Effect of Genetic Variants Associated With Plasma Homocysteine Levels on Stroke Risk

Ioana Cotlarciuc, PhD*; Rainer Malik, PhD*; Elizabeth G. Holliday, PhD; Kourosh R. Ahmadi, PhD; Guillaume Paré, MD; Bruce M. Psaty, MD; Myriam Fornage, PhD; Nazeeha Hasan, PhD; Paul E. Rinne, MSc; M. Arfan Ikram, PhD; Hugh S. Markus, DM; Jonathan Rosand, MD; Braxton D. Mitchell, PhD; Steven J. Kittner, MD; James F. Meschia, MD; Joyce B.J. van Meurs, PhD; Andre G. Uitterlinden, PhD; Bradford B. Worrall, MD; Martin Dichgans, MD; Pankaj Sharma, PhD, MD, FRCP; on behalf of METASTROKE and the International Stroke Genetics Consortium

Background and purpose—Elevated total plasma homocysteine (tHcy) levels are known to be associated with increased risk of ischemic stroke (IS). Given that both tHcy and IS are heritable traits, we investigated a potential genetic relationship between homocysteine levels and stroke risk by assessing 18 polymorphisms previously associated with tHcy levels for their association with IS and its subtypes.

Methods—Previous meta-analysis results from an international stroke collaborative network, METASTROKE, were used to assess association of the 18 tHcy-associated single-nucleotide polymorphisms (SNPs) in 12,389 IS cases and 62,004 controls. We also investigated the associations in regions located within 50 kb from the 18 tHcy-related SNPs and the association of a genetic risk score, including the 18 SNPs.

Results—One SNP located in the RASIP1 gene and a cluster of 3 SNPs located at and near SLC17A3 were significantly associated with IS (P=0.0003) after correcting for multiple testing. For stroke subtypes, the sentinel SNP located upstream of MUT was significantly associated with small-vessel disease (P=0.0022), whereas 1 SNP located in MTHFR was significantly associated with large-vessel disease (P=0.00019). A genetic risk score, including the 18 SNPs, did not show significant association with IS or its subtypes.

Conclusions—This study found several potential associations with IS and its subtypes: an association of an MUT variant with small-vessel disease, an MTHFR variant with large-vessel disease, and associations of RASIP1 and SLC17A3 variants with overall IS. *(Stroke. 2014;45:1920-1924.)*

Key Words: genetic association studies ■ genetic risk score ■ homocysteine ■ stroke

The relationship between total plasma homocysteine (tHcy) levels and stroke risk has been investigated by numerous observational studies, which together provide compelling evidence that elevated homocysteine levels are associated with an increased risk of ischemic stroke (IS).1 However, residual confounding and reverse causation impair causal inference from the results of observational studies.2 Mendelian randomization studies investigating a potential causal relationship between tHcy and IS risk have yielded inconclusive results.3

Received from February 17, 2014; final revision received March 31, 2014; accepted April 4, 2014.

From the Department of Medicine, Imperial College Cerebrovascular Research Unit, Imperial College London, London, United Kingdom (I.C., N.H., P.E.R., P.S.); Institute for Stroke and Dementia Research, Medical Centre, Klinikum der Universität München, Ludwig-Maximilians-University, Munich, Germany (R.M., M.D.); Center for Clinical Epidemiology and Biostatistics, School of Medicine and Public Health, University of Newcastle, Callaghan, New South Wales, Australia (E.G.H.); Department of Nutritional Sciences, Faculty of Health and Medical Sciences, University of Surrey, Guildford, United Kingdom (K.R.A.); Department of Pathology, McMaster University, Hamilton, Ontario, Canada (G.P.); Departments of Epidemiology, Medicine, and Health Services, University of Washington, Seattle (B.M.P.); Group Health Research Institute, Group Health, Seattle, WA (B.M.P.); University of Texas Health Science Center at Houston (M.F.); Departments of Epidemiology, Neurology, and Radiology (M.A.I.) and Internal Medicine (J.B.J.v.M., A.G.U.), Erasmus MC University Medical Center, Rotterdam, The Netherlands; Department of Clinical Neurosciences, University of Cambridge, Cambridge, United Kingdom (H.S.M.); Department of Neurology (J.R.) and Center for Human Genetic Research (J.R.), Massachusetts General Hospital, Boston; Program in Medical and Population Genetics, Broad Institute, Cambridge, MA (J.R.); Veterans Affairs Medical Center, Baltimore, MD (B.D.M., S.J.K.); Departments of Medicine (B.D.M., S.J.K.) and Neurology (S.J.K.), University of Maryland School of Medicine, Baltimore; Department of Neurology, Mayo Clinic, Jacksonville, FL (J.F.M.); Departments of Neurology (B.B.W.) and Public Health Science (B.B.W.), University of Virginia, Charlottesville; and Munich Cluster for Systems Neurology, Munich, Germany (M.D.).

* Drs Cotlarciuc and Malik contributed equally.

The online-only Data Supplement is available with this article at http://stroke.ahajournals.orglookup/suppl/doi:10.1161/STROKEAHA.114.005208/-/DC1.

Correspondence to Ioana Cotlarciuc, PhD, Department of Medicine, Imperial College Cerebrovascular Research Unit, Imperial College London, Fulham Palace Rd, London W6 8RF, United Kingdom. E-mail i.cotlarciuc@imperial.ac.uk

© 2014 American Heart Association, Inc.

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

DOI: 10.1161/STROKEAHA.114.005208
Recent studies showed that lowering homocysteine levels through vitamin B (folic acid and vitamin B12) intervention reduced the risk of stroke in patients with normal renal function and with normal vitamin B12 metabolism.4–6 These findings may explain the negative results obtained by earlier studies,7,8 where renal function and vitamin B12 metabolism status were not taken into account. In addition, the folic acid fortification of grains in the United States and some European countries may have also reduced the benefit of folic acid intervention trials.9 Thus, the relationship between stroke and homocysteine is complex, and careful consideration is required in designing future clinical trials.10

Recently, a genome-wide association study (GWAS) meta-analysis of plasma homocysteine levels in 44,147 individuals of European ancestry identified 13 associated genetic loci at genome-wide significance (P<5×10−8).11 Within the 13 associated loci, this meta-analysis identified 18 independent single-nucleotide polymorphisms (SNPs) explaining 6% of variation in tHcy levels. Since both tHcy and IS have a genetic component, the reported association between tHcy and stroke risk may result from shared genetic risk factors. Thus, we investigated the 18 SNPs previously associated with homocysteine levels for their association with IS and its subtypes: large-vessel disease (LVD), small-vessel disease (SVD), and cardioembolic stroke. We also evaluated association of a genotype risk score (GRS), including the 18 tHcy-associated SNPs for association with IS and its subtypes.

Materials and Methods

Study Population

The study population included 12,389 IS cases and 62,004 controls of European ancestry from 15 cohorts contributing to the META STROKE collaboration.11 Details of the study designs for each participating study are included in the online-only Data Supplement.

SNP Selection

Plasma homocysteine-associated SNPs reaching genome-wide significance (P<5×10−8) in a published meta-analysis of GWAS11 were selected for inclusion. In total, 18 independent SNPs (r²<0.2) were found to be significantly associated with tHcy at a P<5×10−8 and selected for assessment in ischemic stroke and its subtypes (Table I in the online-only Data Supplement). As a secondary analysis, to account for potential population differences in linkage disequilibrium (LD) between functional and tag-SNPs, we included additional SNPs located ±50 kb around the 18 tHcy-associated SNPs, and a total of 3,160 SNPs were selected of which 166 variants were independent (r²<0.2).

Statistical Analysis

Summary statistics for association of the 18 tHcy-associated SNPs with IS and its subtypes were provided by META STROKE Consortium. Details on genotyping, imputation, and quality control methods are provided in the Methods in the online-only Data Supplement. Logistic regression was performed to test association of individual SNPs with IS and its subtypes assuming an additive model and adjusting for study-specific covariates: age, sex, and ancestry principal components (Methods in the online-only Data Supplement).

Considering that stroke subtypes are independent of each other,12–16 we prespecified a Bonferroni-adjusted significance threshold of α=0.0027 (where α=0.05/18 SNPs) to adjust for primary analyses testing 18 independent SNPs for association with IS and its subtypes.

For the secondary analysis of SNPs located within ±50 kb of the 18 tHcy-related SNPs, we specified a Bonferroni-adjusted significance threshold of 0.0003 (α=0.05/166=0.0003, for 166 independent SNPs tested).

We then tested the association between an additive GRS of the 18 homocysteine-associated SNPs and increased risk of IS and its subtypes. The GRS was calculated using a previously described method (Methods in the online-only Data Supplement).

Using the CaTS (Power Calculator for Two Stage Association Studies) genetic power calculator,14 our study had 80% power to detect odds ratio (OR) of 1.06 to 1.11 for IS, and subtype-specific OR of 1.13 to 1.23 for variants with allele frequency 10% to 50% at a Bonferroni-corrected threshold of 0.0027 (α=0.05/18).

Results

Study Population Characteristics

The discovery meta-analysis of IS studies included 12,389 cases and 62,004 controls of European descent. Stroke subtypes include cardioembolic, LVD, and SVD accounted for 19% (n=2365), 17.4% (n=2167), and 15.2% (n=1894), respectively, of all IS cases. Detailed characteristics of the participating studies have been summarized previously.15

Association of tHcy-Associated SNPs

With Overall IS

First, we assessed the association of each tHcy-associated SNP with IS risk and subsequently with its subtypes. Then, we tested the combined effect of the 18 sentinel tHcy-associated SNPs on risk of IS and its subtypes.

For overall IS, the OR for the 18 tested SNPs ranged from 0.96 to 1.04 (Table II in the online-only Data Supplement). Of the 18 SNPs tested, 2 SNPs located at/near FUT1 (rs838133: OR, 1.04; 95% confidence interval [CI], 1.00–1.07; P=0.045) and CPS1 (rs7422339: OR, 0.96; 95% CI, 0.92–0.99; P=0.045) were associated with IS at a nominal P value (0.05), but did not pass Bonferroni-corrected P=0.0027 (P=0.05/18).

The combined GRS, including the 18 independent tHcy SNPs, did not show a significant association with IS (OR, 1.02; 95% CI, 0.91–1.15; P=0.63; Table III in the online-only Data Supplement).

Association of tHcy-Associated SNPs

With IS Subtypes

Next, we investigated the association of the tHcy-related SNPs with IS subtypes. For SVD, 2 SNPs located near MUT (rs9369898: OR, 1.12; 95% CI, 1.04–1.21; P=0.0022) and FUT1 (rs838133: OR, 1.07; 95% CI, 1.00–1.15; P=0.04), respectively, were nominally associated at P<0.05 (Table IV in the online-only Data Supplement). The variant located near MUT (rs9369898) also passed Bonferroni-corrected value P=0.0027. The major allele A of rs9369898 associated with higher tHcy levels was also associated with increased risk of SVD. There was no evidence of between-study heterogeneity for rs9369898 (P=7.4%; P-het=0.37). The GRS, including the 18 independent tHcy SNPs, did not show an association with SVD risk (OR, 1.1; 95% CI, 0.85–1.42; P=0.43; Table III in the online-only Data Supplement).

For LVD, 1 SNP (rs838133: OR, 1.08; 95% CI, 1.01–1.16; P=0.018) located near FUT1 gene was associated at a nominal P value (P<0.05) but did not pass Bonferroni correction for
multiple testing (Table V in the online-only Data Supplement). The GRS of the 18 independent tHcy SNPs did not show an association with LVD risk (OR, 1.06; 95% CI, 0.82–1.35; \( P=0.64 \); Table III in the online-only Data Supplement). None of the tested SNPs were associated with cardioembolic risk even at a nominal \( P \) value (\( P<0.05 \); Table VI in the online-only Data Supplement). In addition, the GRS of the 18 independent tHcy SNPs did not show an association with cardioembolic risk (OR, 0.9; 95% CI, 0.71–1.14; \( P=0.4 \); Table III in the online-only Data Supplement).

### Investigation of Associations in Regions Located ±50 kb Around the 18 tHcy-Associated SNPs

As a secondary analysis, we assessed the associations with IS and its subtypes for SNPs located within ±50 kb of the 18 tHcy-associated SNPs.

For overall IS, 3 variants (rs9379800, rs17271121, and rs12664474) located within 50 kb from the tHcy-associated polymorphism, rs548987, were associated with IS at \( P \) values lower than the Bonferroni-corrected threshold (\( P=0.05/166=0.0003 \); Table). Two SNPs (rs9379800 and rs12664474) located upstream of \( SLC17A3 \) were highly correlated (\( r=0.766 \)) with each other, and moderately correlated with the third SNP, rs17271121, located in an intron of \( SLC17A3 \) (\( r^{2}=0.545 \)). None of the 3 SNPs were in LD with the tHcy-associated polymorphism, rs548987 (\( r^{2}=0.035 \)).

In addition, another SNP rs2287921 located in an intron of \( RASPI1 \) gene, within 50 kb from the \( FUT1 \) polymorphism, rs838133, was associated with IS at a \( P=0.0002 \) (OR, 0.94; 95% CI, 0.91–0.97), lower than Bonferroni correction for multiple testing (\( P<0.0003 \)). This SNP was in low LD with the 2 tHcy sentinel SNPs (\( r^{2}=0.658 \)) with the sentinel SNP rs838133, which may suggest that this could be a broader risk region spanning the 2 neighboring genes \( RASPI1 \) and \( FUT1 \).

For LVD, 1 SNP (rs1801131), a missense variant located in \( MTHFR \) gene, near the 2 sentinel SNPs in this gene, rs1801133 and rs12134663, was associated with LVD with \( P=1.92 \times 10^{-4} \) (OR, 1.15; 95% CI, 1.06–1.23) lower than the Bonferroni correction for multiple testing (\( P<0.0003 \)). This missense SNP was in low LD with the 2 tHcy sentinel SNPs (\( r^{2} \mid rs1801131 \) and rs1801133\)=0.19 and \( r^{2} \mid rs1801131 \) and rs12134663\)=0.268).

For SVD, 1 SNP (rs566295) located upstream of \( MUT \), 44 kb from the tHcy-associated polymorphism, rs9369898, was associated with SVD with \( P=2.2 \times 10^{-4} \) (OR, 0.87; 95% CI, 0.80–0.93) passing Bonferroni correction for multiple testing. This SNP was in low LD with the \( MUT \) polymorphism, rs9369898 (\( r^{2}=0.264 \)).

