University of Dundee

Integrated GWAS and Gene Expression Suggest ORM1 as a Potential Regulator of Plasma Levels of Cell-Free DNA and Thrombosis Risk

Lopez, Sonia; Martinez-Perez, Angel; Rodriguez-Rius, Alba; Viñuela, Ana; Brown, Andrew A.; Martin-Fernandez, Laura

[Link to publication in Discovery Research Portal](#)

Integrated GWAS and Gene Expression Suggest *ORM1* as a Potential Regulator of Plasma Levels of Cell-Free DNA and Thrombosis Risk

Sonia Lopez[1] Angel Martinez-Perez[1] Alba Rodriguez-Rius[1] Ana Viñuela[2] Andrew A. Brown[3]
Laura Martin-Fernandez[1,4,5,6] Noelia Vilalta[7] Marc Arús[7] Nikolaos I. Panousis[8,9] Alfonso Buil[10]
Maria Sabater-Lleal[1,11] Juan Carlos Souto[7] Emmanouil T. Dermitzakis[9] Jose Manuel Soria[1]

[1] Genomics of Complex Diseases Unit, Research Institute Hospital de la Santa Creu i Sant Pau, IIB Sant Pau, Barcelona, Spain
[2] Biosciences Institute, Faculty of Medicine, Newcastle University, Newcastle Upon Tyne, United Kingdom
[3] Population Health and Genomics, University of Dundee, Dundee, Scotland, United Kingdom
[4] Fundación Española de Trombosis y Hemostasia (FETH), Madrid, Spain
[5] Congenital Coagulopathies Laboratory, Banc de Sang i Teixits, Barcelona, Spain
[6] Transfusional Medicine, Vall d'Hebron Research Institute, Universitat Autònoma de Barcelona (VHIR-UAB), Barcelona, Spain
[7] Haemostasis and Thrombosis Unit, Department of Hematology, Hospital de la Santa Creu i Sant Pau, Barcelona, Spain

Address for correspondence Sonia Lopez, PhD, Genomics of Complex Diseases Unit, Research Institute Hospital de la Santa Creu i Sant Pau, IIB Sant Pau, Sant Quintí 77-79, 3rd Floor, 08041 Barcelona, Spain (e-mail: slopezm@santpau.cat).

[8] Wellcome Sanger Institute, Wellcome Genome Campus, Hinxton, South Cambridgeshire, United Kingdom
[9] Department of Genetic Medicine and Development, University of Geneva, Geneva, Switzerland
[10] Institute of Biological Psychiatry, Mental Health Sct. Hans Hospital, Roskilde, Denmark
[11] Cardiovascular Medicine Unit, Department of Medicine, Center for Molecular Medicine, Karolinska Institutet, Stockholm, Sweden

Thromb Haemost

Abstract

Plasma cell-free DNA (cfDNA) is a surrogate marker of neutrophil extracellular traps (NETs) that contribute to immunothrombosis. There is growing interest about the mechanisms underlying NET formation and elevated cfDNA, but little is known about the factors involved. We aimed to identify genes involved in the regulation of cfDNA levels using data from the Genetic Analysis of Idiopathic Thrombophilia (GAIT-2) Project.

Imputed genotypes, whole blood RNA-Seq data, and plasma cfDNA quantification were available for 935 of the GAIT-2 participants from 35 families with idiopathic thrombophilia. We performed heritability and GWAS analysis for cfDNA. The heritability of cfDNA was 0.26 ($p = 3.7 \times 10^{-6}$), while the GWAS identified a significant association (rs1687391, $p = 3.55 \times 10^{-10}$) near the *ORM1* gene, on chromosome 9. An eQTL (expression quantitative trait loci) analysis revealed a significant association between the lead GWAS variant and the expression of *ORM1* in whole blood ($p = 6.14 \times 10^{-9}$). Additionally, *ORM1* expression correlated with levels of cfDNA ($p = 4.38 \times 10^{-4}$). Finally, genetic correlation analysis between cfDNA and thrombosis identified a suggestive association ($\rho_g = 0.43$, $p = 0.089$).

All in all, we show evidence of the role of *ORM1* in regulating cfDNA levels in plasma, which might contribute to the susceptibility to thrombosis through mechanisms of immunothrombosis.

Keywords
► immunothrombosis
► neutrophil extracellular traps
► α-1-acid glycoprotein 1
► expression quantitative trait loci

Visual Summary. Schematic overview illustrating the results for cfDNA in GAIT-2. Our results suggest that plasma levels of cfDNA are under the influence of the genetic variability close to or within *ORM1*, which could contribute to thrombosis risk through mechanisms that might be related to immunothrombosis. In particular, the minor allele rs1687391-T of the lead SNP near *ORM1* identified in the GWAS for cfDNA in our study increases (+) the amount of cfDNA in plasma, probably through the decrease (−) of the *ORM1* expression observed for this allele. Consistently, *ORM1* decreases (−) cfDNA levels, suggesting a protective effect against elevated cfDNA. The elevated amount of cfDNA in plasma could contribute (+) to thrombosis risk through the activation of blood coagulation and fibrin formation.
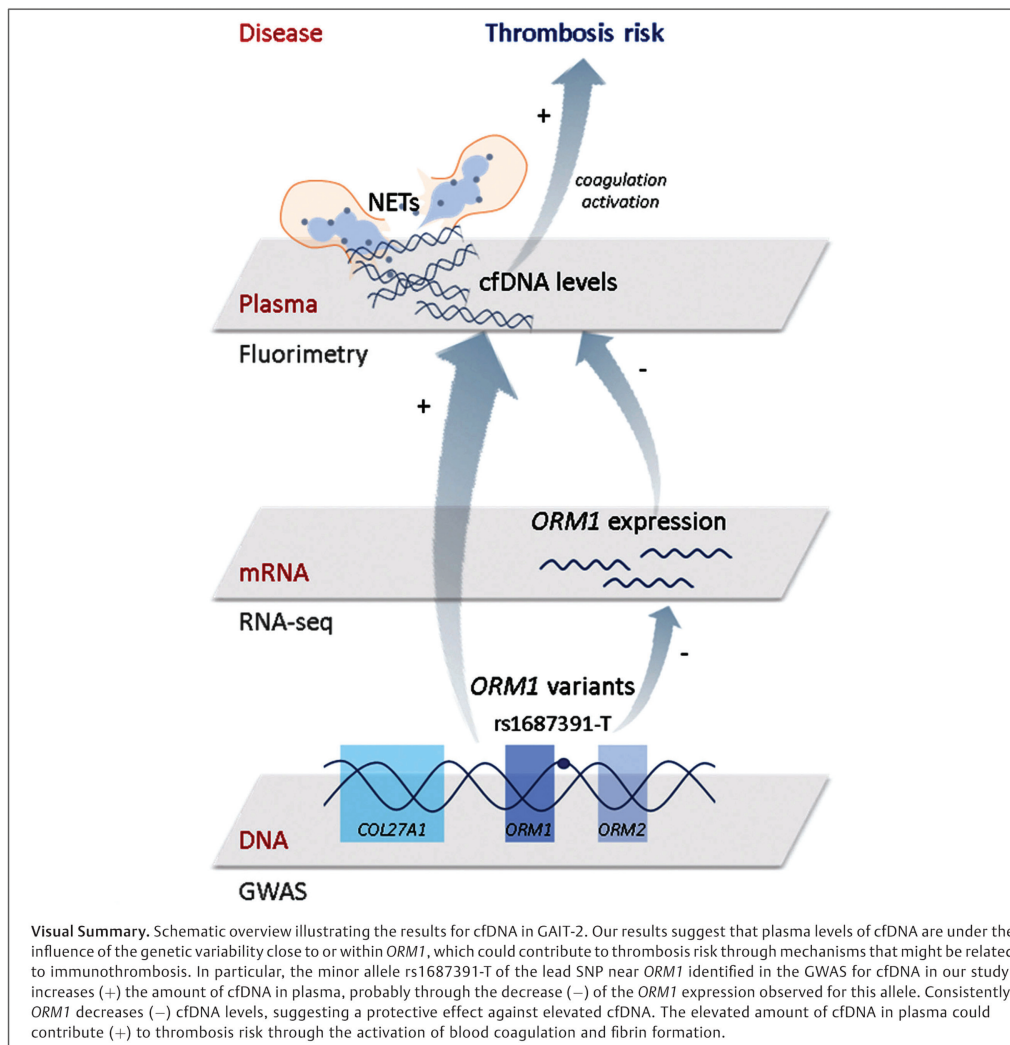
Introduction

Cell-free DNAs (cfDNAs) are extracellular small double-stranded DNA molecules found circulating in blood. They have gained importance recently in a wide variety of pathologies as a promising disease biomarker.[1,2] cfDNA is considered a surrogate marker of neutrophil extracellular traps (NETs), since circulating DNA is its major structural component that increases with NET deposition. NETs consist of a complex extracellular network of nuclear DNA fibers, decorated with histones and antimicrobial proteins that trap and kill pathogens to prevent microbial dissemination into the tissue or the vasculature. The release of DNA into the circulation from neutrophils was described in 2004 by Brinkmann et al[3] as a novel mechanism of defense of the innate immune system to kill pathogens. This mechanism is called NETosis. It is a unique form of programmed cell death by which active neutrophils extrude their cell material to form NETs.

