Single nucleotide variations in ZBTB46 are associated with post-thrombolytic parenchymal haematoma

Carrera, Caty

2021-08

Carrera , C , Carcel-Marquez , J , Cullell , N , Torres-Aguila , N , Muino , E , Castillo , J , Sobrino , T , Campos , F , Rodriguez-Castro , E , Llucia-Carol , L , Millan , M , Munoz-Narbona , L , Lopez-Cancio , E , Bustamante , A , Ribo , M , Alvarez-Sabin , J , Jimenez-Conde , J , Roquer , J , Giralt-Steinhauer , E , Soriano-Tarraga , C , Mola-Caminal , M , Vives-Bauza , C , Diaz Navarro , R , Tur , S , Obach , V , Francisco Arenillas , J , Segura , T , Serrano-Heras , G , Marti-Fabregas , J , Delgado-Mederos , R , Freijo-Guerrero , M M , Moniche , F , Antonio Cabezas , J , Castellanos , M , Gallego-Fabrega , C , Gonzalez-Sanchez , J , Krupinsky , J , Strbian , D , Tatlisumak , T , Thijs , V , Lemmens , R , Slowik , A , Pera , J , Kittner , S , Cole , J , Heitsch , L , Ibanez , L , Cruchaga , C , Lee , J-M , Montaner , J & Fernandez-Cadenas , I 2021 , ' Single nucleotide variations in ZBTB46 are associated with post-thrombolytic parenchymal haematoma ' , Brain : a journal of neurology , vol. 144 , pp. 2416-2426 . https://doi.org/10.1093/brain/awab090

http://hdl.handle.net/10138/341674
https://doi.org/10.1093/brain/awab090

acceptedVersion

Single nucleotide variations in *ZBTB46* are associated with post-thrombolytic parenchymal haematoma

Caty Carrera,[1,2] Jara Cárcel-Márquez,[2] Natalia Cullell,[2,3] Nuria Torres-Águila,[2,3] Elena Muiño,[2] José Castillo,[4] Tomás Sobrino,[4] Francisco Campos,[4] Emilio Rodríguez-Castro,[5] Laia Llucia-Carol,[2] Mònica Millán,[6] Lucía Muñoz-Narbona[6] Elena López-Cancio,[7] Alejandro Bustamante,[1] Marc Ribó,[8] José Álvarez-Sabín,[9] Jordi Jiménez-Conde,[10] Jaume Roquer,[10] Eva Giralt-Steinhauer,[10] Carolina Soriano-Tárraga,[10] Marina Mola-Caminal,[10] Cristófol Vives-Bauza,[11] Rosa Díaz Navarro,[12] Silvia Tur,[12] Victor Obach,[13] Juan Francisco Arenillas,[14] Tomás Segura,[15] Gemma Serrano-Heras,[16] Joan Martí-Fàbregas,[17] Raquel Delgado-Mederos,[17] M. Mar Freijo-Guerrero,[18] Francisco Moniche,[19] Juan Antonio Cabezas,[19] Mar Castellanos,[20] Cristina Gallego-Fabrega,[2,3] Jonathan González-Sanchez,[3,21] Jurek Krupinsky,[21,22] Daniel Strbian,[23] Turgut Tatlisumak,[24] Vincent Thijs,[25,26] Robin Lemmens,[27] Agnieszka Slowik,[28] Johanna Pera,[28] Steven Kittner,[29] John Cole,[29] Laura Heitsch,[30,31] Laura Ibañez,[32] Carlos Cruchaga,[32] Jin-Moo Lee,[31] Joan Montaner,[1,19] Israel Fernández-Cadenas,[2] on behalf of the International Stroke Genetic Consortium and the Spanish Stroke Genetic Consortium

ABSTRACT

Hemorrhagic transformation is a complication of recombinant tissue-plasminogen activator (rtPA) treatment. The most severe form, parenchymal hematoma, can result in neurological deterioration, disability, and death.

Our objective is to identify single nucleotide variations associated with a risk of parenchymal hematoma following thrombolytic therapy in acute ischemic stroke patients.

A fixed-effect genome-wide metanalysis was performed combining two-stage Genome Wide Association studies (GWAs) (n=1,904). The Discovery Stage (3 cohorts) comprised 1,324 ischemic stroke individuals, of whom 5.4% had a parenchymal hematoma. Genetic variants yielding a p-value $<1x10^{-5}$ were analyzed in the Validation Stage (6 cohorts), formed by 580 ischemic stroke patients with 12.1% hemorrhagic events. All the participants received rtPA; cases

were parenchymal hematoma type 1 or 2 as defined by the ECASS criteria. Genome-wide significant findings ($p<5x10^{-8}$) were characterized by in-silico functional annotation, gene expression, and DNA regulatory elements.

We analyzed 7,989,272 single nucleotide polymorphisms (SNPs) and identified a Genome-wide association locus on chromosome 20 in the Discovery Cohort; functional annotation indicated that the *ZBTB46* gene was driving the association for Chrosome 20. The top SNP was rs76484331 in the *ZBTB46* gene ($p=2.49x10^{-8}$; odds ratio (OR): 11.21; 95% confidence interval (CI): 4.82-26.55). In the Replication Cohort (n=580), the rs76484331 polymorphism was associated with parenchymal hematoma (p=0.01), and the overall association after meta-analysis increased ($p=1.61x10^{-8}$; OR: 5.84; 95%CI: 3.16-10.76). *ZBTB46* codes the Zinc Finger and BTB domain-containing protein 46 that acts as a transcription factor. In-silico studies indicated that *ZBTB46* is expressed in brain tissue by neurons and endothelial cells. Moreover, rs76484331 interacts with the promoter sites located at 20q13.

In conculsion, we identified single nucleotide variants in the *ZBTB46* gene associated with a higher risk of parenchymal hematoma following rtPA treatment.

Author affiliations:

1 Neurovascular Research Laboratory, VHIR, Universitat Autònoma de Barcelona, Barcelona, Spain

2 Stroke Pharmacogenomics and Genetics, IIB-Sant Pau, Barcelona, Spain

3 Stroke Pharmacogenomics and Genetics, Fundació Docència i Recerca Mútua Terrassa, Terrassa, Spain

4 Clinical Neurosciences Research Laboratory, IDIS, Santiago de Compostela, Spain

5 Department of Neurology, CHUS, Santiago de Compostela, Spain

6 Department of Neuroscience, HUGTP, Badalona, Spain

7 Stroke Unit, HUCA, Oviedo, Spain

8 Stroke Unit, HUVH, Barcelona, Spain

9 Department of Neurology, HUVH, UAB, Barcelona, Spain

10 Department of Neurology, Neurovascular Research Group, IMIM-Hospital del Mar, Barcelona, Spain
11 Neurobiology Laboratory, IdISPa, Mallorca, Spain
12 Department of Neurology, HUSE, Mallorca, Spain
13 Department of Neurology, Hospital Clínic i Provincial de Barcelona, Barcelona, Spain
14 Department of Neurology, Hospital Clínico Universitario, University of Valladolid, Valladolid, Spain
15 Department of Neurology, CHUA, Albacete, Spain
16 Experimental Research Unit, CHUA, Albacete, Spain
17 Department of Neurology, Hospital de la Santa Creu i Sant Pau, IIB-Sant Pau, Barcelona, Spain
18 Neurovascular Unit, Biocruces Bizkaia Health Research Institute, Bilbao, Spain
19 Department of Neurology, Virgen del Rocío, IBIS, Seville, Spain
20 Department of Neurology, CHUAC, A Coruña, Spain
21 School of Healthcare Science, Manchester Metropolitan University, Manchester, UK
22 Neurology Unit, Hospital Universitari Mútua Terrassa, Terrassa, Spain
23 Department of Neurology, Helsinki University Hospital, Helsinki, Finland
24 Sahlgrenska Academy at University of Gothenburg and Sahlgrenska University Hospital, Gothenburg, Sweden
25 Stroke Division, Florey Institute of Neuroscience and Mental Health, University of Melbourne, Heidelberg, Victoria, Australia
26 Department of Neurology, Austin Health, Heidelberg, Victoria, Australia
27 Department of Neurology, University Hospitals Leuven, Campus Gasthuisberg, Leuven, Belgium
28 Department of Neurology, Jagiellonian University Medical College, Kraków, Poland
29 Department of Neurology, University of Maryland School of Medicine and Baltimore, Baltimore, USA
30 Division of Emergency Medicine, Washington University School of Medicine, St. Louis, USA
31 Department of Neurology, Washington University School of Medicine, St. Louis, USA
32 Department of Psychiatry, Washington University School of Medicine, St. Louis, USA

Correspondence to: Israel Fernández-Cadenas
Stroke Pharmacogenomics and Genetics. IIB-Sant Pau
c/ Sant Antoni Mª Claret, 167, 08025, Barcelona, Spain
E-mail: israelcadenas@yahoo.es

Running head: *ZBTB46* and parenchymal haematoma risk

Keywords: GWAs; ischemic stroke; pharmacogenetics; parenchymal hematoma; thrombolysis

INTRODUCTION

Despite the effectiveness of thrombolytic recombinant tissue-plasminogen activator (rtPA) treatment for acute ischemic stroke (AIS), a six- to seven-fold increased risk of intracerebral hemorrhage remains a serious therapeutic limitation[1]. The most severe presentation of rtPA-induced intracerebral bleeding – parenchymal hematoma (PH) – worsens the patient's outcome and increases the risk of 3-month mortality from 2.4 to 4.5 times[2].

