

1 **Rare and low-frequency coding variants alter human adult height**

2

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

4

5 **Correspondence to:**

6 Joel N. Hirschhorn (joelh@broadinstitute.org)

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

8 Guillaume Lettre (guillaume.lettre@umontreal.ca)

9

10 *Summary: 149 words*

11 *Main text: 2,653 words*

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

13 **SUMMARY**

14 Height is a highly heritable, classic polygenic trait with ~700 common associated variants
15 identified so far through genome-wide association studies. Here, we report 83 height-associated
16 coding variants with lower minor allele frequencies (range of 0.1-4.8%) and effects of up to 2
17 cm/allele (*e.g.* in *IHH*, *STC2*, *AR* and *CRISPLD2*), >10 times the average effect of common
18 variants. In functional follow-up studies, rare height-increasing alleles of *STC2* (+1-2 cm/allele)
19 compromised proteolytic inhibition of PAPP-A and increased cleavage of IGFBP-4 *in vitro*,
20 resulting in higher bioavailability of insulin-like growth factors. These 83 height-associated
21 variants overlap genes mutated in monogenic growth disorders and highlight new biological
22 candidates (*e.g.* *ADAMTS3*, *IL11RA*, *NOX4*) and pathways (*e.g.*
23 proteoglycan/glycosaminoglycan synthesis) involved in growth. Our results demonstrate that
24 sufficiently large sample sizes can uncover rare and low-frequency variants of moderate to large
25 effect associated with polygenic human phenotypes, and that these variants implicate relevant
26 genes and pathways.

27

28

29 INTRODUCTION

30 Human height is a highly heritable, polygenic trait^{1,2}. The contribution of common DNA
31 sequence variation to inter-individual differences in adult height has been systematically
32 evaluated through genome-wide association studies (GWAS). This approach has thus far
33 identified 697 independent variants located within 423 loci that together explain ~20% of the
34 heritability of height³. As is typical of complex traits and diseases, most of the height alleles
35 discovered so far are common (minor allele frequency (MAF) >5%) and are mainly located
36 outside coding regions, complicating the identification of the relevant genes or functional
37 variants. Identifying coding variants associated with a complex trait in new or known loci has the
38 potential to pinpoint causal genes. Furthermore, the extent to which rare (MAF <1%) and low-
39 frequency (1% < MAF ≤ 5%) coding variants also influence complex traits and diseases remains
40 an open question. Many recent DNA sequencing studies have identified only few such variants⁴⁻
41 ⁸, but this limited success could be due to their modest sample size⁹. Some studies have
42 suggested that common sequence variants may explain the majority of the heritable variation in
43 adult height¹⁰, making it timely to assess whether and to what extent rare and low-frequency
44 coding variation contributes to the genetic landscape of this model polygenic trait.

45

46 In this study, we used an ExomeChip¹¹ to test the association between 241,453 variants (83%
47 coding with MAF ≤5%) and adult height variation in 711,428 individuals (discovery and
48 validation sample sizes were 458,927 and 252,501, respectively). The ExomeChip is a
49 genotyping array designed to query in very large sample sizes coding variants identified by
50 whole-exome DNA sequencing of ~12,000 participants. The main goals of our project were to
51 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

55 RESULTS

56 *32 rare and 51 low-frequency coding variants associated with adult height*

57 We conducted single-variant meta-analyses in a discovery sample of 458,927 individuals, of
58 whom 381,625 were of European ancestry. We validated our association results in an
59 independent set of 252,501 participants. We first performed standard single-variant association
60 analyses; technical details of the discovery and validation steps are in **Methods (Extended Data**
61 **Figs 1-3, Supplementary Tables 1-11)**. In total, we found 606 independent ExomeChip variants
62 at array-wide significance ($P < 2 \times 10^{-7}$), including 252 non-synonymous or splice site variants
63 (**Methods and Supplementary Table 11**). Focusing on non-synonymous or splice site variants
64 with MAF $< 5\%$, our single-variant analyses identified 32 rare and 51 low-frequency height-
65 associated variants (**Extended Data Tables 1-2**). To date, these 83 height variants (MAF range
66 0.1-4.8%) represent the largest set of validated rare and low-frequency coding variants associated
67 with any complex human trait or disease. Among these 83 variants, there are 81 missense, one
68 nonsense (in *CCND3*), and one essential acceptor splice site (in *ARMC5*) variants.

69
70 We observed a strong inverse relationship between MAF and effect size (**Fig. 1**). Although
71 power limits our capacity to find rare variants of small effects, we know that common variants
72 with effect sizes comparable to the largest seen in our study would have been easily discovered
73 by prior GWAS, but were not detected. Our results agree with a model based on accumulating
74 theoretical and empirical evidences that suggest that variants with strong phenotypic effects are
75 more likely to be deleterious, and therefore rarer^{12,13}. The largest effect sizes were observed for
76 four rare missense variants, located in the androgen receptor gene *AR* (rs137852591,
77 MAF=0.21%, $P_{\text{combined}}=2.7 \times 10^{-14}$), in *CRISPLD2* (rs148934412, MAF=0.08%, $P_{\text{combined}}=2.4 \times 10^{-$
78 ²⁰), in *IHH* (rs142036701, MAF=0.08%, $P_{\text{combined}}=1.9 \times 10^{-23}$), and in *STC2* (rs148833559,

79 MAF=0.1%, $P_{\text{combined}}=1.2 \times 10^{-30}$). Carriers of the rare *STC2* missense variant are ~2.1 cm taller
80 than non-carriers, whereas carriers of the remaining three variants (or hemizygous men that carry
81 the X-linked *AR*-rs137852591 rare allele) are ~2 cm shorter than non-carriers. In comparison, the
82 mean effect size of common height alleles is ten times smaller in the same dataset. Across all 83
83 rare and low-frequency non-synonymous variants, the minor alleles were evenly distributed
84 between height-increasing and -decreasing effects (48% vs. 52%, respectively) (**Fig. 1** and
85 **Extended Data Tables 1-2**).

86

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

88 Many of the height-associated variants in this ExomeChip effort are located near common
89 variants previously associated with height. Of the 83 rare and low-frequency non-synonymous
90 variants, two low-frequency missense variants were previously identified (in *CYTL1* and
91 *IL11*)^{3,14} and 47 fell within 1 Mb of a known height signal; the remaining 34 define new loci. We
92 used conditional analysis in the UK Biobank dataset and confirmed that 38 of these 47 variants
93 were independent from the previously described height SNPs (**Supplementary Table 12**). We
94 validated the UK Biobank conditional results using an orthogonal imputation-based methodology
95 implemented in the full discovery set (**Extended Data Fig. 4** and **Supplementary Table 12**). In
96 addition, we found a further 85 common variants and one low-frequency synonymous variant (in
97 *ACHE*) that define novel loci (**Supplementary Table 12**). Thus, our study identified a total of
98 120 new height loci (**Supplementary Table 11**).

99

100 We used the UK Biobank dataset to estimate the contribution of the new height variants to
101 heritability, which is $h^2 \sim 80\%$ for adult height². In combination, the 83 rare and low-frequency

102 variants explained 1.7% of the heritability of height. The newly identified novel common
103 variants accounted for another 2.4%, and all independent variants, known and novel together
104 explained 27.4% of heritability. By comparison, the 697 known height SNPs explain 23.3% of
105 height heritability in the same dataset (vs. 4.1% by the new height variants identified in this
106 ExomeChip study). We observed a modest positive association between MAF and heritability
107 explained per variant ($P=0.012$, **Extended Data Fig. 5**), with each common variant explaining
108 slightly more heritability than rare or low-frequency variants (0.036% vs. 0.026%, **Extended**
109 **Data Fig. 5**).

110

111 *Gene-based association results*

112 To increase power to find rare or low-frequency coding variants associated with height, we
113 performed gene-based analyses (**Methods** and **Supplementary Tables 13-15**). After accounting
114 for gene-based signals explained by a single variant driving the association statistics, we
115 identified ten genes with $P < 5 \times 10^{-7}$ that harbor more than one coding variant independently
116 associated with height variation (**Supplementary Tables 16-17**). These gene-based results
117 remained significant after conditioning on genotypes at nearby common height-associated
118 variants present on the ExomeChip (**Table 1**). Using the same gene-based tests in an independent
119 dataset of 59,804 individuals genotyped on the same exome array, we replicated three genes at
120 $P < 0.05$ (**Table 1**). Further evidence for replication in these genes was seen at the level of single
121 variants (**Supplementary Table 18**). From the gene-based results, three genes – *CSAD*, *NOX4*,
122 and *UGGT2* – fell outside of the loci found by single-variant analyses and are implicated in
123 human height for the first time.

