ARTICLE

doi:10.1038/nature21039

Rare and low-frequency coding variants alter human adult height

A full list of authors and affiliations appears in the online version of the paper.

Height is a highly heritable, classic polygenic trait with approximately 700 common associated variants identified through genome-wide association studies so far. Here, we report 83 height-associated coding variants with lower minor-allele frequencies (in the range of 0.1–4.8%) and effects of up to 2 centimetres per allele (such as those in *IHH*, *STC2*, *AR* and *CRISPLD2*), greater than ten times the average effect of common variants. In functional follow-up studies, rare height-increasing alleles of *STC2* (giving an increase of 1–2 centimetres per allele) compromised proteolytic inhibition of PAPP-A and increased cleavage of IGFBP-4 *in vitro*, resulting in higher bioavailability of insulin-like growth factors. These 83 height-associated variants overlap genes that are mutated in monogenic growth disorders and highlight new biological candidates (such as *ADAMTS3*, *IL11RA* and *NOX4*) and pathways (such as proteoglycan and glycosaminoglycan synthesis) involved in growth. Our results demonstrate that sufficiently large sample sizes can uncover rare and low-frequency variants of moderate-to-large effect associated with polygenic human phenotypes, and that these variants implicate relevant genes and pathways.

Human height is a highly heritable, polygenic trait[1,2]. The contribution of common DNA sequence variation to inter-individual differences in adult height has been systematically evaluated through genome-wide association studies (GWAS). This approach has thus far identified 697 independent variants located within 423 loci that together explain around 20% of the heritability of height[3]. As is typical of complex traits and diseases, most of the alleles that affect height that have been discovered so far are common (with a minor allele frequency (MAF) > 5%) and are mainly located outside coding regions, complicating the identification of the relevant genes or functional variants. Identifying coding variants associated with a complex trait in new or known loci has the potential to help pinpoint causal genes. Furthermore, the extent to which rare (MAF < 1%) and low-frequency (1% < MAF ≤ 5%) coding variants also influence complex traits and diseases remains an open question. Many recent DNA sequencing studies have identified only a few of these variants[4–8], but this limited success could be due to their modest sample size[9]. Some studies have suggested that common sequence variants may explain the majority of the heritable variation in adult height[10]. It is therefore timely to assess whether and to what extent rare and low-frequency coding variations contribute to the genetic landscape of this model polygenic trait.

In this study, we used an ExomeChip[11] to test the association between 241,453 variants (of which 83% are coding variants with a MAF ≤ 5%) and adult height variation in 711,428 individuals (discovery and validation sample sizes were 458,927 and 252,501, respectively). The ExomeChip is a genotyping array designed to query in very large sample sizes coding variants identified by whole-exome DNA sequencing of approximately 12,000 participants. The main goals of our project were to determine whether rare and low-frequency coding variants influence the architecture of a model complex human trait (in this case, adult height) and to discover and characterize new genes and biological pathways implicated in human growth.

Coding variants associated with height

We conducted single-variant meta-analyses in a discovery sample of 458,927 individuals, of whom 381,625 were of European ancestry. We validated our association results in an independent set of 252,501 participants. We first performed standard single-variant association analyses (Extended Data Figs 1–3 and Supplementary Tables 1–11; technical details of the discovery and validation steps are presented in the Methods). In total, we found 606 independent ExomeChip variants at array-wide significance ($P < 2 \times 10^{-7}$), including 252 non-synonymous or splice-site variants (Methods and Supplementary Table 11). Focusing on non-synonymous or splice-site variants with a MAF < 5%, our single-variant analyses identified 32 rare and 51 low-frequency height-associated variants (Extended Data Tables 1, 2). To our knowledge, these 83 height variants (MAF range of 0.1–4.8%) represent the largest set of validated rare and low-frequency coding variants associated with any complex human trait or disease to date. Among these 83 variants, there are 81 missense, one nonsense (in *CCND3*), and one essential acceptor splice site (in *ARMC5*) variants.

We observed a strong inverse relationship between MAF and effect size (Fig. 1). Although power limits our capacity to find rare variants with small effects, we know that common variants with effect sizes comparable to the largest seen in our study would have been easily discovered by prior GWAS, but were not detected. Our results agree with a model based on accumulating theoretical and empirical evidence that suggest that variants with strong phenotypic effects are more likely to be deleterious, and therefore rarer[12,13]. The largest effect sizes were observed for four rare missense variants, located in the androgen receptor gene *AR* (NCBI single nucleotide polymorphism (SNP) reference ID: rs137852591; MAF = 0.21%, $P_{combined} = 2.7 \times 10^{-14}$), in *CRISPLD2* (rs148934412; MAF = 0.08%, $P_{combined} = 2.4 \times 10^{-20}$), in *IHH* (rs142036701, MAF = 0.08%, $P_{combined} = 1.9 \times 10^{-23}$), and in *STC2* (rs148833559, MAF = 0.1%, $P_{combined} = 1.2 \times 10^{-30}$). Carriers of the rare *STC2* missense variant are approximately 2.1 cm taller than non-carriers, whereas carriers of the remaining three variants (or hemizygous men that carry a rare X-linked *AR* allele at rs137852591) are approximately 2 cm shorter than non-carriers. By comparison, the mean effect size of common height alleles is ten times smaller in the same dataset. Across all 83 rare and low-frequency non-synonymous variants, the minor alleles were evenly distributed between height-increasing and height-decreasing effects (48% and 52%, respectively) (Fig. 1 and Extended Data Tables 1, 2).

Coding variants in new and known height loci

Many of the height-associated variants discovered in this study are located near common variants previously associated with height.

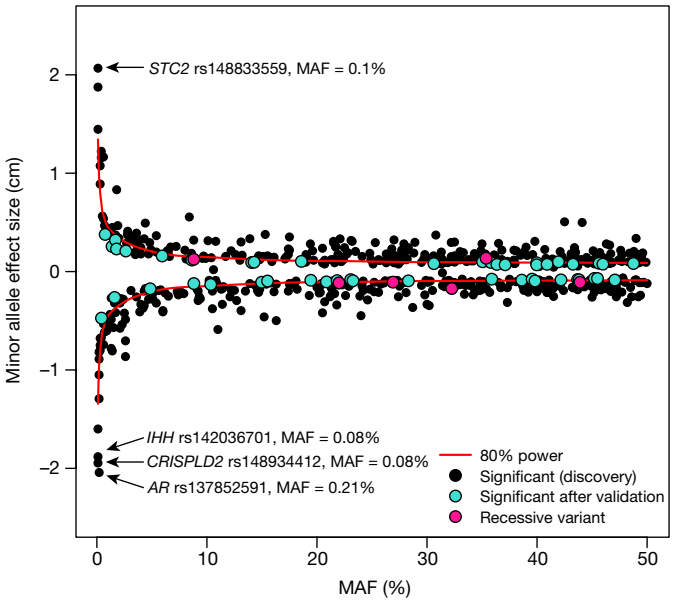


Figure 1 | Variants with a larger effect size on height variation tend to be rarer. An inverse relationship between the effect size (from the combined 'discovery and validation' analysis, in centimetres on the *y* axis) and the MAF for the height variants (*x* axis, from 0 to 50%) can be observed. Included in this figure are the 606 height variants with a $P < 2 \times 10^{-7}$.

Of the 83 rare and low-frequency non-synonymous variants, 2 low-frequency missense variants were previously identified (in *CYTL1* and *IL11*)[3,14] and 47 fell within 1 Mb of a known height signal; the remaining 34 define new loci. We used conditional analysis of the UK Biobank dataset and confirmed that 38 of these 47 variants were independent of the previously described height SNPs (Supplementary Table 12). We validated the UK Biobank conditional results using an orthogonal imputation-based methodology implemented in the full discovery set (Extended Data Fig. 4 and Supplementary Table 12). In addition, we found a further 85 common variants and one low-frequency synonymous variant (in *ACHE*) that define novel loci (Supplementary Table 12). Thus, our study identified a total of 120 new height-associated loci (Supplementary Table 11).

We used the UK Biobank dataset to estimate the contribution of the new height variants to heritability, which is $h^2 \approx 80\%$ for adult height[2]. In combination, the 83 rare and low-frequency variants explained 1.7% of the heritability of height. The newly identified novel common variants accounted for another 2.4% and all independent variants, known and novel, together explained 27.4% of heritability. By comparison, the 697 known height-associated SNPs explain 23.3% of height heritability in the same dataset (versus the 4.1% explained by the new height-associated variants identified in this study). We observed a modest positive association between MAF and heritability for each variant ($P = 0.012$, Extended Data Fig. 5), with each common variant explaining slightly more heritability than rare or low-frequency variants (0.036% versus 0.026%, Extended Data Fig. 5).

Gene–based association results

To increase the power to find rare or low-frequency coding variants associated with height, we performed gene-based analyses (Methods and Supplementary Tables 13–15). After accounting for gene-based signals explained by a single variant driving the association statistics, we identified ten genes with $P < 5 \times 10^{-7}$ that harboured more than one coding variant independently associated with height variation (Supplementary Tables 16, 17). These gene-based results remained significant after conditioning on genotypes at nearby common height-associated variants present on the ExomeChip (Table 1). Using the same gene-based tests in an independent dataset of 59,804 individuals genotyped on the same exome array, we replicated three genes at $P < 0.05$ (Table 1). Further evidence for replication in these genes was seen at the level of single variants (Supplementary Table 18). From the gene-based results, three genes—*CSAD*, *NOX4*, and *UGGT2*—are outside of the loci found by single-variant analyses and are implicated in human height for the first time to our knowledge.

Coding variants implicate pathways in skeletal growth

Previous pathway analyses of height loci identified by GWAS have highlighted gene sets related to both general biological processes (such as chromatin modification and regulation of embryonic size) and skeletal-growth-specific pathways (such as chondrocyte biology, extracellular matrix and skeletal development)[3]. We used two different methods, DEPICT[15] and PASCAL[16] (see Methods), to perform pathway analyses using the ExomeChip results to test whether coding variants could independently confirm the relevance of these previously highlighted pathways (and further implicate specific genes in these pathways) or identify new pathways. To compare the pathways emerging from coding and non-coding variation, we

Table 1 | Ten height genes implicated by gene-based testing

Gene	Discovery gene-based *P* value				Validation *P* value*	Combined *P* value*	Conditional *P* value†	Note‡
	SKAT-broad	VT-broad	SKAT-strict	VT-strict				
OSGIN1	$\mathbf{4.3 \times 10^{-11}}$	4.5×10^{-5}	0.19	0.18	0.048	2.6×10^{-12}	7.7×10^{-11}	Known locus. No predicted causal genes.
CRISPLD1	2.2×10^{-7}	$\mathbf{6.7 \times 10^{-11}}$	8.5×10^{-6}	8.9×10^{-7}	0.50	1.2×10^{-12}	NA	Known locus, sentinel GWAS SNP not tested on ExomeChip. Predicted to be causal.
CSAD	2.3×10^{-8}	$\mathbf{2.4 \times 10^{-9}}$	0.83	0.59	0.54	2.0×10^{-9}	NA	New locus.
SNED1	1.9×10^{-5}	$\mathbf{4.3 \times 10^{-9}}$	NA	NA	0.083	4.5×10^{-10}	1.4×10^{-9}	Known locus. Not predicted to be causal.
G6PC	1.3×10^{-5}	$\mathbf{3.6 \times 10^{-8}}$	5.5×10^{-6}	1.3×10^{-6}	0.24	5.2×10^{-8}	3.9×10^{-8}	Known locus. Not predicted to be causal. Mutated in glycogen storage disease type 1a.
NOX4	5.1×10^{-6}	$\mathbf{1.4 \times 10^{-7}}$	NA	NA	0.013	5.5×10^{-9}	NA	New locus.
UGGT2	3.0×10^{-5}	$\mathbf{2.6 \times 10^{-7}}$	2.3×10^{-5}	4.8×10^{-7}	0.64	3.4×10^{-7}	NA	New locus.
FLNB	2.2×10^{-6}	5.1×10^{-4}	$\mathbf{2.4 \times 10^{-9}}$	3.2×10^{-6}	0.016	8.6×10^{-11}	3.6×10^{-9}	Known locus. Predicted to be causal; mutated in atelosteogenesis type 1.
B4GALNT3	2.4×10^{-5}	1.9×10^{-5}	1.8×10^{-5}	$\mathbf{3.1 \times 10^{-7}}$	0.79	4.3×10^{-7}	7.7×10^{-7}	Known locus. Predicted to be causal.
CCDC3	6.3×10^{-4}	6.3×10^{-6}	3.0×10^{-7}	$\mathbf{5.4 \times 10^{-9}}$	0.080	1.2×10^{-9}	1.6×10^{-9}	Known locus. Predicted to be causal.

These genes meet our three criteria for statistical significance: (1) gene-based $P < 5 \times 10^{-7}$; (2) the gene does not include variants with $P < 2 \times 10^{-7}$; and (3) the gene-based *P* value is at least two orders of magnitude smaller than the *P* value for the most significant variant within the gene. For each gene, we provide *P* values for the four different gene-based tests applied. *P* values in bold are the most significant results for a given gene. NA, not applicable.

*Validation ($n = 59{,}804$) and combined results using the same test and (when possible) variants.

†When the gene is located in a locus identified by our single-variant analysis (1-Mb window), we conditioned the gene-based association result on genotypes at the single variant(s).

‡If the gene falls within a known GWAS height locus, we mention whether it was predicted to be causal using bioinformatic tools[3].

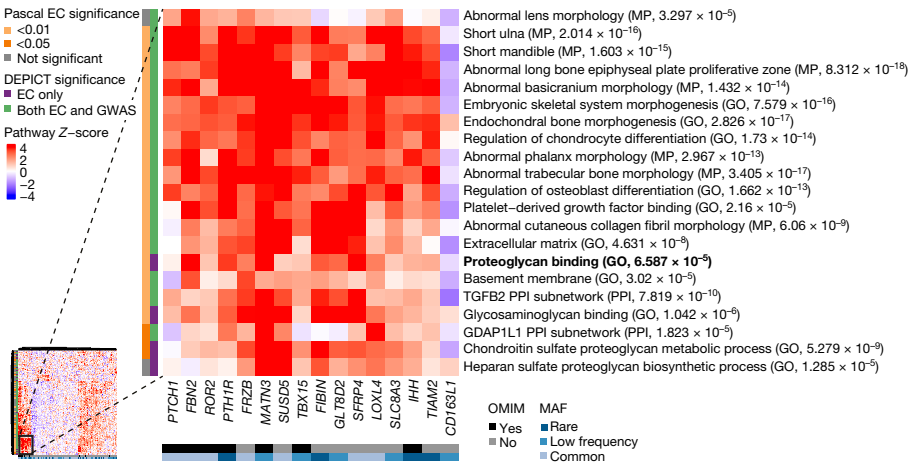


Figure 2 | Heat map showing subset of DEPICT gene set enrichment results. The full heat map is available in Extended Data Fig. 7. For any given square, the colour indicates how strongly the corresponding gene (shown on the *x* axis) is predicted to belong to the reconstituted gene set (*y* axis). This value is based on the gene's *Z* score for gene set inclusion in DEPICT's reconstituted gene sets, where red indicates a higher *Z* score and blue indicates a lower one. The proteoglycan-binding pathway (bold) was uniquely implicated by coding variants by DEPICT and PASCAL. To visually reduce redundancy and increase clarity, we chose one representative meta-gene set for each group of highly correlated gene sets, based on affinity propagation clustering (Supplementary Information). Heat map intensity and DEPICT *P* values correspond to the most significantly enriched gene set within the meta-gene set; meta-gene sets are listed with their database source. Annotations for the genes indicate whether the gene has OMIM annotation as underlying a disorder of skeletal growth (black and grey) and the MAF of the significant ExomeChip (EC) variant (shades of blue; if multiple variants, the lowest-frequency variant was kept). Annotations for the gene sets indicate if the gene set was also found significant for ExomeChip by PASCAL (yellow, orange and grey) and if the gene set was found significant by DEPICT for ExomeChip only or for both ExomeChip and GWAS (purple and green). GO, Gene Ontology; MP, mouse phenotype in the Mouse Genetics Initiative; PPI, protein–protein interaction in the InWeb database.

applied DEPICT separately onto exome-array-wide associated coding variants independent of known GWAS signals and onto non-coding GWAS loci, excluding all novel height-associated genes implicated by coding variants. We identified a total of 496 and 1,623 enriched gene sets, respectively, at a false discovery rate < 1% (Supplementary Tables 19, 20); similar analyses with PASCAL yielded 362 and 278 enriched gene sets, respectively (Supplementary Tables 21, 22). Comparison of the results revealed a high degree of shared biology for coding and non-coding variants (for DEPICT, gene set *P* values compared between coding and non-coding results had a Pearson's $r = 0.583$, $P < 2.2 \times 10^{-16}$; for PASCAL, Pearson's $r = 0.605$, $P < 2.2 \times 10^{-16}$). However, some pathways were more strongly enriched for either coding or non-coding genetic variation. In general, coding variants more strongly implicated pathways specific to skeletal growth (such as extracellular matrix and bone growth), whereas GWAS signals highlighted more global biological processes (such as transcription factor binding and embryonic size or lethality) (Extended Data Fig. 6). The two significant gene sets identified by DEPICT and PASCAL that uniquely implicated coding variants were the BCAN protein–protein interaction sub-network and the proteoglycan-binding set. Both of these pathways relate to the biology of proteoglycans, which are proteins (such as aggrecan) that contain glycosaminoglycans (such as chrondroitin sulfate) and that have well established connections to skeletal growth[17].

We also investigated which height-associated genes identified by ExomeChip analyses were driving enrichment of pathways such as proteoglycan binding. Using unsupervised clustering analysis, we observed that a cluster of 15 height-associated genes was strongly implicated in a group of correlated pathways that include biology related to proteoglycans and glycosaminoglycans (Fig. 2 and Extended Data Fig. 7). Seven of these 15 genes overlap a previously curated list of 277 genes annotated in OMIM (http://omim.org/) as causing skeletal growth disorders[3]; genes in this small cluster are enriched for OMIM annotations relative to genes outside the cluster (odds ratio = 27.6, Fisher's exact $P = 1.1 \times 10^{-5}$). As such, the remaining genes in this cluster may harbour variants that cause Mendelian growth disorders. Within this group are genes that are largely uncharacterized (*SUSD5*), have relevant biochemical functions (*GLT8D2*, a glycosyltransferase studied mostly in the context of the liver[18]; *LOXL4*, a lysyl oxidase expressed in cartilage[19]), modulate pathways known to affect skeletal growth (*FIBIN*, *SFRP4*)[20,21] or lead to increased body length when knocked out in mice (*SFRP4*)[22].

Functional characterization of rare *STC2* variants

To investigate whether the identified rare coding variants affect protein function, we performed *in vitro* functional analyses of two rare coding variants in a particularly compelling and novel candidate gene, *STC2*. Overexpression of *STC2* diminishes growth in mice by covalent binding to and inhibition of the proteinase PAPP-A, which specifically cleaves insulin growth factor binding protein 4 (IGFBP-4), leading to reduced levels of bioactive insulin-like growth factors[23] (Fig. 3a). Although there was no prior genetic evidence implicating *STC2* variation in human growth, the *PAPPA* and *IGFBP4* genes have both been implicated in height GWAS[3], and rare mutations in *PAPPA2* cause severe short stature[24], emphasizing the likely relevance of this pathway in humans. The two *STC2* height-associated variants are rs148833559 (p.R44L, MAF = 0.096%, $P_{discovery} = 5.7 \times 10^{-15}$) and rs146441603 (p.M86I, MAF = 0.14%, $P_{discovery} = 2.1 \times 10^{-5}$). These rare alleles increase height by an average of 1.9 and 0.9 cm, respectively, suggesting that they both partially impair STC2 activity. In functional studies, STC2 variants with these amino acid substitutions were expressed at similar levels to wild-type STC2, but showed clear, partial defects in binding to PAPP-A and in inhibition of PAPP-A-mediated cleavage of IGFBP-4 (Fig. 3b–d). Thus, the genetic analysis successfully identified rare coding alleles that have demonstrable and predicted functional consequences, strongly confirming the role of these variants and the *STC2* gene in human growth.

