

<https://doi.org/10.1038/s41525-025-00464-w>

Insights from the largest diverse ancestry sex-specific disease map for genetically predicted height

Check for updates

A. Papadopoulou¹, E. M. Litkowski^{2,3}, M. Graff⁴, Z. Wang^{5,6}, R. A. J. Smit^{5,6,7,8}, G. Chittoor⁹, I. Dinsmore¹⁰, N. S. Josyula⁹, M. Lin¹¹, J. Shortt¹¹, W. Zhu¹², S. L. Vedantam^{13,14}, L. Yengo¹⁵, A. R. Wood¹⁶, S. I. Berndt¹⁷, I. A. Holm¹⁸, F. D. Mentch¹⁹, H. Hakonarson¹⁹, K. Kiryluk²⁰, C. Weng²¹, G. P. Jarvik²², D. Crosslin²³, D. Carrell²⁴, I. J. Kullo²⁵, O. Dikilitas²⁵, M. G. Hayes²⁶, W. -Q. Wei²⁷, D. R. V. Edwards²⁸, T. L. Assimes²⁹, J. N. Hirschhorn^{14,30,31}, J. E. Below¹², C. R. Gignoux¹¹, A. E. Justice⁹, R. J. F. Loos^{5,6,8,32}, Y. V. Sun^{33,34}, S. Raghavan^{2,3}, P. Deloukas^{1,36}, K. E. North^{4,36} & E. Marouli^{1,35,36} ✉

We performed ancestry and sex specific Phenome Wide Association Studies (PheWAS) to explore disease related outcomes associated with genetically predicted height. This is the largest PheWAS on genetically predicted height involving up to 840,000 individuals of diverse ancestry. We explored European, African, East Asian ancestries and Hispanic population groups. Increased genetically predicted height is associated with hyperpotassemia and autism in the male cross-ancestry analysis. We report male-only European ancestry associations with anxiety disorders, post-traumatic stress and substance addiction and disorders. We identify a signal with benign neoplasm of other parts of digestive system in females. We report associations with a series of disorders, several with no prior evidence of association with height, involving mental disorders and the endocrine system. Our study suggests that increased genetically predicted height is associated with higher prevalence of many clinically relevant traits which has important implications for epidemiological and clinical disease surveillance and risk stratification.

Adult height is an easily measured anthropometric trait that is complex and highly heritable^{1,2}. Several factors contribute to adult height including both genetic and environmental, such as nutrition, socio-economic status, and physical activity^{3–7}.

Several observational studies have been performed to better understand the association between height and disease. In individuals of European ancestry, increased height has been associated with a reduced risk of several circulatory diseases, including coronary artery disease (CAD), aortic valve stenosis (AS), heart failure (HF), hypertension and stroke⁶. In addition to these observational studies, increased genetically predicted height has been associated with decreased risk of hypertension, diaphragmatic hernia, and gastro-esophageal reflux disease (GERD)⁶. A recent study in the Million Veteran Program (MVP) used a polygenic score based on 3290 height-associated Single Nucleotide Polymorphisms (SNPs)² to show that increased genetically predicted height is associated with an increased risk of atrial fibrillation (AF) and decreased risk of CAD, hypertension, hyperlipidemia. They also reported potential novel associations with peripheral

neuropathy and infections of the skin and bones, both in European and African ancestries individuals⁸. Furthermore, increased genetically predicted height has been associated with longer PR interval and QRS duration⁹, venous thromboembolism^{6,10}, AF, intervertebral disc disorder, hip fracture, vasculitis, breast cancer^{6,11} and colorectal cancer^{6,12} in European ancestry.

The Genetics of Anthropometric Traits - (GIANT) - consortium has performed increasingly larger meta-analyses of genome-wide association studies (GWAS) of height over the years^{2,13,14}. In the present study we are using a multi-ancestry polygenic score (PGS) for height in six study populations of diverse ancestries to explore the association with a comprehensive set of health-related outcomes. The PGS for height was constructed using genetic variants taken from the most recent GIANT GWAS for adult height, excluding data from 23andMe¹⁴. We employed a Phenome-Wide Association Analyses (PheWAS) approach; a hypothesis-free analysis, with no prior assumptions, to detect phenotypes associated with the height PGS^{15,16}, followed by meta-analysis (meta-PheWAS) of the individual

A full list of affiliations appears at the end of the paper. ✉ e-mail: e.marouli@qmul.ac.uk

PheWAS in each study population both within and across-ancestry groupings, to potentially identify new diseases associated with genetically predicted height. Sex-stratified cross-ancestry analyses were also considered.

Results

We performed a PheWAS in each cohort using the PGS of height as exposure and tested its association with disease outcomes available in each of them (Methods). The sex-combined cross-ancestry meta-PheWAS in up to 839,872 participants, interrogating 1768 traits (available in at least 2 cohorts), yielded 254 significant associations below Bonferroni threshold (p -value = $2.83\text{E-}05$) (Table 1, Fig. 1). All phecode categories harbored multiple significant associations with the tested height PGS; circulatory system (62), congenital anomalies (4), dermatologic (17), digestive (18), endocrine/metabolic (35), genitourinary (8), hematopoietic (11), infectious diseases (6), injuries & poisonings (10), mental disorders (7), musculoskeletal (29), neoplasms (16), neurological (11), respiratory (7), sense organs (6) and symptoms (7) (Supplementary Data 6).

Table 1 | Sample size and number of traits in the cross-ancestry meta-PheWAS

Population	Sample	Number of traits	Significant traits	Bonferroni threshold
Sex-combined	839,872	1768	254	2.83×10^{-5}
Males	471,395	1582	173	3.16×10^{-5}
Females	267,576	1499	56	3.34×10^{-5}

The traits that displayed the strongest associations with height PGS are shown in Table 2. The results from the PheWAS performed in each cohort, along with the full results from the meta-PheWAS, are presented in Supplementary Data 7– Supplementary Data 14 and Supplementary Data 6, respectively.

From the cross-ancestry meta-PheWAS analysis, six traits exhibited evidence of heterogeneity (defined as when the p -value of the Cochran's heterogeneity test is below Bonferroni threshold) as shown in Table 3, Supplementary Figs. 1–7. For example, in *cardiac dysrhythmias* (427) the signal indicated strong evidence for association in European ancestry (p -value = 1.29×10^{-91}) but not in the other ancestral groups (Supplementary Data 4). We also observed evidence of heterogeneity of effects across cohorts for *Cardiac dysrhythmias* (Supplementary Fig. 3). Another notable example of heterogeneity was for *Chronic ulcer of skin* (707) (Supplementary Fig. 5).

Cross-ancestry analyses revealed 30 additional signals that were not present in the European ancestry meta-analyses (Supplementary Data 5, Supplementary Fig. 8).

We further performed sex-specific meta-PheWAS analyses in the UKB, MVP, BioVU and BioMe cohorts (males: Supplementary Data 21, 23, 25, 27 and females: Supplementary Data 32, 34, 36, 38). The meta-PheWAS analysis for males in up to 471,395 participants, interrogated 1582 traits (available in at least two cohorts) and yielded 173 statistically significant trait associations below the Bonferroni threshold ($p\text{-value} = 3.16 \times 10^{-5}$) (Table 1) (Supplementary Data 19). The identified categories included the circulatory system (50), congenital anomalies (4), dermatologic (14), digestive (10), endocrine/metabolic (22), genitourinary (6), hematopoietic

