

# **Distinct proteomic profiles in monozygotic twins discordant for ischaemic stroke**

**Nirmal Vadgama, PhD<sup>1,2</sup>; Douglas Lamont, BSc<sup>3</sup>; John Hardy, PhD<sup>1</sup>; Jamal Nasir, PhD<sup>2,5</sup>; Ruth C. Lovering, PhD<sup>4</sup>**

*1) Institute of Neurology, University College London, London, UK*

*2) Cell Biology and Genetics Research Centre, St. George's University of London, London, UK*

*3) College of Life Sciences, University of Dundee, Dundee, UK*

*4) Centre for Cardiovascular Genetics, Institute of Cardiovascular Science, University College London, London, UK.*

*5) Present Address: Molecular Biosciences Research Group, Faculty of Health & Society, University of Northampton, UK.*

**Cover title:** Serum proteome comparison of MZ twins discordant for stroke

**Tables and figures:** Tables 4; Figures 1

**Indexing terms:** Proteomics, Gene Ontology, Stroke, Biomarker, Monozygotic twins

**Subject terms:** Ischaemic stroke, proteomics, biomarkers, fibulin 1, MZ discordant twins

**Word count:** 5679

**Corresponding author:**

Jamal Nasir<sup>2</sup>

Email: [jamal.nasir@northampton.ac.uk](mailto:jamal.nasir@northampton.ac.uk).

## **Abstract**

Background and Purpose: Stroke is a common disorder with significant morbidity and mortality, and complex aetiology involving both environmental and genetic risk factors. Although some of the major risk factors for stroke, such as smoking and hypertension, are well-documented, the underlying genetic and detailed molecular mechanisms remain elusive. Exploring the relevant biochemical pathways may contribute to the clinical diagnosis of stroke and shed light on its aetiology.

Methods: A comparative proteomic analysis of blood serum of a pair of monozygotic (MZ) twins discordant for ischaemic stroke (IS) was performed using a label-free quantitative proteomics approach. To overcome the limit of reproducibility in the serum preparation, two separate runs were performed, each consisting of three technical replicates per sample. Biological processes associated with proteins differentially expressed between the twins were explored with Gene Ontology (GO) classification using the functional analysis tool g:Profiler.

Results: ANOVA test performed in Progenesis LC-MS identified 179 (run 1) and 209 (run 2) proteins as differentially expressed between the affected and unaffected twin ( $P < 0.05$ ). Furthermore, the level of serum fibulin 1, an extracellular matrix protein associated with arterial stiffness, was on average 13.37-fold higher in the affected twin. Each dataset was then analysed independently, and the proteins were classified according to GO terms. The categories overrepresented in the affected twin predominantly corresponded to stroke-relevant processes, including wound healing, blood coagulation and haemostasis, with a high proportion of the proteins overexpressed in the affected twin associated with these terms. By contrast, in the unaffected twin diagnosed with atopic dermatitis, there were increased levels of keratin proteins and GO terms associated with skin development.

Conclusion: The identification of cellular pathways enriched in IS as well as the up-regulation of fibulin 1 sheds new light on the underlying disease-causing mechanisms at the molecular level. Our findings of distinct proteomic signatures associated with IS and atopic dermatitis suggest proteomic profiling could be used as a general approach for improved diagnostic, prognostic and therapeutic strategies.

## **Introduction**

Stroke affects up to 15 million people per year [1] and is a leading cause of death. The aetiology of stroke is complex, involving both environmental and genetic risk factors. Despite numerous large-scale genetic studies, the genetic causes of stroke remain poorly understood reflecting the complex nature of the disease [1]. Studies aimed at identifying the underlying genes are limited due to enormous genetic variability between individuals.

Advances in next-generation sequencing technologies have revolutionised studies of genetic diseases [2], but for complex diseases, whole exome and genome sequencing have proved less beneficial, and attention is increasingly switching to RNA sequencing and epigenetics to explore disease-causing mechanisms. However, these and other methods, such as those which involve analysis of structural variants, have also proved to be limited, highlighting the complexity of the underlying genetic mechanisms. Here, we describe quantitative label-free proteomic analysis as an alternative approach, which potentially could capture all the variation arising from genomic, transcriptomic and epigenomic changes.

Although advanced prevention, accurate diagnosis and appropriate treatment for stroke are crucial, ischaemic stroke (IS) is misdiagnosed in more than 1 in 7 patients during initial presentation in the emergency department, and increases to two thirds in cases where only non-traditional symptoms are present [3]. Clinical assessment and neuroimaging remain the

mainstay of stroke diagnosis, however this is reliant on probability rather than certainty and imaging facilities not always readily available. The development of molecular biomarkers, in the form of proteins, lipids and metabolites, to support the clinical diagnosis of stroke, as well as to provide prognostic indicators, would be of great value [4].

Several promising biomarkers have been identified for stroke, including vascular cell adhesion molecule 1, intercellular cell adhesion molecule 1, von Willebrand factor and matrix metalloproteinases, however none are currently used in clinical practice due to lack of sufficient evidence [5]. Thus, further study and validation are warranted.

Monozygotic (MZ) twins that are discordant with respect to a genetic disease offer a unique opportunity to explore the genetics of complex diseases [6]. We previously described a unique pair of then 57-year-old MZ twins discordant for IS with raised  $\gamma$ -glutamyl transferase (GGT) and erythrocyte sedimentation rate (ESR) in the affected twin [7]. Here, we describe proteomic analysis of serum samples from the twins.

