

Defining the role of common variation in the genomic and biological architecture of adult human height

Using genome-wide data from 253,288 individuals, we identified 697 variants at genome-wide significance that together explained one-fifth of the heritability for adult height. By testing different numbers of variants in independent studies, we show that the most strongly associated ~2,000, ~3,700 and ~9,500 SNPs explained ~21%, ~24% and ~29% of phenotypic variance. Furthermore, all common variants together captured 60% of heritability. The 697 variants clustered in 423 loci were enriched for genes, pathways and tissue types known to be involved in growth and together implicated genes and pathways not highlighted in earlier efforts, such as signaling by fibroblast growth factors, WNT/ β -catenin and chondroitin sulfate-related genes. We identified several genes and pathways not previously connected with human skeletal growth, including mTOR, osteoglycin and binding of hyaluronic acid. Our results indicate a genetic architecture for human height that is characterized by a very large but finite number (thousands) of causal variants.

Height is a classical polygenic trait that has provided general insights into the genetic architecture of common human traits and diseases and into the prospects and challenges of different methods used to identify genetic risk factors. Studies consistently estimate that the additive genetic contribution to normal variation in adult height ('narrow-sense heritability') is approximately 80% (refs. 1–3). Previous analysis of genome-wide association studies (GWAS) of adult height showed that common variants together account for 50% of this heritable contribution to variation in height^{4,5}. The most recent GWAS of adult height identified 180 loci, which together highlighted many genes relevant to human skeletal growth that had not been implicated in previous studies⁶. Common variants in these loci, however, only accounted for 10% of the phenotypic variation (~12% of heritability). Here we report the results from a GWAS meta-analysis of adult height in 253,288 individuals of European ancestry. We show that the additive contributions of fewer than 10,000 SNPs (at $P < 5 \times 10^{-3}$) can account for 36% of the heritability of adult height. Variants reaching genome-wide significance ($P < 5 \times 10^{-8}$) in this larger study (697 SNPs) clustered in loci, were substantially enriched for regulatory variants and implicated multiple known and previously unknown genes and pathways relevant to growth. More broadly, our results provide evidence that increasing GWAS sample sizes to the order of 10^5 individuals, now plausible for many common traits, will likely continue to identifying associated variants and loci while improving knowledge of the biology of these traits.

RESULTS

The overall analysis strategy is illustrated in **Supplementary Figure 1**. We first performed a GWAS meta-analysis of adult height using the summary statistics from 79 studies consisting of 253,288 individuals of European ancestry (Online Methods). We identified 697 SNPs that reached genome-wide significance ($P < 5 \times 10^{-8}$) using an approximate conditional and joint multiple-SNP (COJO) analysis⁷

in GCTA⁸ (Online Methods), which takes linkage disequilibrium (LD) between SNPs into account (**Supplementary Figs. 2 and 3**, and **Supplementary Table 1**). The 697 SNPs clustered in 423 loci, with a locus defined as one or multiple jointly associated SNPs located within a 1-Mb window. Most of these 697 SNPs were uncorrelated, although those in close physical proximity (for example, <1 Mb apart) might be in partial LD (see **Supplementary Table 1** for LD between adjacent pairs of the 697 SNPs). The clustering of signals was non-random (empirical enrichment of 1.4-fold; $P < 1 \times 10^{-4}$) with 90, 26 and 31 loci containing 2, 3 and ≥ 4 signals, respectively (**Supplementary Tables 1 and 2**, and **Supplementary Note**). We observed strong evidence of clustering of association signals within loci across a range of locus sizes, from 100 kb to 1.25 Mb, but the clustering was almost entirely driven by variants within 250 kb of index SNPs (**Supplementary Table 2** and **Supplementary Note**). In some loci, multiple signals clustered tightly around a single gene, whereas in other cases the clustering of associated variants was likely due to multiple different height-related genes being in close proximity (**Fig. 1** and **Supplementary Fig. 4**).

Of the 697 SNPs, 403 were represented on the MetaboChip array⁹. Using data from 80,067 individuals genotyped on the MetaboChip array from 37 independent studies, we observed very strong evidence of concordance of effect sizes between the MetaboChip and GWAS samples ($P = 1.9 \times 10^{-160}$): >99% of variants had directionally consistent effects in the MetaboChip and GWAS data (Online Methods, **Supplementary Table 3** and **Supplementary Note**).

We observed a large genome-wide 'inflation' of the test statistic for association, even after we corrected each study's test statistics by its individual inflation factor (single $\lambda_{GC} = 1.94$). At least two phenomena could have contributed to this observation. First, as described previously¹⁰, highly polygenic models of inheritance are expected to increase the genomic inflation factor to levels comparable to what we observe. Second, height is particularly susceptible to confounding by population ancestry (stratification), which can also lead to inflation

A full list of authors and affiliations appears at the end of the paper.

Received 18 December 2013; accepted 29 August 2014; published online 5 October 2014; doi:10.1038/ng.3097

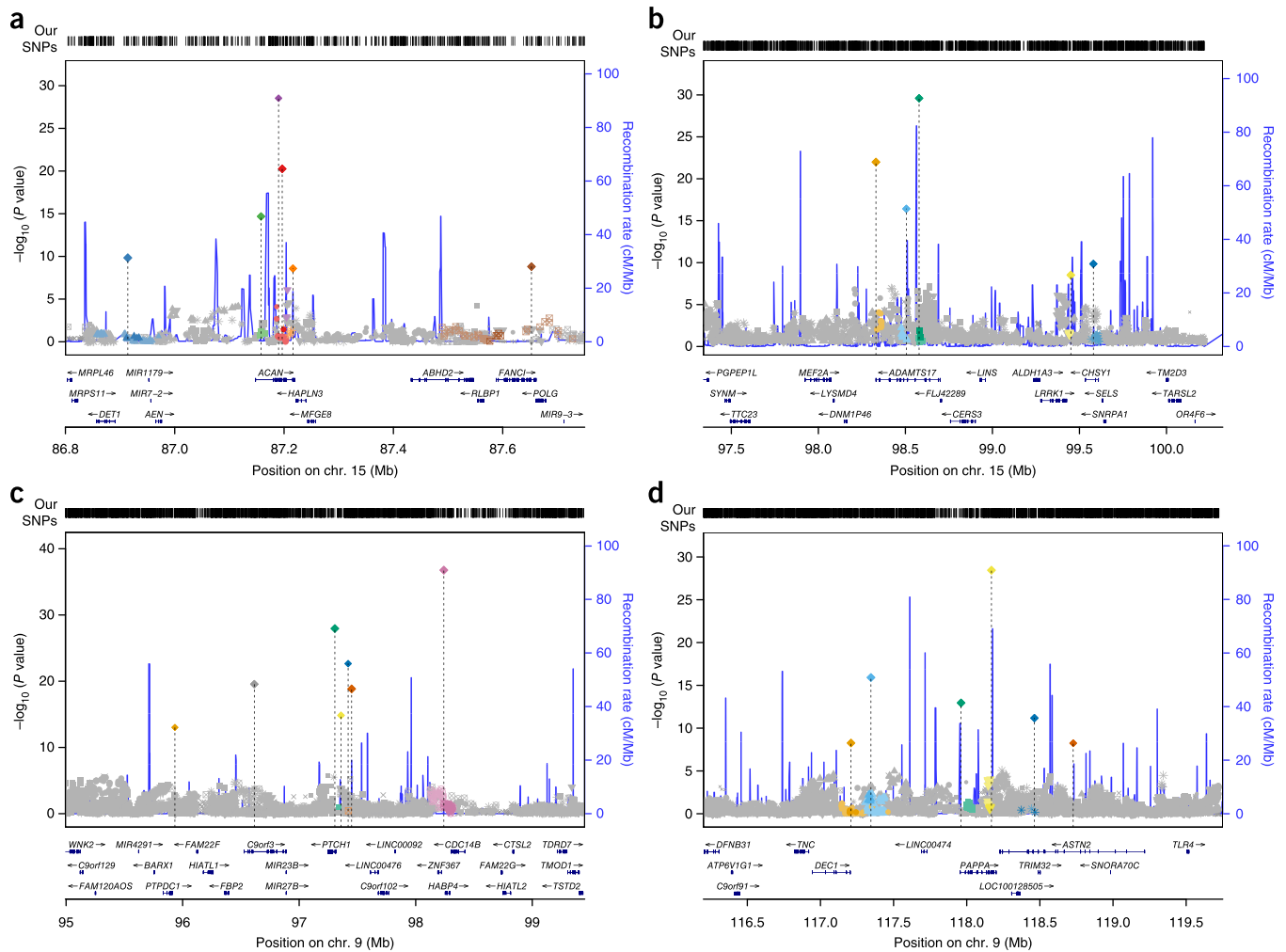


Figure 1 Regional association plots for loci with multiple association signals. (a–d) Examples of multiple signals after approximate conditional joint multiple-SNP (GCTA-COJO) analysis. SNPs are represented by colored symbols according to the index SNP with which they are in strongest LD ($r^2 > 0.4$). In some loci, the majority of signals cluster in and around a single gene, for example, at *ACAN* (a), *ADAMTS17* (b) and *PTCH1* (c), whereas at some loci multiple signals cluster through proximity (d).

of the test statistics. We addressed these possibilities by comparing our results with those obtained using more stringent corrections for stratification (linear mixed models; LMMs) and with results obtained in subsets of studies in which a purely family-based analysis was feasible, and by performing a within-family prediction analysis that partitioned the variance in the genetic predictor into the contributions of true associations and population stratification.

Our LMM analyses, performed in a subset of 15 individual studies comprising 59,380 individuals, provided strong evidence that the inflated statistics were driven predominantly by the highly polygenic nature of the trait. This approach uses a genomic relationship matrix (GRM) calculated from genome-wide SNP data to correct for distant relatedness between all pairs of individuals in a study. We obtained a single λ_{GC} of 1.20. This value is entirely consistent with the single λ_{GC} of 1.20 obtained from the standard GWAS analysis of the same individuals and a single λ_{GC} of 1.94 obtained from the analysis of all 253,288 individuals (Supplementary Table 4). Because this approach might be overly conservative for a strongly genetic and highly polygenic trait, each study additionally repeated the analyses for each chromosome using a GRM generated from the remaining 21 chromosomes or, in the case of the largest study (WGHS), repeated the

analyses for all odd-numbered chromosomes using a GRM generated from the even-numbered chromosomes and vice versa. The single λ_{GC} inflation factor for this analysis, 1.23, is also entirely consistent with the standard GWAS results (Online Methods, Supplementary Table 4 and Supplementary Note).

Our family-based analyses also provided strong evidence that the inflated statistics were driven predominantly by the highly polygenic nature of height. We assessed whether variants that reached genome-wide significance after single λ_{GC} correction replicated in family-based analyses of up to 25,849 samples (effective sample size of 14,963; using methods that are immune to stratification (Online Methods, Supplementary Tables 5 and 6, and Supplementary Note)). We identified genome-wide significant associations from a meta-analysis that excluded the family-based samples and tested these associations for replication in the family-based samples; a lower rate of replication than expected could be due to inflation of effect sizes in the discovery sample from ‘winner’s curse’ and/or stratification. Of the 416 genome-wide significant SNPs representing multiple signals selected after the exclusion of the family-based studies, 371 SNPs had a consistent direction of effect (in comparison with 208 expected by chance and 400 expected in the absence of any inflation of estimated

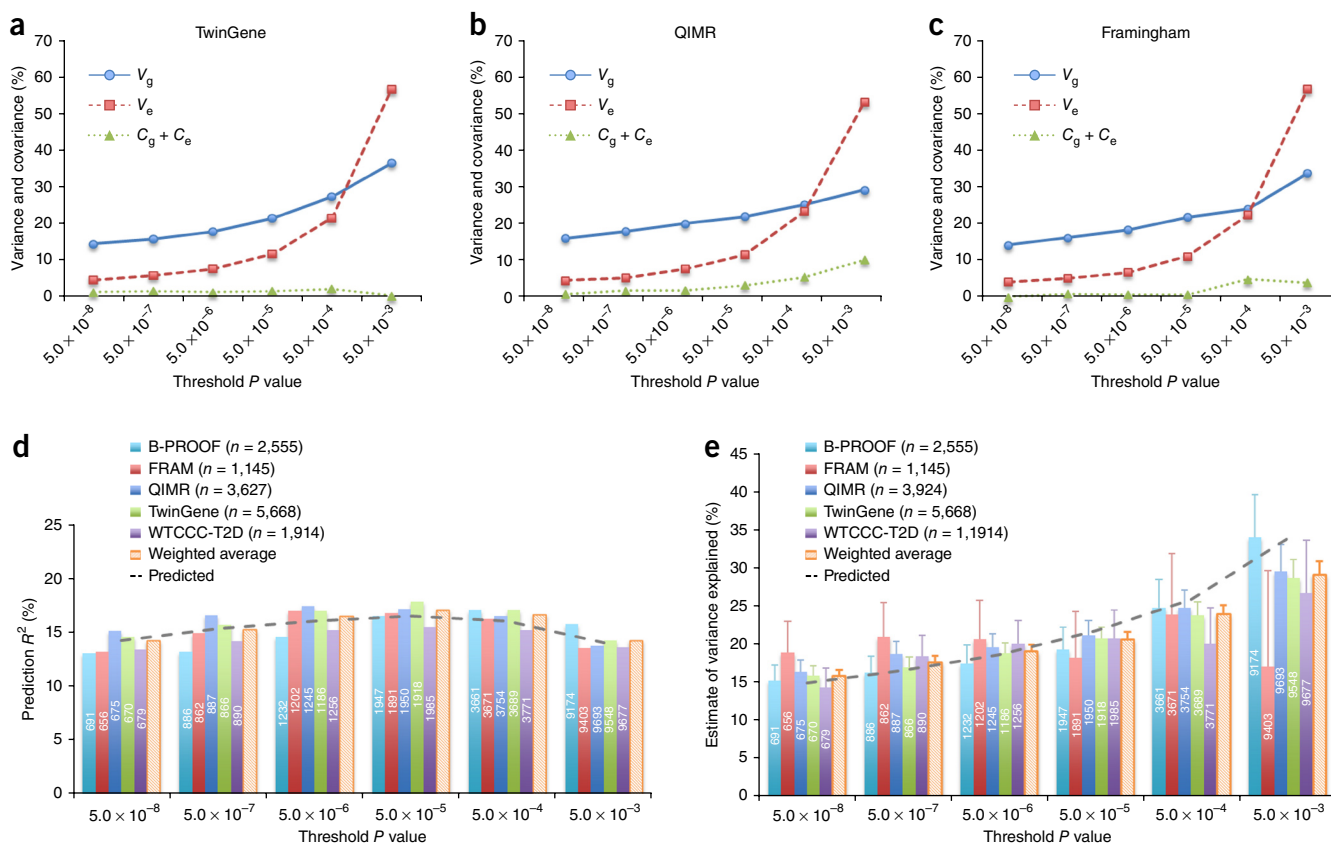


Figure 2 Quantifying the variance explained by height-associated SNPs at different levels of significance. SNPs were selected from the GCTA-COJO analysis with the target cohort excluded from the meta-analysis. **(a–c)** Partitioning of the variance of the SNP-derived genetic predictor using a within-family analysis in three studies: TwinGene **(a)**, QIMR **(b)** and Framingham **(c)**. The SNP-based predictor was adjusted by the first 20 principal components. The four variance and covariance components V_g , V_e , C_g and C_e are defined in the Online Methods. **(d)** Accuracy of predicting phenotype with the genetic predictor in unrelated individuals. The prediction R^2 value shown on the y axis is the squared correlation between the phenotype and predictor. The SNP-based predictor was adjusted by the first 20 principal components. The solid line is the average prediction R^2 value weighted by sample size over the five cohorts. The dashed line is the prediction accuracy inferred from the within-family prediction analysis (equation (19) in the Online Methods). **(e)** Estimate of variance explained by the selected SNPs from the GCTA-GREML analysis. The phenotype was adjusted by the first 20 principal components. Each error bar represents the standard error of the estimate. The estimates from all five cohorts (B-PROOF, FRAM, QIMR, TwinGene and WTCCC-T2D) were averaged by the inverse variance approach. The dashed line is the variance explained inferred from the within-family prediction analysis. In **d** and **e**, the number shown in each column is the number of SNPs used in the analysis.

effect sizes) and 142 SNPs replicated with $P < 0.05$ (in comparison with 21 expected by chance and 210 expected in the absence of effect size inflation; **Supplementary Table 5**). These analyses (particularly the directional consistency) show that most of the loci represent true associations but also show that there is a modest inflation in the effect size estimates, due to stratification and/or winner's curse. To distinguish between these possibilities, we repeated this analysis, substituting for the family-based samples a random set of studies with similar total effective sample sizes. The number of replicating loci was only slightly lower in the family-based cohorts than in the random samples (12–17 fewer replications attributable to stratification at different P -value thresholds; **Supplementary Table 5**). This finding indicates that most of the modest inflation in effect estimates is due to winner's curse, that a small amount of inflation is due to residual stratification and that few (upper limit of ~15–25; **Supplementary Table 5** and **Supplementary Note**) if any of the loci that reach genome-wide significance after single- λ_{GC} correction are likely to be complete false positives due to stratification (that is, with no real association whatsoever with height).

Variance explained by SNPs at different significance levels

Having established that single- λ_{GC} correction is sufficient to identify SNPs that are likely to be truly associated with height, we next performed a series of analyses using the GWAS data from five independent validation studies to quantify the fraction of phenotypic variance explained by SNPs selected from the GCTA-COJO analyses⁷ of the meta-analysis data, which excluded data from the validation studies, at a range of statistical thresholds and to quantify the accuracy of predicting height using these selected SNPs (Online Methods). We first developed a new method that uses within-family prediction to partition the variance of the SNP-based predictor into components due to real SNP effects, errors in estimating SNP effects and population stratification (Online Methods), and we applied the method to data on full-sibling pairs from three of the five validation studies (Online Methods). Consistently across the three studies, all the partitioned variance components increased in size as a less stringent significance level was used for SNP selection in the discovery sample, and the error variance increased more dramatically than the genetic variance when more SNPs selected at a less significant level were included

Table 1 Estimates of variance explained by SNPs selected at different significance levels.

