Article

Using proteomic biomarkers to estimate acute ischaemic stroke onset time

Check for updates

Qianyun Li [1,6], Xiaodan Zhang[1,6], Yilin Zhang[1], Rex Pui Kin Lam[1], Weinv Fan[2], Yulan Jin[3], Chengcheng Ji[4], Jonathan Wilson Johnson[5] & Timothy Hudson Rainer [1] ✉

Abstract

Background Current guidelines recommend intravenous thrombolysis (IVT) for acute ischaemic stroke (AIS) patients within 4.5 hour (h) of symptom onset. Our study aims to use proteomic biomarkers to identify AIS patients with an onset time within 4.5 h when the history is not clear.

Methods We conducted a retrospective case-control study between June 2022 and July 2023 in Ningbo No. 2 Hospital, recruiting 30 AIS patients and 12 controls. Patients with AIS were grouped into early-onset (ES, symptom onset time ≤ 4.5 h, $n = 16$) and late-onset (LS, symptom onset 4.5–24 h, $n = 14$). Plasma proteome were identified using mass spectrometry. A stepwise analysis was conducted to screen for candidate proteins. Multiple logistic regression was used to construct various combinations.

Results Here we show six proteins discriminate ES from LS, with the area under curve (AUC) ranging from 0.897 to 0.951. Protein 4.2 (EPB42) achieves the highest AUC of 0.951 (95% confidence interval 0.882–1), a specificity of 0.929 (0.714–1) and a sensitivity of 0.875 (0.750–1). Ten combinations are derived from these six proteins, of which EPB42 and Phosphatidylethanolamine-binding protein 1 achieve an AUC of 0.991 (0.970–1), a specificity of 0.929 (0.857–1), and a sensitivity of 1 (0.875–1) in differentiating ES from LS.

Conclusions The six proteins and their combinations show promise as molecular clocks for determining the onset time of AIS in patients whose symptom onset time are unknown, potentially increasing their chances of receiving effective IVT to improve stroke outcomes.

Plain language summary

A stroke happens when blood flow to the brain is blocked, causing brain cells to die. It can be deadly, but quick treatment can save lives. One treatment, called intravenous thrombolysis, should be started within 4.5 h of stroke symptoms starting. However, some people don't know when their symptoms begin, which makes it hard to know if this treatment would be suitable for them. This study looked at blood samples from stroke patients to find out if certain proteins could show how recently the stroke happened. We found six proteins that could tell if a stroke occurred within the 4.5-hour window. Testing for these proteins could help doctors make faster, better treatment decisions and give more patients the care they need at an appropriate time.

Acute ischaemic stroke (AIS) is a common cerebrovascular disease affecting 68.16 million patients worldwide, with 7.59 million new cases and 3.48 million deaths yearly[1]. Administrating intravenous thrombolysis (IVT) therapy within 4.5 hour (h) of symptom onset is the most effective treatment for AIS[2]. However, a 20–25% of AIS patients, including those experiencing wake-up strokes and daytime-unwitnessed strokes, are not aware of the onset time[3]. Clinicians often resort to presuming that the last known well time as the onset time, leading to an overestimation of the onset-to-presentation time that often exceeds the therapeutic window of 4.5 h. As a result, these patients usually miss the opportunity for IVT, leading to poorer outcomes and increased post-discharge dependence on their families[4]. A personalised approach is required to reduce the risk of missing IVT treatment in AIS patients whose symptom onset time is unclear.

Several large clinical trials have confirmed the beneficial outcomes of advanced neuroimaging-based mismatch in identifying salvageable brain tissue for IVT in AIS patients with unknown time of onset[3]. However, timely access to these advanced neuroimaging is limited in many hospitals, even in developed regions[5]. Some patients also have restrictions related to the performing of neuroimaging or the usage of contrast medium[6]. Moreover, the discriminatory performance of the mismatch method is not ideal, with a sensitivity of 62% and a specificity of 78%[7]. Thus, there is a pressing need for more accessible and efficient complementary solutions to address these limitations.

AIS triggers changes in blood proteins over time[8], wherein distinct temporal alterations in certain proteins may serve as biomarkers to estimate the stroke onset time, resembling a molecular clock. However, no study

[1]Department of Emergency Medicine, The University of Hong Kong, Hong Kong SAR, China. [2]Department of Neurology, Ningbo No.2 Hospital, Ningbo, Ningbo, Zhejiang Province, China. [3]Department of Clinical Laboratory, Ningbo No.2 Hospital, Ningbo, Ningbo, Zhejiang Province, China. [4]School of Medicine, Shaoxing University, Shaoxing, Shaoxing, Zhejiang Province, China. [5]School of Humanities, the University of Hong Kong, Hong Kong SAR, China. [6]These authors contributed equally: Qianyun Li, Xiaodan Zhang. ✉e-mail: thrainer@hku.hk

raises this concept of molecular clock in determining the onset time of AIS. Studies that explore blood biomarkers in this field are also scarce. Although there is an animal study that identifies a panel of 5 metabolites as biomarkers[9], the panel is originated from rat and the performance is sub-optimal with an area under curve (AUC) of 0.87. While in this study, we directly use AIS patients' plasma samples and apply high-throughput proteomic profiling. We identify six potential proteins with AUCs ranging from 0.897 to 0.951 in estimating the onset time of AIS within 4.5 h. The various combinations of the six proteins achieve even better AUCs ranging from 0.978 to 0.991.

Methods

Study design, setting and population

This was an exploratory retrospective case-control study. A convenience blood sample of patient participants was collected between June 2020 and July 2023 from the emergency and neurology departments of Ningbo No. 2 Hospital, China. Patients were identified based on one or more neurological symptoms suggestive of stroke with a sudden onset, such as weakness or numbness in the limbs or face, sensory deficits, dizziness, aphasia, dysarthria, dysphagia, ataxia, visual field defects or neglect, and cognitive impairment. The diagnosis of AIS was established by two neurologists and confirmed by the presence of a Diffusion Weighted Imaging (DWI)-positive lesion on Magnetic Resonance Imaging (MRI) or the detection of a new hypodense lesion on Computed Tomography (CT) scans. Researchers were blinded to biomarker information when making the diagnosis. Controls were individuals undergoing health checkup at the hospital's health examination center. They had no stroke symptoms but had diabetes, hypertension, or hyperlipidaemia.

The inclusion criteria were: (1) AIS patients within 24 h of symptom onset; (2) age $\geq$ 18 years; (3) informed consent. Patients who were (1) not compliant with neurological examination and follow-up or (2) had an unclear symptom onset time were excluded.

The study protocol was approved by the ethics committee of Ningbo No. 2 Hospital (Approval Number YJ-NBEY-KY-2023-099-01), with all participants providing written informed consent. The study was conducted with full compliance with the Declaration of Helsinki, and relevant local guidelines and regulations.

Reference standard

The reference standard was defined as neurologists' diagnosis of AIS with a clear onset time within 4.5 h, as reported by the patients. The pre-defined cutoff of $\leq$4.5 h was selected to align with the therapeutic window of IVT for AIS. The AIS patients were split into two groups accordingly: the early AIS group (ES), with an onset time within 4.5 h (0–4.5 h), and the late AIS group (LS), with an onset time beyond 4.5 h (4.5–24 h).

Index test

The index test comprised tests for single proteins or their combinations in patients' plasma for discrimination between ES and LS. The trends of protein concentrations after stroke onset over time were evaluated when determining the positivity of the index test. If a protein showed a consistent upward trend within 24 h of onset, values below the specified cutoff were considered positive for ES. Conversely, for proteins displaying a consistent downward trend, values surpassing the cutoff were considered positive for ES.

