scientific reports

Check for updates

OPEN

Candidate biomarkers to identify mesothelioma patients at risk of developing venous thromboembolism post-surgery

Adnan Shami-shah[1,2,4,5], Shira Roth[1,2,4,5], Shad R. Morton[1,5], Bogdan Budnik[1], William G. Richards[3,4], Raphael Bueno[3,4], Rushdy Ahmad[1]✉ & David R. Walt[1,2,4]✉

Pleural Mesothelioma (PM) is an aggressive cancer that attacks thousands of people every year. One of the common treatments is surgery to remove the tumor. Unfortunately, around 11% of patients die from blood clots post-surgery. Current predictive factors such as C-reactive protein, D-Dimer, and abnormal platelet count lack specificity. To date, no blood-based protein biomarkers have been identified to reliably predict PM patients at risk of developing Venous Thromboembolism (VTE) post-surgery. In this study, we present a set of host-response plasma protein candidate biomarkers that could predict patients at risk of developing VTE. We employed a quantitative mass spectrometry-based proteomics approach integrated with a multilayered, structured, and systematic evaluation of candidate biomarkers in a cohort of 18 patients, comprising six mesothelioma cases, six mesothelioma controls, and six lung cancer controls. This is the first step towards personalized treatment plans for PM patients undergoing surgery. This study's findings can potentially guide subsequent, larger-scale investigations, highlighting the value of small-scale exploratory research.

Keywords Pleural mesothelioma, Venous thromboembolism, Biomarker discovery, Mass spectrometry, Proteomics

Venous Thromboembolism is a serious and potentially life-threatening condition that often complicates the post-operative recovery of patients with mesothelioma. In the US, approximately 2000 to 3000 new pleural mesothelioma cancer cases are diagnosed annually. The 30-day mortality rate post-surgery can reach up to 11.8%[1]. Blood clots are the main cause of death, as the incidence of VTE among patients after pleurectomy for mesothelioma patients is high (32%), and much higher compared to the incidence of VTE following other major surgeries in non-mesothelioma patients (~ < 2%)[2]. In addition, diagnosing VTE in mesothelioma patients could be challenging, as up to 33% of patients with deep vein thrombosis are asymptomatic at the time of diagnosis[3].

All Brigham and Women's Hospital, mesothelioma surgical patients undergo noninvasive vascular studies (using ultrasound) before surgery and on day seven post-surgery to identify patients at risk of developing VTE and determine who requires anticoagulants. Many other institutions administer anti-coagulants for several weeks post-surgery as a preventive action[1]. Predictive factors for developing thrombotic events include abnormal platelet count, low hemoglobin, elevated white blood cell count, and non-epithelial tumor histology[4–7]. The D-Dimer test is the conventional method used to assess the presence of blood clots. This test has a high negative predictive value (97–99%), but it decreases when combined with pretest results (89%), and its specificity is low (45–61%)[8,9]. C-reactive protein (CRP) levels are also measured for an inflammation baseline. Other biomarkers for thromboembolism exist, such as Fibrinogen and FVIII[8,10–12]. No specific blood-based protein biomarkers to identify PM patients at risk have been described to date. Identifying reliable biomarkers for predicting VTE risk in this patient population is crucial for improving clinical outcomes and tailoring individualized treatment plans. In the event a patient tests positive, they would be carefully monitored and may require a prophylactic filter to be placed prior to surgery. If they test negative, they would not need weekly monitoring.

[1]Wyss Institute for Biologically Inspired Engineering, Harvard University, Boston, MA, USA. [2]Department of Pathology, Brigham and Women's Hospital, Boston, MA, USA. [3]Division of Thoracic and Cardiac Surgery, Brigham and Women's Hospital, Boston, MA, USA. [4] Harvard Medical School, Harvard University, Boston, USA. [5] Adnan Shami-shah, Shira Roth and Shad R. Morton contributed equally to this work. ✉email: rushdy.ahmad@wyss.harvard.edu; dwalt@bwh.harvard.edu

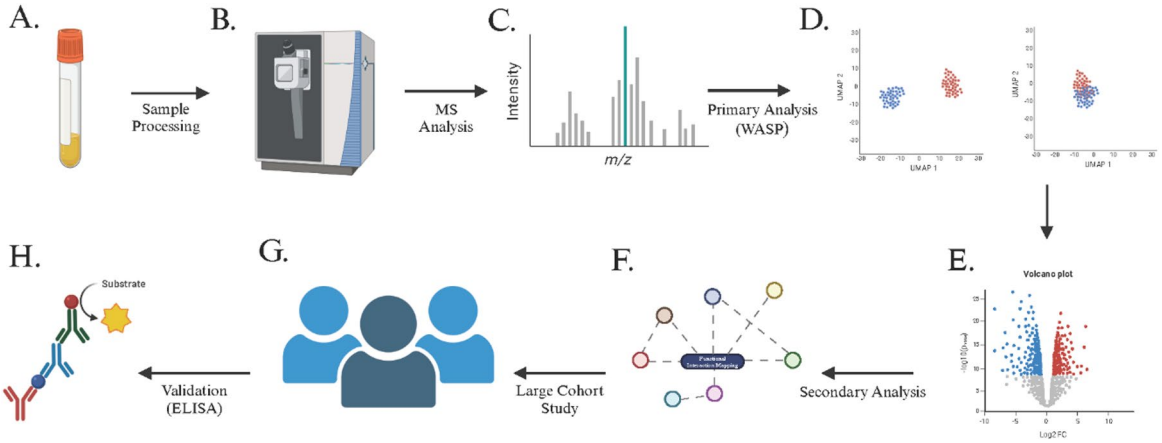


Fig. 1. A schematic diagram of the workflow used in this study for biomarker discovery. (**A**) Plasma collected before surgery from study participants is separated from whole blood. (**B**) Samples are processed for bottom-up, LC-MS/MS proteomics. (**C**) Spectra obtained from mass spectrometry analysis are searched against a spectral library, identified, and quantified. (**D**) Protein abundances are analyzed by our internal pipeline (WASP) to determine if sample groups differentiate. (**E**–**F**) Significantly differentially expressed proteins are selected for secondary analysis. (**G**) To validate the top candidate biomarkers, a larger cohort will be tested in the future (**H**) Validation of the candidate biomarkers by detecting their concentrations in the plasma of a larger cohort study and developing a classification algorithm. The figure was created with https://BioRender.com.

