

Leveraging cell type specific regulatory regions to detect SNPs associated with tissue factor pathway inhibitor plasma levels

Jessica Dennis, Alejandra Medina-Rivera, Vinh Truong, Lina Antounians, Nora Zwingerman, Giovana Carrasco, Lisa Strug, Phil Wells, David-alexandre Trégouët, Pierre-Emmanuel Morange, et al.

▶ **To cite this version:**

Jessica Dennis, Alejandra Medina-Rivera, Vinh Truong, Lina Antounians, Nora Zwingerman, et al.. Leveraging cell type specific regulatory regions to detect SNPs associated with tissue factor pathway inhibitor plasma levels. Genetic Epidemiology, 2017, 41 (5), pp.455-466. 10.1002/gepi.22049 . hal-01766535

HAL Id: hal-01766535
https://amu.hal.science/hal-01766535

Submitted on 13 Apr 2018

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Leveraging cell type specific regulatory regions to detect SNPs associated with tissue factor pathway inhibitor plasma levels

Jessica Dennis Alejandra Medina-Rivera Vinh Truong Lina Antounians Nora Zwingerman Giovana Carrasco Lisa Strug Phil Wells David-Alexandre Trégouët Morange PE, Wilson MD, Gagnon F

ABSTRACT

Tissue factor pathway inhibitor (TFPI) regulates the formation of intravascular blood clots, which manifest clinically as ischemic heart disease, ischemic stroke, and venous thromboembolism (VTE). TFPI plasma levels are heritable, but the genetics underlying TFPI plasma level variability are poorly understood. Herein we report the first genome-wide association scan (GWAS) of TFPI plasma levels, conducted in 251 individuals from five extended French-Canadian Families ascertained on VTE. To improve discovery, we also applied a hypothesis-driven (HD) GWAS approach that prioritized single nucleotide polymorphisms (SNPs) in (1) hemostasis pathway genes, and (2) vascular endothelial cell (EC) regulatory regions, which are among the highest expressers of *TFPI*. Our GWAS identified 131 SNPs with suggestive evidence of association (P-value $< 5 \times 10^{-8}$), but no SNPs reached the genome-wide threshold for statistical significance. Hemostasis pathway genes were not enriched for TFPI plasma level associated SNPs (global hypothesis test P-value = 0.147), but EC regulatory regions contained more TFPI plasma level associated SNPs than expected by chance (global hypothesis test P-value = 0.046). We therefore stratified our genome-wide SNPs, prioritizing those in EC regulatory regions via stratified false discovery rate (sFDR) control, and reranked the SNPs by q-value. The minimum q-value was 0.27, and the top-ranked SNPs did not show association evidence in the MARTHA replication sample of 1,033 unrelated VTE cases. Although this study did not result in new loci for TFPI, our work lays out a strategy to utilize epigenomic data in prioritization schemes for future GWAS studies.

1 INTRODUCTION

Thrombosis—the formation of a blood clot inside a vein or artery—is a leading cause of death worldwide. Arterial thrombosis, which includes ischemic heart disease and ischemic stroke, accounts for one in four deaths (GBD 2013 Mortality and Causes of Death Collaborators, 2015), and venous thromboembolism (VTE), which comprises deep vein thrombosis and pulmonary embolism, is the third leading cause of cardiovascular disease death after ischemic heart disease and stroke (Goldhaber & Bounameaux, 2012).

Low tissue factor pathway inhibitor (TFPI) plasma levels are a risk factor for both arterial thrombosis (Morange et al., 2004) and VTE (Zakai, Lutsey, Folsom, Heckbert, & Cushman, 2010). Vascular endothelial cells (ECs) are an important site of TFPI expression (Bajaj, Kuppuswamy, Saito, Spitzer, & Bajaj, 1990; Lupu, Lupu, Dennehy, Kakkar, & Scully, 1995; Lupu et al., 1999), and are fundamental to the coagulation and inflammatory responses that lay at the heart of thrombosis (Aird, 2012; Verhamme & Hoylaerts, 2006).

TFPI regulates the blood coagulation cascade and limits the intravascular formation of fibrin clots by inhibiting two procoagulant complexes: TF-FVIIa (Novotny, Girard, Miletich, & Broze, 1989), and early forms of prothrombinase (Wood et al., 2013). Up to 52% of the

variability in TFPI plasma levels is attributable to genetics (Almasy et al., 2005; Bladbjerg et al., 2006; Dennis et al., 2016; Warren et al., 2005), but few genes have been implicated. Most studies have been of candidate coagulation genes (*TFPI*, *PROS1*, *F5*), yet variants in these genes explained little of TFPI plasma level variability (Dennis, Kassam, Morange, Tregouet, & Gagnon, 2015), and *in vitro* and mouse models suggested that genes outside of the hemostatic pathway may be important (Lupu, Zhu, Popescu, Wren, & Lupu, 2011; Sanchez-Solana, Motwani, Li, Eswaran, & Kumar, 2012). Genome-wide linkage studies of TFPI plasma levels have been reported (Almasy et al., 2005; Dennis et al., 2016), and fine mapping of the most significant linkage signal identified three SNPs 25.1 Mb upstream of *TFPI* associated with TFPI plasma level variability and VTE risk (Dennis et al., 2016). Much remains to be learned about the genetics and biology of thrombosis (Morange, Suchon, & Tregouet, 2015), and genome-wide association scans (GWAS) of TFPI plasma levels may contribute to this goal by identifying additional putative loci.

Herein we present results of the first reported GWAS of TFPI plasma levels. Analyses were conducted in the F5L Family Study, which consisted of 251 individuals from five French-Canadian families ascertained on VTE. Advantages of the study sample are the inclusion of mostly healthy individuals who were deeply phenotyped, a reduced risk of population stratification (Mersha et al., 2015), and better genotyping quality control afforded by the family-based design (Wijsman, 2012). A limitation of the study is power due to its sample size. We attempted to overcome this limitation by applying a hypothesis-driven (HD) GWAS using the stratified false discovery rate (sFDR) control (Sun, Craiu, Paterson, & Bull, 2012) for prioritization. sFDR control is an efficient and potentially powerful way to incorporate prior information into GWAS, and consists of assigning individual SNPs to high- and low-priority strata based on external information, and applying FDR control separately in each stratum (Sun et al., 2006; Yoo et al., 2010). The FDR is controlled overall and within each stratum, and no power is lost if the hypothesis is uninformative.

We hypothesized that SNPs in gene regulatory regions in vascular EC would be more likely to be associated with TFPI plasma levels than expected by chance. Over 90% of top-scoring GWAS SNPs are nonprotein coding, regardless of phenotype, and three-quarters of this 90% lie within an active gene regulatory region in at least one cell type, or are in complete linkage disequilibrium (LD) with an SNP in an active regulatory region (Maurano et al., 2012). Moreover, GWAS hits across multiple phenotypes systematically fall into gene regulatory regions in a pathologically or physiologically relevant tissue or cell type (Dimas et al., 2009; Ernst et al., 2011; Maurano et al., 2012), and the cell type specific enrichment is observed across the range of *P*-values, not simply for top-scoring SNPs (Maurano et al., 2012).

The objectives of our study were to conduct a GWAS of TFPI plasma levels and to test an HD-GWAS approach in which we prioritized (1) SNPs in hemostasis pathway genes, and (2) SNPs in active enhancer/promoter regions used by ECs. We first tested our hypotheses *globally*, using the collection of SNPs in hemostasis pathway genes or in enhancer/promoter regions in ECs. We then applied sFDR control genome wide to determine if any individual SNPs in the prioritized regions were significantly associated with TFPI plasma levels after prioritization.

2 MATERIALS AND METHODS

2.1 Study populations

Our discovery sample was the French-Canadian Family Study on Factor V Leiden (F5L) Thrombophilia (F5L Family Study), which included 369 individuals from five multigenerational French-Canadian families (Antoni et al., 2010; Gagnon, Aissi, Carrie, Morange, & Tregouet, 2014). Probands were patients at the Thrombosis Clinic at the Ottawa Hospital between 2005 and 2006, had a history of idiopathic, objectively diagnosed VTE, and carried the F5L mutation. Probands' biological living relatives 10 years of age and older were eligible to participate, unless the relative was deceased or lived far from the study's catchment area. All invited participants consented (n = 255), completed an interviewer-administered questionnaire on their personal and medical information, and provided blood samples for the measurement of hemostatic traits, lipid-related traits, and genotyping. The research ethics boards of the University of Toronto and the Ottawa Hospital Research Institute approved this study.