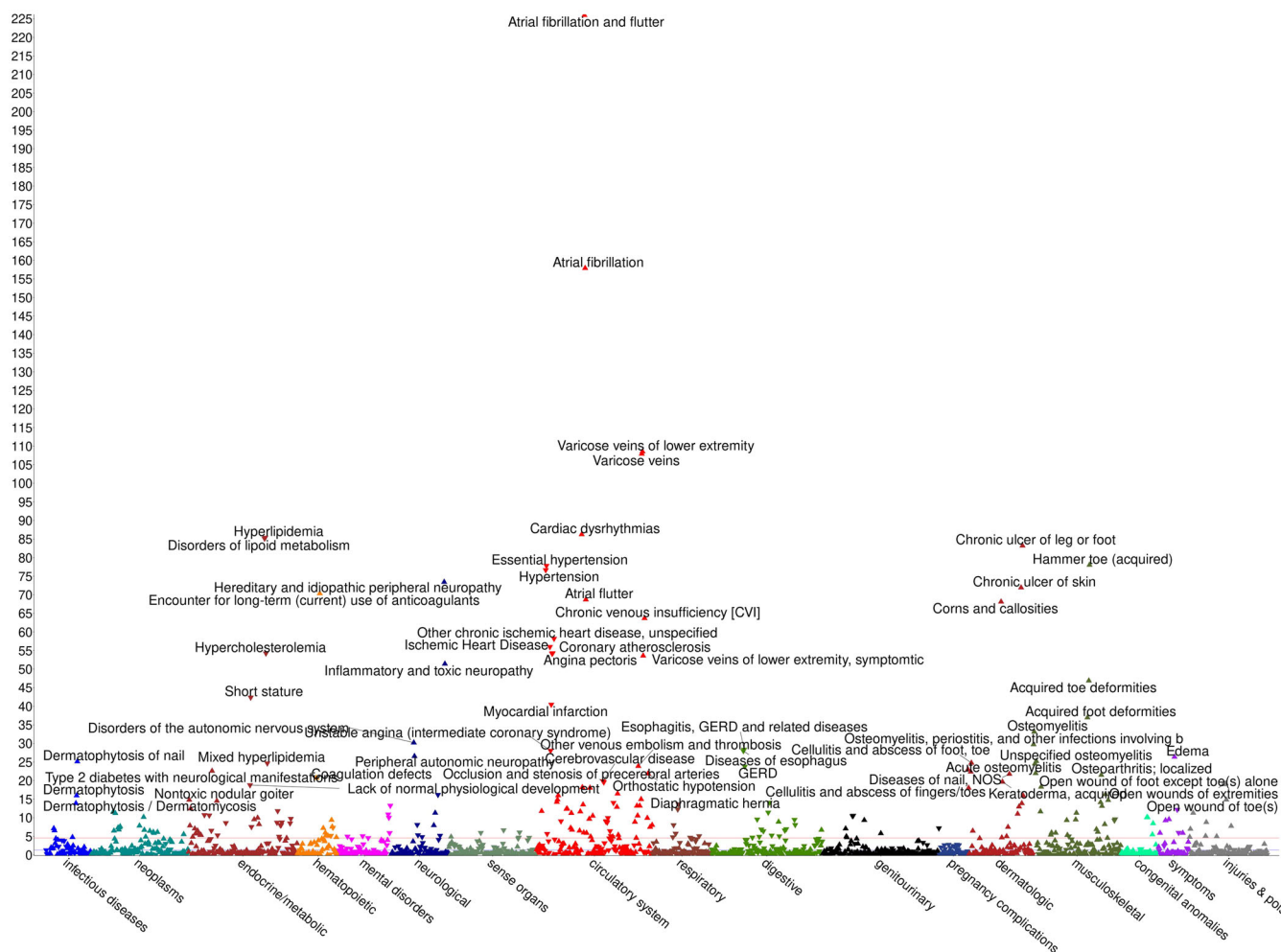


Fig. 1 | Phenome Wide Association Study (PheWAS). Manhattan plot showing the significant phecodes per category for the sex-combined cross-ancestry PheWAS meta-analysis of European (EUR), African (AFR), East Asian (EAS) ancestries and Hispanic (HIS) population groups.

Table 2 | Top 20 significant hits from the sex-combined cross-ancestry PheWAS meta-analysis of EUR, AFR, EAS, HIS

phecode	Description	Category	OR	Lower	Upper	P-value	pval het
427.2	Atrial fibrillation and flutter	Circulatory system	1.16	1.15	1.17	1.08×10^{-226}	4.50×10^{-4}
427.21	Atrial fibrillation	Circulatory system	1.15	1.14	1.17	1.24×10^{-158}	1.27×10^{-3}
454.1	Varicose veins of lower extremity	Circulatory system	1.16	1.15	1.18	2.87×10^{-109}	4.17×10^{-2}
454	Varicose veins	Circulatory system	1.15	1.14	1.17	1.23×10^{-108}	5.85×10^{-2}
427	Cardiac dysrhythmias	Circulatory system	1.07	1.06	1.07	5.68×10^{-87}	3.63×10^{-10}
272.1	Hyperlipidemia	Endocrine/metabolic	0.94	0.94	0.95	4.04×10^{-86}	1.34×10^{-4}
272	Disorders of lipid metabolism	Endocrine/metabolic	0.94	0.94	0.95	7.13×10^{-86}	1.28×10^{-4}
707.2	Chronic ulcer of leg or foot	Dermatologic	1.19	1.17	1.21	6.29×10^{-84}	7.27×10^{-3}
735.21	Hammer toe (acquired)	Musculoskeletal	1.18	1.16	1.20	9.23×10^{-79}	3.12×10^{-3}
401.1	Essential hypertension	Circulatory system	0.95	0.94	0.95	1.51×10^{-78}	1.29×10^{-3}
401	Hypertension	Circulatory system	0.95	0.94	0.95	2.20×10^{-77}	1.31×10^{-3}
356	Hereditary and idiopathic peripheral neuropathy	Neurological	1.15	1.13	1.16	3.71×10^{-74}	1.21×10^{-1}
707	Chronic ulcer of skin	Dermatologic	1.14	1.12	1.16	1.19×10^{-72}	1.05×10^{-5}
286.2	Encounter for long-term (current) use of anticoagulants	Hematopoietic	1.13	1.12	1.15	4.40×10^{-71}	8.36×10^{-4}
427.22	Atrial flutter	Circulatory system	1.17	1.15	1.19	1.96×10^{-69}	1.18×10^{-1}
700	Corns and callosities	Dermatologic	1.15	1.13	1.17	7.51×10^{-69}	2.42×10^{-5}
456	Chronic venous insufficiency [CVI]	Circulatory system	1.16	1.14	1.18	2.04×10^{-64}	6.65×10^{-2}
411.8	Other chronic ischemic heart disease, unspecified	Circulatory system	0.93	0.92	0.94	6.38×10^{-59}	2.80×10^{-1}
411	Ischemic Heart Disease	Circulatory system	0.95	0.94	0.95	1.03×10^{-56}	6.53×10^{-3}
411.4	Coronary atherosclerosis	Circulatory system	0.94	0.94	0.95	5.14×10^{-55}	2.07×10^{-3}

Table 3 | Heterogeneous traits in the cross-ancestry meta-PheWAS analysis

phecode	Description	Category	OR	Lower	Upper	p-value	pval het	I ²
351	Other peripheral nerve disorders	Neurological	0.97	0.96	0.97	6.52×10^{-17}	1.15×10^{-17}	88.3%
427	Cardiac dysrhythmias	Circulatory system	1.07	1.06	1.07	5.68×10^{-87}	3.63×10^{-10}	81.9%
700	Corns and callosities	Dermatologic	1.15	1.13	1.17	7.51×10^{-69}	2.42×10^{-5}	73.2%
707	Chronic ulcer of skin	Dermatologic	1.14	1.12	1.16	1.19×10^{-72}	1.05×10^{-5}	73.3%
735	Acquired foot deformities	Musculoskeletal	1.06	1.05	1.07	1.11×10^{-37}	2.62×10^{-10}	82.1%
735.2	Acquired toe deformities	Musculoskeletal	1.10	1.09	1.11	1.32×10^{-47}	1.93×10^{-5}	72.4%

(6), infectious diseases (5), injuries & poisonings (6), mental disorders (8), musculoskeletal (21), neoplasms (5), neurological (7), respiratory (3), sense organs (3) and symptoms (3). In total, 10 traits, spanning across different categories, were significant only in the male cross-ancestry meta-PheWAS and not in the sex-combined cross-ancestry meta-PheWAS (Supplementary Data 42). For example, increased genetically predicted height was associated with decreased risk of *Hyperpotassemia* (276.13) (OR = 0.95, 95% CI [0.93, 0.97], p -value = 1.23×10^{-6} , het p -value = 9.46×10^{-1}) in males but showed a null association in females (Supplementary Data 42).

Looking at the ancestry level, 10 traits were significant (p -value < 3.16×10^{-5}) only in European ancestry males but not in the male cross-ancestry analysis; 3 of them from the mental disorders category, with increased height PGS having a decreased risk of: *Anxiety disorders* (300), *Posttraumatic stress disorder* (300.9) and *Substance addiction and disorders* (316) (Supplementary Data 16) (Fig. 2). In the other populations, *Drusen (degenerative) of retina* (362.27) and *Fracture of lower limb* (800) were significantly associated with increased height PGS in the male African (decreased risk) and Hispanic (increased risk) populations, respectively, but not in the cross-ancestry analyses (Supplementary Data 17 and 18).

