UNIVERSITAT DE BARCELONA

Study of the Regulatory Mechanisms of Gene Expression in Venous Thromboembolic Disease: microRNAs

Alba Rodríguez Rius

UNIVERSIDAD DE BARCELONA

FACULTAD DE FARMACIA Y CIENCIAS DE LA ALIMENTACIÓN

Study of the Regulatory Mechanisms of Gene Expression in Venous Thromboembolic Disease: microRNAs

ALBA RODRÍGUEZ RIUS

2020

UNIVERSIDAD DE BARCELONA

FACULTAD DE FARMACIA Y CIENCIAS DE LA ALIMENTACIÓN

DOCTORADO EN BIOTECNOLOGÍA

Study of the Regulatory Mechanisms of Gene Expression in Venous Thromboembolic Disease: microRNAs

Memoria presentada por Alba Rodríguez Rius para optar al Título de Doctor por la Universidad de Barcelona

Dr. José Manuel Soria Fernández
Director de Tesis

Dra. Sonia López Moreno
Directora de Tesis

Dra. Josefa Badía Palacín
Tutora

Alba Rodríguez Rius
Doctoranda

Barcelona, 2020

Abstract

Venous Thrombosis (VT) is a frequent complex disease that involves a disruption of the balance of the hemostatic system. It is a highly heritable disorder in which de known genetic causes only account 15% of genetic susceptibility. microRNAs (miRNAs) are small non-coding RNAs that regulate gene expression. The relevance of miRNAs in human diseases is due both to their pathogenic function through the regulation of gene expression, and to their possible use as clinical tools. In this regard, promising findings in other complex diseases encourage their use as clinical biomarkers. However, up to now, the implication of miRNAs in VT has not been studied in-depth.

The main objective of this Thesis was to study the regulatory mechanisms of the gene expression through miRNAs in VT. Therefore, the first specific objective was to identify a plasma miRNA profile associated with VT and to analyze its suitability as biomarker. The second aim was, given the function of miRNAs as regulators of the gene expression, to dissect the interactions between biological layers in the biological context of VT. Finally, in order to help the translation of miRNAs into the clinic landscape, the third objective was to identify factors affecting either expression or quantification of miRNAs in plasma.

In the first article, the differential expression of miRNAs in plasma was analyzed in the 'Genetic Analysis of Idiopathic Thrombophilia' 2 (GAIT-2) Project, which involves 935 individuals belonging to 35 extended Spanish families with idiopathic thrombophilia. First, we conducted a discovery phase, in which 752 miRNAs were measured in plasma by quantitative Polymerase Chain Reaction in 104 individuals of GAIT-2 (52 VT cases and 52 controls). Sixteen miRNAs were selected: nine associated with VT and seven with clinical VT phenotypes of interest. Therefore, the validation phase included 16 miRNAs,

which were measured in the entire GAIT-2 (n=935). Four of the miRNAs were significantly associated with VT (false discovery rate <0.1): hsa-miR-885-5p, hsa-miR-126-3p, hsa-miR-192-5p and hsa-miR-194-5p. All the four miRNAs returned significant odds ratio for VT, in the range 1.3-2.12. The discriminatory ability of the profile including the four miRNAs, age and sex retuned an area under the curve (AUC) of 0.7. In addition, significant correlations were found between the miRNAs and clinical VT phenotypes, such as thrombin generation. Finally, we found that the four miRNAs had predicted or validated target genes in the clotting cascade pathway.

In the second article, the four miRNAs identified above were integrated with the gene expression levels of 260 genes of the blood coagulation pathway and 14 clinical VT phenotypes. 51 VT cases and 51 controls of the GAIT-2 Project were included. Feature selection was conducted by the building of linear models for VT discrimination, which were then optimized using penalized regression. We obtained three models with AUC >0.7. The first model (VT ~ *GATA2* + von Willebrand Factor) showed that the expression of *GATA2* in blood was inversely correlated with the blood levels of von Willebrand Factor, and that the disruption of this relationship represents a prothrombotic phenotype. The second model (VT ~ Factor IX, *ANXA2* + *ENTPD1* + *ILK* + *PDPK1* + *PRKAR1A* + *STXBP3* +hsa-miR-885-5p + hsa-miR-192-5p) represented an interaction between the fibrinolytic system and platelet activation through the αIIbβ3 signaling pathway, as a mechanism underlying VT. The third model (VT ~*CSRP1*+ *LYN* + hsa-miR-192-5p+ hsa-miR-885-5p) revealed the interaction between two group of genes involved in platelet activation, correlated with Protein S, and two miRNAs, in the biological context of VT.

In the third article, we used the miRNA expression data of the discovery phase (103 miRNAs in 104 subjects) to analyze the effect of biological and technical factors in their expression and/or quantification. First, we found that, while all the sample were in the

safe range of hemolysis, the hemolysis marker represented ~10% of the shared variability of miRNA expression. Therefore, using this value as a continuous covariate, beyond as a categorical control, could help to increase consistency across miRNA studies. Second, we found that the expression of miRNAs in plasma was not biased by any blood cell count and that the expression levels of miRNAs synthetized in a specific blood cell are not a mere mirror of its corresponding cell count. Then, we identified 1,323 genetic variants associated with the expression of 16 miRNA genes, that represent 158 independent *loci*. Finally, we found that these *loci* were enriched in promoter regions from several tissues, though not in blood tissue. This finding is in agreement with the results regarding blood cell counts, and encourages the role of circulating miRNAs as biomarkers of tissue specific conditions.

In conclusion, the results on this Thesis have allow to identify a miRNA profile which expression in plasma is associated with VT, and to prove its suitability as biomarker for VT risk assessment. In addition, we have identified interactions between gene expression levels, clinical VT phenotypes and miRNAs underlying the disease and thus, we have shed light on the biological signature of VT. Finally, we have identified technical factors affecting the quantification of miRNAs in plasma and genetic variants associated with their expression.

Resumen

La trombosis venosa (VT) es una enfermedad de alta prevalencia que implica un desequilibrio en el sistema hemostático. Se trata de una enfermedad compleja con alta heredabilidad, en la que las variantes genéticas de riesgo conocidas apenas explican el 15% de la susceptibilidad genética. Los microRNAs (miRNAs) son moléculas de RNA no codificante que regulan la expresión génica. La relevancia de los miRNAs en las enfermedades humanas reside tanto en su papel patogénico mediante la regulación de la expresión génica, como en su potencial uso clínico. En este sentido, hallazgos prometedores en otras patologías alientan el uso de los miRNAs como biomarcadores. Sin embargo, la implicación de los miRNAs en la VT no se ha estudiado en profundidad todavía.

El objetivo de esta Tesis fue estudiar los mecanismos de regulación de la expresión génica mediante miRNAs en la VT. Para ello, el primer objetivo fue identificar un perfil de expresión plasmática de miRNAs asociado a la VT y explorar su utilidad como biomarcador. El segundo, basándonos en la función de los miRNAs como reguladores de la expresión génica, estudiar las interacciones entre distintas capas biológicas en el contexto de la VT. Por último, para favorecer la inclusión de los miRNAs en el ámbito clínico, identificar factores que afecten a la expresión y/o cuantificación de miRNAs en plasma.

En el primer artículo, se explora la expresión diferencial de miRNAs en VT utilizando la población del proyecto ‘Genetic Analysis of Idiopathic Thrombophilia’ 2 (GAIT-2), compuesta por 935 individuos de 35 familias extensas españolas con trombosis idiopática. El diseño experimental implicó una fase de descubrimiento en la que 752 miRNAs se cuantificaron, por reacción en cadena de la polimerasa cuantitativa, en plasma de 104

individuos del GAIT2 (52 casos y 52 controles). Se seleccionaron 16 miRNAs (nueve asociados con la enfermedad y siete asociados con fenotipos clínicos de interés) para la fase de validación, en la que se cuantificaron en toda la población GAIT-2 (n=935). Cuatro de los miRNAs se asociaron significativamente con la enfermedad (false discovery rate <0.1): hsa-miR-885-5p, hsa-miR-126-3p, hsa-miR-192-5p y hsa-miR-194-5p. Las odds ratio respecto a VT fueron significativas para los cuatro miRNAs, en el rango 1.3-2.12. La capacidad discriminatoria del perfil de los cuatro miRNAs, edad y sexo, alcanzó un área bajo la curva (AUC) de 0.77. Además, se identificaron correlaciones significativas con fenotipos clínicos de VT, como la generación de trombina. Por último, encontramos que los cuatro miRNAs tienen como posibles dianas genes de la cascada de la coagulación.

El segundo artículo supone la integración de los cuatro miRNAs identificados con los niveles de expresión en sangre de 260 genes de la vía de la coagulación y con 14 fenotipos clínicos de VT. Se incluyeron 51 casos de VT y 51 controles del GAIT-2 y se llevó a cabo la selección de variables mediante la construcción de modelos lineales para la discriminación de VT, y su optimización mediante regresión penalizada. Obtuvimos tres modelos con AUC >07. De la interpretación del primer modelo (VT ~ GATA2 + Factor von Willebrand), se desprende que los niveles de expresión de *GATA2* están inversamente correlacionados con los niveles de Factor von Willebrand y que la disrupción de dicha relación representa un fenotipo protrombótico. El segundo modelo (VT ~ Factor IX, *ANXA2* + *ENTPD1* + *ILK* + *PDPK1* + *PRKAR1A* + *STXBP3* +hsa-miR-885-5p + hsa-miR-192-5p) pone de manifiesto la interacción entre el sistema fibrinolítico y la activación plaquetaria mediante la vía de señalización αIIbβ3, como mecanismo subyacente a la VT. El último modelo (VT ~*CSRP1*+ *LYN* + hsa-miR-192-5p+ hsa-miR-885-5p) identifica dos grupos de genes implicados en la activación plaquetaria,

correlacionados con la Proteína S, que junto a dos miRNAs representan un perfil biológico característico de VT.

En el tercer artículo, utilizando los datos de la fase de descubrimiento (103 miRNAs en 104 sujetos), se analiza el efecto de variables biológicas y técnicas en la expresión de miRNAs en plasma. Primero, identificamos que aun cuando todas las muestras estaban dentro del rango seguro de hemólisis, el marcador de hemólisis representó un 10% de la variabilidad común de la expresión de miRNAs. Por lo tanto, usar dicho marcador como covariable continua además de como control categórico, podría ayudar a mejorar la concordancia entre estudios. Segundo, de nuestros datos se desprende que los conteos celulares no sesgan sistemáticamente la expresión de miRNAs en plasma y que los niveles de expresión de miRNAs sintetizados en un determinado tipo celular no son un mero reflejo del conteo celular correspondiente. A continuación, identificamos 1,323 variantes genéticas asociadas con la expresión de 16 genes de miRNAs, que suponen 158 *loci* independientes. Por último, identificamos que dichos *loci* están significativamente enriquecidos en regiones promotoras de diversos tejidos, pero no de sangre. Este hallazgo, en línea con los resultados de conteos celulares, alientan el uso de los miRNAs circulantes como biomarcadores de condiciones específicas de tejido.

Como conclusión, el trabajo realizado en esta Tesis ha permitido identificar un perfil de miRNAs cuya expresión en plasma está asociada con la VT y hemos demostrado que dicho perfil puede ser utilizado como biomarcador para el riesgo de VT. A su vez, hemos identificado interacciones entre niveles de expresión génica, fenotipos clínicos y miRNAs subyacentes a la VT, arrojando así luz sobre la firma biológica de la VT. Por último, hemos identificado factores técnicos que causan variabilidad en la cuantificación de miRNAs en plasma y variantes genéticas que regulan su expresión.

Table of Content

List of Tables

List of Figures

Abbreviations

Abbreviation	Description
AGO2	Argonaute 2
APCR	Activated Protein C Resistance
aPTT	activated Partial Thromboplastin Time
AUC	Area Under the Curve
BMI	Boddy Mass Index
cDNA	complementary-DNA
DNA	DeoxyRibonucleic Acid
DVT	Deep Vein Thrombosis
eQTL	expression Quantitative Trait Loci
EXP5	Exportin 5
FDR	False Discovery Rate
FII	Coagulation Factor II, prothrombin
FIIa	Coagulation Factor II activated, thrombin
FIX	Coagulation Factor IX
FIXa	Coagulation Factor IX activated
FV	Coagulation Factor V
FVa	Coagulation Factor V activated
FVII	Coagulation Factor VII
FVIIa	Coagulation Factor VII activated
FVIII	Coagulation Factor VIII
FVIIIa	Coagulation Factor VIII activated
FX	Coagulation Factor X
FXa	Coagulation Factor X activated
FXI	Coagulation Factor XI
FXIa	Coagulation Factor XI activated
FXII	Coagulation Factor XII
FXIIa	Coagulation Factor XII activated
FXIII	Coagulation Factor XIII
FXIIIa	Coagulation Factor XIII activated
GAIT	Genetic Analysis of Idiopathic Thrombophilia
GWAS	Genome-Wide Association Study
GRS	Genetic Risk Score
LASSO	Least Absolute Shrinkage and Selection Operator
miR-eQTL	miRNA - Expression quantitative trait loci
miRISC	miRNA-Induced Silencing Complex
miRNA(s)	microRNA(s)
mRNA(s)	messenger-RNA(s)
NGS	Next Generation Sequencing
nt	Nucleotides
OLS	Ordinary Least Squares

OR	Odds Ratio
PC(s)	Principal Component(s)
PCA	Principal Component Analysis
PE	Pulmonary Embolism
pre-miRNA	precursor miRNA
pri-miRNA	primary miRNA
PT	Prothrombin Time
qPCR	quantitative Polymerase Chain Reaction
ROC	Receiver Operating Characteristic
RNA(s)	Ribonucleic Acid(s)
RNApol-II	RNA polymerase II
RNA-seq	RNA-sequencing
RT	Reverse Transcription
SNP(s)	Single Nucleotide Polymorphism(s)
TF	Tissue Factor
TGT	Thrombin Generation test
t-PA	tissue Plasminogen Activator

1

INTRODUCTION

1. Venous Thromboembolic Disease

1.1. Definition and epidemiology

Venous thromboembolism or venous thrombosis (VT) is a complex disease that involves two different clinical manifestations: deep vein thrombosis (DVT) and pulmonary embolism (PE). DVT occurs when a blood clot (i.e., a thrombi) is formed in a deep vein, usually in large veins of legs or pelvis. PE happens when the blood clot breaks, moves into the lung vessels through the bloodstream and leads to a blockage of the lung artery [1].

In general population, VT incidence rate for the first-event is estimated in ~1.4 per 1,000 person-years, with DVT as the most frequent presentation [2]. The recurrence rate is about 30% at 10 years [3]. The main factor that affects the incidence is the age, with a higher risk in the older age [2]. Moreover, the risk, and therefore the incidence, increases dramatically in some patient cohorts; for example, cancer patients have 7-fold risk of VT [4] and hospitalized patients have >100-fold risk [5].

The mortality rate of VT is estimated around 10-30% within one month, mostly due to PE [2,6]. It should be noted that VT is the leading cause of preventable in-hospital mortality [7]. Regarding morbidity, beyond the risk of recurrence, suffering a thrombotic event implies, in 20-50% of the cases, the development of post-thrombotic syndrome [8], which includes a range of symptoms such as swelling, edema, or ulceration. Finally, VT supposes a significant economic burden for health systems. It is difficult to estimate a cost due to the recurrence, the misdiagnosis and because it is frequently related to other diseases. As a representative example, it is estimated that the cost of direct treatment of a primary VT event ranges between $10,000-16,000 per patient [9], while the 5-years cost increases in 1.5 fold in major surgery patients compared to those that did not develop VT

($55,956 vs. $32,718) [10]. Overall, in the United States, the total economic impact in the health system of VT is estimated in a range of $7-10 billion per year [11].

1.2. Etiology and biology: Hemostasis and coagulation

The term ‘hemostasis’ refers to the biological processes through which the body maintains a normal blood flow. In this way, the hemostatic system attempts to reach a proper balance between bleeding due to a hypocoagulable state, and thrombosis due to a hypercoagulable state [12]. The hemostatic system involves three different processes that could be explained as three sequential phases: primary hemostasis, secondary hemostasis and fibrinolysis. Primary hemostasis is the immediate response to a vessel injury to form a platelet plug at the site of injury. Secondary hemostasis implies the activation of the coagulation cascade to produce a cross-linked fibrin clot to stabilize the platelet plug. Finally, the fibrinolytic system breaks the fibrin clot and promotes tissue repair [13,14].

An injury in the vessel exposes the subendothelial matrix, which contains ligands for platelets, such as collagen or von Willebrand Factor (vWF). The formation of the platelet plug begins because platelets bind those exposed ligands (platelet adhesion). This binding promotes the activation of platelets, which release compounds that trigger their aggregation and the activation of the coagulation cascade [15]. The coagulation cascade (Figure 1) implies the sequential activation of the proteins involved, the coagulation factors. The coagulation factors are mainly serine proteases that circulate in an inactive form (zymogens). The activated form is usually represented by the name of the protein followed by ‘a’ (e.g., protein coagulation factor VII [FVII] activated is FVIIa) [16]. As shown in Figure 1, the initiation of the coagulation cascade is divided in two main pathways: intrinsic and extrinsic. The intrinsic pathway is activated by blood components with negative charge released due to the vessel injury, such as collagen or platelet

components. Then, FXII, FXI and FIX are sequentially activated. FIXa forms a complex with FVIIIa (FVIII circulates associated with vWF and is usually activated by thrombin.) [16,17]. In the extrinsic pathway, the tissue factor (TF) exposed by subendothelial cells activates FVII [18]. Both intrinsic and extrinsic pathways are followed by the 'common' pathway. It starts with FX, which can be activated by FIXa + FVIIIa complex from the intrinsic path or by FVIIa + TF from the extrinsic path. FXa joins other cellular components, including FVa, to conform the prothrombinase complex. This complex converts FII (prothrombin) to FIIa (thrombin). Thrombin cleaves the fibrinogen into fibrin and activates FXIII, which crosslinks fibrin polymers in the platelet plug to conform a stable fibrin clot [13]. Finally, in the third step, the fibrinolytic system breaks this stable clot. First, tissue plasminogen activator (t-PA) or urokinase plasminogen activator (u-PA), activate plasminogen into plasmin. Then, plasmin digests the crosslinked fibrin chains and releases degradation products and D-Dimer [13,19].

This schematic and simplified explanation of the hemostatic system reflects the complexity of this pathway, in which not only a large number of molecules are involved, but also the three phases are (and need to be) strongly coordinated. When we define VT as a disorder, we are talking about a disruption of the balance of the hemostatic system that leads to a prothrombic (hypercoagulable) state and consequently, to an increased risk of suffering a thrombotic episode.

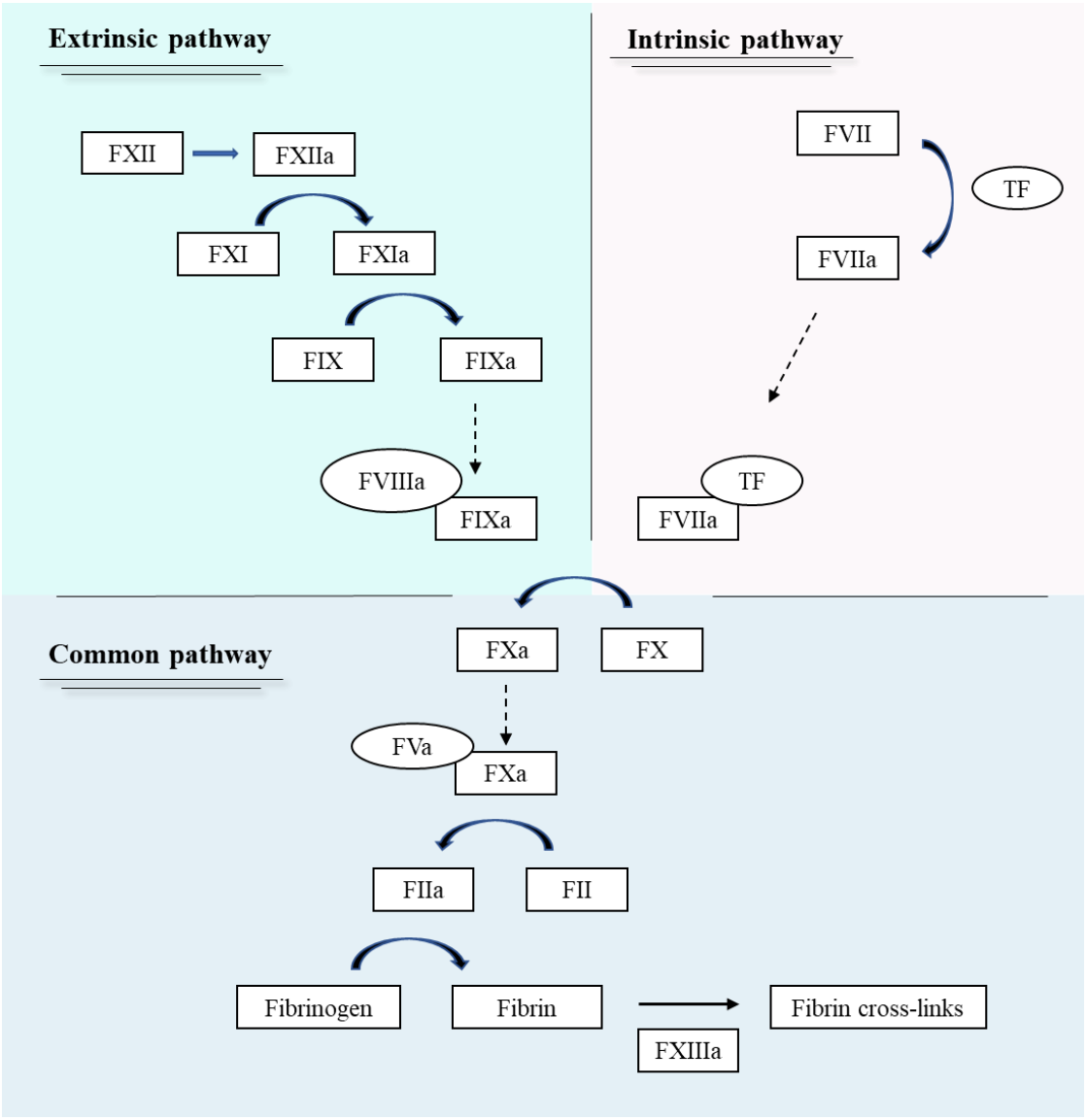


Figure 1: The clotting cascade. Schematic representation of the main steps of the clotting cascade. The extrinsic pathway is triggered by FXII activation, primarily due to negatively charged blood components released from either platelets or vessels after a vessel injury. The extrinsic pathway begins with the activation of FVII by TF, usually exposed by subendothelial cells, although platelets and other cells can also express TF. Both paths converge on the common pathway, which starts with the activation of FX and leads to form a stable fibrin clot.

1.3. Risk factors

VT is a complex (i.e., multifactorial) disease and therefore, the risk factors that predispose one to its development are both, genetic and environmental, as well as their interaction.

Table 1 summarizes the most stablished environmental risk factors [20–22], grouped into risk categories. Briefly, some of the environmental risk factors correspond to transient condition (e.g., pregnancy or immobilization), while other could be considered continuous (e.g., smoking or obesity). Also, they could be classified into modifiable or non-modifiable. As shown in Table 1, some of them are hospitalized-related situations, such as immobilization, bed rest or surgery. These risk factors justify the relevance of VT not only as an own disease, but also as a common complication for a range of health problems.

Regarding the genetic component, the heritability of VT is estimated between 50-60% [23–26]. The heritability is defined as the variability on the risk to suffer VT that become determined by the genetic component. Consequently, a range of genetic variants have been associated with VT and related to a considerable increase in the risk of VT. Table 2 summarizes the main genetic factors identified and their odds ratio (OR) for VT [27,28]. The most relevant examples are FV-Leiden, the prothrombin mutation G20210A, and the ABO blood group.

Table 1: Main environmental risk factors for VT. Classification attending to the odds ratio (OR) into risk: high (OR >10), medium (OR 2-9) and low (OR <2). Selection of the risk factors considered in Ref. [20–22].

High	Medium	Low
Immobilization	Oral contraception	Age
Surgery	Cancer	Smoking
	Pregnancy	Obesity
	Previous VT	Bed rest

FV-Leiden [29] is the common name for the rs6025, located in the *F5* gene. This single nucleotide polymorphism (SNP) encodes a change from arginine to glutamine in the sequence of FV. Activated Protein C, an anticoagulant factor, cannot inactivate FV without this amino acid. This lack of response to the action of Protein C is named Activated Protein C Resistance (APCR), which is an independent risk factor for VT. Thus, the FV-Leiden mutation is associated with an increased risk of VT through an increase in APCR. As shown in Table 2, whereas it is not a frequent mutation, the risk for the carriers of the risk allele is high, especially in homozygotes. On the other hand, the mutation G20210A in the prothrombin gene (*F2* gene, rs1799963), implies also a high VT risk [30]. This mutation leads to an overactive *F2* gene, which produces more prothrombin and consequently, more thrombin. Therefore, it causes a hypercoagulable state because of the elevated prothrombin levels. Finally, the ABO group locus [31]: A and B alleles (and mainly, the A1 allele) are associated with increased levels of vWF. Given that vWF is a protective factor for circulant FVIII, the concentration of both of them (vWF and FVIII) are increased in non-O blood group individuals, whom therefore have a greater risk of VT.

Table 2: Main genetic variants associated with the risk of VT. Selection of the genetic variants and their associated frequencies and odds ratio (OR) from Ref. [27,28].

Locus	SNP	Frequency [a]	OR
AB0	rs9411377	0.24	1.68
	rs2519093	0.31	1.36
F5	rs6025 (FV-Leiden)	0.05	3
F11	rs2036914	0.52	1.35
F2	rs1799963	0.02	2.50
TSPAN15	rs78707713	0.87	1.34
FGG	rs2066865	0.25	1.47
PROCR	rs867186	0.07	1.22
SLC44A2	rs2288904	0.78	1.12

SNP= single nucleotide polymorphism, OR= odds ratio, [a] Frequency of the effect allele.

Moreover, the interaction between environmental and genetic risk factors should be considered. For example, a study in a cohort of women [32] discovered that, while the risk of VT increased 4-fold in oral contraceptive users and 8-fold in FV-Leiden carriers, the risk of VT in users of oral contraceptives who were carriers of the FV-Leiden mutation increased 30-fold.

However, it is estimated that around 20% of the VT events are not related to any known risk factor and thus, are considered idiopathic [3]. Moreover, the known genetic variants only explain around 15% [33] of the genetic susceptibility to VT. This gap of knowledge is commonly referred to as 'missing heritability' and, as explained later, unraveling this missing heritability is one of the primary goals of current VT research.

1.4. Clinical management: current situation and needs

The recommended procedures for the clinical management of VT, regarding diagnosis, treatment and prophylaxis, are discussed and summarized into published guidelines for clinicians, such as the ones released by the American Society of Hematology [34], or the European Society of Cardiology [35].

For diagnosis of a VT event [35,36], the first step is to evaluate the symptoms (which sometimes are unspecific) and evaluate the risk factors, usually with the Well's score [37] or modified versions. These scores return a level of risk, considering mainly the additive effect of environmental factors such as cancer, immobilization or previous VT events. If there is a subsequent suspicion of VT, the D-Dimer is usually assayed [38]. The level of D-Dimer is a biomarker with high sensitivity but low specificity. Thus, a positive value could be a false positive, but a negative value discards a VT event. Finally, if D-Dimer is positive or if the first risk assessment revealed a high-risk situation, image is used to

confirm or discard the diagnosis: computed tomography angiographic for PE [39] and ultrasonography for DVT [36].

Treatment [34,35] usually includes initial heparin-based treatment for a few days, followed by oral anticoagulant therapy for 3-6 months, either direct oral anticoagulants or vitamin-K antagonists. It should be noted that anticoagulant therapies involve a risk of bleeding. In this way, anticoagulant doses, either for treatment or prophylaxis, must be adjusted to achieve a proper balance of the hemostatic system.

Prophylaxis [34–36] consists of prescribing anticoagulant treatment, usually at low doses, to prevent VT when a risk factor is present. Therefore, patients are classified into a certain risk group and received a proportional prophylactic treatment. In general, the guidelines evaluate particular risk factors in different groups of patients, mainly: thrombophilia, hospitalized, cancer, pregnancy or surgery patients. When a prophylactic treatment is recommended, it may be based on comprehension stocking and/or parenteral or oral pharmacologic prophylaxis (e.g., low-molecular-weight heparins, unfractionated heparins, vitamin K antagonists, aspirin or direct oral anticoagulants). It should be notice that a previous VT event is a risk factor and thus, in this 'prophylaxis' statement we are including also the prevention of recurrence. In last years, the prophylactic strategies recommendations have been improved, regarding the treatment but also in terms of *how* and *when* to assess the risk.

Despite these advances, the incidence of VT remains unchanged during last decades [40,41]. It should be considered that some risk factors, such as obesity, cancer or the older age, are more prevalent each day. Thus, the aforementioned advances in prophylactic strategies have been useful, otherwise the global incidence would have increased [41,42]. However, in order to reduce the global incidence of VT, more accurate methods for VT

risk assessment are needed, so that prophylactic strategies can be improved and personalized.

1.5. Intermediate phenotypes in VT

1.5.1. Clinical phenotypes

Clinical phenotypes are those quantitative variables that are used in the clinic landscape of VT. Briefly, they are:

(1) Blood concentration of proteins involved in the coagulation system, such as: FII, FV, FVII, FVIII, FIX, FX, FXI, FXII, FXIII, vWF, fibrinogen, Protein S or Protein C. Most of them are usually considered only as the concentration of the functional form in blood. Some of them are useful to test the VT risk, for example, basal deficiencies of Protein S, Protein C or fibrinogen, as well as increased levels of FVIII are known risk factors for VT.

(2) Parameters of clinical assays to evaluate the activity of hemostatic processes. Prothrombin time (PT), thrombin time (TT) and activated partial thromboplastin time (aPTT) evaluate the time that a citrated plasma needs to form a clot after different stimulus are provided [43]. All of them and their combinations serve to identify and measure abnormalities in the coagulation pathway, mainly in the initiation step. The thrombin generation test (TGT) also provides information about the propagation and termination phases [44]. In addition, APCR is usually assayed by a dual determination of the aPTT with and without adding protein C activated [45].

These clinical phenotypes are a mainstay in the clinical management of VT, with respect to the assessment of the VT risk and to monitor the response to anticoagulant drugs. The combination of these phenotypes (e.g., the TT parameter must be evaluated considering

the fibrinogen concentration), as well as their combination with the genetic variants (e.g., APCR associated with FV-Leiden or not) help to characterize the pathogenesis in a given patient. This role as intermediate features between the disease and the biological cause is also very valuable in basic and translational research. In addition, it should be noted that clinical phenotypes are standardized measured included in almost all VT cohorts worldwide, which further increases their utility.

1.5.2. Genetic variants

As in all biomedical fields, the genetic research in VT has been driven by the technological advances. The candidate-gene approach in the decade of the 90s, focused on genes of the clotting cascade, has moved into the era of the Genome Wide Association Study (GWAS). From the first GWAS of VT, published in 2009 [46], several GWAS have been conducted regarding both VT and clinical VT phenotypes. The GWAS with the largest sample size so far was published by Klarin et al. in 2019 [28]. Their results could serve to illustrate the current knowledge on VT genetics. First, there are 11 loci that have been associated with VT in several independent studies, which include the main genetic causes aforementioned, such as FV-Leiden or prothrombin G20210A mutation. Second, it is gaining relevance the effect of genetic variants that involve less increase in the risk (fewer OR for VT) but represent a cross-talk with other related processes, such as inflammation, platelet activation or risk factors, such as body mass index (BMI). However, despite the number of studies performed, it is estimated that known genetic causes only account 15% of genetic risk of VT [33]. Therefore, more studies are still needed to unravel the genetic factors involved in the development of VT.

Finally, in recent years, this knowledge of the genetic component of VT has been translated into genetic risk scores (GRS) for VT risk assessment. There are two main examples: (1) in 2012, de Haan et al. [47] published a 5-SNP GRS that returned an area

under the curve (AUC) of 0.69 for VT outcome, (2) in 2014, Soria et al. [33] developed the Thrombo inCode® kit that evaluates 7 genetic variants and returned an AUC of 0.7, with a better performance that the one of de Haan et al. in the studied population. While more studies are needed before they can be used extensively in the clinic, the path to 'personalized medicine' in VT begins with these GRS, so that the risk of VT for a given patient can be estimated accurately, either the individual baseline risk or the total individual risk when a transient risk situation occurs.

1.5.3. Other intermediate phenotypes

Beyond these two main intermediate phenotypes, other promising biological layers have been explored in last years. The advances in both basic knowledge (e.g., discovery of microRNAs) and laboratory technologies (e.g., high-throughput sequencing platforms) have allowed to increase the number of molecular phenotypes studied. Regarding those explored in VT, we could point out:

Transcriptomic. The study of RNA molecules in a given tissue or cell, either whole RNAs or only messenger RNAs (mRNAs). Up to now, two studies have studied this signature in whole blood as a quantitative trait in relation to VT: (1) Lewis et al. [48] studied blood gene expression in VT patients on warfarin, including 23 single and 17 recurrent VT patients and identified a set of 50 genes to distinguish between recurrent and single VT patient; and (2) the same authors [49], compared patients in four different groups of VT risk and identified over 3,000 genes differentially expressed between controls and high-risk patient. Recently [50], a Transcriptome-Wide Association Study (TWAS) (consisting of associating genetic variants with transcriptomic expression in a given tissue [51]) has identified that some loci associated with VT are also associated with the expression of certain genes never related to VT before and thus, has shed light on novel mechanisms underlying the disease.

Proteomic. It is the large-scale study of the protein set produced by a given organism, tissue or biologic context. Different study designs have been developed in VT, for example, regarding the plasma proteome of carriers of the prothrombin G20210A mutation [52], the proteomic signature of plasma microparticles in VT patients [53], or the proteome in clots obtained *ex vivo* using plasma of DVT patients [54]. One of the most relevant example so far is the study of Jensen et al. [55], that compared the whole plasma proteome before and after a DVT event and in contrast to non-VT patients, and identified some potential blood biomarkers for stablish the risk of VT.

Epigenomics. While genetics studies changes in the DNA sequence, epigenetics refers to the mechanisms to modify gene expression without altering the DNA sequence, that is, the processes through which the genotype is expressed in the phenotype [56]. The three main mechanisms are: DNA methylation, histone modification and non-coding RNAs. All of them are gaining relevance in VT given their inheritance (thus, they could explain a part of the missing heritability) and because their mechanisms could be easily reversible, so are potential therapeutic targets [57,58]. Promising findings regarding DNA methylation and histone modification have been done in relation to mechanisms underlying the VT disease [59], but also to some risk factors, such as smoking or obesity [60,61]. Finally, non-coding RNAs regulate the gene expression by interfering with RNA molecules. Main examples are microRNAs, long non-coding RNAs or piwi-interacting RNAs. Among them, the most studied up to now are microRNAs, whom relation with VT will be explained in Introduction-2.6. given the relevance for this Thesis.

Microbiomic. It is the study of the microbial composition of the body, for example, regarding the bacterial strains present, the relative proportion of strains, their metabolic status, etc. The findings in VT are still few, but some promising candidate process have been suggested. For example, the synthesis of forms of vitamin K by intestinal bacteria ,

given that Vitamin K is a needed cofactor for many coagulation factors [62], or the synthesis of the gut microbiota metabolite trimethylamine, which could affect platelet function in VT [63].

All these intermediate phenotypes (and others, such as exposomics or metabolomics) could help to better understand the biological mechanisms underlying VT. It seems logical to think that novel biomarkers and therapeutic targets could be identified. Also, it should be noted the growing interest in their integration into multi-omic studies and/or 'personalized medicine' strategies.

1.6. The Genetic Analysis of Idiopathic Thrombophilia Project

The Genetic Analysis of Idiopathic Thrombophilia (GAIT) Project was designed with the aim to identify and dissect novel genes, genetic variants and intermediate phenotypes related to VT. The GAIT-1 study was recruited between 1995 and 1997 at the Unit of Thrombosis and Hemostasis of Hospital de la Santa Creu i Sant Pau (Barcelona, Spain). GAIT-1 included 398 individuals of 21 Spanish families. Among them, 12 families were recruited via a proband with idiopathic thrombophilia and the remaining 9 were selected only on the basis of family size, without considering any phenotype. Given the relevance of the results obtained with the GAIT-1 study [64–66], the GAIT-2 Project was conceived to increase the sample size and to consider a larger range of phenotypes.

Therefore, the GAIT-2 study was recruited between 2006 and 2010 with the approval of the Review Board of Hospital de la Santa Creu i Sant Pau (Barcelona, Spain). All procedures followed the Declaration of Helsinki and all adults gave written informed consents for themselves and their minor children.

All the families of GAIT-2 were recruited via a proband with idiopathic thrombophilia. The proband must met one of these three conditions: (1) VT age of onset under 45, (2)

recurrent VT events, which at least one was spontaneous, or (3) single spontaneous VT event with a first-degree relative affected. The VT events were considered idiopathic when all biological causes of thrombosis were excluded: deficiencies of Protein S, Protein C, plasminogen, antithrombin or heparin cofactor II; APCR, FV-Leiden, antiphospholipid antibodies, lupus anticoagulant and dysfibrinogenemia. Moreover, the presence of transient risk conditions should be excluded during the three months prior to the VT event: surgery, immobilization, bone fracture, hospitalization and pregnancy. In addition, the VT event should not be secondary to neoplastic conditions, Behcet disease, or inflammatory bowel disease. Also, each family must had at least 10 participants of three generations to be included.

Finally, the GAIT-2 study recruited 935 individuals of 35 extended Spanish families. Basic characteristics of the population are given in Table 3. Briefly, the GAIT-2 involves a similar number of males and females (465 and 470, respectively), with an age range of 2.6-101 years, from whom 85 have suffered a VT event. The average number of individuals per family is 26.7, in a range of 10 to 68, and they belong to between three and five generations.

In addition, in order to successfully model the pedigrees, descriptive data of other 128 non-participants subjects (e.g., death individuals) were obtained. A representative example of the final pedigrees is shown in Figure 2.

To study the basal-risk state rather than the acute phase, the collection of biological samples was performed at least three months after the last VT event, and at least one month after any acute inflammatory process. Moreover, to avoid interferences with any of the measures performed, all subjects suspended the on-going therapies: heparins (at least 24 hours before), antiplatelet drugs (at least 15 days before), oral anticoagulants (at least 15 days before), and anti-inflammatory drugs (only if possible, 15 days before).

The GAIT-2 Project is understood as a dynamic research tool and therefore, in addition to a large range of phenotypes measured on fresh blood at the moment of the inclusion in the study, frozen samples of serum, plasma, DNA and RNA of each subject are available.

Up to now, the main intermediate phenotypes covered on GAIT-2 have been genetic variants and clinical phenotypes. While both fields continue to be studied, in this Thesis the GAIT-2 Project begins to explore other biological layers, focusing on gene expression and the mechanisms involved in its regulation.

Table 3: Characteristics of the GAIT-2 population.

