Clinical and Genetic Determinants of Varicose Veins
Prospective, Community-Based Study of ≈500,000 Individuals

BACKGROUND: Varicose veins are a common problem with no approved medical therapies. Although it is believed that varicose vein pathogenesis is multifactorial, there is limited understanding of the genetic and environmental factors that contribute to their formation. Large-scale studies of risk factors for varicose veins may highlight important aspects of pathophysiology and identify groups at increased risk for disease.

METHODS: We applied machine learning to agnostically search for risk factors of varicose veins in 493,519 individuals in the UK Biobank. Predictors were further studied with univariable and multivariable Cox regression analyses (2,441 incident events). A genome-wide association study of varicose veins was also performed among 337,536 unrelated individuals (9,577 cases) of white British descent, followed by expression quantitative loci and pathway analyses. Because height emerged as a new candidate risk factor, we performed mendelian randomization analyses to assess a potential causal role for height in varicose vein development.

RESULTS: Machine learning confirmed several known (age, sex, obesity, pregnancy, history of deep vein thrombosis) and identified several new risk factors for varicose vein disease, including height. After adjustment for traditional risk factors in Cox regression, greater height remained independently associated with varicose veins (hazard ratio for upper versus lower quartile, 1.74; 95% CI, 1.51–2.01; P<0.0001). A genome-wide association study identified 30 new genome-wide significant loci, identifying pathways involved in vascular development and skeletal/limb biology. Mendelian randomization analysis provided evidence that increased height is causally related to varicose veins (inverse-variance weighted: odds ratio, 1.26; P=2.07×10^{-16}).

CONCLUSIONS: Using data from nearly a half-million individuals, we present a comprehensive genetic and epidemiological study of varicose veins. We identified novel clinical and genetic risk factors that provide pathophysiological insights and could help future improvements of treatment of varicose vein disease.

© 2018 American Heart Association, Inc.
Clinical Perspective

What Is New?

- In this population-based study of ≈500,000 individuals, greater height appeared as a novel predictor of varicose vein disease in machine learning analyses and was independently associated in multivariable-adjusted Cox regression.
- Using mendelian randomization, we demonstrate that greater height has a causal role in varicose vein development.
- We identify 30 genetic loci in the first large-scale genome-wide analysis of varicose veins, including 337,536 individuals (9577 cases and 327,959 controls), and discover a strong genetic correlation between varicose veins and deep vein thrombosis.

What Are the Clinical Implications?

- Varicose veins are a common condition that are increasingly being associated with serious health risks such as venous thromboembolism; yet little is known about the genetic and environmental risk factors for varicose vein disease.
- Our study demonstrates the causal role of height and identifies novel genetic and clinical associations with varicose veins.
- This knowledge greatly expands our understanding of disease pathophysiology and may help future improvements in the management of varicose veins and their associated complications.

Varicose veins are a common manifestation of chronic venous disease. It is estimated that >30 million adults in the United States have varicose veins, with interventions consuming more than $1 billion in direct healthcare resources per year. Although sometimes dismissed as a cosmetic finding, varicose veins can culminate with life-limiting ulceration in up to 20% of affected individuals and have increasingly been associated with serious health risks. Recent studies show that patients with varicose veins have up to a 5-fold increased risk for developing deep vein thrombosis (DVT). Moreover, the association of varicose veins with peripheral artery disease (PAD) and additional vascular diseases is also being elucidated.

The factors influencing varicose vein formation are not fully understood. Epidemiological studies have established multiple risk factors such as age, female sex, pregnancy, obesity, and prior DVT, but several other putative factors remain unconfirmed. There is also a clear familial component to varicose vein disease, but prior genetic studies have been small and have provided conflicting results. Accordingly, it remains difficult to identify those at risk, and our lack of biological insights into pathogenesis may help explain why no approved therapies exist to prevent or delay disease progression.

To address the need for larger-scale studies of varicose vein disease, we investigated the clinical, environmental, and genetic bases of varicose veins in up to 502,619 individuals from the UK Biobank using a constellation of modalities, including agnostic machine learning; traditional epidemiological analyses; genetic studies, including a genome-wide association study (GWAS) and pathway analyses; and mendelian randomization (MR) to address causality.

