Genetic Factors Influencing Coagulation Factor XIII B-Subunit Contribute to Risk of Ischemic Stroke

Ken B. Hanscombe, PhD; Matthew Traylor, PhD; Pirro G. Hysi, PhD; Stephen Bevan, PhD; Martin Dichgans, MD; Peter M. Rothwell, FMedSci; Bradford B. Worrall, MD; Sudha Seshadri, MD; Cathie Sudlow, FRCPE; the METASTROKE Consortium; the Wellcome Trust Case Control Consortium 2; Frances M.K. Williams, PhD*; Hugh S. Markus, DM*; Cathryn M. Lewis, PhD*

Background and Purpose—Abnormal coagulation has been implicated in the pathogenesis of ischemic stroke, but how this association is mediated and whether it differs between ischemic stroke subtypes is unknown. We determined the shared genetic risk between 14 coagulation factors and ischemic stroke and its subtypes.

Methods—Using genome-wide association study results for 14 coagulation factors from the population-based TwinsUK sample (N=2000 for each factor), meta-analysis results from the METASTROKE consortium ischemic stroke genome-wide association study (12 389 cases, 62 004 controls), and genotype data for 9520 individuals from the WTCCC2 ischemic stroke study (3548 cases, 5972 controls—the largest METASTROKE subsample), we explored shared genetic risk for coagulation and stroke. We performed three analyses: (1) a test for excess concordance (or discordance) in single nucleotide polymorphism effect direction across coagulation and stroke, (2) an estimation of the joint effect of multiple coagulation-associated single nucleotide polymorphisms in stroke, and (3) an evaluation of common genetic risk between coagulation and stroke.

Results—One coagulation factor, factor XIII subunit B (FXIIIB), showed consistent effects in the concordance analysis, the estimation of polygenic risk, and the validation with genotype data, with associations specific to the cardioembolic stroke subtype. Effect directions for FXIIIB-associated single nucleotide polymorphisms were significantly discordant with cardioembolic disease (smallest \( P = 5.7 \times 10^{-04} \)); the joint effect of FXIIIB-associated single nucleotide polymorphisms was significantly predictive of ischemic stroke (smallest \( P = 1.8 \times 10^{-06} \)) and the cardioembolic subtype (smallest \( P = 1.7 \times 10^{-04} \)). We found substantial negative genetic covariation between FXIIIB and ischemic stroke (\( r_G = -0.71, P = 0.01 \)) and the cardioembolic subtype (\( r_G = -0.80, P = 0.03 \)).

Conclusions—Genetic markers associated with low FXIIIB levels increase risk of ischemic stroke cardioembolic subtype. (Stroke. 2015;46:00-00. DOI: 10.1161/STROKEAHA.115.009387.)

Key Words: coagulation ■ genome-wide association study ■ stroke
risk in the general population. Similarly, twin and family-based epidemiological studies, as well as more recent estimates of heritability from genome-wide association study (GWAS) arrays, suggest a substantial heritability for ischemic stroke.3–5

In a previous GWAS, we found that common genetic variants in the ABO gene were associated with ischemic stroke subtypes large-vessel disease (LVD) and cardioembolic (CE) stroke, but not with small-vessel disease (SVD).6 However, even though the GWAS approach surveys the entire genome for genetic associations, it typically queries one single nucleotide polymorphism (SNP) at a time for association with the disease or trait of interest. Evidence from GWAS suggests that common complex traits, such as coagulation and ischemic stroke, are polygenic, influenced by many SNPs each contributing a relatively small effect. Recently developed approaches test a polygenic model of common genetic liability by considering aggregate SNP effects.7,8 We used GWAS results for multiple different coagulation factors and hypothesized that by aggregating the evidence for association in each coagulation factor, we would be able to identify those coagulation factors that increase stroke risk and determine whether these differentially associate with stroke subtype.

