

Multi-ancestry Genome-wide Association Study of Varicose Veins Reveals Polygenic Architecture, Genetic Overlap with Arterial and Venous Disease, and Novel Therapeutic Opportunities

Michael G. Levin, MD^{1,2}; Jennifer E. Huffman, PhD³; Anurag Verma, PhD^{2,4}; Kyle A. Sullivan, PhD⁵; Alexis A. Rodriguez, PhD⁶; David Kainer, PhD⁵; Michael R. Garvin, PhD⁵; Matthew Lane, MS⁷; Hyejung Won, PhD⁸; Binglan Li, PhD⁹; Yuan Luo, PhD¹⁰; Gail P. Jarvik, MD, PhD¹¹; Hakon Hakonarson, MD, PhD¹²; Elizabeth A. Jasper, PhD¹³; Alexander G. Bick, MD, PhD¹⁴; Marylyn D. Ritchie, PhD^{4,15}; Daniel A. Jacobson, PhD⁵; Ravi K. Madduri, MS⁶; Scott M. Damrauer, MD^{2,16}

on behalf of the VA Million Veteran Program

¹Division of Cardiovascular Medicine, Department of Medicine, University of Pennsylvania Perelman School of Medicine, Philadelphia, PA

²Corporal Michael J. Crescenz VA Medical Center, Philadelphia, PA

³Massachusetts Veterans Epidemiology Research and Information Center (MAVERIC), VA Boston Healthcare System, Boston, MA

⁴Department of Genetics, University of Pennsylvania Perelman School of Medicine, Philadelphia, PA

⁵Computational and Predictive Biology, Oak Ridge National Laboratory, Oak Ridge, TN

⁶Data Science and Learning Division, Argonne National Laboratory, Lemont, IL

⁷Bredesen Center for Interdisciplinary Research and Graduate Education, University of Tennessee, Knoxville, TN

⁸Department of Genetics, University of North Carolina, Chapel Hill, NC

⁹Department of Biomedical Data Science, Stanford University, Stanford, CA

¹⁰Department of Preventive Medicine, Feinberg School of Medicine, Northwestern University, Chicago, IL

¹¹Departments of Medicine (Division of Medical Genetics) and Genome Sciences, University of Washington Medical Center, Seattle, WA

¹²Center for Applied Genomics, The Children's Hospital of Philadelphia, Philadelphia, Pennsylvania

¹³Division of Quantitative Sciences, Department of Obstetrics and Gynecology, Vanderbilt University Medical Center, Nashville, TN

¹⁴Division of Genetic Medicine, Department of Medicine, Vanderbilt University School of Medicine, Nashville, TN

¹⁵Institute for Biomedical Informatics, University of Pennsylvania Perelman School of Medicine, Philadelphia, PA

¹⁶Department of Surgery, University of Pennsylvania Perelman School of Medicine, Philadelphia, PA

Key Words: Varicose veins, genetics, vascular medicine, cardiovascular medicine

Corresponding Author:

Scott M. Damrauer
3900 Woodland Avenue
Philadelphia PA 19104
Phone: 215-823-5880
damrauer@upenn.edu

ABSTRACT

Background: Varicose veins represent a common cause of cardiovascular morbidity, with limited available medical therapies. Although varicose veins are heritable and epidemiologic studies have identified several candidate varicose veins risk factors, the molecular and genetic basis remains uncertain. Here, we analyzed the contribution of common genetic variants to varicose veins using data from the VA Million Veteran Program and other large multi-ancestry biobanks. Among 49,765 individuals with varicose veins and 1,334,301 disease-free controls, we identified 139 risk loci. We identified genetic overlap between varicose veins, other vascular diseases, and dozens of anthropometric factors. Using Mendelian randomization, we prioritized novel therapeutic targets via integration of proteomic and transcriptomic data. Finally, topological enrichment analyses confirmed the biologic roles of endothelial shear flow disruption, inflammation, vascular remodeling, and angiogenesis. These findings may facilitate future efforts to develop non-surgical therapies for varicose veins.

INTRODUCTION

Varicose veins are a manifestation of chronic venous disease, characterized by venous dilation resulting from venous valvular incompetence, thrombosis, or non-thrombotic venous occlusion. While population prevalence varies, increasing with age and BMI, some estimates suggest varicose veins affect nearly 40% of the adult population.¹⁻⁸ Varicose veins may initially present asymptotically as superficial dilation and tortuosity, but may progress to include symptoms of pain, edema, skin changes, and ulceration.⁹ Serious downstream cardiovascular consequences have been ascribed to varicose veins, including risks of venous thromboembolism and peripheral artery disease.^{10,11}

Several molecular mechanisms have been implicated in varicose veins.¹²⁻¹⁵ The generally accepted molecular pathology of varicose veins involves five major processes: (1) disrupted vascular endothelium homeostasis due to altered shear stress from increases in venous pressure and/or valve failure resulting in (2) immune cell activation, adhesion, and infiltration of the blood vessel wall causing (3) local inflammatory processes and cytokine production followed by (4) remodeling of the extracellular matrix and (5) angiogenesis.

Observational studies have identified several risk factors for varicose veins, including age, sex, measures of obesity (eg. body mass index, waist-to-hip ratio), height, prior venous thromboembolism, pregnancy, and family history of varicose veins, among others.^{16,17} Recent genome wide association studies (GWAS) have begun to elucidate common genetic risk factors for varicose veins, identifying more than 30 risk loci.^{16,18,19} However, these studies have been limited to individuals of European ancestry, and a substantial proportion of varicose veins heritability remains unexplained.

Here, we investigated the genetic basis of varicose veins by conducting the largest multi-ancestry GWAS of varicose veins to date. Using downstream bioinformatic techniques we aimed to (1) prioritize causal risk variants, (2) identify traits with shared genetic mechanisms, (3) define putative causal risk factors, (4) identify circulating proteins and drug targets that may represent therapeutic targets, and (5) provide mechanistic understanding of varicose veins biology using a multi-omic integration method.

RESULTS

Multi-ancestry GWAS

Overall GWAS and study design are presented in Error! Reference source not found.. The VA Million Veteran Program (MVP) is a national research program established in 2011, designed to understand how genes,

lifestyle, and military exposures affect health and illness.²⁰ Using electronic health record diagnosis codes (**SUPPLEMENTAL METHODS**), we identified 18,977 MVP participants with varicose veins and 561,203 disease-free individuals. Leveraging the diverse genetic composition of MVP, we performed ancestry-specific GWAS. Next, we assembled genetic association data for 49,765 individuals with varicose veins and 1,334,301 disease-free controls, across four major genetic ancestral populations (African, East-Asian, European, and Hispanic) by supplementing our MVP analysis with summary data from other large-scale biobanks (eMERGE, UK Biobank, FinnGen, and BioBank Japan). We combined genetic association data first within, and then across ancestries, using fixed-effects meta-analysis.

Following multi-ancestry meta-analysis, we identified 8,336 DNA sequence variants at 139 loci ($r^2 < 0.1$, distance = 250kb) associated with varicose veins at genome-wide significance ($p < 5 \times 10^{-8}$) (Error! Reference source not found.; **SUPPLEMENTAL FIGURE 1; SUPPLEMENTAL TABLES 1-2**). We replicated 30/36 previously-reported risk loci^{16,18,19} (27/36 loci at genome-wide significance, and 3/9 remaining loci at nominal significance [$p < 0.05$]; **SUPPLEMENTAL TABLE 3**). Overall, novel associations included variants with lower frequency and of greater effect size than previously reported associations, including 11 rare variants (minor allele frequency < 0.01) (**SUPPLEMENTAL FIGURE 2**). Using linkage disequilibrium score regression²¹, we found the majority of genomic inflation was due to polygenicity rather than population stratification (LD score intercept = 1.08 [0.01]; Ratio: 0.14 [0.02]). We estimate the heritability of varicose veins at 15.7% (95% Confidence Interval [CI] 13.6% to 17.8%), based on an estimated population prevalence of 23%.²²

Fine-mapping of risk loci

To refine the set of genetic variants associated with varicose veins, we leveraged the multi-ancestry nature of our meta-analysis to perform multi-ancestry fine-mapping using MR-MEGA.²³ This statistical approach performs meta-regression accounting for the heterogeneity in allelic effects that is correlated with ancestry to construct credible sets. The median size of the 99% credible sets was 1 variant, while the mean size was 36. We identified 70 varicose veins risk loci where a single variant was included in the 99% credible set (**SUPPLEMENTAL TABLE 4**). At 63 of these 70 loci, multi-ancestry fine-mapping prioritized a different variant than the fixed-effects meta-analysis. Novel risk variants included rs6025, which represents the thrombophilic mutation F5 p.Arg534Gln (factor V Leiden). This association reinforces a pathophysiologic role for venous thromboembolism, which is a clinically-recognized secondary etiology of varicose veins within the CEAP (Clinical-Etiology-Anatomy-Pathophysiology) classification of chronic venous disorders.⁹ Similarly, rs62033413

at the *FTO* locus (associated with increased weight, body mass index, and hip circumference, among others), highlights an association between obesity and risk of varicose veins.

Tissue and Gene-set Enrichment

We evaluated whether specific tissues were enriched for varicose veins GWAS loci using MAGMA, which integrates GWAS summary statistics with gene expression (RNAseq) data from the Genotype Tissue Expression Project (GTEx v8). We identified significant enrichment in vascular tissues including the aorta, tibial and coronary arteries; reproductive tissues including uterus, ovary, cervix (endo- and ectocervix), and breast; as well as tibial nerve, subcutaneous adipose, and sigmoid colon (**SUPPLEMENTAL FIGURE 3A**). MAGMA gene-set enrichment prioritized significant (Bonferroni adjusted $p < 0.05$) roles for vascular smooth muscle cell differentiation, skeletal system morphogenesis, and Wnt signaling, among others (**SUPPLEMENTAL FIGURE 3B**).

As an alternative to the polygenic mapping approach utilized by MAGMA, we also tested whether genes located nearest to genome-wide significant loci were associated with specific biological processes. Here, we found significant (FDR < 0.05) enrichment for pathways involving vascular smooth muscle cell differentiation (based on associations near *PDGFB*, *NFATC2*, *NFATC1*, and *NOTCH1*), as well as pathways related to vascular, cardiac, and muscle growth/development/differentiation (**SUPPLEMENTAL TABLE 5**). Of 24 genes previously implicated in Mendelian diseases associated with varicose veins in the Human Phenotype Ontology database (**SUPPLEMENTAL TABLE 6**), only 1 (*PIEZO1*) was the closest gene to a genome-wide significant GWAS locus (**SUPPLEMENTAL TABLE 2**).

Shared genetic architecture of vascular traits

Given the polygenic architecture of varicose veins and other vascular traits, as well as observational associations linking varicose veins with other vascular diseases, we estimated genetic correlations between varicose veins and other common vascular diseases. Varicose veins were significantly (FDR $q < 0.05$) genetically correlated with both venous (venous thromboembolism $r_g = 0.27$, $p = 4.8 \times 10^{-9}$) and arterial traits (abdominal aortic aneurysm $r_g = 0.1$, $p = 2.1 \times 10^{-4}$; peripheral artery disease $r_g = 0.21$, $p = 8.6 \times 10^{-8}$), although the degree of correlation was modest (**SUPPLEMENTAL FIGURE 4**).