Our replication sample was the MARseille THrombosis Association study (MARTHA), which included 1,542 patients with idiopathic, objectively diagnosed VTE seen at the Thrombophilia Centre of La Timone Hospital, Marseille, France, between January 1994 and October 2005 (Antoni et al., 2011; Oudot-Mellakh et al., 2012). Participants completed a physician-administered questionnaire on their personal and medical histories and provided a fasting blood sample. All participants were white, and the majority reported French ancestry going back at least three generations. The Ethics Committee at La Timone Hospital approved this study.

2.2 Genotyping and imputation

F5L Family Study participants were genotyped using the Illumina Human660W-Quad beadchip. Quality control and imputation steps have been described elsewhere (Antoni et al., 2010; Dennis et al., 2016). In brief, we used data on 490,083 genotyped autosomal SNPs with at least 20 observations of the minor allele (minor allele frequency (MAF) 0.04). We estimated haplotypes while accounting for family structure with SHAPEIT v2.r778 (Howie, Fuchsberger, Stephens, Marchini, & Abecasis, 2012; O'Connell et al., 2014) and imputed genotypes into these haplotypes using the 1000 Genomes Phase 1 Total European Ancestry (EUR) reference population (2010-11-23 release) (Delaneau, Marchini, 1000 Genomes Project Consortium, & 1000 Genomes Project Consortium, 2014) and IMPUTE2 v2.3.1 (Howie, Donnelly, & Marchini, 2009). We filtered imputed SNPs with MAF < 0.04 and poor imputation quality (info <0.3), for a total of 6,159,088 imputed and genotyped SNPs. We used the imputed allele dose (a continuous value between 0 and 2) in all analyses.

Genotyping and imputation in MARTHA have been described elsewhere (Dennis et al., 2016; Oudot-Mellakh et al., 2012). MARTHA subjects were genotyped with either the Illumina Human610-Quad or the Illumina Human660W-Quad beadchips, and additional genotypes were imputed relative to the 1000 Genomes Total European Ancestry (EUR) population (2010-08 release) using MACH v1.0.18.c (Li, Willer, Sanna, & Abecasis, 2009; Li, Willer, Ding, Scheet, & Abecasis, 2010). We filtered imputed SNPs with MAF < 0.01 and poor imputation quality (r^2 < 0.3, analogous to the info <0.3 criterion from IMPUTE2, above), and analyzed the imputed allele dose. In total, 7,804,320 genotyped and imputed SNPs were available for replication analyses.

2.3 Biological measurements

Free TFPI plasma levels were measured in the F5L Family and MARTHA studies using the Asserachrom® Free TFPI enzyme immunoassay from Diagnostica Stago (Asnières, France). TFPI plasma levels were measured in 251 individuals from the F5L Family Study and in 1,170 individuals from MARTHA. Heparin had been administered to 137 MARTHA subjects undergoing biological thrombophilia testing at the time of blood draw. Heparin induces the release of TFPI from the EC surface, resulting in a large spike in TFPI plasma levels (Morange et al., 2007), and so these individuals were excluded from the replication analyses.

2.4 Epigenetic data generation and analysis

Cell culture: Two biological replicates (15 year old healthy Black male and 60 year old healthy Caucasian male) of primary aortic ECs (Human Aortic ECs, HAEC) were obtained from Cell Applications (catalogue #304-05a) and cultured in EC Growth Media (catalogue #211-500). Cells were grown to passage 6 and cross-linked in 1% formaldehyde prior to proceeding with the chromatin immunoprecipitation. ChIP-seq for H3K27ac was performed with 10 μg of mouse anti-H3K27ac (Millipore, 05–1334 monoclonal), as previously described (Schmidt et al., 2009). Sequencing reads were aligned to the hg19 assembly as previously described (Ballester et al., 2014) (see supplementary Table S1 for sequencing read depth and quality control metrics).

ChIP-seq peaks were called relative to data obtained from nonimmunoprecipitated whole cell extracts using MACS2 (broad peaks setting) with a cut-off of FDR q-value ≤ 0.05 (Zhang et al., 2008). These data have been deposited into ArrayExpress (www.ebi.ac.uk/arrayexpress; accession number E-MTAB-4935).

3 STATISTICAL ANALYSES

The F5L Family Study sample served as the primary discovery sample in all analyses. Most individuals in the study were healthy, and our HAEC regulatory region data were also from healthy individuals. The MARTHA sample, on the other hand, consisted of subjects with a recent history of VTE, and the regulatory landscape in MARTHA subjects may not have been comparable to that in healthy individuals. We therefore did not test our global hypotheses in the MARTHA sample. Instead, we tested individual SNPs for replication in MARTHA, because MARTHA has served as a valuable replication sample in previous studies of TFPI plasma levels (Dennis et al., 2016), and other hemostatic traits (Antoni et al., 2010, 2011).

3.1 Genome-wide association analysis

We conducted a GWAS of TFPI plasma levels in the F5L Family Study using a mixed linear model to account for family structure, as implemented in GEMMA (Zhou & Stephens, 2012). We modeled the natural logarithm of TFPI plasma levels under an additive genetic effect and included age and sex as covariates. In sensitivity analyses, we also modeled an empirical normal quantile transformation of TFPI plasma levels in which ranked values were scaled and fit to a normal distribution (Peng, Yu, Dehoff, & Amos, 2007). Although the empirical normal quantile transformation accounts for non-normal outcome distributions, information may be lost in the transformation.

3.2 Global hypothesis test

We evaluated our hypothesis globally by testing whether SNPs in prioritized regions had smaller *P*-values (larger association statistics) than expected by chance (Sun et al., 2012). Our test statistic was the sum of association Wald values for SNPs in prioritized regions, and we compared the observed sum to the distribution of the sum generated by 1,000 permutations under the null hypothesis of no association. We used a semiparametric bootstrapping method (MVNpermute; Abney, 2015) to permute the phenotype 1,000 times while accounting for familial relatedness, and ran 1,000 association tests between each permuted phenotype and prioritized SNPs. We summed the Wald values from each permutation run to generate the null distribution of 1,000 summed Wald values. We calculated an empirical *P*-value for the hypothesis (global hypothesis *P*-value), as the proportion of summed Wald values from permutation runs that exceeded the observed summed Wald value. *P*-values < 0.05 suggested that the prioritization scheme significantly enriched for associated SNPs.

We also tested whether additional subsetting of the primary hypothesis improved enrichment. As above, we summed the Wald statistics of SNPs prioritized under our primary hypothesis, and for subsets of SNPs. We weighted each sum by the number of variants in each group, and computed the difference between the proportional sum statistics (Soave et al., 2015). The observed difference was compared to the difference from 1,000 permutation runs, as above, and an empirical *P*-value < 0.05 indicated that the subsetting added additional information.

3.3 HD genome-wide association analysis via sFDR control

Ultimately, the vascular EC hypothesis showed the most enrichment of TFPI plasma level associated SNPs. We therefore assigned SNPs in H3K27ac-bound regions in vascular EC to the high priority stratum and applied sFDR control to our GWAS results (Sun et al., 2006; Yoo et al., 2010). For comparison, we also applied weighted (w) FDR control to our single-SNP association tests (Roeder, Bacanu, Wasserman, & Devlin, 2006). SNPs in prioritized regions were assigned a weight proportional to the MACS2 peak *P*-value (representing the confidence of the peak call) (Zhang et al., 2008), transformed to stabilize the variance (sqrt(–$\log_{10}$(*P*-value)), and SNPs in nonprioritized regions were assigned a weight of 1 (Li et al., 2013). All weights were then scaled to have a mean of 1, the only necessary assumption to the wFDR, and FDR control was applied to the weighted *P*-values (Roeder et al., 2006). The wFDR is marginally more powerful than the sFDR when the prior is informative; when the prior is uninformative or misleading, however, the sFDR outperforms the wFDR (Yoo et al., 2010).

We conservatively based our interpretations on sFDR *q*-values, and SNPs with sFDR *q*-values < 0.10 were deemed statistically significant and were tested for replication in MARTHA (*q*-values > 0.05 are required in GWAS to strike a balance between true discoveries and non-discoveries, although the optimal threshold will depend on the proportion of true alternative hypotheses in a study; Craiu & Sun, 2008). Statistical models in MARTHA included age, sex, medication use, and the first four principal components as covariates. We also tested top SNPs for association with TFPI plasma levels in the F5L family study under quantile regression models, with quantiles set at 0.50 and 0.25. Quantile regression differentially weights associations at different values of the outcome distribution, and may be more powerful than mean regression when the outcome is skewed (White et al., 2015). We used the R package lqmm (Geraci, 2014), which accounted for clustering by family, but did not explicitly model family structures as in our primary (mean regression) analysis.

