What is currently known about the genetics of venous thromboembolism at the dawn of next generation sequencing technologies

David-Alexandre Trégouët¹,², Pierre-Emmanuel Morange³,⁴

¹ Sorbonne Universités, UPMC Univ. Paris 06, Institut National pour la Santé et la Recherche Médicale (INSERM), Unité Mixte de Recherche en Santé (UMR_S) 1166, Department of Genomics & Pathophysiology of Cardiovascular Diseases, Paris, France.
² ICAN Institute for Cardiometabolism and Nutrition, Paris, France.
³ Laboratory of Haematology, La Timone Hospital, Marseille, France.
⁴ INSERM UMR_S 1062, Nutrition Obesity and Risk of Thrombosis, Aix-Marseille University, Marseille, France.

The genetic contribution to the risk of venous thromboembolism (VTE) has been established since 1960s with the observation of an association of antithrombin deficiency and ABO blood type with clinical risk (Egeberg, 1965; Jick et al, 1969). Since then, improved molecular technology has expanded the scientific community’s ability to better identify and characterize the loci and their functional impact on VTE risk. In this review, we provide an overview of what is currently known about the genetics of VTE, and summarize in a table (Table 1) the main genetic associations that have been robustly validated so far, either experimentally or using epidemiological data from several independent studies.

**Genes with well characterized functions and with identified causal coding variations**

As could be expected from population genomics, genetic variants with strong impact on VTE risk are uncommon and are often private mutations (i.e present in less than 0.1% of the population). There are dozens of loss-of-function variants in anticoagulant genes coding for the three natural coagulation inhibitors: antithrombin (SERPINC1 gene), protein C (PROC) and protein S (PROS1) (Egeberg, 1965; Griffin et al, 1981; Schwarz et al, 1984). The risk of developing the disease in carriers of such mutations is ~10 fold increased. Interestingly, two of these genes (PROS1 and SERPINC1) have been found to harbour less uncommon polymorphisms that also modulate the susceptibility to VTE. As far PROS1 is concerned, a meta-analysis of 4 French case-control studies for VTE demonstrated that the rs121918472, also referred to as the Protein S Heerlen mutation, a non synonymous Serine to Proline (Ser501Pro) substitution of frequency <1% in the general population which was previously considered has a benign polymorphism, was associated with ~6 fold increased risk of VTE (Suchon et al, 2017). The rs121918474 (Lys196Glu) is another PROS1 coding mutation associated with increased risk of VTE (OR ~5) by lowering protein S anticoagulant activity but this mutation is specific to the Japanese population, approximately 1/12,000 individuals being homozygous for the 196Glu allele (Liu et al, 2013; Kimura et al, 2006). Specifically, Banno et al (Banno et al, 2015) generated PS-Lys196Glu knock-in mice and analyzed phenotypes comparing homozygous vs heterozygous PS-deficient mice. Using 2 animal models, they showed that the murine PS-Lys196Glu mutation reduces its APC anticoagulant cofactor activity in plasma and that PS-Lys196Glu mice and heterozygous PS deficiency are more vulnerable to VTE than wild-type mice, proving pathogenic causality for the Lys196Glu mutation. In a recent meta-analysis of 12 case-control studies for VTE conducted on behalf of the INVENT consortium (Germain et al, 2015) and
totalling ~7,500 cases and ~52,000 controls, the SERPINC1 rs2227624, a nonsynonymous Val30Glu variation (also called the antithrombin Dublin mutation) (Navarro-Fernández et al, 2016) with frequency of ~0.4% in whites, was associated with a ~2 fold increased risk for the disease. A frequent (~0.10) SERPINC1 polymorphism, the intronic rs2227589, has also been proposed as a susceptibility polymorphism for VT (Bezemer et al, 2008). However this association was not robustly confirmed in subsequent studies aimed at replicating this association (Trégouët et al, 2009; Austin et al, 2011) nor showed up as a significant signal in the INVENT project (Germain et al, 2015; Trégouët et al, 2016). Nevertheless, the rs2227589 has been shown to be functional and associated with mild reduced anticoagulant activity (Antón et al, 2009). Whether it has a true effect on VT risk per se or in combination of additional factor(s) remained to be clarified.

