HHS Public Access

Author manuscript

Circ Cardiovasc Genet. Author manuscript; available in PMC 2018 April 01.

Published in final edited form as:

Circ Cardiovasc Genet. 2017 April ; 10(2): . doi:10.1161/CIRCGENETICS.116.001643.

Genetic Analysis of Venous Thromboembolism in UK Biobank Identifies the ZFPM2 Locus and Implicates Obesity as a Causal Risk Factor

Derek Klarin, MD[1,2,3], **Connor A. Emdin, DPhil**[1,2], **Pradeep Natarajan, MD, MMSc**[1,2], **Mark F. Conrad, MD**[4], **Sekar Kathiresan, MD**[1,2], and **the INVENT Consortium**

[1]Center for Human Genetic Research, Cardiovascular Research Center, Massachusetts General Hospital, Harvard Medical School, Boston

[2]Program in Medical and Population Genetics, Broad Institute, Cambridge

[3]Department of Surgery, Massachusetts General Hospital, Boston MA

[4]Division of Vascular and Endovascular Surgery, Massachusetts General Hospital, Boston MA

Abstract

Background—UK Biobank is the world's largest repository for phenotypic and genotypic information for individuals of European ancestry. Here, we leverage UK Biobank to understand the inherited basis for venous thromboembolism (VTE), a leading cause of cardiovascular mortality.

Methods and Results—We identified 3290 VTE cases and 116,868 controls through billing-code based phenotyping. We performed a genome-wide association study (GWAS) for VTE with ~9,000,000 imputed SNPs. We performed a phenome-wide association study (PheWAS) for a genetic risk score (GRS) of 10 VTE associated variants. To assess if obesity is a causal factor for VTE, we performed Mendelian randomization analysis using a GRS instrument composed of 68 body mass index (BMI)-associated variants. The GWAS for VTE replicated previous findings at the *F5*, *F2*, *ABO*, *F11*, and *FGG* loci. We identified one new locus - *ZFPM2* rs4602861 - at genome-wide significance [OR = 1.11 (95%CI: 1.07–1.15, P = 4.9×10^{-10})] and a new independent variant at the *F2* locus [rs3136516, OR = 1.10 (95%CI: 1.06–1.13, P = 7.60×10^{-9})]. In a PheWAS, a 10 SNP VTE GRS was associated with coronary artery disease [OR = 1.08 (95%CI: 1.05–1.10 per unit increase in VTE odds, P = 1.08×10^{-9}]. In a Mendelian randomization analysis, genetically-elevated BMI (a one standard deviation increase) was associated with 57% higher risk of VTE [OR = 1.57 (95%CI: 1.08–1.97, P = 0.003)].

Correspondence: Sekar Kathiresan, MD, Program in Medical and Population Genetics, Broad Institute, Center for Human Genetics Research, Massachusetts General Hospital, 185 Cambridge Street, CPZN 5.252, Boston, MA 02114, Tel: 617 643 6120, Fax: 8779915996, skathiresan1@mgh.harvard.edu.

Conflict of Interest Disclosures: Dr. Kathiresan reports grant support from Regeneron and Bayer, personal fees from Regeneron Genetics Center, Merck, Celera, Novartis, Bristol-Myers Squibb, Sanofi, AstraZeneca, Alnylam, Eli Lilly, Noble Insights, Ionis, Genomics PLC and Leerink Partners, personal fees and other support from Catabasis, and other support from San Therapeutics outside the submitted work.

Conclusions—For common diseases like VTE, biobanks provide potential to perform genetic discovery, explore the phenotypic consequences for disease-associated variants, and test causal inference.

Keywords

venous thromboembolism; genetics; epidemiology; single nucleotide polymorphism

Journal Subject Terms

Genetic; Association Studies; Genetics; Thrombosis

Introduction

Venous thromboembolism (VTE) is a leading cause of cardiovascular-related mortality and a disorder with significant genetic predisposition[1, 2]. VTE is a common complication of hospitalization, and millions of dollars are spent yearly on VTE prevention and treatment[3–5]. Whereas environmental risk factors contributing to the development of VTE were first identified over a century ago[6], the genetic determinants of the disease remain under active investigation.

The genome-wide association study (GWAS) approach is a validated method to identify genetic variants associated with common diseases. Prior genetic studies have identified multiple loci for VTE including the *F5* Leiden[7] (rs6025) and prothrombin G20210A[8] (rs1799963) mutations[9]. Identification of additional genetic loci may aid our understanding of the pathophysiology underlying VTE, help identify high-risk individuals, and pinpoint new therapeutic targets.

Large-scale biobanks bringing together phenotypes from the electronic health record (EHR) and genetic data are being developed all around the world[10]. With more than 500,000 participants enrolled, UK Biobank is the world's largest single repository for phenotypic and genotypic information in individuals of European ancestry[11]. However, the reliability of EHR-based phenotypes has been challenged[12]. Here, we leverage the UK Biobank resource to: 1) estimate the heritability for VTE; 2) conduct a GWAS for VTE; 3) perform a phenome-wide association study (PheWAS) for a VTE genetic risk score; and 4) test if obesity, an independent epidemiologic risk factor for VTE[13], is causally related to risk for VTE.