Pleiotropic effects

Previous GWAS studies have reported pleiotropic or secondary effects on other phenotypes for many common variants associated with adult height[3,25]. Using association results from 17 human complex phenotypes for which well-powered meta-analysis results are available, we investigated whether rare and low-frequency height variants are

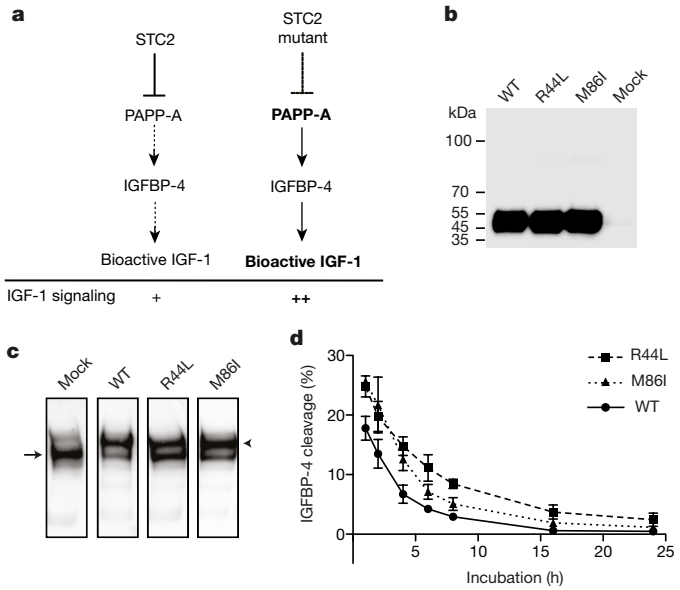


Figure 3 | STC2 mutants p.Arg44Leu (R44L) and p.Met86Ile (M86I) show compromised proteolytic inhibition of PAPP-A. a, Schematic representation of the role of STC2 in IGF-1 signalling. Partial inactivation of STC2 by height-associated DNA sequence variation could increase bioactive IGF-1 through reduced inhibition of PAPP-A. **b**, Western blot analysis of recombinant wild-type (WT) STC2 and variants R44L and M86I. **c**, Covalent complex formation between PAPP-A and wild-type STC2 or variants R44L and M86I. Separately synthesized proteins were analysed by PAPP-A western blotting following incubation for 8 h. In the absence of STC2 (Mock), PAPP-A appears as a single 400-kDa band, indicated by an arrow. Following incubation with wild-type STC2, the majority of PAPP-A is present as the approximately 500-kDa covalent PAPP-A–STC2 complex, indicated by an arrowhead, in which PAPP-A is devoid of proteolytic activity towards IGFBP-4. Under similar conditions, incubation with variants R44L or M86I appeared to cause a lesser degree of covalent complex formation with PAPP-A. The gels are representative of at least three independent experiments. **d**, PAPP-A proteolytic cleavage of IGFBP-4 following incubation with wild-type STC2 or variants for 1–24 h. Wild-type STC2 causes reduction in PAPP-A activity, with complete inhibition of activity following a 24-h incubation. Both STC2 variants show increased IGFBP-4 cleavage (that is, less inhibition) for all time points analysed. Mean ± s.d. of three independent experiments are shown. One-way repeated measures analysis of variance followed by Dunnett's post-test showed significant differences between STC2 wild-type and variants R44L ($P < 0.001$) and M86I ($P < 0.01$).

also pleiotropic. We found one rare and five low-frequency missense variants associated with at least one of the other investigated traits at array-wide significance ($P < 2 \times 10^{-7}$) (Extended Data Fig. 8 and Supplementary Table 23). The minor alleles at rs77542162 (*ABCA6*, MAF = 1.7%) and rs28929474 (*SERPINA1*, MAF = 1.8%) are associated with increased height and increased levels of low-density lipoprotein (LDL) cholesterol and total cholesterol, whereas the minor allele at rs3208856 in *CBLC* (MAF = 3.4%) is associated with increased height, high-density lipoprotein (HDL) cholesterol and triglyceride, but decreased LDL cholesterol and total cholesterol levels. The minor allele at rs141845046 (*ZBTB7B*, MAF = 2.8%) was associated with both increased height and body mass index (BMI). The minor alleles at the other two missense variants associated with shorter stature, rs201226914 in *PIEZO1* (MAF = 0.2%) and rs35658696 in *PAM* (MAF = 4.8%), were associated with decreased glycated haemoglobin (HbA1c) and increased risk of type 2 diabetes (T2D), respectively.

Discussion

We undertook an association study of nearly 200,000 coding variants in 711,428 individuals, and identified 32 rare and 51 low-frequency coding variants associated with adult height. Furthermore, gene-based testing discovered 10 genes that harbour several additional rare or low-frequency variants associated with height, including three genes (*CSAD*, *NOX4* and *UGGT2*) in loci not previously implicated in height. Given the design of the ExomeChip, which did not consider variants with a MAF < 0.004% (corresponding to approximately one allele in 12,000 participants), our gene-based association results do not rule out the possibility that additional genes with such rarer coding variants also contribute to height variation; deep DNA sequencing in very large sample sizes will be required to address this question. In total, our results highlight 89 genes (10 from gene-based testing and 79 from single-variant analyses (4 genes have 2 independent coding variants)) that are likely to modulate human growth, and 24 alleles segregating in the general population that affect height by more than 1 cm (Table 1 and Extended Data Tables 1, 2). The rare and low-frequency coding variants explain 1.7% of the heritable variation in adult height. When considering all rare, low-frequency and common height-associated variants validated in this study, we can now explain 27.4% of the heritability of height.

Our analyses revealed many coding variants in genes mutated in monogenic skeletal growth disorders, confirming the presence of allelic series (from familial penetrant mutations to mild effect common variants) in the same genes for related growth phenotypes in humans. We used gene-set-enrichment-type analyses to demonstrate the functional connectivity between the genes that harbour coding height variants, highlighting both known and novel biological pathways that regulate height in humans (Fig. 2, Extended Data Fig. 7 and Supplementary Tables 19–22), and implicating genes such as *SUSD5*, *GLT8D2*, *LOXL4*, *FIBIN* and *SFRP4* that have not been previously connected with skeletal growth. Additional noteworthy height candidate genes include *NOX4*, *ADAMTS3*, *ADAMTS6*, *PTH1R* and *IL11RA* (Extended Data Tables 1, 2 and Supplementary Tables 17, 24). *NOX4*, identified through gene-based testing, encodes NADPH oxidase 4, an enzyme that produces reactive oxygen species, a biological pathway not previously implicated in human growth. *Nox4*$^{-/-}$ mice display higher bone density and a reduced number of osteoclasts, a cell type that is essential for bone repair, maintenance and remodelling[12]. We also found rare coding variants in *ADAMTS3* and *ADAMTS6*, genes that encode metalloproteinases that belong to the same family as several other human growth syndromic genes (such as *ADAMTS2*, *ADAMTS10* and *ADAMTSL2*). Moreover, we discovered a rare missense variant in *PTH1R* that encodes a receptor for parathyroid hormone; parathyroid hormone–PTH1R signalling is important for bone resorption, and mutations in *PTH1R* cause chondrodysplasia in humans[26]. Finally, we replicated the association between a low-frequency missense variant in the cytokine gene *IL11*, but also found a low-frequency missense variant in the gene encoding its receptor, *IL11RA*. The IL11–IL11RA axis has been shown to play an important role in bone formation in the mouse[27,28]. Thus, our data confirm that this signalling cascade is also relevant in human growth.

Overall, our findings provide strong evidence that rare and low-frequency coding variants contribute to the genetic architecture of height, a model complex human trait. This conclusion has implications for the prediction of complex human phenotypes in the context of precision medicine initiatives. Although rare, large effect-size variants might not explain most of the heritable disease risk at the population level, they are important for predicting the risk of disease development for the individuals that carry them. Our findings also seem to contrast markedly with results from the recent large-scale T2D association study, which found only six variants with a MAF < 5% (ref. 29.). This apparent difference could be explained simply by the large difference in sample sizes between the two studies (711,428 for height versus 127,145 for T2D). When we consider the fraction of associated variants with a MAF < 5% among all confirmed variants for height and T2D, we find that it is similar (9.7% for height versus 7.1% for T2D). This supports the strong probability that rarer T2D

alleles and, more generally, rarer alleles for other polygenic diseases and traits will be uncovered as sample sizes continue to increase.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 11 July; accepted 4 December 2016.
Published online 1 February 2017.

1. Fisher, R. A. The correlation between relatives on the supposition of Mendelian inheritance. *Trans. R. Soc. Edinb.* **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. Wood, A. R. *et al.* Defining the role of common variation in the genomic and biological architecture of adult human height. *Nat. Genet.* **46,** 1173–1186 (2014).
4. Flannick, J. *et al.* Loss-of-function mutations in SLC30A8 protect against type 2 diabetes. *Nat. Genet.* **46,** 357–363 (2014).
5. Steinthorsdottir, V. *et al.* Identification of low-frequency and rare sequence variants associated with elevated or reduced risk of type 2 diabetes. *Nat. Genet.* **46,** 294–298 (2014).
6. Gudmundsson, J. *et al.* A study based on whole-genome sequencing yields a rare variant at 8q24 associated with prostate cancer. *Nat. Genet.* **44,** 1326–1329 (2012).
7. Sidore, C. *et al.* Genome sequencing elucidates Sardinian genetic architecture and augments association analyses for lipid and blood inflammatory markers. *Nat. Genet.* **47,** 1272–1281 (2015).
8. Danjou, F. *et al.* Genome-wide association analyses based on whole-genome sequencing in Sardinia provide insights into regulation of hemoglobin levels. *Nat. Genet.* **47,** 1264–1271 (2015).
9. Zuk, O. *et al.* Searching for missing heritability: designing rare variant association studies. *Proc. Natl Acad. Sci. USA* **111,** E455–E464 (2014).
10. Yang, J. *et al.* Genetic variance estimation with imputed variants finds negligible missing heritability for human height and body mass index. *Nat. Genet.* **47,** 1114–1120 (2015).
11. Grove, M. L. *et al.* Best practices and joint calling of the HumanExome BeadChip: the CHARGE Consortium. *PLoS One* **8,** e68095 (2013).
12. Kryukov, G. V., Pennacchio, L. A. & Sunyaev, S. R. Most rare missense alleles are deleterious in humans: implications for complex disease and association studies. *Am. J. Hum. Genet.* **80,** 727–739 (2007).
13. Tennessen, J. A. *et al.* Evolution and functional impact of rare coding variation from deep sequencing of human exomes. *Science* **337,** 64–69 (2012).
14. Lanktree, M. B. *et al.* Meta-analysis of dense genecentric association studies reveals common and uncommon variants associated with height. *Am. J. Hum. Genet.* **88,** 6–18 (2011).
15. Pers, T. H. *et al.* Biological interpretation of genome-wide association studies using predicted gene functions. *Nat. Commun.* **6,** 5890 (2015).
16. Lamparter, D., Marbach, D., Rueedi, R., Kutalik, Z. & Bergmann, S. Fast and rigorous computation of gene and pathway scores from SNP-based summary statistics. *PLOS Comput. Biol.* **12,** e1004714 (2016).
17. Schwartz, N. B. & Domowicz, M. Chondrodysplasias due to proteoglycan defects. *Glycobiology* **12,** 57R–68R (2002).
18. Wei, H. S., Wei, H. L., Zhao, F., Zhong, L. P. & Zhan, Y. T. Glycosyltransferase GLT8D2 positively regulates ApoB100 protein expression in hepatocytes. *Int. J. Mol. Sci.* **14,** 21435–21446 (2013).
19. Ito, H. *et al.* Molecular cloning and biological activity of a novel lysyl oxidase-related gene expressed in cartilage. *J. Biol. Chem.* **276,** 24023–24029 (2001).
20. Wakahara, T. *et al.* Fibin, a novel secreted lateral plate mesoderm signal, is essential for pectoral fin bud initiation in zebrafish. *Dev. Biol.* **303,** 527–535 (2007).
21. Kawano, Y. & Kypta, R. Secreted antagonists of the Wnt signalling pathway. *J. Cell Sci.* **116,** 2627–2634 (2003).
22. Mastaitis, J. *et al.* Loss of SFRP4 alters body size, food intake, and energy expenditure in diet-induced obese male mice. *Endocrinology* **156,** 4502–4510 (2015).
23. Jepsen, M. R. *et al.* Stanniocalcin-2 inhibits mammalian growth by proteolytic inhibition of the insulin-like growth factor axis. *J. Biol. Chem.* **290,** 3430–3439 (2015).
24. Dauber, A. *et al.* Mutations in pregnancy-associated plasma protein A2 cause short stature due to low IGF-I availability. *EMBO Mol. Med.* **8,** 363–374 (2016).
25. Lango Allen, H. *et al.* Hundreds of variants clustered in genomic loci and biological pathways affect human height. *Nature* **467,** 832–838 (2010).
26. Karaplis, A. C. *et al.* Inactivating mutation in the human parathyroid hormone receptor type 1 gene in Blomstrand chondrodysplasia. *Endocrinology* **139,** 5255–5258 (1998).
27. Sims, N. A. *et al.* Interleukin-11 receptor signaling is required for normal bone remodeling. *J. Bone Miner. Res.* **20,** 1093–1102 (2005).
28. Takeuchi, Y. *et al.* Interleukin-11 as a stimulatory factor for bone formation prevents bone loss with advancing age in mice. *J. Biol. Chem.* **277,** 49011–49018 (2002).
29. Fuchsberger, C. *et al.* The genetic architecture of type 2 diabetes. *Nature* **536,** 41–47 (2016).

Supplementary Information is available in the online version of the paper.

Acknowledgements A full list of acknowledgments appears in the Supplementary Information. Part of this work was conducted using the UK Biobank resource.

Author Contributions Writing group (wrote and edited manuscript): P.D., T.M.F., M.Gr., J.N.H., G.L., K.S.L., Y.Lu., E.M., C.M.-G., F.Ri. All authors contributed and discussed the results, and commented on the manuscript. Data preparation group (checked and prepared data from contributing cohorts for meta-analyses and replication): T.Es., M.Gr., H.M.H., A.E.J., T.Ka., K.S.L., A.E.L., Y.Lu., E.M., N.G.D.M., C. M.-G., P.Mu., M.C.Y.N., M.A.R., C.S., K.St., V.T., S.V., T.W.W., K.L.Y. This work was done under the auspices of the GIANT, CHARGE, BBMRI, UK ExomeChip, and GOT2D consortia. Height meta-analyses (discovery and replication, single-variant and gene-based): P.D., T.M.F., M.Gr., J.N.H., G.L., D.J.L., K.S.L., Y.Lu, E.M., C.M.-G., F.Ri., A.R.W. UK Biobank-based integration of height association signals group and heritability analyses: P.D., T.M.F., G.L., Z.K., K.S.L., E.M., S.R., A.R.W. Pleiotropy working group: G.A., M. Bo., J.P.C., P.D., F.D., J.C.F., H.M.H., S. Kat., C.M.L., D.J.L., R.J.F.L., A.Ma., E.M., M.I.M., P.B.M., G.M.P., J.R.B.P., K.S.R., C.J.W. Biological and clinical enrichment and pathway analyses: R.S.F., J.N.H., Z.K., D.L., G.L., K.S.L., T.H.P. Functional characterization of STC2: T.R.K., C.O.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to J.N.H. (joelh@broadinstitute.org), P.D. (p.deloukas@qmul.ac.uk) or G.L. (guillaume.lettre@umontreal.ca).

Reviewer Information *Nature* thanks J. Barrett, D. Hinds and D. Hunter for their contribution to the peer review of this work.