For cardioembolic stroke, no significant associations were observed at a threshold exceeding Bonferroni correction for multiple testing (\( P<0.0003 \)).

### Discussion

This large study of 12389 IS cases and 62004 controls has identified several potential novel associations with IS and its subtypes by testing previously reported associations with homocysteine levels in stroke. We found evidence of an association of \( MUT \) gene with SVD, an association of \( MTHFR \) gene with LVD, and associations of \( RASPI1 \) and \( SLC17A3 \) with overall IS.

Of the 18 tHcy polymorphisms investigated, 1 polymorphism located upstream of \( MUT \) gene was significantly associated with SVD, whereas none of the 18 tHcy-related SNPs were significantly associated with LVD, cardioembolic, or overall IS. The allele correlated with increased tHcy levels at \( MUT \) gene showed to be also associated with increased risk of SVD, suggesting a potential small but significant effect on SVD risk.

On a closer inspection of this region, another SNP located 44 kb from the sentinel SNP and in low LD with the sentinel SNP was also associated with SVD. This polymorphism was also significantly associated with homocysteine levels at a genome-wide significance level (\( P=2.27 \times 10^{-9} \)), but conditional analysis has not been conducted to establish if these 2 polymorphisms were independently influencing homocysteine levels. These 2 polymorphisms may thus potentially be correlated with either a single or multiple regulatory variants in this region that modulate both tHcy levels and SVD risk.

### Table. Association With IS and Its Subtypes of SNPs Located ±50 kb From the 18 Total Plasma Homocysteine–Associated SNPs at a \( P<0.0003 \) Obtained After Adjustment for Multiple Testing

<table>
<thead>
<tr>
<th>SNP</th>
<th>Sentinel SNPs</th>
<th>Chr</th>
<th>Risk Locus</th>
<th>CA</th>
<th>Freq</th>
<th>( \beta )</th>
<th>SE</th>
<th>( P ) Value</th>
<th>OR</th>
<th>OR (95 L)</th>
<th>OR (95 U)</th>
<th>( P ) _Het</th>
<th>( P )</th>
<th>Q</th>
</tr>
</thead>
<tbody>
<tr>
<td>IS</td>
<td>rs9379800</td>
<td>6</td>
<td>SLC17A3</td>
<td>G</td>
<td>0.17</td>
<td>0.087</td>
<td>0.0233</td>
<td>1.65 \times 10^{-4}</td>
<td>1.09</td>
<td>1.13</td>
<td>1.05</td>
<td>0.63</td>
<td>0</td>
<td>9.74</td>
</tr>
<tr>
<td></td>
<td>rs17271121</td>
<td>6</td>
<td>SLC17A3</td>
<td>C</td>
<td>0.13</td>
<td>0.089</td>
<td>0.0246</td>
<td>2.8 \times 10^{-4}</td>
<td>1.09</td>
<td>1.13</td>
<td>1.05</td>
<td>0.42</td>
<td>0</td>
<td>13.24</td>
</tr>
<tr>
<td></td>
<td>rs12664474</td>
<td>6</td>
<td>SLC17A3</td>
<td>G</td>
<td>0.17</td>
<td>0.082</td>
<td>0.0226</td>
<td>2.86 \times 10^{-4}</td>
<td>1.08</td>
<td>1.12</td>
<td>0.89</td>
<td>0.07</td>
<td>2</td>
<td>10.21</td>
</tr>
<tr>
<td></td>
<td>rs2287921</td>
<td>19</td>
<td>FUT1</td>
<td>C</td>
<td>0.55</td>
<td>0.061</td>
<td>0.0166</td>
<td>2.09 \times 10^{-4}</td>
<td>1.06</td>
<td>1.09</td>
<td>1.03</td>
<td>0.08</td>
<td>27.2</td>
<td>20.61</td>
</tr>
<tr>
<td>LVD</td>
<td>rs1801131</td>
<td>1</td>
<td>MTHFR</td>
<td>T</td>
<td>0.68</td>
<td>0.141</td>
<td>0.0378</td>
<td>1.92 \times 10^{-4}</td>
<td>1.15</td>
<td>1.06</td>
<td>1.23</td>
<td>0.89</td>
<td>0</td>
<td>6.20</td>
</tr>
<tr>
<td></td>
<td>rs1801133</td>
<td>12</td>
<td>MUT</td>
<td>C</td>
<td>0.54</td>
<td>0.138</td>
<td>0.0374</td>
<td>2.2 \times 10^{-4}</td>
<td>1.13</td>
<td>1.2</td>
<td>1.07</td>
<td>0.84</td>
<td>0</td>
<td>4.88</td>
</tr>
</tbody>
</table>

\( \beta \) indicates \( \beta \) coefficient; 95 L and 95 U, lower and upper 95 percentile; CA, coded allele; Chr, chromosome; Freq, frequency of the coded allele; \( P \), Higgins heterogeneity index; IS, ischemic stroke; LVD, large-vessel disease; OR, odds ratio; \( P \) _Het, \( P \) value for the Cochran heterogeneity statistic; Q, Cochran heterogeneity statistic; SNP, single-nucleotide polymorphism; and SVD, small-vessel disease.
The *MUT* gene is known to encode the mitochondrial enzyme methylmalonyl Coenzyme A mutase, a vitamin B12–dependent enzyme. Considering that vitamin B12 is an important cofactor in homocysteine metabolism, a potential pleiotropic effect of *MUT* gene on both plasma homocysteine and vitamin B12 levels was suggested previously. In addition, vitamin B12 deficiency is highly prevalent in SVD and may underlie blood–brain barrier damage, leading to small-vessel dysfunction, especially periventricular white matter lesions, suggesting a potential role of *MUT* polymorphisms in SVD via mechanisms involving vitamin B12 deficiency.

Moreover, hyperhomocysteinemia as an independent risk factor for SVD may act via endothelial dysfunction, suggesting that homocysteine-lowering therapy may be particularly effective in this stroke subgroup. In support of this, the VITamins TO Prevent Stroke (VITATOPS) trial found that after a 3-year vitamin B supplementation period, the risk of stroke was reduced only in patients with symptomatic SVD. Therefore, considering that in our study the only homocysteine significant association was with SVD, brings more evidence to support the hypothesis that homocysteine may be a risk factor in particular to SVD.

Our study also found an association with LVD, a missense variant located in *MTHFR* (A1298C), 2 kb from the well-studied homocysteine-associated polymorphism *MTHFR* C677T. Previous studies showed that A1298C was associated with a decrease in *MTHFR* activity but was not associated with increased homocysteine levels. The functional differences between the 2 polymorphisms could be explained by their location: C677T is located within the N-terminal catalytic region, whereas A1298C is located within the C-terminal regulatory domain, being involved in enzyme regulation.

Given that high folate levels have been associated with a reduced effect of *MTHFR C677T* on homocysteine levels and stroke, the nonassociation of *MTHFR C677T* with stroke could be explained by the inclusion in the METAstudy of individuals from countries where folic acid fortification has already been implemented. The relationship between *MTHFR* A1298C and folate status has not yet been investigated, and it is not known whether this polymorphism is influenced by folate levels. However, if this SNP is not influenced or influenced at a lesser extent by folate status when compared with C677T, it may explain why A1298C polymorphism, and not C677T, was found to be associated with LVD in our stroke cohort.

For overall IS, 2 highly correlated SNPs located upstream of *SLC17A3* and 1 SNP located in an intron of *SLC17A3* were found to be associated with IS after correction for multiple testing. None of these SNPs were in LD with the sentinel polymorphism, suggesting the presence of another independent potential risk locus in this region.

*SLC17A3* encodes a voltage-driven transporter that excretes intracellular urate and organic anions from the blood into renal tubule cells. A significant association between a polymorphism (rs1165205) located in intron 1 of *SLC17A3* and serum uric acid concentrations has been found by a GWAS study. This SNP was located ≤80 kb from our investigated SNPs but was in low LD (r^2<0.2) with all our *SLC17A3* SNPs. Therefore, the possibility exists that these polymorphisms are correlated with another variant that has a causal role on influencing urate levels. Alternatively, there may be several independent loci associated with urate levels in this region as it has been suggested previously, and our investigated SNPs may be correlated with another polymorphism influencing urate levels independent from the GWAS reported association. Epidemiological studies have shown that elevated serum uric acid is a strong independent risk factor for hypertension, and considering that hypertension is a risk factor for stroke, it has been suggested that increased uric acid levels may be involved in predicting stroke risk.

Another association with IS was a variant located in *RASIP1* gene. *RASIP1* is required for the proper formation of vascular structures that develop via both vasculogenesis and angiogenesis. As it is well known that insufficient vessel growth or maintenance can lead to stroke among other disorders, we provide a possible link between *RASIP1* and stroke risk. Furthermore, this polymorphism has been reported to be significantly associated with retinal vascular caliber in a previous GWAS. Therefore, considering that changes in retinal vascular caliber are associated with cardiovascular diseases, including IS, it provides more evidence of a potential role of this gene in stroke.

To date, several GWAS have been conducted on IS and its subtypes, which have identified associations of common polymorphisms specific to each stroke subtype, endorsing the fact that different stroke subtypes have different risk factor profiles and pathophysiological mechanisms. Our study supported this hypothesis and identified a potential association of *MUT* with SVD and of *MTHFR* with LVD. In addition, we also identified potential associations of *RASIP1* and *SLC17A3* with overall IS. The reason why we did not find significant associations of *RASIP1* and *SLC17A3* with any of the stroke subtypes could be because of the smaller sample sizes and thus reduced power of the stroke subtype cohorts. However, it is important to highlight that our findings need to be validated by replication in independent cohorts to avoid considering spurious results.

In addition, the combined genetic score of the 18 independent tHcy SNPs did not show a significant association with overall IS or its subtypes. The low power of our study to detect small effects, together with the fact that stroke and homocysteine may only partly share their allelic architecture, may explain the lack of association of the tHcy GRS.

However, our results apply to the European population and to countries with established policies of folic acid fortification of grains. A recent study has shown that in these countries, the evidence from genetic studies and from randomized trials with folic acid suggested no benefit from lowering homocysteine levels for stroke prevention. Thus, to elucidate the controversial role of homocysteine in stroke, additional studies should be conducted in regions with low folate levels where homocysteine-lowering interventions may have an important effect in reducing stroke risk.

To conclude, our study provides evidence of several potential associations with IS and its subtypes: an association of *MUT* gene with small-vessel stroke, an association of *MTHFR* gene with large-vessel stroke, and associations of *RASIP1* and *SLC17A3* with overall IS, highlighting possible roles of these genes in IS and its subtypes.
Acknowledgments

We gratefully acknowledge the METASTROKE Collaboration.12

Sources of Funding

This study was supported by the Department of Health (United Kingdom), Henry Smith Charity, British Council, United Kingdom-India Education and Research Institute, and the Qatar National Research Foundation. Genotyping of the cohorts within METASTROKE was supported by the Wellcome Trust, UK Medical Research Council, Australian National and Medical Health Research Council, National Institutes of Health, including National Heart, Lung, and Blood Institute, the National Institute on Aging, the National Human Genome Research Institute, the National Institute of Neurological Disorders and Stroke, and the American Heart Association-Buher Foundation Centers for Stroke Prevention.

Disclosures

Dr Worrall received support from US National Institute of Neurological Disorders and Stroke and National Human Genome Research Institute. Dr Rosand received support from National Institutes of Health and he is a consultant for Boehringer Ingelheim. Drs Cotfaruci and Sharma received support from the Qatar National Research Fund NPRP grant no. 4-154-3-050. The other authors report no conflicts.

References

Effect of Genetic Variants Associated With Plasma Homocysteine Levels on Stroke Risk

Stroke. 2014;45:1920-1924; originally published online May 20, 2014;
doi: 10.1161/STROKEAHA.114.005208

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/45/7/1920

Data Supplement (unedited) at:
http://stroke.ahajournals.org/content/suppl/2014/05/20/STROKEAHA.114.005208.DC1
ONLINE SUPPLEMENT

Effect of genetic variants associated with plasma homocysteine levels on stroke risk

Ioana Cotlarchiu, PhD*, Rainer Malik, PhD*, Elizabeth G. Holliday, PhD, Kouros Ahmadi, PhD, Guillaume Pare, MD, Bruce M. Psaty, MD, Myriam Fornage, PhD, Nazeeha Hasan, PhD, Paul Rinne, MSc, M. Arfan Ikram, MD, PhD, Hugh S Markus, DM, Jonathan Rosand, MD, MSc, Braxton D. Mitchell PhD, Steven J. Kittner MD, MPH, James F. Meschia MD, Joyce BJ van Meurs, PhD, Andre G Uitterlinden, PhD, Bradford B. Worrall MD, MSc, Martin Dichgans, MD, Pankaj Sharma, PhD, MD, FRCP, on behalf of METASTROKE and the International Stroke Genetics Consortium

* These authors contributed equally to this work.

Corresponding author: i.cotlarchiu@imperial.ac.uk
Contents

Supplemental Methods........................................................................................................................................3
Study Population..................................................................................................................................................3
Discovery Stroke Cohorts ..................................................................................................................................3
Genotyping and quality control of the stroke cohorts .........................................................................................13
Statistical Analysis.............................................................................................................................................19
Supplemental tables..........................................................................................................................................21
Table I. Association results of the top independent associated SNPs with plasma homocysteine levels at a genome-wide significance level ($P<5\times10^{-8}$)..................................................................................................................21
Table II. Association results of the 18 tHcy associated SNPs with Ischemic Stroke (IS) .................................22
Table III. Association results of a Multi-SNP Genotypic Risk Score (consisting of 18 tHcy associated SNPs) with Ischemic Stroke (IS) and its subtypes LVD, SVD, CE.................................................................26
Table IV. Association results of the 18 tHcy associated SNPs with small vessel disease (SVD)........23
Table V. Association results of the 18 tHcy associated SNPs with large vessel disease (LVD)..................24
Table VI. Association results of the 18 tHcy associated SNPs with cardioembolic (CE) stroke..............25
Supplemental references..................................................................................................................................27
Supplemental Methods

Study Population

The study population included 12,389 ischemic stroke cases and 62,004 controls from 15 cohorts contributing to the METASTROKE collaboration. The 15 participating cohorts are: The Atherosclerosis Risk in Communities study (ARIC), Australian Stroke Genetics Collaborative (ASGC), Bio-Repository of DNA in Stroke (BRAINS), Cardiovascular Health Study (CHS), deCODE Genetics, Framingham Heart Study (FHS), Genetics of Early-Onset Stroke (GEOS), Heart Protection Study (HPS), The Heart and Vascular Health Study (HVH), The Ischemic Stroke Genetics Study/Sibling with Ischaemic Stroke Study (ISGS/SWISS), The MGH Genes Affecting Stroke Risk and Outcome Study (MGH-GASROS), Milano, Rotterdam, The Wellcome Trust Case-Control Consortium II Munich (WTCCC2-Munich), The Wellcome Trust Case-Control Consortium II UK (WTCCC2-UK). All participants were of European ancestry and were based in Europe, North America, and Australia. Controls were free of stroke and of matching ancestry to cases.

