RESEARCH

Genomics & Informatics





Genome-wide association study-driven identification of thrombomodulin and factor V as the best biomarker combination for deep vein thrombosis

Usi Sukorini^{1,2}, Gisca Ajeng Widya Ninggar^{1,2}, Mohammad Hendra Setia Lesmana³, Lalu Irham⁴, Wirawan Adikusuma^{5,6}, Hegaria Rahmawati^{7,8}, Nur Imma Fatimah Harahap^{1,9}, Chiou-Feng Lin^{10,11} and Rahmat Dani Satria^{1,2*}

Abstract

Deep vein thrombosis (DVT) is a clinically significant condition characterized by the formation of thrombi in deep venous structures, leading to high morbidity and potential mortality. Identifying reliable biomarkers for DVT risk prediction remains challenging due to the intricate genetic and molecular mechanisms underlying the disease. This study aims to investigate the best biomarker for DVT. Our study utilized genome-wide association studies (GWAS) findings coupled with a functional annotation scoring system to identify and prioritize genetic markers with strong associations to DVT. Furthermore, gene expression levels were analyzed to determine the most promising genetic markers. Several databases were utilized, including the GWAS Catalog, HaploReg 4.2, WebGestalt, Enrichr, and the GTEx Portal. Through the comprehensive analysis, we found 5 potential biomarkers and highlighted thrombomodulin (THBD) and Factor V (F5) as the best blood-based biomarkers. THBD and F5 genes were selected based on their elevated expression levels in blood and the presence of eQTLs. Functionally, THBD modulates coagulation via protein C activation, while F5 is pivotal in thrombin formation and clot stabilization, underscoring their mechanistic relevance to DVT pathogenesis, and rendering them suitable for non-invasive clinical assessment. Our findings emphasize the potential of genetic biomarkers to transform DVT risk assessment and support advancements in precision medicine for thrombotic disorders.

Keywords Deep vein thrombosis, Biomarkers, Genome-wide association studies, Single nucleotide polymorphisms, Thrombomodulin, Factor V, THBD, F5

*Correspondence: Rahmat Dani Satria rahmat.dani.s@ugm.ac.id Full list of author information is available at the end of the article



© The Author(s) 2025. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.

1 Introduction

Deep vein thrombosis (DVT) or venous thromboembolism is a complex disorder involving the formation of blood clots in deep veins and is a leading cause of morbidity and mortality worldwide. It often precedes severe complications such as pulmonary embolism (PE) and poses significant healthcare burdens globally [1, 2]. The incidence was estimated approximately to be more than 10 million cases per year worldwide and increased with age [3, 4]. Over 1.2 million people were reported to have venous thromboembolism in the USA in 2016, and 40% presented as pulmonary embolism. The Mortality rate of a pulmonary embolism at 1 year is 19,6%, which unfortunately has been unchanged since 1999 [5].

A wide range of biomarkers has been explored for the diagnosis and prognosis of DVT. These include traditional coagulation markers such as prothrombin time (PT), activated partial thromboplastin time (APTT), and coagulation factors, as well as markers of coagulation activation like prothrombin fragment 1+2 (F1+2), thrombin-antithrombin complexes (TAT), and fibrin degradation products [6, 7]. Markers of fibrinolysis, such as plasminogen activator inhibitor-1 (PAI-1) and tissue plasminogen activator (t-PA), along with fibrinolysis activation markers like D-dimer and fibrin degradation products (FDP), provide insight into the dynamic balance between clot formation and breakdown [8]. Inhibitors such as antithrombin and other regulatory proteins further contribute to the complex hemostatic system [8, 9]. Despite the abundance of biomarkers, there is still no consensus on the optimal combination of biomarkers for DVT risk assessment.

Given these considerations, this study aims to systematically identify and validate the most promising biomarkers for deep vein thrombosis (DVT). By leveraging genome-wide association study (GWAS) datasets, functional annotation scoring, and gene expression profiling, we seek to prioritize genetic markers with strong mechanistic relevance to DVT [10]. We hope that the findings from this study will contribute to advancing global health and well-being, aligning with the United Nations' Sustainable Development Goal (SDG) 3: Good Health and Well-being, by improving the diagnosis of DVT through a biomarker-based strategy.

2 Methods

2.1 Dataset selection

Our investigation commenced by identifying genomic variants, particularly single-nucleotide polymorphisms (SNPs), that exhibit significant associations with DVT. For this purpose, we utilized data from the GWAS catalogue, applying a stringent selection criterion with a *p*-value < 5.10^{-8} to ensure the inclusion of only the

strongest associated SNPs. This rigorous threshold was crucial to narrowing our focus on variants with strong statistical significance. A database search was conducted on platform accessed on October 24, 2024, from GWAS catalog (https://www.ebi.ac.uk/gwas/), to identify SNPs related to venous thromboembolism and deep vein thrombosis trait. A summary of the research workflow is shown in Fig. 1.

