


**Please cite the Published Version**

Cullell, Natalia, Gallego-Fábrega, Cristina, Cárcel-Márquez, Jara, Muiño, Elena, Lluçia-Carol, Laia, Lledós, Miquel, Martín-Campos, Jesús M, Molina, Jessica, Casas, Laura, Almeria, Marta, Fernández-Cadenas, Israel and Krupinski, Jerzy  (2022) ICA1L is associated with small vessel disease: a proteome-wide association study in small vessel stroke and intracerebral haemorrhage. *International Journal of Molecular Sciences*, 23 (6). 3161 ISSN 1661-6596

**DOI:** <https://doi.org/10.3390/ijms23063161>

**Publisher:** MDPI

**Version:** Published Version

**Downloaded from:** <https://e-space.mmu.ac.uk/634367/>

**Usage rights:**  [Creative Commons: Attribution 4.0](https://creativecommons.org/licenses/by/4.0/)

**Additional Information:** This is an open access article which originally appeared in *International Journal of Molecular Sciences*, published by MDPI

**Data Access Statement:** The datasets used and/or analysed during the current study are available from the corresponding author upon reasonable request.

**Enquiries:**

If you have questions about this document, contact [openresearch@mmu.ac.uk](mailto:openresearch@mmu.ac.uk). Please include the URL of the record in e-space. If you believe that your, or a third party's rights have been compromised through this document please see our Take Down policy (available from <https://www.mmu.ac.uk/library/using-the-library/policies-and-guidelines>)



Article

# ICA1L Is Associated with Small Vessel Disease: A Proteome-Wide Association Study in Small Vessel Stroke and Intracerebral Haemorrhage

Natalia Cullell <sup>1,2,3,4</sup>, Cristina Gallego-Fàbrega <sup>1</sup>, Jara Cárcel-Márquez <sup>1,5</sup>, Elena Muiño <sup>1</sup>, Laia Lluçia-Carol <sup>1</sup>, Miquel Lledós <sup>1</sup>, Jesús M. Martín-Campos <sup>1</sup>, Jessica Molina <sup>2</sup>, Laura Casas <sup>2</sup>, Marta Almeria <sup>2</sup>, Israel Fernández-Cadenas <sup>1,2,\*</sup> and Jerzy Krupinski <sup>2,\*</sup>

<sup>1</sup> Stroke Pharmacogenomics and Genetics Group, Biomedical Research Institute Sant Pau, 08041 Barcelona, Spain; ncullell@mutuaterrassa.cat (N.C.); cristina.gallego.fabrega@gmail.com (C.G.-F.); jara.carcel@gmail.com (J.C.-M.); elena.muinho@gmail.com (E.M.); laialluciacarol@gmail.com (L.L.-C.); miquel.lledos@gmail.com (M.L.); jmartinca@santpau.cat (J.M.M.-C.)

<sup>2</sup> Neurology Department, Hospital Universitari MútuaTerrassa, 08221 Terrassa, Spain; jessicamolina@mutuaterrassa.cat (J.M.); lcasas@mutuaterrassa.cat (L.C.); malmeria@mutuaterrassa.cat (M.A.)

<sup>3</sup> Fundacio Docència i Recerca MutuaTerrassa, 08221 Terrassa, Spain

<sup>4</sup> Department of Medicine, Universitat de Barcelona, 08036 Barcelona, Spain

<sup>5</sup> Department of Medicine, Universitat Autònoma de Barcelona, 08193 Barcelona, Spain

\* Correspondence: israelcadenas@yahoo.es (I.F.-C.); jkrupinski@mutuaterrassa.es (J.K.)



Citation: Cullell, N.;

Gallego-Fàbrega, C.; Cárcel-Márquez, J.; Muiño, E.; Lluçia-Carol, L.; Lledós, M.; Martín-Campos, J.M.; Molina, J.; Casas, L.; Almeria, M.; et al. ICA1L Is Associated with Small Vessel Disease: A Proteome-Wide Association Study in Small Vessel Stroke and Intracerebral Haemorrhage. *Int. J. Mol. Sci.* **2022**, *23*, 3161. <https://doi.org/10.3390/ijms23063161>

Academic Editor: Teresa Gasull

Received: 10 February 2022

Accepted: 10 March 2022

Published: 15 March 2022

**Publisher's Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Copyright:** © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

**Abstract:** Small vessel strokes (SVS) and intracerebral haemorrhages (ICH) are acute outcomes of cerebral small vessel disease (SVD). Genetic studies combining both phenotypes have identified three loci associated with both traits. However, the genetic cis-regulation at the protein level associated with SVD has not been studied before. We performed a proteome-wide association study (PWAS) using FUSION to integrate a genome-wide association study (GWAS) and brain proteomic data to discover the common mechanisms regulating both SVS and ICH. Dorsolateral prefrontal cortex (dPFC) brain proteomes from the ROS/MAP study (N = 376 subjects and 1443 proteins) and the summary statistics for the SVS GWAS from the MEGASTROKE study (N = 237,511) and multi-trait analysis of GWAS (MTAG)-ICH-SVS from Chung et al. (N = 240,269) were selected. We performed PWAS and then a co-localization analysis with COLOC. The significant and nominal results were validated using a replication dPFC proteome (N = 152). The replicated results ( $q$ -value < 0.05) were further investigated for the causality relationship using summary data-based Mendelian randomization (SMR). One protein (ICA1L) was significantly associated with SVS ( $z$ -score =  $-4.42$  and  $p$ -value =  $9.6 \times 10^{-6}$ ) and non-lobar ICH ( $z$ -score =  $-4.8$  and  $p$ -value =  $1.58 \times 10^{-6}$ ) in the discovery PWAS, with a high co-localization posterior probability of 4. In the validation PWAS, ICA1L remained significantly associated with both traits. The SMR results for ICA1L indicated a causal association of protein expression levels in the brain with SVS ( $p$ -value =  $3.66 \times 10^{-5}$ ) and non-lobar ICH ( $p$ -value =  $1.81 \times 10^{-5}$ ). Our results show that the association of ICA1L with SVS and non-lobar ICH is conditioned by the cis-regulation of its protein levels in the brain.

**Keywords:** small vessel disease; small vessel stroke; intracerebral haemorrhage; PWAS; proteome-wide association study; genome-wide association study (GWAS)

## 1. Introduction

Cerebral small vessel disease (SVD) is a chronic and progressive heterogeneous disorder affecting small arteries, arterioles, venules, and capillaries in the brain [1,2]. Among the different outcomes of SVD, small vessel stroke (SVS) and spontaneous intracerebral haemorrhage (ICH) are the most acute and devastating manifestations [3–5]. One-quarter

of all ischemic strokes (IS) are SVS [6], and ICH is the most common cause of haemorrhagic stroke [7]. The incidence of SVD increases with age, and it is, thus, expected that with the increase in life expectancy, the frequency of SVD manifestations will also increase in following years [1,3].

SVD causes parenchyma lesions, such as white matter hyperintensities (WMHs), enlarged perivascular spaces, lacunes, small subcortical infarcts, cerebral microbleeds, and brain atrophy, among others [1,3]. These parenchyma lesions are used as surrogate markers of SVD and can be quantified using brain imaging techniques. The extent of these lesions predicts poor outcomes after ICH [8] and SVS [9]. The two major pathophysiological mechanisms beyond SVD are arteriosclerosis and cerebral amyloid angiopathy (CAA) [1]. Cardiovascular risk factors, mainly hypertension, are associated with arteriosclerotic SVD. It can cause the rupture or occlusion of small arteries and arterioles, producing non-lobar (or deep) ICH and SVS, respectively [10].

