

A New Value Proposition For A Genetic Test Of Hereditary Thrombophilia

Beatriz Melo Silveira
beatriz.silveira@tecnico.ulisboa.pt

Instituto Superior Técnico, Lisboa, Portugal

November 2018

Abstract

Having a genetic predisposition to the formation of blood clots - hereditary thrombophilia - comes with potentially fatal consequences, being one of the risk factors for developing venous thromboembolism (VTE). VTE is characterized by the pathological formation of blood clots in veins, which can possibly enter blood circulation and obstruct another vessel in the body - embolism. The chances of survival after a VTE event tend to decrease over time, being 41.5% eight years after the first pulmonary embolism (PE) [1]. The annual incidence rate of VTE among people of European ancestry varies between 104 and 183 per 100,000 person-years [1]. Given the negative impact of VTE in health, a diagnostic tool which estimates the genetic predisposition to abnormal coagulation is of great value. The goal of this dissertation was to develop a new value proposition for the TromboGene Kit (TRB kit), a genetic test of hereditary thrombophilia developed by Heartgenetics, Genetics Biotechnology, S.A.. First, a literature review about hereditary thrombophilia was performed. A statistical analysis followed involving two cohorts of VTE patients and of recurrent miscarriage (RM) patients, a clinical manifestation also of interest, aiming to investigate the association between the test's genetic panel and those manifestations. The previous steps led to the proposal of alterations to the TRB kit's current panel, whose impact was then assessed in the product's future risk prediction model. This work paved the way for the development of a new solid value proposition for the TRB kit.

Keywords: Hereditary Thrombophilia, Venous Thromboembolism, Recurrent Miscarriage, Genetic Tests, Risk Predicting Model

1. Introduction

Hemostasis comprises a set of vital physiological mechanisms that, in normal conditions, stops bleeding at the site of an injury while simultaneously keeping normal blood flow elsewhere in the circulation [2]. Upon damage to a blood vessel, components present under its internal layer are exposed to the blood, consequently triggering hemostatic processes that lead to the formation of a blood clot, ultimately stopping the bleeding [2, 3]. Both the processes leading to the formation of a blood clot, also known as a thrombus, and the mechanism by which it is dissolved - fibrinolysis - are crucial to maintain hemostasis. When the delicate balance between these mechanisms is disrupted, abnormal bleeding can occur or, on the contrary, the pathological formation of clots in blood vessels - thrombosis [2, 3].

Venous thromboembolism (VTE) is a condition affecting the veins which encompasses both the occurrence of thrombosis and the possibility of a subsequent embolism, the latter implying that the thrombus dislodges from its original site and en-

ters the circulation, potentially blocking a vessel anywhere in the circulatory system. For instance, it may block a vessel in the lungs, causing a pulmonary embolism (PE), or a vessel irrigating the brain, therefore compromising blood circulation and oxygen supply to this organ, an event known as ischemic stroke (IS) [2, 3]. There are a number of risk factors for the development of VTE, which can be acquired or inherited. Acquired risk factors include major surgery, pregnancy, or malignancy, among others. Plus, having an inherent predisposition to forming blood clots - thrombophilia - is also a risk factor for VTE [4]. Thrombophilia can be inherited or acquired, the focus of this work being on hereditary thrombophilia.

VTE is a multifactorial disorder resulting from the complex interaction between genetic and environmental factors. It is estimated that the average annual incidence rate of overall VTE among people of European ancestry varies between 104 and 183 per 100,000 person-years [1]. Incidence rates appear to be higher in women during childbearing years, though they are generally higher in men after

the age of 45 years [1].

Survival after a VTE event depends of the nature of the event itself. For instance, the survival rate after a PE event is much worse when compared with that of deep vein thrombosis (DVT) alone, meaning a thrombus forming in a deep vein of the leg, for example. The chances of survival decrease as time passes, being 65.2% and 41.5% eight years after the first DVT and PE event, respectively [1]. Plus, for almost one-quarter of PE patients, the initial and only clinical presentation of the condition is sudden death [1].

Given the negative impact of VTE in human health, having ways of knowing whether one is prone to developing the disease is certainly valuable. Since genetics plays a role in VTE's etiology, testing for the presence of specific genetic markers that are indicative of an inherent predisposition to forming blood clots - hereditary thrombophilia - can make a difference regarding prophylaxis. Heartgenetics, Genetics Biotechnology, S.A., is a company that has developed a genetic test in the cardiovascular field, namely the TromboGene Kit (TRB kit), which performs a genetic study of hereditary thrombophilia. By genotyping a set of genetic variants, this currently qualitative product provides an estimate of the genetically determined propensity to thrombotic events, addressing as well the condition's impact on pregnancy.