METHODS

All data reported in this article are made available to other researchers via application to the UK Biobank for purposes of reproducing the results or replicating the procedure. The data are publicly available at the UK Biobank repository.

Study Design and Participants

We used data from the UK Biobank, a longitudinal cohort study of 502,619 individuals 40 to 69 years of age from 21 centers across the United Kingdom. Between 2006 and 2010, individuals 40 to 69 years of age underwent baseline assessment and have since been carefully investigated for extensive phenotype, lifestyle, health outcomes, and genetic data. The UK Biobank protocol has been described in detail previously and can be found online (https://www.ukbiobank.ac.uk).

In our observational analyses, we excluded individuals with prevalent varicose veins (n = 9100), leaving 493,519 individuals eligible. In the GWAS, we studied unrelated individuals of white British ancestry with available phenotype data (n = 337,536). The UK Biobank study was approved by the North West Multi-Center Research Ethics Committee, and all participants provided written informed consent to participate in the UK Biobank.

Definitions of Potential Risk Factors

To identify risk factors for varicose veins, we considered all variables from the UK Biobank, spanning domains of demographics, socioeconomic data, lifestyle behaviors, medical history, and a range of physical and cognitive measures. We excluded administrative, temporary, and unspecific International Classification of Diseases (ICD) codes (chapters U and Z), as well as diagnoses likely to represent varicose veins, including phlebitis, thrombophlebitis, and other diseases of veins and lymphatic vessels (ICD, 10th Revision codes 180–182, 185–189).

A total of 2715 predictor variables were extracted from the UK Biobank database and used in our phenome-wide scan of varicose vein risk factors (Table I in the online-only Data Supplement). Medical history and smoking status were assessed with self-reported questionnaires. Blood pressure and body mass index (BMI) were measured in a standard fashion (full details available at UK Biobank data showcase; http://biobank.ox.ac.uk/crystal). Systolic blood pressure was defined as the mean of 2 measurements.
Follow-Up and Outcome
The primary outcome was the diagnosis of varicose veins, defined by ICD, Ninth Revision (454*) and/or ICD, 10th Revision (I-83*). For observational analyses, prevalent cases were excluded (n=9100), leaving 2441 cases of incident varicose veins during a median follow-up of 6.2 years (maximum, 9.1 years). For GWAS and MR analyses, both incident and prevalent events were included (n=11541). After the exclusion of individuals as a result of genetic relatedness or non-European descent (n=1964), 9577 cases were included.

Machine Learning Analyses
For agnostic discovery of novel risk factors for varicose veins, we used a gradient boosting machine (GBM) model. On the basis of the input predictor variables, this machine learning algorithm consecutively fits new decision trees to provide a more accurate prediction of a response variable (ie, outcome, varicose veins). The principal idea of this algorithm is the learning procedure that results in consecutive error fitting, with each decision tree chosen to minimize the loss function. In addition to classification, GBM ranks the importance (weight) of input variables to the outcome based on decision trees. We used incident cases of varicose veins as the response variable. Results are presented as variable importance. All analyses were conducted with the machine learning platform (H2O 3.10.4.1). Detailed descriptions of the GBM model can be found in Methods in the online-only Data Supplement.

Statistical Analysis
We analyzed the associations of traditional varicose vein risk factors and the top 10 predictors identified through GBM on incident varicose veins using univariable and multivariable Cox regression analyses. For categorical predictors, Kaplan-Meier estimates were used. For continuous predictors, the predictor was entered as a restricted cubic spline to account for any nonlinear associations. The top 10 novel predictors were further assessed with 2 multivariable-adjusted Cox models: adjusting for age and sex and additionally adjusting for established varicose vein risk factors, including BMI, waist-hip ratio (WHR), history of DVT, and pregnancy. As a secondary analysis, we assessed variables in a third model that included the first 5 principal components of ancestry to account for genetic ancestry. The discriminatory ability of the models was assessed with C indexes. All regression analyses were conducted with R 3.3.0 and 3.3.3.