2.2 Computational analysis

After extracting SNP data from the GWAS catalogue, we constructed a Manhattan plot to highlight the most significant SNPs. From this visualization, we identified the top identified genes most strongly associated with DVT. However, our focus extended beyond just these top identified genes, aiming to prioritize those with the highest biological risk potential. It is essential to recognize that not all SNPs act independently in the expression of a gene. SNPs can be influenced by neighboring variants, and there are often intricate interactions both between genes and between proteins that contribute to disease processes [10-12]. For this reason, we used functional annotation to prioritize genes, ensuring that those selected had a strong, biologically meaningful connection to DVT development.

After the initial SNP identification, we expanded our analysis by integrating additional SNPs linked to genes. To achieve this, we employed HaploReg version 4.2, a well-established tool that provides in-depth information about regulatory variants, linked genetic variants, and other relevant annotations for gene loci [13]. This tool facilitated a more comprehensive examination of the genomic landscape surrounding DVT-associated SNPs. We specifically targeted genetic data from the Asian population. This focus on a population-specific dataset allowed for a more nuanced understanding of the genetic underpinnings of DVT within this demographic, thereby improving the relevance of our findings to populations that may exhibit different genetic susceptibilities compared to other ethnic groups [14].

2.3 Functional annotations

To enhance the accuracy and reliability of our DVT risk gene identification, we implemented a scoring system based on several criteria. Our methodology builds on a foundational approach from Okada's study with our additional novel modifications to include the top DVT associated genes identified through the most significant SNPs from GWAS study. 10 We applied seven criteria for functional annotation: (1) prioritizing the top genes identified from GWAS-derived DVT-associated SNPs; (2) identifying missense mutations in genes with DVT-associated SNPs that exhibit strong linkage disequilibrium



Fig. 1 Workflow of the study for identifying blood-based biomarkers of DVT. **1** Data Extraction: SNPs associated with DVT were extracted from the EBI-NHGRI GWAS Catalog, with a Manhattan plot used to visualize the significant SNPs. A total of 689 SNPs were identified, with 458 SNPs passing the threshold of $p < 5.10^{-8}$. **2** Computational analysis: using HaploReg, SNP variants were analyzed, identifying 4430 SNP variants linked to 334 genes. **3** Functional Annotations: Genes were prioritized based on seven biological risk criteria to identify those with a strong association with DVT, resulting in 28 risk genes. **4** Enrichment and Tissue Expression: GO (Gene Ontology) analysis, network analysis, and tissue-specific gene expression highlighting THBD and F5 as the most promising blood-based biomarkers for DVT

(r2>0.80); (3) examining eQTL effects in whole blood for genes carrying DVT-associated SNPs; (4) assessing involvement in biological processes; (5) identifying associations with cellular components; (6) exploring roles in molecular functions. These last three criteria are based on Gene Ontology (GO) classifications to provide a systematic framework for function assessment. For GO, we employed the WEB-based GEne SeT AnaLysis Toolkit to focus on genes with strong functional connections to DVT risk based on GO terms. We conducted an overlap analysis with genes related to primary immunodeficiency (PID) from the 2022 International Union of Immunological Societies (IUIS) list [15], enabling us to explore potential overlaps between immunological pathways and DVT-related genes. This multi-faceted approach offers valuable insights into the genetic factors underlying DVT, with a particular focus on possible immune-related contributions to disease susceptibility.

2.4 Enrichment, network analysis, and tissue expression

We employed a hypergeometric test to assess the statistical significance of gene enrichment within our prioritized list. This test enabled us to determine whether specific genes were disproportionately represented among our DVT risk genes. A false discovery rate (FDR) threshold of 0.05 was applied to identify genes with a statistically significant overrepresentation, ensuring that our selected genes were biologically relevant and likely to contribute to DVT pathogenesis. We conducted a network analysis to further elucidate the functional roles of these genes and their associations with DVT-related pathways. We analyze gene expression profiles across a variety of tissues, providing insights into the tissue-specific expression patterns of these genes, which are critical for their potential as biomarkers.

2.5 Statistical analysis

Statistical analyses were performed using RStudio version 2024.09.0+375 (RStudio, Boston, MA, USA). The haploR package (https://cran.r-project.org/web/packa ges/haploR/index.html) was utilized to identify missense variants and Whole Blood eQTLs. Gene Ontology (GO) enrichment analyses, including biological processes, cellular components, and molecular functions, were performed using the WEB-based GEne SeT AnaLysis Toolkit (https://www.webgestalt.org). Network analysis was conducted with Enrichr (https://maayanlab.cloud/enrichrkg) to investigate gene-disease interactions. GTEx portal (https://gtexportal.org/home/) was used to analyze gene tissue expression.