There is no effective treatment for SVD, apart from controlling cardiovascular risk factors [4]. Genome-wide association studies (GWAS) have been helpful in identifying genes involved in SVD, analysing magnetic resonance image (MRI) SVD surrogate markers [11–21], and, moreover, specifically studying SVS and ICH.

Two GWAS have been performed with ICH patients. The first study identified a risk locus for non-lobar ICH in 1q22, overlapping with *PMF1* and *SLC25A44* genes [5,11,12]. The second performed a meta-analysis that allowed for the identification of *APOE* associated with lobar ICH [4]. In order to improve the statistical power of these studies, Chung et al. applied a novel strategy, using a multi-trait analysis of GWAS (MTAG) and combining genotype data from two related phenotypes, ICH and SVS. This approach permitted the identification of two novel associations in 2q33 (*ICA1L*) and 13q34 (*COL4A2* and *COL4A1*) associated with non-lobar ICH [4]. Interestingly, these loci presented a nominal association ( $p$ -value  $< 5 \times 10^{-3}$ ) with non-lobar ICH and SVS in the corresponding GWAS but not at a genome-wide level [4]. Three different GWAS in SVS have identified significant associations with this phenotype in 16q24.2 (*ZCCHC14*) [6], 2q33.2 (*ICA1L* and *WDR12*) [22], *ULK4*; *SPI1-SLC39A13-PSMC3-RAPSN*, and *ZBTB14-EPB41L3* [23]. Recently, an MTAG combining SVS and WMH allowed for the identification of seven loci associated with SVS for the first time but some of them were previously associated with ICH (*SLC25A44-PMF1-BGLAP*, *LOX-ZNF474-LOC100505841*, *FOXF2 FOXQ1*, *SH3PXD2A*, *VTA1-GPR126*, *HTRA1-ARMS2*, and *COL4A2*) [23].

Proteome-wide association studies (PWAS) have been proposed as a novel strategy to recover causal protein-coding genes and to identify new associations for a disease. Since proteins are the focus of PWAS [24], it is a potentially beneficial tool for the identification of drug targets for disease [25]. This method, combined with Bayesian or Mendelian randomization, allows for the identification of gene-phenotype associations causally mediated by protein levels [24–26]. We performed a proteome-wide association study (PWAS) to integrate the genomic and proteomic data to discover common mechanisms regulating the two main acute manifestations of SVD: SVS and ICH.

## 2. Results

### 2.1. Stringent Criterion PWAS Analysis

From the 1475 proteins evaluated in the discovery cohort, only *ICA1L* reached a significant  $q$ -value in association with SVS ( $z$ -score =  $-4.93$ ;  $q$ -value =  $1.42 \times 10^{-2}$ ) and with non-lobar ICH ( $z$ -score =  $-4.8$ ;  $q$ -value =  $2.33 \times 10^{-3}$ ) (Table 1). The results from the Bayesian colocalization analysis showed a high posterior probability for the rs7582720 to be the causal variant for both the *ICA1L* levels and the significant GWAS traits (SVS and non-lobar ICH) (posterior probability 4 (PP4) = 0.9 for SVS and non-lobar ICH) (Table 1). The association of *ICA1L* with SVS and non-lobar ICH was the only one to reach the stringent criterion, and it was considered for the replication stage. Both associations resulted as significant in the replication analysis (the  $z$ -score =  $-3.55$ ;  $q$ -value =  $5.08 \times 10^{-3}$ , and

PP4 = 0.99 for SVS and the z-score =  $-4.07$ ,  $q$ -value =  $6.19 \times 10^{-4}$ , and PP4 = 0.99 for non-lobar ICH) (Table 1).

**Table 1.** PWAS results for the stringent criteria. Significant PWAS associations using the stringent criteria in the discovery and replication. P0–P1: initial-final position for the analysed protein; pQTL ID: ID for the pQTL strongly associated with ICA1L; pQTL z-score: z-score for the pQTL strongly associated with ICA1L; pQTL z-score: z-score in the correspondent GWAS for the best pQTL; NSNP: number of SNPs in the locus; TWAS z-score: z-score for the PWAS association; PWAS  $p$ -value:  $p$ -value for the PWAS association; PP4: posterior probability 4 in the COLOC analysis;  $q$ -value: FDR-adjusted PWAS  $p$ -value.

Trait	Protein	Chr	P0–P1	pQTL ID	pQTL z-Score	pQTL GWAS z-Score	Discovery				Replication				
							NSNP	PWAS z-Score	PWAS $p$ -Value	PP4	$q$ -Value	PWAS z-Score	PWAS $p$ -Value	PP4	$q$ -Value
SVS	ICA1L	2	203640690	rs7582720	7.21	−4.426	81	−4.43	$9.60 \times 10^{-6}$	0.90	$1.42 \times 10^{-2}$	−3.55	$3.91 \times 10^{-4}$	0.99	$5.08 \times 10^{-3}$
Non-lobar ICH			203736708			−4.801		−4.80	$1.58 \times 10^{-6}$		0.93	$2.33 \times 10^{-3}$	−4.07		$4.76 \times 10^{-5}$

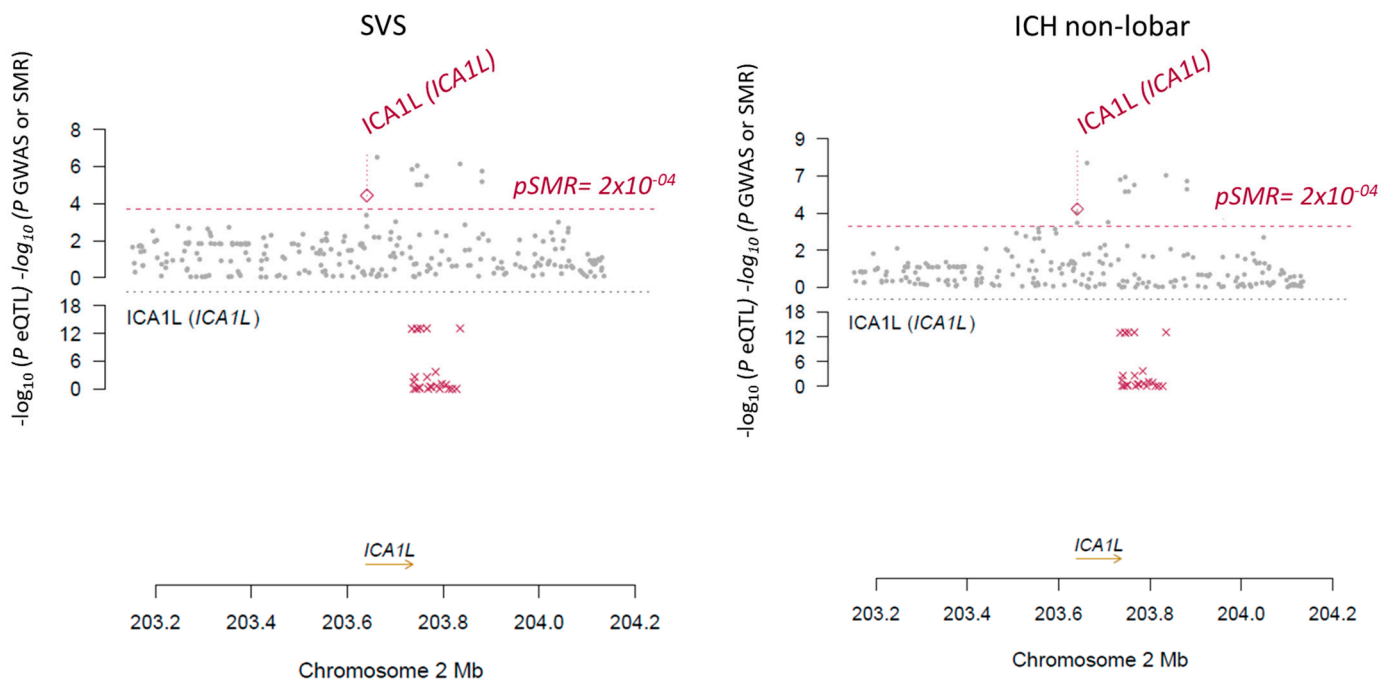
We also tested, specifically, those proteins from loci associated previously with SVD. The PWAS aimed to interrogate 26 proteins corresponding to 13 different loci previously associated with ICH (4 loci) and/or SVS (12 loci). From these, only ICA1L was present among the proteins analysed in the PWAS.