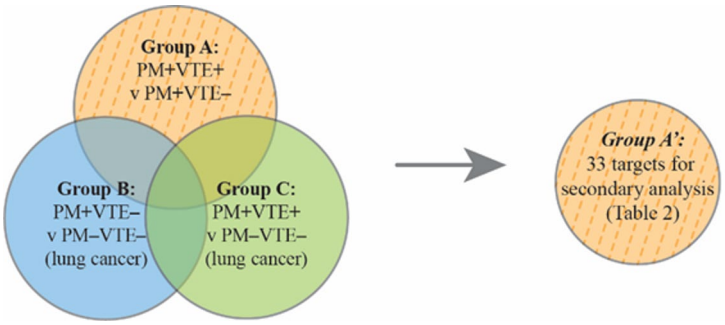


Fig. 2. Venn diagram of the three groups. Groups A, B, and C overlapped, and group A′ with 33 targets, was selected for the secondary mechanistic and structured analysis.

Here, we identify several plasma protein biomarker candidates from a carefully curated cohort of 18 patients, comprising six mesothelioma cases, six mesothelioma controls, and six lung cancer controls. We employed a quantitative mass spectrometry-based proteomics discovery approach integrated with a multilayered, structured, mechanistic, and systematic evaluation of candidate biomarkers. We demonstrate that small-scale studies with proper methodological rigor, clear clinical questions, and close collaboration with clinical experts can produce valuable information as a first step for discovering candidate biomarkers that can be used in a subsequent large-scale investigations to validate these candidates and compare them to existing predictive tests.

Methods and materials for biomarker discovery

This study integrates mass spectrometry results with pathway and mechanistic analyses (Fig. 1). As a first step, we adopt a signal differentiation approach with WASP[13] (Wyss Analysis Software for Proteomics) to identify a predictive signal in the form of protein biomarkers for downstream evaluation. The WASP proteomics tool enhances reproducibility and interpretability, is catered for 18-plex, Tandem Mass Tag (TMT)-labelled experiments (Thermo Scientific A52045), and provides precise outputs through statistical analyses for visualizing signal changes in small cohorts (Fig. 1D). We then overlapped the proteins identified in the experiment with proper controls and narrowed down a small number of targets for individual evaluation. In the case of VTE post WASP analysis, we generated an A′ group by overlapping with mesothelioma cases that did not develop VTE (PM + VTE–, control for VTE) and lung cancer cases that rarely develop VTE (PM– VTE–, control for mesothelioma cancer) (Figs. 1E and 2).

After identifying the target group of biomarkers in A', we incorporated a mechanistically informed secondary analysis to further evaluate the targets for their known associations with VTE. By identifying biomarkers

previously associated with VTE, this step functioned as a downstream verification of the targets shortlisted in A′ from our WASP-based proteomics analysis, thereby increasing confidence in the new targets revealed by this pipeline. Through this approach, we systematically evaluated each biomarker individually, assessing target pathway enrichment, disease-related mechanistic analysis, and mesothelioma stages, ensuring a comprehensive understanding of their significance (Fig. 1E). This method maximizes already available data utility, allowing for robust biomarker identification despite the limited statistical power inherent in small cohorts. Future work would include repeating steps A–E on a larger cohort (Fig. 1, F-G).

Sample collection and preparation

The 18 samples used in this study were identified and pulled from inventory by the team at Brigham and Women's Hospital (BWH) (Figure S1 and Table S1). These samples were collected pre-operatively (2010–2011), before the onset and diagnosis of VTE—which occurs post-surgery—and before any VTE-related treatment began. All human samples were collected in accordance with the Brigham and Women's Hospital (BWH) human subjects' protection policy. All patients were consented under Dana Farber Cancer Institute (DFCI) Institutional Review Board (IRB) 98 – 063. All experiments were conducted according to the approved experimental protocol of Harvard University Committee on Microbiological Safety (COMS) ID 22–051, named "Ultrasensitive Detection of Biomarkers for Disease Diagnosis".

Sample vials contain primary aliquots of plasma, labeled with secondary identifiers and barcodes. The cases are male, aged between 65 and 75 years with identical match of age and histological subtype between Venous Thromboembolism positive (VTE (+)) and paired VTE (–) pleural Mesothelioma (PM) cases on the day of surgical tumor resection. All blood samples were drawn before surgery in light blue sodium citrate tubes, which were then inverted several times and spun at 2200 rpm in a non-refrigerated centrifuge for 12 min. Plasma was pipetted from above the buffy coat layer and frozen in barcode-labeled cryovials at −80 C. The VTE (+) and VTE (–) outcomes for all patients and associated relevant data were collected and provided in an anonymized spreadsheet. It is important to note that we included six lung cancer patients who did not develop VTE as a negative control (VTE–). This allows us to provide a broader comparative framework that enhances the robustness of the biomarker discovery process. By comparing the proteomic profiles of mesothelioma patients with and without VTE to those of lung cancer patients without VTE, we can better distinguish disease-specific signals from those that might be common across different cancer types. This negative control helps ensure that the identified putative biomarkers are not only specific to the risk of VTE but also potentially specific to mesothelioma, thereby improving the specificity and relevance of the findings in the context of mesothelioma-associated thrombotic events. See SI Table 1 and SI Fig. 1 for the clinical characteristics of all patients.

Protein depletion and digestion

Plasma was depleted using the top 14 abundant protein depletion resin (Thermo Scientific) according to the manufacturer's instructions. For digestion, depleted flow-through was applied to 10 kDa filters (PALL OD010C34), reduced in 10 mM tris (2-carboxyethyl)-phosphine) (TCEP) at 25 °C for 45 min and alkalized in 10 mM iodoacetamide (IAA) in the dark at 25 °C for 20 min. The filters were saturated with 50 mM tetraethylammonium bromide (TEAB), and proteins were digested overnight at 37 °C with mass-spectrometry grade Platinum Trypsin (Promega VA9000) at a 1:50 enzyme-to-substrate ratio.

Tandem mass Tag (TMT) labeling

Peptides were labeled with TMTpro 18-plex isobaric labeling kit (Thermo Scientific A52045) according to the manufacturer's instructions, pooled, and dried on an Eppendorf Vacufuge at 60 °C and 1400 rpm under vacuum pressure until completely dry.

High-pH reversed-phase fractionation

TMT-labeled peptides were fractionated into 20 fractions using high-pH reversed-phase spin columns (Pierce 84868) according to the manufacturer's instructions.

LC-MS/MS analysis

Peptides were reconstituted in 0.1% formic acid and analyzed using an Orbitrap Exploris 240 mass spectrometer (Thermo Scientific) coupled with an EASY-nLC 1200 nanoLC system. Peptides were separated on a 150-µm × 15-cm C18 column (Bruker PepSep 1893471) with a linear gradient of 5% to 35% acetonitrile in 0.1% formic acid over 60 min at a flow rate of 300 nL/min. The mass spectrometer was operated in data-dependent acquisition mode, with a full scan range of 300-1,500 m/z at a resolution of 70,000, followed by HCD fragmentation of the top 20 most intense ions at a resolution of 17,500.

