A Saturated Map of Common Genetic Variants Associated with Human Height from 5.4 Million Individuals of Diverse Ancestries

ABSTRACT

Common SNPs are predicted to collectively explain 40-50% of phenotypic variation in human height, but identifying the specific variants and associated regions requires huge sample sizes. Here we show, using GWAS data from 5.4 million individuals of diverse ancestries, that 12,111 independent SNPs that are significantly associated with height account for nearly all of the common SNP-based heritability. These SNPs are clustered within 7,209 non-overlapping genomic segments with a median size of ~90 kb, covering ~21% of the genome. The density of independent associations varies across the genome and the regions of elevated density are enriched for biologically relevant genes. In out-of-sample estimation and prediction, the 12,111 SNPs account for 40% of phenotypic variance in European ancestry populations but only ~10%-20% in other ancestries. Effect sizes, associated regions, and gene prioritization are similar across ancestries, indicating that reduced prediction accuracy is likely explained by linkage disequilibrium and allele frequency differences within associated regions. Finally, we show that the relevant biological pathways are detectable with smaller sample sizes than needed to implicate causal genes and variants. Overall, this study, the largest GWAS to date, provides an unprecedented saturated map of specific genomic regions containing the vast majority of common height-associated variants.

INTRODUCTION

Since 2007, genome-wide association studies (GWAS) have identified thousands of associations between common single nucleotide polymorphisms (SNPs) and height, primarily using studies of European ancestry. The largest GWAS published to date for adult height focussed on common variation and reported up to 3,290 independent associations in 712 loci using a sample size of up to 700,000 individuals.1 To date, 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 GWAS 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 GWAS 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 GWAS. 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 prioritised genes and gene sets. Furthermore, because most GWAS continue to be performed largely in populations of European ancestry, it is important 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 GWAS, rather than implicating specific subsets of genes and biological pathways.²

Using data from 5,380,080 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, explore signal density across the genome, perform out-of-sample estimation and prediction analyses within European and non-European ancestry studies, and prioritise 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 European-ancestry populations and largely saturates both the genomic regions associated with height and broad categories of likely relevant gene sets; future work remains to extend prediction accuracy to non-European ancestries and to more definitively connect associated regions with individual likely causal genes and variants.

RESULTS

An overview of our study design and analysis strategy is illustrated in Suppl. Fig. 1.

Multi-ancestry GWAS 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, Inc. including 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-Americans 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/cohorts by the shorthand EUR, EAS, HIS, AFR, and SAS, respectively; yet recognising that these commonly used groupings oversimplify the actual genetic diversity among participants. Cohort-specific information is provided in Suppl. Tables 1 – 3. We tested the association between standing height and 1,385,132 autosomal bi-allelic SNPs from the HM3 tagging panel, which contains >1,095,888 SNPs with a minor allele frequency (MAF) >1% in each of the five ancestral groups included in our meta-analysis. Suppl. Fig. 2 shows the frequency distribution of HM3 SNPs across all five groups of cohorts.

We first performed individual meta-analyses in each of the five groups of cohorts. We identified 9863, 1888, 918, 493 and 69 quasi-independent genome-wide significant (GWS; P<5×10⁻⁸) SNPs in the EUR, HIS, EAS, AFR and SAS groups, respectively (Table 1; Suppl. Tables 4 – 8). Quasi-independent associations were obtained after performing approximate conditional and joint multiple-SNP (COJO) analyses,⁴ as implemented in GCTA⁵ (Suppl. Methods). Previous studies have shown that confounding due to population stratification may remain uncorrected in large EUR GWAS meta-analyses.⁶,⁷ Therefore, we specifically investigated confounding effects in our EUR GWAS and found no evidence that these GWAS results are driven by population stratification (Suppl. Note 1, Suppl. Fig. 3).
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 non-EUR groups. We found that over 83% of GWS SNPs detected in non-EUR are in strong linkage disequilibrium (LD; $r^2_{LD}>0.8$) with at least one variant reaching marginal genome-wide significance in EUR (Suppl. Tables 5 – 8) and over 87% of associations detected in non-EUR meta-analyses fall within 100 kb of at least one GWS SNP identified in EUR (Suppl. Fig. 4a). In contrast, a randomly sampled HM3 SNP falls within 100 kb of a EUR GWS SNP only about 68% of the time (standard error; S.E.=0.5% over 10,000 draws). Next, we quantified the cross-ancestry correlation of allele substitution effects ($\rho_b$) at GWS SNPs for all pairs of ancestry groups. We estimated $\rho_b$ using five sets of GWS SNPs identified in each of ancestry group. After correction for winner’s curse,$^6$9 we found $\rho_b$ to range between 0.64 and 0.99 across all pairs of ancestry groups and all sets of GWS SNPs (Suppl. Fig. 5 – 9). Thus, the observed GWS height associations are substantially shared across major ancestral groups, consistent with previous studies based on smaller sample sizes.$^{10,11}$

To find signals that are specific to certain groups, we tested if any individual SNPs detected in non-EUR GWAS are conditionally independent of signals detected in EUR GWAS. 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 ($r^2_{LD}>0.8$ in either ancestry group), we found that 2, 19, 49 and 143 of the GWS SNPs detected in SAS, AFR, EAS and HIS GWAS respectively are conditionally independent of GWS SNPs identified in EUR GWAS (Suppl. 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 increased statistical power of detection. The largest frequency difference relative to EUR was observed for rs2463169 (height-increasing G allele frequency: 23% in AFR vs. 84% in EUR) within the intron of PAWR, which codes for the prostate apoptosis response-4 protein. Interestingly, rs2463169 is located within the 12q21.2 locus, where a strong signal of positive selection in West-African Yoruba populations was previously reported.$^{12}$ The estimated effect at rs2463169 is $\beta \sim 0.034$ standard deviation (SD) per G allele in AFR vs. $\beta \sim 0.002$ SD/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 demonstrate a strong genetic overlap of GWAS signals across ancestries, we performed a fixed-effect meta-analysis of all five ancestry groups to maximise statistical power for discovering associations due to shared causal variants. The mean Cochran’s heterogeneity Q-statistic is ~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_FE) is ~36, and ~18% of all HM3 SNPs are marginally GWS. Moreover, we found allele frequencies in our META_FE to be very similar to that of EUR (mean $F_{ST}$ across SNPs between EUR and META_FE is ~0.001), as expected because our META_FE consists of >75% EUR participants and ~14% participants with admixed European and non-European ancestries (i.e. HIS and AFR). To further assess if LD in our META_FE could be reasonably approximated by the LD from EUR, we performed LD score regression analysis of our META_FE using LD scores estimated in EUR. In this analysis, we focused on the attenuation ratio statistic ($R_{LDSC-EUR}$), for which values >20% classically indicate strong LD inconsistencies between a given reference and GWAS summary statistics. For example, using EUR LD scores in the GWAS of HIS, which is the non-EUR group genetically closest to EUR ($F_{ST}~0.02$), yields an estimated $R_{LDSC-EUR}$ of ~25% (S.E. 1.8%), consistent with strong LD
differences between HIS and EUR. By contrast, in our META, we found an estimated R_{LDSC-EUR} of 
~4.5% (S.E. 0.8%), which is significantly lower than 20% and also not statistically different from 
3.8% (S.E. 0.8%) in our EUR meta-analysis. Altogether, our LD score regression analyses suggest 
that LD in our META 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, using genotypes from ~350,000 unrelated 
EUR participants of 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 (Suppl. Table 10). Of 
the GWS SNPs obtained from the non-EUR meta-analyses above that were conditionally 
independent of the EUR GWS SNPs, 0/2 in SAS, 5/19 in AFR, 27/49 in EAS, and 39/143 in HIS 
remained statistically significant in our META (Suppl. Table 9), meaning that a small number of 
additional signals were only identified in the ancestry-specific analyses.

