A genome wide search for non-additive allele effects identifies *PSKH2* as involved in the variability of Factor V activity.

Blandine Gendre[1], Angel Martinez-Perez[2,3], Marcus E Kleber[4,5], Astrid van Hylckama Vlieg[6], Anne Boland[7,8], Robert Olaso[7,8], Marine Germain[1], Gaëlle Munsch[1], Angela Patricia Moissl[4], Pierre Suchon[9], Juan Carlos Souto[,3,10], José Manuel Soria[2,3], CHARGE Hemostasis working group, Jean-François Deleuze[7,8], Winfried März[4,11,12], Frits R Rosendaal[6], Maria Sabater-Lleal[2,3,13], Pierre-Emmanuel Morange[9], David-Alexandre Trégouët[1]

[1] University of Bordeaux, INSERM, Bordeaux Population Health Research Center, UMR 1219, ELEANOR, Bordeaux, France.

[2] Unit of Genomics of Complex Diseases, Institut de Recerca Sant Pau (IR SANT PAU), Barcelona, Spain

[3] Centre for Biomedical Network Research on Rare Diseases (CIBERER), Instituto de Salud Carlos III, Madrid, Spain.

[4] Department of Medicine V, Medical Faculty Mannheim, University of Heidelberg, Mannheim, Germany.

[5] SYNLAB Center of Human Genetics Mannheim, Manheim Germany

[6] Department of Clinical Epidemiology, Leiden University Medical Center, Leiden, Netherlands.

[7] Université Paris-Saclay, CEA, Centre National de Recherche en Génomique Humaine (CNRGH), 91057, Evry, France.

[8] Laboratory of Excellence GENMED (Medical Genomics), Evry, France.

9 Cardiovascular and Nutrition Research Center (C2VN), INSERM, INRAE, Aix-Marseille University, Marseille, France

10 Thrombosis and Haemostasis Unit, Hospital de la Santa Creu i Sant Pau and Institut d'Investigació Biomèdica Sant Pau (IIB-Sant Pau), Barcelona, Spain.

11 Clinical Institute of Medical and Chemical Laboratory Diagnostics, Medical University of Graz, Graz, Austria.

12 SYNLAB Academy, SYNLAB Holding Germany, Mannheim and Augsburg, Germany.

13 Cardiovascular Medicine Unit, Department of Medicine, Karolinska Institutet, Stockholm, Sweden.

Short title (max 50 caractères incluant espaces)

PSKH2, a new molecular player of FV regulation

Total word count of the manuscript (including Title Page, Abstract, Text, References, Tables and Figures Legends) : 5523

Abstract

Background: Factor V (FV) is a key molecular player in the coagulation cascade. FV plasma levels have been associated with several human diseases, including thrombosis, bleeding and diabetic complications. So far, two genes have been

robustly found through genome wide association analyses to contribute in the inter-individual variability of plasma FV levels: structural F5 gene and PLXDC2.

Methods: We used the underestimated Brown-Forsythe methodology implemented in the Quicktest software to search for non-additive genetic effects that could contribute to the inter-individual variability of FV plasma activity. QUICKTEST was applied to 4 independent GWAS studies (LURIC, MARTHA, MEGA and RETROVE) totaling 4,505 participants of European ancestry with measured FV plasma levels. Results obtained in the 4 cohorts were meta-analyzed using a fixed-effect model. Additional analyses involved exploring haplotype and gene×gene interactions in downstream investigations.

Results: We observed a genome-wide significant signal at *PSKH2* locus, on chr8q21.3 with lead variant rs75463553 with no evidence for heterogeneity across cohorts ($p = 0.518$). Although rs75463553 did not show association with mean FV levels ($p = 0.49$), it demonstrated a robust significant ($p = 8.4 \ 10^{-9}$) association with the variance of FV plasma levels. Further analyses confirmed the reported association of *PSKH2* with neutrophil biology and revealed that rs75463553 likely interact with two loci, *GRIN2A* and *POM121L12*, known for their involvement in smoking biology.

Conclusions: This comprehensive approach identifies the role of *PSKH2* as a novel molecular player in the genetic regulation of FV, shedding light on the contribution of neutrophils to FV biology.

Introduction

Factor V (FV) is a central protein of the coagulation cascade. By acting as a co-factor for activated Factor X, FV facilitates the conversion of prothrombin to thrombin [1], the latter converting then fibrinogen into fibrin, the main component of blood clots and which also activates platelets. Mainly expressed in the liver, FV can also be stored and released by platelets [2] which provide a surface for the coagulation reactions to occur and which can contribute to amplify the coagulation process. There is natural variability in FV levels and increased/decreased FV levels have been observed in several conditions including mainly bleeding [3–6] and thrombotic disorders [7,8] but also in infections [9], inflammation [10], pregnancy [11,12], hormone contraceptives usage [13], and impaired liver dysfunction [14].

