A saturated map of common genetic variants associated with human height

Since 2007, genome-wide association studies (GWASs) have identified thousands of associations between common SNPs and height, mainly using studies with participants of European ancestry. The largest GWAS published so far for adult height focused on common variation and reported up to 3,290 independent associations in 712 loci using a sample size of up to 700,000 individuals. Adult height, which is highly heritable and easily measured, has provided a larger number of common genetic associations than any other human phenotype. In addition, a large collection of genes has been implicated in disorders of skeletal growth, and these are enriched in loci mapped by GWASs of height in the normal range. These features make height an attractive model trait for assessing the role of common genetic variation in defining the genetic and biological architecture of polygenic human phenotypes.

As available sample sizes continue to increase for GWASs of common variants, it becomes important to consider whether these larger samples can ‘saturate’ or nearly completely catalogue the information that can be derived from GWASs. This question of completeness can take several forms, including prediction accuracy compared with heritability attributable to common variation, the mapping of associated genomic regions that account for this heritability, and whether increasing sample sizes continue to provide additional information about the identity of prioritized genes and gene sets. Furthermore, because most GWASs continue to be performed largely in populations of European ancestry, it is necessary to address these questions of completeness in the context of multiple ancestries. Finally, some have proposed that, when sample sizes become sufficiently large, effectively every gene and genomic region will be implicated by GWASs, rather than certain subsets of genes and biological pathways being specified.

Here, using data from 5.4 million individuals, we set out to map common genetic associations with adult height, using variants catalogued in the HapMap 3 project (HM3), and to assess the saturation of this map with respect to variants, genomic regions and likely causal genes and gene sets. We identify significant variants, examine signal density across the genome, perform out-of-sample estimation and prediction analyses within studies of individuals of European ancestry and other ancestries and prioritize genes and gene sets as likely mediators of the effects on height. We show that this set of common variants reaches predicted limits for prediction accuracy within populations of European ancestry and largely saturates both the genomic regions associated with height and broad categories of gene sets that are likely to be relevant; future work will be required to extend prediction accuracy to populations of other ancestries, to account for rarer genetic variation and to more definitively connect associated regions with individual probable causal genes and variants.

An overview of our study design and analysis strategy is provided in Extended Data Fig. 1.
Meta-analysis identifies 12,111 height-associated SNPs

We performed genetic analysis of up to 5,380,080 individuals from 281 studies from the GIANT consortium and 23andMe. Supplementary Fig. 1 represents projections of these 281 studies onto principal components reflecting differences in allele frequencies across ancestry groups in the 1000 Genomes Project (1KGP). Altogether, our discovery sample includes 4,080,687 participants of predominantly European ancestries (75.8% of total sample); 472,730 participants with predominantly East Asian ancestries (8.8%); 455,180 participants of Hispanic ethnicity with typically admixed ancestries (8.5%); 293,593 participants of predominantly African ancestries—mostly African American individuals with admixed African and European ancestries (5.5%); and 77,890 participants of predominantly South Asian ancestries (1.4%).

We refer to these five groups of participants or cohorts as EUR, EAS, HIS, AFR and SAS, respectively, while recognizing that these commonly used groupings oversimplify the actual genetic diversity among participants. Cohort-specific information is provided in Supplementary Tables 1–3.

We tested the association between standing height and 1,385,132 autosomal bi-allelic SNPs from the HM3 targeting panel5, which contains more than 1,095,888 SNPs with a minor allele frequency (MAF) greater than 1% in each of the five ancestral groups included in our meta-analysis. Supplementary Fig. 2 shows the frequency and imputation quality distribution of HM3 SNPs across all five groups of cohorts.

We first performed separate meta-analyses in each of the five groups of cohorts. We identified 9,863, 1,511, 918, 453 and 69 quasi-independent genome-wide significant (GWAS; \(P \leq 5 \times 10^{-8}\)) SNPs in the EUR, HIS, EAS, AFR and SAS groups, respectively (Table 1 and Supplementary Tables 4–8). Quasi-independent associations were obtained after performing approximate conditional and joint (COJO) multiple-SNP analyses6, as implemented in GCTA7 (Methods). Supplementary Note 1 presents sensitivity analyses of theseCoefficient of determination (LD) in admixed AFR and HJS individuals8 (Supplementary Fig. 3), and shows how to correct those biases by varying the GCTA input parameters (Supplementary Fig. 4). Moreover, previous studies have shown that confounding due to population stratification may remain uncorrected in large GWAS meta-analyses9,10,11. Therefore, we specifically investigated confounding effects in all ancestry-specific GWASs, and found that our results are minimally affected by population stratification (Supplementary Note 2 and Supplementary Figs. 5–7).

To compare results across the five groups of cohorts, we examined the genetic and physical colocalization between SNPs identified in the largest group (EUR) with those found in the other (non-EUR) groups. We found that more than 85% of GWS SNPs detected in the non-EUR groups are in strong LD (\(r^2_D > 0.8\)) with at least one variant reaching marginal genome-wide significance (GWAS; \(P \leq 5 \times 10^{-8}\)) in EUR (Supplementary Tables 5–8). Furthermore, more than 91% of associations detected in non-EUR meta-analyses fall within 100 kb of a GWS SNP identified in EUR (Extended Data Fig. 2). By contrast, a randomly sampled HM3 SNP (matched with GWS SNPs identified in non-EUR meta-analyses on 24 functional annotations; Methods) falls within 100 kb of a EUR GWS SNP 35% of the time on average (s.d. = 1% over 1,000 draws).

Next, we quantified the cross-ancestry correlation of marginal allele substitution effects (\(\rho_{sb}\)) at GWS SNPs for all pairs of ancestry groups. We estimated \(\rho_{sb}\) using five subsets of GWS SNPs identified in each of the ancestry groups, which also reached marginal genome-wide significance in at least one group. After correction for winner’s curse12, we found that \(\rho_{sb}\) ranged between 0.64 and 0.99 across all pairs of ancestry groups and all sets of GWS SNPs (Supplementary Figs. 8–12). We also extended the estimation of \(\rho_{sb}\) for SNPs that did not reach genome-wide significance and found that \(\rho_{sb} > 0.5\) across all comparisons (Supplementary Fig. 13). Thus, the observed GWS height associations are substantially shared across major ancestral groups, consistent with previous studies based on smaller sample sizes13,14.

To find signals that are specific to certain groups, we tested whether any individual SNPs detected in non-EUR GWASs are conditionally independent of signals detected in EUR GWASs. We fitted an approximate joint model that includes GWS SNPs identified in EUR and non-EUR, using LD reference panels specific to each ancestry group. After excluding SNPs in strong LD (\(r^2_D > 0.8\) in either ancestry group), we found that 2, 17, 49 and 63 of the GWS SNPs detected in SAS, AFR, EAS and HIS GWASs, respectively, are conditionally independent of GWS SNPs identified in EUR GWASs (Supplementary Table 9). On average, these conditionally independent SNPs have a larger MAF and effect size in non-EUR than in EUR cohorts, which may have contributed to an increased statistical power of detection. The largest frequency difference relative to EUR was observed for rs2463169 (height-increasing G allele frequency: 23% in AFR versus 84% in EUR) within the intron of PAWR, which encodes the prostate apoptosis response-4 protein. Of note, rs2463169 is located within the 12q21.2 locus, where a strong signal of positive selection in West African Yoruba populations was previously reported15. The estimated effect at rs2463169 is \( \beta = 0.034 \) s.d. per G allele in AFR versus \( \beta = -0.002 \) s.d. per G allele in EUR, and the \( P \) value of marginal association in EUR is \( P_{EUR} = 0.08 \), suggesting either a true difference in effect size or nearby causal variant(s) with differing LD to rs2463169.

Given that our results show a strong genetic overlap of GWAS signals across ancestries, we performed a fixed-effect meta-analysis of all five ancestry groups to maximize statistical power for discovering associations due to shared causal variants. The mean Cochran’s heterogeneity
Q-statistic is around 34% across SNPs, which indicates moderate heterogeneity of SNP effects between ancestries. The mean chi-square association statistic in our fixed-effect meta-analysis (hereafter referred to as META_ε) is around 36, and around 18% of all HM3 SNPs are marginally GWS. Moreover, we found that allele frequencies in our META_ε were very similar to that of EUR (mean fixation index of genetic differentiation (F_ST) across SNPs between EUR and META_ε is around 0.001), as expected because our META_ε consists of more than 75% EUR participants and around 14% participants with admixed European and non-European ancestries that is, HIS and AFR). To further assess whether LD in our META_ε could be reasonably approximated by the LD from EUR, we performed an LD score regression analysis of our META_ε using LD scores estimated in EUR. In this analysis, we focused on the attenuation ratio statistic (R^2_LDSC-EUR), for which large values can also indicate strong LD inconsistencies between a given reference and GWAS summary statistics. A threshold of R^2_LDSC-EUR > 0.2 was recommended by the authors of the LDSC software as a rule-of-thumb to detect such inconsistencies. Using EUR LD scores in the GWAS of HIS, which is the non-EUR group that is genetically closest to EUR (F_ST = 0.02), yields an estimated R^2_LDSC-EUR of around 25% (standard error (s.e.) 1.8%), consistent with strong LD differences between HIS and EUR. By contrast, in our META_ε, we found an estimated R^2_LDSC-EUR of around 4.5% (s.e. 0.8%), which is significantly lower than 20% and not statistically different from 3.8% (s.e. 0.8%) in our EUR meta-analysis. Furthermore, we show in Supplementary Note 1 that using a composite LD reference containing samples from various ancestries (with proportions matching that in our META_ε) does not improve signal detection over using an EUR LD reference. Altogether, these analyses suggest that LD in our META_ε can be reasonably approximated by LD from EUR.

We therefore proceeded to identify quasi-independent GWS SNPs from the multi-ancestry meta-analysis by performing a COJO analysis of our META_ε using genotypes from around 350,000 unrelated EUR participants in the UK Biobank (UKB) as an LD reference. We identified 12,111 quasi-independent GWS SNPs, including 9,920 (82%) primary signals with a GWS marginal effect and 2,191 secondary signals that only reached GWS in a joint regression model (Supplementary Table 10). Figure 1 represents the relationship between frequency and joint effect sizes of minor alleles at these 12,111 associations. The GWS SNPs obtained from the non-EUR meta-analyses above that were conditionally independent of the EUR GWS SNPs, 0/2 in SAS, 5/17 in AFR, 27/49 in EAS and 27/63 in HIS were marginally significant in our META_ε (Supplementary Table 9), and 24 of those (highlighted in Fig. 2) overlapped with our list of 12,111 quasi-independent GWS SNPs.

