의학석사 학위논문

고분해능 질량분석기 기반의 정량적 단백체 기법을 활용한 피질하 경색과 백질뇌증을 동반하는 상염색체 우성 뇌동맥질환 특이적 혈장 바이오마커 발굴 연구

Biomarker discovery of cerebral autosomal dominant arteriopathy with subcortical infarcts and leukoencephalopathy using quantitative proteomics approach based on high resolution mass spectrometry

울 산 대 학 교 대 학 원

의 과 학 과

홍 정 연

고분해능 질량분석기 기반의 정량적 단백체 기법을 활용한 유전성 다발경색치매 특이적 혈장 바이오마커 발굴 연구

지도교수 김 경 곤

이 논문을 의학석사 학위 논문으로 제출함

2023 년 8 월

울 산 대 학 교 대 학 원
의 과 학 과
홍 정 연

홍정연의 의학석사학위 논문을 인준함

심사위원 이 은 재 인

심사위원 김 상 엽 인

심사위원 김 경 곤 인

울 산 대 학 교 대 학 원

2023 년 8 월

Abstract

Cerebral Autosomal Dominant Arteriopathy with Subcortical Infarcts and Leukoencephalopathy (CADASIL) is a rare genetic disorder that affects small blood vessels in the brain, leading to various neurological symptoms. The most common symptoms are abnormal cerebral white matter morphology and leukoencephalopathy, and cerebral ischemia, cognitive impairment, emotional lability, migraine, and stroke also occur frequently. The best method for diagnosing CADASIL is through genetic testing for NOTCH3 mutations, but genetic testing may not always be possible or feasible, and not all CADASIL patients have pathogenic NOTCH3 mutations. Therefore, there is a need for research to discover blood-based biomarkers that can aid in the diagnosis and treatment of CADASIL.

This study aims to contribute to the improvement of the quality of life for CADASIL patients by identifying biomarkers for diagnosis and monitoring through plasma protein analysis. The study also aims to provide evidence for the management and treatment of CADASIL. Plasma samples were collected from three groups of subjects with similar ages (67 ± 12.5 years) with 5 males and 5 females in each group (CADASIL patient group, stroke patient group, and healthy control group, total of 10 subjects for each group). The collected plasma samples underwent pretreatment, in which the top 14 high-concentration plasma proteins were removed using the Multi Affinity Removal System Human 14 (MARS14) column (100 × 4.6 mm, Agilent Technology, Palo Alto, CA, USA). The remaining plasma proteins that were not removed by the column were harvested and digested using Suspension Trap (S-Trap) digestion method with trypsin/LysC at 37℃ for 16 hours. Using the Dionex UltiMate 3000 RSLCnano system (Thermo Fisher Scientific, Waltham, MA, USA) and Q Exactive HF-X mass spectrometer (Thermo Fisher Scientific), spectrum was acquired for quantitative analysis of 1,902 plasma proteins using Proteome Discoverer (Thermo Fisher Scientific, ver 2.3) equipped with the Sequest HT algorithm. From this analysis, quantitative and qualitative information for each group was obtained, and differentially expressed proteins (DEPs) were identified. And principal component analysis, Venn diagram analysis, and volcano plot analysis were performed for additional statistical result. Candidate pathways were analyzed through gene ontology (GO) analysis using DEP information. The expression results of potential CADASIL plasma biomarkers identified in this study were compared

with those from previous studies. From the result, small glutamine-rich tetratricopeptide repeat-containing protein alpha (SGTA) was found to be significantly increased in CADASIL patients compared to both groups, while Treacle protein (TCOF1) showed a decreasing trend. These protein candidates may be utilized for CADASIL screening and monitoring using patient plasma with further validation studies and are expected to contribute to the management and treatment of CADASIL patients.

Keyword: CADASIL, Proteomics, Biomarker, TCOF1, SGTA, LC-MS

TABLE OF CONTENTS

List of Tables

LIST OF FIGURES

1. Introduction

1.1. Cause and characteristics of CADASIL

Cerebral autosomal dominant arteriopathy with subcortical infarcts and leukoencephalopathy (CADASIL) is an uncommon hereditary condition that impacts the small blood vessels in the brain, resulting in various neurological symptoms [1, 2]. The most common symptoms include abnormal cerebral white matter morphology and leukoencephalopathy, as well as frequent occurrences of cerebral ischemia, cognitive impairment, emotional lability, migraines, and stroke [3]. It was first described in 1976 [4, 5] and has since been recognized as a distinct entity among genetic disorders that affect the brain [4, 5, 6]. CADASIL was first identified more than two decades ago, yet it remains a challenging disease to manage due to several medical unmet needs[2, 4]. CADASIL is caused by mutations in the NOTCH3 gene, which codes for a transmembrane receptor protein involved in cell signaling and homeostasis [1, 2, 4, 5, 6, 7, 8, 9].

Despite being a rare disease, CADASIL is a significant cause of stroke in young and middle-aged individuals, with an estimated prevalence of stroke [4, 5, 6, 9]. It has the distinct clinical and imaging features [1, 4], but the diagnosis of CADASIL-related stroke remains a challenge [2, 7, 8, 9]. The gold standard for diagnosing CADASIL is genetic testing for NOTCH3 mutations [4, 8, 10]. Pathogenic mutations in CADASIL are characterized by their distinctive impact on the cysteine residues within the EGFr domains of the NOTCH3 protein [11]. These mutations often result in the loss or gain of a cysteine residue in one of the 34 EGFr domains [11]. While most mutations are missense mutations, there have been reports of small deletions, insertions, and splice-site mutations that also lead to alterations in cysteine residues [11]. The presence of cysteine-altering mutations as a common feature in CADASIL is a matter of debate, as there are conflicting reports suggesting the pathogenicity of other types of mutations as well [11]. In such cases, a diagnosis can be made through a skin biopsy; however, it has low sensitivity. [11, 12]. Due to these reasons, genetic testing is not always feasible. Furthermore, it is important to note that not everyone with NOTCH3 mutations develops CADASIL [10].

There is a lack of consensus on the clinical criteria for diagnosing CADASIL, and the differential diagnosis of CADASIL includes other types of strokes, such as small vessel disease and cardioembolic

stroke [4, 13, 14]. Therefore, there is a need for research on biomarkers that can aid in the diagnosis and treatment of CADASIL. There are several ways for identifying for biomarker such as proteomics analysis, metabolomics analysis [15]. Proteomic analysis has emerged as a promising research direction for CADASIL, as well as for various other diseases [5, 7, 15, 16]. For these reasons, a study was conducted to identify biomarkers in the plasma of CADASIL patients using liquid chromatography-mass spectrometry (LC-MS), and this thesis presents the findings of that research.

1.2. Biomarkers of CADASIL and its application in clinics

Biomarkers are defined as measurable indicators of biological processes, such as genes, proteins, or metabolites, which can provide information about the status of a disease or the response to treatment [17, 18]. Identifying and validating biomarkers require multidisciplinary approach from various fields such as genomics, proteomics, metabolomics, and bioinformatics knowledge and techniques [17, 18, 19]. These technologies allow for the simultaneous analysis of thousands of molecules, which can provide a comprehensive view of the molecular changes associated with disease [18].

One of the major challenges in clinical medicine is the accurate diagnosis and monitoring of diseases. Biomarkers not only make it possible, but also useful for predicting treatment response, prognosis and identifying patients who are at higher risk of developing a particular condition [17, 18, 20].

In recent years, significant progress has been made in the identification of biomarkers for a wide range of diseases, including CADASIL. Several studies have investigated the proteomic profile of blood samples from CADASIL patients to identify potential biomarkers [7, 21, 22, 23]. Table 1,2 presents potential biomarkers for CADASIL provided by Jeju National University. The development and validation of reliable biomarkers for CADASIL could have significant clinical implications, including early diagnosis, accurate disease monitoring, and the development of new therapeutic strategies.

Blood biomarkers			
Name [References]	**Origin**	**Findings in CADASIL**	**Biological function**
Neurofilament light chain [21, 22, 23]	Patients with CADASIL	Elevated in serum Associated with disability Predicted stroke and cognitive decline	Main element of neuroaxonal cytoskeleton
Notch3 ECD [7, 24]	Mouse model of CADASIL and patients with CADASIL	Decreased in blood Correlated with white matter hyperintensities	Unknown
HTRA1 [7]	Mouse model of CADASIL	Elevated	Serine protease TGF signaling
Endostatin [7]	Mouse model of CADASIL	Elevated	Inhibitor of angiogenesis

Table 1. Potential biomarkers detected in the blood of CADASIL provided by Jeju National University

Biomarkers found in proteomic analysis of brain vessels			
Name [References]	**Origin**	**Findings in CADASIL**	**Biological function**
Serum amyloid P-component [25, 26, 27]	Postmortem or biopsied brain vessels of CADASIL patients	Elevated Co-localized with NOTCH3	Amyloid formation
TIMP3 [26]	Postmortem brain vessels of CADASIL patients	Elevated Co-localized with NOTCH3	Inhibitor of matrix metalloproteinases
Vitronectin [25, 26, 27]	Postmortem or biopsied brain vessels of CADASIL patients	Elevated Co-localized with NOTCH3	Cell adhesion
Other extracellular matrix proteins [25, 26, 27]	Postmortem or biopsied brain vessels of CADASIL patients	Elevated	Collagen (1α2, 8α1, 12α1, 14α1, 18α1), Laminin (α5, γ1) Lactadherin, Leucin rich repeat proteoglycan, Norrin, Biglycan, Microfibril-associated glycoprotein 4, Annexin A2, and Periostin

Table 2. Potential biomarkers detected in the brain vessels of CADASIL provided by Jeju National University

1.3. Proteomics research of CADASIL

Proteomics is a rapidly growing field of study that aims to comprehensively understand the structure, function, and interactions of proteins within biological systems [20, 28]. This field has emerged as a powerful tool for discovering and characterizing biomolecules in complex samples such as blood, tissues, and cells, and has wide-ranging applications in the fields of medicine, biotechnology [20].

Proteomic analysis involves the identification, quantification, and characterization of proteins using a variety of techniques such as MS, protein microarrays, and gel electrophoresis [20]. With advances in technology and computational tools, proteomics has become increasingly high-throughput and sensitive, enabling researchers to explore complex biological systems in unprecedented detail [20, 28].

The potential of proteomics for advancing our understanding of disease mechanisms and identifying new therapeutic targets has been widely recognized. For example, several studies have investigated the proteomic profile of blood samples from CADASIL to identify potential biomarkers [7, 21, 22, 23, 24, 25, 26, 27]. These studies have used various proteomic techniques such as two-dimensional gel electrophoresis, MS, and protein microarrays to identify differentially expressed proteins in CADASIL patients compared to healthy controls [7, 25, 27].

2. Materials and Methods

2.1. Information of clinical samples

Plasma samples for this study were obtained from cohorts of CADASIL, stroke, and healthy control groups registered at Jeju National University Hospital. The study was approved by the Institutional Review Board (IRB) of Jeju National University (Jeju, Korea; IRB-e no. 2020-08-010). 10 people blood samples were collected from each group, and enough protein suitable for analysis were extracted from all plasma samples, making them suitable for the final analysis. Table 3 is CADASIL patient information provided by Jeju National University. The Korean mini-mental state examination (K-MMSE) was used for cognitive function evaluation [29, 30]. K-MMSE is a standard that reflects age and education level based on the existing MMSE and has a maximum score of 30 points [30]. Scores of 23 points or less are considered suspicious of dementia [30]. The modified Rankin Scale (mRS) is a measure used to assess the overall disability status of patients with stroke or neurological disorders [29, 31]. The term "modified" was added because the scale now includes a Grade 0 to indicate no symptoms and a Grade 6 to indicate death, in addition to the original Grade 1-5 [31]. Grade 1 indicates "No significant disability despite symptoms," Grade 2 indicates "Slight disability," and Grade 3 indicates "Moderate disability" [31]. Lacunes refer to small cavities filled with cerebrospinal fluid (CSF) that range in size from 3 to 15 mm [32]. These cavities are typically found in the basal ganglia or white matter of the brain and are often detected incidentally on medical imaging in older individuals [29, 32]. Lacunes are not typically associated with specific neurological symptoms, although they may contribute to cognitive decline or other age-related neurological changes [32]. The brain parenchymal fraction (BPF) is a measure of the degree of brain atrophy [29]. It is information that can be obtained through brain MRI, such as Lacunes [29].

Study ID	Gender	Age	K-MMSE	mRS	Total lacunes	BPF
Patient1	Male	46	30	1	19	0.695
Patient2	Male	60	30	0	6	0.770
Patient3	Male	65	30	1	10	0.661
Patient4	Male	79	24	3	25	0.552
Patient5	Male	85	20	1	6	0.593
Patient6	Female	52	30	0	5	0.863
Patient7	Female	59	29	0	2	0.730
Patient8	Female	68	24	0	8	0.728
Patient9	Female	73	25	0	16	0.650
Patient10	Female	78	24	2	3	0.654

* K-MMSE: Korean mini-mental state examination, mRS: modified Rankin Scale, BPF: brain parenchymal fraction

Table 3. CADASIL patient group's patient information provided by Jeju National University

The age information for each group was presented in a table 4. The mean age was 66.5 (±11.9) years for the healthy control group, 65.9 (±12.4) years for the stroke patient group, and 66.5 (±11.9) years for the CADASIL patient group. Although there were slight differences in the standard deviations, the age composition was matched across the groups. Specifically, the age composition of the healthy control group was completely matched with the CADASIL patient group, and the composition of the stroke patient group was matched with a difference of only 1-2 years, ensuring a statistically equivalent condition for the study.

Study ID	Sex	Age	Study ID	Sex	Age	Study ID	Sex	Age
Control 1	Male	46	**Stroke 1**	Male	42	**Patient 1**	Male	46
Control 2	Male	60	**Stroke 2**	Male	61	**Patient 2**	Male	60
Control 3	Male	65	**Stroke 3**	Male	63	**Patient 3**	Male	65
Control 4	Male	79	**Stroke 4**	Male	79	**Patient 4**	Male	79
Control 5	Male	85	**Stroke 5**	Male	85	**Patient 5**	Male	85
Control 6	Female	52	**Stroke 6**	Female	52	**Patient 6**	Female	52
Control 7	Female	59	**Stroke 7**	Female	59	**Patient 7**	Female	59
Control 8	Female	68	**Stroke 8**	Female	69	**Patient 8**	Female	68
Control 9	Female	73	**Stroke 9**	Female	73	**Patient 9**	Female	73
Control 10	Female	78	**Stroke 10**	Female	76	**Patient 10**	Female	78
Mean age	66.5 (±11.9)		**Mean age**	65.9 (±12.4)		**Mean age**	66.5 (±11.9)	

Table 4. Age information by group provided by Jeju National University

In this study, stroke patients were selected as the comparative control group for CADASIL patients. Stroke is a condition that occurs when the blood vessels in the brain are blocked or burst, causing bleeding [33]. When blood flow to a part of the brain is blocked, the brain tissue that does not receive blood stops functioning and dies [33]. This can result in paralysis, speech and language disorders, cognitive impairment, and even death [33]. CADASIL is a rare genetic disorder that increases the risk of stroke. While it shares similarities in phenotype and symptoms with stroke, it is a distinct disease [4, 5, 8, 13]. Therefore, this study aimed to investigate the differences in protein expression between CADASIL and stroke, and furthermore help in differential diagnosis.

2.2. Sample Preparation

2.2.1. Plasma samples depletion

For protein analysis, 30 clinical plasma samples were prepared for LC-MS analysis. 40 µL of plasma sample and 125 µL of Buffer A (Agilent Technologies, USA) were put in a spin filter (0.22µm, cellulose acetate, Agilent Technologies, USA) and centrifuged at 16,000 g for 2 minutes. After transferring 160 µL of the filtered sample into a vial, we removed 14 highly abundant proteins by loading it onto a Multi Affinity Removal System Human 14 (MARS14) column (100 × 4.6 mm, Agilent Technology, Palo Alto, CA, USA) of a Shimadzu binary HPLC system (20A Prominence, Shimadzu, Tokyo, Japan).