Diagnostic devices such as the TRB kit are a reflection of how genetics has brought a new light to how we view health care. Forecasting the odds of disease based on an individual's genetic profile gives one the chance of incorporating appropriate lifestyle changes much earlier than possible otherwise, thus representing a paradigm change in the management of conditions such as VTE, and consequently supporting the importance of continuously improving this sort of technology. The goal of this dissertation was to develop a new value proposition for the TRB kit, which essentially translated into proposing a number of alterations to the current version of the test, aiming to improve its risk-predicting ability.

2. Hereditary Thrombophilia Background

Thrombophilia can be defined as an abnormality in blood coagulation leading to a state of hypercoagulability that increases the risk of thrombotic events. As already mentioned, this condition can result from genetic factors or acquired changes in the clotting mechanisms, resulting more commonly from the interaction between both [5]. Its genetic basis has been attributed to changes in the amount and/or function of proteins involved in the coagulation process.

The coagulation and fibrinolytic systems are two

separate, yet linked, enzyme cascades whose combined functions aim to regulate the production and breakdown of fibrin, an insoluble protein whose deposition during blood coagulation strengthens and stabilizes the clot. In the coagulation cascade, each enzyme involved is present in the plasma as a precursor molecule which, upon activation, releases its active factor (or coagulation factor). The final product of this cascade is thrombin, which in turn converts soluble fibrinogen into insoluble fibrin, activates other coagulation factors and can, on the contrary, even exert an anticoagulant effect [5, 6].

2.1. Classic Inherited Thrombophilias

There are different main inherited thrombophilias. These can essentially be divided in loss-of-function mutations of genes encoding natural anticoagulant proteins - antithrombin (AT), protein C (PC), and protein S (PS) - and in gain-of-function mutations of genes encoding coagulation factors II (FII) and V (FV).

AT Deficiency AT exerts a major inhibitory effect on thrombin and on other prothrombotic factors such as activated factor X (FXa). Over 250 loss-of-function mutations have been identified in the AT gene, *SERPINC1*, causing AT plasma levels to decrease or compromising its anticoagulant ability [6]. Though extremely rare in the Caucasian population - estimated prevalence between 0.02 and 0.2% - AT deficiency increases the risk of developing VTE more than 50-fold comparatively to non-carriers, thus being considered the most severe inherited thrombophilia [6].

PC Deficiency PC is activated in the presence of thrombin. Activated PC (APC) together with its cofactor, PS, reduces thrombin production through the inactivation of activated factors V (FVa) and VIII (FVIIIa). More than 200 loss-of-function mutations in the PC gene, *PROC*, have been identified as causing this protein's deficiency [6]. PC deficiency is also very rare in the Caucasian population - estimated prevalence around 0.2% - and carriers have a 15-fold increased risk of developing VTE [6].

PS Deficiency Besides acting as a cofactor of APC, PS also functions as a cofactor in the inhibition of FXa. Nearly 200 loss-of-function mutations have been identified in the PS gene, *PROS1*, leading to PS deficiency [6]. PS deficiency has an estimated prevalence between 0.03 and 0.1% in the Caucasian population and carriers have a 10-fold increased risk of VTE, comparatively to non-carriers [6].

Factor V Leiden (FVL) This gain-of-function mutation in the FV gene, *F5*, causes the mutant

molecule - FVL - to become resistant to inactivation by APC, hence continuing to exert its full procoagulant activity. The FVL mutation is considered the most common prothrombotic mutation in the Caucasian population, with a prevalence in heterozygosity of 5% [6]. While heterozygous carriers have a 7-fold increased VTE risk, compared to non-carriers, the same risk is increased by 80-fold in homozygous carriers [5, 6].

Prothrombin G20210A This gain-of-function mutation in the prothrombin gene, *F2*, leads to an increase of around 30% in prothrombin plasma levels, the precursor molecule of thrombin. This mutation is the second most common prothrombotic mutation in the Caucasian population - estimated prevalence in heterozygosity between 2 and 3% [6, 5]. Heterozygosity for this mutation is associated with a 3- to 4-fold increased VTE risk, whereas homozygosity confers a 30-fold increased risk [6].