Conversely, gain-of-functions in procoagulant genes also contribute to VTE risk, but with less impactful disease risk. The most well known gain-of-function variation associated with VTE is probably the FV Leiden mutation, also referred to as F5 rs6025, an Arginine to Glutamine substitution at amino acid 506 (R506Q) of the coagulation Factor V protein. Factor V promotes conversion of prothrombin to thrombin as well as degradation of activated factor VIII (Segers & Castoldi, 2009). The Q506 isoform, with an allele frequency of ~3% in the Caucasian population, leads to resistance to activated protein C (Bertina et al, 1994) and is associated with an increased risk of ~3 (Morange & Tregouet, 2011). Of note, F5 also harbors a loss of functin coding variation which, in turn, protects against the risk of VTE. This is the rs4524 A/G, a Lysine to Arginine substitution at amino acid 858(K858R), that leads to APC resistance. The G allele, with frequency ~0.25, is associated with a protective odds ratio for VTE of ~0.83 (Smith et al, 2007). Its effect is independent of that of the rs6025 with which it is in complete negative linkage disequilibrium (LD).

KNG1 encodes high molecular weight fibrinogen (HK) that participates to blood coagulation by positioning prekallikrein and FXI near factor XII (Sainz et al, 2007). The non-synonymous rs7100446 T/C variant is an Isoleucine to Threonine substitution at amino acid 581 (Ile581Thr) whose minor allele, with frequency of ~0.40, is associated with an increased VTE risk of ~1.20 (Morange et al, 2011a). The impact of this variant on the disease is likely mediated by its influence on decreased activated partial thromboplastin time (aPTT) (Houlihan et al, 2010).

PROCR encodes endothelial protein C receptor (ePCR) whose elevated soluble levels are associated with increased VTE risk (Uitte de Willige et al, 2004). ePCR enhances the activation of protein C (PROC), one of the main inactivators of Factors Va and Villia under the activation by the thrombin-thrombomodulin complex (Rezaie, 2010). Soluble ECPR levels are genetically controlled at 75% by the rs867186 A/G, a Serine to Glycine substitution at amino acid 219 (Ser219Gly), whose Gly219 allele renders the receptor more sensitive to cleavage through a splicing mechanism (Saposnik et al, 2008). The Gly219 isoform has a frequency of ~0.10 in Caucasians and is associated with a 1.22 fold increased VTE risk (Dennis et al, 2012). A second polymorphism at the PROCR locus, independent of the rs867186, is also modulating the risk of VTE. This could be the rs6088735 C/T or any other polymorphism in strong linkage disequilibrium with it. The rs6088735 T allele, with frequency ~0.23, is associated with a 1.11 fold increased risk (Germain et al, 2015). The rs6088735 maps to the EDEM2 gene, nearby PROCR, and is associated with plasma levels of protein C (Tang et al, 2010). Further deep genetic investigations are required to better characterize the genetic burden at this locus with respect to VTE risk.
F9 encodes coagulation factor IX which, upon activation, contributes with its cofactor, FVIII, to convert Factor X into Factor Xa, the latter participating to cleave prothrombin into thrombin. Only one genetic variation at the F9 locus has been reported to associate with VTE risk. This is a private gain-of-function mutation, R338L (also called the FIX Padua mutation), characterized by a markedly enhanced factor IX activity (Simioni et al, 2009) and causing familial form of VTE (thrombophilia).

Secreted from activated endothelial cells, von Willebrand factor (vWF) is a key endothelial player participating to both platelet aggregation and regulation of coagulation FVIII levels in the blood (Vischer, 2006). Increased plasma levels of vWF are associated with increased risk of VTE (Lazzari et al, 2012), making any genetic variation controlling the regulation of plasma vWF levels an obvious candidate for VTE. The latest genome-wide association study for vWF levels has revealed 8 loci associated with vWF levels (Smith et al, 2010). Three of them, VWF, ABO and STXBPS, are now considered as well-established susceptibility loci for VTE and are also contributing to FVIII plasma levels. The VWF gene encoding the vWF protein harbours one non-synonymous variation, rs1063856 T/C, associated with both vWF levels (Smith et al, 2010) and VTE risk. This is a Threonine-to-Alanine substitution at amino acid 789 (Thr789Ala) whose minor allele has a frequency of ~0.40 and is associated with increased vWF levels and risk of VTE (OR ~1.20) (Smith et al, 2011b).

The STXBPS gene encoding the Syntaxin Binding Protein 5 harbours one non-synonymous variation, rs1039084 A/G. This is a Asparagin-to-Serine substitution at amino acid 436 (Asn436Ser) whose uncommon form (frequency ~0.45) is associated with decreased vWF levels and with a protective effect against VTE (OR ~0.80) (Smith et al, 2011b). STXBPS associated mechanisms involve, not exclusively, endothelial exocytosis of vWF and platelet secretion (Zhu et al, 2014).

