MULTI-ANCESTRY GENETIC STUDY IN 5,876 PATIENTS IDENTIFIES AN ASSOCIATION BETWEEN EXCITOTOXIC GENES AND EARLY OUTCOMES AFTER ACUTE ISCHEMIC STROKE.

Laura Ibanez, PhD[1,2]; Laura Heitsch, MD[3,4]; Caty Carrera, MD[5]; Fabiana H.G. Farias, PhD[1,2]; Rajat Dhar, MD[3]; John Budde[1,2]; Kristy Bergmann[1,2]; Joseph Bradley[1,2]; Oscar Harari, PhD[1,2,7,8]; Chia-Ling Phuah, MD[3]; Robin Lemmens PhD[8]; Alessandro A. Viana Oliveira Souza, MD[9,10], Francisco Moniche, PhD[11]; Antonio Cabezas-Juan[11,12]; Juan Francisco Arenillas, PhD[13]; Jerzy Krupinksi, PhD[14,15]; Natalia Cullell[15,16]; Nuria Torres-Aguila[15,16]; Elena Muiño, PhD[16]; Jara Cárcel-Márquez [16]; Joan Marti-Fabregas, PhD[16]; Raquel Delgado-Mederos, PhD[16]; Rebeca Marin-Bueno, RN[16]; Alejandro Hornick, MD[17]; Cristofol Vives-Bauza PhD[18]; Rosa Diaz Navarro, MD[19]; Silvia Tur, MD[19]; Carmen Jimenez, MD[19]; Victor Obach, MD[20]; Tomas Segura, MD[21]; Gemma Serrano-Heras, PhD[21]; Jong-Won Chung, PhD[22]; Jaume Roquer, PhD[23]; Carol Soriano-Tarraga PhD[1,2,23]; Eva Giralt-Steinhauer, PhD[23]; Marina Mola-Caminal, PhD[23,24]; Joanna Pera, PhD[25]; Katarzyna Lapicka-Bodzioch[25]; Justyna Derbisz[25]; Antoni Davalos, PhD[26]; Elena Lopez-Cancio, PhD[27]; Lucia Muñoz, PhD[26]; Turgut Tatlisumak, MD[28,29]; Carlos Molina, PhD[5]; Marc Ribo, MD[5]; Alejandro Bustamante, PhD[26]; Tomas Sobrino, PhD[30]; Jose Castillo-Sanchez, PhD[30]; Francisco Campos, PhD[30]; Emilio Rodriguez-Castro, PhD[30]; Susana Arias-Rivas, PhD[30]; Manuel Rodríguez-Yáñez, PhD[30]; Christina Herbosa, MD[3], Andria L. Ford, MD[3,6,31]; Antonio Arauz, PhD[32]; Iscia Lopes-Cendes, PhD[9,10]; Theodore Lowenkopf, MD[33]; Miguel A. Barboza, MD[34]; Hajar Amini, PhD[35]; Boryana Stamova, PhD[35]; Bradley P. Ander, PhD[35]; Frank R Sharp, MD[35]; Gyeong Moon Kim, MD[22]; Oh Young Bang, PhD[22]; Jordi Jimenez-Conde, PhD[23]; Agnieszka Slowik, PhD[25]; Daniel Stribian, PhD[36];; Ellen A. Tsai, PhD[37]; Linda C. Burkly, PhD[38]; Joan Montaner, PhD[5,11,12]; Israel Fernandez-Cadenas, PhD[5,16]; Jin-Moo Lee*, MD[3,6,31,39,40] and Carlos Cruchaga*, PhD[1,2,3,6,7,41]

*Co-Corresponding Authors

1. Department of Psychiatry, Washington University School of Medicine, 660 S. Euclid Avenue, Saint Louis (63110), Missouri, US
2. NeuroGenomics and Informatics, Washington University School of Medicine, 425 S. Euclid Avenue, Saint Louis (63110), Missouri, US
3. Department of Neurology, Washington University School of Medicine, 660 S. Euclid Avenue; Campus Box 8111; Saint Louis (63110), Missouri, US
4. Emergency Medicine, Washington University School of Medicine, 660 S. Euclid Avenue; Campus Box 8072; Saint Louis (63110), Missouri, US
5. Stroke Unit, Vall d'Hebron University Hospital, Universitat de Barcelona, Passeig de la Vall d'Hebron, 1198; Barcelona (08035), Spain
6. Hope Center for Neurological Disorders, Washington University School of Medicine, 660 S. Euclid Avenue; Campus Box 8111; Saint Louis (63110), Missouri, US
7. The Charles F. and Joanne Knight Alzheimer Disease Research Center, Washington University School of Medicine, 4488 Forest Park Avenue; Saint Louis (63110), Missouri, US
8. Department of Neuroscience, Katholieke Universiteit Leuven, Campus Gasthuisberg O&N2; Herestraat 49 box 1021; Leuven (BE-3000), Belgium
9. Department of Neurology, School of Medical Sciences, University of Campinas (UNICAMP), R. Tessalia Viera de Camargo, 126; Cidade Universitaria, Campinas (13083-887), Brazil
10. Brazilian Institute of Neuroscience and Neurotecnology (BRAINN), R. Tessalia Viera de Camargo, 126; Cidade Universitaria, Campinas (13083-887), Brazil
11. Department of neurology, Hospital Virgen del Rocio, University of Seville, Avenida Manuel Siurot, s/n; Seville (41013), Spain
12. Hospital Virgen de la Macarena, University of Seville, Calle Dr. Fedriani, 3; Seville (41009), Spain
13. Department of Neurology, Hospital Clinico Universitario Valladolid, Valladolid University, Avenida Ramon y Cajal, 3; Valladolid (47003), Spain
14. Department of Neurology, Mutua Terrassa University Hospital, Universitat de Barcelona, Plaça del Dr. Robert, 5; Terrassa (08221), Spain
15. Fundacio Docencia i Recerca Mutua Terrassa, Universitat de Barcelona, Carrer Sant Antoni, 19; Terrassa (08221), Spain
16. Department of Neurology, Hospital de la Santa Creu i Sant Pau, Universitat Autonoma de Barcelona, Carrer de Sant Quinti, 89; Barcelona (08041), Spain

17. Department of Neurology, Southern Illinois Healthcare Memorial Hospital of Carbondale, 405 W Jackson Street, Carbondale (62901), Illinois, US
18. Department of Biology, Universitat de les Illes Balears, Carretera de Valldemossa, km 7,5, Palma (07122), Spain
19. Department of Neurology, Hospital Universitari Son Espases, Universitat de les Illes Balears, Carretera de Valldemossa, 79, Palma (07120), Spain
20. Department of Neurology, Hospital Clinic de Barcelona, Universitat de Barcelona, Carrer Villarroel, 170, Barcelona (08036), Spain
21. Research Unit, Complejo Hospitalario Universitario de Albacete. Calle Laurel s/n. Albacete (02008), Spain
22. Department of Neurology, Samsung Medical Center, 81 Irwon-ro, Gangnam-gu, Seoul, South Korea
23. Neurovascular Research Group, Institut Hospital del Mar de Investigacions Mediques, Passeig Maritim, 25-29, Barcelona (08003), Spain
24. Department of Surgical Sciences, Orthopedics, Uppsala University, Uppsala, 75185, Sweden
25. Department of Neurology, Jagiellonian University, Golebia, 24, Krakow(31-007), Poland
26. Department of Neurology, Hospital Germans Trias i Pujol, Universitat Autonoma de Barcelona, Carretera de Canyet, s/n; Badalona (08916), Spain
27. Department of Neurology, Hospital Universitario Central de Asturias, Oviedo, Spain
28. Department of Neurology, Sahlgrenska University Hospital, University of Gothenburg, Bla straket, 5; Gothenburg (413 45), Sweden
29. Department of Clinical Neuroscience, Institute of Neuroscience and Physiology, Sahlgrenska Academy at University of Gothenburg, Gothenburg, Sweden
30. Clinical Neurosciences Research Laboratory, Health Research Institute of Santiago de Compostela (IDIS), Avda. Travesa da Choupana s/n; Santiago de Compostela (15706), Spain.
31. Department of Radiology, Washington University School of Medicine, 660 S. Euclid Avenue, Saint Louis (63110), Missouri, US
32. Instituto Nacional de Neurologia y Neurocirurgia de Mexico, Avenida Insurgentes Sur 3877, Ciudad de Mexico (14269), Mexico
33. Department of Neurology, Providence St. Vincent Medical Center, 9205 SW Barnes Rd, Portland (97225), Oregon, US
34. Neurosciences Department, Hospital Rafael A. Calderon Guardia, Avenidas 7 y 9, calles 15 y 17, Aranjuez, San José, Costa Rica
35. Department of Neurology and MIND Institute, University of California at Davis, 2825 5th street, Sacramento (95817), California, US
36. Department of Neurology, Helsinki University Hospital, Haartmaninkatu 4 Rakennus 1, Helsinki (00290), Finland
37. Translational Biology, Biogen, Inc, 115 Brodway, Cambridge (02142), Massachusetts, US
38. Genetics and Neurodevelomental Disease Research Unit, Biogen, Inc, 115 Brodway, Cambridge (02142), Massachusetts, US
39. Department of Biomedical Engineering, Washington University School of Medicine, 660 S. Euclid Avenue, Saint Louis (63110), Missouri, US
40. Stroke and Cerebrovascular Center, Washington University School of Medicine, One Barnes-Jewish Hospital Plaza, Saint Louis (63110), Missouri, US
41. Department of Genetics, Washington University School of Medicine, 4515 McKinley Ave, Saint Louis (63110), Missouri, US

Correspondence:
Carlos Cruchaga
Washington University School of Medicine
660 South Euclid Avenue
Campus Box 8134
Saint Louis, MO 63110
Tel: 314-286-0546 - Fax: 314-747-2983
cruchagac@wustl.edu

Jin-Moo Lee
Washington University School of Medicine
660 South Euclid Avenue
Campus Box 8111
St. Louis, MO 63110
Tel: 314-362-7382
leejm@wustl.edu

Cover title: GWAS reveals excitotoxicity associated with acute ischemic stroke
Word count: 4,971
Figures: 3
Tables: 2
Key words: Ischemic Stroke, Neuroprotection, Genetics, NIHSS
Term list: Ischemic Stroke, Genetics, Neuroprotection

ABSTRACT

During the first hours after stroke onset neurological deficits can be highly unstable: some patients rapidly improve, while others deteriorate. This early neurological instability has a major impact on long-term outcome. Here, we aimed to determine the genetic architecture of early neurological instability measured by the difference between NIH stroke scale (NIHSS) within six hours of stroke onset and NIHSS at 24h (ΔNIHSS). A total of 5,876 individuals from seven countries (Spain, Finland, Poland, United States, Costa Rica, Mexico and Korea) were studied using a multi-ancestry meta-analyses. We found that 8.7% of ΔNIHSS variance was explained by common genetic variations, and also that early neurological instability has a different genetic architecture than that of stroke risk. Seven loci (2p25.1, 2q31.2, 2q33.3, 4q34.3, 5q33.2, 6q26 and 7p21.1) were genome-wide significant and explained 2.1% of the variability suggesting that additional variants influence early change in neurological deficits. We used functional genomics and bioinformatic annotation to identify the genes driving the association from each loci. eQTL mapping and SMR indicate that *ADAM23* (log Bayes Factor (LBF)=6.34) was driving the association for 2q33.3. Gene based analyses suggested that *GRIA1* (LBF=5.26), which is predominantly expressed in brain, is the gene driving the association for the 5q33.2 locus. These analyses also nominated *PARK2* (LBF=5.30) and *ABCB5* (LBF=5.70) for the 6q26 and 7p21.1 loci. Human brain single nuclei RNA-seq indicates that the gene expression of *ADAM23* and *GRIA1* is enriched in neurons. *ADAM23*, a pre-synaptic protein, and *GRIA1*, a protein subunit of the AMPA receptor, are part of a synaptic protein complex that modulates neuronal excitability. These data provides the first evidence in humans that excitotoxicity may contribute to early neurological instability after acute ischemic stroke.