Methods

Participants

Over a 5-year period beginning in 2006, 503,325 individuals aged 45 to 69 years old were recruited from across the United Kingdom for participation in UK Biobank. At the time of recruitment, a trained healthcare provider conducted verbal interviews to gather baseline characteristics and medical history. In addition, UK Biobank is integrating each participant's EHR into their database, including inpatient International Classification of Diseases (ICD-10) diagnosis codes and Office of Population and Censuses and Surveys (OPCS-4)

procedure codes. Informed consent was obtained for all participants, and UK Biobank received ethical approval from the Research Ethics Committee (reference number 11/NW/0382). Our study was approved by a local Institutional Review Board at Partners Healthcare (protocol 2013P001840).

VTE Phenotype

Individuals were defined as a VTE case based on at least one of the following criteria:

1. VTE ascertained at baseline by self-report, followed by a verbal interview with a trained nurse to confirm diagnosis;
2. Hospitalization for ICD-10 Code I80.1 or I80.2 - phlebitis and thrombophlebitis of the femoral vein or other deep vessels of lower extremities;
3. Hospitalization for ICD-10 Code I82.2 - embolism and thrombosis of vena cava;
4. Hospitalization for ICD-10 Code I26.0 or I26.9 - pulmonary embolism with or without acute cor pulmonale;
5. Hospitalization for OPCS-4 Procedures Codes L79.1 or L90.2 - open venous thrombectomy of lower limb vein or inferior vena cava filter insertion.

The British Society of Interventional Radiology inferior vena cava filter registry reports that the majority of filters were inserted following an antecedent VTE event[14]. Individuals were excluded if they had an ICD-10 diagnosis code of portal vein thrombosis (I81), Budd-Chiari syndrome (I82.0), superficial or unclear site of thrombophlebitis (I80.0, I80.3, I80.8, I80.9), or known coagulation defects (D68). All other individuals passing quality control (QC) in the genetic analysis (see below) were defined as control participants free of VTE.

UK Biobank Genetic Data

UK Biobank samples were genotyped using either the UK Bileve or UK Biobank Axiom Array. In total, genotypes were available for 152,732 participants, having been performed in 33 separate batches of ~4700 samples by Affymetrix (High Wycombe, UK). After phasing of individuals using SHAPEIT-2, genotypes from a reference panel (a combined 1000 Genomes/UK10K reference panel) were imputed (statistically transferred) into UK Biobank participants via IMPUTE2 software[15].

Individual-level QC has been described elsewhere[16]. Briefly, individuals with relatedness of 3rd degree or higher were removed, and an additional 480 individuals were excluded due to more heterozygosity than expected or an excess of missing genotype calls. Of the remaining participants, we included 120,286 individuals of European ancestry (self-identified as "Caucasian" and European ancestry based on principal components of ancestry)[17].

Variant level QC exclusion metrics included: call rate < 95%, Hardy-Weinberg equilibrium P value $<1\times10^{-6}$, posterior call probability < 0.9, imputation quality <0.4, and minor allele frequency (MAF) < 0.005. Sex chromosome and mitochondrial genetic data were excluded from this analysis.

Heritability Calculation

Previous twin-twin based studies have estimated the heritability of VTE to be approximately 50–60%[18, 19], however these are estimates are potentially inflated owing to their inclusion of related individuals[20, 21]. For this study, h_g^2 is defined as the proportion of population variance in disease liability - assuming a liability threshold model[22] - explained by the best linear predictor using genotyped variants. Previous work has demonstrated the need for stringent QC in case-control heritability estimation[23], and in addition to the above QC, the following single nucleotide polymorphism (SNP) exclusion criteria were applied to genotyped SNPs: MAF < 0.01, Hardy-Weinberg equilibrium P < 0.01, case-control SNP differential call-rate P-value < 0.05, and variant call rate < 99.5%. We excluded subjects so that no pair of individuals in our analysis had a relatedness coefficient > 0.05 as calculated by the GCTA software[24]. Heritability calculations were performed with the BOLT-REML software[25], and genotyped SNPs were used for calculation given decreased computational demands while providing minimal deflation in heritability estimates compared to analyses with imputed data[26]. In our REML (restricted maximum likelihood) analysis, genotyped SNPs were divided into five separate MAF bins (0.01–0.1, 0.1–0.2, 0.2–0.3, 0.3–0.4, 0.4–0.5) to account for potential biases[27], and we further adjusted for age, sex, chip array at run-time, and twenty principal-component covariates. Raw REML parameter estimates were transformed to h_g^2 using linear transformation[22] assuming the case ascertainment in the UK Biobank reflected population risk with an estimated disease prevalence (K) of 0.5%[28, 29].

We defined h_{GWAS}^2 as the sum of the variance in liability (V_g) explained for each of the known genome-wide VTE SNPs again invoking the liability threshold method[30]. We used the previously published[9, 31, 32] risk allele frequencies and odds ratios for each variant to calculate the liability threshold for each genotype. We present values estimated using a prevalence estimate (K) of 0.5%.

