SHORT THESIS FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY (PHD)

Genetic and Environmental Risk Factors Associated With Venous Thrombosis in the Hungarian Population

By Shewaye Fituma Natae

Supervisor: Dr.Szilvia Fiatal, MD, PhD

UNIVERSITY OF DEBRECEN

DOCTORAL SCHOOL OF HEALTH SCIENCES

DEBRECEN, 2024

Genetic and Environmental Risk Factors Associated with Venous Thrombosis in the Hungarian Population

By Shewaye Fituma Natae, PH, MPH

Supervisor: Dr.Szilvia Fiatal, MD, PhD

Doctoral School of Health Sciences, University of Debrecen

Head of the defence board: Prof. Dr. Endre Nagy, PhD, DSc

Reviewers: Prof. Dr. István Kiss, PhD, DSc
Dr. Zsolt Oláh, PhD

Members of the defence board: Prof. Dr. Istvan Balogh, PhD, DSc
Dr. Regina Molnar, PhD

The **PhD defence** takes place at the Lecture Hall of the Department of Internal Medicine, **Building 'A''**, Faculty of Medicine, University of Debrecen on **November 05, 2024, at 13:00.**

INTRODUCTION

Thrombosis is a multifactorial trait that contributes to the burden of cardiovascular diseases globally. It is a severe medical condition resulting from the formation of a blood clot in the lumen of the venous or arterial blood vessels, which leads to life-threatening illnesses such as pulmonary embolism (PE), stroke, and heart attack. Thrombosis is the primary cause of cardiovascular disease (CVD) related morbidity and mortality, which shares about 25% of all deaths globally. Based on the involved blood vessels; thrombosis broadly categorized as arterial and venous thrombosis.

Venous thrombosis (VT) is usually present as venous thromboembolism (VTE), which comprehends deep venous thrombosis (DVT) and pulmonary embolism (PE). DVT occurs when clotting blood forms in the large veins, mainly in the leg, thigh, and pelvis veins. Sometimes, it resolves spontaneously without causing further complications. However, as a result of the migration of thrombus into the pulmonary arteries, the DVT would result in PE, which results in the occlusion of blood flow into the lung and leads to severe medical conditions such as disability or death.

An overall estimated VTE incidence rate among people of European ancestry ranges from 104 to 183 per 100,000 person-years, with higher incidence in African-Americans, and lower in Asians. African-Americans are more frequently diagnosed with PE than DVT compared to Caucasians ancestry; 60% of VTE cases were presented with DVT. In European ancestry, DVT is more prevalent among the young, while PE is frequent among the elderly.

In Europe although the overall CVD-related morbidity is decreasing, mortality is substantially high. CVD is the principal cause of mortality in Europe and is attributed to over 3.9 million deaths annually. A higher burden of CVD-related mortality was reported in central and eastern European countries. Hungary shares the highest proportion of this mortality and CVD remain the prominent cause of death in Hungary. As of 2014, approximately 35,000 women and 27,000 men died of CVD, accounting for 55% and 45% of all deaths for women and men, respectively. The age-standardized death rate from CVD in Hungary is over twice the European Union (EU) average reported in 2014.

Around mid-19th century, Rudolf Virchow identified the triad of risk factors that contributes to thrombosis formation:-blood flow stasis, injury to the endothelium, and hypercoagulability. All VTE risk factors reflect this fundamental pathophysiologic processes. VTE risk factors broadly classified into heritable and non-heritable/acquired: the heritable prothrombotic risk factors

influence the VT risk via the coagulation process, whereas the non-heritable risk factors result in VT either through stasis or endothelial injury. Various studies have established the impact of heritable factors on VT risk. The incidence of hospitalization due to VT was two-fold higher for persons with affected families than for the general population.

Although abundant single nucleotide polymorphisms (SNPs) provoke the susceptibility of an individual to VT, it seems that the most strongly associated SNPs: rs6025 (Leiden mutation) in the *F5* gene, rs1799963 (prothrombin G20210A) in the coagulation factor two gene (*F2*), rs8176719 (non-O blood type) in the *ABO* gene, rs2036914 in the coagulation factor eleven gene (*F11*), rs2066865 in the fibrinogen gamma gene (*FGG*), and the rs121909567(ATBp3 mutation) in the *SERPINC1* gene, which play the prominent role in the determination of VT incidence and recurrence in genetically vulnerable individuals.

The Leiden mutation is one of the most dominant heritable VT risk factors that increase the burden of VT in genetically vulnerable individuals. The prevalence of Leiden mutation /*F5* is highest in European descent populations (3-15%), followed by Caucasian Americans (5.2%). However, it is rare in African-Americans (1.2%) and Asian-Americans (0.45%), and is not present at all in African populations. Likewise, prothrombin gene mutation / *F2*, often known as the G20210A mutation , is the second most common heritable VT risk in Caucasian ancestry (3-17%), particularly in European ancestry (4%) and Caucasian Americans (2%). However, it is less frequent in African Americans, which is approximately 0.4% (1 in 250), and extremely rare in African and Asian ancestry.

Often, due to their joint presence and their possible gene-gene interaction, the prothrombin gene mutation (rs1799963) and Leiden mutation (rs6025) SNPs were studied together. Furthermore, studies have shown that the ancestry distribution of coagulation Factor 11 (rs2036914) is similar in both Caucasian ancestry and African American ancestry. Studies have indicated that O blood type individuals are at lower risk of VT disease than other non-O blood type individuals, whereas the possibility of VT disease is higher among non-O blood type subjects. In addition, Fang *et al.* reported that the risk of venous thromboembolism (VTE) is higher for the African American ancestry and non-O blood type individuals than the Caucasian ancestry and O blood type individuals. Moreover, studies indicated that FGG (rs2066865) is more prevalent in Caucasian ancestry than African American ancestry.

Furthermore, studies have found that antithrombin deficiency plays a vital role in the pathogenesis of VTE. Antithrombin is an essential inhibitor of blood coagulation proteases;

individuals with hereditary AT deficiency have elevated thrombotic risk. Studies have revealed that the mutation profile of the AT gene (*SERPINC1*) is heterozygous. Formerly, it found that the prevalence of ATBp3 mutation was relatively high in the Roma population but not in the general Hungarian population.

