# Multi-ancestry GWAS reveals excitotoxicity associated with outcome after ischaemic stroke

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

1 Department of Psychiatry, Washington University School of Medicine, Saint Louis (63110), Missouri, USA

2 NeuroGenomics and Informatics, Washington University School of Medicine, Saint Louis (63110), Missouri, US

3 Department of Neurology, Washington University School of Medicine, Saint Louis (63110), Missouri, US

4 Department of Emergency Medicine, Washington University School of Medicine, Saint Louis (63110), Missouri, USA

5 Stroke Unit, Vall d'Hebron University Hospital, Universitat de Barcelona, Barcelona (08035), Spain

6 Hope Center for Neurological Disorders, Washington University School of Medicine, Saint Louis (63110), Missouri, USA

7 The Charles F. and Joanne Knight Alzheimer Disease Research Center, Washington University School of Medicine, Saint Louis (63110), Missouri, USA

8 Department of Neuroscience, Katholieke Universiteit Leuven, Campus Gasthuisberg O&N2; Leuven (BE-3000), Belgium

9 Department of Neurology, School of Medical Sciences, University of Campinas (UNICAMP), Cidade Universitaria, Campinas (13083-887), Brazil

10 Brazilian Institute of Neuroscience and Neurotecnology (BRAINN), R. Tessalia Viera de Camargo, Campinas (13083-887), Brazil
11 Department of neurology, Hospital Virgen del Rocio, University of Seville, Seville (41013), Spain
12 Hospital Virgen de la Macarena, University of Seville, Seville (41009), Spain
13 Department of Neurology, Hospital Clinico Universitario Valladolid, Valladolid University, Valladolid (47003), Spain
14 Department of Neurology, Mutua Terrassa University Hospital, Universitat de Barcelona, Terrassa (08221), Spain
15 Fundacio Docencia i Recerca Mutua Terrassa, Universitat de Barcelona, Terrassa (08221), Spain
16 Department of Neurology, Hospital de la Santa Creu i Sant Pau, Universitat Autonoma de Barcelona, Barcelona (08041), Spain
17 Department of Neurology, Southern Illinois Healthcare Memorial Hospital of Carbondale, Carbondale (62901), Illinois, USA
18 Department of Biology, Universitat de les Illes Balears, Palma (07122), Spain
19 Department of Neurology, Hospital Universitari Son Espases, Universitat de les Illes Balears, Palma (07120), Spain
20 Department of Neurology, Hospital Clinic de Barcelona, Universitat de Barcelona, Barcelona (08036), Spain
21 Research Unit, Complejo Hospitalario Universitario de Albacete. Albacete (02008), Spain
22 Department of Neurology, Samsung Medical Center, Seoul, South Korea
23 Neurovascular Research Group, Institut Hospital del Mar de Investigacions Mediques, Barcelona (08003), Spain
24 Department of Surgical Sciences, Orthopedics, Uppsala University, Uppsala, 75185, Sweden
25 Department of Neurology, Jagiellonian University, Krakow(31-007), Poland
26 Department of Neurology, Hospital Germans Trias i Pujol, Universitat Autonoma de Barcelona, Badalona (08916), Spain
27 Department of Neurology, Hospital Universitario Central de Asturias, Oviedo, Spain
28 Department of Neurology, Sahlgrenska University Hospital, University of Gothenburg, 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), Santiago de Compostela (15706), Spain
31 Department of Radiology, Washington University School of Medicine, Saint Louis (63110), Missouri, USA
32 Instituto Nacional de Neurologia y Neurocirurgia de Mexico, Ciudad de Mexico (14269), Mexico
33 Department of Neurology, Providence St. Vincent Medical Center, Portland (97225), Oregon, USA
34 Neurosciences Department, Hospital Rafael A. Calderon Guardia, Aranjuez, San José, Costa Rica
35 Department of Neurology and MIND Institute, University of California at Davis, Sacramento (95817), California, USA
36 Department of Neurology, Helsinki University Hospital, Helsinki (00290), Finland
37 Translational Biology, Biogen, Inc, Cambridge (02142), Massachusetts, USA
38 Genetics and Neurodevelomental Disease Research Unit, Biogen, Inc, Cambridge (02142), Massachusetts, USA
39 Department of Biomedical Engineering, Washington University School of Medicine, Saint Louis (63110), Missouri, USA
40 Stroke and Cerebrovascular Center, Washington University School of Medicine, Saint Louis (63110), Missouri, US
41 Department of Genetics, Washington University School of Medicine, Saint Louis (63110), Missouri, USA

Correspondence to: Carlos Cruchaga
Washington University School of Medicine
660 South Euclid Avenue
Campus Box 8134
Saint Louis, MO 63110, USA
E-mail: cruchagac@wustl.edu

Correspondence may also be addressed to: Jin-Moo Lee
Washington University School of Medicine
660 South Euclid Avenue
Campus Box 8111
St. Louis, MO 63110, USA
E-mail: leejm@wustl.edu

**Running title**: Excitotoxicity and ischaemic stroke

# 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. Eight loci (1p21.1, 1q42.2, 2p25.1, 2q31.2, 2q33.3, 5q33.2, 7p21.2,and 13q31.1) were genome-wide significant and explained 1.8% 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)=5.41) was driving the association for 2q33.3. Gene based analyses suggested that *GRIA1* (LBF=5.19), which is predominantly expressed in brain, is the gene driving the association for the 5q33.2 locus. These analyses also nominated *GNPAT* (LBF=7.64)*ABCB5* (LBF=5.97) for the 1p21.1 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 genetic evidence in humans that excitotoxicity may contribute to early neurological instability after acute ischemic stroke.

