Chromosome 9p21 in Ischemic Stroke
Population Structure and Meta-Analysis

Christopher D. Anderson, MD*; Alessandro Biffi, MD*; Natalia S. Rost, MD; Lynelle Cortellini, MSc; Karen L. Furie, MD, MPH; Jonathan Rosand, MD, MSc

Background and Purpose—Sequence variants on chromosome 9p21.3 are implicated in coronary artery disease and myocardial infarction, but studies in ischemic stroke have produced inconsistent results. We investigated whether these conflicting findings were due to false-positive studies confounded by population stratification or false-negative studies that failed to account for effects specific to certain stroke subtypes.

Methods—After assessing for population stratification at 9p21.3 using genomewide data, we meta-analyzed 8 ischemic stroke studies. This analysis focused on 2 single nucleotide polymorphisms, rs1537378 and rs10757278, because these variants are in strong linkage disequilibrium with most single nucleotide polymorphisms analyzed in prior studies of the region.

Results—Principal component analysis of the genomewide data showed no evidence of population stratification at that locus. Meta-analysis confirmed that both rs1537378 and rs10757278 are risk factors for ischemic stroke (ORs, 1.09 [P=0.0014] and 1.11 [P=0.001], respectively). Subtype analysis revealed a substantial increase in the effect of each single nucleotide polymorphism for risk of large artery stroke, achieving an effect size similar to that seen in coronary artery disease/myocardial infarction.

Conclusions—Variants on 9p21.3 are associated with ischemic stroke, and restriction of analysis to large artery stroke increases effect size toward that observed in prior association studies of coronary artery disease/myocardial infarction. Previous inconsistent findings are best explained by this subtype specificity rather than any unmeasured confounding by population stratification. (Stroke. 2010;41:1123-1131.)

Key Words: atherosclerosis ■ cardiac disease ■ genetics ■ ischemic stroke ■ meta-analysis ■ methodology
All individuals were of self-reported European ancestry, which might otherwise lead to confounding by population stratification at chromosome 9p21.3. We therefore performed principal component analysis of GWAS data. The Institutional Review Boards of participating institutions approved the protocols for each of these studies. Both cohorts were genotyped using the Affymetrix 6.0 platform at the Broad Institute (Cambridge, Mass). A total of 60,451 SNPs was selected to perform this analysis based on the following filters: missingness <0.1%, minor allele frequency >5%, Hardy-Weinberg equilibrium probability value <0.0001, and LD of $r^2 <0.2$ with every other SNP in the data set. To capture rs10757278 and rs1537378 (not directly genotyped on the Affymetrix 6.0 platform), we imputed SNPs at the 9p21.3 locus (range 21920Kb to 22125kb) using MACH 1.0 (www.sph.umich.edu/csg/abecasis/MaCH/index.html) and HapMap Phase 2 (Release 22) phased haplotypes. Both rs10757278 and rs1537378 achieved very high imputation quality with observed: expected allele dosage ratios >0.95. Multidimensional scaling procedures were used to assess population stratification with principal components using PLINK Version 1.07 (http://pngu.mgh.harvard.edu/~purcell/plink).

Genotyping Methods
In all included candidate gene studies, rs10757278 and rs1537378 (or their proxies in LD with $r^2 >0.8$) were captured through targeted genotyping. In included GWAS, both rs10757278 and rs1537378 were imputed, because these SNPs are not covered by any commercially available genomewide platform.

Meta-Analysis
Results from allele-based additive model logistic regression analyses in individual studies of rs10757278 or rs1537378 were meta-analyzed using a conservative random-effects pooling method (DerSimonian-Laird). Data published using nonadditive genetic models were excluded. Cochran Q test was used to estimate heterogeneity followed by calculation of $I^2$ (percentage of effect size attributable to heterogeneity). Effect size heterogeneity was significant for heterogeneity probability values <0.10 or $I^2 >0.20$. To assess the potential impact of any population stratification at rs10757278 or rs1537378 on association tests for ischemic stroke, we substituted our newly published candidate gene data from Boston for both loci in the following meta-analyses (Figures 3 and 4).