124

125 *Coding variants implicate biological pathways in human skeletal growth*

126 Prior pathway analyses of height loci identified by GWAS have highlighted gene sets related to
127 both general biological processes (such as chromatin modification and regulation of embryonic
128 size) and more skeletal growth-specific pathways (chondrocyte biology, extracellular matrix
129 (ECM), and skeletal development)³. We used two different methods, DEPICT¹⁵ and PASCAL¹⁶
130 (**Methods**), to perform pathway analyses using the ExomeChip results to test whether coding
131 variants could either independently confirm the relevance of these previously highlighted
132 pathways (and further implicate specific genes in these pathways), or identify new pathways. To
133 compare the pathways emerging from coding and non-coding variation, we applied DEPICT
134 separately on (1) exome array-wide associated coding variants independent of known GWAS
135 signals and (2) non-coding GWAS loci, excluding all novel height-associated genes implicated
136 by coding variants. We identified a total of 496 and 1,623 enriched gene sets, respectively, at a
137 false discovery rate (FDR) <1% (**Supplementary Tables 19-20**); similar analyses with PASCAL
138 yielded 362 and 278 enriched gene sets (**Supplementary Tables 21-22**). Comparison of the
139 results revealed a high degree of shared biology for coding and non-coding variants (for
140 DEPICT, gene set P-values compared between coding and non-coding results had Pearson's $r =$
141 0.583 , $P < 2.2 \times 10^{-16}$; for PASCAL, Pearson's $r = 0.605$, $P < 2.2 \times 10^{-16}$). However, some pathways
142 showed stronger enrichment with either coding or non-coding genetic variation. In general,
143 coding variants more strongly implicated pathways specific to skeletal growth (such as ECM and
144 bone growth), while GWAS signals highlighted more global biological processes (such as
145 transcription factor binding and embryonic size/lethality)(**Extended Data Fig. 6**). The two
146 significant gene sets identified by DEPICT and PASCAL that uniquely implicated coding
147 variants were “BCAN protein protein interaction subnetwork” and “proteoglycan binding.” Both
148 of these pathways relate to the biology of proteoglycans, which are proteins (such as aggrecan)

149 that contain glycosaminoglycans (such as chondroitin sulfate) and that have well-established
150 connections to skeletal growth¹⁷.

151

152 We also examined which height-associated genes identified by ExomeChip analyses were
153 driving enrichment of pathways such as proteoglycan binding. Using unsupervised clustering
154 analysis, we observed that a cluster of 15 height-associated genes is strongly implicated in a
155 group of correlated pathways that include biology related to proteoglycans/glycosaminoglycans
156 (**Fig. 2** and **Extended Data Fig. 7**). Seven of these 15 genes overlap a previously curated list of
157 277 genes annotated in OMIM as causing skeletal growth disorders¹⁸; genes in this small cluster
158 are enriched for OMIM annotations relative to genes outside the cluster (odds ratio=27.6,
159 Fisher's exact $P=1.1 \times 10^{-5}$). As such, the remaining genes in this cluster may be strong candidates
160 for harboring variants that cause Mendelian growth disorders. Within this group are genes that
161 are largely uncharacterized (*SUSD5*), have relevant biochemical functions (*GLT8D2*, a
162 glycosyltransferase studied mostly in the context of the liver¹⁹; *LOXLA*, a lysyl oxidase expressed
163 in cartilage²⁰), modulate pathways known to affect skeletal growth (*FIBIN*, *SFRP4*)^{21,22} or lead to
164 increased body length when knocked out in mice (*SFRP4*)²³.

165

166 ***Functional characterization of rare STC2 variants***

167 To begin exploring whether the identified rare coding variants affect protein function, we
168 performed *in vitro* functional analyses of two rare coding variants in a particularly compelling
169 and novel candidate gene, *STC2*. Over-expression of *STC2* diminishes growth in mice by
170 covalent binding and inhibition of the proteinase PAPP-A, which specifically cleaves IGF
171 binding protein-4 (IGFBP-4), leading to reduced levels of bioactive insulin-like growth factors

172 (Fig. 3A)²⁴. Although there was no prior genetic evidence implicating *STC2* variation in human
173 growth, the *PAPPA* and *IGFBP4* genes were both implicated in height GWAS³, and rare
174 mutations in *PAPPA2* cause severe short stature²⁵, emphasizing the likely relevance of this
175 pathway in humans. The two *STC2* height-associated variants are rs148833559 (p.Arg44Leu,
176 MAF=0.096%, $P_{\text{discovery}}=5.7 \times 10^{-15}$) and rs146441603 (p.Met86Ile, MAF=0.14%,
177 $P_{\text{discovery}}=2.1 \times 10^{-5}$). These rare alleles increase height by 1.9 and 0.9 cm, respectively, suggesting
178 that they both partially impair *STC2* activity. In functional studies, *STC2* with these amino acid
179 substitutions were expressed at similar levels to wild-type, but showed clear, partial defects in
180 binding to PAPP-A and in inhibition of PAPP-A-mediated cleavage of IGFBP-4 (Fig. 3B-D).
181 Thus, the genetic analysis successfully identified rare coding alleles that have demonstrable and
182 predicted functional consequences, strongly confirming the role of these variants and the *STC2*
183 gene in human growth.

184

185 *Pleiotropic effects*

186 Previous GWAS studies have reported pleiotropic or secondary effects on other phenotypes for
187 many common variants associated with adult height^{3,26}. Using association results from 17 human
188 complex phenotypes for which well-powered meta-analysis results were available, we explored
189 if rare and low-frequency height variants are also pleiotropic. We found one rare and five low-
190 frequency missense variants associated with at least one of the other investigated traits at array-
191 wide significance ($P < 2 \times 10^{-7}$) (Extended Data Fig. 8 and Supplementary Table 23). The minor
192 alleles at rs77542162 (*ABCA6*, MAF=1.7%) and rs28929474 (*SERPINA1*, MAF=1.8%) were
193 associated with increased height and increased levels of LDL-cholesterol (LDL-C) and total
194 cholesterol (TC), whereas the minor allele at rs3208856 in *CBLC* (MAF=3.4%) was associated
195 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

202 DISCUSSION

203 We undertook an association study of nearly 200,000 coding variants in 711,428 individuals, and
204 identified 32 rare and 51 low-frequency coding variants associated with adult height.

205 Furthermore, gene-based testing discovered 10 genes that harbor several additional rare/low-
206 frequency variants associated with height, including three genes (*CSAD*, *NOX4*, *UGGT2*) in loci
207 not previously implicated in height. Given the design of the ExomeChip, which did not consider
208 variants with MAF <0.004% (or one allele in ~12,000 participants), our gene-based association
209 results do not rule out the possibility that additional genes with such rarer coding variants also
210 contribute to height variation; deep DNA sequencing in very large sample sizes will be required
211 to address this question. In total, our results highlight 89 genes (10 from gene-based testing and
212 79 from single-variant analyses (four genes have 2 independent coding variants)) that are likely
213 to modulate human growth, and 24 alleles segregating in the general population that affect height
214 by more than 1 cm (**Extended Data Tables 1-2** and **Table 1**). The rare and low-frequency
215 coding variants explain 1.7% of the heritable variation in adult height. When considering all rare,
216 low-frequency, and common height-associated variants validated in this study, we can now
217 explain 27.4% of the heritability.

218
219 Our analyses revealed many coding variants in genes mutated in monogenic skeletal growth
220 disorders, confirming the presence of allelic series (from familial penetrant mutations to mild
221 effect common variants) in the same genes for related growth phenotypes in humans. We used
222 gene set enrichment-type analyses to demonstrate the functional connectivity between the genes
223 that harbor coding height variants, highlighting known as well as novel biological pathways that
224 regulate height in humans (**Fig. 2**, **Extended Data Fig. 7** and **Supplementary Tables 19-22**),
225 and newly implicating genes such as *SUSD5*, *GLT8D2*, *LOXLA*, *FIBIN*, and *SFRP4* that have not

226 been previously connected with skeletal growth. Additional interesting height candidate genes
227 include *NOX4*, *ADAMTS3* and *ADAMTS6*, *PTH1R*, and *IL11RA* (**Extended Data Tables 1-2**,
228 **Supplementary Tables 17** and **24**). *NOX4*, identified through gene-based testing, encodes
229 NADPH oxidase 4, an enzyme that produces reactive oxygen species, a biological pathway not
230 previously implicated in human growth. *Nox4*^{-/-} mice display higher bone density and reduced
231 numbers of osteoclasts, a cell type essential for bone repair, maintenance, and remodelling¹². We
232 also found rare coding variants in *ADAMTS3* and *ADAMTS6*, genes that encode
233 metalloproteinases that belong to the same family than several other human growth syndromic
234 genes (e.g. *ADAMTS2*, *ADAMTS10*, *ADAMTSL2*). Moreover, we discovered a rare missense
235 variant in *PTH1R* that encodes a receptor of the parathyroid hormone (PTH): PTH-PTH1R
236 signaling is important for bone resorption and mutations in *PTH1R* cause chondrodysplasia in
237 humans²⁷. Finally, we replicated the association between a low-frequency missense variant in the
238 cytokine gene *IL11*, but also found a new low-frequency missense variant in its receptor gene
239 *IL11RA*. The IL11-IL11RA axis has been shown to play an important role in bone formation in
240 the mouse^{28,29}. Thus, our data confirm the relevance of this signaling cascade in human growth
241 as well.

242

243 Overall, our findings provide strong evidence that rare and low-frequency coding variants
244 contribute to the genetic architecture of height, a model complex human trait. This conclusion
245 has strong implications for the prediction of complex human phenotypes in the context of
246 precision medicine initiatives. However, it also seems to contrast sharply with results from the
247 recent large-scale T2D association study, which found only six variants with MAF <5% (ref. ³⁰).
248 This apparent difference could simply be explained by the large difference in sample sizes

249 between the two studies (711,428 for height vs. 127,145 for T2D). When we consider the
250 fraction of associated variants with $MAF < 5\%$ among all confirmed variants for height and T2D,
251 we find that it is similar (9.7% for height vs. 7.1% for T2D). This supports the strong probability
252 that rarer T2D alleles and more generally, rarer alleles for other polygenic diseases and traits,
253 will be uncovered as sample sizes continue to increase.

254 **SUPPLEMENTARY INFORMATION**

255 **Supplementary Information** is linked to the online version of the paper at

256 www.nature.com/nature.

257

258 **ACKNOWLEDGMENTS**

259 A full list of acknowledgments appears in the **Supplementary Information**. Part of this work

260 was conducted using the UK Biobank resource.

