HGG Advances

ARTICLE

Whole-exome sequence analysis of anthropometric traits illustrates challenges in identifying effects of rare genetic variants

Kristin L. Young,[1,38,40,*] Virginia Fisher,[2,38] Xuan Deng,[2,38] Jennifer A. Brody,[3] Misa Graff,[1] Elise Lim,[2] Bridget M. Lin,[4] Hanfei Xu,[2] Najaf Amin,[5] Ping An,[6] Stella Aslibekyan,[7] Alison E. Fohner,[8,9] Bertha Hidalgo,[7] Petra Lenzini,[6] Robert Kraaij,[10] Carolina Medina-Gomez,[10] Ivana Prokić,[5] Fernando Rivadeneira,[10] Colleen Sitlani,[3] Ran Tao,[11,12] Jeroen van Rooij,[10] Di Zhang,[13] Jai G. Broome,[14,15] Erin J. Buth,[14] Benjamin D. Heavner,[14] Deepti Jain,[14]

(Author list continued on next page)

Summary

Anthropometric traits, measuring body size and shape, are highly heritable and significant clinical risk factors for cardiometabolic disorders. These traits have been extensively studied in genome-wide association studies (GWASs), with hundreds of genome-wide significant loci identified. We performed a whole-exome sequence analysis of the genetics of height, body mass index (BMI) and waist/hip ratio (WHR). We meta-analyzed single-variant and gene-based associations of whole-exome sequence variation with height, BMI, and WHR in up to 22,004 individuals, and we assessed replication of our findings in up to 16,418 individuals from 10 independent cohorts from Trans-Omics for Precision Medicine (TOPMed). We identified four trait associations with single-nucleotide variants (SNVs; two for height and two for BMI) and replicated the *LECT2* gene association with height. Our expression quantitative trait locus (eQTL) analysis within previously reported GWAS loci implicated *CEP63* and *RFT1* as potential functional genes for known height loci. We further assessed enrichment of SNVs, which were monogenic or syndromic variants within loci associated with our three traits. This led to the significant enrichment results for height, whereas we observed no Bonferroni-corrected significance for all SNVs. With a sample size of ~20,000 whole-exome sequences in our discovery dataset, our findings demonstrate the importance of genomic sequencing in genetic association studies, yet they also illustrate the challenges in identifying effects of rare genetic variants.

Introduction

Anthropometric traits, measuring body size and shape, are highly heritable and significant clinical risk factors for cardiometabolic disorders. Body mass index (BMI) is a standard measure of obesity, and waist/hip ratio (WHR) measures central adiposity, each of which are independent risk factors for both coronary heart disease and type 2 diabetes.[1,2] Although height has been associated with lower odds of coronary artery disease and hypertension,

[1]Department of Epidemiology, Gillings School of Global Public Health, University of North Carolina at Chapel Hill, Chapel Hill, NC 27516, USA; [2]Department of Biostatistics, School of Public Health, Boston University, Boston, MA 02118, USA; [3]Cardiovascular Health Research Unit, Department of Medicine, University of Washington, Seattle, WA 98101, USA; [4]Department of Biostatistics, Gillings School of Global Public Health, University of North Carolina at Chapel Hill, Chapel Hill, NC 27516, USA; [5]Department of Epidemiology, Erasmus MC, University Medical Center, Rotterdam 3015CN, the Netherlands; [6]Department of Genetics, Washington University School of Medicine, St. Louis, MO 63110, USA; [7]Department of Epidemiology, University of Alabama at Birmingham, Birmingham, AL 35294, USA; [8]Department of Epidemiology, University of Washington, Seattle, WA 98101, USA; [9]Institute for Public Health Genetics, University of Washington, Seattle, WA 98101, USA; [10]Department of Internal Medicine, Erasmus MC, University Medical Center, Rotterdam 3015CN, the Netherlands; [11]Department of Biostatistics, Vanderbilt University Medical Center, Nashville, TN 37232, USA; [12]Vanderbilt Genetics Institute, Vanderbilt University Medical Center, Nashville, TN 37232, USA; [13]Department of Molecular and Human Genetics, Baylor College of Medicine, Houston, TX 77030, USA; [14]Department of Biostatistics, University of Washington, Seattle, WA 98105, USA; [15]Division of Medical Genetics, Department of Medicine, University of Washington, Seattle, WA 98105, USA; [16]Department of Biostatistics, University of Michigan, Ann Arbor, MI, USA; [17]Division of Biomedical Informatics and Personalized Medicine, University of Colorado Anschutz Medical Campus, Aurora, CO 80045, USA; [18]Tempus Labs, Chicago, IL 60654, USA; [19]Division of Cardiology, Department of Medicine, University of Illinois at Chicago, Chicago, IL, USA; [20]Health Quantitative Sciences Research, Mayo Clinic, Rochester, MN, USA; [21]The Institute for Translational Genomics and Population Sciences, Department of Pediatrics, The Lundquist Institute for Biomedical Innovation at Harbor-UCLA Medical Center, Torrance, CA, USA; [22]Division of Public Health Sciences, Fred Hutchinson Cancer Research Center, Seattle, WA, USA; [23]Division of Endocrinology, Department of Medicine, Johns Hopkins University School of Medicine, Baltimore, MD, USA; [24]Department of Epidemiology, University of Michigan, Ann Arbor, MI, USA; [25]Channing Division of Network Medicine, Brigham and Women's Hospital, Boston, MA, USA; [26]Division of Allergy and Clinical Immunology, Department of Medicine, Johns Hopkins University, Baltimore, MD, USA; [27]Division of Pulmonary, Allergy and Critical Care Medicine, Department of Medicine, University of Alabama at Birmingham, Birmingham, AL, USA; [28]Department of Medicine, University of Maryland, Baltimore, MD, USA; [29]Department of Medicine, National Jewish Health, Denver, CO, USA; [30]Department of Medicine, Brigham & Women's Hospital, Boston, MA, USA; [31]Center for Public Health Genomics, University of Virginia, Charlottesville, VA, USA; [32]Department of Medicine, University of Washington, Seattle, WA, USA; [33]Division of General Internal Medicine, Department of Medicine, Johns Hopkins University School of Medicine, Baltimore, MD, USA; [34]College of Public Health, University of Kentucky, Lexington, KY, USA; [35]Department of Neurology, Boston University, Boston, MA, USA; [36]Department of Neurology, Columbia University, New York City, NY, USA; [37]Nuffield Department of Population Health, St. Cross

(Affiliations continued on next page)

Albert V. Smith,[16] Kathleen Barnes,[17,18] Meher Preethi Boorgula,[17] Sameer Chavan,[17] Dawood Darbar,[19] Mariza De Andrade,[20] Xiuqing Guo,[21] Jeffrey Haessler,[22] Marguerite R. Irvin,[7] Rita R. Kalyani,[23] Sharon L.R. Kardia,[24] Charles Kooperberg,[22] Wonji Kim,[25] Rasika A. Mathias,[26] Merry-Lynn McDonald,[27] Braxton D. Mitchell,[28] Patricia A. Peyser,[24] Elizabeth A. Regan,[29] Susan Redline,[30] Alexander P. Reiner,[8,22] Stephen S. Rich,[31] Jerome I. Rotter,[21] Jennifer A. Smith,[24] Scott Weiss,[25] Kerri L. Wiggins,[32] Lisa R. Yanek,[33] Donna Arnett,[34] Nancy L. Heard-Costa,[35] Suzanne Leal,[36] Danyu Lin,[4] Barbara McKnight,[14] Michael Province,[6] Cornelia M. van Duijn,[37] Kari E. North,[1,38] L. Adrienne Cupples,[2,38,39] and Ching-Ti Liu[2,38,*]