Data collection

Clinical data, including demographic information and medical history, were extracted from medical records. The severity of stroke on admission was assessed using the National Institutes of Health Stroke Scale (NIHSS)[10]. The aetiology of AIS was determined using the Trial of Org 10172 in Acute Stroke Treatment (TOAST) classification, which included categories containing large artery atherosclerosis (LAA), cardio-embolism (CE), small vessel occlusion (SVO), stroke of other determined aetiology, and stroke of undetermined aetiology[11]. Outcome evaluation at 3-month following AIS

was conducted using the modified Rankin Scale (mRS)[12]. An mRS score of 0–2 indicated a good outcome, and a score of 3–6 indicated a poor outcome. The infarct volume (mm^3) was calculated using the $(a \times b \times c)/2$ formula based on MRI-DWI or CT images (in cases where MRI was unavailable)[13].

Blood sampling

Potassium ethylenediaminetetraacetic acid (EDTA) Vacutainer tubes (Becton, Dickinson and Company, New Jersey, US) were used to collect blood samples from the patient participants at hospital admission. The blood was quickly transferred on ice to the hospital laboratory and centrifuged at 1500 g for 10 min at 4 °C. The plasma was then extracted and immediately stored at −80 °C. The haemolysed samples were excluded. Time interval from stroke onset to blood sampling was recorded.

Proteomic analyses

Plasma proteome profiling was performed at Shanghai Lu-Ming Biotech Co., Ltd. (Shanghai, China). The operators performing the proteomic analysis were not blinded to the grouping information. Measurements were taken from distinct samples in both AIS and control groups. The processing and detection of all samples were followed the same standardized procedures.

Initially, proteins extracted from plasma were enzymatically hydrolysed into peptide fragments which were subsequently analysed using mass spectrometry (MS). Specifically, plasma samples underwent proteomic profiling using liquid chromatography (Agilent 1100 HPLC system) coupled with a tandem MS (LC-MS/MS) (Bruker timsTOF Pro, Bremen, Germany). The MS analysis process comprised two primary steps. First, the traditional data-dependent acquisition (DDA) method was employed to establish a protein spectral library. Subsequently, 4-dimensional-data-independent acquisition (4D-DIA) technology was utilised to gather MS data for each sample, facilitating spectrum matching, quantitative information extraction, and subsequent statistical analysis based on the aforementioned DDA database.

Following the retrieval of raw data from the database, proteins were qualified and quantified. Proteins exhibiting expression in over 50% of samples within any group were retained. Any absent protein expression values were imputed using one tenth minimal expression value of the respective protein within the same group. Subsequently, the protein expression values underwent median normalization and log2 transformation. The values obtained was referred as relative abundance (RA) values which was utilised to denote the concentration of the detected proteins. RA was served as the unit.

Statistical analysis

Statistical analyses were conducted using R software (version 4.0.3, R Core Team) and GraphPad Prism 9.1 software (GraphPad Software, Boston, Massachusetts, US). The sample size for the diagnostic test was determined following the method proposed by Buderer[14], with the expected specificity set at 0.95, the prevalence of ES in the AIS population within 24 h at 0.2, a 95% confidence level (CI), and a 95% CI width of no more than 10%. The calculated total sample size was 23.

To ensure that the control group and the AIS group were comparable in terms of age, gender, diabetes, hypertension, and hyperlipidaemia, a matching process was performed between the two groups. Participants in the ES and LS groups were additionally matched on smoking and alcohol consumption. A student's t-test was used to compare the ages of the two groups and a chi-square test was used to compare categorical variables, including gender, diabetes, hypertension, hyperlipidaemia, smoking and alcohol consumption. If $p > 0.05$, the variable was considered comparable between the two groups.

For statistical methods on identifying the differential proteins between two groups, student's t-test was utilised to assess statistical significance of the abundance of proteins, with p-values < 0.05 indicating statistical significance. Fold change (FC) analysis was used to calculate the ratio of the average abundance of a protein in one group compared to another.

Bioinformatics analyses of the differential proteins were carried out using the tools available at https://cloud.oebiotech.com/, encompassing gene ontology (GO) enrichment analysis, Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway enrichment analysis, and protein-protein interactions (PPI).

Then it was the stepwise identification of potential proteins as biomarkers. Binary logistic regression was used to calculate AUC, sensitivity, specificity, accuracy, positive predictive value (PPV), negative predictive value (NPV) with 95% CI to present the classification performance of the proteins between two groups. The dynamic changes of proteins were modeled using nonlinear regression. Two machine learning techniques were employed for feature selection in biomarker screening: variable importance analysis with random forest and least absolute shrinkage and selection operator (LASSO). Variable importance analysis assessed the contribution of predictor variables to the model's predictive performance, ranking variables based on their importance score[15]. LASSO, a regression

Fig. 1 | Selection process flowchart for potential protein biomarkers. Initially, 950 proteins were identified through untargeted mass spectrometry detection. Subsequent filtering excluded 881 proteins that did not meet the criteria of significant difference ($p < 0.05$) and fold changes (FC) > 1.2 or <1/1.2 between the AIS and control groups, as well as the ES and LS groups. Ten proteins were further eliminated due to an AUC p-value > 0.05 between the ES and LS groups. Among the remaining 59 proteins, 36 were removed for having an AUC < 0.8 in differentiating the ES and LS groups. Of the 23 remaining proteins, two were excluded as their dynamic changes did not exhibit a consistent unidirectional trend. Subsequently, RF variable importance analysis ranked the remaining 21 proteins, revealing the top 10 candidates. LASSO selection narrowed down the pool to eight proteins. An intersection analysis aligning the top 10 proteins from RF with the eight proteins from LASSO identified a common subset of eight proteins. Finally, two proteins were excluded due to an AUC less than 0.890, resulting in the final selection of six proteins. AIS acute ischaemic stroke, AUC Area Under the Curve, ES early AIS with onset time within 4.5 h, FC fold change, LASSO least absolute shrinkage and selection operator, LS late AIS with onset time between 4.5 and 24 h, RF random forest.
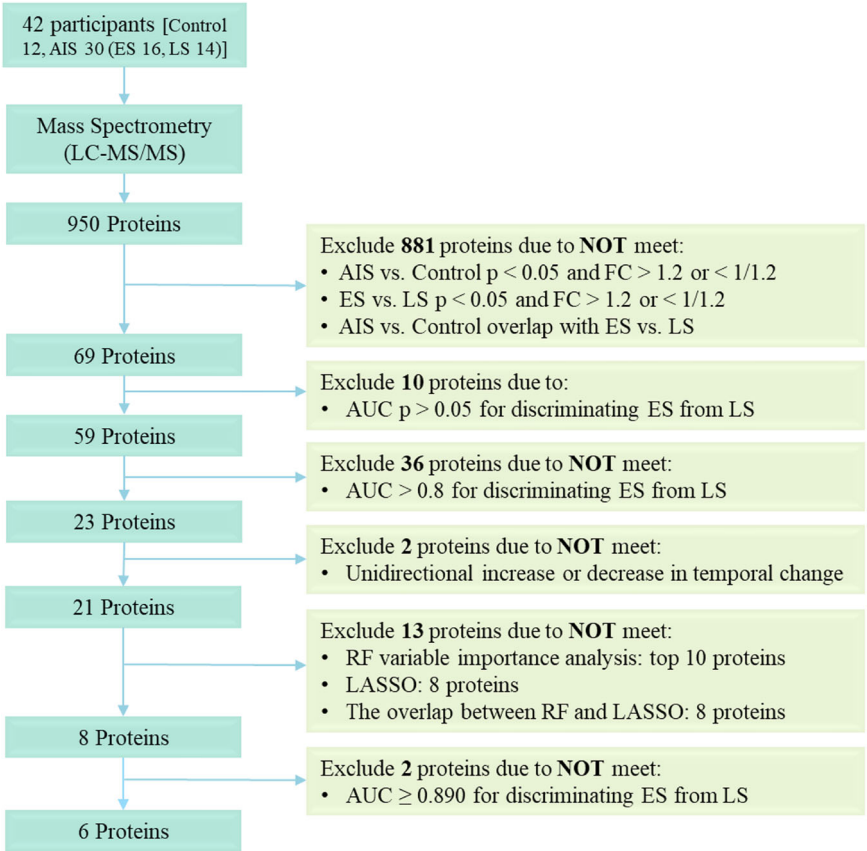