Genome Wide Association Analysis

We studied the outcome variable of VTE case status and predictor variable of DNA sequence variant with covariates of age, gender, and chip array at run-time. DNA sequence variant was modeled in an additive fashion (carriage of 0, 1, or 2 copies of the risk allele). We utilized BOLT-LMM[33] software to perform linear mixed models (LMMs) within the selected individuals. BOLT-LMM was chosen given its ability to adjust for population structure and relatedness whilst efficiently performing association analysis in a population-based cohort. In addition, we performed a secondary analysis where previously associated SNPs were included as covariates (i.e., conditional analysis). To derive effect estimates for each variant that exceeded genome-wide significance or was taken forward for replication, we performed logistic regression (score test) with outcome variable of case status, predictor variable of genotype coded in an additive fashion, and covariates of age, sex, chip array at run-time, and principal components utilizing R and SNPTEST (https://mathgen.stats.ox.ac.uk/genetics_software/snptest/snptest.html) statistical software programs[34].

Replication

For 11 SNPs with publicly available summary statistics or that exceeded a P value of 10^{-5} in UK Biobank, we obtained replication evidence in collaboration with the International Network on Venous Thrombosis (INVENT) Consortium. In April 2015, INVENT published a VTE GWAS with over 65,000 participants, and summary statistics were obtained from this analysis[9]. Association analysis in INVENT consisted of logistic regression, outcome of VTE, SNPs coded in an additive model, and covariates of age, sex, and principal components. BOLT-LMM results in UK Biobank and replication results in INVENT were meta-analyzed via a weighted z-score method, adjusting for an unbalanced ratio of cases to controls. Logistic regression results for the same top SNPs were meta-analyzed using inverse-variance weighting with fixed effects to refine effect size estimates. The above analyses were performed using the METAL analysis software[35]. A one-sided P-value of <0.05 was required for replication, and we set a statistical threshold of $P < 5 \times 10^{-8}$ for genome-wide significance.

Phenome-wide Association Study

To explore the range of phenotypic consequences associated with VTE risk variants, we conducted a PheWAS utilizing publicly available GWAS summary data as well as individual level data for 112,338 participants of European ancestry within UK Biobank. For this analysis we used a genetic risk score composed of 10 independent ($R^2 < 0.01$) VTE-associated variants (supplementary table S1).

To maximize discovery power, we first investigated whether complete SNP data was available for a given phenotype in GWAS summary statistics, and then performed an association analysis. For the summary level data, this approach is equivalent to a inverse-variance weighted fixed effects meta-analysis of the effect of each SNP on the trait or outcome of interest, divided by the effect of each SNP on VTE, mathematically equivalent to a previously described genetic risk score based approach[36]. Of the publicly available disease GWAS summary statistics, only the CARDIoGRAMplusC4D[37] coronary artery disease (CAD) summary data contained all 10 VTE associated variants in the genetic risk score (http://www.cardiogramplusc4d.org/). For phenotypes meeting the discovery threshold P value using GWAS summary statistics, replication was then performed in UK Biobank (see below).

In UK Biobank, we performed the PheWAS analysis using individual level data across 37 disorders. Disease status was ascertained at baseline by self-report, followed by a verbal interview with a trained nurse to confirm the diagnosis, or appropriate EHR codes. We created a genetic risk score weighting by the effect of each SNP on VTE as previously described[38]. We then examined the effect of the risk score on each trait and outcome per unit increase in VTE odds using a logistic model adjusting for age, sex, principal components, and chip array at run-time.

In discovery, phenotypes were declared to be significantly associated with the genetic risk score if they met a Bonferroni corrected P value of < 0.0014 (0.05/37 traits). For replication, a one-sided P-value of <0.05 was required for statistical significance.

Body Mass Index and VTE: Mendelian Randomization Analysis

We identified 68 independent, genome-wide significant SNPs that were associated with BMI^{adj} among Europeans across both sexes from the GIANT consortium (supplementary table S2). Our primary instrument was a weighted genetic risk score (GRS^{BMI}) constructed as previously described[39]. We examined the association of a one standard deviation increase in the GRS^{BMI}, corresponding to a one standard deviation increase in BMI in UK Biobank, with venous thromboembolism. BMI^{adj} was derived in the UK Biobank through inverse normal transformation of BMI after adjustment for age, age^2, and sex as performed in the GIANT consortium BMI GWAS[40]. We then performed logistic regression adjusting for age, sex, ten principal components of ancestry, and chip array at run-time.

Results

Characteristics of the UK Biobank participants according to case-control status are presented in Table 1. VTE cases tended to be older, female, have a higher BMI, and more likely to be smokers, diabetic, and have a history of cancer. The proportion of overall phenotypic variance explained by genotyped SNPs is estimated to be 29.5% (i.e., h_g^2 of 29.5% ± 6.5%).

After QC, 9,155,762 imputed autosomal variants remained for GWAS in a total of 3,290 participants with VTE and 116,868 controls. A total of 214 variants at five distinct loci met the genome-wide significance threshold ($P < 5\times10^{-8}$) (supplementary figures S1 and S2). All five have been previously reported[9]. In UK Biobank, the *F5* Leiden mutation (rs6025; R534Q) was the top association result (2.3% frequency for T allele; OR = 2.34; 95%CI: 2.08–2.62; $P = 7.10\times10^{-50}$); the other four loci were *FGG*, *F11*, *ABO*, and *F2* (Table 2). For ten of the autosomal SNPs previously associated with VTE, a comparison of the odds ratios for these SNPs in the literature versus the odds ratio in the current UK Biobank analysis is shown in supplementary table S3 (Pearson Correlation coefficient = 0.98 for odds ratios from UK Biobank compared to published literature)[9, 31, 41, 42]. Each of these ten SNPs previously associated with VTE were replicated and directionally consistent in UK Biobank.

