Shared genetic basis for migraine and ischemic stroke

A genome-wide analysis of common variants

Rainer Malik, PhD
Tobias Freilinger, MD
Bendik S. Winsvold, PhD
Verneri Anttila, PhD
Jason Vander Heiden, MSc
Matthew Traylor, PhD
Boukje de Vries, PhD
Elizabeth G. Holliday, PhD
Gisela M. Terwindt, MD, PhD
Jonathan Sturm, PhD
Joshua C. Bis, PhD
Jemma C. Hopewell, PhD
Michel D. Ferrari, MD, PhD
Kristiina Rannikmae, MD
Maija Wessman, PhD
Mikko Kallela, MD, PhD
Christian Kubisch, MD
Myriam Fornage, PhD
James F. Meschia, MD
Terho Lehtimäki, MD
Cathie Sudlow, FRCP
Robert Clarke, FRCP
Daniel I. Chasman, PhD
Braxton D. Mitchell, PhD
Jane Maguire, PhD
Jaakko Kaprio, MD, PhD
Martin Farrall, FRCPath
Olli T. Raitakari, MD, PhD
Tobias Kurth, MD, ScD
M. Arfan Ikram, MD
Alex P. Reiner, PhD
W.T. Longstreth, Jr., MD
Peter M. Rothwell, FMedSci
David P. Strachan, MD
Pankaj Sharma, PhD

Author list continued on next page

ABSTRACT

Objective: To quantify genetic overlap between migraine and ischemic stroke (IS) with respect to common genetic variation.

Methods: We applied 4 different approaches to large-scale meta-analyses of genome-wide data on migraine (23,285 cases and 95,425 controls) and IS (12,389 cases and 62,004 controls). First, we queried known genome-wide significant loci for both disorders, looking for potential overlap of signals. We then analyzed the overall shared genetic load using polygenic scores and estimated the genetic correlation between disease subtypes using data derived from these models. We further interrogated genomic regions of shared risk using analysis of covariance patterns between the 2 phenotypes using cross-phenotype spatial mapping.

Results: We found substantial genetic overlap between migraine and IS using all 4 approaches. Migraine without aura (MO) showed much stronger overlap with IS and its subtypes than migraine with aura (MA). The strongest overlap existed between MO and large artery stroke (LAS; $p = 6.4 \times 10^{-28}$ for the LAS polygenic score in MO) and between MO and cardioembolic stroke (CE; $p = 2.7 \times 10^{-20}$ for the CE score in MO).

Conclusions: Our findings indicate shared genetic susceptibility to migraine and IS, with a particularly strong overlap between MO and both LAS and CE pointing towards shared mechanisms. Our observations on MA are consistent with a limited role of common genetic variants in this subtype. *Neurology*® 2015;84:2132–2145

GLOSSARY

CE = cardioembolic stroke; **CPSM** = cross-phenotype spatial mapping; **GWAS** = genome-wide association studies; **IHGC** = International Headache Genetics Consortium; **IS** = ischemic stroke; **LAS** = large artery stroke; **LD** = linkage disequilibrium; **MA** = migraine with aura; **MO** = migraine without aura; **SNP** = single nucleotide polymorphism; **SVD** = small vessel disease.

Migraine is a primary headache disorder characterized by recurrent attacks of severe, often throbbing, headache associated with autonomic dysfunction. Although the majority of patients have migraine without aura (MO), one third have headaches preceded by transient neurologic disturbances (migraine with aura [MA]).[1] Ischemic stroke (IS) is etiologically heterogeneous and a leading cause of premature death and disability.[2]

Results of epidemiologic studies show increased risk of IS in migraine patients.[3] A large meta-analysis of case-control and observational cohort studies reported an increased risk of IS for patients with MO and MA,[4] whereas more recent meta-analyses reported the association to be restricted to MA.[3,5,6] Pathophysiology linking these neurovascular disorders remains poorly understood; suggested mechanisms include cortical spreading depression,[7] endothelial dysfunction,[8] enhanced platelet activation,[9] and vasoconstriction.[10]

Recent genome-wide association studies (GWAS) identified common genetic variants associated with migraine[11] and its subtypes MO[12] and MA.[13] Similarly, GWAS results point to variants associated with IS subtypes such as large artery atherosclerotic[14,15] and cardioembolic.[16] We combined GWAS from the International Headache Genetics Consortium (IHGC)[11] and METASTROKE[15] to assess shared genetic bases for migraine and IS. We first

Author affiliations are provided at the end of the article.

Coinvestigators are listed on the *Neurology*® Web site at Neurology.org.

