

RESEARCH ARTICLE

A Novel Prioritization Method in Identifying Recurrent Venous Thromboembolism-Related Genes

Jing Jiang^{1☯‡}, Wan Li^{1☯‡}, Binhua Liang^{2☯‡}, Ruiqiang Xie¹, Binbin Chen¹, Hao Huang¹, Yiran Li¹, Yuehan He¹, Junjie Lv¹, Weiming He^{3*}, Lina Chen^{1*}

1 College of Bioinformatics Science and Technology, Harbin Medical University, Harbin, Hei Longjiang Province, China, Postal code: 150081, **2** National Microbiology Laboratory, Public Health Agency of Canada, Winnipeg, Manitoba, Canada, **3** Institute of Opto-electronics, Harbin Institute of Technology, Harbin, Hei Longjiang Province, China

☯ These authors contributed equally to this work.

‡ These authors are joint first authors on this work.

* chenlina@ems.hrbmu.edu.cn (LC); hewm@hit.edu.cn (WH)



OPEN ACCESS

Citation: Jiang J, Li W, Liang B, Xie R, Chen B, Huang H, et al. (2016) A Novel Prioritization Method in Identifying Recurrent Venous Thromboembolism-Related Genes. PLoS ONE 11(4): e0153006. doi:10.1371/journal.pone.0153006

Editor: Yu Xue, Huazhong University of Science and Technology, CHINA

Received: November 29, 2015

Accepted: March 21, 2016

Published: April 6, 2016

Copyright: © 2016 Jiang et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: All profile files are available from the GEO database (accession number (s) 19151).

Funding: This work was supported in part by the National Natural Science Foundation of China (Grant No. 61272388 and 31301040); the Natural Science Foundation of Heilongjiang Province (Grant No. F201237); the Science & Technology Research Project of the Heilongjiang Ministry of Education (Grant No. 12541476); the Health Department Funds of Heilongjiang Province China (Grant No. 2012-810); and the Master Innovation Funds of Heilongjiang Province (Grant No. YJSCX2014-18HYD).

Abstract

Identifying the genes involved in venous thromboembolism (VTE) recurrence is important not only for understanding the pathogenesis but also for discovering the therapeutic targets. We proposed a novel prioritization method called Function-Interaction-Pearson (FIP) by creating gene-disease similarity scores to prioritize candidate genes underlying VTE. The scores were calculated by integrating and optimizing three types of resources including gene expression, gene ontology and protein-protein interaction. As a result, 124 out of top 200 prioritized candidate genes had been confirmed in literature, among which there were 34 antithrombotic drug targets. Compared with two well-known gene prioritization tools Endeavour and ToppNet, FIP was shown to have better performance. The approach provides a valuable alternative for drug targets discovery and disease therapy.

Introduction

Venous thromboembolism (VTE) is the third most common cardiovascular disease with a high risk of recurrence and mortality [1–5]. It was reported that around one-third of patients suffering from a first episode of deep venous thrombosis (DVT) or pulmonary embolism (PE) developed a VTE recurrence within 10 years [6]. Even during warfarin anticoagulant therapy, VTE-experienced patients still face risks of recurrent VTE [7–9]. In clinical practice, it is helpful to identify biomarkers that aid the early diagnosis of patients at a high or low risk of primary and recurrent VTE, and assess therapy [10].

In the past, efforts had been exerted on seeking these biomarker [11]. Through whole blood gene expression analysis, the D-dimer [12], the soluble p-selectin [13], and the thrombin [14] were found to be strongly associated with an increased risk of recurrent VTE and thus were accepted as biomarkers of recurrent VTE [15,16]. However, there were limitations in

Competing Interests: The authors have declared that no competing interests exist.

determining biomarkers of recurrent VTE through whole blood gene expression analysis. At first, the VTE patient population was a heterogeneous mixture of patients with provoked and non-provoked VTEs. Secondly, the two groups of VTE patients differed in the duration of time since their last VTE as well as duration of warfarin therapy. At last, some patients with a single VTE would likely be vulnerable to a recurrent event if anticoagulant therapy discontinued, resulting in reclassification of any affected individual [16]. Differential expression analysis might not determine which genes were more important or could neglect some potential disease-related genes [17].

Alternately, the computational methods such as prioritization methods, including ToppNet (<https://toppgene.cchmc.org/network.jsp>) [18] and Endeavour (<http://homes.esat.kuleuven.be/~biouser/endeavour/tool/endeavourweb.php>) [19], were deployed to investigate potential disease genes [20–22]. These methods assume that both the potential disease-related genes and the known genes share functions, interact with each other, and are involved in similar phenotypes. The studied genes were assigned similarity or confidence scores with disease followed by the ranking based on the descending order of the scores. In general, these prioritization methods rely on functional annotations [23,24], network properties [25–28] and gene expression data [29–31]. ToppNet ranks or prioritizes genes based on topological features in the protein-protein interaction network (PPIN). ToppNet has been applied with good performances in a few studies [32–34]. For example, Lascorz et al. applied ToppNet tool in identifying markers of colorectal cancer. The three overrepresenting genes was found to be closely related to the mitogen-activated protein kinase (MAPK) signaling pathways, which is well-known to increase the risk of colorectal cancer [35]. In another study by using ToppNet, the OPRM1 gene was shown to be significantly differently expressed between different HIV groups [36]. The weakness of ToppNet is only one data source used for ranking genes which affects its robustness for candidate gene identification.

Inspired by the fact that that integrative strategy in combining distinct resources showed a better performance in discovery of disease-related genes [37–46], Endeavour was developed [47]. Endeavour integrates 19 distinct data sources, including annotation (Gene Ontology, Swissprot, Interpro, Kegg, EnsemblEst), Interaction (Bind, String, BioGrid, Hprd, InNetDb, Intact, Mint), Expression (SonEtAl, SuEtAl), Pre-calculated (Ouzounis, Prospectr), Motif, Blast, and Text mining. The rankings of the candidates derived from each source were further combined into one global ranking. Robert et al. ranked the differentially expressed genes through Endeavour and identified P2rx7 (the 2nd ranked) and P2rx4 (the 3rd ranked) responsible for impaired blood pressure control in rat. The result was confirmed by Western analysis which was consistent with the previous congenic studies [48]. In Kamron et al.'s study, candidate genes of congenital cataract were prioritized using Endeavour and the three top-ranked genes were confirmed to be associated with the disease by literature [49]. The limitation of Endeavour was that it did not take disease samples into account [50]. In fact, the accuracy of prioritization methods is directly correlated with the quality of data [51]. Moreover, Endeavour solely depends on the protein interactions defined in the databases for gene prioritization. However, many protein-protein links in the databases are very loose since structural or chemical properties and functionalities were not taken into consideration, leading to reduced protein interaction reliabilities.

In this study, we present FIP (Function-Interaction-Pearson), a novel prioritization method designed for identifying Recurrent Venous Thromboembolism-related Genes. FIP addressed the limitations of the current commonly used methods in prioritizing genes. Potential VTE recurrence related genes were identified as the top-ranked genes. Our study would provide a valuable alternative for enhancing our understanding of the complex molecular mechanism of VTE recurrence at a system level.

Materials and Methods

Data Source

A gene expression profile of whole blood was downloaded from the publicly available Gene Expression Omnibus (GEO, <http://www.ncbi.nlm.nih.gov/geo/>) [52]. The profile GSE19151 in the platform GPL571 was selected for downstream analysis in this study. GSE19151 contains 13785 genes derived from the 133 samples in different groups, including normal subjects (63), single event VTE patients (32), and recurrent VTE patients (38) who are on warfarin therapy. The differentially expressed genes were identified by using the Significance Analysis of Microarrays (SAM) between normal and recurrent samples. The 119 thrombosis disease-related genes were obtained from Online Mendelian Inheritance in Man (OMIM, <http://omim.org/>) [53], Genetic Association Database (GAD, <http://geneticassociationdb.nih.gov/>) [54] and Disease Ontology (DO, <http://disease-ontology.org/>) [55]. The interaction network used in this study was downloaded from STRING (<http://string-db.org/>) [56]. In the network, gene association datasets were either directly derived from physical interactions or functional links from experimental evidence and computational methods [57,58]. The network composes of 5260 nodes (disease-related genes and differential genes in the interaction network) and 42087 edges, which represent genes and interactions between them, respectively. In our study, 108 disease-related genes (excluded 11 genes not in the STRING database and the profile GSE19151) were selected as seed genes and other genes as candidate genes.

The FIP method

A novel prioritization method FIP was developed to prioritize VTE candidate genes by calculating gene-disease similarity scores, also called disease relevance score q . Briefly, the disease relevance score for each gene was measured by considering the overall similarity with its neighboring genes in the disease-related network based on the separated data sources: gene ontology, protein-protein interaction, and gene expression. The workflow of the method and its validations were described below (Fig 1).

The score vector Q ($n \times 1$; n —the total number of genes) represented disease relevance scores for all genes in the disease-related network, which was formulated as follows:

$$Q = \frac{(1 - d)e}{I - dW^T D^{-1}} \quad (1)$$

where q_i ($q_i \in Q$) is the ranking score of gene i , I denotes an identity matrix of $n \times n$, e is the expression score vector of $n \times 1$ where e_i is defined as the absolute value of the difference between the sum of expression values of gene i in normal and recurrent samples, d denotes a control parameter in the range of $[0,1]$ which is to adjust the weight of disease-related network in calculating ranking scores (here we chose $d = 0.9$ [59]), and D corresponds to a diagonal matrix of $n \times n$ where d_{ii} is the sum of weights of interactions between gene i and its neighboring genes in the network. The weights are contained in the matrix W ($n \times n$), where w_{ij} is used to measure the overall similarity between gene i and its neighbor gene j from the aspects of interaction, expression and function. Thus, W was characterized as:

$$W = \alpha \cdot S(i, j) + \beta \cdot P(i, j) + \gamma \cdot F(i, j) \quad (2)$$

Here $S(i,j)$, $P(i,j)$ and $F(i,j)$ denote the interaction credibility score, Pearson correlation coefficient, and shared functional significance score between gene i and gene j , respectively. Three coefficients α , β , and γ in the range of $[0,1]$ were used to assess the importance of $S(i,j)$, $P(i,j)$, and $F(i,j)$ in formula, respectively.

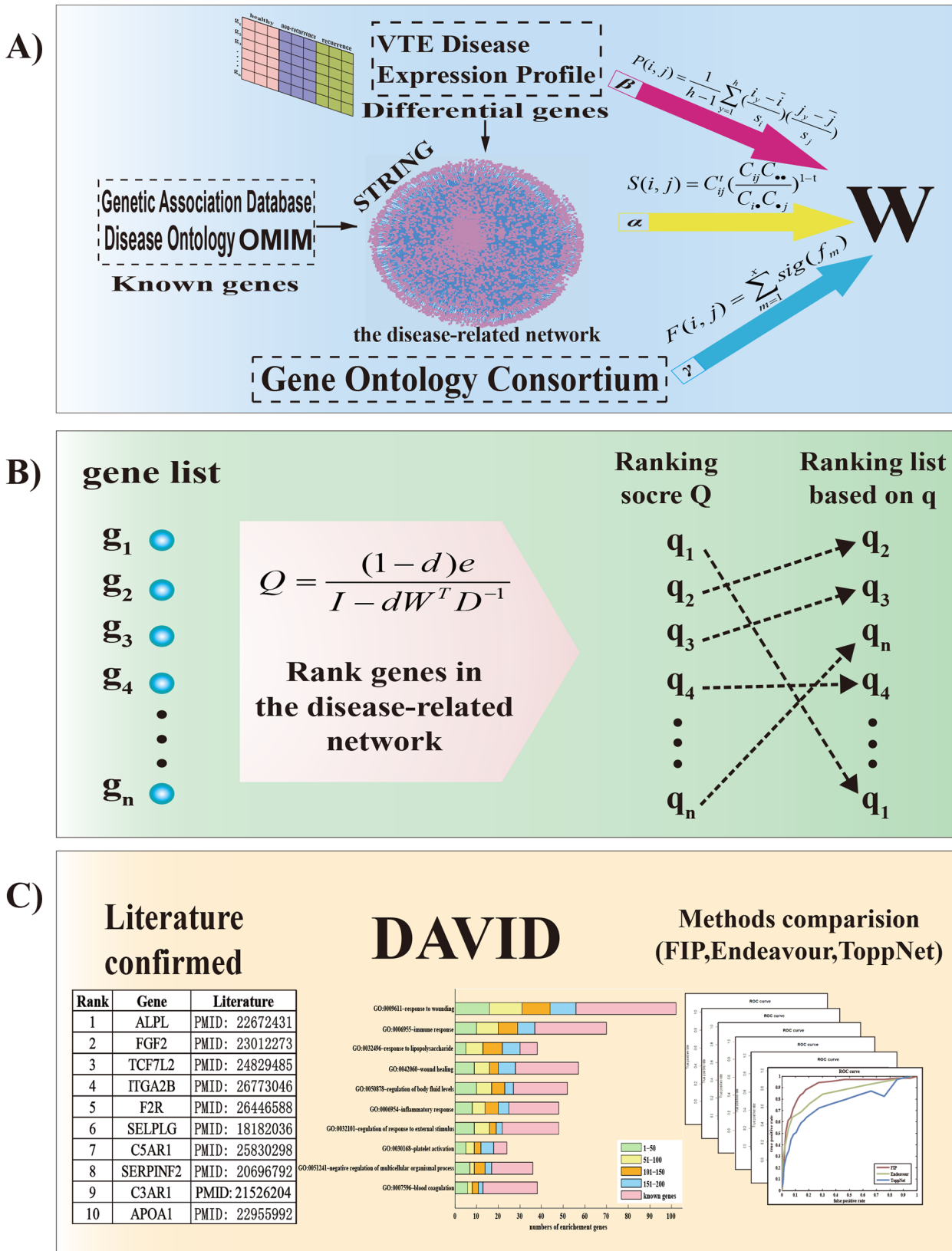


Fig 1. The workflow of FIP method and its validations. A: measurement of the overall similarity between genes. B: calculation of ranking scores of candidate genes. C: verification of the performance of the results.

