



**HAL**  
open science

## Sex-specific DNA methylation marks associated with sex-biased risk of recurrence in unprovoked venous thromboembolism

Ohanna C L Bezerra, Marc Rodger, Gaelle Munsch, Michael J Kovacs, Gregoire Le Gal, Pierre E Morange, David-Alexandre Tregouet, Celia M T Greenwood, France Gagnon

### ► To cite this version:

Ohanna C L Bezerra, Marc Rodger, Gaelle Munsch, Michael J Kovacs, Gregoire Le Gal, et al.. Sex-specific DNA methylation marks associated with sex-biased risk of recurrence in unprovoked venous thromboembolism. *Journal of Thrombosis and Haemostasis*, 2025, 10.1016/j.jtha.2025.01.004 . hal-04956649

**HAL Id: hal-04956649**

**<https://hal.science/hal-04956649v1>**

Submitted on 19 Feb 2025

**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.



Distributed under a Creative Commons Attribution 4.0 International License

## ORIGINAL ARTICLE

# Sex-specific DNA methylation marks associated with sex-biased risk of recurrence in unprovoked venous thromboembolism

Ohanna C. L. Bezerra<sup>1</sup>   | Marc Rodger<sup>2</sup> | Gaëlle Munsch<sup>3</sup> | Michael J. Kovacs<sup>4</sup> | Grégoire Le Gal<sup>5</sup> | Pierre-Emmanuel Morange<sup>6</sup> | David-Alexandre Trégouët<sup>3</sup> | Celia M. T. Greenwood<sup>7,8</sup> | France Gagnon<sup>1,9</sup>

<sup>1</sup>Dalla Lana School of Public Health, University of Toronto, Toronto, Ontario, Canada

<sup>2</sup>Department of Medicine, McGill University, Montreal, Quebec, Canada

<sup>3</sup>Bordeaux Population Health Research Center, Institut national de la santé et de la recherche médicale (INSERM), Unité Mixte de Recherche (UMR) 1219, F-33000, University of Bordeaux, Bordeaux, France

<sup>4</sup>Department of Medicine, Western University, London, Ontario, Canada

<sup>5</sup>Department of Medicine, Ottawa Hospital Research Institute, University of Ottawa, Ottawa, Ontario, Canada

<sup>6</sup>Cardiovascular and Nutrition Research Center, Aix-Marseille University, Marseille, France

<sup>7</sup>Lady Davis Institute for Medical Research, Jewish General Hospital, Montreal, Quebec, Canada

<sup>8</sup>Department of Epidemiology, Biostatistics and Occupational Health, McGill University, Montreal, Quebec, Canada

<sup>9</sup>Office of the Vice-Principal of Research and Innovation, University of Toronto Mississauga, Mississauga, Ontario, Canada

## Correspondence

France Gagnon, Office of the Vice-Principal of Research and Innovation, University of Toronto Mississauga, 3359 Mississauga Road, Mississauga, ON L5L1C6, Canada.  
Email: [france.gagnon@utoronto.ca](mailto:france.gagnon@utoronto.ca)

## Abstract

**Background:** Whether to stop oral anticoagulants after a first unprovoked venous thromboembolism (VTE) is challenging, partially due to an intriguingly higher risk of VTE recurrence (rVTE) in men after therapy discontinuation. DNA methylation (DNAm) differences between men and women might underlie this sex-biased rVTE risk difference.

**Objectives:** To investigate sex-specific associations between DNAm at cytosine-phosphate-guanine (CpG) sites and rVTE.

**Methods:** In 417 unprovoked VTE patients, including 101 experiencing recurrences over a 5-year follow-up (REcurrent VEnous thromboembolism Risk Stratification Evaluation [REVERSE] I), we analyzed blood DNAm using the Illumina EPIC array and performed a sex-stratified epigenome-wide association study. We further examined 181 major provoked VTE patients, including 36 recurrences over a 14-year follow-up (the MARseille THrombosis Association [MARTHA]), to investigate whether DNAm is a risk factor for rVTE after anticoagulation therapy.

**Results:** Hypomethylated CpGs at genes *TBC1D22B*-cg01060850 and *ZHX2*-cg07808424 in men and *DIP2B*-ch.12.1038646R and *DENND3*-cg03401656 in women were associated with rVTE at genome-wide level ( $P < 7 \times 10^{-8}$ ). Though not statistically significant, *DENND3*-cg03401656 had the same direction of effect in MARTHA women. Sensitivity analysis confirmed the robustness of the estimates, including potential confounders, adaptations of the Cox model, non-Europeans, and proximal methylation quantitative trait loci in the association. The associated CpGs were situated at genes for membrane trafficking, corroborating the participation of Rab regulatory proteins in rVTE and transcription factors.

**Conclusion:** We identified DNAm marks as potential risk factors for sex-biased recurrence in unprovoked VTE. Further replication and experimental validation could

Manuscript handled by: Karl C. Desch

Final decision: Karl C. Desch, 6 January 2025

© 2025 The Author(s). Published by Elsevier Inc. on behalf of International Society on Thrombosis and Haemostasis. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

**Funding information** This study was supported by a research grant from the Canadian Institutes of Health Research (grant PJT-162233) and a fellowship from the CANSSI (Canadian Statistical Sciences Institute) Ontario STAGE (Strategic Training for Advanced Genetic Epidemiology) program (grant 2021-02-22-50090040-31). The MARseille THrombosis Association (MARTHA) project was supported by a grant from the Program Hospitalier de Recherche Clinique. The H450M epityping in MARTHA was partially funded by the Canadian Institutes of Health Research (grant MOP 86466) and by the Heart and Stroke Foundation of Canada (grant T6484).

refine our understanding of the regulation of the identified DNAm sites and help optimize personalized decision-making for long-term anticoagulation after a first VTE.

#### KEYWORDS

DNA methylation, genome-wide association study, recurrence, sex differences, venous thromboembolism

## 1 | INTRODUCTION

Venous thromboembolism (VTE), which encompasses deep vein thrombosis (DVT) and pulmonary embolism, is a multifactorial, potentially fatal condition with substantial morbidity and mortality [1]. Initial treatment of a first VTE is based on 3 to 6 months of anticoagulant therapy [2]. After discontinuation, a higher risk of recurrence (5%-10% per year) is typically observed in individuals who had an unprovoked VTE or weakly provoked VTE compared with those whose VTE was provoked by a major transient risk factor (eg, major surgery or immobilization) [3,4]. In unprovoked VTE, where there is no clear triggering factor and weakly provoked VTE (eg, travel-associated), the optimal duration of treatment is uncertain and often extended (ie, indefinite) in patients who are not at high bleeding risks. The decision of whether to continue the anticoagulants beyond the short-term course entails balancing the benefits of the therapy and the risk of hemorrhagic complications in the presence or absence of provoking risk factors [2].

Men and women are at similar risk of developing a first VTE. In contrast, recurrent VTE (rVTE) has a consistently higher rate in men than in women after anticoagulant cessation: up to 10% per year in men vs 2% to 5% per year in women [5]. The best-known rVTE risk assessment models are based on clinical predictors [6,7]. The “HERDOO2” decision rule, based on age, obesity, D-dimer, and clinical symptoms, identifies women at low risk of rVTE who can discontinue anticoagulants after completing 5 to 12 months of therapy, but not men (all men are under high risk) [8,9]. Genetic risk factors, including inherited thrombophilia (eg, factor [F]V Leiden [FVL], FII G20210A, protein C and S, and antithrombin deficiency), confer a higher risk for a first VTE [10]. However, while the most common thrombophilia has been included in predictive models (FVL and FII G20210A), it does not appear to predict the risk of recurrence, prompting an investigation into other molecular mechanisms underlying rVTE [11–13].

DNA methylation (DNAm) is an epigenetic mark by which a methyl bonds to cytosines in 5′-C-phosphate-G-3′ (CpG) dinucleotides and regulates gene expression [14]. While DNA sequence is invariant across cell types, DNAm is dynamic and varies over time or in response to exposures (eg, medication) between cell types within an individual and between individuals [15]. DNAm variation is strongly influenced by sex and has emerged as a molecular candidate for

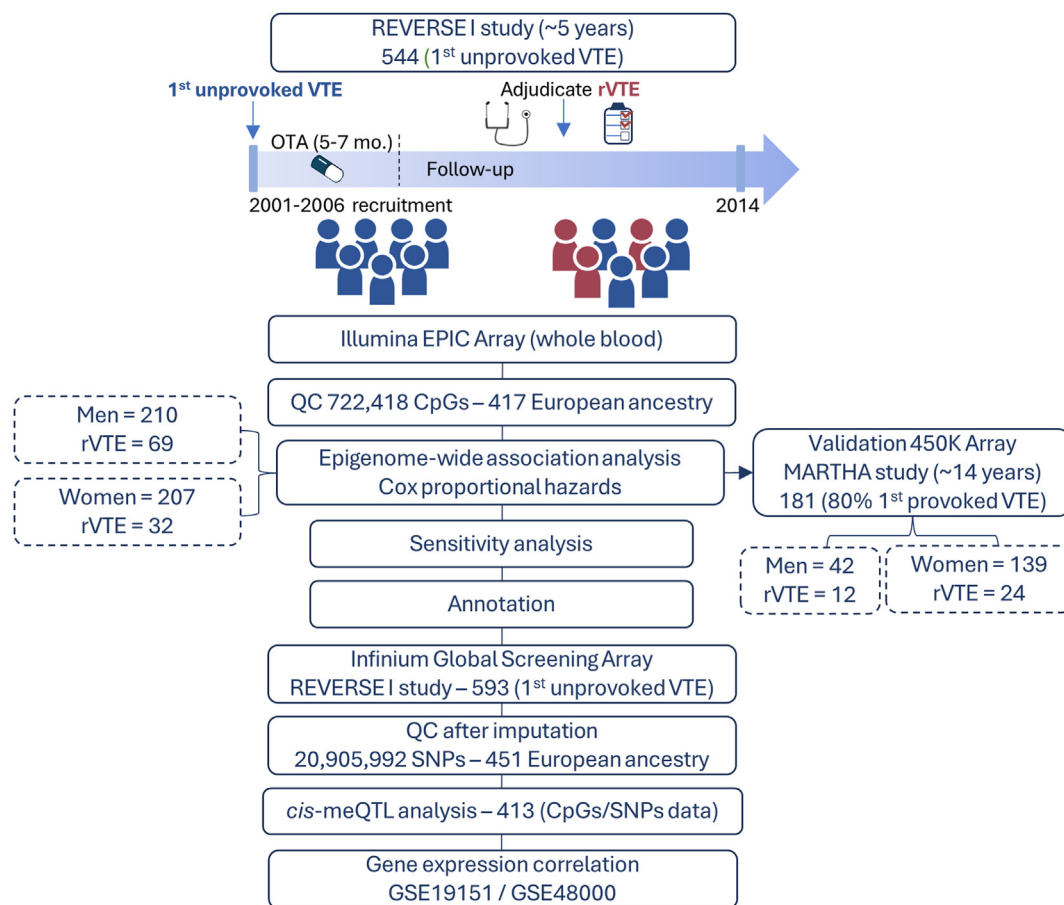
mediating the relationship between sex-biased complex diseases and prognostic biomarkers or disease-related exposures [16–19]. Epigenome-wide association studies (EWAS) have revealed many autosomal loci displaying sex-biased DNAm differences in several tissues [20–23]. Although most DNAm differences and associated changes in gene expression are small, they can still be associated with large effects on disease risks [20]. DNAm biomarkers may account for parts of the “missing heritability” [24] and hold the potential for improving personalized risk assessment for first and rVTE [25,26].