**Keywords:** ischaemic stroke; neuroprotection; genetics; NIHSS

**Abbreviations**: AfA=African-American descent; BMI=Body Mass Index; eQTL=Expression Quantitative Trait Loci; EuA=European Descent; GCTA=Genome-wide Complex Traits Analysis; GENISIS=Genetics of Early Neurological InStability after Ischemic Stroke; GTEx=Genotype-Tissue Expression; GWAS=Genome-Wide Association Studies; LBF=Log Bayes Factor; LD=Linkage Disequilibrium; MAF=Minor Allele Frequency; NIH=National Institute of Health; NIHSS=NIH Stroke Scale; SMR=Summary-data-based Mendelian Randomization; SNP=Single Nucleotide Polymorphism

# 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. All available inpatient data, including history, clinical exam, lab values, diagnostic tests, imaging, and discharge diagnosis were utilized to confirm the diagnosis of ischemic stroke. 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 (Fig. 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) $> 5$[19,20], were annotated using bioinformatics tools to identify the gene driving the genetic signal (Fig. 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.[21] 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[22] and IMPUTE2.[23] 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.[24] 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 (Fig. 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.[25] 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), tPA, and baseline NIHSS to adjust for stroke severity. The primary focus of the GWAS was on early neurological change; thus, baseline NIHSS was included as a covariate in the model. 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.[26] Second, we meta-analyzed the populations with similar ethnic backgrounds using with fixed effect meta-analyses using METAL.[27] 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.[19,20]

To ensure that the loci were related to ΔNIHSS in all ischemic strokes and were not specific to a stroke subtype (defined by TOAST criteria), we also performed joint analyses for cardioembolic stroke (N=2,149), large-artery atherosclerosis (N=980), small-vessel disease (618), undetermined (N=1,926) and other (N=222). No significant loci were found associated with specific stroke subtypes.

As both time to evaluation and time to treatment with tPA have been shown to be predictors of stroke outcome, we conducted sensitivity analyses for subjects that had available information regarding elapsed time to evaluation (n=4,477) and elapsed time to tPA (n=2,312). In both instances, the results of the joint GWAS with and without the time variable of interest demonstrated highly correlated beta and p-values and did not reveal any additional potential loci associated with ΔNIHSS.

# Functional annotation

We annotated all the variants with suggestive associations (LBF>4) with ANNOVAR[28] and SnpEff[29] 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.[30]

DEPICT[31] and FUMA[32] were used to perform gene ontology and pathways analyses. We also leveraged brain single nuclei RNA expression data (http://ngi.pub/snuclRNA-seq/)[21], 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 (Fig. 1B). Finally, we accessed blood RNA expression data taken at different times after stroke onset (3h, 5h and 24h) from

the CLEAR trial[33] (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).[34]

## 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 12/09/2021), the Brain eQTL Almanac (Braineac) and an in-house dataset that includes brain expression data for 613 brains[35]. We used the summary-data-based Mendelian Randomization (SMR)[36] and colocalization[37] 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 (Fig. 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[38] 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[39] 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[40], LDSC[41] and GNOVA[42] (Fig. 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[43], lipid levels[44] and body mass index (BMI)[45]) are genetically correlated to ΔNIHSS.

## Data availability

Summary statistics of the GENISIS dataset used for these analyses along with 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, 54% were treated with tPA, 20% had a prior history of stroke. 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 Fig. 1).

## 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) (Fig. 1A). We identified eight GWAS significant loci (Fig. 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=6.58), was located in a region comprised by several long non-coding RNAs and microRNAs (Supplementary Fig. 2E). 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 Fig. 2F).

The second locus, rs13403787 ($MAF_A$=0.16; LBF=5.13), was also located on chr2 in a region with more than 20 genes (Supplementary Fig. 2G). The minor allele was associated with higher (or more positive) ΔNIHSS in all cohorts (Supplementary Fig. 2H). The last genome wide significant locus in chr2 was rs16838349 ($MAF_A$=0.07; LBF=5.41), located in a region that includes *ADAM23*, *CREB1, DYTN, NRP2, MDH1B,* among many others (Fig. 2C). The signal is driven by the Non-Hispanic Whites (meta-analysis $p=8.74\times10^{-06}$), but virtually all ethnic groups contributed to this association, as the directionality was consistent across Hispanic, non-Hispanic White and AfA ethnic groups (Fig. 2D and Supplementary Table 1). However, the SNPs in this locus were monomorphic in the Asian population. We did not observe any significant correlation between ΔNIHSS with the genotype in this locus ($R^2$=0.063, p=0.09; Supplementary Table 2).

