Prevalence of the Factor V Leiden Mutation Arg534Gln in Western Region of Saudi Arabia: Functional Alteration and Association Study With Different Populations

Mohammad Athar, PhD[1,2], Zainularifeen Abduljaleel, PhD[1,2], Ibrahim S. Ghita, MD[3], Amani A. Albagenny, MD[4], Saeed H. Halawani, PhD[5], Mohammad M. Alkazmi, MD[3], Wafa M. Elbjeirami, PhD[6], Khalid Alquthami, PhD[4], Mohammad M. Alkhuzae, BLM[4], Fadel M. Ragab, MS[4], and Faisal A. Al-Allaf, PhD[1,2]

Abstract

The rare Gln534 (Factor V Leiden; FVL) allele (1:169,519,049 T>C) is associated with an increased risk of venous thrombosis. The purpose of this study was to measure the prevalence of Factor V Leiden mutation in thrombophilia patients with deep vein thrombosis. Also, we investigated the functional and structural characteristics of this mutation p.(Arg534Gln) to be examined the cumulative impact on venous thrombosis risk as well correlated with different populations by Genome Wide Association Studies (GWAS). A total of 108 patients with idiopathic deep vein thrombosis were examined for Factor V Leiden gene mutation. Our preliminary data show that about 10% of patients were detected with the heterozygous and homozygous form of the Factor V Leiden mutation. An association analysis confirmed that the *Factor V* SNP variant (rs6025) was highly associated (*P*-value 4.91 x10-^ -39) with an increased risk of venous thrombosis. Also, we found that the recognized SNP was important among HapMap populations. Our results indicated that among the 3 populations (Asian, African, and American) studied, this association was highest in the African population based on the r(2) significant threshold (*P*-value 5e-190). In addition, this mutation was located at the domain F5/8 type A 2, which can disturb this domain and abolish its function. Because of aspartic acid nearby wild type position as form in the salt bridge due to this discharge will disturb the ionic interaction made by the wild type residue Arg534. This residue was not found to be in contact with other domains of which the function was known. However, contact with other molecules or domains (THPH2: MIM: 188055) were still possible and might be affected by this mutation that may cause thrombophilia due to activated protein C resistance.

[1] Department of Medical Genetics, Faculty of Medicine, Umm Al-Qura University, Makkah, Saudi Arabia
[2] Science and Technology Unit, Umm Al-Qura University, Makkah, Saudi Arabia
[3] Hematology Department, Al-Noor Specialist Hospital, Makkah, Saudi Arabia
[4] Laboratory and Blood Bank Department, Al-Noor Specialist Hospital, Makkah, Saudi Arabia
[5] Department of Hematology and Immunology, Faculty of Medicine, Umm Al-Qura University, Makkah, Saudi Arabia
[6] Molecular Diagnostics Unit, Department of Laboratory and Blood Bank, King Abdullah Medical City, Makkah, Saudi Arabia

Corresponding Author:
Zainularifeen Abduljaleel, Department of Medical Genetics, Faculty of Medicine, Umm Al-Qura University, Al-Abdia campus, P.O. Box 715, Makkah 21955, Saudi Arabia. Emails: zaabduljaleel@uqu.edu.sa; zainulbio@gmail.com

Introduction

The occurrence of venous thrombosis is approximately 1 in 1000 people every year.[1] Venous thromboembolic disease (VTE) is a multifactorial disorder, attributed to one or a combination of genetic, acquired, or environmental risk factors. The risk of developing VTE increases as the number of risk factors increase; however VTE can also occur in the absence of known risk factors.[2] Inherited defects causing thrombosis include factor V Leiden p.(Arg534Gln) mutation historically reported as p.Arg506Gln, the *Factor V* R2 polymorphism in individuals with *Factor V* Leiden heterozygosity, the prothrombin 20210A mutation, and deficiencies of antithrombin, protein C and protein S. This mutation occurs at one of 3 sites causing *Factor V* to be inactivated by a naturally occurring anticoagulant called activated protein C (APC). This inactivation of factor V causes low anticoagulant response to APC (APC resistance).[3] Also other than Deep Vein Thrombosis (DVT) is the formation of a blood clot in a vein deep within the body, usually in the legs. This condition is related to a more severe condition called pulmonary embolism (PE), which occurs if the clot breaks free and travels through the circulatory system.

In additional, thrombophilic mutations or a disorder is not uncommon in individuals with *Factor V* Leiden, and risk increases with increasing number of mutations present.[2] In families with thrombophilia, 13% have more than 1 mutation or thrombophilic disorder.[4] For example, in thrombophilic families, *Factor V* Leiden is found in 27% of individuals with antithrombin deficiency and 19% of individuals with protein C deficiency. A large number of missense polymorphisms in the *Factor V* gene coding for *Factor V* has been reported.[5] Among these, 2 genetic variations are now well established to affect the risk of venous thrombosis (VT): FV Leiden (FVL, rs6025, R534Q) identified[6] and the Lysine to Arginine substitution at amino acid 858 (rs4524, K858 R) identified.[7] The Q534 allele is the major genetic risk factor of VT and has a frequency of ∼5% in the general population of European descent.

The aim of this study was to measure the prevalence of Factor V Leiden mutation in thrombophilia patients and to describe the structural and functional association of *Factor V* allele (1:169,519,049 T>C) of rare genetic variation with an increased risk of venous thrombosis. To test this hypothesis, we presented a genome-wide association study on the same Venous Thrombosis (VT) case and control in different 3 populations (Asian, African, and American) used for our candidate of *Factor V* gene study, and replicated all SNPs exceeding genome-wide significance in a separate venous thrombosis case and control duplication population.

Materials and Methods

Factor V Leiden p.(Arg534Gln) Mutation Screening

About 5 ml of whole peripheral blood was collected from 108 patients with idiopathic deep vein thrombosis, from a referral Al-Noor hospital, Makkah, Kingdom of Saudi Arabia. All patients were consented and given the correct information about the study and its impact on healthcare services before samples collection. Ethical approval was taken from the Institutional Review Board (IRB), Makkah region, Saudi Arabia. Sample collection and studies were performed in accordance with the Research Ethics Committee's regulation. Blood samples were obtained from the subjects after obtaining informed written consent from the patient or their representative. The study's subjects were representing families that lived in the western areas of Saudi Arabia. Deep venous thrombosis diagnosis was based on patient's history, D-dimer test, and clinical findings confirmed by Doppler ultrasonography. Genomic DNA was extracted from whole blood samples using Invitrogen PureLink genomic DNA mini kit according to the manufacturer's instructions. FVL mutation screening was performed by real-time polymerase chain reaction-based method using DUPLICα Real Time Factor V G1691A Genotyping Kit (EUROCLON, Siziano (PV), Italy).