## **Materials and methods:**

The authors declare that all supporting data are available within the article and its online supplementary file.

### Medical history

The affected twin (KG) presented with a stroke in 2007 (aged 55 years). This has left her with persistent right-sided weakness affecting her balance, speech, and walking. In 2007 an MRI scan confirmed a left ganglion infarct, ischaemic changes in both cerebral hemispheres, and a small right temporal infarct. A coincidental finding of a right-sided intracavernous internal carotid artery aneurysm measuring 9 mm was made in 2007, prior to the stroke. This has been

managed conservatively ever since, consisting of routine periodic follow-up imaging with MR or CT angiography. Patients with significant aneurysm growth are strongly considered for interventional treatment, however the affected twin refused this intervention despite hers measuring 10 mm in 2009. The affected twin has not smoked for seven years since the stroke, whereas the unaffected twin (HG), despite sharing the same risk factors and still smoking, has not had a stroke.

In addition, the affected twin was diagnosed with a single patch of alopecia areata in 1995 (aged 44), and more recently with seborrhoeic dermatitis capitis in 2012 (aged 60). The unaffected twin was diagnosed with atopic dermatitis in 2004 (aged 53). These problems might be linked to a family history of psoriasis (in father) and suggest an autoimmune disorder in the family as shown in Supplementary Figure I. We describe MZ twins discordant for IS with a family history of hypertension, depression and psoriasis. To our knowledge, this is the first proteomics study of MZ twins discordant for stroke and intracranial aneurysm.

#### Blood collection:

Written informed consent was obtained from both patients. Blood samples of both twins were collected seven years after stroke onset. Samples were drawn from the twins by standard venesection into plain tubes and fluoride and EDTA anticoagulated tubes, just prior to lunch, and processed in parallel. From each plasma sample, 2ml blood was allowed to clot at 4°C for at least 2hrs and then centrifuged at 1500g for 10mins to sediment the clotted cells. Plasma was then collected, divided into aliquots, and stored frozen at -80°C until the analysis was carried out. The control sample was created by taking 5µl of sample from each twin's serum after processing to generate peptides and pooling these together. As the control serum contains all the features of both twin samples, it was used to align the data effectively – that is, to correct for any drift in the retention time of the peptides analysed.

### Gel electrophoresis and in-gel digestion:

1µl of a 1:4 dilution of each serum sample was run on SDS-PAGE using a 4-12% bis-tris gel and MOPS buffer system. Briefly 10µl of sample buffer and 4µl of reducing agent were added to each sample prior to addition of 25µl of milliQ water. The gel lanes from each serum sample were processed as follows. Samples were processed by in-gel reduction with 10mM dithiothreitol and alkylation with 50mM iodoacetamide prior to overnight (16hrs) trypsin digestion (Modified Sequencing Grade, Roche) at 30°C. Peptides were extracted from the gel by 1% formic acid, then 100% acetonitrile, and dried in a SpeedVac concentrator (Thermo Scientific). The resulting peptides were then resuspended in 50µl 1% formic acid, centrifuged and transferred to the high-performance liquid chromatography (HPLC) vial. 7.5µl of each twin and pooled sample was used for analysis.

### Sample separation

Technical replicates (3 x 7.5µl) of trypsin-digested peptides were separated using an UltiMate 3000 RSLCnano UHPLC System (*Thermo Scientific*) coupled to a LTQ Orbitrap Velos Pro (*Thermo Scientific*). The nLC buffers used were buffer A (0.1% formic acid), and buffer B (80% acetonitrile in 0.08% formic acid). Technical replicates (3 x 2.5µg) of each sample were initially trapped on an Acclaim PepMap100 C18, 5µM (100µM x 2cm nanoViper). After trap enrichment, peptides were separated on an Easy-Spray PepMap RSLC C18, 2µM column (75µM x 50cm nanoViper) (*Thermo Scientific*) with a linear gradient of 2-40% buffer B over 120mins with a constant flow of 0.3µL/min.

The UHPLC system was coupled to a LTQ Orbitrap Velos Pro (*Thermo Scientific*), via a nano electrospray ion source (*Proxeon Biosystems*). A Top 15 method was used to acquire Data Dependant Acquisition data. Briefly, a 60,000-resolution full-scan mass spectrometry (MS)

survey spectra (m/z 335-1800) were attained in the Orbitrap with an automatic gain control target of 1,000,000 ions. Lock mass was set at 445.120024 and the spray voltage set to 1.8kV. This was followed by ion-trap MS/MS scans for the 15 most intense peptide ions.

Maximal filling times were 500ms for the full scans and 100ms for the MS/MS scans. Precursor ion charge state screening was enabled and all unassigned charge states, as well as singly charged species, were rejected. The dynamic exclusion was set to an exclusion list size of 500 with a maximum retention period of 45secs and an exclusion mass width of  $\pm 10$  ppm. The lock mass option was enabled for survey scans to improve mass accuracy (Olsen, 2005). Data were acquired using the XCalibur software.