doi:10.1371/journal.pone.0153006.g001

The interaction credibility scores $S(i, j)$ for each pair of gene i and j was calculated as follows [58]:

$$S(i, j) = C_{ij}^t \left(\frac{C_{ij} C_{..}}{C_{i.} C_{.j}} \right)^{1-t} \tag{3}$$

where $C_{i.}$ and $C_{.j}$ are the sums over all pairs involving i or j and another entity, $C_{..}$ is the sum over all pairs of entities, C_{ij} represents the sums over all pairs involving both i and j , and $t = 0.6$ [58]. The parameters were optimized on the KEGG benchmark set [58]. The co-occurrence score C_{ij} was defined as:

$$C_{ij} = \sum_{k=1}^b (\delta_{dijk} v_d + \delta_{pijk} v_p + \delta_{sijk} v_s) \tag{4}$$

where $v_d = 1$, $v_p = 2$, and $v_s = 0.2$ are the weights for co-occurrence genes within the same document, paragraph, and sentence based on literature mining, respectively. The delta functions δ_{dijk} , δ_{pijk} , and δ_{sijk} are 1 if the genes i and j are both mentioned in the document k , a paragraph of k or a sentence of k , otherwise they are 0 [58].

The Pearson correlation coefficient $P(i, j)$, which is used to represent the co-expression relationship between gene i and gene j , was defined as follows:

$$P(i, j) = \frac{1}{h-1} \sum_{y=1}^h \left(\frac{i_y - \bar{i}}{s_i} \right) \left(\frac{j_y - \bar{j}}{s_j} \right) \tag{5}$$

where h is the number of normal samples adding recurrent samples in the expression profile, \bar{i} , \bar{j} , s_i , s_j , i_y and j_y represent the average expression value of normal and recurrent samples, standard deviation and observed values of i and j , respectively.

For the shared functional significance score $F(i, j)$ between gene i and gene j , one function was represented by one GO term f_m . $F(i, j)$ is defined as the total sum of the significance of the functions shared:

$$F(i, j) = \sum_{m=1}^x sig(f_m) \tag{6}$$

where x is the number of common GO terms annotated by genes i and j , and $sig(f_m)$ denotes the significance of a function f_m , which was defined as follows:

$$sig(f_m) = \frac{1}{|Gene(f_m)|} \tag{7}$$

here $Gene(f_m)$ is genes annotated on GO term f_m , $|Gene(f_m)|$ is the number of genes annotated to f_m . We calculated the ranking score q for each gene in the disease-related network and ranked these genes in the descending order of q .

In the formula (1) and (2), all the combinations of α , β , γ , and d were used to rank candidate genes. The best α , β , γ combination was determined according to the seed genes identified in the top 50 and 100 ranking list. The best d value was selected based on the α , β , and γ combination which showed the best performance in ranking candidate genes.

Validation

The comparison of FIP with ToppNet/Endeavour was carried out using the same data. The performance of them was assessed using the Leave-One-Out Cross Validation (LOOCV). For

all the seed genes, one seed gene was removed as a test gene each time, and then added to candidate genes. All the candidate genes were ranked by our method to determine the ranking of the test gene. This procedure was repeated until all the seed genes were used up as test genes. Receiver Operating Characteristic (ROC) curves were then plotted and the area under ROC curve (AUC) values were used to compare the performances of the three methods.

Results

Optimization of ranking coefficient parameters

As described in method, score vector Q for all the genes was calculated based on the rankings from the separated data sources such as gene ontology, protein-protein interaction, and gene expression in their corresponding coefficients α , β , and γ , respectively. Candidate genes, which were the common genes in the disease-related network and the differentially expressed genes identified using SAM, were then ranked in the descending order of Q value. For the top 50 and 100 genes in the ranking list, we calculated the number of matched seeds against single, two and three these parameter combination, respectively. There was a significant difference between single and multiple parameter combinations both in top 50 and 100, as well as between two and three parameter combination in top 50 of ranking gene list (t-test, $p < 0.05$) (Fig 2).

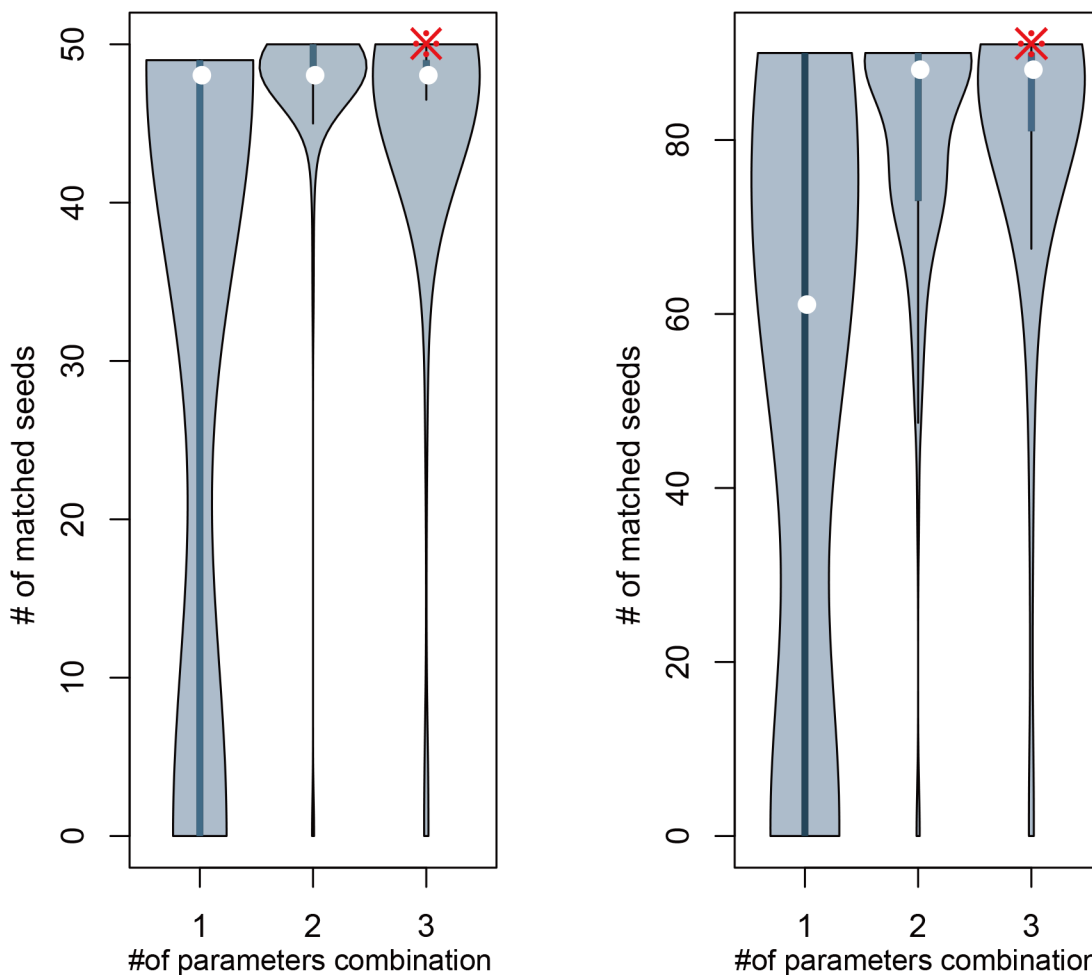


Fig 2. Violin plots of the number of matched seeds identified in top 50 and 100 of the ranking list. The number 1, 2 and 3 represent the number of parameter(s) in the parameter combinations, respectively. The asterisks and white circles present our results and the medians of each combination.

doi:10.1371/journal.pone.0153006.g002

The LOOCV has been further applied for all parameter combinations and four parameter combinations ($\alpha = 0.8, \beta = 0.5, \gamma = 0.9$ (AUC = 0.9107); $\alpha = 0.7, \beta = 0.5, \gamma = 0.8$ (AUC = 0.9187); $\alpha = 0.9, \beta = 0.5, \gamma = 0.8$ (AUC = 0.8955); $\alpha = 0.9, \beta = 0.6, \gamma = 0.8$ (AUC = 0.8763)) were shown to be better than the rest. Since no other independent dataset of VTE could be obtained, 10-fold cross-validation was carried out to further select the optimized parameter values in Formulas 1 and 2 from these four parameter combinations ($\alpha = 0.8, \beta = 0.5, \gamma = 0.9$ (AUC = 0.9013); $\alpha = 0.7, \beta = 0.5, \gamma = 0.8$ (AUC = 0.8948); $\alpha = 0.9, \beta = 0.5, \gamma = 0.8$ (AUC = 0.8527); $\alpha = 0.9, \beta = 0.6, \gamma = 0.8$ (AUC = 0.8416)).

The optimal parameter combination of $\alpha = 0.8, \beta = 0.5,$ and $\gamma = 0.9,$ was achieved (Fig 3).

For all parameter combinations, genes were also ranked according to the calculated q scores with five different d values (d = 0.1, 0.3, 0.5, 0.7, and 0.9). The matching numbers of genes were applied to assess the effectiveness of FIP (Fig 4). The number of matched seeds among top 500 in the ranking list of d = 0.9 was higher than those of other d-values.

Finally, the parameter combination of $\alpha = 0.8, \beta = 0.5, \gamma = 0.9,$ and d = 0.9 was selected to calculate vector Q so as to obtain the ranking results.

Prioritization of candidate genes and validation by literature review

In the disease-related network, all the genes were prioritized by FIP according to vector Q in the optimal ranking coefficient parameter combination. As a result, a total 200 of top candidate genes were generated (S1 Table). We manually searched these top 200 candidate genes for drug targets in literature of PubMed (<http://www.ncbi.nlm.nih.gov/pubmed>). There were 34 antithrombotic drug targets among the top 200 candidates, including thrombin-activated factor 2 receptor (F2R; rank 5), SELPLG (rank 6), APOA1 (rank 10), SCARB1 (rank 17), TTR (rank 30), and F10 (rank 37) (S1 Table). Thrombin-activated factor 2 receptor (F2R) was reported to link thrombosis to inflammation modulating interleukin 6 (IL6) synthesis [60,61]. Treatment of rats with APOA1 Milano (the mutant form of human APOA1) was shown to markedly delay thrombus formation, inhibit platelet aggregation, and to reduce weight of the thrombus [62]. FX protein was encoded by gene F10, and its mutations gave rise to severe Factor X (FX) deficiency. Anti-FX inhibitor had been approved by FDA for the prevention of venous thromboembolism surgical intervention and as an initial treatment for deep venous thrombosis and pulmonary embolism [63–65].