Recently, NETs have been the focus of intensive investigation due to their role in innate immunity and its association with various diseases, either of infection origin or sterile inflammation, such as sepsis,[4] cancer,[5] autoimmune diseases,[6] and atherosclerosis.[7] Elevated levels of NETs, through excessive production or defective removal of NETs, may lead to these clinical conditions.[8] Moreover, damaging effects and impaired tissue function may correlate with NET formation.[9] NETs have been identified as an important link between innate immunity, inflammation, and coagulation, which describes the concept of immunothrombosis.[10] In vitro and in vivo studies have shown that DNA and histones of NETs provide scaffolding for platelet adhesion and promote fibrin deposition in the vasculature by enhancing the activity of blood coagulation factors and promoting thrombosis.[11–13] Accordingly, cfDNA levels have been found elevated in deep vein thrombosis (DVT) both in animals and humans.[11,14–16] Furthermore, in vivo DNase 1 treatment is effective in eliminating cfDNA from the blood stream and prevent DVT in mice.[16] Also, elevated levels of cfDNA have been able to predict mortality in venous thrombosis (VT) patients aged >65 years.[17] Most recent findings relate the systemic coagulopathy and thromboembolic events, frequently observed in patients with severe acute respiratory syndrome-coronavirus 2 infection, to elevated NET formation.[18] Indeed,

patients with coronavirus disease 2019 (COVID-19) have shown elevated serum levels of cfDNA and other biomarkers of NETs that correlate with disease severity.[19]

cfDNA of NETs is a biomarker for disease, but may also act as a causal pathogenic mediator that contributes to the procoagulant phenotype observed in various diseases. Thus, the identification of the mechanisms that control DNA release and maintenance into the circulation could help to understand the functional role of cfDNA in physiological and pathological conditions. Also, it would help provide insights for the development of therapies against harmful amounts of cfDNA.

The mechanisms and genetic factors that predispose to elevated levels of cfDNA are largely unknown. To the best of our knowledge, there is only one study that aimed to identify the genetic determinants underlying the control of cfDNA levels. In this study, the gene *UDP glucuronosyltransferase family 1 member A1 (UGT1A1)* on chromosome 2 was identified as a regulator of cfDNA levels in serum by a genome-wide association study (GWAS) in a cohort of young adults participating in the Cardiovascular Risk in Young Finns Study.[20]

In our study, we conducted a GWAS to identify genes that regulate plasma levels of cfDNA using related individuals of the Genetic Analysis of Idiopathic Thrombophilia (GAIT-2) Project. In addition, we attempted to identify a candidate effector transcript by using gene expression data of these individuals and elucidate the role of the locus mediating the GWAS effect.

Methods

Population
The GAIT-2 project included 935 individuals from 35 extended Spanish families selected through a proband with idiopathic thrombophilia. All of the individuals were recruited at the Hospital de la Santa Creu i Sant Pau (Barcelona, Spain) from 2006 to 2010. The number of members per family ranged from 11 to 79 (mean ± standard deviation [SD]: 33 individuals ± 17). All of the families included in the project were required to have at least 10 living individuals in three or more generations. Idiopathic thrombophilia was defined as: (1) recurrent venous and/or arterial thrombosis (AT; with at least one spontaneous thrombotic event), (2) a single spontaneous thrombotic event with a first-degree relative also affected, or (3) early-onset thrombosis (<45 years), with the causes of thrombophilia excluded at inclusion. Accordingly, probands with deficiencies of coagulation proteins or cofactors including protein S, protein C, antithrombin, plasminogen, factor V Leiden, heparin cofactor II, activated protein C resistance, dysfibrinogenemia, lupus anticoagulant, and antiphospholipid antibodies were not included. In the GAIT-2 study, 118 individuals had been diagnosed with thrombosis, with a total of 187 thrombotic events (121 venous and 66 arterial) registered. Among the individuals with thrombosis, 71 had VT only, 33 had AT only, and 14 had both VT and AT. Detailed composition of the families and the collection of lifestyle, medical, and family history data have been previously described.[21]

Laboratory Measurements

Sample Recruitment
Thrombophilic participants were not using oral anticoagulants and antiplatelet drugs at the time of sampling (not used at least 15 days before the extraction), and were not under heparin treatment (for at least 24 hours before). Blood was obtained at least 3 months after the thrombotic event in affected individuals and at least 1 month after an acute inflammatory event to minimize the influence of the acute phase. Blood was collected from the antecubital vein from fasting subjects by venipuncture in 1/10 volume of 0.129 mol/L sodium citrate. After centrifugation at 2,000 g for 25 minutes at room temperature, platelet-poor plasma was obtained and used for the biochemical assays. The remaining plasma was stored at −80°C until used. The fraction of blood cells was washed (1:1) with sodium chloride 0.9% and centrifuged for 10 minutes at 1,580 g to obtain the buffy coat, which was stored at −20°C until DNA extraction. In addition, 2.5 mL of blood was collected in a PAXgene Blood RNA Tube (QIAGEN GmbH, Germany) and stored at −80°C until RNA extraction according to the manufacturer's instructions.

DNA Isolation
Total DNA from the buffy coats of GAIT-2 samples was isolated by a standard salting-out procedure.[22] An aliquot of DNA was stored at 4°C until used for subsequent genotyping of the subjects. The remaining genetic material was kept at −20°C for long-term storage.

RNA Isolation
For each individual of the GAIT-2 study, intracellular RNA from whole blood collected in PAXgene Blood RNA Tubes was isolated following the protocol of the PAXgene Blood RNA Kit (QIAGEN GmbH, Germany). Total RNA was stored at −80°C until used.

cfDNA Measurement
The cfDNA was quantified in duplicate in plasma from each individual by fluorescence as previously documented by Fuchs et al.[11,15] In brief, 100 μL of 1/10 diluted plasma in sterile ultra-pure grade phosphate-buffered saline (PBS; Amrek) was mixed with 100 μL of PBS containing Sytox Green (Life Technologies) at 2 μM final concentration to label DNA. Autofluorescence was background and determined in samples mixed with PBS without Sytox Green. Samples were incubated for 15 minutes with Sytox Green in 96-well microplates (FluoroNunc F96 MicroWell plates, Nunc) protected from light. After incubation, fluorescence was measured at 485 nm excitation and 538 nm emission wavelengths in a microplate reader fluorometer (Tecan Infinite 200Pro). cfDNA concentrations were calculated through a standard curve of known DNA concentrations (Human genomic DNA 100 μg/mL, Novagen) and expressed in ng/mL. Measurements of cfDNA were available in 934

GAIT-2 subjects. As differences in mean and dispersion of cfDNA values (ng/mL) were observed among plates, the median value of cfDNA of each plate was subtracted from the value of each sample, and divided by the SD of each plate for normalization. Plate-normalized cfDNA values were used in the analyses in which cfDNA was a covariate, while original values were used when mixed models including plate as random effect were tested. The assessed inter- and intra-assay variation coefficients for cfDNA measurement were 6.6 and 3.4%, respectively.

Genotyping and Imputation of Samples

All of the individuals in the GAIT-2 study were genotyped with a combination of two human genotyping arrays from Illumina (San Diego, California, United States): HumanOmniExpressExome-8v1.2 (324 individuals and coverage 964,193 variants) and HumanCoreExome-12v1.1 (610 individuals and coverage 542,585 variants). The variants with a genotype call rate <0.98, a minor allele frequency (MAF) <0.01, or failing the Hardy-Weinberg equilibrium (HWE) test ($p < 10^{-6}$) were excluded from the analysis. HWE was calculated by a standard χ^2-test with 1 degree of freedom and it was tested using parental data only. After filtering the datasets, we merged the data and obtained 485,672 variants in all the samples. We then imputed genotypes to the 1000 Genomes Project phase 1 panel. After imputing, variants with MAF <0.005 were excluded. After quality control (QC), 10,844,567 variants remained. Imputed variants were then mapped along the genome according to the human genome assembly GRCh37 (hg19) from the Genome Reference Consortium.

RNA Sequencing

The messenger RNA (mRNA) levels of the genes expressed in whole blood were obtained in the GAIT-2 population by RNA sequencing (RNA-Seq). In brief, 928 GAIT-2 individuals with high-quality RNA samples were eligible for RNA-Seq analysis. RNA samples were prepared for sequencing with the Illumina TruSeq sample preparation kit (Illumina, San Diego, California, United States) according to manufacturer's instructions and were sequenced on a HiSeq2000 Sequencing System (Illumina). The 49-bp sequenced paired-end reads were mapped to the GRCh37 reference genome with the Burrows–Wheeler Aligner v0.5.9.[23] We used genes defined as protein coding in the GENCODE 10 annotation. We excluded samples that failed in the library preparation or sequence process, those with less than 10 million reads, and those with sequences that did not correspond to actual genotype data. After QC, 16,748 mRNAs from whole blood of 915 GAIT-2 individuals were available for further transcriptome analyses. RNA-Seq data were normalized by the effect of technical covariates, age, sex, and blood cell counts (by eliminating the variance explained by the first four principal components in the principal component analysis) and then, used in the analyses as residualized values.