2.2.2. Peptidization for LC-MS analysis

The unbound fraction was subjected to lyophilization using an evaporator (CentriVap Cold Traps; Labconco, Kansas City, MO, USA). The resulting dried samples were then resuspended in 200 µL of 5% SDS in 50 mM TEAB. To reduce disulfide bonds, dithiothreitol (DTT) was added to a final concentration of 20 mM and the samples were incubated at 95 and 750 rpm for 10 minutes. Next, a final concentration of 40 mM iodoacetamide (IAA) was added and the samples were incubated in the dark at room temperature for 30 minutes. The acidified samples were then added to 350 µL of Suspension Trap digestion (S-Trap) binding buffer (90% methanol, 100 mM Triethylammonium bicarbonate buffer (TEAB) (pH 7.55)), prepared by a 10-fold dilution of 12% phosphoric acid. The S-Trap spin column (ProtiFi, Long Island, New York, USA) was then used to perform centrifugation at 4,000 g for 30 seconds. After washing the spin column with 400 µL of S-trap binding buffer and centrifuging at 4,000 g for 30 seconds, the washing step was repeated three times. Finally, the spin column was transferred to a new 2mL sample tube and a trypsin/Lys-C Mix with a protein to Trypsin/Lys-C mixture ratio of 25:1 (Promega, Madison, WI, USA) dissolved in 50 mM TEAB was added to the S-trap spin column. The column containing Trypsin/Lys-C was incubated at 37 for 16 hours without shaking [34]. Peptide elution was performed three times, first by adding 80 µL of 50 mM TEAB, centrifuging at 1,000 g for 1 minute, and then by adding 80 µL of 0.2% formic acid, followed by centrifugation at 1,000 g for 1 minute. In the final elution step, 100 µL of 0.2% formic acid and 50%

acetonitrile were added, and the sample was centrifuged at 4,000 g for 1 minute to elute the peptides. The eluted peptides were dried using an evaporator combined with a cold trap and stored at -80 until use.

2.3. Nano-LC-ESI-MS/MS Analysis

The dried peptide samples were dissolved in buffer A (0.1% formic acid in HPLC water), and the total peptide concentration was measured using a UV/Vis spectrophotometer (NanoDrop One, Thermo Fisher Scientific) at a wavelength of 280 nm, with the sample type option set to "1 Abs = 1 mg/mL". Based on the measured values, 40 µg of all samples were resuspended in 40 µL. Based on the measured values, 40 µg of all samples were resuspended in 40 µL. Used LC system was an Dionex UltiMate 3000 RSLCnano system (Thermo Fisher Scientific, Waltham, MA, USA). Mobile phase A was 0.1% formic acid and 5% DMSO in water and mobile phase B was 0.1% formic acid, 5% DMSO and 80% acetonitrile in water. The samples were reconstituted with 25 µL of mobile phase A, injected with a full sample loop injection of 5 µL into a C18 Pepmap trap column (20 × 100 µm i.d., 5 µm, 100 Å; Thermo Fisher Scientific), and separated in an Acclaim™ Pepmap 100 C18 column (500 × 75 µm i.d., 3 µm, 100 Å; Thermo Fisher Scientific) over 200 minutes (at a flow rate of 250 nL/min) at 50 °C. The column was pre-equilibrated with 95% mobile phase A and 5% mobile phase B. A gradient of 5–40% B was applied for 150 minutes, followed by a gradient of 40–95% B for 2 minutes, and then held at 95% B for 23 minutes. This was followed by a gradient of 95–5% B for 10 minutes, and then held at 5% B for 15 minutes. The LC system was coupled to a Q Exactive HF-X mass spectrometer (Thermo Fisher Scientific) with a nano-ESI source. The instrument was operated in data-dependent mode, with one scan cycle including one MS1 scan at a resolution of 70,000 at m/z 400, followed by 20 MS2 scans in higher energy collisional dissociation mode to fragment the 20 most abundant precursor ions identified in the MS1 spectrum. The target value for MS1 by Orbitrap was 3×10^6 with a maximum injection time of 100 ms. The ion target value for MS2 was set to 1×10^6 with a maximum injection time of 50 ms and a resolution of 17,500 at m/z 400. The dynamic exclusion was enabled with the following settings: repeat count = 1 and exclusion duration = 20 s.

2.4. Protein Identification by Database Search

For the MS analysis, each individual raw file was searched against the reviewed Human Uniprot-SwissProt protein database (released on July 2019) using SEQUEST-HT on Proteome Discoverer (Version 2.3, Thermo Fisher Scientific) [35]. The search parameters included a 10-ppm tolerance for precursor ion mass and 0.02 Da for fragmentation mass, while tryptic peptides were allowed up to two false cleavages. Fixed modification was set as carbamidomethylation of cysteines, while variable modifications included N-terminal acetylation and methionine oxidation. The false discovery rate (FDR) was calculated using the target-decoy search strategy, and peptides within 1% of the FDR were selected using the semi-supervised learning tool Percolator based on the SEQUEST results [36]. Label-free quantitation (LFQ) of proteins was determined using the precursor ion peak intensity for unique and razor peptides of each protein, and peptides with methionine oxidation were excluded.

3. Results

3.1. Plasma protein expression between CADASIL, Stroke and Healthy groups

3.1.1. Commonalities and differences in protein expression among three groups

In a study involving 30 participants, depleted plasma samples were analyzed using single LC-MS/MS runs to identify constitutive proteins. Using the Dionex UltiMate 3000 RSLCnano system (Thermo Fisher Scientific, Waltham, MA, USA) and Q Exactive HF-X mass spectrometer (Thermo Fisher Scientific), were acquired for 1,902 plasma proteins using equipment equipped with Proteome Discoverer (Thermo Fisher Scientific, ver 2.3) and the Sequest HT algorithm. The spectrum abundance of each sample through Proteome Discoverer was represented in a box plot (Fig. 1). After analyzing the -spectrum abundance of each file, it was observed that the median values align after normalization. This observation was utilized to perform statistical analysis (Fig. 1).

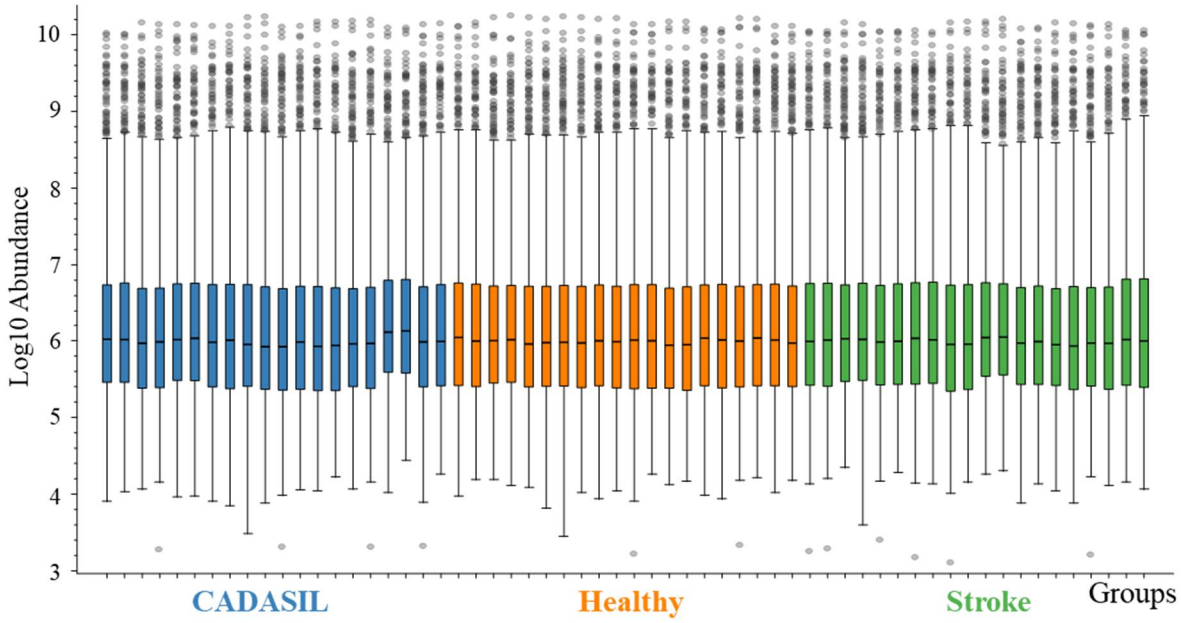


Figure 1. Box plot showing the spectrum abundance of each sample by Proteome Discoverer

CADASIL patient group and healthy control group were compared (Fig. 2). When compared with Venn diagrams, a total of 1,835 proteins in the CADASIL patient group and 1687 proteins in the healthy control group were quantified. Among them, 1664 proteins were quantified in both groups, and the number of CADASIL specific proteins was 171 and the number of proteins specific to the healthy control group was 23 (Fig. 2A). The volcano plot is a graphical representation that shows the relationship between the $\log_2$ fold changes and the *p*-values of all proteins in a dataset, comparing the CADASIL patient group and the healthy control group (Fig. 2B). The plot allows for visual identification of significant changes in protein expression between the two groups, with proteins that have high levels of significance (low *p*-values) and large fold changes appearing farther to the right or left on the plot, resembling the shape of a volcano (Fig. 2B). Plasma proteins that were upregulated with a fold change greater than two and a *p*-value of less than 0.05 are represented as red circles on the volcano plot, while proteins that were downregulated with the same fold change and *p*-value are represented as green circles. The gray circles correspond to plasma proteins that did not show statistically significant differences. The number of proteins in the CADASIL patient group showing significant change was 239, and the number of proteins in the healthy control group was 41 (Fig. 2B).

CADASIL patient group and stroke patient group were compared (Fig. 3). A total of 1,687 proteins in the stroke patient group were quantified. Among them, 1,672 proteins were quantified in both groups, and the number of CADASIL specific proteins was 163 and the number of proteins specific to the stroke patient group was 15 (Fig. 3A). The volcano plot displays the Log_2 fold changes and the corresponding *p*-values of all proteins between the CADASIL patient group and the stroke patient group (Fig. 3B). The number of proteins in the CADASIL patient group showing significant change was 211, and the number of proteins in the stroke patient group was 67 (Fig. 3B).

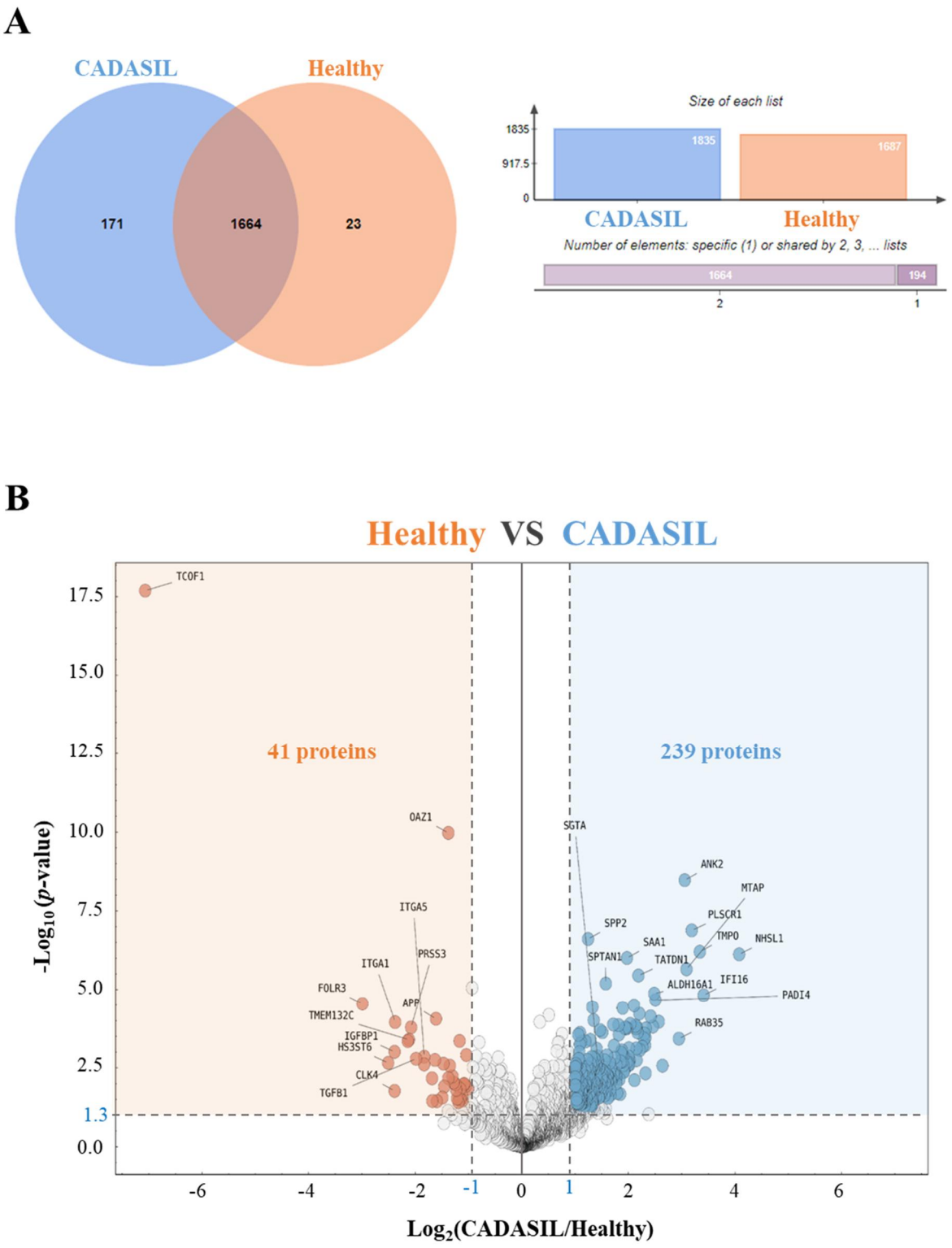


Figure 2. Comparison of proteins from a CADASIL patient group with a healthy control group (**A**) Venn diagram showing the number of common and specific proteins in each group (**B**) Volcano plot showing significant specific proteins in each group

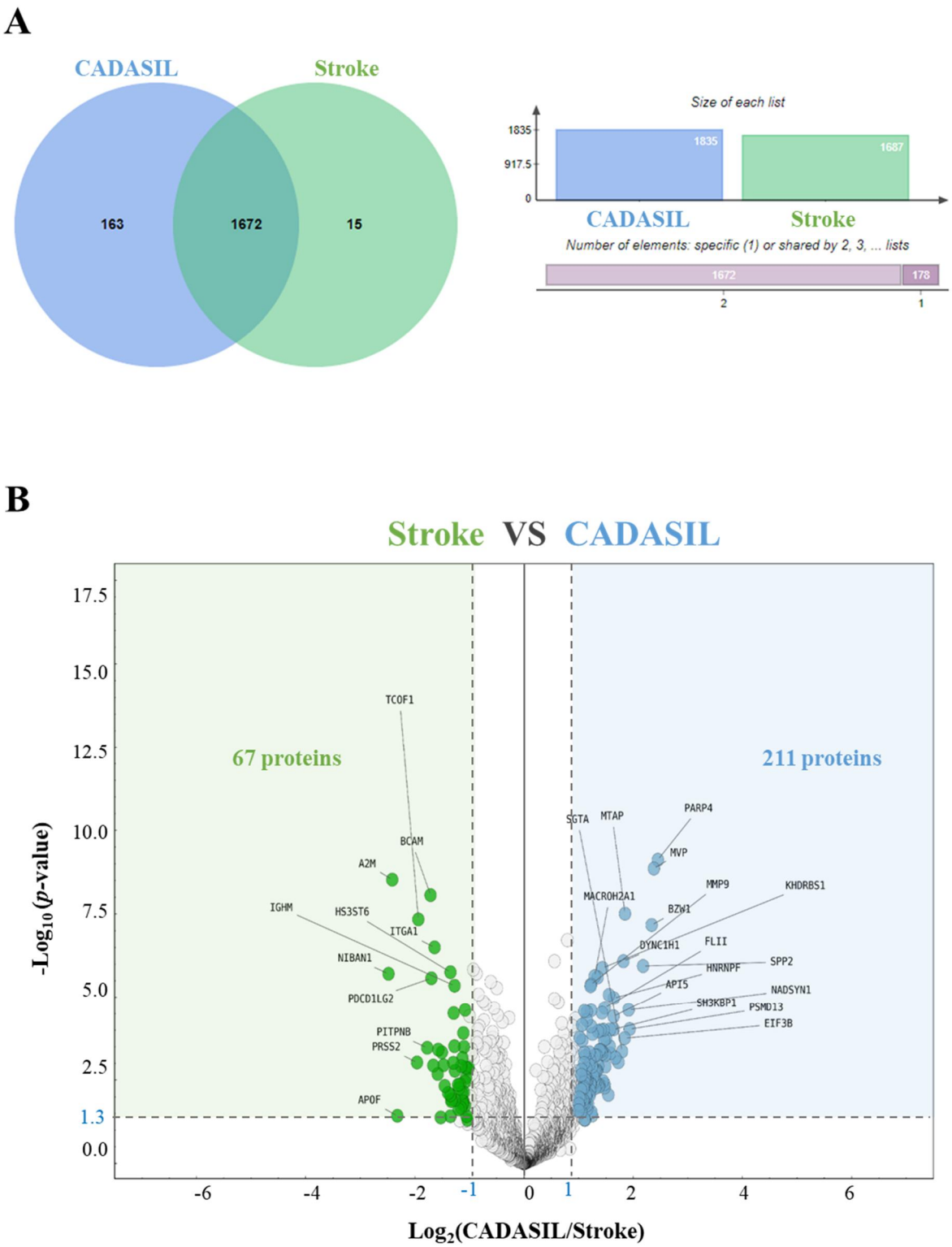


Figure 3. Comparison of proteins from a CADASIL patient group with a stroke patient group (**A**) Venn diagram showing the number of common and specific proteins in each group (**B**) Volcano plot showing significant specific proteins in each group

CADASIL patient group, healthy control group, stroke patient group were compared (Fig. 4). When each of the three groups was compared with a Venn diagram, a total of 1835 proteins in the CADASIL patient group, 1687 proteins in the stroke patient group, and 1687 proteins in the healthy control group were quantified (Fig. 4A). Among them, 1613 proteins were quantified in all three groups, and the number of specific proteins was 112 in the CADASIL patient group, 7 in the stroke patient group, and 15 in the healthy control group (Fig. 4A). Based on these results, 112 CADASIL patient group-specific proteins (Table 5) and 7 stroke patient group-specific proteins (Table 6) were organized into a table.