3 Results

3.1 Visualizing DVT-associated SNPS: GWAS Manhattan plot analysis

The Manhattan plot illustrates the results of total 689 SNPs from a genome-wide association study (GWAS) aimed at identifying SNPs associated with DVT (Supplementary file 1). SNPs are plotted based on their chromosomal positions and association levels with DVT risk, with higher -log10 (p-value) values indicating stronger associations (Fig. 2). Notable peaks were observed at SNP loci on chromosomes 1, 4, 9, and 13 that exceed the threshold for statistical significance, suggesting their possible roles in DVT susceptibility. These areas of high association may indicate regions that contain candidate genes or regulatory elements influencing DVT development, warranting further exploration to understand the biological mechanisms involved. This analysis highlights key genomic regions that may contribute to the genetic basis of DVT, with certain SNPs showing a stronger association with the condition. The peaks on specific chromosomes may suggest genetic variations affecting pathways related to DVT pathogenesis. Identifying these SNPs provides a foundation for further studies to clarify their functions and assess their potential as biomarkers. The decision to focus on the top 40 SNPs was guided by both their visual prominence in the Manhattan plot (i.e., those appearing above the red dashed line) and the consistency of their functional annotation scores. Expanding the selection beyond the top 40 SNPs did not reveal any additional genes with a functional annotation score ≥ 3 , thereby supporting the choice of 40 SNPs as a balanced and justified cut-off for further analysis.

The list showcases the SNPs with the strongest associations and their mapped genes, based on the highest peaks observed in the Manhattan plot (Table 1). Notably, the ABO gene on chromosome 9 appears frequently, indicating a potential major role in influencing DVT risk [16]. Other key genes include F5 on chromosome 1, known for its role in coagulation processes and its established link to thrombosis. Additional genes, such as ATP1B1, NME7, and FGB, are also listed, hinting at their possible involvement in DVT. This analysis reinforces the idea that DVT has a polygenic basis, with multiple genes potentially contributing to its development. The frequent identification of coagulation-related genes aligns with the known pathophysiology of DVT, while the association with the ABO gene may point to mechanisms tied to blood type and vascular health [3, 16]. Further investigation is necessary to elucidate the biological roles of these SNPs and understand how genetic variations contribute to DVT risk.

3.2 Identification of genes associated with DVT

This study identified 458 SNPs from the GWAS catalog that met the inclusion criterion of $p < 5.10^{-8}$ out of a total of 689 SNPs (Supplementary file 1). It is important



Fig. 2 Manhattan plot of SNPs associated with DVT. 689 SNPs across chromosomes are shown with -log10 (*p*-value) on the y-axis, highlighting significant associations with DVT. Peaks on chromosomes 1, 4, 9, and 13 indicate SNPs loci that may influence DVT susceptibility. The black dashed line represents the genome-wide significance threshold ($p = 5 \times 10^{-8}$). The red dashed line indicates an additional threshold corresponding to the top 40 SNPs, based on visual prominence

Table 1 Top DVT-associated genes

No	SNP	Region	Mapped gene
1	rs1046205	13q34	MCF2L
2	rs2073826	9q34.2	ABO
3	rs10901252	9q34.2	ABO
4	rs710446	3q27.3	KNG1, HRG-AS1
5	rs6025	1q24.2	F5
6	rs3756011	4q35.2	F11
7	rs6025	1q24.2	F5
8	rs9411377	9q34.2	ABO
9	rs6025	1q24.2	F5
10	rs6025	1q24.2	F5
11	rs6025	1q24.2	F5
12	rs2227402	4q31.3	FGB
13	rs6025	1q24.2	F5
14	rs6025	1q24.2	F5
15	rs6025	1q24.2	F5
16	rs6025	1q24.2	F5
17	rs6025	1q24.2	F5
18	rs2769071	9q34.2	ABO
19	rs2769071	9q34.2	ABO
20	rs2769071	9q34.2	ABO
21	rs687289	9q34.2	ABO
22	rs687289	9q34.2	ABO
23	rs687289	9q34.2	ABO
24	rs687289	9q34.2	ABO
25	rs2519093	9q34.2	ABO
26	rs2040445	1q24.2	ATP1B1, NME7
27	rs1209731	1q24.2	NME7
28	rs145163454	1q24.2	ATP1B1
29	rs579459	9q34.2	ABO - Y_RNA
30	rs6025	1q24.2	F5
31	rs1063856	12p13.31	VWF
32	rs6025	1q24.2	F5
33	rs3756011	4q35.2	F11
34	rs2066865	4q32.1	FGA - FGG
35	rs2066865	4q32.1	FGA - FGG
36	rs28446901	9q34.2	ADAMTS13
37	rs2057655	5q31.1	IRF1-AS1, IRF1
38	rs9411377	9q34.2	ABO
39	rs579459	9q34.2	ABO - Y_RNA
40	rs7265317	20q11.22	PROCR

This table presents top genes derived from SNPs associated DVT as identified from the Manhattan plot. Highlighted among the top genes is F5 emphasizing its potential as key biomarkers for DVT.

to understand that SNPs do not always function independently in gene expression. They can be influenced by nearby SNP variants, and complex interactions frequently occur between genes and proteins, collectively contributing to disease mechanisms [17]. To ensure comprehensive coverage and minimize the possibility of missing relevant genetic variations, we expanded the initial 458 SNPs using data from the Human Genome Project via HaploReg version 4.2. This expansion applied a linkage disequilibrium threshold of $r^2 > 0.8$ in the Asian population, enabling the inclusion of additional variants that are strongly linked to the originally identified SNPs. This process yielded 4430 SNPs, which were responsible for 334 genes associated with DVT (Supplementary file 2). These genes were subsequently selected for further analysis.