Statistical Analyses

All analyses were calculated using the R (v.3.6.0; http://www.R-project.org) package Solarius,[24] as the interface of the Sequential Oligogenic Linkage Analysis Routines (SOLAR-ECLIPSE, version 8.1.1),[25] which is specifically designed for the genetic analyses in extended pedigrees.

Heritability

The heritability for cfDNA levels is the estimated proportion of total variance of the cfDNA concentration that is attributable to the additive effects of genes. It was calculated in GAIT-2 using variance components methods adjusted for covariates age, age squared (age^2), and sex.

Genetic and Environmental Correlation of cfDNA with Thrombosis

Genetic and environmental correlation coefficients were quantified in GAIT-2 between cfDNA concentration and thrombosis (including VT, AT, and both), as well as with VT only and AT only. Genetic correlation is of special interest because it quantifies the additive genetic effects that are common to both traits. For that, we tested a bivariate variance component model adjusted for the significant covariates specific for each trait. Significance was calculated using the likelihood ratio test (LRT).[26] All of the models were corrected for family structure and proband ascertainment bias. This ascertainment scheme was used to obtain the best estimation for general population parameters.

Genome-Wide Association Study

A GWAS with the imputed genotypes was conducted in two phases in GAIT-2 for transformed values of cfDNA levels to identify single-nucleotide polymorphisms (SNPs) regulating interindividual variance of this quantitative trait. First, we performed a fast screen of variants with Matrix eQTL[27] using a linear model, adjusted for age, age^2, sex, and including the kinship matrix to correct for relatedness. Variants with $p < 10^{-3}$ were then recalculated in a second phase with the Solarius package[24] using the same covariates. All p-values were calculated by LRT.[26] Only the p-values calculated with SOLAR are reported and are represented in a Manhattan plot. The genome-wide significance threshold for SNP–phenotype associations was set at the commonly accepted level of $p \leq 5 \times 10^{-8}$. Regional plots of the significant associated loci with cfDNA levels were generated by LocusZoom.[28]

Conditional Association Analysis

A second round of SNP–phenotype association analysis for cfDNA levels in GAIT-2 was performed by further conditioning on the top associated SNP identified in the GWAS to identify secondary association signals in a particular genomic region.

cis-eQTL Analysis for Candidate Genes

The identification of SNPs associated with the variability of gene expression levels are known as expression quantitative trait loci (eQTLs), which helps to understand the functional role of the SNP–phenotype associations of a GWAS. cis-eQTL variants likely affecting proximal gene expression in whole blood were analyzed for the genes of interest in our study (ORM1, ORM2, and COL27A1). For that, linear mixed models

adjusted for the significant covariates age, age^2, sex, and kinship matrix were conducted in GAIT-2 using FastQTL[29] to evaluate associations between imputed SNPs located within 1 megabase (Mb) upstream or downstream of the corresponding gene transcription start site and expression levels of the candidate gene transcripts. p-Values were adjusted for multiple testing by applying the Bonferroni correction. Regional plots of the significant eQTLs were generated by LocusZoom.[28] The R package LocusCompareR was used to generate plots for visualization of GWAS-eQTL colocalization events. The cis-eQTLs identified in GAIT-2 for candidate genes were compared with data from the Genotype-Tissue Expression (GTEx) Portal V8.

Associations of cfDNA Levels with Gene Expression of Candidate Genes

Linear associations were performed using mixed models between cfDNA levels and the gene expression in blood of the candidate genes in our study, considering significance at a Benjamini–Hochberg (B–H) false discovery rate (FDR) <0.05. This analysis is useful for prioritizing the best candidate genes to explain associations with circulating cfDNA levels.

Pathway Enrichment Analysis

Biological process and pathway analysis can increase the understanding of gene function. Accordingly, the Database for Annotation, Visualization and Integrated Discovery (DAVID) v6.8 (https://david.ncifcrf.gov/) is an algorithm useful for discovering enriched functional-related gene groups for pathway analysis.[30] DAVID annotation tool was used to extract biological meaning from large gene lists obtained by our RNA-Seq.

Results

Population Description

Characteristics of the GAIT-2 participants are provided in ►**Table 1**. The mean age of the GAIT-2 population ($N=935$) was 39.51 years $\pm$ 21.43 of SD, ranging from the youngest individual with 2.6 years to the oldest one with 101.1 years. There were nearly an equal number of males (49.73%) and females (50.27%) in the study, with 24.06% of subjects considered current smokers.

The covariates that showed a significant effect on cfDNA levels in GAIT-2 were age ($\beta=-0.015$; $p=0.0023$), age^2 ($\beta=0.652$; $p=1.75 \times 10^{-7}$), and sex ($\beta=0.164$; $p=0.0074$). Further analyses were adjusted for the effect of these covariates.

The heritability for cfDNA in GAIT-2 was 0.26 ± 0.06 of standard error ($p=3.7 \times 10^{-6}$) (►**Table 1**). This indicates that a considerable proportion of cfDNA variance is due to the additive effect of genes.

The mean value of cfDNA levels measured in the GAIT-2 individuals was 335.97 ng/mL $\pm$ 66.39 of SD (►**Table 1**). A genetic correlation was found between cfDNA and thrombosis (considering cases of VT, AT, and both VT and AT), although not significant ($\rho_g=0.43$, $p=0.089$). Also, a lack of significant genetic correlation with cfDNA was observed

Table 1 Demographic characteristics of the GAIT-2 subjects

Characteristics	GAIT-2
No. of subjects, n	935
No. of families	35
No. of subjects with thrombosis, n (%)	118 (12.62)
• VT only, n (%)	71 (7.60)
• AT only, n (%)	33 (3.52)
• Both VT and AT, n (%)	14 (1.50)
Mean age, years $\pm$ SD, range	39.51 $\pm$ 21.43, [2.6–101.1]
Males, n (%)	465 (49.73)
Current smokers[a], n (%)	225 (24.06)
Oral contraceptives[b], n (%)	39 (8.30)
BMI, mean $\pm$ SD	24.77 $\pm$ 5.58
cfDNA (ng/mL), mean $\pm$ SD	335.97 $\pm$ 66.39
Heritability of cfDNA $\pm$ SE (p)	0.264 $\pm$ 0.065 ($p=3.7 \times 10^{-6}$)

Abbreviations: AT, arterial thrombosis; BMI, body mass index; cfDNA, cell-free DNA; GAIT-2, Genetic Analysis of Idiopathic Thrombophilia; SD, standard deviation; SE, standard error; VT, venous thrombosis.
Note: Heritability was calculated adjusting for age, age^2, and sex. p-Value $\leq$0.05 was considered statistically significant.
[a]Individuals in the study were considered as current smokers when they smoked, independently of the number of cigarettes.
[b]Oral contraceptives use at inclusion. The percentage of oral contraception is expressed with respect to the total number of females included in the study ($N=470$).

for VT ($\rho_g=0.41$, $p=0.135$) or AT ($\rho_g=0.58$, $p=0.186$) (►**Supplementary Table S1** [available in the online version]).