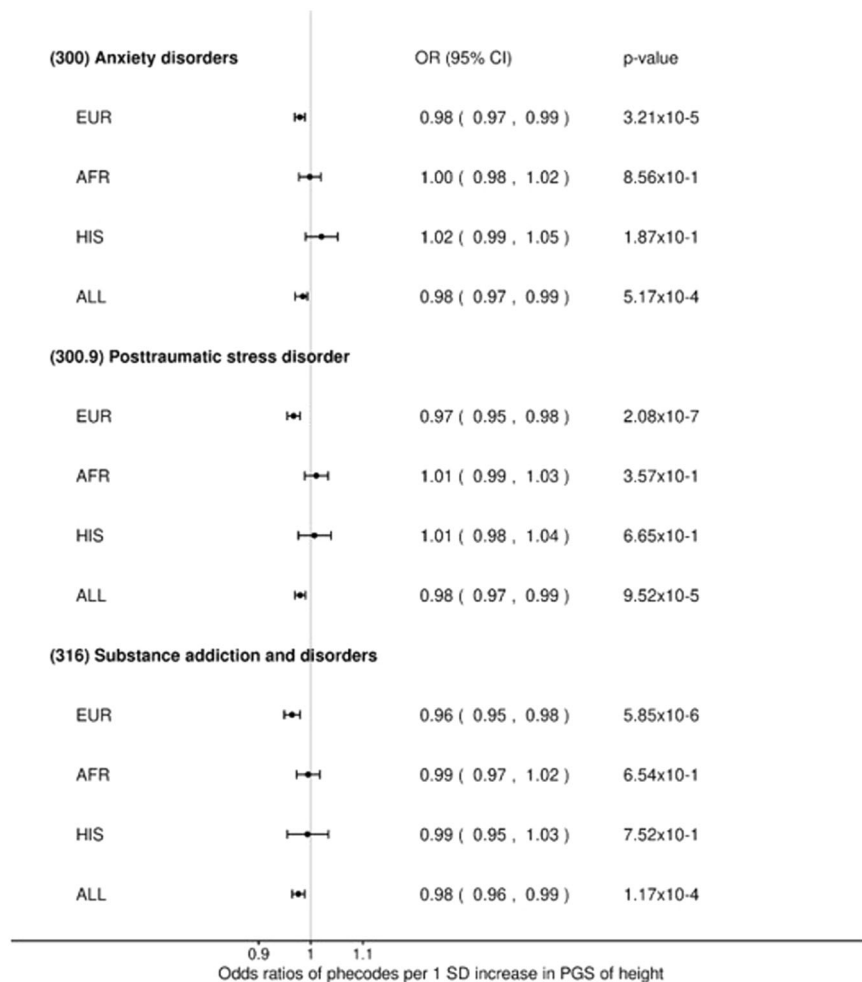
The meta-PheWAS analysis in up to 267,576 female individuals, interrogated 1499 traits (available in at least two cohorts) and yielded 56 significant associations below Bonferroni threshold (p -value = 3.34×10^{-5}) (Table 1) (Supplementary Data 30). The identified categories included the circulatory system (23), dermatologic (2), digestive (7), endocrine/

metabolic (6), genitourinary (1), hematopoietic (1), musculoskeletal (5), neoplasms (8) and neurological (3). Only 1 association identified as significant in the females meta-PheWAS and was not observed in the sex-combined meta-PheWAS; *Benign neoplasm of other parts of digestive system* (211) (OR = 0.95, 95% CI [0.92, 0.97], p -value = 1.53×10^{-5} , het p -value = 5.48×10^{-1}) (Supplementary Data 41). Seven associations were significant in European ancestry but not in the cross-ancestry analysis, mainly from musculoskeletal and infectious diseases categories, such as *Osteoporosis* (743.1) and *Dermatophytosis / Dermatomycosis* (110), respectively (Supplementary Data 29).

Comparing males to females, the meta-PheWAS yielded 126 significant associations only in males, primarily from the circulatory system, endocrine/metabolic and musculoskeletal categories (Supplementary Data 44). Ninety-three percent of the traits had concordant effect sizes and were larger for males. On the other hand, comparing females with males, the meta-PheWAS yielded 13 significant associations in females only, with the digestive and neoplasms categories including the most traits (Supplementary Data 43). Ninety-two percent of the traits were concordant in direction, and the effect sizes in females were larger. Examining the heterogeneity between males and females in the cross-ancestry meta-PheWAS, 7 associations were identified; 4 of them from the musculoskeletal category, such as *Acquired foot deformities* (735) (Supplementary Data 40).

We performed a meta-PheWAS analysis excluding UKB in the cross-ancestry sex-combined meta-PheWAS (Supplementary Data 46) and the

Fig. 2 | Estimates per ancestry in the male meta-analysis of Phenome Wide Association Studies (meta-PheWAS), for signals from the mental disorders category that were identified as significant only in the European males meta PheWAS. SD standard deviation of the mean, PGS polygenic score.



sex-specific ones for males and females (Supplementary Data 50 and 54). For the sex-combined (Supplementary Fig. 9, Supplementary Data 46) and the males (Supplementary Fig. 10, Supplementary Data 50) meta-PheWAS the estimates are concordant as presented in the plots. In the cross-ancestry female meta-PheWAS (Supplementary Fig. 11, Supplementary Data 54) three traits were identified as discordant: Benign neoplasm of other parts of digestive system (211), Other disorders of circulatory system (459), Gastritis and duodenitis (535).

Replication analyses were performed in an independent sample of the Colorado Biobank. Comparing the European ancestry meta-PheWAS with the European PheWAS in Colorado biobank we observe that the ORs are concordant in their majority; Colorado biobank has larger error bars due to the smaller sample size than the meta-PheWAS analysis (Supplementary Figs. 12–14). Colorado biobank also provided PheWAS results using both weighted and unweighted PGS (Supplementary Figs. 15–17).

Discussion

We performed a large ancestrally diverse meta-PheWAS for height in six cohorts including up to 840,000 individuals. Of the 1768 disease traits that were in common across cohorts and were meta-analysed, we identified 254 significant PGS-trait associations (p -value = 2.83×10^{-5}). The largest number and most precise phenotypic associations were observed for the circulatory system, endocrine/metabolic and musculoskeletal categories.

From the circulatory system category, increased genetically predicted height was associated with an increased risk of *Chronic venous insufficiency (CVI)* (456) (OR = 1.16 95% CI [1.14, 1.18], p -value = 2.04×10^{-64}) (Supplementary Data 6), with no evidence of heterogeneity across cohorts (het p -value = 6.65×10^{-3}). These findings were concordant with a recent study in

MVP which reported an association between increased genetically predicted height and increased risk of CVI in European American (EA) (OR = 1.366, p -value = 1.6×10^{-35}) and in African American (AA) individuals (OR = 1.469, p -value = 3.1×10^{-4})⁸. The effect was similar in both males and females in our analyses. Failure of the femoral vein valves may lead to CVI, with severe consequences. However, for the valves to be replaced, the femoral vein diameter (FVD) must be known. A recent study by Keiler et al.¹⁷ reported that height was positively correlated with FVD; this correlation was attenuated when the sample was stratified by sex. In addition, failure of the venous valve can lead to varicose veins¹⁷. In our study, increased genetically predicted height was associated with increased risk of *Varicose veins (VV)* (454) (OR = 1.15, 95% CI [1.14, 1.17], p -value = 1.23×10^{-108}) (Supplementary Data 6), with no evidence of heterogeneity across cohorts (het p -value = 5.85×10^{-2}), again a finding in agreement with the MVP PheWAS⁸. Moreover, Mendelian Randomisation (MR) studies in European ancestry have supported a causal association between genetically predicted height and VV^{18,19}.

Within the circulatory system category, the strongest association was for *Atrial fibrillation and flutter (AF)* (427.2) (OR = 1.16, 95% CI [1.15, 1.17], p -value = 1.08×10^{-226}) (Supplementary Data 6), with no evidence of heterogeneity across cohorts (het p -value = 4.50×10^{-4}), and similar effect sizes in the sex-stratified meta-PheWAS. The aforementioned MVP study similarly reported an increased risk of AF in EA (OR = 1.381, p -value = 5.70×10^{-84}) and in AA (OR = 1.352, p -value = 3.3×10^{-4})⁸. Significant causal associations from MR analysis have been reported in two previous studies^{6,20}.