RESEARCH INTO CONTEXT

Evidence before this study: No previous genome-wide association studies have investigated the genetic architecture of early outcomes after ischemic stroke.
Added Value of this study: This is the first study that investigated genetic influences on early outcomes after ischemic stroke using a genome-wide approach, revealing seven genome-wide significant loci. A unique aspect of this genetic study is the inclusion of all of the major ethnicities by recruiting from participants throughout the world. Most genetic studies to date have been limited to populations of European ancestry.
Implications of all available evidence: The findings provide the first evidence that genes implicating excitotoxicity contribute to human acute ischemic stroke, and demonstrates proof of principle that GWAS of acute ischemic stroke patients can reveal mechanisms involved in ischemic brain injury.

Introduction

Stroke is the second most common cause of death and the most common cause of disability, worldwide.[1] Ischemic stroke, the most common subtype[2], is caused by the occlusion of an artery in the brain, resulting in the abrupt development of cerebral ischemia and neurological deficits.[3] During the first hours after stroke onset, neurological deficits can be highly unstable with some patients demonstrating rapid deterioration, while others rapidly improve.[4] In fact, early change in neurological deficits have a major influence on long-term outcome. NIH stroke scale (NIHSS) changes from baseline (within 6 hours of stroke onset) to 24 hours after acute ischemic stroke (ΔNIHSS) have a significant and independent association with favorable 90-day outcome, accounting for more than 30% of the explained variance.[4-6] A number of mechanisms are thought to account for these early changes including fibrinolysis and reperfusion, hemorrhagic transformation, etiology, and endogenous neuroprotective mechanisms.[7-14]

Prior genome wide association studies (GWAS), mostly in populations of European descent, have identified numerous loci associated with stroke risk. In 2018, the MEGASTROKE consortia performed one of the largest GWAS to date, combining most of the available GWAS for stroke risk in a unique multi-ancestry meta-analysis including 67,162 cases and 454,450 controls. This analysis led to the discovery of 22 novel loci, bringing the total stroke risk loci to 32. Many loci were previously linked to other vascular traits (blood pressure, cardiac phenotypes, venous thromboembolism); while others had no obvious connection with stroke, warranting further investigation to identify potentially novel mechanisms.[15] A similar approach, used to decipher the genetics of long term disability after ischemic stroke in 6,165 non-Hispanic Whites, identified one locus that was been not replicated so far.[16,17] However, to date there have been no genetic studies examining early neurological change after ischemic stroke.

To our knowledge, the Genetics of Early Neurological InStability after Ischemic Stroke (GENISIS) is the largest well-characterized study for early outcomes quantified by ΔNIHSS.[18] To increase the power to detect genetic associations, our study recruited patients from multiple diverse ancestry groups. We leveraged the GENISIS cohort using ΔNIHSS as a quantitative phenotype, to identify novel variants, genes and pathways associated with early neurological instability after ischemic stroke.

Materials and Methods

Study Design

A detailed description of the acute ischemic stroke patients recruited from 21 sites from seven countries throughout the world, has been published elsewhere.[6] Briefly, adult acute ischemic stroke patient with measurable deficit on the NIHSS that presented within 6 hours of stroke onset (or last known normal) were enrolled in the study after obtaining informed consent, including patients treated with tPA. Patients who underwent a thrombectomy, were enrolled in other treatment trials, or for whom consent and/or a blood sample could not be obtained were excluded. Demographics, co-morbidities, acute treatment variables, imaging data and TOAST classification were collected.

To accommodate the difference in the genetic architecture intrinsic to the country of origin, we performed a three-stage analysis (Figure 1A). First, we used an additive model to perform a GWAS in each country individually, except for the United States, where the population was stratified into European and African ancestry cohorts. We then performed a fixed effects meta-analyses within the same ethnic cohorts. Finally, we used a multi-ancestry Bayesian meta-analysis to collapse all the ethnic backgrounds. Unlike a fixed effect meta-analysis, the Bayesian approach is able to account for population structure differences.[19] Genetic loci that passed multiple test correction, a threshold set at Log Bayes Factor (LBF) > 5, were annotated using bioinformatics tools to identify the gene driving the genetic signal (Figure 1B). We used functional annotation, multi-tissue expression quantitative trait loci (eQTL) data, and summary-data-based Mendelian randomization (SMR) to map the genome-wide to specific genes. Single nuclei-RNA-seq data derived from cortex samples was used to determine potential correlation between the transcripts of the identified genes and determine in which brain cell types the genes are expressed.[20]

The study was approved by the Institutional Review Boards at every participating site. Written informed consent was obtained from all participants or their family members. All research was performed according to the approved protocols and consents.

Genotyping

All participants were genotyped using Illumina SNP array technology. Samples were genotyped in seven batches during the GENISIS recruitment (see Supplementary Methods). Genotyping quality control and imputation were

performed separately for each genotyping round using SHAPEIT[21] and IMPUTE2.[22] For each genotyping batch, SNPs with a call rate lower than 98% and autosomal SNPs that were not in Hardy-Weinberg equilibrium ($P<1\times10^{-06}$) were removed from the dataset. The X chromosome SNPs were used to determine sex based on heterozygosity rates, and samples with discordant inferred sex and reported sex were removed. Only samples with call rate greater than 98% were considered to pass quality control. Finally, the genotype batches were merged in a single file to perform the analyses.

Additional QC was performed in the merged dataset. We tested pairwise genome-wide estimates of proportion identity-by-descent, the presence of unexpected duplicates, and cryptically relatedness (PI-HAT>0.30). Of the pairs of these samples flagged, the sample with higher genotyping rate was kept for downstream analysis. Principal component analysis (PCA) was performed using HapMap as an anchor to remove ethnic outliers and keep the populations as homogeneous as possible for each of the participant countries. Principal components were also used to cluster and identify ancestry populations for US participants with European descent (EuA) and African-American descent (AfA). Samples outside two standard deviations from the center of the Non-Hispanic White or the Asian cluster were considered outliers for Spain, Finland and Poland. We confirmed the ethnicity of the AfA and Hispanic populations, however, due to the genetic heterogeneity present in these populations we did not remove the samples outside two standard deviations from the mean.

Analysis of variance

We used genome-wide complex traits analysis (GCTA) to determine the heritability of ΔNIHSS.[23] GCTA estimates the amount of phenotypic variance in a given complex trait explained by all the SNPs and fits the effects of these SNPs as random effects in a linear mixed model. Because it relies on a large, homogeneous populations for accurate results, we only included the individuals with non-Hispanic White ancestry.

Single Variant Analyses

To mitigate the effects of genetic heterogeneity due to the diverse ancestry of participants enrolled in the GENISIS study, we used a multi-step study design (Figure 1A). First, we performed single variant analyses each participant country separately. We tested the association of SNPs across the genome with ΔNIHSS using an additive linear model with PLINK 1.9.[24] Sex, age, and the two Principal Components calculated for each population were included in the model. Additional covariates include the SNP genotyping batch, TOAST classification (using dummy variables to incorporate all subtypes), and baseline NIHSS to adjust for stroke severity. Although baseline NIHSS was used to calculate ΔNIHSS, it does not fully explain the observed variance in ΔNIHSS; further, there is no multicollinearity between these two variables, permitting their inclusion in the model.[25] Fifty four percent of study participants were treated with tPA, however treatment was not included as covariate. While tPA influences ΔNIHSS[6], it has little impact on the genetic architecture of ΔNIHSS. The inclusion of tPA as covariate lead to highly correlated results when compared to a model without tPA (r^2=0.96, Supplementary Figure 1). Second, we meta-analyzed the populations with similar ethnic backgrounds using with fixed effect meta-analyses using METAL.[26] We performed two meta-analyses, one for the non-Hispanic Whites (Spain, Finland, Poland, and United States EuA) and one for the Hispanics (Costa Rica and Mexico). Finally, we analyzed the four available ethnicities non-Hispanic Whites (meta-analysis), Hispanics (meta-analysis), Asians (Korea) and African Americans (United States AfA) using MANTRA, a Bayesian-based multi-ancestry meta-analyses.[19] Log Bayes Factor (LBF) greater than 5 was considered to be genome wide significant after multiple test correction.