The genetic locus impacting the most on the plasma variability of vWF is the ABO gene, with ABO blood groups explaining about 25% of vWF plasma variability (Morange et al, 2005). Compared to ABO O blood group, A1 and B blood groups are associated with increased vWF levels (Morange et al, 2005) and also with a ~1.5 fold increased risk for VTE (Wu et al, 2008; Trégouët et al, 2009). Conversely, ABO A2 blood group does not appear to modulate either vWF levels nor VTE risk. ABO A1 blood group can be tagged by the rs579459, B group by the rs8176749 and A2 by the rs8176704. Whether other polymorphisms at the ABO locus contribute to VTE risk remains an open question.

Genes with well characterized functions and with causal variations in untranslated regions

F2 encodes thrombin the major enzyme of the coagulation cascade. The F2 G20210A, rs1799963 G/A, variation in the 3’UTR region of the gene is uncommon (frequency ~2%) in Caucasians but associated with a rather strong risk of VT, with odds ratio of ~2.3. The rare A allele is associated with elevated plasma prothrombin (FII) levels induced by a more efficient translational machinery of the mRNA harbouring the A allele (Gehring et al, 2001). It is also associated with thrombin generation as is the intronic F2 rs3136516 G/A, in complete negative linkage disequilibrium with the rs1799963 (Rocanin-Arjo et al, 2014). In the INVENT study, the rs3136516 A allele, with frequency ~0.40, demonstrated strong statistical evidence (p = 5.65 × 10^-6) for association with VTE, after adjusting for the effect of the rs1799963 polymorphism. The OR for VT associated with rs3136516-A allele is ~1.12. While the genetic contribution of F2 gene to VT susceptibility is mainly known through the impact attributable to the F2 G20210A variant (sometimes referred to as F2 Leiden mutation), recent studies revealed that the F2 gene could also harbour private mutations causing resistance to antithrombin
and leading to inherited thrombophilia. So far, three non synonymous mutations affecting the same codon have been reported to cause inherited thrombophilia, R596L (Miyawaki et al, 2012), R596Q (Djordjevic et al, 2013) and R596W (Bulato et al, 2016). The first one detected in a Japanese family is referred to as the Yukuashi mutation, the second as the Belgrade mutation following its identification in Serbian families and the third one as the Padua 2 mutation, due to its discovery in Italian families.

$\textit{FGG}$ encodes the fibrinogen $\gamma$ chain, one of the three polypeptides composing the fibrinogen molecule. Under the catalytic action of thrombin, fibrinogen is converted into fibrin, the main component of the thrombus (i.e. the fibrin clot). The rs2066865 C/T polymorphism located in the $\textit{FGG}$ 3UTR region is associated with VTE risk (OR ~1.3 for the rs2066865-T allele) through its decreasing effect on $\gamma \Lambda/\gamma' \text{fibrinogen levels}$ (Uitte de Willige et al, 2005). The alternative $\gamma'$ chain is the result of a polyadenylation mechanism in intron 9 of the $\textit{FGG}$ gene and reduced $\gamma'$fibrinogen /total fibrinogen ratio has been observed in VTE patients compared to healthy individuals (Uitte de Willige et al, 2005). In the INVENT study, a second strong association signal, independent of the rs2066865 effect, was observed, also in the 3UTR region of the $\textit{FGG}$ gene. The rs2227421 (or the rs1044291 in complete association with it) achieved strong statistical association ($p = 5.20 \times 10^{-6}$) after conditioning on the rs2066865 effect. Whether such polymorphism also impact VTE risk by modulating $\gamma'$ chain levels or another mechanism remains to be elucidated. Of note, there has been strong interest in the association of the non synonymous $\textit{FGA}$ Thr331Ala, rs6050, with VTE (Carter et al, 2000; Ko et al, 2006). The $\textit{FGA}$ gene belongs to a cluster of tightly linked genes, $\textit{FGA}$/$\textit{FGB}$/$\textit{FGG}$ on chromosome 4q31.3 and the $\textit{FGA}$ rs6050 is in complete association with the $\textit{FGG}$ rs2066865. As a consequence, the association of the former with VTE completely disappears after adjusting for the effect of the latter.

$\textit{THBD}$ encodes thrombomodulin (TM) that has a dual role in the coagulation cascade. By acting as a cofactor for thrombin, TM drastically enhances thrombin's ability to activate protein C, leading to decreased coagulation. Conversely, complexed with thrombin, TM could also promote coagulation by inhibiting fibrinolysis through activation of the thrombin activated fibrinolysis inhibitor (TAFI) molecule. One polymorphism, rs16984852, in the 5'UTR region of the $\textit{THBD}$ gene has been found to associate with a 3 fold increased risk of VTE. This polymorphism has a frequency of ~1% in the Chinese population where it was discovered (Tang et al, 2013). It has never been observed in non Asian populations. Several common $\textit{THBD}$ polymorphisms with minor allele frequency ~0.20 and in strong LD with each other have recently been proposed to associate with a 2.30 increased risk of VTE in African Americans (Hernandez et al, 2016). However, such association did not replicate in a larger African Americans population of the ARIC study (Folsom et al, 2017). More work is needed to assess whether common polymorphisms at this locus really associate with VTE.