Eirini Marouli[1]*, Mariaelisa Graff[2]*, Carolina Medina-Gomez[3,4]*, Ken Sin Lo[5]*, Andrew R. Wood[6]*, Troels R. Kjaer[7]*, Rebecca S. Fine[8,9,10]*, Yingchang Lu[11,12,13]*, Claudia Schurmann[12,13], Heather M. Highland[2,14], Sina Rüeger[15,16], Gudmar Thorleifsson[17], Anne E. Justice[2], David Lamparter[16,18], Kathleen E. Stirrups[1,19], Valérie Turcot[5], Kristin L. Young[2], Thomas W. Winkler[20], Tõnu Esko[8,10,21], Tugce Karaderi[22], Adam E. Locke[23,24], Nicholas G. D. Masca[25,26], Maggie C. Y. Ng[27,28], Poorva Mudgal[27], Manuel A. Rivas[8,29], Sailaja Vedantam[8,9,10], Anubha Mahajan[22], Xiuqing Guo[30], Goncalo Abecasis[23], Katja K. Aben[31,32], Linda S. Adair[33], Dewan S. Alam[34], Eva Albrecht[35], Kristine H. Allin[36], Matthew Allison[37], Philippe Amouyel[38,39,40], Emil V. Appel[36], Dominique Arveiler[41,42], Folkert W. Asselbergs[43,44,45], Paul L. Auer[46], Beverley Balkau[47], Bernhard Banas[48], Lia E. Bang[49], Marianne Benn[50,51], Sven Bergmann[16,18], Lawrence F. Bielak[52], Matthias Blüher[53,54], Heiner Boeing[55], Eric Boerwinkle[56,57], Carsten A. Böger[48], Lori L. Bonnycastle[58], Jette Bork-Jensen[36], Michiel L. Bots[59], Erwin P. Bottinger[12], Donald W. Bowden[27,28,60], Ivan Brandslund[61,62], Gerome Breen[63], Murray H. Brilliant[64], Linda Broer[4], Amber A. Burt[65], Adam S. Butterworth[66,67], David J. Carey[68], Mark J. Caulfield[1,69], John C. Chambers[70,71,72], Daniel I. Chasman[8,73,74,75], Yii-Der Ida Chen[30], Rajiv Chowdhury[66], Cramer Christensen[76], Audrey Y. Chu[74,77], Massimiliano Cocca[78], Francis S. Collins[58], James P. Cook[79], Janie Corley[80,81], Jordi Corominas Galbany[82], Amanda J. Cox[27,28,83], Gabriel Cuellar-Partida[84,85], John Danesh[66,67,86,87], Gail Davies[80,81], Paul I. W. de Bakker[59,88], Gert J. de Borst[89], Simon de Denus[5,90], Mark C. H. de Groot[91,92], Renée de Mutsert[93], Ian J. Deary[80,81], George Dedoussis[94], Ellen W. Demerath[95], Anneke I. den Hollander[96], Joe G. Dennis[97], Emanuele Di Angelantonio[66,67], Fotios Drenos[98,99], Mengmeng Du[100,101], Alison M. Dunning[102], Douglas F. Easton[97,102], Tapani Ebeling[103,104], Todd L. Edwards[105], Patrick T. Ellinor[106,107], Paul Elliott[108], Evangelos Evangelou[71,109], Aliki-Eleni Farmaki[94], Jessica D. Faul[110], Mary F. Feitosa[111], Shuang Feng[23], Ele Ferrannini[112,113], Marco M. Ferrario[114], Jean Ferrieres[115], Jose C. Florez[106,107,116], Ian Ford[117], Myriam Fornage[118], Paul W. Franks[119,120,121], Ruth Frikke-Schmidt[51,122], Tessel E. Galesloot[32], Wei Gan[22], Ilaria Gandin[123], Paolo Gasparini[123,124], Vilmantas Giedraitis[125], Ayush Giri[105], Giorgia Girotto[123,124], Scott D. Gordon[85], Penny Gordon-Larsen[126,127], Mathias Gorski[20,48], Niels Grarup[36], Megan L. Grove[56], Vilmundur Gudnason[128,129], Stefan Gustafsson[130], Torben Hansen[36], Kathleen Mullan Harris[126,131], Tamara B. Harris[132], Andrew T. Hattersley[133], Caroline Hayward[134], Liang He[135,136], Iris M. Heid[20,35], Kauko Heikkilä[136,137], Øyvind Helgeland[138,139], Jussi Hernesniemi[140,141,142], Alex W. Hewitt[143,144,145], Lynne J. Hocking[146,147], Mette Hollensted[36], Oddgeir L. Holmen[148], G. Kees Hovingh[149], Joanna M. M. Howson[66], Carel B. Hoyng[96], Paul L. Huang[106], Kristian Hveem[150], M. Arfan Ikram[3,151,152], Erik Ingelsson[130,153], Anne U. Jackson[23], Jan-Håkan Jansson[154,155], Gail P. Jarvik[65,156], Gorm B. Jensen[157], Min A. Jhun[52], Yucheng Jia[30], Xuejuan Jiang[158,159], Stefan Johansson[139,160], Marit E. Jørgensen[161,162], Torben Jørgensen[51,163,164], Pekka Jousilahti[165], J. Wouter Jukema[166,167], Bratati Kahali[168,169,170], René S. Kahn[171], Mika Kähönen[172], Pia R. Kamstrup[50], Stavroula Kanoni[1], Jaakko Kaprio[136,137,165], Maria Karaleftheri[173], Sharon L. R. Kardia[52], Fredrik Karpe[174,175], Frank Kee[176], Renske Keeman[177], Lambertus A. Kiemeney[32], Hidetoshi Kitajima[22], Kirsten B. Kluivers[32], Thomas Kocher[178], Pirjo Komulainen[179], Jukka Kontto[165], Jaspal S. Kooner[70,72,180], Charles Kooperberg[181], Peter Kovacs[53], Jennifer Kriebel[182,183,184], Helena Kuivaniemi[68,185], Sébastien Küry[186], Johanna Kuusisto[187], Martina La Bianca[188], Markku Laakso[187], Timo A. Lakka[179,189], Ethan M. Lange[190], Leslie A. Lange[190], Carl D. Langefeld[191], Claudia Langenberg[192], Eric B. Larson[65,193,194], I-Te Lee[195,196,197], Terho Lehtimäki[141,142], Cora E. Lewis[198], Huaixing Li[199], Jin Li[200], Ruifang Li-Gao[93], Honghuang Lin[201], Li-An Lin[118], Xu Lin[199], Lars Lind[202], Jaana Lindström[165], Allan Linneberg[51,164,203], Yeheng Liu[30], Yongmei Liu[204], Artitaya Lophatananon[205], Jian'an Luan[192], Steven A. Lubitz[106,107], Leo-Pekka Lyytikäinen[141,142], David A. Mackey[144], Pamela A. F. Madden[206], Alisa K. Manning[106,107,116], Satu Männistö[165], Gaëlle Marenne[86], Jonathan Marten[134], Nicholas G. Martin[85], Angela L. Mazul[2], Karina Meidtner[182,207], Andres Metspalu[21], Paul Mitchell[208], Karen L. Mohlke[190], Dennis O. Mook-Kanamori[93,209], Anna Morgan[123], Andrew D. Morris[210], Andrew P. Morris[22,79], Martina Müller-Nurasyid[35,211,212], Patricia B. Munroe[1,69], Mike A. Nalls[213], Matthias Nauck[214,215], Christopher P. Nelson[25,26], Matt Neville[174,175], Sune F. Nielsen[50,51], Kjell Nikus[216], Pål R. Njølstad[138,139], Børge G. Nordestgaard[50,51], Ioanna Ntalla[1], Jeffrey R. O'Connel[217], Heikki Oksa[218], Loes M. Olde Loohuis[219], Roel A. Ophoff[171,219], Katharine R. Owen[174,175], Chris J. Packard[117], Sandosh Padmanabhan[117], Colin N. A. Palmer[220], Gerard Pasterkamp[221,222], Aniruddh P. Patel[8,75,106], Alison Pattie[81], Oluf Pedersen[36], Peggy L. Peissig[64], Gina M. Peloso[106,107], Craig E. Pennell[223], Markus Perola[165,224,225], James A. Perry[217], John R. B. Perry[192], Thomas N. Person[64], Ailith Pirie[102], Ozren Polasek[210,226], Danielle Posthuma[227,228], Olli T. Raitakari[229,230], Asif Rasheed[231], Rainer Rauramaa[179,232], Dermot F. Reilly[233], Alex P. Reiner[181,234], Frida Renström[119,235], Paul M. Ridker[74,75,236], John D. Rioux[5,237], Neil Robertson[22,174], Antonietta Robino[188], Olov Rolandsson[154,238], Igor Rudan[210], Katherine S. Ruth[6], Danish Saleheen[231,239], Veikko Salomaa[165], Nilesh J. Samani[25,26], Kevin Sandow[30], Yadav Sapkota[85], Naveed Sattar[117], Marjanka K. Schmidt[177], Pamela J. Schreiner[240], Matthias B. Schulze[182,207], Robert A. Scott[192], Marcelo P. Segura-Lepe[71], Svati Shah[241], Xueling Sim[23,242], Suthesh Sivapalaratnam[106,243,244], Kerrin S. Small[245], Albert Vernon Smith[128,129], Jennifer A. Smith[52], Lorraine Southam[22,86], Timothy D. Spector[245], Elizabeth K. Speliotes[168,169,170], John M. Starr[80,246], Valgerdur Steinthorsdottir[17], Heather M. Stringham[23], Michael Stumvoll[53,54], Praveen Surendran[66], Leen M. 't Hart[247,248,249], Katherine E. Tansey[250,251], Jean-Claude Tardif[5,237], Kent D. Taylor[30], Alexander Teumer[252], Deborah J. Thompson[97], Unnur Thorsteinsdottir[17,128], Betina H. Thuesen[164], Anke Tönjes[253], Gerard Tromp[68,254], Stella Trompet[166,255], Emmanouil Tsafantakis[256], Jaakko Tuomilehto[165,257,258,259], Anne Tybjaerg-Hansen[51,122], Jonathan P. Tyrer[102], Rudolf Uher[260], André G. Uitterlinden[3,4], Sheila Ulivi[188], Sander W. van der Laan[222], Andries R. Van Der Leij[261], Cornelia M. van Duijn[3], Natasja M. van Schoor[247], Jessica van Setten[43], Anette Varbo[50,51], Tibor V. Varga[119], Rohit Varma[159], Digna R. Velez Edwards[262], Sita H. Vermeulen[32], Henrik Vestergaard[36], Veronique Vitart[134], Thomas F. Vogt[263], Diego Vozzi[124], Mark Walker[264], Feijie Wang[199], Carol A. Wang[223], Shuai Wang[265], Yiqin Wang[199], Nicholas J. Wareham[192], Helen R. Warren[1,69], Jennifer Wessel[266], Sara M. Willems[192], James G. Wilson[267], Daniel R. Witte[268,269], Michael O. Woods[270], Ying Wu[190], Hanieh Yaghootkar[6], Jie Yao[30], Pang Yao[199], Laura M. Yerges-Armstrong[217,271], Robin Young[66,117], Eleftheria Zeggini[86], Xiaowei Zhan[272], Weihua Zhang[70,71], Jing Hua Zhao[192], Wei Zhao[239], Wei Zhao[52], He Zheng[199], Wei Zhou[168,169], The EPIC-InterAct Consortium†, EPIC-CVD Consortium†, CHD Exome+ Consortium†, ExomeBP Consortium†, T2D-Genes Consortium†, GoT2D Genes Consortium†, Global Lipids Genetics Consortium†, ReproGen Consortium†, MAGIC Investigators†, Jerome I Rotter[30], Michael Boehnke[23], Sekar Kathiresan[8,75,106], Mark I. McCarthy[22,174,175], Cristen J. Willer[168,169,273], Kari Stefansson[17,128], Ingrid B. Borecki[111], Dajiang J. Liu[274], Kari E. North[275], Nancy L. Heard-Costa[77,276], Tune H. Pers[36,277], Cecilia M. Lindgren[22,278], Claus Oxvig[7]§, Zoltán Kutalik[15,16]§, Fernando Rivadeneira[3,4]§, Ruth J. F. Loos[12,13,279]§, Timothy M. Frayling[6]§, Joel N. Hirschhorn[8,10,280]§, Panos Deloukas[1,281]§ & Guillaume Lettre[5,237]§

[1]William Harvey Research Institute, Barts and The London School of Medicine and Dentistry, Queen Mary University of London, London EC1M 6BQ, UK. [2]Department of Epidemiology, University of North Carolina, Chapel Hill, North Carolina 27514, USA. [3]Department of Epidemiology, Erasmus Medical Center, Rotterdam, 3015 GE, The Netherlands. [4]Department of Internal Medicine, Erasmus Medical Center, Rotterdam, 3015 GE, The Netherlands. [5]Montreal Heart Institute, Montreal, Quebec H1T 1C8, Canada. [6]Genetics of Complex Traits, University of Exeter Medical School, University of Exeter, Exeter EX2 5DW, UK. [7]Department of Molecular Biology and Genetics, Aarhus University, Aarhus, 8000, Denmark. [8]Broad Institute of MIT and Harvard, Cambridge, Massachusetts 02142, USA. [9]Department of Genetics, Harvard Medical School, Boston, Massachusetts 02115, USA. [10]Division of Endocrinology and Center for Basic and Translational Obesity Research, Boston Children's Hospital, Boston, Massachusetts 02115, USA. [11]Division of Epidemiology, Department of Medicine, Vanderbilt-Ingram Cancer Center, Vanderbilt Epidemiology Center, Vanderbilt University School of Medicine, Nashville, Tennessee 37203, USA. [12]The Charles Bronfman Institute for Personalized Medicine, Icahn School of Medicine at Mount Sinai, New York, New York 10029, USA. [13]The Genetics of Obesity and Related Metabolic Traits Program, Icahn School of Medicine at Mount Sinai, New York, New York 10069, USA. [14]Human Genetics Center, The University of Texas School of Public Health, The University of Texas Graduate School of Biomedical Sciences at Houston, The University of Texas Health Science Center at Houston, Houston, Texas 77030, USA. [15]Institute of Social and Preventive Medicine, Lausanne University Hospital, Lausanne, 1010, Switzerland. [16]Swiss Institute of Bioinformatics, Lausanne, 1015, Switzerland. [17]deCODE Genetics/Amgen inc., Reykjavik, 101, Iceland. [18]Department of Computational Biology, University of Lausanne, Lausanne, 1011, Switzerland. [19]Department of Haematology, University of Cambridge, Cambridge CB2 0PT, UK. [20]Department of Genetic Epidemiology, University of Regensburg, Regensburg, D-93051, Germany. [21]Estonian Genome Center, University of Tartu, Tartu, 51010, Estonia. [22]Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford OX3 7BN, UK. [23]Department of Biostatistics and Center for Statistical Genetics, University of Michigan, Ann Arbor, Michigan 48109, USA. [24]McDonnell Genome Institute, Washington University School of Medicine, Saint Louis, Missouri 63108, USA. [25]Department of Cardiovascular Sciences, Univeristy of Leicester, Glenfield Hospital, Leicester LE3 9QP, UK. [26]NIHR Leicester Cardiovascular Biomedical Research Unit, Glenfield Hospital, Leicester LE3 9QP, UK. [27]Center for Diabetes Research, Wake Forest School of Medicine, Winston-Salem, North Carolina 27157, USA. [28]Center for Genomics and Personalized Medicine Research, Wake Forest School of Medicine, Winston-Salem, North Carolina 27157, USA. [29]Nuffield Department of Clinical Medicine, Oxford OX3 7BN, UK. [30]Institute for Translational Genomics and Population Sciences, LABioMed at Harbor-UCLA Medical Center, Torrance, California 90502, USA. [31]Netherlands Comprehensive Cancer Organisation, Utrecht, 3501 DB, The Netherlands. [32]Radboud University Medical Center, Nijmegen, 6500 HB, The Netherlands. [33]Department of Nutrition, University of North Carolina, Chapel Hill, North Carolina 27599, USA. [34]Centre for Control of Chronic Diseases (CCCD), Dhaka, 1212, Bangladesh. [35]Institute of Genetic Epidemiology, Helmholtz Zentrum München - German Research Center for Environmental Health, Neuherberg, D-85764, Germany. [36]The Novo Nordisk Foundation Center for Basic Metabolic Research, Faculty of Health and Medical Sciences, University of Copenhagen, Copenhagen, 2100, Denmark. [37]Department of Family Medicine & Public Health, University of California, San Diego, La Jolla, California 92093, USA. [38]INSERM U1167, Lille, F-59019, France. [39]Institut Pasteur de Lille, U1167, Lille, F-59019, France. [40]Universite de Lille, U1167 - RID-AGE - Risk factors and molecular determinants of aging-related diseases, Lille, F-59019, France. [41]Department of Epidemiology and Public Health, University of Strasbourg, Strasbourg, F-67085, France. [42]Department of Public Health, University Hospital of Strasbourg, Strasbourg, 67081, France. [43]Department of Cardiology, Division Heart & Lungs, University Medical Center Utrecht, Utrecht, The Netherlands. [44]Durrer Center for Cardiogenetic Research, ICIN-

