



(51) International Patent Classification:

C12Q 1/6883 (2018.01)

(21) International Application Number:

PCT/EP2019/087038

(22) International Filing Date:

26 December 2019 (26.12.2019)

(25) Filing Language:

English

(26) Publication Language:

English

(30) Priority Data:

18306857.6 27 December 2018 (27.12.2018) EP

19305977.1 24 July 2019 (24.07.2019) EP

(71) Applicants: INSERM (INSTITUT NATIONAL DE LA SANTÉ ET DE LA RECHERCHE MÉDICALE)

[FR/FR]; 101, rue de Tolbiac, 75013 Paris (FR).

UNIVERSITÉ DE BORDEAUX [FR/FR]; 35 Place Pey

Berland, 33000 BORDEAUX (FR). UNIVERSITÉ D'AIX

MARSEILLE [FR/FR]; 58 Boulevard Charles Livon,

13284 Marseille Cedex 07 (FR). ASSISTANCE PUB-

LIQUE HÔPITAUX DE MARSEILLE [FR/FR]; 80 Rue

Brochier, 13005 Marseille (FR). COMMISSARIAT À

L'ÉNERGIE ATOMIQUE ET AUX ÉNERGIES AL-

TERNATIVES (CEA) [FR/FR]; 25 Rue Leblanc, 75015

PARIS (FR). SORBONNE UNIVERSITÉ [FR/FR]; 21

rue de l'Ecole de Médecine, 75006 Paris (FR). INSTITUT

NATIONAL DE LA RECHERCHE AGRONOMIQUE

(INRA) [FR/FR]; 75007 Paris (FR).

(72) Inventors: TREGOUET, David-Alexandre; INSERM

UMR 1219, Bordeaux Population Health Research Center,

146 rue Léo Saignat, 33076 Bordeaux (FR). MORANGE,

Pierre-Emmanuel; INSERM 1263, Aix-Marseille Uni-

versité, Faculté de Médecine Timone, 27 Boulevard Jean

Moulin, 13385 cedex 05 Marseille (FR). DELEUZE, Jean-

François; Centre National de Recherche en Génomique

Humaine (CNRGH), 2 rue Gaston Crémieux, 91057 Evry

cedex (FR).

(74) Agent: INSERM TRANSFERT; 7 rue Watt, 75013 Paris

(FR).

(81) Designated States (unless otherwise indicated, for every

kind of national protection available): AE, AG, AL, AM,

AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DJ, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IR, IS, JO, JP, KE, KG, KH, KN, KP, KR, KW, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.

(84) Designated States (unless otherwise indicated, for every kind of regional protection available): ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, ST, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, RU, TJ, TM), European (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG).

Published:

— with international search report (Art. 21(3))

(54) Title: METHODS OF IDENTIFYING SUBJECTS HAVING OR AT RISK OF HAVING A COAGULATION RELATED DISORDER

(57) Abstract: Coagulation factor V (FV) plays an important and dual role in the regulation of blood coagulation by exhibiting both pro- and anticoagulant functions. In order to detect novel genetic loci participating to the regulation of Factor V (FV) plasma levels, the inventors conducted the first Genome Wide Association Study on this hemostatic phenotype in a sample of 510 individuals and replicated the main findings in an independent samples of 1156 individuals. In addition to genetic variations at the F5 locus, they identified novel associations at the *PLXDC2* locus with the *PLXDC2* rs927826 polymorphism explaining ~3.7% ($p = 7.5 \cdot 10^{-15}$ in the combined discovery and replication samples) of the variability of FV plasma levels. siRNA experiments in human hepatocellular carcinoma cell line confirmed the role of *PLXDC2* in modulating factor F5 gene expression. Accordingly, detecting a genetic variant in the *PLXDC2* gene would be suitable for identifying a subject having or at risk of having a coagulation related disorder.



METHODS OF IDENTIFYING SUBJECTS HAVING OR AT RISK OF HAVING A COAGULATION RELATED DISORDER

5 **FIELD OF THE INVENTION:**

The present invention relates to methods of identifying subjects having or at risk of having a coagulation related disorder.

BACKGROUND OF THE INVENTION:

10 Coagulation factor V (FV) plays an important and dual role in the regulation of blood coagulation by exhibiting both pro- and anticoagulant functions (reviewed in [1]). In plasma, single-chain FV expresses anticoagulant activity as a cofactor of both TFPI and activated protein C. In situations where the coagulation system is triggered, FV is converted to a highly effective procoagulant cofactor to activated factor X (FXa), which activates prothrombin to
15 thrombin, which in turn catalyzes fibrin deposition and activates platelets. While FV deficiency is known to associate with bleeding tendency, it has recently been proposed that FV plasma levels were associated with the risk of venous thrombosis (VT) in an ambiguous pattern. Indeed, in 2,377 VT patients and 2,943 controls of the MEGA study, individuals with either high (>1.22 IU/mL) or low (<0.57 IU/mL) FV levels were at higher risk of disease [2]. Until now, only one
20 genetic factor, the HR2 haplotype located in the *F5* gene, has been robustly found associated with FV plasma levels [3] even though it has been hypothesized that FV Leiden genotype could also modulate FV plasma levels [2,4].

SUMMARY OF THE INVENTION:

25 As defined by the claims, the present invention relates to methods of identifying subjects having or at risk of having a coagulation related disorder.

DETAILED DESCRIPTION OF THE INVENTION:

30 Factor V (FV) is a circulating protein primarily synthesized in the liver, and mainly present in plasma. It is a major component of the coagulation process. The objective of the inventors was to detect novel genetic loci participating to the regulation of FV plasma levels. They conducted the first Genome Wide Association Study on FV plasma levels in a sample of 510 individuals and replicated the main findings in an independent sample of 1156 individuals. In addition to genetic variations at the *F5* locus, the inventors identified novel associations at

the *PLXDC2* locus, with the lead *PLXDC2* rs927826 polymorphism explaining ~3.7% ($p = 7.5 \times 10^{-15}$ in the combined discovery and replication samples) of the variability of FV plasma levels. In silico transcriptomic analyses in various cell types confirmed that *PLXDC2* expression is positively correlated to *F5* expression. SiRNA experiments in human
5 hepatocellular carcinoma cell line confirmed the role of *PLXDC2* in modulating factor *F5* gene expression, and revealed further influences on *F2* and *F10* expressions. In conclusion the study identified *PLXDC2* as a new molecular player of the coagulation process.

Thus the first object of the present invention relates to a method of identifying a subject
10 having or at risk of having a coagulation related disorder due to a Factor V excess or deficiency comprising determining in a nucleic acid sample obtained from the subject the presence or absence of at least one genetic variant in the *PLXDC2* gene.

As used herein, the term “coagulation related disorder” comprises bleeding disorders
15 and clotting disorders that results from a deficiency or excess of Factor V.

The term “bleeding disorder” is meant to include uncontrolled and excessive bleeding. Bleeding episodes may be a major problem both in connection with surgery and other forms of tissue damage.

The term “clotting disorder” refers to any disorder in which there is a tendency toward
20 excessive clotting inside a vascular vessel (artery or vein).

As used herein, the term ‘Factor V’ or ‘FV’ has its general meaning in the art and refers to a protein of the coagulation system. The term is also known as coagulation factor V. The molecule circulates in plasma as a single-chain molecule.

As used herein, the term “Factor V deficiency” or “FV deficiency” refers to a state
25 wherein the Factor V plasma level is lower than the level measured in the general population.

As used herein, the term “Factor V excess” or “FV excess” refers to a state wherein the Factor V plasma level is higher than the level measured in the general population.

In particular, the method of the present of the present invention is particularly suitable
30 for identifying a subject having or at risk of having a bleeding disorder due to a Factor V deficiency. More particularly, the method of the present invention is particularly suitable for determining whether the subject is at risk of having haemorrhage.

In particular, the method of the present invention is particularly suitable for identifying a subject having or at risk of having a clotting disorder due to a Factor V deficiency or excess.

More particularly, the method of the present invention is particularly suitable for determining whether the subject is at risk of having a thromboembolic event selected from the group consisting of arterial thrombosis, fatal- and non-fatal myocardial infarction, stroke, transient ischemic attacks, cerebral venous thrombosis, peripheral arteriopathy, deep venous thrombosis and pulmonary embolism.

The subject can be male or female. A subject can be one who has been previously diagnosed as having some symptoms of a coagulation related disorder. A subject can be one who has been previously diagnosed as having a FV deficiency or excess. Alternatively, the subject can also be one who has not been previously diagnosed as having symptoms of a coagulation related disorder. For example, a subject can be one who exhibits one or more risk factors for coagulation related disorder, or a subject who does not exhibit risk factors, or a subject who is asymptomatic for coagulation related disorder.