Stroke subtypes were assigned by using the TOAST classification system into 3 major aetiological categories: cardioembolic stroke (CE), large artery disease (LVD) and small vessel disease (SVD). Of the 15 participating studies, only HPS and Rotterdam did not perform TOAST subtyping. In addition, CHS had only classified CE patients, while FHS had information only on two stroke subtypes: CE and SVD.

Discovery Stroke Cohorts

The Atherosclerosis Risk in Communities Study (ARIC)

The ARIC study is a prospective population-based study of atherosclerosis and clinical atherosclerotic diseases in 15,792 men and women, including 11,478 non-Hispanic white participants, drawn from 4 U.S. communities (Suburban Minneapolis, Minnesota; Washington County, Maryland; Forsyth County, North Carolina, and Jackson, Mississippi). In the first three communities, the sample reflects the demographic composition of the community. In Jackson, only black residents were enrolled. Ancestry was self-reported during an interview. Over 99% self-identified as either white or black. Only self-identified whites were included in the analyses. Hospitalized strokes that occurred by December 31, 2007 (median follow-up, 18.7 years) were included in the analyses. During annual telephone contacts, trained interviewers asked each ARIC participant to list all hospitalizations during the past year. Hospital records for any hospitalizations identified were then obtained. Moreover, all local hospitals annually provided lists of stroke discharges (International Classification of Diseases, Ninth Revision, Clinical Modification codes 430 to 438), which were surveyed for ARIC participant discharges. Details on quality assurance for ascertainment and classification of stroke are described elsewhere. Briefly, the stroke diagnosis was assigned according to criteria adapted from the National Survey of Stroke. Strokes secondary to trauma, neoplasm, hematologic abnormality, infection, or vasculitis were excluded, and a focal deficit lasting <24 hours was not considered to
be a stroke. Out-of-hospital stroke was not ascertained and validated; thus, these potential stroke events are not included. Strokes were classified into hemorrhagic stroke (subarachnoid and intracerebral hemorrhage) and ischemic stroke (thrombotic and embolic brain infarction). A stroke was classified as ischemic when a brain CT or MRI revealed acute infarction and showed no evidence of hemorrhage. All definite ischemic strokes were further classified as lacunar, nonlacunar thrombotic, or cardioembolic on the basis of the recorded neuroimaging results. A stroke was classified as “lacunar” when 2 criteria were met: (1) typical location of the infarct (basal ganglia, brain stem, thalamus, internal capsule, or cerebral white matter); and (2) infarct size of ≤2 cm or unstated size.6 Definite or probable “cardioembolic” stroke required either (1) autopsy evidence of an infarcted area in the brain and a source of possible cerebral emboli in a vessel or the presence of an embolus in the brain or (2) medical record evidence of a possible noncarotid source of embolus such as moderate or greater valvular heart disease, atrial fibrillation, cardiac or arterial procedure (eg, cardiac catheterization, open heart surgery, cerebral angiography, and carotid endarterectomy), or intracardiac thrombus. Definite or probable ischemic strokes that were not classified as lacunar or cardioembolic, including atherothrombotic and unclassified thrombotic strokes, were labeled “nonlacunar.” Hemorrhagic strokes identified by ARIC were censored at the time of their occurrence.

**Australian Stroke Genetics Collaborative (ASGC)**

ASGC stroke cases comprised stroke patients of European ancestry who were admitted to four clinical centers across Australia (The Neurosciences Department at Gosford Hospital, Gosford; the Neurology Department at John Hunter Hospital, Newcastle; The Queen Elizabeth Hospital, Adelaide; and the Royal Perth Hospital, Perth) between 2003 and 2008.7 Stroke was defined by World Health Organization criteria as a sudden focal neurological deficit of vascular origin, lasting more than 24 h and confirmed by imaging, such as computerized tomography (CT) and/or magnetic resonance imaging (MRI) brain scan. Other investigative tests such as electrocardiogram, carotid doppler and trans-esophageal echocardiogram were conducted to define ischemic stroke mechanism as clinically appropriate. Cases were excluded from participation if they were aged <18 years, were diagnosed with hemorrhagic stroke or had transient ischemic attack rather than ischemic stroke or if they were unable to undergo baseline brain imaging. On the basis of these criteria, a total of 1,230 ischemic stroke cases were included in the current study. Ischemic stroke subtypes were assigned using TOAST criteria on the basis of clinical, imaging and risk factor data. ASGC controls were participants in the Hunter Community Study (HCS), a population-based cohort of individuals aged 55–85 years, predominantly of European ancestry and residing in the Hunter Region in New South Wales, Australia. Detailed recruitment methods for the HCS have been previously described25. Briefly, participants were randomly selected from the New South Wales State electoral roll and were contacted by mail between 2004 and 2007. Consenting participants completed five detailed self-report questionnaires and attended the HCS data collection center, at which time a series of clinical measures were obtained. A total of 1,280 HCS participants were genotyped for the current study. All study participants gave informed consent for participation in genetic studies. Approval for the individual studies was obtained from the relevant institutional ethics committees.
Bio-Repository of DNA in stroke (BRAINS)

BRAINS is an ongoing, multicentre, in-hospital study which recruits consenting acute stroke patients into a highly characterized biobank. All adult (>18 years of age) stroke patients are recruited with either ischaemic or haemorrhagic pathology MRI confirmed lesions. Ischaemic stroke subtypes are further sub-classified according to TOAST criteria. Known monogenic causes of stroke are excluded. BRAINS has two principal arms. The first arm recruits UK European stroke patients while the second arm recruits South Asian stroke patients from multiple sites in the UK and also from sites in India. Control data for the European arm is provided by the Wellcome Trust Case Control Consortium while control subjects for the South Asian arm are recruited simultaneously as the affected stroke patient and usually is the proband's spouse.

Cardiovascular Health Study (CHS)

The CHS is a population-based cohort study of risk factors for coronary heart disease (CHD) and stroke in adults ≥ 65 years conducted across four field centers in the United States. The original predominantly white cohort of 5,201 persons (4,964 whites) was recruited in 1989-1990 from a random sample of people on Medicare eligibility lists and an additional 687 African-Americans were enrolled subsequently (1992-93) for a total sample of 5,888. Race was determined by self-identification at interview. In addition to the 5 categories used in the ARIC study, participants were also asked a second question as to whether they considered themselves to be of Hispanic origin. To reduce the possibility of confounding by population structure, these analyses were limited to participants of self-described European-ancestry. The study sample for these analyses includes participants who were free of CVD at baseline, had blood samples drawn at their baseline examination, consented to genetic testing, had DNA available, and had successful genome-wide genotyping assay.

Participants were examined annually from enrollment to 1999, and since then continue to be under surveillance for stroke. Since baseline, participants have also been contacted twice a year to identify potential cardiovascular events, including stroke. In addition, all hospitalizations were screened for potential stroke events. For suspected events, information was collected from the participant or next of kin, from medical records, and, if needed, from the participant's physician. When available, CT and/or MRI scans or reports were reviewed centrally. Final adjudication of the occurrence of stroke, stroke types, and subtypes was undertaken by vascular neurologists at a consensus conference using all available information.

Strokes were classified as ischemic if there was imaging (CT or MRI within 4 weeks), surgical or autopsy evidence excluding a hemorrhage, or in the absence of such direct evidence (in <10% of cases in FHS and Rotterdam Study, none in CHS) if the preponderance of indirect evidence (e.g. deficit limited to one limb or completely resolved within 72 hours, atrial fibrillation in persons not on anticoagulants) suggested the event was an ischemic rather than a hemorrhagic stroke. A stroke was classified as hemorrhagic if there was imaging, surgical, lumbar puncture or autopsy evidence of hemorrhage, and in the absence of direct evidence to the contrary, when the participant lost consciousness permanently or died within hours after onset of focal signs. The stroke type was defined as unknown if there was insufficient information available to categorize the event as ischemic or hemorrhagic. All ischemic and hemorrhagic strokes and strokes of unknown type were included in the analyses of total stroke with one exception: subarachnoid
hemorrhages (n=28 across all studies) were excluded from these analyses since the heritability, risk factors and pathophysiologic mechanisms underlying subarachnoid hemorrhages are distinctly different from other stroke subtypes. Persons with a subarachnoid hemorrhage were censored at the time of the event.

Only known ischemic strokes were included in the analysis of ischemic stroke. In secondary analyses we related those SNPs which reached genome-wide significance in our initial GWAS to the specific ischemic stroke subtype of atherothrombotic stroke, also called atherothrombotic brain infarction (ABI). We used the best available definitions of definite and possible ABI in each cohort; both large artery atherosclerotic strokes and small-vessel or lacunar strokes were included in this phenotype, events known to be cardioembolic were excluded. For the analysis of ABI, participants were censored when they developed an alternative type of stroke.

**Framingham Heart Study (FHS)**

The FHS is a three-generation, single-site, community-based, ongoing cohort study that was initiated in 1948 to investigate prospectively the risk factors for CVD including stroke. It now comprises 3 generations of participants (N=10,333): the Original cohort followed since 1948, their Offspring and spouses of the Offspring, followed since 1971; and children from the largest Offspring families enrolled in 2000 (Gen 3). Gen 3 participants were not included in this analysis since they are young (mean age 40±9 years) and few have suffered strokes. The Original cohort enrolled 5209 men and women who comprised two-thirds of the adult population then residing in Framingham, MA. Survivors continue to receive biennial examinations. The Offspring cohort comprises 5124 persons (including 3514 biological offspring) who have been examined approximately once every 4 years. The population of Framingham was virtually entirely white (Europeans of English, Scots, Irish and Italian descent) in 1948 when the Original cohort was recruited. At the initial examination participants were asked for country of birth and whether or not they had any Italian ancestry. At a later examination (the 8th) the Offspring cohort participants were asked to identify their race from the following choices: Caucasian or white, African-American or black, Asian, Native Hawaiian or other Pacific Islander, American Indian or Alaska native or ‘prefer not to answer’. They were also asked to identify their ethnicity as either ‘Hispanic or Latino’ or neither. Almost all the FHS Original and Offspring participants are white/Caucasian and none were excluded from the discovery cohort.

At each clinic exam, participants receive questionnaires, physical examinations and laboratory testing; between examinations they remain under surveillance (regardless of whether or not they live in the vicinity) via physician referrals, record linkage and annual telephone health history updates. Incident strokes have been identified since 1948 through this ongoing system of FHS clinic and local hospital surveillance and methods used have been detailed previously; they include review of medical records and collaboration with local general practitioners, emergency rooms and imaging facilities. If a participant saw a physician or was admitted to the hospital, visited an emergency room or obtained any brain imaging between biennial examinations for symptoms suggestive of TIA or stroke, a stroke neurologist from the Heart Study attempted to visit the person within 48 hours and recorded a complete history and neurological examination; this was repeated at 1, 3 and 6 months. All medical records from practitioners, hospitals, imaging centers, rehabilitation centers and nursing homes were procured for review. A panel of 3
investigators (at least 2 neurologists) adjudicated the diagnosis of stroke and determined stroke subtype in each case based on the Framingham evaluations and external records. The recruitment of Original and Offspring cohort participants at FHS had occurred long before the DNA collection with the result that a large number of stroke events in the FHS (although ascertained prospectively) were prevalent at the time of DNA collection and were excluded from these analyses. While this reduced the sample size from FHS, the meta-analyses presented here focused on incident events.

Strokes were classified as ischemic if there was imaging (CT or MRI within 4 weeks), surgical or autopsy evidence excluding a hemorrhage, or in the absence of such direct evidence (in <10% of cases in FHS and Rotterdam Study, none in CHS) if the preponderance of indirect evidence (e.g. deficit limited to one limb or completely resolved within 72 hours, atrial fibrillation in persons not on anticoagulants) suggested the event was an ischemic rather than a hemorrhagic stroke. A stroke was classified as hemorrhagic if there was imaging, surgical, lumbar puncture or autopsy evidence of hemorrhage, and in the absence of direct evidence to the contrary, when the participant lost consciousness permanently or died within hours after onset of focal signs. The stroke type was defined as unknown if there was insufficient information available to categorize the event as ischemic or hemorrhagic. All ischemic and hemorrhagic strokes and strokes of unknown type were included in the analyses of total stroke with one exception: subarachnoid hemorrhages (n=28 across all studies) were excluded from these analyses since the heritability, risk factors and pathophysiologic mechanisms underlying subarachnoid hemorrhages are distinctly different from other stroke subtypes. Persons with a subarachnoid hemorrhage were censored at the time of the event.

Only known ischemic strokes were included in the analysis of ischemic stroke. In secondary analyses we related those SNPs which reached genome-wide significance in our initial GWAS to the specific ischemic stroke subtype of atherothrombotic stroke, also called atherothrombotic brain infarction (ABI). We used the best available definitions of definite and possible ABI in each cohort; both large artery atherosclerotic strokes and small-vessel or lacunar strokes were included in this phenotype, events known to be cardioembolic were excluded. For the analysis of ABI, participants were censored when they developed an alternative type of stroke. A lacunar stroke was diagnosed based on either clinical presentation of a typical lacunar syndrome, or a clinical picture compatible with a lacunar lesion and imaging showing a small ischemic lesion in the territory of the deep penetrating arteries. An ischemic stroke was categorized as cardioembolic if there was evidence of a cardiac or aortic source of embolization based on the clinical presentation, brain imaging and review of all available cardiac assessments including ECG, echocardiography and cardiac monitoring.