In addition, studies have indicated that dual exposure to VT risk factors (genetic and non-heritable) increases the susceptibility of an individual to VT diseases. Aging and obesity are well-known non-heritable VT risk factors that hasten the occurrence of VT. As a result of multiple anatomical and pathophysiological changes elderly individuals are prone to age-related cardiovascular morbidity and mortality. Aging plays a principal role in the higher incidence of VT risk (1%) in elderly individuals.

Correspondingly, obese individuals were at higher risk of VT than normal-weight individuals. Formerly conducted studies showed that the risk of VT was 2-6-fold higher in obese individuals than in normal-weight individuals (BMI 20 to 24.9 kg/m^2). A study indicated that the risk of VT was higher among aged (>50 years old) and obese individuals. Furthermore, the existence other non-communicable diseases (CNCDs) such as cancer, diabetes mellitus (DM), chronic kidney diseases (CKDs), and coronary artery diseases (CADs) increased the likelihood of VT among individuals..

Some populations are susceptible to CVDs due to the coexistence of genetic and environmental risk factors. The Roma population is the most marginalized ethnic group in Central-Eastern European countries, with an estimated population of 8 to 12 million. They experience social exclusion, which intensely affects their health. A higher burden of disease, low life expectancy, low socioeconomic status, low education, and hazardous practices are common among Roma minorities. Consequently, cardiovascular risk factors are prevalent in the Roma population. On the other hand, due to the restriction of health-related data collection by ethnicity in the Hungarian Roma population, there are no data related to the incidence or prevalence of VT for this population. However, recent studies indicate that the Roma population is at higher risk of VT due to an elevated prevalence of metabolic syndrome and several heritable risk factors. Our previous study concluded that the Roma population seems to have increased genetic susceptibility to VT. Further investigation also suggested further study to compare the gene-environmental interaction (GxE) for VTE risk in the general Hungarian and Roma populations. Understanding how genetic and environmental risk factors interact provides insight for the early identification of risk groups within populations, allowing appropriate preventive and therapeutic measures to be taken.

To our knowledge, scant evidence compares gene and environmental risk factors associated with venous thrombosis among the general Hungarian and Roma populations. Thus, the main aim of the current study was to explore the interaction of environmental risk factors with six prothrombotic SNPs [(rs121909567 (*SERPINC1*), rs1799963 (*F2*), rs2036914 (*F11*), rs2066865 (*FGG*), rs6025 (*F5*), and rs8176719 (*ABO*)]. In addition, we further aimed to investigate the distribution of the rs121909567 (ATBp3) mutation in the *SERPINC1* gene in the Hungarian population, which contributes to the vast majority of antithrombin (AT) deficiencies in the Hungarian population due to its founder effect.

Moreover, the stratification of higher-risk individuals based on their genetic profiling for thromboprophylaxis is essential for efficient and effective utilization of available resources. Even though the 5-SNP impact on venous thrombosis risk is high, association and risk prediction models are seldom establish by using merely 5 strongly associated SNPs: [rs6025 (*F5 Leiden*), rs2066865 (*FGG*), rs2036914 (*F11*), rs8176719 (*ABO*), and rs1799963 (*F2*)]. To our knowledge, there is a lack of studies investigating the VT risk prediction ability of the combined five strongly associated prothrombotic SNPs in clinically confirmed VT subjects in the Hungarian population. Consequently, we aimed to explore the VT risk prediction ability of the combined five SNPs and conventional VT risk factors in the Hungarian population.

AIM AND OBJECTIVES

General objective

The main aim of our work was to investigate the role of selected genetic and environmental risk factors and the combined risk predictability of the five prothrombotic SNPs in the Hungarian population.

Specific Objectives:

To assess VT prevalence in the Roma and general Hungarian populations

To identify and compare the GxE and VTE risk in the Roma and general Hungarian populations

To explore the heritable VT risk in the Hungarian population

To determine the VT risk predictability of the combined five prothrombotic SNPs and conventional VT risk factors in the Hungarian population.

Methods and Materials

Study design, sample size, and data source

In this study, we used a comparative cross-sectional (study 1) and case-control study design (study 2). In total, 1532 subjects were identified to participate in this study (832 for comparative study and 700 for case-control study). However, due to incomplete data: 31 subjects (20 Roma and 11 subjects from the general Hungarian population) were excluded from the final analyses of the first study. Only those subjects who had complete genotype and covariate data (395 Roma and 406 general Hungarian subjects=801 subjects) were involved in this study (Flowchart 1A). Data management was similar in Study 2, i.e., two subjects were excluded due to missed genotype and covariate data. Finally, 698 (298 clinically confirmed VT cases and 400 healthy controls) participants were involved in the second study (Flowchart 1B).

The two counties found in northeast Hungary namely Szabolcs-Szatmár-Bereg and Hajdú-Bihar were taken as the source population. The population where the Hungarian Roma predominantly resided and the segregated Roma colonies located are taken as the study population. All the study populations involved in both studies (a comparative cross-sectional (1st study) and control group of a case-control (2nd study) were taken from a comprehensive health survey database collected for a cardio-metabolic comparative study among the Roma and general Hungarian population.

A comprehensive health survey database has 1000 study participants (500 general Hungarian and 500 Roma population) who met the selection criteria (Age 20-64 years old and residence of the indicated location). The study subjects who met the selection criteria were randomly selected from each household of the identified colony. In total, 25 colonies were randomly selected and 20 households from each colony were randomly identified for the study. One eligible individual was involved in the study from each household.

Data collection was made using structured questionnaires, physical examinations, and laboratory-based investigations from each study unit. The questionnaire-based data were collected through face-to-face interviews with data collectors who previously had data collection experience. Data collection related to physical examination and blood samples was done by general practitioners (GPs), and the whole data collection process was supervised by public health coordinators daily. The details of data collection and subject recruitment process were published by Ádány et al (2020) (134).