FDR methods assume that all tests (i.e., SNPs) are independent. In a sensitivity analysis, we therefore restricted the HD-GWAS to genotyped tag SNPs where there was less LD. We prioritized SNPs that lay directly in a prioritized region, and those in LD with an SNP in a prioritized region. We used LD estimates from SNPs in the 1000 Genomes Phase 1 Total European Ancestry (EUR) reference population (2010-11-23 release), and annotated a genotyped SNP to a prioritized region if it was in LD ($r^2 > 0.80$ and within 1 Mb) with a second SNP that lay in a prioritized region.

The stratified and weighted FDR analyses described in this article are implemented as a standalone tool (*hdgwas_analysis.py*), available in the following git repository: https://github.com/gcarrasco/HD_GWAS_Package.

4 RESULTS

Our GWAS of TFPI plasma levels included 251 individuals from the F5L Family Study. Most individuals in the study were apparently healthy, and only 15 had a history of VTE (Table 1). Compared to the MARTHA replication sample, which included only VTE cases, individuals in the F5L Family Study were younger and included equal proportions of males and females. Thrombosis risk factors, however, were well represented in both data samples: nearly one-quarter of the F5L Family and MARTHA study samples smoked, and the mean body mass index in both samples was above the overweight threshold of 25 kg/m^2. A small number of individuals in the F5L Family Study used antiplatelet (6.8%) or anticoagulant (3.2%) mediations, and TFPI plasma levels did not vary by medication status. Conversely, 6.7% and 22.7% of MARTHA subjects used antiplatelet or anticoagulant medications, and medication status was significantly associated with TFPI plasma levels ($P < 0.001$). TFPI plasma levels were slightly skewed in both datasets, even after taking the natural logarithm, with Shapiro-Wilk normality test *P*-values of 0.005 in the F5L Family Study and 0.03 in MARTHA (supplementary Fig. S1).

Table 1. Study sample characteristics

Characteristic	F5L Family Study (*N* = 251)	MARTHA (*N* = 1,033)
Female sex, *N* (%)	126 (50.2)	684 (66.2)
Mean age (SD)	40.5 (18.0)	47.5 (15.2)
History of venous thromboembolism, *N* (%)	15 (6.0)	1,033 (100)
Mean TFPI (SD), in ng/mL	8.3 (3.8)	13.5 (6.5)
F5L mutation carrier, *N* (%)	63 (25.1)	237 (22.9)
Current smoker	61 (24.3)	234 (22.7)
Female hormone therapy use, *N* (%)	18 (7.2)	66 (6.4)
Mean body mass index (SD), in kg/m^2	26.9 (5.9)	25.4 (4.8)
Antiplatelet use, *N* (%)	17 (6.8)	69 (6.7)

Characteristic	F5L Family Study (*N* = 251)	MARTHA (*N* = 1,033)
Anticoagulant use (vitamin K antagonists and low molecular weight or unfractioned heparin), *N* (%)[a]	8 (3.2)	235 (22.7)

- *N*, number; SD, standard deviation; TFPI, tissue factor pathway inhibitor; F5L, the F5 Leiden mutation.
- a Information was not available on the specific types of anticoagulant drugs used by individuals in the F5L Family Study. MARTHA individuals using heparin were excluded from all analyses.

In total, 131 SNPs with suggestive (*P*-value < 1×10^{-5}) evidence for association were identified in the GWAS, but no SNPs reached the genome-wide significance threshold of *P*-value < 5×10^{-8} (Fig. 1). Results were similar after applying the empirical normal quantile transformation (supplementary Fig. S2), and when using genotyped SNPs only (supplementary Fig. S3).

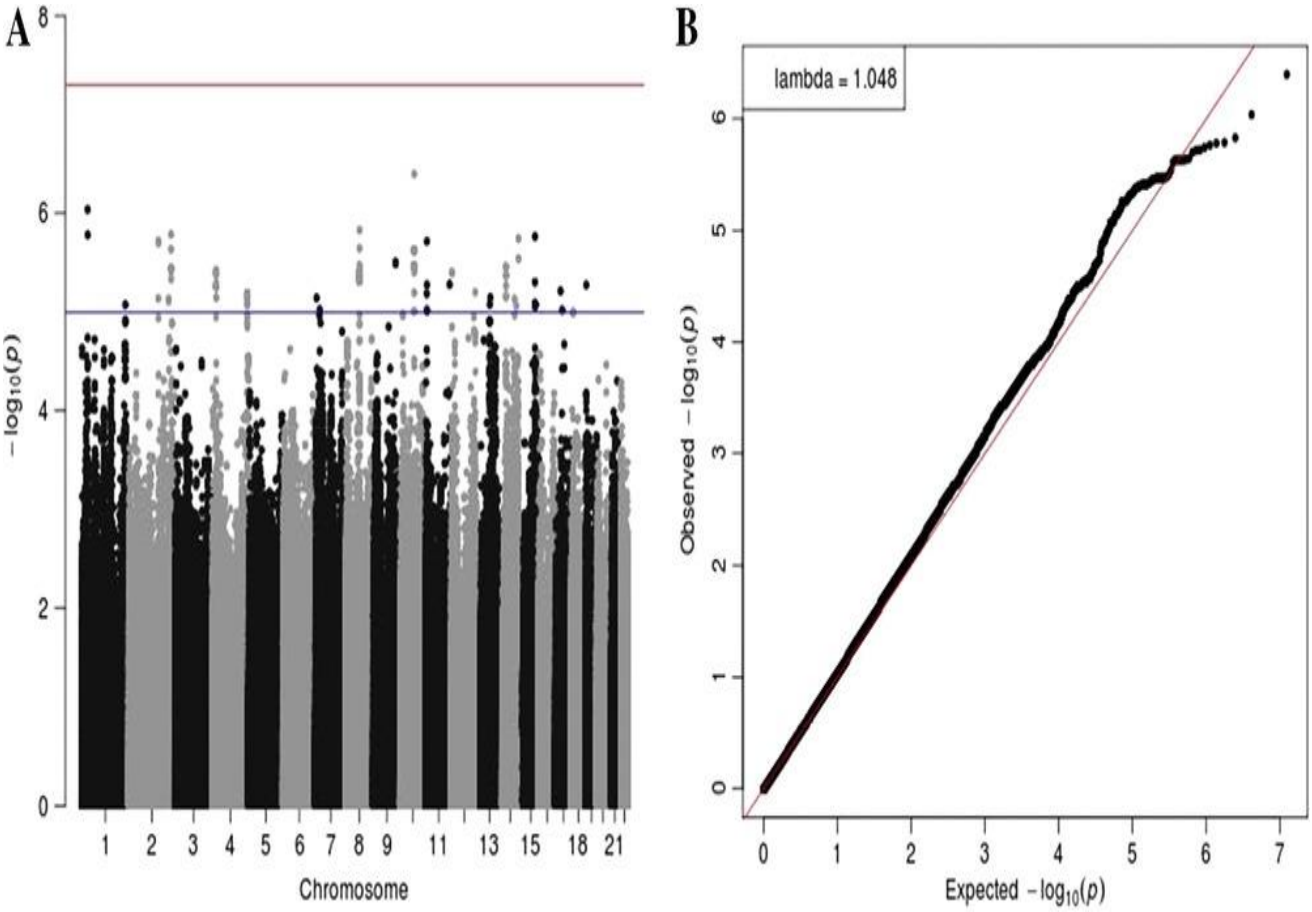


Figure 1

Manhattan plot (A) and q-q plot (B) of results from a genome-wide association scan of TFPI plasma levels in the F5L Family Study. Included in the Manhattan plot are the suggestive (blue; *P*-value < 1×10^{-5}) and genome-wide (red; *P*-value < 5×10^{-8}) thresholds for statistical significance

We defined hemostasis pathway genes using Reactome (Croft et al., 2011), which is a curated and peer-reviewed pathways database with 512 proteins annotated to the "Hemostasis" pathway (R-HSA-109582.1; 2015-03-17). We mapped these proteins to 495 genes in Ensembl and defined gene regions according to the genes' position in hg19 ± 100 kb. The 495 Hemostasis genes covered 278,493 SNPs (4.5% of total), but did not contain more TFPI plasma level associated SNPs than expected by chance (global hypothesis test *P*-value = 0.147; Fig. 2A).