Netherlands Heart Institute, Utrecht, The Netherlands. [45]Institute of Cardiovascular Science, Faculty of Population Health Sciences, University College London, London, UK. [46]Zilber School of Public Health, University of Wisconsin-Milwaukee, Milwaukee, Wisconsin 53201, USA. [47]INSERM U1018, Centre de recherche en Épidemiologie et Sante des Populations (CESP), Villejuif, France. [48]Department of Nephrology, University Hospital Regensburg, Regensburg, 93042, Germany. [49]Department of Cardiology, Rigshospitalet, Copenhagen University Hospital, Copenhagen, 2100, Denmark. [50]Department of Clinical Biochemistry, Herlev and Gentofte Hospital, Copenhagen University Hospital, Herlev, 2730, Denmark. [51]Faculty of Health and Medical Sciences, University of Copenhagen, Copenhagen, 2200, Denmark. [52]Department of Epidemiology, School of Public Health, University of Michigan, Ann Arbor, Michigan 48109, USA. [53]IFB Adiposity Diseases, University of Leipzig, Leipzig, 04103, Germany. [54]University of Leipzig, Department of Medicine, Leipzig, 04103, Germany. [55]Department of Epidemiology, German Institute of Human Nutrition Potsdam-Rehbruecke (DIfE), Nuthetal, 14558, Germany. [56]School of Public Health, Human Genetics Center, The University of Texas Health Science Center at Houston, Houston, Texas 77030, USA. [57]Human Genome Sequencing Center, Baylor College of Medicine, Houston, Texas 77030, USA. [58]Medical Genomics and Metabolic Genetics Branch, National Human Genome Research Institute, National Institutes of Health, Bethesda, Maryland 20892, USA. [59]Julius Center for Health Sciences and Primary Care, University Medical Center Utrecht, Utrecht, The Netherlands. [60]Department of Biochemistry, Wake Forest School of Medicine, Winston-Salem, North Carolina 27157, USA. [61]Department of Clinical Biochemistry, Lillebaelt Hospital, Vejle, 7100, Denmark. [62]Institute of Regional Health Research, University of Southern Denmark, Odense, 5000, Denmark. [63]MRC Social Genetic and Developmental Psychiatry Centre, Institute of Psychiatry, Psychology and Neuroscience, King's College London & NIHR Biomedical Research Centre for Mental Health at the Maudsley, London, SE5 8AF, UK. [64]Marshfield Clinic Research Foundation, Marshfield, Wisconsin 54449, USA. [65]Department of Medicine, University of Washington, Seattle, Washington 98195, USA. [66]MRC/BHF Cardiovascular Epidemiology Unit, Department of Public Health and Primary Care, University of Cambridge, Cambridge CB1 8RN, UK. [67]NIHR Blood and Transplant Research Unit in Donor Health and Genomics, University of Cambridge, Cambridge CB1 8RN, UK. [68]The Sigfried and Janet Weis Center for Research, Danville, Pennsylvania 17822, USA. [69]NIHR Barts Cardiovascular Research Unit, Barts and The London School of Medicine & Dentistry, Queen Mary University, London EC1M 6BQ, UK. [70]Department of Cardiology, London North West Healthcare NHS Trust, Ealing Hospital, Middlesex UB1 3HW, UK. [71]Department of Epidemiology and Biostatistics, School of Public Health, Imperial College London, London W2 1PG, UK. [72]Imperial College Healthcare NHS Trust, London W12 0HS, UK. [73]Division of Genetics, Brigham and Women's Hospital and Harvard Medical School, Boston, Massachusetts 02115, USA. [74]Division of Preventive Medicine, Brigham and Women's and Harvard Medical School, Boston, Massachusetts 02215, USA. [75]Harvard Medical School, Boston, Massachusetts 02115, USA. [76]Medical Department, Lillebaelt Hospital, Vejle, 7100, Denmark. [77]NHLBI Framingham Heart Study, Framingham, Massachusetts 01702, USA. [78]Department of Medical, Surgical and Health Sciences, University of Trieste, Trieste, 34100, Italy. [79]Department of Biostatistics, University of Liverpool, Liverpool L69 3GL, UK. [80]Centre for Cognitive Ageing and Cognitive Epidemiology, University of Edinburgh, Edinburgh EH8 9JZ, UK. [81]Department of Psychology, University of Edinburgh, Edinburgh EH8 9JZ, UK. [82]Department of Human Genetics, Radboud University Medical Center, Nijmegen, 6500 HB, The Netherlands. [83]Menzies Health Institute Queensland, Griffith University, Southport, Queensland, Australia. [84]Diamantina Institute, University of Qeensland, Brisbane, Queensland, 4072, Australia. [85]QIMR Berghofer Medical Research Institute, Brisbane, Queensland, 4006, Australia. [86]Wellcome Trust Sanger Institute, Wellcome Genome Campus, Hinxton, Cambridge CB10 1SA, UK. [87]British Heart Foundation, Cambridge Centre of Excellence, Department of Medicine, University of Cambridge, Cambridge CB2 0QQ, UK. [88]Department of Genetics, Center for Molecular Medicine, University Medical Center Utrecht, Utrecht, 3584 CX, The Netherlands. [89]Department of Vascular Surgery, Division of Surgical Specialties, University Medical Center Utrecht, Utrecht, 3584 CX, The Netherlands. [90]Faculty of Pharmacy, Université de Montréal, Montreal, Quebec, H3T 1J4, Canada. [91]Department of Clinical Chemistry and Haematology, Division of Laboratory and Pharmacy, University Medical Center Utrecht, Utrecht, 3508 GA, The Netherlands. [92]Utrecht Institute for Pharmaceutical Sciences, Division Pharmacoepidemiology & Clinical Pharmacology, Utrecht University, Utrecht, 3508 TB, The Netherlands. [93]Department of Clinical Epidemiology, Leiden University Medical Center, Leiden, 2300 RC, The Netherlands. [94]Department of Nutrition and Dietetics, School of Health Science and Education, Harokopio University, Athens, 17671, Greece. [95]Division of Epidemiology & Community Health, School of Public Health, University of Minnesota, Minneapolis, Minnesota 55454, USA. [96]Department of Ophthalmology, Radboud University Medical Center, Nijmegen, 6500 HB, The Netherlands. [97]Centre for Cancer Genetic Epidemiology, Department of Public Health and Primary Care, University of Cambridge, Cambridge CB1 8RN, UK. [98]Institute of Cardiovascular Science, University College London, London WC1E 6JF, UK. [99]MRC Integrative Epidemiology Unit, School of Social & Community Medicine, University of Bristol, Bristol BS8 2BN, UK. [100]Fred Hutchinson Cancer Research Center, Public Health Sciences Division, Seattle, Washington 98109, USA. [101]Memorial Sloan Kettering Cancer Center, Department of Epidemiology and Biostatistics, New York, New York 10017, USA. [102]Centre for Cancer Genetic Epidemiology, Department of Oncology, University of Cambridge, Cambridge CB1 8RN, UK. [103]Department of Medicine, Oulu University Hospital, Oulu, 90029, Finland. [104]Research Unit of Internal Medicine, University of Oulu, Oulu, FI-90014, Finland. [105]Division of Epidemiology, Department of Medicine, Institute for Medicine and Public Health, Vanderbilt Genetics Institute, Vanderbilt University, Nashville, Tennessee 37203, USA. [106]Massachusetts General Hospital, Boston, Massachusetts 02114, USA. [107]Medical and Population Genetics Program, Broad Institute, Cambridge, Massachusetts 02141, USA. [108]Department of Epidemiology and Biostatistics, MRC-PHE Centre for Environment and Health, School of Public Health, Imperial College London, London W2 1PG, UK. [109]Department of Hygiene and Epidemiology, University of Ioannina Medical School, Ioannina, 45110, Greece. [110]Survey Research Center, Institute for Social Research, University of Michigan, Ann Arbor, Michigan 48104, USA. [111]Division of Statistical Genomics, Department of Genetics, Washington University School of Medicine, StLouis, Missouri 63108, USA. [112]CNR Institute of Clinical Physiology, Pisa, Italy. [113]Department of Clinical & Experimental Medicine, University of Pisa, Pisa, Italy. [114]Research Center on Epidemiology and Preventive Medicine, Department of Clinical and Experimental Medicine, University of Insubria, Varese, 21100, Italy. [115]Toulouse University School of Medicine, Toulouse, TSA 50032 31059, France. [116]Department of Medicine, Harvard University Medical School, Boston, Massachusetts 02115, USA. [117]University of Glasgow, Glasgow G12 8QQ, UK. [118]Institute of Molecular Medicine, The University of Texas Health Science Center at Houston, Houston, Texas 77030, USA. [119]Department of Clinical Sciences, Genetic and Molecular Epidemiology Unit, Lund University, Malmö, SE-20502, Sweden. [120]Department of Nutrition, Harvard School of Public Health, Boston, Massachusetts 02115, USA. [121]Department of Public Health and Clinical Medicine, Unit of Medicine, Umeå University, Umeå, 901 87, Sweden. [122]Department of Clinical Biochemistry, Rigshospitalet, Copenhagen University Hospital, Copenhagen, 2100, Denmark. [123]Department of Medical Sciences, University of Trieste, Trieste, 34137, Italy. [124]Division of Experimental Genetics, Sidra Medical and Research Center, Doha, 26999, Qatar. [125]Geriatrics, Department of Public Health, Uppsala University, Uppsala, 751 85, Sweden. [126]Carolina Population Center, University of North Carolina, Chapel Hill, North Carolina 27514, USA. [127]Department of Nutrition, Gillings School of Global Public Health, University of North Carolina, Chapel Hill, North Carolina 27514, USA. [128]Faculty of Medicine, University of Iceland, Reykjavik, 101, Iceland. [129]Icelandic Heart Association, Kopavogur, 201, Iceland. [130]Department of Medical Sciences, Molecular Epidemiology and Science for Life Laboratory, Uppsala University, Uppsala, 751 41, Sweden. [131]Department of Sociology, University of North Carolina, Chapel Hill, North Carolina 27514, USA. [132]Laboratory of Epidemiology and Population Sciences, National Institute on Aging, Intramural Research Program, National Institutes of Health, Bethesda, Maryland 20892, USA. [133]University of Exeter Medical School, University of Exeter, Exeter EX2 5DW, UK. [134]MRCHGU, Institute of Genetics and Molecular Medicine, University of Edinburgh, Edinburgh EH4 2XU, UK. [135]Biodemography of Aging Research Unit, Social Science Research Institute, Duke University, Durham, North Carolina 27708, USA. [136]Department of Public Health, University of Helsinki, Helsinki, FI-00014, Finland. [137]Institute for Molecular Medicine Finland (FIMM), University of Helsinki, Helsinki, FI-00014, Finland. [138]Department of Pediatrics, Haukeland University Hospital, Bergen, 5021, Norway. [139]KG Jebsen Center for Diabetes Research, Department of Clinical Science, University of Bergen, Bergen, 5020, Norway. [140]Department of Cardiology, Heart Center, Tampere University Hospital, Tampere, 33521, Finland. [141]Department of Clinical Chemistry, Fimlab Laboratories, Tampere, 33520, Finland. [142]Department of Clinical Chemistry, University of Tampere School of Medicine, Tampere, 33014, Finland. [143]Centre for Eye Research Australia, Royal Victorian Eye and Ear Hospital, University of Melbourne, Melbourne, Victoria, 3002, Australia. [144]Centre for Ophthalmology and Vision Science, Lions Eye Institute, University of Western Australia, Perth, Western Australia, 6009, Australia. [145]Menzies Research Institute Tasmania, University of Tasmania, Hobart, Tasmania, 7000, Australia. [146]Generation Scotland, Centre for Genomic and Experimental Medicine, University of Edinburgh, Edinburgh EH4 2XU, UK. [147]Musculoskeletal Research Programme, Division of Applied Medicine, University of Aberdeen, Aberdeen, AB25 2ZD, UK. [148]K.G. Jebsen Center for Genetic Epidemiology, Department of Public Health, NTNU, Norwegian University of Science and Technology, Trondheim, 7600, Norway. [149]AMC, Department of Vascular Medicine, Amsterdam, 1105 AZ, The Netherlands. [150]HUNT Research Centre, Department of Public Health and General Practice, Norwegian University of Science and Technology, Levanger, 7600, Norway. [151]Department of Neurology, Erasmus Medical Center, Rotterdam, 3015 GE, The Netherlands. [152]Department of Radiology, Erasmus Medical Center, Rotterdam, 3015 GE, The Netherlands. [153]Department of Medicine, Division of Cardiovascular Medicine, Stanford University School of Medicine, Stanford, California 94305, USA. [154]Department of Public Health & Clinical Medicine, Umeå University, Umeå, SE-90185, Sweden. [155]Research Unit Skellefteå, Skellefteå, SE-93141, Sweden. [156]Department of Genome Sciences, University of Washington, Seattle, Washington 98195, USA. [157]The Copenhagen City Heart Study, Frederiksberg Hospital, Frederiksberg, 2000, Denmark. [158]Department of Preventive Medicine, Keck School of Medicine of the University of California, Los Angeles, California 90089, USA. [159]USC Roski Eye Institute, Department of Ophthalmology, Keck School of Medicine of the University of Southern California, Los Angeles, California 90089, USA. [160]Center for Medical Genetics and Molecular Medicine, Haukeland University Hospital, Bergen, 5021, Norway. [161]National Institute of Public Health, University of Southern Denmark, Copenhagen, 1353, Denmark. [162]Steno Diabetes Center, Gentofte, 2820, Denmark. [163]Aalborg University, Aalborg, DK-9000, Denmark. [164]Research Center for Prevention and Health, Capital Region of Denmark, Glostrup, DK-2600, Denmark. [165]National Institute for Health and Welfare, Helsinki, FI-00271, Finland. [166]Department of Cardiology, Leiden University Medical Center, Leiden, 2333, The Netherlands. [167]The Interuniversity Cardiology Institute of the Netherlands, Utrecht, 2333, The Netherlands. [168]Department of Computational Medicine and Bioinformatics, University of Michigan, Ann Arbor, Michigan 48109, USA. [169]Department of Internal Medicine, University of Michigan, Ann Arbor, Michigan 48109, USA. [170]Division of Gastroenterology, University of Michigan, Ann Arbor, Michigan 48109, USA. [171]Department of Psychiatry, Brain Center Rudolf Magnus, University Medical Center Utrecht, Utrecht, 3584 CG, The Netherlands. [172]Department of Clinical Physiology, University of Tampere School of Medicine, Tampere, 33014, Finland. [173]Echinos Medical Centre, Echinos, Greece. [174]Oxford Centre for Diabetes, Endocrinology and Metabolism, Radcliffe Department of Medicine, University of Oxford, Oxford OX3 7LE, UK. [175]Oxford NIHR Biomedical Research Centre, Oxford University Hospitals Trust, Oxford OX3 7LE, UK. [176]UKCRC Centre of Excellence for Public Health Research, Queens University Belfast, Belfast BT12 6BJ, UK. [177]Netherlands Cancer Institute - Antoni van Leeuwenhoek hospital, Amsterdam, 1066 CX, The Netherlands. [178]Department of Restorative Dentistry, Periodontology and Endodontology, University Medicine Greifswald, Greifswald, 17475, Germany. [179]Foundation for Research in Health Exercise and Nutrition, Kuopio Research Institute of Exercise Medicine, Kuopio, 70100, Finland. [180]National Heart and Lung Institute, Imperial College London, Hammersmith Hospital Campus, London W12 0NN, UK. [181]Division of Public Health Sciences, Fred Hutchinson Cancer Research Center, Seattle, Washington 98109, USA. [182]German Center for Diabetes Research, München-Neuherberg, 85764, Germany. [183]Institute of Epidemiology II, Helmholtz Zentrum München - German Research Center for Environmental Health, Neuherberg, D-85764, Germany.

[184]Research Unit of Molecular Epidemiology, Helmholtz Zentrum München—German Research Center for Environmental Health, Neuherberg, D-85764, Germany. [185]Department of Psychiatry, and Division of Molecular Biology and Human Genetics, Department of Biomedical Sciences, Faculty of Medicine and Health Sciences, Stellenbosch University, Tygerberg, Western Cape, 7505, South Africa. [186]CHU Nantes, Service de Génétique Médicale, Nantes, 44093, France. [187]Institute of Clinical Medicine, Internal Medicine, University of Eastern Finland and Kuopio University Hospital, Kuopio, 70210, Finland. [188]Institute for Maternal and Child Health - IRCCS 'Burlo Garofolo', Trieste, 34137, Italy. [189]Institute of Biomedicine & Physiology, University of Eastern Finland, Kuopio, 70210, Finland. [190]Department of Genetics, University of North Carolina, Chapel Hill, North Carolina 27514, USA. [191]Department of Biostatistical Sciences and Center for Public Health Genomics, Wake Forest School of Medicine, Winston-Salem, North Carolina 27157, USA. [192]MRC Epidemiology Unit, University of Cambridge School of Clinical Medicine, Institute of Metabolic Science, Cambridge CB2 0QQ, UK. [193]Group Health Research Institute, Seattle, Washington 98101, USA. [194]Department of Health Services, University of Washington, Seattle, Washington 98101, USA. [195]Division of Endocrinology and Metabolism, Department of Internal Medicine, Taichung Veterans General Hospital, Taichung 407, Taiwan. [196]School of Medicine, National Yang-Ming University, Taipei 112, Taiwan. [197]School of Medicine, Chung Shan Medical University, Taichung 402, Taiwan. [198]Division of Preventive Medicine University of Alabama at Birmingham, Birmingham, Alabama 35205, USA. [199]Key Laboratory of Nutrition and Metabolism, Institute for Nutritional Sciences, Shanghai Institutes for Biological Sciences, Chinese Academy of Sciences, University of the Chinese Academy of Sciences, Shanghai, People's Republic of China, Shanghai, 200031, China. [200]Department of Medicine, Division of Cardiovascular Medicine, Stanford University School of Medicine, Palo Alto, California 94304, USA. [201]Department of Medicine, Boston University School of Medicine, Boston, Massachusetts 02118, USA. [202]Uppsala University, Uppsala, 75185, Sweden. [203]Department of Experimental Medicine, Rigshospitalet, Copenhagen, DK-2200, Denmark. [204]Division of Public Health Sciences, Wake Forest School of Medicine, Winston-Salem, North Carolina 27157, USA. [205]Division of Health Sciences, Warwick Medical School, Warwick University, Coventry CV4 7AL, UK. [206]Department of Psychiatry, Washington University, Saint Louis, Missouri 63110, USA. [207]Department of Molecular Epidemiology, German Institute of Human Nutrition Potsdam-Rehbruecke (DIfE), Nuthetal, 14558, Germany. [208]Westmead Millennium Institute of Medical Research, Centre for Vision Research and Department of Ophthalmology, University of Sydney, Sydney, New South Wales, 2022, Australia. [209]Department of Public Health and Primary Care, Leiden University Medical Center, Leiden, 2300 RC, The Netherlands. [210]Centre for Global Health Research, Usher Institute of Population Health Sciences and Informatics, University of Edinburgh, Edinburgh EH8 9AG, UK. [211]Department of Medicine I, Ludwig-Maximilians-Universität, Munich, 81377, Germany. [212]DZHK (German Centre for Cardiovascular Research), partner site Munich Heart Alliance, Munich, 80802, Germany. [213]Laboratory of Neurogenetics, National Institute on Aging, NIH, Bethesda, Maryland 20892, USA. [214]DZHK (German Centre for Cardiovascular Research), partner site Greifswald, Greifswald, 17475, Germany. [215]Institute of Clinical Chemistry and Laboratory Medicine, University Medicine Greifswald, Greifswald, 17475, Germany. [216]Department of Cardiology, Heart Center, Tampere University Hospital and School of Medicine, University of Tampere, Tampere, 33521, Finland. [217]Program in Personalized Medicine, Department of Medicine, University of Maryland School of Medicine, Baltimore, Maryland 21201, USA. [218]Department of Medicine, Tampere University Hospital, Tampere, 33521, Finland. [219]Center for Neurobehavioral Genetics, UCLA, Los Angeles, California 90095, USA. [220]Pat Macpherson Centre for Pharmacogenetics and Pharmacogenomics, Medical Research Institute, Ninewells Hospital and Medical School, Dundee DD1 9SY, UK. [221]Laboratory of Clinical Chemistry and Hematology, Division Laboratories and Pharmacy, University Medical Center Utrecht, Utrecht, 3584 CX, The Netherlands. [222]Laboratory of Experimental Cardiology, Division Heart & Lungs, University Medical Center Utrecht, Utrecht, 3584 CX, The Netherlands. [223]School of Women's and Infants' Health, The University of Western Australia, Perth, Western Australia, 6009, Australia. [224]University of Helsinki, Institute for Molecular Medicine (FIMM) and Diabetes and Obesity Research Program, Helsinki, FI00014, Finland. [225]University of Tartu, Estonian Genome Center, Tartu, Estonia, Tartu, 51010, Estonia. [226]School of Medicine, University of Split, Split, 21000, Croatia. [227]Center for Neurogenomics and Cognitive Research, Department Complex Trait Genetics, VU University, Amsterdam, 1081 HV, The Netherlands. [228]Neuroscience Campus Amsterdam, Department Clinical Genetics, VU Medical Center, Amsterdam, 1081 HV, The Netherlands. [229]Department of Clinical Physiology and Nuclear Medicine, Turku University Hospital, Turku, 20521, Finland. [230]Research Centre of Applied and Preventive Cardiovascular Medicine, University of Turku, Turku, 20520, Finland. [231]Centre for Non-Communicable Diseases, Karachi, Pakistan. [232]Department of Clinical Physiology and Nuclear Medicine, Kuopio University Hospital, Kuopio, 70210, Finland. [233]MRL, Merck & Co., Inc., Genetics and Pharmacogenomics, Boston, Massachusetts 02115, USA. [234]Department of Epidemiology, University of Washington, Seattle, Washington 98195, USA. [235]Department of Biobank Research, Umeå University, Umeå, SE-90187, Sweden. [236]Division of Cardiovascular Medicine, Brigham and Women's Hospital and Harvard Medical School, Boston, Massachusetts 02115, USA. [237]Department of Medicine, Faculty of Medicine, Université de Montréal, Montreal, Quebec, H3T 1J4, Canada. [238]Department of Public Health and Clinical Medicine, Unit of Family Medicine, Umeå University, Umeå, 90185, Sweden. [239]Department of Biostatistics and Epidemiology, Perelman School of Medicine, University of Pennsylvania, Philadelphia, Pennsylvania 19104, USA. [240]Division of Epidemiology & Community Health University of Minnesota, Minneapolis, Minnesota 55454, USA. [241]Duke University, Durham, North Carolina 27703, USA. [242]Saw Swee Hock School of Public Health, National University of Singapore, National University Health System, Singapore, Singapore. [243]Departement of Haematology, University of Cambridge, Cambridge CB2 OPT, UK. [244]Department of Vascular Medicine, AMC, Amsterdam, 1105 AZ, The Netherlands. [245]Department of Twin Research and Genetic Epidemiology, Kingís College London, London SE1 7EH, UK. [246]Alzheimer Scotland Dementia Research Centre, University of Edinburgh, Edinburgh EH8 9JZ, UK. [247]Department of Epidemiology and Biostatistics, VU University Medical Center, Amsterdam, 1007MB, The Netherlands. [248]Department of Molecular Cell Biology, Leiden University Medical Center, Leiden, 2333ZC, The Netherlands. [249]Department of Molecular Epidemiology, Leiden University Medical Center, Leiden, 2333ZC, The Netherlands. [250]College of Biomedical and Life Sciences, Cardiff University, Cardiff, CF14 4EP, UK. [251]MRC Integrative Epidemiology Unit, School of Social and Community Medicine, University of Bristol, Bristol BS8 2BN, UK. [252]Institute for Community Medicine, University Medicine Greifswald, Greifswald, 17475, Germany. [253]Center for Pediatric Research, Department for Women's and Child Health, University of Leipzig, Leipzig, 04103, Germany. [254]Division of Molecular Biology and Human Genetics, Department of Biomedical Sciences, Faculty of Medicine and Health Sciences, Stellenbosch University, Tygerberg, Western Cape, 7505, South Africa. [255]Department of Gerontology and Geriatrics, Leiden University Medical Center, Leiden, 2333, The Netherlands. [256]Anogia Medical Centre, Anogia, Greece. [257]Centre for Vascular Prevention, Danube-University Krems, Krems, 3500, Austria. [258]Dasman Diabetes Institute, Dasman, 15462, Kuwait. [259]Diabetes Research Group, King Abdulaziz University, Jeddah, 21589, Saudi Arabia. [260]Department of Psychiatry, Dalhousie University, Halifax B3H 4R2, Canada. [261]University of Amsterdam, Department of Brain & Cognition, Amsterdam, 1018 WS, The Netherlands. [262]Department of Obstetrics and Gynecology, Institute for Medicine and Public Health, Vanderbilt Genetics Institute, Vanderbilt University, Nashville, Tennessee 37203, USA. [263]MRL, Merck & Co., Inc., Cardiometabolic Disease, Kenilworth, New Jersey 07033, USA. [264]Institute of Cellular Medicine, The Medical School, Newcastle University, Newcastle NE2 4HH, UK. [265]Department of Biostatistics, Boston University School of Public Health, Boston, Massachusetts 02118, USA. [266]Departments of Epidemiology & Medicine, Diabetes Translational Research Center, Fairbanks School of Public Health & School of Medicine, Indiana University, Indiana 46202, USA. [267]Department of Physiology and Biophysics, University of Mississippi Medical Center, Jackson, Mississippi 39216, USA. [268]Danish Diabetes Academy, Odense, 5000, Denmark. [269]Department of Public Health, Aarhus University, Aarhus, 8000, Denmark. [270]Memorial University, Faculty of Medicine, Discipline of Genetics, St. John's, Newfoundland A1B 3V6, Canada. [271]GlaxoSmithKlein, King of Prussia, Pennsylvania 19406, USA. [272]Department of Clinical Sciences, Quantitative Biomedical Research Center, Center for the Genetics of Host Defense, University of Texas Southwestern Medical Center, Dallas, Texas 75390, USA. [273]Department of Human Genetics, University of Michigan, Ann Arbor, Michigan 48109, USA. [274]Department of Public Health Sciences, Institute for Personalized Medicine, the Pennsylvania State University College of Medicine, Hershey, Pennsylvania 17033, USA. [275]Department of Epidemiology and Carolina Center of Genome Sciences, Chapel Hill, North Carolina 27514, USA. [276]Department of Neurology, Boston University School of Medicine, Boston, Massachusetts 02118, USA. [277]Department of Epidemiology Research, Statens Serum Institut, Copenhagen, 2200, Denmark. [278]Li Ka Shing Centre for Health Information and Discovery, The Big Data Institute, University of Oxford, Oxford OX3 7BN, UK. [279]The Mindich Child Health and Development Institute, Ichan School of Medicine at Mount Sinai, New York, New York 10069, USA. [280]Departments of Pediatrics and Genetics, Harvard Medical School, Boston, Massachusetts 02115, USA. [281]Princess Al-Jawhara Al-Brahim Centre of Excellence in Research of Hereditary Disorders (PACER-HD), King Abdulaziz University, Jeddah, 21589, Saudi Arabia.
*These authors contributed equally to this work.
†A list of members and their affiliations appears in the Supplementary Information.
§These authors jointly supervised this work.