The case-control study design was conducted among 400 healthy control and 298 clinically confirmed VT cases to explore the heritable VT risk and to estimate the VT risk predictability of the combined strongly associated five SNPs (rs6025 (*F5 Leiden*), rs2066865 (*FGG*), rs2036914 (*F11*), rs8176719 (*ABO*), and rs1799963 (*F2*) and well-known non-heritable VT risk factors (age, obesity, and sex) in the Hungarian population. The control subjects were recruited from a comprehensive health survey database (134), whereas, the VT cases were recruited by the Division of Clinical Laboratory Science (DCLS), Department of Laboratory Medicine, Faculty of Medicine, University of Debrecen for one year. Color Doppler ultrasound or phlebography was used to diagnose the VT cases at the Department of Internal Medicine, Faculty of Medicine, University of Debrecen.

Personal and Environmental risk factors

For a comparative cross sectional study survey, the environmental, and personal risk factors data that affect the VTE risk of an individual were extracted from a comprehensive health survey database. Generally, the following data were included:

Chronic diseases such as cancer, DM, CAD, and CKD

Lipid parameters data (TC, LDL-C, HDL-C, and TG),

Lifestyle factors and anthropometric data such as smoking status, BMI, and WC were collected from the comprehensive health survey database.

However, for study 2 (case-control study) we only consider the well-known conventional VT risk factors (Age, obesity, and Sex). To verify the risk of VT among elderly and obese subjects, we used age ≥60 years, and BMI ≥30 kg/m^2 as cut-off values during the data analysis and we used ages below 60 years and BMI below 30 kg/m2 as reference.

DNA extraction and Genotyping

DNA was extracted from the peripheral blood using a MagNA Pure LC system (Roche Diagnostics, Basel, Switzerland) with a MagNA Pure LC DNA Isolation Kit–Large Volume according to the manufacturer's instructions. Extracted DNA was eluted in 200 µl MagNA Pure LC DNA Isolation Kit- Large Volume elution buffer. The assay design and the genotyping were performed by the Karolinska University Hospital, Stockholm, Sweden Mutation Analysis Core Facility (MAF). A MassARRAY platform (Sequenom, CA, USA) with iPLEX Gold chemistry was used for genotyping. Quality control, validation, and concordance analysis were conducted by MAF.

SNPs selection

For both studies, the five strongly associated and influential prothrombotic SNPs: rs1799963 (*F2*), rs6025 (*F5*, Leiden), rs2066865 (*FGG*), rs2036914 (*F11*), and rs8176719 (*ABO*) were considered due to their large effect size on the GWAs and strong relationships with the higher incidence and recurrence rate of VT risk in the formerly conducted studies. In addition, due to the indicative result of a previous studies, which identified *SERPINC1* (rs121909567)/ Antithrombin Budapest3 (ATBp3 mutation; p.Leu131Phe) as the common cause of the antithrombin (AT) deficiency in Hungary; particularly the Roma population we consider it in our comparative cross-sectional study.

Genetic risk score (GRS) computations

To establish the pooled effect of the included prothrombotic SNPs: rs6025 (*F5* Leiden), rs2066865 (*FGG*), rs2036914 (*F11*), rs8176719 (*ABO*), and rs1799963 (*F2*) on VT risk we also computed weighted and unweighted genetic risk score. In the genetic risk score (GRS), the individuals were assigned a score based on the number of risk alleles they carried. Accordingly, "0", "1", and "2" codes were given for the absence of risk alleles, heterozygous, and homozygous for risk alleles, respectively. When the risk allele was found protective, the coding for the homozygous risk allele becomes "0", while "2" for the other homozygous allele. Unweighted-GRS (unGRS) was simply computed by adding all risk alleles in the loci by

assuming that all alleles have the same effect. Whereas, weighted genetic risk score (wGRS) was computed with the assumption that, SNPs with larger effects contributed more to the GRS. Weights were derived from the risk coefficient for each allele depending on the odds ratio (OR) reported in the former genetics associations study. Median values of wGRS and unGRS were used to compare the association between genetics risk score and VTE risk factors in the study populations.

Statistical Analysis used

Study 1

Statistical tests were computed using IBM SPSS Version 25 statistical software. The Shapiro-Wilk normality test was used to test the distribution of quantitative variables. Non-normally distributed variables were transformed using a two-step transforming approach proposed by Templeton. The presence of Hardy-Weinberg Equilibrium (HWE) and allele frequency differences of all included SNPs between the two study populations was evaluated by using PLINK statistical software Version 1.9. Logistic regression analysis was used to determine the associations between individual SNPs, environmental risk factors, and VTE. A multivariable linear regression analysis with 95% CI, was used to test the impact of GxE on VTE risk after interaction terms between 6 SNPs and environmental risk factors were generated. An interaction term was created between each SNP and environmental risk factor to assess the combined effect of them on VTE risk. All analyses were adjusted for age and reported p-values were two-sided, and α level of 0.05 was used to define statistical significance.

Study 2

Statistical tests were computed using PLINK (version 1.9) and IBM SPSS (version 26.0) statistical software. The Mann–Whitney U test was used to compare the age, BMI, and GRS distribution in the study populations. The Shapiro–Wilk normality test was used to test the distribution of quantitative variables. The presence of Hardy-Weinberg equilibrium (HWE) and risk allele frequency differences between VT cases and controls were estimated using the X^2 test. The association between the five SNPs and VT disease risk was assessed by the odds ratio (ORs) with their respective 95% confidence interval (CI) under all genetic models. Likewise, logistic regression analysis was also used to compute the OR with 95% of individual SNPs, genetic, non-heritable, and combined VT risk factors.

In addition, the receiver-operating characteristic (ROC) curve was determined to assess how well its score classifies venous thrombosis patients and control subjects. To determine which SNP is more influential in their discriminatory accuracy of the area under the receiver operating characteristics curve (AUC), we added each SNP one-by-one into a model.

