Common Genetic Variants on Chromosome 9p21 Confers Risk of Ischemic Stroke
A Large-Scale Genetic Association Study

J. Gustav Smith, MD; Olle Melander, MD, PhD; Håkan Löfvqvist, MSc; Bo Hedblad, MD, PhD; Gunnar Engström, MD, PhD; Peter Nilsson, MD, PhD; Joyce Carlson, MD, PhD; Göran Berglund, MD, PhD; Bo Norrving, MD, PhD; Arne Lindgren, MD, PhD

Background—Epidemiological studies indicate a genetic contribution to ischemic stroke risk, but specific genetic variants remain unknown, with the exception of a few rare variants. Recent genome-wide association studies identified and replicated common genetic variants on chromosome 9p21 to confer risk of coronary heart disease. We examined whether these variants are associated with ischemic stroke.

Methods and Results—We genotyped 6 common genetic variants on chromosome 9p21, previously associated with coronary artery disease in genome-wide association studies, in 2 population-based studies in southern Sweden, the Lund Stroke Register (n=1837 cases, 947 controls) and the Malmö Diet and Cancer study (MDC; n=888 cases, 893 controls). We examined association in each study and in the pooled dataset. Adjustments were made for cardiovascular risk factors and further for previous myocardial infarction in MDC. We found a modest increase in ischemic stroke risk for 2 common (minor allele frequencies 0.46 to 0.49) variants, rs2383207 (P=0.04 in Lund Stroke Register, P=0.01 in MDC) and rs10757274 (P=0.03 in Lund Stroke Register, P=0.03 in MDC), in each sample independently. The strength of the association increased when samples were pooled with an odds ratio of 1.15 (95% CI, 1.05 to 1.25; P=0.002) for the strongest variant rs2383207. Results were similar after adjustment for clinical covariates. rs1333049 also showed significant association in MDC, which increased in the pooled sample (P=0.004).

Conclusions—In this large sample (n=4565), we detected common genetic determinants for ischemic stroke on chromosome 9p21. Our findings indicate that ischemic stroke shares pathophysiological determinants with coronary heart disease and other arterial diseases and highlight the need for large sample sizes in stroke genetics. (Circ Cardiovasc Genet. 2009;2:159-164.)

Key Words: genetics ■ atherosclerosis ■ cardiovascular diseases ■ cerebral infarction ■ stroke

Ischemic stroke is a leading cause of death and disability worldwide. Family and twin studies have established a genetic contribution to ischemic stroke but, with the exception of a small number of rare, monogenic disorders manifesting as ischemic stroke, few if any genetic variants have been irrefutably associated with stroke.1 Common diseases have been hypothesized to result from common genetic variants with individually modest impact on disease risk.2 Consequently, large sample sizes are necessary to detect or exclude association3 as has been shown in association studies for other complex diseases, including myocardial infarction4 and type 2 diabetes.5 The typical odds ratios (ORs) of common genetic variants seen in these studies have been in the range of 1.1 to 1.5, necessitating sample sizes of several thousand individuals for adequate statistical power.6 Calls for larger studies of ischemic stroke genetics have also been raised with a published power calculation for ischemic stroke suggesting that sample sizes in the range of 2444 to 4587 individuals are necessary for statistical power of 0.8 in studies of genetic variants conferring relative risks in the order of 1.2 with minor allele frequencies 0.05 to 0.10.6 This study represents such a large-scale association study of ischemic stroke, with 4565 individuals from 2 population-based studies.

Clinical Perspective see p 164

A recent development in the field of complex genetics is the use of genome-wide association (GWA) studies, which systematically examine common variants throughout the
entire genome. Two GWA studies for ischemic stroke have been published so far,7,8 of which one identified an association with a locus on chromosome 4q25 previously associated with atrial fibrillation,6 but were limited in statistical power by small sample sizes to detect additional associations with variants of small and moderate effect. However, the finding of several single-nucleotide polymorphisms (SNPs) on chromosome 9p21 in 4 GWA studies9–12 for coronary heart disease has been comprehensively replicated.13,14 We hypothesized the association of these SNPs with ischemic stroke, because previous studies have shown overlapping heritability of myocardial infarction and ischemic stroke,15 most likely due to a common atherosclerotic pathogenesis.