2.2. Clinical Manifestations

Even though the clinical picture may differ between individuals with heterozygosity or homozygosity for the defects described above, one may say that the main consequence of having these inherited thrombophilias is VTE, often leading to DVT with or without PE [7].

Throughout the years, hereditary thrombophilia has been systematically implicated in other conditions in the literature. Among the most common are events or complications related with cardiovascular disease (CVD), such as IS or myocardial infarction (MI), for example. Plus, also the link between poor pregnancy outcome (PPO) and thrombophilia has been a topic of extensive research. However, since both CVD and PPO are multifactorial conditions, proving an association, let alone causality, is not straightforward, sometimes giving rise to contradictory evidence.

Thrombotic events, whether occurring in veins or arteries - arterial thromboembolism (ATE) - are often addressed in the context of CVD, the latter encompassing a group of diseases affecting both the heart and blood vessels. Mahmoodi *et al.* investigated the role of the five classic inherited thrombophilias in ATE, usually linked to atherosclerosis and ultimately leading to MI or IS. It was concluded that these thrombophilias were associated with a higher risk of ATE in the presence of traditional cardiovascular risk factors, such as hypertension, diabetes mellitus, or hypercholesterolemia [8]. Another author subscribed the previous conclusion in the case of IS, stating that the combination of thrombophilic defects, particularly FVL and prothrombin G20210A, with common cardiovascular risk factors increased the risk of younger patients (before the age of 55 years) developing IS [9]. Nev-

ertheless, thrombophilic defects are not considered the primary cause when investigating the etiology of IS [10, 11].

Pregnancy is a prothrombotic state by itself, meaning that all three factors involved in the etiology of thrombosis - stasis, endothelial damage, and hypercoagulability - are naturally present. This hypercoagulability is thought to be a protection mechanism developed during this period, given the bleeding risks associated with childbirth or miscarriage [12, 13]. The inherent hypercoagulability of pregnancy means that the risk of VTE in women increases 5- to 6-fold compared with age matched controls [14]. Though there appears to be consensus about the fact that hereditary thrombophilia has a compounding effect on the inherent hypercoagulability of pregnancy, hence putting pregnant women at a higher risk of developing VTE, some controversy remains regarding its connection with RM [12, 15, 16, 17].

RM can be defined as the occurrence of two or more consecutive fetal losses before twenty weeks of gestation [13]. RM affects 1 to 5% of couples and has a multifactorial nature [16]. It can be due to chromosomal abnormalities, anatomical, hormonal, genetic, and unexplained causes, the latter representing 30 to 40% of cases [15, 18]. Growing evidence throughout the years has suggested that inherited thrombophilias, besides augmenting the inherent prothrombotic state of pregnancy, compromise adequate fetomaternal circulation, through the formation of microthrombi in the placental vessels, and even compromise the process of placentation in the developing embryo, ultimately leading to a failed pregnancy [13, 19, 18]. Genetic defects linked to RM include classic thrombophilic variants such as FVL and prothrombin G20210A. Moreover, also variants of the methylenetetrahydrofolate reductase (*MTHFR*) gene have been linked to RM [13, 19]. Gao *et al.* published a systematic review and meta-analysis assessing the role of prothrombin G20210A in RM. The main conclusion was that this mutation increased the risk of RM, particularly in European women over the age of 29 years old [20]. Similarly to the previous source, many others, with different sample sizes and methodologies, can be found in the literature featuring results that support a connection between hereditary thrombophilia and RM. However, many others can also be found reaching contrary conclusions, as it is the case in a review by Unterscheider *et al.*, which stated that more recent studies have failed to confirm the otherwise extensively described association between hereditary thrombophilia and PPO in general [19]. The general picture regarding this topic is that some authors support a connection while others do not, and therefore do not recommend screening or thrombo-

prophylaxis aiming to improve pregnancy outcome. An example of that is given by Pritchard *et al.*, who explained that the inability to generalize results discourages anticoagulation among women with hereditary thrombophilia, hoping to improve pregnancy outcome, adding, however, that there are sufficient evidence supporting VTE prevention treatment in the same group [12].