Metric	Count
Total PSMs	32,703
Unique peptides	5694
Proteins identified	593

Table 1. Summary of TMT-labelled mass spectrometry data metrics for six PM + VTE+ samples, six PM + VTE– (control) samples, and six PM–VTE– (or lung cancer) control samples.

Data processing and analysis

Raw MS data files were processed using Proteome Discoverer (Version 3.0) with the Sequest HT and Chimerys (Version 1.8) search Engine. Spectra were searched against the UniProt human proteome database (2019). Enzyme specificity was set to trypsin with a maximum of two missed cleavages. Carbamidomethylation of cysteine was set as a fixed modification, while oxidation of methionine and TMT modifications of lysine and peptide N-termini were set as variable modifications. Peptide and protein false discovery rates (FDR) were set to 1%. TMT reporter ion quantification was performed using a 20-ppm integration tolerance (Fig. 1).

Primary analysis

Quantitative, TMT-labeled proteomics data were pre-processed using the Wyss Analysis Software for Proteomics (WASP)[13] for normalization, data frame partitioning, imputation, dimensionality-reduction visualization, and differential expression analysis (Fig. 1). TMT-labeled plasma samples were normalized in silico to a pseudo-reference channel, defined as the average protein abundance across all samples. Normalization followed the median-of-ratios method implemented in pyDESeq2[14] to correct for isotopic interference of peptides co-eluted and co-isolated with target peptides[15].

After normalization, the samples were split by their diagnostic group for k-nearest neighbors (k-NN) imputation by the KNNImputer module in scikit-learn v1.7, Proteins with more than 50% missing values in both groups were removed. Dimensionality reduction for visualization was performed using Uniform Manifold Projection and Approximation (UMAP) to identify potential protein-level signal structure. Differential protein expression was assessed using pyDESeq2[14]. Significance was determined using an unadjusted, two-sided Wald p-value; proteins with p-value < 0.05 and absolute log2 fold change > 0.58 were considered differentially expressed; these cutoffs are consistent with thresholds used in previous studies (Table S2). Initial pathway enrichment analysis was performed using PantherDB[16] (Fig. 1).

Comparison between PM + VT+ and controls Post WASP analysis, three groups of comparisons (Fig. 2 and Table S2) overlapped to create a new Group A′ with thirty-three down selected targets that are found in the VTE only samples and absent in two controls (Table 2; Fig. 2):

Group A: PM + VTE+ vs. PM + VTE–.

Group B: PM + VTE+ vs. PM– VTE– (lung cancer control).

Group C: PM + VTE– vs. PM– VTE– (lung cancer control).

Group A′: Only found in VTE+ samples but absent in two controls.

We analyzed the targets in A′ (Table 2) and only used those for further secondary mechanistic and structured analysis.

Secondary analysis

Based on the comparison above, group A′ was created and 33 protein targets were identified and subject to downstream analysis. To understand prior associations of identified targets with VTE or other blood clotting disorders, targets were further analyzed on three criteria – (a) functional interaction between targets, (b) enrichment of targets in stages of mesothelioma, and (c) prior mechanistic associations with VTE. This assessment of targets enabled us to evaluate biomarkers in a comprehensive way, comparing multiple parameters and further providing confidence for the new markers identified through our pipeline (Figs. 1, 2 and 3).

(a) Functional interaction mapping

Understanding the functional connectivity of proteins is crucial for a rational discovery of biomarker candidates. We used STRING V12.0 that maps interactions based on functional association, i.e. a relationship between two proteins that contribute jointly to a specific biological function. In this case, the associated proteins do not necessarily need to interact physically[17]. We utilized the k-means clustering parameter within the STRING interface to identify enriched clusters in the entered target pool. For this study, we set a threshold of STRING strength score of > 0.8. The STRING false discovery rate describes how significant the enrichment is and is reported as a p-value corrected for multiple testing using the Benjamini-Hochberg approach. We set the FDR to $p < 0.05$ for this study. Functionally enriched targets were further assessed using the KEGG (Kyoto Encyclopedia of Genes and Genomes) pathways for associations with processes relevant to the blood clotting phenomena[18–20].

(b) Mesothelioma cancer stage analysis of targets in A' subgroup

The goal of this analysis was to identify biomarkers whose expression is unaffected by the stages of mesothelioma, suggesting minimal contribution to the marker due to cancer. By understanding stage-related variability, this approach increases confidence that any observed differential expression truly reflects VTE status and should further be investigated. The Cancer Genome Atlas (TCGA) molecularly characterized over 20,000 primary cancers and matched normal samples spanning 33 cancer types. The data generated by TCGA is publicly available. The analysis was done using the University of Alabama at Birmingham Cancer (UALCAN) data analysis Portal[21]. The portal allows the user to define the expression profile based on histological subtype, individual cancer stages, patient gender, and more. Transcripts per million (TPM) values employed for the generation of boxplots were also used to estimate the significance of the difference in gene expression levels between groups. The t-test was performed previously by UALCAN developers using a PERL script with Comprehensive Perl Archive Network (CPAN)module "Statistics::TTest" (http://search.cpan.org/~yunfang/Statistics-TTest-1.1.0/TTest.pm)[21].

Results

Mass spectrometry results

A total of 32,703 high-confidence peptide-spectrum matches (PSMs) were identified with a false-discovery rate (FDR) of less than 1%. From these PSMs, 5694 unique peptides and 593 unique proteins were identified (Table 1).