Methods

Samples and Definitions

We used 2 population-based samples from the same geographic region in southern Sweden, the Lund Stroke Register (LSR) and the Malmö Diet and Cancer study (MDC). Characteristics of sample, data collection, and clinical definitions in both LSR16 and MDC17 have been described. Briefly, LSR is a prospective, epidemiological register initiated in 2001, which consecutively includes all patients with a first ever stroke from the catchment area of Lund University Hospital (N = 234 505 inhabitants on December 31, 2001). Control subjects were randomly selected from the same region and matched to LSR cases of 2001 by age and gender using the Swedish Population Register. Blood samples and information on risk factors were acquired from patients with ischemic stroke between 2001 and 2006 for this study along with 947 control subjects.

MDC is a prospective, population-based cohort study that included 28 449 randomly selected men (born between 1923 and 1945) and women (born between 1923 and 1950) at baseline examination between 1991 and 1996. Subjects with ischemic stroke after the baseline examination were identified in the MDC17 stroke examinations between 1991 and 1996. Subjects with ischemic stroke before the baseline examination were excluded. Controls were drawn from the same cohort matched for gender, age, and time of baseline investigation. The diagnosis of ischemic stroke in all cases of both studies was ascertained in accordance with World Health Organization criteria.19 Individuals with intracerebral or subarachnoid hemorrhage or unspecified stroke were excluded. In MDC, myocardial infarction before screening was ascertained by record linkage to the Swedish Hospital Discharge Register. Myocardial infarction was defined on the basis of International Classification of Diseases, 9th Revision code 410. Clinical covariates (age, gender, hypertension, diabetes mellitus, and smoking) were examined at disease onset in LSR and at the baseline screening visit for both cases and controls in MDC. Hypertension was defined as systolic blood pressure ≥160, diastolic blood pressure ≥90 mm Hg, or the use of antihypertensive medications. Diabetes was defined as self-reported history of a physicians diagnosis of diabetes or antidiabetic treatment. Smoking was defined as current smoking (yes/no). Informed consent was obtained from all participants, and the study was approved by the local ethics committee.

Selection, Genotyping, and Quality Control of SNPs

We selected for genotyping the 6 SNPs recently associated with coronary artery disease in GWA studies: the 2 SNPs from the study by McPherson et al20 (rs10757274 and rs2383206), the 3 SNPs from the study by Helgadottir et al10 (rs10116277, rs1333040, and rs2383207), and the most significant SNP from the study by Samani et al11,12 (rs1333049). DNA extracted from EDTA-treated blood samples were randomized to batches within each sample. Both samples were genotyped with the same set of reagents on a matrix-assisted laser desorption/ionization time of flight mass spectrometer (MassArray, Sequenom Inc) using Sequenom reagents and protocols with 10-ng DNA as PCR template. Automatic allele calls by the Sequenom software were validated by manual evaluation. All laboratory analyses were performed by the SWEGENE Resource Center for Profiling Polygenic Disease (Malmö University Hospital, Malmö, Sweden), with no access to case/control status or personal identities of the samples. Data were stored at the Bioinformatics Unit at the same facility. We excluded all SNPs with minor allele frequency <0.05 or >0.5 missing genotypes.

Statistical Methods

Genetic association was examined in each sample separately and in the pooled sample using a prespecified analysis plan. Association was examined using unconditional logistic regression analysis in an additive inheritance model where the additive effect per copy of the minor allele on stroke risk is tested. We prespecified significance levels of P<0.05 in both samples or P<0.05, with adjustment for the number of tests performed using Bonferroni correction in the pooled sample. Analysis was also performed with adjustment for cardiovascular risk factors (hypertension, current smoking, and diabetes mellitus) in the MDC sample, we further adjusted for myocardial infarction before the baseline examination. Statistical analyses were performed in SPSS version 15 (SPSS Inc). Linkage disequilibrium (LD) and Hardy–Weinberg equilibrium were examined in Haploview version 4.0.20

Results

We included 2784 individuals from LSR and 1781 individuals from MDC. Characteristics of the 2 samples are shown in Table 1. Ischemic stroke occurred on an average of 6.3 years...
after inclusion in MDC (SD, 7). Results of genotyping are shown in Table 2. Four SNPs passed quality control criteria whereas 2 were excluded (rs10116277 and rs2383206) due to a call rate <0.95 in either sample. One SNP differed significantly from Hardy-Weinberg equilibrium in the full sample, rs1333040 ($P = 7 \times 10^{-4}$) but did not differ significantly when examined in controls only ($P = 0.11$). Minor allele frequencies for all SNPs were high, in the range of 0.45 to 0.49, which is in good agreement with frequencies seen for the same SNPs in the sample of European ancestry from the HapMap project (HapMap CEU).$^{21}$