To identify new variants, we took forward for replication 11 SNPs from the discovery UK Biobank GWAS (supplementary table S4). Following replication, two variants, rs4602861 within Zinc Finger Protein, Multitype-2 protein (*ZFPM2*), OR = 1.11 (95%CI: 1.07–1.15, $P = 4.9\times10^{-10}$) and rs3136516 within Factor 2 (*F2*), OR = 1.10 (95%CI: 1.06–1.13, $P = 7.60\times10^{-9}$), achieved genome-wide significance (Table 2). The rs3136516 variant was identified after conditioning on the prothrombin G20210A (rs1799963) mutation ($R^2 = 0.007$).

The proportion of overall phenotypic variance explained by 14 VTE SNPs ((12 previously published VTE SNPs as well as the two new ones in this study (rs3136516 and rs4602861)) was 5.24% (i.e., h_{GWAS}^2). Supplementary table S5 lists each of the GWAS SNPs with their allele frequencies, odds ratios, and associated V_g utilized for this calculation.

We conducted a PheWAS for a 10 SNP VTE genetic risk score. Besides VTE, the genetic risk score was associated with coronary artery disease in CARDIoGRAMplusC4D summary data [OR = 1.08 (95%CI: 1.05–1.11 per unit increase in VTE odds, $P = 7.9\times10^{-9}$] and was

replicated in UK Biobank ($P < 0.05$) for a combined OR = 1.08 [95%CI: 1.05–1.10 per unit increase in VTE odds, combined $P = 1.08\times10^{-9}$, Figure 1]. No other phenotype demonstrated statistically significant association with these variants after Bonferroni correction (supplementary Figure 3). PheWAS phenotype definitions are listed in supplementary table S6.

In observational studies, BMI has been shown to be an independent risk factor for VTE[43, 44], however prior genetic analyses assessing causality have failed to provide conclusive results[45]. In UK Biobank, we find that BMI^{adj} is associated with higher odds for VTE (OR = 1.37 for every 1 SD higher BMI^{adj}, 95%CI: 1.33–1.41, $P < 2\times10^{-16}$). However, it is uncertain if this association is causal or due to any of several methodological limitations of observational studies including reverse causation or confounding.

Mendelian randomization leverages the fact that genotypes are randomly assigned at meiosis, are independent of non-genetic confounding, and are unmodified by disease processes; as such, Mendelian randomization is useful for causal inference. We developed an instrument consisting of 68 SNPs previously associated with BMI^{adj}. We first verified that GRS^{BMI} associated with BMI^{adj} in UK Biobank ($\beta = 0.84$ per SD in GRS^{BMI}, $P < 2\times10^{-16}$). A one SD increase in BMI^{adj} corresponded to an absolute increase in BMI of 4.6 kg/m^2. Genetically-elevated BMI (a one SD increase) increased the risk of VTE by 57% (OR = 1.57, 95%CI: 1.08–1.97, $P = 0.003$) after adjustment for age, sex, chip-array, and principal components (Table 3).

Discussion

In this study, we leveraged the UK Biobank resource to explore the inherited basis of VTE. We estimated the heritability of VTE to be nearly 30%, performed a GWAS for VTE, which confirmed previous results, and identified two new variants. We showed the phenotypic consequences of a VTE genetic risk score. Finally, we provide genetic support for a causal relationship between BMI and VTE.

These findings permit several conclusions. First, EHR-based biobanks provide an enormous potential to aid genetic discovery for complex, common disease. To date the largest VTE GWAS analyses have been limited to ascertained case-control cohorts[9] or direct-to-consumer genetic repositories[31]. In this study, we utilized a single resource to identify over 3,000 VTE cases and 116,000 controls through EHR and medical-code based phenotyping, representing approximately one-half the number of cases and twice as many controls as the most recent INVENT VTE GWAS analysis. Recent work has questioned the fidelity of such phenotypes[12], however, we find that for VTE, cases ascertained through EHR are of high fidelity, as evidenced by the robust replication of previous genetic findings.

Second, these GWAS results provide evidence that genes not related to coagulation may contribute to VTE risk. Early genetic work identified VTE risk variants in genes limited to the coagulation cascade[8], and later GWAS[41] confirmed these findings. In a meta-analysis, INVENT Consortium investigators provided the first indication that genes outside of the coagulation cascade, *TSPAN15* and *SLC44A2*[9], contribute to VTE risk. In this study, we

find that rs4602861 within *ZFPM2* also contributes to VTE risk. Zinc Finger Protein, Multitype-2 protein is a known transcription factor critical for hematopoiesis and cardiac development[46, 47]. The locus has been associated with circulating Vascular Endothelial Growth Factor (VEGF) levels[48], and recent evidence suggests that VEGF may be critical for venous thrombus resolution[49, 50]. Taken together, *ZFPM2* may influence VTE risk through modulation of circulating VEGF and disruption of the thrombosis balance within the venous system.

Third, variants conferring risk of venous thromboembolism may also increase the risk of CAD. Previous work has demonstrated common associations between VTE and CAD at the *ABO* and *SLC44A2* loci[31]. Building in these findings, in our study a PheWAS utilizing a 10 variant VTE-associated genetic risk score demonstrated an association with increased coronary artery disease risk. While the causal mechanisms underlying atherosclerosis and venous thrombosis are quite different, the downstream activation of the coagulation cascade is common to both myocardial infarction (a direct sequelae of CAD) and VTE. Taken together, our findings suggest that there may be greater overlap in genetic risk variants associated with arterial and venous disease than originally thought.