We next sought to replicate the 12,111 META_ε signals using GWAS data from 49,160 participants in the Estonian Biobank (EBB). We first re-assessed the consistency of allele frequencies between our META_ε and the EBB set. We found a correlation of allele frequencies of around 0.98 between the two datasets and a mean F_ST across SNPs of around 0.005, similar to estimates that were obtained from populations from the same continent. Of the 12,111 GWS SNPs identified through our COJO analysis, 11,847 were available in the EBB dataset, 97% of which (11,529) have a MAF greater than 1% (Supplementary Table 10). Given the large difference in sample size between our discovery and replication samples, direct statistical replication of individual associations at GWS is not achievable for most SNPs identified (Extended Data Fig. 3a). Instead, we assessed the correlation of SNP effects between our discovery and replication GWASs as an overall metric of replicability. Among the 11,529 out of 11,847 SNPs that had a MAF greater than 1% in the EBB, we found a correlation of marginal SNP effects of ρ = 0.93 (jackknife standard error; s.e. 0.01) and a correlation of conditional SNP effects using the same LD reference panel of ρ = 0.80 (s.e. 0.03; Supplementary Fig. 14). Although we had limited power to replicate associations with 238 GWS variants that are rare in the EBB (MAF < 1%), we found, consistent with expectations (Methods and Extended Data Fig. 3b), that 60% of them had a marginal SNP effect that was sign-consistent with that from our discovery GWAS (Fisher’s exact test; P = 0.001). The proportion of sign-consistent SNP effects was greater than 75% (Fisher’s exact test; P < 10^-50) for variants with a MAF greater than 1%—also consistent with expectations (Extended Data Fig. 3b). Altogether, our analyses demonstrate the robustness of our findings and show their replicability in an independent sample.

**Genomic distribution of height-associated SNPs**

To examine signal density among the 12,111 GWS SNPs detected in our META_ε, we defined a measure of local density of association signals for each GWS SNP on the basis of the number of additional independent associations within 100 kb (Supplementary Fig. 15). Supplementary Fig. 16 shows the distributions of density of height-associated SNPs across the entire genome is 2.0 (s.e. 0.14), consistent with a non-random genomic distribution of GWS SNPs. Next, we evaluated signal density around 462 autosomal genes curated from the Online Mendelian Inheritance in Man (OMIM) database as containing pathogenic mutations that cause syndromes of abnormal skeletal growth (‘OMIM genes’; Methods and Supplementary Table 11). We found that a high density of height-associated SNPs is significantly correlated with the presence of an OMIM gene nearby (enrichment fold of OMIM gene when density is greater than 1.25×; P < 0.001; Methods and Extended Data Fig. 4a). Notably, the enrichment of OMIM genes almost linearly increases with the density of height-associated SNPs (Extended Data Fig. 4b). Thus, these 12,111 GWS SNPs nonrandomly cluster near each other and near known skeletal growth genes.

The largest density of conditionally independent associations was observed on chromosome 15 near Acan, a gene mutated in short stature and skeletal dysplasia syndromes, where 25 GWS SNPs co-localize within 100 kb of one another (Fig. 2 and Supplementary Fig. 17).
Supplementary Note 3 and Extended Data Fig. 5a–d, using haplotype-and simulation-based analyses, that a multiplicity of independent causal variants is the most likely explanation of this observation. We also found that signal density is partially explained by the presence of a recently identified12,13 height-associated variable-number tandem repeat (VNTR) polymorphism at this locus (Supplementary Note 3). In fact, the 25 independent GWS SNPs clustered within 100 kb of rs4932198 explain more than 40% of the VNTR length variation in multiple ancestries (Extended Data Fig. 5e), and an additional approximately 0.24% ($P = 8.7 \times 10^{-55}$) of phenotypic variance in EUR above what is explained by the VNTR alone (Extended Data Fig. 5f). Altogether, our conclusion is consistent with previous evidence of multiple types of common variation influencing height through ACAN gene function, involving multiple enhancers23, missense variants24 and tandem repeat polymorphisms21,22.

Variance explained by SNPs within identified loci

To quantify the proportion of height variance that is explained by GWS SNPs identified in our META FE, we stratified all HM3 SNPs into two groups: SNPs in the close vicinity of GWS SNPs, hereafter denoted GWS loci; and all remaining SNPs. We defined GWS loci as non-overlapping genomic segments that contain at least one GWS SNP, such that GWS SNPs in adjacent loci are more than 2 × 35 kb away from each other (that is, a 35-kb window on each side). We chose this size window because it was predicted that causal variants are located within 35 kb of GWS SNPs with a probability greater than 80% (ref. 23). Accordingly, we grouped the 12,111 GWS SNPs identified in our META FE into 7,209 non-overlapping loci (Supplementary Table 12) with lengths ranging from 70 kb (for loci containing only one signal) to 711 kb (for loci containing up to 25 signals). The average length of GWS loci is around 90 kb (s.d. 46 kb). The cumulative length of GWS loci represents around 647 Mb, or about 21% of the genome (assuming a genome length of around 3,039 Mb)24.

To estimate the fraction of heritability that is explained by common variants within the 21% of the genome overlapping GWS loci, we calculated two genomic relationship matrices (GRMs)—one for SNPs within these loci and one for SNPs outside these loci—and then used both matrices to estimate a stratified SNP-based heritability ($h^2_{SNP}$) of height in eight independent samples of all five population groups represented in our META FE (Fig. 3 and Methods). Altogether, our stratified estimation of SNP-based heritability shows that SNPs within these 7,209 GWS loci explain around 100% of $h^2_{SNP}$ in EUR and more than 90% of $h^2_{SNP}$ across all non-EUR groups, despite being drawn from less than 21% of the genome (Fig. 3). We also varied the window size used to define GWS loci and found that 35 kb was the smallest window size for which this level of saturation of SNP-based heritability could be achieved (Supplementary Fig. 18).

To further assess the robustness of this key result, we tested whether the 7,209 height-associated GWS loci are systematically enriched for trait heritability. We chose body-mass index (BMI) as a control trait, given its small genetic correlation with height ($r_g = -0.1$, ref. 27) and found no significant enrichment of SNP-based heritability for BMI within height-associated GWS loci (Supplementary Fig. 19). Furthermore, we repeated our analysis using a random set of SNPs matched with the 12,111 height-associated GWS SNPs on EUR MAF and LD scores. We found that this control set of SNPs explained only around 27% of $h^2_{SNP}$ for height, consistent with the proportion of SNPs within the loci defined by this random set of SNPs (Supplementary Figs. 18 and 19).

Finally, we extended our stratified estimation of SNP-based heritability to all well-imputed common SNPs (that is, beyond the HM3 panel) and found, consistently across population groups, that although more genetic variance can be explained by common SNPs that are not included in the HM3 panel, all information remains concentrated within these 7,209 GWS loci (Extended Data Fig. 6). Thus, with this large GWAS, nearly all of the variability in height that is attributable to common genetic variants can be mapped to regions comprising around 21% of the genome. Further work is required in cohorts of non-European ancestries to map the remaining 5–10% of the SNP-based heritability that is not captured within those regions.

Out-of-sample prediction accuracy

We quantified the accuracy of multiple polygenic scores (PGSs) for height on the basis of GWS SNPs (hereafter referred to as PGS GWS) and on the basis of all HM3 SNPs (hereafter referred to as PGS HM3). PGS GWS were calculated using joint SNP effects from COJO, and PGS HM3 using joint effects calculated using the SBayesC method29 (Methods). We denote $R^2_{GWS}$ and $R^2_{HM3}$ as the prediction accuracy of PGS GWS and PGS HM3, respectively. For conciseness, we also use the abbreviations PGS GWS X and PGS HM3 X (and $R^2_{GWS-X}$ and $R^2_{HM3-X}$) to specify which GWAS
meta-analysis each PGS (and corresponding prediction accuracy) was trained from. For example, PGS_{GWS-META} refers to PGSs based on 12,111 GWS SNPs identified from our META_{AFR}.

We first present results from PGS_{GWS} across different ancestry groups. PGS_{GWS-META} yielded prediction accuracies greater than or equal to that of all other PGS_{GWS} (Fig. 4a), partly reflecting sample size differences between ancestry-specific GWASs and also consistent with previous studies\textsuperscript{26}. PGS_{GWS-EUR} (based on 9,863 SNPs) was the second best of all PGS_{GWS} across ancestry groups except in AFR. Indeed, PGS_{GWS-AFR} (based on 453 SNPs) yielded an accuracy of 8.5% (s.e. 0.6%) in AFR individuals from UKB and PAGE; that is, significantly larger than the 5.9% (s.e. 0.6%) and 7.0% (s.e. 0.6%) achieved by PGS_{GWS-EUR} in these two samples, respectively (Fig. 4a). PGS_{GWS-META} was the best of all PGS_{GWS} in AFR participants with an accuracy $R^2_{GWS-META} = (12.3\% + 9.9\%)/2 = 10.8\%$ (s.e. 0.5%) on average between UKB and PAGE (Fig. 4a). Across ancestry groups, the highest accuracy of PGS_{GWS-META} was observed in EUR participants ($R^2_{GWS-META} = 40\%$; s.e. 0.6%) and the lowest in AFR participants from the UKB ($R^2_{GWS-META} = 9.4\%$; s.e. 0.7%). Note that the difference in $R^2_{GWS-META}$ between the EUR and AFR ancestry cohorts is expected because of the over-representation of EUR in our META_{AFR}, and consistent with a relative accuracy ($R^2_{GWS-META}$ in AFR)/($R^2_{GWS-META}$ in EUR) of around 25% that was previously reported\textsuperscript{30}. We extended analyses of PGS_{GWS} to PGS based on SNPs identified with COJO at lower significance thresholds (Extended Data Fig. 7). As in previous studies\textsuperscript{120,310}, the inclusion of sub-significant SNPs increased the accuracy of ancestry-specific PGSs. However, lowering the significance thresholds in our META_{AFR} mostly improved accuracy in EUR (from 40% to 42%), whereas it slightly decreased the accuracy in AFR.

Overall, ancestry-specific PGS_{HM3} consistently outperform their corresponding PGS_{GWS} in most ancestry-groups. However, PGS_{HM3} was sometimes less transferable across ancestry groups than PGS_{GWS}, in particular in AFR and HIS individuals from PAGE. In EUR, PGS_{HM3} reaches an accuracy of 44.7% (s.e. 0.6%), which is higher than previously published SNP-based predictors of height derived from individual-level data\textsuperscript{31-33} and from GWAS summary statistics\textsuperscript{28,34,35} across various experimental designs (different SNP sets, different sample sizes and so on). Finally, the largest improvement of PGS_{HM3} over PGS_{GWS} was observed in AFR individuals from the PAGE study ($R^2_{GWS-AFR} = 8.5\%$ versus $R^2_{HM3} = 15.4\%$; Fig. 4a) and the UKB ($R^2_{GWS-AFR} = 8.5\%$ versus $R^2_{HM3} = 14.4\%$; Fig. 4a).

Furthermore, we sought to evaluate the prediction accuracy of PGSs relative to that of familial information as well as the potential improvement in accuracy gained from combining both sources of information. We analysed 981 unrelated EUR trios (that is, two parents and one child) and 17,492 independent EUR sibling pairs from the UKB, who were excluded from our META_{AFR}. We found that height of any first-degree relative yields a prediction accuracy between 25% and 30% (Fig. 4b). Moreover, the accuracy of the parental average is around 43.8% (s.e. 3.2%), which is lower than yet not significantly different from the accuracy of PGS_{HM3-EUR}. In EUR, we found that a linear combination of the average height of parents and of the child’s PGS yields an accuracy of 54.2% (s.e. 3.2%) with PGS_{GWS-EUR} and 55.2% (s.e. 3.2%) with PGS_{HM3-EUR}. This observation reflects the fact that PGSs can explain within-family differences between siblings, whereas average parental height cannot. To show this empirically, we estimate that our PGSs based on GWS SNPs explain around 33% (s.e. 0.7%) of height variance between siblings (Methods). Finally, we show that the optimal weighting between parental average and PGS can be predicted theoretically as a function of the prediction accuracy of the PGS, the full narrow sense heritability and the phenotypic correlation between spouses (Supplementary Note 4 and Supplementary Fig. 20).