Five additional loci were identified outside chr2. Two independent loci in chr1. Both, rs1451040 ($MAF_T$=0.16, LBF=6.56) and rs9660272 ($MAF_T$=0.16, LBF=7.64) were in gene rich regions (Supplementary Fig. 2A and 2C). These two loci are highly significant in the Latino populations (Mexico and Costa Rica), with large negative effect sizes (Supplementary Fig. 2B and 2D). However, they are not significant in any of the other populations, except for rs1451040, that is nominally significant in the Korean population, but has the opposite direction of effect. The locus identified on chr5 is located on a region containing nine genes (*LOC101927134, GRIA1, FAM114A2, SAP30L, SAP30L-AS1, MFAP3, GALNT10, HAND1* and *MIR3141;* Fig. 2F). The minor allele for the top hit in this locus, rs17115057 ($MAF_T$=0.06; LBF=5.19) was associated with greater (more positive) ΔNIHSS across most of the cohorts, and was significant in the Spanish ($p=1.35\times10^{-07}$) and Finnish cohorts (p=0.03; Fig. 2G). Another locus on chr7, tagged by the variant rs10807797 ($MAF_G$=0.42; LBF=5.97), and located in a gene rich region with 15 genes, including *TWISTNB, MACC1, TMEM196, ABCB5, RPL23P8* (Supplementary Fig. 2I). 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 Fig. 2J, Supplementary Table 1). Finally, we identified a locus on chr13, tagged by rs9545725 located in a gene desert. No genes were identified in this locus (Supplementary Fig. 2K). The variants in the region were significant in the Latino cohorts but the direction of effect was not consistent across all cohorts (Supplementary Fig. 2L). Moreover, the MAF for these variants ranged between 1% in the Korean population to 15% in the African American and Spanish 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.

# 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). GCTA 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 8 genome-wide significant loci, defined as 500 bp upstream or downstream of the top signal, explained 1.8% of the total variance ($p=2.18\times10^{-04}$) of ΔNIHSS, or just 20.7% 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 (Fig. 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 Fig. 3.

Gene-based analyses using FUMA suggested that *DYTN* ($p=2.55\times10^{-04}$, Z=3.47) and *ADAM23* ($p=2.04\times10^{-03}$, Z=2.87) 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 (esophageal mucosa: $p=2.00\times10^{-06}$; Cultured Fibroblasts: $p=5.90\times10^{-05}$). MR analyses indicated that *ADAM23* (p=0.04) 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 (Fig. 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* ($p=3.70\times10^{-04}$) in the occipital cortex, but *GALNT10* was less significant in the gene-based analysis (p=0.04, Z=1.79). 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 (Fig. 2H and Supplementary 3A). *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 in peripheral blood was also nominally associated with ΔNIHSS in the CLEAR trial dataset ($p=0.002$, $r^2=0.22$).

Of the remaining six loci, we were able to map five (Supplementary Results). Briefly, eQTL analysis, revealed that 1p21.1 was likely to be driven by *COL11A1* or AMY2B. Gene-based and MR analyses suggested that the locus 1q42.2 was driven by *GNPAT.* No eQTLs were identified for 2p25.1, but gene-based analyses suggested that the signal is likely driven by *AGPS* or *TTC30A*. Regarding 2p31.2, it is likely driven by *DFNB59,* while 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 cardiovascular system ($1.8\times10^{-03}$) and the central nervous system ($p=2.0\times10^{-03}$), including the brain ($p=0.01$) and some brain regions: occipital lobe ($p=2.00\times10^{-03}$), cerebral cortex ($p=4.80\times10^{-03}$), and temporal lobe ($p=6.33\times10^{-03}$; Supplementary Table 6). The most significant pathways in the gene-set enrichment were the regulation of the heart contraction ($p=5.80\times10^{-06}$), the sodium ion transmembrane transport ($p=6.27\times10^{-06}$), the circulatory system process ($p=6.39\times10^{-06}$) learning or memory ($p=7.11\times10^{-06}$) and abnormal CNS synaptic transmission ($p=2.88\times10^{-05}$; Supplementary Table 7). Several genome-wide significant candidate genes fell within these networks, of special interest, *GRIA1* (5q33.2, LBF=5.19) in the sodium ion transmembrane transport, which adds evidence to the involvement of *GRIA1* in ΔNIHSS. MAGMA gene-set analyses did not reveal any enriched gene set associated with ΔNIHSS (Supplementary Table 8).

# Unique genetic architecture of early outcomes after stroke

We examined the genetic architecture of ΔNIHSS for shared genetic variation 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 Supplementary Fig. 4), LDSC (Supplementary Table 10) and GNOVA (Supplementary Table 11). Although the p value for PRSice was significant in the comparison of

stroke risk and ΔNIHSS, the amount of variance explained was very small ($R^2$=0.009). Additionally, this finding was not supported by LDSC or GNOVA analyses, suggesting that there is no genetic overlap, as reported in a previous work.[18] Similarly, no overlap with age at death, lipid levels, or BMI was identified by LDSC or GNOVA. Even though PRSice found significant correlations with several stroke risk factors, HDL levels (p=0.01), TG levels ($p=8.97\times10^{-04}$), total cholesterol (p=0.02), body mass index ($p=1.89\times10^{-06}$) and age at death (p=0.01), the amount of variance explained were all below 0.5%, suggesting that the overlap is minimal. LDSC was unable to calculate the heritability estimate for ΔNIHSS. GNOVA, was successful at estimating the heritability for ΔNIHSS, but could not calculate the genetic correlation estimate. Several of the heritability estimates for ΔNIHSS for overlap were 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.[4,46-48] Not only is early neurological change (as reflected by ΔNIHSS) 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 eight genome-wide significant loci that are related to ΔNIHSS. However, they explain only 1.8% of the variance, indicating that 6.9% 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 suggests that *ADAM23* is the functional gene for the locus,2q33.3. 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.[49] ADAM23 is a transmembrane member without catalytic domain, and is involved in cell-cell and cell-matrix interactions.[49,50]