Results
Population Stratification at Chromosome 9p21.3
We performed principal component analysis on data from the MGH and the MIGen samples and extracted principal components 1 and 2 (PC1 and PC2). At rs10757278, neither PC1 nor PC2 was associated with genotype (Spearman $P >0.05$). At rs1537378, there was minimal association between principal components and genotype. PC1 showed a correlation coefficient of $-0.10$ (Spearman, $P =0.0067$), whereas PC2 showed a correlation coefficient of 0.08 (Spearman, $P =0.011$).

To further evaluate the possibility that this minimal population stratification at 9p21.3 had confounded published association studies of stroke, we substituted our newly available MGH–MIGen GWAS results for the previously published candidate gene data from Boston for both loci in the following meta-analyses (Figures 3 and 4).
Meta-Analysis of rs10757278 in Ischemic Stroke

Meta-analysis results (Figure 3A–C) of 8 studies (9632 cases and 30,716 control subjects) demonstrated an association between rs10757278 and ischemic stroke (OR = 1.11, 95% CI 1.05 to 1.17, P < 0.0001), although significant heterogeneity was present (I² = 0.37, 95% CI 0.12 to 0.68, P = 0.03). To reduce heterogeneity, we performed a second meta-analysis restricted to studies with sufficient information to allow subtype-specific evaluation. We substituted MGH–MIGen data for Boston MGH data in Gschwendtner et al, as described previously. Despite a substantial reduction in sample size (2993 cases and 21,533 control subjects in 5 studies), we observed an increase in OR to 1.15 (95% CI 1.08 to 1.23, P < 0.0001) with minimal heterogeneity (I² = 0.02, 95% CI 0.00 to 0.25, P = 0.68). Restricting the analysis further to the LA subtype exclusively (1290 cases and 5496 control subjects in 3 cohorts) resulted in a larger effect with an OR = 1.20 (95% CI 1.08 to 1.33, P = 0.0006) and further reduced heterogeneity (I² = 0.00, 95% CI 0.00 to 0.02, P = 0.92), demonstrating that this variant’s effect on LA stroke is consistent across studies (Figure 5A).

To assess whether the observed association of rs10757278 with ischemic stroke was due to confounding by underlying CAD, we performed a sensitivity analysis using only data that had previously been controlled for coincident CAD/MI (either by adjustment or by removal of affected individuals). This restricted meta-analysis (5152 cases and 19,678 control subjects) revealed an OR = 1.16 (95% CI 1.09 to 1.25, P < 0.0001) and minimal heterogeneity (I² = 0.00, 95% CI 0.00 to 0.32, P = 0.78). Restriction of this sensitivity analysis to LA strokes (938 cases and 4262 control subjects) yielded an OR of 1.18 (95% CI 1.04 to 1.31, P = 0.0068) with very limited heterogeneity (I² = 0.00, 95% CI 0.00 to 0.23, P = 0.87). The results of the most recent meta-analysis of rs10757278 in CAD/MI are shown for reference in Figure 5A to allow comparison with our observations.

