Identification of critical genes and pathways related to venous thromboembolism in non-small cell lung cancer patients using integrated bioinformatics analysis

Sen Han
Peking University Cancer Hospital
Qibin Luo
Technische Universität München and Beijing QYNODE Institute of Biotechnology Co., Ltd.
Yi Ren
Beijing QYNODE Institute of Biotechnology Co., Ltd.
Xu Ma (maxu1985@bjmu.edu.cn)
Peking University Cancer Hospital
Jian Fang
Peking University Cancer Hospital

Research Article

Keywords: non-small cell lung cancer, venous thromboembolism, genome-wide association study, bioinformatics

Posted Date: February 3rd, 2021

DOI: https://doi.org/10.21203/rs.3.rs-155640/v1

Abstract

Background: Non-small cell lung cancer (NSCLC) patients have a significantly higher risk of developing venous thromboembolism (VTE). Although many driver mutations as well as some susceptibility loci for VTE have been identified in NSCLC, the critical genome atlas of NSCLC patients complicated with VTE and relevant molecular mechanisms are yet to be fully understood. The aim of this study is to investigate the critical genes and pathways related to VTE in NSCLC patients using integrated bioinformatics analysis.

Results: RNA-Seq data of 1014 NSCLC samples were collected and analyzed systematically to identify the potential risk genes of NSCLC, while VTE risk genes were obtained from the associated study. The VTE risk genes with differential expression among NSCLC were further subjected to pathway enrichment analysis and protein-protein interaction network (PPI) analysis to determine the potential pathways associated with developing VTE in NSCLC patients. This study identified a total of 22 genes with differential expression in the NSCLC group, which can be defined as VTE risk genes based on previous studies. The PPI network analysis demonstrated that among the 22 encoded proteins, 21 including VWF, F2, F8, and four blood coagulation associated proteins were located in the central part of the network. The bioinformatics analysis further identified enriched pathways with statistical significance ($P < 0.05$), most of which were associated with blood coagulation.

Conclusions: This study reveals that potential risk genes of NSCLC-associated VTE are enriched in blood coagulation associated pathways, suggesting a significant role of blood coagulation in VTE development of NSCLC patients.

Introduction

Venous thromboembolism events (VTE) includes deep vein thrombosis (DVT), migratory thrombophlebitis (Trousseau syndrome), and pulmonary embolism (PE). The correlation between cancer and VTE was discovered by Trousseau in 1865. It has been reported that as many as 20% of VTE cases derive from cancerous patients[1, 2]. Moreover, cancer-associated VTE is more aggressive and usually associated with worse prognosis[3]. Lung cancer is one of the most common malignancies worldwide, with a low 5-year survival rate[4, 5]. Non-small cell lung cancer (NSCLC) patients account for 80–85% of lung cancer cases [6]. Relative risk of developing VTE in NSCLC patients is almost 2 times as high as that in those with small-cell lung cancer (SCLC)[7]. Notably, 7.3–13.6% of patients with NSCLC are related to VTE[8, 9]. As a VTE risk factor, NSCLC plays an important role in pathogenesis and development of VTE[10].

Once being complicated in NSCLC patients, VTE could lead to devastating consequences. The prognosis of NSCLC patients complicated with VTE is influenced by not only the thrombotic event itself, but also the adverse effects of anti-thrombosis treatment, including bleeding, delay in delivering chemotherapy, and interactions between the anti-thrombosis treatment and other medications[11]. In recent years, a growing

number of studies have focused on thrombosis events and related risk factors of NSCLC patients[12–14]. Although many driver mutation genes as well as some susceptibility loci for VTE have been identified in NSCLC[15], critical and common genome atlas of NSCLC patients complicated with VTE and relevant molecular mechanisms have yet to be fully understood.

Microarray technology, usually with bioinformatics analysis, has been widely used to profile gene expression in cancer. Many studies have already employed this technology to investigate the differentially expressed genes (DEGs) in distinct pathways, biological processes, molecular functions, and cellular components[16, 17]. The combination of microarray technology with bioinformatics analysis enables comprehensive analysis of changes in gene expression, providing a new and effective way to investigate the molecular mechanisms of various diseases[18]. In this study, we carried out bioinformatics analysis to examine characteristic changes of gene expression in NSCLC patients and those with VTE, and then try to find their overlaps and correlations. The findings in the study could provide some novel insights into the critical genes and pathways in both NSCLC and VTE.

Methods

Database search

RNA-seq and clinical data of NSCLC were downloaded from LUAD and LUSC datasets in the TCGA Research Network (www.cancer.gov/tcga). Articles on Genome-Wide Association Studies (GWAS) towards VTE risk were searched for identifying the potential VTE risk genes.

Study Selection Criteria

The clinical data and matched gene expression data were selected according to the gene expression profile, while samples in TCGA database with the gene expression data were retained.

Bioinformatics Analysis

Additional quality control of the NSCLC samples was conducted according to the gene expression values prior to differential gene expression (DGE) analysis. DESeq2 package of R program (www.bioconductor.org/; version 3.1) was employed to carry out the DGE analysis[19], while DAVID (https://david.ncifcrf.gov/home.jsp), an online program for identification of enriched biological themes and gene pathways[20], was used for conducting gene set and pathway enrichment analysis (GSEA) based on GO (www.geneontology.org/) functions and Kyoto Encyclopedia of Genes and Genomes (KEGG; www.genome.jp/kegg/pathway.html). Protein-protein interaction network (PPI) analysis was performed to identify the association between the potential target genes on the online database STRING (https://string-db.org/; version 11.0).

