Title page

Integrative multi-omics analysis to characterize human brain ischemia

Laura Ramiro*[1], Teresa García-Berrocoso*[1,2,#], Ferran Briansó[3], Leire Goicoechea[1], Alba Simats[1], Víctor Llombart[1], Ricardo Gonzalo[4,5], Alexandre Hainard[6], Elena Martínez-Saez[7], Francesc Canals[8], Jean-Charles Sanchez[9], Alex Sánchez-Pla[4], Joan Montaner[1,10].

1. Neurovascular Research Laboratory, Vall d'Hebron Institute of Research (VHIR), Universitat Autònoma de Barcelona, Barcelona, Spain.

2. CSIC/UAB Proteomics Laboratory, Institute of Biomedical Research of Barcelona, Spanish National Research Council (IIBB-CSIC/IDIBAPS), Barcelona, Spain.

3. Genetics, Microbiology and Statistics Department, Universitat de Barcelona, Barcelona, Spain.

4. Statistics and Bioinformatics Unit, Vall d'Hebron Institute of Research (VHIR), Universitat Autònoma de Barcelona, Barcelona, Spain.

5. High Technology Unit, Vall d'Hebron Institute of Research (VHIR), Universitat Autònoma de Barcelona, Barcelona, Spain.

6. Proteomics Core Facility, Faculty of medicine, University Medical Center, University of Geneva, Geneva, Switzerland.

7. Neuropathology, Pathology department, Hospital Vall d'Hebron, Universitat Autònoma de Barcelona, Barcelona, Spain.

8. Proteomics Laboratory, Vall d'Hebron Institute of Oncology (VHIO), Universitat Autònoma de Barcelona, Barcelona, Spain.

9. Translational biomarker group, Internal Medicine department, University Medical Center, University of Geneva, Geneva, Switzerland.

10. Institute de Biomedicine of Seville, IBiS/Hospital Universitario Virgen del Rocío/CSIC/University of Seville & Department of Neurology, Hospital Universitario Virgen Macarena, Seville.

* These authors contributed equally to this work.

Corresponding author at: Neurovascular Research Laboratory, Vall d'Hebron Institute of Research (VHIR), Hospital Vall d'Hebron, Pg. Vall d'Hebron 119-129, 08035, Barcelona, Spain. teresa.garcib@gmail.com, +34 93 489 4029.

Acknowledgments

Under a collaborative agreement, Affymetrix – ThermoFisher Scientific kindly supplied the gene expression microarrays used in this study without being involved in any part of the study development.

We thank Dr. Pilar Delgado for providing blood samples from the ISSYS cohort.

Microarray sample processing and hybridization were carried out at the High Technology Unit (UAT) from VHIR. We acknowledge the Molecular Diagnosis Platform staff for their contribution to this project's development. Bioinformatics analysis has been carried out in the Statistics and Bioinformatics Unit (UEB) at Vall d'Hebron Research Institute (VHIR)

Abstract

Stroke is a major cause of death and disability. A better comprehension of stroke pathophysiology is fundamental to reduce its dramatic outcome. The use of high-throughput unbiased *omics* approaches and the integration of these data might deepen the knowledge of stroke at the molecular level depicting the interaction between different molecular units. We aimed to identify protein and gene expression changes in the human brain after ischemia through an integrative approach to join the information of both *omics* analyses. The translational potential of our results was explored in a pilot study with blood samples from ischemic stroke patients.

Proteomics and transcriptomics discovery studies were performed in human brain samples from six deceased stroke patients comparing the infarct core with the corresponding contralateral brain region, unveiling 128 proteins and 2716 genes significantly dysregulated after stroke. Integrative bioinformatics analyses joining both datasets exposed canonical pathways altered in the ischemic area, highlighting the most influential molecules. Among the molecules with the highest fold-change, 28 genes and 9 proteins were selected to be validated in five independent human brain samples using orthogonal techniques. Our results were confirmed for NCDN, RAB3C, ST4A1, DNM1L, A1AG1, A1AT, JAM3, VTDB, *ANXA1*, *ANXA2* and *IL8*. Finally, circulating levels of the validated proteins were explored in ischemic stroke patients. Fluctuations of A1AG1 and A1AT, both up-regulated in the ischemic brain, were detected in blood along the first week after onset. In summary, our results expand the knowledge of ischemic stroke pathology, revealing key molecules to be further explored as biomarkers and/or therapeutic targets.

Keywords: Ischemic stroke, proteomics, transcriptomics, integrative analysis

Declarations

Funding

This work has been funded by Instituto de Salud Carlos III (PI15/00354, PI18/00804), MINECO (MTM2015-64465-C2-1R) and GRBIO (2014-SGR-464) and co-financed by the European Regional Development Fund (FEDER). Neurovascular Research Laboratory takes part in the Spanish stroke research network INVICTUS+ (RD16/0019/0021). L.R is supported by a pre-doctoral fellowship from the Instituto de Salud Carlos III (IFI17/00012).

Conflicts of interest/Competing interests

None

Availability of data and material

The mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium via the PRIDE partner repository with the dataset identifier PXD022850. Microarrays raw data can be accessed through the Gene Expression Omnibus (GEO) data repository with the accession number GSE162955.

Ethics approval

The whole study was approved by the Ethics Committee of Vall d'Hebron Hospital (PR[HG]85/04, PR[HG]89/03 and PR[IR]87/10).

Consent to participate

Written informed consent was acquired from all participants or relatives in agreement with the Declaration of Helsinki.

Abbreviations

BCA: bicinchoninic acid

CL: corresponding contralateral brain area

CV: coefficient of variation

FDR: false discovery rate

GEO: Gene Expression Omnibus

GIS: gene influential score

GPF: gas phase fractionation

GSS: gene set scores

IC: infarct core

LC-ESI-MS/MS: liquid chromatography coupled to electrospray ionization – tandem mass spectrometry

LogFC: Logarithmic base 2 fold-change

LTQ: linear trap quadrupole

MCAO: middle cerebral artery occlusion

moGSA: multiple omics gene set analysis

mRS: modified Rankin Scale

NIHSS: National Institutes of Health stroke scale

RQ: Relative quantification

rt-PA: recombinant tissue plasminogen activator

1. Introduction

Stroke is one of the most frequent causes of morbidity and mortality worldwide [1]. The currently approved therapy consists of restoring cerebral blood flow via the intravenous administration of thrombolytics such as recombinant tissue plasminogen activator (rt-PA), or through mechanical thrombectomy [2]. Beyond these therapeutic approaches at the vascular level, the use of neuroprotective drugs to prevent infarct growth is still far from being applied in daily clinics and needs further research [3, 4]. In addition to these therapeutic interventions, it is indispensable to have rapid diagnostic techniques to optimize stroke patients' management. In this line, blood biomarkers are thought to be promising tools to complement the current clinical methods used to diagnose stroke, which are still based on patients' medical history and neurological and neuroimaging explorations [5].

On this basis, a better understanding of stroke pathophysiology would help to improve preventive, diagnostic and therapeutic strategies. Thanks to the rapid development of *omics* techniques, various studies [6–9] have taken advantage of these high-throughput approaches to identify molecular pathways altered after ischemic stroke, improving enormously the knowledge on stroke pathophysiology [6, 10]. Several proteomics, transcriptomics, genomics and metabolomics studies have been published in the field of ischemic stroke. For example, the brain proteome after ischemic stroke has been described [7] and key molecules in stroke pathology such as matrix metalloproteinases (MMPS) [11] or glial fibrillary acidic protein (GFAP) [12], among others, have been identified thanks to these approaches. Probably in the near future, MMPS measurement will allow identifying ischemic stroke patients with a high risk of suffering hemorrhagic transformation [13], while GFAP could be used to discriminate between hemorrhagic and ischemic stroke patients [14, 15]. Besides, thanks to genomics studies, 42 loci associated with stroke at a genome-wide significant level have been identified to date [6] and together with different proposals of gene transcripts panels could help to diagnose ischemic stroke in a near future [16–18]. However, the vast majority of these studies focus on the analysis of a single *omics* dataset, missing and underestimating the crosstalk and interplay that occur between the different molecular components.