In summary, the estimation of variance explained and prediction accuracies in samples with European ancestry show that the set of 12,111 GWS SNPs accounts for nearly all of $h^2_{SNP}$, and that combining SNP-based PGS with family history significantly improves prediction accuracy. By contrast, both estimation and prediction results show clear attenuation in samples with non-European ancestry, consistent with previous studies\textsuperscript{30,33-38}.
GWAS discoveries, sample size and ancestry diversity

Our large study offers the opportunity to quantify empirically how much increasing GWAS sample sizes and ancestry diversity affects the discovery of variants, genes and biological pathways. To address this question, we re-analysed three previously published GWASs of height19,20,21 and also down-sampled our meta-analysis into four subsets (including our EUR and META_eu GWASs). Altogether, we analysed seven GWASs with a sample size increasing from around 0.13 million up to around 5.3 million individuals (Table 2).

For each GWAS, we quantified eight metrics grouped into four variant- and locus-based metrics (number of GWS SNPs; number of GWS loci; prediction accuracy \(R^2\) of PGS based on GWS SNPs; and proportion of the genome covered by GWS loci), a functional-annotation-based metric (enrichment statistics from stratified LDSC29,30), two gene-based metrics (number of genes prioritized by summary-data-based Mendelian randomization4 (SMR; Methods) and proximity of variants with OMIM genes) and a gene-set-based metric (enrichment within clusters of gene sets or pathways). Overall, we found different patterns for the relationship between those metrics and GWAS sample size and ancestry composition, consistent with varying degrees of saturation achieved at different sample sizes.

We observed the strongest saturation for the gene-set and functional-annotation metrics, which capture how well general biological functions can be inferred from GWAS results using currently available computational methods. Using two popular gene-set prioritization methods (DEPICT42 and MAGMA43), we found that the same broad clusters of related gene sets (including most of the clusters enriched for OMIM genes) are prioritized at all GWAS sample sizes (Supplementary Fig. 21, Extended Data Fig. 8, Supplementary Tables 13–15 and Supplementary Note 5). Similarly, stratified LDSC estimates of heritability enrichment within 97 functional annotations also remain stable across the range of sample sizes (Extended Data Fig. 9). Overall, we found no significant improvement for all these higher-level metrics from adding non-EUR samples to our analyses. The latter observation is consistent with other analyses showing that GWASs expectedly implicate similar biology across major ancestral groups (Supplementary Note 5 and Supplementary Fig. 22).

For the gene-level metric, the excess in the number of OMIM genes that are proximate to a GWS SNP (compared with matched sets of random genes) plateaus at sample sizes of larger than 1.5 million, whereas the relative enrichment of GWS SNPs near OMIM genes first decreases with sample size, then plateaus when \(n\) is greater than 1.5 million (Supplementary Fig. 23a–c). Notably, the decrease observed for \(n\) values of less than 1.5 million reflects the preferential localization of larger effect variants (those identified with smaller sample sizes) closer to OMIM genes (Supplementary Fig. 23d) and, conversely, that more recently identified variants with smaller effects tend to localize further away from OMIM genes (Supplementary Fig. 23e). We also investigated the number of genes prioritized using SMR (hereafter referred to as SMR genes; Methods) using expression quantitative trait loci (eQTLs) as genetic instruments (Supplementary Table 16) as an alternative gene-level metric and found it to saturate for \(n\) values greater than 4 million (Supplementary Fig. 23f). Note that saturation of SMR genes is partly affected by the statistical power of current eQTL studies, which do not always survey biologically relevant tissues and cell types for height. Therefore, we can expect more genes to be prioritized when integrating GWAS summary statistics from this study with those from larger eQTL studies that may be available in the future and may involve more tissue types. Gene-level metrics were also not substantially affected by adding non-EUR samples, again consistent with broadly similar sets of genes affecting height across ancestries.

At the level of variants and genomic regions, we saw a steady and almost linear increase in the number of GWS SNPs as a function of sample size, as previously reported44. However, given that newly identified variants tend to cluster near ones identified at smaller sample sizes, we also saw a saturation in the number of loci identified for \(n\) values greater than 2.5 million, where the upward trend starts to weaken (Supplementary
Table 2 Overview of five European-ancestry GWASs re-analysed in our study to quantify the relationship between sample size and discovery

| Down-sampled GWAS | Max n (mean n) | Number of GWS COJO SNPs | Percentage of the genome covered by GWS loci (35 kb) (%) |
|-------------------|---------------|-------------------------|------------------------------------------------------|
| Lango Allen et al. (2016) | 130,010 (128,942) | 240 | 0.5 |
| Wood et al. (2014) | 247,724 (239,227) | 633 | 1.4 |
| Yengo et al. (2018) | 695,648 (688,927) | 2,794 | 5.8 |
| GIANT-EUR (no 23andMe) | 1,632,839 (1,502,499) | 4,867 | 9.7 |
| 23andMe-EUR | 2,502,262 (2,498,336) | 7,020 | 13.6 |

Summary statistics from the three published GWASs were imputed using the ImpG-Summary software to maximize the coverage of HM3 SNPs (Methods). GWAS loci are defined as in the legend of Table 1.

Discussion

By conducting one of the largest GWASs so far in 5.4 million individuals, with a primary focus on common genetic variation, we have provided insights into the genetic architecture of height—including a saturated genomic map of 12,111 genetic associations for height. Consistent with previous studies, we have shown that signal density of associations (known and novel) is not randomly distributed across the genome; rather, associated variants are more likely to be detected around genes that have been previously associated with Mendelian disorders of growth. Furthermore, we observed a strong genetic overlap of association across cohorts with various ancestries. Effect estimates of associated SNPs are moderately to highly correlated (minimum = 0.64; maximum = 0.99), suggesting even larger correlations of effect sizes of underlying causal variants. Moreover, although there are significant differences in power to detect an association between cohorts with European and non-European ancestries, most genetic associations for height observed in populations with non-European ancestry lie in close proximity and in linkage disequilibrium to associations identified within populations of European ancestry.

By increasing our experimental sample size to more than seven times that of previous studies, we have explained up to 40% of the inter-individual variation in height in independent European-ancestry samples using GWS SNPs alone, and more than 90% of $h^2_{SNP}$ across diverse populations when incorporating all common SNPs within 35 kb of GWS SNPs. This result highlights that future investigations of common (MAF > 1%) genetic variation associated with height in many ancestries will be most likely to detect signals within the 7,209 GWS loci that we have identified in the present study. A question for the future is whether rare genetic variants associated with height are also concentrated within the same loci. We provide suggestive evidence supporting this hypothesis from analysing imputed SNPs with 0.1% < MAF < 1% (Supplementary Note 6, Extended Data Fig. 10 and Supplementary Fig. 25).

Table 2: Overview of five European-ancestry GWASs re-analysed in our study to quantify the relationship between sample size and discovery

| Down-sampled GWAS | Max n (mean n) | Number of GWS COJO SNPs | Percentage of the genome covered by GWS loci (35 kb) (%) |
|-------------------|---------------|-------------------------|------------------------------------------------------|
| Lango Allen et al. (2016) | 130,010 (128,942) | 240 | 0.5 |
| Wood et al. (2014) | 247,724 (239,227) | 633 | 1.4 |
| Yengo et al. (2018) | 695,648 (688,927) | 2,794 | 5.8 |
| GIANT-EUR (no 23andMe) | 1,632,839 (1,502,499) | 4,867 | 9.7 |
| 23andMe-EUR | 2,502,262 (2,498,336) | 7,020 | 13.6 |

Summary statistics from the three published GWASs were imputed using the ImpG-Summary software to maximize the coverage of HM3 SNPs (Methods). GWAS loci are defined as in the legend of Table 1.

*Summary statistics from the Lango Allen et al. study, initially over-corrected for population stratification using a double genomic control correction, were re-inflated such that the LD score regression intercept estimated from re-inflated test statistics equals $1$.

**Fig. 24a.** We found a similar pattern for the percentage of the genome covered by GWS loci, with the degree of saturation varying as a function of the window size used to define loci (Supplementary Fig. 24b). The observed saturation in PGs prediction accuracy (both within ancestry—that is, in EUR—and multi-ancestry) was more noticeable than that of the number and genomic coverage of GWS loci. In fact, increasing the sample size from 2.5 million to 4 million by adding another 1.5 million EUR samples increased the number of GWS SNPs from 7,020 to 9,863—that is, an increase of around 1.4-fold (9,863 – 7,020)/7,020—but the absolute increase in prediction accuracy is less than 2.7%.

This improvement is mainly observed in EUR but remains lower than 1.3% in individuals of the EAS and AFR ancestry groups. However, adding another approximately 1 million participants of non-EUR improves the multi-ancestry prediction accuracy by more than 3.4% (Supplementary Fig. 24c), highlighting the value of including non-EUR populations.

Altogether, these analyses show that increasing the GWAS sample size not only increases the prediction accuracy, but also sheds more light on the genomic distribution of causal variants and, at all but the largest sample sizes, the genes proximal to these variants. By contrast, enrichment of higher-level, broadly defined biological categories such as gene sets and pathways and functional annotations can be identified using relatively small sample sizes ($n = 0.25$ million for height). Of note, we confirm that increased genetic diversity in GWAS discovery samples significantly improves the prediction accuracy of PGs in under-represented ancestries.

Our study provides a powerful genetic predictor of height based on 12,111 GWS SNPs, for which accuracy reaches around 40% (that is, $80\%$ of $h^2_{SNP}$) in individuals of European ancestries and up to around $10\%$ in individuals of predominantly African ancestry. Notably, we show using a previously developed method that LD and MAF differences between European and African ancestries can explain up to around 84% (s.e. 1.5%) of the loss of prediction accuracy between these populations (Methods), with the remaining loss being presumably explained by differences in heritability between populations and/or differences in effect sizes across populations (for example, owing to gene-by-gene or gene-by-environment interactions). This observation is consistent with common causative variants for height being largely shared across ancestries. Therefore, we anticipate that fine-mapping of GWS loci identified in this study, ideally using methods that can accommodate dense sets of signals and large populations with African ancestries, would substantially improve the accuracy of a derived height PG for populations of non-European ancestry. Our study has a large number of participants with African ancestries as compared with previous efforts. However, we emphasize that further increasing the size of GWASs in populations of non-European ancestry, including those with diverse African ancestries, is essential to bridge the gap in prediction accuracy—particularly as most studies only partially capture the wide range of ancestral diversity both within Africa and globally. Such increased sample sizes would help to identify potential ancestry-specific causative variants, to facilitate ancestry-specific fine-mapping and to inform gene–environment and gene–ancestry interactions. Another important finding of our study is to show how individual PGs can be optimally combined with familial information and thereby improve the overall accuracy of height prediction to above 54% in populations of European ancestry.