Genes with well characterized functions and with causal variations to be identified

$\textit{F8}$ encodes the coagulation factor VIII that, upon activation by thrombin, interacts with factor IX to activate factor X which, in turn, boost thrombin activation even more. Elevated plasma FVIII levels are well established risk factor for VTE (Jenkins et al, 2012). So far, only one genetic variation in the $\textit{F8}$ gene has been reported to associate with the disease (Hinds et al, 2016). This is the rs114209171, upstream the $\textit{F8}$ gene, and located within an intronic region of the nearby $\textit{FUNDC2}$ gene. The frequency of the minor allele is ~0.20, this allele being protective for VT with allelic OR ~0.90.
According to public database, the rs114209171 is not in strong LD with any other polymorphism at its locus. Further investigations are needed to assess whether this polymorphism is truly functional and by which it affects disease risk.

F11 encodes the coagulation Factor XI that, upon activation by Factor XIIa, activates Factor IX. Elevated plasma FXI levels are risk factor for VTE (Key, 2014) and a recent large scale association study revealed that plasma FXI levels are under the genetic control of 3 loci, F11, KNG1 and GCKR (Sennblad et al, 2017). The association at the KNG1 locus was mainly due to the non synonymous rs7100446 variant that is also associated with the disease as already discussed above. The pattern of association at the F11 locus is slightly more complex with two polymorphisms, rs4253417 and rs425321, additively contributing to influence FXI levels (Sennblad et al, 2017). These two polymorphisms are in strong LD with several other F11 polymorphisms including the often studied rs2289252 and rs2036914. These two latter polymorphisms, with minor allele frequency ~0.40, have been found associated with VTE in several studies, with allelic ORs of~1.40 (Li et al, 2009; Germain et al, 2011, 2015). There is a need to better investigate the haplotype structure of the F11 locus in order to better characterize the functional units responsible for the genetic association with VTE risk. Of note, no GCKR polymorphism have been reported so far to associate with VTE.

Gene with yet uncharacterized functions and with identified causal coding variation

SLC44A2 encoding the choline transporter-like protein-2 (CTL-2) was recently found to harbour a non synonymous variation, rs2288904 G/A, associated with VTE (Germain et al, 2015). This is one of the two main readouts of the INVENT consortium meta-analysis. This variation is an Arginine-to- Glutamine substitution at amino acid (Arg154Gln) whose minor allele, of frequency ~0.24, was found protective against VTE (OR ~0.78). Even though no clear mechanism has been proposed to explain the observed association with the disease, CTL-2 has been described as a new binding partner for VWF to promote neutrophil activation (Bayat et al, 2015). In addition, the at-risk Arg154 isoform codes the human neutrophil alloantigen-3a that is the target of antibodies circulating in the blood to provoke transfusion related acute lung injury (TRALI), a life-threatening complication of blood transfusion. A pathophysiological mechanism involving leucocyte activation and aggregation (Greinacher et al, 2010) is highly suspected.

Genes with yet uncharacterized functions and with causal variations to be identified

TSPAN15 is the second main redouts of the INVENT consortium project (Germain et al, 2015). The intronic rs78707713 T/C at this completely novel gene was found associated with VTE, the minor C allele with frequency ~0.10 being associated with a protective OR for the disease of ~0.76. This polymorphism was observed to associate with TSPAN15 expression in different cell types, including endothelial cells, but so far no clear evidence suggest that this polymorphism could be the functional one. The biological link between TSPAN15 and VTE is still obscure. Of note, TSPAN15 has been shown to partner with ADAM10 to modulate the N-cadherin adhesion molecule and to control vascular development (Alabi et al, 2016) as well as to act as a negative regulator of the Notch signaling pathway (Matthews et al, 2017), a key pathway controlling angiogenesis (Benedito & Hellström,
In addition, TSPAN15 contributes to the cleavage of platelets amyloid precursor protein (Matthews et al., 2017) whose regulation is proposed as a new player in the pathophysiology of VTE (Canobbio et al., 2017).