#### Abundance quantification

Orbitrap Velos Pro .RAW data files were imported into Progenesis LC-MS (version 4.0) for label-free differential analysis and subsequent identification and quantification of relative ion abundance ratios. Following alignment of MS data, principal component analysis and preliminary filtering (power <80% and  $P > 0.05$ , corresponding to an ion score threshold of 35), datasets were exported from Progenesis as an .mgf file. These files were subsequently used to map the individual peptide sequences to corresponding UniProt identifiers using the UniProt database via Mascot Daemon (version 2.4.1). Enzyme specificity was set to trypsin, with fixed (quantitative) modifications set to carbamidomethyl cysteine (as it is assumed that all cysteines are modified so this does not alter the number of potential peptides). To increase the number of potential peptides, parameters for the variable (non-quantitative) modifications were set to methionine oxidation, methionine dioxidation, protein N-acetylation, and Gln  $\rightarrow$  pyro-Glu. Other parameters used were as follows: Peptide mass tolerance,  $\pm 10$ ppm; fragment mass tolerance,  $\pm 0.6$  Da; minimum peptide length, 6; and maximum missed cleavages, 2. Statistical P-values as shown in Tables 1, 2 and Supplementary Table 1 were automatically generated

using Progenesis software through a one way ANOVA on the ArcSinh transform of the normalised data.

### Peptide/protein identification

The filtered Mascot search results were reimported into the Progenesis LC-MS and conflicts for peptide assignments at protein level were examined and resolved appropriately.

### Pathway and Network analysis

To explore the biological processes associated with the differentially expressed proteins, we performed Gene Ontology (GO) pathway analysis using the functional analysis tool g:Profiler, gOST [8] (February 2017). Only proteins with an average of >1.5-fold higher level in both runs for each twin were used to seed each network, using Cytoscape [9] version 3.3.0 (April 2017).

## **Results:**

A comparative proteomic analysis of blood serum of MZ twins discordant for IS was performed using a label-free proteomics method to identify qualitative and quantitative differences at the protein level.

Proteins were extracted from serum samples collected from each twin. Following staining of the gel, a prominent band at approximately 60 kDa, corresponding mainly to albumin, was cut out from the gel to enhance the detection of less abundant proteins as albumin represents approximately 55% of total cellular protein [10] (see Supplementary Figure II). The proteins were eluted from the gel for proteomic studies, as described (see Materials and Methods).



With the aim to overcome the limit of reproducibility in the serum preparation, two separate runs were performed, each consisting of three technical replicates per sample. The samples were analysed using nano-scale liquid chromatographic tandem mass spectrometry (nLC-MS/MS).

The raw data were matched against the UniProt database for peptide identification using Mascot (version 2.4.1). To identify the proteins that were differentially expressed between the affected and unaffected twin blood serum, an ANOVA test was performed in Progenesis LC-MS that identified 179 and 209 proteins ( $P < 0.05$ ) in runs 1 and 2 respectively.

The twins had distinctively different serum profiles. Some of the proteins were present at higher levels in the affected twin's serum (Table 1). Many of these, including carbonic anhydrase I [11], hemoglobin subunit alpha 1 [12], alpha-2-macroglobulin [13], hyaluronan binding protein 2 [14], S100 calcium binding protein 8 [15], ficolin 3 [16], and coagulation factor II, thrombin [17] (Table 1), have previously been associated with biological pathways relevant to IS.

The highest average fold increase (13.37) was observed in the expression of fibulin 1, an extracellular matrix protein associated with arterial stiffness [18]. Intriguingly, for some of the proteins showing low levels of expression in the affected twin (Table 2), including desmoplakin [19] and serpin family A member 4 (kallistatin) [20], reduced expression appears to be associated with stroke. In addition, desmoglein 1, which is also present at lower levels in the affected twin's serum, associates with desmoplakin as a heterodimer to form the desmosome, a key adhesion molecule, and is also associated with keratin filaments [21].

Functional analysis was performed independently on all 4 datasets of differentially expressed proteins, to classify the proteins according to 'biological process', 'molecular function' and 'cellular component'. The g:Profiler web server [8] (<http://biit.cs.ut.ee/gprofiler/>) comprises

several tools to perform functional enrichment analysis and mine additional information. The statistical method used by this tool, g:SCS (Set Counts and Sizes, [8]) takes into account the ontology structure supporting the annotations and enables significantly enriched GO terms to be identified within short lists of proteins (down to lists of only 11 proteins).

Proteins were only included in the analysis if they were identified at higher levels in either the affected or unaffected twin in both runs, or only one run. Proteins with high levels in one run, but low levels in the equivalent other run were excluded from the analysis. After applying this exclusion criteria, the number of proteins remaining were 125 and 156 in runs 1 and 2, respectively. Some of the protein identifiers associated with the differentially expressed peptides did not have any associated GO terms, and thus were not included in the g:Profiler analysis. In the majority of cases these peptides were predicted to be derived from an immunoglobulin molecule. Overall, 95 and 139 proteins, from runs 1 and 2 respectively, identified as differentially expressed and associated with GO terms, and were thus included in the analysis (Supplemental Table 1).

With a fold-difference average threshold greater than 1.5, 19 proteins were identified with higher levels in the affected twin's serum in both runs, and 13 in the unaffected twin's serum in both runs. Despite the differences in the proteins identified in each run, several GO terms were significantly enriched consistently within either the high or low protein lists, suggesting that they share common biological pathways. Although there are also GO terms which do not show consistency (Supplemental Table 2).

Surprisingly, the GO terms enriched in the serum of each twin were relevant to very different biological pathways. The 'biological process' GO categories overrepresented in the affected twin related to cytolysis, response to stress, clot formation and blood coagulation (Table 3 and Supplemental Table 2), all of which have previously been implicated in IS. Between 15-30%

of the proteins differentially overexpressed in the affected twin associated with wound healing and/or blood coagulation. By contrast, in the unaffected twin, the enriched GO terms were almost exclusively related to skin development (Table 4 and Supplemental Table 2).