Although large sample sizes are needed to pinpoint the variants responsible for the heritability of height (and larger samples in multiple ancestries will probably be required to map these at finer scale), the prioritization of relevant genes and gene sets is feasible at smaller
sample sizes than that required to account for the common variant heritability. Thus, the sample sizes required for saturation of GWAS are smaller for identifying enriched gene sets, with the identification of genes implicated as potentially causal and mapping of genomic regions containing associated variants requiring successively larger sample sizes. Furthermore, unlike prediction accuracy, prioritization of genes that are likely to be causal and even mapping of associated regions is consistent across ancestries, reflecting the expected similarity in the biological architecture of human height across populations. Recent studies using UKB data predicted that GWAS sample sizes of just over 3 million individuals are required to identify 6,000–7,000 GWAS SNPs explaining more than 90% of the SNP-based heritability of height. We showed empirically that these predictions are downwardly biased given that around 10,000 independent associations are, in fact, required to explain 80–90% of the SNP-based heritability of height in EUR individuals. Discrepancies between observed and predicted levels of saturation could be explained by several factors, such as (i) heterogeneity of SNP effects between cohorts and background ancestries, which may have reduced the statistical power of our study as compared to a homogenous sample like UKB; (ii) inconsistent definitions of GWS SNPs (using COJO in this study versus standard clumping in ref. 52); and, most importantly, (iii) misspecification of the SNP-effects distribution assumed to make these predictions. Nevertheless, if these predictions reflect proportional levels of saturation between traits, then we could expect that two- to tenfold larger samples would be required for GWAS of inflammatory bowel disease (>2, that is, n = 10 million), schizophrenia (>7; n = 35 million) or BMI (>10; n = 50 million) to reach a similar saturation of 80–90% of SNP-based heritability.

Our study has a number of limitations. First, we focused on SNPs from the HM3 panel, which only partially capture common genetic variation. However, although a significant fraction of height variance can be explained by common SNPs outside the HM3 SNPs panel, we showed that the extra information (also referred to as ‘hidden heritability’) remains concentrated within GWS loci identified in our HM3-SNP-based analyses (Extended Data Fig. 6). This result underlines the widespread allelic heterogeneity at height-associated loci. Another limitation of our study is that we determined conditional associations using a EUR LD reference (n = 350,000), which is sub-optimal given that around 24% of our discovery sample is of non-European ancestry. We emphasize that no analytical tool with an adequately large multi-ancestry reference panel is at present available to properly address how to identify conditionally independent associations in a multi-ancestry study. Fine-mapping of variants remains a particular challenge when attempted across ancestries in loci containing multiple signals (as is often the case for height). A third limitation of our study is our inability to perform well-powered replication analyses of genetic associations specific to populations with non-European ancestries, owing to the current limited availability of such data. Finally, as with all GWASs, definitive identification of effectors genes and the mechanisms by which genes and variants influence phenotype remains a key bottleneck. Therefore, progress towards identifying causal genes from GWAS of height may be achieved by a combination of increasingly large whole-exome sequencing studies, allowing straightforward SNP-to-gene mapping, the use of relevant complementary data (for example, context-specific eQTLs in relevant tissues and cell types) and the development of computational methods that can integrate these data.

In summary, our study has been able to show empirically that the combined additive effects of tens of thousands of individual variants, detectable with a large enough experimental sample size, can explain substantial variation in a human phenotype. For human height, we show that studies of the order of around 5 million participants of various ancestries provide enough power to map more than 90% (around 100 in populations of European ancestry) of genetic variability explained by common SNPs down to around 21% of the genome. Mapping the missing 5–10% of SNP-based heritability not accounted for in the four non-European ancestries studied here will require additional and directed efforts in the future.

Height has been used as a model trait for the study of human polygenic traits, including common diseases, because of its high heritability and relative ease of measurement, which enable large sample sizes and increased power. Conclusions about the genetic architecture, sample size requirements for additional GWAS discovery and scope for polygenic prediction that were initially made for height have by-and-large agreed with those for common disease. If the results from this study can also be extrapolated to disease, this would suggest that substantially increased sample sizes could largely resolve the heritability attributed to common variation to a finite set of SNPs (and small genomic regions). These variants and regions would implicate a particular subset of genes, regulatory elements and pathways that would be most relevant to address questions of function, mechanism and therapeutic intervention.

**Online content**

Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at https://doi.org/10.1038/s41586-022-05275-y.

1. Yang, J. et al. Common SNPs explain a large proportion of the heritability for human height. *Nat. Genet.* 42, 559–564 (2010).
2. The International HapMap 3 Consortium. Integrating common and rare genetic variation in diverse human populations. *Nature* 467, 52–58 (2010).
3. Yengo, L. et al. Meta-analysis of genome-wide association studies for height and body mass index in ~700,000 individuals of European ancestry. *Hum. Mol. Genet.* 27, e61–e64 (2018).
4. Flint, J. & Ideker, T. The great hairball gambit. *PLoS Genet.* 15, e1008519 (2019).
5. The 10,000 Genomes Project Consortium. A global reference for human genetic variation. *Nature* 526, 68–74 (2015).
6. 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).
7. 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).
8. Luo, Y. et al. Estimating heritability and its enrichment in tissue-specific gene sets in admixed populations. *Hum. Mol. Genet.* 30, 1521–1534 (2021).
9. Berg, J. J. et al. Reduced signal for polygenic adaptation of height in UK Biobank. *elife* 8, e39725 (2019).
10. Sohail, M. et al. Polygenetic adaptation on height is overestimated due to uncorrected stratification in genome-wide association studies. *elife* 8, e39702 (2019).
11. Palmer, C. & Peter, I. Statistical correction of the Winner’s Curse explains replication variability in quantitative trait genome-wide association studies. *PLoS Genet.* 13, e1006998 (2017).
12. Zhang, H. & Prentice, R. L. Bias-reduced estimators and confidence intervals for odds ratios in genome-wide association studies. *Biostatistics* 9, 621–634 (2008).
13. Galsinsky, K. J. et al. Estimating cross-population genetic correlations of causal effect sizes. *Genet. Epidemiol.* 43, 180–188 (2019).
14. Guo, J. et al. Quantifying genetic heterogeneity between continental populations for human height and body mass index. *Sci. Rep.* 11, 5240 (2021).
15. Voight, B. F., Kudaravalli, S., Wen, X. & Pritchard, J. K. A map of recent positive selection in the human genome. *PLoS Biol.* 4, e72 (2006).
16. Bulik-Sullivan, B. K. et al. LD score regression distinguishes confounding from polygenicity in genome-wide association studies. *Nat. Genet.* 47, 291–295 (2015).
17. Turley, P. et al. Multi-trait analysis of genome-wide association summary statistics using MTAG. *Nat. Genet.* 50, 239–247 (2018).
18. Lui, J. C. et al. Synthesizing genome-wide association studies and expression microarray reveals novel genes that act in the human growth plate to modulate height. *Hum. Mol. Genet.* 21, 5193–5201 (2012).
19. Lango Allen, H. et al. Hundreds of variants clustered in genomic loci and biological pathways affect human height. *Nature* 467, 832–838 (2010).
20. Wood, A. R. et al. Defining the role of common variation in the genomic and biological architecture of adult human height. *Nat. Genet.* 46, 1173–1184 (2014).
21. Beyrer, D. et al. Long-read sequencing of 3,622 Icelanders provides insights into the role of structural variants in human diseases and other traits. *Nat. Genet.* 53, 779–786 (2021).
22. Mukamel, R. E. et al. Protein-coding repeat polymorphisms strongly shape diverse human structural variants in human diseases and other traits. *Science* 373, 1499–1505 (2021).
23. Hu, G., Codina, M. & Fisher, S. Multiple enhancers associated with ACAN suggest highly redundant transcriptional regulation in cartilage. *Matrix Biol.* 31, 328–337 (2012).
24. Marouli, E. et al. Rare and low-frequency coding variants alter human adult height. *Nature* 542, 185–190 (2017).
25. Wu, Y., Zheng, Z., Visscher, P. M. & Yang, J. Quantifying the mapping precision of genome-wide association studies using whole-genome sequencing data. *Genome Biol.* 18, 86 (2017).
26. Sun, H., Ding, J., Piednoël, M. & Schneeberger, K. findGSE: estimating genome size variation within human and Arabidopsis using k-mer frequencies. Bioinformatics 34, 550–557 (2018).
27. Bulik-Sullivan, B. et al. An atlas of genetic correlations across human diseases and traits. Nat. Genet. 47, 1236–1241 (2015).
28. Lloyd-Jones, L. R. et al. Improved polygenic prediction by Bayesian multiple regression on summary statistics. Nat. Commun. 10, 5086 (2019).
29. Graham, S. E. et al. The power of genetic diversity in genome-wide association studies of lipids. Nature 600, 675–679 (2021).
30. Martin, A. R. et al. Current clinical use of polygenic scores will risk exacerbating health disparities. Nat. Genet. 51, 584–591 (2019).
31. Lello, L. et al. Accurate genomic prediction of human height. Genetics 210, 477–497 (2018).
32. Privé, F. et al. Portability of 245 polygenic scores when derived from the UK Biobank and applied to 9 ancestry groups from the same cohort. Am. J. Hum. Genet. 109, 12–23 (2022).
33. Patxot, M. et al. Probabilistic inference of the genetic architecture underlying functional enrichment of complex traits. Nat. Commun. 12, 6972 (2021).
34. Ge, T., Chen, C.-Y., Ni, Y., Feng, Y.-C. A. & Smoller, J. W. Polygenic prediction via Bayesian multiple regression and continuous shrinkage priors. Nat. Commun. 10, 1776 (2019).
35. Márquez-Luna, C. et al. Incorporating functional priors improves polygenic prediction accuracy in UK Biobank and 23andMe data sets. Nat. Commun. 12, 6052 (2021).
36. Duncan, L. et al. Analysis of polygenic risk score usage and performance in diverse human populations. Nat. Commun. 10, 3328 (2019).
37. Marnetto, D. et al. Ancestry deconvolution and partial polygenic score can improve susceptibility predictions in recently admixed individuals. Nat. Commun. 11, 1628 (2020).
38. Wang, Y. et al. Theoretical and empirical quantification of the accuracy of polygenic scores in ancestry divergent populations. Nat. Commun. 11, 3865 (2020).
39. Finucane, H. K. et al. Partitioning heritability by functional annotation using genome-wide association study statistics. Nat. Genet. 47, 1228–1235 (2015).
40. Gazal, S. et al. Linkage disequilibrium-dependent architecture of human complex traits shows action of negative selection. Nat. Genet. 49, 1421–1427 (2017).
41. Zhu, Z. et al. Integration of summary data from GWAS and eQTL studies predicts complex trait gene targets. Nat. Genet. 48, 481–487 (2016).
42. Pers, T. H. et al. Biological interpretation of genome-wide association studies using predicted gene functions. Nat. Commun. 6, 5890 (2015).
43. de Leeuw, C. A., Mooij, J. M., Heskes, T. & Posthuma, D. MAGMA: generalized gene-set analysis of GWAS data. PLoS Comput. Biol. 11, e1004219 (2015).
44. Visscher, P. M. et al. 10 years of GWAS discovery: biology, function, and translation. Am. J. Hum. Genet. 101, 5–22 (2017).
45. Backman, J. D. et al. Some sequencing and analysis of 454,787 UK Biobank participants. Nature 599, 628–634 (2021).
46. Das, S. et al. Next-generation genotype imputation service and methods. Nat. Genet. 48, 1284–1287 (2016).
47. Si, Y., Vanderwerf, B. & Zöllner, S. Why are rare variants hard to impute? Coalescent models reveal theoretical limits in existing algorithms. Genetics 217, iyab011 (2021).
48. Hallbörsön, B. V. et al. The sequences of 150,119 genomes in the UK Biobank. Nature 607, 732–740 (2022).
49. Taliun, D. et al. Sequencing of 53,831 diverse genomes from the NHLBI TOPMed Program. Nature 590, 290–299 (2022).
50. Wainschtein, P. et al. Assessing the contribution of rare variants to complex trait heritability from whole-genome sequence data. Nat. Genet. 54, 263–273 (2022).
51. Yang, J. et al. Genome partitioning of genetic variation for complex traits using common SNPs. Nat. Genet. 43, 519–525 (2011).
52. O’Connor, L. J. The distribution of common-variant effect sizes. Nat. Genet. 53, 1243–1249 (2021).