Gene names	Uniprot ID	Protein name	Log2 fold change	Fold change	95% Confidence interval
CFHR2 (FHR2)	P36980	Complement factor H-related protein 2	1.46	2.75	1.94–3.92
KNG1	P01042	Kininogen-1	1.35	2.55	1.85–3.53
BASP1	P80723	Brain acid soluble protein 1	0.94	1.92	1.51–2.43
CLEC3B	P05452	Tetranectin	0.91	1.88	1.52–2.33
VASP	P50552	Vasodilator-stimulated phosphoprotein	0.83	1.78	1.45–2.18
LMAN2	Q12907	Vesicular integral-membrane protein VIP36	0.73	1.66	1.42–1.95
ADA2 (CECR1)	Q9NZK5	Adenosine deaminase 2	0.67	1.59	1.36–1.87
RFC1	P35251	Replication factor C subunit 1	0.66	1.58	1.4–1.8
ARHGEF7	Q14155	Rho guanine nucleotide exchange factor 7	1.07	2.10	1.59–2.78
ENPP2	Q13822	Autotaxin	0.89	1.85	1.46–2.35
PCYOX1	Q9UHG3	Prenylcysteine oxidase 1	0.79	1.73	1.4–2.13
MLEC	Q14165	Malectin	0.68	1.61	1.33–1.94
PMP2	P02689	Myelin P2 protein	0.75	1.68	1.34–2.11
HLA-C	P10321	HLA class I histocompatibility antigen, C alpha chain	0.75	1.68	1.35–2.09
CKM	P06732	Creatine kinase M-type	0.63	1.55	1.29–1.85
PGM1	P36871	Phosphoglucomutase	0.62	1.53	1.27–1.85
F12	P00748	Coagulation factor XII	0.78	1.69	1.31–2.19
CAP1	Q01518	Adenylyl cyclase-associated protein 1	0.60	1.52	1.22–1.88
VNN1	O95497	Pantetheinase	−1.24	0.42	0.35–0.52
ALDOC	P09972	Fructose-biphosphate aldolase C	−0.91	0.53	0.43–0.66
C4A	P0C0l4	Complement C4-A	−0.91	0.53	0.39–0.73
COL18A1	P39060	Collagen alpha-1(XVIII) chain	−0.86	0.55	0.44–0.69
IGFBP7	Q16270	Insulin-like growth factor-binding protein 7	−0.78	0.58	0.47–0.72
FGL1	Q08830	Fibrinogen-like protein 1	−0.77	0.59	0.45–0.76
PKM	P14618	Pyruvate kinase PKM	−0.69	0.62	0.5–0.77
IGHV5-51	A0A0C4DH38	Immunoglobulin heavy variable 5–51	−0.68	0.62	0.52–0.75
SDE2 (C1ORF55)	Q6IQ49	Splicing regulator SDE2	−0.67	0.63	0.53–0.74
L1CAM	P32004	Neural cell adhesion molecule L1	−0.67	0.63	0.53–0.75
CFHR5	Q9BXR6	Complement factor H-related protein 5	−0.65	0.64	0.54–0.76
PF4V1	P10720	Platelet factor 4 variant	−0.63	0.65	0.52–0.8
ART3	Q13508	Ecto-ADP-ribosyltransferase 3	−0.62	0.65	0.52–0.8
PLXNB1	O43157	Plexin-B1	−0.60	0.66	0.57–0.76
CAPN1	P07384	Calpain-1 catalytic subunit	−0.60	0.66	0.56–0.78

Table 2. The 33 targets that were found in the VTE-only samples and absent in controls labelled as subgroup A' subjected to secondary analysis.

Pair-wise comparisons and group results

A pair-wise comparison between PM patients who developed VTE post-surgery and those who did not (Group A, Fig. 3A) identified 63 up-regulated proteins and 29 down-regulated proteins. A pair-wise comparison between PM patients who developed VTE post-surgery and lung cancer patients who did not develop VTE (Group B, Fig. 3B) identified 54 up-regulated proteins and 38 down-regulated proteins. A pairwise comparison of PM patients who did not develop VTE and lung cancer patients who did not develop VTE (Group C, Fig. 3C) identified ten up-regulated proteins and 34 down-regulated proteins (SI Table 2). Individual patient samples are separated on the peptide level according to their local structure (Fig. 3d).

After WASP analysis, 33 targets (Table 2) found in the VTE only samples and absent in controls (PM + VTE– and PM–VTE– or lung cancer) were labelled as subgroup A' (Fig. 2) and subjected to secondary analysis.

Secondary analysis for target assessment

After defining the 33 target biomarkers in subgroup A' (Table 2), we applied a mechanistically informed secondary analysis to further assess each target's known associations with VTE. By highlighting biomarkers previously linked to VTE, this step served as a downstream verification of the candidates prioritized in A' from our WASP-based proteomics workflow, thereby strengthening confidence in the newly identified targets emerging from this pipeline. The analyses from each criterion highlighting some of the biomarkers identified are detailed below:

(a) Functional interaction mapping

Protein-protein interaction (PPI) analysis of the 33 protein targets identified four probable functional PPI clusters based on a k-means clustering approach[17] (Table 2; Fig. 4). Cluster 1 identified eight targets with

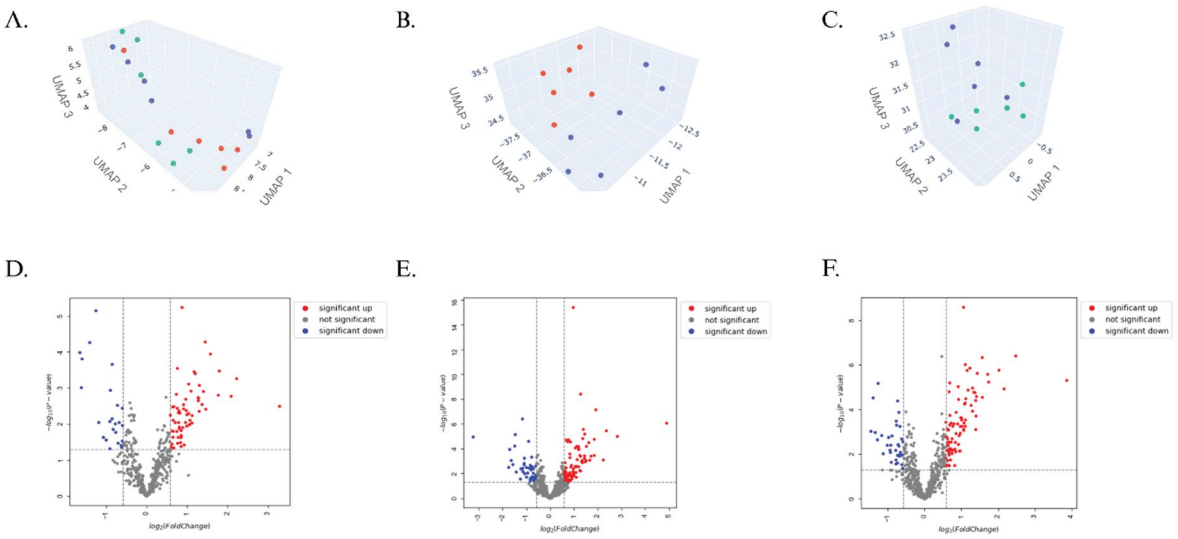
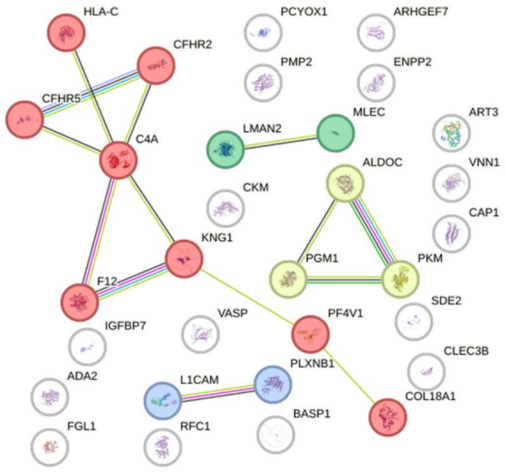


Fig. 3. Primary WASP analysis of identified peptides and protein targets. (**A**) UMAP embedding for three groups (blue: PM + VTE+, red: PM + VTE–, green: PM– VTE– (lung cancer). (**B**) Group A: UMAP embedding for PM + VTE+ (blue) vs. PM + VTE– (red). (**C**) Group C: PM + VTE+ (blue) vs. PM– VTE– (green). (**D**) Volcano plots for Group A: PM + VTE+ vs. PM + VTE– (**E**) Group B: PM + VTE+ vs. PM– VTE– and (**F**) Group C: PM + VTE– vs. PM– VTE– showing regulated and significant (P-value < 0.05) proteins. Significant proteins have Fold Change > 0.58 (log2) or < – 0.58 (log2) and p-value > 1.3 (log10). See SI Table 2 for the list of enriched proteins in all three groups.