Threshold	QIMR (<i>n</i> = 3,924)			FRAM (<i>n</i> = 1,145)			TwinGene (<i>n</i> = 5,668)			WTCCC-T2D (<i>n</i> = 1,914)			B-PROOF (<i>n</i> = 2,555)			Weighted average ^a		Pred. ^b
	Number of SNPs	h_g^2	SE	Number of SNPs	h_g^2	SE	Number of SNPs	h_g^2	SE	Number of SNPs	h_g^2	SE	Number of SNPs	h_g^2	SE	h_g^2	SE	h_g^2
5×10^{-8}	675	0.164	0.016	656	0.190	0.040	670	0.159	0.013	679	0.143	0.025	691	0.152	0.021	0.159	0.008	0.149
5×10^{-7}	887	0.187	0.017	862	0.210	0.045	866	0.170	0.013	890	0.184	0.028	886	0.162	0.022	0.176	0.009	0.166
5×10^{-6}	1,245	0.196	0.018	1,202	0.207	0.050	1,186	0.188	0.014	1,256	0.201	0.030	1,232	0.175	0.024	0.190	0.009	0.186
5×10^{-5}	1,950	0.212	0.020	1,891	0.183	0.060	1,918	0.208	0.015	1,985	0.208	0.037	1,947	0.194	0.029	0.206	0.010	0.218
5×10^{-4}	3,754	0.248	0.024	3,671	0.239	0.080	3,689	0.239	0.017	3,771	0.201	0.047	3,661	0.248	0.037	0.240	0.013	0.259
5×10^{-3}	9,693	0.297	0.035	9,403	0.171	0.126	9,548	0.287	0.025	9,677	0.267	0.070	9,174	0.341	0.055	0.292	0.018	0.339
HM3 ^c	1.08	0.473	0.086	1.06	0.313	0.291	1.12	0.522	0.060	0.97	0.534	0.170	1.09	0.463	0.126	0.498	0.044	
	million			million			million			million			million					

SNPs were selected by an approximate conditional and joint multiple-SNP analysis (GCTA-COJO) of the summary statistics from the meta-analysis. The target cohort for variance estimation was excluded from the meta-analysis. h^2 , heritability; SE, standard error. Values in bold highlight the averages of five studies.

^aThe estimates from all five cohorts were averaged by the inverse variance approach, where $\sum_i (h_{g(i)}^2 / SE_i^2) / \sum_i (1 / SE_i^2)$. ^bThe predicted variance explained by the selected SNPs (V_g) from the within-family prediction analysis. ^cSNPs from the HapMap 3 project¹¹.

in the predictor (Fig. 2a–c). We demonstrated the partitioning of variance due to population stratification by the within-family prediction analyses with and without adjusting for principal components (Supplementary Fig. 5). The results again confirmed that the impact of population stratification on the top associated SNPs was minor and demonstrated that the variation in the predictor due to true SNP effect, estimation error and population stratification was quantifiable. We next inferred, using these partitioned variance components from the within-family prediction analysis, how well different selected sets of SNPs would predict height in independent samples. We showed that the observed prediction accuracy (squared correlation between phenotype and predictor, or R^2) in five different population-based cohorts was highly consistent with the values inferred from the within-family analyses, with prediction accuracy peaking at ~17% using the ~1,900 SNPs reaching $P < 5 \times 10^{-5}$ (Fig. 2d). Finally, we estimated the variance explained by the selected SNPs in population-based studies using the GCTA-GREML method^{4,8} (Fig. 2e). The results showed that ~670 SNPs at $P < 5 \times 10^{-8}$ and ~9,500 SNPs at $P < 5 \times 10^{-3}$ captured ~16% and ~29% of phenotypic variance, respectively (Table 1), findings that were also consistent with the estimates inferred from the within-family prediction analysis. As shown in equation (19), prediction R^2 is not equal to the variance explained but is a function of the variance of true SNP effects and the error variance in estimating SNP effects, in the absence of population structure. At thresholds below genome-wide significance, the variance explained is higher than the prediction accuracy because the latter is deflated both by imprecise estimates of effect sizes (estimation errors) and the inclusion of SNPs that are not associated with height (Fig. 2). The estimate of variance explained by all the HapMap 3 (ref. 11) SNPs without SNP selection was ~50% (Table 1), consistent with previous estimates^{4,5}. Thus, a group of ~9,500 SNPs (representing <1% of common SNPs) selected at $P < 5 \times 10^{-3}$ explained ~29% of phenotypic variance. Because ~50% of phenotypic variance is explained by all common SNPs the selected set of SNPs, despite being limited to <1% of common SNPs, accounts for the majority of the variance attributable to all common SNPs (29/50; ~60%). This set of ~9,500 SNPs strongly clustered with the newly established height loci: 1,704 (19%) variants were located within 250 kb of one of the 697 genome-wide associated SNPs, suggesting that a substantial fraction of ‘missing heritability’ is within already-identified loci. This clustering of additional variants within identified loci was confirmed in a parallel analysis based on two left-out studies where we observed that SNPs in closer physical proximity to the top associated SNPs explained disproportionately more variance (Online Methods and Supplementary Fig. 6).

The larger GWAS identifies new biologically relevant genes and pathways

Having shown that ~1% of variants can account for the majority of heritability attributable to common variation, we next considered whether the expanded set of height-associated variants could be used to identify the genomic features and biological pathways of most relevance to normal variation in adult height. To test whether our GWAS could implicate new biology, we used established and novel approaches to test whether the height-associated loci were enriched for functionally relevant variants, genes, pathways and tissues.

Similar to the 180 variants identified in our previous analysis, the 697 variants were non-randomly distributed with respect to functional and putatively functional regions of the genome (Online Methods). We observed that the height-associated variants were enriched for nonsynonymous SNPs (empirical enrichment of 1.2-fold; $P = 0.02$), *cis*-regulatory effects in blood (empirical enrichment of 1.5-fold; $P = 0.03$), overlap with a curated list of genes that underlie monogenic syndromes of abnormal skeletal growth¹² (empirical enrichment of 1.4-fold; $P = 0.013$), associations with apparently unrelated complex traits in the National Human Genome Research Institute (NHGRI) GWAS catalog (empirical enrichment of 2.6-fold; $P < 1 \times 10^{-4}$) and functional chromatin annotations in multiple tissues and cell types (empirical enrichment of 1.8-fold; $P < 1 \times 10^{-3}$) (Supplementary Tables 7–11 and Supplementary Note).

The greater resolution for the height-associated variants provided by increased sample size, in combination with improved gene prioritization and gene set enrichment approaches, resulted in the identification of multiple new tissues, gene sets and specific genes that were highly likely to be involved in the biology of skeletal growth. Specifically, using a variety of established and novel pathway methods, we identified ~3 times as many enriched pathways and prioritized ~5 times as many genes (including genes newly prioritized in previously identified loci) in comparison to the results derived from identical pathway methods applied to the previous GWAS of 133,000 individuals (Table 2).

We first focused on existing pathway and gene prioritization methods: (i) MAGENTA¹³, a method designed to identify gene sets enriched in GWAS data, and (ii) GRAIL¹⁴, which uses the published literature to highlight connections between likely relevant genes within GWAS-identified loci. As expected, the GRAIL and MAGENTA analyses confirmed several previously identified gene sets and pathways clearly relevant to skeletal growth, but in the larger sample they also provided evidence for additional known and new

Table 2 Comparison of prioritized variants, loci, biology and variance explained from GWAS on human stature

	Height GWAS with 130,000 samples (Lango Allen <i>et al.</i> ⁶ , Yang <i>et al.</i> ⁴) ^a	Height GWAS with 253,288 samples (this study)
SNP-based comparisons		
GWAS significant SNPs	199	697
Genomic loci ^b (± 1 Mb)	180	423
Loci ^b with multiple signals	19	147
Secondary associations in loci ^b	19	273
Biological annotation (DEPICT at FDR < 0.05)		
Prioritized genes	92	649
Loci ^c with a prioritized gene	74 (43%)	422 (75%)
Pruned gene sets and protein-protein complexes ^d	813	2,330
Tissues and cell types	5	43
Variance explained		
GWAS significant SNPs	10%	16%
Deep list of SNPs at $P < 1 \times 10^{-3}$	13%	29%
All common SNPs	45% ^e	50%
Heritability explained		
GWAS significant SNPs	12.5%	20%
Deep list of SNPs at $P < 1 \times 10^{-3}$	16%	36%
All common SNPs	56% ^e	62.5%

Comparison is shown for data from GWAS on human stature with 130,000 individuals (previously published) and with 250,000 individuals (this study).

^aCounts, numbers and estimates for Lango Allen *et al.*⁶ are taken from the respective publication. ^bGenomic loci defined by distance: ± 1 Mb from the index SNP for height. ^cGenomic loci defined by LD: $r^2 > 0.5$ with the index SNP for height. ^dAfter clumping of similar gene sets and pathways. ^eYang *et al.*⁴.

associated genes, gene sets and protein complexes not identified in our previous smaller study (for example, fibroblast growth factor (FGF) signaling, WNT signaling, osteoglycin and other genes related to bone or cartilage development) (Supplementary Fig. 7 and Supplementary Tables 12 and 13).

To obtain more detailed insight into height biology, we applied DEPICT, a novel data-driven, integrative method that uses gene sets reconstituted on the basis of large-scale expression data to prioritize genes and gene sets and also to identify tissues enriched in highly expressed genes from associated loci (T.H.P., J. Karjalainen, Y. Chan, H. Westra and A.R.W. *et al.*, unpublished data; Online Methods and Supplementary Note). The DEPICT analysis highlighted 2,330 reconstituted gene sets (after pruning for high levels of redundancy). These gene sets both confirmed and extended the MAGENTA and GRAIL findings and identified new pathways not identified in our previous height GWAS (for example, regulation of β -catenin, biology related to glycosaminoglycans such as chondroitin sulfate and hyaluronic acid, and mTOR signaling) (Supplementary Table 14). The gene sets identified on the basis of 327 strictly new height variants (>1 Mb from the 180 known variant loci) highly resembled the gene sets highlighted by the 180 already-known loci (Spearman's rank correlation coefficient between gene set enrichment z scores, $r = 0.91$; $P = 2 \times 10^{-16}$). Thus, the variants discovered through increased sample size continued to highlight specific and relevant growth-associated gene sets, and the combined analysis of both old and new loci provided the additional power needed to identify new gene sets (Table 3 and Supplementary Table 14).

The DEPICT analysis also prioritized tissues and individual genes. We found that genes within loci associated with height were enriched

for expression in tissues related to chondrocytes (cartilage, joint capsule, synovial membrane and joints; $P < 5.5 \times 10^{-9}$, false discovery rate (FDR) < 0.001) and other musculoskeletal, cardiovascular and endocrine tissue types (FDR < 0.05) (Fig. 3, Supplementary Fig. 8 and Supplementary Table 15). We also showed that a subset of the 697 height-associated SNPs that represented lead *cis* expression quantitative trait loci (eQTLs) in blood defined 75 genes that were collectively enriched for expression in cartilage ($P = 0.008$) (Supplementary Table 8 and Supplementary Note).

We used DEPICT to prioritize 649 genes (at FDR < 0.05) within height-associated loci (Table 3 and Supplementary Table 16). Of these 649 genes, 202 genes (31%) were significant in the GRAIL analysis (Supplementary Tables 13 and 16) and/or overlapped with a list of genes involved in abnormal skeletal growth syndromes that we assembled from the Online Mendelian Inheritance in Man (OMIM) database¹² ($n = 40$; Supplementary Tables 9 and 16). Many other newly prioritized genes had additional supporting evidence (Supplementary Table 16), including specific expression in the growth plate¹² and/or connections to relevant pathways (for example: *GLI2* and *LAMA5* (Hedgehog signaling); *FRS2* (FGF signaling); *AXIN2*, *NFATC1*, *CTNGB1*, *FBXW11*, *WNT4*, *WNT5A* and *VANGL2* (WNT/ β -catenin signaling); *SMAD3* and *MTOR* (transforming growth factor (TGF)- β and/or mTOR signaling); *WWP2-MIR140*, *IBSP*, *SHOX2* and *SP3* (the corresponding genes are required in mice for proper bone and cartilage formation); *CHSY1*, *DSE* and *PCOLCE2* (glycosaminoglycan/collagen metabolism); and *SCARA3*, *COPZ2*, *TBX18*, *CRISPLD1* and *SLIT3* (differential expression in growth plate and predicted to be in highly relevant pathways)).

DEPICT also prioritized the genes that were new candidates for having a role in skeletal growth. The genes newly and strongly implicated in this study included not only genes with obvious relationships to skeletal biology, such as *SOX5* and collagen genes, but also genes that have no clear published connection to skeletal growth and likely represent as-yet-unknown biology (Table 3 and Supplementary Table 16). DEPICT strongly prioritized genes that did not have published annotations related to growth pathways but were predicted to be in gene sets that were both enriched in the associated loci and clearly connected to growth. These included genes newly predicted to be in pathways related to cartilage or bone development (*FAM101A*, *CRISPLD1* and the noncoding RNA *LINC00476*), collagen or extracellular matrix (*GLT8D2*, *CCDC3* and *ZCCHC24*), histone demethylation (*ATAD2B* and *TSTD2*) and other genes predicted to have skeletal phenotypes but not currently annotated as belonging to relevant pathways (*ARSJ*, *PSKH1*, *COPZ2*, *ADAMTS17* and the microRNA cluster *MIR17HG*). Of note, mutations in both *ADAMTS17* and *MIR17HG* have been identified as causes of syndromic short stature in humans^{15,16}.

As suggested by the prioritization of *ADAMTS17* and *MIR17HG*, it is possible that some of the newly highlighted genes might also underlie new syndromes of abnormal skeletal growth. As a further proof of principle, the second entry on our list of prioritized genes (Table 3 and Supplementary Table 16), *CHSY1*, was not a known monogenic gene in the OMIM database¹² when we assembled our list, but mutations in this gene have since been shown to cause a syndrome including brachydactyly and short stature^{17,18}. Thus, the novel DEPICT method, applied to the larger GWAS data set, not only identified similar biology to GRAIL and MAGENTA but also implicated a large number of additional genes, gene sets and pathways that are likely important in skeletal biology and human growth.

Table 3 Significantly prioritized new human growth-associated genes

Locus (height SNP)	Gene	New locus	Prioritization P value	Lines of supporting evidence	Top-ranking reconstituted gene sets
Genes with previous literature support (GRAIL)					
rs10748128	<i>FRS2</i>	No	1.0×10^{-16}	7	PI3K cascade (REACTOME, $P = 6.2 \times 10^{-13}$) Chronic myeloid leukemia (KEGG, $P = 1.6 \times 10^{-12}$) Response to fibroblast growth factor stimulus (GO, $P = 5.4 \times 10^{-11}$)
rs2166898	<i>GLI2</i>	Yes	4.4×10^{-16}	7	Growth factor binding (GO, $P = 2.6 \times 10^{-14}$) Regulation of osteoblast differentiation (GO, $P = 2.3 \times 10^{-11}$) WNT protein binding (GO, $P = 1.9 \times 10^{-12}$)
rs526896-rs9327705	<i>TBX4</i>	No	9.9×10^{-9}	7	Short mandible (MP, $P = 3.3 \times 10^{-19}$) Respiratory system development (GO, $P = 3.1 \times 10^{-17}$) Abnormal ulna morphology (MP, $P = 1.9 \times 10^{-15}$)
rs16860216	<i>SOX8</i>	No	0.016	7	Small thoracic cage (MP, $P = 6.9 \times 10^{-14}$) Short ribs (MP, $P = 2.7 \times 10^{-8}$) Short sternum (MP, $P = 6.5 \times 10^{-7}$)
rs1199734	<i>LATS2</i>	Yes	1.0×10^{-16}	6	Partial lethality throughout fetal growth and development (MP, $P = 1.2 \times 10^{-18}$) Growth factor binding (GO, $P = 2.6 \times 10^{-14}$) TGFB1 protein complex (InWeb, $P = 6.3 \times 10^{-12}$)
rs12323101	<i>PDS5B</i>	No	1.0×10^{-16}	6	Chromatin binding (GO, $P = 6.4 \times 10^{-17}$) Nuclear hormone receptor binding (GO, $P = 2.4 \times 10^{-12}$) RBBP4 protein complex (InWeb, $P = 1.3 \times 10^{-11}$) WNT16 protein complex (InWeb, $P = 1.9 \times 10^{-8}$)
rs6746356	<i>SP3</i>	Yes	1.0×10^{-16}	6	BCOR protein complex (InWeb, $P = 2.7 \times 10^{-17}$) AFF2 protein complex (InWeb, $P = 4.5 \times 10^{-7}$) Intracellular steroid hormone receptor signaling pathway (GO, $P = 9.0 \times 10^{-6}$)
rs3923086	<i>AXIN2</i>	Yes	2.2×10^{-16}	6	Signaling by transforming growth factor β (KEGG, $P = 3.8 \times 10^{-15}$) WNT receptor signaling pathway (GO, $P = 6.9 \times 10^{-14}$) Polydactyly (MP, $P = 1.5 \times 10^{-10}$)
rs3790086	<i>LTBP1</i>	No	1.3×10^{-13}	6	Abnormal skeleton morphology (MP, $P = 1.1 \times 10^{-15}$) TGF β signaling pathway (KEGG, $P = 3.8 \times 10^{-15}$) Growth factor binding (GO, $P = 2.6 \times 10^{-14}$)
rs2034172	<i>WNT5A</i>	Yes	4.3×10^{-13}	6	Partial lethality throughout fetal growth and development (MP, $P = 1.2 \times 10^{-18}$) Tissue morphogenesis (GO, $P = 4.1 \times 10^{-20}$) Abnormal skeleton morphology (MP, $P = 1.1 \times 10^{-15}$)
rs3915129	<i>CTNNB1</i>	Yes	3.5×10^{-12}	6	AR protein complex (InWeb, $P = 8.9 \times 10^{-17}$) TCEB1 protein complex (InWeb, $P = 1.5 \times 10^{-11}$) GTF2I protein complex (InWeb, $P = 4.6 \times 10^{-11}$)
rs12330322	<i>BMP2</i>	No	5.6×10^{-10}	6	Transcription factor binding (GO, $P = 4.7 \times 10^{-26}$) Complete embryonic lethality during organogenesis (MP, $P = 4.9 \times 10^{-21}$) Short mandible (MP, $P = 3.3 \times 10^{-19}$)
rs10958476-rs6999671	<i>BMP6</i>	No	2.9×10^{-8}	6	Small basisphenoid bone (MP, $P = 8.9 \times 10^{-17}$) TGF β signaling pathway (KEGG, $P = 3.8 \times 10^{-15}$) Growth factor binding (GO, $P = 2.6 \times 10^{-14}$)
rs564914	<i>SOX5</i>	Yes	4.6×10^{-7}	6	Disproportionate dwarf (MP, $P = 1.8 \times 10^{-13}$) Abnormal cartilage morphology (MP, $P = 1.9 \times 10^{-13}$) Short limbs (MP, $P = 2.8 \times 10^{-13}$)
rs17807185	<i>WNT4</i>	Yes	4.6×10^{-7}	6	Morphogenesis of an epithelium (GO, $P = 2.3 \times 10^{-17}$) Gland development (GO, $P = 5.4 \times 10^{-16}$) Basal cell carcinoma (KEGG, $P = 1.5 \times 10^{-12}$)
New genes without previous evidence					
rs8042424	<i>CHSY1</i>	No	1.0×10^{-16}	7	Abnormal cartilage morphology (MP, $P = 1.9 \times 10^{-13}$) Abnormal bone ossification (MP, $P = 2.1 \times 10^{-12}$) Signaling by transforming growth factor β (REACTOME, $P = 5.9 \times 10^{-9}$)
rs7652177	<i>FNDC3B</i>	No	1.0×10^{-16}	5	Abnormal spongiotrophoblast layer morphology (MP, $P = 3.2 \times 10^{-16}$) Decreased length of long bones (MP, $P = 2.7 \times 10^{-12}$) ITGB1 protein complex (InWeb, $P = 5.2 \times 10^{-8}$)
rs7284476	<i>TRIOBP</i>	Yes	1.0×10^{-16}	5	Negative regulation of cell proliferation (GO, $P = 4.3 \times 10^{-17}$) Abnormal vitelline vasculature morphology (MP, $P = 1.7 \times 10^{-15}$) β -catenin binding (GO, $P = 3.0 \times 10^{-5}$)
rs2149163-rs3927536	<i>BNC2</i>	No	1.0×10^{-16}	5	Short ulna (MP, $P = 4.7 \times 10^{-13}$) Abnormal joint morphology (MP, $P = 8.6 \times 10^{-11}$) Regulation of chondrocyte differentiation (GO, $P = 2.9 \times 10^{-9}$)
rs3790086	<i>WWP2</i>	Yes	1.0×10^{-16}	5	Cartilage development (GO, $P = 2.0 \times 10^{-19}$) Chondrocyte differentiation (GO, $P = 3.0 \times 10^{-15}$) Signaling by platelet-derived growth factor (REACTOME, $P = 4.8 \times 10^{-10}$)

The table lists 20 genes prioritized by DEPICT. Genes are ranked by the number of lines of supporting evidence and the DEPICT P value (Supplementary Table 16). Lines of supporting evidence for a gene included (1) annotation by GRAIL, (2) differential expression within the growth plate, (3) specific expression within the growth plate, (4) a mouse skeletal growth phenotype, (5) DEPICT FDR < 0.05, (6) membership in a DEPICT-prioritized gene set, (7) being the nearest gene to the lead SNP, (8) the lead SNP being an eQTL for the gene and (9) the lead SNP being a missense SNP for the gene. Because 20 of the 30 top-ranked genes were in a curated list of genes known to cause syndromes of abnormal skeletal growth¹², these 'OMIM genes' are not shown here. The top 15 genes with previous literature support (based on GRAIL) are shown, followed by the top 5 new genes. Each gene is accompanied by the significantly enriched reconstituted gene sets in which it appears (DEPICT gene set enrichment analysis). GO, Gene Ontology; MP, Mice Phenotypes from Mouse Genome Informatics database; InWeb, protein-protein interaction complexes; KEGG and REACTOME databases.