Phenome-wide scan and colocalization reveal pleiotropic associations

To characterize the broader spectrum of traits sharing risk variants with varicose veins, we performed a phenome-wide association study (pheWAS) followed by colocalization. We queried associations among GWAS

of diagnoses/operations/medications/surveys/measurements (n = 2514) among UK Biobank participants, GWAS of circulating proteins (n = 4489), and GWAS of circulating metabolites (n = 974) from the MRC IEU Open GWAS Project. Among categories of UK Biobank traits sharing genome-wide significant associations ($p < 5 \times 10^{-8}$) with sentinel varicose veins SNPs were 81 anthropometric traits, 20 diagnoses, 15 operations, 5 medications/treatments, and 71 other survey/self-reported traits (**FIGURE 2**). We further identified 77 circulating proteins, and 216 circulating metabolites that shared associations with sentinel varicose vein SNPs (**FIGURE 2**).

Risk variants may be shared across traits based on linkage disequilibrium or shared causal genetic mechanisms.²⁴ To prioritize traits that share causal genetic mechanisms with varicose veins, we performed Bayesian colocalization.²⁴ We performed colocalization between each varicose veins risk locus (+/- 500kb) and individual 1) UKB traits, 2) proteins, and 3) metabolites sharing genome-wide significant associations at that locus. We identified evidence of colocalization for 826 loci-trait pairs, across 70 distinct loci and 408 distinct traits. Anthropometric traits represented 46% (n = 72 unique traits at 37 loci) of UK Biobank traits sharing causal risk variants with varicose veins (n = 158 unique traits at 68 loci) (**FIGURE 3**). We also detected evidence of shared risk variants with circulating metabolites and proteins. These associations were concentrated at small number of pleiotropic loci (55 proteins at 4 loci, 195 metabolites at 6 loci) (**FIGURE 3**). At the *ABO* locus surrounding rs507666 (type A1 histo-blood group antigen),^{25,26} genes encoding 41 circulating proteins demonstrated evidence of colocalization with varicose veins, including several vascular proteins like tissue factor, E-selectin, ICAM1/2/4/5, PECAM1, P-selectin, and VEGFR2/3 among others.

Next, multi-trait colocalization was performed, similarly testing for shared genetic mechanisms between varicose veins and clusters of 1) UKB traits, 2) proteins, and 3) metabolites (**SUPPLEMENTAL FIGURE 5**). We identified evidence of shared risk variants between varicose veins and UK Biobank traits at 34 loci (**SUPPLEMENTAL FIGURE 5**). We also found loci associated with varicose veins and clusters of circulating metabolites (3 loci) and circulating proteins (2 loci) (**SUPPLEMENTAL FIGURE 5**).

Shared polygenetic architecture of vascular and anthropometric traits

While the shared genetic associations between varicose veins risk loci and anthropometric traits that were identified in pheWAS and colocalization analyses represent evidence of shared genetic architecture at the locus-level, varicose veins and anthropometric traits are largely polygenic traits.²¹ Therefore, we estimated genetic correlations between varicose veins, other vascular traits, and anthropometric traits to better understand shared polygenic patterns of genetic risk. A few notable findings emerged. Measures of height (eg.

sitting height, standing height) were positively correlated with varicose veins and other vascular traits (eg. VTE and AAA), but negatively correlated with CAD (**FIGURE 4A**), consistent with prior reports.^{16,27} Measures of body composition (eg. body mass index, weight, fat mass, fat-free mass, body fat percentage) were largely positively correlated across all vascular traits. While systolic blood pressure was positively correlated with atherosclerotic traits like PAD, CAD, and stroke, we observed an inverse correlation with risk of varicose veins. Finally, markers of bone mineral density were uniquely correlated with PAD, but not other vascular traits.

Anthropometric traits as risk factors for varicose veins

To evaluate for putative casual effects of anthropometric traits on risk of varicose veins and other vascular disease, we performed inverse variance-weighted Mendelian randomization (**FIGURE 4B**). Measures of height (sitting height, standing height) and body composition (eg. body mass index, waist circumference, fat mass, fat-free mass, body water) were strongly associated with increased risk of varicose veins. Results were similar using the weighted-median MR method, which makes different assumptions about the presence of pleiotropy (**SUPPLEMENTAL TABLE 10**).

Proteome-wide Mendelian randomization reveals circulating effectors of venous disease

We hypothesized that circulating proteins, which exist in close contact with the venous vasculature, may represent causal effectors of venous disease and thus novel treatment targets. To characterize the role of circulating proteins on risk of varicose veins, we performed proteome-wide Mendelian randomization. We selected previously established high-confidence genetic proxies of circulating proteins as instrumental variables within the Mendelian randomization framework.²⁸ Of 714 circulating proteins with available *cis*-pQTL genetic proxies, 35 were significantly (FDR $q < 0.05$) associated with risk of varicose veins (**FIGURE 5; SUPPLEMENTAL TABLE 11**). These significant associations were enriched for proteins with molecular functions including extracellular matrix structure (FDR $q = 1.2 \times 10^{-6}$) and collagen binding (FDR $q = 9.3 \times 10^{-3}$); cellular components of the extracellular region (FDR $q = 2 \times 10^{-9}$); and biological processes including cell adhesion (FDR $q = 9.1 \times 10^{-4}$) and anatomical morphogenesis (FDR $q = 1.2 \times 10^{-3}$), among others (**SUPPLEMENTAL TABLES 12-14**). Examples of specific associations include adrenomedullin (*ADM*), a vasodilatory protein released by vascular endothelial cells in response to stress, which was associated with increased risk of varicose veins (OR 1.34, 95% CI 1.25 to 1.44, $p = 1 \times 10^{-16}$). Microfibrillar-associated protein 2 (*MFAP2*), an extracellular matrix protein which interacts with fibrillin (implicated in vascular diseases like Marfan syndrome), and associated with vascular phenotypes in zebrafish, was also associated with increased risk of varicose veins (OR 1.24, 95% CI 1.16 to 1.33, $p = 2 \times 10^{-10}$). We also identified proteins where higher circulating levels were associated with decreased risk of varicose veins, including Periostin (*POSTN*), a secreted

extracellular matrix protein which binds integrins to support migration and adhesion of epithelial cells (OR 0.83, 95% CI 0.78 to 0.89, $p = 4 \times 10^{-9}$).

Drug-repurposing Mendelian randomization

We next evaluated whether drug targets of currently approved or clinical-stage therapeutics might be repurposed to modify risk of varicose veins. Two-sample Mendelian randomization was performed using previously defined genetic instruments proxying gene expression levels (*cis*-eQTLs) of drug targets in ChEMBL v.26.²⁹ Of 983 actionable therapeutic targets with corresponding SNP effects for varicose veins, expression levels of 69 unique targets were significantly associated (FDR $q < 0.05$) with risk of varicose veins (**FIGURE 6; SUPPLEMENTAL TABLE 15**). Drug targets significantly associated with varicose veins included *PDE4C*, which encodes phosphodiesterase 4C. Although no medications are broadly available for treating varicose veins, pentoxifylline (a non-specific phosphodiesterase inhibitor) has been therapeutically applied for treatment of venous ulceration.³⁰ Other drug targets with repurposing potential include *ITGB3* and *F2*, which encode integrin beta 3 and thrombin, respectively. Both targets implicate roles for thrombosis and coagulation, and can be therapeutically modulated using GPIIb/IIIa and direct-acting thrombin inhibitors, respectively.

Topological enrichment of varicose veins-associated biological pathways

Finally, we constructed an integrated mechanistic conceptual model of varicose vein pathophysiological processes (**FIGURE 7**). Genome-wide significant SNPs were mapped to genes using H-MAGMA (**SUPPLEMENTARY FIGURE 6A**) to integrate publicly available chromatin interaction data from human umbilical vein endothelial cells (GEO: GSE121520) with traditional proximity-based SNP-gene mapping.³¹ Genes were tested for standard enrichment of GO terms, as well as rank-based topological enrichment using random walk with restart (RWR) algorithms that traverse a multiplex network integrating 12 layers of omics data across relevant human tissues (**SUPPLEMENTARY FIGURE 6A**). We specifically considered gene enrichment in six GO biological processes associated with varicose veins pathophysiology: angiogenesis, blood pressure regulation, cytokine production, extracellular matrix (ECM) disassembly, leukocyte migration, and response to fluid shear stress. GWAS genes assigned by H-MAGMA were more topologically enriched in these six biological processes compared to random gene sets as determined by area under precision-recall curve (AUPRC; **SUPPLEMENTARY FIGURE 6B, 6C**). Furthermore, the GWAS genes were more topologically enriched for these six biological processes than nine GO biological processes not known to be associated with varicose vein pathophysiology (**SUPPLEMENTARY FIGURE 6C**).

As seen in the mechanistic conceptual model depicted in **FIGURE 7**, using gene set and topological enrichment we identified GWAS genes related to altered blood pressure and endothelial shear stress (9

genes), angiogenesis and wound healing (12 genes), ECM and vascular wall remodeling (13 genes), leukocyte adhesion and invasion (7 genes), and inflammation and T-cell activation (23 genes). We highlight several examples of enriched gene sets which recapitulate molecular mechanisms of varicose veins. For example, endothelin-1 (*EDN1*), the endothelin-A receptor (*EDNRA*), and the transcription factor *GATA2*, are known to contribute to both angiogenesis and hypertension in varicose vein patients through a vascular endothelial growth factor (VEGF)-mediated mechanism that is sensed by *PIEZO1*.³²⁻³⁵ The transcription factor *KLF5* similarly promotes angiogenesis by its transcriptional activation of *VEGFA* (one of the genes encoding VEGF) and increases vascular wall remodeling by causing transcription of the matrix metalloprotease genes *MMP2* and *MMP3*.³⁶ Additional matrix metalloproteases (*ADAM15*, *ADAMTS4*, *ADAMTS13*) associated with increased endothelial permeability in venous disease were also implicated.³⁷⁻⁴⁰ Immune-related genes included *DOCK8* (dendritic cell migration), *AIF1* (dendritic cell differentiation and macrophage migration), and several genes implicated in T-cell activation and inflammatory processes including *PPP3R1* (calcineurin subunit), *CABIN1* (calcineurin binding protein), two NFAT subunits (*NFATC1* and *NFATC2*), MHC class I genes (*HLA-A*, *HLA-B*, *HLA-C*, *MICA*, and *MICB*) that lead to antigen presentation to T-cells, and several butyrophilin genes (*BTN1A1*, *BTN2A1*, *BTN3A1*, *BTN3A2*, *BTN3A3*) which modulate T-cell cytokine release.⁴¹⁻⁴⁵ Finally, the pro-inflammatory cytokines *IL6* and *LTA* were identified in tandem with *IL27*, a cytokine that induces T-cell differentiation and has been implicated in cardiovascular disease more broadly.⁴⁶ These topological enrichment findings demonstrate that genetic variants present in varicose vein patients influence the major pathophysiologic processes that arise in the course of disease.

DISCUSSION

This GWAS meta-analysis represents the largest multi-ancestry analysis of varicose veins to date. Overall, we identified 139 risk loci, establishing novel credible risk variants. We characterized the shared genetic architecture of varicose veins with other traits at the locus- and polygenic levels. Mendelian randomization defined putative causal clinical risk factors and identified circulating proteins and drug targets that may represent causal treatment targets. Finally, our integrative topological enrichment analysis systematically linked human genetic data with major biological processes implicated in varicose veins.