Here, we investigated sex-specific epigenetic signatures underlying rVTE risk differences between men and women with unprovoked or weakly provoked VTE after stopping anticoagulant therapy. We conducted a time-to-event EWAS of almost 723 000 autosomal CpGs from the blood of unprovoked VTE participants followed-up over 5 years for rVTE. Additionally, because genetic variants have strong effects on the DNA methylome, we incorporated genotype data on the same individuals into our analyses as another step toward understanding gene regulatory variation and modulation in rVTE. A deep understanding of the molecular downstream mechanisms acting in rVTE may uncover risk predictors to help classify patients at high risk in whom thromboprophylaxis treatment decisions must be made.

## 2 | METHODS

### 2.1 | Overall analytic strategy

Hypothesizing that differential blood DNAm signatures underlie the sex-biased rVTE, we first tested for an association between time to rVTE in unprovoked VTE patients of the REcurrent VEnous thromboembolism Risk Stratification Evaluation (REVERSE) I study and DNAm levels from the Illumina Infinium EPIC array. We aimed for a complementary validation of our identified associations with rVTE by analyzing a study of provoked first VTE, the MARseille THrombosis Association (MARTHA), where methylation was assayed using the Illumina Infinium 450 K array. Second, to better understand the epigenetic molecular mechanisms of rVTE, we mapped the methylation quantitative trait loci (meQTLs) of the rVTE-associated CpGs in REVERSE I. Third, to explore whether DNAm could repress or activate the loci under the association, we examined the



**FIGURE 1** Study overview. Blood samples of unprovoked venous thromboembolism (VTE) individuals followed-up for recurrence over an average of 5 years were collected while on oral anticoagulants (OTA; REcurrent VENous thromboembolism Risk Stratification Evaluation [REVERSE] I study). From the Illumina EPIC array, we ended up with 722 418 probes, genome-wide analyzed in 417 European-ancestry individuals. We performed a sex-stratified epigenome-wide association study using the Cox regression model adjusted for the principal components of methylation and cell type proportions plus age at the first VTE. After sensitivity analysis and annotation of the cytosines in 5'-C-phosphate-G-3' (CpGs), we used the MARseille THrombosis Association [MARTHA] study, where 80% of the first VTE were provoked and followed-up for an average of 14 years, as a validation step. To evaluate the genetic influence around the associated methylation loci, we integrated methylation and genotyping data of REVERSE I and examined the methylation quantitative trait loci (*cis*-meQTL) SNPs ( $\pm$  250 kb) within the associated CpGs, modeling their effect on VTE recurrence (rVTE). Finally, we compared the transcript levels of the associated genes between individuals with no VTE, VTE, and rVTE using Gene Expression Omnibus datasets. QC, quality control; SNP, single nucleotide polymorphisms.

expression profiles of these genes in individuals with no VTE, first VTE, and rVTE. The overall schema is in [Figure 1](#).

## 2.2 | Study population

REVERSE I is a multicenter prospective cohort study designed to derive a clinical prediction rule for patients at low risk for rVTE after completion of 5 to 7 months of anticoagulant therapy for a first unprovoked VTE [8,27]. All participants were initially treated with >5 days of heparin or low-molecular-weight heparin after their index unprovoked VTE, followed by 5 to 7 months of oral anticoagulants. Enrolment occurred from October 2001 to March 2006, with the last follow-up in 2014. Event-free patients were followed-up for 8 years. During a mean follow-up of 5 years (range, 1 month to 9.77 years), episodes of suspected rVTE were independently adjudicated. Details on recruitment criteria and baseline assessment are in the [Supplementary Methods](#) and elsewhere [8,28].

Gender was reported upon enrolment. All individuals retained for analysis were *cis*-gendered; therefore, we use “sex” throughout. Research ethics board approvals were obtained by all participating centers (Ottawa Hospital Research Ethics Board was the lead institution). The study of genetic and epigenetic data in REVERSE I was approved by the University of Toronto Research Ethics Board. As validation, we used the MARTHA study ([Supplementary Methods](#)). Definitions for a first VTE in REVERSE I and MARTHA are in [Supplementary Table S1](#) and MARTHA characteristics are in [Supplementary Table S2](#).

## 2.3 | Measurements

### 2.3.1 | DNAm processing

Genome-wide DNAm was profiled by the Illumina EPIC array in 544 stored blood samples from REVERSE I individuals. The quality control

features of the R package “meffil” [29] were exploited to identify problematic samples based on I) unusual intensity patterns for control probes and methylated/unmethylated probes, II) failing correlation with matched genotyping data, III) outliers of principal component (PC) analysis, and IV) mismatch between provided gender and estimated sex [29]. Sex was predicted by “wateRmelon” to check for sample mix-ups [30]. Before normalization, we removed probes with a detection  $P$  value  $> .01$  and  $< 3$  beads count in more than 20% of samples. Normalization was based on the method of “meffil.” Control probes in the array were used to guarantee that each batch was normalized in the same way. After normalization, we excluded probes overlapping any single nucleotide polymorphisms (SNP) with a minor allele frequency  $> 5\%$ , probes located in X and Y chromosomes, and cross-hybridizing probes [31–33]. We retained 417 inferred Europeans for the primary analyses at a total of 722 418 autosomal CpGs and included the remaining participants in a multi-ancestry analysis to explore ancestry influence. The methylation level of each CpG was represented by the  $\beta$ -value (proportion of total signal intensity from the methylated-specific probe, ranging from 0 unmethylated to 1 methylated).  $\beta$ -values were logit transformed into methylation levels (ie, M-values) because they give more resolution to methylation near 0 or 1 [34]. Methylation data of the MARTHA study are in [Supplementary Methods](#) and elsewhere [35]. Genotyping measurements of REVERSE I are detailed in [Supplementary Methods](#).

### 2.3.2 | Cell type proportions

DNAm patterns are cell type-specific. Cell type proportions for primary blood cell types of REVERSE I were inferred using the robust partial correlations framework of the package “EpiDISH,” using the IDOLOptimizedCpGs for reference-based deconvolution in whole blood [36]. To reduce collinearity, we computed PC analysis on the cell proportion estimates using the *prcomp* R function and used the first 2 PCs for model adjustments.

## 2.4 | EWAS

### 2.4.1 | Quality control

To capture unmeasured confounders in the methylation data (eg, hidden batch effects), we computed PCs using all autosomal estimated methylation proportions, both using all participants and separately for each sex. The genomic inflation factor ( $\lambda$ ) was initially estimated by comparing the observed test statistics with those expected under a uniform distribution, as commonly used in genome-wide association studies. However, given that genomic inflation in EWAS usually exceeds that observed in genome-wide association studies, possibly due to many small true associations or bias driven by confounding, we also applied a Bayesian method implemented in the R package “bacon” to estimate and correct for residual confounding, thereby improving the estimation of the genomic inflation of methylation data [37,38].

### 2.4.2 | Sex-specific EWAS

To estimate sex-specific risks for each CpG with rVTE, we performed a sex-stratified EWAS fitting Cox proportional hazards model using the R package “survival” [39]. The hazard ratio can be interpreted as the instantaneous risk ratio for rVTE at time  $t$ , considering a 1-unit change in the M-value for the CpG being analyzed and controlling for the effects of the covariates. A hazard ratio  $> 1$  indicates that increased DNAm of a CpG is associated with a reduced probability of survival (increased risk of rVTE). The validity of the proportional hazards assumption was confirmed using Schoenfeld residuals. Covariates included age at baseline VTE, PCs of methylation data (mPCs; mPC1-3), and PCs of cell type proportions (ctPCs; ctPC1-2). mPCs from the general cohort or stratified by sex were used as sensitivity analysis. We applied the Bonferroni correction ( $P = .05/722\,418 = 7 \times 10^{-8}$ ), and a suggestive threshold ( $P < 7 \times 10^{-6}$ ) was also considered for female analysis. As a secondary investigation of the associated loci, we also fit models including interactions between sex and each CpG’s DNAm to test for differences by sex. To test whether the meQTLs played an indirect effect in rVTE mediated by the CpGs, we included meQTL genotypes as predictors in the models and evaluated the evidence for the causal relationship of genotypes, methylation, and outcome [40]. To visualize survival curves using the package “survminer” [41], individuals were dichotomized into high and low methylation groups based on the sex-specific median cutoff for each CpG’s methylation proportions. CpG annotation was acquired from Illumina EPIC manifest 1.0 B5 (GRCh38). Attempting to improve statistical power to identify additional candidate loci, we also performed a hypothesis-driven EWAS, ie, up-weighting CpGs in genes targeted by oral anti-coagulants [42,43]. This approach, along with the proof-of-concept study details that led to its hypothesis, are in the [Supplementary Methods](#) and [Supplementary Figures S1 and S2](#).

### 2.4.3 | Sensitivity analysis

Sensitivity analyses were carried out to assess the robustness of the regression coefficients and to account for potentially unmeasured confounders. For the candidate associations, we checked for sensitivity to the choice of the Cox model by reanalyzing the data with (1) Cox proportional hazard models using methylation residuals after adjusting for mPCs and ctPCs adjusted by age (since we had few recurrences and many covariates, we fit residuals to improve the model); (2) case-base sampling models, which allows smooth baseline risk estimates [44] and (3) Cox robust models that address robustness to outliers [45]. Second, we assessed sensitivity to our choice of covariates by adjusting our standard model for (4) body mass index (BMI), (5) type of first VTE, (6) D-dimer, and (7) including more than 2 ctPCs. Finally, (8) including genome-wide genotyping PCs as adjustments, we ran the model in the multi-ancestry dataset with a total of 452 individuals (416 Europeans, 17 Africans, 16 Native Americans, 2 East Asians, and 1 South Asian).

