

# A Large Multiethnic Genome-Wide Association Study of Adult Body Mass Index Identifies Novel Loci

Thomas J. Hoffmann,<sup>\*,†,1</sup> Hélène Choquet,<sup>‡</sup> Jie Yin,<sup>‡</sup> Yambazi Banda,<sup>\*</sup> Mark N. Kvale,<sup>\*</sup> Maria Glymour,<sup>†</sup> Catherine Schaefer,<sup>‡</sup> Neil Risch,<sup>\*,†,‡</sup> and Eric Jorgenson<sup>‡,1</sup>

<sup>\*</sup>Institute for Human Genetics, University of California San Francisco, California 94143, <sup>†</sup>Department of Epidemiology and Biostatistics, University of California San Francisco, California 94158, and <sup>‡</sup>Division of Research, Kaiser Permanente Northern California, Oakland, California 94612

ORCID ID: 0000-0001-9839-8667 (H.C.)

**ABSTRACT** Body mass index (BMI), a proxy measure for obesity, is determined by both environmental (including ethnicity, age, and sex) and genetic factors, with > 400 BMI-associated loci identified to date. However, the impact, interplay, and underlying biological mechanisms among BMI, environment, genetics, and ancestry are not completely understood. To further examine these relationships, we utilized 427,509 calendar year-averaged BMI measurements from 100,418 adults from the single large multiethnic Genetic Epidemiology Research on Adult Health and Aging (GERA) cohort. We observed substantial independent ancestry and nationality differences, including ancestry principal component interactions and nonlinear effects. To increase the list of BMI-associated variants before assessing other differences, we conducted a genome-wide association study (GWAS) in GERA, with replication in the Genetic Investigation of Anthropomorphic Traits (GIANT) consortium combined with the UK Biobank (UKB), followed by GWAS in GERA combined with GIANT, with replication in the UKB. We discovered 30 novel independent BMI loci ( $P < 5.0 \times 10^{-8}$ ) that replicated. We then assessed the proportion of BMI variance explained by sex in the UKB using previously identified loci compared to previously and newly identified loci and found slight increases: from 3.0 to 3.3% for males and from 2.7 to 3.0% for females. Further, the variance explained by previously and newly identified variants decreased with increasing age in the GERA and UKB cohorts, echoed in the variance explained by the entire genome, which also showed gene–age interaction effects. Finally, we conducted a tissue expression QTL enrichment analysis, which revealed that GWAS BMI-associated variants were enriched in the cerebellum, consistent with prior work in humans and mice.

**KEYWORDS** adult body mass index; adult BMI; genome-wide association study; obesity

**B**ODY mass index (BMI) is a proxy measure for obesity, and high BMI ( $\geq 30$  kg/m<sup>2</sup>) increases the risk of many health problems (Bhaskaran *et al.* 2014; Ortega *et al.* 2016; Benjamin *et al.* 2017). BMI is determined by both genetic and environmental factors, though their individual and combined contributions to risk are not completely understood. Recent BMI heritability estimates are ~40%, over one-half of which are due to common genetic variation (Hemani *et al.* 2013; Yang *et al.* 2015). To date, meta-analyses of genome-wide

association studies (GWAS) have identified 426 independent BMI-associated variants (Liu *et al.* 2008; Thorleifsson *et al.* 2009; Willer *et al.* 2009; Speliotes *et al.* 2010; Kim *et al.* 2011; Ng *et al.* 2012, 2017; Okada *et al.* 2012; Wen *et al.* 2012, 2014; Yang *et al.* 2012, 2014; Berndt *et al.* 2013; Gong *et al.* 2013; Monda *et al.* 2013; Scannell Bryan *et al.* 2014; Hägg *et al.* 2015; Horikoshi *et al.* 2015; Locke *et al.* 2015; Winkler *et al.* 2015; Ahmad *et al.* 2016; Bakshi *et al.* 2016; Minster *et al.* 2016; Ried *et al.* 2016; Salinas *et al.* 2016; Wang *et al.* 2016; Akiyama *et al.* 2017; Graff *et al.* 2017; Justice *et al.* 2017; Nagy *et al.* 2017; Tachmazidou *et al.* 2017; Turcot *et al.* 2018) and 676 independent variants associated with measures of adiposity phenotypes, including BMI (Scuteri *et al.* 2007; Chambers *et al.* 2008; Cotsapas *et al.* 2009; Heard-Costa *et al.* 2009; Lindgren *et al.* 2009; Meyre *et al.* 2009; Heid *et al.* 2010; Scherag *et al.* 2010; Jiao *et al.* 2011; Kilpeläinen *et al.* 2011; Kraja *et al.* 2011; Paternoster *et al.*

Copyright © 2018 by the Genetics Society of America

doi: <https://doi.org/10.1534/genetics.118.301479>

Manuscript received May 10, 2018; accepted for publication August 8, 2018; published Early Online August 14, 2018.

Supplemental material available at Figshare: <https://doi.org/10.25386/genetics.6957152>.

<sup>1</sup>Corresponding authors: University of California San Francisco, 513 Parnassus Ave., San Francisco, CA 94117. E-mail: thomas.hoffmann@ucsf.edu; and Division of Research, Kaiser Permanente Northern California, 2000 Broadway, Oakland, CA 94612. E-mail: eric.jorgenson@kp.org

2011; Wang *et al.* 2011; Bradfield *et al.* 2012; Comuzzie *et al.* 2012; Melka *et al.* 2012; Graff *et al.* 2013; Liu *et al.* 2013; Namjou *et al.* 2013; Wheeler *et al.* 2013; Pei *et al.* 2014, 2017; Shungin *et al.* 2015; Felix *et al.* 2016; Sung *et al.* 2016; Wen *et al.* 2016; Chu *et al.* 2017; Southam *et al.* 2017). These variants account for only ~3% of the variance for this complex trait (Speliotes *et al.* 2010; Wen *et al.* 2014; Horikoshi *et al.* 2015; Locke *et al.* 2015). The vast majority of these loci were identified in studies of European- or Asian-ancestry (Okada *et al.* 2012; Wen *et al.* 2012, 2014; Scannell Bryan *et al.* 2014; Yang *et al.* 2014; Akiyama *et al.* 2017; Graff *et al.* 2017; Justice *et al.* 2017; Turcot *et al.* 2018) populations, as sample sizes have been somewhat smaller in Hispanic/Latino (Salinas *et al.* 2016) or African-ancestry populations (Ng *et al.* 2012, 2017; Gong *et al.* 2013; Monda *et al.* 2013; Salinas *et al.* 2016). Previous work has implicated ancestral differences, with some conflicting results (Hu *et al.* 2015), but has not yet assessed ancestry variation within nationality subgroups.