**deCODE Genetics**

Cases, irrespective of age, were identified from a registry of individuals diagnosed with ischemic stroke or TIA at Landspitali University Hospital in Reykjavik, the only tertiary referral centre in Iceland, during the years 1993 to 2006. The ischemic stroke or TIA diagnoses were based on standard WHO criteria and imaging evidence (either CT or MRI), and were clinically confirmed by neurologists. Eligible patients who survived the stroke were invited to participate the genetic study, either by attending a recruitment centre for deCODE’s genetic studies, or they were
visited at their home by a study nurse. Control subjects were participants from a large variety of genetic programs at deCODE. Individuals with confirmed stroke (identified by cross-matching with hospital lists), who had participated in genetic studies other than those of cardiovascular diseases (CVD) (but not participated in CVD studies) were excluded as controls.

The Genetics of Early Onset Stroke (GEOS) Study, Baltimore, USA

GEOS is a population-based case-control study designed to identify genes associated with early-onset ischemic stroke and to characterize interactions of identified stroke genes and/or SNPs with environmental risk factors. Participants were recruited from the greater Baltimore-Washington area in 4 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, Stroke Prevention in Young Men (SPYM) conducted from 2003-2007, and Stroke Prevention in Young Adults (SPYA) conducted in 2008. Case participants were hospitalized with a first cerebral infarction identified by discharge surveillance from one of the 59 hospitals in the greater Baltimore-Washington area and direct referral from regional neurologists. The abstracted hospital records of cases were reviewed and adjudicated for ischemic stroke subtype by a pair of neurologists according to previously published procedures with disagreements resolved by a third neurologist. 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. Control participants without a history of stroke were identified by random-digit dialing and were balanced to cases by age and region of residence in each recruitment periods. Genomic DNA was isolated from a variety of sample types, including cell line, whole blood, mouth wash and buccal swab. Samples were genotyped at the Johns Hopkins Center for Inherited Disease Research (CIDR) using the Illumina HumanOmni1-Quad_v1-0_B BeadChip (Illumina, San Diego, CA, USA). Individuals were excluded if they were unexpected duplicates, gender discrepancy and unexpected relatedness.

Heart Protection Study (HPS)

The Heart Protection Study (HPS) was a large randomized trial involving individuals at increased risk of vascular events. Between 1994-1997 20,536 men and women aged 40-80 years were recruited from 69 collaborating hospitals in the United Kingdom (with ethics committee approval). Participants were eligible for inclusion provided they had non-fasting blood total cholesterol concentrations of at least 135 mg/dL (3.5 mmol/L) and either a previous diagnosis of coronary disease, ischemic stroke, other occlusive disease of non-coronary arteries, diabetes mellitus, or if were men 65 years or older) treated hypertension. None of them was on statin therapy. At the initial screening visit, all participants provided written consent and began a “run-in” phase involving 4 weeks of placebo followed by 4 to 6 weeks of 40 mg simvastatin daily, after which compliant and eligible individuals were randomly allocated 40 mg simvastatin daily or matching placebo for approximately 5 years. Individuals entering HPS with a clinical diagnosis of ischemic stroke were used as cases in the METASTROKE study. Individuals entering HPS with pre-existing diabetes but no history of cerebrovascular disease, coronary heart disease or peripheral vascular disease were used as controls.
The Heart and Vascular Health Study (HVH)

The setting for this study was Group Health (GH), a large integrated health care system in western Washington State. Data were utilized from an ongoing case-control study of incident myocardial infarction (MI) and stroke cases with a shared common control group. Methods for the study have been described previously17-19 and are briefly summarized below. The study was approved by the human subjects committee at GH, and written informed consent was provided by all study participants.

All study participants were GH members and aged 30-79 years. MI and stroke cases were identified from hospital discharge diagnosis codes and were validated by medical record review. Controls were a random sample of GH members frequency matched to MI cases on age (within decade), sex, treated hypertension, and calendar year of identification. The index date for controls was a computer-generated random date within the calendar year for which they had been selected. For stroke cases, the index date was the date of admission for the first acute stroke. Participants were excluded if they were recent enrollees at GH, had a history of prior stroke, or if the incident event was a complication of a procedure or surgery.

Trained medical record abstractors collected eligibility and risk factor information from a review of the GH medical record using only data available prior to the index date and through a telephone interview. Medication use was ascertained using computerized GH pharmacy records. A venous blood sample was collected from all consenting subjects, and DNA was extracted from white blood cells using standard procedures. Diagnostic criteria for ischemic stroke were adopted from the Cardiovascular Health Study.20 These criteria included (1) rapid onset of neurologic deficit or subarachnoid hemorrhage, (2) deficit persisting for longer than 24 hours unless computed tomography or magnetic resonance imaging show evidence of permanent damage, and (3) no underlying brain trauma, tumor, or infection to cause symptoms. These analyses were limited to ischemic stroke cases, namely those satisfying 1 or more of the following criteria: (a) Focal deficit, without evidence of blood on CT or MRI, (b) Focal deficit, with mottled appearance in the appropriate location on CT, or (c) surgery or autopsy evidence of infarction.

Among ischemic strokes, the subtypes were defined as follows: Lacunar stroke (“SVD”) required either: (a) CT/MRI demonstrates a deep area of infarction (decreased density) less than 2 cm. across, or (b) A normal CT, but the clinical syndrome is typical of a lacunar infarction, that is: a pure motor stroke, a pure sensory stroke, hemiparesis plus ataxia, or dysarthria plus a clumsy hand. Embolic stroke (“CE”) required either (a) a recognized source of emboli such as atrial fibrillation, endocarditis, mitral stenosis, thrombus in heart, recent MI or cardiac surgery, or (b) a mottled appearance consistent with infarction on the CT. Atherosclerotic infarction (“LAA”) when there is no apparent source of emboli or evidence of lacunar infarction and there is evidence of large vessel atherosclerosis by carotid ultrasound or angiography.

ISGS/SWISS

ISGS is a multicenter inception cohort study of first-ever ischemic stroke in adult men and women.21 Cases were recruited from inpatient stroke services at five academic medical centers in Florida, Georgia, Virginia and Minnesota. The diagnosis of ischemic stroke was confirmed by a study neurologist on the basis of medical history, physical examination and CT or MR imaging
of the brain. Cases had to be enrolled within 30 days of onset of stroke symptoms. Cases were excluded if they had a mechanical aortic or mitral valve, central nervous system vasculitis, or bacterial endocarditis at the time of the stroke. They were also excluded if they were known to have: cerebral autosomal dominant arteriopathy with subcortical infarcts and leukoencephalopathy (CADASIL), Fabry disease, homocystinuria, mitochondrial encephalopathy with lactic acidosis and stroke-like episodes (MELAS), or sickle cell anemia. Stroke severity at enrollment was assessed using the NIH Stroke Scale (NIHSS) and outcomes at 90-days were assessed by telephone using the Barthel Index, Glasgow Outcome Scale, and the modified Rankin scale. Diagnostic evaluation included: head CT (95% of individuals enrolled) or MRI (83%), electrocardiography (92%), cervical arterial imaging (86%), and echocardiography (74%). A vascular neurology committee reviewed the medical records of every case and assigned ischemic stroke subtype diagnoses according to criteria from the Trial of ORG10172 (TOAST), the Oxfordshire Community Stroke Project, and the Baltimore-Washington Young Stroke Study. DNA was donated to the NINDS DNA Repository (Coriell Institute, Camden, NJ) for eligible samples with appropriate written informed consent. A separate certified neurologist adjudicator additionally assigned a subtype diagnosis using the standardized Causative Classification of Stroke web-based algorithm.

**SWISS** is a multicenter affected sibling pair study. Probands with ischemic stroke were enrolled at 66 US medical centers and 4 Canadian medical centers. Probands are adult men and women over the age of 18 years diagnosed with ischemic stroke confirmed by a study neurologist on the basis of history, physical examination and CT or MR imaging of the brain. Probands were required to have a history of at least one living sibling with a history of stroke. Probands were excluded if they had a mechanical aortic or mitral valve, central nervous system vasculitis, or bacterial endocarditis at the time of the index ischemic stroke. Probands were also excluded if they were known to have cerebral autosomal dominant arteriopathy with subcortical infarcts and leukoencephalopathy (CADASIL), Fabry disease, homocystinuria, mitochondrial encephalopathy with lactic acidosis and stroke-like episodes (MELAS), or sickle cell anemia. Siblings were enrolled using proband-initiated contact or direct contact when permitted by Institutional Review Boards. Concordant (affected) siblings had their diagnosis of ischemic stroke confirmed by review of medical records by a vascular neurology committee. Concordant siblings had the same eligibility criteria as probands. Subtype diagnoses were assigned to the index strokes of probands and concordant siblings according to TOAST criteria. Discordant siblings of the proband were confirmed to be stroke-free using the Questionnaire for Verifying Stroke-free Status. Lymphoblastoid cell lines were created on all subjects. A certified neurologist adjudicator additionally assigned a subtype diagnosis using the standardized Causative Classification of Stroke web-based algorithm to all concordant siblings and a subset of probands for whom medical records were available.

**The MGH Genes Affecting Stroke Risk and Outcome Study (MGH-GASROS)**

Cases were all consecutive patients aged ≥ 18 years presenting with ischemic stroke and admitted to the Massachusetts General Hospital (MGH) Stroke Unit through the Emergency Department, or evaluated in the MGH Neurology outpatient clinics, as well as on the inpatient Medical and Vascular Surgical services from January 2003 to July 2008. Only patients of
European ancestry (confirmed by principal component analysis using genome-wide SNP data) were included in the present analysis.

Ischemic stroke was defined as either (1) a radiographically proven (head CT or MRI) infarct associated with the appropriate clinical stroke syndrome, or (2) a fixed neurological deficit persisting more than 24 hours, consistent with a vascular pattern of involvement and without radiographic evidence of demyelinating disease, or other non-vascular structural disease. Patients with specific vascular disorders (vasculitis, subacute bacterial endocarditis, fibromuscular dysplasia, vasospasm) were excluded from the study. All subjects were evaluated by a neurologist upon presentation and provided informed consent. Clinical and laboratory data were collected during the admission for qualifying ischemic stroke event. Diagnostic work-up included: head CT (100%), brain MRI (90%), cervical and intracranial vessel imaging using CTA or MRA (75%), carotid and/or transcranial ultrasound (24%), echocardiography (86%), and Holter monitoring (16%). Controls were recruited among the stroke-free adults presenting to the MGH outpatient clinics and matched with the stroke cases on the basis of age, sex and ancestry information obtained from principal component analysis of GWAS data.

**Milano Study**

This study includes consecutive Italian patients referred to Besta Institute from 2000 to 2009 with stroke and included in the Besta Cerebrovascular Diseases Registry (CEDIR). Ischemic stroke cases, first ever or recurrent, confirmed on brain imaging, were selected for this study. All cases were of self-reported Caucasian ancestry and had clinically relevant diagnostic workup performed. All cases were phenotyped by an experienced stroke neurologist according to TOAST criteria, based on relevant clinical imaging and available information on cardiovascular risk factors. Controls are Italian individuals enrolled within the PROCARDIS Study, with no personal or sibling history of coronary heart disease before age 66 years.

**Rotterdam Study**

The Rotterdam Study is a population-based cohort study among inhabitants of a district of Rotterdam (Ommoord), The Netherlands, and aims to examine the determinants of disease and health in the elderly with a focus on neurogeriatric, cardiovascular, bone, and eye disease. In 1990-1993, 7,983 persons participated and were re-examined every 3 to 4 years. After enrollment in the Rotterdam Study, participants are continuously monitored for incident stroke through automated linkage of the study database with files from general practitioners. Nursing home physicians' files and files from general practitioners of participants who moved out of the district are scrutinized as well. Additional information is obtained from hospital records. Potential strokes are reviewed by research physicians and verified by an experienced stroke neurologist. Stroke is defined as rapidly developing clinical signs of focal or global disturbance of cerebral function with no apparent cause other than a vascular origin. History of stroke at baseline was assessed during the baseline interview and verified in medical records. Strokes are further classified as cerebral infarction or intracerebral hemorrhage based on neuroimaging reports. If neuroimaging is lacking, a stroke is classified as unspecified. Subarachnoid hemorrhages were excluded. Ischemic strokes are subtyped into large-vessel disease, small vessel disease, or cardio-embolic based on all available clinical and imaging data.
If insufficient information is present to reliably subtype into one of three categories, the ischemic stroke is classified as unspecified ischemic stroke. For the current study, too few cases of ischemic stroke subtypes were present to reliably perform statistical analyses.

Participants were followed from baseline to stroke, death, last health status update when they were known to be stroke-free, or January 1, 2005, whichever came first. Follow-up was complete up to January 1, 2005, for 99.1% of potential person-years.

Wellcome Trust Case-Control Consortium 2 (WTCCC2)

The WTCCC2 samples were genotyped as part of the WTCCC2 ischemic stroke study. Stroke cases included samples recruited by investigators at St. George's University London (SGUL) and University of Oxford in the UK and the Department of Neurology, Klinikum Großhadern, Ludwig-Maximilians-University, Munich. The SGUL collection comprised 1224 ischemic stroke samples from a hospital based setting. All cases were of self-reported Caucasian ancestry. Ischemic stroke subtypes were determined according to TOAST criteria based on relevant clinical imaging and available information on cardiovascular risk factors. The University of Oxford collection comprised 896 ischemic stroke cases, consecutively collected as part of the Oxford vascular study (OXVASC). Cases were of self-reported Caucasian ancestry, and ischemic stroke subtypes were determined according to TOAST criteria based on relevant clinical imaging. The Munich samples included 1383 ischemic stroke cases. Cases were consecutive European Caucasians recruited from a single dedicated Stroke Unit at the Department of Neurology, Klinikum Großhadern, Ludwig-Maximilians-University, Munich. Ischemic stroke subtypes were determined according to TOAST criteria based on relevant clinical and imaging data. Controls for the UK samples were drawn from shared WTCCC controls obtained from the 1958 Birth Cohort. This is a prospectively collected cohort of individuals born in 1958 (http://www.b58cgene.sgul.ac.uk/), and ascertained as part of the national child development study (http://www.cls.ioe.ac.uk/studies.asp?section=000100020003). Data from this cohort are available as a common control set for a number of genetic and epidemiological studies. For the German samples controls were Caucasians of German origin participating into the population KORAgen study (www.gsf.de/kora/en/english.html). This survey represents a gender- and age stratified random sample of all German residents of the Augsburg area and consists of individuals 25 to 74 years of age, with about 300 subjects for each 10-year increment. All controls were free of a history of stroke or transient ischemic attack.
Genotyping and quality control of the stroke cohorts

The Atherosclerosis Risk in Communities Study (ARIC)

Methods of genotyping, quality-control assessment and filtering, and genotype imputation have been described in 3. Briefly, genotyping was performed with the GeneChip SNP Array 6.0 (Affymetrix). Subject specific quality control filters included filters for call rate, heterozygosity, sex mismatch. SNP specific quality control filters included filters for call rate, minor allele frequency (MAF), Hardy-Weinberg equilibrium (HWE), and differential missingness by outcome or genotype. The set of genotyped input SNPs used for imputation was selected based on high quality GWA data. We used a callrate >95%, HWE p-value>5x10^-6; MAF>0.01. A total of 704,588 SNPs passing QC criteria were used for imputation, which was performed with the MaCH v1.0.16 software.