Previous studies have shown that *ADAM23* is expressed in pre-synaptic membranes, linked by the extracellular protein LGI1 to post-synaptic ADAM22.[51,52] We found that *ADAM2*3 was expressed primarily in excitatory neurons of the cerebral cortex, based on our human brain single-nuclei transcriptomics dataset[21], and confirmed by the Human Transcriptomic Cell Types dataset from the Allen Brain Map.[53] Several lines of evidence suggest that ADAM23 is important for pathological synaptic excitability: 1) *adam23* is a common risk gene for canine idiopathic epilepsy;[54-56] 2) mutations in its binding partner, *LGI1*, cause the neurological syndrome, ADPEAF (autosomal dominant partial epilepsy with auditory features)[57]; 3) autoimmunity against LGI1 (as seen in limbic encephalitis) results in seizures and encephalopathy.[58]

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).[52] 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.[59,60] 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.[61-64] 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, [33,34] 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 six 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. Furthermore, specific medication information (such as type of anticoagulation medication, if being used for secondary prevention at the time of stroke) were not collected, and therefore cannot be accounted for. 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.

# Acknowledgements

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.

# Funding

This work was supported by grants from the 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); NIH/NIA-K99-AG062723 (LI),; Barnes-Jewish Hospital Foundation (JML); Biogen (CC, JML); Bright Focus Foundation, US Department of Defense, 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. 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.

# Competing interests

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.

# Supplementary material

Supplementary material is available at *Brain* online

# 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*. Mar 20 2018;137(12):e67-e492. doi:10.1161/CIR.0000000000000558
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*. Oct 2008;18(10):665. doi:10.2008/JCPSP.665665
4. Saver JL, Altman H. Relationship between neurologic deficit severity and final functional outcome shifts and strengthens during first hours after onset. *Stroke*. Jun 2012;43(6):1537-41. doi:10.1161/STROKEAHA.111.636928
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*. Dec 14 1995;333(24):1581-7. doi:10.1056/NEJM199512143332401
6. Heitsch L, Ibanez L, Carrera C, et al. 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*. Dec 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*. Sep 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*. Jun 2008;79(6):625-9. doi:10.1136/jnnp.2007.132100
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*. Jan 2009;30(1):165-70. doi:10.3174/ajnr.A1276
11. Ali LK, Saver JL. The ischemic stroke patient who worsens: new assessment and management approaches. *Reviews in neurological diseases*. Spring 2007;4(2):85-91.
12. Dreier JP. The role of spreading depression, spreading depolarization and spreading ischemia in neurological disease. *Nat Med*. Apr 2011;17(4):439-47. doi:10.1038/nm.2333
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*. Aug 2008;39(8):2249-56. doi:10.1161/STROKEAHA.107.510321
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*. Dec 2019;50(12):3618-3621. doi:10.1161/STROKEAHA.119.026593
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*. Apr 2018;50(4):524-537. doi:10.1038/s41588-018-0058-3
16. Soderholm M, Pedersen A, Lorentzen E, et al. Genome-wide association meta-analysis of functional outcome after ischemic stroke. *Neurology*. Mar 19 2019;92(12):e1271-e1283. doi:10.1212/WNL.0000000000007138
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*. Jan 4 2019;124(1):114-120. doi:10.1161/CIRCRESAHA.118.313533
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*. Jun 2019;50(6):1339-1345. doi:10.1161/STROKEAHA.118.023097
19. Morris AP. Transethnic meta-analysis of genomewide association studies. *Genetic epidemiology*. Dec 2011;35(8):809-22. doi:10.1002/gepi.20630
20. Wang X, Chua HX, Chen P, et al. Comparing methods for performing trans-ethnic meta-analysis of genome-wide association studies. *Hum Mol Genet*. Jun 1 2013;22(11):2303-11. doi:10.1093/hmg/ddt064
21. 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*. Aug 9 2019;11(1):71. doi:10.1186/s13195-019-0524-x
22. Delaneau O, Coulonges C, Zagury JF. Shape-IT: new rapid and accurate algorithm for haplotype inference. *BMC bioinformatics*. Dec 16 2008;9:540. doi:10.1186/1471-2105-9-540
23. Howie B, Marchini J, Stephens M. Genotype imputation with thousands of genomes. *G3*. Nov 2011;1(6):457-70. doi:10.1534/g3.111.001198
24. Yang J, Lee SH, Goddard ME, Visscher PM. GCTA: a tool for genome-wide complex trait analysis. *Am J Hum Genet*. Jan 7 2011;88(1):76-82. doi:10.1016/j.ajhg.2010.11.011
25. 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. doi:10.1186/s13742-015-0047-8
26. 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*. Aug 01 2005;162(3):267-78. doi:10.1093/aje/kwi187
27. Willer CJ, Li Y, Abecasis GR. METAL: fast and efficient meta-analysis of genomewide association scans. *Bioinformatics*. Sep 01 2010;26(17):2190-1. doi:10.1093/bioinformatics/btq340
28. Yang H, Wang K. Genomic variant annotation and prioritization with ANNOVAR and wANNOVAR. *Nature protocols*. Oct 2015;10(10):1556-66. doi:10.1038/nprot.2015.105
29. 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*. Apr-Jun 2012;6(2):80-92. doi:10.4161/fly.19695
30. Boyle AP, Hong EL, Hariharan M, et al. Annotation of functional variation in personal genomes using RegulomeDB. *Genome research*. Sep 2012;22(9):1790-7. doi:10.1101/gr.137323.112
31. Pers TH, Karjalainen JM, Chan Y, et al. Biological interpretation of genome-wide association studies using predicted gene functions. *Nature communications*. Jan 19 2015;6:5890. doi:10.1038/ncomms6890
32. Watanabe K, Taskesen E, van Bochoven A, Posthuma D. Functional mapping and