For that reason, a comprehensive integrative analysis of multilevel *omics* data would enable to decipher the molecular changes triggered by stroke, ultimately leading to the identification of the key players of this complex disease and even exposing altered

pathways that are not revealed when examining a single data source [19, 20]. Little is known about *omics* integration approaches in the context of ischemic stroke. Heterogeneous data and numerous sources of information are difficult to integrate in a global analysis, making challenging to discern between biologically relevant and irrelevant molecules [21]. Nonetheless, recently, a study integrating for the first time proteomic and transcriptomic data obtained from mouse brains at 2 hours after cerebral ischemia was published by our group [22]. However, to date, any study of these characteristics has been performed in stroke patients.

Thus, this study aimed to massively identify and verify changes that occur in the human brain after ischemic stroke at different molecular levels. Specifically, we were interested in exploring the main proteomics and transcriptomics alterations as well as joining all this information through an integrative analysis to interpret the results in their biological context. Finally, with a more translational aim in view, we also explored the blood circulating levels of some of the validated candidates in ischemic stroke patients at different time-points within the acute phase of the disease.

2. Materials and methods

All reagents were purchased from Sigma-Aldrich (USA) unless contrary stated.

Experimental design

The present study is divided into 3 main sections: a first Discovery Phase performed in human brain samples using mass spectrometry and RNA microarrays, with further integrative bioinformatics analyses; a second Replication Phase for selected candidates conducted in different/independent human brain samples employing Western Blot and qRT-PCR; and a third Qualification Phase developed in human blood samples using ELISA techniques. A flowchart summarizing all these steps can be seen in Figure 1.

Individuals and sample collection

Ethics statement

The whole study was approved by the Ethics Committee of Vall d'Hebron Hospital (PR[HG]85/04, PR[HG]89/03 and PR[IR]87/10). Written informed consent was acquired from all participants or relatives in agreement with the Declaration of Helsinki.

Brain samples

Eleven ischemic stroke patients who died during hospitalization in the Vall d'Hebron Hospital (Barcelona, Spain) were included in our brain tissue collection between March 2004 and July 2011. Tissue collection was done under the supervision of an experienced neuropathologist. Brain pieces from the infarct core (IC) and the corresponding contralateral (CL) areas were gathered within the first hours after death, snap-frozen in liquid nitrogen and stored at -80ºC. Table 1 summarizes the demographic and clinical information of these patients.

Blood samples

Patients with symptoms of acute ischemic stroke admitted to the emergency department of the Vall d'Hebron Hospital within the first 4.5 hours after symptoms onset were prospectively recruited from September 2006 to January 2009. All patients underwent a standardized protocol of brain imaging and neurological assessment. Evaluation of neurological severity using the National Institutes of Health stroke scale (NIHSS) [23], of functional state with the modified Rankin Scale (mRS) [24] and etiological classification following the TOAST definitions [25] were carried out. All patients received the standard thrombolytic treatment (intravenous 0.9 mg/Kg recombinant tissue-plasminogen activator, rt-PA; Actilyse, Boehringer Ingelheim International GmbH, Germany). From this cohort, anonymized samples from 11 stroke patients (5 males and 6 females, age and sex matched) were randomly selected (see Supplementary Table 1 for demographic and clinical details).

Peripheral blood samples were drawn on admission (before any treatment was administered), 24 hours and 1 week later. EDTA plasma was separated by centrifugation at 1,500 *g* for 15 min at 4ºC and stored at -80ºC. Additionally, plasma samples from 5 subjects free from brain lesions from the ISSYS cohort [26] were included as control reference (40% males, 67.4 ± 1.1 years old).

Proteomics analysis

Sample preparation

Alkaline cold lysis buffer containing 50 mM Tris-HCl, 150 mM NaCl, 5 mM $CaCl_2$, 0.05% Brij-35 and 0.02% NaN3 was used to prepare 0.1% Rapigest (Waters, USA) solution. Protease inhibitors (1% phenylmethylsulfonyl fluoride – PSMF – and 0.5% aprotinin) were added. Frozen brain samples from IC and CL areas were immersed in this

solution and homogenized with a dounce tissue grinder with glass pestles. Homogenates were centrifuged for 12 min at 15,300 *g* at 4ºC and supernatants were stored at -80ºC.
The protein content of brain homogenates was determined by the bicinchoninic acid (BCA) assay (ThermoFisher Scientific, USA). For each sample, 25 µg total protein was brought to 30 µL 0.1% Rapigest in 0.1 M triethylammonium bicarbonate (TEAB) before a standard reduction, alkylation, trypsin digestion and sample processing was carried out [27].

LC-MS/MS analysis

Brain digests were analyzed by liquid chromatography (NanoAcquity LC system; Waters) coupled to electrospray ionization – tandem mass spectrometry (LC-ESI-MS/MS) on a linear trap quadrupole (LTQ) Orbitrap Velos Pro (ThermoFisher) following a gas phase fractionation (GPF)-4 approach. Details about the procedure can be found elsewhere[27].

Protein identification and quantitative analysis

Mass spectrometry data were analyzed with Progenesis LC-MS® software v4.0 (Nonlinear dynamics, UK) using default settings and automatic processing of the runs. Each GPF range was analyzed independently, as previously described [27], before combining them into one single experiment. For those proteins identified with a minimum of 2 unique peptides, protein abundance was quantified in a label-free manner based on the sum of all peak areas for each peptide ion and normalized to all proteins.
Further normalization was done by adding a value of 1 to every data point (to avoid 0 values in the analysis for missing proteins in some brain samples and/or areas). Analysis of data heterogeneity showed a skewed distribution, with the top 3 highly abundant proteins interfering in the distribution of the data; these 3 proteins corresponded to hemoglobin α and β and albumin and were excluded from further analysis. Finally, data were log-transformed and used for a statistical paired analysis to compare IC and CL areas based on a linear model (limma package; R software, Austria). Furthermore, the Benjamini-Hochberg procedure to calculate the false discovery rate (FDR; multtest R package) was applied to obtain q-values. We considered differentially abundant proteins those with a q-value <0.1. Logarithmic base 2 fold-change (logFC) was computed to show the average protein changes in the IC with respect to the CL areas; a minus sign (-) indicates lower protein abundance in the IC area, and vice versa.

Transcriptomics analysis

Sample preparation

Total RNA was isolated from frozen brain samples using the Fatty Tissue RNA Purification kit (Norgen Biotek Corp., Canada), following the manufacturer's instructions. DNase treatment was performed in solution with the DNA-free DNA removal kit (Ambion, ThermoFisher). All RNAs were kept at -80ºC.

The integrity of the isolated RNA was determined using the Bioanalyzer 2100 platform (Agilent, UK), with an average integrity number of 6.4 ± 1.5.

Microarrays analysis

For each sample, 150 ng total RNA was amplified and transformed to biotinylated sense-strand DNA with the GeneChip® WT PLUS Reagent Kit (Affymetrix, ThermoFisher). The appropriate GeneChip Hybridization, Wash, and Stain Kit (Affymetrix) were employed before hybridization onto GeneChip Human Transcriptome Array 2.0 cartridges (Affymetrix) following the manufacturer's instructions.