Results from association analysis for each sample and in the pooled sample are shown in Table 3. Three SNPs were significantly associated with ischemic stroke in the pooled sample, of which 2 SNPs were significant in each sample separately. Only rs1333040 was not significant in any analysis. Association of the 3 SNPs remained significant after Bonferroni correction for 4 SNPs in the pooled sample ($P < 0.0125$). Adjustment for matching variables (age and sex) did not alter the magnitude of the OR or probability value, and neither did further adjustment for cardiovascular risk factors (hypertension, current smoking, and diabetes mellitus), as shown in Table 4. Adjustment for acute myocardial infarction before the baseline examination in MDC decreased ORs slightly and widened CIs resulting in a modest decrease in probability values, which remained significant for the strongest SNP, rs2383207, as shown in Table 4.

The SNP with the strongest association, rs2383207, was in strong LD with the second strongest SNP rs10757274 ($r^2 = 0.89$) and only marginally weaker LD with the third significantly associated SNP rs1333049 ($r^2 = 0.86$). The 2 less significant SNPs were also in strong LD ($r^2 = 0.86$) whereas the nonsignificant SNP, rs1333040, was in more modest LD with the others, rs10757274 ($r^2 = 0.54$), rs2383207 ($r^2 = 0.49$), and rs1333049 ($r^2 = 0.46$).

### Discussion

In this large genetic association study of ischemic stroke, we found significant association with a genetic locus on chromosome 9p21 previously showing a strong association with coronary heart disease. We used 2 samples to test the hypothesis of association with ischemic stroke and achieve replication, defined as association of the same allele with the same trait under the same genetic model.$^{22}$ The samples were drawn from studies of different design; one consisting of incident stroke cases and matched controls nested from a population-based cohort from the medium-sized city of Malmö and the other consisting of consecutive cases of ischemic stroke and randomly sampled matched controls from the smaller city of Lund. We then pooled the samples to achieve a sample size with good statistical power in line with previous power calculations for ischemic stroke$^6$ and found increased evidence of association. Three previous studies have examined 9p21 in ischemic stroke of which 2 small studies found no association$^{23,24}$ whereas a modestly sized study found a weak but significant association,$^{25}$ interestingly with the same unadjusted OR as for the strongest SNP in our study but with a higher standard error of the log-OR and with similarly high minor allele frequency.

This study has 2 major strengths. First, the sample size is large enough to allow adequate statistical power to detect an association. Virtually all previous studies of common variants in stroke have been underpowered to generate any reliable conclusions. Second, our study includes only individuals from the same geographic region in southern Sweden, a region with a uniform population.

### Table 2. Genotyping Results for 6 SNPs at 9p21

<table>
<thead>
<tr>
<th>SNP</th>
<th>Position</th>
<th>Minor/Major Allele</th>
<th>MAF LSR</th>
<th>MAF MDC</th>
<th>HWE LSR</th>
<th>HWE MDC</th>
<th>Call Rate LSR</th>
<th>Call Rate MDC</th>
</tr>
</thead>
<tbody>
<tr>
<td>rs10116277</td>
<td>22 071 397</td>
<td>T/G</td>
<td>0.46</td>
<td>0.45</td>
<td>...</td>
<td>...</td>
<td>0.75</td>
<td>0.97</td>
</tr>
<tr>
<td>rs10757274</td>
<td>22 086 055</td>
<td>G/A</td>
<td>0.47</td>
<td>0.46</td>
<td>0.37</td>
<td>0.38</td>
<td>0.96</td>
<td>0.98</td>
</tr>
<tr>
<td>rs1333040</td>
<td>22 073 404</td>
<td>C/T</td>
<td>0.45</td>
<td>0.45</td>
<td>0.21</td>
<td>0.32</td>
<td>0.95</td>
<td>0.95</td>
</tr>
<tr>
<td>rs1333049</td>
<td>22 115 503</td>
<td>G/C</td>
<td>0.47</td>
<td>0.47</td>
<td>0.54</td>
<td>0.13</td>
<td>0.98</td>
<td>0.98</td>
</tr>
<tr>
<td>rs2383206</td>
<td>22 105 026</td>
<td>G/A</td>
<td>0.49</td>
<td>0.49</td>
<td>...</td>
<td>...</td>
<td>0.86</td>
<td>0.89</td>
</tr>
<tr>
<td>rs2383207</td>
<td>22 105 959</td>
<td>G/A</td>
<td>0.49</td>
<td>0.49</td>
<td>0.16</td>
<td>0.42</td>
<td>0.95</td>
<td>0.97</td>
</tr>
</tbody>
</table>