## 2.5 | Molecular characterization

### 2.5.1 | Proximal SNP-CpG associations

We conducted meQTL analyses to evaluate the influence of proximal SNPs within  $\pm 250$  kb (*cis*-meQTL) on methylation M-values at our top associated CpGs. We analyzed biallelic SNPs with minor allele frequency  $> 5\%$  in the European samples of REVERSE I with genotyping and methylation data and tested the association of DNAm levels (dependent variable) with each SNP (coded 0, 1, or 2 for the number of minor alleles) in the respective window using linear models ( $P < .05$ ). Independent variables were age, sex, mPC1-3, and ctPC1-2 as well as the SNPs. We then estimated the total genetic contribution at each locus by multiple penalized linear regression (least absolute shrinkage and selection operator) using the R package “glmnet,” with the penalty parameter chosen to minimize the residual error after 10-fold cross-validation [46]. For visualization, CpG methylation levels were compared across genotypes. The package “biomaRt” was used for SNP annotation [47]. We validated our meQTL results by comparing them with databases of the Genetics of DNA Methylation Consortium (GoDMC) and eQTLGen Consortium and checking reports as expression quantitative trait locus (eQTL) in the Genotype-Tissue Expression (GTEx) portal [48–50]. Details of further analysis on loci characterization are in the [Supplementary Methods](#).

## 3 | RESULTS

### 3.1 | Study population characteristics

Of 455 unprovoked VTE subjects from REVERSE I passing quality control, 417 with European ancestry were included in our primary analyses, of whom 207 were women, and 210 were men ([Figure 1](#), [Table 1](#)). Over a 5-year follow-up, men had more incident rVTE events ( $n = 69$ , 32% vs  $n = 32$ , 15%;  $P = 5.53 \times 10^{-5}$ ) and manifested recurrent episodes younger than women ( $52.9 \pm 14.8$  vs  $57 \pm 20.6$  years;  $P = .15$ ). Higher D-dimer after 5 to 7 months of anticoagulation was detected in those who developed rVTE during follow-up. A higher percentage of individuals with DVT as a first VTE type had a recurrent event (72% in men and 60% in women). Time to recurrence after the end of anticoagulant therapy was similar between men (2.4 years) and women (2.5 years). The validation subset of MARTHA ( $n = 181$ ) involved  $\sim 80\%$  provoked VTE individuals, mostly women (77%), on average 15 years younger than men ([Supplementary Table S2](#)). Higher recurrence was also observed in men ( $n = 28\%$  vs  $n = 17\%$ ). MARTHA participants had, on average, a lower BMI (24.1) compared with REVERSE I participants (29.6). Time to recurrence in MARTHA was longer than in REVERSE I, with an average of 5.2 years for men compared with 6.1 for women.

### 3.2 | Sex-specific hypomethylated CpGs associated with rVTE

Our sex-based EWAS predicting rVTE were adjusted for PCs derived from all methylation probes to capture technical factors and residual confounding and PCs from the estimated cell type composition to capture cell type heterogeneity. Sensitivity analyses using sex-specific PCs showed no notable differences in association results compared with global PCs ([Supplementary Figure S3](#)). QQ plots ascertained whether models were well adjusted ([Figure 2](#), [Supplementary Figure S4](#)). To assess for widespread uncaptured confounding, we estimated  $\lambda$  using both the conventional and “bacon” methods. After “bacon,” the distribution of corrected  $P$  values showed no evidence of inflation when analyses were adjusted by age + mPC1-3 + ctPC1-2 in men ( $\lambda = 1.02$ ; Bayesian  $\lambda = 0.99$ ) and women ( $\lambda = 1.36$ ; Bayesian  $\lambda = 1.00$ ), even though inflation was observed before the correction in women ([Figure 2](#)).

In men, DNAm levels at 2 CpG sites were significantly associated with rVTE at our genome-wide significance level ( $P < 7 \times 10^{-8}$ ; [Figure 2A](#)). The top hit was cg01060850, detected with power  $> 99\%$  ([Supplementary Table S3](#)), located in the *TBC1D22B* gene body (coefficient =  $-2.66$ ;  $P = 3.23 \times 10^{-8}$ ), followed by cg07808424 (coefficient =  $-1.55$ ;  $P = 2.47 \times 10^{-8}$ ) in the 5'UTR of the *ZHX2* gene ([Table 2](#)). In women, the only statistically significant CpG site, ch.12.1038646R (coefficient =  $-0.76$ ;  $P = 2.79 \times 10^{-8}$ ), was annotated to the *DIP2B* gene body ([Figure 2B](#)). Since the small number of recurrent events in women will limit statistical power to detect true associations, using a suggestive threshold ( $P < 7 \times 10^{-6}$ ), we also examined the top 5 probes in the female EWAS, adjusting for either mPC1-3 or mPC1-4. Therefore, we also investigated a second suggestively significant association in women: cg03401656 (coefficient =  $-2.15$ ;  $P = 5.46 \times 10^{-7}$ ) lies in the gene body of *DENND3* ([Supplementary Figure S4](#)). When fitting models containing the interaction term with sex, coefficients for the associated probes followed the same direction as in the sex-specific analysis with significant  $P$  values for 3 out of the 4 probes (except cg07808424), suggesting moderation by sex of DNAm association and recurrence ([Supplementary Table S4](#)). The same top CpGs were associated in the hypothesis-driven EWAS, whereas no prioritized probe or additional loci were revealed ([Supplementary Figure S5](#)). The associated probes located genes implicated in vesicular trafficking, signal transduction, and methylation/transcriptional regulation ([Supplementary Table S5](#)). In the estimated survival curves, patients exhibiting shorter overall survival (ie, earlier rVTE) tended to have lower methylation levels of cg01060850 and cg07808424 in men but not women ([Figure 3A, B](#)). Correspondingly, shorter survival probabilities were observed for hypomethylated ch.12.1038646 and cg03401656 in women but not men ([Figure 3C, D](#)). Significant sex-specific hypomethylation on a global level was observed in rVTE for cg07808424, ch.12.1038646R, and cg03401656 ([Figure 3](#)).

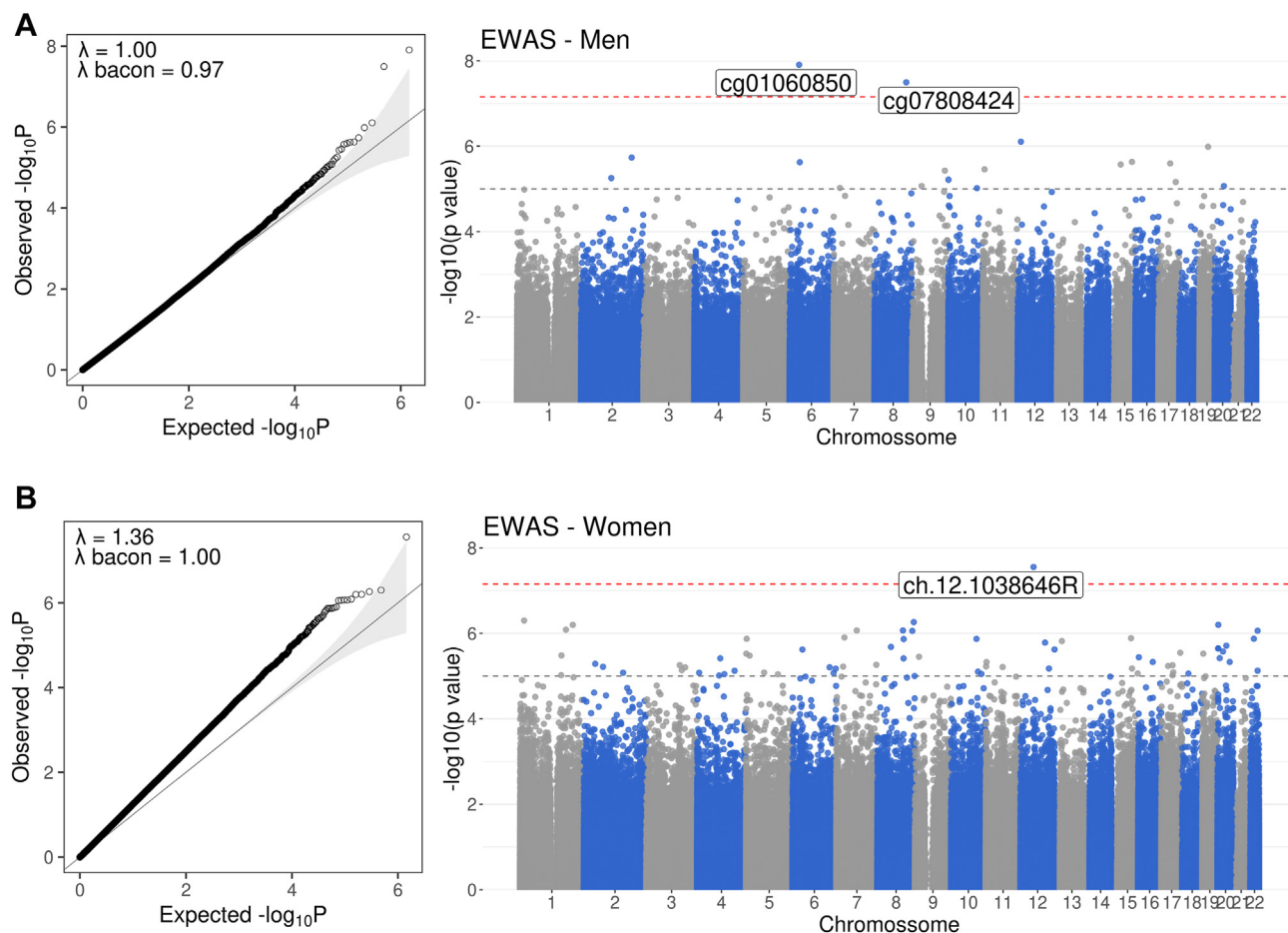
TABLE 1 The REcurrent VEnous thromboembolism Risk Stratification Evaluation (REVERSE) I study sample characteristics.