The damageon the blood vessel caused by ischemia and reperfusion[3], the breakdown of the blood-brain barrier[4], and the coagulopathy produced by rtPA[5] result in blood extravasation within the ischemic core. Despite the efforts to identify risk factors, such as older age, stroke severity, high blood pressure, circulating glucose levels, leukoaraiosis, and prediction scores[2,6,7], 6% of patients still suffer a PH after thrombolysis[8] .

Part of the inter-individual variability in the response to AIS treatment could be explained by the genetic architecture of the patients. Previous results used a candidate gene approach that recognized the polymorphisms involved in the risk of PH, which included, but were not limited to: Alpha-2-macroglobulin (*A2M*), Coagulation Factor XII (*FXII*)[6], and Survivin (*BIRC5*)[9]. However,the results of some of these findings need to be confirmed by replication in larger populations. Nowadays, no genome-wide association studies (GWAs) have been published in this field. The hypothesis-free method has been able to find and replicate genetic variants and

biological pathways associated with the risk of stroke[10], and a similar success is expected when studying AIS phenotypes, such as PH.

This study aimed to identify polymorphisms associated with PH in AIS patients undergoing thrombolytic therapy through a GWAs approach.

SUBJECTS/MATERIALS AND METHODS

Study population

The study consisted of AIS patients admitted to an Emergency Department and treated with intravenous rtPA. Thrombolysis was performed within 4.5 hours of symptom onset at a standard dose of 0.9 mg/kg/dose (10% bolus and 90% continuous 1-hour infusion). Eligible participants were at least 18 years-old, who suffered a neurological deficit, diagnosed by an expert neurologist, and confirmed by neuroimaging. The exclusion criteria were: patients with remote parenchymal hematoma (as different physiological mechanisms may be involved in this condition, e.g. cerebral microbleeds, amyloid angiopathy); cases in which information on the presence of hemorrhagic transformation (HT) was not available; onset to treatment time over 4.5 hours; and patients who had undergone endovascular therapy. Furthermore, an extreme phenotype approach was implemented, then we excluded hemorrhagic infarctions (HI) as previous studies did not show an association of hemorrhagic infarction (HI) types 1 and 2 with worsened outcomes[11].

The Discovery Cohort (n=3,217) consisted of AIS patients recruited via hospital-based studies between 2003 and 2017. The participants were part of the Genetics of Early Neurological Instability After Ischemic Stroke (GENISIS), Genetic contribution to Functional Outcome and Disability after Stroke (GODS) and the Genotyping Recurrence Risk of Stroke (GRECOS) studies. Participants in the Replication Stage (n=1172) were enrolled through the effort of collaborative networks: The International Stroke Genetics Consortium (ISGC), the Spanish Stroke Genetics Consortium (Genestroke), the Spanish Stroke Research Network (INVICTUS plus), the Stroke Genetics Network (SiGN), the Genetic Study in Ischemic Stroke Patients treated with tPA (GenoTPA), BAse de Datos de ICtus del hospital del MAR (BASICMAR) [*Stroke database of the*

Hospital del Mar], Leuven Stroke Genetics Study (LSGS), Helsinki 2000 Ischemic Stroke Genetics Study and GENISIS study. In particular, the Replication Study was intentionally strengthened with PH cases in order to ensure a sufficient number for conducting studies (please see "Online Methods" for more detailed information about these studies and patient recruitment).

Clinical Protocol

Demographic data, past medical history, cardiovascular risk factors, clinical examination, stroke severity assessed with the National Institutes of Health Stroke Score (NIHSS) at initial evaluation, and treatment decisions were retrieved from the medical records. CT scans were obtained prior to thrombolytic administration (baseline), and 24 hours after symptom onset (follow-up) or whenever a neurological deterioration was detected. Neurological deterioration was defined as an increase of at least 4 points in the NIHSS score. All brain images were reviewed by a radiologist or neuro-radiologist. The radiological, clinical and genetic evaluations were mutually blinded.

Phenotype definition

The presence of HT was assessed in the follow-up CT scan and radiologically-classified according to the European Cooperative Acute Stroke Study (ECASS) criteria[12] into HI types 1 and 2 and PH types 1 and 2. PH was defined as bleeding in less than 30 percent of the infarcted area with mild space-occupying effect (PH-1) or hematoma over 30 percent of this area and a significant mass effect (PH-2). Symptomatic Intracerebral Hemorrhage (sICH) was defined according to the SITS-MOST criteria as local PH-2 on the CT brain scan combined with an increase of ≥ 4 points in the NIHSS score within 22-36 hours of treatment[13].

Standard protocol approvals, registrations, and patient consents

An Institutional Review Board or Ethics Committee approved the study at each participating site. All patients or their relatives provided written informed consent.

Genome-wide genotyping and imputation

DNA samples were genotyped on commercial arrays from Illumina (San Diego, CA) (Table 1). Stringent quality controls were performed (Online Methods), such as removing genetic variants based on the genotyping call rate (<97%), minor allele frequency (MAF <1%), significant deviations from the Hardy-Weinberg equilibrium ($p=1x10^{-6}$), and Identity by State to analyze relatedness among subjects. Studies genotyped on the same platforms were combined with the exception of the Finnish participants, resulting in one stratum for the Discovery and 3 strata for the Replication Stages (Online Figure I). Quality controls were applied yet again[14], then phase haplotypes by stratum were estimated and imputation was performed using the Michigan Imputation Server Portal[15] based on the 1000 Genomes Phase 3v5 panel. The Finnish cohort was imputed separately due to their genetic background, which differs from other European populations.

Association analysis and meta-analysis

We performed a Discovery case-control association analysis. Frequentist association test for each allele were performed in SNPTEST v2.5.2[16], under an additive genetic model adjusted for age, sex, three principal components (PCs) (Online Figure II), and clinical variables associated with PH after a logistic regression: baseline NIHSS and Diabetes ($p<0.05$). Polymorphisms that achieved genome-wide significance and nominal significance, defined as $p<5x10^{-8}$ and $p<1x10^{-5}$, respectively, were selected for replication in new independent cohorts. Furthermore, to obtain enough cases, all participants of the Replication Study were combined into a single dataset. Any possible bias created by population substructure was then checked through analysis with and without adjustment for the PCs.

Subsequently, we examined the SNPs identified during the first stage using an association study adjusted for age, sex, baseline NIHSS, genotyping platform and 10 PCs (Online Figure III). A different number of PCs compared with the Discovery Cohort was required due to the different genetic background of the populations studied. In order to homogenize the Replication Cohort (Online Figure III) we used this different number of PCs. This difference was mainly due to the Finnish population. Lastly, we conducted a fixed effect inverse-variance meta-analysis using

METAL[17]. Standard error and genomic control options were applied. The genome-wide significance was set at $p < 5 \times 10^{-8}$. In addition, we performed gene-based analysis using the Multi-marker Analysis of GenoMic Annotation (MAGMA)[18]; the study included variants located within 2 KB of the gene at 3' and 5'UTR. Additionally we ran S-MultiXcan[19], the multi-tissue tool of the S-PrediXcan software to confirm the MAGMA findings.

We performed GNOVA[20] analysis to evaluate the genetic architecture shared with other diseases. We tested genetic correlation using MEGASTROKE European analysis data[10] for the ischemic stroke phenotypes. The sample sizes of the stroke subtypes were: for ischemic stroke (n=440,328), large artery-atherosclerosis stroke (n=301,663), cardioembolic stroke (n=362,661), and small vessel stroke (n=348,946). For Intracerebral Hemorrhages (ICH) we used the summary statistics available from the ICH GWAs 2014[21] (n=6,965) and its lobar (n= 1,148) and deep (n= 2,075) ICH subtypes. Additionally, we evaluated genetic correlation for the white matter hyperintensity volume (WMHv) phenotype (n=11,226) using published data[22]. All datasets were obtained from http://cerebrovascularportal.org.

We generated a Polygenic Risk Score (PRS) with the data from our study. We used the PRSice-2 software that runs logistic regressions to determine the p-threshold with the largest variance explained by the PRS, assessed as the increment in Nagelkerke's pseudo-R2.

Bioinformatic functional analysis

After meta-analysis, novel SNPs significantly associated with PH were examined to identify their biological function using publicly available online bioinformatic tools. We tested tissue-specific expression for the polymorphism-containing gene where the polymorphism is located through the Genotype-Tissue Expression (GTEx) Project[23], BRAINEAC dataset[24] and by mapping the gene expression in brain-tissue using the single-nuclei RNA-sequencing expression browser[25]. In addition, we checked the effect of the top genetic variant over the gene expression levels in a 1MB window (eQTL, expression quantitative trait loci) in different tissues using the Summary data-based Mendelian Randomization and the Heterogeneity in Dependent Instruments Test (SMR & HEIDI)[26], the eQTL consortium[27], and the GTEx project[23]. An extended eQTL analysis in vascular and brain tissues was performed as well. Interactions between the top SNPs with distal DNA

regulatory elements and gene promoter sites were evaluated by Capture HiC Plotter[28] in the lymphoblastoid cell line GM12878 and by the RegulomeDB database[29]. Furthermore, previous associations of the most significant SNP and proxies (r^2>0.6) with stroke, cardiovascular traits, and any genome-wide association studies were investigated with the Cerebrovascular disease knowledge portal[30] and the PhenoScanner[31] web tool.

