Rare and low-frequency coding variants alter human adult height

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

Correspondence to:

Joel N. Hirschhorn (joelh@broadinstitute.org)

Panos Deloukas (p.deloukas@qmul.ac.uk)

Guillaume Lettre (guillaume.lettre@umontreal.ca)

Summary: 149 words

Main text: 2,653 words

One figure, three tables, 10 extended data figures, 2 extended data tables

SUMMARY

Height is a highly heritable, classic polygenic trait with ~700 common associated variants identified so far through genome-wide association studies. Here, we report 83 height-associated coding variants with lower minor allele frequencies (range of 0.1-4.8%) and effects of up to 2 cm/allele (*e.g.* in *IHH*, *STC2*, *AR* and *CRISPLD2*), >10 times the average effect of common variants. In functional follow-up studies, rare height-increasing alleles of *STC2* (+1-2 cm/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 mutated in monogenic growth disorders and highlight new biological candidates (*e.g. ADAMTS3, IL11RA, NOX4*) and pathways (*e.g.* proteoglycan/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.

INTRODUCTION

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 ~20% of the heritability of height[3]. As is typical of complex traits and diseases, most of the height alleles discovered so far are common (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 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 few such 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], making it timely to assess whether and to what extent rare and low-frequency coding variation contributes 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 (83% coding with 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 ~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

52 complex human trait, such as adult height, and to discover and characterize new genes and
53 biological pathways implicated in human growth.
54

RESULTS

32 rare and 51 low-frequency coding variants associated with adult 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; technical details of the discovery and validation steps are in **Methods** (**Extended Data Figs 1-3**, **Supplementary Tables 1-11**). In total, we found 606 independent ExomeChip variants at array-wide significance ($P<2x10^{-7}$), including 252 non-synonymous or splice site variants **(Methods** and **Supplementary Table 11**). Focusing on non-synonymous or splice site variants with MAF <5%, our single-variant analyses identified 32 rare and 51 low-frequency height-associated variants (**Extended Data Tables 1-2**). To date, these 83 height variants (MAF range 0.1-4.8%) represent the largest set of validated rare and low-frequency coding variants associated with any complex human trait or disease. 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 of 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 evidences 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* (rs137852591, MAF=0.21%, $P_{combined}=2.7x10^{-14}$), in *CRISPLD2* (rs148934412, MAF=0.08%, $P_{combined}=2.4x10^{-20}$), in *IHH* (rs142036701, MAF=0.08%, $P_{combined}=1.9x10^{-23}$), and in *STC2* (rs148833559,

MAF=0.1%, $P_{combined}$=1.2x10^{-30}). Carriers of the rare *STC2* missense variant are ~2.1 cm taller than non-carriers, whereas carriers of the remaining three variants (or hemizygous men that carry the X-linked *AR*-rs137852591 rare allele) are ~2 cm shorter than non-carriers. In 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 -decreasing effects (48% vs. 52%, respectively) (**Fig. 1** and **Extended Data Tables 1-2**).

Coding variants in new and known height loci, and heritability explained

Many of the height-associated variants in this ExomeChip effort are located near common variants previously associated with height. Of the 83 rare and low-frequency non-synonymous variants, two 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 in the UK Biobank dataset and confirmed that 38 of these 47 variants were independent from 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 loci (**Supplementary Table 11**).

We used the UK Biobank dataset to estimate the contribution of the new height variants to heritability, which is h^2~80% 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 SNPs explain 23.3% of height heritability in the same dataset (vs. 4.1% by the new height variants identified in this ExomeChip study). We observed a modest positive association between MAF and heritability explained per variant (P=0.012, **Extended Data Fig. 5**), with each common variant explaining slightly more heritability than rare or low-frequency variants (0.036% vs. 0.026%, **Extended Data Fig. 5**).

Gene-based association results

To increase 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 $P<5x10^{-7}$ that harbor 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* – fell outside of the loci found by single-variant analyses and are implicated in human height for the first time.

Coding variants implicate biological pathways in human skeletal growth

Prior 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 more skeletal growth-specific pathways (chondrocyte biology, extracellular matrix (ECM), and skeletal development)[3]. We used two different methods, DEPICT[15] and PASCAL[16] (**Methods**), to perform pathway analyses using the ExomeChip results to test whether coding variants could either 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 applied DEPICT separately on (1) exome array-wide associated coding variants independent of known GWAS signals and (2) 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 (FDR) <1% (**Supplementary Tables 19-20**); similar analyses with PASCAL yielded 362 and 278 enriched gene sets (**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 Pearson's r = 0.583, $P<2.2\text{x}10^{-16}$; for PASCAL, Pearson's r=0.605, $P<2.2\text{x}10^{-16}$). However, some pathways showed stronger enrichment with either coding or non-coding genetic variation. In general, coding variants more strongly implicated pathways specific to skeletal growth (such as ECM and bone growth), while GWAS signals highlighted more global biological processes (such as transcription factor binding and embryonic size/lethality)(**Extended Data Fig. 6)**. The two significant gene sets identified by DEPICT and PASCAL that uniquely implicated coding variants were "BCAN protein protein interaction subnetwork" and "proteoglycan binding." 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 examined 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 is strongly implicated in a group of correlated pathways that include biology related to proteoglycans/glycosaminoglycans (**Fig. 2** and **Extended Data Fig. 7**). Seven of these 15 genes overlap a previously curated list of 277 genes annotated in OMIM as causing skeletal growth disorders[18]; genes in this small cluster are enriched for OMIM annotations relative to genes outside the cluster (odds ratio=27.6, Fisher's exact P=1.1x10^{-5}). As such, the remaining genes in this cluster may be strong candidates for harboring 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[19]; *LOXL4*, a lysyl oxidase expressed in cartilage[20]), modulate pathways known to affect skeletal growth (*FIBIN*, *SFRP4*)[21,22] or lead to increased body length when knocked out in mice (*SFRP4*)[23].