We next sought replication of the 12,111 META signals using GWAS data from 49,160 
participants of the Estonian Biobank (EBB). We first re-assessed the consistency of allele 
frequencies between our META and the EBB set. We found a correlation of allele frequencies of 
~0.98 between the two datasets and a mean F_{ST} across SNPs of ~0.005, similar to estimates 
obtained between 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 
MAF>1% (Suppl. 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 (Suppl. Fig. 10a). Instead, we assessed the correlation of SNP effects 
between our discovery and replication GWAS as an overall metric of replicability. Over the 
11,529/11,847 SNPs with a MAF>1% in the EBB, we found a correlation of marginal SNP effects of 
\( \rho_b = 0.93 \) (jackknife standard error; S.E. 0.01) and a correlation of conditional SNP effects using the 
same LD reference panel of \( \rho_B = 0.80 \) (S.E. 0.03; Suppl. Fig. 11). 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 (Suppl. Methods; Suppl. Fig. 10b), that 60% of them have a marginal 
SNP effect that is sign-consistent with that from our discovery GWAS (Fisher exact test; \( P=0.001 \)). 
The proportion of sign-consistent SNP effects was >75% (Fisher exact test; \( P<10^{-50} \)) for variants 
with a MAF>1%, also consistent with expectations (Suppl. Fig. 10b). 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, we defined a 
measure of local density of association signals for each GWS SNP based on the number of additional 
independent associations within 100 kb (Suppl. Fig. 12). We observed that 69% of GWS SNPs 
shared their location with another associated, conditionally independent, GWS SNP (Fig. 1). The 
mean signal density across the entire genome is 2.0 (LOCO-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 
harbouring pathogenic mutations causing syndromes of abnormal skeletal growth (“OMIM genes”; 
Suppl. Methods; Suppl. 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 $>1$: 2.5×; $P<0.001$; Suppl. Methods, Suppl. Fig. 13a).\textsuperscript{15,16} Interestingly, the enrichment of OMIM genes almost linearly increases with the density of height-associated SNPs (Suppl. Fig. 13b). Thus, these 12,111 GWS SNPs nonrandomly cluster near each other and also 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-localise within 100 kb of one another (Fig. 1; Suppl. Fig. 14). We show in Suppl. Note 2 and Suppl. Figs. 14-15, using haplotype- and simulation-based analyses, that a multiplicity of independent causal variants is the most likely explanation of this observation. Interestingly, we also found that signal density is partially explained by the presence of a recently identified\textsuperscript{17,18} height-associated variable-number-of-tandem-repeat (VNTR) polymorphism at this locus (Suppl. Note 2). In fact, the 25 independent GWS SNPs clustered within 100 kb of rs4932198 explain $>40\%$ of the VNTR length variation in multiple ancestries (Suppl. Fig. 15e) and an additional $\sim0.24\%$ ($P=8.7 \times 10^{-55}$) phenotypic variance in EUR above what is explained by the VNTR alone (Suppl. Fig. 15f). Altogether, our conclusion is consistent with prior evidence of multiple types of common variation influencing height through ACAN gene function, involving multiple enhancers,\textsuperscript{19} missense variants\textsuperscript{20} and tandem repeat polymorphisms.\textsuperscript{17,18}

Variance explained by SNPs within identified loci

To quantify the proportion of height variance explained by GWS SNPs identified in our META\textsubscript{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 containing at least 1 GWS SNP, such that GWS SNPs in adjacent loci are $>2\times35$ kb away from each other (i.e. 35 kb window on each side). We chose a 35 kb threshold based on findings from Wu et al.\textsuperscript{21} who previously showed that causal common variants are located within 35 kb of GWS SNPs with $>80\%$ probability. Accordingly, we grouped the 12,111 GWS SNPs identified in our META\textsubscript{FE} into 7,209 non-overlapping loci (Suppl. Table 12) with lengths ranging from 70 kb (for loci containing only 1 signal) to 711 kb (for loci containing up to 25 signals). The average length of GWS loci is $\sim90$ kb (SD 46 kb). The cumulative length of GWS loci represent $\sim647$ Mb, or $\sim21\%$ of the genome (assuming a genome length of $\sim3039$ Mb).\textsuperscript{22}

To estimate what fraction of heritability 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 ($h^2_{\text{SNP}}$) of height in 8 independent samples of all five population groups represented in our META\textsubscript{FE} (Fig. 2; Suppl. Methods). Altogether, our stratified estimation of SNP-based heritability shows that SNPs within these 7,209 GWS loci explain $\sim100\%$ of $h^2_{\text{SNP}}$ in EUR and $\sim90\%$ of $h^2_{\text{SNP}}$ across all non-EUR groups, despite being drawn from less than a quarter of the genome (Fig. 2). 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 (Suppl. Fig. 16).

To further assess the robustness of this key result, we tested if 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 ($r_h=0.1$, ref.\textsuperscript{23}) and found no significant
enrichment of SNP-based heritability for BMI within height-associated GWS loci (Suppl. Fig. 17). Furthermore, we repeated our analysis using a random set of SNPs with similar EUR MAF and LD scores as the 12,111 height-associated GWS SNPs. We found this control set of SNPs to explain only ~27% of $h^2_{\text{SNP}}$ for height, consistent with the proportion of SNPs within the loci defined by this random set of SNPs (Suppl. Figs. 16 - 17). Finally, we extended our stratified estimation of SNP-based heritability to all well-imputed common SNPs (i.e. beyond the HM3 panel) and found, consistently across population groups, that although more genetic variance can be explained by common SNPs not included in the HM3 panel, all information remains concentrated within these 7,209 GWS loci (Suppl. Fig. 18). 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 ~21% of the genome.

Out-of-sample prediction accuracy

We quantified the accuracy of polygenic scores (PGS) for height based on GWS SNPs in 61,095 unrelated individuals from 3 studies, including 33,001 participants of the UKB who were not included in our discovery GWAS (i.e. 14,587 EUR; 9,257 SAS; 6,911 AFR and 2,246 EAS; Suppl. Methods), 14,058 EUR participants from the Lifelines cohort study; and 8,238 HIS and 5,798 AFR participants from the PAGE study. Prediction accuracy ($R^2_{\text{GWS}}$) was defined as the squared correlation between the PGS and actual height (corrected for mean and variance sex differences and 20 genotypic principal components). We found that PGS based on 12,111 GWS SNPs from our META$^2$FE systematically outperformed those based on GWS identified in ancestry-specific meta-analyses (Fig. 3a). The only exception was in EUR where both PGS performed equally. The largest prediction accuracy was observed in EUR participants ($R^2_{\text{GWS}}$~40%; S.E. 0.6%) and the smallest one in AFR participants from the UKB ($R^2_{\text{GWS}}$~9.4%; S.E. 0.7%). Note that the difference in $R^2_{\text{GWS}}$ between the EUR and AFR ancestry cohorts is expected because of the over-representation of EUR in our META$^2$FE and consistent with a relative accuracy ($R^2_{\text{GWS}}$ in AFR)/($R^2_{\text{GWS}}$ in EUR) of ~25% previously reported.24 Nevertheless, we found the accuracy of PGS based on GWS from our multi-ancestry META$^2$FE to be consistently larger than that of PGS based on GWS SNPs from a EUR GWAS (Fig. 3a). The largest improvement was observed in AFR, where the meta-analysed accuracy in AFR participants of UKB and PAGE was increased from $R^2_{\text{GWS}}$=6.6% (S.E. 0.4%) to $R^2_{\text{GWS}}$=10.8% (S.E. 0.5%), i.e. almost a ~1.6-fold improvement. This observation is partly explained by the increased statistical power but also by the refined estimation of SNP effects due to the inclusion of shorter and ancestry-specific LD blocks in AFR cohorts.

Furthermore, we sought to evaluate the prediction accuracy of PGS 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 (i.e. two parents and one offspring) and 17,492 independent EUR sibling pairs from the UKB, who were excluded from our META$^2$FE. We found that height of any first-degree relative yields a prediction accuracy between 25% and 30% (Fig. 3b). Moreover, the accuracy of the parental average is ~44% (S.E. 3.2%), which is larger but not significantly different from $R^2_{\text{GWS}}$ in EUR. In addition, we found that a linear combination of the average height of parents and of the offspring’s PGS yields an unprecedented accuracy of 54% (S.E. 3.2%). This observation reflects the fact that PGS can explain within-family differences between siblings, while average parental height cannot. To show this empirically, we estimate that our PGS based on GWS SNPs explain ~33% (S.E. 0.7%) of height variance between siblings (Suppl. Methods). Finally, we demonstrate that the optimal weighting between parental average and PGS
can be predicted theoretically as function of $R^2_{GWS}$, the full narrow sense heritability and the phenotypic correlation between spouses (Suppl. Note 3, Suppl. Fig. 19).

In summary, the estimation of variance explained and prediction analyses in European-ancestry samples show that the set of 12,111 GWS SNPs account for nearly all of $h^2_{SNP}$ and that combining SNP-based PGS with family history significantly improves prediction accuracy. In contrast, both estimation and prediction results show clear attenuation in samples with non-European ancestry, consistent with previous studies.\textsuperscript{24-27}

**Relationship between GWAS discoveries, sample size and ancestry diversity**

Our large study offers a unique opportunity to empirically quantify how increasing GWAS sample sizes and ancestry diversity affects discovery of variants, genes and biological pathways. To address this question, we re-analysed 3 previously published GWAS of height\textsuperscript{1,15,16} and also down-sampled our meta-analysis into 4 subsets (including our EUR and META\textsubscript{FE} GWAS). Altogether we analysed 7 GWAS with a sample size increasing from ~0.13 M up to ~5.3 M individuals (Table 2).