Statistical Analyses

Statistical Analyses for clinical variables were performed using SPSS statistical package version 17.0 (IBM, Chicago, US). Univariate analysis for case-controls was evaluated by $\chi 2$ or Fisher's exact test. T-test, Mann-Whitney U or Kruskal-Wallis tests were used for continuous variables. Logistic regression was conducted with the forward-stepwise method to select clinical variables, such as covariates in the association studies. Meta-analysis heterogeneity was calculated using Cochran's Q-test.

Data availability

Any qualified investigator may request the summary statistics and dataset.

Detailed methods are available in the "Supplementary Material" section.

RESULTS

Genome-Wide Association Analysis and Meta-analysis

After quality controls and imputation, 1,324 AIS patients treated with rt-PA met the inclusion criteria and 7,989,272 polymorphisms were tested (Online Figure I). PH occurred in 5.4% (n=71) (Table 1), 2.4% (n=32) had PH-1 and 2.9% (n=39) had PH-2. Moreover, the incidence of sICH was 1.4% (n=18). The sICH was associated with in-hospital mortality, disability and mortality at 3 months (p<0.001); significantly, PH-1 and PH-2 hemorrhages excluding sICH were also

associated with disability (p=0.007; OR: 7.19; 95%CI: 3.18-16.25) and mortality at 3 months ($p=4.48x10^{-8}$; OR: 6.05; 95%CI: 3.22-11.36) (Online Table I). The univariate analysis showed that Diabetes (DM), cardioembolism etiology (CE) and baseline NIHSS appeared to be associated with PH risk (p=0.02; p=0.04; $p=1.79x10^{-10}$). After logistic regression, only DM and baseline NIHSS were significantly associated with PH and used for the GWAs adjustment (Online Table II). The clinical characteristics are summarized in Table 2.

The Manhattan plot obtained is presented in Online Figure IV. We did not observe overall inflation of p-values (genomic inflation factor λ =1.018) (Online Figure IV). The Discovery analysis identified one polymorphism that reached genome-wide significance: rs76484331 located on chromosome 20, in intron 1 of Zinc Finger and BTB Domain-Containing protein 46 (*ZBTB46)*, ($p=2.49x10^{-8}$; OR: 11.31; 95%CI:4.82-26.55). Furthermore, 182 SNPs were selected for the follow-up stage based on the p-value cut-off $1x10^{-5}$, including 43 polymorphisms at *ZBTB46* gene (Online Table III).

Replication was performed in 580 AIS patients. The prevalence of PH was 12.1% (n=70); PH-1 developed in 6.4% (n=37) and PH-2 in 5.7% (n=33). The higher occurrence of bleeding events can be attributed to the Genot-PA project, a study designed to understand the genetic variability of thrombolytic response. The number of PH samples included in this study was larger than expected. Age, sex and NIHSS were required as minimum clinical covariates (Table 3). CE stroke was not included as a covariate because it was not significant after logistic regression. Moreover, 10 PCs accounted for the GWAs population substructure adjustment. The genome-wide significant SNP rs76484331 ($p<5x10^{-8}$) selected from the Discovery Phase was significant in the Validation Cohort (p=0.01). When we included CE stroke as a covariable the resulst did not change significantly (Only Table IV). Additionally, we analyzed 182 polymorphisms with nominally significant p-values ($p<1x10^{-5}$), obtaining the lowest p-value at rs1962779 on chromosome 7 ($p=3.34x10^{-3}$). Three of these SNPs could not be included in the analysis because of failed imputation quality controls (Online Table III).

The meta-analysis of both stages revealed one locus on chromosome 20q13 that reached genome-wide significance in association with PH: rs76484331 ($p=1.61x10^{-8}$; OR: 5.84, 95%CI: 3.16-10.76; heterogeneity p=0.03), formed by 43 polymorphisms that exceeded the nominal significance cut-

off ($p<1x10^{-5}$) (Figure 1). Moreover, we found 5 loci (overall 45 SNPs) in Syntrophin Gamma 2 *(SNTG2*), RUN and FYVE Domain-Containing protein 4 (*RUFY4*), Semaphorin-3A (*SEMA3A)*, Down Syndrome Cell Adhesion Molecule-Like Protein 1 (*DSCAML1)*, and PDZ Domain-Containing Ring Finger 4 (*PDZRN4)* that reached p-values lower than $1x10^{-5}$. All the top SNPs had a consistent direction of effect (Table 4); the best association obtained was for *SEMA3A,* which addressed a p-value close to the significance cut-off point ($p=7.85x10^{-8}$; OR: 2.43; 95%CI: 1.76-3.37; heterogeneity p=0.25).

On the other hand, the gene-based analysis computed one statistically significant gene associated with our phenotype: Chromosome 20 Open Reading Frame 181 (*C20orf181;* p=0.05/18647; $p <2.68x10^{-6}$) (Online Table V), located at 20q13.33. In addition, *ZBTB46* showed a p-value of 0.0035. S-MultiXcan using GTEx [23] v8 data confirmed the results for *ZBTB46* ($p=3.16x10^{-4}$), however the C20orf181 transcript was not covered by the 22,313 transcripts evaluated with S-MultiXcan and it could not be analyzed (Online Table VI).

Genetic correlation analysis revealed a shared genetic background of PH and deep ICH, lobar ICH and WMH ($p<9.2 \times 10^{-3}$) (Online Table VII). After Bonferroni correction, only lobar ICH was significantly associated with PH.

The best-fit for the PRS estimation of disability after three months using the genetic data from the Discovery Cohort was observed for a threshold of p=0.00530005, pseudo-R2= 0.0684721, and composed of 3,506 SNPs. A significant association was observed with PH and mortality after 3 months, but no association was found with in-hospital mortality (Online Table VIII, Online Table IX). The most significant association was with disability after three months (mRS 0-2 vs. 3-6; $p=1.5 \times 10^{-06}$). Interestingly, in a multivariable logistic regression for disability after three months the PRS remained significant in the logistic regression after inclusion of the clinical variables: baseline NIHSS, sex and age (Online Table X).

Functional analysis

The GTEx and BRAINEAC portal revealed gene expression of *ZBTB46* in different tissues, with the highest expression detected in the brain, specifically in the region of the cerebellum (Online Figure V, Online Figure VI). Besides, single-nuclei RNA-sequencing confirmed *ZBTB46* RNA-

expression in neurons and endothelial cells (Figure 2). Furthermore, the eQTL study showed a nominal association of rs76484331 with expression levels of *ZBTB46* in blood tissue ($p=2.2x10^{-6}$) (Online Table XI); similarly, interactions between rs76484331 and promoter sites at the *ZBTB46* gene were evidenced by Capture HiC Plotter (Online Figure VII). The RegulomeDB classified rs76484331 as transcription factor-binding or DNase peak, chromatin states included quiescent action in blood cells and weak transcription in the brain (Online Table XII). However, rs76484331 did not affect a single gene exclusively. We observed additional cis-eQTL and promoter sites located in the near window of 1MB associated with the presence of the polymorphism including Lck Interacting Transmembrane Adaptor 1 (*LIME1)* and DnaJ Heat Shock Protein Family (Hsp40) Member C5 *(DNAJC5)*, further interactions are detailed in Online Table XIII and Online Figure VII.

Traits that have been reported as being associated with rs76484331 include cardiomyopathy, blood pressure, and coronary atherosclerosis. Moreover, using the Cerebrovascular disease knowledge portal (https://cerebrovascularportal.org/), we could not find any association with hemorrhagic stroke and ischemic stroke subtypes (Online Table XIV). Additionally, we did not find evidence of association within rs76484331 and hypertension or blood pressure levels in our cohort (Online Table XV and XVI).

DISCUSSION

This study found a genome-wide polymorphism associated with PH in an intron of the Zinc Finger and BTB Domain-Containing protein 46 (*ZBTB46*). The SNP, rs7648433, was initially identified in 1,324 patients and further validation was sought in 580 participants; both stages included AIS treated with intravenous thrombolysis.

ZBTB46 is a member of the Poxvirus and Zinc Finger and Krüppel-type (POK) protein family that contains a POZ domain for protein-protein interactions and zinc fingers for DNA binding[32]. The POK proteins act as transcription factors, facilitating the recruitment of corepressors to promoter

regions[33] and play a role in the development of hematopoietic[34], dendritic[35], and endothelial cells[35], and in lymphocyte differentiation during the immune response[36]. No studies have associated *ZBTB46* with ischemic or hemorrhagic stroke; however, involvement in stroke mechanisms, such as shear stress and atherosclerosis, has been reported previously[37]. Moreover, *ZBTB46* was also linked with prostate cancer[38] and multiple sclerosis[39].