Functional Annotation

We annotated all the variants with suggestive associations (LBF>4) with ANNOVAR[27] and SnpEff[28] to identify the nearest gene and to determine if any variant is predicted to change protein sequence (non-synonymous variants) or could affect expression. We also confirmed if any of the SNPs were possible regulatory elements or DNA features using RegulomeDB.[29]

DEPICT[30] and FUMA[31] were used to perform gene ontology and pathways analyses. We also leveraged brain single nuclei RNA expression data (http://ngi.pub/snuclRNA-seq/)[20], to determine if the gene expression of the genes located in each identified loci was expressed in brain. For the ones expressed in brain, we also investigated if they were expressed in any specific brain cells (Figure 1B). Finally, we accessed blood RNA expression data taken at different times after stroke onset (3h, 5h and 24h) from the CLEAR trial[32] (NCT00250991 at www.Clinical-Trials.gov) to test if the expression of genes located in the identified loci were associated with NIHSS or ΔNIHSS ($NIHSS_{5h}$-$NIHSS_{24h}$). We extracted the correlation between ΔNIHSS and gene expression (meausered using Affymetrix U133 Plus 2.0 array).[33]

eQTL mapping, Mendelian Randomization and Colocalization

To identify the most likely functional gene, we accessed available expression quantitative trait (eQTL) datasets: the Genotype-Tissue Expression (GTEx) Project V8 (accessed on 10/10/2019), the Brain eQTL Almanac (Braineac) and an in-house dataset that includes brain expression data for 613 brains[34]. We used the summary-data-based Mendelian Randomization (SMR)[35] and colocalization[36] to test for pleiotropic association between the expression level of a gene and a complex trait to evaluate if the effect size of a genetic variant on the phenotype is mediated by gene expression (Figure 1B). We tested GWAS-significant and -suggestive loci from the ΔNIHSS analysis in two datasets: selected GTEx tissues (brain anterior cortex, cerebellum, brain cerebellar hemisphere, substantia nigra, hippocampus, frontal cortex, and putamen) and the Westra *et al*. dataset[37] derived from whole blood. Both SMR and colocalization require effect sizes and the respective standard error to test the causal relationship, but MANTRA does not provide effect sizes. As a consequence, we used the summary statistics from the joint analysis for all populations to perform these analyses that are correlated with the results from MANTRA ($r=-0.57$; $p<1.07\times10^{-05}$ – data not shown). To complement the Mendelian randomization analyses with the posterior probability of a variant being causal in both GWAS and eQTL studies accounting for the genetic heterogeneity and *linkage disequilibrium* (LD), we used eCAVIAR[38] which will consider several variants within the GWAS significant loci to perform the test.

Genetic Correlation

We examined similarities in the genetic architectures of stroke early outcomes (ΔNIHSS) and stroke risk[15] using PRSice[39], LDSC[40] and GNOVA[41] (Figure 1B). Briefly, PRSice calculates polygenic risk scores at different p value thresholds by weighting each SNP by their effect size estimates. SNPs present in one dataset, ambiguous SNPs (A/T or C/G) and all SNPs in LD are removed prior to polygenic risk score calculation. LDSC and GNOVA estimate the genetic covariance and the variant-based heritability for two sets of summary statistics, each one corresponding to one trait of interest. These two parameters are used to calculate the genetic correlation and covariance respectively between the two traits. We limited our comparisons to the non-Hispanic White population to keep the population genetically homogeneous and use the 1000 Genomes European population-derived reference dataset. We calculated the genetic correlation between the European ischemic stroke summary statistics of the MEGASTROKE[15] study and the non-Hispanic Whites meta-analysis summary statistics from the GENISIS study. We also determined if traits related to cardiovascular and general health (age at death[42], lipid levels[43] and body mass index (BMI)[44]) are genetically correlated to ΔNIHSS.

Summary statistics of the GENISIS dataset used for these analyses are available upon request. Individual data for the full GENISIS dataset will be uploaded to dbGAP titled: "*Genetics of Early Neurological Instability After Ischemic Stroke (GENISIS)"*.

Results

The GENISIS study recruited 5,876 acute ischemic stroke patients from seven countries (Spain, Finland, Poland, United States, Costa Rica, Mexico and Korea). The mean patient age was 73 years; 45% of the patients were females; and 54% were treated with tPA. No significant differences in age or sex were found across sites. The distribution of TOAST classification of stroke etiology was also similar across sites. Significant differences were observed in baseline NIHSS and tPA treatment rates, likely due to differences in practices across the sites (Table 1).[6] ΔNIHSS approximated a normal distribution, similar to that of each of the ethnic groups (non-Hispanic whites, Hispanics, African descent, and Asians) (Supplementary Figure 2).

Identification of novel loci associated with stroke early outcomes

We performed single variant analyses for each individual cohort separately; then we combined cohorts with similar ethnic backgrounds; finally, we performed a multi-ancestry meta-analysis with the four ethnic groups available in this study (Non-Hispanic Whites, Hispanics, Asians and African Americans) (Figure 1A). We identified seven GWAS significant loci (Figure 2A and Table 2) associated with ΔNIHSS.

Three independent loci were identified in chr2. The first locus, tagged by rs58763243 ($MAF_G=0.07$; LBF=6.51), was located in a region comprised by several long non-coding RNAs and microRNAs (Supplementary Figure 3A). For this locus, all of the populations contributed to the association with negative betas, indicating that the minor allele was associated with lower (or more negative) ΔNIHSS. In addition this locus reached genome-wide significance in the US AfA population and was nominally significant in the Finnish population (Supplementary Figure 3B).

The second locus, rs13403787 (MAF_A=0.16; LBF=5.57), was also located on chr2 in a region with more than 20 genes between two recombination sites (Supplementary Figure 3C). The minor allele was associated with higher (or more positive) ΔNIHSS in all cohorts (Supplementary Figure 3D). The last genome wide significant locus in chr2 was rs72958644 (MAF_T=0.04; LBF=6.34), located in a region that includes *ADAM23*, *CREB1*, *DYTN*, *NRP2*, *MDH1B*, among many others (Figure 2C). The signal is driven by the Non-Hispanic Whites (meta-analysis $p=3.44\times10^{-08}$), but virtually all ethnic groups contributed to this association, as the directionality was consistent across Hispanic, non-Hispanic White and AfA ethnic groups (Figure 2D and Supplementary Table 1). However, the SNPs in this locus were monomorphic in the Asian population. We observed a trend for positive correlation between ΔNIHSS with the genotype in this locus (R^2=0.005, p=0.04; Supplementary Table 2).

Four additional loci were identified outside chr2. One locus at 4q34.3, rs12641856 (MAF_T=0.05; LBF=5.50), in a region that includes four genes (*LINC00290*, *MGC45800*, *MIR1205* and *TENM3*; Supplementary Figure 3E). The signal was driven mainly by the Korean cohort ($p=1.38\times10^{-07}$) (Supplementary Figure 3F). The locus identified on chr5 is located on a region containing nine genes (*LOC101927134*, *GRIA1*, *FAM114A2*, *SAP30L*, *SAP30L-AS1*, *MFAP3*, *GALNT10*, *HAND1* and *MIR3141;* Figure 2F). The minor allele for the top hit in this locus, rs114248865 (MAF_T=0.05; LBF=5.29) was associated with greater (more positive) ΔNIHSS across all cohorts, and was significant in the Spanish ($p=8.41\times10^{-07}$) and Finnish cohorts (p=0.03; Figure 2G). Another locus on chr6, tagged by rs6930598 (MAF_T=0.06; LBF=5.30), was located in a region with the genes *PARK2*, *PACRG*, *MAP3K4*, *AGPAT4*, *AGPAT4-IT1* and *LOC101929239* (Supplementary Figure 3G). The variants in the region were significant in the Mexican and Korean cohorts but the direction of effect was not consistent (Supplementary Figure 3H). Moreover, the MAF for these variants ranged between 16% in the AfA population to 6% in the Asian and Finnish populations, suggesting that the region is very polymorphic depending on ethnicity. Thus, even though the locus is important for ΔNIHSS, it is possible that it is not the causal variant. Finally, we identified a locus on chr7, tagged by the variant rs10807797 (MAF_G=0.34; LBF=5.70), and located in a gene rich region with 15 genes, including *TWISTNB*, *MACC1*, *TMEM196*, *ABCB5*, *RPL23P8* (Supplementary Figure 3I). This locus is tightly encompassed by two recombination sites. The top signal was significant or suggestive in all populations except the Polish and Mexican cohorts. Consistently, the direction of effect was the same in all cohorts except the Mexican cohort (Supplementary Figure 3J, Supplementary Table 1).

Genetic contribution to early outcomes after ischemic stroke

We used GCTA to quantify the phenotypic variance explained by common SNPs. Because GCTA exploits LD patterns to calculate the explained variance, we restricted our analysis to non-Hispanic Whites. Due to founder effects present in the Finnish population, we also removed this cohort from the variance calculation (Final N=4,573). GTCA revealed that common genetic variants explained 8.7% of the variance of ΔNIHSS (p=0.001), confirming that genetic variants and genes are implicated on stroke outcomes. Next we determine what proportion of the genetic component is explained by the GWAS signals.

The SNPs comprised within the 7 genome-wide significant loci, defined as 500 bp upstream or downstream of the top signal, explained 2.1% of the total variance ($p=1.80\times10^{-06}$) of ΔNIHSS, or just 24.1% of the genetic component of ΔNIHSS. This suggests that there are additional loci associated with ΔNIHSS yet to be discovered. Thus, studies with larger sample size and more statistical power are needed to identify these additional loci.

Functional Annotation of the Genome-Wide Significant Loci

Identifying the likely causal gene from each loci driving the association is a multi-step process (Figure 1B). We first annotated the suggestive variants (LBF>4), but none of them were predicted to change the protein sequence, comprise a regulatory element, or affect the chromatin architecture. Next, we explored publicly available datasets to investigate if any of the SNPs with suggestive LBFs were eQTLs (Supplementary Table 3). We performed gene-based analyses (Supplementary Table 4) and Mendelian Randomization (MR) analyses to identify possible causal relationships between gene expression and ΔNIHSS (Supplementary Table 5). Summary results can be found in Figure 3.

Gene-based analyses using FUMA suggested that *DYTN* ($p=2.86\times10^{-05}$, Z=4.02) and *ADAM23* ($p=1.59\times10^{-04}$, Z=3.60) were the genes driving the association at 2q33.3. Several variants in the *ADAM23* region were strong eQTLs for this gene in multiple tissues, based on the GTEx data (esophagus mucosa: $p=1.90\times10^{-06}$; Cultured Fibroblasts: $p=3.70\times10^{-05}$). Several SNPs in this region were also eQTLs for *NDFUS1* (thyroid $p=1.20\times10^{-04}$) although with less significant p-values than those for *ADAM23*. The variant was also significant in an independent dataset from the Braineac eQTL study, where rs13422013 (LFB=4.05) was also associated with *ADAM23* gene expression (thalamus $p=7.60\times10^{-04}$) and *NRP2* (cerebellum $p=4.40\times10^{-04}$). MR analyses indicated that *ADAM23* (p=0.05) and not *NRP2* (p>0.05) was the

gene driving the association in this locus. Human brain single nuclei RNA-seq data indicate that *ADAM23* expression is enriched in neurons ($p<2.20\times10^{-16}$); compared to all the other brain cell types. More specifically, its expression is enriched in excitatory neurons (Figure 2E).