METHODS

Study design and participants. The discovery cohort consisted of 147 studies comprising 458,927 adult individuals of the following ancestries: (1) European descent ($n = 381{,}625$); (2) African ($n = 27{,}494$); (3) South Asian ($n = 29{,}591$); (4) East Asian ($n = 8{,}767$); (5) Hispanic ($n = 10{,}776$) and (6) Saudi Arabian ($n = 695$). All participating institutions and coordinating centres approved this project, and informed consent was obtained from all subjects. Discovery meta-analysis was carried out in each ancestry group (except the Saudi Arabian) separately as well as in the All group. Validation was undertaken in individuals of European ancestry only (Supplementary Tables 1–3). Conditional analyses were undertaken only in the European descent group (106 studies, $n = 381{,}625$). The SNPs we identify are available from the NCBI dbSNP database of short genetic variations (https://www.ncbi.nlm.nih.gov/projects/SNP/). No statistical methods were used to predetermine sample size. The experiments were not randomized and the investigators were not blinded to allocation during experiments and outcome assessment.

Phenotype. Height (in centimetres) was corrected for age and the genomic principal components (derived from GWAS data, the variants with a MAF > 1% on ExomeChip (http://genome.sph.umich.edu/wiki/Exome_Chip_Design), or ancestry-informative markers available on the ExomeChip), as well as any additional study-specific covariates (for example, recruiting centre), in a linear regression model. For studies with non-related individuals, residuals were calculated separately by sex, whereas for family-based studies sex was included as a covariate in the model. Additionally, residuals for case/control studies were calculated separately. Finally, residuals were subject to inverse normal transformation.

Genotype calling. The majority of studies followed a standardized protocol and performed genotype calling using the designated manufacturer's software, which was then followed by zCall[30]. For ten studies participating in the Cohorts for Heart and Aging Research in Genomic Epidemiology (CHARGE) Consortium, the raw intensity data for the samples from seven genotyping centres were assembled into a single project for joint calling[11]. Study-specific quality-control measures of the genotyped variants was implemented before association analysis (Supplementary Tables 1–2).

Study-level statistical analyses. Individual cohorts were analysed separately for each ancestry population, with either RAREMETALWORKER (http://genome.sph.umich.edu/wiki/RAREMETALWORKER) or RVTEST (http://zhanxw.github.io/rvtests/), to associate inverse normal transformed height data with genotype data taking potential cryptic relatedness (kinship matrix) into account in a linear mixed model. These software are designed to perform score-statistics based rare-variant association analysis, can accommodate both unrelated and related individuals, and provide single-variant results and variance-covariance matrix. The covariance matrix captures linkage disequilibrium relationships between markers within 1 Mb, which is used for gene-level meta-analyses and conditional analyses[31]. Single-variant analyses were performed for both additive and recessive models (for the alternate allele).

Centralized quality control. The individual study data were investigated for potential existence of ancestry population outliers based on the 1000 Genome Project phase 1 ancestry reference populations. A centralized quality control procedure implemented in EasyQC[32] was applied to individual study association summary statistics to identify outlying studies: (1) assessment of possible problems in height transformation; (2) comparison of allele frequency alignment against 1000 Genomes Project phase 1 reference data to pinpoint any potential strand issues; and (3) examination of quantile–quantile plots per study to identify any problems arising from population stratification, cryptic relatedness and genotype biases. We excluded variants if they had a call rate <95%, Hardy–Weinberg equilibrium $P < 1 \times 10^{-7}$, or large allele frequency deviations from reference populations (>0.6 for all ancestry analyses and >0.3 for ancestry-specific population analyses). We also excluded from downstream analyses markers not present on the Illumina ExomeChip array 1.0, variants on the Y chromosome or the mitochondrial genome, indels, multiallelic variants, and problematic variants based on the Blat-based sequence alignment analyses. Meta-analyses were carried out in parallel by two different analysts at two sites.

Single-variant meta-analyses. *Discovery analyses.* We conducted single-variant meta-analyses in a discovery sample of 458,927 individuals of different ancestries using both additive and recessive genetic models (Extended Data Fig. 1 and Supplementary Tables 1–4). Significance for single-variant analyses was defined at an array-wide level ($P < 2 \times 10^{-7}$, Bonferroni correction for 250,000 variants). The combined additive analyses identified 1,455 unique variants that reached array-wide significance ($P < 2 \times 10^{-7}$), including 578 non-synonymous and splice-site variants (Supplementary Tables 5–7). Under the additive model, we observed a high genomic inflation of the test statistics (for example, a λ_{GC} of 2.7 in European ancestry studies for common markers, Extended Data Fig. 2 and Supplementary Table 8), although validation results (see below) and additional sensitivity analyses (see below) suggested that it is consistent with polygenic inheritance as opposed to population stratification, cryptic relatedness, or technical artefacts (Extended Data Fig. 2). The majority of these 1,455 association signals (1,241; 85.3%) were found in the European ancestry meta-analysis (85.5% of the discovery sample size) (Extended Data Fig. 2). Nevertheless, we discovered eight associations within five loci in our all-ancestry analyses that are driven by African studies (including one missense variant in the growth hormone gene *GH1* (rs151263636), Extended Data Fig. 3), three height variants found only in African studies, and one rare missense marker associated with height in South Asians only (Supplementary Table 7).

Genomic inflation and confounding. We observed a marked genomic inflation of the test statistics even after adequate control for population stratification (linear mixed model) arising mainly from common markers; λ_{GC} in European ancestry was 1.2 and 2.7 for all and common markers, respectively (Extended Data Fig. 2 and Supplementary Table 8). Such inflation is expected for a highly polygenic trait like height, and is consistent with our very large sample size[3,33]. To confirm this, we applied the recently developed linkage disequilibrium score regression method to our height ExomeChip results[34], with the caveats that the method was developed (and tested) with >200,000 common markers available. We restricted our analyses to 15,848 common variants (MAF ≥ 5%) from the European-ancestry meta-analysis, and matched them to pre-computed linkage disequilibrium scores for the European reference dataset[34]. The intercept of the regression of the χ^2 statistics from the height meta-analysis on the linkage disequilibrium score estimates that the inflation in the mean χ^2 is due to confounding bias, such as cryptic relatedness or population stratification. The intercept was 1.4 (s.e.m. = 0.07), which is small when compared to the λ_{GC} of 2.7. Furthermore, we also confirmed that the linkage disequilibrium score regression intercept is estimated upward because of the small number of variants on the ExomeChip and the selection criteria for these variants (that is, known GWAS hits). The ratio statistic of (intercept − 1)/(mean χ^2 − 1) is 0.067 (s.e.m. = 0.012), well within the normal range[34], suggesting that most of the inflation (~93%) observed in the height association statistics is due to polygenic effects (Extended Data Fig. 2).

Furthermore, to exclude the possibility that some of the observed associations between height and rare and low-frequency variants could be due to allele calling problems in the smaller studies, we performed a sensitivity meta-analysis with primarily European ancestry studies totalling >5,000 participants. We found very concordant effect sizes, suggesting that smaller studies do not bias our results (Extended Data Fig. 2).

Conditional analyses. The RAREMETAL R package[35] and the GCTA v1.24 (ref. 36) software were used to identify independent height association signals across the European descent meta-analysis results. RAREMETAL performs conditional analyses by using covariance matrices in order to distinguish true signals from those driven by linkage disequilibrium at adjacent known variants. First, we identified the lead variants ($P < 2 \times 10^{-7}$) based on a 1-Mb window centred on the most significantly associated variant and performed linkage disequilibrium pruning ($r^2 < 0.3$) to avoid downstream problems in the conditional analyses due to co-linearity. We then conditioned on the linkage disequilibrium-pruned set of lead variants in RAREMETAL and kept new lead signals at $P < 2 \times 10^{-7}$. The process was repeated until no additional signal emerged below the pre-specified *P*-value threshold. The use of a 1-Mb window in RAREMETAL can obscure dependence between conditional signals in adjacent intervals in regions of extended linkage disequilibrium. To detect such instances, we performed joint analyses using GCTA with the ARIC and UK ExomeChip reference panels, both of which comprise >10,000 individuals of European descent. With the exception of a handful of variants in a few genomic regions with extended linkage disequilibrium (for example, the HLA region on chromosome 6), the two pieces of software identified the same independent signals (at $P < 2 \times 10^{-7}$).

To discover new height variants, we conditioned the height variants found in our ExomeChip study on the previously published GWAS height variants[3] using the first release of the UK Biobank imputed dataset and regression methodology implemented in BOLT-LMM[37]. Because of the difference between the sample size of our discovery set ($n = 458{,}927$) and the UK Biobank (first release, $n = 120{,}084$), we applied a threshold of $P_{conditional} < 0.05$ to declare a height variant as independent in this analysis. We also explored an alternative approach based on approximate conditional analysis[36]. This latter method (SSimp) relies on summary statistics available from the same cohort, thus we first imputed summary statistics[38] for exome variants, using summary statistics from a previous study[3]. Conversely, we imputed the top variants from this study[3] using the summary statistics from the ExomeChip. Subsequently, we calculated effect sizes for each exome variant conditioned on the top variants of this study[3] in two ways. First, we conditioned the imputed summary statistics of the exome variant on the summary statistics of the top variants that fell within 5 Mb of the target ExomeChip variant. Second, we conditioned the summary statistics of the ExomeChip variant on the imputed summary statistics of the hits of this study[3]. We then selected the option that yielded a higher imputation quality. For poorly tagged variants ($\hat{r}^2 < 0.8$), we simply

used up-sampled HapMap summary statistics for the approximate conditional analysis. Pairwise SNP-by-SNP correlations were estimated from the UK10K data (TwinsUK[39] and ALSPAC[40] studies, $n = 3{,}781$).

Validation of the single-variant discovery results. Several studies, totalling 252,501 independent individuals of European ancestry, became available after the completion of the discovery analyses, and were thus used for validation of our experiment. We validated the single-variant association results in eight studies, totalling 59,804 participants, genotyped on the ExomeChip using RAREMETAL[31]. We sought additional evidence for association for the top signals in two independent studies in the UK (UK Biobank) and Iceland (deCODE), comprising 120,084 and 72,613 individuals, respectively. We used the same quality control and analytical methodology as described above. Genotyping and study descriptions are provided in Supplementary Tables 1–3. For the combined analysis, we used the inverse-variance-weighted fixed effects meta-analysis method using METAL[41]. Significant associations were defined as those with a combined meta-analysis (discovery and validation) $P_{combined} < 2 \times 10^{-7}$.

We considered 81 variants with suggestive association in the discovery analyses ($2 \times 10^{-7} < P_{discovery} \leq 2 \times 10^{-6}$). Of those 81 variants, 55 reached significance after combining discovery and replication results based on a $P_{combined} < 2 \times 10^{-7}$ (Supplementary Table 9). Furthermore, recessive modelling confirmed seven new independent markers with a $P_{combined} < 2 \times 10^{-7}$ (Supplementary Table 10). One of these recessive signals is due to a rare X-linked variant in the *AR* gene (rs137852591, MAF = 0.21%). Because of its frequency, we only tested hemizygous men (we did not identify homozygous women for the minor allele) so we cannot distinguish between a true recessive mode of inheritance or a sex-specific effect for this variant. To test the independence and integrate all height markers from the discovery and validation phase, we used conditional analyses and GCTA 'joint' modelling[36] in the combined discovery and validation set. This resulted in the identification of 606 independent height variants, including 252 non-synonymous or splice-site variants (Supplementary Table 11). If we consider only the initial set of lead SNPs with $P < 2 \times 10^{-7}$, we identified 561 independent variants. Of these 561 variants (selected without the validation studies), 560 have concordant direction of effect between the discovery and validation studies, and 548 variants have a $P_{validation} < 0.05$ (466 variants with $P_{validation} < 8.9 \times 10^{-5}$, Bonferroni correction for 561 tests), suggesting a very low false discovery rate (Supplementary Table 11).

Gene-based association meta-analyses. For the gene-based analyses, we applied two different sets of criteria to select variants, based on coding variant annotation from five prediction algorithms (PolyPhen2 HumDiv and HumVar, LRT, MutationTaster and SIFT)[42]. The mask labelled 'broad' included variants with a MAF < 0.05 that are nonsense, stop-loss, splice site, as well as missense variants that are annotated as damaging by at least one program mentioned above. The mask labelled 'strict' included only variants with a MAF < 0.05 that are nonsense, stop-loss, splice-site, as well as missense variants annotated as damaging by all five algorithms. We used two tests for gene-based testing, namely the SKAT[43] and VT[44] tests. Statistical significance for gene-based tests was set at a Bonferroni-corrected threshold of $P < 5 \times 10^{-7}$ (threshold for 25,000 genes and four tests). The gene-based discovery results were validated (same test and variants, when possible) in the same eight studies genotyped on the ExomeChip ($n = 59{,}804$ participants) that were used for the validation of the single-variant results (see above, and Supplementary Tables 1–3). Gene-based conditional analyses were performed in RAREMETAL.

Pleiotropy analyses. We accessed ExomeChip data from GIANT (BMI, waist:hip ratio), GLGC (total cholesterol, triglycerides, HDL-cholesterol, LDL-cholesterol), IBPC (systolic and diastolic blood pressure), MAGIC (glycaemic traits), REPROGEN (age at menarche and menopause), and DIAGRAM (type 2 diabetes) consortia. For coronary artery disease, we accessed 1000 Genomes Project-imputed GWAS data released by CARDIoGRAMplusC4D[45].

Pathway analyses. DEPICT (http://www.broadinstitute.org/mpg/depict/) is a computational framework that uses probabilistically defined reconstituted gene sets to perform gene set enrichment and gene prioritization[15]. For a description of gene set reconstitution, refer to refs 15, 46. In brief, reconstitution was performed by extending pre-defined gene sets (such as Gene Ontology terms, canonical pathways, protein-protein interaction subnetworks and rodent phenotypes) with genes co-regulated with genes in these pre-defined gene set using large-scale microarray-based transcriptomics data. In order to adapt the gene set enrichment part of DEPICT for ExomeChip data (https://github.com/RebeccaFine/height-ec-depict), we made two principal changes. First, because DEPICT for GWAS incorporates all genes within a given linkage disequilibrium block around each index SNP, we modified DEPICT to take as input only the gene directly impacted by the coding SNP. Second, we adapted the way DEPICT adjusts for confounders (such as gene length) by generating null ExomeChip association results using Swedish ExomeChip data (Malmö Diet and Cancer (MDC), All New Diabetics in Scania (ANDIS), and Scania Diabetes Registry (SDR) cohorts, $n = 11{,}899$) and randomly assigning phenotypes from a normal distribution before conducting association analysis (see Supplementary Information). For the gene set enrichment analysis of the ExomeChip data, we used significant non-synonymous variants statistically independent of known GWAS hits (and that were present in the null ExomeChip data; see Supplementary Information for details). For gene set enrichment analysis of the GWAS data, we used all loci with a non-coding index SNP and that did not contain any of the novel ExomeChip genes. In visualizing the analysis, we used affinity propagation clustering[47] to group the most similar reconstituted gene sets based on their gene memberships (see Supplementary Information). Within a 'meta-gene set', the best *P* value of any member gene set was used as representative for comparison. DEPICT for ExomeChip was written using the Python programming language and the code can be found at https://github.com/RebeccaFine/height-ec-depict.

We also applied the PASCAL (http://www2.unil.ch/cbg/index.php?title=Pascal) pathway analysis tool[16] to association summary statistics for all coding variants. In brief, the method derives gene-based scores (both SUM and MAX statistics) and subsequently tests for the over-representation of high gene scores in predefined biological pathways. We used standard pathway libraries from KEGG, REACTOME and BIOCARTA, and also added dichotomized (Z score > 3) reconstituted gene sets from DEPICT[15]. To accurately estimate SNP-by-SNP correlations even for rare variants, we used the UK10K data (TwinsUK[39] and ALSPAC[40] studies, $n = 3781$). To separate the contribution of regulatory variants from the coding variants, we also applied PASCAL to association summary statistics of only regulatory variants (20 kb upstream, gene body excluded) from a previous study[3]. In this way, we could classify pathways driven principally by coding, regulatory or mixed signals.

STC2 functional experiments. For the generation of STC2 mutants (R44L and M86I), wild-type STC2 cDNA contained in pcDNA3.1/Myc-His(−) (Invitrogen)[23] was used as a template. Mutagenesis was carried out using Quickchange (Stratagene), and all constructs were verified by sequence analysis. Recombinant wild-type STC2 and variants were expressed in human embryonic kidney (HEK) 293T cells (293tsA1609neo, ATCC CRL-3216) maintained in high-glucose DMEM supplemented 10% fetal bovine serum, 2 mM glutamine, nonessential amino acids, and gentamicin. The cells are routinely tested for mycoplasma contamination. Cells (6×10^6) were plated onto 10-cm dishes and transfected 18 h later by calcium phosphate co-precipitation using 10 μg plasmid DNA. Medium was collected 48 h after transfection, cleared by centrifugation, and stored at −20 °C until use. Protein concentrations (58–66 nM) were determined by TRIFMA using antibodies described previously[23]. PAPP-A was expressed stably in HEK293T cells as previously reported[48]. Expressed levels of PAPP-A (27.5 nM) were determined by a commercial ELISA (AL-101, Ansh Labs).

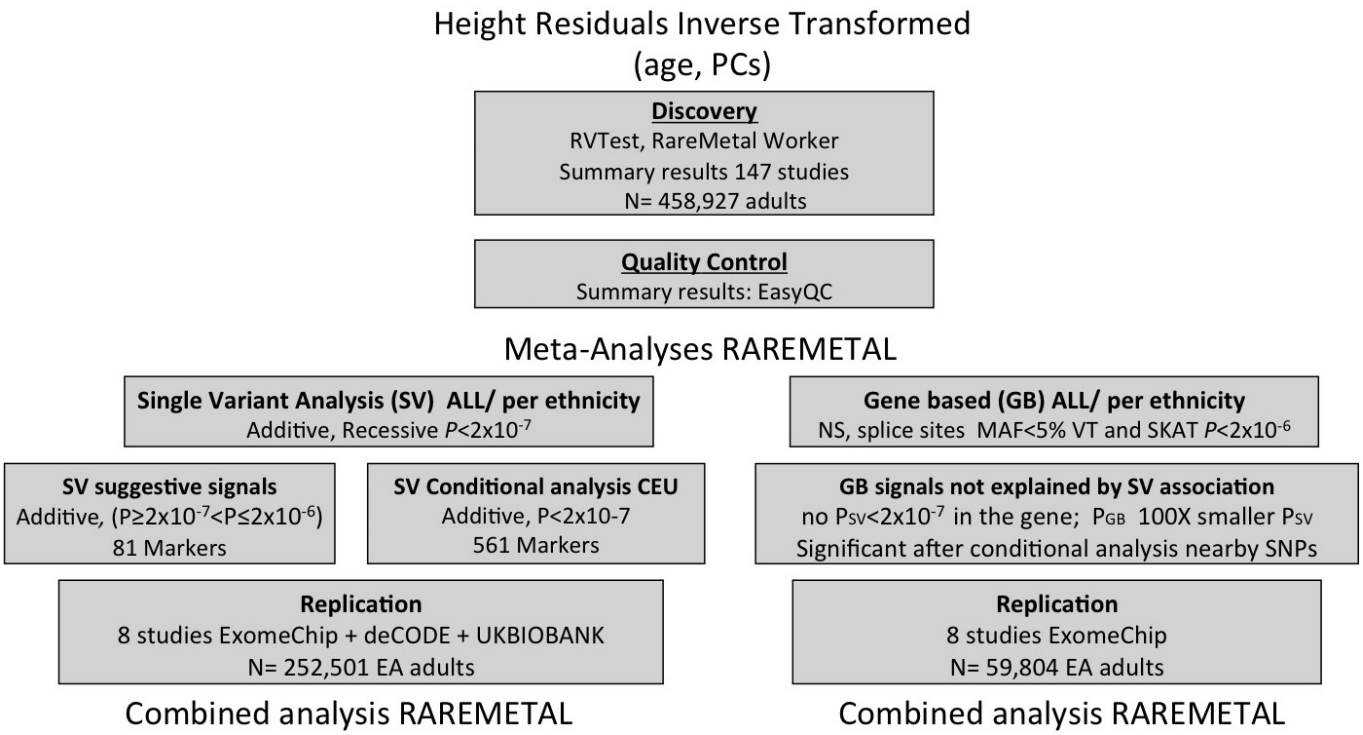
Culture supernatants containing wild-type STC2 or variants were adjusted to 58 nM, added an equal volume of culture supernatant containing PAPP-A corresponding to a 2.1-fold molar excess, and incubated at 37 °C. Samples were taken at 1, 2, 4, 6, 8, 16, and 24 h and stored at −20 °C.