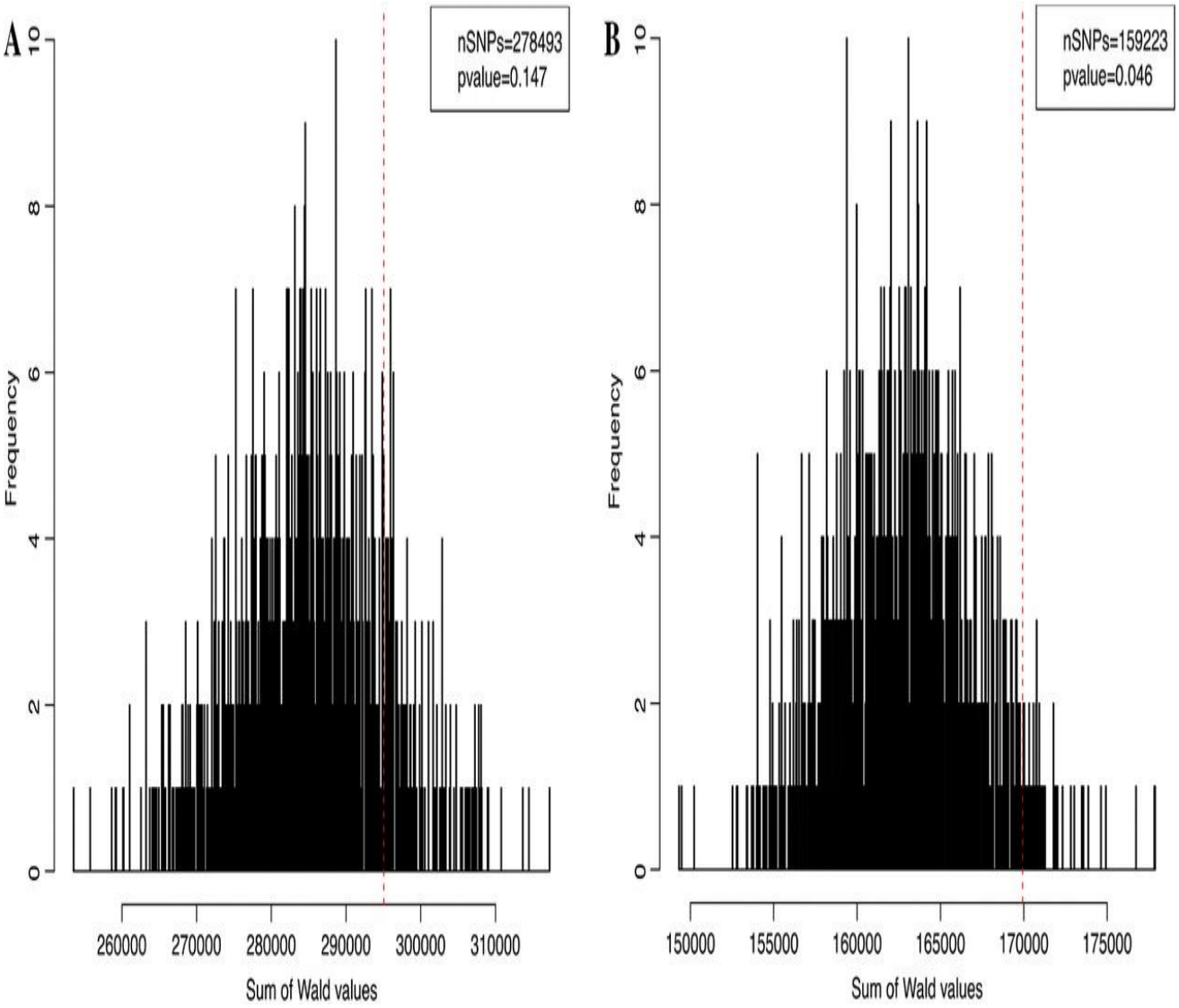


Figure 2

Global hypothesis test results for genotyped and imputed SNPs. We tested SNPs in (A) Reactome "Hemostasis" genes, and (B) H3K27ac-bound regions in vascular endothelial cells for association with TFPI plasma levels in the F5L Family Study and compared the sum of observed Wald values (red dotted line) to the distribution of test statistics possible when the null hypothesis of no association was true

To identify putative vascular EC enhancers/promoters, we focused on gene regulatory regions that were common to both arterial and venous EC. These regions are more likely to harbor sequence variation that is relevant to EC processes, including TFPI plasma level regulation.

We therefore retained all HAEC peaks that overlapped (by at least 1 bp) an H3K27ac peak in human umbilical vein EC (HUVEC) from the ENCODE consortium (ENCFF000BSI, ENCFF000BSD, ENCFF000BRV) (The ENCODE Project Consortium, 2012). Of the 58,408 regions we identified bound by H3K27ac in HAEC, 52,164 (89%) overlapped an H3K27ac-bound region in HUVEC, including the transcription start site of *TFPI* (Fig. 3). The 52,164 H3K27ac-bound regions in vascular EC included 159,223 genotyped and imputed SNPs (2.6% of total), and the global hypothesis test revealed that these regions were enriched for TFPI plasma level associated SNPs (*P*-value = 0.046; Fig. 2B).

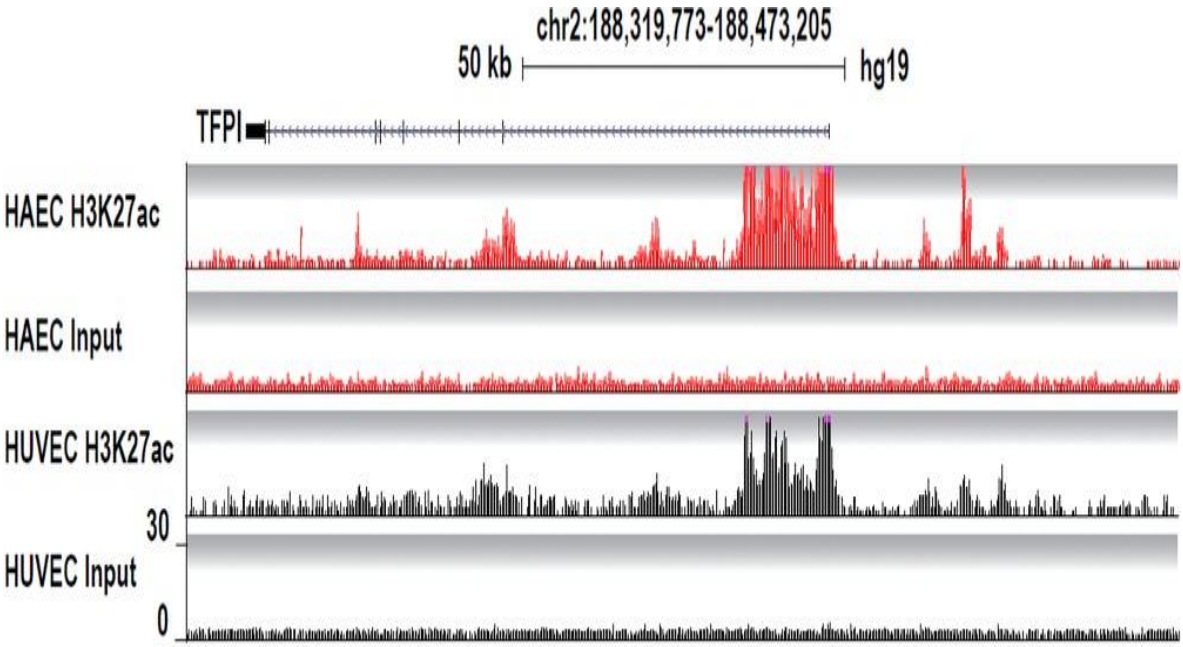


Figure 3

H3K27ac enrichment around *TFPI* in human aortic endothelial cells (HAEC: red; this study) and human umbilical vein endothelial cells (HUVEC; black; ENCODE consortium). Sequence enrichments obtained from DNA isolated from nonimmunoprecipitated whole cell extract are plotted below each H3K27ac experiment. The total size of the genomic window shown (*x*-axis) is 154 kb with *TFPI* on the minus strand. The *y*-axis represents number of sequencing reads and has been arbitrarily capped at 30 reads. Both HAEC and HUVEC H3K27ac ChIP-seq experiments produced peaks at the transcription start site of TFPI that exceeded 30 reads and these regions are indicated with pink lines above the red and black peaks

We additionally conducted two negative control hypothesis tests. In the first, we reasoned that SNPs associated with TFPI plasma levels would *not* be enriched in active regulatory regions of melanocytes, a cell type unrelated to thrombosis and with minimal *TFPI* expression. We identified 95,164 melanocyte regulatory regions from H3K27ac peaks in Roadmap Epigenomics melanocytes (GSM1127072 and GSM1127073; input controls GSM941740, GSM941741; supplementary Table S1). Most melanocyte regions (*N* = 74,736; 79%) did not overlap (by at least 1 bp) one of the 52,164 vascular EC regions, and these 74,736 melanocyte-specific regions were used in subsequent analyses. As expected, H3K27ac-bound regions in melanocytes did not contain an excess of TFPI plasma level associated SNPs beyond what was expected by chance (global hypothesis test *P*-value = 0.339).