Fourth, these findings lend genetic support for a causal relationship of obesity with VTE. Previous observational epidemiologic work has demonstrated BMI as an independent risk factor for VTE[13]. However, whether this association reflects a causal relationship has been unclear. Klovaite et al performed a Mendelian randomization study examining the BMI exposure and VTE; however, the statistical evidence was limited to a sub-phenotype of VTE, namely deep vein thrombosis complicated by pulmonary embolism, which represented only 14% of their VTE cases[45]. In this study, human genetic evidence supports a causal role for obesity in risk for VTE. Recent work has suggested that the pro-inflammatory state induced by obesity may result in impaired fibrinolysis and an increased risk of VTE[51]. As such, therapies targeting the obesity epidemic to reduce the burden of cardiometabolic disease may also mitigate the rising incidence of VTE events[44].

These data should be interpreted in the context of the study's limitations. We focused on participants of European ancestry within UK Biobank and therefore results may not be generalizable to other populations. The PheWAS may have been underpowered to detect association for the VTE genetic risk score with certain diseases. Lastly, our phenotype definitions are based largely on ICD and OPCS codes and this may result in misclassification of case status. However, such misclassification should reduce statistical power for discovery and bias results toward the null.

Conclusions

EHR-based biobanks provide enormous potential to uncover the genetic basis of common diseases. For the common disease of VTE, the UK Biobank resource enabled us to estimate heritability, perform genetic discovery, explore the range of phenotypic consequences for disease-associated variants, and test causal inference all within a single resource.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

This research has been conducted using the UK Biobank resource. The sponsors had no role in the design and conduct of the study; collection, management, analysis, and interpretation of the data; preparation, review, or approval of the manuscript; and decision to submit the manuscript for publication.

Sources of Funding: Data provided by the International Network against Thrombosis (INVENT) collaboration was financially supported by the French National Institute for Health and Medical Research; French Ministry of Health; French Medical Research Foundation; French Agency for Research; National Heart, Lung, and Blood Institute (USA); National Institute for Health Research (USA); National Human Genome Research Institute (USA); National Cancer Institute (USA); the Donald W. Reynolds Foundation; Fondation Leducq; The European Commission; Netherlands Organisation for Scientific Research; Netherlands Heart Foundation; Dutch Cancer Foundation; Canadian Institutes of Health Research; Heart and Stroke Foundation of Canada; British Heart Foundation; and ICAN Institute for Cardiometabolism and Nutrition. Dr. Klarin is supported by the National Heart, Lung, and Blood Institute of the National Institutes of Health [T32 HL007734]. Connor Emdin is supported by the Rhodes Trust. Dr. Kathiresan is supported by a Research Scholar award from the Massachusetts General Hospital (MGH), the Donovan Family Foundation, and the National Institutes of Health [R01HL127564].

References

1. Naess IA, Christiansen SC, Romundstad P, Cannegieter SC, Rosendaal FR, Hammerstrøm J. Incidence and mortality of venous thrombosis: a population-based study. J Thromb Haemost. 2007; 5:692–699. [PubMed: 17367492]
2. Zoller B, Li X, Sundquist J, Sundquist K. Familial transmission of venous thromboembolism: a cohort study of 80 214 Swedish adoptees linked to their biological and adoptive parents. Circ Cardiovasc Genet. 2014; 7:296–303. [PubMed: 24795348]
3. Alikhan R, Bedenis R, ATC. Heparin for the prevention of venous thromboembolism in acutely ill medical patients (excluding stroke and myocardial infarction). Cochrane Database of Systematic Reviews. 2014; Art. No.: CD003747. doi: 10.1002/14651858.CD003747.pub4.
4. Rasmussen MSJL, Wille-Jørgensen P. Prolonged thromboprophylaxis with Low Molecular Weight heparin for abdominal or pelvic surgery (Review). Cochrane Database of Systematic Reviews. 2009; Art. No.: CD004318. doi: 10.1002/14651858.CD004318.pub2.
5. Grosse SD, Nelson RE, Nyarko KA, Richardson LC, Raskob GE. The economic burden of incident venous thromboembolism in the United States: A review of estimated attributable healthcare costs. Thromb Res. 2016; 137:3–10. [PubMed: 26654719]
6. Kyrle PA, Eichinger S. Is Virchow's triad complete? Blood. 2009; 114:1138–9. [PubMed: 19661277]
7. Bertina RM, Koeleman BP, Koster T, Rosendaal FR, Dirven RJ, de Ronde H, et al. Mutation in blood coagulation factor V associated with resistance to activated protein C. Nature. 1994; 369:64–7. [PubMed: 8164741]
8. Poort SR, Rosendaal FR, Reitsma PH, Bertina RM. A common genetic variation in the 3′-untranslated region of the prothrombin gene is associated with elevated plasma prothrombin levels and an increase in venous thrombosis. Blood. 1996; 88:3698–703. [PubMed: 8916933]
9. Germain M, Chasman DI, de Haan H, Tang W, Lindstrom S, Weng LC, et al. Meta-analysis of 65,734 individuals identifies TSPAN15 and SLC44A2 as two susceptibility loci for venous thromboembolism. Am J Hum Genet. 2015; 96:532–42. [PubMed: 25772935]
10. Schneider D, Riegman PH, Cronin M, Negrouk A, Moch H, Balling R, et al. Accelerating the Development and Validation of New Value-Based Diagnostics by Leveraging Biobanks. Public Health Genomics. 2016; 19:160–169. [PubMed: 27237867]
11. Collins R. What makes UK Biobank special? Lancet. 2012; 379:1173–1174. [PubMed: 22463865]