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 license, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons license 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 license, visit http://creativecommons.org/licenses/by/4.0/.
© The Author(s) 2022.
Transdisciplinary Research Initiatives, Osaka University, Osaka, Japan. Departments of Pediatrics and Genetics, Harvard Medical School, Boston, MA, USA. Present address: Department of Mathematics and Statistics, St Cloud State University, St Cloud, MN, USA. Present address: Laboratory for Systems Genetics, RIKEN Center for Integrative Medical Sciences, Kanagawa, Japan. Department of Genome Informatics, Graduate School of Medicine, The University of Tokyo, Tokyo, Japan. These authors contributed equally: Loïc Yengo, Sailaja Vedantam, Eirini Marouli. These authors jointly supervised this work: Yukinori Okada, Andrew R. Wood, Peter M. Visscher, Joel N. Hirschhorn. *A list of authors and their affiliations appears online.
23andMe Research Team

Adam Auton12, Gabriel Cuellar Partida12, Yunxuan Jiang12 & Jingchunzi Shi12

Full lists of members and their affiliations appear in the Supplementary Information.

VA Million Veteran Program

Saiju Pyarajan57 & Yan Sun484,485

Full lists of members and their affiliations appear in the Supplementary Information.

DiscovEHR (DiscovEHR and MyCode Community Health Initiative)

Jason E. Miller57, Shefali S. Verma67 & Anne E. Justice6,885

Full lists of members and their affiliations appear in the Supplementary Information.

eMERGE (Electronic Medical Records and Genomics Network)

Damien Croteau-Chonka35

Full lists of members and their affiliations appear in the Supplementary Information.

Lifelines Cohort Study

Ilja M. Nolte41, Harold Snieder41, Peter M. Visscher1 & Judith M. Vonk41

Full lists of members and their affiliations appear in the Supplementary Information.

The PRACTICAL Consortium

Sonja I. Berndt15, Stephen Chanock15, Christopher Haiman339 & Loic Le Marchand390

Full lists of members and their affiliations appear in the Supplementary Information.

Understanding Society Scientific Group

Meena Kumari386

Full lists of members and their affiliations appear in the Supplementary Information.
**Methods**

A summary of the methods, together with a full description of genome-wide association analyses and follow-up analyses is described below. Written informed consent was obtained from every participant in each study, and the study was approved by relevant ethics committees (Supplementary Table 1).

**Quality control checks of individual studies**

All study files were checked for quality using the software EasyQC

**GWAS meta-analysis**

We first performed ancestry-group-specific GWAS meta-analyses of 173 studies of EUR, 56 studies of EAS, 29 studies of AFR, 11 studies of HIS and 12 studies of SAS. Meta-analyses within ancestry groups were performed as described before using a modified version of RAREM-ETAL

**Hold-out sample from the UK Biobank**

We excluded 56,477 UK Biobank (UKB) participants from our discovery GWAS for following analyses including quantification of population stratification. More precisely, our hold-out EUR sample consists of 17,942 sibling pairs and 981 trios (two parents and one child) plus all UKB participants with an estimated genetic relationship larger than 0.05 with our set of sibling pairs and trios. We identified 14,587 individuals among these 56,477 UKB participants who were unrelated (unrelatedness was determined as when the genetic relationship coefficient estimated from HM3 SNPs was lower than 0.05) to each other and used their data to quantify the variance explained by SNPs within GWS loci (described below) and the prediction accuracy of PGSSs.

**COJO analyses**

We performed COJO analyses of each of the five ancestry group-specific GWAS meta-analyses using the software GCTA (version v.1.93)

**FST calculation and (stratified) LD score regression**

We used two statistics to evaluate whether an EUR LD reference could approximate well enough the LD structure in our trans-ancestry GWAS meta-analysis. The first statistic that we used is the Wright fixation index, which measures allele frequency divergence between two populations. We used the Hudson's estimator of $F_{ST}$ as previously recommended to compare allele frequencies from our $\text{META}_{\text{EUR}}$ with that from our EUR GWAS meta-analysis and an independent replication sample from the EBB. The other statistic that we used is the attenuation ratio statistic from the LD score regression methodology. These LD score regression analyses were performed using version 1.0 of the LDSC software and using LD scores calculated from EUR participants in the 1KGP (see ‘URLs’ section). Moreover, we performed a stratified LD score regression analysis to quantify the enrichment of height heritability in 97 genomic annotations curated and described previously as the baseline-LD model. Annotation-weighted LD scores used for those analyses were also calculated using data from 1KGP (see ‘URLs’ section).

**Density of GWS signal and enrichment near OMIM genes**

We defined the density of independent signals around each GWS SNP as the number of other independent associations identified with COJO within a 100-kb window on both sides. Therefore, a SNP with no other associations within 100 kb has a density of 0, whereas a SNP colocating with 20 other GWS associations within 100 kb will have a density of 20.
We quantified the standard error of the mean signal density across the genome using a leave-one-chromosome-out jackknife procedure. We then quantified the enrichment of 462 curated OMIM genes near GWS SNPs with a large signal density, by counting the number of OMIM genes within 100 kb of a GWS SNP, then comparing that number for SNPs with a density of 0 and those with a density of at least 1. The strength of the enrichment was measured using an odds ratio calculated from a 2×2 contingency table: ‘presence/absence of an OMIM gene’ versus ‘density of 0 or larger than 0’. To assess the significance of the enrichment, we simulated the distribution of enrichment statistics for a random set of 0 or larger than 0’. To assess the significance of the enrichment, we simulated the distribution of enrichment statistics for a random set of 22 length classes (<10 kb; between 10 and 50 kb; between 50 and 100 kb; between 100 and 200 kb; between 200 and 1 Mb; and >1 Mb) to match OMIM genes with random genes. OMIM genes within a given length class were matched with the same number of non-OMIM genes present in the class. We sampled 1,000 random sets of genes and calculated for each them an enrichment statistic. Enrichment P value was calculated as the number of times enrichment statistics of random genes exceeded that of OMIM genes. The list of OMIM genes is provided in Supplementary Table 11.

**Genomic colocalization of GWS SNPs identified across ancestries**

We assessed the genomic colocalization between 2,747 GWS SNPs identified in non-EUR (Supplementary Tables 5–8) and 9,863 GWS SNPs identified in EUR (Supplementary Table 4) by quantifying the proportion of EUR GWS SNPs identified within 100 kb of any non-EUR GWS SNP. We tested the statistical significance of this proportion by comparing it with the proportion of EUR GWS SNPs identified within 100 kb of random HM3 SNPs matched with non-EUR GWS SNPs on 24 binary functional annotations.

These 24 annotations (for example, coding or conserved) are thoroughly described in a previous study and were downloaded from https://alkesgroup.broadinstitute.org/LDSCORE/baselineLD_v2.1_annots/.

Our matching strategy consists of three steps. First, we calibrated a statistical model to predict the probability for a given HM3 SNP to be GWS in any of our non-EUR GWAS meta-analyses as a function of their annotation. For that, we used a logistic regression of the non-EUR GWS status (1 = if the SNP is GWS in any of the non-EUR GWAS; 0 = otherwise) onto the 24 annotations as regressors. Second, we used that model to predict the probability to be GWS in non-EUR. Thirdly, we used the predicted probability to sample (with replacement) 1,000 random sets of 2,747 SNPs. Finally, we estimated the proportion of EUR GWS SNPs within 100 kb of SNPs in each sampled SNP set. We report in the main text the mean and s.d. over these 1,000 proportions.

To validate our matching strategy, we compared the mean value of each of these 24 annotations (for example, proportion of coding SNPs) between non-EUR EUR GWS SNPs and each of the 1,000 random sets of SNPs, using a Fisher’s exact test. For each of the 24 annotations, both the mean and median P value were greater than 0.6 and the proportion of P values < 5% was less than 1%, suggesting no significant differences in the distribution of these 24 annotations between non-EUR GWS SNPs and matched SNPs.

**Replication analyses**

To assess the replicability of our results, we tested whether the correlation $\rho_b$ of estimated SNP effects between our discovery GWAS and our replication sample of 49,160 participants of the EBB was statistically different from 1. We used the estimator of $\rho_b$ from a previous study, which accounts for sampling errors in both discovery and replication samples. Standard errors were calculated using a leave-one-SNP-out jackknife procedure. We quantified the correlation of marginal and also that of joint SNP effects. Joint SNP effects in our replication sample were obtained by performing a single-step COJO analysis of GWAS summary statistics from our EBB sample, using the same LD reference as in the discovery GWAS. Correlation of SNP effects were calculated after correcting SNP effects for winner’s curse using a previouly described method. We provide the R scripts used to apply these corrections and estimate the correlation of SNP effects (see ‘URLs’ section). The expected proportion, $E[P]$, of sign-consistent SNP effects between discovery and replication was calculated using the quadrant probability of a standard bivariate Gaussian distribution with correlation $E[\rho_b]$, denoting the expected correlation between estimated SNP effects in the discovery and replication sample:

$$E[P] = \frac{1}{2} + \frac{\sin^{-1}(E[\rho_b])}{\pi},$$

where $\sin^{-1}$ denotes the inverse of the sine function and $E[\rho_b]$ the expectation of the $\rho_b$ statistic under the assumption that the true SNP effects are the same across discovery and replications cohorts. $E[\rho_b]$ was calculated as

$$E[\rho_b] = \frac{\sigma_b^2}{\sqrt{(\sigma_b^2 + (1 - \sigma_h^2)/(N_h^2 h^2)) (\sigma_b^2 + (1 - \sigma_h^2)/(N_h^2 h^2))}},$$

where $N_h$ and $N_e$ denote the sizes of the discovery and replication samples, respectively; $h$ and $h_e$ the average heterozygosity under Hardy–Weinberg equilibrium (that is, $2 \times MAF \times (1 - MAF)$) across GWS SNPs in the discovery and replication samples, respectively; and $\sigma_b^2$ the mean per-SNP variance explained by GWS SNPs, which we calculated (as per ref. 60) as the sample variance of estimated SNP effects in the discovery sample minus the median squared standard error.