Go to Neurology.org for full disclosures. Funding information and disclosures deemed relevant by the authors, if any, are provided at the end of the article.

Sudha Seshadri, MD
Lydia Quaye, PhD
Lynn Cherkas, PhD
Markus Schürks, MD, MSc
Jonathan Rosand, MD
Lannie Ligthart, PhD
Giorgio B. Boncoraglio, MD
George Davey Smith, DSc
Cornelia M. van Duijn, PhD
Kari Stefansson, MD, PhD
Bradford B. Worrall, MD, MSc
Dale R. Nyholt, PhD
Hugh S. Markus, FRCP
Arn M.J.M. van den Maagdenberg, PhD
Chris Cotsapas, PhD
John A. Zwart, MD, PhD
Aarno Palotie, MD
For the International Headache Genetics Consortium
Martin Dichgans, MD
For the METASTROKE Collaboration of the International Stroke Genetics Consortium

Correspondence to Dr. Dichgans: martin.dichgans@med.uni-muenchen.de

examined known genome-wide risk loci in the respective phenotypes. Using 2 methodologies, we then evaluated shared genetic risk for migraine with IS: (1) analysis of shared polygenic risk with subsequent estimation of genetic correlation between phenotypes and (2) detailed investigation of overlapping regions.

METHODS

Standard protocol approvals, registrations, and patient consents. ***Ethics statement.*** For all study cohorts, all participants gave informed consent and local research ethics boards approved all protocols.[11,15]

Cohorts. Investigators of the IHGC study, a meta-analysis of GWAS data, enrolled 23,285 patients with migraine and 95,425 population-based or headache-free controls from 29 studies.[11] When possible, researchers considered subclassifications of migraine with (MA: 5,118 cases vs 74,239 controls) or without aura (MO: 7,107 cases vs 69,427 controls). The METASTROKE study consists of combined data from 15 GWAS of IS (12,389 cases vs 62,004 controls).[15] We used TOAST criteria[17] to classify IS as large artery stroke (LAS) (2,167 cases/49,159 controls from 11 studies), cardioembolic stroke (CE) (2,365 cases/56,140 controls from 13 studies), and small vessel disease (SVD) (1,894 cases/51,976 controls from 12 studies) (tables e-1 and e-2 on the *Neurology*® Web site at Neurology.org).[11,15] We removed overlapping controls between the migraine and stroke samples from deCODE, WTCCC2 (B58C and KORA), and the Rotterdam studies from the stroke datasets for polygenic score analyses, cross-phenotype spatial mapping (CPSM), and correlation analyses to avoid inflation of statistics.

Genome-wide association studies. Both the IHGC migraine[11] and METASTROKE[15] studies consisted of independently performed genome-wide single nucleotide polymorphism (SNP) genotyping using standard technologies and imputation to HapMap release 21 or 22 CEU phased genotype[18] or 1000 Genome reference panels. Investigators contributed summary statistical data from association analyses using frequentist additive models for meta-analysis after application of appropriate quality control measures (e-Methods). Subtle differences in allele frequencies between migraine and stroke could lead to deviation from the expected test statistic. Thus, as a final quality control step, we meta-analyzed results from the IHGC study and the METASTROKE study and constructed quantile-quantile plots (figure e-1).

Statistical analysis. For analysis of previously discovered risk loci for IS or migraine, we extracted relevant loci from the literature and the 2 described consortia.[11,15] We examined all SNPs within a window of ±50 kb surrounding the original reported risk SNP (coordinates from human genome build hg18) and reported the most significant p values of all genotyped or imputed SNPs within this window. We applied Bonferroni correction for association, integrating all tested SNPs for IS risk loci (650 tested SNPs), migraine risk loci (1,175 tested SNPs), and MO risk loci (213 tested SNPs) with resulting p value thresholds of 7.69E-5, 4.25E-5, and 2.30E-4, respectively.

Polygenic scores reveal combined effects of multiple nonsignificant variants derived from a derivation sample and tested in an independent replication sample. We derived polygenic scores for multiple p value cutoffs (0.5, 0.25, 0.1, 0.05, 0.01, 0.001, and 0.0001) in derivation samples. Further, we performed testing on summary statistics using the grs.summary function of the gtx package for *R*, a technique previously used in multiple studies, which estimates the polygenic component with high reliability.[19] We use the term replication to describe analyses across phenotypes.

Use of linkage disequilibrium (LD) pruned data ($r^2 > 0.25$) ensured approximate independence of SNPs. We retained the SNP with the lowest p value in an independent region and calculated the proportion of variance explained in the testing set by the polygenic scores using Nagelkerke's pseudo R^2. Outcome measures include the p value of the association of the polygenic score in the testing dataset and the variance explained.