Functional characterization of rare STC2 variants

To begin exploring 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*. Over-expression of *STC2* diminishes growth in mice by covalent binding and inhibition of the proteinase PAPP-A, which specifically cleaves IGF binding protein-4 (IGFBP-4), leading to reduced levels of bioactive insulin-like growth factors

(**Fig. 3A**)[24]. Although there was no prior genetic evidence implicating *STC2* variation in human growth, the *PAPPA* and *IGFBP4* genes were both implicated in height GWAS[3], and rare mutations in *PAPPA2* cause severe short stature[25], emphasizing the likely relevance of this pathway in humans. The two *STC2* height-associated variants are rs148833559 (p.Arg44Leu, MAF=0.096%, $P_{discovery}$=5.7x10^{-15}) and rs146441603 (p.Met86Ile, MAF=0.14%, $P_{discovery}$=2.1x10^{-5}). These rare alleles increase height by 1.9 and 0.9 cm, respectively, suggesting that they both partially impair STC2 activity. In functional studies, STC2 with these amino acid substitutions were expressed at similar levels to wild-type, 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[3,26]. Using association results from 17 human complex phenotypes for which well-powered meta-analysis results were available, we explored if rare and low-frequency height variants are 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<2x10^{-7}) (**Extended Data Fig. 8** and **Supplementary Table 23**). The minor alleles at rs77542162 (*ABCA6*, MAF=1.7%) and rs28929474 (*SERPINA1*, MAF=1.8%) were associated with increased height and increased levels of LDL-cholesterol (LDL-C) and total cholesterol (TC), whereas the minor allele at rs3208856 in *CBLC* (MAF=3.4%) was associated with increased height, HDL-cholesterol (HDL-C) and triglyceride (TG), but lower LDL-C and

196 TC levels. The minor allele at rs141845046 (*ZBTB7B*, MAF=2.8%) was associated with both
197 increased height and body mass index (BMI). The minor alleles at the other two missense
198 variants associated with shorter stature, rs201226914 in *PIEZO1* (MAF=0.2%) and rs35658696
199 in *PAM* (MAF=4.8%), were associated with decreased glycated haemoglobin (HbA1c) and
200 increased type 2 diabetes (T2D) risk, respectively.
201

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 harbor several additional rare/low-frequency variants associated with height, including three genes (*CSAD*, *NOX4*, *UGGT2*) in loci not previously implicated in height. Given the design of the ExomeChip, which did not consider variants with MAF <0.004% (or 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 (four 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 (**Extended Data Tables 1-2** and **Table 1**). 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.

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 harbor coding height variants, highlighting known as well as novel biological pathways that regulate height in humans (**Fig. 2**, **Extended Data Fig. 7** and **Supplementary Tables 19-22**), and newly implicating genes such as *SUSD5, GLT8D2*, *LOXL4*, *FIBIN*, and *SFRP4* that have not

been previously connected with skeletal growth. Additional interesting height candidate genes include *NOX4*, *ADAMTS3* and *ADAMTS6, PTH1R,* and *IL11RA* (**Extended Data Tables 1-2**, **Supplementary Tables 17** and **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 reduced numbers of osteoclasts, a cell type 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 than several other human growth syndromic genes (e.g. *ADAMTS2*, *ADAMTS10, ADAMTSL2*). Moreover, we discovered a rare missense variant in *PTH1R* that encodes a receptor of the parathyroid hormone (PTH): PTH-PTH1R signaling is important for bone resorption and mutations in *PTH1R* cause chondrodysplasia in humans[27]. Finally, we replicated the association between a low-frequency missense variant in the cytokine gene *IL11*, but also found a new low-frequency missense variant in its receptor gene *IL11RA*. The IL11-IL11RA axis has been shown to play an important role in bone formation in the mouse[28,29]. Thus, our data confirm the relevance of this signaling cascade in human growth as well.

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 strong implications for the prediction of complex human phenotypes in the context of precision medicine initiatives. However, it also seems to contrast sharply with results from the recent large-scale T2D association study, which found only six variants with MAF <5% (ref. [30]). This apparent difference could simply be explained by the large difference in sample sizes

between the two studies (711,428 for height vs. 127,145 for T2D). When we consider the fraction of associated variants with MAF<5% among all confirmed variants for height and T2D, we find that it is similar (9.7% for height vs. 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.

SUPPLEMENTARY INFORMATION

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

ACKNOWLEDGMENTS

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)

Panos Deloukas, Timothy M. Frayling, Mariaelisa Graff, Joel N. Hirschhorn, Guillaume Lettre, Ken Sin Lo, Yingchang Lu, Eirini Marouli, M. Carolina Medina-Gomez, Fernando Rivadeneira. 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)

Tonu Esko, Mariaelisa Graff, Heather Highland, Anne Justice, Tugce Karaderi, Ken Sin Lo, Adam E. Locke, Yingchang Lu, Eirini Marouli, Nicholas G.D. Masca, M. Carolina Medina-Gomez, Poorva Mudgal, Maggie C.Y. Ng, Manuel A. Rivas, Claudia Schurmann, Kathy Stirrups, Valérie Turcot, Sailaja Vedantam, Thomas W. Winkler, Kristin L. Young.

Height meta-analyses (discovery and replication, single-variant and gene-based)

Panos Deloukas, Timothy M. Frayling, Mariaelisa Graff, Joel N. Hirschhorn, Guillaume Lettre, Daijiang J. Liu, Ken Sin Lo, Yingchang Lu, Eirini Marouli, M. Carolina Medina-Gomez,

Fernando Rivadeneira, Andrew R. Wood.