Our study confirmed that increased genetically predicted height is inversely associated with cardiovascular diseases^{21–23}. Increased genetically

predicted height was associated with decreased risk of *hypertension* (401) (OR = 0.950, 95% CI [0.944, 0.955], p -value = 2.20×10^{-77}) (Supplementary Data 6), with no evidence of heterogeneity across cohorts (het p -value = 1.31×10^{-3}), and with similar effect sizes in males and females. This finding is in accordance with previous studies, although our effect sizes were slightly attenuated, possibly due to a lack of coding “hypertension” using ICD codes^{8,24}. According to World Health Organisation (WHO) “hypertension is diagnosed if, when it is measured on two different days, the systolic blood pressure readings on both days is ≥ 140 mmHg and/or the diastolic blood pressure readings on both days is ≥ 90 mmHg”²⁵. A study in the Finnish population examining blood pressure found that shorter participants had higher SBP than taller ones, and this could be partially the reason for observing inverse association between height and cardiovascular disease²¹. A study in the USA reported that height was inversely associated with DBP in older males and females, in contrast to SBP that was positively associated²². A recent systematic review concluded that there was a potentially inverse association of stature and BP²⁶. An MR analysis conducted in European ancestry individuals showed that an increase in adult height was causally associated with a lower risk of coronary heart disease, with one potential mechanism including BP²⁷.

Epidemiological and genetic studies suggest that increased height is associated with decreased risk of CAD^{6,23,28}. In a meta-analysis of European ancestry participants, genetically predicted increased height was associated with decreased risk of CAD (OR = 0.88, 95% CI [0.82, 0.95], p -value $< 1.00 \times 10^{-3}$)²⁸. Similar findings were reported in several MR studies^{6,23}. CAD is a broad category including diseases such as ischemic heart disease, myocardial infarction and coronary atherosclerosis. For instance, *Ischemic heart disease* (411) (OR = 0.948, 95%CI [0.942, 0.954], p -value = 1.03×10^{-56} , het p -value = 6.53×10^{-3}), and *Myocardial infarction* (MI) (411.2) (OR = 0.93, 95%CI [0.92, 0.94], p -value = 3.54×10^{-41} , het p -value = 1.32×10^{-1}) (Supplementary Data 6) were identified as significant among the cardiovascular diseases and with similar effect at the sex-stratified meta-PheWAS; all these have been confirmed in previous studies^{8,29}.

In the endocrine/metabolic category, several health-related outcomes were identified. Our study identified decreased risk of *Hyperlipidemia* (272.1) (OR = 0.942, 95% CI [0.936, 0.947], p -value = 4.04×10^{-86} , het p -value = 1.34×10^{-4}) and *Hypercholesterolemia* (272.11) (OR = 0.946, 95% CI [0.939, 0.953], p -value = 5.67×10^{-55} , het p -value = 4.17×10^{-2}) (Supplementary Data 6), with similar effect at the sex-stratified meta-PheWAS. These findings have also been reported by MVP⁸, and in a Korean population^{30,31}. Our meta-analysis confirmed the well-established association between 1 SD increase in genetically predicted height and decreased risk of *Type 2 diabetes* (T2D) (250.2) (OR = 0.98, 95% CI [0.97, 0.99], p -value = 2.27×10^{-11} , het p -value = 9.11×10^{-3}) (Supplementary Data 6)^{32,33}. In addition, we observed an association between increased genetically predicted height and the increased risk of *Hypothyroidism* (244) (OR = 1.022, 95% CI [1.014, 1.031], p -value = 8.58×10^{-8} , het p -value = 3.30×10^{-1}) (Supplementary Data 6). This is an interesting insight towards the known epidemiological links between hypothalamic-pituitary-thyroid (HPT) axis dysregulation and stature³⁴.

Several health outcomes from the musculoskeletal category were associated with genetically predicted height. *Acquired foot deformities* (735) (OR = 1.06, 95% CI [1.05, 1.07], p -value = 1.11×10^{-37}) were associated with higher genetically predicted height, with strong evidence of heterogeneity across cohorts (het p -value = 2.62×10^{-10}) (Supplementary Data 6). In the present study, EA descent individuals presented the strongest signal in MVP, followed by eMERGE and in AA only in MVP (Supplementary Fig. 6). We found this association in males only, which is supported by a previous study reporting foot deformities to be significantly more prevalent in male veterans versus male non-veterans in USA³⁵. In contrast, *Osteoarthritis; localized* (740.1) (OR = 1.033, 95% CI [1.026, 1.039], p -value = 3.13×10^{-22} , het p -value = 4.88×10^{-2}) (Supplementary Data 6) was found to have a similar effect in both males and females. This finding is supported by the MVP PheWAS⁸ and is widely supported in the epidemiological literature,

that taller individuals have an increased risk of knee osteoarthritis, that remained significant for both sexes, after adjusting for confounders³⁶. A recent meta-analysis of GWAS studies for osteoarthritis, in Icelanders and European ancestry from UKB, found that a large proportion of osteoarthritis risk variants are associated with height³⁷.

We identified several notable associations in the neoplasms category. There has been a significant body of literature studying the association between height and risk of breast cancer (BC) and the results are controversial. Several PheWAS and MR studies reported null associations between height PGS and BC^{38,39}. In contrast, several studies, including ours, confirm the association of height and risk of BC. An observational study, using data from EPIC and the Women’s Health Initiative (WHI) in the USA, observed that for every 10 cm increase in height there was an 18% increased risk of ER + BC; null association was found for ER- BC⁴⁰. Another observational study, analysing post-menopausal women from the Netherlands Cohort Study (1986–2006), observed that for every 5 cm increase in height there was a 7% increased risk of BC (95% CI: 1.01–1.13); an association that remained significant for the ER + BC but not for ER- BC⁴¹.

We observed an attenuated, non-significant association, between increased genetically predicted height and *Colorectal cancer* (153) (OR = 1.02, 95% CI [1.00, 1.04], p -value = 2.19×10^{-2} , het p -value = 6.89×10^{-1}) (Supplementary Data 6). This finding contrasts with the majority of PheWAS and MR studies that describe an association between increased adult height and increased risk of colorectal cancer^{11,12,42}.

We identified a significant association between increased genetically predicted height and decreased risk of *Hyperpotassemia* (276.13) (OR = 0.95, 95% CI [0.93, 0.97], p -value = 1.23×10^{-6} , het p -value = 9.46×10^{-1}) (Supplementary Data 19) in males. Additionally, increased genetically predicted height was associated with 3 traits from the mental disorders category in the males meta-PheWAS: *Pervasive developmental disorders* (313) (OR = 1.06, 95% CI [1.03, 1.09], p -value = 6.11×10^{-6} , het p -value = 5.25×10^{-3}), *Attention deficit hyperactivity disorder* (ADHD) (313.1) (OR = 1.06, 95% CI [1.03, 1.09], p -value = 2.35×10^{-5} , het p -value = 3.93×10^{-1}) and *Autism* (313.3) (OR = 1.215, 95% CI [1.222, 1.316], p -value = 1.64×10^{-6} , het p -value = 2.58×10^{-1}) (Supplementary Data 19). Similarly, the traits were concordant in the sex-combined meta-PheWAS but showed null association in the female meta-PheWAS. Previous PheWAS provided suggestive support of these findings, with the exception of autism⁸. A study by Yackobovitch-Gavan et al.⁴³ employing data from Israel Clalit Health Services, reported that drug treatment for ADHD was associated with greater decline of height z-score in boys than girls, with 66% of the participants being boys. Additionally two studies in the US, one for children⁴⁴, and one for both children and adolescents⁴⁵, confirmed a decline of height z-scores for patients using stimulants and it is confirmed by a study in Netherlands⁴⁶. However, these studies have examined the case in which the participants are medicated. Nevertheless, there is evidence suggesting that there are more males diagnosed than females, which is in accordance with our results. Our results relating to autism are in accordance with the literature; a study in Spanish pre-school children showed that autism spectrum disorder (ASD) had increased height in contrast to children with typical development⁴⁷. Additionally, in Australia, male babies with ASD were born smaller, but grew taller in comparison to children with typical development⁴⁸. Therefore, for these disease traits, it seems that males drive the association.

In males, 10 phenotypes displayed significant associations with height PGS in European descent individuals only; 3 of them belong to the mental disorders category: *Anxiety disorders* (300) (OR = 0.98, 95% CI [0.97, 0.99], p -value = 3.21×10^{-5} , het p -value = 1.40×10^{-1}), *Posttraumatic stress disorder disorders* (300.9) (OR = 0.97, 95% CI [0.96, 0.98], p -value = 2.08×10^{-7} , het p -value = 2.86×10^{-1}) and *Substance addiction and disorders* (316) (OR = 0.96, 95% CI [0.95, 0.98], p -value = 5.85×10^{-6} , het p -value = 9.10×10^{-1}) (Supplementary Data 16).

