Common genetic variants on chromosome 9p21 confers risk of ischemic stroke: a large-scale genetic association study

Short title: Smith et al, Ischemic stroke and 9p21

Manuscript ID#: CIRCULATIONAHA/2008/835173

Authors:

J Gustav Smith, MD, Olle Melander, MD, PhD, Håkan Lökvist, 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

Affiliations:

Program in Medical and Population Genetics, Broad Institute of Harvard and Massachusetts Institute of Technology (JGS), Cambridge, MA, USA. Department of Clinical Sciences (JGS, HL, BN, AL) and Department of Laboratory Medicine (JC), Lund University, Lund, Sweden. Department of Clinical Sciences (JGS, OM, BH, GE, PN, GB), Lund University, Malmö, Sweden.

Journal subject codes:


Total word count: 4940 (Max 6000 words)
Summary word count: 249 (Max 250 words)
Reference count: 36 (Max 50 references)
Clinical perspective word count: 211 (Max 250 words)

Correspondence addressed to:

Arne Lindgren, MD, PhD
Department of Neurology
Lund University Hospital
SE-221 85, Lund, Sweden
Tel: +46 46 171424
Fax: +46 46 158919
Email: arne.lindgren@med.lu.se
Abstract

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 six common genetic variants on chromosome 9p21, previously associated with coronary artery disease in genome-wide association studies, in two population-based studies in southern Sweden; Lund Stroke Register (LSR, n=1837 cases, 947 controls) and 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 two common (minor allele frequencies 0.46-0.49) variants, rs2383207 (p=0.04 in LSR, p=0.01 in MDC) and rs10757274 (p=0.03 in LSR, p=0.03 in MDC), in each sample independently. The strength of the association increased when samples were pooled with an OR of 1.15 (95%CI=1.05-1.25, 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.

Key words
Genetics, atherosclerosis, cardiovascular diseases, cerebral infarction.
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. Common diseases have been hypothesized to result from common genetic variants with individually modest impact on disease risk. Consequently, large sample sizes are necessary to detect or exclude association as has been shown in association studies for other complex diseases, including myocardial infarction and type 2 diabetes. The typical odds ratios 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. 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 - 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 0.05-0.10. The current study represents such a large-scale association study of ischemic stroke, with 4565 individuals from two population-based studies.

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, of which one identified an association with a locus on chromosome 4q25 previously associated with atrial fibrillation, 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 four GWA studies for coronary heart disease has been comprehensively replicated. We hypothesized the association of these SNPs with ischemic stroke, as previous studies have shown overlapping
heritability of myocardial infarction and ischemic stroke\textsuperscript{15}, most likely due to a common atherosclerotic pathogenesis.

\textbf{Materials and Methods}

\textit{Samples and definitions}

We used two population-based samples from the same geographic region in southern Sweden, Lund Stroke Register (LSR) and Malmö Diet and Cancer study (MDC). Sample characteristics, data collection and clinical definitions in both LSR\textsuperscript{16} and MDC\textsuperscript{17} have been described. Briefly, LSR is a prospective, epidemiologic 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\textsuperscript{st} 2001). Control subjects were randomly selected from the same region, 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 which included 28 449 randomly selected men (born between 1923-1945) and women (born between 1923-1950) at baseline examinations between 1991 and 1996. Subjects with ischemic stroke after the baseline examination were identified in the Stroke Register of Malmö\textsuperscript{18}. Subjects with 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 were ascertained in accordance with WHO criteria\textsuperscript{19}. Individuals with intracerebral or subarachnoid hemorrhage and unspecified stroke were excluded. In MDC, myocardial infarction prior to 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 (ICD9) 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 use of antihypertensive medications. Diabetes was defined as self-reported history of a physician’s 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 six SNPs recently associated with coronary artery disease in GWA studies; the two SNPs from the study by McPherson et al9 (rs10757274 and rs2383206), the three SNPs from the study by Helgadottir et al10 (rs10116277, rs1333040 and rs2383207) and the most significant SNP from the study by Samani et al11 (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 MALDI-TOF mass spectrometer (SEQUENOM MassArray) 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 >0.05 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 prior to the baseline exam. Statistical analyses were performed in SPSS (SPSS v15, SPSS Inc., Chicago, Illinois).