**Variance explained by GWS SNPs and loci**

We estimated the variance explained by GWS SNPs using the genetic relationship-based restricted maximum likelihood (GREML) approach implemented in GCTA. This approach involves two main steps: (i) calculation of genetic relationships matrices (GRM); and (ii) estimation of variance components corresponding to each of these matrices using a REML algorithm. We partitioned the genome in two sets containing GWS loci on the one hand and all other HM3 SNPs on the other hand. GWS loci were defined as non-overlapping genomic segments containing at least one GWS SNP and such that GWS SNPs in adjacent loci are more than $2 \times 35$ kb away from each other (that is, a 35-kb window on each side). We then calculated a GRM based on each set of SNPs and estimated jointly a variance explained by GWS alone and that explained by the rest of the genome. We performed these analyses in multiple samples independent of our discovery GWAS, which include participants of diverse ancestry. Details about the samples used for these analyses are provided below. We extended our analyses to also quantify the variance explained by GWS loci using alternative definitions based on a window size of 0 kb and 10 kb around GWS SNPs (Supplementary Figs. 18 and 19).

We also repeated our analyses using a random set of 12,111 SNPs matched with GWS SNPs on MAF and LD. Loci for these 12,111 random SNPs were defined similarly as for GWS loci. To match random SNPs with GWS SNPs on MAF and LD, we first created 28 MAF-LD classes of HM3 SNPs (7 MAF classes × 4 LD score classes). MAF classes were defined as <1%; between 1% and 5%; between 5% and 10%; between 10% and 20%; between 20% and 30%; between 30% and 40%; and between 40% and 50%. LD score classes were defined using quartiles of the HM3 LD score distribution. We next matched GWS SNPs in each of the 28 MAF-LD classes, with the same number of SNPs randomly sampled from that MAF-LD class.

**Prediction analyses**

Height was first mean-centred and scaled to variance 1 within each sex. We quantified the prediction accuracy of height predictors as the difference between the variance explained by a linear regression model of sex-standardized height regression on the height predictor, age, 20
genotypic principal components and study-specific covariates (full model) minus that explained by a reduced linear regression not including the height predictor. Genetic principal components were calculated from LD pruned HM3 SNPs ($r^2_{LD} < 0.1$). We used height of siblings or parents as a predictor of height as well as various polygenic scores (PGSs) calculated as a weighted sum of height-increasing alleles. The direction and magnitude of these weights was determined by estimated SNP effects from our discovery GWAS meta-analyses. No calibration of tuning parameters in a validation was performed.

**Between-family prediction.** We analysed two classes of PGS. The first class is based on SNPs ascertained using GCTA-COJO. We applied GCTA-COJO to ancestry-specific and cross-ancestry GWAS meta-analyses using ancestry-matched and EUR LD reference, respectively. We compared PGSs based on SNPs ascertained at different significance thresholds: $P < 5 \times 10^{-8}$ (GWS: reported in the main text) and $P < 5 \times 10^{-6}$ and $P < 5 \times 10^{-5}$ for all COJO-based PGS, we used estimated joint effects to calculate the PGS. The second class of PGS uses weights for all HM3 SNPs obtained from applying the SBayesC method to ancestry-specific and cross-ancestry GWAS meta-analyses with ancestry-matched and EUR specific LD matrices, respectively. The SBayesC method is a Bayesian PGS-method implemented in the GCTB software (v. 2.0), which uses the same prior as the LDpred method. In brief, SBayesC models the distribution of joint effects of all SNPs using a two-component mixture distribution. The first component is a point-mass Dirac distribution on zero and the other component a Gaussian distribution (for each SNP) with mean 0 and a variance parameter to estimate. Full LD matrices (that is, not sparse) were calculated using GCTB across around 250 overlapping (50% overlap) blocks of around 8,000 SNPs (average size is around 20 Mb). These LD matrices were calculated using the same sets of genotypes used for COJO analyses (described above). We ran SBayesC in each block separately with 100,000 Monte Carlo Markov Chain iterations. In each run, we initialized the proportion of causal SNPs in a block at 0.0001 and the heritability explained by SNPs in the block at 0.001. Posterior SNP effects of SNPs present in two blocks were meta-analysed using inverse-variance meta-analysis.

Prediction accuracy was quantified in 61,095 unrelated individuals from three studies, including 33,001 participants of the UKB who were not included in our discovery GWAS (that is, 14,587 EUR; 9,257 SAS; 6,911 AFR and 2,246 EAS; Methods section ‘Samples used for prediction and estimation of variance explained’): 14,058 EUR participants from the Lifelines cohort study; and 8,238 HIS and 5,798 AFR participants from the PAGE study.

**Within-family prediction.** The prediction accuracy of sibling’s height was assessed in 17,942 unrelated sibling pairs from the UKB. Those pairs were determined by intersecting the list of UKB sibling pairs determined by Bycroft et al. with a list of genetically determined European ancestry participants from the UKB also described previously. We then filtered the resulting list for SNP-based genetic relationship between members of different families to be smaller than 0.05. The prediction accuracy of parental height (each parent and their average) was assessed in 981 unrelated trios obtained as described above by crossing information from Bycroft et al. calling (of relatives) with that from Yengo et al. (of European ancestry participants). We quantified the within-family variance explained by PGS as the squared correlation of height difference between siblings with PGS difference between siblings. We describe in Supplementary Note 4 how familial information and PGS were combined to generate a single predictor.

**Samples used for prediction and estimation of variance explained** We quantified the accuracy of a PGS based on GWS SNPs as well as the variance explained by SNPs within GWS loci, in eight different datasets independent of our discovery GWAS meta-analyses. These datasets include two samples of EUR from the UKB ($n = 14,587$) and the Lifelines study ($n = 14,058$), two samples of AFR from the UKB ($n = 6,911$) and the PAGE study ($n = 8,238$), two samples of SAS and the China Kadoorie Biobank (CKB; $n = 47,693$), one sample of SAS from the UKB ($n = 9,257$) and one sample of HIS from the PAGE study ($n = 4,939$). Analyses were adjusted for age, sex, 20 genotypic principal components and study-specific covariates (for example, recruitment centres). Genotypes of EUR UKB participants were imputed to the HaploReference Consortium (HRC) and to a combined reference panel including haplotypes from the IKG Project and the UK10K Project. To improve variant coverage in non-EUR participants of UKB, we re-imputed their genotypes to the IKG reference panel, as described previously. Lifelines samples were imputed to the HRC panel. PAGE and CKB were imputed to the IKG reference panel. Standard quality control ($r^2_{INFO} > 0.3, P_{\text{HW}} > 10^{-5} \text{ and MAC} > 5$) were applied to imputed genotypes in each dataset.

**Contribution of LD and MAF to the loss of prediction accuracy** We defined the EUR-to-AFR relative accuracy as the ratio of prediction accuracies from an AFR sample over that from a EUR sample. We used a previously published method to quantify the expectation of that relative accuracy under the assumption that causal variants and their effects are shared between EUR and AFR, whereas MAF and LD structures can differ. In brief, this method contrasts LD and MAF patterns within 100-kb windows around each GWS SNPs and uses them to predict the expected loss of accuracy. As previously described, we used genotypes from 503 EUR and 661 AFR participants of the IKG as a reference sample to estimate ancestry-specific MAF and LD correlations between GWS SNPs and SNPs in their close vicinity, and defined candidate causal variants as any sequenced SNP with an $r^2_{LD} > 0.45$ with a GWS SNP within that 100-kb window. Standard errors were calculated using a delta-method approximation as previously described.

**Down-sampled GWAS analyses** In addition to our EUR GWAS meta-analysis and our trans-ancestry meta-analysis (META), we re-analysed five down-sampled GWASs as shown in Table 2. These down-sampled GWASs include various iterations of previous efforts of the GIANT consortium and have a sample size varying between around 130,000 and 2.5 million EUR participants from 23andMe. To ensure sufficient genomic coverage of HM3 SNPs we imputed GWAS summary statistics from Lango Allen et al. Wood et al. and Yengo et al. with ImpG Summary (v.1.0.1) using haplotypes from IKG as a LD reference. GWAS summary statistics from Lango Allen et al. only contain $P$ values ($P$), height-increasing alleles and per-SNP sample sizes ($N$). Therefore, we first calculated $Z$-scores ($Z$) from $P$ values assuming that $Z$-scores are normally distributed, then derived SNP effects ($\beta$) and corresponding standard errors (s.e.) using linear regression theory as $\beta = Z / \sqrt{2 \text{MAF} \times (1 - \text{MAF}) \times (N + Z^2)}$ and $SE = \beta / Z$. Imputed GWAS summary statistics from these three studies are made publicly available on the GIANT consortium website (see ‘URLs’ section). We next performed a COJO analysis of all down-sampled GWAS using genotypes of 349,501 unrelated EUR participants in the UKB as a LD reference panel, as for our META and EUR GWAS meta-analysis.

**Gene prioritization using SMR** We used SMR to identify genes whose expression could mediate the effects of SNPs on height. SMR analyses were performed using the SMR software v.1.03. We used publicly available gene eQTLs identified from two large eQTL studies; namely, the GTEx v.8 and the eQTLgen studies (see ‘URLs’ section). To ensure that our SMR results robustly reflect causality or pleiotropic effects of height-associated SNPs on gene expression, we only report here significant SMR results (that is, $P < 5 \times 10^{-8}$), which do not pass the heterogeneity in dependent instrument (HEIDI) test (that is, $P > 0.01$; Methods). The significance threshold for the HEIDI test was chosen on the basis of recommendations from another study.
To generate a list of genes that are known to underlie syndromes of abnormal skeletal growth, we queried the Online Mendelian Inheritance in Man database (OMIM; https://www.omim.org/). From July 2019 to August 2020, we performed queries using search terms of “short stature”, “tall stature”, “overgrowth”, “skeletal dysplasia” and “brachydactyly.” We then used the free text descriptions in OMIM to manually curate the resulting combined list of genes, as well as genes in our earlier list from Wood et al. and all genes listed as causing skeletal disease in an online endocrine textbook (https://www.endotext.org/, accessed September 2020). For short stature, we only included genes that underlie syndromes in which short stature was either consistent (less than −2 s.d. in the vast majority of patients with data recorded), or present in multiple families or sibships and accompanied by (a) more severe short stature (−3 s.d.), (b) presence of skeletal dysplasia (beyond poor bone quality/fractures); or (c) presence of brachydactyly, shortened digits, disproportionate short stature or limb shortening (not simply absence of specific bones). We removed genes underlying syndromes in which short stature was likely to be attributable to failure to thrive, specific metabolic disturbances, intestinal failure or enteropathy and/or very severe disease (for example, early lethality or severe neurological disease). For tall stature or overgrowth, we only included genes underlying syndromes in which tall stature was consistent (more than +2 s.d. in the vast majority of patients with data recorded) or present in multiple families or sibships and accompanied by either (a) more severe tall stature (≥+3 s.d.) or (b) arachnodactyly. For brachydactyly, we required more than only fifth finger involvement, and that brachydactyly be either consistent (present in the vast majority of patients) or accompanied by consistent short stature or other skeletal dysplasias. For skeletal dysplasias, we only considered genes that underlie syndromes in which the skeletal dysplasia involved long bones or the spine and was accompanied by short stature, brachydactyly or limb or digit shortening. We also included all genes in a list we generated in Lango Allen et al., which was curated using similar criteria. The resulting list contained 536 genes, of which 462 (Supplementary Table 36) were also included in Lango Allen et al., which was curated using similar criteria.