Methods

Cohorts Studied

TwinsUK
The UK Adult Twin Registry (TwinsUK, http://www.twinsuk.ac.uk)9,10 was the sampling frame for a European Union–funded multicentre study that aimed to identify common genetic risk for blood coagulation and ischemic stroke (EuroCLOT, http://cordis.europa.eu/result/rcn/52015_en.html). Coagulation and fibrinolytic factors were selected on the basis of evidence for association with risk for atherothrombotic disorders, in a previous study estimating heritability of hemostatic factors.1 These hemostatic proteins are components of the coagulation pathway that is part of the defense mechanism against severe blood loss.11 The plasma protein assay protocols for the coagulation factors are described in detail elsewhere.12 Genotyping was performed in 3 batches on the Illumina Human Hap300 and Human Hap610-Quad arrays. Full details of genotyping and quality control were described elsewhere. The sample sizes for the 14 coagulation factors for which GWAS results were available are shown in Table. We used the coagulation GWAS results as a discovery sample.

METASTROKE
METASTROKE is a 15-cohort meta-analysis of ischemic stroke GWASs.12 The GWASs included ischemic stroke patients of European ancestry (Europe, North America, and Australia) with ancestry-matched controls. Eleven studies used case–control methodology and were cross-sectional; the remaining 4 were prospective, population-based cohorts. Full details of the populations and analysis are described elsewhere.12 We used METASTROKE meta-analysis results as the target sample.

WTCCC2
The Wellcome Trust Case Control Consortium 2 (WTCCC2) ischemic stroke sample—the largest METASTROKE subsample—included a German (1174 cases and 797 controls) and a UK sample (2374 cases and 5175 controls).13 After combining the UK and German samples, we removed 16 population outliers. We used the WTCCC2 stroke sample genotypes as our validation sample.

In all METASTROKE studies, including the WTCCC2 subsample, stroke was defined as a typical clinical syndrome with radiological confirmation. Stroke subtyping was performed with the Trial of Org 10172 in Acute Stroke Treatment (TOAST) classification system.14 Brain imaging with computed tomography or magnetic resonance imaging was undertaken for >95% of cases in all the METASTROKE cohorts (sample sizes are included in Table).

Study Design

SNP Selection
We used a thinned set of coagulation-associated risk SNPs for both the concordance and polygenic risk analyses because SNP correlation because of linkage disequilibrium may inflate test statistics from polygenic approaches.

ClumpingFrom the TwinsUK SNPs imputed to HapMap2, we first selected only SNPs common to both genotyping platforms used. For all SNPs with a minor allele frequency ≥0.05, we then used the clumping routine implemented in PLINK,15 which selects SNPs based on association P value and linkage disequilibrium, using standard parameter values (linkage disequilibrium threshold of r² ≤0.25 in the HapMap CEU reference panel; distance window of 300 kb). Clumping resulted in a subset of ≈70000 near-independent SNPs for each coagulation factor. The ratio of estimated number of independent SNPs to observed (clumped) SNPs, the so-called effective ratio,17 was ≈0.93.

Inclusion ThresholdsWe repeated each coagulation factor-stroke subtype analysis at multiple SNP-association P value inclusion thresholds. We ranked the clumped SNPs by their