Table. Individual Studies Included in the Meta-Analysis

<table>
<thead>
<tr>
<th>Study</th>
<th>SNP</th>
<th>Genotype</th>
<th>Ethno</th>
<th>Site</th>
<th>Pheno</th>
<th>Case</th>
<th>Ctrl</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ding et al12</td>
<td>rs10757278</td>
<td>Direct</td>
<td>CH</td>
<td>Wuhan</td>
<td>LA+SV</td>
<td>558</td>
<td>554</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Direct</td>
<td>CH</td>
<td>Hubel</td>
<td>LA+SV</td>
<td>440</td>
<td>490</td>
</tr>
<tr>
<td>Gschwendtner et al7</td>
<td>rs10757278</td>
<td>Direct</td>
<td>EU-AA</td>
<td>Multiple</td>
<td>AIS</td>
<td>4095</td>
<td>4262</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Direct</td>
<td>EU</td>
<td>Aberdeen</td>
<td>LA</td>
<td>113</td>
<td>500</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Direct</td>
<td>EU</td>
<td>Baltimore</td>
<td>LA</td>
<td>31</td>
<td>269</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Direct</td>
<td>EU</td>
<td>Jacksonville</td>
<td>LA</td>
<td>20</td>
<td>105</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Direct</td>
<td>EU</td>
<td>London</td>
<td>LA</td>
<td>229</td>
<td>873</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Direct</td>
<td>EU</td>
<td>Munich</td>
<td>LA</td>
<td>311</td>
<td>1167</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Direct</td>
<td>AA</td>
<td>Baltimore</td>
<td>LA</td>
<td>28</td>
<td>226</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Direct</td>
<td>AA</td>
<td>Jacksonville</td>
<td>LA</td>
<td>19</td>
<td>104</td>
</tr>
<tr>
<td></td>
<td>rs1537378</td>
<td>Direct</td>
<td>EU-AA</td>
<td>Multiple</td>
<td>AIS</td>
<td>4091</td>
<td>4202</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Direct</td>
<td>EU</td>
<td>Aberdeen</td>
<td>LA</td>
<td>122</td>
<td>504</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Direct</td>
<td>EU</td>
<td>Baltimore</td>
<td>LA</td>
<td>26</td>
<td>379</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Direct</td>
<td>EU</td>
<td>Jacksonville</td>
<td>LA</td>
<td>19</td>
<td>104</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Direct</td>
<td>EU</td>
<td>London</td>
<td>LA</td>
<td>236</td>
<td>853</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Direct</td>
<td>EU</td>
<td>Munich</td>
<td>LA</td>
<td>314</td>
<td>1185</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Direct</td>
<td>AA</td>
<td>Baltimore</td>
<td>LA</td>
<td>28</td>
<td>266</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Direct</td>
<td>AA</td>
<td>Jacksonville</td>
<td>LA</td>
<td>97</td>
<td>312</td>
</tr>
<tr>
<td>Helgadottir et al5</td>
<td>rs10757278</td>
<td>Imputed</td>
<td>EU</td>
<td>Iceland</td>
<td>LA+CE</td>
<td>415</td>
<td>278*</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Imputed</td>
<td>EU</td>
<td>Sweden</td>
<td>LA+CE</td>
<td>290</td>
<td>213*</td>
</tr>
<tr>
<td>Hu et al8</td>
<td>rs10757278</td>
<td>Proxy</td>
<td>CH</td>
<td>China</td>
<td>LA</td>
<td>352</td>
<td>423</td>
</tr>
<tr>
<td>Ikram et al15</td>
<td>rs1537378</td>
<td>Imputed</td>
<td>EU</td>
<td>Multiple</td>
<td>AIS</td>
<td>1164</td>
<td>18438</td>
</tr>
<tr>
<td>Karvanen et al13</td>
<td>rs10757278</td>
<td>Proxy</td>
<td>EU</td>
<td>Multiple</td>
<td>AIS</td>
<td>109</td>
<td>2064</td>
</tr>
<tr>
<td>MGH–MIGen7,15</td>
<td>rs10757278</td>
<td>Imputed</td>
<td>EU</td>
<td>MGH</td>
<td>LA</td>
<td>105</td>
<td>1417</td>
</tr>
<tr>
<td></td>
<td>rs1537378</td>
<td>Imputed</td>
<td>EU</td>
<td>MGH</td>
<td>LA</td>
<td>105</td>
<td>1417</td>
</tr>
<tr>
<td>Lemmens et al10</td>
<td>rs10757278</td>
<td>Direct</td>
<td>EU</td>
<td>Multiple</td>
<td>AIS</td>
<td>648</td>
<td>828</td>
</tr>
<tr>
<td>Smith et al14</td>
<td>rs10757278</td>
<td>Proxy</td>
<td>EU</td>
<td>Scandinavia</td>
<td>AIS</td>
<td>2725</td>
<td>1840</td>
</tr>
<tr>
<td>Wahlstrand et al10</td>
<td>rs10757278</td>
<td>Direct</td>
<td>EU</td>
<td>Scandinavia</td>
<td>AIS</td>
<td>526</td>
<td>1†</td>
</tr>
</tbody>
</table>