UK Biobank-based integration of height association signals group and heritability analyses

Panos Deloukas, Timothy M. Frayling, Guillaume Lettre, Zoltán Kutalik, Ken Sin Lo, Eirini Marouli, Sina Rüeger, Andrew R. Wood.

Pleiotropy working group

Gonçalo Abecasis, Michael Boehnke, James P. Cook, Panos Deloukas, Fotios Drenos, Jose C. Florez, Heather Highland, Sekar Kathiresan, Cecilia M. Lindgren, Dajiang J. Liu, Ruth J.F. Loos, Anubha Mahajan, Eirini Marouli, Mark I. McCarthy, Patricia B. Munroe, Gina M. Peloso, John R. B. Perry, Katherine S. Ruth, Cristen J. Willer.

Biological and clinical enrichment, and pathway analyses

Rebecca S. Fine, Joel N. Hirschhorn, Zoltán Kutalik, David Lamparter, Guillaume Lettre, Ken Sin Lo, Tune H. Pers.

Functional characterization of STC2

Troels R. Kjaer, Claus Oxvig.

AUTHOR INFORMATION

Summary genetic association results are available on the GIANT website: http://portals.broadinstitute.org/collaboration/giant/index.php/GIANT_consortium. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Correspondence and requests should be addressed to J.N.H.

(joelh@broadinstitute.org), P.D. (p.deloukas@qmul.ac.uk), or G.L.

(guillaume.lettre@umontreal.ca).

References

1 Fisher, R. A. The Correlation Between Relatives on the Supposition of Mendelian Inheritance. *Transactions of the Royal Society of Edinburgh* **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 U S A* **111**, E455-464 (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. *American journal of human genetics* **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. *American journal of human genetics* **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 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).

19 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).

20 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).

21 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).

22 Kawano, Y. & Kypta, R. Secreted antagonists of the Wnt signalling pathway. *J Cell Sci* **116**, 2627-2634 (2003).

23 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).

24 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).

25 Dauber, A. *et al.* Mutations in pregnancy-associated plasma protein A2 cause short stature due to low IGF-I availability. *EMBO Mol Med* (2016).

26 Lango Allen, H. *et al.* Hundreds of variants clustered in genomic loci and biological pathways affect human height. *Nature* **467**, 832-838 (2010).

27 Karaplis, A. C. *et al.* Inactivating mutation in the human parathyroid hormone receptor type 1 gene in Blomstrand chondrodysplasia. *Endocrinology* **139**, 5255-5258 (1998).

28 Sims, N. A. *et al.* Interleukin-11 receptor signaling is required for normal bone remodeling. *J Bone Miner Res* **20**, 1093-1102 (2005).

29 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).

30 Fuchsberger, C. *et al.* The genetic architecture of type 2 diabetes. *Nature* **536**, 41-47 (2016).

31 Bulik-Sullivan, B. K. *et al.* LD Score regression distinguishes confounding from polygenicity in genome-wide association studies. *Nat Genet* **47**, 291-295 (2015).

32 Goldstein, J. I. *et al.* zCall: a rare variant caller for array-based genotyping: genetics and population analysis. *Bioinformatics* **28**, 2543-2545 (2012).

33 Liu, D. J. *et al.* Meta-analysis of gene-level tests for rare variant association. *Nat Genet* **46**, 200-204 (2014).

34 Winkler, T. W. & Day, F. R. Quality control and conduct of genome-wide association meta-analyses. **9**, 1192-1212 (2014).

35 Yang, J. *et al.* Genomic inflation factors under polygenic inheritance. *European Journal of Human Genetics* **19**, 807-812 (2011).

36 Feng, S., Liu, D., Zhan, X., Wing, M. K. & Abecasis, G. R. RAREMETAL: fast and powerful meta-analysis for rare variants. *Bioinformatics (Oxford, England)* **30**, 2828-2829 (2014).

37 Yang, J. *et al.* Conditional and joint multiple-SNP analysis of GWAS summary statistics identifies additional variants influencing complex traits. *Nature genetics* **44**, 369-S363 (2012).

38 Loh, P. R. *et al.* Efficient Bayesian mixed-model analysis increases association power in large cohorts. *Nat Genet* **47**, 284-290 (2015).

39 Pasaniuc, B. *et al.* Fast and accurate imputation of summary statistics enhances evidence of functional enrichment. *Bioinformatics* **30**, 2906-2914 (2014).

40 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).

41 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).

42 Willer, C. J., Li, Y. & Abecasis, G. R. METAL: Fast and efficient meta-analysis of genomewide association scans. *Bioinformatics* **26**, 2190-2191 (2010).

43 Purcell, S. M. *et al.* A polygenic burden of rare disruptive mutations in schizophrenia. *Nature* **506**, 185-190 (2014).

44 Wu, M. C. *et al.* Rare-variant association testing for sequencing data with the sequence kernel association test. *American journal of human genetics* **89**, 82-93 (2011).

45 Price, A. L. *et al.* Pooled association tests for rare variants in exon-resequencing studies. *Am J Hum Genet* **86**, 832-838 (2010).

46 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).

47 Fehrmann, R. S. *et al.* Gene expression analysis identifies global gene dosage sensitivity in cancer. *Nat Genet* **47**, 115-125 (2015).

48 Frey, B. J. & Dueck, D. Clustering by passing messages between data points. *Science* **315**, 972-976 (2007).

49 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. *The Journal of biological chemistry* **275**, 31128-31133 (2000).

50 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).

51 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. *Biochimica et biophysica acta* **1201**, 415-423 (1994).