Gene-based analyses using FUMA revealed that *GRIA1* located in 5q33.2 was the gene most likely driving the association in that region (p=0.03, Z=1.83). However, Braineac identified several eQTLs for *GALNT10* ($p=3.60\times10^{-04}$) in the occipital cortex, but *GALNT10* did not reach statistical significance in the gene-based analysis. GTEx portal and the protein atlas reveals that *GRIA1* is mainly expressed in brain tissue. While *GALNT10* is also expressed in the brain, it has higher expression in other tissues. The human brain single nuclei RNA-seq data confirmed that both *GRIA1* and *GALNT10* are expressed in divergent brain cell types (Figures 2H and Supplementary 4A). *GRIA1* is highly-expressed in neurons compared to other cell types ($p<2.20\times10^{-16}$), but not expressed in oligodendrocytes ($p<2.20\times10^{-16}$) or astrocytes ($p<2.20\times10^{-16}$). In contrast, *GALNT10* is expressed in microglia, oligodendrocytes and astrocytes, but expression in neurons is low ($p<2.20\times10^{-16}$). *GRIA1* expression was also nominally associated with ΔNIHSS in the CLEAR trial dataset (p=0.002, r2=0.22).

Of the remaining five loci, we were able to map four (Supplementary Results). Briefly, eQTL analysis, revealed that 2q31.2 was likely to be driven by *DFNB59*, and 4q34.3 by *MGC45800*. No eQTLs were identified in 6q26, however, the locus falls within the boundaries of *PARK2*. Finally, 7p21.1 contains several eQTLs for *TWISTNB* and *ABCB5* (Supplementary Results).

Pathway analyses

Gene ontology and pathway analyses using DEPICT and summary statistics for ΔNIHSS revealed consistent suggestive associations with functions relating to the brain and central nervous system. The top tissue enrichment from DEPICT identified the central nervous system ($p=4.9\times10^{-03}$), including the brain ($p=4.9\times10^{-03}$) and some brain regions: occipital lobe ($p=1.52\times10^{-03}$), parietal lobe ($p=5.31\times10^{-03}$), and hippocampus ($p=5.07\times10^{-03}$; Supplementary Table 6). The most significant pathways in the gene-set enrichment were the AKAP5 protein-protein interaction subnetwork ($p=5.11\times10^{-04}$), the CAMK1D protein-protein interaction subnetwork ($p=5.27\times10^{-04}$), the RALA protein-protein interaction subnetwork ($p=6.32\times10^{-04}$) and neurotransmitter uptake ($p=6.74\times10^{-04}$; Supplementary Table 7). Several genome-wide significant candidate genes fell within these networks, including *AGPS* (2q31.2, LBF=5.57) in the RALA subnetwork and *GPRS1* (2q33.3, LBF=6.34) in the CAMK1D subnetwork. AKAP5, expressed almost uniquely in brain (GTEx portal and the protein atlas.), interacts with post-synaptic density protein 95 (PSD-95) in dendritic spines and modulates synaptic transmission.[45-47] Both CAMK1D and RALA, expressed in most tissues (CAMK1D demonstrates higher expression in brain), are involved in signal transduction functions. MAGMA gene-set analyses revealed the gene-set "*meissner_brain_hcp_with_h3k4me3_and_h3k27me3*" ($p=9.41\times10^{-06}$, Supplementary Table 8). Unfortunately, there is no function associated with this gene set that has been described in animal models. Together these results indicate that our unbiased genetic screening has identified genes that are largely expressed in the brain and in some cases involved in pathways associated with neuronal and synaptic function.

Unique genetic architecture of early outcomes after stroke

We examined the genetic architecture of ΔNIHSS for shared genetic variation and variance with other cardiovascular and aging-related traits, including stroke risk, age at death, plasma lipid levels and body mass index using PRSice (Supplementary Table 9 and Figure 5), LDSC (Supplementary Table 10) and GNOVA (Supplementary Table 11). Supported by a previous report[18], PRSice did not reveal any genetic overlap between the genetic architecture of ΔNIHSS and stroke risk. Similarly, no overlap with age at death, lipid levels, or BMI was found. LDSC was unable to calculate the heritability estimate for ΔNIHSS. GNOVA, was successful at estimating the heritability for ΔNIHSS, and revealed significant covariance with stroke risk, but only when correcting for sample overlap ($p=4.67\times10^{-05}$, correlation=0.80) and age at death ($p=9.26\times10^{-04}$, correlation=1). However, the heritability estimate for ΔNIHSS for overlap with lipid levels and body mass index was negative, likely due to the low number of variants included in the analyses. Because both GNOVA and LDSC require larger sample sizes, the results of these analyses were inconclusive.

Discussion

The first 24 hours after stroke onset is a period of great neurological instability, which may reflect brain tissue at risk for infarction but with the potential for salvageability.[4,48-50] Not only is early neurological change common, it is influenced by known mechanisms involved in early deterioration/improvement, and has a strong influence on long-term functional outcome.[6] Here, we performed a GWAS using ΔNIHSS as a quantitative phenotype in 5,876 acute

ischemic stroke patients. We found that ΔNIHSS is heritable: common SNPs account for 8.7% of its variance. We have found seven genome-wide significant loci that are related to ΔNIHSS. However, they explain only 2.1% of the variance, indicating that 6.6% of the variance is explained by genes below the genome-wide significant threshold. Through functional annotation, we have linked each locus to specific genes, some of which are uniquely expressed in the brain.

Of all the loci showing association with ΔNIHSS, functional annotation analyses strongly suggest functional genes for two of the seven loci: *ADAM23* for 2q33.3 and *GRIA1* for 5q33.2. ADAM23 belongs to the ADAM (a disintegrin and metalloproteinase) family of proteins, defined by a single-pass transmembrane structure with a metallopeptidase domain (some inactive). This protein family is involved in cell adhesion, migration, proteolysis and signaling.[51] ADAM23 is a transmembrane member without catalytic domain, and is involved in cell-cell and cell-matrix interactions.[51,52] Previous studies have shown that *ADAM23* is expressed in pre-synaptic membranes, linked by the extracellular protein LGI1 to post-synaptic ADAM22.[53,54] We found that *ADAM23* was expressed primarily in excitatory neurons of the cerebral cortex, based on our human brain single-nuclei transcriptomics dataset[20], and confirmed by the Human Transcriptomic Cell Types dataset from the Allen Brain Map.[55] Several lines of evidence suggest that ADAM23 is important for pathological synaptic excitability: 1) *adam23* is a common risk gene for canine idiopathic epilepsy;[56-58] 2) mutations in its binding partner, *LGI1*, cause the neurological syndrome, ADPEAF (autosomal dominant partial epilepsy with auditory features)[59]; 3) autoimmunity against LGI1 (as seen in limbic encephalitis) results in seizures and encephalopathy.[60]

Indeed, ADAM23 is also known to be a binding partner (via ADAM22 and PSD95) of the protein product of another one of our genome-wide significant associated genes, *GRIA1*, which encodes for the α-amino-3-hydroxy-5-methyl-4-isoxazolepropionic acid receptor subunit 1 (AMPAR1).[54] It has long been known that AMPA receptors, along with other glutamate receptors, are mediators of excitotoxic neuronal death, hypothesized to play an important role in ischemic brain injury.[61,62] The failure of numerous older clinical trials examining the efficacy of anti-excitotoxic drugs has cast doubt on the relevance of excitotoxicity in human acute ischemic stroke, although questions about the quality of these early clinical trials have been raised.[63-66] Thus, the association between *ADAM23* and *GRIA1* with ΔNIHSS provides the first genetic evidence that excitotoxicity may contribute to ischemic brain injury in humans.

The plausible roles that *ADAM23* and *GRIA1* play in acute brain ischemia mechanisms lend support to the idea that GWAS using ΔNIHSS as a quantitative phenotype can identify novel mechanisms and potential drug targets to mitigate neurological deterioration or enhance early improvement after stroke. From the CLEAR dataset, [32,33] expression levels of *GRIA1* in peripheral blood of ischemic stroke patients was associated with ΔNIHSS between 5h and 24h post stroke onset, supporting a link between increased expression of GRIA1 and improved outcomes. In addition to the two genes discussed above, our GWAS identified five other loci—the functional genes remain to be identified. Acute ischemic stroke patients are extremely well-phenotyped, as part of standard of care, with both clinical assessments and structural/physiological imaging. Thus, there is great potential for additional quantitative phenotypes to expand understanding of the genetic architecture of acute ischemic stroke, promising to identify novel mechanisms and drug targets. Larger and more comprehensive genetic studies of acute ischemic stroke are needed.

There are several limitations to this study. GENISIS enrolled a heterogeneous group of stroke patients without regard to underlying etiology, stroke localization and genetic and environmental background. Although we have previously demonstrated that etiology (TOAST criteria) has little influence on ΔNIHSS, it is likely that mechanisms involved in neurological instability may depend on etiology. Stroke localization may also be an important determinant of mechanisms involved in neurological instability. For example, mechanisms in cortical strokes may differ from those in subcortical or brainstem strokes. False positive findings due to the characteristics of the population is possible, but by using MANTRA we were able to correct by population heterogeneity. Future studies might aim to enroll a more homogeneous cohort of stroke patients to increase power to discover more genetic variants that associate with neurological instability. Finally, most of the patients in GENISIS were enrolled prior to the thrombectomy treatment era, and patients that underwent thrombectomy were excluded from the study to reduce heterogeneity. As a result, genetic interactions with reperfusion are largely unexplored.

Summary

In conclusion, we identified seven novel loci associated with early neurological instability after ischemic stroke. Two of these loci were linked to genes, *ADAM23* and *GRIA1*, involved in synaptic excitability at glutamatergic synapses.

These results provide some of the first evidence that excitotoxic mechanisms are relevant to acute ischemic stroke in humans. Moreover, these results provide proof of principle that GWAS using ΔNIHSS, a quantitative phenotype of early neurological instability, may reveal underlying mechanisms involved in ischemic brain injury.

Acknowledgments

We would like to thank the patients and their families for making possible all the genetic studies included in this manuscript. We also thank the MEGASTROKE consortium for access to the data (see full list of MEGASTROKE authors in supplementary data), the Genotype-Tissue Expression (GTEx) Project (supported by the Common Fund of the Office of the Director of the National Institutes of Health, and by NCI, NHGRI, NHLBI, NIDA, NIMH, and NINDS) and the Brain eQTL Almanac (Braineac) resource to access the UK Brain Expression Consortium (UKBEC) dataset.