We compared the AUCs of genetic, non-heritable, and combined risk models. We included each non-heritable risk factor and their combination with genetic VT risk factors into a model to verify the difference in AUC value and their VT risk predictability in the study population. A logistic regression model was used to generate a combined risk score of genetic and non-heritable VT risk factors. SPSS version 26.0 was used to calculate ROC curves and AUCs. Bonferroni multiple testing was employed to prevent the problem associated with multiple comparisons (0.05/5= $p<0.01$). Statistically significant variables were declared at a conventional p value of 0.05.

Ethical approval

The committee of the Hungarian Scientific Council on Health Research approved the protocol (61327-2017/EKU). Written informed consent was obtained from all study subjects.

Results

Characteristics of the study participants

In total, 1499 subjects (406= general Hungarian, 395=Roma population, 400= Healthy control, and 298=Clinically confirmed VT cases) who had complete genotype and covariate data were included in both surveys. In a study 1, the female proportion was higher in the Roma population and the difference is statistically significant (55.4% vs. 73.9%; $p<0.001$). However, in a study 2; the elderly and male subjects were highly prevalent in the cases group than in the control group. The observed differences were statistically significant. The mean age of the VT cases subjects was higher than that of the control group (63.4±16.4 years vs. 43.8±12.6 years, $p<0.001$), respectively.

VTE morbidity in the study populations

In a study1, only 6 (1.5%) and 12 (3.0%) of the general Hungarian and Roma populations reported VTE during the survey, respectively. The proportion of VTE cases was higher among the Roma population; in particular, the VTE risk was higher among female subjects. However, in study 2, the VT cases were higher among the male participants than the female participants.

Risk allele frequency comparison in the study population

After multiple correction testing emplaced, the allele frequency of the *SERPINC1* SNP (rs121909567) remains statistically significant. The *SERPINC1* (rs121909567) SNP is only present in the Roma population and does not occur in the general population. However, in a study 2, the risk allele frequencies of the rs6025 (*F5*), rs2036914 (*F11*), and the rs8176719 *(ABO* gene) were higher in the case group than in the control group, and the differences remained statistically significant after multiple correction testing ($p<0.01$). Nevertheless, before the multiple correction testing except for *F2* gene or Prothrombin mutation (rs1799963) the allele frequencies of the remaining SNPs: rs6025 (*F5*), rs2066865 (*FGG*), rs2036914 (*F11*), rs8176719 (*ABO*) were significantly distinct ($p<0.05$) among cases and control group.

Associations between SNPs and VT risk in the study populations

Associations of individual SNPs with VTE (Study 1)

In this study, only the rs6025 (Leiden mutation) and the rs2066865 (*FGG*) were significantly associated with VTE risk and found to be nominally significant for the Roma population but not for the general population. Study subjects who had homozygous risk alleles for the *FGG* gene were 5.9 times more at risk for VTE than subjects without a risk allele (OR=5.9; 95% CI: [1.23,28.4]). Furthermore, people with the heterozygous risk allele of the Leiden mutation (rs6025) were 3.8 times more likely to develop VTE than individuals without the risk allele, and the risk was higher for the Hungarian Roma population.

Gene-environmental interaction (GxE) and VTE risk

The finding of the GxE and VTE risk in our study was reported using the standardized beta value of the multivariate linear regression analysis. We found several statistically significant multiplicative interactions on VTE risk for the majority of the included SNPs, such as rs2036914 (*F11*), rs2066865 (*FGG*), rs6025 (*F5*, Leiden), and rs8176719 (*ABO*). Since the ATBp3 mutation was not present in the general population, regression analysis was only performed for the Roma population. Although the trend of their relationships indicated the possibility of higher VTE risk, the multiplicative interaction was not statistically significant for the *SERPINC1* (rs121909567) and F2 (rs1799963) genes (p-value for interaction >0.05).

According to the regression coefficient of the multivariate linear regression analysis, the likelihood of VTE was higher among the study subjects who had cancer, DM, CAD, and CKD.

Furthermore, cigarette smoking, a high level of LDL-C, and obesity increased the VTE risk for the study subjects. The observed multiplicative interaction and VTE risk were bidirectional: a positive beta (β) value indicated a VTE risk increment as a result of GxE whereas a negative beta (β) value indicated the reverse (risk decrements). The presence of high levels of LDL-C and the rs2066865 (*FGG*) risk variant makes Roma subjects to be at higher risk of VTE (β= 0.389, p= 0.002); however, the joint presence of those risk factors did not increase the VTE risk in the general subjects (β=0.048, p=0.70). The existence of a multiplicative interaction between CAD and rs2036914 (*F11*) increases the VTE risk among both populations: for the Roma population β= 0.280 (p= 0.001) and for the general Hungarian population β= 0.423 (p= 0.001)

As a result of the multiplicative interaction between rs2066865 (*FGG*) and CAD, VTE risk was higher for the Roma population (β= 0.143, p= 0.046) but not for the general Hungarian population (β= -0.329, p <0.001). The interaction between rs6025 (*F5, Leiden*) and smoking (β= 0.172, p= 0.008) as well as Leiden and LDL-C (β= 0.368, p= 0.001) increased the risk of VTE for the general population only but not for the Roma population (β= -0.014, p= 0.86 and β= -0.150, p= 0.55, respectively).

Our study also identifies the higher risk of VTE as a result of a multiplicative interaction between rs8176719 (*ABO*) and cancer, and the risk was higher for the Roma population (β=0.370, p<0.001) than for the general Hungarian population (β=-0.042, p=0.6). Nevertheless, the interaction of rs8176719 (*ABO*) with CAD, (β= 0.197, p= 0.009) significantly increased VTE risk only for the general Hungarian population. The risk of VTE was higher for general Hungarian subjects (β= 0.194, p <0.01) who had diabetes mellitus and non O blood type but not for the Roma subjects (β=-0.039, p=0.63).

Case control study (study 2)

Associations between SNPs and VT risk in the study population using genetic association models (study 2)

The strengths of association regarding VT risk using complete genetic association models [multiplicative, additive, dominant, recessive, and genotypic model] were estimated. Only the Leiden mutation (rs6025) and *F11* (rs2036914) remained significant after multiple testing adjustments (p<0.01). In particular, the Leiden mutation variant strongly influenced the VT disease risk in the Hungarian population (p<0.001): the odds of VT disease risk among VT cases due to the Leiden mutation ranged from 3.25 (heterozygous genotypic for risk variant (OR=3.25, 95% CI: 2.22; 4.76) to 19.67 (OR=19.6, 95% CI: 2.57; 150.4) in the recessive model/homozygous for risk variant.