GWAS of Varicose Veins
A GWAS was performed using imputed genotype data from the second data release of the UK Biobank (9577 cases and 327959 controls). A detailed description of genotyping, quality control, and imputation can be found in the Methods in the online-only Data Supplement. We included 10972371 markers belonging to the Haplopreference Reference Consortium 1.1 subset with a minor allele count ≥20 in cases and controls and MaCH r² ≥ 0.8. Assuming an additive model, the associations between genotype dosages of each marker and varicose veins were tested in logistic regression (Firth penalized logistic regression in nonconvergence) with PLINK215 and adjusted for age, sex, and the first 15 principal components of ancestry. A value of P<5x10⁻⁸ was considered to be statistically significant. Independent lead single nucleotide polymorphisms (SNPs) were identified by conditional analysis. Regional plots were created for the association test results of significant loci with Locuszoom version 1.4. All other plots and tables were created in R 3.3.0.

Linkage Disequilibrium Score Regression
Linkage disequilibrium score regression (LDSC) is a tool used to estimate SNP heritability, genetic correlation, and genomic inflation using GWAS summary data sets. Because height appeared as a novel predictor in our analyses, we applied LDSC to estimate the genetic correlation between height and varicose veins using summary statistics from our GWAS for varicose veins and publically available GWAS results for height from the GIANT consortium (Genetic Investigation of Anthropometric Traits). As a comparison, we also examined the genetic correlations with PAD and traditional risk factors for varicose veins: BMI, WHR, and history of DVT. For BMI and WHR, we used GWAS summary data from the GIANT consortium limited to individuals of European ancestry. For history of DVT and PAD, GWAS results from the UK Biobank were used. We additionally applied LDSC to estimate SNP heritability and to test for genomic inflation (expected to some degree for polygenic traits and large sample sizes). SNP heritability is reported on the liability scale to take into account sample prevalence of varicose vein disease.

Expression Quantitative Loci Analysis
To explore the potential functional significance of our findings, we evaluated all genome-wide significant loci for evidence of expression quantitative loci (eQTL). Lookups were performed in the Genotype-Tissue Expression database (53 tissue types), and the FUMA GWAS platform was used to interrogate whole-blood data repositories, BIOSQTL, and the blood eQTL browser.

Pathway and Functional Analyses
We used 2 data-driven integrative platforms to perform pathway analyses. After clumping to identify independent loci using PLINK2, we performed analysis using DEPICT. We also used the FUMA GWAS platform to investigate loci for pathway and tissue enrichment. A detailed description of pathway analysis can be found in Methods in the online-only Data Supplement.

Height and Varicose Veins: MR Analysis
We performed a 2-sample MR analysis using height-associated SNPs from the GIANT study, a GWAS meta-analysis of height in 253,288 individuals of European ancestry. A genetic risk score was created with height-associated SNPs as the instrumental variable and our GWAS of varicose veins as the outcome. Our main genetic risk score instrument included 512 SNPs without evidence of pleiotropy (see Methods in the online-only Data Supplement). Analyses were conducted with the R package TwoSampleMR. Power was calculated with the online tool (https://sb452.shinyapps.io/power/).
RESULTS

Machine Learning and Cox Analyses

Baseline characteristics of the study sample are shown in Table 1. The 20 variables with the highest importance in the GBM model are presented in Figure 1. Age, pregnancy, history of DVT, and obesity are previously reported risk factors that were again observed in our study (Table II in the online-only Data Supplement). In addition, several novel predictors of varicose veins were identified, including leg bioimpedance, height, and history of surgery on leg arteries or other major operation. Our GBM model had similar areas under the receiver-operating characteristics curve in the training, validation, and test sets: 0.76, 0.71, and 0.70, with each set having a mean squared error of 0.02, indicating that the model had acceptable discriminatory performance and low error. A number of putative predictors that have been inconsistently reported in the literature, including hypertension (variable 238), age at menopause (variable 405), and smoking (variable 487), were not observed among the top predictors by our machine learning approaches.