3.3 Mapping biological risk genes in DVT: a functional annotation approach

To determine the biological roles of these 334 genes, we performed functional annotation using seven databases to prioritize these genes. This approach was based on the rationale that gene–disease associations may be influenced not only by gene expression but also by various other biological processes. The underlying assumption is that genes annotated in multiple functional annotations are more likely to play a crucial role in DVT development.

We used a minimum score criterion of 3 to determine genes considered to play a significant role in the pathology of DVT. Based on functional annotation criteria, we identified 28 biological risk genes associated with DVT (Table 2). Among the genes listed, IL6R, F5, and ABO stand out with high scores, totaling 6, 5, and 5, respectively. These scores indicate that these genes meet multiple functional criteria, suggesting a significant role in DVT development. The extensive annotation profile of IL6R highlights it as a key candidate in DVT research, likely due to its involvement in inflammatory signaling pathways that influence vascular function. Inflammation is closely associated with thrombosis, and IL6R's role in modulating immune responses may impact both vascular health and blood clotting dynamics, which are critical factors in DVT risk [18].

Factor V (F5) and ABO also receive high scores (5 each) and are among the top DVT- associated genes based on SNP significance. This distinction highlights their strong association with DVT-related genetic variations. Factor V, known for its role in coagulation, is particularly relevant as it directly influences blood clotting pathways-a central aspect of DVT pathology [19]. Genetic variations in F5, such as the well-known Factor V Leiden mutation, have been widely studied for their impact on clotting risk, reinforcing the gene's importance in DVT [20]. The ABO gene has been linked to DVT through blood group associations. Certain blood types are known to carry a higher risk of thrombosis, and genetic variations in ABO may contribute to this predisposition [16]. The strong annotation scores for both F5 and ABO emphasize their roles as central players in DVT pathogenesis, further justifying their prioritization in functional analysis.

Other notable genes such as F2, F9, F11, IL1RN, KLKB1, SH2B3, FGB, VWF, FGA, ADAMTS13, and THBD have scored 4, meeting several annotation criteria. These genes are often linked to pathways essential to DVT, including those governing coagulation and immune function. Supporting our findings, F9, F11, and KLKB1

		Functional Annotation							
No	Gene	Missense variant	Whole Blood eQTL	Biological process	Cellular component	Molecular function	PID 2022	Top Identified Genes from GWAS	Score
1	IL6R								6
2	F5								5
3	ABO								5
4	F11								4
5	IL1RN								4
6	THBD								4
7	SH2B3								4
8	F9								4
9	KLKB1								4
10	FGB								4
11	VWF								4
12	FGA								4
13	ADAMTS13								4
14	F2								4
15	NLRP3								3
16	PLCG2								3
17	STX2								3
18	CPS1								3
19	ATXN2								3
20	CLEC4M								3
21	ST3GAL4								3
22	PROC								3
23	MICA								3
24	CDK2AP1								3
25	PLEK								3
26	PARP10								3
27	SH2B1								3
20	EUT2								2

 Table 2
 Functional annotation scores of top DVT-associated genes

This table provides a functional annotation of the top genes associated with DVT, scored based on seven criteria: missense variants, whole blood eQTL presence, involvement in biological processes, cellular components, molecular functions, overlap with primary immunodeficiency (PID) genes, and inclusion in the top identified DVT-associated genes.

are well-known for their roles in blood coagulation pathways [21, 22] Their presence in this list underscores the importance of coagulation in DVT pathophysiology. IL1RN, SH2B3, and THBD, with known functions in immune regulation and endothelial health further emphasize the role of the immune system and vascular integrity in DVT [23]. Abnormalities in these pathways can increase clotting risk and impact vascular health, key elements in DVT development.

Overall, the table effectively organizes DVT-associated genes by their functional annotations, highlighting those with the strongest potential roles in DVT pathogenesis. By using a multi-criteria approach to prioritize genes, the analysis helps pinpoint those most relevant for further study. This table, therefore, serves as a foundational step in translating genetic findings into actionable clinical insight.

3.4 Identification coagulation-related genes

We performed enrichment analysis on the 28 genes that had a minimum functional annotation score of 3 to determine whether these genes play a significant role in the pathophysiology of DVT. The results confirmed this hypothesis, as the enrichment analysis highlighted their strong involvement in coagulation-related pathways, providing valuable insights into the genetic factors underlying DVT (Fig. 3). This analysis shows significant enrichment in coagulation-associated pathways,



Fig. 3 Gene ontology (GO) enrichment analysis of biological processes related to DVT-associated genes. This bar plot illustrates the enriched biological processes for genes associated with DVT, indicated by enrichment ratio and adjusted by false discovery rate (FDR). Notably, 15 genes involved in coagulation, regulation of response to wounding, and wound healing (FDR \leq 0.05). These processes highlight the involvement of these genes in clot formation and tissue repair mechanisms, providing valuable insights into their roles in DVT pathogenesis

suggesting that these genes primarily influence the processes of clot formation. Pathways related to regulation of response to wounding, wound healing are notably prominent, indicating that these genes may help maintain the balance between clot-promoting and clot-preventing forces in blood vessels.