Based on the PD search results, principal component analysis (PCA) was performed for the three groups (Fig. 4B). PCA of the CADASIL patient group, stroke patient group, and healthy control group apparently grouped them in PC1 (28.8% explained variance) and PC2 (6.8% explained variance) (Fig. 4B). Based on the results of principal component analysis, it was observed that the three groups were not well separated. However, in this state, proteins that show significant differences through additional statistical analysis are more meaningful as potential biomarkers.

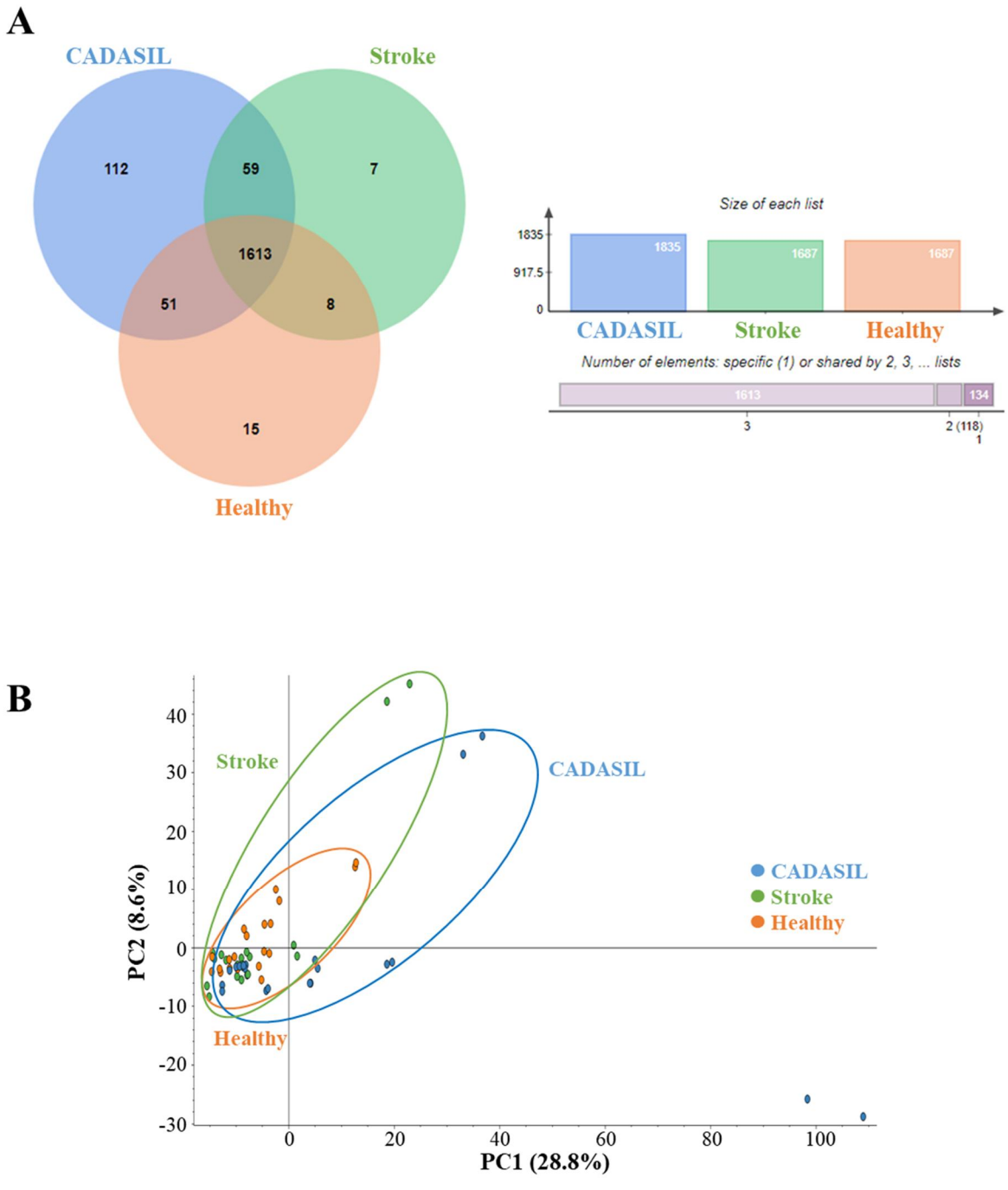


Figure 4. Comparison of proteins from a CADASIL patient group with a stroke patient group and a healthy control group (**A**) Venn diagram showing the number of common and specific proteins in each group (**B**) Principal component analysis (PCA) of the three groups by the proteins

Elements only in CADASIL	Protein name	Peptide count
P40763-2	Isoform Del-701 of Signal transducer and activator of transcription 3	6
P06753	Tropomyosin alpha-3 chain	6
O60437	Periplakin	5
P18085	ADP-ribosylation factor 4	4
O43143	Pre-mRNA-splicing factor ATP-dependent RNA helicase DHX15	4
Q92817	Envoplakin	4
Q9UMS4	Pre-mRNA-processing factor 19	3
Q92614	Unconventional myosin-XVIIIa	2
Q92499	ATP-dependent RNA helicase DDX1	2
Q8WVY7	Ubiquitin-like domain-containing CTD phosphatase 1	2
P27482	Calmodulin-like protein 3	2
Q96B97	SH3 domain-containing kinase-binding protein 1	2
P12081	Histidine--tRNA ligase, cytoplasmic	2
Q15208	Serine/threonine-protein kinase 38	2
Q13277	Syntaxin-3	2
P10412	Histone H1.4	2
Q8IZ07	Ankyrin repeat domain-containing protein 13A	2
Q92835	Phosphatidylinositol 3,4,5-trisphosphate 5-phosphatase 1	2
P07902	Galactose-1-phosphate uridylyltransferase	1
Q9Y678	Coatomer subunit gamma-1	1
O00487	26S proteasome non-ATPase regulatory subunit 14	1
Q9NTI5	Sister chromatid cohesion protein PDS5 homolog B	1
Q01844	RNA-binding protein EWS	1
Q86VS8	Protein Hook homolog 3	1
O95319	CUGBP Elav-like family member 2	1
P30626	Sorcin	1
Q9Y5L0-1	Isoform 1 of Transportin-3	1
Q14677	Clathrin interactor 1	1
P62913	60S ribosomal protein L11	1
Q9Y333	U6 snRNA-associated Sm-like protein LSm2	1
Q8TCU6	Phosphatidylinositol 3,4,5-trisphosphate-dependent Rac exchanger 1 protein	1
Q6ZVX7	F-box only protein 50	1
O00560	Syntenin-1	1
O60826	Coiled-coil domain-containing protein 22	1
Q9P016	Thymocyte nuclear protein 1	1
Q9Y371	Endophilin-B1	1
P62714	Serine/threonine-protein phosphatase 2A catalytic subunit beta isoform	1
P56192	Methionine--tRNA ligase, cytoplasmic	1
Q9UQE7	Structural maintenance of chromosomes protein 3	1
P80748	Immunoglobulin lambda variable 3-21	1
Q01433	AMP deaminase 2	1
Q99719	Septin-5	1
P62277	40S ribosomal protein S13	1
P21964	Catechol O-methyltransferase	1
P05198	Eukaryotic translation initiation factor 2 subunit 1	1
O00244	Copper transport protein ATOX1	1
Q6P4A8	Phospholipase B-like 1	1
P36578	60S ribosomal protein L4	1
Q14498	RNA-binding protein 39	1
Q2TAA2	Isoamyl acetate-hydrolyzing esterase 1 homolog	1
P26599	Polypyrimidine tract-binding protein 1	1
Q5T5C0	Syntaxin-binding protein 5	1
Q8IVF2	Protein AHNAK2	1
Q99986	Serine/threonine-protein kinase VRK1	1
P35611	Alpha-adducin	1
Q8WVV4-1	Isoform 1 of Protein POF1B	1
P46776	60S ribosomal protein L27a	1
P62701	40S ribosomal protein S4, X isoform	1
O75884	Serine hydrolase RBBP9	1
Q9NUV9	GTPase IMAP family member 4	1

Table 5. (continue)

Elements only in CADASIL	Protein name	Peptide count
Q99418	Cytohesin-2	1
P53609	Geranylgeranyl transferase type-1 subunit beta	1
P54619	5'-AMP-activated protein kinase subunit gamma-1	1
Q8NFT8	Delta and Notch-like epidermal growth factor-related receptor	1
P51665	26S proteasome non-ATPase regulatory subunit 7	1
Q9NS98	Semaphorin-3G	1
Q5T1M5	FK506-binding protein 15	1
Q9P000	COMM domain-containing protein 9	1
Q9Y2Z0	Protein SGT1 homolog	1
Q8WWW0	Ras association domain-containing protein 5	1
P56377	AP-1 complex subunit sigma-2	1
Q9Y5K8	V-type proton ATPase subunit D	1
Q5VT79	Annexin A8-like protein 1	1
Q9BSJ8	Extended synaptotagmin-1	1
P07332	Tyrosine-protein kinase Fes/Fps	1
P49589	Cysteine--tRNA ligase, cytoplasmic	1
Q7Z4H3	5'-deoxynucleotidase HDDC2	1
P06576	ATP synthase subunit beta, mitochondrial	1
P25705	ATP synthase subunit alpha, mitochondrial	1
P62316	Small nuclear ribonucleoprotein Sm D2	1
Q9Y262	Eukaryotic translation initiation factor 3 subunit L	1
G9CGD6	CNK3/IPCEF1 fusion protein	1
P42224	Signal transducer and activator of transcription 1-alpha/beta	1
Q9UI12	V-type proton ATPase subunit H	1
Q9NZT2	Opioid growth factor receptor	1
Q96QH2	PML-RARA-regulated adapter molecule 1	1
O95998	Interleukin-18-binding protein	1
O75494	Serine/arginine-rich splicing factor 10	1
P18124	60S ribosomal protein L7	1
P41250	Glycine--tRNA ligase	1

Elements only in CADASIL	Protein name	Peptide count
Q13451	Peptidyl-prolyl cis-trans isomerase FKBP5	1
Q9Y3I0	RNA-splicing ligase RtcB homolog	1
Q9Y295	Developmentally-regulated GTP-binding protein 1	1
O95967	EGF-containing fibulin-like extracellular matrix protein 2	1
Q7L2H7	Eukaryotic translation initiation factor 3 subunit M	1
Q8N4F0	BPI fold-containing family B member 2	1
Q5JRA6	Transport and Golgi organization protein 1 homolog	1
P15880	40S ribosomal protein S2	1
P19256	Lymphocyte function-associated antigen 3	1
P15586	N-acetylglucosamine-6-sulfatase	1
Q9NZL9	Methionine adenosyltransferase 2 subunit beta	1
P29466	Caspase-1	1
Q9Y266	Nuclear migration protein nudC	1
P47897	Glutamine--tRNA ligase	1
P50148	Guanine nucleotide-binding protein G(q) subunit alpha	1
P31153	S-adenosylmethionine synthase isoform type-2	1
Q9H3H3	UPF0696 protein C11orf68	1
Q16774	Guanylate kinase	1
O75663	TIP41-like protein	1
Q96H20	Vacuolar-sorting protein SNF8	1
Q7Z4V5	Hepatoma-derived growth factor-related protein 2	1
Q9Y6K9	NF-kappa-B essential modulator	1

Table 5. 112 proteins specific to the CADASIL patient group shown by Venn diagram

Elements only in Stroke	Protein name	Peptide count
Q9C0C9	(E3-independent) E2 ubiquitin-conjugating enzyme	1
Q86UL3	Glycerol-3-phosphate acyltransferase 4	1
P01903	HLA class II histocompatibility antigen, DR alpha chain	1
Q6UXF1	Transmembrane protein 108	1
Q6Q788	Apolipoprotein A-V	1
O43491	Band 4.1-like protein 2	2
Q96MK3	Pseudokinase FAM20A	1

Table 6. 7 proteins specific to the stroke patient group shown by Venn diagram

3.1.2. Downregulation of TCOF1 expression in CADASIL patients; however, the expression of SGTA increases.

Based on the previous results, as a result of finding the specific protein expression of the CADASIL patient group, Treacle protein (TCOF1) showed a marked decrease and Small glutamine-rich tetratricopeptide repeat-containing protein alpha (SGTA) a tendency to increase markedly. The relative protein abundance of TCOF1 was presented for each group. (Fig. 5). The x-axis means samples for each group, and the y-axis means relative protein abundance. The gray bar represents the relative protein abundance in each sample, and the blue bar represents the average in each group. The expression of TCOF1 was highest in the healthy control group, slightly weaker in stroke patient group than in the healthy control group and absent in the CADASIL patient group (Fig. 5).

ROC curves and interactive plots were shown of TCOF1 in CADASIL patient group (Fig. 6). The ROC curve showed a solid line indicating the sensitivity and 100-specificity values, while the Y-axis of the interactive plot represented the normalized concentration of TCOF1. The sensitivity and specificity at the cutoff concentration are displayed on the right side of the plot (Fig. 6). ROC curves and interactive plot of TCOF1 in CADASIL patient group versus two groups were shown (Fig. 6A). When comparing the CADASIL patient group with the rest of the groups, TCOF1 showed an AUC value of 0.913, a sensitivity of 82.5%, and a specificity of 100% (Fig. 6A). ROC curves and interactive plot of TCOF1 in CADASIL patient group versus stroke patient group were shown (Fig. 6B). When comparing the CADASIL patient group with the stroke patient group, TCOF1 showed an AUC value of 0.850, a sensitivity of 70.0%, and a specificity of 100% (Fig. 6B).