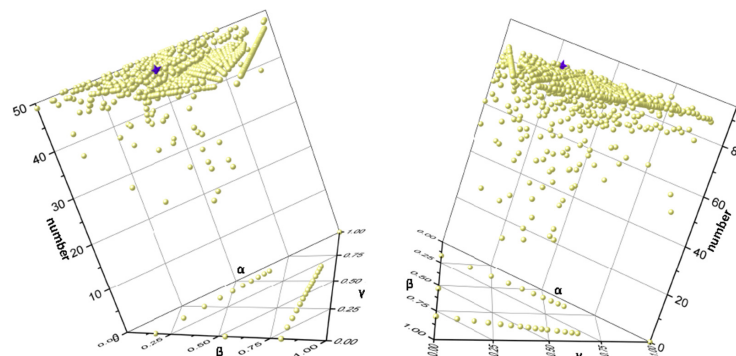


Fig 3. The 3-D distribution of seed genes in top 50 (left) and top 100 (right) of the ranking gene list against all parameter combinations. The three sides of the triangle coordinate system represent the three parameters, respectively. The perpendicular axis of the triangular coordinate system represents the number of seed genes. The purple five-pointed star and yellow ball present the optimal parameter combination and all other parameter combinations.

doi:10.1371/journal.pone.0153006.g003

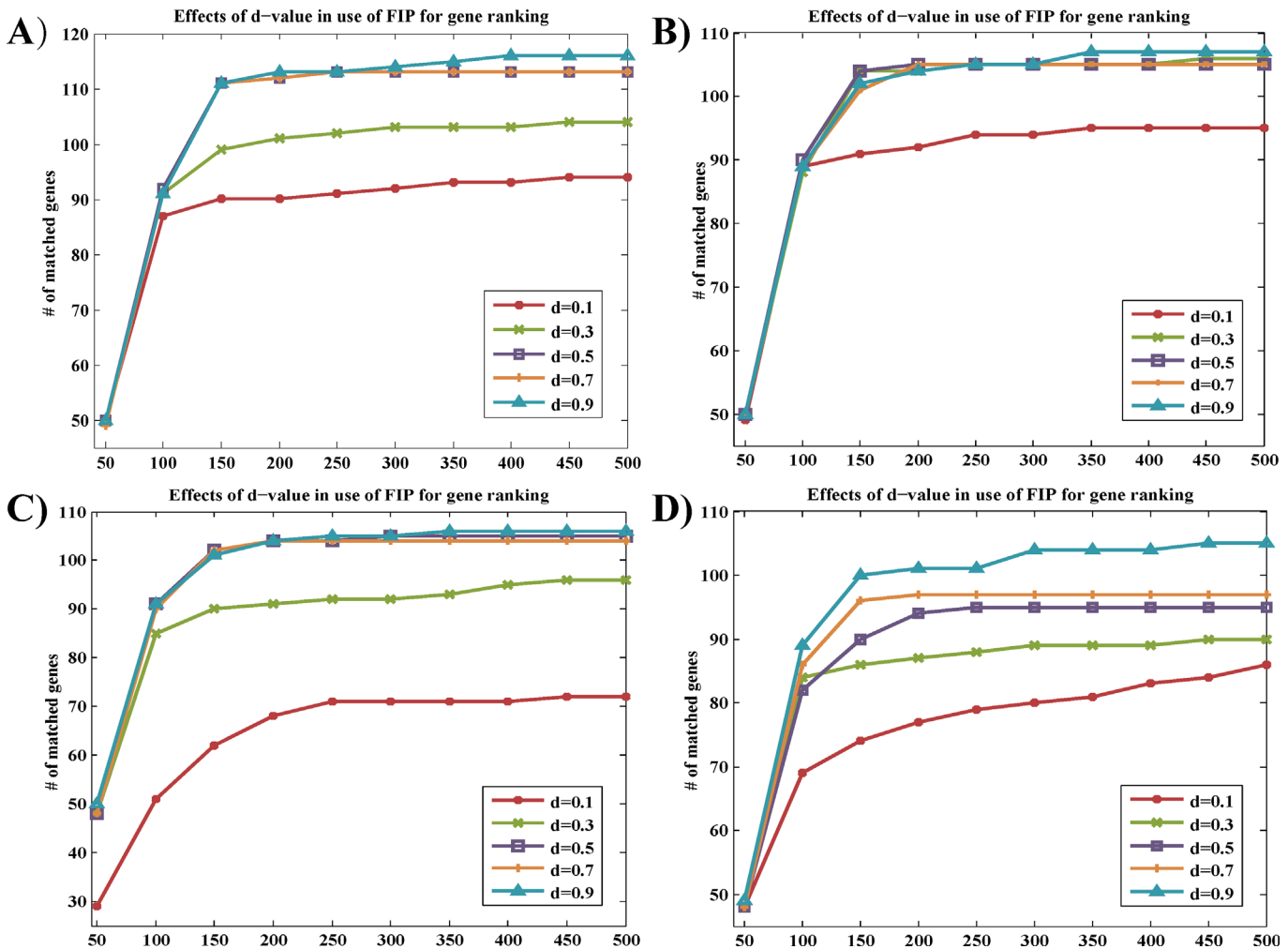


Fig 4. Comparison of performance of FIP-based gene ranking with different d values for four parameter combinations. The number of matched genes identified using FIP with five different d values (d = 0.1, 0.3, 0.5, 0.7, and 0.9) for four parameter combinations were counted and plotted ((A) $\alpha = 0.8$, $\beta = 0.5$, $\gamma = 0.9$; (B) $\alpha = 0.7$, $\beta = 0.5$, $\gamma = 0.8$; (C) $\alpha = 0.9$, $\beta = 0.5$, $\gamma = 0.8$; (D) $\alpha = 0.9$, $\beta = 0.6$, $\gamma = 0.8$). Y-axis: the number of matched genes identified using FIP; X-axis: the number of ranked genes.

doi:10.1371/journal.pone.0153006.g004

Non-drug target candidate genes of the top 200 candidates were also reported to be associated with thrombosis. For instance, SNP could be used in the prediction of recurrent thrombosis such as susceptibility gene ALPL (rank 1) with SNP [66,67]. The coagulation factor III gene (F3; rank 11) was suggested to produce tissue factor, which could initiate thrombosis on disrupted atherosclerotic plaques [68]. The loss of CYP2C19 (rank 22) function triggered platelet reactivity, which was a predictor of stent thrombosis [69,70]. Variation of VTN (rank 28) promoter haplotype, causing transcription factor binding activity increased, was proposed to be a novel genetic marker for deep venous thrombosis [71]. Sex hormone-binding globulin (SHBG; rank 51), easily measured in routine laboratories, could serve as a marker for the risk of venous thrombosis [72].

Taken together, of the top 200 candidate genes in the ranking list, 124 candidate genes predicted by our method had been confirmed to be correlated with thrombosis in PubMed literature, which have not been recorded in disease databases (OMIM, GAD and DO) (S1 Table).

Table 1. The confirmation rate of top 200 candidate genes in the ranking list.

Top n	Confirmation Number	Confirmation Rate
10	10	100.00%
20	16	80.00%
40	30	75.00%
50	35	70.00%
100	66	66.00%
150	100	66.67%
200	124	62.00%

The confirmation rate was calculated by dividing the confirmation number by the corresponding number of top n. It represented the effectiveness of the confirmation.

doi:10.1371/journal.pone.0153006.t001

Top-ranked candidates were found to have a high confirmation rate in terms of their association with thrombosis, especially top 10 candidates (Table 1).

Validation of FIP through Functional and pathway analysis

We conducted DAVID (<http://david.abcc.ncifcrf.gov/>) [73] and Gene Ontology (GO, <http://geneontology.org/>) (Biological Process and Molecular Function) [74] analysis to assess the functional enrichment of the identified candidate genes. In this way, the biological features/or meanings of the candidate genes can be extracted in order to improve the classification of these genes in terms of their functionalities. The classification was further interpreted in KEGG (<http://www.genome.jp/kegg/>) [75] pathways (FDR <0.05). Top 200 candidates were selected and divided into four groups with 50 genes in each, followed by KEGG and GO analysis in DAVID. As a result, 10 significant functional categories were identified and associated with thrombotic disease (Fig 5) [76–83]. For instance, GO: 0007596~blood coagulation was reported to be the main cause of thrombosis and recurrence. Blood coagulation, causing damage to the vascular endothelium, was suggested to initiate acute venous thrombus generation [84]. The maximum number of candidate and seed genes were found in GO: 0009611~respond to wounding functional category. The most common sites of wounding in conflict were extremities, which were associated with a significant incidence of vascular trauma, and had a high complication rate (graft thrombosis) [85]. ‘GO: 0030168~platelet activation’, leading to severe end-organ damage, was shown to increase the risk of thrombosis, implying that platelet reactivity was an important pathological mechanism of thrombosis [86,87].

We counted the number of the candidate and seed genes among the 10 functional categories which each gene was annotated to. Ten candidate genes (95% confidence interval) appeared in more than 8 functional categories and were confirmed by literature (Fig 6).

Moreover, 6 of these candidates were drug targets, and 3 of them were at top 50 candidate genes (Table 2).

Furthermore, the known disease-related genes and top 200 candidate genes were obviously enriched in four common pathways: Hematopoietic cell lineage, cytokine-cytokine receptor interaction, Cell adhesion molecules (CAMs) and complement and coagulation cascades pathway (FDR<0.05). The coagulation cascade pathway appeared to be a critical determinant of atherosclerotic plaque thrombogenicity [88]. Cell adhesion molecules (CAMs), hematopoietic cell lineage and cytokine-cytokine receptor interaction were also associated with thrombosis [85,89,90]. We mapped the enriched genes, including the known disease-related genes and candidate genes, in the coagulation cascades pathway [91,92] (Fig 7).

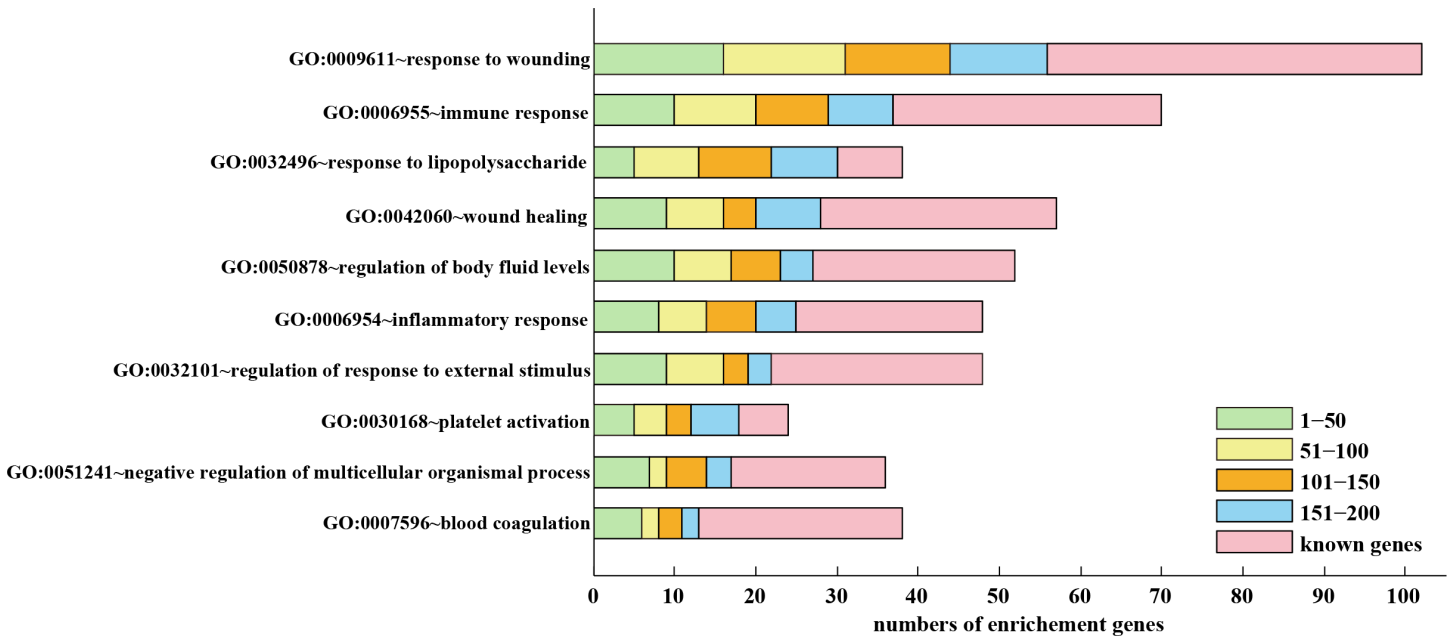


Fig 5. The top 200 candidate genes and known genes involved in the identified 10 functional categories. The genes were analyzed in GO and KEGG with DAVID and classified into 10 VTE-related functional categories.

doi:10.1371/journal.pone.0153006.g005

In the map, there are 19 known genes and 9 candidate genes, respectively. Among these 9 candidate genes, each of them was annotated to no less than four functional categories, especially F2R, SERPINF2, and A2M, which were annotated to more than eight functional categories (Fig 6). Triggering tissue factor (F3) and F2R (coagulation pathway sensors) have been shown to influence the vascular microenvironment and angiogenesis respective of clinically apparent thrombosis [93,94]. The mutations of other two genes, PROC and PROS1, were shown to increase risk of recurrent thromboembolic events if they were combined with other genetic or environmental thrombosis factors [95,96]. A2M was reported to inhibit the known

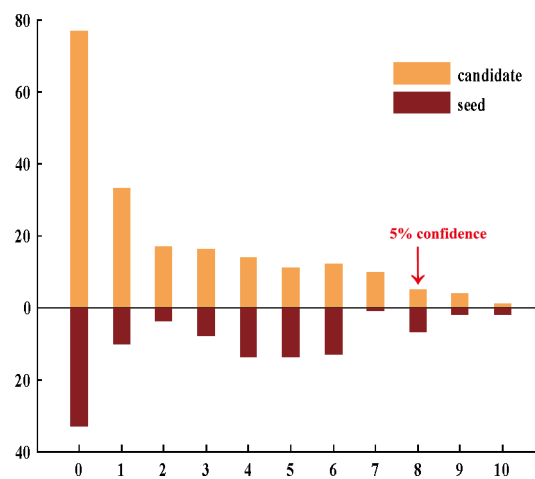


Fig 6. The distribution of the top 200 candidates and seeds among different functional categories. X-axis and y-axis represent the number of the functional categories and enriched genes, respectively.

doi:10.1371/journal.pone.0153006.g006

Table 2. The candidate genes in more than eight functional categories.