For each GWAS, we quantified 8 metrics grouped into 4 variant- and locus-based metrics (number of GWS SNPs, number of GWS loci, prediction accuracy ($R^2_{GWS}$) of PGS based on GWS SNPs, the proportion of the genome covered by GWS loci), a functional annotation-based metric (enrichment statistics from stratified LDSC\textsuperscript{28,29}), 2 gene-based metrics (number of genes prioritised by Summary data based Mendelian Randomization\textsuperscript{30} (SMR; Suppl. Methods), proximity of variants with OMIM genes), and a gene-set-based metric (enrichment within clusters of gene sets/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 prioritisation methods (DEPICT\textsuperscript{31} and MAGMA\textsuperscript{32}), we found that the same broad clusters of related gene sets (including most of the clusters enriched for OMIM genes) are prioritised at all GWAS sample sizes (Suppl. Figs. 20-21; Suppl. Tables 13 – 15; Suppl. Note 4). Similarly, stratified LDSC estimates of heritability enrichment within 97 functional annotations also remain stable across the range of sample sizes (Suppl. Fig. 22). 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 demonstrating that GWAS expectedly implicate similar biology across major ancestral groups (Suppl. Note 4; Suppl. Fig. 23).

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 N>1.5M; while the relative enrichment of GWS SNPs near OMIM genes first decreases with sample size, then plateaus when N>1.5M (Suppl. Figs. 24a-c). Interestingly, the decrease observed for N<1.5M reflects the preferential localization of larger effect variants (those identified with smaller sample sizes) closer to OMIM genes (Suppl. Fig. 24d) and, conversely, that more recently identified variants with smaller effects tend to localize further away from OMIM genes (Suppl. Fig. 24e). We also investigated the number of genes prioritised using Summary-data based Mendelian...
Randomization (hereafter referred to as SMR genes; Suppl. Methods) using expression quantitative trait loci (eQTL) as genetic instruments (Suppl. Table 16) as an alternative gene-level metric and found it to saturate for N>4M (Suppl. Fig. 24f). Note that saturation of SMR genes is partly affected by the biological relevance and statistical power of eQTL studies. Therefore, we can expect more genes to be prioritised when integrating GWAS summary statistics from this study with that 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 reported. 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>2.5M, where the upward trend starts to weaken (Suppl. Fig. 25a). 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 (Suppl. Fig. 25b). The observed saturation in PGS prediction accuracy (both within ancestry, i.e. in EUR; and multi-ancestry) was more noticeable than that of the number and genomic coverage of GWS loci. In fact, increasing sample size from 2.5M to 4M by adding another 1.5M EUR samples increased the number of GWS SNPs from 7,020 to 9,863 (i.e. (9,863-7,020)/7,020 = ~1.4-fold increase) 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 EAS and AFR individuals. However, adding another ~1M participants of non-EUR improves the multi-ancestry prediction accuracy by over +3.4% (Suppl. Fig. 25c), highlighting the value of non-EUR populations for this purpose.

Altogether, these analyses show that increasing GWAS sample size not only increases 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/pathways and functional annotations can be identified using relatively small sample sizes (N~0.25M for height). Importantly, we confirm that increased genetic diversity in GWAS discovery samples significantly improves the prediction accuracy of PGS in under-represented ancestries.

DISCUSSION

By performing the largest GWAS to date in 5,380,080 individuals with a primary focus on common genetic variation, we have provided new 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 signal density of associations (known and novel) are not randomly distributed across the genome; rather, associated variants are more likely detected around genes previously associated with Mendelian disorders of growth. Furthermore, we observed strong genetic overlap of association across cohorts of various continental ancestries. Effect estimates are moderately to highly correlated (min=0.64, max=0.99), and while there are significant differences in power to detect an association between cohorts with European and non-European ancestries, the majority of 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 >7-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 >90% of $h^2_{SNP}$ across diverse populations when incorporating all common SNPs within 35 kb of GWS SNPs. This result is important as it highlights that future investigation of common (MAF>1%) genetic variation associated with height in many ancestries will most likely detect signals within the 7,209 GWS loci identified in the present study. An interesting future question is whether rare genetic variants associated with height are also concentrated within the same loci. Of note, previous studies have reported significant enrichment of height heritability near genes as compared to intergenic regions (e.g. up to >50 kb away from start/stop genomic position of genes). Our findings are consistent but not reducible to that observation, given that up to ~31% of GWS SNPs identified in this study lie >50 kb away from any gene.

Our study provides a powerful genetic predictor of height based on 12,111 GWS SNPs, for which accuracy reaches ~40% (i.e. 80% of $h^2_{SNP}$) in individuals of European ancestries and up to ~10% in individuals of predominantly African ancestries. Importantly, we show using a new method developed by Wang and colleagues\(^{27}\) that LD and MAF differences between European and African ancestries can explain up to ~84% (S.E. 1.5%) of the loss of prediction accuracy between these populations (Suppl. Methods), with the remaining loss being presumably explained by heritability differences between populations and/or differences in effect sizes across populations (e.g., due to gene-by-gene or gene-by-environment interactions). This observation is consistent with common causal 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 PGS for non-European ancestry populations. Our study has a large number of participants with African ancestries as compared with previous efforts. However, we emphasise that further increasing the size of GWAS in non-European ancestry populations, 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 samples size would help to identify potential ancestry-specific causal variants, to facilitate ancestry-specific fine mapping, and to inform gene by environment/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 European ancestry populations.

Although large sample sizes are needed to pinpoint the common variants responsible for the heritability of height (and larger samples in multiple ancestries will likely 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 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 likely causal genes and even mapping of associated regions is consistent across ancestries, reflecting the expected similarity in the biological architecture of human height across populations.

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 from our HM3 SNPs based analyses (Suppl. Fig. 18). This result underlines the widespread allelic heterogeneity at height-associated loci. Another limitation of our study is that we determined conditional associations using an EUR LD reference (N~350,000), which is sub-optimal given that ~24% of our discovery sample is of non-EUR. We emphasise that no analytical tool with an adequately large multi-ancestry reference panel currently is 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, due to current limited availability of such data. Finally, as with all GWAS, definitive identification of effector 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 will be mostly driven by the availability of relevant complementary data (e.g., context-specific eQTL in relevant tissues and cell-types) and the power of computational methods that can integrate these data.

In summary, our study has been able to demonstrate 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 ~5 million participants of various ancestries provide enough power to map >90% of genetic variance explained by common SNPs down to ~21% of the genome. 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 enabling 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.
REFERENCES


