Genome-wide association analysis and replication in 810,625 individuals identifies novel therapeutic targets for varicose veins

First author's surname: Ahmed

Short title: The genetic architecture of varicose veins

Waheed-Ul-Rahman Ahmed¹, Akira Wiberg *DPhil MRCS*¹, Michael Ng *DPhil*¹, Wei Wang *PhD*², Adam Auton *DPhil*², 23andMe Research Team², Regent Lee *DPhil FRCS (Vasc)*³, Ashok Handa *FRCS (Vasc)*³, Krina T Zondervan *DPhil*^{4,5}, Dominic Furniss *DM FRCS (Plast)*^{1,*}

1. Nuffield Department of Orthopaedics, Rheumatology and Musculoskeletal Sciences, University of Oxford, Botnar Research Centre, Windmill Road, Oxford, OX3 7LD, UK.
2. 23andMe, Inc., Sunnyvale, CA, USA.
3. Nuffield Department of Surgical Sciences, University of Oxford, John Radcliffe Hospital, Oxford, OX3 9DU, UK.
4. Nuffield Department of Women's & Reproductive Health, University of Oxford, John Radcliffe Hospital, Oxford, OX3 9DU, UK.
5. Wellcome Centre for Human Genetics, University of Oxford, Old Road Campus, Roosevelt Drive, Oxford, OX3 7BN, UK.

***Correspondence:** Professor Dominic Furniss, Nuffield Department of Orthopaedics, Rheumatology and Musculoskeletal Science, University of Oxford, Botnar Research Centre, Windmill Road, Oxford, OX3 7LD, United Kingdom. Telephone: +44 1865 227233. E-mail: dominic.furniss@ndorms.ox.ac.uk

ORCID Details: WUR.A, orcid.org/0000-0003-0880-0355; A.W, orcid.org/0000-0003-4944-123X; A.A, orcid.org/0000-0002-1630-1225; R.L, orcid.org/0000-0001-8621-5165; A.H, orcid.org/0000-0001-7888-7180 K.Z, orcid.org/0000-0002-0275-9905; D.F, orcid.org/0000-0003-2780-7173.

Subject Terms: Genetic, Association Studies; Genetics; Vascular disease

Abstract

Background: Varicose veins (VVs) affect one-third of Western society, with a significant subset of patients developing venous ulceration, and ongoing management of venous leg ulcers costing around \$14.9 billion annually in the USA. There is no current medical management for VVs, with approaches limited to compression stockings, ablation techniques, or open surgery for more advanced disease. A significant proportion of patients report a positive family history, and heritability is ~17%, suggesting a strong genetic component. We aimed to identify novel therapeutic targets by improving our understanding of the aetiopathology and genetic architecture of VVs.

Methods: We performed the largest two-stage genome-wide association study of VVs in 401,656 subjects from UK Biobank, and replication in 408,969 subjects from 23andMe (total 135,514 varicose veins cases and 675,111 controls). We constructed a genetic risk score for VVs to investigate its use as a prognostic tool. Genes and pathways were prioritised using a suite of bioinformatic tools, and therapeutic targets identified using the Open Targets Platform.

Results: We discovered 49 signals at 46 susceptibility loci associated with VVs, including 29 previously unreported genetic associations (28 susceptibility loci). We demonstrated that patients with VVs requiring surgery have a higher genetic risk score than those managed non-surgically. We map 237 genes to these loci, many of which are biologically relevant and tractable to therapeutic targeting or repurposing (notably

VEGFA, *COL27A1*, *EFEMP1*, *PPP3R1* and *NFATC2*). Tissue enrichment analyses implicated vascular tissue, and several genes were enriched in biological pathways relating to extracellular matrix biology, inflammation, angiogenesis, lymphangiogenesis, vascular smooth muscle cell migration, and apoptosis.

Conclusions: Genes and pathways identified represent biologically plausible contributors to the pathobiology of VVs, identifying promising candidates for further investigation of venous biology and potential therapeutic targets. We have provided the proof-of-principle that genetic risk score correlates with disease severity, which represents a first step in personalised medicine approaches to varicose veins.

Key words: genetics; genome-wide association study; varicose veins

Introduction

Varicose veins (VVs) are a very common manifestation of chronic venous disease, affecting over 30% of the population in Western countries.¹ In the USA, chronic venous disease affects over 11 million males and 22 million females aged 40-80 years², meaning it is twice as prevalent as coronary heart disease.³ Chronic venous insufficiency leads to serious complications in 10% of cases, including lipodermatosclerosis, venous ulceration, and rarely amputation. Despite best care, 25-50% of venous leg ulcers remain unhealed after 6 months of treatment.⁴ Ongoing management of venous leg ulcers costs around \$14.9 billion annually and 4.5 million work days per year are lost to venous-related illness in the USA.^{5,6} Despite this, at present there are no medical treatments for VVs. For symptomatic patients, endovenous ablation is the first-line treatment approach.⁷ However, recurrence following surgery is 20%, with no difference in rate compared to conventional open surgery.⁸

VVs are thought to develop from a combination of valvular insufficiency, venous wall alterations, and haemodynamic changes that precipitate venous reflux, stasis, and hypertension of the venous network, causing varicosities.^{9,10} Risk factors for VVs include older age, female sex, pregnancy, a positive family history, obesity, tall height, and previous DVT.^{3,11,12} Many patients with VVs report a positive family history¹³, and amongst offspring with one affected parent the familial standardised incidence ratio is 2.39¹⁴, with a heritability of 17%¹⁵, suggesting a genetic component to aetiology. Two

recent genome-wide association studies (GWAS) of VVs have been described. Ellinghaus *et al.*¹⁶ tested for associations in 323 cases and 4,619 controls, with suggestive associations examined in an independent cohort totalling 1,946 cases and 3,146 controls. They reported two associations, mapped to *EFEMP1* and *KCNH8*. Fukaya *et al.*¹⁷ used UK Biobank to identify a further 30 putative associations (27 loci) associated with VVs. However, their cases were defined only by the International Classification of Diseases (ICD) diagnostic codes, meaning that thousands of cases defined by operative intervention codes were misclassified as controls. Moreover, the genetic associations discovered were not replicated in an independent cohort.

By undertaking the largest two-stage GWAS of VVs to date, we aimed to advance substantially our understanding of the aetiopathology and genetic architecture of VVs; discover clinically-relevant biologic pathways and prioritise targets for therapeutic development; and through genetic risk scoring pave the way for personalised medicine approaches to VV management.

Methods

Ethical approval

UK Biobank obtained ethical approval from the North West Multi-Centre Research Ethics Committee (MREC) (11/NW/0382). This study was conducted under UK Biobank study ID 10948. All 23andMe research participants included in this study provided informed consent for their genotype data to be utilised for research purposes under a protocol approved by the external AAHRPP-accredited IRB, Ethical & Independent Review Services.

Population and phenotype definition

The characteristics of the full UK Biobank cohort are described in detail elsewhere.¹⁸ In the discovery analysis, VV cases were identified from the UK Biobank data showcase using diagnostic, operative and self-report codes (Supplementary Table 1). Following quality control, we identified 22,473 VV cases and the remaining 379,183 subjects were defined as controls.

In the 23andMe replication cohort, participants answered the question *'Do you have varicose veins on your legs? (Yes/Not Sure/No)*. VV cases were identified if they answered 'Yes', whilst controls were identified by answering 'No'. 113,041 VV cases and 295,928 controls were included in the final replication analysis.

Genotyping

UK Biobank subjects were genotyped on UK BiLeve Axiom and UK Biobank Axiom arrays, with ~95% shared content. 805,426 directly genotyped variants were available prior to quality control. The 23andMe replication cohort was genotyped on one of four custom arrays (v1/v2, v3, v4, v5). Illumina HumanHap550+ BeadChip was used for v1/v2 (1,680 cases, 4,882 controls) and the Illumina OmniExpress+ BeadChip was used for v3 (21,342 cases, 56,448 controls). For v4 a fully customized array (58,883 cases, 148,637 controls) was used, and for v5, the Illumina Infinium Global Screening Array was used (31,136 cases, 85,961 controls). Successive arrays contained significant overlap with all previous arrays.

Quality control

Quality control (QC) for the UK Biobank discovery cohort was conducted using PLINK v1.9¹⁹ and R v3.3.1, as previously described (Supplementary Figure 1).²⁰ Following the QC protocol, our final discovery GWAS consisted of 401,667 subjects and 547,011 genotyped single-nucleotide polymorphisms (SNPs).

For the 23andMe replication analysis, samples were restricted to individuals from European ancestry determined through an analysis of local ancestry.²¹ A maximal set of unrelated individuals was chosen for each analysis using a segmental identity-by-descent (IBD) estimation algorithm, defining related individuals if they shared more than 700 cM IBD, including regions where the two individuals share either one or both genomic segments IBD. Cases were preferentially retained in the analysis. Variant QC was applied independently to genotyped and imputed GWAS results. The SNPs failing

QC were flagged based on multiple criteria, such as HWE P-value, call rate, imputation R-square and test statistics of batch effects.

Imputation

The phasing and imputation of UK Biobank has been previously described.²² In 23andMe, out-of-sample modified versions of the Beagle graph-based haplotype phasing algorithm²³ and Eagle2²⁴ algorithm were used to phase samples. We imputed samples against a single unified imputation reference panel combining the 1000 Genomes Phase 3 haplotypes²⁵ with the UK10K imputation reference panel²⁶ using Minimac3.²⁷

Association analysis

In the UK Biobank, we performed GWAS using a linear mixed non-in infinitesimal model implemented in BOLT-LMM v2.323²⁸ across 547,011 genotyped SNPs (minor allele frequency (MAF) ≥ 0.01) and 8,397,536 imputed SNPs (MAF ≥ 0.01 and INFO score ≥ 0.90), adjusting for genotyping platform and genetic sex. Conditional regression analysis was performed at the top signal at each of 109 associated loci in BOLT-LMM²⁸, excluding the MHC region.

In 23andMe, summary statistics were generated via logistic regression assuming an additive model for allelic effects. Association analysis was performed adjusting for age, sex, the first five principal components, and genotyping platform. The top 118 independent variants from the discovery GWAS were tested for their association with

VVs in the 23andMe cohort. The Bonferroni-corrected significance threshold for replication was set at $P < 4.24 \times 10^{-4}$ ($0.05/118$). Data for 108 of 118 variants was available; nine variants failed to meet the SNP QC within 23andMe, and one (10:79677281_CA_C) was not identified in 23andMe. A fixed-effects meta-analysis was performed in GWAMA.²⁹

Genetic risk score

Using the 49 replicated signals, we calculated a weighted genetic risk score (wGRS) for individuals in the UK Biobank cohort as previously described.³⁰ The effect alleles for each variant were used to compute SNP dosage, using QCTOOL v2. wGRS calculations and unpaired t-testing between the different subgroups was performed in R v3.3.1.