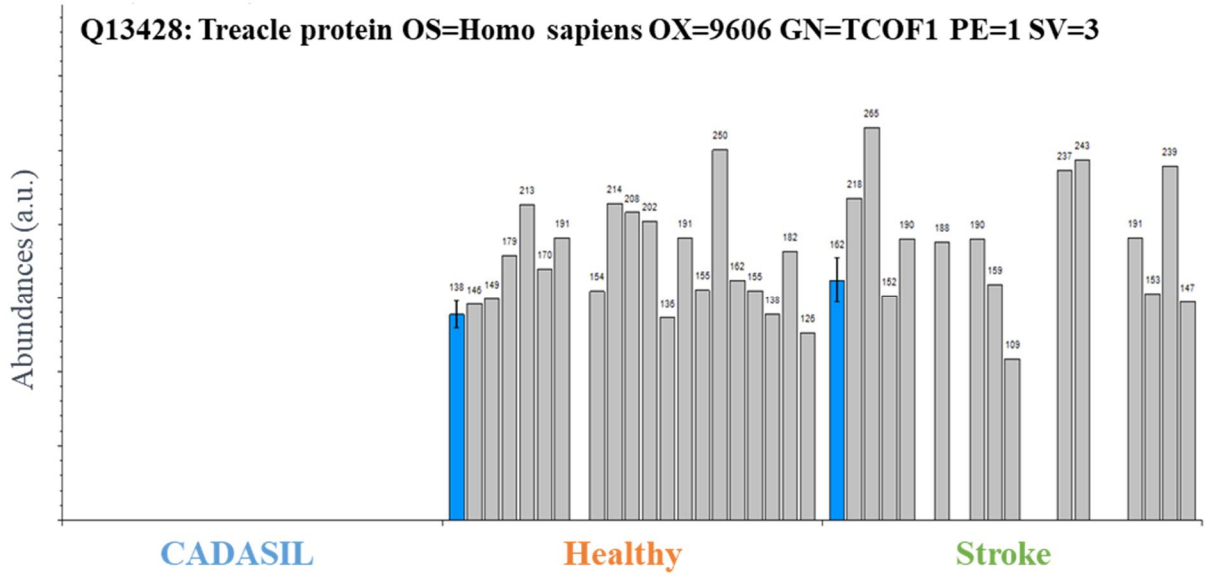


Figure 5. Bar graph showing the relative protein abundance of TCOF1 per sample

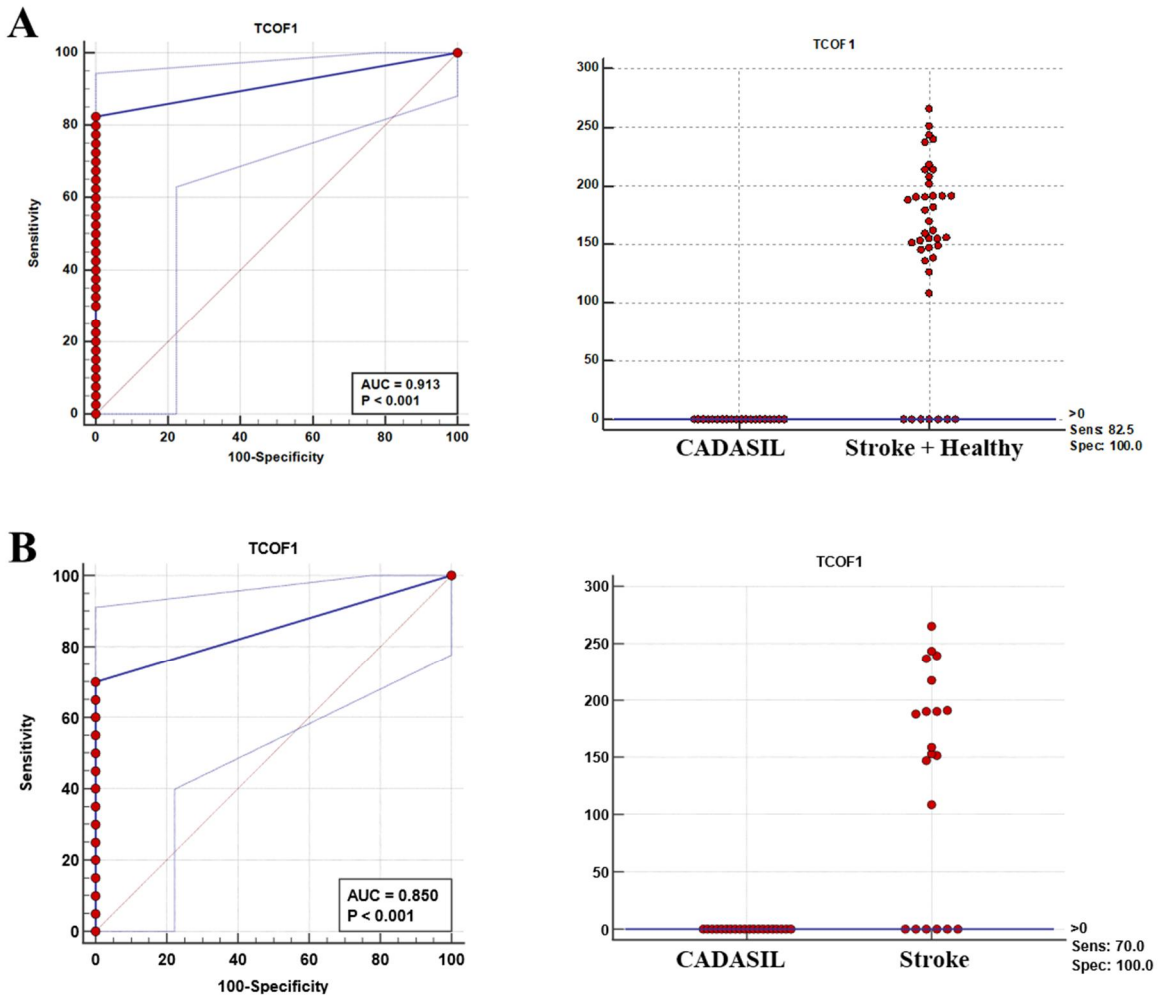


Figure 6. ROC curves and interactive plots of TCOF1 in CADASIL patient group (**A**) ROC curve and interactive plot of TCOF1 in CADASIL patient group versus two groups (**B**) ROC curve and interactive plot of TCOF1 in CADASIL patient group versus stroke patient group

The relative protein abundance of SGTA was presented for each group (Fig. 7). The expression of SGTA was highest in the CADASIL patient group (Fig. 7). ROC curves and interactive plots were shown of SGTA in CADASIL patient group (Fig. 8). The ROC curve showed a solid line indicating the sensitivity and 100-specificity values, while the Y-axis of the interactive plot represented the normalized concentration of SGTA. The sensitivity and specificity at the cutoff concentration are displayed on the right side of the plot (Fig. 8). ROC curves and interactive plot of SGTA in CADASIL patient group versus two groups were shown (Fig. 8A). When comparing the CADASIL patient group with the rest of the groups, SGTA showed an AUC value of 0.767, a sensitivity of 97.5%, and a specificity of 55% (Fig. 8A). ROC curves and interactive plot of SGTA in CADASIL patient group versus stroke patient group were shown (Fig. 8B). When comparing the CADASIL patient group with the stroke patient group, SGTA showed an AUC value of 0.759, a sensitivity of 95.0%, and a specificity of 55% (Fig. 8B).

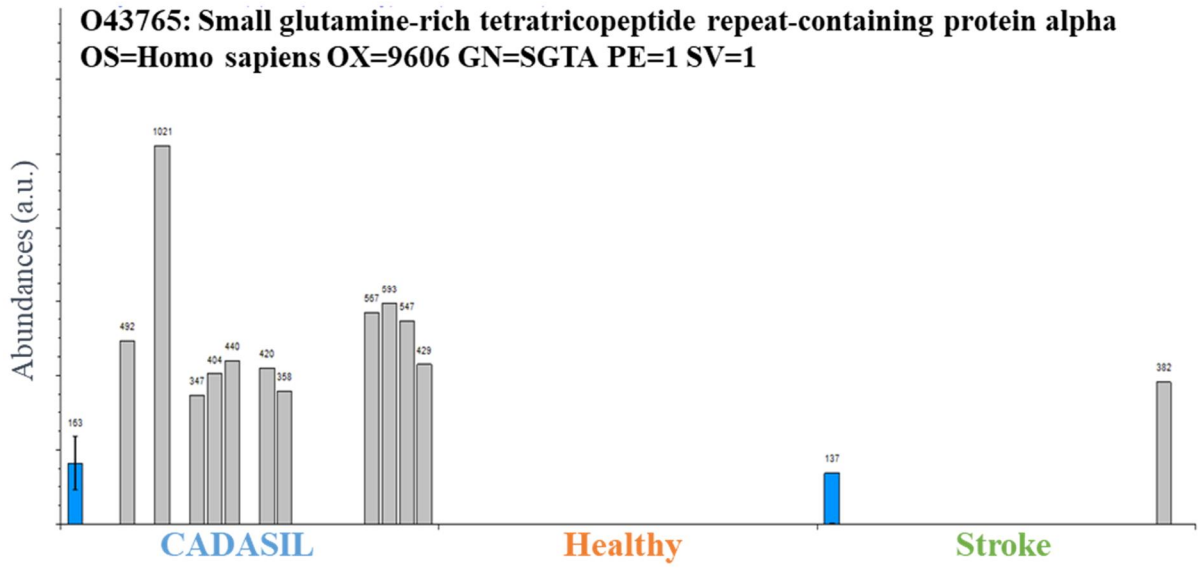


Figure 7. Bar graph showing the relative protein abundance of Small glutamine-rich tetratricopeptide repeat-containing protein alpha (SGTA) per sample

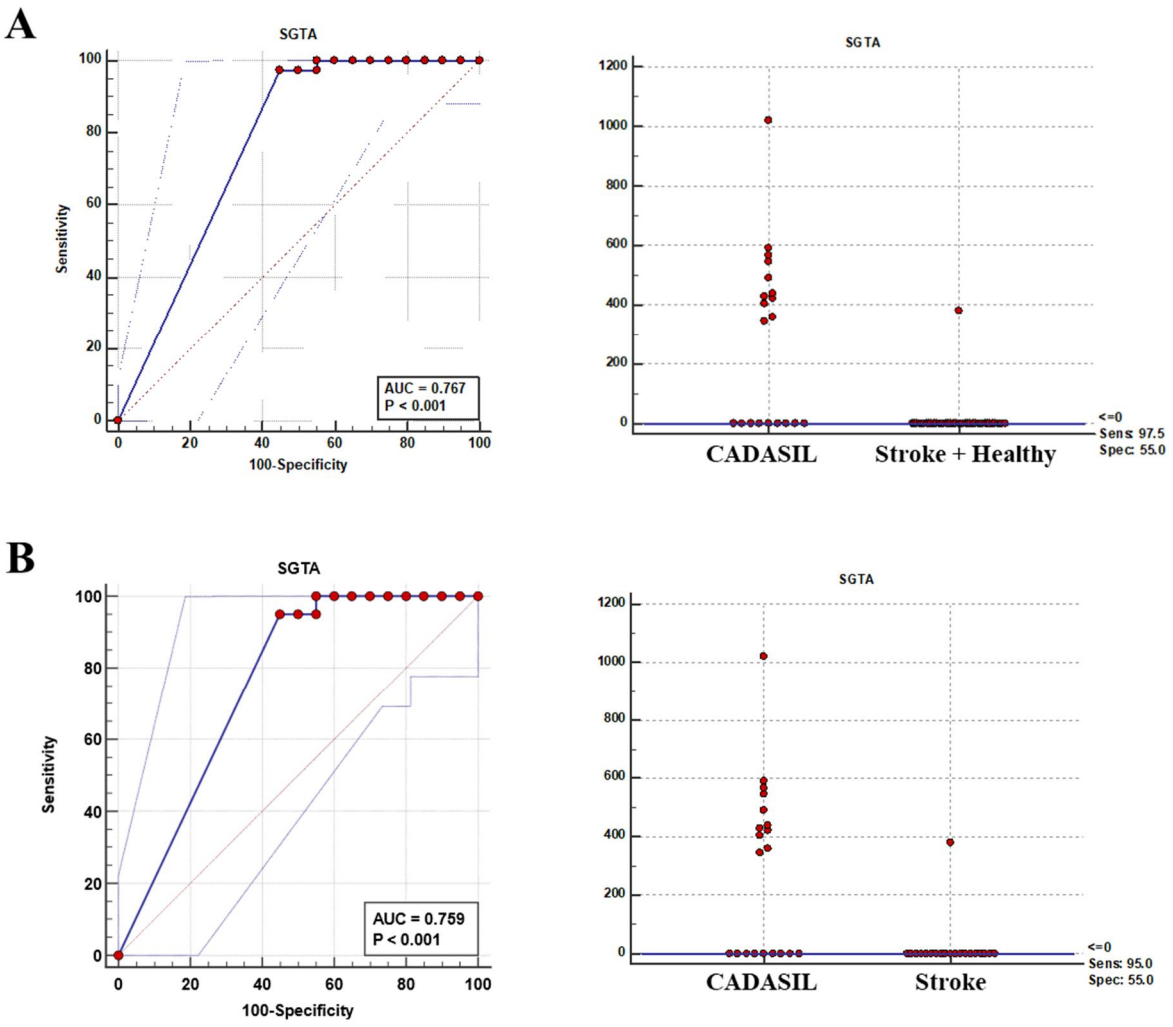
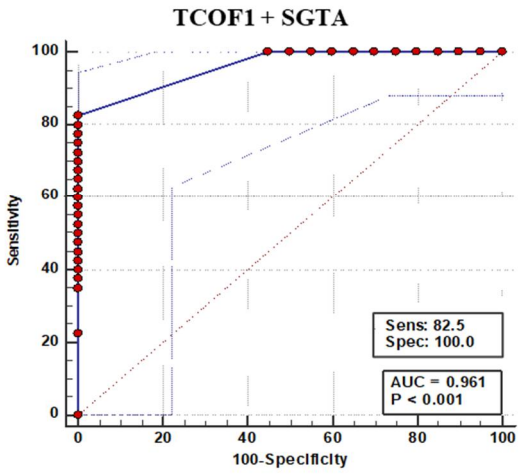


Figure 8. ROC curves and interactive plots of SGTA in CADASIL patient group (**A**) ROC curves and interactive plots of SGTA in CADASIL patient group versus two groups (**B**) ROC curves and interactive plots of SGTA in CADASIL patient group versus stroke patient group

ROC curves and interactive plots were shown of TCOF1 and SGTA in CADASIL patient group (Fig. 9). The ROC curve showed a solid line indicating the sensitivity and 100-specificity values, while the Y-axis of the interactive plot represented the normalized concentration of TCOF1 and SGTA. The sensitivity and specificity at the cutoff concentration are displayed on the right side of the plot (Fig. 9). ROC curves and interactive plot of TCOF1 and SGTA in CADASIL patient group versus two groups were shown (Fig. 9A). When comparing the CADASIL patient group with the rest of the groups, SGTA showed an AUC value of 0.961, a sensitivity of 82.5%, and a specificity of 100% (Fig. 9A). The ROC curve was generated using logistic regression to analyze the panel (Fig. 9B).

A



B

Overall Model Fit

Null model -2 Log Likelihood	76.382
Full model -2 Log Likelihood	21.930
Chi-squared	54.452
DF	2
Significance level	P < 0.0001
Cox & Snell R^2	0.5965
Nagelkerke R^2	0.8284

Coefficients and Standard Errors

Variable	Coefficient	Std. Error	Wald	P
SGTA	-0.058277	35.38026	0.000002713	0.9987
TCOF1	0.18464	85.25683	0.000004690	0.9983
Constant	-0.25131	0.50395	0.2487	0.6180

Classification table (cut-off value p=0.5)

Actual group	Predicted group		Percent correct
	0	1	
Y = 0	20	0	100.00%
Y = 1	7	33	82.50%
Percent of cases correctly classified			88.33%

Figure 9. ROC curves and interactive plots of TCOF1 and SGTA in CADASIL patient group (**A**) ROC curves and interactive plots of TCOF1 and SGTA in CADASIL patient group versus two groups (**B**) Logistic regression of TCOF1 and SGTA

Mutations in TCOF1 are known to cause Treacher Collins syndrome [37]. In this study, TCOF1 was identified specifically in relation to CADASIL. However, while TCOF1 is known to influence cranial development [37], its impact on cerebral vasculature has been scarcely studied. Therefore, Biological General Repository for Interaction Datasets (BioGRID) analysis was performed to investigate protein-protein interactions, genetic interactions, chemical interactions, and post-translational modifications (Fig. 10). From the overall (Fig. 10A) and zoomed-in (Fig. 10B) BioGRID analysis of TCOF1, the yellow lines indicated associations with physical evidence. The green lines indicated associations with genetic evidence (Fig. 10C), and the purple lines indicated associations with both genetic and physical evidence (Fig. 10B).

The RNA and protein expression of TCOF1 was shown through The Human Protein Atlas (Fig. 11). The bar on the left represents RNA expression, and the bar on the right represents protein expression (Fig. 11A). Both RNA and protein are most highly expressed in the brain (Fig. 11A). Among the brains that occupied the most ratio, the cerebral cortex occupied an overwhelmingly high proportion, followed by white matter (Fig. 11B).

SGTA has been shown to facilitate the biogenesis and quality control of hydrophobic proteins, prevent protein misfolding and improper folding, and transport proteins to desired secure locations [38]. However, no specific research has been conducted on the association between SGTA and CADASIL. To investigate the mechanism, mechanistic studies were conducted by confirming the interaction of SGTA through BioGRID, as done for TCOF1 (Fig. 12). From the overall (Fig. 12A) and zoomed-in (Fig. 12B) BioGRID analysis of SGTA, the confirmed interacting proteins included collagen, which was previously mentioned as a candidate biomarker. Collagen was represented as a circle in the figure (Fig. 12B).