A network analysis of the overexpressed proteins in each twin was undertaken. Only proteins which were overexpressed in both runs and with an average fold difference greater than 1.5 were included in the analysis (Tables 1 and 2). This analysis demonstrated that 8 of the proteins overexpressed in the affected twin's serum can be located within a single network, with 11 proteins not connected to any network. Further investigation is needed to clarify whether any of these unconnected proteins are associated. All but one of the 13 proteins overexpressed in the unaffected twin's serum were associated with a close network (Figure 1).

## **Discussion**

Plasma proteomes between MZ twins discordant for IS were compared to identify putative disease-associated markers in the affected twin. This approach provided a snapshot of proteins expressed in blood and identified differences in protein expression between the affected and unaffected twin. To the best of our knowledge this is the first proteomic study of MZ twins discordant for IS.

Several of the proteins present at high levels in the affected twin (Table 1) have previously been associated with IS.

Huang et al. [12] found that hemoglobin  $\alpha$  and  $\beta$ -chains were differentially expressed between stroke patients and controls ( $P < 0.0001$ ), and conclude that serum free hemoglobin may serve as a potential biomarker for the diagnosis of acute IS. In our study, only the subunit alpha 1

was identified at high levels in the patient; however, this does support the suggestion that this protein could be a useful biomarker for IS.

Nezu et al. [13] identified alpha-2-macroglobulin as an acute IS biomarker, and Gori et al. [22] found that this protein was significantly associated with poor outcome (death within three months), suggesting that it contributes to cerebral damage after IS and thrombolysis. The authors conclude that high levels of alpha-2-macroglobulin may be a useful prognostic biomarker to identify patients with an increased probability of death after the administration of tissue plasminogen activator [22].

Hyaluronan binding protein 2 is found to accumulate in lipid-rich areas within the necrotic core of atherosclerotic plaques, but not in normal vessels [23]; and a single nucleotide polymorphism in the gene, which attenuates its capacity to activate pro-urokinase, is associated with carotid stenosis [24]. In a more recent study, increased hyaluronan binding protein 2 antigen levels and activity were independently associated with all main aetiologic subtypes of IS, suggesting a potential role in the pathophysiology of the condition through prothrombotic mechanisms [14].

It has been postulated that upregulation and signalling of S100 calcium binding protein A8 contributes to neuroinflammation and the progression of ischaemic damage [15]. Further, Wang et al. [25] identified a role of this protein in the molecular pathway of thrombosis, suggesting that targeting it has potential for treating atherothrombotic disorders, including myocardial infarction and stroke. Interestingly, blockers for the S100A8/A9 heterodimer (calprotectin) have been developed and are approved for clinical testing [26].

Several of the proteins that were at low levels in the affected twin (Table 2) could also play an important role in the pathogenesis of IS. For instance, it has been shown that serum serpin

family A member 4 (kallistatin) levels are considerably reduced in hypertension, diabetes, cardiovascular and renal injury [27]. Thus, the downregulation of kallistatin in the affected twin may indicate its usefulness as biomarker for patients with cardiovascular disorders.

Desmoplakin is an important component of desmosome structures in cardiac muscle and epidermal cells [28], and abnormalities in its expression results in cardiomyocyte death, changes in lipid metabolism, and defects in cardiac development [29]. In a proteomics study set out to identify novel plasma biomarkers associated with vascular recurrence in post-stroke patients, it was found that desmoplakin may be protective against a new vascular event, lending its usefulness as a prognostic biomarker [19]. Decreased levels of plasma desmoplakin in the affected twin could be a risk factor for the reoccurrence of vascular events.

Moreover, the higher levels of fibulin 1 expression in the affected twin suggests it could be a biomarker for IS. Fibulin 1 is a secreted calcium-binding glycoprotein [30] that stabilises extracellular matrix integrity surrounding vascular smooth muscle. It is known to play important roles in wound repair [31] and mediates platelet adhesion by cross-linking with fibrinogen, forming part of a general mechanism by which platelets interact with exposed subendothelial matrices following vascular injury [32]. Increased plasma concentrations of fibulin 1 are predictive of cardiovascular mortality in type 2 diabetes [18], and a recent study found that common polymorphisms in the *FBLN1* gene predict aortic stiffness in young healthy subjects [33]. Aortic stiffness is an important independent risk factor for IS, suggesting the potential usefulness of fibulin 1 as a biomarker [34].

Each twin exhibited unique enriched GO terms. In the affected twin, biological processes were predominantly related to wound healing and blood coagulation, whereas in the unaffected twin processes relevant to epidermis development were enriched. Interestingly, the unaffected twin was previously diagnosed with atopic dermatitis (whereas the affected twin was diagnosed with

seborrheic dermatitis capitis). The enrichment of ‘skin development’ GO terms associated with the serum of the unaffected twin suggests she was experiencing an episode of atopic dermatitis at the time the blood sample was taken.

Proteomics has the potential to bridge the gap between genetics and biochemistry, identify functional biomarkers for complex conditions, and provide important insights into the mechanisms of disease. The identification of cellular pathways enriched in IS as well as the up-regulation of fibulin 1 sheds new light on the underlying disease-associated mechanisms at the molecular level, which could lead to the development of new diagnostic and therapeutic approaches in the future. However, the changes reported here don’t necessarily represent the cause of IS, as protein levels were measured several years after the stroke, which may reflect changes associated with addressing the lesion (e.g. wound healing) and not necessarily predict the susceptibility to stroke. In addition, there is great variability in protein expression between individuals, due to differences in lifestyle and dietary habits, among other factors. However, the 60-year-old twins in this study have lived together since birth, indicating that the effect of environmental factors is minimal.