Seven traits were identified as significant in the female meta-PheWAS for European descent individuals and not in the cross-ancestry analysis. Amongst them, increased height PGS was associated with decreased risk of

Osteoporosis (743.1) (OR = 0.93, 95% CI [0.90, 0.96], p -value = 6.58×10^{-6} , $\text{het } p$ -value = 9.65×10^{-1}) (Supplementary Data 29). Post-menopausal European ancestry females had an increased risk of osteoporotic fractures, in contrast to African and Asian ancestries^{49,50}.

Our study had several important limitations. Although we used the recently published cross-ancestry GWAS from GIANT, the study populations were predominantly of European ancestry. Thus, we observed a poorer prediction performance of the height PGS in our study populations that were ancestrally diverse, diminishing the power in populations with substantial non-European admixture. It is also possible that some of the signals observed may be driven by differences in phenotype prevalence across cohorts. The differences in sample size by sex and ancestry complicate interpretation of differences across these strata. This limitation is not new for genetic studies but likely limits our inference on true sex and ancestry differences in the phenotype associations with genetically predicted height at phenome-wide significance. We included all available cohort data as a discovery meta-analysis to increase power. Trait associations with genetically predicted height may be particularly influenced by indirect genetic effects and assortative mating. A recent study showed that population estimates are larger than within-sibship meta-analysis GWAS estimates for height⁵¹. The authors presented strong evidence of polygenic adaptation on taller height in European ancestry individuals, suggesting that demographic effects, such as assortative mating, could vary between populations^{51,52}. Additionally, previous work in the UK Biobank has reported an association between stature and socio-economic status in both sexes, therefore this could serve as a mediator of the reported associations rather than the actual direct effect of height⁶. Lastly, we did not consider obvious reasons for differences across studies, sex, and ancestry. Social factors have a powerful influence on many of the phenotype-genetically predicted height associations described herein. By including data from diverse populations in future investigation of the role of genetically predicted height across the phenome, future research might be able to address the limitations of this study. It might be possible to better understand the genetic and environmental factors that affect height by more broadly interpreting the results. The study's power would be enhanced, and more precise results would be produced by expanding the sample size and providing more in-depth information on lifestyle factors. Finding associations between height and disease using data from different ancestries would improve the generalizability of our findings and offer a more thorough understanding of the genetic and environmental factors affecting height and disease risk. Additional approaches could include carrying out population-specific studies, which would enable the investigation of height-disease relationships in particular ethnic groups. This could be accomplished by enlisting volunteers from particular ethnic groups and gathering thorough data on disease outcomes, height, and other pertinent covariates like lifestyle variables. In the process of creating new treatments and preventative measures for a variety of diseases, this could assist in the identification of novel genetic variants and pathways.

Methods

PheWAS is used to identify the effects of genetic variation already associated with a trait of interest across a larger array of phenotypes, using a hypothesis-free approach, with no prior assumptions⁵³. We employed Bonferroni correction to determine statistical significance. Despite this, our large sample size facilitated the replication of known associations and even the discovery of new ones⁵⁴.

To assess the associations of the PGS with hospital-record data, we used the PheWAS library⁵³ implemented in R⁵⁵. The package converts International Classification of Diseases (ICD) codes to 'PheWAS codes' or phecodes, which represent 1866 phenotypes in total formed from grouped ICD codes using the "Phecode Map 1.2 ICD-10-CM" (https://phewascatalog.org/phecodes_icd10cm). Each phenotype case is accompanied by accurate controls, meaning participants who have similar disease with the phenotype case are excluded. For instance, if the phenotype case under investigation is T2D, then participants who have T1D are excluded from the control group. This built-in exclusion feature, that prevents contamination of the controls,

is essential to preserve statistical power to identify associations^{53,56}. The phecodes are divided in 17 distinct categories: circulatory system, endocrine/metabolic, mental disorders, neurological, respiratory, infectious diseases, neoplasms, hematopoietic, sense organs, digestive, genitourinary, pregnancy complications, dermatologic, musculoskeletal, congenital anomalies, symptoms and injuries & poisonings⁵⁶. Next, binary logistic regression models are employed to examine the association of the exposure, the PGS of height (independent variable), with the trait of interest with each phecode. As covariate adjustments in each study population, we used age, sex, genotype batch, to reduce model variability. Each study population (described in Supplementary Information) also adjusted for principal components for ancestry to control for confounding via population stratification (details per study on ancestry determination and exclusion in Supplementary Data 1A).

Details regarding compliance with all relevant ethical regulations including the Declaration of Helsinki can be found in the information and references for each participating cohort below. The PheWAS and meta-PheWAS summary statistics results that are discussed in the manuscript are included in the Supplementary Data 3–57.

Polygenic score

We performed a conditional and joint analysis (GCTA-COJO) to select quasi-independent height-associated SNPs for the construction of the PGS^{57,58}. A stepwise procedure was used for SNP selection and the joint effects of all selected SNPs were estimated after the model was optimized. The genetic variants are still genome-wide significant, independent and the variance explained by them is larger than considering only the leading SNP at each locus. This conditional analysis was performed in the recent cross-ancestry GWAS for adult height, excluding data from 23andMe¹⁴, using 50,000 unrelated and randomly sampled European participants of UKB as the LD reference panel. We performed analyses using p -value threshold $p = 5 \times 10^{-9}$ to declare a genome-wide significant hit. Also, SNPs with allele frequency differences larger than 0.2 as compared to a UKB reference panel, were excluded from the analysis along with SNPs having MAF ≤ 0.001 . The GCTA-COJO analysis resulted in a list of 6797 SNPs. As covariate adjustments we used age, sex, genotype batch, to reduce model variability. We also adjusted for principal components for ancestry to control for confounding via population stratification (Supplementary Data 1B). The PGS of height was constructed as the unweighted sum of the height-increasing alleles within each study (Supplementary Data 1) and afterwards is scaled (using scale function in R).

All herein reported ORs are per one standard deviation increase in PGS.

Meta-analysis

Meta-analysis is a popular statistical technique used to increase the power to detect new effects by combining the information from independent studies. In addition, heterogeneity among the studies can be assessed, employing the beta estimates and standard errors from each study. For a small number of similar studies, the most common technique is the fixed-effect inverse variance weighted meta-analysis, which uses as a hypothesis that a common underlying effect exists for all studies^{59,60}. We performed a meta-PheWAS, combined in a fixed-effect meta-analysis for UKB, MVP, BioVU, BioMe, MyCode and eMERGE cohorts, using the phecodes derived from the PheWAS in each cohort (Supplementary Data 3–Supplementary Data 14). For the sex-specific analysis, we employed data from the UKB, MVP, BioVU and BioMe (Supplementary Data 15–38). The examined ancestries were European, African, East Asian ancestries and Hispanic population groups, and the sample size per ancestry and per study are included in Table 4. For more details the reader is referred to Supplementary Data 1 and 2.

The sample size and examined number of traits for the sex-combined and sex-specific cross-ancestry meta-PheWAS are detailed in Table 1, and for the specific ancestries in Supplementary Data 2. For the meta-analysis we employed the statistical software R 3.6.1 and the library *metafor*⁶¹.

Table 4 | Sample size per ancestry and per study

	EUR	AFR	EAS	HIS
BioMe	7993	6224		9007
BioVU	71,986	15,593		
Emerge	40,088			
MVP	221,927	57,974	5069	26,845
MyCode	46,922			
UKB	330,965	6682	1936	

Replication

Replication analyses were performed in an independent sample of the Colorado Biobank. We also performed a replication PheWAS in the same biobank using a score weighted for the effects of the height-associated SNPs in the GWAS meta-analysis. Details are provided in the Supplementary Material.

UK Biobank (UKB). The UKB is a prospective cohort of 502,504 participants, aged 40–69 years old, who were recruited between 2006 and 2010. The cohort includes information regarding a variety of phenotypes like blood measurements, clinical assessments, anthropometry, cognitive function, hearing, arterial stiffness, hand grip strength, spirometry, ECG, data on cancer and death registries, health and lifestyle medical conditions, operations, mental health, sociodemographic factors, lifestyle, family history, psychosocial factors and dietary intake, described in more detail elsewhere⁶². Hospital episode statistics (HES) is a database containing details of all admissions at NHS hospitals in UK, which has been linked to the UKB⁶³.

Million Veteran Program (MVP). The Department of Veterans Affairs (VA) created in 2011 a national cohort across USA: the MVP. This cohort was created as a representative, national and longitudinal study of Veterans for genomic and non-genomic research, employing responses to questionnaires, blood specimens and electronic health records (EHR). The blood specimens were collected for genotyping, and these were linked to the EHR, which coded the diagnosis in ICD9 and ICD10, up until September 2019. As expected, most of the participants are males, aged between 50 and 69 years old at recruitment. Regarding ethnicity, European Americans and African Americans are well represented; Hispanics and Asian descent participants are also included⁶⁴.