Malmö, Sweden, 157Department of Computer Science, University of Toronto, Toronto, ON M5S 2E4, Canada, 158Department of Anthropology, University of Toronto at Mississauga, Mississauga, ON L5L 1C6, Canada, 159Institute of Health and Biomedical Innovation, The University of Queensland Genomics Research Centre, Centre for Genomics and Personalised Health, School of Biomedical Sciences, Qld University of Technology, 60 Musk Ave, Kelvin Grove Qld, Australia, 160Oneomics, Soonchunhyang Mirai Medical Center, Gyeonggi-do, 14585, Korea, 161Laboratory of Neurogenetics, National Institute on Aging, National Institutes of Health, Bethesda, MD 20892, USA, 162Center for Alzheimer’s and Related Dementias, National Institutes of Health, Bethesda, MD 20892, USA, 163Data Tecnica International, Glen Echo, MD 20812, USA, 164Department of Population and Quantitative Health Sciences, Case Western Reserve University, Cleveland, OH 44106, USA, 165current address: Department of Mathematics and Statistics, St. Cloud State University, St Cloud, MN 56301, USA, 166Section of Computational Biomedicine, Department of Medicine, Boston University School of Medicine, Boston, MA 02118, USA, 167Center for Geriatrics and Gerontology, Taichung Veterans General Hospital, Taichung, Taiwan, 168Montreal Heart Institute, Montreal, Quebec, H1T 1C8, Canada, 169Department of Ophthalmology, Radboud University Medical Center, Nijmegen, 6525 GA, the Netherlands, 170MRC Cardiovascular Epidemiology Unit, Department of Public Health and Primary Care, University of Cambridge, Cambridge, 171Department of Nutritional Sciences, Lund University Diabetes Centre, Malmö, Sweden, 172Department of Clinical Science, Center for Diabetes Research, University of Bergen, Bergen, Norway, 173Department of Clinical Sciences, Lund University Diabetes Centre, Malmö, Sweden, 174Department of Clinical Chemistry, Fimlab Laboratories, Tampere 33520, Finland, 175Department of Clinical Chemistry, Finnish Cardiovascular Research Center - Tampere, Faculty of Medicine and Health Technology, Tampere University, Tampere 33014, Finland, 176Department of Cardiology, Heart Center, Tampere University Hospital, Tampere 33521, Finland, 177National and Kapodistrian University of Athens, Dromokaitieo Psychiatric Hospital, Athens, Greece, 178Department of Twin Research and Genetic Epidemiology, King’s College London, London SE1 7EH, UK, 179NIHR Biomedical Research Centre at Guy’s and St Thomas’ Foundation Trust, London SE1 9RT, UK, 180Center for Public Health Genomics, University of Virginia School of Medicine, Charlottesville, VA 22908, USA, 181MRC Human Genetics Unit, Institute of Genetics and Cancer, University of Edinburgh, Western General Hospital, Edinburgh EH4 2XU, Scotland, 182Department of Psychiatry and Department of Community Health and Epidemiology, Dalhousie University, Halifax, NS, Canada, 183Institute of Psychiatric Phenomics and Genomics (IPPG), University Hospital, LMU Munich, Munich, Germany, 184Institute for Medical Informatics, Biometry and Epidemiology, University Hospital Essen, Essen, 45130, Germany, 185Programs in Metabolism and Medical and Population Genetics, Broad Institute of Harvard and MIT, Cambridge, MA 02142, USA, 186Diabetes Unit, Massachusetts General Hospital, Boston, MA 02114, USA, 187Center for Genomic Medicine, Massachusetts General Hospital, Boston, MA 02114, USA, 188Department of Psychiatry, Amsterdam Public Health and Amsterdam Neuroscience, Amsterdam UMC/Vrije Universiteit, Amsterdam, 1081 HL, the Netherlands, 189Biomedical and Translational Informatics Institute, Geisinger, Danville, PA 17821, USA, 190Department of Genetics, Institute for Biomedical Informatics, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA 19104, USA, 191MRC Population Health Research Unit, Nuffield Department of Population Health, University of Oxford, Oxford OX3 7LF, UK, 192Population Health Sciences, Bristol Medical School, University of Bristol, Bristol, BS8 2BN, UK, 193Institute for Cardiogenetics, University of Lübeck, 23562, Lübeck, Germany, 194Public Health Informatics Unit, Department of Integrated Health Sciences, Nagoya University Graduate School of Medicine, Nagoya, 461-8673, Japan, 195Centre for Bone and Arthritis Research, Department of Internal Medicine and Clinical Nutrition, Institute of Medicine, Sahlgrenska Academy, University of Gothenburg, Gothenburg, Sweden, 196Bioinformatics Core Facility, Sahlgrenska Academy, University of Gothenburg, Gothenburg, Sweden, 197Korea Institute of Science and Technology, Gangneung Institute of Natural Products, Gangneung, Gangwon-do, 25451, Republic of Korea, 198Department of Clinical Biochemistry, Lillebaelt Hospital, Kolding, Denmark, 199Department of Human Genetics, Wellcome Sanger Institute, Hinxton, CB10 1SA, UK, 199Department of Internal Medicine, Section of Gerontology and Geriatrics, Leiden University Medical Center, Leiden, the Netherlands, 200Institute of Genetics and Biophysics A. Buzzati-Traverso, CNR, Naples, 80131 Italy, 201Institute of Genomics, University of Tartu, 51010, Tartu, Estonia, 202Department of Clinical Biochemistry and Immunology, Hospital of Southern Jutland, 6200 Aabenraa, Denmark, 203The National Centre for Register-based Research, University of Aarhus, Aarhus, Denmark, 204Centre for Population Health Research, University of Turku and Turku University Hospital, Turku 20014, Finland, 205Research Centre of Applied and Preventive Cardiovascular Medicine, University of Turku, Turku 20014, Finland, 206Medical School, University of Split, Šoltanska 2, 21000 Split, Croatia, 207Algebra University College, Ilica 242, Zagreb, Croatia, 208The Charles Bronfman Institute for Personalized Medicine, Icahn School of Medicine at Mount Sinai, New York, NY 10029, USA, 209Department of Environmental Medicine and Public Health, Icahn School of Medicine at Mount Sinai, New York, NY 10029, USA, 210Center for Non-Communicable Diseases, Karachi,
Vincent's Hospital, Darlinghurst, NSW, 2010, Australia, 320Faculty of Medicine, UNSW Sydney, Kensington, NSW 2052, Australia, 321Department of Preventive Medicine, Keck School of Medicine of USC, Los Angeles, CA 90089, USA, 322Translational Gerontology Branch, National Institute on Aging, National Institutes of Health, Baltimore, MD 21224, USA, 323Robertson Center for Biostatistics, University of Glasgow, UK, 324Human Genetics Center, School of Public Health, University of Texas Health Science Center at Houston, Houston, TX 77030, USA, 325Department of Public Health and Clinical Medicine, Umeå University, Umeå, Sweden, 326Department of Nutrition, Harvard T.H. Chan School of Public Health, Boston, MA, 327Department of Internal Medicine, Wake Forest School of Medicine, Medical Center Boulevard, Winston-Salem, NC 27157, USA, 328Faculty of Veterinary and Agricultural Science, Na, Parkville, Victoria, 3010, Australia, 329Agriculture Victoria Research, Department of Jobs, Precincts and Regions, Bundoora, Victoria, 3083, Australia, 330Injury Prevention Research Center, University of North Carolina at Chapel Hill, Chapel Hill, NC 27599, USA, 331Division of Physical Therapy, University of North Carolina at Chapel Hill, Chapel Hill, NC 27599, USA, 332Centro de Investigacion en Salud Poblanol Instituto Nacional de Salud Publica and Centro de Estudios en Diabetes, Ahuacatlan CP 62100 Cuernavaca Morelos, Mexico, 333Division of Human Genetics, Children’s Hospital of Philadelphia, Philadelphia, PA 19104, USA, 334Departments of Pediatrics and Genetics, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA 19104, USA, 335Division of Endocrinology and Diabetes, Children’s Hospital of Philadelphia, Philadelphia, PA 19104, USA, 336Faculty of Medicine, University of Iceland, 101 Reykjavik, Iceland, 337Department of Pediatrics, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA 19104, USA, 338Division of Pulmonary Medicine, Children’s Hospital of Philadelphia, Philadelphia, PA 19104, USA, 339Institute of Biomedical and Clinical Science, University of Exeter Medical School, Exeter, UK, 340Cardiovascular Health Research Unit, Department of Epidemiology, University of Washington, Seattle, WA 98195, USA, 341Department of Paediatrics, Yong Loo Lin School of Medicine, National University of Singapore, Singapore, 342Khooo Teck Puat - National University Children’s Medical Institute, National University Health System, Singapore, 343Department of Cardiology, Internal Medicine II, Medical University of Vienna, 1090 Vienna, Austria, 344Menzies Research Institute Tasmania, University of Tasmania, Hobart, 7000, Tasmania, Australia, 345Centre for Eye Research Australia, Royal Victorian Eye and Ear Hospital, University of Melbourne, Melbourne, 3002, Victoria, Australia, 346Centre for Ophthalmology and Vision Science, Lions Eye Institute, University of Western Australia, Perth, 6009, Western Australia, Australia, 347Cardiology Division, Massachusetts General Hospital, Boston, MA 02114, USA, 348Department of Genetics, Shanghai-MOST Key Laboratory of Health and Disease Genomics, Chinese National Human Genome Center and Shanghai Industrial Technology Institute, Shanghai, China, 349Department of Internal Medicine, University of Utah, Salt Lake City, UT 84132, USA, 350Australian Centre for Precision Health, Clinical and Health Sciences, University of South Australia, Adelaide, Australia, 351South Australian Health and Medical Research Institute, Adelaide, Australia, 352Department of Environmental and Preventive Medicine, Jichi Medical University School of Medicine, Shimotsuke, 329-0498, Japan, 353Department of Epidemiology and Population Health, Albert Einstein College of Medicine, Bronx, NY 10461, USA, 354Division of Endocrinology, Diabetes and Metabolism, School of Medicine, Ohio State University, Columbus OH 43210, USA, 355Center for Life Course Health Research, Faculty of Medicine, University of Oulu, FI-90014 Oulun yliopisto, Finland, 356Unit of Primary Health Care, Oulu University Hospital, OYS, 90220 Oulu, Finland, 357Department of Life Sciences, College of Health and Life Sciences, Brunel University London, Uxbridge, Middlesex UB8 3PH, UK, 358Beijing Institute of Ophthalmology, Beijing Tongren Eye Center, Beijing Tongren Hospital, Capital Medical University, Beijing Ophthalmology and Visual Sciences Key Laboratory, 100730 Beijing, China, 359The Eye Hospital, School of Ophthalmology & Optometry, Wenzhou Medical University, Wenzhou, Zhejiang 325027, China, 360Einhoven Laboratory for Experimental Vascular Medicine, LUMC, Leiden, the Netherlands, 361Netherlands Heart Institute, Utrecht, the Netherlands, 362Department of Clinical Physiology, Tampere University Hospital, Tampere 33521, Finland, 363Department of Clinical Physiology, Finnish Cardiovascular Research Center - Tampere, Faculty of Medicine and Health Technology, Tampere University, Tampere 33014, Finland, 364Laboratory of Complex Trait Genomics, Department of Computational Biology and Medical Sciences, Graduate School of Frontier Sciences, The University of Tokyo, Tokyo, Japan, 365Department of Ophthalmology, The Catholic University of Korea Incheon St. Mary’s Hospital, Incheon, 21431, Republic of Korea, 366NIHR Oxford Biomedical Research Centre, Churchill Hospital, Oxford, Oxford, UK, 367German Centre for Cardiovascular Research (DZHK), partner site Munich Heart Alliance, Munich, Germany, 368Radboud University Medical Center, Radboud Institute for Health Sciences, Department of Urology, Nijmegen, the Netherlands, 369Yonsei University Severance Eye Hospital, Seodaemun-gu, Seoul, 03718, Republic of Korea, 370Department of Biochemistry, College of Medicine, Ewha Womans University, Seoul, 07804, Republic of Korea, 371Department of Cardiology, University Heart and Vascular Center UKE Hamburg, Hamburg, Germany, 372Institute of Cardiovascular Sciences, College of Medical and Dental Sciences, University of Birmingham,
Genomics Network, 484 Lifelines Cohort Study, Groningen, the Netherlands, 485 Regeneron Genetics Center, Tarrytown, NY 10591, USA, 486 The Institute of Cancer Research, London, SM2 5NG, UK, 487 Understanding Society Scientific Group, 488 Vanderbilt Genetics Institute, Division of Genetic Medicine, Vanderbilt University Medical Center, Nashville, TN 37232, USA, 489 Department of Epidemiology, Emory University Rolls School of Public Health, Atlanta, GA 30322, USA, 490 Atlanta VA Health Care System, Decatur, GA 30033, USA, 491 Princess Al-Jawhara Al-Braham Centre of Excellence in Research of Hereditary Disorders (PACER-HD), King Abdulaziz University, Jeddah, Saudi Arabia, 492 Department of Population Health Sciences, Geisinger, Danville, PA 17822, USA, 493 The Mindich Child Health and Development Institute, Icahn School of Medicine at Mount Sinai, New York, NY 10029, USA, 494 The Novo Nordisk Center for Basic Metabolic Research, University of Copenhagen, Copenhagen, Denmark, 495 School of Life Sciences, Westlake University, Hangzhou, Zhejiang 310024, China, 496 Westlake Laboratory of Life Sciences and Biomedicine, Hangzhou, Zhejiang 310024, China, 497 Department of Human Genetics, University of Michigan, Ann Arbor, MI 48109, USA, 498 McDonnell Genome Institute and Department of Medicine, Washington University School of Medicine, St. Louis, MO, 63110, USA, 499 Laboratory of Statistical Immunology, Immunology Frontier Research Center (WPI-IFReC), Osaka 565-0871, Japan, 500 Integrated Frontier Research for Medical Science Division, Institute for Open and Transdisciplinary Research Initiatives, Osaka University, Osaka 565-0871, Japan, 501 Programs in Metabolism and Medical and Population Genetics, Broad Institute of MIT and Harvard, Cambridge, MA 02142, USA, 502 Departments of Pediatrics and Genetics, Harvard Medical School, Boston, MA 02115.