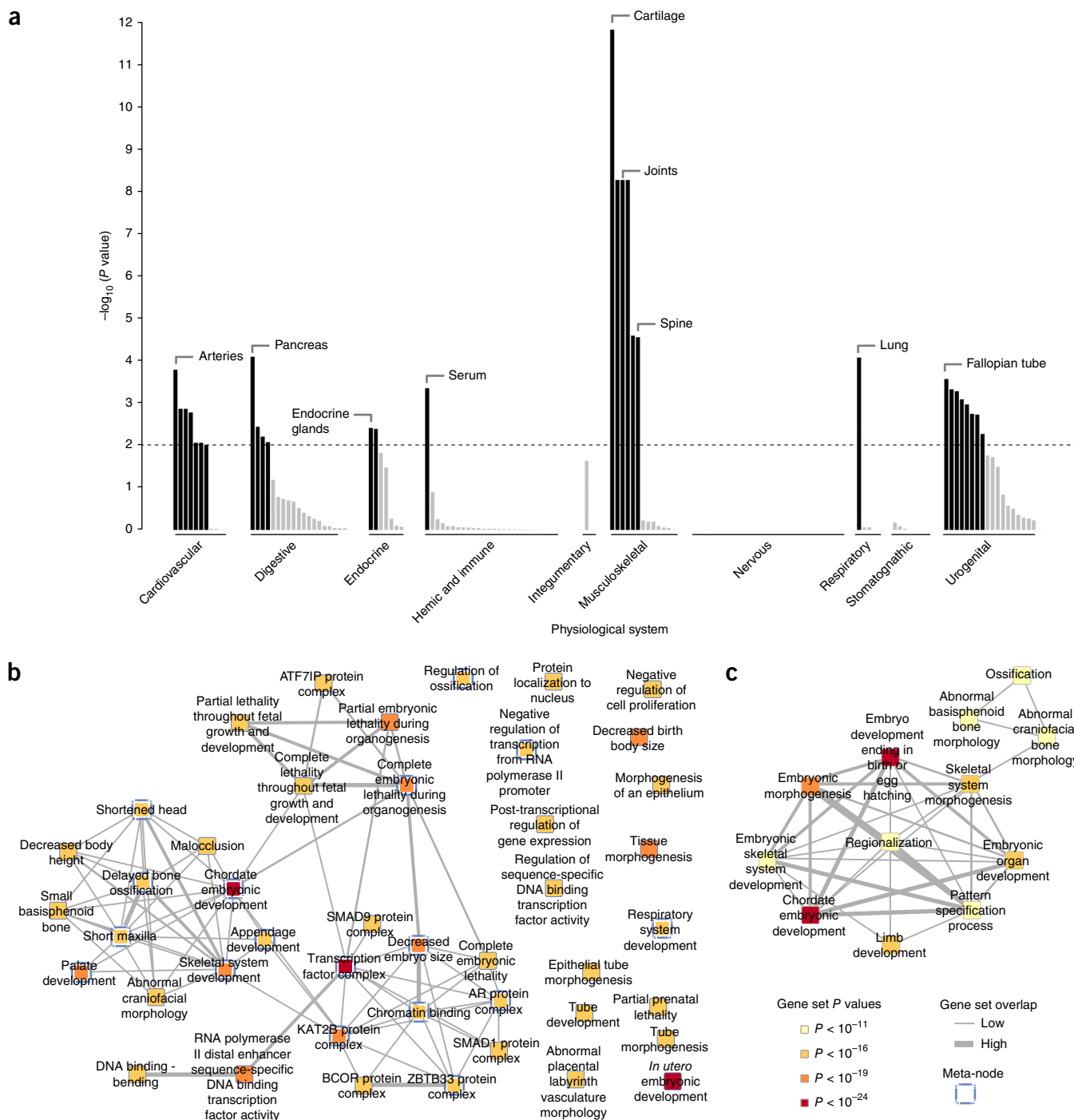


Figure 3 Tissue enrichment combined with pruned gene set network synthesis. Genes within genome-wide significant height-associated loci were enriched for several relevant tissue annotations as well as gene sets. **(a)** Genes in associated loci tended to be highly expressed in tissues related to chondrocytes and osteoblasts (cartilage, joints and spine) and other musculoskeletal, cardiovascular and endocrine tissue types. The analysis was conducted using the DEPICT method and 37,427 human microarray samples. Tissue annotations are sorted by physiological system and significance. Data for significantly enriched ($FDR < 0.05$) tissues are shown in black. **(b)** Significantly enriched reconstituted gene sets ($P < 1 \times 10^{-11}$, $FDR < 1 \times 10^{-5}$) identified by DEPICT. Nodes represent reconstituted gene sets and are colored by statistical significance. Edge thickness between nodes is proportional to the degree of gene overlap as measured by the Jaccard index. Nodes with gene overlap greater than 25% were collapsed into single meta-nodes and are marked by blue borders. **(c)** Reconstituted gene sets comprised by the chordate embryonic development meta-node, which represented several gene sets relevant to human height (for example, ossification, embryonic skeletal system development and limb development).

DISCUSSION

By performing a large GWAS on adult height, a highly heritable polygenic trait, we have provided answers to several current questions of

relevance to the genetic study of polygenic diseases and traits. First, we showed that, by conducting larger GWAS, we can identify SNPs that explain a substantial proportion of the heritability attributable

to common variants. As hypothesized by Yang *et al.*⁵, the heritability directly accounted for by variants identified by GWAS and inferred by whole-genome estimation approaches are converging with increasing sample size. The variance explained by genome-wide significant SNPs has increased from 3–5% with discovery samples of ~25,000 (ref. 19) to 10% with a discovery sample size of ~130,000 (ref. 6) and 16% with a discovery sample size of 250,000 (this study), and the variance explained from all captured common SNPs is ~50% (refs. 4,5). The variance explained by genome-wide significant SNPs on a chromosome is also proportional to chromosome length, consistent with the conclusion made by Yang *et al.*⁵ using all SNPs (**Supplementary Fig. 9**). Our new results show that ~21%, ~24% and ~29% of phenotypic variance in independent validation samples is captured by the best ~2,000, ~3,700 and ~9,500 SNPs, respectively, selected in the discovery samples (**Table 1**), and that the correlation between actual and predicted height in independent samples from the same population has increased to 0.41 (maximum prediction $R^2 = 0.41^2 = 0.17$; **Fig. 2d**). The results are consistent with a genetic architecture for human height that is characterized by a very large but finite number (thousands) of causal variants, located throughout the genome but clustered in both a biological and genomic manner. Such a genetic architecture may be described as pseudo-infinitesimal and may characterize many other polygenic traits and diseases. There is also strong evidence of multiple alleles at the same locus segregating in the population and for associated loci overlapping with mendelian forms, suggesting a large but finite genomic mutational target for height and effect sizes ranging from minute (<1 mm; ~0.01 s.d.) to gigantic (>300 mm; >3 s.d.; in the case of monogenic mutations).

It has been argued that the biological information emerging from GWAS will become less relevant as sample sizes increase because, as thousands of associated variants are discovered, the range of implicated genes and pathways will lose specificity and cover essentially the entire genome²⁰. If this were the case, then increasing sample sizes would not help to prioritize follow-up studies aimed at identifying and understanding new biology and the associated loci would blanket the entire genome. Our study provides strong evidence to the contrary: the identification of many hundred and even thousand associated variants can continue to provide biologically relevant information. In other words, the variants identified in larger sample sizes both display a stronger enrichment for pathways clearly relevant to skeletal growth and prioritize many additional new and relevant genes. Furthermore, the associated variants are often non-randomly and tightly clustered (typically separated by <250 kb), resulting in the frequent presence of multiple associated variants in a locus. The observations that genes and especially pathways are now beginning to be implicated by multiple variants suggests that the larger set of results retain biological specificity but that, at some point, a new set of associated variants will largely highlight the same genes, pathways and biological mechanisms as have already been seen. This endpoint (which we have clearly not reached for height) could be considered analogous to reaching 'saturation' in model organism mutagenesis screens, where new alleles typically map to previously identified genes²¹.

We have identified a large number of gene sets and pathways that are enriched for associations with height. Although the number of gene sets and pathways is large, many overlap and likely represent multiple annotations of a much smaller set of core biological mechanisms. We also highlight individual genes within associated loci as being relevant to skeletal growth, including candidates that might contribute to syndromes of abnormal skeletal growth; for example, we strongly implicated *CHSY1*, recently identified as an underlying cause of a monogenic syndrome with short stature and brachydactyly^{17,18}.

The lists of prioritized genes and pathways should therefore provide a rich trove of data for future studies of skeletal growth; to facilitate such studies, we have made our results (including the genome-wide association results and the complete list of highlighted genes and pathways) publicly available. On the basis of the results of large genetic studies of height, we anticipate that increasing the number of associated loci for other traits and diseases could yield similarly rich lists that would generate new biological hypotheses and motivate future research into the basis of human biology and disease.

URLs. Genetic Investigation of Anthropometric Traits (GIANT) Consortium, http://www.broadinstitute.org/collaboration/giant/index.php/GIANT_consortium; Mouse Genome Informatics, <http://www.informatics.jax.org/>.

METHODS

Methods and any associated references are available in the [online version of the paper](#).

Note: Any Supplementary Information and Source Data files are available in the online version of the paper.

ACKNOWLEDGMENTS

A full list of acknowledgments appears in the **Supplementary Note**.

AUTHOR CONTRIBUTIONS

Steering committee overseeing the consortium: G.R.A., T.L.A., I.B., S.I.B., M. Boehnke, I.B.B., P.D., C.S.F., T.M.F., L.C.G., I.M.H., J.N.H., D.J.H., E.I., R.C.K., R.J.F.L., M.I.M., K.L. Mohlke, K.E.N., J.R.O., D. Schlessinger, D.P.S., U.T. and C.M.v.D. Writing group (wrote, edited and commented on the manuscript): S.I.B., D.I.C., A.Y.C., T.E., T.M.F., J.N.H., E.I., T.H.P., S.V., P.M.V., M.N.W., A.R.W. and J.Y. Data preparation group (checked and prepared data from contributing cohorts for meta-analyses): D.C.C.-C., F.R.D., T.E., T. Fall, T. Ferreira, S.G., I.M.H., Z.K., C.M.L., A.E.L., R.J.F.L., J. Luan, R.M., J.C.R., A. Scherag, E.K.S., S.V., T.W.W., A.R.W. and T. Workalemahu. Height meta-analyses group (GWAS and Metabochip) (analyses specific to the manuscript): T.E., T.M.F. (chair), S.V., P.M.V., A.R.W. (lead meta-analyses) and J.Y. (lead joint-effects and approximate conditional analyses). Mixed linear model analyses: J.S.B., M. Boehnke, D.I.C., A.Y.C., K.E., T.M.F. (chair), S.G., J.N.H., J.H.Z., E.I., A.U.J., Z.K., R.J.F.L., J. Luan, A. Metspalu, E.M., J.R.O., A.L.P., A.G.U., S.V., P.M.V., M.N.W., A.R.W. (lead) and J.Y. Large λ group: T.M.F., J.N.H., P.M.V., M.E. Goddard, A.L.P., M.N.W., J.Y. and G.R.A. Family transmission analyses: G.R.A., N.A., I.B.B., Y.D., C.M.v.D., J.N.H. (chair), E.I., J.R.O., E.P., S.V. (lead), P.M.V. and J.Y. Variance, heritability and prediction analyses: K.E., M.E. Goddard, M.I.M., A.A.E.V., P.M.V. (chair), M.N.W., A.R.W. and J.Y. (lead). Biological enrichment and pathway analyses: T.E. (lead biological enrichment analyses), J.N.H. (chair) and T.H.P. (lead pathway analyses). ENCODE working group: M.L.B., G.L. (chair) and K.S.L. Gene expression (eQTL) working group: J. Baron, T.E. (chair), L. Franke, J. Karjalainen, J.C.L., A. Metspalu, E.R., J.E.P. and H.-J.W. (lead). Other contributions: (DEPICT) R.F., L. Franke, J. Karjalainen and T.H.P.

Project design, management and coordination of contributing studies

Previous GWAS: (AGES) V. Gudnason, T.B.H.; (AMISH) A.R.S.; (ARIC) K.E.N.; (B58C T1D CONTROLS) D.P.S.; (B58C WTCCC) D.P.S.; (BRIGHT) M.J.B., N.J.S.; (CAPS) E.I.; (CHS) J.I.R.; (COLAUS) J.S.B., S. Bergmann; (CROATIA-Vis) I.R.; (deCODE) K. Stefansson, U.T.; (DGI) L.C.G.; (EGCUT) A. Metspalu; (EPIC-Norfolk) N.J.W.; (FENLAND) N.J.W.; (Finnish Twin Cohort) J. Kaprio, K. Silventoinen; (FRAM) L.A.C.; (FUSION) R.N.B., M. Boehnke; (GerMIFS I) J.E., C. Hengstenberg; (GerMIFS II) H. Schunkert; (H2000) S. Koskinen; (HFPS) D.J.H.; (KORA S4) C.G., A.P.; (MICROS) A.A.H., P.P.P.; (NFC66) M.-R.J., S. Sebert; (NHS) D.J.H.; (NSPHS) U.G.; (NTRNESDA) D.I.B.; (ORCADES) H.C.; (PLCO) S.I.B., S.J.C.; (RS I) C.M.v.D., A. Hofman, M. Kayser, F. Rivadeneira, A.G.U.; (RUNMC) L.A.K.; (SARDINIA) G.R.A.; (SASBAC) E.I.; (SHIP) R.B., H.V.; (WGHS) P.M.R.; (WTCCC-CAD) A.S.H., N.J.S.; (WTCCC-T2D) C.M.L., M.I.M.; (Young Finns Study (YFS)) T.L., O.T.R.

New GWAS: (ASCOT) M.J.C., P.S.; (ATCG) P.I.W.d.B., D.W.H.; (Athero-Express Biobank Studies) F.W.A., H.M.d.R., F.L.M., G.P.; (B-PROOF) R.D.-R., L.C.P.G.M.d.G., N.M.v.S., N.v.d.V.; (BLSA) L. Ferrucci; (CLHNS) K.L. Mohlke; (COROGENE) M.P., J. Sinisalo; (DESIR) S.C., P.F.; (DNBS) M. Melbye, J.C.M.; (EGCUT) A. Metspalu; (eMERGE) M.G.H.; (ERF) B.A.O., C.M.v.D.; (FamHS) I.B.B.; (FINGESTURE) J.-C.T.; (GOOD) C.O.; (HBCS) J.G.E.;

(Health ABC) T.B.H., Y. Liu; (HERITAGE Family Study) C. Bouchard, D.C.R., M.A. Sarzynski; (InCHIANTI) L. Ferrucci, T.M.F.; (IPM) E.P.B., R.J.F.L.; (LLS) P.E.S.; (LOLIPOP) J.C.C., J.S.K.; (MGS) P.V.G.; (NELSON) P.I.W.d.B., P.Z.; (PLCO2) S.I.B., S.J.C.; (PREVEND) P.v.d.H.; (PROCARDIS) H.W.; (PROSPER/PHASE) I.F., J.W.J.; (QFS) C. Bouchard, A. Marette, L.P., M.-C.V.; (QIMR) A.C.H., N.G.M., G.W.M.; (RISC) E.F., T.M.F., A. Golay, M. Walker; (RS II) A. Hofman, M. Kayser, F. Rivadeneira, A.G.U.; (RS III) A. Hofman, M. Kayser, F. Rivadeneira, A.G.U.; (SHIP-TREND) R.B., H.V.; (SORBS) A. Tönjes; (TRAILS) A.J.O., H. Snieder; (TWINGENE) E.I.; (TwinsUK) T.D.S.