Specific proteolytic cleavage of ^{125}I-labeled IGFBP-4 is described in detail elsewhere[49]. In brief, the PAPP-A–STC2 complex mixtures were diluted (1:190) to a concentration of 72.5 pM PAPP-A and mixed with pre-incubated ^{125}I-IGFBP4 (10 nM) and IGF-1 (100 nM) in 50 mM Tris-HCl, 100 mM NaCl, 1 mM $CaCl_2$. Following 1 h incubation at 37 °C, reactions were terminated by the addition of SDS–PAGE sample buffer supplemented with 25 mM EDTA. Substrate and co-migrating cleavage products were separated by 12% non-reducing SDS–PAGE and visualized by autoradiography using a storage phosphor screen (GE Healthcare) and a Typhoon imaging system (GE Healthcare). Band intensities were quantified using ImageQuant TL 8.1 software (GE Healthcare).

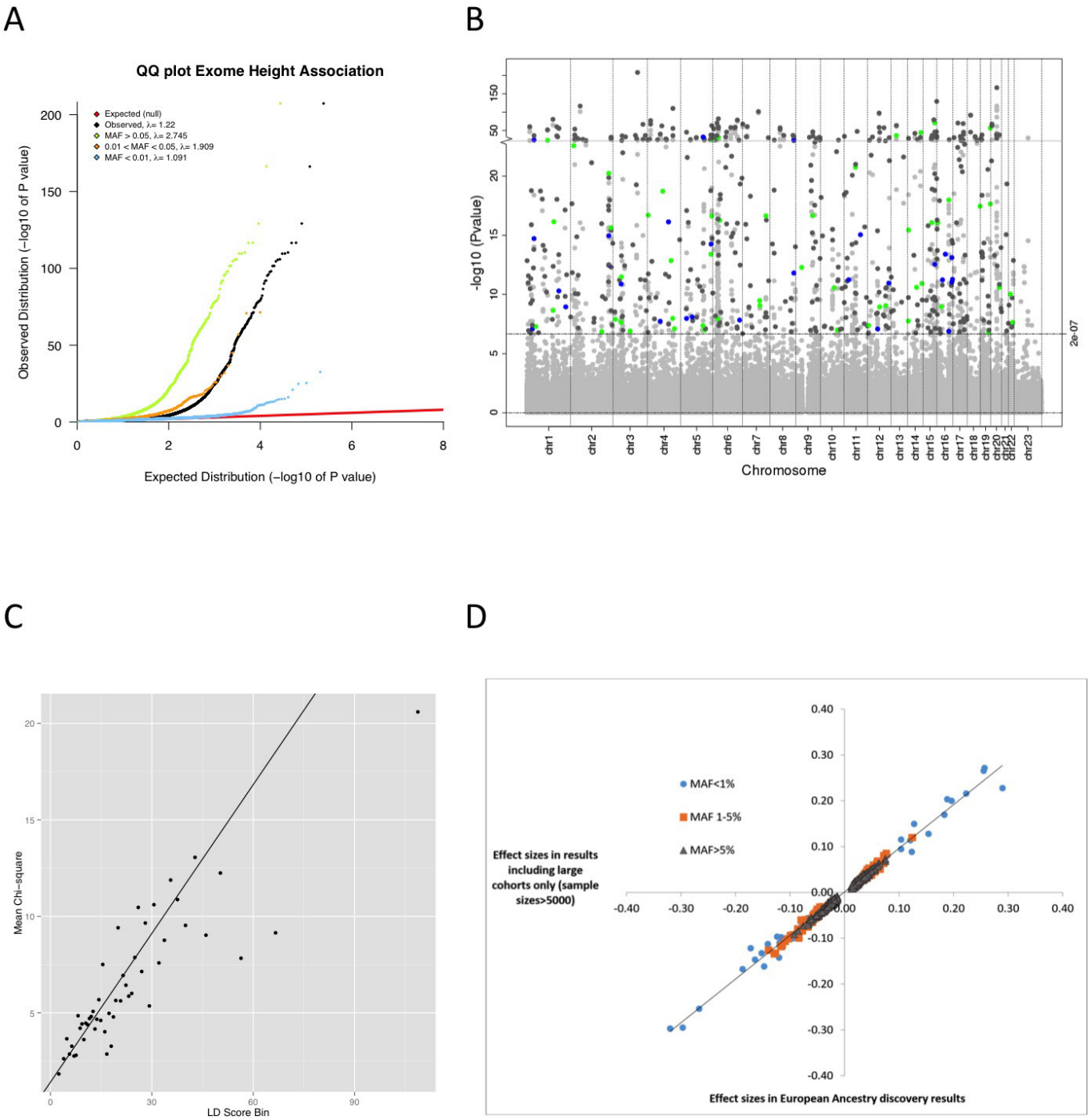
STC2 and covalent complexes between STC2 and PAPP-A were blotted onto PVDF membranes (Millipore) following separation by 3–8% SDS–PAGE. The membranes were blocked with 2% Tween-20, and equilibrated in 50 mM Tris-HCl, 500 mM NaCl, 0.1% Tween-20; pH 9 (TST). For STC2, the membranes were incubated with goat polyclonal anti-STC2 (R&D systems, AF2830) at 0.5 μg ml^{-1} in TST supplemented with 2% skimmed milk for 1 h at 20 °C. For PAPP-A–STC2 complexes, the membranes were incubated with rabbit polyclonal anti-PAPP-A[50] at 0.63 μg ml^{-1} in TST supplemented with 2% skimmed milk for 16 h at 20 °C. Membranes were washed with TST and subsequently incubated with polyclonal rabbit anti-goat IgG[en rule]horseradish peroxidase (DAKO, P0449) or polyclonal swine anti-rabbit IgG[en rule]horseradish peroxidase (DAKO, P0217), respectively, diluted 1:2,000 in TST supplemented with 2% skimmed milk for 1 h at 20 °C. Following washing with TST, membranes were developed using enhanced chemiluminescence (ECL Prime, GE Healthcare). Images were captured using an ImageQuant LAS 4000 instrument (GE Healthcare).

Data availability. Summary genetic association results are available on the GIANT website (http://portals.broadinstitute.org/collaboration/giant/index.php/GIANT_consortium).

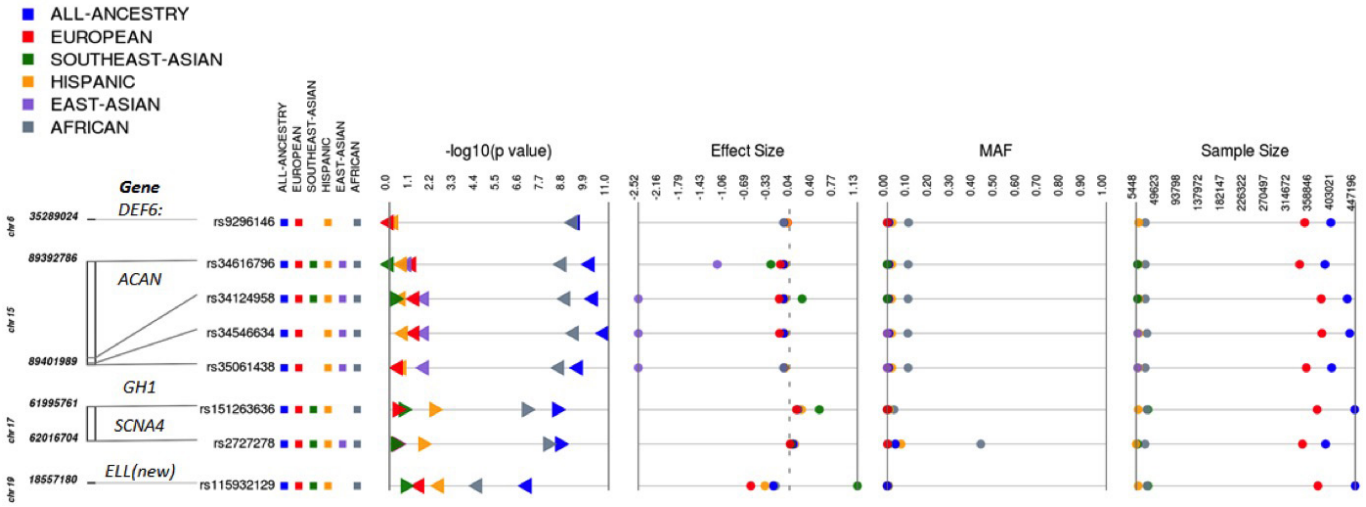
30. Goldstein, J. I. *et al.* zCall: a rare variant caller for array-based genotyping: genetics and population analysis. *Bioinformatics* **28,** 2543–2545 (2012).
31. Liu, D. J. *et al.* Meta-analysis of gene-level tests for rare variant association. *Nat. Genet.* **46,** 200–204 (2014).
32. Winkler, T. W. & Day, F. R. Quality control and conduct of genome-wide association meta-analyses. *Nat. Protocols* **9,** 1192–1212 (2014).
33. Yang, J. *et al.* Genomic inflation factors under polygenic inheritance. *Eur. J. Hum. Genet.* **19,** 807–812 (2011).
34. Bulik-Sullivan, B. K. *et al.* LD score regression distinguishes confounding from polygenicity in genome-wide association studies. *Nat. Genet.* **47,** 291–295 (2015).
35. Feng, S., Liu, D., Zhan, X., Wing, M. K. & Abecasis, G. R. RAREMETAL: fast and powerful meta-analysis for rare variants. *Bioinformatics* **30,** 2828–2829 (2014).
36. 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).
37. Loh, P. R. *et al.* Efficient Bayesian mixed-model analysis increases association power in large cohorts. *Nat. Genet.* **47,** 284–290 (2015).
38. Pasaniuc, B. *et al.* Fast and accurate imputation of summary statistics enhances evidence of functional enrichment. *Bioinformatics* **30,** 2906–2914 (2014).
39. Moayyeri, A., Hammond, C. J., Valdes, A. M. & Spector, T. D. Cohort Profile: TwinsUK and healthy ageing twin study. *Int. J. Epidemiol.* **42,** 76–85 (2013).
40. Boyd, A. *et al.* Cohort Profile: the 'children of the 90s'—the index offspring of the Avon Longitudinal Study of Parents and Children. *Int. J. Epidemiol.* **42,** 111–127 (2013).
41. Willer, C. J., Li, Y. & Abecasis, G. R. METAL: fast and efficient meta-analysis of genomewide association scans. *Bioinformatics* **26,** 2190–2191 (2010).
42. Purcell, S. M. *et al.* A polygenic burden of rare disruptive mutations in schizophrenia. *Nature* **506,** 185–190 (2014).
43. Wu, M. C. *et al.* Rare-variant association testing for sequencing data with the sequence kernel association test. *Am. J. Hum. Genet.* **89,** 82–93 (2011).
44. Price, A. L. *et al.* Pooled association tests for rare variants in exon-resequencing studies. *Am. J. Hum. Genet.* **86,** 832–838 (2010).
45. Nikpay, M. *et al.* A comprehensive 1,000 Genomes-based genome-wide association meta-analysis of coronary artery disease. *Nat. Genet.* **47,** 1121–1130 (2015).
46. Fehrmann, R. S. *et al.* Gene expression analysis identifies global gene dosage sensitivity in cancer. *Nat. Genet.* **47,** 115–125 (2015).
47. Frey, B. J. & Dueck, D. Clustering by passing messages between data points. *Science* **315,** 972–976 (2007).
48. Overgaard, M. T. *et al.* Expression of recombinant human pregnancy-associated plasma protein-A and identification of the proform of eosinophil major basic protein as its physiological inhibitor. *J. Biol. Chem.* **275,** 31128–31133 (2000).
49. Gyrup, C. & Oxvig, C. Quantitative analysis of insulin-like growth factor-modulated proteolysis of insulin-like growth factor binding protein-4 and -5 by pregnancy-associated plasma protein-A. *Biochemistry* **46,** 1972–1980 (2007).
50. Oxvig, C., Sand, O., Kristensen, T., Kristensen, L. & Sottrup-Jensen, L. Isolation and characterization of circulating complex between human pregnancy-associated plasma protein-A and proform of eosinophil major basic protein. *Biochim. Biophys. Acta* **1201,** 415–423 (1994).



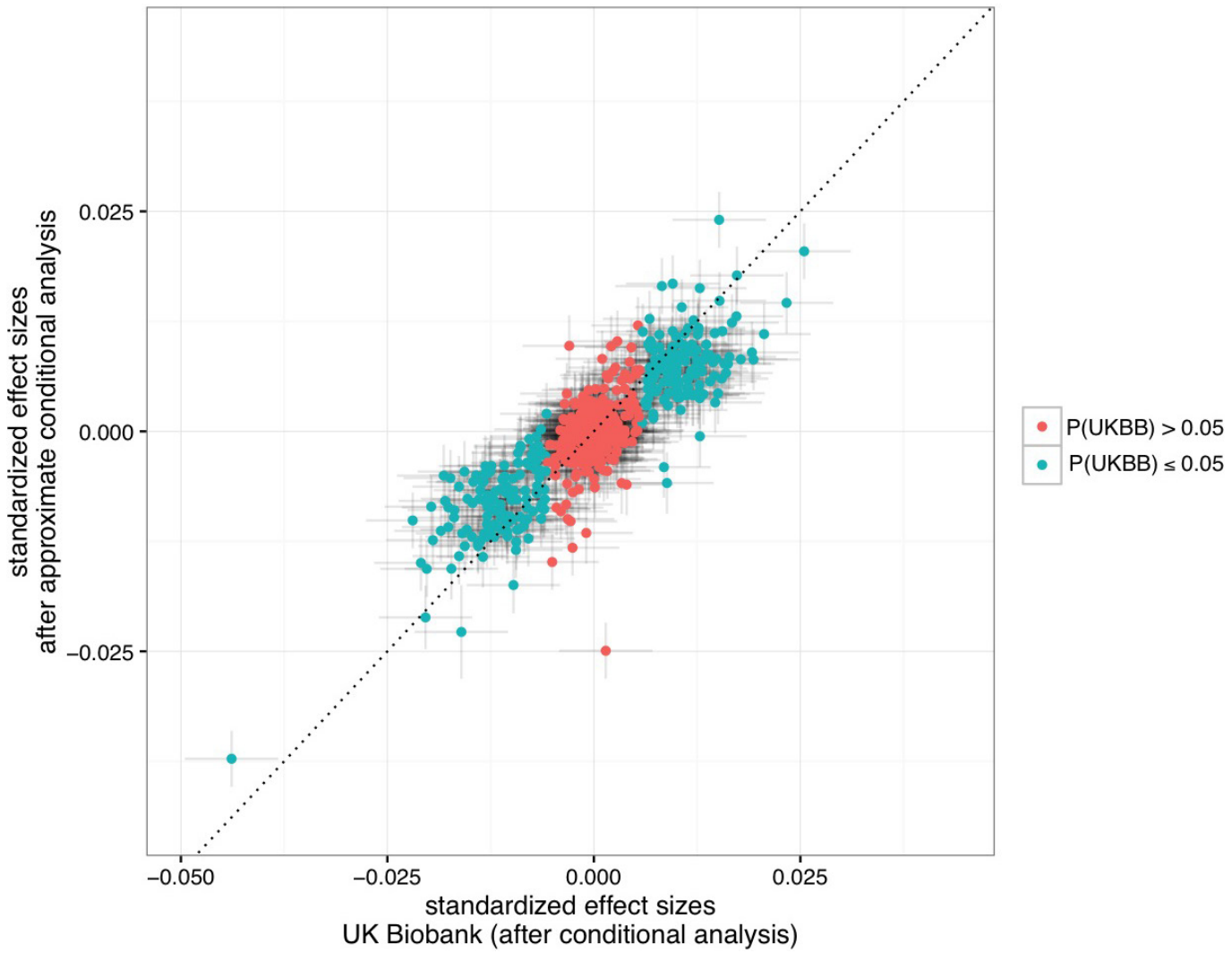
Extended Data Figure 1 | Flowchart of the GIANT ExomeChip height study design.



Extended Data Figure 2 | Height ExomeChip association results. **a**, Quantile–quantile plot of ExomeChip variants and their association to adult height under an additive genetic model in individuals of European ancestry. We stratified results on the basis of allele frequency. **b**, Manhattan plot of all ExomeChip variants and their association to adult height under an additive genetic model in individuals of European ancestry with a focus on the 553 independent SNPs, of which 469 have a MAF > 5% (grey), 55 have MAF between 1–5% (green), and 29 have a MAF < 1% (blue). **c**, Linkage disequilibrium (LD) score regression analysis for the height association results in European-ancestry studies. In the plot, each point represents a linkage disequilibrium score quantile, where the *x* axis of the point is the mean linkage disequilibrium score of variants in that quantile and the *y* axis is the mean χ^2 statistic of variants in that quantile. The linkage disequilibrium score regression slope of the black line is calculated using equation 1 in ref. 34, which is estimated upwards owing to the small number of common variants ($n = 15{,}848$) and the design of the ExomeChip. The linkage disequilibrium score regression intercept is 1.4, the λ_{GC} is 2.7, the mean χ^2 is 7.0, and the ratio statistic of (intercept − 1)/(mean χ^2 − 1) is 0.067 (s.e.m. = 0.012). **d**, Scatter plot comparison of the effect sizes for all variants that reached significance in the European-ancestry-discovery results ($n = 381{,}625$) and results including only studies with sample sizes of more than 5,000 individuals ($n = 241{,}453$).

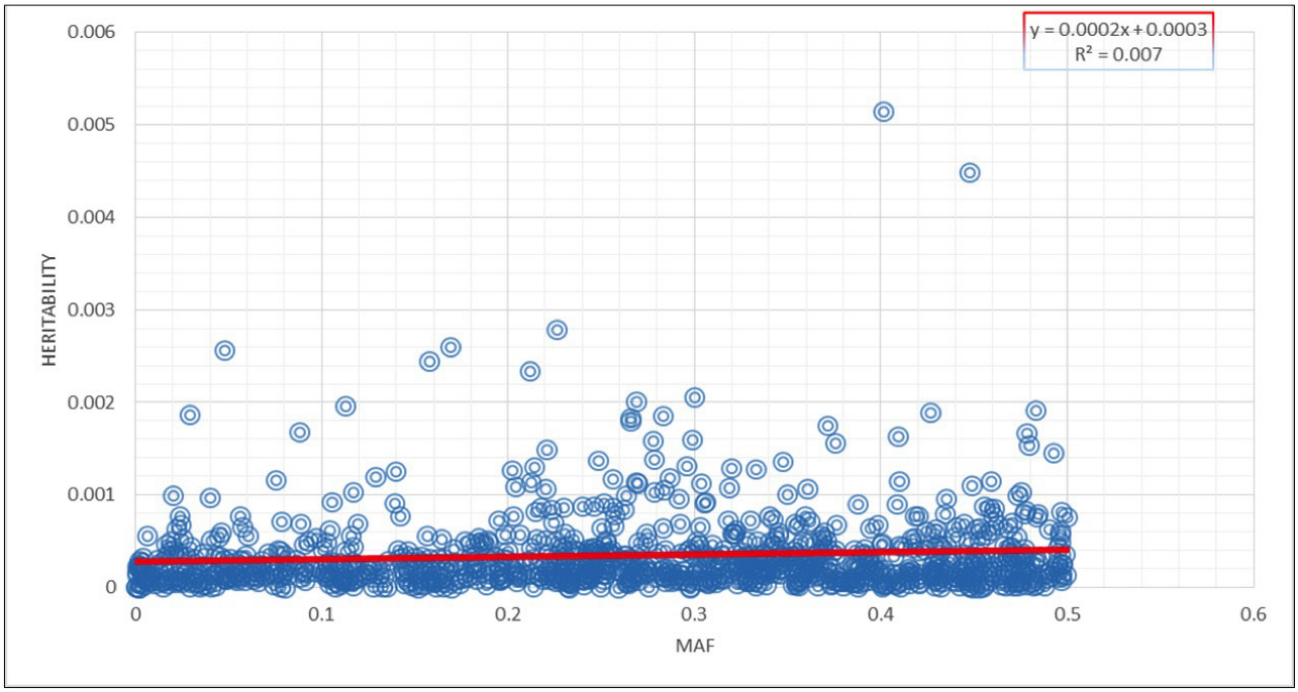


Extended Data Figure 3 | Height ExomeChip association results in African-ancestry populations. Among the all-ancestry results, we found eight variants for which the genetic association with height is mostly driven by individuals of African ancestry. The MAF of these variants is <1% (or monomorphic) in all ancestries except African ancestry. In individuals of African ancestry, the variants had allele frequencies between 9 and 40%.