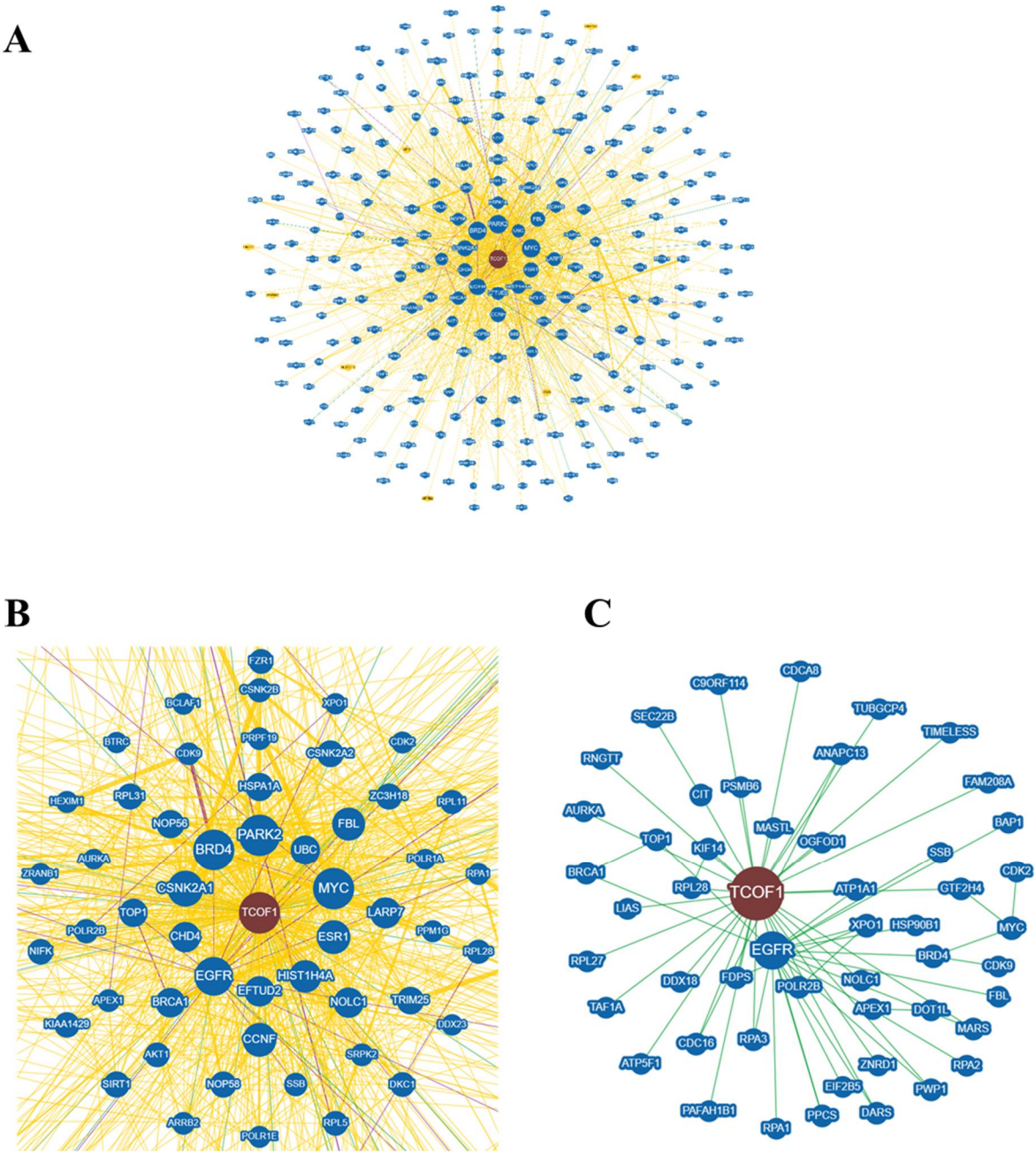


Figure 10. The biological general repository for interaction datasets (BioGRID) analysis suggests potential substrates of TCOF1 (https://thebiogrid.org/ (accessed on 19 April 2023)) (**A**) The overall BioGRID analysis of TCOF1 (**B**) A zoomed-in figure of the entire BioGRID (**C**) Only the green lines representing association with genetic evidence are depicted in the entire BioGRID

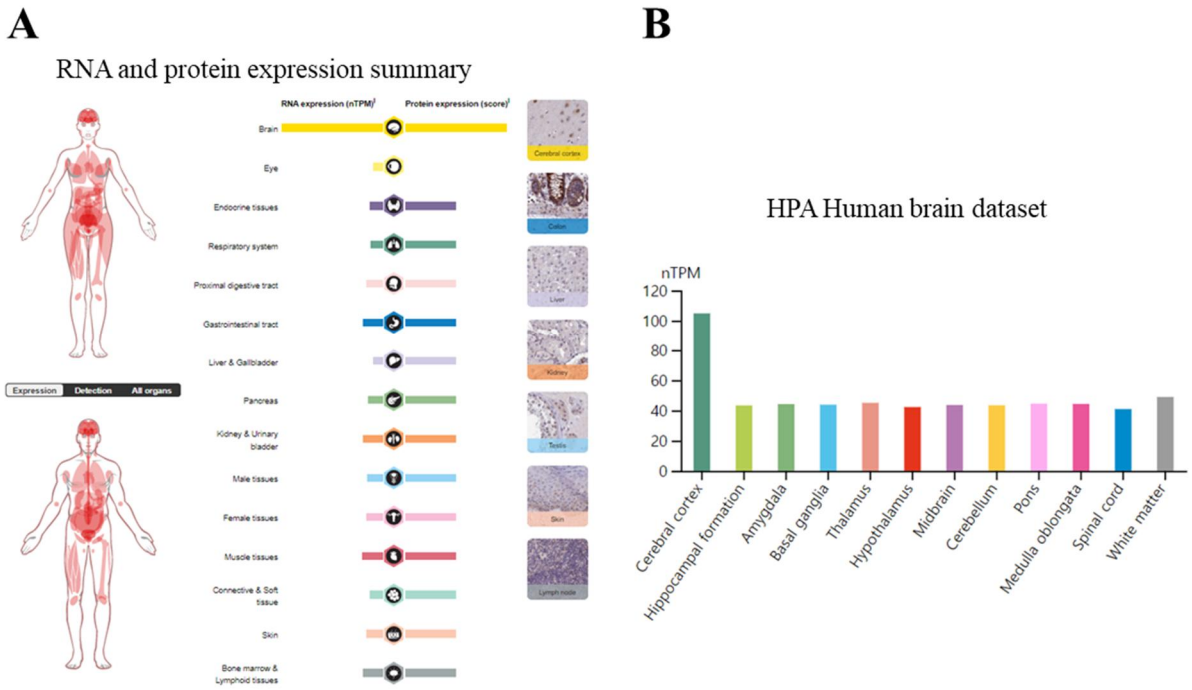
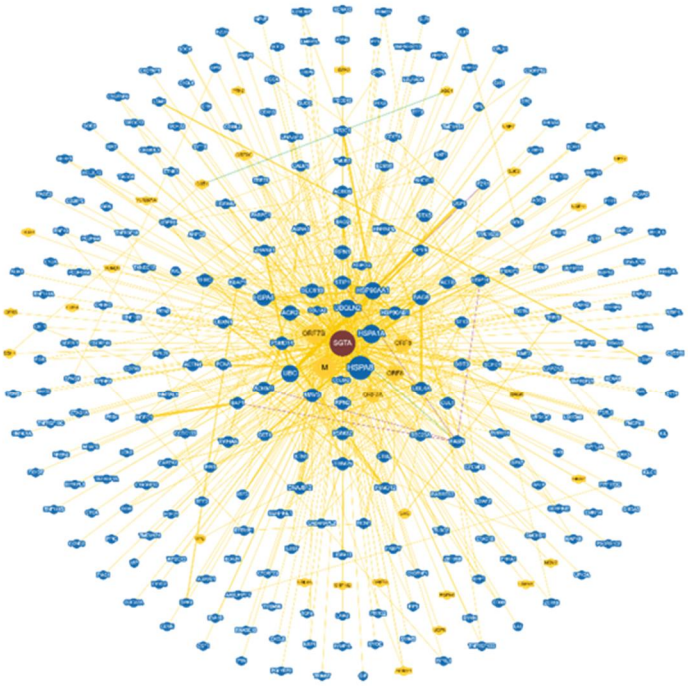


Figure 11. RNA and protein expression of TCOF1 (https://www.proteinatlas.org/ (accessed on 19 April 2023)) (**A**) RNA and protein expression summary of TCOF1. The bar on the left represents RNA expression, and the bar on the right represents protein expression. Both RNA and protein are most highly expressed in the brain (**B**) Brain RNA expression of TCOF1

A



B

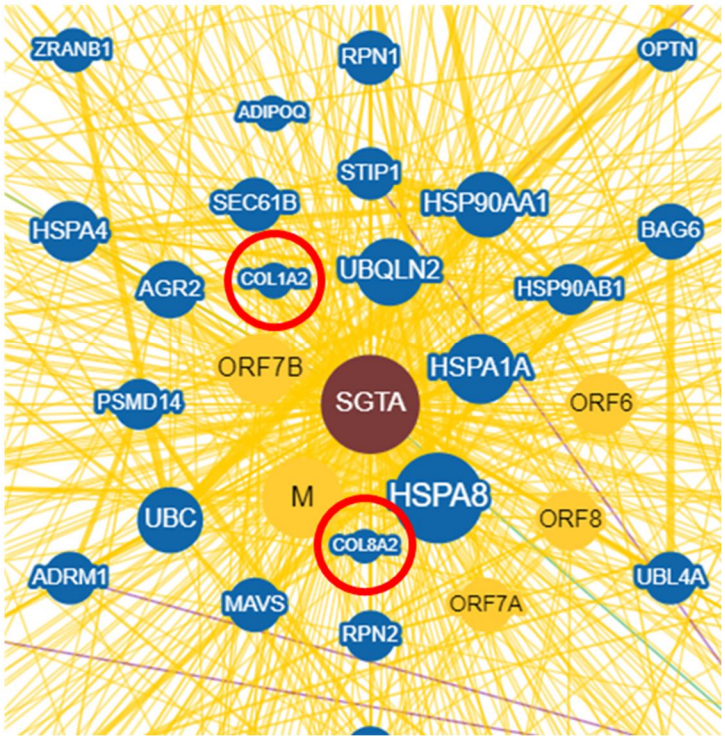


Figure 12. The biological general repository for interaction datasets (BioGRID) analysis suggests potential substrates of SGTA (https://thebiogrid.org/ (accessed on 19 April 2023)) (**A**) The overall BioGRID analysis of SGTA (**B**) A zoomed-in figure of the entire BioGRID

3.2. The CADASIL patient group had distinct differences from other groups.

In addition to the above-mentioned proteins, specific proteins were expressed in the CADASIL patient group and in the other two groups (Fig. 13). Unconventional myosin-XVIIIa (MYO18A), RNA-binding protein EWS (EWSR1), Sorcin (SRI), Isoform 1 of Transportin-3 (TNPO3) were specifically expressed in the CADASIL patient group (Fig. 13A). Iduronate 2-sulfatase (IDS), Putative HLA class I histocompatibility antigen, alpha chain H (HLA-H) were specifically expressed in the other two groups (Fig. 13B).

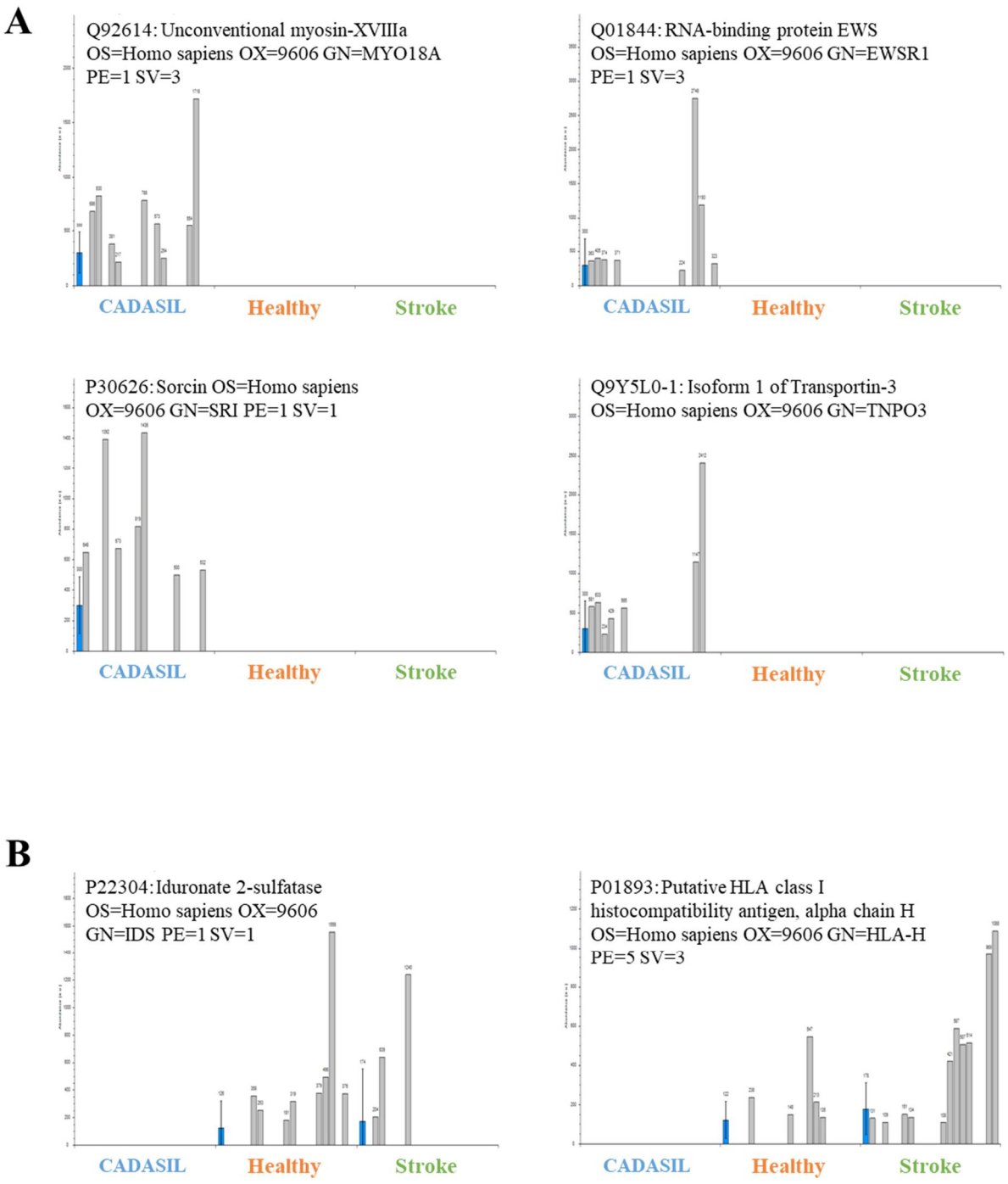


Figure 13. Besides TCOF1 and SGTA, some proteins were identified in the CADASIL patient group, while others were only expressed in other groups (**A**) Proteins that showed increased expression in the CADASIL patient group (**B**) Proteins that showed decreased expression in the CADASIL patient group

Analysis Gene ontology (GO) analysis was performed on CADASIL patient group and healthy control group (Fig. 14). Similarly, the stroke patient group and the healthy control group were analyzed by GO analysis (Fig. 14). GO levels with a multiple of 1.5 or more based on the healthy control group were shown (Fig. 14). GO biological process analysis provides insights into the biological processes or activities in which genes participate (Fig. 14A). When the GO biological process analysis of proteins showing a fold change of 1.5 or greater between the CADASIL and stroke patient groups was compared with the healthy control group, both groups were predominantly associated with genes related to neutrophil function (neutrophil activation involved in immune response, neutrophil degranulation, neutrophil mediated immunity) (Fig. 14A). However, in the case of CADASIL, genes involved in mRNA splicing were more prominent, while in the case of stroke, negative regulation of receptor-mediated endocytosis and response to glucose and transmembrane transport were more prominent (Fig. 14A). GO molecular function analysis provides insights into the molecular-level functions performed by genes (Fig. 14B). When the GO molecular function analysis of proteins showing a fold change of 1.5 or greater between the CADASIL and stroke patient groups was compared with the healthy control group, a notable difference emerged (Fig. 14B). CADASIL was characterized by a predominant expression of proteins associated with RNA binding, while stroke exhibited a higher representation of proteins involved in kinase binding (Fig. 14B). GO cellular component analysis provides insights into the cellular structures or locations where genes are present (Fig. 14C). When the GO cellular component analysis of proteins showing a fold change of 1.5 or greater between the CADASIL and stroke patient groups was compared with the healthy control group, it was observed that in the case of CADASIL, proteins were predominantly located in the secretory granule lumen, while in the case of stroke, proteins were primarily located in the ficolin 1-rich granule (Fig. 14C).