As used herein, the term "risk" relates to the probability that an event will occur over a specific time period, as in the conversion to a coagulation related disorder, and can mean a subject's "absolute" risk or "relative" risk. Absolute risk can be measured with reference to either actual observation post-measurement for the relevant time cohort, or with reference to index values developed from statistically valid historical cohorts that have been followed for the relevant time period. Relative risk refers to the ratio of absolute risks of a subject compared either to the absolute risks of low risk cohorts or an average population risk, which can vary by how clinical risk factors are assessed. Odds ratios, the proportion of positive events to negative events for a given test result, are also commonly used (odds are according to the formula $p/(1-p)$ where p is the probability of event and $(1-p)$ is the probability of no event) to no-conversion. "Risk evaluation," or "evaluation of risk" in the context of the present invention encompasses making a prediction of the probability, odds, or likelihood that an event or disease state may occur, the rate of occurrence of the event or conversion from one disease state to another, i.e., from a normal condition to a coagulation related disorder or to one at risk of developing a coagulation related disorder. Risk evaluation can also comprise prediction of future clinical parameters, traditional laboratory risk factor values, or other indices of a coagulation related disorder, either in absolute or relative terms in reference to a previously measured population. The methods of the present invention may be used to make continuous or categorical measurements of the risk of conversion to a coagulation related disorder, thus diagnosing and defining the risk spectrum of a category of subjects defined as being at risk for a coagulation

related disorder. In the categorical scenario, the invention can be used to discriminate between normal and other subject cohorts at higher risk for a coagulation related disorder. In some embodiments, the present invention may be used so as to discriminate those at risk for developing a coagulation related disorder from normal.

5 As used herein, the term "nucleic acid sample" refers to any biological sample isolated from the subject liable to contain nucleic acid for the purpose of the present invention. Samples can include by way of example and not limitation, body fluids (e.g. saliva) and/or tissue extracts such as homogenates or solubilized tissue obtained from the subject. In some embodiments, the sample is a blood sample. The term "blood sample" means any blood sample derived from the
10 patient that contains nucleic acids. Peripheral blood is preferred, and mononuclear cells (PBMCs) are the preferred cells. The term "PBMC" or "peripheral blood mononuclear cells" or "unfractionated PBMC", as used herein, refers to whole PBMC, i.e. to a population of white blood cells having a round nucleus, which has not been enriched for a given sub-population. Typically, these cells can be extracted from whole blood using Ficoll, a hydrophilic
15 polysaccharide that separates layers of blood, with the PBMC forming a cell ring under a layer of plasma. Additionally, PBMC can be extracted from whole blood using a hypotonic lysis which will preferentially lyse red blood cells. Such procedures are known to the expert in the art. The template nucleic acid need not be purified. Nucleic acids may be extracted from a sample by routine techniques such as those described in Diagnostic Molecular Microbiology:
20 Principles and Applications (Persing et al. (eds), 1993, American Society for Microbiology, Washington D.C.).

 As used herein, the term "*PLXDC2*" refers to the gene encoding for the plexin domain-containing protein 2. The *PLXDC2* gene is known per and is available under the reference ENSG00000120594 in the Ensembl Gene Database.

25 As used herein, the term "genetic variant" has its general meaning in the art and denotes any of two or more alternative forms of a gene occupying the same chromosomal locus. The alteration typically consists in a substitution, an insertion, and/or a deletion, at one or more (e.g., several) positions in the gene. Genetic variation arises naturally through mutation, and may result in phenotypic polymorphism within populations. Gene mutations can be silent (no change
30 in the encoded polypeptide) or may encode polypeptides having altered amino acid sequence. The term is also known as "polymorphism".

 In some embodiments, the genetic variant is located in the promoter.

 In some embodiments, the genetic variant is located in an intron.

In some embodiments, the genetic variant is located in an exon.

In some embodiments, the genetic variant is present is heterozygous (i.e. present in only one allele) or homozygous (i.e. present in the 2 alleles).

5 In some embodiments, the method of the present invention comprises detecting one or more single nucleotide polymorphism (SNP).

In some embodiments, the genetic variant is a single nucleotide polymorphism. As used herein, the term "single nucleotide polymorphism" or "SNP" has its general meaning in the art and refers to a single nucleotide variation in a genetic sequence that occurs at appreciable frequency in the population.

10 In some embodiments, the method of the present invention comprises detecting at least one SNP selected in the group consisting of rs965787, rs200970303, rs1887035, rs927826, rs9651367, rs2148299, rs1018635, rs1409338, rs7895470, rs10159712, rs7070003. Preferably, said SNP is rs927826.

15 In some embodiments, the rs927826-G allele is associated with decreased FV plasma levels.

In some embodiments, the method of the present invention comprises detecting the any SNP in linkage disequilibrium with rs927826 yielding a pairwise linkage disequilibrium (LD) $r^2 > 0.8$ therewith. Preferably, the linkage disequilibrium yields a $r^2 > 0.90, 0.91, 0.92, 0.93, 0.94, 0.95, 0.96, 0.97, 0.98, \text{ or } 0.99$. In some embodiments, $r^2 = 1$ (i.e. SNPs are in complete linkage disequilibrium).

20

Detecting the genetic variant may be determined according to any genotyping method known in the art. Typically, common genotyping methods include, but are not limited to, TaqMan assays, molecular beacon assays, nucleic acid arrays, allele-specific primer extension, allele-specific PCR, arrayed primer extension, homogeneous primer extension assays, primer extension with detection by mass spectrometry, sequencing, multiplex primer extension sorted on genetic arrays, ligation with rolling circle amplification, homogeneous ligation, OLA, multiplex ligation reaction sorted on genetic arrays, restriction-fragment length polymorphism, single base extension-tag assays, and the Invader assay. Such methods may be used in combination with detection mechanisms such as, for example, luminescence or chemiluminescence detection, fluorescence detection, time-resolved fluorescence detection, fluorescence resonance energy transfer, fluorescence polarization, mass spectrometry, and electrical detection. Various methods for detecting polymorphisms include, but are not limited to, methods in which protection from cleavage agents is used to detect mismatched bases in

25

30

RNA/RNA or RNA/DNA, comparison of the electrophoretic mobility of variant and wild type nucleic acid molecules, and assaying the movement of polymorphic or wild-type fragments in polyacrylamide gels containing a gradient of denaturant using denaturing gradient gel electrophoresis. Sequence variations at specific locations can also be assessed by nuclease protection assays such as RNase and SI protection or chemical cleavage methods.

In some embodiments, genotyping is performed using the TaqMan assay, which is also known as the 5' nuclease assay. The TaqMan assay detects the accumulation of a specific amplified product during PCR. The TaqMan assay utilizes an oligonucleotide probe labeled with a fluorescent reporter dye and a quencher dye. The reporter dye is excited by irradiation at an appropriate wavelength, it transfers energy to the quencher dye in the same probe via a process called fluorescence resonance energy transfer (FRET). When attached to the probe, the excited reporter dye does not emit a signal. The proximity of the quencher dye to the reporter dye in the intact probe maintains a reduced fluorescence for the reporter. The reporter dye and quencher dye may be at the 5' most and the 3' most ends, respectively, or vice versa.

Alternatively, the reporter dye may be at the 5' or 3' most end while the quencher dye is attached to an internal nucleotide, or vice versa. In yet another embodiment, both the reporter and the quencher may be attached to internal nucleotides at a distance from each other such that fluorescence of the reporter is reduced. During PCR, the 5' nuclease activity of DNA polymerase cleaves the probe, thereby separating the reporter dye and the quencher dye and resulting in increased fluorescence of the reporter. Accumulation of PCR product is detected directly by monitoring the increase in fluorescence of the reporter dye. The DNA polymerase cleaves the probe between the reporter dye and the quencher dye only if the probe hybridizes to the target SNP-containing template which is amplified during PCR, and the probe is designed to hybridize to the target SNP site only if a particular SNP allele is present. Preferred TaqMan primer and probe sequences can readily be determined using the SNP and associated nucleic acid sequence information provided herein. A number of computer programs, such as Primer Express (Applied Biosystems, Foster City, Calif.), can be used to rapidly obtain optimal primer/probe sets. It will be apparent to one of skill in the art that such primers and probes for detecting the nucleic acids of the present invention are useful in diagnostic assays for stenosis and related pathologies, and can be readily incorporated into a kit format.

Another method for genotyping the nucleic acids of the present invention is the use of two oligonucleotide probes in an Oligonucleotide Ligation Assay (OLA). In this method, one probe hybridizes to a segment of a target nucleic acid with its 3' most end aligned with the

nucleic acid site. A second probe hybridizes to an adjacent segment of the target nucleic acid molecule directly 3' to the first probe. The two juxtaposed probes hybridize to the target nucleic acid molecule, and are ligated in the presence of a linking agent such as a ligase if there is perfect complementarity between the 3' most nucleotide of the first probe with the nucleic acid site. If there is a mismatch, efficient ligation cannot occur. After the reaction, the ligated probes are separated from the target nucleic acid molecule, and detected as indicators of the presence of a nucleic acid sequence. OLA may also be used for performing nucleic acid detection using universal arrays, wherein a zipcode sequence can be introduced into one of the hybridization probes, and the resulting product, or amplified product, hybridized to a universal zip code array. Alternatively OLA may be used where zipcodes are incorporated into OLA probes, and amplified PCR products are determined by electrophoretic or universal zipcode array readout. Alternatively one may use SNPLEX methods and software for multiplexed SNP detection using OLA followed by PCR, wherein zipcodes are incorporated into OLA probes, and amplified PCR products are hybridized with a zipchute reagent, and the identity of the SNP determined from electrophoretic readout of the zipchute. In some embodiments, OLA is carried out prior to PCR (or another method of nucleic acid amplification). In some other embodiments, PCR (or another method of nucleic acid amplification) is carried out prior to OLA.