Understanding the exact sources of variability of FV levels is crucial for a better identification of individuals at higher risk of clotting disorders and for better targeting appropriate preventive and therapeutic strategies. Age, sex, smoking [15], obesity [16] and to lesser extent medication use [17], are the main environmental variables known to influence FV plasma levels. Genetic factors have also been demonstrated to contribute to the inter-individual FV variability including single nucleotide polymorphisms (SNPs) at *F5* and *PLXDC2* loci [18]. The implication of *F5* SNPs in the regulation of FV plasma levels dates back to the end of the 90s [19] where the HR2 haplotype tagged by rs6027 was identified. More recently, the *F5* rs4524 was also shown to influence plasma FV levels independently of the rs6027 [18]. The first genome wide association study (GWAS) on FV levels, based on ~1700 individuals identified the *PLXDC2* locus as a second genetic player in FV regulation [18]. Altogether, these three loci explain less than 15% of the variability in plasma FV levels, suggesting that additional molecular determinants could be involved in its

regulation. With the aim of characterizing novel genomic regulators of FV plasma levels, we here deployed a large scale agnostic genome wide search for non-additive genetic effects associated with FV plasma levels using the Brown-Forsythe (BF) methodology implemented in the Quicktest software [20]. While initially developed for detecting parental-of-origin effects (POE), this methodology can also detect loci prone to gene x gene or gene x environment interactions, making it a valuable tool to complement standard genome wide association analysis. POE is a specific kind of genomic imprinting [21,22] and several studies suggested that such epigenetic mechanisms could impact key genes of the coagulation cascade [23,24], including *PLXDC2* [25]. In this work, the BF methodology was applied to genome wide genotype data available in 4 study populations totaling 4,505 individuals with measured FV plasma levels.

Materials & Methods

This work builds on four independent study populations of unrelated individuals, all of European ancestry, that are part of the *Cohorts for Heart and Aging Research in Genomic Epidemiology* (CHARGE) Consortium [26] : the *LUdwigshafen RIsk and Cardiovascular Health* (LURIC) study [27] composed of 1,833 individuals, the *MARseille THrombosis Association* (MARTHA) study [28] composed of 1,011 patients with venous thrombosis (VT), the *Multiple Environmental and Genetic Assessment* (MEGA) study [29] of 865 VT cases and the *Riesgo de Enfermedad TROmboembolica VEnosa* (RETROVE) [30] composed of a sample of 398 VT cases and of 398 controls. All participants were phenotyped for plasma FV activity and genotyped for genome wide polymorphisms using high-throughput DNA arrays. Genotype data were further imputed using different reference panels. Detailed descriptions of the phenotype and genotype measurements are given in

Supplementary Table S1 together with additional details on imputation and genotype quality controls.

All participating studies were approved by the respective institutional Ethics Committees. Written informed consent was obtained from all participants to be included in such genetic investigations.

Statistical Analysis

The Brown-Forsythe methodology implemented in the Quicktest program

Standard GWAS are generally performed for detecting SNPs having additive allele effects on a trait of interest. The statistical modeling can thus be expressed, in case of a quantitative trait Y, as

$$E(Y|G) = \alpha + \beta \ G \quad (1)$$

with G={0,1,2} according to the number of tested alleles carried by an individual. This model assumes the absence of POE for the tested allele while POE would imply that the effect of the tested allele would depend on whether it has been inherited from the father or from the mother. In that case, a POE model could be written as:

$$E(Y|G) = \alpha + \frac{1}{2}(\beta_m + \beta_p)G \quad (2)$$

where β_m and β_p corresponds to the maternal and paternal effects, respectively, if they are identifiable. In absence of family data, these effects cannot be distinguished and therefore cannot be estimated. To solve this problem, Hoggart et al. [20]

proposed an appealing methodology to allow the detection of POE in genotype data of unrelated individuals only. By rewriting model (2) as

$$E(Y|G = 0) = \alpha$$

$$E(Y|G = 1) = \alpha + \pi\beta_m + (1 - \pi)\beta_p$$

$$E(Y|G = 2) = \alpha + \beta_m + \beta_p$$

where π is a random variable following a Bernoulli distribution with parameter ½ (50% of alleles coming from the paternal and 50% from the maternal transmission), they observed that, in presence of POE, the phenotypic variance in heterozygous individuals should be higher than the phenotypic variances in the two groups of homozygotes:

$$Var(Y|G = 0) = Var(Y|G = 2) = \sigma^2$$

$$Var(Y|G = 1) = \sigma^2 + \frac{1}{4}(\beta_m - \beta_p)^2$$

They then proposed to use the Brown-Forsythe (BF) test [31], a robust version of the Levene test, to test whether the phenotypic variance in heterozygote carriers of a given SNP is significantly higher than the phenotypic variances observed in homozygotes. They further showed that this BF test is equivalent to performing a linear model where the absolute deviation of the phenotype from the intra-genotype median is regressed on a binary variable indicating whether an individual i is heterozygote or not at the tested j SNP.

$$|y_{ij} - \tilde{y}_j| = \mu + \gamma \mathbb{1}_{(\text{Individual } i \text{ is heterozygous for genotype } j)} + \varepsilon$$

A positive and significant value for the γ regression coefficient associated to this indicator variable is a sign for POE.