College, Oxford, UK
[38]Senior author
[39]Deceased
[40]Lead contact
*Correspondence: kristin.young@unc.edu (K.L.Y.), ctliu@bu.edu (C.-T.L.)
https://doi.org/10.1016/j.xhgg.2022.100163.

increased height is linked to greater odds of atrial fibrillation, venous thromboembolism, and cancer,[3] and the broad-sense heritability of height is estimated at >75% in numerous populations with a polygenic architecture of many small genetic effects.[4,5] These traits have been extensively studied in genome-wide association studies (GWASs), with hundreds of genome-wide significant loci identified,[6–9] but interpretation of these results in terms of functional genes remains a challenge. More recently, the Exome Chip genotyping array and target sequencing facilitated studies of association between rare and low-frequency coding variants and anthropometric traits.[10–13] This study presents the results of a whole-exome sequence association analysis of height, BMI, and WHR. In contrast with a GWAS design, where many common variants are genotyped or imputed, whole-exome sequencing focuses on the exons of protein-coding genes to identify variants that are not included in genotyping arrays or imputation reference panels. We meta-analyzed SNV and gene-based associations of whole-exome sequence variation with height, BMI, and WHR in up to 22,004 individuals. To take advantage of the high density of the exome sequence data, we investigated associations in candidate gene sets defined by known anthropometric GWAS loci, seeking to identify genes within those loci with evidence of a functional role in the trait outcome. We also examined genes harboring known monogenic or syndromic variants influencing related phenotypes (height, obesity for BMI, and lipids/insulin resistance for WHR) to identify additional low-frequency (minor allele frequency [MAF], 1%–5%) or rare (MAF, <1%) variants influencing our traits of interest.

Subjects and methods

Discovery studies

Our discovery sample included participants from seven studies: Atherosclerosis Risk in Communities (ARIC) Study, Cardiovascular Health Study (CHS), cardiovascular disease (CVD) case-control datasets from the National Heart, Lung, and Blood Institute (NHLBI) "Grand Opportunity:" Exome Sequencing Project (GO-ESP HeartGO), Erasmus Rucphen Family (ERF) Study, Framingham Heart Study (FramHS), Genetics of Lipid Lowering Drugs and Diet Network (GOLDN), and the Rotterdam Study (RS) totaling 22,004 adult individuals of European (EA:n = 17,105) or African (AA:n = 4,899) descent.[14] Individual study-level analyses were carried out in each ancestry and in ancestries combined for sex-specific and sexes combined analyses. Informed consent was obtained from participants for each parent study, and data for participants who withdrew consent or declined to participate in genetic research were excluded. Overlapping participants found in GO-ESP and other discovery cohorts (ARIC, CHS, FramHS; see supplemental information) were excluded from each parent study sample prior to analysis. All studies received prior approval from their respective institutional review boards.

Phenotypes

In each study, height was measured to the nearest centimeter using standard protocols. BMI was calculated from measured weight (kg) and height (m). WHR was calculated as waist circumference (cm) divided by hip circumference (cm). Study-specific phenotype descriptive statistics are provided in Table S1. All phenotypes were adjusted for covariates listed in the study-level statistical analyses section in linear regression models for the studies with unrelated samples and mixed models for the studies with related samples.

Exome sequencing and QC

As part of CHARGE-S (Cohorts for Heart and Aging Research in Genomic Epidemiology Sequencing), exome sequencing in ARIC, CHS, and FramHS was conducted at the Baylor College of Medicine Human Genome Sequencing Center (https://www.hgsc.bcm.edu/human/charge-consortium) as previously described.[15] In brief, each single-nucleotide variant (SNV) called using the Atlas2 suite was filtered based on the following criteria to produce a high-quality variant list: low SNV posterior probability (<0.95), low variant read count (<3), variant read ratio <0.25 or >0.75, strand-bias of >99% variant reads in a single strand direction, or total coverage <10-fold. All variant calls filtered by these criteria and reference calls with <10-fold coverage were set to missing. The variant call filters were the same for indels, except a total coverage <30-fold was used for variant sites.

Variant-level quality control steps excluded variants outside the exon capture regions (VCRrome 2.1),[16] with missing rate >20%, mappability score <0.8, and mean depth of coverage >500-fold. Variants not meeting Hardy-Weinberg equilibrium expectations ($p < 5 \times 10^{-6}$) in ancestry-specific groups were also excluded. Sample-level quality control metrics were calculated by cohort and ancestry group. A sample was excluded for missingness >20%, or if compared with the other samples it fell <6 standard deviations (SDs) for mean depth, more than 6 SDs for singleton count,

or outside of 6 SDs for heterozygote/homozygote ratio or transition/transversion (Ti/Tv) ratio.

The final sample for CHARGE contained 11,263 EA individuals (1,751 for CHS, 7,810 for ARIC, and 1,702 for FHS) and 3,180 AA from ARIC. In total, there were 2,556,859 SNVs and 76,133 indels after QC. The mean depth of coverage was 78-fold.

Exome sequencing for GO-ESP HeartGO was conducted at the University of Washington or the Broad Institute with quality control performed at the University of Michigan as part of the NHLBI GO-ESP HeartGO as previously described.[17,18] In brief, samples were excluded from exome sequencing if DNA mass or concentration was low, genotyping of 96 common SNVs (MAF, 30%–50%) revealed genotype call rates <90%, or there were sex mismatches. Samples >99% concordant at the 96 genotyped SNVs and with targeted read coverage >8x for >90% of the exome target sequence (University of Washington) or targeted read coverage >20x for >70% of the exome target sequence (Broad Institute) were considered complete. SNVs were then jointly called for all ESP samples at the University of Michigan using the UMAKE pipeline.[19] Genotypes were excluded if the read depth was <10, the average per-SNV read depth was >500, or the genotype-calling quality score was <20. Variants identified as likely false-positives using a support vector machine classifier, with missing rate ≥10% in any of the four capture arrays used in GO-ESP, with HWE $p < 5 \times 10^{-20}$ across all samples or HWE $p < 1 \times 10^{-6}$ in control samples, found on only one array, and singletons, were excluded. Samples with homozygosity >0.3, sex mismatches, low concordance with GWAS array data, or of indeterminate genetic ancestry were excluded.

Exome sequencing was conducted separately in the ERF study, GOLDN, and RS, and exome sequencing, variant analysis, and quality control protocols are available elsewhere.[20–22] In ERF, sequencing was performed at a median depth of 57x and after preprocessing, variant calling was completed using the Unified Genotyper tool from the Genome Analysis Toolkit (GATK) v.2.314.[23] In GOLDN, variants were called with SAMtools version r963 and VarScan v.2.3.6.[24–26] In the RS, variants were called using GATK's Haplotypecaller.[27] All discovery studies were aligned to the human reference genome (hg19/GRCh37). Additional details on variant calling and QC for studies used in discovery are included in the supplemental information. Samples with low genotype array concordance (<95%), with ≥4 SDs from the mean number of variants detected per sample and high heterozygote/homozygote ratio or Ti/Tv ratio, or low call rate (<90%) were excluded from further analysis.[20] We further excluded autosomal variants with call rates <0.8 and Hardy-Weinberg equilibrium $p < 10^{-4}$ from study-level analyses.

To identify study-specific quality control issues, including phenotype transformation errors using plots of the inverse of the median standard error (SE) of the effect estimates in all variants included in a study divided by the square root of the study sample size, strand flips by comparing allele frequencies across SNPs with an ancestry-specific reference (1000 Genomes European and 1000 Genomes African reference panels), and genomic inflation, we implemented a centralized procedure for cohort summary statistics in EasyQC.[28] This procedure ensured that any study-specific data errors were identified and corrected prior to meta-analysis.