Functional annotation of SNPs

To annotate SNPs at our susceptibility loci, SNP2GENE was performed in Functional Mapping and Annotation of GWAS (FUMA) v1.3.3³¹ using summary statistics from the UK Biobank discovery cohort and default settings. FUMA defined risk loci borders by using all independent genome-wide significant SNPs ($r^2 < 0.6$), and identified all SNPs that were in linkage disequilibrium with one of these candidate SNPs. Using ANNOVAR, candidate SNPs within the replicated loci were annotated on the basis of genomic location. Exonic SNPs were investigated further using Gnomad and Ensembl genome browsers to uncover non-synonymous variants. All candidate SNPs were annotated with Combined Annotation Dependent Depletion (CADD)³², RegulomeDB³³, and 15-

core chromatin states (predicted by the Roadmap ChromHMM model³⁴) to predict any regulatory or transcription effects from chromatin states at each SNP.

Candidate gene mapping

Four gene mapping approaches – positional mapping, eQTL mapping, MAGMA gene mapping, and summary-based mendelian randomisation (SMR) – were used to map putative genes at the replicated loci. For FUMA positional mapping, all genome-wide significant SNPs at each locus were mapped to genes within a positional window of 10Kb.³¹ eQTL gene mapping was used to map genes based on having at least one genome-wide significant cis-eQTL in GTEx v8 tibial artery tissue.³¹ Using MAGMA v.1.07³⁵, we conducted a genome-wide, gene-based association study, testing 17,966 protein-coding genes. The significance threshold was set at $P < 2.78 \times 10^{-6}$ (0.05/17966).

To identify gene expression levels associated with VVs due to pleiotropy, SMR and HEIDI analyses were performed.³⁶ Summary statistics from the UK Biobank discovery GWAS were used, alongside eQTL data for GTEx V7 tibial artery. The top associated eQTL for each gene was used as an instrumental variable to examine association with VVs. A Bonferroni-corrected significance for all SMR probes was set at $P_{\text{SMR}} < 1.01 \times 10^{-5}$ (0.05/4946 probes). Subsequently, a HEIDI test was conducted across 44 SMR-significant probes to examine for heterogeneity in SMR estimates (significance set at $P < 1.12 \times 10^{-3}$ (0.05/44)).

Gene set, tissue-specific, and pathway enrichment analysis

Gene set analysis was implemented in MAGMA v1.07³⁵ across 15,496 gene sets derived from MSigDB v8.0³⁷ with a significance threshold of $P < 3.23 \times 10^{-6}$ (0.05/15496). Tissue-specific gene property analysis was also performed in MAGMA v1.07 to determine the expression of the protein-coding genes in VV-related tissue types in GTEx v8.0.³⁸

Pathway analysis was performed in eXploring Genomic Relations (XGR).³⁹ XGR ontology enrichment analysis was performed across all mapped genes, in 'canonical pathways' with the following settings: hypergeometric distribution testing, any number of genes annotated, any overlap with input genes, and an adjusted FDR < 0.05.

SNP-based heritability and genetic correlation

Linkage Disequilibrium Score (LDSC) regression⁴⁰ was used to calculate the LDSC intercept, mean chi-squared test and attenuation score, and SNP-based heritability (h^2_g).⁴¹ LDSC regression was also used to calculate the genetic correlation between VVs and 51 preselected traits, across nine trait categories: metabolites, glycaemic traits, autoimmune diseases, anthropometric traits, smoking behaviour, lipids, cardiometabolic traits, reproductive traits and haematological traits, from publicly-available GWAS data within LDHub. LDHub is a centralised database which contains summary-level GWAS data from 855 GWAS studies across 41 GWAS consortia.⁴¹ Traits were selected based on putative epidemiological associations with VVs from the literature.

Drug target enrichment analysis

The prioritised genes at our replicated loci were queried on the Open Targets Platform,⁴² assessing whether encoded proteins were tractable to small molecule or antibody targeting, or drug targets in any phase of clinical trial. Genes were also analysed for enriched drug pathways, with a nominal $P < 0.05$ threshold of significance.

Results

The overall analytic workflow is shown in Figure 1.

Association analysis

GWAS of the UK Biobank discovery cohort (22,473 cases and 379,183 controls) yielded genome-wide significant associations ($P < 5 \times 10^{-8}$) at 109 risk loci (12,391 variants; Supplementary Data 1). Conditional regression yielded a further nine independent signals at eight of the 109 loci. As expected, the large discovery sample led to the λ_{GC} showing nominal inflation (1.25). The linkage disequilibrium score (LDSC) regression intercept (1.06), and attenuation ratio of 0.13 is fully in keeping with expectations of polygenicity (Figure 2).⁴⁰

We tested the top 118 signals at the 109 risk loci in the 23andMe dataset (113,041 cases and 295,928 controls). Again, the LDSC intercept demonstrated moderate inflation ($\lambda = 1.13$, S.E. = 0.01) consistent with polygenicity and large sample size. Forty-nine of 118 variants demonstrated significant association at a conservative Bonferroni-corrected threshold of $P < 4.24 \times 10^{-4}$. Thus, we identified 49 independent significant associations at 46 risk loci (Figure 2; regional association plots for all 49 signals can be found in Supplementary Figure 2). Allelic effects were concordant across both cohorts at all 49 replicated variants, with minimal evidence of heterogeneity between the two GWAS at all loci (Q-statistic > 0.05). Eighteen loci were previously

reported, and 28 are novel (Table 1). Sixty-nine variants at 63 risk loci did not replicate (Supplementary Table 2).

Genetic risk score

As expected, we found the weighted genetic risk score (wGRS) for VV cases (5.179) was higher than in controls (4.986; $P = 9.90 \times 10^{-324}$; Table 2). We hypothesized that VV cases that had required surgery would have a higher wGRS. In keeping with this, we discovered a higher wGRS in the operation group (5.185) compared with the non-operation group (5.076; $P = 2.46 \times 10^{-13}$). These results strongly suggest that the wGRS derived from our susceptibility loci correlate with more severe VVs.

***In silico* annotation**

FUMA identified 5,315 genome-wide significant candidate SNPs from our discovery cohort associated with VVs at 45 of the 46 replicated loci (Figure 3; Supplementary Data 2). Around 2% of candidate SNPs ($n = 103$) were exonic, of which 56 were non-synonymous (52 missense, two stop gain, one splice site variant, and one frameshift variant). Of the non-synonymous variants, four missense variants were predicted to affect protein structure or function, and were in moderate linkage disequilibrium ($R^2 \geq 0.22$ and $D' \geq 0.75$) with the index SNP at three loci – 12q13.12 (index SNP: rs7308356), 16q24.3 (index SNP: rs2002833) and 17q24.3 (index SNP: rs9895127) (Supplementary Table 3). This includes rs7184427 (A/G) ($P = 9.10 \times 10^{-40}$, OR = 1.19, $R^2_{\text{index}} = 0.22$, $D'_{\text{index}} = 0.75$), which causes a predicted deleterious p.Val250Ala substitution within *PIEZO1* (SIFT: 0). *PIEZO1* encodes a mechanically active ion

channel involved in the detection of vascular shear stress, and was previously associated with VVs and lymphoedema.^{17,43} Furthermore, rs12303082 ($P = 2 \times 10^{-9}$, OR = 1.06, $R^2_{\text{index}} = 0.27$, $D'_{\text{index}} = 0.92$) resides within a highly conserved region (Genomic evolutionary rate profiling (GERP) score: 0.83) in *FAM186A*, and causes a non-conservative p.Lys187Gln amino acid substitution that is predicted to be deleterious and alter protein function (SIFT: 0.01, PolyPhen-2: 0.91).

Of the 4,294 intronic and intergenic variants identified by FUMA (Figure 3), 3,735 (87.0%) resided in open chromatin regions (Supplementary Data 2), and 163 demonstrated evidence of functionality with CADD (Combined Annotation-Dependent Depletion³²) score ≥ 12.37 , the threshold suggested for deleterious variants (Supplementary Table 4). Using RegulomeDB (RDB³³) to investigate their regulatory functions, 17 had a RDB score of at least 2b (likely to affect binding) and eight had an RDB score of at least 1f (likely to affect binding and linked to expression of a gene target) (Supplementary Table 4).

Gene mapping

Positional mapping in FUMA SNP2GENE highlighted 204 genes based on genomic position at 38 loci (Supplementary Table 5).³¹ eQTL mapping, based on GTEx v8 tibial artery tissue³⁸, mapped a total of 80 genes (Supplementary Table 5). Genome-wide, gene-based association analysis implemented in MAGMA v1.07³⁵ identified 248 protein-coding genes significantly associated with VVs ($P < 2.67 \times 10^{-6}$); 117 were within our replicated loci (Supplementary Table 6; Supplementary Figure 3).

In the summary-based mendelian randomisation (SMR) analysis³⁶, 44 genes passed the set significance threshold ($P_{\text{SMR}} < 1.01 \times 10^{-5}$). To exclude SMR associations due to linkage, we performed HEIDI analysis across 44 significant genes — 27 passed the HEIDI test ($P_{\text{HEIDI}} \geq 1.12 \times 10^{-3}$; Supplementary Table 7), 14 of which were within our VVs susceptibility loci, highlighting an association through pleiotropy rather than linkage disequilibrium and co-localisation.

In summary, 237 unique genes were mapped by at least one mapping approach. Significant overlap between the mapping strategies was seen, with the majority of genes (54.9%, $n = 130$) being prioritised by two or more mapping approaches. Thirty-six genes were prioritised by three mapping approaches, and six genes (*ATF1*, *AP1M1*, *DNAH10OS*, *FBLN7*, *LBH*, *WDR92*) were prioritised by all four approaches (Supplementary Table 5).

Gene set, tissue-specific and pathway enrichment

Following MAGMA gene set enrichment analysis, four Gene Ontology (GO) gene sets were significantly over-represented in our data: Cardiovascular Development ($P = 1.56 \times 10^{-8}$, $n = 666$); Tube Morphogenesis ($P = 9.35 \times 10^{-8}$, $n = 778$); Blood Vessel Morphogenesis ($P = 9.39 \times 10^{-7}$, $n = 555$); and Tube Development ($P = 1.68 \times 10^{-6}$, $n = 956$) (Supplementary Table 8). Further, tissue-specific gene property analysis demonstrated significant gene expression in all three vascular tissue types present in GTEx 54 tissue types: coronary artery ($P = 6.23 \times 10^{-7}$, 2nd most enriched), tibial artery (P

= 1.05×10^{-6} , 3rd most enriched) and aorta ($P = 3.92 \times 10^{-5}$, 8th most enriched; Figure 4). MAGMA analysis of GTEx 30 general tissue types also demonstrated blood vessels to be a highly enriched tissue ($P = 3.8 \times 10^{-4}$, 3rd most enriched; Figure 4). Using XGR³⁹, six canonical pathways were significantly enriched, including genes in pathways related to extracellular matrix biology, the VEGF and VEGFR signalling network, and intracellular Ca^{2+} signalling in the T-Cell Receptor (TCR) Pathway (Table 3).