They implemented this BF framework in the Quicktest program [https://wp.unil.ch/sgg/program/quicktest/] that can easily be applied to large GWAS genotype datasets to detect POE acting on a quantitative trait. Of note, as highlighted by Hoggart et al, while the presence of POE can lead to significant BF test, the inverse is not necessary true as a significant BF test can also be due to other phenomena such as haplotype effects, gene x gene or gene x environment interactions.

For the present work, all SNPs with imputation quality $r^2 > 0.5$, minor allele frequency > 0.005, and a number of heterozygous individuals greater than 20, were tested through the BF methodology in relation to FV activity. Analyses were adjusted for age, sex, and genetically derived principal components. Additional adjustment on case-control status was performed in RETROVE.

The Quicktest software was applied in each study and results were then meta-analyzed using a fixed-effect meta-analysis as implemented In the GWAMA software [32].

Heterogeneity across study populations was assessed by the Cochran's Q statistic and the I^2 index.

Genome wide statistical significance was considered at BF p-values $<5 \ 10^{-8}$.

Search for gene x gene interactions

To further investigate the possible source explaining each genome–wide significant BF signal, we sought for SNPs that could modulate FV activity differentially according to the heterozygote status at the lead SNP identified by the QuickTest analysis. For this, in each contributing study population, we conducted a genome-wide interaction analysis based on a linear model where FV activity were regressed for age, gender, genetically derived principal components, the heterozygote status at the lead BF SNP, any SNP, and an interaction term between the latter two. All SNPs with imputation quality $r^2 < 0.5$ and minor allele frequency < 0.005 were excluded from the analyses. These genome wide interaction analyses were conducted using Plink2 (www.cog-genomics.org/plink/2.0/) [33].

For each tested SNP, interaction terms were then meta-analyzed across the 4 studies using a fixed-effect meta-analysis as implemented in the GWAMA software [32].

Results

In total, 4,505 individuals were studied in this work. A brief description of the general characteristics of the four contributing studies is given in Table 1.

7,300,264 SNPs were tested in relation with FV activity through the BF framework. A Manhattan plot summarizing the statistical findings is shown in Figure 1. The associated Quantile-Quantile plot is given in Supplementary Figure S1. One locus, chr8q21.3, reached the prespecified genome wide statistical threshold of 5×10^{-8}. The lead SNP was rs75463553 and its POE γ coefficient was 0.128 ± 0.022 (p = 3.45×10^{-9}). As shown in Table 2, the POE γ coefficients were very homogeneous across

the 4 contributing studies as were the allele frequencies. We observed that the variance in FV activity was higher in carriers of the G/T genotype compared to the combined groups of G/G and T/T genotypes while no association with mean FV levels was observed (p=0.49 in the combined 4 studies).

rs75463553 maps to an intronic region of the non-coding RNA LOC105375623, downstream to *SLC7A13* and upstream to *ATP6V0D2*. It is in strong linkage disequilibrium (LD) with 6 other nearby SNPs with genome wide significant BF p-value (Supplementary Figure S2 – Supplementary Table S2). These 6 SNPs generate 3 haplotypes with frequency > 1%. None of them associated with mean FV activity (Supplementary Table S3) suggesting that the detected BF signal was unlikely due to LD effects between nearby SNPs.

By interrogating various online genomic resources, we observed that:

- rs75463553 is an expression quantitative trait locus (eQTL) for LOC105375623 in testis but also for nearby *SLC7A13* (testis) and *WWP1* (cultured fibroblasts) genes (https://gtexportal.org/home/snp/rs75463553). It is also moderately associated with the expression of more distant genes on 8q21.3 such as *RMDN1*, *CPNE3* and *NTAN1P2*.

- rs75463553 also associated (p = 1.76 x 10^{-8}) with plasma levels of RMDN1 protein in the Fenland study [34], but it was mainly a protein quantitative trait locus (pQTL) (p = 2.92 x 10^{-13}) for Carbonic anhydrase XIII whose structural gene (*CA13*), located on 8q21.2, is more than 1 Mb from rs75463553.

- While moderately associated (p = 5.9 x 10^{-7}) with whole blood DNA methylation levels at *CA13* cg0571334 CpG site in the GoDMC database (http://mqtldb.godmc.org.uk/index) [35], rs75463553 demonstrated strong

associations with whole blood DNA methylation at several CpG sites at the *PSKH2* locus (Supplementary Figure S3) such as cg00001099 (p = 5.4 x 10^{-291}), cg26186954 (p = 7.54 x 10^{-185}) and cg20982735 (p = 1.13 x 10^{-157}).