3. Methodology

3.1. Database Search

Besides the classic inherited thrombophilias described, several other genetic variants have been linked to hereditary thrombophilia in the literature. The first stage was to perform an extensive search for studies reporting other variants, using mainly PubMed and Web of Science search engines. Given the difficulty in finding studies focusing explicitly on hereditary thrombophilia, its clinical manifestations became the focus instead. Thus, the majority of the studies selected - essentially genetic association studies in different populations - investigated the genetic basis of VTE (including DVT and PE) and RM. Only studies published from 2010 onward were eligible. As VTE and RM are conditions or events of multifactorial nature, the genetic panels featured in these studies often included variants that had no known association with the coagulation process. Moreover, when reviewing each study, only variants known to be related to blood coagulation were selected, except for *MTHFR* genetic variants, which were also included for having been repeatedly linked to hereditary thrombophilia in the literature.

3.2. Statistical Analysis

A statistical analysis (or, in other words, a genetic association analysis) was performed using a set of internal samples made available by the company. All analyses were performed using R software (www.r-project.org). The samples used were of patients who took Heartgenetics' TRB test and whose medical reasons behind taking it were VTE (DVT, PE, and thrombophlebitis) or RM. Thus, two separate analysis cohorts were created - a cohort of VTE patients and a cohort of RM patients. All samples included were Caucasian. The VTE cohort was composed of 57 female cases and 21 male cases, whereas the RM cohort was composed of 128 female cases. In both analysis, the control samples were retrieved from the 1000 Genomes Project (www.1000genomes.org) - a total of 503 Caucasian samples, 263 of which were female and 240 were male. In the RM analysis, only the 263 Caucasian female samples were used. Overall, the VTE cohort was comprised of a total of 581 samples and the RM cohort of a total of 391 samples.

The genetic variants up for analysis included

eight of the fifteen variants featured in the TRB kit, which are mostly of genes affecting blood coagulation. Seven variants were excluded because either there was no genotypic data on them in the 1000 Genomes Project, resulting in null genotype counts for the controls, or all cases and controls had the wild type genotype, or there were very few samples for which the variant had been genotyped.

As a quality control step, deviations from Hardy-Weinberg equilibrium (HWE) were assessed in the control population. The R function "HWEexact" was used, which performs, for every variant, a two-sided exact test on a contingency table. A second quality control step was to ensure that there was no population stratification, as it can lead to false-positive associations, or even to missing true associations. To do so, a clustering-based method was employed. The *K*-means clustering algorithm was used, which aims to find the cluster centroids that minimize the distance between data points and the nearest centroid [21, 22]. The Bayesian Information Criterion (BIC) was used to infer the number of *K* clusters of the *K*-means algorithm, therefore indicating how many subpopulations existed in both cohorts. This is a criterion for model selection, meaning that, given a set of models, the one with the lowest BIC value should be selected as the best model. Thus, the *K*-means algorithm was run for different *K* values, using the R function "kmeans", and the model whose *K* value had the lowest associated BIC value was selected. The genetic data was subjected to principal component analysis (PCA), using the R function "prcomp", as a pre-processing step. Although PCA is commonly known as a dimensionality reduction technique, in this case, it was not used for feature selection purposes, but only to remove correlations from the data, which can bias the results of the *K*-means algorithm [22].

To assess the association between the genetic panel and VTE and RM, a two-sided Fisher exact test was performed using the R function "fisher.test". This function tests the null hypothesis of independence between the rows and columns of a contingency table featuring the genotype counts, for the cases and controls, of a given variant. The first tests were followed by more association tests in order to determine which genetic model suited best each variant. Five genetic models were tested - full (corresponding to the 2×3 matrix first tested), dominant, recessive, allelic, and additive - and the one with the smallest associated p-value was selected as the appropriate genetic model.

After the association tests, a logistic regression model was built having as explanatory variables the genotypes of the final set of genetic variants:

$$\text{logit}(p) = \ln \frac{p}{1-p} = \beta_0 + \beta_1 x_1 + \dots + \beta_8 x_8 \quad (1)$$

where p is the probability of having the disease (or trait) and $x_1 \dots x_8$ contain information concerning the genotypes of the eight SNPs included in the model - the number of risk alleles, hence in accordance with the additive model. β_0 is the intercept term and $\beta_1 \dots \beta_8$ represent the effects of each marker on the trait.

To train the model and estimate parameters $\beta_0 \dots \beta_8$ for each cohort, a leave-one-out cross-validation method was employed, a special case of k -fold cross-validation in which k equals the total number of samples [23]. The predictions for all the samples, obtained following the leave-one-out cross-validation process, were stored in two separate data frames, together with their respective true classifications (1 for case and 0 for control).