MetaboChip studies: (ADVANCE) T.L.A., T.Q.; (AMC-PAS) G.K.H., P.D.; (ARIC) E.B., K.E.N.; (B1958C) E.H., C.P.; (BHS) J. Beilby, J. Hui; (CARDIOGENICS) P.D., W.H.O., H. Schunkert; (DESIR) S.C., P.F.; (DGE DietGeneExpression) B.J.; (DIAGEN) S.R.B., P.E.H.S.; (DILGOM) P.J., A.M.J., S. Männistö, M.P., V. Salomaa; (DPS) M.U.; (DR's EXTRA) T.A.L., R. Rauramaa; (DUNDEE-GoDARTS) C.N.A.P.; (EAS) J.F.P.; (EGCUT) A. Metspalu; (EMIL (SWABIA)) B.O.B.; (FBPP) A.C., R.S.C., S.C.H.; (FIN-D2D 2007) S.M.K.-K., T.E.S.; (FUSION 2) F.S.C., J. Saramies, J.T.; (GLACIER) P.W.F.; (GxE) R.S.C., J.N.H., C.A.M.; (HNR) R.E., P. Hoffmann, S. Moebus; (HUNT 2) K.H.; (IMPROVE) U.d.F., A. Hamsten, S.E.H., E.T.; (KORA S3) T.M., H.-E.W.; (KORA S4) K. Strauch; (Leipzig) M.S.; (LURIC) M.W.; (MEC) C.A. Haiman, L.L.M.; (METSIM) J. Kuusisto, M. Laakso; (MORGAM) P.A., D. Arveiler, P. Brambilla, J.F., F.K., J.V.; (NSHD) D.K.; (PIVUS) E.I.; (PROMIS) J. Danesh, P.D., D. Saleheen; (ScarfSheep) A. Hamsten; (SPT) R.S.C., J.N.H., C.A.M.; (STR) E.I., (Tandem) M. Bochud, P. Bovet; (THISEAS) G. Dedoussis, P.D.; (Tromsø) I.N.; (ULSAM) E.I.; (WHI) T.C.M., C.K., U.P.; (Whitehall) A.D.H., M. Kivimaki, N.J.W.; (WTCCC-T2D) C.M.L., M.I.M.

Genotyping of contributing studies

Previous GWAS: (AGES) A.V. Smith; (B58C T1D CONTROLS) W.L.M.; (B58C WTCCC) W.L.M.; (CAPS) H. Grönberg; (CROATIA-Vis) C. Hayward; (EGCUT) M. Nelis; (EPIC-Norfolk) N.J.W.; (FENLAND) N.J.W.; (Finnish Twin Cohort) J. Kaprio; (KORA S3) T.L., M.M.-N.; (MICROS) A.A.H.; (NFBC66) M.-R.J.; (ORCADES) A.F.W.; (PLCO) S.J.C.; (RS I) K.E., C.M.-G., F. Rivadeneira, A.G.U.; (SASBAC) P. Hall; (SHIP) A. Hannemann, M. Nauck; (WGHS) D.I.C., L.M.R.; (WTCCC-CAD) A.S.H., N.J.S.; (WTCCC-T2D) A.T.H., M.I.M.; (Young Finns Study (YFS)) T.L., O.T.R.

New GWAS: (ASCOT) P.B.M.; (ATCG) P.I.W.d.B., D.W.H., P.J.M.; (Athero-Express Biobank Study) S.W.v.d.L.; (CLHNS) D.C.C.-C.; (DESIR) E.E., S. Lobbens; (EGCUT) T.E., L.M.; (eMERGE) D.C.C., M.G.H.; (ERF) A.I., B.A.O., C.M.v.D.; (FamHS) I.B.B., E.W.D., M.F.F., A.T.K., M.K.W., Q.Z.; (GOOD) C.O., M. Lorentzon; (Health ABC) Y. Liu; (HERITAGE Family Study) M.A. Sarzynski; (HYPERGENES) S. Lupoli; (IPM) E.P.B.; (LifeLines) M.A. Swertz; (LLS) J. Deelen, Q.H.; (LOLIPOP) J.C.C., J.S.K.; (NELSON) J. Smolonska; (PLCO2) S.J.C., K.B.J., Z.W.; (PREVEND) P.v.d.H., I.M.L.; (PROCARDIS) M.F., A. Goel; (PROSPER/PHASE) J.W.J., D.J.S., S.T.; (QFS) C. Bellis, J. Blangero; (QIMR) A.K.H.; (SHIP-TREND) A. Hannemann, M. Nauck; (RS II) K.E., C.M.-G., F. Rivadeneira, A.G.U.; (RS III) K.E., C.M.-G., F. Rivadeneira, A.G.U.; (TRAILS) M. Bruinenberg, C.A. Hartman; (TWINGENE) A. Hamsten, N.L.P.; (TwinsUK) M. Mangino, A. Moayeri; (WGHS) D.I.C., L.M.R.

MetaboChip studies: (ADVANCE) D. Absher, T.L.A., T.Q.; (AMCPAS) K. Stirrups; (ARIC) E.B., K.E.N.; (B1958C) N.R.R., C.J.G., T.J.; (BHS) G.M.A., J. Hui; (CARDIOGENICS) K. Stirrups; (DESIR) E.E., S. Lobbens; (DGE DietGeneExpression) B.J.; (DIAGEN) M.A.M.; (DUNDEE-GoDARTS) A.J.B., C.N.A.P., N.W.R.; (EAS) J.F.W.; (EGCUT) T.E., L.M.; (ELY) N.G.F., C.L., R.J.F.L., K.K.O., R.A.S., N.J.W.; (EMIL (SWABIA)) B.O.B.; (EPIC-Norfolk) N.G.F., C.L., R.J.F.L., K.K.O., R.A.S., N.J.W.; (FBPP) A.C.; (FENLAND) N.G.F., C.L., R.J.F.L., K.K.O., R.A.S., N.J.W.; (FIN-D2D 2007) P.S.C.; (FUSION 2) L.K.; (GLACIER) I.B.; (HNR) M.M.N.; (HUNT 2) N.N.; (KORA S3) N.K., M. Waldenberger; (KORA S4) H. Grallert, P.L.; (Leipzig) Y.B., P.K.; (LURIC) M.E.K.; (MEC) C.A. Haiman, L.A.H.; (NSHD) D.K., K.K.O., A.W.; (PIVUS) E.I., C. Berne, L.L., J. Sundström; (PROMIS) K. Stirrups; (STR) N.L.P.; (Tandem) G.B.E., M. Maillard; (THISEAS) K. Stirrups; (Tromsø) P.S.C.; (ULSAM) J.Ä., E.I., A.-C.S.; (WHI) C.K., U.P.; (Whitehall) C.L.; (WTCCC-T2D) A.T.H., M.I.M.

Phenotype coordination of contributing studies

Previous GWAS: (AMISH) A.R.S.; (B58C T1D CONTROLS) D.P.S.; (B58C WTCCC) D.P.S.; (BRIGHT) M.J.B., N.J.S.; (CAPS) H. Grönberg; (CHS) R.C.K.; (CROATIA-Vis) I.R.; (DGI) V. Lyssenko; (EGCUT) A. Metspalu; (EPIC-Norfolk) N.J.W.; (FENLAND) N.J.W.; (Finnish Twin Cohort) J. Kaprio; (KORA S4) A.P.; (NFBC66) M.-R.J.; (NTRNESDA) J.H.S.; (ORCADES) A.F.W.; (PLCO) S.I.B.; (RS I) A. Hofman, F. Rivadeneira, A.G.U.; (SASBAC) P. Hall; (SHIP) M. Dörr, W.H., T.K.; (UKBS-CC) J. Jolley; (WGHS) D.I.C., L.M.R., A.Y.C.; (WTCCC-CAD) A.S.H., N.J.S.; (WTCCC-T2D) A.B., A.T.H.; (Young Finns Study (YFS)) T.L., O.T.R.

New GWAS: (ASCOT) M.J.C., P.S., A.V. Stanton; (ATCG) D.W.H.; (Athero-Express Biobank Study) F.L.M.; (BSLA) S. Bandinelli; (DESIR) R. Roussel; (DNBC) H.A.B., B.F., F.G.; (EGCUT) T.E., A. Metspalu; (eMERGE) J.C.D., A.N.K.; (ERF)

B.A.O., C.M.v.D.; (FamHS) I.B.B., M.F.F.; (FINGESTURE) J. Juntila; (GOOD) C.O., M. Lorentzon; (HBSC) J.G.E.; (Health ABC) M.E. Garcia, T.B.H., M.A.N.; (HERITAGE Family Study) C. Bouchard; (HYPERGENES) P.M.; (InCHIANTI) S. Bandinelli, L. Ferrucci; (IPM) O.G.; (LifeLines) S. Scholtens, M.A. Swertz, J.M.V.; (LLS) D.v.H.; (LOLIPOP) J.C.C., J.S.K., U.A., L.O., J. Sehmi; (NELSON) P.A.D.J.; (PLCO2) S.I.B.; (PREVEND) S.J.L.B., R.T.G., H.L.H.; (PROCARDIS) R. Clarke, R. Collins, M.F., A. Hamsten; (PROSPER/PHASE) J.W.J., I.F., B.M.B.; (QFS) A. Tremblay; (QIMR) A.K.H., A.C.H., P.A.F.M., N.G.M., G.W.M.; (RS II) A. Hofman, F. Rivadeneira, A.G.U.; (RS III) A. Hofman, F. Rivadeneira, A.G.U.; (SORBS) A. Tönjes; (SHIP-TREND) M. Dörr, W.H., T.K.; (TRAILS) C.A. Hartman, R.P.S., F.V.A.v.O.; (TWINGENE) P.K.E.M., N.L.P.; (TwinsUK) M. Mangino, C.M.; (WGHS) D.I.C., L.M.R.

MetaboChip studies: (ADVANCE) A.S.G., M.A.H.; (AMCPAS) J.J.P.K.; (ARIC) E.B.; (B1958C) E.H., C.P.; (BHS) A.L.J., A.W.M.; (DESIR) R. Roussel; (DGE DietGeneExpression) B.J., I.H.C.; (DIAGEN) J. Gräßler, G.M.; (DPS) J. Lindström; (DR's EXTRA) M.H.; (DUNDEE-GoDARTS) A.S.F.D., A.D.M., C.N.A.P.; (EAS) S. McLachlan; (EGCUT) T.E., A. Metspalu; (EMIL (SWABIA)) B.O.B., S.C.-B., W.K., S. Merger, T.S., R.W.; (FBPP) R.S.C., S.C.H.; (GLACIER) G. Hallmans; (GxE) T. Forrester, B.O.T.; (HNR) R.E., S. Moebus; (HUNT 2) O.H.; (KORA S3) H.-E.W.; (Leipzig) M. Blüher; (MEC) L.R.W.; (METSIM) H.M.S.; (NSHD) D.K.; (PIVUS) C. Berne, E.I., L.L., J. Sundström; (PROMIS) D. Saleheen; (SPT) T. Forrester, B.O.T.; (STR) N.L.P.; (Tandem) M. Bochud, P. Bovet; (THISEAS) S. Kanoni; (Tromsø) T. Wilsgaard; (ULSAM) J.Ä., V. Giedraitis, E.I.; (WHI) C.K., U.P.; (Whitehall) M. Kumari; (WTCCC-T2D) A.B., A.T.H.

Data analysis

Previous GWAS: (AGES) A.V. Smith; (ARIC) K.L. Monda, K.E.N.; (B58C T1D CONTROLS) D.P.S.; (B58C WTCCC) D.P.S.; (CAPS) E.I.; (CHS) R.C.K., B.M.; (COLAUS) S. Bergmann, Z.K.; (CROATIA-Vis) C. Hayward; (deCODE) V. Steinthorsdottir, G.T.; (EGCUT) M. Nelis; (EPIC-Norfolk) J.H.Z.; (FENLAND) J. Luan; (FRAM) L.A.C., N.L.H.-C.; (FUSION) C.J.W.; (GerMIFS II) C.W.; (H2000) N.E.; (HPFS) L.Q.; (NHS) L.Q.; (NSPHS) Å.J.; (PLCO) S.I.B.; (RS I) K.E., C.M.-G., F. Rivadeneira, A.G.U.; (RUNMC) S.H.V.; (SardiNIA) S. Sanna; (SASBAC) E.I.; (SEARCH) J.P.T.; (SHIP) A. Teumer; (WGHS) D.I.C., L.M.R., A.Y.C.; (WTCCC-T2D) A.P.M., T. Ferreira, A. Mahajan, R.M.

New GWAS: (ATCG) P.I.W.d.B., P.J.M., S.R.; (Athero-Express Biobank Studies) S.W.v.d.L.; (B-PROOF) S.v.D.; (BHS) M.C.; (BSLA) T.T.; (CLHNS) D.C.C.-C.; (DESIR) S.C., L.Y.; (DNBC) B.F., F.G.; (EGCUT) T.E., K.F., T.H., R.M.; (eMERGE) M.G.H.; (ERF) N.A., A.D.; (FamHS) M.F.F.; (GOOD) C.O., M. Lorentzon; (HBSC) N.E.; (Health ABC) M.A.N.; (HERITAGE Family Study) C. Bouchard, M.A. Sarzynski, D.C.R., T.R., T.K.R., Y.J.S.; (HYPERGENES) S. Lupoli; (InCHIANTI) D.P., T.T., A.R.W.; (IPM) J. Jeff, V. Lotay, Y. Lu; (LifeLines) I.M.N., J.V.V.V.-O.; (LLS) M. Beekman, J.J.H.-D.; (LOLIPOP) W.Z.; (MGS) J. Shi; (NELSON) S.R., J.v.S.; (PLCO2) S.I.B., Z.W.; (PREVEND) P.v.d.H., I.M.L., N.V.; (PROCARDIS) A. Goel; (PROSPER/PHASE) I.F., B.M.B., S.T.; (QFS) J. Blangero, L.P.; (QIMR) G. Hemani, D.R.N., J.E.P.; (RISC) D.P., A.R.W.; (RS II) K.E., C.M.-G., F. Rivadeneira, A.G.U.; (RS III) K.E., C.M.-G., F. Rivadeneira, A.G.U.; (SHIP-TREND) A. Teumer; (SORBS) R.M.; (TRAILS) H. Snieder; (TWINGENE) E.I., S.G.; (TwinsUK) M. Mangino; (WGHS) D.I.C., L.M.R.

MetaboChip studies: (ADVANCE) D. Absher, T.L.A., L.L.W.; (AMCPAS) S. Kanoni; (ARIC) S. Buyske, A.E.J., K.E.N.; (B1958C) T. Ferreira; (BHS) D. Anderson; (CARDIOGENICS) S. Kanoni; (DESIR) S.C., L.Y.; (DGE DietGeneExpression) I.H.C.; (DIAGEN) A.U.J., G.M.; (DILGOM) K.K.; (DUNDEE) T. Ferreira; (EAS) J.L.B., R.M.F.; (EGCUT) T.E., K.F., E.M.; (ELY) J. Luan; (EMIL (SWABIA)) B.O.B.; (EPIC-Norfolk) J. Luan; (FBPP) A.C., G.B.E.; (FENLAND) J. Luan; (GLACIER) F. Renstrom, D. Shungin; (GxE) C.D.P.; (HNR) S. Pechlivanis, A. Scherag; (IMPROVE) L. Folkersen, R.J.S.; (KORA S3) J.S.R.; (KORA S4) E.A.; (Leipzig) A. Mahajan, I.P.; (LURIC) G. Delgado, T.B.G., M.E.K., S. Pilz, H. Schnarag; (MEC) U.L., F.R.S.; (METSIM) A. Stancáková; (NSHD) A.W., J. Luan; (PIVUS) S.G., E.I.; (PROMIS) S. Kanoni; (ScarfSheep) R.J.S.; (SPT) C.D.P.; (STR) E.I., S.G.; (TANDEM) G.B.E.; (THISEAS) M. Dimitriou; (ULSAM) S.G., E.I.; (WHI) J. Gong, J. Haessler, M.R.; (Whitehall) J. Luan; (WTCCC-T2D) A.P.M., T. Ferreira, A. Mahajan, R.M.

COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests.

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

1. Fisher, R.A. The correlation between relatives on the supposition of Mendelian inheritance. *Trans. R. Soc.* **52**, 399–433 (1918).
2. Silventoinen, K. *et al.* Heritability of adult body height: a comparative study of twin cohorts in eight countries. *Twin Res.* **6**, 399–408 (2003).

3. Visscher, P.M. *et al.* Assumption-free estimation of heritability from genome-wide identity-by-descent sharing between full siblings. *PLoS Genet.* **2**, e41 (2006).
4. Yang, J. *et al.* Common SNPs explain a large proportion of the heritability for human height. *Nat. Genet.* **42**, 565–569 (2010).
5. Yang, J. *et al.* Genome partitioning of genetic variation for complex traits using common SNPs. *Nat. Genet.* **43**, 519–525 (2011).
6. Lango Allen, H. *et al.* Hundreds of variants clustered in genomic loci and biological pathways affect human height. *Nature* **467**, 832–838 (2010).
7. 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).
8. 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).
9. Voight, B.F. *et al.* The Metachip, a custom genotyping array for genetic studies of metabolic, cardiovascular, and anthropometric traits. *PLoS Genet.* **8**, e1002793 (2012).
10. Yang, J. *et al.* Genomic inflation factors under polygenic inheritance. *Eur. J. Hum. Genet.* **19**, 807–812 (2011).
11. Altshuler, D.M. *et al.* Integrating common and rare genetic variation in diverse human populations. *Nature* **467**, 52–58 (2010).
12. 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).
13. Segre, A.V., Groop, L., Mootha, V.K., Daly, M.J. & Altshuler, D. Common inherited variation in mitochondrial genes is not enriched for associations with type 2 diabetes or related glycemic traits. *PLoS Genet.* **6**, e1001058 (2010).
14. Raychaudhuri, S. *et al.* Identifying relationships among genomic disease regions: predicting genes at pathogenic SNP associations and rare deletions. *PLoS Genet.* **5**, e1000534 (2009).
15. de Pontual, L. *et al.* Germline deletion of the miR-17–92 cluster causes skeletal and growth defects in humans. *Nat. Genet.* **43**, 1026–1030 (2011).
16. Morales, J. *et al.* Homozygous mutations in *ADAMTS10* and *ADAMTS17* cause lenticular myopia, ectopia lentis, glaucoma, spherophakia, and short stature. *Am. J. Hum. Genet.* **85**, 558–568 (2009).
17. Li, Y. *et al.* Temtamy preaxial brachydactyly syndrome is caused by loss-of-function mutations in chondroitin synthase 1, a potential target of BMP signaling. *Am. J. Hum. Genet.* **87**, 757–767 (2010).
18. Tian, J. *et al.* Loss of CHSY1, a secreted FRINGE enzyme, causes syndromic brachydactyly in humans via increased NOTCH signaling. *Am. J. Hum. Genet.* **87**, 768–778 (2010).
19. Visscher, P.M. Sizing up human height variation. *Nat. Genet.* **40**, 489–490 (2008).
20. Goldstein, D.B. Common genetic variation and human traits. *N. Engl. J. Med.* **360**, 1696–1698 (2009).
21. Nüsslein-Volhard, C., Wieschaus, E. & Kluding, H. Mutations affecting the pattern of the larval cuticle in *Drosophila melanogaster*: zygotic loci on the second chromosome. *Roux's Arch. Dev. Biol.* **193**, 267–282 (1984).