PSKH2 is located upstream *ATP6V0D2* gene and has been shown to harbor SNPs associated with neutrophil count [36] whose expression has been shown to correlate with that of *F5* in individuals with inflammatory disorders [37]. We then investigated the association of rs75463553 with neutrophil counts in LURIC, MARTHA and RETROVE. We observed a trend for neutrophil counts being higher in homozygote carriers of the rs75463553-T allele (Table 3). However, this association was mainly observed in individuals with high platelets counts. As shown in Table 3, when the combined sample was divided according to the median of platelets observed in the global population, rs75463553-TT carriers with platelets count over the median exhibited significantly (p = 0.0023) higher neutrophil count while no association (p = 0.59) was observed in the group of individuals with lower platelets. Surprisingly, we further observed that the significant association was mainly restricted to smokers (Table 4). However, no such interactive effects were observed on FV activity (data not shown) suggesting that the complex relationship between rs75463553, platelets and neutrophils count would unlikely explain the statistical BF signal observed on FV activity. Consistent with this hypothesis is the observation that the effect of rs75463553 on FV variability, as assessed by the BF methodology, still holds according to platelets and smoking (Supplementary Table S4), except in smokers with high platelet counts.

Of note, in LURIC, MARTHA and RETROVE where neutrophils and platelets count were measured, FV activity did not exhibit significant correlation with them (Supplementary Table S5).

We then further explored whether the original detected BF signal could be explained by the interaction of rs75463553 with other SNPs. The genome wide scan conducted in the four contributing studies identified one genome wide significant (p = 2.6 10^{-8}) interaction (Supplementary Table S6, Supplementary Figure S4). In heterozygous carriers of the rs75463553-T allele, carrying the rs7190785-A allele at *GRIN2A* on chromosome 16p13.2 was associated with increased FV activity (β = 0.05 $\pm$ 0.01, p = 2.52 10^{-8}). By contrast, no association was observed for the rs7190785 allele in individuals with GG or TT genotypes at rs75463553.(β = -0.01 $\pm$ 0.01, p = 0.22). This phenomenon was consistent in LURIC, MARTHA and MEGA, but not in RETROVE (Supplementary Table S7). It is worthy to note that a second interaction signal nearly reached genome wide significance (p = 6.27 x 10^{-8}) mapping to *POM121L12* locus on chr7p12.1 which, as *GRIN2A*, has been observed in several GWAS to associate with smoking phenotypes [38–40].

Discussion

This work was motivated by the search of non-additive genetic effects that could contribute to the inter individual variability of FV plasma activity. To achieve this objective, we used an underestimated methodology, but with great potential, that allows to leverage on existing GWAS data in a very efficient and quick way as it is

implemented in the easy-to-use Quicktest software [20]. Even if the method was initially proposed to detect POE effects, it also has potential to detect non-additive genetic effects that could be due to gene x gene or gene x environment interactions. Its application here provides evidence for the presence of gene x smoking interaction in the modulation of FV plasma activity.

The application of this methodology on 4,505 individuals phenotyped for FV activity and with GWAS data identified SNPs at the 8q21.3 locus significantly associated with the variability of FV activity. Using publicly available resources, we observed that the lead SNP at this locus, rs75463553, associates with DNA methylation levels of this locus, in particular with several CpG sites at the *PSKH2* gene. *PSKH2* encodes a protein serine kinase about which, little is known. Some genetic studies have linked *PSKH2* SNPs with neutrophil [36,41] and myeloid leukocyte [36] counts whose role in the coagulation and thrombotic pathways have been highly discussed in the literature [42]. In our work, we did observe an association between rs75463553 and neutrophil counts, but this association could not explain the genome-wide signal we detected. We sought to investigate whether this signal could be caused by POE in a family study but we were only able to assess this hypothesis in a sample of 21 families from the GAIT1 study [43]. Unfortunately, only 26 informative meioses were available to test for a differential paternal – maternal effect of the rs75463553 and no statistical association was observed (p>0.50) (data not shown). Of note, *PSKH2* was not detected to be prone to POE using an alternative methodology based on sequencing data and applied to several phenotypes, including neutrophil counts, from the UKbiobank [44]. This would suggest that the signal we detected using the QuickTest software could be due to other phenomena rather than POE. In line with this hypothesis are the candidate gene x gene interactions we identified with two loci,

GRIN2A and *POM121L12*, that have been proposed to be involved in smoking phenotypes in previous GWAS [38–40]. Unfortunately, smoking status at the time of blood sampling was not available in all contributing studies and it was then not possible to assess whether the observed signal was due to complex interactions between several polymorphisms and smoking. Similarly, our studies had very limited information about additional environmental covariates that prevented us from performing more exhaustive gene x environment interaction analyses and from determining whether the *PSKH2* locus statistical signal could underline such epistasis phenomena.

In conclusion, this work provides strong statistical argument supporting the role of the *PSKH2* locus in the variability of FV activity. But more in-depth investigations are now needed to characterize the exact underlying mechanisms. Besides, this work also emphasizes how to leverage on existing large GWAS datasets to detect non additive allele effects using the Brown-Forsyth methodology as implemented in the QuickTest program which could explain part of the missing heritability still existing for most complex phenotypes.