Another method for genotyping is based on mass spectrometry. Mass spectrometry takes advantage of the unique mass of each of the four nucleotides of DNA. Nucleic acids can be unambiguously genotyped by mass spectrometry by measuring the differences in the mass of nucleic acids having alternative nucleic acid alleles. MALDI-TOF (Matrix Assisted Laser Desorption Ionization—Time of Flight) mass spectrometry technology is preferred for extremely precise determinations of molecular mass, such as for SNPs. Numerous approaches to genotype analysis have been developed based on mass spectrometry. Preferred mass spectrometry-based methods of nucleic acid genotyping include primer extension assays, which can also be utilized in combination with other approaches, such as traditional gel-based formats and microarrays. Typically, the primer extension assay involves designing and annealing a primer to a template PCR amplicon upstream (5') from a target nucleic acid position. A mix of dideoxynucleotide triphosphates (ddNTPs) and/or deoxynucleotide triphosphates (dNTPs) are added to a reaction mixture containing template. For example, in some embodiments this is a SNP-containing nucleic acid molecule which has typically been amplified, such as by PCR. Primer and DNA polymerase may further be added. Extension of the primer terminates at the first position in the template where a nucleotide complementary to one of the ddNTPs in the

mix occurs. The primer can be either immediately adjacent (i.e., the nucleotide at the 3' end of the primer hybridizes to the nucleotide next to the target SNP site) or two or more nucleotides removed from the nucleic acid position. If the primer is several nucleotides removed from the target nucleic acid position, the only limitation is that the template sequence between the 3' end of the primer and the nucleic acid position cannot contain a nucleotide of the same type as the one to be detected, or this will cause premature termination of the extension primer. Alternatively, if all four ddNTPs alone, with no dNTPs, are added to the reaction mixture, the primer will always be extended by only one nucleotide, corresponding to the target SNP position. In this instance, primers are designed to bind one nucleotide upstream from the SNP position (i.e., the nucleotide at the 3' end of the primer hybridizes to the nucleotide that is immediately adjacent to the target SNP site on the 5' side of the target SNP site). Extension by only one nucleotide is preferable, as it minimizes the overall mass of the extended primer, thereby increasing the resolution of mass differences between alternative SNP nucleotides. Furthermore, mass-tagged ddNTPs can be employed in the primer extension reactions in place of unmodified ddNTPs. This increases the mass difference between primers extended with these ddNTPs, thereby providing increased sensitivity and accuracy, and is particularly useful for typing heterozygous base positions. Mass-tagging also alleviates the need for intensive sample-preparation procedures and decreases the necessary resolving power of the mass spectrometer. The extended primers can then be purified and analyzed by MALDI-TOF mass spectrometry to determine the identity of the nucleotide present at the target SNP position. In one method of analysis, the products from the primer extension reaction are combined with light absorbing crystals that form a matrix. The matrix is then hit with an energy source such as a laser to ionize and desorb the nucleic acid molecules into the gas-phase. The ionized molecules are then ejected into a flight tube and accelerated down the tube towards a detector. The time between the ionization event, such as a laser pulse, and collision of the molecule with the detector is the time of flight of that molecule. The time of flight is precisely correlated with the mass-to-charge ratio (m/z) of the ionized molecule. Ions with smaller m/z travel down the tube faster than ions with larger m/z and therefore the lighter ions reach the detector before the heavier ions. The time-of-flight is then converted into a corresponding, and highly precise, m/z . In this manner, SNPs can be identified based on the slight differences in mass, and the corresponding time of flight differences, inherent in nucleic acid molecules having different nucleotides at a single base position.

Detecting the genetic variant may also be performed by sequencing. A variety of automated sequencing procedures can be used, including sequencing by mass spectrometry. The nucleic acid sequences of the present invention enable one of ordinary skill in the art to readily design sequencing primers for such automated sequencing procedures. Commercial instrumentation, such as the Applied Biosystems 377, 3100, 3700, 3730, and 3730×1 DNA Analyzers (Foster City, Calif.), is commonly used in the art for automated sequencing. Nucleic acid sequences can also be determined by employing a high throughput mutation screening system, such as the SpectruMedix system.

Other methods that can be used to genotype the nucleic acids of the present invention include single-strand conformational polymorphism (SSCP), and denaturing gradient gel electrophoresis (DGGE). SSCP identifies base differences by alteration in electrophoretic migration of single stranded PCR products. Single-stranded PCR products can be generated by heating or otherwise denaturing double stranded PCR products. Single-stranded nucleic acids may refold or form secondary structures that are partially dependent on the base sequence. The different electrophoretic mobilities of single-stranded amplification products are related to base-sequence differences at nucleic acid positions. DGGE differentiates nucleic acid alleles based on the different sequence-dependent stabilities and melting properties inherent in polymorphic DNA and the corresponding differences in electrophoretic migration patterns in a denaturing gradient gel.

Sequence-specific ribozymes can also be used to score nucleic acids, in particular SNPs, based on the development or loss of a ribozyme cleavage site. Perfectly matched sequences can be distinguished from mismatched sequences by nuclease cleavage digestion assays or by differences in melting temperature. Thus, for example, if the SNP affects a restriction enzyme cleavage site, the SNP can be identified by alterations in restriction enzyme digestion patterns, and the corresponding changes in nucleic acid fragment lengths determined by gel electrophoresis. Genotyping can include the steps of, for example, collecting the sample, isolating nucleic acids (e.g., genomic DNA, mRNA or both) from the cells of the sample, contacting the nucleic acids with one or more primers which specifically hybridize to a region of the isolated nucleic acid containing a target nucleic acid region of interest under conditions such that hybridization and amplification of the target nucleic acid region occurs, and determining the nucleotide present at the nucleic acid position of interest, or, in some assays, detecting the presence or absence of an amplification product (assays can be designed so that hybridization and/or amplification will only occur if a particular nucleic acid sequence allele is present or absent). In some assays, the size of the amplification product is detected and

compared to the length of a control sample; for example, deletions and insertions can be detected by a change in size of the amplified product compared to a normal genotype. Methods of comparing the identity of two or more sequences may be performed by any reasonable means, including programs available in the Wisconsin Sequence Analysis Package version 9.1 (Genetics Computer Group, Madison, Wis., USA). Other programs such as BESTFIT may be used to find the “local homology” algorithm of Smith and Waterman and finds the best single region of similarity between two sequences. Further, programs such as GAP may be used, which aligns two sequences finding a “maximum similarity.” Preferably, % identities and similarities are determined when the two sequences being compared are optimally aligned. Other programs for determining identity and/or similarity between sequences are also known in the art, for instance the BLAST family of programs, available from the National Center for Biotechnology Information (NCBI), Bethesda, Md., USA) and FASTA, available as part of the Wisconsin Sequence Analysis Package.

In some embodiments, the method of the present invention further comprises measuring at least one other risk factor.

In some embodiments, the method of the present invention further comprises a step consisting of calculating a score, representing an estimation value of the risk that the subject has a coagulation related disorder. Typically, the score is based the presence or absence of at least one genetic variant in the PLXDC2 gene and may typically include another risk factor as well as the level of Factor V measured in a plasma sample obtained from the subject. Typically, other risk factors may include clinical features such as age, gender, diabetes mellitus, smoking, family history of coagulation related disorder, pregnancy, and body mass index. Based the above input features obtained from the subject, an operator can calculate a numerical function of the above list of inputs by applying an algorithm. For instance this numerical function may return a number, i.e. risk score (R), for instance between zero and one, where zero is the lowest possible risk indication and one is the highest. This numerical output may also be compared to a threshold (T) value between zero and one. If the risk score exceeds the threshold T, it is meant than the subject has a high risk of having a coagulation related disorder and if the risk score is under the threshold T, it is meant than the subject has a low risk of having a coagulation related disorder.

In some embodiments, the method of the invention thus comprises the use of an algorithm.