Characteristics	Total N = 417			Men n = 210			Women n = 207		
	VTE n = 316	rVTE n = 101	P value (95% CI)	VTE n = 141	rVTE n = 69	P value (95% CI)	VTE n = 175	rVTE n = 32	P value (95% CI)
Age at first VTE (y), mean (SD)	53.5 (17.9)	54.2 (16.9)	.72 (-4.90; 3.42)	56.3 (15.4)	52.9 (14.8)	.13 (-1.18; 7.91)	51.3 (19.4)	57.0 (20.6)	.10 (-14.4; 1.50)
BMI (kg/m <sup>2</sup> ), mean (SD)	29.0 (7.2)	30.3 (8.2)	.30 (-2.01; 0.64)	29.5 (6.0)	29.3 (7.3)	.30 (-0.70; 2.24)	28.6 (8.0)	32.6 (9.6)	.03 (-6.37; -0.22)
D-dimer (ng/mL), mean (SD)	298.6 (321.4)	310.2 (268.6)	.22 (-49.9; 11.9)	278.9 (255.2)	284.3 (261.5)	.85 (-32; 40.9)	314.5 (366.4)	365.3 (279.2)	.03 (-152; -5.99)
First-type VTE									
DVT only, n (%)	156 (49)	69 (68)	1x10 <sup>-3</sup> (-0.30; -0.07)	78 (55)	50 (72)	.02 (-0.31; -0.02)	78 (45)	19 (60)	.17 (-0.35; 0.05)
PE or PE + DVT, n (%)	160 (51)	32 (32)		63 (45)	19 (28)		97 (55)	13 (40)	
FV Leiden heterozygous, n (%)	47 (14.8)	18 (17.8)	.58 (-0.06; 0.11)	20 (14.1)	8 (11.5)	.76 (-0.13; 0.07)	27 (15.4)	10 (31.2)	.06 (-0.03; 0.34)
FII G20210A, heterozygous, n (%)	19 (4)	4 (0.9)	.59 (-0.07; 0.03)	9 (4)	3 (1)	.77 (-0.09; 0.05)	10 (4)	1 (0.4)	.86 (-0.11; 0.06)
Family history, n (%)	77 (24.3)	28 (27.7)	.62 (-0.07; 0.13)	28 (19.8)	19 (27.5)	.30 (-0.06; 0.20)	49 (28)	9 (28)	1 (-0.17; 0.16)
Time to recurrence (y), mean (SD)	6.0 (2.9) <sup>a</sup>	2.4 (2.3)	2x10 <sup>-16</sup> (3.32; 5.26)	5.6 (3.0) <sup>a</sup>	2.4 (2.4)	1x10 <sup>-12</sup> (2.42; 5.04)	6.7 (2.8) <sup>a</sup>	2.5 (2.4)	1x10 <sup>-10</sup> (2.90; 6.24)

Numbers might not be 100% due to missing information. D-dimer was measured by VIDAS at the moment of the first VTE.

BMI, body mass index; DVT, deep vein thrombosis; FII/FV, factor II/V; PE, pulmonary embolism; rVTE, recurrent venous thromboembolism; VIDAS, Versatile Immunoassay Detection Assay System; VTE, unprovoked venous thromboembolism.

<sup>a</sup> Time to last follow-up in years. The univariate analysis applied the Wilcoxon test and the chi-square test for continuous and categorical variables, respectively.



**FIGURE 2** QQ and Manhattan plots of the sex-stratified time-to-event epigenome-wide association studies (EWAS) of DNA methylation and recurrent VTE. QQ plots show the distribution of the observed  $\log_{10}$ -fold transformed  $P$  values against the expected null distribution in (A) men and (B) women. The gray shaded area represents the 95% CI for equality. Cox models were adjusted by age and principal components of methylation and cell type proportions. The conventional genomic inflation factor  $\lambda$  and the Bayesian  $\lambda$  from “bacon” are given in each QQ plot to quantify the statistical inflation of the  $P$  values. Manhattan plots display the  $P$  values from the Cox regression testing the association of each cytosine in 5′-C-phosphate-G-3′ with recurrent VTE. The horizontal axis represents chromosomal position, and the vertical axis shows the  $P$  values on a  $-\log_{10}$ -fold scale. Dashed red and gray lines mark Bonferroni ( $P < 7 \times 10^{-8}$ ) and suggestive significance thresholds ( $P < 1 \times 10^{-5}$ ).

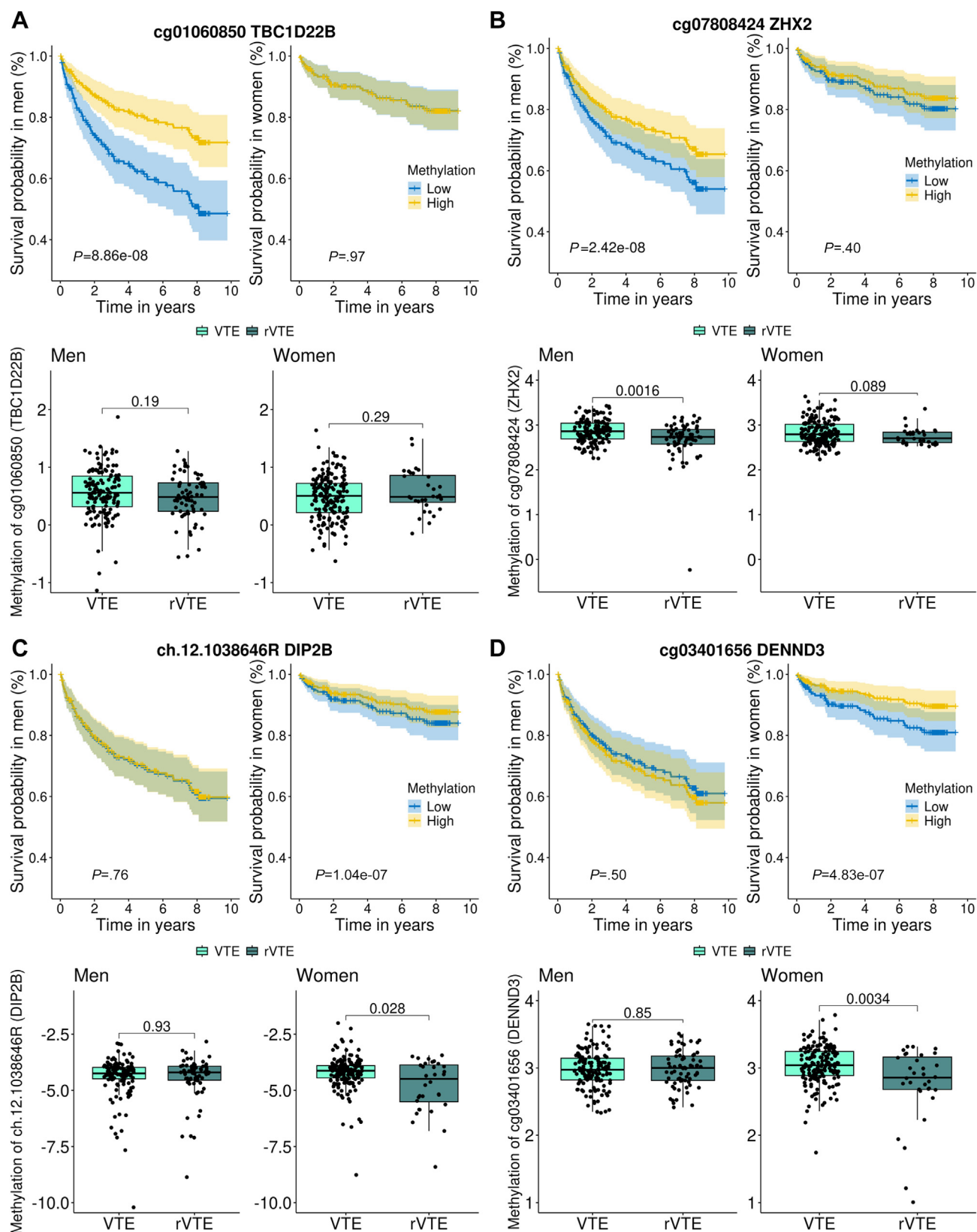
**TABLE 2** Cytosine-phosphate-Guanine (CpG) sites whose methylation levels showed sex-specific association at Bonferroni-corrected significance levels in the time-to-event EWAS in the REcurrent VENous thromboembolism Risk Stratification Evaluation (REVERSE) I.

CpG Chr:position	Gene annotation	Location at the gene	Men (n = 210)			Women (n = 207)		
			Coeff.	HR (95% CI)	P value	Coeff.	HR (95% CI)	P value
cg01060850 chr6:37260002	<i>TBC1D22B</i>	Body	-2.66	0.06 (0.02; 0.17)	$3.23 \times 10^{-8}$	0.09	1.10 (0.33; 3.62)	.87
cg07808424 chr8:122802461	<i>ZHX2</i>	5′ UTR	-1.55	0.21 (0.12; 0.36)	$2.47 \times 10^{-8}$	-0.09	0.40 (0.07; 2.31)	.30
ch.12.1038646R chr12:50731142	<i>DIP2B</i>	Body	-0.04	0.95 (0.76; 1.20)	.71	-0.76	0.46 (0.35; 0.60)	$2.79 \times 10^{-8}$
cg03401656 chr8:141151888	<i>DENND3</i>	Body	0.38	1.46 (0.54; 3.91)	.44	-2.15	0.11 (0.04; 0.26)	$5.46 \times 10^{-7}$

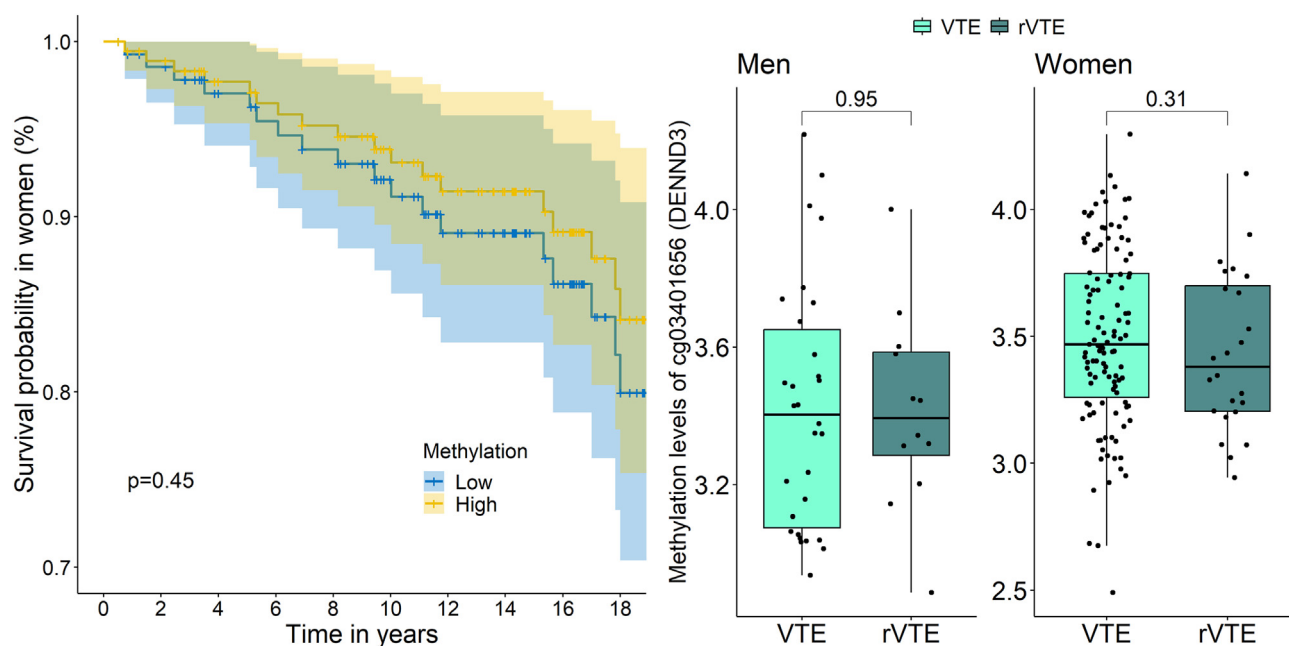
Gene annotation is according to EPIC manifest v1.0 B5 (University of California, Santa Cruz [UCSC] Genome Browser). HR were calculated using Cox regression with natural log-transformed values.