Genetic correlations with other phenotypes

Using LDSC regression⁴⁰, we estimated the total SNP heritability (h^2_g) for VVs in UK Biobank to be 5.03% (S.E. = 0.30%) and in 23andMe to be 5.40% (S.E. = 0.30%). Of the nine trait categories tested for genetic correlation, two (anthropometric and autoimmune) contained traits that met our Bonferroni-corrected significance threshold ($P < 5.56 \times 10^{-3}$; Supplementary Table 9). All twelve significantly correlated traits were positively correlated with VVs (r_g range: 9- 21%; Figure 5). Eleven traits belonged to the anthropometric category and pertained to height and weight phenotypes, which are well-known, established risk factors for VVs.^{3,11,12} In the autoimmune trait category, we discovered systemic lupus erythematosus (SLE) to be correlated with VVs, sharing ~19% genetic overlap ($P = 4.2 \times 10^{-3}$, $r_g = 0.195$). Of note, the C allele of variant rs17321999 at 2p23.1 (*LBH*), which is associated with an increased risk of SLE ($P = 2.22 \times 10^{-16}$, OR = 1.20)⁴⁴, was also significantly associated with VVs in our discovery cohort ($P_{\text{disc}} = 3.20 \times 10^{-14}$, OR = 1.09), and in high linkage disequilibrium with lead SNP at 2p23.1 in the meta-analysis (rs9967884; $P_{\text{meta}} = 1.40 \times 10^{-14}$, OR = 1.11; $r^2 = 0.87$).

Drug target enrichment analysis

Of the 237 mapped genes, 200 genetic targets were identified.⁴² 42 drug pathways reached nominal significance ($P < 0.05$), with the Butyrophillin family interactions pathway being most enriched ($P = 2.6 \times 10^{-7}$, six targets), followed by the Calcineurin-NFAT pathway ($P = 9.4 \times 10^{-4}$, two targets), and the Transcriptional regulation by RUNX1 pathway, which possessed the highest number of targets ($P = 1.6 \times 10^{-3}$, 14 targets; Supplementary Table 10). Tractability information for 105 gene targets was available, with 65 of 237 genes predicted to be tractable to antibody targeting to a high confidence, and 26 genes predicted to be tractable to small molecule targeting (Supplementary Table 11). Eight of the 237 genes overlapped with pharmacologically active targets with known pharmaceutical interactions (*CDK10*, *COL27A1*, *GABBR1*, *KCNJ2*, *MAPK10*, *OPRL1*, *TNC* and *VEGFA*) (Supplementary Table 12). Of note, *VEGFA* is a target for several antibody, protein and oligosaccharide agents which are currently in phase 2, 3 and 4 clinical trials, including for several ocular vascular disorders (Supplementary Table 12).

Discussion

VVs cause significant morbidity and large healthcare costs. There is an urgent need to understand the biology of VVs in order to develop new therapeutic strategies. Our analysis represents the largest and most comprehensive GWAS to date of VVs, including 135,514 patients and 675,111 controls. We discovered 28 new risk loci (29 new signals), and independently replicated 18 of 29 previously reported but non-replicated loci (20 of 32 known signals). Our weighted genetic risk score correlates with more severe prognosis – a first step in facilitating personalised medicine approaches to management. Furthermore, our *in silico* analyses demonstrated strong evidence of functional variants in VV-associated genomic regions. Our pathway analyses establish a strong enrichment for genes expressed in the extracellular matrix, immune cell signalling and circulatory system development. We have also identified a novel genetic correlation between VVs and systemic lupus erythematosus (SLE). Several prioritised genes demonstrate the potential for pharmacological targeting, and are currently under active investigation in other diseases.

Genetic risk score

Over two million people in the USA have advanced chronic venous disease², and around 500,000 per year require invasive surgical procedures. Our genetic risk score represents a proof-of-principle, demonstrating the feasibility of enhanced prognostication in enabling the subset of VVs patients that will require surgical intervention to be defined. Preventive strategies in high risk individuals could include

prophylactic compression stocking use, or early ablation procedures to mitigate risk of venous ulceration. The benefits of early endovenous ablation in improving healing of venous leg ulcers has been demonstrated.⁴⁵ Future research may also enable the identification of those at high risk of recurrence following surgery, a significant problem in current management of VVs patients.

Biological drug targets

This study has identified several biological pathways as mediators of VV pathophysiology that may be of relevance to pharmacological targeting (Supplementary Table 13).

Extracellular matrix regulation

VVs demonstrate an increased luminal diameter and intimal hypertrophy, features closely related to disruption of the extracellular matrix (ECM).⁴⁶ Deposition of ECM in the perivascular space in VVs is also recognised – possibly a compensatory mechanism to reinforce an already weakened wall.⁴⁷ Venous dilatation and valve ring enlargement seen in VVs is thought to affect the ability of venous valves to co-apt, which in turn contributes to venous reflux and hypertension.¹⁰ Indeed, there is a noted imbalance of ECM proteins in VVs, specifically collagen and elastin - with a preponderance of collagen compared to normal vein.⁴⁸ It is therefore possible that disruption in intrinsic connective tissue components of the vein wall or valves may contribute to VVs.

Our prioritised genes significantly overlapped with canonical pathways relating to ECM components, including Collagen Type XXVII Alpha 1 Chain (*COL27A1*) and EGF-containing Fibulin-like Extracellular Matrix Protein 1 (*EFEMP1*).

rs753085 ($P = 2.17 \times 10^{-11}$, OR = 1.07) is in an intron of *COL27A1*. *COL27A1* is a fibrillar collagen in the extracellular matrices of several tissues, including blood vessels.⁴⁹ *COL27A1* expression has been demonstrated to be reduced in VV samples.⁵⁰ Moreover, our drug enrichment analysis demonstrated *COL27A1* to be a pharmacologically active target, with pharmaceutical agents currently being investigated in several clinical trials (Supplementary Table 12), demonstrating its potential candidacy as a therapeutic target for VV prevention or treatment.

In our GWAS, we replicated the previously described association between rs3791679 and VVs ($P = 1.59 \times 10^{-13}$, OR = 1.08).¹⁶ rs3791679 resides in an enhancer region of *EFEMP1*, which encodes the ECM glycoprotein fibulin-3.⁵¹ Fibulin-3 plays a key role in maintaining the integrity of elastic tissues¹⁶, and is highly expressed in vein endothelial cells. Fibulin-3 has been found to antagonise vascular development by reducing the expression of the matrix metalloproteases, MMP2 and MMP3, and increasing expression of tissue inhibitors of metalloproteases (TIMPs) in endothelial cells.⁵² The saphenofemoral junction in VVs demonstrates reduced expression of MMP2, and heightened expression of TIMP1 and MMP1 protein levels.⁵³ Alterations in expression of these enzymes may therefore result in weakness in the venous wall which may predispose patients to VVs. Our drug-enrichment analysis found fibulin-3 to be tractable

to antibody targeting to a high confidence; moreover, metformin has been demonstrated to down-regulate fibulin-3 through downregulation of MMP2.⁵⁴ Fibulin-3 therefore represents another potential pharmacological target for VVs.

Immune response

Chronic Inflammation in the venous wall has been proposed to play a role in VV aetiopathology.⁴⁶ When compared to normal veins, enhanced expression of inflammatory mediators has been observed.¹⁰ VVs contain increased mast cells, monocytes and macrophages compared to normal veins.¹⁰

We defined five inflammation-associated risk loci in our GWAS. Of particular note is rs78216177 ($P_{\text{meta}} = 5.80 \times 10^{-14}$, OR = 1.10), in an intron of *DOCK8* that has roles in both innate and adaptive immune systems. Deletion of *DOCK8* is strongly associated with Hyper-IgE syndrome, a type of primary immunodeficiency that affects multiple systems, including the vasculature.⁵⁵ Vascular abnormalities in hyper-IgE syndrome include aneurysmal changes and abnormalities in great vessels.

We also discovered a significant genetic overlap between VVs and SLE. Lead variant rs1471251 ($P = 8.33 \times 10^{-11}$, OR = 1.06) is a known eQTL of *AFF1*, which is associated with SLE.⁵⁶ Reinforcing this shared genetic basis, an associated variant at *LBH*, rs17321999 ($P_{\text{disc}} = 3.20 \times 10^{-14}$, OR = 1.09), also increases the risk of SLE in a GWAS by Morris *et al* ($P = 2.22 \times 10^{-16}$, OR = 1.20).⁴⁴

XGR analysis demonstrated enrichment of “intracellular calcium signalling in the CD4+ T-Cell Receptor (TCR) pathway” ($P = 1.9 \times 10^{-3}$, $Z = 3.5$), specifically highlighting genes *NFATC2* and *PPP3R1* which are intimately involved in this pathway. rs3787184 ($P_{\text{meta}} = 2.51 \times 10^{-36}$, OR = 1.16) is in an intron of *NFATC2*, and rs2861819 ($P_{\text{meta}} = 2.65 \times 10^{-77}$, OR = 1.20) is in an intergenic region ~19kb upstream of *PPP3R1*. *PPP3R1* encodes calcium binding B (CnB), a subunit of calcineurin, a Ca^{2+} influx-activated serine/threonine-specific phosphatase that interacts with NFAT transcription factors in the regulation of naive T-cell activation.⁵⁷ VVs are defined by clustering and infiltration of T lymphocytes⁴⁷, which are predominantly distributed close to the venous valve agger⁵⁸, a fibroelastic structure located at the base of venous valves where media meets adventitia. Therefore, we hypothesize that aberrant *PPP3R1* and *NFATC2* expression could alter calcium signalling in T-cells which may contribute to the valvular pathology seen in VVs. The Calcineurin-NFAT pathway was also the second most enriched in our drug-target enrichment analysis, and may represent a novel therapeutic avenue in VVs.

Angiogenesis

Disruption in normal angiogenic processes can lead to VVs⁵⁹, potentially because of a failure to develop properly-formed venous walls and valves, or to repair defects following vascular stress injury. MAGMA gene set analysis revealed enrichment of several gene sets relating to tube formation and morphology, with the VEGF/VEGFR signalling network also being enriched in the XGR analysis.

rs11967262 at 6p21.1 ($P_{\text{meta}} = 1.45 \times 10^{-19}$, OR = 1.09) lies in an intergenic region ~7kb upstream of vascular endothelial growth factor A (*VEGFA*). *VEGFA* is a critical regulator of angiogenesis, and is fundamental to maintaining the integrity and functionality of the vessel wall.⁶⁰ *VEGFA* is a selective endothelial mitogen, binding to its receptor VEGFR2 to induce endothelial cell proliferation, migration and differentiation. *VEGFA* and VEGFR2 expression are significantly enhanced in the wall of VVs compared to normal veins, particularly in VVs complicated by thrombophlebitis.⁶¹ Plasma levels of *VEGFA* have also been demonstrated to be significantly increased in patients with VVs.⁶² *VEGFA* also causes vasodilatation, which may decrease vessel tone, and lead to stasis and the release of oxygen free radicals, which contribute to vein wall weakness.⁶³ Additionally, *VEGFA* functions as a potent vascular permeability factor⁶⁴, which may lead to VV progression and complications.⁶³ Intriguingly, *VEGFA* also promotes inflammation via expression of intercellular and vascular cell adhesion molecules, linking angiogenesis to immune dysregulation.⁶⁵ Of note, anti-*VEGFA* agents are currently being investigated in several clinical trials for the treatment of retinal vein occlusion (Supplementary Table 12). Furthermore, rs2713575 ($P_{\text{meta}} = 1.82 \times 10^{-36}$, OR = 1.12) at 3q21.3 lies in an intronic region near *GATA2*. *GATA2* is a haematopoietic transcription factor necessary for vascular integrity that acts downstream of VEGF, and has been demonstrated to regulate VEGF-induced angiogenesis and lymphangiogenesis.⁶⁶ The VEGF axis may therefore be a promising candidate for therapeutic targeting in the treatment of VVs.