First, our findings build on prior studies demonstrating the heritability of varicose veins and extend these findings using larger cohorts of diverse genetic ancestry. Our identification of 139 varicose veins risk loci represents a 3.8-fold improvement in genomic discovery compared with prior reports, likely a result of increased sample size and cohort diversity. These findings are broadly consistent with similarly large and

diverse genome-wide analyses of complex traits.⁴⁷ Further, we find evidence of genetic overlap between both venous and arterial disease. This finding is supported at the polygenic level by significant genetic correlation between varicose veins and both arterial (peripheral artery disease and abdominal aortic aneurysm), and venous disease (venous thromboembolism). We demonstrate additional overlap at the locus-level. For instance, F5 p.Arg534Gln (factor V Leiden), a commonly-recognized risk factor for venous thromboembolism, is now identified as a novel risk locus for varicose veins. Recent evidence further suggests this risk variant contributes to risk of peripheral artery disease, blurring the traditional differences in arterial and venous risk factors.⁴⁸

Second, our findings implicate pathways relevant to both primary and secondary etiologies of varicose veins. Clinically, varicose veins are classified based on the CEAP classification scheme, where varicose veins occur based on congenital, primary, secondary, or no identifiable venous etiologies.⁹ Although we did not detect substantial overlap with Mendelian diseases associated with varicose veins, we did find genetic architecture relevant to both primary and secondary etiologies. Focusing on genome-wide significant loci, both pathway enrichment and colocalization analyses implicated vascular pathways and proteins relevant to primary varicose veins. For example, our prioritization of vascular smooth muscle cell pathways is consistent with *in vitro* evidence finding dysregulated smooth muscle cell phenotypes in varicose veins in comparison to healthy control vessels.⁴⁹ Similarly, MAGMA tissue and pathway enrichment at the polygenic level highlighted vascular tissues and pathways, while Mendelian randomization identified causal vascular proteins and drug targets. At both the locus and polygenic levels we also identified genetic overlap with secondary causes of varicose veins (Eg. the *FTO* locus and genetic correlation with anthropometric markers like obesity; the *F5* locus and genetic correlation with venous thromboembolism). While these findings support biological pathways relevant to the current varicose veins classification scheme, future study will be required to determine whether increased polygenic risk specifically within vascular pathways is necessary/sufficient for disease, and whether gene-environment interactions (particularly obesity) alter disease risk.

Third, our findings highlight the pleiotropic effects of obesity and related anthropometric traits. Our genetic correlation and Mendelian randomization analyses demonstrate that vascular and anthropometric traits are closely linked, and that body composition adversely influences vascular disease risk more broadly. While some markers of obesity and body composition have been previously linked to varicose veins,^{16,18} we extend those findings to demonstrate shared patterns of risk across common vascular diseases. Although varicose veins (in contrast to diseases like coronary/peripheral artery disease) are not likely to be the driving force behind public health efforts to curb obesity, our findings suggest that targeting obesity and related markers of body

composition would be expected to generally improve the vascular risk profile, reducing the burden of varicose veins as a beneficial consequence.

Fourth, we identify novel pathways relevant to the pathogenesis of varicose veins. For example, we identify a novel risk locus (rs55750792) near microfibrillar associated protein 2 (MFAP2; also known as MAGP1). Our protein Mendelian randomization analysis identified a strong association between circulating levels MFAP2, and increased risk of varicose veins, a finding reinforced by significant colocalization between varicose veins and circulating MFAP2 (posterior probability 0.95). MFAP2 is an extracellular matrix protein with a carboxy-terminal fibrillin and tropoelastin-binding domains and an amino-terminal domain capable of binding active transforming growth factor beta (TGF- β).⁵⁰⁻⁵² Several plausible mechanisms link MFAP2 to risk of varicose veins. At the extracellular matrix, disorganized fibrillin distribution and increased TGF expression have been observed within varicose veins in comparison with health control vessels.⁵³ Consistent with these findings, we observed a novel risk variant (rs12123737) at the TGFBR3 locus, and find that genetically-proxied circulating levels of TGFBI (TGF- β induced protein) were also associated with increased risk of varicose veins. Prior observations found significantly increased TGFBI levels in varicose veins in comparison with normal control vessels.⁵⁴ MFAP2 also binds NOTCH1, and crosstalk between notch, TGF, and vascular endothelial growth factor pathways play key roles in vascular morphogenesis. We identify a novel varicose veins risk locus (rs507666), which also represents a pleiotropic *trans*-pQTL for members of the vascular endothelial growth factor receptor family (VEGFR2 and VEGFR3). Further, rs72775784 at the NOTCH1 locus represents another novel risk variant for varicose veins. In addition to the functions of MFAP2 in the extracellular matrix and vascular morphogenesis which may lead to primary varicose veins, recent evidence suggests MAFP2 binds several hemostatic factors including fibronectin, fibrinogen, and von Willebrand factor, contributing to hemostasis and thrombosis, suggesting an additional mechanism for MFAP2 relevant to secondary causes of varicose veins.⁵⁵

Fifth, we identify drug-repurposing opportunities, providing human genetic evidence to motivate additional mechanistic studies and clinical trials. For example, our prioritization of *PDE4C* provides human genetic evidence to study the role of pentoxifylline for the treatment of varicose veins, while prioritization of obesity (eg. *LEPR*), thrombosis (eg. *ITGB3*), and coagulation (eg. *F2*) targets reinforce the role of these pathways for secondary forms of varicose veins. Future mechanistic evaluation of these targets for varicose veins is warranted, however the genetic evidence for these pathways increases the likelihood of successful clinical translation.^{56,57}

Finally, our systems biology-inspired enrichment analysis expands the number of loci linked to biological processes known to be disrupted in varicose vein pathology. By leveraging tissue/cell-type specific epigenetic data, we refine SNP-to-gene assignment compared to methods applied in previous varicose vein GWAS.^{16,18} By combining standard GO enrichment with a novel RWR approach, we demonstrate that genes from the present GWAS are enriched for functional connectivity to biological pathways important to varicose veins. This novel gene set enrichment analysis differs from standard gene set enrichment in that it does not rely entirely on the intersection of GWAS genes with genes in a GO term. Instead, enrichment is determined by demonstrating that both GWAS genes and genes pertaining to a GO term exhibit similar network topological connectivity across a multiplex network integrating 12 layers of multi-omics data using ranks determined by RWR, indicating that they are involved in similar biological processes. Thus, by using our network-based approach, here we link genetic variants from GWAS into a systems-level integration of biological pathways involved in varicose vein pathophysiology.

Limitations

Our findings should be interpreted within the context of several limitations. First, we sought to assemble the broadest and most diverse set of individuals with varicose veins to improve power for genetic discovery, recognizing that concurrent formal replication in a similarly diverse and well-powered cohort would be unlikely. However, we apply complementary colocalization, genetic correlation, Mendelian randomization, and systems biology methods to prioritize loci and pathways for further study.⁵⁸ Second, our definition of varicose veins was based on prevalent diagnosis codes ascertained from the electronic health record. Although several varicose veins classifications/subtypes have been described based on different pathophysiologic mechanisms, identifying varicose veins subtypes at scale using the electronic health record remains limited. While we identify pathways relevant to both primary and secondary subtypes of varicose veins, interactions between these pathways and their relevance across individual subtypes requires further study. Further, the VA Million Program cohort is >90% male, and we do not explore sex-specific differences. While Mendelian randomization has successfully predicted the therapeutic effects of drug targets, the method formally proxies lifelong genetic exposure rather than short-term pharmacologic interventions. The therapeutic effect sizes predicted in our analyses may therefore differ in comparison to pharmacologic treatment.

Conclusion

In the largest and most diverse genome wide association study of varicose veins to-date, we identify novel genetic risk loci for varicose veins, highlight shared risk factors with other vascular diseases, connect genetic

variants to biological pathways with multi-omic network analysis, and identify novel therapeutic targets and drug repurposing opportunities.

METHODS

Study Populations

Individuals of diverse ancestry with and without varicose veins were identified using electronic health record diagnosis codes linked to genetic data from the VA Million Veteran Program,²⁰ UK Biobank,⁵⁹ eMERGE (<https://emerge-network.org>), FinnGen (<https://finngen.gitbook.io/documentation/>), and BioBank Japan (<http://jenger.riken.jp/en/>). The VA Million Veteran program recruits adults age >18 from Veterans Affairs Medical Centers across the USA. From MVP, we evaluated 18,977 cases (15,240 European-ancestry; 2511 African-ancestry; 1,226 Hispanic-ancestry) and 561,203 controls (409,037 European-ancestry; 106,998 African-ancestry; 45,168 Hispanic-ancestry) from Release 4. From UK Biobank, eMERGE, FinnGen, and BioBank Japan, we evaluated an additional 30,788 cases and 773,098 controls. Details of cohorts and phenotype definitions are available in the **Supplemental Methods**. All studies received ethical and protocol approval from the relevant institutional review boards, and all participants provided informed consent. The VA Million Veteran Program was approved by the VA Central Institutional Review Board

Genome Wide Association Study Meta-Analysis

Details of study-specific genotyping and quality control are available in the **Supplemental Methods**. Briefly, ancestry-specific genome wide association studies were performed within each cohort, including covariates for age, sex, and population structure. MVP analyses were performed using logistic regression in plink2a,^{60,61} eMERGE analyses were performed using regenie,⁶² and publicly-available GWAS from UK Biobank, FinnGen and BioBank Japan were performed using SAIGE,⁶³ which are computationally-efficient methods for performing GWAS within biobank-scale populations. GWASinspector⁶⁴ was used to perform study-level quality control (using default settings), and to harmonize effect alleles across studies. Fixed-effects inverse variance-weighted GWAS meta-analysis was performed using METAL,⁶⁵ first within- and then across ancestries to generate a multi-ancestry meta-analysis. Independent variants and loci associated with varicose veins were identified using FUMA v1.3.6d (<https://fuma.ctglab.nl/>),⁶⁶ using a genome wide-significance threshold of $p < 5 \times 10^{-8}$, and default parameters for pruning independent variants/loci (merge-distance 250kb; linkage disequilibrium $r^2=0.6$, $r_2^2 = 0.1$). All genomic positions are reported using coordinates from the GRCh37 build of the human genome.

Fine-mapping

Multi-ancestry fine-mapping was performed using MR-MEGA.²³ This statistical method performs meta-regression, using principal components to account for heterogeneity in allelic effects that is correlated with ancestry. We considered study-level associations at each genome-wide significant locus, using the default settings to account for 4 genetic principal components. At each genome-wide significant locus (+/- 500kb of each lead variant) we derived 99% credible sets by ranking variants by Bayes' factor and including variants until the cumulative posterior probability reached 0.99.

Tissue and Gene-set Enrichment

Tissue and gene-set enrichment were evaluated using MAGMA v1.08, via the FUMA web application.^{66,67} MAGMA aggregates SNP-level associations into gene-level test statistics. For tissue enrichment analysis, MAGMA performs gene property analysis, accounting for average tissue expression across 54 tissues (identified using GTEx v8).⁶⁸ For gene-set enrichment, 15,485 gene sets ("Curated gene sets" and "GO terms" from MSigDB v7.0)⁶⁹ were tested for enrichment with varicose veins risk loci.

Phenome-wide association study (pheWAS)

Lead variants at each of the 139 loci significantly associated with varicose veins were queried using the UK Medical Research Council Integrative Epidemiology Unit OpenGWAS project (<https://gwas.mrcieu.ac.uk/>).^{70,71} We retrieved associations between these 139 varicose veins-associated variants and traits for which prior genome-wide association studies had been performed, focusing on UK Biobank diagnoses/operations/medications/surveys/measurements (n = 2514; from the "ukb-b" dataset), circulating proteins (n = 4489; from the "prot-" datasets), and GWAS of circulating metabolites (n = 974; from the "met-" datasets). Given some overlap of proteins and metabolites across individual datasets, we considered the conventional genome-wide association study significance threshold of $p < 5 \times 10^{-8}$, which is similar to p-value threshold of $0.05/[139 \times (2514 + 4489 + 974)]$ when using the Bonferroni adjustment for multiple testing.