Chr, chromosome; CpG, name of the cytosine in 5′-C-phosphate-G-3′ site; Coeff., regression coefficient (the Cox model coefficient represents the expected change in the log HR for a unit change in the methylation level [ie, M-value]); HR, hazard ratio (HR per 1-unit M-value increase); Position, the genomic location of each site in genome build GRCh38.





**FIGURE 3** Overall survival curves for first and recurrent venous thromboembolism (rVTE) in the REcurrent VEinous thromboembolism Risk Stratification Evaluation (REVERSE) I. Graphics show independent analyses for cytosine in 5'-C-phosphate-G-3 (A) cg01060850, (B) cg07808424, (C) ch.12.1038646, and (D) cg03401656. Patients were divided into low and high methylation groups using the median cutoff of methylation levels (ie, *M-values*). The estimated survival curves were calculated based on Cox models for time to rVTE as a function of methylation residuals of principal components and cell type proportions adjusted by age. The x- and y-axes denote survival time in years and the probability of patient survival, respectively. Boxplots of methylation levels showing Wilcoxon test *P* values for the comparison of DNA methylation between single venous thromboembolism (VTE; men = 141 and women = 175) and recurrent venous thromboembolism (rVTE; men = 69 and women = 32).



**FIGURE 4** Survival curves and methylation levels of *DENND3*-cg03401656 in MARTHA. Patients were divided into low and high methylation groups using the median cutoff of methylation levels (ie, *M-values*). The estimated survival curves were calculated based on Cox models for time to recurrent venous thromboembolism (rVTE) as a function of methylation residuals of principal components and cell type proportions adjusted by age. The x- and y-axes denote survival time in years and the probability of patient survival, respectively. Survival curves were not calculated for men in MARTHA due to the small number of events. Boxplots, along with the Wilcoxon test, were plotted to compare DNA methylation between single venous thromboembolism (VTE; men = 30 and women = 115) and rVTE (men = 12 and women = 24).

Among these sex-specific rVTE-associated CpGs, only 1 probe was available for validation in the independent set of MARTHA. *DENND3*-cg03401656 was suggestively hypomethylated in rVTE compared with VTE in women (Figure 4). The regression coefficient had the same sign as in REVERSE I but showed essentially no evidence of statistical significance (coefficient =  $-0.56$ ;  $P = .45$ ). Although neighboring correlated CpGs with the male-associated sites were suggestively hypomethylated in men (Supplementary Figure S6), male-derived analysis in MARTHA was limited due to sample size and absence of our top probes.

### 3.3 | Sensitivity analyses

In our sensitivity analyses, associations remained robust after evaluating the impact of including more than 2 ctPCs, BMI, D-dimer, first-type VTE, and FVL in the models (Supplementary Table S6). In the multiancestry dataset, cg07808424 and cg01060850 remained the top 2 hits in men, although cg01060850 did not pass Bonferroni, and ch.12.1038646R continued to be associated in women (Supplementary Figure S7). The inclusion of the ancestral genotype PCs did not change the detected effect sizes of the CpGs. All significant or suggestive associations displayed negative effect estimates, indicating that the CpGs are more likely to be hypomethylated in rVTE. Also, we obtained similar estimates after applying robust adaptations of the Cox model to reduce the potential influence of methylation outliers (Supplementary Table S7).

### 3.4 | Genetic effects on the association between CpGs and rVTE

We carried out a *cis*-meQTL analysis to investigate the role of proximal genetic variation (within 250 kb) on DNAm in REVERSE I (Supplementary Figures S8 and S9). After least absolute shrinkage and selection operator regression, SNPs were pruned for high linkage disequilibrium (LD) ( $R^2 > .7$ ) and defined as *cis*-meQTLs if the minor allele significantly modulated DNAm. All associated probes had at least 4 SNPs affecting DNAm, many of them previously confirmed as *cis*-meQTLs for the gene of interest or associated with additional CpGs (100-650 kb) from the candidate site, attesting that single genetic variants are often associated with methylation at numerous nearby CpG sites (Supplementary Table S8). The proportion of variance explained by meQTLs for the methylation sites was 6% (ch.12.1038646R), 20% (cg07808424 and cg03401656), and 54% (cg01060850). In eQTLGen and GoDMC, our identified *cis*-meQTLs for cg01060850 were reported at genes in the vicinity of *TBC1D22B*. rs55800130 was the only meQTL that also acts as an eQTL for *CMTR1* in fibroblasts, which play an important role in maintaining microenvironment homeostasis [51]. Many meQTLs near *ZHX2* were strongly associated with cg07808424 (false discovery rate  $< .05$ ); 9 of them were validated as *ZHX2* meQTLs in public resources. The meQTLs of ch.12.1038646 were also eQTLs for *DIP2B* with relevance in whole blood and cultured fibroblasts (GTEx). At cg03401656-*DENND3*, we corroborated the effect of numerous meQTLs at *DENND3* and adjacent genes, including the synonymous variant rs1045303, also characterized as an eQTL of whole blood and

fibroblasts. To study whether the meQTLs played an indirect effect on rVTE due to DNAm, we modeled the individual and joint effects of the meQTLs of each gene as predictors. Despite some meQTLs demonstrating an association with rVTE, neither *P* values nor regression coefficients of the CpGs were significantly altered when including them, suggesting that the rVTE association depends on the methylation and not the meQTLs (Supplementary Table S9).

### 3.5 | Differentially expressed genes associated with rVTE

To gain insight into the relationship between DNAm differences and gene expression in VTE, we examined the messenger RNA levels for our identified genes between no VTE and VTE groups in Gene Expression Omnibus microarray datasets. The enriched Gene Ontology terms for the top 1000 genes that most vary between the groups of GSE48000 and GSE19151 attested to the relevance of vesicular and transport-related ontologies in VTE (protein binding, cytosol, cytoplasm, extracellular exosome, endoplasmic reticulum, Golgi apparatus, and endosome; Supplementary Figure S10).

*TBC1D22B* is included among the 1285 upregulated genes in VTE compared with no VTE in GSE48000, and *DENND3* among the 187 downregulated genes in the GSE19151 dataset by Fan et al. [52]. *TBC1D22B* was found to be significantly overexpressed in unprovoked VTE with >2 recurrences compared with no VTE or provoked VTE and displayed a suggestive overexpression ( $P = .06$ ) compared with unprovoked VTE with 1 recurrence in GSE48000 (Supplementary Figure S11A). The same trend was observed in GSE19151 (Supplementary Figure S11G). The expression of *TBC1D22B* was considerably increased in rVTE women compared with rVTE men in reference to no VTE in their respective group, suggesting a sex-biased regulation in the context of VTE and a potential association with recurrence. Interestingly, *RAB27A*, a downregulated gene in recurrent unprovoked VTE at GSE48000, displayed an opposite direction of expression in VTE and rVTE compared with *TBC1D22B* (Supplementary Figure S11E). The protein interaction network revealed that *TBC1D22B* directly interacts with *CMTR1* and indirectly interacts with *RAB27A* through *TBC1D10A* (TBC1 domain family member 10A; Supplementary Figure S12). Furthermore, probes at *CMTR1* showed a significant correlation ( $.4 < R^2 < .6$ ) with *TBC1D22B*-cg01060850. Because they colocalize in the same topological association domains, it is likely that they share functional relevance and regulatory processes (Supplementary Figure S13).

## 4 | DISCUSSION

Limitations of current clinical decision-making tools to predict high-risk men and women for rVTE after stopping anticoagulation therapy challenge the clinical management of unprovoked VTE. A deeper understanding of the underlying processes of recurrence may help tailor the duration of anticoagulation to high-risk individuals. In a

preliminary candidate gene analysis (see Supplementary Figure S1), we observed sex-specific blood DNAm marks in vitamin K cycle genes (*VKORC1* and *UBIAD1*) relevant to the coagulation cascade and vascular integrity. Here, interrogating almost 730 000 DNAm sites on the genome of unprovoked VTE patients, we identified male- and female-independent DNA hypomethylation marks favoring recurrence in a sex-specific manner, providing a mechanistic explanation for the role of genes previously hypothesized in VTE. These associations remained significant after adjusting for potential confounders and inclusion of meQTLs, suggesting that rVTE might be driven, at least in part, by DNAm levels at vesicle-mediated transport-related genes. Determining whether these associated CpGs can predict rVTE warrants further validation.