Andrew R Wood^{1,3,23}, Tonu Esko^{2–5,3,23}, Jian Yang^{6,7,3,23}, Sailaja Vedantam^{3,4,3,23}, Tune H Pers^{3–5,8,3,23}, Stefan Gustafsson^{9,10,3,23}, Audrey Y Chu¹¹, Karol Estrada^{4,12,13}, Jian'an Luan¹⁴, Zoltán Kutalik^{15–17}, Najaf Amin¹⁸, Martin L Buchkovich¹⁹, Damien C Croteau-Chonka^{19,20}, Felix R Day¹⁴, Yanan Duan²¹, Tove Fall^{9,10,22}, Rudolf Fehrmann²³, Teresa Ferreira²⁴, Anne U Jackson²⁵, Juha Karjalainen²³, Ken Sin Lo²⁶, Adam E Locke²⁵, Reedik Mägi^{2,24}, Evelin Mihailov^{2,27}, Eleonora Porcu²⁸, Joshua C Randall^{24,29}, André Scherag^{30,31}, Anna A E Vinkhuyzen⁶, Harm-Jan Westra²³, Thomas W Winkler³², Tsegaselassie Workalemahu³³, Jing Hua Zhao¹⁴, Devin Absher³⁴, Eva Albrecht³⁵, Denise Anderson³⁶, Jeffrey Baron³⁷, Marian Beekman^{38,39}, Ayse Demirkan^{18,40}, Georg B Ehret^{41,42}, Bjarke Feenstra⁴³, Mary F Feitosa⁴⁴, Krista Fischer², Ross M Fraser⁴⁵, Anuj Goel^{24,46}, Jian Gong⁴⁷, Anne E Justice⁴⁸, Stavroula Kanoni⁴⁹, Marcus E Kleber^{50,51}, Kati Kristiansson⁵², Unhee Lim⁵³, Vaneet Lotay⁵⁴, Julian C Lui³⁷, Massimo Mangino⁵⁵, Irene Mateo Leach⁵⁶, Carolina Medina-Gomez^{12,57,58}, Michael A Nalls⁵⁹, Dale R Nyholt⁶⁰, Cameron D Palmer^{3,4}, Dorota Pasko¹, Sonali Pechlivanis³⁰, Inga Prokopenko^{24,61,62}, Janina S Ried³⁵, Stephan Ripke^{13,63}, Dmitry Shungin^{64–66}, Alena Stancáková⁶⁷, Rona J Strawbridge⁶⁸, Yun Ju Sung⁶⁹, Toshiko Tanaka⁷⁰, Alexander Teumer⁷¹, Stella Trompet^{72,73}, Sander W van der Laan⁷⁴, Jessica van Setten⁷⁵, Jana V Van Vliet-Ostaptchouk⁷⁶, Zhaoming Wang^{77–80}, Loïc Yengo^{81–83}, Weihua Zhang^{84,85}, Uzma Afzal^{84,85}, Johan Ärnlöv^{9,10,86}, Gillian M Arscott⁸⁷, Stefania Bandinelli⁸⁸, Amy Barrett⁶¹, Claire Bellis⁸⁹, Amanda J Bennett⁶¹, Christian Berne⁹⁰, Matthias Blüher^{91,92}, Jennifer L Bolton⁴⁵, Yvonne Böttcher⁹¹, Heather A Boyd⁴³, Marcel Bruinenberg⁹³, Brendan M Buckley⁹⁴, Steven Buyske^{95,96}, Ida H Caspersen⁹⁷, Peter S Chines⁹⁸, Robert Clarke⁹⁹, Simone Claudi-Boehm¹⁰⁰, Matthew Cooper³⁶, E Warwick Daw⁴⁴, Pim A De Jong¹⁰¹, Joris Deelen^{38,39}, Graciela Delgado⁵⁰, Josh C Denny¹⁰², Rosalie Dhonukshe-Rutten¹⁰³, Maria Dimitriou¹⁰⁴, Alex S F Doney¹⁰⁵, Marcus Dörr^{77,106}, Niina Eklund^{52,107}, Elodie Eury^{81–83}, Lasse Folkersen⁶⁸, Melissa E Garcia¹⁰⁸, Frank Geller⁴³, Vilmantas Giedraitis¹⁰⁹, Alan S Go¹¹⁰, Harald Grallert^{35,111,112}, Tanja B Grammer⁵⁰, Jürgen Gräßler¹¹³, Henrik Grönberg²², Lisette C P G M de Groot¹⁰³, Christopher J Groves⁶¹, Jeffrey Haessler⁴⁷, Per Hall²², Toomas Haller², Goran Hallmans¹¹⁴, Anke Hannemann⁷⁸, Catharina A Hartman¹¹⁵, Maija Hassinen¹¹⁶, Caroline Hayward¹¹⁷, Nancy L Heard-Costa^{118,119}, Quinta Helmer^{38,120,121}, Gibran Hemani^{6,7}, Anjali K Henders⁶⁰, Hans L Hillege^{56,122}, Mark A Hlatky¹²³, Wolfgang Hoffmann^{77,124}, Per Hoffmann^{125–127}, Oddgeir Holmen¹²⁸, Jeanine J Houwing-Duistermaat^{38,120}, Thomas Illig^{111,129}, Aaron Isaacs^{18,130}, Alan L James^{131,132}, Janina Jeff⁵⁴, Berit Johansen⁹⁷, Åsa Johansson¹³³, Jennifer Jolley^{134,135}, Thorhildur Juliusdottir²⁴, Juhani Juntila¹³⁶, Abel N Kho¹³⁷, Leena Kinnunen⁵², Norman Klopp^{111,129}, Thomas Kocher¹³⁸, Wolfgang Kratzer¹³⁹, Peter Lichtner¹⁴⁰, Lars Lind¹⁴¹, Jaana Lindström⁵², Stéphane Lobbens^{81–83}, Mattias Lorentzon¹⁴², Yingchang Lu^{54,143}, Valeriya Lyssenko¹⁴⁴, Patrik K E Magnusson²², Anubha Mahajan²⁴, Marc Maillard¹⁴⁵, Wendy L McArdle¹⁴⁶, Colin A McKenzie¹⁴⁷, Stela McLachlan⁴⁵, Paul J McLaren^{148,149}, Cristina Menni⁵⁵, Sigrun Merger¹⁰⁰, Lili Milani², Alireza Moayyeri⁵⁵, Keri L Monda^{48,150}, Mario A Morken⁹⁸, Gabriele Müller¹⁵¹, Martina Müller-Nurasyid^{35,152–154}, Arthur W Musk¹⁵⁵, Narisu Narisu⁹⁸, Matthias Nauck^{77,78}, Ilja M Nolte¹²², Markus M Nöthen^{126,127}, Laticia Oozageer⁸⁴, Stefan Pilz^{156,157},

Nigel W Rayner^{24,29,61}, Frida Renstrom⁶⁴, Neil R Robertson^{24,61}, Lynda M Rose¹¹, Ronan Roussel^{158–160}, Serena Sanna²⁸, Hubert Scharnagl¹⁶¹, Salome Scholtens¹²², Fredrick R Schumacher¹⁶², Heribert Schunkert^{154,163}, Robert A Scott¹⁴, Joban Sehmi^{84,85}, Thomas Seufferlein¹³⁹, Jianxin Shi¹⁶⁴, Karri Silventoinen¹⁶⁵, Johannes H Smit^{166,167}, Albert Vernon Smith^{168,169}, Joanna Smolonska^{23,122}, Alice V Stanton¹⁷⁰, Kathleen Stirrups^{29,49}, David J Stott¹⁷¹, Heather M Stringham²⁵, Johan Sundström¹⁴¹, Morris A Swertz²³, Ann-Christine Syvänen^{9,172}, Bamidele O Tayo¹⁷³, Gudmar Thorleifsson¹⁷⁴, Jonathan P Tyrer¹⁷⁵, Suzanne van Dijk¹², Natasja M van Schoor¹⁵⁶, Nathalie van der Velde^{12,176}, Diana van Heemst^{38,73}, Floor V A van Oort¹⁷⁷, Sita H Vermeulen^{178,179}, Niek Verweij⁵⁶, Judith M Vonk¹²², Lindsay L Waite³⁴, Melanie Waldenberger¹¹¹, Roman Wennauer¹⁸⁰, Lynne R Wilkens⁵³, Christina Willenborg^{181,182}, Tom Wilsgaard¹⁸³, Mary K Wojczynski⁴⁴, Andrew Wong¹⁸⁴, Alan F Wright¹¹⁷, Qunyuan Zhang⁴⁴, Dominique Arveiler¹⁸⁵, Stephan J L Bakker¹⁸⁶, John Beilby^{87,187}, Richard N Bergman¹⁸⁸, Sven Bergmann^{16,17}, Reiner Biffar¹⁸⁹, John Blangero⁸⁹, Dorret I Boomsma¹⁹⁰, Stefan R Bornstein¹¹³, Pascal Bovet^{191,192}, Paolo Brambilla¹⁹³, Morris J Brown¹⁹⁴, Harry Campbell⁴⁵, Mark J Caulfield^{195,196}, Aravinda Chakravarti⁴¹, Rory Collins⁹⁹, Francis S Collins⁹⁸, Dana C Crawford^{197,198}, L Adrienne Cupples^{118,199}, John Danesh²⁰⁰, Ulf de Faire²⁰¹, Hester M den Ruijter^{74,202}, Raimund Erbel²⁰³, Jeanette Erdmann^{181,182}, Johan G Eriksson^{52,204,205}, Martin Farrall^{24,46}, Ele Ferrannini^{206,207}, Jean Ferrières²⁰⁸, Ian Ford²⁰⁹, Nita G Forouhi¹⁴, Terrence Forrester¹⁴⁷, Ron T Gansevoort¹⁸⁶, Pablo V Gejman²¹⁰, Christian Gieger³⁵, Alain Golay²¹¹, Omri Gottesman⁵⁴, Vilmundur Gudnason^{168,169}, Ulf Gyllenstein¹³³, David W Haas²¹², Alistair S Hall²¹³, Tamara B Harris¹⁰⁸, Andrew T Hattersley²¹⁴, Andrew C Heath²¹⁵, Christian Hengstenberg^{154,163}, Andrew A Hicks²¹⁶, Lucia A Hindorf²¹⁷, Aroon D Hingorani²¹⁸, Albert Hofman^{57,58}, G Kees Hovingh²¹⁹, Steve E Humphries²²⁰, Steven C Hunt²²¹, Elina Hyponen^{222–225}, Kevin B Jacobs^{79,80}, Marjo-Riitta Jarvelin^{85,226–230}, Pekka Jousilahti⁵², Antti M Jula⁵², Jaakko Kaprio^{52,107,231}, John J P Kastelein²¹⁹, Manfred Kayser^{57,232}, Frank Kee²³³, Sirkka M Keinänen-Kiukaanniemi^{234,235}, Lambertus A Kiemeny^{178,236}, Jaspal S Kooner^{84,237,238}, Charles Kooperberg⁴⁷, Seppo Koskinen⁵², Peter Kovacs^{91,92}, Aldi T Kraja⁴⁴, Meena Kumari²³⁹, Johanna Kuusisto²⁴⁰, Timo A Lakka^{116,241,242}, Claudia Langenberg^{14,239}, Loic Le Marchand⁵³, Terho Lehtimäki²⁴³, Sara Lupoli^{244,245}, Pamela A F Madden²¹⁵, Satu Männistö⁵², Paolo Manunta^{246,247}, André Marette^{248,249}, Tara C Matisse⁹⁶, Barbara McKnight²⁵⁰, Thomas Meitinger¹⁵⁴, Frans L Moll²⁵¹, Grant W Montgomery⁶⁰, Andrew D Morris¹⁰⁵, Andrew P Morris^{2,24,252}, Jeffrey C Murray²⁵³, Mari Nelis², Claes Ohlsson¹⁴², Albertine J Oldehinkel¹¹⁵, Ken K Ong^{14,184}, Willem H Ouwehand^{134,135}, Gerard Pasterkamp⁷⁴, Annette Peters^{111,154,254}, Peter P Pramstaller^{216,255}, Jackie F Price⁴⁵, Lu Qi^{20,33}, Olli T Raitakari^{256,257}, Tuomo Rankinen²⁵⁸, D C Rao^{44,69,215}, Treva K Rice^{69,215}, Marylyn Ritchie²⁵⁹, Igor Rudan^{45,260}, Veikko Salomaa⁵², Nilesh J Samani^{261,262}, Jouko Saramies²⁶³, Mark A Sarzynski²⁵⁸, Peter E H Schwarz^{113,264}, Sylvain Sebert²³⁰, Peter Sever²⁶⁵, Alan R Shuldiner^{266–268}, Juha Sinisalo²⁶⁹, Valgerdur Steinthorsdottir¹⁷⁴, Ronald P Stolk¹²², Jean-Claude Tardif^{26,270}, Anke Tönjes^{91,92}, Angelo Tremblay²⁷¹, Elena Tremoli²⁷², Jarmo Virtamo⁵², Marie-Claude Vohl^{249,273}, The Electronic Medical Records and Genomics (eMERGE) Consortium²⁷⁴, The MIGen Consortium²⁷⁴, The PAGE Consortium²⁷⁴, The LifeLines Cohort Study²⁷⁴, Philippe Amouyel²⁷⁵, Folkert W Asselbergs^{218,276,277}, Themistocles L Assimes¹²³, Murielle Bochud^{191,192}, Bernhard O Boehm^{100,278}, Eric Boerwinkle²⁷⁹, Erwin P Bottinger⁵⁴, Claude Bouchard¹²⁵⁸, Stéphane Cauchi^{81–83}, John C Chambers^{84,85,237}, Stephen J Chanock⁷⁹, Richard S Cooper¹⁷³, Paul I W de Bakker^{75,280,281}, George Dedoussis¹⁰⁴, Luigi Ferrucci⁷⁰, Paul W Franks^{33,64,65}, Philippe Froguel^{62,81–83}, Leif C Groop^{107,282,283}, Christopher A Haiman¹⁶², Anders Hamsten⁶⁸, M Geoffrey Hayes¹³⁷, Jennie Hui^{87,187,222,223}, David J Hunter^{20,33,284}, Kristian Hveem¹²⁸, J Wouter Jukema^{72,277,285}, Robert C Kaplan²⁸⁶, Mika Kivimäki²³⁹, Diana Kuh¹⁸⁴, Markku Laakso²⁴⁰, Yongmei Liu²⁸⁷, Nicholas G Martin⁶⁰, Winfried März^{50,161,288}, Mads Melbye^{43,123}, Susanne Moebus³⁰, Patricia B Munroe^{195,196}, Inger Njølstad¹⁸³, Ben A Oostra^{18,130,289}, Colin N A Palmer¹⁰⁵, Nancy L Pedersen²², Markus Perola^{2,52,107}, Louis Pérusse^{249,271}, Ulrike Peters⁴⁷, Joseph E Powell^{6,7}, Chris Power²²⁵, Thomas Quertermous¹²³, Rainer Rauramaa^{116,242}, Eva Reinmaa², Paul M Ridker^{11,290}, Fernando Rivadeneira^{12,57,58}, Jerome I Rotter²⁹¹, Timo E Saaristo^{292,293}, Danish Saleheen^{200,294,295}, David Schlessinger²⁹⁶, P Eline Slagboom^{38,39}, Harold Snieder¹²², Tim D Spector⁵⁵, Konstantin Strauch^{35,153}, Michael Stumvoll^{91,92}, Jaakko Tuomilehto^{52,297–299}, Matti Uusitupa^{300,301}, Pim van der Harst^{23,56,277}, Henry Völzke^{77,124}, Mark Walker³⁰², Nicholas J Wareham¹⁴, Hugh Watkins^{24,46}, H-Erich Wichmann^{303–305}, James F Wilson⁴⁵, Pieter Zanen³⁰⁶, Panos Deloukas^{29,49,307}, Iris M Heid^{32,35}, Cecilia M Lindgren^{4,24}, Karen L Mohlke¹⁹,

Elizabeth K Speliotes^{308,309}, Unnur Thorsteinsdottir^{174,310}, Inês Barroso^{29,311,312}, Caroline S Fox¹¹⁸, Kari E North^{48,313}, David P Strachan³¹⁴, Jacques S Beckmann^{16,17,315}, Sonja I Berndt⁷⁹, Michael Boehnke²⁵, Ingrid B Borecki⁴⁴, Mark I McCarthy^{24,61,316}, Andres Metspalu^{2,27}, Kari Stefansson^{174,310}, André G Uitterlinden^{12,57,58}, Cornelia M van Duijn^{18,57,58,130}, Lude Franke²³, Cristen J Willer^{309,317,318}, Alkes L Price^{4,284,319}, Guillaume Lettre^{26,270}, Ruth J F Loos^{14,54,143,320}, Michael N Weedon¹, Erik Ingelsson^{9,10,24}, Jeffrey R O'Connell^{266,267}, Goncalo R Abecasis^{25,324}, Daniel I Chasman^{11,290,324}, Michael E Goddard^{321–324}, Peter M Visscher^{6,7,324}, Joel N Hirschhorn^{3–5,324} & Timothy M Frayling^{1,324}

