Observational Study

Medicine®

OPEN

Identification of biomarkers of venous thromboembolism by bioinformatics analyses

Guiming Wang, MD[a], Wenbo Zhao, MD[a], Yudong Yang, MM[a], Gaochao Yang, MM[a], Zhigang Wei, MD[b,*], Jiansheng Guo, MM[b,*]

Abstract

Venous thromboembolism (VTE) is a common vascular disease and a major cause of mortality. This study intended to explore the biomarkers associated with VTE by bioinformatics analyses.

Based on Gene Expression Omnibus (GEO) database, the GSE19151 expression profile data were downloaded. The differentially expressed genes (DEGs) between single VTE (sVTE)/recurrent VTE (rVTE) and control were identified. Then, pathway enrichment analysis of DEGs were performed, followed by protein–protein interaction (PPI) network construction.

Total 433 upregulated and 222 downregulated DEGs were obtained between sVTE and control samples. For rVTE versus control, 625 upregulated and 302 downregulated DEGs were identified. The overlap DEGs were mainly enriched in the pathways related to ribosome, cancer, and immune disease. The DEGs specific to rVTE were enriched in several pathways, such as nod-like receptor signaling pathway. In the PPI network, 2 clusters of VTE genes, including ribosomal protein family genes and NADH family-ubiquinol-cytochrome genes, were identified, such as ribosomal protein L9 (*RPL9*), *RPL5*, *RPS20*, *RPL23*, and tumor protein p53 (*TP53*).

The nod-like receptor signaling pathway, ribosomal protein family genes, such as *RPL9*, *RPL5*, *RPS20*, and *RPL23*, and DEG of *TP53* may have the potential to be used as targets for diagnosis and treatment of VTE.

Abbreviations: DEG = differentially expressed gene, GEO = Gene Expression Omnibus, IL = interleukin, PPI = protein–protein interaction, RPL9 = ribosomal protein L9, rVTE = recurrent venous thromboembolism, sVTE = single venous thromboembolism, TP53 = tumor protein p53, VTE = venous thromboembolism.

Keywords: differentially expressed gene, pathway enrichment analysis, protein–protein interaction network, venous thromboembolism

1. Introduction

Venous thromboembolism (VTE) is one of the most common vascular disease, as well as a major cause of mortality.[1,2] The incidence of VTE increases dramatically over 45 years old.[3] Its annual incidence is about 1 or 2 cases per 1000 persons in the general population.[4] The most frequent clinical manifestations are deep vein thrombosis of the leg, and pulmonary embolism. VTE is a chronic disease with acute exacerbations, and the major outcomes of VTE are recurrence, postthrombotic syndrome, and death.[1] Approximately, as many as 30% patients develop to recurrent venous thromboembolism (rVTE) within 8 years after stopping a standard anticoagulant therapy.[5] In consideration of the increasing life expectancy, adopting strategies to prevent and treat VTE is becoming more and more important all over the world.[6]

VTE is a multicausal disease involving environmental and genetic risk factors.[7] Family and twin studies have found that genetics play important roles in the occurrence of VTE.[8] An increasing number of genetic variants involved in the fibrinolysis and coagulation pathways are consistently linked to VTE.[9–11] Reitsma and Rosendaal[12] reported that some inflammatory mediators, such as interleukin (IL)-6 and tumor necrosis factor α, represented risk determinants for VTE. In addition, a recent study suggested that tissue factor and tissue factor pathway inhibitor played key roles in coagulation and acted as direct risk factors for VTE.[13] Although progresses have been achieved for the pathogenesis of VTE, the genetic mechanisms are far from being understood. Determination of the recurrence risk factors for VTE is still a vital health concern.[2]

In the present study, the microarray data GSE19151[2] including single venous thromboembolism (sVTE), rVTE, and control samples were downloaded from Gene Expression Omnibus (GEO) database for differential expression analysis. In addition, functional enrichment analysis and protein–protein interaction (PPI) network construction were performed for differentially expressed genes (DEGs). The present study aimed to explore the genes and pathways associated with the occurrence and development of sVTE and rVTE, which may be helpful to find out novel biomarkers for the diagnosis and treatment of VTE.

Editor: Ovidiu Constantin Baltatu.

GW and WZ are first coauthors.