The top 10 new predictors were selected, together with previously established risk factors, for additional analysis and validation by 2 Cox models (age and sex adjusted and adjusted for established varicose vein risk factors). Among the independently associated variables were established risk factors such as history of DVT (hazard ratio, 2.60; 95% CI, 2.18–3.11; \( P < 0.0001 \)) and BMI (hazard ratio, 1.10; 95% CI 1.03–1.17; \( P = 0.001 \); Table 2) but also novel markers such as leg bioimpedance, greater height, and surgery on leg arteries (Figure 2 and Table 2 and Figure I in the online-only Data Supplement). Each new predictor from our GBM model improved the predictive capacity of the Cox models when added to models including age, sex, and traditional risk factors and remained independently associated in our secondary analysis that also adjusted for genetic ancestry (Table III in the online-only Data Supplement).

The largest improvement in the C index was observed for leg bioimpedance and height (from 0.56 to 0.65 and 0.61, respectively; Table IV in the online-only Data Supplement). Of note, debated risk factors such as oral contraceptive use were not found to be independently associated with varicose veins in Cox analysis, whereas some were, including systolic blood pressure.

GWAS for Varicose Veins

We conducted a GWAS to identify genetic variants that influence varicose vein susceptibility. Overall, the GWAS results showed modest deviation in test statistics (\( \lambda = 1.09 \)). The constructed quantile-quantile plot showed minimal inflation except for in the upper tail of distribution (Figure II in the online-only Data Supplement), suggesting adequate correction for population stratification. LDSC supported this and provided evidence that inflation was caused by a polygenic signal in which heritability accounted for 28% of the population variance in varicose vein disease (LDSC intercept, 1.032, SE=0.0088; \( h^2 = 0.28 \), SE=0.024).

We identified 855 new SNPs associated with varicose veins that exceeded genome-wide significance (Figure 3). After conditional analyses, we found 30 independent genetic variants associated with varicose vein disease (Table V in the online-only Data Supplement). The strongest association was located on chromosome 1 in the CASZ1 gene, in a known blood pressure locus (rs11121615; \( P = 3.71 \times 10^{-15} \)).24 Other lead variants were located in the 16q24 region, in or near vascular mechano-sensory channel PIEZO1 and galactosamine-sulfatase enzyme GALNS (rs2911461, \( P = 4.81 \times 10^{-29} \); rs8053350, \( P = 8.93 \times 10^{-18} \)). Regional plots for the 30 newly discovered loci are shown in Figure III in the online-only Data Supplement. Full results for all SNPs with values of \( P < 5 \times 10^{-8} \) are shown in Table VI in the online-only Data Supplement.

Pathway and Functional Analyses

eQTL analyses identified several genes significantly regulated by varicose vein SNPs. For example, one of the lead variants, rs2861819, had several eQTL associations, including PNO1, WDR92, PLEK, and PPP3R1, which expresses calcineurin, a critical signaling component during angiogenesis.25 We also identified significant eQTLs for lead variants rs2911461 (PIEZO1,
P = 1.50 \times 10^{-5}; \text{ GALNS, } P = 1.50 \times 10^{-5}) \text{ and rs3101725 (FBN2, } P = 9.50 \times 10^{-7}) \text{. These loci are particularly interesting because mutations in GALNS and FBN2 have been linked to a hereditary form of skeletal dysplasia}^{26} \text{ and the Marfan-like syndrome congenital contractual arachnodactyly,}^{27} \text{ respectively. Detailed results from eQTL analysis are given in Table VII in the online-only Data Supplement.}

FUMA prioritized tissues, pathways, and gene sets enriched in varicose vein loci. Tissue enrichment was greatest for adipose tissue (P = 7.58 \times 10^{-5}; Table VIII in the online-only Data Supplement). Evaluation for enriched biological processes identified several gene sets related to vascular development and structure (vascular development, P = 1.96 \times 10^{-6}; endothelial cell differentiation, P = 8.95 \times 10^{-6}) \text{ and highlighted canonical pathways that may influence varicose vein disease: Gata3 pathway (P = 3.06 \times 10^{-6}), vascular endothelial growth factor signaling (P = 1.11 \times 10^{-5}), and nuclear factor of activated T cells pathway (P = 1.29 \times 10^{-4}; Table IX in the online-only Data Supplement).}