Study-level statistical analyses

For each ancestry and sex stratum, residuals were obtained by regressing each phenotype (height, BMI, and WHR) on age, phenotype-associated principal components to control for population stratification, and study site (for multisite studies). Additionally, BMI was regressed on age,[2] and WHR was regressed on age[2] and BMI to identify associations independent of overall adiposity. For population-based and case-control studies of unrelated individuals, residuals were calculated separately for each sex, whereas we included sex as a covariate in sexes combined models for family-based studies. Residuals for case-control studies were calculated separately. Study-specific residuals were inverse-normally transformed prior to genetic association analyses. For each trait, transformed residuals were regressed on each variant assuming an additive genetic model in either RAREMETALWORKER or RVTESTs.[29,30] In GO-ESP, many individuals were selected because of extreme values of BMI, blood pressure, and low-density lipoprotein cholesterol. We used SCORE-SeqTDS[31] to account for trait-dependent sampling when performing genetic association analyses in GO-ESP. We used PreMeta[32] to convert the summary statistics in SCORE-SeqTDS format to RAREMETAL format to facilitate subsequent meta-analyses with the other studies.

Meta-analyses

Meta-analyses of single-variant results were performed using RAREMETAL.[30] Only variants with a cumulative minor allele count (cMAC) ≥20 across studies were considered for single-variant meta-analysis. The exome-wide significance threshold was determined by Bonferroni correction as 0.05 divided by the number of variants examined in the full sample, which varied by trait because of sample size. For BMI (Figures S1 and S2), we considered 88,168 variants at a significance threshold of 5.67×10^{-7}. For height (Figures S3 and S4), we considered 455,318 variants at a significance level of 1.10×10^{-7}. For WHR (Figures S5 and S6), we considered 67,633 variants at a significance level of 7.39×10^{-7}.

Gene-based analyses

For the gene-based analyses of our meta-analysis results, we selected variants with a MAF <5% in either ancestry, with annotation as a non-synonymous, splice-site, stop or coding indel variant. These analyses were performed in RAREMETAL using the burden (GRANVIL) and sequence kernel association test (SKAT) methods.[30] We excluded those genes with a cMAC <20 and a sample size less than one-third of the total sample for that stratum. Statistical significance was determined using a Bonferroni-corrected threshold of $p < 3.83 \times 10^{-6}$ (0.05/13,055 genes tested).

Validation

We attempted to validate our single-variant and gene-based results that reached suggestive significance ($p \leq 1 \times 10^{-6}$) in the TransOmics for Precision Medicine (TOPMed) Freeze5b data release, lifted over to hg19/GRCh37 to match the build used in the discovery analyses.[33] Specifically, we followed up with four SNVs for BMI, eight SNVs for height, zero SNVs for WHR, as well as one gene for each of three traits. Replication analyses included 16,418 participants for BMI and height from 10 studies (Amish, BAGS, CFS, COPDGene, GENOA, GeneSTAR, HVH, HyperGEN, Mayo_VTE, and VU_AF) and 4,776 participants for WHR drawn from five cohorts (Amish, CFS, GENOA, GeneSTAR, and HyperGEN). Replication study descriptions are provided in the supplemental information. We implemented linear mixed-effects models with adjustment for an empirical kinship matrix to account for relatedness and population stratification. All analyses

were adjusted for the same covariates as the discovery models. Residuals were created within each study by ancestry and sex and inverse-normal transformed, and SNV replication significance was determined using a p value of 0.05/number of variants tested per trait (BMI p = 0.0125, height p = 0.0063). For gene-based tests, variants were included if they were annotated as either loss-of-function, missense, or protein-altering indels with MAF <5%. SKAT tests used the default weights of 1 and 25 for beta distribution.[34] Analyses were restricted to those with African and European ancestry (n = 6,644, 6,644, and 3,547 African American and n = 9,774, 9,774, and 1,229 European ancestry for BMI, height, and WHR, respectively) and were conducted separately by sex and ancestry, as well as in the combined sample.

Statistical power

We calculated statistical power to reject the null hypothesis for single-variant tests at the exome-wide significance level of 1.09×10^{-7}, in a sample of 22,000 participants, with MAFs ranging from 0.01 to 0.45, and anthropometric trait allelic effect sizes of 0.007–0.2 (Figure S7).[6,7,9] Power calculations were performed in R using the gwas-power functions described in Visscher et al.[35] Given our maximum sample size of ~22,000 participants and a common MAF of 0.45, the allelic effect size must be >0.060 to achieve 80% power and >0.055 to achieve 50% power. For low-frequency variants with MAF = 0.03, the allelic effect size must be ≥0.176 to achieve 80% power and ≥0.167 to achieve 50% power.

Expression quantitative trait locus (eQTL) analysis

To identify potential functional genes within known GWAS loci for our three traits, we performed eQTL lookups based on previously reported SNVs from the National Human Genome Research Institute-European Bioinformatics Institute (NHGRI-EBI) GWAS catalog. Associations between these SNVs and gene expression levels in relevant tissues (skeletal muscle for height, five brain tissues for BMI [cerebellar hemisphere, cerebellum, cortex, nucleus accumbens, and putamen], and subcutaneous and visceral omentum adipose for BMI and WHR) were downloaded from Genotype-Tissue Expression (GTEx) v6p and filtered based on $p < 1 \times 10^{-10}$.[36,37] This provided a set of genes with evidence of regulatory relationships to SNVs previously associated with anthropometry traits.[6,7,9] We hypothesized that if the expression of these genes is involved in the causal mechanism of these GWAS SNVs, rare variants within their exons may also affect the same outcome. Thus, we considered the rare variant burden and SKAT tests for genes overlapping the ±500-kb region of the query SNV previously identified by GWAS.

Monogenic analysis

To identify low-frequency and rare variants that may contribute to anthropometric traits at the population level in genes known to influence these phenotypes, we curated lists of genes for each of our traits associated with monogenic forms of each phenotype (Tables S6–S8). For height, we included genes associated with monogenic short stature, pituitary disorders, and overgrowth syndromes. For BMI, we included genes associated with monogenic forms of obesity, while for WHR, our list included genes associated with monogenic lipodystrophy or insulin resistance. We report the association of low-frequency and rare variants within the curated lists of genes that showed nominal significance in our gene-based results. We further evaluated the significance of the excess of nominally significant variants (with 5% expected) using one-sided binomial tests.

Results

In this study, we used whole-exome sequencing to identify SNVs and genes associated with three anthropometric traits (BMI, height, and WHR) in seven discovery cohorts (ARIC, CHS, GO-ESP, ERF, FramHS, GOLDN, and RS), with validation in TOPMed Freeze5b cohorts. Phenotype sample means (SD) and sample sizes for each cohort are presented in Table S1.

Single-variant analyses

Two SNVs were identified at the exome-wide significance level associated with BMI (Table 1). The first, rs1682 4283-C (Beta [SE] = 0.084 [0.017], $p = 5.40 \times 10^{-7}$ in AA + EA women; Figure S8), is a nonsynonymous variant in *SPHKAP* (SPHK1-interactor and AKAP domain-containing protein [MIM: 611646]) on chromosome 2, and the second is a nonsynonymous variant, rs144506740-C (Beta [SE] = −1.128 [0.221], $p = 3.22 \times 10^{-7}$ in European sexes combined; Figure S9), in the 3′ UTR of *PIEZO2* (piezo-type mechanosensitive ion channel component 2 [MIM: 613629]) on chromosome 18. These findings were directionally consistent in the TOPMed sample (rs16824283-C Beta [SE] = 0.0217 [0.023], p = 0.343; rs144506740-C Beta [SE] = −0.396 [0.407], p = 0.330).