The authors have no funding and conflicts of interest to disclose.

[a] Department of Vascular Surgery, [b] Department of Surgery, The First Hospital of Shanxi Medical University, Taiyuan, PR, China.

[*] Correspondence: Zhigang Wei and Jiansheng Guo, Department of Surgery, The First Hospital of Shanxi Medical University, No.85, Jiefang South Road, Yingze District, Taiyuan, Shanxi 030001, China (e-mails: weizhigangwz@hotmail.com, jianshengguo@sina.com).

Medicine (2018) 97:14(e0152)

Received: 22 December 2016 / Received in final form: 21 August 2017 / Accepted: 22 February 2018

http://dx.doi.org/10.1097/MD.0000000000010152

2. Methods

2.1. Affymetrix microarray data

The GSE19151 gene expression profile data were downloaded from GEO, (http://www.ncbi.nlm.nih.gov/geo/), a public functional genomics database. The gene expression profile platform was Affymetrix Human Genome U133A 2.0 Array (Affymetrix Inc., Santa Clara, CA). A total of 63 healthy control, 32 sVTE, and 38 rVTE were available for analysis.

2.2. Data preprocessing and DEGs identification

The original data were performed background correction, quantile normalization, and log transformation using the robust multiarray average algorithm.[14] The probe was removed when it corresponded to multiple gene symbols. Finally, the gene expression matrix including 12,754 genes was obtained.

The DEGs of sVTE versus control and rVTE versus control were screened out based on the limma package in R/Bioconductor.[15] The *P*-values regarding the significance of differences were calculated using paired-samples *t* test. The obtained *P*-values were adjusted with Benjamini and Hochberg (BH) method,[16] obtaining the false discovery rate (FDR). Additionally, the $\log_2$-fold change ($\log_2$FC) was calculated. $|\log_2 FC| > 1.5$ and $FDR < 0.05$ were considered as the cutoff values for DEGs screening.

2.3. Functional enrichment analysis

ClueGO is a plugin of cytoscape used to facilitate the biological interpretation and to visualize functionally grouped terms in the form of networks.[17] In the present study, we performed Kyoto Encyclopedia of Genes and Genomes pathway enrichment analysis based on the ClueGO plugin with *P*-value $< .05$ and kappa score $= 0.4$. The DEGs enriched in the pathways were displayed using CluePedia[18] plugin in cytoscape.

2.4. PPI network construction

We downloaded the comprehensive interaction information of human proteins from the Search Tool for the Retrieval of Interacting Genes database (http://string-db.org/).[19] The interaction relationship pairs of DEGs associated with VTE with combined score > 0.9 were extracted to construct the PPI network using cytoscape (http://www.cytoscape.org/).[20]

3. Results

3.1. Identification of DEGs

After preprocessing, 655 DEGs including 433 upregulated and 222 downregulated DEGs were obtained between sVTE and control samples. For rVTE versus control group, there were 625 upregulated and 302 downregulated DEGs. Additionally, 357 DEGs were upregulated and 176 ones were downregulated in both sVTE and rVTE samples (Fig. 1).

3.2. Functional enrichment analysis

As shown in Fig. 1, 357 overlapped upregulated and 176 overlapped downregulated DEGs were identified in between sVTE and rVTE samples. Functional enrichment analysis of these DEGs revealed that they were significantly enriched in