## **Conclusion**

The pathogenesis of IS is very complex, involving multiple molecular mechanisms. This makes it unlikely for a single biomarker to sufficiently reflect the underlying complexity. Multiple biomarkers combined in a panel may be required to capture the dysregulation of several processes, which converge to lead to susceptibility to an ischaemic event, thereby improving the diagnostic sensitivity and specificity.

Although the potential roles of fibulin 1 and the other identified proteins in IS is not clear, further investigations into the expression levels of these proteins in stroke patients may add weight to their potential usefulness as biomarkers. In addition, large-scale proteomic studies may further elucidate the underlying common biochemical pathways and mechanisms.

## **Acknowledgments**

We would like to thank the FingerPrints Proteomics Facility, School of Life Sciences, University of Dundee for the proteomic and mass spectrometry analysis, and The Leverhulme Trade Charities Trust for a bursary to NV.

## **Sources of Funding**

RCL: Parkinson's UK Grant G-1307, British Heart Foundation (RG/13/5/30112), the National Institute for Health Research University College London Hospitals Biomedical Research Centre.

## **Conflict(s)-of-Interest/Disclosure(s)**

The authors report no conflicts.

## References

1. Raffeld MR, Debette S, Woo D (2016) International Stroke Genetics Consortium Update. *Stroke* 47:1144–5. <https://doi.org/10.1161/STROKEAHA.116.012682>
2. Goodwin S, McPherson JD, McCombie WR (2016) Coming of age: ten years of next-generation sequencing technologies. *Nat Rev Genet* 17:333–351. <https://doi.org/10.1038/nrg.2016.49>
3. Lever NM, Nyström K V, Schindler JL, et al (2013) Missed opportunities for recognition of ischemic stroke in the emergency department. *J Emerg Nurs* 39:434–9. <https://doi.org/10.1016/j.jen.2012.02.011>
4. Makris K, Haliassos A, Chondrogianni M, Tsivgoulis G (2018) Blood biomarkers in ischemic stroke: potential role and challenges in clinical practice and research. *Crit Rev Clin Lab Sci* 55:294–328. <https://doi.org/10.1080/10408363.2018.1461190>
5. Jickling GC, Sharp FR (2015) Biomarker panels in ischemic stroke. *Stroke* 46:915–20. <https://doi.org/10.1161/STROKEAHA.114.005604>
6. Zwijnenburg PJG, Meijers-Heijboer H, Boomsma DI (2010) Identical but not the same: The value of discordant monozygotic twins in genetic research. *Am J Med Genet Part B Neuropsychiatr Genet* 153B:n/a-n/a. <https://doi.org/10.1002/ajmg.b.31091>
7. Vadgama N, Gaze D, Ranson J, et al (2015) Elevated  $\gamma$ -glutamyltransferase and erythrocyte sedimentation rate in ischemic stroke in discordant monozygotic twin study. *Int J Stroke* 10:. <https://doi.org/10.1111/ijvs.12440>
8. Reimand J, Kull M, Peterson H, et al (2007) g:Profiler--a web-based toolset for functional profiling of gene lists from large-scale experiments. *Nucleic Acids Res*



- 35:W193-200. <https://doi.org/10.1093/nar/gkm226>
9. Smoot ME, Ono K, Ruscheinski J, et al (2011) Cytoscape 2.8: new features for data integration and network visualization. *Bioinformatics* 27:431–2. <https://doi.org/10.1093/bioinformatics/btq675>
  10. Walker HK, Hall WD, Hurst JW (1990) *Clinical Methods*. Butterworths
  11. Connor DE, Chaitanya G V., Chittiboina P, et al (2017) Variations in the cerebrospinal fluid proteome following traumatic brain injury and subarachnoid hemorrhage. *Pathophysiology* 24:169–183. <https://doi.org/10.1016/j.pathophys.2017.04.003>
  12. Huang P, Lo L-H, Chen Y-C, et al (2009) Serum free hemoglobin as a novel potential biomarker for acute ischemic stroke. *J Neurol* 256:625–31. <https://doi.org/10.1007/s00415-009-0133-x>
  13. Nezu T, Hosomi N, Aoki S, et al (2013) Alpha2-macroglobulin as a promising biomarker for cerebral small vessel disease in acute ischemic stroke patients. *J Neurol* 260:2642–2649. <https://doi.org/10.1007/s00415-013-7040-x>
  14. Hanson E, Kanse SM, Joshi A, et al (2012) Plasma factor VII-activating protease antigen levels and activity are increased in ischemic stroke. *J Thromb Haemost* 10:848–856. <https://doi.org/10.1111/j.1538-7836.2012.04692.x>
  15. Ziegler G, Harhausen D, Schepers C, et al (2007) TLR2 has a detrimental role in mouse transient focal cerebral ischemia. *Biochem Biophys Res Commun* 359:574–9. <https://doi.org/10.1016/j.bbrc.2007.05.157>
  16. Füst G, Munthe-Fog L, Illes Z, et al (2011) Low ficolin-3 levels in early follow-up serum samples are associated with the severity and unfavorable outcome of acute ischemic stroke. *J Neuroinflammation* 8:185. <https://doi.org/10.1186/1742-2094-8-185>