Statistical analysis

Log (fold change) for each gene between the control and case groups was calculated before *T*-test, and *P*-value < 0.05 was considered statistically significant. While the identified genes were ranked according

to P-value, genes or pathways with *P*-value < 0.05 were selected as statistically significant ones. And GO: BP and KEGG pathways were ranked based on the number of genes enriched in pathways with statistical significance.

The schematic diagram of bioinformatics analysis was presented in Fig. 1.

Results

General characteristics

Among a total of 1144 samples with NSCLC gene expression data, 1014 with matched clinical data consisted of 955 cases (primary tumor) and 59 controls (normal solid tissue) with a median age of 66 years (range, 33–90); out of them, 609 cases were male, while 439 were diagnosed as squamous cell carcinoma. All the clinical characteristics were presented in Table 1.

Table 1
Demographic and clinical characteristics of samples in NSCLC and control cohorts

Category	N	NSCLC	Control
Gender, no. (%)	**1014**		
Female	405	381 (39.9)	24 (40.6)
Male	609	574 (60.1)	35 (59.4)
Age, years	**986**		
Max		90	86
Min		33	45
Mean		66.14	67.46
Med		67	68
TNM stage, no. (%)	**944**		
Stage		493 (52.2)	
Stage		265 (28.1)	
Stage		154 (16.3)	
Stage		32 (3.4)	
Data on distinct characteristics are missing in a small proportion of samples as follows: Gender (− 130), Age (− 158), TNM stage (− 141).			

DGE Analysis Of NSCLC Group

A total of 16,383 genes with significantly differential expression ($P < 0.05$, Log > 1) in NSCLC patients were identified. When log of 3 or greater and − 3 or lower with *P*-value < 0.001 were set as the thresholds, the number of DEGs was reduced from 16,383 to 2457, of which 2197 were upregulated, and 260 genes were downregulated in comparison with the control group (Fig. 2).

GWAS-derived VTE Risk Genes

The literature search led to identification of a total of 44 risk genes for VTE. Among them, the major genes play a dominant role in VTE, including SERPINC1, PROC, PROS1, and F5, while the minor genes exert a comprehensive effect towards VTE. The risk genes were listed in Table 2.

Table 2
VTE risk genes

Risk genes	Descriptions	Reference
A4GALT	Minor risk gene	[21]
ABO	Minor risk gene	[21–23]
ARID4A	Minor risk gene	[21]
BCOR	Minor risk gene	[22]
BRCC3	Minor risk gene	[22]
C1orf198	Minor risk gene	[21]
C4BPA	Minor risk gene	[22]
CATSPERB	Minor risk gene	[22]
CD93	Minor risk gene	[22]
CSRNP1	Minor risk gene	[22]
CYP27C1	Minor risk gene	[22]
EDEM2	Minor risk gene	[22]
EIF5A	Minor risk gene	[21]
EPCR	Minor risk gene	[24]
F10	Minor risk gene	[22]
F11	Minor risk gene	[21–23]
F2	Classic risk gene	[25]
F5	Minor risk gene	[21, 26]
F8	Minor risk gene	[21]
F9	Minor risk gene	[22]
FGG	Minor risk gene	[21–23]
GP6	Minor risk gene	[21, 22]
GRK5	Minor risk gene	[21, 22]
HLA-C	Minor risk gene	[22]
MPHOSPH9	Minor risk gene	[21]
MYRF	Minor risk gene	[22]
NUGGC	Minor risk gene	[21]

Risk genes	Descriptions	Reference
OSMR-AS1	Minor risk gene	[21]
PEPD	Minor risk gene	[22]
PLCG2	Minor risk gene	[21, 22]
PLEK	Minor risk gene	[21, 22]
PROC	Classic risk gene	[25–27]
PROCR	Minor risk gene	[21, 23]
PROS1	Classic risk gene	[22, 25]
RP11-122C5.1	Minor risk gene	[22]
SBNO1	Minor risk gene	[22]
SCARA5	Minor risk gene	[22]
SERPINC1	Classic risk gene	[25–27]
SLC44A2	Minor risk gene	[21–23]
SMG6	Minor risk gene	[21, 22]
STXBP5	Minor risk gene	[21, 22]
TSPAN15	Minor risk gene	[21–23]
VWF	Minor risk gene	[21, 22]
ZFPM2	Minor risk gene	[21–23]

Potential Target Genes Of Nsclc Patients Complicated With VTE

VTE risk genes with differential expression among DGE analysis in NSCLC group were considered potential target genes of NSCLC patients with VTE. The results for DGE analysis among NSCLC group and the risk genes for VTE were comprehensively arranged. Here, a total of 22 genes were selected as the target genes (Table 3).

Table 3
Potential risk genes of NSCLC patients with VTE

Genes	log2FoldChange (NSCLC)	P-adj (NSCLC)
C1orf198	1.429057666	5.51E-106
C4BPA	1.475964587	2.31E-09
CATSPERB	-1.671113149	7.23E-24
CD93	2.583538868	7.35E-143
CSRNP1	2.79589866	5.40E-170
CYP27C1	-3.238089143	3.38E-66
F10	1.945099423	3.92E-48
F11	3.411995172	1.00E-41
F2	-4.724801311	3.46E-33
F5	-1.691358458	9.02E-18
F8	2.250624177	5.83E-146
GRK5	2.607366709	9.39E-191
MPHOSPH9	-1.058150144	5.16E-52
MYRF	2.229430068	1.80E-39
PLEK	-2.344300302	1.82E-86
PROS1	1.561909917	2.03E-53
SCARA5	3.06624216	2.47E-45
SLC44A2	1.059030594	1.53E-69
SMG6	-1.611735302	0.004736288
STXBP5	-5.150809183	2.61E-58
VWF	2.761452583	7.57E-141
ZFPM2	-5.313752823	6.27E-158

GSEA Analysis Of The Potential Target Genes

GSEA analysis of the 22 potential target genes revealed 16 significantly enriched pathways that were subsequently ranked according to the number of genes included. And the top 5 pathways involved

complement and coagulation cascades, blood coagulation, blood coagulation-intrinsic pathway, platelet degranulation, and vesicle-mediated ER to Golgi transport (Fig. 3).