In addition, FUMA analysis revealed that our lead SNPs were previously associated with a variety of traits, including varicose vein predictors identified in our observational analyses: WHR (P = 2.79 \times 10^{-3}; Table VIII in the online-only Data Supplement). Evaluation for enriched biological processes identified several gene sets related to vascular development and structure (vascular development, P = 7.58 \times 10^{-5}; Table VIII in the online-only Data Supplement). Evaluation for enriched biological processes identified several gene sets related to vascular development and structure (vascular development, P = 1.96 \times 10^{-6}; endothelial cell differentiation, P = 8.95 \times 10^{-6}) \text{ and highlighted canonical pathways that may influence varicose vein disease: Gata3 pathway (P = 3.06 \times 10^{-6}), vascular endothelial growth factor signaling (P = 1.11 \times 10^{-5}), and nuclear factor of activated T cells pathway (P = 1.29 \times 10^{-4}; Table IX in the online-only Data Supplement).}

Genetic Correlation Between Height and Varicose Veins

After integrating data from our GWAS and publicly available GWAS summary results on height, we found evidence for a strong genetic correlation between height and varicose veins (r = 0.16; SE = 0.031; P = 2.99 \times 10^{-7}). This suggests that there is \approx 16% genetic overlap between height and varicose veins, which was larger than the estimated genetic correlation with BMI (r = 0.095; SE = 0.032; P = 0.00260) and WHR (r = 0.078; SE = 0.052; P = 0.131). History of DVT was found to have the greatest genetic correlation with varicose vein disease (r = 0.36; SE = 0.0805; P = 5.57 \times 10^{-7}) \text{ compared with PAD (r = 0.20; SE = 0.22; P = 0.37) and traditional factors such as BMI. The estimated genetic correlation of varicose veins and PAD was lower than that of DVT and was nonsignificant.}

Mendelian Randomization

Because we observed a strong association of height with varicose veins, we next performed multiple MR analyses to investigate causal effects. We used a genetic risk score instrument composed of 512 independent SNPs previously associated with height, without evidence of pleiotropy. The results provided strong evidence of a causal effect of height on varicose veins (inverse-variance weighted: odds ratio [OR], 1.26, P = 2.07 \times 10^{-16}; weighted median: OR, 1.28, P = 1.79 \times 10^{-8}; maximum likelihood: OR, 1.26, P = 9.12 \times 10^{-8}; MR Egger: OR, 1.43, P = 6.74 \times 10^{-6}; Table X in the online-only Data Supplement). These results remained consistent in a leave-one-out sensitivity analysis, and a funnel plot did not indicate any directional pleiotropy (Figures IV and V in the online-only Data Supplement). Sensitivity analyses including all 536 variants (without exclusion of potentially pleiotropic variants) provided almost identical results.
**DISCUSSION**

Leveraging data from 493,519 individuals, this study is the largest to investigate the genetic and environmental bases of varicose veins. First, we used “hypothesis-free” machine learning approaches to identify clinical and environmental predictors of disease. This powerful method confirmed a number of well-established relationships and identified a variety of novel risk factors that have not been previously reported. Second, we used multivariable Cox regression methods to validate our epidemiological observations and confirm their importance for predicting disease risk. Third, we performed a GWAS and pathway analyses to discover genetic determinants of varicose veins. Finally, we used MR to establish height as a causal risk factor for disease.

**Environmental Predictors of Varicose Veins**

Prior studies have yielded inconsistencies in environmental predictors of varicose veins. Machine learning confirmed several of the more accepted risk factors (eg, age, history of DVT, obesity, pregnancy),\(^1\,^2\,^6\,^8\) in addition to providing insight into previously debated predictors. Age when full-time education was completed (reflecting length of education) was among the most important variables identified in the GBM model, which supports the association between education and varicose veins that has been observed in smaller studies.\(^8\,^28\)

It is possible that this association is related to the ability to obtain medical treatment because level of education may reflect socioeconomic circumstances and access to health care. Individuals with higher education may be more inclined to seek and receive medical care for varicose veins and thus receive a formal diagnosis. Previous studies have also found a significant association with hormone replacement therapy,\(^2\,^28\) which again appeared as an important variable in our GBM model and was positively related to incident varicose veins in multivariable analysis. Supplemental estrogen may alter venous wall compliance; a previous study showed that elevated levels of endogenous estradiol are associated with increased venous distensibility.\(^29\)

On the contrary, machine learning did not prioritize several other controversial factors (eg, oral contraceptive use; Table 2),\(^2\,^6\,^8\,^28\) which remained consistent in subsequent formal Cox regression analyses (investigating 1 risk factor at a time). Given the size of this study, it is likely that these previously debated factors have limited utility as traditional risk discriminators.