Gene	Frequency	Rank	Drug	PubMedID
TLR4	10	105	✓	PMID:24488914
APOA1	9	10	✓	PMID:18801202
CX3CL1	9	191		PMID:25795074
F2R	9	5	✓	PMID:19404549
THBS1	9	57	✓	PMID:25343959
A2M	8	41		PMID:20156641
F2RL1	8	31		PMID:12069753
SERPINF2	8	8		PMID:20696792
SYK	8	143		PMID:24376657
VEGFA	8	45	✓	PMID:25006132

doi:10.1371/journal.pone.0153006.t002

genes PROS1 and PROC in the coagulation cascade pathway, which could be associated with recurrent thrombosis.

Comparison of FIP to ToppNet and Endeavour

To evaluate the performance of the proposed FIP method in predicting novel recurrent thrombosis genes by prioritizing candidate genes, we carried out LOOCV on the known disease-

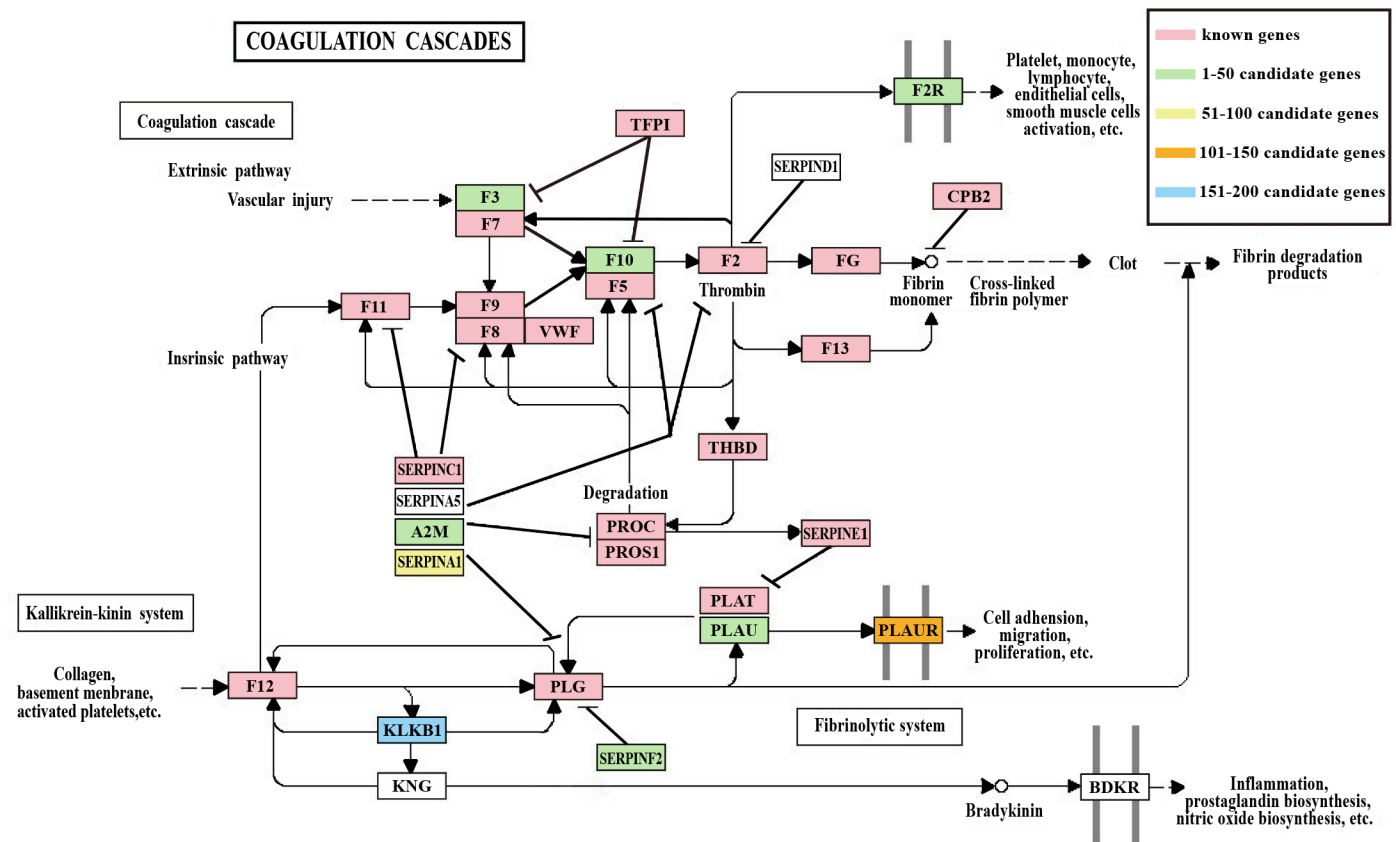


Fig 7. Coagulation cascades pathway. The red, green, pink, orange and blue rectangles present known genes in the top 1–50, top 51–100, top 101–150 and top 151–200 candidate genes.

doi:10.1371/journal.pone.0153006.g007

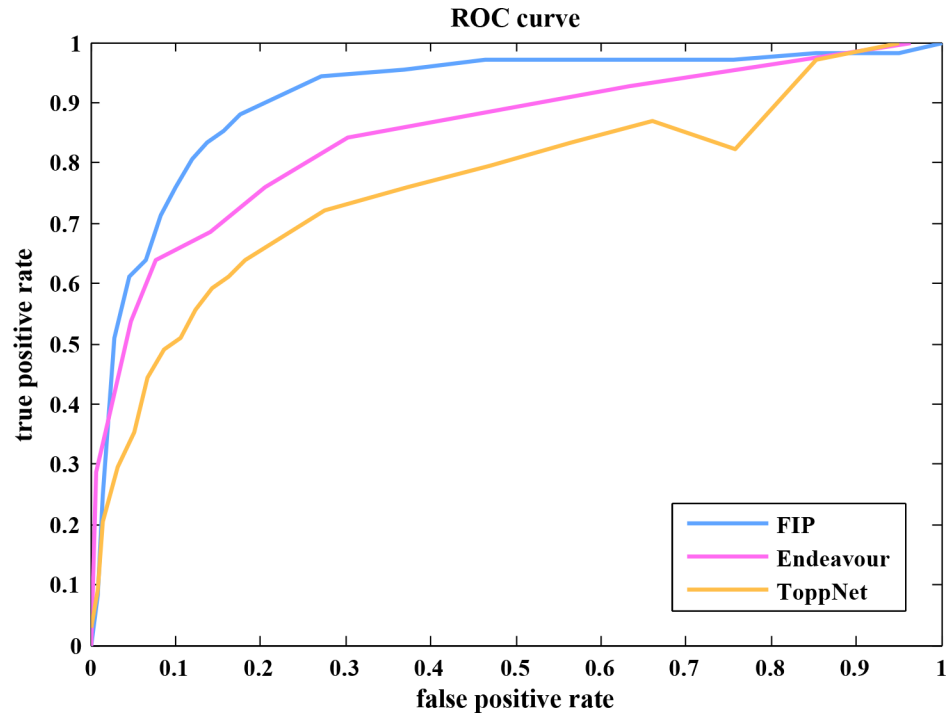


Fig 8. The ROC curves of FIP, ToppNet, and Endeavour methods.

doi:10.1371/journal.pone.0153006.g008

related genes. In this validation, the same training and testing gene sets were used in the FIP, ToppNet, and Endeavour methods. The ROC curves were plotted to compare the performance of the three methods (Fig 8).

The AUC value of FIP method was 0.9107, which was much higher than ToppNet (0.7150) and Endeavour (0.8127). Thus, FIP method provided a good performance in efficiently identifying known disease-related genes in the prioritization gene list and was more sensitive and specific in ranking the test genes.

To further verify the top-ranked candidates as novel disease recurrence genes, support vector machine (SVM) was applied to classify normal and recurrent samples with top-ranked candidates as classification characteristics. The outcome of FIP was then compared with those of ToppNet and Endeavour methods with the top 50 and top 100 candidates as classification characteristics, respectively. Four performance measurements, false positive rate (FPR), true positive rate (TPR), best cutoff curve, and AUC, were calculated (Fig 9).

The AUC values of FIP were higher than those of ToppNet and Endeavour methods using either the top 50 or 100 candidates as classification characteristics. In the meantime, the AUC values of each method using the top 50 candidates as characteristics were higher than those of each corresponding method using top 100 candidates as characteristics.

To explore the factors which may affect the performance of FIP, we first assessed the correlation between specific expression profile and outcome of gene prioritization. $P(i,j)$ in formula (2) were assigned randomly from all correlation coefficients using sampling with and without replacement, respectively. The disease relevance score Q was recalculated and genes were ranked according to the q value. The seed numbers in the top 50 and top 100 ranking list were calculated to evaluate the performance of our method. Each process was repeated 100 times. The results showed that the performance of our method was better than that of random