Functional Analysis of Single Nucleotide Polymorphisms

For the bulk of mutant variants (single amino acid adjustments or nsSNPs) in humans, its impact on protein function is unknown. This process provides binary classifications (impact/neutral) accompanied by a more detailed score. In addition, we gained insights about the protein's stability using the program Schrodinger, USA, considering the mutant stability predictions on a protein of unknown structure. In case of the mutant variant, the location of the mutated residue was specified, apart from the wild type amino acids. Several known disease-associated nsSNPs with a known 3-dimensional protein structure have structural influence on key residues and sites that are associated with protein function. Furthermore, The functional analyses were determined by HOPE,[8] SIFT,[9] and PolyPhen2[10] programs. From the SIFT and PolyPhen functional deleterious or non-pathological significant variants were identified. The HOPE report will evaluate the effect of the mutation based on the following features. 1) Contacts made by the mutated residue, 2) structural domains in which the residue is located, modifications on our targeted residues and known variants for this residues.

Genotypic and Phenotypic Analysis

We selected our genotype one single nucleotide polymorphism based on association end result characterizing common genetic variation in the *Factor V* Leiden gene using LD select[1] as carried out on the phenotypic and genotypic variation dataset server (http://pheweb.sph.umich.edu/SAIGE-UKB/).[11] The minor-allele frequency is calculated and rounded to the closest integer. When the integer is bigger than or equal to the allele frequency cutoff, the SNP is retained. The precise frequency cutoff is thus 0.5% below the allele frequency cutoff. The cutoff to 3-5% results in a precise cutoff of 4.5%. For genotypes, linkage disequilibrium, and SNP, because we glance multiple population groups based on our Allele, the frequency

and no-monomorphic filters are applied to the merged set of genotypes.

SNP Association and Assortment

An association analyses used the data from UKBiobank ICD PheWeb project throughout completely Analysis of 1403 ICD-based traits using SAIGE[11] to search for *Factor V* Leiden gene eQTLs variants and the marker linked with it. With choose a reference file of Human Genome Assembly GRCh38.p13 (https://www.ncbi.nlm.nih.gov/assembly/GCF_000001405.39/). After finding the probes, a genotype to expression pairs was selected for various populations and genotype (GenCord/ GRCh38.p13/1:169,519,049 T>C (rs6025) (MAF > 5%). The probe end result matched with transcript probe ID NM_000130.4. The evaluation parameters were fixed for associated SNPs for correlation or regression based on Spearman's rank correlation coefficient (rho) by preserving distance regulated to not more than 1 Mb at a fixed P-value default threshold 0.001.

Population Study

For population study, we retrieve dataset published by,[12] a comprehensive haplotype analysis of the rs6025 in a case-control based on totaling ∼3719 Venous Thrombosis patients and 4086 controls in order to better estimate their true impact on Venous Thrombosis risk. There are more than 1000 SNPs usage for these association analyses. Lastly which were extremely associated SNPs within a full-length gene according to the threshold P-value. Further we proposed study to link with other populations (African, Asian, and American) by used as the reference of HapMap 3 is the third phase of the International HapMap project (The International HapMap 3 Consortium 2010).[13] Parameter to set of Hardy-Weinberg p > 0.000001 (per population) missingness <0.05 (per population) <3 Mendel errors per population; only applies SNP must have a rsID and map to a unique genomic location. The "consensus" data set contains in addition for also included all individuals (168 Case and 16401 controls) in both males and females.

Statistical Methods

Observed genotype frequencies have been in compared with expected frequencies to test for deviations from Hardy-Weinberg equilibrium using Chi-Square test. SNP frequency of our main analysis was restricted based on.[14] The minor allele frequency of the 1-nucleotide polymorphisms from the full-size gene of Factor V analyzed. SNP alleles were in Hardy-Weinberg equilibrium by the Chi-Square test except link with Associated with an Increased Risk of Venous Thrombosis, which was then excluded from the analyses. In more-than 1000 SNPs within a single gene was enough to give us a power of 80%. The false-positive report probability for statistically significant observations was estimated using the methods described.[15] The linkage disequilibrium (LD) status

among polymorphisms was examined utilizing Haploview.[16] All analyses had been performed using the Statistical Analysis System for Linux in Unix platform, model 2.1.1

System Requirement

We incorporate our local server to UKBiobank ICD PheWeb database for the integrative analysis and visualization of SNP associations in eQTL studies. EHR-derived PheWAS codes in MGI library[17] to map object oriented models onto MySQL database.[18] A gene-centered Manhattan plot represents noticed SNP to gene associations a round gene of curiosity, and an SNP-centered line chart illustrates observed eQTLs surrounding SNPs of interest. Tested on a Bluegene super computing with ∼32 GB of RAM in response to set a parameters based on load samples and component for dynamic balancing of virtual machines within the cloud infrastructure resource, as soon as it is uploaded, our server can fetch per SNP-probe pairs from these >75 people in <0.0257s from the database, and calculates Spearman's rhos and nominal *P*-values for 486 SNP-probe pairs in 3 s.

Results

Analysis of Factor V Leiden p.(Arg534Gln) Mutation

A total of ∼108 patients with idiopathic deep vein thrombosis were examined for Factor V Leiden gene mutation. The genotype frequency of Factor V Leiden mutation in this study population is summarized in Table 1. The Factor V Leiden mutation in its homozygous and heterozygous forms was found in 11 out of 108 patients for an overall prevalence of 10.2%. However, 9 (8.3%) patients were heterozygous and 2 (1.9%) patients were homozygous for Factor V Leiden mutation.

Deep Vein Thrombosis and Organ Damaged

We given this reference clinical information was very useful for understand our genotypic SNP and correlate with different populations by GWAS association and statistical analysis. A blood clot inside the vein i.e. deep vein thrombosis, usually deep within leg having the risk is that part of the clot can break off and transportable through the bloodstream and could get stuck lungs and block blood flow, causing organ damage or death. All collateral veins are better detected this way of sonography, including perforator veins, but of most importance is

Table 1. Genotype Frequency of Factor V Leiden Mutation in the Studied Saudi Patients With Deep Vein Thrombosis.