261

262 **AUTHOR CONTRIBUTIONS**

263 *Writing Group (wrote and edited manuscript)*

264 Panos Deloukas, Timothy M. Frayling, Mariaelisa Graff, Joel N. Hirschhorn, Guillaume Lettre,

265 Ken Sin Lo, Yingchang Lu, Eirini Marouli, M. Carolina Medina-Gomez, Fernando Rivadeneira.

266 All authors contributed and discussed the results, and commented on the manuscript.

267

268 *Data preparation group (checked and prepared data from contributing cohorts for meta-*

269 *analyses and replication)*

270 Tonu Esko, Mariaelisa Graff, Heather Highland, Anne Justice, Tugce Karaderi, Ken Sin Lo,

271 Adam E. Locke, Yingchang Lu, Eirini Marouli, Nicholas G.D. Masca, M. Carolina Medina-

272 Gomez, Poorva Mudgal, Maggie C.Y. Ng, Manuel A. Rivas, Claudia Schurmann, Kathy

273 Stirrups, Valérie Turcot, Sailaja Vedantam, Thomas W. Winkler, Kristin L. Young.

274

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

276 Panos Deloukas, Timothy M. Frayling, Mariaelisa Graff, Joel N. Hirschhorn, Guillaume Lettre,

277 Daijiang J. Liu, Ken Sin Lo, Yingchang Lu, Eirini Marouli, M. Carolina Medina-Gomez,

278 Fernando Rivadeneira, Andrew R. Wood.
279
280 *UK Biobank-based integration of height association signals group and heritability analyses*
281 Panos Deloukas, Timothy M. Frayling, Guillaume Lettre, Zoltán Kutalik, Ken Sin Lo, Eirini
282 Marouli, Sina Rüeger, Andrew R. Wood.
283
284 *Pleiotropy working group*
285 Gonçalo Abecasis, Michael Boehnke, James P. Cook, Panos Deloukas, Fotios Drenos, Jose C.
286 Florez, Heather Highland, Sekar Kathiresan, Cecilia M. Lindgren, Dajiang J. Liu, Ruth J.F.
287 Loos, Anubha Mahajan, Eirini Marouli, Mark I. McCarthy, Patricia B. Munroe, Gina M. Peloso,
288 John R. B. Perry, Katherine S. Ruth, Cristen J. Willer.
289
290 *Biological and clinical enrichment, and pathway analyses*
291 Rebecca S. Fine, Joel N. Hirschhorn, Zoltán Kutalik, David Lamparter, Guillaume Lettre, Ken
292 Sin Lo, Tune H. Pers.
293
294 *Functional characterization of STC2*
295 Troels R. Kjaer, Claus Oxvig.
296
297 **AUTHOR INFORMATION**
298 Summary genetic association results are available on the GIANT website:
299 http://portals.broadinstitute.org/collaboration/giant/index.php/GIANT_consortium. Reprints and
300 permissions information is available at www.nature.com/reprints. The authors declare no
301 competing financial interests. Correspondence and requests should be addressed to J.N.H.

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

303 (guillaume.lettre@umontreal.ca).

304 **References**

- 305 1 Fisher, R. A. The Correlation Between Relatives on the Supposition of Mendelian
306 Inheritance. *Transactions of the Royal Society of Edinburgh* **52**, 399-433 (1918).
- 307 2 Silventoinen, K. *et al.* Heritability of adult body height: a comparative study of twin
308 cohorts in eight countries. *Twin Res* **6**, 399-408 (2003).
- 309 3 Wood, A. R. *et al.* Defining the role of common variation in the genomic and biological
310 architecture of adult human height. *Nat Genet* **46**, 1173-1186 (2014).
- 311 4 Flannick, J. *et al.* Loss-of-function mutations in SLC30A8 protect against type 2 diabetes.
312 *Nat Genet* **46**, 357-363 (2014).
- 313 5 Steinthorsdottir, V. *et al.* Identification of low-frequency and rare sequence variants
314 associated with elevated or reduced risk of type 2 diabetes. *Nat Genet* **46**, 294-298
315 (2014).
- 316 6 Gudmundsson, J. *et al.* A study based on whole-genome sequencing yields a rare variant
317 at 8q24 associated with prostate cancer. *Nat Genet* **44**, 1326-1329 (2012).
- 318 7 Sidore, C. *et al.* Genome sequencing elucidates Sardinian genetic architecture and
319 augments association analyses for lipid and blood inflammatory markers. *Nat Genet* **47**,
320 1272-1281 (2015).
- 321 8 Danjou, F. *et al.* Genome-wide association analyses based on whole-genome sequencing
322 in Sardinia provide insights into regulation of hemoglobin levels. *Nat Genet* **47**, 1264-
323 1271 (2015).
- 324 9 Zuk, O. *et al.* Searching for missing heritability: designing rare variant association
325 studies. *Proc Natl Acad Sci U S A* **111**, E455-464 (2014).
- 326 10 Yang, J. *et al.* Genetic variance estimation with imputed variants finds negligible missing
327 heritability for human height and body mass index. *Nat Genet* **47**, 1114-1120 (2015).

328 11 Grove, M. L. *et al.* Best practices and joint calling of the HumanExome BeadChip: the
329 CHARGE Consortium. *PloS one* **8**, e68095 (2013).

330 12 Kryukov, G. V., Pennacchio, L. A. & Sunyaev, S. R. Most rare missense alleles are
331 deleterious in humans: implications for complex disease and association studies.
332 *American journal of human genetics* **80**, 727-739 (2007).

333 13 Tennessen, J. A. *et al.* Evolution and functional impact of rare coding variation from deep
334 sequencing of human exomes. *Science* **337**, 64-69 (2012).

335 14 Lanktree, M. B. *et al.* Meta-analysis of Dense Genecentric Association Studies Reveals
336 Common and Uncommon Variants Associated with Height. *American journal of human*
337 *genetics* **88**, 6-18 (2011).

338 15 Pers, T. H. *et al.* Biological interpretation of genome-wide association studies using
339 predicted gene functions. *Nat Commun* **6**, 5890 (2015).

340 16 Lamparter, D., Marbach, D., Rueedi, R., Kutalik, Z. & Bergmann, S. Fast and Rigorous
341 Computation of Gene and Pathway Scores from SNP-Based Summary Statistics. *PLoS*
342 *Comput Biol* **12**, e1004714 (2016).

343 17 Schwartz, N. B. & Domowicz, M. Chondrodysplasias due to proteoglycan defects.
344 *Glycobiology* **12**, 57R-68R (2002).

345 18 Wood, A. R. *et al.* Defining the role of common variation in the genomic and biological
346 architecture of adult human height. *Nat Genet* **46**, 1173-1186 (2014).

347 19 Wei, H. S., Wei, H. L., Zhao, F., Zhong, L. P. & Zhan, Y. T. Glycosyltransferase
348 GLT8D2 positively regulates ApoB100 protein expression in hepatocytes. *Int J Mol Sci*
349 **14**, 21435-21446 (2013).

350 20 Ito, H. *et al.* Molecular cloning and biological activity of a novel lysyl oxidase-related
351 gene expressed in cartilage. *J Biol Chem* **276**, 24023-24029 (2001).

352 21 Wakahara, T. *et al.* Fibin, a novel secreted lateral plate mesoderm signal, is essential for
353 pectoral fin bud initiation in zebrafish. *Dev Biol* **303**, 527-535 (2007).

354 22 Kawano, Y. & Kypta, R. Secreted antagonists of the Wnt signalling pathway. *J Cell Sci*
355 **116**, 2627-2634 (2003).

356 23 Mastaitis, J. *et al.* Loss of SFRP4 Alters Body Size, Food Intake, and Energy Expenditure
357 in Diet-Induced Obese Male Mice. *Endocrinology* **156**, 4502-4510 (2015).

358 24 Jepsen, M. R. *et al.* Stanniocalcin-2 inhibits mammalian growth by proteolytic inhibition
359 of the insulin-like growth factor axis. *J Biol Chem* **290**, 3430-3439 (2015).

360 25 Dauber, A. *et al.* Mutations in pregnancy-associated plasma protein A2 cause short
361 stature due to low IGF-I availability. *EMBO Mol Med* (2016).

362 26 Lango Allen, H. *et al.* Hundreds of variants clustered in genomic loci and biological
363 pathways affect human height. *Nature* **467**, 832-838 (2010).

364 27 Karaplis, A. C. *et al.* Inactivating mutation in the human parathyroid hormone receptor
365 type 1 gene in Blomstrand chondrodysplasia. *Endocrinology* **139**, 5255-5258 (1998).

366 28 Sims, N. A. *et al.* Interleukin-11 receptor signaling is required for normal bone
367 remodeling. *J Bone Miner Res* **20**, 1093-1102 (2005).

368 29 Takeuchi, Y. *et al.* Interleukin-11 as a stimulatory factor for bone formation prevents
369 bone loss with advancing age in mice. *J Biol Chem* **277**, 49011-49018 (2002).

370 30 Fuchsberger, C. *et al.* The genetic architecture of type 2 diabetes. *Nature* **536**, 41-47
371 (2016).

372 31 Bulik-Sullivan, B. K. *et al.* LD Score regression distinguishes confounding from
373 polygenicity in genome-wide association studies. *Nat Genet* **47**, 291-295 (2015).

374 32 Goldstein, J. I. *et al.* zCall: a rare variant caller for array-based genotyping: genetics and
375 population analysis. *Bioinformatics* **28**, 2543-2545 (2012).