Two SNVs (rs11205303 and rs4911494) reached exome-wide significance in our AA + EA sexes combined analysis of height (Figures S10 and S11). The first, rs11205303-C (Beta [SE] = 0.065 [0.011], $p = 3.85 \times 10^{-9}$), is a missense variant previously associated with height[8] in the gene *MTMR11* (a non-receptor myotubularin-related protein) on chromosome 1. This finding was validated in the AA + EA TOPMed sample (Beta [SE] = 0.044 [0.013], $p = 5.83 \times 10^{-4}$). The second SNV reaching exome-wide significance, rs4911494-T (Beta [SE] = −0.055 [0.010], $p = 6.09 \times 10^{-8}$), is a missense variant in the height-associated gene *UQCC1* (ubiquinol-cytochrome *c* reductase complex assembly factor 1 [MIM: 611797]). This finding was also validated in the AA + EA TOPMed sample (Beta [SE] = −0.055 [0.012], $p = 2.21 \times 10^{-6}$). Of 34 associations reported in/near *UQCC1* in the GWAS catalog, 12 SNVs are associated with height. Although the specific variant identified in this study was previously associated with the WHR adjusted for BMI (WHRadjBMI),[11] it is in high linkage disequilibrium (LD) with one of the variants previously associated with height in this region (rs6088813: $R^2 = 0.998$, $D' = 1.0$) and moderate LD with five other height SNVs (rs6060355: $R^2 = 0.695$, $D' = 0.853$; rs878639: $R^2 = 0.660$, $D' = 0.909$; rs6060369: $R^2 = 0.665$, $D' = 0.913$; rs6060373: $R^2 = 0.668$, $D' = 0.913$; rs224329: $R^2 = 0.677$, $D' = 0.854$) in the 1000 Genomes phase 3 African and European (AFR + EUR) reference populations.[40]

Table 1. Single-variant results passing significance threshold of 0.05 per number of SNVs with minor allele count (MAC) >20 for each anthropometric trait

					Discovery					Replication					Previous findings
SNV	Chr:Pos	Ref/Alt	Gene	Stratum	N	MAF	Beta	SE	p	N	MAF	Beta	SE	p	
BMI (exome-wide threshold p = 5.6E−7)															
rs16824283	2: 228, 855,866	G/C	*SPHKAP*	all ancestry women	11,251	0.224	0.084	0.017	5.40E−7	8,374	0.166	0.022	0.023	0.343	not significant in Turcot et al.,[10] Ng et al. (2017)[38]
rs144506740	18: 10, 696,097	A/C	*PIEZO2*	EA sexes combined	12,319	0.001	−1.128	0.221	3.22E−7	9,774	3.10E−4	−0.396	0.407	0.330	not significant in Turcot et al.,[10] Ng et al. (2017)[38]
Height (exome-wide threshold p = 1.09E−7)															
rs11205303	1: 149, 906,413	T/C	*MTMR11*	all ancestry sexes combined	20,232	0.328	0.065	0.011	3.85E−9	16,418	0.279	0.044	0.013	5.83E−4	Berndt et al.[8]
rs4911494	20: 33, 971,914	C/T	*UQCC1*	all ancestry sexes combined	20,338	0.451	−0.055	0.01	6.09E−8	16,418	0.514	−0.055	0.012	2.21E−6	Soranzo (2009)[39]
WHRadjBMI (exome-wide threshold p = 7.86E−7)															
None															

Chr, chromosome; Pos, position; Ref, reference allele; Alt, alternate allele

No SNVs reached exome-wide significance for WHR in our analyses.

Gene-based analyses

Results of our gene-based analyses and their contributing SNVs for significant gene-based results are presented in Tables 2 and S2, respectively. The gene *LECT2* (leukocyte cell-derived chemotaxin 2 [MIM: 602882]) was associated with height in European ancestry men (burden $p = 2.56 \times 10^{-6}$, SKAT $p = 1.69 \times 10^{-5}$). Both this gene and the top variant in our gene-based association (rs62623707: MAF = 0.044, Beta [SE] = −0.142 [0.040], $p = 3.60 \times 10^{-4}$) have been previously associated with height in a sample of 381,625 European ancestry individuals (*LECT2*: SKAT $p = 9.30 \times 10^{-8}$; rs62623707: MAF = 0.044, Beta [SE] = −0.30 [0.006], $p = 1.02 \times 10^{-7}$).[12] In our full sample, this gene attained nominal significance (burden p = 0.017, SKAT p = 0.047). This finding was significant in TOPMed for the burden test in EA men (p = 0.020), but not for the SKAT test (0.084).

The gene *AGBL1* (ATP/GTP-binding protein-like 1 [MIM: 615496]) was associated with BMI in the full sample (burden $p = 9.59 \times 10^{-7}$, SKAT $p = 5.02 \times 10^{-5}$). The most significant variant in this gene was rs73459659 (MAF = 0.036, Beta [SE] = −0.099 [0.030], p = 0.001), with an additional 14 of the 174 SNVs included in the test with $p < 0.05$. Dropping the top variant from the gene-based burden association reduced the significance to $p = 1.15 \times 10^{-4}$ (above our significance threshold of 3.83×10^{-6}), suggesting that much of this signal is driven by rs73459659. Our *AGBL1*gene-based result for BMI was not replicated for either burden (p = 0.533) or SKAT (p = 0.569) tests in TOPMed EA + AA. *KRT20* (keratin 20, type 1 [MIM: 608218]) was significantly associated with WHR in our AA women gene-based burden test results ($p = 3.63 \times 10^{-6}$), but not the SKAT results ($p = 3.20 \times 10^{-4}$). Of the 22 SNVs included in this gene-based test that were polymorphic in at least one cohort, only three were found in both AA women cohorts, and only one was significant (rs34006883, $p = 4.58 \times 10^{-4}$). This result was not validated in TOPMed AA women (burden p = 0.263, SKAT p = 0.145).

eQTL analysis

For height, 697 genome-wide significant SNVs were available for analysis,[6,12,44] and we had gene-based exome sequencing and GTEx eQTL results for 25 genes around those loci. The eQTL data in skeletal muscle suggest *CEP63* (centrosomal protein, 63 kDa [MIM: 614724]), as implicated in the GWAS signal at rs4974480/rs6762606 on chromosome 3 (eQTL $p = 6.61 \times 10^{-24}$, burden test p = 0.007, SKAT p = 0.088) (Table 3). Additionally, *RFT1* (*RFT1* homolog [MIM: 611908]) in the rs2240919/rs2581830-tagged locus on chromosome 3 attained locus-specific significance (eQTL $p = 3.09 \times 10^{-17}$, burden test p = 0.002, SKAT p = 0.021). Full height eQTL results for all genes with $p < 0.05$ in our gene-based results are presented in Table S3.

For BMI, 217 genome-wide significant SNVs were available,[7,10,45–48] and we had gene-based exome sequencing

Table 2. Gene-based results passing significance threshold of 0.05 per number of genes with cumulative minor allele count (cMAC) >20 = 3.83E−6

			Discovery					Replication				
Trait	Gene	Stratum	N	SNVs (N)	cMAC	Burden p	SKAT p	N	SNVs (N)	Burden p	SKAT p	Previous findings
BMI	*AGBL1*	all ancestries sexes combined	18,740	174	5,925	9.59E−7	5.02E−5	4,756	40	0.533	0.569	Walford (2016)[41] - BMI interaction
Height	*LECT2*	EA men	5,450	11	7,415	2.56E−6	1.69E−5	585	4	0.020	0.084	Marouli[12]
WHR	*KRT20*	AA women	2,941	22	63	3.63E−6	3.20E−4	4,116	14	0.263	0.145	–

and GTEx eQTL results for 11 genes around those loci. For WHR, we identified 102 previously reported GWAS SNVs[9,47,49–51] and had gene-based exome sequencing and GTEx eQTL results for seven genes around those loci. We had no gene-based exome sequencing results with cMAC >20 for any of the eQTLs identified in the five brain tissues tested (data not shown) and no gene-based exome sequencing results for BMI or WHR in either subcutaneous or visceral fat tissue that met the threshold for locus-wide significance. BMI and WHR eQTL results with nominally significant gene-based results are presented in Tables S4 and S5, respectively.