Exosomes regulate intercellular communication and signal transduction, and differential methylation at vesicle trafficking regulatory genes might ultimately impact VTE [48,49]. *TBC1D22B* and *DENND3* are differentially expressed in individuals who had a first VTE or rVTE compared with those without [52]. *TBC1D22B*, a TBC1 domain family (member 22B) gene, codes a member of the guanosine triphosphate (GTP)ase-activating proteins for Rab proteins, which are predominantly expressed in bone marrow [53]. Sex-biased expression has been observed for other members of this family (*TBC1D1* and *TBC1D4*) [54]. *DENND3*, expressed in blood and immune cells, codes the largest family of Rab-activating guanine nucleotide exchange factors (GEFs) [55]. Rab GTPases are crucial for vesicle-mediated membrane trafficking and cycling between inactive (guanosine diphosphate-bound) and active (GTP-bound) conformations. GEFs mediate their activation, while GTPase-activating proteins terminate their activity [56]. Rab GTPases regulate Weibel–Palade bodies exocytosis, endothelial organelles involved in the release of von Willebrand factor that increases endothelial permeability and vascular tone in angiogenesis and thrombosis. One Rab gene that acts on exome secretion is *RAB27A*, which has been shown to be highly expressed in VTE patients with 2 or more sustained, unprovoked events compared with those who had 1 or more provoked VTE events [57]. Biesemann et al. [58] revealed that overexpression of *TBC1D22B*, *TBC1D14*, *TBC1D10A*, and *TBC1D10B* reduced von Willebrand factor secretion, indicating an inhibitory effect on Weibel–Palade bodies' exocytosis. While it is unclear whether *TBC1D22B* directly modulates *RAB27A*, we observed an upregulation of *TBC1D22B* and downregulation of *RAB27A* in unprovoked VTE with >2 recurrences, pointing toward an inhibitory effect of *TBC1D22B* on *RAB27A*, demanding further evaluation. Engagingly, *TBC1D3C*, also involved in Rab-regulated trafficking, was differentially expressed in DVT patients manifesting early recurrent events postanticoagulation compared with those without recurrence [59]. Furthermore, by activating Rab12, *DENND3* mediates autophagy [55]. Studies have shed light on the role of autophagy in hemostasis and thrombosis, but no direct link between GEF-mediated autophagy and VTE has been documented. In our study, even prioritizing coagulation genes, some with sex-biased methylation patterns, no genes of coagulation were associated with rVTE. Rather, we found DNAm marks in genes involved in vesicle trafficking and transcriptional regulation.

Studies have implicated the *DIP2B* gene in cardiovascular diseases and VTE through its modulation of circulating fibrinogen [60]. Elevated fibrinogen increases the risk of residual VTE, but the association with rVTE remains questionable [61,62]. *DIP2B* codes a disc-interacting protein homolog 2 protein family containing binding sites for DNA methyltransferase 1 associated protein 1 and adenosine monophosphate binding sites involved in transcriptional regulation. *ZHX2*, a zinc-finger and homeobox family of transcription factors, acts as a transcriptional repressor and oncogene [63]. Near *ZHX2* is another TBC1 family member gene, *TBC1D31*, involved in cilium biogenesis [64]. Several other transcription factors have sex-biased regulatory targeting patterns [20,65]. In mice, *ZHX2* is a regulator of sex-biased CYP gene expression, a key drug metabolism enzyme [66]. Besides, *ZHX2* positively regulates the hypoxia-inducible factor 1A activity, a trigger compound of VTE [67,68]. Despite detecting the cg07808424-*ZHX2* association in men, we do not see it when fitting interactions between sexes. This probe was suggestively associated with rVTE when no sex stratification was done in REVERSE I (data not shown). This might have occurred due to a higher number of recurrences in men, increasing the power to detect associations or potential uncontrolled variables.

There are several strengths to this study. First, we used a cohort of a first unprovoked VTE followed-up for several years, enabling robust time-to-event analysis. While others linked higher levels of global DNAm (before treatment) with rVTE in males [69], we evidenced locus-specific DNAm associated with sex-biased risk of rVTE when participants were under anticoagulants. Second, our sensitivity analyses showed consistent estimates after adjustments for several potential confounders, running adaptations of the Cox model, and considering the cohort that included multiethnic individuals adjusting for heterogeneity on a global scale. The inclusion of different ethnic compositions did not expressively alter the replication of the top probes, although studies with larger proportions of non-European descent individuals are required for generalizability. Third, using the genotype data available in the same individuals, we screened *cis*-meQTL effects in the associated genes. Note that distal genetic influence was not considered in our study, and *cis*-regulatory SNPs might have a pleiotropic effect on multiple genes, not just the sites investigated [58]. To explore the gene regulatory effect on DNAm and rVTE, we evaluated the influence of *cis*-meQTLs on regression coefficients assuming potential causal relationships: (I) the meQTL contributes to rVTE (or to CpG) either directly (SNP → rVTE or SNP → CpG), (II) indirectly through the presence of the CpG (SNP → CpG → rVTE), or (III) the CpG has a direct effect on rVTE (CpG → rVTE). The last condition better represents our findings, suggesting that while some meQTLs were independently associated with DNAm, the association with rVTE was most likely caused by DNAm.

Limitations are also acknowledged. We lacked another methylation dataset of unprovoked VTE. (1) In MARTHA, we were unable to explore independent validation in unprovoked VTE; (2) the MARTHA sample size was insufficient for the sex-stratified analysis in men and had low power for women; (3) MARTHA lacked most of the associated probes of our discovery cohort; and (4) MARTHA had considerable

differences in age and time-to-event compared with REVERSE, which might have impacted the validation of the single probe available in the dataset. Even though cg03401656 exhibited the same direction of association in REVERSE and MARTHA, it is not clear whether methylation regulation plays a major role in rVTE in provoked VTE; any trends could also be attributed to the transient risk factors of a first VTE. In unprovoked VTE, however, since there is no clear risk factor, molecular (dys)regulation acting through methylation and, ultimately, gene expression could prominently influence rVTE risk. Likewise, etiologic disparities between provoked and unprovoked VTE may confound the gene expression level comparisons between VTE and rVTE in Gene Expression Omnibus datasets [70].

A well-established statistical method to remove the effect of cell type heterogeneity was applied in our analyses, but residual confounding might still be possible. The smaller sample size of rVTE in women, which was expected since rates of rVTE are higher in men, inevitably leads to lower statistical power. Our results warrant follow-up, ideally in methylation datasets with larger sample sizes. This is important because little is known about the plausible EWAS effect sizes in complex traits due to the decreasing longitudinal stability of the epigenetic marks and their variance within biological samples [71]. We were unable to elucidate whether the sex-associated CpGs play a causal role in driving rVTE or whether the DNAm changes represent a downstream effect of the anticoagulation therapy. While alterations in DNAm at warfarin regulatory genes, or perhaps direct oral anticoagulants, might influence drug efficacy [72], we could not assess the impact of anticoagulants on methylation in our candidate genes because all participants were under treatment. If anticoagulant activity does alter DNAm, it would affect gene expression. Therefore, analyzing both methylation and gene expression in the same samples is crucial to evaluate if DNAm can serve as a biomarker for high-risk VTE patients prior to their withdrawal from anticoagulant therapy. Notwithstanding, we cannot exclude the possibility that the DNAm differences reflect the influence of other factors that were not controlled, such as environmental factors and other exposures (eg, smoking and medications). It is worth considering that ~20% of the women in REVERSE were on estrogen therapy at the time of the index event, and while the risk of recurrence was not higher in users of estrogen compared with nonusers, we cannot exclude that oral contraceptives could impact methylation [73].

Overall, our findings generated new hypotheses on putative mechanisms by which DNAm regulates rVTE in a sex-biased manner. While our findings warrant further validation, it appears very likely that genes outside the coagulation pathway play a significant role in VTE pathophysiology. Comparing our results with those of a similar European-based VTE cohort and a multiethnic study will be particularly informative. Moreover, the lack of gene expression data in REVERSE I limited our capacity to assess if DNAm levels were followed by a direct impact on the expression of the associated genes. Further transcriptional characterization at those loci would be relevant to better understand the sex-biased risk of rVTE and optimize decision-making on thromboprophylaxis after a first VTE event.

## ACKNOWLEDGMENTS

This study was supported by a research grant from the Canadian Institutes of Health Research (grant PJT-162233) and a fellowship from the CANSSI (Canadian Statistical Sciences Institute) Ontario STAGE (Strategic Training for Advanced Genetic Epidemiology) program. The authors thank Lee Vernich for her diligent support in coordinating the array analyses of the REcurrent VENous thromboembolism Risk Stratification Evaluation (REVERSE) I study samples, Elham Sabri for sharing her deep knowledge of the REVERSE I data, and Liz Li, Jingjing Cao, and Nicole Roslin of The Centre for Applied Genomics, The Hospital for Sick Children, Toronto, Canada, for assistance in the processing of the methylation and genotyping data from REVERSE I. The MARseille THrombosis Association (MARTHA) project was supported by a grant from the Program Hospitalier de Recherche Clinique. The H450M epityping in MARTHA was partially funded by the Canadian Institutes of Health Research (grant MOP 86466) and by the Heart and Stroke Foundation of Canada (grant T6484). Statistical analyses of the REVERSE and MARTHA datasets were performed using the Niagara supercomputer at the SciNET High Performance Computing Consortium of the University of Toronto and the Centre de Bioinformatique de Bordeaux (CBIB) computing center of the University of Bordeaux, respectively. The authors are thankful to the individuals who participated in the study.

## AUTHOR CONTRIBUTIONS

F.G. designed the study, obtained the funding in collaboration with M.R., and led the overall project. O.C.L.B., C.M.T.G., and F.G. designed the study analytic plan and interpreted the study findings. O.C.L.B. led the implementation of the analyses and drafted and led the manuscript writing. C.M.T.G. and F.G. supported O.C.L.B. in drafting the manuscript. M.R., M.J.K., and G.L.G. designed and led the REcurrent VENous thromboembolism Risk Stratification Evaluation I cohort study data collection. D.-A.T. and P.E.M. designed and led the MARseille THrombosis Association study data collection. G.M. conducted statistical analysis in the MARTHA study. All authors provided feedback on the manuscript and approved its final version for publication.

## DECLARATION OF COMPETING INTERESTS

There are no competing interests to disclose.