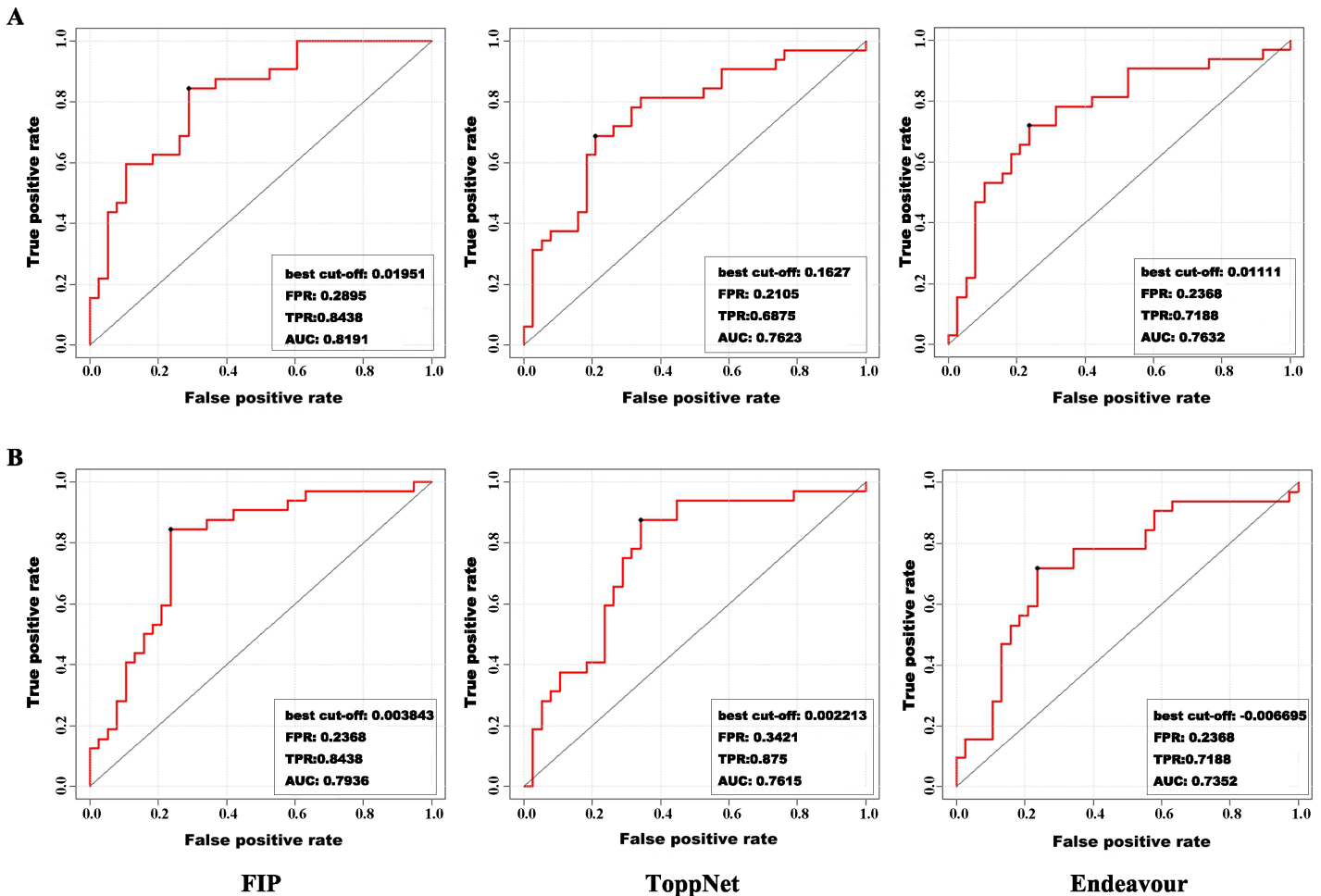


Fig 9. The performance of sample classification with top-ranked candidates of FIP, ToppNet, and Endeavour methods. (A) ROC curves of FIP, ToppNet, and Endeavour methods with the top 50 candidates as classification characteristics (B) ROC curves of FIP, ToppNet, and Endeavour methods with the top 100 candidates as classification characteristics.

doi:10.1371/journal.pone.0153006.g009

sampling of specific expression data (Fig 10). It suggested that specific expression profile did affect the performance of gene prioritization methods.

Secondly, we evaluated the importance of protein interaction reliability. We altered the $S(i, j)$ in formula (2) to 1 (no protein interaction reliability) and recalculated the disease relevance score Q . Genes were ranked according to the q value. LOOCV was used to assess the performance using the new weights. Its AUC (0.6878) was lower than that of the original weights (AUC = 0.9107) (Fig 11).

To evaluate the robustness of FIP, 10-fold CV was also applied to ToppNet and Endeavour. There was a statistical significance between FIP and ToppNet (one-side t-test, p -value < 0.05) as well as FIP and Endeavour (one-side t-test, p -value < 0.05) (S1 Fig).

We performed literature validation, function annotation and pathway analysis for top 200 candidates of ToppNet and Endeavour (S2–S4 Figs). In general, the performances of FIP were better than those of ToppNet and Endeavour.

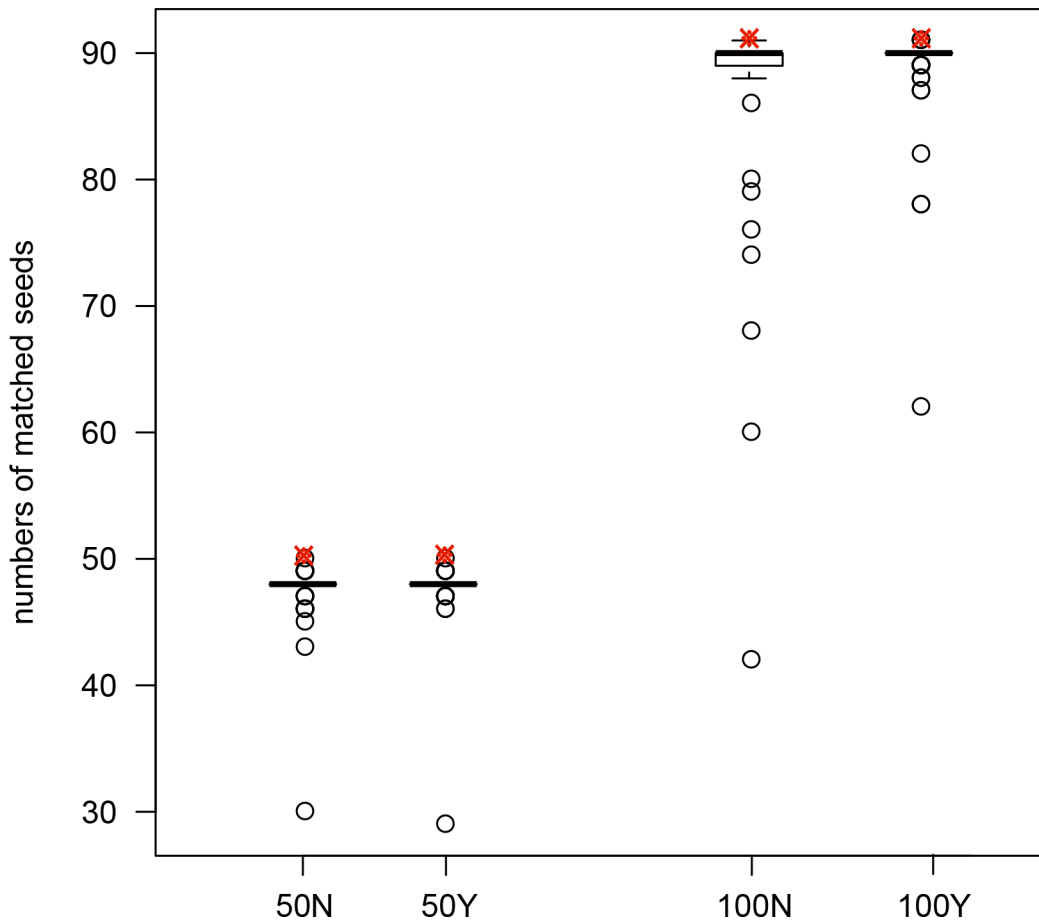


Fig 10. Comparison of the performance of the FIP with and without replacement in random sampling. 50N and 100N were top 50 and top 100 genes in the ranking list with and without replacement, respectively. 50Y and 100Y were top 50 and top 100 genes in the ranking list with back, respectively. Asterisks present our results.

doi:10.1371/journal.pone.0153006.g010

Discussion

In this study, we devised and implemented a novel algorithm called FIP to prioritize candidate genes involved in VTE. This algorithm is based on overall similarity with its neighboring genes by taking into account three aspects: expressions, functions, and interactions. In this way, we were able to prioritize the genes involved in VTE. For the top 200 candidates, we manually searched in PubMed literature and 124 genes were confirmed, in which 34 were drug targets. Furthermore, we conducted KEGG and GO analysis to functionally enrich the identified candidate genes. More candidates not confirmed by literature were classified into 10 significant functional categories which were associated with thrombotic disease (Fig 5). Overall FIP had a better predictive performance and achieved a reliable AUC value.

In reality, multiple properties of genes could be associated with each other in disease states contributing to the formation of disease. Integrating multiple data sources of genes has been reported to be better than a single data source in terms of sensitivity and accuracy of gene prioritization [97]. In our study, we compared the performance of integrating three data sources with those of integrating the two. As a result, there was no significant difference of the number of the matched seed genes between the combinations of integrating two data sources, while the combinations of integrating three data sources produced the much better performance than

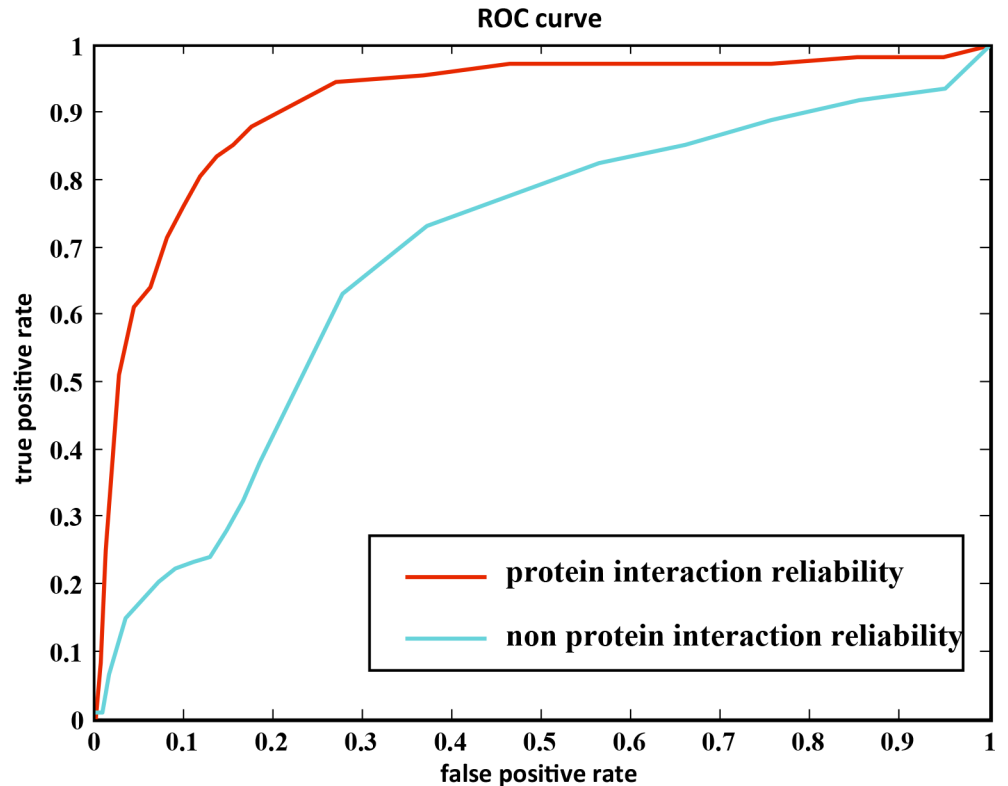


Fig 11. The ROC of the weight with protein interaction reliability and non-protein interaction reliability.

doi:10.1371/journal.pone.0153006.g011

those of integrating two data sources (t-test, $p < 0.05$) in terms of the number of the matched seed genes (S2 Table). Moreover, coefficients such as α , β , and γ and the control parameter d were shown to affect the performance of gene prioritization. According to the number of matched seed genes, LOOCV and 10-fold cross-validation, the best performance of gene prioritization was achieved in the parameter combination of $\alpha = 0.8$, $\beta = 0.5$, $\gamma = 0.9$, and $d = 0.9$ in prioritizing VTE-related genes in this study.

ToppNet and Endeavour are currently commonly used prioritization methods. According to network properties-based knowledge, ToppNet employs three algorithms (PageRank, Hyperlink-Induced Topic Search-HITS, and K-step Markov) to prioritize disease-related candidate genes by estimating their relative importance in PPIN [98,99]. Thus, ToppNet ranks or prioritizes genes based on topological features in PPIN with only one data type. As described above, the performance of integrating more data sources was better than those of integrating less ones. Thus, it is not surprised that FIP outperformed ToppNet in prioritizing genes involved in VTE in this study (Fig 8).

Endeavour takes the similar three data types as what we used in this study to rank candidate genes except its expression data background (high-density gene expression database). As compared to Endeavour, FIP applied disease-specific expression data, including recurrent VTE sample data, in our study. In theory, whether disease-specific or non disease-specific expression data through random sampling gene expression data could affect the performance of gene prioritization. This was confirmed by our results that disease-specific gene expression data did affect the performance of FIP (Fig 10). It was shown that FIP using VTE-specific gene expression data achieved the better performance than Endeavour using non disease-specific

expression data since the identified top candidates by FIP through VTE-specific gene expression analysis were more likely to be associated with VTE.

On the other hand, protein interaction databases used by the commonly used prioritization methods, including Endeavour, don't provide the details enough to assess whether a protein binds its interaction partner(s) which share similar structural or chemical properties and functionalities since many protein-protein links are loose because of random or unspecific bindings of proteins collected in databases. Thus, the reliability of protein-protein interaction is interrogated, resulting in low accuracy of ranking genes. In fact, the edge weight in the disease-related network can provide reasonable and consistent values to quantify the strength of connection of proteins. In our study, we took this feature into account in prioritizing candidates. As a baseline for weighted networks, we constructed a non-weighted network with the same protein interaction pairs and assessed the performance of FIPs using a weighted or a non-weighted network. It was showed FIP using a weighted network achieved a better performance (Fig 11). This result implied that the improved reliability of protein interaction applied by FIP might enhance its performance compared to Endeavour in prioritizing candidates related to VTE.

