, 2008.
From the Department of Neurology, Mayo Clinic, Jacksonville, Fla.
Correspondence to James F. Meschia, MD, Department of Neurology, Mayo Clinic, 4500 San Pablo Road, Jacksonville, FL 32224. E-mail meschia.james@mayo.edu
© 2009 American Heart Association, Inc.

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

DOI: 10.1161/STROKEAHA.108.533182

S61
neurologically normal adult men and women who had been previously genotyped. Adjusting for conventional stroke risk factors, no SNP attained clear genomewide significance. After attempting to harvest low-hanging fruit, investigators returned with an empty bushel. Nonetheless, the data set has proved useful. In a focused analysis, the enormous data set showed that the recently discovered cardiovascular risk factor locus on chromosome 9p21 also appeared to be a risk factor for ischemic stroke. The data set has been made publicly available through dbGaP to the extent allowable by the informed consent process (http://www.ncbi.nlm.nih.gov/sites/entrez?db=gap).

Assuming that ischemic stroke is no less genetically heterogeneous than T2D and that the effects of individual SNPs in imparting stroke risk are no greater than those in T2D, future stroke genetics studies will need samples from several thousand patients. The Wellcome Trust is funding a multistage genomewide association study in ischemic stroke as a collaboration between the International Stroke Genetics Consortium and the WTCCC-2 under the direction of Dr Hugh Markus. An initial genomewide screen will be performed on 4189 stroke cases, which will be compared with an equal number of previously genotyped control subjects from the WTCCC and the German KORA-gen population-based biobank. A replication stage will be done using 4432 cases from 2 additional UK and European populations. A third and final replication will be performed in 2873 cases and 2743 control subjects from 4 US populations.

There is intense interest among US investigators to conduct a similarly large genomewide association study. Preparatory to research, investigators have pooled data on individuals who have donated DNA. Samples are potentially available from a Phase 3 clinical trial, population- and hospital-based observational studies, and family-based studies. The Figure shows the age distribution of cases and control subjects of potentially available samples. Along with identifying sources

### Table. Consensus Common Data Elements for Planned Ischemic Stroke Genomewide Association Study

<table>
<thead>
<tr>
<th>Type of Data Element</th>
<th>Data Element</th>
<th>Categories</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Demographic</strong></td>
<td>Age</td>
<td>In years</td>
</tr>
<tr>
<td></td>
<td>Gender</td>
<td>Male/female</td>
</tr>
<tr>
<td></td>
<td>Race</td>
<td>White/black or African American/Asian/American Indian or Alaska Native/Native Hawaiian or Pacific Islander/other/unknown</td>
</tr>
<tr>
<td></td>
<td>Ethnicity</td>
<td>Hispanic or Latino/non-Hispanic or Latino/unknown</td>
</tr>
<tr>
<td></td>
<td>Name of principal investigator</td>
<td>(Free text field)</td>
</tr>
<tr>
<td></td>
<td>Name of study</td>
<td>(Free text field)</td>
</tr>
<tr>
<td></td>
<td>Subject no.</td>
<td>(Unique identifier that links DNA sample to clinical data elements)</td>
</tr>
<tr>
<td><strong>Risk factor</strong></td>
<td>Atrial fibrillation</td>
<td>Yes/no/unknown</td>
</tr>
<tr>
<td></td>
<td>Coronary artery disease</td>
<td>Yes/no/unknown</td>
</tr>
<tr>
<td></td>
<td>Diabetes mellitus</td>
<td>Yes/no/unknown</td>
</tr>
<tr>
<td></td>
<td>Hypertension</td>
<td>Yes/no/unknown</td>
</tr>
<tr>
<td></td>
<td>Cigarette smoking</td>
<td>Yes/no/unknown</td>
</tr>
<tr>
<td><strong>Stroke information</strong></td>
<td>Index stroke status</td>
<td>First-ever stroke/recurrent stroke/unknown/not applicable, control</td>
</tr>
<tr>
<td></td>
<td>Subject status</td>
<td>Ischemic stroke/control</td>
</tr>
<tr>
<td><strong>Stroke diagnostic evaluation</strong></td>
<td>Was imaging of the cervical carotid arteries performed?</td>
<td>Yes/no/unknown/not applicable, control</td>
</tr>
<tr>
<td></td>
<td>Was electrocardiography performed?</td>
<td>Yes/no/unknown/not applicable, control</td>
</tr>
<tr>
<td></td>
<td>What head imaging was done to confirm the diagnosis of ischemic stroke?</td>
<td>CT and MRI/CT only/MRI only/none/unknown/not applicable, control</td>
</tr>
<tr>
<td></td>
<td>What is the subtype of the index ischemic stroke?*</td>
<td>Large vessel atherosclerosis/small vessel/cardioembolic/other identified etiology/unknown etiology/not applicable, control</td>
</tr>
</tbody>
</table>

*Categories of ischemic stroke are defined using TOAST criteria.

Figure. Number of study participants in a North American subset of the International Stroke Genetics Consortium. Age distribution data for ischemic stroke cases and control subjects were provided by the following investigators (studies): Karen Furie, Massachusetts General Hospital (Genes Associated with Stroke Risk and Outcomes Study); Steven Kittner, University of Maryland (Stroke Prevention in Young Women 1, Stroke Prevention in Young Women 2, and Stroke Prevention in Young Men); Lynda Lisabeth, University of Michigan (Brain Attack Surveillance in Corpus Christi study); James F. Meschia, Mayo Clinic Jacksonville (Ischemic Stroke Genetics Study and Siblings with Ischemic Stroke Study); Ralph L. Sacco, University of Miami (Northern Manhattan Study); James Toole, Wake Forest University (Vitamin Intervention in Stroke Prevention study); and Daniel Woo, University of Cincinnati (GERFHS, CCC, Epi1999, and Epi2005).
of genetic material to test, investigators have worked together to create a uniform set of core common data elements across diverse studies in the hope of expediting large collaborative efforts (Table). It is the expectation that these core data elements will be richly supplemented in future efforts with additional clinimetric and radiographic phenotypic data to the extent feasible.

Acknowledgments
I thank Hugh Markus, MA, BMBch, DM, FRCP, for details regarding the planned WTCCC genomewide association study.

Sources of Funding
J.F.M. receives funding from an unrestricted grant from the National Institute of Neurological Disorders and Stroke for SWISS (the Siblings with Ischemic Stroke Study) (R01 NS39987).

Disclosures
None.

References


Whole Genome Approaches in Ischemic Stroke
James F. Meschia

Stroke. 2009;40:S61-S63; originally published online December 8, 2008;
doi: 10.1161/STRK.108.533182

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/40/3_suppl_1/S61

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/