Factor V Leiden (FVL)	c.1601G>A, p.(Arg534Gln)
Chromosome position	chr1:169519049
SNP ID	rs6025
Sample size (n)	108
Genotype: n (frequency %)	G/G: 97 (89.8)
	G/A: 9 (8.3)
	A/A: 2 (1.9)

the detection of venous thrombosis. The most reliable sign of thrombosis is the absence of compressibility (Figure 1). A rubber pipe cannot be compressed if the water inside is frozen and in the same way a vein cannot be compacted when the blood is in a solid state, as with a thrombus. However, if the probe was parallel to the vein axis, when the examiner compresses it, the probe can slide to the right or to the left giving a false negative for thrombosis as the probe has moved away and the vein will not then be evident. The probe will be presented parallel to the vein axis. Due to these conditions, it can damage the legs and other organs in the body and cause death. Blood clots in the thighs are more likely to break off and cause PE than blood clots in the lower legs or other parts of the body.

Lead to Changes in Protein Structure and Function Arg534Gln

This amino acid has its specific size, charge, and hydrophobicity-value. The original wild-type residue Arg534 and newly introduced mutant residue 534Gln often differ in these properties, which is based on the PDB ID: 2R7E. The wild-type residue charge was positive and mutant residue charge was neutral. The difference in charge will disturb the ionic interaction made by the original, wild type residue. The mutation is located within a domain, annotated in UniProt as: *Factor V* /8 type A 2 and the mutation introduces an amino acid with different properties, which can disturb this domain and abolish its function (Figure 2A-H). The wild-type residue forms a salt bridge with Aspartic Acid at position 532, the difference in charge will disturb the ionic interaction made by the wild-type residue. This mutation was associated with thrombophilia due to activated protein C resistance (THPH2) [MIM: 188055]. Here we given reference by ExPASy about this variant: VAR_001213 annotated with severity: DISEASE. The wild-type residue is very conserved, but a few other residue types have been observed at this position too. This mutant residue was among the residues at this position observed in other sequences. Which means that homologous proteins exist with the same residue type, as mutant at this position and this mutation is possibly not damaging to the protein. The mutated residue was located on the surface of a domain with unknown function. The residue was not found to be in contact with other
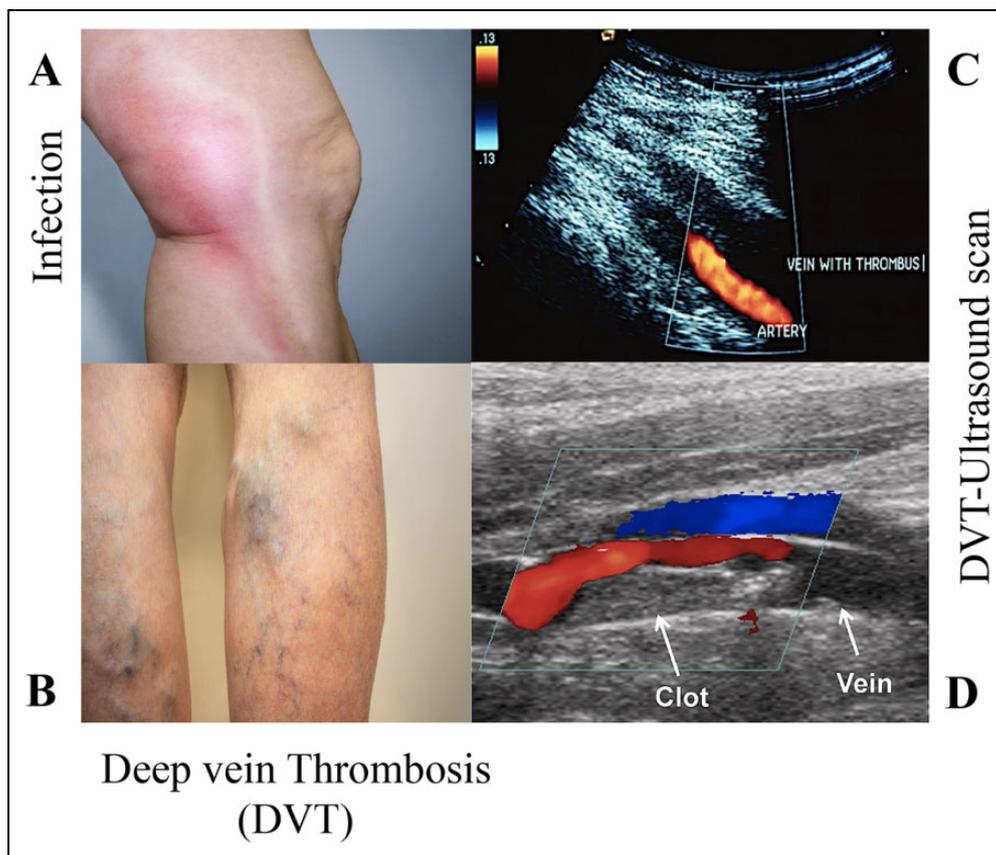