The polymorphisms identified lay within intronic regions, portions of gene that are spliced out prior to protein translation. Introns are part of the non-coding variants that represent over 90% of all the hits identified by GWAs in complex diseases. Finding the causality of these SNPs could be challenging; however, these genome regions may regulate gene expression[40]. We tested whether the GWAs significant polymorphism could influence other neighboring genes using in-silico analysis tools. Considering the importance of the integrity of the blood-brain barrier in the pathophysiology of hemorrhagic events after thrombolysis, we have emphasized the SNP effect on gene expression in brain, blood and blood vessel tissues.

We found a nominal association between rs7648433 and *ZBTB46* gene expression reported in arterial vessel tissue; in addition, *ZBTB46* RNA expression was identified in endothelial cells and neurons in brain tissue. Moreover, cis-eQTL of rs7648433 within a neighboring region of 1MB identified six differential gene expressions in blood tissue. None of the cis-eQTL genes identified have been previously associated with cerebrovascular traits.

Regulation of gene expression by *ZBTB46* is mediated through binding of the zinc finger domain to a chromatin complex[41]. The POK proteins act over a large and assorted group of genes which includes transcription factors, RNA processing factors, chromatin regulators, kinases, peptidases, ubiquitin ligases, and phosphatases[42]. The current study exhibited an interaction among rs7648433, and several promoter sites located in chromosome 20, including *ZBTB46.* Moreover, the gene-based analysis identified the open read frame *C20orf181*, located at 231pb from the upstream region of *ZBTB46*; open read frames can mediate the protein synthesis through the translation regulation of genes[43].

Our results suggest the modulation of genetic targets affects the PH risk and could influence the response to rtPA therapy. On the other hand, rs76484331 has been associated with

cardiomyopathies and systolic blood pressure levels in the United Kingdom Biobank (UKBB) project. Polymorphism pleiotropy could highlight the biological effect of the variant; however, replications should be performed in independent cohorts and a secondary analysis is needed in order to clarify this. In addition, *ZBTB46* plays a role in keeping cells in a quiescent state (G0-G1)[37] and could be activated through several proteins or events like the Toll-like receptor proteins (TLR)[42] or disturbed vessel flow[37]. Little is known about how *ZBTB46* responds under ischemic conditions and with the data available, we are not able to determine which molecular process is modified in the PH risk by rs76484331 or the Zinc Finger and BTB domain-containing protein 46.

Our study also revealed 5 loci that reached a nominal association with PH; the second most significant locus, Semaphorin-3A (*SEMA3A),* is related to the vascular permeability of the blood-brain barrier and brain damage after cerebral ischemia in murine models[44]. Moreover, *SEMA3A* is expressed by the ischemic brain and the ischemic core during reperfusion[45]. It seems that the genetics of Semaphorin-3A could modulate the response to ischemia and promote the onset of bleeding events. Studies with larger sample sizes are required to determine whether this GWAs hit is associated with the onst of PH.

Interestingly, genetic correlation analysis revelead a shared genetic correlation of PH with several traits, such as lobar and deep ICH or WMH. These findings indicate that ICH genetic risk factors could play a role in the risk of PH, however further studies are needed to confirm this hypothesis.

The consequences of sICH on the stroke outcome at 3 months and in-hospital mortality[13] are well established. In our cohort the occurrence of PH-1 and non-symptomatic PH-2 also had a damaging effect on disability after stroke. In view of this, efforts to translate causal variants and genes to clinical practice should be encouraged. This is supported by our results with the PRS generated using the data from the genetic analysis of PH. The PRS was associated in the independent Replication Cohort with PH, disability and mortality after three months. In the case of disability it remains significantly associated after logistic regression indicating a potential use in clinical practice as a possible biomarker or as something to be explored in further studies looking for drug targets.

The GWAs findings could feasibly allow the identification of new drug targets or the repurposing of existing drug components used for other diseases that address the newly-identified causal genes or pathways; for example, the identification of *IL23R* polymorphisms associated with ankylosing spondylitis[46] generated the repositioning of secukinumab, an anti-IL-17A monoclonal antibody involved in the *IL23* pathway. Secukinumab is widely used in the treatment of psoriasis and psoriatic arthritis and is currently used succesfully in ankylosing spondylitis too[47]. Another strategy for translating discoveries from genome-wide association studies includes the development of tools (e.g. risk scores) for measuring the individual predisposition to a trait. The Genot-PA score, based on clinical characteristics and two polymorphisms, demonstrated its prediction capability of bleeding events in patients undergoing thrombolysis alone or in combination with mechanical thrombectomy[6,48]. The implementation of predictive scores could help physicians in decision-making, avoiding treatment delays, adjusting the drug dose, or implementing additional therapies.

We performed the first genome-wide association study of the hemorrhagic transformation risk after thrombolysis, based on a relatively small sample size in comparison with large-scale GWAs. To address this issue, petechial bleeding cases were excluded, thus through an extreme phenotyping study approach, we aimed at increasing the variant effect sizes[49] to detect PH-associated polymorphisms. Furthermore, in order to reduce false-positive results due to low-frequency variants, we removed extremely large beta estimates[50], applied a stringent filter of MAF and evaluated the presence of alelles in order to guarantee a total of at least 25 alleles[51] in each stage.

The limitations of the study included not being able to analyze the Finnish cohort because of the small number of cases available. The Finnish population is considered genetically isolated[52] and its genetic population structure could generate a bias; therefore, we attempted to reduce this limitation by including a sufficient number of PCs in the association study[53]. Based on isolation by distance, we used 10 eigenvectors in our analysis. Another limitation is the absence of patients treated with mechanical thrombectomy, a treatment that is currently available for certain AIS patients[54]. Moreover, our results cannot be extrapolated to non-European populations. Further studies are required to validate our findings in diverse populations and patients who underwent mechanical recanalization.

In summary, our findings identified a previously unreported polymorphism in the *ZBTB46* gene associated with the development of parenchymal hematoma following rtPA treatment. Besides, suggestive loci were identified, which require confirmation in future independent studies. Further functional studies are required to clarify the genetic mechanism involved in the bleeding risk of thrombolysis.

ACKNOWLEDGEMENTS

We are grateful to Lucía Muñoz (Hospital Germans Trias i Pujol), Anna Penalba (Vall d'Hebron Research Institute), Uxue Lazkano (IMIM-Hospital del Mar), Carmen Jimenez (Hospital Universitari Son Espases), Elisa Cortijo (Hospital Clínico Universitario), Rebeca Marín Bueno (Hospital de la Santa Creu i Sant Pau, IIB-Sant Pau), Esther Sarasola Diez (Hospital de Basurto), Carmen Gubern (Doctor Josep Trueta University Hospital), Aki Havulinna (Institute for Molecular Medicine Finland), Veikko Salomaa (Institute for Molecular Medicine Finland), and Antoni Ferens (Jagiellonian University), for their contribution to patient recruitment; and to Agustin Ruiz and Oscar Sotolongo (Fundació ACE) for their technical support.

FUNDING

The GENISIS study is funded by the National Institute of Health (K23 NS099487, and R01NIH NS085419). The Neurovascular Research Laboratory is supported by the Spanish stroke research network (INVICTUS plus). The Stroke Pharmacogenomics and Genetics Laboratory is supported by the Spanish stroke research network (INVICTUS plus); the Generation Project (PI15/01978) and the Pre-Test Stroke Project (PMP15/00022) are funded by the Instituto de Salud Carlos III and Fondo Europeo de Desarrollo Regional (ISCIII-FEDER). The Neurovascular Research Group, IMIM is supported by INVICTUS plus (RD16/0019/0002). The Biomedical Research Institute Hospital de la Santa Creu i Sant Pau, IIB Sant Pau is supported by INVICTUS plus (RD16/0019/0010). Ibiostroke (AC19/00106), Maestro Project (PI18/01338) is supported by the Instituto de Salud Carlos III and Fondo Europeo de Desarrollo Regional (ISCIII-FEDER) and the Epigenisis project is supported by Marató TV3 (201711.30).

I. Fernandez (CPII17/00021), T. Sobrino (CPII17/00027) and F. Campos (CPII19/00020) are recipients of a research contract from the Miguel Servet Program from the Instituto de Salud Carlos III. A. Bustamante is supported by a Juan Rodes research contract from the Carlos III Health Institute (JR16/00008). J. Cárcel-Márquez is supported by an AGAUR Contract (agència de gestió d'ajuts universitaris i de recerca; FI_DGR 2019, grant number 2019_FI_B 00853) co-financed with Fons Social Europeu (FSE).

The Genotype-Tissue Expression (GTEx) Project was 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. The data used for the analyses described in this manuscript were obtained from the GTEx Portal on 06/20/2019. Schofield EC, Carver T, Achuthan P, Freire-Pritchett P, Spivakov M, Todd JA, Burren OS. CHiCP: a web-based tool for the integrative and interactive visualization of promoter capture Hi-C datasets. Bioinformatics. (2016) 15:32(16):2511-3. Accessed on 10/29/2019. Cerebrovascular Disease Knowledge Portal, NINDS grant # 1R24NS09*f*2983. Accessed on 06/20/2019.

CONFLICT OF INTEREST

The authors report no conflict of interest.