In summary, our FIP method combined experimental data with mathematical modeling and provided an alternative system biology approach in promising to tackle complex VTE disease for aiding diagnosis of recurrent VTE. This method could also be applied to other complex diseases to reveal disease mechanism and provide new perspective for diagnosis and drug development.

Supporting Information

S1 Fig. The boxplot of FIP, ToppNet and Endeavour.

(TIF)

S2 Fig. The Venn diagrams of literature validation among three methods at top 100 (left) and top 200 (right) candidates. The numbers in the slash left and right present the number of confirmed genes and the number of candidate genes, respectively.

(TIF)

S3 Fig. The comparison of literature validation for top 200 candidates generated from three methods on ten function categories.

(TIF)

S4 Fig. The comparison of literature validation among three methods at four pathways for top 200 candidates.

(TIF)

S5 Fig. The Venn diagrams of literature validation among four parameter combinations for top 200 candidates. The numbers in the slash left and right present the number of confirmed genes and the number of candidate genes, respectively.

(TIF)

S1 Table. Top 200 candidate genes identified by FIP.

(DOC)

S2 Table. The significance between combinations among the integration of the two and three data sources.

(DOC)

S3 Table. The literature validation of top 200 candidate genes among three methods.

(DOC)

S4 Table. The top 200 candidates of three methods on ten function categories.
(DOC)

S5 Table. The top 200 candidates among three methods at four pathways.
(DOC)

S6 Table. The literature validation among four parameter combinations for top 200 candidates.
(DOC)

Author Contributions

Conceived and designed the experiments: LNC WMH. Performed the experiments: JJJ YHH. Analyzed the data: YRL HH. Contributed reagents/materials/analysis tools: BBC RQX. Wrote the paper: JJ WL BHL.

References

1. Kooiman J, van Hagen N, Iglesias Del Sol A, Planken EV, Lip GY, van der Meer FJ, et al. (2015) The HAS-BLED Score Identifies Patients with Acute Venous Thromboembolism at High Risk of Major Bleeding Complications during the First Six Months of Anticoagulant Treatment. *PLoS One* 10: e0122520. doi: [10.1371/journal.pone.0122520](https://doi.org/10.1371/journal.pone.0122520) PMID: [25905638](https://pubmed.ncbi.nlm.nih.gov/25905638/)
2. Tatebe S (2015) Cardiologists and the management of obstetric venous thromboembolism. *Circ J* 79: 1215–1217. doi: [10.1253/circj.CJ-15-0367](https://doi.org/10.1253/circj.CJ-15-0367) PMID: [25902743](https://pubmed.ncbi.nlm.nih.gov/25902743/)
3. Chew TW, Gau CS, Wen YW, Shen LJ, Mullins CD, Hsiao FY. (2015) Epidemiology, clinical profile and treatment patterns of venous thromboembolism in cancer patients in Taiwan: a population-based study. *BMC Cancer* 15: 298. doi: [10.1186/s12885-015-1200-6](https://doi.org/10.1186/s12885-015-1200-6) PMID: [25925555](https://pubmed.ncbi.nlm.nih.gov/25925555/)
4. Hamidi S, Riazi M (2015) Cutoff values of plasma d-dimer level in patients with diagnosis of the venous thromboembolism after elective spinal surgery. *Asian Spine J* 9: 232–238. doi: [10.4184/asj.2015.9.2.232](https://doi.org/10.4184/asj.2015.9.2.232) PMID: [25901235](https://pubmed.ncbi.nlm.nih.gov/25901235/)
5. Millan Longo C (2014) [Oral apixaban for the treatment of acute venous thromboembolism]. *Rev Clin Esp (Barc)* 214: 164.
6. Schulman S, Lindmarker P, Holmstrom M, Larfars G, Carlsson A, Nicol P, et al. (2006) Post-thrombotic syndrome, recurrence, and death 10 years after the first episode of venous thromboembolism treated with warfarin for 6 weeks or 6 months. *J Thromb Haemost* 4: 734–742. PMID: [16634738](https://pubmed.ncbi.nlm.nih.gov/16634738/)
7. Rubio-Terres C, Soria JM, Morange PE, Souto JC, Suchon P, Mateo J, et al. (2015) Economic analysis of thrombo inCode, a clinical-genetic function for assessing the risk of venous thromboembolism. *Appl Health Econ Health Policy* 13: 233–242. doi: [10.1007/s40258-015-0153-x](https://doi.org/10.1007/s40258-015-0153-x) PMID: [25652150](https://pubmed.ncbi.nlm.nih.gov/25652150/)
8. Cai J, Preblich R, Zhang Q, Kwong WJ (2014) Utilization of parenteral anticoagulants and warfarin: impact on the risk of venous thromboembolism recurrence in the outpatient setting. *Am Health Drug Benefits* 7: 444–451. PMID: [25558306](https://pubmed.ncbi.nlm.nih.gov/25558306/)
9. Schulman S, Kakkar AK, Goldhaber SZ, Schellong S, Eriksson H, Mismetti P, et al. (2014) Treatment of acute venous thromboembolism with dabigatran or warfarin and pooled analysis. *Circulation* 129: 764–772. doi: [10.1161/CIRCULATIONAHA.113.004450](https://doi.org/10.1161/CIRCULATIONAHA.113.004450) PMID: [24344086](https://pubmed.ncbi.nlm.nih.gov/24344086/)
10. Moll S, Mackman N (2008) Venous thromboembolism: a need for more public awareness and research into mechanisms. *Arterioscler Thromb Vasc Biol* 28: 367–369. doi: [10.1161/ATVBAHA.108.163097](https://doi.org/10.1161/ATVBAHA.108.163097) PMID: [18296590](https://pubmed.ncbi.nlm.nih.gov/18296590/)
11. Pabinger I, Ay C (2009) Biomarkers and venous thromboembolism. *Arterioscler Thromb Vasc Biol* 29: 332–336. doi: [10.1161/ATVBAHA.108.182188](https://doi.org/10.1161/ATVBAHA.108.182188) PMID: [19228607](https://pubmed.ncbi.nlm.nih.gov/19228607/)
12. Verhovsek M, Douketis JD, Yi Q, Shrivastava S, Tait RC, Baglin T, et al. (2008) Systematic review: D-dimer to predict recurrent disease after stopping anticoagulant therapy for unprovoked venous thromboembolism. *Ann Intern Med* 149: 481–490, W494. PMID: [18838728](https://pubmed.ncbi.nlm.nih.gov/18838728/)
13. Kyrle PA, Hron G, Eichinger S, Wagner O (2007) Circulating P-selectin and the risk of recurrent venous thromboembolism. *Thromb Haemost* 97: 880–883. PMID: [17549288](https://pubmed.ncbi.nlm.nih.gov/17549288/)
14. Hron G, Kollars M, Binder BR, Eichinger S, Kyrle PA (2006) Identification of patients at low risk for recurrent venous thromboembolism by measuring thrombin generation. *JAMA* 296: 397–402. PMID: [16868297](https://pubmed.ncbi.nlm.nih.gov/16868297/)

15. Lewis DA, Suchindran S, Beckman MG, Hooper WC, Grant AM, Heit JA, et al. (2015) Whole blood gene expression profiles distinguish clinical phenotypes of venous thromboembolism. *Thromb Res* 135: 659–665. doi: [10.1016/j.thromres.2015.02.003](https://doi.org/10.1016/j.thromres.2015.02.003) PMID: [25684211](https://pubmed.ncbi.nlm.nih.gov/25684211/)
16. Lewis DA, Stashenko GJ, Akay OM, Price LI, Owzar K, Ginsburg GS, et al. (2011) Whole blood gene expression analyses in patients with single versus recurrent venous thromboembolism. *Thromb Res* 128: 536–540. doi: [10.1016/j.thromres.2011.06.003](https://doi.org/10.1016/j.thromres.2011.06.003) PMID: [21737128](https://pubmed.ncbi.nlm.nih.gov/21737128/)
17. Liebner DA, Huang K, Parvin JD (2014) MMAD: microarray microdissection with analysis of differences is a computational tool for deconvoluting cell type-specific contributions from tissue samples. *Bioinformatics* 30: 682–689. doi: [10.1093/bioinformatics/btt566](https://doi.org/10.1093/bioinformatics/btt566) PMID: [24085566](https://pubmed.ncbi.nlm.nih.gov/24085566/)
18. Chen J, Xu H, Aronow BJ, Jegga AG (2007) Improved human disease candidate gene prioritization using mouse phenotype. *BMC Bioinformatics* 8: 392. PMID: [17939863](https://pubmed.ncbi.nlm.nih.gov/17939863/)
19. Tranchevent LC, Barriot R, Yu S, Van Vooren S, Van Loo P, Coessens B, et al. (2008) ENDEAVOUR update: a web resource for gene prioritization in multiple species. *Nucleic Acids Res* 36: W377–384. doi: [10.1093/nar/gkn325](https://doi.org/10.1093/nar/gkn325) PMID: [18508807](https://pubmed.ncbi.nlm.nih.gov/18508807/)
20. Adie EA, Adams RR, Evans KL, Porteous DJ, Pickard BS (2006) SUSPECTS: enabling fast and effective prioritization of positional candidates. *Bioinformatics* 22: 773–774. PMID: [16423925](https://pubmed.ncbi.nlm.nih.gov/16423925/)
21. Nitsch D, Tranchevent LC, Goncalves JP, Vogt JK, Madeira SC, Moreau Y (2011) PINTA: a web server for network-based gene prioritization from expression data. *Nucleic Acids Res* 39: W334–338. doi: [10.1093/nar/gkr289](https://doi.org/10.1093/nar/gkr289) PMID: [21602267](https://pubmed.ncbi.nlm.nih.gov/21602267/)
22. Chang JT, Nevins JR (2006) GATHER: a systems approach to interpreting genomic signatures. *Bioinformatics* 22: 2926–2933. PMID: [17000751](https://pubmed.ncbi.nlm.nih.gov/17000751/)
23. Foong J, Girdea M, Stavropoulos J, Brudno M (2015) Prioritizing Clinically Relevant Copy Number Variation from Genetic Interactions and Gene Function Data. *PLoS One* 10: e0139656. doi: [10.1371/journal.pone.0139656](https://doi.org/10.1371/journal.pone.0139656) PMID: [26437450](https://pubmed.ncbi.nlm.nih.gov/26437450/)
24. Taniya T, Tanaka S, Yamaguchi-Kabata Y, Hanaoka H, Yamasaki C, Maekawa H, et al. (2012) A prioritization analysis of disease association by data-mining of functional annotation of human genes. *Genomics* 99: 1–9. doi: [10.1016/j.ygeno.2011.10.002](https://doi.org/10.1016/j.ygeno.2011.10.002) PMID: [22019378](https://pubmed.ncbi.nlm.nih.gov/22019378/)
25. Wang X, Gulbahce N, Yu H (2011) Network-based methods for human disease gene prediction. *Brief Funct Genomics* 10: 280–293. doi: [10.1093/bfpg/elr024](https://doi.org/10.1093/bfpg/elr024) PMID: [21764832](https://pubmed.ncbi.nlm.nih.gov/21764832/)
26. Luo J, Liang S (2015) Prioritization of potential candidate disease genes by topological similarity of protein-protein interaction network and phenotype data. *J Biomed Inform* 53: 229–236. doi: [10.1016/j.jbi.2014.11.004](https://doi.org/10.1016/j.jbi.2014.11.004) PMID: [25460206](https://pubmed.ncbi.nlm.nih.gov/25460206/)
27. Goncalves JP, Francisco AP, Moreau Y, Madeira SC (2012) Interactogeneous: disease gene prioritization using heterogeneous networks and full topology scores. *PLoS One* 7: e49634. doi: [10.1371/journal.pone.0049634](https://doi.org/10.1371/journal.pone.0049634) PMID: [23185389](https://pubmed.ncbi.nlm.nih.gov/23185389/)
28. Magger O, Waldman YY, Ruppin E, Sharan R (2012) Enhancing the prioritization of disease-causing genes through tissue specific protein interaction networks. *PLoS Comput Biol* 8: e1002690. doi: [10.1371/journal.pcbi.1002690](https://doi.org/10.1371/journal.pcbi.1002690) PMID: [23028288](https://pubmed.ncbi.nlm.nih.gov/23028288/)
29. Antanaviciute A, Daly C, Crinnion LA, Markham AF, Watson CM, Bonthron DT, et al. (2015) GeneTIER: prioritization of candidate disease genes using tissue-specific gene expression profiles. *Bioinformatics*.
30. Oellrich A, Sanger Mouse Genetics P, Smedley D (2014) Linking tissues to phenotypes using gene expression profiles. *Database (Oxford)* 2014: bau017.
31. Xiao Y, Xu C, Ping Y, Guan J, Fan H, Li Y, et al. (2011) Differential expression pattern-based prioritization of candidate genes through integrating disease-specific expression data. *Genomics* 98: 64–71. doi: [10.1016/j.ygeno.2011.04.001](https://doi.org/10.1016/j.ygeno.2011.04.001) PMID: [21515357](https://pubmed.ncbi.nlm.nih.gov/21515357/)
32. Rao BS, Gupta KK, Karanam P, Peruri A (2014) Alzheimer disease: An interactome of many diseases. *Ann Indian Acad Neurol* 17: 48–54. doi: [10.4103/0972-2327.128551](https://doi.org/10.4103/0972-2327.128551) PMID: [24753659](https://pubmed.ncbi.nlm.nih.gov/24753659/)
33. Barh D, Kamapantula B, Jain N, Nalluri J, Bhattacharya A, Juneja L, et al. (2015) miRegulome: a knowledge-base of miRNA regulomics and analysis. *Sci Rep* 5: 12832. doi: [10.1038/srep12832](https://doi.org/10.1038/srep12832) PMID: [26243198](https://pubmed.ncbi.nlm.nih.gov/26243198/)
34. O'Brien MA, Costin BN, Miles MF (2012) Using genome-wide expression profiling to define gene networks relevant to the study of complex traits: from RNA integrity to network topology. *Int Rev Neurobiol* 104: 91–133. doi: [10.1016/B978-0-12-398323-7.00005-7](https://doi.org/10.1016/B978-0-12-398323-7.00005-7) PMID: [23195313](https://pubmed.ncbi.nlm.nih.gov/23195313/)
35. Lascorz J, Forsti A, Chen B, Buch S, Steinke V, Rahner N, et al. (2010) Genome-wide association study for colorectal cancer identifies risk polymorphisms in German familial cases and implicates MAPK signalling pathways in disease susceptibility. *Carcinogenesis* 31: 1612–1619. doi: [10.1093/carcin/bgq146](https://doi.org/10.1093/carcin/bgq146) PMID: [20610541](https://pubmed.ncbi.nlm.nih.gov/20610541/)