Gene–environment interaction may explain an additional portion of the missing heritability. Previous work has discovered several variants that differ between sexes (Locke *et al.* 2015) and age (Winkler *et al.* 2015), as well as overall heritability differences in age (Robinson *et al.* 2017), but found that this was driven only by young (ages 18–40) vs. old (age  $\geq$  60) in the AHTHEL composite cohort, with no evidence of interaction between ages 46–73 in the UK Biobank (UKB).

To further investigate the relationships among BMI, ancestry, sex, and age, we utilized 427,509 calendar year-averaged BMI measurements from electronic health records (EHRs) of 100,418 members from the Genetic Epidemiology Research in Adult Health and Aging (GERA) cohort. Our goal was to utilize advantages of having a single, large multiethnic cohort to obtain a more comprehensive picture of the genetic landscape of BMI. To this end, we first comprehensively assessed ancestry and nationality effects on BMI in the GERA cohort. Then, to test for ancestry differences in polygenic risk factors and further characterize single-nucleotide polymorphisms (SNPs), including age and sex differences, we first attempted to increase the list of associated variants by searching for additional BMI-associated variants in GERA, and further meta-analyzed the GERA data with the Genetic Investigation of Anthropomorphic Traits (GIANT) consortium for improved discovery. We then used previously reported plus our novel variants to test for age, sex, and ancestry differences, by both testing the individual variants themselves and testing them together in polygenic risk scores, and through gene–age coheritability estimates.

## Materials and Methods

All statistical tests were two-sided.

### Participants and phenotype

Our primary analysis used data on adult Research Program on Genes, Environment and Health (RPGEH) GERA cohort

participants (Banda *et al.* 2015; Kvale *et al.* 2015) who were, on average, 62.7 years old (at specimen collection) and members of Kaiser Permanente Northern California (KPNC) for 23 years, and had comprehensive EHRs available for the retrieval of height and weight. Each individual's height was calculated as the mode (or median if no mode), and outpatient weight measurements from 2005 to 2010 were averaged within each calendar year, excluding outlier weights  $< 70$  or  $> 500$  lb ( $< 31.7$  or  $> 226.8$  kg). Within a calendar year, if the range exceeded 100 lb (45.3 kg) (3759 occurrences), weights  $\geq 75$  pounds (34.0 kg) from the median were discarded. Across calendar years, if a difference between the average of two consecutive years was  $> 175$  lb (79.4 kg), the one further from the median was excluded. BMI was calculated by definition: weight (kg)/height (m)<sup>2</sup>. We further excluded BMI measures  $< 10$  or  $> 100$  kg/m<sup>2</sup>, where age  $< 18$  year, 6 months prior to/after childbirth, or after bariatric surgery. In total, 427,509 calendar-year BMI measurements were available for 100,418 individuals. To define nationality/geography within ethnicity groups, participants endorsed all applicable from 23 groups (Banda *et al.* 2015). The KPNC Institutional Review Board and the University of California San Francisco Human Research Protection Program Committee on Human Health approved this project. Written informed consent was obtained from all subjects.

### Genotyping, quality control, and imputation

Individuals were genotyped at over 650,000 SNPs on four custom Affymetrix arrays optimized for individuals of European, Latino, East Asian, and African American ancestry (Hoffmann *et al.* 2011a,b), and South Asians were genotyped on the European array. Genotype quality control procedures were performed array-wise, as described previously by Kvale *et al.* (2015), plus removal of SNPs with call rates  $< 90\%$ . Individuals were prephased with Shape-it v2.r727 (Delaneau *et al.* 2011) and imputed from the 1000 Genomes Project (<http://1000genomes.org>) cosmopolitan reference panel with Impute2 v2.3.0 (Howie *et al.* 2009, 2011, 2012). After excluding variants with  $r_{\text{info}}^2 < 0.3$  (Marchini and Howie 2010) and minor allele count  $< 20$ , we retained 24,149,855, 20,828,585, 15,248,462, 21,485,958, and 8,607,429 SNPs in non-Hispanic whites, Latinos, East Asians, African Americans, and South Asians (28,613,428 unique SNPs), respectively.

### Phenotype–ancestry distributions

To visualize the distribution of BMI by the ancestry principal components (PCs; PCs calculated from only the highest-performing call rate  $> 99.5\%$  SNPs) (Banda *et al.* 2015), we created a smoothed distribution of each individual's age and sex-adjusted first BMI<sub>*i*</sub> measurement using a radial kernel density estimate weighted on the distance to each other  $j^{\text{th}}$  individual,  $\sum_j \phi(\{d(i,j)/\max_{i^*,j^*}[d(i^*,j^*)]*15\})$ , where  $\phi(\cdot)$  is the standard normal density distribution and  $d(i,j)$  is the Euclidean distance of the first two PCs. Ethnicity and/or nationality subgroup labels were derived from the

GERA cohort (described above) or the Human Genome Diversity Project (Banda *et al.* 2015).

### **GERA GWAS analysis and covariate adjustment**

We first analyzed each of the five ethnicity groups (non-Hispanic whites, Latinos, East Asians, African Americans, and South Asians) separately, modeling each SNP using additive dosages (Zheng *et al.* 2011). For computational reasons, first a mixed model on BMI was fitted, adjusting for age (at corresponding BMI measurement), sex, and ancestry covariates. We then inverse normally transformed the residuals, as has been done in large meta-analyses (Locke *et al.* 2015), to be comparable. We then averaged these transformed residuals for each individual, and ran a linear regression on each SNP in a mixed model framework using estimated kinship matrices with Bolt-LMM v2.1 (Loh *et al.* 2015). We then undertook a fixed-effects meta-analysis to combine the five ethnicity groups with Metasoft v2.0 (Han and Eskin 2011).