Table 1 | Demographic and Clinical Characteristics of participants

	Total			AIS only		
	Controls ($n = 12$)	AIS ($n = 30$)	P value	ES ($n = 16$)	LS ($n = 14$)	P value
Age, mean (SD), y	66.17 ± 6.25	70.6 ± 7.68	0.08	73 ± 9.25	71 ± 9	0.56
Female (%)	5 (41.67%)	7 (23.33%)	0.42	4 (25%)	3 (21.43%)	1.00
Diabetes mellitus (%)	3 (25%)	9 (30%)	1.00	4 (25%)	5 (35.71%)	0.69
Hypertension (%)	5 (41.67%)	16 (53.33%)	0.50	11 (68.75%)	5 (35.71%)	0.14
Hyperlipidaemia (%)	1 (8.33%)	3 (10%)	1.00	2 (12.5%)	1 (7.14%)	1.00
Smoking (%)				8 (50%)	3 (21.43%)	0.09
Alcohol (%)				3 (18.75%)	4 (28.57%)	1.00
Onset-sampling-time (IQR), h				1.52 (0.93, 2.21)	15.18 (6.22, 19.84)	0.000000014
NIHSS-admission (IQR)				5 (4, 6.5)	11 (5.5, 17.5)	0.045
Infarct-volume, mean (SD), mm³				49.76 ± 41.71	43.05 ± 32.33	0.72
TOAST-classification						0.08
Cardioembolism (%)				7 (43.75%)	4 (28.57%)	
Large-artery atherosclerosis (%)				6 (37.5%)	7 (50%)	
Small-vessel occlusion (%)				3 (18.75%)	3 (21.43%)	

AIS acute ischaemic stroke, *ES* early AIS with onset time within 4.5 h, *IQR* interquartile range, *LS* late AIS with onset time 4.5–24 h, *NIHSS* National Institute of Health Stroke Scale, *SD* standard deviation, *TOAST* Trial of Org 10172 in Acute Stroke Treatment. There was missing value of infarct volume for 7 participates in ES group and 6 participates in LS group.

technique incorporating regularisation, retained key features by shrinking some coefficients to zero[16]. Multivariable logistic regression was utilised to create combinations from the final selected proteins for identifying AIS with an onset time within 4.5 h.

　　　Then it was the statistical method on the comparison of demographic and clinical characteristics, the abundance comparison of the final selected proteins in different groups, and their association with clinical metrics. For continuous data, Shapiro normality test was employed to determine data normality. When the data followed a normal distribution, we represented it using mean ± standard deviation (SD), and compared two groups using independent student t-test. For comparisons involving three or more groups, we utilized One-way ANOVA, with Fisher's LSD test for post-hoc pairwise comparisons. When the data did not follow a normal distribution, we presented it using the median (interquartile range, IQR), and compared two groups using the Wilcoxon test. For comparisons consisting three or more groups in this condition, we employed non-parametric test (Kruskal-

Wallis test) and conducted Dunn's test as post-hoc analysis. Categorical data were described using frequency (percentage, %) for statistical description, and group comparisons were made using the χ^2 test. A two-tailed p-value < 0.05 was considered statistically significant. Correlation analysis was performed using Pearson's correlation to evaluate the associations of selected proteins with infarct volume and NIHSS scores.

Reporting summary
Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Results
Participants
Thirty patients diagnosed with AIS, including 16 patients with ES and 14 with LS. And 12 control individuals were recruited in this study (Fig. 1). The participant flow was illustrated in Supplementary Fig. 1. Table 1
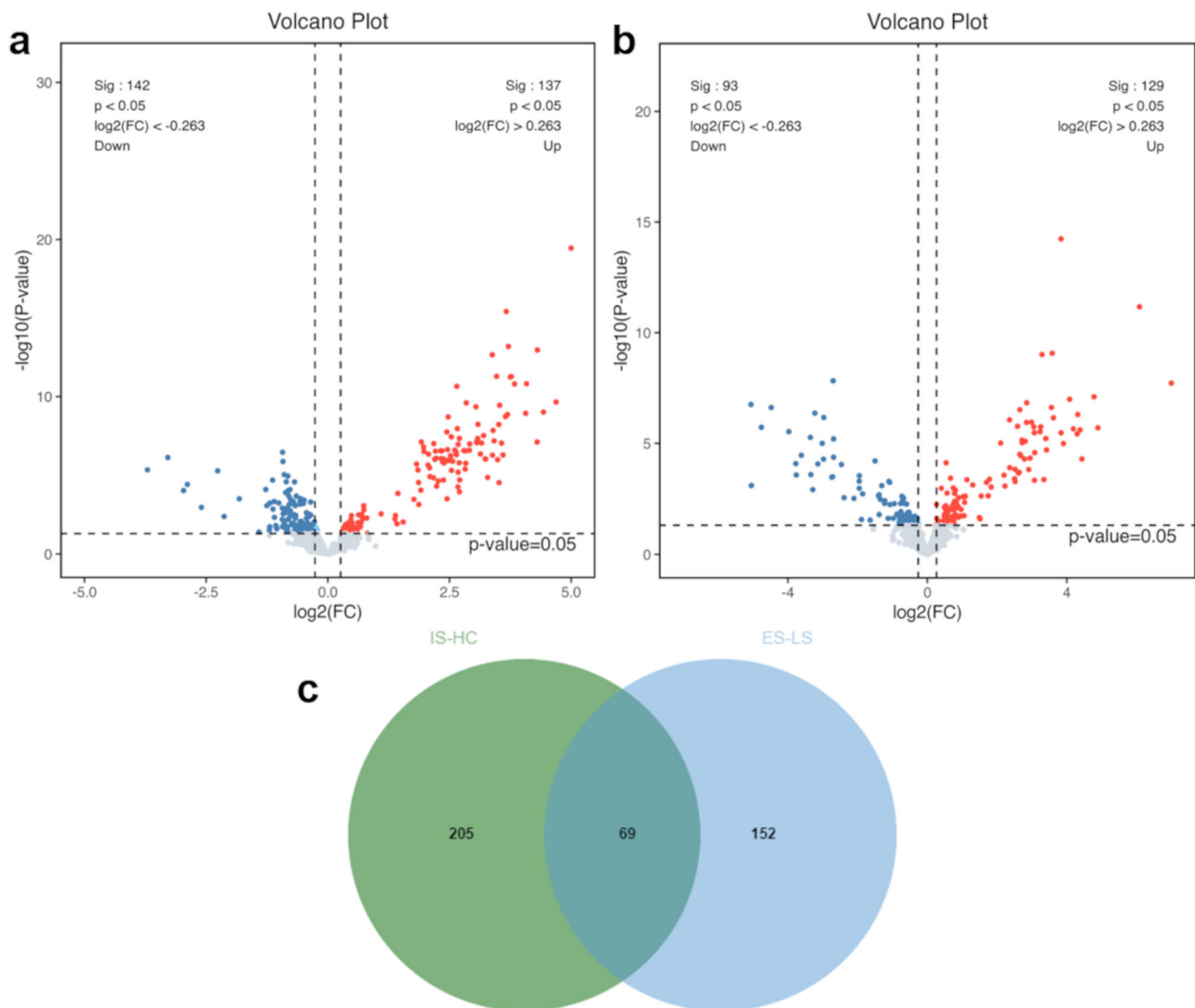