Based on the significant enrichment in coagulation, regulation of response to wounding and wound healing (Fig. 3), we identified 15 genes (*F5; F11; THBD; SH2B3; F9; KLKB1; PLCG2; FGB; VWF; FGA; ADAMTS13; F2; ST3GAL4; PROC; PLEK*) for further analysis to elucidate their potential roles in DVT pathogenesis.

The prominent role of these coagulation pathways implies that many of these genes might act as regulators or modulators of thrombin and fibrin activity, both crucial to clot formation. Disruptions in these pathways could lead to excessive clotting or inadequate clot resolution, both of which are risk factors for DVT.

3.5 Uncovering five genetic drivers of DVT: insights from network analysis

To investigate the connections between these genes and DVT-related phenotypes, we selected 15 genes from the previous enrichment analysis to perform network analysis based on the Jensen Disease and Human Phenotype Ontology databases. This analysis reveals five key genes—F2, F5, PROC, F9, and THBD—with significant links to DVT, as illustrated in Fig. 4. Exploring genetic relationships and associated phenotypes, this provides a clearer understanding of the molecular basis of DVT. The network diagrams show specific interactions between these genes and pathways related to coagulation, providing insight into their functions in thrombotic conditions.

Each of these 5 genes shows a meaningful association with DVT and thrombophilia phenotypes, highlighting

their critical roles in thrombosis. F2 and F5 are fundamental to blood clotting and are directly linked to processes associated with DVT, underlining their importance in clot stability and formation [19, 24]. PROC serves as a natural anticoagulant by deactivating certain clotting factors, and alterations in this gene have been associated with a higher risk of thrombosis [25]. F9, essential in the intrinsic coagulation pathway, plays a vital role in thrombin generation [26]. Thrombomodulin (THBD), known for its dual function in promoting anticoagulation and preserving vascular health, appears linked to DVT, suggesting that disruptions in THBD could undermine these protective roles and increase the likelihood of thrombosis [27].

Identifying these genes as central elements in the DVT network highlights their value as potential biomarkers. Given their strong associations with DVT and coagulation pathways, these genes are likely valuable indicators for evaluating DVT risk and understanding broader coagulation-related issues. As biomarkers, they could play a vital role in early detection, risk evaluation, and tracking the progression or recurrence of DVT, contributing to better management and outcomes for individuals with thrombotic disorders.

3.6 F2, F5, F9, PROC, and THBD gene expression

across different tissues: focus on liver and whole blood Based on the gene expression patterns displayed for Prothrombin/Factor II (F2), Factor V (F5), Factor IX (F9), Protein C (PROC), and Thrombomodulin (THBD) in the liver and whole blood, the selection of effective biomarkers for deep DVT can be refined (Fig. 5). An ideal biomarker should not only be relevant to DVT but also detectable in a sample that is easy to collect, such as blood. From the visual data, THBD and F5 emerge as



Fig. 4 Network analysis of key genes associated with DVT and thrombophilia. This network illustrates the relationships between genes (green nodes), human phenotypes (blue nodes), and disease ontologies linked to DVT and thrombophilia. The highlighted genes *F5, F2, F9, PROC*, and *THBD* show direct connections (red lines) to DVT, underscoring their central role in coagulation pathways and thrombotic risk. This network highlights gene-phenotype and gene-disease associations that may contribute to the genetic susceptibility and pathogenesis of thrombotic disorders

top candidates due to their substantial expression in vein (whole blood) tissue, as shown by the broad distribution seen in the data. This high level of expression indicates that these genes are actively involved in coagulation processes in the bloodstream, making them well-suited for blood-based detection relevant to DVT.

In contrast, F2, F9, and PROC exhibit higher expression predominantly in liver tissue, suggesting a more localized function in that organ. However, it is also noteworthy that the expression of F2, F9, and PROC in blood varies among normal individuals, even though their average levels are lower than in liver tissue. This variation suggests that, while these genes are not expressed consistently at high levels in the blood, some degree of activity does occur in the bloodstream. Yet, due to the variability and generally lower expression in blood compared to the liver, F2, F9, and PROC may be less practical as direct blood-based biomarkers for DVT.

As shown in Table 2, there are expression quantitative trait loci (eQTLs) for THBD and F5 in whole blood,

(See figure on next page.)

Fig. 5 Gene expression profiles of DVT-associated genes in liver and whole blood. Violin plots representing the bulk tissue gene expression of five key genes associated with DVT: F9, F2, PROC, F5, and THBD. Each plot shows expression levels in liver (green) and whole blood (pink) tissues, as analyzed from GTEx data. THBD and F5 exhibit significant expression in whole blood, suggesting their potential as blood-based biomarkers for DVT. In contrast, F2, F9, and PROC show higher expression levels in the liver, indicating a more organ-specific role. The broad expression of THBD and F5 in whole blood highlights their suitability biomarker of DVT



Fig. 5 (See legend on previous page.)

reinforcing their applicability as blood-based biomarkers for DVT. In contrast, no eQTLs are observed in blood for PROC, F2, and F9, suggesting that these genes may not be as effective for direct detection in blood. This distinction further supports THBD and F5 as promising candidates for blood-based DVT biomarkers, as their genetic variants actively influence expression levels in the blood, providing a more practical basis for clinical testing.