<table>
<thead>
<tr>
<th>Variable</th>
<th>N</th>
<th>Study Sample Sizes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Phenotype</td>
<td></td>
<td></td>
</tr>
<tr>
<td>TwinUK (genotype, phenotype, and GWAS results data)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Factor VII antigen, IU/mL</td>
<td>1374</td>
<td></td>
</tr>
<tr>
<td>Factor VII percent, %</td>
<td>1787</td>
<td></td>
</tr>
<tr>
<td>Factor VIII antigen, IU/mL</td>
<td>2504</td>
<td></td>
</tr>
<tr>
<td>Factor X antigen, IU/mL</td>
<td>2600</td>
<td></td>
</tr>
<tr>
<td>Factor XII antigen, IU/mL</td>
<td>2595</td>
<td></td>
</tr>
<tr>
<td>Factor XIII activity, %</td>
<td>2598</td>
<td></td>
</tr>
<tr>
<td>Factor XIII subunit A antigen, IU/mL</td>
<td>2599</td>
<td></td>
</tr>
<tr>
<td>Factor XIII subunit B antigen, IU/mL</td>
<td>2601</td>
<td></td>
</tr>
<tr>
<td>Von Willebrand factor antigen, IU/mL</td>
<td>2584</td>
<td></td>
</tr>
<tr>
<td>Prothrombin antigen, IU/mL</td>
<td>2596</td>
<td></td>
</tr>
<tr>
<td>D-dimer, ng/mL</td>
<td>2762</td>
<td></td>
</tr>
<tr>
<td>Plasminogen activator inhibitor antigen, ng/mL</td>
<td>2590</td>
<td></td>
</tr>
<tr>
<td>Tissue-type plasminogen activator antigen, ng/mL</td>
<td>2564</td>
<td></td>
</tr>
<tr>
<td>Fibrinogen, g/L</td>
<td>2607</td>
<td></td>
</tr>
<tr>
<td>METASTROKE (GWAS meta-analysis results data)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cardioembolic subtype</td>
<td>2365 cases</td>
<td></td>
</tr>
<tr>
<td>Large vessel disease</td>
<td>2167 cases</td>
<td></td>
</tr>
<tr>
<td>Small vessel disease</td>
<td>1894 cases</td>
<td></td>
</tr>
<tr>
<td>All ischemic stroke</td>
<td>12389 cases; 62004 controls</td>
<td></td>
</tr>
<tr>
<td>WTCCC2 (genotype and phenotype data)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cardioembolic subtype</td>
<td>790 cases</td>
<td></td>
</tr>
<tr>
<td>Large vessel disease</td>
<td>844 cases</td>
<td></td>
</tr>
<tr>
<td>Small vessel disease</td>
<td>580 cases</td>
<td></td>
</tr>
<tr>
<td>All ischemic stroke</td>
<td>3548 cases; 5972 controls</td>
<td></td>
</tr>
</tbody>
</table>

GWAS indicates genome-wide association study.
evidence for association in the TwinsUK discovery GWASs and created 7 overlapping SNP subsets for all SNPs significant at $P \leq \{0.01, 0.05, 0.1, 0.2, 0.3, 0.4, 0.5\}$, for each factor. These $P$ values define increasingly liberal inclusion thresholds. Polygenic approaches show that more liberal inclusion thresholds benefit from true associations that exist all the way down a $P$ value ranked list of genetic markers, far below the level usually considered as genome-wide significant.7

Multiple Testing
We estimated the number of independent coagulation variables using a Nyholt correction18 of the coagulation factor polygenic risk scores (PRSs). The number of independent tests from the 14 coagulation PRSs was estimated at 13.4, which highlights the low level of polygenic correlation among the coagulation factors themselves. We used a conservative Bonferroni-corrected probability value of $P \leq 0.05/(13.4\times4)$, where 4 is the number of stroke subtypes, to control for multiple testing in the discovery and target samples using GWAS summary statistics; we used $P \leq 0.05$ for focused follow-up on genotype data in the validation subsample.

Statistical Analysis
We used 3 approaches to test for shared genetic influence on coagulation traits and ischemic stroke and its subtypes, as follows:

Concordance of Genetic Effects
We first tested for greater than expected concordance (or discordance) in direction of effects ($\beta$-estimate, odds ratio) between each coagulation factor–stroke subtype pair. At each $P$ value inclusion threshold, we used an Exact Binomial Test to assess whether there was a greater than chance agreement in effect direction. This approach simply counts the number of SNPs acting in the same direction, regardless of magnitude of effect and tests the expectation under the null hypothesis of 50% agreement.

Polygenic Risk Scores
Next we assessed shared genetic risk (pleiotropy) between each coagulation factor and each stroke subtype by summing over all SNP effects in the target sample and weighting by evidence for association in the coagulation discovery sample. For a given set of SNPs, an individual’s PRS is the sum of risk alleles they have weighted by allele effect size in the discovery trait. It represents their genetic risk for the discovery trait. The method described below does not explicitly calculate the coagulation PRS for each individual but rather estimates the association between such a score and the target trait.