Physical activity, smoking, and hypertension are other debated risk factors for varicose vein disease. Although duration of moderate daily activity was not found to be a strong predictor in our GBM model and

---

**Table 2. Cox Models of The Top Novel Predictors From the Gradient Boosting Machine approach, Along With Debated Predictors for Varicose Veins**

<table>
<thead>
<tr>
<th>Variable</th>
<th>Age- and Sex-Adjusted Hazard Ratio (95% CI)</th>
<th>Fully Adjusted(^*) Hazard Ratio (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Traditional</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>History of deep vein thrombosis</td>
<td>2.71 (2.28–3.23)</td>
<td>2.60 (2.18–3.11)</td>
</tr>
<tr>
<td>Birth weight of first child</td>
<td>1.23 (1.05–1.44)</td>
<td>1.21 (1.03–1.42)</td>
</tr>
<tr>
<td>Number of live births</td>
<td>1.12 (0.99–1.27)</td>
<td>1.12 (0.99–1.27)</td>
</tr>
<tr>
<td>Body mass index</td>
<td>1.11 (1.00–1.24)</td>
<td>1.11 (0.99–1.24)</td>
</tr>
<tr>
<td>Waist-hip ratio</td>
<td>1.10 (0.97–1.26)</td>
<td>1.01 (0.88–1.16)</td>
</tr>
<tr>
<td><strong>Novel</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Leg biompedance</td>
<td>0.56 (0.51–0.63)</td>
<td>0.43 (0.39–0.50)</td>
</tr>
<tr>
<td>Height</td>
<td>1.70 (1.48–1.96)</td>
<td>1.74 (1.51–2.01)</td>
</tr>
<tr>
<td>Major operations</td>
<td>0.75 (0.69–0.82)</td>
<td>0.77 (0.70–0.84)</td>
</tr>
<tr>
<td>Surgery on leg arteries</td>
<td>3.71 (2.46–5.62)</td>
<td>3.50 (2.30–5.33)</td>
</tr>
<tr>
<td>Ulceration of lower limb</td>
<td>4.65 (2.36–9.16)</td>
<td>3.00 (1.36–6.61)</td>
</tr>
<tr>
<td>Other soft tissue disorders</td>
<td>2.01 (1.58–2.55)</td>
<td>1.66 (1.30–2.12)</td>
</tr>
<tr>
<td>Cellulitis (ICD-10 code L03)</td>
<td>2.09 (1.52–2.87)</td>
<td>1.79 (1.29–2.48)</td>
</tr>
<tr>
<td><strong>Debated</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Smoking</td>
<td>1.04 (0.92–1.19)</td>
<td>1.04 (0.91–1.19)</td>
</tr>
<tr>
<td>Systolic blood pressure</td>
<td>0.81 (0.72–0.90)</td>
<td>0.79 (0.71–0.88)</td>
</tr>
<tr>
<td>Age completed full-time education</td>
<td>0.80 (0.70–0.91)</td>
<td>0.82 (0.72–0.93)</td>
</tr>
<tr>
<td>Oral contraceptive use</td>
<td>0.67 (0.44–1.02)</td>
<td>0.68 (0.44–1.04)</td>
</tr>
<tr>
<td>Hormone replacement therapy</td>
<td>1.25 (1.04–1.49)</td>
<td>1.25 (1.04–1.50)</td>
</tr>
<tr>
<td>Days per week walked &gt;10 min</td>
<td>1.32 (1.18–1.47)</td>
<td>1.33 (1.19–1.48)</td>
</tr>
<tr>
<td>Duration of moderate daily activity</td>
<td>1.00 (0.87–1.14)</td>
<td>1.00 (0.88–1.15)</td>
</tr>
</tbody>
</table>

Associations between candidate variables and varicose veins are reported as hazard ratio and 95% CI for the upper versus lower quartile for continuous variables or presence versus absence for binary variables. Separate models were used for each variable.

ICD-10 indicates International Classification of Diseases, 10th Revision.

