Whole Exome Sequencing Identifies genetic variants in Chinese Han pregnant women with Venous thromboembolism-a case control study

Yupei Shen
SIPPR: Shanghai Institute of Planned Parenthood Research
Yan Zhang
SIPPR: Shanghai Institute of Planned Parenthood Research
Ying Xiong
Shanghai Jiaotong University School of Medicine Xinhua Hospital
Zhiping Zhang
Shanghai Pudong New Area Healthcare Hospital for Women and Children
Baohua Zhang
Shanghai Punan Hospital
Aihong Li
Minghang Maternal and Child Hospital
Zhaofeng Zhang
SIPPR: Shanghai Institute of Planned Parenthood Research
Jing Du (dujing42@126.com)
SIPPR: Shanghai Institute of Planned Parenthood Research https://orcid.org/0000-0002-6767-2763
Yan Che
SIPPR: Shanghai Institute of Planned Parenthood Research

Research article

Keywords: KLF7, WDR72 and GRK5 genes may be associated with VTE in pregnancies and CDC7, MCM6 may be related to pathogenic process.

Posted Date: September 30th, 2020

DOI: https://doi.org/10.21203/rs.3.rs-83203/v1

Abstract

Background: Venous thromboembolism (VTE) is a common health problem, causing considerable morbidity and mortality. The incidence of VTE is higher in pregnant women than in those who are not pregnant. However, genetic factors for VTE in pregnant women are largely unknown.

Methods: We performed a large-scale prospective cohort study of 65138 pregnancies. Pregnant patients with VTE and pregnant women without VTE were enrolled in the study and sequenced by whole exome sequencing. Functional and enrichment analyses were performed using the DAVID online database. The protein-protein interaction network was constructed using the STRING database.

Results: 5810 significant variants were associated with pregnant patients with VTE, including 4874 single nucleotide variants and 936 short deletions which were annotated in 3417 genes ($P < 0.05$). Fifty-six variants annotated in 46 genes ($P < 0.001$) and the top 3 variants, including rs2706258 (RNA *LOC102724050*, $p = 1.25 \times 10^{-6}$), rs17057520 (*SCARA3*, $p = 4.64 \times 10^{-5}$), and rs3739550 (*CAAP1*, $p = 4.64 \times 10^{-5}$), were identified. Fourteen low frequency variants had a minor allele frequency (MAF) of less than 1%. Logistic analysis revealed that rs7099478 (*GRK5*), rs8041208 (*WDR72*), rs17215792 (*KLF7*), rs13035688 (*KLF7*), rs6725221 (*KLF7*), and rs3214417 (*KLF7*) were associated with an increased risk of developing VTE ($P < 0.05$, $OR > 1$). In addition, combined pathway and PPI analyses revealed that CDC7 and MCM6 involved with DNA replication were associated with VTE in pregnant individuals.

Conclusion: The study identified a series of variants and genes that may contribute to VTE in the Chinese pregnant population. Several genes may be risk factors for VTE including *KLF7, GRK5,* and *WDR72*. CDC7 and MCM6 may be related to the potential functions of VTE in pregnant women. Notably, the *KLF7* gene had 4 genetic variants that were found to be associated with lipid metabolism and cardiovascular diseases. Therefore, further validation is required to reveal the KLF7 mechanism in pregnant women with VTE.

Introduction

Venous thromboembolism (VTE) is defined as a blood clot forming condition, consisting of both deep vein thrombosis (DVT) and pulmonary embolism (PE). It is a major public health issue that causes considerable morbidity and mortality in the population, especially in pregnant women [1–3].

Genetic variation has been regarded as a significant determinant of thrombosis risk [4–6]. The genetic contribution to the risk of VTE was first established in the 1960s, when it was found that the association of antithrombin deficiency and ABO blood type lead to increased VTE risk[7]. Moreover, five genetic factors (factor V Leiden [*FVL F5*], prothrombin gene [*PT*] mutation G20210A, and deficiencies of protein C [*PC*], protein S [*PS*], and antithrombin [*AT*]) were observed, however these factors underlie a minority of VTE cases [8].

Recently, next generation sequencing and whole exome sequencing (WES) studies have provided solid evidence for genetic contributions to VTE pathogenesis. For example, a 55 gene panel and a 63 gene panel have been established in VTE studies [8, 9]. Furthermore, several large genome-wide association studies have also identified single nucleotide polymorphisms (SNPs) associated with VTE [10–15]. However, known VTE-associated genes account for only a small fraction of VTEs and a large number of patients with VTE lack known thrombophilia mutations, especially in pregnant women [16–18]. This suggests that other causative or susceptible factors may play significant roles in the development of VTE, particularly in pregnant women. It must be mentioned that there are still no VTE studies in pregnant women from the Chinese mainland, but there is one study focusing on genetic factors of VTE disease in pregnant women in Hong Kong [19].

In this study, we performed a case-control study of 25 pregnant Chinese women with VTE and 17 healthy pregnant women using WES. We aimed to identify novel or rare variants that may contribute to the pathogenesis of pregnant women with VTE.

Materials And Methods

2.1 Study population

A prospective cohort study was conducted in Shanghai, China between 2014 and 2017. 10 hospitals, including 5 general hospitals and 3 maternal and child hospitals, joined this study. Pregnant women whose first antenatal care (within 16–24 gestational weeks) was at these participating hospitals were recruited and followed-up until 42 days after delivery. Data on women's background information and physical examination were collected during the first interviews. They were then followed-up after 42 days when they returned to the hospitals for routine postnatal care. A nest case-control study was conducted, in which cases were women who were diagnosed with VTE and controls were those at similar gestational age but without VTE. All VTE cases were confirmed by B-ultrasound scan or CT examination, in addition to clinical symptoms and signs. The demographic characteristics, surgery history, trauma history, gestational diabetes history, and chronic obstructive pulmonary disease history of cases and controls were recorded via face-to-face interviews, and their medical examination data were recorded. Blood samples were taken immediately once women were diagnosed with VTE. Blood serum and plasma were separated within 12 hours and stored at -80 °C (Table 1). All participants gave their written informed consent before recruitment into the study. This study was approved by the ethics committee of the Shanghai Institute of Planned Parenthood Research and ethical committees of participating hospitals in the People's Republic of China.