Colocalization

Bayesian colocalization was performed using the default parameters of HyPrColoc.²⁴ HyPrColoc enables colocalization between arbitrary-sized groups of traits, remaining computationally efficient even when evaluating hundreds of traits simultaneously. We performed colocalization between each varicose veins risk locus (+/- 500kb) and traits that shared risk loci (identified using our pheWAS). Colocalization was first performed pairwise, individually testing for colocalization between varicose veins and each trait sharing a risk locus. Next, multi-trait colocalization was performed, simultaneously evaluating for colocalization between

varicose veins and all traits sharing a given risk locus. The default posterior probability threshold $P_{RPA} > 0.49$ was used to identify significant colocalization.

Genetic Correlation

Genetic correlations were estimated using cross-trait LD-score regression, using LDSC v1.0.1 (<https://github.com/bulik/ldsc>).^{21,72} When estimating correlations with vascular traits, summary statistics were obtained from large genome wide association studies of coronary artery disease, peripheral artery disease, ischemic stroke, venous thromboembolism, and abdominal aortic aneurysm.^{48,73–76} When estimating correlations with anthropometric traits, summary statistics were obtained from publicly-available genome-wide association studies of anthropometric traits measured among UK Biobank participants, identified from the “ukb-b” dataset of the MRC-IEU OpenGWAS project (<https://gwas.mrcieu.ac.uk/>).^{70,71}

Anthropometric Mendelian Randomization

Genetic instruments for anthropometric traits were constructed from independent ($r^2 < 0.001$, distance = 10,000kb) genetic variants associated with each anthropometric trait at genome-wide significance ($p < 5 \times 10^{-8}$), from publicly available GWAS (described in Genetic Correlation section above). Corresponding SNP effects in each vascular trait (varicose veins, coronary artery disease, peripheral artery disease, venous thromboembolism, and abdominal aortic aneurysm) were identified, and harmonized to ensure concordant effect alleles. Inverse variance-weighted two-sample Mendelian randomization was then performed using the *TwoSampleMR* package in R.⁷⁰ As a sensitivity analysis we applied the weighted-median method, which makes different assumptions about the presence of pleiotropy and invalid instruments.⁷⁷ To account for multiple testing, the false discovery rate was controlled at 5%.

pQTL Mendelian Randomization

Details of genetic instrument selection and validation have been previously reported.²⁸ Briefly, we considered high-confidence (tier 1) genetic instruments which were identified based on independent ($r^2 < 0.001$) *cis*-pQTLs which were associated with circulating protein levels at a GWAS p-value threshold $\leq 5 \times 10^{-8}$, and passed both specificity and consistency tests.²⁸ Of genetic instruments for 738 circulating proteins, corresponding varicose veins outcome data was available for 714 (96.7%). For each of the 714 genetic instruments (each representing a single locus/protein), Mendelian randomization was performed using the Wald ratio method.⁷⁸ To account for multiple testing, the false discovery rate was controlled at 5%.

Drug-Repurposing Mendelian Randomization

Details of genetic instrument selection have been previously reported.²⁹ Briefly, genes encoding the targets of approved or clinical-stage therapeutics were identified using ChEMBL v.26.⁷⁹ Conditionally-independent *cis*-eQTLs for each gene was then identified from GTEx v8.⁶⁸ When multiple instruments from different GTEx v8 tissues were present, each tissue was tested independently. Two sample Mendelian Randomization was then performed to estimate the effect of transcript levels of each drug target on varicose veins. Inverse-variance weighting was used to combine effects when multiple genetic instruments for a transcript were available. To account for multiple testing, the false discovery rate was controlled at 5%.

Hi-C data processing and H-MAGMA SNP-to-gene assignment

Hi-C data from control human umbilical vein endothelial cells (HUVECs) were downloaded from publicly available data (GEO: GSE121520)⁸⁰ aligned to reference genome hg19/GRCh37. Significant chromatin contacts were determined at 5kb resolution using Fit-Hi-C using 2 spline passes with lower bound threshold of 10kb, upper bound threshold of 2Mb, and significant contacts defined as $p < 0.01$.^{81,82} We used this information to generate an annotation file for H-MAGMA for HUVEC-specific SNP-to-gene assignment.³¹ 8,123 SNPs exceeding genome-wide significance ($p < 5 \times 10^{-8}$) were included for H-MAGMA gene assignment, resulting in 638 unique genes identified (**Supplementary Table 16**).

Multiplex network generation

A multiplex network (a homogenous multilayer network where individual layer topologies are retained) containing gene nodes was created from 12 layers of omics data using the RandomWalkRestartMH R package.⁸³ Protein-protein interactions were merged into a single network layer from HumanNetv2 and STRING (v11) databases.^{84,85} Co-citation, co-essentiality, co-expression, pathway database, gene neighborhood, interlog, and phylogenetic network layers from HumanNetv2 were also included as individual layers. Transcription factor - target networks were downloaded from hTFtarget⁸⁶ on 2/20/2021 and merged into a single network layer derived from the following 15 HUVEC ChIP-seq GEO datasets: BRD4, RELA: GSE53998;⁸⁷ CTCF: GSE29611;⁸⁸ ELK3: GSE60156;⁸⁹ EP300, ETS1: GSE41166;⁸⁹ EWSR1, FLI1: GSE31838;⁹⁰ FOS, GATA2, JUN, MAX: GSE31477;⁸⁸ MYC: GSE33213; PGR: GSE43786;⁹¹ PPAR/PPARG: GSE50144.⁹² Three tissue type-specific predictive gene expression networks were generated using the explainable-AI algorithm iterative random forest leave one-out prediction (iRF-LOOP)^{93,94} from the following GTEx v8 RNA-seq data sets downloaded from the GTEx Portal on 10/13/2020: tibial artery, whole blood/EBV-transformed lymphocytes, and liver.⁹⁵

Network-based topological functional enrichment

The set of significantly associated GWAS genes from H-MAGMA was tested for enrichment for genes belonging to six key Gene Ontology (GO) annotations by evaluating their connectivity in the multiplex network: GO:0001525: angiogenesis; GO: 0002367: cytokine production involved in immune response; GO:0008217: regulation of blood pressure; GO:0022617: extracellular matrix disassembly; GO:0034405: response to fluid shear stress; GO:0050900: leukocyte migration. The entire GWAS geneset was used as seed genes for the Random Walk with Restart (RWR) algorithm, producing a ranking of all other genes in the multiplex according to their joint topological connectivity to the GWAS genes. Area Under Precision Recall Curve (AUPRC) was calculated and used to score the ability for the random walker to effectively find each target geneset when starting from the GWAS genes. The same was done for 20 random genesets of the same size as each target gene set to determine if the score was better than random. Nine non-target, brain-related GO biological process gene sets were also scored for comparison purposes: GO:0001963: synaptic transmission, dopaminergic; GO:0007411: axon guidance; GO:0021766: hippocampus development; GO:0021854: hypothalamus development; GO:0035249: synaptic transmission, glutamatergic; GO:0042165: neurotransmitter binding; GO:0051932: synaptic transmission, GABAergic; GO:0098664: G protein-coupled serotonin receptor signaling pathway; GO:1903859: regulation of dendrite extension. GWAS genes were also tested for standard GO enrichment using ToppGene (<https://toppgene.cchmc.org/>).⁹⁶ GO-enriched and topologically enriched gene lists from H-MAGMA genes are included in **Supplementary Tables 17-18**.

ACKNOWLEDGEMENTS

The authors thank the participants of the VA Million Veteran Program, UK Biobank, eMERGE, FinnGen, and BioBank Japan studies.

DATA AVAILABILITY

GWAS meta-analysis summary statistics will be deposited in dbGaP upon final publication. GWAS summary statistics from FinnGen, BioBank Japan, and UK Biobank are publicly available from <https://r4.finnngen.fi/>, <https://pheweb.jp/>, and <https://pan.ukbb.broadinstitute.org>.

FUNDING

MGL is supported by the Institute for Translational Medicine and Therapeutics of the Perelman School of Medicine at the University of Pennsylvania and the NIH/NHLBI National Research Service Award postdoctoral fellowship (T32HL007843). SMD is supported by the US Department of Veterans Affairs award #IK2-CX001780. This publication does not represent the views of the Department of Veterans Affairs or the United States

Government. SMD receives research support from RenalytixAI and personal consulting fees from Calico Labs, both outside the scope of the current manuscript. DAJ, KAS, DK, MRG, ML, AMC, and JR are supported by funding from the Million Veteran Program Computational Health Analytics for Medical Precision to Improve Outcomes Now (CHAMPION) initiative and NIH grants DA051908 and DA051913. eMERGE is supported by U01HG8657 (University of Washington); U01HG8685 (Brigham and Women's Hospital); U01HG8672 (Vanderbilt University Medical Center); U01HG8666 (Cincinnati Children's Hospital Medical Center); U01HG6379 (Mayo Clinic); U01HG8679 (Geisinger Clinic); U01HG8680 (Columbia University Health Sciences); U01HG8684 (Children's Hospital of Philadelphia); U01HG8673 (Northwestern University); U01HG8701 (Vanderbilt University Medical Center serving as the Coordinating Center); U01HG8676 (Partners Healthcare/Broad Institute); and U01HG8664 (Baylor College of Medicine). This research used resources of the Oak Ridge Leadership Computing Facility, which is a DOE Office of Science User Facility supported under Contract DE-AC05-00OR22725. This manuscript has been co-authored by UT-Battelle, LLC under contract no. DE-AC05-00OR22725 with the U.S. Department of Energy. The United States Government retains and the publisher, by accepting the article for publication, acknowledges that the United States Government retains a nonexclusive, paid-up, irrevocable, world-wide license to publish or reproduce the published form of this manuscript, or allow others to do so, for United States Government purposes. The Department of Energy will provide public access to these results of federally sponsored research in accordance with the DOE Public Access Plan (<http://energy.gov/downloads/doe-public-access-plan>, last accessed September 16, 2020).