Figure legends

Figure 1. Variants with a larger effect size on height variation tend to be rarer. We observed an inverse relationship between the effect size (from the combined "discovery+validation" analysis, in cm on the *y*-axis) and the minor allele frequency (MAF) for the height variants (*x*-axis, from 0 to 50%). We included in this figure the 606 height variants with $P<2x10^{-7}$.

Figure 2. Heat map showing subset of DEPICT gene set enrichment results. The full heat map is available as **Extended Data Fig. 7**. For any given square, the color 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 minor allele frequency 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 EC by PASCAL (yellow and grey) and if the gene set was found significant by DEPICT for EC only or for both EC and GWAS (purple and green). Abbreviations: GO: Gene

Ontology; MP: mouse phenotype in the Mouse Genetics Initiative; PPI: protein-protein interaction in the InWeb database.

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 signaling. 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 STC2 wild-type and variants R44L and M86I. (**C**) Covalent complex formation between PAPP-A and STC2 wild-type or variants R44L and M86I. Separately synthesized proteins were analyzed by PAPP-A Western blotting following incubation for 8 h. In the absence of STC2 (Mock lane), PAPP-A appears as a single 400 kDa band (*). Following incubation with wild-type STC2, the majority of PAPP-A is present as the approximately 500 kDa covalent PAPP-A:STC2 complex (#), in which PAPP-A is devoid of proteolytic activity towards IGFBP-4. Under similar conditions, incubation with variants R44L or M86I appeared to cause less 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 24 h incubation. Both STC2 variants show increased IGFBP-4 cleavage (*i.e.* less inhibition) for all time points analyzed. Mean and standard deviations 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 ($P<0.01$).

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 based on 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 MAF>5% (grey), 55 have MAF between 1 and 5% (green), and 29 have 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 an LD Score quantile, where the x-axis of the point is the mean LD Score of variants in that quantile and the y-axis is the mean χ^2 statistic of variants in that quantile. The LD Score regression slope of the black line is calculated based on Equation 1 in Bulik-Sullivan et *al*.[31] which is estimated upwards due to the small number of common variants (N=15,848) and the design of the ExomeChip. The LD 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 (standard error=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 >5000 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 minor allele frequency of these variants is <1% (or monomorphic) in all ancestries except African-ancestry

individuals. 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.

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 minor allele frequency (MAF) and heritability explained (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 data set, 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 (EC) or non-coding variation from genome-wide association study data (GWAS). The x-axis indicates the P-value for enrichment of a given gene set using DEPICT adapted for EC data, where the input to DEPICT is the genes implicated by coding EC 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. Colors correspond to whether the meta-gene set was significant for EC only (blue), GWAS only (green), both but more significant for EC (purple), or both but more significant for GWAS (orange), and the most significant gene sets within each category are labeled. A line is drawn at x = y for ease of comparison.

Extended Data Figure 7. Heat map showing entire DEPICT gene set enrichment results (analogous to **Fig. 2** in the main text). For any given square, the color 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 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 minor allele frequency of the significant 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 EC by the Pascal method (yellow and grey) and if the gene set was found significant by DEPICT for EC only or for both EC and GWAS (purple

and green). Abbreviations: GO: Gene Ontology; KEGG: Kyoto encyclopedia of genes and genomes; MP: mouse phenotype in the Mouse Genetics Initiative; PPI: protein-protein interaction in the InWeb database.

Extended Data Figure 8. Heatmaps showing associations of the height variants to other complex traits; –log10(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 $<2x10^{-7}$ and, orange to red shading for hits with negative beta in the other trait and P-values between 0.05 to $<2x10^{-7}$. 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**) Variants for which the minor allele is the height-decreasing allele. (**B**) Variants for which the minor allele is the height-increasing allele.

METHODS

Study design & 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 (N=695). All participating institutions and coordinating centers approved this project, and informed consent was obtained from all subjects. Discovery meta-analysis was carried out in each ancestry group (except the Saudi) 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).

Phenotype

Height (in centimeters) was corrected for age and the genomic principal components (derived from GWAS data, the variants with MAF >1% on ExomeChip, or ancestry informative markers available on the ExomeChip), as well as any additional study-specific covariates (e.g. recruiting center), 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 software, which was then followed by zCall[32]. For 10 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 centers were assembled into a single project for joint calling[11]. Study-specific quality control (QC) measures of the genotyped variants was implemented before association analysis (**Supplementary Tables 1-2**).

Study-level statistical analyses

Individual cohorts were analyzed 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 (LD) relationships between markers within 1 Mb, which is used for gene-level meta-analyses and conditional analyses[33]. Single-variant analyses were performed for both additive and recessive models.

Centralized quality-control

The individual study data were investigated for potential existence of ancestry population outliers based on 1000 Genome Project phase 1 ancestry reference populations. A centralized QC procedure implemented in EasyQC[34] 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 (QQ) plots per study to identify any problems arising from population stratification, cryptic relatedness and genotype biases. We excluded variants if they had call rate <95%, Hardy-Weinberg equilibrium $P<1x10^{-7}$, 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 array-wide level ($P<2x10^{-7}$, Bonferroni correction for 250,000 variants). The combined additive analyses identified 1,455 unique variants that reached array-wide significance ($P<2x10^{-7}$), 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 (*e.g.* 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 artifacts (**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[3,35]. To confirm this, we applied the recently developed linkage disequilibrium (LD) score regression method to our height ExomeChip results[31], 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 LD scores for the European reference dataset[31]. The intercept of the regression of the χ^2 statistics from the height meta-analysis on the LD score estimate the inflation in the mean χ^2 due to confounding bias, such as cryptic relatedness or population stratification. The intercept was 1.4 (standard error = 0.07), which is small when compared to the λ_{GC} of 2.7. Furthermore, we also confirmed that the LD score regression intercept is estimated upward because of the small number of variants on the ExomeChip and the selection criteria for these