Figure 1. A-D. Clinical condition of deep vein thrombosis. A) A poor blood flow in the lower infection leg due to malfunctioning valves in the veins (venous insufficiency), the result can be another common mimic of cellulitis: venous stasis dermatitis. B) The veins have a bulging, lumpy or twisted appearance and are usually purple or dark blue. Present as swollen veins on the back of the calf or sometimes on the inside of the leg. The deep veins are located within the muscles whereas the superficial veins are closest to the skin and collect blood from capillaries and the surrounding subcutaneous tissue. The blood from the superficial veins drain into the perforator veins and ultimately the deep veins. C) Ultrasound shown deep vein thrombosis is a blood clot that forms inside a vein, usually deep within a leg. D) Ultrasonography of deep vein thrombosis color were shown in red (clot) and (vein) blue (https://res.cloudinary.com).
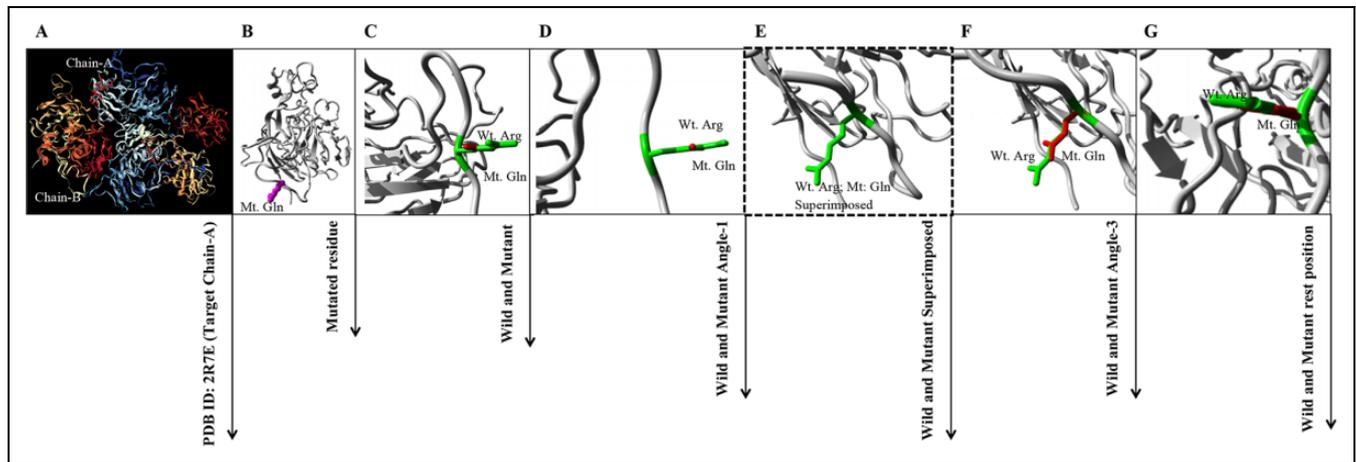
Figure 2. A-G. Structural impact. A) The PDB ID: 2R7E (Crystal Structure Analysis of Coagulation Factor VIII) contain 2 chains (Chain A & B) but in our target residue on chain A in ribbon-presentation. The protein was colored by element; α-helix (blue), β-strand (red), turn (green), 3/10 helix (yellow) and random coil (cyan). Other molecules in the complex are colored gray. B) Overview of the protein in ribbon-presentation. The protein was colored gray, the side chain of the mutated residue was colored magenta and shown as small balls. C) Close-up of the mutation. The protein was gray colored, the side chains of both the wild type and the mutant residue are shown and colored green and red respectively. D & E) Close-up of the mutation was seen from a slightly different angle). The protein was gray colored, the side chains of both the wild type and the mutant residue are shown and colored green and red respectively. F) Close-up of the mutation. Both the wild type and mutant side chain are shown in green and red respectively. The rest of the protein is show in gray. G) Close-up of the mutation, as same colors and alternating the wild-type side chain and the mutant side chain.

domains of which the function is known within the used structure. However, contact with other molecules or domains are still possible and might be affected by this mutation. Moreover, there are main amino acids properties are highly noticed due to the structure changes. This changes in charge of the wild-type residue was lost by this mutation. This can cause loss of interactions with other molecules and further cause a possible loss of external interactions. In-addition, according to the functional studies by SIFT and Polyphen2 shown both are not damaged but in clinical variant shown as pathogenic.

SNPs Association Based on GWAS and Populations

All associations between SNPs and venous thrombosis in all populations and the average estimated effect size based on the available datasets (PheWeb based on SAIGE analysis of ICD-derived traits) (Figure 3A). Not all SNPs were associated with venous thrombosis in our study populations quiet, we included entire *Factor V* gene SNPs in the genetic risk score because these SNPs had been associated with venous thrombosis in other studies. We investigated expression quantitative trait loci (eQTL) associations within a gene locus of interest in real time as functional variants. We performed gene expression profiling and linked with genetic variants of single-nucleotide polymorphisms from 3 cell varieties of 75 individuals. The samples had been explored in depth for the position of Cis-regulatory variation mostly based on huge samples totaling 3719 Venous Thrombosis patients and 4086 controls in order to better estimate their true impact on Venous Thrombosis risk. Which have been obtained simultaneously

from a subset of well phenotype of the PheWeb resource. *Factor V* gene structure studied for linkage disequilibrium (LD) patterns with haplotype and genotypic SNP associated with an increased risk of Venous Thrombosis predisposition among phenotypic African population with 0.996 allele frequencies. The LD report was obtained for the entire *Factor V* isoform 1 ($\sim$1q24–25) current in reverse orientation (3′-5′) on the long arm of chromosome 1.

A particular nucleotide polymorphism rs6025 confirmed with strong association (*P*-value 4.91×10^{-39}) in Factor V gene regions based on their implicit position on HapMap African data set (phase 3). The ideogram of chromosome 2(P21) showed the place of *Factor V* gene utilizing PheWeb (SAIGE analysis of ICD-derived traits) and through a haplotype association scanning. Furthermore, to study the association of sub-haplotypes in subsets of alleles (1:169,519,049 T>C) from the full haplotype with activity of venous thrombosis and elevated susceptibility-risk in African population, we considered a "window" of alleles by looking out a fixed-width window (1-SNPs) all over the whole haplotype in this gene.