376 33 Liu, D. J. *et al.* Meta-analysis of gene-level tests for rare variant association. *Nat Genet*
377 **46**, 200-204 (2014).

378 34 Winkler, T. W. & Day, F. R. Quality control and conduct of genome-wide association
379 meta-analyses. **9**, 1192-1212 (2014).

380 35 Yang, J. *et al.* Genomic inflation factors under polygenic inheritance. *European Journal*
381 *of Human Genetics* **19**, 807-812 (2011).

382 36 Feng, S., Liu, D., Zhan, X., Wing, M. K. & Abecasis, G. R. RAREMETAL: fast and
383 powerful meta-analysis for rare variants. *Bioinformatics (Oxford, England)* **30**, 2828-
384 2829 (2014).

385 37 Yang, J. *et al.* Conditional and joint multiple-SNP analysis of GWAS summary statistics
386 identifies additional variants influencing complex traits. *Nature genetics* **44**, 369-S363
387 (2012).

388 38 Loh, P. R. *et al.* Efficient Bayesian mixed-model analysis increases association power in
389 large cohorts. *Nat Genet* **47**, 284-290 (2015).

390 39 Pasaniuc, B. *et al.* Fast and accurate imputation of summary statistics enhances evidence
391 of functional enrichment. *Bioinformatics* **30**, 2906-2914 (2014).

392 40 Moayyeri, A., Hammond, C. J., Valdes, A. M. & Spector, T. D. Cohort Profile: TwinsUK
393 and healthy ageing twin study. *Int J Epidemiol* **42**, 76-85 (2013).

394 41 Boyd, A. *et al.* Cohort Profile: the 'children of the 90s'--the index offspring of the Avon
395 Longitudinal Study of Parents and Children. *Int J Epidemiol* **42**, 111-127 (2013).

396 42 Willer, C. J., Li, Y. & Abecasis, G. R. METAL: Fast and efficient meta-analysis of
397 genomewide association scans. *Bioinformatics* **26**, 2190-2191 (2010).

398 43 Purcell, S. M. *et al.* A polygenic burden of rare disruptive mutations in schizophrenia.
399 *Nature* **506**, 185-190 (2014).

400 44 Wu, M. C. *et al.* Rare-variant association testing for sequencing data with the sequence
401 kernel association test. *American journal of human genetics* **89**, 82-93 (2011).

402 45 Price, A. L. *et al.* Pooled association tests for rare variants in exon-resequencing studies.
403 *Am J Hum Genet* **86**, 832-838 (2010).

404 46 Nikpay, M. *et al.* A comprehensive 1,000 Genomes-based genome-wide association
405 meta-analysis of coronary artery disease. *Nat Genet* **47**, 1121-1130 (2015).

406 47 Fehrmann, R. S. *et al.* Gene expression analysis identifies global gene dosage sensitivity
407 in cancer. *Nat Genet* **47**, 115-125 (2015).

408 48 Frey, B. J. & Dueck, D. Clustering by passing messages between data points. *Science*
409 **315**, 972-976 (2007).

410 49 Overgaard, M. T. *et al.* Expression of recombinant human pregnancy-associated plasma
411 protein-A and identification of the proform of eosinophil major basic protein as its
412 physiological inhibitor. *The Journal of biological chemistry* **275**, 31128-31133 (2000).

413 50 Gyruup, C. & Oxvig, C. Quantitative analysis of insulin-like growth factor-modulated
414 proteolysis of insulin-like growth factor binding protein-4 and -5 by pregnancy-
415 associated plasma protein-A. *Biochemistry* **46**, 1972-1980 (2007).

416 51 Oxvig, C., Sand, O., Kristensen, T., Kristensen, L. & Sottrup-Jensen, L. Isolation and
417 characterization of circulating complex between human pregnancy-associated plasma
418 protein-A and proform of eosinophil major basic protein. *Biochimica et biophysica acta*
419 **1201**, 415-423 (1994).
420
421

422 **Figure legends**

423 **Figure 1.** Variants with a larger effect size on height variation tend to be rarer. We observed an
424 inverse relationship between the effect size (from the combined “discovery+validation” analysis,
425 in cm on the *y*-axis) and the minor allele frequency (MAF) for the height variants (*x*-axis, from 0
426 to 50%). We included in this figure the 606 height variants with $P < 2 \times 10^{-7}$.

427

428 **Figure 2.** Heat map showing subset of DEPICT gene set enrichment results. The full heat map is
429 available as **Extended Data Fig. 7**. For any given square, the color indicates how strongly the
430 corresponding gene (shown on the *x*-axis) is predicted to belong to the reconstituted gene set (*y*-
431 axis). This value is based on the gene’s *Z*-score for gene set inclusion in DEPICT’s reconstituted
432 gene sets, where red indicates a higher *Z*-score and blue indicates a lower one. The proteoglycan
433 binding pathway (bold) was uniquely implicated by coding variants by DEPICT and PASCAL.
434 To visually reduce redundancy and increase clarity, we chose one representative "meta-gene set"
435 for each group of highly correlated gene sets based on affinity propagation clustering
436 (**Supplementary Information**). Heat map intensity and DEPICT *P*-values correspond to the
437 most significantly enriched gene set within the meta-gene set; meta-gene sets are listed with their
438 database source. Annotations for the genes indicate whether the gene has OMIM annotation as
439 underlying a disorder of skeletal growth (black and grey) and the minor allele frequency of the
440 significant ExomeChip (EC) variant (shades of blue; if multiple variants, the lowest-frequency
441 variant was kept). Annotations for the gene sets indicate if the gene set was also found
442 significant for EC by PASCAL (yellow and grey) and if the gene set was found significant by
443 DEPICT for EC only or for both EC and GWAS (purple and green). Abbreviations: GO: Gene

444 Ontology; MP: mouse phenotype in the Mouse Genetics Initiative; PPI: protein-protein
445 interaction in the InWeb database.
446
447 **Figure 3.** STC2 mutants p.Arg44Leu (R44L) and p.Met86Ile (M86I) show compromised
448 proteolytic inhibition of PAPP-A. (A) Schematic representation of the role of STC2 in IGF-1
449 signaling. Partial inactivation of STC2 by height-associated DNA sequence variation could
450 increase bioactive IGF-1 through reduced inhibition of PAPP-A. (B) Western blot analysis of
451 recombinant STC2 wild-type and variants R44L and M86I. (C) Covalent complex formation
452 between PAPP-A and STC2 wild-type or variants R44L and M86I. Separately synthesized
453 proteins were analyzed by PAPP-A Western blotting following incubation for 8 h. In the absence
454 of STC2 (Mock lane), PAPP-A appears as a single 400 kDa band (*). Following incubation with
455 wild-type STC2, the majority of PAPP-A is present as the approximately 500 kDa covalent
456 PAPP-A:STC2 complex (#), in which PAPP-A is devoid of proteolytic activity towards IGFBP-
457 4. Under similar conditions, incubation with variants R44L or M86I appeared to cause less
458 covalent complex formation with PAPP-A. The gels are representative of at least three
459 independent experiments. (D) PAPP-A proteolytic cleavage of IGFBP-4 following incubation
460 with wild-type STC2 or variants for 1-24 h. Wild-type STC2 causes reduction in PAPP-A
461 activity, with complete inhibition of activity following 24 h incubation. Both STC2 variants
462 show increased IGFBP-4 cleavage (*i.e.* less inhibition) for all time points analyzed. Mean and
463 standard deviations of three independent experiments are shown. One-way repeated measures
464 analysis of variance followed by Dunnett's post-test showed significant differences between
465 STC2 wild-type and variants R44L ($P<0.001$) and M86I ($P<0.01$).
466

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

468

469 **Extended Data Figure 2.** Height ExomeChip association results. (A) Quantile-quantile plot of
470 ExomeChip variants and their association to adult height under an additive genetic model in
471 individuals of European ancestry. We stratified results based on allele frequency. (B) Manhattan
472 plot of all ExomeChip variants and their association to adult height under an additive genetic
473 model in individuals of European ancestry with a focus on the 553 independent SNPs, of which
474 469 have MAF>5% (grey), 55 have MAF between 1 and 5% (green), and 29 have MAF<1%
475 (blue). (C) Linkage disequilibrium (LD) score regression analysis for the height association
476 results in European-ancestry studies. In the plot, each point represents an LD Score quantile,
477 where the x-axis of the point is the mean LD Score of variants in that quantile and the y-axis is
478 the mean χ^2 statistic of variants in that quantile. The LD Score regression slope of the black line
479 is calculated based on Equation 1 in Bulik-Sullivan et al.³¹ which is estimated upwards due to the
480 small number of common variants (N=15,848) and the design of the ExomeChip. The LD score
481 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)
482 / (mean χ^2 -1) is 0.067 (standard error=0.012). (D) Scatter plot comparison of the effect sizes for
483 all variants that reached significance in the European-ancestry discovery results (N=381,625) and
484 results including only studies with sample sizes >5000 individuals (N=241,453).

485

486 **Extended Data Figure 3.** Height ExomeChip association results in African-ancestry
487 populations. Among the all-ancestry results, we found eight variants for which the genetic
488 association with height is mostly driven by individuals of African ancestry. The minor allele
489 frequency of these variants is <1% (or monomorphic) in all ancestries except African-ancestry

490 individuals. In individuals of African ancestry, the variants had allele frequencies between 9 and
491 40%.