	GAIT-2	VT patients	Controls
N	935	85	850
Sex-female (n, %)	470, 50.3%	55, 64.7%	415, 48.8%
Age – years	2.6 - 101.1	20.3-71.9	2.6 – 101.1
(range, average)	39.5	55.1	37.6
Body Mass Index	13.1 - 49.8	16 – 49.8	13.2 – 47.9
(range, average)	24.8	28.6	24.4
Oral contracep.[a, b] (n, %)	39, 8.3%	0	39, 9.4%
Smoking [b, c] (n, % - yes)	225, 24%	18, 21.2%	207, 24.4%
Hypertension [b] (n, %)	136,14.5%	31, 36.5%	105, 12.4%
Diabetes mellitus [b] (n, %)	42, 4.5%	9, 10.6%	33, 3.9%
Arterial thrombosis [b] (n, %)	47, 4.7%	13, 15.3%	34, 4%

n: numbers of individuals; VT: venous thrombosis; Oral contracept.: oral contraceptive active therapy. [a] Only in females, [b] codified as 'yes' or 'no', number and percentage refer to 'yes', [c] Active smokers

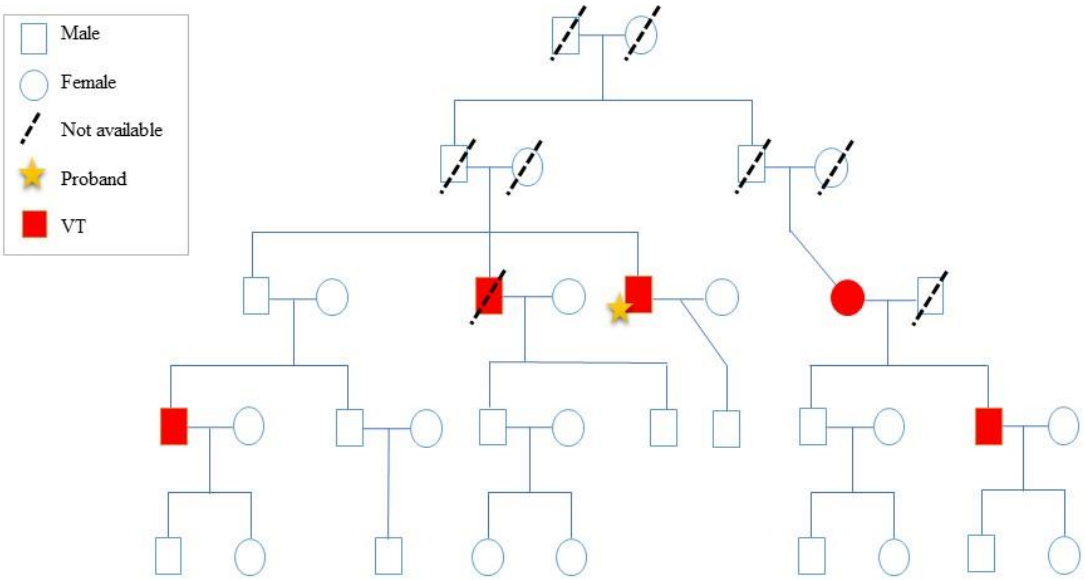


Figure 2: Example of a pedigree of the GAIT-2 Project. Real example of one of the families included in the GAIT-2 Project. Twenty-seven members of three generation of the family are participants of the study, of whom four have suffered a VT event. Also, the descriptive data of eight non-participants members were collected to model the pedigree. Highlighted the proband, the subject with idiopathic thrombophilia from which the family was recruited.

2. microRNAs

2.1. Definition and history

MicroRNAs (miRNAs) are small endogenous non coding RNAs that regulate gene expression [67]. They bind complementary sequences on the 3' untranslated region (UTR) of mRNAs and inhibit their expression by translational repression or mRNA degradation [68].

miRNAs were discovered in 1993 by Lee et al. [69] in *C. elegans*. While studying the role of the gene *lin-4* in postembryonic development, they found that the gene did not code for a protein, but for two small molecules of RNA that regulated the expression of the gene *lin-14*. Moreover, they described that the mRNA of *lin-4* bound a complementary sequence in the 3'-UTR of the mRNA of *lin-14* and defined the process as an 'antisense RNA-RNA interaction'. It took seven years until the second miRNA was discovered: *let-7* was the second miRNA gene described in *C.elegans*, [70]. Finally, the state of art of miRNAs changed dramatically in 2001, when two promising studies were published. First, Lee et al. [71] reported for first time a catalogue of 15 miRNA genes in *C.elegans* and demonstrated that the gene *let-7* was conserved across species. At the same time, Lagos-Quintana et al. [67] suggested the name of 'microRNAs' for these small non-coding RNAs and demonstrated that miRNAs existed in several vertebrates and invertebrate species. From this point, miRNAs were understood as a general mechanism for the regulation of gene expression and the number of studies conducted to elucidate the biological mechanisms surrounding the expression and function of miRNAs began to increase dramatically.

Nowadays, two decades after their discovery, at least 48,860 miRNAs have been described in 271 organisms, including *Homo sapiens*, with a total of 2,654 miRNAs identified [72].

2.2. Biogenesis of miRNAs

The biogenesis of miRNAs from their genes into their functional mature forms involves three sequential steps. Figure 3 shows a graphic representation of the process.

First, the miRNA genes are transcribed in the nucleus by RNA polymerase II (RNApol-II) [73] into the primary miRNAs (pri-miRNAs). The pri- miRNA is a long molecule (around 1Kb) that includes a stem of ~35 nucleotides (nt), a terminal loop and single-RNA segments at 3' and 5' sides. This first step become determined by the structure of the miRNA gene and the regulation of the RNApol-II. The miRNA genes are classified attending to their location, into: intronic, embedded into a gene that consequently is called 'host-gene', and intergenic [74]. Furthermore, often a group of miRNAs are encoded in proximal genetic locations and thus, the group is called a 'genetic cluster' of miRNAs [75]. In both cases, the miRNAs can be transcribed into a shared transcription unit with their host genes or clustered miRNAs, or in independent ones [76]. Moreover, even when transcribed in a single unit, they can be co-expressed or not, depending on post-transcriptional regulatory mechanisms [77]. On the other hand, regulatory factors can modulate the transcription rate of RNApol-II through binding transcription factors in the core promoter, binding directly the RNApol-II complex or interacting with histone or nucleosome modifiers, among others [78].

Second, the pri-miRNA is processed by Drosha, a nuclease that cuts the stem-loop to release the precursor miRNA (pre-miRNA), a small hairpin-shaped RNA of about ~65 nt of length [79]. The specificity of Drosha cut is ensured by DGCR8 [80]. This first post-

transcriptional step is also highly regulated. First, the efficiency of the catalytic activity of Drosha is affected by the sequence and structure of the pri-miRNA. For example, the efficiency of Drosha seems to be lower in non-conserved miRNA genes [81] , in stems with mismatching in the central region [82] or when the pri-miRNA adopts a globular tertiary structure [83]. Second, several factor can be released into the nucleus and affect to the catalytic activity of Drosha in a fast and direct way, such as Smad proteins or p53 mutants [84,85]. Furthermore, the abovementioned factors can affect differently to the Drosha activity only for certain miRNAs and therefore, regulate the expression of a specific subset of miRNAs. The same specific-miRNA regulation can be found when some factors bind directly the structure of the pri-miRNA and interact with Drosha complex [86].

Finally, in the third step, the pre-miRNA is exported to the cytoplasm by exportin 5 (EXP5). Once there, it is cleaved by Dicer, which produces a small RNA duplex that contains the -5p and -3p strands of the same miRNA gene. While one of the strands will be degraded (and thus, called ‘passenger’ strand), the other strand (the ‘guide’ strand) will be loaded into Argonaute 2 (AGO2), to conform the miRNA-induced silencing complex (miRISC), the most common active form of mature miRNAs [87]. Regarding Dicer, mechanisms of regulation have been reported at transcriptional [88,89], post-transcriptional [90] and proteomic levels [91]. On the other hand, many factors determine which strand will be the ‘guide’ and which one the ‘passenger’. Sometimes, it depends on the sequence of each strand, for example, the strand most stable in 5’ or the one with an uracil in 5’ seem to be more likely to be loaded. However, this selection may be affected by several cellular factors, such as the relative concentration of both miRNA and AGO2 in the cell, if AGO2 is joining Dicer as a complex, or if in this complex are present also other cofactors [92,93].

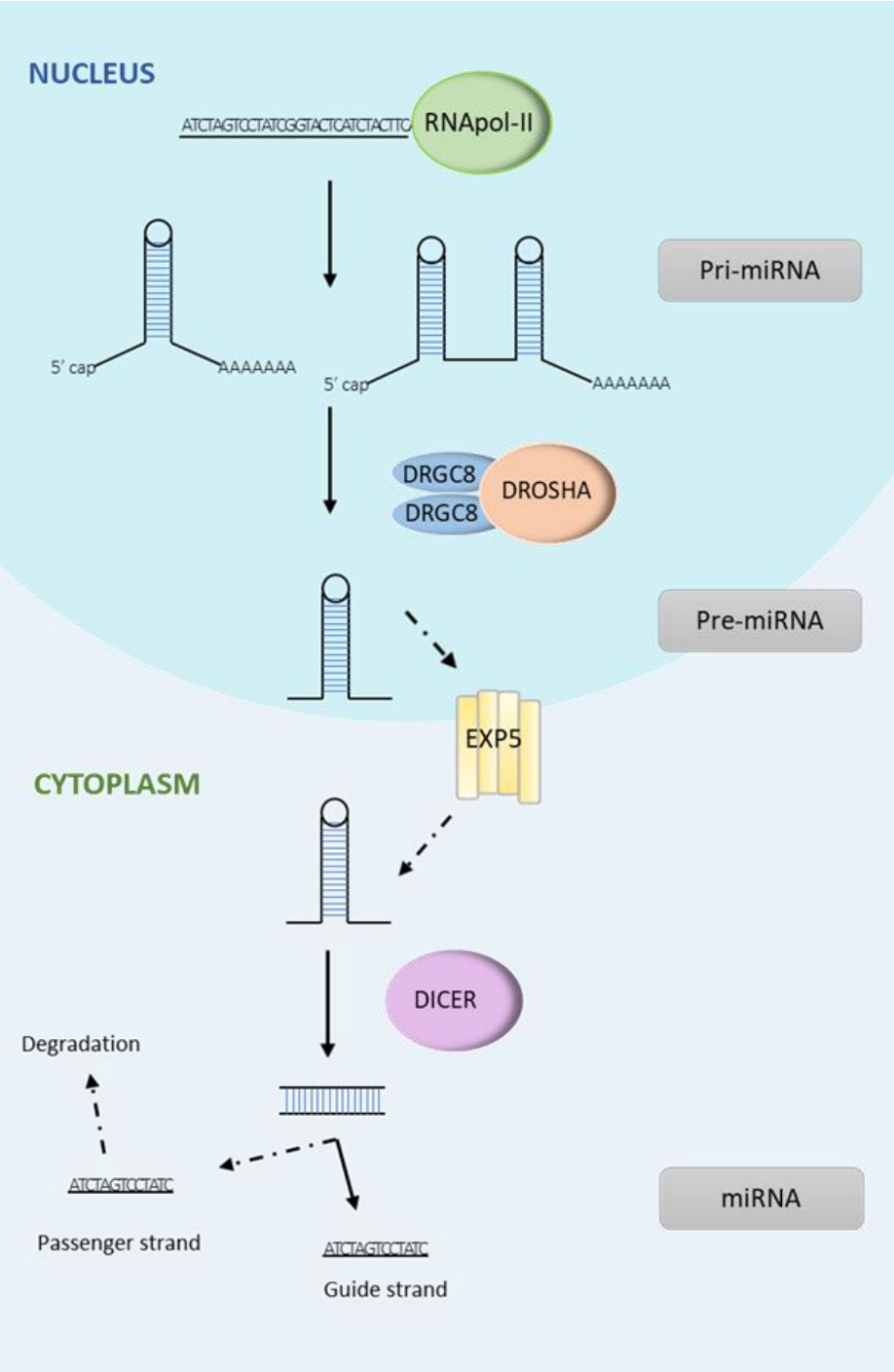


Figure 3: Biogenesis of microRNAs (miRNAs). The genes are transcribed into primary-miRNA (pri-miRNA) by the RNA polymerase II (RNApol-II) in the nucleus. DROSHA cuts the stem loop to release the precursor-miRNA (pre-miRNA), which is then exported to the cytoplasm by exportin 5 (EXP5). Finally, the pre-miRNA is cleaved by DICER, which produces a small RNA duplex. One strand will be immediately degraded (‘passenger’ strand), while the other one (‘guide’ strand) will be the final mature miRNA, which is usually loaded into an AGO2 protein.

This pathway is the most common for the biogenesis of miRNAs, although there are some less frequent and therefore called ‘non-canonical’ pathways; detailed descriptions can be found in Ref. [94–96]. Throughout this description, we have pointed out only the most frequent ways of regulation and from the simplest possible point of view. Therefore, it is clear that the regulation of the expression of miRNAs is highly regulated by different processes and depends on the interaction of a wide range of stimuli. This complexity leads to a highly specific expression of miRNAs, which justifies their use as biomarkers or therapeutic targets on certain diseases or biological conditions.

2.3. Mechanisms of action

The mechanism of action of miRNAs implies the binding between the ‘seed sequence’ of the miRNA and a complementary sequence in a mRNA. The genes regulated by each miRNA are commonly called their ‘target’ genes. Most commonly, the binding occurs in the 3’-UTR of the target mRNA, inhibiting its expression, either by degradation of the mRNA or translational repression.

Target degradation [97–99] is caused by the exonucleolytic decay of the mRNA after de-adenylation and decapping steps. This path is usually triggered by the action of a GW182 protein, associated with AGO2 protein in the miRISC complex. Thus, once the target mRNA is cleaved, GW182 is able to recruit de-adenylation and decapping factors. Also, depending on the cell stimuli, GW182 may act through other mechanisms, such as the dissociation of the poly(A) binding proteins increasing the efficiency of the de-adenylation, or promoting the decapping without affecting the poly(A) tail.

On the other hand, the mechanisms for translation repression [100–102] affect the translation of the target gene at either initiation, elongation or termination steps. Sometimes, it is due also to the action of GW182, which may interact with eukaryotic

translation initiation factors affecting the efficiency of the translation. Also, translation can be blocked by interaction with ribonucleoproteins, for example, avoiding the assembly of the 40S initiation complex or its join with the 60S subunit. Furthermore, proteases can be recruited, causing a co-translation degradation.

However, miRNAs can sometimes act through other mechanisms, for example, promoting the expression of the target gene or interacting with complementary sequences in the 5' UTR or coding regions of the target gene [103–105]. Furthermore, all of these mechanisms can occur simultaneously or sequentially, in an independent or coordinated manner, for the same or different miRNAs and/or target genes.

2.3.1. Target genes

Beyond the differences between the subsequent mechanisms to regulate the expression of a target gene, the very first requirement is that the miRNA binds to the corresponding mRNA. As aforementioned, the major determinant is the complementarity between a sequence in the 3' UTR of the mRNA and the 'seed' sequence of the miRNAs, which involves the nucleotides at position 2-7/8 of the miRNAs [106]. Therefore, since the sequence of the entire genome is known, algorithms have been developed to predict which genes might be regulated by each miRNA (hereinafter, 'predicted targets'). Besides the complementarity, each algorithm considers other factors that determine the binding [107,108]. For example, the binding is more likely if the interaction miRNA-mRNA is thermodynamically stable [109], if there is an additional pairing in the 3' side of the miRNA [110] or if the target sequence is evolutionary conserved. Overall, around 60% of human genes are predicted targets for a miRNA, and each miRNA has hundreds of predicted targets [111]. However, the algorithms produce a considerable rate of false positives for *in vivo* condition [112]. In this regard, although these algorithms provide invaluable information, they are actually modeling only if a given binding is structurally

possible. In this way, beyond this 'basal' possibility, the binding is determined by several environmental conditions that cannot be modeled from this general point of view.

Briefly, a large range of cell-environment conditions determine the target selection. For example, the concentration of the miRNA, given that the selection of the target is dose-dependent [113]. Also, it is important the relative abundance of the miRNA and the target, as well as of other targets of the miRNA, and the proteins involved in the mechanism, such as AGO2 [114]. Moreover, there are both additive and synergistic mechanisms for targets with more than one binding site [115]. Also, sometimes different miRNAs are coordinated in a given pathway [116,117]. Finally, the interaction with targets also triggers different feed-back regulation for both miRNA and target gene expression and consequently, modifies all the aforementioned factors [118,119].

Therefore, the mechanism of action for a given miRNA can be elucidated by functional assays but must be interpreted for the cell or tissue studied and for the exact environmental conditions tested, while it should be always considered the variable and adaptable nature of both the expression and the action of miRNAs.

2.3.2. Role of circulating miRNAs

miRNAs are known to be expressed in almost all human biofluids, such as blood, serum, plasma, saliva, urine, synovial fluids, breast milk, etc. [120]. Early theories suggested that circulating miRNAs were simply bioproducts passively released as consequence of cell processes or cell death. However, it was found that, in contrast to cellular miRNAs, those expressed in bio-fluids were remarkably stable against fridge-thaw cycles, RNAases or pH changes [121]. Thus, it seems logical to think that if cells protect miRNAs in some way to be stable in bio-fluids, the release could be selective and have a function, instead of simply being a passive release of bioproducts.

Regarding *how* miRNAs are released in a stable form, circulating miRNAs can be loaded into proteins, such as AGO2 [122], associated with high-density lipoproteins [123] , or released in exosomes [124], microvesicles [125] or apoptotic bodies [126]. Up to now, there are contradictory results about whether there is a majority release mechanism. Then, released miRNAs can act as direct cell-cell communication mediators or as 'hormone-like' signaling pathways. For example, Le at al. [127] found that breast metastatic cells release vesicles with miR-200 and that they are up taken by other proximal cells, which become then more potentially metastatic. Updated reviews regarding the biology surrounding circulating miRNAs can be found in Ref. [120,128–130]

Elucidating the role of circulating miRNAs, as well as regulatory pathways for both donor and recipient cells, is one of the main goals in current studies regarding miRNAs. A deeper knowledge of the biological bases is needed to understand their function, but also their use as clinical tool and their possible limitations of use.

2.4. miRNAs and disease: biomarkers.

If in 2001 miRNAs started to be understood as a common way for gene regulation, only one year later the root of future (current) studies regarding the implication of miRNAs in certain diseases was published. In a promising and innovative research, Cali et al. [131] proposed that a frequent mutation in a *loci* ligated to leukemia could affect the genes of two miRNAs in this region, mir-15 and miR-16. Moreover, they conducted functional assays to demonstrate that the mutation was associated with aberrant levels of both miRNAs, and that the cells with higher expression of the miRNAs are those with lower expression of the putative complementary target. From this very first article published in 2002 to date, more than 25,000 studies can be found by searching for 'miRNA' plus

‘disease’ in Pubmed, an average of more than 1,500 studies published each year worldwide in this regard.

During these years, as in any emerging field, the trend has been to move from basic toward translational research. Two important advances that encouraged this transition could be highlighted. First, in 2008, miRNAs were found to be expressed in blood and in a stable manner [132]. Second, from 2005-2010, the state of art of the laboratory technologies, especially for genetic studies, changed dramatically, for example, with the development of the next generation sequencing (NGS) platforms. This change also affected the quantification of miRNAs: different techniques were developed for their quantification, which were then standardized into commercial kits. Therefore, analyzing the expression of miRNA became more accessible both financially and technically. Thus, as of 2010, miRNAs began to be understood not only as a general gene regulation mechanism, but also disease-specific, stable in many biofluids and measurable in an easy and non-invasive manner; that is, became clear that miRNAs fulfilled the desirable properties of a biomarker [133].

During last years, distinct miRNA signatures have been associated to human diseases [134–136], both to acute conditions and complex chronic disorders, as useful biomarkers not only for the diagnosis of a disease, but also for the prognosis, the risk of suffer the disease, or the response to a pharmacological treatment.

Up to now, miRNAs have reached the clinical environment in a low-rate [137]. Currently, some miRNA-profiling tools are commercialized, but only for research purposes. For example, osteomiR® is available to discriminate between low and high risk of osteoporosis, and ThyraMIR® to stratify the risk of thyroid nodules after an indeterminate cytology. Around 10 years is relatively little time to move a completely new field from

the bench to the clinic and thus, the transition is expected to be finally achieved in a few years.

2.5. Basic concepts in miRNA studies

2.5.1. Nomenclature of miRNAs

Currently, there is a standardized nomenclature for miRNAs [72].

- The first three letters identify the specie. ‘hsa’ corresponds to *Homo sapiens*. Great exception is let-7 family, which because of its historical value, is named let-7 for any specie.
- ‘miR’ to recognize that it is a miRNA
- A number that is given sequentially.
- A letter (optional). If present, indicates that there exist closely related mature sequences. For example, mature forms of hsa-miR-320a and hsa-miR-320b only differ in one nt.
- Final ‘5p’ or ‘3p’ indicates the strand of the pre-miRNA from which become the mature miRNA. For some miRNAs, it is not stated, as only one of them exists (e.g., the other one is not stable and thus, never exists at RNA level).
- A final number (optional). When a miRNA gene has two different position on the genome, the final number indicate to which position is referred. For example, hsa-miR-194 is codified in two different positions of the genome, hsa-miR-194-1 encoded in chromosome 1 and hsa-miR-194-2 located in chromosome 11.

2.5.2. Quantification of miRNAs

Three main techniques can be used to quantify miRNAs [138,139]: quantitative polymerase chain reaction (qPCR), microarrays, and small RNA sequencing (RNA-seq).

For all the techniques, some characteristics of the miRNAs suppose a challenge to their measurement. Regarding the sensitivity, it should be considered that miRNAs are expressed in a very low concentration over the total mRNAs, specially the circulating ones. Second, the specificity could be limited given the small length (~22 nt) of miRNAs and that certain miRNAs only differ in one single base. Moreover, no 'house-keeping' gene is available for miRNA assays and thus, inter-individual normalization methods must be developed and tested for each research. Nowadays, qPCR is considered the gold standard technique [138], mainly used to validate results of other platforms because of its financial and time cost; notwithstanding, the aforementioned limitations must be considered also in qPCR assays.

In the present study, miRNAs will be quantified using qPCR. Thus, other techniques will not be further introduced. There are different commercial kits for qPCR measurement of miRNAs, and we will use miRCURY (Exiqon) line of products for biofluids. Briefly, the quantification process involves three steps:

Step1. Isolation of miRNAs. Purified small RNAs are isolated from plasma using a chromatography column with a special resin as separation matrix, in which only RNAs with less than 100 nt are retained. First, after removing possible contaminant, the cells are lysed and the proteins precipitated. The supernatant is loaded into the column, with subsequent wash cycles to remove any other component fixed in the column. Finally, the small-RNAs retained in the columns are eluted. They can be stored long-term at -80ºC.

Step2. Reverse transcription (RT) to complementary-DNA (cDNA). A first-strand cDNA synthesis is conducted using the reverse transcriptase. As miRNAs are not poly(A), the poly(A) polymerase is added to polyadenylate the miRNA, that will be then reverse transcribed using oligo-dT primers. The buffer provides the best

conditions for the enzyme activity, which is activate for 60 min at 42ºC ant then heat-inactivated for 5 min at 95ºC.

Step3. Amplification and measurement of miRNAs by qPCR. The qPCR is conducted under the usual conditions, that is, consecutive cycles of temperature and time that triggers denaturation, annealing and elongation processes. The amount of DNA after each cycle is measured using SYBR green as dye, which binds double stranded DNA molecules. Consequently, the intensity of the fluorescence is proportional to the amount of DNA molecules. Once finished, the cycle in which each assay reached the half of the amplification curve (Ct) is representative of the amount of the cDNA before que qPCR started.

There are some quality controls to ensure the quality of the quantification:

- Synthetic controls. Before to the isolation, three synthetic controls are added, which are in decreased concentrations: UniSp2, UniSp4 and UniSp5. Thus, obtaining similar Ct values for each of them ensure that the efficiency of the isolation process was similar across samples and for miRNAs present in both, high and low concentrations. UniSp6 is the synthetic control for the RT, added after the isolation and prior to the RT. Similar Ct values should be obtained across samples; otherwise, the efficiency of the RT cannot be warranted in the given sample. Finally, UniSp3 is the synthetic control to check the quality of the qPCR, and it is added directly in the qPCR. A highly different value of UniSp3 may indicate that the sample had any compound that interferes with the qPCR (e.g., heparins).

- Hemolysis marker. Hemolysis in the starting sample could comprise the quantification of miRNAs [140]. To estimate the hemolysis degree, it is considered the increase between the endogenous miRNAs hsa-miR-451a and hsa-miR-23a-3p. This hemolysis marker is

based on that hsa-miR-451a is only synthetized in erythrocytes, while hsa-miR-23a-3p is widely stable in plasma. Thus, if hemolysis dCt is under 5, hemolysis can be discarded in that sample. In contrast, samples with dCt >7 must be discarded, while those in the range 5-7 must be consider as in modest hemolysis risk.

Finally, once the quality of the quantification is warranted, two main corrections must be performed on the expression data before can be analyzed:

- Inter-plate calibration. To avoid variability among the qPCR plates due to technical covariates. It could be used UniSp3, considering the average value of UniSp3 in each plate in contrast to the average value across all plates. Also, it could be applied other frequent methods for qPCR assays, such as consider the median or average Ct values of the PCR plate. Given the differences expected if each plate included one or more samples, and if each sample is located in one or more plates, it should be optimized for each experimental design so that the variability due to qPCR plate is removed.
- Inter-individual normalization. To remove differences in global expression across individual. It corresponds to the classical correction by 'house-keeping' genes (genes whom expression is known to be highly stable across individuals) in other mRNA assays, but in the case of circulating miRNAs there is not an endogenous stable reference. Thus, the normalization strategy must be critically considered and selected for each assay. Some tools are available to select the best normalizer in a given assay, such as Normfinder [141] or GeNorm [142]. For experiments involving a considerable number of miRNAs and samples, the most common approach supposes the use of the miRNAs expressed in all the samples (i.e., 'global mean'). Otherwise, a miRNA or a set of miRNAs expected to be stable (for example, according to the literature considering the type of sample and the disease studied) should be selected in the design.

2.5.3. Guidelines for study circulating miRNAs as biomarkers

The aforementioned transition to the clinic is challenged by a lack of concordance between studies and a high inter and intra-assays variability, which is not surprising in such a new field. In this way, a deeper understanding of the factors causing variability in either their expression or their quantification is still needed. During last years, the relevance of some factors (e.g., the centrifugation of the starting sample or the time of storing) has been highlighted and follow subsequent recommendations is critical to ensure the results. Review of common pitfalls and caveats can be found in Ref. [143–145]. Recently, de Ronde et al. [146] published a detailed review of both pitfalls and suggested solutions for studies analyzing miRNAs as biomarkers of diseases.

Regarding the starting sample, it is now clear that different results must be expected in different biofluids. Currently, serum and plasma are preferred over whole blood because of the effect of the cellular fraction [147]. Using plasma, the poor-platelet form is more stable because it avoids miRNAs released from platelets [148] and seems to be more stable against thaw-fridge cycles and other processing confounders [149]. However, the centrifugation protocol to obtain the poor-platelet fraction and the anticoagulant in the tubes also affects the miRNA profile, as mild contamination by platelet miRNAs can be found, for example, if using EDTA as anticoagulant [150].

On the other hand, there are also some global recommendations for the design of the study. Due to the high variability caused by the aforementioned factors, the need of use a considerable sample size and conducted both discovery and validation phases is of special relevance on this field [146]. Also gain special relevance the patient's characteristics, regarding on going medication and potential confounding factors [146,148]. The quantification technique must be selected considering the expected differences among

platforms [138] and it is always recommended to validated by qPCR at least the most relevant results.

Finally, during the analysis of the results, the most critical step is the normalization strategy[151]. Several authors have pointed out the differences in the results and their reproducibility when using different normalization strategies [152,153]. Thus, it should be exhaustively selected and reported in detail. Also, the correction for remove the inter-plate variability must be critically selected, as well as considered prior so that patients and controls, as well as, for example, subjects of different sex or age, overlapped in plates as far as possible [146]. Finally, it is preferred to report a miRNA profile rather than a single miRNA as biomarker to ensure the specificity.

2.6. Current knowledge on miRNAs and VT

Up to now, only three studies have sought to identify a miRNA profile associated with VT. First, Qin et al. [154] compared patients of Chinese Han ethnicity which underwent orthopedic surgery of knee or hip. After the surgery, they identified 18 patients who developed DVT and 20 who not (controls). They quantified 736 serum miRNAs using microarray in pooled samples and selected 19 miRNAs to be validated in 14 DVT patients and 14 controls by qPCR. They concluded that three miRNAs were differentially expressed in orthopedic-surgery patients who develop an DVT in contrast to those who not. One year later, in 2015, Starikova et al. [155] quantified 742 plasma miRNAs by qPCR in 20 healthy controls and 20 patients with unprovoked VT, and reported a set of nine dysregulated miRNAs. Finally, in 2016, Wang et al. [156] studied 248 subjects, 53 of whom had developed DVT. They performed a discovery phase with pooled samples of 12 DVT patients and 12 controls, measuring 742 miRNAs in plasma by qPCR. They

selected 13 miRNAs to be validated in the whole cohort and concluded that two miRNAs were associated with DVT.

While these previous studies encourage the potential role of miRNAs in VT, they are mainly exploratory and preliminary approaches, not in-depth enough to identified neither dissect the implication of miRNAs in VT. In this regard, some observations could be done about them: (1) the heterogenicity of the VT patients included (surgery, unprovoked VT, different ethnics, etc.), (2) the small sample size (the maximum sample size was 248, and all of them included more miRNAs than patients), (3) pooled samples in discovery step, or without validation step; and (4) lack of concordance among their results.

Given the promising results of miRNAs in other complex diseases, as well as these preliminary approaches in VT, further studies are needed to identify and dissect, if any, the role of miRNAs in VT.

3. Statistical and computational background

3.1. Regression Linear Models and their applications

To answer the biological question 'is there any relationship among these two variables?' we usually describe the joint distribution of the two variables, that is, their covariance [157]. Also, with interpretability purpose, it is usual to report the standardized correlation coefficient, which simply rescale the covariance to the magnitude of x and y, given a r value dimensionless with 0+-1 limits. Thus, given two quantitative variables x and y measured in n individuals, the correlation coefficient between them can be expressed as (1):

$$(1)\ r(x,y) = \frac{Cov\ (x,y)}{\sqrt[2]{Var(x) - Var(y)}}$$

To further explore their conditional expectations, that is, how does y (i.e., the dependent variable) change for a change in x (i.e., the independent variable), a linear model can always serve as a first approximation [158]. For both univariate (i.e., only one predictor x) or multivariate (i.e., considering x1 to xn predictors), we assume that the conditional distribution of y in x can be modeled under a straight-line equation, and for n additive predictors (2):

$$(2)\ \eta = \beta_0 + \beta_1\, x_1 \ +\ \beta_2\, x_2 \ldots \beta_n\, n$$

Consequently, the objective of linear regression is to estimate the β parameters that give the best fit for the joint distribution.

Among the assumptions that we made by applying linear models, one raises particular relevance: both independent and dependent variables are normally distributed. While most of the quantitative biological variables are expected to meet this requirement, it

should be checked. If they do not follow a normal distribution, non-parametric test can be used, or the distribution of the variable can be transformed.

However, while quantitative continuous variables can be transformed to meet a normal distribution if necessary, the disease condition is usually codified as a dichotomic variable (yes/no, 0/1). Thus, the disease variable cannot fit a normal distribution because is not continuous so that, it follows a discrete probability distribution. In this case, it can be applied a generalize linear model, which are generalized options of the regression linear models that allow to work with outcomes of certain distribution using a link function g, such as y = g(p) [158]. For categorical outcomes, the most usual is the logistic linear regression, in which the link function can be for example, probit, logit or inverse logit. Using a logit function, it takes the form (3):

$$(3)\ \eta = g(p) = \ln\left(\frac{p}{1-p}\right) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 \ldots \beta_n n$$

The logistic linear models are easily interpretable as odds ratio (OR), which represent the constant effect of a predictor (x) in the likelihood that the outcome (y) occurs. So that, the OR of x for y outcome can be easily interpret from equation (3), given that the OR for a predictor x is simply (4):

$$(4)\ OR = \exp(\beta_x)$$

When the predictor variable is a quantitative variable, it is easily interpretable as the increase or decrease in the probability of the outcome for the increase of one unit of the predictor x. If the x variable is a categorical one, it supposes the increase or decrease of the probability of the outcome when the variable x takes a given value.

Other common approach in biological science is interpret the linear models for the disease status in term of discrimination ability. Receiver Operating Characteristic (ROC) curves

[159,160], have been widely used for measuring the diagnostic ability of a test, as they represent the probability of a diseased person to have higher values than a non-diseased one. The diagnosis value can be a simple laboratory value, but also the outcome of a linear model including different predictors (the estimated *y* outcome, here named p). The probability density function of the value of predictor (p) can be plotted for diseased and healthy. If the predictor variable is able to completely difference those group of subjects, the distribution curves will not share any value of p. Otherwise, the distribution curves will overlap. Different thresholds in p will led to more diseased considered as healthy or healthy considered diseased. The ROC curve is the graphical representation of the true positive fraction versus the false positive fraction for all the possible values of p. The most common parameter to estimate the accuracy using ROC curves is the AUC, which not depends on the threshold of p value, given that all possible p values are plotted and thus, represents the overall probability of classify correctly any random subject. Although different thresholds may be then assigned attending to the biological use of the discriminatory test, the AUC is a global accuracy measure that allows also to compare different models. Consequently, it is possible to test whether the difference between the accuracy of two ROC curves is significant, either when adding a new variable as predictor or when comparing independent diagnostic tools.

3.2. Family-based studies

To work in a family-based cohort allows to explore the genetic correlations among phenotypes, their segregation patterns or to obtain a better representation of rare mutations related to the disease. In addition, some data-related advantages can be found, such as to avoid stratification bias on the cohort, or to ensure the quality of the genotyping process by checking for Mendelian inheritance. In this Thesis, the analyses regarding the entire GAIT-2 population will be done using the variance component method. The

software SOLAR implements this method for genetics on extended pedigrees and the R package solarius is an R interface for this software. Detailed explanations of both the method and its implementations can be found in Ref. [161,162].

Briefly, the variance component method consists in that the total phenotypic variance (σ_p^2) of a given trait, is decomposed into two components: the genetics variance (σ_g^2), and the environmental variance (σ_e^2). Both them can be further divided, so that genetic component could be divided into additive, dominance and epistatic interactions effects. Also, the environmental variance can be decomposed into the shared environmental effects (modeled mainly by the house-hold term) and the residual not shared environmental effect. We will assume the simplest model attending to the additive genetic effect (σ_a^2) and the environmental variance.

To decompose the phenotypic variance into the genetic and environmental components, the genetic relationship between all the individuals of the population must be known. This relationship is quantified as a kinship coefficient, which assigns a value in the range 0-1 considering the strength of the genetic relationship (e.g., 0 for independent individuals and 1 for twins). The kinship coefficient of each pair of individuals in the population of *n* subjects are collected into a *n x n* kinship matrix (Φ). Thus, once known the structure of the pedigrees and the observed variance of a trait of interest (y), it can be modeled using a linear mixed model, which is an extension of the generalized linear models that includes as predictors both fixed and random effects. In this way, for an observed matrix of the fixed variable *X* and its fixed effect β, and a matrix of random variables *Z* and its random effect *u*, it takes the general form (1):

$$(1) y = X\beta + Zu$$

Considering the previous definition, the general form (1) can be applied for the observed covariance among family members for a given trait (Ω), using as predictors the fixed effect of covariates, the random effect due to the additive genetic effect modeled with the additive genetic covariance multiplied by twice the kinship matrix (2Φ) and the residual or environmental effect, defined as the identity matrix (I) multiplied by the environmental covariance, so that (2):

$$(2)\ \Omega = X\beta + 2\Phi\sigma_a^2 + \mathrm{I}\,\sigma_e^2$$

Once the phenotypic variance of a given trait is divided into genetic and environmental components, the heritability (h^2) of the trait can be estimated as (3):

$$(3)\ h^2 = \frac{\sigma_a^2}{\sigma_p^2}$$

Therefore, as shown in (3), the heritability can be defined as the proportion of the phenotypic variance of a trait that can be attributable to the additive genetic factors.

To test for correlations among traits, it is assumed that each trait can be modeled through the univariate model in (2). So that, the correlations among two traits (that is, their shared covariance) can be modeled using a multivariate model of the same form. Thus, it allows to decompose the covariance between two traits also into the genetic and environmental components, that is, to obtain their genetic correlation coefficient (ρ_g) and environmental correlation coefficient (ρ_e). From both them and considering also the heritability of each trait (h^2_1 and h^2_2 for traits 1 and 2 respectively), the phenotypic correlation coefficient (ρ_p) can be calculated given the formula (4):

$$(4)\ \rho_p = \sqrt{h_1^2}\sqrt{h_2^2}\rho_g + \sqrt{1 - h_1^2}\sqrt{1 - h_2^2}\rho_e$$

3.3. Feature selection: penalized regression

Integrate different molecular phenotypes is a promising strategy to understand the interactions and dynamics among different biological layers. However, it implies an statistical challenge not only because of the high amount of data, but also considering, for example, the different distribution of missing values across biological layers, the expected collinearity among variables or the interpretability of the results in a biological context [163]. To address this integration, most common approaches are based on classify, cluster, or select biological variables given their suitability to represent a desired outcome (for example, health or illness or different type of tumors). In this Thesis, the integration of different intermediate phenotypes will be done using feature selection through penalized regression. Therefore, other analyses will not be further explained, although reviews can be found in Ref. [164,165].

Feature selection methods [166] attempt to identify and select the variables that gives more information for a given outcome (for example, the disease status). Using penalized regression, a multivariate linear regression model is applied, but considering a shrinkage factor. One of the main advantages of penalized regression over other feature selection techniques, is that it avoids the overfitting of the models. There are different methods for penalized regression, and the most common are ridge regression and Least Absolute Shrinkage and Selection Operator (LASSO) regression [167–169]. Both them are based on considering a shrinkage factor to make the coefficients as small as possible without compromising the fit of the model. The main difference between them is that ridge regression does not zeroed any coefficient, so it might be used when all variables are expected to be informative and in a similar weight. Otherwise, when no assumption about the informative capability of the variable is done, LASSO regression is most suitable.

Briefly, as usually done for linear models, the coefficients (β) are estimated using the ordinary least squares (OLS) approach, which consists in using a loss function to minimize the sum of the square of the error between the observed (y) and the estimated values ($\hat{y}$), which therefore (1):

$$(1)\ L_{OLS}\left(\hat{\beta}\right) = \sum_{i=1}^{m}(y - \hat{y_i}) = \sum_{i=1}^{m}(yi - \hat{\beta}X)$$

In penalized regression, a second term is introduced into the general form (1), so that the loss function minimizes also the β coefficients. In LASSO, this penalty term is the value of the absolute sum of the regression coefficients multiplied by a parameter lambda (λ). Thus, it takes the form (2):

$$(2)\ L_{LASSO}\left(\hat{\beta}\right) = \sum_{i=1}^{n}(yi - \hat{\beta}X) + \lambda\sum_{j=1}^{m}\left|\hat{\beta}_j\right|$$

Therefore, a critical step to ensure the robustness of the results is the estimation of the λ parameter. As shown in (2), λ=0 implies that the correlation coefficients are not modified and the bigger λ, the smaller the coefficients. There are different implementations of LASSO algorithm for feature selection, such as R packages mlr[170] or glmnet. These packages allow also to estimate the best value for λ. A common approach is to test and train the model under a cross-validation scheme, testing different values of λ in each iteration. Then, the λ value that minimizes the mean cross-validation error can be selected. Also, it can be selected the λ value that corresponded to the mean error plus one standard error, which therefore will make more coefficients to zero (because the λ is bigger), so that although the error is mild higher, the model is simpler.

Some consideration must be done when using LASSO, from both statistical and biological points of view. First, it makes no sense to include hundreds or thousands of

variables as predictors. Thus, if needed, a pre-selection step must be done, which could be based on biological knowledge, in statistical process (for example, performing a clustering or a classification analysis before), or both them. Also, given that the shrinking attends to the sum of coefficients, the scale of the variables must be comparable so that their coefficients are also. Finally, it should be considered that if the information about the outcome provided by two closely related features is similar, LASSO will keep only one of them.

2

OBJECTIVES

1. Objectives

The principal objective of this Thesis was to identify and study mechanisms of regulation of the gene expression through miRNAs underlying the development of VT.

In order to do that, the main hypothesis and their subsequent specific objectives were:

Hypothesis (1) miRNAs could be involved in the pathogenic processes underlying VT and circulating miRNAs may be suitable as biomarkers for VT.

Objective (1): To analyze the plasma miRNA signature in patients with VT, in contrast to non-VT controls.

1.1. To identify, if any, a miRNA profile differentially expressed in VT patients.

1.2. To determine their suitability as biomarker for VT risk.

1.3. To dissect their relationship with other intermediate phenotypes.

1.4. To stablish, as far as possible, their potential targets or mechanism of action.

Hypothesis (2): miRNAs have an intermediate role between genetics, gene expression and clinical phenotypes. The integrated analysis of these layers could help to reach deeper knowledge about each single layer, as well as their interactions.

Objective (2): To integrate miRNAs with other intermediate phenotypes in VT patients in contrast to non-VT controls, in order to reach novel knowledge regarding VT biological signature.

2.1. To identify genes and clinical phenotypes which expression in blood may be of use in VT risk assessment.

2.2. To identify interactions among genes, miRNAs and clinical phenotypes in the biological context of VT.

2.3. To characterize, as far as possible, the biological basis underlying the blood signature of VT.

Hypothesis (3): There is still a lack of knowledge in the factors affecting miRNA expression and quantification. To dissect miRNA expression data in relation to biological factors could help to increase our understanding of the biology surrounding circulating miRNAs, as well as their clinical utility.

Objective (3): To dissect biological and technical features affecting the expression of circulating miRNAs.