In some embodiments, the algorithm is a classification algorithm typically selected from Multivariate Regression Analysis, Linear Discriminant Analysis (LDA), Topological Data Analysis (TDA), Neural Networks, Support Vector Machine (SVM) algorithm and Random Forests algorithm (RF) such as described in the Example. In some embodiments, the method of the invention comprises the step of determining the subject response using a classification algorithm. As used herein, the term "classification algorithm" has its general meaning in the art and refers to classification and regression tree methods and multivariate classification well known in the art such as described in US 8,126,690; WO2008/156617. As used herein, the term "support vector machine (SVM)" is a universal learning machine useful for pattern recognition, whose decision surface is parameterized by a set of support vectors and a set of corresponding weights, refers to a method of not separately processing, but simultaneously processing a plurality of variables. Thus, the support vector machine is useful as a statistical tool for classification. The support vector machine non-linearly maps its n-dimensional input space into a high dimensional feature space, and presents an optimal interface (optimal parting plane) between features. The support vector machine comprises two phases: a training phase and a testing phase. In the training phase, support vectors are produced, while estimation is performed according to a specific rule in the testing phase. In general, SVMs provide a model for use in classifying each of n subjects to two or more disease categories based on one k-dimensional vector (called a k-tuple) of biomarker measurements per subject. An SVM first transforms the k-tuples using a kernel function into a space of equal or higher dimension. The kernel function projects the data into a space where the categories can be better separated using hyperplanes than would be possible in the original data space. To determine the hyperplanes with which to discriminate between categories, a set of support vectors, which lie closest to the boundary between the disease categories, may be chosen. A hyperplane is then selected by known SVM techniques such that the distance between the support vectors and the hyperplane is maximal within the bounds of a cost function that penalizes incorrect predictions. This hyperplane is the one which optimally separates the data in terms of prediction (Vapnik, 1998 Statistical Learning Theory. New York: Wiley). Any new observation is then classified as belonging to any one of the categories of interest, based where the observation lies in relation to the hyperplane. When more than two categories are considered, the process is carried out pairwise for all of the categories and those results combined to create a rule to discriminate between all the categories. As used herein, the term "Random Forests algorithm" or "RF" has its general meaning in the art and refers to classification algorithm such as described in US 8,126,690; WO2008/156617. Random Forest is a decision-tree-based classifier that is constructed using an algorithm

originally developed by Leo Breiman (Breiman L, "Random forests," Machine Learning 2001, 45:5-32). The classifier uses a large number of individual decision trees and decides the class by choosing the mode of the classes as determined by the individual trees. The individual trees are constructed using the following algorithm: (1) Assume that the number of cases in the training set is N, and that the number of variables in the classifier is M; (2) Select the number of input variables that will be used to determine the decision at a node of the tree; this number, m should be much less than M; (3) Choose a training set by choosing N samples from the training set with replacement; (4) For each node of the tree randomly select m of the M variables on which to base the decision at that node; (5) Calculate the best split based on these m variables in the training set. In some embodiments, the score is generated by a computer program.

In some embodiments, the method of the present invention thus comprises a) detecting at least one genetic variant in the PLXDC2 gene; b) implementing an algorithm on data comprising the presence or absence of the genetic variant so as to obtain an algorithm output; c) determining the probability that the subject will have a coagulation related disorder.

Once the subject is identified as having or being at risk of having a coagulation related disorder the physician can optionally further prescribe preventive therapy. Typically the treatment includes, for example, weight management, physical activity, limitation of alcohol and cigarette consumption, or prescription of a particular drug. For instance, if the patient is at risk of having a thromboembolic event, preventive administration of anticoagulants and/or antiplatelet drugs may be prescribed. Non-limiting examples of anticoagulants include: coumarins, heparin, warfarin, acenocoumarol, phenprocoumon, atromentin, phenindione, fondaparinux, idraparinux, direct factor Xa inhibitors, direct thrombin inhibitors, antithrombin protein therapeutics, batroxobin, and hementin. Non-limiting examples of antiplatelet drugs include: irreversible cyclooxygenase inhibitors (e.g., aspirin or triflusal), adenosine diphosphate receptor inhibitors (e.g., clopidogrel, prasugrel, ticagrelor, or ticlopidine), phosphodiesterase inhibitors (e.g., cilostazol), glycoprotein IIB/IIIA inhibitors (e.g., abciximab, eptifibatide, or tirofiban), adenosine reuptake inhibitors (e.g., dipyridamole), and thromboxane inhibitors (e.g., thromboxane synthase inhibitors or thromboxane receptor antagonists).

A further object of the present invention relates to a method of treating a coagulation related disorder due to a Factor V excess in patient in need thereof comprising administering to the patient a therapeutically effective amount of an inhibitor of PLXDC2 expression.

An "inhibitor of expression" refers to a natural or synthetic compound that has a biological effect to inhibit the expression of a gene. In some embodiments, said inhibitor of gene expression is a siRNA, an antisense oligonucleotide or a ribozyme. For example, antisense oligonucleotides, including anti-sense RNA molecules and anti-sense DNA molecules, would act to directly block the translation of the mRNA by binding thereto and thus preventing protein translation or increasing mRNA degradation, thus decreasing the level of the polypeptide (e.g. PLXDC2), and thus activity, in a cell. For example, antisense oligonucleotides of at least about 15 bases and complementary to unique regions of the mRNA transcript sequence encoding the polypeptide (e.g. PLXDC2) can be synthesized, e.g., by conventional phosphodiester techniques. Methods for using antisense techniques for specifically inhibiting gene expression of genes whose sequence is known are well known in the art (e.g. see U.S. Pat. Nos. 6,566,135; 6,566,131; 6,365,354; 6,410,323; 6,107,091; 6,046,321; and 5,981,732). Small inhibitory RNAs (siRNAs) can also function as inhibitors of expression for use in the present invention. Gene expression can be reduced by contacting a patient or cell with a small double stranded RNA (dsRNA), or a vector or construct causing the production of a small double stranded RNA, such that gene expression is specifically inhibited (i.e. RNA interference or RNAi). Antisense oligonucleotides, siRNAs, shRNAs and ribozymes of the invention may be delivered in vivo alone or in association with a vector. In its broadest sense, a "vector" is any vehicle capable of facilitating the transfer of the antisense oligonucleotide, siRNA, shRNA or ribozyme nucleic acid to the cells. Typically, the vector transports the nucleic acid to cells with reduced degradation relative to the extent of degradation that would result in the absence of the vector. In general, the vectors useful in the invention include, but are not limited to, plasmids, phagemids, viruses, other vehicles derived from viral or bacterial sources that have been manipulated by the insertion or incorporation of the antisense oligonucleotide, siRNA, shRNA or ribozyme nucleic acid sequences. Viral vectors are a preferred type of vector and include, but are not limited to nucleic acid sequences from the following viruses: retrovirus, such as moloney murine leukemia virus, harvey murine sarcoma virus, murine mammary tumor virus, and rous sarcoma virus; adenovirus, adeno-associated virus; SV40-type viruses; polyoma viruses; Epstein-Barr viruses; papilloma viruses; herpes virus; vaccinia virus; polio virus; and RNA virus such as a retrovirus. One can readily employ other vectors not named but known to the art. In some embodiments, the inhibitor of expression is an endonuclease. In a particular embodiment, the endonuclease is CRISPR-cas. In some embodiment, the endonuclease is CRISPR-cas9, which is from *Streptococcus pyogenes*. The CRISPR/Cas9 system has been described in US 8697359 B1 and US 2014/0068797. In some embodiment, the endonuclease is

CRISPR-Cpf1, which is the more recently characterized CRISPR from Provotella and Francisella 1 (Cpf1) in Zetsche et al. ("Cpf1 is a Single RNA-guided Endonuclease of a Class 2 CRISPR-Cas System (2015); Cell; 163, 1-13).

5 A "therapeutically effective amount" refers to an amount effective, at dosages and for periods of time necessary, to achieve a desired therapeutic result. A therapeutically effective amount of drug may vary according to factors such as the disease state, age, sex, and weight of the individual, and the ability of drug to elicit a desired response in the individual. A therapeutically effective amount is also one in which any toxic or detrimental effects of the
10 antibody or antibody portion are outweighed by the therapeutically beneficial effects. The efficient dosages and dosage regimens for drug depend on the disease or condition to be treated and may be determined by the persons skilled in the art. A physician having ordinary skill in the art may readily determine and prescribe the effective amount of the pharmaceutical composition required. For example, the physician could start doses of drug employed in the
15 pharmaceutical composition at levels lower than that required in order to achieve the desired therapeutic effect and gradually increase the dosage until the desired effect is achieved. In general, a suitable dose of a composition of the present invention will be that amount of the compound, which is the lowest dose effective to produce a therapeutic effect according to a particular dosage regimen. Such an effective dose will generally depend upon the factors
20 described above. For example, a therapeutically effective amount for therapeutic use may be measured by its ability to stabilize the progression of disease. A therapeutically effective amount of a therapeutic compound may decrease tumour size, or otherwise ameliorate symptoms in a subject. One of ordinary skill in the art would be able to determine such amounts based on such factors as the subject's size, the severity of the subject's symptoms, and the
25 particular composition or route of administration selected. An exemplary, non-limiting range for a therapeutically effective amount of drug is about 0.1-100 mg/kg, such as about 0.1-50 mg/kg, for example about 0.1-20 mg/kg, such as about 0.1-10 mg/kg, for instance about 0.5, about such as 0.3, about 1, about 3 mg/kg, about 5 mg/kg or about 8 mg/kg. An exemplary, non-limiting range for a therapeutically effective amount of an antibody of the present invention
30 is 0.02-100 mg/kg, such as about 0.02-30 mg/kg, such as about 0.05-10 mg/kg or 0.1-3 mg/kg, for example about 0.5-2 mg/kg.

The invention will be further illustrated by the following figures and examples. However, these examples and figures should not be interpreted in any way as limiting the scope of the present invention.