Therefore, THBD and F5 stand out as practical biomarkers due to their robust expression in the blood, enabling easier, minimally invasive sampling and testing. Monitoring THBD and F5 levels in blood could provide a reliable approach for assessing DVT risk, offering valuable insights for clinical diagnostics and risk management in thrombotic conditions.

4 Discussion

This study contributes to the understanding of genetic factors associated with DVT by identifying key biomarkers that could enhance both risk evaluation and clinical diagnostics. Through the integration of network analysis and functional annotation, we highlighted five significant genes: F2, F5, PROC, F9, and THBD. Among these, THBD and F5 stood out as particularly suitable bloodbased biomarkers. Their substantial expression levels in blood (Fig. 5), with the presence of expression quantitative trait loci (eQTLs) in whole blood (Table 2), make them viable candidates for clinical use. The accessibility of blood samples supports non-invasive, repeatable testing, making THBD and F5 preferable for clinical application over other tissue-specific markers.

The functional roles of THBD and F5 emphasize their importance as biomarkers and elucidate their mechanisms in DVT pathophysiology. Thrombomodulin, expressed on endothelial cells, binds to thrombin, facilitating the activation of protein C (PC) into its activated form (aPC). Activated protein C acts as a natural anticoagulant by inactivating Factor Va, the active form of Factor V, thus modulating thrombin generation and preventing excessive clotting [19, 28]. This anticoagulant pathway is vital for maintaining a balance in coagulation, safeguarding against excessive clot formation that could lead to thrombosis. Therefore, genetic variants affecting THBD expression or function may disrupt this equilibrium, raising thrombotic risk [27, 29].

Factor V plays a central role in clot formation and stability. Variants in F5, such as the Factor V Leiden mutation, are associated with increased thrombotic risk due to their resistance to degradation by activated protein C [30, 31]. The functional interplay between THBD and F5 is crucial, as thrombomodulin-mediated protein C activation provides a natural mechanism to counteract Factor V's pro-coagulant effects [27]. Disruptions in this regulatory process could result in a pro-thrombotic state, underlining the relevance of both THBD and F5 in DVT pathogenesis and their suitability as biomarkers.

In contrast, while F2, PROC, and F9 are essential components of the coagulation cascade, their high expression levels in liver tissue and lack of blood-based eQTLs make them less suitable as blood biomarkers for DVT [26, 32]. Despite their significant functions in coagulation, these genes may be more applicable for liver-specific analyses rather than blood-based diagnostics. The observed variability in their blood expression among healthy individuals further limits their reliability as consistent blood biomarkers, as such fluctuations may not correlate directly with thrombotic risk (Fig. 5).

The polygenic nature of DVT is affirmed by this study, with multiple genes contributing to coagulation and thrombotic risk across various pathways. The combined use of GWAS data and functional annotation analysis allowed for the selection of biomarkers that are not only biologically significant but also practical in a clinical setting. The presence of eQTLs for THBD and F5 in blood highlights their potential application in an ELISA-based assay to detect their protein levels in plasma. Such a diagnostic tool would offer a non-invasive, cost-effective, and practical approach for early DVT detection in hospital settings.

The limitation of this study is that the potential biomarkers, F5 and THBD, may represent only a small fraction of the total genetic variation contributing to DVT. While this approach aimed to prioritize genes based on functional annotation, it is possible that other relevant genes or regulatory regions were not captured. Therefore, future studies involving more comprehensive and integrative analyses are warranted to validate the utility of THBD and F5 as biomarkers for DVT across diverse populations. Such studies should evaluate how their expression interacts with established risk factors, including age, immobility, and metabolic conditions. By integrating THBD and F5 into clinical risk models, we may improve diagnostic precision and enable more personalized approaches to DVT management.

5 Conclusion

In conclusion, this study identifies THBD and F5 as promising blood-based biomarkers for assessing DVT risk. Their strong association with DVT-related pathways, high blood expression levels, and the presence of eQTLs support their clinical applicability for non-invasive testing. This research lays a foundation for future work to bridge genomic findings with practical clinical applications in DVT.

Supplementary Information

The online version contains supplementary material available at https://doi. org/10.1186/s44342-025-00047-2.

Supplementary Material 1. The identified 458 SNPs from the GWAS catalog that met the inclusion criterion of $p < 5.10^{-8}$ out of a total of 689 SNPs

Supplementary Material 2. The process yielded 4430 SNPs, which were responsible for 334 genes associated with DVT.

Authors' contributions

US: study design, interpretation, and manuscript drafting. GAWN and HR: manuscript revision, and proofreading. WA and LI: experiment performing and data analyzing, experiment instruction. MHSL and NIFHA: data analyzing and material preparation. CFL: data analyzing and interpretation. RDS: study design, experiment instruction, data analyzing and interpretation, manuscript drafting, and resource providing. All authors read and approved the final manuscript.

Funding

This research was conducted without any external funding. No financial support was received from any funding agency, commercial entity, or institutional grant.

Data availability

No datasets were generated or analysed during the current study.

Declarations

Ethics approval and consent to participate

For research involving secondary data analysis of public datasets, consent and ethical clearance were deemed not applicable, as the data were anonymized and publicly available for research use.