ACKNOWLEDGEMENTS

We gratefully acknowledge the participants in each cohorts contributing to this study. Additional acknowledgements are listed in Supplementary information. Support for title page creation and format was provided by AuthorArranger, a tool developed at the National Cancer Institute.

This research was supported by the following funding bodies.

US National Institutes of Health:
75N92021D00001, 75N92021D00002, 75N92021D00003, 75N92021D00004, 75N92021D00005, AA07535, AA10248, AA014041, AA13220, AA13321, AA13326, DA12854, U01 DK062418, HHSN26820180005I, HHSN26820180007I, HHSN26820180003I, HHSN26820180006I, HHSN26820180004I, R01 CA55069, R35 CA53890, R01 CA80205, R01 CA14403, HHSN268201200008I, EY022310, 1X01HG006934-01, R01DK118427, R21DK105913, HHSN268201200036C, HHSN26820080007C, HHSN268200800009C, HHSN268201800081C, N01HC55222, N01HC58079, N01HC58059, N01HC58081, N01HC58082, N01HC58083, N01HC58086, 75N92021D0006, U01HL080295, R01HL085251, R01HL087652, R01HL105756, R01HL103612, R01HL120939, U01HL130114, R18AG023629, U1LRR01881, DK063491, R01 HL09506, 1R01HL139731, 1R01HL092577 (P.T.E.), K24HL105780 (P.T.E.), HHSRC200872096C, R01 DK087914, R01 DK066358, R01 DK053591, K101HG010155 (A.V.K.), U01HG011719 (A.V.K.), U01 HG004436, P30 DK072488, HHSN26820072096C, U01 HG00444, R01 NS45012, U01 NS609280-01, R01-NS114045 (J.W.C.), R01-NS100178 (J.W.C.), R01-NS105150 (J.W.C.), HL043851, HL084067, CA04798, UM1CA182913, U01HG008657, U01HG008685, U10HG008672, U10HG008666, U10HG006379, U10HG008679, U10HG008680, U10HG008673, U10HG008701, U10HG008676, U10HG008664, U54MD007593, U1LRR01878, R01-DK062370 (M.B.), R01-DK072193 (K.L.M.), intramural project number 1201-HG000024 (F.S.C.), N01-HG-65403, DA044283, DA02755, DA037904, AA09367, DA05147, DA036216, 5-P60-AR03701, 5-P60-AR49465, N01-AG-1-2100, HHSN27120120002C, National Institute on Aging Intramural Research Program, R-35-HL135824 (C.J.W.), AA-12502, AA-00145, AA-09203, AA15416, K02AA180755, UM1-CA186107, P01 CA87969, R01 CA49449, U01 CA176726, CA141298, U01CA055075, CA141298, HL54471, HL54472, HL54473, HL54474, HL54495, HL54496, HL54509, HL54515, U24MH08457-06, R01D0042157-01A1, R01MH58979-03, MO01880, 1RC2MH09951-01, 1RC2 MH09995, R01DK102127-04, R01DK110113 (R.J.F.L.), R01DK107786 (R.J.F.L.), R01HL142302 (R.J.F.L.), R01DK10297 (R.J.F.L.), R01DK124097 (R.J.F.L.), R01HL15152 (R.J.F.L.), R01-ML046380, K22-ML046380, R05-ML135818, R01-ML113338, R35HL135818 (Su.R.), HL046389 (Su.R.), and HL113338 (Su.R.), R01 HL135405 (B.E.C.), R03 HL154284 (B.E.C.), R01HL086718, HG011052 (X.Zhu), N01-RC-6-2478, R01-DK122503, U01AG023746, U01AG023712, U01AG023749, U01AG023755, U01AG023744, U19AG03893, R01-DK089256, R01HL117078, R01HL093571, R01 HL093571, R01 HL104135, R37-HL045508, R01-HL053353, R01-DK075787.
German Federal Ministry of Education and Research:
01ZZ9603, 01ZZD103, 01ZZD403, 03SZT061A, 03ZK012, 01EA1801A (G.E.D.), 01ER0804 (K-U.E.), BMBF 01ER1206 and BMBF 01ER1507 (I.M.H.), BMBF projects 01EG0401, 01GI0856, 01GI0860, 01GS0820_WB2-C, 01ER1001D, 01GI0205.