The MVP study from Raghavan et al.⁸ uses different sample than the one we are using in the current study.

BioVU. The Vanderbilt Institutional Review Board (IRB) approved the creation of Vanderbilt DNA databank, that collected DNA samples from 2007 until 2010. During the past years, the Vanderbilt University Medical Center has developed a comprehensive electronic medical record (EMR) system that covers all inpatient and outpatient data, including labs, drug ordering, and diagnostic imaging, including over 1.4 million records⁶⁵. Regarding ethnicity, there is large concordance between race assignment and genetic ancestry for Europeans and African Americans, in contrast to lower concordance for Hispanics, East Asians and South Asians⁶⁶.

BioMe. The Icahn School of Medicine at Mount Sinai's Institutional Review Board approved in 2007 the construction of BioMe biobank. This EMR-linked biorepository enrolls participants non-selectively from the Mount Sinai Health System, which serves a diverse group of communities across the greater New York City area. At enrolment, participants provided informed consent to link their DNA and plasma sample to their EMR. This is further complemented by a questionnaire on demographic and lifestyle factors. At present, the cohorts comprise over 60,000 participants. 58% of the participants are females; participants were aged between 18 and 89+ years old at recruitment. Regarding ethnicity,

European Americans, African Americans and Hispanics are well represented⁶⁷.

Geisinger's MyCode Community Health Initiative Study (MyCode). The Geisinger Health System (GHS) includes a large percentage of stable participants from Pennsylvania, from more than 70 care facilities. In 2007 GHS initiated the MyCode Community Health Initiative (MyCode) to create a biobank of blood, serum, DNA samples along with genotype and exome sequence data. These data were linked to the EMR data for research purposes. By 2015, MyCode reported more than 90,000 participants and an ongoing monthly enrolment of around 2000, across the age spectrum (0 to >89 years old). Regarding ethnicity, more than 95% of the population are self-identified white or European American⁶⁸.

Electronic Medical Records and Genomics (eMERGE) network. In 2007 the electronic Medical Records and Genomics (eMERGE) Network is a National Human Genome Research Institute (NHGRI) created to employ EHR for genomic research purposes. Today, eMERGE Network includes nine research groups across US, that they have connected the DNA samples to EHR. The majority of the studied participants have European ancestry, but also African, Asian and Hispanic descent participants are included in a smaller percent^{69,70}.

Colorado Center for Personalized Medicine (CCPM Biobank). The biobank at the Colorado Center for Personalized Medicine (CCPM Biobank) was jointly developed by the University of Colorado Anschutz Medical Campus and UCHHealth to serve as a unique, dual-purpose research and clinical resource accelerating personalized medicine. As a resource comprising electronic health records (EHRs), genotype data, and other integrated data sources (e.g., geocoded data and survey data), the CCPM Biobank had more than 200,000 enrolled participants and 33,674 genotyped participants as of March 2022. The latter formed the freeze 2 research dataset. More details about the CCPM Biobank are described in Wiley et al.⁷¹.

Data availability

Individual level data could be accessed upon request and approval from the respective biobanks. The summary statistics results that are discussed in the manuscript are included in the Supplementary Data.

Code availability

All code for the analysis employed in this paper is publicly available (GCTA-COJO version: 1.93.0., PheWAS and metafor in R 3.6.1.).

Received: 14 June 2023; Accepted: 20 January 2025;

Published online: 27 February 2025

References

1. Tachmazidou, I. et al. Whole-genome sequencing coupled to imputation discovers genetic signals for anthropometric traits. *Am. J. Hum. Genet.* **100**, 865–884 (2017).
2. Yengo, L. et al. Meta-analysis of genome-wide association studies for height and body mass index in approximately 700,000 individuals of European ancestry. *Hum. Mol. Genet.* **27**, 3641–3649 (2018).
3. Gudbjartsson, D. F. et al. Many sequence variants affecting diversity of adult human height. *Nat. Genet.* **40**, 609–615 (2008).
4. Zhou, W. et al. Sex-specific relationship between adult height and the risk of stroke: a dose-response meta-analysis of prospective studies. *J. Clin. Hypertens.* **21**, 262–270 (2019).
5. Perkins, J. M., Subramanian, S. V., Davey Smith, G. & Ozaltin, E. Adult height, nutrition, and population health. *Nutr. Rev.* **74**, 149–165 (2016).
6. Lai, F. Y. et al. Adult height and risk of 50 diseases: a combined epidemiological and genetic analysis. *BMC Med.* **16**, 187 (2018).
7. Shrestha, S. et al. Associations between adult height and type 2 diabetes mellitus: a systematic review and meta-analysis of