In the phenotypic analysis of GWAS catalog y-axis indicated -log (10) *P* values on a minus logarithmic scale, whereas the x-axis showed the SNPs present on the chromosome based on their positions. The red spot designate the peaks of detrimental logarithm of P values (-log (10) P) of "fixed-width window" haplotype association (Figure 3B). This procedure utilized 1 successive SNPs to study the sub-haplotype rating and analyze the significance; where the 1-SNPs (the sub-haplotype) could have a single P value. The violet spot symbolize a major SNPs with a *P*-value 4.91×10^{-39} and its positions on
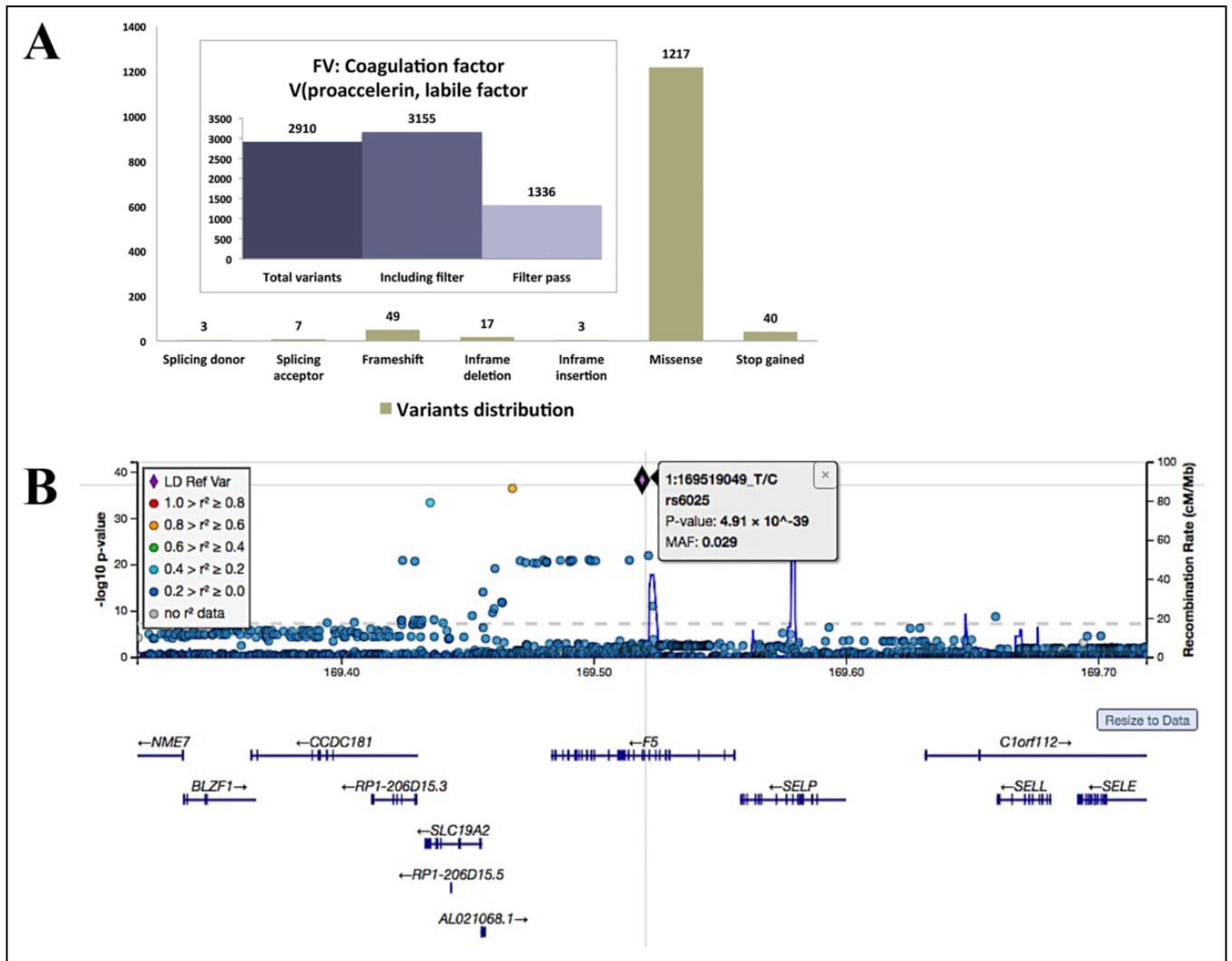
Figure 3. A-B. Factor V variants distribution and coordination of genome wide association. A) Distribution of Factor V genetic variants according to ClinVar interpretation. B) Genome-wide scan for association with rs6025. MS is replicated from prior analyses. The dashed line represents of Factor V gene [1: 169519049_T/C (*P*-value: 4.91x10^-39) and MAF: 0.029], the dotted line represents the Bonferroni correction.

the linkage disequilibrium are proven by black squares where because the D, an estimate of linkage disequilibrium parameter (R-Sq.). GWAS frequencies shown of the HapMap were chosen for the SNPs within the *Factor V* gene in African population. In every haplotype allele 1, whereas characterize allele 2 for correlated SNPs. Numbers subsequent to every haplotype frequencies. The outcomes indicate that haplotypes 1 and 100 are highly important variation blocks. The relative MAF frequency was 0.020 with which a given haplotype within the adjoining block (boldface sort = frequency of >5%).

Manhattan Plots and GWAS Based on Association Studies

GWAS results are normally checked for possible false positives with a Quantile-Quantile (QQ) plot, where the distribution of the observed p-values is compared to the expected under the null hypothesis of no SNP-phenotype association.