Cluster Number	Cluster Color	Gene Count	Protein Names
1	Red	8	C4A, CFHR2, CFHR5, COL18A1, F12, HLA-C, KNG1, PF4V1
2	Yellow	3	ALDOC, PGM1, PKM
3	Green	2	LMAN2, MLEC
4	Blue	2	L1CAM, PLXNB1

Fig. 4. PPI cluster map. Four functional clusters colored in red, yellow, green, and blue were identified based on k-means clustering for 33 target entries. Proteins with no direct functional enrichments within these four clusters are shown in white.

significantly enriched biological processes (PPI enrichment $p = 7.76e^{-10}$). This enrichment suggests that the proteins identified in this network are at least partially and putatively functionally connected.

Cluster 2 identified several significant enrichments (PPI enrichment $p = 1.26e^{-4}$) of biological processes, although all were in the metabolic cascades. Cluster 3 identified two proteins that are both in the lectin family (PPI enrichment $p = 1.46e^{-3}$) with no functional enrichment, while cluster 4 identified two proteins significantly enriched (PPI enrichment $p = 1.70e^{-2}$) for processes in the axon guidance system. Of note, cluster 1 identified complement activation and blood coagulation pathways, pathways frequently associated with VTE at a mechanistic level, as two of the most enriched connections (Fig. 4).

The complement and coagulation cascades are closely interconnected due to shared synergistic roles and cross-activation between them[22]. Furthermore, we used KEGG to analyze the targets in one of the clusters – Cluster 1 for their associations with complement and coagulation pathways. Intriguingly, five (C4A, CFHR2, CFHR5, F12, and KNG1) of eight proteins identified in cluster 1 are enriched in the complement and coagulation

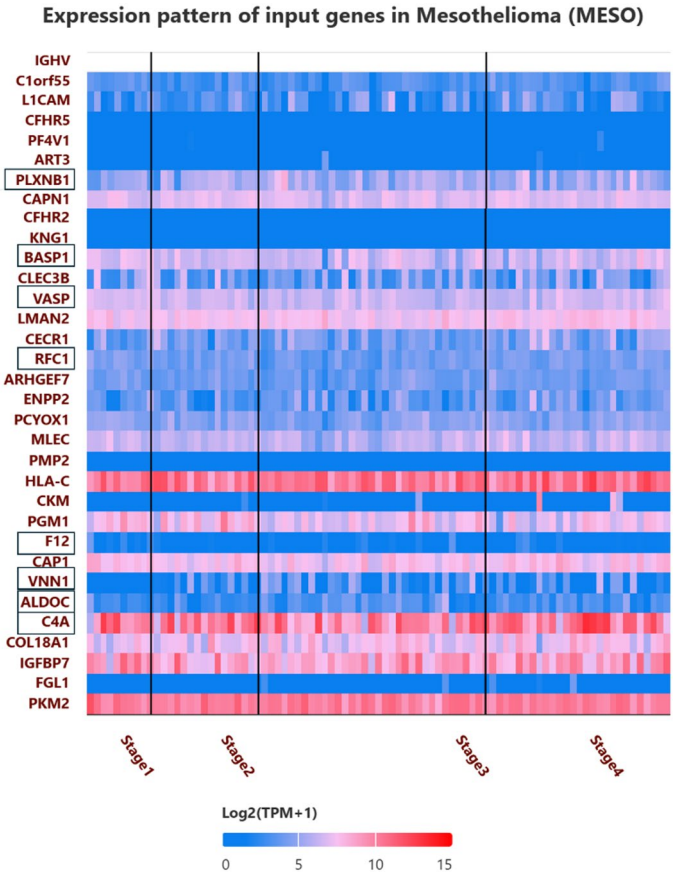


Fig. 5. shows the heatmap of the expression pattern of all the 33 genes from subgroup A' (Fig. 2) in mesothelioma according to stage. The following genes show different levels of fluctuations between cancer stages: PLXNB1, BASP1, VASP, RFC1, F12, VNN1, ALDOC, and C4A (Figure S2).

cascades (Fig. 4, Figure S1). These identified targets are also all upstream of the complement and coagulation cascades (See marked stars in Figure S1).

(b) Mesothelioma cancer stage analysis

Figure 5. Heatmap generated from TCGA data from the UALCAN portal[21] of the expression pattern of the 33 potential biomarker genes in various stages of mesothelioma. The purpose was to identify biomarkers whose expression remains constant regardless of the mesothelioma cancer stage. This approach aims to remove the influence of cancer stage on biomarker expression changes, thereby enhancing the reliability that the observed differential expression is a true indicator of VTE. Genes marked with a rectangle are the ones whose expression was significantly different between stages according to their expression profile based on individual cancer stages (Figure S2).

(c) Association between targets and VTE

Cross-referencing the identified biomarkers in subgroup A' (Fig. 2) using PubMed and running a research task using the agent "Biomarker Discovery from Omics" in Scispace.com identified several targets in our list that have prior known associations with VTE[23–39] (Table S3).

Discussion

PM patients undergoing surgery to remove the tumor have an elevated risk of developing blood clots post-surgery. To reduce these patients' mortality rate, there is an unmet need to develop a minimally invasive early blood test that will identify patients most at risk for developing blood clots after surgery. As there are no known biomarkers to predict mesothelioma patients at risk of developing blood clots, we collaborated with an expert clinician and used quantitative mass-spectrometry based untargeted proteomics to identify promising blood protein candidate biomarkers, the first step toward developing such a diagnostic test.