Microarrays were scanned in a GeneChip scanner 3000 7G (Applied Biosystems, ThermoFisher) and these images were processed with the GeneChip Command Console software (Affymetrix) to obtain .CEL files containing gene expression intensities. Data quality was assessed with the arrayQualityMetrics R package [28] before further analyses were conducted without the exclusion of any sample.

Differential gene expression analysis

Raw gene expression intensities were processed using the Robust Multi-array Average algorithm [29], which accounts for background correction, normalization and summarization of probe set values for each gene. Besides, non-specific filtering was conducted to remove genes with low intensity and/or low variability among samples. Only genes coding for proteins were considered for further analysis in this study. A linear model analysis (limma R package) was used for a statistical paired analysis to compare IC and CL brain areas, together with further FDR multiple testing adjustment (multtest R package) to give q-values. We considered differentially expressed genes those with q-values <0.1. LogFC was computed to show the average gene alterations in the IC in relation to the CL areas; a minus sign (-) applies for genes underexpressed in the IC area, and vice versa.

Integrative bioinformatics analyses

Integrative bioinformatics analysis was developed using proteomics and transcriptomics data obtained from the discovery phase comparing paired infarcted and contralateral brain areas from same patients. Filtered lists of the most relevant proteins and genes from both *omics* datasets (q-value <0.1 and |logFC|>1.5) were combined using appropriate multivariate methods for dimensionality reduction to find common expression trends. RV coefficients were obtained as a measure of global similarity between protein and gene datasets. Co-inertia analysis [30] (made4 R package, default parameters) yielded patterns to explain the maximum covariance between global protein and gene datasets, whereas regularized canonical correlations [31] (mixOmics R package, using *seq(0.1, 2,length=50)* and *seq(0.000001,0.002,length=20)* for tuning parameters λ1 and λ2 respectively) inferred relevance networks based on the relationship between particular proteins and genes. In this case, the cut-off to highlight the most relevant correlations was set at R ≥0.9.

Additionally, further multidimensional exploration of both *omics* datasets was completed with biological annotations by unsupervised multiple *omics* gene set analysis [32] (moGSA R package using *inertia* with *statis* parameters for the *mogsa* function). Proteins and genes were weighted against gene sets from the Canonical Pathways sub-collection of the Molecular Signature database (MSigDB [33], file *msigdb.v5.2.symbols.gmt*, currently available at https://www.gsea-msigdb.org/gsea/downloads_archive.jsp, from Broad Institute Inc., USA) to compute gene set scores (GSS), which were further decomposed to evaluate the contribution of proteins and genes to the overrepresentation of a particular gene set under brain ischemia. Gene sets were considered significantly relevant in the disease context when most GSS showed a p-value <0.1. Moreover, the gene influential score (GIS) was calculated to elucidate which molecules were most relevant for a given GSS. A maximum GIS equal to |1| indicates that the molecule contributes a high proportion to the overall GSS; only those molecules with GIS >|0.5|were considered as relevant.

Replication phase

From the lists of meaningful proteins and genes dysregulated after ischemic stroke in the human brain, top molecules (considering its |logFC|) both up- and down-regulated were selected for further independent replication in independent samples using orthogonal methodologies (Table 1). In addition to statistical significance (q-values<0.1 and top

|logFC|), selection criteria were complemented by a comprehensive review of related bibliography on candidates' previous knowledge.

Western Blot

Frozen brain samples were homogenized by means of a dounce tissue grinder using cold lysis buffer containing 50 mM Tris-HCl, 150 mM NaCl, 5 mM CaCl2, 0.05% Brij-35, 0.02% NaN3, 1% Triton X-100, 1% PSMF and 0.5% aprotinin. After centrifugation (12 min at 15,300 *g* at 4ºC), supernatants were collected and stored at -80ºC. Before use, the protein content was determined by BCA assay.

Following standard procedures, 12 µg of brain homogenates mixed with 2X Laemmli buffer (Bio-Rad, USA) were resolved by 10-14% sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE) under reducing conditions. Afterward, separated proteins were transferred onto nitrocellulose membranes (GE Healthcare, UK) using a mini Trans-Blot Electrophoretic Transfer Cell (Bio-Rad) for 1h at 100V. Membranes were stained with Ponceau S solution for the detection of total protein content in each sample and were scanned in a Perfection V39 scanner (Epson, Japan). Non-specific binding was blocked for 1h with 10% non-fat milk before membranes were incubated overnight at 4ºC with optimized anti-human antibodies against protein candidates (Supplementary Table 2). Appropriate secondary antibodies conjugated to horseradish peroxidase (GE Healthcare) were incubated at room temperature for 1h. The substrate reaction was developed with peroxide and luminol solutions (ThermoFisher) and visualized with an Odyssey Fc imaging system (LI-COR Biosciences, USA).

Protein bands and total protein lane images were quantified using Image-J free software. Positive band signal was corrected by total loading signal to calculate the candidate protein level in each sample [34].

qRT-PCR

Total RNA was isolated from frozen brain samples using the Fatty Tissue RNA Purification kit (Norgen), including in-column DNase treatment with the RNase-Free DNase I kit (Norgen). Before freezing samples at -80ºC, successful RNA extraction was checked by NanoDrop® ND-1000 spectrophotometer (Nucliber, Spain).

Reverse transcription for cDNA synthesis was performed using the High-Capacity cDNA Reverse Transcription kit (Applied), following the manufacturer's instructions. A standard real-time quantitative PCR amplification was carried out in 384-well plates, with samples run in triplicates and using an unrelated brain sample as endogenous calibrator

control in all the plates. We selected TaqMan™ Gene Expression Assays (Applied) for each candidate gene (Supplementary Table 3) and for *PPIA* (Hs99999904_m1), which was used as a housekeeping gene to normalize the results. Plates were analyzed using the 7900HT Fast Real-Time PCR system (Applied) and SDS 2.4 and RQ Manager softwares (Applied). Relative quantification (RQ) values were calculated by using the Livak equation: $RQ = 2^{-\Delta\Delta Ct}$, using both calibrator and housekeeping controls.

Statistical analyses

Protein and gene quantifications from the replication phase were analyzed using SPSS® 20.0 (IBM Corp., USA). Considering a paired samples design, differences between IC and CL brain areas were assessed by Wilcoxon signed-rank test. In all cases, p-values <0.05 were considered significant. LogFC were calculated as detailed above.

Qualification study

For those proteins validated in the replication phase, a pilot exploration in blood samples was conducted.

ELISAs

Commercial immunoassays were used following manufacturer's instructions to quantify plasma levels of A1AG1 (1:100 dilution; MBS763331, MyBioSource), A1AT (1:20 dilution; ELH-SerpinA1, RayBiotech), DNM1L (undiluted; MBS9358522, MyBioSource), JAM3 (undiluted; ELH-JAMC, RayBiotech, USA), NCDN (1:2 dilution; abx392242, Abbexa), RAB3C (undiluted; MBS9317641, MyBioSource), ST4A1 (undiluted; MBS9339095, MyBioSource) and VTDB (1:50 dilution; MBS763939, MyBioSource, USA). Optical density was measured using a Synergy™ Mx microplate reader (BioTek Instruments Inc, USA).

Duplicates were assayed for each sample and the mean value was used for further analyses, excluding those samples with an intra-assay coefficient of variation (CV) higher than 20%. In the case of protein values under the detectable range of the assay, the limit of the detection value of the assay was consigned. Besides, protein values over the detectable range were given the highest readable value in each particular assay.