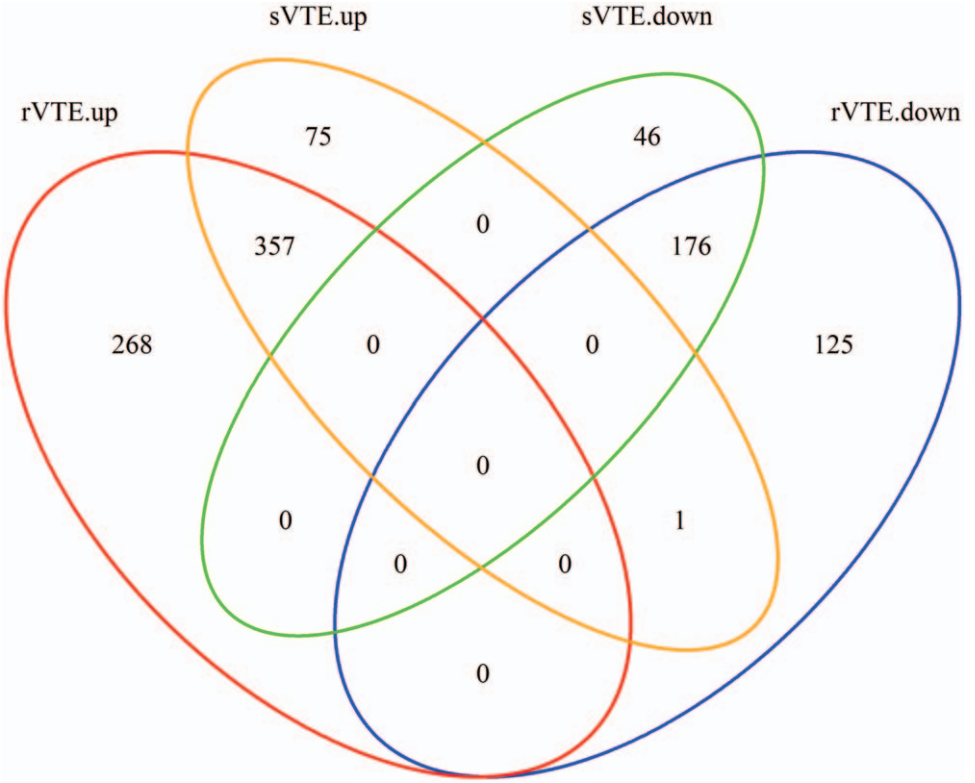

Figure 1. Venn plot of DEGs. Up represents upregulated DEGs; down represents downregulated DEGs. DEG = differentially expressed gene, sVTE = single venous thromboembolism, rVTE = recurrent venous thromboembolism.