To analyze proteomics data generated in this project, we employed a custom-built software suite WASP. WASP has been optimized for 18-plex, TMT-labeled proteomics experiments, with tailored strategies for normalization, data quality assessment, and data cleaning. To ensure accessibility and open availability, the software is hosted on GitHub and can be run in interactive Jupyter Notebooks, complete with high-quality

annotations. WASP has been created to deliver accurate results through its carefully chosen statistical analysis and visualization tools. For example, we elected to use UMAP for dimensionality reduction to show separation between groups, as it is superior to Principal Component Analysis (PCA) at preserving local structures by using local manifold approximations and patching together their local fuzzy simplicial set representations[40] (Fig. 3).

The field of biomarker discovery often faces significant challenges due to limited sample size[41]. This can be a bottleneck for identifying and validating reliable biomarkers, thus hindering the advancement of translational research[41,42]. Carefully choosing the controls for the study and combining WASP analysis with secondary mechanistic and pathway analyses, we effectively discovered and assessed candidate biomarkers using a small cohort of VTE samples. For example, we used lung cancer patient samples as an additional negative control. We compared the proteomic signals of samples of mesothelioma patients with and without VTE to those of lung cancer patients who did not develop VTE to better differentiate VTE-specific signals from those that might be common across various cancer types. We also ensured that the VTE signal was not confused with mesothelioma stage. This analysis provided a comprehensive and broader comparative framework for a more robust analytic approach toward biomarker discovery from small cohorts. Our approach can serve as a first step in the biomarker discovery process, allowing for a go/no-go decision before investing in larger-scale biomarker discovery studies using precious resources and valuable hard-to-access clinical samples. For example, if the initial WASP analysis fails to identify distinct proteomic signals highlighting separation between the experimental and control groups, or if the overlap with mechanistic and pathway analyses does not yield a coherent set of potential biomarkers, then further downstream analysis would not be justified due to the lack of compelling evidence supporting the existence of predictive biomarkers within the small cohort.

Our secondary analysis using STRING found targets enriched in coagulation and complement cascades (cluster 1).CFHR5, one of the proteins identified from our pipeline, is a complement activation protein that causes alternative complement pathway activation. While Mendelian randomization does not necessarily demonstrate causality between CFHR5 and VTE[37], CFHR5 was identified in a large-scale proteomic profiling as a potential diagnostic/risk predictive biomarker for VTE[37], with elevated plasma levels. CFHR2 also belongs to the family of complement factor H-related genes, involved in regulation of the complement pathway. The *CFHR2* locus has been identified as a novel susceptibility locus for VTE in the recently published international effort on VTE genetics[38]. Other proteins in cluster 1, KNG1 and F12, are part of the contact system. This system activates two distinct pathways: the intrinsic pathways of coagulation and inflammation (Figure S1). KNG1 functions as a precursor that is alternatively spliced to yield high-molecular-weight kininogen (HMWK) and low-molecular-weight kininogen (LMWK). HMWK is essential for proper activation of the coagulation cascade through interaction with F12, which initiates the intrinsic coagulation pathway[43]. While further mechanistic studies are needed to understand the role of KNG1 in VTE, Wang et al. report genetic associations of KNG1 with VTE post-surgery[43]. Intriguingly, our study identified the HMWK peptide of KNG1, further bolstering the power of our study design, even though it comprised a small sample set. Although F12 was found as differentially expressed and has strong mechanistic links to VTE[23–26], we found some fluctuations in the expression pattern during different stages of mesothelioma, suggesting a likely contribution from cancer itself, as our cohort included different cancer stage samples.

Cluster 2 identified several significant enrichments in the metabolic cascades. The metabolic enzyme PKM2 from this cluster was shown to regulate platelet function and contribute to arterial thrombosis in experimental studies[33]. Limiting PKM2 dimer formation using the small molecule inhibitor ML265, reduced thrombus formation and clot retraction in vitro and in vivo[33]. Other proteins related to platelet function (not from cluster 2) are VASP and PF4V1. VASP is a protein involved in regulating the actin cytoskeleton within platelets. Its function is primarily controlled by phosphorylation, which is mediated by the cyclic AMP (cAMP) and cyclic GMP (cGMP) signaling pathways[44]. Reduced VASP function, typically indicated by a low VASP phosphorylation index, is an established biomarker associated with increased platelet activity and a higher risk of thrombotic events[32]. However, we found some fluctuations in the expression pattern during different stages of mesothelioma. PF4V1 is an inhibitor of angiogenesis[45]. It was found to be an independent predictor of VTE in patients with pancreatic adenocarcinoma[31].

Interestingly, while no functional enrichment or known prior association with VTE was observed for biomarkers such as FGL1, FGL1 is a known upstream modulator of D-dimer levels in plasma[39]. D-dimer is the current pre-operative screening test used for VTE[46]. It is possible that at a mechanistic scale, levels of FGL1 could also fluctuate together with D-dimer, and a combined test could provide better predictive power. Therefore, further investigation of targets such as FGL1 together with known VTE biomarkers could open new avenues for standard pre-operative screening tests with better predictive power.

Of note, we identified several genes – CFHR2, CFHR5, KNG1, PKM2, ENPP2, and PF4V1 with known associations to VTE[27,29,33–35,37,38,43,47] whose expression levels do not change significantly with different cancer stages. It is likely that assays developed against these targets could have the potential to distinguish true VTE signal from general cancer contribution. Although our multi-layer analysis provides a comprehensive evaluation of the identified targets, it is important to recognize that not all MS-detected proteins will ultimately serve as disease-specific biomarkers for VTE. For example, C4A emerged as a top hit based on peptide-spectrum matches (PSMs); however, it is a known broad inflammatory marker. Consistent with this finding, our TCGA analysis in mesothelioma revealed that C4A levels fluctuate across different disease stages, underscoring its lack of specificity, and with levels changing as a combination of both cancer and VTE.

The study has several limitations. First, it includes a small sample size, which may not capture the full spectrum of disease heterogeneity and does not enable further downstream testing with more specific methods like ELISA to verify and validate the potential identified biomarkers. In addition, because of the small sample size, no correction for multiple testing was made in the primary analyses. Second, STRING associations on default settings include edges between proteins and other weaker evidence for potential interactions, and therefore,

it cannot be concluded definitively that these proteins interact. Third, the current study did not compare the identified biomarkers to current pre-operative tests like D-dimer, and therefore, we cannot conclude that the identified biomarkers are more specific than current ones.

To summarize, our pipeline offers a simple yet valuable starting point, laying the groundwork for further exploration and discovery and can guide future studies. We found several promising biomarkers associated with VTE that may predict mesothelioma patients at risk of developing blood clot post-surgery. Validation with larger and more diverse cohorts will undoubtedly strengthen these findings and expand their potential impact.

Data availability

All data generated or analyzed during this study are included in this published article (and its Supplementary Information files). The Supporting Information includes a diagram showing the clinical characteristics of all patients, a table with clinical characteristics of all patients, a table with potential biomarkers associated with VTE, results of up-and down-regulated proteins, and Figures S1 and S2.