PPI Networks Of The Potential Target Genes

PPI network analysis was performed on 21 out of the 22 potential targets genes, excluding C1orf198 whose gene function and encoded protein are not completely understood, to determine the associations of their encoded proteins. As shown in Fig. 4, among the 21 encoded proteins, VWF, F2, F8, and other four blood coagulation associated proteins were positioned in the central part of the sub-network, while being associated with many other encoded proteins.

Discussion

Lung cancer is the leading cause of cancer death worldwide and has been associated with VTE in 7.3–13.6% of patients with lung cancer[8, 9, 28]. Many factors have been shown to be involved in the formation of VTE, including individual factors and tumor-related factors. A large prospective study showed that the elder patients had a 10-fold higher incidence of VTE than the young patients[29]. Moreover, gender and ethnicity can serve as the risk factors[30, 31]. In a large retrospective study conducted by Khorana et al, African American patients had the highest rate of VTE with about 5%, while Asian/Pacific Islander patients displayed the lowest rate of VTE with only 3%[31]. Besides, surgery and chemotherapy for cancer could result in an increase in the risk of VTE. Notably, cancer patients have a 6 to 7-fold increased risk of developing chemotherapy-associated thrombosis[32]. The first evidence came from cisplatin, a platinum-based chemotherapy drug widely used in many tumors and usually administered in combination with other drugs[33]. Other immunosuppressive or cytotoxic agents, such as tamoxifen, L-asparganase, and lenalidomide have also been reported to increase the risk[34]. Recent studies have indicated that certain driver mutations of NSCLC are related to VTE, including those in anaplastic lymphoma kinase (ALK)[14] and ROS-1[13].

However, the underlying mechanisms and biomarkers related to VTE in NSCLC are yet to be fully understood. To our knowledge, this study is the first to employ the bioinformatics strategy to identify DEGs and possible pathways in NSCLC that complicated with VTE. The findings in the study indicated that most of potential risk genes of NSCLC and VTE are significantly enriched in coagulation associated pathways, including complement and coagulation cascades (8 genes), blood coagulation (8), blood coagulation intrinsic pathway (5), and platelet degranulation (5), while genes F8, VWF, F5, F2 and F10 are more important in the PPI network than other genes within most of the pathways(Fig. 4). Thus, we reason that proteins encoded by the above-mentioned genes are connected with plenty of other proteins, acting as the vital nodes in the network. In these cases, F8 gene encodes coagulation factor VIII that participates in the coagulation by adhering to blood platelets in healthy people, whereas a higher activity of coagulation FXIII has been detected in cancer patients in which the tumor cell acts as a stimulus for coagulation, supporting metastatic progression by limiting NK clearance[35, 36]. Here, we showed a significantly increased expression of F8 gene in NSCLC patients (Log Fold Change = 2.25, p = 5.83E-146),

which was in accordance with the previous study[35]. Prothrombin protein, also known as coagulation FII, encoded by F2 gene, enhances tumor cell proliferation and tumor cell adherence to platelets, promoting tumor angiogenesis and the reconstruction of tumor microenvironment by up-regulating the expression of VEGF, MMP-2, IL-8 via MAPK, ERK and PI3-K signaling pathways, and ultimately leading to increased tumor growth, invasion and metastasis[37, 38].

Our study shows the upregulation of F2 and F5 gene in VTE, indicating a high risk of VTE, while they are downregulated in NSCLC. The possible reason is that some NSCLC patients were treated with chemotherapy and radiotherapy, which may cause impaired synthesis of normal clotting factors in liver. Another opposite regulated gene in our study is PROS1, which is downregulated in VTE but upregulated in NSCLC. PROS1 encodes vitamin K-dependent plasma protein S, which acts as an anticoagulant protease cofactor to activate protein C and inhibits blood coagulation. Thus, the decreased expression of S protein will lead to an increased risk of thrombosis. However, the explanation for the upregulation of PROS1 in NSCLC may due to tumor associated macrophages and glial cells can also secrete protein S[39].

A recent study analyzed the differential gene expression and pathway enrichment among 24 colorectal cancer patients with or without VTE, and found that the coagulation system pathway was associated with VTE[40]. Another study involving 12 lung cancer patients with or without VTE identified 1037 genes with differential expression[41]. In this study, the identified top genes with the highest Log were associated with innate immunity and inflammatory pathways, suggesting a significance of the immunity and inflammatory pathways for thrombosis of lung cancer patients. Although the present study revealed a difference between the differentially expressed genes and potential risk genes, multiple studies including the present one demonstrated that thrombosis, the increased coagulation factors, and abnormal coagulation can serve as risk factors for developing VTE in cancer patients.

The findings in this study suggest that coagulation, the pathological basis of the thrombotic events, is associated with the occurrence of VTE in NSCLC patients based on bioinformatics analysis at the gene level. Moreover, we identified and partly discussed the potential risk genes for developing VTE in NSCLC patients in the study. More importantly, we employed multiple flexible bioinformatics approaches including DGE, PPI and pathway enrichment analysis, together with an open-source bioinformatics database to undertake statistical analysis on the limited samples.