GWAS for cfDNA in GAIT-2

We conducted a GWAS to identify genetic variants associated with cfDNA levels in plasma from the subjects of the GAIT-2 study. Among the >10 million of imputed variants in our GWAS, we found one significant locus, with rs1687391 as lead SNP (rs1687391-T, $\beta=0.804$, $p=3.55 \times 10^{-10}$). The percentage of the variance in cfDNA levels explained by the effect of one copy of the rs1687391-T allele was estimated in 0.0495%. The SNP rs1687391 is an intergenic variant located 1.25 kb downstream ORM1 (orosomucoid 1) gene, and 2.06 kb upstream ORM2 (orosomucoid 2) gene on chromosome 9q32. A Manhattan plot showing the SNP association signal with cfDNA levels at the ORM locus is shown in ►**Fig. 1**, along with a regional association plot. A box plot of the association of cfDNA levels in the GAIT-2 sample with the genotype of the lead SNP is given in ►**Supplementary Fig. S1** (available in the online version). The top SNP associations with cfDNA levels of the GWAS are shown in ►**Table 2**. Further GWAS data including all SNP–phenotype associations with $p<10^{-4}$ are shown in ►**Supplementary Table S2** (available in the online version). A conditional analysis was conducted on chromosome 9 conditioning on the lead SNP rs1687391 at the ORM locus to identify secondary association signals in this genomic region. The results of this analysis revealed that no other SNPs were significantly associated with cfDNA
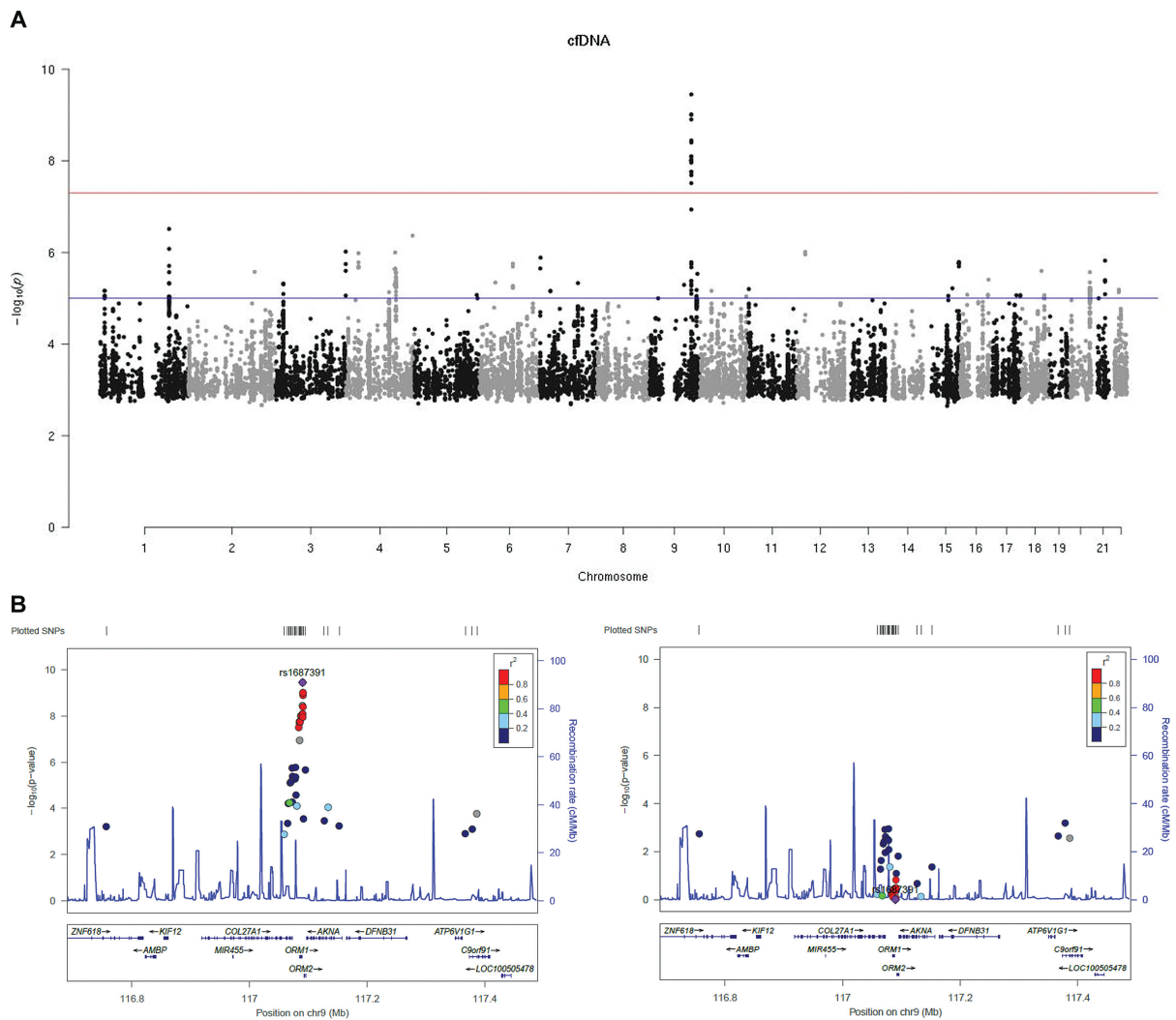
Fig. 1 Manhattan plot of the GWAS for cfDNA levels in GAIT-2. (**A**) Manhattan plot displaying the strength of genetic association (*y*-axis $= -\log_{10}$(*p*-value)) of cfDNA levels versus chromosomal position (*x*-axis = Chromosome). Each dot represents a genotyped or imputed SNP that remained significant after filtering through Matrix eQTL, the second phase of genomic association. The *red horizontal line* represents the threshold for genome-wide significance ($p = 5 \times 10^{-8}$) and the *blue line* indicates the threshold for suggestive associations ($p = 10^{-6}$). The most genome-wide significant SNP is the rs1687391, which is located on chromosome 9 at the intergenic region between *ORM1* and *ORM2* loci. (**B**) Regional plots to depict the independent association signal at the *ORM* locus. Each dot represents one SNP. Dot colors signify the degree of pairwise correlation (r^2) with the top SNP (*purple dot*), as presented in the color key. From left to right, LocusZoom of the association signal in the *ORM* locus for the GWAS; association signal conditioned on the lead SNP (rs1687391). The SNPs (colored dots) significantly associated with cfDNA dropped below significance after conditioning, indicating these SNPs belong to the same association signal. cfDNA, cell-free DNA; eQTL, expression quantitative trait loci; GWAS, genome-wide association study; SNP, single nucleotide polymorphism.

(►**Fig. 1**), indicating that rs1687391, which shows the strongest statistical evidence for association, may capture most of the phenotypic variation at this locus.

cis-eQTL Mapping for *ORM* Genes in GAIT-2
To investigate whether *ORM1* or *ORM2* would be mediating the activity of the significant GWAS locus, *cis*-eQTLs were analyzed in the GAIT-2 population for the *ORM* locus. We identified 83 *cis*-eQTL variants for *ORM1* expression that reached Bonferroni significance ($p < 0.05/28{,}133$) (►**Supplementary Table S3** [available in the online version]) and 146 for *ORM2* (►**Supplementary Table S4** [available in the online version]). Regional plots showing significant *cis*-

eQTLs for *ORM* genes are shown in ►**Fig. 2**. The lead *cis*-eQTL variant for *ORM1* expression in our sample was rs1249761 (rs1249761-C; $\beta = 0.73$; $p = 8.17 \times 10^{-65}$), an intron variant for *COL27A1*, which was also an eQTL for *ORM2* in GAIT-2 (rs1249761-C; $\beta = 0.53$; $p = 1.18 \times 10^{-28}$). The variant was also associated with cfDNA levels in the GWAS (rs1249761-C; $\beta = -0.234$; $p = 7.98 \times 10^{-6}$). The lead eQTL variant for *ORM2* expression in GAIT-2 was rs1631637 (rs1631637-T; $\beta = 0.53$; $p = 4.36 \times 10^{-31}$), an intergenic variant located 3.71 kb downstream *COL27A1* and 8.62 kb upstream *ORM1*. This variant was also significantly associated with the expression of *ORM1* in our sample (rs1631637-T; $\beta = 0.69$; $p = 2.42 \times 10^{-63}$), and showed association with cfDNA levels

Table 2 Top SNP associations with cfDNA levels in GAIT-2 from the GWAS

SNP	Chr.	Position (bp)[a]	Type	Closest gene	strand	Alleles[b]	Minor Allele	MAF[c]	β[d]	Var[e]	p-Value[f]
rs1687391	9	117090006	Intergenic	ORM1	+	C/T	T	0.057	0.804	0.0495	3.55×10^{-10}
rs10982163	9	117091033	Intergenic	ORM2	+	G/T/A	T	0.059	0.764	0.0473	9.78×10^{-10}
rs10982164	9	117091067	Intergenic	ORM2	+	G/A	A	0.058	0.764	0.0472	9.8×10^{-10}
rs112164771	9	117091002	Intergenic	ORM2	+	C/T/A	T	0.059	0.759	0.0469	1.25×10^{-9}
rs1687390	9	117089888	Intergenic	ORM1	+	G/A	A	0.063	0.711	0.0442	3.59×10^{-9}
rs7040440	9	117091074	Intergenic	ORM2	+	C/T/G	T	0.068	0.679	0.0427	3.97×10^{-9}
rs146077223	9	117089921	Intergenic	ORM1	+	G/T	T	0.063	0.697	0.0425	7.98×10^{-9}
rs3762056	9	117090434	Intergenic	ORM2	+	C/T	T	0.063	0.696	0.0425	8.10×10^{-9}
rs10982154	9	117087212	Intron	ORM1	+	T/A	A	0.063	0.693	0.0419	9.67×10^{-9}
rs10982155	9	117087254	Intron	ORM1	+	C/T	T	0.063	0.693	0.0419	9.68×10^{-9}
rs145835687	9	117087745	Intron	ORM1	+	G/A/T	A	0.063	0.692	0.0419	9.84×10^{-9}
rs10982156	9	117088064	Intron	ORM1	+	T/A	A	0.063	0.691	0.0418	1.0×10^{-8}
rs3762055	9	117090575	Intergenic	ORM2	+	T/C	C	0.063	0.686	0.0416	1.09×10^{-8}
rs116994374	9	117084672	Promoter	ORM1	+	G/A	A	0.062	0.693	0.0395	1.73×10^{-8}
rs113354603	9	117086241	Intron	ORM1	+	G/A	A	0.066	0.665	0.0391	1.76×10^{-8}
rs188924106	9	117084722	Promoter	ORM1	+	G/A	A	0.062	0.689	0.0392	2.06×10^{-8}
rs150611042	9	117083803	Promoter	ORM1	+	C/A/T	A	0.063	0.679	0.0379	3.07×10^{-8}
rs140041983	9	117084971	Promoter	ORM1	+	C/G/T	G	0.054	0.700	0.0357	1.16×10^{-7}
rs10921528	1	194259458	Intergenic	CDC73	+	G/C	C	0.387	−0.260	0.0352	3.07×10^{-7}
rs115078510	4	185917640	Intergenic	HELT	+	C/T	T	0.020	−0.995	0.0334	4.30×10^{-7}
rs1325429	1	194262018	Intergenic	CDC73	+	C/T	T	0.405	−0.248	0.0320	8.36×10^{-7}
rs60345835	3	196688877	Intron	PIGZ	−	C/A	A	0.315	0.250	0.0291	9.58×10^{-7}
rs7301073	12	25262868	Intron	CASC1	−	T/C	C	0.018	−0.976	0.0309	9.73×10^{-7}
rs10519385	4	137249906	Intergenic	PCDH18	−	C/T	T	0.077	−0.437	0.0303	1.00×10^{-6}

Abbreviations: CASC1: cancer susceptibility candidate 1; CDC73: cell division cycle 73; Chr, chromosome; HELT: helt bHLH transcription factor; ; ORM1: orosomucoid 1; ORM2: orosomucoid 2; PIGZ: phosphatidylinositol glycan anchor biosynthesis, class Z; PCDH18: protocadherin 18.

