Supporting Information

Novel exomic rare variants associated with venous thrombosis

Hiroshi Deguchi[a], Meenal Shukla[a], Mohammed Hayat[a], Ali Torkamani[a, b], Darlene J. Elias[a,c], and John H. Griffin[a,d1]

[a]Department of Molecular Medicine, The Scripps Research Institute, La Jolla, CA, 92037 USA.

[b]Scripps Research Translational Institute and Department of Integrative Structural and Computational Biology, The Scripps Research Institute, La Jolla, CA, 92037 USA

[c]Scripps Clinic, La Jolla, California, 92037 USA.

[d]Division of Hematology, Department of Medicine, University of California, San Diego, CA 92093 USA

Material and Design

Patients.

The Scripps Venous Thrombosis Registry (**Table 1S**) is a case-control study of risk factors for VTE as previously described (Deguchi*, et al* 2017). Inclusion criteria for this study included age at first thrombosis < 55 years old, > 3 months since diagnosis of acute thrombosis, a life expectancy of at least three years and no lipid lowering medications or cancer. A total of 154 VTE patients (73 male and 81 female) were recruited. 130 VTE patients were age and sex matched with controls. 24 VTE patients (15 male and 9 female) were not control matched because their age was not represented in the normal blood donation service. 17 of the 130 matched VTE patients (10 male and 7 female) were excluded from analysis due to the post enrollment identification of statin use. This resulted in a total of 113 VTE patients and 113 age and sex matched controls. Further, an additional 113 age and sex matched controls who were recruited later to achieve a 1:2 VTE to control ratio for statistical power in other investigations. For this study, only Caucasian samples were analyzed and the exomic genotyping VTE cohort comprised 104 VTE cases and 211 controls. Age matched (± 2 years old) healthy controls were recruited through the Scripps General Clinical Research Center's (GCRC) blood donation program. Blood was collected in the GCRC at least three months after VTE diagnosis and after 12 hr fasting. Serum, EDTA-plasma and DNA were prepared, and they were stored at -70ºC. Participants in the blood donation program had normal CBC and negative HIV, hepatitis B and C testing. Some controls were from the community but most were employees or former employees of Scripps. Clinical data collection included detailed medical history and the presence of risk factors for venous thrombosis. The protocol was approved by the Institutional Review Board of Scripps Clinic and subjects provided written informed consent. African American ancestry was excluded from the exomic analysis due to the potential difference in genetic risk factors among races.

Exomic array analysis.

The study population consisted of 104 VTE cases and 211 controls from the Scripps Venous Thrombosis Registry (**Supplemental Materials, Table 1S**). The quality of each DNA stored in the freezer was re-evaluated according to the manufacture's protocol for performing Affymetrix Axiom® Exome Genotyping Array analysis without any errors. Subjects were genotyped at DNA Core Laboratory of the Scripps Research Institute using the Axiom® Exome Genotyping Array (Affymetrix) containing ~ 300,000 exomic markers. Genotyping included three sample replicates and three Affymetrix controls in each array plate for quality control. In addition, case and control DNA samples were randomly assigned across the 96-well plates, ensuring approximately equal ratio of case and control DNA samples (1:2) by each stratum to avoid potential plate and chip effects, respectively. The genotype for each SNP for each subject was determined using the Affymetrix Genotyping Console (version 4.1.4). Overall, 99.4% of assayed samples passed quality-control standards. Both the mean SNP call rates and the mean sample call rates were 99.7%.

Post-genotyping SNP filtering was performed using SNPolisher (version 1.3.6.6) to filter out SNPs with the potential to produce false positive association due to genotyping mis-clustering and/or unexpected variation about the SNP sites in the genomes in the study set. The SNP filtering included unacceptably high rates of missing genotype calls (>5%) and Fisher's Linear Discriminant for SNP cluster (≤ 3.6). SNP heterozygous strength offset excluded 7,217 SNPs (2.2%). A total of 3,900 SNPs had unacceptably high rates of deviation from Hardy-Weinberg equilibrium (HWE, $P<10^{-5}$) to leave 308,166 SNPs for the association analysis. The cluster pattern of all candidate SNPs associated with targets were visually analyzed to ensure the assigned clusters were acceptable, and the genotyping was corrected if required.