The rs8176719 (*ABO*) remained statistically significant only in the multiplicative (OR=1.33, 95% CI: 1.07; 1.64**)** and genotypic models (OR=1.77, 95% CI: 1.14; 2.73); nevertheless, it lost its significance in other models after multiple testing adjustments. Similarly, the protective variant of rs8176719 (*ABO*) remained statistically significant only in multiplicative (OR=0.75, 95% CI: 0.61; 0.93**).** Additionally, the *FGG* (rs2066865) expressed a significant association with VT risk in the multiplicative, additive, and dominant models before multiple testing; however, it lost its significance after adjustment was performed. Conversely, the *F2* (rs1799963) did not show any statistically significant association with VT directly with any of the used models**.**

Comparison of genetic risk scores and risk of VT in study populations

Comparative cross sectional study (study 1)

In this study, wGRS was computed only using 5 SNPs which showed a strong association with VTE from previously conducted studies (63,128). The wGRS ranges from 0.0 to 4.7 and 0.0 to 4.6 for the general Hungarian and Roma populations, respectively. The mean wGRS was 1.8 ± 0.8; 95% CI [1.7, 1.9] for the general Hungarian and 1.9 ± 0.76; 95% CI [1.8, 1.9] for the Roma population. The unweighted GRS was calculated for 6 SNPs, and ranges from 0.0 to 7.0 for both populations, with a mean of 2.7 ± 1.2; 95% CI [2.6, 2.8] for Hungarian general and 2.8 ± 1.2; 95% CI [2.7, 2.9] for the Roma population.

Case-control study (study 2)

The unGRS and wGRS for 5 SNPs were computed for 298 VT patients and 400 healthy controls. The unGRS ranged from 0 to 6 (3.46±1.31) and 0 to 7(2.77±1.28) for VT patients and healthy controls, respectively**.** The wGRS ranged from 0 to 4.6 (1.93±0.97) for VT patients and 0 to 4.7 (1.37±0.78) for healthy control.

Associations of GRSs with VT risk

The distributions of other covariates including GRSs were significantly distinct ($p<0.001$) between study groups**.** The test revealed significant differences (cases vs. control) in age (median=65; 44, $p<0.001$), BMI (median=28.72; 26.75, $p<0.001$), unGRS (median=3; 3, $p<0.001$), and wGRS (median=1.79; 1.34, $p<0.001$). Although the median values of unGRS for the VT patients and healthy controls were the same, higher value of unGRS (3) was more frequent in the cases than in the healthy controls.

Of the well-known non-heritable VT risk factors; age and obesity were significantly associated with VT risk in the study population, and the risk was higher in the case groups than in the control group. The risk of VT among elderly subjects aged ≥60 years old was 12.8 times higher than that of among subjects aged below 60 years (AOR=12.83, 95%CI: 8.38; 19.63). Likewise, the VT risk was 2.3 times higher for obese subjects **(**BMI>30 kg/m2) than for normal-weight subjects (AOR=2.28, 95%CI: 1.51**;** 3.42). Furthermore, the wGRS remained statistically significant after we adjusted for both sex and age [AOR=2.69, 95%CI: 1.74; 4.19 and AOR=2.24, 95% CI: 1.51; 3.32], respectively. However, the unGRS lost its statistical significance (AOR=0.88, 95% CI: 0.65; 1.18) in the multivariate regression analyses model.

Venous thrombosis risk prediction in a study population (study 2)

For a case-control study, we calculated the receiver operating characteristic curve (ROC) to assess how well the score classified venous thrombosis in patients and control subjects. The AUC value of the included SNPs ranges from 0.51 [95%CI: 0.47; 0.55, p-value 0.64] for rs1799963 in *F2* up to 0.62 [95%CI: 0.57; 0.66, p value<0.001] for rs6025 in F5. The discriminative accuracy of the model improved with the addition of each SNPError! Reference source not found.. We started addition with the Leiden mutation (rs6025) with the highest effect size and ended with rs2036914 (*F11*) with the lowest effect size of the included 5 SNPs. The addition of each SNP increases the AUC value after *F2* (rs1799963).

The AUC of the 5-SNP risk score was 0.68 [95% CI: 0.64; 0.72]. The proportions of variability explained by the Leiden mutation (rs6025) were higher than that of the 5-SNP risk score (8% vs. 7%). Furthermore, approximately 39% of the variability observed was attributed to the combination of genetic and non-heritable risk factors, which is higher than those factors independently. Likewise, the ROC curve for the weighted 5-SNP risk score had an AUC of 0.68 [95% CI: 0.64; 0.72], i.e., there is a 68% probability that a randomly selected patient will have a higher score than a randomly chosen control subject. The wGRS was a better predictor than the unGRS [AUC= 0.65, 95% CI: 0.60; 0.69].

Risk prediction based on a combination of genetic and non-heritable risk factors

We also evaluated the discriminative accuracy of well-known non-heritable VT risk factors such as age, sex, and obesity to explore their independent and combined VT risk predictability in the study subjects. The independent AUCs of age and obesity were 0.84 (p value<0.0001)

and 0.59 (p value<0.001), respectively. The combination of all well-known VT risk factors changes the discriminative accuracy of the AUC in to 0.85, p value<0.000. Likewise, when we added the non-heritable risk factors into the genetic risk factors, the AUC significantly projected to 0.89 [95% CI:0.86;0.91] compared to genetic (AUC=0.68) or non-heritable risk factor predictability (AUC=0.85; p<0.0001). The AUC difference of the combined and non-heritable risk factors were significant (AUC=0.039, 95% CI: 0.02; 0.059, p value<0.0001).