36. Dever SM, Costin BN, Xu R, El-Hage N, Balinang J, Samoshkin A, et al. (2014) Differential expression of the alternatively spliced OPRM1 isoform mu-opioid receptor-1K in HIV-infected individuals. *AIDS* 28: 19–30. doi: [10.1097/QAD.000000000000113](https://doi.org/10.1097/QAD.000000000000113) PMID: [24413261](https://pubmed.ncbi.nlm.nih.gov/24413261/)
37. Chapman NH, Wijsman EM (1998) Genome screens using linkage disequilibrium tests: optimal marker characteristics and feasibility. *Am J Hum Genet* 63: 1872–1885. PMID: [9837839](https://pubmed.ncbi.nlm.nih.gov/9837839/)
38. Zheng G, Freidlin B, Gastwirth JL (2006) Robust genomic control for association studies. *Am J Hum Genet* 78: 350–356. PMID: [16400614](https://pubmed.ncbi.nlm.nih.gov/16400614/)
39. Sasieni PD (1997) From genotypes to genes: doubling the sample size. *Biometrics* 53: 1253–1261. PMID: [9423247](https://pubmed.ncbi.nlm.nih.gov/9423247/)
40. Manolio TA (2010) Genomewide association studies and assessment of the risk of disease. *N Engl J Med* 363: 166–176. doi: [10.1056/NEJMra0905980](https://doi.org/10.1056/NEJMra0905980) PMID: [20647212](https://pubmed.ncbi.nlm.nih.gov/20647212/)
41. Che J, Shin M (2015) A meta-analysis strategy for gene prioritization using gene expression, SNP genotype, and eQTL data. *Biomed Res Int* 2015: 576349. doi: [10.1155/2015/576349](https://doi.org/10.1155/2015/576349) PMID: [25874220](https://pubmed.ncbi.nlm.nih.gov/25874220/)
42. Jiang R, Wu M, Li L (2015) Pinpointing disease genes through phenomic and genomic data fusion. *BMC Genomics* 16 Suppl 2: S3. doi: [10.1186/1471-2164-16-S2-S3](https://doi.org/10.1186/1471-2164-16-S2-S3) PMID: [25708473](https://pubmed.ncbi.nlm.nih.gov/25708473/)
43. Vafaee F, Rosu D, Broackes-Carter F, Jurisica I (2013) Novel semantic similarity measure improves an integrative approach to predicting gene functional associations. *BMC Syst Biol* 7: 22. doi: [10.1186/1752-0509-7-22](https://doi.org/10.1186/1752-0509-7-22) PMID: [23497449](https://pubmed.ncbi.nlm.nih.gov/23497449/)
44. Wu C, Zhu J, Zhang X (2012) Integrating gene expression and protein-protein interaction network to prioritize cancer-associated genes. *BMC Bioinformatics* 13: 182. PMID: [22838965](https://pubmed.ncbi.nlm.nih.gov/22838965/)
45. Hulsege I, Woelders H, Smits M, Schokker D, Jiang L, Sorensen P (2013) Prioritization of candidate genes for cattle reproductive traits, based on protein-protein interactions, gene expression, and text-mining. *Physiol Genomics* 45: 400–406.
46. Zhang SW, Shao DD, Zhang SY, Wang YB (2014) Prioritization of candidate disease genes by enlarging the seed set and fusing information of the network topology and gene expression. *Mol Biosyst* 10: 1400–1408. doi: [10.1039/c3mb70588a](https://doi.org/10.1039/c3mb70588a) PMID: [24695957](https://pubmed.ncbi.nlm.nih.gov/24695957/)
47. Aerts S, Lambrechts D, Maity S, Van Loo P, Coessens B, De Smet F, et al. (2006) Gene prioritization through genomic data fusion. *Nat Biotechnol* 24: 537–544. PMID: [16680138](https://pubmed.ncbi.nlm.nih.gov/16680138/)
48. Menzies RI, Unwin RJ, Dash RK, Beard DA, Cowley AW Jr., Carlson BE, et al. (2013) Effect of P2X4 and P2X7 receptor antagonism on the pressure diuresis relationship in rats. *Front Physiol* 4: 305. doi: [10.3389/fphys.2013.00305](https://doi.org/10.3389/fphys.2013.00305) PMID: [24187541](https://pubmed.ncbi.nlm.nih.gov/24187541/)
49. Khan K, Al-Maskari A, McKibbin M, Carr IM, Booth A, Mohamed M, et al. (2011) Genetic heterogeneity for recessively inherited congenital cataract microcornea with corneal opacity. *Invest Ophthalmol Vis Sci* 52: 4294–4299. doi: [10.1167/iovs.10-6776](https://doi.org/10.1167/iovs.10-6776) PMID: [21474777](https://pubmed.ncbi.nlm.nih.gov/21474777/)
50. Oliver KL, Lukic V, Thorne NP, Berkovic SF, Scheffer IE, Bahlo M (2014) Harnessing gene expression networks to prioritize candidate epileptic encephalopathy genes. *PLoS One* 9: e102079. doi: [10.1371/journal.pone.0102079](https://doi.org/10.1371/journal.pone.0102079) PMID: [25014031](https://pubmed.ncbi.nlm.nih.gov/25014031/)
51. Mughini-Gras L, Enserink R, Friesema I, Heck M, van Duynhoven Y, van Pelt W (2014) Risk factors for human salmonellosis originating from pigs, cattle, broiler chickens and egg laying hens: a combined case-control and source attribution analysis. *PLoS One* 9: e87933. doi: [10.1371/journal.pone.0087933](https://doi.org/10.1371/journal.pone.0087933) PMID: [24503703](https://pubmed.ncbi.nlm.nih.gov/24503703/)
52. Barrett T, Wilhite SE, Ledoux P, Evangelista C, Kim IF, Tomashevsky M, et al. (2013) NCBI GEO: archive for functional genomics data sets—update. *Nucleic Acids Res* 41: D991–995. doi: [10.1093/nar/gks1193](https://doi.org/10.1093/nar/gks1193) PMID: [23193258](https://pubmed.ncbi.nlm.nih.gov/23193258/)
53. Kanehisa M, Goto S (2000) KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Res* 28: 27–30. PMID: [10592173](https://pubmed.ncbi.nlm.nih.gov/10592173/)
54. Becker KG, Barnes KC, Bright TJ, Wang SA (2004) The genetic association database. *Nat Genet* 36: 431–432. PMID: [15118671](https://pubmed.ncbi.nlm.nih.gov/15118671/)
55. Kibbe WA, Arze C, Felix V, Mitraka E, Bolton E, Fu G, et al. (2015) Disease Ontology 2015 update: an expanded and updated database of human diseases for linking biomedical knowledge through disease data. *Nucleic Acids Res* 43: D1071–1078. doi: [10.1093/nar/gku1011](https://doi.org/10.1093/nar/gku1011) PMID: [25348409](https://pubmed.ncbi.nlm.nih.gov/25348409/)
56. Szklarczyk D, Franceschini A, Wyder S, Forslund K, Heller D, Huerta-Cepas J, et al. (2015) STRING v10: protein-protein interaction networks, integrated over the tree of life. *Nucleic Acids Res* 43: D447–452. doi: [10.1093/nar/gku1003](https://doi.org/10.1093/nar/gku1003) PMID: [25352553](https://pubmed.ncbi.nlm.nih.gov/25352553/)
57. Jensen LJ, Kuhn M, Stark M, Chaffron S, Creevey C, Muller J, et al. (2009) STRING 8—a global view on proteins and their functional interactions in 630 organisms. *Nucleic Acids Res* 37: D412–416. doi: [10.1093/nar/gkn760](https://doi.org/10.1093/nar/gkn760) PMID: [18940858](https://pubmed.ncbi.nlm.nih.gov/18940858/)