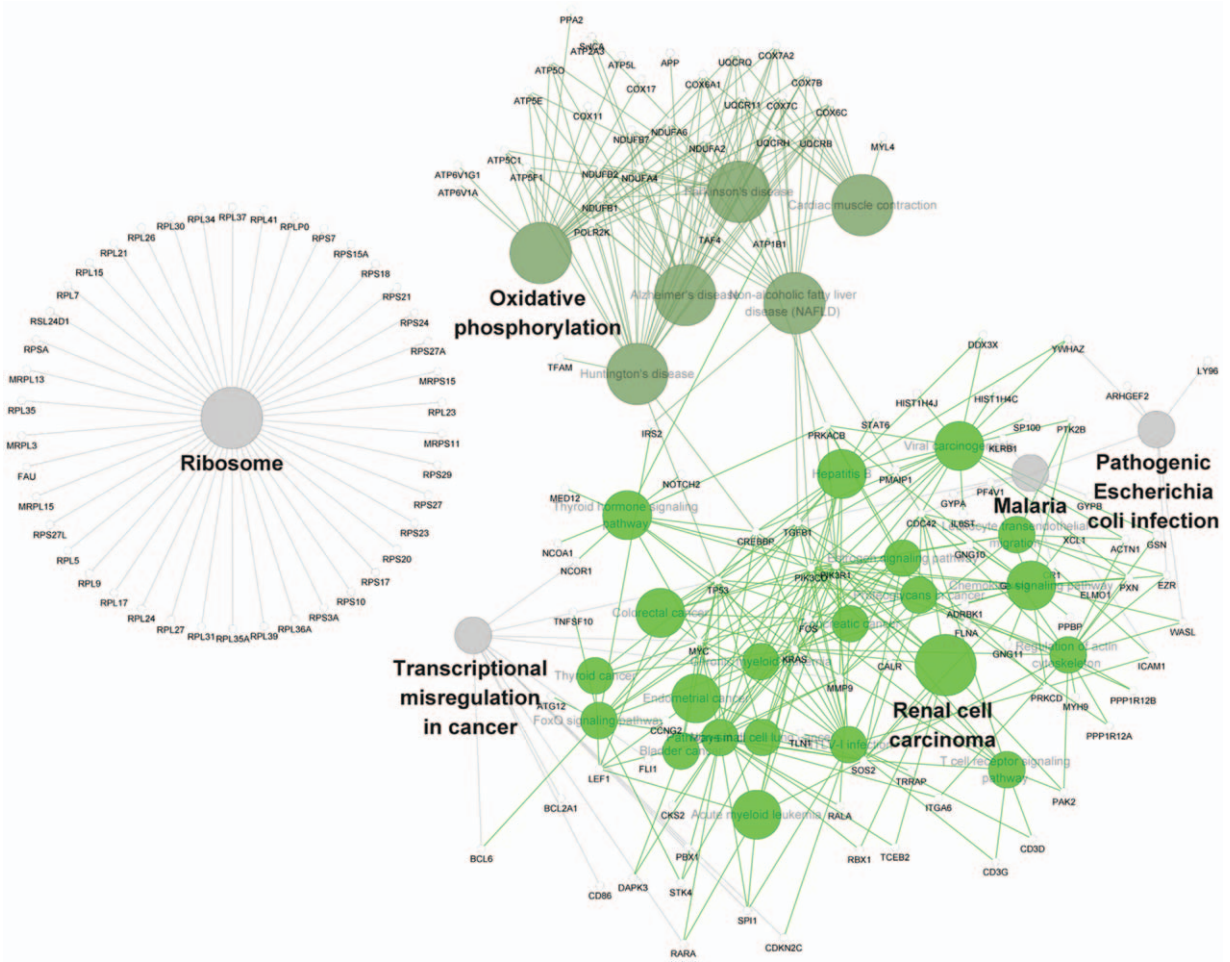

Figure 2. Pathways enriched by the overlap DEGs of rVTE versus control, and sVTE versus control. Larger dot represents pathway; small dot represents DEG enriched in the pathway. Pathway node size represents the significant level, and a bigger size represents a more significant level. Pathway dot color represents function group, and the same group has the same color. Gene node color is the same with its function group. DEG = differentially expressed gene, sVTE = single venous thromboembolism, rVTE = recurrent venous thromboembolism.

Table 1

The pathways enriched by overlapped up- and downregulated differentially expressed genes of sVTE and rVTE group.

	Pathway ID	Pathway name	Count	*P*
Up	hsa03010	Ribosome	42	9.56E-36
	hsa00190	Oxidative phosphorylation	24	2.37E-14
	hsa05012	Parkinson's disease	21	6.69E-11
	hsa05010	Alzheimer's disease	20	9.11E-09
	hsa05016	Huntington's disease	21	1.51E-08
	hsa04932	Nonalcoholic fatty liver disease (NAFLD)	17	3.62E-07
	hsa04260	Cardiac muscle contraction	11	8.03E-06
	hsa04062	Chemokine signaling pathway	10	2.94E-02
Down	hsa05161	Hepatitis B	11	6.38E-05
	hsa05221	Acute myeloid leukemia	6	1.52E-03
	hsa04919	Thyroid hormone signaling pathway	8	1.66E-03
	hsa05200	Pathways in cancer	15	2.34E-03
	hsa05210	Colorectal cancer	6	2.40E-03
	hsa05205	Proteoglycans in cancer	10	3.25E-03
	hsa04510	Focal adhesion	10	3.95E-03
	hsa04810	Regulation of actin cytoskeleton	10	4.63E-03
	hsa05166	HTLV-I infection	11	5.33E-03
	hsa05213	Endometrial cancer	5	7.79E-03

rVTE = recurrent venous thromboembolism, sVTE = single venous thromboembolism.