Discussion

Overview of the studies

In our present study, we tried to investigate the VT risk background in the Hungarian population using comparative cross sectional and case-control study design to verify the gene-environmental interaction (GxE) and VT risk predictability of strongly associated VTE SNPs and other comorbidities. Our study is the first to explore and compare GxE and VTE risk in this population. The present study revealed evidence as the Roma population was at higher risk of VTE due to dual exposure and the possibility of the synergistic interaction of heritable and non-heritable VTE risk factors.

Multiplicative interaction and VTE risk

Our study explores the possibility of interaction between the genetic and environmental risk factors that are divergent in both populations. A significant multiplicative interaction was observed between the rs8176719 (*ABO*) gene and diabetes mellitus and CAD for the general Hungarian population but only with cancer for the Roma population. The regression coefficient of multiplicative interactions of DM and rs8176719 (*ABO*) was synergistic and significant. The finding demonstrated that non-O blood group individuals with diabetes had higher VTE risk than O blood group subjects. Our study is in line with a systematic review and meta-analysis of cohort study results that show the risk of VTE is 1.4 times higher among diabetes mellitus individuals than non-diabetic individuals.

Likewise, the risk of developing VTE was 6 to 7 times higher in cancer patients. The result is along with our study findings that reveal the VTE risk is higher for the Roma subjects who had cancer. The risk of VTE is higher among Roma subjects who had cancer and non-O blood group. Our present work findings also support former related studies' results which demonstrated the existence of cancer increases the likelihood of VTE. Further studies on cancer and prothrombotic genotypes highlight VTE risk increased by 11-12 fold due to the concurrent presence of cancer and rs8176719 (ABO) risk variant. The authors also found that 39% and 30% of VTE risk was attributed to the joint presence of non-O-blood group and cancer.

Furthermore, our study revealed a synergistic and statistically significant association between cigarette smoking and the Leiden mutation (rs6025) variant carrier which increases the risk of VTE in those subjects. The risk of VTE is higher among general Hungarian subjects who carry the risk variant for Leiden mutation (rs6025) and smokers. Prior studies also showed that the combined effect of rs6025 (*F5*) and smoking increased the risk of VTE. A large population-based case-control study also supported our finding, where the joint effect of rs6025 (*F5*) and current smoking resulted in a 5.0-fold increased VTE risk. Another cohort study revealed that the concurrent existence of smoking other than rs6025 (*F5*) increased the VTE risk by 51% and 10% at 10 years for homozygous and heterozygous risk variants, respectively. Crous-Bou et al. also found an additive interaction between prothrombotic SNPs and smoking that increased VTE risk.

Likewise, individuals with CAD and rs2036914 (*F11*) or rs8176719 (*ABO*) were at higher risk of VTE than individuals without CAD and those variants. The existence of an interaction between rs2036914 (*F11*) and CAD increased the VTE risk among general Hungarian and Roma populations who carry the risk variant for coagulation factors eleven (rs2036914) and rs8176719 (*ABO*). Our finding was also consistent with a study that found that myocardial infarction patients with ≥1 risk allele at rs2036914 (*F11*) had a 1.8-fold higher risk of pulmonary embolism than non-carriers. Besides, the risk of VTE was 1.5-fold higher among individuals with non-O blood type and myocardial infarction.

Additive interaction and VTE risk

Our study also explores an additive interaction between a high level of LDL-C, migraine, current cigarette smoking, and ≥3 wGRS value. The risk of VTE increased by 3.2-fold for the Roma subjects with a high level of LDL-C and ≥3 risk alleles. This result was consistent with a GWAs which revealed one standard deviation (SD) of elevated LDL-C was associated with an increased risk of VTE. The finding of current cigarette smoking was in agreement with the study of Crous-Bou et al., in which the relationships between current smoking and VTE genetic factors were additive. In our study, individuals who had experience with migraine in addition to a wGRS value ≥3 had a 3.9 times higher risk for VTE than individuals with either a wGRS ≥3 or migraine. This finding was in harmony with a study by Peng et al., which revealed that migraine headaches increased the risk of VTE. Altogether, our comparative study among general Hungarian and Roma populations demonstrated the simultaneous presence of comorbidities such as CAD, DM, cancer, cigarette smoking and genetic VTE risk factors increased the VTE risk among exposed individuals. Furthermore, our study confirmed as the rs121909567 (*SERPINC1*, ATBp3) is a founder mutation among the Roma population but not for the general Hungarian population.

Although the finding was highly subjected to selection and observation biases due to the small number of cases and the study design, our study reveals some clues about the burden of the joint presence of genetic and environmental risk factors on VTE risk. Due to higher genetic load and gene-environmental interactions, this minority Roma population is at higher risk of VTE than the general Hungarian population. Thus, our results suggest that an intensive search for the rs121909567 (*SERPINC1*; ATBp3) founder mutation might be an important factor for the assessment of thrombotic disease susceptibility among the Roma population. In addition, we strongly recommend further studies among a large number of VTE cases to explore the more precise impact of genetic and environmental risk factors on VTE in the Hungarian populations.

The fact of strongly associated SNPs and VT risk determination in a study population

Based on our former study finding we further conduct a case-control study using 298 clinically confirmed VT cases and 400 healthy control to explore the genetic background of VT risk in the Hungarian population and to determine the combined VT risk predictability of strongly associated SNPs: rs6025 (*F5 Leiden*), rs2066865 (*FGG*), rs2036914 (*F11*), rs8176719 (*ABO*), and rs1799963 (*F2*) and well-known conventional VT risk factors (Age, Sex, and Obesity) in the population. In this study, we found three SNPs: rs6025 (*F5*), rs2036914 (*F11*), and rs8176719 (*ABO*), which is remained statistically significant after adjustment for multiple testing ($p<0.01$). The highest VT risk was detected among Leiden mutation carriers/rs6025 [OR=3.52, 95% CI: 2.50; 4.95. Its allele frequency was approximately 3-fold higher in VT cases (20%) than in control (6.8%) groups. Our findings are consistent with previously conducted studies that indicated the odds of VT risk are 3.5 and 4.38 times higher for rs6025 (*F5*) variant carriers than for noncarriers.