[a]All chromosome positions were based on the human GRCh37/hg19 assembly.
[b]Alleles aligned to +strand. Underlined alleles are very rare alleles in some populations, with minor allele frequency (MAF) < 0.01, not present in our sample.
[c]Minor allele frequency of the SNP in our sample.
[d]Effect size of one copy of the minor allele on cfDNA levels.
[e]Percentage of the variance in cfDNA levels that is explained by the effect of one copy of the minor allele.
[f]p-Value of the association with cfDNA levels. Genome-wide significant threshold was set at $p \leq 5 \times 10^{-8}$. Most significant SNP–phenotype associations, with p-value $\leq 10^{-6}$, are shown in the table.
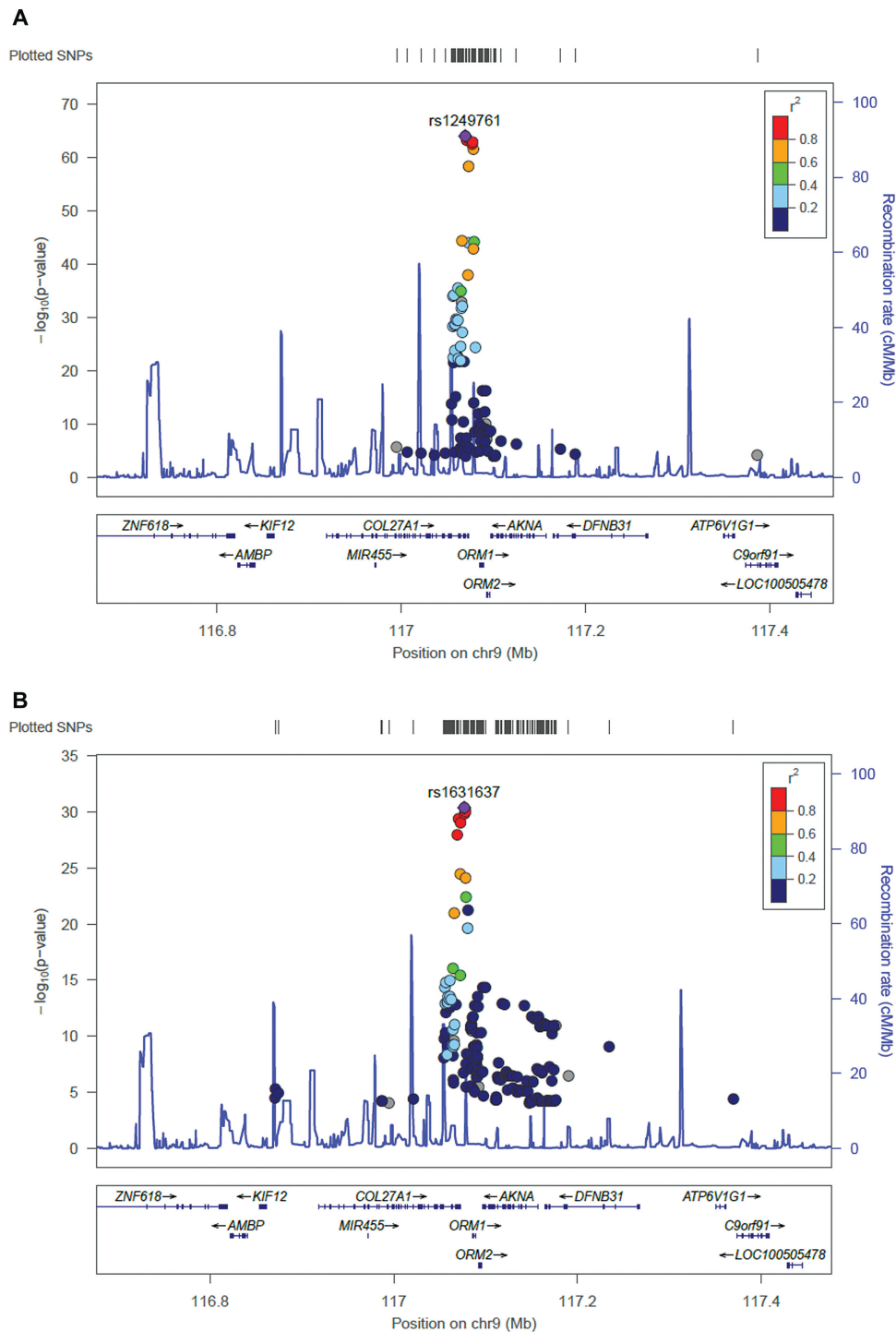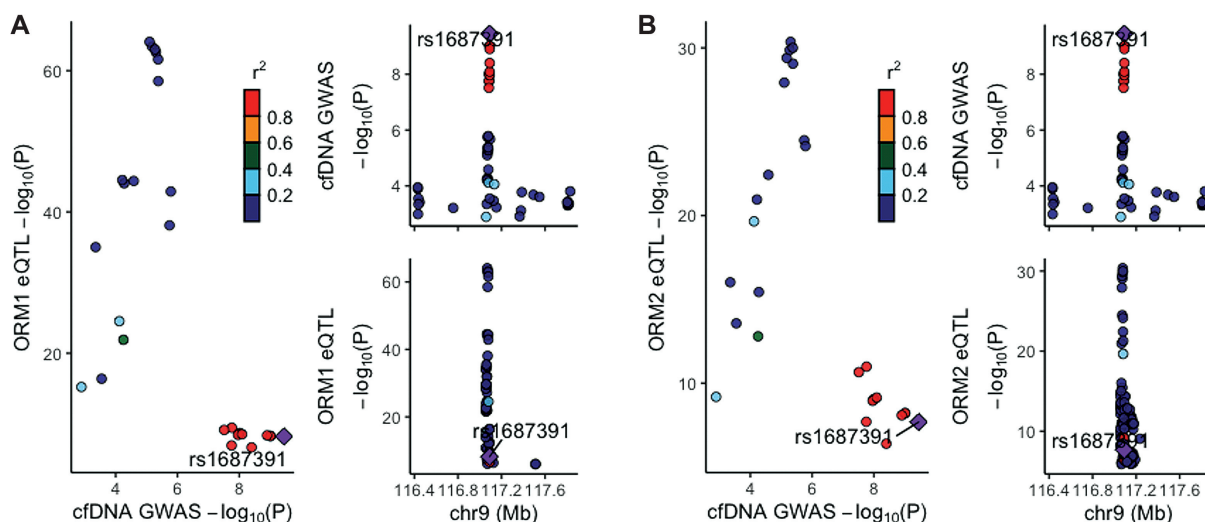
Fig. 2 Regional plots of *cis*-eQTLs for *ORM1* and *ORM2* genes. Regional association plots displaying the genomic context of the *cis*-eQTL signal for *ORM1* (**A**) and *ORM2* (**B**). Each dot represents a genotyped or imputed SNP. Dot colors signify the degree of pairwise correlation (r^2) with the top SNP (*purple dot*), as presented in the color key. eQTL, expression quantitative trait loci; SNP, single nucleotide polymorphism.

in the GWAS (rs1631637-T; $\beta = -0.229$; $p = 4.97 \times 10^{-6}$). In addition, the lead SNP rs1687391 in the GWAS for cfDNA levels was also a significant *cis*-eQTL variant for *ORM1* (rs1687391-T; $\beta = -0.673$; $p = 6.14 \times 10^{-9}$) and *ORM2* expression (rs1687391-T; $\beta = -0.691$; $p = 2.07 \times 10^{-8}$). This indicates that *ORM1* and *ORM2* are the main candidate genes that explain the association of cfDNA with genetic variation. Box plots in ►**Supplementary Fig. S2** (online only) illustrate

the expression levels of *ORM1* and *ORM2* genes as a function of the genotypes for the most significant *cis*-eQTL variants in whole blood for both genes.