annotation of genetic associations with FUMA. *Nature communications*. Nov 28 2017;8(1):1826. doi:10.1038/s41467-017-01261-5
33. 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*. Dec 2008;39(12):3268-76. doi:10.1161/STROKEAHA.108.517656
34. Stamova B, Xu H, Jickling G, et al. Gene expression profiling of blood for the prediction of ischemic stroke. *Stroke*. Oct 2010;41(10):2171-7. doi:10.1161/STROKEAHA.110.588335
35. 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*. Jun 8 2018;10(1):43. doi:10.1186/s13073-018-0551-4
36. Zhu Z, Zhang F, Hu H, et al. Integration of summary data from GWAS and eQTL studies predicts complex trait gene targets. *Nat Genet*. May 2016;48(5):481-7. doi:10.1038/ng.3538
37. 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*. Apr 2009;10(2):327-34. doi:10.1093/biostatistics/kxn039
38. Westra HJ, Peters MJ, Esko T, et al. Systematic identification of trans eQTLs as putative drivers of known disease associations. *Nat Genet*. Oct 2013;45(10):1238-1243. doi:10.1038/ng.2756
39. Hormozdiari F, van de Bunt M, Segre AV, et al. Colocalization of GWAS and eQTL Signals Detects Target Genes. *Am J Hum Genet*. Dec 1 2016;99(6):1245-1260. doi:10.1016/j.ajhg.2016.10.003
40. Euesden J, Lewis CM, O'Reilly PF. PRSice: Polygenic Risk Score software. *Bioinformatics*. May 01 2015;31(9):1466-8. doi:10.1093/bioinformatics/btu848
41. Bulik-Sullivan BK, Loh PR, Finucane HK, et al. LD Score regression distinguishes confounding from polygenicity in genome-wide association studies. *Nat Genet*. Mar 2015;47(3):291-5. doi:10.1038/ng.3211
42. 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*. Dec 7 2017;101(6):939-964. doi:10.1016/j.ajhg.2017.11.001
43. Pilling LC, Kuo CL, Sicinski K, et al. Human longevity: 25 genetic loci associated in 389,166 UK biobank participants. *Aging (Albany NY)*. Dec 6 2017;9(12):2504-2520. doi:10.18632/aging.101334
44. Liu DJ, Peloso GM, Yu H, et al. Exome-wide association study of plasma lipids in >300,000 individuals. *Nat Genet*. Dec 2017;49(12):1758-1766. doi:10.1038/ng.3977
45. 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*. Jan 2018;50(1):26-41. doi:10.1038/s41588-017-0011-x
46. 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*. Jul 13 1999;53(1):126-31.
47. 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*. Nov 2013;22(8):e590-6. doi:10.1016/j.jstrokecerebrovasdis.2013.07.024
48. 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*. Jan 2017;48(1):105-110. doi:10.1161/STROKEAHA.116.014456
49. Goldsmith AP, Gossage SJ, ffrench-Constant C. ADAM23 is a cell-surface glycoprotein expressed by central nervous system neurons. *Journal of neuroscience research*. Dec 1 2004;78(5):647-58. doi:10.1002/jnr.20320
50. Mitchell KJ, Pinson KI, Kelly OG, et al. Functional analysis of secreted and transmembrane proteins critical to mouse development. *Nat Genet*. Jul 2001;28(3):241-9. doi:10.1038/90074
51. 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. doi:10.7150/ijbs.4.387
52. 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*. Jul 28 2015;112(30):E4129-37. doi:10.1073/pnas.1511910112
53. Hawrylycz MJ, Lein ES, Guillozet-Bongaarts AL, et al. An anatomically comprehensive atlas of the adult human brain transcriptome. *Nature*. Sep 20 2012;489(7416):391-399. doi:10.1038/nature11405
54. Koskinen LL, Seppala EH, Belanger JM, et al. Identification of a common risk haplotype for canine idiopathic epilepsy in the ADAM23 gene. *BMC genomics*. Jun 18 2015;16:465. doi:10.1186/s12864-015-1651-9
55. Koskinen LL, Seppala EH, Weissl J, et al. ADAM23 is a common risk gene for canine idiopathic epilepsy. *BMC genetics*. Jan 31 2017;18(1):8. doi:10.1186/s12863-017-0478-6
56. 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. doi:10.1371/journal.pone.0033549
57. Berkovic SF, Izzillo P, McMahon JM, et al. LGI1 mutations in temporal lobe epilepsies. *Neurology*. Apr 13 2004;62(7):1115-9. doi:10.1212/01.wnl.0000118213.94650.81
58. 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*. Oct 2016;15(10):970-4. doi:10.1016/j.autrev.2016.07.018
59. Lee JM, Zipfel GJ, Choi DW. The changing landscape of ischaemic brain injury mechanisms. *Nature*. Jun 24 1999;399(6738 Suppl):A7-14. doi:10.1038/399a007
60. 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*. Jun 2005;57(2):253-77. doi:10.1124/pr.57.2.7
61. Chamorro A, Dirnagl U, Urra X, Planas AM. Neuroprotection in acute stroke: targeting excitotoxicity, oxidative and nitrosative stress, and inflammation. *The Lancet Neurology*. Jul 2016;15(8):869-881. doi:10.1016/S1474-4422(16)00114-9
62. Wahlgren NG, Ahmed N. Neuroprotection in cerebral ischaemia: facts and fancies--the need for new approaches. *Cerebrovasc Dis*. 2004;17 Suppl 1:153-66. doi:10.1159/000074808
63. Neuhaus AA, Couch Y, Hadley G, Buchan AM. Neuroprotection in stroke: the importance of collaboration and reproducibility. *Brain : a journal of neurology*. Aug 1 2017;140(8):2079-2092. doi:10.1093/brain/awx126
64. Castillo J, Davalos A, Noya M. Progression of ischaemic stroke and excitotoxic aminoacids. *Lancet*. Jan 11 1997;349(9045):79-83. doi:10.1016/S0140-6736(96)04453-4