There are some limitations in the present study. Firstly, a limited number of transcriptome data for NSCLC were enrolled, while no transcriptome data for VTE or VTE with NSCLC were used in the study due to limited data sources. Secondly, all data used in the study are obtained from open sources, consisting of NSCLC patients and VTE risk genes separately, which lack of specificity towards the NSCLC patients with VTE. Thirdly, the risk genes were only analyzed and discussed in the level of transcriptome and genome, while further studies at different levels such as proteome, methylome, and metabolome are required. Given that this study is limited to preliminary screening results, further study based on large samples from NSCLC with VTE expanding multi-omics is needed to identify risk factors for developing thrombotic events in NSCLC patients.

Conclusions

This study reveals that potential risk genes of NSCLC-associated VTE are enriched in blood coagulation related pathways, suggesting a vital role of the blood coagulation in development of VTE in NSCLC patients. Our findings may lead to identification of potential targets for prevention and therapy of NSCLC-associated VTE.

Abbreviations

NSCLC: non-small cell lung cancer ; VTE: venous thromboembolism; DVT: deep vein thrombosis; PE: pulmonary embolism; PFS: progression-free survival; OS: overall survival; SCLC: small-cell lung cancer; ALK: anaplastic lymphoma kinase; PPI: protein interaction network; DGE: differentially gene expression; GWAS: Genome-Wide-Association Studies; GSEA: Gene Set Enrichment Analysis.

Declarations

Ethics approval and consent to participate

Ethical approval: This article does not contain any studies with human participants or animals performed by any of the authors.

Consent for publication

Written informed consent for publication was obtained from all participants.

Availability of data and materials

All data used for the systematic review and meta-analysis is present in the main manuscript in the Table.

Competing interests

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Funding

This research was supported by the Science Foundation of Peking University Cancer Hospital□2020-27□.

Author’s contribution

SH: study design and drafting the manuscript, QL and YR: literature review and analysis, XM and JF: critical review and finalizing the manuscript. All authors have read and approved the manuscript for submission.

Acknowledgement

Not applicable.

References

1. Wun T, White RH: **Venous thromboembolism (VTE) in patients with cancer: epidemiology and risk factors**. *Cancer Invest* 2009, **27 Suppl 1**:63-74.
2. Corrales-Rodriguez L, Blais N: **Lung cancer associated venous thromboembolic disease: a comprehensive review**. *Lung Cancer* 2012, **75**(1):1-8.
3. Sorensen HT, Mellemkjaer L, Olsen JH, Baron JA: **Prognosis of cancers associated with venous thromboembolism**. *N Engl J Med* 2000, **343**(25):1846-1850.
4. Torre LA, Bray F, Siegel RL, Ferlay J, Lortet-Tieulent J, Jemal A: **Global cancer statistics, 2012**. *CA Cancer J Clin* 2015, **65**(2):87-108.
5. Siegel RL, Miller KD, Jemal A: **Cancer statistics, 2020**. *CA Cancer J Clin* 2020, **70**(1):7-30.
6. Osmani L, Askin F, Gabrielson E, Li QK: **Current WHO guidelines and the critical role of immunohistochemical markers in the subclassification of non-small cell lung carcinoma (NSCLC): Moving from targeted therapy to immunotherapy**. *Semin Cancer Biol* 2018, **52**(Pt 1):103-109.
7. Vathiotis I, Dimakakos EP, Boura P, Ntineri A, Charpidou A, Gerotziafas G, Syrigos K: **Khorana Score: Nuew Predictor of Early Mortality in Patients With Lung Adenocarcinoma**. *Clin Appl Thromb Hemost* 2018, **24**(8):1347-1351.
8. Tagalakis V, Levi D, Agulnik JS, Cohen V, Kasymjanova G, Small D: **High risk of deep vein thrombosis in patients with non-small cell lung cancer: a cohort study of 493 patients**. *J Thorac Oncol* 2007, **2**(8):729-734.
9. Hicks LK, Cheung MC, Ding K, Hasan B, Seymour L, Le Maitre A, Leighl NB, Shepherd FA: **Venous thromboembolism and nonsmall cell lung cancer: a pooled analysis of National Cancer Institute of Canada Clinical Trials Group trials**. *Cancer* 2009, **115**(23):5516-5525.
10. Khorana AA, Kuderer NM, Culakova E, Lyman GH, Francis CW: **Development and validation of a predictive model for chemotherapy-associated thrombosis**. *Blood* 2008, **111**(10):4902-4907.
11. Prandoni P, Lensing AW, Piccioli A, Bernardi E, Simioni P, Girolami B, Marchiori A, Sabbion P, Prins MH, Noventa F *et al*: **Recurrent venous thromboembolism and bleeding complications during anticoagulant treatment in patients with cancer and venous thrombosis**. *Blood* 2002, **100**(10):3484-3488.
12. Ay C, Unal UK: **Epidemiology and risk factors for venous thromboembolism in lung cancer**. *Curr Opin Oncol* 2016, **28**(2):145-149.
13. Chiari R, Ricciuti B, Landi L, Morelli AM, Delmonte A, Spitaleri G, Cortinovis DL, Lamberti G, Facchinetti F, Pilotto S *et al*: **ROS1-rearranged Non-small-cell Lung Cancer is Associated With a High Rate of Venous Thromboembolism: Analysis From a Phase II, Prospective, Multicenter, Two-arms Trial (METROS)**. *Clin Lung Cancer* 2020, **21**(1):15-20.