Tail Clip Procedure

All animal protocols were approved by the Institutional Animal Care and Use Committee of The Scripps Research Institute. C57BL/6J mice acquired from the Rodent Breeding Colony at the Scripps Research Institute were anesthetized with isoflurane. Intravenous injection of anti-FVIII antibody (GMA-8015, 0.25 mg/kg) (Green Mountain Antibodies, Burlington, VT) or control vehicle was given retro-orbitally to wild type 8-12 weeks old C57BL/6J mice at 2 hours prior to tail cutting (Wyseure*, et al* 2018). Rabbit skeletal muscle myosin (Cytoskeletal Inc or Sigma) (5.4 mg/kg, estimated to give ≤ 200 nmol/L final plasma concentration) was given i.v. at 15 minutes prior to tail cutting. Mice were anaesthetized with isoflurane and placed on a temperature-controlled flow heating pad (37 °C), and the distal portion of the tail was surgically removed via a scalpel blade at 1.5-mm tail diameter to induce a moderate bleeding effect (Wyseure*, et al* 2018). Tails were immersed in 50 mL of saline, and blood was collected for 10 minutes. Body weight-corrected blood loss was determined as described (von Drygalski*, et al* 2014). A final volume of 150 µL from the saline vials was mixed with equal volume of 2% acetic acid solution to induce erythrocyte lysis for measurement of hemoglobin concentration. Then 150 µL for each hemoglobin sample was placed in a PolySorp 96-well plate and the absorbance of 490 nm was measured. Data were obtained using a VERSAmax Microplate Reader (Molecular Devices Corporation, Menlo Park, CA). Total blood loss was expressed in microliter per gram of mouse and was obtained using a standard derived from defined blood volumes lysed and processed similarly as tail bleeding volumes above.

Plasma Glucosyl Ceramide analysis

Plasma glucosyl ceramide (GlcCer) levels were measured with minor modification as described (Deguchi*, et al* 2001, Deguchi*, et al* 2017). Briefly, plasma lipids extracted from citrated plasma with chloroform/methanol (2:1, v/v) were analyzed using a high performance liquid chromatography (HPLC) system (Waters Corp., Milford, MA) coupled to a Sedex-55 evaporative light scattering detector system (SEDERE, Alfortville Cedex, France). A µPorosil column (300 mm x 3.9 mm) was used with isocratic chloroform/methanol/water (88:11:1).

Statistical analysis.

The primary outcome was VTE status, a binary measure, and we tested for an association between each SNP and VTE using PLINK v 1.07 (association test adjustment for multiple testing using Benjamini and Hochberg step up false discovery rate (FDR) control (FDR-BH)). For variant collapsing burden analyses, a meta-SNP was defined as having an alternative allele defined by the presence of an alternative allele at any SNP within the variant set. This meta-SNP was then tested for association with VTE or recurrent VTE by using a simple allele counting approach using PLINK association test. Mann Whitney test was performed using Prism™ 7.04 software (Graph Pad Software Inc., San Diego, CA). Plasma GlcCer values for both control and VTE were classified into binary measure and the association of SNPs were tested using PLINK considering low GlcCer samples (<10%-ile of control) as cases and high GlcCer samples as controls.

Table S1. Study Population Characteristics for the Scripps Venous Thrombosis Registry that was used for rare variant genotyping for association of SNPs with VTE.

Variables	*Controls*	*VTE*	*p-value*
	N=211	***N=104***	
Female, n (%)	116 (55)	61 (59)	0.55
Age, yr (SD)	44.3 (9.7)	45.0 (10.0)	0.40
Age at 1st VTE, yr (SD)	-	40.3 (10.2)	
female		38.4 (10.4)	
male		42.7 (9.6)	0.035*
Recurrent VTE, n (%)	-	32 (30.8)	
Female, n (%)	-	18 (56.3)	
Age at 1st VTE, yr (SD)	-	37.6 (11.1)	
female	-	35.1 (11.1)	
male	-	41.0 (10.5)	0.14*
	N=104	***N=104***	
GlcCer (IQR), µg/mL**	5.8 (5.2–6.8)	5.0 (4.4–5.9)	p<0.0001

* The differences in age at first VTE or in percentage for recurrent VTE between male and female groups were compared by unpaired t-test.