## 2.2. SMR and Conditional Analysis

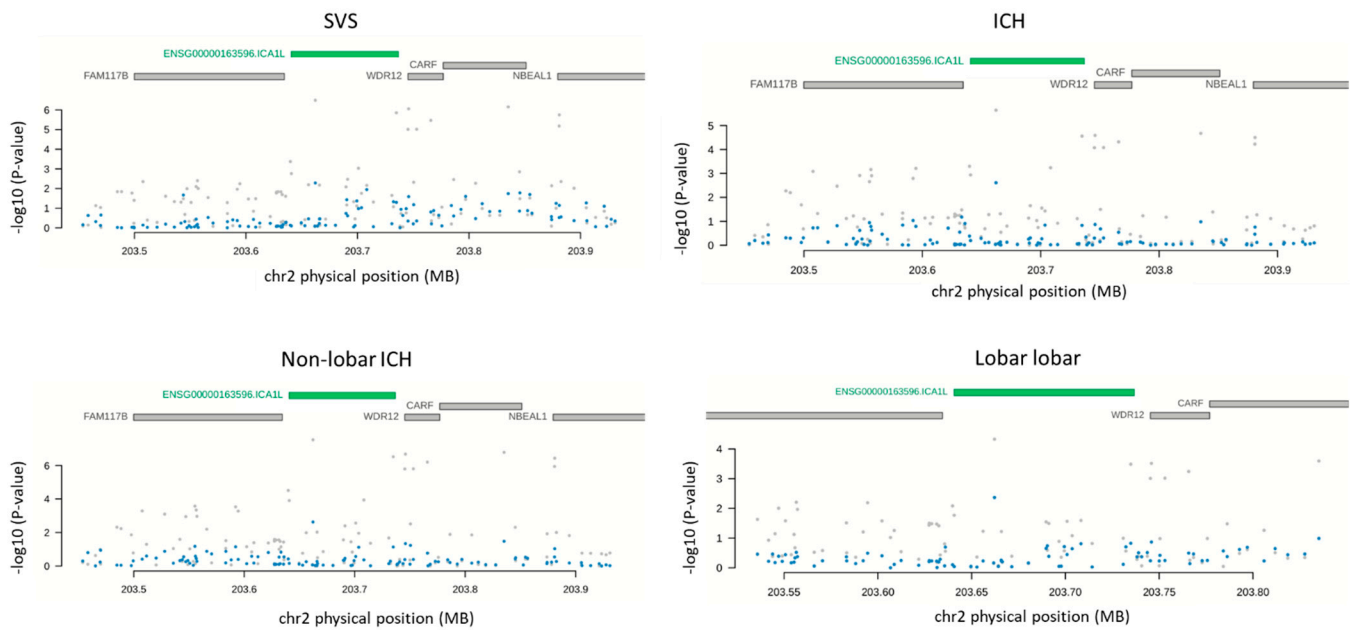
In order to ensure that the genetic association of ICA1L with SVS and non-lobar ICH was mediated by the cis-regulation of the ICA1L protein levels, SMR and a conditional analysis were performed. The SMR results indicated that the pQTL was associated with SVS ( $p$ -value =  $3.66 \times 10^{-5}$ ) and non-lobar ICH ( $p$ -value =  $1.81 \times 10^{-5}$ ) through the ICA1L protein expression regulation (Table 2, Figure 1). Similarly, the conditional analysis plot showed a conditional effect by the expression levels of ICA1L on the GWAS results (Figure 2). Table 2 shows the Summary statistics for the SMR results with the beta ( $b_{SMR}$ ) and  $p$ -values ( $p_{SMR}$ ) for all the significant and replicated associations from PWAS.

**Table 2.** SMR results.

Trait	Protein	$b_{SMR}$	$p_{SMR}$
SVS	ICA1L	−2.78	$3.66 \times 10^{-5}$
SVS	SPATA20	−0.24	$3.69 \times 10^{-1}$
ICH	ICA1L	−1.55	$1.81 \times 10^{-5}$
Non-lobar ICH	ICA1L	−1.55	$1.81 \times 10^{-5}$
Lobar ICH	SPATA20	−0.05	$7.98 \times 10^{-1}$



**Figure 1.** Significant SMR results for the significant and replicated PWAS results. Each “x” in figure represents a pQTL for ICA1L.