Strengths and limitations

Certain limitations of this study must be acknowledged. While the discovery GWAS in UK Biobank used a combination of hospital diagnostic codes, operation codes and self-report codes, VV cases in 23andMe were identified based on self-report alone, meaning that the phenotyping for the replication GWAS was necessarily less stringent. Moreover, rather than undertake a formal meta-analysis between the discovery and replication GWAS, we tested the association only for the 118 independent lead SNPs that were genome-wide significant in the UK Biobank discovery GWAS. Thus, sub-threshold signals in the discovery GWAS that may have reached the significance threshold in the replication GWAS, or under meta-analysis, were not identified. Moreover, not having access to the full summary statistics for the replication GWAS also meant that our *in silico* analyses were performed on the summary statistics from the discovery GWAS alone. Finally, whilst we aimed to show the proof-of-principle behind genetic risk scoring and its clinical utility in the management of VVs, an independent validation cohort would enable improvement in its predictive capabilities.

However, several strengths of the paper mitigate these limitations. We have performed the largest GWAS of VVs to date by a considerable margin, with 135,514 cases and 675,111 controls. We have stringently controlled our false positive rate by reporting only the loci that were genome-wide significant in the discovery GWAS *and* that subsequently replicated, so we can be confident in the veracity of these 49 signals. This is reflected in the plethora of biologically plausible genes, gene clusters and biological pathways that were associated with these loci. This has also highlighted several novel therapeutic avenues that are currently under investigation. Furthermore, by including

surgical codes for phenotyping in the UK Biobank discovery GWAS, we were able to identify a considerably greater number of cases than a previous GWAS that also used the UK Biobank resource but relied on ICD diagnostic codes alone (22,473 vs 9,577 cases).¹⁷ As a general principle of case ascertainment, we believe there is much to be gained by seeking out individuals who have undergone surgery for a disease: given the inevitable risk of complications, surgery is generally reserved for individuals at the more phenotypically severe end of the spectrum who have failed non-surgical treatment. We have demonstrated through our genetic risk score that the individuals who had undergone surgery for VVs were also at the *genotypically* more severe end of the spectrum. This finding lends further credence to the validity of the 49 identified loci, and opens the door to the future use of genetic risk scores in improving prognostication and guiding decision-making in the management of VVs patients.

Conclusion

In conclusion, we have described the largest GWAS to date of VVs, a highly prevalent disease with a substantial health and socioeconomic cost. We discovered 49 variants at 46 loci that predispose to VVs. Importantly, our genetic risk score was higher in VV patients requiring surgery and, more immediately, represents a first step towards better prognostication. Furthermore, we have identified new associated pathways and genes involved in extracellular matrix regulation, inflammation, vascular and lymphatic development, smooth muscle cell activity, and apoptosis. These genes and pathways are biologically plausible contributors to the pathobiology of VVs, and provide excellent candidates for further investigation of venous biology. Several genes appear tractable to pharmacological targeting (notably *VEGFA*, *COL27A1*, *EFEMP1*, *PPP3R1* and *NFATC2*), and may represent viable therapeutic targets in the future management of VV patients.

URLs

ANNOVAR, www.annovar.openbioinformatics.org/en/latest/; BOLT-LMM, www.data.broadinstitute.org/alkesgroup/BOLT-LMM/; CADD, cadd.gs.washington.edu/; ENSEMBL, www.ensembl.org/index.html; flashpca, github.com/gabraham/flashpca/; FUMA, www.fuma.ctglab.nl/; GERP, <http://mendel.stanford.edu/SidowLab/downloads/gerp/>; GTEx Portal, www.gtexportal.org/home/; GWAMA, <https://genomics.ut.ee/en/tools/gwama>; Human Genome Variation Society (HGVS), <http://varnomen.hgvs.org/>; HRC, [www.haplotype-reference-consortium.org/](http://reference-consortium.org/); LD Hub, www.ldsc.broadinstitute.org/ldhub/; LD Link, www.ldlink.nci.nih.gov/; MAGMA, www.ctg.cncr.nl/software/magma; Open Targets Platform, www.targetvalidation.org/; PLINK, www.pngu.mgh.harvard.edu/~purcell/plink/; Polyphen-2, www.genetics.bwh.harvard.edu/pph2/; QCTOOL, www.well.ox.ac.uk/~gav/qctool_v2/#overview; R, www.r-project.org/; RegulomeDB, www.regulomedb.org/; SHAPEIT3, imarchini.org/shapeit3/; SIFT, www.sift.bii.a-star.edu.sg/; UK Biobank, www.ukbiobank.ac.uk/; XGR, www.galahad.well.ox.ac.uk:3040; 1000 Genomes Project, www.1000genomes.org/; 23andMe, <https://research.23andme.com/>

Data availability

Full UK Biobank data can be accessed by direct application (see URLs). Discovery GWAS summary statistics from UK Biobank can be found in Supplementary Data 4.

Acknowledgements

The data analysed in the present study was in part provided by the UK Biobank (www.ukbiobank.ac.uk), received under UK Biobank application no. 10948. The replication cohort was provided by 23andMe, Inc. (California).

D.F, A.W and K.Z. contributed to the conception, study design and supervision of this work. W.A, A.W and M.N contributed to data analysis. A.W and M.N. contributed to QC and imputation. W.W., A.A., and the 23andMe Research Team contributed to the replication study and analysis. R.L and A.H contributed to case ascertainment. W.A, A.W, D.F prepared the first draft of the manuscript. All co-authors made substantial contributions to data acquisition, data interpretation, and revised the work critically for important intellectual content.

The following members of the 23andMe Research Team contributed to this study:

Michelle Agee, Stella Aslibekyan, Adam Auton, Robert K. Bell, Katarzyna Bryc, Sarah K. Clark, Sarah L. Elson, Kipper Fletez-Brant, Pierre Fontanillas, Nicholas A. Furlotte, Pooja M. Gandhi, Karl Heilbron, Barry Hicks, David A. Hinds, Karen E. Huber, Ethan M. Jewett, Yunxuan Jiang, Aaron Kleinman, Keng-Han Lin, Nadia K. Litterman, Marie K. Luff, Jennifer C. McCreight, Matthew H. McIntyre, Kimberly F. McManus, Joanna L. Mountain, Sahar V. Mozaffari, Priyanka Nandakumar, Elizabeth S. Noblin, Carrie A.M. Northover, Jared O'Connell, Aaron A. Petrakovitz, Steven J. Pitts, G. David Poznik, J. Fah Sathirapongsasuti, Anjali J. Shastri, Janie F. Shelton, Suyash Shringarpure, Chao Tian, Joyce Y. Tung, Robert J. Tunney, Vladimir Vacic, Xin Wang, Amir S. Zare.

Sources of Funding

W.A. is supported by the Aziz Foundation, Wolfson Foundation, Royal College Surgeons of England and Oxford NIHR Biomedical Research Centre (Musculoskeletal theme). W.A was previously supported by grants from the British Association of Plastic and Reconstructive Surgeons (BAPRAS) and Royal College Surgeons of Edinburgh to complete part of this work. A.W. is an NIHR Academic Clinical Lecturer and was previously supported by an MRC Clinical Research Training Fellowship (MR/N001524/1). M.N and D.F are supported by the Oxford NIHR Biomedical Research Centre.

Disclosures

W.W., A.A., and members of the 23andMe Research Team disclose a 'significant' relationship with 23andMe, Inc. as employees, and hold stock or stock options in 23andMe, Inc. All other authors have no relationships to disclose.

References

1. Evans CJ, Fowkes FGR, Ruckley C V., Lee AJ. Prevalence of varicose veins and chronic venous insufficiency in men and women in the general population: Edinburgh Vein Study. *J Epidemiol Community Health*. 1999;53(3):149-153. doi:10.1136/jech.53.3.149
2. Gloviczki P, Comerota AJ, Dalsing MC, et al. The care of patients with varicose veins and associated chronic venous diseases: Clinical practice guidelines of the Society for Vascular Surgery and the American Venous Forum. *J Vasc Surg*. 2011;53(5 SUPPL.):2S-48S. doi:10.1016/j.jvs.2011.01.079
3. Brand FN, Dannenberg AL, Abbott RD, Kannel WB. The epidemiology of varicose veins: The Framingham Study. *Am J Prev Med*. 1988;4(2):96-101. doi:10.1016/s0749-3797(18)31203-0
4. Singer AJ, Tassiopoulos A, Kirsner RS. Evaluation and Management of Lower-Extremity Ulcers. *N Engl J Med*. 2017;377(16):1559-1567. doi:10.1056/NEJMra1615243
5. Rice JB, Desai U, Cummings AKG, Birnbaum HG, Skornicki M, Parsons N. Burden of venous leg ulcers in the United States. *J Med Econ*. 2014;17(5):347-356. doi:10.3111/13696998.2014.903258
6. Stanley JC, Barnes RW, Ernst CB, Hertzner NR, Mannick JA, Moore WS. Vascular surgery in the United States: Workforce issues: Report of the Society for Vascular Surgery and the International Society for Cardiovascular Surgery, North American Chapter, Committee on Workforce Issues. *J Vasc Surg*. 1996;23(1):172-181. doi:10.1016/S0741-5214(05)80050-3