¹Genetics of Complex Traits, University of Exeter Medical School, University of Exeter, Exeter, UK. ²Estonian Genome Center, University of Tartu, Tartu, Estonia. ³Division of Endocrinology, Genetics and Basic and Translational Obesity Research, Boston Children's Hospital, Boston, Massachusetts, USA. ⁴Broad Institute of MIT and Harvard University, Cambridge, Massachusetts, USA. ⁵Department of Genetics, Harvard Medical School, Boston, Massachusetts, USA. ⁶Queensland Brain Institute, University of Queensland, Brisbane, Queensland, Australia. ⁷University of Queensland Diamantina Institute, Translation Research Institute, Brisbane, Queensland, Australia. ⁸Center for Biological Sequence Analysis, Department of Systems Biology, Technical University of Denmark, Lyngby, Denmark. ⁹Science for Life Laboratory, Uppsala University, Uppsala, Sweden. ¹⁰Department of Medical Sciences, Molecular Epidemiology, Uppsala University, Uppsala, Sweden. ¹¹Division of Preventive Medicine, Brigham and Women's Hospital, Boston, Massachusetts, USA. ¹²Department of Internal Medicine, Erasmus Medical Center, Rotterdam, the Netherlands. ¹³Analytic and Translational Genetics Unit, Massachusetts General Hospital and Harvard Medical School, Boston, Massachusetts, USA. ¹⁴Medical Research Council (MRC) Epidemiology Unit, University of Cambridge, Institute of Metabolic Science, Addenbrooke's Hospital, Cambridge, UK. ¹⁵Institute of Social and Preventive Medicine (IUMSP), Centre Hospitalier Universitaire Vaudois (CHUV), Lausanne, Switzerland. ¹⁶Swiss Institute of Bioinformatics, Lausanne, Switzerland. ¹⁷Department of Medical Genetics, University of Lausanne, Lausanne, Switzerland. ¹⁸Genetic Epidemiology Unit, Department of Epidemiology, Erasmus University Medical Center, Rotterdam, the Netherlands. ¹⁹Department of Genetics, University of North Carolina, Chapel Hill, North Carolina, USA. ²⁰Channing Division of Network Medicine, Department of Medicine, Brigham and Women's Hospital and Harvard Medical School, Boston, Massachusetts, USA. ²¹Department of Genetics, Division of Statistical Genomics, Washington University School of Medicine, St. Louis, Missouri, USA. ²²Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, Stockholm, Sweden. ²³Department of Genetics, University Medical Center Groningen, University of Groningen, Groningen, the Netherlands. ²⁴Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford, UK. ²⁵Center for Statistical Genetics, Department of Biostatistics, University of Michigan, Ann Arbor, Michigan, USA. ²⁶Montreal Heart Institute, Montreal, Quebec, Canada. ²⁷Institute of Molecular and Cell Biology, University of Tartu, Tartu, Estonia. ²⁸Istituto di Ricerca Genetica e Biomedica (IRGB), Consiglio Nazionale delle Ricerche, Cagliari, Italy. ²⁹Wellcome Trust Sanger Institute, Hinxton, UK. ³⁰Institute for Medical Informatics, Biometry and Epidemiology (IMIBE), University Hospital Essen, Essen, Germany. ³¹Clinical Epidemiology, Integrated Research and Treatment Center, Center for Sepsis Control and Care (CSCC), Jena University Hospital, Jena, Germany. ³²Department of Genetic Epidemiology, Institute of Epidemiology and Preventive Medicine, University of Regensburg, Regensburg, Germany. ³³Department of Nutrition, Harvard School of Public Health, Harvard University, Boston, Massachusetts, USA. ³⁴HudsonAlpha Institute for Biotechnology, Huntsville, Alabama, USA. ³⁵Institute of Genetic Epidemiology, Helmholtz Zentrum München—German Research Center for Environmental Health, Neuherberg, Germany. ³⁶Telethon Institute for Child Health Research, Centre for Child Health Research, University of Western Australia, Perth, Western Australia, Australia. ³⁷Section on Growth and Development, Program in Developmental Endocrinology and Genetics, Eunice Kennedy Shriver National Institute of Child Health and Human Development, US National Institutes of Health, Bethesda, Maryland, USA. ³⁸Netherlands Consortium for Healthy Aging (NCHA), Leiden University Medical Center, Leiden, the Netherlands. ³⁹Department of Molecular Epidemiology, Leiden University Medical Center, Leiden, the Netherlands. ⁴⁰Department of Human Genetics, Leiden University Medical Center, Leiden, the Netherlands. ⁴¹Center for Complex Disease Genomics, McKusick-Nathans Institute of Genetic Medicine, Johns Hopkins University School of Medicine, Baltimore, Maryland, USA. ⁴²Department of Specialties of Internal Medicine, Division of Cardiology, Geneva University Hospital, Geneva, Switzerland. ⁴³Department of Epidemiology Research, Statens Serum Institut, Copenhagen, Denmark. ⁴⁴Department of Genetics, Washington University School of Medicine, St. Louis, Missouri, USA. ⁴⁵Centre for Population Health Sciences, University of Edinburgh, Edinburgh, UK. ⁴⁶Radcliffe Department of Medicine, Division of Cardiovascular Medicine, University of Oxford, Oxford, UK. ⁴⁷Division of Public Health Sciences, Fred Hutchinson Cancer Research Center, Seattle, Washington, USA. ⁴⁸Department of Epidemiology, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina, USA. ⁴⁹William Harvey Research Institute, Barts and The London School of Medicine and Dentistry, Queen Mary University of London, London, UK. ⁵⁰Vth Department of Medicine (Nephrology, Hypertensiology, Endocrinology, Diabetology, Rheumatology), Medical Faculty of Mannheim, University of Heidelberg, Heidelberg, Germany. ⁵¹Department of Internal Medicine II, Ulm University Medical Center, Ulm, Germany. ⁵²National Institute for Health and Welfare, Helsinki, Finland. ⁵³Epidemiology Program, University of Hawaii Cancer Center, Honolulu, Hawaii, USA. ⁵⁴Charles Bronfman Institute for Personalized Medicine, Icahn School of Medicine at Mount Sinai, New York, New York, USA. ⁵⁵Department of Twin Research and Genetic Epidemiology, King's College London, London, UK. ⁵⁶Department of Cardiology, University Medical Center Groningen, University of Groningen, Groningen, the Netherlands. ⁵⁷Netherlands Consortium for Healthy Aging (NCHA), Rotterdam, the Netherlands. ⁵⁸Department of Epidemiology, Erasmus Medical Center, Rotterdam, the Netherlands. ⁵⁹Laboratory of Neurogenetics, National Institute on Aging, US National Institutes of Health, Bethesda, Maryland, USA. ⁶⁰QIMR Berghofer Medical Research Institute, Brisbane, Queensland, Australia. ⁶¹Oxford Centre for Diabetes, Endocrinology and Metabolism, University of Oxford, Oxford, UK. ⁶²Department of Genomics of Common Disease, School of Public Health, Imperial College London, Hammersmith Hospital, London, UK. ⁶³Stanley Center for Psychiatric Research, Broad Institute of MIT and Harvard University, Cambridge, Massachusetts, USA. ⁶⁴Department of Clinical Sciences, Genetic and Molecular Epidemiology Unit, Lund University Diabetes Center, Skåne University Hospital, Malmö, Sweden. ⁶⁵Department of Public Health and Clinical Medicine, Unit of Medicine, Umeå University, Umeå, Sweden. ⁶⁶Department of Odontology, Umeå University, Umeå, Sweden. ⁶⁷Department of Medicine, University of Eastern Finland, Kuopio, Finland. ⁶⁸Atherosclerosis Research Unit, Center for Molecular Medicine, Department of Medicine, Karolinska Institutet, Stockholm, Sweden. ⁶⁹Division of Biostatistics, Washington University School of Medicine, St. Louis, Missouri, USA. ⁷⁰Translational Gerontology Branch, National Institute on Aging, Baltimore, Maryland, USA. ⁷¹Interfaculty Institute for Genetics and Functional Genomics, University Medicine Greifswald, Greifswald, Germany. ⁷²Department of Cardiology, Leiden University Medical Center, Leiden, the Netherlands. ⁷³Department of Gerontology and Geriatrics, Leiden University Medical Center, Leiden, the Netherlands. ⁷⁴Experimental Cardiology Laboratory, Division of Heart and Lungs, University Medical Center Utrecht, Utrecht, the Netherlands. ⁷⁵Department of Medical Genetics, University Medical Center Utrecht, Utrecht, the Netherlands. ⁷⁶Department of Endocrinology, University of Groningen, University Medical Center Groningen, Groningen, the Netherlands. ⁷⁷DZHK (Deutsches Zentrum für Herz-Kreislaufforschung—German Centre for Cardiovascular Research), partner site Greifswald, Greifswald, Germany. ⁷⁸Institute of Clinical Chemistry and Laboratory Medicine, University Medicine Greifswald, Greifswald, Germany. ⁷⁹Division of Cancer Epidemiology and Genetics, National Cancer Institute, US National Institutes of Health, Bethesda, Maryland, USA. ⁸⁰Core Genotyping Facility, SAIC-Frederick, Inc., NCI-Frederick, Frederick, Maryland, USA. ⁸¹CNRS UMR 8199, Lille, France. ⁸²European Genomic Institute for Diabetes, Lille, France. ⁸³Université de Lille 2, Lille, France. ⁸⁴Ealing Hospital National Health Service (NHS) Trust, Middlesex, UK. ⁸⁵Department of Epidemiology and Biostatistics, Imperial College London, London, UK. ⁸⁶School of Health and Social Studies, Dalarna University, Falun, Sweden. ⁸⁷PathWest Laboratory Medicine of Western Australia, Nedlands, Western Australia, Australia. ⁸⁸Geriatric Unit, Azienda Sanitaria Firenze (ASF), Florence, Italy. ⁸⁹Department of Genetics, Texas Biomedical Research Institute, San Antonio, Texas, USA. ⁹⁰Department of Medical Sciences, Endocrinology, Diabetes and Metabolism, Uppsala University, Uppsala, Sweden. ⁹¹Integrated Research and Treatment Center (IFB) Adiposity Diseases, University of Leipzig, Leipzig, Germany. ⁹²Department of Medicine, University of Leipzig, Leipzig, Germany. ⁹³LifeLines, University Medical Center Groningen, University of Groningen, Groningen, the Netherlands. ⁹⁴Department of Pharmacology and Therapeutics, University College Cork, Cork, Ireland. ⁹⁵Department of Statistics and Biostatistics, Rutgers University, Piscataway, New Jersey, USA. ⁹⁶Department of Genetics, Rutgers University, Piscataway, New Jersey, USA. ⁹⁷Department of Biology, Norwegian University of Science and Technology, Trondheim, Norway. ⁹⁸Genome Technology Branch, National Human Genome Research Institute, US National Institutes of Health, Bethesda, Maryland, USA. ⁹⁹Clinical Trial Service Unit, Epidemiological Studies Unit, Nuffield Department of Population Health, University of Oxford, Oxford, UK. ¹⁰⁰Division of

Endocrinology, Diabetes and Metabolism, Ulm University Medical Centre, Ulm, Germany. ¹⁰¹Department of Radiology, University Medical Center Utrecht, Utrecht, the Netherlands. ¹⁰²Department of Biomedical Informatics, Vanderbilt University, Nashville, Tennessee, USA. ¹⁰³Department of Human Nutrition, Wageningen University, Wageningen, the Netherlands. ¹⁰⁴Department of Dietetics–Nutrition, Harokopio University, Athens, Greece. ¹⁰⁵Medical Research Institute, University of Dundee, Ninewells Hospital and Medical School, Dundee, UK. ¹⁰⁶Department of Internal Medicine B, University Medicine Greifswald, Greifswald, Germany. ¹⁰⁷Institute for Molecular Medicine, University of Helsinki, Helsinki, Finland. ¹⁰⁸Laboratory of Epidemiology and Population Sciences, National Institute on Aging, US National Institutes of Health, Bethesda, Maryland, USA. ¹⁰⁹Department of Public Health and Caring Sciences, Geriatrics, Uppsala University, Uppsala, Sweden. ¹¹⁰Kaiser Permanente, Division of Research, Oakland, California, USA. ¹¹¹Research Unit of Molecular Epidemiology, Helmholtz Zentrum München–German Research Center for Environmental Health, Neuherberg, Germany. ¹¹²German Center for Diabetes Research (DZD), Neuherberg, Germany. ¹¹³Department of Medicine III, University Hospital Carl Gustav Carus, Technische Universität Dresden, Dresden, Germany. ¹¹⁴Unit of Nutritional Research, Department of Public Health and Clinical Medicine, Umeå University, Umeå, Sweden. ¹¹⁵Department of Psychiatry, University of Groningen, University Medical Center Groningen, Groningen, the Netherlands. ¹¹⁶Kuopio Research Institute of Exercise Medicine, Kuopio, Finland. ¹¹⁷MRC Human Genetics Unit, Institute of Genetics and Molecular Medicine, University of Edinburgh, Western General Hospital, Edinburgh, UK. ¹¹⁸National Heart, Lung, and Blood Institute, Framingham Heart Study, Framingham, Massachusetts, USA. ¹¹⁹Department of Neurology, Boston University School of Medicine, Boston, Massachusetts, USA. ¹²⁰Department of Medical Statistics and Bioinformatics, Leiden University Medical Center, Leiden, the Netherlands. ¹²¹Faculty of Psychology and Education, VU University Amsterdam, Amsterdam, the Netherlands. ¹²²Department of Epidemiology, University Medical Center Groningen, University of Groningen, Groningen, the Netherlands. ¹²³Department of Medicine, Stanford University School of Medicine, Stanford, California, USA. ¹²⁴Institute for Community Medicine, University Medicine Greifswald, Greifswald, Germany. ¹²⁵Department of Biomedicine, Division of Medical Genetics, University of Basel, Basel, Switzerland. ¹²⁶Department of Genomics, Life and Brain Center, University of Bonn, Bonn, Germany. ¹²⁷Institute of Human Genetics, University of Bonn, Bonn, Germany. ¹²⁸Department of Public Health and General Practice, Norwegian University of Science and Technology, Trondheim, Norway. ¹²⁹Hannover Unified Biobank, Hannover Medical School, Hannover, Germany. ¹³⁰Center for Medical Systems Biology, Leiden, the Netherlands. ¹³¹Department of Pulmonary Physiology and Sleep Medicine, Nedlands, Western Australia, Australia. ¹³²School of Medicine and Pharmacology, University of Western Australia, Crawley, Western Australia, Australia. ¹³³Department of Immunology, Genetics and Pathology, SciLifeLab, Rudbeck Laboratory, Uppsala University, Uppsala, Sweden. ¹³⁴Department of Haematology, University of Cambridge, Cambridge, UK. ¹³⁵NHS Blood and Transplant, Cambridge, UK. ¹³⁶Department of Medicine, University of Oulu, Oulu, Finland. ¹³⁷Department of Medicine, Northwestern University Feinberg School of Medicine, Chicago, Illinois, USA. ¹³⁸Unit of Periodontology, Department of Restorative Dentistry, Periodontology and Endodontology, University Medicine Greifswald, Greifswald, Germany. ¹³⁹Department of Internal Medicine I, Ulm University Medical Centre, Ulm, Germany. ¹⁴⁰Institute of Human Genetics, Helmholtz Zentrum München–German Research Center for Environmental Health, Neuherberg, Germany. ¹⁴¹Department of Medical Sciences, Cardiovascular Epidemiology, Uppsala University, Uppsala, Sweden. ¹⁴²Centre for Bone and Arthritis Research, Department of Internal Medicine and Clinical Nutrition, Institute of Medicine, Sahlgrenska Academy, University of Gothenburg, Gothenburg, Sweden. ¹⁴³Genetics of Obesity and Related Metabolic Traits Program, Icahn School of Medicine at Mount Sinai, New York, New York, USA. ¹⁴⁴Steno Diabetes Center A/S, Gentofte Denmark. ¹⁴⁵Service of Nephrology, Department of Medicine, Lausanne University Hospital (CHUV), Lausanne, Switzerland. ¹⁴⁶School of Social and Community Medicine, University of Bristol, Bristol, UK. ¹⁴⁷Tropical Metabolism Research Unit, Tropical Medicine Research Institute, University of the West Indies, Mona, Kingston, Jamaica. ¹⁴⁸Global Health Institute, Department of Life Sciences, Ecole Polytechnique Fédérale de Lausanne, Lausanne, Switzerland. ¹⁴⁹Institute of Microbiology, University Hospital and University of Lausanne, Lausanne, Switzerland. ¹⁵⁰Center for Observational Research, Amgen, Inc., Thousand Oaks, California, USA. ¹⁵¹Center for Evidence-Based Healthcare, University Hospital Carl Gustav Carus, Technische Universität Dresden, Dresden, Germany. ¹⁵²Department of Medicine I, University Hospital Großhadern, Ludwig Maximilians Universität, Munich, Germany. ¹⁵³Chair of Genetic Epidemiology, Institute of Medical Informatics, Biometry and Epidemiology, Ludwig Maximilians Universität, Neuherberg, Germany. ¹⁵⁴DZHK (Deutsches Forschungszentrum für Herz-Kreislaufkrankungen–German Research Centre for Cardiovascular Research), Munich Heart Alliance, Munich, Germany. ¹⁵⁵Department of Respiratory Medicine, Sir Charles Gairdner Hospital, Nedlands, Western Australia, Australia. ¹⁵⁶Department of Epidemiology and Biostatistics, Institute for Research in Extramural Medicine (EMGO) Institute for Health and Care Research, VU University Medical Center, Amsterdam, the Netherlands. ¹⁵⁷Department of Internal Medicine, Division of Endocrinology and Metabolism, Medical University of Graz, Graz, Austria. ¹⁵⁸Diabetology-Endocrinology-Nutrition, Public Hospital System of the City of Paris (AP-HP), Bichat Hospital, Paris, France. ¹⁵⁹INSERM U872, Centre de Recherche des Cordeliers, Paris, France. ¹⁶⁰Paris Diderot University, Paris, France. ¹⁶¹Clinical Institute of Medical and Chemical Laboratory Diagnostics, Medical University of Graz, Graz, Austria. ¹⁶²Department of Preventive Medicine, Keck School of Medicine, University of Southern California, Los Angeles, California, USA. ¹⁶³Deutsches Herzzentrum München, Technische Universität München, Munich, Germany. ¹⁶⁴National Cancer Institute, Bethesda, Maryland, USA. ¹⁶⁵Department of Sociology, University of Helsinki, Helsinki, Finland. ¹⁶⁶EMGO Institute for Health and Care Research, VU University, Amsterdam, the Netherlands. ¹⁶⁷Department of Psychiatry, Neuroscience Campus, VU University Amsterdam, Amsterdam, the Netherlands. ¹⁶⁸Icelandic Heart Association, Kopavogur, Iceland. ¹⁶⁹University of Iceland, Reykjavik, Iceland. ¹⁷⁰Molecular and Cellular Therapeutics, Royal College of Surgeons in Ireland, Dublin, Ireland. ¹⁷¹Institute of Cardiovascular and Medical Sciences, Faculty of Medicine, University of Glasgow, Glasgow, UK. ¹⁷²Department of Medical Sciences, Molecular Medicine, Uppsala University, Uppsala, Sweden. ¹⁷³Department of Public Health Sciences, Stritch School of Medicine, Loyola University of Chicago, Maywood, Illinois, USA. ¹⁷⁴deCODE Genetics, Amgen, Inc., Reykjavik, Iceland. ¹⁷⁵Department of Oncology, University of Cambridge, Cambridge, UK. ¹⁷⁶Section of Geriatrics, Department of Internal Medicine, Academic Medical Center, Amsterdam, the Netherlands. ¹⁷⁷Department of Child and Adolescent Psychiatry, Psychology, Erasmus University Medical Center, Rotterdam, the Netherlands. ¹⁷⁸Department for Health Evidence, Radboud University Medical Center, Nijmegen, the Netherlands. ¹⁷⁹Department of Genetics, Radboud University Medical Center, Nijmegen, the Netherlands. ¹⁸⁰Department of Clinical Chemistry, Ulm University Medical Centre, Ulm, Germany. ¹⁸¹DZHK (Deutsches Forschungszentrum für Herz-Kreislaufkrankungen–German Research Centre for Cardiovascular Research), partner site Hamburg-Lübeck-Kiel, Lübeck, Germany. ¹⁸²Institut für Integrative und Experimentelle Genomik, Universität zu Lübeck, Lübeck, Germany. ¹⁸³Department of Community Medicine, Faculty of Health Sciences, UiT The Arctic University of Tromsø, Tromsø, Norway. ¹⁸⁴MRC Unit for Lifelong Health and Ageing at University College London, London, UK. ¹⁸⁵Department of Epidemiology and Public Health, University of Strasbourg, Faculty of Medicine, Strasbourg, France. ¹⁸⁶Department of Internal Medicine, University Medical Center Groningen, University of Groningen, Groningen, the Netherlands. ¹⁸⁷Pathology and Laboratory Medicine, University of Western Australia, Perth, Western Australia, Australia. ¹⁸⁸Cedars-Sinai Diabetes and Obesity Research Institute, Los Angeles, California, USA. ¹⁸⁹Department of Prosthetic Dentistry, Gerostomatology and Dental Materials, University Medicine Greifswald, Greifswald, Germany. ¹⁹⁰Biological Psychology, VU University Amsterdam, Amsterdam, the Netherlands. ¹⁹¹Institute of Social and Preventive Medicine (IUMSP), Centre Hospitalier Universitaire Vaudois (CHUV) and University of Lausanne, Lausanne, Switzerland. ¹⁹²Ministry of Health, Victoria, Republic of Seychelles. ¹⁹³Laboratory Medicine, Hospital of Desio, Department of Health Sciences, University of Milano, Bicocca, Italy. ¹⁹⁴Clinical Pharmacology Unit, University of Cambridge, Addenbrooke's Hospital, Cambridge, UK. ¹⁹⁵Clinical Pharmacology, William Harvey Research Institute, Barts and The London School of Medicine and Dentistry, Queen Mary University of London, London, UK. ¹⁹⁶Barts and The London Genome Centre, William Harvey Research Institute, Barts and The London School of Medicine and Dentistry, Queen Mary University of London, London, UK. ¹⁹⁷Center for Human Genetics Research, Vanderbilt University Medical Center, Nashville, Tennessee, USA. ¹⁹⁸Department of Molecular Physiology and Biophysics, Vanderbilt University, Nashville, Tennessee, USA. ¹⁹⁹Department of Biostatistics, Boston University School of Public Health, Boston, Massachusetts, USA. ²⁰⁰Department of Public Health and Primary Care, University of Cambridge, Cambridge, UK. ²⁰¹Division of Cardiovascular Epidemiology, Institute of Environmental Medicine, Karolinska Institutet, Stockholm, Sweden. ²⁰²Julius Center for Health Sciences and Primary Care, University Medical Center Utrecht, Utrecht, the Netherlands. ²⁰³Clinic of Cardiology, West German Heart Centre, University Hospital Essen, Essen, Germany. ²⁰⁴Department of General Practice and Primary Health Care, University of Helsinki, Helsinki, Finland. ²⁰⁵Unit of General Practice, Helsinki University Central Hospital, Helsinki, Finland. ²⁰⁶Department of Internal Medicine, University of Pisa, Pisa, Italy. ²⁰⁷National Research Council (CNR) Institute of Clinical Physiology, University of Pisa, Pisa, Italy. ²⁰⁸Department of Cardiology, Toulouse University School of Medicine, Rangueil Hospital, Toulouse, France. ²⁰⁹Robertson Center for Biostatistics, University of Glasgow, Glasgow, UK. ²¹⁰NorthShore University HealthSystem, University of Chicago, Evanston, Illinois, USA. ²¹¹Service of Therapeutic Education for Diabetes, Obesity and Chronic Diseases, Geneva University Hospital, Geneva, Switzerland. ²¹²Department of Medicine, Pharmacology, Pathology, Microbiology and Immunology, Vanderbilt University School of Medicine, Nashville, Tennessee, USA. ²¹³Leeds MRC Medical Bioinformatics Centre, University of Leeds, Leeds, UK. ²¹⁴Institute of Biomedical and Clinical Science, University of Exeter, Exeter, UK. ²¹⁵Department of Psychiatry, Washington University School of Medicine, St. Louis, Missouri, USA. ²¹⁶Center for Biomedicine, European Academy Bozen, Bolzano (EURAC), Bolzano, Italy (affiliated institute of the University of Lübeck, Lübeck, Germany). ²¹⁷Division of Genomic Medicine, National Human Genome Research Institute, US National Institutes of Health, Bethesda, Maryland, USA. ²¹⁸Institute of Cardiovascular Science, University College London, London, UK. ²¹⁹Department of Vascular Medicine, Academic Medical Center, Amsterdam, the Netherlands. ²²⁰Centre for Cardiovascular Genetics, Institute of Cardiovascular Sciences, University