Linkage Disequilibrium (LD) and Hardy-Weinberg equilibrium (HWE) were examined in Haploview v4.0.

The authors had full access to the data and take responsibility for its integrity. All authors have read and agree to the manuscript as written.

**Results**

We included 2784 individuals from LSR and 1781 individuals from MDC. Characteristics of the two samples are shown in Table 1. Ischemic stroke occurred on average 6.3 years after inclusion in MDC (SD 7). Results of genotyping are shown in Table 2. Four SNPs passed quality control criteria while two were excluded (rs10116277 and rs2383206) due to a call rate <0.95 in either sample. One SNP differed significantly from HWE in the full sample, rs1333040 ($p=7\times10^{-4}$), 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 0.45-0.49, which is in good agreement with frequencies seen for the same SNPs in the sample of European ancestry (CEU) in the HapMap project.
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 two SNPs were significant in each sample separately. Only rs1333040 was not significant in any analysis. Association of the three SNPs remained significant after Bonferroni correction for four SNPs in the pooled sample (p<0.0125). Adjustment for matching variables (age, sex) did not alter the magnitude of the odds ratio or p-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 prior to the baseline exam in MDC decreased odds ratios slightly and widened confidence intervals resulting in a modest decrease in p-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 two less significant SNPs were also in strong LD (r^2=0.86) whereas the non-significant 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 two 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. 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 and found increased evidence of association.

Three previous studies have examined 9p21 in ischemic stroke of which two small studies found no association whereas a modestly sized study found a weak but significant association, interestingly with the same unadjusted odds ratio as for the strongest SNP in our study but with a higher standard error of the log-odds ratio and with similarly high MAF.

The current study has two 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. Secondly, our study includes only individuals from the same geographic region in southern Sweden, a region with a uniform population 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 the current study is the lack of comprehensive data on concurrent coronary heart disease. Data on MI were available only from one of the cohorts (MDC) and at baseline of screening rather than at the time of stroke. In patients with TIA 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. 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, i.e. 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 non-replication of reported findings, mirroring results for other complex traits. 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 etiological subclassifications of ischemic stroke, such as the TOAST classification, are necessary to identify common genetic determinants but so far have resulted mostly in additional multiple testing burden in already underpowered studies. 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 1.2-1.3), consistent with a more complex phenotype and the effect size of 9p21 SNPs may very well be larger in etiological subgroups. Third, the association between ischemic stroke and SNPs at a locus previously associated with coronary heart disease and both intracranial and abdominal aneurysms suggest that ischemic stroke shares common pathophysiological pathways with these diseases of the arterial wall. The nature of this pathway remains unknown since 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 atherosclerosis 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. 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, non-coding antisense RNA gene, ANRIL, which has been shown to regulate expression of CDKN2B and to be expressed in atherosclerotic...
plaques\textsuperscript{14}. All SNPs in our study are in intronic regions of ANRIL 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 endothelial\textsuperscript{34} and immunological\textsuperscript{35} cells which are thought to play central roles in plaque instability\textsuperscript{36}. Finally, establishment that common genetic variants confer risk of ischemic stroke gives hope for ongoing adequately powered genome-wide association 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 insights into pathophysiology and novel therapeutic targets.

**Acknowledgements**

The authors wish to thank Hélène Pessah-Rasmussen, MD, PhD, Dept of Neurology, Malmö University Hospital for assessment of stroke patients in the Stroke Register of Malmö and Ann-Cathrin Jönsson, RN, PhD, Dept of Neurology, Lund University Hospital for help with inclusion of stroke patients in the LSR study. We also wish to thank all participants in Malmö Diet and Cancer and Lund Stroke Register for making this study possible.

**Funding sources**

We gratefully acknowledge financial support from 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; the Swedish Stroke Association and Lund University.