Table 1
Participant characteristics for WES.

Maternal variables	Normal (n = 14)	VTE (n = 18)	P-value
Age (years)	31 ± 3.5 (24–37)	33 ± 3.9 (25–41)	0.074
Height (cm)	161.6 ± 2.3 (157–165)	161.3 ± 5.5 (152–169)	0.822
Weight (kg)	61.4 ± 11.2 (45–85)	59.5 ± 6.1 (46–75)	0.567
BMI (kg/m2)	23.5 ± 4.3 (16.5–33.2)	22.8 ± 2.4 (19.9–29.3)	0.64
High fat food intake before pregnancy(cups/week)	1.4 ± 0.9	1.3 ± 0.6	0.93
High fat food intake during pregnancy(cups/week)	1.3 ± 0.7	1.3 ± 0.6	0.973
Passive smoking before pregnancy	6	7	0.828
Passive smoking during pregnancy	3	7	0.306
Surgery history	0	1	0.387
Gestational diabetes	1	1	0.819
Hypertension	0	1	0.391
Chronic obstructive pulmonary disease	0	1	0.391
Varicose vein of lower extremity	0	1	0.387

2.2 DNA extraction

Blood samples were obtained from pregnant patients with VTE (n = 25) and controls from patients without VTE (n = 17). Genomic DNA was extracted using a DNA extraction kit (Qiagen, Hilden, Germany) from whole blood and stored at -80 °C.

2.3 Whole exome sequencing

Library construction, WES, and data analyses were carried out by iGeneTech in Shanghai. According to the manufacturer's protocol, 200 ng of genomic DNA from each individual was sheared by Biorupter (Diagenode, Belgium) to acquire 150 ~ 200 bp fragments. The ends of the DNA fragments were repaired and Illumina Adapter was added (Fast Library Prep Kit, iGeneTech, Beijing, China). After sequencing, libraries were constructed and the whole exomes were captured with AIExome Enrichment Kit V1 (iGeneTech,Beijing, China) and sequenced on an Illumina NovaSeq 6000 (Illumina, San Diego, CA) next generation sequencing platform with 150 base paired-end reads.

2.4. Exome data analyses

WES data were analyzed with the company's standard pipeline. First, the raw reads were filtered to remove low quality reads using FastQC. Clean reads were then mapped to the reference genome GRCh37 (hg19) using BWA (Burrows–Wheeler Alignment). After removing duplications, single nucleotide variants (SNVs) and indels, were identified using the Genome Analysis Toolkit (GATK). All variants were annotated using ANNOVAR [20].

The minor allele frequencies of the most promising candidate pathogenic variants were screened using public databases, including dbSNP, 1000 Genomes Project (http://www.1000genomes.org), Exome Aggregation Consortium (ExAC) (http://exac.broadinstitute.org), and the Genome Aggregation Database (gnomAD).

Variant damage was predicted by different tools: Polyphen-2 (genetics.bwh.harvard.edu/pph2), Mutation Taster (www.mutationtaster.org/), SIFT (sift.jcvi.org/), Combined Annotation Dependent Depletion (cadd.gs.washington.edu), and PROVEAN (http://provean.jcvi.org/index.php).

2.5 Statistical analysis

The intersection of the mutation sets of all patients was analyzed. The cutoff criteria were used to screen for single nucleotide variation with statistical significance ($P < 0.05$).

A series of filtering criteria were applied to the variant candidates to finally identify significantly associated SNVs and indels: (1) Rare and novel protein-altering variants (frameshift, essential splicing-site, and missense) were preferred; (2) synonymous mutations were discarded; (3) only variants within the exonic or splicing site were preferred and variants in intronic site were filtered out when selecting for VTE associated variants; (4) variants reported in more than 1% of the population in the public 1000 Genomes Project database were additionally filtered out by strict criteria as they were regarded as SNPs. A Manhattan plot of $-\log_{10}P$ of significant SNVs or indels was generated using R3.6.1. Enrichment analysis analysis were performed using the web tool DAVID (https://david.ncifcrf.gov/). Protein–protein interaction analyses were conducted using the STRING online tool (https://string-db.org/) We performed a logistic regression analysis to explore the association between pregnant individuals with VTEs and selected variants by using R3.6.1. Results were expressed in terms of odds ratios (ORs) with their 95% confidence intervals (95% CI).

Results

3.1 Study population

Twenty-five positive cases and 17 controls were recruited for WES. The basic characteristics of the participants are shown in Table 1. Twenty-seven women were diagnosed with VTE during the study period. The basic population characteristics were shown, including average age, BMI, high fat food intake, and passive smoking. Controls were younger on average than the positive cases.

3.2 SNVs and indel association analysis

Among pregnant women with follow-up outcomes, 42 participants were available for experimental analysis (25:17). A summary of the main whole exome sequencing findings and analysis flowchart is shown in Fig. 1. In total, 5810 significant SNVs or short indels ($P < 0.05$) were annotated in 3417 genes. All 5705 sites were obtained with the rs number shown in the Manhattan plot (Fig. 2). 4874 sites were SNVs and 936 sites were indels.

Fifty-six variants were annotated in 46 genes including 43 SNVs and 13 indels, which were identified to be significantly differentiated between VTE cases and controls based on the threshold of $P < 0.001$ (Supplemental Table 1). Three significant variants ranked at the top were listed ($p < 10^{-4}$). The significant variants included one in the long non-coding RNA *LOC102724050* (12:54767613 C > T, rs2706258, $p = 1.25 \times 10^{-6}$), which was detected in 2 VTE cases and 14 controls, a significant variant in *SCARA3* (chr8: 27528321 C > T, rs17057520, $p = 4.64 \times 10^{-5}$) and a variant in *CAAP1* (chr9: 26892301A > G, rs3739550, $p = 4.64 \times 10^{-5}$).