We considered loci novel if  $\geq 0.5$  Mb was not present in any previously or newly described loci [with larger distances in regions of strong linkage disequilibrium (LD)]; specifically, we determined this both via visual inspection of plots of each locus and correlation structure, and conditional analysis with surrounding variants to confirm (Han and Eskin 2011)].

Then, to find additional independent genome-wide significant SNPs at each previously and newly described locus, we ran a stepwise regression analysis using all SNPs with  $r_{\text{info}}^2 > 0.8$  within a 1-Mb window ( $\pm 0.5$  Mb, or an expanded window size for regions with longer-LD stretches as just described) of the lead SNP. In this analysis, we adjusted only for ancestry PCs [top 10 for non-Hispanic whites, top six for other groups (Banda *et al.* 2015)] instead of the mixed model approach for simplicity and computational efficiency.

We estimated the amount that the genomic inflation factor was due to causes other than polygenicity via LD score regression with LDSC v1.0.0 (Bulik-Sullivan *et al.* 2015). We used LD score estimates from the 1000 Genomes Project European data supplied by the authors; as such, we report the ratio in GERA European-ancestry individuals and in the meta-analysis of all of GERA data (which may slightly inflate the estimate of the ratio since GERA is only 81% European ancestry).

### **SNPs previously identified**

To determine if our loci were novel, we identified 528 non-independent SNPs from previously reported studies to date for adult BMI (Liu *et al.* 2008; Thorleifsson *et al.* 2009; Willer *et al.* 2009; Speliotes *et al.* 2010; Kim *et al.* 2011; Ng *et al.* 2012, 2017; Okada *et al.* 2012; Wen *et al.* 2012, 2014; Yang *et al.* 2012, 2014; Berndt *et al.* 2013; Gong *et al.* 2013; Monda *et al.* 2013; Scannell Bryan *et al.* 2014; Hägg *et al.* 2015; Horikoshi *et al.* 2015; Locke *et al.* 2015; Winkler *et al.* 2015; Ahmad *et al.* 2016; Bakshi *et al.* 2016; Minster *et al.* 2016; Ried *et al.* 2016; Salinas *et al.* 2016; Wang *et al.* 2016; Akiyama *et al.* 2017; Graff *et al.* 2017; Justice *et al.* 2017;

Nagy *et al.* 2017; Turcot *et al.* 2018) (426 variants with all pairwise  $r^2 < 0.3$  in European ancestry) and, separately, 1304 nonindependent SNPs more broadly associated with adiposity-related phenotypes (Scuteri *et al.* 2007; Chambers *et al.* 2008; Cotsapas *et al.* 2009; Heard-Costa *et al.* 2009; Lindgren *et al.* 2009; Meyre *et al.* 2009; Heid *et al.* 2010; Scherag *et al.* 2010; Jiao *et al.* 2011; Kilpeläinen *et al.* 2011; Kraja *et al.* 2011; Paternoster *et al.* 2011; Wang *et al.* 2011; Bradfield *et al.* 2012; Comuzzie *et al.* 2012; Melka *et al.* 2012; Graff *et al.* 2013; Liu *et al.* 2013; Namjou *et al.* 2013; Wheeler *et al.* 2013; Pei *et al.* 2014, 2017; Shungin *et al.* 2015; Felix *et al.* 2016; Sung *et al.* 2016; Wen *et al.* 2016; Chu *et al.* 2017; Justice *et al.* 2017; Southam *et al.* 2017) (including the adult BMI, as well as childhood BMI, obesity, adiposity, weight, waist-hip ratio, waist circumference, fat body mass; 676 variants with all pairwise  $r^2 < 0.3$  in European ancestry). We required our novel loci to be  $> 0.5$  Mb from all of the more general previously reported adiposity-related phenotypes (or of greater distance in regions of strong LD, as described above).

### **Replication of GERA- and GERA+GIANT-identified SNPs**

To determine if any novel GERA genome-wide significant results failed to replicate, we evaluated their association in a meta-analysis of the GIANT and UKB data.

We used 234,069 European-ancestry individuals from the GIANT consortium (Locke *et al.* 2015). We restricted to European ancestry so we could extend GIANT results from the smaller HapMap v22 reference panel to the 1000 Genomes Project reference panel used here for GERA, using ImpG v1.01 (Pasaniuc *et al.* 2014). After removing SNPs with  $\leq 200,000$  individuals (Pasaniuc *et al.* 2014), 2,300,072 autosomal SNPs remained for the imputation backbone. We imputed 21,691,898 SNPs with frequency  $\geq 0.01$  (the approach performs poorly for low-frequency variants). In particular, note that using ImpG assumes all HapMap SNPs were imputed without error; likely dampening the results. Effect sizes were estimated using allele frequency and Hardy-Weinberg assumptions (Hoffmann *et al.* 2017).

The multiethnic UKB cohort (Sudlow *et al.* 2015) was imputed to the Haplotype Reference Consortium (HRC; [www.ukbiobank.ac.uk](http://www.ukbiobank.ac.uk), version 1 of the imputed data, HRC-only sites); non-HRC imputation was done by prephasing with Eagle (Loh *et al.* 2016) and imputing with Minimac3 (Das *et al.* 2016) with the 1000 Genomes Project described above. BMI was calculated from measured weight and height (UKB data record #21001). After excluding first-degree relatives, we identified 431,743 individuals who reported their ancestry as any white group and with global ancestry  $PC_1 \leq 70$  and  $PC_2 \geq -80$ , where  $PC_1$  and  $PC_2$  were calculated from the entire cohort, in addition to 7620 mixed/other, 9275 South Asian, 1822 East Asian, and 8261 African British, totaling 458,721 individuals. Ancestry PCs were recalculated within each ethnicity group, and using 50,000 random white individuals with the remaining subjects projected in for whites, as previously shown to work well (Banda *et al.*

2015). Variants were analyzed as in GERA, except using linear regression (as opposed to a mixed model), since we were only testing for replication on a few dozen variants.

### **GERA+GIANT meta-analysis and replication**

To further our discovery, we meta-analyzed GERA with the 235,069 GIANT cohort individuals as described above, genome-wide. To determine if the genome-wide significant GERA+GIANT meta-analysis SNPs replicated, we tested in the UKB.

### **Characterizing SNP effects**

To better characterize the effects of previously and newly identified SNPs, we ran a series of analyses, as follows.