641 variants (*i.e.* known GWAS hits). The ratio statistic of (intercept -1) / (mean χ^2 -1) is 0.067
642 (standard error = 0.012), well within the normal range[31], suggesting that most of the inflation
643 (~93%) observed in the height association statistics is due to polygenic effects (**Extended Data**
644 **Fig. 2**).
645
646 Furthermore, to exclude the possibility that some of the observed associations between height
647 and rare/low-frequency variants could be due to allele calling problems in the smaller studies, we
648 performed a sensitivity meta-analysis with primarily Europe-ancestry studies totaling >5,000
649 participants. We found very concordant effect sizes, suggesting that smaller studies do not bias
650 our results (**Extended Data Fig. 2**).
651
652 *Conditional analyses*. The RAREMETAL R-package[36] and the GCTA v1.24[37] software were
653 used to identify independent height association signals across the European descent meta-
654 analysis results. RAREMETAL performs conditional analyses by using covariance matrices in
655 order to distinguish true signals from those driven by LD at adjacent known variants. First, we
656 identified the lead variants ($P<2x10^{-7}$) based on a 1 Mb window centered on the most
657 significantly associated variant and performed LD pruning ($r^2<0.3$) to avoid downstream
658 problems in the conditional analyses due to co-linearity. We then conditioned on the LD-pruned
659 set of lead variants in RAREMETAL and kept new lead signals at $P<2x10^{-7}$. The process was
660 repeated until no additional signal emerged below the pre-specified P-value threshold. The use of
661 a 1Mb window in RAREMETAL can obscure dependence between conditional signals in
662 adjacent intervals in regions of extended LD. To detect such instances, we performed joint
663 analyses using GCTA with the ARIC and UK ExomeChip reference panels, both of which

664 comprise >10,000 individuals of European descent. With the exception of a handful of variants
665 in a few genomic regions with extended LD (*e.g.* the HLA region on chromosome 6), the two
666 software identified the same independent signals (at $P<2x10^{-7}$).
667
668 To discover new height variants, we conditioned the height variants found in our ExomeChip
669 study on the previously published GWAS height variants[3] using the first release of the UK
670 Biobank imputed dataset and regression methodology implemented in BOLT-LMM[38]. Because
671 of the difference between the sample size of our discovery set (N=458,927) and the UK Biobank
672 (first release, N=120,084), we applied a threshold of $P_{conditional}$ <0.05 to declare a height variant
673 as independent in this analysis. We also explored an alternative approach based on approximate
674 conditional analysis[37]. This latter method (SSimp) relies on summary statistics available from the
675 same cohort, thus we first imputed summary statistics[39] for exome variants, using summary
676 statistics from the Wood et *al*. 2014 study[3]. Conversely, we imputed the top variants from the
677 Wood et *al*. 2014 study using the summary statistics from the ExomeChip. Subsequently, we
678 calculated effect sizes for each exome variant conditioned on the Wood et *al*. 2014 top variants
679 in two ways. First, we conditioned the imputed summary statistics of the exome variant on the
680 summary statistics of the Wood et *al*. 2014 top variants that fell within 5 Mb of the target
681 ExomeChip variant. Second, we conditioned the summary statistics of the ExomeChip variant on
682 the imputed summary statistics of the Wood et *al*. 2014 hits. We then selected the option that
683 yielded a higher imputation quality. For poorly tagged variants ($\hat{r}^2$< 0.8), we simply used up-
684 sampled HapMap summary statistics for the approximate conditional analysis. Pairwise SNP-by-
685 SNP correlations were estimated from the UK10K data (TwinsUK[40] and ALSPAC[41] studies ,
686 N=3,781).

687

688 *Validation of the single-variant discovery results.* Several studies, totaling 252,501 independent
689 individuals of European ancestry, became available after the completion of the discovery
690 analyses, and were thus used for validation of our experiment. We validated the single-variant
691 association results in eight studies, totaling 59,804 participants, genotyped on the Exomechip
692 using RAREMETAL[33]. We sought additional evidence for association for the top signals in two
693 independent studies in the UK (UK Biobank) and Iceland (deCODE), comprising 120,084 and
694 72,613 individuals, respectively. We used the same QC and analytical methodology as described
695 above. Genotyping and study descriptives are provided in **Supplementary Tables 1-3**. For the
696 combined analysis, we used the inverse-variance weighted fixed effects meta-analysis method
697 using METAL[42]. Significant associations were defined as those with a combined meta-analysis
698 (discovery and validation) $P_{combined}<2x10^{-7}$.
699

700 We considered 81 variants with suggestive association in the discovery analyses ($2x10^{-7}<P_{discovery}\leq2x10^{-6}$). Of those 81 variants, 55 reached significance after combining discovery and
702 replication results based on $P_{combined}<2x10^{-7}$ (**Supplementary Table 9**). Furthermore, recessive
703 modeling confirmed seven new independent markers with $P_{combined}<2x10^{-7}$ (**Supplementary**
704 **Table 10**). One of these recessive signals is due to a rare X-linked variant in the *AR* gene
705 (rs137852591, MAF=0.21%). Because of its frequency, we only tested hemizygous men (we did
706 not identify homozygous women for the minor allele) so we cannot distinguish between a true
707 recessive mode of inheritance or a sex-specific effect for this variant. To test the independence
708 and integrate all height markers from the discovery and validation phase, we used conditional
709 analyses and GCTA "joint" modeling[37] 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 only consider the initial set of lead SNPs with $P < 2x10^{-7}$, 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 $P_{validation} < 8.9x10^{-5}$, 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)[43]. The mask labeled *"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 labeled *"strict"* included only variants with 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[44] and VT[45] tests. Statistical significance for gene-based tests was set at a Bonferroni-corrected threshold of $P<5x10^{-7}$ (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 (TC), triglycerides (TG), HDL-cholesterol (HDL-C), LDL-cholesterol (LDL-C)), IBPC (systolic and diastolic blood pressure), MAGIC (glycaemic traits), REPROGEN (age at menarche and menopause), and DIAGRAM (type 2 diabetes). For coronary artery disease, we accessed 1000 Genomes Project-imputed GWAS data released by CARDIoGRAMplusC4D[46].