** Subgroup of published data (Deguchi*, et al* 2017)

Table S2. Twenty eight variant SNPs with FDR-adjusted p < 0.05 were associated with recurrent venous thrombosis based on exomic genotyping of subjects in the Scripps Venous Thrombosis Registry. For this exomic study, 315 subjects of the Scripps VTE Registry were genotyped using the Axiom® Exome Genotyping Array. The allele frequencies were compared for 32 recurrent VTE cases with 211 controls. The unadjusted and FDR adjusted p values were calculated using the PLINK test. Population frequency was obtained from gnomAD database (https://gnomad.broadinstitute.org/).

chr	Gene	SNP	MAF		p value	p value	Population frequency	
			VTE	control	unadjusted	FDR-adjusted	Allele number	MAF
5	***FAM134B***	**rs78314670**	**0.078**	**0.002**	**3.37×10^{-7}**	**7.20×10^{-3}**	**128630**	**0.0009**
1	*F5*	rs6687813	0.219	0.050	1.20×10^{-6}	1.83×10^{-2}	15416	0.065
1	*F5*	rs6025	0.141	0.021	2.68×10^{-6}	3.23×10^{-2}	129056	0.026
2	*CCDC93*	rs17512204	0.403	0.156	3.54×10^{-6}	3.23×10^{-2}	24632	0.072
5	*PPARGC1*	rs7732671	0.219	0.055	4.24×10^{-6}	3.23×10^{-2}	112764	0.083
5	*PPARGC1*	rs17572019	0.219	0.055	4.24×10^{-6}	3.23×10^{-2}	112760	0.083
12	*E2F7*	rs310831	0.281	0.088	5.75×10^{-6}	3.23×10^{-2}	128880	0.116
3	*TGM4*	rs1395388	0.188	0.043	7.81×10^{-6}	3.23×10^{-2}	128720	0.063
17	*BZRAP1*	rs9913145	0.313	0.108	7.90×10^{-6}	3.23×10^{-2}	125884	0.116
17	***MYH8***	**rs111567318**	**0.047**	**0**	**8.55×10^{-6}**	**3.23×10^{-2}**	**129190**	**0.0076**
1	*DDX20*	rs41310098	0.047	0	8.55×10^{-6}	3.23×10^{-2}	128720	0.063
4	*PABPC4L*	rs6830036	0.047	0	8.55×10^{-6}	3.23×10^{-2}	75108	0.0078
5	*GEMIN5*	rs115551140	0.047	0	8.55×10^{-6}	3.23×10^{-2}	129064	0.010
6	*HIST1H1C*	rs61748580	0.047	0	8.55×10^{-6}	3.23×10^{-2}	129190	0.0087
6	*HIST1H1C*	rs41266787	0.047	0	8.55×10^{-6}	3.23×10^{-2}	129168	0.00002
13	*KIAA0226*	rs139968164	0.047	0	8.55×10^{-6}	3.23×10^{-2}	127792	0.0025
16	*SRRM2*	rs117133016	0.047	0	8.55×10^{-6}	3.23×10^{-2}	129066	0.011
16	*WDR59*	rs16948255	0.047	0	8.55×10^{-6}	3.23×10^{-2}	15394	0.00071
18	*GRP*	rs55796466	0.047	0	8.55×10^{-6}	3.23×10^{-2}	129126	0.0021
13	*MIPEP*	rs7327620	0.047	0	8.98×10^{-6}	3.23×10^{-2}	15424	0.0029
4	*ZNF718*	rs144152312	0.063	0.002	9.38×10^{-6}	3.23×10^{-2}	128206	0.014
9	*HRCT1*	rs141107455	0.063	0.002	9.38×10^{-6}	3.23×10^{-2}	125640	0.0059
11	*ROBO4*	rs138481093	0.063	0.002	9.38×10^{-6}	3.23×10^{-2}	128520	0.0078
12	*KRT74*	rs139723680	0.063	0.002	9.38×10^{-6}	3.23×10^{-2}	129160	0.0054
16	*PDZD9*	rs140219446	0.063	0.002	9.38×10^{-6}	3.23×10^{-2}	129166	0.012
10	*NKX6-2*	rs2804003	0.156	0.031	1.14×10^{-5}	3.70×10^{-2}	127034	0.059
13	*KATNAL1*	rs143925753	0.125	0.019	1.14×10^{-5}	3.70×10^{-2}	110126	0.036
23	*RGAG1*	rs2073787	0.660	0.337	1.22×10^{-5}	3.82×10^{-2}	92080	0.407