12. Newton KM, Peissig PL, Kho AN, Bielinski SJ, Berg RL, Choudhary V, et al. Validation of electronic medical record-based phenotyping algorithms: results and lessons learned from the eMERGE network. J Am Med Inform Assoc. 2013; 20:e147–54. [PubMed: 23531748]
13. Heit JA, Spencer FA, White RH. The epidemiology of venous thromboembolism. J Thromb Thrombolysis. 2016; 41:3–14. [PubMed: 26780736]
14. Uberoi R, Tapping CR, Chalmers N, Allgar V. British Society of Interventional Radiology (BSIR) Inferior Vena Cava (IVC) Filter Registry. Cardiovasc Intervent Radiol. 2013; 36:1548–61. [PubMed: 23511992]
15. UK Biobank Analysis Team. [Accessed October 1, 2015] Genotype imputation and genetic association studies of UK Biobank. 2015. http://www.ukbiobank.ac.uk/wp-content/uploads/2014/04/imputation_documentation_May2015.pdf
16. UK Biobank Analysis Team. [Accessed September 16, 2015] UK Biobank genotype collections and QC on 150k participants. Apr. 2014 http://www.ukbiobank.ac.uk/wp-content/uploads/2014/04/UKBiobank_genotyping_QC_documentation-web.pdf
17. Price AL, Patterson NJ, Plenge RM, Weinblatt ME, Shadick NA, Reich D. Principal components analysis corrects for stratification in genome-wide association studies. Nat Genet. 2006; 38:904–9. [PubMed: 16862161]
18. Heit JA, Phelps MA, Ward SA, Slusser JP, Petterson TM, De Andrade M. Familial segregation of venous thromboembolism. J Thromb Haemost. 2004; 2:731–6. [PubMed: 15099278]
19. 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. Am J Hum Genet. 2000; 67:1452–1459. [PubMed: 11038326]
20. Zaitlen N, Pasaniuc B, Sankararaman S, Bhatia G, Zhang J, Gusev A, et al. Leveraging population admixture to characterize the heritability of complex traits. Nat Genet. 2014; 46:1356–62. [PubMed: 25383972]
21. Yang J, Bakshi A, Zhu Z, Hemani G, Vinkhuyzen AA, Lee SH, et al. Genetic variance estimation with imputed variants finds negligible missing heritability for human height and body mass index. Nat Genet. 2015; 47:1114–20. [PubMed: 26323059]
22. Lee SH, Wray NR, Goddard ME, Visscher PM. Estimating missing heritability for disease from genome-wide association studies. Am J Hum Genet. 2011; 88:294–305. [PubMed: 21376301]
23. Lee SH, DeCandia TR, Ripke S, Yang J, et al. Schizophrenia Psychiatric Genome-Wide Association Study C, International Schizophrenia C. Estimating the proportion of variation in susceptibility to schizophrenia captured by common SNPs. Nat Genet. 2012; 44:247–50. [PubMed: 22344220]
24. Yang J, Lee SH, Goddard ME, Visscher PM. GCTA: a tool for genome-wide complex trait analysis. Am J Hum Genet. 2011; 88:76–82. [PubMed: 21167468]
25. Loh PR, Bhatia G, Gusev A, Finucane HK, Bulik-Sullivan BK, Pollack SJ, et al. Contrasting genetic architectures of schizophrenia and other complex diseases using fast variance-components analysis. Nat Genet. 2015; 47:1385–92. [PubMed: 26523775]
26. Bhatia G, Gusev A, Loh P-R, Vilhjálmsson BJ, Ripke S, Purcell S, et al. Haplotypes of common SNPs can explain missing heritability of complex diseases. 2015 bioRxiv.
27. Lee SH, Yang J, Chen GB, Ripke S, Stahl EA, Hultman CM, et al. Estimation of SNP heritability from dense genotype data. Am J Hum Genet. 2013; 93:1151–5. [PubMed: 24314550]
28. Deitelzweig SB, Johnson BH, Lin J, Schulman KL. Prevalence of clinical venous thromboembolism in the USA: current trends and future projections. Am J Hematol. 2011; 86:217–20. [PubMed: 21264912]
29. Tagalakis V, Patenaude V, Kahn SR, Suissa S. Incidence of and mortality from venous thromboembolism in a real-world population: the Q-VTE Study Cohort. Am J Med. 2013; 126:832e13–21.
30. So HC, Gui AH, Cherny SS, Sham PC. Evaluating the heritability explained by known susceptibility variants: a survey of ten complex diseases. Genet Epidemiol. 2011; 35:310–7. [PubMed: 21374718]