Subsequently, all 5810 mutations significantly associated with disease presentation in five or more patients were analyzed ($p < 0.05$). The results showed that 306 candidate variants were annotated in 196 genes, present in > 6 patients (7–14) but not present in the control samples. 265 mutations were identified in > 8 patients (9–15) and only 1 in control individuals. 216 mutations were in > 10 patients (11–16) and only 2 in control individuals. Further studies regarding these mutations are needed.

Moreover, 443 mutations were found with a minor allele frequency (MAF) of less than 5% in the 1000 genomes project 2015 (Supplemental Table 2). Forty-nine mutations (MAF < 0.05) were obtained, which were not found in normal controls but presented only in patients (7–11). As variations with MAF > 1% were regarded as SNPs, we applied stringent screening criteria (MAF < 1%, $P < 0.05$) and 94 variants were obtained (Supplemental Table 3). Next, variants in intronic sites were filtered out and only variants within the exonic or 5 UTR sites were preferred. Variations absent in the positive cases were also filtered out. Finally, 14 sites were obtained shown in Supplemental Table 4.

3.2Association analysis between variants and pregnant patients with VTE: logistic regression

The logistic regression analysis for 56 and 14 selected variants regarding their potential associations are reported in Table 2. Six variants were annotated in 3 genes, including rs7099478 (*GRK5*), rs8041208 (*WDR72*), rs17215792 (*KLF7*), rs13035688 (*KLF7*), rs6725221 (*KLF7*), and rs3214417 (*KLF7*) and these were strongly related to VTE in pregnancy ($P < 0.05$, $OR > 1$). ORs for all genotypes were estimated with respect to the reference genotype. For example, the calculated OR for rs7099478 mutation annotated in *GRK5* for a heterozygous mutation from the wild type was 19.8 (95% CI3.84. 160.04, $p = 0.001$). The OR for a homozygous mutation was 27.5 ($p = 0.013$). Eighteen pregnant patients with VTE had the rs7099478 C > A heterozygous mutation and 5 pregnant patients with VTE had the rs7099478 C > A homozygous mutation, whereas 6 had this mutation in control group. The ORs for rs8041208 C > T in *WDR72* gene were 13.87 (95% CI 3.10–81.97, $p = 0.001$) and 10.4($p = 0.058$). Notably, rs17215792, rs13035688, rs6725221 and, rs3214417 homozygous variants annotated in *KLF7* were associated with a higher likelihood of pregnant patients with VTE. The OR for this site was 20.36 (95% CI 3.35–396.88, $p = 0.006$). The likelihood of VTE in pregnancy was 20 times higher in variants than the reference genotype. It was suggested that *KLF7* may have a putative impact for VTE occurrence during pregnancy. Furthermore, other sites were associated with a decreased risk of VTE in pregnancy ($P < 0.05$, $OR < 1$).

Table 2
Logistic regression analysis of selected SNVs identifies 6 significant risk factors for VTE in pregnant patients.

Gene	Chr	SNP	Mut_type	Mutation	Location	Position	Func.ref	Genotype	Control(n = 17)	Case(n = 25)	OR	95% CI	P
GRK5	chr10	rs7099478	SNP	C/A	flank150	121191186	intronic	C/C	11 (64.7)	2 (8.0)			
								C/A	5 (29.4)	18 (72.0)	19.8	3.84–160.04	0
								A/A	1 (5.9)	5 (20.0)	27.5	2.71–725.33	0
WDR72	chr15	rs8041208	SNP	C/T	ontarget	53808157	UTR3	C/C	13 (76.5)	5 (20.0)			
								C/T	3 (17.6)	16 (64.0)	13.87	3.10–81.97	0
								T/T	1 (5.9)	4 (16.0)	10.4	1.19–232.78	0
KLF7	chr2	rs17215792	SNP	A/C	ontarget	207941909	UTR3	A/A	16 (94.1)	11 (44.0)			
								A/C	1 (5.9)	14 (56.0)	20.36	3.35–396.88	0
KLF7	chr2	rs13035688	SNP	T/A	ontarget	207942862	UTR3	T/T	16 (94.1)	11 (44.0)			
								T/A	1 (5.9)	14 (56.0)	20.36	3.35–396.88	0
KLF7	chr2	rs6725221	SNP	C/G	ontarget	207942894	UTR3	C/C	16 (94.1)	11 (44.0)			
								C/G	1 (5.9)	14 (56.0)	20.36	3.35–396.88	0
KLF7	chr2	rs3214417	InDel	CA/C	ontarget	207944475	UTR3	CA/CA	16 (94.1)	11 (44.0)			
								CA/C	1 (5.9)	14 (56.0)	20.36	3.35–396.88	0

3.3 Enrichment and PPI of the identified VTE associated genes in pregnancy

To assess whether VTE associated genes in pregnant patients were enriched in certain pathways or biological processes, we conducted gene ontology (GO) and pathway analysis of 46 significantly associated genes ($P < 0.001$) (Fig. 3). 3 enriched biological processes, 6 enriched molecular function clusters, 2 enriched cellular components, and 2 enriched Reactome pathways were identified with a cut-off of 0.1 (Supplemental Table 5). The biological processes showed that these significantly differentiated genes were strongly related to DNA replication initiation, positive regulation of cell proliferation, and the release of sequestered calcium ions into the cytosol ($P < 0.1$). The molecular function clusters showed that these genes were strongly related to functions involving ions ($P < 0.1$), such as ATP binding, single-stranded DNA binding, calcium-release channel activity, DNA helicase activity, ATP-dependent DNA helicase activity, and ion channel activity. Cellular component enriched these genes in the perikarya and mitochondria ($P < 0.1$). The most highly enriched class of Reactome pathway was for the activation of the pre-replicative complex ($P = 0.058$) with 2 genes, including *CDC7* and *MCM6*.