College London, London, UK. ²²¹Cardiovascular Genetics Division, Department of Internal Medicine, University of Utah, Salt Lake City, Utah, USA. ²²²School of Population Health, University of South Australia, Adelaide, South Australia, Australia. ²²³Sansom Institute for Health Research, University of South Australia, Adelaide, South Australia, Australia. ²²⁴South Australian Health and Medical Research Institute, Adelaide, South Australia, Australia. ²²⁵Centre for Paediatric Epidemiology and Biostatistics, University College London Institute of Child Health, London, UK. ²²⁶National Institute for Health and Welfare, Oulu, Finland. ²²⁷MRC Health Protection Agency (HPA) Centre for Environment and Health, School of Public Health, Imperial College London, London, UK. ²²⁸Unit of Primary Care, Oulu University Hospital, Oulu, Finland. ²²⁹Biocenter Oulu, University of Oulu, Oulu, Finland. ²³⁰Institute of Health Sciences, University of Oulu, Oulu, Finland. ²³¹Hjelt Institute Department of Public Health, University of Helsinki, Helsinki, Finland. ²³²Department of Forensic Molecular Biology, Erasmus Medical Center, Rotterdam, the Netherlands. ²³³UK Clinical Research Collaboration Centre of Excellence for Public Health (Northern Ireland), Queens University of Belfast, Belfast, UK. ²³⁴Faculty of Medicine, Institute of Health Sciences, University of Oulu, Oulu, Finland. ²³⁵Unit of General Practice, Oulu University Hospital, Oulu, Finland. ²³⁶Department of Urology, Radboud University Medical Centre, Nijmegen, the Netherlands. ²³⁷Imperial College Healthcare NHS Trust, London, UK. ²³⁸National Heart and Lung Institute, Imperial College London, London, UK. ²³⁹Department of Epidemiology and Public Health, University College London, London, UK. ²⁴⁰Department of Medicine, Kuopio University Hospital and University of Eastern Finland, Kuopio, Finland. ²⁴¹Department of Physiology, Institute of Biomedicine, University of Eastern Finland, Kuopio Campus, Kuopio, Finland. ²⁴²Department of Clinical Physiology and Nuclear Medicine, Kuopio University Hospital and University of Eastern Finland, Kuopio, Finland. ²⁴³Department of Clinical Chemistry, Fimlab Laboratories and School of Medicine, University of Tampere, Tampere, Finland. ²⁴⁴Department of Health Sciences, University of Milano, Milan, Italy. ²⁴⁵Fondazione Filarete, Milan, Italy. ²⁴⁶Division of Nephrology and Dialysis, San Raffaele Scientific Institute, Milan, Italy. ²⁴⁷Università Vita-Salute San Raffaele, Milan, Italy. ²⁴⁸Institut Universitaire de Cardiologie et de Pneumologie de Québec, Faculty of Medicine, Laval University, Quebec City, Quebec, Canada. ²⁴⁹Institute of Nutrition and Functional Foods, Laval University, Quebec City, Quebec, Canada. ²⁵⁰Department of Biostatistics, University of Washington, Seattle, Washington, USA. ²⁵¹Department of Surgery, University Medical Center Utrecht, Utrecht, the Netherlands. ²⁵²Department of Biostatistics, University of Liverpool, Liverpool, UK. ²⁵³Department of Pediatrics, University of Iowa, Iowa City, Iowa, USA. ²⁵⁴Institute of Epidemiology II, Helmholtz Zentrum München–German Research Center for Environmental Health, Neuherberg, Germany. ²⁵⁵Department of Neurology, General Central Hospital, Bolzano, Italy. ²⁵⁶Department of Clinical Physiology and Nuclear Medicine, Turku University Hospital, Turku, Finland. ²⁵⁷Research Centre of Applied and Preventive Cardiovascular Medicine, University of Turku, Turku, Finland. ²⁵⁸Human Genomics Laboratory, Pennington Biomedical Research Center, Baton Rouge, Louisiana, USA. ²⁵⁹Center for Systems Genomics, Pennsylvania State University, University Park, Pennsylvania, USA. ²⁶⁰Croatian Centre for Global Health, Faculty of Medicine, University of Split, Split, Croatia. ²⁶¹Department of Cardiovascular Sciences, University of Leicester, Glenfield Hospital, Leicester, UK. ²⁶²National Institute for Health Research (NIHR) Leicester Cardiovascular Biomedical Research Unit, Glenfield Hospital, Leicester, UK. ²⁶³South Carelia Central Hospital, Lappeenranta, Finland. ²⁶⁴Paul Langerhans Institute Dresden, German Center for Diabetes Research (DZD), Dresden, Germany. ²⁶⁵International Centre for Circulatory Health, Imperial College London, London, UK. ²⁶⁶Program for Personalized and Genomic Medicine, University of Maryland School of Medicine, Baltimore, Maryland, USA. ²⁶⁷Division of Endocrinology, Diabetes and Nutrition, University of Maryland School of Medicine, Baltimore, Maryland, USA. ²⁶⁸Geriatric Research and Education Clinical Center, Veterans Administration Medical Center, Baltimore, Maryland, USA. ²⁶⁹Helsinki University Central Hospital Heart and Lung Center, Department of Medicine, Helsinki University Central Hospital, Helsinki, Finland. ²⁷⁰Montreal Heart Institute, Université de Montréal, Montreal, Quebec, Canada. ²⁷¹Department of Kinesiology, Laval University, Quebec City, Quebec, Canada. ²⁷²Dipartimento di Scienze Farmacologiche e Biomolecolari, Università di Milano and Centro Cardiologico Monzino, Istituto di Ricovero e Cura a Carattere Scientifico (IRCCS), Milan, Italy. ²⁷³Department of Food Science and Nutrition, Laval University, Quebec City, Quebec, Canada. ²⁷⁴A full list of members and affiliations appears in the **Supplementary Note**. ²⁷⁵Institut Pasteur de Lille, INSERM U744, Université de Lille 2, Lille, France. ²⁷⁶Department of Cardiology, Division of Heart and Lungs, University Medical Center Utrecht, Utrecht, the Netherlands. ²⁷⁷Durrer Center for Cardiogenetic Research, Interuniversity Cardiology Institute Netherlands–Netherlands Heart Institute, Utrecht, the Netherlands. ²⁷⁸Lee Kong Chian School of Medicine, Imperial College London and Nanyang Technological University, Singapore. ²⁷⁹Health Science Center at Houston, University of Texas, Houston, Texas, USA. ²⁸⁰Division of Genetics, Department of Medicine, Brigham and Women’s Hospital, Harvard Medical School, Boston, Massachusetts, USA. ²⁸¹Department of Epidemiology, University Medical Center Utrecht, Utrecht, the Netherlands. ²⁸²Lund University Diabetes Centre, Lund University, Malmö, Sweden. ²⁸³Diabetes and Endocrinology Unit, Department of Clinical Science, Lund University, Malmö, Sweden. ²⁸⁴Department of Epidemiology, Harvard School of Public Health, Harvard University, Boston, Massachusetts, USA. ²⁸⁵Interuniversity Cardiology Institute of the Netherlands (ICIN), Utrecht, the Netherlands. ²⁸⁶Department of Epidemiology and Population Health, Albert Einstein College of Medicine, Belfer, New York, USA. ²⁸⁷Center for Human Genetics, Division of Public Health Sciences, Wake Forest School of Medicine, Winston-Salem, North Carolina, USA. ²⁸⁸Synlab Academy, Synlab Services, Mannheim, Germany. ²⁸⁹Department of Clinical Genetics, Erasmus University Medical Center, Rotterdam, the Netherlands. ²⁹⁰Harvard Medical School, Boston, Massachusetts, USA. ²⁹¹Institute for Translational Genomics and Population Sciences, Los Angeles BioMedical Research Institute at Harbor–University of California, Los Angeles Medical Center, Torrance, California, USA. ²⁹²Finnish Diabetes Association, Tampere, Finland. ²⁹³Pirkanmaa Hospital District, Tampere, Finland. ²⁹⁴Center for Non-Communicable Diseases, Karachi, Pakistan. ²⁹⁵Department of Medicine, University of Pennsylvania, Philadelphia, Pennsylvania, USA. ²⁹⁶Laboratory of Genetics, National Institute on Aging, Baltimore, Maryland, USA. ²⁹⁷Instituto de Investigación Sanitaria del Hospital Universitario La Paz (IdiPAZ), Madrid, Spain. ²⁹⁸Diabetes Research Group, King Abdulaziz University, Jeddah, Saudi Arabia. ²⁹⁹Centre for Vascular Prevention, Danube University Krems, Krems, Austria. ³⁰⁰Department of Public Health and Clinical Nutrition, University of Eastern Finland, Kuopio, Finland. ³⁰¹Research Unit, Kuopio University Hospital, Kuopio, Finland. ³⁰²Institute of Cellular Medicine, Newcastle University, Newcastle, UK. ³⁰³Chair of Epidemiology, Institute of Medical Informatics, Biometry and Epidemiology, Ludwig Maximilians Universität, Munich, Germany. ³⁰⁴Klinikum Großhadern, Munich, Germany. ³⁰⁵Institute of Epidemiology I, Helmholtz Zentrum München–German Research Center for Environmental Health, Neuherberg, Germany. ³⁰⁶Department of Pulmonology, University Medical Center Utrecht, Utrecht, the Netherlands. ³⁰⁷King Abdulaziz University, Jeddah, Saudi Arabia. ³⁰⁸Department of Internal Medicine, Division of Gastroenterology, University of Michigan, Ann Arbor, Michigan, USA. ³⁰⁹Department of Computational Medicine and Bioinformatics, University of Michigan, Ann Arbor, Michigan, USA. ³¹⁰Faculty of Medicine, University of Iceland, Reykjavik, Iceland. ³¹¹University of Cambridge Metabolic Research Laboratories, Institute of Metabolic Science, Addenbrooke’s Hospital, Cambridge, UK. ³¹²NIHR Cambridge Biomedical Research Centre, Institute of Metabolic Science, Addenbrooke’s Hospital, Cambridge, UK. ³¹³Carolina Center for Genome Sciences, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina, USA. ³¹⁴Division of Population Health Sciences and Education, St George’s, University of London, London, UK. ³¹⁵Service of Medical Genetics, CHUV University Hospital, Lausanne, Switzerland. ³¹⁶Oxford NIHR Biomedical Research Centre, Oxford University Hospitals NHS Trust, Oxford, UK. ³¹⁷Department of Internal Medicine, Division of Cardiovascular Medicine, University of Michigan, Ann Arbor, Michigan, USA. ³¹⁸Department of Human Genetics, University of Michigan, Ann Arbor, Michigan, USA. ³¹⁹Department of Biostatistics, Harvard School of Public Health, Boston, Massachusetts, USA. ³²⁰Mindich Child Health and Development Institute, Icahn School of Medicine at Mount Sinai, New York, New York, USA. ³²¹Biosciences Research Division, Department of Primary Industries, Melbourne, Victoria, Australia. ³²²Department of Food and Agricultural Systems, University of Melbourne, Melbourne, Victoria, Australia. ³²³These authors contributed equally to this work. ³²⁴These authors jointly directed this work. Correspondence should be addressed to P.M.V. (peter.visscher@uq.edu.au), J.N.H. (joelh@broadinstitute.org) or T.M.F. (t.m.frayling@exeter.ac.uk).

ONLINE METHODS

A summary of the methods, together with a full description of genome-wide association analyses and follow-up analyses, can be found here and in the **Supplementary Note**. Written informed consent was obtained from every participant in each study, and the study was approved by relevant ethics committees.

Genome-wide association study meta-analysis. We combined the height summary association statistics from 79 GWAS in a meta-analysis of 253,288 individuals using the same methods and studies as previously described⁶ and additional studies as described in **Supplementary Tables 17–19**. Meta-analysis was performed on a total of 2,550,858 autosomal SNPs using the inverse variance fixed-effects method with METAL²².

GCTA-COJO: conditional and joint multiple-SNP analysis. We used GCTA-COJO analysis^{7,8} to select the top associated SNPs. This method uses the summary statistics from the meta-analysis and LD correlations between SNPs estimated from a reference sample to perform a conditional association analysis⁷. The method starts with an initial model of the SNP that shows the strongest evidence of association across the whole genome. It then implements the association analysis conditioning on the selected SNP(s) to search for the top SNPs iteratively one by one via a stepwise model selection procedure until no SNP has a conditional *P* value that passes the significance level. Finally, all the selected SNPs are fitted jointly in the model for effect size estimation. We used 6,654 unrelated individuals from the ARIC cohort as the reference sample for LD estimation. There were ~3.0 million SNPs included in the original meta-analysis. We included in this analysis only the SNPs (~2.48 million) in HapMap 2 for which we had a sample size of >50,000. We used the genome-wide significance level $P < 5 \times 10^{-8}$ (as reported in **Supplementary Table 1**).

Metachip replication. We combined the height summary association statistics from 37 independent studies genotyped using the Illumina Metachip array⁹ in a meta-analysis of 80,067 individuals of European ancestry (**Supplementary Tables 20–22**). Each study tested the association for each genotyped SNP using the same quality control procedures, height transformation, adjustment and inheritance model as described for the GWAS analysis. Genomic control correction was applied to the results for each study before meta-analysis, using a set of 4,427 SNPs associated with QT interval to control study-specific inflation factors. We used the inverse variance fixed-effects meta-analysis method.

Validation: linear mixed model-based association analysis. Each of 15 studies (59,380 individuals) used genome-wide SNP information to calculate a GRM for all pairs of individuals and used this matrix to correct association statistics for cryptic relatedness and population stratification. Each study used an LMM as implemented in the software EMMAX²³. Meta-analysis was performed as described for the standard GWAS using a single- λ_{GC} correction. Each study additionally repeated the analyses for each chromosome using a GRM generated from the remaining 21 chromosomes or, in the case of the largest study (WGHS), repeated the analysis for all odd-numbered chromosomes using a GRM generated from the even-numbered chromosomes and vice versa. Each study then combined the association results from the 22 or 2 parts of the genome into a single set of data, and the single- λ_{GC} meta-analysis was repeated.