Sources of Funding

Emergency Medicine Foundation Career Development Grant; AHA Mentored Clinical & Population Research Award (14CRP18860027); NIH/NINDS-R01-NS085419 (CC, JML); NIH/NINDS-R37-NS107230, NIH/NINDS U24-NS107230 (JML); NIH/NINDS-K23-NS099487 (LH); Barnes-Jewish Hospital Foundation (JML); Biogen (CC, JML); Helsinki University Central Hospital; Finnish Medical Foundation; Finland government subsidiary funds; Spanish Ministry of Science and Innovation; Instituto de Salud Carlos III (grants "Registro BASICMAR" Funding for Research in Health (PI051737), "GWALA project" from Fondos de Investigación Sanitaria ISC III (PI10/02064, PI12/01238 and PI15/00451), JR18/00004); Fondos FEDER/EDRF Red de Investigación Cardiovascular (RD12/0042/0020); Fundació la Marató TV3; Genestroke Consortium (76/C/2011); Recercaixa'13 (JJ086116). Tomás Sobrino (CPII17/00027), Francisco Campos (CPII19/00020) and Israel Fernandez are supported by Miguel Servet II Program from Instituto de Salud Carlos III and Fondos FEDER. Israel Fernandez is also supported by Maestro project (PI18/01338) and Pre-test project (PMP15/00022) from Instituto de Salud Carlos III and Fondos Feder, Agaur; and Epigenesis project from Marató TV3 Foundation. José Castillo, Joan Montaner, Antonio Dávalos, Joan Martí-Fábregas, Juan Arenillas and Israel Fernández,are supported by Invictus plus Network (RD16/0019) from Instituto de Salud Carlos III and Fondos Feder. Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP-2013/07559-3) (ILC), Sigrid Juselius Foundation. Sigrid Juselius Foundation. The MEGASTROKE project received funding from sources specified at http://www.megastroke.org/acknowledgments.html.Boryana Stamova, Bradley Ander and Frank Sharp are supported by NIH awards: NS097000, NS101718, NS075035, NS079153 and NS106950.

Disclosures

CC receives research support from: Biogen, EISAI, Alector and Parabon, and is a member of the advisory board of ADx Healthcare and Vivid Genomics. JML receives research support from Biogen, and is a consultant for Regenera. EAT and LCB are employed by Biogen. JFA has received speaker or consultant honoraria from Bayer, Boehringer Ingelheim, Pfizer-BMS, Daiichi Sankyo, Amgen and Medtronic. TT receives or has received research support from Bayer, Boehringer Ingelheim and Bristol Myers Squibb; he is a member of advisory boards for Bayer, Boehringer Ingelheim, Bristol Myers Squibb and Portola Pharmaceuticals; and he is granted international patents: new therapeutic uses (method to prevent brain edema and reperfusion injury), and thrombolytic compositions (method to prevent post-thrombolytic hemorrhage formation). The funders of the study had no role in the collection, analysis, or interpretation of data; in the writing of the report; or in the decision to submit the paper for publication.

References

1. Organization WH. The Global Burden of Disease: 2004 update. 2008.
2. Benjamin EJ, Virani SS, Callaway CW, et al. Heart Disease and Stroke Statistics-2018 Update: A Report From the American Heart Association. *Circulation* 2018; **137**(12): e67-e492.
3. Farooq MU, Chaudhry AH, Amin K, Majid A. The WHO STEPwise Approach to Stroke Surveillance. *Journal of the College of Physicians and Surgeons--Pakistan : JCPSP* 2008; **18**(10): 665.
4. Saver JL, Altman H. Relationship between neurologic deficit severity and final functional outcome shifts and strengthens during first hours after onset. *Stroke* 2012; **43**(6): 1537-41.
5. Tissue plasminogen activator for acute ischemic stroke. The National Institute of Neurological Disorders and Stroke rt-PA Stroke Study Group. *N Engl J Med* 1995; **333**(24): 1581-7.
6. Laura Heitsch LI, Caty Carrera, Michael Binkley, Daniel Strbian, Turgut Tatlisumak, Alejandro Bustamante, Marc Ribo, Carlos Molina, Antoni Davalos, Elena Lopez-Cancio, Luci¬a Muñoz-Narbona, Carolina Soriano-Tárraga, Eva Giralt Steinhauer, Victor Obach, Agnieszka Slowik, Joanna Pera, Katarzyna Lapicka-Bodzioch, Justyna Derbisz, Tomas Sobrino, Jose Castillo, Francisco Campos, Emilio Rodríguez-Castro, Susana Arias, Tomas Segura, Gemma Serrano-Heras, Cristofol Vives-Bauza, Rosa Diaz-Navarro, Silvia Tur, Carmen Jiménez, Joan Martí-Fàbregas, Raquel Delgado-Mederos, Juan Arenillas, Jerzy Krupinski, Natalia Cullell, Nuria Torres-Aguila, Elena Muiño, Jara Carcel-Marquez, Francisco Moniche, Juan Cabezas, Andria Ford, Rajat Dhar, Jaume Roquer, Pooja Khatri, Jordi Jiménez-Conde, Israel Fernandez-Cadenas, Joan Montaner, Jonathan Rosand, Carlos Cruchaga, and Jin-Moo Lee. Early change in NIH stroke scale after ischemic stroke captures brain injury mechanisms and predicts 90-day outcome. *Stroke* 2020; **In Press**.
7. Barber PA, Davis SM, Infeld B, et al. Spontaneous reperfusion after ischemic stroke is associated with improved outcome. *Stroke* 1998; **29**(12): 2522-8.
8. Jorgensen HS, Sperling B, Nakayama H, Raaschou HO, Olsen TS. Spontaneous reperfusion of cerebral infarcts in patients with acute stroke. Incidence, time course, and clinical outcome in the Copenhagen Stroke Study. *Archives of neurology* 1994; **51**(9): 865-73.
9. Bang OY, Saver JL, Buck BH, et al. Impact of collateral flow on tissue fate in acute ischaemic stroke. *Journal of neurology, neurosurgery, and psychiatry* 2008; **79**(6): 625-9.
10. Christoforidis GA, Karakasis C, Mohammad Y, Caragine LP, Yang M, Slivka AP. Predictors of hemorrhage following intra-arterial thrombolysis for acute ischemic stroke: the role of pial collateral formation. *AJNR American journal of neuroradiology* 2009; **30**(1): 165-70.
11. Ali LK, Saver JL. The ischemic stroke patient who worsens: new assessment and management approaches. *Reviews in neurological diseases* 2007; **4**(2): 85-91.
12. Dreier JP. The role of spreading depression, spreading depolarization and spreading ischemia in neurological disease. *Nat Med* 2011; **17**(4): 439-47.
13. Paciaroni M, Agnelli G, Corea F, et al. Early hemorrhagic transformation of brain infarction: rate, predictive factors, and influence on clinical outcome: results of a prospective multicenter study. *Stroke* 2008; **39**(8): 2249-56.
14. Torres-Aguila NP, Carrera C, Giese AK, et al. Genome-Wide Association Study of White Blood Cell Counts in Patients With Ischemic Stroke. *Stroke* 2019; **50**(12): 3618-21.
15. Malik R, Chauhan G, Traylor M, et al. Multiancestry genome-wide association study of 520,000 subjects identifies 32 loci associated with stroke and stroke subtypes. *Nat Genet* 2018; **50**(4): 524-37.
16. Soderholm M, Pedersen A, Lorentzen E, et al. Genome-wide association meta-analysis of functional outcome after ischemic stroke. *Neurology* 2019; **92**(12): e1271-e83.
17. Mola-Caminal M, Carrera C, Soriano-Tarraga C, et al. PATJ Low Frequency Variants Are Associated With Worse Ischemic Stroke Functional Outcome. *Circ Res* 2019; **124**(1): 114-20.
18. Ibanez L, Heitsch L, Dube U, et al. Overlap in the Genetic Architecture of Stroke Risk, Early Neurological Changes, and Cardiovascular Risk Factors. *Stroke* 2019; **50**(6): 1339-45.
19. Morris AP. Transethnic meta-analysis of genomewide association studies. *Genetic epidemiology* 2011; **35**(8): 809-22.
20. Del-Aguila JL, Li Z, Dube U, et al. A single-nuclei RNA sequencing study of Mendelian and sporadic AD in the human brain. *Alzheimer's research & therapy* 2019; **11**(1): 71.
21. Delaneau O, Coulonges C, Zagury JF. Shape-IT: new rapid and accurate algorithm for haplotype inference. *BMC bioinformatics* 2008; **9**: 540.
22. Howie B, Marchini J, Stephens M. Genotype imputation with thousands of genomes. *G3* 2011; **1**(6): 457-70.
23. Yang J, Lee SH, Goddard ME, Visscher PM. GCTA: a tool for genome-wide complex trait analysis. *Am J Hum Genet* 2011; **88**(1): 76-82.