In the second negative control hypothesis test, we tested whether the 52,164 vascular EC regulatory regions were enriched for SNPs associated with age, a phenotype with no genetic liability. The global hypothesis test *P*-value was 0.538.

Nearly 10% (*N* = 4,875) of the 52,164 vascular EC regions overlapped a hemostasis pathway gene by at least 1 bp. These overlapping regions included 14,819 SNPs, but subsetting to these regions did not better define the associated regulatory regions than all 52,164 vascular EC regions (subsetting *P*-value = 0.537).

Sixty two percent of the 52,164 vascular EC regions (*N* = 32,311) did not overlap (by at least 1 bp) a melanocyte region. Nonetheless, these vascular EC-specific regions (containing 79,232 SNPs) did not additionally enrich for TFPI plasma level associated SNPs compared to all vascular EC regions (subsetting *P*-value = 0.238).

Because our primary hypothesis yielded the most promising global hypothesis test results, we proceeded with an HD-GWAS of TFPI plasma levels, prioritizing SNPs in H3K27ac regions in vascular EC, and computing sFDR and wFDR *q*-values. No SNPs, however, reached our threshold for genome-wide statistical significance. The minimum sFDR *q*-value was 0.27, and only two SNPs (rs55778554 and rs7904454, with pairwise r^2 = 0.92 in the 1000 Genomes reference population) had sFDR *q*-values < 0.30. The SNPs did not replicate in MARTHA (Table 2). The sFDR *q*-value was consistently smaller than the wFDR *q*-value, implying that the MACS2 peak *P*-values added little information. Association evidence for these SNPs was unchanged after applying the empirical quantile normal transformation to TFPI plasma levels and under the quantile regression models (supplementary Table S2). Associations were still null in MARTHA when anticoagulant and/or antiplatelet users were excluded.

Table 2. Association Results for SNPs with sFDR *q*-values <0.30 in the F5L Family Study

				F5L Family Study									MARTHA		
SNP	**CHR**	**Position (hg19)**	**Alleles [a]**	**MAF**	**Effect [b] (SE)**	***P***	**FDR *q***	**wFDR *q***	**sFDR *q***	**FDR Rank**	**wFDR Rank**	**sFDR Rank**	**MAF**	**Effect (SE)**	***P***
rs55778554	10	73619685	A/G	0.12	−0.23 (0.05)	2.34×10^{-6}	0.391	0.386	0.270	13	13	1	0.13	−0.014 (0.025)	0.568
rs7904454 [c]	10	73631361	C/A	0.10	−0.25 (0.05)	3.39×10^{-6}	0.391	0.386	0.270	24	24	2			

- MAF, minor allele frequency; SE, standard error; FDR, false discovery rate.
- a Allele format is major/minor.
- b Effect is relative to the minor allele.
- c rs7904454 (pairwise r^2 = 0.92 with rs55778554 in the 1000 Genomes reference population) was not imputed in the MARTHA study sample.

To minimize the potential bias in FDR estimates introduced by including correlated (imputed) SNPs, we restricted our HD-GWAS to genotyped SNPs (N = 77,374). We prioritized SNPs in vascular EC regulatory regions, as well as those in LD with a SNP in a vascular EC regulatory region. The global hypothesis P-value for these SNPs was 0.106, but no SNPs had sFDR q-values < 0.10. The minimum sFDR q-value was 0.13, and only two SNPs (rs4746097 and rs4237333, also in LD with rs55778554 and rs7904454) had sFDR q-values < 0.15.

5 DISCUSSION

We conducted the first reported GWAS of TFPI plasma levels, and incorporated experimental epigenetic data for mapped active enhancers and promoters in vascular EC, which were marginally enriched for TFPI plasma level associated SNPs. In contrast, genes annotated to the hemostasis pathway did not contain more TFPI plasma level associated SNPs than expected by chance. Prioritization of SNPs in vascular EC regulatory regions, however, did not identify individual genome-wide significant variants.

Previous candidate gene studies of TFPI plasma levels have focused on hemostatic- and lipid-related genes (reviewed in Dennis et al., 2015). Increasingly, however, genes with no previously known roles in clotting are being implicated in VTE (Antoni et al., 2010), arterial thrombosis, and thrombosis-related phenotypes (Antoni et al., 2010; Morange et al., 2010; Rocanin-Arjo et al., 2014). A recent meta-analysis of over 7,000 VTE cases and their controls identified two genes never associated with cardiovascular phenotypes (Germain et al., 2015). Likewise, the 9p21 locus has been robustly implicated in coronary artery disease, but the biological mechanism and gene underlying the association remains elusive (Miller, Assimes, Montgomery, & Quertermous, 2014). The null results from our hemostatic pathway hypothesis align with these large thrombosis studies. Active enhancers and promoters in vascular EC showed the most promising global enrichment results, and although these results need to be replicated in an independent sample of healthy individuals, our analysis suggests that a cell type specific regulatory region prioritization approach is relevant to TFPI biology.

Detecting causal SNPs is a persistent challenge to GWAS. Although many successes are reported (Welter et al., 2014), there are likely SNPs that are truly trait associated, but whose effect sizes and P-values fall below the multiple testing significance threshold (Schork et al., 2013). Indeed, the proportion of trait heritability explained by all GWAS SNPs combined is much greater than the proportion explained by top GWAS hits (Gusev et al., 2014; Yang et al., 2010, 2011), and meta-analyses have identified SNPs that were only suggestive associations in individual studies (Germain et al., 2015).

These challenges have prompted the development of methods that improve discovery in GWAS by incorporating prior information (Carbonetto & Stephens, 2013; Darnell, Duong, Han, & Eskin, 2012; Nicolae et al., 2010; Zablocki et al., 2014). We used the sFDR (Sun et al., 2006) and wFDR (Roeder et al., 2006), which were among the first methods proposed. Both the sFDR and wFDR are flexible, easy to apply and interpret, and have been validated with a variety of priors, including linkage scores (Roeder et al., 2006; Sun et al., 2006; Wright et al., 2011), MAF (Xing, Cohen, & Boerwinkle, 2010), SNPs in disease-relevant gene pathways (Sun et al., 2012), and SNPs with pleiotropic effects on related phenotypes (Andreassen et al., 2014). The sFDR and wFDR have also been successfully used to prioritize putative regulatory variants, namely, eQTL (Li et al., 2013), and SNPs in regulatory genic categories (e.g., 5′ and 3′ UTR) (Schork et al., 2013).

Machine learning algorithms are another class of methods for integrating GWAS and gene regulatory data, and have been heavily investigated in recent years (Gagliano, Barnes, Weale, & Knight, 2014; Iversen, Lipton, Clyde, & Monteiro, 2014; Kircher et al., 2014; Pickrell, 2014; Ritchie, Dunham, Zeggini, & Flicek, 2014). These methods use massive amounts of functional data (e.g., multiple cell types and functional annotations from ENCODE; The ENCODE Project Consortium, 2012) to predict the most relevant functional class of variants for a given phenotype, and then to prioritize SNPs in these functional categories. Power gains with machine learning algorithms, however, are mixed (Gagliano, Ravji, Barnes, Weale, & Knight, 2015; Ritchie, Holzinger, Li, Pendergrass, & Kim, 2015). Whether this is a function of the input data or of underlying biology is unclear, as few machine learning algorithms explicitly modeled cell-type specificity.