31. Hinds DA, Buil A, Ziemek D, Martinez-Perez A, Malik R, Folkersen L, et al. Genome-wide association analysis of self-reported events in 6135 individuals and 252 827 controls identifies 8 loci associated with thrombosis. Hum Mol Genet. 2016; 25:1867–74. [PubMed: 26908601]
32. Morange PE, Suchon P, Tregouet DA. Genetics of Venous Thrombosis: update in 2015. Thromb Haemost. 2015; 114:910–9. [PubMed: 26354877]
33. Loh PR, Tucker G, Bulik-Sullivan BK, Vilhjalmsson BJ, Finucane HK, Salem RM, et al. Efficient Bayesian mixed-model analysis increases association power in large cohorts. Nat Genet. 2015; 47:284–90. [PubMed: 25642633]
34. R Development Core Team. R: A language and environment for statistical computing. R Foundation for Statistical Computing; Vienna, Austria: 2008. http://www.R-project.org
35. Willer CJ, Li Y, Abecasis GR. METAL: fast and efficient meta-analysis of genomewide association scans. Bioinformatics. 2010; 26:2190–1. [PubMed: 20616382]
36. Burgess S, Butterworth A, Thompson SG. Mendelian randomization analysis with multiple genetic variants using summarized data. Genet Epidemiol. 2013; 37:658–65. [PubMed: 24114802]
37. CARDIoGRAMplusC4D Consortium. A comprehensive 1000 Genomes-based genome-wide association meta-analysis of coronary artery disease. Nat Genet. 2015; 47:1121–30. [PubMed: 26343387]
38. Mega JL, Stitziel NO, Smith JG, Chasman DI, Caulfield MJ, Devlin JJ, et al. Genetic risk, coronary heart disease events, and the clinical benefit of statin therapy: an analysis of primary and secondary prevention trials. Lancet. 2015; 385:2264–2271. [PubMed: 25748612]
39. Voight BF, Peloso GM, Orho-Melander M, Frikke-Schmidt R, Barbalic M, Jensen MK, et al. Plasma HDL cholesterol and risk of myocardial infarction: a mendelian randomisation study. Lancet. 2012; 380:572–580. [PubMed: 22607825]
40. Locke AE, Kahali B, Berndt SI, Justice AE, Pers TH, Day FR, et al. Genetic studies of body mass index yield new insights for obesity biology. Nature. 2015; 518:197–206. [PubMed: 25673413]
41. Li Y, Bezemer ID, Rowland CM, Tong CH, Arellano AR, Catanese JJ, et al. Genetic variants associated with deep vein thrombosis: the F11 locus. J Thromb Haemost. 2009; 7:1802–8. [PubMed: 19583818]
42. Heit JA, Cunningham JM, Petterson TM, Armasu SM, Rider DN, MDEA. Genetic variation within the anticoagulant, procoagulant, fibrinolytic and innate immunity pathways as risk factors for venous thromboembolism. J Thromb Haemost. 2011; 9:1133–42. [PubMed: 21463476]
43. Severinsen MT, Kristensen SR, Johnsen SP, Dethlefsen C, Tjonneland A, Overvad K. Anthropometry, body fat, and venous thromboembolism: a Danish follow-up study. Circulation. 2009; 120:1850–7. [PubMed: 19858417]
44. Heit JA. Epidemiology of venous thromboembolism. Nat Rev Cardiol. 2015; 12:464–74. [PubMed: 26076949]
45. Klovaite J, Benn M, Nordestgaard BG. Obesity as a causal risk factor for deep venous thrombosis: a Mendelian randomization study. J Intern Med. 2015; 277:573–84. [PubMed: 25161014]
46. Tevosian SG, Deconinck AE, Cantor AB, Rieff HI, Fujiwara Y, Corfas G, et al. FOG-2: A novel GATA-family cofactor related to multitype zinc-finger proteins Friend of GATA-1 and U-shaped. Proc Natl Acad Sci U S A. 1999; 96:950–5. [PubMed: 9927674]
47. Garnatz AS, Gao Z, Broman M, Martens S, Earley JU, Svensson EC. FOG-2 mediated recruitment of the NuRD complex regulates cardiomyocyte proliferation during heart development. Dev Biol. 2014; 395:50–61. [PubMed: 25196150]
48. Choi SH, Ruggiero D, Sorice R, Song C, Nutile T, Vernon Smith A, et al. Six Novel Loci Associated with Circulating VEGF Levels Identified by a Meta-analysis of Genome-Wide Association Studies. PLoS Genet. 2016; 12:e1005874. [PubMed: 26910538]
49. Waltham M, Burnand KG, Collins M, McGuinness CL, Singh I, Smith A. Vascular endothelial growth factor enhances venous thrombus recanalisation and organisation. Thromb Haemost. 2003; 89:169–76. [PubMed: 12540967]
50. Waltham M, Burnand KG, Collins M, Smith A. Vascular endothelial growth factor and basic fibroblast growth factor are found in resolving venous thrombi. J Vasc Surg. 2000; 32:988–96. [PubMed: 11054231]

51. Orenes-Pinero E, Pineda J, Roldan V, Hernandez-Romero D, Marco P, Tello-Montoliu A, et al. Effects of Body Mass Index on the Lipid Profile and Biomarkers of Inflammation and a Fibrinolytic and Prothrombotic State. J Atheroscler Thromb. 2015; 22:610–7. [PubMed: 25737193]

Clinical Perspective

Venous thromboembolism (VTE) is a common disease with both genetic and environmental contributors. In our study, we identified 2 additional DNA sequence variants that contribute to VTE risk. At the current time, at least 14 DNA sequence variants across the genome have been shown to confer risk for VTE. We demonstrate that a subset of these appear to contribute to coronary artery disease risk as well. Lastly, obesity has been implicated as an independent epidemiologic risk factor for VTE, however, whether this association reflects a causal relationship has been unclear. We found that human genetic evidence supports a causal role for obesity in risk for VTE. Future efforts to decrease obesity may prove effective in reducing risk for VTE.