Statistical analyses

Analyses were performed with SPSS 20.0. Longitudinal analysis of protein circulating levels over time after stroke was conducted with the Friedman test. Differences between time-points were assessed with the Wilcoxon signed-rank test corrected by Bonferroni.

Reference protein values in the control population were calculated with Tukey's hinges. P-values <0.05 were considered significant at 95% confidence.

3. Results

A chart of the study workflow is shown in Figure 1.

Protein and gene dysregulation in the brain after stroke

The unbiased molecular analysis of human brain samples from six deceased stroke patients through mass spectrometry and microarrays techniques allowed the identification and quantification of 1902 proteins and 6822 genes (Supplementary Data 1). After cleaning the datasets and focusing attention on proteins identified with at least two unique peptides and protein-coding genes, we determined the differential proteome and transcriptome between paired IC and CL brain areas from the same patients. Statistical analyses revealed 128 proteins and 2716 genes significantly dysregulated under ischemia (i.e. FDR q-values <0.1), with 38 common molecules among these lists (details about q-values and logFC are given in Supplementary Table 4).

Integrative multidimensional molecular landscape of stroke

To obtain more meaningful molecular information, further filtering in both datasets was accomplished by applying a cut-off of |logFC|>1.5. Thus, 95 proteins and 244 genes were deemed fit for integrative purposes. Data blending of these significantly dysregulated proteins and genes showed a global similarity on the profiles of both datasets of 60% (RV coefficient = 0.608). Based on co-inertia analysis, the projection of both molecular datasets identified a greater correlation between protein and gene expression levels in the CL than in the IC area. The IC region, instead, showed a higher molecular divergence together with the expected spread of individual variation (Figure 2A). Additionally, a projection-based approach allowed us to identify trends of co-expression of proteins and genes and to highlight the key molecular players in human brain ischemia. Different members of several molecular families involved in canonical processes typically associated with stroke appeared among the highest-correlated molecules ($R \geq 0.9$); e.g. annexins, GABA receptors and neurofilaments (Figure 2B).

Besides, the multiple *omics* data-based gene set analysis showed that genes contributed almost double than proteins in the molecular variance between IC and CL areas (Figure 2C). Seven gene sets were found significantly overrepresented under ischemia (Figure

2D). Among these canonical pathways, protein changes were highly relevant for the up-regulation of platelets activity (GSS = 0.004) and extracellular matrix composition (known as matrisome; GSS = 0.003). Also, dysregulated genes outweighed the down-regulation of chemical synaptic transmission (GSS = -0.017) and other related neuronal system processes (GSS = -0.021). Genes and proteins performed oppositely in regards to their contribution to the up-regulation of hemostasis (GSS = 0.003) and signaling through G proteins (GSS = 0.001) but similarly in the down-regulation of the neurotransmitter release (GSS = -0.006). The driver molecules in each significant gene set together with their GIS values are depicted in the corresponding panel of Figure 2E.

Replication study

From the lists of meaningful proteins and genes dysregulated after ischemic stroke in the human brain, 28 genes and 9 proteins (Table 2) were selected as candidates among the differential molecules with the highest logFC when IC and CL were compared. In addition to statistical values (q-values<0.1 and top |logFC|), selection criteria were complemented by a bibliographic review on candidates' previous knowledge. A further independent replication was conducted following the same design of paired brain areas in a new set of 5 brain stroke samples by means of Western Blot and qRT-PCR. From all the molecules selected, five molecules (NCDN, RAB3C, ST4A1, DNM1L and MYPR) were common in both datasets and were therefore selected to be replicated at both protein and gene levels. Three out of these five candidates (NCDN, RAB3C and ST4A1) were confirmed to display a globally reduced expression in the IC of a new series of samples. The results of DNM1L were only replicated regarding protein changes showing decreased expression in the IC, while MYPR was not validated at all (Table 2, Figure 3A).

Moreover, four candidates were selected to be replicated exclusively at the protein level (A1AG1, A1AT, JAM3 and VTDB) (Table 2, Figure 3B). Successfully, all four were validated, showing increased expression in the IC. Finally, 23 candidates were selected to be replicated only at the gene level. Of these, three (*ANXA1*, *ANXA2* and *IL8*) were validated displaying increased levels in the IC. We also observed a trend in up-regulation for *SRGN* (Table 2, Figure 3C).

Qualification study

With a translational aim in view, further exploration of the protein verified candidates in blood samples from 11 independent acute ischemic stroke patients treated with rt-PA (median age 71 (63.5-79) years, 45.5% males) was conducted in a pilot experiment using

ELISA techniques to determine whether they were detectable in the blood (see Supplementary Table 1 for demographic and clinical data). Protein levels were assessed in blood at hospital admission, 24 hours and 1 week after the event. Here, longitudinal analyses at these three time points exposed different profiles of protein fluctuations. Baseline reference values were depicted by using blood samples from 5 control subjects without brain damage. Globally, significant changes over time were detected for A1AG1 and A1AT blood levels (Friedman test: p=0.018 and p=0.050 respectively) whereas a slight trend appeared for DNM1L (Friedman test: p=0.072). VTDB, RAB3C, ST4A1, JAM3 and NCDN were also present in the blood without a significant variation in their levels in the acute phase of ischemic stroke (Friedman test: p=0.264, p=0.122, p=0.202, p=0.121 and p=0.159, respectively) (Figure 4).

Deeper analysis showed that A1AG1 displayed higher circulating levels on admission with a quick and sustained remission from 24h whereas A1AT reverted to normal levels one week after the ischemic event. Moreover, a modest reduction of RAB3C, JAM3 and ST4A1 in plasma levels at 24h was revealed (Figure 4).

4. Discussion

The study presented here identifies key molecular pathways altered in the human brain after stroke through an integrative analysis of the main transcriptomics and proteomics changes triggered by cerebral ischemia. To do so, we have combined for the first time in human brain samples information of both *omics* datasets and created a complex network of inter-connecting genes and proteins that may play an important role in stroke pathophysiology. The integrative analysis of our proteomics and transcriptomics data revealed a 60% of resemblance between datasets, disclosing important crosstalk between the different molecular features in response to the ischemic event. The other 40% molecular divergence suggested that proteins and genes may also bring distinct and complementary information about the underlying stroke mechanisms. In this regard, several biological reasons could be considered. Among others, differential post-transcriptional mechanisms, the tight regulation of translation, the existence of regulatory proteins and differences in the turnover of proteins and mRNAs could be factors highly influencing the observed divergence [35, 36]. Besides, many RNAs are actively transcribed but not translated, a fact that can partially explain, why there are more

dysregulated genes than proteins (40% vs 7%) [37], reinforcing the importance of performing studies integrating *omics* data such as the one presented here.

Interestingly, the co-inertia analysis revealed that gene and protein expression were less synchronized in the IC than in the CL, which might be related to the well-known burst of molecular changes arising after stroke and mainly triggered in the hypoperfused area [38]. Moreover, the variation in the IC area appeared to be individual-dependent and might be highly influenced by inter-patient variability and/or differences in the elapsed time from stroke onset to death. Beyond this observed variance, there are common changes that clearly define the IC region. Theoretically, a high statistical correlation between the abundance level of molecules might be interpreted as a functional relationship [39]. In fact, some molecules from several families previously described to play an important role in stroke showed the highest correlation coefficients in our correlation-based analysis, such as the GABA receptors, which are well known to be involved in neuronal excitability and contribute to the consequent neuronal death [40].