Monogenic results

For height, we compiled a list of 187 autosomal genes with monogenic or syndromic effects on height (Table S6) and first examined our gene-based results, filtering out those genes with cMAC ≤20. A further 10 genes were not found in our results, leaving 132 genes, 7 of which were nominally significant for both gene-based tests (*ACAN* [MIM: 155760], *ADAMTS10* [MIM: 608990], *COL2A1* [MIM: 120140], *CYTL1* [MIM: 607930], *GLI3* [MIM: 165240], *PEX7* [MIM: 601757], and *RNF135* [MIM: 611358]), 6 of which were nominally significant for the burden test (*COL11A2* [MIM: 120290], *EXT1* [MIM: 608177], *FGFR3* [MIM: 134934], *NPR2* [MIM: 607072], *ORC4* [MIM: 603056], *SDHA* [MIM: 600857]), and 2 for the SKAT test (*GJA1* [MIM: 121014], *TNFRSF11B* [MIM: 602643]) (Table S9). Seventy-eight out of 1,201 SNVs in those 15 genes were nominally significant (Table S10), yielding a significant enrichment p value of 0.024 based on the binomial distribution, whereas none was significant after correction for the number of variants tested (Bonferroni-corrected $p < 0.05/1{,}201 = 4.2 \times 10^{-5}$).

For BMI, we identified 35 autosomal genes associated with monogenic or syndromic obesity (Table S9), 29 of which passed our filtering criteria (Table S11). Two genes were nominally significant for both burden and SKAT tests (*LEPR* [MIM: 601007] and *KSR2* [MIM: 610737]), while *CEP290* [MIM: 610142] was also nominally significant for the burden test. For those three nominally significant genes, we extracted results for 414 low-frequency or rare SNVs. Twenty-two SNVs were nominally significant, but none reached our Bonferroni-corrected significance threshold ($p = 0.05/414 = 1.21 \times 10^{-4}$) (Table S12).

For WHR, we identified 55 genes associated with monogenic or syndromic forms of lipodystrophy or maturity-onset diabetes of the young (MODY) (Table S13). Of those 55 genes, 39 were present in our results and passed our filtering criteria (Table S14). Two genes were nominally significant for the burden test (*LMNA* [MIM: 150330] and *HNF4A* [MIM: 600281]), and two genes were nominally significant in our SKAT results (*WFS1* [MIM: 606201] and *POLD1* [MIM: 174761]). We extracted 408 low-frequency or rare variants in these genes. Twenty-one SNVs were nominally significant in these four genes (Table S14), but none was significantly associated with WHR after Bonferroni correction for the number of variants tested ($p \leq 1.23 \times 10^{-4}$).

Discussion

In this analysis of anthropometric traits using whole-exome sequencing data from the CHARGE consortium of up to 22,004 individuals of European or African ancestry, we identified two variants associated with BMI (rs16824283 in *SPHKAP* and rs144506740 in the 3′ UTR of *PIEZO2*) and two SNVs associated with height (rs11205303 in *MTMR11* and rs4911494 in *UQCC1*). The eQTL analysis implicated two potentially functional genes for GWAS loci of height (*CEP63* and *RFT1*). Also, leveraging our aggregated association results for height, we found that the nominally significant genes with monogenic or syndromic effect showed enrichment for rare variants of nominal significance.

Two variants were associated with BMI (rs16824283 in *SPHKAP* in our women-only sample and rs144506740 in the 3′ UTR of *PIEZO2* in the sexes combined European ancestry sample), which were directionally consistent in our TOPMed validation sample, although not statistically significant (rs16824283: p = 0.343, rs144506740: p = 0.330). rs16824283 is a nonsynonymous variant in the 11th of 12 exons in *SPHKAP*, which is highly expressed in the cerebellum and heart.[37,52] This gene codes for an anchoring protein involved in transport of the type I regulatory subunit of cAMP-dependent protein kinases.[53,54] Several intronic variants in this gene have been previously associated with BMI (rs4973618 and rs6436755) and predicted visceral adipose tissue (rs4500930) or type 2 diabetes (rs7561798),[55–58] although these variants were not

Table 3. Significant eQTL lookup results in all ancestries sexes combined analyses for variants in height GWAS loci

Gene	chr	GWAS locus top SNV	eQTL tissue	eQTL p	Genes in locus	Test	N	p	Previous findings
CEP63	3	rs4974480	skeletal muscle	6.61E−24	5	burden	14,801	0.0066	Sir et al.[42]
RFT1	3	rs2581830	skeletal muscle	3.09E−17	11	burden	14,801	0.0019	Haeuptle[43]

No significant associations were noted for BMI or WHR.

in strong LD with our lead variant rs16824283 (1000 Genomes phase 3 AFR + EUR LD: $r^2 = 0.047–0.10$, $D' = 0.35–0.59$). *SPHKAP* knockout mice are obese, with increased subcutaneous, visceral, and total fat mass, and have increased plasma insulin levels and improved glucose tolerance compared with wild-type mice.[59] rs144506740 is a nonsynonymous variant in the 3′ UTR of *PIEZO2* on chromosome 18, which encodes a transmembrane protein that is involved in mechanically activated cation channels in somatosensory neurons and is expressed in the endothelial and enteroendocrine cells of the human gut.[60]

For height, two SNVs (rs11205303 and rs4911494) reached exome-wide significance in our full sample analysis and were validated in TOPMed. The first, rs11205303, is a known missense height variant[8] in *MTMR11* on chromosome 1. The second, rs4911494, is a missense variant the second exon of *UQCC1*, a gene previously associated with human height that codes for a ZIC-binding transmembrane protein that is repressed by basic fibroblast growth factor.[61] Of the dozen variants in or near this gene that have been previously reported for height, rs4911494 is in moderate LD ($r^2 \geq 0.6$) with half of them in 1000 Genomes AFR + EUR (rs224329, rs878639, rs6088813, rs6060355, rs6060369, and rs6060373), suggesting that these previously reported intronic or intergenic variants may be tagging our exonic variant.

The small number of sequence variants attaining exome-wide significance in the present study, relative to the findings of studies based on the Exome Chip, is consistent with the lower statistical power to detect associations due to differences in sample size. Our discovery sample sizes were between 17x and 25x smaller than those of the GIANT Exome Chip-based studies.[10–12] However, the Exome Chip genotyping array was based on targeted exome sequencing of approximately 12,000 subjects and likely missed rare or private variants segregating in the population. In theory, the exome sequence data analyzed here did not suffer from this limitation, because all exonic variants in the study sample are observed. The majority of the additional sequence variants available in the present study have relatively low minor allele frequencies, and the rare variants may be monomorphic or not present in all studies, resulting in lower statistical power to detect associations at expected effect sizes. Extensive simulations to evaluate statistical power of rare variant tests in exome sequencing studies have found only modest power for most genetic models in samples of up to 10,000 individuals.[62]

To combine the effects of rare variants and reduce the burden of multiple testing, we identified candidate gene sets based on previously published GWAS loci and eQTL associations. This approach implicated potentially functional genes in two loci for height on chromosome 3, *CEP63* in the GWAS signal at rs4974480 and *RFT1* in the rs2581830-tagged locus. rs4974480 is an intronic variant in *ANAPC13* (MIM: 614484), the nearest gene tagged in Wood et al.,[6] which codes for a protein involved in mitosis. The candidate causal gene in this region identified in our eQTL analysis (*CEP63*) has been linked to Seckel syndrome, a rare disorder characterized by microcephaly and short stature.[42] rs2581830 is an intronic variant in *RFT1*, the nearest gene tagged in Wood et al.,[6] and identified as a candidate causal gene in our eQTL analysis. This gene codes for a protein involved in N-linked glycosylation, an essential post-translational modification of proteins, but it is unclear how this may affect human height.[43] We also investigated potential causal low-frequency variants in genes associated with monogenic or syndromic forms of our traits or related phenotypes but found no additional significant SNVs in genes associated with monogenic or syndromic stature, obesity, MODY, or lipodystrophy after correction for multiple testing.