Consent for publication

Not applicable.

Competing interests

The authors declare no competing interests.

Author details

¹Department of Clinical Pathology and Laboratory Medicine, Faculty of Medicine, Public Health and Nursing, Universitas Gadjah Mada, Yogyakarta 55281, Indonesia.²Integrated Laboratory Installation, Dr. Sardjito General Hospital, Yoqvakarta 55281, Indonesia. ³Department of Nursing, Faculty of Medicine, Public Health and Nursing, Universitas Gadjah Mada, Yogyakarta 55281, Indonesia. ⁴Faculty of Pharmacy, Ahmad Dahlan University, Yogyakarta, Indonesia. ⁵Research Center for Computing, Research Organization for Electronics and Informatics, National Research and Innovation Agency (BRIN, Cibinong 16911, Indonesia. ⁶Department of Pharmacy, University of Muhammadiyah Mataram, Mataram 83127, Indonesia. ⁷PKU Muhammadiyah Hospital, Gombong, Kebumen, Central Java, Indonesia.⁸Faculty of Medicine, Universitas 'Aisyiyah Yogyakarta, Yogyakarta, Indonesia.⁹Integrated Clinical Laboratory Installation, Universitas Gadjah Mada Academic Hospital, Yogyakarta 55291, Indonesia. ¹⁰Department of Microbiology and Immunology, School of Medicine, College of Medicine, Taipei Medical University, Taipei, Taiwan.¹¹Graduate Institute of Medical Sciences, Taipei Medical University, Taipei, Taiwan.

Received: 16 February 2025 Accepted: 22 April 2025 Published online: 15 May 2025

References

- Wendelboe A, Weitz JI. Global health burden of venous thromboembolism. Arterioscler Thromb Vasc Biol. 2024;44(5):1007–11.
- Raskob GE, Angchaisuksiri P, Blanco AN, Buller H, Gallus A, Hunt BJ, et al. Thrombosis: a major contributor to global disease burden. Arterioscler Thromb Vasc Biol. 2014;34(11):2363–71.

- Lo Faro V, Johansson T, Höglund J, Hadizadeh F, Johansson Å. Polygenic risk scores and risk stratification in deep vein thrombosis. Thromb Res. 2023;228(June):151–62.
- 4. Lutsey PL, Zakai NA. Epidemiology and prevention of venous thromboembolism. Nat Rev Cardiol. 2023;20(4):248–62.
- Baylis RA, Smith NL, Klarin D, Fukaya E. Epidemiology and genetics of venous thromboembolism and chronic venous disease. Circ Res. 2021;128(12):1988–2002.
- Shaw JR, Nopp S, Stavik B, Youkhana K, Michels AL, Kennes S, et al. Thrombosis, translational medicine, and biomarker research: moving the needle. J Am Heart Assoc. 2025;14(1). Available from: https://www. ahajournals.org/doi/10.1161/JAHA.124.038782.
- Lundbech M, Krag AE, Christensen TD, Hvas AM. Thrombin generation, thrombin-antithrombin complex, and prothrombin fragment F1+2 as biomarkers for hypercoagulability in cancer patients. Thromb Res. 2020;186:80–5 Available from: https://linkinghub.elsevier.com/retrieve/ pii/S0049384819305511.
- Hvas CL, Larsen JB. The fibrinolytic system and its measurement: history, current uses and future directions for diagnosis and treatment. Int J Mol Sci. 2023;24(18):14179 Available from: https://www.mdpi.com/ 1422-0067/24/18/14179.
- 9. Rezaie AR, Giri H. Anticoagulant and signaling functions of antithrombin. J Thromb Haemost. 2020;18(12):3142–53.
- Okada Y, Wu D, Trynka G, Raj T, Terao C, Ikari K, et al. Genetics of rheumatoid arthritis contributes to biology and drug discovery. Nature. 2014;506(7488):376–81.
- Bezemer ID. Gene variants associated with deep vein thrombosis. J Am Med Assoc. 2008;299(11):1306 Available from: http://jama.jaman etwork.com/article.aspx?doi=10.1001/jama.299.11.1306.
- Virolainen SJ, VonHandorf A, Viel KCMF, Weirauch MT, Kottyan LC. Gene–environment interactions and their impact on human health. Genes Immun. 2023;24(1):1–11.
- Ward LD, Kellis M. HaploReg v4: systematic mining of putative causal variants, cell types, regulators and target genes for human complex traits and disease. Nucleic Acids Res. 2016;44(D1):D877-81 Available from: https://academic.oup.com/nar/article-lookup/doi/10.1093/nar/ gkv1340.
- Žakai NA, Mcclure LA. Racial differences in venous thromboembolism. J Thromb Haemost. 2011;9(10):1877–82. https://doi.org/10.1111/j.1538-7836.2011.04443.x.
- Tangye SG, Al-Herz W, Bousfiha A, Cunningham-Rundles C, Franco JL, Holland SM, et al. Human inborn errors of immunity: 2022 update on the classification from the international union of immunological societies expert committee. Journal of Clinical Immunology. 2022;42:1473– 507. https://doi.org/10.1007/s10875-022-01289-3. Springer US.
- Christiansen MK, Larsen SB, Nyegaard M, Neergaard-Petersen S, Würtz M, Grove EL, et al. The ABO locus is associated with increased platelet aggregation in patients with stable coronary artery disease. Int J Cardiol. 2019;286(2019):152–8. https://doi.org/10.1016/j.ijcard.2019.01.090.
- Ghouse J, Tragante V, Ahlberg G, Rand SA, Jespersen JB, Leinøe EB, et al. Genome-wide meta-analysis identifies 93 risk loci and enables risk prediction equivalent to monogenic forms of venous thromboembolism. Nat Genet. 2023;55(3):399–409 Available from: https://www.nature. com/articles/s41588-022-01286-7.
- Zhang Y, Zhang Z, Wei R, Miao X, Sun S, Liang G, et al. IL (Interleukin)-6 contributes to deep vein thrombosis and is negatively regulated by miR-338-5p. Arterioscler Thromb Vasc Biol. 2020;40(2):323–34.
- Mohapatra AK, Todaro AM, Castoldi E. Factor V variants in bleeding and thrombosis. Res Practn Thrombosis and Haemostasis. 2024;8(1): 102330. https://doi.org/10.1016/j.rpth.2024.102330. Available from:
- Sokol J, Skerenova M, Ivankova J, Simurda T, Stasko J. Association of genetic variability in selected genes in patients with deep vein thrombosis and platelet hyperaggregability. Clin Appl Thromb. 2018;24(7):1027–32.
- Li Y, Bezemer ID, Rowland CM, Tong CH, Arellano AR, Catanese JJ, et al. Genetic variants associated with deep vein thrombosis: the F11 locus. J Thromb Haemost. 2009;7(11):1802–8.
- 22. Motta G, Juliano L, Chagas JR. Human plasma kallikrein: roles in coagulation, fibrinolysis, inflammation pathways, and beyond. Front Physiol. 2023;14(August):1–12.

