ONLINE SUPPLEMENT

Heritability Estimates Identify a Substantial Genetic Contribution to Risk and Outcome of Intracerebral Hemorrhage.

Devan WJ et. al.

Content:

SUPPLEMENTAL METHODS

APOE gene region capture

Given the limited capture of APOE ε2 / ε3 / ε4 alleles on commercially available genotyping array,[1] these APOE variants were genotyped directly according to previously published methods.[2] To avoid overestimation of the contribution from the APOE locus to heritability estimates, we removed all array markers in the APOE gene +/- 55 kilobase pairs on each side prior to performing all analyses for non-APOE loci. This ensured that no markers with linkage disequilibrium relationships with the ε2 / ε3 / ε4 alleles would be included in the non-APOE loci analysis, thus unduly influencing results.

Sensitivity analyses were performed by including all the array markers in the APOE gene region that qualified for removal from analysis based on the criteria detailed above. These repeat analyses showed no change in heritability estimates for both all ICH and lobar ICH (data not shown).

Data Quality Control (QC)

DNA from cases and controls from all centers was isolated from fresh or frozen blood and sent to Coordinating Center (Massachusetts General Hospital) for further steps. DNA was then quantified with a quantification kit (Qiagen, Valencia, CA, USA) and normalized to a concentration of 30 ng/μL. GW genotyping was performed using Illumina 610k chip (Illumina HumanHap 610-Quad SNP array, San Diego, CA, USA) at the Broad Institute. GW data went through a series of quality control metrics aimed at identifying individual subjects with poor quality genotyping, as well as individual markers (also known as Single Nucleotide Polymorphisms, SNPs) with evidence of inadequate / incomplete capture. QC procedures excluded SNPs missing in > 5% of samples, with a MAF < 1%, or with an excessive deviation from HWE p<1E-06.

Individuals were removed if their inferred genotype gender was discordant with the recorded gender, or they had more than 5% missing genotype data. Individuals were also excluded if they were found to have familial relatedness - first cousin relationship or closer (one of each pair demonstrating identity-by-descent pi-hat > 0.15 was removed) - as their inclusion might unduly influence our findings. Additionally, Principal Component Analysis (PCA) was performed on GW data using HapMap Phase 3 populations to identify subjects of non-European descent. These subjects were therefore removed from all further analysis. Principal components were subsequently extracted for use in heritability estimation as covariates (see below). QC and PCA were performed using PLINK v1.6 using previously validated protocols.[3]

ICH Risk

ICH risk heritability was estimated by analyzing the proportion of variance in case/control status accounted for by APOE and non-APOE loci. Principal components 1 and 2, age at time of enrollment, and sex were entered into the ICH risk heritability model to account for the environmental portion of the trait variance. Initial analyses included all ICH patients (defined as ICH cases and compared to controls). For location-defined subset analyses, we included only ICH cases in the lobar or deep location (thus excluding cases with cerebellar, mixed lobar/deep, primary IVH, and undetermined ICH locations); all available controls were included in these subset analyses (Figure 1).

We subsequently repeated location-defined subset heritability analyses after removal of known genetic risk factors. For lobar ICH, a variant in the CR1 gene (SNP rs6656401) was previously

shown to be associated with cerebral amyloid angiopathy (CAA)-ICH.[4] We therefore removed the entire CR1 gene region (the CR1 gene +/- 55 kilobase pairs on each side), prior to repeating lobar ICH risk heritability estimation.

Similarly, a gene-score including all variants associated with systolic and diastolic blood pressure (at $p < 5.0 \times 10^{-5}$ from the NHGRI GWAS association catalogue) was found to associated with deep-ICH risk.[5] We therefore removed all these variants prior to repeating deep ICH heritability estimation. Furthermore, to ensure no remaining signal from correlated variants could be left behind, we also removed variants in linkage disequilibrium with the BP variants ($r^2>0.5$).

ICH 90-day Mortality and Admission Volumes

Since outcome and ICH volume data were only available for subjects enrolled in GOCHA, mortality and hematoma size heritability was calculated only for these individuals.

For ICH mortality, cases were defined as ICH patients who died within 90 days of their acute ICH event; controls were defined as ICH cases who survived greater than 90 days after the acute ICH event. Principal components 1 through 2, age, admission ICH volume, sex, warfarin, aspirin, hypertension, diabetes mellitus, hyperlipidemia, coronary heart disease, alcohol abuse, and smoking were entered into the model to account for the environmental portion of the variance.