24. Chang CC, Chow CC, Tellier LC, Vattikuti S, Purcell SM, Lee JJ. Second-generation PLINK: rising to the challenge of larger and richer datasets. *GigaScience* 2015; **4**: 7.
25. Glymour MM, Weuve J, Berkman LF, Kawachi I, Robins JM. When is baseline adjustment useful in analyses of change? An example with education and cognitive change. *Am J Epidemiol* 2005; **162**(3): 267-78.
26. Willer CJ, Li Y, Abecasis GR. METAL: fast and efficient meta-analysis of genomewide association scans. *Bioinformatics* 2010; **26**(17): 2190-1.
27. Yang H, Wang K. Genomic variant annotation and prioritization with ANNOVAR and wANNOVAR. *Nature protocols* 2015; **10**(10): 1556-66.
28. Cingolani P, Platts A, Wang le L, et al. A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of Drosophila melanogaster strain w1118; iso-2; iso-3. *Fly* 2012; **6**(2): 80-92.
29. Boyle AP, Hong EL, Hariharan M, et al. Annotation of functional variation in personal genomes using RegulomeDB. *Genome research* 2012; **22**(9): 1790-7.
30. Pers TH, Karjalainen JM, Chan Y, et al. Biological interpretation of genome-wide association studies using predicted gene functions. *Nature communications* 2015; **6**: 5890.
31. Watanabe K, Taskesen E, van Bochoven A, Posthuma D. Functional mapping and annotation of genetic associations with FUMA. *Nature communications* 2017; **8**(1): 1826.
32. Pancioli AM, Broderick J, Brott T, et al. The combined approach to lysis utilizing eptifibatide and rt-PA in acute ischemic stroke: the CLEAR stroke trial. *Stroke* 2008; **39**(12): 3268-76.
33. Stamova B, Xu H, Jickling G, et al. Gene expression profiling of blood for the prediction of ischemic stroke. *Stroke* 2010; **41**(10): 2171-7.
34. Li Z, Del-Aguila JL, Dube U, et al. Genetic variants associated with Alzheimer's disease confer different cerebral cortex cell-type population structure. *Genome Med* 2018; **10**(1): 43.
35. Zhu Z, Zhang F, Hu H, et al. Integration of summary data from GWAS and eQTL studies predicts complex trait gene targets. *Nat Genet* 2016; **48**(5): 481-7.
36. Plagnol V, Smyth DJ, Todd JA, Clayton DG. Statistical independence of the colocalized association signals for type 1 diabetes and RPS26 gene expression on chromosome 12q13. *Biostatistics* 2009; **10**(2): 327-34.
37. Westra HJ, Peters MJ, Esko T, et al. Systematic identification of trans eQTLs as putative drivers of known disease associations. *Nat Genet* 2013; **45**(10): 1238-43.
38. Hormozdiari F, van de Bunt M, Segre AV, et al. Colocalization of GWAS and eQTL Signals Detects Target Genes. *Am J Hum Genet* 2016; **99**(6): 1245-60.
39. Euesden J, Lewis CM, O'Reilly PF. PRSice: Polygenic Risk Score software. *Bioinformatics* 2015; **31**(9): 1466-8.
40. Bulik-Sullivan BK, Loh PR, Finucane HK, et al. LD Score regression distinguishes confounding from polygenicity in genome-wide association studies. *Nat Genet* 2015; **47**(3): 291-5.
41. Lu Q, Li B, Ou D, et al. A Powerful Approach to Estimating Annotation-Stratified Genetic Covariance via GWAS Summary Statistics. *Am J Hum Genet* 2017; **101**(6): 939-64.
42. Pilling LC, Kuo CL, Sicinski K, et al. Human longevity: 25 genetic loci associated in 389,166 UK biobank participants. *Aging (Albany NY)* 2017; **9**(12): 2504-20.
43. Liu DJ, Peloso GM, Yu H, et al. Exome-wide association study of plasma lipids in >300,000 individuals. *Nat Genet* 2017; **49**(12): 1758-66.
44. Turcot V, Lu Y, Highland HM, et al. Protein-altering variants associated with body mass index implicate pathways that control energy intake and expenditure in obesity. *Nat Genet* 2018; **50**(1): 26-41.
45. Weisenhaus M, Allen ML, Yang L, et al. Mutations in AKAP5 disrupt dendritic signaling complexes and lead to electrophysiological and behavioral phenotypes in mice. *PLoS One* 2010; **5**(4): e10325.
46. Zhang M, Patriarchi T, Stein IS, et al. Adenylyl cyclase anchoring by a kinase anchor protein AKAP5 (AKAP79/150) is important for postsynaptic beta-adrenergic signaling. *J Biol Chem* 2013; **288**(24): 17918-31.
47. Li X, Nooh MM, Bahouth SW. Role of AKAP79/150 protein in beta1-adrenergic receptor trafficking and signaling in mammalian cells. *J Biol Chem* 2013; **288**(47): 33797-812.
48. Adams HP, Jr., Davis PH, Leira EC, et al. Baseline NIH Stroke Scale score strongly predicts outcome after stroke: A report of the Trial of Org 10172 in Acute Stroke Treatment (TOAST). *Neurology* 1999; **53**(1): 126-31.
49. Yeo LL, Paliwal P, Teoh HL, et al. Early and continuous neurologic improvements after intravenous thrombolysis are strong predictors of favorable long-term outcomes in acute ischemic stroke. *J Stroke Cerebrovasc Dis* 2013; **22**(8): e590-6.
50. Sajobi TT, Menon BK, Wang M, et al. Early Trajectory of Stroke Severity Predicts Long-Term Functional Outcomes in Ischemic Stroke Subjects: Results From the ESCAPE Trial (Endovascular Treatment for Small Core and

Anterior Circulation Proximal Occlusion With Emphasis on Minimizing CT to Recanalization Times). *Stroke; a journal of cerebral circulation* 2017; **48**(1): 105-10.
51. Goldsmith AP, Gossage SJ, ffrench-Constant C. ADAM23 is a cell-surface glycoprotein expressed by central nervous system neurons. *Journal of neuroscience research* 2004; **78**(5): 647-58.
52. Mitchell KJ, Pinson KI, Kelly OG, et al. Functional analysis of secreted and transmembrane proteins critical to mouse development. *Nat Genet* 2001; **28**(3): 241-9.
53. Sagane K, Ishihama Y, Sugimoto H. LGI1 and LGI4 bind to ADAM22, ADAM23 and ADAM11. *International journal of biological sciences* 2008; **4**(6): 387-96.
54. Lovero KL, Fukata Y, Granger AJ, Fukata M, Nicoll RA. The LGI1-ADAM22 protein complex directs synapse maturation through regulation of PSD-95 function. *Proc Natl Acad Sci U S A* 2015; **112**(30): E4129-37.
55. Hawrylycz MJ, Lein ES, Guillozet-Bongaarts AL, et al. An anatomically comprehensive atlas of the adult human brain transcriptome. *Nature* 2012; **489**(7416): 391-9.
56. Koskinen LL, Seppala EH, Belanger JM, et al. Identification of a common risk haplotype for canine idiopathic epilepsy in the ADAM23 gene. *BMC genomics* 2015; **16**: 465.
57. Koskinen LL, Seppala EH, Weissl J, et al. ADAM23 is a common risk gene for canine idiopathic epilepsy. *BMC genetics* 2017; **18**(1): 8.
58. Seppala EH, Koskinen LL, Gullov CH, et al. Identification of a novel idiopathic epilepsy locus in Belgian Shepherd dogs. *PLoS One* 2012; **7**(3): e33549.
59. Berkovic SF, Izzillo P, McMahon JM, et al. LGI1 mutations in temporal lobe epilepsies. *Neurology* 2004; **62**(7): 1115-9.
60. van Sonderen A, Schreurs MW, Wirtz PW, Sillevis Smitt PA, Titulaer MJ. From VGKC to LGI1 and Caspr2 encephalitis: The evolution of a disease entity over time. *Autoimmunity reviews* 2016; **15**(10): 970-4.
61. Lee JM, Zipfel GJ, Choi DW. The changing landscape of ischaemic brain injury mechanisms. *Nature* 1999; **399**(6738 Suppl): A7-14.
62. Palmer CL, Cotton L, Henley JM. The molecular pharmacology and cell biology of alpha-amino-3-hydroxy-5-methyl-4-isoxazolepropionic acid receptors. *Pharmacol Rev* 2005; **57**(2): 253-77.
63. Chamorro A, Dirnagl U, Urra X, Planas AM. Neuroprotection in acute stroke: targeting excitotoxicity, oxidative and nitrosative stress, and inflammation. *The Lancet Neurology* 2016; **15**(8): 869-81.
64. Wahlgren NG, Ahmed N. Neuroprotection in cerebral ischaemia: facts and fancies--the need for new approaches. *Cerebrovasc Dis* 2004; **17 Suppl 1**: 153-66.
65. Neuhaus AA, Couch Y, Hadley G, Buchan AM. Neuroprotection in stroke: the importance of collaboration and reproducibility. *Brain : a journal of neurology* 2017; **140**(8): 2079-92.
66. Castillo J, Davalos A, Noya M. Progression of ischaemic stroke and excitotoxic aminoacids. *Lancet* 1997; **349**(9045): 79-83.
67. Adams HP, Jr., Bendixen BH, Kappelle LJ, et al. Classification of subtype of acute ischemic stroke. Definitions for use in a multicenter clinical trial. TOAST. Trial of Org 10172 in Acute Stroke Treatment. *Stroke* 1993; **24**(1): 35-41.

Figure Legends

Figure 1. Study design. Summarized description of the multi-step approach used to account for the genetic heterogeneity intrinsic to the multi-ancestry nature of the GENISIS study (**A**). We performed single variant analysis in each of the participating countries separately. Then we meta-analyzes all the non-Hispanic whites (blue) and Hispanic (green) ethnicities. Finally we analyzed the non-Hispanic whites, Hispanics, Korea (orange) and US participants with African Descent (US AfA – yellow) using a Bayesian model. The variants with genome-wide significant or suggestive results were annotated using sequential steps to elucidate the gene driving the association (**B**). We performed gene-based and pathway analyses, we collected the information available in publicly available datasets and we performed Mendelian randomization. We also performed genetic architecture overlap tests to examine overlap with known genetic risk factors.

Figure 2. Association and annotation results. A. Manhattan plot shows Log Bayes factor (LBF) values from the multi-ancestry meta-analysis in each genomic location. The red line indicates the GWAS significant threshold (LBF>5) and the blue line the GWAS suggestive threshold (LBF>4). The genome-wide significant loci are highlighted. Local Manhattan plots are shown for rs72958644 (**C**) and rs114248865 (**F**) along with the corresponding forest plots (**D** and **G)**, showing the contribution of each population to the overall signal. As part of the functional gene mapping, we accessed an in-house single nuclei dataset (**B**) to describe the expression patterns in human brain cortical cell populations of the driving genes identified for rs72958644, *ADAM23* (**E**) and rs114248865, *GRIA1* (**H**).

Figure 3. Gene prioritizing summary. Summary table showing the seven genome-wide significant loci from the multi-ancestry analysis (first column), the total number of genes identified in each of the locus (second column) and gene name for genes for which we have found some kind of evidence (third column). We have included the results from the gene-based analyses, the presence of any eQTL in GTEx portal or Braineac for any of the genome-wide or suggestive variants, if the gene is differentially expressed in any bran region according to the snRNAseq data and the results from Mendelian randomization using Westra dataset (whole blood) or GTEx portal (all tissues). Black dots indicate that the gene was not found, red is that it was found but was not significant, yellow it was moderately significant ($0.05<p<1\times10^{-03}$) and green shows a significant association ($p<1\times10^{-03}$).

Table 1. Demographic Characteristics of the GENISIS cohort by country

	Spain (N=3,419)	Finland (N=490)	Poland (N=356)	US-EuA (N=798)	Costa Rica (N=141)	Mexico (N=63)	Korea (N=285)	US-AfA (N=324)	GENISIS (N=5,876)
Age (years)*	76.0 (66.0-83.0)	68.0 (58.0-76.0)	71.0 (63.0-80.0)	70.0 (60.0-79.0)	67.0 (56.0-78.0)	67.0 (50.5-75.5)	69.0 (58.0-78.0)	63.0 (54.0-74.3)	73.0 (62.0-81.0)
Sex (Females, %)	1,554 (45.5%)	193 (39.4%)	159 (44.7%)	337 (42.2%)	57 (40.4%)	28 (44.4%)	91 9 (31.9%)	169 (52.2%)	2,588 (44.0%)
Baseline NIHSS*	10.0 (5.0-17.0)	5.0 (2.0-9.0)	6.0 (3.0-12.0)	6.0 (3.0-8.2)	13.0 (9.0-18.0)	11.0 (6.0-14.5)	4.0 (2.0-8.0)	7.0 (4.0-12.0)	8.90 (4.0-15.0)
tPA Treatment (%)	48.32%	48.37%	59.55%	73.81%	100%	46.03%	28.07%	75.62%	54.20%
ΔNIHSS	2.77±5.42	2.34±5.68	2.12±3.40	2.11±5.98	6.00±7.14	3.40±4.90	1.17±3.40	2.37±6.29	2.56±5.52
TOAST Classification (%):**									
Cardioembolic	38.32%	41.63%	29.21%	37.72%	21.28%	23.81%	30.53%	29.01%	36.50%
Large Artery	17.17%	16.53%	12.36%	13.03%	39.01%	25.40%	24.56%	8.64%	16.76%
Small Vessel Disease	9.15%	6.73%	3.09%	13.16%	12.77%	14.29%	17.89%	16.98%	10.14%
Other	2.46%	8.16%	2.81%	3.13%	2.13%	15.87%	13.68%	3.09%	3.76%
Undetermined	32.90%	26.94%	52.53%	32.96%	24.11%	20.63%	13.33%	42.28%	32.83%

Values are expressed as mean±Standard Deviation. *Values expressed as median (95% confidence interval). EuA: European American Ancestry, AfA: African American Ancestry.