7. Marsden G, Perry M, Kelley K, Davies AH. Diagnosis and management of varicose veins in the legs: Summary of NICE guidance. *BMJ*. 2013;347(7918). doi:10.1136/bmj.f4279
8. O'Donnell TF, Balk EM, Dermody M, Tangney E, Iafrati MD. Recurrence of varicose veins after endovenous ablation of the great saphenous vein in randomized trials. *J Vasc Surg Venous Lymphat Disord*. 2016;4(1):97-105. doi:10.1016/j.jvsv.2014.11.004
9. Raffetto JD, Khalil RA. Mechanisms of varicose vein formation: Valve dysfunction and wall dilation. *Phlebology*. 2008;23(2):85-89. doi:10.1258/phleb.2007.007027
10. Lim CS, Davies AH. Pathogenesis of primary varicose veins. *Br J Surg*. 2009;96(11):1231-1242. doi:10.1002/bjs.6798
11. Lee AJ, Evans CJ, Allan PL, Ruckley C V, Fowkes FGR. Lifestyle factors and the risk of varicose veins: Edinburg Vein Study. *J Clin Epidemiol*. 2003;56:171-179.
12. Sisto T, Reunanen A, Laurikka J, et al. Prevalence and risk factors of varicose veins in lower extremities: Mini-Finland health survey. *Eur J Surgery, Acta Chir*. 1995;161(6):405-414.
13. Scott TE, LaMorte WW, Gorin DR MJ. Risk factors for chronic venous insufficiency: a dual case-control study. *J Vasc Surg*. 1995;22(5):622-628.
14. Zöller B, Ji J, Sundquist J, Sundquist K. Family history and risk of hospital treatment for varicose veins in Sweden. *Br J Surg*. 2012;99(7):948-953. doi:10.1002/bjs.8779
15. Fiebig A, Krusche P, Wolf A, et al. Heritability of chronic venous disease. *Hum Genet*. 2010;127(6):669-674. doi:10.1007/s00439-010-0812-9

16. Ellinghaus E, Ellinghaus D, Krusche P, et al. Genome-wide association analysis for chronic venous disease identifies EFEMP1 and KCNH8 as susceptibility loci. *Sci Rep*. 2017;7(November 2016):1-9. doi:10.1038/srep45652
17. Fukaya E et al. Clinical and Genetic Determinants of Varicose Veins. *Circulation*. 2018;1-12. doi:10.1161/CIRCULATIONAHA.118.035584
18. Sudlow C, Gallacher J, Allen N, 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*. 2015;12(3). doi:10.1371/journal.pmed.1001779
19. Purcell S, Neale B, Todd-Brown K, et al. PLINK: A tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet*. 2007;81(3):559-575. doi:10.1086/519795
20. Wiberg A, Ng M, Schmid AB, et al. A genome-wide association analysis identifies 16 novel susceptibility loci for carpal tunnel syndrome. *Nat Commun*. 2019;10(1):1-12. doi:10.1038/s41467-019-08993-6
21. Durand EY, Do CB, Mountain JL, Macpherson JM. Ancestry Composition: A Novel, Efficient Pipeline for Ancestry Deconvolution. *bioRxiv*. October 2014:010512. doi:10.1101/010512
22. Bycroft C, Freeman C, Petkova D, et al. The UK Biobank resource with deep phenotyping and genomic data. *Nature*. 2018;562(7726):203-209. doi:10.1038/s41586-018-0579-z
23. Browning SR, Browning BL. Rapid and accurate haplotype phasing and missing-data inference for whole-genome association studies by use of localized haplotype clustering. *Am J Hum Genet*. 2007;81(5):1084-1097.

doi:10.1086/521987

24. Loh PR, Palamara PF, Price AL. Fast and accurate long-range phasing in a UK Biobank cohort. *Nat Genet.* 2016;48(7):811-816. doi:10.1038/ng.3571
25. Auton A, Abecasis GR, Altshuler DM, et al. A global reference for human genetic variation. *Nature.* 2015;526(7571):68-74. doi:10.1038/nature15393
26. Walter K, Min JL, Huang J, et al. The UK10K project identifies rare variants in health and disease. *Nature.* 2015;526(7571):82-89. doi:10.1038/nature14962
27. Das S, Forer L, Schön herr S, et al. Next-generation genotype imputation service and methods. *Nat Genet.* 2016;48(10):1284-1287. doi:10.1038/ng.3656
28. Loh PR, Tucker G, Bulik-Sullivan BK, et al. Efficient Bayesian mixed-model analysis increases association power in large cohorts. *Nat Genet.* 2015;47(3):284-290. doi:10.1038/ng.3190
29. Mägi R, Morris AP. GWAMA: Software for genome-wide association meta-analysis. *BMC Bioinformatics.* 2010;11. doi:10.1186/1471-2105-11-288
30. Riesmeijer SA, Manley OWG, Ng M, et al. A Weighted Genetic Risk Score Predicts Surgical Recurrence Independent of High-Risk Clinical Features in Dupuytren's Disease. *Plast Reconstr Surg.* 2019;143(2):512-518. doi:10.1097/PRS.0000000000005208
31. Watanabe K, Taskesen E, Van Bochoven A, Posthuma D. Functional mapping and annotation of genetic associations with FUMA. *Nat Commun.* 2017;8(1):1-10. doi:10.1038/s41467-017-01261-5
32. Rentzsch P, Witten D, Cooper GM, Shendure J, Kircher M. CADD: predicting the deleteriousness of variants throughout the human genome. *Nucleic Acids Res.*

- 2018;47(D1):D886-D894. doi:10.1093/nar/gky1016
33. Boyle AP, Hong EL, Hariharan M, et al. Annotation of functional variation in personal genomes using RegulomeDB. *Genome Res.* 2012;22(9):1790-1797. doi:10.1101/gr.137323.112
34. Ernst J, Kellis M. Chromatin-state discovery and genome annotation with ChromHMM. *Nat Protoc.* 2017;12(12):2478-2492. doi:10.1038/nprot.2017.124
35. de Leeuw CA, Mooij JM, Heskes T, Posthuma D. MAGMA: Generalized Gene-Set Analysis of GWAS Data. *PLoS Comput Biol.* 2015;11(4):1-20. doi:10.1371/journal.pcbi.1004219
36. Zhu Z, Zhang F, Hu H, et al. Integration of summary data from GWAS and eQTL studies predicts complex trait gene targets. *Nat Genet.* 2016;48(5):481-487. doi:10.1038/ng.3538
37. Liberzon A, Birger C, Thorvaldsdóttir H, Ghandi M, Mesirov JP, Tamayo P. The Molecular Signatures Database Hallmark Gene Set Collection. *Cell Syst.* 2015;1(6):417-425. doi:10.1016/j.cels.2015.12.004
38. Aguet F, Brown AA, Castel SE, et al. Genetic effects on gene expression across human tissues. *Nature.* 2017;550(7675):204-213. doi:10.1038/nature24277
39. Fang H, Knezevic B, Burnham KL, Knight JC. XGR software for enhanced interpretation of genomic summary data, illustrated by application to immunological traits. *Genome Med.* 2016;8(1):1-20. doi:10.1186/s13073-016-0384-y
40. Bulik-Sullivan B, Loh PR, Finucane HK, et al. LD score regression distinguishes confounding from polygenicity in genome-wide association studies. *Nat Genet.*

- 2015;47(3):291-295. doi:10.1038/ng.3211
41. Zheng J, Erzurumluoglu AM, Elsworth BL, et al. LD Hub: A centralized database and web interface to perform LD score regression that maximizes the potential of summary level GWAS data for SNP heritability and genetic correlation analysis. *Bioinformatics*. 2017;33(2):272-279. doi:10.1093/bioinformatics/btw613
42. Carvalho-Silva D, Pierleoni A, Pignatelli M, et al. Open Targets Platform: new developments and updates two years on. *Nucleic Acids Res*. 2019;47(D1):D1056-D1065. doi:10.1093/nar/gky1133
43. Lukacs V, Mathur J, Mao R, et al. Impaired PIEZO1 function in patients with a novel autosomal recessive congenital lymphatic dysplasia. *Nat Commun*. 2015;6(1):1-7. doi:10.1038/ncomms9329
44. Morris DL, Sheng Y, Zhang Y, et al. Genome-wide association meta-analysis in Chinese and European individuals identifies ten new loci associated with systemic lupus erythematosus. *Nat Genet*. 2016;48(8):940-946. doi:10.1038/ng.3603
45. Gohel MS, Heatley F, Liu X, et al. A randomized trial of early endovenous ablation in venous ulceration. *N Engl J Med*. 2018;378(22):2105-2114. doi:10.1056/NEJMoa1801214
46. Oklu R, Habito R, Mayr M, et al. Pathogenesis of varicose veins. *J Vasc Interv Radiol*. 2012;23(1):33-39. doi:10.1016/j.jvir.2011.09.010
47. Segiet OA, Brzozowa-Zasada M, Piecuch A, Dudek D, Reichman-Warmusz E, Wojnicz R. Biomolecular mechanisms in varicose veins development. *Ann Vasc Surg*. 2015;29(2):377-384. doi:10.1016/j.avsg.2014.10.009
48. Gandhi RH, Irizarry E, Nackman GB, Halpern VJ, Mulcare RJ, Tilson MD.

- Analysis of the connective tissue matrix and proteolytic activity of primary varicose veins. *J Vasc Surg.* 1993;18(5):814-820. doi:10.1016/0741-5214(93)90336-K
49. Pace JM, Corrado M, Missero C, Byers PH. Identification, characterization and expression analysis of a new fibrillar collagen gene, COL27A1. *Matrix Biol.* 2003;22(1):3-14. doi:10.1016/S0945-053X(03)00007-6
 50. Markovic JN, Shortell CK. Genomics of varicose veins and chronic venous insufficiency. *Semin Vasc Surg.* 2013;26(1):2-13. doi:10.1053/j.semvascsurg.2013.04.003
 51. Zhang Y, Marmorstein LY. Focus on molecules: Fibulin-3 (EFEMP1). *Exp Eye Res.* 2010;90(3):374-375. doi:10.1016/j.exer.2009.09.018
 52. Albig AR, Neil JR, Schiemann WP. Fibulins 3 and 5 antagonize tumor angiogenesis in vivo. *Cancer Res.* 2006;66(5):2621-2629. doi:10.1158/0008-5472.CAN-04-4096
 53. Parra JR, Cambria RA, Hower CD, et al. Tissue inhibitor of metalloproteinase-1 is increased in the saphenofemoral junction of patients with varices in the leg. *J Vasc Surg.* 1998;28(4):669-675. doi:10.1016/S0741-5214(98)70093-X
 54. Gao L-B, Tian S, Gao H-H, Xu Y-Y. Metformin inhibits glioma cell U251 invasion by downregulation of fibulin-3. *Neuroreport.* 2013;24(10):504-508. doi:10.1097/WNR.0b013e32836277fb
 55. Engelhardt KR, McGhee S, Winkler S, et al. Large deletions and point mutations involving the dedicator of cytokinesis 8 (DOCK8) in the autosomal-recessive form of hyper-IgE syndrome. *J Allergy Clin Immunol.* 2009;124(6). doi:10.1016/j.jaci.2009.10.038