Fig. 2 | Proteomic profiling and differential proteins. a. Volcano plot of differential proteins between AIS patients and control. In this plot, 142 proteins were downregulated while 137 proteins were upregulated (totaling 279 differential proteins) in patients with AIS ($n = 30$ biologically independent samples) compared to controls ($n = 12$ biologically independent samples), meeting the criteria of $p < 0.05$ and fold change (FC) > 1.2 or <1/1.2. The x-axis denotes the log2 (FC) values, while the y-axis represents the -log10 (p-value) values. Red dots signify significantly upregulated proteins, blue dots indicate significantly downregulated proteins, and lower gray dots represent proteins with no significant difference. **b** Volcano plot of differential proteins between ES and LS. This plot illustrates 93 downregulated and 129 upregulated proteins (totaling 222 differential proteins) between ES ($n = 16$ biologically independent samples) and LS groups ($n = 14$ biologically independent samples), meeting the criteria of $p < 0.05$ and FC > 1.2 or <1/1.2. **c** Venn diagram overlap of differential proteins. The Venn diagram showcases the overlap of differential proteins between AIS-control (IS-HC) and ES-LS groups, revealing 69 proteins in common. AIS acute ischaemic stroke, ES early AIS with onset time within 4.5 h, FC fold change, HC control, LS late AIS with onset time > 4.5 h.

Table 2 | The performance of 6 individual proteins in identifying acute ischaemic stroke within 4.5 h

Protein	FC	AUC (95% CI)	Accuracy (95% CI)	Cutoff value	Sensitivity (95% CI)	Specificity (95% CI)	PPV (95% CI)	NPV (95% CI)
EPB42	19.874	0.951 (0.882–1.000)	0.900 (0.833–1.000)	16.243	0.875 (0.750–1.000)	0.929 (0.714–1.000)	0.933 (0.800–1.000)	0.867 (0.750–1.000)
PEBP1	20.827	0.942 (0.867–1.000)	0.867 (0.800–1.000)	16.471	0.750 (0.686–1.000)	1.000 (0.714–1.000)	1.000 (0.800–1.000)	0.778 (0.722–1.000)
PIK3IP1	0.082	0.911 (0.810–1.000)	0.867 (0.767–0.967)	15.944	1.000 (0.625–1.000)	0.714 (0.571–1.000)	0.800 (0.727–1.000)	1.000 (0.700–1.000)
CD36	27.487	0.906 (0.774–1.000)	0.933 (0.833–1.000)	15.914	1.000 (1.000–1.000)	0.857 (0.643–1.000)	0.889 (0.762–1.000)	1.000 (1.000–1.000)
DDR1	4.298	0.906 (0.767–1.000)	0.933 (0.833–1.000)	12.713	1.000 (1.000–1.000)	0.857 (0.643–1.000)	0.889 (0.762–1.000)	1.000 (1.000–1.000)
IL1R2	10.699	0.897 (0.774–1.000)	0.900 (0.800–1.000)	15.806	0.938 (0.812–1.000)	0.857 (0.643–1.000)	0.882 (0.762–1.000)	0.923 (0.800–1.000)

AUC area under the curve, CD36 Platelet glycoprotein 4, DDR1 Epithelial discoidin domain-containing receptor 1, EPB42 Protein 4.2, FC fold change, IL1R2 Interleukin-1 receptor type 2, NPV negative predictive value, PEBP1 Phosphatidylethanolamine-binding protein 1, PIK3IP1 Phosphoinositide-3-kinase-interacting protein 1, PPV positive predictive value. p-value comparison between early and late acute ischaemic stroke groups.

Table 3 | The top 10 combinations in identifying acute ischaemic stroke within 4.5 h

Combinations	AUC (95% CI)	Change in C-statistic	Accuracy (95% CI)	Sensitivity (95% CI)	Specificity (95% CI)	PPV (95% CI)	NPV (95% CI)
EPB42 + PEBP1	0.991 (0.970–1.000)	0	0.967 (0.933–1.000)	1.000 (0.875–1.000)	0.929 (0.857–1.000)	0.941 (0.889–1.000)	1.000 (0.875–1.000)
CD36 + IL1R2	0.991 (0.970–1.000)	0	0.967 (0.933–1.000)	0.938 (0.875–1.000)	1.000 (0.929–1.000)	1.000 (0.941–1.000)	0.933 (0.875–1.000)
EPB42 + PEBP1 + CD36	0.991 (0.970–1.000)	0	0.967 (0.900–1.000)	1.000 (0.875–1.000)	0.929 (0.857–1.000)	0.941 (0.889–1.000)	1.000 (0.875–1.000)
EPB42 + PIK3IP1 + DDR1	0.991 (0.970–1.000)	0	0.967 (0.933–1.000)	0.938 (0.875–1.000)	1.000 (0.929–1.000)	1.000 (0.941–1.000)	0.933 (0.875–1.000)
PIK3IP1 + CD36 + IL1R2	0.991 (0.970–1.000)	0	0.967 (0.933–1.000)	0.938 (0.875–1.000)	1.000 (0.857–1.000)	1.000 (0.889–1.000)	0.933 (0.875–1.000)
PIK3IP1 + DDR1 + IL1R2	0.991 (0.970–1.000)	0	0.967 (0.933–1.000)	1.000 (0.875–1.000)	0.929 (0.857–1.000)	0.941 (0.889–1.000)	1.000 (0.875–1.000)
PEBP1 + CD36	0.987 (0.960–1.000)	−0.004	0.933 (0.900–1.000)	0.938 (0.812–1.000)	0.929 (0.857–1.000)	0.938 (0.889–1.000)	0.929 (0.824–1.000)
PEBP1 + PIK3IP1 + CD36	0.987 (0.960–1.000)	−0.004	0.933 (0.900–1.000)	0.938 (0.812–1.000)	0.929 (0.857–1.000)	0.938 (0.882–1.000)	0.929 (0.824–1.000)
PEBP1 + PIK3IP1	0.987 (0.957–1.000)	−0.004	0.967 (0.900–1.000)	0.938 (0.812–1.000)	1.000 (0.929–1.000)	1.000 (0.941–1.000)	0.933 (0.824–1.000)
EPB42 + PIK3IP1	0.978 (0.936–1.000)	−0.013	0.933 (0.867–1.000)	0.938 (0.812–1.000)	0.929 (0.786–1.000)	0.938 (0.842–1.000)	0.929 (0.824–1.000)

AUC area under the curve, CD36 Platelet glycoprotein 4, DDR1 Epithelial discoidin domain-containing receptor 1, EPB42 Protein 4.2, IL1R2 Interleukin-1 receptor type 2, NPV negative predictive value, PEBP1 Phosphatidylethanolamine-binding protein 1, PIK3IP1 Phosphoinositide-3-kinase-interacting protein 1, PPV positive predictive value.

summarised the demographic and clinical characteristics of the patients. Age, sex, and cardiovascular risk factors, including diabetes mellitus, hypertension, and hyperlipidaemia, did not differ significantly between the AIS and control groups. Additionally, there were no significant differences on age, sex, risk factors, infarct volume, and TOAST classification between the ES and LS groups. However, the LS group had a significantly higher admission NIHSS score than the ES group.