Protein and protein interaction analysis showed direct, substantial connectivity among the significant VTE-associated genes in pregnant patients (Fig. 4). A total of 5 direct connections among 7 genes were observed. 1 main network was identified, including 3 genes; *KCNC2, CCK*, and *GRK5*. Two more extensive networks were also obtained, the first network included *ABCA4* and *SEMA4A*. The other network was consistent with the pathway results and included *CDC7* and *MCM6*, both of which are related to cell division and DNA replication in tissue.

3.4 Association between putative VTE high-risk variants and VTE in pregnancy

we conducted an association analysis between well-studied VTE high-risk variants and VTE in pregnancy. All 28 variants annotated in 11 genes, including *F5* (coagulation factor V, FVL), *F2* (coagulation factor II), *MTHFR* (methylenetetrahydrofolate reductase), *ACE* (angiotensin I converting enzyme), *FGG* (fibrinogen gamma chain), *FGB* (fibrinogen beta chain), *TES* (testis derived transcript (3 LIM domains)), *PTS* (6-pyruvoyltetrahydropterin synthase), *PTH* (parathyroid hormone), *PROC* (protein C (inactivator of coagulation factors V a and VIII a)), and *PROS1* (protein S (alpha)), were analyzed with P values as shown in Supplemental Table 6. However, the differences between cases and controls were not statistically significant in this study.

Discussion

In this study, we selected 25 Chinese Han pregnancies clinically diagnosed with VTE, and 17 normal controls and performed WES to identify pathogenic variants. WES analysis revealed 5810 variants of 3417 genes in unrelated subjects. 56 variants annotated in 46 genes were significantly different between maternal VTE cases and controls ($P < 0.001$). Furthermore, 306 variants that were absent in controls were obtained. Moreover, some variants were selected by

MAF. By the logistic regression of 70 selected variants, 6 variants including rs7099478 in *GRK5*, rs8041208 in *WDR72*, rs17215792, rs13035688, rs6725221 and rs3214417 in *KLF7*, were associated with an increased risk of VTE in pregnant patients ($P < 0.05$, $OR > 1$).

Moreover, the biological processes showed that these significantly differentiated genes were strongly related to DNA replication initiation. The molecular function clusters showed that these genes were strongly related to functions involving ions. The cellular component enriched these genes in the perikarya and mitochondria. The most highly enriched class of pathway was for the activation of the pre-replicative complex ($P = 0.058$) with 2 genes including *CDC7* and *MCM6*. *CDC7* (cell division cycle 7) encodes a cell division cycle protein with kinase activity that is critical for the G1/S transition. *MCM6* (Mini-chromosome maintenance complex component 6) encodes a putative replicative helicase. There has been Activation of ATR in response to replication stress with the same 2 genes identified in our study. A total of 5 direct connections among 7 genes (*KCNC2, CCK, GRK5, ABCA4, SEMA4A, CDC7, MCM6*) were observed by PPI analysis. We presumed that these variants may contribute to protein function by involving DNA replication and ion mediated molecular signaling, which underlies VTE. Notably, *GRK5* (G protein-coupled receptor kinase 5) encodes a member of the guanine nucleotide-binding protein (G protein)-coupled receptor kinase subfamily of the Ser/Thr protein kinase family and plays a role in regulating the motility of polymorphonuclear leukocytes (PMNs). *GRK5* was also found to be significantly related to VTE in pregnancy by logistic regression. The other extensive network was consistent with the pathway results including *CDC7* and *MCM6*, both of which are related to cell division and DNA replication in tissue. In summary, it has been suggested that *GRK5*, *CDC7*, and *MCM6* may have potential functions involving VTE in pregnancies.

Several well-studied variants were detected, including *F5, F2, MTHFR, ACE, FGG, FGB, TES, PTS, PTH, PROC, and PROS1*. However, the results showed that the differences in these factors between cases and controls were not statistically significant.

To date, there have been no studies on relevant maternal VTE genetic factors focusing on populations in mainland China. In our study, 6 variants were identified by the WES and logistic regression analyses of selected variants for VTE in pregnancy. The OR value were > 2.0, which may be an important reference for thrombosis causing VTE in pregnancies. Most previous genetic studies have found that variations for statistically significant genes show a small odds ratio to disease (usually less than 1.2). However, the 95% CI range for each site was larger in our study because of the small sample size. Another limitation is that the GO and PPI analyses involved only small number of genes, which may miss some important genetic factors.

The common risk factors of VTE include age, pregnancy, delivery, oral contraceptives, hormone replacement therapy (HRT), antiphospholipid antibody syndrome, trauma, surgery, malignant tumor, diabetes, nephrotic syndrome, long-distance travel, and long-term bed rest. In addition, several genetic factors, including *PC, PS, PT* gene mutation G20210A, *FVL F5*, and *AT* were observed [21]. Previous VTE case-control studies of Chinese pregnant women regarding risk factors for VTE showed no significance. To date, a study in Hong Kong (case: control = 44:55) showed that no significance of factor V Leiden and G20210A mutation was identified. Furthermore, a deficiency in the antithrombin gene was observed in only 5 cases (11%) [19]. To date, the largest survey of anticoagulants in the healthy population (n = 3493) in China [21] found that the detection rates of *PC, AT*, and *PS* deficiency were 0.29%, 0.08%, and 0.056%, respectively, and the gene variation rate was 0.43%. Another study showed [22] no mutation of factor V Leiden and prothrombin G20210A in VTE patients, PE patients, and normal controls. These results were consistent with our analysis. Further studies with large sample sizes are needed.

Notably, *KLF7*, with 4 variants present in 14 cases, was strongly associated with VTE in pregnancies. The rs17215792-AC, rs13035688-TA, rs6725221-CG and rs3214417-CA/C genotypes (OR > 1) were associated with increased risk of VTE in pregnant patients. It suggested that *KLF7* may have an impact on VTE pathogenesis.