Summary Statistics
With summary data, the joint effect of the SNPs in each subset was estimated by a weighted mean of individual SNP effects, that is, summing the SNP effects in the stroke target sample weighted by their effects in the coagulation discovery sample. In a regression model predicting the target sample outcome, $y = \epsilon + dPRS + e$, the PRS coefficient can be estimated by

$$\hat{d} = \frac{\sum_i o_i \hat{\beta}_i s_i^2}{\sum_i o_i s_i^2 - \bar{s}_i^2}. $$

where $\hat{\beta}_i$ are the discovery sample SNP effects, $o_i$ are the target sample SNP effects, and $s_i$ are the target sample standard errors for the ith SNP. Without individual-level genotype data, this method does not explicitly estimate a PRS—only its effect. The joint SNP effect estimate inherits covariate adjustment from covariates included in the contributing GWASs, for example, population structure axes.19 The approach (described in detail elsewhere20) is implemented in the package gtx (gtx: Genetics ToolboX, http://cran.r-project.org/web/packages/gtx)21 in the language and statistical computing environment R (http://www.R-project.org/).22

Genetic Covariation
Finally, using genotype-level data for the TwinsUK sample and the UK and German subsamples of the WTCCC2, we validated findings from the concordance and polygenic risk analyses by estimating the phenotypic covariation between the 2 traits (ie, pleiotropy—$\kappa_{\text{poly}}$), as well as genetic contribution to phenotypic variance within trait (ie, univariate SNP heritability—$h^2_{\text{uni}}$). These estimates are based on distant relatedness calculated from common SNPs. Estimation of SNP heritability and pleiotropy between complex traits using SNP-derived genomic relationships and restricted maximum likelihood is implemented in the program Genome-wide Complex Trait Analysis.5,23 Specifically, we used the genetic covariance estimation with genotype data to validate findings from the concordance and polygenic risk analyses with GWAS summary results.

Results
Concordance of Genetic Effects
Three coagulation factors showed excess concordance (or discordance) of effect direction with stroke: SNPs significant at $P \leq 0.3$ in factor XIII subunit B (FXIIIIB) were significantly discordant with the CE subtype (most significant $P=5.7 \times 10^{-14}$); SNPs significant at $P \leq 0.05$ through $P \leq 0.2$ in factor VIII (FVIII) were significantly concordant with the SVD subtype (most significant binomial $P=6.5 \times 10^{-14}$), but not at more liberal inclusion thresholds; SNPs significant at $P \leq 0.3$ only in Von Willebrand factor were significantly concordant with the SVD subtype. Figure 1 summarizes the evidence for excess concordance (or discordance) of effect between the 14 coagulation factors and ischemic stroke and its 3 subtypes.

Polygenic Risk Scores
All ischemic stroke and the CE subtype, but not the LVD or SVD subtypes, were significantly associated with coagulation polygenic risk. For all ischemic stroke, PRSs from 3 coagulation factors were significant: FVIIIC% at inclusion thresholds $P \leq 0.2$ and 0.4 (most significant $P=4.9 \times 10^{-14}$); FX at inclusion threshold greater than $P \leq 0.4$ (most significant $P=1.5 \times 10^{-14}$); and FXIIIIB for inclusion thresholds greater than $P \leq 0.01$ (most significant $P=1.8 \times 10^{-14}$). For the CE subtype, FXIIIIB was significantly predictive at all inclusion thresholds greater than $P \leq 0.2$ (most significant $P=1.7 \times 10^{-14}$), although the proportion of variance explained ($R^2$) was modest (<1%) at all inclusion thresholds. FXIIIIB was the one coagulation factor that showed consistent effects in both the concordance analyses and the PRS analyses. Figure 2 shows the increasing explanatory power of FXIIIIB for all ischemic stroke and CE, as we increased the number of FXIIIIB-associated risk variants included in the stroke prediction model.

In the target sample analyses, we found that genetic variation associated with lower FXIIIIB levels was consistently predictive of higher stroke risk: genetic risk was significantly discordant for FXIIIIB and stroke, and a negative association between polygenic risk based on FXIIIIB indicates that controls had more FXIIIIB-associated SNPs than stroke cases.