The initial association study was performed using PheWeb server, where the nodes have been corresponded to the biological (or) clinical database. The server represents a process of that begins with a SNP and ends within the terminal node with the calculation of its general prioritization score "S." The general score symbolize a cumulative measure of clinical relevance obtained by combining information across multiple domains. Therefore, if a SNP in a gene then it was to link to the gene nodes, which finally enhance the overall score, linking the gene to a biological phenomenon associated to the disease (Figure 4A). The extent of precedence may be achieved on a per gene basis using a numeric prioritization score. This server confirmed the original listing of SNPs of full-length *Factor V* gene ranked primarily based on the combination of evidence for association and the extent of clinical relevance. Thus, as a substitute of rating the SNPs solely based on the *p*-value, SNPs specified genes with missense mutations also have greater precedence.
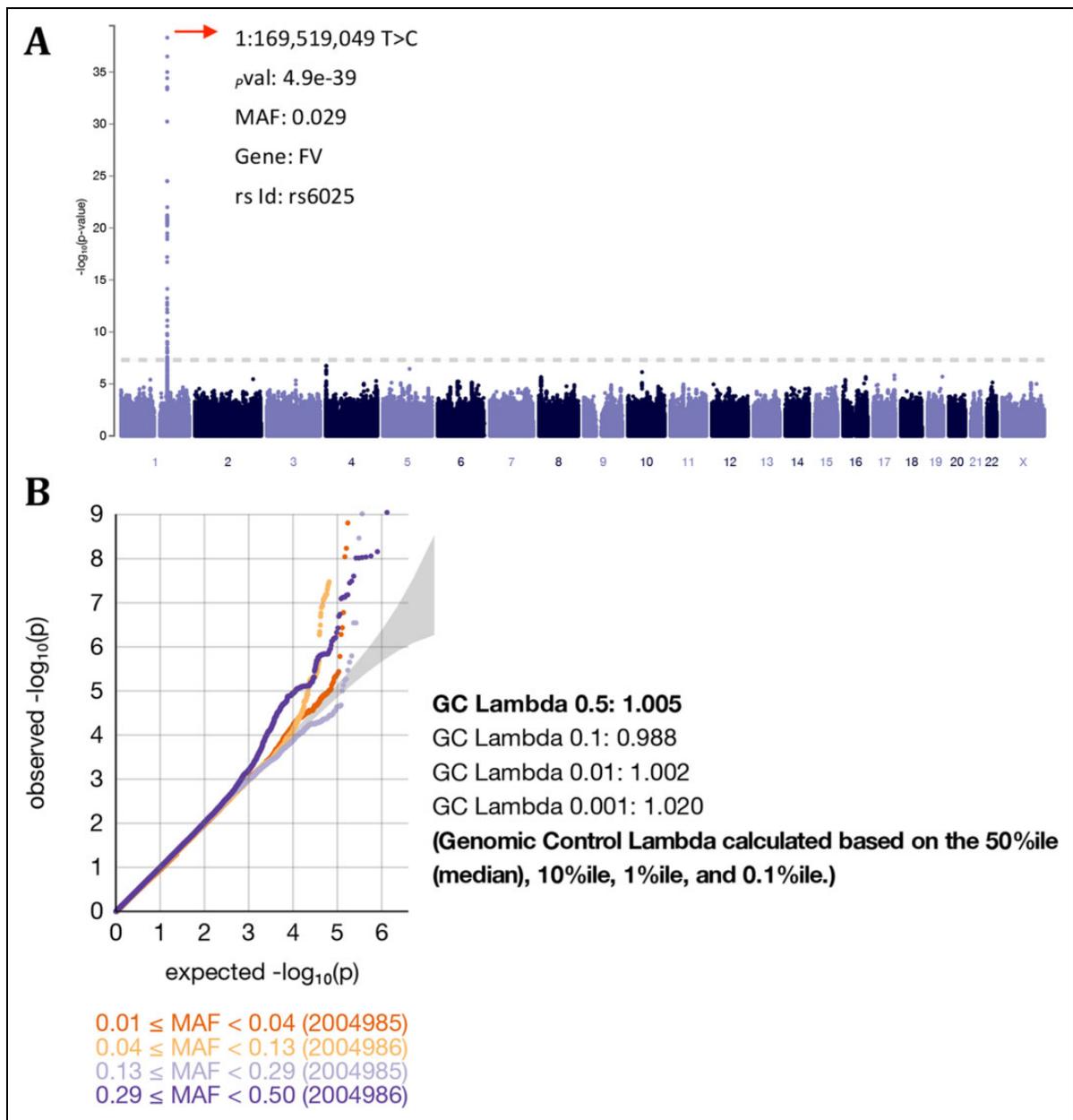
Figure 4. A-B. Analysis for case-control genome-wide association and quantification studies. A) Manhattan plot of genome-wide association analyses of X-axis shows chromosomal positions. Y-axis shows [–log₁₀ P-values] from linear regression adjusted for phenotypic. The horizontal solid line indicates the preset threshold of P-value 5e-190; MAF: 0.029. B) QQ plots showing deviation: QQ plot of genome-wide single-variant association analyses in the UK Biobank PheWeb based on SAIGE of ICD-derived traits cohort. Markers are stratified by MAF between 0.5-0.02 and $5 \times 10 -4 -1 \times 10 -4$ shown with light blue and dark blue. Dots indicate observed P-values (-log₁₀[p]) compared with those expected by chance. The gray line indicates the identity (no association) including the corresponding 95% CI under the null hypothesis (no association).

The Manhattan plots results was significant associations of particular SNP rs6025 have more than *P*-value 5e-190, under the assumption that only 1 million independent tests are performed, even if a larger number of genetic variants was tested. The significance threshold *P*-value 5e-190, also termed "genome-wide significance," was reached by dividing the usual alpha of 0.05 by 1 million. The effective number of tests was performed. Such a Bonferroni correction is conservative, increasing the credibility of loci with a *P* value not less than *P*-value 5e-190. A logical consequence of these requirements

were the need for a large sample size; such highly significant results can be reached only by analyzing large samples in generally ≥1000 participants, and this requirement is an important limitation of the method.

The QQ plot deviation of the observed *P* values from the null hypothesis. The observed P values for each SNP are sorted from largest to smallest and plotted against expected values from a theoretical χ2-distribution (Figure 4B). The quantile-quantile (QQ) plots for the model-based joint analyses of SNP and SNP-environment interaction with the Beck Depression

Table 2. Different Populations Frequencies of Rare Variant rs6025.

Population	Allele count	Allele number	Homozygotes	Allele frequency
African	16 185	16 252	8 059	0.996
Ashkenazi Jewish	9 892	10 078	4 856	0.982
East Asian	18 351	18 352	9 175	0.999
European (Finnish)	21 158	21 624	10 352	0.978
European (Non-Finnish)	110 202	113 632	53 446	0.969
Latino	34 327	34 570	17 044	0.993
South Asian	30 190	30 612	14 891	0.986
Other	6 023	6 130	2 960	0.983
Total	**246 328**	**251 250**	**120 783**	**0.98**
Male	133 083	135 798	65 226	0.98
Female	113 245	115 452	55 557	0.981

Inventory. For comparison, also includes the QQ plot of the marginal SNP association results for BDI. Clearly shows sizable inflation in P values in both joint tests that we considered. To quantify this observation, we calculated the genome-wide inflation factor (λ) for each joint analysis as the median of the observed joint tests divided by the median of a χ^2_2 random variable (which is the asymptotic distribution of the joint Wald test under the null hypothesis). This result is rarely due to thousands of true positives; more often, it is due to population stratification: systematic differences in allele frequencies between subpopulations of the collection of individuals investigated, so that a large number of P values are smaller than expected from chance alone.

Discussions

Factor V Leiden mutation causing a hemostatic disorder called activated protein C resistance (APCR), a major probability factor for the expansion of coagulation (thrombosis). FVL is the utmost usual heritable form of thrombophilia. The incidence of blood coagulation primarily depends on if an individual inherits heterozygous or homozygous for the FVL. There is about 5-10 fold higher risk of thrombosis when individuals are inheriting the heterozygous mutation and up to 50-100 fold higher risk in the presence of the homozygous mutation, compared to the wild-type healthy person. The prevalence of FVL worldwide varies depending on the geographical location and the ethnic background of the population. The prevalence of FVL has shown significant variations (5–27%) in different Arab countries.[19] Only few studies reported, the incidence of carriers of FVL in healthy individuals in Saudi Arabia is less than 2.0%.[20] In this study we were able to detect FVL mutation in 10% Saudi patients with deep vein thrombosis.