TFPI, for Tissue Factor (TF) Pathway Inhibitor, is another regulator of the blood coagulation cascade that participates to limit clot formation through different mechanisms including inhibition of FXa, of complexed FVila-TF and of prothrombin-to-thrombin conversion (Camire, 2016). Low plasma levels of TFPI have been associated with increased risk of VTE (Zakai et al., 2010) suggesting that any genetic variations modulating TFPI plasma levels could be a good candidate for VTE risk. Several polymorphisms have been proposed to modulate TFPI plasma levels, but only one has been found associated both with TFPI levels and VTE risk, with the allele decreasing TFPI levels being associated with increased disease risk. This is the rs62187992 G/A (or any other polymorphism in strong LD with it) mapping to the ERBB4 locus on chromosome 2q34 in a linkage peak region for TFPI levels (Dennis et al., 2016). The rs62187992- A allele associated with increased TFPI levels was associated with a significant ($p = 0.03$) decreased risk of 0.90 in the INVENT consortium. The regulatory mechanisms linking the ERBB4 locus to TFPI remain unknown.

**Additional candidate susceptibility genes for VTE to be definitively validated.**

Several other polymorphisms/genes have been proposed as risk factors for VTE but these lack compelling evidence, including robust statistical replication in independent studies, to be definitively claimed as susceptibility variants/genes for the disease. These include the rs2232710 of SERPINA10 gene encoding the Protein Z-Dependent Protease Inhibitor of coagulation factors Xa and Xla (Folsom et al., 2007; Gorski et al., 2016), the SERPINE1 rs2227631 influencing circulating plasminogen activator inhibitor-1 (PAI-1) levels (Huang et al., 2012), the C4BPB/C4BPA rs3813948 affecting protein S associated phenotypes (Buil et al., 2010), the rs7080536 of the HABP2 gene coding for Factor VII activating protease (Ahmad-Nejad et al., 2012), the HIVEP1 rs169713 with yet no associated intermediate phenotype (Morange et al., 2010), the F13A1 rs5985 (Val34Leu) (Wells et al., 2006) influencing Factor XIII activity (Kohler et al., 1998; Wartiovaara et al., 2000), the APOH rs8178847 associated with thrombin generation (Tang et al., 2014), genetic variations at the STAB2, STX2 and TC2N loci with suggestive impact on VTE risk (Morange et al., 2011b; Germain et al., 2011; Smith et al., 2011b; Desch et al., 2016) via their influence on vWF levels (Smith et al., 2010) and the rs4602861 (Germain et al., 2015; Klarin et al., 2017) at the ZFPM2 locus whose suspected association with VTE could involve platelet count (Gieger et al., 2011) and vascular endothelial growth factor (Debette et al., 2011). Some of these associations, if true, could be restricted to some specific ethnic populations. For example, the effect of the APOH rs8178847 variant has been reported so far only in Chinese populations (Tang et al., 2014) and not in Caucasians, despite similar allele frequencies in both populations. In a similar way, genetic variations at the LEMD3, LY86 and LOC100130298 have recently been proposed as risk factors for VTE in African-Americans (Heit et al., 2017) and their impact in independent studies and in other populations warrants further investigations. Some genetic associations could also be restricted to populations of individuals defined by specific clinical, biological, (epi-)genetic characteristics that are yet to be determined. For instance, an epigenetic mechanism related to parent-of-origin transmission could be proposed, and deserves further works,
to explain the non consistent association of the F13A1 rs5985 variant with VTE (Infante-Rivard & Weinberg, 2005).

This review mainly focuses on genes associated with primary VTE in adults and did not mention genetic factors contributing to other forms of venous thrombosis, including cerebral vein thrombosis, splanchnic vein thrombosis and pediatric VTE. For example, a recent genome-wide association study performed in pediatric VTE cases identified the B3GAT2 locus (Rühle et al, 2017) as a novel susceptibility locus for pediatric forms of VTE. Investigating this locus further in adults warrants efforts.

**Future research avenues**

As of July 2017, the largest genetic association study aimed at identifying genetic factors for VTE assembled ~7,500 cases and more than 50,000 controls (Germain et al, 2015). It led to the conclusion that there are 8 loci with common polymorphisms (i.e with allele frequency > ~0.02 in the general population) moderately or strongly associated with VTE. This does not preclude at all that there does not exist addition undiscovered loci genetically associated with VTE. A follow-up study of the main INVENT findings (Trégouët et al, 2016) suggested that one shall not expect to identify additional common VTE polymorphisms with allelic OR greater than 1.2. However, common polymorphisms with weaker genetic effects or rarer polymorphisms, that are either responsible for inherited form of VTE (Martinelli et al, 2014) or simply susceptibility alleles, could still contribute to VTE susceptibility. Larger sample size studies would then be generally required to identify common polymorphisms with weak effects, the interest of such approach being nicely exemplified by genetic association studies for arterial thrombosis (Nikpay et al, 2015; Myocardial Infarction Genetics and CARDioGRAM Exome Consortium Investigators et al, 2016). Conversely, rare genetic defects with much stronger effect can be detected even in moderate sample size studies enriched in genetic cases. As an example, the rare MYH8 rs111567318 (Glu1838Ala) variation (minor allele frequency ~0.005) was recently identified as a risk factor for VTE using only a sample of 107 cases with multiple events. Experimental works demonstrated that the MYH8 protein binds factors Va and Xa to active prothrombin and influence thrombin generation. (Deguchi et al, 2016)