492

493 **Extended Data Figure 4.** Concordance between direct conditional effect sizes using UK
494 Biobank (x-axis) and conditional analysis performed using a combination of imputation-based
495 methodology and approximate conditional analysis (SSimp, y-axis). The Pearson's correlation
496 coefficient is $r=0.85$. The dashed line indicates the identity line. The 95% confidence interval is
497 indicated in both directions. Red, SNPs with $P_{\text{cond}}>0.05$ in the UK Biobank; Green, SNPs with
498 $P_{\text{cond}}\leq 0.05$ in the UK Biobank.

499

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

506

507 **Extended Data Figure 6.** Comparison of DEPICT gene set enrichment results based on coding
508 variation from ExomeChip (EC) or non-coding variation from genome-wide association study
509 data (GWAS). The x-axis indicates the P-value for enrichment of a given gene set using DEPICT
510 adapted for EC data, where the input to DEPICT is the genes implicated by coding EC variants
511 that are independent of known GWAS signals. The y-axis indicates the P-value for gene set
512 enrichment using DEPICT, using as input the GWAS loci that do not overlap the coding

513 signals. Each point represents a meta-gene set, and the best P-value for any gene set within the
514 meta-gene set is shown. Only significant (false discovery rate < 0.01) gene set enrichment results
515 are plotted. Colors correspond to whether the meta-gene set was significant for EC only (blue),
516 GWAS only (green), both but more significant for EC (purple), or both but more significant for
517 GWAS (orange), and the most significant gene sets within each category are labeled. A line is
518 drawn at $x = y$ for ease of comparison.

519

520 **Extended Data Figure 7.** Heat map showing entire DEPICT gene set enrichment results
521 (analogous to **Fig. 2** in the main text). For any given square, the color indicates how strongly the
522 corresponding gene (shown on the x-axis) is predicted to belong to the reconstituted gene set (y-
523 axis). This value is based on the gene's Z-score for gene set inclusion in DEPICT's reconstituted
524 gene sets, where red indicates a higher Z-score and blue indicates a lower one. The proteoglycan
525 binding pathway was uniquely implicated by coding variants (as opposed to common variants)
526 by both DEPICT and the Pascal method. To visually reduce redundancy and increase clarity, we
527 chose one representative "meta-gene set" for each group of highly correlated gene sets based on
528 affinity propagation clustering (see **Methods** and **Supplementary Information**). Heat map
529 intensity and DEPICT p-values correspond to the most significantly enriched gene set within the
530 meta-gene set; meta-gene sets are listed with their database source. Annotations for the genes
531 indicate whether the gene has OMIM annotation as underlying a disorder of skeletal growth
532 (black and grey) and the minor allele frequency of the significant EC variant (shades of blue; if
533 multiple variants, the lowest-frequency variant was kept). Annotations for the gene sets indicate
534 if the gene set was also found significant for EC by the Pascal method (yellow and grey) and if
535 the gene set was found significant by DEPICT for EC only or for both EC and GWAS (purple

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

539

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

549 **METHODS**

550 *Study design & participants*

551 The discovery cohort consisted of 147 studies comprising 458,927 adult individuals of the
552 following ancestries: 1) European descent (N=381,625), 2) African (N=27,494), 3) South Asian
553 (N=29,591), 4) East Asian (N=8,767); 5) Hispanic (N=10,776) and 6) Saudi (N=695). All
554 participating institutions and coordinating centers approved this project, and informed consent
555 was obtained from all subjects. Discovery meta-analysis was carried out in each ancestry group
556 (except the Saudi) separately as well as in the All group. Validation was undertaken in
557 individuals of European ancestry only (**Supplementary Tables 1-3**). Conditional analyses were
558 undertaken only in the European descent group (106 studies, N=381,625).

559

560 *Phenotype*

561 Height (in centimeters) was corrected for age and the genomic principal components (derived
562 from GWAS data, the variants with MAF >1% on ExomeChip, or ancestry informative markers
563 available on the ExomeChip), as well as any additional study-specific covariates (e.g. recruiting
564 center), in a linear regression model. For studies with non-related individuals, residuals were
565 calculated separately by sex, whereas for family-based studies sex was included as a covariate in
566 the model. Additionally, residuals for case/control studies were calculated separately. Finally,
567 residuals were subject to inverse normal transformation.

568

569 *Genotype calling*

570 The majority of studies followed a standardized protocol and performed genotype calling using
571 the designated manufacturer software, which was then followed by zCall³². For 10 studies

572 participating in the Cohorts for Heart and Aging Research in Genomic Epidemiology
573 (CHARGE) Consortium, the raw intensity data for the samples from seven genotyping centers
574 were assembled into a single project for joint calling¹¹. Study-specific quality control (QC)
575 measures of the genotyped variants was implemented before association analysis
576 **(Supplementary Tables 1-2).**

577

578 *Study-level statistical analyses*

579 Individual cohorts were analyzed separately for each ancestry population, with either
580 RAREMETALWORKER (<http://genome.sph.umich.edu/wiki/RAREMETALWORKER>) or
581 RVTEST (<http://zhanxw.github.io/rvtests/>), to associate inverse normal transformed height data
582 with genotype data taking potential cryptic relatedness (kinship matrix) into account in a linear
583 mixed model. These software are designed to perform score-statistics based rare-variant
584 association analysis, can accommodate both unrelated and related individuals, and provide
585 single-variant results and variance-covariance matrix. The covariance matrix captures linkage
586 disequilibrium (LD) relationships between markers within 1 Mb, which is used for gene-level
587 meta-analyses and conditional analyses³³. Single-variant analyses were performed for both
588 additive and recessive models.

589

590 *Centralized quality-control*

591 The individual study data were investigated for potential existence of ancestry population
592 outliers based on 1000 Genome Project phase 1 ancestry reference populations. A centralized QC
593 procedure implemented in EasyQC³⁴ was applied to individual study association summary
594 statistics to identify outlying studies: (1) assessment of possible problems in height

595 transformation, (2) comparison of allele frequency alignment against 1000 Genomes Project
596 phase 1 reference data to pinpoint any potential strand issues, and (3) examination of quantile-
597 quantile (QQ) plots per study to identify any problems arising from population stratification,
598 cryptic relatedness and genotype biases. We excluded variants if they had call rate <95%, Hardy-
599 Weinberg equilibrium $P < 1 \times 10^{-7}$, or large allele frequency deviations from reference populations
600 (>0.6 for all ancestry analyses and >0.3 for ancestry-specific population analyses). We also
601 excluded from downstream analyses markers not present on the Illumina ExomeChip array 1.0,
602 variants on the Y-chromosome or the mitochondrial genome, indels, multiallelic variants, and
603 problematic variants based on the Blat-based sequence alignment analyses. Meta-analyses were
604 carried out in parallel by two different analysts at two sites.

605

606 *Single-variant meta-analyses*

607 *Discovery analyses.* We conducted single-variant meta-analyses in a discovery sample of
608 458,927 individuals of different ancestries using both additive and recessive genetic models
609 (**Extended Data Fig. 1** and **Supplementary Tables 1-4**). Significance for single-variant
610 analyses was defined at array-wide level ($P < 2 \times 10^{-7}$, Bonferroni correction for 250,000 variants).
611 The combined additive analyses identified 1,455 unique variants that reached array-wide
612 significance ($P < 2 \times 10^{-7}$), including 578 non-synonymous and splice site variants
613 (**Supplementary Tables 5-7**). Under the additive model, we observed a high genomic inflation
614 of the test statistics (*e.g.* λ_{GC} of 2.7 in European-ancestry studies for common markers, **Extended**
615 **Data Fig. 2** and **Supplementary Table 8**), although validation results (see below) and additional
616 sensitivity analyses (see below) suggested that it is consistent with polygenic inheritance as
617 opposed to population stratification, cryptic relatedness, or technical artifacts (**Extended Data**

618 **Fig. 2).** The majority of these 1,455 association signals (1,241; 85.3%) were found in the
619 European-ancestry meta-analysis (85.5% of the discovery sample size) (**Extended Data Fig. 2**).
620 Nevertheless, we discovered eight associations within five loci in our all-ancestry analyses that
621 are driven by African studies (including one missense variant in the growth hormone gene *GHI*
622 (rs151263636), **Extended Data Fig. 3**), three height variants found only in African studies, and
623 one rare missense marker associated with height in South Asians only (**Supplementary Table**
624 **7**).

625
626 *Genomic inflation and confounding.* We observed a marked genomic inflation of the test
627 statistics even after adequate control for population stratification (linear mixed model) arising
628 mainly from common markers; λ_{GC} in European-ancestry was 1.2 and 2.7 for all and common
629 markers, respectively (**Extended Data Fig. 2** and **Supplementary Table 8**). Such inflation is
630 expected for a highly polygenic trait like height, and is consistent with our very large sample
631 size^{3,35}. To confirm this, we applied the recently developed linkage disequilibrium (LD) score
632 regression method to our height ExomeChip results³¹, with the caveats that the method was
633 developed (and tested) with >200,000 common markers available. We restricted our analyses to
634 15,848 common variants (MAF $\geq 5\%$) from the European-ancestry meta-analysis, and matched
635 them to pre-computed LD scores for the European reference dataset³¹. The intercept of the
636 regression of the χ^2 statistics from the height meta-analysis on the LD score estimate the inflation
637 in the mean χ^2 due to confounding bias, such as cryptic relatedness or population stratification.
638 The intercept was 1.4 (standard error = 0.07), which is small when compared to the λ_{GC} of 2.7.
639 Furthermore, we also confirmed that the LD score regression intercept is estimated upward
640 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 range³¹, 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-package³⁶ and the GCTA v1.24³⁷ 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 ($P < 2 \times 10^{-7}$) 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 $P < 2 \times 10^{-7}$. 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 $P < 2 \times 10^{-7}$).