Pathway analyses

DEPICT is a computational framework that uses probabilistically-defined reconstituted gene sets to perform gene set enrichment and gene prioritization[15]. For a description about gene set reconstitution please refer to references [15] and [47]. 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, we made two principal changes. First and foremost, because DEPICT for GWAS incorporates all genes within a given LD 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 adjust 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 (1) with a non-coding index SNP and (2) that did not contain any of the novel ExomeChip genes. In visualizing the analysis, we used affinity propagation clustering[48] 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 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[40] and ALSPAC[41] studies , N=3781). In order 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 the Wood et *al*. study[3]. In this way, we could classify pathways driven principally by coding, regulatory or mixed signals.

STC2 functional experiments

778 *Mutagenesis, cell culture and transfection.* For the generation of STC2 mutants (R44L and
779 M86I), wild-type STC2 cDNA contained in pcDNA3.1/Myc-His(-) (Invitrogen)[24] was used as a
780 template. Mutagenesis was carried out using Quickchange (Stratagene), and all constructs were
781 verified by sequence analysis. Recombinant wild-type STC2 and variants were expressed in
782 human embryonic kidney (HEK) 293T cells (293tsA1609neo, ATCC CRL-3216) maintained in
783 high-glucose DMEM supplemented 10% fetal bovine serum, 2 mM glutamine, nonessential
784 amino acids, and gentamicin. The cells are routinely tested for mycoplasma contamination. Cells
785 ($6x10^6$) were plated onto 10 cm-dishes and transfected 18 h later by calcium phosphate
786 coprecipitation using 10 μg plasmid DNA. Media were harvested 48 h post transfection, cleared
787 by centrifugation, and stored at -20°C until use. Protein concentrations (58-66 nM) were
788 determined by TRIFMA using antibodies described previously[24]. PAPP-A was expressed stably
789 in HEK293T cells as previously reported[49]. Expressed levels of PAPP-A (27.5 nM) were
790 determined by a commercial ELISA (AL-101, Ansh Labs, TX).
791
792 *STC2 and PAPP-A complex formation.* Culture supernatants containing wild-type STC2 or
793 variants were adjusted to 58 nM, added an equal volume of culture supernatant containing
794 PAPP-A corresponding to a 2.1-fold molar excess, and incubated at 37°C. Samples were taken at
795 1, 2, 4, 6, 8, 16, and 24 h and stored at -20°C.
796
797 *Analysis of proteolytic activity.* Specific proteolytic cleavage of ^{125}I-labeled IGFBP-4 is
798 described in detail elsewhere[50]. Briefly, the PAPP-A:STC2 complex mixtures were diluted
799 (1:190) to a concentration of 145 pM PAPP-A and mixed with preincubated ^{125}I-IGFBP4 (10
800 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% nonreducing 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).

Western blotting. 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 in TST supplemented with 2% skim milk for 1 h at 20°C. For PAPP-A:STC2 complexes, the membranes were incubated with rabbit polyclonal anti-PAPP-A[51] at 0.63 μg/ml in TST supplemented with 2% skim milk for 16 h at 20°C. Membranes were washed with TST and subsequently incubated with polyclonal swine anti-rabbit IgG-HRP (DAKO, P0217) or polyclonal rabbit anti-goat IgG-HRP (DAKO, P0449), respectively, diluted 1:2000 in TST supplemented with 2% skim 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 STATEMENT

Summary genetic association results are available on the GIANT website:

http://portals.broadinstitute.org/collaboration/giant/index.php/GIANT_consortium.

URLs

ClinVar, http://www.ncbi.nlm.nih.gov/clinvar/

DEPICT, http://www.broadinstitute.org/mpg/depict/

ExomeChip, http://genome.sph.umich.edu/wiki/Exome_Chip_Design

ExomeDEPICT, https://github.com/RebeccaFine/height-ec-depict

OMIM, http://omim.org/

PASCAL, http://www2.unil.ch/cbg/index.php?title=Pascal

RAREMETALWORKER, http://genome.sph.umich.edu/wiki/RAREMETALWORKER

RVTEST, http://zhanxw.github.io/rvtests/

Table 1. Ten height genes implicated by gene-based testing. These genes meet our three criteria for statistical significance: (1) gene-based $P<5x10^{-7}$, (2) the gene does not include variants with $P<2x10^{-7}$, 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. [1]Validation (N=59,804) and combined results using the same test and (when possible) variants. [2]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). [3]If the gene falls within a known GWAS height locus, we mention if it was predicted to be causal using bioinformatic tools (ref. [3]). NA, not applicable.