Acknowledgments

Disclosures

None

References

[1] Bouchard BA, Chapin J, Brummel-Ziedins KE, Durda P, Key NS, Tracy PB. Platelets and platelet-derived factor Va confer hemostatic competence in complete factor V deficiency. Blood 2015;125:3647–50. https://doi.org/10.1182/blood-2014-07-589580.

[2] Camire RM, Bos MHA. The molecular basis of factor V and VIII procofactor activation. J Thromb Haemost JTH 2009;7:1951–61. https://doi.org/10.1111/j.1538-7836.2009.03622.x.

[3] Lippi G, Favaloro E, Montagnana M, Manzato F, Guidi G, Franchini M. Inherited and acquired factor V deficiency. Blood Coagul Fibrinolysis Int J Haemost Thromb 2011;22. https://doi.org/10.1097/MBC.0b013e3283424883.

[4] Huang JN, Koerper MA. Factor V deficiency: a concise review. Haemoph Off J World Fed Hemoph 2008;14:1164–9. https://doi.org/10.1111/j.1365-2516.2008.01785.x.

[5] Nichols WC, Seligsohn U, Zivelin A, Terry VH, Hertel CE, Wheatley MA, et al. Mutations in the ER-Golgi intermediate compartment protein ERGIC-53 cause combined deficiency of coagulation factors V and VIII. Cell 1998;93:61–70. https://doi.org/10.1016/s0092-8674(00)81146-0.

[6] Duckers C, Simioni P, Spiezia L, Radu C, Dabrilli P, Gavasso S, et al. Residual platelet factor V ensures thrombin generation in patients with severe congenital factor V deficiency and mild bleeding symptoms. Blood 2010;115:879–86. https://doi.org/10.1182/blood-2009-08-237719.

[7] Suehisa E, Kawasaki T, Toku M, Hidaka Y. Low level of factor V is associated with development of deep-vein thrombosis in Japanese patients. Thromb Res 2010;125:128–33. https://doi.org/10.1016/j.thromres.2009.05.010.

[8] 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. https://doi.org/10.1161/01.atv.20.5.1382.

[9] Stefely JA, Christensen BB, Gogakos T, Cone Sullivan JK, Montgomery GG, Barranco JP, et al. Marked factor V activity elevation in severe COVID-19 is associated with venous thromboembolism. Am J Hematol 2020;95:1522–30. https://doi.org/10.1002/ajh.25979.

[10] Esmon CT. The interactions between inflammation and coagulation. Br J Haematol 2005;131:417–30. https://doi.org/10.1111/j.1365-2141.2005.05753.x.

[11] Brenner B. Haemostatic changes in pregnancy. Thromb Res 2004;114:409–14. https://doi.org/10.1016/j.thromres.2004.08.004.

[12] Hellgren M. Hemostasis during normal pregnancy and puerperium. Semin Thromb Hemost 2003;29:125–30. https://doi.org/10.1055/s-2003-38897.

[13] Movérare S, Skrtic S, Lindberg MK, Dahlman-Wright K, Ohlsson C. Estrogen increases coagulation factor V mRNA levels via both estrogen receptor-alpha and -beta in murine bone marrow/bone. Eur J Endocrinol 2004;151:259–63. https://doi.org/10.1530/eje.0.1510259.

[14] Lefrère JJ, Conard J, Mavier P, Bettan L, Beaugrand M, Gozin D, et al. Coagulation assays as diagnostic markers of hepatocellular carcinoma. Thromb Haemost 1988;60:468–70.

[15] Csordas A, Bernhard D. The biology behind the atherothrombotic effects of cigarette smoke. Nat Rev Cardiol 2013;10:219–30. https://doi.org/10.1038/nrcardio.2013.8.

[16] Nagai N, Hoylaerts MF, Cleuren ACA, Van Vlijmen BJM, Lijnen HR. Obesity promotes injury induced femoral artery thrombosis in mice. Thromb Res 2008;122:549–55. https://doi.org/10.1016/j.thromres.2007.12.017.

[17] Paret N, Dautriche A, Bera APJ, Gouraud A, Millaret A, Descotes J, et al. Decrease in factor V activity in patients treated with azathioprine or 6-mercaptopurine. Clin Res Hepatol Gastroenterol 2014;38:426–31. https://doi.org/10.1016/j.clinre.2013.12.003.

[18] Thibord F, Hardy L, Ibrahim-Kosta M, Saut N, Pulcrano-Nicolas A-S, Goumidi L, et al. A Genome Wide Association Study on plasma FV levels identified PLXDC2 as a new modifier of the coagulation process. J Thromb Haemost JTH 2019;17:1808–14. https://doi.org/10.1111/jth.14562.

[19] Lunghi B, Iacoviello L, Gemmati D, Dilasio MG, Castoldi E, Pinotti M, et al. Detection of new polymorphic markers in the factor V gene: association with factor V levels in plasma. Thromb Haemost 1996;75:45–8.