Disclosures

None.
References


Table 1  *Description of 2784 LSR participants and 1781 MDC participants by case-control status.*

<table>
<thead>
<tr>
<th></th>
<th>Lund Stroke Register</th>
<th>Malmö Diet and Cancer</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Stroke cases</td>
<td>Controls</td>
</tr>
<tr>
<td>Count</td>
<td>1837</td>
<td>947</td>
</tr>
<tr>
<td>Age</td>
<td>73.4 (12.0)</td>
<td>73.2 (11.9)</td>
</tr>
<tr>
<td>Gender</td>
<td>46 % male</td>
<td>43 % male</td>
</tr>
<tr>
<td>Diabetes</td>
<td>26 %</td>
<td>7 %</td>
</tr>
<tr>
<td>Hypertension</td>
<td>66 %</td>
<td>48 %</td>
</tr>
<tr>
<td>Current smoking</td>
<td>19 %</td>
<td>10 %</td>
</tr>
<tr>
<td>Previous myocardial</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>infarction</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Age is shown as mean (standard deviation). MDC characteristics were recorded at baseline.
Table 2. *Genotyping results for six 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>C / G</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 21,957,751-21,984,490 and 21,992,902-21,999,312 respectively whereas the noncoding antisense RNA is located at 21,984,790-22,111,094. MAF indicates Minor Allele Frequency. Hardy-Weinberg Equilibrium (HWE) was examined in controls using Haploview with p-values presented for SNPs with call rate ≥ 0.95 in both samples.
Table 3  Results of association analysis of ischemic stroke and four SNPs at 9p21

<table>
<thead>
<tr>
<th>SNP</th>
<th>Lund Stroke Register</th>
<th>Malmö Diet and Cancer</th>
<th>Pooled</th>
<th>Source</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>OR (95% CI)</td>
<td>p-value</td>
<td>OR (95% CI)</td>
<td>p-value</td>
</tr>
<tr>
<td>rs10757274</td>
<td>1.14 (1.01-1.27)</td>
<td>0.029</td>
<td>1.16 (1.02-1.33)</td>
<td>0.029</td>
</tr>
<tr>
<td>rs2383207</td>
<td>1.12 (1.01-1.25)</td>
<td>0.044</td>
<td>1.19 (1.04-1.36)</td>
<td>0.010</td>
</tr>
<tr>
<td>rs1333049</td>
<td>1.11 (0.99-1.24)</td>
<td>0.064</td>
<td>1.18 (1.03-1.35)</td>
<td>0.017</td>
</tr>
<tr>
<td>rs1333040</td>
<td>0.94 (0.84-1.05)</td>
<td>0.28</td>
<td>0.93 (0.81-1.06)</td>
<td>0.28</td>
</tr>
</tbody>
</table>

Results are presented as crude odds ratios per SNP (95% confidence intervals and p-values).
Table 4  Results of association analyses after adjustment for clinical covariates

<table>
<thead>
<tr>
<th>Model</th>
<th>rs10757274</th>
<th>rs2383207</th>
<th>rs1333049</th>
<th>rs1333040</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pooled sample</td>
<td>1.15 (1.05-1.26, p=0.003)</td>
<td>1.16 (1.05-1.27, p=0.002)</td>
<td>1.14 (1.04-1.25, p=0.006)</td>
<td>0.93 (0.85-1.02, p=0.13)</td>
</tr>
<tr>
<td>MDC</td>
<td>1.14 (0.99-1.32, p=0.07)</td>
<td>1.19 (1.03-1.38, p=0.02)</td>
<td>1.15 (1.00-1.34, p=0.06)</td>
<td>0.94 (0.81-1.09, p=0.40)</td>
</tr>
</tbody>
</table>

Results are presented as adjusted odds ratios (95% CI and p-values). In the pooled sample adjustments for matching variables (age, sex) and cardiovascular risk factors (hypertension, diabetes and current smoking) were made. In MDC further adjustment for baseline myocardial infarction was made.
Common Genetic Variants on Chromosome 9p21 Confers Risk of Ischemic Stroke: A Large-Scale Genetic Association Study

J. Gustav Smith, Olle Melander, Häkan Lökvist, Bo Hedblad, Gunnar Engström, Peter Nilsson, Joyce Carlson, Göran Berglund, Bo Norrving and Arne Lindgren

_Circ Cardiovasc Genet._ published online February 12, 2009;

_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/early/2009/02/12/CIRCGENETICS.108.835173

Permissions: Requests for permissions to reproduce figures, tables, or portions of articles originally published in _Circulation: Cardiovascular Genetics_ 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 _Circulation: Cardiovascular Genetics_ is online at:
http://circgenetics.ahajournals.org/subscriptions/