**TOAST classification criteria [67]

Table 2. Summary Statistics for the Multi-Ancestry Meta-Analysis top hits by cohort

SNP		rs58763243		rs13403787		rs72958644		rs12641856		rs114248865		rs6930598		rs10807797	
	MAF*	0.070		0.158		0.035		0.074		0.054		0.100		0.579	
	Effect Allele	G		A		T		A		T		T		A	
	Chr:Position	2:7762999		2:178129146		2:207515652		4:182609865		5:153073742		6:162515526		7:19995629	
		Beta	**P**	**Beta**	**P**	**Beta**	**P**	**Beta**	**P**	**Beta**	**P**	**Beta**	**P**	**Beta**	**P**
Non-Hispanic Whites Cohorts	Spain	-0.393	0.188	0.716	5.05×10^{-05}	1.609	1.00×10^{-06}	-0.185	0.438	1.230	8.41×10^{-07}	-0.311	0.126	0.531	4.80×10^{-05}
	Finland	-1.308	8.03×10^{-03}	0.248	0.580	0.955	0.183	-0.136	0.782	1.509	0.029	-0.396	0.486	0.634	0.018
	Poland	-0.360	0.527	0.458	0.427	1.664	0.053	0.499	0.403	NA	NA	0.395	0.531	0.130	0.701
	US EuA	-0.847	0.084	NA	NA	0.792	0.141	-0.659	0.212	0.064	0.928	-0.139	0.781	0.530	0.069
Non-Hispanic White META		-0.640	2.34×10^{-03}	0.639	5.23×10^{-05}	1.430	3.44×10^{-08}	-0.170	0.367	1.143	02.79×10^{-07}	-0.246	0.153	0.509	8.98×10^{-07}
Hispanic Cohorts	Costa Rica	0.485	0.691	2.812	0.053	2.862	0.211	0.872	0.685	1.760	0.381	-0.561	0.698	1.880	0.019
	Mexico	-0.058	0.953	2.897	0.102	NA	NA	NA	NA	0.370	0.847	-4.864	2.66×10^{-03}	-0.881	0.298
Hispanic META		0.159	0.836	2.847	0.010	2.862	0.207	0.872	0.684	1.032	0.455	-2.576	0.014	0.574	0.319
Additional Cohorts	Korea	-0.282	0.456	0.841	3.23×10^{-03}	NA	NA	3.654	1.38×10^{-07}	NA	NA	2.615	1.34×10^{-06}	0.487	0.092
	US AfA	-6.555	6.59×10^{-08}	NA	NA	NA	NA	NA	NA	NA	NA	-0.512	0.381	0.424	0.378
		Effect†	**LBF‡**	**Effect**	**LBF**	**Effect**	**LBF**	**Effect**	**LBF**	**Effect**	**LBF**	**Effect**	**LBF**	**Effect**	**LBF**
MANTRA**		**-+--**	**6.51**	**+++?**	**5.573**	**++??**	**6.343**	**-++?**	**5.495**	**++??**	**5.289**	**--+-**	**5.300**	**++++**	**5.697**

*MAF=Minor Allele Frequency;† Direction of effect are showed in the following order: Non-Hispanic Whites, Hispanic, Korea and US AfA; ‡Log Bayes Factor; NA=Not Available due to MAF below the inclusion threshold (0.03)

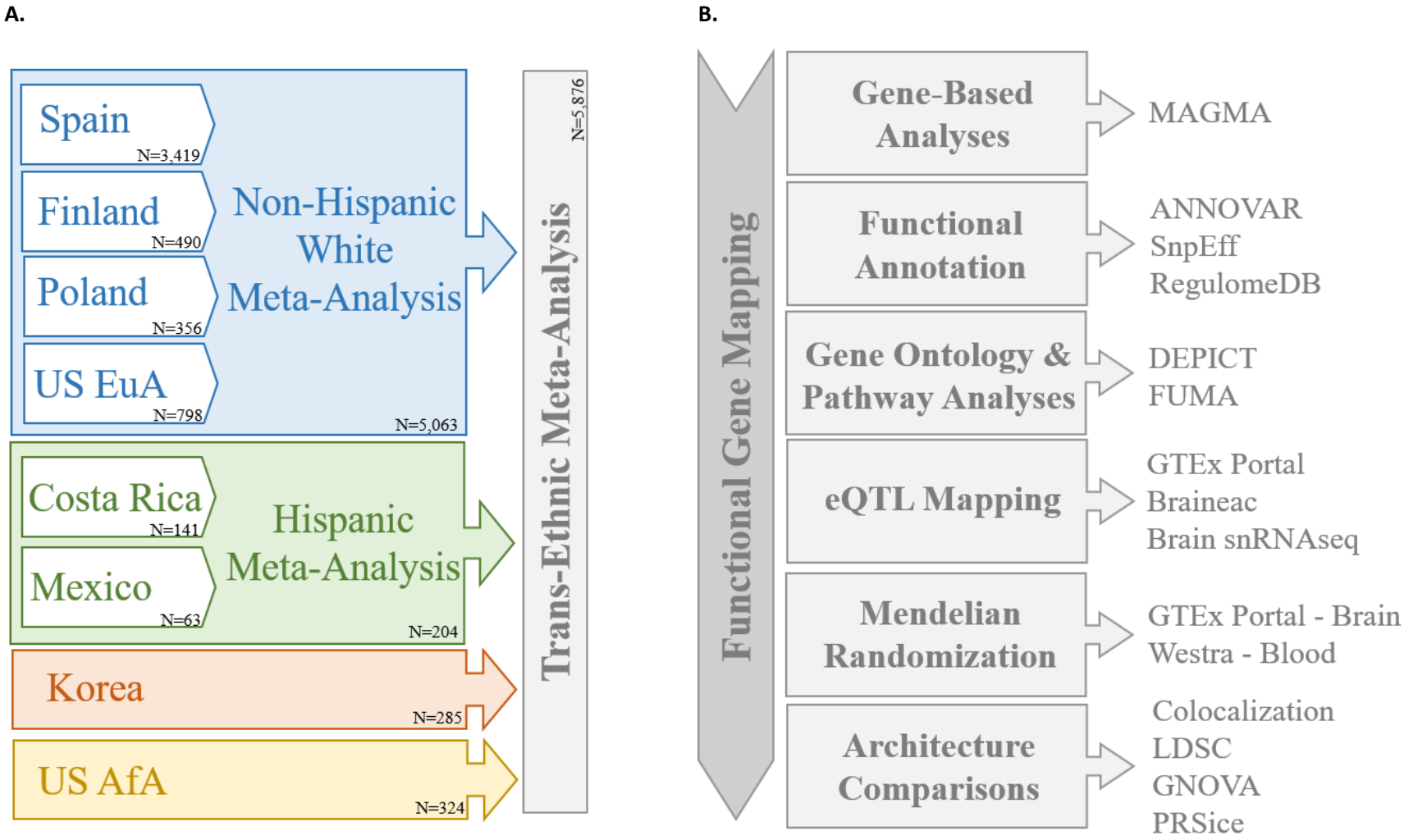
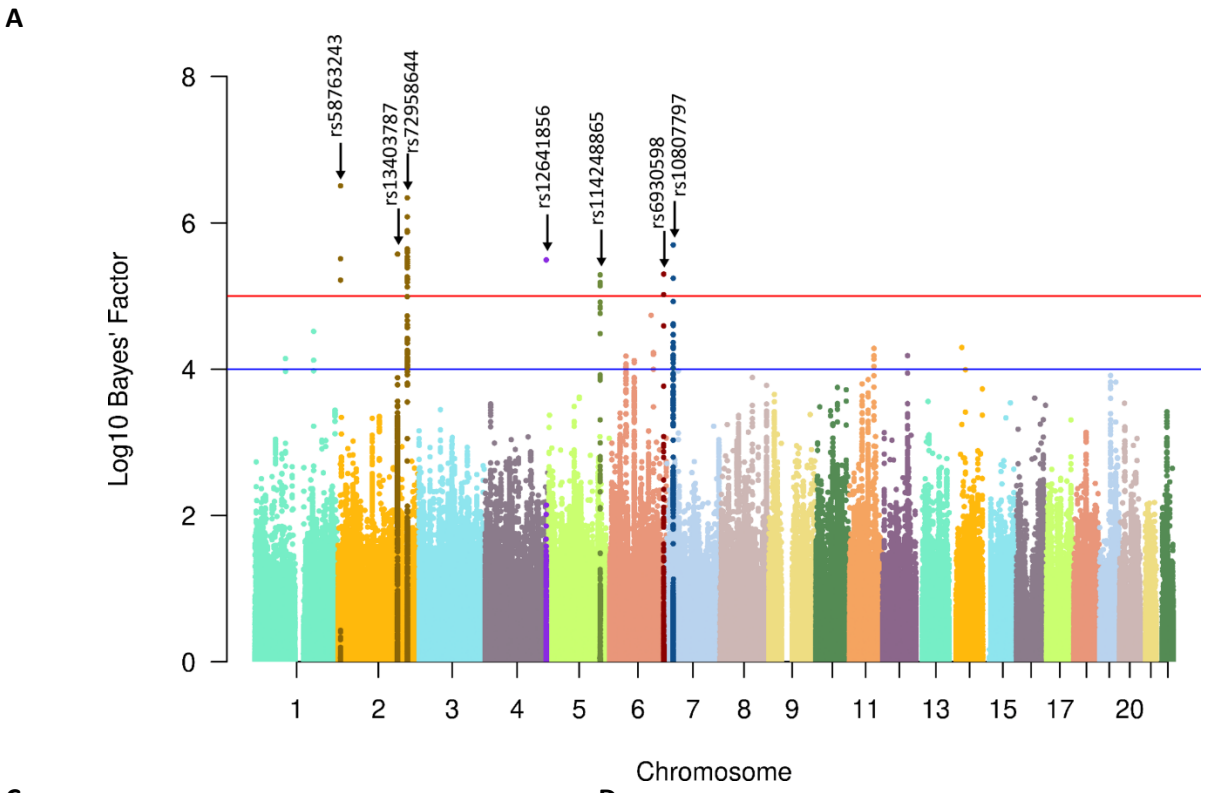
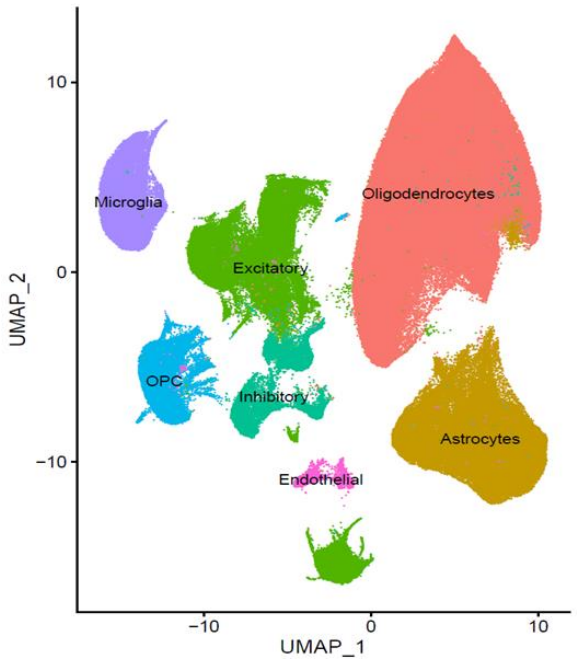
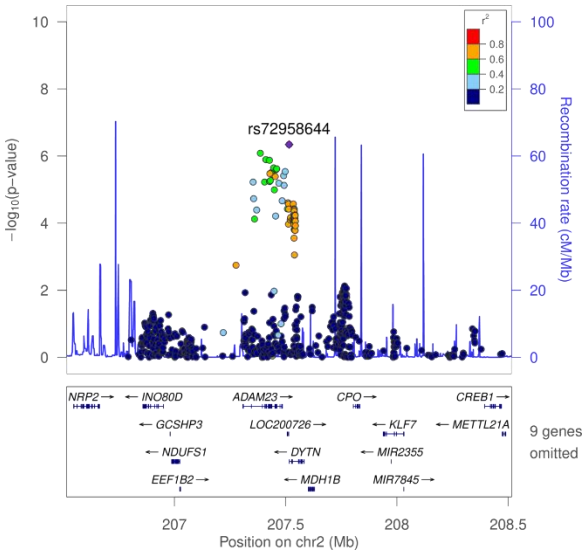