14. Zer A, Moskovitz M, Hwang DM, Hershko-Klement A, Fridel L, Korpanty GJ, Dudnik E, Peled N, Shochat T, Leighl NB *et al*: **ALK-Rearranged Non-Small-Cell Lung Cancer Is Associated With a High Rate of Venous Thromboembolism**. *Clin Lung Cancer* 2017, **18**(2):156-161.
15. Tabchi S, Kourie HR, Klastersky J: **Concurrent driver mutations/rearrangements in non-small-cell lung cancer**. *Curr Opin Oncol* 2017, **29**(2):118-122.
16. Chen R, Li WX, Sun Y, Duan Y, Li Q, Zhang AX, Hu JL, Wang YM, Gao YD: **Comprehensive Analysis of lncRNA and mRNA Expression Profiles in Lung Cancer**. *Clin Lab* 2017, **63**(2):313-320.
17. Goh F, Yang IA, Bowman RV, Fong KM: **Subtype variation and actionability of telomere length abnormality in lung cancer**. *Transl Lung Cancer Res* 2018, **7**(Suppl 3):S251-S253.
18. Rhodes DR, Yu J, Shanker K, Deshpande N, Varambally R, Ghosh D, Barrette T, Pandey A, Chinnaiyan AM: **ONCOMINE: a cancer microarray database and integrated data-mining platform**. *Neoplasia* 2004, **6**(1):1-6.
19. Gentleman RC, Carey VJ, Bates DM, Bolstad B, Dettling M, Dudoit S, Ellis B, Gautier L, Ge Y, Gentry J *et al*: **Bioconductor: open software development for computational biology and bioinformatics**. *Genome Biol* 2004, **5**(10):R80.
20. Jiao X, Sherman BT, Huang da W, Stephens R, Baseler MW, Lane HC, Lempicki RA: **DAVID-WS: a stateful web service to facilitate gene/protein list analysis**. *Bioinformatics* 2012, **28**(13):1805-1806.
21. Lindstrom S, Wang L, Smith EN, Gordon W, van Hylckama Vlieg A, de Andrade M, Brody JA, Pattee JW, Haessler J, Brumpton BM *et al*: **Genomic and transcriptomic association studies identify 16 novel susceptibility loci for venous thromboembolism**. *Blood* 2019, **134**(19):1645-1657.
22. Klarin D, Busenkell E, Judy R, Lynch J, Levin M, Haessler J, Aragam K, Chaffin M, Haas M, Lindstrom S *et al*: **Genome-wide association analysis of venous thromboembolism identifies new risk loci and genetic overlap with arterial vascular disease**. *Nat Genet* 2019, **51**(11):1574-1579.
23. Germain M, Chasman DI, de Haan H, Tang W, Lindstrom S, Weng LC, de Andrade M, de Visser MC, Wiggins KL, Suchon P *et al*: **Meta-analysis of 65,734 individuals identifies TSPAN15 and SLC44A2 as two susceptibility loci for venous thromboembolism**. *Am J Hum Genet* 2015, **96**(4):532-542.
24. Li YY, Wu JJ, Yang XX, Geng HY, Gong G, Kim HJ: **EPCR Gene Ser219Gly Polymorphism and Venous Thromboembolism: A Meta-Analysis of 9,494 Subjects**. *Front Physiol* 2017, **8**:339.
25. Mannucci PM, Franchini M: **Classic thrombophilic gene variants**. *Thromb Haemost* 2015, **114**(5):885-889.
26. Rosendaal FR, Reitsma PH: **Genetics of venous thrombosis**. *J Thromb Haemost* 2009, **7 Suppl 1**:301-304.
27. Zoller B, Garcia de Frutos P, Hillarp A, Dahlback B: **Thrombophilia as a multigenic disease**. *Haematologica* 1999, **84**(1):59-70.
28. Blom JW, Osanto S, Rosendaal FR: **The risk of a venous thrombotic event in lung cancer patients: higher risk for adenocarcinoma than squamous cell carcinoma**. *J Thromb Haemost* 2004, **2**(10):1760-1765.

29. Tsai AW, Cushman M, Rosamond WD, Heckbert SR, Polak JF, Folsom AR: **Cardiovascular risk factors and venous thromboembolism incidence: the longitudinal investigation of thromboembolism etiology**. *Arch Intern Med* 2002, **162**(10):1182-1189.
30. Khorana AA, Francis CW, Culakova E, Fisher RI, Kuderer NM, Lyman GH: **Thromboembolism in hospitalized neutropenic cancer patients**. *J Clin Oncol* 2006, **24**(3):484-490.
31. Khorana AA, Francis CW, Culakova E, Kuderer NM, Lyman GH: **Frequency, risk factors, and trends for venous thromboembolism among hospitalized cancer patients**. *Cancer* 2007, **110**(10):2339-2346.
32. Khorana AA, Dalal M, Lin J, Connolly GC: **Incidence and predictors of venous thromboembolism (VTE) among ambulatory high-risk cancer patients undergoing chemotherapy in the United States**. *Cancer* 2013, **119**(3):648-655.
33. Doll DC, List AF, Greco FA, Hainsworth JD, Hande KR, Johnson DH: **Acute vascular ischemic events after cisplatin-based combination chemotherapy for germ-cell tumors of the testis**. *Ann Intern Med* 1986, **105**(1):48-51.
34. Heit JA: **Epidemiology of venous thromboembolism**. *Nat Rev Cardiol* 2015, **12**(8):464-474.
35. Lee SH, Suh IB, Lee EJ, Hur GY, Lee SY, Lee SY, Shin C, Shim JJ, In KH, Kang KH *et al*: **Relationships of coagulation factor XIII activity with cell-type and stage of non-small cell lung cancer**. *Yonsei Med J* 2013, **54**(6):1394-1399.
36. Palumbo JS, Potter JM, Kaplan LS, Talmage K, Jackson DG, Degen JL: **Spontaneous hematogenous and lymphatic metastasis, but not primary tumor growth or angiogenesis, is diminished in fibrinogen-deficient mice**. *Cancer Res* 2002, **62**(23):6966-6972.
37. Ay C, Vormittag R, Dunkler D, Simanek R, Chiriac AL, Drach J, Quehenberger P, Wagner O, Zielinski C, Pabinger I: **D-dimer and prothrombin fragment 1 + 2 predict venous thromboembolism in patients with cancer: results from the Vienna Cancer and Thrombosis Study**. *J Clin Oncol* 2009, **27**(25):4124-4129.
38. Radjabi AR, Sawada K, Jagadeeswaran S, Eichbichler A, Kenny HA, Montag A, Bruno K, Lengyel E: **Thrombin induces tumor invasion through the induction and association of matrix metalloproteinase-9 and beta1-integrin on the cell surface**. *J Biol Chem* 2008, **283**(5):2822-2834.
39. Sadahiro H, Kang KD, Gibson JT, Minata M, Yu H, Shi J, Chhipa R, Chen Z, Lu S, Simoni Y *et al*: **Activation of the Receptor Tyrosine Kinase AXL Regulates the Immune Microenvironment in Glioblastoma**. *Cancer Res* 2018, **78**(11):3002-3013.
40. Unlu B, van Es N, Arindrarto W, Kielbasa SM, Mei H, Westerga J, Middeldorp S, Kuppen PJK, Otten J, Cannegieter S *et al*: **Genes associated with venous thromboembolism in colorectal cancer patients**. *J Thromb Haemost* 2018, **16**(2):293-302.
41. Sussman TA, Abazeed ME, McCrae KR, Khorana AA: **RNA expression and risk of venous thromboembolism in lung cancer**. *Res Pract Thromb Haemost* 2020, **4**(1):117-123.