Positions refer to chromosome 9 in NCBI build 36. For comparison, the closest protein coding genes CDKN2A and CDKN2B are located at 21 957 751 to 21 984 490 and 21 992 902 to 21 999 312, respectively, whereas the noncoding antisense RNA is located at 21 984 790 to 22 111 094. Hardy-Weinberg equilibrium (HWE) was examined in controls using Haploview with $P$ values presented for SNPs with call rate >0.95 in both samples. MAF indicates minor allele frequency.

### Table 3. Results of Association Analysis of Ischemic Stroke and 4 SNPs at 9p21

<table>
<thead>
<tr>
<th>SNP</th>
<th>LSR</th>
<th>MDC</th>
<th>Pooled</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>OR (95% CI)</td>
<td>$P$</td>
<td>OR (95% CI)</td>
</tr>
<tr>
<td>rs10757274</td>
<td>1.14 (1.01 to 1.27)</td>
<td>0.029</td>
<td>1.14 (1.02 to 1.33)</td>
</tr>
<tr>
<td>rs2383207</td>
<td>1.12 (1.01 to 1.25)</td>
<td>0.044</td>
<td>1.19 (1.04 to 1.36)</td>
</tr>
<tr>
<td>rs1333049</td>
<td>1.11 (0.99 to 1.24)</td>
<td>0.064</td>
<td>1.18 (1.03 to 1.35)</td>
</tr>
<tr>
<td>rs1333040</td>
<td>0.94 (0.84 to 1.05)</td>
<td>0.28</td>
<td>0.93 (0.81 to 1.06)</td>
</tr>
</tbody>
</table>

Data are presented as crude ORs per SNP, with 95% CIs and $P$ values.
history. This reduces the risk of the much debated problem of population stratification, the potential problem in genetic association studies of differences in ancestry between cases and controls leading to false-positive association of any genetic variants with frequencies that differ by 
ancestry.

A weakness in this study is the lack of comprehensive data on concurrent coronary heart disease. Data on myocardial infarction were available only from one of the cohorts (MDC) and at baseline of screening rather than at the time of stroke. In patients with transient ischemic attack and stroke, the prevalence of coronary heart disease is high; about one quarter of patients with stroke have a history of a symptomatic coronary event, and an additional 25% to 50% have asymptomatic coronary plaques, stenoses, or silent myocardial infarcts.26,27 Adjustment for myocardial infarction in the MDC sample attenuated the associations slightly but the strongest SNP remained significant. However, the rationale of adjusting for myocardial infarction in genetic stroke studies on determinants known to be associated with coronary heart disease is uncertain as myocardial infarction reflects the same proposed pathophysiological background (ie, atherosclerosis).

We suggest that these findings have several general implications. First, the genetics of ischemic stroke is a field that has long been plagued by nonrepllication of reported findings, mirroring results for other complex traits.3 Our results confirm that, as for other phenotypes, large-scale association studies can identify common genetic variants for ischemic stroke. Second, there has been some debate over whether etiologic subclassifications of ischemic stroke, such as the Trial of Org 10172 in Acute Treatment classification,28 are necessary to identify common genetic determinants but so far have resulted mostly in additional multiple testing burden in already underpowered studies.29 Our results show that common genetic variants can have an impact on the phenotype of ischemic stroke. However, the observed effect size (OR, 1.15) is smaller than for coronary heart disease (OR, 1.2 to 1.3),9–12 consistent with a more complex phenotype, and the effect size of 9p21 SNPs may very well be larger in etiologic subgroups. Third, the association between ischemic stroke and SNPs at a locus previously associated with coronary heart disease and both intracranial and abdominal aneurysms25 suggest that ischemic stroke shares common pathophysiological pathways with these diseases of the arterial wall. The nature of this pathway remains unknown because the specific causal variant and consequently causal gene is unknown but the negative results of a recent large-scale association study of 9p21 and markers of atherosclerosis30 suggests that the association may reflect plaque instability rather than development of atherosclerosis. Indeed, plaque instability has been proposed to be a systemic phenomenon and a major mechanism in ischemic stroke.31 As shown in Table 2, the SNPs are distant from any known protein-coding gene, the closest being the well-known tumor suppressor genes CDKN2B and CDKN2A. However, the SNPs are located in a recently annotated large, noncoding antisense RNA gene,32 ANRIL, which has been shown to regulate expression of CDKN2B33 and to be expressed in atherosclerotic plaques.14 All SNPs in our study are in intronic regions of ANRI IS except rs1333049, which showed the weakest association in our study and is located downstream of the RNA gene. CDKN2B is a good candidate gene as it regulates cellular proliferation and senescence in both endothelial14 and immunologic35 cells, which are thought to play central roles in plaque instability.36 Finally, establishment that common genetic variants confer risk of ischemic stroke gives hope for ongoing adequately powered GWA studies of ischemic stroke. Even though common variants seem to have such modest effects that contribution to clinical decision making seems unlikely, potential gains include novel insight into pathophysiology and novel therapeutic targets.