REFERENCES

1. Whiteley WN, Emberson J, Lees KR, et al. Risk of intracerebral haemorrhage with alteplase after acute ischaemic stroke: a secondary analysis of an individual patient data meta-analysis. *Lancet Neurol*. 2016;15(9):925-933. doi:10.1016/S1474-4422(16)30076-X

2. Strbian D, Sairanen T, Meretoja A, et al. Patient outcomes from symptomatic intracerebral hemorrhage after stroke thrombolysis. *Neurology*. 2011;77(4):341-348. doi:10.1212/WNL.0b013e3182267b8c

3. Khatri R, McKinney AM, Swenson B, Janardhan V. Blood-brain barrier, reperfusion injury, and hemorrhagic transformation in acute ischemic stroke. *Neurology*. 2012;79(13 SUPPL.

1). doi:10.1212/WNL.0b013e3182697e70

4. Richard Leigh M, Shyian S. Jen M, Argye E. Hillis M, John W. Krakauer M, Peter B. Barker DpM, Imaging on behalf of the S and V. Pretreatment Blood Brain Barrier Damage and Post Treatment Intracranial Hemorrhage in Patients Receiving IV tPA. *Bone*. 2014;23(1):1-7. doi:10.1161/STROKEAHA.114.005249.Pretreatment

5. Panni P, Gory B, Xie Y, et al. Acute Stroke With Large Ischemic Core Treated by Thrombectomy: Predictors of Good Outcome and Mortality. *Stroke*. 2019;50(5):1164-1171. doi:10.1161/STROKEAHA.118.024295

6. Del Río-Espínola A, Fernández-Cadenas I, Giralt D, et al. A predictive clinical-genetic model of tissue plasminogen activator response in acute ischemic stroke. *Ann Neurol*. 2012;72(5):716-729. doi:10.1002/ana.23664

7. Lou M, Safdar A, Mehdiratta M, et al. The HAT Score: A simple grading scale for predicting hemorrhage after thrombolysis. *Neurology*. 2008;71(18):1417-1423. doi:10.1212/01.wnl.0000330297.58334.dd

8. Team SIC. JStroke Update Conference 11-13 November 2018 | SITS International. https://sitsinternational.azurewebsites.net/news/join-eso-karolinska-stroke-update-conference-11-13-november-2018/. Accessed February 11, 2021.

9. Mallolas J, Rodríguez R, Gubern C, Camós S, Serena J, Castellanos M. A Polymorphism in the Promoter Region of the Survivin Gene is Related to Hemorrhagic Transformation in Patients with Acute Ischemic Stroke. *NeuroMolecular Med*. 2014;16(4):856-861. doi:10.1007/s12017-014-8333-7

10. 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):524-537. doi:10.1038/s41588-018-0058-3

11. Molina CA, Alvarez-Sabín J, Montaner J, et al. Thrombolysis-related hemorrhagic infarction: A marker of early reperfusion, reduced infarct size, and improved outcome in patients with proximal middle cerebral artery occlusion. *Stroke*. 2002;33(6):1551-1556. doi:10.1161/01.STR.0000016323.13456.E5

12. Larrue V, Von Kummer R, Del Zoppo G, Bluhmki E. Hemorrhagic transformation in acute

ischemic stroke: Potential contributing factors in the European Cooperative Acute Stroke Study. *Stroke*. 1997;28(5):957-960. doi:10.1161/01.STR.28.5.957

13. Wahlgren N, Ahmed N, Dávalos A, et al. Thrombolysis with alteplase for acute ischaemic stroke in the Safe Implementation of Thrombolysis in Stroke-Monitoring Study (SITS-MOST): an observational study. *Lancet*. 2007;369(9558):275-282. doi:10.1016/S0140-6736(07)60149-4

14. Rebecca L. Zuvich1, Loren L. Armstrong2, Suzette J. Bielinski3, Yuki Bradford, Carlson CS, Crawford DC, Crenshaw AT, et al. Pitfalls of Merging GWAS Data: Lessons Learned in the eMERGE Network and Quality Control Procedures to Maintain High Data Quality. *Genet Epidemiol*. 2011;35(8):887-898. doi:10.1002/gepi.20639.Pitfalls

15. Das S, Forer L, Schönherr S, et al. Next-generation genotype imputation service and methods. *Nat Genet*. 2016;48(10):1284-1287. doi:10.1038/ng.3656

16. Marchini J, Howie B, Myers S, McVean G, Donnelly P. A new multipoint method for genome-wide association studies by imputation of genotypes. *Nat Genet*. 2007;39(7):906-913. doi:10.1038/ng2088

17. Willer CJ, Li Y, Abecasis GR. METAL: Fast and efficient meta-analysis of genomewide association scans. *Bioinformatics*. 2010;26(17):2190-2191. doi:10.1093/bioinformatics/btq340

18. de Leeuw CA, Mooij JM, Heskes T, Posthuma D. MAGMA: Generalized Gene-Set Analysis of GWAS Data. *PLoS Comput Biol*. 2015;11(4). doi:10.1371/journal.pcbi.1004219

19. Barbeira AN, Pividori MD, Zheng J, Wheeler HE, Nicolae DL, Im HK. Integrating predicted transcriptome from multiple tissues improves association detection. *PLoS Genet*. 2019;15(1). doi:10.1371/journal.pgen.1007889

20. 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):939-964. doi:10.1016/j.ajhg.2017.11.001

21. Woo D, Falcone GJ, Devan WJ, et al. Meta-analysis of genome-wide association studies identifies 1q22 as a susceptibility locus for intracerebral hemorrhage. *Am J Hum Genet*. 2014;94(4):511-521. doi:10.1016/j.ajhg.2014.02.012

22. Traylor M, Tozer DJ, Croall ID, et al. Genetic variation in PLEKHG1 is associated with white matter hyperintensities (n = 11,226). *Neurology*. 2019;92(8):E749-E757.

doi:10.1212/WNL.0000000000006952

23. Aguet F, Brown AA, Castel SE, et al. Genetic effects on gene expression across human tissues. *Nature*. 2017;550(7675):204-213. doi:10.1038/nature24277

24. Ramasamy A, Trabzuni D, Guelfi S, et al. Genetic variability in the regulation of gene expression in ten regions of the human brain. *Nat Neurosci*. 2014;17(10):1418-1428. doi:10.1038/nn.3801

25. Del-Aguila JL, Li Z, Dube U, et al. A single-nuclei RNA sequencing study of Mendelian and sporadic AD in the human brain. *bioRxiv*. 2019:1-16. doi:10.1101/593756

26. 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):481-487. doi:10.1038/ng.3538

27. Võsa U, Claringbould A, Westra H-J, et al. Unraveling the polygenic architecture of complex traits using blood eQTL metaanalysis. *bioRxiv*. October 2018:447367. doi:10.1101/447367

28. Schofield EC, Carver T, Achuthan P, et al. CHiCP: A web-based tool for the integrative and interactive visualization of promoter capture Hi-C datasets. *Bioinformatics*. 2016;32(16):2511-2513. doi:10.1093/bioinformatics/btw173

29. Boyle AP, Hong EL, Hariharan M, et al. Annotation of functional variation in personal genomes using RegulomeDB. *Genome Res*. 2012;22(9):1790-1797. doi:10.1101/gr.137323.112

30. Crawford KM, Gallego-Fabrega C, Kourkoulis C, et al. Cerebrovascular disease knowledge portal an open-access data resource to accelerate genomic discoveries in stroke. *Stroke*. 2018;49(2):470-475. doi:10.1161/STROKEAHA.117.018922

31. Kamat MA, Blackshaw JA, Young R, et al. PhenoScanner V2: an expanded tool for searching human genotype-phenotype associations. *Bioinformatics*. 2019;35(22):4851-4853. doi:10.1093/bioinformatics/btz469

32. Lee SU, Maeda T. POK/ZBTB proteins: An emerging family of proteins that regulate lymphoid development and function. *Immunol Rev*. 2012;247(1):107-119. doi:10.1111/j.1600-065X.2012.01116.x

33. Perissi V, Jepsen K, Glass CK, Rosenfeld MG. Deconstructing repression: Evolving

models of co-repressor action. *Nat Rev Genet*. 2010;11(2):109-123. doi:10.1038/nrg2736

34. Bilic I, Ellmeier W. The role of BTB domain-containing zinc finger proteins in T cell development and function. *Immunol Lett*. 2007;108(1):1-9. doi:10.1016/j.imlet.2006.09.007

35. Satpathy AT, KC W, Albring JC, et al. Zbtb46 expression distinguishes classical dendritic cells and their committed progenitors from other immune lineages. *J Exp Med*. 2012;209(6):1135-1152. doi:10.1084/jem.20120030

36. Chevrier S, Corcoran LM. BTB-ZF transcription factors, a growing family of regulators of early and late B-cell development. *Immunol Cell Biol*. 2014;92(6):481-488. doi:10.1038/icb.2014.20

37. Wang Y, Sun HY, Kumar S, Puerta M del M, Jo H, Rezvan A. ZBTB46 is a shear-sensitive transcription factor inhibiting endothelial cell proliferation via gene expression regulation of cell cycle proteins. *Lab Investig*. 2019;99(3):305-318. doi:10.1038/s41374-018-0060-5