Figures

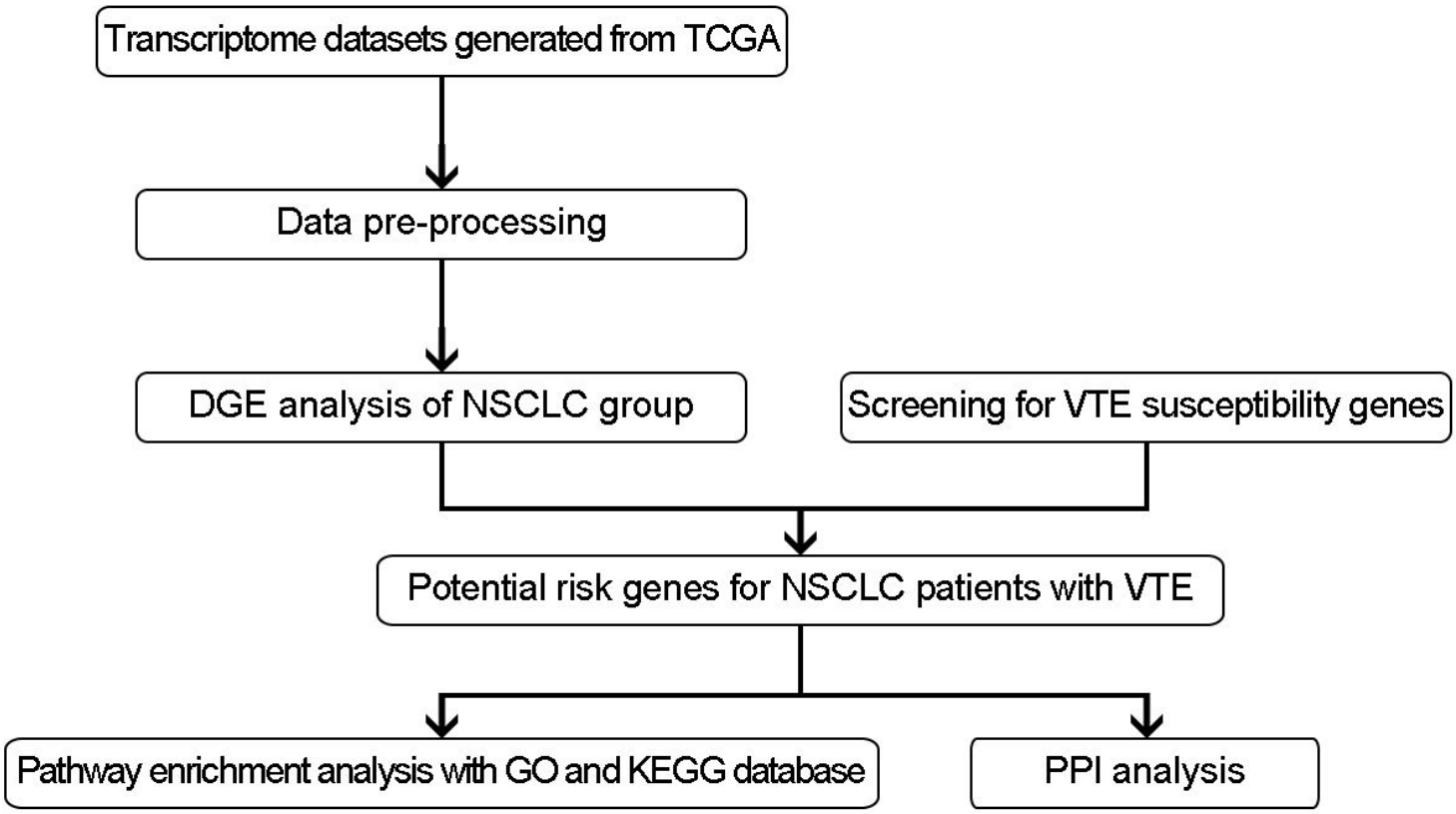

Figure 1

The schematic diagram of the bioinformatics analysis.