Australian Stroke Genetics Collaborative (ASGC)

The ASGC sample was genotyped using the Illumina HumanHap610-Quad array. Quality control excluded SNPs with genotype call rate <0.95, deviation from Hardy-Weinberg equilibrium (P<10^-6) or minor allele frequency <0.01. At the sample level, quality control excluded individuals with: (i) genotype call rate <95% (n=4); ii) genome-wide heterozygosity < 23.3% or > 27.2% (n=9); iii) inadequate clinical data or inconsistent clinical and genotypic gender (n=45) and; iv) an inferred first or second-degree relative in the sample based on pairwise allele sharing estimates (estimated genome proportion shared identical by descent (IBD): pi-hat >0.1875 : n=37). Following these exclusions, Eigenstrat principal components analysis (PCA) was performed, incorporating genotype data from Phase 3 HapMap populations (CEU, CHB, JPT, TSI, YRI). In eigenvector plots, the majority of ASGC samples clustered closely with European (CEU and TSI) reference populations. Eighteen samples (16 cases and 2 controls) showed prominent evidence of Asian ancestry and were removed. Principal component and IBD analyses were performed using a pruned subset of quasi-independent SNPs (~130,000 SNPs) to avoid confounding by linkage disequilibrium (LD). Following quality control, 1162 cases and 1244 controls were available for association analyses at 551,514 SNPs. Genotype imputation in the filtered sample was performed using MACH v1.0.16 (http://www.sph.umich.edu/csg/yli/mach/index.html), based on HapMap Phase 2 (release #24) phased haplotypes for European-ancestry (CEU) samples. Subsequent quality control excluded imputed SNPs with MAF <0.01 or ratio of observed dosage variance to expected binomial variance of r^2 <0.3.

Bio-Repository of DNA in Stroke (BRAINS)

The BRAINS sample was genotyped using the Illumina HumanHap610-Quad array. Quality control excluded SNPs not genotyped on all case and control collections and SNPs with genotype call rate <0.95, deviation from Hardy-Weinberg equilibrium (P<10^-6) or minor allele frequency <0.01. Individual samples were excluded due to low call rates (<95%), gender discrepancy, unexpected relatedness or evidence of non-European ancestry. Genotype imputation was performed using MACH v1.0.16 (http://www.sph.umich.edu/csg/yli/mach/index.html), based on HapMap Phase 2 (release #22) phased haplotypes for European-ancestry (CEU) samples. Quality control removed imputed SNPs with MAF <0.01 or ratio of observed dosage variance of r^2 <0.3.
variance to expected binomial variance of \( r^2 < 0.3 \). Analyses were performed using PLINK v 1.6 (http://pngu.mgh.harvard.edu/~purcell/plink/).

**Cardiovascular Health Study (CHS)**

In 2007-2008, genotyping was performed at the General Clinical Research Center's Phenotyping/Genotyping Laboratory at Cedars-Sinai using the Illumina 370CNV Duo® BeadChip system on the first 2,427 of 3,980 CHS participants who were free of CVD at baseline. The 1,908 persons excluded for prevalent CVD had prevalent coronary heart disease (n=1195), congestive heart failure (n=86), peripheral vascular disease (n=93), valvular heart disease (n=20), stroke (n=166) or transient ischemic attack (n=56). Some persons had more than one reason to be excluded and for these individuals only the initial exclusionary event is listed. The 2,427 participants who had been genotyped by July 2008 (time of the meta-analysis) were a stratified probability sample that included all cases of myocardial infarction (MI), stroke, atrial fibrillation, dementia, and heart failure, and a selection of controls sampled based on the age-, and sex distribution of MI cases. Sampling weights were used in the analysis to account for the study design and to weight back to the underlying cohort selected for genotyping (N=3,980). Because the other cohorts were predominantly white, the African American participants were excluded from the initial meta-analysis to limit the potential for false positive associations due to population stratification. At the time of the meta-analysis, genotyping had been attempted in 2,101 white participants, and was successful in 2,022 persons; the latter constitute the CHS sample for this study. QC criteria used to define successful genotyping are listed elsewhere, 35 persons were excluded for a subject specific call rate <95%.

The set of genotyped input SNPs used for imputation was selected based on the highest quality GWA data. We used a call rate >95%; a minor allele frequency >0.01; a Hardy-Weinberg \( p > 1 \times 10^{-9} \); and a test of differential missingness by the “mishap” test in PLINK \( p > 1 \times 10^{-9} \). We used BIM-BAM to impute to the plus strand of NCBI build 35. For each imputed SNP a reliability of imputation was estimated (as the ratio of the empirically observed dosage variance to the expected binomial dosage variance: O/E ratio).

**deCODE Genetics**

The Icelandic chip-genotyped samples were assayed with the Illumina Human Hap300, Hap CNV370, Hap 610, 1M or Omni-1 Quad bead chips at deCODE genetics. SNPs were excluded if (i) yield was lower than 95%, (ii) minor allele frequency was less than 1% in the population, (iii) significant deviation from Hardy-Weinberg equilibrium was observed in the controls \( P < 0.001 \), (iv) an excessive inheritance error rate (over 0.001) was produced or (v) there was a substantial difference in allele frequency between chip types (from just a single chip if that resolved all differences, but from all chips, otherwise). All samples with a call rate below 97% were excluded from the analysis. Imputation was performed using IMPUTE.

**Framingham Heart Study (FHS)**

FHS participants had DNA extracted and provided consent for genotyping in the 1990s. All available eligible participants were genotyped at Affymetrix (Santa Clara, CA) through an NHLBI funded SNPHealth Association Resource (SHARE) project using the Affymetrix
GeneChip® Human Mapping 500K Array Set and 50K Human Gene Focused Panel. In 272 persons (31 with stroke), small amounts of DNA were extracted from stored whole blood and required whole genome amplification prior to genotyping. Cell lines were available for most of the remaining participants. Genotyping was attempted in 5293 participants, and 4,519 persons met QC criteria. Failures (call rate <97%, extreme heterozygosity or high Mendelian error rate) were largely restricted to persons with whole-genome amplified DNA and DNA extracted from stored serum samples. We also excluded 156 participants who were less than 45 years old at the time of the DNA draw, 135 persons with prevalent stroke and 97 persons who did not have stroke surveillance on follow-up after genotyping; the remaining 4,131 subjects constitute the FHS sample for this study.

The set of genotyped input SNPs used for imputation was selected based on the highest quality GWA data. We used a callrate >97%; a minor allele frequency >0.01; a Hardy-Weinberg p>1x10^{-6}; and a test of differential missingness by the “mishap” test in PLINK p>1x10^{-9} in each study. We used the Markov Chain Haplotyping (MaCH) package (http://www.sph.umich.edu/csg/abecasis/MACH) version 1.0.15 software to impute to plus strand of NCBI build 36, HapMap release #22. For each imputed SNP a reliability of imputation was estimated (as the ratio of the empirically observed dosage variance to the expected binomial dosage variance: O/E ratio). For the primary meta-analysis using inverse variance weighting less weight is given to imputed SNPs with low observed dosage variance (resulting in higher variance of the estimate). For the secondary meta-analysis using the inverse square root (N) weighting the ratio was used to compute an effective sample size.

**The Genetics of Early Onset Stroke (GEOS) Study, Baltimore, USA**

Samples were genotyped at the Johns Hopkins Center for Inherited Disease Research (CIDR), and genotyping was performed using the Illumina HumanOmni1-Quad_v1-0_B BeadChip (Illumina, San Diego, CA, USA). Case and control samples were balanced across the plates. Allele cluster definitions for each SNP were determined using Illumina BeadStudio Genotyping Module version 3.3.7, Gentrain version 1.0 and the combined intensity data from all released samples. Genotypes were not called if the quality threshold (Gencall score) was below 0.15. All samples had a genotype call rate > 98%. Genotyping concordance rate was 99.996% based on study duplicates. Samples were excluded due to unexpected duplicates, gender discrepancy, unexpected relatedness or evidence of non-European ancestry based on principal components analysis. Individual SNPs were excluded from analysis if they had excessive deviation from Hardy-Weinberg Equilibrium (HWE) proportions (P < 1.0x10^{-6}), genotype call rates <97.5% or minor allele frequency < 1%. Departure from HWE was assessed by chi-square test among controls only.

**Heart Protection Study (HPS)**

DNA was extracted from stored white cells and genotyping was carried out at the Centre National de Génotypage in Evry, France. Genotypes were measured using the Illumina 610K Quad panel, called using Illumina BeadStudio software, and imputed with reference to HapMap2 CEU release 22 (build 36) using MACH. Single nucleotide polymorphisms with <97.5% call rate, significant deviation from Hardy-Weinberg equilibrium (p<1x10^{-6}) or low minor allele
frequency (<0.01) were excluded. Genotype data were available for 578 stroke cases and 468 controls after quality control exclusions for discrepant sex, repeated samples and non-European ancestry. Statistical analyses were performed using MACH2QTL.

The Heart and Vascular Health Study (HVH)

Genotyping was performed at the General Clinical Research Center's Phenotyping/Genotyping Laboratory at Cedars-Sinai using the Illumina 370CNV BeadChip system. Genotypes were called using the Illumina BeadStudio software. Samples were excluded from analysis for sex mismatch or call rate < 95%. The following exclusions were applied to identify a final set of 301,321 autosomal SNPs: call rate < 97%, HWE P < 10^{-5}, > 2 duplicate errors or Mendelian inconsistencies (for reference CEPH trios), heterozygote frequency = 0, SNP not found in HapMap, inconsistencies across genotyping batches. Imputation was performed using BIMBAM with reference to HapMap CEU using release 22, build 36 using one round of imputations and the default expectation-maximization warm-ups and runs.

Logistic regression was used to investigate the association of each SNP with the risk of stroke and MI, adjusting for the matching factors of age, sex, hypertension status and index year. We used linear additive models with robust standard errors and estimated risk for each additional copy of the variant allele, using R. SNPs were excluded from analysis for variance on the allele dosage ≤0.01.

The Ischemic Stroke Genetics Study (ISGS)/ Siblings With Ischemic Stroke Study (SWISS)

Both the ISGS/SWISS cases were genotyped using the Illumina 610 or 660 genotyping arrays, while control series used in the ISGS/SWISS dataset were genotyped using the Illumina HumanHap 550Kv1 or 550Kv3 genotyping arrays. Genotypes were called using Illumina GenomeStudio software, with all alleles called on the forward strands. All A/T and G/C SNPs were removed prior to merging case and control sample sets. Also SNPs with discordant minor alleles on the same strand across microarrays were removed prior to merging datasets. Preliminary exclusion criteria per sample included genomewide SNP call rates <95% and discordance between self-reported gender and sex determined from X chromosome heterozygosity. SNPs were excluded from the merged case-control series if genotyping success rate <95%, minor allele frequency (MAF) <0.01, Hardy-Weinberg equilibrium (HWE) P<1E-4 in controls and P<1E-7 in cases, nonrandom missingness per haplotype P<1E-5 and missingness in cases compared to controls (from chi-squared test) P<1E-5.

ISGS/SWISS cases and controls were merged with a subset of samples from HapMap 3 (ASW, CEU, CHB, JPT, TSI and YRI populations) and underwent multidimensional scaling analyses to verify European ancestry with principal component vector 1 (PC1) and 2 (PC2) values greater than 3 standard deviations from the combined CEU/TSI means for each vector were excluded as outliers. Samples were excluded if they shared greater than a 0.125 proportion of alleles (pi_hat >0.125). Basic quality control of genotyped SNP data was carried out using PLINKv1.07. The SNPs passing quality control for the ISGS/SWISS were imputed using a two-stage procedure implemented in Markov Chain based haplotyper (MACH; version 1.0.16) under default settings. For this study, the August 2010 release of the 1000 Genomes European ancestry haplotypes was utilized as a reference for SNP imputation.
The MGH Genes Affecting Stroke Risk and Outcome Study (MGH-GASROS)

Cases and controls were genotyped using the Affy 6.0 array. Quality control procedures excluded SNPs with >5% missingness, minor allele frequency <0.01, or Hardy-Weinberg p-value < 10^{-7}. Individual samples were excluded if they exhibited genotype missingness >5%, cryptic relatedness (one of each pair demonstrating pi_hat [estimated proportion of genome shared identical by descent] > 0.15 was removed), or non-European ancestry based on multidimensional scaling analysis using HapMap Phase 3 populations. Analyses were performed using PLINK v 1.6 (http://pngu.mgh.harvard.edu/~purcell/plink/). Imputation was performed using MACH v 1.0.16 (http://www.sph.umich.edu/csg/yli/mach/index.html) and the HapMap 3 CEU+TSI training set.

Milano Study

Italian cases were genotyped using Illumina Human610-Quad v1_B or Human660W-Quad v1_A chips. Italian controls were genotyped with the Illumina HumanHap610-Quad chip. PCA with HapMap 3 on the Italian cases showed that Italian PROCARDIS controls had similar ancestry to the cases. All samples had a genotype call rate > 95%. Samples were excluded due to unexpected duplicates or evidence of non-European ancestry based on principal components analysis. Quality control procedures excluded SNPs with MAF <0.01 or Hardy-Weinberg P-value <5x10^{-6} in either the case or control collections.