The analysis of main molecular functions and biological processes altered due to stroke exposed a variety of canonical pathways that have been previously described, reinforcing the robustness of our results. In general, proteins and genes performed similarly, being both datasets dysregulated in the same direction in the vast majority of pathways. For example, the down-regulation in the ischemic region of the neurotransmitter release, the neuronal system activity and the transmission across chemical synapses were depicted. Due to the high energy demand of the synapse, the lack of glucose and oxygen derived from the blockage of cerebral blood flow might impair synapses function and viability, ultimately altering the release of neurotransmitters and even leading to neuronal death [41]. Moreover, genes and proteins related to the matrisome were also found to be up-regulated. Several changes occur in the extracellular matrix composition of the brain after stroke. While some proteins of the blood-brain-barrier (BBB) are degraded, new extracellular matrix proteins are deposited into the brain parenchyma. However, the biological significance of their regulation may be different depending on the time point after the event and thus contribute to the BBB disruption, the inflammatory response or the remodeling and repairing processes in the parenchyma [42–44]. For example, the overexpression of TGF-β after stroke contributes to remodeling processes [44], and it is one of the most influential molecules of the matrisome pathway in our study, together with other interesting candidates (Figure 2E). All in all, our canonical pathway analysis

brought to light several pathways altered after the ischemic event that might be key players of stroke pathophysiology and need to be taken into consideration to be further explored.

Beyond the integrative data analysis, from the 9 proteins and 28 genes selected for replication 6 genes and 8 proteins were verified. Some of the validated molecules resulted to be among the most influential of the overrepresented canonical pathways in the ischemic brain, such as JAM3 and SRGN which participate in hemostasis processes, or annexins (ANXA1 and ANXA2) that play a role in the matrisome regulation. Furthermore, it is worth noting that the genes and proteins found to be altered after ischemic stroke in this study are interesting molecules to be further studied as potential therapeutic targets for stroke management. As an example, several studies show that the modulation of DNM1L as well as ANXA2, two of the molecules validated in the present study, protects the brain after cerebral ischemia. In brief, DNM1L is a GTPase enzyme implicated in mitochondrial division and distribution, vesicle endocytosis and mitochondria-related necrosis and apoptosis [45], and its dysregulation can cause energy production disruption leading to cell death [46]. In agreement with our results, previous studies showed a reduction of DNM1L in the brain 24 hours after cerebral ischemia in rats [46]. Interestingly, it has been also shown that the inhibition of DNM1L provides neuroprotection both in *in vitro* and *in vivo* stroke models, by reducing the infarct volume [47, 48]. Regarding Annexin 2, it was found up-regulated in the infarcted brain region of our cohort of stroke patients. Briefly, ANXA2 is a calcium regulated phospholipid binding protein involved in cell cycle regulation, cell division, proliferation, cell survival and neo-angiogenesis [49, 50] that can increase the catalytic efficiency of rt-PA about 60-fold [51, 52]. Various studies pointed out that the administration of ANXA2 in combination with rt-PA after experimental ischemic stroke, reduced hemorrhagic transformation risk and the infarct volume as well as improved the functional recovery after stroke [52–55]. Considering the encouraging results available regarding DNM1L and ANXA2 as potential therapeutic targets, further exploration and/or modulation of these and other dysregulated molecules presented in this study could shed light on the search for an appropriate treatment for ischemic stroke. It is worth mentioning stroke is an extremely complex process, so it seems plausible that targeting a simple pathway may not be sufficient to attenuate brain damage [56]. To that end, the simultaneous modulation of the most influential molecules and/or canonical pathways presented here through the

combination of neuroprotective agents might be an interesting approach to achieve optimal therapies in forthcoming studies.

With the idea of a future translation into the clinics of our findings, we explored whether the successfully replicated proteins that were deregulated in the brain were also detectable in the peripheral circulation. If so, these molecules could be promising candidates to be further studied as potential blood biomarkers for stroke diagnosis and/or prognosis in the future. To the best of our knowledge, it is the first time that ST4A1, RAB3C and NCDN, proteins almost exclusively expressed in the brain, are detected in the blood. Interestingly, A1AT and A1AG1 circulating levels showed a substantial increase within the first 24 hours from stroke onset, and returned to normal 1 week after the ischemic event. In fact, plasma concentrations of these acute-phase response proteins were already known to raise several folds in response to acute inflammatory processes, including cerebral ischemia [57]. A1AG1 is thought to suppress the immune response and to have pro-angiogenic properties [58], while A1AT inhibits neutrophil elastases, and regulates inflammation and proteostasis [59]. In the present study, we found significant overexpression of both A1AG1 and A1AT in the IC in comparison with the CL area, complementing what we previously found in an independent brain samples set [7]. Two main factors could be influencing this observed increase in the infarcted region. On the one hand, as shown in the temporal circulating profile, after stroke there is a massive inflammatory response that consequently increases the expression of acute-phase response proteins in the blood, which can reach the brain parenchyma due to the stroke-induced BBB disruption [38]. On the other hand, resident brain cells could also be overexpressing these proteins in response to stroke. In fact, brain endothelial cells are known to express A1AG1 [58],so the observed A1AG1 increased levels of could be reflecting the urge to reestablish the normal blood flow through the formation of new vessels to ensure oxygen availability in the damaged brain tissue. All the same, the fact that all the selected candidates have been detectable in circulation reinforces the idea that they, and others identified here, can be further explored to determine their plausible role as stroke biomarkers.

Altogether, we have integrated for the first time data from proteomics and transcriptomics techniques through an innovative biostatistical approach to identify new key players in human stroke pathophysiology. On this basis, the dysregulated molecules presented in this study might be interesting candidates to become blood biomarkers and to help in the diagnosis of stroke or to predict patients' outcome. Moreover, these molecules and the

canonical pathways in which they have a role might also be further explored in the future as potential therapeutic targets to mitigate or even reverse stroke pathology.

The present study has some limitations that need to be considered and should be addressed in future studies. To begin with, brain samples used in our study were from deceased patients, and the time elapsed from death to sample collection was on average 8h (Table 1). Over this period of time, some molecules could be degraded or even modified. Moreover, the affected brain region and the time from stroke onset to death were different among patients, variables that can be influencing protein and gene expression. However, with this limitation in mind, we made use of the contralateral hemisphere as the respective individual control and thus maintain the inherent individual variability of this disease. Besides, the number of patients used in the discovery, replication and qualification phases was relatively small, so although our results are consistent with previous studies, further research in larger cohorts has to be conducted. Moreover, new studies exploring blood circulating levels of the candidates should include stroke mimics subjects to determine the plausible role of these proteins as biomarkers for stroke diagnosis. Nonetheless, several strengths have to be also highlighted. Foremost, the study presented here provides novel information about changes in protein and gene expression that occur in the brain after ischemic stroke.

During the following years, it is highly likely that integrative approaches like the one presented here will emerge and grow, shedding light on the understanding of complex diseases such as stroke. In this regard, the integration of other *omics* techniques such as genomics or metabolomics could complement and enrich even more the present study, providing additional information that could aid in depicting the complex stroke phenotype. Integration of omics data could also become a crucial tool in the development of personalized therapies through the alignment of clinical phenotypes with multilevel molecular networks, easing the identification of biological signatures of clinical manifestations [6]. These methods will ultimately enable the triage of targets in the design of diagnostic, prognostic and therapeutic approaches. Besides, future research could focus on creating *in silico* models of stroke through the integration of these multi-omics data, facilitating drug repositioning strategies or the identification of key molecules and processes of stroke pathophysiology. Finally, there is also the need to further comprehend tissue-specific crosstalk integrating brain stroke models with other tissues and with gut microbiota, to simulate stroke in a global biological context.