Genetic Covariation
Finally, to validate the association with FXIIIIB using an alternative approach, we directly estimated from genotype data the genetic covariation between stroke and FXIIIIB. We found evidence of a significantly high level of shared genetic risk for
FXIIIB–all ischemic stroke (rG SNP=−0.71, SE=0.81, \(P\) =0.01, \(N=10,668\); \(h^2\) SNP FXIIIB =0.13, SE=0.28; \(h^2\) SNP ischemic stroke =0.45, SE=0.04). Similarly, bivariate GREML analysis for FXIIIB–CE suggested a significantly high degree of common genetic risk (rG SNP=−0.80, SE=0.90, \(P\) =0.03, \(N=7917\); \(h^2\) SNP FXIIIB =0.13, SE=0.28; \(h^2\) SNP CE =0.34, SE=0.05).
The negative genetic covariance between FⅩⅢB and both all ischemic stroke and the CE subtype indicates that genetic variation predisposing to low levels of FⅩⅢB is associated with increased risk for stroke. A negative genetic correlation is consistent with our results above: FⅩⅢB had discordant SNP effects with CE stroke and the FⅩⅢB-associated PRS for stroke was lower in cases.

The large standard errors associated with these estimates are typical for small sample sizes in bivariate GREML analyses. From the combined TwinsUK-WTCCC2 sample, we excluded one individual from each pair with an estimated relatedness above 10%. We chose this cut-off (which removes up to first cousin relationships) because of the coagulation sample size: for example, of the 10 668 individuals in the FⅩⅢB-ischemic stroke analysis, 1940 were TwinsUK samples. Using the default Genome-wide Complex Trait Analysis relatedness cut-off of 2.5%, we found similarly high point estimates with much larger standard errors (FⅩⅢB-ischemic stroke: \( r_{\text{G}}=0.85, \text{SE}=2.69, \text{P}=0.06 \)), suggesting that a larger sample would yield similar pleiotropy estimates with narrower confidence intervals.

**Discussion**

Here, we used statistical genetic approaches to evaluate the genetic influence of coagulation traits on ischemic stroke and its subtypes. Our results indicate that the aggregate effect of risk SNPs for the plasma protein FⅩⅢB has a small, but significant effect on risk for ischemic stroke, specifically for the CE subtype, but not LVD or SVD stroke. Genetic factors that increased levels of FⅩⅢB were associated with decreased ischemic stroke risk. This result was identified using 3, but related, statistical analysis approaches.

A common genetic contribution to higher levels of FⅩⅢB and lower risk of ischemic stroke might seem counterintuitive. Two A subunits (FⅩⅢA) and 2 B subunits (FⅩⅢB) make up the FⅩⅢ tetramer whose main function is to strengthen and protect the fibrin clot against degradation during clot formation. Individuals with FⅩⅢ deficiency manifest a severe susceptibility to hemorrhage (or bleeding diathesis). However, as well as its clot stabilizing or prothrombotic effects, FⅩⅢ also has an antithrombotic effect by inhibiting platelet aggregation. That said, it is not clear what effect high plasma levels of free-floating FⅩⅢB (measured by the assay) would have on levels of the FⅩⅢ tetramer and how a change in levels of FⅩⅢ would affect the balance of its pro- and antithrombotic effects. At a 2.8-year poststroke follow up for mortality, FⅩⅢB and FⅩⅢA have been found to be present at significantly higher levels in survivors compared with those that had died. One particular polymorphism, FⅩⅢVal34Leu, has been reported to provide a protective effect against venous thromboembolism and myocardial infarction in Caucasians (in the absence of insulin resistance). The same polymorphism conveys no protective effect in South Asians, known to have higher levels of insulin resistance. FⅩⅢB was in fact found to be higher among South Asian ischemic stroke cases even after accounting for insulin resistance. However, as polygenic risk scoring approaches by design capture net genetic effects, the risk score based on FⅩⅢB-associated SNPs is the sum of many SNPs (of which the base change at the FⅩⅢVal34Leu polymorphism is just one), potentially acting both to increase and decrease levels of FⅩⅢB. The biological role of FⅩⅢB is complex, and further research is required to understand the specific mechanism underlying the observed protective effect.