The datasets analyzed during the current study are available in the MASSIve repository, MSV000100724, MassIVE Dataset Summary.

Received: 23 January 2025; Accepted: 6 February 2026
Published online: 16 February 2026

References

1. Bueno, R., Opitz, I. & Taskforce, I. M. Surgery in malignant pleural mesothelioma. *J. Thorac. Oncol.* **13**, 1638–1654. https://doi.org/10.1016/j.jtho.2018.08.001 (2018).
2. Yhim, H. Y. et al. Incidence of venous thromboembolism following major surgery in korea: from the health insurance review and assessment service database. *J. Thromb. Haemost.* **12**, 1035–1043. https://doi.org/10.1111/jth.12611 (2014).
3. De León, L. E. et al. Routine surveillance for diagnosis of venous thromboembolism after pleurectomy for malignant pleural mesothelioma. *J. Thorac. Cardiovasc. Surg.* **160**, 1064–1073. https://doi.org/10.1016/j.jtcvs.2019.12.115 (2020).
4. Pass, H. I. Biomarkers and prognostic factors for mesothelioma. *Annals Cardiothorac. Surg.* **1**, 449–456 (2012).
5. Sugarbaker, D. J. et al. Clinical and pathological features of three-year survivors of malignant pleural mesothelioma following extrapleural pneumonectomy. *Eur. J. Cardiothorac. Surg.* **40**, 298–303. https://doi.org/10.1016/j.ejcts.2010.12.024 (2011).
6. Steele, J. P. C. Prognostic factors for mesothelioma. *Hematol. Oncol. Clin. N. Am.* **19**, 1041–1052. https://doi.org/10.1016/j.hoc.2005.09.009 (2005).
7. Edwards, J. G. et al. Prognostic factors for malignant mesothelioma in 142 patients: validation of CALGB and EORTC prognostic scoring systems. *Thorax* **55**, 731–735. https://doi.org/10.1136/thorax.55.9.731 (2000).
8. Ganesh, S. K. et al. Detection of venous thromboembolism by proteomic serum biomarkers. *PLoS ONE.* **2**, e544. https://doi.org/10.1371/journal.pone.0000544 (2007).
9. Gilens, J. F. & Khodaee, M. Does a negative d-dimer test result rule out deep-vein thrombosis? *Can. Fam Physician*. **68**, 195. https://doi.org/10.46747/cfp.6803195 (2022).
10. Dranichnikov, P., Mahteme, H., Cashin, P. H. & Graf, W. Coagulopathy and venous thromboembolic events following cytoreductive surgery and hyperthermic intraperitoneal chemotherapy. *Ann. Surg. Oncol.* **28**, 7772–7782. https://doi.org/10.1245/s10434-021-09941-9 (2021).
11. Lippi, G., Cervellin, G., Franchini, M. & Favaloro, E. J. Biochemical markers for the diagnosis of venous thromboembolism: the past, present and future. *J. Thromb. Thrombolysis*. **30**, 459–471. https://doi.org/10.1007/s11239-010-0460-x (2010).
12. Matilainen, S., Kask, G., Nieminen, J., Lassila, R. & Laitinen, M. Preoperative coagulation biomarkers associate with survival and pulmonary embolism after surgical treatment of non-spinal skeletal metastases. *Thromb. J.* https://doi.org/10.1186/s12959-022-00431-w (2022).
13. Lewandowski, M. et al. In *in Mass Spectrometry Based Single Cell Proteomics* (eds Vegvari, A., Teppo, J., Roman, A., Zubarev et al.) 157–175 (Springer US, 2024).
14. Muzellec, B., Teleńczuk, M., Cabeli, V. & Andreux, M. PyDESeq2: a python package for bulk RNA-seq differential expression analysis. *Bioinformatics* https://doi.org/10.1093/bioinformatics/btad547 (2023).
15. Iwasaki, M., Tabata, T., Kawahara, Y., Ishihama, Y. & Nakagawa, M. Removal of interference MS/MS spectra for accurate quantification in isobaric Tag-Based proteomics. *J. Proteome Res.* **18**, 2535–2544. https://doi.org/10.1021/acs.jproteome.9b00078 (2019).
16. Thomas, P. D. et al. Making genome-scale phylogenetics accessible to all. *Protein Sci.* **31**, 8–22. https://doi.org/10.1002/pro.4218 (2022).
17. Szklarczyk, D. et al. The STRING database in 2023: protein-protein association networks and functional enrichment analyses for any sequenced genome of interest. *Nucleic Acids Res.* **51**, D638–d646. https://doi.org/10.1093/nar/gkac1000 (2023).
18. Kanehisa, M. & Goto, S. KEGG: Kyoto encyclopedia of genes and genomes. *Nucleic Acids Res.* **28**, 27–30. https://doi.org/10.1093/nar/28.1.27 (2000).
19. Kanehisa, M. Toward Understanding the origin and evolution of cellular organisms. *Protein Sci.* **28**, 1947–1951. https://doi.org/10.1002/pro.3715 (2019).
20. Kanehisa, M., Furumichi, M., Sato, Y., Matsuura, Y. & Ishiguro-Watanabe, M. KEGG: biological systems database as a model of the real world. *Nucleic Acids Res.* **53**, D672–D677. https://doi.org/10.1093/nar/gkae909 (2025).
21. Chandrashekar, D. S. et al. A portal for facilitating tumor subgroup gene expression and survival analyses. *Neoplasia* **19**, 649–658. https://doi.org/10.1016/j.neo.2017.05.002 (2017).
22. Oikonomopoulou, K., Ricklin, D., Ward, P. A. & Lambris, J. D. Interactions between coagulation and complement–their role in inflammation. *Semin Immunopathol.* **34**, 151–165. https://doi.org/10.1007/s00281-011-0280-x (2012).
23. Nilsson, T., Mellbring, G. & Hedner, U. Relationship between factor XII, von Willebrand factor and postoperative deep vein thrombosis. *Acta Chir. Scand.* **152**, 347–349 (1986).
24. Nickel, K. F., Long, A. T., Fuchs, T. A., Butler, L. M. & Renné, T. Factor XII as a therapeutic target in thromboembolic and inflammatory diseases. *Arterioscler. Thromb. Vasc Biol.* **37**, 13–20. https://doi.org/10.1161/atvbaha.116.308595 (2017).
25. Maas, C. & Renné, T. Coagulation factor XII in thrombosis and inflammation. *Blood* **131**, 1903–1909. https://doi.org/10.1182/blood-2017-04-569111 (2018).
26. Kenne, E. et al. Factor XII: a novel target for safe prevention of thrombosis and inflammation. *J. Intern. Med.* **278**, 571–585. https://doi.org/10.1111/joim.12430 (2015).
27. Park, S., Gu, J. Y. & Kim, H. K. Contact system activation in disseminated intravascular coagulation: activities of prekallikrein and high-molecular-weight kininogen are significant risk factors. *J. Thromb. Thrombolysis*. **54**, 11–14. https://doi.org/10.1007/s11239-021-02598-x (2022).