The significant *cis*-eQTL variants identified in our study associated with *ORM1* and *ORM2* expression mostly were regulatory variants of the transcript levels of the two genes. In this regard, 70 variants that reside mainly close to or within *COL27A1* (54%), *ORM1* (27%), and *ORM2* (17%)

Fig. 3 LocusCompare plots of *cis*-eQTLs for *ORM1* and *ORM2* genes. Visualization plots displaying colocalization of GWAS and *cis*-eQTLs for *ORM1* (A) and *ORM2* (B) in the GAIT-2 study. Only *cis*-eQTL variants reaching the Bonferroni threshold of significance (0.05/28,133) are represented in LocusZoom plots showing eQTLs. The labeled SNP (*purple dot*) is the lead SNP in the GWAS that is a significant eQTL variant for both *ORM1* and *ORM2* genes. Other SNPs are colored according to their LD r^2 with the lead SNP, as presented in the color key. eQTL, expression quantitative trait loci; GWAS, genome-wide association study; LD, linkage disequilibrium; SNP, single nucleotide polymorphism.

(►**Supplementary Table S5** [available in the online version]) were *cis*-eQTL variants shared by the two genes. Of note, although more than half of the regulatory variants for *ORM1* and *ORM2* lay close to or within *COL27A1*, also being the most significant, none of them were regulatory variants of *COL27A1* in whole blood (according to our data in GAIT-2— ►**Supplementary Table S6** [available in the online version]— and in agreement to GTEx), therefore the gene *COL27A1* is not considered a mediator of our GWAS signal. Noteworthy, these *cis*-eQTL variants in common for the two *ORM* genes were completely overlapping with the variants associated with cfDNA levels. This indicates that *cis*-eQTLs for *ORM* genes capture 100% of the GWAS signal. The overlapping *cis*-eQTL variants are listed in ►**Supplementary Table S7** (available in the online version). Colocalization of the GWAS signal and eQTLs were illustrated through LocusCompare (►**Fig. 3**).

cis-eQTLs for *ORM* Genes in GTEx

As genotype and expression data in multiple tissues are available in GTEx, we conducted a search for *cis*-eQTLs for *ORM* genes in GTEx in whole blood and compared with our results. Lead *cis*-eQTL variants for *ORM1* (rs1249761) and *ORM2* (rs1631637) that we identified in whole blood were also eQTLs in GTEx for these genes in the same tissue. The lead variant rs1687391 was not reported in GTEx; however, the rs10982154 (an intron variant of *ORM1*) was used as a proxy SNP (LD $r^2 = 1$) and showed association only with *ORM1* ($p = 1.8 \times 10^{-5}$) and *ORM2* ($p = 1.0 \times 10^{-8}$) in whole blood in GTEx.

A total of 79 *cis*-eQTL variants for *ORM1* in whole blood were reported in GTEx, of which 46 (58.2%) were significant regulatory SNPs identified in GAIT-2. Twenty-five of them were *cis*-eQTL variants overlapping with the associated locus with cfDNA levels in our GWAS, of which 10 were genome-wide significant variants. For *ORM2*, 103 *cis*-eQTL variants in whole

blood were reported in GTEx, of which 80 (77.7%) were significant regulatory SNPs identified in GAIT-2. Thirty-one of them were *cis*-eQTL variants overlapping with our GWAS signal, of which 14 were genome-wide significant SNPs.

cfDNA and mRNA Levels of Candidate Genes

Finally, associations between cfDNA levels and mRNA levels for the candidate genes from the GWAS were calculated to identify a potential effector transcript and elucidate the role of the locus mediating the GWAS effect. The expression levels of *ORM1* in whole blood showed an inverse significant correlation with cfDNA levels at a FDR < 0.05 ($\beta = -0.098$; $p = 4.38 \times 10^{-4}$; FDR = 0.026). This indicates that *ORM1* expression decreased with the increase of cfDNA in plasma. However, no significant correlation was observed with *ORM2* ($\beta = -0.071$; $p = 0.0245$; FDR = 0.199). This supports *ORM1* as the main candidate gene in the control of the levels of cfDNA in plasma. Since most significant eQTL variants for *ORM1* and *ORM2* genes were SNPs located close to or within *COL27A1* gene, we tested also its correlation with cfDNA. However, it was not statistically significant ($\beta = -0.069$; $p = 0.0275$; FDR = 0.211). These data are shown in ►**Table 3**.

Pathway Enrichment Analysis

To identify functionally related genes from the large list of mRNA transcripts that correlated with cfDNA in our study, and delineate a potential biological pathway for our studied phenotype, we conducted a functional enrichment analysis by using the DAVID annotation tool. Accordingly, we performed the gene set enrichment analysis with all the transcripts filtered by FDR <0.1 in our RNA-Seq. Thus, a total of 983 mRNAs showed significant correlation with cfDNA levels. All of these transcripts were considered for the enrichment analysis and were uploaded into DAVID. A total of 52 transcripts (5.29%) were not recognized by the enrichment tool, as they mainly

Table 3 Associations between mRNA levels of candidate genes and plasma levels of cfDNA in GAIT-2

Gene	Ensembl Gene ID	Chr. position (bp)[a]	β[b]	p-Value[c]	FDR 5%
ORM1	ENSG00000229314	9:117085336–117088755	−0.098	4.38×10^{-4}	0.026
ORM2	ENSG00000228278	9:117092149–117095532	−0.071	0.0245	0.199
COL27A1	ENSG00000196739	9:116917840–117074791	−0.069	0.0275	0.211

Abbreviations: Chr.: chromosome; COL27A1: collagen type XXVII α 1; ORM1: orosomucoid 1; ORM2: orosomucoid 2.
[a]All chromosome positions were based on the human GRCh37/hg19 assembly.
[b]Effect size of one copy of the minor allele on gene expression levels.
[c]p-Value of the association of gene expression with cfDNA levels. Significance was considered at a false discovery rate (FDR) < 0.05.

were long intergenic noncoding RNAs (lincRNAs) or genes coding uncharacterized proteins. Then, 931 transcripts were finally analyzed by the Functional Annotation Clustering Tool according to the following functional annotation categories: Gene Ontology (GO) Biological Process, GO Molecular Function, GO Cellular Component, KEGG pathway, and UniProt Keywords. The results show that our gene set is mainly enriched in pathways related to the innate immunity, type I interferon (IFN) signaling pathway, defense response to virus, and cell division. ►**Supplementary Table S8** (available in the online version) shows the functional annotation clusters that have significant group enrichment scores, showing the genes associated with each enriched biology term.

Finally, a summary of our study design indicating our most relevant findings for cfDNA in GAIT-2 is provided in ►**Supplementary Fig. S3** (available in the online version).

Discussion

The potential of plasma levels of cfDNA as a biomarker of disease and pathogenic mediator is currently being investigated. However, the factors that regulate cfDNA levels in physiological and pathological conditions are poorly understood. Our aim was to identify genes that control the variation of cfDNA in families with idiopathic thrombophilia from the GAIT-2 study (**Supplementary Material** [available in the online version]). For this purpose, we performed an integrated genomic and transcriptomic data analysis. Our results show that cfDNA is a heritable trait, with 26% of cfDNA variance in plasma under genetic control. To the best of our knowledge, this is the first time that the heritability of cfDNA has been reported. It is important to note that cfDNA levels show an important genetic correlation with thrombosis, although it did not reach statistical significance mainly due to limited sample size, since only the 12.62% of the total GAIT-2 population had thrombosis. This suggests the presence of common genetic pathways underlying both phenotypes.

The GWAS shows that a SNP near the ORM1 gene influences the plasma levels of cfDNA, explaining nearly 0.05% of its variance (rs1687391-T). The lead SNP showed no association with thrombosis in the whole GAIT-2 population, maybe due to limited sample size of cases with thrombosis in our study as mentioned above (data not shown). The conditional analysis on the lead SNP shows that it is an independent GWAS signal. This indicates that a causal variant within this signal or in strong LD (linkage disequilibrium) with the

associated locus does may exist. The gene UGT1A1 on chromosome 2 was identified recently by GWAS as a regulator of cfDNA levels in serum.[20] However, we did not find significant or suggestive SNP associations at this locus. Several nongenetic factors that influence the phenotype variance, and thus might contribute to different findings, have been controlled in the analyses of both studies, although others may remain uncontrolled (**Supplementary Material** [available in the online version]). Some of them may be related to technical issues (different fluorochrome and assay conditions) or differences in populations (e.g., case–control vs. family-based, different demographic and clinical characteristics). On the other hand, the higher sample size of the study of Jylhävä et al could contribute with more statistical power to the GWAS study. In addition, both studies might have different pools of cfDNA in plasma or serum (from different origins or cell sources) thus showing different regulatory mechanisms. Furthermore, our study of families with hereditary idiopathic thrombophilia could be genetically enriched in a specific pathway for cfDNA regulation (e.g., immuno-thrombosis) and thus, it could show different locus than those observed in the study with healthy young individuals. In this regard, studies of extended pedigrees are likely to represent a more homogeneous and limited set of causative genes and pathways.