A second model was also taken into consideration in this analysis. Developed internally by the company, the scores model is to be included in a future version of the TRB kit and includes all fifteen genetic variants of the TRB kit's genetic panel. Based on a subject's genotypes for the fifteen variants, the model returns a risk score which represents the overall, genetically determined, susceptibility to thrombotic events. This risk is graphically represented through a risk bar featured in the test's report. Since the scores model is completely specified, thus not requiring any training to fit model parameters, it was readily used to obtain the risk scores of all the samples of both cohorts. These risk scores were then stored in two separate data frames together with their respective true classifications.

To compare the performance of both models in the VTE and RM cohorts, a receiver operating characteristics (ROC) curve was used. ROC curves are a useful way of visualizing a model's performance, x axis representing the false positive rate and the y axis depicting the true positive rate. To build the ROC curves, the R function "roc" was used, which essentially receives the model predictions (scores, later sorted) and corresponding true classifications to build a ROC curve from them. Once the ROC curves were built, the R function "roc.test" was used in order to compare both curves in terms of their AUCs. This function performs a statistical test to assess if the AUCs of the curves being tested are significantly different or not, which is indicated by the p-value returned. A two-sided test was performed and the bootstrap method was employed, meaning that n replicates or permutations are drawn from the original data and, for each replicate, the AUC of both curves is computed and the difference between them is stored. Then, the difference between the original AUCs of the two ROC

curves is divided by the standard deviation of the bootstrap differences and the resulting statistic is compared to the normal distribution.

3.3. Development of a new value proposition

Based on the results of the database search and of the statistical analysis, a number of alterations to the current genetic panel of the TRB kit was suggested. A resulting new set of genetic variants was then proposed as the genetic panel of the future version of the test.

3.3.1 Scores model impact assessment

Since the new genetic panel included the addition of a new variant that had not been genotyped in the case samples of both cohorts, it had to be simulated. As there were genotypic data on the variant available in the 1000 Genomes Project, the genotypes of the control samples were retrieved from its database. The genotypes of the samples retrieved were converted to the number of risk alleles - 0, 1, or 2 - in accordance with the additive genetic model. The genotypes of the case samples of both cohorts were simulated. To do so, a number of wildtype (0), heterozygous (1), and homozygous (2) genotypes were generated and randomly attributed to the case samples, so that the total number of cases was in accordance with the variant's genotype frequencies indicated in the 1000 Genomes Project, similarly to the controls.

The assessment of the impact of the new panel on the final risk scores produced was performed by computing the relative difference (%), for each sample, between the risk score produced by the model featuring the new panel and the risk score resulting from the original panel. As the cases for the new variant were simulated, the procedure involving randomly attributing (0), (1), and (2) to the samples and computing the relative differences (%) was repeated 100 times. Then, for each sample, the mean of the resulting 100 relative differences (%) was computed. The means of the relative differences (%) for the low risk samples, medium risk samples, and high risk samples were represented in separate histograms.

4. Results

Regarding the database search, 27 studies were selected and a total of 55 genetic variants were reported. The variants reported are of genes predominantly encoding proteins related to blood coagulation - anticoagulant proteins, coagulation factors, subunits of fibrinogen, among others. The final list includes, as expected, well-established genetic defects, such as FVL (rs6025) or prothrombin G20210A (rs1799963), with several sources reporting an association between these variants and mostly VTE. On the contrary, sources reporting

variants of genes encoding natural anticoagulants, or leading to their deficiency, were difficult to find given the rarity of these deficiencies, making hardly possible to conduct large clinical studies concerning them [24].

Thirteen *ABO* genetic variants were included in the list of thrombophilic variants. The *ABO* gene encodes proteins related to the ABO blood group system, which classifies human blood as type A, B, AB, or O, depending on the presence or absence of A and B antigens on the surface of red blood cells. Antigen H also belongs to the group of ABO antigens. While the codominant A and B alleles of the *ABO* gene encode glycosyltransferases that convert the common H precursor into A or B antigens, the recessive O allele does not encode a functional enzyme, causing OO carriers to keep the unaltered H structure [6]. The presence of ABH antigenic structures on circulating von Willebrand factor (vWF), a protein that binds to factor VIII (FVIII) and plays an important role in platelet adhesion, influences the molecules half-life in plasma, hence explaining the role of the ABO blood group system in hemostasis. This differentiated expression causes the vWF plasma half-life to be 10 hours for group O, while rising to 25 hours for non-O groups. As a result, individuals of non-O blood groups have vWF levels, and consequently FVIII levels, approximately 25% higher comparatively to O blood group individuals [6]. Since increased vWF and FVIII levels are considered risk factors for the development of VTE, non-O blood type has been linked to the development of this condition. In fact, non-O blood type has been shown to increase the risk of VTE by approximately 2-fold. [6, 25]. The list of variants of the *ABO* gene includes variants related to blood type, as well as others whose association with VTE is independent of blood type, therefore suggesting that there may be other mechanisms underlying the connection between *ABO* genetic variants and VTE besides blood type.