A

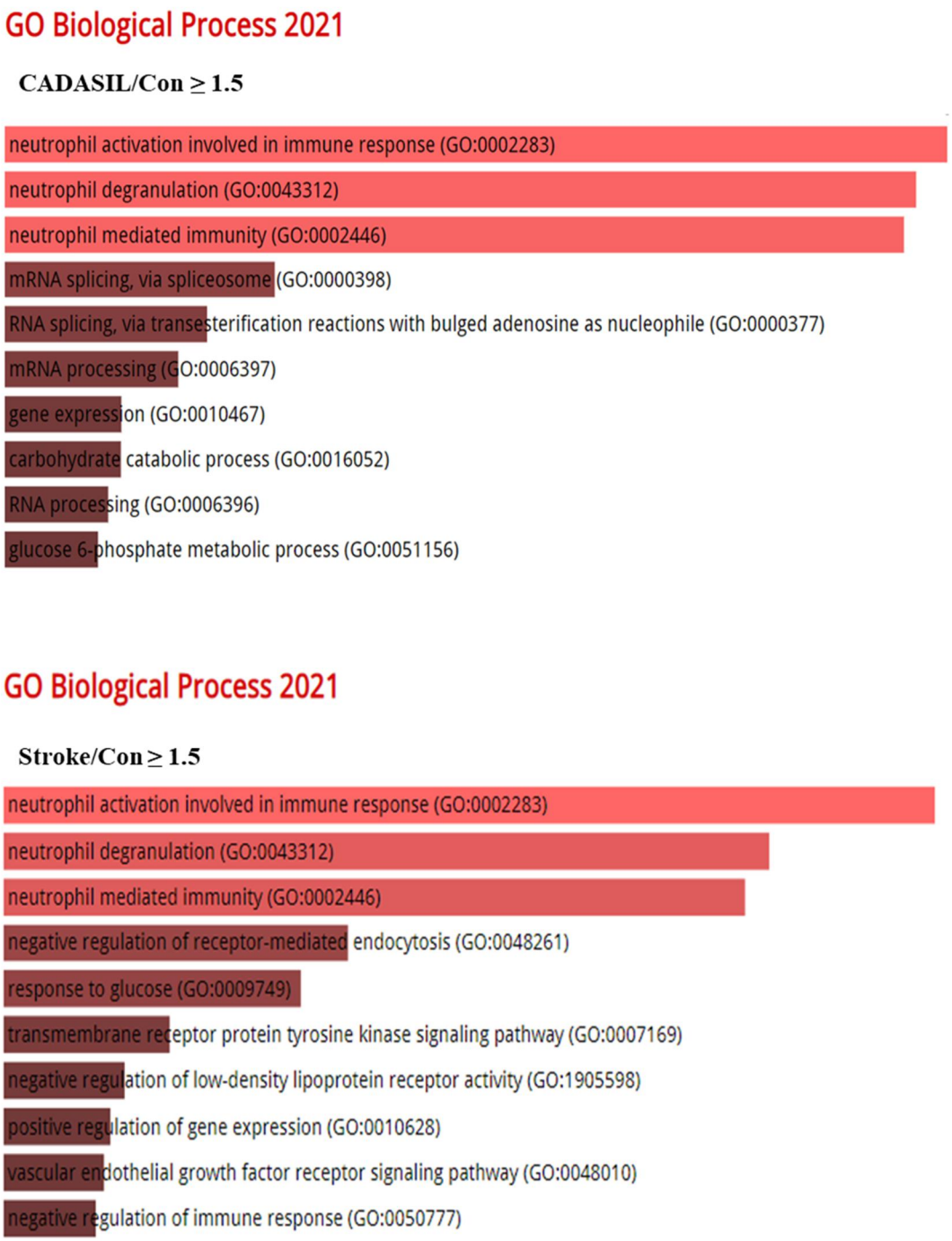


Figure 14. (continue)

B

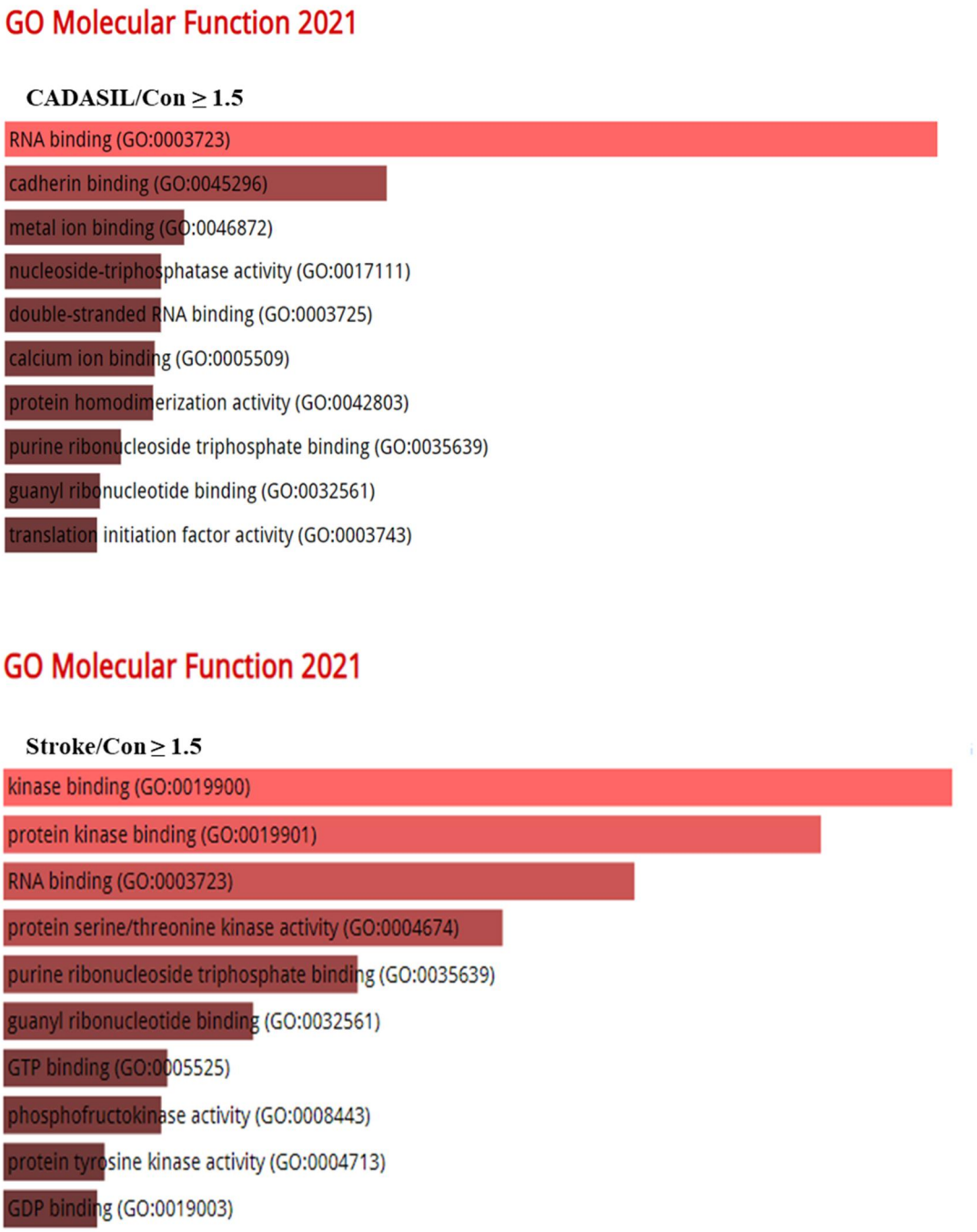


Figure 14. (continue)

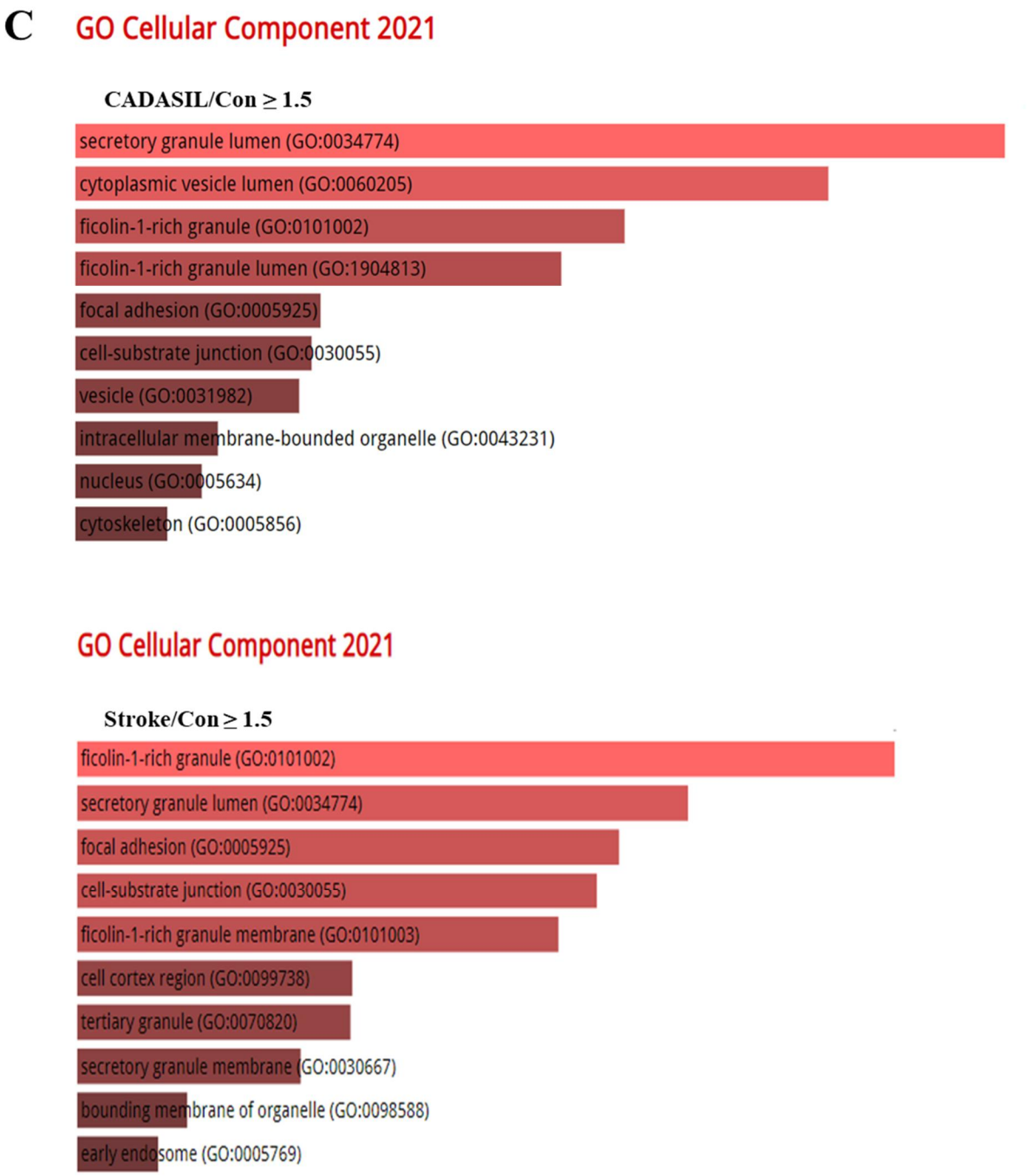


Figure 14. Gene ontology (GO) analysis of the CADASIL patient group and stroke patient group The CADASIL patient group and the stroke patient group were compared with the healthy control group, respectively (**A**) GO biological process analysis of proteins showing a fold change of 1.5 or greater between the CADASIL patient group and the stroke patient group, compared to the healthy control group (**B**) GO molecular function analysis of proteins showing a fold change of 1.5 or greater between the CADASIL patient group and the stroke patient group, compared to the healthy control group (**C**) GO cellular component analysis of proteins showing a fold change of 1.5 or greater between the CADASIL patient group and the stroke patient group, compared to the healthy control group

3.3. Discussion

Figure 15 depicts the comparison of the biomarkers studied thus far, as mentioned in Table 1, with the results obtained in this study. The graph includes not only the given biomarkers but also their related markers. Data showing significant changes with p-values below 0.05 were marked with boxes. Additionally, potential biomarkers exhibiting specific changes only in the CADASIL patient group were marked with both boxes and arrows. As a result, we observed markers, such as Notch 3, that exhibited changes consistent with the previously reported potential biomarkers. However, related markers showed contrasting changes. For instance, collagen-related markers demonstrated varying patterns of change. (Fig. 15). As a result, a total of 28 related proteins were quantified (HtrA Serine Peptidase 1 (HTRA1), Vitronectin (VTN), Serum amyloid P-component, Microfibril-associated glycoprotein 4, Annexin A2, Isoform 2 of Periostin, Isoform 5 of Periostin, Metalloproteinase inhibitor 1, Metalloproteinase inhibitor 2, Neurogenic locus notch homolog protein 3, Collagen alpha-1(XII) chain, Procollagen C-endopeptidase enhancer 1, Collagen alpha-1(V) chain, Collagen alpha-1(XV) chain, Collagen alpha-1(VI) chain, 72 kDa type IV collagenase, Collagen alpha-2(I) chain, Collagen alpha-1(XVIII) chain, Collagen alpha-1(I) chain, Neutrophil collagenase, Collagen alpha-2(VI) chain, Collagen alpha-3(VI) chain, Collagen alpha-1(III) chain, Laminin subunit beta-2, Isoform 1 of Laminin subunit alpha-3, Laminin subunit gamma-1, Laminin subunit beta-1, Laminin subunit alpha-2 (Fig. 15).

The data with significant values are displayed on the graph with their corresponding *p*-values ($p < 0.05$) (Fig. 16). Biomarkers that showed significant expression differences were HTRA1, Isoform 5 of Periostin, Metalloproteinase inhibitor 1, Neurogenic locus notch homolog protein 3, Procollagen C-endopeptidase enhancer 1, Collagen alpha-1(V) chain, Collagen alpha-1(VI) chain, Collagen alpha-1(XVIII) chain, Collagen alpha-1(I) chain, Laminin subunit beta-2, Isoform 1 of Laminin subunit alpha-3, Laminin subunit gamma-1, and Laminin subunit alpha-2, totaling 13. Other 15 proteins, VTN, Serum amyloid P-component, Microfibril-associated glycoprotein 4, Annexin A2, Isoform 2 of Periostin, Metalloproteinase inhibitor 2, Collagen alpha-1(XII) chain, Collagen alpha-1(XV) chain, 72 kDa type IV collagenase, Collagen alpha-2(I) chain, Neutrophil collagenase, Collagen alpha-2(VI) chain, Collagen alpha-3(VI) chain, Collagen alpha-1(III) chain, Laminin subunit beta-1 showed no significant

difference (Fig. 16).

HTRA1 showed a significant difference between CADASIL patient group and healthy control group ($p = 0.007$) (Fig. 16A). Isoform 5 of Periostin showed significant expression differences between healthy control group and CADASIL patient group ($p = 0.002$), and between healthy control group and stroke patient group ($p = 0.04$) (Fig. 16G). Metalloproteinase inhibitor showed significant expression differences between the healthy control group and the CADASIL patient group ($p = 0.005$), and between the CADASIL patient group and the stroke patient group ($p = 0.001$) (Fig. 16H). VTN, Serum amyloid P-component, Microfibril-associated glycoprotein 4, Annexin A2, and Isoform 2 of Periostin showed no significant difference in expression between groups.

Neurogenic locus notch homolog protein 3 showed significant expression differences in healthy control group and CADASIL patient group ($p = 0.03$), CADASIL patient group and stroke patient group ($p = 0.001$), and healthy control group and stroke patient group ($p = 0.03$) (Fig. 16J). Procollagen C-endopeptidase enhancer 1 showed significant expression differences between the healthy control group and the CADASIL patient group ($p = 0.03$) (Fig. 16L). Collagen alpha-1(V) chain showed a significant difference in expression between the CADASIL patient group and the stroke patient group ($p = 0.02$) (Fig. 16M). Collagen alpha-1(VI) showed significant expression differences between the healthy control group and the CADASIL patient group ($p = 0.03$) (Fig. 16O). Metalloproteinase inhibitor 2, collagen alpha-1(XII) chain, collagen alpha-1(XV) chain, 72 kDa type IV collagenase, showed no significant difference in expression.

Collagen alpha-1(XVIII) chain showed a significant difference in expression between the healthy control group and the stroke patient group ($p = 0.009$) (Fig. 16R). Collagen alpha-1(I) chain showed a significant difference in expression between the healthy control group and the CADASIL patient group ($p = 0.04$), and between the healthy control group and the stroke patient group ($p = 0.02$) (Fig. 16S). Laminin subunit beta-2 showed significant expression differences between the healthy control group and the CADASIL patient group ($p = 0.0008$), and between the CADASIL patient group and the stroke patient group ($p = 0.000002$) (Fig. 16X). Collagen alpha-2(I) chain, Neutrophil collagenase, Collagen alpha-2(VI) chain, Collagen alpha-3(VI) chain, Collagen alpha-1(III) chain showed no significant difference.