The median time from symptom onset to sampling was 1.52 h in the ES group and 15.18 h in the LS group. The average time interval between clinical diagnosis and LC-MS/MS analysis was 8 months. No adverse events related to the index test were reported.

Stepwise screening of potential proteins for discriminating ES and LS

We conducted a stepwise screening process, as depicted in Fig. 1. Initially, 1124 proteins were detected, among which 950 proteins were identified as credible proteins through the untargeted MS analysis. We then identified 279 proteins that exhibited differential expression between the AIS and control groups at a significance level of $p < 0.05$ and FC > 1.2 or <1/1.2, comprising 142 downregulated and 137 upregulated proteins in the AIS group (Fig. 2a). In the comparison between the ES and LS groups, 222 proteins showed differential expression, with 93 downregulated and 129 upregulated proteins in the ES group (Fig. 2b). The intersection of these two sets of differential proteins revealed 69 common proteins that met the criteria of significant differences ($p < 0.05$) and FC > 1.2 or <1/1.2 in both comparisons (Fig. 2c).

Bioinformatic analyses of these differential proteins were detailed in the supplemental materials, encompassing GO enrichment analysis (Supplementary Fig. 2), KEGG enrichment analysis (Supplementary Fig. 3), and PPI network analysis (Supplementary Fig. 4).

The 69 proteins were further evaluated based on their ability to differentiate the ES and LS groups using the AUC, resulting in 59 proteins with AUC p-values < 0.05. Subsequently, a stricter criterion of AUC > 0.8 led to 23 proteins being retained. Further assessment for unidirectional temporal

changes narrowed down the selection to 21 proteins, which were ranked by their AUC in differentiating ES from LS (see Supplementary Data 1).

RF variable importance analysis was then applied to rank the 21 proteins, with the top 10 being selected. LASSO was subsequently used, resulting in the selection of 8 proteins from the initial 21. An overlap analysis between the top 10 proteins from random forest and the 8 proteins from LASSO yielded 8 proteins. Following the exclusion of two proteins with an AUC of less than 0.890, 6 high-performance proteins emerged: protein 4.2 (EPB42), phosphatidylethanolamine-binding protein 1 (PEBP1), phosphoinositide-3-kinase-interacting protein 1 (PIK3IP1), platelet glycoprotein 4 (CD36), epithelial discoidin domain-containing receptor 1 (DDR1), and interleukin-1 receptor type 2 (IL1R2).

The individual discriminatory performances of these 6 proteins in distinguishing between the ES and LS groups were outlined in Table 2, ranked by AUC values. Notably, EPB42 showed the highest AUC of 0.951 (95% CI 0.882–1). At a cutoff value of 16.243 RA, it demonstrated a specificity of 0.929 (95% CI 0.714–1), a sensitivity of 0.875 (95% CI 0.750–1), an accuracy of 0.900 (95% CI 0.833–1), a PPV of 0.933 (95% CI 0.800-1), and an NPV of 0.867 (95% CI 0.750–1). PEBP1 exhibited a slightly lower AUC of 0.942 (95% CI 0.867–1), a specificity of 1 (95% CI 0.714–1), a sensitivity of 0.750 (95% CI 0.686–1), an accuracy of 0.867 (95% CI 0.800–1.000), a PPV of 1 (95% CI 0.800-1), and an NPV of 0.778 (95% CI 0.722–1). The third one was PIK3IP1 with an AUC of 0.911 (95% CI 0.810–1), a specificity of 0.714 (95% CI 0.571–1), a sensitivity of 1 (95% CI 0.625–1), an accuracy of 0.867 (95% CI 0.767–0.967), a PPV of 0.800 (95% CI 0.727–1), and an NPV of 1 (95% CI 0.700–1). The fourth to fifth protein were CD36, DDR1, and IL1R2, respectively, with the detailed performance specified in Table 2.

Identification of optimal combinations from the six proteins

Table 3 shows the results of the top 10 panels with the highest AUC value after multivariable logistic regression in distinguishing between ES and LS groups.

Among these combinations, the panel of EPB42 and PEBP1 yielded an AUC of 0.991 (95% CI 0.970–1), with an accuracy of 0.967 (95% CI