17. Carcaillon L, Alhenc-Gelas M, Bejot Y, et al (2011) Increased Thrombin Generation Is Associated With Acute Ischemic Stroke but Not With Coronary Heart Disease in the Elderly. *Arterioscler Thromb Vasc Biol* 31:1445–1451.  
<https://doi.org/10.1161/ATVBAHA.111.223453>
18. Cangemi C, Skov V, Poulsen MK, et al (2011) Fibulin-1 Is a Marker for Arterial Extracellular Matrix Alterations in Type 2 Diabetes. *Clin Chem* 57:1556–1565.  
<https://doi.org/10.1373/clinchem.2011.162966>
19. López-Farré AJ, Zamorano-León JJ, Segura A, et al (2012) Plasma desmoplakin I biomarker of vascular recurrence after ischemic stroke. *J Neurochem* 121:314–25.  
<https://doi.org/10.1111/j.1471-4159.2012.07683.x>
20. Chao J, Bledsoe G, Yin H, Chao L (2006) The tissue kallikrein-kinin system protects against cardiovascular and renal diseases and ischemic stroke independently of blood pressure reduction. *Biol Chem* 387:665–75. <https://doi.org/10.1515/BC.2006.085>
21. Garrod D, Chidgey M (2008) Desmosome structure, composition and function. *Biochim Biophys Acta - Biomembr* 1778:572–587.  
<https://doi.org/10.1016/j.bbamem.2007.07.014>
22. Gori AM, Giusti B, Piccardi B, et al (2017) Inflammatory and metalloproteinases profiles predict three-month poor outcomes in ischemic stroke treated with thrombolysis. *J Cereb Blood Flow Metab* 37:3253–3261.  
<https://doi.org/10.1177/0271678X17695572>
23. Kannemeier C, Al-Fakhri N, Preissner KT, Kanse SM (2004) Factor VII-activating protease (FSAP) inhibits growth factor-mediated cell proliferation and migration of vascular smooth muscle cells. *FASEB J* 18:728–730. <https://doi.org/10.1096/fj.03-0898fje>

24. Willeit J, Kiechl S, Weimer T, et al (2003) Marburg I polymorphism of factor VII—activating protease: a prominent risk predictor of carotid stenosis. *Circulation* 107:667–70
25. Wang L, Luo H, Chen X, et al (2014) Functional Characterization of S100A8 and S100A9 in Altering Monolayer Permeability of Human Umbilical Endothelial Cells. *PLoS One* 9:e90472. <https://doi.org/10.1371/journal.pone.0090472>
26. Schiopu A, Cotoi OS (2013) S100A8 and S100A9: DAMPs at the crossroads between innate immunity, traditional risk factors, and cardiovascular disease. *Mediators Inflamm* 2013:828354. <https://doi.org/10.1155/2013/828354>
27. Chao J, Bledsoe G, Chao L (2016) Protective Role of Kallistatin in Vascular and Organ Injury. *Hypertens (Dallas, Tex 1979)* 68:533–41. <https://doi.org/10.1161/HYPERTENSIONAHA.116.07861>
28. Stokes DL (2007) Desmosomes from a structural perspective. *Curr Opin Cell Biol* 19:565–71. <https://doi.org/10.1016/j.ceb.2007.09.003>
29. Yang Z, Bowles NE, Scherer SE, et al (2006) Desmosomal Dysfunction due to Mutations in Desmoplakin Causes Arrhythmogenic Right Ventricular Dysplasia/Cardiomyopathy. *Circ Res* 99:646–655. <https://doi.org/10.1161/01.RES.0000241482.19382.c6>
30. Argraves WS, Tran H, Burgess WH, Dickerson K (1990) Fibulin is an extracellular matrix and plasma glycoprotein with repeated domain structure. *J Cell Biol* 111:3155–64
31. Liu G, Cooley MA, Jarnicki AG, et al (2016) Fibulin-1 regulates the pathogenesis of tissue remodeling in respiratory diseases. *JCI Insight* 1:.

<https://doi.org/10.1172/jci.insight.86380>

32. Godyna S, Diaz-Ricart M, Argraves WS (1996) Fibulin-1 mediates platelet adhesion via a bridge of fibrinogen. *Blood* 88:2569–77
33. Yasmin, Maskari R Al, McEniery CM, et al (2018) The matrix proteins aggrecan and fibulin-1 play a key role in determining aortic stiffness. *Sci Rep* 8:8550.  
<https://doi.org/10.1038/s41598-018-25851-5>
34. Lee Y-B, Park J-H, Kim E, et al (2014) Arterial stiffness and functional outcome in acute ischemic stroke. *J Cerebrovasc Endovasc Neurosurg* 16:11–9.  
<https://doi.org/10.7461/jcen.2014.16.1.11>

## Figure Legends

**Figure 1.** Predicted protein networks associated with the proteins upregulated in the serum of the affected and unaffected twins. Cytoscape networks were constructed using (A) 19 proteins upregulated in the affected twin's serum (Table 1) and (B) 13 proteins upregulated in the unaffected twin's serum (Table 2). Yellow nodes indicate the seed proteins; blue nodes indicate interacting proteins. There was no interaction data for four of the proteins, CD5L, CD7, GPLD1 and IGLV3-19. Edges describe experimentally supported interactions.