Gene	Discovery gene-based P-value				Validation P-value[1]	Combined P-value[1]	Conditional P-value[2]	Note[3]
	SKAT-broad	VT-broad	SKAT-strict	VT-strict				
OSGIN1	**$4.3x10^{-11}$**	$4.5x10^{-5}$	0.19	0.18	0.048	$2.6x10^{-12}$	$7.7x10^{-11}$	Known locus. No predicted causal genes.
CRISPLD1	$2.2x10^{-7}$	**$6.7x10^{-11}$**	$8.5x10^{-6}$	$8.9x10^{-7}$	0.50	$1.2x10^{-12}$	NA	Known locus, sentinel GWAS SNP not tested on ExomeChip. *CRISPLD1* was predicted to be causal.
CSAD	$2.3x10^{-8}$	**$2.4x10^{-9}$**	0.83	0.59	0.54	$2.0x10^{-9}$	NA	New locus.
SNED1	$1.9x10^{-5}$	**$4.3x10^{-9}$**	NA	NA	0.083	$4.5x10^{-10}$	$1.4x10^{-9}$	Known locus. *SNED1* was not predicted to be causal.
G6PC	$1.3x10^{-5}$	**$3.6x10^{-8}$**	$5.5x10^{-6}$	$1.3x10^{-6}$	0.24	$5.2x10^{-8}$	$3.9x10^{-8}$	Known locus, *G6PC* was not predicted to be causal. *G6PC* is mutated in glycogen storage disease Ia.
NOX4	$5.1x10^{-6}$	**$1.4x10^{-7}$**	NA	NA	0.013	$5.5x10^{-9}$	NA	New locus.
UGGT2	$3.0x10^{-5}$	**$2.6x10^{-7}$**	$2.3x10^{-5}$	$4.8x10^{-7}$	0.64	$3.4x10^{-7}$	NA	New locus.
FLNB	$2.2x10^{-6}$	$5.1x10^{-4}$	**$2.4x10^{-9}$**	$3.2x10^{-6}$	0.016	$8.6x10^{-11}$	$3.6x10^{-9}$	Known locus. *FLNB* was predicted to be causal. *FLNB* is mutated in atelosteogenesis type I.
B4GALNT3	$2.4x10^{-5}$	$1.9x10^{-5}$	$1.8x10^{-5}$	**$3.1x10^{-7}$**	0.79	$4.3x10^{-7}$	$7.7x10^{-7}$	Known locus. *B4GALNT3* was predicted to be causal.
CCDC3	$6.3x10^{-4}$	$6.3x10^{-6}$	$3.0x10^{-7}$	**$5.4x10^{-9}$**	0.080	$1.2x10^{-9}$	$1.6x10^{-9}$	Known locus. *CCDC3* was predicted to be causal.

Authors
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[8-10]*, Yingchang Lu[11-13]*, 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 GD Masca[25,26], Maggie CY Ng[27,28], Poorva Mudgal[27], Manuel A Rivas[8,29], Sailaja Vedantam[8-10], 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[38-40], Emil V Appel[36], Dominique Arveiler[41,42], Folkert W Asselbergs[43-45], 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[70-72], Daniel I Chasman[8,73-75], 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 IW de Bakker[59,88], Gert J. de Borst[89], Simon de Denus[5,90], Mark CH 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[119-121], 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[140-142], Alex W Hewitt[143-145], Lynne J Hocking[146,147], Mette Hollensted[36], Oddgeir L Holmen[148], G. Kees Hovingh[149], Joanna MM 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[168-170], René S Kahn[171], Mika Kähönen[172], Pia R Kamstrup[50], Stavroula Kanoni[1], Jaakko Kaprio[136,137,165], Maria Karaleftheri[173], Sharon LR 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[182-184], 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[195-197], 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 AF 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 NA 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[168-170], John M Starr[80,246], Valgerdur Steinthorsdottir[17], Heather M Stringham[23], Michael Stumvoll[53,54], Praveen Surendran[66], Leen M ‘t Hart[247-249], 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[165,257-259], 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], EPIC-CVD Consortium[¶], The EPIC-InterAct 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 JF Loos[12,13,279§], Timothy M Frayling[6§], Joel N Hirschhorn[8,10,280§], Panos Deloukas[1,281§], Guillaume Lettre[5,237§]

*These authors contributed equally to this work.
[¶]A full list of members appears in the **Supplementary Information.**
[§]These authors jointly supervised this work.

Correspondence should be addressed to JNH (joelh@broadinstitute.org), PD (p.deloukas@qmul.ac.uk) or GL (guillaume.lettre@umontreal.ca).