*These analyses were performed after exclusion of ischemic stroke cases with a history of CAD/MI.
†The cohort in Wahlstrand et al was prospectively followed for incident ischemic stroke, and only the total sample size at baseline was reported in the original publication.
AA indicates African American; AIS, all ischemic strokes; CE, cardioembolic stroke; CH, Chinese Han; Ctrl, control; EU, European; Ethno, ethnicity; Pheno, phenotype.
Meta-Analysis of rs1537378 in Ischemic Stroke

Results for rs1537378 in stroke were only reported in 2 studies.7,15 Meta-analysis for rs1537378 in these studies (5255 cases and 22640 control subjects) demonstrated an association with ischemic stroke (OR = 1.09, 95% CI 1.03 to 1.14, \( P = 0.0014 \)) with limited heterogeneity across studies (\( I^2 = 0.00, 95\% \) CI 0.0 to 0.20, \( P = 0.67 \); Figure 4A–C). We subsequently restricted the meta-analysis to LA stroke reported by Gschwendtner and colleagues7 (MGH–MIGen GWAS data substituted for the original Boston data) and to GWAS results for all ischemic stroke subtypes published by Ikram and colleagues15 (total sample size of 2105 cases and 22259 control subjects). We observed an increase in OR to 1.12 (95% CI 1.05 to 1.18, \( P = 0.0003 \)) at the expense of an increase in heterogeneity (\( I^2 = 0.08, 95\% \) CI 0.00 to 0.65, \( P = 0.36 \)).

Further restriction of the meta-analysis to include only LA strokes required the removal of data from Ikram and colleagues due to absence of TOAST subtypes. Novel meta-
analysis results after inclusion of MGH–MI Gen GWAS data (OR = 1.23, 95% CI 1.12 to 1.33, \( P < 0.001 \); Figure 4C) were comparable to previous findings by Gschwendtner et al (OR = 1.20, 95% CI 1.07 to 1.34, \( P < 0.001 \)) without evidence of heterogeneity as evaluated by computation of \( I^2 \) (0.00, 95% CI 0.00 to 0.00, \( P = 0.99 \)). All samples in this LA stroke analysis were controlled for CAD/MI, and so a separate analysis was not needed for this SNP. For comparison of the association between rs1537378 and stroke with the association for CAD/MI, the result of a recent CAD GWAS at rs7865618 (in high LD with rs1537378) is shown in Figure 5B.4

**Discussion**

Our analyses provide novel evidence that variants on chromosome 9p21.3 influence risk of stroke and that this risk is largely related to LA stroke. These findings suggest that discordance between published results at this locus is likely to be due to lack of subtype information in most of the studies rather than population stratification. Restriction to 1 subtype reduced heterogeneity, thereby improving the ability of our meta-analysis to estimate the true effect size. As a result, the magnitude of the OR increases substantially when analysis is restricted to LA cases (Figure 5).

The 9p21.3 locus has been associated with atherosclerotic disease of the coronary arteries and abdominal aorta.3,20 An LA subtype-specific effect for 9p21.3 therefore seems biologically reasonable. Furthermore, the size of the effect on LA stroke appears to be similar to that in MI/CAD.4,19 Although environmental risk factors predispose to both LA disease and CAD/MI, our observation that the effect on stroke persists in meta-analysis restricted to studies controlling for underlying CAD/MI makes confounding by cardiac disease unlikely.
The impact of subtype-specific effects on design of GWAS has substantial implications for study power. If a variant were to have an all-stroke risk of 1.05 and an LA stroke risk of 1.2 (assuming 33% of strokes are LA subtype), 7800 incident ischemic strokes would be required to detect any effect in a study without subtype information. In a population-based study, this would require a longitudinal cohort of 156,000 individuals followed over 10 years.