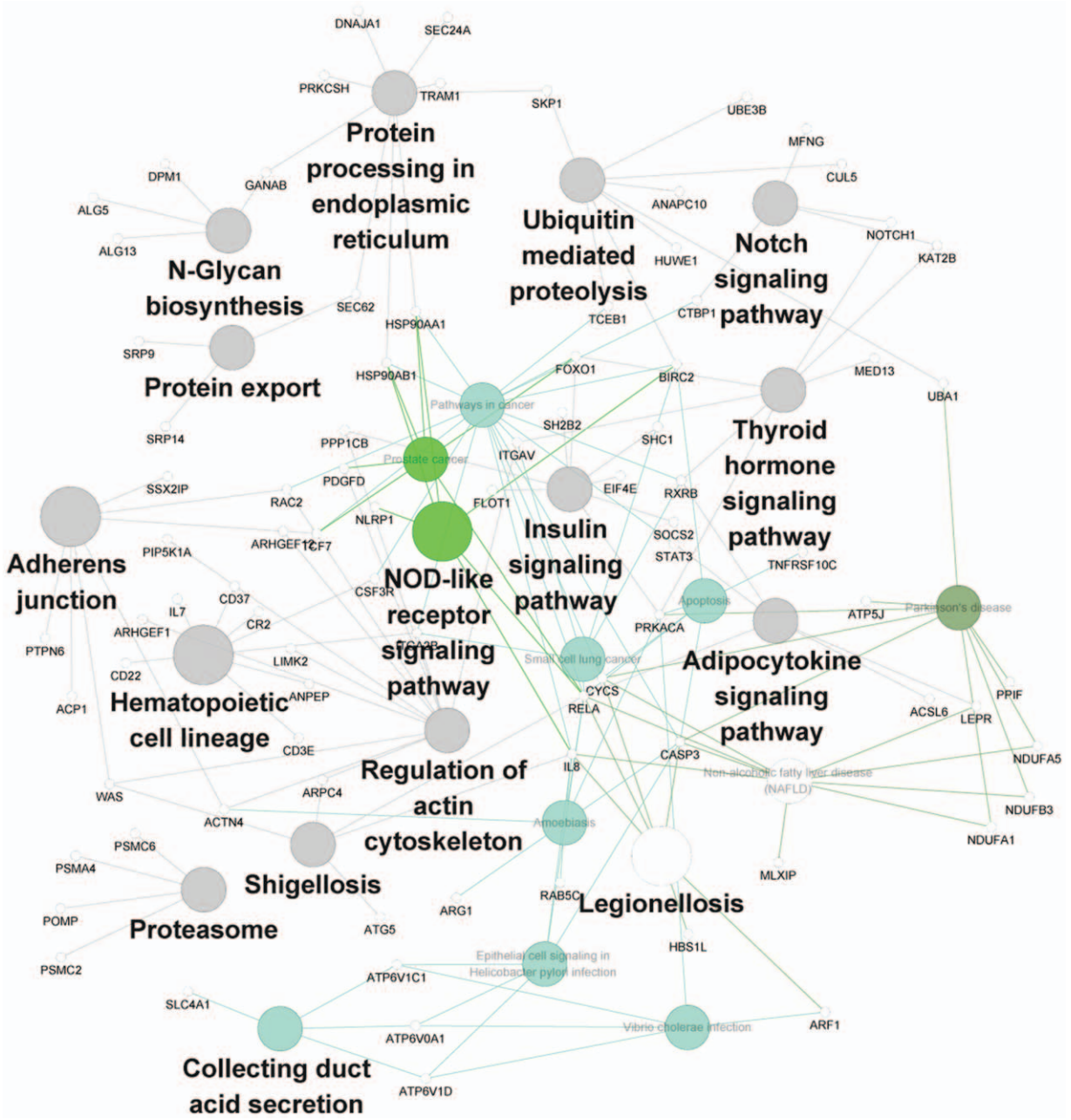

Figure 3. Pathways enriched by rVTE specific DEGs. Larger dot represents pathway; small dot represents DEG enriched in the pathway. Pathway node size represents the significant level, and a bigger size represents a more significant level. Pathway dot color represents function group, and the same group has the same color. Gene node color is the same with its function group. DEG = differentially expressed gene, sVTE = single venous thromboembolism, rVTE = recurrent venous thromboembolism.

some pathways which could be divided into several function groups including ribosome, oxidative phosphorylation, cancer, and immune disease. Specially, some DEGs were involved in multiple function groups, such as tumor protein p53 (*TP53*), transforming growth factor beta 1, and v-myc avian myelocytomatosis viral oncogene homolog (Fig. 2). In addition, we analyzed the pathways enriched by up- and downregulated DEGs, respectively. As shown in Table 1, the upregulated DEGs were significantly enriched in ribosome (hsa03010), oxidative phosphorylation (hsa00190), and Parkinson disease (hsa05012), and the downregulated DEGs were involved in hepatitis B (hsa05161), acute myeloid leukemia (hsa05221), thyroid hormone signaling pathway (hsa04919), and pathways in cancer (hsa05200).