The *KLF7* (Krüppel-like factor 7) gene is located on human chromosome 2q33.3and contains 11 exons according to the NCBI database. The KLF7 protein belongs to the Krüppel-like transcriptional regulator family, also known as ubiquitous Krüppel-like factor (UKLF), which plays an important role in regulating cell proliferation, differentiation, and survival. The KLF7 protein contains three C2H2 zinc fingers at the C-terminus that mediate binding to guanine-cytosine rich sites. The KLF7 protein is ubiquitously expressed in the bone marrow, endometrium, fat, and other various tissues.

In 1998, *KLF7* was initially cloned from human vascular endothelial cells by PCR. According to its wide expression in various adult tissues, it was named the ubiquitous KLF by Matsumoto [23].

Many factors of the Krüppel family (KLFs) participate in the physiological regulation of muscle cells. KLF7 plays an important role in regulating nerve development and myeloid cell proliferation and is expressed in skeletal muscles and myoblasts. A study found that KLF7 may affect the myoblast cell cycle.

In addition, variants of KLF7 have been previously reported to cause obesity and type 2 diabetes, involving fat cell differentiation [24, 25]. It was reported that KLF7 can regulate a number of genes related to fat synthesis and metabolism, including lipoprotein lipase (*LPL*) and leptin (*LEP*), which are known to be associated with VTE [26, 27].

KLF7 is also involved in coronary artery disease (CAD) [28]. KLF7 may bind to the promoter region of CAD-associated genes (factor VII, platelet derived growth factor, leptin, and plasminogen) as shown by bioinformatics prediction. Moreover, KLF7 may play important roles in ischemic heart disease (IHD) with hypercholesterolemia, arterial hypertension, and type 2 diabetes [29].

These findings indicate that lipid metabolism genes, such as *KLF7*, may be involved in the formation of cardiovascular continuum comorbidity by regulating glucose and lipid metabolism.

In our study, 4 variants in KLF7 were all located in the 3’ UTR region. This indicates that these mutations may alter KLF7 expression by regulating its binding to the miRNA of fat or muscle function genes.

Therefore, further in-depth research is needed to explore how the *KLF7* gene affects cardiovascular disease, as well as VTE, obesity, and glucose dysfunction disorders.

Conclusions

In this study, we identified 6 variants related to the risk of VTE, including rs7099478 in *GRK5*, rs8041208 in *WDR72*, rs17215792, rs13035688, rs6725221, and rs3214417 in *KLF7*. *GRK5*, *CDC7*, and *MCM6* may be related to certain functions of VTE in pregnant patients. Notably, KLF7 is involved in lipid metabolism, type 2 diabetes, and cardiovascular diseases, further indicating that KLF7 may be involved in the formation of cardiovascular continuum comorbidities, which includes VTE, by regulating glucose and lipid metabolism. Further research is needed to validate *KLF7* and other identified associated genes.

Abbreviations

VTE: Venous thromboembolism

DVT: deep vein thrombosis

PE: pulmonary embolism

WES: whole exome sequencing

SNPs: single nucleotide polymorphisms

SNVs: single nucleotide variants

PPI: protein-protein interaction

GO: gene ontology

CDC7: cell division cycle 7

MCM6: Mini-chromosome maintenance complex component 6

GRK5: G protein-coupled receptor kinase 5

PMNs: polymorphonuclear leukocytes

KLF7: Krüppel-like factor 7

Declarations

Acknowledgements

We thank the many women who participated in the pregnant VTE study and the funding.

Competing interests

None declared.

Funding

This work was supported by grants from the National Natural Science Foundation of China (No.81472990) and Clinical Research Project of Shanghai Municipal Commission of Health and Family Planning (No.201840095).

Author' s Contributions

YPS and YZ were responsible for acquisition of data, analysis and interpretation of data, drafting the article and revising it critically for important intellectual content. YX, ZPZ, BHZ, AHL and ZFZ contributed to acquisition of data and revision of the manuscript. JD and YC were responsible for the conception and design of the study, revising it critically for important intellectual content and final approval of the version to be submitted. All authors read and approved the final manuscript.

Ethics approval and consent to participate

Ethical approval was obtained from the medical ethical committee, shanghai institute of planned parenthood research (PJ2014-11) on 26 February 2014. The study was conducted according to good clinical practice guidelines and the principles of the Declaration of Helsinki.

Availability of data and materials

The authors declare that the data supporting the findings of this study are available within the article and on request from the corresponding author.

Consent for publication

Not applicable.