- observational studies. *J. Epidemiol. Community Health* **73**, 681–688 (2019).
8. Raghavan, S. et al. A multi-population phenome-wide association study of genetically-predicted height in the Million Veteran Program. *PLoS Genet.* **18**, e1010193 (2022).
9. Kofler, T. et al. Relationships of measured and genetically determined height with the cardiac conduction system in healthy adults. *Circ. Arrhythm Electrophysiol.* **10**, <https://doi.org/10.1161/CIRCEP.116.004735> (2017).
10. Roetker, N. S. et al. Taller height as a risk factor for venous thromboembolism: a Mendelian randomization meta-analysis. *J. Thromb. Haemost.* **15**, 1334–1343 (2017).
11. Vithayathil, M. et al. Body size and composition and risk of site-specific cancers in the UK Biobank and large international consortia: a Mendelian randomisation study. *PLoS Med.* **18**, e1003706 (2021).
12. Cornish, A. J. et al. Modifiable pathways for colorectal cancer: a Mendelian randomisation analysis. *Lancet Gastroenterol. Hepatol.* **5**, 55–62 (2020).
13. Wood, A. R. et al. Defining the role of common variation in the genomic and biological architecture of adult human height. *Nat. Genet.* **46**, 1173–1186 (2014).
14. Yengo, L. et al. A saturated map of common genetic variants associated with human height. *Nature* **610**, 704–712 (2022).
15. Shen, X. et al. A phenome-wide association and Mendelian randomisation study of polygenic risk for depression in UK Biobank. *Nat. Commun.* **11**, 2301 (2020).
16. Millard, L. A. C., Davies, N. M., Gaunt, T. R., Davey Smith, G. & Tilling, K. Software application profile: PHESANT: a tool for performing automated phenome scans in UK Biobank. *Int J. Epidemiol.* **47**, 29–35 (2018).
17. Keiler, J., Schulze, M., Claassen, H. & Wree, A. Human femoral vein diameter and topography of valves and tributaries: a post mortem analysis. *Clin. Anat.* **31**, 1065–1076 (2018).
18. Yuan, S., Bruzelius, M., Damrauer, S. M. & Larsson, S. C. Cardiometabolic, lifestyle, and nutritional factors in relation to varicose veins: a Mendelian randomization study. *J. Am. Heart Assoc.* **10**, e022286 (2021).
19. Fukaya, E. et al. Clinical and genetic determinants of varicose veins. *Circulation* **138**, 2869–2880 (2018).
20. Levin, M. G. et al. Genetics of height and risk of atrial fibrillation: a Mendelian randomization study. *PLoS Med.* **17**, e1003288 (2020).
21. Korhonen, P. E., Kautiainen, H. & Eriksson, J. G. The shorter the person, the higher the blood pressure: a birth cohort study. *J. Hypertens.* **35**, 1170–1177 (2017).
22. Bourgeois, B. et al. Associations between height and blood pressure in the United States population. *Medicine* **96**, e9233 (2017).
23. Nikpay, M. & Mohammadzadeh, S. Phenome-wide screening for traits causally associated with the risk of coronary artery disease. *J. Hum. Genet.* **65**, 371–380 (2020).
24. Nguyen, X. T. et al. Baseline characterization and annual trends of body mass index for a mega-biobank cohort of US Veterans 2011–2017. *J. Health Res. Rev. Dev. Ctries* **5**, 98–107 (2018).
25. Organization, W. H. *Hypertension*, <https://www.who.int/news-room/fact-sheets/detail/hypertension> (2022).
26. Cochran, J. M. et al. The relationship between adult height and blood pressure. *Cardiology* **146**, 345–350 (2021).
27. Nuesch, E. et al. Adult height, coronary heart disease and stroke: a multi-locus Mendelian randomization meta-analysis. *Int J. Epidemiol.* **45**, 1927–1937 (2016).
28. Nelson, C. P. et al. Genetically determined height and coronary artery disease. *N. Engl. J. Med.* **372**, 1608–1618 (2015).
29. Schmidt, M., Botker, H. E., Pedersen, L. & Sorensen, H. T. Adult height and risk of ischemic heart disease, atrial fibrillation, stroke, venous thromboembolism, and premature death: a population based 36-year follow-up study. *Eur. J. Epidemiol.* **29**, 111–118 (2014).
30. Lee, M. Y. et al. Association between height and hypercholesterolemia in adults: a nationwide population-based study in Korea. *Lipids Health Dis.* **18**, 198 (2019).
31. Oh, N. K., Song, Y. M., Kim, S. H. & Park, M. J. Short stature is associated with increased risk of dyslipidemia in Korean adolescents and adults. *Sci. Rep.* **9**, 14090 (2019).
32. Wittenbecher, C., Kuxhaus, O., Boeing, H., Stefan, N. & Schulze, M. B. Associations of short stature and components of height with incidence of type 2 diabetes: mediating effects of cardiometabolic risk factors. *Diabetologia* **62**, 2211–2221 (2019).
33. Yuan, S. & Larsson, S. C. An atlas on risk factors for type 2 diabetes: a wide-angled Mendelian randomisation study. *Diabetologia* **63**, 2359–2371 (2020).
34. Rivkees, S. A., Bode, H. H. & Crawford, J. D. Long-term growth in juvenile acquired hypothyroidism: the failure to achieve normal adult stature. *N. Engl. J. Med.* **318**, 599–602 (1988).
35. Shibuya, N., Jupiter, D. C., Ciliberti, L. J. Jr., VanBuren, V. & La Fontaine, J. Prevalence of podiatric medical problems in veterans versus nonveterans. *J. Am. Podiatr. Med Assoc.* **101**, 323–330 (2011).
36. Welling, M. et al. Association between height and osteoarthritis of the knee and hip: the Northern Finland Birth Cohort 1966 Study. *Int. J. Rheum. Dis.* **20**, 1095–1104 (2017).
37. Styrkarsdottir, U. et al. Meta-analysis of Icelandic and UK data sets identifies missense variants in SMO, IL11, COL11A1 and 13 more new loci associated with osteoarthritis. *Nat. Genet.* **50**, 1681–1687 (2018).
38. Escala-Garcia, M. et al. Breast cancer risk factors and their effects on survival: a Mendelian randomisation study. *BMC Med.* **18**, 327 (2020).
39. Khalis, M. et al. Body size, silhouette trajectory and the risk of breast cancer in a Moroccan case-control study. *Breast Cancer* **27**, 748–758 (2020).
40. Li, K. et al. Risk prediction for estrogen receptor-specific breast cancers in two large prospective cohorts. *Breast Cancer Res.* **20**, 147 (2018).
41. Elands, R. J. J. et al. Associations of adult-attained height and early life energy restriction with postmenopausal breast cancer risk according to estrogen and progesterone receptor status. *Int. J. Cancer* **144**, 1844–1857 (2019).
42. Clinton, S. K., Giovannucci, E. L. & Hursting, S. D. The World Cancer Research Fund/American Institute for cancer research third expert report on diet, nutrition, physical activity, and cancer: impact and future directions. *J. Nutr.* **150**, 663–671 (2020).
43. Yackobovitch-Gavan, M. et al. Sex-specific long-term height and body mass index trajectories of children diagnosed with attention-deficit/hyperactivity disorder and treated with stimulants. *J. Pediatr.* **238**, 296–304.e294 (2021).
44. Davallow Ghajar, L. & DeBoer, M. D. Children with attention-deficit/hyperactivity disorder are at increased risk for slowed growth and short stature in early childhood. *Clin. Pediatr.* **59**, 401–410 (2020).
45. Schneider, G. et al. Weight and height in children and adolescents with attention-deficit/hyperactivity disorder: a longitudinal database study assessing the impact of guanfacine, stimulants, and no pharmacotherapy. *J. Child Adolesc. Psychopharmacol.* **29**, 285–304 (2019).
46. Lentferink, Y. E., van de Garde, E. M. W., Knibbe, C. A. J. & van der Vorst, M. M. J. Psychostimulants: influence on body mass index and height in a pediatric population with attention-deficit/hyperactivity disorder? *J. Child Adolesc. Psychopharmacol.* **28**, 530–536 (2018).
47. Esteban-Figuerola, P., Morales-Hidalgo, P., Arijia-Val, V. & Canals-Sans, J. Are there anthropometric and body composition differences between children with autism spectrum disorder and children with typical development? Analysis by age and spectrum severity in a school population. *Autism* **25**, 1307–1320 (2021).

48. Green, C. C., Dissanayake, C., Loesch, D. Z., Bui, M. & Barbaro, J. Skeletal growth dysregulation in Australian male infants and toddlers with autism spectrum disorder. *Autism Res.* **11**, 846–856 (2018).
49. Cauley, J. A. Defining ethnic and racial differences in osteoporosis and fragility fractures. *Clin. Orthop. Relat. Res.* **469**, 1891–1899 (2011).
50. Barrett-Connor, E. et al. Osteoporosis and fracture risk in women of different ethnic groups. *J. Bone Min. Res.* **20**, 185–194 (2005).
51. Howe, L. J. et al. Within-sibship genome-wide association analyses decrease bias in estimates of direct genetic effects. *Nat. Genet.* **54**, 581–592 (2022).
52. Border, R. et al. Cross-trait assortative mating is widespread and inflates genetic correlation estimates. *Science* **378**, 754–761 (2022).
53. Carroll, R. J., Bastarache, L. & Denny, J. C. R PheWAS: data analysis and plotting tools for phenome-wide association studies in the R environment. *Bioinformatics* **30**, 2375–2376 (2014).
54. Millard, L. A. C., Davies, N. M., Tilling, K., Gaunt, T. R. & Davey Smith, G. Searching for the causal effects of body mass index in over 300 000 participants in UK Biobank, using Mendelian randomization. *PLoS Genet.* **15**, e1007951 (2019).
55. Team, R. C. R: A Language and Environment for Statistical Computing, <https://www.R-project.org/> (2019).
56. Wu, P. et al. Mapping ICD-10 and ICD-10-CM codes to phecodes: workflow development and initial evaluation. *JMIR Med. Inf.* **7**, e14325 (2019).
57. Yang, J. et al. Conditional and joint multiple-SNP analysis of GWAS summary statistics identifies additional variants influencing complex traits. *Nat. Genet.* **44**, 369–375 (2012).
58. Yang, J., Lee, S. H., Goddard, M. E. & Visscher, P. M. GCTA: a tool for genome-wide complex trait analysis. *Am. J. Hum. Genet.* **88**, 76–82 (2011).
59. Cheung, M. W. & Vijayakumar, R. A guide to conducting a meta-analysis. *Neuropsychol. Rev.* **26**, 121–128 (2016).
60. Michailidou, K. Meta-analysis of common and rare variants. *Methods Mol. Biol.* **1793**, 73–88 (2018).
61. Viechtbauer, W. Conducting meta-analyses in R with the metafor package. *J. Stat. Softw.* **36**, 1–48 (2010).
62. Bycroft, C. et al. The UK Biobank resource with deep phenotyping and genomic data. *Nature* **562**, 203–209 (2018).
63. Digital, N. Hospital Episode Statistics (HES), <https://digital.nhs.uk/data-and-information/data-tools-and-services/data-services/hospital-episode-statistics> (2021).
64. Gaziano, J. M. et al. Million Veteran Program: a mega-biobank to study genetic influences on health and disease. *J. Clin. Epidemiol.* **70**, 214–223 (2016).
65. Roden, D. M. et al. Development of a large-scale de-identified DNA biobank to enable personalized medicine. *Clin. Pharm. Ther.* **84**, 362–369 (2008).
66. Hall, J. B., Dumitrescu, L., Dilks, H. H., Crawford, D. C. & Bush, W. S. Accuracy of administratively-assigned ancestry for diverse populations in an electronic medical record-linked biobank. *PLoS ONE* **9**, e99161 (2014).
67. Tayo, B. O. et al. Genetic background of patients from a university medical center in Manhattan: implications for personalized medicine. *PLoS ONE* **6**, e19166 (2011).
68. Carey, D. J. et al. The Geisinger MyCode community health initiative: an electronic health record-linked biobank for precision medicine research. *Genet. Med.* **18**, 906–913 (2016).
69. McCarty, C. A. et al. The eMERGE network: a consortium of biorepositories linked to electronic medical records data for conducting genomic studies. *BMC Med. Genomics* **4**, 13 (2011).
70. Gottesman, O. et al. The Electronic Medical Records and Genomics (eMERGE) network: past, present, and future. *Genet. Med.* **15**, 761–771 (2013).
71. Wiley, L. K. et al. Building a vertically integrated genomic learning health system: the biobank at the Colorado Center for personalized medicine. *Am. J. Hum. Genet.* **111**, 11–23 (2024).