# 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 rs16838349 (**C**) and rs17115057 (**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 rs16838349, *ADAM23* (**E**) and rs17115057, *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 ($p<1\times10^{-03}$).

| | 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)[a] | 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[a] | 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[b] | 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[c] (%): | | | | | | | | | |
| 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% |

**Table 1 Demographic Characteristics of the GENISIS cohort by country**

[a]Values are expressed as median (95% confidence interval).
[b]Values are expressed as mean ± Standard Deviation
[c]TOAST classification criteria [1]

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

| SNP | rs1451040 | | rs9660272 | | rs58763243 | | rs13403787 | | rs16838349 | | rs17115054 | | rs10807797 | | rs9545725 | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| MAF | 0.160 | | 0.161 | | 0.070 | | 0.158 | | 0.067 | | 0.059 | | 0.579 | | 0.108 | |
| Effect Allele | T | | T | | G | | A | | G | | T | | A | | A | |
| Chr:Position | 1:103158738 | | 1:232253211 | | 2:7762999 | | 2:178459146 | | 2:207427437 | | 5:153074938 | | 7:19995629 | | 13:82056977 | |
| | **Beta** | **P** | **Beta** | **P** | **Beta** | **P** | **Beta** | **P** | **Beta** | **P** | **Beta** | **P** | **Beta** | **P** | **Beta** | **P** |
| **Non-Hispanic Whites Cohorts** | | | | | | | | | | | | | | | | |
| Spain | −0.095 | 0.565 | −0.060 | 0.712 | −0.413 | 0.158 | 0.687 | $7.50 \times 10^{-5}$ | 0.803 | 0.001 | 1.267 | $1.35 \times 10^{-7}$ | 0.560 | $1.33 \times 10^{-5}$ | 0.096 | 0.578 |
| Finland | −0.046 | 0.905 | −0.273 | 0.481 | −1.426 | 0.003 | 0.133 | 0.763 | 0.674 | 0.250 | 1.455 | 0.032 | 0.635 | 0.016 | 0.015 | 0.973 |
| Poland | 0.360 | 0.411 | −0.427 | 0.320 | −0.285 | 0.613 | 0.310 | 0.588 | 1.893 | 0.005 | NA[a] | NA[a] | 0.127 | 0.705 | 0.074 | 0.892 |
| US EuA | 0.254 | 0.510 | −0.586 | 0.117 | −0.844 | 0.084 | NA[a] | NA[a] | 0.793 | 0.130 | −0.032 | 0.963 | 0.505 | 0.082 | −0.012 | 0.977 |
| Non-Hispanic White META[b] | −0.004 | 0.977 | −0.187 | 0.159 | 0.661 | $1.41 \times 10^{-3}$ | 0.591 | $1.43 \times 10^{-4}$ | 0.884 | $8.74 \times 10^{-6}$ | 1.162 | $6.41 \times 10^{-8}$ | 0.524 | $2.82 \times 10^{-7}$ | 0.073 | 0.614 |
| **Hispanic Cohorts** | | | | | | | | | | | | | | | | |
| Costa Rica | −3.236 | 0.014 | −5.203 | 0.002 | 0.485 | 0.691 | 2.812 | 0.053 | 2.674 | 0.029 | 0.249 | 0.893 | 1.880 | 0.019 | −8.735 | $1.33 \times 10^{-5}$ |
| Mexico | −4.569 | $7.45 \times 10^{-6}$ | −6.335 | $1.17 \times 10^{-6}$ | 0.168 | 0.825 | 2.519 | 0.066 | 1.451 | 0.203 | 0.616 | 0.678 | −0.624 | 0.343 | 1.424 | $6.76 \times 10^{-4}$ |
| Hispanic META[b] | −4.131 | $3.45 \times 10^{-8}$ | −5.953 | $1.81 \times 10^{-10}$ | −0.257 | 0.690 | 2.655 | $6.74 \times 10^{-3}$ | 2.019 | 0.001 | 0.473 | 0.681 | 0.385 | 0.444 | −6.42 | $2.07 \times 10^{-8}$ |
| **Additional Cohorts** | | | | | | | | | | | | | | | | |
| Korea | 1.276 | 0.023 | −0.566 | 0.364 | −0.277 | 0.462 | 0.827 | 0.004 | NA[a] | NA[a] | NA[a] | NA | 0.500 | 0.084 | NA | NA |
| US AfA | −0.919 | 0.071 | 0.078 | 0.887 | −6.491 | $8.24 \times 10^{-8}$ | NA[a] | NA[a] | −2.896 | 0.020 | −1.230 | 0.158 | 0.416 | 0.387 | −0.514 | 0.452 |
| | **Effect[c]** | **LBF** | **Effect[c]** | **LBF** | **Effect[c]** | **LBF** | **Effect[c]** | **LBF** | **Effect[c]** | **LBF** | **Effect[c]** | **LBF** | **Effect[c]** | **LBF** | **Effect[c]** | **LBF** |
| MANTRA[b] | −−+− | 6.56 | −−−+ | 7.64 | −+−− | 6.58 | +++? | 5.13 | ++?+ | 5.41 | ++?− | 5.19 | ++++ | 5.97 | +−−? | 5.50 |