The MTHFR enzyme is encoded by the *MTHFR* gene, which has two main polymorphisms, namely MTHFR rs1801133 and MTHFR rs1801131, both reported following the database search. These variants have repeatedly been associated in the literature with hyperhomocysteinemia, a condition characterized by elevated plasma levels of homocysteine. Hyperhomocysteinemia occurs because a reduction in the MTHFR enzyme activity disrupts the balance of the conversion of homocysteine to methionine [26]. This condition may have different consequences, such as an increase in oxidative stress or an impact in thrombus formation. The latter occurs primarily because elevated levels of homocysteine in the blood may cause vascular endothelial injury, one of the three factors involved in the

etiology of thrombosis. In other words, the tendency of homocysteine to promote cellular toxicity through oxidative stress contributes to the onset of an endothelial lesion [27]. This explains the relevance of *MTHFR* genetics variants in hereditary thrombophilia, despite the *MTHFR* gene not having a specific role in coagulation *per se*.

Regarding the statistical analysis, the results of the HWE tests revealed that, with the exception of one variant, all the remaining variants were in HWE. Moreover, the results of the population stratification analysis confirmed that there was only one population in both cohorts, as the lowest BIC value obtained in both cases corresponded to $K = 1$.

5. Conclusions

Hereditary thrombophilia is a condition with potentially serious health consequences, therefore being of relevance to study its genetic basis in order to come up with tools, such as Heartgenetics' TRB kit, that help estimate an individual's genetic predisposition to conditions such as VTE.

The aim of this dissertation was essentially to develop a new value proposition for the TRB kit, a genetic test that provides an estimate of the genetic predisposition to thrombotic events, based on genotypic data of set of genetic markers.

The first step was to attempt to list all the genetic variants involved in this condition. As a result, 55 genetic variants were reported, the majority of them associated VTE and RM. Despite the large number of variants reported, it is worth noting that variants of other potentially interesting genes with a role in coagulation, such as the (*vWF*) gene, for example, are not part of the list due to the lack of relevant sources found.

The second step was to perform a statistical analysis, or a genetic association analysis, involving eight genetic variants of the test's current panel, in order to investigate the association between that set of variants and the two traits of interest, VTE and RM.

When testing for deviations from HWE in the control group, it was observed that one variant - FGB rs1800790 - appeared not to be in HWE. Since it was demonstrated that there was no population stratification in both cohorts, the low p-value obtained for FGB rs1800790 may possibly be due to low levels of genotyping errors, since no genetic data set is completely free from errors [28]. Thus, the deviation from HWE observed can be interpreted as an artifact.

Based on the results of the database search and of the statistical analysis, as well as taking into account the input from the company's scientific team, the third step was to suggest alterations to the TRB kit's current genetic panel, which included the ad-

dition of a new variant.

The fourth and final step was to assess the impact of the panel being proposed on the TRB kit's risk prediction model. When analyzing the mean relative differences between the final risk scores produced by both panels, it was observed that the alterations proposed did not impact the risk scores produced by the model severely, as expected, given the small magnitude of the alterations proposed.

This work has some limitations, one of them being the small sample size of both cohorts, which certainly had an impact on the overall results of the statistical analysis. Nevertheless, the results of this work represent interesting contributions to the continuous development a new solid value proposition for the TRB kit.