**Testing for dominance and epistasis:** We tested for dominance deviation from additivity in the previously and newly identified independent SNPs by fitting a model similar to above, with an additive term for the genotype, plus an additional term for dominance (tested for significance), coded as 1 for both of the homozygote genotypes and  $-2$  for the heterozygote genotype (here we used the best guess genotype for the imputed data, rather than the dosages, as elsewhere; Bonferroni correction for 457 SNPs,  $P < 0.00011$ ).

We tested for epistasis at all pairwise sets of previously and newly identified independent SNPs. For each SNP pair, we fitted a model similar to above, with a coefficient for both genotypes (each coded additively), plus an interaction term of the two (tested for significance; Bonferroni correction for all 104,196 interactions of 353 SNPs,  $P < 4.7 \times 10^{-7}$ ).

**Effects of sex and age on BMI-associated loci:** Within GERA, we analyzed for male and female heterogeneity at all SNPs. We also analyzed BMI associations stratified by younger/older individuals using the first BMI measurement [age  $\leq 50$  year ( $n = 20,848$ ) or age  $> 50$  ( $n = 79,111$ )], as in Winkler *et al.* (2015), and then tested for heterogeneity.

### **In silico analyses**

We conducted several in silico analyses to prioritize the potentially causal variant at each of the 30 newly identified BMI loci.

**Credible sets of variants:** We used the Bayesian approach CAVIARBF (v2017-03-27) to derive the smallest set of variants that includes the causal variant with 95% probability (Chen *et al.* 2015).

**Functional variant analysis using RegulomeDB:** We used RegulomeDB (Xie *et al.* 2013; Boyle *et al.* 2014) to identify variants at each loci that likely influence regulation of gene expression, incorporating data from the Roadmap Epigenomics (Roadmap Epigenomics Consortium *et al.* 2015) and ENCODE (ENCODE Project Consortium 2012) projects. SNPs showing the most functional evidence (RegulomeDB score  $\leq 4$ ) were then investigated regarding their protein-binding capacity.

**Expression QTL analysis:** Lastly, we examined associations with gene expression using expression QTLs (eQTLs) from 44 Genotype-Tissue Expression (GTEx) v6 tissues, including subcutaneous and visceral adipose tissues (GTEx Consortium 2015). *Cis*-eQTLs were defined as variants associated with gene expression within a 2-Mb window. If no eQTLs were found, variants in high LD ( $r^2 > 0.8$ ) with the independent SNPs were examined for expression association.

### **Genetic risk score**

To test for aggregate group differences in the genetic burden of variants currently known and for variance-explained calculations described later, we additionally constructed a BMI risk score for sets of independent SNPs at previously and newly described associated loci. For each GERA and UKB individual, we summed up the additive coding of each SNP weighted by effect sizes from the UKB meta-analysis (for GERA) and the GERA meta-analysis (for UKB), so the estimate is independent from the cohort being used to test, respectively, stratified by sex. We removed previously reported nonindependent BMI variants such that no two pairwise SNPs had  $r^2 > 0.3$ .

### **Heritability**

To test for the aggregate genetic BMI burden, we first estimated familial correlations and heritability by intraclass correlations for spouse-pairs and sib-pairs, and Pearson correlations for parent-offspring relationships.

We additionally estimated the additive array heritability of all genotyped and imputed SNPs using Gear v0.7.7 (Chen 2014). As array heritability estimates can be more sensitive to artifacts than GWAS results (Lee *et al.* 2011), we restricted our analysis here to the largest set of non-Hispanic whites run on the same reagent kit and microarray. We used only autosomal data (common in array heritability estimation) and LD-filtered our data so no two pairwise SNPs had  $r^2 > 0.8$  using plink v1.90 (Chang *et al.* 2015), resulting in 547,922 genotyped and 3,796,606 imputed SNPs. Since there was population stratification within the non-Hispanic whites, we used PC-Relate (Conomos *et al.* 2016) to estimate the kinship coefficients rather than GCTA estimates (Yang *et al.* 2011a), which assume a homogeneous population. We also compared the results to this standard GCTA estimate, adjusting for PCs as described above. We additionally used Gear instead of GCTA for the PC-Relate-based heritability estimate, as the PC-Relate kinship matrix estimate was not positive definite (this can happen as the kinship estimates are based on different allele frequencies, *i.e.*, those from the PC analysis that depend on ancestry). Finally, we removed individuals so that no two individuals had kinship  $> 0.025$ , using a greedy algorithm to maximize sample size (Chang *et al.* 2015), resulting in 62,791 individuals.

We also estimated genotype-age heritability interaction effects using the genotype covariate interaction genome-based restricted maximum likelihood (GCI-GREML) model implemented in GCTA, as previously described for BMI (Robinson *et al.* 2017).

**Table 1 Descriptive factors**

Factors	Non-Hispanic white	Latino	East Asian	African American	South Asian
N (%)	81,278 (80.9%)	8322 (8.3%)	7290 (7.3%)	3069 (3.1%)	459 (0.5%)
Female, N (%)	47,335 (58.2%)	5045 (60.6%)	4264 (58.5%)	1834 (59.8%)	189 (41.2%)
Average number measured (SD)	4.3 (1.3)	4.2 (1.4)	4.0 (1.4)	4.3 (1.3)	4.0 (1.4)
Age (years)					
Male mean (SD)	63.9 (12.1)	59.1 (13.7)	59.2 (13.5)	61.6 (11.7)	55.3 (13.8)
Female mean (SD)	60.3 (13.6)	52.9 (14.9)	53.4 (14.7)	56.5 (14.2)	48.0 (13.9)
BMI (kg/m <sup>2</sup> )					
Male mean (SD)	28.0 (4.6)	29.1 (4.9)	26.1 (4.0)	29.3 (5.2)	25.8 (3.7)
Female mean (SD)	27.3 (6.0)	28.6 (6.3)	24.5 (4.5)	30.8 (6.9)	25.2 (4.2)

Descriptive factors for the GERA subjects used in the genome-wide association study of BMI by ethnicity group at first BMI measurement (for age and BMI). Abbreviations: N, number; BMI, body mass index.