56. Okada Y, Shimane K, Kochi Y, et al. A Genome-Wide Association Study Identified AFF1 as a Susceptibility Locus for Systemic Lupus Erythematosus in Japanese. McCarthy MI, ed. *PLoS Genet.* 2012;8(1):e1002455.
doi:10.1371/journal.pgen.1002455
57. Lewis RS. Calcium Signaling Mechanisms in T Lymphocytes. *Annu Rev Immunol.* 2001;19(1):497-521. doi:10.1146/annurev.immunol.19.1.497
58. Sayer GL, Smith PDC. Immunocytochemical Characterisation of the Inflammatory Cell Infiltrate of Varicose Veins. *Eur J Vasc Endovasc Surg.* 2004;28(5):479-483.
doi:10.1016/j.ejvs.2004.07.023
59. Chang MY, Chiang PT, Chung YC, et al. Apoptosis and Angiogenesis in Varicose Veins Using Gene Expression Profiling. *Fooyin J Heal Sci.* 2009;1(2):85-91.
doi:10.1016/S1877-8607(10)60005-7
60. Ferrara N. Molecular and biological properties of vascular endothelial growth factor. *J Mol Med.* 1999;77(7):527-543. doi:10.1007/s001099900019
61. Kowalewski R, Małkowski A, Sobolewski K, Gacko M. *Vascular Endothelial Growth Factor and Its Receptors in the Varicose Vein Wall Czynnik Wzrostu Śródbłonna Naczyniowego i Jego Receptory w Ścianie Żyłaków Kończyn.* Vol 17.; 2011.
62. Howlader MH, Coleridge Smith PD. Relationship of plasma vascular endothelial growth factor to CEAP clinical stage and symptoms in patients with chronic venous disease. *Eur J Vasc Endovasc Surg.* 2004;27(1):89-93.
doi:10.1016/j.ejvs.2003.10.002
63. Hollingsworth SJ, Powell GL, Barker SGE, Cooper DG. Primary varicose veins:

- Altered transcription of VEGF and its receptors (KDR, flt-1, soluble flt-1) with sapheno-femoral junction incompetence. *Eur J Vasc Endovasc Surg*. 2004;27(3):259-268. doi:10.1016/j.ejvs.2003.12.015
64. Esser S, Wolburg K, Wolburg H, Breier G, Kurzchalia T, Risau W. Vascular endothelial growth factor induces endothelial fenestrations in vitro. *J Cell Biol*. 1998;140(4):947-959. doi:10.1083/jcb.140.4.947
 65. Kim I, Moon SO, Kim SH, Kim HJ, Koh YS, Koh GY. Vascular Endothelial Growth Factor Expression of Intercellular Adhesion Molecule 1 (ICAM-1), Vascular Cell Adhesion Molecule 1 (VCAM-1), and E-selectin through Nuclear Factor- κ B Activation in Endothelial Cells. *J Biol Chem*. 2001;276(10):7614-7620. doi:10.1074/jbc.M009705200
 66. Spinner MA, Sanchez LA, Hsu AP, et al. GATA2 deficiency: A protean disorder of hematopoiesis, lymphatics, and immunity. *Blood*. 2014;123(6):809-821. doi:10.1182/blood-2013-07-51552

Table 1. Forty-nine variants at 46 susceptibility loci significantly associated with varicose veins

SNP					Discovery GWAS in UK Biobank				Replication GWAS in 23andMe				Meta-Analysis		
Chr	SNP	Position [*]	EA [†]	NEA [‡]	EA [§]	INFO	OR [#]	P-Value	EA [§]	INFO	OR [#]	P-Value	OR [#]	P-Value	Candidate genes
1	rs11121615	10825577	C	T	0.31	0.984	1.32 (1.30-1.35)	1.40×10 ⁻¹⁵⁸	0.32	0.944	1.57 (1.48-1.67)	2.72×10 ⁻⁵⁰	1.35 (1.32-1.38)	6.95×10 ⁻²⁰¹	<i>CASZ1</i>
1	rs7518191	115603413	T	C	0.30	0.986	1.07 (1.05-1.09)	2.30×10 ⁻¹⁰	0.31	0.998	1.14 (1.08-1.21)	5.72×10 ⁻⁶	1.08 (1.06-1.10)	6.90×10 ⁻¹⁴	<i>TSPAN2</i>
1	rs17712208**	214150445	A	T	0.04	G	1.13 (1.07-1.19)	3.00×10 ⁻⁶	0.03	0.687	1.50 (1.25-1.81)	1.71×10 ⁻⁵	1.15 (1.10-1.21)	1.68×10 ⁻⁸	<i>PROX1</i>
1	rs340875	214158986	G	C	0.48	0.993	1.07 (1.05-1.09)	2.30×10 ⁻¹¹	0.50	0.993	1.27 (1.20-1.34)	3.22×10 ⁻¹⁸	1.09 (1.07-1.11)	4.22×10 ⁻²⁰	<i>PROX1</i>
1	rs2820464	219693220	A	G	0.34	0.997	1.07 (1.04-1.09)	3.60×10 ⁻¹⁰	0.32	0.997	1.16 (1.09-1.22)	6.91×10 ⁻⁷	1.07 (1.05-1.10)	4.48×10 ⁻¹⁴	-
2	rs9967884	30488183	A	G	0.77	0.995	1.10 (1.07-1.12)	1.60×10 ⁻¹⁵	0.78	0.986	1.20 (1.12-1.28)	4.99×10 ⁻⁸	1.11 (1.08-1.13)	1.40×10 ⁻²⁰	<i>LBH</i>
2	rs3791679	56096892	A	G	0.77	G	1.07 (1.04-1.09)	1.90×10 ⁻⁸	0.76	G	1.22 (1.15-1.30)	4.42×10 ⁻¹⁰	1.08 (1.06-1.11)	1.59×10 ⁻¹³	<i>EFEMP1</i>
2	rs2861819	68489221	C	G	0.66	0.996	1.17 (1.15-1.20)	2.10×10 ⁻⁵⁶	0.65	0.925	1.40 (1.32-1.49)	1.04×10 ⁻²⁹	1.20 (1.17-1.22)	2.65×10 ⁻⁷⁷	<i>PPP3R1</i>
2	rs4849044	112898933	C	T	0.51	0.991	1.07 (1.05-1.09)	2.00×10 ⁻¹¹	0.50	0.990	1.12 (1.06-1.18)	5.78×10 ⁻⁵	1.07 (1.05-1.09)	1.98×10 ⁻¹⁴	<i>FBLN7</i>
2	rs17819430	118886398	C	A	0.96	0.991	1.15 (1.10-1.21)	5.00×10 ⁻⁹	0.96	0.979	1.40 (1.22-1.61)	1.63×10 ⁻⁶	1.18 (1.13-1.23)	1.51×10 ⁻¹²	<i>INSIG2</i>
2	rs55889669	173194747	A	G	0.19	0.994	1.08 (1.05-1.11)	3.60×10 ⁻¹⁰	0.19	0.977	1.20 (1.12-1.29)	2.62×10 ⁻⁷	1.09 (1.07-1.11)	2.89×10 ⁻¹⁴	-

					1.11)				1.29)				1.12)		
3	rs844176	14827080	G	A	0.89	0.974	1.14	1.60×10⁻¹⁰	0.89	0.966	1.29)	1.81×10⁻⁴	1.11 (1.08-1.15)	3.39×10⁻¹³	FGD5
3	rs2713575	128294355	G	A	0.50	0.980	1.11 (1.09-1.13)	7.30×10 ⁻²⁸	0.50	0.999	1.21 (1.15-1.28)	5.02×10 ⁻¹²	1.12 (1.10-1.14)	1.82×10 ⁻³⁶	GATA2
3	rs9877579	188058716	C	T	0.26	0.986	1.07 (1.05-1.10)	7.90×10⁻¹¹	0.26	0.983	1.14 (1.07-1.21)	3.21×10⁻⁵	1.08 (1.06-1.10)	5.84×10⁻¹⁴	LPP
4	rs28558138	26818080	G	C	0.58	0.979	1.14 (1.12-1.16)	2.00×10 ⁻⁴⁰	0.58	0.890	1.21 (1.14-1.28)	8.71×10 ⁻¹¹	1.15 (1.13-1.17)	9.35×10 ⁻⁴⁹	TBC1D19
4	rs56155140	57824451	A	G	0.19	G	1.09 (1.07-1.12)	6.50×10⁻¹³	0.19	0.996	1.21 (1.13-1.29)	5.17×10⁻⁸	1.11 (1.08-1.13)	8.30×10⁻¹⁸	IGFBP7
4	rs1471251	87976359	A	T	0.60	0.993	1.06 (1.04-1.08)	4.50×10⁻⁸	0.61	0.987	1.12 (1.06-1.18)	5.31×10⁻⁵	1.06 (1.04-1.08)	8.33×10⁻¹¹	AFFI
4	rs34154818	89726823	A	AT	0.55	0.980	1.05 (1.03-1.08)	4.90×10⁻⁸	0.53	0.951	1.14 (1.08-1.21)	2.25×10⁻⁶	1.06 (1.04-1.08)	2.13×10⁻¹¹	FAM13A
4	rs10007409	120142306	C	T	0.69	0.999	1.06 (1.04-1.08)	3.20×10⁻⁸	0.68	0.990	1.21(1.14-1.28)	1.12×10⁻¹⁰	1.07 (1.05-1.10)	2.09×10⁻¹³	USP53
4	rs11728719	186696172	A	C	0.76	0.978	1.08 (1.05-1.10)	3.00×10⁻¹¹	0.77	0.958	1.15 (1.08-1.23)	1.63×10⁻⁵	1.09 (1.06-1.11)	1.65×10⁻¹⁴	SORBS2
5	rs57253948	38754162	G	A	0.06	G	1.12 (1.07-1.16)	3.70×10⁻⁸	0.06	0.989	1.32 (1.18-1.47)	1.15×10⁻⁶	1.14 (1.09-1.18)	1.05×10⁻¹¹	-
5	rs3749748	127350549	T	C	0.25	0.992	1.16 (1.13-1.18)	5.60×10 ⁻³⁹	0.24	0.964	1.42 (1.33-1.51)	3.69×10 ⁻²⁷	1.18 (1.16-1.21)	1.06×10 ⁻⁵⁶	FBN2, SLC12A2
5	rs11135046	158230013	G	T	0.46	0.995	1.12 (1.10-1.14)	1.60×10 ⁻³²	0.45	0.992	1.16 (1.10-1.23)	5.81×10 ⁻⁸	1.13 (1.11-1.15)	1.33×10 ⁻³⁸	EBF1