For brachydactyly, we required more than only fifth finger involvement, and that brachydactyly be either consistent (present in the vast majority of patients) or accompanied by consistent short stature or other skeletal dysplasias. For skeletal dysplasias, we only considered genes that underlie syndromes in which the skeletal dysplasia involved long bones or the spine and was accompanied by short stature, brachydactyly or limb or digit shortening. We also included all genes in a list we generated in Lango Allen et al., which was curated using similar criteria.

The resulting list contained 536 genes, of which 462 (Supplementary Table 36) were also included in Lango Allen et al., which was curated using similar criteria.

For brachydactyly, we required more than only fifth finger involvement, and that brachydactyly be either consistent (present in the vast majority of patients) or accompanied by consistent short stature or other skeletal dysplasias. For skeletal dysplasias, we only considered genes that underlie syndromes in which the skeletal dysplasia involved long bones or the spine and was accompanied by short stature, brachydactyly or limb or digit shortening. We also included all genes in a list we generated in Lango Allen et al., which was curated using similar criteria.

The resulting list contained 536 genes, of which 462 (Supplementary Table 36) were also included in Lango Allen et al., which was curated using similar criteria.

For brachydactyly, we required more than only fifth finger involvement, and that brachydactyly be either consistent (present in the vast majority of patients) or accompanied by consistent short stature or other skeletal dysplasias. For skeletal dysplasias, we only considered genes that underlie syndromes in which the skeletal dysplasia involved long bones or the spine and was accompanied by short stature, brachydactyly or limb or digit shortening. We also included all genes in a list we generated in Lango Allen et al., which was curated using similar criteria.

The resulting list contained 536 genes, of which 462 (Supplementary Table 36) were also included in Lango Allen et al., which was curated using similar criteria.
Scientific Research (NWO); the Government of Canada through Genome Canada and the Canadian Institutes of Health Research grant GPH-129344; the Ministére de l’Économie et de l’Innovation du Québec through Genome Quebec grant PSGR8801701; the Quebec Breast Cancer Foundation; the US Department of Defence grant W81XWH-10-1-0347; the Canadian Institutes of Health Research (CIHR) for the CIHR Team in Familial Risks of Breast Cancer; Komen Foundation for the Cure; the Breast Cancer Research Foundation; and the Ovarian Cancer Research Fund; the Economic and Social Research Council grant number ES/M 001580/1; Wellcome Investigator and NIHR Senior Investigator (M.I.M.); Council of Scientific and Industrial Research, Government of India grant number BSC012/2; the Department of Science and Technology, Government of India through PURSE II CDST/SR/PURSE PHASE II/11 provided to Jawaharlal Nehru University; the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) Projektfinanzierung 209855388 - SFB 1052; B03, C01; SPP 1629 TO 718/2-1; the Competitive Research Funding of the Tampere University Hospital grants 90048 and 90035; the Finnish Cultural Foundation; the Finnish Foundation for Cardiovascular Research; the Emil Aaltonen Foundation, Finland; Juho Vainio Foundation; Finnish Cardiac Research Foundation; Finnish Ministry of Education and Culture; Yrjö Jahnsson Foundation; C.G. Sundell Foundation; Special Governmental Grants for Health Sciences Research, Turku University Hospital; Foundation for Pediatric Research; and Turku University Foundation, the Social Insurance Institution of Finland, Competitive State Research Financing of the Expert Responsibility area of Kuopio, Tampere and Turku University Hospitals grant KSF001; Paavo Nurmi Foundation, Signe and Ane Gyllenberg Foundation, Diabetes Research Foundation of Finnish Diabetes Association; Tampere University Hospital Supporting Foundation; and Finnish Society of Clinical Chemistry, the Italian Ministry of Health—RC 01/21 (M.P.C.) and D70-RESRICGROTO (G.G.); $5 per mille 2015 sentences CUP: C90F170003560001 (P.G.), the Helmholtz Zentrum Münchener—German Research Center for Environmental Health, which is funded by the German Federal Ministry of Education and Research (BMBF) and by the State of Bavaria, the Department of Innovation, Research, and University of the Autonomous Province of Bolzano-South Tyrol; the Croatian National Center of Research Excellence in Personalized Healthcare grant number KK.0111.01.0016 (O. Polašek) and the Center of Competence in Molecular Diagnostics grant number KK.012/2.03.0006 (O. Polašek); the Norwegian Research Council Mobility Grant 24014 and Young Research Talent grant 287086; the South-Eastern Health Authorities PhD grant 2019/012; Vestre Viken Hospital Trust PhD-grant, afb.no - the Norwegian Atrial Fibrillation Research Network, Indremedisinsk Forskningsfond at Baerum Hospital, the Foundation for the National Institutes of Health Accelerating Medicines Partnership award no. HART7AMP; the Dutch Pearls of Invention; the Amsterdam Hospital; the Foundation for the National Institutes of Health Accelerating Medicines Partnership award no. HART7AMP; the Dutch Pearls of Invention; the Amsterdam Hospital; the Foundation for the National Institutes of Health Accelerating Medicines Partnership award no. HART7AMP.

From Amgen, Sanofi, MSD and Elpen. S. A. Lubitz previously received sponsored research support from Bristol Myers Squibb, Pfizer, Bayer AG, Boehringer Ingelheim, Fidia and IBM, and has consulted for Bristol Myers Squibb, Pfizer, Bayer AG and Blackstone Life Sciences. S. A. Lubitz is a current employee of Novartis Institute of Biomedical Research. W.M. reports grants and personal fees from AMGEN, BASF, Sanofi, Siemens Diagnostics, Aegerion Pharmaceuticals, Astrazeneca, Danone Research, Numera, Pfizer and Hoffmann Laroche; personal fees from MSD and Alexion; and grants from Abbott Diagnostics, all outside the submitted work. W.M. is employed with Synlab Holding Deutschland. M.A.N. receives support from a consulting contract between Data Tecnica International and the National Institute on Aging (NIA), National Institutes of Health (NIH). S.N. is a current employee of Novartis of Biomedical Research. ADAS software, CardioSolv and ImriCor receives support from Biostereose, ADAG software and ImriCor. H. Schunkert has received honoraria for consulting from Astrazeneca, MSD, Merck, Daichii, Servier, Amgen and Takeda Pharma. He has further received honoraria for lectures and/or chairs from AstraZeneca, Bayer-Vital, BRAHMS, Daichi, Medtronic, Novartis, Sanofi and Servier. P.S. has received research awards from Pfizer. Members of the 23andMe Research Team are employed by and hold stock or stock options in 23andMe. The views expressed in this article are those of the author(s) and not necessarily those of the NHS, the NIHR or the Department of Health. M. I. McCarthy is employed on advisory panels for Pfizer, Novo Nordisk and Zee Global and has received honoraria from Merck, Pfizer, Novo Nordisk and Eli Lilly and research funding from Abbvie, AstraZeneca, Boehringer Ingelheim, Eli Lilly, Janssen, Merck, Novo Nordisk, Pfizer, Roche, Sanofi Aventis, Servier and Takeda. As of June 2019, M. I. McCarthy is an employee of Genentech, and a holder of Roche stock. C.J.O. is a current employee of Novartis Institute of Biomedical Research. U.T. is employed by deCODE Genetics (Amgen). K.S. is employed by deCODE Genetics (Amgen). A. Auton is employed by and holds stock or stock options in 23andMe. G.R.A. is an employee of Regeneron Pharmaceuticals and owns stock and stock options for Regeneron Pharmaceuticals. C.J.W.’s spouse is employed by Regeneron. A.E.L. is currently employed by and holds stock in Regeneron Pharmaceuticals. J.N.H. holds equity in Camp4 Therapeutics. The remaining authors declare no competing interests.

Additional information Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41586-022-05275-y. Correspondence and requests for materials should be addressed to Loic Yengo, Yukinori Okada, Andrew R. Wood, Peter M. Visscher or Joel N. Hirschhorn.

Peer review information Nature thanks Elizabeth Atkinson and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. Peer reviewer reports are available.

Reprints and permissions information is available at http://www.nature.com/reprints.
GWAS meta-analysis of height in 281 studies

European:
- n = 4,080,687

East Asian:
- n = 472,730

Hispanic:
- n = 455,180

African/African American:
- n = 293,593

South Asian:
- n = 77,890

Replication:
- n = 49,160

Genetic discoveries:
- Heritability estimation
- Conditional analysis
- Effect size comparison

Genomic distribution:
- Signal density analysis
- OMIM enrichment

Polygenic prediction:
- Out-of-sample prediction
- Trans-ancestry comparison
- Within-family analyses

Saturation of discovery from GWAS:
- Down-sampling analysis
- Variant-, functional-, gene-, and pathway-based metrics
- Cross-population comparison

Extended Data Fig. 1 | Broad ancestries composition. Geographical mapping and ancestries composition of 281 studies meta-analysed in this study. Various analyses were performed including (1) detection of height-associated SNPs (Genetic discoveries box), (2) quantification of the genomic distribution of height-associated loci (Genomic distribution box), (3) assessment of the performances of polygenic predictors of height (Polygenic prediction box), and (4) assessment of the relationship between GWAS sample size and discoveries (Saturation of discovery from GWAS box).
Extended Data Fig. 2 | Colocalization of height-associated signals across ancestries. Proportion (y-axis) of GWS SNPs identified in our GWAS meta-analyses of non-European (non-EUR: African – AFR; East Asian – EAS; South Asian – SAS; Hispanic – HIS) ancestry/ethnicity participants that are located within a certain distance (x-axis) of GWS SNPs identified in our GWAS meta-analysis of EUR participants only.
Extended Data Fig. 3 | Replication of marginal associations in the EBB.