### Tissue eQTL enrichment

We also used the 44 GTEx tissues to test for tissue enrichment of all lead previously and newly identified variants. Similar to previous work (Hoffmann *et al.* 2018), we constructed 10<sup>6</sup> sets of frequency-matched ( $\pm 0.5\%$ ) SNPs with respect to the lead SNP. For each tissue, we calculated the proportion of eQTL SNPs that were lead genome-wide significant SNPs (to avoid bias due to varying numbers of eQTLs in each tissue, owing to different sample sizes of each tissue). A *P*-value for enrichment was calculated with a *z*-score using the overall median tissue proportion and the SD of the null distribution of that tissue.

### Data availability

Summary statistics will be made publicly available from the National Human Genome Research Institute-European Bioinformatics Institute (NHGRI-EBI) GWAS Catalog, <https://www.ebi.ac.uk/gwas/downloads/summary-statistics>. The complete GERA data are available upon application to the Kaiser Permanente Research Bank Portal, <http://research-bank.kaiserpermanente.org/our-research/for-researchers>. The UKB data are available upon application to the UKB, [www.ukbiobank.ac.uk](http://www.ukbiobank.ac.uk). The GIANT summary statistics are available online, [http://portals.broadinstitute.org/collaboration/giant/index.php/GIANT\\_consortium\\_data\\_files](http://portals.broadinstitute.org/collaboration/giant/index.php/GIANT_consortium_data_files). Grand Opportunity (GO) Project Institutional Review Board (IRB): CN-09CScha-06-H. Supplemental material available at Figshare: <https://doi.org/10.25386/genetics.6957152>.

### Results

Our main study sample included 100,418 GERA individuals who had at least one recorded BMI measurement during the years of available EHR data. Of those, 81,278 (81.0%) were non-Hispanic white, 8322 (8.3%) Latino, 7290 (7.3%) East Asian, 3069 (3.1%) African American, and 459 (0.5%) South Asian. BMI varied by ethnicity group, with African Americans averaging the highest BMI [although this group also has greater bone mineral density and body protein content, so this is not necessarily an indication of greater adiposity (Wagner and Heyward 2000)], and East Asians and South Asians averaging the lowest (Table 1). Men generally had higher BMI

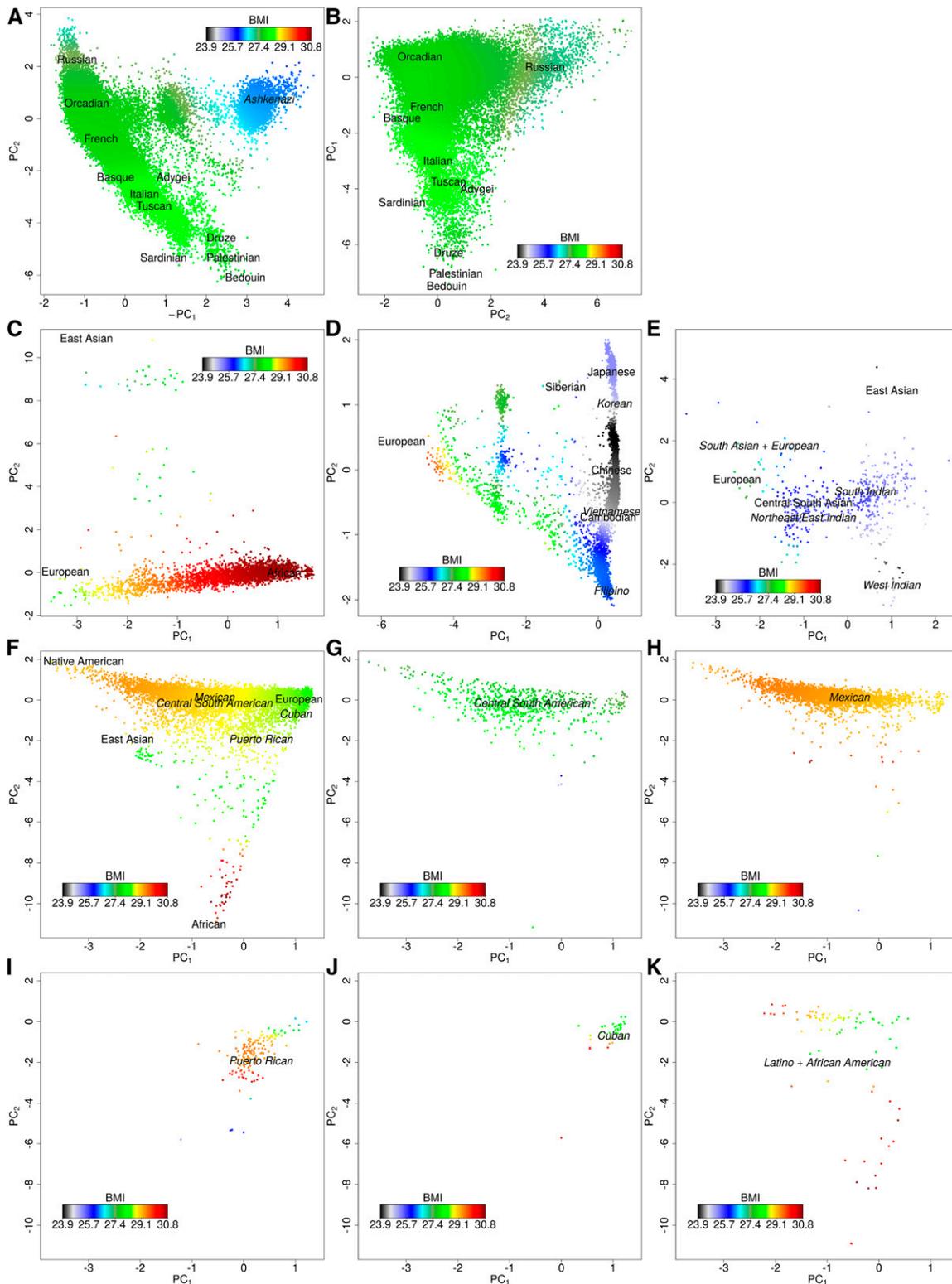
than women, except in African Americans, consistent with previous findings (Robert and Reither 2004; Flegal *et al.* 2012).