For ICH admission volume heritability estimation, hematoma size was measured using a published semi-automated method.[6–8] ICH volume measurements were subsequently log-transformed to achieve normality, and analyzed as a continuous phenotype. Principal components 1 through 2, age, time from symptom onset to CT scan, sex, warfarin, aspirin, hypertension, diabetes mellitus, hyperlipidemia, coronary heart disease, alcohol abuse, and smoking were entered into the model to account for the environmental portion of the variance.

Heritability Estimates after Stringent QC

We investigated the impact of SNP missingness and MAF on ICH heritability. When we impose more stringent SNP missingness and MAF thresholds the number of SNPs drops from 490,328 (MAF > 0.05 & missingness > 0.05) to 402,765 (MAF > 0.001 & missingness > 0.05) (Supplemental Table S1). The heritability estimates for all ICH and both subtypes (deep and lobar) slightly decrease as well. However as noted previously in a previous publication this small decrease is most likely due to the number of SNPs used to estimate the heritability rather than an artifact of insufficient SNP quality control.[9]

SUPPLEMENTAL REFERENCES

1. Valant V, Keenan B, Anderson C, Shulman J, Devan W, Ayres A. TOMM40 in Cerebral Amyloid Angiopathy Related Intracerebral Hemorrhage: Comparative Genetic Analysis with Alzheimer's Disease. *Translational Stroke Research*. 2012;3:102–112.

2. Greenberg SM, Rebeck GW, Vonsattel JP, Gomez-Isla T, Hyman BT. Apolipoprotein E epsilon 4 and cerebral hemorrhage associated with amyloid angiopathy. *Ann. Neurol.* 1995;38:254–259.

3. Biffi A, Anderson CD, Desikan RS, Sabuncu M, Cortellini L, Schmansky N. Genetic variation and neuroimaging measures in Alzheimer disease. *Arch. Neurol.* 2010;67:677–685.

4. Biffi A, Shulman JM, Jagiella JM, Cortellini L, Ayres AM, Schwab K. Genetic variation at CR1 increases risk of cerebral amyloid angiopathy. *Neurology*. 2012;78:334–341.

5. Ehret GB, Munroe PB, Rice KM, Bochud M, Johnson AD, Chasman DI. Genetic variants in novel pathways influence blood pressure and cardiovascular disease risk. *Nature*. 2011;478:103–109.

6. Biffi A, Sonni A, Anderson CD, Kissela B, Jagiella JM, Schmidt H. Variants at APOE influence risk of deep and lobar intracerebral hemorrhage. *Ann. Neurol.* 2010;68:934–943.

7. Brouwers HB, Biffi A, Ayres AM, Schwab K, Cortellini L, Romero JM. Apolipoprotein E Genotype Predicts Hematoma Expansion in Lobar Intracerebral Hemorrhage. *Stroke*. 2012;43:1490–1495.

8. Flibotte JJ, Hagan N, O'Donnell J, Greenberg SM, Rosand J. Warfarin, hematoma expansion, and outcome of intracerebral hemorrhage. *Neurology*. 2004;63:1059–1064.

9. Lee SH, Wray NR, Goddard ME, Visscher PM. Estimating missing heritability for disease from genome-wide association studies. *Am. J. Hum. Genet.* 2011;88:294–305.

Supplemental Table S1. ICH Risk Heritability Estimates Stratified by MAF and SNP Missingness

			All ICH			Deep ICH			Lobar ICH		
MAF Threshold	**Genotype Missingness (Miss) Threshold**	**No. SNP**	**N**	**Heritability Estimate (%, SE)**	**p-value**	**N**	**Heritability Estimate (%, SE)**	**p-value**	**N**	**Heritability Estimate (%, SE)**	**p-value**
MAF > 0.01	Miss > 0.05	490328	1667	29 (11)	0.001	1242	30 (17)	0.023	1214	48 (18)	0.002
	Miss > 0.01	482860	1667	29 (11)	0.001	1242	30 (17)	0.024	1214	49 (18)	0.001
	Miss > 0.005	474163	1667	29 (11)	0.001	1242	29 (17)	0.026	1214	48 (18)	0.002
	Miss > 0.001	423355	1667	29 (10)	0.001	1242	29 (17)	0.024	1214	48 (18)	0.002
MAF > 0.05	Miss > 0.05	467845	1667	28 (10)	0.002	1242	27 (16)	0.034	1214	47 (17)	0.002
	Miss > 0.01	460391	1667	28 (10)	0.001	1242	26 (16)	0.036	1214	47 (17)	0.002
	Miss > 0.005	451762	1667	27 (10)	0.002	1242	26 (16)	0.039	1214	46 (17)	0.002
	Miss > 0.001	402765	1667	26 (10)	0.002	1242	25 (16)	0.041	1214	46 (17)	0.002