58. Franceschini A, Szklarczyk D, Frankild S, Kuhn M, Simonovic M, Roth A, et al. (2013) STRING v9.1: protein-protein interaction networks, with increased coverage and integration. *Nucleic Acids Res* 41: D808–815. doi: [10.1093/nar/gks1094](https://doi.org/10.1093/nar/gks1094) PMID: [23203871](https://pubmed.ncbi.nlm.nih.gov/23203871/)
59. Shin M, Lee H (2011) Prioritizing candidate genes by weighted network structure for the identification of disease marker genes. *BioChip Journal* 5.
60. Gigante B, Vikstrom M, Meuzelaar LS, Chernogubova E, Silveira A, Hooft FV, et al. (2009) Variants in the coagulation factor 2 receptor (F2R) gene influence the risk of myocardial infarction in men through an interaction with interleukin 6 serum levels. *Thromb Haemost* 101: 943–953. PMID: [19404549](https://pubmed.ncbi.nlm.nih.gov/19404549/)
61. Misumida N, Kobayashi A, Saeed M, Fox JT, Kanei Y (2015) Prevalence and outcomes of non-ST-segment elevation myocardial infarction resulting from stent thrombosis. *Cardiovasc Revasc Med*.
62. Li D, Weng S, Yang B, Zander DS, Saldeen T, Nichol WW, et al. (1999) Inhibition of arterial thrombus formation by ApoA1 Milano. *Arterioscler Thromb Vasc Biol* 19: 378–383. PMID: [9974422](https://pubmed.ncbi.nlm.nih.gov/9974422/)
63. Livnat T, Shenkman B, Kenet G, Tamarin I, Gillis S, Varon D, et al. (2011) Severe factor X deficiency in three unrelated Palestinian patients is caused by homozygosity for the mutation c302delG—correlation with thrombin generation and thromboelastometry. *Blood Coagul Fibrinolysis* 22: 673–679. doi: [10.1097/MBC.0b013e32834ad785](https://doi.org/10.1097/MBC.0b013e32834ad785) PMID: [22008904](https://pubmed.ncbi.nlm.nih.gov/22008904/)
64. Lippi G, Harenberg J, Mattiuzzi C, Favalaro EJ (2015) Next generation antithrombotic therapy: focus on antisense therapy against coagulation factor XI. *Semin Thromb Hemost* 41: 255–262. doi: [10.1055/s-0035-1546466](https://doi.org/10.1055/s-0035-1546466) PMID: [25703390](https://pubmed.ncbi.nlm.nih.gov/25703390/)
65. Luan D, Zai M, Varner JD (2007) Computationally derived points of fragility of a human cascade are consistent with current therapeutic strategies. *PLoS Comput Biol* 3: e142. PMID: [17658944](https://pubmed.ncbi.nlm.nih.gov/17658944/)
66. Tregouet DA, Sabater-Lleal M, Bruzelius M, Emmerich J, Amouyel P, Dartigues JF, et al. (2012) Lack of association of non-synonymous FUT2 and ALPL polymorphisms with venous thrombosis. *J Thromb Haemost* 10: 1693–1695. doi: [10.1111/j.1538-7836.2012.04807.x](https://doi.org/10.1111/j.1538-7836.2012.04807.x) PMID: [22672431](https://pubmed.ncbi.nlm.nih.gov/22672431/)
67. van Hylckama Vlieg A, Flinterman LE, Bare LA, Cannegieter SC, Reitsma PH, Arellano AR, et al. (2014) Genetic variations associated with recurrent venous thrombosis. *Circ Cardiovasc Genet* 7: 806–813. doi: [10.1161/CIRCGENETICS.114.000682](https://doi.org/10.1161/CIRCGENETICS.114.000682) PMID: [25210051](https://pubmed.ncbi.nlm.nih.gov/25210051/)
68. Smith NL, Heit JA, Tang W, Teichert M, Chasman DI, Morange PE, et al. (2012) Genetic variation in F3 (tissue factor) and the risk of incident venous thrombosis: meta-analysis of eight studies. *J Thromb Haemost* 10: 719–722. doi: [10.1111/j.1538-7836.2012.04665.x](https://doi.org/10.1111/j.1538-7836.2012.04665.x) PMID: [22340074](https://pubmed.ncbi.nlm.nih.gov/22340074/)
69. Xie C, Ding X, Gao J, Wang H, Hang Y, Zhang H, et al. (2014) The effects of CES1A2 A(-816)C and CYP2C19 loss-of-function polymorphisms on clopidogrel response variability among Chinese patients with coronary heart disease. *Pharmacogenet Genomics* 24: 204–210. doi: [10.1097/FPC.000000000000035](https://doi.org/10.1097/FPC.000000000000035) PMID: [24535487](https://pubmed.ncbi.nlm.nih.gov/24535487/)
70. Dahabreh IJ, Moorthy D, Lamont JL, Chen ML, Kent DM, Lau J (2013). Testing of CYP2C19 Variants and Platelet Reactivity for Guiding Antiplatelet Treatment. Rockville (MD).
71. Wang Y, Xu J, Chen J, Fan X, Zhang Y, Yu W, et al. (2013) Promoter variants of VTN are associated with vascular disease. *Int J Cardiol* 168: 163–168. doi: [10.1016/j.ijcard.2012.09.100](https://doi.org/10.1016/j.ijcard.2012.09.100) PMID: [23041018](https://pubmed.ncbi.nlm.nih.gov/23041018/)
72. Raps M, Helmerhorst F, Fleischer K, Thomassen S, Rosendaal F, Rosing J, et al. (2012) Sex hormone-binding globulin as a marker for the thrombotic risk of hormonal contraceptives. *J Thromb Haemost* 10: 992–997. doi: [10.1111/j.1538-7836.2012.04720.x](https://doi.org/10.1111/j.1538-7836.2012.04720.x) PMID: [22469296](https://pubmed.ncbi.nlm.nih.gov/22469296/)
73. Huang da W, Sherman BT, Lempicki RA (2009) Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat Protoc* 4: 44–57. doi: [10.1038/nprot.2008.211](https://doi.org/10.1038/nprot.2008.211) PMID: [19131956](https://pubmed.ncbi.nlm.nih.gov/19131956/)
74. Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, et al. (2000) Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat Genet* 25: 25–29. PMID: [10802651](https://pubmed.ncbi.nlm.nih.gov/10802651/)
75. Kanehisa M, Goto S, Sato Y, Kawashima M, Furumichi M, Tanabe M (2014) Data, information, knowledge and principle: back to metabolism in KEGG. *Nucleic Acids Res* 42: D199–205. doi: [10.1093/nar/gkt1076](https://doi.org/10.1093/nar/gkt1076) PMID: [24214961](https://pubmed.ncbi.nlm.nih.gov/24214961/)
76. Palta S, Saroa R, Palta A (2014) Overview of the coagulation system. *Indian J Anaesth* 58: 515–523. doi: [10.4103/0019-5049.144643](https://doi.org/10.4103/0019-5049.144643) PMID: [25535411](https://pubmed.ncbi.nlm.nih.gov/25535411/)
77. Thushara RM, Hemshekhar M, Basappa, Kemparaju K, Rangappa KS, Girish K. S. (2015) Biologicals, platelet apoptosis and human diseases: An outlook. *Crit Rev Oncol Hematol* 93: 149–158. doi: [10.1016/j.critrevonc.2014.11.002](https://doi.org/10.1016/j.critrevonc.2014.11.002) PMID: [25439323](https://pubmed.ncbi.nlm.nih.gov/25439323/)
78. Mollard JM, Prevot JM, Baulande M (1985) [Prevention of deep venous thrombosis by physical methods. Use of an external electrical stimulator. Initial results in surgery of the hip]. *Phlebologie* 38: 293–305. PMID: [3875112](https://pubmed.ncbi.nlm.nih.gov/3875112/)
79. Koupenova M, Freedman JE (2015) Platelets: the unsung hero of the immune response. *J Thromb Haemost* 13: 268–270. doi: [10.1111/jth.12806](https://doi.org/10.1111/jth.12806) PMID: [25471925](https://pubmed.ncbi.nlm.nih.gov/25471925/)

80. Gardiner EE, Andrews RK (2014) Structure and function of platelet receptors initiating blood clotting. *Adv Exp Med Biol* 844: 263–275. doi: [10.1007/978-1-4939-2095-2_13](https://doi.org/10.1007/978-1-4939-2095-2_13) PMID: [25480646](https://pubmed.ncbi.nlm.nih.gov/25480646/)
81. Marcus AJ, Safier LB (1993) Thromboregulation: multicellular modulation of platelet reactivity in hemostasis and thrombosis. *FASEB J* 7: 516–522. PMID: [8472890](https://pubmed.ncbi.nlm.nih.gov/8472890/)
82. Liu H, Hou J, Hu S, Du X, Fang Y, Jia H, et al. (2014) A rabbit model of spontaneous thrombosis induced by lipopolysaccharide. *J Atheroscler Thromb* 21: 1075–1086. PMID: [24898380](https://pubmed.ncbi.nlm.nih.gov/24898380/)
83. Loren CP, Aslan JE, Rigg RA, Nowak MS, Healy LD, Gruber A, et al. (2015) The BCR-ABL inhibitor ponatinib inhibits platelet immunoreceptor tyrosine-based activation motif (ITAM) signaling, platelet activation and aggregate formation under shear. *Thromb Res* 135: 155–160. doi: [10.1016/j.thromres.2014.11.009](https://doi.org/10.1016/j.thromres.2014.11.009) PMID: [25527332](https://pubmed.ncbi.nlm.nih.gov/25527332/)
84. Myers DD Jr. (2015) Pathophysiology of venous thrombosis. *Phlebology* 30: 7–13. doi: [10.1177/0268355515569424](https://doi.org/10.1177/0268355515569424) PMID: [25729062](https://pubmed.ncbi.nlm.nih.gov/25729062/)
85. Brown KV, Ramasamy A, Tai N, MacLeod J, Midwinter M, Clasper JC (2009) Complications of extremity vascular injuries in conflict. *J Trauma* 66: S145–149. doi: [10.1097/TA.0b013e31819cdd82](https://doi.org/10.1097/TA.0b013e31819cdd82) PMID: [19359958](https://pubmed.ncbi.nlm.nih.gov/19359958/)
86. Versteeg HH, Heemskerk JW, Levi M, Reitsma PH (2013) New fundamentals in hemostasis. *Physiol Rev* 93: 327–358. doi: [10.1152/physrev.00016.2011](https://doi.org/10.1152/physrev.00016.2011) PMID: [23303912](https://pubmed.ncbi.nlm.nih.gov/23303912/)
87. Li S, Chen H, Ren J, Geng Q, Song J, Lee C, et al. (2014) MicroRNA-223 inhibits tissue factor expression in vascular endothelial cells. *Atherosclerosis* 237: 514–520. doi: [10.1016/j.atherosclerosis.2014.09.033](https://doi.org/10.1016/j.atherosclerosis.2014.09.033) PMID: [25463083](https://pubmed.ncbi.nlm.nih.gov/25463083/)
88. Tatsumi K, Mackman N (2015) Tissue Factor and Atherothrombosis. *J Atheroscler Thromb*.
89. Han J, Shuvaev VV, Davies PF, Eckmann DM, Muro S, Muzykantov VR (2015) Flow shear stress differentially regulates endothelial uptake of nanocarriers targeted to distinct epitopes of PECAM-1. *J Control Release* 210: 39–47. doi: [10.1016/j.jconrel.2015.05.006](https://doi.org/10.1016/j.jconrel.2015.05.006) PMID: [25966362](https://pubmed.ncbi.nlm.nih.gov/25966362/)
90. Kahner BN, Dorsam RT, Mada SR, Kim S, Stalker TJ, Brass LF, et al. (2007) Hematopoietic lineage cell specific protein 1 (HS1) is a functionally important signaling molecule in platelet activation. *Blood* 110: 2449–2456. PMID: [17579181](https://pubmed.ncbi.nlm.nih.gov/17579181/)
91. Kanehisa M, Sato Y, Morishima K (2015) BlastKOALA and GhostKOALA: KEGG tools for functional characterization of genome and metagenome sequences. *J Mol Biol*.
92. Kanehisa M, Goto S, Sato Y, Furumichi M, Tanabe M (2012) KEGG for integration and interpretation of large-scale molecular data sets. *Nucleic Acids Res* 40: D109–114. doi: [10.1093/nar/gkr988](https://doi.org/10.1093/nar/gkr988) PMID: [22080510](https://pubmed.ncbi.nlm.nih.gov/22080510/)
93. D'Asti E, Kool M, Pfister SM, Rak J (2014) Coagulation and angiogenic gene expression profiles are defined by molecular subgroups of medulloblastoma: evidence for growth factor-thrombin cross-talk. *J Thromb Haemost* 12: 1838–1849. doi: [10.1111/jth.12715](https://doi.org/10.1111/jth.12715) PMID: [25163932](https://pubmed.ncbi.nlm.nih.gov/25163932/)
94. Zhang F, Wen Y, Guo X, Zhang Y, Wang S, Yang T, et al. (2015) Genome-wide pathway-based association study implicates complement system in the development of Kashin-Beck disease in Han Chinese. *Bone* 71: 36–41. doi: [10.1016/j.bone.2014.09.025](https://doi.org/10.1016/j.bone.2014.09.025) PMID: [25305519](https://pubmed.ncbi.nlm.nih.gov/25305519/)
95. Heeb MJ (2008) Role of the PROS1 gene in thrombosis: lessons and controversies. *Expert Rev Hematol* 1: 9–12. doi: [10.1586/17474086.1.1.9](https://doi.org/10.1586/17474086.1.1.9) PMID: [19809585](https://pubmed.ncbi.nlm.nih.gov/19809585/)
96. Wypasek E, Undas A (2013) Protein C and protein S deficiency—practical diagnostic issues. *Adv Clin Exp Med* 22: 459–467. PMID: [23986205](https://pubmed.ncbi.nlm.nih.gov/23986205/)
97. Petrillo A, Fusco R, Petrillo M, Granata V, Sansone M, Avallone A, et al. (2015) Standardized Index of Shape (SIS): a quantitative DCE-MRI parameter to discriminate responders by non-responders after neoadjuvant therapy in LARC. *Eur Radiol* 25: 1935–1945. doi: [10.1007/s00330-014-3581-3](https://doi.org/10.1007/s00330-014-3581-3) PMID: [25577525](https://pubmed.ncbi.nlm.nih.gov/25577525/)
98. Chen J, Bardes EE, Aronow BJ, Jegga AG (2009) ToppGene Suite for gene list enrichment analysis and candidate gene prioritization. *Nucleic Acids Res* 37: W305–311. doi: [10.1093/nar/gkp427](https://doi.org/10.1093/nar/gkp427) PMID: [19465376](https://pubmed.ncbi.nlm.nih.gov/19465376/)
99. Junker BH, Koschutski D, Schreiber F (2006) Exploration of biological network centralities with CentiBiN. *BMC Bioinformatics* 7: 219. PMID: [16630347](https://pubmed.ncbi.nlm.nih.gov/16630347/)