38. Fararjeh AFS, Liu YN. ZBTB46, SPDEF, and ETV6: Novel potential biomarkers and therapeutic targets in castration-resistant prostate cancer. *Int J Mol Sci*. 2019;20(11). doi:10.3390/ijms20112802

39. Lill CM, Schjeide BMM, Graetz C, et al. MANBA, CXCR5, SOX8, RPS6KB1 and ZBTB46 are genetic risk loci for multiple sclerosis. *Brain*. 2013;136(6):1778-1782. doi:10.1093/brain/awt101

40. Nishizaki SS, Boyle AP. Mining the Unknown: Assigning Function to Noncoding Single Nucleotide Polymorphisms Toward the Goal of Understanding Variation. *Trends Genet*. 2017;33(1):34-45. doi:10.1016/j.tig.2016.10.008.Mining

41. Beaulieu AM, Sant'Angelo DB. The BTB-ZF Family of Transcription Factors: Key Regulators of Lineage Commitment and Effector Function Development in the Immune System. *J Immunol*. 2011;187(6):2841-2847. doi:10.4049/jimmunol.1004006

42. Meredith MM, Liu K, Darrasse-Jeze G, et al. Expression of the zinc finger transcription factor zDC (Zbtb46, Btbd4) defines the classical dendritic cell lineage. *J Exp Med*. 2012;209(6):1153-1165. doi:10.1084/jem.20112675

43. Leppek K, Das R, Barna M. Functional 5′ UTR mRNA structures in eukaryotic translation regulation and how to find them. *Nat Rev Mol Cell Biol*. 2018;19(3):158-174.

doi:10.1038/nrm.2017.103

44. Hou ST, Nilchi L, Li X, et al. Semaphorin3A elevates vascular permeability and contributes to cerebral ischemia-induced brain damage. *Sci Rep*. 2015;5. doi:10.1038/srep07890

45. Fujita H, Zhang B, Sato K, Tanaka J, Sakanaka M. Expressions of neuropilin-1, neuropilin-2 and semaphorin 3A mRNA in the rat brain after middle cerebral artery occlusion. *Brain Res*. 2001;914(1-2):1-14. doi:10.1016/S0006-8993(01)02765-2

46. Burton PR, Clayton DG, Cardon LR, et al. Association scan of 14,500 nonsynonymous SNPs in four diseases identifies autoimmunity variants. *Nat Genet*. 2007;39(11):1329-1337. doi:10.1038/ng.2007.17

47. Sieper J, Deodhar A, Marzo-Ortega H, et al. Secukinumab efficacy in anti-TNF-naive and anti-TNF-experienced subjects with active ankylosing spondylitis: Results from the MEASURE 2 Study. *Ann Rheum Dis*. 2017;76(3):571-575. doi:10.1136/annrheumdis-2016-210023

48. Carrera C, Cullell N, Torres-Águila N, et al. Validation of a clinical-genetics score to predict hemorrhagic transformations after rtPA. *Neurology*. 2019;93(9):e851-e863. doi:10.1212/WNL.0000000000007997

49. Tesi N, van der Lee SJ, Hulsman M, et al. Centenarian controls increase variant effect sizes by an average twofold in an extreme case–extreme control analysis of Alzheimer's disease. *Eur J Hum Genet*. 2019;27(2):244-253. doi:10.1038/s41431-018-0273-5

50. Winkler TW, Day FR, Croteau-Chonka DC, et al. Quality control and conduct of genome-wide association meta-analyses. *Nat Protoc*. 2014;9(5):1192-1212. doi:10.1038/nprot.2014.071

51. Howrigan D. Details and considerations of the UK Biobank GWAS — Neale lab. http://www.nealelab.is/blog/2017/9/11/details-and-considerations-of-the-uk-biobank-gwas. Accessed February 11, 2021.

52. Lim ET, Würtz P, Havulinna AS, et al. Distribution and Medical Impact of Loss-of-Function Variants in the Finnish Founder Population. *PLoS Genet*. 2014;10(7). doi:10.1371/journal.pgen.1004494

53. Crosslin DR, Tromp G, Burt A, et al. Controlling for population structure and genotyping platform bias in the eMERGE multi-institutional biobank linked to electronic health records. *Front Genet*. 2014;5(SEP):1-14. doi:10.3389/fgene.2014.00352

54. Powers WJ, Rabinstein AA, Ackerson T, et al. 2018 Guidelines for the Early Management of Patients With Acute Ischemic Stroke: A Guideline for Healthcare Professionals From the American Heart Association/American Stroke Association. *Stroke*. 2018;49(3):e46-e110. doi:10.1161/STR.0000000000000158

FIGURE LEGENDS

Figure 1. Genome-wide association meta-analysis of Parenchymal Hematoma

SNPs were represented by dots and were plotted based on their GWAs meta-analysis p-values.

A. Manhattan plot of genome-wide association meta-analysis. SNPs were represented by dots and plotted based on their genome-wide association meta-analysis p-values. The red line shows genome-wide significance ($p < 5\times10^{-8}$) and the blue line represents the suggestive association significance threshold ($p<1x10^{-5}$). Results were adjusted for age, sex, baseline NIH Stroke Scale, diabetes, and principal components.

B. Regional association plot centered on rs76484331 (*ZBTB46*). The regional plot was drawn using LocusZoom software v0.4.8. Linkage disequilibrium with the top SNP is represented by the colored dots. Genes were characterized by horizontal lines. The *ZNF512B* gene was omitted.

Figure 2. RNA-seq expression of *ZBTB46* in the brain

A. The T-distributed Stochastic Neighbor Embedding (tSNE) plots represent 26,331 nuclei and clusters. ZBTB46 gene expression level is represented by color: high expression is presented in purple and low expression in gray.

B. Clusters are identified by cell type: Ex: excitatory neurons; In: inhibitory neurons; oligodendrocytes, microglia, OPC: oligodendrocyte precursor cells, and endothelial cells. Data were extracted from the web-based application http://ngi.pub/snuclRNA-seq/

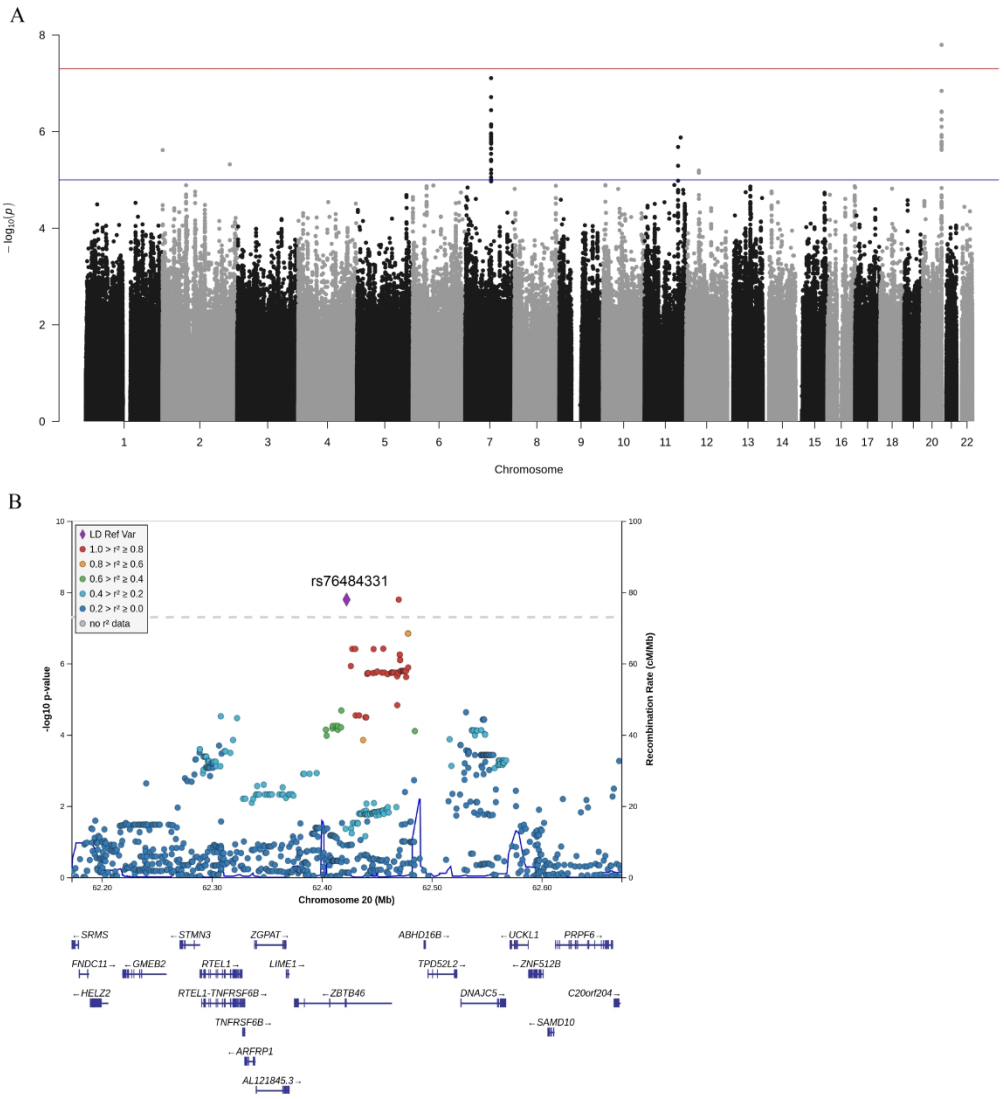


Figure 1. Genome-wide association meta-analysis of Parenchymal Hematoma. SNPs were represented by dots and were plotted based on their GWAs meta-analysis p-values.