Moreover, numerous studies suggest that the Leiden mutation is vastly prevalent in Caucasian ancestry, particularly in European descent. It is one of the most influential heritable VT risk factors that increase the burden of VT in genetically vulnerable individuals. These findings support our study result that showed that F5 is highly prevalent in cases (20%). Furthermore, it was strongly associated with the trait in all genetic association models. This highlights the fact that the Leiden mutation is an independent predictor of venous thrombosis risk, and its contribution to the burden of VT disease is remarkable, particularly for genetically susceptible individuals and Caucasian ancestry.

Likewise, the risk allele frequency and VT risk for rs2036914 (*F11*) and rs8176719 (*ABO*) were higher for cases even after adjustment for multiple testing. A previously conducted study revealed that *F11* (rs2036914) is an independent predictor of VT, which is supported by our result that the risk was 1.38 times higher for cases than for controls. Studies have shown that VT risk distribution due to F11 (rs2036914) is similar in Caucasians and African Americans. The allele frequency of the ABO SNP (rs8176719) was more prevalent in the control group (47.4% vs. 40.4%). Additionally, it was revealed that the risk of VT was lower (OR= 0.75, 95% CI: 0.61; 0.93) among subjects with the rs8176719 variant. Furthermore, the VT risk was 1.33 times higher for risk variant carriers. Studies have indicated that individuals with the O blood type are at the lowest risk of VT compared with individuals without the O blood type.

On the other hand, numerous studies have shown that non-O blood type subjects (A, AB, and B) were at higher risk of developing venous thrombosis compared with O blood type subjects. Our study findings are also consistent with those of previously conducted studies. Fang *et al.* reported that the risk of VTE is higher for the African American race and non-O blood type individuals than the Caucasian race and O blood type individuals.

Studies have indicated that pooled variants have more impact on VT risk determination than a single variant. Our findings showed that the wGRS is an independent predictor of VT risk in the study population, and the value was 2.37 times higher for VT cases than for controls. Formerly conducted studies also support our findings.

Non-heritable VT risk factors and their impact

The impact of non-heritable risk factors on VT risk is also appreciable. Our study showed that VT was more prevalent in elderly ($\geq$ 60 years) subjects (58.1% vs. 10.5%; $p<0.0001$). Likewise, the odds of VT risk for elderly subjects were 12-fold higher than for those aged <60 years. The risk of VT increases with age due to different factors, such as anatomical, pathophysiological, and hormonal derangement. Consequently, it hastens and increases the vulnerability of elderly individuals to VT risk and other CVDs. Furthermore, our findings show that the risk of VT is 2.28 times higher for obese subjects than for normal-weight subjects. This finding is in line with the results of previously conducted studies that show that obesity is an independent predictor of VT risk.

VT risk predictability: independent SNPs, non-heritable, and their combination in the Hungarian population

The ability to predict the risk of a certain event before its occurrence is important in clinical epidemiology. Precise risk prediction helps control an event at as early a stage as possible and offers to use the available resources effectively and efficiently. We used ROC curves to establish individual and combined VT risk predictability of the SNPs and non-heritable VT risk factors to develop a risk stratification tool.

In our study, the highest AUC result was obtained for the Leiden mutation (AUC=0.62), whereas the lowest AUC was obtained for prothrombin mutation /F2 (0.52). The addition of each SNP into the model after *F5* increases the AUC value in general. Our finding is in line with studies that showed that the addition of more SNPs into the model increases the AUC value to a certain extent, but after a certain level, the AUC value does not change, although they added more explanatory variables into the model.

We also found that the wGRS is a better predictor of VT risk than individual SNPs (0.68 vs. 0.62), and their combination with non-heritable risk factors yields a larger AUC value with higher discriminatory accuracy (AUC=0.89). This finding is consistent with previous studies showing that the combination of clinical and genetic risk factors increases the risk of VT 8 times more than either the genetic or the clinical model alone.

Conclusion and Recommendations

In general, the Hungarian population is at higher risk of VT due to aging and the presence of heritable VT risks such as Leiden mutation, *F11*, and *ABO*. Particularly, the Leiden mutation influences VT in the Hungarian population with the largest risk variants frequency. The Leiden mutation is the most prevalent genetic VT risk in the Hungarian population. The pooled variant predicts the VT risk in the Hungarian population. Moreover, the combination of non-modifiable VT risk and the heritable risk increases the VT risk predictability accuracy of the model. Therefore, the stratification of highly vulnerable individuals based on their genetic profiling and well-known nonmodifiable VT risk factors is important for the effective and efficient utilization of preventive and control measures of VT risk.

New Findings

Risk allele frequencies, GXE

The SERPNC1 risk allele frequency is only present in the Roma population but not at all in the general Hungarian population. Our study also confirmed the result of former study which suggested the origin of SERPINC1 mutation is Roma population.

The presence of heritable and non-heritable risk factors increases the VT risk in exposed subjects. Our study verified that as the result of the synergistic multiplicative interaction between the comorbidities such as DM, CAD, CKD, cancer, obesity, cigarette smoking, and heritable risk factors is increasing VT risk among dual exposed subjects.

In our case control study we explore the three SNPs (rs6025 (*F5*), rs2036914 (*F11*), and rs8176719 (*ABO*),) risk allele is remained statistically significant after multiple correction testing and the risk is higher for cases than control group. The highest VT risk was detected among Leiden mutation carriers/rs6025 [OR=3.52, 95% CI: 2.50; 4.95. Its allele frequency was approximately 3-fold higher in VT cases (20%) than in control (6.8%) groups.

Aging (>60yrs); AOR=12.83, 95%CI: 8.38; 19.63) and obesity (AOR=2.28, 95%CI: 1.51; 3.42) independently determine the VT risk in the Hungarian population.

Strongly associated VT SNPs, non-heritable VT risk factors, and their VT risk predictability in the Hungarian population.