REFERENCES

1. Criqui, M. H. *et al.* Chronic venous disease in an ethnically diverse population: the San Diego Population Study. *Am. J. Epidemiol.* **158**, 448–456 (2003).
2. Evans, C. J., Fowkes, F. G., Ruckley, C. V. & Lee, A. J. Prevalence of varicose veins and chronic venous insufficiency in men and women in the general population: Edinburgh Vein Study. *J. Epidemiol. Community Health* **53**, 149–153 (1999).
3. Brand, F. N., Dannenberg, A. L., Abbott, R. D. & Kannel, W. B. The epidemiology of varicose veins: the Framingham Study. *Am. J. Prev. Med.* **4**, 96–101 (1988).
4. Zahariev, T. *et al.* Prevalence of primary chronic venous disease: the Bulgarian experience. *Int. Angiol. J. Int. Union Angiol.* **28**, 303–310 (2009).
5. Callam, M. J. Epidemiology of varicose veins. *Br. J. Surg.* **81**, 167–173 (1994).
6. Carpentier, P. H., Maricq, H. R., Biro, C., Ponçot-Makinen, C. O. & Franco, A. Prevalence, risk factors, and clinical patterns of chronic venous disorders of lower limbs: a population-based study in France. *J. Vasc. Surg.* **40**, 650–659 (2004).
7. Cesarone, M. R. *et al.* ‘Real’ epidemiology of varicose veins and chronic venous diseases: the San Valentino Vascular Screening Project. *Angiology* **53**, 119–130 (2002).
8. Sisto, T. *et al.* Prevalence and risk factors of varicose veins in lower extremities: mini-Finland health survey. *Eur. J. Surg. Acta Chir.* **161**, 405–414 (1995).
9. Lurie, F. *et al.* The 2020 update of the CEAP classification system and reporting standards. *J. Vasc. Surg. Venous Lymphat. Disord.* **8**, 342–352 (2020).
10. Müller-Bühl, U. *et al.* Varicose veins are a risk factor for deep venous thrombosis in general practice patients. *VASA Z. Gefasskrankheiten* **41**, 360–365 (2012).
11. Chang, S.-L. *et al.* Association of Varicose Veins With Incident Venous Thromboembolism and Peripheral Artery Disease. *JAMA* **319**, 807–817 (2018).
12. Bergan, J. J. *et al.* Chronic venous disease. *N. Engl. J. Med.* **355**, 488–498 (2006).
13. Jacobs, B. N., Andraska, E. A., Obi, A. T. & Wakefield, T. W. Pathophysiology of varicose veins. *J. Vasc. Surg. Venous Lymphat. Disord.* **5**, 460–467 (2017).
14. Anwar, M. A., Shalhoub, J., Lim, C. S., Gohel, M. S. & Davies, A. H. The effect of pressure-induced mechanical stretch on vascular wall differential gene expression. *J. Vasc. Res.* **49**, 463–478 (2012).
15. Raffetto, J. D. Pathophysiology of Chronic Venous Disease and Venous Ulcers. *Surg. Clin. North Am.* **98**, 337–347 (2018).
16. Fukaya, E. *et al.* Clinical and Genetic Determinants of Varicose Veins. *Circulation* **138**, 2869–2880 (2018).
17. Anwar, M. A. *et al.* A Review of Familial, Genetic, and Congenital Aspects of Primary Varicose Vein Disease. *Circ. Cardiovasc. Genet.* **5**, 460–466 (2012).
18. Shadrina, A. S., Sharapov, S. Z., Shashkova, T. I. & Tsepilov, Y. A. Varicose veins of lower extremities: Insights from the first large-scale genetic study. *PLoS Genet.* **15**, e1008110 (2019).
19. Ellinghaus, E. *et al.* Genome-wide association analysis for chronic venous disease identifies EFEMP1 and KCNH8 as susceptibility loci. *Sci. Rep.* **7**, 45652 (2017).
20. Gaziano, J. M. *et al.* Million Veteran Program: A mega-biobank to study genetic influences on health and disease. *J. Clin. Epidemiol.* **70**, 214–223 (2016).
21. Bulik-Sullivan, B. *et al.* LD score regression distinguishes confounding from polygenicity in genome-wide association studies. *Nat. Genet.* (2015) doi:10.1038/ng.3211.
22. Piazza, G. Varicose Veins. *Circulation* **130**, 582–587 (2014).
23. Mägi, R. *et al.* Trans-ethnic meta-regression of genome-wide association studies accounting for ancestry increases power for discovery and improves fine-mapping resolution. *Hum. Mol. Genet.* **26**, 3639–3650 (2017).
24. Foley, C. N. *et al.* A fast and efficient colocalization algorithm for identifying shared genetic risk factors across multiple traits. *Nat. Commun.* **12**, 764 (2021).

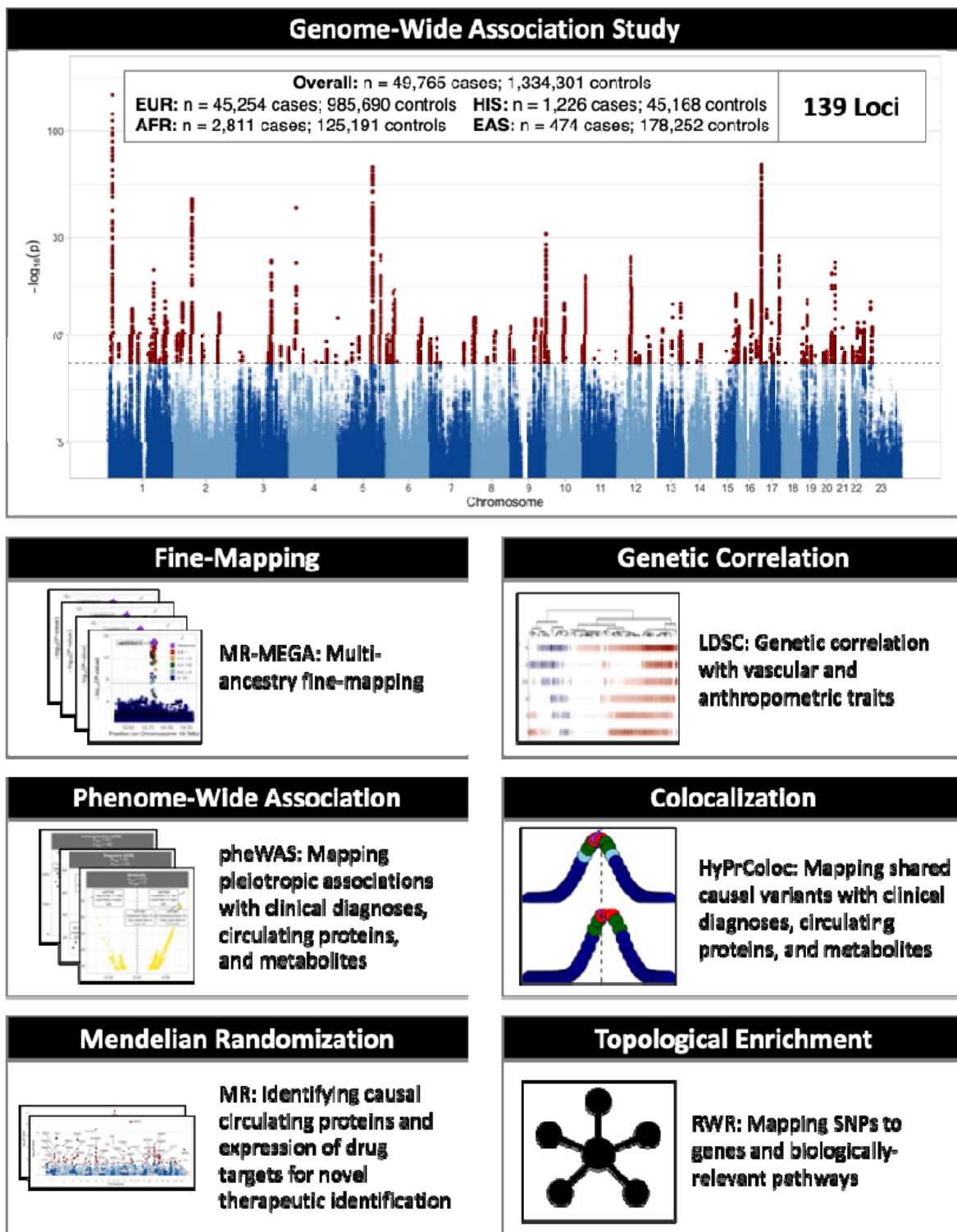
25. Paré, G. *et al.* Novel Association of ABO Histo-Blood Group Antigen with Soluble ICAM-1: Results of a Genome-Wide Association Study of 6,578 Women. *PLOS Genet.* **4**, e1000118 (2008).
26. Kiechl, S. *et al.* Association of variation at the ABO locus with circulating levels of soluble intercellular adhesion molecule-1, soluble P-selectin, and soluble E-selectin: a meta-analysis. *Circ. Cardiovasc. Genet.* **4**, 681–686 (2011).
27. Nelson, C. P. *et al.* Genetically Determined Height and Coronary Artery Disease. *N. Engl. J. Med.* **372**, 1608–1618 (2015).
28. Zheng, J. *et al.* Phenome-wide Mendelian randomization mapping the influence of the plasma proteome on complex diseases. *Nat. Genet.* **52**, 1122–1131 (2020).
29. Gaziano, L. *et al.* Actionable druggable genome-wide Mendelian randomization identifies repurposing opportunities for COVID-19. *Nat. Med.* **27**, 668–676 (2021).
30. Jull, A. B., Arroll, B., Parag, V. & Waters, J. Pentoxifylline for treating venous leg ulcers. *Cochrane Database Syst. Rev.* **2012**, CD001733 (2012).
31. Sey, N. Y. A. *et al.* A computational tool (H-MAGMA) for improved prediction of brain-disorder risk genes by incorporating brain chromatin interaction profiles. *Nat. Neurosci.* **23**, 583–593 (2020).
32. Hyman, A. J., Tumova, S. & Beech, D. J. Piezo1 Channels in Vascular Development and the Sensing of Shear Stress. *Curr. Top. Membr.* **79**, 37–57 (2017).
33. Purgatorio, G. *et al.* Germline GATA2 variant disrupting endothelial eNOS cell function and angiogenesis can be restored by c-Jun/AP-1 upregulation. *Haematologica* (2021) doi:10.3324/haematol.2021.278450.
34. Mangiafico, R. A., Malatino, L. S., Santonocito, M., Spada, R. S. & Benedetto, F. A. Plasma endothelin-1 release in normal and varicose saphenous veins. *Angiology* **48**, 769–774 (1997).
35. Neves, K. B., Montezano, A. C., Lang, N. N. & Touyz, R. M. Vascular toxicity associated with anti-angiogenic drugs. *Clin. Sci. Lond. Engl.* **1979** **134**, 2503–2520 (2020).
36. Xie, Z. *et al.* Current knowledge of Krüppel-like factor 5 and vascular remodeling: providing insights for therapeutic strategies. *J. Mol. Cell Biol.* **13**, 79–90 (2021).
37. Pagliari, M. T. *et al.* ADAMTS13 activity, high VWF and FVIII levels in the pathogenesis of deep vein thrombosis. *Thromb. Res.* **197**, 132–137 (2021).
38. Kenagy, R. D., Min, S.-K., Mulvihill, E. & Clowes, A. W. A link between smooth muscle cell death and extracellular matrix degradation during vascular atrophy. *J. Vasc. Surg.* **54**, 182-191.e24 (2011).
39. Chatterjee, V. *et al.* MicroRNA-147b regulates vascular endothelial barrier function by targeting ADAM15 expression. *PLoS One* **9**, e110286 (2014).
40. Charrier-Hisamuddin, L., Laboisse, C. L. & Merlin, D. ADAM-15: a metalloprotease that mediates inflammation. *FASEB J. Off. Publ. Fed. Am. Soc. Exp. Biol.* **22**, 641–653 (2008).
41. Kunimura, K., Uruno, T. & Fukui, Y. DOCK family proteins: key players in immune surveillance mechanisms. *Int. Immunol.* **32**, 5–15 (2020).
42. Elizondo, D. M. *et al.* Allograft Inflammatory Factor-1 Governs Hematopoietic Stem Cell Differentiation Into cDC1 and Monocyte-Derived Dendritic Cells Through IRF8 and RelB in vitro. *Front. Immunol.* **10**, 173 (2019).
43. Yang, Z. F. *et al.* Allograft inflammatory factor-1 (AIF-1) is crucial for the survival and pro-inflammatory activity of macrophages. *Int. Immunol.* **17**, 1391–1397 (2005).
44. Sun, L. *et al.* Cabin 1, a negative regulator for calcineurin signaling in T lymphocytes. *Immunity* **8**, 703–711 (1998).
45. Afrache, H., Gouret, P., Ainouche, S., Pontarotti, P. & Olive, D. The butyrophilin (BTN) gene family: from milk fat to the regulation of the immune response. *Immunogenetics* **64**, 781–794 (2012).
46. van der Heijden, T., Bot, I. & Kuiper, J. The IL-12 cytokine family in cardiovascular diseases. *Cytokine* **122**, 154188 (2019).
47. Zhou, W. *et al.* Global Biobank Meta-analysis Initiative: powering genetic discovery across human diseases. 2021.11.19.21266436 (2021) doi:10.1101/2021.11.19.21266436.