References

1. Beckman MG, Hooper WC, Sara EC, Thomas LO: **Venous Thromboembolism: A Public Health Concern**. *Am J Prev Med* 2010, **38**(4-supp-S):S495-S501.
2. Martinelli I, De Stefano V, Taioli E, Paciaroni K, Rossi E, Pier M: **Inherited Thrombophilia and First Venous Thromboembolism during Pregnancy and Puerperium**. *Thromb Haemost* 2002, **87**(05):791-795.
3. Nelson SM, Greer IA: **Thrombophilia and the Risk for Venous Thromboembolism during Pregnancy, Delivery, and Puerperium**. *Obstet Gynecol Clin North Am* 2006, **33**(3):413-427.
4. Souto JC, Laura A, Montserrat B, Francisco B-V, José M, José Manuel S, Inma C, Rosa F, William S, Jordi F: **Genetic Susceptibility to Thrombosis and Its Relationship to Physiological Risk Factors: The GAIT Study**. *Am J Hum Genet* 2000, **67**(6):1452-1459.
5. Rosendaal FR: **Venous thrombosis: a multicausal disease**. *Lancet* 1999, **353**(9159):0-1173.
6. Seligsohn U, Lubetsky A: **Genetic susceptibility to venous thrombosis**. *New Engl J Med* 2001, **344**(16):1222-1231.
7. Jick H, Slone D, Westerholm B, Inman WHW, Worcester J: **Venous thromboembolic disease and ABO blood type. A cooperative study**. *Lancet* 1969, **1**(7594):539-542.
8. Lee EJ, Dykas DJ, Leavitt AD, Camire RM, Lee AI: **Whole-exome sequencing in evaluation of patients with venous thromboembolism**. *Blood Adv* 2017.
9. Germain M, Saut N, Greliche N, Dina C, Lambert JC, Perret C, Cohen W, Oudot-Mellakh T, Antoni G, Alessi MC: **Genetics of Venous Thrombosis: Insights from a New Genome Wide Association Study**. *Plos One* 2011, **6**(9):e25581.
10. Weihong T, Saonli B, Xiaoxiao K, Pankow JS, Nena A, Adrian T, Mary C, Eric B, Folsom AR: **Genome-wide association study identifies novel loci for plasma levels of protein C: the ARIC study**. *Blood* 2010, **116**(23):5032-5036.
11. Morange PE, N S, G A, J E, D.-A TT: **Impact on venous thrombosis risk of newly discovered gene variants associated with FVIII and VWF plasma levels**. *J Thromb Haemost* 2011, **9**(1):229-231.
12. De Visser MCH, van Minkelen R, van Marion V, den Heijer M, Eikenboom J, Vos HL, Slagboom PE, Houwing-Duistermaat JJ, Rosendaal FR, Bertina RM: **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**(8):1474-1484.
13. Lotta LA, Tuana G, Yu J, Martinelli I, Wang M, Yu F, Passamonti SM, Pappalardo E, Valsecchi C, Scherer SE: **Next-generation sequencing study finds an excess of rare, coding single-nucleotide variants of ADAMTS13 in patients with deep vein thrombosis**. *J Thromb Haemost* 2013, **11**(7):1228-1239.
14. Buil A, Trégouët DA, Souto JC, Saut N, Germain M, Rotival M, Tiret L, Cambien F, Lathrop M, Zeller T: **C4BPB/C4BPA is a new susceptibility locus for venous thrombosis with unknown protein S–independent mechanism: results from genome-wide association and gene expression analyses followed by case-control studies**. *Blood* 2010, **115**(23):4644-4650.
15. NL. S, MH. C, A. D, DP. S, S. B, N. S: **Novel Associations of Multiple Genetic Loci With Plasma Levels of Factor VII, Factor VIII, and von Willebrand Factor: The CHARGE (Cohorts for Heart and Aging Research in Genome Epidemiology) Consortium**. *Circulation* 2010, **121**(12):1382-1392.
16. Sluis GLV, M SH, D. Y. El K, M. W T, V. E. A G, H. R B: **Family history and inherited thrombophilia**. *J Thromb Haemost* 2010, **4**(10):2182-2187.
17. Bezemer ID, van der Meer FJM, Eikenboom JCJ, Rosendaal FR, Doggen CJM: **The Value of Family History as a Risk Indicator for Venous Thrombosis**. *Arch Intern Med* 2009, **169**(6):610.
18. Bengt ZL, Li X, Ohlsson H, Ji J, Sundquist J, Sundquist K: **Family history of venous thromboembolism as a risk factor and genetic research tool**. *Thromb Haemost* 2015, **114**(11):890-900.
19. Tam W-H, Ng MH-L, Yiu AK-W, Lau K-M, Cheng GY-M, Li C-Y: **Thrombophilia among Chinese women with venous thromboembolism during pregnancy**. *Gyneco Obstet Inves* 2012, **73**(3):183-188.
20. Wang K, Li M, Hakonarson H: **ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data**. *Nucleic acids research* 2010, **38**(16):e164.
21. Zhu T, Ding Q, Bai X, Wang X, Kaguelidou F, Alberti C, Wei X, Hua B, Yang R, Wang X: **Normal ranges and genetic variants of antithrombin, protein C and protein S in the general Chinese population. Results of the Chinese Hemostasis Investigation on Natural Anticoagulants Study I Group**. *Haematologica* 2011, **96**(7):1033-1040.
22. Jun ZJ, Ping T, Lei Y, L LI, Ming SY, Jing W: **Prevalence of factor V Leiden and prothrombin G20210A mutations in Chinese patients with deep venous thrombosis and pulmonary embolism**. *Clin Lab Haematol* 2006, **28**(2):111-116.
23. Matsumoto N, Laub F, Aldabe R, Zhang W, Ramirez F, Yoshida T, Terada M: **Cloning the cDNA for a New Human Zinc Finger Protein Defines a Group of Closely Related Krüppel-like Transcription Factors**. *Journal of Biological Chemistry* 1998, **273**(43):28229-28237.
24. Kanazawa A KY, Sekine A: **Single nucleotide polymorphisms in the gene encoding Krüppel-like factor 7 are associated with type 2 diabetes**. *Diabetologia* 2005, **48**(7):1315-1322.
25. D, Zobel, C, Andreasen, K, Burgdorf, E: **Variation in the gene encoding Kruppel-like factor 7 influences body fat: studies of 14 818 Danes**. *European Journal of Endocrinology* 2009.
26. Zee RYL, Cook NR, Cheng S, Erlich HA, Lindpaintner K, Ridker PM: **Polymorphism in the β2-Adrenergic Receptor and Lipoprotein Lipase Genes as Risk Determinants for Idiopathic Venous Thromboembolism**. *Digest of the World Core Medical Journals* 2006, **113**(18):2193.

27. Umesh, Yadav, Ailiman, Mahemuti, Xuemei, Hu, Kailibinure, Abudureheman, Yuning, Xia: **Single nucleotide polymorphisms in interleukin-6 and their association with venous thromboembolism**. *Molecular Medicine Reports* 2015.
28. Vangala RK, Ravindran V, Ghatge M, Shanker J, Arvind P, Bindu H, Shekar M, Rao VS: **Integrative bioinformatics analysis of genomic and proteomic approaches to understand the transcriptional regulatory program in coronary artery disease pathways**. *Plos One* 2013, **8**(2):e57193.
29. Makeeva OA, Sleptsov AA, Kulish EV, Barbarash OL, Puzyrev VP: **Genomic Study of Cardiovascular Continuum Comorbidity**. *Acta Naturae* 2015, **7**(3):89-99.