### Variation in BMI by ethnicity and ancestry/nationality

We next examined how BMI varied within each ethnicity group. The first two ancestry PCs (Banda *et al.* 2015), calculated within each group separately, generally represent geographic origin. In non-Hispanic whites, we initially found PC<sub>1</sub> ( $P = 10^{-72}$ ) and PC<sub>2</sub> ( $P = 10^{-45}$ ) associated with BMI, representing Northwest to Southeast European ancestry and Northwestern to Southeastern Europe, respectively. To better visualize this association of BMI with ancestry groups, we smoothed the phenotype distribution over the PCs (each ethnicity group and nationality subgroup in Figure 1, similar pattern seen in age- and sex-adjusted residuals in Supplemental Material, Figure S1). Refitting non-Hispanic whites with Ashkenazi's removed, and then projected, along with a term for Ashkenazi ancestry (Banda *et al.* 2015), we found an interaction effect between the first two PCs ( $P = 0.00031$ ; Table S1), representing Northern vs. Southeastern ( $P = 7.7 \times 10^{-11}$ ) and Northern vs. Southwestern ( $P = 4.5 \times 10^{-11}$ ), and Ashkenazi ancestry was also associated with lower BMI ( $P = 10^{-49}$ ; mean values by group in Table S2). In Latinos, Native American ancestry was associated with higher BMI compared to European ancestry ( $P = 1.4 \times 10^{-6}$ ), but African ancestry was not associated ( $P = 0.98$ ); in addition, Central South American nationality was associated with lower BMI ( $P = 7.0 \times 10^{-8}$ ). In East Asians, there was also an interaction effect between the first two PCs ( $P = 0.00038$ ), representing the amount of European ancestry ( $P = 10^{-67}$ ) and Northern vs. Southern East Asian ancestry, including dramatic nonlinear effects with linear ( $P = 5.2 \times 10^{-11}$ ) and quadratic ( $P = 10^{-35}$ ) terms. In African Americans, African ancestry was associated with higher BMI compared to European ancestry ( $P = 2.7 \times 10^{-9}$ ), but PC<sub>2</sub>, representing East Asian ancestry, was not associated ( $P = 0.34$ ), although there were not many individuals with significant East Asian ancestry. In South Asians, PC<sub>1</sub> was associated ( $P = 0.0067$ ) but was difficult to geographically interpret.

### GWAS of BMI in GERA, and replication in GIANT and UKB combined

Before characterizing individual and aggregate SNP effects with age, sex, and ancestry, we next sought to find additional



**Figure 1** BMI distribution in GERA ethnicity groups using the first calendar year-averaged measurement. The phenotype distribution was smoothed over the PCs (within the individuals in each respective figure), which were divided by their SD for interpretability (see *Materials and Methods*). Human Genome Diversity Project populations are in a plain font and GERA populations are in an italic font. (A) Non-Hispanic whites including individuals with Ashkenazi ancestry ( $n = 81,377$ ); (B) Non-Hispanic whites excluding individuals with Ashkenazi ancestry ( $n = 76,088$ ); (C) African Americans ( $n = 3069$ ); (D) East Asians excluding Indo Fijians ( $n = 7235$ ); (E) South Asians ( $n = 459$ ); (F) Latinos ( $n = 8322$ ); (G) Latinos: Central and South American ( $n = 612$ ); (H) Latinos: Mexican ( $n = 3048$ ); (I) Latinos: Puerto Rican ( $n = 148$ ); (J) Latinos: Cuban ( $n = 40$ ); and (K) Latinos: reporting as Latino and African American ( $n = 112$ ). BMI: body mass index; GERA, Genetic Epidemiology Research on Adult Health and Aging cohort; PC, principle component.

**Table 2 Novel loci associated with BMI ( $P \leq 5 \times 10^{-8}$ ) in the GERA multiethnic meta-analysis that replicate in GIANT+UKB**

SNP <sup>a</sup>	Chr.	Bp	Locus	Allele	GERA				GIANT+UKB	
					Freq.	Info.	$\beta$	$P$	$\beta$	$P$
rs200870675 <sup>i</sup>	3	77,646,862	ROBO2	TG/T	0.398	0.97	0.026	$4.7 \times 10^{-9}$	0.012	$3.6 \times 10^{-10}$
rs513357 <sup>i</sup>	6	69,558,698	ADGRB3	A/G	0.124	0.99	0.036	$4.4 \times 10^{-8}$	0.011	$9.8 \times 10^{-5}$
rs7938308 <sup>i</sup>	11	13,320,533	ARNTL	C/T	0.712	0.97	0.027	$2.7 \times 10^{-8}$	0.012	$1.0 \times 10^{-9}$

Genomic contexts within genes, or surrounding genes, are given. Replication test is given by the GIANT+UKB meta-analysis. GERA  $P$ -values are based on the multiethnic meta-analysis of GWAS (81,278 non-Hispanic whites, 8322 Hispanic/Latinos, 7290 East Asians, 3069 African Americans, and 459 South Asians) from the GERA sex-combined data set; GIANT+UKB  $P$ -values are based on the multiethnic meta-analysis of the European ancestry GIANT cohort members and the UKB (GIANT: 234,069 non-Hispanic whites; UKB: 431,743 non-Hispanic whites, 7620 mixed/other, 9275 South Asians, 1822 East Asians, and 8261 African British). GERA, Genetic Epidemiology Research on Adult Health and Aging cohort; GIANT, Genetic Investigation of Anthropomorphic Traits consortium; UKB, UK Biobank; Chr., chromosome; Bp, base pair (based on University of California Santa Cruz Genome Browser Assembly February 2009: GRCh37/hg19); Freq., frequency; Info., information;  $\beta$ ,  $\beta$  coefficient.

<sup>a</sup> Genomic context: i, intron; no label is intergenic.