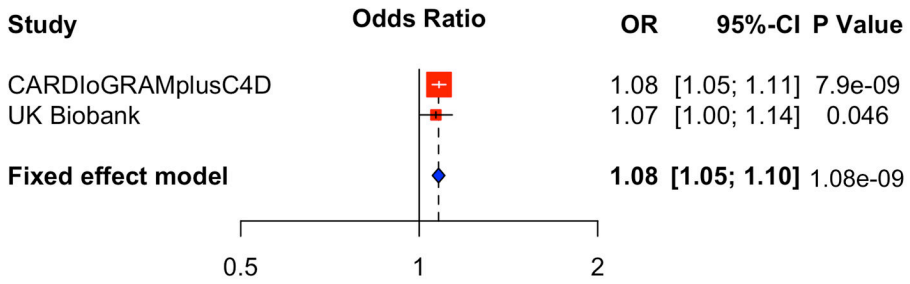

Figure 1.

Phenome-wide association results for the ten variant VTE genetic risk score and coronary artery disease. Odds ratio per unit increase in VTE odds on coronary artery disease within CARDIoGRAMplusC4D data and UK Biobank. For the CARDIOGRAMplusC4D coronary artery disease summary statistics, we performed an inverse-variance weighted fixed effects meta-analysis. Estimates in UK Biobank were derived using logistic regression adjusting for age, sex, chip-array, and ten principal components of ancestry. Abbreviations: OR, odds ratio

Table 1

Demographic and clinical characteristics for venous thromboembolism cases and controls in UK Biobank

	Cases	Controls
N Individuals	3290	116,868
Age ± SD, years	59.5 ± 7.2	56.8 ± 7.9
Male, n (%)	1426 (43.3%)	55,446 (47.4%)
Prevalent Cancer, n (%)	471 (14.3%)	9757 (8.3%)
Ever Smoker, n (%)	1252 (38%)	39799 (34%)
Hypertension, n (%)	1073 (32.6%)	26,037 (22.3%)
Diabetes Mellitus, n (%)	317 (9.6%)	6074 (5.2%)
Hyperlipidemia, n (%)	1022 (31.1%)	22,273 (19.1%)
Body Mass Index ± SD, kg/m^2	29.3 ± 5.8	27.5 ± 4.8

SD = Standard Deviation

Table 2

Genetic loci associated with venous thromboembolism after discovery and replication

UK Biobank Discovery Analysis								Replication		Combined	
Lead SNP	**Chr**	**Gene**	**Description**	**EA**	**EAF**	**OR[*] (95% CI)**	**P[†]**	**OR (95% CI)**	**P**	**OR (95% CI)**	**P**
Previously Described Loci											
rs6025	1	*F5*	missense	T	0.02	3.5 (2.96–4.11)	7.10×10^{-50}	-	-		
rs2066865[‡]	4	*FGG*	3′ UTR	A	0.24	1.21 (1.15–1.29)	3.10×10^{-11}	-	-		
rs4253416	4	*F11*	intronic	T	0.45	1.18 (1.12–1.24)	2.0×10^{-10}	-	-		
rs2519093	9	*ABO*	intronic	T	0.19	1.41 (1.32–1.50)	6.00×10^{-26}	-	-		
rs8176645	9	*ABO*	intronic	A	0.33	1.28 (1.22–1.35)	4.40×10^{-21}	-	-		
rs1799963	11	*F2*	3′ UTR	A	0.01	2.63 (2.03–3.40)	4.90×10^{-13}	-	-		
New Independent Genome-Wide Significant Variants											
rs4602861	8	*ZFPM2*	intronic	A	0.73	1.08 (1.03–1.15)	0.0045	1.13 (1.08–1.19)	5.04×10^{-7}	1.11 (1.07–1.15)	4.88×10^{-10}
rs3136516[§]	11	*F2*	intronic	G	0.48	1.10 (1.04–1.15)	0.00033	1.10 (1.06–1.15)	5.65×10^{-6}	1.10 (1.06–1.13)	7.60×10^{-9}

[*] Odds ratio calculated from logistic regression analysis

[†] P value based on linear mixed modeling analysis

[‡] Reporting the known causal variant at this locus

[§] Results after conditioning on rs1799963

Chr = Chromosome, CI = Confidence Interval, EA = Effect Allele, EAF = Effect Allele Frequency, OR = Odds Ratio

Table 3

Instrumental variable analysis estimate of the association of genetically raised BMI and VTE

Instrument	VTE OR (95% CI) Per SD Increase in BMI	P Value
Observational Epidemiology	1.37 (1.33–1.41)	2×10^{-16}
BMI Genetic Risk Score	1.57 (1.08–1.97)	0.003

BMI = Body Mass Index, CI = Confidence Interval, OR = Odds Ratio, SD = Standard Deviation