48. Klarin, D. *et al.* Genome-wide association study of peripheral artery disease in the Million Veteran Program. *Nat. Med.* **25**, 1274–1279 (2019).
49. Xiao, Y., Huang, Z., Yin, H., Lin, Y. & Wang, S. In vitro differences between smooth muscle cells derived from varicose veins and normal veins. *J. Vasc. Surg.* **50**, 1149–1154 (2009).
50. Finnis, M. L. & Gibson, M. A. Microfibril-associated glycoprotein-1 (MAGP-1) binds to the pepsin-resistant domain of the alpha3(VI) chain of type VI collagen. *J. Biol. Chem.* **272**, 22817–22823 (1997).
51. Brown-Augsburger, P. *et al.* Microfibril-associated glycoprotein binds to the carboxyl-terminal domain of tropoelastin and is a substrate for transglutaminase. *J. Biol. Chem.* **269**, 28443–28449 (1994).
52. Broekelmann, T. J., Bodmer, N. K. & Mecham, R. P. Identification of the growth factor-binding sequence in the extracellular matrix protein MAGP-1. *J. Biol. Chem.* **295**, 2687–2697 (2020).
53. Buján, J. *et al.* Expression of elastic components in healthy and varicose veins. *World J. Surg.* **27**, 901–905 (2003).
54. Barallobre-Barreiro, J. *et al.* Extracellular matrix remodelling in response to venous hypertension: proteomics of human varicose veins. *Cardiovasc. Res.* **110**, 419–430 (2016).
55. Werneck, C. C. *et al.* Mice lacking the extracellular matrix protein MAGP1 display delayed thrombotic occlusion following vessel injury. *Blood* **111**, 4137–4144 (2008).
56. King, E. A., Davis, J. W. & Degner, J. F. Are drug targets with genetic support twice as likely to be approved? Revised estimates of the impact of genetic support for drug mechanisms on the probability of drug approval. *PLOS Genet.* **15**, e1008489 (2019).
57. Nelson, M. R. *et al.* The support of human genetic evidence for approved drug indications. *Nat. Genet.* **47**, 856–860 (2015).
58. Huffman, J. E. Examining the current standards for genetic discovery and replication in the era of mega-biobanks. *Nat. Commun.* **9**, 5054 (2018).
59. Sudlow, C. *et al.* UK Biobank: An Open Access Resource for Identifying the Causes of a Wide Range of Complex Diseases of Middle and Old Age. *PLoS Med.* **12**, e1001779 (2015).
60. Purcell, S. *et al.* PLINK: A tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.* **81**, 559–575 (2007).
61. Chang, C. C. *et al.* Second-generation PLINK: Rising to the challenge of larger and richer datasets. *GigaScience* **4**, (2015).
62. Mbatchou, J. *et al.* Computationally efficient whole-genome regression for quantitative and binary traits. *Nat. Genet.* 1–7 (2021) doi:10.1038/s41588-021-00870-7.
63. Zhou, W. *et al.* Efficiently controlling for case-control imbalance and sample relatedness in large-scale genetic association studies. *Nat. Genet.* **50**, 1335–1341 (2018).
64. Ani, A., van der Most, P. J., Snieder, H., Vaez, A. & Nolte, I. M. GWASinspector: comprehensive quality control of genome-wide association study results. *Bioinformatics* **37**, 129–130 (2021).
65. Willer, C. J., Li, Y. & Abecasis, G. R. METAL: Fast and efficient meta-analysis of genomewide association scans. *Bioinformatics* **26**, 2190–2191 (2010).
66. Watanabe, K., Taskesen, E., van Bochoven, A. & Posthuma, D. Functional mapping and annotation of genetic associations with FUMA. *Nat. Commun.* **8**, 1826 (2017).
67. de Leeuw, C. A., Mooij, J. M., Heskes, T. & Posthuma, D. MAGMA: Generalized Gene-Set Analysis of GWAS Data. *PLOS Comput. Biol.* **11**, e1004219 (2015).
68. GTEx Consortium. Human genomics. The Genotype-Tissue Expression (GTEx) pilot analysis: multitissue gene regulation in humans. *Science* **348**, 648–660 (2015).
69. Subramanian, A. *et al.* Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl. Acad. Sci.* **102**, 15545–15550 (2005).
70. Hemani, G. *et al.* The MR-base platform supports systematic causal inference across the human phenome. *eLife* **7**, e34408 (2018).
71. Elsworth, B. *et al.* The MRC IEU OpenGWAS data infrastructure. 2020.08.10.244293 (2020) doi:10.1101/2020.08.10.244293.

72. Bulik-Sullivan, B. *et al.* An atlas of genetic correlations across human diseases and traits. *Nat. Genet.* **47**, 1236–1241 (2015).
73. Nikpay, M. *et al.* A comprehensive 1000 Genomes-based genome-wide association meta-analysis of coronary artery disease. *Nat. Genet.* **47**, 1121–1130 (2015).
74. Klarin, D. *et al.* Genome-wide association analysis of venous thromboembolism identifies new risk loci and genetic overlap with arterial vascular disease. *Nat. Genet.* **51**, 1574–1579 (2019).
75. Klarin, D. *et al.* Genetic Architecture of Abdominal Aortic Aneurysm in the Million Veteran Program. *Circulation* **142**, 1633–1646 (2020).
76. Malik, R. *et al.* Multiancestry genome-wide association study of 520,000 subjects identifies 32 loci associated with stroke and stroke subtypes. *Nat. Genet.* **50**, 524–537 (2018).
77. Bowden, J., Davey Smith, G., Haycock, P. C. & Burgess, S. Consistent Estimation in Mendelian Randomization with Some Invalid Instruments Using a Weighted Median Estimator. *Genet. Epidemiol.* **40**, 304–14 (2016).
78. Lawlor, D. A., Harbord, R. M., Sterne, J. A. C., Timpson, N. & Davey Smith, G. Mendelian randomization: using genes as instruments for making causal inferences in epidemiology. *Stat. Med.* **27**, 1133–1163 (2008).
79. Mendez, D. *et al.* ChEMBL: towards direct deposition of bioassay data. *Nucleic Acids Res.* **47**, D930–D940 (2019).
80. Higashijima, Y. *et al.* Coordinated demethylation of H3K9 and H3K27 is required for rapid inflammatory responses of endothelial cells. *EMBO J.* **39**, e103949 (2020).
81. Ay, F., Bailey, T. L. & Noble, W. S. Statistical confidence estimation for Hi-C data reveals regulatory chromatin contacts. *Genome Res.* **24**, 999–1011 (2014).
82. Kaul, A., Bhattacharyya, S. & Ay, F. Identifying statistically significant chromatin contacts from Hi-C data with FitHiC2. *Nat. Protoc.* **15**, 991–1012 (2020).
83. Valdeolivas, A. *et al.* Random walk with restart on multiplex and heterogeneous biological networks. *Bioinforma. Oxf. Engl.* **35**, 497–505 (2019).
84. Hwang, S. *et al.* HumanNet v2: human gene networks for disease research. *Nucleic Acids Res.* **47**, D573–D580 (2019).
85. Szklarczyk, D. *et al.* STRING v11: protein-protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. *Nucleic Acids Res.* **47**, D607–D613 (2019).
86. Zhang, Q. *et al.* hTFtarget: A Comprehensive Database for Regulations of Human Transcription Factors and Their Targets. *Genomics Proteomics Bioinformatics* **18**, 120–128 (2020).
87. Brown, J. D. *et al.* NF- κ B directs dynamic super enhancer formation in inflammation and atherogenesis. *Mol. Cell* **56**, 219–231 (2014).
88. ENCODE Project Consortium. An integrated encyclopedia of DNA elements in the human genome. *Nature* **489**, 57–74 (2012).
89. Zhang, B. *et al.* A dynamic H3K27ac signature identifies VEGFA-stimulated endothelial enhancers and requires EP300 activity. *Genome Res.* **23**, 917–927 (2013).
90. Patel, M. *et al.* Tumor-specific retargeting of an oncogenic transcription factor chimera results in dysregulation of chromatin and transcription. *Genome Res.* **22**, 259–270 (2012).
91. Goddard, L. M. *et al.* Progesterone receptor in the vascular endothelium triggers physiological uterine permeability preimplantation. *Cell* **156**, 549–562 (2014).
92. Inoue, T. *et al.* Cross-enhancement of ANGPTL4 transcription by HIF1 alpha and PPAR beta/delta is the result of the conformational proximity of two response elements. *Genome Biol.* **15**, R63 (2014).
93. Cliff, A. *et al.* A High-Performance Computing Implementation of Iterative Random Forest for the Creation of Predictive Expression Networks. *Genes* **10**, E996 (2019).
94. Basu, S., Kumbier, K., Brown, J. B. & Yu, B. Iterative random forests to discover predictive and stable high-order interactions. *Proc. Natl. Acad. Sci. U. S. A.* **115**, 1943–1948 (2018).

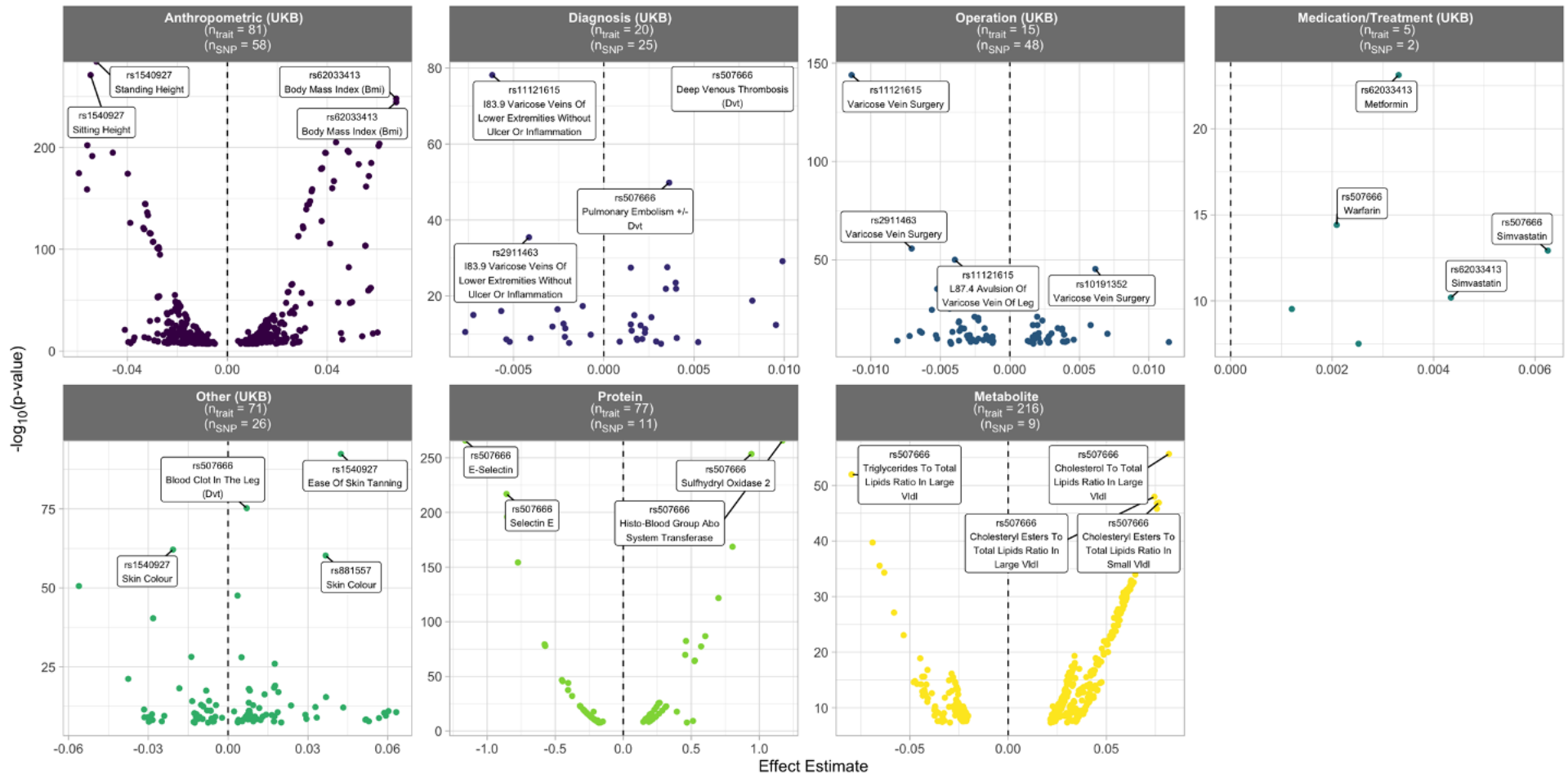
95. GTEx Consortium. The GTEx Consortium atlas of genetic regulatory effects across human tissues. *Science* **369**, 1318–1330 (2020).
96. Chen, J., Bardes, E. E., Aronow, B. J. & Jegga, A. G. ToppGene Suite for gene list enrichment analysis and candidate gene prioritization. *Nucleic Acids Res.* **37**, W305-311 (2009).