BMI-associated variants. For our discovery GWAS meta-analysis across GERA ethnicity groups, our genomic inflation factor was 1.11, which is reasonable for a polygenic trait with a sample size this large (Yang *et al.* 2011b) (Figures S2 and S3); indeed, LD-score regression (Bulik-Sullivan *et al.* 2015) estimates that for the non-Hispanic whites, all but 0.22% was due to polygenicity (using European ancestry LD scores), and for all of GERA all but 3.6% (using European ancestry LD scores; since this reflects only 81% of our data, this may inflate the overall GERA estimate somewhat). Of note, in the individual ethnicity group analyses, we identified genome-wide significant loci in only the non-Hispanic white group (the majority of the cohort). In the multiethnic meta-analysis, we identified a total of 48 genome-wide significant loci, seven of which (Table S3) were not previously reported for BMI or any other adiposity-related phenotype (Table S4). We then tested these seven lead SNPs in a combined meta-analysis of 234,069 European-ancestry individuals from the GIANT consortium (Locke *et al.* 2015) (HapMap summary statistics extended to the 1000 Genomes Project, Figures S4 and S5 comparing the groups) and 458,721 UKB individuals from five ethnicity groups (European, East Asian, South Asian, African British, and mixed ancestries). Three of the seven variants replicated at Bonferroni significance ( $P \leq 0.05/7 = 0.0071$ , same direction of effect) in the GWAS meta-analysis of GIANT and UKB, and are reported in Table 2. Of note, two of the SNPs that failed to replicate imputed poorly ( $r_{\text{info}}^2 < 0.8$ ), while all other genome-wide significant variants imputed well ( $r_{\text{info}}^2 \geq 0.8$ ).

#### Meta-analysis of GERA and GIANT, and replication in UKB

For increased discovery, we then meta-analyzed GERA results with GIANT (replication of GERA-identified variants included these individuals plus others, as described above) (Figure 2). Our genomic inflation factor was 1.071, which is slightly lower than the analysis of GERA alone, likely due to the conservative nature of extending summary statistics (Pasaniuc *et al.* 2014). This analysis revealed an additional 31 genome-wide significant loci not previously reported (Table S5). We then tested these for replication in the UKB. Of the 31 variants, 27 replicated at a Bonferroni level ( $P \leq 0.05/31 = 0.0016$ ; Table 3).

#### Conditional results

A strength of a large cohort is more accurate conditional analysis; we next sought to find additional independent signals at each newly and previously identified locus within our large, single GERA cohort. Only two loci (2p25.3 and 18q21.32) contained an additional genome-wide significant conditional variant (Table S6). One was novel at the *TMEM18* locus, previously reported to be associated with BMI with lead SNP rs13021737 (Willer *et al.* 2009); our lead SNP at that locus was rs10188334 (meta-analysis joint  $P = 3.1 \times 10^{-23}$ ), and additional independent rs62106258 (meta-analysis joint  $P = 5.8 \times 10^{-18}$ ), which had  $r^2 < 0.01$  with rs10188334 (Table S6). We confirmed this conditional association in the UKB (joint  $P_{\text{rs10188334}} = 10^{-84}$ ,  $P_{\text{rs62106258}} = 10^{-77}$ ). We also confirmed a secondary association near *MC4R* that has been previously reported in multiple studies (Speliotes *et al.* 2010; Locke *et al.* 2015).

#### Characterizing SNP effects

**Dominance and epistasis:** We then sought to characterize BMI-associated variants further with two analyses that are typically highly underpowered: dominance and epistasis analyses. However, we still found no evidence of individual SNP dominance (using a  $P < 0.00011$  criterion, Bonferroni for all 457 previously and newly identified independent SNPs), and only a very modest overall distributional departure (Q-Q plot Figure S6,  $\lambda = 1.19$ ). We also found no individual SNP epistasis (Bonferroni  $P < 8.0 \times 10^{-7}$  for all pairwise tests of previously and newly identified SNPs), nor any distributional difference (Figure S7,  $\lambda = 1.018$ ).

**Effects of sex and age on BMI-associated loci:** We next sought to characterize the individual SNP effects by sex. We did not observe any additional genome-wide significant associations for either of the sexes that were not found in the main GERA meta-analysis, as might be expected due to the reduced sample size in each analysis and resulting loss in statistical power, nor did we find any genome-wide significant differences between the sexes. In addition, none of the previously and newly identified SNPs were different between the sexes after Bonferroni correction ( $P < 0.00014$ , Figure S8 and Table S7), although there was a moderate overall



**Table 3 Novel loci associated with BMI ( $P < 5 \times 10^{-8}$ ) in the GERA+GIANT multiethnic meta-analysis that replicate in UKB**