In conclusion, we have integrated and validated changes due to ischemia at protein and gene level in human brain samples. Some of the proposed candidates show potential as stroke biomarkers, while some flagship molecules might be promising therapeutic candidates to be further explored in the future.

5. References

1. Benjamin EJ, Virani SS, Callaway CW, et al (2018) Heart Disease and Stroke Statistics—2018 Update: A Report From the American Heart Association. Circulation e67–e492. https://doi.org/10.1161/CIR.0000000000000558

2. Lo EH, Dalkara T, Moskowitz MA (2003) Neurological diseases: Mechanisms, challenges and opportunities in stroke. Nat Rev Neurosci 4:399–414. https://doi.org/10.1038/nrn1106

3. Karsy M, Brock A, Guan J, et al (2017) Neuroprotective strategies and the underlying molecular basis of cerebrovascular stroke. Neurosurg Focus 42:E3. https://doi.org/10.3171/2017.1.FOCUS16522

4. O'Collins VE, Macleod MR, Donnan GA, et al (2006) 1,026 Experimental treatments in acute stroke. Ann Neurol. https://doi.org/10.1002/ana.20741

5. Yew KS, Cheng EM (2015) Diagnosis of acute stroke. Am Fam Physician 91:528–536. https://doi.org/http://www.aafp.org/afp/2015/0415/p528.html. Accessed September 16,2016

6. Montaner J, Ramiro L, Simats A, et al (2020) Multilevel omics for the discovery of biomarkers and therapeutic targets for stroke. Nat. Rev. Neurol.

7. Cuadrado E, Rosell A, Colomé N, et al (2010) The proteome of human brain after ischemic stroke. J Neuropathol Exp Neurol 69:1105–1115. https://doi.org/10.1097/NEN.0b013e3181f8c539

8. Malik R, Chauhan G, Traylor M, et al (2018) Multiancestry genome-wide association study of 520,000 subjects identifies 32 loci associated with stroke and stroke subtypes. Nat Genet. https://doi.org/10.1038/s41588-018-0058-3

9. Tae Hwan Shin, Da Yeon Lee, Shaherin Basith, Balachandran Manavalan, Man Jeong Paik, Igor Rybinnik, M Maral Mouradian, Jung Hwan Ahn GL (2020)

Metabolome Changes in Cerebral Ischemia. Cells 9:

10. Ludhiadch A, Vasudeva K, Munshi A (2020) Establishing molecular signatures of stroke focusing on omic approaches: a narrative review. Int. J. Neurosci.

11. Cuadrado E, Rosell A, Penalba A, et al (2009) Vascular MMP-9/TIMP-2 and neuronal MMP-10 up-regulation in human brain after stroke: A combined laser microdissection and protein array study. J Proteome Res 8:3191–3197. https://doi.org/10.1021/pr801012x

12. Yuan D, Liu C, Hu B (2018) Dysfunction of Membrane Trafficking Leads to Ischemia-Reperfusion Injury After Transient Cerebral Ischemia. Transl. Stroke Res.

13. Wang L, Wei C, Deng L, et al (2018) The Accuracy of Serum Matrix Metalloproteinase-9 for Predicting Hemorrhagic Transformation After Acute Ischemic Stroke: A Systematic Review and Meta-Analysis. J Stroke Cerebrovasc Dis. https://doi.org/10.1016/j.jstrokecerebrovasdis.2018.01.023

14. Lawrence F. Eng, Roopa S. Ghirnikar, Yuen L. Lee (2000) Glial Fibrillary Acidic Protein: GFAP-Thirty-One Years (1969–2000). Neurochem Res

15. Perry LA, Lucarelli T, Penny-Dimri JC, et al (2018) Glial fibrillary acidic protein for the early diagnosis of intracerebral hemorrhage: Systematic review and meta-analysis of diagnostic test accuracy. Int J Stroke. https://doi.org/10.1177/1747493018806167

16. Moore DF, Li H, Jeffries N, et al (2005) Using peripheral blood mononuclear cells to determine a gene expression profile of acute ischemic stroke: A pilot investigation. Circulation. https://doi.org/10.1161/01.CIR.0000152105.79665.C6

17. Stamova B, Xu H, Jickling G, et al (2010) Gene expression profiling of blood for the prediction of ischemic stroke. Stroke. https://doi.org/10.1161/STROKEAHA.110.588335

18. Stamova B, Ander BP, Jickling G, et al (2018) The intracerebral hemorrhage blood transcriptome in humans differs from the ischemic stroke and vascular risk factor control blood transcriptomes. J. Cereb. Blood Flow Metab.

19. Haas R, Zelezniak A, Iacovacci J, et al (2017) Designing and interpreting “multi-

omic" experiments that may change our understanding of biology. Curr. Opin. Syst. Biol.

20. Subramanian I, Verma S, Kumar S, et al (2020) Multi-omics Data Integration, Interpretation, and Its Application. Bioinform. Biol. Insights

21. de Tayrac M, Lê S, Aubry M, et al (2009) Simultaneous analysis of distinct Omics data sets with integration of biological knowledge: Multiple Factor Analysis approach. BMC Genomics 10:32. https://doi.org/10.1186/1471-2164-10-32

22. Simats A, Ramiro L, García-Berrocoso T, et al (2020) A Mouse Brain-based Multi-omics Integrative Approach Reveals Potential Blood Biomarkers for Ischemic Stroke. Mol Cell Proteomics. https://doi.org/10.1074/mcp.RA120.002283

23. Brott T, Adams HP, Olinger CP, et al (1989) Measurements of acute cerebral infarction: A clinical examination scale. Stroke. https://doi.org/10.1161/01.STR.20.7.864

24. van Swieten JC, Koudstaal PJ, Visser MC, et al (1988) Interobserver agreement for the assessment of handicap in stroke patients. Stroke 19:604–7

25. Adams H., Adams H., Bendixen B., et al (1993) Classification of Subtype of Acute Ischemic Stroke. Stroke 23:35–41. https://doi.org/10.1161/01.STR.24.1.35

26. Riba-Llena I, Jarca CI, Mundet X, et al (2013) Investigating silent strokes in hypertensives: a magnetic resonance imaging study (ISSYS): rationale and protocol design. BMC Neurol 13:130–137. https://doi.org/10.1186/1471-2377-13-130

27. García-Berrocoso T, Llombart V, Colàs-Campàs L, et al (2018) Single Cell Immuno-Laser Microdissection Coupled to Label-Free Proteomics to Reveal the Proteotypes of Human Brain Cells After Ischemia. Mol Cell Proteomics 17:175–189. https://doi.org/10.1074/mcp.ra117.000419

28. Kauffmann A, Gentleman R, Huber W (2009) arrayQualityMetrics—a bioconductor package for quality assessment of microarray data. Bioinformatics 25:415–416. https://doi.org/10.1093/bioinformatics/btn647

29. Irizarry RA, Hobbs B, Collin F, et al (2003) Exploration, normalization, and summaries of high density oligonucleotide array probe level data. Biostatistics 4:249–264. https://doi.org/10.1093/biostatistics/4.2.249

30. Meng C, Kuster B, Culhane AC, Gholami AM (2014) A multivariate approach to the integration of multi-omics datasets. BMC Bioinformatics 15:. https://doi.org/10.1186/1471-2105-15-162

31. Gonz I, Martin PGP, Baccini A (2008) CCA: An R Package. J Stat Softw 23:

32. Meng C, Kuster B, Peters B, et al (2016) moGSA: integrative single sample gene-set analysis of multiple omics data. bioRxiv 046904. https://doi.org/10.1101/046904