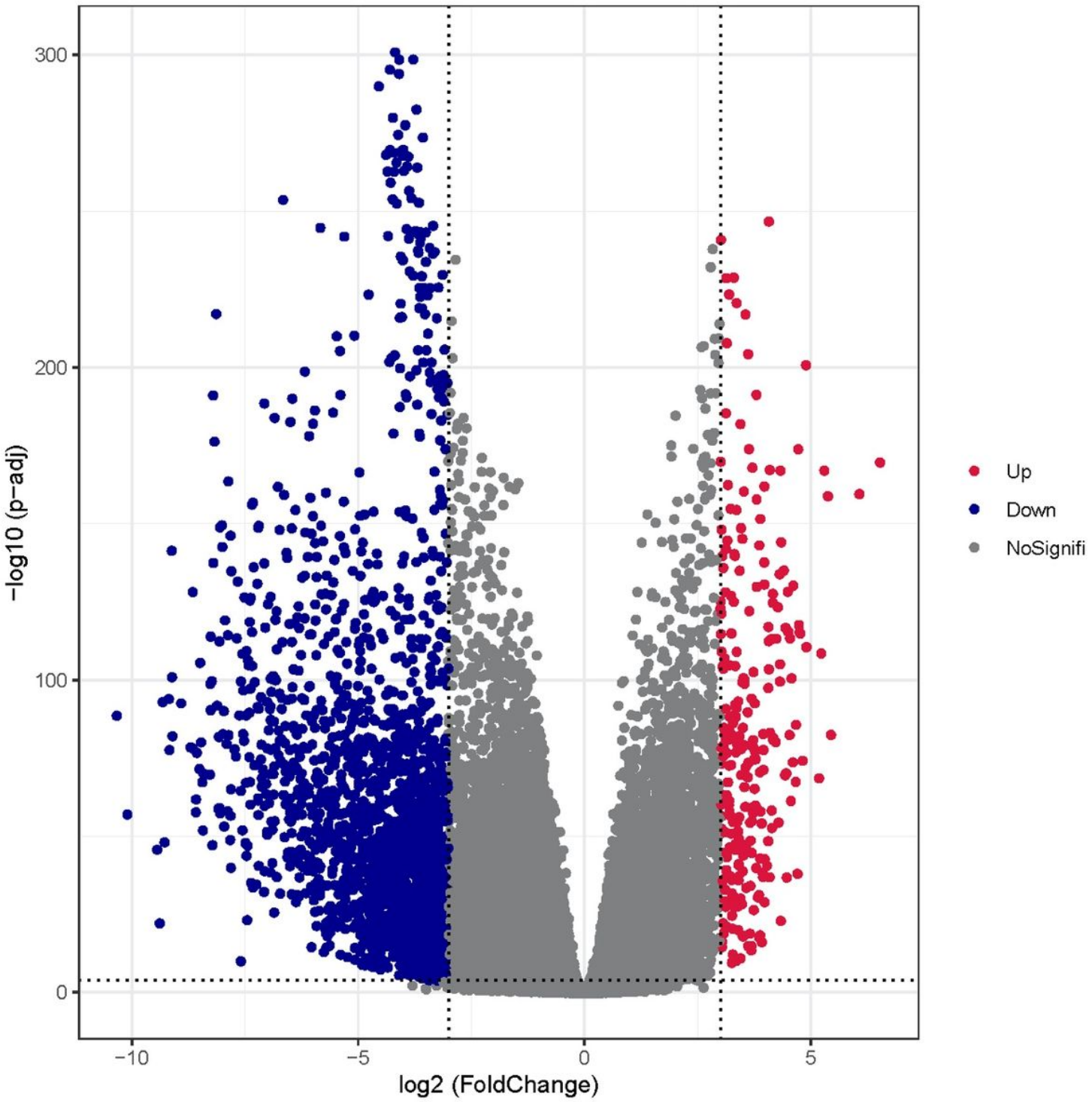

Figure 2

Identification of 2457 differentially expressed genes between NSCLC group and the control using R program. Among these DGEs, 260 are downregulated, and 2197 are upregulated.