FIGURES:

Figure 1. Impact of *PLXDC2* gene silencing on *F5* gene expression in liver

Abbreviation: KD=siRNA KnockDown

Quantification of *PLXDC2* and *F5*, *F2*, *F7* and *F10* mRNA levels in Hep3B cells after 24h treatment with 50nM siRNA directed to *PLXDC2* gene (n = 9 for both controls and siRNA KD). A 79% decrease ($p < 10^{-8}$) in *PLXDC2* gene expression was followed by a significant decrease in *F5* (-27%, $p = 0.02$), *F2* (-21%, $p = 0.004$) and *F10* (-23%, $p = 0.0003$) gene expression, while *F7* expression was not significantly decreased (-12%, $p = 0.11$).

EXAMPLE:

Methods:

Study description

This work builds on two independent samples of unrelated VT patients of European ancestry recruited at the Thrombophilia center of La Timone Hospital (Marseille, France), the MARTHA and MARTHA12 cohorts [5]. MARTHA patients were used for the discovery GWAS while MARTHA12 individuals were considered for the replication step. All participants provided written informed consent, and the protocol was approved by the ethics committee of the participating institution.

The MARseille THrombosis Association (MARTHA) project has already been extensively described [6-7]. It is composed of unrelated subjects of European origin, with the majority being of French ancestry, consecutively recruited at the Thrombophilia center of La Timone hospital (Marseille, France) between January 1994 and October 2012. All patients had a documented history of VT and are free of well characterized genetic risk factors including AT, PC, or PS deficiency, homozygosity for FV Leiden or FII G20210A, and lupus anticoagulant. The MARTHA12 study is an independent sample of 1,245 VT patients recruited between 2010 and 2012 according to the same criteria as the MARTHA patients [5,8].

Haemostatic traits measurements

FV activity plasma levels (FV:C) were measured using human FV deficient plasma on automated coagulometers (STA® analysers from Diagnostica Stago). Activated partial thromboplastin time (aPTT) and prothrombin time (PT) were measured in plasma with the use

of automated coagulometers. Tests were conducted on the same day as blood collection or within a few weeks (after freezing). PT was recorded as a ratio of the patient's PT to the mean normal PT, and aPTT was recorded in seconds (s). FV:C plasma levels were available in 510 participants of the discovery phase and 1156 participants of the replication. FV antigen plasma levels (FV:Ag) were measured by ELISA on freezing samples using the ZYMUTEST Factor V kit (Hyphen Biomed, France) in a random sample of 60 patients from the MARTHA study (n=20 per rs927826 genotypes). Factor II (FII) and Factor X (FX) activity levels in plasma were measured in MARTHA and MATHA12 cohorts by using human FII or FX deficient plasma on automated coagulometers in a total of 738 patients with no anti vitamin K antagonists (VKA) at the time of blood sampling.

Genotyping

DNA samples of the MARTHA participants were typed with high-density genotyping arrays and imputed for SNPs available in the 1000G reference as previously described [5]. In MARTHA12, wet-lab genotyping was performed using Taqman assay. Genotyping was done using Taqman 5' nuclease assay (ThermoFischer Scientific, ref. C__2386120_10 (rs27218, *MAST4*), C__8879235_10 (rs1409338, *PLXDC2*)), following the supplier instructions, on a LC480® Real time PCR Instrument (Roche Life sciences), using dedicated Dual Color Hydrolysis Probe program. Reactions were made in a 10µl final volume in a 1X final concentration of Genotyping Master Mix (ThermoFischer Scientific ref. 4371355), with 25ng of DNA, and 0,1µl of 20X Genotyping Assay containing probes, nucleotides and DNA Taq Polymerase.

Genetic Association Analyses

For the GWAS discovery phase, association analyses of imputed SNPs with plasma FV levels were performed using linear regression analysis adjusted for age, sex and the four first principal components derived from the GWAS data as implemented in the Mach2QTL program [9]. Only SNPs with acceptable imputation quality ($r^2 > 0.5$) and minor allele frequency > 0.05 were considered. At the replication stage, associations of SNPs with plasma FV levels were assessed using standard linear regression analysis adjusted for age and sex. Haplotype association analyses were conducted using the THESIAS software [10]. Results from the discovery and replication studies were meta-analyzed via a fixed-effects model based on the inverse-variance weighting method. A replicated SNP was further tested for association with additional haemostatic traits including PT, aPTT, FII and FX circulating levels. For these analyses, patients with VKA at the time of blood sampling were excluded.

RNA interference mediated PLXDC2 silencing experiments

Small interference RNA (siRNA) *PLXDC2* gene silencing was conducted in human hepatocytes, a key cell type for *F5* regulation, to validate *in vitro* the association between *PLXDC2* and *F5* gene expressions. Additional expression of coagulation factors *F2*, *F7* and *F10* were also measured to assess the specificity of the *PLXDC2* association with *F5* expression.

The human hepatocellular carcinoma cell line Hep3B (American Type Culture Collection, Rockville MD, USA) was grown at 37°C in 5% CO₂ in Dulbecco's Modified Eagle's Medium containing 10% fetal calf serum, 2 mmol/L glutamine and 100 U/mL penicillin/streptomycin. Cells were seeded on 12-well plates at 100,000 cells per well and were transfected 72 hours later with 50 nM control siRNA or an siRNA targeting the *PLXDC2* gene, s39607 (sequence ctacagaagatgataccaa from Thermofisher) using lipofectamine RNAiMax (Life Technologies) according to the manufacturer's instructions. Twenty-four hours following transfection, total RNA was extracted (NucleoSpin RNA II kit, Macherey-Nagel) and reverse transcribed (High-capacity cDNA reverse transcription kit, Life Technologies) and gene expression was analyzed by real-time qPCR using a LightCycler LC480 (Roche). Primers used for quantification of *PLXDC2* and *F5*, *F2*, *F7* and *F10* are provided as supplemental material. Expression of mRNA levels was normalized to human non-POU domain containing, octamer-binding housekeeping gene (NONO), human alanin (ALA) and human heat shock protein 90kDa alpha (cytosolic) class B member 1 (HSP90AB1). Data (Table S1) were expressed as a fold change in mRNA expression relative to control cells.

The impact of *PLXDC2* knock-down on gene expression was tested via two-way analysis of variance (ANOVA).

Results:

FV activity plasma levels (FV:C) were measured using human FV deficient plasma on automated coagulometers. The discovery GWAS was composed of 510 patients with venous thromboembolism (VTE) assessed for 6,264,382 single nucleotide polymorphisms (SNPs) and main significant findings were tested for replication in an independent sample of 1,156 patients (see Materials & Methods). Main clinical and biological characteristics of the studied samples are shown in **Table 1**.

About 70 SNPs achieved genome-wide significance ($p < 5 \times 10^{-8}$), with all SNPs mapping to the *F5* locus on chromosome 1q24.2. The lead SNP, rs72708008, with association p -value = 8.43×10^{-12} , was in high linkage disequilibrium (LD) ($r^2 > 0.80$) with several SNPs including the rs6027 (p.Asp2222Gly, $p = 1.27 \times 10^{-11}$) known to tag the *F5* HR2 haplotype whose association with FV:C plasma levels has already been established [11]. After conditioning on rs6027, no additional signal reached genome-wide significance (**data not**

shown), the smallest p-value being now $p = 3.05 \times 10^{-7}$ and observed at the *PLXDC2* locus. As two other SNPs at the *F5* locus, rs6025 (p.Arg534Gln) and rs4524 (p.Lys858Arg), have been extensively studied in relation to VT risk [5], we further analyzed their joint association with rs6027 with respect to FV:C plasma levels in both the discovery and replication cohorts using
 5 haplotype analysis. This analysis revealed that in addition to the strong decreasing effect of the rs6027-C allele ($\beta = -0.151 \pm 0.02$, $p = 1.93 \times 10^{-12}$), the rs4524-C allele was associated with a slight decrease ($\beta = -0.033 \pm 0.01$, $p = 0.003$) in FV:C plasma levels (**data not shown**). Altogether, these two *F5* SNPs explained ~7% of the variability of FV:C plasma levels.

Aside the aforementioned *F5* association signal, two other loci with several SNPs in LD
 10 exhibited suggestive statistical evidence for association with FV:C plasma levels in the discovery cohort, with p-values ranging between $\sim 10^{-6} - 10^{-5}$ (**data not shown**). These loci were *MAST4* and *PLXDC2* where lead SNPs, rs27218 (intronic; $p = 9.02 \times 10^{-7}$) and rs927826 (intronic; $p = 1.10 \times 10^{-6}$), respectively, were looked for replication. While the association of *MAST4* rs27218 did not replicate (**Table 2**), the association of *PLXDC2* rs927826 with plasma
 15 FV:C levels was confirmed ($p = 2.89 \times 10^{-10}$). In both the discovery and replication samples, the rs927826-G allele was associated with decreased FV:C plasma levels, $\beta = -0.076 \pm 0.016$ ($p = 1.10 \times 10^{-6}$) and $\beta = -0.062 \pm 0.010$ ($p = 2.89 \times 10^{-10}$), respectively. When both samples were combined together, the overall statistical evidence for association of rs927826 with FV:C levels reached $p = 7.54 \times 10^{-15}$, with no evidence for heterogeneity across the two samples ($p =$
 20 0.458), and with the rs927826 explaining ~3.5% of FV:C plasma variability. As a reminder, the *PLXDC2* locus was the top locus in the GWAS analysis conditioned on *F5* rs6027 (**data not shown**) and the rs927826 ranked 4th ($p = 3.13 \times 10^{-7}$) in this conditional GWAS.