Another appealing strategy to identify some additional weak genetic contributors to the disease consists in addressing the genetic architecture of an hypothesized quantitative risk factor and how it relates to VTE risk. As an example, Lindström et al (Lindström et al, 2017) demonstrated that a genetic risk score derived from 95 BMI-associated SNPs was associated with ~1.60 fold increased risk of VTE risk. Of note, the BMI-associated SNP showing individually the stronger influence on VTE was the FTO rs1558902 but its impact on the disease was rather weak (allelic OR of 1.07) (Lindström et al, 2017).Beyond the identification of SNPs contributing to VTE susceptibility, such strategy conducted in the context of a Mendelian Randomization approach (Davey Smith & Hemani, 2014) allows to determine whether a risk factor could be causal or just a disease marker.

However, the source of the inter-individual susceptibility to VTE does not lie exclusively in genetics. Epigenetics mechanisms are now widely believed to participate to the susceptibility to complex
diseases, such as VTE. As mentioned above, the \textit{FIIIA1} rs5985 variant is prone to parental transmission distortion (Infante-Rivard & Weinberg, 2005). The expression of several genes encoding hemostatic proteins such as Factor VII (Friso \textit{et al}, 2012), FVIII (El-Maarri \textit{et al}, 2007), and tissue-type plasminogen activator (t-PA) (Zwingerman \textit{et al}, 2015) have been shown to be subject to DNA methylation mechanisms. And BMI, a causal risk factor for VTE, is influenced by epigenetic marks (Dick \textit{et al}, 2014). All these observations strongly suggest that properly designed studies investigated DNA methylation marks associated with VTE deserve to be conducted. In a similar way, the expression of several VTE associated genes are predicted to be regulated by microRNAs (Marchand \textit{et al}, 2012; Vossen \textit{et al}, 2017; Sennblad \textit{et al}, 2017; Wang \textit{et al}, 2016). Several candidate microRNAs have been proposed as novel biomarkers for VTE as part of a case-control study on plasma samples (Starikova \textit{et al}, 2015). Similar studies should be conducted to replicate these findings and to discover other microRNAs associated with the disease.

Another proof, if needed, that genetics is not the panacea and that circulating biomarkers can be of clinical utility independently of any genetic regulation is illustrated by the D-dimers case. D-dimers are now widely used as a tool to diagnose VTE whereas the main genetic regulator of its plasma levels, the \textit{FIIIA1} rs12029080 polymorphism (Smith \textit{et al}, 2011a) is not associated with VTE risk (Smith \textit{et al}, 2012). This should encourage agnostic searches for circulating biomarkers, such as microRNA, metabolites, proteins, associated with VTE by ignoring, in the first instance, any genetic context. A first pilot study for agnostic plasma proteomic investigation with consistent replication in an independent study led to the identification of PDGF-B as a novel circulating markers for VTE (Bruzelius \textit{et al}, 2016). Other similar studies are ongoing and should lead to novel findings.

\textbf{Clinical Perspectives}

The clinical perspective of genetic testing for VTE has been nicely and recently reviewed (Middeldorp, 2016). Briefly, in general, screening tests allow to prevent the occurrence of pathologies or their complications in individuals at risk. They must be carried out in principle only if their positivity results in a change in preventive or curative therapeutic attitude. In the context of VTE, genetic testing commonly includes levels of antithrombin, protein C, or protein S to identify a deficiency, as well as factor V Leiden and prothrombin G20210A (this is usually called the thrombophilia screening). This list has not been modified since the discovery of the prothrombin G20210A mutation in 1996. Despite the publication of several recommendations, there is no consensus on which patients should benefit of this screening (Stevens \textit{et al}, 2016). In patients with a personal history of VTE, the positivity of this screening does not influence the dosage or duration of anticoagulant therapy since it only modestly increases the risk of recurrence. Genetic counseling in families with VTE is also still debated. Indeed, the identification of the defect in the relatives of a proband with a VTE history would theoretically identify those at high risk to offer them specific care in high risk situations. However, even if they carry the same defect, relatives often do not present the same clinical manifestations as the proband. In addition, in families with thrombophilia, the risk of VT is increased among relatives without the defect which is probably due to the existence of other genetic risk factors for VTE still unidentified. Finally, a family history of VT is itself a risk factor after adjustment for thrombophilia (Bezemer \textit{et al}, 2009). Thus, given all those considerations, it is difficult to
dichotomize carriers and non carriers in terms of their risk for a first episode of VTE with the classical thrombophilia screening.