3.1. To dissect the effect of technical covariates in the quantification of circulating miRNAs.

3.2. To identify biological factors affecting the expression of miRNAs.

3.3. To identify and characterize genetic factors that regulate the expression of miRNAs

3

RESULTS

1. Informe del Director de Tesis

José Manuel Soria Fernández y Sonia López Moreno, como Directores de la Tesis Doctoral titulada 'Study of the Regulatory Mechanisms of Gene Expression in Venous Thromboembolic Disease: microRNAs' presentada por Alba Rodríguez Rius, declaramos que todos los resultados recogidos en esta Tesis son resultados originales que se han obtenido durante la realización de la presente Tesis.

Esta Tesis se presenta como compendio de dos artículos publicados y un manuscrito en preparación. En todos ellos, la doctoranda es primera autora y ninguno se presenta en régimen de coautoría. A continuación, detallamos la clasificación de las revistas donde se han publicado, de acuerdo a *Web of Science – Journal Citation Reports*, así como la contribución de la doctoranda a cada publicación:

1. **'Identification of a Plasma microRNA Profile Associated with Venous Thrombosis'.**

Artículo publicado en *Arteriosclerosis, Thrombosis and Vascular Biology*.

Factor de Impacto (2019): 6.6. Mejor clasificación: Q1 en *Peripheral Vascular Disease*.

La doctoranda ha realizado el trabajo de laboratorio necesario para la cuantificación de microRNAs plasmáticos y ha llevado a cabo los análisis estadísticos. A su vez, ha interpretado y discutido los resultados y redactado el manuscrito.

Los resultados recogidos en este artículo han dado lugar a la solicitud de una **Patente Europea, con número de referencia 20382151.7-1118, titulada 'MicroRNAs Markers of Thrombosis Conditions'**, propiedad de Fundació Insititut de Recerca de

l'Hospital de la Santa Creu i Sant Pau. Inventores: José Manuel Soria Fernández, Alba Rodríguez Rius y Juan Carlos Souto Andrés.

2. **'Whole Blood Gene Expression in Venous Thrombosis: An Integrative Analysis with Clinical Phenotypes and MicroRNAs'.**

Manuscrito en preparación.

Este artículo utiliza los datos de microRNAs obtenidos por la doctoranda en el artículo anterior, así como otras variables cuantificadas previamente. La doctoranda ha diseñado y llevado a cabo los análisis estadísticos, ha interpretado y discutido los resultados y ha escrito el manuscrito.

3. **'Expression of microRNAs in human platelet-poor plasma: analysis of the factors affecting their expression and association with proximal genetic variants'.**

Artículo publicado en *Epigenetics*.

Factor de Impacto (2019): 4.25. Mejor clasificación: Q1 en *Genetics and Heredity*.

Este artículo utiliza los datos de microRNAs obtenidos por la doctoranda en el primer artículo. La doctoranda ha diseñado y realizado todos los análisis estadísticos, ha interpretado y discutido los resultados y ha redactado el manuscrito.

2. Article 1

Identification of a Plasma microRNA Profile Associated with Venous Thrombosis

Arteriosclerosis, Thrombosis, and Vascular Biology

CLINICAL AND POPULATION STUDIES

Identification of a Plasma MicroRNA Profile Associated With Venous Thrombosis

Alba Rodriguez-Rius, Sonia Lopez, Angel Martinez-Perez, Juan Carlos Souto, Jose Manuel Soria

OBJECTIVE: Venous thrombosis (VT) is a complex condition with a highly heritable genetic component that predisposes one to its development. Certain microRNAs (miRNAs) might be used as biomarkers of VT, but few studies have examined miRNA expression in this respect. The aim of the present work was to identify a plasma miRNA profile associated with VT.

APPROACH AND RESULTS: miRNAs were analyzed by quantitative polymerase chain reaction in plasma samples from members of the GAIT-2 (Genetic Analysis of Idiopathic Thrombophilia 2) population (n=935). A discovery phase involving the screening of 752 miRNAs from a subset of 104 GAIT-2 subjects was followed by an internal validation phase in which the selected miRNAs were quantified in the whole GAIT-2 population. In the discovery phase, 16 miRNAs were selected, including 9 associated with VT and 7 that correlated with an intermediate phenotype of VT. In the next phase, 4 miRNAs were validated as differentially expressed (false discovery rate, <0.1) in VT: hsa-miR-126-3p, hsa-miR-885-5p, hsa-miR-194-5p, and hsa-miR-192-5p. The 4 miRNAs each returned a significant (P<0.05) odds ratio for VT (range of 1.3–1.8). A risk model including the 4 miRNAs, age, and sex returned an area under the receiver operating characteristic curve of 0.77. Moreover, all 4 miRNAs showed significant correlations with intermediate phenotypes of VT (eg, protein S and factor VII). The targets of the miRNAs in the blood coagulation pathway and their interactions are also discussed.

CONCLUSIONS: The present results suggest a 4-miRNA plasma profile associated with VT is of potential use in predicting the risk of this condition.

VISUAL OVERVIEW: An online visual overview is available for this article.

Key Words: biomarkers ■ epigenomics ■ microRNAs ■ risk ■ thrombosis

Venous thrombosis (VT) is a complex condition with a prevalence of 1 to 2 events per 1000 person-years; it is associated with high rates of morbidity and mortality.[1] The risk of VT depends on environmental and genetic factors and on their interactions.[2] The heritability of VT (ie, the proportion of the variance in the liability of developing the condition attributable to genetic factors) has been estimated at around 60%,[3,4] but the genetic variants known to be involved explain only some 15% of this variance.[5] Genetic and epigenetic studies are, therefore, needed to determine and characterize this part of the risk of developing VT, to discover biomarkers of the condition, and to identify their clinical applications.

MicroRNAs (miRNAs) are small noncoding RNAs that regulate gene expression at complementary mRNA targets via posttranscriptional inhibition or target degradation.[6] Circulating miRNAs have been extensively studied as biomarkers for the diagnosis/prognosis of many diseases, including other complex disorders with a highly heritable component.[7,8] Circulating miRNAs might, therefore, be reliable biomarkers for use in predicting the risk of VT.

To our knowledge, only 3 studies have quantified miRNAs in populations of subjects with VT in attempts to describe a VT-associated miRNA profile. Qin et al[9] compared 18 patients of Chinese Han ethnicity with postoperative (orthopedic) deep vein thrombosis to 20 controls and suggested 3 serum miRNAs to be predictors of this condition. Starikova et al[10] later quantified plasma miRNAs in 20 healthy controls and 20 patients with unprovoked

Correspondence to: Alba Rodriguez-Rius, MSc, IIB Sant Pau, Genomics of Complex Diseases, Carrer de Sant Quint , 77, 08041 Barcelona, Spain. Email arodriguezri@santpau.cat

The Data Supplement is available with this article at https://www.ahajournals.org/doi/suppl/10.1161/ATVBAHA.120.314092.

For Sources of Funding and Disclosures, see page 1398.

Arterioscler Thromb Vasc Biol is available at www.ahajournals.org/journal/atvb

Nonstandard Abbreviations and Acronyms

AUC	area under the curve
FDR	false discovery rate
GAIT-2	Genetic Analysis of Idiopathic Thrombophilia 2
miRNA	microRNA
VT	venous thrombosis

Highlights

- A total of 752 microRNAs were analyzed by quantitative polymerase chain reaction in the Genetic Analysis of Idiopathic Thrombophilia population (n=935) following a discovery and internal validation approach.
- Four plasma microRNAs are associated with venous thrombosis and with related intermediate phenotypes (eg, protein S or factor VII).
- This microRNA profile is of use for predicting the risk of venous thrombosis; a model including the microRNAs, age, and sex returned an area under the curve of 0.77.

VT and reported a set of 9 dysregulated plasma miRNAs. Finally, Wang et al[11] studied 248 subjects, 53 of whom had developed VT, and reported 2 miRNAs to be associated with the condition. Other studies have focused on related topics, such as miRNAs as markers of pulmonary embolism,[12] recurrent VT,[13] and in a rat model of VT.[14]

Despite the importance of these exploratory studies, their small sample sizes and the lack of agreement between the results shows that larger and more in-depth studies are required if the role of miRNAs in VT, and their contribution to the genetic risk of this condition, is to be understood.

The aims of the present work were to identify plasma miRNAs associated with VT in subjects belonging to the GAIT-2 (Genetic Analysis of Idiopathic Thrombophilia 2; n=935) and to determine how they might affect cell function leading to VT.

MATERIALS AND METHODS

The data that support the findings of this study are available from the corresponding author upon reasonable request. For detailed information about the laboratory reagents, please see the Major Resources Table in the Data Supplement.

Population and Study Scheme

This work involved the population taking part in the GAIT-2 study—an exploration of the genetics of thrombosis involving 35 extended Spanish families. The recruitment criteria, ethics statements, and general characteristics of the population have been described elsewhere.[15] The study was performed according to the Declaration of Helsinki and reviewed and approved by the Institutional Review Board of the Hospital de la Santa Creu i Sant Pau (Barcelona, Spain). All participants gave written informed consent for themselves and for their minor children. Briefly, families were recruited via a proband who experienced an episode of idiopathic thrombosis (ie, excluding those with antithrombin deficiency or who had the factor V Leiden mutation). A detailed description of the study and the recruitment criteria is given in Methods in the Data Supplement.

A 2-step process was designed to optimize the identification of VT-related miRNAs. In the first—the discovery—phase, 752 miRNAs were screened for in a subset of 104 subjects (52 with VT [excluding paraneoplastic VT] and 52 genetically unrelated, sex- and age-matched [±5 years] control subjects with no VT) to identify miRNAs of interest, that is, those differentially expressed in VT and which correlated with known intermediate phenotypes of VT. Table I in the Data Supplement shows the main characteristics of the discovery phase subjects. This was followed by an internal validation phase to quantify the selected miRNAs in the plasma of the whole GAIT-2 population (n=935). Though described before,[15] Table I in the Data Supplement also summarizes the characteristics of this complete population as whole.

Descriptions of, and methodologies for measuring, the analytes involved in intermediate VT phenotypes, and the genotyping process used in the GAIT-2 project, have been extensively described elsewhere.[3,15]

miRNA Quantification

Before the collection of the blood samples, the patients suspended the treatments with oral anticoagulants and antiplatelet drugs (at least 15 days before the extraction), as well as heparin treatments (at least 24 hours before). Platelet-poor plasma was collected in citrated tubes, centrifugated (2000*g* for 25 minutes at room temperature), and frozen at −80°C until use.

miRNAs were extracted from the plasma using the miRCURY RNA Isolation Kit-Biofluids (Exiqon) and frozen at −80°C. Before each study phase, the extracted miRNAs were reverse transcribed using the miRCURY LNA Universal RT microRNA PCR kit (Exiqon). In both phases, synthetic spike-in (UniSp3, UniSp4, UniSp5, UniSp6, and UniSp3), hemolysis,[16] and nontemplate controls were included.

Before the discovery phase, quality control was performed using a QC panel (Exiqon) including 5 miRNAs and 3 spike-ins to ensure the quality of the screening process. In the discovery phase, 752 miRNAs were quantified using the miRCURY LNA miRNA miRNOME PCR Panel-Human I+II (Exiqon), in a LightCycler 480 Real-Time PCR (polymerase chain reaction) System (Roche). In the internal validation phase, the miRNAs of interest were quantified using Pick&Mix custom panels (Exiqon) of 384 wells and amplification performed in a HT-7900 Fast qPCR (quantitative-PCR) system (Applied Biosystems). In accordance with the manufacturer's protocol, the Ct values above 37 were considered as undetectable.

The miRNA Ct values were corrected according to the manufacturer's protocol using a plate calibrator factor to avoid qPCR interplate differences and an interindividual normalization factor. In the discovery phase, these factors were, respectively, UniSp3 and the mean expression level for the miRNAs expressed in all individuals. In the internal validation phase, the

factors were the mean Ct value per qPCR plate and the mean expression level of the miRNA expressed in at least 90% of individuals. Therefore, unlike the raw Ct values, after interplate calibration and normalization, the final dCq values are directly proportional to the expression of miRNAs. Given that we used a different interplate calibrator for each of the 2 phases, we performed additional analyses to ensure that similar results would be obtained if the same calibrator was used (Table II in the Data Supplement).

Statistical and Bioinformatic Analyses

In the discovery phase, principal component analysis was performed for those miRNAs that were expressed in 90% of subjects. Since this phase included unrelated subjects, associations between the miRNAs and VT plus its intermediate phenotypes were calculated by simple linear regression. Significance was set at $P<0.05$.

In the internal validation phase, the population included 35 families with extended pedigrees; heritability and correlation analyses were, therefore, performed, adjusting for family structure and an ascertainment correction, using the Solarius package[17] as the R interface to SOLAR.[18] Age, sex, and age-squared were tested as covariates affecting miRNA expression and taken into account in correlations with phenotypes when significant. In correlations with both VT and intermediate phenotypes, the focus was on the phenotypic component of the correlation. Multiple testing correction was performed using the qvalue package.[19] Significance was set at a false discovery rate (FDR) <0.1.

Odds ratios were determined using a logistic linear regression model for each miRNA independently. Univariate models included each miRNA expression as a predictor of VT. Multivariate models included miRNA expression, age, sex, and body mass index as additive predictors of VT. The predictive power of the model (ie, to discriminate between cases and controls) that included all the miRNAs of interest together was examined using receiver operating characteristic curves, following a logistic linear model. Significance was determined using the DeLong test. Three models were tested for VT case discrimination: (1) the miRNAs; (2) the miRNAs, age, and sex; and (3) the miRNAs, age, sex, and the 7 genetic variants included in the Thrombo InCode kit[5] (AB0/A1 allele, rs1801020, rs2232698, rs121909548, rs6025, rs5985, and rs1799963). To ensure that the accuracy of the model was not due to the subjects involved in both phases, the subjects including in the discovery phase were removed. Then, the remaining subjects were under-resampled (attending to the VT condition),[20] and the accuracy was estimated with 1000 interactions of bootstrap resampling, using R package mlr.[21] Target predictions were searched for and downloaded for each miRNA using TargetScan v.7.2 software[22] (http://www.targetscan.org/vert_72/; accessed May 6, 2019). Validated miRNA-target interactions were consulted using miRTarBase software v.8[23] (downloaded from http://mirtarbase.cuhk.edu.cn/php/index.php; Homo sapiens file, accessed January 22, 2020). Genes annotated in the gene ontology classification[24,25] blood coagulation pathway (GO:0007596) were downloaded using amiGO[26] (http://amigo.geneontology.org/amigo; accessed May 6, 2019). Network representation was undertaken including the interactions held in the STRING v.11[27] database and represented graphically using Cytoscape software.[28] All statistical and bioinformatic analyses were performed using R software v.3.6.0 (R Foundation for Statistical Computing; http://www.R-project.org).

RESULTS

Discovery Phase

Of the 752 miRNAs quantified in the 104 subjects of the discovery phase, 582 were expressed in at least 1 individual, and 40 were expressed in all of them. After normalization of the data, the 103 miRNAs expressed in at least 90% of the individuals were selected for further analysis. The number of nondetected assays for each of the 103 miRNAs is shown in Table III in the Data Supplement. In principal component analysis, the first 4 principal components explained the 45.6% of the variance of the expression of all of these 103 miRNAs (Figure I in the Data Supplement). Their final residual values were, therefore, calculated by eliminating this explained variance.

Table 1 shows that 9 miRNAs were significantly associated with VT: hsa-miR-23b-3p, hsa-miR-27a-3p, hsa-miR-548c-5p, hsa-miR-221-3p, hsa-miR-197-3p,

Table 1. Set of 16 miRNAs Selected in the Discovery Phase (Screening 752 miRNAs in 104 Subjects)

miRNA	VT Association: β* (*P* Value)	Associated Intermediate Phenotypes†
hsa-miR-192-5p	0.58 (1.9×10^{-03})	Thrombin time
hsa-miR-885-5p	0.58 (2.7×10^{-03})	von Willebrand factor
hsa-miR-23b-3p	−0.49 (8.4×10^{-02})	Factor VIII
		Total protein S
hsa-miR-27a-3p	−0.44 (1.9×10^{-02})	Thrombin time
hsa-miR-194-5p	0.40 (3.8×10^{-02})	Thrombin time
hsa-miR-221-3p	−0.39 (3.9×10^{-02})	Factor VIII
		von Willebrand factor
hsa-miR-197-3p	−0.39 (4.0×10^{-02})	Factor VIII
hsa-miR-548c-5p	−0.39 (4.6×10^{-02})	Factor VIII
		von Willebrand factor
hsa-miR-320a	−0.37 (4.9×10^{-02})	Clot formation rate
hsa-miR-142-3p	−0.13 (0.48, NS)	Factor XII
hsa-miR-146a-5p	−0.31 (0.09, NS)	Prothrombin time
		Fibrinogen
hsa-miR-148a-3p	0.04 (0.82, NS)	Average RNA by platelet
hsa-miR-28-3p	0.07 (0.69, NS)	Thrombin generation–peak of thrombin
hsa-miR-320b	−0.06 (0.71, NS)	Functional protein S
hsa-miR-342-3p	−0.16 (0.37, NS)	Factor VIII
		von Willebrand factor
		Fibrinogen
hsa-miR-126-3p	0.34 (0.06, NS)	Factor VII

miRNA indicates microRNA; NS, nonsignificant; and VT, venous thrombosis.
*Correlation coefficient for the linear association.
†Main intermediate phenotypes related to VT that resulted in a significant ($P<0.05$) linear correlation with the miRNA.

and hsa-miR-320a were downregulated and hsa-miR-194-5p, hsa-miR-192-5p, and hsa-miR-885-5p were upregulated. These 9 miRNAs were selected for the internal validation phase. Correlations with intermediate VT phenotypes were then explored and 7 additional miRNAs selected: hsa-miR-320b, hsa-miR-342-3p, hsa-miR-142-3p, hsa-miR-28-3p, hsa-miR-148a-3p, hsa-miR-146a-5p, and hsa-miR-126-3p. Thus, 16 miRNAs were selected for the internal validation phase (Table 1). An extended report of these associations is given in Table IV in the Data Supplement.

Internal Validation Phase

Heritability of miRNA Expression

The heritability of the expression of each miRNA (ie, the proportion of the variance attributable to genetic factors) was calculated, with age, sex, and age-squared introduced as covariates. Only 2 miRNAs showed nonsignificant heritability (hsa-miR-197-3p and hsa-miR-148a-3p). The heritability of the remaining miRNAs varied from 0.1 to 0.38 ($P<0.05$). Table V in the Data Supplement summarizes the heritabilities and significant covariates for each miRNA.

Differential Expression in VT

The differential expression of the miRNAs in VT was next examined. Table 2 shows the 4 miRNAs returning a significant, positive correlation with the condition (FDR<0.1). Three of these miRNAs (hsa-miR-885-5p, hsa-miR-192-5p, and hsa-miR-194-5p) returned such correlations in both study phases, while the fourth (hsa-miR-126-3p) correlated with an intermediate phenotype in the discovery phase and with VT in the validation phase.

Risk Prediction

To further explore the role of these 4 miRNAs in the risk of VT, odds ratio analysis was performed for each; all returned significant results (Table 3). Since age, sex, and body mass index influence the appearance of VT and are significant covariates affecting miRNA expression, the odds ratio values were calculated again taking these variables into account: they remained significant (Table 3).

Since the 4 miRNAs showed independent predictive power with respect to VT, a preliminary risk model was constructed to analyze their potential use as a combined set. This risk model returned an area under the curve (AUC) of 0.66 ([95% CI, 0.59–0.74] sensitivity, 85.7%; specificity, 41.1%)—significant compared with the random model AUC of 0.5 ($P=2.19\times10^{-05}$). Moreover, a further risk model including the 4 miRNAs, age, and sex (Figure 1) returned an AUC of 0.77 ([95% CI, 0.71–0.82] sensitivity, 85.7%; specificity, 54.5%). Given that the 104 subjects of the discovery phase were included also in the internal validation phase, this risk model was tested also excluding those 104 subjects, and similar accuracy measures were obtained (AUC, 0.79; sensitivity, 82%; specificity, 45.9%).

The effect of adding the 7 known genetic risk variants included in the Thrombo InCode diagnostic kit (which currently provides the most accurate genetic risk score for VT[5]) was also examined. This resulted in an AUC of 0.80 ([95% CI, 0.74–0.85] sensitivity, 74.5%; specificity, 72.0%; no significant improvement [$P>0.05$]). This model including the miRNAs and the genetic variants was compared also with the accuracy of the model including only the genetic variants (AUC, 0.61 [95% CI, 0.56–0.68]), and in this case, the improvement was significant ($P=1.87\times10^{-05}$).

Associations With Intermediate Phenotypes

The biological implications of the expression of the 4 miRNAs with respect to VT were next examined. Table 4 shows the significant correlations detected with intermediate phenotypes (FDR<0.1). Notably, all of the miRNAs returned at least 1 significant correlation, and 3 correlated with factor VII concentration.

Calculations were also made to determine whether any of the miRNAs were differentially related to an intermediate phenotype in subjects with VT and in VT-free controls. The interaction *hsa-miR-194-5p x VT* significantly influenced erythrocyte folate ($P=3.1\times10^{-04}$; FDR, 0.09); *hsa-miR-885-5p x VT* significantly influenced the fibrinogen level ($P=1.6\times10^{-03}$; FDR, 0.19); and

Table 2. Differentially Expressed miRNAs in VT

	Discovery Phase (n=104)		Internal Validation Phase (n=935)			
miRNA	β*	*P* Value	n†	ρ_{phe}‡	*P* Value	FDR
hsa-miR-192-5p	0.58	1.9×10^{-03}	749	0.24	2.72×10^{-04}	7.8×10^{-03}
hsa-miR-194-5p	0.40	3.8×10^{-02}	725	0.21	1.22×10^{-03}	2.3×10^{-02}
hsa-miR-885-5p	0.58	2.7×10^{-03}	576	0.19	6.26×10^{-03}	8.1×10^{-02}
hsa-miR-126-3p	…	>0.05	818	0.18	5.94×10^{-03}	7.81×10^{-02}

Associations between VT and miRNA expression in both experimental phases for those miRNAs that showed a positive association (FDR<0.1) in the internal validation phase. FDR indicates false discovery rate; miRNA, microRNA; and VT, venous thrombosis.

*Correlation coefficient of the simple linear model.

†Number of individuals who expressed the miRNA in the internal validation phase.

‡Correlation coefficient of the phenotypic component of the linear model including the family structure.

Table 3. OR for Each miRNA With Respect to VT

	Model 1*			Model 2†		
miRNA (dCq)	OR	95% CI	*P* Value	OR	95% CI	*P* Value
hsa-miR-885-5p	1.33	1.15–1.60	1.7×10^{-03}	1.29	1.07–1.56	9×10^{-03}
hsa-miR-194-5p	1.37	1.11–1.70	4×10^{-03}	1.39	1.12–1.75	3.6×10^{-03}
hsa-miR-192-5p	1.31	1.08–1.59	5.7×10^{-03}	1.46	1.20–1.79	2.5×10^{-04}
hsa-miR-126-3p	2.12	1.41–3.22	3.3×10^{-04}	1.72	1.11–2.67	0.01

BMI indicates body mass index; miRNA, microRNA; OR, odds ratio; and VT, venous thrombosis.
*Model 1: univariate model testing only the miRNA as a predictor of VT.
†Model 2: multivariate model, adding age, sex, and BMI as covariates to the logistic linear model.

hsa-miR-126-3p x VT significantly influenced prothrombin time ($P=9.9\times10^{-07}$; FDR, 0.19×10^{-03}), anti-cardiolipin antibody IgG ($P=2.9\times10^{-05}$; FDR, 0.02), and anti-β 2-glycoprotein I antibody isotype (IgG subclass; $P=1.8\times10^{-04}$; FDR, 0.07). The interaction *hsa-miR-192-5p x VT* had no influence on any intermediate phenotype.

Target Prediction and Network Modeling

The network of biochemical interactions surrounding these 4 miRNAs in VT was explored by identifying their putative targets and the interactions between these targets. For this, genes annotated in the gene ontology blood coagulation pathway were examined to see whether they had been validated as, or were predicted to be, targets of miRNAs. Table 5 summarizes the results. Notably, all of the genes involved in the blood coagulation pathway had either been validated as, or were predicted to be, the targets of miRNAs. The complete network of these targets represented graphically with the intermediate phenotypes and with the interactions in the STRING[27] database (Figure 2).

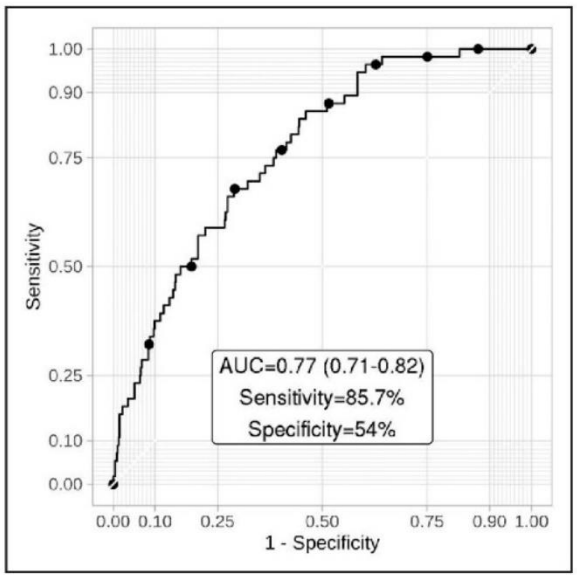


Figure 1. Receiver operating characteristic curve for venous thrombosis outcome.
Risk model including the 4 microRNAs, age, and sex. AUC indicates area under the curve.

DISCUSSION

This analysis of the association between miRNA production and VT involves the largest sample to date. The discovery phase identified 16 miRNAs of interest, and the internal validation phase showed 4 of these to be differentially expressed in patients with VT: hsa-miR-126-3p, hsa-miR-885-5p, hsa-miR-194-5p, and hsa-miR-192-5p. Among the 12 miRNAs not validated in the latter phase, 6 had been previously associated with VT or with related traits: hsa-miR-320a,[10,12] hsa-miR-320b,[10] hsa-miR-197-3p,[13] hsa-miR-28-3p,[12] hsa-miR-146a,[9] and hsa-miR-23b-3p.[14] Although in the present internal validation phase these miRNAs were not associated with VT for an FDR <0.1, it is remarkable that different studies have suggested the involvement of these miRNAs in this condition. Further studies that focus on different tissues or patient characteristics could help to confirm whether these miRNAs are truly involved.

hsa-miR-126-3p was earlier reported to be dysregulated in VT in a screening phase that included plasma pools of 12 deep vein thrombosis patients and 12 controls, although it was not validated for the entire study cohort.[11] A previous study also reported the involvement of hsa-miR-126 in the proangiogenic activity of endothelial progenitor cells, which is required for the resolution of venous thrombi.[29] The authors indicated this miRNA to target PIK3R2 (phosphatidylinositol 3-kinase regulatory subunit beta), interfering with the PI3K/Akt (phosphatidylinositol 3-kinase/protein kinase B) signal pathway. These findings from both population-based and in vitro studies support the involvement of hsa-miR-126-3p in VT, although in the present work it was predicted to target the fewest mRNAs involved in the blood coagulation pathway. However, its mechanism of action in VT might involve targets of other related pathways.

No previous study has associated the expression of hsa-miR-885-5p with VT or any other cardiovascular or circulatory condition (although it has been associated with different types of cancer,[30,31] liver disease,[32] and preeclampsia[33]). However, a study in colorectal cancer cell lines has recently shown this miRNA to inhibit the expression of von Willebrand factor and IGFBP5 (insulin-like growth factor-binding protein 5) in these cells.[34] Interestingly, the present discovery phase results suggest the expression of this miRNA to correlate with blood

Table 4. Significant Correlations (FDR<0.1) Between the miRNAs Differentially Expressed in Venous Thrombosis and Intermediate Phenotypes of Venous Thrombosis

miRNA	Intermediate Phenotype	ρ_{phe}*	*P* Value	FDR
hsa-miR-885-5p	Thrombin generation–lag time	0.18	2.36×10^{-05}	5.61×10^{-03}
	Lupus anticoagulant antibody	0.17	3.68×10^{-05}	7.15×10^{-03}
	Thrombin generation–thrombin peak	−0.18	3.93×10^{-05}	7.41×10^{-03}
	Factor VII	0.16	1.01×10^{-04}	1.25×10^{-02}
	Protein S free	0.13	1.97×10^{-03}	7.64×10^{-02}
hsa-miR-192-5p	ADAMTS13	0.14	1.94×10^{-04}	1.82×10^{-02}
	Factor VII	0.12	1.46×10^{-03}	6.24×10^{-02}
hsa-miR-126-3p	Thrombin generation test–ETP	0.13	5.58×10^{-04}	3.39×10^{-02}
	Factor XI	0.11	1.93×10^{-03}	7.55×10^{-02}
hsa-miR-194-5p	Factor VII	0.18	1.90×10^{-06}	1.20×10^{-03}
	Thrombin generation test–lag time	0.15	8.06×10^{-05}	1.10×10^{-02}
	Protein S total	0.14	1.55×10^{-04}	1.60×10^{-02}
	Lupus anticoagulant antibody	0.14	4.31×10^{-04}	2.90×10^{-02}

ADAMTS13 indicates a disintegrin and metalloproteinase with a thrombospondin type 1 motif, member 13; ETP, endogenous thrombin potential; FDR, false discovery rate; and miRNA, microRNA.
*Coefficient of the phenotypic correlation.

protein levels of von Willebrand Factor (an intermediate phenotype), although this was not confirmed in the internal validation phase. The present results show a significant correlation between hsa-miR-885-5p and thrombin generation lag time, thrombin peak, and the concentration of protein S (all intermediate phenotypes). Interestingly, our group previously reported *F2* to be the main gene affecting thrombin generation,[15] and both FII (coagulation factor II) and F2RL (coagulation factor II-receptor like) are predicted targets of hsa-miR-885-5p. Moreover, Protein S is a hemostatic factor that affects thrombin generation.[35] In addition, the correlation between hsa-miR-885-5p and fibrinogen was significant in VT but not in controls. Altogether, the expression of this miRNA, the correlated intermediate phenotypes, the significant influence of the interaction *hsa-miR-885-5p x VT* on some of these intermediate phenotypes, and the predicted targets reveal a network suggesting that thrombin generation is regulated by hsa-miR-885-5p in VT.

In an earlier study, hsa-miR-194-5p was reported upregulated in plasma pools from patients with deep vein thrombosis (12 patients compared with 12 controls), although this was not confirmed in the whole cohort.[11] In addition, a study in hepatoma cells reported this miRNA to affect fibrinogen production.[36] hsa-miR-194-5p is encoded at 2 positions in the genome. Interestingly, one of these genes is located just 200 base pairs from the *hsa-miR-192-5p* gene. Their shared regulation or expression might, therefore, be contemplated.

Finally, hsa-miR-192-5p has never before been associated with VT or any related trait or phenotype. Although

Table 5. Validated and Predicted Targets in the Blood Coagulation Pathway for the 4 miRNAs Differentially Expressed in VT: Results for the Genes Annotated in the Gene Ontology Blood Coagulation Pathway

miRNA	Functional Validated Targets*	Predicted Targets
hsa-miR-192-5p	*CLIC1*, *THBD*, *RAB27A*, *PLAU*, *ENPP4*, *LMAN1*, *PPRKAR1A*, *ANO6*, *PABPC4*, *SLC7A11*, *SH2B3*, *UBASH3B*, *CAV1*, *ITGB3*, *H3F3A*, *SERPINE1*	*PRKAR1A*, *PRKG1*, *PRKACB*, *MAPK1*
hsa-miR-194-5p	*RAC1*, *CAPZA1*, *FZD6* (2), *HIST2H3A*, *CAV1*	*PRKAR1A*, *RAP2B*, *THBS1*, *MAPK1*, *GNA13*, *PHF21A*, *CBX5*, *PDGFA*, *GNG2*, *DGKH*, *SHH*, *JMJD1C*, *SLC7A11*
hsa-miR-126-3p	*MERTK*, *PIK3CG* (4), *PIK3R1*	*GNA13*
hsa-miR-885-5p	*RAC1* (6)	*RAC1*, *DGKE*, *IL6*, *CAPZA1*, *CAPZA2*, *GNAS*, *DGKH*, *MAPK1*, *GNA14*, *UBASH3B*, *LMAN1*, *PROCR*, *DOCK11*, *VCL*, *HPS5*, *GNAQ*, *F9*, *MAFK*, *DOCK8*, *GNG2*, *ADRA2A*, *RCOR1*, *PRKCA*, *PTPN11*, *TBXA2R*, *CD59*, *PAFAH2*, *F2RL2*, *DGKG*, *CYP4F11*, *DGKB*, *PIK3CG*, *SLC7A11*, *AP3B1*, *C1GALT1C1*, *TEC*, *ENTPD1*, *HDAC2*, *DGKI*, *SERPING1*, *RAF1*, *F2*, *METAP1*, *TLR4*

miRNA indicates microRNA; and VT, venous thrombosis.
*The number in brackets indicates the number of studies that have validated the binding, when >1.

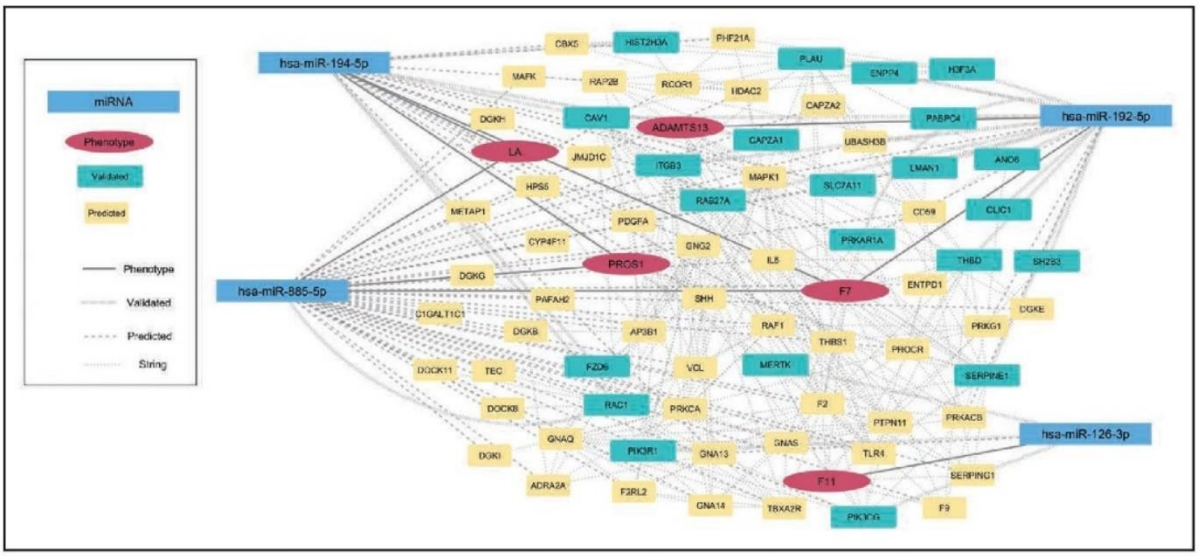


Figure 2. Network representation for the 4 microRNAs (miRNAs) in relation to the blood coagulation pathway.
Representation of the interactions (provided by the STRING database) among the predicted targets, validated targets, and intermediate phenotypes associated with each miRNA.

it has the most validated targets in the blood coagulation pathway, most of these were identified in a single study that examined the effect of p53-inducible miRNAs in the regulation of the cell cycle.[37]

The 4 validated miRNAs not only showed a significant correlation with VT but are also potential predictors of the condition, as shown by the detected significant odds ratios and AUC value of 0.77. This potential predictive power is interesting since the genetic scores available for VT only return a maximum AUC of around 0.67.[5] When the known genetic variants included in the TiC[5] kit were introduced into the risk model, there was no significant improvement in the AUC. However, it should be remembered that the GAIT-2 population is based on probands with idiopathic thrombophilia and their families, so genetic risk variants (eg, factor V Leiden) were excluded. The latter, extended risk model might, therefore, be valuable for use with other populations. Overall, the present findings support the idea that the comprehensive integration of omic data, including miRNAs and other epigenetic features, holds promise as a means of predicting VT and for explaining its missing heritability.[38]

The present work suffers the limitation that the population of the discovery phase was included in the validation phase. However, we have demonstrated that the accuracy of the risk model remains unchanged even when excluding the subjects of the discovery phase. Another limitation is that the expression of miRNAs is tissue and time specific[39]; different results might, therefore, be returned by other tissues (eg, differences between circulating and tissue miRNAs) or different patients (ie, in whom the time elapsed since a VT event is different or in whom the time left to elapse before an event occurs is different). Also, different expression levels could be obtained if other procedures were used to obtain platelet-poor plasma or for the quantification of miRNAs. Finally, the plasma miRNA profile reported to be associated with VT needs to be validated in an independent population, ideally in a longitudinal study, although it should be remembered that the present sample is the largest to have been examined, that the entire project was performed using gold standard technique for measuring miRNAs, and that confounding variables and FDRs were rigorously taken into account.

In conclusion, 4 plasma miRNAs differentially expressed in VT are reported, their potential as predictors of VT discussed, and some light is shed on their possible biological involvement via their correlation with intermediate phenotypes of VT. These findings are not without clinical significance.

ARTICLE INFORMATION

Received December 20, 2019; accepted February 28, 2020.

Affiliations

From the Genomics of Complex Diseases Group, Research Institute Hospital de la Santa Creu i Sant Pau, IIB Sant Pau, Barcelona, Spain (A.R.-R., S.L., A.M.-P., J.M.S.); and Unit of Thrombosis and Hemostasis, Hospital de la Santa Creu i Sant Pau, Barcelona, Spain (J.C.S.).

Sources of Funding

This study was funded by a grant of the Spanish Government (FIS_PI14_00582, J.M. Soria), the Government of Catalonia (predoctoral fellowship FI_2017B_00673, A. Rodriguez-Rius; Consolidate Research Group SGR_1736, J.M. Soria; CERCA [Centres de Recerca de Catalu a] Programme), the Spanish Society of Thrombosis and Hemostasis, the nonprofit association Activa´TT por la Salud, and STAGO.

Disclosures

None.