9p21.3 contains multiple genes, several of which reside in a single LD block (Figure 2, bottom panel). Our results provide new evidence for multiple causal variants at this locus, because we identified independent associations with ischemic stroke for 2 SNPs in weak LD with one another. Targeted resequencing of 9p21.3 will likely be required to elucidate the roles played by genes at this locus in CAD, MI, and ischemic stroke.

Our study included an analysis of population stratification at the 9p21.3 locus because this phenomenon could explain the lack of GWAS replication for findings from targeted genotyping studies. The MGH–MIGen data used to perform this analysis were not available at the time of prior studies of 9p21.3. We found no population stratification at rs10757278 and negligible population stratification at rs1537378. Adjustment for population stratification did not alter previously published results or introduce additional heterogeneity in the meta-analysis. This result does not support the hypothesis that the lack of association for 9p21.3 in published stroke GWAS reflected that prior candidate gene studies were false-positive due to failure to control for unmeasured population stratification.

Our study has limitations. Several publications identified in our initial literature search could not be included in our...
Figure 5. Effect size estimates (OR), sample size, and meta-analysis heterogeneity for rs10757278 (A) and rs1537378 (B). Restriction of cases to LA only results in increased effect size and decreased heterogeneity across studies for both SNPs. LA’s only effect persists despite controlling for MI and approaches the OR observed in association studies for CAD/MI.4,19 Candidate Studies indicates Gschwendner et al7; Ikram GWAS, Ikram et al15; LA, LA stroke.
meta-analysis (Figure 1). Of the included studies, the majority could not be used for subtype analyses due to lack of recording of TOAST subtypes. Despite the large number of cases and controls amassed in meta-analysis of rs10757278 and rs1537378, neither SNP achieved genomewide significance levels ($P < 5 \times 10^{-8}$). Therefore, this meta-analysis cannot be considered definitive. Finally, although results from our unpublished genomewide data reject the hypothesis of population stratification affecting associations at 9p21.3, we cannot fully rule out this possibility in other published cohorts. However, the effect size concordance and observed minimal heterogeneity between MGH–MI Gen GWAS data and candidate gene study results from European and non-European cohorts makes additional unidentified population stratification unlikely.

Appendix: PubMed Search Terms

9p21
9p21.3
Stroke
Atherosclerotic stroke
Large artery stroke
Cerebral infarction
CAD
CHD
CVD
MI
Myocardial infarction
Atherosclerosis
Thrombosis
CDKN2A/CDKN2B
INK4
MTAP
ARF
ANRIL

All possible combinations of listed search terms were used and results compared.

Acknowledgments

We thank the Myocardial Infarction Genetics Consortium (MI Gen) study for the use of their genotype data as control data in our study. The MI Gen study was funded by the US National Institutes of Health and National Heart, Lung, and Blood Institute’s STAMPEED genomics research program (R01 HL087676) and a grant from the National Center for Research Resources.

Sources of Funding

Supported by the American Heart Association/Bugher Foundation Centers for Stroke Prevention Research (0775010N), Deane Institute for Integrative Study of Atrial Fibrillation and Stroke, US National Institutes of Health and National Heart, Lung, and Blood Institute’s STAMPEED genomics research program (R01 HL087676), and The National Center for Research (US4 RR020278).

Disclosures

None.


Chromosome 9p21 in Ischemic Stroke: Population Structure and Meta-Analysis
Christopher D. Anderson, Alessandro Biffi, Natalia S. Rost, Lynelle Cortellini, Karen L. Furie and Jonathan Rosand

Stroke. 2010;41:1123-1131; originally published online April 15, 2010;
doi: 10.1161/STROKEAHA.110.580589
Stroke is published by the American Heart Association, 7272 Greenville Avenue, Dallas, TX 75231
Copyright © 2010 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/41/6/1123

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/