**Figure 2.** GWAS results before conditioning by ICA1L levels (grey) or after conditioning by protein levels (blue).

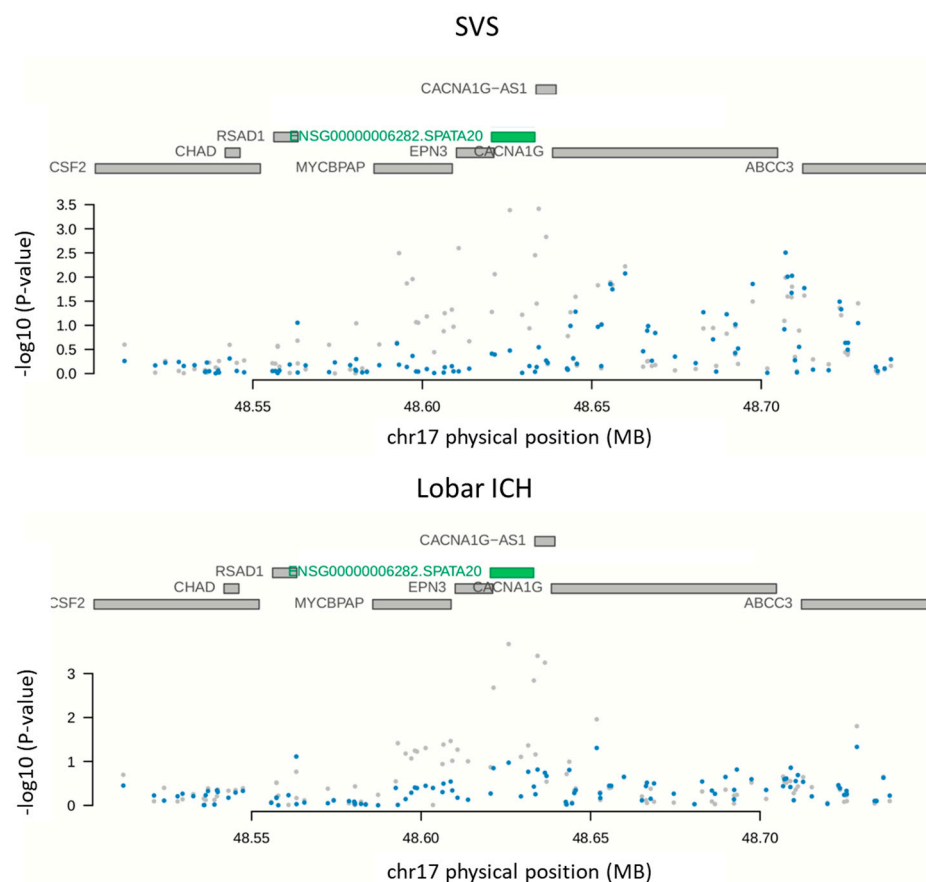
### 2.3. Open Criterion PWAS Analysis

In considering the open criterion ( $p$ -value  $< 1 \times 10^{-5}$  in the discovery PWAS), five proteins reached a nominal association with all ICH, four proteins with SVS (excluding ICA1L) and lobar ICH, and two proteins with non-lobar ICH (excluding ICA1L) (Table 3). In the replication of the open analysis, only SPATA20 and ICA1L were significant ( $p$ -value  $< 0.05$ ), with a PP4 higher than 0.5 in the replication proteome panel. SPATA20 was significantly associated with SVS and lobar ICH ( $q$ -value =  $1.95 \times 10^{-3}$  and PP4 = 0.97 for SVS;  $8.82 \times 10^{-3}$  and PP4 = 0.55 for lobar ICH). ICA1L was significantly associated with all ICH ( $q$ -value =  $1.25 \times 10^{-2}$  and PP4 = 0.99 for all ICH). Two proteins (NRBF2 and OSBPL11) could not be evaluated in the replication because they were not present in the panel (Table 3). Results from the SMR indicated that an ICA1L abundance was mediating the association through the pQTL and all ICH ( $p$ -value =  $3.66 \times 10^{-5}$ ). However, the SMR results for the SPATA20 were not significant ( $p$ -value =  $3.69 \times 10^{-1}$  for SVS and  $7.98 \times 10^{-1}$  for lobar ICH) (Table 2). The conditional analyses for both proteins showed that the genetic results were conditionally affected by a protein abundance (Figures 2 and 3).

**Table 3.** PWAS results for the open criteria. The results for the discovery and replication of PWAS associations for proteins with a nominal  $p$ -value in the discovery (open criteria).

Trait	Protein	Chr	pQTL ID	Discovery						Replication				
				pQTL z-Score	pQTL GWAS z-Score	NSNP	PWAS z-Score	PWAS p-Value	PP4	q-Value	PWAS Z-Score	PWAS p-Value	PP4	q-Value
SVS	SPATA20	17	rs878619	-10.89	-2.917	102	3.71	$2.10 \times 10^{-4}$	0.17	$3.10 \times 10^{-1}$	3.79	$1.50 \times 10^{-4}$	0.97*	$1.95 \times 10^{-3}$
SVS	ALDH2	12	rs4648328	-7.25	-3.437	61	3.58	$3.38 \times 10^{-4}$	0.29	$4.99 \times 10^{-1}$	3.38	$7.28 \times 10^{-4}$	0.27	$9.46 \times 10^{-3}$
SVS	EXOC6	10	rs980204	4.98	-2.469	210	-3.58	$3.41 \times 10^{-4}$	0.19	$5.03 \times 10^{-1}$	3.37	$7.36 \times 10^{-1}$	-	1
SVS	OSBPL11	3	rs2922170	4.11	3577	111	3.58	$3.48 \times 10^{-4}$	0.35	$5.14 \times 10^{-1}$	-	-	-	-
ICH	ICA1L	2	rs7582720	7.21	-3.935	81	-3.94	$8.32 \times 10^{-5}$	0.68	$1.23 \times 10^{-1}$	-3.30	$9.66 \times 10^{-4}$	0.99*	$1.25 \times 10^{-2}$
ICH	NRBF2	10	rs4379723	-4.19	-3.96	77	3.85	$1.19 \times 10^{-4}$	0.18	$1.76 \times 10^{-1}$	-	-	-	-
ICH	DLGAP2	8	rs7842425	4.32	-3.742	225	-3.74	$1.83 \times 10^{-4}$	0.12	$2.70 \times 10^{-1}$	1.14	$2.54 \times 10^{-1}$	-	1
ICH	SPATA20	17	rs878619	-10.89	-2.89	102	3.58	$3.47 \times 10^{-4}$	0.12	$5.12 \times 10^{-1}$	3.27	$1.09 \times 10^{-3}$	0.33	$1.41 \times 10^{-2}$
ICH	MADD	11	rs11570115	-9.09	-2.565	99	3.54	$4.06 \times 10^{-4}$	0.03	$5.99 \times 10^{-1}$	1.88	$6.04 \times 10^{-2}$	-	0.78
Non-lobar ICH	SPATA20	17	rs878619	-10.89	-2.672	102	3.69	$2.19 \times 10^{-4}$	0.42	$3.23 \times 10^{-1}$	3.51	$4.44 \times 10^{-4}$	0.27	$5.77 \times 10^{-3}$
Non-lobar ICH	NRBF2	10	rs4379723	-4.19	-3.569	77	3.43	$6.00 \times 10^{-4}$	0.29	$8.90 \times 10^{-1}$	-	-	-	-
Lobar ICH	MRV11	11	rs753002	-5.36	2.452	211	-3.69	$2.20 \times 10^{-4}$	0.02	$3.25 \times 10^{-1}$	-1.87	$6.22 \times 10^{-2}$	-	0.80
Lobar ICH	NRBF2	10	rs4379723	-4.19	-3.804	77	3.68	$2.33 \times 10^{-4}$	0.11	$3.44 \times 10^{-1}$	-	-	-	-
Lobar ICH	SPATA20	17	rs878619	-10.89	-3.186	102	3.51	$4.44 \times 10^{-4}$	0.28	$6.55 \times 10^{-1}$	3.40	$6.79 \times 10^{-4}$	0.55*	$8.82 \times 10^{-3}$
Lobar ICH	ICA1L	2	rs7582720	7.21		81	-3.30	$9.6 \times 10^{-4}$	0.2	1	-2.5692	$1.02 \times 10^{-2}$	0.9	0.13

\* Significant PP4 in the replication. pQTL ID: ID for the pQTL strongly associated with proteins; pQTL z-score: z-score for the pQTL strongly associated with proteins; pQTL z-score: z-score in the correspondent GWAS for the best pQTL; NSNP: number of SNPs in the locus; TWAS z-score: z-score for the PWAS association; PWAS p-value: p-value for the PWAS association; PP4: posterior probability 4 in the COLOC analysis; q-value: FDR-adjusted PWAS p-value.



**Figure 3.** GWAS results before conditioning by the SPATA20 levels (grey) or after conditioning by protein levels (blue).

### 3. Discussion

In the present study, we performed a PWAS for the two main acute manifestations of SVD: SVS and ICH. We identified an association at the proteomic level of ICA1L with SVS and non-lobar ICH. Moreover, ICA1L was also associated with lobar ICH in our study. However, it was not a statistically significant association after a multivariable correction.

Although different loci have previously been found to be associated with ICH and SVS, how these genes contribute to their risk is not fully understood. One approach used to prioritize the likely causal gene within an associated locus with multiple genes is the use of a transcriptome-wide association study (TWAS) [27]. The TWAS from the SVS GWAS results found that the expression of *SLC25A44* in the arteries and the 2q33-2 locus (*CARF*, *FAM117B*, *ICA1L*, and *NBEAL1*) in the arteries and brain was causally associated with an SVS risk [23]. Very similar results regarding the 2q33-2 locus were previously obtained by Persyn et al. and Sargurupremraj et al. in their TWAS of WMH [16,17]. In contrast, the *ULK4* expression was negatively associated with an SVS risk in the arterial tissues, whole blood, and in the brain [23].