REFERENCES

1. Beckman MG, Hooper WC, Critchley SE, Ortel TL. Venous thromboembolism: a public health concern. *Am J Prev Med.* 2010;38(4 suppl):S495–S501. doi: 10.1016/j.amepre.2009.12.017
2. Rosendaal FR. Venous thrombosis: a multicausal disease. *Lancet.* 1999;353:1167–1173. doi: 10.1016/s0140-6736(98)10266-0
3. Souto JC, Almasy L, Borrell M, Blanco-Vaca F, Mateo J, Soria JM, Coll I, Felices R, Stone W, Fontcuberta J, et al. Genetic susceptibility to thrombosis and its relationship to physiological risk factors: the GAIT study. Genetic analysis of idiopathic thrombophilia. *Am J Hum Genet.* 2000;67:1452–1459. doi: 10.1086/316903
4. Heit JA, Phelps MA, Ward SA, Slusser JP, Petterson TM, De Andrade M. Familial segregation of venous thromboembolism. *J Thromb Haemost.* 2004; 2:731–736. doi: 10.1111/j.1538-7933.2004.00660.x
5. Soria JM, Morange PE, Vila J, Souto JC, Moyano M, Tr gou t DA, Mateo J, Saut N, Salas E, Elosua R. Multilocus genetic risk scores for venous thromboembolism risk assessment. *J Am Heart Assoc.* 2014;3:e001060. doi: 10.1161/JAHA.114.001060
6. Huntzinger E, Izaurralde E. Gene silencing by microRNAs: contributions of translational repression and mRNA decay. *Nat Rev Genet.* 2011;12:99–110. doi: 10.1038/nrg2936
7. Zampetaki A, Willeit P, Tilling L, Drozdov I, Prokopi M, Renard JM, Mayr A, Weger S, Schett G, Shah A, et al. Prospective study on circulating MicroRNAs and risk of myocardial infarction. *J Am Coll Cardiol.* 2012;60:290–299. doi: 10.1016/j.jacc.2012.03.056
8. Wang GK, Zhu JQ, Zhang JT, Li Q, Li Y, He J, Qin YW, Jing Q. Circulating microRNA: a novel potential biomarker for early diagnosis of acute myocardial infarction in humans. *Eur Heart J.* 2010;31:659–666. doi: 10.1093/eurheartj/ehq013
9. Qin J, Liang H, Shi D, Dai J, Xu Z, Chen D, Chen X, Jiang Q. A panel of microRNAs as a new biomarkers for the detection of deep vein thrombosis. *J Thromb Thrombolysis.* 2015;39:215–221. doi: 10.1007/s11239-014-1131-0
10. Starikova I, Jamaly S, Sorrentino A, Blondal T, Latysheva N, Sovershaev M, Hansen JB. Differential expression of plasma miRNAs in patients with unprovoked venous thromboembolism and healthy control individuals. *Thromb Res.* 2015;136:566–572. doi: 10.1016/j.thromres.2015.07.005
11. Wang X, Sundquist K, Elf JL, Strandberg K, Svensson PJ, Hedelius A, Palmer K, Memon AA, Sundquist J, Z ller B. Diagnostic potential of plasma microRNA signatures in patients with deep-vein thrombosis. *Thromb Haemost.* 2016;116:328–336. doi: 10.1160/TH16-01-0071
12. Zhou X, Wen W, Shan X, Qian J, Li H, Jiang T, Wang W, Cheng W, Wang F, Qi L, et al. MiR-28-3p as a potential plasma marker in diagnosis of pulmonary embolism. *Thromb Res.* 2016;138:91–95. doi: 10.1016/j.thromres.2015.12.006
13. Wang X, Sundquist K, Svensson PJ, Rastkhani H, Palm r K, Memon AA, Sundquist J, Z ller B. Association of recurrent venous thromboembolism and circulating microRNAs. *Clin Epigenetics.* 2019;11:1–12. doi: 10.1186/s13148-019-0627-z
14. Jin QQ, Sun JH, Du QX, Lu XJ, Zhu XY, Fan HL, H lscher C, Wang YY. Integrating microRNA and messenger RNA expression profiles in a rat model of deep vein thrombosis. *Int J Mol Med.* 2017;40:1019–1028. doi: 10.3892/ijmm.2017.3105
15. Martin-Fernandez L, Ziyatdinov A, Carrasco M, Millon JA, Martinez-Perez A, Vilalta N, Brunel H, Font M, Hamsten A, Souto JC, et al. Genetic determinants of thrombin generation and their relation to venous thrombosis: results from the GAIT-2 project. *PLoS One.* 2016;11:e0146922. doi: 10.1371/journal.pone.0146922
16. Blondal T, Jensby Nielsen S, Baker A, Andreasen D, Mouritzen P, Wrang Teilum M, Dahlsveen IK. Assessing sample and miRNA profile quality in serum and plasma or other biofluids. *Methods.* 2013;59:S1–S6. doi: 10.1016/j.ymeth.2012.09.015
17. Ziyatdinov A, Brunel H, Martinez-Perez A, Buil A, Perera A, Soria JM. Genetics and population analysis solarius : an R interface to SOLAR for variance component analysis in pedigrees. *Bioinformatics.* 2016;32:1901–1902.
18. Almasy L, Blangero J. Multipoint quantitative-trait linkage analysis in general pedigrees. *Am J Hum Genet.* 1998;62:1198–1211. doi: 10.1086/301844
19. Storey JD. A direct approach to false discovery rates. *J R Stat Soc Ser B Stat Methodol.* 2002;64:479–498.
20. He H, Garcia EA. Learning from imbalanced data. *Knowl Data Eng IEEE Trans.* 2009;21:1263–1284.
21. Bischl B, Lang M, Kotthoff L, Schiffner J, Richter J, Studerus E, Casalicchio G, Jones ZM. Mlr: Machine learning in R. *J Mach Learn Res.* 2016;17:1–5.
22. Agarwal V, Bell GW, Nam JW, Bartel DP. Predicting effective microRNA target sites in mammalian mRNAs. *Elife.* 2015;4:1–38. doi: 10.7554/eLife.05005
23. Huang HY, Lin YC, Li J, Huang KY, Shrestha S, Hong HC, Tang Y, Chen YG, Jin CN, Yu Y, et al. miRTarBase 2020: updates to the experimentally validated microRNA-target interaction database. *Nucleic Acids Res.* 2020;48(D1):D148–D154. doi: 10.1093/nar/gkz896
24. Carbon S, Douglass E, Dunn N, et al. The gene ontology resource: 20 years and still going strong. *Nucleic Acids Res.* 2019;47: D330–D338.
25. Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, Davis AP, Dolinski K, Dwight SS, Eppig JT, et al. Gene ontology: tool for the unification of biology. The gene ontology consortium. *Nat Genet.* 2000;25:25–29. doi: 10.1038/75556
26. Carbon S, Ireland A, Mungall CJ, Shu S, Marshall B, Lewis S; AmiGO Hub; Web Presence Working Group. AmiGO: online access to ontology and annotation data. *Bioinformatics.* 2009;25:288–289. doi: 10.1093/bioinformatics/btn615
27. Szklarczyk D, Gable AL, Lyon D, Junge A, Wyder S, Huerta-Cepas J, Simonovic M, Doncheva NT, Morris JH, Bork P, et al. STRING v11: protein-protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. *Nucleic Acids Res.* 2019;47(D1):D607–D613. doi: 10.1093/nar/gky1131
28. Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, Amin N, Schwikowski B, Ideker T. Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res.* 2003;13:2498–2504. doi: 10.1101/gr.1239303
29. Meng Q, Wang W, Yu X, Li W, Kong L, Qian A, Li C, Li X. Upregulation of microRNA-126 contributes to endothelial progenitor cell function in deep vein thrombosis via its target PIK3R2. *J Cell Biochem.* 2015;116:1613–1623. doi: 10.1002/jcb.25115
30. Afanasyeva EA, Mestdagh P, Kumps C, Vandesompele J, Ehemann V, Theissen J, Fischer M, Zapatka M, Brors B, Savelyeva L, et al. MicroRNA miR-885-5p targets CDK2 and MCM5, activates p53 and inhibits proliferation and survival. *Cell Death Differ.* 2011;18:974–984. doi: 10.1038/cdd.2010.164
31. Lam CS, Ng L, Chow AK, Wan TM, Yau S, Cheng NS, Wong SK, Man JH, Lo OS, Foo DC, et al. Identification of microRNA 885-5p as a novel regulator of tumor metastasis by targeting CPEB2 in colorectal cancer. *Oncotarget.* 2017;8:26858–26870. doi: 10.18632/oncotarget.15844
32. Gui J, Tian Y, Wen X, Zhang W, Zhang P, Gao J, Run W, Tian L, Jia X, Gao Y. Serum microRNA characterization identifies miR-885-5p as a potential marker for detecting liver pathologies. *Clin Sci (Lond).* 2011;120:183–193. doi: 10.1042/CS20100297
33. Sandrim VC, Luizon MR, Palei AC, Tanus-Santos JE, Cavalli RC. Circulating microRNA expression profiles in pre-eclampsia: evidence of increased miR-885-5p levels. *BJOG.* 2016;123:2120–2128. doi: 10.1111/1471-0528.13903
34. Zhang T, Guo J, Gu J, Wang Z, Wang G, Li H, Wang J. Identifying the key genes and microRNAs in colorectal cancer liver metastasis by bioinformatics analysis and *in vitro* experiments. *Oncol Rep.* 2019;41:279–291. doi: 10.3892/or.2018.6840
35. Ser KM, Rosing J, Hackeng TM. Inhibition of thrombin generation by protein S at low procoagulant stimuli : implications for maintenance of the hemostatic balance Inhibition of thrombin generation by protein S at low procoagulant stimuli: implications for maintenance of the hemostatic. *Blood.* 2011;104:3624–3630.
36. Fort A, Borel C, Migliavacca E, Antonarakis SE, Fish RJ, Neerman-Arbez M. Regulation of fibrinogen production by microRNAs. *Blood.* 2010;116:2608–2615. doi: 10.1182/blood-2010-02-268011
37. Georges SA, Biery MC, Kim SY, Schelter JM, Guo J, Chang AN, Jackson AL, Carleton MO, Linsley PS, Cleary MA, et al. Coordinated regulation of cell cycle transcripts by p53-Inducible microRNAs, miR-192 and miR-215. *Cancer Res.* 2008;68:10105–10112. doi: 10.1158/0008-5472.CAN-08-1846
38. Benincasa G, Costa D, Infante T, Lucchese R, Donatelli F, Napoli C. Interplay between genetics and epigenetics in modulating the risk of venous thromboembolism: a new challenge for personalized therapy. *Thromb Res.* 2019;177:145–153. doi: 10.1016/j.thromres.2019.03.008
39. Liang Y, Ridzon D, Wong L, Chen C. Characterization of microRNA expression profiles in normal human tissues. *BMC Genomics.* 2007;8:166. doi: 10.1186/1471-2164-8-166

3. Article 2

Whole Blood Gene Expression in Venous Thrombosis: An Integrative Analysis with Clinical Phenotypes and MicroRNAs

Manuscript in preparation.

Whole Blood Gene Expression in Venous Thrombosis: An Integrative Analysis with Clinical Phenotypes and MicroRNAs.

Abstract

The integrative analysis of different molecular phenotypes could help to increase our knowledge about venous thrombosis (VT). In this study, we aimed to identify genes associated with VT and to dissect their interactions with other molecular phenotypes. We studied 102 subjects of the GAIT-2 ('Genetic Analysis of Idiopathic Thrombophilia') project. We integrated 14 clinical VT phenotypes, 260 genes of the blood coagulation pathway, and 4 microRNAs related to VT in linear predictive models for VT. The phenotypes in the most accurate models were explored. First, 49 significant (False Discovery Rate <0.1) correlations including 11 clinical phenotypes and 41 genes were identified. Therefore, 11 models were built and after optimization, three of them resulted in Area Under the Curve >0.7. Model 1 identified the role of *GATA2* gene in VT, via an inverse correlation with von Willebrand Factor that was stronger in controls; thus, the disruption of this correlation may reflect a prothrombic phenotype. Model 2 included Factor IX protein, 6 genes and 2 microRNAs, which represented an interaction between the fibrinolytic system (e.g., *ANXA2*) and the αIIbβ3 outside-in signaling platelet pathway (e.g., *ILK*, *ENTPD1*), as a mechanism underlying VT. Model 3 included *LYN* and *CSRP1* genes (correlated with Protein S) and two microRNAs, and suggested a platelet-based mechanism, involving both aggregation and activation processes. In conclusion, we have identified genes which expression in blood is associated with VT and their interaction with other molecular phenotypes.

Introduction

Venous thrombosis (VT) is the leading cause of preventable in-hospital mortality[1]. A further understanding of the disease could help to improve the prophylactic strategies. Driven by this need, basic and translational investigations are being conducted to identify new biomarkers[2], develop risk scores[3] or characterize the biological mechanisms underlying the disease[4].

Recently, studies in other complex diseases have demonstrated that the integration of different molecular phenotypes (e.g., genomic, transcriptomic, proteomic, etc.) is a promising strategy that helps to identify factors underlying the disease[5,6]. Although each biologic layer provides comprehensive information, studying multiple layers simultaneously provides understanding of their cross-talks. In addition, this approach could reveals dynamics of each single layer that could not be found through a more simplistic approach[7].

Until now, the most studied molecular phenotypes in relation to VT have been genetics[8], environmental factors[9] and clinical phenotypes, as hemostasis parameters[10]. Moreover, recent studies have explored other molecular phenotypes, such as DNA-methylation[11] or non-coding RNAs[12], identifying novel VT biomarkers and mechanisms underlying VT.

Notwithstanding, little is known about gene expression in blood in individuals with VT. Lewis et al[13] studied blood gene expression in VT patients on warfarin, including 23 single and 17 recurrent VT patients. They identified a set of 50 genes to distinguish between recurrent and single VT patients, which were enriched in genes of splicing processes. Later, also Lewis et al[14], compared individuals in four different groups of VT risk. They identified 3,131 genes differentially expressed between controls and high-risk patients and 446 between high and moderate risk patients; and pointed out those genes involved in the coagulation system. Finally, other authors have used their data to conduct alternative bioinformatic analysis[15]. To our knowledge, no more previous studies have explored blood gene expression in VT; thus, further analyses of this signature in human blood are still needed to increase our knowledge about VT.

In this study we aimed to identify genes differentially expressed in VT cases and to dissect their interactions with other molecular phenotypes. We analyzed blood gene expression from 102 individuals, including 51 VT patients. We assayed 260 genes related to the

blood coagulation pathway and conducted an integrative analysis including clinical phenotypes and microRNAs (miRNAs) related to VT.

Methods

Cohort

This study is based on the individuals of the GAIT-2 ('Genetic Analysis of Idiopathic Thrombophilia') Project, composed by 935 individuals of 35 Spanish families recruited via a proband with idiopathic thrombophilia[16]. A total of 104 GAIT-2 participants were included in this study and their characteristics have been described before[12]. Briefly, it included 52 subjects with VT and 52 non-VT controls, who were matched for age and sex and not genetically related. Two of the subjects were excluded for failed to measure gene expression and thus, final sample size was 102 and its characteristics are summarized in Supplemental Table 1. The collection of samples was performed at least 3 months after the last VT event and ongoing treatments (heparins, oral anticoagulants and anti-platelet drugs) were suspended prior.

Molecular phenotypes

We included 14 clinical VT phenotypes (i.e. hemostasis-related measurements), previously reported by Souto *et.al.*[17] and available in at least the 90% of the 102 GAIT-2 subjects: Factor (F) VII , FVIII, FIX, FXI, FXII, von Willebran Factor (vWF), Functional Protein S (funcPS), Fibrinogen (FIB), Functional Protein C (Pc), Antithrombin III (ATIII), Homocysteine (HCY), Prothrombin Time (PT), Activated Partial Thromboplastin Time (aPTT) and Activated Protein C Resistance (APCR). The methodologies for their quantification have been described extensively elsewhere[17].

The gene expression in whole blood was assessed for the genes annotated in blood coagulation closure (i.e. the main pathway and its direct and indirect interactors) in according to AmiGO[18] database. Samples were prepared for sequencing with the Illumina TruSeq sample preparation kit (Illumina, San Diego, CA) according to manufacturer's instructions and were sequenced on a HiSeq2000 machine using 49 basis pairs reads. The reads were mapped to GRCh37 reference genome with BWA v0.5.9[19] and we use genes defined as protein coding in the GENCODE 10[20] annotation. A total of 260 genes of the blood coagulation closure were successfully expressed in 102 individuals. The expression values were corrected for technical covariates and blood cell counts.

Finally, the four miRNAs previously reported by our group[12] as upregulated in VT were included: hsa-miR-885-5p, hsa-miR-192-5p, hsa-miR-194-5p and hsa-miR-126-3p. The miRNAs were quantified by qPCR from platelet-poor plasma using miRCURY LNA™ microRNA system (Exiqon). For each miRNA, the predicted targets (i.e. the genes that match the criteria to be regulated by each miRNA) were obtained from TargetScan[21] and the functionally-validated targets from miRTarBase[22]; they will be named as 'putative' target, meaning predicted and/or validated.

Statistical analyses

The workflow of this study is summarized in Figure 1 and included three steps: (1) Building of the models for VT prediction, (2.) Optimization of the models trough penalized regression and, (3) Training and test of the accuracy of the models.

All data were scaled by inverse normal transformation. Missing values were imputed with predictive mean matching method, using R package mice[23]. First, we calculated Pearson's correlations between 14 clinical phenotypes and the 260 genes. Linear logistic models for VT discrimination were built adding each phenotype and its correlated genes (False Discovery Rate [FDR] <0.1) as additive predictor for VT outcome. If any of the genes included in the model was a putative target for any of the four miRNAs, those were then added also as additive linear predictors.

Optimization of the models was performed using penalized regression by the Least Absolute Shrinkage and Selection Operator (LASSO)[24] algorithm, using R package glmnet[25]. It was applied in a 5-fold cross-validation resample strategy, and the models were adjusted based on the minimal lambda; non-significant features were subsequently removed from the model.

Finally, the accuracies of the optimized models were tested by Receiver-Operating Characteristics (ROC) analysis, with 1,000 interactions of bootstrap resampling, using R package mlr[26]. The models with the greatest discriminatory ability (i.e., Area Under the Curve [AUC] >0,7) were explored to dissect their biological implications.

First, it should be considered that if there is collinearity among phenotypes (e.g., genes co-expressed), it is likely that the penalized regression kept only one of them. Thus, principal component analysis (PCA) was performed using package FactoMineR[27] to allow considering the whole co-expressed group for the biological interpretation. Then, the 102 individuals were grouped by biological factors including categorical factors Sex

(Female/Male) and VT (Cases/Controls) and continuous variables Age and BMI (Body Mass Index) transformed into binomial variables by the median value (High/Low). Among these groups, differences between accuracies of the models were tested by the DeLong test[28], and differences between single correlations were tested attending to the Z-Fisher transformation of the correlation coefficients. Finally, for the models involving a large number of phenotypes, correlations among them were calculated based on Pearson distances and represented graphically using Cytoscape[29].

All statistical analyses were performed using R, version 3.6.1.

Results

Model building

Pearson correlations between the 14 clinical phenotypes and the expression levels of 260 genes identified 49 significant correlations (FDR <0.1), including 11 phenotypes and 41 genes (Supplemental Table 2).

Among the 41 genes included in the models, 13 were putative targets of any of the four miRNAs. Two genes were putative targets for two of the miRNAs (*PIK3CG* for hsa-miR-192-5p and hsa-miR-194-5p, and *PRKAR1A* for hsa-miR-126-3p and hsa-miR-885-5p) and *MAPK* was a putative target for three of them (hsa-miR-192-5p, hsa-miR-194-5p and hsa-miR-885-5p).

Consequently, the 11 linear models summarized in Table 1 were built, with each phenotype, its significantly correlated genes and the related miRNAs as additive predictors of VT.

Model optimization and accuracy test

The 11 models were then optimized using LASSO penalized regression, that subsequently removed the non-significant features. After excluding redundant models and those that failed to retain more than one feature as predictor, five optimized models were obtained.

The accuracies of the five optimized models were trained and tested using bootstrap resampling. Table 2 summarizes the mean accuracy measures of the test sample, which identified three models (Figure 2) with AUC >0.7.

Model 1

Model 1 (VT ~ vWF + *GATA2*) was based on a single correlation between vWF and the expression of the *GATA2* gene. The accuracy of the model did not show significant ($p < 0.05$) differences among the different groups of individuals (i.e., by age, sex or BMI). However, the inverse correlation between vWF and *GATA2* expression was significantly stronger in controls than in cases (βcontrols =-0,51; βcases =-0.20; $p = 3.1 \times 10^{-3}$); while no other biological factor gave a significant difference. Given the relationship between FvW levels and ABO blood-group, it was added as covariate in the model (considering bivariate O and non-O) and it had non-significant effect ($p<0.05$).

Model 2

Model 2 (VT ~ FIX + *ANXA2* + *ENTPD1* + *ILK* + *PDPK1* + *PRKAR1A* + *STXBP3* + hsa-miR-885-5p + hsa-miR-192-5p) was the most complex, involving 2 miRNAs, 6 genes and the blood concentration of coagulation factor IX. The result of PCA revealed three different expression groups (Supplemental Figure 1).

The accuracy of the model was significantly higher in the high-BMI group than in the low-BMI group (AUC_{lowBMI} =0,74, $AUC_{highBMI}$ =0.92, p =0.03). No other biological factor returned significant differences. However, the inverse correlation between: *ILK* and *PRKAR1A* genes was stronger in controls than in cases (βcontrols =-0.75, βcases =-0.37, FDR =0.13).

Finally, given the complexity of the model, a correlation matrix was calculated including also the other clinical phenotypes correlated to any gene in the model (those in Supplemental Table 2) and the genes removed in the optimization step (those in Table 1). Figure 3 represents the global network of this model.

Model 3

Model 3 (VT~ *CSRP1* + *LYN* + hsa-miR-192-5p + hsa-miR-885-5p) involved two genes and two miRNAs, with no VT clinical phenotype. Figure 4 shows the PCA analysis including all of the genes in the preliminary model, which revealed two different groups of genes and that one gene of each group remained in the final model.

No significant differences were found between the accuracies of the model in the different groups of individuals. However, the association between *CSRP1* and hsa-miR-192-5p

reached a nominally significant difference between male and females ($\beta_{females}$ =-0.23, β_{males} =0.19, p =0.03).

Discussion

To improve understanding of VT, in this study we have conducted a comprehensive analysis of blood gene expression levels related to VT and their interactions with clinical phenotypes and miRNAs.

First, a model including only vWF and gene expression of *GATA2* showed an AUC of 0.78. In last years, vWF has been revealed as one of the main blood indicators of VT risk. A recent study determined that vWF and FVIII were the procoagulant factors associated with the highest VT risk and that both of them explained a large part of the risk associated with other procoagulant factors[30]. In accordance, our results indicate also this central role of vWF concentration in blood for VT risk assessment. On the other hand, *GATA2* is an endothelial transcription factor implicated in several biological processes, including a large range of hematological disorders[31]. *GATA2* is a transcription factor for vWF in endothelial cells[32]. A study[31] that involved GATA2-deficiency patients found a high incidence (25%) of VT in these patients. Moreover, the genes of blood coagulation pathway were found to be dysregulated in murine models of GATA2-deficiency[33]. It should be highlighted that this association was independent of the ABO-blood group and that, given that gene expression levels were corrected by blood cells count, it can not be biased by red cells count. Our results show for first time that levels of *GATA2* expression in human whole blood are associated with VT and that are anti-correlated to vWF protein in blood. Furthermore, this correlation was stronger in controls compared to VT patients, suggesting that, either as a cause or as a consequence, the disruption of this inverse association represents a prothrombotic signature.

Second model was the most complex and the PCA analysis revealed three groups of features. The first group included FIX protein and the expression of *ILK* and *ANXA2* genes. *ANXA2* expression in blood has been found previously as dysregulated between low and high-risk VT cases[14]. *ANXA2* gene codifies for annexin A2, a co-receptor for plasminogen and tissue plasminogen activator and thus, has a central role in plasmin generation[34,35]. A recent study has suggested for first time aberrant expression of annexin A2 as an independent risk factor for VT, *via* an affection of the plasmin generation capacity[36]. Our results support this role of *ANXA2* in VT and, while plasmin is known to

reduce FIX levels and activity[37], we also found that *ANXA2* expression is inversely correlated with FIX. Moreover, plasmin also regulates platelet activity and aggregation[38]. *ILK* gene, the last member of this first group, has been reported as a major determinant in platelet activation needed for thrombus stability *in vivo*, mainly *via* the αIIbβ3 outside-in signaling pathway[39]. All of the genes in the second group have been linked also to the αIIbβ3 signaling pathway: *PDK1* is a major αIIbβ3 interactor[40], *ENTPD1* attenuates the receptor activity[41]; and *PRKAR1A* is a regulatory subunit of cAMP/PKA signaling, associated with inhibition of αIIbβ3[42]. The last member of the group was *STXBP3*, a member of the family Sec1/Munc18, from which other members have been reported as necessary for αIIbβ3 signaling[43]. In addition, a previous study reported that low-risk VT patients expressed *STXBP3* more in whole blood than high-risk patients[14]. Finally, two miRNAs composed the third group, with *PRKAR1A* as predicted[21] and validated[44] target for hsa-miR-192-5p, and *ENTPD1* as predicted[21] target for hsa-miR-885-5p. Overall, we hypothesize that this model with AUC= 0.77, identified the cross-talk between FIX protein, the fibrinolytic system via annexin-2 receptor, and the platelet αIIbβ3 outside-in signaling pathway (Figure 3). Moreover, this model was more accurate in the BMI-high group and both, platelet activation and fibrinolytic system, have been associated with obesity[45,46].

The third model included *LYN* and *CSRP1* genes and two miRNAs. *LYN* is a kinase protein of the Src-family involved in platelet activation, mainly affecting some platelets receptors, as the collagen adhesion to glycoprotein VI receptor in platelets[47] or the receptor of vWF[48]. To note, *LYN* was included in the model as it was correlated with funcPS, while Protein S is a cofactor for tissue factor pathway inhibitor released from endothelium and platelets[49]. The other gene in the model was *CSRP1*, never associated to VT before, but also involved in platelet activity and aggregation. As aforementioned, these two features fall into two groups of genes (as shown in Figure 4), and most of the genes in each group also fitted this hypothesis; for example, *CECAM1* in *LYN* group was related also to the interaction collagen-platelets[50] and *ANO6*, in the second group, has been implicated in platelet aggregation and activation[51]. This model, that suggests a platelet activation-based mechanism, was completed by two miRNAs, including hsa-miR-192-5p, with three validated targets in the correlated genes: *ANO6, H3F3A* and *PABPC4*.

Finally, beyond the discussion of each of the three models, some global interpretations could be extracted from the findings of this study. First, most of the genes identified were implicated on endothelium-specific processes and platelet-related mechanisms. We could expect to find more elements implicated in local (i.e., blood or endothelium) processes than those implicated in the distal processes, such as the regulation of coagulation factors uniquely synthetized in the liver. However, these findings encourage current trend in VT research, in which other processes beyond the clotting cascade are raising relevance. In this regard, it should be noted that most of the genes that we identified are involved in platelet related mechanisms, which implication in VT has gained relevance recently[52,53]. On the other hand, we could point out the role of *GATA* transcription factors family in VT risk. Although only *GATA2* remained significant in the most accurate models, *GATA6* was significantly correlated with APCR and *GATA3* with HCY (as shown in Supplemental Table 2). Therefore, our results support previous finding that suggested the importance of the genes of the *GATA* family in VT[54–56]. Finally, two of the three final hit models included miRNAs, which remained significant for VT discrimination even after penalized regression. Thus, our study supports their utility as informative intermediate features in the mechanisms underlying VT.

Some limitations of this study must be discussed. First, the pre-selection of the features based on prior knowledge limited the number of relationships interrogated and therefore, perhaps some important variables underlying VT remain hidden. However, dimension reduction was needed to avoid major issues regarding multiple testing penalization, missing values or highly inter-correlated structures of the datasets[57]. Also, as a cautionary note, several studies have shown that gene expression levels in blood are poorly correlated with their corresponding protein levels[58] and therefore, all interpretations of the results must regard exclusively the gene expression level. Finally, we acknowledge that all findings must be validated in an independent population.

In conclusion, in this study we have shed light into the blood signature of VT, employing gene expression levels of genes in the blood coagulation pathway and their interactions with miRNAs and clinical phenotypes related to VT.

References

1. Beckman MG, Hooper WC, Critchley SE, Ortel TL. Venous Thromboembolism. A Public Health Concern. Am J Prev Med 2010;38(4 Suppl.):S495–S501.
2. Farah R, Nseir W, Kagansky D, Khamisy-farah R. The role of neutrophil-lymphocyte ratio, and mean platelet volume in detecting patients with acute venous thromboembolism. J Clin Lab Anal 2019;n/a(n/a):e23010.
3. Timp JF, Braekkan SK, Lijfering WM, et al. Prediction of recurrent venous thrombosis in all patients with a first venous thrombotic event: The Leiden Thrombosis Recurrence Risk Prediction model (L-TRRiP). PLOS Med 2019;16(10):e1002883.
4. Moschonas IC, Tselepis AD. The pathway of neutrophil extracellular traps towards atherosclerosis and thrombosis. Atherosclerosis 2019;2889–16.
5. Hasin Y, Seldin M, Lusis A. Multi-omics approaches to disease. Genome Biol 2017;18(1):1–15.
6. Wu Y, Zeng J, Zhang F, et al. Integrative analysis of omics summary data reveals putative mechanisms underlying complex traits. Nat Commun;9(1):.
7. Sun Y V, Hu Y-J. Integrative Analysis of Multi-omics Data for Discovery and Functional Studies of Complex Human Diseases. Adv Genet 2016;93147–190.
8. Morange PE, Suchon P, Trégouët DA. Genetics of venous thrombosis: Update in 2015. Thromb Haemost 2015;114(5):910–919.
9. Crous-Bou M, Harrington LB, Kabrhel C. Environmental and Genetic Risk Factors Associated with Venous Thromboembolism. Semin Thromb Hemost 2016;42(8):808–820.
10. Lipets EN, Ataullakhanov FI. Global assays of hemostasis in the diagnostics of hypercoagulation and evaluation of thrombosis risk. Thromb J;13(1):.
11. Aïssi D, Dennis J, Ladouceur M, et al. Genome-Wide investigation of DNA methylation marks associated with FV Leiden mutation. PLoS One;9(9):.
12. Rodriguez-Rius A, Lopez S, Martinez-Perez A, Souto JC, Soria JM. Identification of a Plasma MicroRNA Profile Associated with Venous Thrombosis. Arterioscler Thromb Vasc Biol 2020;(May):1392–1399.
13. Lewis DA, Stashenko GJ, Akay OM, et al. Whole blood gene expression analyses in patients with single versus recurrent venous thromboembolism. Thromb Res 2011;128(6):536—540.
14. Lewis DA, Suchindran S, Beckman MG, et al. Whole blood gene expression profiles distinguish clinical phenotypes of venous thromboembolism. Thromb Res 2015;135(4):659–665.
15. Zhou W, Zhang K, Chen D, Gao P, Wang Q. Gene microarray analyses for potential biomarkers of single and recurrent venous thromboembolism. Mol Med Rep 2015;12(5):7358–7366.
16. Martin-Fernandez L, Ziyatdinov A, Carrasco M, et al. Genetic determinants of

thrombin generation and their relation to venous thrombosis: Results from the GAIT-2 project. PLoS One 2016;11(1):1–12.

17. Souto JC, Almasy L, Borrell M, et al. Genetic determinants of hemostasis phenotypes in Spanish families. Circulation 2000;101(13):1546–1551.

18. Carbon S, Ireland A, Mungall CJ, et al. AmiGO: Online access to ontology and annotation data. Bioinformatics 2009;25(2):288–289.

19. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics 2009;25(14):1754–1760.

20. Harrow J, Frankish A, Gonzalez JM, et al. GENCODE : The reference human genome annotation for The ENCODE Project. Genome Res 2012;1760–1774.

21. Agarwal V, Bell GW, Nam JW, Bartel DP. Predicting effective microRNA target sites in mammalian mRNAs. Elife 2015;4(e05005):1–38.

22. Huang H-Y, Lin Y-C-D, Li J, et al. miRTarBase 2020: updates to the experimentally validated microRNA–target interaction database. Nucleic Acids Res 2019;48(D1):D148–D154.

23. van Buuren S, Groothuis-Oudshoorn K. mice: Multivariate Imputation by Chained Equations in R. J Stat Software; Vol 1, Issue 3 [Epub ahead of print].

24. Tibshirani R. Regression Shrinkage and Selection via the Lasso. J R Stat Soc Ser B 1996;58(1):267–288.

25. Friedman JH, Hastie T, Tibshirani R. Regularization Paths for Generalized Linear Models via Coordinate Descent. J Stat Software; Vol 1, Issue 1 [Epub ahead of print].

26. Bischl B, Lang M, Kotthoff L, et al. Mlr: Machine learning in R. J Mach Learn Res 2016;171–5.

27. Lê S, Josse J, Husson F. FactoMineR: An R Package for Multivariate Analysis. J Stat Software; Vol 1, Issue 1 [Epub ahead of print].

28. DeLong ER, DeLong DM, Clarke-Pearson DL. Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. Biometrics 1988;44(3):837–845.

29. Shannon P, Markiel A, Ozier O, et al. Cytoscape: a software environment for integrated models of biomolecular interaction networks. Genome Res 2003;13(11):2498–2504.

30. Rietveld IM, Lijfering WM, le Cessie S, et al. High levels of coagulation factors and venous thrombosis risk: strongest association for factor VIII and von Willebrand factor. J Thromb Haemost 2019;17(1):99–109.

31. Spinner MA, Sanchez LA, Hsu AP, et al. GATA2 deficiency: a protean disorder of hematopoiesis, lymphatics, and immunity. Blood 2014;123(6):809–821.

32. Jahroudi N, Lynch DC. Endothelial-cell-specific regulation of von Willebrand factor gene expression. Mol Cell Biol 1994;14(2):999–1008.

33. Johnson KD, Hsu AP, Ryu M-J, et al. Cis-element mutated in GATA2-dependent

immunodeficiency governs hematopoiesis and vascular integrity. J Clin Invest 2012;122(10):3692–3704.

34. Flood EC, Hajjar KA. The annexin A2 system and vascular homeostasis. Vascul Pharmacol 2011;54(3–6):59–67.

35. Roda O, Valero ML, Peiró S, Andreu D, Real FX, Navarro P. New insights into the tPA-annexin A2 interaction: Is annexin A2 Cys 8 the sole requirement for this association? J Biol Chem 2003;278(8):5702–5709.

36. Fassel H, Chen H, Ruisi M, De Sancho MT, Hajjar KA. Markedly Reduced Expression of Annexin A2 Is Associated with Hypofibrinolysis and Provoked and Unprovoked Venous Thromboembolism. Blood 2019;134(Supplement_1):443.

37. Samis JA, Ramsey GD, Walker JB, Nesheim ME, Giles AR. Proteolytic processing of human coagulation factor IX by plasmin. Blood 2000;95(3):943–951.

38. van der Vorm L, Remijn J, de Laat B, Huskens D. Effects of Plasmin on von Willebrand Factor and Platelets: A Narrative Review. TH Open 2018;02(02):e218–e228.

39. Jones CI, Tucker KL, Sasikumar P, et al. Integrin-linked kinase regulates the rate of platelet activation and is essential for the formation of stable thrombi. J Thromb Haemost 2014;12(8):1342–1352.

40. Chen X, Zhang Y, Wang Y, et al. PDK1 regulates platelet activation and arterial thrombosis. Blood 2013;121(18):3718–3726.

41. Huttinger ZM, Milks MW, Nickoli MS, et al. Ectonucleotide triphosphate diphosphohydrolase-1 (CD39) mediates resistance to occlusive arterial thrombus formation after vascular injury in mice. Am J Pathol 2012;181(1):322–333.

42. Ya F, Xu XR, Shi Y, et al. Coenzyme Q10 Upregulates Platelet cAMP/PKA Pathway and Attenuates Integrin αIIbβ3 Signaling and Thrombus Growth. Mol Nutr Food Res 2019;19006621900662.

43. Xiang B, Zhang G, Ye S, et al. VPS33B, a Member of the Sec1/Munc18 Protein Family, Binds to Integrin Beta Subunits and Is Required for αIIbβ3 Outside-in Signaling. Blood 2014;124(21):340.

44. Chou C, Shrestha S, Yang C, et al. miRTarBase update 2018 : a resource for experimentally validated microRNA-target interactions. Nucleic Acids Res 2018;46(D1):D296–D302.

45. Mertens I, Van Gaal LF. Obesity, haemostasis and the fibrinolytic system. Obes Rev 2002;3(2):85–101.

46. Anfossi G, Russo I, Trovati M. Platelet dysfunction in central obesity. Nutr Metab Cardiovasc Dis 2009;19(6):440–449.

47. Suzuki-Inoue K, Tulasne D, Shen Y, et al. Association of Fyn and Lyn with the proline-rich domain of glycoprotein VI regulates intracellular signaling. J Biol Chem 2002;277(24):21561–21566.

48. Yin H, Liu J, Li Z, Berndt MC, Lowell CA, Du X. Src family tyrosine kinase Lyn mediates VWF/GPIb-IX-induced platelet activation via the cGMP signaling pathway. Blood 2008;112(4):1139—1146.

49. Wood JP, Ellery PER, Maroney SA, Mast AE. Protein S is a cofactor for platelet and endothelial tissue factor pathway inhibitor-alpha but not for cell surface-associated tissue factor pathway inhibitor. Arterioscler Thromb Vasc Biol 2014;34(1):169–176.

50. Wong C, Liu Y, Yip J, et al. CEACAM1 negatively regulates platelet-collagen interactions and thrombus growth in vitro and in vivo. Blood 2009;113(8):1818–1828.

51. Liu G, Liu G, Chen H, et al. Involvement of Ca2+ Activated Cl- Channel Ano6 in Platelet Activation and Apoptosis. Cell Physiol Biochem 2015;37(5):1934–1944.

52. Vázquez-Santiago M, Vilalta N, Cuevas B, et al. Short closure time values in PFA–100® are related to venous thrombotic risk. Results from the RETROVE Study. Thromb Res 2018;16957–63.

53. Marcus L, M. SR, J. KP, et al. Platelets Drive Thrombus Propagation in a Hematocrit and Glycoprotein VI–Dependent Manner in an In Vitro Venous Thrombosis Model. Arterioscler Thromb Vasc Biol 2018;38(5):1052–1062.

54. Minami T, Aird WC. Thrombin Stimulation of the Vascular Cell Adhesion Molecule-1 Promoter in Endothelial Cells Is Mediated by Tandem Nuclear Factor-κB and GATA Motifs. J Biol Chem 2001;276(50):47632–47641.

55. Petit L, Lesnik P, Dachet C, et al. The Promoter of Human Tissue Factor Pathway Inhibitor Gene. Thromb Res 1999;95(5):255–262.

56. Ameri A, Kuppuswamy MN, Basu S, Bajaj SP. Expression of tissue factor pathway inhibitor by cultured endothelial cells in response to inflammatory mediators. Blood 1992;79(12):3219–3226.

57. de Maturana EL, Alonso L, Alarcón P, et al. Challenges in the integration of omics and non-omics data. Genes (Basel);10(3):.

58. Maier T, Güell M, Serrano L. Correlation of mRNA and protein in complex biological samples. FEBS Lett 2009;583(24):3966–3973.

Tables

Table 1. Preliminary linear models for VT discrimination. For each phenotype, all the mRNAs significantly correlated (FDR<0.1) and the miRNAs with a putative target in the model. Number in subscript indicates the putative target of each miRNA that was, consequently, included in the model.

Phenotype	mRNAs	miRNAs
PSfunc	*ANO6* (1), *C1QBP, CEACAM1, CSRP1, ENTPD1, GGCX, H3F3A* (1), *HMG20B, HPS4, ILK, IRF2, LYN, P2RX6, PABPC4* (2), *PAFAH2* (2), *PHF21A* (3), *SH2B2, VPS45*	hsa-miR-192-5p (1), hsa-miR-885-5p (2), hsa-miR-194-5p (3)
FXI	*ENTPD1* (1), *MAPK1* (2), *STXBP3, TLR4* (3), *VAV3, VKORC1*	hsa-miR-885-5p (1, 3), hsa-miR-192-5p (2), hsa-miR-194-5p (2)
FIX	*ANXA2, ENTPD1* (1), *ILK, JMJD1C, MYL12A, PDPK1, PIK3CA, PRKAR1A* (2), *STXBP3, THBD* (3)	hsa-miR-885-5p (1), hsa-miR-194-5p (2), hsa-miR-192-5p (2, 3)
FXI	*DOCK11* (1), *PIK3CG* (2)	hsa-miR-885-5p (1, 2), hsa-miR-126-3p (2)
HCY	*GATA3, HPS6, PRTN3*	
APTT	*STXBP3*	
APCR	*GATA6*	
FIB	*ARRB2, CD34*	
Pc	*DGKZ, VKORC1, ZFPM1*	
FVII	*THBD* (1), *ZFPM1*	hsa-miR-192-5p (1)
FvW	*GATA2*	

Table 2. Accuracy measures of the five optimized linear models for VT discrimination. Results of the ROC analysis with the 1,000 bootstrapping interactions, reporting the mean accuracy measures of the test sample.

	Model	AUC	TPR	TNR
1	VT ~ FvW + *GATA2*	0.78	0.73	0.70
2	VT ~ FIX + *ANXA2* + *ENTPD1* + *ILK* + *PDPK1* + *PRKAR1A* + *STXBP3* + hsa-miR-885-5p + hsa-miR-192-5p	0.75	0.71	0.67
3	VT~ *CSRP1* + *LYN* + hsa-miR-192-5p + hsa-miR-885-5p	0.73	0.70	0.64
4	VT~ FXI + *STXBP3* + *TLR4*+ *VKORC1* + hsa-miR-885-5p + hsa-miR-192-5p	0.68	0.67	0.60
5	VT~ APTT + *STXBP3*	0.63	0.59	0.58

* TPR: True Positive Rate
† TNR: True Negative Rate

Figures and Figure Legends

Figure 1. Workflow of the study. Schematic representation of the regression-based strategy followed in this study.

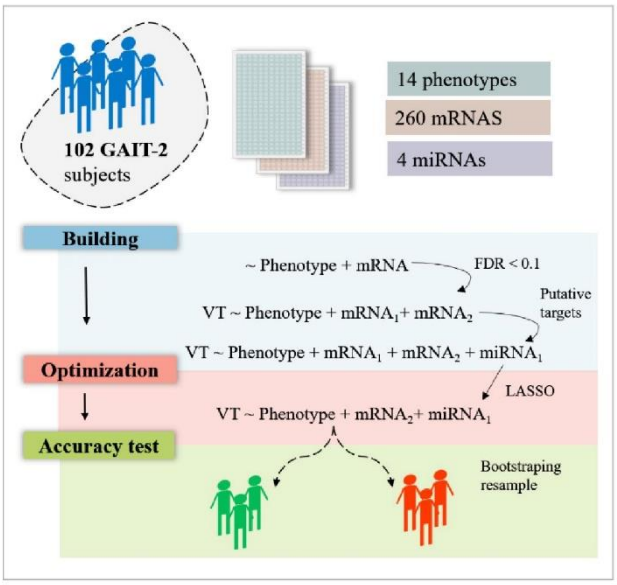


Figure 2. Receiver Operating Characteristic curves of the three models with AUC > 0.7. Accuracy of the three hit models, represented by the mean probability value for each subject in the 1,000 bootstrapping interactions.