Additional funding:
The University of Newcastle Strategic Initiatives Fund; the Gladys M Brawn Senior Research Fellowship scheme; Vincent Fairfax Family Foundation; The Hunter Medical Research Institute; the Nagahama City Office and the Zeroji Club; the Center of Innovation Program, the Global University Project from the Ministry of Education, Culture, Sports, Science and Technology of Japan; the Practical Research Project for Rare/Intractable Diseases (ek0109070, ek0109283, ek0109196, ek0109348), and the Program for an Integrated Database of Clinical and Genomic Information (ek0205008), for the Japan Agency for Medical Research and Development; Takeda Medical Research Foundation; Astellas Pharma, Inc.; Daiichi Sankyo Co., Ltd.; Mitsubishi Tanabe Pharma Corporation; Otsuka Pharmaceutical Co., Ltd.; Taisho Pharmaceutical Co., Ltd.; and Takeda Pharmaceutical Co., Ltd.; Type 1 Diabetes Genetics Consortium; the French Ministry of Research; the Chief Scientist Office of the Scottish Government #CZB/4/276 and #CZB/4/710; Arthritis Research UK; Royal Society URF (J.F.W.); the Atlantic Philanthropies; the UK Economic and Social Research Council awards ES/L008459/1 and ES/L008459/1; the UKCRC Centre of Excellence for Public Health Northern Ireland; the Centre for Ageing Research and Development in Ireland; the Office of the First Minister and Deputy First Minister; the Health and Social Care Research and Development Division of the Public Health Agency; the Wellcome Trust/Wolfgang Foundation; and Queen's University Belfast; the Science Foundation Ireland-Department for the Economy Award 15/IA/3152 (NICOLA); NI HSC R&D division STL/5569/19 (L.Sm.); the Italian Ministry of Education, University and Research (MIUR) number 5571/DSPAR/2002 (OGP study); GlaxoSmithKline; the Faculty of Biology and Medicine of Lausanne; the Swiss National Science Foundation grants 33CSCO-122661, 33CS30-139468, 33CS30-148401 and 33CS30-177535/1; the Montreal Heart Institute Biobank; the Canadian Institutes of Health Research Project PT #156248; the Canada Research Chair Program, Genome Quebec and Genome Canada, and the Montreal Heart Institute Foundation (G.L.); the Strategic Priority CAS Project grant number XDB38000000, Shanghai Municipal Health and Family Planning Commission grant number 17SHZX00301, and the National Natural Science Foundation of China grant number 81970684; the National Medical Research Council (grants 0796/2003, 1176/2008, 1149/2008, STA/R/0003/2008, 1249/2010, CG/SERI/2010, CIRG/1371/2013, and CIRG/1417/2015) and the Biomedical Research Council (grants 08/1/35/19/550 and 09/1/35/19/616) of Singapore; the Ministry of Health, Singapore; the National University of Singapore and the National University Health System, Singapore; the Agency for Science, Technology and Research, Singapore; Merck Sharp & Dohme Corp., Whitehouse Station, NJ, USA; Kuwait Foundation for Advancements of Sciences (The KODGP); the Oogfonds, MaculaFonds, Landelijke Stichting voor Blinden en Slechtzienden, Stichting Blindenhulp, Stichting A.F. Deutman Oogheelkunde Researchfonds; in Mexico the Fondo Sectorial de Investigación en Salud y Seguridad Social SSA/IMSS/ISSESTECONACYT project 150352; Temas Prioritarios de Salud Infantil del Seguro Social 2014-FIS/IMSS/PROT/PRIO/14/34; the Fundación IMSS; Compute Ontario (www.computeontario.ca) and Compute Canada (www.compute.canada.ca); CHRI Operating grants and a CHR New Investigator Award (E.J.P.); the Westlake Education Foundation (Jian Y.); Astrazeneca; a Miguel Servet Grant from the ISCIII Spanish Health Institute number CP17/00142 and co-financed by the European Social Fund (M.S.-L.); the Dutch Ministry of Justice; the European Science Foundation EuroSTRESS project FP-006; Biobanking and Biomolecular Resources Research Infrastructure BMBRI-NL award CP 32; Accare Centre for Child and Adolescent Psychiatry; and the Dutch Brain Foundation; the Federal Ministry of Science, Germany award 01 EA 9401; German Cancer Aid award 70-2488-Ha I; the participating Departments, the Division and the Board of Directors of the Leiden University Medical Centre and the Leiden University, Research Profile Area ‘Vascular and Regenerative Medicine’; Research Project For Excellence IKY/SIEMENS; the Wake Forest School of Medicine grant M01 RR07122 and Venture Fund; the Greek General Secretary of Research and Technology award PENED 2003; the MRC-PHE Centre for Environment and Health; the Singapore Ministry of Health’s National Medical Research Council under its Singapore Translational Research Investigator (STaR) Award NMRC/STaR/0028/2017 (J.C.C.); the German Research Foundation Project-ID 431984000 - SFB 1453 (M.Wu., Anna K.); the KHI Foundation for Preventive Medicine, and Bayer Pharma AG; the German Research Foundation grant KO 3598/5-1 (Anna K.); the Leipzig Research Center for Civilization Diseases; the Medical Faculty of the University of Leipzig; the Free State of Saxony; the Medical Research Funds from Kengbuk Samsung Hospital (H.-N.K.); the Division of Adult and Community Health, Centers for Disease Control and Prevention; Astra-Zeneca (P.M.R., D.I.C.); Amgen (P.M.R., D.I.C.); a gift from the Smilow family; the Perelman School of Medicine at the University of Pennsylvania; the University of Bristol; a comprehensive list of grants funding is available on the ALSPAC website; the US Centers for Disease Control and Prevention/Association of Schools of Public Health awards S043, S1734, and S3486, and US Centers for Disease Control and Prevention awards U01 DP003206 and U01 DP006266; the Netherlands Genomics Initiative’s Netherlands Consortium for Healthy Aging grant 050-060-B10; the Netherlands Heart Foundation grant 2001 D 032 (J.W.); the Chief Scientist Office of the Scottish Government Health Directorates award CZD/16/6, the Scottish Funding Council award HR03006; the Stiftelsen Kristian Gerhard Jebsen; Faculty of Medicine and Health Sciences, Norwegian University of Science and Technology; Central Norway Regional
Foundation (A.Po.); VIAGenomics number SP/19/2/344612; the Strategic Cardiovascular Program of Karolinska Institutet and Stockholm County Council; the Foundation for Strategic Research and the Stockholm County Council number 560283; the ALF/LUA research grant in Gothenburg; the Torsten Soderberg Foundation; the ERC grants ES/S007253/1, ES/T002611/1, and ES/T014083/1 (M.Ku.); Beijing Municipal of Health Reform and Development Project #2019-4 (Beijing Eye Study); the Children’s Hospital of Philadelphia; a Research Development Award from the Cosfowment; the Children’s Hospital of Philadelphia Endowed Chair in Genomic Research; the Daniel B. Burke Endowed Chair for Diabetes Research; the Italian Ministry of Universities grant IDF SHARID ARS01_01270; the Assessorato Ricerca Regione Campania grant POR CAMPANIA 2000/2006 MISURA 3.16; the Dutch Ministry of Health, Welfare and Sport; the Dutch Ministry of Economic Affairs; the University Medical Center Groningen (UMCG the Netherlands); University of Groningen and the Northern Provinces of the Netherlands; the UMCG Genetics Lifelines Initiative supported by a Spinoza Grant from NWO; University of Michigan discretionary funds; National Institute of Health, Republic of Korea grants 4845–301, 4851–302, 4851–307; Korea National Institute of Health intramural grant 2019-NG-053-02; the Korea Healthcare Technology R&D Project, Ministry of Health and Welfare, Republic of Korea award A102065; the National Research Foundation of Korea Grant 2020RI11A2075302 (Y.S.C.); the National Research Foundation of Korea Grant NRF-2020RI1A2C1012931; the Republic of Croatia Ministry of Science, Education and Sports research grant 108-1080315-0302; the Eye Birth Defects Foundation Inc.; the National Science Council, Taiwan grant NSC 98-2314-B-075A-002-MY3; the Taichung Veterans General Hospital, Taichung, Taiwan grant TCVGH-1003001C; AFNET; EHRA; German Centre for Cardiovascular Research (DZHK); German heart Foundation (DSF); the State of Brandenburg DZD grant 82DD00302; Sanofi; Abbott; the Victor Chang Cardiac Research Institute; NSW Health; the Center for Translational Molecular Medicine, the University Medical Center Groningen; the Dutch Kidney Foundation grant E0.13; the Netherlands Cardiovascular Research Initiative; the Dell Loy Hansen Heart Foundation (M.J.Cu.); Biosense Webster, ImriCor, and ADAS software (S.N.); the Swedish Heart-Lung Foundation grant 2019-0526; Swedish Foundation for Strategic Research grant IRC15-0067; Skåne University Hospital; governmental funding of clinical research within the Swedish National Health Service; the Knut and Alice Wallenberg Foundation (J.G.S.); the Boettcher Foundation Webb Waring Biomedical Research Award (Sr.R.); the Translational Genomics Research Institute; the Singapore National Medical Research Council grant 1270/2010, and the National Research Foundation, Singapore project 370062002; the Genetic Laboratory of the Department of Internal Medicine, Erasmus MC; the Research Institute for Diseases in the Elderly grant 014-93-015; the Netherlands Genomics Initiative (NGI)/Netherlands Organisation for Scientific Research (NWO) Netherlands Consortium for Healthy Aging project 050-060-810; the Dutch Dairy Association NZO; Netherlands Consortium Healthy Aging, Ministry of Economic Affairs, Agriculture and Innovation project KB-15-004-003; Wageningen University; VU University Medical Center; and Erasmus MC; The Folkhalsan Research Foundation; Nordic Center of Excellence in Disease Genetics; Finnish Diabetes Research Foundation; Foundation for Life and Health in Finland; Finnish Medical Society; Helsinki University Central Hospital Research Foundation; Perkéli Foundation; Oljquist Foundation; Narpes Health Care Foundation; Municipal Health Care Center and Hospital in Jakobstad; and Health Care Centers in Vasa, Narpes and Korsholm; the Institute of Cancer Research and The Everyman Campaign; The Prostate Cancer Research Foundation; Prostate Research Campaign UK (now PCUK); The Orchid Cancer Appeal; Rosettes Trust; The National Cancer Research Network UK; The National Cancer Research Institute (NCRI) UK; the Movember Foundation grants D2013-36 and D2013-17; the Morris and Horowitz Families Endowed Professorship; the Swedish Cancer Foundation; Ligue Nationale Contre le Cancer, Institut National du Cancer (INCa); Fondation ARC; Fondation de France; Agence Nationale de sécurité sanitaire de l’alimentation, de l’environnement et du travail (ANSES); Ligue départementale du Val de Marne; the Baden Württemberg Ministry of Science, Research and Arts; The Ronald and Rita McAulay Foundation; Cancer Australia; AICR Netherlands A10-0227; Cancer Council Tasmania; Cancer Councils of Victoria and South Australia; Philanthropic donation to Northshore University Hospital System Health; FWO Vlaanderen grants G.0684.12N and G.0830.13N; the Belgian federal government grant KPC_29.023; a Concerted Research Action of the KU Leuven grant GOA/15/017; the Spanish Ministry Council Instituto de Salud Carlos III-FEDER grants PI08/1770, PI09/00773-Cantabria, PI11/01899-FEDER, PI12/00265, PI12/01270, PI12/00715, PI15/0069, and RD09/0076/0036; the Fundación Marqués de Valdecilla grant API 10/09; the Spanish Association Against Cancer (AECC) Scientific Foundation; the Catalan Government DURSI grant 2009SGR1489; the Xarxa de Bancs de Tumors de Catalunya sponsored by Pla Director d’Oncologia de Catalunya (XBTC); the Spanish Ministry of Science and Innovation grant CEX2018-000806-S; the Generalitat de Catalunya; the VicHealth and Cancer Council Victoria; Programa Grupos Emergentes; Cancer Genetics Unit, CHIUV Vigo Hospital; Instituto de Salud Carlos III, Spain; Cancer Australia PaCCRS and Cancer Council Queensland; the Canada Cancer Research Fund grant 99-00527V-10182; US Public Health Service grants U10CA37429 and 5UM1CA182883; Canadian Cancer Society Research Institute Career Development Award in Cancer Prevention grant 2013-702108; the German Cancer Aid (Deutsche Krebshilfe); The Anthony DeNovi Fund; the Donald C. McGraw Foundation; and the St. Louis Men’s Group Against Cancer; UK Biobank project 12505; the Australian Research Council grant DE200100425 (Lo.Y.); the Australian Research Council grant FL180100072 (P.M.Vis.); Westlake Education Foundation (Jian Y.); a Burroughs Wellcome Fund Career Award, the Next Generation Fund at the Broad Institute of MIT and Harvard, and a Sloan Research Fellowship (P.-R.L.); the Consortium for Systems Biology (NCSB), the Netherlands Genomics Initiative (NGI)/Netherlands Organisation for Scientific Research (NWO); the Government of Canada through Genome Canada and the Canadian Institutes of Health Research grant GPH-129344; the Ministère de l’Economie et de l’Innovation du Québec through Genome Québec grant P5RSIIRI-701; the Quebec Breast Cancer Foundation; the US Department of Defense grant W81XWH-10-1-0341; 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/M001660/1; Wellcome Investigator and NIHR Senior Investigator (M.I.M.); Council of Scientific and
Industrial Research, Government of India grant number BSC0122; the Department of Science and Technology, Government of India through PURSE II CDST/SR/PURSE PHASE II/11 provided to Jawaharlal Nehru University, New Delhi, INDIA; the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) Projektnumer 209933838 – SFB 1052; B03, C01; SPP 1629 TO 718/2- 1; the Competitive Research Funding of the Tampere University Hospital
grants 9M048 and 9N035; 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 X51001; 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-
RESRIGIROTTO (G.G.); 5 per mille 2015 senses CUP: C92F17003560001 (P.G.); the Helmholtz Zentrum München –
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.01.1.01.0010 (O.Po.) and the Center of Competence in Molecular Diagnostics grant
number KK.01.2.2.03.0006 (O.Po.); the Norwegian Research Council Mobility Grant 24014 and Young Research Talent
grant 287086; the South-Eastern Health Authorities PhD-grant 2019122; Vestre Viken Hospital Trust PhD-grant;
afib.no - the Norwegian Atrial Fibrillation Research Network; "Indremedisinsk Forskningsfond" at Bærum Hospital.
AUTHOR CONTRIBUTIONS