Our study implemented a similar strategy, focused on proteins, that allowed for a better characterization of a previously known association of ICA1L with non-lobar ICH, SVS [4,23], and related SVD traits, such as WMH [16,17] and fractional anisotropy (FA) [16]. Based on our results, genetics are responsible for the cis-regulation of an ICA1L abundance in brain tissue. Specifically, the top pQTL for ICA1L determines a higher abundance in ICA1L, which, in turn, reduces the risk for non-lobar ICH and SVS.

Although no previous PWAS has investigated the association of a protein abundance with SVS nor ICH, an ICA1L abundance was previously found to be associated with Alzheimer's disease (AD) risk using a PWAS approach [25,28]. These studies used the same proteomic data as in our work and found that a higher expression of the ICA1L protein

was linked with a decrease in AD risk [25,28]. Thus, both results for AD and SVD showed that elevated ICA1L levels are a protective factor. These AD studies, using single-cell RNA-sequencing, also identified enrichment in the ICA1L expression in glutamatergic excitatory neurons [25,28]. Excitatory processes involving glutamate are related to the cell death features in stroke and ICH [29,30]. Furthermore, there is a clear relationship between AD and SVD. Both have common features that lead to dementia, such as a blood–brain barrier breakdown [31]. Moreover, cerebral amyloid angiopathy and arteriosclerosis, common SVD mechanisms, worsen the threshold for AD dementia. Different SVD features have been found in post-mortem brains from AD patients, such as small cortical and subcortical infarcts, microbleeds, perivascular spacing, and WMH [32]. Additionally, *ApoE* is the major genetic factor associated both with lobar ICH and AD [33]. The *Ica1l* results suggest that this protein could also be commonly regulated in AD and SVD.

SPATA20 was also identified as a potential protein affecting SVS and lobar ICH. Previous genetic studies using a GWAS found polymorphisms in this gene associated with platelet distribution width [34], reticulocyte count [34], and subcortical volume [35]. However, pleiotropy was not found in the SMR analysis, indicating that different genetic regulations could be mediating the protein abundance and the trait risk.

The main limitation of our study is the number of proteins that we were able to evaluate. In this sense, only ICA1L was present in the proteomic panel from all proteins from loci previously associated with SVS or ICH. Moreover, we only had access to proteomic data from dPFC, and other brain regions could be relevant in SVD.

The results from this PWAS, combined with the previous results at the genomic and transcriptomic levels, highlight the relevance of ICA1L in the risk of SVD and specific SVD outcomes. As all these outcomes (SVS, non-lobar ICH, and lobar ICH) are common manifestations of SVD, the ICA1L regulation could be a common mechanism in SVD. Given that this regulation is also produced at the protein level and in brain tissue, future studies could focus on the investigation of ICA1L as a potential drug target.

## 4. Materials and Methods

### 4.1. Proteome-Wide Associations Study (PWAS)

We used FUSION software (Boston, MA, USA) [36] to estimate the association of brain protein levels with SVS and ICH. FUSION is implemented to identify associations between a functional phenotype (i.e., protein expression levels) and a GWAS phenotype. FUSION takes two inputs: 1) Precomputed functional weights, and 2) GWAS summary statistics unified to a reference single-nucleotide polymorphism (SNP) panel. In the PWAS, the functional weights consist of estimations of the strength of association between proteomic expression levels and the SNPs from the cis-locus of the protein (only considered in the heritable proteins). The SNPs associated with cis-locus protein expression levels are protein quantitative-traits (or pQTLs).

In our PWAS analysis, we selected, as functional weights, the freely available pre-computed weights for the proteome expression analysis published by Wingo et al. [28]. As mentioned, in this case, the functional weights consisted of a summary-based file with data for the association between protein levels and SNPs (for the same locus of the protein). In the study from Wingo et al., from which we selected the functional weights, post-mortem dorsolateral prefrontal cortex (dPFC) human brain proteomes from 400 participants of European ancestry from the ROS/MAP cohorts were generated [37]. Proteomic sequencing was performed using isobaric tandem mass tag (TMT) peptide labelling combined with liquid chromatography coupled to mass spectrometry. They carried out different quality controls (QCs), which are described in the original article [28]. First, they randomized the 400 samples and 100 global internal standards into 50 batches before TMT labeling with the TMT10plex kit (Thermo Fisher Scientific, Waltham, MA, USA). Each batch contained two exact global internal standards. After running mass spectrometry, they assigned a peptide spectral using the Proteome Discoverer suite v.2.3 (Thermo Fisher Scientific) and a search in the canonical UniProtKB human proteome database. Then, the peptide spectral



matches were collated into proteins. They did not include, in the analysis, the peptides whose protein quantification was outside the 95% confidence interval of the global internal standard, nor proteins with missing values in more than 50% of the samples. After the QCs, 8356 proteins remained for analysis. The abundance of these proteins was log<sub>2</sub> transformed. Then, QC for the samples was also performed, removing samples outstanding more than four standard deviations from the first two principal components. The effect of the batch, sex, age of death, study, mass spectrometry reporter quantification mode, postmortem interval, and clinical diagnosis was detected and removed [28].

They coupled this proteomic data with genetic data for weight calculation purposes, and the genotype data from 1,190,321 HapMap SNPs present in 489 individuals from the same cohort was processed. The genotypes were obtained from whole-genome sequencing or genome-wide genotyping. Classical QCs were applied to the genotype data: removal of variants deviating from the HardyWeinberg equilibrium, with a missing genotype rate of >5%, and rare variants with a minor allele frequency of <1%. Individuals with >5% genotyping missingness and related individuals were also excluded from the analysis. For genome-wide genotyped data, imputation was performed, and only HapMap SNPs selected from the 1000 Genomes Project were included. After considering all the QCs for proteomic and genotyping data, the functional weight included dPFC proteome abundance from heritable proteins (a total of 1475 heritable proteins) in 376 subjects, combined with 1,190,321 SNPs [28].