Alternatively, the genome may contain many redundancies, and gene regulatory data may still be too nonspecific to use as informative priors. Using epigenetic data as priors will most likely benefit from knowing which tissue(s), and more specifically which cell types within specific tissues, play a role in the biology of the trait measured by GWAS. For complex traits especially, this question is not trivial. Although *TFPI* is constitutively expressed by vascular ECs (Bajaj et al., 1990; Lupu et al., 1995, 1999), other physiologically relevant sources exist including megakaryocytes (Novotny et al., 1988), monocytes (Bajaj, Ghosh, & Bajaj, 2007; Basavaraj et al., 2010), cardiomyocytes, vascular smooth muscle cells, and fibroblasts (Bajaj, Steer, Kuppuswamy, Kisiel, & Bajaj, 1999; Caplice et al., 1998; Kereveur, Enjyoji, Masuda, Yutani, & Kato, 2001). Moreover, expression changes during inflammation add additional complexities that could be relevant to include as biological priors.

Clinical differences between the F5L Family and MARTHA study samples may have also contributed to the lack of individual SNP replications. The F5L Family Study included mostly nondiseased individuals, whereas the MARTHA study included only VTE cases. Although TFPI plasma levels did not differ by VTE status in the F5L Family Study, others have reported that TFPI plasma levels are elevated immediately following a VTE event (Sidelmann, Bladbjerg, Gram, Munster, & Jespersen, 2008), and the cell types and SNPs involved in this pathological TFPI expression may differ from those that regulate endogenous TFPI production (Bjorkegren, Kovacic, Dudley, & Schadt, 2015). HAEC regulatory regions were ascertained in cells from healthy individuals, so the F5L Family Study was ideally suited for testing our global hypotheses, but our lack of individual SNP replication in MARTHA suggests systemic differences in gene regulation post-VTE.

In the first GWAS of TFPI plasma levels, we report a slight enrichment of TFPI plasma level associated SNPs in the active enhancers and promoters of vascular ECs, the predominant cell type to express *TFPI*. We leveraged the observed enrichment in our GWAS, prioritizing SNPs in vascular EC regulatory regions via sFDR control, but did not identify any individual SNPs associated with TFPI plasma levels. The overrepresentation of trait-associated SNPs in cell type specific regulatory regions is now well documented, but methods that efficiently incorporate functional epigenomic data into GWAS are lacking. Our approach provides a framework for integrating genetic and epigenetic data, and may be useful in larger study samples.

ACKNOWLEDGMENTS

This work was supported by the Canadian Institutes of Health Research (Grant MOP 86466) and by the Heart and Stroke Foundation of Canada (Grant T6484) to F.G. and Canadian

Institutes of Health Research Catalyst grant to F.G. and M.D.W. (CEN-132382). J.D. was supported by a Vanier Canada Graduate Scholarship and the Canadian Institutes of Health Research Strategic Training for Advanced Genetic Epidemiology (CIHR STAGE) program. F.G. and M.D.W. were supported by Canada Research Chairs. M.D.W., is supported by an Early Researcher Award from the Ontario Ministry of Research, Innovation and Science, and L.A., and A.M.-R. were supported by Heart and Stroke Foundation of Ontario (Bridge Grant 7486). L.A. was supported by Alexander Graham Bell Canada Graduate Scholarship. A.M.R. was supported by CIHR STAGE and fellowships. The A.M.R. laboratory was supported by a CONACYT Infrastructure grant (269449).

CONFLICT OF INTEREST

The authors have no conflicts of interest to declare

REFERENCES

- Abney, M. (2015). Permutation testing in the presence of polygenic variation. *Genetic Epidemiology*, 39(4), 249–258.
- Aird, W. C. (2012). Endothelial cell heterogeneity. *Cold Spring Harbor Perspectives in Medicine*, 2(1), a006429, 1–13.
- Almasy, L., Soria, J. M., Souto, J. C., Warren, D. M., Buil, A., Borrell, M., … Blangero, J. (2005). A locus on chromosome 2 influences levels of tissue factor pathway inhibitor: Results from the GAIT study. *Arteriosclerosis, Thrombosis, and Vascular Biology*, 25(7), 1489–1492.
- Andreassen, O. A., McEvoy, L. K., Thompson, W. K., Wang, Y., Reppe, S., Schork, A. J., … Dale, A. M. (2014). Identifying common genetic variants in blood pressure due to polygenic pleiotropy with associated phenotypes. *Hypertension*, 63(4), 819–826.
- Antoni, G., Morange, P. E., Luo, Y., Saut, N., Burgos, G., Heath, S., … Gagnon, F. (2010). A multi-stage multi-design strategy provides strong evidence that the BAI3 locus is associated with early-onset venous thromboembolism. *Journal of Thrombosis and Haemostasis*, 8(12), 2671–2679.
- Antoni, G., Oudot-Mellakh, T., Dimitromanolakis, A., Germain, M., Cohen, W., Wells, P., … Tregouet, D. A. (2011). Combined analysis of three genome-wide association studies on vWF and FVIII plasma levels. *BMC Medical Genetics*, 12, 102, 1–11.
- Bajaj, M. S., Ghosh, M., & Bajaj, S. P. (2007). Fibronectin-adherent monocytes express tissue factor and tissue factor pathway inhibitor whereas endotoxin-stimulated monocytes primarily express tissue factor: Physiologic and pathologic implications. *Journal of Thrombosis and Haemostasis*, 5(7), 1493–1499.
- Bajaj, M. S., Kuppuswamy, M. N., Saito, H., Spitzer, S. G., & Bajaj, S. P. (1990). Cultured normal human hepatocytes do not synthesize lipoprotein-associated coagulation inhibitor: Evidence that endothelium is the principal site of its synthesis. *Proceedings of the National Academy of Sciences of the United States of America*, 87(22), 8869–8873.
- Bajaj, M. S., Steer, S., Kuppuswamy, M. N., Kisiel, W., & Bajaj, S. P. (1999). Synthesis and expression of tissue factor pathway inhibitor by serum-stimulated fibroblasts, vascular smooth muscle cells and cardiac myocytes. *Thrombosis and Haemostasis*, 82(6), 1663–1672.

- Ballester, B., Medina-Rivera, A., Schmidt, D., Gonzalez-Porta, M., Carlucci, M., Chen, X., … Wilson, M. D. (2014). Multi-species, multi-transcription factor binding highlights conserved control of tissue-specific biological pathways. *Elife*, 3, e02626, 1–29.
- Basavaraj, M. G., Gruber, F. X., Sovershaev, M., Appelbom, H. I., Osterud, B., Petersen, L. C., & Hansen, J. B. (2010). The role of TFPI in regulation of TF-induced thrombogenicity on the surface of human monocytes. *Thrombosis Research*, 126(5), 418–425.
- Bjorkegren, J. L., Kovacic, J. C., Dudley, J. T., & Schadt, E. E. (2015). Genome-wide significant loci: How important are they? Systems genetics to understand heritability of coronary artery disease and other common complex disorders. *Journal of the American College of Cardiology*, 65(8), 830–845.
- Bladbjerg, E. M., de Maat, M. P., Christensen, K., Bathum, L., Jespersen, J., & Hjelmborg, J. (2006). Genetic influence on thrombotic risk markers in the elderly—A Danish twin study. *Journal of Thrombosis and Haemostasis*, 4(3), 599–607.
- Caplice, N. M., Mueske, C. S., Kleppe, L. S., Peterson, T. E., Broze, G. J., & Simari, R. D. (1998). Expression of tissue factor pathway inhibitor in vascular smooth muscle cells and its regulation by growth factors. *Circulation Research*, 83(12), 1264–1270.
- Carbonetto, P., & Stephens, M. (2013). Integrated enrichment analysis of variants and pathways in genome-wide association studies indicates central role for IL-2 signaling genes in type 1 diabetes, and cytokine signaling genes in Crohn's disease. *PLoS Genetics*, 9(10), e1003770, 1–19.
- Craiu, R. V., & Sun, L. (2008). Choosing the lesser evil: Trade-off between false discovery rate and non-discovery rate. *Statistica Sinica*, 18, 861–869.
- Croft, D., O'Kelly, G., Wu, G., Haw, R., Gillespie, M., Matthews, L., … Stein, L. (2011). Reactome: A database of reactions, pathways and biological processes. *Nucleic Acids Research*, 39(Database issue), D691–D697.
- Darnell, G., Duong, D., Han, B., & Eskin, E. (2012). Incorporating prior information into association studies. *Bioinformatics*, 28(12), i147–i153.
- Delaneau, O., Marchini, J., & The 1000 Genomes Project Consortium. (2014). Integrating sequence and array data to create an improved 1000 Genomes Project haplotype reference panel. *Nature Communications*, 5, 3934.
- Dennis, J., Kassam, I., Morange, P. E., Tregouet, D. A., & Gagnon, F. (2015). Genetic determinants of tissue factor pathway inhibitor plasma levels. *Thrombosis and Haemostasis*, 114(2), 245–257.
- Dennis, J., Truong, V., Aissi, D., Medina-Rivera, A., Blankenberg, S., Germain, M., … Gagnon, F. (2016). Single nucleotide polymorphisms in an intergenic chromosome 2q region associated with tissue factor pathway inhibitor plasma levels and venous thromboembolism. *Journal of Thrombosis and Haemostasis*, 14(10), 1960–1970.
- Dimas, A. S., Deutsch, S., Stranger, B. E., Montgomery, S. B., Borel, C., Attar-Cohen, H., … Antonarakis, E. (2009). Common regulatory variation impacts gene expression in a cell type-dependent manner. *Science*, 325(5945), 1246–1250.
- Ernst, J., Kheradpour, P., Mikkelsen, T. S., Shoresh, N., Ward, L. D., Epstein, C. B., … Bernstein, B. E. (2011). Mapping and analysis of chromatin state dynamics in nine human cell types. *Nature*, 473(7345), 43–49.
- Gagliano, S. A., Barnes, M. R., Weale, M. E., & Knight, J. (2014). A Bayesian method to incorporate hundreds of functional characteristics with association evidence to improve variant prioritization. *PLoS One*, 9(5), e98122, 1–14.