*Fully adjusted model includes age, sex, body mass index, history of deep vein thrombosis, and waist-hip ratio. Pregnancy-specific factors, oral contraceptive use, and hormone replacement therapy were assessed only in female study participants.

†Weight in pounds.

but with a larger degree of heterogeneity, as expected (Table X in the online-only Data Supplement). At an \(\alpha\) threshold of 0.05, power for MR analyses was estimated to be 100% with an OR of 1.51 per 1-SD increase in height (corresponding to the effect from our observational analyses).
Cox analyses, the number of days per weeks of walking ≥10 minutes was positively associated. This difference may reflect a strong influence of immobility and nonambulatory status on disease rather than routine aerobic exercise per se. Some studies have reported higher smoking rates in subjects with varicose veins compared with those without disease.6,30 However, the majority of studies have observed no association between varicose veins and smoking.31–35 In our study, smoking was not associated with varicose vein development after adjustment for traditional risk factors but became marginally associated in our secondary analysis that adjusted for genetic ancestry. Studies of the association between varicose veins and hypertension have also shown various relationships, including a positive association,6 no association,7 and a lower prevalence of hypertension among subjects with varicose veins.8 In multivariable analysis, we observed a slight

Figure 2. Univariable analysis demonstrating relationships between varicose veins and other variables. DVT indicates deep vein thrombosis.

Figure 3. Manhattan plot of genome-wide association study analysis. Gray line represents the threshold of genome-wide significance, $P<5\times10^{-8}$. Gene names correspond to the gene in closest proximity to the variant.
inversion association between systolic blood pressure and varicose veins. Given this modest association of varicose veins with both smoking and blood pressure, which could reflect reverse causation or confounding, more information is needed to evaluate their potential shared pathophysiology with varicose veins.

Because our machine learning algorithms are not restricted to a priori hypotheses, they had the capacity to identify novel factors associated with disease. By using this “agnostic” phenotype-wide approach, we were able to identify several new strong predictors, including leg bioimpedance and height. Height was suggested to be a potential risk factor in an early epidemiological study several decades ago but has been inconsistently reported since then. Bioimpedance, defined as the ability of tissue to impede electric current, is related to the amount of fluid accumulation in body tissue. Combined with the recent finding that leg bioimpedance is associated with risk for heart failure, these novel predictors suggest a link between high-volume venous reflux, increased hydrostatic pressure, and resulting venous hypertension.

Genetics of Varicose Veins

Although varicose veins are known to have a strong heritable component, their pathophysiological underpinnings remain poorly understood. To date, genetic research on varicose veins has been restricted to syndromic forms of disease (eg, Klippel-Trenaunay syndrome) and candidate gene studies in patients with primary venous disease. These studies have suggested that mutations in FOXC2, thrombomodulin (THBD), and desmoplakin (SYNM) may promote the development of varicose veins by altering vein function. However, these studies were typically small (18–700 patients), and few confirmation studies have been reported.

In our GWAS, we included nearly 10 000 cases and 300 000 controls; thus, our analyses were uniquely powered to screen the whole genome in an unbiased manner. We identified 30 genetic loci robustly associated with varicose veins, with the strongest associations occurring in the intron region of CASZ1 (rs11121615; \( P=3.71 \times 10^{-65} \)), which has been implicated in blood pressure, and in 16q24, where, among other genes, PIEZO1, encoding a vascular mechanosensory channel, is located (rs2911463; \( P=4.81 \times 10^{-29} \)). Disruption of PIEZO1 has previously been shown to result in significant disorganization of the vascular system, suggesting the importance of PIEZO1 for mature vascular development. In addition, the lead SNP rs7773004 on 6q14 is located within 50 kb upstream of the hemochromatosis gene HFE, in which mutations have previously been associated with both venous ulceration and venous thromboembolism (an aggregate of DVT and pulmonary embolism). These loci appear particularly interesting given the strong genetic overlap that we discovered between varicose veins and DVT. Genes regulating vascular dysfunction may promote changes in venous architecture and chronic blood stasis (and could represent new translational targets).