Affiliations

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, NC, 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, Université de Montréal, 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, MA, 02142, USA
9. Department of Genetics, Harvard Medical School, Boston, MA, 02115, USA
10. Division of Endocrinology and Center for Basic and Translational Obesity Research, Boston Childrenís Hospital, Boston, MA, 02115, USA
11. Division of Epidemiology, Department of Medicine, Vanderbilt-Ingram Cancer Center, Vanderbilt Epidemiology Center, Vanderbilt University School of Medicine, Nashville, TN, 37203, USA
12. The Charles Bronfman Institute for Personalized Medicine, Icahn School of Medicine at Mount Sinai, New York, NY, 10029, USA
13. The Genetics of Obesity and Related Metabolic Traits Program, Ichan School of Medicine at Mount Sinai, New York, NY, 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, TX, 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, MI, 48109, USA
24. McDonnell Genome Institute, Washington University School of Medicine, Saint Louis, MO, 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, NC, 27157, USA
28. Center for Genomics and Personalized Medicine Research, Wake Forest School of Medicine, Winston-Salem, NC, 27157, USA
29. Nuffield Department of Clinical Medicine, Oxford, OX37BN, UK
30. Institute for Translational Genomics and Population Sciences, LABioMed at Harbor-UCLA Medical Center, Torrance, CA, 90502, USA
31. Netherlands Comprehensive Cancer Organisation, Utrecht, 3501 DB, The Netherlands
32. Dept of obstetrics and gynaecology, Radboud University Medical Center, Nijmegen, 6500 HB, The Netherlands
33. Department of Nutrition, University of North Carolina, Chapel Hill, NC, 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, CA, 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, WI, 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, MI, 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, TX, 77030, USA
57. Human Genome Sequencing Center, Baylor College of Medicine, Houston, TX, 77030 USA
58. Medical Genomics and Metabolic Genetics Branch, National Human Genome Research Institute, National Institutes of Health, Bethesda, MD, 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, NC 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, Kingís College London & NIHR Biomedical Research Centre for Mental Health at the Maudsley, London, SE5 8AF, UK
64. Marshfield Clinic Research Foundation, Marshfield, WI, 54449, USA
65. Department of Medicine, University of Washington, Seattle, WA, 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, PA, 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, MA, 02115, USA
74. Division of Preventive Medicine, Brigham and Women's and Harvard Medical School, Boston, MA, 02215, USA
75. Harvard Medical School, Boston, MA, 02115, USA
76. Medical department, Lillebaelt Hospital, Vejle, 7100, Denmark
77. NHLBI Framingham Heart Study, Framingham, MA, 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, QLD, 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, Dvision Pharmacoepidemiology & Clinical Pharmacology, Utrecht University, Utrecht, 3508 TB, The Netherlands
93. Department of Clinical Epidemiology, Leiden University Medical Center, Leiden, 2300RC, 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, MN, 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 Bristo, Bristol, BS8 2BN, UK
100. Fred Hutchinson Cancer Research Center, Public Health Sciences Division, Seattle, WA, 98109, USA
101. Memorial Sloan Kettering Cancer Center, Department of Epidemiology and Biostatistics, New York, NY, 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, TN, 37203, USA
106. Massachusetts General Hospital, Boston, MA, 02114, USA
107. Medical and Population Genetics Program, Broad Institute, Cambridge, MA, 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, MI, 48104, USA
111. Division of Statistical Genomics, Department of Genetics, Washington University School of Medicine, St. Louis, MO, 63108, USA
112. CNR Institute of Clinical Physiology, Pisa, Italy
113. Department of Clinical & Experimental Medicine, University of Pisa, Italy
114. Research Center on Epidemiology and Preventive Medicine, Dept. 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, MA, 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, TX, 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, MA, 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, NC, 27514, USA
127. Department of Nutrition, Gillings School of Global Public Health, University of North Carolina, Chapel Hill, NC, 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, NC, 27514, USA
132. Laboratory of Epidemiology and Population Sciences, National Institute on Aging, Intramural Research Program, National Institutes of Health, Bethesda, MD, 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, NC, 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, 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, CA, 943 05, 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, WA, 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, USA, 90089, USA
159. USC Roski Eye Institute, Department of Ophthalmology, Keck School of Medicine of the University of Southern California, Los Angeles, CA, 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, MI, 48109, USA
169. Department of Internal Medicine, University of Michigan, Ann Arbor, MI, 48109, USA
170. Division of Gastroenterology, University of Michigan, Ann Arbor, MI, 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, UK, 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, USA
181. Division of Public Health Sciences, Fred Hutchinson Cancer Research Center, Seattle WA, 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, NC, 27514, USA
191. Department of Biostatistical Sciences and Center for Public Health Genomics, Wake Forest School of Medicine, Winston-Salem, NC, 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, WA, 98101, USA
194. Department of Health Services, University of Washington, Seattle WA 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, AL, AL 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, CA, 94304, USA
201. Department of Medicine, Boston University School of Medicine, Boston, MA, 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, NC, 27157, USA
205. Division of Health Sciences, Warwick Medical School, Warwick University, Coventry, CV4 7AL, UK
206. Department of Psychiatry, Washington University, Saint Louis, MO, 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, 2300RC, The Netherland
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, MD, 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, MD, 21201, USA
218. Department of Medicine, Tampere University Hospital, Tampere, 33521, Finland
219. Center for Neurobehavioral Genetics, UCLA, Los Angeles, CA, 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. Kuopio University Hospital, Department of Clinical Physiology and Nuclear Medicine, Kuopio, 70210, Finland

233. Merck Research Laboratories, Genetics and Pharmacogenomics, Boston, MA, 02115, USA
234. Department of Epidemiology, University of Washington, Seattle, WA, 98195, USA
235. Department of Biobank Research, Umeå University, Umeå, SE-90187, Sweden
236. Division of Cardiovascular Medicine, Brigham and Womenís Hospital and Harvard Medical School, Boston, MA, 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, PA, 19104, USA
240. Division of Epidemiology & Community Health University of Minnesota, Minneapolis, MN, 55454, USA
241. Duke University, Durham, NC, 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, 1007MB, 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 Womenís 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, TN, 37203, USA
263. Merck Research Laboratories, Cardiometabolic Disease, Kenilworth, NJ, 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, MA, 02118, USA
266. Departments of Epidemiology & Medicine, Diabetes Translational Research Center, Fairbanks School of Public Health & School of Medicine, Indiana University, Indiana, IN, 46202, USA
267. Department of Physiology and Biophysics, University of Mississippi Medical Center, Jackson, MS, 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, NL, A1B 3V6, Canada
271. GlaxoSmithKlein, King of Prussia, PA, 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, TX, 75390, USA
273. Department of Human Genetics, University of Michigan, Ann Arbor, MI, 48109, USA
274. Department of Public Health Sciences, Institute for Personalized Medicine, the Pennsylvania State University College of Medicine, Hershey, PA, 17033, USA
275. Department of Epidemiology and Carolina Center of Genome Sciences, Chapel Hill, NC, 27514, USA
276. Department of Neurology, Boston University School of Medicine, Boston, MA, 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, NY, 10069, USA
280. Departments of Pediatrics and Genetics, Harvard Medical School, Boston, MA, 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