3.3. rVTE-specific pathway

After functional annotation for the 393 DEGs specific to rVTE, we found that they were enriched in several pathways such as notch signaling pathway, nod-like receptor signaling pathway, and adherens junction (Fig. 3). Furthermore, the upregulated genes were enriched in pathways of ribosome, proteasome, and

Table 2

The pathways enriched by the up- and downregulated differentially expressed genes in recurrent venous thromboembolism (rVTE) group.

	Pathway ID	Pathway name	Count	*P*
Up	hsa03010	Ribosome	7	2.65E-02
	hsa03050	Proteasome	4	3.71E-02
	hsa05016	Huntington's disease	8	4.22E-02
Down	hsa05200	Pathways in cancer	14	2.68E-04
	hsa04640	Hematopoietic cell lineage	7	2.96E-04
	hsa04062	Chemokine signaling pathway	8	3.88E-03
	hsa04662	B cell receptor signaling pathway	5	6.50E-03
	hsa04144	Endocytosis	9	6.56E-03
	hsa04520	Adherens junction	5	7.19E-03
	hsa04810	Regulation of actin cytoskeleton	8	7.66E-03
	hsa04666	Fc gamma R-mediated phagocytosis	5	1.28E-02
	hsa04919	Thyroid hormone signaling pathway	5	3.49E-02

Huntington disease. Downregulated genes involved in pathways in cancer, hematopoietic cell lineage, chemokine signaling pathway, and adherens junction (Table 2).

3.4. PPI network construction

The PPI network was constructed with 466 nodes and 1356 edges. Among these nodes, 166 were specific to rVTE, 37 were specific to sVTE, and 263 were overlap genes. For the 1356 edges, 567 ones were specific to rVTE, 110 ones were specific to sVTE, and 679 ones belong to overlap genes. The PPI network included 2 clusters of VTE genes, including ribosomal protein family genes such as ribosomal protein L9 (*RPL9*), *RPL5*, *RPS20* and *RPL23*, and NADH family-ubiquinol-cytochrome genes. In addition, the top 5 hub genes of each class (overlap/rVTE/sVTE) were shown in Table 3, such as *RPL9*, *TP53*, and eukaryotic initiation factor 4e.

4. Discussion

In the present study, 655 DEGs were identified between sVTE and control samples, and 927 DEGs were obtained in rVTE versus control group. The overlap DEGs of the 2 comparison groups were significantly enriched in several pathway function groups: ribosome, cancer, and immune disease. *TP53* was involved in multiple function groups. The rVTE specific DEGs were mainly enriched in nod-like receptor signaling pathway and notch signaling pathway. *RPL9*, *TP53*, *RPL5*, *RPS20*, and *RPL23* were top 5 hub genes in the PPI network.

Table 3

Degree top 5 genes in each class.

Node	Type	Up/down	Degree
RPL9	Overlap	Up	40
TP53	Overlap	Down	39
RPL5	Overlap	Up	36
RPS20	Overlap	Up	34
RPL23	Overlap	Up	34
EIF4E	rVTE	Up	27
SRP14	rVTE	Up	24
RPS25	rVTE	Up	20
STAT3	rVTE	Down	19
KAT2B	rVTE	Up	18
RPL37A	sVTE	Up	25
NCOA3	sVTE	Up	9
MED13L	sVTE	Up	8
SF3B1	sVTE	Up	7
PCBP2	sVTE	Up	6

rVTE represents DEGs in recurrent venous thromboembolism, sVTE represents DEGs in single venous thromboembolism, and overlap represents the overlap DEGs of rVTE and sVTE. Up/down represents up/downregulation. DEG = differentially expressed gene, sVTE = single venous thromboembolism, rVTE = recurrent venous thromboembolism.

Among the top 5 hub genes in the PPI network, 4 ones including *RPL9*, *RPL5*, *RPS20*, and *RPL23* belonged to the ribosomal protein family, besides they were upregulated in both rVTE and sVTE samples. Early studies have identified ribosomes in platelet preparations.[21,22] The ribosomal protein RPS6, for instance, has been reported to present in circulating platelets, the unique cellular effectors of hemostasis and thrombosis.[23,24] Platelets support the accumulation of innate immune cells and promote the formation of neutrophil extracellular trap, which contribute to the propagation of deep vein thrombosis.[3] Currently, there are few reports about the relationships between these ribosomal protein family genes and thrombosis, so we speculate that DEGs of *RPL9*, *RPL5*, *RPS20*, and *RPL23* and VTE may be closely linked in light of their relationships with platelets.