6	rs7773004	26267755	A	G	0.51	0.998	1.09 (1.07-1.11)	6.00×10 ⁻²⁰	0.50	0.984	1.18 (1.12-1.25)	1.38×10 ⁻⁹	1.10 (1.08-1.12)	2.33×10 ⁻²⁶	<i>HFE</i>
6	rs11967262	43760327	C	G	0.51	0.996	1.06 (1.04-1.08)	5.20×10⁻⁹	0.51	0.990	1.34 (1.27-1.42)	8.13×10⁻²⁷	1.09 (1.07-1.11)	1.45×10⁻¹⁹	<i>VEGFA</i>
6	rs1936800	127436064	C	T	0.47	G	1.06 (1.04-1.08)	2.60×10 ⁻⁹	0.49	0.986	1.20 (1.13-1.26)	7.26×10 ⁻¹¹	1.07 (1.05-1.09)	8.29×10 ⁻¹⁵	<i>RSPO3</i>
8	rs34022079	6648676	C	T	0.64	0.975	1.07 (1.05-1.09)	8.20×10⁻¹¹	0.63	0.888	1.62 (1.52-1.71)	4.59×10⁻⁵⁸	1.11 (1.09-1.14)	1.61×10⁻²⁹	-
8	rs10504825	87567848	C	A	0.41	G	1.07 (1.05-1.09)	7.20×10 ⁻¹³	0.41	0.997	1.14 (1.08-1.21)	1.55×10 ⁻⁶	1.08 (1.06-1.10)	6.51×10 ⁻¹⁷	<i>CPNE3</i>
9	rs78216177	232148	C	G	0.14	0.992	1.10 (1.07-1.13)	3.70×10⁻¹¹	0.14	0.988	1.16 (1.08-1.25)	1.28×10⁻⁴	1.10 (1.08-1.13)	5.80×10⁻¹⁴	<i>DOCK8</i>
9	rs753085	117045447	G	A	0.73	0.994	1.06 (1.04-1.09)	1.50×10⁻⁸	0.73	0.995	1.14 (1.07-1.21)	4.11×10⁻⁵	1.07 (1.05-1.09)	2.17×10⁻¹¹	<i>COL27A1</i>
9	rs10817762	118161597	A	C	0.52	0.994	1.08 (1.06-1.10)	1.80×10 ⁻¹⁶	0.52	0.994	1.14 (1.08-1.20)	2.05×10 ⁻⁶	1.09 (1.07-1.11)	1.02×10 ⁻²⁰	<i>TNC</i>
10	rs61863928	64449549	G	T	0.62	0.955	1.06 (1.04-1.08)	3.00×10⁻⁸	0.64	0.887	1.37 (1.29-1.45)	4.22×10⁻²⁵	1.09 (1.07-1.11)	1.51×10⁻¹⁷	-
11	rs79465012	128258136	C	T	0.93	G	1.13 (1.09-1.17)	9.70×10⁻¹¹	0.93	0.842	1.28 (1.14-1.44)	2.46×10⁻⁵	1.14 (1.10-1.18)	1.08×10⁻¹³	-
12	rs7308356	50539611	G	A	0.63	0.997	1.08 (1.06-1.11)	4.10×10⁻¹⁶	0.62	0.997	1.14 (1.08-1.21)	2.93×10⁻⁶	1.09 (1.07-1.11)	3.02×10⁻²⁰	<i>CERS5</i>
12	rs1054852	124496316	G	A	0.38	0.904	1.06 (1.04-1.08)	1.60×10⁻⁸	0.37	0.927	1.29 (1.21-1.36)	1.44×10⁻¹⁷	1.08 (1.06-1.11)	2.87×10⁻¹⁶	<i>DNAH10OS</i>
13	rs41286076	73634859	T	C	0.26	0.998	1.08 (1.05-1.10)	1.10×10⁻¹¹	0.25	0.977	1.14 (1.07-1.21)	5.64×10⁻⁵	1.08 (1.06-1.10)	1.06×10⁻¹⁴	<i>KLF5</i>

					1.10)				1.21)				1.11)		
					1.22 (1.14-				2.38 (1.90-				1.28 (1.20-		
14	rs72683923	50735947	C	T	0.02	G	1.30)	1.40×10 ⁻⁸	0.02	0.721	2.99)	1.06×10 ⁻¹³	1.37)	3.62×10 ⁻¹⁴	CDKL1
15	rs11852492	96167544	T	C	0.84	0.999	1.12 (1.09- 1.15)	8.90×10 ⁻¹⁸	0.83	0.994	1.20 (1.11- 1.29)	1.49×10 ⁻⁶	1.13 (1.10- 1.15)	3.30×10 ⁻²²	-
16	rs11076178	57146402	T	C	0.11	0.986	1.09 (1.06- 1.12)	2.10×10 ⁻⁸	0.11	0.905	1.18 (1.08- 1.28)	3.16×10 ⁻⁴	1.10 (1.07- 1.13)	1.05×10 ⁻¹⁰	CPNE2
16	rs111350029**	88796770	G	GGGAG GC	0.14	0.910	1.24 (1.20- 1.27)	1.90×10 ⁻⁴⁸	0.14	0.810	1.57 (1.44- 1.71)	7.82×10 ⁻²⁵	1.27 (1.23- 1.30)	7.39×10 ⁻⁶⁶	PIEZO1, GALNS
16	rs11646394**	88812279	C	A	0.87	0.995	1.19 (1.16- 1.23)	2.30×10 ⁻³⁴	0.88	0.924	1.47 (1.35- 1.61)	5.00×10 ⁻¹⁸	1.22 (1.18- 1.25)	3.62×10 ⁻⁴⁶	PIEZO1, GALNS
16	rs2002833	88842117	G	A	0.33	0.988	1.19 (1.17- 1.22)	1.10×10 ⁻⁶⁵	0.31	0.988	1.43 (1.35- 1.52)	8.55×10 ⁻³⁴	1.22 (1.19- 1.24)	2.47×10 ⁻⁹⁰	PIEZO1, GALNS
17	rs6503321	2096580	A	G	0.38	0.999	1.06 (1.04- 1.08)	3.80×10 ⁻⁸	0.37	0.990	1.14 (1.08- 1.21)	3.28×10 ⁻⁶	1.07 (1.05- 1.08)	1.77×10 ⁻¹¹	SMG6
17	rs638538	68216128	A	C	0.27	0.979	1.11 (1.09- 1.14)	6.90×10 ⁻²³	0.28	0.994	1.19 (1.12- 1.26)	1.79×10 ⁻⁸	1.12 (1.10- 1.14)	5.85×10 ⁻²⁹	KCNJ2
17	rs9895127	70029808	T	C	0.43	0.987	1.10 (1.08- 1.12)	6.40×10 ⁻²²	0.44	0.989	1.16 (1.10- 1.22)	1.48×10 ⁻⁷	1.11 (1.09- 1.13)	2.88×10 ⁻²⁷	AC007461.1
19	rs12609241**†	16360926	G	A	0.75	0.990	1.07 (1.05- 1.10)	3.00×10 ⁻¹⁰	0.75	0.950	1.33 (1.25- 1.42)	1.84×10 ⁻¹⁸	1.10 (1.08- 1.12)	1.32×10 ⁻¹⁸	KLF2
20	rs3787184	50157837	A	G	0.83	0.978	1.16 (1.13- 1.19)	3.10×10 ⁻³²	0.82	0.949	1.17 (1.09- 1.26)	1.32×10 ⁻⁵	1.16 (1.14- 1.19)	2.51×10 ⁻³⁶	NFATC2
20	rs76602912	57459868	T	C	0.98	G	1.25 (1.18- 1.33)	7.50×10 ⁻¹³	0.97	0.857	1.52 (1.26- 1.83)	1.41×10 ⁻⁵	1.28 (1.20- 1.36)	3.56×10 ⁻¹⁶	GNAS

20	rs6062619	62683002	A	G	1.10 (1.08-				1.27 (1.19-				1.11 (1.09-		
					0.73	0.952	1.12)	1.90×10^{-17}	0.73	0.824	1.36)	4.97×10^{-13}	1.14)	5.48×10^{-25}	<i>SOX18</i>

*Based on NCBI Genome Build 37 (hg19). [†]The effect allele. [‡]The alternate (non-effect) allele. [§]The effect allele frequency in the study population.

^{||}The imputation quality score; G= genotyped SNP. [#]Odds ratio (95% confidence intervals). OR > 1 indicative of increased risk with effect allele.

denotes four residual significant signals following conditional regression analysis at the lead SNP. ^{*†}At this locus, 19p13.11, the lead SNP in the UK Biobank cohort was rs451367 ($P_{\text{discovery}} = 2.10 \times 10^{-10}$), however, this did not replicate in 23andMe ($P_{\text{replication}} = 0.47$; Supplementary Table 1) - the independent residual signal, rs12609241 shown here, did however replicate. Bold variants represent loci not previously reported.

Table 2. Weighted genetic risk score for varicose veins in the UK Biobank cohort

Group	VV		P-Value [†]	VV cases with operation code	VV cases without operation code	P-Value [§]
	cases	Controls				
N	22,473	379,183		21,407	1,066	
Mean wGRS*	5.179	4.986	9.90×10^{-324}	5.185 (0.453)	5.076 (0.468)	2.46×10^{-13}
(standard deviation)	(0.454)	(0.451)				

*wGRS: weighted genetic risk score. [†]Unpaired two-tailed t-test between varicose veins cases and controls. [§]Unpaired two-tailed t-test between varicose veins cases with an operation code and varicose veins cases without an operation code.

Table 3. Gene-based enrichment analysis.

Biological Process	Z-Score	P-Value	FDR	Number of overlapped genes	Genes
Alpha9 beta1 integrin signalling events	3.85	0.0012	0.032	2	<i>TNC, VEGFA</i>
Genes encoding structural ECM glycoproteins	3.39	0.0012	0.032	6	<i>EFEMP1, FBLN7, FBN2, IGFBP7, RSPO3, TNC</i>
Calcium signalling in the CD4+ TCR pathway	3.5	0.0019	0.032	2	<i>NFATC2, PPP3R1</i>
Ensemble of genes encoding core extracellular matrix including ECM glycoproteins, collagens and proteoglycans	3.11	0.0019	0.032	7	<i>COL27A1, EFEMP1, FBLN7, FBN2, IGFBP7, RSPO3, TNC</i>
Non-canonical WNT signalling pathway	3.29	0.0025	0.035	2	<i>MAPK10, NFATC2</i>
VEGF and VEGFR signalling network	3.11	0.0032	0.036	1	<i>VEGFA</i>

Figure 1. Study design and analysis workflow. A discovery GWAS was performed in the UK Biobank cohort, following which, the top independent lead variants were tested within the 23andMe replication cohort. Of the 118 tested variants, data on 117 variants was available for replication in the 23andMe Cohort, of which 108 passed quality control within the replication cohort (see Methods). In total, 49 independent variants at 46 loci met the Bonferroni-corrected threshold in the replication cohort (presented in Table 1), and were interrogated further in subsequent analyses.

Figure 2. Results of genome-wide association study in varicose veins. A) QQ plot of observed vs. expected P-values. B) Manhattan plot showing genome-wide P-values plotted against position on each of the autosomes. The dark blue, light blue, and green dots refer to the discovery UK Biobank Cohort, with the red dots corresponding to the 49 variants from the 23andMe cohort at each replicated locus (shown in Table 1). The dark blue peaks correspond to the 46 loci that replicated in the 23andMe cohort at a Bonferroni-corrected threshold of $P < 4.24 \times 10^{-4}$. Candidate genes at each locus are named above each signal, with newly discovered genetic loci in blue, and previously described loci in black.