In GWAS studies, the identification of genes mediating the activity of SNPs with functional effect is usually a challenging task. In this regard, the analysis of eQTLs in disease-relevant tissues is useful to detect associations of genetic variation with changes in gene expression that may help to understand the functional role of the SNP–phenotype associations.[31] Accordingly, cis-eQTLs, which have a large effect size on the expression of nearby genes, were mapped in GAIT-2 for the ORM locus and lead to the identification of the variants that likely affect proximal ORM1 and ORM2 expression in whole blood. Among the number of cis-eQTL variants identified in our study for ORM genes, the lead variant rs1687391 in our GWAS was significantly associated with both ORM1 and ORM2 expression. This indicates that the rs1687391 might be a regulatory variant that captures part of the variability of the expression of ORM genes and, thus it could influence cfDNA levels.

We found that most cis-eQTL variants for ORM1 and ORM2 in blood are regulatory SNPs shared by the two genes that reside mainly close to or within COL27A1 and ORM genes. Among these cis-eQTL variants, the most significant SNPs

modulating gene expression of the *ORM* locus lie close to or within *COL27A1* (located 10.5 kb upstream *ORM1*); however, none of them were eQTL variants for *COL27A1*. Of note, all the *cis*-eQTL variants shared by *ORM* genes were shared with the GWAS signal, thus providing more evidence of the functional role of the *ORM* locus in the control of cfDNA levels. In addition, most of the SNPs of the GWAS signal lie in regulatory regions (noncoding and promoter) of *ORM1* expression, which suggests *ORM1* as the main candidate gene in the control of cfDNA levels. These results are reinforced by the integration of expression data from the main candidate genes, which reveals that among the genes within the associated locus, only *ORM1* expression correlates negatively with plasma levels of cfDNA. Accordingly, we hypothesize that *ORM1* could have a protective role against harmful effects of elevated cfDNA. While functional assays will be necessary to disclose the causal SNP, our study has provided strong evidence for a novel gene that regulates cfDNA levels. Furthermore, our RNA-Seq data indicate that the gene set associated with cfDNA is mainly related to pathways of innate immunity, type I IFN signaling, defense response to virus, and cell division. Type I IFNs have diverse effects on innate and adaptive immune cells during infection, are important for host defense against viruses, and enhance NETosis. IFN-α primes mature neutrophils to form NETs to bind to pathogens.[32] This points to immunothrombosis as a mechanism by which *ORM1* could contribute to the susceptibility of thrombosis.

Various genetic variants close to or within *ORM1*, which have been associated with cfDNA levels in our GWAS, have been found associated with traits related to coagulation or cardiovascular disease in the NHGRI-EBI GWAS Catalog. For example, the lead variant rs1687391 has been associated with the serine protease inhibitor Kazal-type 2 (SPINK2),[33] a protein associated with stroke,[34] mean platelet volume and platelet count,[35] and leukocyte count.[36] The intergenic rs3762056 and the intron variant rs10982156 have been associated with the plasma levels of the coagulation factor X (FX).[33,37] The rs1687390 has been associated with variance in warfarin dose,[38] and the rs150611042 in the promoter of *ORM1* has been reported to influence the interindividual capacity to generate thrombin.[39] Also, SNPs at *ORM1* gene (rs150611042) were recently found associated with protein S levels through a meta-analysis using GWAS data from the CHARGE Consortium.[40] These data, along with our results, suggest that *ORM1* is a susceptibility locus in thrombosis. Also, it might be linked to cfDNA-mediated inflammation as an underlying pathological pathway. The *ORM* locus conforms a complex and very polymorphic genomic region with two highly homologous genes, *ORM1* and *ORM2*, that are differentially regulated.[41] *ORM1* is located 3.55 kb upstream *ORM2*, and code for α-1-acid glycoprotein 1 (AGP1) and AGP2, respectively. Both proteins are present in plasma, and AGP1 is the most abundant, where it functions as transport proteins. AGP1 is an acute phase protein[42] mainly expressed in liver and secreted in the blood, although it can be produced also in extrahepatic tissues.[43] The exact biological function of AGP (also known as orosomucoid) is un-

known, although it has been related to multiple modulating activities.[43] AGP is considered a natural anti-inflammatory and immunoregulatory protein, with proangiogenic and antiapoptotic properties that could be important for cardiovascular disease.[44,45] Accordingly, some evidence supports a role of AGP in coagulation. As mentioned above, AGP has been associated with FX,[33,37] thrombin generation potential,[39] and protein S levels.[40] Also, AGP interacts with plasminogen activator inhibitor 1, enhancing fibrinolysis inhibition.[46] In an experimental model of septic peritonitis in rats, AGP administration suggests a protective effect of AGP against a procoagulant phenotype by normalizing some clot parameters.[47] This is in line with the protective effect of *ORM1* that we suggest. Consistent with our results, Liu et al reported that AGP reduced in vitro the formation of DNA-based NETs for sperm entanglement. The study showed that AGP reduced the phagocytic activity of neutrophils in a dose-dependent manner, reducing NET formation and superoxide generation of activated neutrophils.[48] Also, elevated AGP in body fluids and altered patterns of glycosylation have been associated with several diseases.[49] In particular, elevated gene expression and plasma levels of AGP have been associated with coronary disease and stroke.[50,51]

Some limitations in our study should be considered. First, the measurement of ORM1 protein in plasma is not available in the GAIT-2 sample. Plasma levels of ORM1 protein and its correlation with the lead SNP, mRNA *ORM1* levels or cfDNA, could have added an extra piece of support to our study. Moreover, the low number of cases with thrombosis in relation to the whole participants in the family-based study may have limited the association of cfDNA or *ORM1* variants with thrombosis risk. Accordingly, a replication cohort with cases and controls is essential. Also, a modest coefficient of variance is attributable to the lead SNP that could be in part explained by the complex nature of the plasma cfDNA. However, while not clinically meaningful for prediction of the risk of the disease, our finding of a genome-wide significant signal in the *ORM1* gene clearly points to a novel regulatory mechanism of cfDNA. On the other hand, although direct fluorescent Sytox Green staining is not a validated method for the quantification of cfDNA in plasma, it is extensively used and might serve as reliable alternative to real-time quantitative polymerase chain reaction for routine quantification of cfDNA and multicenter standardization.[52] Also, it has shown higher sensitivity for DNA in plasma than other intercalating fluorochromes.[17] Furthermore, diverse forms of DNA may circulate in plasma derived from NETs or mitochondria, but may also be actively released from living cells either encapsulated in vesicles or in the form of complexes, triggering different biological mechanisms associated with health and disease. However, fluorescence quantification of cfDNA is an unspecific approach that measures the total amount of extracellular DNA, and thus it may not be considered a single biomarker of disease.

In summary, to the best of our knowledge, we have presented the first integrated study that links genetic variation with gene expression and the levels of cfDNA in plasma in a study of extended families affected with idiopathic thrombophilia. Our results should shed light on the genetic

control of cfDNA levels, which is considered a risk factor for VT and suggests that *ORM1* is a novel locus regulating cfDNA levels in plasma. This locus could contribute to the susceptibility for VT through a mechanism of immunothrombosis. Further studies will be essential to evaluate our results.

What is known about this topic?

- Plasma cfDNA is a surrogate marker of NETs, and its elevated levels have been associated with increased thrombosis risk.
- The knowledge about the factors implicated in the regulation of plasma levels of cfDNA is scarce.

What does this paper add?

- This is the first study linking genetic variation with gene expression and cfDNA in thrombosis to identify genes regulating cfDNA levels in plasma.
- *ORM1* is identified as a new gene involved in the control of cfDNA levels that could have a protective role against elevated cfDNA. A suggestive genetic association between cfDNA and thrombosis is shown.
- *ORM1* arises as a susceptibility locus for thrombosis, which might be linked to mechanisms of immunothrombosis.

References

1 Aucamp J, Bronkhorst AJ, Badenhorst CPS, Pretorius PJ. The diverse origins of circulating cell-free DNA in the human body: a critical re-evaluation of the literature. Biol Rev Camb Philos Soc 2018;93(03):1649–1683

2 Pös O, Biró O, Szemes T, Nagy B. Circulating cell-free nucleic acids: characteristics and applications. Eur J Hum Genet 2018;26(07): 937–945

3 Brinkmann V, Reichard U, Goosmann C, et al. Neutrophil extracellular traps kill bacteria. Science 2004;303(5663):1532–1535

4 Denning N-L, Aziz M, Gurien SD, Wang P. DAMPs and NETs in sepsis. Front Immunol 2019;10:2536

5 Demers M, Wagner DD. NETosis: a new factor in tumor progression and cancer-associated thrombosis. Semin Thromb Hemost 2014;40(03):277–283

6 Lee KH, Kronbichler A, Park DD-Y, et al. Neutrophil extracellular traps (NETs) in autoimmune diseases: a comprehensive review. Autoimmun Rev 2017;16(11):1160–1173

7 Döring Y, Soehnlein O, Weber C. Neutrophil extracellular traps in atherosclerosis and atherothrombosis. Circ Res 2017;120(04): 736–743

8 Zawrotniak M, Rapala-Kozik M. Neutrophil extracellular traps (NETs) - formation and implications. Acta Biochim Pol 2013;60 (03):277–284