667

668 To discover new height variants, we conditioned the height variants found in our ExomeChip
669 study on the previously published GWAS height variants³ using the first release of the UK
670 Biobank imputed dataset and regression methodology implemented in BOLT-LMM³⁸. 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 $P_{\text{conditional}} < 0.05$ to declare a height variant
673 as independent in this analysis. We also explored an alternative approach based on approximate
674 conditional analysis³⁷. This latter method (SSimp) relies on summary statistics available from the
675 same cohort, thus we first imputed summary statistics³⁹ for exome variants, using summary
676 statistics from the Wood *et al.* 2014 study³. 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 ($r^2 < 0.8$), 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 (TwinsUK⁴⁰ and ALSPAC⁴¹ 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 RAREMETAL³³. 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 METAL⁴². Significant associations were defined as those with a combined meta-analysis
698 (discovery and validation) $P_{\text{combined}} < 2 \times 10^{-7}$.

699
700 We considered 81 variants with suggestive association in the discovery analyses (2×10^{-6}
701 $< P_{\text{discovery}} \leq 2 \times 10^{-5}$). Of those 81 variants, 55 reached significance after combining discovery and
702 replication results based on $P_{\text{combined}} < 2 \times 10^{-7}$ (**Supplementary Table 9**). Furthermore, recessive
703 modeling confirmed seven new independent markers with $P_{\text{combined}} < 2 \times 10^{-7}$ (**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” modeling³⁷ in the combined discovery and validation set. This

710 resulted in the identification of 606 independent height variants, including 252 non-synonymous
711 or splice site variants (**Supplementary Table 11**). If we only consider the initial set of lead
712 SNPs with $P < 2 \times 10^{-7}$, we identified 561 independent variants. Of these 561 variants (selected
713 without the validation studies), 560 have concordant direction of effect between the discovery
714 and validation studies, and 548 variants have a $P_{\text{validation}} < 0.05$ (466 variants with $P_{\text{validation}}$
715 $< 8.9 \times 10^{-5}$, Bonferroni correction for 561 tests), suggesting a very low false discovery rate
716 (**Supplementary Table 11**).

717

718 *Gene-based association meta-analyses*

719 For the gene-based analyses, we applied two different sets of criteria to select variants, based on
720 coding variant annotation from five prediction algorithms (PolyPhen2 HumDiv and HumVar,
721 LRT, MutationTaster and SIFT)⁴³. The mask labeled “*broad*” included variants with a MAF
722 < 0.05 that are nonsense, stop-loss, splice site, as well as missense variants that are annotated as
723 damaging by at least one program mentioned above. The mask labeled “*strict*” included only
724 variants with MAF < 0.05 that are nonsense, stop-loss, splice site, as well as missense variants
725 annotated as damaging by all five algorithms. We used two tests for gene-based testing, namely
726 the SKAT⁴⁴ and VT⁴⁵ tests. Statistical significance for gene-based tests was set at a Bonferroni-
727 corrected threshold of $P < 5 \times 10^{-7}$ (threshold for 25,000 genes and four tests). The gene-based
728 discovery results were validated (same test and variants, when possible) in the same eight studies
729 genotyped on the ExomeChip (N=59,804 participants) that were used for the validation of the
730 single-variant results (see above, and **Supplementary Tables 1-3**). Gene-based conditional
731 analyses were performed in RAREMETAL.

732

733 *Pleiotropy analyses*

734 We accessed ExomeChip data from GIANT (BMI, waist-hip ratio), GLGC (total cholesterol
735 (TC), triglycerides (TG), HDL-cholesterol (HDL-C), LDL-cholesterol (LDL-C)), IBPC (systolic
736 and diastolic blood pressure), MAGIC (glycaemic traits), REPROGEN (age at menarche and
737 menopause), and DIAGRAM (type 2 diabetes). For coronary artery disease, we accessed 1000
738 Genomes Project-imputed GWAS data released by CARDIoGRAMplusC4D⁴⁶.

739

740 *Pathway analyses*

741 DEPICT is a computational framework that uses probabilistically-defined reconstituted gene sets
742 to perform gene set enrichment and gene prioritization¹⁵. For a description about gene set
743 reconstitution please refer to references ¹⁵ and ⁴⁷. In brief, reconstitution was performed by
744 extending pre-defined gene sets (such as Gene Ontology terms, canonical pathways, protein-
745 protein interaction subnetworks and rodent phenotypes) with genes co-regulated with genes in
746 these pre-defined gene set using large-scale microarray-based transcriptomics data. In order to
747 adapt the gene set enrichment part of DEPICT for ExomeChip data, we made two principal
748 changes. First and foremost, because DEPICT for GWAS incorporates all genes within a given
749 LD block around each index SNP, we modified DEPICT to take as input only the gene directly
750 impacted by the coding SNP. Second, we adapted the way DEPICT adjust for confounders (such
751 as gene length) by generating null ExomeChip association results using Swedish ExomeChip
752 data (Malmö Diet and Cancer (MDC), All New Diabetics in Scania (ANDIS), and Scania
753 Diabetes Registry (SDR) cohorts, N=11,899) and randomly assigning phenotypes from a normal
754 distribution before conducting association analysis (see **Supplementary Information**). For the
755 gene set enrichment analysis of the ExomeChip data, we used significant non-synonymous

756 variants statistically independent of known GWAS hits (and that were present in the null
757 ExomeChip data; see **Supplementary Information** for details). For gene set enrichment analysis
758 of the GWAS data, we used all loci (1) with a non-coding index SNP and (2) that did not contain
759 any of the novel ExomeChip genes. In visualizing the analysis, we used affinity propagation
760 clustering⁴⁸ to group the most similar reconstituted gene sets based on their gene memberships
761 (see **Supplementary Information**). Within a “meta-gene set”, the best P-value of any member
762 gene set was used as representative for comparison. DEPICT for ExomeChip was written using
763 the Python programming language and the code can be found at
764 <https://github.com/RebeccaFine/height-ec-depict>.

765
766 We also applied the PASCAL pathway analysis tool¹⁶ to association summary statistics for all
767 coding variants. In brief, the method derives gene-based scores (both SUM and MAX statistics)
768 and subsequently tests for the over-representation of high gene scores in predefined biological
769 pathways. We used standard pathway libraries from KEGG, REACTOME and BIOCARTA, and
770 also added dichotomized (Z-score>3) reconstituted gene sets from DEPICT¹⁵. To accurately
771 estimate SNP-by-SNP correlations even for rare variants, we used the UK10K data (TwinsUK⁴⁰
772 and ALSPAC⁴¹ studies, N=3781). In order to separate the contribution of regulatory variants
773 from the coding variants, we also applied PASCAL to association summary statistics of only
774 regulatory variants (20 kb upstream, gene body excluded) from the Wood et al. study³. In this
775 way, we could classify pathways driven principally by coding, regulatory or mixed signals.

776

777 ***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 (6×10^6) were plated onto 10 cm-dishes and transfected 18 h later by calcium phosphate
786 coprecipitation using 10 μ g 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 previously²⁴. PAPP-A was expressed stably
789 in HEK293T cells as previously reported⁴⁹. 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 ¹²⁵I-labeled IGFBP-4 is
798 described in detail elsewhere⁵⁰. Briefly, the PAPP-A:STC2 complex mixtures were diluted
799 (1:190) to a concentration of 145 pM PAPP-A and mixed with preincubated ¹²⁵I-IGFBP4 (10
800 nM) and IGF-1 (100 nM) in 50 mM Tris-HCl, 100 mM NaCl, 1 mM CaCl₂. Following 1 h

801 incubation at 37°C, reactions were terminated by the addition of SDS-PAGE sample buffer
802 supplemented with 25 mM EDTA. Substrate and co-migrating cleavage products were separated
803 by 12% nonreducing SDS-PAGE and visualized by autoradiography using a storage phosphor
804 screen (GE Healthcare) and a Typhoon imaging system (GE Healthcare). Band intensities were
805 quantified using ImageQuant TL 8.1 software (GE Healthcare).

806

807 *Western blotting.* STC2 and covalent complexes between STC2 and PAPP-A were blotted onto
808 PVDF membranes (Millipore) following separation by 3-8% SDS-PAGE. The membranes were
809 blocked with 2% Tween-20, and equilibrated in 50 mM Tris-HCl, 500 mM NaCl, 0.1% Tween-
810 20, pH 9 (TST). For STC2, the membranes were incubated with goat polyclonal anti-STC2
811 (R&D systems, AF2830) at 0.5 µg/ml in TST supplemented with 2% skim milk for 1 h at 20°C.
812 For PAPP-A:STC2 complexes, the membranes were incubated with rabbit polyclonal anti-
813 PAPP-A⁵¹ at 0.63 µg/ml in TST supplemented with 2% skim milk for 16 h at 20°C. Membranes
814 were washed with TST and subsequently incubated with polyclonal swine anti-rabbit IgG-HRP
815 (DAKO, P0217) or polyclonal rabbit anti-goat IgG-HRP (DAKO, P0449), respectively, diluted
816 1:2000 in TST supplemented with 2% skim milk for 1 h at 20°C. Following washing with TST,
817 membranes were developed using enhanced chemiluminescence (ECL Prime, GE Healthcare).
818 Images were captured using an ImageQuant LAS 4000 instrument (GE Healthcare).