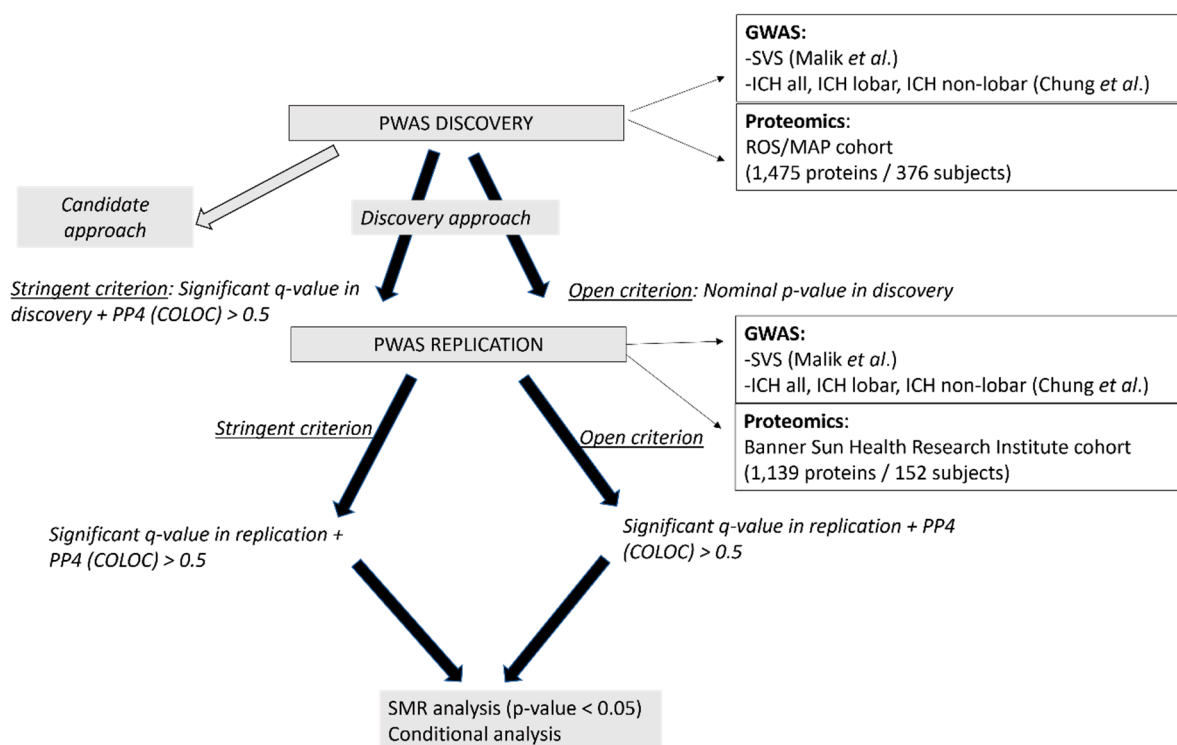
For the PWAS, we selected the GWAS summary statistics from the traits of interest considering the largest sample size published. For SVS, we selected summary statistics for the European ancestry MEGASTROKE SVS GWAS (N = 5386 SVS patients and 406,111 controls) [22]. For ICH, we selected the three different ICH subtypes (all ICH, ICH lobar, and ICH non-lobar) GWAS summary statistics available from the MTAG published by Chung et al. (N = 6255 ICH cases and 233,058 control subjects) [4].

Using FUSION, we first unified the GWAS SNPs (total SNPs SVS = 8,177,651; total SNPs all ICH = 7,136,230; total SNPs non-lobar ICH = 7,136,333; and total SNPs lobar ICH = 7,495,490) with the reference panel from the 1000 Genomes Project (Linkage disequilibrium (LD) reference panel). Secondly, SNPs from the GWAS not present in the LD reference panel were imputed. Finally, the association between the functional weights (1475 proteins) and the GWAS summary statistics was estimated. We corrected the results, considering multiple testing by 1475 proteins with a False Discovery Rate (FDR) multivariable test correction (*q*-values).

#### 4.2. Colocalization Analysis

We performed a Bayesian colocalization analysis for the PWAS results with a *p*-value of <0.05. The objective of this analysis was to assess whether the association of protein levels with the GWAS phenotype was caused by the same variant (or pQTL). We used COLOC, currently implemented in FUSION, for this purpose. Five different COLOC hypotheses were tested: no association (posterior probability (PP0), functional association only (PP1), GWAS association only (PP2), independent functional/GWAS associations (PP3), or colocalized functional/GWAS associations (PP4)). We considered plausible causal relationships for those with a posterior PP4 > 0.5.

We considered two criteria to assign significant terms to the PWAS results from the discovery analysis. First, using a stringent criterion, we considered significant PWAS results, those proteins with a significant *q*-value (<0.05) in the discovery PWAS, and a PP4 > 0.5 in the colocalization analysis [28]. Proteins following this criterion were considered in the replication analysis. Second, using an open criterion, we selected for the replication of all the proteins with a nominal *p*-value in the discovery PWAS (*p*-value < 10<sup>-5</sup>) (Figure 4).



**Figure 4.** Schematic study workflow considering the discovery and the replication PWAS but also the sensitivity analyses.

#### 4.3. Replication Analysis

For replication purposes, we selected a confirmation dPFC brain proteome panel obtained in 198 participants of European ancestry from the Banner Sun Health Research Institute [38]. Proteomic profiles were obtained using the same methodology used in the discovery panel [28]. After applying the QCs previously described, a total of 152 patients with proteomic and genetic data were used in the functional weight construction. Finally, proteomic weights included 1139 heritable proteins [28] that were combined with the above-mentioned GWAS summary statistics using the same protocol we used for the discovery panel.

All the proteins evaluated were considered significant in the replication study when the  $q$ -value was <0.05 (adjusting for all the proteins analysed in the replication) and the  $PP4$  was >0.5 (Figure 1).

#### 4.4. Summary-Based Mendelian Randomization (SMR)

We applied SMR to test whether the cis-regulated protein abundance was pleiotropically regulating the relationship between the SNPs and the trait of interest [39]. To perform the analysis for proteins achieving the stringent PWAS criterion, we followed the predetermined criteria of at least one pQTL at a  $p$ -value <  $5 \times 10^{-8}$  and included only SNPs with pQTL  $p$ -values <  $1.5654 \times 10^{-3}$ . SNPs with LD  $r^2$  between the top SNP > 0.9 or < 0.05 were excluded [39]. For the results from the open criterion, we included SNPs with pQTL  $p$ -values < 0.05.

We considered that protein levels were mediating the effect of an SNP on the phenotype when they had a  $p$ -value < 0.05 in SMR.

#### 4.5. Conditional Analysis

For the significant PWAS proteins, we identified SNPs from the GWAS ( $p$ -value < 0.05) in the regions cis-overlapped with the selected proteins. In order to assess whether the significant associations between these polymorphisms and the trait were mediated by the

cis-regulated protein abundance, we applied conditional tests using FUSION [27]. We plotted the results to observe that mediation visually, and we calculated conditional statistics.

**Author Contributions:** I.F.-C. and J.K. were responsible for the study design and the analytical accuracy of the manuscript. N.C. performed the statistical analyses and prepared the manuscript. C.G.-F., J.C.-M., L.L.-C., M.L., E.M., J.M.M.-C., J.M., L.C. and M.A. reviewed the manuscript and contributed to the data interpretation. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was funded by The EPIGENESIS project (Carlos III Institute-PI17/02089, Marató TV3 and Fundació MútuaTerrassa), the MAESTRO project- PI18/01338 (Carlos III Institute), the iBioStroke project (Eranet-Neuron, European research grants), the EPINEXO project-PI20/00678 (Carlos III Institute/Fondo Europeo de Desarrollo Regional (FEDER), the SEDMAN Study (Boehringer Ingelheim), the APHAS project (Pfizer/BMS), and the European Regional Development Fund (ERDF)/Fondo FEDER and 2017SGR-1427 (AGAUR). I. Fernández is the recipient of a research contract from the Miguel Servet Program (CP12/03298) from the Carlos III Institute. 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 by Fons Social Europeu (FSE). M. Lledós is supported by a PFIS Contract (Contratos Predoctorales de Formación en Investigación en Salud): FI19/00309. C. Gallego-Fabrega is supported by a Sara Borrell Contract (CD20/00043) from the Carlos III Institute and Fondo Europeo de Desarrollo Regional (ISCIII-FEDER). E. Muiño is supported by a Río Hortega Contract (CM18/00198) from the Carlos III Institute.

**Institutional Review Board Statement:** The study was conducted in accordance with the Declaration of Helsinki guidelines. Each Institutional Review Board from previously published studies was responsible for approving the various studies included in this manuscript.