Figures

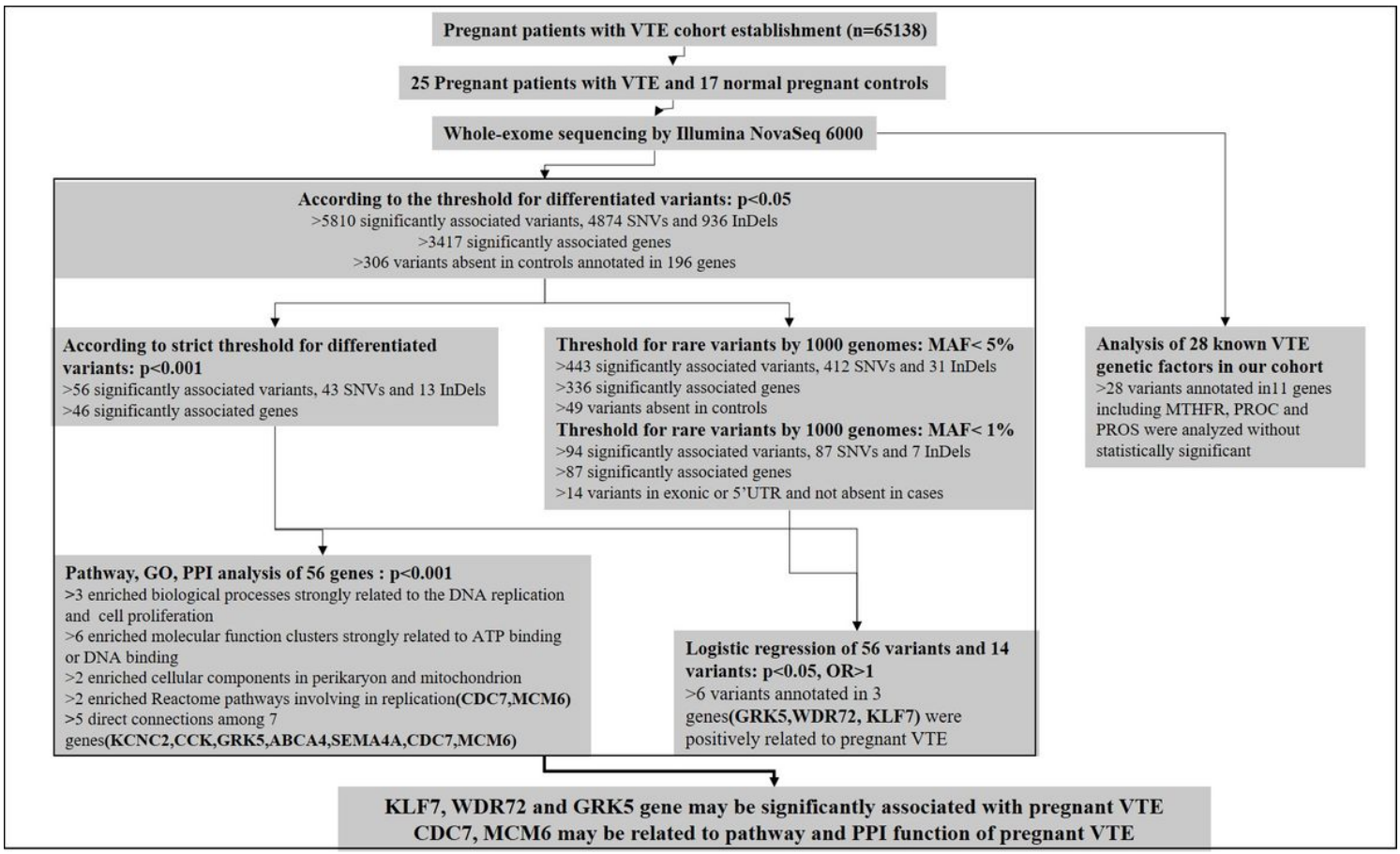

Figure 1

Summary of the main whole exome-sequencing findings and analysis flowchart.