Fig. 3 | Box plots illustrating the relevance abundance of the six proteins. a EPB42 relative abundance in ES ($n = 16$ biologically independent samples), LS ($n = 14$ biologically independent samples), and control ($n = 12$ biologically independent samples) groups. **b** PEBP1 relative abundance across the three groups. **c** PIK3IP1 relative abundance among the three groups. **d** CD36 relative abundance in the three groups. **e** DDR1 relative abundance across the three groups. **f** IL1R2 relative abundance among the three groups. The center line: median; the error bars, upper and lower quartiles. CD36 Platelet glycoprotein 4, DDR1 Epithelial discoidin domain-containing receptor 1, ES early acute ischaemic stroke group with onset time within 4.5 h, EPB42 Protein 4.2, IL1R2 Interleukin-1 receptor type 2, LS late acute ischaemic stroke group with onset time between 4.5 and 24 h, PEBP1 Phosphatidylethanolamine-binding protein 1, PIK3IP1 Phosphoinositide-3-kinase-interacting protein 1.
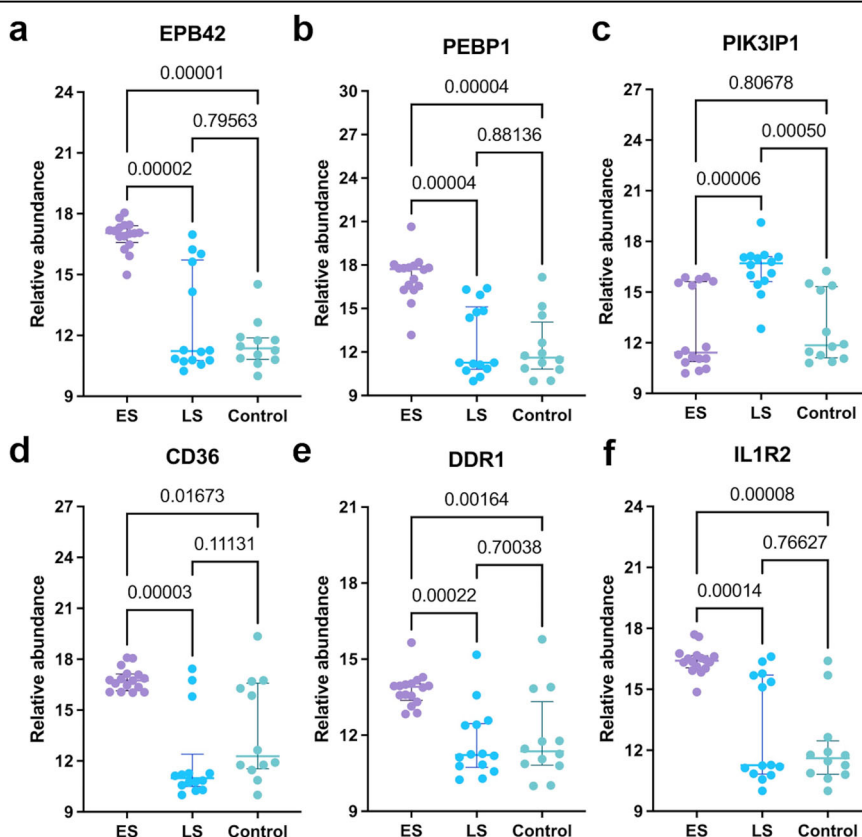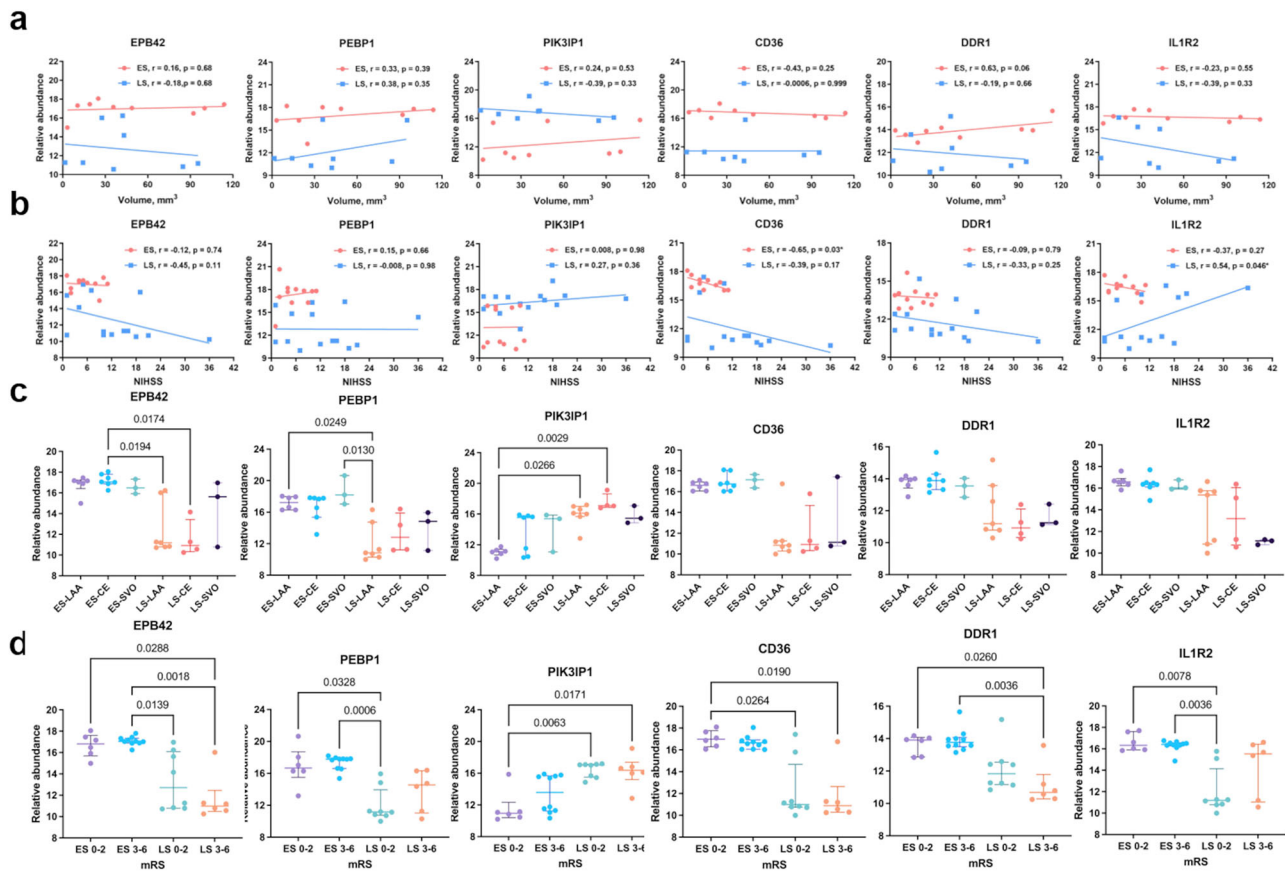
Fig. 4 | Associations of the six proteins with clinical metrics. a Protein correlation with the infarct volume. ES: $n = 9$ biologically independent samples; LS: $n = 8$ biologically independent samples. **b** Protein correlation with NIHSS score. ES: $n = 11$ biologically independent samples; LS: $n = 14$ biologically independent samples. **c** Protein levels in ES or LS subgroups of the TOAST classification (LAA, CE, SVO). ES-LAA: $n = 6$; ES-CE: $n = 7$; ES-SVO: $n = 3$; LS-LAA: $n = 7$; LS-CE: $n = 4$; LS-SVO: $n = 3$. The center line: median; the error bars: upper and lower quartiles. **d** Protein levels in ES or LS subgroups of prognosis (Good: 0–2, Poor: 3–6) according to the mRS score at 3 months follow-up. ES 0–2: $n = 6$; ES 3–6: $n = 10$; LS 0–2: $n = 8$; LS 3–6:

$n = 6$. The center line: median; the error bars, upper and lower quartiles. Abbreviations: CD36 Platelet glycoprotein 4, CE cardio-embolism, DDR1 Epithelial discoidin domain-containing receptor 1, EPB42 Protein 4.2, ES early acute ischaemic stroke group with onset time within 4.5 h, IL1R2 Interleukin-1 receptor type 2, LAA large artery atherosclerosis, LS late acute ischaemic stroke group with onset time between 4.5 and 24 h, mRS modified Rankin Scale, NIHSS National Institutes of Health Stroke Scale, PEBP1 Phosphatidylethanolamine-binding protein 1, PIK3IP1 Phosphoinositide-3-kinase-interacting protein 1, SVO small vessel occlusion.

0.933–1), a specificity of 0.929 (95% CI 0.857–1), a sensitivity of 1 (95% CI 0.875–1), a PPV of 0.941 (95% CI 0.889–1), and an NPV of 1 (95% CI 0.875–1). Similarly, the combination of CD36 and IL1R2 exhibited an AUC of 0.991 (95% CI 0.970–1), with an accuracy of 0.967 (95% CI 0.933–1), a specificity of 1 (95% CI 0.929–1), a sensitivity of 0.938 (95% CI 0.875–1), a PPV of 1 (95% CI 0.941–1), and an NPV of 0.933 (95% CI 0.875–1). Table 3 presented the remaining eight combinations with their classification performance.

Associations of the six proteins with clinical metrics
Figure 3 illustrated the RA of the six proteins in the ES, LS, and control groups. Notably, EPB42 (Kruskal-Wallis test, $p = 0.00002$), PEBP1(Kruskal-Wallis test, $p = 0.00004$), CD36 (Kruskal-Wallis test, $p = 0.00003$), DDR1(Kruskal-Wallis test, $p = 0.00022$), and IL1R2 (Kruskal-Wallis test, $p = 0.00014$) exhibited higher levels in the ES group compared to the LS group, while PIK3IP1 (Kruskal-Wallis test, $p = 0.00006$) demonstrated elevated levels in the LS group compared to the ES group. The RAs of these proteins did not exhibit correlations with the infarct volume in either group, as depicted in Fig. 4a.

CD36 in the ES group (Pearson's correlation, r = −0.65, $p = 0.03$) and IL1R2 in the LS group (Pearson's correlation, r = 0.54, $p = 0.046$) displayed moderate to weak correlations with the NIHSS score, as shown in Fig. 4b. When considering the TOAST classification, the levels of these proteins did not show significant differences in pairwise comparisons

among the LAA, CE, and SVO subgroups within the ES or LS group, as depicted in Fig. 4c.

No distinctions were observed in the protein levels between good prognosis (mRS 0–2) and poor prognosis (mRS 3–6) at three months within the ES or LS groups, as illustrated in Fig. 4d.

Discussion
This study identified potential proteomic biomarkers for categorising the onset time of AIS as either within or beyond 4.5 h. Using high-throughput plasma proteome profiling alongside methodologies such as random forest, LASSO, and logistic regression, we identified six proteins with individual AUCs ranging from 0.897 to 0.951, PPVs ranging from 0.800 to 1, and NPVs ranging from 0.778 to 1 for distinguishing patients with AIS onset within 4.5 h. By constructing protein combinations, we achieved improved AUCs ranging from 0.978 to 0.991, PPVs ranging from 0.938 to 1, and NPVs ranging from 0.929 to 1. These individual proteins or protein panels hold promise in identifying AIS cases within 4.5 h of symptom onset, potentially assisting clinicians in determining eligibility for IVT in AIS patients with an unknown time of onset.