819

820

821 **DATA AVAILABILITY STATEMENT**

822 Summary genetic association results are available on the GIANT website:

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

824

825 **URLs**

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

827 DEPICT, <http://www.broadinstitute.org/mpg/depict/>

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

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

830 OMIM, <http://omim.org/>

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

832 RAREMETALWORKER, <http://genome.sph.umich.edu/wiki/RAREMETALWORKER>

833 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 $P < 5 \times 10^{-7}$, (2) the gene does not include variants with $P < 2 \times 10^{-7}$, and (3) the gene-based P-value is at least two orders of magnitude smaller than the P-value for the most significant variant within the gene. For each gene, we provide P-values for the four different gene-based tests applied. P-values in bold are the most significant results for a given gene. ¹Validation (N=59,804) and combined results using the same test and (when possible) variants. ²When 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). ³If 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				
<i>OSGIN1</i>	4.3x10⁻¹¹	4.5x10 ⁻⁵	0.19	0.18	0.048	2.6x10 ⁻¹²	7.7x10 ⁻¹¹	Known locus. No predicted causal genes.
<i>CRISPLD1</i>	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. <i>CRISPLD1</i> was predicted to be causal.
<i>CSAD</i>	2.3x10 ⁻⁸	2.4x10⁻⁹	0.83	0.59	0.54	2.0x10 ⁻⁹	NA	New locus.
<i>SNED1</i>	1.9x10 ⁻⁵	4.3x10⁻⁹	NA	NA	0.083	4.5x10 ⁻¹⁰	1.4x10 ⁻⁹	Known locus. <i>SNED1</i> was not predicted to be causal.
<i>G6PC</i>	1.3x10 ⁻⁵	3.6x10⁻⁸	5.5x10 ⁻⁶	1.3x10 ⁻⁶	0.24	5.2x10 ⁻⁸	3.9x10 ⁻⁸	Known locus, <i>G6PC</i> was not predicted to be causal. <i>G6PC</i> is mutated in glycogen storage disease Ia.
<i>NOX4</i>	5.1x10 ⁻⁶	1.4x10⁻⁷	NA	NA	0.013	5.5x10 ⁻⁹	NA	New locus.
<i>UGGT2</i>	3.0x10 ⁻⁵	2.6x10⁻⁷	2.3x10 ⁻⁵	4.8x10 ⁻⁷	0.64	3.4x10 ⁻⁷	NA	New locus.
<i>FLNB</i>	2.2x10 ⁻⁶	5.1x10 ⁻⁴	2.4x10⁻⁹	3.2x10 ⁻⁶	0.016	8.6x10 ⁻¹¹	3.6x10 ⁻⁹	Known locus. <i>FLNB</i> was predicted to be causal. <i>FLNB</i> is mutated in atelosteogenesis type I.
<i>B4GALNT3</i>	2.4x10 ⁻⁵	1.9x10 ⁻⁵	1.8x10 ⁻⁵	3.1x10⁻⁷	0.79	4.3x10 ⁻⁷	7.7x10 ⁻⁷	Known locus. <i>B4GALNT3</i> was predicted to be causal.
<i>CCDC3</i>	6.3x10 ⁻⁴	6.3x10 ⁻⁶	3.0x10 ⁻⁷	5.4x10⁻⁹	0.080	1.2x10 ⁻⁹	1.6x10 ⁻⁹	Known locus. <i>CCDC3</i> 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 Thorleifsson¹⁷, Anne E Justice², David Lamparter^{16,18}, Kathleen E Stirrups^{1,19}, Valérie Turcot⁵, Kristin L Young², Thomas W Winkler²⁰, Tõnu Esko^{8,10,21}, Tugce Karaderi²², Adam E Locke^{23,24}, Nicholas GD Masca^{25,26}, Maggie CY Ng^{27,28}, Poorva Mudgal²⁷, Manuel A Rivas^{8,29}, Sailaja Vedantam⁸⁻¹⁰, Anubha Mahajan²², Xiuqing Guo³⁰, Goncalo Abecasis²³, Katja K Aben^{31,32}, Linda S Adair³³, Dewan S Alam³⁴, Eva Albrecht³⁵, Kristine H Allin³⁶, Matthew Allison³⁷, Philippe Amouyel³⁸⁻⁴⁰, Emil V Appel³⁶, Dominique Arveiler^{41,42}, Folkert W Asselbergs⁴³⁻⁴⁵, Paul L Auer⁴⁶, Beverley Balkau⁴⁷, Bernhard Banas⁴⁸, Lia E Bang⁴⁹, Marianne Benn^{50,51}, Sven Bergmann^{16,18}, Lawrence F Bielak⁵², Matthias Blüher^{53,54}, Heiner Boeing⁵⁵, Eric Boerwinkle^{56,57}, Carsten A Böger⁴⁸, Lori L Bonnycastle⁵⁸, Jette Bork-Jensen³⁶, Michiel L Bots⁵⁹, Erwin P Bottinger¹², Donald W Bowden^{27,28,60}, Ivan Brandslund^{61,62}, Gerome Breen⁶³, Murray H Brilliant⁶⁴, Linda Broer⁴, Amber A Burt⁶⁵, Adam S Butterworth^{66,67}, David J Carey⁶⁸, Mark J Caulfield^{1,69}, John C Chambers⁷⁰⁻⁷², Daniel I Chasman^{8,73-75}, Yii-Der Ida Chen³⁰, Rajiv Chowdhury⁶⁶, Cramer Christensen⁷⁶, Audrey Y Chu^{74,77}, Massimiliano Cocca⁷⁸, Francis S Collins⁵⁸, James P Cook⁷⁹, Janie Corley^{80,81}, Jordi Corominas Galbany⁸², 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 Borst⁸⁹, Simon de Denus^{5,90}, Mark CH de Groot^{91,92}, Renée de Mutsert⁹³, Ian J Deary^{80,81}, George Dedoussis⁹⁴, Ellen W Demerath⁹⁵, Anneke I den Hollander⁹⁶, Joe G Dennis⁹⁷, Emanuele Di Angelantonio^{66,67}, Fotios Drenos^{98,99}, Mengmeng Du^{100,101}, Alison M Dunning¹⁰², Douglas F Easton^{97,102}, Tapani Ebeling^{103,104}, Todd L Edwards¹⁰⁵, Patrick T Ellinor^{106,107}, Paul Elliott¹⁰⁸, Evangelos Evangelou^{71,109}, Aliko-Eleni Farmaki⁹⁴, Jessica D Faul¹¹⁰, Mary F Feitosa¹¹¹, Shuang Feng²³, Ele Ferrannini^{112,113}, Marco M Ferrario¹¹⁴, Jean Ferrieres¹¹⁵, Jose C Florez^{106,107,116}, Ian Ford¹¹⁷, Myriam Fornage¹¹⁸, Paul W Franks¹¹⁹⁻¹²¹, Ruth Frikke-Schmidt^{51,122}, Tessel E Galesloot³², Wei Gan²², Iliaria Gandin¹²³, Paolo Gasparini^{123,124}, Vilmantas Giedraitis¹²⁵, Ayush Giri¹⁰⁵, Giorgia Giroto^{123,124}, Scott D Gordon⁸⁵, Penny Gordon-Larsen^{126,127}, Mathias Gorski^{20,48}, Niels Grarup³⁶, Megan L. Grove⁵⁶, Vilmundur Gudnason^{128,129}, Stefan Gustafsson¹³⁰, Torben Hansen³⁶, Kathleen Mullan Harris^{126,131}, Tamara B Harris¹³², Andrew T Hattersley¹³³, Caroline Hayward¹³⁴, Liang He^{135,136}, Iris M Heid^{20,35}, Kauko Heikkilä^{136,137}, Øyvind Helgeland^{138,139}, Jussi Hernesniemi¹⁴⁰⁻¹⁴², Alex W Hewitt¹⁴³⁻¹⁴⁵, Lynne J Hocking^{146,147}, Mette Hollensted³⁶, Oddgeir L Holmen¹⁴⁸, G. Kees Hovingh¹⁴⁹, Joanna MM Howson⁶⁶, Carel B Hoyng⁹⁶, Paul L Huang¹⁰⁶, Kristian Hveem¹⁵⁰, M. Arfan Ikram^{3,151,152}, Erik Ingelsson^{130,153}, Anne U Jackson²³, Jan-Håkan Jansson^{154,155}, Gail P Jarvik^{65,156}, Gorm B Jensen¹⁵⁷, Min A Jhun⁵², Yucheng Jia³⁰, Xuejuan Jiang^{158,159}, Stefan Johansson^{139,160}, Marit E Jørgensen^{161,162}, Torben Jørgensen^{51,163,164}, Pekka Jousilahti¹⁶⁵, J Wouter Jukema^{166,167}, Bratati Kahali¹⁶⁸⁻¹⁷⁰, René S Kahn¹⁷¹, Mika Kähönen¹⁷², Pia R Kamstrup⁵⁰, Stavroula Kanoni¹, Jaakko Kaprio^{136,137,165}, Maria Karaleftheri¹⁷³, Sharon LR Kardina⁵², Fredrik Karpe^{174,175}, Frank Kee¹⁷⁶, Renske Keeman¹⁷⁷, Lambertus A Kiemeny³², Hidetoshi Kitajima²², Kirsten B Kluivers³², Thomas Kocher¹⁷⁸, Pirjo Komulainen¹⁷⁹, Jukka Kontto¹⁶⁵, Jaspal S Kooner^{70,72,180}, Charles Kooperberg¹⁸¹, Peter Kovacs⁵³, Jennifer Kriebel¹⁸²⁻¹⁸⁴, Helena Kuivaniemi^{68,185}, Sébastien Küry¹⁸⁶, Johanna Kuusisto¹⁸⁷, Martina La Bianca¹⁸⁸, Markku Laakso¹⁸⁷, Timo A Lakka^{179,189}, Ethan M Lange¹⁹⁰, Leslie A Lange¹⁹⁰, Carl D Langefeld¹⁹¹, Claudia Langenberg¹⁹², Eric B Larson^{65,193,194}, I-Te Lee¹⁹⁵⁻¹⁹⁷, Terho Lehtimäki^{141,142}, Cora E Lewis¹⁹⁸, Huaixing Li¹⁹⁹, Jin