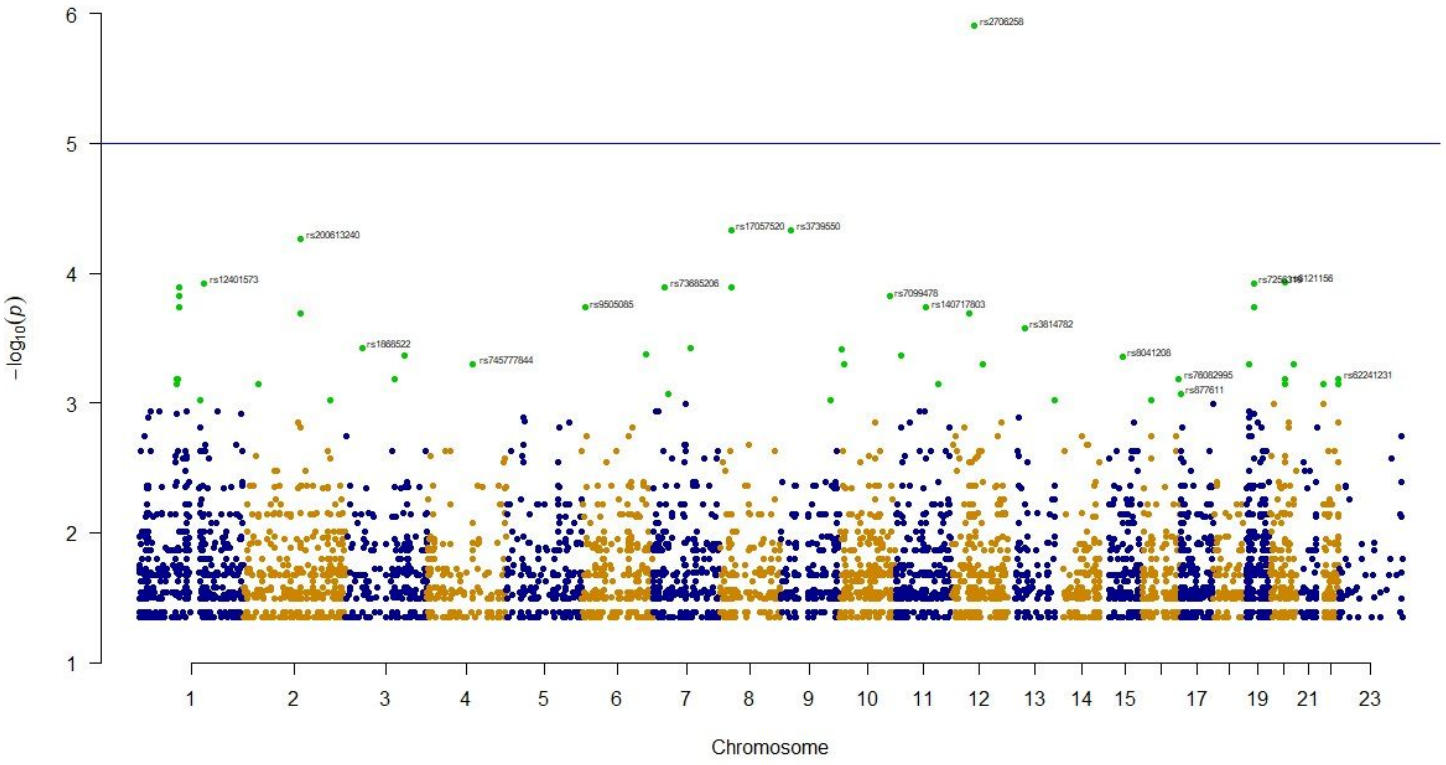

Figure 2

Manhattan plot of whole exome association results. (A) Manhattan plot for 5705 individual variants between VTE cases and normal controls (P < 0.05). The plot shows the p-values from the discovery stage for association with VTE.

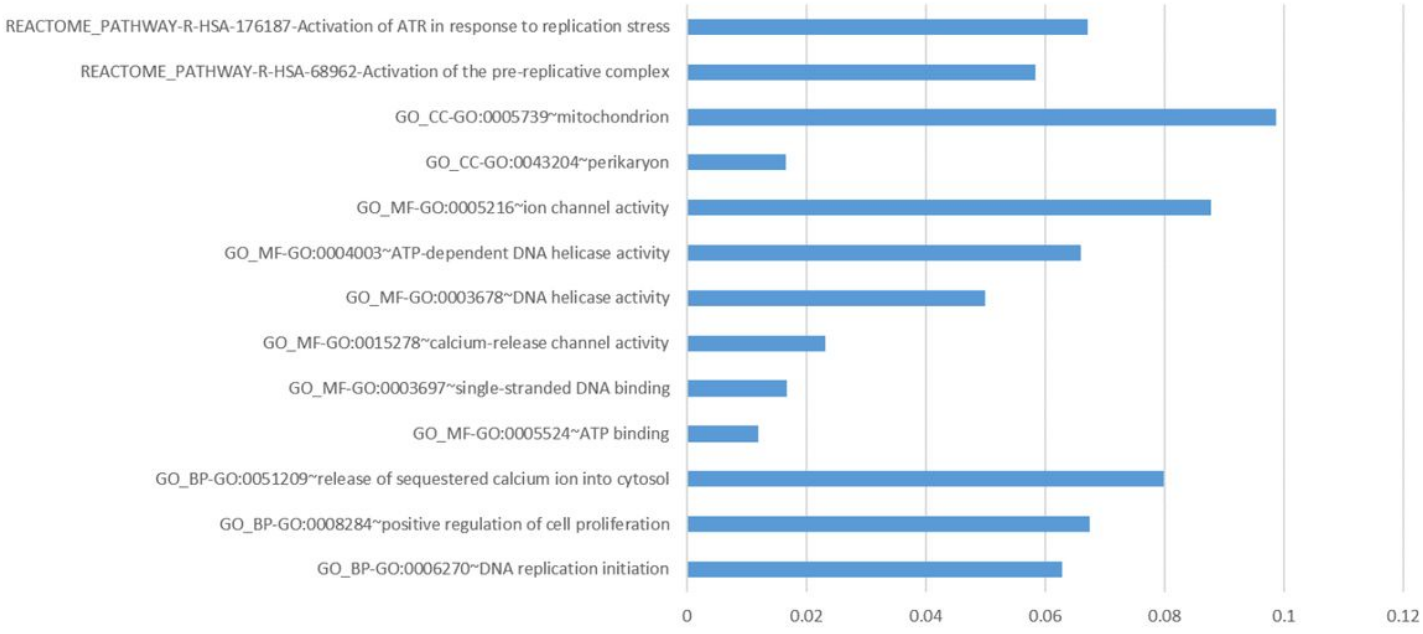

Figure 3

Gene ontology analysis of significantly associated genes annotated by altered SNVs between VTE and normal controls.

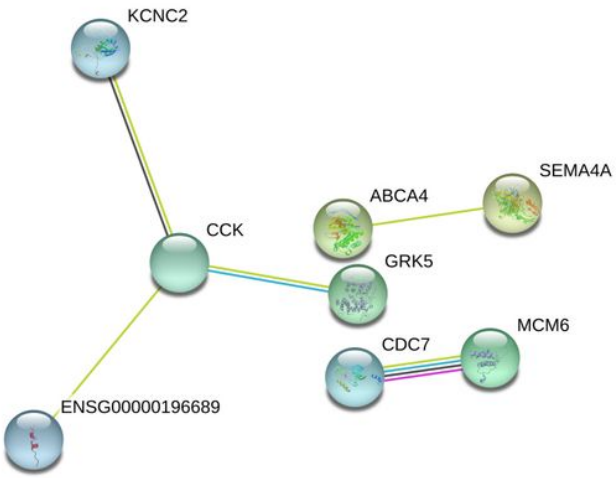

Figure 4

The protein-protein interaction (PPI) network of VTE associated genes.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- supplementarytables.docx