## ORCID

Ohanna C.L. Bezerra  <https://orcid.org/0000-0002-0798-778X>

## X

Ohanna C.L. Bezerra  @ohannaclb

## REFERENCES

- [1] Scheres LJJ, Lijfering WM, Cannegieter SC. Current and future burden of venous thrombosis: not simply predictable. *Res Pract Thromb Haemost.* 2018;2:199–208.
- [2] Kearon C, Kahn SR. Long-term treatment of venous thromboembolism. *Blood.* 2020;135:317–25.
- [3] Khan F, Rahman A, Carrier M, Kearon C, Weitz JI, Schulman S, Couturaud F, Eichinger S, Kyrle PA, Becattini C, Agnelli G, Brighton TA, Lensing AWA, Prins MH, Sabri E, Hutton B, Pinede L, Cushman M, Palareti G, Wells GA, et al. Long term risk of symptomatic recurrent venous thromboembolism after discontinuation of anticoagulant treatment for first unprovoked venous thromboembolism event: systematic review and meta-analysis. *BMJ.* 2019;366:14363. <https://doi.org/10.1136/bmj.l4363>
- [4] Ainle FN, Kevane B. Which patients are at high risk of recurrent venous thromboembolism (deep vein thrombosis and pulmonary embolism)? *Blood Adv.* 2020;4:5595–606.
- [5] Roach REJ, Cannegieter SC, Lijfering WM. Differential risks in men and women for first and recurrent venous thrombosis: the role of genes and environment. *J Thromb Haemost.* 2014;12:1593–600.
- [6] Tosetto A, Iorio A, Marcucci M, Baglin T, Cushman M, Eichinger S, Palareti G, Poli D, Tait RC, Douketis J. Predicting disease recurrence in patients with previous unprovoked venous thromboembolism: a proposed prediction score (DASH). *J Thromb Haemost.* 2012;10:1019–25.
- [7] Eichinger S, Heinze G, Kyrle PA. D-dimer levels over time and the risk of recurrent venous thromboembolism: an update of the Vienna prediction model. *J Am Heart Assoc.* 2014;3:e000467. <https://doi.org/10.1161/JAHA.113.000467>
- [8] Rodger MA, Kahn SR, Wells PS, Anderson DA, Chagnon I, Le Gal G, Solymoss S, Crowther M, Perrier A, White R, Vickars L, Ramsay T, Betancourt MT, Kovacs MJ. Identifying unprovoked thromboembolism patients at low risk for recurrence who can discontinue anticoagulant therapy. *CMAJ.* 2008;179:417–26.
- [9] Rodger MA, Le Gal G, Anderson DR, Schmidt J, Pernod G, Kahn SR, Righini M, Mismetti P, Kearon C, Meyer G, Elias A, Ramsay T, Ortel TL, Huisman MV, Kovacs MJ, REVERSE II Study Investigators. Validating the HERDOO2 rule to guide treatment duration for women with unprovoked venous thrombosis: multinational prospective cohort management study. *BMJ.* 2017;356:j1065. <https://doi.org/10.1136/bmj.j1065>
- [10] Morange PE, Suchon P, Tréguët DA. Genetics of venous thrombosis: update in 2015. *Thromb Haemost.* 2015;114:910–9.
- [11] Baglin T, Luddington R, Brown K, Baglin C. Incidence of recurrent venous thromboembolism in relation to clinical and thrombophilic risk factors: prospective cohort study. *Lancet.* 2003;362:523–6.
- [12] Stevens H, Peter K, Tran H, McFadyen J. Predicting the risk of recurrent venous thromboembolism: current challenges and future opportunities. *J Clin Med.* 2020;9:1582. <https://doi.org/10.3390/jcm9051582>
- [13] Timp JF, Braekkan SK, Lijfering WM, van Hylckama Vlieg A, Hansen JB, Rosendaal FR, le Cessie S, Cannegieter SC. Prediction of recurrent venous thrombosis in all patients with a first venous thrombotic event: the Leiden Thrombosis Recurrence Risk Prediction model (L-TRRIP). *PLOS Med.* 2019;16:e1002883. <https://doi.org/10.1371/journal.pmed.1002883>
- [14] Mattei AL, Bailly N, Meissner A. DNA methylation: a historical perspective. *Trends Genet.* 2022;38:676–707.
- [15] Portela A, Esteller M. Epigenetic modifications and human disease. *Nat Biotechnol.* 2010;28:1057–68.
- [16] Cappozzo A, McCrory C, Robinson O, Freni Sterrantino A, Sacerdote C, Krogh V, Panico S, Tumino R, Iacoviello L, Ricceri F, Sieri S, Chiodini P, McKay GJ, McKnight AJ, Kee F, Young IS, McGuinness B, Crimmins EM, Arpawong TE, Kenny RA, et al. A blood DNA methylation biomarker for predicting short-term risk of cardiovascular events. *Clin Epigenet.* 2022;14:121. <https://doi.org/10.1186/s13148-022-01341-4>
- [17] Cuadrat RRC, Kratzer A, Arnal HG, Rathgeber AC, Wreczycka K, Blume A, Gündüz IB, Ebenal V, Mauno T, Osberg B, Moobed M, Hartung J, Jakobs K, Seppelt C, Meteva D, Haghikia A, Leistner DM, Landmesser U, Akalin A. Cardiovascular disease biomarkers derived from circulating cell-free DNA methylation. *NAR Genom Bioinform.* 2023;5:lqad061. <https://doi.org/10.1093/nargab/lqad061>

- [18] Jurmeister P, Glöß S, Roller R, Leitheiser M, Schmid S, Mochmann LH, Payá Capilla E, Fritz R, Dittmayer C, Friedrich C, Thieme A, Keyl P, Jarosch A, Schallenberg S, Bläker H, Hoffmann I, Vollbrecht C, Lehmann A, Hummel M, Heim D, et al. DNA methylation-based classification of sinonasal tumors. *Nat Commun.* 2022;13:7148. <https://doi.org/10.1038/s41467-022-34815-3>
- [19] van Dongen J, Nivard MG, Willemsen G, Hottenga JJ, Helmer Q, Dolan CV, Ehli EA, Davies GE, van Iterson M, Breeze CE, Beck S, BIOS Consortium, Suchiman HE, Jansen R, van Meurs JB, Heijmans BT, Slagboom PE, Boomsma DI. Genetic and environmental influences interact with age and sex in shaping the human methylome. *Nat Commun.* 2016;7:11115. <https://doi.org/10.1038/ncomms11115>
- [20] Lopes-Ramos CM, Chen CY, Kuijjer ML, Paulson JN, Sonawane AR, Fagny M, Platig J, Glass K, Quackenbush J, DeMeo DL. Sex differences in gene expression and regulatory networks across 29 human tissues. *Cell Rep.* 2020;31:107795. <https://doi.org/10.1016/j.celrep.2020.107795>
- [21] Gatev E, Inkster AM, Negri GL, Konwar C, Lussier AA, Skakkebaek A, Sokolowski MB, Gravholt CH, Dunn EC, Kobor MS, Aristizabal MJ. Autosomal sex-associated co-methylated regions predict biological sex from DNA methylation. *Nucleic Acids Res.* 2021;49:9097–116.
- [22] Andrews SV, Yang IJ, Froehlich K, Oskotsky T, Sirota M. Large-scale placenta DNA methylation integrated analysis reveals fetal sex-specific differentially methylated CpG sites and regions. *Sci Rep.* 2022;12:9396. <https://doi.org/10.1038/s41598-022-13544-z>
- [23] Liu J, Morgan M, Hutchison K, Calhoun VD. A study of the influence of sex on genome wide methylation. *PLoS One.* 2010;5:e10028. <https://doi.org/10.1371/journal.pone.0010028>
- [24] Maher B. Personal genomes: The case of the missing heritability. *Nature.* 2008;456:18–21.
- [25] Feinberg AP. The key role of epigenetics in human disease prevention and mitigation. *N Engl J Med.* 2018;378:1323–34.
- [26] Ahmad A, Sundquist K, Palmér K, Svensson PJ, Sundquist J, Memon AA. Risk prediction of recurrent venous thromboembolism: a multiple genetic risk model. *J Thromb Thrombolysis.* 2019;47:216–26.
- [27] Rodger MA, Scarvelis D, Kahn SR, Wells PS, Anderson DA, Chagnon I, Le Gal G, Gandara E, Solymoss S, Sabri E, Kovacs J, Kovacs MJ. Long-term risk of venous thrombosis after stopping anticoagulants for a first unprovoked event: a multi-national cohort. *Thromb Res.* 2016;143:152–8.
- [28] LE Gal G, Carrier M, Kovacs MJ, Betancourt MT, Kahn SR, Wells PS, Anderson DA, Chagnon I, Solymoss S, Crowther M, Righini M, Delluc A, White RH, Vickers L, Rodger M. Residual vein obstruction as a predictor for recurrent thromboembolic events after a first unprovoked episode: data from the REVERSE cohort study. *J Thromb Haemost.* 2011;9:1126–32.
- [29] Min JL, Hemani G, Smith GD, Relton C, Suderman M, Meffli: efficient normalization and analysis of very large DNA methylation datasets. *Bioinformatics.* 2018;34:3983–9.
- [30] Pidsley R, Wong CCY, Volta M, Lunnon K, Mill J, Schalkwyk CL. A data-driven approach to preprocessing Illumina 450K methylation array data. *BMC Genomics.* 2013;14:293. <https://doi.org/10.1186/1471-2164-14-293>
- [31] Pidsley R, Zotenko E, Peters TJ, Lawrence MG, Risbridger GP, Molloy P, Van Dijk S, Muhlhäuser B, Stirzaker C, Clark SJ. Critical evaluation of the Illumina MethylationEPIC BeadChip microarray for whole-genome DNA methylation profiling. *Genome Biol.* 2016;17:208. <https://doi.org/10.1186/s13059-016-1066-1>
- [32] Chen YA, Lemire M, Choufani S, Butcher DT, Grafodatskaya D, Zanke BW, Gallinger S, Hudson TJ, Weksberg R. Discovery of cross-reactive probes and polymorphic CpGs in the Illumina Infinium HumanMethylation450 microarray. *Epigenetics.* 2013;8:203–9.
- [33] McCartney DL, Walker RM, Morris SW, McIntosh AM, Porteous DJ, Evans KL. Identification of polymorphic and off-target probe binding sites on the Illumina Infinium MethylationEPIC BeadChip. *Genom Data.* 2016;9:22–4.
- [34] Du P, Zhang X, Huang CC, Jafari N, Kibbe WA, Hou L, Lin SM. Comparison of Beta-value and M-value methods for quantifying methylation levels by microarray analysis. *BMC Bioinformatics.* 2010;11:587. <https://doi.org/10.1186/1471-2105-11-587>
- [35] Truong V, Huang S, Dennis J, Lemire M, Zwingerman N, Aïssi D, Kassam I, Perret C, Wells P, Morange PE, Wilson M, Tréguët DA, Gagnon F. Blood triglyceride levels are associated with DNA methylation at the serine metabolism gene PHGDH. *Sci Rep.* 2017;7:11207. <https://doi.org/10.1038/s41598-017-09552-z>
- [36] Teschendorff AE, Breeze CE, Zheng SC, Beck S. A comparison of reference-based algorithms for correcting cell-type heterogeneity in epigenome-wide association studies. *BMC Bioinformatics.* 2017;18:105. <https://doi.org/10.1186/s12859-017-1511-5>
- [37] Devlin B, Roeder K. Genomic control for association studies. *Biometrics.* 1999;55:997–1004.
- [38] van Iterson M, van Zwet EW, BIOS Consortium, Heijmans BT. Controlling bias and inflation in epigenome- and transcriptome-wide association studies using the empirical null distribution. *Genome Biol.* 2017;18:19. <https://doi.org/10.1186/s13059-016-1131-9>
- [39] Therneau TM, Grambsch PM. *Modeling Survival Data: Extending the Cox Model.* New York: Springer; 2000. ISBN 0-387-98784-3.
- [40] Baron RM, Kenny DA. The moderator-mediator variable distinction in social psychological research: conceptual, strategic, and statistical considerations. *J Pers Soc Psychol.* 1986;51:1173–82.
- [41] Kassambara A, Kosinski M. Przemyslaw Biecek. *survminer: Drawing Survival Curves using 'ggplot2'.* R package version 0.4.9. 2021. <https://CRAN.R-project.org/package=survminer>. [accessed February 10, 2025].
- [42] Sun L, Craiu RV, Paterson AD, Bull SB. Stratified false discovery control for large-scale hypothesis testing with application to genome-wide association studies. *Genet Epidemiol.* 2006;30:519–30.
- [43] Sun L, Rommens JM, Corvol H, Li W, Li X, Chiang TA, Lin F, Dorfman R, Busson PF, Parekh RV, Zelenika D, Blackman SM, Corey M, Doshi VK, Henderson L, Naughton KM, O'Neal WK, Pace RG, Stonebraker JR, Wood SD, et al. Multiple apical plasma membrane constituents are associated with susceptibility to meconium ileus in individuals with cystic fibrosis. *Nat Genet.* 2012;44:562–9.
- [44] Hanley JA, Miettinen OS. Fitting smooth-in-time prognostic risk functions via logistic regression. *Int J Biostat.* 2009;5. <https://doi.org/10.2202/1557-4679.1125>
- [45] Lin DY, Wei LJ. The robust inference for the Cox proportional hazards model. *J Am Stat Assoc.* 1989;84:1074–8.
- [46] Friedman J, Hastie T, Tibshirani R. Regularization paths for generalized linear models via coordinate descent. *J Stat Softw.* 2010;33:1–22.
- [47] Durinck S, Spellman P, Birney E, Huber W. Mapping identifiers for the integration of genomic datasets with the R/Bioconductor package biomaRt. *Nat Protoc.* 2009;4:1184–91.
- [48] Min JL, Hemani G, Hannon E, Dekkers KF, Castillo-Fernandez J, Luijk R, Carnero-Montoro E, Lawson DJ, Burrows K, Suderman M, Bretherick AD, Richardson TG, Klughammer J, Iotchkova V, Sharp G, Al Khleifat A, Shatunov A, Iacoangeli A, McArdle WL, Ho KM, et al. Genomic and phenotypic insights from an atlas of genetic effects on DNA methylation. *Nat Genet.* 2021;53:1311–21.
- [49] Vösa U, Claringbould A, Westra HJ, Bonder MJ, Deelen P, Zeng B, Kirsten H, Saha A, Kreuzhuber R, Yazar S, Brugge H, Oelen R, de Vries DH, van der Wijst MGP, Kasela S, Pervjakova N, Alves I, Favé MJ, Agbessi M, Christiansen MW, et al. Large-scale cis- and trans-eQTL analyses identify thousands of genetic loci and polygenic scores that regulate blood gene expression. *Nat Genet.* 2021;53:1300–10.