A. Manhattan plot of genome-wide association meta-analysis. SNPs were represented by dots and plotted based on their genome-wide association meta-analysis p-values. The red line shows genome-wide significance (p <5×10−8) and the blue line represents the suggestive association significance threshold (p<1x10−5). Results were adjusted for age, sex, baseline NIH Stroke Scale, diabetes, and principal components.

B. Regional association plot centered on rs76484331 (ZBTB46).

The regional plot was drawn using LocusZoom software v0.4.8. Linkage disequilibrium with the top SNP is represented by the colored dots. Genes were characterized by horizontal lines. The ZNF512B gene was omitted.

371x419mm (300 x 300 DPI)

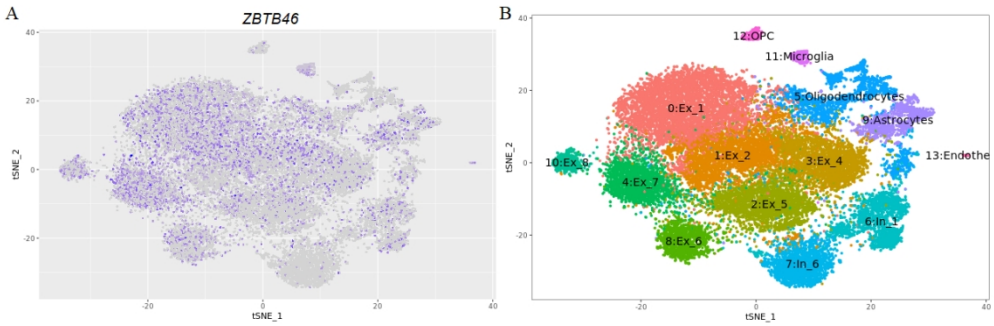


Figure 2. RNA-seq expression of ZBTB46 in the brain.

A. The T-distributed Stochastic Neighbor Embedding (tSNE) plots represent 26,331 nuclei and clusters. ZBTB46 gene expression level is represented by color: high expression is presented in purple and low expression in gray.
B. Clusters are identified by cell type: Ex: excitatory neurons; In: inhibitory neurons; oligodendrocytes, microglia, OPC: oligodendrocyte precursor cells, and endothelial cells.
Data were extracted from the web-based application http://ngi.pub/snuclRNA-seq/

107x36mm (300 x 300 DPI)

STREGA statement: Reporting guidelines checklist for genetic association studies

SECTION	ITEM NUMBER	CHECKLIST ITEM	REPORTED ON PAGE NUMBER:
TITLE AND ABSTRACT			
	1a	Indicate the study's design with a commonly used term in the title or the abstract	3
	1b	Provide in the abstract an informative and balanced summary of what was done and what was found	3
INTRODUCTION			
Background and objectives	2	Explain the scientific background and rationale for the investigation being reported	4
	3	State specific objectives, including any pre-specified hypotheses. State if the study is the first report of a genetic association, a replication effort, or both	4
METHODS			
Study design	4	Present key elements of study design early in the paper	5,6
Setting	5	Describe the setting, locations, and relevant dates, including periods of recruitment, exposure, follow-up, and data collection	5
Participants	6a	Give information on the criteria and methods for selection of subsets of participants from a larger study, when relevant	5
Variables	7	Clearly define genetic exposures (genetic variants) using a widely-used nomenclature system. Identify variables likely to be associated with population stratification (confounding by ethnic origin)	5,6
Data sources/measurements	8*	Describe laboratory methods, including source and storage of DNA, genotyping methods and platforms (including the allele calling algorithm used, and its version), error rates and call rates. State the laboratory/centre where genotyping was done. Describe comparability of laboratory methods if there is more than one group. Specify whether genotypes were assigned using all of the data from the study simultaneously or in smaller batches	5,6,7
Bias	9	For quantitative outcome variables, specify if any investigation of potential bias resulting from pharmacotherapy was undertaken. If relevant, describe the nature and magnitude of the potential bias, and explain what approach was used to deal with this	NA

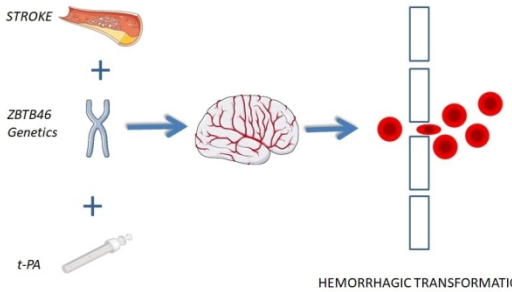
SECTION	ITEM NUMBER	CHECKLIST ITEM	REPORTED ON PAGE NUMBER:
Study size	10	Explain how the study size was arrived at	5
Quantitative variables	11	If applicable, describe how effects of treatment were dealt with	8
Statistical methods	12a	Describe all statistical methods, including those used to control for confounding. State software version used and options (or settings) chosen	8
	12b	Describe any methods used to examine subgroups and interactions	5,6,7
	12c	Explain how missing data were addressed	NA
	12d	Cohort study—If applicable, explain how loss to follow-up was addressed Case-control study—If applicable, explain how matching of cases and controls was addressed Cross-sectional study—If applicable, describe analytical methods taking account of sampling strategy	5
	12e	Describe any sensitivity analyses	NA
	12f	State whether Hardy–Weinberg equilibrium was considered and, if so, how	6
	12g	Describe any methods used for inferring genotypes or haplotypes	6
	12h	Describe any methods used to assess or address population stratification	5,6
	12i	Describe any methods used to address multiple comparisons or to control risk of false-positive findings	7,9
	12j	Describe any methods used to address and correct for relatedness among subjects	6
RESULTS			
Participants	13a	Report numbers of individuals in whom genotyping was attempted and numbers of individuals in whom genotyping was successful	8
	13b	Give reasons for non-participation at each stage	5
	13c	Consider use of a flow diagram	done
Descriptive Data	14a	Give characteristics of study participants (e.g. demographic, clinical, social) and information on exposures and potential confounders. Consider giving information by genotype	**6,8,9**

SECTION	ITEM NUMBER	CHECKLIST ITEM	REPORTED ON PAGE NUMBER:
	14b	Indicate number of participants with missing data for each variable of interest	5, 8
	14c	Cohort study—Summarise follow-up time (eg, average and total amount)	5,6
Outcome Data	15*	Cohort study— Report outcomes (phenotypes) for each genotype category over time Case-control study— Report numbers in each genotype category Cross-sectional study— Report outcomes (phenotypes) for each genotype category	26
Main Results	16a	Give unadjusted estimates and, if applicable, confounder-adjusted estimates and their precision (e.g. 95% confidence interval). Make clear which confounders were adjusted for and why they were included	done
	16b	Report category boundaries when continuous variables were categorized	NA
	16c	If relevant, consider translating estimates of relative risk into absolute risk for a meaningful time period	NA
	16d	Report results of any adjustments for multiple comparisons	done
Other Analyses	17a	Report other analyses done—e.g. analyses of subgroups and interactions, and sensitivity analyses	done
	17b	If numerous genetic exposures (genetic variants) were examined, summarize results from all analyses undertaken	done
	17c	If detailed results are available elsewhere, state how they can be accessed	NA
DISCUSSION			
Key Results	18	Summarise key results with reference to study objectives	11-14
Limitations	19	Discuss limitations of the study, taking into account sources of potential bias or imprecision. Discuss both direction and magnitude of any potential bias	13,14
Interpretation	20	Give a cautious overall interpretation of results considering objectives, limitations, multiplicity of analyses, results from similar studies, and other relevant evidence	11-14
Generalisability	21	Discuss the generalisability (external validity) of the study results Other information	11-14
FUNDING			
	22	Give the source of funding and the role of the funders for the present study and, if applicable, for the original study on which the present article is based	14,15

SECTION	ITEM NUMBER	CHECKLIST ITEM	REPORTED ON PAGE NUMBER:

*Give information separately for cases and controls in case-control studies and, if applicable, for exposed and unexposed groups in cohort and cross-sectional studies.

Reference: von Elm E, Altman DG, Egger M, Pocock SJ, Gotzsche PC, Vandenbroucke JP. The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) Statement: guidelines for reporting observational studies. This guideline was published simultaneously in 8 journals. Ann Intern Med. 2007; 147(8):573-577. PMID: 17938396, PLoS Med. 2007;4(10):e296. PMID: 17941714, BMJ. 2007;335(7624):806-808. PMID: 17947786, Prev Med. 2007;45(4):247-251. PMID: 17950122, Epidemiology. 2007;18(6):800-804. PMID: 18049194, Lancet. 2007;370(9596):1453-1457. PMID: 18064739 J Clin Epidemiol. 2008;61(4):344-349. PMID: 18313558, Bull World Health Organ. 2007;85(11):867-872. PMID: 18038077



217x144mm (192 x 192 DPI)

Carrera *et al.* describe the first genome-wide association study on ischaemic stroke patients treated with pharmacological thrombolysis. They show that polymorphisms in the *ZBTB46* gene are associated with the risk of parenchymal haematoma, and replicate the finding in an independent cohort.