- Devallière J, Charreau B. The adaptor Lnk (SH2B3): an emerging regulator in vascular cells and a link between immune and inflammatory signaling. Biochem Pharmacol. 2011;82(10):1391–402 Available from: https://linki nghub.elsevier.com/retrieve/pii/S0006295211004060.
- Zhang S, Taylor AK, Huang X, Luo B, Spector EB, Fang P, et al. Venous thromboembolism laboratory testing (factor V Leiden and factor II c.*97G>A), 2018 update: a technical standard of the American College of Medical Genetics and Genomics (ACMG). Genet Med. 2018;20(12):1489– 98. https://doi.org/10.1038/s41436-018-0322-z.
- Reiner AP, Carty CL, Jenny NS, Nievergelt C, Cushman M, Stearns-Kurosawa DJ, et al. PROC, PROCR and PROS1 polymorphisms, plasma anticoagulant phenotypes, and risk of cardiovascular disease and mortality in older adults: the cardiovascular health study. J Thromb Haemost. 2008;6(10):1625–32 Available from: https://linkinghub.elsevier.com/retri eve/pii/S1538783622139668.
- 26. Mackman N. Role of tissue factor in hemostasis, thrombosis, and vascular development. Arterioscler Thromb Vasc Biol. 2004;24(6):1015–22.
- Lobato RL, White WD, Mathew JP, Newman MF, Smith PK, McCants CB, et al. Thrombomodulin gene variants are associated with increased mortality after coronary artery bypass surgery in replicated analyses. Circulation. 2011;124(SUPPL. 1: S143-S148). https://doi.org/10.1161/CIRCULATIO NAHA.110.00833.
- Stone J, Hangge P, Albadawi H, Wallace A, Shamoun F, Knuttien MG, et al. Deep vein thrombosis: pathogenesis, diagnosis, and medical management. Cardiovasc Diagn Ther. 2017;7(Suppl 3):S276–84.
- Wang Q, Tang L, Hu B, Hu Y, Liu H. Thrombomodulin gene mutations associated with thromboembolic diseases. Blood. 2015;126(23):3497– 3497 Available from: https://ashpublications.org/blood/article/126/23/ 3497/90955/Thrombomodulin-Gene-Mutations-Associated-with.
- Moore GW, Castoldi E, Teruya J, Morishita E, Adcock DM. Factor V Leidenindependent activated protein C resistance: communication from the plasma coagulation inhibitors subcommittee of the International Society on Thrombosis and Haemostasis Scientific and Standardisation Committee. J Thromb Haemost. 2023;21(1):164–74. https://doi.org/10.1016/j.jtha. 2022.10.014.
- Kujovich JL. Factor V Leiden thrombophilia. Genet Med. 2011;13(1):1-16. https://doi.org/10.1097/GIM.0b013e3181faa0f2. PMID: 21116184.
- Zhang Y, Zhang Z, Shu S, Niu W, Xie W, Wan J, et al. The genetics of venous thromboembolism: a systematic review of thrombophilia families. J Thromb Thrombolysis. 2021;51(2):359–69 Available from: https://link. springer.com/10.1007/s11239-020-02203-7.

Publisher's Note

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