References

- [1] John A. Heit *et al.* The epidemiology of venous thromboembolism. *J Thromb Thrombolysis*, 41, 2016.
- [2] Andrew J. Gale. Current understanding of hemostasis. *Toxicol Pathol.*, 39(1), 2011.
- [3] Ewelina M. Golebiewska *et al.* Platelet secretion: From haemostasis to wound healing and beyond. *Blood Reviews*, 29, 2015.
- [4] Fatemeh Moheimani *et al.* Venous thromboembolism: Classification, risk factors, diagnosis, and management. *ISRN Hematology*, 2011.
- [5] S. Khan *et al.* Hereditary thrombophilia. *Thromb. J.*, 4:1–17, 2006.
- [6] P. M. Mannucci *et al.* Classic thrombophilic gene variants. *Thromb. Haemost.*, 2015.
- [7] V. De Stefano *et al.* Inherited thrombophilia: Pathogenesis, clinical syndromes, and management. *J Am Soc Hematol.*, 87(9), 1996.
- [8] B. K. Mahmoodi *et al.* Interaction of hereditary thrombophilia and traditional cardiovascular risk factors on the risk of arterial thromboembolism. *Circ. Cardiovasc. Genet.*, 9(1):79–85, 2016.
- [9] Dragoni F. Inherited and acquired thrombophilic factors in young patients with acute coronary syndrome or ischemic stroke. *Austin J Clin Cardiol.*, 2(2), 2015.
- [10] K. W. P. Ng *et al.* Role of investigating thrombophilic disorders in young stroke. *Stroke Res. Treat.*, 2011.
- [11] S. H. Pahus *et al.* Thrombophilia testing in young patients with ischemic stroke. *Thromb. Res.*, 2016.
- [12] Ashley M. Pritchard *et al.* Hereditary thrombophilia and recurrent pregnancy loss. *Clinical Obstetrics and Gynecology*, 59(3), 2016.
- [13] M. Chatzidimitriou *et al.* Thrombophilic gene polymorphisms and recurrent pregnancy loss in greek women. *Int. J. Lab. Hematol.*, 2017.
- [14] F. N. Croles *et al.* Pregnancy, thrombophilia, and the risk of a first venous thrombosis: systematic review and bayesian meta-analysis. *BMJ*, 2017.
- [15] C. Karadag *et al.* Obstetric outcomes of recurrent pregnancy loss patients diagnosed with inherited thrombophilia. *Ir J Med Sci*, 186, 2017.
- [16] Ana-Luisa Stefanski *et al.* Maternal thrombophilia and recurrent miscarriage is there evidence that heparin is indicated as prophylaxis against recurrence? *GebFra Science*, 78, 2018.
- [17] N. Roozbeh *et al.* Potential role of factor v leiden mutation in adverse pregnancy outcomes: An updated systematic review. *Biomed Res Ther*, 4(12):18321846, 2017.
- [18] Nadja Bogdanova *et al.* Hereditary thrombophilic risk factors for recurrent pregnancy loss. *J Community Genet*, 1, 2010.
- [19] J. Unterscheider *et al.* The role of thrombophilia testing in women with adverse pregnancy outcomes. *Obstet. Gynaecol.*, 2017.
- [20] Hui Gao *et al.* Prothrombin g20210a mutation is associated with recurrent pregnancy loss: A systematic review and meta-analysis update. *Thrombosis Research*, 135, 2015.
- [21] Chih Lee *et al.* Pca-based population structure inference with generic clustering algorithms. *BMC Bioinformatics*, 10, 2009.
- [22] Adam Coates *et al.* *Neural Networks: Tricks of the Trade*, chapter Learning Feature Representations with K-means. Springer, 2012.
- [23] Rosa J. Meijer *et al.* Efficient approximate k-fold and leave-one-out cross-validation for ridge regression. *Biometrical Journal*, 55(2):141–155, 2013.
- [24] R. Gindele *et al.* Clinical and laboratory characteristics of antithrombin deficiencies: A large cohort study from a single diagnostic center. *Thromb. Res.*, 160:119128, 2017.
- [25] M. F. Francesco Dentali *et al.* Non-o blood type is the commonest genetic risk factor for vte: Results from a meta-analysis of the literature. *Thromb. Haemost.*, 2012.

- [26] R. Stefanov *et al.* Application of single-nucleotide polymorphism-related risk estimates in identification of increased genetic susceptibility to cardiovascular diseases: A literature review. *Frontiers in Public Health*, 5, 2018.
- [27] Gilberto Vizcano *et al.* Homocistena: bases genéticas y sus implicaciones cardiovasculares y cognitivas como factor de riesgo. *Invest Clin*, 58(4), 2017.
- [28] Phillip A. Morin *et al.* Significant deviations from hardyweinberg equilibrium caused by low levels of microsatellite genotyping errors. *Molecular Ecology Resources*, 9, 2009.