Table S3.

Literature search for the potential association of candidate gene or gene's protein product with thrombosis including blood coagulation/fibrinolysis, lipid metabolism

Name - Gene or Protein	Reference	Association(s)
FAM134B	(Yuan, et al 2014)	Protein's enzyme activity linked to acetyl-CoA carboxylase and fatty acid synthase
FAM134B	(Kong, *et al* 2011)	FAM134B SNP associated with Vascular dementia
Skeletal muscle myosin (SkM)	(Machovich, *et al* 1997)	Myosin binding to plasminogen and tissue Plasminogen Activator
SkM	(Erlacher, *et al* 2001) (Chung, *et al* 2014)	● SkM is elevated in polymyositis or dermatomyositis patients ● Polymyositis is associated with high VTE risk
Cardiac myosin (high homology with SkM)	(Kolev, *et al* 2003)	• Cardiac myosin stabilizes fibrin • Myosin was found in arterial thrombi
Actin (myosin binding protein)	(Wang, *et al* 1990) (Furmaniak-Kazmierczak, *et al* 1995)	Binds to factors V and Va

Table S4. Evaluation of the 356 variant SNPs in chromosome 17p.13.1 *MYH* genes (*MYH1*, *MYH2*, *MYH4*, *MYH8*, and *MYH13)* for association with recurrent VTE. Eleven of 356 SNPs in the MYH chromosome 17 gene cluster showed a significant difference between recurrent VTE cases and controls (unadjusted p < 0.05). FDR-adjustment was for comparisons of 32 recurrent VTE cases to 211 controls for 356 MYH SNPs representing all *MYH* SNPs that were interrogated using the Axiom Exome array. Bold font indicates FDR-adjusted p value < 0.05. A prediction of functional effect was made using PolyPhen-2 (http://genetics.bwh.harvard.edu/pph2/) (Adzhubei*, et al* 2013) which predicts the possible impact of an amino acid substitution on the structure and function of a human protein. The score (HumDiv) represents the probability that a substitution is damaging. Variants with scores 0.0 to 0.15 or 0.15-1.0 are predicted to be benign or possibly damaging, respectively. Variants with scores 0.85-1.0 (bold) are more confidently predicted to be likely damaging. Variants with MAF < 0.05 for controls were used to construct sets for further analyses (See **Table 1**). Population frequency was obtained from the gnomAD database (https://gnomad.broadinstitute.org/).

Gene	SNP	variant	prediction of functional effect: PolyPhen-2 score	MAF		p-value		Population frequency	
				recurrent VTE	control	unadjusted	FDR-adjusted	Allele number	MAF
MYH8	**rs111567318**	**Glu1838Ala**	**0.986**	**0.047**	**0**	**8.55×10^{-6}**	**8.55×10^{-4}**	**129190**	**0.0076**
MYH2	**rs11658164**	**Ser1043Ala**	0.376	**0.031**	**0**	**2.83×10^{-4}**	**1.42×10^{-2}**	**129080**	**0.0019**
MYH8	rs34693726	Ala636Val	0.000	0.125	0.045	9.60×10^{-3}	1.29×10^{-1}	129032	0.055
MYH1	rs147039410	Arg1321Ser	**0.909**	0.016	0	1.03×10^{-2}	1.29×10^{-1}	129178	0.00046
MYH2	rs143341678	Arg1295His	0.034	0.016	0	1.03×10^{-2}	1.29×10^{-1}	127726	0.0012
MYH1	rs149329893	Arg1283Cys	**1.000**	0.016	0	1.03×10^{-2}	1.29×10^{-1}	129176	0.00067
MYH1	rs142884848	Arg1386Cys	**0.941**	0.016	0	1.03×10^{-2}	1.29×10^{-1}	128890	0.00053
MYH8	rs142606252	Gln318His	0.103	0.016	0	1.03×10^{-2}	1.29×10^{-1}	129160	0.0013
MYH4	rs12949680	Ala594Thr	0.000	0.125	0.048	1.35×10^{-2}	1.50×10^{-1}	129176	0.054
MYH13	rs2074876	Asp1076Glu	0.001	0.016	0.093	3.63×10^{-2}	3.30×10^{-1}	126416	0.10