Validation: within-family (transmission) association analyses. A pure transmission-based analysis was performed in 7 cohorts for SNPs representing 416 signals of association (**Supplementary Note**), selected after repeating meta-analysis excluding these studies with single- λ_{GC} correction. The filtering out of SNPs with low imputation quality in the studies was followed by applying the inverse variance method of meta-analysis to the family-based results. Because of the presence of related individuals, family-based studies have lower power at a given sample size. For each study, we calculated the effective sample size (the size of a sample of unrelated individuals that would have equivalent power; see the **Supplementary Note** and Winkler *et al.*²⁴). Estimation of winner's curse in our data set was performed by repeating the meta-analysis excluding either the family-based studies or random sets of studies from GIANT matched by effective sample size to the family-based studies. Independent genome-wide significant loci were selected from each meta-analysis. Power for replication

in the excluded samples was estimated at different *P*-value thresholds, and the deficit in replication (number of replications expected minus the number observed replications) was calculated. The contribution of winner's curse to the deficit in replication was estimated as the average deficit across the three sets of random non-family-based cohorts. By subtracting this deficit from the deficit observed for the family-based cohorts, we estimated the lack of replication that could be attributed to stratification (either inflation of effect size for true associations or false positive associations).

Variance and heritability explained. We used GCTA-COJO analysis to select the top associated SNPs at a range of stringent significance levels (5×10^{-3} , 5×10^{-4} , 5×10^{-5} , ..., 5×10^{-8}) for estimation and prediction analyses. We then quantified the variance explained by the selected SNPs using a three-stage analysis—within-family prediction, GCTA-GREML analysis and population-based prediction—in five validation studies (B-PROOF, FRAM, QIMR, TwinGene and WTCCC-T2D). To avoid sample overlap, we repeated the main GWAS meta-analysis and the multiple-SNP analysis five times, each time excluding one of the five validation studies. This approach ensured complete independence between the data used to discover SNPs and the data used to estimate how much variance in height these SNPs explained and how well they predicted height. For the within-family prediction analyses, we selected 1,622, 2,758 and 1,597 pairs of full siblings from the QIMR, TwinGene and FRAM cohorts, respectively, with 1 sibling pair per family. For the whole-genome estimation and prediction analyses, we used GCTA-GRM⁸ to estimate the genetic relatedness between individuals and selected unrelated individuals with pairwise genetic relatedness of <0.025 in each of the 5 studies: B-PROOF ($n = 2,555$), FRAM ($n = 1,145$), QIMR ($n = 3,627$), TwinGene ($n = 5,668$) and WTCCC-T2D ($n = 1,914$).

Within-family prediction analysis. We used the SNPs selected from GCTA-COJO analysis to create a genetic predictor (also called a 'genetic profile score') for each of the full-sibling pairs using PLINK²⁵. We then adjusted the genetic predictor by the first 20 principal components generated from principal-component analysis (PCA)²⁶. By comparing the predictors within and between families, we partitioned the variance in the predictor analysis into components due to real SNP effects (V_g), errors in estimating SNP effect (V_e) and population structure ($C_g + C_e$).

We calculated the weighted average of each of the four (co)variance components over the three cohorts by their sample size:

$$\sum_i V_{g(i)} n_i / \sum_i (n_i)$$

with the subscript *i* indicating the cohort and *n* indicating the sample size. From the results of these partitioning analyses within families, we can infer what the prediction R^2 value (equation (19)) and what the proportion of the variance explained by SNPs (V_g/V_p , with V_p being the phenotypic variance) would be in a sample of unrelated individuals when using the same set of SNPs. We then tested these inferred values in unrelated samples.

GCTA-GREML analysis. We performed GREML analysis⁴ in GCTA⁸ to estimate the variance explained by the selected SNPs (h_g^2) in each of the five validation studies. This method fits the effects of a set of SNPs simultaneously in a model as random effects and estimates the genetic variance captured by all the fitted SNPs without testing the significance of the association of any single SNPs. We combined the estimates of h_g^2 from the five studies by the inverse variance approach:

$$\sum_i (h_{g(i)}^2 / SE_{g(i)}^2) / \sum_i (1 / SE_{g(i)}^2)$$

Population-based prediction analysis. We created a genetic predictor using the selected SNPs for the unrelated individuals in each of the five validation studies. We then calculated the squared correlation (R^2) between the phenotype and predictor in each validation study and calculated the average prediction R^2 value weighted by the sample size across the five studies:

$$\sum_i (R_i^2 n_i) / \sum_i (n_i)$$

Theory and method to partition the variance in a genetic predictor. Under the assumption of an additive genetic model, the phenotype of a quantitative trait can be written as:

$$y = g + \varepsilon = \sum_i x_i b_i + \varepsilon \quad (1)$$

where y is the trait phenotype, g is the total genetic effect of all SNPs, x is an indicator variable for SNP genotypes, b is the SNP effect and ε is the residual.

From this model, the additive genetic variance is:

$$\text{var}(g) = \sum_i \text{var}(x_i) b_i^2 + \sum_i \sum_{j(i \neq j)} \text{cov}(x_i, x_j) b_i b_j \quad (2)$$

with the first component being the expected value of additive genetic variance under linkage equilibrium (LE) and the second component being deviation from the expected value that could be caused by LD, population structure or selection²⁷.

Considering a pair of full siblings in a family, the additive genetic covariance between the siblings is:

$$\begin{aligned} \text{cov}(g_1, g_2) &= \text{cov}\left(\sum_i x_{1i} b_i, \sum_i x_{2i} b_i\right) \\ &= \sum_i \text{cov}(x_{1i}, x_{2i}) b_i^2 + \sum_i \sum_{j(i \neq j)} \text{cov}(x_{1i}, x_{2j}) b_i b_j \end{aligned} \quad (3)$$

For full siblings:

$$\text{cov}(x_{1i}, x_{2i}) = 1/2 \text{var}(x_i)$$

$$\text{cov}(x_{1i}, x_{2j}) = 1/2 \text{cov}(x_i, x_j)$$

for SNPs that are in LD and

$$\sum_i \sum_{j(i \neq j)} \text{cov}(x_{1i}, x_{2j}) b_i b_j = \sum_i \sum_{j(i \neq j)} \text{cov}(x_i, x_j) b_i b_j$$

for SNPs that are not in LD (as shown by both empirical and simulation results).

Let:

$$V_g = \sum_i \text{var}(x_i) b_i^2 + \sum_i \sum_{j(i \neq j)} \text{cov}(x_i, x_j) b_i b_j \quad | \text{ SNPs are in LD}$$

$$C_g = \sum_i \sum_{j(i \neq j)} \text{cov}(x_i, x_j) b_i b_j \quad | \text{ SNPs are not in LD but are correlated owing}$$

to population structure

Then, the genetic variance is:

$$\text{var}(g) = V_g + C_g \quad (4)$$

The genetic covariance between a pair of full siblings is:

$$\text{cov}(g_1, g_2) = 1/2 V_g + C_g \quad (5)$$

If we take a set of SNPs with their effects estimated from GCTA-COJO analysis and create a predictor using these SNPs in an independent validation sample, we can write the predictor as:

$$\hat{g} = \sum_i x_i \hat{b}_i \quad (6)$$

where \hat{b} is the estimate of b , with $\hat{b} = b + e$ where e is the error in estimating b .

If we assume b and e to be independent and denote

$$V_e = \sum_i \text{var}(x_i) e_i^2$$

and

$$C_e = \sum_i \sum_{j(i \neq j)} \text{cov}(x_i, x_j) e_i e_j$$

the variance of the predictor is:

$$\begin{aligned} \text{var}(g) &= \text{var}\left(\sum_i x_i \hat{b}_i\right) = \sum_i \text{var}(x_i) \hat{b}_i^2 + \sum_i \sum_{j(i \neq j)} \text{cov}(x_i, x_j) \hat{b}_i \hat{b}_j \\ &= \sum_i \text{var}(x_i) b_i^2 + \sum_i \text{var}(x_i) e_i^2 + \sum_i \sum_{j(i \neq j)} \text{cov}(x_i, x_j) b_i b_j + \sum_i \sum_{j(i \neq j)} \text{cov}(x_i, x_j) e_i e_j \\ &= V_g + V_e + C_g + C_e \end{aligned} \quad (7)$$

The covariance between the predictors of a pair of full siblings is:

$$\begin{aligned} \text{cov}(\hat{g}_1, \hat{g}_2) &= \text{cov}\left(\sum_i x_{1i} \hat{b}_i, \sum_i x_{2i} \hat{b}_i\right) = 1/2 \sum_i \text{var}(x_i) \hat{b}_i^2 + \sum_i \sum_{j(i \neq j)} \text{cov}(x_{1i}, x_{2j}) \hat{b}_i \hat{b}_j \\ &= 1/2 V_g + 1/2 V_e + C_g + C_e \end{aligned} \quad (8)$$

The covariance between the true phenotype and the predictor for the same individual is:

$$\text{cov}(y, \hat{g}) = \text{cov}(g + \varepsilon, g + e) = \text{var}(g) = V_g + C_g \quad (9)$$

The covariance between the true phenotype of one sibling and the predictor of the other sibling is:

$$\begin{aligned} \text{cov}(y_1, \hat{g}_2) &= \text{cov}(g_1 + \varepsilon_1, g_2 + e_2) \\ &= 1/2 \sum_i \text{var}(x_i) b_i^2 + \sum_i \sum_{j(i \neq j)} \text{cov}(x_i, x_j) b_i b_j = 1/2 V_g + C_g \end{aligned} \quad (10)$$

If we define $\Delta \hat{g} = \hat{g}_1 - \hat{g}_2$ and $\Delta y = y_1 - y_2$,

$$\text{var}(\Delta \hat{g}) = \text{var}(\hat{g}_1) + \text{var}(\hat{g}_2) - 2 \text{cov}(\hat{g}_1, \hat{g}_2) = V_g + V_e \quad (11)$$

$$\text{cov}(\Delta y, \Delta \hat{g}) = \text{cov}(y_1, \hat{g}_1) + \text{cov}(y_2, \hat{g}_2) - \text{cov}(y_1, \hat{g}_2) - \text{cov}(y_2, \hat{g}_1) = V_g \quad (12)$$

We then can calculate the four parameters as:

$$V_g = \text{cov}(\Delta y, \Delta \hat{g}) \quad (13)$$

$$V_e = \text{var}(\Delta \hat{g}) - V_g \quad (14)$$

$$C_g = \text{cov}(y, \hat{g}) - V_g \quad (15)$$

$$C_e = 2 \text{cov}(\hat{g}_1, \hat{g}_2) - \text{var}(\hat{g}) - C_g \quad (16)$$

where V_g can be interpreted as the variance explained by real SNP effects, C_g is the covariance between predictors attributed to the real effects of SNPs that are not in LD but are correlated owing to population stratification, V_e is the accumulated variance due to errors in estimating SNP effects and C_e is the covariance between predictors attributed to errors in estimating the effects of SNPs that are correlated owing to population stratification.

To assess prediction accuracy, we usually perform a regression analysis of the real phenotype against the predictor:

$$y = \beta_0 + \hat{g} \beta_1 + \varepsilon \quad (17)$$

so that the regression slope is actually

$$\beta = \text{cov}(y, \hat{g}) / \text{var}(\hat{g}) = (V_g + C_g) / (V_g + V_e + C_g + C_e) \quad (18)$$

with the regression R^2 being

$$R^2 = (V_g + C_g)^2 / (V_g + V_e + C_g + C_e) \quad (19)$$

In the absence of population structure:

$$R^2 = V_g^2 / (V_g + V_e) \quad (20)$$

Variance explained by SNPs in proximity to the top associated SNPs. We performed analyses to quantify the variance explained by SNPs in close physical proximity to the top associated SNPs in 9,500 unrelated individuals (pairwise genetic relatedness < 0.025) from a combined data set of the QIMR and TwinGene cohorts. As in previous analyses, to avoid sample overlap between the discovery and validation studies, we repeated the discovery meta-analysis excluding the QIMR and TwinGene cohorts and identified 643 genome-wide significant SNPs from the GCTA-COJO analysis of the summary statistics using ARIC data for LD estimation. We used GCTA-GREML analysis^{4,8} to quantify the phenotypic variance explained by all the common SNPs (minor allele frequency (MAF) > 0.01) within 100 kb, 500 kb or 1 Mb of the 643 genome-wide significant SNPs. There are 104,000, 423,000 and 745,000 SNPs within 100 kb, 500 kb and 1 Mb of the top associated SNPs, respectively, which explain 20.8% (SE = 1.3%), 25.7% (SE = 1.8%) and 29.5% (SE = 2.2%) of phenotypic variance (Supplementary Fig. 6a). We then applied a regression-based approach²⁸ to adjust for LD between SNPs. The estimates of variance explained after LD adjustment were slightly higher than those without adjustment, and the ratio of the estimates with and without LD adjustment was consistently ~ 1.05 , regardless of the window size (Supplementary Fig. 6a). However, this difference is small.

We then sought to investigate whether there was an enrichment of additional associated signals at the top associated loci. We varied the window size, using windows of 20 kb, 50 kb, 100 kb, 150 kb, 200 kb, 300 kb, 400 kb, 500 kb, 750 kb and 1 Mb, and fitted a two-component model in GCTA-GREML analysis, with the first component being the top associated SNPs and the second component being the rest of the SNPs in the window. We found that the per-SNP variance explained excluding the top SNPs (variance explained by the second component divided by the number of SNPs included in this component) decreased with the size of the window (Supplementary Fig. 6b), implying that SNPs in closer physical proximity to the top associated SNPs tend to explain disproportionately more variance.

Enrichment of associated SNPs in ENCODE regions, loci containing OMIM genes, eQTLs and nonsynonymous SNPs. To identify putative causal variants among the height-associated markers, we explored whether the height-associated SNPs were in strong LD ($r^2 > 0.8$) with nonsynonymous coding variants in 1000 Genomes Project CEU Phase 1 data (Utah residents of Northern and Western European ancestry), showed an effect on whole-blood gene expression levels, were located within Encyclopedia of DNA Elements (ENCODE)-annotated regions, were within loci harboring monogenic growth genes or had previously been associated with other complex traits in the NHGRI GWAS catalog ($P < 5 \times 10^{-8}$) (Supplementary Tables 7–11). To estimate the empirical assessment of enrichment for listed features, we used 10,000 permutations of random sets of SNPs matched to the LD-pruned ($r^2 > 0.1$) 628 height-associated SNPs by the number of nearby genes (within an LD distance of $r^2 > 0.5$), the physical distance to the nearest gene and MAF.

Enrichment of genes in associated loci in known and new pathways. *Data-Driven Expression-Prioritized Integration for Complex Traits (DEPICT) analysis.* The DEPICT method (T.H.P., J. Karjalainen, Y. Chan, H. Westra and A.R.W. *et al.*, unpublished data; see Geller *et al.*²⁹ for an earlier application of DEPICT)

relies on precomputed predictions of gene function based on a heterogeneous panel of 77,840 expression arrays (Fehrmann *et al.*, unpublished data; ref. 30), 5,984 molecular pathways (based on 169,810 high-confidence experimentally derived protein-protein interactions³¹), 2,473 phenotypic gene sets (based on 211,882 gene-phenotype pairs from the Mouse Genome Informatics; see URLs), 737 REACTOME pathways³², 5,083 Gene Ontology terms¹⁴ and 184 Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways³³. The method leverages these predictions to extend the functional annotations of genes, including genes that previously had only a few or no functional annotations. DEPICT facilitates the analysis of GWAS data by (i) assessing whether genes in associated loci are enriched in tissue-specific expression, (ii) identifying reconstituted gene sets that are enriched in genes from associated loci and (iii) systematically identifying the most likely causal gene(s) at a given locus (see the **Supplementary Note** for a more detailed description of DEPICT). To run DEPICT, we first clumped the summary statistics from the meta-analysis using 500-kb flanking regions with $r^2 > 0.1$ and excluded SNPs with $P \geq 5 \times 10^{-8}$, which yielded 628 SNPs. We then mapped genes to each of the 628 most strongly associated SNPs. For a given SNP, this mapping was accomplished by including all genes that resided within the boundaries of $r^2 > 0.5$ of that SNP and always including the nearest gene to its locus gene set. We used a locus definition that was calibrated using the GWAS data for height levels presented in this report and optimized the capture of known monogenic genes for those traits. We merged overlapping loci and excluded loci that mapped near or within the major histocompatibility complex locus (chromosome 6, 20–40 Mb), which resulted in a list of 566 non-overlapping loci that were used as input for DEPICT. HapMap Project Phase 2 CEU genotype data were used for all LD calculations.

GRAIL and MAGENTA analyses. The GRAIL¹⁴ algorithm was run using the LD-pruned ($r^2 > 0.1$) 628 SNPs, without correcting for gene size and using text-mining data available up to December 2006 (default setting). MAGENTA¹³ was run with the adjusted summary statistics with single λ_{GC} as input using default settings and excluding the human leukocyte antigen (HLA) region.

22. Willer, C.J., Li, Y. & Abecasis, G.R. METAL: fast and efficient meta-analysis of genomewide association scans. *Bioinformatics* **26**, 2190–2191 (2010).
23. Kang, H.M. *et al.* Variance component model to account for sample structure in genome-wide association studies. *Nat. Genet.* **42**, 348–354 (2010).
24. Winkler, T.W. *et al.* Quality control and conduct of genome-wide association meta-analyses. *Nat. Protoc.* **9**, 1192–1212 (2014).
25. Purcell, S. *et al.* PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.* **81**, 559–575 (2007).
26. Price, A.L. *et al.* Principal components analysis corrects for stratification in genome-wide association studies. *Nat. Genet.* **38**, 904–909 (2006).
27. Lynch, M. & Walsh, B. *Genetics and Analysis of Quantitative Traits* (Sinauer Associates, Inc., 1998).
28. Gusev, A. *et al.* Quantifying missing heritability at known GWAS loci. *PLoS Genet.* **9**, e1003993 (2013).
29. Geller, F. *et al.* Genome-wide association analyses identify variants in developmental genes associated with hypospadias. *Nat. Genet.* **46**, 957–963 (2014).
30. Cvejic, A. *et al.* *SMIM1* underlies the Vel blood group and influences red blood cell traits. *Nat. Genet.* **45**, 542–545 (2013).
31. Lage, K. *et al.* A human phenome-interactome network of protein complexes implicated in genetic disorders. *Nat. Biotechnol.* **25**, 309–316 (2007).
32. Croft, D. *et al.* Reactome: a database of reactions, pathways and biological processes. *Nucleic Acids Res.* **39**, D691–D697 (2011).
33. Kanehisa, M., Goto, S., Sato, Y., Furumichi, M. & Tanabe, M. KEGG for integration and interpretation of large-scale molecular data sets. *Nucleic Acids Res.* **40**, D109–D114 (2012).