## Tables

<i>Gene symbol</i>	<i>Approved gene name</i>	<i>UniProt ID</i>	<i>Run 1</i>		<i>Run 2</i>		<i>Average fold difference</i>
			<i>fold difference</i>	<i>P-value</i>	<i>fold difference</i>	<i>P-value</i>	
<i>FBLN1</i>	fibulin 1	P23142	2.79	8.27E-05	23.96	4.49E-02	13.37
<i>CA1</i> [11]	carbonic anhydrase 1	P00915	5.64	1.12E-05	4.84	5.89E-08	5.24
<i>PIGR</i>	polymeric immunoglobulin receptor	P01833	8.29	3.97E-02	1.96	7.36E-05	5.12
<i>IGKV1-5</i>	immunoglobulin kappa variable 1-5	P01602	2.22	1.31E-06	5.14	1.06E-05	3.68
<i>GPLD1</i>	glycosylphosphatidylinositol specific phospholipase D1	P80108	3.85	6.56E-06	2.08	1.82E-04	2.96
<i>IGKV3-20</i>	immunoglobulin kappa variable 3-20	P01620	2.38	3.47E-07	2.88	3.09E-03	2.63
<i>PZP</i>	PZP, alpha-2-macroglobulin like	P20742	2.91	4.16E-06	2.03	2.43E-05	2.47
<i>C5</i>	complement C5	P01031	3.44	4.77E-07	1.39	7.35E-05	2.42
<i>APOLI</i>	apolipoprotein L1	O14791	1.47	3.61E-03	2.94	2.17E-04	2.21
<i>HBA1</i> [12]	hemoglobin subunit alpha 1	P69905	1.64	6.91E-07	2.69	5.39E-09	2.17
<i>A2M</i> [13]	alpha-2-macroglobulin	P01023	2.78	3.70E-09	1.48	1.00E-05	2.13
<i>HABP2</i> [14]	hyaluronan binding protein 2	Q14520	2.17	8.05E-04	1.89	1.60E-02	2.03
<i>IGLV3-19</i>	immunoglobulin lambda variable 3-19	P01714	1.54	2.92E-06	2.14	4.06E-07	1.84
<i>C4BPA</i>	complement component 4 binding protein alpha	P04003	2.20	1.05E-08	1.39	3.56E-04	1.80
<i>C7</i>	complement C7	P10643	2.06	1.38E-04	1.49	1.06E-02	1.78
<i>FCN3</i> [16]	ficolin 3	O75636	1.76	6.90E-03	1.69	3.64E-04	1.72
<i>S100A8</i> [15]	S100 calcium binding protein A8	P05109	1.43	1.59E-02	1.89	1.27E-02	1.66

<i>CD5L</i>	CD5 molecule like	O43866	1.16	6.30E-03	1.91	2.36E-05	1.53
<i>F2</i> [17]	coagulation factor II, thrombin	P00734	1.29	1.20E-03	1.72	4.26E-05	1.50

**Table 1.** The top proteins identified at high levels in the affected twin compared with the unaffected twin (ranked by fold change) that were present in both runs. A fold change threshold of >1.5-fold was used as a cut off. References are provided next to the gene name for the proteins which have been associated with stroke or stroke risk factors by other studies.



<i>Gene name</i>	<i>Protein name</i>	<i>UniProt ID</i>	<i>Run 1</i>		<i>Run 2</i>		<i>Average fold difference</i>
			<i>fold difference</i>	<i>P-value</i>	<i>fold difference</i>	<i>P-value</i>	
<i>KRT17</i>	keratin 17	Q04695	13.33	4.42E-05	4.08	1.70E-05	8.71
<i>SPTB</i>	spectrin beta, erythrocytic	P11277	13.81	5.55E-07	2.60	7.52E-06	8.20
<i>SERPINA4</i> [20]	serpin family A member 4	P29622	6.95	3.33E-03	1.25	2.05E-02	4.10
<i>KRT16</i>	keratin 16	P08779	4.60	1.47E-06	2.99	6.10E-05	3.79
<i>DSP</i> [19]	desmoplakin	P15924	2.72	4.11E-05	1.82	1.45E-02	2.27
<i>IGHG4</i>	immunoglobulin heavy constant gamma 4 (G4m marker)	P01861	2.80	2.71E-06	1.60	1.18E-03	2.20
<i>DSG1</i>	desmoglein 1	Q02413	1.95	1.69E-06	1.99	2.96E-03	1.97
<i>KRT9</i>	keratin 9	P35527	2.21	4.44E-06	1.60	3.29E-05	1.91
<i>CPNI</i>	carboxypeptidase N subunit 1	P15169	2.10	8.88E-03	1.50	4.96E-03	1.80
<i>IGHA2</i>	immunoglobulin heavy constant alpha 2 (A2m marker)	P01877	1.80	1.59E-05	1.55	1.34E-03	1.68
<i>GAPDH</i>	glyceraldehyde-3-phosphate dehydrogenase	P04406	1.68	2.62E-03	1.63	2.27E-03	1.65
<i>HPX</i>	hemopexin	P02790	1.25	6.11E-02	1.81	2.15E-05	1.53
<i>IGHA1</i>	immunoglobulin heavy constant alpha 1	P01876	1.45	2.37E-04	1.58	6.44E-04	1.52

**Table 2.** The top proteins identified at high levels in the unaffected twin compared with the affected twin (ranked by fold change) that were present in both runs. A fold change threshold of >1.5-fold was used as a cut off. References are provided next to the gene name for the proteins which had been associated with stroke by other studies.