Of the five studied strongly associated SNPs: (rs6025 (*F5 Leiden*), rs2066865 (*FGG*), rs2036914 (*F11*), rs8176719 (*ABO*), and rs1799963 (*F2*); the Leiden mutation SNP (rs6025) AUC=0.62, 95%CI: 0.57; 0.66, p value<0.001) is higher and more predict the VT risk in the Hungarian population compare to other studied SNPs. The pooled variants, particularly, the weighted genetic risk score (wGRS) (AUC=0.68, 95% CI: 0.64; 0.72) more predict the VT risk than the independent SNPs. The combination of both heritable and non-heritable VT risk factors was best to predict the VT risk in the Hungarian population (AUC=0.89, 95% CI: 0.86; 0.91) compared to genetic (AUC=0.68). Also, their differences is statistically significant (AUC=0.039, 95% CI: 0.02; 0.059, p value<0.0001).

Generally, our study revealed the fact that efficient and effective utilization of available resource could be used to predict the disease risk in a given population and take certain measure to control diseases burden. Accordingly, we propose stratifying highly vulnerable individuals based on their genetic profiling and well-known VT risk for the efficient and effective utilization of preventive and control measures

Acknowledgements

First and foremost, I would like to thank my Lord and His Mother St. Virgin Mary for their support and unexplained gift during my entry life and in my PhD journey.

I would also like to extend my thanks and appreciation to my supervisor Dr. Szilvia Fiatal for her unreserved support, guidance, and encouragement to reach at this stage.

I would also like to extend my sincere gratitude to the Dean of the Doctoral School of Health Sciences, Professor Mariann Harangi, the secretarial, Dr.Mónika Katkó and the head of the Department of Public Health and Epidemiology, Professor János Sándor, and all staff of the Department of Public Health and Epidemiology who support me in a different perspective for the fruitfulness of my study.

My sincere thanks to my special and strong mother Aregash Guddisa, my brother Jinenus Fituma, and my children Meskerem Tarressa, Soreti Dereje, and Bona Ayele for their patience and unexplained encouragement for the success of my PhD Journey.

Lastly, I dedicate this dissertation to the Soul of my Hero Father Fituma Natae Murti; who has always been my energy, confidence, value, and motivator to be a risk taker, influential, and successful woman in any journey in addition to my study. I always pray and ask my GOD to rest his Soul in heaven. Amen, Amen, Amen.

Funding

This work was supported by the Stipendium Hungaricum Scholarship Programme of the Tempus Public Foundation, the European Union under the European Regional Development Fund (GINOP-2.3.2-15-2016-00005), and the National Research, Development and Innovation Office, Hungarian Ministry of Innovation and Technology (Grant No. OTKA K139293).

Publications

UNIVERSITY AND NATIONAL LIBRARY
UNIVERSITY OF DEBRECEN
H-4002 Egyetem tér 1, Debrecen
Phone: +3652/410-443, email: publikaciok@lib.unideb.hu

Registry number: DEENK/507/2023.PL
Subject: PhD Publication List

Candidate: Shewaye Natae
Doctoral School: Doctoral School of Health Sciences
MTMT ID: 10086242

List of publications related to the dissertation

1. **Natae, S.**, Merzah, M., Sándor, J., Ádány, R., Bereczky, Z., Fiatal, S.: A combination of strongly associated prothrombotic single nucleotide polymorphisms could efficiently predict venous thrombosis risk.
   Front. Cardiovasc. Med. 10, 1-11, 2023.
   DOI: http://dx.doi.org/10.3389/fcvm.2023.1224462
   IF: 3.6 (2022)
2. **Natae, S.**, Kósa, Z., Sándor, J., Merzah, M., Bereczky, Z., Pikó, P., Ádány, R., Fiatal, S.: The Higher Prevalence of Venous Thromboembolism in the Hungarian Roma Population Could Be Due to Elevated Genetic Risk and Stronger Gene-Environmental Interactions.
   Front. Cardiovasc. Med. 8, 1-13, 2021.
   DOI: http://dx.doi.org/10.3389/fcvm.2021.647416
   IF: 5.846

List of other publications

3. Merzah, M., Póliska, S., Balogh, L., Sándor, J., Szász, I., **Natae, S.**, Fiatal, S.: A Transcriptomic Analysis of Smoking-Induced Gene Expression Alterations in Coronary Artery Disease Patients.
   Int. J. Mol. Sci. 24 (18), 1-14, 2023.
   DOI: http://dx.doi.org/10.3390/ijms241813920
   IF: 5.6 (2022)
4. **Natae, S.**, Nigatu, D. T., Negawo, M. K., Mengesha, W. W.: Cervical cancer screening uptake and determinant factors among women in Ambo town, Western Oromia, Ethiopia: community-based cross-sectional study.
   Cancer Medicine. 10 (23), 8651-8661, 2021.
   DOI: http://dx.doi.org/10.1002/cam4.4369
   IF: 4.711

UNIVERSITY AND NATIONAL LIBRARY
UNIVERSITY OF DEBRECEN
H-4002 Egyetem tér 1, Debrecen
Phone: +3652/410-443, email: publikaciok@lib.unideb.hu

5. Merzah, M., Kósa, Z., Sándor, J., **Natae, S.**, Pikó, P., Ádány, R., Fiatal, S.: Roma Socioeconomic Status Has a Higher Impact on Smoking Behaviour than Genetic Susceptibility.
   Int. J. Environ. Res. Public Health. 18 (6), 1-12, 2021.
   DOI: http://dx.doi.org/10.3390/ijerph18063206
   IF: 4.614

6. Nigatu, D. T., Seme, A., **Natae, S.**, Segni, M. T.: Sexual initiation and associated factors among young women in West Shoa, Ambo Town, Ethiopia: a community-based cross-sectional study.
   BMC Womens Health. 18 (1), 1-6, 2018.
   DOI: http://dx.doi.org/10.1186/s12905-018-0563-7
   IF: 1.592

Total IF of journals (all publications): 25,963
Total IF of journals (publications related to the dissertation): 9,446

The Candidate's publication data submitted to the iDEa Tudóstér have been validated by DEENK on the basis of the Journal Citation Report (Impact Factor) database.

13 November, 2023

DEBRECENI EGYETEM
Egyetemi és Nemzeti Könyvtár