To improve performance of thrombophilia screening in clinical practice, new approaches based on common polymorphisms recently selected from GWAS have been developed. The predictive value of 31 polymorphisms on the risk of recurrence of VTE has been tested (van Hylckama Vlieg et al, 2014). The use of a genetic risk score based on the five polymorphisms located at the ABO, F11, F2, F5 and FGG genes have been classified according to their risk: after a six-year follow-up, the cumulative incidence of recurrence was high when the score was greater than or equal to five at risk alleles (20.3%), while it was low when patients presented a score less than or equal to an at risk allele (9.4%). In the future, the combination of these genetic variables together with other clinical and plasma variables in the context of algorithms validated by intervention studies should make possible a better assessment of the risk of VTE recurrence. However, up to now, the only genetic factors that have been tested for association with recurrence were those already found associated with a first VTE event. There is need to deploy agnostic approaches to identify genetic risk factors specific to recurrence, especially since its pathophysiology and its associated risk factors do not totally overlap with those of primary VTE. Some international initiatives relying on GWAS strategy have been initiated to identify new recurrence associated polymorphisms and should soon deliver their first results.

Given the evolution of the molecular technologies and the widespread of the next generation sequencing at low price, the other possibility would be to systematically screen known VTE genes or even the whole exome/genome, instead of the limited thrombophilia screening presented above. This is particularly relevant for the search of genetic defect located in AT, PC and PS genes as the surrogate measurement of the molecules in plasma does not always allow to detect the causal mutation as this is the case for the PS Heerlen (Suchon et al, 2017) or AT Dublin (Navarro-Fernández et al, 2016) variations mentioned above. A recent pilot study highlighted the promises of the whole exome sequencing approach in the context of clinical VTE (Lee et al, 2017). In 40 patients for which clinical thrombophilia testing remained unfruitful, 2 patients (5%) were found to harbour probably disease-causing mutations when the genetic data analysis was restricted to 55 candidate VTE genes. This percentage increases to 50% if one includes variations of unknown significance. These results suggest that more efficient screening could be achieved when such analysis is extended to the whole coding genes (or even over the genome) provided one has efficient tools to assess the functionality of the identified variations.

Acknowledgments

We are grateful to the INVENT consortium for providing us some unpublished data at the FGG rs2227421