Acknowledgements

UK Biobank participants (Application 53723). **MVP** This research is based on data from the Million Veteran Program, Office of Research and Development, Veterans Health Administration, and was supported by awards MVP001 I01-BX004821 and MVP003/028 I01-BX003362. This publication does not represent the views of the Department of Veteran Affairs or the United States Government. This work was supported by funding from the US Department of Veterans Affairs MVP Program awards MVP001 I01-BX004821 (YVS) and MVP003/028 I01-BX003362 (TLA); by the US Department of Veterans Affairs award IK2-CX00190 (SR); and by funds from the Boettcher Foundation's Webb-Waring Biomedical Research Program (SR, EL). **The Mount Sinai BioMe Biobank** has been supported by The Andrea and Charles Bronfman Philanthropies and in part by Federal funds from the NHLBI and NHGRI (U01HG00638001; U01HG007417; X01HL134588). We thank all participants and all our recruiters who have assisted and continue to assist in data collection and management. We are grateful for the computational resources and staff expertise provided by Scientific Computing at the Icahn School of Medicine at Mount Sinai. **Vanderbilt University Medical Center's BioVU (BIOVU)** projects are supported by numerous sources: institutional funding, private agencies, and federal grants. These include NIH funded Shared Instrumentation Grant S10OD017985, S10RR025141, and S10OD025092; CTSA grants UL1TR002243, UL1TR000445, and UL1RR024975. Genomic data are also supported by investigator-led projects that include U01HG004798, R01NS032830, RC2GM092618, P50GM115305, U01HG006378, U19HL065962, and R01HD074711. **eMERGE Network (Phase III)** This phase of the eMERGE Network was initiated and funded by the NHGRI through the following grants: U01HG008657 (Group Health Cooperative/University of Washington); U01HG008685 (Brigham and Women's Hospital); U01HG008672 (Vanderbilt University Medical Center); U01HG008666 (Cincinnati Children's Hospital Medical Center); U01HG006379 (Mayo Clinic); U01HG008679 (Geisinger Clinic); U01HG008680 (Columbia University Health Sciences); U01HG008684 (Children's Hospital of Philadelphia); U01HG008673 (Northwestern University); U01HG008701 (Vanderbilt University Medical Center serving as the Coordinating Center); U01HG008676 (Partners Healthcare/Broad Institute); U01HG008664 (Baylor College of Medicine); and U54MD007593 (Meharry Medical College). **AP** was supported by National Institute for Health and Care Research Barts Biomedical Research Centre (NIHR203330). This research was supported by the British Heart Foundation (BHF) Accelerator Award AA/18/5/34222. **AEJ** was funded in part by NIH NHGRI U01 HG 008679 and NIH NIDDK R01 DK 122503. **GC, ID, NSJ** were funded in part by NIH NIDDK R01 DK 122503. **KK** and **CW** were funded by NIH 2U01-HG008680 and NIH R01-LM013061. **IJK** was additionally funded by U01 HG11710 and K24 HL137010. **JEB** and **WZ** were supported by NIH funded grants R01DK139598, R01HL142302 and R01HL163262. **RJFL** was funded by NIH (R01HG010297, R01HL151152, R01DK075787), and Novo Nordisk Foundation (NNF20OC0059313). This work was done under the auspices of the GIANT consortium.

Author contributions

A.P. conduction of analyses, interpretation of results, writing the original draft. E.L., Z.W., G.C., I.D., N.S.J., W.Z., A.E.J., M.L., J.S. conduction of analyses. M.G., R.A.J.S., S.V., L.Y., A.R.W., S.I.B., I.H., F.M., H.H., K.K., C.W., G.P.J., D.C., D.S.C., I.K., O.D., M.G.H., W.W., D.R.V.E., T.A., J.N.H., J.E.B., C.R.G., A.E.J., R.J.F.L., Y.V.S., S.R., M.L., J.S. interpretation of results, critical editing, and article approval. P.D., K.E.N., E.M. study design, supervision and interpretation of results, critical editing, and article approval. All authors have read and approved the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41525-025-00464-w>.

Correspondence and requests for materials should be addressed to E. Marouli.

Reprints and permissions information is available at <http://www.nature.com/reprints>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2025

¹William Harvey Research Institute, Faculty of Medicine and Dentistry, Queen Mary University of London, London, UK. ²VA Eastern Colorado Health Care System, Aurora, CO, USA. ³Division of Biomedical Informatics and Personalized Medicine, Department of Medicine, University of Colorado Anschutz Medical Campus, Aurora, CO, USA. ⁴Department of Epidemiology, University of North Carolina at Chapel Hill, Chapel Hill, NC, USA. ⁵The Charles Bronfman Institute for Personalized Medicine, Icahn School of Medicine at Mount Sinai, New York, NY, USA. ⁶The Mindich Child Health and Development Institute, Icahn School of Medicine at Mount Sinai, New York, NY, USA. ⁷Department of Clinical Epidemiology, Leiden University Medical Center Leiden, Leiden, NL, The Netherlands. ⁸The Genetics of Obesity and Related Metabolic Traits Program, Icahn School of Medicine at Mount Sinai, New York, NY, USA. ⁹Department of Population Health Sciences, Geisinger, Danville, PA, USA. ¹⁰Department of Genomic Health, Geisinger, Danville, PA, USA. ¹¹Colorado Center for Personalized Medicine, Department of Biomedical Informatics, University of Colorado Anschutz Medical Campus, Aurora, USA. ¹²Vanderbilt Genetics Institute, Vanderbilt University Medical Center, Nashville, TN, USA. ¹³Program in Medical and Population Genetics, Broad Institute of Harvard and MIT, Cambridge, MA, USA. ¹⁴Division of Endocrinology and Center for Basic and Translational Obesity Research, Boston Children's Hospital, Boston, MA, USA. ¹⁵Institute for Molecular Bioscience, The University of Queensland, Brisbane, Australia. ¹⁶Department of Biomedical Science, Centre of Membrane Interactions and Dynamics, University of Sheffield, Western Bank, Sheffield, UK. ¹⁷Division of Cancer Epidemiology and Genetics, National Cancer Institute, NIH, Bethesda, MD, USA. ¹⁸Division of Genetics and Genomics and Manton Center for Orphan Diseases Research, Boston Children's Hospital, Department of Pediatrics, Harvard Medical School, Boston, MA, USA. ¹⁹The Center for Applied Genomics, Children's Hospital of Philadelphia, Philadelphia, PA, USA. ²⁰Department of Medicine, Division of Nephrology, Vagelos College of Physicians & Surgeons, Columbia University, New York, NY, USA. ²¹Department of Biomedical Informatics, Vagelos College of Physicians & Surgeons, Columbia University, New York, NY, USA. ²²Department of Medicine (Medical Genetics) and Genome Sciences, University of Washington Medical Center, Seattle, WA, USA. ²³Division of Biomedical Informatics and Genomics, John W. Deming Department of Medicine, Tulane University, School of Medicine, New Orleans, LA, USA. ²⁴Kaiser Permanente Washington Health Research Institute, Seattle, WA, USA. ²⁵Department of Cardiovascular Medicine and the Gonda Vascular Center, Mayo Clinic, Rochester, MN, USA. ²⁶Division of Endocrinology, Metabolism, and Molecular Medicine, Feinberg School of Medicine, Northwestern University, Chicago, IL, USA. ²⁷Department of Biomedical Informatics, Vanderbilt University Medical Center, Nashville, TN, USA. ²⁸Division of Quantitative Sciences, Department of Obstetrics and Gynecology, Vanderbilt University Medical Center, Nashville, TN, USA. ²⁹VA Palo Alto Health Care System, Palo Alto, CA, USA. ³⁰Program in Medical and Population Genetics, Broad Institute, Boston, MA, USA. ³¹Departments of Genetics and Pediatrics Harvard Medical School, Boston, MA, USA. ³²Novo Nordisk Foundation Center for Basic Metabolic Research, Faculty of Health and Medical Sciences, University of Copenhagen, Copenhagen, Denmark. ³³Atlanta VA Health Care System, Decatur, GA, USA. ³⁴Department of Epidemiology, Emory University Rollins School of Public Health, Atlanta, GA, USA. ³⁵Digital Environment Research Institute, Queen Mary University of London, London, UK. ³⁶These authors contributed equally: P. Deloukas, K. E. North, E. Marouli.

✉ e-mail: e.marouli@qmul.ac.uk