28. Schmaier, A. H. The contact activation and kallikrein/kinin systems: pathophysiologic and physiologic activities. *J. Thromb. Haemost.* **14**, 28–39. https://doi.org/10.1111/jth.13194 (2016).
29. Sabater-Lleal, M. et al. A Genome-Wide association study Identifies < i>KNG1 as a genetic determinant of plasma factor XI level and activated partial thromboplastin time. *Arterioscler. Thromb. Vasc. Biol.* **32**, 2008–2016. https://doi.org/10.1161/ATVBAHA.112.248492 (2012).
30. Emsley, J., Ma, Y. & Meijers, J. C. M. Structure and interactions of the proteins from the contact system. *Blood* **146**, 1422–1430. https://doi.org/10.1182/blood.2024025322 (2025).
31. Poruk, K. E. et al. Serum platelet factor 4 is an independent predictor of survival and venous thromboembolism in patients with pancreatic adenocarcinoma. *Cancer Epidemiol. Biomarkers Prev.* **19**, 2605–2610. https://doi.org/10.1158/1055-9965.Epi-10-0178 (2010).
32. Blindt, R. et al. The significance of vasodilator-stimulated phosphoprotein for risk stratification of stent thrombosis. *Thromb. Haemost*. **98**, 1329–1334 (2007).
33. Nayak, M. K. et al. The metabolic enzyme pyruvate kinase M2 regulates platelet function and arterial thrombosis. *Blood* **137**, 1658–1668. https://doi.org/10.1182/blood.2020007140 (2021).
34. Pamuklar, Z. et al. Autotaxin/lysopholipase D and lysophosphatidic acid regulate murine hemostasis and thrombosis. *J. Biol. Chem.* **284**, 7385–7394. https://doi.org/10.1074/jbc.M807820200 (2009).
35. Han, L. et al. The binding of autotaxin to integrins mediates hyperhomocysteinemia-potentiated platelet activation and thrombosis in mice and humans. *Blood Adv.* **6**, 46–61. https://doi.org/10.1182/bloodadvances.2021004572 (2022).
36. Wang, H., Ricklin, D. & Lambris, J. D. Complement-activation fragment C4a mediates effector functions by binding as untethered agonist to protease-activated receptors 1 and 4. *Proceedings of the National Academy of Sciences* 114, 10948–10953, doi: (2017). https://doi.org/10.1073/pnas.1707364114
37. Iglesias, M. J. et al. Elevated plasma complement factor H related 5 protein is associated with venous thromboembolism. *Nat. Commun.* **14**, 3280. https://doi.org/10.1038/s41467-023-38383-y (2023).
38. Thibord, F. et al. Cross-Ancestry investigation of venous thromboembolism genomic predictors. *Circulation* **146**, 1225–1242. https://doi.org/10.1161/circulationaha.122.059675 (2022).
39. Thibord, F. et al. FGL1 as a modulator of plasma D-dimer levels: Exome-wide marker analysis of plasma tPA, PAI-1, and D-dimer. *J. Thromb. Haemost.* **19**, 2019–2028. https://doi.org/10.1111/jth.15345 (2021).
40. Leland McInnes, J. H. & Melville, J. UMAP: Uniform manifold approximation and projection for dimension reduction. doi:arXiv:1802.03426 (2020).
41. Taube, S. E. et al. A perspective on challenges and issues in biomarker development and drug and biomarker codevelopment. *J. Natl. Cancer Inst.* **101**, 1453–1463. https://doi.org/10.1093/jnci/djp334 (2009).
42. McDermott, J. E. et al. Challenges in biomarker discovery: combining expert insights with statistical analysis of complex omics data. *Expert Opin. Med. Diagn.* **7**, 37–51. https://doi.org/10.1517/17530059.2012.718329 (2013).
43. Wang, Q. et al. Genetic effects of BDKRB2 and KNG1 on deep venous thrombosis after orthopedic surgery and the potential mediator. *Sci. Rep.* **8**, 17332. https://doi.org/10.1038/s41598-018-34868-9 (2018).
44. Aszódi, A. et al. The vasodilator-stimulated phosphoprotein (VASP) is involved in cGMP- and cAMP-mediated Inhibition of agonist-induced platelet aggregation, but is dispensable for smooth muscle function. *Embo j.* **18**, 37–48. https://doi.org/10.1093/emboj/18.1.37 (1999).
45. Struyf, S., Burdick, M. D., Proost, P., Van Damme, J. & Strieter, R. M. Platelets release CXCL4L1, a nonallelic variant of the chemokine platelet factor-4/CXCL4 and potent inhibitor of angiogenesis. *Circ. Res.* **95**, 855–857. https://doi.org/10.1161/01.Res.0000146674.38319.07 (2004).
46. Pulivarthi, S. & Gurram, M. K. Effectiveness of d-dimer as a screening test for venous thromboembolism: an update. *N Am. J. Med. Sci.* **6**, 491–499. https://doi.org/10.4103/1947-2714.143278 (2014).
47. Han, S., Luo, Q., Ren, Y., Ma, X. & Fang, J. Identification of critical genes and pathways related to venous thromboembolism in non-small cell lung cancer patients using integrated bioinformatics analysis. *Res. Square*. https://doi.org/10.21203/rs.3.rs-155640/v1 (2021).

Acknowledgements

The authors thank all the researchers involved with Biomarker Discovery Lab (BDL) of the Diagnostic Accelerator (DxA) at the Wyss Institute at Harvard University. Biomarker discovery is possible because of the generosity of patients' samples, so we thank all the patients whose samples made this study possible. All human samples were collected in accordance with the Brigham and Women's Hospital (BWH) human subjects' protection policy and participants were appropriately consented to Dana Farber Cancer Institute protocol 98 – 063. This work was supported by the Wyss Diagnostics Accelerator's platform budget and Dr. Bueno's funding source at BWH.

Author contributions

AS, SR, RB, and RA conceptualized the study. Experiments were performed by AS, SR, SRM, and BB. Data analysis pipeline was developed by AS, SR, RA, and SRM. The analysis code was written by SRM. Samples were provided by WGR and RB. Study supervision and funding acquisition by RA and DRW. AS, SR, SRM, and RA contributed to the manuscript writing with feedback from all authors.

Declarations

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1038/s41598-026-39805-9.

Correspondence and requests for materials should be addressed to R.A. or D.R.W.