The alteration for assured complex diseases was due to relatively highly penetrant of rare variant rs6025, the allele frequency of which was typically not more less than 1%, most of which are recently derived alleles in the human populations (Supplementary Table 1). Interestingly, genetic as for the reported based on these studies by Ibrahim-Kosta 2019, the patient revealed the presence of a homozygous *Factor V* Leiden mutation (858K/K homozygotes (1.92 ±1.61 "vs"

2.81 ± 1.57, $p = 0.025$).[12] Demonstrate that the R858 allele of the *Factor V* rs4524 variant protects from venous thrombosis only in non-carriers of the Q534 allele (1:169,519,049 T>C) of the F5 rs6025, but in the normal coagulation, *Factor V* is normally inhibited by activated protein C (APC) that acts to regulate the clotting process. In *Factor V* Leiden thrombophilia, the variant factor V is resistant to APC, which increases the risk of abnormal clotting and deep venous thromboembolism (VTE).[21] Its protective effect is mediated by reduced *Factor VIII* levels and reduced activated protein C resistance.

Entire-gene testing is a biologically plausible approach to the problem of identifying functionally significant genetic variations because the ultimate unit of biological activity is the gene or its protein product.[22] In the potential molecular level of *Factor V* mutation R534Q results by the structural and functional analysis. These studies showed that the biological properties of proteins depend on cumulatively cooperative interactions of amino acids.[23] In this study, we explore not only the conservation of the salt bridge itself but also the conservation of the local sequence context and its influence on protein stability. In this aspect, wild-type residue forms a salt bridge with aspartic acid at position 532, which may play an important role in the thermodynamic stability not only as a partner in the salt bridge with Aspartic acid 532 but also by stabilizing the alpha-helix from residues in between 532 to 536 in Factor V protein. On the basis of the present results, this difference in charge will disturb the ionic interaction made by the wild-type residue.[24] Because of in this mutation particularly located on the domain F5/8 type A 2, so that can disturb this domain and abolish its function.

The genetics and genomics data associated analysis of our results from genome-wide association studies. GWAS generate genotype information on hundreds of thousands of SNPs that can be ranked according to their evidence for association with a given trait. Making sense of these data requires a number of steps related to the prioritization of variants (SNPs or CNVs) showing association to related disease which used as reference data set including the HapMap 3, and NCBI (dbSNP 137), For example, the multi-locus measure of disequilibrium, used to summarize some of the reported LD relationships among SNPs and to characterize coverage of genes by product, was

calculated using Testing Untyped Alleles (TUNA),[25] with data from HapMap. This prediction was used to estimate un-typed allele frequencies, and to perform association tests. The methods estimate of the great accuracy frequency of 1 allele of T is given by, where the haplotype frequencies are estimated from the available genotypes. The rs6025 (Arg534Gln) case described above, we test based on this hypothesis that the frequencies $h000 + 0.048h101$ are equal in cases and controls.[25]

$$U_1 = h_{000} + \frac{0.017}{0.017 + 0.333} h_{101}$$

It is very interesting to see convincing, replicated evidence of association for a R534Q with such strong biological relevance. However, it is still not clear to what extent known biology will predict variants that influence disease. This prioritization method is not designed to act as a predictor, but to preferentially select biologically relevant signals when resources are limited, either for genotyping or for functional studies in the laboratory. In this method does not incorporate information regarding the number of potential associations detected in or near a gene. For example, it is known that even for *Factor V* gene that was associated with increased risk of venous thrombosis, there are over 500 different mutant alleles.[26] It would be useful to integrate an additional mechanism into the prioritization process that somehow gives additional weight to SNP with multiple SNPs associations in the number of associations would have to be corrected for LD. This will be studied in future iterations of the GIN prioritization method. Moreover, The result of R534Q phenotypic information of correlate with different populations that were [African, Ashkenazi Jewish, East Asian, European (Finnish), European (Non-Finnish), Latino, and South Asian] based on *P*-value and (MAF > = .05) (Table 2), that is served by PheWeb comes from a series of publications describing studies characterizing eQTLs in cell lines from HapMap for different transcript levels. But as per the ratio of African population maximum allele count was 16185, Homozygotes was 8059, and allele frequency was 0.996 compared with other populations.

However, when appropriately conducted, large and small studies should give, theoretically, the same results, with just a more precise effect measure estimate from the larger ones.[1,27] Secondly, the associations we observed in our investigation of specific study hypotheses may reflect chance because a number of comparisons were made. Thus, replication in other studies is important to confirm or disprove our results. Another potential that we only analyzed 1-candidate loci of *Factor V* gene polymorphisms, that were only 1 SNP R534Q got highly associated with an increased risk of venous thrombosis and studies (*P*-value 4.91 x10^{-39}) on the relationship among genetic variations of other mutations in the pathways is required in the future.

Conclusions

This study shows the prevalence of Factor V Leiden mutation in the western region of Saudi Arabia. Our work based on genotypic and phenotypic supports the complete penetrance

of the *Factor V* rs6025 that could be the result of complex haplotype effects at the *Factor 5* locus, or interaction at this locus with other genetic or environmental exposures, such investigations may require alternative study designs to detect different effects. Therefore, more research involving the FVL and deep vein thrombosis in order to better define the genetic profile of this population in relation to the presence of mutation in *Factor V*.