Isoform 1 of Laminin subunit alpha-3 showed significant expression differences between the healthy control group and the CADASIL patient group (p = 0.006), and between the CADASIL patient group and the stroke patient group (p = 0.0007) (Fig. 16Y). Laminin subunit gamma-1 showed significant expression differences between healthy control group and stroke patient group (p = 0.01), CADASIL patient group and stroke patient group (p = 0.0001) (Fig. 16Z). Laminin subunit alpha-2 showed significant expression differences between the healthy control group and the CADASIL patient group (p = 0.006), and between the healthy control group and the stroke patient group (p = 0.02) (Fig. 16AB). Laminin subunit beta-1 showed no significant difference.

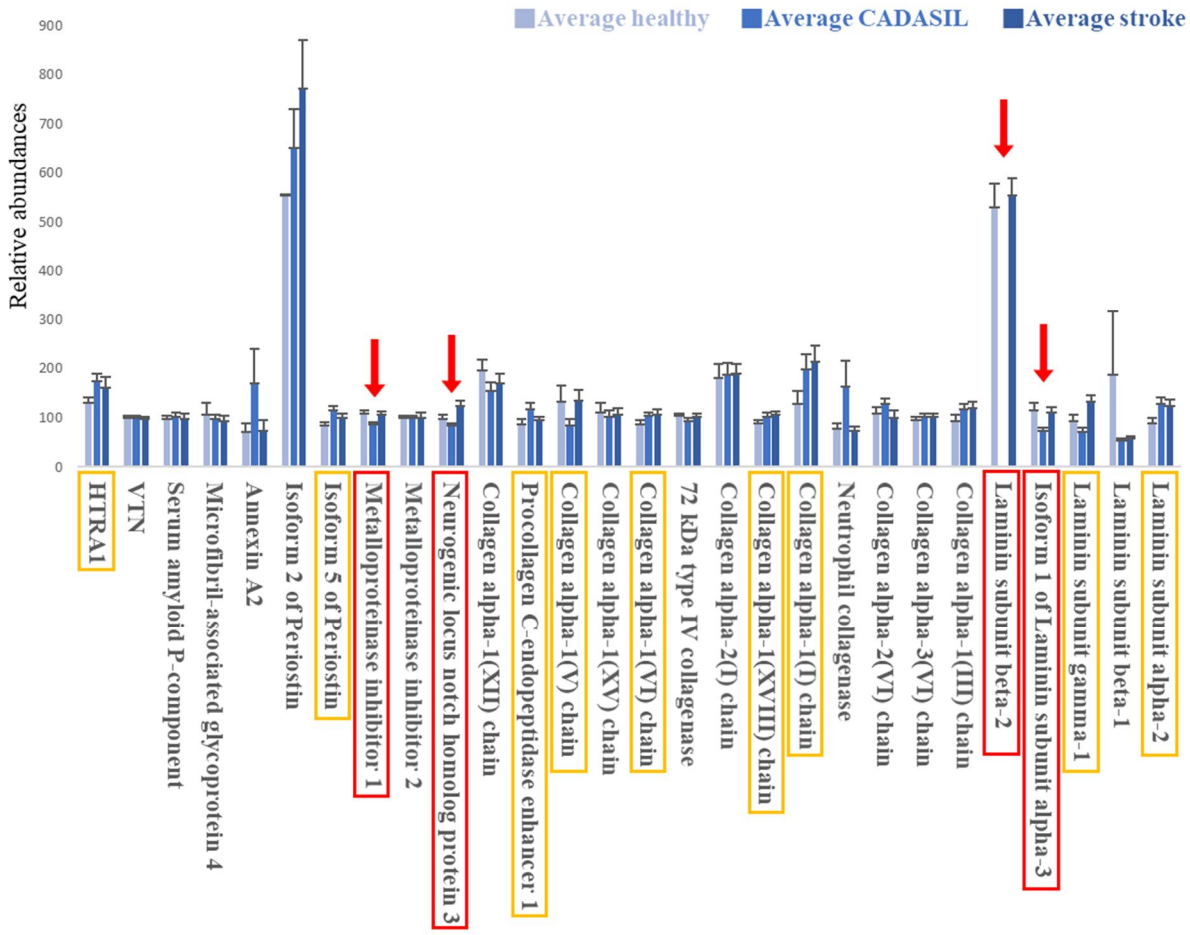


Figure 15. Potential biomarkers detected in the blood of CADASIL mentioned in Table 1 and their related proteins were compared with the results of this study

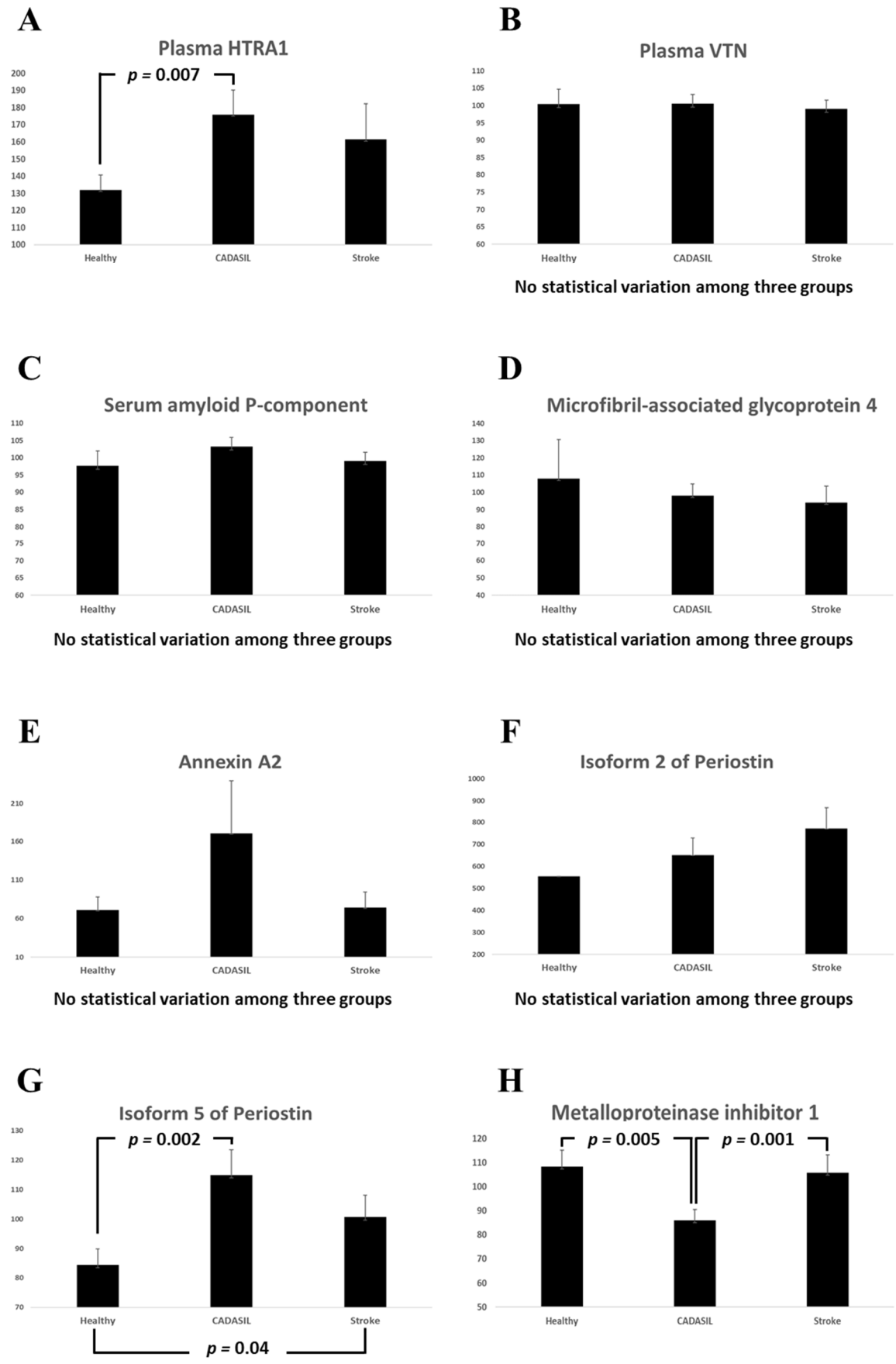


Figure 16. (continue)

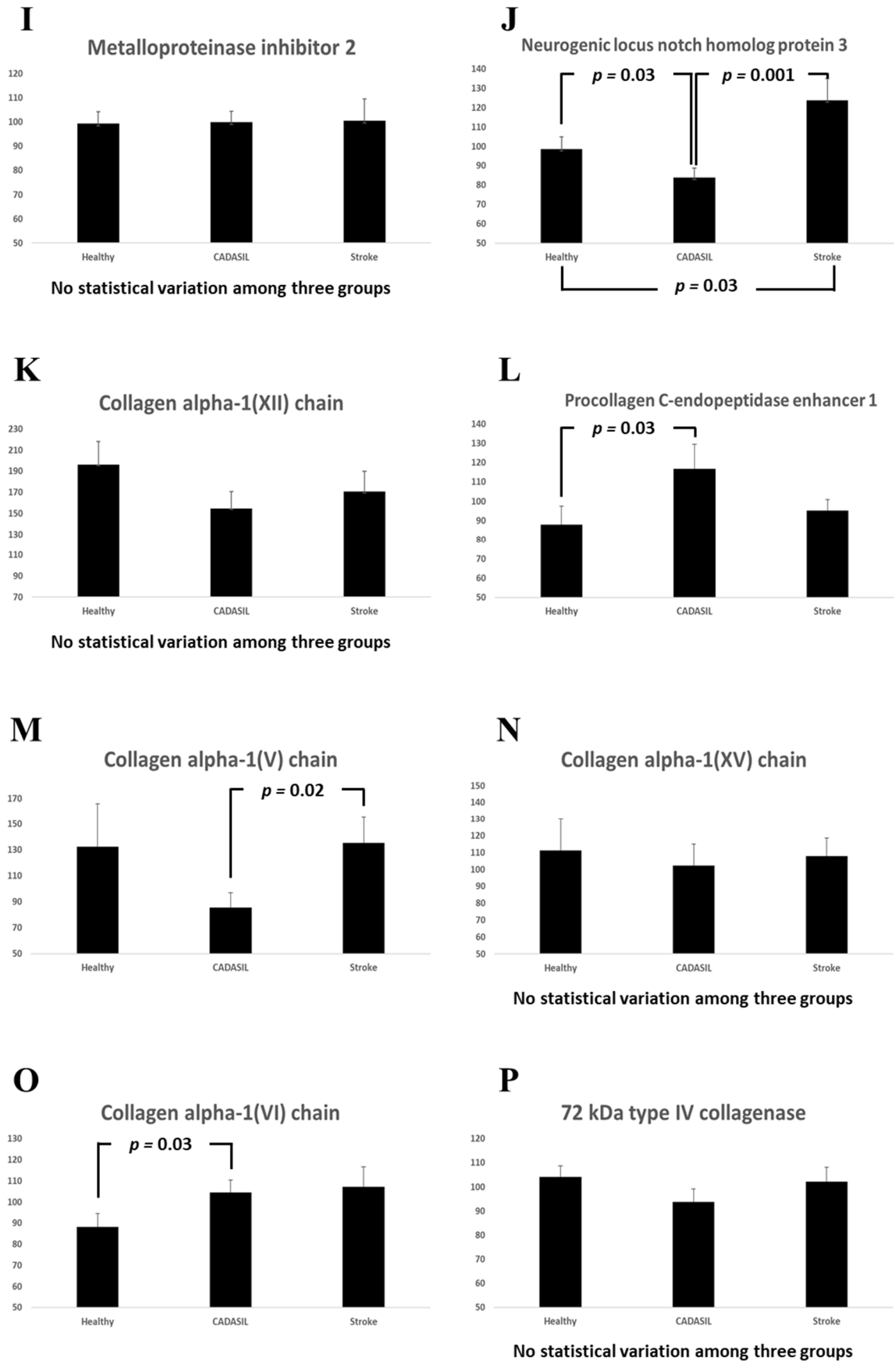


Figure 16. (continue)

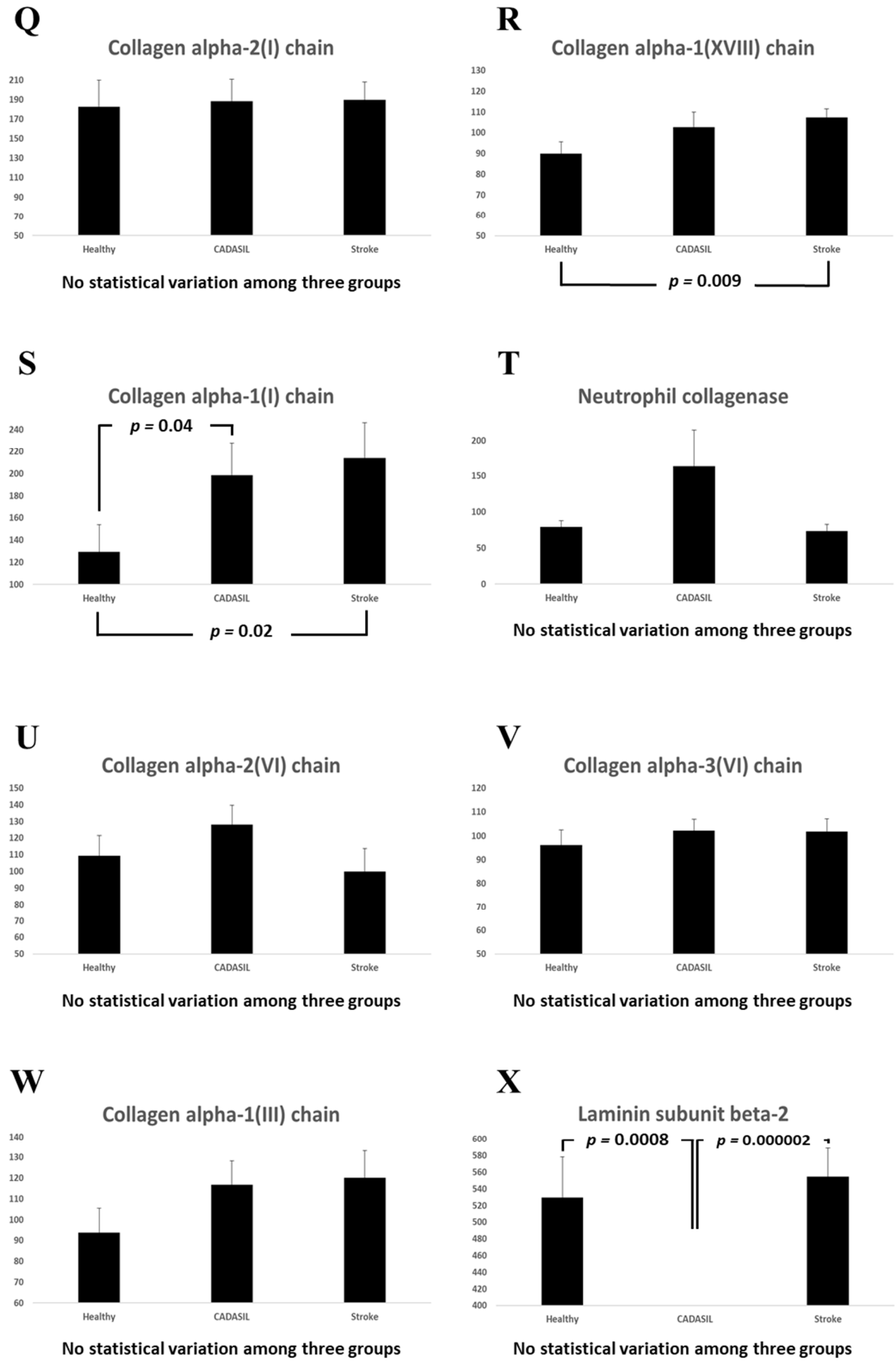


Figure 16. (continue)