In addition, the present study showed that *TP53* involved in multiple function groups, besides it was a downregulated hub gene in the PPI network. *TP53* encodes a tumor suppressor protein p53 which regulates cell cycle and negatively regulates cell division, and acts as a tumor suppressor in many tumor types.[25] Mutations of *TP53* are the most common genetic alterations in human tumors, which have been suggested as a molecular marker for the prognosis of tumor.[26] Interestingly, VTE occurs commonly in patients with cancer. For instance, the disseminated intravascular coagulation, a kind of clinical manifestation of VTE, is most commonly observed in patients with widespread metastatic cancer and hematological malignant disorders.[27] In addition, as Cermak et al[28] reported that tissue factor contributed to the development of thrombosis and acted as an direct risk factor for VTE. Importantly, tissue factor has been found to be a target of the inactivation of p53.[29] Therefore, we speculated that *TP53* might be a key biomarker of VTE.

Nod-like receptor signaling pathway was a significant pathway specific to rVTE. The nod-like receptors are a family of intracellular receptors that represent critical components of the innate immune responses and inflammation in mammals.[30] Chen et al[31] have reported that nod-like receptors regulate nuclear factor-kappa B signaling, and IL-1β production,

indicating their crucial role in the pathogenesis of inflammatory diseases. Specially, Fox and Kahn[32] reported that patients with VTE manifested 4 common signs of inflammation, namely redness, heat, pain, and swelling, which might indicate that inflammation play an important role in VTE. Furthermore, in our study, *IL8*, encoding a proinflammatory mediator, was found to enrich in the nod-like receptor signaling pathway. Specially, IL-8 has been reported to be capable of activating coagulation.[33] Taken together, nod-like receptor signaling pathway and *IL8* may play important roles in the progression of VTE.

In summary, we have used a comprehensive bioinformatics method to analyze the DEGs and pathways related to VTE. The nod-like receptor signaling pathway and its enriched DEGs including *IL8* may play important roles in the VTE progression. *RPL9*, *RPL5*, *RPS20*, TP53, and *RPL23* may have the potential to be used as targets for VTE diagnosis and treatment. However, further genetic and experimental studies with larger sample size are still needed to confirm our results.

References

[1] Visser M, Minkelen R, Marion V, et al. Genome-wide linkage scan in affected sibling pairs identifies novel susceptibility region for venous thromboembolism: Genetics In Familial Thrombosis study. J Thromb Haemost 2013;11:1474–84.

[2] Lewis DA, Stashenko GJ, Akay OM, et al. Whole blood gene expression analyses in patients with single versus recurrent venous thromboembolism. Thromb Res 2011;128:536–40.

[3] von Brühl M-L, Stark K, Steinhart A, et al. Monocytes, neutrophils, and platelets cooperate to initiate and propagate venous thrombosis in mice in vivo. J Exp Med 2012;209:819–35.

[4] Spencer FA, Emery C, Lessard D, et al. The worcester venous thromboembolism study: a population-based study of the clinical epidemiology of venous thromboembolism. J Gen Intern Med 2006; 21:722–7.

[5] Heit JA, Mohr DN, Silverstein MD, et al. Predictors of recurrence after deep vein thrombosis and pulmonary embolism: a population-based cohort study. Arch Intern Med 2000;160:761–8.

[6] Silverstein MD, Heit JA, Mohr DN, et al. Trends in the incidence of deep vein thrombosis and pulmonary embolism: a 25-year population-based study. Arch Intern Med 1998;158:585–93.

[7] Rosendaal F. Venous thrombosis: a multicausal disease. Lancet 1999; 353:1167–73.

[8] Larsen TB, Sørensen HT, Skytthe A, et al. Major genetic susceptibility for venous thromboembolism in men: a study of Danish twins. Epidemiology 2003;14:328–32.

[9] Blondon M, Hwang M, Smith NL. Genotyping in prothrombotic states: implications for the clinician. Curr Cardiovasc Risk Rep 2011;5:525–32.

[10] Germain M, Saut N, Greliche N, et al. Genetics of venous thrombosis: insights from a new genome wide association study. PloS One 2011;6: e25581.