Reference for Supporting Information

Adzhubei, I., Jordan, D.M. & Sunyaev, S.R. (2013) Predicting functional effect of human missense mutations using PolyPhen-2. *Curr Protoc Hum Genet,* **Chapter 7,** Unit7.20.

Chung, W.S., Lin, C.L., Sung, F.C., Lu, C.C. & Kao, C.H. (2014) Increased risk of venous thromboembolism in patients with dermatomyositis/polymyositis: a nationwide cohort study. *Thromb Res,* **134,** 622-626.

Deguchi, H., Fernandez, J.A., Pabinger, I., Heit, J.A. & Griffin, J.H. (2001) Plasma glucosylceramide deficiency as potential risk factor for venous thrombosis and modulator of anticoagulant protein C pathway. *Blood,* **97,** 1907-1914.

Deguchi, H., Navarro, S., Payne, A.B., Elias, D.J., Dowling, N.F., Austin, H.D., Espana, F., Medina, P., Hooper, W.C. & Griffin, J.H. (2017) Low level of the plasma sphingolipid, glucosylceramide, is associated with thrombotic diseases. *Res Pract Thromb Haemost,* **1,** 33-40.

Erlacher, P., Lercher, A., Falkensammer, J., Nassonov, E.L., Samsonov, M.I., Shtutman, V.Z., Puschendorf, B. & Mair, J. (2001) Cardiac troponin and beta-type myosin heavy chain concentrations in patients with polymyositis or dermatomyositis. *Clin Chim Acta,* **306,** 27-33.

Furmaniak-Kazmierczak, E., Nesheim, M.E. & Cote, G.P. (1995) Coagulation factor Va is an actin filament binding and cross-linking protein. *Biochem Cell Biol,* **73,** 105-112.

Kolev, K., Tenekedjiev, K., Ajtai, K., Kovalszky, I., Gombas, J., Varadi, B. & Machovich, R. (2003) Myosin: a noncovalent stabilizer of fibrin in the process of clot dissolution. *Blood,* **101,** 4380-4386.

Kong, M., Kim, Y. & Lee, C. (2011) A strong synergistic epistasis between FAM134B and TNFRSF19 on the susceptibility to vascular dementia. *Psychiatr Genet,* **21,** 37-41.

Machovich, R., Ajtai, K., Kolev, K. & Owen, W.G. (1997) Myosin as cofactor and substrate in fibrinolysis. *FEBS Lett,* **407,** 93-96.

von Drygalski, A., Cramer, T.J., Bhat, V., Griffin, J.H., Gale, A.J. & Mosnier, L.O. (2014) Improved hemostasis in hemophilia mice by means of an engineered factor Va mutant. *J Thromb Haemost,* **12,** 363-372.

Wang, D.L., Annamalai, A.E., Ghosh, S., Gewirtz, A.M. & Colman, R.W. (1990) Human platelet factor V is crosslinked to actin by FXIIIa during platelet activation by thrombin. *Thromb Res,* **57,** 39-57.

Wyseure, T., Cooke, E.J., Declerck, P.J., Behrendt, N., Meijers, J.C.M., von Drygalski, A. & Mosnier, L.O. (2018) Defective TAFI activation in hemophilia A mice is a major contributor to joint bleeding. *Blood,* **132,** 1593-1603.