In order to assess whether *PLXDC2* could be a true determinant of FV levels and not a determinant of the clotting assay used to measure FV:C levels, we measured FV antigen
 25 (FV:Ag) levels by ELISA in a random sample of 60 patients from the MARTHA study. The ELISA data confirmed the specificity of the association of *PLXDC2* genotypes with FV plasma levels as the rs927826-G allele was associated with significant ($p = 0.028$) decrease of FV:Ag levels (**Table 3**). Note, in this subsample, the correlation between the two FV measurements was 0.76.

30 We further examined the correlation between *PLXDC2* and *F5* gene expression in several genome-wide gene expression data from multiple cell lines and tissues [12-17]. In all investigated resources except macrophages, we observed positive correlations between *PLXDC2* and *F5* expressions, the strongest correlation ($r = 0.45$, $p = 3.94 \times 10^{-14}$) being observed in liver (**data not shown**) where FV synthesis mainly occurs (**Table 4**). We also

assessed the correlation of liver *PLXDC2* expression with that of other coagulation genes expressed in the liver (*F2*, *F7* and *F10*). We observed a significant positive, but less strong, correlation with *F2* expression ($r = 0.17$, $p = 0.007$) and a very low correlation, if any, with *F7* and *F10* expressions (**Table 4**).

To strengthen the perspective of a specific effect of the rs927826 variant on FV levels, we tested its association with FII and FX plasma levels in 738 participants. Even though FII and FX were moderately correlated with FV plasma levels ($r = 0.35$, $p = 4.8 \times 10^{-17}$ and $r = 0.32$, $p = 4.40 \times 10^{-14}$ for FII and FX, respectively), we did not find any association between rs927826 and both FII ($\beta = -0.002 \pm 0.01$, $p = 0.83$) and FX ($\beta = 0.003 \pm 0.01$, $p = 0.81$) plasma levels **data not shown**).

To follow-up on these genetic epidemiological findings, we conducted preliminary *in vitro* study to assess whether *PLXDC2* gene expression could associate with *F5* gene expression in human liver cells (see Materials and Methods). As shown in **Figure 1**, knock-down expression of *PLXDC2* using siRNA was associated with a significant decrease liver expression of *F5* ($p = 0.020$). Concurrently, expressions of other coagulation genes (*F2*, *F7* and *F10*) were also measured, and we also observed a significant decrease of *F2* ($p = 0.004$) and *F10* ($p = 0.0003$) expressions when *PLXDC2* is silenced. *F7* expression remained unchanged ($p = 0.1$).

Discussion:

Altogether, these results strongly support the role of *PLXDC2* in the regulation of *F5* gene, and more generally in the coagulation cascade. Interestingly, the *PLXDC2* rs927826 has recently been found associated with activated partial thromboplastin time in a Japanese population but with no formal replication of the statistical findings nor experimental validation [18]. We did not observe such association in our population but did observe an association of rs927826 with prothrombin time (**data not shown**). This polymorphism is in very strong LD ($r^2 > 0.80$) with other *PLXDC2* SNPs (**data not shown**), all located in intronic regions subject to epigenetic regulation [19] but only the rs927826 is so far predicted to affect transcription factors' binding [20].

Little is known about the *PLXDC2* gene/protein. It is recognized as a cell surface transmembrane receptor expressed in various tissues (see, e.g., GTEx Portal [21]) without so far clear arguments that it could be directly involved in mRNA regulation. We showed that *PLXDC2* expression correlated with *F5* expression in different cell lines and that *F5* mRNA expression was significantly reduced in the liver in the *PLXDC2* knock-down experiment. Interestingly, the siRNA experiment revealed that the *PLXDC2* gene expression could also

modulate the liver expression of *F2* and *F10* genes that might indicate a more generalized effect on coagulation. However, we did not find any association between the rs927826 variant and either FII or FX plasma levels.

Very interestingly, in a recent plasma proteomic profiling study [22], the *F5* rs6027 and *PLXDC2* rs927826 variants we found here associated with plasma FV levels were both found influencing the plasma levels of the same 4 proteins (CD3E, CDH7, SLC22A16 and TOR1AIP1), such observations adding support for *PLXDC2* and *F5* belonging to a same regulatory pathway.

The strength of our study lies in its novelty. Indeed, there are limited information currently available regarding the regulation of plasma levels of Factor V. We performed the first GWAS on Factor V plasma levels and replicated the findings in an independent sample of subjects which limited the risk of spurious findings. Moreover, we increased the specificity of the finding by replicating the observed association between *PLXDC2* rs927826 genotypes and FV activity plasma levels by using an ELISA measuring FV antigen levels in plasma. Several limitations must be acknowledged. First, this study was performed in VT patients and validation of the observed genetic association deserves to be investigated in healthy individuals in order to assess whether the observed effect size also holds in non-diseased individuals. Second, the size of our discovery cohort was relatively modest compared to what is currently done in a GWAS context which has likely hampered our chance to detect additional loci participating to FV regulation. Additional efforts would be needed to measure FV plasma levels in independent cohorts with both available plasma and GWAS data in order to better characterize the genetic regulation underlying FV plasma variability. With large samples, it would also be interesting to assess the contribution of rare variants which was not possible in the current GWAS study mainly focusing on common polymorphisms. Finally, *PLXDC2* is recognized as a cell surface transmembrane receptor and it is then unclear how it could be involved in the regulation of *F5* gene expression. Moreover, the correlation between *PLXDC2* and *F2* expression in the liver and the significant decrease of *F2* and *F10* gene expressions in the *PLXDC2* knock-down experiment might indicate a more generalized effect on coagulation. Further experimental studies are mandatory to decipher the underlying mechanism.

In conclusion, all these observations point out the existence of a new player in the coagulation cascade whose exact molecular contribution needs to be extensively investigated. Its impact on FV mediated coagulation related disorders also warrants further investigations as whole blood *PLXDC2* expression levels have been reported to be associated with stroke [23].

TABLES:

Table 1. Description of the MARTHA and MARTHA12 cohorts

	Discovery MARTHA N = 510	Replication MARTHA12 N = 1156
Age	47.3 ± 16.0	50.4 ± 15.4
Gender (M/F)	176 / 334	527 / 629
Current smoker (%)	24.1 %	20.2 %
FV Leiden carriers (%)	15.5 %	11.0 %
BMI, kg/m ²	25.34 ± 4.96	26.26 ± 5.06
Vitamin K antagonist users (%)	3.7 %	28 %
FV:C plasma levels (IU/mL)	1.068 ± 0.235	1.089 ± 0.217
Activated partial thromboplastin time, sec	31.68 ± 3.71	33.61 ± 6.95
Prothrombine Time (%)	99.11 ± 14.72	83.64 ± 30.03

Mean ± Standard Deviation

5 **Table 2. Association with FV activity plasma levels of lead SNPs at two loci that reached suggestive evidence for association in the discovery GWAS**

Discovery	MAST4 rs27218 ^a				PLXDC2 rs927826 ^{a,b}			
	CC	CT	TT	AF	TT	T	G	AF
Mean FV plasma levels ± SD	1.0 3 ± 0.22 N = 271	1.0 9 ± 0.23 N = 199	1.2 3 ± 0.29 N = 40	.274	1.1 1 ± 0.22 N = 249	1. 04 ± 0.24 N = 224	0. 95 ± 0.19 N = 37	.293
Additive Allele effect ^c	+0.075 ± 0.015 p = 9.02 10 ⁻⁷				-0.076 ± 0.016 p = 1.10 10 ⁻⁸			
Replication	CC	CT	TT	AF	TT	T	G	AF
	CC	CT	TT	AF	TT	T	G	AF
Mean FV plasma levels ± SD	1.1 02 ± 0.21 N = 605	1.0 7 ± 0.22 N = 434	1.1 01 ± 0.21 N = 89	.271	1.1 3 ± 0.22 N = 537	1. 05 ± 0.21 N = 461	1. 02 ± 0.20 N = 87	.293
Additive Allele effect	-0.019 ± 0.010 p = 0.054				-0.062 ± 0.010 p = 2.89 10 ⁻¹⁰			

Abbreviations: MAF = Minor Allele Frequency; SD = Standard Deviation

^a Imputation quality criterion was 0.984 and 0.996 for MAST4 rs27218 and PLXDC2 rs927826, respectively, in the discovery GWAS.

10 ^b In the replication study, the rs927826 was substituted by the rs1409338 that served as a proxy ($r^2 = 0.97$) due to some technical issues in wet-lab genotyping the rs927826.