LBF = Log Bayes factor.

[a]NA=Not Available due to MAF below the inclusion threshold (0.03) or non-convergence of the statistical model.

[b]The results from the meta-analysis of the single populations.

[c]Direction of effect are showed in the following order: Non-Hispanic Whites, Hispanic, Korea and US AfA; +/−/? = positive beta in given population/negative beta in given population/Not present in given population.

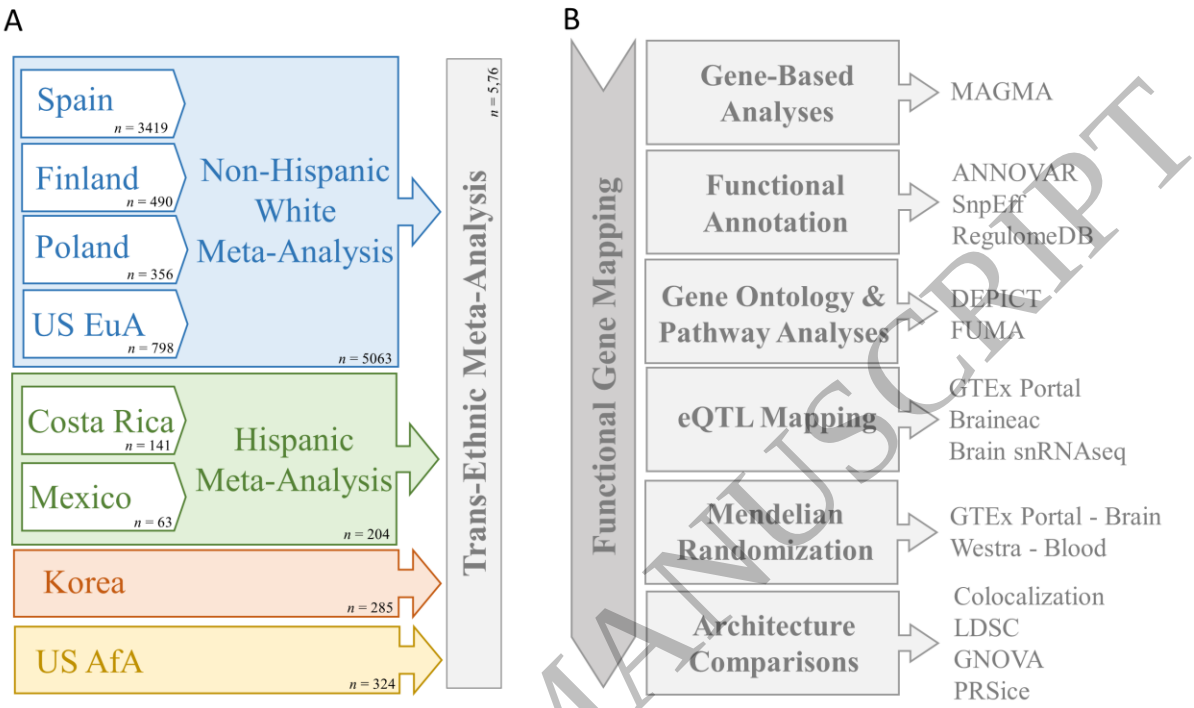


**Figure 1**
**165x103 mm (.15 x DPI)**

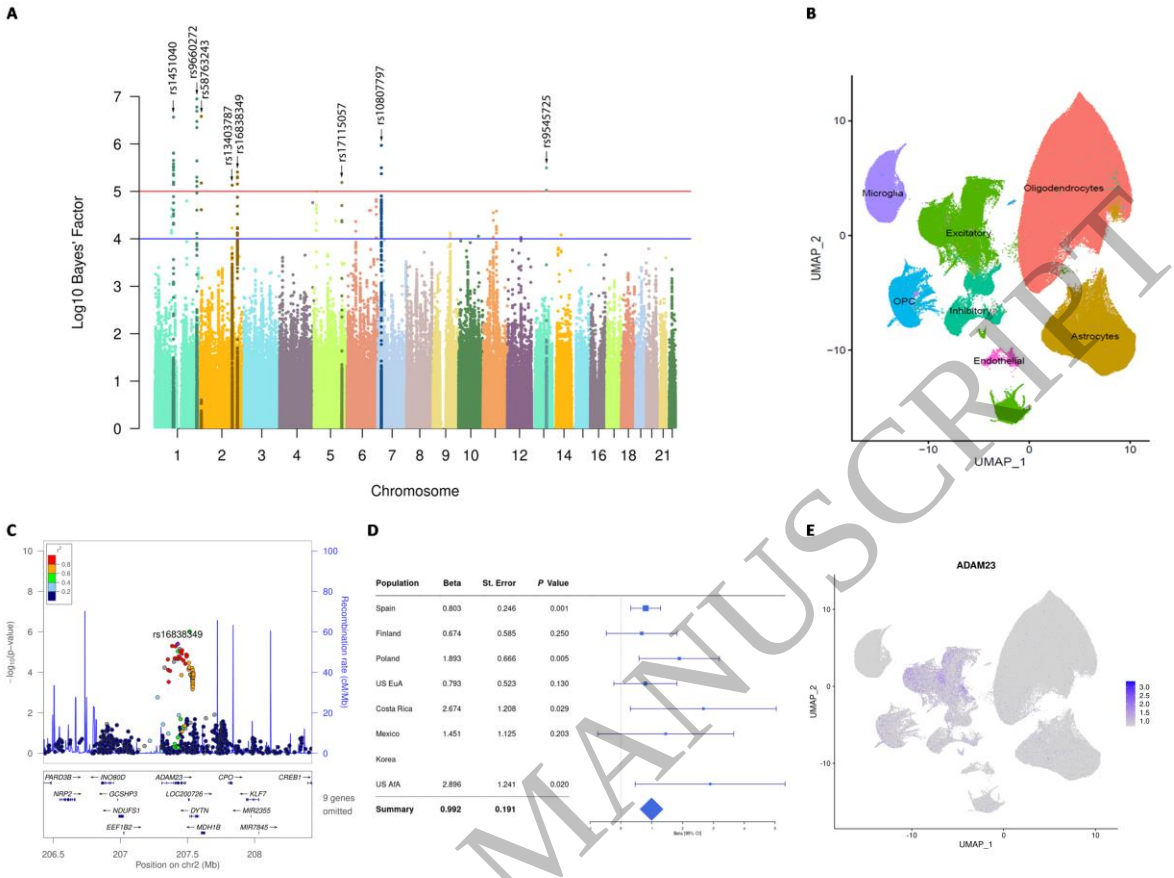


**Figure 2**

**165x115 mm (.15 x DPI)**

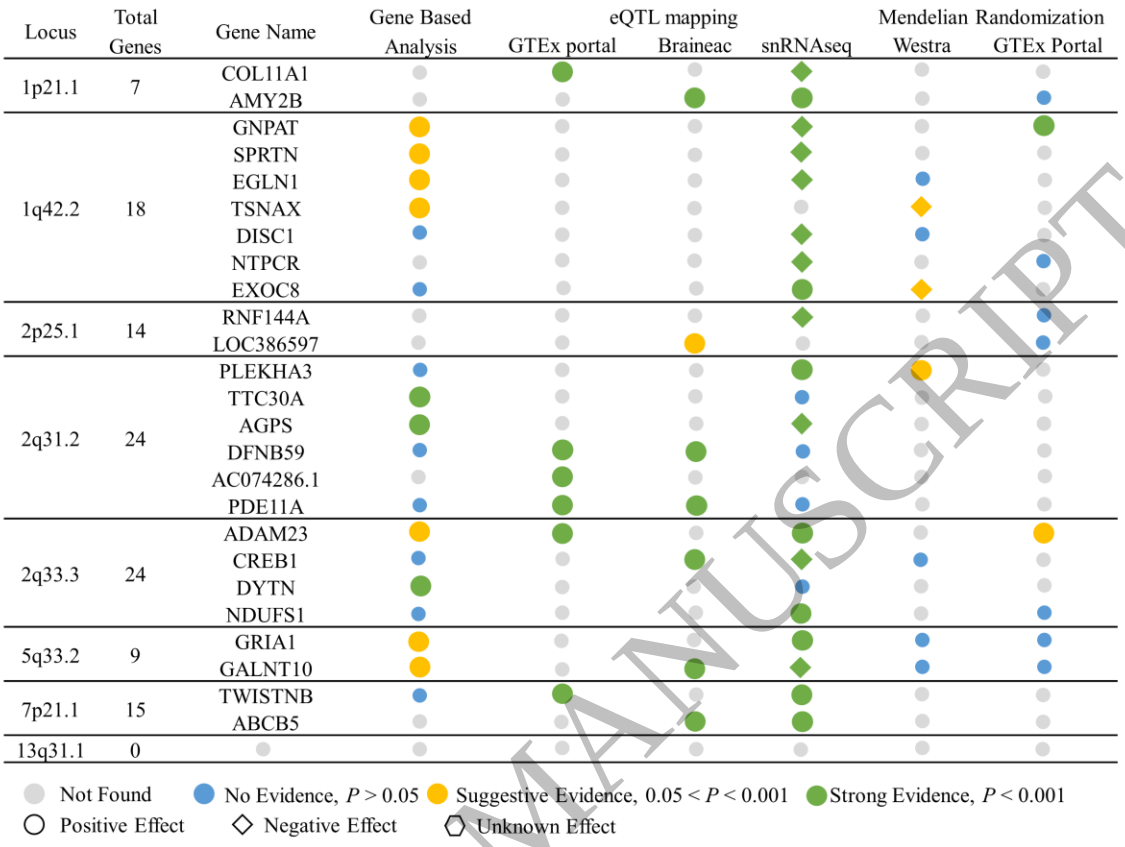


**Figure 3**
**165x119 mm (.15 x DPI)**