Table 1 Cohorts included in the meta-analysis

Stage	Cohort	*n*	PH (%)	Location	Array
Discovery	GENISIS	1062	52 (4.9)	Spain	HumanCore ExomeChip
	GODS	234	14 (6)	Spain	HumanCore ExomeChip
	GRECOS	28	5 (17.9)	Spain	HumanCore ExomeChip
	Overall	**1324**	**71 (5.4)**	-	-
Replication[a]	GENOTPA[b]	157	36 (22.9)	Spain	Omni 1.5M
	BASICMAR[b]	91	8 (8.8)	Spain	Omni 5M
	LSGS[b]	45	8 (17.8)	Belgium	Omni 5M
	HELSINKI2000[c]	164	12 (7.3)	Finland	HumanCore ExomeChip
	GENISIS[c]	70	2 (2.9)	Finland	HumanCore ExomeChip
	GENISIS	53	4 (7.5)	Poland	HumanCore ExomeChip
	Overall	**580**	**70 (12.1)**	-	-
Overall	-	**1904**	**141 (7.4)**	-	-

[a]Contributing cohorts were not analyzed by ancestry stratum due to the limited samples available.

[b,c]The shaded cohorts in the Replication Stage indicate groups pooled and imputed based on genotyping array.

PH: Parenchymal hematoma; GENISIS: Genetics of Early Neurological Instability after Ischemic Stroke study; GODS: Genetic contribution to Functional Outcome and Disability after Stroke study; GRECOS: Genotyping Recurrence Risk of Stroke; GenoTPA: Genetic study in ischemic stroke patients treated with t-PA; BASICMAR: Base de Datos de Ictus del Hospital del Mar [*Hospital del Mar Stroke Database*]; LSGS: Leuven Stroke Genetics Study; HELSINKI2000: Helsinki 2000 Ischemic Stroke Genetics Study.

Table 2 Clinical findings and univariate analysis of the Discovery Study subjects

		PH			
	Total (n = 1324)	Absence (n = 1253)	Presence (n = 71)	P	OR (95%CI)
Sex, male (%)	733 (55.4)	692 (55.2)	41 (57.7)	0.7	1.11 (0.68–1.79)
AF (%)	361 (27.3)	336 (26.9)	25 (35.2)	0.13	1.48 (0.89–2.44)
DM (%)	329 (24.8)	303 (24.2)	26 (36.6)	0.02*	1.81 (1.10–2.99)
HTN (%)	867 (65.8)	817 (65.5)	50 (71.4)	0.31	1.32 (0.78–2.24)
ST (%)	374 (34.5)	356 (34.7)	18 (31)	0.57	0.85 (0.48–1.50)
TOAST (%)					
CE	556 (43,4)	518 (42.6)	38 (55.1)	0.04*	1.65 (1.02–2.69)
LAA	219 (17.1)	215 (17.7)	4 (5.9)	0.01*	0.29 (0.07–0.79)
SVO	56 (4.4)	55 (4.5)	1 (1.5)	0.36	0.31 (0.01–1.89)
OT	24 (1.9)	23 (1.9)	1 (1.5)	1	0.77 (0.02–4.91)
UND	426 (33.3)	402 (33.1)	24 (35.3)	0.71	1.1 (0.6–1.8)
Age, years (IQR)	76 (65–82)	75 (65–82)	77 (70–82)	0.28	
Baseline NIHSS (IQR)	12 (7–18)	11 (7–18)	18 (14–22)	<0.001*	
Glucose, mg/dl (IQR)	120 (103–147)	119 (103–146)	130 (105–168)	0.18	
OTT, min (IQR)[a]	130 (90–180)	128 (90–180)	140 (90–180)	0.54	
SBP, mm Hg (IQR)	155 (138–172)	154 (138–172)	158 (144–173)	0.36	
DBP, mm Hg (IQR)	82 (71–98)	81 (71–97)	85 (70–112)	0.34	

*P-value <0.05. P-values after logistic regression: DM: $P = 0.04$; CE: $P = 0.29$; Baseline NIHSS: $P = 4.8 \times 10^{-9}$.

[a]OTT was available in 788 participants.

For categorical variables, frequencies were described as percentages. For continuous variables, the median values and interquartile range (IQR) were calculated.

(95%CI): 95% Confidence Interval; AF: Atrial Fibrillation; CE: Cardioembolism; DBP: Diastolic blood pressure; DM: Diabetes; HTN: Hypertension; LAA: Large–artery atherosclerosis; NIHSS: National Institutes of Health Stroke Score; OR: Odds–ratio; OT: Other etiology; OTT: Time from onset to treatment; PH: Parenchymal hematoma; SBP:Systolic blood pressure; ST: Statins; SVO: Small vessel occlusion; TOAST: Trial of Org 1072 in Acute Stroke Treatment; UND: Undetermined etiology.

Table 3 Clinical characteristics of the replication cohort

		PH (%)			
	Total (n = 580)	Absence (n = 510)	Presence (n = 70)	*P*	OR (95%CI)
Sex, male (%)	287 (49.5)	256 (50.2)	31 (44.3)	0.35	0.79 (0.48–1.30)
CE (%)	286 (49.8)	239 (47.3)	47 (68.1)	0.001*	2.38 (1.39–4.06)
Age, years (IQR)	71 (60–78)	70 (59–78)	74 (68–80)	0.004*	
DM, yes (%)	42 (12.9)	35 (12.7)	7 (13.7)	0.84	1.09 (0.46–2.61)
Baseline NIHSS (IQR)	10 (6–17)	9 (5–15)	16 (11–19)	<0.001*	

*P-value <0.05.

For categorical variables, frequencies were described as percentages. For continuous variables, median values and interquartile range (IQR) were calculated.

95%CI = 95% confidence interval; CE = Cardioembolism etiology by Trial of Org 1072 in Acute Stroke Treatment classification; DM = diabetes; NIHSS = National Institutes of Health Stroke Score; OR = odds-ratio; PH = parenchymal haematoma.

Table 4 Independent leading SNPs and the most significant associations with PH in the meta-analysis

SNP	CHR	Position (bp)	Location	Gene	EA/NEA	No. variants	EAF	Stage	OR (95%CI)	*P*
rs77557904	2	1047076	intronic	*SNTG2*	G/C	1	0.06	Meta-analysis	3.82 (2.19–6.68)	2.43×10^{-6}
								Discovery	6.56 (3.10–13.89)	8.62×10^{-7}
								Replication	2.01 (0.88–4.59)	0.1
rs112541215	2	218916291	intronic	*RUFY4*	A/T	1	0.08	Meta-analysis	3.01 (1.88–4.84)	4.79×10^{-6}§
								Discovery	4.67 (2.43–8.96)	3.76×10^{-6}
								Replication	1.90 (0.96–3.74)	0.07
rs1962779	7	83837734	intronic	*SEMA3A*	C/G	40	0.19	Meta-analysis	2.43 (1.76–3.37)	7.85×10^{-8}§
								Discovery	2.94 (1.87–4.61)	2.77×10^{-6}
								Replication	2.00 (1.26–3.18)	3.34×10^{-3}
rs4356265	11	117301818	intronic	*DSCAML1*	T/C	1	0.11	Meta-analysis	2.75 (1.83–4.15)	1.33×10^{-6}§
								Discovery	3.92 (2.23–6.88)	1.89×10^{-6}
								Replication	1.87 (1.03–3.38)	0.04
rs564865745	12	41626444	intronic	*PDZRN4*	G/A	57	0.05	Meta-analysis	4.23 (2.37–7.54)	1.03×10^{-6}
								Discovery	9.89 (4.20–23.28)	1.56×10^{-7}
								Replication	2.12 (0.97–4.62)	0.06
rs76484331	**20**	**62422504**	**intronic**	***ZBTB46***	**A/C**	**43**	**0.1**	**Meta-analysis**	**5.84 (3.16–10.76)**	**1.61×10^{-8}***
								Discovery	**11.31 (4.82–26.55)**	**2.49×10^{-8}***
								Replication	2.97 (1.24–7.09)	0.01

**P*-value < 5×10^{-8} §heterogeneity *P*-value > 0.05.

(95%CI): 95% Confidence Interval; bp: Base pair; CHR: Chromosome; EA: Effect allele; EAF: Effect allele frequency; No. Variants: Number of variants reaching p>$1x10^{-5}$; NEA: Non-effect allele; OR: Odds-ratio; SNP: Single nucleotide polymorphism.

Only independent SNPs (r2<0.1, within 1MB window) with a p-value <1x10-5 are shown. Alleles and chromosomal positions were identified on the basis of the 1000 Genomes Phase 3 Project. Location was described following the ANNOVAR system.