^c Association testing was adjusted for age, sex and the 4 main genetics components in the discovery cohort and on age and sex in the replication study.

Table 3. Association of rs927826 genotypes with plasma FV antigen (FV:Ag) and activity (FV:C) in a sample of 60 patients from the MARTHA cohort

rs927826	FV:Ag (IU/mL)	FV:C (IU/mL)
TT (n = 20)	0.647 ± 0.163	1.182 ± 0.271
TG (n = 20)	0.621 ± 0.133	1.204 ± 0.265
GG (n = 20)	0.545 ± 0.140	1.006 ± 0.200
Pvalue ^a	p = 0.0286	p = 0.0263

Mean ± Standard Deviation

^a Associations were tested using a linear regression model adjusted for age and sex, under the assumption of an additive model.

Table 4: Correlations between *PLXDC2* and *F2* / *F5* / *F7* / *F10* gene expressions in 6 different tissues

	Macrophages ^a N = 684	Whole Blood ^b N = 78	Monocytes ^c N = 849	Endothelial Cells ^d N = 157	Adipose Tissue ^e N = 200	Liver ^f N = 253
<i>F2</i>	NA	-0.27 p = 0.02	NA	0.16 p = 0.05	NA	0.17 p = 0.007
<i>F5</i>	0.03 p = 0.45	0.39 p = 0.001	0.17 p = 5.83 × 10 ⁻⁶	0.33 p = 4.87 × 10 ⁻⁵	0.31 p = 6.32 × 10 ⁻⁶	0.45 p = 3.94 × 10 ⁻¹⁴
<i>F7</i>	-0.02 p = 0.56	-0.33 p = 0.003	-0.05 p = 0.18	-0.08 p = 0.31	-0.06 p = 0.37	0.03 p = 0.61
<i>F10</i>	NA	NA	NA	-0.06 p = 0.49	-0.08 p = 0.26	0.08 p = 0.21

NA: Expressions were not available

Corresponding publication for each study: ^a [12]; ^b [13]; ^c [14]; ^d [15]; ^e [16]; ^f [17]

REFERENCES:

Throughout this application, various references describe the state of the art to which this invention pertains. The disclosures of these references are hereby incorporated by reference into the present disclosure.

1. Dahlbäck B. Novel insights into the regulation of coagulation by factor V isoforms, tissue factor pathway inhibitor α , and protein S. *J Thromb Haemost* 2017; 15: 1241–50.

2. Rietveld IM, Bos MHA, Lijfering WM, Li-Gao R, Rosendaal FR, Reitsma PH, Cannegieter SC. Factor V levels and risk of venous thrombosis: The MEGA case-control study. *Res Pract Thromb Haemost* 2018; 2: 320–6.

3. Vos HL. Inherited defects of coagulation Factor V: the thrombotic side. *J Thromb Haemost* 2006; 4: 35–40.

4. Kamphuisen PW, Rosendaal FR, Eikenboom JC, Bos R, Bertina RM. Factor V antigen levels and venous thrombosis: risk profile, interaction with factor V leiden, and relation with factor VIII antigen levels. *Arterioscler Thromb Vasc Biol* 2000; 20: 1382–6.

5. Germain M, Chasman DI, de Haan H, Tang W, Lindström S, Weng L-C, de Andrade M, de Visser MCH, Wiggins KL, Suchon P, Saut N, Smadja DM, Le Gal G, van Hylekama Vlieg A, Di Narzo A, Hao K, Nelson CP, Rocanin-Arjo A, Folkersen L, Monajemi R, et al. Meta-analysis of 65,734 individuals identifies TSPAN15 and SLC44A2 as two susceptibility
5 loci for venous thromboembolism. *Am J Hum Genet* 2015; 96: 532–42.
6. Antoni G, Oudot-Mellakh T, Dimitromanolakis A, Germain M, Cohen W, Wells P, Lathrop M, Gagnon F, Morange P-E, Tregouet D-A. Combined analysis of three genome-wide association studies on vWF and FVIII plasma levels. *BMC Med Genet* 2011; 12: 102.
7. Oudot-Mellakh T, Cohen W, Germain M, Saut N, Kallel C, Zelenika D, Lathrop M,
10 Trégouët D-A, Morange P-E. Genome wide association study for plasma levels of natural anticoagulant inhibitors and protein C anticoagulant pathway: the MARTHA project. *Br J Haematol* 2012; 157: 230–9.
8. Suchon P, Germain M, Delluc A, Smadja D, Jouven X, Gyorgy B, Saut N, Ibrahim M, Deleuze JF, Alessi MC, Morange PE, Trégouët DA. Protein S Heerlen mutation
15 heterozygosity is associated with venous thrombosis risk. *Sci Rep* 2017; 7: 45507.
9. Li Y, Willer CJ, Ding J, Scheet P, Abecasis GR. MaCH: using sequence and genotype data to estimate haplotypes and unobserved genotypes. *Genet Epidemiol* 2010; 34: 816–34.
10. Tregouet DA, Garelle V. A new JAVA interface implementation of THESIAS: testing haplotype effects in association studies. *Bioinformatics* 2007; 23: 1038–9.
- 20 11. Lunghi B, Iacoviello L, Gemmati D, Dilasio MG, Castoldi E, Pinotti M, Castaman G, Redaelli R, Mariani G, Marchetti G, Bernardi F. Detection of new polymorphic markers in the factor V gene: association with factor V levels in plasma. *Thromb Haemost* 1996; 75: 45–8.
12. Codoni V, Blum Y, Civelek M, Proust C, Franzén O, Björkegren JLM, Le Goff W,
25 Cambien F, Lusis AJ, Trégouët D-A. Preservation Analysis of Macrophage Gene Coexpression Between Human and Mouse Identifies PARK2 as a Genetically Controlled Master Regulator of Oxidative Phosphorylation in Humans. *G3 (Bethesda)* 2016; 6: 3361–71.
13. Pulcrano-Nicolas A-S, Proust C, Clarençon F, Jacquens A, Perret C, Roux M, Shotar E, Thibord F, Puybasset L, Garnier S, Degos V, Trégouët D-A. Whole-Blood miRNA
30 Sequencing Profiling for Vasospasm in Patients With Aneurysmal Subarachnoid Hemorrhage. *Stroke* 2018; 49: 2220–3.
14. Rotival M, Zeller T, Wild PS, Maouche S, Szymczak S, Schillert A, Castagné R, Deiseroth A, Proust C, Brocheton J, Godefroy T, Perret C, Germain M, Eleftheriadis M, Sinning CR, Schnabel RB, Lubos E, Lackner KJ, Rossmann H, Münzel T, et al. Integrating Genome-

Wide Genetic Variations and Monocyte Expression Data Reveals Trans-Regulated Gene Modules in Humans. *PLoS Genet* 2011; 7.

15. Erbilgin A, Civelek M, Romanoski CE, Pan C, Hagopian R, Berliner JA, Lusis AJ. Identification of CAD candidate genes in GWAS loci and their expression in vascular cells. *J Lipid Res* 2013; 54: 1894–905.

16. Stancáková A, Civelek M, Saleem NK, Soininen P, Kangas AJ, Cederberg H, Paananen J, Pihlajamäki J, Bonnycastle LL, Morken MA, Boehnke M, Pajukanta P, Lusis AJ, Collins FS, Kuusisto J, Ala-Korpela M, Laakso M. Hyperglycemia and a common variant of GCKR are associated with the levels of eight amino acids in 9,369 Finnish men. *Diabetes* 2012; 61: 1895–902.

17. Folkersen L, van't Hooft F, Chernogubova E, Agardh HE, Hansson GK, Hedin U, Liska J, Syvänen A-C, Paulsson-Berne G, Paulsson-Berne G, Franco-Cereceda A, Hamsten A, Gabrielsen A, Eriksson P, BiKE and ASAP study groups. Association of genetic risk variants with expression of proximal genes identifies novel susceptibility genes for cardiovascular disease. *Circ Cardiovasc Genet* 2010; 3: 365–73.

18. Kanai M, Akiyama M, Takahashi A, Matoba N, Momozawa Y, Ikeda M, Iwata N, Ikegawa S, Hirata M, Matsuda K, Kubo M, Okada Y, Kamatani Y. Genetic analysis of quantitative traits in the Japanese population links cell types to complex human diseases. *Nat Genet* 2018; 50: 390–400.

19. Ward LD, Kellis M. HaploReg: a resource for exploring chromatin states, conservation, and regulatory motif alterations within sets of genetically linked variants. *Nucleic Acids Res* 2012; 40: D930-934.

20. Boyle AP, Hong EL, Hariharan M, Cheng Y, Schaub MA, Kasowski M, Karczewski KJ, Park J, Hitz BC, Weng S, Cherry JM, Snyder M. Annotation of functional variation in personal genomes using RegulomeDB. *Genome Res* 2012; 22: 1790–7.

21. GTEx Consortium. Human genomics. The Genotype-Tissue Expression (GTEx) pilot analysis: multitissue gene regulation in humans. *Science* 2015; 348: 648–60.