References

1. Kujovich JL. Factor V Leiden thrombophilia. *Genet Med*. 2011; 13(1):1-16. doi:10.1097/GIM.0b013e3181faa0f2

2. Zöller B, Svensson PJ, Dahlbäck B, Lind-Hallden C, Hallden C, Elf J. Genetic risk factors for venous thromboembolism. *Expert Rev Hematol*. 2020;13(9):971-981. doi:10.1080/17474086.2020. 1804354

3. Levkova M, Hachmeriyan M, Stoyanova M, Miteva V, Angelova L. Comparison between thrombophilic gene polymorphisms among high risk patients. *Rom J Intern Med*. 2020;58(1):20-26. doi:10.2478/rjim-2019-0021

4. Huang X, Xu F, Assa CR, Shen L, Chen B, Liu Z.Recurrent pulmonary embolism associated with deep venous thrombosis diagnosed as protein s deficiency owing to a novel mutation in PROS1: a case report. *Medicine (Baltimore)*. 2018;97(19):e0714. doi:10.1097/MD.0000000000010714

5. Vos HL. Inherited defects of coagulation Factor V: the thrombotic side. *J Thromb Haemost*. 2006;4(1):35-40. doi:10.1111/j.1538-7836.2005.01572.x

6. Smith NL, Hindorff LA, Heckbert SR, et al. Association of genetic variations with nonfatal venous thrombosis in postmenopausal women. *JAMA*. 2007;297(5):489-498. doi:10.1001/jama. 297.5.489

7. Morange PE, Tregouet DA. Lessons from genome-wide association studies in venous thrombosis. *J Thromb Haemost*. 2011; 9(suppl 1):258-264. doi:10.1111/j.1538-7836.2011.04311.x

8. Venselaar H, Te Beek TA, Kuipers RK, Hekkelman ML, Vriend G.Protein structure analysis of mutations causing inheritable

diseases. An e-Science approach with life scientist friendly interfaces. *BMC Bioinformatics*. 2010;11:548. doi:10.1186/1471-2105-11-548

9. Vaser R, Adusumalli S, Leng SN, Sikic M, Ng PC. SIFT missense predictions for genomes. *Nat Protoc*. 2016;11(1):1-9. doi:10.1038/nprot.2015.123

10. Adzhubei I, Jordan DM, Sunyaev SR. Predicting functional effect of human missense mutations using PolyPhen-2. *Curr Protoc Hum Genet*. 2013. Chapter 7: Unit 7.20. doi:10.1002/0471142905.hg0720s76

11. Zhou W, Nielsen JB, Fritsche LG, et al. Efficiently controlling for case-control imbalance and sample relatedness in large-scale genetic association studies. *Nat Genet*. 2018;50(9):1335-1341. doi:10.1038/s41588-018-0184-y

12. Ibrahim-Kosta M, Suchon P, Couturaud F, et al. Minor allele of the factor V K858 R variant protects from venous thrombosis only in non-carriers of factor V Leiden mutation. *Sci Rep*. 2019;9(1):3750. doi:10.1038/s41598-019-40172-x

13. Altshuler DM, Gibbs RA, Peltonen L, et al. Integrating common and rare genetic variation in diverse human populations. *Nature*. 2010;467(7311):52-58. doi:10.1038/nature09298

14. Rohlfs RV, Weir BS. Distributions of Hardy-Weinberg equilibrium test statistics. *Genetics*. 2008;180(3):1609-1616. doi:10.1534/genetics.108.088005

15. Wacholder S, Chanock S, Garcia-Closas M, El Ghormli L, Rothman N. Assessing the probability that a positive report is false: an approach for molecular epidemiology studies. *J Natl Cancer Inst*. 2004;96(6):434-442. doi:10.1093/jnci/djh075

16. Colhoun HM, McKeigue PM, Davey Smith G. Problems of reporting genetic associations with complex outcomes. *Lancet*. 2003;361(9360):865-872. doi:10.1016/s0140-6736(03)12715-8

17. Denny JC, Bastarache L, Roden DM. Phenome-Wide association studies as a tool to advance precision medicine. *Annu Rev Genomics Hum Genet*. 2016;17:353-373. doi:10.1146/annurev-genom-090314-024956

18. Sánchez-de-Madariaga R, Muñoz A, Castro AL, Moreno O, Pascual M. Executing complexity-increasing queries in relational (MySQL) and NoSQL (MongoDB and EXist) size-growing ISO/EN 13606 standardized EHR databases. *J Vis Exp*. 2018;(133):57439. doi:10.3791/57439

19. Dashti AA, Jadaon MM, Lewis HL. Factor V Leiden mutation in Arabs in Kuwait by real-time PCR: different values for different Arabs. *J Hum Genet*. 2010;55(4):232-235. doi:10.1038/jhg.2010.11

20. Gawish G E-H. The Prevalence of inherited thrombophilic polymorphisms in Saudi females with recurrent pregnancy loss confirmed using different screening protocols of PCR. *J Mol Genet Med*. 2015;9:156. doi:10.4172/1747-0862.1000156

21. Santamaria S, Reglińska-Matveyev N, Gierula M, et al. Factor V has an anticoagulant cofactor activity that targets the early phase of coagulation. *J Biol Chem*. 2017;292(22):9335-9344. doi:10.1074/jbc.M116.769570

22. Potkin SG, Turner JA, Guffanti G, et al. Genome-wide strategies for discovering genetic influences on cognition and cognitive disorders: methodological considerations. *Cogn Neuropsychiatry*. 2009;14(4-5):391-418. doi:10.1080/13546800903059829

23. Gribenko AV, Makhatadze GI. Role of the charge-charge interactions in defining stability and halophilicity of the CspB proteins. *J Mol Biol*. 2007;366(3):842-856. doi:10.1016/j.jmb.2006.11.061

24. Kawamura S, Tanaka I, Yamasaki N, Kimura M. Contribution of a salt bridge to the thermostability of DNA binding protein HU from Bacillus stearothermophilus determined by site-directed mutagenesis. *J Biochem*. 1997;121(3):448-455. doi:10.1093/oxfordjournals.jbchem.a021609

25. Wen X, Nicolae DL. Association studies for untyped markers with TUNA. *Bioinformatics*. 2008;24(3):435-437. doi:10.1093/bioinformatics/btm603

26. Angchaisuksiri P, Pingsuthiwong S, Aryuchai K, et al. Prevalence of the G1691A mutation in the factor V gene (factor V Leiden) and the G20210A prothrombin gene mutation in the Thai population. *Am J Hematol*. 2000;65(2):119-122. doi:10.1002/1096-8652(200010)65:2<119:: aid-ajh5>3.0.co;2-8

27. Boukhris I, Abdallah M, Bouslema K, et al. Venous thromboembolism: what other risk factors in a North African population? A study of 276 cases [in French]. *Ann Cardiol Angeiol (Paris)*. 2018;67(1):41-47. doi:10.1016/j.ancard.2016.12.001