[20] Hoggart CJ, Venturini G, Mangino M, Gomez F, Ascari G, Zhao JH, et al. Novel approach identifies SNPs in SLC2A10 and KCNK9 with evidence for parent-of-origin effect on body mass index. PLoS Genet 2014;10:e1004508. https://doi.org/10.1371/journal.pgen.1004508.

[21] Reik W, Walter J. Genomic imprinting: parental influence on the genome. Nat Rev Genet 2001;2:21–32. https://doi.org/10.1038/35047554.

[22] Lokody I. Gene expression: Consequences of parent-of-origin effects. Nat Rev Genet 2014;15:145. https://doi.org/10.1038/nrg3677.

[23] Berger M, Mattheisen M, Kulle B, Schmidt H, Oldenburg J, Bickeböller H, et al. High factor VIII levels in venous thromboembolism show linkage to imprinted loci on chromosomes 5 and 11. Blood 2005;105:638–44. https://doi.org/10.1182/blood-2004-05-2018.

[24] Fry AE, Griffiths MJ, Auburn S, Diakite M, Forton JT, Green A, et al. Common variation in the ABO glycosyltransferase is associated with susceptibility to severe Plasmodium falciparum malaria. Hum Mol Genet 2008;17:567–76. https://doi.org/10.1093/hmg/ddm331.

[25] Ribarska T, Bastian K-M, Koch A, Schulz WA. Specific changes in the expression of imprinted genes in prostate cancer--implications for cancer progression and epigenetic regulation. Asian J Androl 2012;14:436–50. https://doi.org/10.1038/aja.2011.160.

[26] Psaty BM, O'Donnell CJ, Gudnason V, Lunetta KL, Folsom AR, Rotter JI, et al. Cohorts for Heart and Aging Research in Genomic Epidemiology (CHARGE) Consortium: Design of prospective meta-analyses of genome-wide association studies from 5 cohorts. Circ Cardiovasc Genet 2009;2:73–80. https://doi.org/10.1161/CIRCGENETICS.108.829747.

[27] Moissl AP, Delgado GE, Krämer BK, Dawczynski C, Stojakovic T, März W, et al. Alcohol consumption and mortality: The Ludwigshafen Risk and Cardiovascular Health (LURIC) study. Atherosclerosis 2021;335:119–25. https://doi.org/10.1016/j.atherosclerosis.2021.08.014.

[28] Oudot-Mellakh T, Cohen W, Germain M, Saut N, Kallel C, Zelenika D, et al. 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. https://doi.org/10.1111/j.1365-2141.2011.09025.x.

[29] Rietveld IM, Bos MHA, Lijfering WM, Li-Gao R, Rosendaal FR, Reitsma PH, et al. Factor V levels and risk of venous thrombosis: The MEGA case-control study. Res Pract Thromb Haemost 2018;2:320–6. https://doi.org/10.1002/rth2.12091.

[30] Llobet D, Vallvé C, Tirado I, Vilalta N, Carrasco M, Oliver A, et al. Platelet hyperaggregability and venous thrombosis risk: results from the RETROVE project. Blood Coagul Fibrinolysis Int J Haemost Thromb 2021;32:122–31. https://doi.org/10.1097/MBC.0000000000001006.

[31] Brown MB, Forsythe AB. Robust Tests for the Equality of Variances. J Am Stat Assoc 1974;69:364–7. https://doi.org/10.2307/2285659.

[32] Mägi R, Morris AP. GWAMA: software for genome-wide association meta-analysis. BMC Bioinformatics 2010;11:288. https://doi.org/10.1186/1471-2105-11-288.

[33] Chang CC, Chow CC, Tellier LC, Vattikuti S, Purcell SM, Lee JJ. Second-generation PLINK: rising to the challenge of larger and richer datasets. GigaScience 2015;4:7. https://doi.org/10.1186/s13742-015-0047-8.

[34] Lindsay T, Westgate K, Wijndaele K, Hollidge S, Kerrison N, Forouhi N, et al. Descriptive epidemiology of physical activity energy expenditure in UK adults (The Fenland study). Int J Behav Nutr Phys Act 2019;16:126. https://doi.org/10.1186/s12966-019-0882-6.

[35] Min JL, Hemani G, Hannon E, Dekkers KF, Castillo-Fernandez J, Luijk R, et al. Genomic and phenotypic insights from an atlas of genetic effects on DNA methylation. Nat Genet 2021;53:1311–21. https://doi.org/10.1038/s41588-021-00923-x.

[36] Chen M-H, Raffield LM, Mousas A, Sakaue S, Huffman JE, Moscati A, et al. Trans-ethnic and Ancestry-Specific Blood-Cell Genetics in 746,667 Individuals from 5 Global Populations. Cell 2020;182:1198-1213.e14. https://doi.org/10.1016/j.cell.2020.06.045.