In a comparison of the genetic correlation of varicose veins with traditional risk factors and PAD, LDSC demonstrated the strongest genetic correlation with history of DVT. Several genes have previously been linked to both varicose veins and DVT, including THBD and methylentetrahydrofolate reductase (MTHFR). Mutations in these genes are classically associated with inherited hypercoagulability, but patients with varicose veins have also been found to have a higher prevalence of thrombophilias and increased levels of systemic inflammatory and prothrombotic markers. In addition, a nationwide genetic study in Sweden demonstrated that varicose veins and venous thromboembolism share familial susceptibility. These findings similarly support the shared genetic associations between varicose veins and DVT and suggest an overlap in their underlying pathophysiology.

To gain further biological insights from our GWAS, we performed a series of analyses to identify genes and pathways that may be dysregulated in individuals predisposed to varicose vein formation (Tables VII through IX in the online-only Data Supplement). eQTL analyses suggested that several SNPs are associated with altered in vivo expression of genes that have been related to vascular development/integrity (eg, PPP3R1, PIEZO1, SOX18, THBD), limb development (eg, LBR), and conditions associated with skeletal abnormalities (eg, GALNS, FBN2) (Table VII in the online-only Data Supplement). The potential importance of genes related to vascular development (\( P=1.96 \times 10^{-6} \)), limb development (\( P=3.55 \times 10^{-3} \)), and height (\( P=2.30 \times 10^{-3} \)) was also highlighted in pathway analyses (Table IX in the online-only Data Supplement). In particular, genetic variants associated with varicose veins were found to be enriched for the vascular endothelial growth factor pathway and nuclear factor of activated T cells pathway. Vascular endothelial growth factor signaling is required for vessel formation and homeostasis, whereas the nuclear factor of activated T cells pathway regulates intracellular calcium signaling in osteoblasts and osteoclasts during bone growth and remodeling. These findings warrant future studies to experimentally evaluate the reported genetic loci.

Because height was associated with varicose veins in our observational analyses, we concluded our studies with LDSC and MR analyses that demonstrated a shared genetic pathogenesis and confirmed a strong causal relationship between increased height and varicose veins (Figures IV and V and Table X in the online-only Data Supplement). Height has also been found to be an independent predictor for venous thromboembolism.
bolism. Considering this, our findings suggest that height may have consequences on long-term risk for both varicose veins and the associated complications.

**Study Strengths and Limitations**

Strengths of our study include the large sample size and use of combined epidemiological and genetic studies, making it the largest and most comprehensive study of varicose veins to date. Nevertheless, this study has important limitations. The UK Biobank includes predominantly white middle-aged to elderly individuals, so our conclusions may have limited generalizability to people of other ethnicities and age groups (such as younger people who have not yet developed disease). Our epidemiological analysis identified incident events with hospital ICD codes. Although this allowed standardized identification of varicose vein events, the true incidence of varicose veins may also be underrepresented in this data set because diagnosis required a formal ICD code input and does not capture data from individuals who did not seek medical attention or include cases of self-reported varicose veins. Thus, it is possible that the controls in our analysis included undocumented cases of varicose veins. This may introduce bias and underestimate some of the associations, driving them toward the null. It is thought that the potential bias could lead to false-positive associations. In addition, limitations in characterizing candidate variables such as physical activity, job, and diet may have altered the strength of associations in our observational analyses, which could be a reason why some of the debated risk factors did not show significant associations. Lastly, our analysis does not include the relation of our findings to disease management. The goal of this study was to identify environmental and genetic risk factors for varicose veins. Thus, although our discoveries may help predict or treat varicose veins in the future, the study did not focus on their impact on current patient care.

**Conclusions**

Using a large study to investigate varicose vein disease, our study substantially extends our understanding of the genetic and environmental bases of varicose veins. We demonstrate that greater height is associated with greater susceptibility to varicose vein disease. Several risk factors with importance for predicting disease risk were identified, including previously debated (hormone replacement therapy, length of education) and other novel (leg bioimpedance, history of cellulitis, surgery on leg arteries) predictors. Our discovery of genetic determinants of disease also sets a path for identification of new targets in the effort to develop therapy for varicose veins.

**REFERENCES**

Fukaya et al. Clinical and Genetic Determinants of Varicose Veins


doi: 10.1038/nature14177