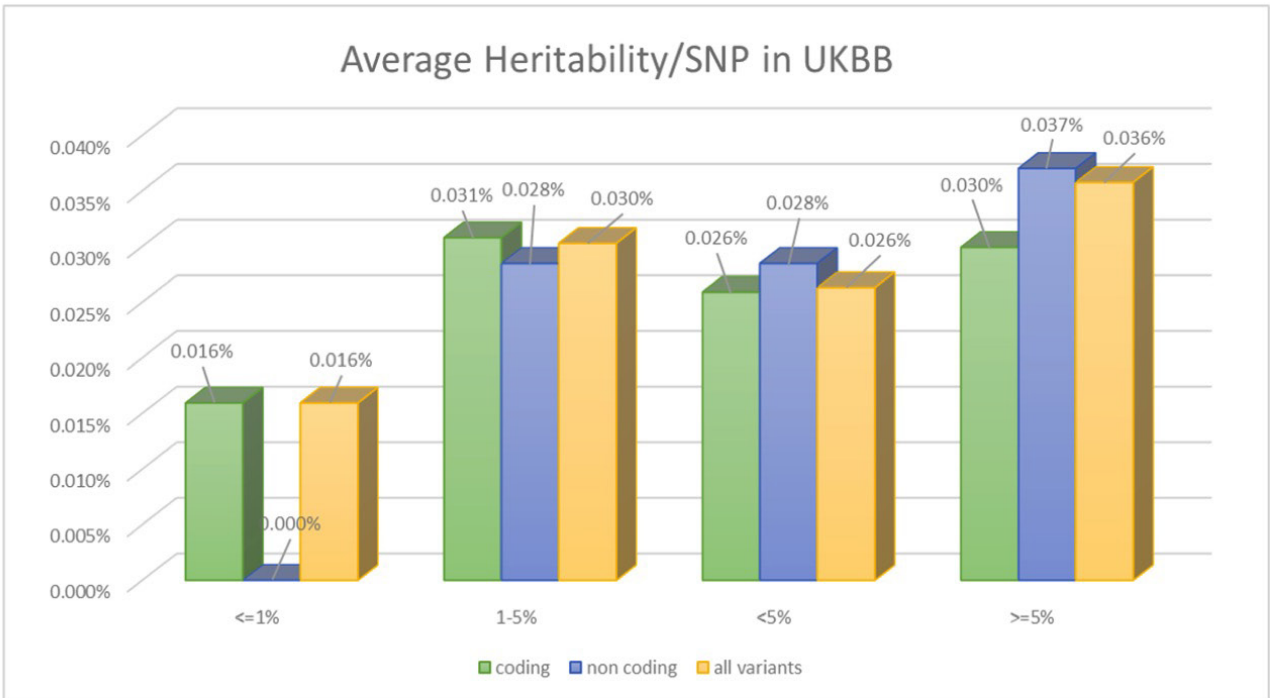


Extended Data Figure 4 | Concordance between direct conditional effect sizes using UK Biobank (*x* axis) and conditional analysis performed using a combination of imputation-based methodology and approximate conditional analysis (SSimp, *y* axis). The Pearson's correlation coefficient is $r = 0.85$. The dashed line indicates the identity line. The 95% confidence interval is indicated in both directions. Red, SNPs with $P_{cond} > 0.05$ in the UK Biobank; green, SNPs with $P_{cond} \leq 0.05$ in the UK Biobank.

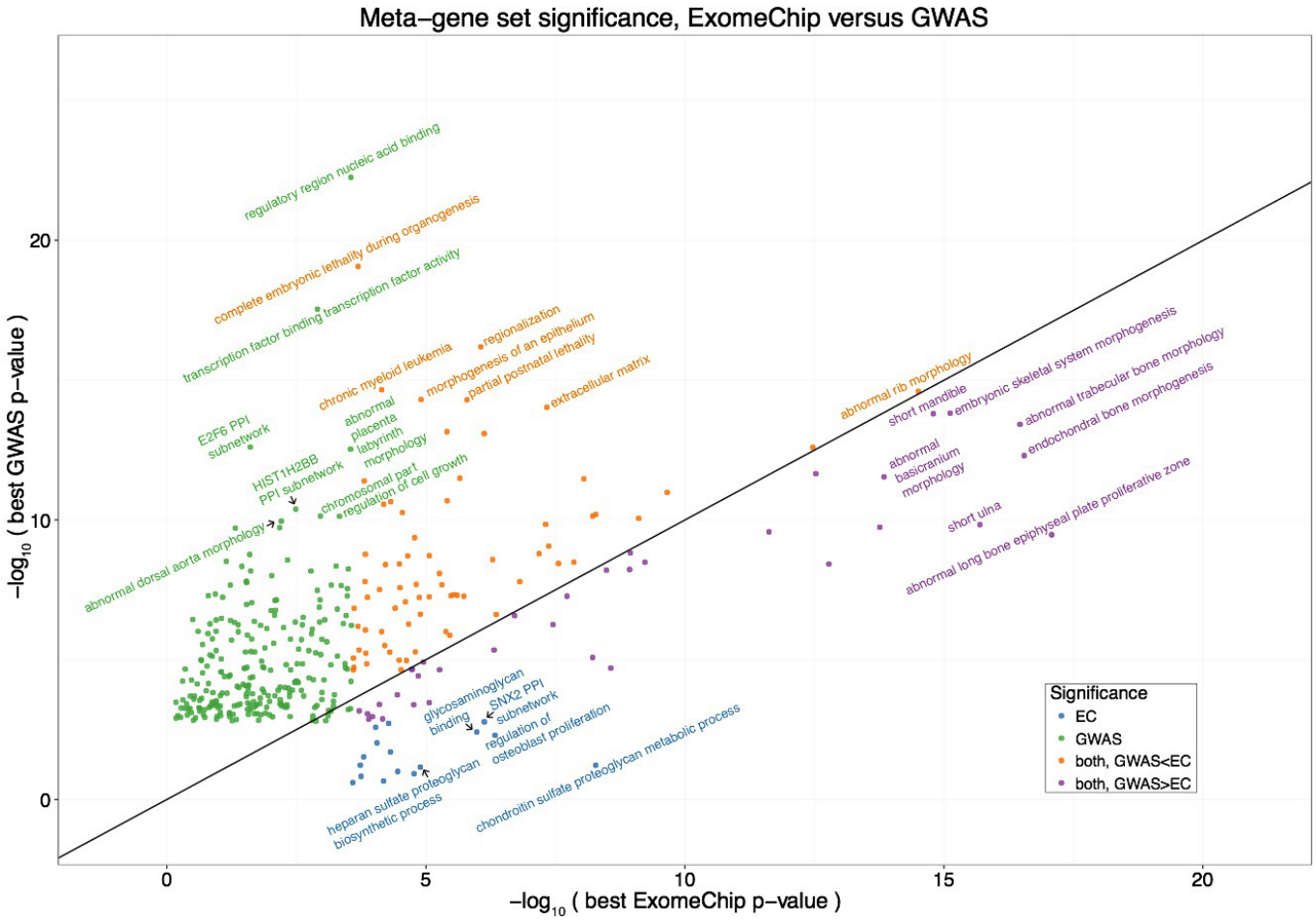
A



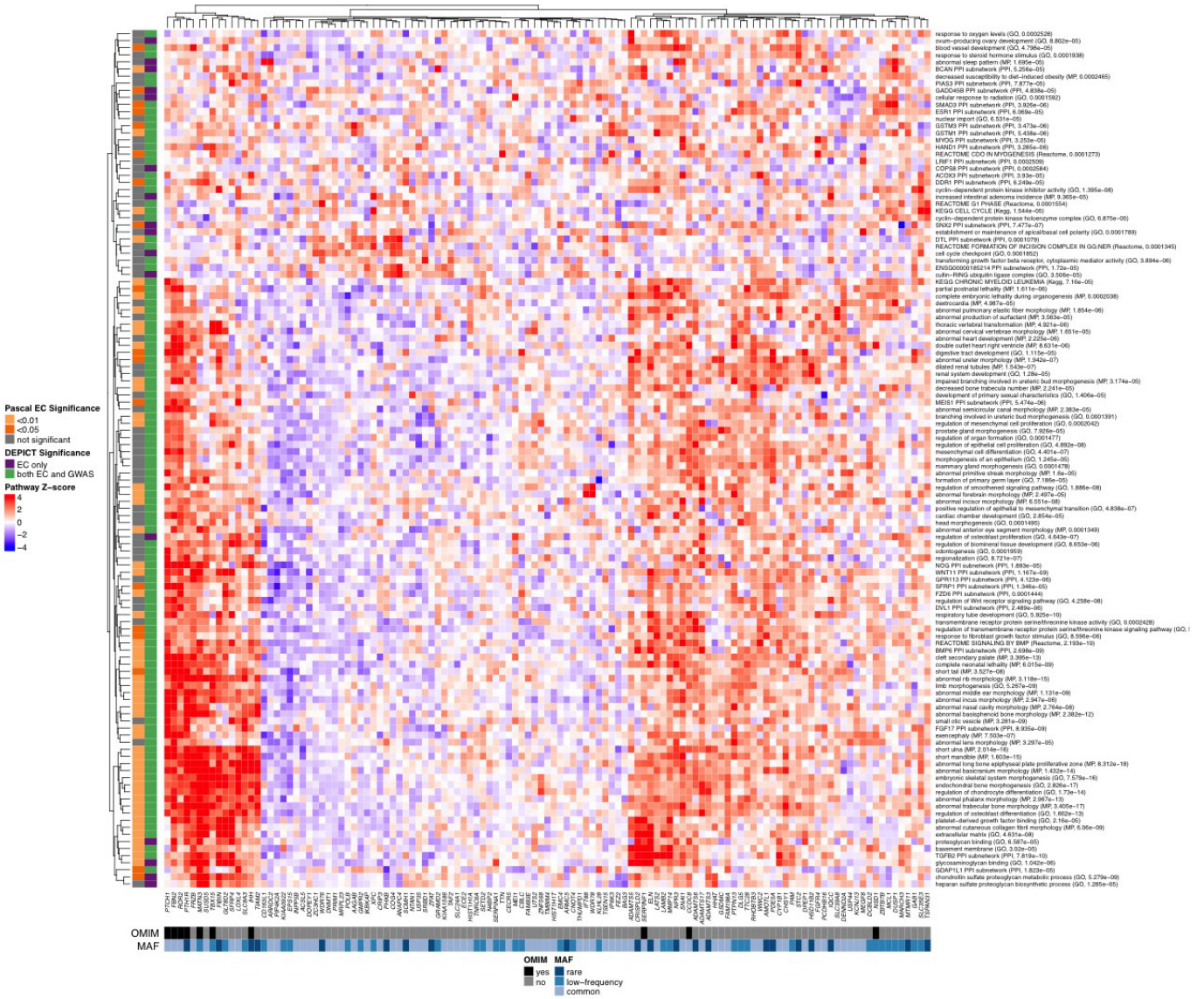
B



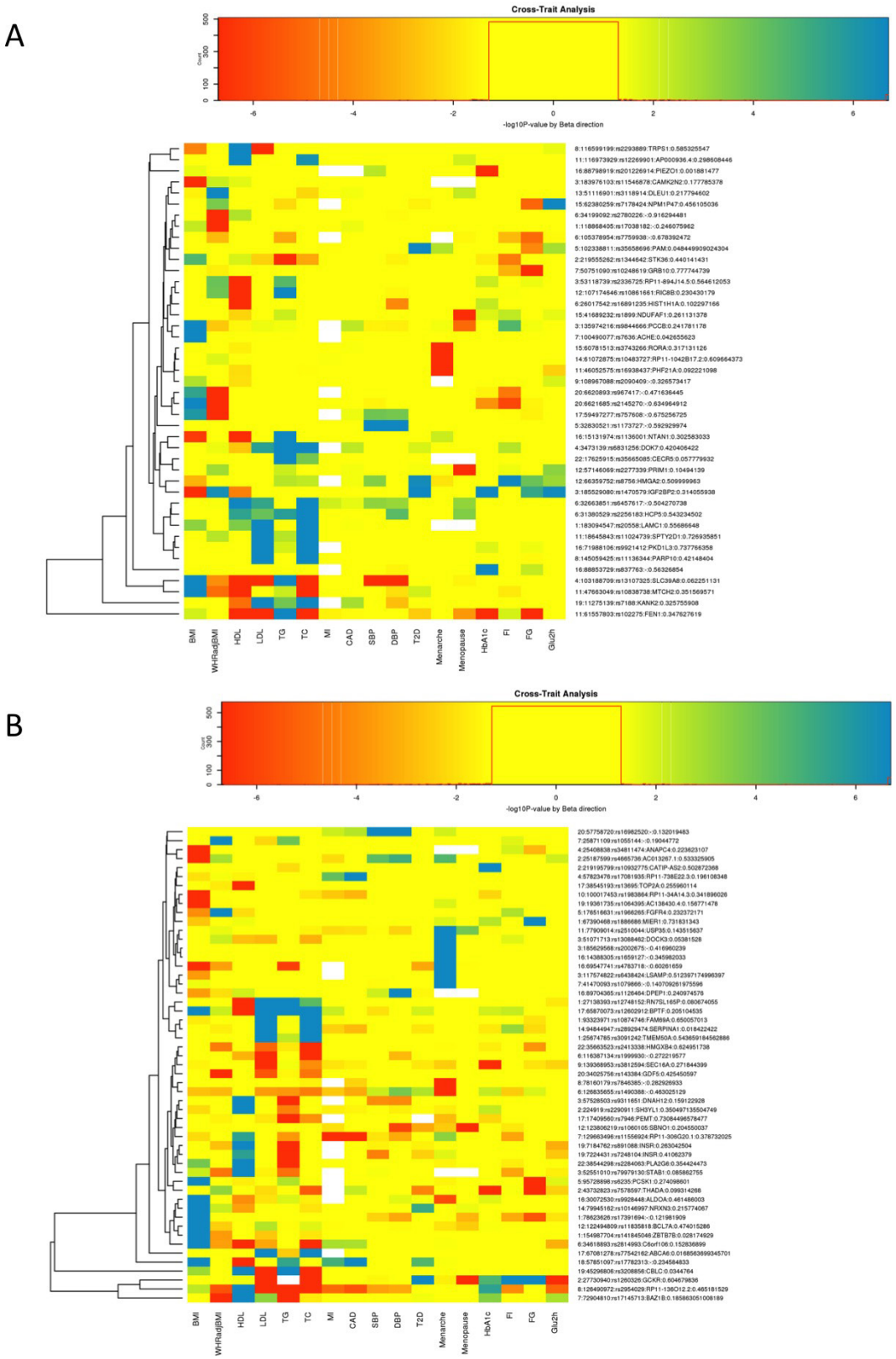
Extended Data Figure 5 | Heritability estimated for all known height variants in the first release of the UK Biobank dataset. a, We observed a weak but significant positive trend between MAF and heritability ($P = 0.012$). **b**, Average heritability explained per variant when stratifying the analyses by allele frequency or genomic annotation. For heritability estimations in UKBB, variants were pruned to $r^2 < 0.2$ in the 1000 Genomes Project dataset, and the heritability figures are based on $h^2 = 80\%$ for height.



Extended Data Figure 6 | Comparison of DEPICT gene set enrichment results based on coding variation from ExomeChip or non-coding variation from GWAS data. The *x* axis indicates the *P* value for enrichment of a given gene set using DEPICT adapted for ExomeChip (EC) data, where the input to DEPICT is the genes implicated by coding ExomeChip variants that are independent of known GWAS signals. The *y* axis indicates the *P* value for gene set enrichment using DEPICT, using as input the GWAS loci that do not overlap the coding signals. Each point represents a meta-gene set and the best *P* value for any gene set within the meta-gene set is shown. Only significant (false discovery rate < 0.01) gene set enrichment results are plotted. Colours correspond to whether the meta-gene set was significant for ExomeChip only (blue), GWAS only (green), both but more significant for ExomeChip (purple), or both but more significant for GWAS (orange), and the most significant gene sets within each category are labelled. A line is drawn at $x = y$ for ease of comparison.



Extended Data Figure 7 | Heat map showing entire DEPICT gene set enrichment results. This figure is analagous to Fig. 2. For any given square, the colour indicates how strongly the corresponding gene (shown on the *x* axis) is predicted to belong to the reconstituted gene set (*y* axis). This value is based on the *Z* score of the gene for gene set inclusion in DEPICT's reconstituted gene sets, where red indicates a higher *Z* score and blue indicates a lower one. The proteoglycan-binding pathway was uniquely implicated by coding variants (as opposed to common variants) by both DEPICT and the Pascal method. To visually reduce redundancy and increase clarity, we chose one representative 'meta-gene set' for each group of highly correlated gene sets based on affinity propagation clustering (see Methods and Supplementary Information). Heat map intensity and DEPICT *P* values correspond to the most significantly enriched gene set within the meta-gene set; meta-gene sets are listed with their database source. Annotations for the genes indicate whether the gene has OMIM annotation as underlying a disorder of skeletal growth (black and grey) and the MAF of the significant ExomeChip variant (shades of blue; if multiple variants, the lowest-frequency variant was kept). Annotations for the gene sets indicate if the gene set was also found significant for ExomeChip by the Pascal method (yellow and grey) and if the gene set was found significant by DEPICT for ExomeChip only or for both ExomeChip and GWAS (purple and green). GO, Gene Ontology; KEGG, Kyoto encyclopaedia of genes and genomes; MP, mouse phenotype in the Mouse Genetics Initiative; PPI, protein–protein interaction in the InWeb database.



Extended Data Figure 8 | Coding height variants are pleiotropic. **a**, **b**, Heat maps showing associations of the height variants to other complex traits; $-\log_{10}(P$ values) are oriented with beta effect direction for the alternate allele, white are missing values, yellow are non-significant ($P > 0.05$), green to blue shading for hits with positive beta in the other trait and P values between 0.05 and $< 2 \times 10^{-7}$ and orange to red shading for hits with negative beta in the other trait and P values between 0.05 and $< 2 \times 10^{-7}$. Short and tall labels are given for the minor alleles. Clustering is done by the complete linkage method with Euclidean distance measure for the loci. Clusters highlight SNPs that are more significantly associated with the same set of traits. **a** shows variants for which the minor allele is the height-decreasing allele. **b** shows variants for which the minor allele is the height-increasing allele.