[37] Malengier-Devlies B, Metzemaekers M, Gouwy M, Van Nieuwenhove E, Betrains A, Cockx M, et al. Phenotypical and Functional Characterization of

Neutrophils in Two Pyrin-Associated Auto-inflammatory Diseases. J Clin Immunol 2021;41:1072–84. https://doi.org/10.1007/s10875-021-01008-4.

[38] Saunders GRB, Wang X, Chen F, Jang S-K, Liu M, Wang C, et al. Genetic diversity fuels gene discovery for tobacco and alcohol use. Nature 2022;612:720–4. https://doi.org/10.1038/s41586-022-05477-4.

[39] Liu M, Jiang Y, Wedow R, Li Y, Brazel DM, Chen F, et al. Association studies of up to 1.2 million individuals yield new insights into the genetic etiology of tobacco and alcohol use. Nat Genet 2019;51:237–44. https://doi.org/10.1038/s41588-018-0307-5.

[40] Karlsson Linnér R, Biroli P, Kong E, Meddens SFW, Wedow R, Fontana MA, et al. Genome-wide association analyses of risk tolerance and risky behaviors in over 1 million individuals identify hundreds of loci and shared genetic influences. Nat Genet 2019;51:245–57. https://doi.org/10.1038/s41588-018-0309-3.

[41] Vuckovic D, Bao EL, Akbari P, Lareau CA, Mousas A, Jiang T, et al. The Polygenic and Monogenic Basis of Blood Traits and Diseases. Cell 2020;182:1214-1231.e11. https://doi.org/10.1016/j.cell.2020.08.008.

[42] Pentz MA, Brannon BR, Charlin VL, Barrett EJ, MacKinnon DP, Flay BR. The power of policy: the relationship of smoking policy to adolescent smoking. Am J Public Health 1989;79:857–62. https://doi.org/10.2105/ajph.79.7.857.

[43] Souto JC, Almasy L, Borrell M, Blanco-Vaca F, Mateo J, Soria JM, 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;67:1452–9. https://doi.org/10.1086/316903.

[44] Hofmeister RJ, Rubinacci S, Ribeiro DM, Buil A, Kutalik Z, Delaneau O. Parent-of-Origin inference for biobanks. Nat Commun 2022;13:6668. https://doi.org/10.1038/s41467-022-34383-6.

Tables

Table 1 : Brief description of the studied populations

	LURIC	MARTHA	MEGA	RETROVE
N	1,833	1,011	865	796
% of VT cases	0%	100%	100%	50%
Age, years (SD)	62.3 (10.8)	47.6 (15.7)	47.6 (12.9)	54.4 (19.9)
Female sex, N (%)	537 (29%)	633 (63%)	444 (51%)	406 (51%)
Factor V (U/dL)	1.13 (0.22)	1.07 (0.23)	0.95 (0.19)	0.99 (0.20)
Platelets count	231.1 (66.7)	256.4 (68.9)	NA	235.3 (62.4)
Neutrophil counts*	59.6 (9.46)	61.1 (8.91)	NA	58.1 (9.7)
Smoking at sampling (%)	1,163 (65%)	167 (18%)	NA	129 (16%)

*percentage compared to total white blood cells. In LURIC, smoking status includes both active and former smokers.

Table 2 : Association of rs75463553 with phenotypic mean and variance of FV activity

		LURIC	MARTHA	MEGA	RETROVE
N		1,833	1,011	865	796
Minor Allele frequency (G/T)		0.145	0.117	0.156	0.106
Imputation r^2		0.980	0.960	0.975	0.982
N					
	GG	1338	779	614	635
	GT	463	222	235	153
	TT	32	10	16	8
FV activity means					
	GG	1.132	1.056	0.958	0.990
	GT	1.142	1.097	0.946	0.998
	TT	1.059	1.082	0.912	1.119
FV activity standard deviation					
	GG	0.212	0.224	0.176	0.192
	GT	0.228	0.264	0.217	0.220
	TT	0.185	0.261	0.157	0.292
POE γ effect		0.098±0.033	0.181±0.049	0.152±0.050	0.115±0.055
P		1.693×10^{-3}	1.042×10^{-4}	1.307×10^{-3}	1.913×10^{-2}

EAF: Estimated allele frequency

The POE γ effects did not show any evidence for heterogeneity across cohorts (I^2=0, P = 0.518)

Table 3 : Association of rs75463553 with neutrophil counts in LURIC, MARTHA and RETROVE studies according platelets count.