9 Saffarzadeh M, Juenemann C, Queisser MA, et al. Neutrophil extracellular traps directly induce epithelial and endothelial cell death: a predominant role of histones. PLoS One 2012;7 (02):e32366

10 Kimball AS, Obi AT, Diaz JA, Henke PK. The emerging role of NETs in venous thrombosis and immunothrombosis. Front Immunol 2016;7:236

11 Fuchs TA, Brill A, Duerschmied D, et al. Extracellular DNA traps promote thrombosis. Proc Natl Acad Sci U S A 2010;107(36): 15880–15885

12 von Brühl M-L, Stark K, Steinhart A, et al. Monocytes, neutrophils, and platelets cooperate to initiate and propagate venous thrombosis in mice in vivo. J Exp Med 2012;209(04):819–835

13 Schulz C, Engelmann B, Massberg S. Crossroads of coagulation and innate immunity: the case of deep vein thrombosis. J Thromb Haemost 2013;11(Suppl 1):233–241

14 van Montfoort ML, Stephan F, Lauw MN, et al. Circulating nucleosomes and neutrophil activation as risk factors for deep vein thrombosis. Arterioscler Thromb Vasc Biol 2013;33(01):147–151

15 Diaz JA, Fuchs TA, Jackson TO, et al; for the Michigan Research Venous Group* Plasma DNA is elevated in patients with deep vein thrombosis. J Vasc Surg Venous Lymphat Disord 2013;1(04):341. e1–348.e1

16 Brill A, Fuchs TA, Savchenko AS, et al. Neutrophil extracellular traps promote deep vein thrombosis in mice. J Thromb Haemost 2012;10(01):136–144

17 Jiménez-Alcázar M, Limacher A, Panda R, et al. Circulating extracellular DNA is an independent predictor of mortality in elderly patients with venous thromboembolism. PLoS One 2018;13(02): e0191150

18 Middleton EA, He XY, Denorme F, et al. Neutrophil extracellular traps contribute to immunothrombosis in COVID-19 acute respiratory distress syndrome. Blood 2020;136(10):1169–1179

19 Zuo Y, Yalavarthi S, Shi H, et al. Neutrophil extracellular traps in COVID-19. JCI Insight 2020;5(11):e138999

20 Jylhävä J, Lyytikäinen L-P, Kähönen M, et al. A genome-wide association study identifies UGT1A1 as a regulator of serum cell-free DNA in young adults: the cardiovascular risk in young Finns study. PLoS One 2012;7(04):e35426

21 Martin-Fernandez L, Ziyatdinov A, Carrasco M, et al. Genetic determinants of thrombin generation and their relation to venous thrombosis: results from the GAIT-2 Project. PLoS One 2016;11 (01):e0146922

22 Miller SA, Dykes DD, Polesky HF. A simple salting out procedure for extracting DNA from human nucleated cells. Nucleic Acids Res 1988;16(03):1215

23 Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics 2009;25(14):1754–1760

24 Ziyatdinov A, Brunel H, Martinez-Perez A, Buil A, Perera A, Soria JM. solarius: an R interface to SOLAR for variance component analysis in pedigrees. Bioinformatics 2016;32(12):1901–1902

25 Almasy L, Blangero J. Multipoint quantitative-trait linkage analysis in general pedigrees. Am J Hum Genet 1998;62(05): 1198–1211

26 Self SG, Liang K-Y. Asymptotic properties of maximum likelihood estimators and likelihood ratio tests under nonstandard conditions. J Am Stat Assoc 1987;82(398):605–610

27 Shabalin AA. Matrix eQTL: ultra fast eQTL analysis via large matrix operations. Bioinformatics 2012;28(10):1353–1358

28 Pruim RJ, Welch RP, Sanna S, et al. LocusZoom: regional visualization of genome-wide association scan results. Bioinformatics 2010;26(18):2336–2337

29 Ongen H, Buil A, Brown AA, Dermitzakis ET, Delaneau O. Fast and efficient QTL mapper for thousands of molecular phenotypes. Bioinformatics 2016;32(10):1479–1485

30 Huang DW, Sherman BT, Tan Q, et al. The DAVID Gene Functional Classification Tool: a novel biological module-centric algorithm to functionally analyze large gene lists. Genome Biol 2007;8(09):R183

31 Joehanes R, Zhang X, Huan T, et al. Integrated genome-wide analysis of expression quantitative trait loci aids interpretation of genomic association studies. Genome Biol 2017;18(01):16

32 Stegelmeier AA, Darzianiazizi M, Hanada K, et al. Type I interferon-mediated regulation of antiviral capabilities of neutrophils. Int J Mol Sci 2021;22(09):4726

33 Sun BB, Maranville JC, Peters JE, et al. Genomic atlas of the human plasma proteome. Nature 2018;558(7708):73–79

34 Carty CL, Keene KL, Cheng Y-C, et al; COMPASS and METASTROKE Consortia. Meta-analysis of genome-wide association studies identifies genetic risk factors for stroke in African Americans. Stroke 2015;46(08):2063–2068

35 Astle WJ, Elding H, Jiang T, et al. The allelic landscape of human blood cell trait variation and links to common complex disease. Cell 2016;167(05):1415–1429

36 Kichaev G, Bhatia G, Loh P-R, et al. Leveraging polygenic functional enrichment to improve GWAS power. Am J Hum Genet 2019;104(01):65–75

37 Williams FMK, Carter AM, Hysi PG, et al; EuroCLOT Investigators Wellcome Trust Case Control Consortium 2 MOnica Risk, Genetics, Archiving and Monograph MetaStroke International Stroke Genetics Consortium. Ischemic stroke is associated with the ABO locus: the EuroCLOT study. Ann Neurol 2013;73(01):16–31

38 Wadelius M, Chen LY, Eriksson N, et al. Association of warfarin dose with genes involved in its action and metabolism. Hum Genet 2007;121(01):23–34

39 Rocanin-Arjo A, Cohen W, Carcaillon L, et al; CardioGenics Consortium. A meta-analysis of genome-wide association studies identifies ORM1 as a novel gene controlling thrombin generation potential. Blood 2014;123(05):777–785

40 Sabater-Lleal M, Ji Y, Temprano G, et al. Genome-wide association analyses of natural anticoagulants using TOPMED reference panel reveal novel loci associated with Antithrombin, Protein C and Protein S. (Abstract #2131). Paper presented at: The 70th Annual Meeting of The American Society of Human Genetics. October 27–30, 2020. Virtual Meeting

41 Sai K, Kurose K, Koizumi T, et al. Distal promoter regions are responsible for differential regulation of human orosomucoid-1 and -2 gene expression and acute phase responses. Biol Pharm Bull 2014;37(01):164–168

42 Fournier T, Medjoubi-N N, Porquet D. Alpha-1-acid glycoprotein. Biochim Biophys Acta 2000;1482(1–2):157–171

43 Luo Z, Lei H, Sun Y, Liu X, Su D-F. Orosomucoid, an acute response protein with multiple modulating activities. J Physiol Biochem 2015;71(02):329–340

44 Irmak S, Oliveira-Ferrer L, Singer BB, Ergün S, Tilki D. Pro-angiogenic properties of orosomucoid (ORM). Exp Cell Res 2009;315(18):3201–3209

45 de Vries B, Walter SJ, Wolfs TGAM, et al. Exogenous alpha-1-acid glycoprotein protects against renal ischemia-reperfusion injury by inhibition of inflammation and apoptosis. Transplantation 2004;78(08):1116–1124

46 Boncela J, Papiewska I, Fijalkowska I, Walkowiak B, Cierniewski CS. Acute phase protein alpha 1-acid glycoprotein interacts with plasminogen activator inhibitor type 1 and stabilizes its inhibitory activity. J Biol Chem 2001;276(38):35305–35311

47 Osikov MV, Makarov EV, Krivokhizhina LV. Effects of α1-acid glycoprotein on hemostasis in experimental septic peritonitis. Bull Exp Biol Med 2007;144(02):178–180

48 Liu J, Marey MA, Kowsar R, et al. An acute-phase protein as a regulator of sperm survival in the bovine oviduct: alpha 1-acid-glycoprotein impairs neutrophil phagocytosis of sperm in vitro. J Reprod Dev 2014;60(05):342–348

49 Zhang D, Huang J, Luo D, Feng X, Liu Y, Liu Y. Glycosylation change of alpha-1-acid glycoprotein as a serum biomarker for hepatocellular carcinoma and cirrhosis. Biomarkers Med 2017;11(05): 423–430

50 Berntsson J, Östling G, Persson M, Smith JG, Hedblad B, Engström G. Orosomucoid, carotid plaque, and incidence of stroke. Stroke 2016;47(07):1858–1863

51 Barr TL, Conley Y, Ding J, et al. Genomic biomarkers and cellular pathways of ischemic stroke by RNA gene expression profiling. Neurology 2010;75(11):1009–1014

52 Szpechcinski A, Struniawska R, Zaleska J, et al. Evaluation of fluorescence-based methods for total vs. amplifiable DNA quantification in plasma of lung cancer patients. J Physiol Pharmacol 2008;59(Suppl 6):675–681