**a**, Each dot represents one of the 12,111 SNPs detected in our trans-ancestry meta-analysis. The x-axis represents the expected statistical power to replicate each association ($P<0.005/9,473 = 5.3\times10^{-6}$); where 9,473 is the number of associations reaching marginal genome-wide significance in our discovery trans-ancestry GWAS and with a minor allele frequency $>1\%$ in the EBB sample. The y-axis represents the $-\log_{10}$ of the association p-value in the EBB multiplied by the product of signs of estimated SNP effects in the discovery and in the EBB. Horizontal dotted line represents replication at $P<0.001$ and the vertical dotted line indicates 80% of statistical power. SNPs highlighted in green have an expected statistical power for replication $>80\%$. One outlier (rs11100870), highlighted in red, does not replicate in the EBB sample. **b**, Proportion ($P$) of SNPs with sign-consistent estimated effect sizes between discovery GWAS (N≈5.3M) and EBB. Expected proportions ($E[P]$) are calculated using equation (2) in the Methods. Error bars are defined as $1.96\times\sqrt{P(1-P)/m}$, where $m$ is the number of SNPs in the corresponding MAF interval.
Extended Data Fig. 4 | Enrichment of genes containing pathogenic mutations that cause extreme height or abnormal skeletal growth syndromes near hotspots of GWS SNPs. Four hundred and sixty-two (462) autosomal genes were curated from the Online Mendelian Inheritance in Man (OMIM) database. a, Red arrow indicates the observed enrichment statistic (OR = 2.5-fold) measuring the odds ratio of the presence of an OMIM gene within 100 kb of a GWS SNP with a density > 1. The blue histogram represents the distribution of enrichment statistics from 1,000 random genes matched, which length distribution matches that of the OMIM genes. b, Enrichment of OMIM genes near high density GWS SNPs. High density is defined by on the x-axis by the minimum number of other independent GWS SNPs detected within 100 kb.
Extended Data Fig. 5 | Haplotypic analysis at the ACAN locus. a, Distribution of estimated haplotype effects from 14,117 haplotypes covering a 100 kb long genomic region near the ACAN gene (hg19 genomic coordinates: chr15:89,307,521-89,407,521). b, Quantile-quantile plot of associations between these 14,117 haplotypes and height. c, Distribution of the variance explained by each of the 14,117 haplotypes. d, Mean signals density (y-axis) across simulated data where 1 causal SNP within the locus explains between 0.5% and 5% (x-axis) of trait variance. Causal variants were sampled from a pool of 13 SNPs with a $1.4\times10^{-5} < \text{MAF} < 1\%$ genotyped in 291,683 unrelated EUR participants of the UKB, with no missing values at these 13 SNPs. Standard errors were calculated as the standard deviation (s.d.) of signal density across 100 simulation replicates. GCTA-COJO analyses to identify independent signals were performed using a subset of 10,000 unrelated EUR participants of the UKB to mimic the large discrepancy between the size of the discovery GWAS and that of the LD reference used in our real data analyses. e, Proportion of VNTR length explained by 25 GWS SNPs identified near ACAN in 4 ancestries (European: EUR; South Asian: SAS; East Asian: EAS; African: AFR). f, Proportion of height variance explained in a sample of EUR UK Biobank participants by various sets of polymorphisms at the ACAN locus. rs3817428 and rs34949187 are two missense variants and rs7176941 is an intronic variant with high posterior causal probability identified in ref. 28. In e and f, error bars represent standard error (s.e.).
Extended Data Fig. 6 | Variance of height explained by common SNPs within 35 kb of GWS SNPs. Stratified SNP-based heritability ($h_{SNP}^2$) estimates were obtained from a partition of the genome into two sets of 1000 Genomes imputed SNPs with a minor allele frequency (MAF) >1%: (1) SNPs within +/− 35 kb of GWS (GWS loci) vs. all other SNPs. Analyses were performed in samples of five different ancestry groups: European (EUR; UK Biobank only), African (AFR), East Asian (EAS) and South Asian (SAS) as described in the legend of Fig. 3. Estimates from stratified analyses were compared with SNP-based heritability estimates obtained from analysing HM3 SNPs only (dotted horizontal violet bar).
Extended Data Fig. 7 | Accuracy of PGSs derived from joint effects of SNPs ascertained at various significance thresholds. The six panels show on their y-axes the prediction accuracy ($R^2$) of multiple PGS across five target samples. The ancestry group and size of each target sample is indicated in the panel title. The top-left panel shows the averaged prediction accuracy in two European ancestry (EUR) target samples from the UK Biobank (UKB) and Lifelines Biobank (LLB). The other panels show prediction accuracies in individual target samples of African ancestry (AFR) from UKB and the PAGE study, East Asian ancestry (EAS) and South Asian ancestry (SAS) ancestry from the UKB and Hispanic ethnicity from the PAGE study. Each panel is divided in four columns representing the four significance levels used to ascertain SNPs using the GCTA-COJO algorithm. GCTA-COJO was applied to each ancestry-group specific GWAS meta-analysis with an ancestry-match linkage disequilibrium (LD) reference. We used genotypes from 50,000 (vs 350,000 for results reported in the main text) unrelated EUR participants as LD reference to run GCTA-COJO on the EUR- and the cross-ancestry GWAS meta-analysis. For the other ancestry groups, we used genotypes from 10,636 AFR individuals, 5,875 EAS individuals, 4,883 HIS individuals and 9,448 SAS individuals as LD reference (as described in Methods). Error bars are standard error (s.e.). The number of SNPs used in each PGS is indicated (in white) within each bar.
Extended Data Fig. 8 | Enrichment of height-associated genes identified at various GWAS sample sizes within 20 clusters of gene sets representing broad categories of biological pathways. Gene-set enrichment was performed with MAGMA and DEPICT across seven GWAS with increasing sample sizes. Samples used (Lango Allen et al. (2010), $n = 0.13\ M$; Wood et al. (2014), $n = 0.24\ M$; Yengo et al. (2018), $n = 0.7\ M$; GIANT-EUR (no 23andMe), $n = 1.63\ M$; 23andMe-EUR, $n = 2.5\ M$; European-ancestry meta-analysis, $n = 4.08\ M$; and cross-ancestry meta-analysis, $n = 5.31\ M$) are described in Tables 1–2.

The degree of enrichment of gene sets (MAGMA, DEPICT) of known skeletal growth disorder genes catalogued in the Online Mendelian Inheritance in Man (OMIM) database among 20 clusters of gene sets (see Methods section in Supplementary Note 5) is indicated by the blue-red colour scale. Enrichment for MAGMA and DEPICT was defined to be the number of prioritized gene sets (top 10% of gene sets) in each cluster divided by the 10% of the number of gene sets in the cluster. Enrichment for OMIM was defined to be the number of OMIM genes in a gene set ($Z > 1.96$) divided by the size of the gene set divided by the proportion of all genes in OMIM, then averaged across the cluster. Significant enrichment (compared to shuffled prioritization of gene sets or genes) is marked with *.
Extended Data Fig. 9 | Annotation-level saturation of GWAS discoveries as a function of sample size. Increase in sample size from ~4 million to ~5 million is achieved by including ~1 million participants of non-European ancestry. 

a. Number of annotations showing a significant heritability enrichment as a function of the sample size of the GWAS used to estimate these enrichment. Heritability enrichment was detected using a stratified LD score regression (LDSC) analysis of 97 genomic annotations included in the “baseline + LD” model from Gazal et al.

b. Correlation between Z-scores measuring the statistical significance of heritability enrichments of 97 annotations (each dot is an annotation) in our largest GWAS (x-axis) as compared to down-sampled GWAS (y-axis). Sample size is denoted by the colour-code.

c. Distribution of estimated enrichment statistics for 21 annotations found significantly enriched ($P < 0.05/97$) in at least 6 of the 7 GWAS analysed here. LoF-i genes: Loss of function intolerant genes; TSS: Transcription Start Sites; DGF: Digital genomic footprint; TFBS: Transcription Factor Binding Sites; DHS: DNAse I hypersensitive sites; GERP (NS): GERP++ score (number of substitutions).
Extended Data Fig. 10 | Partitioning of low-frequency SNP-based heritability within GWS loci. Panels b–d represent partitioned SNP-based heritability estimates from three samples (EBB: Estonian Biobank; UKB: UK Biobank; LLB: Lifelines Biobank) of unrelated European ancestry individuals independent of our discovery GWAS. a, Partitioned SNP-based heritability estimates obtained from an inverse-variance weighted meta-analysis of estimates shown in b–d. SNPs were partitioned into four classes according to their minor allele frequency (MAF: 0.1% < MAF < 1% vs. MAF > 1%) and their position within versus outside GWS loci. The SNP-based heritability contributed by SNPs within GWS loci is denoted \( h^2_{\text{GWS}} \), and that contributed by SNPs outside these loci is denoted \( h^2_{\text{other}} \). These results are further discussed in Supplementary Note 6.
Reporting Summary

Nature Portfolio wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Portfolio policies, see our Editorial Policies and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

- The exact sample size (n) for each experimental group/condition, given as a discrete number and unit of measurement
- A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
- The statistical test(s) used AND whether they are one- or two-sided
  *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
- A description of all covariates tested
- A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
- A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
- For null hypothesis testing, the test statistic (e.g. F, t, r) with confidence intervals, effect sizes, degrees of freedom and P value noted
  *Give P values as exact values whenever suitable.*
- For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
- For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
- Estimates of effect sizes (e.g. Cohen’s d, Pearson’s r), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection

- Genotyping arrays and calling algorithms are listed for all cohorts in ST2.

Data analysis

- Statistical analyses were performed with R version 4.2.1 [for simulations, visualization, regression and prediction analyses], PLINK v1.9 for standard quality control and GWAS analyses, GCTA v1.93 for estimating SNP-based heritability and for conditional analyses [COJO algorithm], GCTB v2.0 for running the SBayesC model used in prediction analyses, BOLT-LMM v2 and rvtest for mixed-model based GWAS analyses (versions listed in ST2), KING v2.2.5 for identifying relatives in certain cohorts, RAREMETAL v4.13.1 for running within-ancestry GWAS meta-analyses, LDSC v1.0.0 for LD score regression analyses, ImpG-Summary v1.0.1 for imputing GWAS summary statistics, SMR v1.03 for eQTL-based Mendelian Randomization analyses. Additional scripts used for analyses are listed in the Code Availability Statement.

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. Github). See the Nature Portfolio guidelines for submitting code & software for further information.

Data

Policy information about availability of data

- All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:
  - Accession codes, unique identifiers, or web links for publicly available datasets
  - A description of any restrictions on data availability
  - For clinical datasets or third party data, please ensure that the statement adheres to our policy

Summary statistics for ancestry-specific and multi-ancestry GWAS [excluding data from 23andMe] as well as SNP weights for polygenic scores derived in this study are made publicly available at https://portals.broadinstitute.org/collaboration/giant/index.php/GIANT_consortium_data_files. GWAS summary statistics derived
Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size: We meta-analyzed GWAS data from 281 cohorts, with the aim to reach the largest possible sample size (here, N=5.4 million). Our sample size is larger than any previous GWAS of height, thereby expected to deliver additional associations.

Data exclusions: Samples were excluded based on heterozygosity, low call rate, relatedness relative to other participants of the same cohort. Details for each cohort are given in ST2. We analyzed SNPs with a minor allele frequency >1% in at least one of five ancestry groups. Other quality control exclusions are listed in ST2.

Replication: We replicated associations detected in our study in N=49,160 individuals in the Estonian Biobank. Quantification of variance explained by height-associated SNPs and prediction accuracy was performed in 61,095 individuals independent of our discovery dataset from four cohorts (hold-out sample from the UK Biobank, the Lifelines Study, the Chinese Kadoorie Biobank and the PAGE Study).

Randomization: N/A. Our study is observational and used data from all available participants. No intervention was implemented in any of the study participants, therefore randomization was not required.

Blinding: N/A. No intervention was implemented in any of the study participants, therefore blinding was not required.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

| n/a | Involved in the study |
|-----|-----------------------|
| XXX | Antibodies            |
| XXX | Eukaryotic cell lines |
| XXX | Palaeontology and archaeology |
| XXX | Animals and other organisms |
| XX  | Human research participants |
| XXX | Clinical data         |
| XXX | Dual use research of concern |

Methods

| n/a | Involved in the study |
|-----|-----------------------|
| XX  | ChiP-seq              |
| XX  | Flow cytometry        |
| XX  | MRI-based neuroimaging |

Human research participants

Policy information about: studies involving human research participants

Population characteristics: The mean/standard deviation/ minimum and maximum values for height and age is given separately for males and females of each cohort in ST3.

Recruitment: Study participants were recruited under various designs including population based, prospective studies, birth cohorts, family-based, hospital-based, (nested) case-controls, case-cohorts, clinical trials. Designs for each cohort is listed in ST1.

Ethics oversight: Written informed consent was obtained from every participant in each study, and the study was approved by relevant ethics committees for each cohort. We provide a list of Institutional Review Boards in ST1.

Note that full information on the approval of the study protocol must also be provided in the manuscript.