FIGURE 1 Study Design



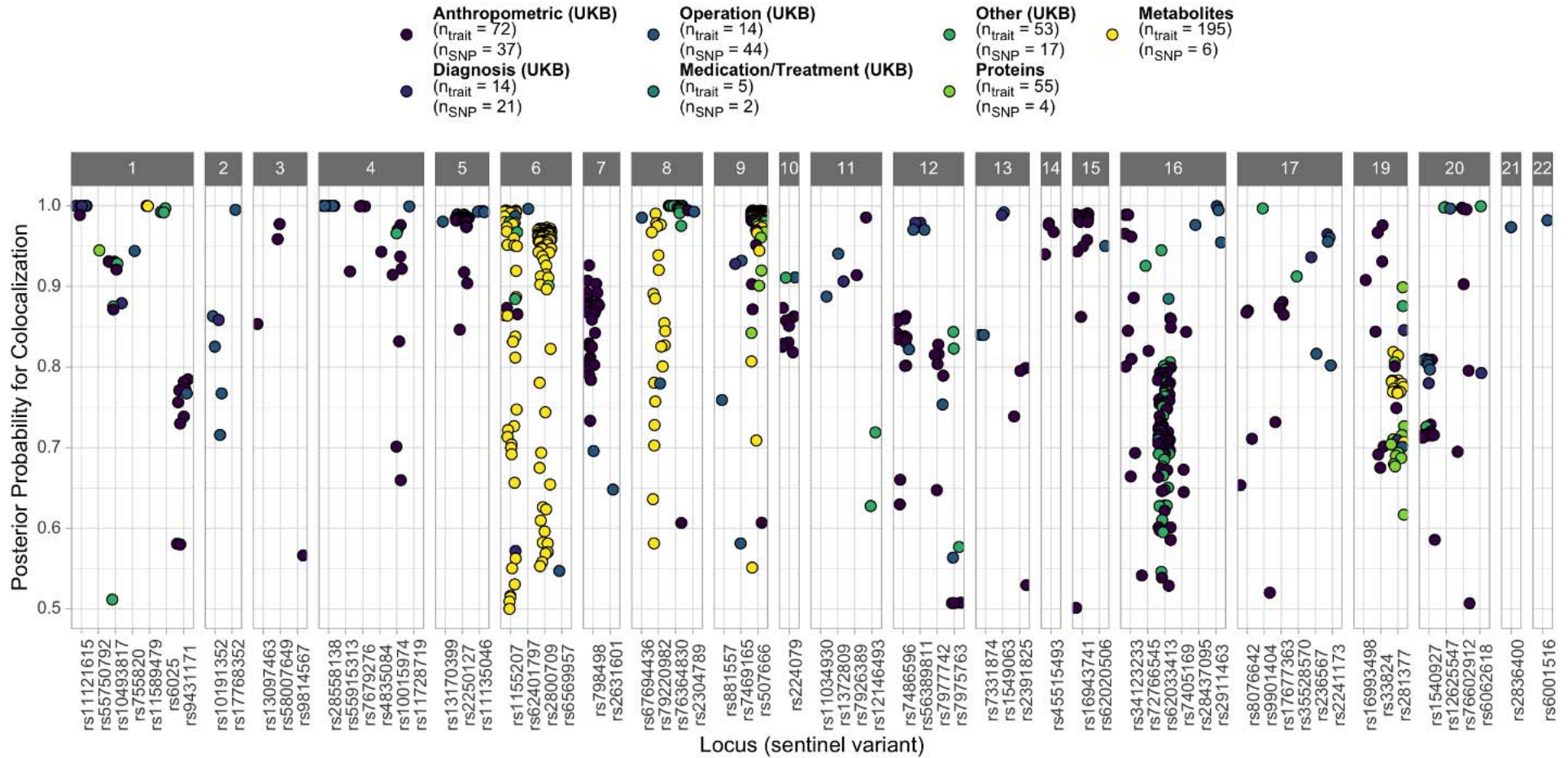
Overview of study design. A multi-ancestry genome wide association study meta-analysis is highlighted in the top portion of the figure. Downstream bioinformatic analyses were then performed to identify causal loci, characterize genetic relationships with other traits/diseases, understand novel therapeutic opportunities, and implicate biologically-relevant pathways. EUR = European-ancestry, AFR = African-ancestry, HIS = Hispanic, EAS = East-Asian ancestry, LDSC = linkage disequilibrium score regression, PheWAS = phenome-wide association study, MR = Mendelian randomization, RWR = random-walk with restart.

FIGURE 2: Phenome-Wide Associations of Varicose Veins Risk Variants



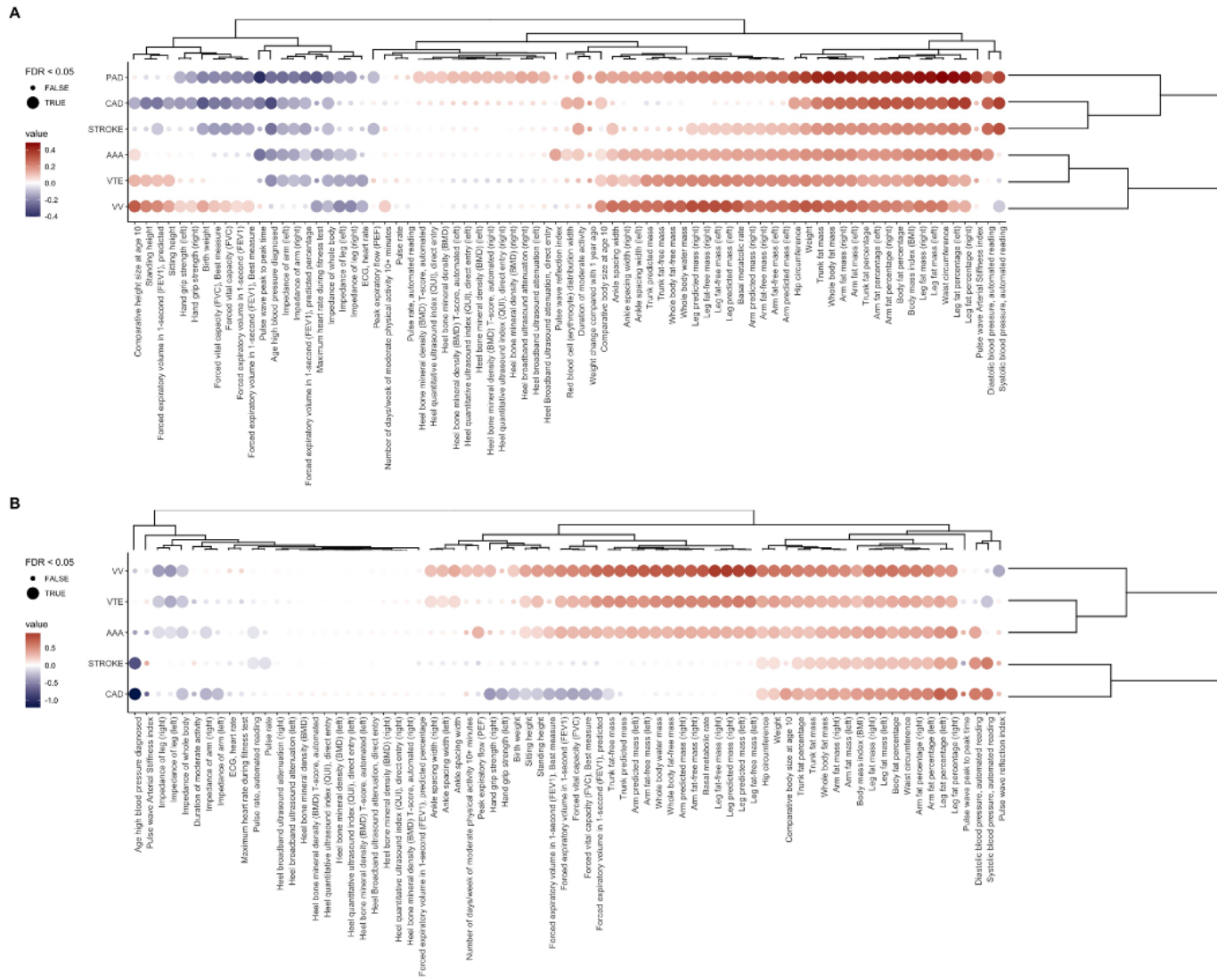
PheWAS was performed to identify associations between sentinel varicose veins-associated variants and traits with previously reported GWAS summary statistics available in the MRC IEU Open GWAS Project. Each point represents a genome-wide significant SNP-trait association ($p < 5 \times 10^{-8}$), with $-\log_{10}(p\text{-value})$ plotted as a function of the effect estimate. The top 4 associations by absolute z-score ($z = \text{effect}/\text{standard error}$) are labeled for within each category. The number of unique traits (n_{trait}) sharing risk variants with varicose veins, and the number of unique varicose veins-associated SNPs (n_{SNP}) associated with traits within each trait category are noted in the title of each facet.