CPSM identifies genomic windows exhibiting similar association patterns across 2 phenotypes using a signal processing approach. CPSM allows analysis of pleiotropy across multiple diseases. Peak heights serve as an intuitive measure for description of shared risk loci in different phenotypes. This method corrects for background noise in the p value distribution and is thus superior to comparisons of single p values. We computed Pearson covariance between p values from 2 traits across a 60-kb sliding window. In each iteration, the window slides to the next SNP; thus, we obtained a covariance coefficient for each SNP in the analysis. We then detected signal peaks across the genome in the covariance trace and calculated the signal s_n for a given SNP with index n, position b_n (base pairs), and association p values $p1_{,n}$, $p2_{,n}$ for 2 phenotypes as follows:

$$x = -\log(10)p_{1,j}, \ldots. -\log(10)\,p_{1,k}$$

$$y = -\log(10)p_{2,j}, \ldots\ldots -\log(10)\,p_{2,k}$$

$$w = \left[1 - \frac{|b_j - b_n|}{b_k - b_j}, \ldots, 1 - \frac{|b_k - b_n|}{b_k - b_j}\right]$$

$$w' = \frac{1}{\sum_{i=j}^{k} w_i}.w$$

$$s_n = \frac{\sum_{i=j}^{k} w'_i x_i y_i}{1 - \sum_{i=j}^{k} w'^2_i},$$

where each $b_i \in b_j \ldots b_k$ is the position of SNP_i within the window of $SNPs_{j,\ldots,k}$ containing SNPn. For a given window size d (base pairs), the window of $SNPs_{j,\ldots,k}$ is defined such that j is the smallest SNP index where $b_n - b_j \leq \frac{d}{2}$ and k is the largest SNP index where $b_k - b_n \leq \frac{d}{2}$.

After constructing the CPSM signal for all SNPs, we corrected for strong associations identified in only one phenotype by permuting the association p values for one phenotype 1,000 times while holding the other phenotype constant, and then recalculating CPSM signals. From the total set of 2,000 permutation signals (1,000 per phenotype), we subtracted the upper 0.95 quantile at each SNP as the background signal threshold from the observed covariance as a correction. We then identified regions of shared association as peaks above the 99.95 (approximately corresponding to a height of 1.5) percentile of the covariance signal. We highlighted regions with a height >2.5 (corresponding to approximately 99.985 percentile) and with a height >5 (corresponding to approximately 99.998 percentile). CPSM only provides a signal when the effect in 2 traits is the same, implying shared causality in the discovered regions.

Utilizing a recently developed framework for polygenic analyses and based on the number of SNPs, the dataset sample sizes, and estimates of disease prevalence and pseudo-heritability, we estimated the power to detect an association indexing on a given degree of genetic correlation between the 2 phenotypes. We used the same framework, including p values from polygenic analysis, to estimate the overall degree of genetic correlation between each of the IS and migraine phenotypes, a posteriori to the polygenic