**Acknowledgments**

We thank Hélène Pessah-Rasmussen, MD, PhD (Department of Neurology, Malmö University Hospital), for assessment of stroke patients in the Stroke Register of Malmö and Ann-Cathrin Jönsson, RN, PhD (Department of Neurology, Lund University Hospital), for help with inclusion of stroke patients in the LSR study. We also thank all participants in MDC and LSR for making this study possible.

**Sources of Funding**

This study was supported by the Albert Påhlsson Research Foundation; the County Council of Skåne; the Crafoord Foundation; the Ernhold Lundströms Research Foundation; the Hulda and Conrad Mossfelt Foundation; the King Gustaf V and Queen Victoria Foundation; the Lennart Hanssons Memorial Fund; Malmö University Hospital; the Research Funds of the Department of Neurology, Lund; the Segerfalk Foundation; the Swedish Heart and Lung Foundation; the Swedish Medical Research Council; and the Swedish Stroke Association and Lund University.

**Disclosures**

None.

**References**


13. Smith et al Ischemic Stroke and 9p21


ANRIL, an antisense noncoding RNA whose expression coclusters with ARF. *Cancer Res.* 2007;67:3963–3969.


**CLINICAL PERSPECTIVE**

Ischemic stroke is a common, severely debilitating disease with complex etiology. A genetic contribution has been shown, but specific genetic variants are unknown with the exception of a few rare variants. A genome-wide association study of ischemic stroke recently identified single-nucleotide polymorphisms at a locus on chromosome 4q25 previously associated with atrial fibrillation, highlighting the importance of atrial fibrillation as a risk factor, but was underpowered to detect additional loci. As ischemic stroke has overlapping heritability also with coronary heart disease (CHD), potentially through atherothrombosis, we studied single-nucleotide polymorphisms at a locus on chromosome 9p21 previously found to be strongly associated with CHD in genome-wide association studies. The genes closest to the locus, *CDKN2A* and *CDKN2B*, are well-known tumor suppressor genes with a central role in regulation of cellular proliferation and senescence, potentially affecting stability of atherosclerotic plaques. In a large sample (n=4565) from southern Sweden, we identified an association of single-nucleotide polymorphisms at 9p21 and ischemic stroke. We observed smaller effect sizes than reported in CHD. Our findings highlight the complex etiology of ischemic stroke and suggest common pathophysiological pathways with CHD. Future studies are needed to determine the causal variant and pathway, which may now constitute a potential target for drug development in ischemic stroke and CHD.
Common Genetic Variants on Chromosome 9p21 Confers Risk of Ischemic Stroke: A Large-Scale Genetic Association Study
J. Gustav Smith, Olle Melander, Håkan Lökvist, Bo Hedblad, Gunnar Engström, Peter Nilsson, Joyce Carlson, Göran Berglund, Bo Norrving and Arne Lindgren

Circ Cardiovasc Genet. 2009;2:159-164; originally published online February 12, 2009; doi: 10.1161/CIRCGENETICS.108.835173
Circulation: Cardiovascular Genetics is published by the American Heart Association, 7272 Greenville Avenue, Dallas, TX 75231
Copyright © 2009 American Heart Association, Inc. All rights reserved.
Print ISSN: 1942-325X. Online ISSN: 1942-3268

The online version of this article, along with updated information and services, is located on the World Wide Web at:
http://circgenetics.ahajournals.org/content/2/2/159