33. Liberzon A, Birger C, Thorvaldsdóttir H, et al (2015) The Molecular Signatures Database Hallmark Gene Set Collection. Cell Syst. https://doi.org/10.1016/j.cels.2015.12.004

34. Romero-Calvo I, Ocón B, Martínez-Moya P, et al (2010) Reversible Ponceau staining as a loading control alternative to actin in Western blots. Anal Biochem 401:318–320. https://doi.org/10.1016/j.ab.2010.02.036

35. Fanara P, Wong PYA, Husted KH, et al (2012) Cerebrospinal fluid-based kinetic biomarkers of axonal transport in monitoring neurodegeneration. J Clin Invest. https://doi.org/10.1172/JCI64575

36. Lam MPY, Wang D, Lau E, et al (2014) Protein kinetic signatures of the remodeling heart following isoproterenol stimulation. J Clin Invest. https://doi.org/10.1172/JCI73787

37. Fagan A, Culhane AC, Higgins DG (2007) A multivariate analysis approach to the integration of proteomic and gene expression data. Proteomics. https://doi.org/10.1002/pmic.200600898

38. Simats A, García-Berrocoso T, Montaner J (2016) Neuroinflammatory biomarkers: From stroke diagnosis and prognosis to therapy. Biochim Biophys Acta - Mol Basis Dis 1862:411–424. https://doi.org/10.1016/j.bbadis.2015.10.025

39. González I, Cao KAL, Davis MJ, Déjean S (2012) Visualising associations

between paired “omics” data sets. BioData Min. https://doi.org/10.1186/1756-0381-5-19

40. Amantea D, Bagetta G (2017) Excitatory and inhibitory amino acid neurotransmitters in stroke: from neurotoxicity to ischemic tolerance. Curr. Opin. Pharmacol.

41. Puig B, Brenna S, Magnus T (2018) Molecular communication of a dying neuron in stroke. Int. J. Mol. Sci.

42. Baeten KM, Akassoglou K (2011) Extracellular matrix and matrix receptors in blood-brain barrier formation and stroke. Dev Neurobiol. https://doi.org/10.1002/dneu.20954

43. Gould DB, Phalan FC, Van Mil SE, et al (2006) Role of COL4A1 in small-vessel disease and hemorrhagic stroke. N Engl J Med. https://doi.org/10.1056/NEJMoa053727

44. Ellison JA, Barone FC, Feuerstein GZ (1999) Matrix remodeling after stroke. De novo expression of matrix proteins and integrin receptors. In: Annals of the New York Academy of Sciences

45. Frank S, Gaume B, Bergmann-Leitner ES, et al (2001) The Role of Dynamin-Related Protein 1, a Mediator of Mitochondrial Fission, in Apoptosis. Dev Cell 1:515–525. https://doi.org/10.1016/S1534-5807(01)00055-7

46. Jang A, Koh P (2016) Ischemic brain injury decreases dynamin-like protein 1 expression in a middle cerebral artery occlusion animal model and. Lab Anim Res 6055:194–199

47. Grohm J, Kim SW, Mamrak U, et al (2012) Inhibition of Drp1 provides neuroprotection in vitro and in vivo. Cell Death Differ. https://doi.org/10.1038/cdd.2012.18

48. Zhang N, Wang S, Li Y, et al (2013) A selective inhibitor of Drp1, mdivi-1, acts against cerebral ischemia/reperfusion injury via an anti-apoptotic pathway in rats. Neurosci Lett. https://doi.org/10.1016/j.neulet.2012.12.049

49. Chiang Y, Schneiderman MH, Vishwanatha JK (1993) Annexin II Expression Is Regulated during Mammalian Cell Cycle. Cancer Res

50. Sharma MC (2019) Annexin A2 (ANX A2): An emerging biomarker and potential therapeutic target for aggressive cancers. Int. J. Cancer

51. Kim J, Hajjar KA (2002) Annexin II: a plasminogen-plasminogen activator co-receptor. Front. Biosci.

52. Jiang Y, Fan X, Yu Z, et al (2015) Low dose tPA plus annexin A2 combination attenuates tPA delayed treatment- associated hemorrhage and improves recovery in rat embolic focal stroke. Neurosci Lett. https://doi.org/10.1016/j.neulet.2015.06.050

53. Wang X, Wang X, Fan X, et al (2014) Effects of tissue plasminogen activator and annexin A2 combination therapy on long-term neurological outcomes of rat focal embolic stroke. Stroke. https://doi.org/10.1161/STROKEAHA.113.003823

54. Jiang Y, Fan X, Yu Z, et al (2015) Combination low-dose tissue-type plasminogen activator plus annexin A2 for improving thrombolytic stroke therapy. Front. Cell. Neurosci.

55. Fan X, Jiang Y, Yu Z, et al (2017) Annexin A2 Plus Low-Dose Tissue Plasminogen Activator Combination Attenuates Cerebrovascular Dysfunction After Focal Embolic Stroke of Rats. Transl Stroke Res. https://doi.org/10.1007/s12975-017-0542-6

56. Onwuekwe I, Ezeala-Adikaibe B (2012) Ischemic stroke and neuroprotection. Ann Med Health Sci Res. https://doi.org/10.4103/2141-9248.105669

57. Bartosik-Psujek H, Belniak E, Stelmasiak Z (2003) Markers of inflammation in cerebral ischemia. Neurol Sci 24:279–280. https://doi.org/10.1007/s10072-003-0156-5

58. Irmak S, Oliveira-Ferrer L, Singer BB, et al (2009) Pro-angiogenic properties of orosomucoid (ORM). Exp Cell Res 315:3201–3209. https://doi.org/10.1016/j.yexcr.2009.07.024

59. Kandregula CAB, Smilin Bell Aseervatham G, Bentley GT, Kandasamy R (2016) Alpha-1 antitrypsin: Associated diseases and therapeutic uses. Clin Chim Acta 459:109–116. https://doi.org/10.1016/j.cca.2016.05.028

6. Tables

Table 1. Demographic and clinical information of ischemic stroke patients. Samples from the infarcted and corresponding contralateral brain areas were included in the discovery (proteomics & transcriptomics) or the replication (Western Blot & qRT-PCR) phase, as indicated.

Patient	Sex	Age	T-O-D (h)	PMI (h)	rt-PA	Discovery	Replication
N3	M	59	70	6	No		X
N12	F	77	72	4	Yes		X
N16	F	79	88	5	No	X	
N22	M	67	62	7	Yes	X	
N24	F	83	100	14.5	Yes		X
N29	F	92	67	25	No		X
N32	M	73	360	6	No	X	
N33	M	80	100	4.5	No	X	
N35	M	84	40	7.75	No	X	
N36	F	73	44	4	Yes	X	
N38	M	75	19	5	Yes		X

M: male; F: female; T-O-D: time from onset of symptoms to death; PMI: post-mortem interval; rt-PA: recombinant tissue plasminogen activator treatment.

Table 2. Selection of dysregulated proteins and genes after ischemic stroke for replication. Summary of results in both discovery (microarray & LC-MS) and replication phases (qRT-PCR & Western Blot). Verified candidates are shaded. Data in light grey indicates that the candidate did not achieve our significance criteria (q-value <0.1) or was not identified (n/a) in the discovery phase but so did its molecular counterpart.