Rotterdam Study

All participants had DNA extracted at baseline in 1990-1993. In 2007-2008, genotyping was attempted in participants with high-quality extracted DNA at baseline. Genotyping was done at the Human Genotyping Facility, Genetic Laboratory Department of Internal Medicine, Erasmus MC, Rotterdam, The Netherlands. Genotyping was carried out using the Illumina HumanHap550 Duo BeadChip® according to the manufacturer’s protocols. Participant-specific quality-control included filters for call rate, heterozygosity, and number of Mendelian errors per individual. SNP-specific quality control included filters for call rate, minor allele frequency, Hardy-Weinberg equilibrium, and differential missingness by outcome or genotypes (mishap test in PLINK). Data were screened for latent substructure, including cryptic relatedness, using IBD matrix. The set of genotyped input SNPs used for imputation was selected based on their highest quality GWA data. We used a call rate >98%, a minor allele frequency >0.01, a Hardy-Weinberg p>1x10^{-6}, and a test of differential missingness by the “mishap” test in PLINK p>1x10^{-9} to selected SNPs used for imputation. We used the Markov Chain Haplotyping (MaCH) package (http://www.sph.umich.edu/csg/abecasis/MACH) to impute to the plus strand of NCBI build 36, HapMap release #22. Imputation of genotypes provides a dosage value for every SNP between 0 and 2 indicating the expected value of a SNP being homozygous for the reference allele. For each imputed SNP an assessment of the informativeness of the imputation was estimated (as the ratio of the empirically observed dosage variance to the expected binomial dosage variance, O/E ratio).
Wellcome Trust Case-Control Consortium 2 (WTCCC2)

All WTCCC2 cases were genotyped as part of the WTCCC2 Ischemic Stroke study using the Illumina Human660W-Quad array. British controls were genotyped using the Illumina Human1.2M-Duo. German controls were genotyped on the Illumina Human 550k platform. Quality control procedures in the WTCCC2 excluded SNPs not genotyped on all case and control collections and SNPs with Fisher information measure <0.98, genotype call rate <0.95, MAF <0.01 or Hardy-Weinberg P-value <1x10-20 in either the case or control collections. Samples were excluded if identified as outliers on call rate, heterozygosity, ancestry and average probe intensity based on a Bayesian clustering algorithm. Samples were also removed if they exhibited discrepancies between inferred and recorded gender or cryptic relatedness with other WTCCC2 samples (pairwise identity-by-descent >0.05). Autosomal genotype imputation was performed using MACH based on HapMap Phase 2 European (CEU) reference data.
Statistical Analysis

The METASTROKE project combined data from 15 GWAS in individuals with European ancestry including 12,389 ischemic stroke cases and 62,004 controls.¹ Genome-wide association analysis were performed for all of the participating studies to test the association of individual SNPs with overall ischemic stroke and its subtypes using logistic regression assuming an additive model. Genomic inflation factor was calculated for each study and was used to correct for over-dispersion in each study. Meta-analysis of the association results from each of the 15 participating stroke studies was conducted using a fixed effects model and an inverse variance method of weighted beta coefficients using METAL software.³¹ Statistical heterogeneity was evaluated using Cochran’s test (Q-test). Association analyses were adjusted for the study-specific covariates age, sex and ancestry principal components.

In the primary analysis, we tested the 18 independent SNPs previously associated with homocysteine levels for their association with ischemic stroke and its subtypes. We pre-specified a Bonferroni-adjusted significance threshold of $\alpha=0.0027$ (where $\alpha=0.05/18$ SNPs) to adjust for multiple testing.

For the secondary analysis of SNPs located within ± 50kb of the 18 tHcy related SNPs we specified a Bonferroni-adjusted significance threshold of 0.0003 ($\alpha=0.05/166=0.0003$, for 166 independent SNPs tested).

Genetic risk score analysis using multi-SNP predictors

We also tested the association between an additive genotype risk score (GRS) of the 18 homocysteine associated SNPs and increased risk of ischemic stroke and its subtypes. The GRS was calculated by using a previously described method³² that approximates the average effect of the tested SNPs on stroke risk. The advantage of this method is that it allows the use of meta-analysis summary statistics and does not require individual-level data from participating studies. This method was implemented using the gtx package for the R statistical programming language, available at http://cran.r-project.org/web/packages/gtx.

The genetic risk score was defined by using a set of $m$ SNPs, for the $i$-th SNP in the $j$-th individual denote $x_{ij}$ as the 0/1/2 coded genotype (for directly genotyped SNPs) or expected allele dosage (which takes real values between 0.0 and 2.0 for imputed SNPs). We used the association results from vanMeurs et al³³ to define the set of regression coefficients to be $w_1, w_2, ..., w_m$. Then the risk score for subject $j$ was defined to be

\[
s_j = s_0 + w_1 x_{1j} + w_2 x_{2j} + ... + w_m x_{mj},
\]

where $s_0$ is the intercept. The coefficients $w_1, w_2, ..., w_m$ represented the effect sizes, in standard deviation units per coded allele, estimated in single SNP analyses of plasma homocysteine levels.³³

The calculations are similar to the inverse-variance weighted meta-analysis. The coefficient of the risk score is a weighted mean of the per-SNP regression coefficients, where each is weighted by its corresponding $w_i$. The estimated variance of the risk score is given by weighting the
estimated variances (squared standard errors) of each per-SNP regression coefficient. The tested SNPs should not be in LD to ensure that the contributions are independent. Using this method, we estimated and tested the coefficient of the risk score in the METASTROKE GWA results for ischemic stroke risk and its subtypes using logistic regression.
Supplemental Tables

**Table I.** Association results of the top independent associated SNPs with plasma homocysteine levels at a genome-wide significance level ($P<5\times 10^{-8}$).\(^{33}\)

<table>
<thead>
<tr>
<th>SNP</th>
<th>Risk locus</th>
<th>Chr</th>
<th>Coded Allele</th>
<th>Freq</th>
<th>Beta</th>
<th>SE</th>
<th>P-value</th>
<th>P_het</th>
</tr>
</thead>
<tbody>
<tr>
<td>rs1801133</td>
<td>MTHFR</td>
<td>1</td>
<td>A</td>
<td>0.34</td>
<td>0.1583</td>
<td>0.007</td>
<td>$4.34\times 10^{-104}$</td>
<td>6.90E-08</td>
</tr>
<tr>
<td>rs2275565</td>
<td>MTR</td>
<td>1</td>
<td>T</td>
<td>0.21</td>
<td>-0.0542</td>
<td>0.009</td>
<td>$1.96\times 10^{-10}$</td>
<td>0.05</td>
</tr>
<tr>
<td>rs7422339</td>
<td>CPS1</td>
<td>2</td>
<td>A</td>
<td>0.33</td>
<td>0.0864</td>
<td>0.008</td>
<td>$4.58\times 10^{-27}$</td>
<td>0.61</td>
</tr>
<tr>
<td>rs9369898</td>
<td>MUT</td>
<td>6</td>
<td>A</td>
<td>0.62</td>
<td>0.0449</td>
<td>0.007</td>
<td>$2.17\times 10^{-10}$</td>
<td>0.66</td>
</tr>
<tr>
<td>rs7130284</td>
<td>NOX4</td>
<td>11</td>
<td>T</td>
<td>0.07</td>
<td>-0.1242</td>
<td>0.013</td>
<td>$1.88\times 10^{-20}$</td>
<td>0.35</td>
</tr>
<tr>
<td>rs154657</td>
<td>DPEPI</td>
<td>16</td>
<td>A</td>
<td>0.47</td>
<td>0.0963</td>
<td>0.007</td>
<td>$1.74\times 10^{-43}$</td>
<td>0.09</td>
</tr>
<tr>
<td>rs234709</td>
<td>CBS</td>
<td>21</td>
<td>T</td>
<td>0.45</td>
<td>-0.0718</td>
<td>0.007</td>
<td>$3.90\times 10^{-24}$</td>
<td>0.44</td>
</tr>
<tr>
<td>rs4660306</td>
<td>MMACHC</td>
<td>1</td>
<td>T</td>
<td>0.33</td>
<td>0.0435</td>
<td>0.007</td>
<td>$2.33\times 10^{-09}$</td>
<td>0.18</td>
</tr>
<tr>
<td>rs548987</td>
<td>SLC17A3</td>
<td>6</td>
<td>C</td>
<td>0.13</td>
<td>0.0597</td>
<td>0.01</td>
<td>$1.12\times 10^{-08}$</td>
<td>0.45</td>
</tr>
<tr>
<td>rs42648</td>
<td>GTPB10</td>
<td>7</td>
<td>A</td>
<td>0.4</td>
<td>-0.0395</td>
<td>0.007</td>
<td>$1.97\times 10^{-08}$</td>
<td>0.57</td>
</tr>
<tr>
<td>rs1801222</td>
<td>CUBN</td>
<td>10</td>
<td>A</td>
<td>0.34</td>
<td>0.0453</td>
<td>0.007</td>
<td>$8.43\times 10^{-10}$</td>
<td>0.05</td>
</tr>
<tr>
<td>rs2251468</td>
<td>HNF1A</td>
<td>12</td>
<td>A</td>
<td>0.65</td>
<td>-0.0512</td>
<td>0.007</td>
<td>$1.28\times 10^{-12}$</td>
<td>0.43</td>
</tr>
<tr>
<td>rs838133</td>
<td>FUT1</td>
<td>19</td>
<td>A</td>
<td>0.45</td>
<td>0.0422</td>
<td>0.007</td>
<td>$7.48\times 10^{-09}$</td>
<td>0.31</td>
</tr>
<tr>
<td>rs12134663</td>
<td>MTHFR</td>
<td>1</td>
<td>A</td>
<td>0.8</td>
<td>-0.101</td>
<td>0.011</td>
<td>$2.54\times 10^{-21}$</td>
<td>0.43</td>
</tr>
<tr>
<td>rs12780845</td>
<td>CUBN</td>
<td>10</td>
<td>A</td>
<td>0.65</td>
<td>0.0529</td>
<td>0.009</td>
<td>$7.8\times 10^{-10}$</td>
<td>0.73</td>
</tr>
<tr>
<td>rs957140</td>
<td>NOX4</td>
<td>11</td>
<td>A</td>
<td>0.45</td>
<td>-0.045</td>
<td>0.008</td>
<td>$2.43\times 10^{-08}$</td>
<td>0.97</td>
</tr>
<tr>
<td>rs12921383</td>
<td>DPEPI/FANCA</td>
<td>16</td>
<td>T</td>
<td>0.87</td>
<td>-0.09</td>
<td>0.014</td>
<td>$8.22\times 10^{-11}$</td>
<td>0.96</td>
</tr>
<tr>
<td>rs2851391</td>
<td>CBS</td>
<td>21</td>
<td>T</td>
<td>0.47</td>
<td>0.056</td>
<td>0.008</td>
<td>$1.70\times 10^{-12}$</td>
<td>0.1</td>
</tr>
</tbody>
</table>

**Abbreviations:** SNP=single-nucleotide polymorphism; Chr=chromosome; Freq=frequency of the coded allele; Beta=beta coefficient; SE=standard error; P_het=P-value for the Cochran’s heterogeneity statistic.
**Table II.** Association results of the 18 tHcy associated SNPs with Ischemic Stroke (IS).