Steering committee:

Conveners of GIANT working groups:

Writing Group (drafted, edited, and commented on manuscript):

Coordinated or supervised data collection or analysis specific to manuscript:

Data preparation group (checked and prepared data from contributing cohorts for meta-analyses):

Meta-analysis working group:
J.N.H., E.M., Sa.V., Lo.Y.

Primary height analysis working group (post meta-analysis):

All other authors were involved in the design, management, coordination, or analysis of contributing studies.
COMPETING FINANCIAL INTERESTS

Yu.J. is employed by and hold stock or stock options in 23andMe, Inc. T.S.A. is a shareholder in Zealand Pharma A/S and Novo Nordisk A/S. G.C-P is an employee of 23andMe Inc. M.E.K. is employed by SYNLAB Holding Deutschland GmbH. H.L.L. receives support from a consulting contract between Data Tecnica International and the National Institute on Aging (NIA), National Institutes of Health (NIH). As of January 2020, A.Mah. is an employee of Genentech, and a holder of Roche stock. I.N. is an employee and stock owner of Gilead Sciences; this work was conducted before employment by Gilead Sciences. J.J. is employed by and hold stock or stock options in 23andMe, Inc. C.A.S. is an employee of Regeneron, Inc. Va.S. is employed by deCODE Genetics/Amgen inc. Since completing the work contributed to this paper, D.J.T. has left the University of Cambridge and is now employed by Genomics plc. G.T. is employed by deCODE Genetics/Amgen inc. H.B. has consulting arrangements with Chiesi Pharmaceuticals and Boehringer Ingelheim. M.J.Ca. is Chief Scientist for Genomics England, a UK Government company. M.J.Cu. has served on Advisory Board or Consulted for Biosense Webster, Janssen Scientific Affairs and Johnson & Johnson. S.M.D. receives research support from RenalytixAI and personal consulting fees from Calico Labs, outside the scope of the current research. P.T.E. receives sponsored research support from Bayer AG and IBM Health, and he has served on advisory boards or consulted for Bayer AG, Quest Diagnostics, MyoKardia and Novartis. P.Ki. has received support from several drug and device companies active in atrial fibrillation, and has received honoraria from several such companies in the past, but not in the last three years. P.Ki. is listed as inventor on two patents held by University of Birmingham (Atrial Fibrillation Therapy WO 2015140571, Markers for Atrial Fibrillation WO 2016012783). G.D.K. has given talks, attended conferences and participated in trials sponsored by Amgen, MSD, Lilly, Vianex, Sanofi, and have also accepted travel support to conferences from Amgen, Sanofi, MSD and Elpen. S.A.Lu. receives sponsored research support from Bristol Myers Squibb / Pfizer, Bayer AG, Boehringer Ingelheim, Fitbit, and IBM, and has consulted for Bristol Myers Squibb / Pfizer, Bayer AG, and Blackstone Life Sciences. W.M. reports grants and personal fees from AMGEN, BASF, Sanofi, Siemens Diagnostics, Aegerion Pharmaceuticals, Astrazeneca, Danone Research, Numares, Pfizer, Hoffmann LaRoche; personal fees from MSD, Alexion; grants from Abbott Diagnostics, all outside the submitted work. W.M. is employed with Synlab Holding Deutschland GmbH. 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 scientific advisor to Circle software, ADAS software, CardioSolv, and ImriCor and receives grant support from Biosense Webster, ADAS software, and ImriCor. Her.S. has received honoraria for consulting from AstraZeneca, MSD/Merck, Daiichi, Servier, Amgen and Takeda Pharma. He has further received honoraria for lectures and/or chairs from AstraZeneca, BayerVital, BRAHMS, Daiichi, Medtronic, Novartis, Sanofi and Servier. P.S. has received research awards from Pfizer Inc. 23andMe Research team are employed by and hold stock or stock options in 23andMe, Inc. 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.McC. has served on advisory panels for Pfizer, NovoNordisk and Zoe Global, has received honoraria from Merck, Pfizer, Novo Nordisk and Eli Lilly, and research funding from Abbvie, Astra Zeneca, Boehringer Ingelheim, Eli Lilly, Janssen, Merck, NovoNordisk, Pfizer, Roche, Sanofi Aventis, Servier, and Takeda. As of June 2019, MMcC 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 inc. K.S. is employed by deCODE Genetics/Amgen inc. Ad.A. is employed by and hold stock or stock options in 23andMe, Inc. C.J.W.’s spouse is employed by Regeneron. A.E.L. is currently employed by and holds stock in Regeneron Pharmaceuticals, Inc. J.N.H. holds equity in Camp4 Therapeutics.
Table 1. Summary of results from within-ancestry and trans-ancestry GWAS meta-analyses. N denotes the sample size for each SNP. GWS: Genome-Wide Significant ($P<5\times 10^{-8}$). COJO SNPs: near independent GWS SNPs identified using an approximate conditional and Joint analysis implemented in the GCTA software. $P_{\text{GWAS}}$: P-value from marginal association test. GWS loci were defined as genomic regions centred around each GWS SNP and including all SNPs within 35 kb on each side of the lead GWS SNP. Overlapping GWS loci were merged so that the number and cumulative length of GWS loci are calculated on non-overlapping GWS loci. Percentage of the genome covered was calculated by dividing the cumulative of GWS loci by 3,039 Mb, i.e. the approximated length of the human genome.