- Gagliano, S. A., Ravji, R., Barnes, M. R., Weale, M. E., & Knight, J. (2015). Smoking gun or circumstantial evidence? Comparison of statistical learning methods using functional annotations for prioritizing risk variants. *Scientific Reports*, 5, 13373, 1–11.
- Gagnon, F., Aissi, D., Carrie, A., Morange, P. E., & Tregouet, D. A. (2014). Robust validation of methylation levels association at CPT1A locus with lipid plasma levels. *Journal of Lipid Research*, 55(7), 1189–1191.
- GBD 2013 Mortality and Causes of Death Collaborators. (2015). Global, regional, and national age–sex specific all-cause and cause-specific mortality for 240 causes of death, 1990–2013: A systematic analysis for the Global Burden of Disease Study 2013. *Lancet*, 385(9963), 117–171.
- Geraci, M. (2014). Linear quantile mixed models: The lqmm package for Laplace quantile regression. *Journal of Statistical Software*, 57(13), 1–29.
- Germain, M., Chasman, D. I., de Haan, H., Tang, W., Lindstrom, S., Weng, L. C., … Morange, P. E. (2015). Meta-analysis of 65,734 individuals identifies TSPAN15 and SLC44A2 as two susceptibility loci for venous thromboembolism. *American Journal of Human Genetics*, 96(4), 532–542.
- Goldhaber, S. Z., & Bounameaux, H. (2012). Pulmonary embolism and deep vein thrombosis. *Lancet*, 379(9828), 1835–1846.
- Gusev, A., Lee, S. H., Trynka, G., Finucane, H., Vilhjalmsson, B. J., Xu, H., … Price, A. L. (2014). Partitioning heritability of regulatory and cell-type-specific variants across 11 common diseases. *American Journal of Human Genetics*, 95(5), 535–552.
- Howie, B., Fuchsberger, C., Stephens, M., Marchini, J., & Abecasis, G. R. (2012). Fast and accurate genotype imputation in genome-wide association studies through pre-phasing. *Nature Genetics*, 44(8), 955–959.
- Howie, B. N., Donnelly, P., & Marchini, J. (2009). A flexible and accurate genotype imputation method for the next generation of genome-wide association studies. *PLoS Genetics*, 5(6), e1000529, 1–15.
- Iversen, E. S., Lipton, G., Clyde, M. A., & Monteiro, A. N. (2014). Functional annotation signatures of disease susceptibility loci improve SNP association analysis. *BMC Genomics*, 15, 398, 1–16.
- Kereveur, A., Enjyoji, K., Masuda, K., Yutani, C., & Kato, H. (2001). Production of tissue factor pathway inhibitor in cardiomyocytes and its upregulation by interleukin-1. *Thrombosis and Haemostasis*, 86(5), 1314–1319.
- Kircher, M., Witten, D. M., Jain, P., O'Roak, B. J., Cooper, G. M., & Shendure, J. (2014). A general framework for estimating the relative pathogenicity of human genetic variants. *Nature Genetics*, 46(3), 310–315.
- Li, L., Kabesch, M., Bouzigon, E., Demenais, F., Farrall, M., Moffatt, M. F., … Liang, L. (2013). Using eQTL weights to improve power for genome-wide association studies: A genetic study of childhood asthma. *Frontiers in Genetics*, 4, 103, 1–14.
- Li, Y., Willer, C., Sanna, S., & Abecasis, G. (2009). Genotype imputation. *Annual Review of Genomics and Human Genetics*, 10, 387–406.
- Li, Y., Willer, C. J., Ding, J., Scheet, P., & Abecasis, G. R. (2010). MaCH: Using sequence and genotype data to estimate haplotypes and unobserved genotypes. *Genetic Epidemiology*, 34(8), 816–834.
- Lupu, C., Lupu, F., Dennehy, U., Kakkar, V. V., & Scully, M. F. (1995). Thrombin induces the redistribution and acute release of tissue factor pathway inhibitor from specific granules within human endothelial cells in culture. *Arteriosclerosis, Thrombosis, and Vascular Biology*, 15(11), 2055–2062.
- Lupu, C., Poulsen, E., Roquefeuil, S., Westmuckett, A. D., Kakkar, V. V., & Lupu, F. (1999). Cellular effects of heparin on the production and release of tissue factor

pathway inhibitor in human endothelial cells in culture. *Arteriosclerosis, Thrombosis, and Vascular Biology*, 19(9), 2251–2262.

- Lupu, C., Zhu, H., Popescu, N. I., Wren, J. D., & Lupu, F. (2011). Novel protein ADTRP regulates TFPI expression and function in human endothelial cells in normal conditions and in response to androgen. *Blood*, 118(16), 4463–4471.
- Maurano, M. T., Humbert, R., Rynes, E., Thurman, R. E., Haugen, E., Wang, H., … Stamatoyannopoulos, J. A. (2012). Systematic localization of common disease-associated variation in regulatory DNA. *Science*, 337(6099), 1190–1195.
- Mersha, T. B., Ding, L., He, H., Alexander, E. S., Zhang, X., Kurowski, B. G., … Fardo, D. W. (2015). Impact of population stratification on family-based association in an admixed population. *International Journal of Genomics*, 2015, 501617, 1–5.
- Miller, C. L., Assimes, T. L., Montgomery, S. B., & Quertermous, T. (2014). Dissecting the causal genetic mechanisms of coronary heart disease. *Current Atherosclerosis Reports*, 16(5), 406, 1–12.
- Morange, P. E., Bezemer, I., Saut, N., Bare, L., Burgos, G., Brocheton, J., … Rosendaal, F. R. (2010). A follow-up study of a genome-wide association scan identifies a susceptibility locus for venous thrombosis on chromosome 6p24.1. *American Journal of Human Genetics*, 86(4), 592–595.
- Morange, P. E., Blankenberg, S., Alessi, M. C., Bickel, C., Rupprecht, H. J., Schnabel, R., … Tiret, L. (2007). Prognostic value of plasma tissue factor and tissue factor pathway inhibitor for cardiovascular death in patients with coronary artery disease: The AtheroGene study. *Journal of Thrombosis and Haemostasis*, 5(3), 475–482.
- Morange, P. E., Simon, C., Alessi, M. C., Luc, G., Arveiler, D., Ferrieres, J., … Juhan-Vague, I. (2004). Endothelial cell markers and the risk of coronary heart disease: The Prospective Epidemiological Study of Myocardial Infarction (PRIME) study. *Circulation*, 109(11), 1343–1348.
- Morange, P. E., Suchon, P., & Tregouet, D. A. (2015). Genetics of venous thrombosis: Update in 2015. *Thrombosis and Haemostasis*, 114(1), 1–10.
- Nicolae, D. L., Gamazon, E., Zhang, W., Duan, S., Dolan, M. E., & Cox, N. J. (2010). Trait-associated SNPs are more likely to be eQTLs: Annotation to enhance discovery from GWAS. *PLoS Genetics*, 6(4), e1000888, 1–10.
- Novotny, W. F., Girard, T. J., Miletich, J. P., & Broze, G. J., Jr. (1988). Platelets secrete a coagulation inhibitor functionally and antigenically similar to the lipoprotein associated coagulation inhibitor. *Blood*, 72(6), 2020–2025.
- Novotny, W. F., Girard, T. J., Miletich, J. P., & Broze, G. J., Jr. (1989). Purification and characterization of the lipoprotein-associated coagulation inhibitor from human plasma. *Journal of Biological Chemistry*, 264(31), 18832–18837.
- O'Connell, J., Gurdasani, D., Delaneau, O., Pirastu, N., Ulivi, S., Cocca, M., … Marchini, J. (2014). A general approach for haplotype phasing across the full spectrum of relatedness. *PLoS Genetics*, 10(4), e1004234, 1–21.
- Oudot-Mellakh, T., Cohen, W., Germain, M., Saut, N., Kallel, C., Zelenika, D., … Morange, P. E. (2012). Genome wide association study for plasma levels of natural anticoagulant inhibitors and protein C anticoagulant pathway: The MARTHA project. *British Journal of Haematology*, 157(2), 230–239.
- Peng, B., Yu, R. K., Dehoff, K. L., & Amos, C. I. (2007). Normalizing a large number of quantitative traits using empirical normal quantile transformation. *BMC Proceedings*, 1(Suppl. 1), S156.
- Pickrell, J. K. (2014). Joint analysis of functional genomic data and genome-wide association studies of 18 human traits. *American Journal of Human Genetics*, 94(4), 559–573.