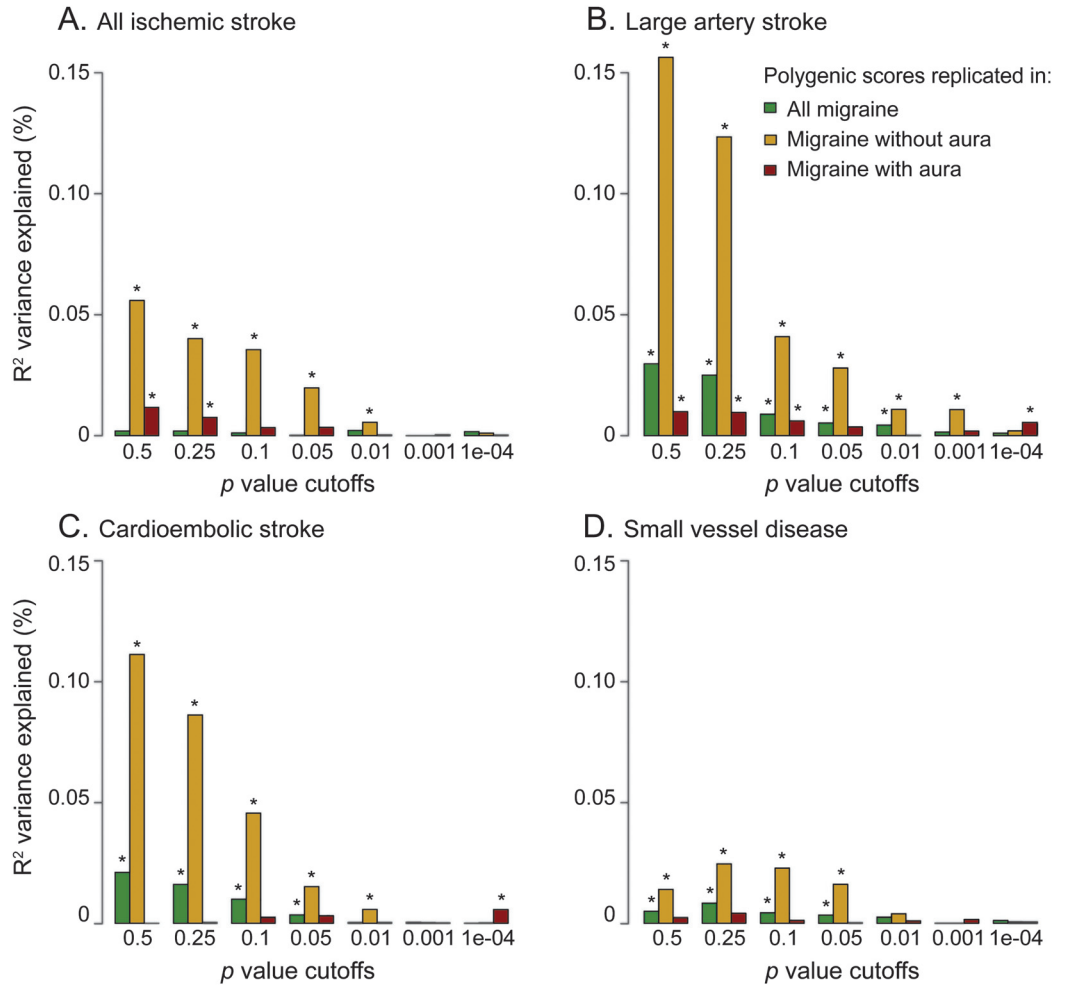
Supplemental data at Neurology.org

analysis. We estimated genetic correlation in both the forward direction (using results from polygenic analysis of IS and subtypes as a discovery sample and migraine and subtypes as a replication sample) and the reverse direction (using results from the polygenic analysis from migraine and subtypes as a discovery sample and stroke and subtypes as a replication sample) to evaluate consistency of results using the estimateCorrFromP method. An implementation of the procedure was downloaded from http://sites.google.com/site/fdudbridge. This method approximates SNP correlation from cross-disorder applications of polygenic scores and can be compared to GREML-SNP genetic correlation. All analyses used *R* statistical software (http://www.R-project.org). Using stroke prevalence data from the British Heart Foundation for IS[20] (1.7% in the United Kingdom) and the proportional incidence of IS events from all stroke events in the OXVASC study[21] (59%), we estimated the prevalence of IS (~1%). We then used the proportion of IS subtypes (CE, LAS, or SVD) from a meta-analysis of population-based incidence studies[22] to estimate the prevalence of each subtype. We estimated stroke heritability on a liability scale.[23] Although we acknowledge that migraine prevalence may vary across countries, we estimated migraine prevalence to be 17% for all migraine, 11% for MO, and 5% for MA based on published data.[1,24] Migraine heritability estimates vary in the literature, with MA being highest. We chose heritability measures of 0.65 for MA[25] and 0.61 for MO[26] and a more conservative measure of 0.57 for all migraine.

RESULTS Information on clinical subtypes was available for 12,225 (52.5%) of the migraine and for 6,426 (51.9%) of the IS patients (tables e-1 and e-2). We identified 38,338 potentially overlapping controls and excluded them from analyses where necessary. QQ plots revealed no inflation of test statistics (lambda inflation factors below 1.05 in all analyses of migraine subtypes vs all IS; figure e-1 and e-Methods).

All migraine. We first evaluated risk loci identified in previous GWAS on IS or its subtypes,[15] in all migraine[11] and vice versa. Although we identified several variants reaching nominal association ($p < 0.05$), when controlling for all tested SNPs, none of the

Figure 1 Results from polygenic score analysis using ischemic stroke as a discovery phenotype



(A) All ischemic stroke. (B) Large artery stroke. (C) Cardioembolic stroke. (D) Small vessel disease. Migraine was used as a replication phenotype. The x-axis describes the *p* value cutoffs used in the polygenic score; the y-axis describes the pseudo-R^2 variance explained by the score. Asterisks on top of a bar designate *p* values < 0.05. Raw values can be found in table e-5.