**Informed Consent Statement:** The informed consent of the patients was obtained for all the studies included in this article.

**Data Availability Statement:** The datasets used and/or analysed during the current study are available from the corresponding author upon reasonable request.

**Acknowledgments:** We would like to thank the International Stroke Genetics Consortium, the Spanish Stroke Genetics Consortium, the International Stroke Genetics Consortium, and the Global Alliance for Acute and Long-term Outcome Studies in stroke and the RETICS-INVICTUS Network (RD16/0019/0011).

**Conflicts of Interest:** The authors declare that they have no competing interests.

## References

1. Pantoni, L. Cerebral Small Vessel Disease: From Pathogenesis and Clinical Characteristics to Therapeutic Challenges. *Lancet Neurol.* **2010**, *9*, 689–701. [[CrossRef](#)]
2. Uniken Venema, S.M.; Marini, S.; Lena, U.K.; Morotti, A.; Jessel, M.; Moomaw, C.J.; Kourkoulis, C.; Testai, F.D.; Kittner, S.J.; Brouwers, H.B.; et al. Impact of Cerebral Small Vessel Disease on Functional Recovery After Intracerebral Hemorrhage. *Stroke* **2019**, *50*, 2722–2728. [[CrossRef](#)] [[PubMed](#)]
3. Ter Telgte, A.; van Leijssen, E.M.C.; Wiegertjes, K.; Klijn, C.J.M.; Tuladhar, A.M.; de Leeuw, F.-E. Cerebral Small Vessel Disease: From a Focal to a Global Perspective. *Nat. Rev. Neurol.* **2018**, *14*, 387–398. [[CrossRef](#)] [[PubMed](#)]
4. Chung, J.; Marini, S.; Pera, J.; Norrving, B.; Jimenez-Conde, J.; Roquer, J.; Fernandez-Cadenas, I.; Tirschwell, D.L.; Selim, M.; Brown, D.L.; et al. Genome-Wide Association Study of Cerebral Small Vessel Disease Reveals Established and Novel Loci. *Brain* **2019**, *142*, 3176–3189. [[CrossRef](#)] [[PubMed](#)]
5. Woo, D.; Falcone, G.J.; Devan, W.J.; Brown, W.M.; Biffi, A.; Howard, T.D.; Anderson, C.D.; Brouwers, H.B.; Valant, V.; Battey, T.W.K.; et al. Meta-Analysis of Genome-Wide Association Studies Identifies 1q22 as a Susceptibility Locus for Intracerebral Hemorrhage. *Am. J. Hum. Genet.* **2014**, *94*, 511–521. [[CrossRef](#)]
6. Traylor, M.; Malik, R.; Nalls, M.A.; Cotlarciuc, I.; Radmanesh, F.; Thorleifsson, G.; Hanscombe, K.B.; Langefeld, C.; Saleheen, D.; Rost, N.S.; et al. Genetic Variation at 16q24.2 Is Associated with Small Vessel Stroke. *Ann. Neurol.* **2017**, *81*, 383–394. [[CrossRef](#)]
7. Qureshi, A.I.; Mendelow, A.D.; Hanley, D.F. Intracerebral Haemorrhage. *Lancet* **2009**, *373*, 1632–1644. [[CrossRef](#)]
8. Sato, S.; Delcourt, C.; Heeley, E.; Arima, H.; Zhang, S.; Al-Shahi Salman, R.; Stapf, C.; Woo, D.; Flaherty, M.L.; Vagal, A.; et al. Significance of Cerebral Small-Vessel Disease in Acute Intracerebral Hemorrhage. *Stroke* **2016**, *47*, 701–707. [[CrossRef](#)]
9. Kim, B.J.; Lee, S.-H. Prognostic Impact of Cerebral Small Vessel Disease on Stroke Outcome. *J. Stroke* **2015**, *17*, 101–110. [[CrossRef](#)]