- Ritchie, G. R., Dunham, I., Zeggini, E., & Flicek, P. (2014). Functional annotation of noncoding sequence variants. *Nature Methods*, 11(3), 294–296.
- D. (2015). Methods of integrating data to uncover genotype-phenotype interactions. *Nature Reviews Genetics*, 16(2), 85–97.
- Rocanin-Arjo, A., Cohen, W., Carcaillon, L., Frere, C., Saut, N., Letenneur, L., … The CardioGenics Consortium. (2014). A meta-analysis of genome-wide association studies identifies ORM1 as a novel gene controlling thrombin generation potential. *Blood*, 123(5), 777–785.
- Roeder, K., Bacanu, S. A., Wasserman, L., & Devlin, B. (2006). Using linkage genome scans to improve power of association in genome scans. *American Journal of Human Genetics*, 78(2), 243–252.
- Sanchez-Solana, B., Motwani, M., Li, D. Q., Eswaran, J., & Kumar, R. (2012). p21-Activated kinase-1 signaling regulates transcription of tissue factor and tissue factor pathway inhibitor. *Journal of Biological Chemistry*, 287(47), 39291–39302.
- Schmidt, D., Wilson, M. D., Spyrou, C., Brown, G. D., Hadfield, J., & Odom, D. T. (2009). ChIP-seq: using high-throughput sequencing to discover protein-DNA interactions. *Methods*, 48(3), 240–248.
- Schork, A. J., Thompson, W. K., Pham, P., Torkamani, A., Roddey, J. C., Sullivan, P. F., … Dale, A. M. (2013). All SNPs are not created equal: Genome-wide association studies reveal a consistent pattern of enrichment among functionally annotated SNPs. *PLoS Genetics*, 9(4), e1003449, 1–13.
- Sidelmann, J. J., Bladbjerg, E. M., Gram, J., Munster, A. M., & Jespersen, J. (2008). Tissue factor pathway inhibitor relates to fibrin degradation in patients with acute deep venous thrombosis. *Blood Coagulation & Fibrinolysis*, 19(5), 405–409.
- Soave, D., Corvol, H., Panjwani, N., Gong, J., Li, W., Boelle, P. Y., … Sun, L. (2015). A joint location-scale test improves power to detect associated SNPs, gene sets, and pathways. *American Journal of Human Genetics*, 97(1), 125–138.
- Sun, L., Craiu, R. V., Paterson, A. D., & Bull, S. B. (2006). Stratified false discovery control for large-scale hypothesis testing with application to genome-wide association studies. *Genetic Epidemiology*, 30(6), 519–530.
- Sun, L., Rommens, J. M., Corvol, H., Li, W., Li, X., Chiang, T. A., … Strug, L. J. (2012). Multiple apical plasma membrane constituents are associated with susceptibility to meconium ileus in individuals with cystic fibrosis. *Nature Genetics*, 44(5), 562–569.
- The ENCODE Project Consortium. (2012). An integrated encyclopedia of DNA elements in the human genome. *Nature*, 489(7414), 57–74.
- Verhamme, P., & Hoylaerts, M. F. (2006). The pivotal role of the endothelium in haemostasis and thrombosis. *Acta Clinica Belgica*, 61(5), 213–219.
- Warren, D. M., Soria, J. M., Souto, J. C., Comuzzie, A. G., Fontcuberta, J., Blangero, J., … Almasy, L. (2005). Heritability of hemostasis phenotypes and their correlation with type 2 diabetes status in Mexican Americans. *Human Biology*, 77(1), 1–15.
- Welter, D., MacArthur, J., Morales, J., Burdett, T., Hall, P., Junkins, H., … Parkinson, H. (2014). The NHGRI GWAS Catalog, a curated resource of SNP-trait associations. *Nucleic Acids Research*, 42(Database issue), D1001–D1006, 1–6.
- White, M. J., Kodaman, N. M., Harder, R. H., Asselbergs, F. W., Vaughan, D. E., Brown, N. J., … Williams, S. M. (2015). Genetics of plasminogen activator inhibitor-1 (PAI-1) in a Ghanaian population. *PLoS One*, 10(8), e0136379, 1–14.
- Wijsman, E. M. (2012). The role of large pedigrees in an era of high-throughput sequencing. *Human Genetics*, 131(10), 1555–1563.

- Wood, J. P., Bunce, M. W., Maroney, S. A., Tracy, P. B., Camire, R. M., & Mast, A. E. (2013). Tissue factor pathway inhibitor-alpha inhibits prothrombinase during the initiation of blood coagulation. *Proceedings of the National Academy of Sciences of the United States of America*, 110(44), 17838–17843.
- Wright, F. A., Strug, L. J., Doshi, V. K., Commander, C. W., Blackman, S. M., Sun, L., … Cutting, G. R. (2011). Genome-wide association and linkage identify modifier loci of lung disease severity in cystic fibrosis at 11p13 and 20q13.2. *Nature Genetics*, 43(6), 539–546.
- Xing, C., Cohen, J. C., & Boerwinkle, E. (2010). A weighted false discovery rate control procedure reveals alleles at FOXA2 that influence fasting glucose levels. *American Journal of Human Genetics*, 86(3), 440–446.
- Yang, J., Benyamin, B., McEvoy, B. P., Gordon, S., Henders, A. K., Nyholt, D. R., … Visscher, P. M. (2010). Common SNPs explain a large proportion of the heritability for human height. *Nature Genetics*, 42(7), 565–569.
- Yang, J., Manolio, T. A., Pasquale, L. R., Boerwinkle, E., Caporaso, N., Cunningham, J. M., … Visscher, P. M. (2011). Genome partitioning of genetic variation for complex traits using common SNPs. *Nature Genetics*, 43(6), 519–525.
- Yoo, Y. J., Bull, S. B., Paterson, A. D., Waggott, D., The Diabetes Control and Complications Trial/Epidemiology of Diabetes Interventions and Complications Research Group, Sun, L. (2010). Were genome-wide linkage studies a waste of time? Exploiting candidate regions within genome-wide association studies. *Genetics Epidemiology*, 34(2), 107–118.
- Zablocki, R. W., Schork, A. J., Levine, R. A., Andreassen, O. A., Dale, A. M., & Thompson, W. K. (2014). Covariate-modulated local false discovery rate for genome-wide association studies. *Bioinformatics*, 30(15), 2098–2104.
- Zakai, N. A., Lutsey, P. L., Folsom, A. R., Heckbert, S. R., & Cushman, M. (2010). Total tissue factor pathway inhibitor and venous thrombosis. The longitudinal investigation of thromboembolism etiology. *Thrombosis and Haemostasis*, 104(2), 207–212.
- Zhang, Y., Liu, T., Meyer, C. A., Eeckhoute, J., Johnson, D. S., Bernstein, B. E., … Liu, X. S. (2008). Model-based analysis of ChIP-Seq (MACS). *Genome Biology*, 9(9), R137, 1–9.
- Zhou, X., & Stephens, M. (2012). Genome-wide efficient mixed-model analysis for association studies. *Nature Genetics*, 44(7), 821–824.