<table>
<thead>
<tr>
<th>Cohort Ancestry/Ethnic Group</th>
<th>Number of studies</th>
<th>Max N (Mean N)</th>
<th>Number of GWS COJO SNPs ($P_{\text{GWAS}}&lt;5\times 10^{-8}$)</th>
<th>Number of GWS loci (35 kb)</th>
<th>Cumulative length of non-overlapping GWS loci (% genome)</th>
</tr>
</thead>
<tbody>
<tr>
<td>European (EUR)</td>
<td>173</td>
<td>4,080,687 (3,612,229)</td>
<td>9,863 (8,382)</td>
<td>6,386</td>
<td>552.5 Mb (18.4%)</td>
</tr>
<tr>
<td>East-Asian (EAS)</td>
<td>56</td>
<td>472,730 (320,570)</td>
<td>918 (807)</td>
<td>821</td>
<td>60.5 Mb (2.0%)</td>
</tr>
<tr>
<td>Hispanic (HIS)</td>
<td>11</td>
<td>455,180 (431,645)</td>
<td>1,888 (1,332)</td>
<td>1,599</td>
<td>121.4 Mb (4.0%)</td>
</tr>
<tr>
<td>African (AFR)</td>
<td>29</td>
<td>293,593 (222,981)</td>
<td>493 (417)</td>
<td>436</td>
<td>32.5 Mb (1.1%)</td>
</tr>
<tr>
<td>South Asian (SAS)</td>
<td>12</td>
<td>77,890 (59,420)</td>
<td>69 (65)</td>
<td>66</td>
<td>4.7 Mb (0.2%)</td>
</tr>
<tr>
<td>Trans-ancestry meta-analysis (META)</td>
<td>281</td>
<td>5,314,291* (4,611,160)</td>
<td>12,111 (9,920)</td>
<td>7,209</td>
<td>647.5 Mb (21.6%)</td>
</tr>
</tbody>
</table>

*The number of individuals in the trans-ancestry meta-analysis (N=5,314,291) is smaller than the sum of ancestry group specific meta-analyses (N=5,380,080) because of variation in per-SNP sample sizes for SNPs included in the final analysis.
Table 2. Overview of 5 European ancestry GWAS re-analysed in our study to quantify the relationship between sample size and discovery. Summary statistics from the 3 published GWAS were imputed using the SSIMP software to maximise coverage of HapMap 3 SNPs (Suppl. Methods). GWS loci are defined as in the legend of Table 1.

<table>
<thead>
<tr>
<th>Down-sampled GWAS</th>
<th>Max N (Mean N)</th>
<th>Number of GWS COJO SNPs</th>
<th>% of the genome covered by GWS loci (35 kb)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lango-Allen et al.(^{15})*</td>
<td>130,010 (128,942)</td>
<td>240</td>
<td>0.5%</td>
</tr>
<tr>
<td>Wood et al.(^{16})</td>
<td>241,724 (239,227)</td>
<td>633</td>
<td>1.4%</td>
</tr>
<tr>
<td>Yengo et al.(^{1})</td>
<td>695,648 (688,927)</td>
<td>2,794</td>
<td>5.8%</td>
</tr>
<tr>
<td>GIANT-EUR (no 23andMe)</td>
<td>1,632,839 (1,502,499)</td>
<td>4,867</td>
<td>9.7%</td>
</tr>
<tr>
<td>23andMe-EUR</td>
<td>2,502,262 (2,498,336)</td>
<td>7,020</td>
<td>13.6%</td>
</tr>
</tbody>
</table>

\(^{*}\)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. 1. Brisbane plot showing the genomic density of independent genetic associations with height. Each dot represents one of the 12,111 quasi-independent genome-wide significant (GWS; \( P < 5 \times 10^{-8} \)) height-associated SNPs identified using approximate conditional and joint multiple-SNP (COJO) analyses of our trans-ancestry GWAS meta-analysis. Density was calculated for each associated SNP as the number of other independent associations within 100 kb. A density of 1 means that a GWS COJO SNP share its location with another independent GWS COJO SNP within <100 kb. The average signal density across the genome is 2 (standard error; S.E. 0.14). S.E. were calculated using a Leave-One-Chromosome-Out jackknife approach (LOCO-S.E.). Sub-significant SNPs are not represented on the figure.
Fig. 2. Variance of height explained by HapMap 3 SNP within genome-wide significant (GWS) loci. **Panel a** shows stratified SNP-based heritability ($h^2_{SNP}$) estimates obtained after partitioning the genome into SNPs within 35 kb of a GWS SNP (“GWS loci” label) vs. SNPs >35 kb away from any GWS SNP. Analyses were performed in samples of five different ancestry/ethnic groups: European (EUR: meta-analysis of UK Biobank (UKB) + Lifelines study), African (AFR: meta-analysis of UKB + PAGE study), East-Asian (EAS: meta-analysis of UKB + China Kadoorie Biobank), South-Asian (SAS: UKB) and Hispanic group (HIS: PAGE). **Panel b** shows that >90% of $h^2_{SNP}$ in all ancestries is explained by SNPs within GWS loci identified in this study. The cumulative length of non-overlapping GWS loci is ~647 Mb, i.e. ~21% of the genome assuming a genome length of ~3039 Mb. The proportion of HapMap 3 SNPs in GWS loci is ~27%. 
Fig. 3. Accuracy of a polygenic predictors of height (PGS) within-family and across ancestries. Prediction accuracy ($R^2$) was measured as the squared correlation between PGS and actual height adjusted for age, sex and 10 genetic principal components. Panel a shows the accuracy of PGSs assessed in participants of 5 different ancestry groups: European (EUR; N=14,587) from the UK Biobank (UKB) and the Lifelines Biobank (LLB; N=14,058) cohorts, South-Asian (SAS; N=9,257) from UKB, East-Asian (EAS; N=2,246) from UKB, Hispanic (HIS; N=8,238) from the PAGE study and admixed African (AFR) from UKB (N=6,911) and PAGE (N=5,798). PGSs used for prediction, in Panel a, are based on genome-wide significant (GWS) SNPs identified in (1) cross-ancestry meta-analysis (green bar), (2) EUR meta-analysis (yellow bar) and (3) ancestry-specific meta-analyses (red bars). Panel b shows the squared correlation of height between first-degree relatives of EUR participants in UKB and the accuracy of a predictor combining PGS (denoted, $\text{PGS}_{\text{GWS}}$ as based on GWS SNPs) and familial information. $\text{PGS}_{\text{GWS}}$ accuracy shown in Panel b is the average accuracy in EUR participants from UKB and LLB from Panel a. Sibling correlation was calculated in 17,492 independent EUR sibling pairs from the UKB and parent-offspring correlations in 981 EUR unrelated trios (i.e. two parents and 1 child) from the UKB.