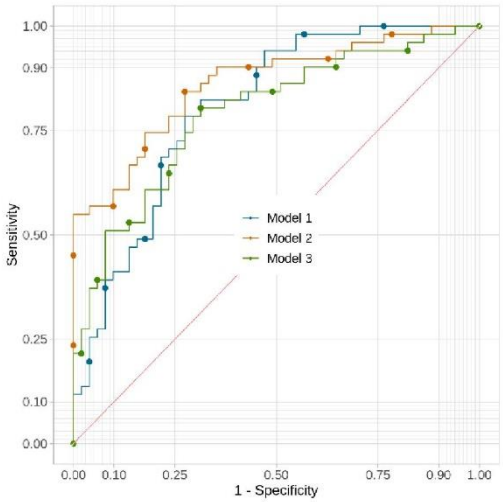


Figure 3. Representation of the interactions among clinical phenotypes and genes related to Model 2. Features retained in final Model 2 were represented in green, while yellow features were those genes removed by penalized regression, and the phenotypes significantly correlated to any of the genes (reported in Table S1 and Table 1). Graphical representation of those significant (adjusted p <0.05) correlations. The distance among features is proportional to their Pearson`s correlation coefficients.

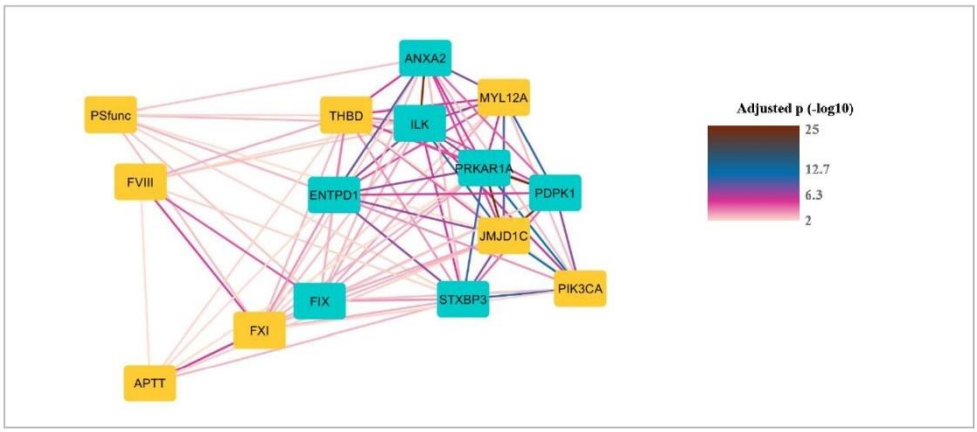
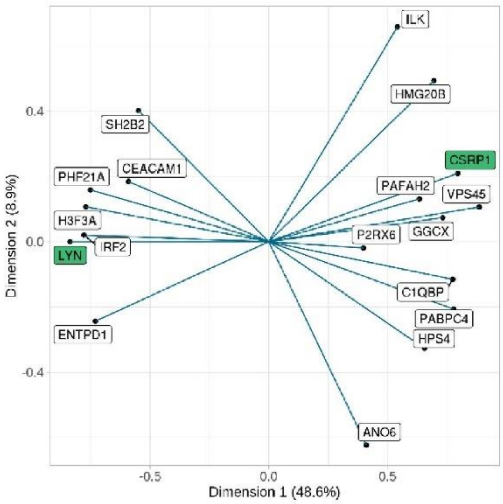


Figure 4. Principal Component Analysis for the genes that led to Model 3. Graphic representation of the first and second principal components, for those 18 genes significantly (FDR <0.1) correlated to funcPS. Highlighted in green, the two features that remained on the model after penalized regression.



4. Article 3

Expression of microRNAs in human platelet-poor plasma: analysis of the factors affecting their expression and association with proximal genetic variants

EPIGENETICS
https://doi.org/10.1080/15592294.2020.1783497

Taylor & Francis
Taylor & Francis Group

RESEARCH PAPER

Check for updates

Expression of microRNAs in human platelet-poor plasma: analysis of the factors affecting their expression and association with proximal genetic variants

Alba Rodriguez-Rius [a], Angel Martinez-Perez [a], Sonia López [a], Maria Sabater-Lleal [a,b], Juan Carlos Souto [c], and José Manuel Soria [a]

[a]Genomics of Complex Diseases Group, Research Institute of Hospital De La Santa Creu I Sant Pau, IIB-Sant Pau, Barcelona, Spain; [b]Cardiovascular Medicine Unit, Department of Medicine, Karolinska Institutet, Center for Molecular Medicine, Karolinska University Hospital, Stockholm, Sweden; [c]Unit of Thrombosis and Hemostasis, Hospital De La Santa Creu I Sant Pau. Barcelona, Spain

ABSTRACT

To translate circulating microRNAs (miRNAs) into the clinic, a deeper understanding of the factors affecting their expression is needed. In this study, we explored the features affecting the expression of miRNAs and their genetic regulation using the expression data of 103 miRNAs obtained by qPCR in the platelet-poor plasma of 104 subjects. The principal components (PCs) of the expression of the miRNAs were associated with technical and biological features (e.g., synthetic controls or sex) and with blood cell counts. Also, the associations with proximal genetics variants were analysed. We found that haemolysis marker (dCt hsa-miR-23a-3p-hsa-miR-451a) was correlated strongly (β = 0.84, p = $2.07x10^{-29}$) with the second PC, which explained 10.1% of the overall variability. Thus, we identified haemolysis as a source of variability for miRNA expression even in mild hemolyzed samples (haemolysis marker dCt <5). In addition to hsa-miR-23a-3p and hsa-miR -451a, the miRNAs most stable and most susceptible to haemolysis were identified. Then, we discovered that the expression of miRNAs in platelet-poor plasma was not biased by any blood cell count, and thus, our results supported their role as biomarkers of tissue-specific conditions. Finally, we identified 1,323 genetic variants that corresponded to 158 miRNA expression quantitative trait loci for 14 miRNAs (FDR <0.2), which were enriched in promoter regions (p = 0.03). This enrichment corresponded to a range of specific tissues (e.g., breast or fat) although not to blood tissue, supporting the concept that the expression of circulating miRNAs is under the genetic control of different tissues.

ARTICLE HISTORY
Received 20 February 2020
Revised 1 June 2020
Accepted 2 June 2020

KEYWORDS
Micrornas; biomarkers; haemolysis; miR-eQTLs; qPCR; platelet-poor plasma; inter-assay variability; blood cell counts

Introduction

MicroRNAs (miRNAs) are small non-coding RNAs that regulate the expression of their target genes binding a complementary seed sequence[1]. Several studies have identified miRNA signatures associated with a specific disease or stages of disease [2]. Among them, circulating miRNAs (i.e., those in whole blood, plasma, or serum) are particularly relevant because these are easy to assess in a non-invasive way, which is the most desirable characteristic of a useful biomarker.

Despite the promising findings in some diseases [2,3], the clinical application is still unachieved [4]. In this regard, a wide range of analytic and pre-analytic conditions cause high inter- and intra-assay variability. For example, the extraction method, type of tubes for collection, time of storage, and normalization method must be reported to promote consistency across studies [5–8]. Also, the biological basis of circulating miRNAs must be considered, in the context of the constant interaction with the abovementioned technical issues. For example, major differences must be expected among circulating miRNAs profiles when released from cells with or without vesicles, exosomes, or lipids [9], between plasma and platelet-poor plasma [10], or between serum and plasma [8]. Moreover, for circulating miRNAs synthetized in a specific blood cell, the effect of the cell count as a confounder must be considered [11]. All of these previous findings have provided a valuable basis to standardize methods for measurement, processing, and normalizing miRNA

CONTACT Alba Rodriguez-Rius arodriguezri@santpau.cat Genomics of Complex Diseases Group, IIB-Sant Pau, Barcelona 08014, Spain
Supplemental data for this article can be accessed here

data. However, further studies are still needed before miRNAs can be utilized in the clinic.

Moreover, the genetic regulation of miRNAs must be studied further. Some genome-wide mapping studies have reported datasets of miRNA expression quantitative trait loci (miR-eQTLs). Most relevant examples are of the mapping studies conducted on whole blood as a part of the Framingham study [12], on lymphoblastoid cell lines [13], and most recently, on the whole blood of European obese subjects [14]. These studies have provided valuable information about genetic biases and trends, such as an enrichment in mRNA-eQTLs or an important effect of distal variants. Moreover, other authors have used their miR-eQTLs datasets, for example, to develop a probability model to classify the genetic variants around a miRNA gene in eQTL or non-eQTL [15]. However, the examples are still too few, so more tissue and cell-specific genome-wide mapping studies are needed. Each new piece of information will help to increase our knowledge about genetic regulation of miRNAs and their role in human diseases.

In the present study, we analyse the expression data of 103 platelet-poor plasma miRNAs in 104 individuals. Technical and biological features (e.g., haemolysis marker or age), as well as blood cell counts, are examined as sources of variability. Moreover, we unravel the proximal miR-eQTLs and their relationship with regulatory regions.

Results

Sources of variability in miRNA expression

The principal component analysis (PCA) of the 103 miRNAs expressed in at least 90% of the 104 individuals revealed that the first four principal components (PCs) explained 44.2% of the overall variability (Figure 1). These first four PCs were extracted and correlated with technical and biological features, namely, synthetic controls (UniSp2, UniSp4, UniSp5, UniSp6, and UniSp3), haemolysis marker (dCt hsa-miR-23a-3p-hsa-miR-451a), age, sex, body mass index (BMI), smoking habit and disease condition. The results are given in Supplemental Data 1.

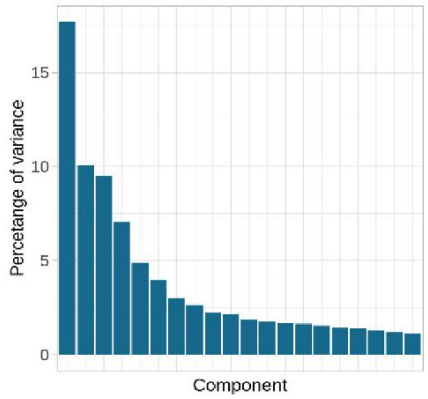
Only one significant (false discovery rate [FDR] <0.1) association was identified, the second PC, which explained 10.1% of the variability was strongly correlated with the haemolysis marker ($\beta = 0.84$, $p = 2.07x10^{-29,}$ Figure 2). It should be noted that the maximum value of the haemolysis marker dCt was 4.46, far from dCt = 7, from which samples should be excluded, and below dCt = 5, from which samples are considered as at low risk of haemolysis (please, see Methods for further details). Furthermore, to ensure that the disease condition was not a confounder, the correlation between disease and haemolysis was calculated and it was not significant ($\beta = -0.11$, $p = 0.27$).

The larger contributors and, therefore, the better-represented miRNAs on the second PC are shown in Figure 3. As expected, the haemolysis marker-miRNAs (i.e., hsa-miR-451a and hsa-miR-23a-3p) were among the greater contributors and on opposite coordinates. However, also other miRNAs were closely related to both of them: the miRNAs most susceptible and most stable to haemolysis. Furthermore, the contributions of these 12 highlighted miRNAs explained ~44% of the second PC, whereas the remaining 91 miRNAs accounted for 56% of the variability represented in the second dimension (Supplemental Data 2). Therefore, even in low hemolyzed samples, haemolysis could be understood as a confounding factor in plasma miRNA expression studies.

Effect of blood cell counts on miRNA expression

The first four PCs were correlated with blood cell counts. However, none of them was associated significantly (FDR >0.1). To explore the effect on the individual variability, single correlations between each of the miRNAs and each blood cell count were tested (Supplemental Data 3). Only one significant correlation was obtained: hsa-miR-150-5p and lymphocyte count ($\beta = 0.39$, $p = 4.8x10^{-05}$, FDR = 0.07). Moreover, the same analysis was performed using the residual values of the miRNAs corrected by the first four PCs (Supplemental Data 4). Only the above correlation reached statistical significance ($\beta = 0.40$, $p = 2.90x10^{-05}$, FDR = 0.06). Therefore, no blood cell count was a major determinant of the overall variability affecting miRNA expression in platelet-poor plasma, and only one of 103 miRNAs

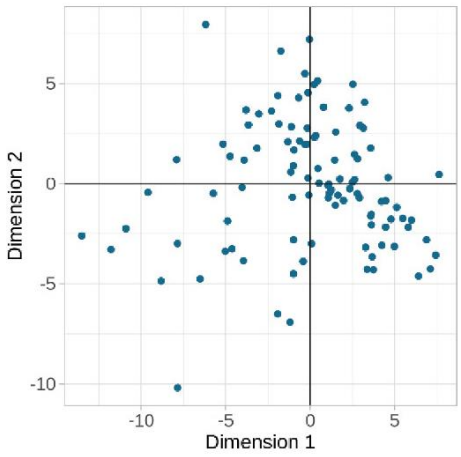
a



b)

Component	Eigen value	Variance (%)	Cum. variance (%)
1	18.22	17.69	17.69
2	10.36	10.06	27.75
3	9.76	9.48	37.23
4	7.25	7.04	44.27

c



d)

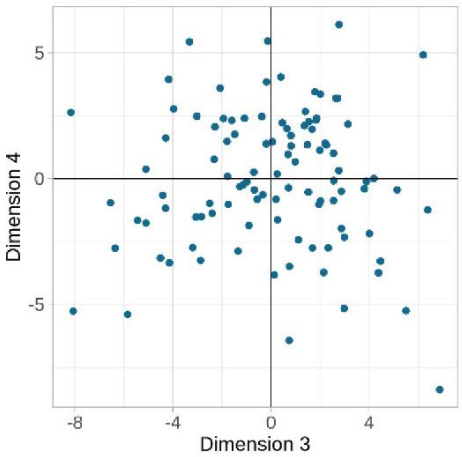


Figure 1. **PCA of the 103 plasma miRNAs in the 104 subjects**. (a) Percentage of variance explained by each of the first 20 components. (b) Eigen value, percentage of variance explained and cumulative (cum) variance explained by each of the first four principal components. (c) Coordinates for each of the 104 subjects, in the first and second PCs. The three most contributing miRNAs were: hsa-miR-103a-3p, hsa-miR-199a-3p and hsa-miR-107 for the first PC, and hsa-miR-451a-3p, hsa-miR-144-3p and hsa-miR-23b-3p for the second PC. (d) Coordinates for the 104 subjects in the third and fourth components. The three most contributing miRNAs were hsa-miR-19a-3p, hsa-miR-19b-3p and hsa-miR-423-5p for the third PC, and hsa-miR-29a-3p, hsa-miR-185-5p and hsa-miR-25-5p for the fourth PC.

analysed was significantly associated with a specific blood cell count.

The correlations between disease and blood cell counts were calculated to ensure that the disease was not a confounding factor. None of the correlations reached statistical significance (FDR >0.1).

Genome-wide mapping of miRNA expression

The expression of each miRNA, corrected for the effect of the aforementioned first four PCs, was associated with the genetic variants located 1MB around the corresponding miRNA gene. As the X chromosome was not imputed in our sample, a total of 98 miRNAs were included.

At FDR <0.2, 1,323 significant genetic variants were identified (Supplemental Data 5). Setting linkage disequilibrium (LD) limit at $r^2 = 0.8$, corresponded with 158 unique *cis*-miR-eQTLs for 14 mature miRNAs and 16 miRNA genes (Supplemental Data 6).

Therefore, *cis*-miR-eQTLs were detected for 14.3% of the miRNAs. None of the signals was shared between more than one miRNA. hsa-miR-7a-5p is encoded at three different positions in the genome, and significant signals were obtained for all of them. For the other miRNAs that are

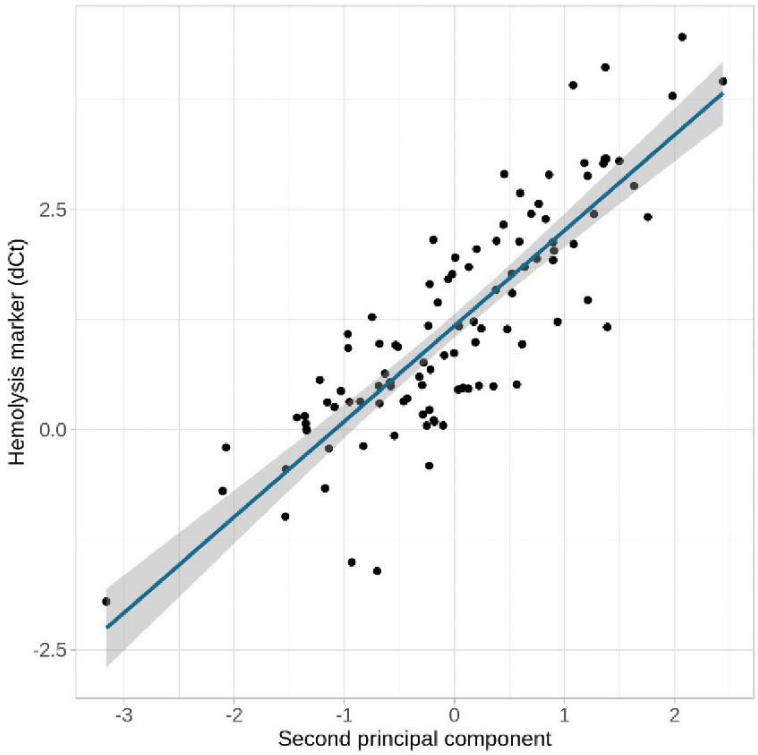


Figure 2. **Correlation between haemolysis marker and second PC**. The second PC of the 103 miRNAs explained a 10.1% of their variability and was strongly correlated (β = 0.84, p = $2.07x10^{-28}$) with the haemolysis marker (dCt hsa-miR-23a-3p – hsa-miR-451a).

encoded at more than one position, *cis*-miR-eQTLs were obtained only for one of the positions. Regarding the 158 *cis*-miR-eQTLs, most of them were located in intronic regions (62.6%) and 50,63% were also mRNA-eQTLs.

Tissue specificity of the miR-eQTLs

First, the positions of the 158 *cis*-miR-eQTLs were analysed in relation to promoter or enhancer regions determined by histone markers, as indicators of active regulatory regions. The frequencies of both enhancer and promoter regions were compared with the expected by chance in non-miR-eQTLs (please see Methods). Though there was not enrichment in enhancer regions (expected = 50.63%, observed = 53.97; p > 0.05), there was a significant enrichment in promoter regions (expected = 15.48%, observed = 21.51%; p = 0.03).

The tissues in which the miR-eQTLs were identified as regulatory regions were analysed by testing the enrichment against the random non-miR-eQTLs. Table 1 summarizes the results for both enhancer and promoter regions. Although there was not a global enrichment in enhancers, the lung and liver enhancers were specifically overrepresented. Moreover, while the great majority of tissue-specific promoters were enriched, only blood and gastrointestinal tissues were not significantly overrepresented (Figure 4). Overall, *cis*-miR-eQTLs for expression of miRNAs in platelet-poor plasma were enriched in promoter regions in several specific tissues but not in blood tissue.

Discussion

In this study, we have shed light on the factors that affect the expression of miRNAs in plasma and on their genetic regulation.

Our results showed that the haemolysis marker was strongly (Figure 2) correlated with the second PC, which explained ~10% of the variability in the 103 widely expressed miRNAs. While some technical and analytical parameters regarding miRNA quantification are not standardized yet [5–7],

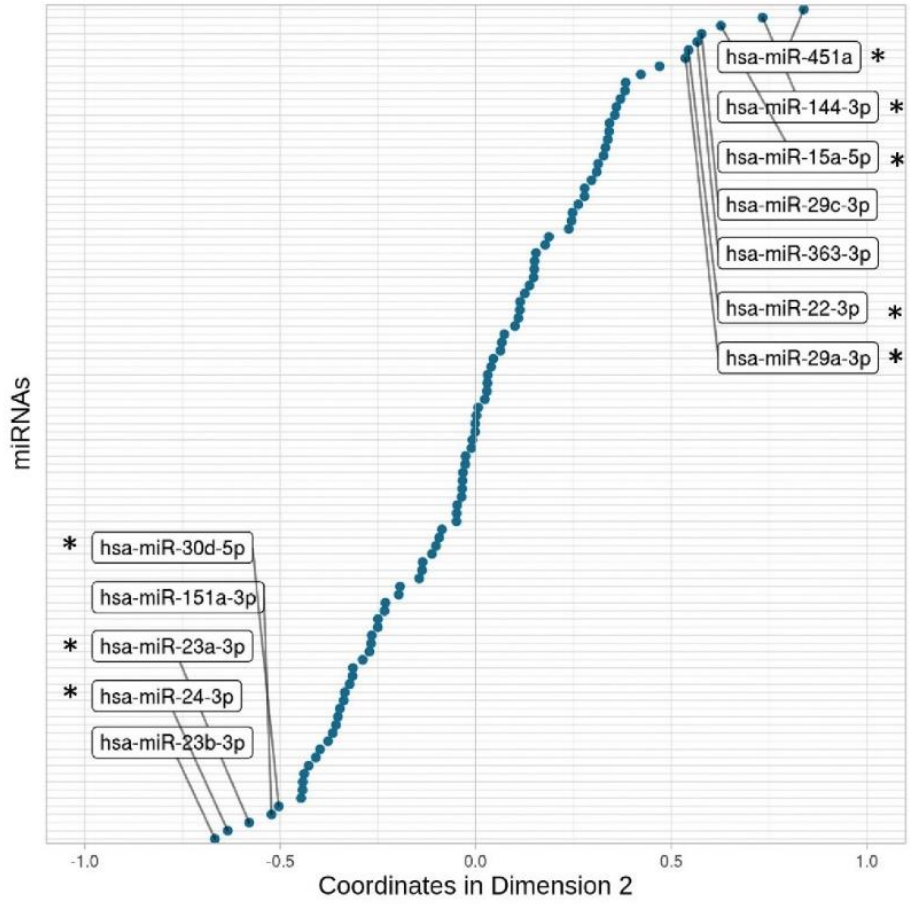


Figure 3. Coordinates of each of the 103 miRNAs in the second dimension. The miRNAs with coordinates less than −0.5 and greater than 0.5 are labelled. It represents the miRNAs that are most stable (negative x axis) and susceptible (positive x axis) to haemolysis. Among them, asterisks show the ones for which previous studies support their susceptibility or stability against haemolysis.

Table 1. Enrichment analysis of *cis*-miR-eQTLs in regulatory regions. Frequencies expected by chance (non-miR-eQTLs) and observed in the 158 *cis*-miR-eQTLs identified, for both enhancers and promoters of each tissue, and the significance of the enrichment test.

	Promoters (%)			Enhancers (%)		
Tissue	non-miR-eQTLs	miR-eQTLs	p	non-miR-eQTLs	miR-eQTLs	p
Blood	46.46	50	>0.05	45.96	45	>0.05
Bone	14.91	32.35	$4.29x10^{-03}$	10.6	15	>0.05
Brain	28.59	55.88	$4.28x10^{-04}$	33.34	32.5	>0.05
Breast	16.27	41.18	$8.32x10^{-05}$	24.17	21.25	>0.05
Fat	23.96	52.94	$7.54x10^{-05}$	24.44	31.25	>0.05
Gastro.[1]	35.92	50	>0.05	38.31	45	>0.05
Liver	22.33	38.24	$2.59x10^{-02}$	21.35	35	$2.89x10^{-03}$
Lung	24.04	41.18	$1.93x10^{-02}$	29.93	40	$4.91x10^{-02}$
Muscle	28.97	47.06	$2x10^{-02}$	38.28	43.75	>0.05
Skin	31.33	47.06	$4.8x10^{-02}$	36.02	42.5	>0.05

haemolysis marker is globally established as the most sensitive method to detect haemolysis in miRNA expression analysis, even more than other available measures, such as spectrophotometry [16]. The rule is to discard those samples with a haemolysis marker above 7, consider as medium-risk those in the range 5–7, and low risk when under 5 [17]. In this regard, some studies have identified great differences in the expression of circulating miRNAs among samples with different grades of haemolysis[18]. Thus, not only the high-risk haemolysis samples excluded, but methodological procedures have been developed to analyse the effect of haemolysis in miRNA when proposed as a disease biomarker [19]. However, all of these previous studies analysed the effect on hemolyzed samples in comparison to non-hemolyzed ones. Our results suggest, for the first time, that haemolysis affects the expression of miRNAs even when all of the samples are in the safe range of haemolysis. Therefore, our data suggest that in addition

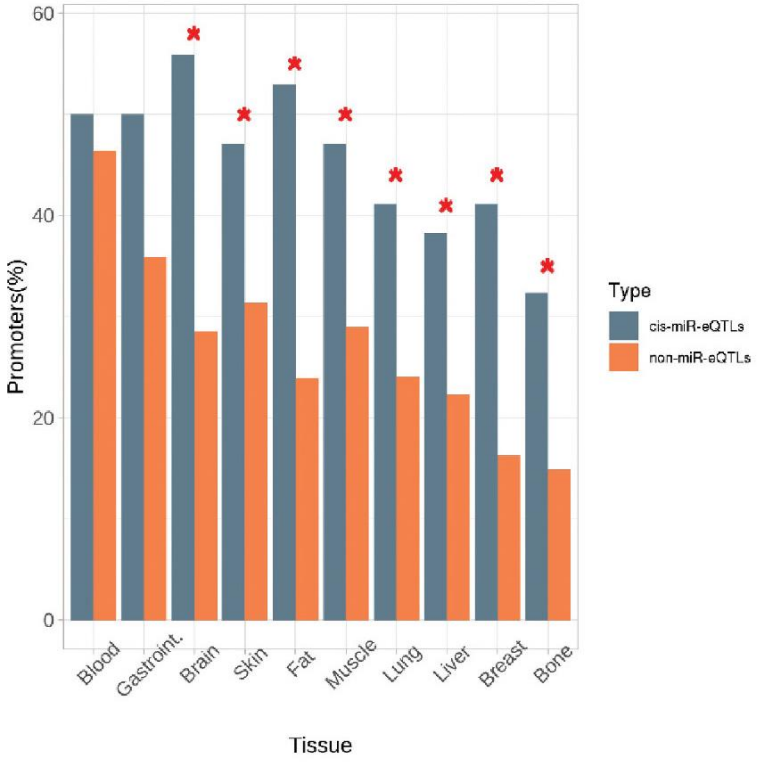


Figure 4. **Tissue enrichment for the *cis*-miR-eQTLs located in promoter regions**. Comparison of the observed frequencies of promoters by tissue and the frequencies expected by chance in the non-miR-eQTLs of the same characteristics.

to being used as a categorical control, the haemolysis marker could be used also as a continuous technical covariate. Using the haemolysis marker as covariate, a ~ 10% of variability would be removed and thus, the consistency across studies may be enhanced.

Furthermore, the analysis of the coordinates of each miRNA in the second PC dimension (Figure 3) identified the miRNAs that are most stable and most susceptible to haemolysis in platelet-poor plasma. The miRNAs in the group of hsa-miR-451a (enriched in red blood cells [20]) included four miRNAs known to be related to haemolysis. First, hsa-miR-144-3p is encoded in the same genetic locus that hsa-miR-451a and the expressions of these two are known to be coordinated [21]. Also, hsa-miR-15a-5p, which is encoded on the genetic cluster hsa-miR-15a/16-1, is highly susceptible to haemolysis [18,19]. For the other two miRNAs in the group, some previous studies have suggested their susceptibility to haemolysis and thus, our results support the implication of hsa-miR-29a [22] and hsa-miR-22-3p [23]. To our knowledge, the two remaining miRNAs, hsa-miR-29 c-3p and hsa-miR-363-3p, have not been related to haemolysis before. It should be noted that some miRNAs known to be related to red blood cells [24] were not into the most affected by haemolysis in our sample, such as hsa-miR-16 (position 25/103 in the ranking by contribution to the second dimension) or hsa-miR-92 (position 35/103). From these results, it seems that hsa-miR-16 and hsa-miR-92 have less influence on haemolysis, or that the methodologies used for their detection have different sensitivities for these two miRNAs. Among the miRNAs that are most stable against haemolysis, besides hsa-miR-23a-3p, we find that previous studies have reported hsa-miR-24-3p and hsa-miR-30d to be unaffected by haemolysis [24,25]. In the course of our literature survey, we have not come across any study that has related the other two stable miRNAs, namely, hsa-miR-151a-3p and hsa-miR-23b-3p, with haemolysis. Overall, we have identified the two subsets of the miRNAs, the most stable and the most susceptible to haemolysis in

the safe range of haemolysis, in platelet-poor plasma. Also, we provide the contribution of each of the 103 miRNAs to this variability (Supplemental Data 2). We expect these data would prove useful for checking any particular miRNA when proposed as a biomarker of a disease, as well as for further studies that analyse the effect of haemolysis on miRNA quantification.

Then, we explored the correlations of blood cell counts with each miRNA and with their PCs. We found that only hsa-miR-150-5p and lymphocyte count reached statistical significance. This association has been reported earlier in an exhaustive study that advised the researchers to be aware of the effect of blood cell counts when measuring circulating miRNAs because some miRNA expression levels could be simply mirroring the corresponding cell count [11]. However, we found that no other miRNA was significantly associated with a blood cell count in our sample. In this sense, a genome-wide mapping of miRNAs in whole blood found no significant differences whether or not the miRNAs were adjusted to the blood cell counts [12]. In accordance, our analysis indicated that, neither the expression of miRNAs in plasma nor their quantification is biased by the blood cell count and therefore, encourage their potential role as circulating biomarkers of certain diseases. Moreover, our results suggest that even for these miRNAs synthetized in a specific blood cell (e.g., hsa-miR-150-5p), miRNA levels do not merely mirror the corresponding cell count but can represent a specific biological process (e.g., hsa-miR-451a has been related to erythrocyte maturation [21,26] and hsa-miR-150-5p to the prognostic of patients with sepsis [27,28]).

Further, we have identified and reported a dataset of 1,323 genetic variants that correspond to 158 *cis*-miR-eQTLs for 14 miRNAs and 16 miRNAs genes. Similar useful datasets have been provided for whole blood or specific cell types [12–14]. We expect that our plasma *cis*-miR-eQTLs dataset will prove to be a useful piece of information for other researches. As in the case of the previous studies, we only obtained *cis*-miR-eQTLs for ~14% of the tested miRNAs. This reflects the complexity of the genetic regulation that underlies miRNA expression.

We found that *cis*-miR-eQTLs were specifically enriched in promoter regions and a specific enrichment in several tissues, whereas not in blood. This result concurs with our expression analysis and suggests that miRNAs expressed in plasma are genetically regulated in different tissues, though not specifically in blood cells. In this regard, how and why miRNAs are released into extracellular biofluids is not clear. While miRNA expression is highly tissue-specific [29] and, therefore, the levels of circulating miRNAs are sometimes poorly associated with their levels in other specific tissues [30], several studies have reported close associations between aberrant circulating miRNA levels and their levels of expression in particular tissues [31,32]. However, a range of theories about their biological basis has been suggested. It is theorized that circulating miRNAs are mere bioproducts of cell activity or cell death, or selectively released as mechanisms of cell-cell communication [9,33,34]. We consider that all of these theories could apply and, although we cannot hypothesize about their function, our results suggest that the origin of circulating miRNAs is genetically controlled by different specific tissues and not especially by the blood cells, which further enhances their value as biomarkers of tissue-specific conditions.

One limitation of our study is that the sample size (n = 104) could limit the ability to detect associations with either genetic variants or with technical and biological features. Thus, some relevant relationships could remain hidden. However, we have included only the proximal genetic variants and used multiple testing corrections to ensure that the results are reliable. Second, we highlight that the results about the relation with blood cell counts must be interpreted only for platelet-poor plasma. It is possible that other results could be found in other biofluids, such as platelet-rich plasma or serum. Also, it should be considered that half of the population included in this study has suffered a thrombotic event. However, the disease has been analysed as a confounding factor throughout the study and none of the highlighted miRNAs are associated with the disease in our cohort [35]. Finally, we acknowledge that all findings need to be validated in independent populations.

In conclusion, the main findings of this study are: (1) the haemolysis marker explains 10% of variability of miRNA expression even in mild-hemolyzed samples and it could be used as a continuous technical covariate, (2) expression of miRNAs in platelet-poor plasma is not widely biased by blood cell counts, (3) a total of 158 *cis*-miR-eQTLs for 14 miRNAs at FDR <0.2 are identified, and (4) the *cis*-miR-eQTLs for plasma miRNAs are enriched in promoter regions, but not specifically in the blood-tissue regions. These results provide valuable knowledge gaining further understanding of the variability in miRNA assays, the genetic regulation of circulating miRNAs, and their potential role as biomarkers of disease.

Methods and patients

Population

This study used the miRNA data of the discovery phase of a venous thrombosis specific research [35]. Therefore, the population was 104 subjects of the 'Genetic Analysis of Idiopathic Thrombophilia' (GAIT-2) cohort [36]. The population included 52 patients who had suffered a thrombotic event in previous years (excluding paraneoplastic thrombosis) and 52 healthy controls, matched for age and sex, and genetically unrelated. Briefly, of the 104 subjects, 68 (65.5%) were females, with an average age of 49.5 years (range 20.3–87.3 years), average BMI of 27.4 and 34 (32.7%) were active smokers.

The GAIT-2 study was reviewed and approved by the Institutional Review Board of the Hospital de la Santa Creu i Sant Pau (Barcelona, Spain). All participants gave written informed consent for themselves and for their minor children. They acknowledge that they could not be identified via the paper and we have ensured their complete anonymity.

miRNA quantification

Before the collection of samples, the anticoagulant, antiplatelet and anti-inflammatory treatments that the subjects had been undergoing, if any, were suspended. Moreover, the samples were collected at least three months after the last thrombotic event and at least one month after any acute inflammatory process. Platelet-poor plasma was collected in citrated tubes, centrifugated (2000 g for 25 minutes at room temperature) and frozen at –80ºC until use.

The isolation of miRNAs was done with miRCURY™ RNA Isolation Kit (Exiqon) and miRNA was reverse transcribed to cDNA using miRCURY™ LNA™ Universal RT (Exiqon). Then, 752 miRNAs were quantified by qPCR in each of the 104 subjects, using miRCURY LNA™ miRNA miRNOME PCR Panel-Human I+ II (Exiqon). Amplification was performed in a LightCycler® 480-Real Time PCR System (Roche). Quality was assayed using UniSp2, UniSp4, and UniSp5 as controls of miRNA extraction, UniSp6 as reverse transcription control, and UniSp3 as inter-plate calibrator. Specificity and contamination were controlled by non-template controls. Haemolysis was assayed using the miRNA haemolysis marker (i.e., dCt hsa-miR-23a-3p – hsa-miR-451a) [17]. Ct values above 37 were considered undetectable. Following manufacturer protocol, raw Ct values were corrected by the inter-plate calibrator (UniSp3) and the most stable value for inter-individual normalization was selected using NormFinder [37].

All the assays met all the quality controls. The 'global mean' of the 40 miRNAs expressed in all the individuals was found as the most suitable value for normalization. Thus, for miRNA a of a subject s, measured in the plate p, the final expression value (dCq) was as follows:

$$Cq_{(a,\,s)} = raw\ Ct_{(a,\,s)} - plate\ calibration\ factor_{(p)},$$

where, plate calibration factor $_{(p)}$ = UniSp3 $_{(p)}$ – mean UniSp3 $_{(all\ plates)}$

$$dCq_{(a,\,s)} = Global\ Mean_{(s)} - Cq_{(a,\,s)},$$

and where, Global Mean $_{(s)}$ = mean Cq value of the 40 miRNAs for individual s.

Blood cell counts

Blood cell counts were quantified on fresh blood using the haematologic analyser Sysmex XE-2100. We included absolute numbers of platelets, leucocytes, lymphocytes, basophils, eosinophils, neutrophils, monocytes (10^9/litre), and erythrocytes (10^{12}/litre).

Statistics of miRNA expression

All statistical analyses were performed using R, version 3.6.1. The miRNAs expressed in more than 90% of the individuals were selected for further analyses, to reduce missing values. The remaining missing values were imputed by the predictive mean matching method, using R package mice [38]. PCA was calculated by Eigen decomposition and the PCs and the corresponding contributions and coordinates for each variable were extracted using the R package FactoMineR [39]. The first four PCs were associated, using Pearson correlations, with technical factors (i.e., haemolysis marker, Unisp2, UniSp3, UniSp4, UniSp5, and Unisp6), and biologic features (i.e., age, sex, BMI, smoking attitude [codified as smokers or non-smokers] and disease condition [if they had suffered a thrombotic event, codified as 0 for 'no' and 1 for 'yes']). Later, the same PCs and the expression of every single miRNA were correlated with the blood cell counts using Pearson correlations. Significance was adjusted for multiple comparisons using FDR [40], with the qvalue R package. Afterwards, for subsequent analyses, miRNA expression values were corrected for the effect of the aforementioned PCs.

Genotyping and identification of *cis*-miR-eQTLs

Genotyping and imputation processes in the GAIT-2 population have been described extensively [36]. Briefly, the genotyping was done with a combination of HumanOmniExpressExome-8v1.2 and HumanCoreExome-12v1.1, and the imputation to the 1000 genomes panel was done using IMPUTE2 [41]. The genetic variants with minor allele frequency (MAF) <5% in the 104 subjects were excluded.

Associations between the expression of each miRNA and its *cis* genetic variants (i.e., those located 1MB around the miRNA gene) were calculated using R package MatrixEQTL [42]. The associations, with age and sex as covariates, were performed using the linear additive model. The associations at FDR <0.2 from the FDR estimates calculated directly by MatrixEQTL, were considered statistically significant. Then, the statistically associated genetic variants were grouped into independent *cis*-miR-eQTLs using a LD threshold of $r^2 = 0.8$. The genetic locations of the miRNA genes were obtained from miRBase [43] v20 in the genome build GRCh37.p5.

Regulatory regions and other genome features

The genomic features regarding the genetics variants identified as *cis*-miR-eQTLs were annotated using R package haploR [44] under the default settings. Enrichment tests were performed for the frequency of active regulatory regions (i.e., enhancers and promoters), based on the 15-state model of ChromHMM [45]. Expected frequencies by chance were calculated by the mean value of 100 random samples of 100 genetic variants bearing the same characteristics. That is, the genetic variants located 1MB around the miRNA genes and with MAF ≥5% (i.e., from those tested but not associated with any miRNA) and named 'non-miR-eQTLs' according to the previous nomenclature [15]. These expected frequencies were compared with the observed frequencies by a two-tailed chi-square test.

Funding

This work was supported by the Government of Catalonia under grants [FI_2017B_00673], (A.R-R.), SGR_1736 (J.M.S.) and CERCA Programme; the Spanish Ministry of Health under grants [FIS_PI14_00582] (J.M.S.) and [ISCIII CP17/00142] (M.S-L.); the Spanish Society of Thrombosis and Hemostasis (SETH); the nonprofit association 'Activa'TT por la Salud'; and STAGO.

Note

a. Gastrointestinal.

Acknowledgments

We thank William H. Stone for reviewing the manuscript, and A. Cardenas and C. Muñoz for their technical support.