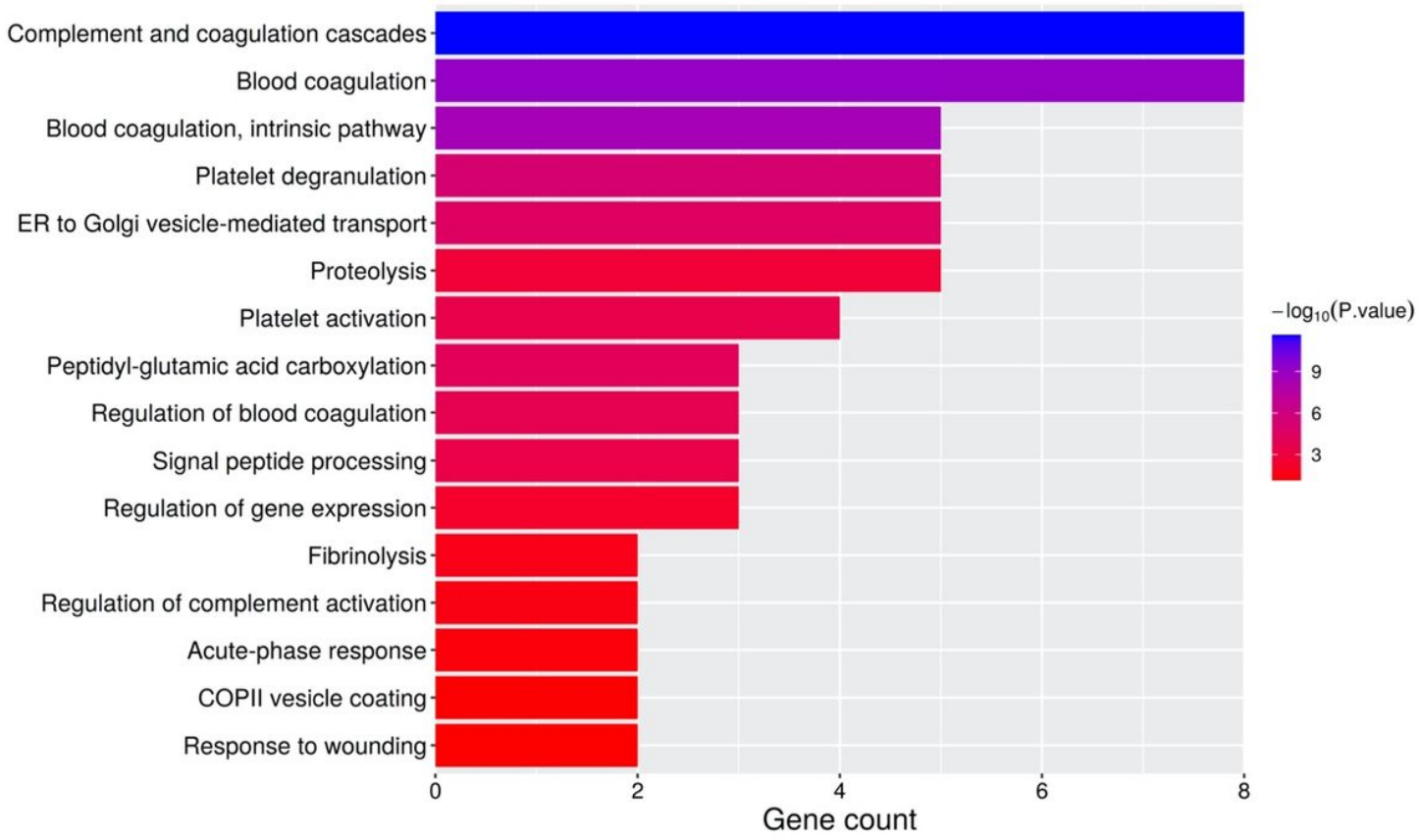

Figure 3

Enriched pathways of potential risk genes of NSCLC patients with VTE.

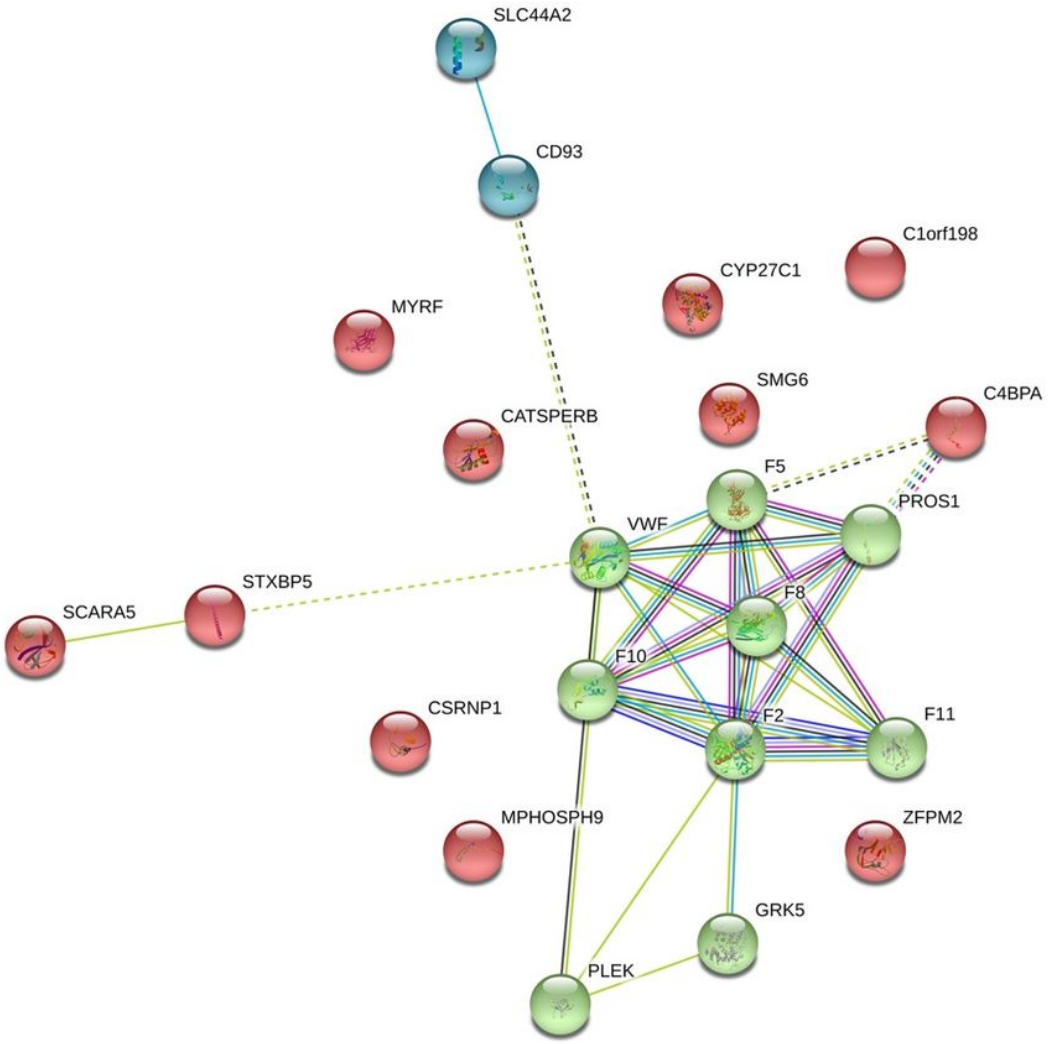

Figure 4

The PPI network of potential risk genes of NSCLC patients with VTE.