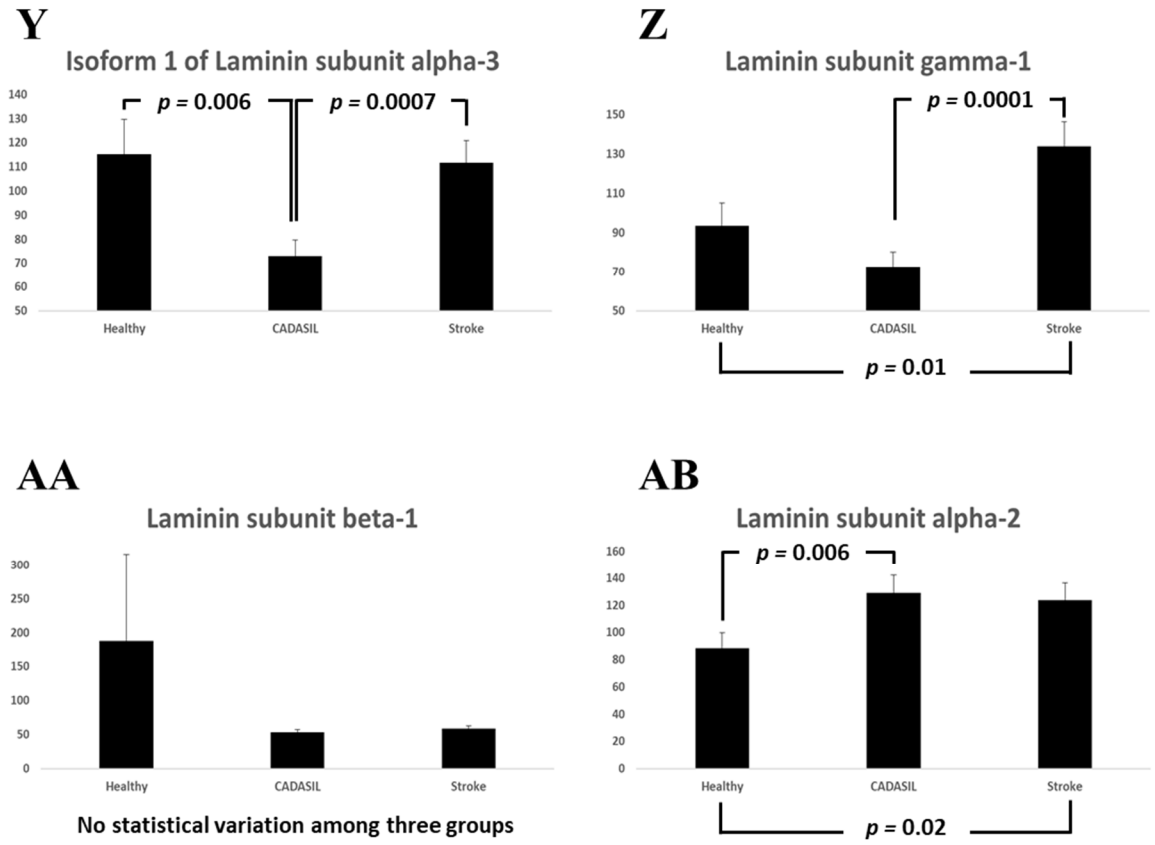


Figure 16. Based on the results in Figure 14, each group's proteins showing significant differences are shown (**A**) Comparison of HtrA Serine Peptidase 1 (HTRA1) expression levels in the three groups (**B**) Comparison of Vitronectin (VTN) expression levels in the three groups (**C**) Comparison of Serum amyloid P-component expression levels in the three groups (**D**) Comparison of Microfibril-associated glycoprotein 4 expression levels in the three groups (**E**) Comparison of Annexin A2 expression levels in the three groups (**F**) Comparison of Isoform 2 of Periostin expression levels in the three groups (**G**) Comparison of Isoform 5 of Periostin expression levels in the three groups (**H**) Comparison of Metalloproteinase inhibitor 1 expression levels in the three groups (**I**) Comparison of Metalloproteinase inhibitor 2 expression levels in the three groups (**J**) Comparison of Neurogenic locus notch homolog protein 3 expression levels in the three groups (**K**) Comparison of Collagen alpha-1(XII) chain expression levels in the three groups (**L**) Comparison of Procollagen C-endopeptidase enhancer 1 expression levels in the three groups (**M**) Comparison of Collagen alpha-1(V) chain expression levels in the three groups (**N**) Comparison of Collagen alpha-1(XV) chain expression levels in the three groups (**O**) Comparison of Collagen alpha-1(VI) chain expression levels in the three groups (**P**) Comparison of 72 kDa type IV collagenase expression levels in the three groups (**Q**) Comparison of Collagen alpha-2(I) chain expression levels in the three groups (**R**) Comparison of Collagen alpha-1(XVIII) chain expression levels in the three groups (**S**) Comparison of Collagen alpha-1(I) chain expression levels in the three groups (**T**) Comparison of Neutrophil collagenase expression levels in the three groups (**U**) Comparison of Collagen alpha-2(VI) chain expression levels in the three groups (**V**) Comparison of Collagen alpha-3(VI) chain expression levels in the three groups (**W**) Comparison of Collagen alpha-1(III) chain expression levels in the three groups (**X**) Comparison of Laminin subunit beta-2 expression levels in the three groups (**Y**) Comparison of Isoform 1 of Laminin subunit alpha-3 expression levels in the three groups (**Z**) Comparison of Laminin subunit gamma-1 expression levels in the three groups (**AA**) Comparison of Laminin subunit beta-1 expression levels in the three groups (**AB**) Comparison of Laminin subunit alpha-2 expression levels in the three groups

In this study, a novel biomarker for CADASIL was identified, TCOF1, and its potential mechanism was proposed. This made the study significant as it presented a biomarker that had not been previously mentioned in other studies.

One of the major strengths of this study was that it analyzed three groups (CADASIL patient group, stroke patient group, healthy control group) of human plasma proteomes to identify candidate biomarkers. This provided a comprehensive analysis of the proteins present in the plasma and increased the chances of identifying relevant biomarkers. Additionally, the proposed pathway would provide insight into the potential mechanism of the disease.

However, the study used a small sample size of 30, and further validation was necessary to confirm the statistical significance of the findings. A validation study on a larger sample size of 60 will be conducted to address this limitation.

The development of a diagnostic kit based on the biomarker could improve the convenience and accuracy of diagnosing CADASIL. This would have significant implications for the quality of life of patients, as early diagnosis and monitoring of the disease could lead to more effective treatment options.

Overall, this study made a valuable contribution to the field of CADASIL research by identifying a novel biomarker and proposing a potential pathway for the disease. With further validation and development, this biomarker has the potential to improve the diagnosis and treatment of CADASIL, ultimately improving the quality of life for affected individuals.

References

1. Kang, C.H., et al., *Pathogenic NOTCH3 Variants Are Frequent Among the Korean General Population.* Neurol Genet, 2021. **7**(6): p. e639.

2. Muino, E., I. Fernandez-Cadenas, and A. Arboix, *Contribution of "Omic" Studies to the Understanding of Cadasil. A Systematic Review.* Int J Mol Sci, 2021. **22**(14).

3. Herve, D. and H. Chabriat, *Cadasil.* J Geriatr Psychiatry Neurol, 2010. **23**(4): p. 269-76.

4. Chabriat, H., et al., *Cadasil.* Lancet Neurol, 2009. **8**(7): p. 643-53.

5. Chabriat, H., et al., *CADASIL: yesterday, today, tomorrow.* Eur J Neurol, 2020. **27**(8): p. 1588-1595.

6. Joutel, A., et al., *Notch3 mutations in CADASIL, a hereditary adult-onset condition causing stroke and dementia.* Nature, 1996. **383**(6602): p. 707-10.

7. Primo, V., et al., *Blood biomarkers in a mouse model of CADASIL.* Brain Res, 2016. **1644**: p. 118-26.

8. Ferrante, E.A., C.D. Cudrici, and M. Boehm, *CADASIL: new advances in basic science and clinical perspectives.* Curr Opin Hematol, 2019. **26**(3): p. 193-198.

9. Choi, J.C., J.S. Lee, and K. Kim, *Importance of CADASIL research in Jeju: a review and update on epidemiology, diagnosis, and clinical spectrum.* Journal of Medicine and Life Science, 2020. **17**(3): p. 65-73.

10. Cho, B.P.H., et al., *NOTCH3 variants are more common than expected in the general population and associated with stroke and vascular dementia: an analysis of 200 000 participants.* J Neurol Neurosurg Psychiatry, 2021. **92**(7): p. 694-701.

11. Rutten, J.W., et al., *Interpretation of NOTCH3 mutations in the diagnosis of CADASIL.* Expert

Rev Mol Diagn, 2014. **14**(5): p. 593-603.

12. Federico, A., S. Bianchi, and M.T. Dotti, *The spectrum of mutations for CADASIL diagnosis.* Neurol Sci, 2005. **26**(2): p. 117-24.

13. Peters, N., et al., *Spectrum of mutations in biopsy-proven CADASIL: implications for diagnostic strategies.* Arch Neurol, 2005. **62**(7): p. 1091-4.

14. Hu, Y., et al., *NOTCH3 Variants and Genotype-Phenotype Features in Chinese CADASIL Patients.* Front Genet, 2021. **12**: p. 705284.

15. Kitteringham, N.R., et al., *Multiple reaction monitoring for quantitative biomarker analysis in proteomics and metabolomics.* J Chromatogr B Analyt Technol Biomed Life Sci, 2009. **877**(13): p. 1229-39.

16. Jacob, M., et al., *Metabolomics toward personalized medicine.* Mass Spectrom Rev, 2019. **38**(3): p. 221-238.

17. Aronson, J.K., *Biomarkers and surrogate endpoints.* Br J Clin Pharmacol, 2005. **59**(5): p. 491-4.

18. Aronson, J.K. and R.E. Ferner, *Biomarkers-A General Review.* Curr Protoc Pharmacol, 2017. **76**: p. 9 23 1-9 23 17.

19. Glaab, E., et al., *Biomarker discovery studies for patient stratification using machine learning analysis of omics data: a scoping review.* BMJ Open, 2021. **11**(12): p. e053674.

20. Cho, W.C., *Proteomics technologies and challenges.* Genomics Proteomics Bioinformatics, 2007. **5**(2): p. 77-85.

21. Chen, C.H., et al., *Plasma neurofilament light chain and glial fibrillary acidic protein predict stroke in CADASIL.* J Neuroinflammation, 2020. **17**(1): p. 124.

22. Duering, M., et al., *Serum Neurofilament Light Chain Levels Are Related to Small Vessel Disease Burden.* J Stroke, 2018. **20**(2): p. 228-238.

23. Gravesteijn, G., et al., *Serum Neurofilament light correlates with CADASIL disease severity and survival.* Ann Clin Transl Neurol, 2019. **6**(1): p. 46-56.

24. Gao, D., et al., *Changes in the Morphology, Number, and Protein Levels of Plasma Exosomes in CADASIL Patients.* J Alzheimers Dis, 2021. **81**(1): p. 221-229.

25. Nagatoshi, A., et al., *Serum amyloid P component: A novel potential player in vessel degeneration in CADASIL.* J Neurol Sci, 2017. **379**: p. 69-76.

26. Arboleda-Velasquez, J.F., et al., *Hypomorphic Notch 3 alleles link Notch signaling to ischemic cerebral small-vessel disease.* Proc Natl Acad Sci U S A, 2011. **108**(21): p. E128-35.

27. Monet-Lepretre, M., et al., *Abnormal recruitment of extracellular matrix proteins by excess Notch3 ECD: a new pathomechanism in CADASIL.* Brain, 2013. **136**(Pt 6): p. 1830-45.

28. Issaq, H.J., Z. Xiao, and T.D. Veenstra, *Serum and plasma proteomics.* Chem Rev, 2007. **107**(8): p. 3601-20.

29. Lee, J.S., et al., *Impact of Brain MRI Markers on Major and Mild Vascular Cognitive Impairment in CADASIL.* Journal of the Korean Neurological Association, 2022. **40**(1): p. 39-46.

30. Kang, I.W., et al., *Accuracy of Korean-Mini-Mental Status Examination Based on Seoul Neuro-Psychological Screening Battery II Results.* Korean J Fam Med, 2016. **37**(3): p. 177-81.

31. Bamford, J.M., et al., *Interobserver agreement for the assessment of handicap in stroke patients.* Stroke, 1989. **20**(6): p. 828.

32. Wardlaw, J.M., *What is a lacune?* Stroke, 2008. **39**(11): p. 2921-2.

33. Zhao, Y., et al., *Neuronal injuries in cerebral infarction and ischemic stroke: From mechanisms to treatment (Review).* Int J Mol Med, 2022. **49**(2).

34. HaileMariam, M., et al., *S-Trap, an Ultrafast Sample-Preparation Approach for Shotgun Proteomics.* J Proteome Res, 2018. **17**(9): p. 2917-2924.

35. UniProt, C., *UniProt: the universal protein knowledgebase in 2021.* Nucleic Acids Res, 2021. **49**(D1): p. D480-D489.

36. Kall, L., et al., *Semi-supervised learning for peptide identification from shotgun proteomics datasets.* Nat Methods, 2007. **4**(11): p. 923-5.

37. Grzanka, M. and A. Piekielko-Witkowska, *The Role of TCOF1 Gene in Health and Disease: Beyond Treacher Collins Syndrome.* Int J Mol Sci, 2021. **22**(5).

38. Benarroch, R., et al., *The roles of cytosolic quality control proteins, SGTA and the BAG6 complex, in disease.* Adv Protein Chem Struct Biol, 2019. **114**: p. 265-313.

국문요약

CADASIL(Cerebral Autosomal Dominant Arteriopathy with Subcortical Infarcts and Leukoencephalopathy)은 뇌의 작은 혈관에 영향을 미쳐 다양한 신경학적 증상을 유발하는 희귀한 유전성 질환이다. 가장 흔한 증상은 비정상적인 대뇌 백질 형태와 백질 뇌증이며, 뇌 허혈, 인지장애, 정서불안, 편두통, 뇌졸중 등도 자주 나타난다. CADASIL을 진단하는 가장 좋은 방법은 NOTCH3 돌연변이에 대한 유전자 검사지만, 유전자 검사가 항상 가능한 것은 아니며 모든 CADASIL 환자가 병원성 NOTCH3 돌연변이를 가지고 있는 것은 아니다. 따라서 CADASIL의 진단과 치료에 도움을 줄 수 있는 혈액 기반 바이오마커 발굴 연구가 필요하다.

본 연구는 CADASIL 환자의 혈장 단백질 분석을 통해 진단 및 모니터링을 위한 바이오마커를 발굴하고, CADASIL 환자의 질병 관리 및 치료에 대한 근거를 제공하여 삶의 질 향상에 기여하는 것을 목표로 한다. 건강한 대조군과 뇌졸중 환자 그룹을 포함한 세 가지 그룹(각 그룹당 남성 5명, 여성 5명, 총 10명)의 유사한 연령(67±12.5세)을 가진 대상자로부터 혈장 샘플을 수집하였다. Multi Affinity Removal System Human 14 (MARS14) column (100 × 4.6 mm, Agilent Technology, Palo Alto, CA, USA)을 사용하여 수집한 혈장 샘플에서 상위 14개의 고농도 혈장 단백질을 제거하고, column에서 제거되지 않은 나머지 혈장 단백질을 이용하여 전처리를 진행하였다. Suspension Trap(S-Trap) 방법으로 정제된 단백질을 trypsin/LysC를 이용하여 37℃에서 16시간 동안 소화시켰다. Dionex UltiMate 3000 RSLCnano 시스템 (Thermo Fisher Scientific, Waltham, MA, USA)과 Q Exactive HF-X 질량분석기 (Thermo Fisher Scientific)를 사용하여 전처리가 완료된 샘플을 분석하였고, Proteome Discoverer (Thermo Fisher Scientific, ver 2.3) 및 Sequest HT 알고리즘을 갖춘 소프트웨어를 사용하여 1,902개의 혈장 단백질에 대한 스펙트럼을 획득하여 정량적 분석을 수행하였다. 이를 통해, 각 그룹에 대한 정량적 및 정성적 정보를 추출하고 차별적으로 발현된 단백질 (differentially expressed proteins, DEPs)을 확인하였다. 또한 주성분 분석, 벤 다이어그램 분석 및 volcano plot 분석을 통해 추가적인 통계 결과를 도출하였고, DEP 정보를 이용한 gene ontology (GO) 분석을 통해 후보 경로를 분석하였다. 이를 토대로 본 연구에서 밝혀진 잠재적인 CADASIL 혈장 바이오마커와 그동안의 연구에서 밝혀진 잠재적 바이오마커의 발현 결과를 비교했다. 그 결과, 이전 연구와 차별적으로 발현된 잠재적 바이오 마커로는 small glutamine-rich tetratricopeptide repeat-containing protein alpha (SGTA)와

Treacle protein (TCOF1)이 있다. SGTA는 두 군에 비해 CADASIL 환자에서 유의하게 증가한 반면, TCOF1은 감소하는 경향을 보였다. 이러한 단백질 후보는 추가 검증 연구와 함께 환자 혈장을 사용하여 CADASIL 관리 및 치료에 활용될 수 있으며 CADASIL 환자의 관리 및 치료에 기여할 것으로 기대된다.