22. Sun BB, Maranville JC, Peters JE, Stacey D, Staley JR, Blackshaw J, Burgess S, Jiang T, Paige E, Surendran P, Oliver-Williams C, Kamat MA, Prins BP, Wilcox SK, Zimmerman ES, Chi A, Bansal N, Spain SL, Wood AM, Morrell NW, et al. Genomic atlas of the human plasma proteome. *Nature* 2018; 558: 73.

23. O'Connell GC, Petrone AB, Treadway MB, Tennant CS, Lucke-Wold N, Chantler PD, Barr TL. Machine-learning approach identifies a pattern of gene expression in peripheral blood that can accurately detect ischaemic stroke. *NPJ Genom Med* 2016; 1: 16038.

CLAIMS:

1. A method of identifying a subject having or at risk of having a coagulation related disorder due to a Factor V excess or deficiency comprising determining in a nucleic acid sample obtained from the subject the presence or absence of at least one genetic variant in the PLXDC2 gene.
5
2. The method of claim 1 wherein the coagulation related disorder is a bleeding disorder or a clotting disorder.
3. The method of claim 1 for determining whether the subject is at risk of having a thromboembolic event selected from the group consisting of arterial thrombosis, fatal- and non-fatal myocardial infarction, stroke, transient ischemic attacks, cerebral venous thrombosis, peripheral arteriopathy, deep venous thrombosis and pulmonary embolism.
10
4. The method of claim 1 wherein the genetic variant is located in the promoter or is located in an intron, or is located in an exon.
5. The method of claim 1 wherein the genetic variant is present is heterozygous or homozygous.
15
6. The method of claim 1 which comprises detecting one or more single nucleotide polymorphisms (SNP).
7. The method of claim 6 which comprises detecting at least one SNP selected in the group consisting of rs965787, rs200970303, rs1887035, rs927826, rs9651367, rs2148299, rs1018635, rs1409338, rs7895470, rs10159712, rs7070003.
20
8. The method of claim 7 wherein the SNP is rs927826.
9. The method of claim 8 wherein the rs927826-G allele is associated with decreased FV plasma levels.
10. The method of claim 6 which comprises detecting any SNP in linkage disequilibrium with rs927826.
25

11. The method of claim 1 which comprises a step consisting of calculating a score, representing an estimation value of the risk that the subject has a coagulation related disorder.
- 5 12. The method of claim 11 which comprises a) detecting at least one genetic variant in the PLXDC2 gene; b) implementing an algorithm on data comprising the presence or absence of the genetic variant so as to obtain an algorithm output; c) determining the probability that the subject will have a coagulation related disorder.
13. The method of claim which comprises prescribing a preventive therapy when it is concluded that the subject has a risk of having a coagulation related disorder.
- 10 14. The method of claim 13 wherein the preventive therapy consisting in administering anticoagulants and/or antiplatelet drugs.
15. A method of treating a coagulation related disorder due to a Factor V excess in patient in need thereof comprising administering to the patient a therapeutically effective amount of an inhibitor of PLXDC2 expression.

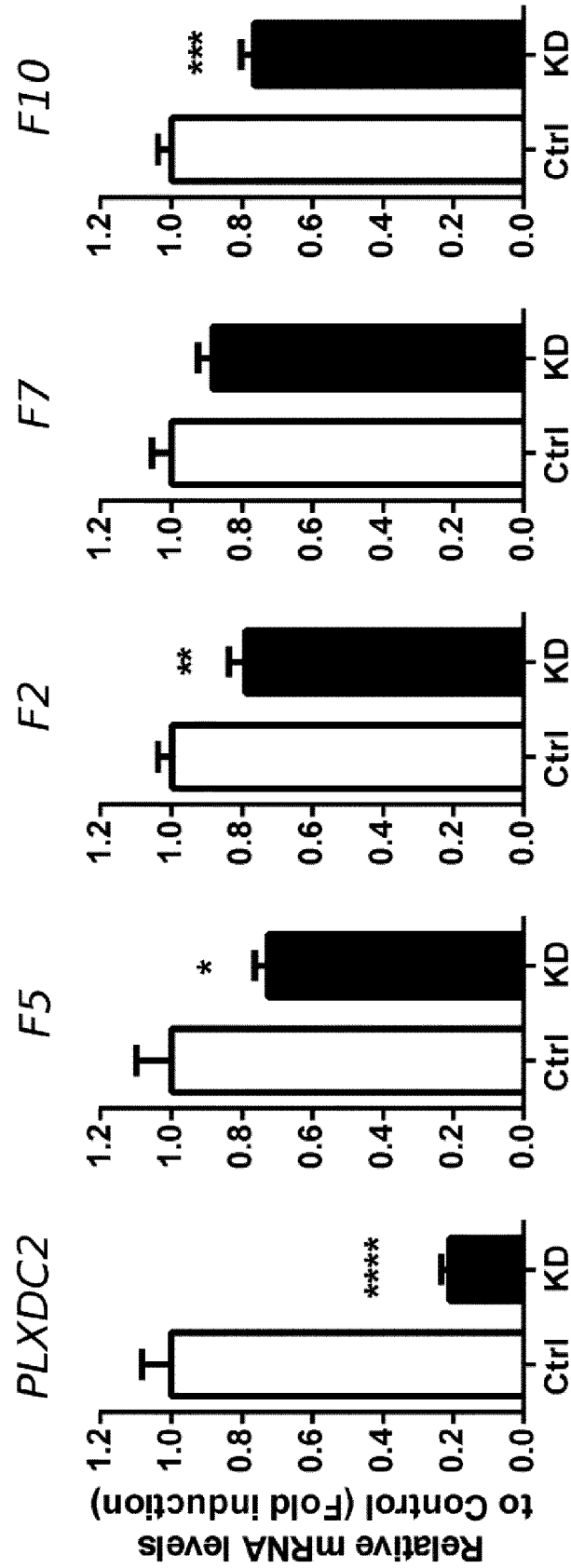


Figure 1

INTERNATIONAL SEARCH REPORT

International application No
PCT/EP2019/087038

A. CLASSIFICATION OF SUBJECT MATTER
INV. C12Q1/6883
ADD.

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
C12Q

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

EPO-Internal, WPI Data, BIOSIS, EMBASE

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	WO 2017/079551 A1 (SCRIPPS RESEARCH INST [US]) 11 May 2017 (2017-05-11) abstract; claim 1	1-15
A	WO 2009/151691 A2 (CELERA CORP [US]; BARE LANCE [US] ET AL.) 17 December 2009 (2009-12-17) abstract; claim 1	1-15
	----- -/--	



Further documents are listed in the continuation of Box C.



See patent family annex.

* Special categories of cited documents :

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier application or patent but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

"&" document member of the same patent family

Date of the actual completion of the international search

3 March 2020

Date of mailing of the international search report

11/03/2020

Name and mailing address of the ISA/

European Patent Office, P.B. 5818 Patentlaan 2
NL - 2280 HV Rijswijk
Tel. (+31-70) 340-2040,
Fax: (+31-70) 340-3016

Authorized officer

Aguilera, Miguel

INTERNATIONAL SEARCH REPORT

International application No
PCT/EP2019/087038

C(Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT		
Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	MARINE GERMAIN ET AL: "Meta-analysis of 65,734 Individuals Identifies TSPAN15 and SLC44A2 as Two Susceptibility Loci for Venous Thromboembolism", AMERICAN JOURNAL OF HUMAN GENETICS, vol. 96, no. 4, 1 April 2015 (2015-04-01), pages 532-542, XP055597411, US ISSN: 0002-9297, DOI: 10.1016/j.ajhg.2015.01.019 cited in the application abstract	1-15
A	----- KANAI MASAHIRO ET AL: "Genetic analysis of quantitative traits in the Japanese population links cell types to complex human diseases", NATURE GENETICS, NATURE PUBLISHING GROUP, NEW YORK, US, vol. 50, no. 3, 5 February 2018 (2018-02-05), pages 390-400, XP036462450, ISSN: 1061-4036, DOI: 10.1038/S41588-018-0047-6 [retrieved on 2018-02-05] cited in the application abstract	1-15
A	----- GRANT C O'CONNELL ET AL: "Machine-learning approach identifies a pattern of gene expression in peripheral blood that can accurately detect ischaemic stroke", NPJ GENOMIC MEDICINE, vol. 1, no. 1, 1 November 2016 (2016-11-01), XP055597901, DOI: 10.1038/npjgenmed.2016.38 cited in the application abstract; figure 2 -----	1-15

INTERNATIONAL SEARCH REPORT

Information on patent family members

International application No

PCT/EP2019/087038

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
WO 2017079551	A1	11-05-2017	NONE

WO 2009151691	A2	17-12-2009	CA 2717045 A1 17-12-2009
			EP 2276855 A2 26-01-2011
			JP 5409658 B2 05-02-2014
			JP 2011517935 A 23-06-2011
			JP 2014003985 A 16-01-2014
			US 2009270332 A1 29-10-2009
			US 2012264692 A1 18-10-2012
			US 2013274195 A1 17-10-2013
			US 2016281164 A1 29-09-2016
			US 2019249253 A1 15-08-2019
			WO 2009151691 A2 17-12-2009