	All population			Platelets<=230			Platelets>230		
	LURIC	MARTHA	RETROVE	LURIC	MARTHA	RETROVE	LURIC	MARTHA	RETROVE
GG	4.21 (1.54) N = 1312	3.95 (1.43) N = 675	3.93 (1.54) N = 635	3.89 (1.38) N = 728	3.62 (1.24) N = 239	3.67 (1.39) N = 318	4.60 (1.63) N = 584	4.13 (1.49) N = 436	4.19 (1.64) N = 317
GT	4.17 (1.57) N = 455	4.08 (1.59) N = 186	3.93 (1.64) N = 153	3.92 (1.52) N = 255	3.59 (1.52) N = 72	3.54 (1.33) N = 73	4.49 (1.58) N = 200	4.38 (1.56) N = 114	4.29 (1.82) N = 80
TT	4.43 (2.11) N = 33	4.87 (2.79) N = 10	3.75 (1.51) N = 8	3.71 (0.94) N = 21	4.02 (0.94) N = 3	3.19 (0.37) N = 6	5.68 (2.94) N =12	5.24 (3.30) N = 7	5.43 (2.79) N = 2
$\beta \pm SE$ P*	$\beta = 0.26 \pm 0.27$ P = 0.351	$\beta = 0.92 \pm 0.47$ P = 0.053	$\beta = -0.16 \pm 0.35$ P = 0.654	$\beta = -0.18 \pm 0.32$ P = 0.562	$\beta = 0.42 \pm 0.76$ P = 0.581	$\beta = -0.24 \pm 0.35$ P = 0.497	$\beta = 1.09 \pm 0.48$ p = 0.023	$\beta = 1.09 \pm 0.59$ P= 0.063	$\beta = 0.53 \pm 0.76$ P = 0.480
Combined β	$\beta = +\ 0.31 \pm 0.22$ P = 0.146			$\beta = -0.137 \pm 0.25$ P = 0.59			$\beta = +\ 1.08 \pm 0.35$ P = 0.0023		

* Association was tested using a linear model adjusted for age and sex under the assumption of recessive effect β

Table 4 : Association of rs75463553 with neutrophil counts according to platelets counts and smoking in LURIC, MARTHA and RETROVE studies combined.

	Platelets<=230		Platelets>230	
	Non Smokers	Smokers	Non Smokers	Smokers
rs75463553				
GG/GT	3.59 (1.30) N = 945	4.04 (1.47) N = 741	4.1 (1.52) N = 1044	4.77 (1.65) N = 688
TT	3.49 (0.83) N = 16	3.81 (0.92) N = 14	4.19 (1.5) N = 11	6.97 (3.44) N = 10
β ± SE P*	β = - 0.032 ± 0.323 P = 0.921	β = - 0.227± 0.394 P = 0.564	β = + 0.084± 0.457 P = 0.853	β = + 2.17 ± 0.538 P = 6.07 10^{-5}

* Association was tested using a linear model adjusted for age, sex and cohort under the assumption of recessive TT effect
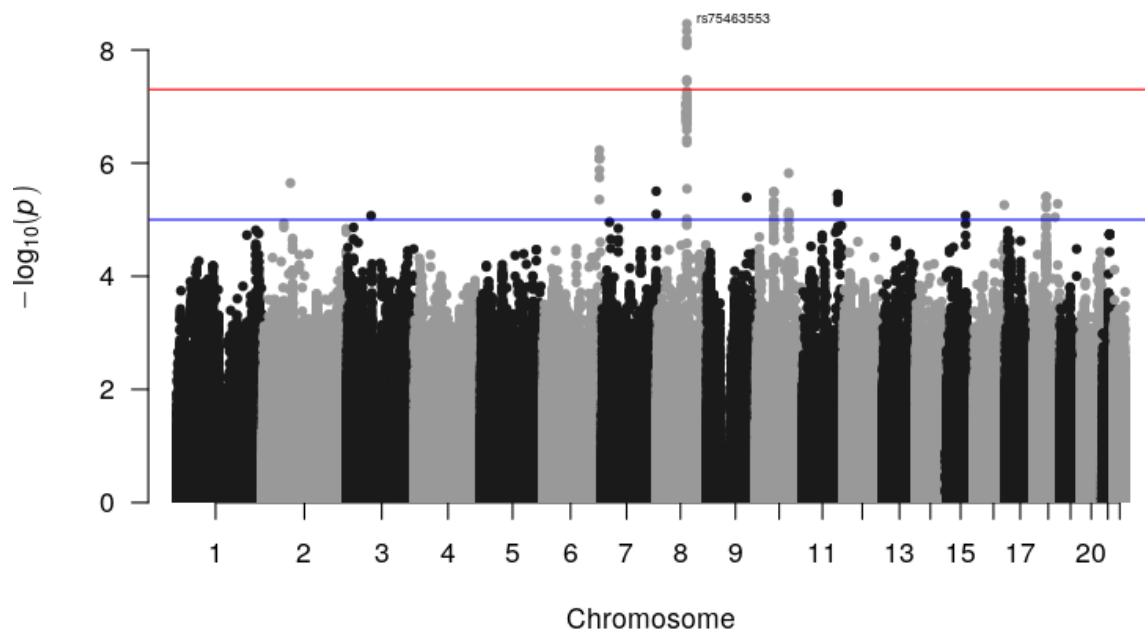
Figure

Figure 1 : Manhattan plot of 4 studies of unrelated Europeans individuals for detecting POE in FV plasma level (n=4,505).

Supplementary Material

Tables S1-S7

Figures S1-S4