FIGURE 3: Bayesian Colocalization Reveals Shared Genetic Mechanisms



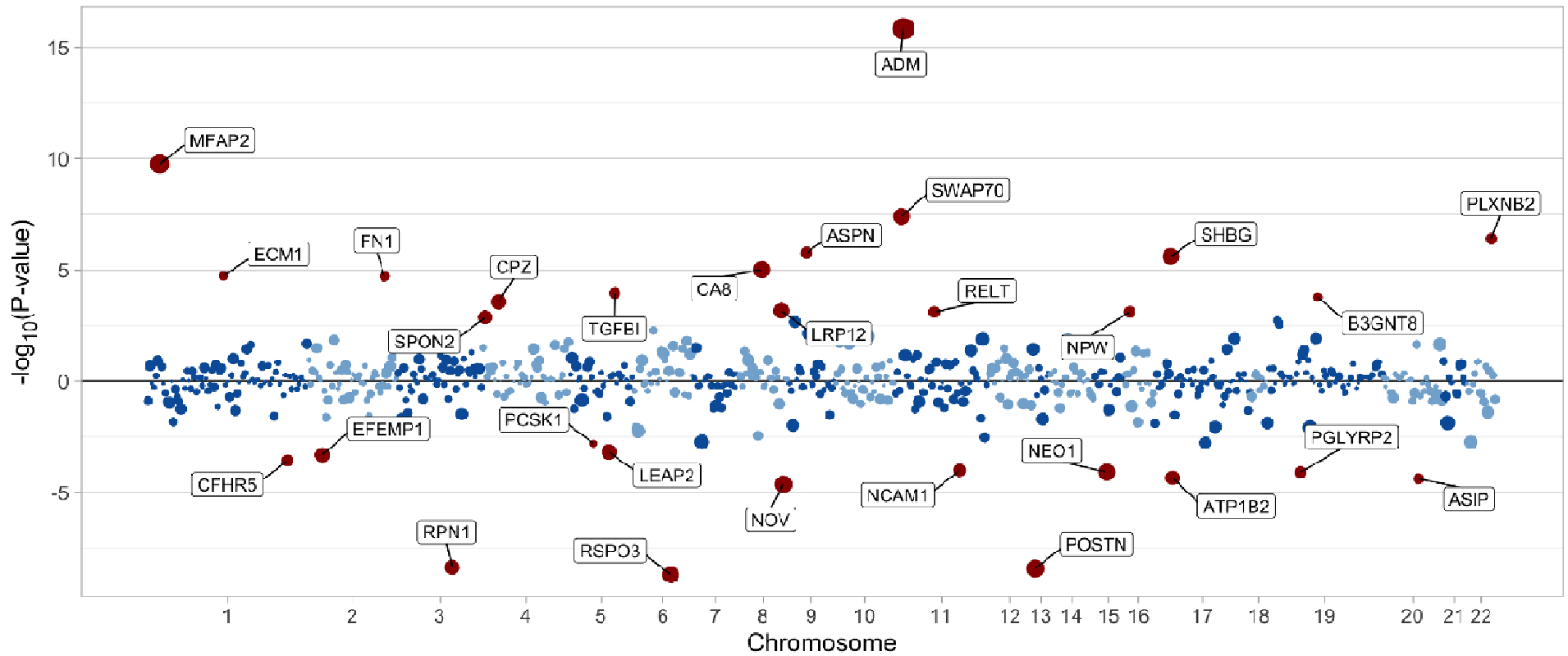
Bayesian colocalization was performed for the locus +/- 500kb surrounding each significant pheWAS SNP-trait association. Each point represents a pairwise association between varicose veins and another trait, at a given locus (x-axis).

FIGURE 4: Shared Heritability of Anthropometric and Vascular Traits



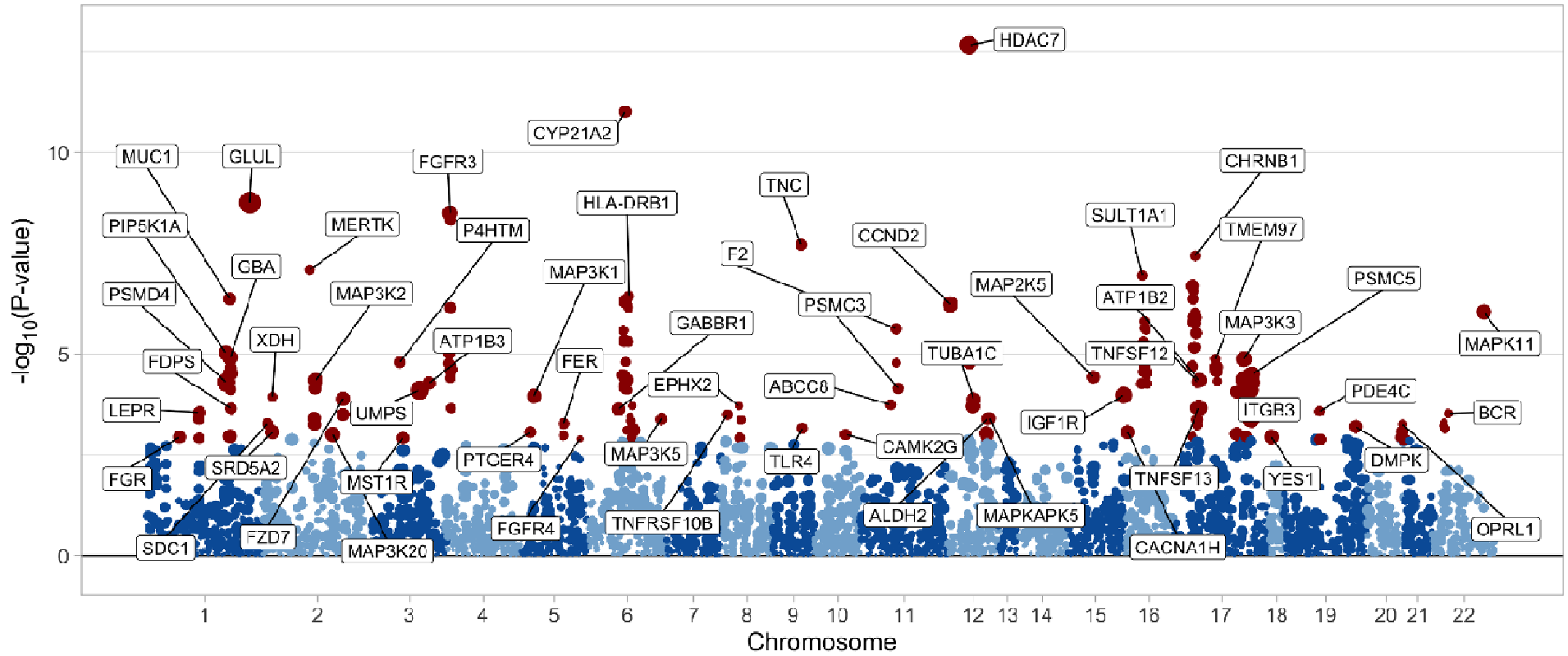
Shared genetic architecture between varicose veins and anthropometric traits. A) Cross-trait LD-score regression was performed to evaluate the shared heritability between varicose veins, other vascular diseases, and anthropometric traits. B) Mendelian randomization was performed to evaluate whether anthropometric traits represent causal risk factors for varicose veins and other vascular diseases. Large circles represent significant findings (FDR < 0.05). Points are colored based on the direction and magnitude of genetic correlation (panel A), or direction of causal effect of each risk factor on each vascular trait (red = positive effect, blue = negative effect). Points are organized by hierarchical clustering of standardized effects (genetic correlation in panel A, z-score in panel B).

FIGURE 5: Proteome-wide Mendelian Randomization



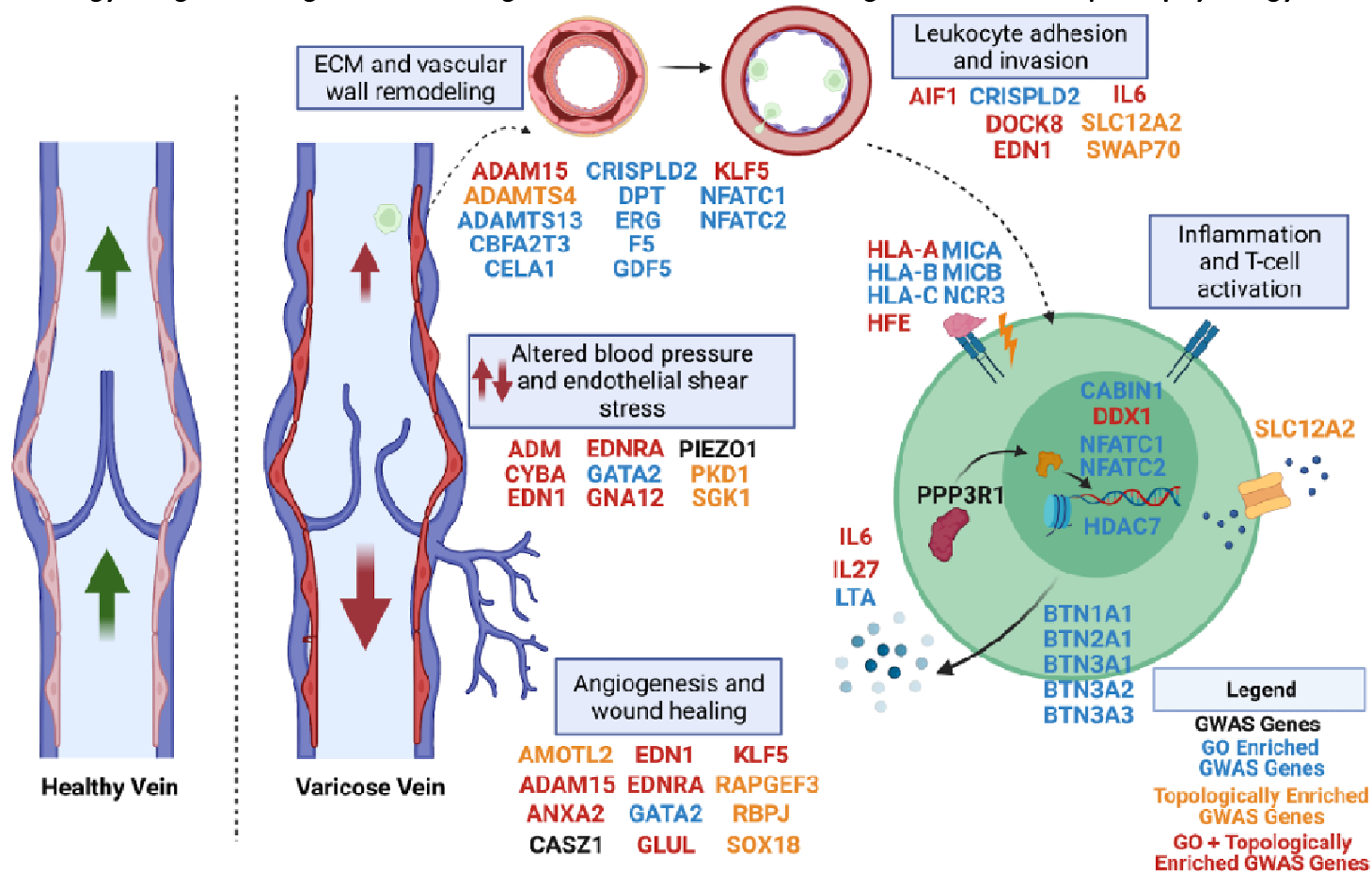
Significant (FDR $q < 0.05$) associations between circulating proteins and varicose veins are highlighted in red and labeled with gene name. Size of points represents absolute effect size (larger points correspond to greater absolute effect size). Points above the origin correspond to proteins where higher circulating levels are associated with increased risk of varicose veins, while those below the origin correspond to proteins where higher circulating levels are associated with decreased risk of varicose veins.

FIGURE 6: Drug-repurposing Mendelian Randomization



Significant (FDR $q < 0.05$) associations between gene expression levels of actionable drug targets and risk of varicose veins are highlighted in red and labeled with gene name. Size of points represents absolute effect size (larger points correspond to greater absolute effect size). Targets are positioned based on the chromosomal location of the protein-coding gene.

FIGURE 7: Systems biology integration of genome-wide significant variants contributing to varicose vein pathophysiology



Known varicose vein pathophysiological processes are implicated by GWAS genes as determined by H-MAGMA (black). GWAS genes also overlapped with traditional GO enrichment (blue), topological enrichment of GO terms (orange), or both GO enrichment and topological enrichment (red). Multiple genes were implicated in angiogenesis and dysregulated blood pressure including GATA2, endothelin-1 (EDN1) and the endothelin A receptor (EDNRA). Vascular wall and extracellular matrix (ECM) remodeling genes were also enriched including ADAM15, ADAMTS4, and ADAMTS13. Variants in leukocyte migration genes were identified, including AIF1 and DOCK8. Finally, pro-inflammatory and T-cell activation genes were also enriched (CABIN1, IL6, IL27, LTA, NFATC1, NFATC3, and PPP3R1) as well as major histocompatibility complex class 1 (MHC-I) and butyrophilin genes.