Disclosure statement

The authors declare no conflict of interest

ORCID

Alba Rodriguez-Rius http://orcid.org/0000-0002-7700-1496
Angel Martinez-Perez http://orcid.org/0000-0002-5934-1454
Sonia López http://orcid.org/0000-0002-5023-5153
Maria Sabater-Lleal http://orcid.org/0000-0002-0128-379X
Juan Carlos Souto http://orcid.org/0000-0003-2092-5142
José Manuel Soria http://orcid.org/0000-0002-6226-4293

References

[1] Huntzinger E, Izaurralde E. Gene silencing by microRNAs: contributions of translational repression and mRNA decay. Nat Rev Genet. 2011;12(2):99–110.
[2] Wang J, Chen J, Sen S. MicroRNA as Biomarkers and Diagnostics. J Cell Physiol. 2016;231(1):25–30.
[3] Pogribny IP. MicroRNAs as biomarkers for clinical studies. Exp Biol Med. 2018;243(3):283–290.
[4] Saliminejad K, Khorram Khorshid HR, Ghaffari SH. Why have microRNA biomarkers not been translated from bench to clinic? Futur Oncol. 2019;15(8):801–803.
[5] Glinge C, Clauss S, Boddum K, et al. Stability of circulating blood-based microRNAs-Pre-Analytic methodological considerations. PLoS One. 2017;12(2):1–16.
[6] Poel D, Buffart TE, Oosterling-Jansen J, et al. Evaluation of several methodological challenges in circulating miRNA qPCR studies in patients with head and neck cancer. Exp Mol Med. 2018;50(3):e454–e454.
[7] Pritchard CC, Cheng HH, Tewari M. MicroRNA profiling: approaches and considerations. Nat Rev Genet. 2012;13(5):358–369.
[8] McDonald JS, Milosevic D, Reddi HV, et al. Analysis of circulating microRNA: preanalytical and analytical challenges. Clin Chem. 2011;57(6):833–840.
[9] Sohel MH. Extracellular/circulating MicroRNAs: release mechanisms, functions and challenges. Achiev Life Sci. 2016;10(2):175–186.
[10] Binderup HG, Houlind K, Madsen JS, et al. Pre-storage centrifugation conditions have significant impact on measured microRNA levels in biobanked EDTA plasma samples. Biochem Biophys Reports. 2016;7:195–200.
[11] Pritchard CC, Kroh E, Wood B, et al. Blood cell origin of circulating MicroRNAs: a cautionary note for cancer biomarker studies. Cancer Prev Res. 2012;5(3):492LP- 497.
[12] Huan T, Rong J, Liu C, et al. Genome-wide identification of microRNA expression quantitative trait loci. Nat Commun. 2015;6:1–9.
[13] Gamazon ER, Ziliak D, Im HK, et al. Genetic architecture of microRNA expression: implications for the transcriptome and complex traits. Am J Hum Genet. 2012;90(6):1046–1063.
[14] Nikpay M, Beehler K, Valsesia A, et al. Genome-wide identification of circulating-miRNA expression quantitative trait loci reveals the role of several miRNAs in the regulation of cardiometabolic phenotypes. Cardiovasc Res. 2019;115(11):1629–1645.
[15] Budach S, Heinig M, Marsico A. Principles of microRNA regulation revealed through modeling microRNA expression quantitative trait loci. Genetics. 2016;203(4):1629–1640.
[16] Shah JS, Soon PS, Marsh DJ. Comparison of methodologies to detect low levels of hemolysis in serum for accurate assessment of serum microRNAs. PLoS One. 2016;11(4):e0153200.
[17] Blondal T, Jensby Nielsen S, Baker A, et al. Assessing sample and miRNA profile quality in serum and plasma or other biofluids. Methods. 2013;59(1):164–169.
[18] Kirschner MB, Edelman JJB, Kao SC-H, et al. The impact of hemolysis on cell-free microRNA biomarkers. Front Genet. 2013;4:94.
[19] Pizzamiglio S, Zanutto S, Ciniselli CM, et al. A methodological procedure for evaluating the impact of hemolysis on circulating microRNAs. Oncol Lett. 2017;13(1):315–320.
[20] Rasmussen KD, Simmini S, Abreu-Goodger C, et al. The miR-144/451 locus is required for erythroid homeostasis. J Exp Med. 2010;207(7):1351–1358.
[21] Kretov DA, Walawalkar IA, Mora-Martin A, et al. Ago2-dependent processing allows miR-451 to evade the global microRNA turnover elicited during erythropoiesis. Mol Cell. 2020;78(2):317–328.e6.
[22] Yamada A, Cox MA, Gaffney KA, et al. Technical factors involved in the measurement of circulating microRNA biomarkers for the detection of colorectal neoplasia. PLoS One. 2014;9(11):e112481–e112481.
[23] Landoni E, Miceli R, Callari M, et al. Proposal of supervised data analysis strategy of plasma miRNAs from hybridisation array data with an application to assess hemolysis-related deregulation. BMC Bioinformatics. 2015;16(1):388.
[24] Kirschner MB, Kao SC, Edelman JJ, et al. Haemolysis during sample preparation alters microRNA content of plasma. PLoS One. 2011;6(9):e24145–e24145.
[25] Meerson A, Ploug T. Assessment of six commercial plasma small RNA isolation kits using qRT-PCR and electrophoretic separation: higher recovery of microRNA following ultracentrifugation. Biol Methods Protoc. 2016;1(1):1–9.
[26] Pase L, Layton JE, Kloosterman WP, et al. miR-451 regulates zebrafish erythroid maturation in vivo via its target gata2. Blood. 2009;113(8):1794–1804.
[27] Vasilescu C, Rossi S, Shimizu M, et al. MicroRNA fingerprints identify miR-150 as a plasma prognostic marker in patients with sepsis. PLoS One. 2009;4(10):e7405.

[28] Ma Y, Liu Y, Hou H, et al. MiR-150 predicts survival in patients with sepsis and inhibits LPS-induced inflammatory factors and apoptosis by targeting NF-κB1 in human umbilical vein endothelial cells. Biochem Biophys Res Commun. 2018;500(3):828–837.
[29] Liang Y, Ridzon D, Wong L, et al. Characterization of microRNA expression profiles in normal human tissues. BMC Genomics. 2007;8:1–20.
[30] Nagy ZB, Barták BK, Kalmár A, et al. Comparison of circulating miRNAs expression alterations in matched tissue and plasma samples during colorectal cancer progression. Pathol Oncol Res. 2019;25(1):97–105.
[31] Zedan AH, Hansen TF, Assenholt J, et al. MicroRNA expression in tumour tissue and plasma in patients with newly diagnosed metastatic prostate cancer. Tumor Biol. 2018;40(5):1–11.
[32] Xu L, Li M, Wang M, et al. The expression of microRNA-375 in plasma and tissue is matched in human colorectal cancer. BMC Cancer. 2014;14:714.
[33] Turchinovich A, Weiz L, Burwinkel B. Extracellular miRNAs: the mystery of their origin and function. Trends Biochem Sci. 2012;37(11):460–465.
[34] Zhao C, Sun X, Li L. Biogenesis and function of extracellular miRNAs. ExRNA. 2019;1(1):1–9.
[35] Rodriguez-Rius A, Lopez S, Martinez-Perez A, et al. Identification of a plasma microRNA profile associated with venous thrombosis. Arterioscler Thromb Vasc Biol. 2020;May:1392–1399. DOI:10.1161/ATVBAHA.120.314092
[36] Martin-Fernandez L, Ziyatdinov A, Carrasco M, et al. Genetic determinants of thrombin generation and their relation to venous thrombosis: results from the GAIT-2 project. PLoS One. 2016;11(1):1–12.
[37] Andersen CL, Jensen JL, Orntoft TF. Normalization of real-time quantitative reverse transcription-PCR data: a model-based variance estimation approach to identify genes suited for normalization, applied to bladder and colon cancer data sets. Cancer Res. 2004;64(15):5245–5250.
[38] van Buuren S, Groothuis-Oudshoorn K. mice: multivariate imputation by chained equations in R. J Stat Software. 2011 December;1(3). https://www.jstatsoft.org/v045/i03.
[39] Lê S, Josse J, Husson F. FactoMineR: an R package for multivariate analysis. J Stat Software. 2008;1(1). DOI:10.18637/jss.v025.i01.
[40] Storey JD. A direct approach to false discovery rates. J R Stat Soc Ser B Stat Methodol. 2002;64(Part 3):479–498.
[41] Howie BN, Donnelly P, Marchini J. A flexible and accurate genotype imputation method for the next generation of genome-wide association studies. PLoS Genet. 2009;5:6.
[42] Shabalin AA. Matrix eQTL: ultra fast eQTL analysis via large matrix operations. Bioinformatics. 2012;28(10):1353–1358.
[43] Griffiths-jones S, Grocock RJ, Dongen SV, et al. miRBase: microRNA sequences, targets and gene nomenclature. Nucleic Acids Res. 2006;34:140–144.
[44] Zhbannikov IY, Arbeev K, Yashin AI. haploR: an R-package for querying web-based annotation tools. F1000Res. 2017;6(1):97.
[45] Ernst J, Kellis M. ChromHMM: automating chromatin-state discovery and characterization. Nat Methods. 2012;9:215.

5. Results

First, in relation to the Objective 1, we analyzed the differential expression of plasma miRNAs in VT patients in contrast to healthy controls, and the results were reported in Article 1.

We conducted a discovery phase that involved 52 VT patients and 52 controls (matched by age and sex, and not genetically related) from the GAIT-2 cohort, in whom 752 plasma miRNAs were measured by qPCR. The 103 miRNAs expressed in at least 90% of subjects were associated using Pearson's correlations with VT and clinical VT phenotypes. Nine miRNAs were associated with VT disease at nominal level ($p < 0.05$): hsa-miR-23b-3p, hsa-miR-27a-3p, hsa-miR-548c-5p, hsa-miR-221-3p, hsa-miR-197-3p, hsa-miR-320a, hsa-miR-194-5p, hsa-miR-192-5p and hsa-miR-885-5p. Also, we selected seven miRNAs that were significantly associated with clinical VT phenotypes of interest: hsa-miR-320b, hsa-miR-342-3p, hsa-miR-146a-5p, hsa-miR-142-3p, hsa-miR-28-3p, hsa-miR-148a-3p and hsa-miR-126-3p.

Therefore, the internal validation phase included 16 miRNAs, which were quantified in the entire GAIT-2 cohort (n=935). Given that this phase included the extended pedigrees of the GAIT-2 Project, all the analyses were conducted using the variance component method to consider the family structure. The heritability of miRNAs varied from 0.1 to 0.38. Only the heritability of two of them did not reach statistical significance (hsa-miR-197-3p and hsa-miR-148a-3p). The differential expression in VT was tested attending to the phenotypic component of the correlations between the expression of each miRNA and VT. Four of them reached statistical significance at false discovery rate (FDR) <0.1: hsa-miR-885-5p, hsa-miR-192-5p, hsa-miR-194-5p and hsa-miR-126-3p. Then, we explored the association of these four miRNAs with clinical VT phenotypes and found that all of

them were significantly associated with some clinical phenotype (e.g., TGT parameters or FVII).

Later, to further explore the suitability of these four miRNAs as biomarkers of VT, we analyzed their discriminatory ability in relation to VT. All the four miRNAs returned significant OR for VT, in a range of 1.31-2.12. We added age, sex and BMI as covariates to remove their effect as confounding factors, and the OR remained significant: hsa-miR-885-5p OR =1.29 ($p = 9x10^{-03}$), hsa-miR-194-5p OR =1.39 ($p = 3.6x10^{-03}$), hsa-miR-192-5p OR =1.46 ($p = 2.5x10^{-04}$) and hsa-miR-126-3p OR =1.72 (p =0.01). The ROC analysis including only the four miRNAs as linear additive predictors for VT outcome, returned an AUC =0.66 (sensitivity 85.7%, specificity 41.1%). When we included also sex and age, the model reached an AUC =0.77 (sensitivity 85.7%, specificity 54.5%). In addition, we tested whether the model improved by also adding the genetic variants included in the Thrombo InCode® kit, and it returned an AUC =0.80 (sensitivity 74.5%, specificity 72.0%). However, this improvement was not significant (De Long test $p < 0.05$).

Finally, the predicted and validated targets of the four miRNAs in the blood coagulation pathway were explored. We found that all of the four miRNAs had predicted or validated targets in this pathway. Their interactions, considering also the associated clinical phenotypes, were graphically represented to shed light on their biological implications.

The four miRNAs identified as associated with VT were then integrated with clinical phenotypes and the expression of genes in blood to further dissect the blood signature of VT, in relation to Objective 2, and the results were reported in Article 2.

This integrative analysis involved the 104 subjects included in the discovery phase in Article 1. Two of them failed in the measurement of gene expression using RNA-seq and thus, final sample size was 102 individuals (51 VT cases and 51 controls). First, 14

clinical phenotypes were associated with the expression of the 260 genes annotated in the blood coagulation pathway. At FDR<0.1, 49 correlations were obtained, which involved 11 phenotypes and 41 genes. Thus, 11 preliminary models were built, which included each phenotype and its correlated genes as linear predictors for VT outcome. To complete the models, we explored whether any of the genes in the models were predicted or validated targets of the any of the four miRNAs and if so, the miRNA was added to the corresponding model. These 11 models were optimized using penalized regression with the LASSO algorithm. After removing the redundant models and those that only kept one feature, five models were obtained. Finally, the accuracy of these models was tested using ROC curves with 1,000 bootstrapping interactions and three of them reached an AUC >0.7.

The first model (AUC =0.78) included vWF and the expression of the *GATA2* gene. The accuracy of the model did not change by attending to sex, age or BMI. However, we found that the inverse correlation among vWF and *GATA2* was significantly stronger in cases than in controls ($\beta_{controls}$ =-0,51; β_{cases} =-0.20; p =3.1x10^{-3}). Given the influence of ABO-blood group in vWF levels, we added it as covariate (codified as categorical O / non-O) and it had not significant effect.

The second model (AUC =0.75) was composed by FIX phenotype, blood expression levels of the genes *ANXA2*, *ENTPD1*, *ILK*, *PDPK1*, *PRKAR1A* and *STXBP3*, and the miRNAs hsa-miR-885-5p and hsa-miR-192-5p. We found that the model was significantly more accurate in the group of high BMI (AUC_{lowBMI} =0,74, $AUC_{highBMI}$ =0.92, p =0.03). The principal component analysis (PCA) revealed that the genes in the model corresponded to two expression clusters, which were anticorrelated. Furthermore, the inverse correlation between *ILK* and *PRKAR1A* genes was stronger in controls than in cases ($\beta_{controls}$ =-0.75, β_{cases} =-0.37, FDR =0.13). The biological network with all the

correlations between phenotypes, genes and miRNAs included in the model was represented graphically in order to allow biological interpretation.

The last model (AUC =0.73) included the expression of the genes *CSRP1* and *LYN* and the miRNAs hsa-miR-192-5p and hsa-miR-885-5p. The association between *CSRP1* and hsa-miR-192-5p reached a nominally significant difference between male and females ($\beta_{females}$ =-0.23, β_{males} =0.19, p =0.03), while no other biological factor affected the accuracy of the model or correlations between the features. We found that 16 genes had been removed from this model in the penalization step. The PCA analysis showed that they fell into two expression clusters and one gene from each group was kept, so they must be interpreted as representatives of these groups.

Finally, we used the miRNA expression data of the discovery phase (103 miRNAs in 104 subjects) to dissect the factors affecting the expression of miRNAs in plasma, and the results were reported in the Article 3, in relation to Objective 3.

We conducted a PCA and found that the first four principal components (PCs) explained 44.2% of the overall variability. These first four PCs were correlated with technical and biological factors: synthetic controls (UniSp2, UniSp4, UniSp5, UniSp6, and UniSp3), hemolysis marker (dCt hsa-miR-23a-3p-hsa-miR-451a), sex, age, BMI, smoking habit and disease condition.

Only one significant (FDR <0.1) association was identified: the second PC, which explained 10.1% of the variability, was strongly correlated with the hemolysis marker (β =0.84, p =2.07x10^{-29}). Of note, all of the samples were into the 'low-risk' of hemolysis category (that is, dCt hemolysis <5). Then, we analyzed the contribution of each miRNA to the second PC. As expected, the hemolysis marker-miRNAs (i.e., hsa-miR-451a and hsa-miR-23a-3p) were among the greater contributors and on opposite coordinates.

However, also other 10 miRNAs were closely related to both them: the miRNAs that are most susceptible and most stable to hemolysis, respectively. In the group of miRNAs most stable against hemolysis, we found: hsa-miR-23a-3p, hsa-miR-30d-5p, hsa-miR-151a-3p, hsa-miR-24-3p and hsa-miR-23b-3p. In contrast, the miRNAs most susceptible to hemolysis were: hsa-miR-451a, hsa-miR-144-3p, hsa-miR-15a-5p, hsa-miR-29c-3p, hsa-miR-363-3p, hsa-miR-22-3p and hsa-miR-29a-3p.

Later, we explored the effect of blood cell counts on the expression of miRNAs. None of the blood cells were significantly associated with any PC. To explore the effect on the individual variability, we tested the single correlations between each of the miRNAs and each blood cell count. Only one significant correlation was obtained: hsa-miR-150-5p and lymphocyte count (β =0.39, p =4.8x10^{-05}, FDR =0.07). Moreover, the same analysis was performed using the residual values of the miRNAs corrected by the first four PCs. Only the above correlation reached statistical significance (β =0.40, p =2.90x10^{-05}, FDR =0.06).

Then, the expression of each miRNA was associated with the genetic variants located 1Mb around the corresponding miRNA gene. At FDR <0.2, 1,323 significant genetic variants were identified. Setting linkage disequilibrium limit at r^2 =0.8, corresponded with 158 unique *cis*-miRNA-expression quantitative trait loci (*cis*-miR-eQTLs) for 14 mature miRNAs and 16 miRNA genes. Thus, *cis*-miR-eQTLs were detected for 14.3 % of the miRNAs. None of the signals was shared between more than one miRNA. Most of the cis-miR-eQTLs were located in intronic regions (62.6%) and 50.63% were also mRNA-eQTLs.

Finally, the positions of the 158 *cis*-miR-eQTLs were analyzed in relation to promoter or enhancer regions determined by histone markers. The frequencies of both enhancer and promoter regions were compared with the expected by chance in genetic variants bearing

the same characteristics. We identified a significant enrichment in promoter regions (expected =15.48%, observed =21.51%; p =0.03). In addition, we analyzed the tissues in which the miR-eQTLs were identified as regulatory regions. Although there was not a global enrichment in enhancers, the lung and liver enhancers were specifically overrepresented. Moreover, while the great majority of tissue-specific promoters were enriched, only blood and gastrointestinal tissues were not significantly overrepresented.

4

DISCUSSION

1. Discussion

First, regarding the Objective 1, we identified that four miRNAs (hsa-miR-885-5p, hsa-miR-192-5p, hsa-miR-194-5p and hsa-miR-126-3p) were differentially expressed in VT cases in contrast to healthy control and demonstrated their suitability as biomarkers for VT risk assessment.

The miRNA that accounted for the greatest increase in VT risk in our study was hsa-miR-126-3p, and previous studies support its involvement in VT. First, Wang et al. [156] identified this miRNA to be differentially expressed in 12 DVT patients in contrast to 12 healthy controls, although it was not replicated in their entire cohort. Also, Meng et al. [171] suggested the role of this miRNA in the recruitment and activation of endothelial cells in the mechanisms for thrombi resolution. These authors suggested this miRNA as a promising therapeutic target in VT, via modulation of the PI3K/Akt signaling pathway. Of note, despite these prior supports, hsa-miR-126-3p is also the miRNA with fewer predicted targets in the blood coagulation pathway. However, its action may be mediated by other related pathways.

Regarding hsa-miR-194-5p, the screening phase reported in the work of Wang et al. [156], which included 12 DVT patients and 12 healthy controls, identified this miRNA as differentially expressed between them. Furthermore, its expression has also been found to reduce fibrinogen production in hepatoma cells [172].

For hsa-miR-192-5p, in our knowledge no previous study has suggested its implication in VT. Of note, it is encoded close (~ 200 nt away) to one of the genes of hsa-miR-194-5p and thus, some shared genetic regulation may be expected. Noteworthy, we found that 16 genes of the blood coagulation pathway are validated targets for hsa-miR-192-5p [173], which reinforces its potential role in VT pathogenesis.

The miRNA profile is completed by hsa-miR-885-5p and in our knowledge, no previous study has related its expression either to VT or to other circulatory affections. However, a previous study found that hsa-miR-885-5p inhibits the expression of vWF in colorectal cancer cells [174]. In this regard, this miRNA was associated with vWF levels in blood in our discovery phase. Also, we found that *F2* and *F2RL* genes are predicted targets of this miRNA and, in the internal validation phase, its expression was significantly associated with thrombin generation and Protein S. Therefore, our results suggest that thrombin generation could be one of the biological mechanisms affected by hsa-miR-885-5p in the context of VT.

The promising discriminatory ability of the four miRNAs evidenced their suitability as biomarkers of VT. It should be noted that, up to now, genetic scores have provided AUC around 0.7. However, these scores improve when considering also clinical risk factors, such as BMI or family history of VT. For example, the score developed by de Haan et al. [47] improved from AUC=0.69 to AUC=0.82 by adding the clinical risk factors. In this regard, we found that the improvement was not significant when adding also the genetic variants included in the Thrombo InCode® kit [33]. However, it should be noted that the GAIT-2 Project excluded all the known genetic causes of VT and consequently, the performance of genetic scores is biased in this population. Therefore, we expect the discriminatory ability to improve further when taking into account known genetic variants and clinical risk factors as well.

Then, it should be highlighted that among the 12 miRNAs selected in the discovery phase but not validated, six of them had been previously associated with VT or related conditions: hsa-miR-320a, hsa-miR-320b, hsa-miR-197-3p, hsa-miR-28-3p, hsa-miR-146a and hsa-miR-23b-3p [154,155,175–177]. Although we only found suggestive associations in the discovery phase, prior support for those miRNAs make them

promising candidates. Therefore, our results encourage further studies to clearly confirm or discard their implication in VT, as well as to explore their role in other specific patients, such as cancer patients who developed VT or patients with recurrent VT.

Overall, the results in Article 1 have led to the identification of a miRNA profile associated with VT and to demonstrate its suitability to assess the risk of VT. We would like to highlight that this study represents a major improvement in current knowledge of miRNAs on VT. In this way, both the sample size and the design (discovery and validation phases, both using qPCR) make of our study the most exhaustive carried out to date in this field. Nevertheless, we acknowledge all the results must be validated in an independent population, as well as the lack of functional studies. However, we have also provided as much information as possible (such as correlations with clinical phenotypes or description of potential targets) in order to assist the design of future functional studies. Also, we consider that these results encourage further research not only on miRNAs, but also on other epigenetic mechanisms. In this sense, our results point out that, beyond being 'complementary' mechanisms, miRNAs have a central role in the development of VT and are promising candidates for improving clinical management of VT risk.

Then, regarding the Objective 2, we have shed light on the blood signature of VT by identifying groups of clinical phenotypes, miRNAs and gene expression levels, which interactions represented pathogenic mechanisms in the context of VT

First, we identified that vWF levels and the expression of *GATA2* in blood were anti-correlated and that the disruption of this correlation represents a prothrombotic signature. The gene *GATA2* encodes an endothelial transcription factor implicated in a variety of hematological disorders [178]. Prior studies have identified that *GATA2* is a transcription factor for vWF in endothelial cells [179], as well as an increased incidence of VT (25%) in patients with *GATA2*-deficiency [178]. Moreover, genes of blood coagulation pathway

were found dysregulated in murine models of *GATA2*-deficiency [180]. On the other hand, the role of vWF as blood biomarker of VT is far known. Noteworthy, a recent study [181] demonstrated that vWF and FVIII are the coagulation factors associated with the highest VT risk. Also, they found that both of them account for most of the risk explained by the rest of coagulation factors usually assayed.

In accordance with these previous studies, our results support the relevance of vWF levels in blood to assess the risk of VT. Furthermore, our results suggest for first time that the expression of *GATA2* gene in blood may be associated with VT through its correlation with vWF levels.

Then, the second model revealed the interactions between two groups of genes, FIX phenotype and two miRNAs. The first group of genes included *ANXA2* and *ILK*, which expression levels were closely related to FIX phenotype. In this regard, *ANXA2* codifies a receptor of the fibrinolytic system named annexin A2, which is a co-receptor for plasminogen and tissue plasminogen activator. Thus, the expression of *ANXA2* is related to plasmin generation [182,183]. In accordance with our results, a previous study reported that the levels of expression of *ANXA2* are dysregulated between patients at low and high risk for VT [49]. Also, a recent study has suggested for first time aberrant levels of annexin A2 in blood as an independent risk factor for VT, through affection to the plasmin generation capacity [184]. These previous studies support our findings, not only regarding the expression of *ANXA2* gene, but also because we found that it was inversely correlated to FIX phenotype, while plasmin is known to reduce both levels and activity of FIX [185]. Plasmin is also known to affect platelet activity [185], while *ILK* gene is as a major determinant for platelet activation needed for thrombus stability *in vivo*, mainly *via* the αIIbβ3 outside-in signaling pathway [186]. In this sense, all of the genes in the second group (*PDK1, ENTPD, PRKAR1A* and *STXBP3)* are involved in the αIIbβ3 outside-in

signaling pathway [187–190]. Also, the expression of *STXBP3* in blood has been reported as dysregulated between patients at low and high risk for VT [49]. Finally, the two miRNAs in the model seem to interact with the second gene group: *PRKAR1A* is a validated target of hsa-miR-192-5p and *ENTPD* a predicted target of hsa-miR-885-5p.

Overall, we hypothesize that the second model uncovered a cross-talk between the fibrinolytic system and the platelet activation through the αIIbβ3 outside-in signaling pathway, linked by FIX levels in blood and its relationship with plasmin generation. Also, it should be noted that the model was significantly more accurate in patients with high BMI, while both pathways have been related to obesity [191,192].

The last model included the expression levels of *LYN* and *CSRP1* genes and the miRNAs hsa-miR-192-5p and hsa-miR-885-5p. First, *LYN* gene encodes a kinase protein involved in platelet activation, mainly affecting platelets receptors, such as receptors for collagen [193] or vWF[194]. On the other hand, whereas we have not found strong evidences about which platelet mechanisms regulates *CSRP1*, most of the genes in the preliminary model have been related also to platelet activity. For example, *CECAM1* has been reported to affect the interaction between collagen and platelets [195]. Furthermore, it should be noted that this model was built from their correlation with functional levels of protein S, which is a cofactor for tissue factor pathway inhibitor released from endothelium and platelets[196]. Therefore, we hypothesize this model is mirroring the activation of platelets via a biological mechanism in which Protein S plays a central role. In regard to the miRNAs that complete the model, it should be pointed out that hsa-miR-192-5p has as predicted targets three of the genes in the preliminary model (*ANO6*, *H3F3A* and *PABPC4*). Given the capability of miRNAs to affect coordinately genes in the same pathway, we hypothesize that this interaction could be one of the pathogenic mechanisms through which hsa-miR-192-5p is implicated in the development of VT.

We acknowledge that our results are limited due to the modest sample size, and that all the results must be validated in an independent population. However, we consider that these results point out three main fields. First, our results evidence the utility of integrating different biological layers ('multi-omic' studies) to unravel pathogenic mechanisms. In this way, considering other intermediate phenotypes and applying other statistical models (clustering, classifications, etc.) would provide other complementary results that could help increase our knowledge about VT. Second, despite the number of features included and after a restrictive penalization strategy, two of the three models included miRNAs, evidencing their central role in the biology underlying VT. Finally, it should be noted that all of the models are someway related to platelet activity, in accordance with recent studies that have highlighted the relevance of platelets in VT [197,198]. Furthermore, it is also in line with current trend in VT research, in which other mechanisms beyond the coagulation cascade (e.g., platelets, immunology, inflammation) are gaining relevance not only for understanding VT, but also its relationship with other related traits, such as arterial thrombosis.

Later, in relation to Objective 3, we have shed light on the factors that affect the expression of miRNAs in plasma and on their genetic regulation.

Our results showed that the hemolysis marker accounted ~10% of the variability in the expression of miRNAs in plasma. While some technical and analytical parameters regarding miRNA quantification are not standardized yet [144,199,200], hemolysis marker is globally established as the most sensitive method to detect hemolysis in miRNA expression analysis [201]. In this regard, some studies have identified great differences in the expression of circulating miRNAs among samples with different grades of hemolysis [202]. Also, procedures have been developed to analyze the effect of hemolysis in a given miRNA when proposed as a disease biomarker [203]. However, all of these

previous studies analyzed the effect on hemolyzed samples in comparison to non-hemolyzed ones. Our results suggest, for the first time, that hemolysis affects the expression of miRNAs even when all of the samples are in the safe range of hemolysis. Therefore, our data suggest that in addition to being used as a categorical control (i.e., those samples with hemolysis marker above 7 must be excluded), the hemolysis marker could be used also as a continuous technical covariate. In this way, using the hemolysis marker as covariate would remove ~10% of variability and thus, the consistency across studies may be enhanced.

Then, we have identified two subsets of miRNAs, the most stable and the most susceptible to hemolysis, in platelet-poor plasma samples in the safe range of hemolysis. The miRNAs in the group of hsa-miR-451a (enriched in red blood cells [204]) included four miRNAs known to be related to hemolysis. First, hsa-miR-144-3p is encoded in the same genetic locus that hsa-miR-451a and the expressions of these two are known to be coordinated [205]. Also, hsa-miR-15a-5p, which is encoded on the genetic cluster hsa-miR-15a/16-1, is highly susceptible to hemolysis [202,203]. For the other two miRNAs in the group, some previous studies have suggested their susceptibility to hemolysis: hsa-miR-29a [206] and hsa-miR-22-3p [207]. To our knowledge, the two remaining miRNAs, hsa-miR-29c-3p and hsa-miR-363-3p, have not been related to hemolysis before. Among the miRNAs that are most stable against hemolysis, besides hsa-miR-23a-3p, we find that previous studies have reported hsa-miR-24-3p and hsa-miR-30d to be unaffected by hemolysis [208,209]. On the other hand, no previous study has reported the stability of the remaining two miRNAs against hemolysis (hsa-miR-151a-3p and hsa-miR-23b-3p). We expect these data would be useful for checking any particular miRNA when proposed as a biomarker and to studies regarding the effect of hemolysis in miRNA expression.

Then, we explored the correlations of blood cell counts with each miRNA and with their PCs. We found that only hsa-miR-150-5p and lymphocyte count reached statistical significance. This association has been reported earlier in an study that advised the researchers to be aware of the effect of blood cell counts when measuring circulating miRNAs because some miRNA expression levels could be simply mirroring the corresponding cell count [147]. However, we found that no other miRNA was significantly associated with a blood cell count in our sample. In this sense, a genome-wide mapping of miRNAs in whole blood found no significant differences whether or not the miRNAs were adjusted to the blood cell counts [210]. In accordance, our analysis indicated that, neither the expression of miRNAs in plasma nor their quantification is biased by the blood cell count. Moreover, our results suggest that even for miRNAs synthetized in a specific blood cell, their expression levels are able to represent a specific biological process (e.g., hsa-miR-451a role in erythrocyte maturation [205,211] or hsa-miR-150-5p implication in sepsis [212,213]) rather than being a merely mirror of the corresponding cell count. Overall, these results encourage the role of circulating miRNAs as biomarkers of certain tissue-specific conditions.

Further, we have identified 1,323 genetic variants that correspond to 158 *cis*-miR-eQTLs for 14 miRNAs and 16 miRNAs genes. Similar useful datasets have been provided for whole blood or specific cell types [210,214,215]. We expect our plasma *cis*-miR-eQTLs dataset would be a useful piece of information for other researchers. The complexity of the genetic regulation of miRNAs become evidenced because we only obtained *cis*-miR-eQTLs for ~14% of the tested miRNAs, according to previous genome-wide mapping studies [210,215]

We found that *cis*-miR-eQTLs were enriched in promoter regions and a specific enrichment in several tissues, whereas not in blood. In this regard, miRNA expression is

known to be highly tissue-specific [216]. Therefore, the levels of circulating miRNAs are sometimes poorly associated with their levels in other specific tissues [217]. However, other authors have found close associations between aberrant circulating miRNA levels and their levels of expression in particular tissues [218,219]. A range of theories about their biological basis have been suggested, from understanding circulating miRNAs as mere bioproducts of cell death to considering them as mechanism of cell-cell communication [120,130,220]. In this context, our results suggest that the expression of circulating miRNAs is genetically controlled by different specific tissues and not especially by the blood cells and thus, encourage their value as biomarkers of tissue-specific conditions

5

CONCLUSIONS

1. Conclusions

1. We have identified that four plasma miRNAs are differentially expressed in plasma of VT patients in contrast to non-VT controls and that they are associated with clinical VT phenotypes. We have demonstrated that this miRNA profile is of potential clinical use for VT risk assessment.
2. We have identified three groups of intermediate phenotypes which expression is of relevance in VT biological signature. These groups revealed the importance of platelet-related processes, as well as of miRNAs, in the biological mechanisms underlying VT development.
3. We have identified that the hemolysis marker could be used as a continuous technical covariate in order to increase consistency across miRNA studies. Also, we have shed light on the genetic regulation of the expression of circulating miRNAs and encourage their role as biomarkers of tissue-specific conditions.

.

References

[1] H.M. Phillippe, Overview of venous thromboembolism., Am. J. Manag. Care. 23 (2017) S376–S382.

[2] I.A. Næss, S.C. Christiansen, P. Romundstad, S.C. Cannegieter, F.R. Rosendaal, J. Hammerstrøm, Incidence and mortality of venous thrombosis: A population-based study, J. Thromb. Haemost. 5 (2007) 692–699. doi:10.1111/j.1538-7836.2007.02450.x.

[3] J.A. Heit, Epidemiology of venous thromboembolism, Nat. Rev. Cardiol. 12 (2015) 464–474. doi:10.1038/nrcardio.2015.83.

[4] J.F. Timp, S.K. Braekkan, H.H. Versteeg, S.C. Cannegieter, Epidemiology of cancer-associated venous thrombosis, Blood. 122 (2013) 1712–1723. doi:10.1182/blood-2013-04-460121.

[5] J.A. Heit, L.J. 3rd Melton, C.M. Lohse, T.M. Petterson, M.D. Silverstein, D.N. Mohr, W.M. O'Fallon, Incidence of venous thromboembolism in hospitalized patients vs community residents., Mayo Clin. Proc. 76 (2001) 1102–1110. doi:10.4065/76.11.1102.

[6] M.G. Beckman, W.C. Hooper, S.E. Critchley, T.L. Ortel, Venous Thromboembolism. A Public Health Concern, Am. J. Prev. Med. 38 (2010) S495–S501. doi:10.1016/j.amepre.2009.12.017.

[7] T.L. Haematology, Thromboembolism: an under appreciated cause of death, Lancet Haematol. 2 (2015) e393. doi:10.1016/S2352-3026(15)00202-1.

[8] S.R. Kahn, A.J. Comerota, M. Cushman, N.S. Evans, J.S. Ginsberg, N.A. Goldenberg, D.K. Gupta, P. Prandoni, S. Vedantham, M.E. Walsh, J.I. Weitz, The postthrombotic syndrome: Evidence-based prevention, diagnosis, and treatment strategies: A scientific statement from the American heart association, Circulation. 130 (2014) 1636–1661. doi:10.1161/CIR.0000000000000130.

[9] A.C. Spyropoulos, J. Lin, Direct medical costs of venous thromboembolism and subsequent hospital readmission rates: An administrative claims analysis from 30 managed care organizations, J. Manag. Care Pharm. 13 (2007) 475–486. doi:10.18553/jmcp.2007.13.6.475.

[10] K.P. Cohoon, C.L. Leibson, J.E. Ransom, A.A. Ashrani, M.S. Park, T.M. Petterson, K.H. Long, K.R. Bailey, J.A. Heit, Direct medical costs attributable to venous thromboembolism among persons hospitalized for major operation: A population-based longitudinal study, Surg. (United States). 157 (2015) 423–431. doi:10.1016/j.surg.2014.10.005.

[11] S.D. Grosse, R.E. Nelson, K.A. Nyarko, L.C. Richardson, G.E. Raskob, The economic burden of incident venous thromboembolism in the United States: A review of estimated attributable healthcare costs, Thromb. Res. 137 (2016) 3–10. doi:10.1016/j.thromres.2015.11.033.

[12] H. Rasche, Haemostasis and thrombosis: An overview, Eur. Hear. Journal, Suppl. 3 (2001) 3–7. doi:10.1016/S1520-765X(01)90034-3.

[13] S. Palta, R. Saroa, A. Palta, Overview of the coagulation system, Indian J. Anaesth. 58 (2014) 515–523. doi:10.4103/0019-5049.144643.

[14] A.J. Gale, Current understanding of hemostasis, Toxicol. Pathol. 39 (2011) 273–280. doi:10.1177/0192623310389474.

[15] K. Broos, H.B. Feys, S.F. De Meyer, K. Vanhoorelbeke, H. Deckmyn, Platelets at work

in primary hemostasis, Blood Rev. 25 (2011) 155–167. doi:https://doi.org/10.1016/j.blre.2011.03.002.

[16] G.D. Boon, An overview of hemostasis, Toxicol. Pathol. 21 (1993) 170–179. doi:10.1177/019262339302100209.

[17] G. David, R. Thomas, Intrinsic Pathway of Coagulation and Arterial Thrombosis, Arterioscler. Thromb. Vasc. Biol. 27 (2007) 2507–2513. doi:10.1161/ATVBAHA.107.155952.

[18] M. Nigel, T.R. E., K.N. S., Role of the Extrinsic Pathway of Blood Coagulation in Hemostasis and Thrombosis, Arterioscler. Thromb. Vasc. Biol. 27 (2007) 1687–1693. doi:10.1161/ATVBAHA.107.141911.

[19] D. Collen, The plasminogen (fibrinolytic) system., Thromb. Haemost. 82 (1999) 259–270.

[20] F.R. Rosendaal, Venous Thrombosis: The Role of Genes, Environment, and Behavior, Hematology. 2005 (2005) 1–12. doi:10.1182/asheducation.V2005.1.1.1.

[21] F.R. Rosendaal, Causes of Venous Thrombosis, Deep Vein Thromb. Pulm. Embolism. 14 (2009) 1–26. doi:10.1002/9780470745007.ch1.

[22] F.A.J. Anderson, F.A. Spencer, Risk factors for venous thromboembolism., Circulation. 107 (2003) I9-16. doi:10.1161/01.CIR.0000078469.07362.E6.

[23] J.C. Souto, L. Almasy, M. Borrell, M. Garí, E. Martínez, J. Mateo, W.H. Stone, J. Blangero, J. Fontcuberta, Genetic determinants of hemostasis phenotypes in Spanish families, Circulation. 101 (2000) 1546–1551. doi:10.1161/01.CIR.101.13.1546.

[24] T.B. Larsen, H.T. Sorensen, A. Skytthe, S.P. Johnsen, J.W. Vaupel, K. Christensen, Major genetic susceptibility for venous thromboembolism in men: a study of Danish twins., Epidemiology. 14 (2003) 328–332.

[25] J.A. Heit, M.A. Phelps, S.A. Ward, J.P. Slusser, T.M. Petterson, M. De Andrade, Familial segregation of venous thromboembolism, J. Thromb. Haemost. 2 (2004) 731–736. doi:10.1111/j.1538-7933.2004.00660.x.

[26] L. Martin-Fernandez, A. Ziyatdinov, M. Carrasco, J.A. Millon, A. Martinez-Perez, N. Vilalta, H. Brunel, M. Font, A. Hamsten, J.C. Souto, J.M. Soria, Genetic determinants of thrombin generation and their relation to venous thrombosis: Results from the GAIT-2 project, PLoS One. 11 (2016) 1–12. doi:10.1371/journal.pone.0146922.

[27] P.E. Morange, P. Suchon, D.A. Trégouët, Genetics of venous thrombosis: Update in 2015, Thromb. Haemost. 114 (2015) 910–919. doi:10.1160/TH15-05-0410.

[28] D. Klarin, E. Busenkell, R. Judy, J. Lynch, M. Levin, J. Haessler, K. Aragam, M. Chaffin, M. Haas, S. Lindström, T.L. Assimes, J. Huang, K. Min Lee, Q. Shao, J.E. Huffman, C. Kabrhel, Y. Huang, Y. V Sun, M. Vujkovic, D. Saleheen, D.R. Miller, P. Reaven, S. DuVall, W.E. Boden, S. Pyarajan, A.P. Reiner, D.-A. Trégouët, P. Henke, C. Kooperberg, J.M. Gaziano, J. Concato, D.J. Rader, K. Cho, K.-M. Chang, P.W.F. Wilson, N.L. Smith, C.J. O'Donnell, P.S. Tsao, S. Kathiresan, A. Obi, S.M. Damrauer, P. Natarajan, I. Consortium, V.A.M.V. Program, Genome-wide association analysis of venous thromboembolism identifies new risk loci and genetic overlap with arterial vascular disease, Nat. Genet. 51 (2019) 1574–1579. doi:10.1038/s41588-019-0519-3.

[29] J.L. Kujovich, Factor V Leiden thrombophilia, Genet. Med. 13 (2011) 1–16. doi:10.1097/GIM.0b013e3181faa0f2.

[30] M.M. Jadaon, Epidemiology of Prothrombin G20210A Mutation in the Mediterranean Region., Mediterr. J. Hematol. Infect. Dis. 3 (2011) e2011054.

doi:10.4084/MJHID.2011.054.

[31] J. O'Donnell, F.E. Boulton, R.A. Manning, M.A. Laffan, Amount of H antigen expressed on circulating von Willebrand factor is modified by ABO blood group genotype and is a major determinant of plasma von Willebrand factor antigen levels., Arterioscler. Thromb. Vasc. Biol. 22 (2002) 335–341. doi:10.1161/hq0202.103997.