In contrast, we found no association with LVD. Thromboembolism is believed to be the primary pathophysiological mechanism in LVD as in CE, but the underlying thrombotic mechanisms may differ in LVD. This is supported by clinical trial data, suggesting antiplatelet agents seem to be the most effective secondary prevention approach in LVD, although it is worth noting that the superiority of antiplatelets or anticoagulants for ischemic stroke of non-CE origin is not unequivocal. In contrast, anticoagulants have consistently been shown to be more effective than antiplatelets in CE stroke. There was also no association with SVD, and this would be consistent with a lack of pathophysiological evidence linking thrombosis with this stroke subtype.

This study does have limitations. First, although our stroke sample was relatively large, numbers of cases in the subtype analyses were necessarily smaller. Similarly, sample sizes for the coagulation factor GWASs were modest. Second, we benefitted from aggregating evidence across many SNPs, including those that did not achieve genome-wide significance in earlier genetic studies of coagulation and stroke, but this provides no resolution at individual SNPs, and we were consequently unable to comment on specific FⅩⅢB polymorphisms previously reported to be associated with thrombotic disease (stroke). Third, our final technique that tested genetic covariation used a subset of the METASTROKE sample and so was not a true replication in an independent data set. However, because the genetic covariation technique used SNP-level information for each individual in both the coagulation and stroke studies (as opposed to summary statistics from the overall GWASs), we included it as a validation of polygenic risk estimation using only GWAS meta-analysis results from the stroke studies.

A recent review of the role of FⅩⅢ in the risk for thrombotic diseases found only 4 studies that investigated FⅩⅢ levels and ischemic stroke. It concluded that no clear picture emerged and called for well-designed studies with clear differentiation of stroke subtype to clarify the problem. By aggregating evidence across the entire genome, our study found a link between the genetic influence on FⅩⅢ levels and the CE subtype, narrowing the focus to a specific subtype and highlighting a direction for further investigation into the mechanism behind the role of FⅩⅢB in aberrant coagulation and the pathogenesis of ischemic stroke.

**Sources of Funding**

Funding for collection, genotyping, and analysis of stroke samples was provided by Wellcome Trust Case Control Consortium-2. We acknowledge the use of the British 1958 Birth Cohort DNA collection and UK National Blood Service controls. The research was funded by the National Institute for Health Research Biomedical Research Centre based at Guy’s and St Thomas’ National Health Service Foundation Trust and King’s College London, and the Biomedical Research Centre for Mental Health at South London and Maudsley NHS Foundation Trust and King’s College London. The views expressed are those of the author(s) and not necessarily...
those of the National Health Service, the National Institute for Health Research, or the Department of Health. H.S. Markus is supported by a National Institute for Health Research Senior Investigator award. H.S. Markus and S. Bevan are supported by the National Institute for Health Research Cambridge University Hospitals Comprehensive Biomedical Research Centre. M. Taylor is supported by a project grant from the Stroke Association (TSA 2013/01). B.B. Worrall is supported by a grant from the National Institutes of Health (NIH).

Disclosures

Dr Worrall is an associate editor for *Neurology*. The other authors report no conflicts.

References


Genetic Factors Influencing Coagulation Factor XIII B-Subunit Contribute to Risk of Ischemic Stroke

Ken B. Hanscombe, Matthew Traylor, Pirro G. Hysi, Stephen Bevan, Martin Dichgans, Peter M. Rothwell, Bradford B. Worrall, Sudha Seshadri, Cathie Sudlow, Frances M.K. Williams, Hugh S. Markus and Cathryn M. Lewis

Stroke. published online July 9, 2015;
Stroke is published by the American Heart Association, 7272 Greenville Avenue, Dallas, TX 75231
Copyright © 2015 American Heart Association, Inc. All rights reserved.
Print ISSN: 0039-2499. Online ISSN: 1524-4628

The online version of this article, along with updated information and services, is located on the World Wide Web at:
http://stroke.ahajournals.org/content/early/2015/07/09/STROKEAHA.115.009387
Free via Open Access