Figure 3. Functional annotation of the 5,315 genome-wide significant SNPs at our 46 replicated loci. Functional consequences of the SNPs on genes were obtained by performing ANNOVAR gene-based annotation using Ensembl genes (build 85) in FUMA. A) CADD scores, B) RegulomeDB scores and C) 15-core chromatin state were annotated to all

5,315 SNPs in 1000G phase 3 by FUMA through matching chromosome, position, reference, and alternative alleles. D) the location of the 5,315 SNPs are shown.

A

B

Figure 4. MAGMA tissue expression analysis. MAGMA Tissue Expression Analysis of varicose veins GWAS-summary data, implemented in FUMA in A) 54 specific and B) 30 general tissue types. This analysis tests the relationship between highly-expressed genes in a specific tissue and the genetic associations from the GWAS. Gene-property analysis is performed using average expression of genes per tissue type as a gene covariate. Gene expression values are log₂ transformed average RPKM (Read Per Kilobase Per Million) per tissue type after winsorization at 50, and are based on GTEx v8 RNA-Seq data across 54 specific tissue types and 30 general tissue types. The dotted line indicates the Bonferroni-corrected α level, and the tissues that meet this significance threshold are highlighted in red.

Figure 5. Genetic correlation between varicose veins and other traits and diseases. Genetic correlation (r_g) between varicose veins and publicly-available phenotypes in LD Hub, using LD Score regression. All twelve traits have met a Bonferroni-corrected significance $P < 5.56 \times 10^{-3}$ and are positively correlated with varicose veins. The consortium and ethnicity for each study is provided in parentheses. EGG, Early Growth Genetics Consortium; GIANT, Genetic

Investigation of ANthropometric Traits Consortium ; Alkes, Alkes Group (Harvard T.H Chan School of Public Health); NA, Not Applicable. Eur, European ethnicity; Mi, Mixed ethnicity. Further details regarding the studies are provided in Supplementary Table 9.

Supplemental Materials (Online Only)

Supplementary Tables 1-13

Supplementary Figures 1-3

Supplementary Data 1-4

References 51-66

1. Supplementary Tables

Supplementary Table 1. Codes used for varicose veins case definition in UK Biobank. The total number of individuals with each of the codes is shown below. A total of 27,165 individuals possessed at least one of the diagnostic codes for varicose veins.

Supplementary Table 2. Additional variants associated with varicose veins in discovery cohort. 69 independent variants at 63 of the 109 loci were genome-wide significant ($P < 5 \times 10^{-8}$) in the UK Biobank cohort, however did not meet the Bonferroni-corrected threshold of $P < 4.24 \times 10^{-4}$ in the 23andMe replication cohort or replication data was not available (for ten tested variants). 59 of the 63 loci have not been previously reported. Using the four mapping strategies implemented in this study (see Methods), 127 genes were mapped to 51 of the 63 loci.

Supplementary Table 3. Varicose veins associated exonic variants at the replicated loci. 103 genome-wide significant exonic SNPs at seventeen of the replicated varicose veins susceptibility loci were identified by FUMA SNP2GENE.

Supplementary Table 4. Predicted functional intronic and intergenic variants at the replicated loci. 163 genome-wide significant intronic and intergenic variants predicted to be deleterious according to a CADD ≥ 12.37 , as identified by FUMA SNP2GENE.

Supplementary Table 5. Genes mapped to the varicose veins associated loci using the four mapping strategies. 237 unique genes were mapped to 39 of 46 replicated loci by one or more gene mapping strategies (see Methods). 204 genes were mapped via positional mapping in FUMA, 80 genes were mapped via eQTL mapping in FUMA, 117 genes were mapped using MAGMA and 14 genes were mapped using summary-based mendelian randomisation. In total, 61 unique genes were mapped to novel loci. Overlap between the four different mapping strategies is shown.

Supplementary Table 6. Genome-wide gene-based association analysis in MAGMA. 248 protein-coding genes met the threshold for genome-wide significance ($P < 2.67 \times 10^{-6}$, $0.05/18,733$) in this analysis. 117 of the 248 genes lay within our replicated loci and are highlighted in red.

Supplementary Table 7. Summary-based Mendelian Randomisation (SMR) using eQTL data from GTEx v7 tibial artery. The twenty-seven probes (genes) that met the Bonferroni-corrected significance threshold $P_{\text{SMR}} < 1.01 \times 10^{-5}$

(0.05/4,946) and passed the HEIDI test ($P_{\text{HEIDI}} \geq 1.12 \times 10^{-3}$ (0.05/44)) are shown. Fourteen probes mapped to our replicated loci.

Supplementary Table 8. Enriched gene sets from genome-wide gene-based enrichment analysis in MAGMA v1.07.

The convergence of 15,496 gene sets (15,381 from MSigDB v7.0) were tested (See Supplementary Data 3 for all tested gene sets). A Bonferroni-corrected threshold of $P < 3.23 \times 10^{-6}$ (0.05/15,496) was set, resulting in four significant Gene Ontology (GO) gene sets and two curated gene sets. This analysis was performed using the SNP2GENE tool in FUMA.

Supplementary Table 9. Genetic correlation between varicose veins and other phenotypes. This analysis was performed using LD score (LDSC) regression, implemented in LD Hub. The traits are shown along with the consortia name, sample size, ethnicity, and PMID of the study from which the LDSC data were derived, the trait category, and the correlation coefficient (r_g). Traits are ranked by P-value, and the twelve traits meeting a Bonferroni-corrected significant threshold of $P < 5.56 \times 10^{-3}$ are shown.

Supplementary Table 10. Enriched drug pathways from the drug target enrichment analysis. Mapped genes were interrogated with the Open Targets Platform to enrich drug pathways relating to the identified gene targets. The 200 gene

targets identified by the Open Targets Platform mapped to 622 drug pathways, of which, 42 reached a nominal significance $P < 0.05$ and are shown below.

Supplementary Table 11. Tractability information for targets in the drug-target enrichment analysis. The 237 mapped genes were interrogated within the Open Targets Platform to determine their tractability to small molecule and antibody targeting. 200 genes targets were identified by the Open Targets Platform of which, tractability information was available for 105 gene targets (left column).

Supplementary Table 12. Pharmacologically-active targets identified in drug-target enrichment Of the 200 varicose-veins associated genes targets identified by the Open Targets Platform, eight gene targets have known pharmaceutical interactions and are presently, or in the past have been, investigated in clinical trials in different phases for the treatment of several diseases. Diseases shown below are those relating to vascular disorders.

Supplementary Table 13. Functional categories of the gene clusters. The varicose veins susceptibility loci map to genes implicated in five functional categories. Several genes map to more than one category. Genes not associated with varicose veins previously are highlighted in bold.

2. Supplementary Figures

Supplementary Figure 1. Overview of Quality Control (QC). **a**, Flowchart summarising QC protocol. Excluded SNPs are in blue panels on the left and excluded individuals are in green panels on the right. ¹Pre-QC exclusions: 3 individuals with invalid IDs and sex, and 8 individuals who have withdrawn from UK Biobank were excluded prior to QC. ²Pre-association exclusions: 11 individuals who were not present in UK Biobank's sample file accompanying the BGEN files were excluded prior to association. **b**, Principal Component Analysis (PCA) for demonstration of ethnicity of UK Biobank individuals. The UK Biobank cohort was merged with publicly available data from the 1000 Genomes Project and PCA was performed using flashpca. Individuals identified by UK Biobank as having white British ancestry are coloured in lime green, and the remaining UK Biobank individuals are in grey. In this graph of principal component 1 vs principal component 2, a near-perfect overlap can be seen between the UK Biobank "white British" individuals and both GBR (British in England and Scotland - light brown) and CEU (Utah residents with Northern and Western European ancestry - magenta) individuals from the 1000 Genomes Project.

Supplementary Figure 2. Regional Locus Zoom plots of all varicose veins associated loci. LocusZoom plots of the 49 independent genome-wide significant SNPs at the 46 replicated varicose veins associated susceptibility loci. Plots are ordered by chromosome number and genomic position. SNP position is shown on the x-axis, and strength of association

on the y-axis ($-\log_{10}$ P-value). The linkage disequilibrium (LD) relationship between the lead SNP and the surrounding SNPs is indicated by the r^2 legend. In the lower panel of each sub-figure, genes within 500kb of the index SNP are shown. The position on each chromosome is shown in relation to Human Genome build hg19.

Supplementary Figure 3. MAGMA Gene-based association analysis Manhattan plot. Association results for varicose veins in MAGMA gene-based association analysis. The dotted red line indicates the threshold for genome-wide significance ($P < 2.68 \times 10^{-6}$). 248 genes reached genome-wide significance in this analysis, with the top-ten genes highlighted in this figure.

3. Supplementary Data

Supplementary Data 1. Genome-wide significant variants at all significant discovery loci. 12,391 genome-wide significant variants ($P < 5 \times 10^{-8}$) at all 109 genomic risk loci identified in the discovery GWAS. Where the column headings pertain to: the SNP ID (RsID), chromosome, base position within NCBI Genome Build 37 (hg19), the effect allele, the alternate (non-effect) allele, the effect allele frequency in the study population, the imputation quality score; where 1 = genotyped SNP, the SNP effect size (BETA), the standard error of the BETA, and the GWAS association P-value. All variants are presented in ascending order according to chromosome and base position.

Supplementary Data 2. Genome-wide significant variants at the replicated susceptibility loci. 5,315 genome-wide significant variants ($P < 5 \times 10^{-8}$) identified by FUMA SNP2GENE at 45 of 46 replicated genomic risk associated with varicose veins. Where the column headings pertain to: the unique ID of the variant, the SNP ID (RsID), chromosome, base position within NCBI Genome Build 37 (hg19), the effect allele, the alternate (non-effect) allele, the effect allele frequency in the study population, the SNP effect size (BETA), the standard error of the BETA, the nearest gene according to positional mapping in FUMA, the functionality of the SNP as derived from ANNOVAR, the Combined Annotation Dependent Depletion (CADD) score of the variant, the RegulomeDB (RDB) score of the variant, the minimum chromatin state of the variant.

Supplementary Data 3. MAGMA gene set analysis. Gene sets were obtained from MsigDB v7.0 and total of 15,496 gene sets (Curated gene sets: 5,500, GO terms: 9,996) were tested. Curated gene sets consists of 9 data resources including KEGG, Reactome and BioCarta (http://software.broadinstitute.org/gsea/msigdb/collection_details.jsp#C2 for details). GO terms consists of three categories, biological processes (bp), cellular components (cc) and molecular functions (mf). All parameters were set as default (competitive test). Gene sets are arranged in order of ascending P-value, with a P-value $< 3.23 \times 10^{-6}$ indicative of a significant gene set, accounting for multiple testing (0.05/15,496).

Supplementary Data 4. UK Biobank Summary Statistics. The full summary statistics for the discovery GWAS of varicose veins in UK Biobank can be found on the Oxford Research Archive (ORA). URL: TBA.

4. Supplementary References

Supplementary References 51- 66