The identification or replication of variants associated with our traits and the findings with less frequency or rare variant enrichment in monogenic or syndromic genes highlight the opportunities and the importance of genomic sequencing to identify susceptibility variants and to better map causal associations at previously identified GWAS signals. However, the dearth of significant results in a sample size of ~20,000 individuals also illustrates the challenges in identifying effects of rare genetic variants. Future studies with much larger sample sizes are warranted.

Data and code availability

Discovery meta-analysis results will be made available through the NHGRI-EBI GWAS catalog. Data from TOPMed replication cohorts are available in dbGaP (database of Genotypes and Phenotypes; accession numbers for each study are provided in the supplemental information and Table S15).

Supplemental information

Supplemental information can be found online at https://doi.org/10.1016/j.xhgg.2022.100163.

Acknowledgments

The authors are grateful to the study participants and staff from all contributing studies. Funding support for "Building on GWAS for NHLBI-diseases: the US CHARGE consortium" was provided by the National Institutes of Health (NIH) through the American Recovery and Reinvestment Act of 2009 (ARRA) (5RC2HL102419). Data for "Building on GWAS for NHLBI-diseases: the US CHARGE consortium" was provided by Eric Boerwinkle on behalf of the ARIC Study, L.A.C., principal investigator for the FramHS, and Bruce Psaty, principal investigator for the CHS. Infrastructure for the CHARGE consortium was supported in part by NHLBI grant R01HL105756. Sequencing was carried out at the Baylor College of Medicine Human Genome Sequencing Center and supported by NHGRI grants U54 HG003273 and UM1 HG008898. The CHARGE Analysis Commons was supported by R01H L131136. K.L.Y. was supported by KL2TR001109 and R21H L14041901. C.-T.L. was supported by R01DK089256 and R01DK122503. F.R. and C.M.-G. were supported by ZonMw grant VIDI 016.136.367. C.M.v.D. used an exchange grant from Personalized Prevention of Chronic Diseases consortium (PRECeDI) (H2020-MSCA-RISE-2014). B.D.H. and D.J. were supported by U01HL-120393, contract HHSN2682018000001I, through the TOPMed Data Coordinating Center (DCC). TOPMed Informatics Research Center (IRC) was supported by contract HHSN268 201800002I. The content is solely the responsibility of the authors and does not necessarily represent the official views of the NHLBI or the NIH.

Declaration of interests

The authors declare no competing interests.

Received: August 1, 2022
Accepted: November 22, 2022

Web resources

bam-readcount, https://github.com/genome/bam-readcount

CHARGE Exome Sequencing, https://www.hgsc.bcm.edu/human/charge-consortium

NHGRI-EBI GWAS catalog, https://www.ebi.ac.uk/gwas/

NCBI dbGaP, https://www.ncbi.nlm.nih.gov/gap/

OMIM, http://www.omim.org

Picard, http://picard.sourceforge.net

RevCov, http://gmt.genome.wustl.edu/packages/refcov/index.html

Seattle Seq, https://snp.gs.washington.edu/SeattleSeqAnnotation138/

TOPMed Freeze 5b WGS QC, https://topmed.nhlbi.nih.gov/topmed-whole-genome-sequencing-project-freeze-5b-phases-1-and-2

References

1. Emdin, C.A., Khera, A.V., Natarajan, P., Klarin, D., Zekavat, S.M., Hsiao, A.J., and Kathiresan, S. (2017). Genetic association of waist-to-hip ratio with cardiometabolic traits, type 2 diabetes, and coronary heart disease. JAMA *317*, 626–634.
2. Lyall, D.M., Celis-Morales, C., Ward, J., Iliodromiti, S., Anderson, J.J., Gill, J.M.R., Smith, D.J., Ntuk, U.E., Mackay, D.F., Holmes, M.V., et al. (2017). Association of body mass index with cardiometabolic disease in the UK Biobank: A Mendelian Randomization Study. JAMA Cardiol. *2*, 882–889.
3. Lai, F.Y., Nath, M., Hamby, S.E., Thompson, J.R., Nelson, C.P., and Samani, N.J. (2018). Adult height and risk of 50 diseases: a combined epidemiological and genetic analysis. BMC Med. *16*, 187.
4. Silventoinen, K., Sammalisto, S., Perola, M., Boomsma, D.I., Cornes, B.K., Davis, C., Dunkel, L., De Lange, M., Harris, J.R., Hjelmborg, J.V.B., et al. (2003). Heritability of adult body height: a comparative study of twin cohorts in eight countries. Twin Res. *6*, 399–408.
5. Visscher, P.M., McEvoy, B., and Yang, J. (2010). From Galton to GWAS: quantitative genetics of human height. Genet. Res. *92*, 371–379.
6. Wood, A.R., Esko, T., Yang, J., Vedantam, S., Pers, T.H., Gustafsson, S., Chu, A.Y., Estrada, K., Luan, J., Kutalik, Z., et al. (2014). Defining the role of common variation in the genomic and biological architecture of adult human height. Nat. Genet. *46*, 1173–1186.
7. Locke, A.E., Kahali, B., Berndt, S.I., Justice, A.E., Pers, T.H., Day, F.R., Powell, C., Vedantam, S., Buchkovich, M.L., Yang, J., et al. (2015). Genetic studies of body mass index yield new insights for obesity biology. Nature *518*, 197–206.
8. Berndt, S.I., Gustafsson, S., Mägi, R., Ganna, A., Wheeler, E., Feitosa, M.F., Justice, A.E., Monda, K.L., Croteau-Chonka, D.C., Day, F.R., et al. (2013). Genome-wide meta-analysis identifies 11 new loci for anthropometric traits and provides insights into genetic architecture. Nat. Genet. *45*, 501–512.
9. Consortium, T.A., Consortium, T.C., Consortium, T.C., Consortium, T.G., Consortium, T.G., GLGC, T., ICBP, T., Consortium, T.I.E., Study, T.L.C., Investigators, T.M., et al. (2015). New genetic loci link adipose and insulin biology to body fat distribution. Nature *518*, 187–196.
10. Turcot, V., Lu, Y., Highland, H.M., Schurmann, C., Justice, A.E., Fine, R.S., Bradfield, J.P., Esko, T., Giri, A., Graff, M., et al. (2018). Protein-altering variants associated with body mass index implicate pathways that control energy intake and expenditure in obesity. Nat. Genet. *50*, 26–41.
11. Justice, A.E., Karaderi, T., Highland, H.M., Young, K.L., Graff, M., Lu, Y., Turcot, V., Auer, P.L., Fine, R.S., Guo, X., et al. (2019). Protein-coding variants implicate novel genes related to lipid homeostasis contributing to body-fat distribution. Nat. Genet. *51*, 452–469.
12. Marouli, E., Graff, M., Medina-Gomez, C., Lo, K.S., Wood, A.R., Kjaer, T.R., Fine, R.S., Lu, Y., Schurmann, C., Highland, H.M., et al. (2017). Rare and low-frequency coding variants alter human adult height. Nature *542*, 186–190.
13. Liu, C.-T., Young, K.L., Brody, J.A., Olden, M., Wojczynski, M.K., Heard-Costa, N., Li, G., Morrison, A.C., Muzny, D., Gibbs, R.A., et al. (2014). Sequence variation in TMEM18 in association with body mass index. Circ. Cardiovasc. Genet. *7*, 344–349.
14. Psaty, B.M., O'Donnell, C.J., Gudnason, V., Lunetta, K.L., Folsom, A.R., Rotter, J.I., Uitterlinden, A.G., Harris, T.B., Witteman, J.C.M., Boerwinkle, E.; and CHARGE Consortium (2009). Cohorts for heart and aging research in genomic epidemiology (CHARGE) consortium: design of prospective meta-analyses of genome-wide association studies from 5 cohorts. Circ. Cardiovasc. Genet. *2*, 73–80.