Li²⁰⁰, Ruifang Li-Gao⁹³, Honghuang Lin²⁰¹, Li-An Lin¹¹⁸, Xu Lin¹⁹⁹, Lars Lind²⁰², Jaana Lindström¹⁶⁵, Allan Linneberg^{51,164,203}, Yeheng Liu³⁰, Yongmei Liu²⁰⁴, Artitaya Lophatananon²⁰⁵, Jian'an Luan¹⁹², Steven A Lubitz^{106,107}, Leo-Pekka Lyytikäinen^{141,142}, David A Mackey¹⁴⁴, Pamela AF Madden²⁰⁶, Alisa K Manning^{106,107,116}, Satu Männistö¹⁶⁵, Gaëlle Marenne⁸⁶, Jonathan Marten¹³⁴, Nicholas G Martin⁸⁵, Angela L Mazul², Karina Meidtnr^{182,207}, Andres Metspalu²¹, Paul Mitchell²⁰⁸, Karen L Mohlke¹⁹⁰, Dennis O Mook-Kanamori^{93,209}, Anna Morgan¹²³, Andrew D Morris²¹⁰, Andrew P Morris^{22,79}, Martina Müller-Nurasyid^{35,211,212}, Patricia B Munroe^{1,69}, Mike A Nalls²¹³, Matthias Nauck^{214,215}, Christopher P Nelson^{25,26}, Matt Neville^{174,175}, Sune F Nielsen^{50,51}, Kjell Nikus²¹⁶, Pål R Njølstad^{138,139}, Børge G Nordestgaard^{50,51}, Ioanna Ntalla¹, Jeffrey R O'Connell²¹⁷, Heikki Oksa²¹⁸, Loes M Olde Loohuis²¹⁹, Roel A Ophoff^{171,219}, Katharine R Owen^{174,175}, Chris J Packard¹¹⁷, Sandosh Padmanabhan¹¹⁷, Colin NA Palmer²²⁰, Gerard Pasterkamp^{221,222}, Aniruddh P Patel^{8,75,106}, Alison Pattie⁸¹, Oluf Pedersen³⁶, Peggy L Peissig⁶⁴, Gina M Peloso^{106,107}, Craig E Pennell²²³, Markus Perola^{165,224,225}, James A Perry²¹⁷, John R.B. Perry¹⁹², Thomas N Person⁶⁴, Ailith Pirie¹⁰², Ozren Polasek^{210,226}, Danielle Posthuma^{227,228}, Olli T Raitakari^{229,230}, Asif Rasheed²³¹, Rainer Rauramaa^{179,232}, Dermot F Reilly²³³, 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 Robino¹⁸⁸, Olov Rolandsson^{154,238}, Igor Rudan²¹⁰, Katherine S Ruth⁶, Danish Saleheen^{231,239}, Veikko Salomaa¹⁶⁵, Nilesh J Samani^{25,26}, Kevin Sandow³⁰, Yadav Sapkota⁸⁵, Naveed Sattar¹¹⁷, Marjanka K Schmidt¹⁷⁷, Pamela J Schreiner²⁴⁰, Matthias B Schulze^{182,207}, Robert A Scott¹⁹², Marcelo P Segura-Lepe⁷¹, Svati Shah²⁴¹, Xueling Sim^{23,242}, Suthesh Sivapalaratnam^{106,243,244}, Kerrin S Small²⁴⁵, Albert Vernon Smith^{128,129}, Jennifer A Smith⁵², Lorraine Southam^{22,86}, Timothy D Spector²⁴⁵, Elizabeth K Speliotes¹⁶⁸⁻¹⁷⁰, John M Starr^{80,246}, Valgerdur Steinthorsdottir¹⁷, Heather M Stringham²³, Michael Stumvoll^{53,54}, Praveen Surendran⁶⁶, Leen M 't Hart²⁴⁷⁻²⁴⁹, Katherine E Tansey^{250,251}, Jean-Claude Tardif^{5,237}, Kent D Taylor³⁰, Alexander Teumer²⁵², Deborah J Thompson⁹⁷, Unnur Thorsteinsdottir^{17,128}, Betina H Thuesen¹⁶⁴, Anke Tönjes²⁵³, Gerard Tromp^{68,254}, Stella Trompet^{166,255}, Emmanouil Tsafantakis²⁵⁶, Jaakko Tuomilehto^{165,257-259}, Anne Tybjaerg-Hansen^{51,122}, Jonathan P Tyrer¹⁰², Rudolf Uher²⁶⁰, André G Uitterlinden^{3,4}, Sheila Ulivi¹⁸⁸, Sander W van der Laan²²², Andries R Van Der Leij²⁶¹, Cornelia M van Duijn³, Natasja M van Schoor²⁴⁷, Jessica van Setten⁴³, Anette Varbo^{50,51}, Tibor V Varga¹¹⁹, Rohit Varma¹⁵⁹, Digna R Velez Edwards²⁶², Sita H Vermeulen³², Henrik Vestergaard³⁶, Veronique Vitart¹³⁴, Thomas F Vogt²⁶³, Diego Vozzi¹²⁴, Mark Walker²⁶⁴, Feijie Wang¹⁹⁹, Carol A Wang²²³, Shuai Wang²⁶⁵, Yiqin Wang¹⁹⁹, Nicholas J Wareham¹⁹², Helen R Warren^{1,69}, Jennifer Wessel²⁶⁶, Sara M Willems¹⁹², James G Wilson²⁶⁷, Daniel R Witte^{268,269}, Michael O Woods²⁷⁰, Ying Wu¹⁹⁰, Hanieh Yaghoobkar⁶, Jie Yao³⁰, Pang Yao¹⁹⁹, Laura M Yerges-Armstrong^{217,271}, Robin Young^{66,117}, Eleftheria Zeggini⁸⁶, Xiaowei Zhan²⁷², Weihua Zhang^{70,71}, Jing Hua Zhao¹⁹², Wei Zhao²³⁹, Wei Zhao⁵², He Zheng¹⁹⁹, 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 Rotter³⁰, Michael Boehnke²³, Sekar Kathiresan^{8,75,106}, Mark I McCarthy^{22,174,175}, Cristen J Willer^{168,169,273}, Kari Stefansson^{17,128}, Ingrid B Borecki¹¹¹, Dajiang J Liu²⁷⁴, Kari E North²⁷⁵, Nancy L Heard-Costa^{77,276}, Tune H Pers^{36,277}, Cecilia M Lindgren^{22,278}, Claus Oxvig⁷⁸, 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 Children's 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, Icahn 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, University 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 Épidémiologie et Santé 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, King's 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 Queensland, Brisbane, Queensland, 4072, Australia
85. QIMR Berghofer Medical Research Institute, Brisbane, Queensland, 4006, Australia
86. Wellcome Trust Sanger Institute, Wellcome Genome Campus, Hinxton, Cambridge, CB10 1SA, UK
87. British Heart Foundation, Cambridge Centre of Excellence, Department of Medicine, University of Cambridge, Cambridge, CB2 0QQ, UK
88. Department of Genetics, Center for Molecular Medicine, University Medical Center Utrecht, Utrecht, 3584 CX, The Netherlands
89. Department of Vascular Surgery, Division of Surgical Specialties, University Medical Center Utrecht, Utrecht, 3584 CX, The Netherlands
90. Faculty of Pharmacy, Université de Montréal, Montreal, Quebec, H3T 1J4, Canada
91. Department of Clinical Chemistry and Haematology, Division of Laboratory and Pharmacy, University Medical Center Utrecht, Utrecht, 3508 GA, The Netherlands
92. Utrecht Institute for Pharmaceutical Sciences, Division of 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 Bristol, 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. Echinus Medical Centre, Echinus, 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 Netherlands
210. Centre for Global Health Research, Usher Institute of Population Health Sciences and Informatics, University of Edinburgh, Edinburgh, EH8 9AG, UK
211. Department of Medicine I, Ludwig-Maximilians-Universität, Munich, 81377, Germany
212. DZHK (German Centre for Cardiovascular Research), partner site Munich Heart Alliance, Munich, 80802, Germany
213. Laboratory of Neurogenetics, National Institute on Aging, NIH, Bethesda, 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 Women's 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. Department 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, King's 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 Women's and Child Health, University of Leipzig, Leipzig, 04103, Germany
254. Division of Molecular Biology and Human Genetics, Department of Biomedical Sciences, Faculty of Medicine and Health Sciences, Stellenbosch University, Tygerberg, Western Cape, 7505, South Africa
255. Department of Gerontology and Geriatrics, Leiden University Medical Center, Leiden, 2333, The Netherlands
256. Anogia Medical Centre, Anogia, Greece
257. Centre for Vascular Prevention, Danube-University Krems, Krems, 3500, Austria
258. Dasman Diabetes Institute, Dasman, 15462, Kuwait
259. Diabetes Research Group, King Abdulaziz University, Jeddah, 21589, Saudi Arabia
260. Department of Psychiatry, Dalhousie University, Halifax, B3H 4R2, Canada
261. University of Amsterdam, Department of Brain & Cognition, Amsterdam, 1018 WS, The Netherlands
262. Department of Obstetrics and Gynecology, Institute for Medicine and Public Health, Vanderbilt Genetics Institute, Vanderbilt University, Nashville, 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