Bibliography


**Table 1 - Established VT-disease genes and their susceptibility variants**

<table>
<thead>
<tr>
<th>Locus</th>
<th>Polymorphism</th>
<th>Genomic location</th>
<th>Allele&lt;sup&gt;1&lt;/sup&gt;</th>
<th>Risk Allele Frequency</th>
<th>Risk Estimate</th>
<th>Level of Certainty&lt;sup&gt;2&lt;/sup&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td>ABO&lt;sup&gt;1&lt;/sup&gt;</td>
<td>rs579459</td>
<td>Intronic</td>
<td>T/C</td>
<td>0.22</td>
<td>1.50</td>
<td>★★★</td>
</tr>
<tr>
<td></td>
<td>rs8176749</td>
<td>Leu310Leu</td>
<td>G/A</td>
<td>0.10</td>
<td>1.50</td>
<td>★★★</td>
</tr>
<tr>
<td>F2</td>
<td>rs1799963</td>
<td>3'UTR</td>
<td>G/A</td>
<td>0.008</td>
<td>2.29</td>
<td>★★★</td>
</tr>
<tr>
<td></td>
<td>rs3136516</td>
<td>Intronic</td>
<td>G/A</td>
<td>0.40</td>
<td>1.12</td>
<td>★★</td>
</tr>
<tr>
<td>F5</td>
<td>rs6025</td>
<td>Arg506Gln</td>
<td>G/A</td>
<td>0.03</td>
<td>3.25</td>
<td>★★★</td>
</tr>
<tr>
<td></td>
<td>rs4524</td>
<td>Lys858Arg</td>
<td>A/G</td>
<td>0.75</td>
<td>1.20</td>
<td>★★★</td>
</tr>
<tr>
<td>F8</td>
<td>rs114209171</td>
<td>Upstream</td>
<td>T/C</td>
<td>0.80</td>
<td>1.11</td>
<td>★★</td>
</tr>
<tr>
<td>F11</td>
<td>rs2289252</td>
<td>Intronic</td>
<td>C/T</td>
<td>0.41</td>
<td>1.35</td>
<td>★★</td>
</tr>
<tr>
<td></td>
<td>rs2036914</td>
<td>Intronic</td>
<td>C/T</td>
<td>0.60</td>
<td>1.35</td>
<td>★★</td>
</tr>
<tr>
<td>FGG</td>
<td>rs2066865</td>
<td>3'UTR</td>
<td>C/T</td>
<td>0.25</td>
<td>1.25</td>
<td>★★★</td>
</tr>
<tr>
<td></td>
<td>rs2274221</td>
<td>3'UTR</td>
<td>A/C</td>
<td>0.34</td>
<td>1.13&lt;sup&gt;4&lt;/sup&gt;</td>
<td>★</td>
</tr>
<tr>
<td>KNG1</td>
<td>rs7100446</td>
<td>Ile581Thr</td>
<td>Ile/Thr</td>
<td>0.40</td>
<td>1.20</td>
<td>★★★</td>
</tr>
<tr>
<td></td>
<td>rs6088735</td>
<td>Upstream</td>
<td>C/T</td>
<td>0.23</td>
<td>1.11</td>
<td>★★</td>
</tr>
<tr>
<td>PROC</td>
<td>rs867186</td>
<td>Ser219Gly</td>
<td>Ser/Gly</td>
<td>0.10</td>
<td>1.22</td>
<td>★★★</td>
</tr>
<tr>
<td>PROS1</td>
<td>rs121918472</td>
<td>Ser501Pro</td>
<td>Ser/Pro</td>
<td>0.002</td>
<td>6.57</td>
<td>★★★</td>
</tr>
<tr>
<td></td>
<td>rs121918474</td>
<td>Lys196Glu</td>
<td>Lys/Glu</td>
<td>0.009&lt;sup&gt;5&lt;/sup&gt;</td>
<td>5.00</td>
<td>★★★</td>
</tr>
<tr>
<td>SERPINC1</td>
<td>rs227624</td>
<td>Val30Glu</td>
<td>Val/Glu</td>
<td>0.004</td>
<td>2.30</td>
<td>★★★</td>
</tr>
<tr>
<td></td>
<td>rs2227589</td>
<td>Intronic</td>
<td>C/T</td>
<td>0.10</td>
<td>1.05</td>
<td>★★</td>
</tr>
<tr>
<td>SLC44A2</td>
<td>rs2288904</td>
<td>Arg154Gln</td>
<td>Arg/Gln</td>
<td>0.76</td>
<td>1.28</td>
<td>★★</td>
</tr>
<tr>
<td>STXB5</td>
<td>rs1039084</td>
<td>Asn436Ser</td>
<td>A/G</td>
<td>0.55</td>
<td>1.11</td>
<td>★★</td>
</tr>
<tr>
<td>THBD</td>
<td>rs16984852</td>
<td>5'UTR</td>
<td>G/T</td>
<td>0.005&lt;sup&gt;6&lt;/sup&gt;</td>
<td>2.80</td>
<td>★★★</td>
</tr>
<tr>
<td>TSPAN15</td>
<td>rs78707713</td>
<td>Intronic</td>
<td>T/C</td>
<td>0.90</td>
<td>1.42</td>
<td>★</td>
</tr>
<tr>
<td>VWF</td>
<td>rs1063856</td>
<td>Thr789Ala</td>
<td>Thr/Ala</td>
<td>0.40</td>
<td>1.20</td>
<td>★★★</td>
</tr>
</tbody>
</table>

Private mutations in known VT genes (eg F2, F9, PROC, PROS1, SERPINC1) are not listed.

1 Common/minor alleles. The allele associated with increased risk of VT is shown in bold.
2 ★★★ variants achieving the three following criteria: definitive statistical evidence - established functionality - well-characterized associated pathophysiology; ★★ variants achieving two of the three above criteria; ★ variants achieving only one of the three above criteria
3 The rs579459 and rs8176749 are tagging the ABO A1 and B blood groups, respectively.
4 Information kindly provided by the INVENT consortium (Germain et al, 2015).
5 variant identified so far only in the Japanese population
6 variant identified so far only in the Chinese population

---

<sup>1</sup> Common/minor alleles. The allele associated with increased risk of VT is shown in bold.
<sup>2</sup> ★★★ variants achieving the three following criteria: definitive statistical evidence - established functionality - well-characterized associated pathophysiology; ★★ variants achieving two of the three above criteria; ★ variants achieving only one of the three above criteria
<sup>3</sup> The rs579459 and rs8176749 are tagging the ABO A1 and B blood groups, respectively.
<sup>4</sup> Information kindly provided by the INVENT consortium (Germain et al, 2015).
<sup>5</sup> variant identified so far only in the Japanese population
<sup>6</sup> variant identified so far only in the Chinese population