15. Schick, U.M., Auer, P.L., Bis, J.C., Lin, H., Wei, P., Pankratz, N., Lange, L.A., Brody, J., Stitziel, N.O., Kim, D.S., et al. (2015). Association of exome sequences with plasma C-reactive protein levels in >9000 participants. Hum. Mol. Genet. *24*, 559–571.
16. Bainbridge, M.N., Wang, M., Wu, Y., Newsham, I., Muzny, D.M., Jefferies, J.L., Albert, T.J., Burgess, D.L., and Gibbs, R.A. (2011). Targeted enrichment beyond the consensus coding DNA sequence exome reveals exons with higher variant densities. Genome Biol. *12*, R68.
17. Lange, L.A., Hu, Y., Zhang, H., Xue, C., Schmidt, E.M., Tang, Z.-Z., Bizon, C., Lange, E.M., Smith, J.D., Turner, E.H., et al. (2014). Whole-exome sequencing identifies rare and low-frequency coding variants associated with LDL cholesterol. Am. J. Hum. Genet. *94*, 233–245.
18. Reiner, A.P., Beleza, S., Franceschini, N., Auer, P.L., Robinson, J.G., Kooperberg, C., Peters, U., and Tang, H. (2012). Genome-wide association and population genetic analysis of C-reactive protein in African American and hispanic American women. Am. J. Hum. Genet. *91*, 502–512.
19. TG and HDL Working Group of the Exome Sequencing Project National Heart Lung and Blood Institute, Crosby, J., Peloso, G.M., Auer, P.L., Crosslin, D.R., Stitziel, N.O., Lange, L.A., Lu, Y., Tang, Z.z., Zhang, H., Hindy, G., et al. (2014). Loss-of-Function mutations in APOC3, triglycerides, and coronary disease. N. Engl. J. Med. *371*, 22–31.
20. Amin, N., Jovanova, O., Adams, H.H.H., Dehghan, A., Kavousi, M., Vernooij, M.W., Peeters, R.P., de Vrij, F.M.S., van der Lee, S.J., van Rooij, J.G.J., et al. (2017). Exome-sequencing in a large population-based study reveals a rare Asn396Ser variant in the LIPG gene associated with depressive symptoms. Mol. Psychiatry *22*, 537–543.
21. Amin, N., Belonogova, N.M., Jovanova, O., Brouwer, R.W.W., van Rooij, J.G.J., van den Hout, M.C.G.N., Svishcheva, G.R., Kraaij, R., Zorkoltseva, I.V., Kirichenko, A.V., et al. (2017). Nonsynonymous variation in NKPD1 increases depressive symptoms in European populations. Biol. Psychiatry *81*, 702–707.
22. Geng, X., Irvin, M.R., Hidalgo, B., Aslibekyan, S., Srinivasasainagendra, V., An, P., Frazier-Wood, A.C., Tiwari, H.K., Dave, T., Ryan, K., et al. (2019). An exome-wide sequencing study of the GOLDN cohort reveals novel associations of coding variants and fasting plasma lipids. Front. Genet. *10*, 158.
23. McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., Garimella, K., Altshuler, D., Gabriel, S., Daly, M., and DePristo, M.A. (2010). The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res. *20*, 1297–1303.
24. Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R.; and Subgroup, 1000 Genome Project Data Processing (2009). The sequence alignment/map format and SAMtools. Bioinformatics *25*, 2078–2079.
25. Koboldt, D.C., Zhang, Q., Larson, D.E., Shen, D., McLellan, M.D., Lin, L., Miller, C.A., Mardis, E.R., Ding, L., and Wilson, R.K. (2012). VarScan 2: somatic mutation and copy number alteration discovery in cancer by exome sequencing. Genome Res. *22*, 568–576.
26. Koboldt, D.C., Chen, K., Wylie, T., Larson, D.E., McLellan, M.D., Mardis, E.R., Weinstock, G.M., Wilson, R.K., and Ding, L. (2009). VarScan: variant detection in massively parallel sequencing of individual and pooled samples. Bioinformatics *25*, 2283–2285.
27. Poplin, R., Ruano-Rubio, V., DePristo, M.A., Fennell, T.J., Carneiro, M.O., Van der Auwera, G.A., Kling, D.E., Gauthier, L.D., Levy-Moonshine, A., Roazen, D., et al. (2018). Scaling accurate genetic variant discovery to tens of thousands of samples. Preprint at bioRxiv. https://doi.org/10.1101/201178.
28. Winkler, T.W., Day, F.R., Croteau-Chonka, D.C., Wood, A.R., Locke, A.E., Mägi, R., Ferreira, T., Fall, T., Graff, M., Justice, A.E., et al. (2014). Quality control and conduct of genome-wide association meta-analyses. Nat. Protoc. *9*, 1192–1212.
29. Zhan, X., Hu, Y., Li, B., Abecasis, G.R., and Liu, D.J. (2016). RVTESTS: an efficient and comprehensive tool for rare variant association analysis using sequence data. Bioinformatics *32*, 1423–1426.
30. Feng, S., Liu, D., Zhan, X., Wing, M.K., and Abecasis, G.R. (2014). RAREMETAL: fast and powerful meta-analysis for rare variants. Bioinformatics *30*, 2828–2829.
31. Lin, D.-Y., Zeng, D., and Tang, Z.-Z. (2013). Quantitative trait analysis in sequencing studies under trait-dependent sampling. Proc. Natl. Acad. Sci. USA *110*, 12247–12252.
32. Tang, Z.-Z., Bunn, P., Tao, R., Liu, Z., and Lin, D.-Y. (2017). PreMeta: a tool to facilitate meta-analysis of rare-variant associations. BMC Genom. *18*, 160–168.
33. Taliun, D., Harris, D.N., Kessler, M.D., Carlson, J., Szpiech, Z.A., Torres, R., Taliun, S.A.G., Corvelo, A., Gogarten, S.M., Kang, H.M., et al. (2019). Sequencing of 53,831 diverse genomes from the NHLBI TOPMed Program. Preprint at bioRxiv. https://doi.org/10.1101/563866.
34. Wu, M.C., Lee, S., Cai, T., Li, Y., Boehnke, M., and Lin, X. (2011). Rare-variant association testing for sequencing data with the sequence kernel association test. Am. J. Hum. Genet. *89*, 82–93.
35. Visscher, P.M., Wray, N.R., Zhang, Q., Sklar, P., McCarthy, M.I., Brown, M.A., and Yang, J. (2017). 10 Years of GWAS discovery: biology, function, and translation. Am. J. Hum. Genet. *101*, 5–22.
36. GTEx Consortium (2013). The genotype-tissue expression (GTEx) project. Nat. Genet. *45*, 580–585.
37. GTEx Consortium (2017). Genetic effects on gene expression across human tissues. Nature *550*, 204–213.
38. Ng, Maggie, Graff, Mariaelisa, Lu, Yingching, Justice, Anne, Mudgal, Poor, Liu, Ching-Ti, et al. (2017). Discovery and fine-mapping of adiposity loci using high density imputation of genome-wide association studies in individuals of African ancestry: African Ancestry Anthropometry Genetics Consortium. PLoS Genetics.
39. Soranzo, Nicole (2009). Meta-Analysis of Genome-Wide Scans for Human Adult Stature Identifies Novel Loci and Associations with Measures of Skeletal Frame Size. PLoS Genetics.
40. Machiela, M.J., and Chanock, S.J. (2015). LDlink: a web-based application for exploring population-specific haplotype structure and linking correlated alleles of possible functional variants. Bioinformatics *31*, 3555–3557.
41. Walford, Geoffrey (2014). Genome-Wide Association Study of the Modified Stumvoll Insulin Sensitivity Index Identifies BCL2 and FAM19A2 as Novel Insulin Sensitivity Loci. Diabetes Care.
42. Sir, J.-H., Barr, A.R., Nicholas, A.K., Carvalho, O.P., Khurshid, M., Sossick, A., Reichelt, S., D'Santos, C., Woods, C.G., and Gergely, F. (2011). A primary microcephaly protein complex forms a ring around parental centrioles. Nat. Genet. *43*, 1147–1153.