10. Regenhardt, R.W.; Das, A.S.; Lo, E.H.; Caplan, L.R. Advances in Lacunar Stroke Pathophysiology: A Review. *JAMA Neurol.* **2018**, *75*, 1273–1281. [[CrossRef](#)]
11. Fornage, M.; Debette, S.; Bis, J.C.; Schmidt, H.; Ikram, M.A.; Dufouil, C.; Sigurdsson, S.; Lumley, T.; DeStefano, A.L.; Fazekas, F.; et al. Genome-Wide Association Studies of Cerebral White Matter Lesion Burden: The CHARGE Consortium. *Ann. Neurol.* **2011**, *69*, 928–939. [[CrossRef](#)] [[PubMed](#)]
12. Haffner, C.; Malik, R.; Dichgans, M. Genetic Factors in Cerebral Small Vessel Disease and Their Impact on Stroke and Dementia. *J. Cereb. Blood Flow Metab.* **2016**, *36*, 158–171. [[CrossRef](#)] [[PubMed](#)]
13. Verhaaren, B.F.J.; Debette, S.; Bis, J.C.; Smith, J.A.; Ikram, M.K.; Adams, H.H.; Beecham, A.H.; Rajan, K.B.; Lopez, L.M.; Barral, S.; et al. Multiethnic Genome-Wide Association Study of Cerebral White Matter Hyperintensities on MRI. *Circ. Cardiovasc. Genet.* **2015**, *8*, 398–409. [[CrossRef](#)] [[PubMed](#)]
14. Traylor, M.; Zhang, C.R.; Adib-Samii, P.; Devan, W.J.; Parsons, O.E.; Lanfranconi, S.; Gregory, S.; Cloonan, L.; Falcone, G.J.; Radmanesh, F.; et al. Genome-Wide Meta-Analysis of Cerebral White Matter Hyperintensities in Patients with Stroke. *Neurology* **2016**, *86*, 146–153. [[CrossRef](#)] [[PubMed](#)]
15. Traylor, M.; Tozer, D.J.; Croall, I.D.; Lisiecka-Ford, D.M.; Olorunda, A.O.; Boncoraglio, G.; Dichgans, M.; Lemmens, R.; Rosand, J.; Rost, N.S.; et al. Genetic Variation in PLEKHG1 Is Associated with White Matter Hyperintensities (n = 11,226). *Neurology* **2019**, *92*, e749–e757. [[CrossRef](#)]
16. Persyn, E.; Hanscombe, K.B.; Howson, J.M.M.; Lewis, C.M.; Traylor, M.; Markus, H.S. Genome-Wide Association Study of MRI Markers of Cerebral Small Vessel Disease in 42,310 Participants. *Nat. Commun.* **2020**, *11*, 2175. [[CrossRef](#)]
17. Sargurupremraj, M.; Suzuki, H.; Jian, X.; Sarnowski, C.; Evans, T.E.; Bis, J.C.; Eiriksdottir, G.; Sakaue, S.; Terzikhan, N.; Habes, M.; et al. Cerebral Small Vessel Disease Genomics and Its Implications across the Lifespan. *Nat. Commun.* **2020**, *11*, 6285. [[CrossRef](#)]
18. Rutten-Jacobs, L.C.A.; Tozer, D.J.; Duering, M.; Malik, R.; Dichgans, M.; Markus, H.S.; Traylor, M. Genetic Study of White Matter Integrity in UK Biobank (N = 8448) and the Overlap with Stroke, Depression, and Dementia. *Stroke* **2018**, *49*, 1340–1347. [[CrossRef](#)]
19. Armstrong, N.J.; Mather, K.A.; Sargurupremraj, M.; Knol, M.J.; Malik, R.; Satizabal, C.L.; Yanek, L.R.; Wen, W.; Gudnason, V.G.; Dueker, N.D.; et al. Common Genetic Variation Indicates Separate Etiologies for Periventricular and Deep White Matter Hyperintensities. *Stroke* **2020**, *51*, 2111–2121. [[CrossRef](#)]
20. Knol, M.J.; Lu, D.; Traylor, M.; Adams, H.H.H.; Romero, J.R.J.; Smith, A.V.; Fornage, M.; Hofer, E.; Liu, J.; Hostettler, I.C.; et al. Association of Common Genetic Variants with Brain Microbleeds: A Genome-Wide Association Study. *Neurology* **2020**, *95*, e3331–e3343. [[CrossRef](#)]
21. Li, H.-Q.; Cai, W.-J.; Hou, X.-H.; Cui, M.; Tan, L.; Yu, J.-T.; Dong, Q. Alzheimer’s Disease Neuroimaging Initiative Genome-Wide Association Study of Cerebral Microbleeds on MRI. *Neurotox. Res.* **2020**, *37*, 146–155. [[CrossRef](#)] [[PubMed](#)]
22. Malik, R.; Chauhan, G.; Traylor, M.; Sargurupremraj, M.; Okada, Y.; Mishra, A.; Rutten-Jacobs, L.; Giese, A.-K.; van der Laan, S.W.; Gretarsdottir, S.; et al. Multiancestry Genome-Wide Association Study of 520,000 Subjects Identifies 32 Loci Associated with Stroke and Stroke Subtypes. *Nat. Genet.* **2018**, *50*, 524–537. [[CrossRef](#)] [[PubMed](#)]
23. Traylor, M.; Persyn, E.; Tomppo, L.; Klasson, S.; Abedi, V.; Bakker, M.K.; Torres, N.; Li, L.; Bell, S.; Rutten-Jacobs, L.; et al. Genetic Basis of Lacunar Stroke: A Pooled Analysis of Individual Patient Data and Genome-Wide Association Studies. *Lancet Neurol.* **2021**, *20*, 351–361. [[CrossRef](#)]
24. Brandes, N.; Linial, N.; Linial, M. PWAS: Proteome-Wide Association Study—Linking Genes and Phenotypes by Functional Variation in Proteins. *Genome Biol.* **2020**, *21*, 173. [[CrossRef](#)]
25. Ou, Y.-N.; Yang, Y.-X.; Deng, Y.-T.; Zhang, C.; Hu, H.; Wu, B.-S.; Liu, Y.; Wang, Y.-J.; Zhu, Y.; Suckling, J.; et al. Identification of Novel Drug Targets for Alzheimer’s Disease by Integrating Genetics and Proteomes from Brain and Blood. *Mol. Psychiatry* **2021**, 1–9. [[CrossRef](#)]
26. Wingo, T.S.; Liu, Y.; Gerasimov, E.S.; Gockley, J.; Logsdon, B.A.; Duong, D.M.; Dammer, E.B.; Lori, A.; Kim, P.J.; Ressler, K.J.; et al. Brain Proteome-Wide Association Study Implicates Novel Proteins in Depression Pathogenesis. *Nat. Neurosci.* **2021**, *24*, 810–817. [[CrossRef](#)] [[PubMed](#)]
27. Malik, R.; Dichgans, M. Challenges and Opportunities in Stroke Genetics. *Cardiovasc. Res.* **2018**, *114*, 1226–1240. [[CrossRef](#)]
28. Wingo, A.P.; Liu, Y.; Gerasimov, E.S.; Gockley, J.; Logsdon, B.A.; Duong, D.M.; Dammer, E.B.; Robins, C.; Beach, T.G.; Reiman, E.M.; et al. Integrating Human Brain Proteomes with Genome-Wide Association Data Implicates New Proteins in Alzheimer’s Disease Pathogenesis. *Nat. Genet.* **2021**, *53*, 143–146. [[CrossRef](#)]
29. Liu, D.-Z.; Sharp, F.R. Excitatory and Mitogenic Signaling in Cell Death, Blood-Brain Barrier Breakdown, and BBB Repair after Intracerebral Hemorrhage. *Transl. Stroke Res.* **2012**, *3*, 62–69. [[CrossRef](#)]
30. Lai, T.W.; Zhang, S.; Wang, Y.T. Excitotoxicity and Stroke: Identifying Novel Targets for Neuroprotection. *Prog. Neurobiol.* **2014**, *115*, 157–188. [[CrossRef](#)]
31. Chagnot, A.; Barnes, S.R.; Montagne, A. Magnetic Resonance Imaging of Blood-Brain Barrier Permeability in Dementia. *Neuroscience* **2021**, *474*, 14–29. [[CrossRef](#)] [[PubMed](#)]
32. Kalaria, R.N.; Sepulveda-Falla, D. Cerebral Small Vessel Disease in Sporadic and Familial Alzheimer Disease. *Am. J. Pathol.* **2021**, *191*, 1888–1905. [[CrossRef](#)] [[PubMed](#)]
33. Bordes, C.; Sargurupremraj, M.; Mishra, A.; Debette, S. Genetics of Common Cerebral Small Vessel Disease. *Nat. Rev. Neurol.* **2022**, 1–18. [[CrossRef](#)]

34. Vuckovic, D.; Bao, E.L.; Akbari, P.; Lareau, C.A.; Mousas, A.; Jiang, T.; Chen, M.-H.; Raffield, L.M.; Tardaguila, M.; Huffman, J.E.; et al. The Polygenic and Monogenic Basis of Blood Traits and Diseases. *Cell* **2020**, *182*, 1214–1231.e11. [[CrossRef](#)] [[PubMed](#)]
35. Van der Meer, D.; Frei, O.; Kaufmann, T.; Shadrin, A.A.; Devor, A.; Smeland, O.B.; Thompson, W.K.; Fan, C.C.; Holland, D.; Westlye, L.T.; et al. Understanding the Genetic Determinants of the Brain with MOSTest. *Nat. Commun.* **2020**, *11*, 3512. [[CrossRef](#)]
36. Gusev, A.; Ko, A.; Shi, H.; Bhatia, G.; Chung, W.; Penninx, B.W.J.H.; Jansen, R.; de Geus, E.J.C.; Boomsma, D.I.; Wright, F.A.; et al. Integrative Approaches for Large-Scale Transcriptome-Wide Association Studies. *Nat. Genet.* **2016**, *48*, 245–252. [[CrossRef](#)]
37. Bennett, D.A.; Buchman, A.S.; Boyle, P.A.; Barnes, L.L.; Wilson, R.S.; Schneider, J.A. Religious Orders Study and Rush Memory and Aging Project. *J. Alzheimer's Dis.* **2018**, *64*, S161–S189. [[CrossRef](#)]
38. Beach, T.G.; Adler, C.H.; Sue, L.I.; Serrano, G.; Shill, H.A.; Walker, D.G.; Lue, L.; Roher, A.E.; Dugger, B.N.; Maarouf, C.; et al. Arizona Study of Aging and Neurodegenerative Disorders and Brain and Body Donation Program. *Neuropathology* **2015**, *35*, 354–389. [[CrossRef](#)]
39. Zhu, Z.; Zhang, F.; Hu, H.; Bakshi, A.; Robinson, M.R.; Powell, J.E.; Montgomery, G.W.; Goddard, M.E.; Wray, N.R.; Visscher, P.M.; et al. Integration of Summary Data from GWAS and eQTL Studies Predicts Complex Trait Gene Targets. *Nat. Genet.* **2016**, *48*, 481–487. [[CrossRef](#)]