		GENES	Microarray		qRT-PCR		PROTEINS	LC-MS		WB	
Symbol	Main function	Entrez ID	logFC	q-value	logFC	p-value	UNIPROT ID	logFC	q-value	logFC	p-value
A1AG1	Transport protein in the blood stream.	5004	n/a	n/a	n/a	n/a	P02763	2.584	0.049	1.850	0.043
A1AT	Inhibitor of serine proteases.	5265	0.541	0.103	n/a	n/a	P01009	2.334	0.048	4.015	0.043
ANXA1	Regulates inflammatory process and actin cytoskeleton.	301	2.259	0.043	1.825	0.043	P04083	2.525	0.116	n/a	n/a
ANXA2	Binds Ca^{2+}.	302	2.304	0.043	2.414	0.043	P07355	2.547	0.158	n/a	n/a
ATRNL1	Melanocortin signaling pathway (energy homeostasis).	26033	-2.713	0.043	-0.678	0.686	Q5VV63	n/a	n/a	n/a	n/a
C5AR1	Receptor of C5a chemotactic peptide.	728	2.336	0.045	2.676	0.225	P21730	n/a	n/a	n/a	n/a
CCL2	Chemokine that attracts monocytes and basophils.	6347	2.329	0.043	0.290	0.686	P13500	n/a	n/a	n/a	n/a
CXCR4	Receptor of SDF-1 chemokine. Mediates neuron survival. Remodeling endothelium.	7852	1.674	0.043	4.339	0.138	P61073	n/a	n/a	n/a	n/a
DNM1L	Mitochondrial and peroxisomal division (membrane fission).	10059	-0.772	0.059	-0.307	0.345	O00429	-2.846	0.028	-1.203	0.043
GABBR2	Receptor of GABA (inhibitor neurotransmitter), mediates G-prot coupling.	9568	-2.299	0.043	-0.502	0.345	O75899	n/a	n/a	n/a	n/a
GABRA1	Receptor of GABA (inhibitor neurotransmitter).	2554	-2.662	0.043	-0.558	0.893	P14867	n/a	n/a	n/a	n/a
GABRG2	Receptor of GABA (inhibitor neurotransmitter).	2566	-2.798	0.043	0.226	0.500	P18507	n/a	n/a	n/a	n/a
GPR183	Receptor from lymphocytes to attract other leucocytes. Receptor of oxysterols.	1880	2.035	0.047	0.580	0.345	P32249	n/a	n/a	n/a	n/a
HCN1	Ion channel.	348980	-2.702	0.043	-0.364	0.500	O60741	n/a	n/a	n/a	n/a
HTR2A	Receptor for serotonin.	3356	-2.266	0.043	-0.594	0.225	P28223	n/a	n/a	n/a	n/a
IL8	Chemokine that attracts neutrophils, basophils and T-cells.	3576	2.048	0.043	4.685	0.043	P10145	n/a	n/a	n/a	n/a
INA	Neuronal intermediate filament (morphogenesis).	9118	-3.169	0.043	-0.395	0.500	Q16352	0.908	0.239	n/a	n/a
JAM3	Cell-cell adhesion. Regulation of PMNs transepithelial migration.	83700	0.182	0.511	n/a	n/a	Q9BX67	3.006	0.067	1.701	0.043
MYPR	Myelin protein.	5354	-0.653	0.073	1.484	0.686	P60201	2.598	0.028	-0.948	0.043
NAMPT	Involved in NAD biosynthesis. Soluble form working as cytokine. Modulation of circadian clock.	10135	1.635	0.043	2.265	0.138	P43490	0.322	0.482	n/a	n/a
NAPB	Vesicular transport from endoplasmic reticulum to Golgi apparatus.	63908	-2.661	0.043	-0.548	0.686	Q9H115	0.568	0.568	n/a	n/a
NCDN	CNS signal transduction. Negative regulator of CaMK2 phosphorylation. Neurite outgrowth.	23154	-1.397	0.043	-1.940	0.043	Q9UBB6	-2.559	0.054	-1.664	0.043
NEFL	Neurofilament light (neuronal caliber).	4747	-3.233	0.043	-0.546	0.345	P07196	0.532	0.530	n/a	n/a
NEFM	Neurofilament medium (neuronal caliber).	4741	-2.786	0.043	-0.975	0.225	P07197	2.053	0.102	n/a	n/a
PLIN2	Maybe, development and maintenance of adipose tissue.	123	1.978	0.043	3.791	0.225	Q99541	n/a	n/a	n/a	n/a
RAB3C	Protein transport. Vesicular traffic.	115827	-1.284	0.044	-1.312	0.043	Q96E17	-3.106	0.028	-2.229	0.043
SCN1A	Na^{+} channel. Regulates the release of neurotransmitters.	6323	-2.715	0.043	-0.993	0.225	P35498	n/a	n/a	n/a	n/a
SCN2A	Na+ channel.	6326	-2.850	0.043	-0.185	0.500	Q99250	n/a	n/a	n/a	n/a
SRGN	Mediates storage in secretory vesicles in T-cells and neutrophils. Mediates processing of MMP2.	5552	1.880	0.043	3.445	0.080	P10124	n/a	n/a	n/a	n/a
ST4A1	Sulfotransferase. Metabolism of neurotransmitters.	25830	-1.529	0.043	-1.398	0.043	Q9BR01	-3.240	0.028	-1.503	0.043
THBS1	Glycoprotein for cell-cell and cell-matrix interaction. Anti-angiogenic.	7057	1.506	0.047	1.762	0.225	P07996	n/a	n/a	n/a	n/a
VTDB	Vitamin D transport and storage. Scavenging extracellular G-actin.	2638	n/a	n/a	n/a	n/a	P02774	3.389	0.048	2.403	0.043

7. Figure legends

Figure 1. Workflow chart. Schematic description summarizing the different phases of our study. LC-MS/MS: liquid chromatography coupled to tandem mass spectrometry; LogFC: logarithmic fold-change; WB: Western Blot; qRT-PCR: quantitative reverse transcription polymerase chain reaction; ELISA: enzyme-linked immunosorbent assay.

Figure 2. Integrative multiomics outline. A. Co-inertia analysis visualization plot of proteomics and transcriptomics patterns across IC (red) and CL (blue) samples; the length of the line joining the tip of the arrow (protein data) and the circle (gene data) is proportional to the divergence between both molecular datasets. **B.** Relevance network plot of differentially expressed proteins (rectangles) and genes (circles) that correlate with an R coefficient ≥ 0.9. Red and green edges indicate positive and negative correlations, respectively. **C.** Contribution of proteins and genes to the variance of each principal component (PC) obtained to discriminate between IC and CL areas. **D.** Canonical pathways overrepresented in the IC of ischemic stroke brains. Decomposed gene set scores (GSS) are weighted for proteins (light grey) and genes (dark grey). Means and 95% of confidence intervals are depicted in bar graphs. **E.** Most influential molecules in each overrepresented canonical pathway from D. Gene influential scores (GIS) are represented for dysregulated proteins (light grey) and genes (dark grey). IC: infarct core brain area; CL: contralateral brain area.

Figure 3. Replicated proteins and genes dysregulated after ischemic stroke. A. Candidates replicated at both molecular levels. **B.** Candidates replicated at the protein level. **C.** Candidates replicated at the gene level. Protein and gene levels represent corrected values. Circles represent outliers while stars stand for extreme values. Statistical analysis results are reported in Table 2. In the case of proteins, representative Western Blots are also depicted. IC: infarct core brain area; CL: contralateral brain area.

Figure 4. Qualification study of protein candidates in blood samples. Longitudinal analysis of molecular plasma levels across 3 different time-points: admission (Adm.), 24h and one week (5-7 d) after stroke (n=11). Circles represent outliers; continuous dashed lines represent quartile values in the control population (n=5) only as a reference. *: $p < 0.05$; #: $0.1 < p > 0.05$ (Wilcoxon test corrected by Bonferroni).