EPB42, a key component of the erythrocyte membrane cytoskeleton, maintained the structural integrity and flexibility of red blood cells[17,18]. Its significant elevation in the ES group compared to the LS group may be linked to red blood cell rupture in the cerebral infarct region.
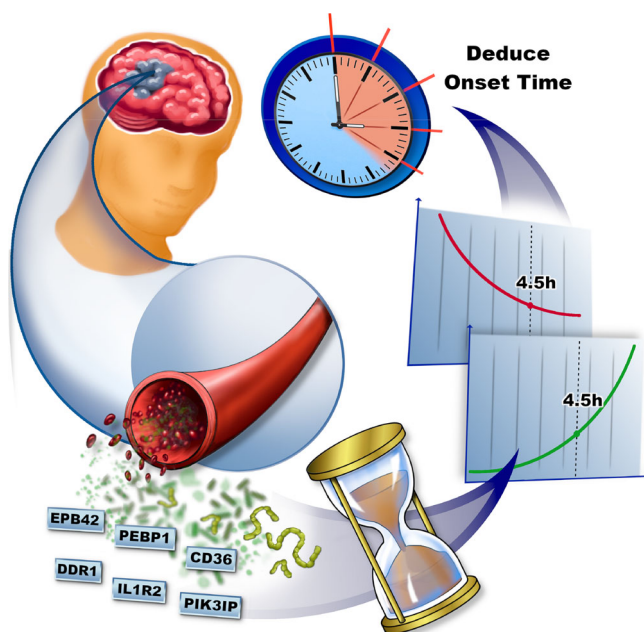
Fig. 5 | The concept of molecular clock. After the occurrence of acute ischaemic stroke (AIS), brain tissue triggers an ischaemic cascade reaction due to ischaemia and hypoxia, leading to changes in the levels of circulating protein molecules over time. The identified six proteins exhibit significant upward or downward trends. Based on their circulating levels, we can infer whether the onset of AIS in patients with unknown onset time occurred within 4.5 h. This method, which uses the levels of biomolecules in the peripheral circulation to retroactively estimate the onset time of the disease, resembles a molecular clock. It is based on the principle that the occurrence of the disease induces changes in the levels of biomolecules in circulation over time. CD36 Platelet glycoprotein 4, DDR1 Epithelial discoidin domain-containing receptor 1, EPB42 Protein 4.2, IL1R2 Interleukin-1 receptor type 2, PEBP1 Phosphatidylethanolamine-binding protein 1, PIK3IP1 Phosphoinositide-3-kinase-interacting protein 1.

PEBP1, located in the cell membrane and cytoplasm, was widely expressed across various tissue types, including the brain[19]. Its functions spanned from modulating neural progenitor cell differentiation[20] to mediating protective effects against cerebral ischaemic injury[21,22], potentially indicating brain self-protection. On the other hand, CD36's activities after stroke were largely detrimental, contributing to various negative outcomes such as blood–brain barrier (BBB) impairment[23], inflammation[24,25], endothelial dysfunction[26], and tissue damage in cerebral ischaemia[27]. DDR1, a member of the discoidin domain receptor family of tyrosine kinases, primarily regulated collagen synthesis and breakdown, affecting the extracellular matrix[28]. Elevated DDR1 levels post-AIS may contribute to BBB disruption and cerebral damage[29].

PIK3IP1 played a protective role against ischaemic injury, particularly through interactions with the phosphoinositide 3-kinase signalling pathway[30]. Its gradual rise, particularly the higher levels in the LS group, suggested an association with later tissue protection mechanisms[31,32]. IL1R2 served as a decoy receptor for IL-1β and inhibited downstream pro-inflammatory signalling pathways critical to the inflammatory response following ischaemic events[33]. The observed circulating upregulation of IL1R2 in this study hinted at its potential role in mitigating post-AIS inflammation.

While these markers were not brain-specific, they were intricately tied to cerebral and circulatory reactions to stroke. When interpreted within the appropriate clinical context, they could complement neuroimaging in aiding AIS management. In addition, these proteins evaluated within 24 h of stroke onset demonstrated limited correlations with infarct volume and stroke severity. This uniformity suggested their potential utility as time indicators.

The early pathophysiological responses following AIS were multifaceted[34]. The six identified proteins were closely linked to these responses. Combining these proteins could offer more comprehensive representations of these responses, thereby enhancing the specificity of ES identification. Particularly in the initial phase of AIS, some patients may not exhibit significant changes in a single protein, emphasising the importance of detecting multiple proteins to enhance the sensitivity of ES detection.

In addition to prior clinical studies on advanced neuroimaging, the molecular clock strategy presented a promising avenue for estimating AIS onset time. The only two studies in this domain included the investigation of targeted proteins in AIS patients[35] and the metabolome in rat models[9]. By directly utilising plasma samples from AIS patients and employing a MS-based high-throughput proteomic approach, our study raised the conception of molecular clock and contributed new evidence in estimating AIS onset time. This concept had been illustrated in Fig. 5.

The individual proteins or protein panels identified in our study may be proved invaluable in scenarios where patients presented with acute stroke symptoms and unclear onset times, such as in wake-up stroke. There were point-of-care testing (POCT) technologies nowadays for rapidly quantify targeted proteins with high sensitivity, such as microfluidic-based electrochemical biosensors[36]. They could achieve simplicity and affordability if these proteins were applied in clinical practice in the future. The integration of these protein biomarkers into routine diagnostic protocols may usher in a new era of precision medicine for AIS, empowering clinicians with enhanced tools to make informed treatment decisions and ultimately improve patient care and outcomes. In addition, as POCT could provide quantitative results within minutes, these proteins had potential to be widely used in settings such as emergency departments, ambulances, and outpatient clinics.

These were a series of limitations in this study. First, this was an exploratory case-control study and necessitated further external validation in larger cohorts. Second, only a single blood sample was drawn from each individual. Serial sampling would enable a better delineation of the dynamic changes of these proteins over time. Future investigations could analyse sequential samples collected at distinct time points after AIS onset. Third, our analysis solely focused on protein trends in the initial 24 h, leaving uncertainties regarding subsequent trends. Nonetheless, it was generally presumed that most stroke cases with an unknown time of onset presented within 24 h. Finally, no information on the pre-stroke dynamics of these markers was assessed. Future animal study may address this issue.

Conclusions

This study has identified six promising protein candidates and established ten panels for the early detection of AIS within 4.5 h of onset. These proteins may potentially help to improve clinical decision making regarding the timely administration of IVT for AIS patients with an unknown time of onset. This may improve patients' prognosis and recovery. Further validation studies in larger cohorts, coupled with longitudinal sampling to capture dynamic protein changes, would be essential steps towards translating these findings into clinical practice.

Data availability

The dataset generated and analysed in this study is available at the repository HKU DataHub: https://doi.org/10.25442/hku.26968519[37]. No accession code is needed. Source data for Fig. 2 can be found in 'Relative abundance of all proteins'. Source data for Figs. 3 and 4 can be found in 'The abundance of the six selected proteins, the grouping and the clinical metrics.' All other data are available from the corresponding author on reasonable request.