43. Haeuptle, M.A., Pujol, F.M., Neupert, C., Winchester, B., Kastaniotis, A.J., Aebi, M., and Hennet, T. (2008). Human RFT1 deficiency leads to a disorder of N-linked glycosylation. Am. J. Hum. Genet. *82*, 600–606.
44. Lanktree, M.B., Guo, Y., Murtaza, M., Glessner, J.T., Bailey, S.D., Onland-Moret, N.C., Lettre, G., Ongen, H., Rajagopalan, R., Johnson, T., et al. (2011). Meta-analysis of dense genecentric association studies reveals common and uncommon variants associated with height. Am. J. Hum. Genet. *88*, 6–18.
45. Speliotes, E.K., Willer, C.J., Berndt, S.I., Monda, K.L., Thorleifsson, G., Jackson, A.U., Lango Allen, H., Lindgren, C.M., Luan, J., Mägi, R., et al. (2010). Association analyses of 249, 796 individuals reveal 18 new loci associated with body mass index. Nat. Genet. *42*, 937–948.
46. Thorleifsson, G., Walters, G.B., Gudbjartsson, D.F., Steinthorsdottir, V., Sulem, P., Helgadottir, A., Styrkarsdottir, U., Gretarsdottir, S., Thorlacius, S., Jonsdottir, I., et al. (2009). Genome-wide association yields new sequence variants at seven loci that associate with measures of obesity. Nat. Genet. *41*, 18–24.
47. Winkler, T.W., Justice, A.E., Graff, M., Barata, L., Feitosa, M.F., Chu, S., Czajkowski, J., Esko, T., Fall, T., Kilpeläinen, T.O., et al. (2015). The influence of age and sex on genetic associations with adult body size and shape: a large-scale genome-wide interaction study. PLoS Genet. *11*, e1005378.
48. Graff, M., Ngwa, J.S., Workalemahu, T., Homuth, G., Schipf, S., Teumer, A., Völzke, H., Wallaschofski, H., Abecasis, G.R., Edward, L., et al. (2013). Genome-wide analysis of BMI in adolescents and young adults reveals additional insight into the effects of genetic loci over the life course. Hum. Mol. Genet. *22*, 3597–3607.
49. Heid, I.M., Jackson, A.U., Randall, J.C., Winkler, T.W., Qi, L., Steinthorsdottir, V., Thorleifsson, G., Zillikens, M.C., Speliotes, E.K., Mägi, R., et al. (2010). Meta-analysis identifies 13 new loci associated with waist-hip ratio and reveals sexual dimorphism in the genetic basis of fat distribution. Nat. Genet. *42*, 949–960.
50. Lindgren, C.M., Heid, I.M., Randall, J.C., Lamina, C., Steinthorsdottir, V., Qi, L., Speliotes, E.K., Thorleifsson, G., Willer, C.J., Herrera, B.M., et al. (2009). Genome-wide association scan meta-analysis identifies three Loci influencing adiposity and fat distribution. PLoS Genet. *5*, e1000508.
51. Monda, K.L., Liu, C.-T., Taylor, K.C., Lange, L., Demerath, E.W., Palmas, W., Wojczynski, M.K., Ellis, J.C., Vitolins, M.Z., Liu, S., et al. (2013). Genome-wide association of body fat distribution in African ancestry populations suggests new loci. PLoS Genet. *9*, e1003681.
52. GTEx Consortium (2015). The Genotype-Tissue Expression (GTEx) pilot analysis: multitissue gene regulation in humans. Science *348*, 648–660.
53. Kovanich, D., van der Heyden, M.A.G., Aye, T.T., van Veen, T.A.B., Heck, A.J.R., and Scholten, A. (2010). Sphingosine kinase interacting protein is an A-kinase anchoring protein specific for type I cAMP-dependent protein kinase. Chembiochem *11*, 963–971.
54. Lacaná, E., Maceyka, M., Milstien, S., and Spiegel, S. (2002). Cloning and characterization of a protein kinase A anchoring protein (AKAP)-related protein that interacts with and regulates sphingosine kinase 1 activity. J. Biol. Chem. *277*, 32947–32953.
55. Karlsson, T., Rask-Andersen, M., Pan, G., Höglund, J., Wadelius, C., Ek, W.E., and Johansson, Å. (2019). Contribution of genetics to visceral adiposity and its relation to cardiovascular and metabolic disease. Nat. Med. *25*, 1390–1395.
56. Kichaev, G., Bhatia, G., Loh, P.-R., Gazal, S., Burch, K., Freund, M.K., Schoech, A., Pasaniuc, B., and Price, A.L. (2019). Leveraging polygenic functional enrichment to improve GWAS power. Am. J. Hum. Genet. *104*, 65–75.
57. Pulit, S.L., Stoneman, C., Morris, A.P., Wood, A.R., Glastonbury, C.A., Tyrrell, J., Yengo, L., Ferreira, T., Marouli, E., Ji, Y., et al. (2019). Meta-analysis of genome-wide association studies for body fat distribution in 694 649 individuals of European ancestry. Hum. Mol. Genet. *28*, 166–174.
58. Xue, A., Wu, Y., Zhu, Z., Zhang, F., Kemper, K.E., Zheng, Z., Yengo, L., Lloyd-Jones, L.R., Sidorenko, J., Wu, Y., et al. (2018). Genome-wide association analyses identify 143 risk variants and putative regulatory mechanisms for type 2 diabetes. Nat. Commun. *9*, 2941.
59. Liu, Y., Harashima, S.-I., Wang, Y., Suzuki, K., Tokumoto, S., Usui, R., Tatsuoka, H., Tanaka, D., Yabe, D., Harada, N., et al. (2019). Sphingosine kinase 1-interacting protein is a dual regulator of insulin and incretin secretion. FASEB J.: Off. Publ. Fed. Am. Soc. Exp. Biol. *33*, 6239–6253.
60. Szczot, M., Nickolls, A.R., Lam, R.M., and Chesler, A.T. (2021). The form and function of PIEZO2. Annu. Rev. Biochem. *90*, 507–534.
61. Sanna, S., Jackson, A.U., Nagaraja, R., Willer, C.J., Chen, W.-M., Bonnycastle, L.L., Shen, H., Timpson, N., Lettre, G., Usala, G., et al. (2008). Common variants in the GDF5-UQCC region are associated with variation in human height. Nat. Genet. *40*, 198–203.
62. Moutsianas, L., Agarwala, V., Fuchsberger, C., Flannick, J., Rivas, M.A., Gaulton, K.J., Albers, P.K., GoT2D Consortium, McVean, G., Boehnke, M., et al. (2015). The power of gene-based rare variant methods to detect disease-associated variation and test hypotheses about complex disease. PLoS Genet. *11*, e1005165.