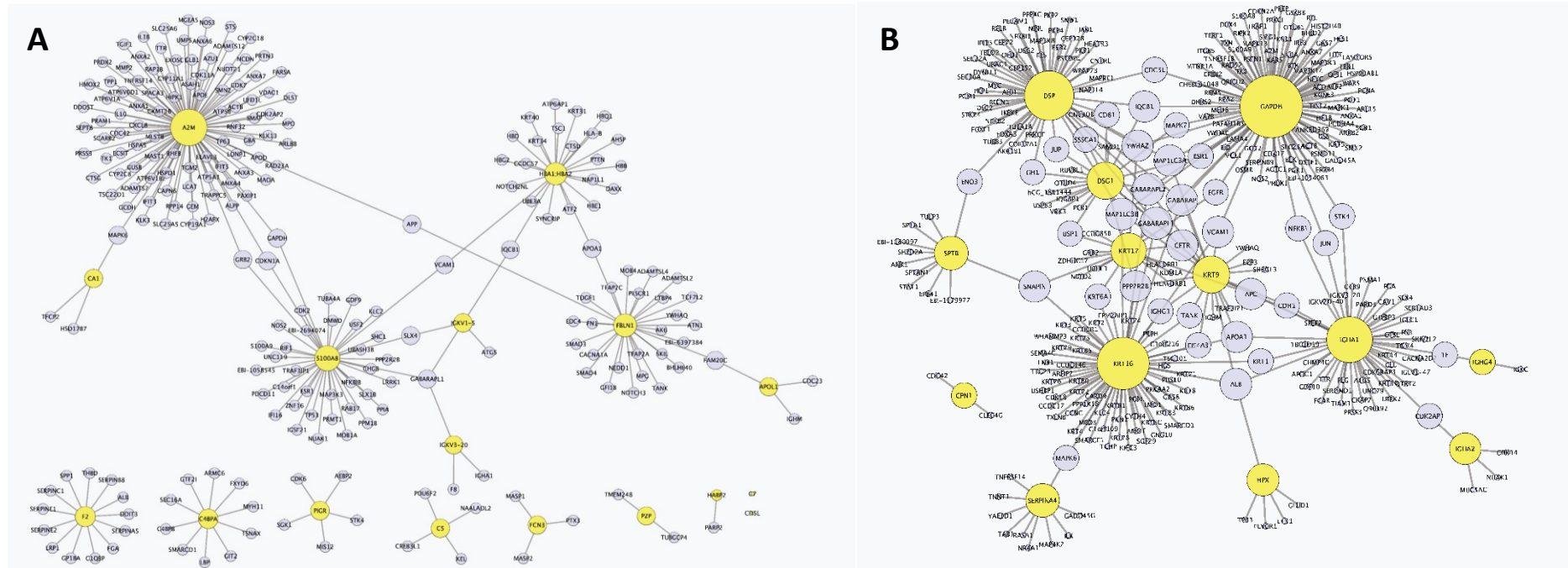
<i>GO term</i>	<b>P-value</b>	<b>S</b>	<b>T</b>	<b>P-value</b>	<b>S</b>	<b>T</b>
		<b>t=50</b>	<b>t=18,711</b>		<b>t=77</b>	<b>t=18,711</b>
	<b>Run 1</b>			<b>Run 2</b>		
<i>response to stress</i>	1.59E-11	35	3579	1.28E-08	42	3579
<i>wound healing</i>	5.06E-11	17	540	1.32E-05	15	540
<i>haemostasis</i>	7.89E-10	14	357	6.87E-05	12	357
<i>blood coagulation</i>	6.51E-10	14	352	5.88E-05	12	352
<i>fibrin clot formation</i>	5.18E-12	8	26	2.75E-08	7	26
<i>lipoprotein particle</i>	1.42E-04	5	39	3.43E-05	6	39

**Table 3.** Selection of enriched GO terms associated with the proteins present at high levels in the affected twin. GO terms were identified as significantly enriched in the high level proteins in the affected twin following g:Profiler analysis. P-values <0.05 are considered as significantly enriched. **S** = number of protein identifiers (IDs) in both the study dataset and GO term group, **T** = number of human protein IDs associated with the GO term, **t** = number of protein IDs in study or GO datasets (Supplemental Table 2).

<i>GO term</i>	<b>P-value</b>	<b>S</b>	<b>T</b>	<b>P-value</b>	<b>S</b>	<b>T</b>
		<b>t=45</b>	<b>t=18,711</b>		<b>t=62</b>	<b>t=18,711</b>
	<b>Run 1</b>			<b>Run 2</b>		
<i>epidermis development</i>	5.01E-06	10	297	1.55E-02	8	297
<i>skin development</i>	1.52E-05	9	242	2.57E-04	9	242
<i>establishment of skin barrier</i>	3.04E-02	3	18	8.50E-04	4	18
<i>intermediate filament cytoskeleton</i>	1.15E-12	14	247	3.96E-03	8	247

**Table 4. Selection of enriched GO terms associated with the proteins at high levels in the unaffected twin. GO terms were identified as significantly enriched in the high level proteins in the unaffected twin following g:Profiler analysis. P-values <0.05 are considered as significantly enriched. S = number of protein identifiers (IDs) in both the study dataset and GO term group, T = number of human protein IDs associated with the GO term, t = number of protein IDs in study or GO datasets (Supplemental Table 2).**

# Figures



**Figure 1**