References

1. Tsao, C. W. et al. Heart disease and stroke statistics-2022 update: a report from the American Heart Association. *Circulation* **145**, e153–e639 (2022).
2. Emberson, J. et al. Effect of treatment delay, age, and stroke severity on the effects of intravenous thrombolysis with alteplase for acute ischaemic stroke: a meta-analysis of individual patient data from randomised trials. *Lancet* **384**, 1929–1935 (2014).
3. Thomalla, G. et al. Intravenous alteplase for stroke with unknown time of onset guided by advanced imaging: systematic review and meta-analysis of individual patient data. *Lancet* **396**, 1574–1584 (2020).
4. Kim, B. J. et al. Ischemic stroke during sleep: its association with worse early functional outcome. *Stroke* **42**, 1901–1906 (2011).
5. Muir, K. W. Treatment of wake-up stroke: stick or TWIST? *Lancet Neurol.* **22**, 102–103 (2023).
6. Simonsen, C. Z., Leslie-Mazwi, T. M. & Thomalla, G. Which Imaging Approach Should Be Used for Stroke of Unknown Time of Onset? *Stroke* **52**, 373–380 (2021).
7. Thomalla, G. et al. DWI-FLAIR mismatch for the identification of patients with acute ischaemic stroke within 4-5 h if symptom onset (PRE-FLAIR): a multicentre observational study. *Lancet Neurol.* **10**, 978–986 (2011).
8. Montaner, J. et al. Multilevel omics for the discovery of biomarkers and therapeutic targets for stroke. *Nat. Rev. Neurol.* **16**, 247–264 (2020).
9. Zhang, Y. et al. Detection of acute ischemic stroke and backtracking stroke onset time via machine learning analysis of metabolomics. *Biomed. Pharmacother.* **155**, 113641 (2022).
10. Brott, T. et al. Measurements of acute cerebral infarction: a clinical examination scale. *Stroke* **20**, 864–870 (1989).
11. Adams, H. P. Jr. et al. Classification of subtype of acute ischemic stroke. Definitions for use in a multicenter clinical trial. TOAST. Trial of Org 10172 in Acute Stroke Treatment. *Stroke* **24**, 35–41 (1993).
12. van Swieten, J. C., Koudstaal, P. J., Visser, M. C., Schouten, H. J. & van Gijn, J. Interobserver agreement for the assessment of handicap in stroke patients. *Stroke* **19**, 604–607 (1988).
13. Kothari, R. U. et al. The ABCs of measuring intracerebral hemorrhage volumes. *Stroke* **27**, 1304–1305 (1996).
14. Buderer, N. M. F. Statistical methodology.1. Incorporating the prevalence of disease into the sample size calculation for sensitivity and specificity. *Acad. Emerg. Med* **3**, 895–900 (1996).
15. Breiman, L. Random forests. *Mach. Learn* **45**, 5–32 (2001).
16. Tibshirani, R. Regression shrinkage and selection via the Lasso. *J. R. Stat. Soc. B* **58**, 267–288 (1996).
17. Bruce, L. J. et al. Absence of CD47 in protein 4.2-deficient hereditary spherocytosis in man: an interaction between the Rh complex and the band 3 complex. *Blood* **100**, 1878–1885 (2002).
18. Delaunay, J. The molecular basis of hereditary red cell membrane disorders. *Blood Rev.* **21**, 1–20 (2007).
19. The Human Protein Atlas https://www.proteinatlas.org/ENSG00000089220-PEBP1 Accessed 20 November 2024.
20. Toyoda, T. et al. Suppression of Astrocyte Lineage in Adult Hippocampal Progenitor Cells Expressing Hippocampal Cholinergic Neurostimulating Peptide Precursor in an In Vivo Ischemic Model. *Cell Transpl.* **21**, 2159–2169 (2012).
21. Su, L., Zhang, R. R., Chen, Y. Y., Ma, C. & Zhu, Z. Y. Raf Kinase Inhibitor Protein Attenuates Ischemic-Induced Microglia Cell Apoptosis and Activation Through NF-κB Pathway. *Cell Physiol. Biochem.* **41**, 1125–1134 (2017).
22. Su, L., Zhao, H. X., Zhang, X. H., Lou, Z. Y. & Dong, X. UHPLC-Q-TOF-MS based serum metabonomics revealed the metabolic perturbations of ischemic stroke and the protective effect of RKIP in rat models. *Mol. Biosyst.* **12**, 1831–1841 (2016).
23. Kim, I. et al. Endothelial cell CD36 mediates stroke-induced brain injury via BBB dysfunction and monocyte infiltration in normal and obese conditions. *J. Cerebr. Blood F. Met.* **43**, 843–855 (2023).
24. Garcia-Bonilla, L. et al. Role of microglial and endothelial CD36 in post-ischemic inflammasome activation and interleukin-1β-induced endothelial activation. *Brain Behav. Immun.* **95**, 489–501 (2021).
25. Garcia-Bonilla, L., Racchumi, G., Murphy, M., Anrather, J. & Iadecola, C. Endothelial CD36 contributes to postischemic brain injury by promoting neutrophil activation via CSF3. *J. Neurosci.* **35**, 14783–14793 (2015).
26. Zong, P. Y. et al. Activation of endothelial TRPM2 exacerbates blood-brain barrier degradation in ischemic stroke. *Cardiovasc. Res.* **120**, 188–202 (2024).
27. Cho, S. et al. The class B scavenger receptor CD36 mediates free radical production and tissue injury in cerebral ischemia. *J. Neurosci.* **25**, 2504–2512 (2005).
28. Vogel, W., Gish, G. D., Alves, F. & Pawson, T. The discoidin domain receptor tyrosine kinases are activated by collagen. *Mol. Cell* **1**, 13–23 (1997).
29. Zhu, M. X. et al. DDR1 may play a key role in destruction of the blood-brain barrier after cerebral ischemia-reperfusion. *Neurosci. Res.* **96**, 14–19 (2015).
30. Park, J. H. et al. Anti-ischemic effects of PIK3IP1 are mediated through its interactions with the ET-PI3Kγ-AKT axis. *Cells-Basel* **11** (2022).
31. DeFrances, M. C., Debelius, D. R., Cheng, J. & Kane, L. P. Inhibition of T-cell activation by PIK3IP1. *Eur. J. Immunol.* **42**, 2754–2759 (2012).
32. Jia, Y. J. et al. PIK3IP1: structure, aberration, function, and regulation in diseases. *Eur. J. Pharmacol*. **977** (2024)
33. Peters, V. A., Joesting, J. J. & Freund, G. G. IL-1 receptor 2 (IL-1R2) and its role in immune regulation. *Brain Behav. Immun.* **32**, 1–8 (2013).
34. Woodruff, T. M. et al. Pathophysiology, treatment, and animal and cellular models of human ischemic stroke. *Mol. Neurodegener.* **6**, 11 (2011).
35. Turck, N. et al. Blood Glutathione S-Transferase-pi as a Time Indicator of Stroke Onset. *PLoS ONE* **7**, e43830 (2012).
36. Islam, T., Hasan, M. M., Awal, A., Nurunnabi, M. & Ahammad, A. J. S. Metal nanoparticles for electrochemical sensing: progress and challenges in the clinical transition of point-of-care testing. *Molecules* **25** (2020).
37. Li, Q. DATA repository for: Using proteomic biomarkers to estimate acute ischemic stroke onset time. HKU Data Repository. Dataset. 2024. https://doi.org/10.25442/hku.26968519.v2 [Internet].

Author contributions

T.R. and Q.L. had the concept and designed the study. Q.L. drafted the manuscript. X.Z., Y.J., C.J., and W.F. participated in acquisition of clinical data. Q.L., Y.Z., R.P., K.L., and X.Z. participated in analysis and interpretation of the data. The funding for this study was obtained by X.Z. and W.F. J.W.J. made the conception graph. All authors participated in the revision of the article. All authors read and approved the final manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.