[32] J.P. Vandenbroucke, T. Koster, E. Briet, P.H. Reitsma, R.M. Bertina, F.R. Rosendaal, Increased risk of venous thrombosis in oral-contraceptive users who are carriers of factor V Leiden mutation., Lancet (London, England). 344 (1994) 1453–1457. doi:10.1016/s0140-6736(94)90286-0.

[33] J.M. Soria, P.E. Morange, J. Vila, J.C. Souto, M. Moyano, D.A. Trégouët, J. Mateo, N. Saut, E. Salas, R. Elosua, Multilocus genetic risk scores for venous thromboembolism risk assessment, J. Am. Heart Assoc. 3 (2014) 1–10. doi:10.1161/JAHA.114.001060.

[34] D.M. Witt, R. Nieuwlaat, N.P. Clark, J. Ansell, A. Holbrook, J. Skov, N. Shehab, J. Mock, T. Myers, F. Dentali, M.A. Crowther, A. Agarwal, M. Bhatt, R. Khatib, J.J. Riva, Y. Zhang, G. Guyatt, American Society of Hematology 2018 guidelines for management of venous thromboembolism: Optimal management of anticoagulation therapy, Blood Adv. 2 (2018) 3257–3291. doi:10.1182/bloodadvances.2018024893.

[35] L. Mazzolai, V. Aboyans, W. Ageno, G. Agnelli, A. Alatri, R. Bauersachs, M.P.A. Brekelmans, H.R. Büller, A. Elias, D. Farge, S. Konstantinides, G. Palareti, P. Prandoni, M. Righini, A. Torbicki, C. Vlachopoulos, M. Brodmann, Diagnosis and management of acute deep vein thrombosis: A joint consensus document from the European Society of Cardiology working groups of aorta and peripheral vascular diseases and pulmonary circulation and right ventricular function, Eur. Heart J. 39 (2018) 4208–4218. doi:10.1093/eurheartj/ehx003.

[36] National Institute for Health and Care Excellence [NICE], Venous thromboembolic diseases: diagnosis, management and thrombophilia testing (NICE clinical guideline [CG144]), (2018). https://www.nice.org.uk/guidance/cg144.

[37] P.S. Wells, J. Hirsh, D.R. Anderson, A.W. Lensing, G. Foster, C. Kearon, J. Weitz, R. D'Ovidio, A. Cogo, P. Prandoni, Accuracy of clinical assessment of deep-vein thrombosis., Lancet (London, England). 345 (1995) 1326–1330. doi:10.1016/s0140-6736(95)92535-x.

[38] P.S. Wells, D.R. Anderson, M. Rodger, M. Forgie, C. Kearon, J. Dreyer, G. Kovacs, M. Mitchell, B. Lewandowski, M.J. Kovacs, Evaluation of D-Dimer in the Diagnosis of Suspected Deep-Vein Thrombosis, N. Engl. J. Med. 349 (2003) 1227–1235. doi:10.1056/NEJMoa023153.

[39] C. Wittram, M.M. Maher, A.J. Yoo, M.K. Kalra, J.-A.O. Shepard, T.C. McLoud, CT Angiography of Pulmonary Embolism: Diagnostic Criteria and Causes of Misdiagnosis, RadioGraphics. 24 (2004) 1219–1238. doi:10.1148/rg.245045008.

[40] J.A. Heit, A. Ashrani, D.J. Crusan, R.D. McBane, T.M. Petterson, K.R. Bailey, Reasons for the persistent incidence of venous thromboembolism, Thromb. Haemost. 117 (2017) 390–400. doi:10.1160/TH16-07-0509.

[41] L.J.J. Scheres, W.M. Lijfering, S.C. Cannegieter, Current and future burden of venous thrombosis: Not simply predictable, Res. Pract. Thromb. Haemost. 2 (2018) 199–208. doi:10.1002/rth2.12101.

[42] A. Delluc, C. Tromeur, F. Le Ven, M. Gouillou, N. Paleiron, L. Bressollette, M. Nonent, P.-Y. Salaun, K. Lacut, C. Leroyer, G. Le Gal, F. Couturaud, D. Mottier, Current incidence of venous thromboembolism and comparison with 1998: a community-based study in

Western France., Thromb. Haemost. 116 (2016) 967–974. doi:10.1160/TH16-03-0205.

[43] M.N. Raber, Coagulation Tests., in: H.K. Walker, W.D. Hall, J.W. Hurst (Eds.), Boston, 1990.

[44] T. Baglin, The measurement and application of thrombin generation., Br. J. Haematol. 130 (2005) 653–661. doi:10.1111/j.1365-2141.2005.05612.x.

[45] H. de Ronde, R.M. Bertina, Laboratory diagnosis of APC-resistance: a critical evaluation of the test and the development of diagnostic criteria., Thromb. Haemost. 72 (1994) 880–886.

[46] D.-A. Trégouët, S. Heath, N. Saut, C. Biron-Andreani, J.-F. Schved, G. Pernod, P. Galan, L. Drouet, D. Zelenika, I. Juhan-Vague, M.-C. Alessi, L. Tiret, M. Lathrop, J. Emmerich, P.-E. Morange, Common susceptibility alleles are unlikely to contribute as strongly as the FV and ABO loci to VTE risk: results from a GWAS approach, Blood. 113 (2009) 5298–5303. doi:10.1182/blood-2008-11-190389.

[47] H.G. De Haan, I.D. Bezemer, C.J.M. Doggen, S. Le Cessie, P.H. Reitsma, A.R. Arellano, C.H. Tong, J.J. Devlin, L.A. Bare, F.R. Rosendaal, C.Y. Vossen, Multiple SNP testing improves risk prediction of first venous thrombosis, Blood. 120 (2012) 656–663. doi:10.1182/blood-2011-12-397752.

[48] D.A. Lewis, G.J. Stashenko, O.M. Akay, L.I. Price, K. Owzar, G.S. Ginsburg, J.-T. Chi, T.L. Ortel, Whole blood gene expression analyses in patients with single versus recurrent venous thromboembolism, Thromb. Res. 128 (2011) 536—540. doi:10.1016/j.thromres.2011.06.003.

[49] D.A. Lewis, S. Suchindran, M.G. Beckman, W.C. Hooper, A.M. Grant, J.A. Heit, M. Manco-Johnson, S. Moll, C.S. Philipp, K. Kenney, C. De Staercke, M.E. Pyle, J.T. Chi, T.L. Ortel, Whole blood gene expression profiles distinguish clinical phenotypes of venous thromboembolism, Thromb. Res. 135 (2015) 659–665. doi:10.1016/j.thromres.2015.02.003.

[50] S. Lindström, L. Wang, E.N. Smith, W. Gordon, A. van Hylckama Vlieg, M. de Andrade, J.A. Brody, J.W. Pattee, J. Haessler, B.M. Brumpton, D.I. Chasman, P. Suchon, M.-H. Chen, C. Turman, M. Germain, K.L. Wiggins, J. MacDonald, S.K. Braekkan, S.M. Armasu, N. Pankratz, R.D. Jackson, J.B. Nielsen, F. Giulianini, M.K. Puurunen, M. Ibrahim, S.R. Heckbert, S.M. Damrauer, P. Natarajan, D. Klarin, T.M.V. Program, P.S. de Vries, M. Sabater-Lleal, J.E. Huffman, T.C.H.W. Group, T.K. Bammler, K.A. Frazer, B.M. McCauley, K. Taylor, J.S. Pankow, A.P. Reiner, M.E. Gabrielsen, J.-F. Deleuze, C.J. O'Donnell, J. Kim, B. McKnight, P. Kraft, J.-B. Hansen, F.R. Rosendaal, J.A. Heit, B.M. Psaty, W. Tang, C. Kooperberg, K. Hveem, P.M. Ridker, P.-E. Morange, A.D. Johnson, C. Kabrhel, D.-A. Trégouët, N.L. Smith, on behalf of the I. Consortium, Genomic and transcriptomic association studies identify 16 novel susceptibility loci for venous thromboembolism, Blood. 134 (2019) 1645–1657. doi:10.1182/blood.2019000435.

[51] A. Gusev, A. Ko, H. Shi, G. Bhatia, W. Chung, B.W.J.H. Penninx, R. Jansen, E.J.C. de Geus, D.I. Boomsma, F.A. Wright, P.F. Sullivan, E. Nikkola, M. Alvarez, M. Civelek, A.J. Lusis, T. Lehtimaki, E. Raitoharju, M. Kahonen, I. Seppala, O.T. Raitakari, J. Kuusisto, M. Laakso, A.L. Price, P. Pajukanta, B. Pasaniuc, Integrative approaches for large-scale transcriptome-wide association studies., Nat. Genet. 48 (2016) 245–252. doi:10.1038/ng.3506.

[52] C. Gelfi, A. Vigano, M. Ripamonti, R. Wait, S. Begum, E. Biguzzi, G. Castaman, E.M. Faioni, A proteomic analysis of changes in prothrombin and plasma proteins associated with the G20210A mutation., Proteomics. 4 (2004) 2151–2159. doi:10.1002/pmic.200300724.

[53] E. Ramacciotti, A.E. Hawley, S.K. Wrobleski, D.D. Myers, J.R. Strahler, P.C. Andrews, K.E. Guire, P.K. Henke, T.W. Wakefield, Proteomics of microparticles after deep venous thrombosis, Thromb. Res. 125 (2010). doi:10.1016/j.thromres.2010.01.019.

[54] A. Stachowicz, J. Siudut, M. Suski, R. Olszanecki, R. Korbut, A. Undas, J.R. Wiśniewski, Optimization of quantitative proteomic analysis of clots generated from plasma of patients with venous thromboembolism, Clin. Proteomics. 14 (2017) 38. doi:10.1186/s12014-017-9173-x.

[55] S.B. Jensen, K. Hindberg, T. Solomon, E.N. Smith, J.D. Lapek, D.J. Gonzalez, N. Latysheva, K.A. Frazer, S.K. Brækkan, J.B. Hansen, Discovery of novel plasma biomarkers for future incident venous thromboembolism by untargeted synchronous precursor selection mass spectrometry proteomics, J. Thromb. Haemost. 16 (2018) 1763–1774. doi:10.1111/jth.14220.

[56] C. Deans, K.A. Maggert, What do you mean, "Epigenetic"?, Genetics. 199 (2015) 887–896. doi:10.1534/genetics.114.173492.

[57] G. Benincasa, D. Costa, T. Infante, R. Lucchese, F. Donatelli, C. Napoli, Interplay between genetics and epigenetics in modulating the risk of venous thromboembolism: A new challenge for personalized therapy, Thromb. Res. 177 (2019) 145–153. doi:10.1016/j.thromres.2019.03.008.

[58] M.D. Patsouras, P.G. Vlachoyiannopoulos, Evidence of epigenetic alterations in thrombosis and coagulation: A systematic review., J. Autoimmun. 104 (2019) 102347. doi:10.1016/j.jaut.2019.102347.

[59] D. Aïssi, J. Dennis, M. Ladouceur, V. Truong, N. Zwingerman, A. Rocanin-Arjo, M. Germain, T.A. Paton, P.E. Morange, F. Gagnon, D.A. Trégouët, Genome-Wide investigation of DNA methylation marks associated with FV Leiden mutation, PLoS One. 9 (2014). doi:10.1371/journal.pone.0108087.

[60] L.P. Breitling, R. Yang, B. Korn, B. Burwinkel, H. Brenner, Tobacco-smoking-related differential DNA methylation: 27K discovery and replication., Am. J. Hum. Genet. 88 (2011) 450–457. doi:10.1016/j.ajhg.2011.03.003.

[61] K.J. Dick, C.P. Nelson, L. Tsaprouni, J.K. Sandling, D. Aissi, S. Wahl, E. Meduri, P.-E. Morange, F. Gagnon, H. Grallert, M. Waldenberger, A. Peters, J. Erdmann, C. Hengstenberg, F. Cambien, A.H. Goodall, W.H. Ouwehand, H. Schunkert, J.R. Thompson, T.D. Spector, C. Gieger, D.-A. Tregouet, P. Deloukas, N.J. Samani, DNA methylation and body-mass index: a genome-wide analysis., Lancet (London, England). 383 (2014) 1990–1998. doi:10.1016/S0140-6736(13)62674-4.

[62] S. Aydin, Can vitamin K synthesis altered by dysbiosis of microbiota be blamed in the etiopathogenesis of venous thrombosis?, Biosci. Microbiota, Food Heal. 36 (2017) 73–74. doi:10.12938/bmfh.17-007.

[63] K. Kiouptsi, W. Ruf, C. Reinhardt, Microbiota-Derived Trimethylamine, Circ. Res. 123 (2018) 1112–1114. doi:10.1038/nm3145.

[64] A. Santamaria, J. Mateo, I. Tirado, A. Oliver, R. Belvis, J. Marti-Fabregas, R. Felices, J.M. Soria, J.C. Souto, J. Fontcuberta, Homozygosity of the T allele of the 46 C->T polymorphism in the F12 gene is a risk factor for ischemic stroke in the Spanish population., Stroke. 35 (2004) 1795–1799. doi:10.1161/01.STR.0000133127.68041.a3.

[65] J.C. Souto, L. Almasy, M. Borrell, F. Blanco-Vaca, J. Mateo, J.M. Soria, I. Coll, R. Felices, W. Stone, J. Fontcuberta, J. Blangero, Genetic susceptibility to thrombosis and its relationship to physiological risk factors: the GAIT study. Genetic Analysis of Idiopathic Thrombophilia., Am. J. Hum. Genet. 67 (2000) 1452–9. doi:10.1086/316903.

[66] M. Sabater-Lleal, J.M. Soria, J. Bertranpetit, L. Almasy, J. Blangero, J. Fontcuberta, F. Calafell, Human F7 sequence is split into three deep clades that are related to FVII plasma levels., Hum. Genet. 118 (2006) 741–751. doi:10.1007/s00439-005-0045-5.

[67] M. Lagos-Quintana, R. Rauhut, W. Lendeckel, T. Tuschl, Identification of novel genes coding for small expressed RNAs., Science (80-.). 294 (2001) 853–858. doi:10.1126/science.1064921.

[68] D.P. Bartel, MicroRNAs: Genomics, Biogenesis, Mechanism, and Function, Cell. 116 (2004) 281–297. doi:https://doi.org/10.1016/S0092-8674(04)00045-5.

[69] R.C. Lee, R.L. Feinbaum, V. Ambros, The C. elegans heterochronic gene lin-4 encodes small RNAs with antisense complementarity to lin-14, Cell. 75 (1993) 843–854. doi:https://doi.org/10.1016/0092-8674(93)90529-Y.

[70] B.J. Reinhart, F.J. Slack, M. Basson, A.E. Pasquinelli, J.C. Bettinger, A.E. Rougvie, H.R. Horvitz, G. Ruvkun, The 21-nucleotide let-7 RNA regulates developmental timing in Caenorhabditis elegans., Nature. 403 (2000) 901–906. doi:10.1038/35002607.

[71] R.C. Lee, V. Ambros, An Extensive Class of Small RNAs in Caenorhabditis elegans, Science (80-.). 294 (2001) 862 LP – 864. doi:10.1126/science.1065329.

[72] A. Kozomara, M. Birgaoanu, S. Griffiths-Jones, miRBase: from microRNA sequences to function, Nucleic Acids Res. 47 (2018) D155–D162. doi:10.1093/nar/gky1141.

[73] Y. Lee, M. Kim, J. Han, K.H. Yeom, S. Lee, S.H. Baek, V.N. Kim, MicroRNA genes are transcribed by RNA polymerase II, EMBO J. 23 (2004) 4051–4060. doi:10.1038/sj.emboj.7600385.

[74] A. Rodriguez, S. Griffiths-Jones, J.L. Ashurst, A. Bradley, Identification of mammalian microRNA host genes and transcription units, Genome Res. 14 (2004) 1902–1910. doi:10.1101/gr.2722704.

[75] Y. Altuvia, P. Landgraf, G. Lithwick, N. Elefant, S. Pfeffer, A. Aravin, M.J. Brownstein, T. Tuschl, H. Margalit, Clustering and conservation patterns of human microRNAs, Nucleic Acids Res. 33 (2005) 2697–2706. doi:10.1093/nar/gki567.

[76] P. Ramalingam, J.K. Palanichamy, A. Singh, P. Das, M. Bhagat, M.A. Kassab, S. Sinha, P. Chattopadhyay, Biogenesis of intronic miRNAs located in clusters by independent transcription and alternative splicing, Rna. 20 (2014) 76–87. doi:10.1261/rna.041814.113.

[77] S.S. Ryazansky, V.A. Gvozdev, E. Berezikov, Evidence for post-transcriptional regulation of clustered microRNAs in Drosophila, BMC Genomics. 12 (2011) 371. doi:10.1186/1471-2164-12-371.

[78] A. Manuscript, RNA polymerase II regulating mechanisms, Nature. 461 (2010) 186–192. doi:10.1038/nature08449.Defining.

[79] A.M. Denli, B.B.J. Tops, R.H.A. Plasterk, R.F. Ketting, G.J. Hannon, Processing of primary microRNAs by the Microprocessor complex., Nature. 432 (2004) 231–235. doi:10.1038/nature03049.

[80] A.C. Partin, T.D. Ngo, E. Herrell, B.C. Jeong, G. Hon, Y. Nam, Heme enables proper positioning of Drosha and DGCR8 on primary microRNAs, Nat. Commun. 8 (2017) 1–10. doi:10.1038/s41467-017-01713-y.

[81] Y. Feng, X. Zhang, Q. Song, T. Li, Y. Zeng, Drosha processing controls the specificity and efficiency of global microRNA expression, Biochim. Biophys. Acta. 1809 (2011) 700–707. doi:10.1016/j.bbagrm.2011.05.015.

[82] H. Sperber, A. Beem, S. Shannon, R. Jones, P. Banik, Y. Chen, S. Ku, G. Varani, S. Yao,

H. Ruohola-Baker, miRNA sensitivity to Drosha levels correlates with pre-miRNA secondary structure., RNA. 20 (2014) 621–631. doi:10.1261/rna.043943.113.

[83] S.G. Chaulk, G.L. Thede, O.A. Kent, Z. Xu, E.M. Gesner, R.A. Veldhoen, S.K. Khanna, I.S. Goping, A.M. MacMillan, J.T. Mendell, H.S. Young, R.P. Fahlman, J.N.M. Glover, Role of pri-miRNA tertiary structure in miR-17~92 miRNA biogenesis., RNA Biol. 8 (2011) 1105–1114. doi:10.4161/rna.8.6.17410.

[84] B.N. Davis, A.C. Hilyard, P.H. Nguyen, G. Lagna, A. Hata, Smad proteins bind a conserved RNA sequence to promote microRNA maturation by Drosha, Mol. Cell. 39 (2010) 373–384. doi:10.1016/j.molcel.2010.07.011.

[85] X.L. Li, M.F. Jones, M. Subramanian, A. Lal, Mutant p53 exerts oncogenic effects through microRNAs and their target gene networks, FEBS Lett. 588 (2014) 2610–2615. doi:https://doi.org/10.1016/j.febslet.2014.03.054.

[86] Y.-J. Kim, S.-J. Park, E.Y. Choi, S. Kim, H.J. Kwak, B.C. Yoo, H. Yoo, S.-H. Lee, D. Kim, J.B. Park, J.H. Kim, PTEN modulates miR-21 processing via RNA-regulatory protein RNH1., PLoS One. 6 (2011) e28308. doi:10.1371/journal.pone.0028308.

[87] M. Ha, V.N. Kim, Regulation of microRNA biogenesis, Nat. Rev. Mol. Cell Biol. 15 (2014) 509. http://dx.doi.org/10.1038/nrm3838.

[88] S.M. Jafarnejad, G.S. Ardekani, M. Ghaffari, M. Martinka, G. Li, Sox4-mediated Dicer expression is critical for suppression of melanoma cell invasion., Oncogene. 32 (2013) 2131–2139. doi:10.1038/onc.2012.239.

[89] C. Levy, M. Khaled, K.C. Robinson, R.A. Veguilla, P.-H. Chen, S. Yokoyama, E. Makino, J. Lu, L. Larue, F. Beermann, L. Chin, M. Bosenberg, J.S. Song, D.E. Fisher, Lineage-specific transcriptional regulation of DICER by MITF in melanocytes., Cell. 141 (2010) 994–1005. doi:10.1016/j.cell.2010.05.004.

[90] S. Singh, S.C. Bevan, K. Patil, D.C. Newton, P.A. Marsden, Extensive variation in the 5'-UTR of Dicer mRNAs influences translational efficiency., Biochem. Biophys. Res. Commun. 335 (2005) 643–650. doi:10.1016/j.bbrc.2005.07.138.

[91] Z. Paroo, X. Ye, S. Chen, Q. Liu, Phosphorylation of the human microRNA-generating complex mediates MAPK/Erk signaling., Cell. 139 (2009) 112–122. doi:10.1016/j.cell.2009.06.044.

[92] H.A. Meijer, E.M. Smith, M. Bushell, Regulation of miRNA strand selection: follow the leader?, Biochem. Soc. Trans. 42 (2014) 1135–1140. doi:10.1042/BST20140142.

[93] C.L. Noland, J.A. Doudna, Multiple sensors ensure guide strand selection in human RNAi pathways, Rna. 19 (2013) 639–648. doi:10.1261/rna.037424.112.

[94] M.A. Havens, A.A. Reich, D.M. Duelli, M.L. Hastings, Biogenesis of mammalian microRNAs by a non-canonical processing pathway, Nucleic Acids Res. 40 (2012) 4626–4640. doi:10.1093/nar/gks026.

[95] G.A. Cipolla, A non-canonical landscape of the microRNA system, Front. Genet. 5 (2014) 1–6. doi:10.3389/fgene.2014.00337.

[96] K. Miyoshi, T. Miyoshi, H. Siomi, Many ways to generate microRNA-like small RNAs: Non-canonical pathways for microRNA production, Mol. Genet. Genomics. 284 (2010) 95–103. doi:10.1007/s00438-010-0556-1.

[97] H. oki Iwakawa, Y. Tomari, The Functions of MicroRNAs: mRNA Decay and Translational Repression, Trends Cell Biol. 25 (2015) 651–665. doi:10.1016/j.tcb.2015.07.011.

[98] A. Eulalio, J. Rehwinkel, M. Stricker, E. Huntzinger, S.-F. Yang, T. Doerks, S. Dorner, P. Bork, M. Boutros, E. Izaurralde, Target-specific requirements for enhancers of decapping in miRNA-mediated gene silencing., Genes Dev. 21 (2007) 2558–2570. doi:10.1101/gad.443107.

[99] J.E. Braun, E. Huntzinger, E. Izaurralde, A molecular link between miRISCs and deadenylases provides new insight into the mechanism of gene silencing by microRNAs., Cold Spring Harb. Perspect. Biol. 4 (2012). doi:10.1101/cshperspect.a012328.

[100] E. Huntzinger, E. Izaurralde, Gene silencing by microRNAs: contributions of translational repression and mRNA decay, Nat Rev Genet. 12 (2011) 99–110. http://dx.doi.org/10.1038/nrg2936.

[101] D.T. Humphreys, B.J. Westman, D.I.K. Martin, T. Preiss, MicroRNAs control translation initiation by inhibiting eukaryotic initiation factor 4E/cap and poly(A) tail function, Proc. Natl. Acad. Sci. U. S. A. 102 (2005) 16961 LP – 16966. doi:10.1073/pnas.0506482102.

[102] T.P. Chendrimada, K.J. Finn, X. Ji, D. Baillat, R.I. Gregory, S.A. Liebhaber, A.E. Pasquinelli, R. Shiekhattar, MicroRNA silencing through RISC recruitment of eIF6., Nature. 447 (2007) 823–828. doi:10.1038/nature05841.

[103] S. Vasudevan, J.A. Steitz, AU-rich-element-mediated upregulation of translation by FXR1 and Argonaute 2., Cell. 128 (2007) 1105–1118. doi:10.1016/j.cell.2007.01.038.

[104] S.S. Truesdell, R.D. Mortensen, M. Seo, J.C. Schroeder, J.H. Lee, O. LeTonqueze, S. Vasudevan, MicroRNA-mediated mRNA translation activation in quiescent cells and oocytes involves recruitment of a nuclear microRNP., Sci. Rep. 2 (2012) 842. doi:10.1038/srep00842.

[105] U.A. Orom, F.C. Nielsen, A.H. Lund, MicroRNA-10a binds the 5'UTR of ribosomal protein mRNAs and enhances their translation., Mol. Cell. 30 (2008) 460–471. doi:10.1016/j.molcel.2008.05.001.

[106] D.P. Bartel, MicroRNAs: Target Recognition and Regulatory Functions, Cell. 136 (2009) 215–233. doi:https://doi.org/10.1016/j.cell.2009.01.002.

[107] S.M. Peterson, J.A. Thompson, M.L. Ufkin, P. Sathyanarayana, L. Liaw, C.B. Congdon, Common features of microRNA target prediction tools, Front. Genet. 5 (2014) 1–10. doi:10.3389/fgene.2014.00023.

[108] Á.L. Riffo-Campos, I. Riquelme, P. Brebi-Mieville, Tools for sequence-based miRNA target prediction: What to choose?, Int. J. Mol. Sci. 17 (2016). doi:10.3390/ijms17121987.

[109] A. Ghoshal, R. Shankar, S. Bagchi, A. Grama, S. Chaterji, MicroRNA target prediction using thermodynamic and sequence curves, BMC Genomics. 16 (2015) 999. doi:10.1186/s12864-015-1933-2.

[110] J.P. Broughton, M.T. Lovci, J.L. Huang, G.W. Yeo, A.E. Pasquinelli, Pairing beyond the Seed Supports MicroRNA Targeting Specificity, Mol. Cell. 64 (2016) 320–333. doi:10.1016/j.molcel.2016.09.004.

[111] V. Agarwal, G.W. Bell, J.W. Nam, D.P. Bartel, Predicting effective microRNA target sites in mammalian mRNAs, Elife. 4 (2015) 1–38. doi:10.7554/eLife.05005.

[112] N. Pinzón, B. Li, L. Martinez, A. Sergeeva, J. Presumey, F. Apparailly, H. Seitz, microRNA target prediction programs predict many false positives, Genome Res. 27 (2017) 234–245. doi:10.1101/gr.205146.116.

[113] J. Shu, Z. Xia, L. Li, E.T. Liang, N. Slipek, D. Shen, J. Foo, S. Subramanian, C.J. Steer, Dose-dependent differential mRNA target selection and regulation by let-7a-7f and miR-

17-92 cluster microRNAs, RNA Biol. 9 (2012) 1275–1287. doi:10.4161/rna.21998.

[114] J.G. Doench, P.A. Sharp, Specificity of microRNA target selection in translational repression, Genes Dev. 18 (2004) 504–511. doi:10.1101/gad.1184404.

[115] T. Saito, P. Sætrom, MicroRNAs – targeting and target prediction, N. Biotechnol. 27 (2010) 243–249. doi:https://doi.org/10.1016/j.nbt.2010.02.016.

[116] R. Ben-Hamo, S. Efroni, MicroRNA regulation of molecular pathways as a generic mechanism and as a core disease phenotype, Oncotarget. 6 (2015) 1594–1604. doi:10.18632/oncotarget.2734.

[117] R. Chhabra, R. Dubey, N. Saini, Cooperative and individualistic functions of the microRNAs in the miR-23a~27a~24-2 cluster and its implication in human diseases, Mol. Cancer. 9 (2010) 232. doi:10.1186/1476-4598-9-232.

[118] J. Tsang, J. Zhu, A. van Oudenaarden, MicroRNA-mediated feedback and feedforward loops are recurrent network motifs in mammals, Mol. Cell. 26 (2007) 753–767. doi:10.1016/j.molcel.2007.05.018.

[119] M.A. Duk, M.G. Samsonova, A.M. Samsonov, Dynamics of miRNA driven feed-forward loop depends upon miRNA action mechanisms, BMC Genomics. 15 (2014) S9. doi:10.1186/1471-2164-15-S12-S9.

[120] M.H. Sohel, Extracellular/Circulating MicroRNAs: Release Mechanisms, Functions and Challenges, Achiev. Life Sci. 10 (2016) 175–186. doi:https://doi.org/10.1016/j.als.2016.11.007.

[121] X. Chen, Y. Ba, L. Ma, X. Cai, Y. Yin, K. Wang, J. Guo, Y. Zhang, J. Chen, X. Guo, Q. Li, X. Li, W. Wang, Y. Zhang, J. Wang, X. Jiang, Y. Xiang, C. Xu, P. Zheng, J. Zhang, R. Li, H. Zhang, X. Shang, T. Gong, G. Ning, J. Wang, K. Zen, J. Zhang, C.-Y. Zhang, Characterization of microRNAs in serum: a novel class of biomarkers for diagnosis of cancer and other diseases, Cell Res. 18 (2008) 997–1006. doi:10.1038/cr.2008.282.

[122] J.D. Arroyo, J.R. Chevillet, E.M. Kroh, I.K. Ruf, C.C. Pritchard, D.F. Gibson, P.S. Mitchell, C.F. Bennett, E.L. Pogosova-Agadjanyan, D.L. Stirewalt, J.F. Tait, M. Tewari, Argonaute2 complexes carry a population of circulating microRNAs independent of vesicles in human plasma, Proc. Natl. Acad. Sci. 108 (2011) 5003 LP – 5008. doi:10.1073/pnas.1019055108.

[123] K.C. Vickers, B.T. Palmisano, B.M. Shoucri, R.D. Shamburek, A.T. Remaley, MicroRNAs are transported in plasma and delivered to recipient cells by high-density lipoproteins, Nat. Cell Biol. 13 (2011) 423–433. doi:10.1038/ncb2210.

[124] H. Valadi, K. Ekström, A. Bossios, M. Sjöstrand, J.J. Lee, J.O. Lötvall, Exosome-mediated transfer of mRNAs and microRNAs is a novel mechanism of genetic exchange between cells, Nat. Cell Biol. 9 (2007) 654–659. doi:10.1038/ncb1596.

[125] Y. Akao, A. Iio, T. Itoh, S. Noguchi, Y. Itoh, Y. Ohtsuki, T. Naoe, Microvesicle-mediated RNA Molecule Delivery System Using Monocytes/Macrophages, Mol. Ther. 19 (2011) 395–399. doi:https://doi.org/10.1038/mt.2010.254.

[126] Z. Zhu, D. Zhang, H. Lee, A.A. Menon, J. Wu, K. Hu, Y. Jin, Macrophage-derived apoptotic bodies promote the proliferation of the recipient cells via shuttling microRNA-221/222., J. Leukoc. Biol. 101 (2017) 1349–1359. doi:10.1189/jlb.3A1116-483R.

[127] M.T.N. Le, P. Hamar, C. Guo, E. Basar, R. Perdigão-Henriques, L. Balaj, J. Lieberman, miR-200-containing extracellular vesicles promote breast cancer cell metastasis, J. Clin. Invest. 124 (2014) 5109–5128. doi:10.1172/JCI75695.

[128] M.A. Mori, R.G. Ludwig, R. Garcia-Martin, B.B. Brandão, C.R. Kahn, Extracellular miRNAs: From Biomarkers to Mediators of Physiology and Disease, Cell Metab. 30 (2019) 656–673. doi:10.1016/j.cmet.2019.07.011.

[129] C. Bar, T. Thum, D. De Gonzalo-calvo, Circulating miRNAs as mediators in cell-to-cell communication, Epigenomics. 11 (2019) 111–113. doi:10.2217/epi-2018-0183.

[130] C. Zhao, X. Sun, L. Li, Biogenesis and function of extracellular miRNAs, ExRNA. 1 (2019) 1–9. doi:10.1186/s41544-019-0039-4.

[131] G.A. Calin, C.D. Dumitru, M. Shimizu, R. Bichi, S. Zupo, E. Noch, H. Aldler, S. Rattan, M. Keating, K. Rai, L. Rassenti, T. Kipps, M. Negrini, F. Bullrich, C.M. Croce, Frequent deletions and down-regulation of micro- RNA genes miR15 and miR16 at 13q14 in chronic lymphocytic leukemia, Proc. Natl. Acad. Sci. 99 (2002) 15524 LP – 15529. doi:10.1073/pnas.242606799.

[132] P.S. Mitchell, R.K. Parkin, E.M. Kroh, B.R. Fritz, S.K. Wyman, E.L. Pogosova-Agadjanyan, A. Peterson, J. Noteboom, K.C. O'Briant, A. Allen, D.W. Lin, N. Urban, C.W. Drescher, B.S. Knudsen, D.L. Stirewalt, R. Gentleman, R.L. Vessella, P.S. Nelson, D.B. Martin, M. Tewari, Circulating microRNAs as stable blood-based markers for cancer detection, Proc. Natl. Acad. Sci. 105 (2008) 10513 LP – 10518. doi:10.1073/pnas.0804549105.

[133] A. Etheridge, I. Lee, L. Hood, D. Galas, K. Wang, Extracellular microRNA: a new source of biomarkers, Mutat. Res. 717 (2011) 85–90. doi:10.1016/j.mrfmmm.2011.03.004.

[134] J. Wang, J. Chen, S. Sen, MicroRNA as Biomarkers and Diagnostics, J. Cell. Physiol. 231 (2016) 25–30. doi:10.1002/jcp.25056.

[135] Y. Li, K. V Kowdley, MicroRNAs in Common Human Diseases, Genomics. Proteomics Bioinformatics. 10 (2012) 246–253. doi:https://doi.org/10.1016/j.gpb.2012.07.005.

[136] D. de Gonzalo-Calvo, E. Iglesias-Gutiérrez, V. Llorente-Cortés, Epigenetic Biomarkers and Cardiovascular Disease: Circulating MicroRNAs, Rev. Esp. Cardiol. 70 (2017) 763–769. doi:10.1016/j.recesp.2017.02.027.

[137] E. Bonneau, B. Neveu, E. Kostantin, G.J. Tsongalis, V. De Guire, How close are miRNAs from clinical practice? A perspective on the diagnostic and therapeutic market, Electron. J. Int. Fed. Clin. Chem. Lab. Med. 30 (2019) 114–127.

[138] P. Mestdagh, N. Hartmann, L. Baeriswyl, D. Andreasen, N. Bernard, C. Chen, D. Cheo, P. D'Andrade, M. DeMayo, L. Dennis, S. Derveaux, Y. Feng, S. Fulmer-Smentek, B. Gerstmayer, J. Gouffon, C. Grimley, E. Lader, K.Y. Lee, S. Luo, P. Mouritzen, A. Narayanan, S. Patel, S. Peiffer, S. Rüberg, G. Schroth, D. Schuster, J.M. Shaffer, E.J. Shelton, S. Silveria, U. Ulmanella, V. Veeramachaneni, F. Staedtler, T. Peters, T. Guettouche, L. Wong, J. Vandesompele, Evaluation of quantitative mirnA expression platforms in the micrornA quality control (mirQC) study, Nat. Methods. 11 (2014) 809–815. doi:10.1038/nMeth.3014.

[139] A. Git, H. Dvinge, M. Salmon-Divon, M. Osborne, C. Kutter, J. Hadfield, P. Bertone, C. Caldas, Systematic comparison of microarray profiling, real-time PCR, and next-generation sequencing technologies for measuring differential microRNA expression., RNA. 16 (2010) 991–1006. doi:10.1261/rna.1947110.

[140] T. Blondal, S. Jensby Nielsen, A. Baker, D. Andreasen, P. Mouritzen, M. Wrang Teilum, I.K. Dahlsveen, Assessing sample and miRNA profile quality in serum and plasma or other biofluids, Methods. 59 (2013) 164–169. doi:10.1016/j.ymeth.2012.09.015.

[141] C.L. Andersen, J.L. Jensen, T.F. Orntoft, Normalization of real-time quantitative reverse transcription-PCR data: a model-based variance estimation approach to identify genes

suited for normalization, applied to bladder and colon cancer data sets., Cancer Res. 64 (2004) 5245–5250. doi:10.1158/0008-5472.CAN-04-0496.

[142] P. Mestdagh, P. Van Vlierberghe, A. De Weer, D. Muth, F. Westermann, F. Speleman, J. Vandesompele, A novel and universal method for microRNA RT-qPCR data normalization, Genome Biol. 10 (2009) R64. doi:10.1186/gb-2009-10-6-r64.

[143] J. Jarry, D. Schadendorf, C. Greenwood, A. Spatz, L.C. van Kempen, The validity of circulating microRNAs in oncology: five years of challenges and contradictions, Mol. Oncol. 8 (2014) 819–829. doi:10.1016/j.molonc.2014.02.009.

[144] C.C. Pritchard, H.H. Cheng, M. Tewari, MicroRNA profiling: approaches and considerations, Nat. Rev. Genet. 13 (2012) 358–369. doi:10.1038/nrg3198.

[145] K.W. Witwer, Circulating MicroRNA Biomarker Studies: Pitfalls and Potential Solutions, Clin. Chem. 61 (2015) 56–63. doi:10.1373/clinchem.2014.221341.

[146] M.W.J. De Ronde, J.M. Ruijter, P.D. Moerland, E.E. Creemers, S.J. Pinto-Sietsma, Study design and qPCR data analysis guidelines for reliable circulating miRNA biomarker experiments: A review, Clin. Chem. 64 (2018) 1308–1318. doi:10.1373/clinchem.2017.285288.

[147] C.C. Pritchard, E. Kroh, B. Wood, J.D. Arroyo, K.J. Dougherty, M.M. Miyaji, J.F. Tait, M. Tewari, Blood Cell Origin of Circulating MicroRNAs: A Cautionary Note for Cancer Biomarker Studies, Cancer Prev. Res. 5 (2012) 492 LP – 497. doi:10.1158/1940-6207.CAPR-11-0370.

[148] N. Sunderland, P. Skroblin, T. Barwari, R.P. Huntley, R. Lu, A. Joshi, R.C. Lovering, M. Mayr, MicroRNA Biomarkers and Platelet Reactivity: The Clot Thickens, Circ. Res. 120 (2017) 418–435. doi:10.1161/CIRCRESAHA.116.309303.

[149] D.C. Muth, B.H. Powell, Z. Zhao, K.W. Witwer, miRNAs in platelet-poor blood plasma and purified RNA are highly stable: a confirmatory study, BMC Res. Notes. 11 (2018) 273. doi:10.1186/s13104-018-3378-6.

[150] A.J. Mitchell, W.D. Gray, S.S. Hayek, Y.-A. Ko, S. Thomas, K. Rooney, M. Awad, J.D. Roback, A. Quyyumi, C.D. Searles, Platelets confound the measurement of extracellular miRNA in archived plasma, Sci. Rep. 6 (2016) 32651. doi:10.1038/srep32651.

[151] H. Schwarzenbach, A.M. da Silva, G. Calin, K. Pantel, Data Normalization Strategies for MicroRNA Quantification, Clin. Chem. 61 (2015) 1333–1342. doi:10.1373/clinchem.2015.239459.

[152] M. Faraldi, M. Gomarasca, V. Sansoni, S. Perego, G. Banfi, G. Lombardi, Normalization strategies differently affect circulating miRNA profile associated with the training status, Sci. Rep. 9 (2019) 1584. doi:10.1038/s41598-019-38505-x.

[153] J. Shen, Q. Wang, I. Gurvich, H. Remotti, R.M. Santella, Evaluating normalization approaches for the better identification of aberrant microRNAs associated with hepatocellular carcinoma, Hepatoma Res. 2 (2016) 305–315. doi:10.20517/2394-5079.2016.28.

[154] J. Qin, H. Liang, D. Shi, J. Dai, Z. Xu, D. Chen, X. Chen, Q. Jiang, A panel of microRNAs as a new biomarkers for the detection of deep vein thrombosis, J. Thromb. Thrombolysis. 39 (2015) 215–221. doi:10.1007/s11239-014-1131-0.

[155] I. Starikova, S. Jamaly, A. Sorrentino, T. Blondal, N. Latysheva, M. Sovershaev, J.B. Hansen, Differential expression of plasma miRNAs in patients with unprovoked venous thromboembolism and healthy control individuals, Thromb. Res. 136 (2015) 566–572. doi:10.1016/j.thromres.2015.07.005.

[156] X. Wang, K. Sundquist, J.L. Elf, K. Strandberg, P.J. Svensson, A. Hedelius, K. Palmér, A. a. Memon, J. Sundquist, B. Zöller, Diagnostic potential of plasma microRNA signatures in patients with deep-vein thrombosis, Thromb. Haemost. 116 (2016) 328–336. doi:10.1160/TH16-01-0071.

[157] M.P. Allen, Understanding Regression Analysis, Springer US, 2007.

[158] M. Lynch, B. Walsh, Genetics and Analysis of Quantitative Traits, Sinauer, 1998.

[159] M.H. Zweig, G. Campbell, Receiver-operating characteristic (ROC) plots: a fundamental evaluation tool in clinical medicine, Clin. Chem. 39 (1993) 561–577. doi:10.1093/clinchem/39.4.561.