<table>
<thead>
<tr>
<th>SNP</th>
<th>Chr</th>
<th>Risk locus</th>
<th>Coded allele</th>
<th>Freq</th>
<th>Beta</th>
<th>SE</th>
<th>P-value</th>
<th>OR</th>
<th>OR (95L)</th>
<th>OR (95U)</th>
<th>P_het</th>
<th>I²</th>
<th>Q</th>
</tr>
</thead>
<tbody>
<tr>
<td>rs838133</td>
<td>19</td>
<td>FUT1</td>
<td>A</td>
<td>0.46</td>
<td>0.0415</td>
<td>0.0169</td>
<td>0.0139</td>
<td>1.04</td>
<td>1.01</td>
<td>1.08</td>
<td>0.03</td>
<td>37.70</td>
<td>22.46</td>
</tr>
<tr>
<td>rs7422339</td>
<td>2</td>
<td>CPS1</td>
<td>A</td>
<td>0.32</td>
<td>-0.0403</td>
<td>0.0202</td>
<td>0.0455</td>
<td>0.96</td>
<td>0.92</td>
<td>1.00</td>
<td>0.83</td>
<td>0.00</td>
<td>6.61</td>
</tr>
<tr>
<td>rs1801133</td>
<td>1</td>
<td>MTHFR</td>
<td>A</td>
<td>0.34</td>
<td>0.0330</td>
<td>0.0172</td>
<td>0.0546</td>
<td>1.03</td>
<td>1.00</td>
<td>1.07</td>
<td>0.99</td>
<td>0.00</td>
<td>3.53</td>
</tr>
<tr>
<td>rs957140</td>
<td>11</td>
<td>NOX4</td>
<td>A</td>
<td>0.45</td>
<td>-0.0300</td>
<td>0.0165</td>
<td>0.0688</td>
<td>0.97</td>
<td>0.94</td>
<td>1.00</td>
<td>0.79</td>
<td>0.00</td>
<td>7.96</td>
</tr>
<tr>
<td>rs9369898</td>
<td>6</td>
<td>MUT</td>
<td>A</td>
<td>0.63</td>
<td>0.0279</td>
<td>0.0172</td>
<td>0.1047</td>
<td>1.03</td>
<td>0.99</td>
<td>1.06</td>
<td>0.97</td>
<td>0.00</td>
<td>4.77</td>
</tr>
<tr>
<td>rs2851391</td>
<td>21</td>
<td>CBS</td>
<td>T</td>
<td>0.43</td>
<td>-0.0247</td>
<td>0.0159</td>
<td>0.1220</td>
<td>0.98</td>
<td>0.95</td>
<td>1.01</td>
<td>0.30</td>
<td>1.20</td>
<td>16.19</td>
</tr>
<tr>
<td>rs42648</td>
<td>7</td>
<td>GTPB10</td>
<td>A</td>
<td>0.39</td>
<td>0.0240</td>
<td>0.0168</td>
<td>0.1528</td>
<td>1.02</td>
<td>0.99</td>
<td>1.06</td>
<td>0.31</td>
<td>0.60</td>
<td>13.92</td>
</tr>
<tr>
<td>rs1801222</td>
<td>10</td>
<td>CUBN</td>
<td>A</td>
<td>0.37</td>
<td>-0.0236</td>
<td>0.0167</td>
<td>0.1571</td>
<td>0.98</td>
<td>0.95</td>
<td>1.01</td>
<td>0.96</td>
<td>0.00</td>
<td>5.70</td>
</tr>
<tr>
<td>rs548987</td>
<td>6</td>
<td>SLC17A3</td>
<td>C</td>
<td>0.12</td>
<td>-0.0335</td>
<td>0.0255</td>
<td>0.1887</td>
<td>0.97</td>
<td>0.92</td>
<td>1.02</td>
<td>0.10</td>
<td>23.90</td>
<td>18.41</td>
</tr>
<tr>
<td>rs12134663</td>
<td>1</td>
<td>MTHFR</td>
<td>A</td>
<td>0.83</td>
<td>-0.0166</td>
<td>0.0218</td>
<td>0.4469</td>
<td>0.98</td>
<td>0.94</td>
<td>1.03</td>
<td>0.12</td>
<td>21.80</td>
<td>17.91</td>
</tr>
<tr>
<td>rs234709</td>
<td>21</td>
<td>CBS</td>
<td>T</td>
<td>0.44</td>
<td>0.0116</td>
<td>0.0171</td>
<td>0.4988</td>
<td>1.01</td>
<td>0.98</td>
<td>1.05</td>
<td>0.23</td>
<td>7.40</td>
<td>14.04</td>
</tr>
<tr>
<td>rs154657</td>
<td>16</td>
<td>DPEP1</td>
<td>A</td>
<td>0.46</td>
<td>-0.0073</td>
<td>0.0160</td>
<td>0.6476</td>
<td>0.99</td>
<td>0.96</td>
<td>1.02</td>
<td>0.55</td>
<td>0.00</td>
<td>12.72</td>
</tr>
<tr>
<td>rs7130284</td>
<td>11</td>
<td>NOX4</td>
<td>T</td>
<td>0.07</td>
<td>-0.0082</td>
<td>0.0329</td>
<td>0.8034</td>
<td>0.99</td>
<td>0.93</td>
<td>1.06</td>
<td>0.11</td>
<td>23.30</td>
<td>18.27</td>
</tr>
<tr>
<td>rs4660306</td>
<td>6</td>
<td>MACHC</td>
<td>T</td>
<td>0.34</td>
<td>0.0032</td>
<td>0.0168</td>
<td>0.8498</td>
<td>1.00</td>
<td>0.97</td>
<td>1.04</td>
<td>0.70</td>
<td>0.00</td>
<td>9.90</td>
</tr>
<tr>
<td>rs2275565</td>
<td>1</td>
<td>MTR</td>
<td>T</td>
<td>0.21</td>
<td>0.0035</td>
<td>0.0194</td>
<td>0.8546</td>
<td>1.00</td>
<td>0.97</td>
<td>1.04</td>
<td>0.00</td>
<td>52.40</td>
<td>33.62</td>
</tr>
<tr>
<td>rs12780845</td>
<td>10</td>
<td>CUBN</td>
<td>A</td>
<td>0.68</td>
<td>0.0017</td>
<td>0.0182</td>
<td>0.9238</td>
<td>1.00</td>
<td>0.97</td>
<td>1.04</td>
<td>0.38</td>
<td>8.50</td>
<td>12.91</td>
</tr>
<tr>
<td>rs12921383</td>
<td>16</td>
<td>DPEP1/FANCA</td>
<td>T</td>
<td>0.89</td>
<td>-0.0025</td>
<td>0.0301</td>
<td>0.9351</td>
<td>1.00</td>
<td>0.94</td>
<td>1.06</td>
<td>0.38</td>
<td>14.00</td>
<td>9.65</td>
</tr>
<tr>
<td>rs2251468</td>
<td>12</td>
<td>HNF1A</td>
<td>A</td>
<td>0.66</td>
<td>0.0001</td>
<td>0.0166</td>
<td>0.9959</td>
<td>1.00</td>
<td>0.97</td>
<td>1.03</td>
<td>0.96</td>
<td>0.00</td>
<td>6.31</td>
</tr>
</tbody>
</table>

*Abbreviations:* SNP=single-nucleotide polymorphism; Chr=chromosome; Freq=frequency of the coded allele; Beta=beta coefficient; SE=standard error; OR=odds ratio; 95L&95U=lower and upper 95 percentile; P_het=P-value for the Cochran’s heterogeneity statistic; Q=Cochran’s heterogeneity statistic; I²=Higgins Heterogeneity index.
**Table III.** Association results of a Multi-SNP Genotypic Risk Score (consisting of 18 tHcy associated SNPs) with Ischemic Stroke (IS) and its subtypes LVD, SVD, CE.

<table>
<thead>
<tr>
<th></th>
<th>SNPs tested</th>
<th>Beta</th>
<th>SE</th>
<th>OR</th>
<th>95% CI</th>
<th>P-value</th>
<th>P_het</th>
<th>Cases/Controls</th>
</tr>
</thead>
<tbody>
<tr>
<td>IS</td>
<td>18</td>
<td>0.02</td>
<td>0.05</td>
<td>1.02</td>
<td>0.91-1.15</td>
<td>0.63</td>
<td>0.03</td>
<td>12,389/62,004</td>
</tr>
<tr>
<td>LVD</td>
<td>18</td>
<td>0.05</td>
<td>0.12</td>
<td>1.06</td>
<td>0.82-1.35</td>
<td>0.64</td>
<td>0.65</td>
<td>2,167/62,004</td>
</tr>
<tr>
<td>SVD</td>
<td>18</td>
<td>0.10</td>
<td>0.12</td>
<td>1.1</td>
<td>0.85-1.42</td>
<td>0.43</td>
<td>0.01</td>
<td>1,894/62,004</td>
</tr>
<tr>
<td>CE</td>
<td>18</td>
<td>-0.09</td>
<td>0.11</td>
<td>0.9</td>
<td>0.71-1.14</td>
<td>0.40</td>
<td>0.95</td>
<td>2,365/62,004</td>
</tr>
</tbody>
</table>

*Abbreviations: SNP=single-nucleotide polymorphism; Beta=beta coefficient; SE=standard error; OR=odds ratio; CI=confidence interval; P_het=P-value for the Cochran’s heterogeneity statistic.*
Table IV. Association results of the 18 tHcy associated SNPs with small vessel disease (SVD).

<table>
<thead>
<tr>
<th>SNP</th>
<th>Chr</th>
<th>Risk locus</th>
<th>Coded allele</th>
<th>Freq</th>
<th>Beta</th>
<th>SE</th>
<th>P-value</th>
<th>OR</th>
<th>OR (95L)</th>
<th>OR (95U)</th>
<th>P_het</th>
<th>(I^2)</th>
<th>Q</th>
</tr>
</thead>
<tbody>
<tr>
<td>rs9369898</td>
<td>6</td>
<td>MUT</td>
<td>A</td>
<td>0.63</td>
<td>0.1174</td>
<td>0.0384</td>
<td>0.0022</td>
<td>1.12</td>
<td>1.04</td>
<td>1.21</td>
<td>0.37</td>
<td>7.40</td>
<td>9.71</td>
</tr>
<tr>
<td>rs838133</td>
<td>19</td>
<td>FUT1</td>
<td>A</td>
<td>0.46</td>
<td>0.0732</td>
<td>0.0368</td>
<td>0.0465</td>
<td>1.08</td>
<td>1.00</td>
<td>1.16</td>
<td>0.02</td>
<td>54.20</td>
<td>21.85</td>
</tr>
<tr>
<td>rs2851391</td>
<td>21</td>
<td>CBS</td>
<td>T</td>
<td>0.44</td>
<td>-0.0664</td>
<td>0.0353</td>
<td>0.0601</td>
<td>0.94</td>
<td>0.87</td>
<td>1.00</td>
<td>0.96</td>
<td>0.00</td>
<td>4.39</td>
</tr>
<tr>
<td>rs957140</td>
<td>11</td>
<td>NOX4</td>
<td>A</td>
<td>0.44</td>
<td>-0.0646</td>
<td>0.0361</td>
<td>0.0737</td>
<td>0.94</td>
<td>0.87</td>
<td>1.01</td>
<td>0.64</td>
<td>0.00</td>
<td>7.00</td>
</tr>
<tr>
<td>rs1801133</td>
<td>1</td>
<td>MTHFR</td>
<td>A</td>
<td>0.34</td>
<td>0.0650</td>
<td>0.0367</td>
<td>0.0766</td>
<td>1.07</td>
<td>0.99</td>
<td>1.15</td>
<td>0.67</td>
<td>0.00</td>
<td>7.58</td>
</tr>
<tr>
<td>rs12921383</td>
<td>16</td>
<td>DPEP1/FANCA</td>
<td>T</td>
<td>0.89</td>
<td>0.1201</td>
<td>0.0691</td>
<td>0.0820</td>
<td>1.13</td>
<td>0.98</td>
<td>1.29</td>
<td>0.59</td>
<td>0.00</td>
<td>5.59</td>
</tr>
<tr>
<td>rs12780845</td>
<td>10</td>
<td>CUBN</td>
<td>A</td>
<td>0.69</td>
<td>-0.0696</td>
<td>0.0403</td>
<td>0.0843</td>
<td>0.93</td>
<td>0.86</td>
<td>1.01</td>
<td>0.17</td>
<td>29.70</td>
<td>12.80</td>
</tr>
<tr>
<td>rs7130284</td>
<td>11</td>
<td>NOX4</td>
<td>T</td>
<td>0.07</td>
<td>0.0952</td>
<td>0.0720</td>
<td>0.1864</td>
<td>1.10</td>
<td>0.96</td>
<td>1.27</td>
<td>0.61</td>
<td>0.00</td>
<td>7.29</td>
</tr>
<tr>
<td>rs42648</td>
<td>7</td>
<td>GTPB10</td>
<td>A</td>
<td>0.39</td>
<td>0.0364</td>
<td>0.0371</td>
<td>0.3265</td>
<td>1.04</td>
<td>0.96</td>
<td>1.12</td>
<td>0.03</td>
<td>52.00</td>
<td>18.75</td>
</tr>
<tr>
<td>rs7422339</td>
<td>2</td>
<td>CPS1</td>
<td>A</td>
<td>0.32</td>
<td>0.0321</td>
<td>0.0457</td>
<td>0.4820</td>
<td>1.03</td>
<td>0.94</td>
<td>1.13</td>
<td>0.85</td>
<td>0.00</td>
<td>4.05</td>
</tr>
<tr>
<td>rs12134663</td>
<td>1</td>
<td>MTHFR</td>
<td>A</td>
<td>0.83</td>
<td>-0.0285</td>
<td>0.0479</td>
<td>0.5518</td>
<td>0.97</td>
<td>0.88</td>
<td>1.07</td>
<td>0.79</td>
<td>0.00</td>
<td>5.44</td>
</tr>
<tr>
<td>rs234709</td>
<td>21</td>
<td>CBS</td>
<td>T</td>
<td>0.45</td>
<td>0.0175</td>
<td>0.0372</td>
<td>0.6380</td>
<td>1.02</td>
<td>0.95</td>
<td>1.09</td>
<td>0.35</td>
<td>10.20</td>
<td>10.02</td>
</tr>
<tr>
<td>rs1801222</td>
<td>10</td>
<td>CUBN</td>
<td>A</td>
<td>0.37</td>
<td>-0.0144</td>
<td>0.0369</td>
<td>0.6959</td>
<td>0.99</td>
<td>0.92</td>
<td>1.06</td>
<td>0.79</td>
<td>0.00</td>
<td>6.32</td>
</tr>
<tr>
<td>rs154657</td>
<td>16</td>
<td>DPEP1</td>
<td>A</td>
<td>0.46</td>
<td>-0.0101</td>
<td>0.0356</td>
<td>0.7759</td>
<td>0.99</td>
<td>0.92</td>
<td>1.06</td>
<td>0.13</td>
<td>32.80</td>
<td>16.36</td>
</tr>
<tr>
<td>rs548987</td>
<td>6</td>
<td>SLC17A3</td>
<td>C</td>
<td>0.13</td>
<td>0.0153</td>
<td>0.0556</td>
<td>0.7838</td>
<td>1.02</td>
<td>0.91</td>
<td>1.13</td>
<td>0.29</td>
<td>16.70</td>
<td>10.80</td>
</tr>
<tr>
<td>rs4660306</td>
<td>1</td>
<td>MMACHC</td>
<td>T</td>
<td>0.34</td>
<td>0.0094</td>
<td>0.0374</td>
<td>0.8022</td>
<td>1.01</td>
<td>0.94</td>
<td>1.09</td>
<td>0.88</td>
<td>0.00</td>
<td>5.11</td>
</tr>
<tr>
<td>rs2275565</td>
<td>1</td>
<td>MTR</td>
<td>T</td>
<td>0.21</td>
<td>-0.0095</td>
<td>0.0433</td>
<td>0.8264</td>
<td>0.99</td>
<td>0.91</td>
<td>1.08</td>
<td>0.03</td>
<td>48.30</td>
<td>21.27</td>
</tr>
<tr>
<td>rs2251468</td>
<td>12</td>
<td>HNF1A</td>
<td>A</td>
<td>0.66</td>
<td>0.0024</td>
<td>0.0366</td>
<td>0.9486</td>
<td>1.00</td>
<td>0.93</td>
<td>1.08</td>
<td>0.83</td>
<td>0.00</td>
<td>6.55</td>
</tr>
</tbody>
</table>