[11] Heit JA, Armasu SM, Asmann YW, et al. A genome-wide association study of venous thromboembolism identifies risk variants in chromosomes 1q24.2 and 9q. J Thromb Haemost 2012;10:1521–31.

[12] Reitsma P, Rosendaal F. Activation of innate immunity in patients with venous thrombosis: the Leiden Thrombophilia Study. J Thromb Haemost 2004;2:619–22.

[13] Kwon A, Jo S-H, Jo Y-A, et al. Genetic polymorphisms and plasma levels of tissue factor and tissue factor pathway inhibitor in venous thromboembolism. Blood Coagul Fibrinolysis 2014;25: 416–21.

[14] Irizarry RA, Bolstad BM, Collin F, et al. Summaries of affymetrix GeneChip probe level data. Nucleic Acids Res 2003;31:e15–15.

[15] Smyth GK. Limma: linear models for microarray data. Bioinformatics and Computational Biology Solutions Using R and Bioconductor 2005; Berlin, Germany: Springer, 397–420.

[16] Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. J R Stat Soc Ser B (Methodol) 1995;289–300.

[17] Bindea G, Mlecnik B, Hackl H, et al. ClueGO: a Cytoscape plug-in to decipher functionally grouped gene ontology and pathway annotation networks. Bioinformatics 2009;25:1091–3.

[18] Bindea G, Galon J, Mlecnik B. CluePedia Cytoscape plugin: pathway insights using integrated experimental and in silico data. Bioinformatics 2013;29:661–3.

[19] Szklarczyk D, Franceschini A, Kuhn M, et al. The STRING database in 2011: functional interaction networks of proteins, globally integrated and scored. Nucleic Acids Res 2011;39(suppl 1): D561–8.

[20] Shannon P, Markiel A, Ozier O, et al. Cytoscape: a software environment for integrated models of biomolecular interaction networks. Genome Res 2003;13:2498–504.

[21] Kieffer N, Guichard J, Farcet JP, et al. Biosynthesis of major platelet proteins in human blood platelets. Eur J Biochem 1987; 164:189–95.

[22] Ts'ao Ch. Rough endoplasmic reticulum and ribosomes in blood platelets. Scand J Haematol 1971;8:134–40.

[23] Weyrich AS, Lindemann S, Tolley ND, et al. Change in protein phenotype without a nucleus: translational control in platelets. Paper presented at: Semin Thromb Hemost 2004.

[24] Zimmerman GA, Weyrich AS. Signal-dependent protein synthesis by activated platelets new pathways to altered phenotype and function. Arterioscler Thromb Vasc Biol 2008;28:s17–24.

[25] Livingstone LR, White A, Sprouse J, et al. Altered cell cycle arrest and gene amplification potential accompany loss of wild-type p53. Cell 1992;70:923–35.

[26] Kringen P, Wang Y, Dumeaux V, et al. TP53 mutations in ovarian carcinomas from sporadic cases and carriers of two distinct BRCA1 founder mutations; relation to age at diagnosis and survival. BMC Cancer 2005;5:134.

[27] Prandoni P, Falanga A, Piccioli A. Cancer and venous thromboembolism. Lancet Oncol 2005;6:401–10.

[28] Cermak J, Key N, Bach R, et al. C-reactive protein induces human peripheral blood monocytes to synthesize tissue factor. Blood 1993;82:513–20.

[29] Joanne LY, May L, Lhotak V, et al. Oncogenic events regulate tissue factor expression in colorectal cancer cells: implications for tumor progression and angiogenesis. Blood 2005;105: 1734–41.

[30] Fritz JH, Ferrero RL, Philpott DJ, et al. Nod-like proteins in immunity, inflammation and disease. Nat Immunol 2006;7:1250–7.

[31] Chen G, Shaw MH, Kim Y-G, et al. NOD-like receptors: role in innate immunity and inflammatory disease. Ann Rev Pathol Mech Dis 2009;4:365–98.

[32] Fox EA, Kahn SR. The relationship between inflammation and venous thrombosis. Thromb Haemost 2005;94:362–5.

[33] van der Poll T, Büller HR, ten Cate H, et al. Activation of coagulation after administration of tumor necrosis factor to normal subjects. N Engl J Med 1990;322:1622–7.