Extended Data Table 1 | Rare variants associated with adult height

					Discovery (N up to 381,625)				Validation (N up to 252,501)				Combined (N up to 634,126)			
Variant	Chr:Pos	Ref/Alt	Gene	Annotation	AF	Beta	SE	P-value	AF	Beta	SE	P-value	AF	Beta	SE	P-value
rs150341307	1:32673514	G/C	*IQCC*	missense	0.002	-0.141	0.026	7.92E-08	0.004	-0.116	0.025	3.83E-06	0.003	-0.128	0.018	1.34E-12
rs143365597	1:41540902	G/A	*SCMH1*	missense	0.004	0.188	0.018	1.58E-25	0.006	0.169	0.024	9.42E-13	0.005	0.181	0.014	1.35E-36
rs114233776	1:41618297	G/A	*SCMH1*	missense	0.006	-0.119	0.015	1.92E-15	0.006	-0.11	0.019	1.32E-08	0.006	-0.116	0.012	1.80E-22
rs145659444	1:149902342	C/T	*MTMR11*	missense	0.007	0.067	0.015	4.16E-06	0.006	0.083	0.019	7.11E-06	0.007	0.073	0.012	3.03E-10
rs144712473	1:183495812	A/G	*SMG7*	missense	0.006	-0.094	0.014	4.97E-11	0.008	-0.067	0.017	8.94E-05	0.007	-0.083	0.011	1.61E-14
rs144673025	1:223178026	T/C	*DISP1*	missense	0.008	-0.078	0.013	1.11E-09	0.007	-0.086	0.018	1.22E-06	0.008	-0.081	0.011	1.27E-14
rs142036701	2:219924961	G/T	*IHH*	missense	0.001	-0.32	0.04	1.09E-15	0.003	-0.263	0.043	1.48E-09	0.002	-0.294	0.029	1.85E-23
rs147445258	2:220078652	C/T	*ABCB6*	missense	0.01	-0.086	0.012	3.43E-13	0.009	-0.064	0.018	4.40E-04	0.01	-0.079	0.01	2.47E-15
rs121434601	3:46939587	C/T	*PTH1R*	missense	0.003	0.154	0.023	1.30E-11	0.003	0.192	0.031	5.48E-10	0.003	0.168	0.019	1.14E-19
rs141374503	4:73179445	C/T	*ADAMTS3*	missense	0.003	-0.119	0.021	1.82E-08	0.004	-0.089	0.023	1.32E-04	0.004	-0.106	0.016	1.30E-11
rs149385790	4:120422407	T/G	*PDE5A*	missense	0.001	0.257	0.031	7.50E-17	0.005	0.19	0.033	1.28E-08	0.003	0.226	0.023	2.65E-23
rs146301345	5:32784907	G/A	*NPR3*	missense	0.003	0.128	0.022	1.05E-08	0.002	0.166	0.035	1.78E-06	0.003	0.139	0.019	7.91E-14
rs61736454	5:64766798	G/A	*ADAMTS6*	missense	0.002	-0.152	0.026	7.82E-09	0.002	-0.182	0.032	1.37E-08	0.002	-0.164	0.02	4.80E-16
rs78727187	5:127668685	G/T	*FBN2*	missense	0.006	0.183	0.015	2.47E-33	0.006	0.181	0.02	5.06E-20	0.006	0.182	0.012	1.47E-52
rs148833559	5:172755066	C/A	*STC2*	missense	0.001	0.29	0.037	5.69E-15	0.001	0.368	0.043	1.32E-17	0.001	0.323	0.028	1.15E-30
rs148543891	6:155450779	A/G	*TIAM2*	missense	0.003	-0.124	0.022	1.45E-08	0.001	-0.016	0.082	8.50E-01	0.003	-0.117	0.021	3.96E-08
rs41511151	7:73482987	G/A	*ELN*	missense	0.004	-0.086	0.018	2.63E-06	0.007	-0.061	0.019	1.51E-03	0.006	-0.074	0.013	2.31E-08
rs112892337	8:135614553	G/C	*ZFAT*	missense	0.004	0.196	0.019	4.42E-26	0.004	0.184	0.024	1.20E-14	0.004	0.191	0.015	6.12E-38
rs75596750	8:135622851	G/A	*ZFAT*	missense	0.001	0.255	0.036	1.54E-12	0.002	0.339	0.039	5.94E-18	0.002	0.293	0.027	2.05E-28
rs138273386	11:27016360	G/A	*FIBIN*	missense	0.004	-0.12	0.017	5.79E-12	0.005	-0.076	0.024	1.56E-03	0.004	-0.105	0.014	3.26E-14
rs138059525	11:94533444	G/A	*AMOTL1*	missense	0.009	-0.096	0.012	9.01E-16	0.007	-0.089	0.017	3.84E-07	0.008	-0.094	0.01	2.84E-21
rs147996581	12:58138971	G/A	*TSPAN31*	missense	0.003	-0.116	0.022	8.26E-08	0.001	-0.268	0.09	2.85E-03	0.003	-0.125	0.021	5.50E-09
rs13141	12:121756084	G/A	*ANAPC5*	missense	0.009	-0.082	0.012	1.09E-11	0.011	-0.105	0.016	1.44E-11	0.01	-0.091	0.01	1.45E-21
rs150494621	15:44153571	C/T	*WDR76*	missense	0.008	0.063	0.013	1.56E-06	0.014	0.054	0.015	3.42E-04	0.011	0.059	0.01	2.32E-09
rs141308595	15:89424870	G/T	*HAPLN3*	missense	0.001	-0.267	0.037	2.84E-13	0.002	-0.234	0.035	2.43E-11	0.002	-0.25	0.025	1.02E-22
rs141923065	16:31474091	A/G	*ARMC5*	splice acceptor	0.006	0.104	0.015	5.88E-12	0.013	0.057	0.018	1.16E-03	0.009	0.084	0.011	1.62E-13
rs34667348	16:47684830	C/A	*PHKB*	missense	0.005	0.121	0.016	3.96E-14	0.005	0.033	0.020	1.04E-01	0.005	0.088	0.013	3.43E-12
rs140385822	16:67470505	G/A	*HSD11B2*	missense	0.002	-0.148	0.028	1.27E-07	0.002	-0.124	0.035	3.38E-04	0.002	-0.139	0.022	1.97E-10
rs149615348	16:84900645	G/A	*CRISPLD2*	missense	0.007	-0.095	0.014	9.13E-12	0.008	-0.098	0.017	4.34E-09	0.008	-0.096	0.011	2.92E-19
rs148934412	16:84902472	G/A	*CRISPLD2*	missense	0.001	-0.297	0.04	7.75E-14	0.001	-0.317	0.058	3.49E-08	0.001	-0.304	0.033	2.36E-20
rs201226914	16:88798919	G/T	*PIEZO1*	missense	0.002	-0.187	0.027	5.27E-12	0.002	-0.241	0.043	1.99E-08	0.002	-0.202	0.023	8.68E-19
rs137852591	23:66941751	C/G	*AR*	missense	0.002	-0.304	0.061	7.05E-07	0.008	-0.333	0.058	7.12E-09	0.005	-0.319	0.042	2.67E-14

Table shows 32 missense or splice site variants with a MAF < 1% in European-ancestry participants that have $P_{combined} < 2 \times 10^{-7}$. The direction of the effect (beta, in units of s.d.) and effect allele frequency (AF) is given for the alternate (Alt) allele. Genomic coordinates are on build 37 of the human genome. For each variant, we provide the most severe annotation using the ENSEMBL Variant Effect Predictor (VEP) tool. N, sample size; Ref, reference allele; SE, s.e.m.

Extended Data Table 2 | Low-frequency variants associated with adult height

					Discovery (N up to 381,625)				Validation (N up to 252,501)				Combined (N up to 634,126)			
Variant	Chr:Pos	Ref/Alt	Gene	Annotation	AF	Beta	SE	P-value	AF	Beta	SE	P-value	AF	Beta	SE	P-value
rs41292521	1:51873967	G/A	*EPS15*	missense	0.020	0.045	0.008	5.07E-08	0.023	0.065	0.010	7.60E-11	0.021	0.053	0.006	2.56E-17
rs61730011	1:119427467	A/C	*TBX15*	missense	0.042	-0.059	0.006	1.61E-24	0.046	-0.056	0.007	4.19E-15	0.044	-0.058	0.005	2.79E-36
rs11580946	1:150551327	G/A	*MCL1*	missense	0.014	0.061	0.010	2.16E-09	0.015	0.085	0.012	7.86E-12	0.015	0.070	0.008	1.55E-19
rs141845046	1:154987704	C/T	*ZBTB7B*	missense	0.028	0.058	0.007	7.30E-17	0.025	0.061	0.010	4.46E-10	0.027	0.059	0.006	3.46E-25
rs79485039	1:180886140	C/T	*KIAA1614*	missense	0.026	0.034	0.007	1.41E-06	0.031	0.030	0.009	4.51E-04	0.028	0.033	0.006	2.63E-09
rs52826764	2:20205541	C/T	*MATN3*	missense	0.026	-0.071	0.007	2.67E-23	0.028	-0.084	0.010	6.60E-19	0.027	-0.076	0.006	3.74E-41
rs16859517	2:219949184	C/T	*NHEJ1*	intron	0.036	0.059	0.006	5.96E-21	0.036	0.064	0.008	1.12E-15	0.036	0.061	0.005	8.20E-37
rs16866412	2:179474668	G/A	*TTN*	missense	0.013	-0.053	0.010	1.35E-07	0.010	-0.019	0.015	2.15E-01	0.012	-0.042	0.008	3.44E-07
rs7571816	2:233077064	A/G	*DIS3L2*	intron	0.025	-0.060	0.007	2.35E-16	0.023	-0.079	0.010	2.58E-15	0.024	-0.066	0.006	6.46E-31
rs2229089	3:14214524	G/A	*XPC*	missense	0.031	-0.038	0.007	1.22E-08	0.035	-0.020	0.008	1.68E-02	0.033	-0.030	0.005	1.29E-08
rs76208147	3:47162886	C/T	*SETD2*	missense	0.019	0.048	0.009	2.24E-08	0.016	0.062	0.012	2.22E-07	0.018	0.053	0.007	1.65E-13
rs35713889	3:49162583	C/T	*LAMB2*	missense	0.039	0.043	0.006	3.28E-12	0.045	0.060	0.007	1.33E-16	0.041	0.050	0.005	3.49E-27
rs9838238	3:98600385	T/C	*DCBLD2*	missense	0.047	0.029	0.005	1.23E-07	0.051	0.027	0.007	5.62E-05	0.048	0.028	0.004	1.68E-12
rs11722554	4:5016883	G/A	*CYTL1*	missense	0.040	-0.049	0.006	2.01E-17	0.034	-0.057	0.009	6.68E-11	0.038	-0.052	0.005	1.86E-25
rs61730641	4:87730980	C/T	*PTPN13*	missense	0.015	-0.086	0.010	1.94E-19	0.016	-0.094	0.012	1.38E-15	0.016	-0.089	0.008	9.43E-32
rs116807401	4:135121721	T/C	*PABPC4L*	missense	0.017	0.065	0.009	1.39E-13	0.016	0.045	0.012	1.33E-04	0.017	0.058	0.007	7.54E-16
rs28925904	4:144359490	C/T	*GAB1*	missense	0.019	-0.048	0.008	1.04E-08	0.023	-0.036	0.010	3.24E-04	0.021	-0.043	0.006	4.29E-12
rs34343821	4:154557616	C/T	*KIAA0922*	missense	0.011	0.059	0.011	7.75E-08	0.015	0.056	0.012	5.75E-06	0.013	0.058	0.008	2.18E-12
rs35658696	5:102338811	A/G	*PAM*	missense	0.048	-0.025	0.005	3.76E-06	0.053	-0.031	0.007	8.47E-06	0.050	-0.027	0.004	1.63E-10
rs34821177	5:126250812	C/T	*MARCH3*	missense	0.036	0.034	0.006	4.25E-08	0.029	0.027	0.009	2.45E-03	0.034	0.032	0.005	1.67E-10
rs62623707	5:135288632	A/G	*LECT2*	missense	0.044	-0.030	0.006	1.02E-07	0.049	-0.024	0.007	4.77E-04	0.046	-0.027	0.005	1.36E-09
rs34471628	5:172196752	A/G	*DUSP1*	missense	0.036	0.048	0.006	4.00E-14	0.042	0.036	0.007	1.26E-06	0.039	0.043	0.005	1.93E-20
rs28932177	5:176637471	G/A	*NSD1*	missense	0.028	0.063	0.007	2.38E-17	0.027	0.065	0.009	2.62E-12	0.028	0.064	0.006	4.27E-30
rs78247455	5:176722005	G/A	*NSD1*	missense	0.023	-0.083	0.008	1.86E-26	0.025	-0.085	0.010	8.42E-18	0.024	-0.084	0.006	2.32E-41
rs7757648	6:30851933	G/A	*DDR1*	intron	0.013	-0.075	0.013	1.11E-08	0.011	-0.079	0.018	1.24E-05	0.012	-0.076	0.011	4.64E-13
rs34427075	6:34730395	C/T	*SNRPC*	synonymous	0.014	-0.117	0.010	9.21E-33	0.016	-0.139	0.012	9.59E-31	0.015	-0.126	0.008	3.45E-60
rs33966734	6:41903798	C/A	*CCND3*	stop_gained	0.013	-0.140	0.017	5.51E-17	0.011	-0.101	0.018	3.41E-08	0.012	-0.122	0.012	1.28E-22
rs17277546	7:99489571	G/A	*TRIM4*	3'UTR	0.049	0.034	0.005	3.28E-10	0.052	0.038	0.007	2.26E-07	0.050	0.035	0.004	1.40E-17
rs7636	7:100490077	G/A	*ACHE*	synonymous	0.043	-0.037	0.006	8.59E-10	0.035	-0.019	0.009	2.92E-02	0.040	-0.031	0.005	2.98E-10
rs17480616	7:135123060	G/C	*CNOT4*	missense	0.028	0.060	0.007	2.31E-17	0.030	0.054	0.009	5.04E-10	0.029	0.058	0.005	3.90E-26
rs3136797	8:42226805	C/G	*POLB*	missense	0.018	0.044	0.009	1.95E-06	0.021	0.026	0.010	1.30E-02	0.019	0.036	0.007	1.88E-07
rs11575580	9:34660864	C/T	*IL11RA*	missense	0.016	-0.064	0.009	5.20E-13	0.020	-0.030	0.011	4.42E-03	0.018	-0.050	0.007	4.01E-13
rs921122	9:95063947	C/T	*NOL8*	missense	0.039	0.041	0.009	2.56E-06	0.040	0.018	0.008	3.45E-02	0.040	0.029	0.006	3.33E-06
rs41274586	10:79580976	G/A	*DLG5*	missense	0.017	-0.058	0.009	2.72E-11	0.017	-0.076	0.012	5.15E-11	0.017	-0.065	0.007	7.66E-20
rs41291604	10:97919011	A/G	*ZNF518A*	missense	0.040	0.031	0.006	9.94E-08	0.040	0.022	0.008	3.05E-03	0.040	0.028	0.005	3.91E-09
rs71455793	11:65715204	G/A	*TSGA10IP*	missense	0.039	-0.058	0.006	1.82E-21	0.046	-0.072	0.007	1.41E-23	0.042	-0.064	0.005	1.52E-43
rs4072796	12:7548996	C/G	*CD163L1*	missense	0.035	0.034	0.006	4.11E-08	0.037	0.015	0.008	6.68E-02	0.036	0.027	0.005	1.87E-08
rs61743810	12:69140339	G/C	*SLC35E3*	missense	0.022	-0.047	0.008	1.13E-09	0.023	-0.036	0.010	5.11E-04	0.022	-0.043	0.006	1.29E-11
rs117801489	12:104408832	T/C	*GLT8D2*	missense	0.017	0.053	0.009	8.72E-10	0.028	0.062	0.010	5.82E-10	0.022	0.057	0.007	1.60E-17
rs2066674	13:50842259	G/A	*DLEU1*	intron	0.044	0.073	0.006	2.33E-37	0.041	0.084	0.008	7.02E-25	0.043	0.077	0.005	5.66E-57
rs17880989	14:23313633	G/A	*MMP14*	missense	0.027	0.041	0.007	1.72E-08	0.029	0.052	0.009	7.81E-09	0.028	0.045	0.006	3.27E-16
rs34354104	14:24707479	G/A	*GMPR2*	missense	0.048	0.045	0.005	3.67E-16	0.050	0.047	0.007	1.34E-11	0.049	0.046	0.004	2.13E-29
rs117295933	14:45403699	C/A	*KLHL28*	missense	0.016	-0.045	0.009	1.55E-06	0.025	-0.036	0.010	4.13E-04	0.020	-0.041	0.007	3.05E-09
rs41286548	14:70633411	C/T	*SLC8A3*	missense	0.021	-0.054	0.008	2.49E-11	0.026	-0.045	0.009	2.02E-06	0.023	-0.050	0.006	2.03E-16
rs28929474	14:94844947	C/T	*SERPINA1*	missense	0.018	0.124	0.009	1.39E-45	0.019	0.139	0.011	2.50E-34	0.019	0.130	0.007	1.72E-75
rs41286560	14:101349454	G/T	*RTL1*	missense	0.024	-0.050	0.007	1.17E-11	0.028	-0.033	0.009	2.12E-04	0.026	-0.044	0.006	2.50E-15
rs116858574	15:34520687	T/C	*EMC4*	missense	0.014	0.047	0.010	1.16E-06	0.014	0.028	0.012	2.19E-02	0.014	0.040	0.008	1.60E-07
rs34815962	15:72462255	C/T	*GRAMD2*	missense	0.019	0.073	0.009	8.72E-17	0.023	0.074	0.010	3.66E-13	0.021	0.073	0.007	1.28E-27
rs16942341	15:89388905	C/T	*ACAN*	synonymous	0.026	-0.129	0.007	4.30E-72	0.028	-0.146	0.009	1.08E-56	0.027	-0.135	0.006	3.79E-130
rs61733564	16:4812705	A/G	*ZNF500*	missense	0.032	0.056	0.007	8.61E-17	0.032	0.044	0.009	2.34E-07	0.032	0.051	0.005	2.89E-21
rs113388806	16:24804954	A/T	*TNRC6A*	missense	0.040	0.036	0.006	1.08E-09	0.047	0.041	0.008	1.65E-07	0.043	0.038	0.005	1.90E-15
rs8052655	16:67409180	G/A	*LRRC36*	missense	0.043	-0.054	0.006	1.08E-18	0.043	-0.055	0.008	3.91E-13	0.043	-0.054	0.005	6.40E-31
rs77542162	17:67081278	A/G	*ABCA6*	missense	0.017	0.049	0.010	2.17E-06	0.023	0.051	0.010	5.58E-07	0.020	0.050	0.007	5.57E-12
rs77169818	18:74980601	A/T	*GALR1*	missense	0.047	-0.048	0.006	3.60E-18	0.038	-0.035	0.008	3.64E-05	0.044	-0.044	0.005	5.11E-19
rs3208856	19:45296806	C/T	*CBLC*	missense	0.034	0.036	0.007	1.48E-07	0.034	0.021	0.008	1.19E-02	0.034	0.030	0.005	2.96E-08
rs4252548	19:55879672	C/T	*IL11*	missense	0.026	-0.114	0.007	1.02E-57	0.022	-0.101	0.010	2.28E-23	0.025	-0.110	0.006	5.32E-81
rs147110934	19:55993436	G/T	*ZNF628*	missense	0.021	-0.084	0.010	2.28E-18	0.022	-0.098	0.011	1.17E-18	0.022	-0.090	0.007	6.33E-34
rs77885044	22:28501414	C/T	*TTC28*	missense	0.012	-0.067	0.010	9.47E-11	0.017	-0.069	0.012	3.24E-09	0.014	-0.068	0.008	3.93E-19
rs147348682	22:42095658	T/G	*MEI1*	missense	0.025	0.041	0.007	2.25E-08	0.034	0.024	0.009	6.59E-03	0.029	0.034	0.006	3.70E-10

Table shows 59 variants (51 missense or nonsense) with minor allele frequency between 1 and 5% in participants of European ancestry that have $P_{combined} < 2 \times 10^{-7}$. For *TTN* rs16866412 and *NOL8* rs921122, the association is significant ($P < 2 \times 10^{-7}$) upon conditional analysis. The direction of the effect (beta, s.d. units) and effect allele frequency (AF) is given for the alternate (Alt) allele. For each variant, we provide the most severe annotation using the ENSEMBL Variant Effect Predictor (VEP) tool. N, sample size; Ref, reference allele; SE, s.e.m.