**Abbreviations:** SNP=single-nucleotide polymorphism; Chr=chromosome; Freq=frequency of the coded allele; Beta=beta coefficient; SE=standard error; OR=odds ratio; 95L&95U=lower and upper 95 percentile; P_het=P-value for the Cochran’s heterogeneity statistic; Q=Cochran’s heterogeneity statistic; \(I^2\)=Higgins Heterogeneity index.
<table>
<thead>
<tr>
<th>SNP</th>
<th>Chr</th>
<th>Risk locus</th>
<th>Coded allele</th>
<th>Freq</th>
<th>Beta</th>
<th>SE</th>
<th>P-value</th>
<th>OR</th>
<th>OR (95L)</th>
<th>OR (95U)</th>
<th>P_het</th>
<th>I²</th>
<th>Q</th>
</tr>
</thead>
<tbody>
<tr>
<td>rs838133</td>
<td>19</td>
<td>FUT1</td>
<td>A</td>
<td>0.45</td>
<td>0.0835</td>
<td>0.0354</td>
<td>0.0185</td>
<td>1.09</td>
<td>1.01</td>
<td>1.17</td>
<td>0.21</td>
<td>9.40</td>
<td>12.14</td>
</tr>
<tr>
<td>rs7422339</td>
<td>2</td>
<td>CPS1</td>
<td>A</td>
<td>0.33</td>
<td>-0.0579</td>
<td>0.0418</td>
<td>0.1660</td>
<td>0.94</td>
<td>0.87</td>
<td>1.02</td>
<td>0.34</td>
<td>10.60</td>
<td>9.04</td>
</tr>
<tr>
<td>rs12134663</td>
<td>1</td>
<td>MTHFR</td>
<td>A</td>
<td>0.83</td>
<td>0.0494</td>
<td>0.0478</td>
<td>0.3016</td>
<td>1.05</td>
<td>0.96</td>
<td>1.15</td>
<td>0.90</td>
<td>0.00</td>
<td>3.43</td>
</tr>
<tr>
<td>rs12780845</td>
<td>10</td>
<td>CUBN</td>
<td>A</td>
<td>0.68</td>
<td>-0.0382</td>
<td>0.0390</td>
<td>0.3278</td>
<td>0.96</td>
<td>0.89</td>
<td>1.04</td>
<td>0.29</td>
<td>3.90</td>
<td>9.63</td>
</tr>
<tr>
<td>rs154657</td>
<td>16</td>
<td>DPEP1</td>
<td>A</td>
<td>0.46</td>
<td>0.0320</td>
<td>0.0343</td>
<td>0.3508</td>
<td>1.03</td>
<td>0.97</td>
<td>1.10</td>
<td>0.35</td>
<td>8.70</td>
<td>11.04</td>
</tr>
<tr>
<td>rs1801222</td>
<td>10</td>
<td>CUBN</td>
<td>A</td>
<td>0.36</td>
<td>-0.0307</td>
<td>0.0362</td>
<td>0.3958</td>
<td>0.97</td>
<td>0.90</td>
<td>1.04</td>
<td>0.73</td>
<td>0.00</td>
<td>6.13</td>
</tr>
<tr>
<td>rs548987</td>
<td>6</td>
<td>SLC17A3</td>
<td>C</td>
<td>0.13</td>
<td>0.0444</td>
<td>0.0537</td>
<td>0.4079</td>
<td>1.05</td>
<td>0.94</td>
<td>1.16</td>
<td>0.14</td>
<td>18.30</td>
<td>12.24</td>
</tr>
<tr>
<td>rs7130284</td>
<td>11</td>
<td>NOX4</td>
<td>T</td>
<td>0.07</td>
<td>0.0507</td>
<td>0.0683</td>
<td>0.4573</td>
<td>1.05</td>
<td>0.92</td>
<td>1.20</td>
<td>0.41</td>
<td>22.00</td>
<td>8.20</td>
</tr>
<tr>
<td>rs1801133</td>
<td>1</td>
<td>MTHFR</td>
<td>A</td>
<td>0.34</td>
<td>0.0265</td>
<td>0.0364</td>
<td>0.4662</td>
<td>1.03</td>
<td>0.96</td>
<td>1.10</td>
<td>0.89</td>
<td>0.00</td>
<td>4.29</td>
</tr>
<tr>
<td>rs234709</td>
<td>21</td>
<td>CBS</td>
<td>T</td>
<td>0.44</td>
<td>-0.0248</td>
<td>0.0360</td>
<td>0.4901</td>
<td>0.98</td>
<td>0.91</td>
<td>1.05</td>
<td>0.84</td>
<td>0.00</td>
<td>4.23</td>
</tr>
<tr>
<td>rs2251468</td>
<td>12</td>
<td>HNF1A</td>
<td>A</td>
<td>0.66</td>
<td>-0.0245</td>
<td>0.0356</td>
<td>0.4913</td>
<td>0.98</td>
<td>0.91</td>
<td>1.05</td>
<td>0.31</td>
<td>2.80</td>
<td>11.68</td>
</tr>
<tr>
<td>rs42648</td>
<td>7</td>
<td>GTPB10</td>
<td>A</td>
<td>0.39</td>
<td>-0.0170</td>
<td>0.0362</td>
<td>0.6395</td>
<td>0.98</td>
<td>0.92</td>
<td>1.06</td>
<td>0.71</td>
<td>0.00</td>
<td>5.42</td>
</tr>
<tr>
<td>rs2851391</td>
<td>21</td>
<td>CBS</td>
<td>T</td>
<td>0.44</td>
<td>0.0150</td>
<td>0.0343</td>
<td>0.6632</td>
<td>1.02</td>
<td>0.95</td>
<td>1.09</td>
<td>0.84</td>
<td>0.00</td>
<td>5.66</td>
</tr>
<tr>
<td>rs12921383</td>
<td>16</td>
<td>DPEP1/FANCA</td>
<td>T</td>
<td>0.90</td>
<td>0.0155</td>
<td>0.0631</td>
<td>0.8064</td>
<td>1.02</td>
<td>0.90</td>
<td>1.15</td>
<td>0.75</td>
<td>0.00</td>
<td>4.28</td>
</tr>
<tr>
<td>rs4660306</td>
<td>1</td>
<td>MMACHC</td>
<td>T</td>
<td>0.34</td>
<td>-0.0062</td>
<td>0.0360</td>
<td>0.8621</td>
<td>0.99</td>
<td>0.93</td>
<td>1.07</td>
<td>0.85</td>
<td>0.00</td>
<td>4.84</td>
</tr>
<tr>
<td>rs2275565</td>
<td>1</td>
<td>MTR</td>
<td>T</td>
<td>0.21</td>
<td>0.0053</td>
<td>0.0408</td>
<td>0.8958</td>
<td>1.01</td>
<td>0.93</td>
<td>1.09</td>
<td>0.01</td>
<td>50.90</td>
<td>24.42</td>
</tr>
<tr>
<td>rs9369898</td>
<td>6</td>
<td>MUT</td>
<td>A</td>
<td>0.63</td>
<td>0.0016</td>
<td>0.0365</td>
<td>0.9639</td>
<td>1.00</td>
<td>0.93</td>
<td>1.08</td>
<td>0.37</td>
<td>15.70</td>
<td>8.64</td>
</tr>
<tr>
<td>rs957140</td>
<td>11</td>
<td>NOX4</td>
<td>A</td>
<td>0.44</td>
<td>-0.0012</td>
<td>0.0351</td>
<td>0.9722</td>
<td>1.00</td>
<td>0.93</td>
<td>1.07</td>
<td>0.25</td>
<td>2.80</td>
<td>10.29</td>
</tr>
</tbody>
</table>

**Abbreviations:** SNP=single-nucleotide polymorphism; Chr=chromosome; Freq=frequency of the coded allele; Beta=beta coefficient; SE=standard error; OR=odds ratio; 95L&95U=lower and upper 95 percentile; P_het=P-value for the Cochran’s heterogeneity statistic; Q=Cochran’s heterogeneity statistic; I²=Higgins Heterogeneity index.
<table>
<thead>
<tr>
<th>SNP</th>
<th>Chr</th>
<th>Risk locus</th>
<th>Coded allele</th>
<th>Freq</th>
<th>Beta</th>
<th>SE</th>
<th>P-value</th>
<th>OR</th>
<th>OR (95L)</th>
<th>OR (95U)</th>
<th>P_het</th>
<th>I²</th>
<th>Q</th>
</tr>
</thead>
<tbody>
<tr>
<td>rs12134663</td>
<td>1</td>
<td>MTHFR</td>
<td>A</td>
<td>0.83</td>
<td>0.0481</td>
<td>0.0444</td>
<td>0.28</td>
<td>0.90</td>
<td>0.82</td>
<td>0.99</td>
<td>0.03</td>
<td>49.90</td>
<td>19.96</td>
</tr>
<tr>
<td>rs7422339</td>
<td>2</td>
<td>CPS1</td>
<td>A</td>
<td>0.33</td>
<td>-0.0424</td>
<td>0.0398</td>
<td>0.29</td>
<td>1.09</td>
<td>1.18</td>
<td>1.00</td>
<td>0.94</td>
<td>0.00</td>
<td>3.56</td>
</tr>
<tr>
<td>rs838133</td>
<td>19</td>
<td>FUT1</td>
<td>A</td>
<td>0.46</td>
<td>0.0334</td>
<td>0.0334</td>
<td>0.32</td>
<td>1.04</td>
<td>0.98</td>
<td>1.11</td>
<td>0.31</td>
<td>13.60</td>
<td>11.57</td>
</tr>
<tr>
<td>rs548987</td>
<td>6</td>
<td>SLC17A3</td>
<td>C</td>
<td>0.12</td>
<td>-0.0488</td>
<td>0.0519</td>
<td>0.35</td>
<td>1.06</td>
<td>1.17</td>
<td>0.96</td>
<td>0.66</td>
<td>0.00</td>
<td>7.64</td>
</tr>
<tr>
<td>rs957140</td>
<td>11</td>
<td>NOX4</td>
<td>A</td>
<td>0.45</td>
<td>-0.0294</td>
<td>0.0329</td>
<td>0.37</td>
<td>0.96</td>
<td>1.01</td>
<td>0.90</td>
<td>0.44</td>
<td>0.10</td>
<td>10.01</td>
</tr>
<tr>
<td>rs2851391</td>
<td>21</td>
<td>CBS</td>
<td>T</td>
<td>0.43</td>
<td>-0.0275</td>
<td>0.0315</td>
<td>0.38</td>
<td>1.06</td>
<td>1.12</td>
<td>1.00</td>
<td>0.53</td>
<td>0.00</td>
<td>10.99</td>
</tr>
<tr>
<td>rs234709</td>
<td>21</td>
<td>CBS</td>
<td>T</td>
<td>0.44</td>
<td>0.0257</td>
<td>0.0344</td>
<td>0.46</td>
<td>0.93</td>
<td>0.88</td>
<td>0.98</td>
<td>0.13</td>
<td>35.30</td>
<td>13.92</td>
</tr>
<tr>
<td>rs154657</td>
<td>16</td>
<td>DPEP1</td>
<td>A</td>
<td>0.46</td>
<td>-0.0232</td>
<td>0.0316</td>
<td>0.46</td>
<td>1.10</td>
<td>1.15</td>
<td>1.05</td>
<td>0.52</td>
<td>0.00</td>
<td>11.13</td>
</tr>
<tr>
<td>rs1801222</td>
<td>10</td>
<td>CUBN</td>
<td>A</td>
<td>0.36</td>
<td>-0.0208</td>
<td>0.0334</td>
<td>0.53</td>
<td>1.05</td>
<td>1.09</td>
<td>1.00</td>
<td>0.83</td>
<td>0.00</td>
<td>6.55</td>
</tr>
<tr>
<td>rs7130284</td>
<td>11</td>
<td>NOX4</td>
<td>T</td>
<td>0.07</td>
<td>-0.0407</td>
<td>0.0656</td>
<td>0.53</td>
<td>0.88</td>
<td>0.96</td>
<td>0.82</td>
<td>0.06</td>
<td>43.10</td>
<td>17.58</td>
</tr>
<tr>
<td>rs4660306</td>
<td>1</td>
<td>MMACHC</td>
<td>T</td>
<td>0.34</td>
<td>-0.0196</td>
<td>0.0333</td>
<td>0.56</td>
<td>1.04</td>
<td>1.09</td>
<td>1.01</td>
<td>0.33</td>
<td>12.10</td>
<td>12.52</td>
</tr>
<tr>
<td>rs12780845</td>
<td>10</td>
<td>CUBN</td>
<td>A</td>
<td>0.68</td>
<td>0.0202</td>
<td>0.0366</td>
<td>0.58</td>
<td>1.05</td>
<td>1.01</td>
<td>1.10</td>
<td>0.69</td>
<td>0.00</td>
<td>7.40</td>
</tr>
<tr>
<td>rs2275565</td>
<td>1</td>
<td>MTR</td>
<td>T</td>
<td>0.21</td>
<td>0.0195</td>
<td>0.0380</td>
<td>0.61</td>
<td>0.95</td>
<td>0.91</td>
<td>0.98</td>
<td>0.02</td>
<td>51.20</td>
<td>24.57</td>
</tr>
<tr>
<td>rs1801133</td>
<td>16</td>
<td>DPEP1/FANCA</td>
<td>T</td>
<td>0.34</td>
<td>0.0165</td>
<td>0.0343</td>
<td>0.63</td>
<td>1.17</td>
<td>1.13</td>
<td>1.21</td>
<td>0.11</td>
<td>36.80</td>
<td>15.81</td>
</tr>
<tr>
<td>rs9369898</td>
<td>6</td>
<td>MUT</td>
<td>A</td>
<td>0.63</td>
<td>0.0151</td>
<td>0.0344</td>
<td>0.66</td>
<td>1.05</td>
<td>1.02</td>
<td>1.08</td>
<td>0.46</td>
<td>0.00</td>
<td>9.81</td>
</tr>
<tr>
<td>rs42648</td>
<td>1</td>
<td>GTPB10</td>
<td>A</td>
<td>0.39</td>
<td>0.0114</td>
<td>0.0338</td>
<td>0.73</td>
<td>0.96</td>
<td>0.94</td>
<td>0.98</td>
<td>0.03</td>
<td>48.80</td>
<td>19.52</td>
</tr>
<tr>
<td>rs12921383</td>
<td>16</td>
<td>DPEP1/FANCA</td>
<td>T</td>
<td>0.90</td>
<td>0.0181</td>
<td>0.0621</td>
<td>0.77</td>
<td>0.91</td>
<td>0.88</td>
<td>0.95</td>
<td>0.72</td>
<td>0.00</td>
<td>4.50</td>
</tr>
<tr>
<td>rs2251468</td>
<td>12</td>
<td>HNF1A</td>
<td>A</td>
<td>0.66</td>
<td>-0.0002</td>
<td>0.0328</td>
<td>1.00</td>
<td>0.95</td>
<td>0.95</td>
<td>0.95</td>
<td>0.33</td>
<td>11.70</td>
<td>13.59</td>
</tr>
</tbody>
</table>

**Abbreviations:** SNP=single-nucleotide polymorphism; Chr=chromosome; Freq=frequency of the coded allele; Beta=beta coefficient; SE=standard error; OR=odds ratio; 95L&95U=lower and upper 95 percentile; P_het=P-value for the Cochran’s heterogeneity statistic; Q=Cochran’s heterogeneity statistic; I²=Higgins Heterogeneity index.
Supplemental References