[160] K. Hajian-Tilaki, Receiver Operating Characteristic (ROC) Curve Analysis for Medical Diagnostic Test Evaluation, Casp. J. Intern. Med. 4 (2013) 627–635.

[161] L. Almasy, J. Blangero, Multipoint Quantitative-Trait Linkage Analysis in General Pedigrees, Am. J. Hum. Genet. 62 (1998) 1198–1211. doi:10.1086/301844.

[162] A. Ziyatdinov, H. Brunel, A. Martinez-Perez, A. Buil, A. Perera, J.M. Soria, Genetics and population analysis solarius : an R interface to SOLAR for variance component analysis in pedigrees, Bioinformatics. 32 (2016) 1901–1902. doi:10.1093/bioinformatics/btw080.

[163] E.L. de Maturana, L. Alonso, P. Alarcón, I.A. Martín-Antoniano, S. Pineda, L. Piorno, M.L. Calle, N. Malats, Challenges in the integration of omics and non-omics data, Genes (Basel). 10 (2019). doi:10.3390/genes10030238.

[164] I.S.L. Zeng, T. Lumley, Review of statistical learning methods in integrated omics studies (An integrated information science), Bioinform. Biol. Insights. 12 (2018). doi:10.1177/1177932218759292.

[165] Y. Hasin, M. Seldin, A. Lusis, Multi-omics approaches to disease, Genome Biol. 18 (2017) 1–15. doi:10.1186/s13059-017-1215-1.

[166] C. Wu, F. Zhou, J. Ren, X. Li, Y. Jiang, S. Ma, A selective review of multi-level omics data integration using variable selection, High-Throughput. 8 (2019) 1–25. doi:10.3390/ht8010004.

[167] J.O. Ogutu, T. Schulz-Streeck, H.-P. Piepho, Genomic selection using regularized linear regression models: ridge regression, lasso, elastic net and their extensions, BMC Proc. 6 (2012) S10. doi:10.1186/1753-6561-6-S2-S10.

[168] R. Tibshirani, Regression shrinkage and selection via the lasso: a retrospective, J. R. Stat. Soc. Ser. B (Statistical Methodol. 73 (2011) 273–282. doi:10.1111/j.1467-9868.2011.00771.x.

[169] R. Tibshirani, Regression Shrinkage and Selection via the Lasso, J. R. Stat. Soc. Ser. B. 58 (1996) 267–288. http://www.jstor.org/stable/2346178.

[170] B. Bischl, M. Lang, L. Kotthoff, J. Schiffner, J. Richter, E. Studerus, G. Casalicchio, Z.M. Jones, Mlr: Machine learning in R, J. Mach. Learn. Res. 17 (2016) 1–5.

[171] Q. Meng, W. Wang, X. Yu, W. Li, L. Kong, A. Qian, C. Li, X. Li, Upregulation of MicroRNA-126 Contributes to Endothelial Progenitor Cell Function in Deep Vein Thrombosis via Its Target PIK3R2, J. Cell. Biochem. 116 (2015) 1613–1623. doi:10.1002/jcb.25115.

[172] A. Fort, C. Borel, E. Migliavacca, S.E. Antonarakis, R.J. Fish, M. Neerman-Arbez, Regulation of fibrinogen production by microRNAs, Blood. 116 (2010) 2608–2615. doi:10.1182/blood-2010-02-268011.

[173] H.-Y. Huang, Y.-C.-D. Lin, J. Li, K.-Y. Huang, S. Shrestha, H.-C. Hong, Y. Tang, Y.-G. Chen, C.-N. Jin, Y. Yu, J.-T. Xu, Y.-M. Li, X.-X. Cai, Z.-Y. Zhou, X.-H. Chen, Y.-Y. Pei, L. Hu, J.-J. Su, S.-D. Cui, F. Wang, Y.-Y. Xie, S.-Y. Ding, M.-F. Luo, C.-H. Chou, N.-W. Chang, K.-W. Chen, Y.-H. Cheng, X.-H. Wan, W.-L. Hsu, T.-Y. Lee, F.-X. Wei, H.-D. Huang, miRTarBase 2020: updates to the experimentally validated microRNA–target interaction database, Nucleic Acids Res. 48 (2019) D148–D154. doi:10.1093/nar/gkz896.

[174] T. Zhang, J. Guo, J. Gu, Z. Wang, G. Wang, H. Li, J. Wang, Identifying the key genes and microRNAs in colorectal cancer liver metastasis by bioinformatics analysis and in vitro experiments, Oncol. Rep. 41 (2019) 279–291. doi:10.3892/or.2018.6840.

[175] X. Zhou, W. Wen, X. Shan, J. Qian, H. Li, T. Jiang, W. Wang, W. Cheng, F. Wang, L. Qi, Y. Ding, P. Liu, W. Zhu, Y. Chen, MiR-28-3p as a potential plasma marker in diagnosis of pulmonary embolism, Thromb. Res. 138 (2015) 91–95. doi:10.1016/j.thromres.2015.12.006.

[176] X. Wang, K. Sundquist, P.J. Svensson, H. Rastkhani, K. Palmér, A.A. Memon, J. Sundquist, B. Zöller, Association of recurrent venous thromboembolism and circulating microRNAs, Clin. Epigenetics. 11 (2019) 1–12. doi:10.1186/s13148-019-0627-z.

[177] Q.Q. Jin, J.H. Sun, Q.X. Du, X.J. Lu, X.Y. Zhu, H.L. Fan, C. Hölscher, Y.Y. Wang, Integrating microRNA and messenger RNA expression profiles in a rat model of deep vein thrombosis, Int. J. Mol. Med. 40 (2017) 1019–1028. doi:10.3892/ijmm.2017.3105.

[178] M.A. Spinner, L.A. Sanchez, A.P. Hsu, P.A. Shaw, C.S. Zerbe, K.R. Calvo, D.C. Arthur, W. Gu, C.M. Gould, C.C. Brewer, E.W. Cowen, A.F. Freeman, K.N. Olivier, G. Uzel, A.M. Zelazny, J.R. Daub, C.D. Spalding, R.J. Claypool, N.K. Giri, B.P. Alter, E.M. Mace, J.S. Orange, J. Cuellar-Rodriguez, D.D. Hickstein, S.M. Holland, GATA2 deficiency: a protean disorder of hematopoiesis, lymphatics, and immunity, Blood. 123 (2014) 809–821. doi:10.1182/blood-2013-07-515528.

[179] N. Jahroudi, D.C. Lynch, Endothelial-cell-specific regulation of von Willebrand factor gene expression., Mol. Cell. Biol. 14 (1994) 999–1008. doi:10.1128/mcb.14.2.999.

[180] K.D. Johnson, A.P. Hsu, M.-J. Ryu, J. Wang, X. Gao, M.E. Boyer, Y. Liu, Y. Lee, K.R. Calvo, S. Keles, J. Zhang, S.M. Holland, E.H. Bresnick, Cis-element mutated in GATA2-dependent immunodeficiency governs hematopoiesis and vascular integrity, J. Clin. Invest. 122 (2012) 3692–3704. doi:10.1172/JCI61623.

[181] I.M. Rietveld, W.M. Lijfering, S. le Cessie, M.H.A. Bos, F.R. Rosendaal, P.H. Reitsma, S.C. Cannegieter, High levels of coagulation factors and venous thrombosis risk: strongest association for factor VIII and von Willebrand factor, J. Thromb. Haemost. 17 (2019) 99–109. doi:10.1111/jth.14343.

[182] E.C. Flood, K.A. Hajjar, The annexin A2 system and vascular homeostasis, Vascul. Pharmacol. 54 (2011) 59–67. doi:10.1016/j.vph.2011.03.003.

[183] O. Roda, M.L. Valero, S. Peiró, D. Andreu, F.X. Real, P. Navarro, New insights into the tPA-annexin A2 interaction: Is annexin A2 Cys 8 the sole requirement for this association?, J. Biol. Chem. 278 (2003) 5702–5709. doi:10.1074/jbc.M207605200.

[184] H. Fassel, H. Chen, M. Ruisi, M.T. De Sancho, K.A. Hajjar, Markedly Reduced Expression of Annexin A2 Is Associated with Hypofibrinolysis and Provoked and Unprovoked Venous Thromboembolism, Blood. 134 (2019) 443. doi:10.1182/blood-2019-127426.

[185] L. van der Vorm, J. Remijn, B. de Laat, D. Huskens, Effects of Plasmin on von Willebrand Factor and Platelets: A Narrative Review, TH Open. 02 (2018) e218–e228. doi:10.1055/s-0038-1660505.

[186] C.I. Jones, K.L. Tucker, P. Sasikumar, T. Sage, W.J. Kaiser, C. Moore, M. Emerson, J.M. Gibbins, Integrin-linked kinase regulates the rate of platelet activation and is essential for the formation of stable thrombi, J. Thromb. Haemost. 12 (2014) 1342–1352. doi:10.1111/jth.12620.

[187] X. Chen, Y. Zhang, Y. Wang, D. Li, L. Zhang, K. Wang, X. Luo, Z. Yang, Y. Wu, J. Liu, PDK1 regulates platelet activation and arterial thrombosis., Blood. 121 (2013) 3718–3726. doi:10.1182/blood-2012-10-461897.

[188] Z.M. Huttinger, M.W. Milks, M.S. Nickoli, W.L. Aurand, L.C. Long, D.G. Wheeler, K.M. Dwyer, A.J.F. D'Apice, S.C. Robson, P.J. Cowan, R.J. Gumina, Ectonucleotide triphosphate diphosphohydrolase-1 (CD39) mediates resistance to occlusive arterial thrombus formation after vascular injury in mice, Am. J. Pathol. 181 (2012) 322–333. doi:10.1016/j.ajpath.2012.03.024.

[189] F. Ya, X.R. Xu, Y. Shi, R.C. Gallant, F. Song, X. Zuo, Y. Zhao, Z. Tian, C. Zhang, X. Xu, W. Ling, H. Ni, Y. Yang, Coenzyme Q10 Upregulates Platelet cAMP/PKA Pathway and Attenuates Integrin αIIbβ3 Signaling and Thrombus Growth, Mol. Nutr. Food Res. 1900662 (2019) 1900662. doi:10.1002/mnfr.201900662.

[190] B. Xiang, G. Zhang, S. Ye, C. Huang, J. Liu, C. Ruan, S.S. Smyth, S.W. Whiteheart, Z. Li, VPS33B, a Member of the Sec1/Munc18 Protein Family, Binds to Integrin Beta Subunits and Is Required for αIIbβ3 Outside-in Signaling, Blood. 124 (2014) 340. doi:10.1182/blood.V124.21.340.340.

[191] I. Mertens, L.F. Van Gaal, Obesity, haemostasis and the fibrinolytic system., Obes. Rev. 3 (2002) 85–101. doi:10.1046/j.1467-789x.2002.00056.x.

[192] G. Anfossi, I. Russo, M. Trovati, Platelet dysfunction in central obesity., Nutr. Metab. Cardiovasc. Dis. 19 (2009) 440–449. doi:10.1016/j.numecd.2009.01.006.

[193] K. Suzuki-Inoue, D. Tulasne, Y. Shen, T. Bori-Sanz, O. Inoue, S.M. Jung, M. Moroi, R.K. Andrews, M.C. Berndt, S.P. Watson, Association of Fyn and Lyn with the proline-rich domain of glycoprotein VI regulates intracellular signaling., J. Biol. Chem. 277 (2002) 21561–21566. doi:10.1074/jbc.M201012200.

[194] H. Yin, J. Liu, Z. Li, M.C. Berndt, C.A. Lowell, X. Du, Src family tyrosine kinase Lyn mediates VWF/GPIb-IX-induced platelet activation via the cGMP signaling pathway, Blood. 112 (2008) 1139—1146. doi:10.1182/blood-2008-02-140970.

[195] C. Wong, Y. Liu, J. Yip, R. Chand, J.L. Wee, L. Oates, B. Nieswandt, A. Reheman, H. Ni, N. Beauchemin, D.E. Jackson, CEACAM1 negatively regulates platelet-collagen interactions and thrombus growth in vitro and in vivo., Blood. 113 (2009) 1818–1828. doi:10.1182/blood-2008-06-165043.

[196] J.P. Wood, P.E.R. Ellery, S.A. Maroney, A.E. Mast, Protein S is a cofactor for platelet and endothelial tissue factor pathway inhibitor-alpha but not for cell surface-associated tissue factor pathway inhibitor., Arterioscler. Thromb. Vasc. Biol. 34 (2014) 169–176. doi:10.1161/ATVBAHA.113.302655.

[197] M. Vázquez-Santiago, N. Vilalta, B. Cuevas, J. Murillo, D. Llobet, R. Macho, N. Pujol-Moix, M. Carrasco, J. Mateo, J. Fontcuberta, J.M. Soria, J.C. Souto, Short closure time values in PFA–100® are related to venous thrombotic risk. Results from the RETROVE Study, Thromb. Res. 169 (2018) 57–63. doi:https://doi.org/10.1016/j.thromres.2018.07.012.

[198] L. Marcus, S.R. M., K.P. J., W.A. M., S.J. R., J.-P. Martine, N.K. B., Platelets Drive Thrombus Propagation in a Hematocrit and Glycoprotein VI–Dependent Manner in an In Vitro Venous Thrombosis Model, Arterioscler. Thromb. Vasc. Biol. 38 (2018) 1052–

1062. doi:10.1161/ATVBAHA.118.310731.

[199] D. Poel, T.E. Buffart, J. Oosterling-Jansen, H.M. Verheul, J. Voortman, Evaluation of several methodological challenges in circulating miRNA qPCR studies in patients with head and neck cancer, Exp. Mol. Med. 50 (2018) e454–e454. doi:10.1038/emm.2017.288.

[200] C. Glinge, S. Clauss, K. Boddum, R. Jabbari, J. Jabbari, B. Risgaard, P. Tomsits, B. Hildebrand, S. Kaèaèb, R. Wakili, T. Jespersen, J. Tfelt-Hansen, Stability of circulating blood-based microRNAs-Pre-Analytic methodological considerations, PLoS One. 12 (2017) 1–16. doi:10.1371/journal.pone.0167969.

[201] J.S. Shah, P.S. Soon, D.J. Marsh, Comparison of Methodologies to Detect Low Levels of Hemolysis in Serum for Accurate Assessment of Serum microRNAs, PLoS One. 11 (2016) e0153200. https://doi.org/10.1371/journal.pone.0153200.

[202] M.B. Kirschner, J.J.B. Edelman, S.C.-H. Kao, M.P. Vallely, N. van Zandwijk, G. Reid, The Impact of Hemolysis on Cell-Free microRNA Biomarkers, Front. Genet. 4 (2013) 94. doi:10.3389/fgene.2013.00094.

[203] S. Pizzamiglio, S. Zanutto, C.M. Ciniselli, A. Belfiore, S. Bottelli, M. Gariboldi, P. Verderio, A methodological procedure for evaluating the impact of hemolysis on circulating microRNAs, Oncol. Lett. 13 (2017) 315–320. doi:10.3892/ol.2016.5452.

[204] K.D. Rasmussen, S. Simmini, C. Abreu-Goodger, N. Bartonicek, M. Di Giacomo, D. Bilbao-Cortes, R. Horos, M. Von Lindern, A.J. Enright, D. O'Carroll, The miR-144/451 locus is required for erythroid homeostasis, J. Exp. Med. 207 (2010) 1351–1358. doi:10.1084/jem.20100458.

[205] D.A. Kretov, I.A. Walawalkar, A. Mora-Martin, A.M. Shafik, S. Moxon, D. Cifuentes, Ago2-Dependent Processing Allows miR-451 to Evade the Global MicroRNA Turnover Elicited during Erythropoiesis, Mol. Cell. 78 (2020) 317-328.e6. doi:10.1016/j.molcel.2020.02.020.

[206] A. Yamada, M.A. Cox, K.A. Gaffney, A. Moreland, C.R. Boland, A. Goel, Technical factors involved in the measurement of circulating microRNA biomarkers for the detection of colorectal neoplasia, PLoS One. 9 (2014) e112481–e112481. doi:10.1371/journal.pone.0112481.

[207] E. Landoni, R. Miceli, M. Callari, P. Tiberio, V. Appierto, V. Angeloni, L. Mariani, M.G. Daidone, Proposal of supervised data analysis strategy of plasma miRNAs from hybridisation array data with an application to assess hemolysis-related deregulation, BMC Bioinformatics. 16 (2015) 388. doi:10.1186/s12859-015-0820-9.

[208] M.B. Kirschner, S.C. Kao, J.J. Edelman, N.J. Armstrong, M.P. Vallely, N. van Zandwijk, G. Reid, Haemolysis during sample preparation alters microRNA content of plasma, PLoS One. 6 (2011) e24145–e24145. doi:10.1371/journal.pone.0024145.

[209] A. Meerson, T. Ploug, Assessment of six commercial plasma small RNA isolation kits using qRT-PCR and electrophoretic separation: Higher recovery of microRNA following ultracentrifugation, Biol. Methods Protoc. 1 (2016) 1–9. doi:10.1093/biomethods/bpw003.

[210] T. Huan, J. Rong, C. Liu, X. Zhang, K. Tanriverdi, R. Joehanes, B.H. Chen, J.M. Murabito, C. Yao, P. Courchesne, P.J. Munson, C.J. O'Donnell, N. Cox, A.D. Johnson, M.G. Larson, D. Levy, J.E. Freedman, Genome-wide identification of microRNA expression quantitative trait loci, Nat. Commun. 6 (2015) 1–9. doi:10.1038/ncomms7601.

[211] L. Pase, J.E. Layton, W.P. Kloosterman, D. Carradice, P.M. Waterhouse, G.J. Lieschke, miR-451 regulates zebrafish erythroid maturation in vivo via its target gata2., Blood. 113 (2009) 1794–1804. doi:10.1182/blood-2008-05-155812.

[212] C. Vasilescu, S. Rossi, M. Shimizu, S. Tudor, A. Veronese, M. Ferracin, M.S. Nicoloso, E. Barbarotto, M. Popa, O. Stanciulea, M.H. Fernandez, D. Tulbure, C.E. Bueso-Ramos, M. Negrini, G.A. Calin, MicroRNA fingerprints identify miR-150 as a plasma prognostic marker in patients with sepsis., PLoS One. 4 (2009) e7405. doi:10.1371/journal.pone.0007405.

[213] Y. Ma, Y. Liu, H. Hou, Y. Yao, H. Meng, MiR-150 predicts survival in patients with sepsis and inhibits LPS-induced inflammatory factors and apoptosis by targeting NF-κB1 in human umbilical vein endothelial cells, Biochem. Biophys. Res. Commun. 500 (2018) 828–837. doi:10.1016/j.bbrc.2018.04.168.

[214] E.R. Gamazon, D. Ziliak, H.K. Im, B. Lacroix, D.S. Park, N.J. Cox, R.S. Huang, Genetic architecture of microRNA expression: Implications for the transcriptome and complex traits, Am. J. Hum. Genet. 90 (2012) 1046–1063. doi:10.1016/j.ajhg.2012.04.023.

[215] M. Nikpay, K. Beehler, A. Valsesia, J. Hager, M.-E. Harper, R. Dent, R. McPherson, Genome-wide identification of circulating-miRNA expression quantitative trait loci reveals the role of several miRNAs in the regulation of cardiometabolic phenotypes, Cardiovasc. Res. 115 (2019) 1629–1645. doi:10.1093/cvr/cvz030.

[216] Y. Liang, D. Ridzon, L. Wong, C. Chen, Characterization of microRNA expression profiles in normal human tissues, BMC Genomics. 8 (2007) 1–20. doi:10.1186/1471-2164-8-166.

[217] Z.B. Nagy, B.K. Barták, A. Kalmár, O. Galamb, B. Wichmann, M. Dank, P. Igaz, Z. Tulassay, B. Molnár, Comparison of Circulating miRNAs Expression Alterations in Matched Tissue and Plasma Samples During Colorectal Cancer Progression, Pathol. Oncol. Res. 25 (2019) 97–105. doi:10.1007/s12253-017-0308-1.

[218] A.H. Zedan, T.F. Hansen, J. Assenholt, M. Pleckaitis, J.S. Madsen, P.J.S. Osther, MicroRNA expression in tumour tissue and plasma in patients with newly diagnosed metastatic prostate cancer, Tumor Biol. 40 (2018) 1–11. doi:10.1177/1010428318775864.

[219] L. Xu, M. Li, M. Wang, D. Yan, G. Feng, G. An, The expression of microRNA-375 in plasma and tissue is matched in human colorectal cancer, BMC Cancer. 14 (2014) 714. doi:10.1186/1471-2407-14-714.

[220] A. Turchinovich, L. Weiz, B. Burwinkel, Extracellular miRNAs: The mystery of their origin and function, Trends Biochem. Sci. 37 (2012) 460–465. doi:10.1016/j.tibs.2012.08.003.

Annexes

Annex 1: Supplemental Material of Article 1.

Annex 2: Supplemental Material of Article 2.

Annex 1: Supplemental Material of Article 1.

SUPPLEMENTAL MATERIAL

Identification of a plasma microRNA profile associated with venous thrombosis

Alba Rodriguez-Rius[1], Sonia Lopez[1], Angel Martinez-Perez[1], Juan Carlos Souto[2] and Jose Manuel Soria[1]

[1] Genomics of Complex Diseases Group, Research Institute Hospital de la Santa Creu i Sant Pau. IIB Sant Pau. Barcelona, Spain.

[2] Unit of Thrombosis and Hemostasis, Hospital de la Santa Creu i Sant Pau. Barcelona, Spain.

Supplemental Methods

Recruitment criteria and description of the population

The GAIT-2 population was recruited via a proband who suffered an event of idiopathic thrombosis. Each family is composed of at least 10 individuals in three generations.

The proband must meet at least one of these criteria:

- Onset of thrombosis before 45 years of age.
- Recurrent thrombotic events, at least one of which was spontaneous.
- Single spontaneous thrombosis event with a first-degree relative also affected.

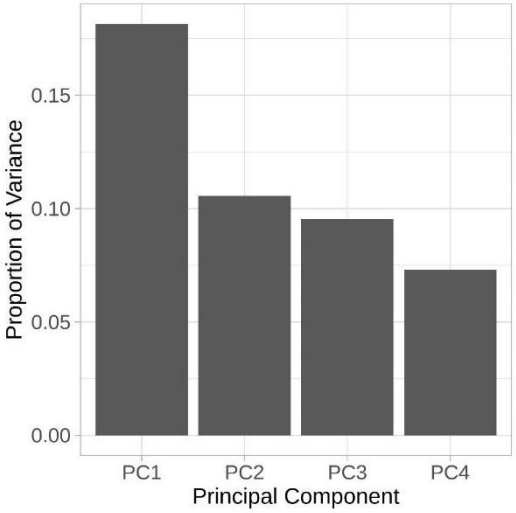
The thrombosis events were considered idiopathic because all know biological causes were excluded: antithrombin deficiency, Protein C deficiency, Protein S deficiency, activated Protein C resistance, plasminogen deficiency, heparin cofactor II deficiency, Factor V Leiden mutation, dysfibrinogenemia, lupus anticoagulant and antiphospholipid antibodies. In addition, the following acquired risk factors in the three months prior to the thrombosis event were excluded: surgery, immobilization, bone fracture, hospitalization, and pregnancy. Finally, the thrombotic events secondaries to the following pathologies were excluded: inflammatory bowel disease, Behçhet disease, and active neoplastic condition.

The collection of the biological samples was performed at least three months after the last thrombotic event, as well as at least one month after any acute inflammatory episode. Finally, prior to the collection, the ongoing treatments were removed, including:

- Heparin treatments, at least 24 hours before.
- Oral anticoagulant treatments, at least 15 days before.
- Anti-platelet drugs, at least 15 days before.
- Anti-inflammatory drugs, at least 15 days before (only if possible).

Supplemental Figures and Figure Legends

Supplemental Figure I: Principal Component Analysis of miRNA expression in the discovery phase.



Supplemental Tables

Supplemental Table I: Main characteristics of the population.

	Discovery phase		Internal validation phase
	Cases	Controls	Whole GAIT-2 population
n	52	52	935
Sex-Female (%)	34 (65.4%)	34 (65.4%)	470 (50.3%)
Age	48.6	50.3	39.5
BMI*	28.3	26.5	24.8
ACO†	0	2 (3.7%)	39 (4.2%)
Smoking‡	17 (32.7%)	17 (32.7%)	225 (24.1%)
VT	--	--	85

* BMI=body mass index
† ACO= oral contraceptive active therapy
‡ Smoking=categorical classification, active smokers or non-smokers.

Supplemental Table II: Results of the comparison of different inter-plate calibrators. While in the discovery phase the inter-plate calibrator used was UniSp3, in the validation phase the mean Ct value per plate was found to represent better the plate variability. Although we considered that this difference was due to the different design of plates in each phase, some analyses were performed to ensure that this change did not affect significantly the data. Here we report the differences among the Ct values corrected for plate, using UniSp3 and the mean Ct value per plate. Correlation coefficient (β) and its corresponding p-value (p) for the linear correlation between them for each miRNA, and the p-value of the t-student test (t-test p) to ensure that the difference was not significant.

miRNA	β	p	t-test p
hsa-miR-126-3p	0.90	$< 1 \times 10^{-300}$	0.86
has-miR-142-3p	0.94	$< 1 \times 10^{-300}$	0.98
hsa-miR146a-5p	0.93	$< 1 \times 10^{-300}$	0.92
hsa-miR-148a-3p	0.94	$< 1 \times 10^{-300}$	0.96
hsa-miR-192-5p	0.92	$< 1 \times 10^{-300}$	0.89
hsa-miR-194-5p	0.9	5.42×10^{-261}	0.91
hsa-miR-197-3p	0.92	$< 1 \times 10^{-300}$	0.84
hsa-miR-221-3p	0.94	$< 1 \times 10^{-300}$	0.95
hsa-miR-23b-3p	0.9	4.99×10^{-296}	0.82
hsa-miR-27a-3p	0.93	$< 1 \times 10^{-300}$	0.93
hsa-miR-28-3p	0.91	7.06×10^{-268}	0.84
hsa-miR-320a	0.9	8.17×10^{-285}	0.95
hsa-miR-320b	0.88	2.36×10^{-266}	0.89
hsa-miR-342-3p	0.88	2.34×10^{-269}	0.88
hsa-miR-548c-5p	0.92	3.42×10^{-299}	0.97
hsa-miR-885-5p	0.92	1.91×10^{-231}	0.73

Supplemental Table III: Number of undetectable assays for each miRNA in the discovery phase.

miRNA	n*	Continue		Continue	
hsa-let-7a-5p	0	miRNA	n	miRNA	n
hsa-let-7b-5p	0	hsa-miR-15b-5p	3	hsa-miR-30d-5p	0
hsa-let-7c-5p	1	hsa-miR-16-5p	0	hsa-miR-30e-5p	0
hsa-let-7d-3p	0	hsa-miR-181a-5p	2	hsa-miR-320a	0
hsa-let-7d-5p	2	hsa-miR-185-5p	0	hsa-miR-320b	0
hsa-let-7f-5p	1	hsa-miR-18a-5p	3	hsa-miR-320c	0
hsa-let-7g-5p	0	hsa-miR-18b-5p	4	hsa-miR-328-3p	6
hsa-let-7i-5p	0	hsa-miR-191-5p	0	hsa-miR-342-3p	0
hsa-miR-101-3p	0	hsa-miR-192-5p	1	hsa-miR-345-5p	7
hsa-miR-103a-3p	0	hsa-miR-194-5p	2	hsa-miR-34a-3p	1
hsa-miR-106a-5p	0	hsa-miR-197-3p	1	hsa-miR-361-5p	0
hsa-miR-106b-5p	1	hsa-miR-199a-3p	0	hsa-miR-363-3p	3
hsa-miR-107	1	hsa-miR-19a-3p	0	hsa-miR-374b-5p	4
hsa-miR-10b-5p	4	hsa-miR-19b-3p	0	hsa-miR-375	5
hsa-miR-122-5p	0	hsa-miR-20a-5p	0	hsa-miR-376a-3p	10
hsa-miR-125b-5p	1	hsa-miR-215-5p	5	hsa-miR-376c-3p	2
hsa-miR-126-3p	0	hsa-miR-21-5p	0	hsa-miR-382-5p	10
hsa-miR-126-5p	0	hsa-miR-221-3p	0	hsa-miR-423-3p	2
hsa-miR-130a-3p	0	hsa-miR-222-3p	0	hsa-miR-423-5p	0
hsa-miR-139-5p	1	hsa-miR-223-3p	0	hsa-miR-424-5p	9
hsa-miR-140-3p	0	hsa-miR-22-3p	0	hsa-miR-425-5p	0
hsa-miR-142-3p	0	hsa-miR-23a-3p	0	hsa-miR-451a	0
hsa-miR-142-5p	0	hsa-miR-23b-3p	0	hsa-miR-484	1
hsa-miR-143-3p	10	hsa-miR-24-3p	0	hsa-miR-486-5p	0
hsa-miR-144-3p	0	hsa-miR-25-3p	0	hsa-miR-548c-5p	8
hsa-miR-144-5p	1	hsa-miR-26a-5p	1	hsa-miR-574-3p	4
hsa-miR-145-5p	1	hsa-miR-26b-5p	3	hsa-miR-584-5p	3
hsa-miR-146a-5p	0	hsa-miR-27a-3p	0	hsa-miR-590-5p	7
hsa-miR-148a-3p	3	hsa-miR-27b-3p	0	hsa-miR-652-3p	4
hsa-miR-148b-3p	4	hsa-miR-28-3p	6	hsa-miR-660-5p	7
hsa-miR-150-5p	0	hsa-miR-29a-3p	0	hsa-miR-885-5p	7
hsa-miR-151a-3p	0	hsa-miR-29c-3p	1	hsa-miR-92a-3p	0
hsa-miR-151a-5p	0	hsa-miR-30a-5p	3	hsa-miR-93-5p	0
hsa-miR-152-3p	2	hsa-miR-30b-5p	0	hsa-miR-99a-5p	2
hsa-miR-15a-5p	0	hsa-miR-30c-5p	0	hsa-miR-378a-3p	4

* n = number of samples in which the miRNA was undetectable in the discovery phase.

Supplemental Table IV: Associations found in the discovery phase for the 16 miRNAs selected. Extended version of Table 1, reporting the correlation coefficients and p-values for each miRNA and phenotype association.

miRNA	VT		Associated intermediate phenotypes*		
	β	p	Phenotype	β	p
hsa-miR-192-5p	0.58	$1.9x10^{-03}$	Thrombin time	0.31	$1.43x10^{-03}$
hsa-miR-885-5p	0.58	$2.7x10^{-03}$	von Willebrand Factor	0.27	$6.69x10^{-03}$
hsa-miR-23b-3p	-0.49	8.4×10^{-02}	Factor VIII	-0.3	$2.18x10^{-03}$
			Total Protein S	-0.28	$3.89x10^{-03}$
hsa-miR-27a-3p	-0.44	$1.9x10^{-02}$	Thrombin time	-0.26	$6.89x10^{-03}$
hsa-miR-194-5p	0.40	$3.8x10^{-02}$	Thrombin time	0.24	$1.56x10^{-02}$
hsa-miR-221-3p	-0.39	$3.9x10^{-02}$	Factor VIII	0.24	$1.53x10^{-02}$
			von Willebrand Factor	0.22	$1.46x10^{-02}$
hsa-miR-197-3p	-0.39	$4.0x10^{-02}$	Factor VIII	-0.2	$4.58x10^{-02}$
hsa-miR-548c-5p	-0.39	$4.6x10^{-02}$	Factor VIII	-0.28	$7.22x10^{-03}$
			von Willebrand factor	-0.29	$4.19x10^{-03}$
hsa-miR-320a	-0.37	$4.9x10^{-02}$	Clot formation rate	0.37	$1.43x10^{-04}$
hsa-miR-142-3p	-0.13	0.48, n.s	Factor XII	-0.26	$8.82x10^{-03}$
hsa-miR-146a-5p	-0.31	0.09, n.s.	Prothrombin time	0.29	$2.82x10^{-03}$
			Fibrinogen	0.27	$6.74x10^{-03}$
hsa-miR-148a-3p	0.04	0.82, n.s.	Average RNA by platelet	0.25	$2.33x10^{-02}$
hsa-miR-28-3p	0.07	0.69, n.s	Thrombin generation – Peak of thrombin	-0.35	$4.50x10^{-04}$
hsa-miR-320b	-0.06	0.71, n.s.	Functional Protein S	-0.25	$1.19x10^{-03}$
hsa-miR-342-3p	-0.16	0.37, n.s.	Factor VIII	-0.41	$1.64x10^{-05}$
			von Willebrand Factor	-0.38	$7.21x10^{-05}$
			Fibrinogen	-0.31	$1.75x10^{-03}$
hsa-miR-126-3p	0.34	0.06, n.s.	Factor VII	0.26	$7.51x10^{-03}$

VT = venous thrombosis; β = correlation coefficient; p = p-value, n.s.=non-significant
* Main intermediate phenotypes related to VT that resulted in linear correlation at p<0.05 with the miRNA.

Supplemental Table V: Heritabilities of miRNAs and significant covariates.

miRNA	Heritability	SE*	p	Covariates
hsa-miR-146a-5p	0.39	0.07	$5.4x10^{-11}$	
hsa-miR-126-3p	0.35	0.07	$2.6x10^{-09}$	Age
hsa-miR-342-3p	0.27	0.06	$3.7x10^{-08}$	
hsa-miR-548c-5p	0.24	0.07	$3.3x10^{-06}$	Age, Age^2
hsa-miR-192-5p	0.24	0.07	$1.5x10^{-05}$	Age, Age^2, Sex
hsa-miR-320a	0.24	0.07	$1.5x10^{-05}$	
hsa-miR-23b-3p	0.22	0.07	$2.3x10^{-05}$	Age, Age^2
hsa-miR-320b	0.24	0.07	$9.1x10^{-05}$	Sex
hsa-miR-221-3p	0.23	0.07	$1.4x10^{-04}$	Sex
hsa-miR-142-3p	0.18	0.08	$1.7x10^{-03}$	
hsa-miR-194-5p	0.19	0.07	$1.8x10^{-03}$	Age, Age^2, Sex
hsa-miR-27a-3p	0.14	0.06	$6.1x10^{-03}$	Age, Age^2, Sex
hsa-miR-885-5p	0.16	0.08	$2x10^{-02}$	Age, Age^2, Sex
hsa-miR-28-3p	0.10	0.06	$2x10^{-02}$	

* SE=standard error

Major Resources Table

Laboratory Reagents for microRNA quantification

Product Name	Vendor or Source	Catalog #
miRCURY™ RNA Isolation Kit - Biofluids	Exiqon	300112
Universal cDNA synthesis kit II, 8-64 rxns	Exiqon	203301
miRCURY LNA™ Universal RT microRNA PCR – RNA Spike-in kit	Exiqon	203203
miRCURY LNA™ miRNA miRNOME PCR Panel-Human I+II	Exiqon	-- Plate R
Pick&Mix custom panels	Exiqon	
miRCURY LNA™ miRNA PCR Panel- Human QC	Exiqon	-- Plate AF
ExiLENT SYBR® Green master mix, 20 ml	Exiqon	203421
MS2 RNA	Roche	10165948001
ROX dye	Termo Fisher	75768500UL

Annex 2: Supplemental Material of Article 2.

SUPPLEMENTAL MATERIAL

Whole Blood Gene Expression in Venous Thrombosis: An Integrative Analysis with Clinical Phenotypes and MicroRNAs

Manuscript in preparation

Supplemental Table 1: Characteristics of the 102 subjects of the GAIT-2 Project included in the study.

	Global	VT patients	Controls
N	102	51	51
Age	49.4	48.9	49.9
Sex-female	66	33	33
BMI	27.5	28.4	26.7
Smoking	33	16	17
ACO	2	0	2
Hypertension	24	13	11

N= number of individuals, BMI= Body Mass Index,
ACO= active contraceptive oral therapy

Supplemental Table 2: Correlations of the 14 clinical phenotypes with the 260 genes annotated in the 'blood coagulation' pathway. Reported the 49 correlations at FDR<0.1.

Gene_ID	Phenotype	β	p	FDR	Gene_Name
ENSG00000136631	PSfunc	0.43	1.02E-05	2.19E-02	VPS45
ENSG00000254087	PSfunc	-0.42	1.33E-05	2.30E-02	LYN
ENSG00000090621	PSfunc	0.39	5.95E-05	3.13E-02	PABPC4
ENSG00000100099	PSfunc	0.39	6.14E-05	3.16E-02	HPS4
ENSG00000115486	PSfunc	0.39	6.64E-05	3.22E-02	GGCX
ENSG00000163041	PSfunc	-0.38	9.19E-05	3.54E-02	H3F3A
ENSG00000138185	FXI	-0.38	9.65E-05	3.59E-02	ENTPD1
ENSG00000116266	FIX	-0.38	9.82E-05	3.61E-02	STXBP3
ENSG00000179588	FVII	0.37	1.45E-04	4.12E-02	ZFPM1
ENSG00000166333	FIX	0.37	1.47E-04	4.13E-02	ILK
ENSG00000079385	PSfunc	-0.37	1.87E-04	4.52E-02	CEACAM1
ENSG00000136869	FXI	-0.36	2.39E-04	4.97E-02	TLR4
ENSG00000135365	PSfunc	-0.36	2.41E-04	4.98E-02	PHF21A
ENSG00000108946	FIX	-0.36	2.42E-04	4.99E-02	PRKAR1A
ENSG00000171988	FIX	-0.36	2.51E-04	5.07E-02	JMJD1C
ENSG00000177119	PSfunc	0.36	2.59E-04	5.13E-02	ANO6
ENSG00000168310	PSfunc	-0.36	2.62E-04	5.16E-02	IRF2
ENSG00000178726	FIX	-0.36	2.73E-04	5.25E-02	THBD
ENSG00000101608	FIX	0.36	2.86E-04	5.34E-02	MYL12A
ENSG00000174059	FIBc	-0.35	3.06E-04	5.49E-02	CD34
ENSG00000140992	FIX	-0.35	3.17E-04	5.58E-02	PDPK1
ENSG00000160999	PSfunc	-0.35	3.44E-04	5.77E-02	SH2B2
ENSG00000064961	PSfunc	0.35	3.46E-04	5.78E-02	HMG20B
ENSG00000159176	PSfunc	0.35	3.83E-04	6.04E-02	CSRP1
ENSG00000105851	FXII	-0.34	3.85E-04	6.05E-02	PIK3CG
ENSG00000147251	FXII	-0.34	4.10E-04	6.22E-02	DOCK11
ENSG00000167397	FXI	0.34	4.23E-04	6.30E-02	VKORC1
ENSG00000167397	Pc	0.34	5.46E-04	7.04E-02	VKORC1
ENSG00000100030	FXI	-0.34	5.78E-04	7.22E-02	MAPK1
ENSG00000178726	FVII	-0.33	6.26E-04	7.47E-02	THBD
ENSG00000182718	FIX	0.34	6.59E-04	7.64E-02	ANXA2
ENSG00000166189	HCY	-0.33	7.00E-04	7.85E-02	HPS6
ENSG00000179348	FvWAg	-0.33	7.62E-04	8.14E-02	GATA2
ENSG00000138185	FIX	-0.33	7.85E-04	8.25E-02	ENTPD1
ENSG00000116266	APTT	0.33	7.92E-04	8.28E-02	STXBP3
ENSG00000138185	PSfunc	-0.33	8.40E-04	8.50E-02	ENTPD1
ENSG00000121879	FIX	-0.33	8.57E-04	8.57E-02	PIK3CA
ENSG00000141448	APCR	0.33	9.00E-04	8.76E-02	GATA6
ENSG00000116266	FXI	-0.32	9.01E-04	8.76E-02	STXBP3
ENSG00000108561	PSfunc	0.33	9.36E-04	8.91E-02	C1QBP
ENSG00000107485	HCY	-0.32	1.01E-03	9.19E-02	GATA3
ENSG00000149091	Pc	0.32	1.05E-03	9.35E-02	DGKZ
ENSG00000134215	FXI	-0.32	1.08E-03	9.48E-02	VAV3
ENSG00000196415	HCY	0.32	1.11E-03	9.58E-02	PRTN3
ENSG00000179588	Pc	0.32	1.12E-03	9.64E-02	ZFPM1
ENSG00000141480	FIBc	0.32	1.13E-03	9.67E-02	ARRB2
ENSG00000158006	PSfunc	0.32	1.17E-03	9.82E-02	PAFAH2
ENSG00000099957	PSfunc	0.32	1.20E-03	9.90E-02	P2RX6
ENSG00000166333	PSfunc	0.32	1.22E-03	1.00E-01	ILK

Supplemental Figure 1: Principal component analysis of the features in Model 2