- [50] GTEx Consortium. The genotype-tissue expression (GTEx) project. *Nat Genet.* 2013;45:580–5.
- [51] Frangogiannis NG. Fact and fiction about fibroblast to endothelium conversion: semantics and substance of cellular identity. *Circulation.* 2020;142:1663–6.
- [52] Fan G, Jin Z, Wang K, Yang H, Wang J, Li Y, Chen B, Zhang H. Identification of four hub genes in venous thromboembolism via weighted gene coexpression network analysis. *BMC Cardiovasc Disord.* 2021;21:577. <https://doi.org/10.1186/s12872-021-02409-4>
- [53] Greninger AL, Knudsen GM, Betegon M, Burlingame AL, DeRisi JL. ACBD3 interaction with TBC1 domain 22 protein is differentially affected by enteroviral and koboviral 3A protein binding. *mBio.* 2013;4:e00098-13. <https://doi.org/10.1128/mBio.00098-13>
- [54] Tobias IS, Lazauskas KK, Siu J, Costa PB, Coburn JW, Galpin AJ. Sex and fiber type independently influence AMPK, TBC1D1, and TBC1D4 at rest and during recovery from high-intensity exercise in humans. *J Appl Physiol.* 2020;128:350–61.
- [55] Xu J, McPherson PS. Regulation of DENND3, the exchange factor for the small GTPase Rab12 through an intramolecular interaction. *J Biol Chem.* 2017;292:7274–82.
- [56] Lamber EP, Siedenburg AC, Barr FA. Rab regulation by GEFs and GAPs during membrane traffic. *Curr Opin Cell Biol.* 2019;59:34–9.
- [57] Lewis DA, Suchindran S, Beckman MG, Hooper WC, Grant AM, Heit JA, Manco-Johnson M, Moll S, Philipp CS, Kenney K, De Staercke C, Pyle ME, Chi JT, Ortel TL. Whole blood gene expression profiles distinguish clinical phenotypes of venous thromboembolism. *Thromb Res.* 2015;135:659–65.
- [58] Biesemann A, Gorontzi A, Barr F, Gerke V. Rab35 protein regulates evoked exocytosis of endothelial Weibel–Palade bodies. *J Biol Chem.* 2017;292:11631–40.
- [59] Montes R, Guruceaga E, González-Porras JR, Reverter JC, Marco P, Pina E, Páramo JA, Hermida J, Lecumberri R, Grupo de Trabajo de Patología Trombótica de la Sociedad Española de Trombosis y Hemostasia (SETH). Identification of new markers of recurrence in patients with unprovoked deep vein thrombosis by gene expression profiling: the retro study. *Eur J Haematol.* 2016;97:128–36.
- [60] Sabater-Lleal M, Huang J, Chasman D, Naitza S, Dehghan A, Johnson AD, Teumer A, Reiner AP, Folkersen L, Basu S, Rudnicka AR, Trompet S, Mälarstig A, Baumert J, Bis JC, Guo X, Hottenga JJ, Shin SY, Lopez LM, Lahti J, et al. Multiethnic meta-analysis of genome-wide association studies in >100 000 subjects identifies 23 fibrinogen-associated loci but no strong evidence of a causal association between circulating fibrinogen and cardiovascular disease. *Circulation.* 2013;128:1310–24.
- [61] Liu Y, Deng X, Zhu F, Zhu W, Wang Z. High fibrinogen and mixed proximal and distal thrombosis are associated with the risk of residual venous thrombosis in patients with posttraumatic deep vein thrombosis. *Front Cardiovasc Med.* 2023;10:1003197. <https://doi.org/10.3389/fcvm.2023.1003197>
- [62] Koster T, Rosendaal FR, Reitsma PH, Van Der Velden PA, Briet E, Vandenbroucke JP. Factor VII and fibrinogen levels as risk factors for venous thrombosis. A case-control study of plasma levels and DNA polymorphisms – the Leiden Thrombophilia Study (LETS). *Thromb Haemost.* 1994;71:719–22.
- [63] Li N, Wu Z, Ma C. ZHX2 in health and disease. *Front Oncol.* 2022;12:1038890. <https://doi.org/10.3389/fonc.2022.1038890>
- [64] Senatore E, Chiuso F, Rinaldi L, Intartaglia D, Delle Donne R, Pedone E, Catalanotti B, Pirone L, Fiorillo B, Moraca F, Giamundo G, Scala G, Raffaeiner A, Torres-Quesada O, Stefan E, Kwiatkowski M, van Pijkeren A, Morleo M, Franco B, Garbi C, et al. The TBC1D31/praja2 complex controls primary ciliogenesis through PKA-directed OFD1 ubiquitylation. *EMBO J.* 2021;40:e106503. <https://doi.org/10.15252/emj.2020106503>
- [65] Maciel NIG, Filiú-Braga LDC, Neves FAR, Rego EM, Lucena-Araujo AR, Saldanha-Araujo F. Low expression of ZHX1 and ZHX2 impacts on the prognosis of chronic lymphocytic leukemia. *Biomark Res.* 2021;9:10. <https://doi.org/10.1186/s40364-021-00263-2>
- [66] Creasy KT, Jiang J, Ren H, Peterson ML, Spear BT. Zinc fingers and homeoboxes 2 (Zhx2) regulates sexually dimorphic Cyp gene expression in the adult mouse liver. *Gene Expr.* 2016;17:7–17.
- [67] Fang W, Liao C, Shi R, Simon JM, Ptacek TS, Zurlo G, Ye Y, Han L, Fan C, Bao L, Ortiz CL, Lin HR, Manocha U, Luo W, Peng Y, Kim WY, Yang LW, Zhang Q. ZHX2 promotes HIF1 $\alpha$  oncogenic signaling in triple-negative breast cancer. *elife.* 2021;10:e70412. <https://doi.org/10.7554/eLife.70412>
- [68] Evans CE. Hypoxia and HIF activation as a possible link between sepsis and thrombosis. *Thromb J.* 2019;17:16. <https://doi.org/10.1186/s12959-019-0205-9>
- [69] Wang X, Memon AA, Palmér K, Svensson PJ, Sundquist J, Sundquist K. The association between blood-based global DNA methylation and venous thromboembolism. *Semin Thromb Hemost.* 2021;47:662–8.
- [70] Lewis DA, Stashenko GJ, Akay OM, Price LI, Owzar K, Ginsburg GS, Chi JT, Ortel TL. Whole blood gene expression analyses in patients with single versus recurrent venous thromboembolism. *Thromb Res.* 2011;128:536–40.
- [71] Battram T, Gaunt TR, Relton CL, Timpson NJ, Hemani G. A comparison of the genes and genesets identified by GWAS and EWAS of fifteen complex traits. *Nat Commun.* 2022;13:7816. <https://doi.org/10.1038/s41467-022-35037-3> [accessed February 10, 2024].
- [72] Luo Z, Liu R, Sun B, Zhou X, Liu Z, Zhou H, Xu H, Li X, Zhang W. Identification of gene modules associated with warfarin dosage by a genome-wide DNA methylation study. *Pharmazie.* 2018;73:288–93.
- [73] Le Gal G, Kovacs MJ, Carrier M, Do K, Kahn SR, Wells PS, Anderson DA, Chagnon I, Solymoss S, Crowther M, Righini M, Lacut K, White RH, Vickers L, Rodger M. Risk of recurrent venous thromboembolism after a first oestrogen-associated episode: data from the REVERSE cohort study. *Thromb Haemost.* 2010;104:498–503.

#### SUPPLEMENTARY MATERIAL

The online version contains supplementary material available at <https://doi.org/10.1016/j.jtha.2025.01.004>.