Figure 1. Study design. Summarized description of **A.** the multi-step approach used to account for the genetic heterogeneity intrinsic to the multi-ancestry nature of the GENISIS study. We performed single variant analysis in each of the participating countries separately. Then we meta-analyzes all the non-Hispanic whites (blue) and Hispanic (green) ethnicities. Finally, we analyzed the non-Hispanic whites, Hispanics, Korea (orange) and US participants with African descent (US AfA – yellow) using a Bayesian model. The variants with genome-wide significant or suggestive results were annotated using **B.** sequential steps to elucidate the gene driving the association. We performed gene-based and pathway analyses, we collected the information available in publicly available datasets and we performed Mendelian randomization. We also performed genetic architecture overlap tests to test if there was any overlap with known risk factors.

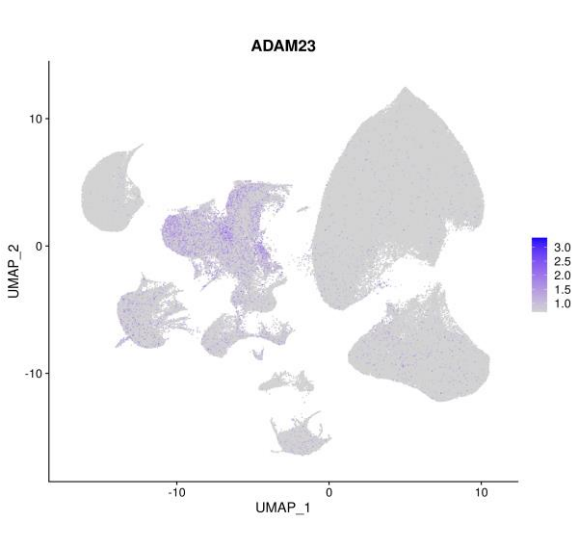






D

Population	Beta	St. Error	P Value
Spain	1.609	0.328	1.001E-06
Finland	0.955	0.716	0.187
Poland	1.664	0.856	0.053
US EuA	0.972	0.660	0.141
Costa Rica	2.862	2.276	0.211
Mexico	MAF<0.03		
Korea	Monomorphic		
US AfA	MAF<0.03		
Summary	**1.450**		



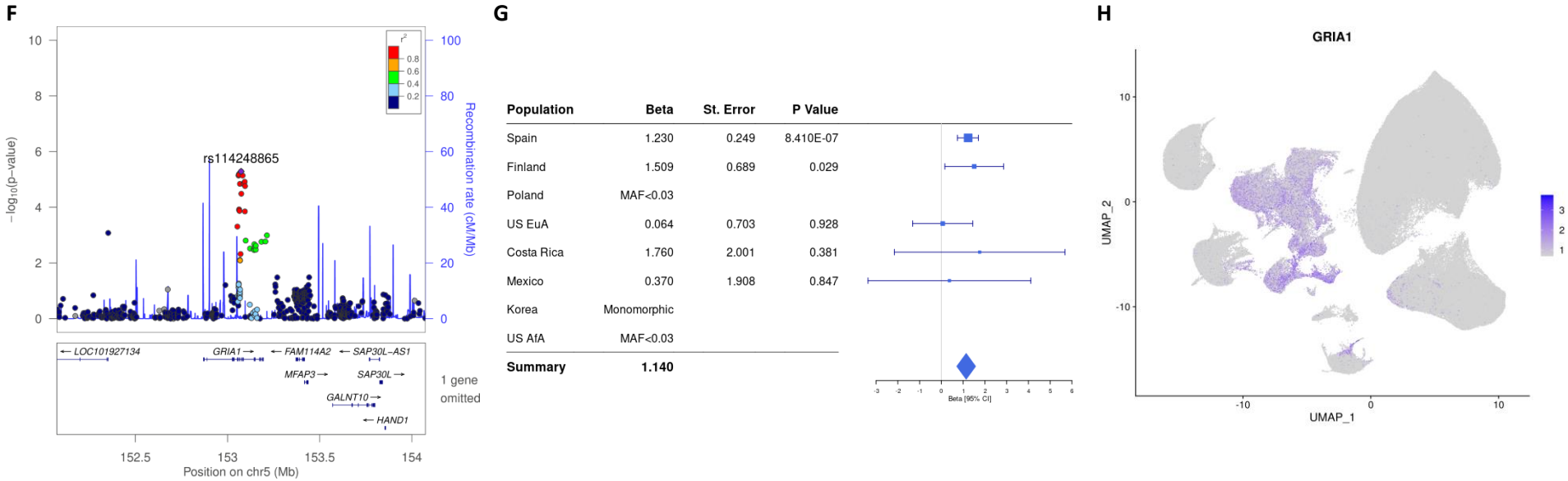


Figure 2. Association and annotation results. A. Association plot for ΔNIHSS. Manhattan plot shows LBF values from the multi-ancestry meta-analysis in each genomic location. The red line indicates the GWAS significant threshold (LBF>5) and the blue line the GWAS suggestive threshold (LBF>4). The genome-wide significant loci are highlighted. Local Manhattan plots are shown for **C.** rs72958644 and **F.** rs114248865 along with the corresponding forest plots, **D** and **G**, showing the contribution of each population to the overall signal. As part of the functional gene mapping, we accessed **B.** an in-house single nuclei data to describe the expression patterns in brain parietal lobe cell populations of the driving genes identified for **E.** rs72958644, *ADAM23* and **H.** rs114248865, *GRIA1*.

Locus	Total Genes	Gene Name	Gene Based Analysis	eQTL mapping			Mendelian Randomization		Total Evidence
				GTEx portal	Braineac	snRNAseq	Westra	GTEx Portal	
2p25.1	14	RSAD2	●	⬢	◆	●	●	●	1
		RNF144A	●	⬢	◆	◆	●	◆	1.5
		LOC386597	●	●	●	●	●	●	0.5
2q31.2	24	AC074286.1	●	◆	◆	●	●	●	1
		DFNB59	●	⬢	●	⬢	●	●	2
2q33.3	24	ADAM23	●	●	◆	●	●	●	4.5
		DYTN	●	⬢	●	⬢	●	●	1
		NRP2	◆	⬢	●	◆	●	●	2
		CREB1	●	⬢	◆	◆	●	●	1.5
		PARD3	◆	⬢	●	⬢	●	●	2
4q34.3	4	MGC45800	●	●	●	●	●	●	0.2
5q33.2	9	GRIA1	●	⬢	●	●	●	●	1.5
		GALNT10	◆	⬢	◆	◆	●	●	2
6q26	6	PARK2	●	⬢	●	⬢	●	●	1
7p21.1	15	TWISTNB	●	●	●	●	●	●	2
		ABCB5	●	⬢	●	⬢	●	●	1

● Not Found ● No Evidence – $p>0.05$ ● Some Evidence – $0.05<p<1\times10^{-03}$ ● Strong Evidence – $p<1\times10^{-03}$

○ Positive Effect ◇ Negative Effect ⬡ Unknown Effect

Figure 3. Gene prioritizing summary. Summary table showing the seven genome-wide significant loci from the multi-ancestry analysis (first column), the total number of genes identified in each of the locus (second column) and gene name for genes for which we have found some kind of evidence (third column). We have included the results from the gene-based analyses, the presence of any eQTL in GTEx portal or Braineac for any of the genome-wide or suggestive variants, if the gene is differentially expressed in neurons according to the snRNAseq data and the results from Mendelian randomization using Westra dataset (whole blood) or GTEx portal (all tissues). Black dots indicate that the gene was not found, red is that it was found but was not significant, yellow it was moderately significant ($0.05<p<1\times10^{-03}$) and green shows a significant association ($p<1\times10^{-03}$). Shape indicates the direction of effect, round positive, diamond negative and hexagon unknown.