SNP	Chr.	BP	Locus <sup>a</sup>	Allele	GERA		GERA+GIANT		UKB	
					Freq	Info	$\beta$	$P$	$\beta$	$P$
rs1074657	1	243,746,634	<i>AKT3<sup>i</sup></i>	T/C	0.346	0.96	0.017	$7.7 \times 10^{-9}$	0.012	$1.9 \times 10^{-7}$
rs1396141	2	41,673,745	<i>ACO10739.1</i>	T/C	0.669	0.98	0.017	$1.2 \times 10^{-8}$	0.010	$2.1 \times 10^{-6}$
rs7580766	2	42,939,351	<i>MTA3</i>	G/A	0.578	1.00	0.015	$4.3 \times 10^{-8}$	0.008	0.00029
rs6710871	2	143,960,593	<i>ARHGAP15<sup>i</sup></i>	A/G	0.146	1.00	0.023	$2.2 \times 10^{-9}$	0.020	$1.9 \times 10^{-11}$
rs4857968	3	20,714,580	<i>U6</i>	G/A	0.749	1.00	0.018	$1.1 \times 10^{-9}$	0.014	$1.1 \times 10^{-8}$
rs1436351	3	104,617,973	<i>ALCAM</i>	T/G	0.736	1.00	0.016	$1.4 \times 10^{-8}$	0.015	$2.5 \times 10^{-10}$
rs4833079	4	38,654,681	<i>RP11-617D20.1<sup>i</sup></i>	T/C	0.654	0.98	0.016	$2.1 \times 10^{-10}$	0.011	$2.1 \times 10^{-7}$
rs10019997	4	137,048,599	<i>RP11-775H9.1</i>	T/C	0.451	1.00	0.015	$7.5 \times 10^{-9}$	0.016	$1.5 \times 10^{-13}$
rs7730898	5	170,459,675	<i>RANBP17<sup>i</sup></i>	A/G	0.739	0.98	0.016	$5.1 \times 10^{-9}$	0.018	$2.2 \times 10^{-14}$
rs947612	6	73,738,661	<i>KCNQ5i</i>	G/A	0.315	0.98	0.018	$1.6 \times 10^{-8}$	0.012	$7.1 \times 10^{-7}$
rs901630	6	98,539,519	<i>MIR2113</i>	C/T	0.617	1.00	0.014	$2.5 \times 10^{-8}$	0.019	$2.5 \times 10^{-19}$
rs6569648	6	130,349,119	<i>L3MBTL3i</i>	C/T	0.205	1.00	0.016	$4.3 \times 10^{-8}$	0.011	$8.7 \times 10^{-6}$
rs9364687	6	163,817,911	<i>QKI</i>	G/T	0.576	0.99	0.016	$1.0 \times 10^{-9}$	0.007	0.0015
rs6471932	8	62,078,904	<i>CLVS1</i>	T/A	0.108	0.95	0.024	$1.6 \times 10^{-8}$	0.017	$6.5 \times 10^{-8}$
rs12352785	9	6,956,850	<i>KDM4C<sup>i</sup></i>	A/C	0.282	0.99	0.018	$5 \times 10^{-11}$	0.009	$9.4 \times 10^{-5}$
rs118067556	10	63,136,165	<i>TMEM26</i>	C/T	0.966	0.89	0.044	$5.6 \times 10^{-9}$	0.030	$1.8 \times 10^{-6}$
rs10742752	11	45,438,374	<i>RP11-430H10.4</i>	C/T	0.628	0.99	0.015	$3.9 \times 10^{-9}$	0.012	$1.8 \times 10^{-8}$
rs11170468	12	39,430,048	<i>RP11-554L12.1</i>	A/C	0.787	0.99	0.017	$1.2 \times 10^{-8}$	0.011	$4.6 \times 10^{-6}$
rs1819844	12	68,205,604	<i>RP11-43N5.1</i>	A/G	0.181	0.99	0.019	$4.4 \times 10^{-9}$	0.014	$4.1 \times 10^{-7}$
rs2372716	12	99,573,426	<i>ANKS1B<sup>i</sup></i>	C/T	0.198	0.99	0.017	$3.1 \times 10^{-8}$	0.015	$1.4 \times 10^{-8}$
rs9595908	13	33,184,288	<i>PDS5B<sup>i</sup></i>	T/C	0.650	0.98	0.016	$6.5 \times 10^{-10}$	0.017	$1.1 \times 10^{-14}$
rs9563576	13	58,670,147	<i>RN5530</i>	C/T	0.816	1.00	0.022	$9.7 \times 10^{-12}$	0.024	$7.1 \times 10^{-18}$
rs7161194	14	101,529,005	<i>MIR377</i>	A/G	0.341	0.90	0.019	$1.5 \times 10^{-9}$	0.018	$1.3 \times 10^{-14}$
rs12899850	15	66,051,299	<i>DENND4A<sup>i</sup></i>	C/T	0.788	1.00	0.020	$4.0 \times 10^{-8}$	0.011	0.00012
rs11081818	18	31,251,088	<i>ASXL3<sup>i</sup></i>	A/G	0.468	0.99	0.016	$6.4 \times 10^{-9}$	0.012	$1.5 \times 10^{-8}$
rs6142096	20	32,686,658	<i>EIF2S2<sup>i</sup></i>	A/G	0.523	1.00	0.016	$1.1 \times 10^{-8}$	0.013	$2.3 \times 10^{-9}$
rs17759796	22	22,190,163	<i>MAPK1<sup>i</sup></i>	A/C	0.133	1.00	0.020	$2.5 \times 10^{-8}$	0.010	0.00067

Genomic context within gene, or surrounding genes, are given. Replication test is given by the UKB meta-analysis. GERA+GIANT  $P$ -values are based on the multiethnic meta-analysis of GWAS (GERA: 81,278 non-Hispanic whites, 8322 Hispanic/Latinos, 7290 East Asians, 3069 African Americans, and 459 South Asians; GIANT: 234,069 non-Hispanic whites) from the GERA sex-combined data set; GIANT+UKB  $P$ -values are based on the multiethnic meta-analysis of the European ancestry GIANT cohort members and the UKB (431,743 non-Hispanic whites, 7620 mixed/other, 9275 South Asians, 1822 East Asians, and 8261 African British). GERA, Genetic Epidemiology Research on Adult Health and Aging cohort; GIANT, Genetic Investigation of Anthropomorphic Traits consortium; UKB, UK Biobank; Chr., chromosome; Bp, base pair (based on University of California Santa Cruz Genome Browser Assembly February 2009: GRCh37/hg19); Freq., frequency; Info., information;  $\beta$ ,  $\beta$  coefficient.

<sup>a</sup> Genomic context: i, intron; no label is intergenic.

corresponding to maximum heritability of ~50–80%. Sex-specific heritability and a higher sibling vs. parent–child correlation were observed in GERA, but not in UKB. For comparison, in GERA, the spouse correlation was 26.6% (95% CI = 24.2–28.9%,  $n = 6064$ ).

We then estimated the GERA array heritability, which was lower than the familial heritability estimates, at 21.0% (95% CI = 19.6–22.4%) using imputed markers and the PC-Relate method (Conomos *et al.* 2016). Using the standard GCTA method, which does not account for population stratification in the kinship estimate, yielded a higher estimate at 35.4% (95% C.I. = 33.1–37.7%). We did not evaluate heritability in the other GERA ethnicity groups as the sample sizes were too small.

Next, we tested for age by heritability interaction. We stratified the sample into five roughly equally-sized age groups, using GCTA, as previously done (Robinson *et al.* 2017). We see a downward trend with increasing age (Figure 3), with a genotype–age interaction that contributes 4.2% (95% C.I. = 0.9–7.7%,  $P = 0.012$ ) to BMI variation and a coheritability estimate of 26.4% (95% C.I. = 24.6–28.2%). We then tested all pairwise comparisons among the five

groups; the comparison between age < 52 year and 66 < age  $\leq 73$  was the most suggestive ( $P = 0.0062$ , Bonferroni  $P < 0.005$ ; Table S11).

Finally, we assessed the proportion of BMI variance explained using a genetic risk score (GRS) of the previously identified BMI SNPs ( $N = 426$ ). Using independent effect size estimates stratified by sex from the UKB meta-analysis, we found that 3.2 and 3.0% of the variation in BMI was explained by previously reported SNPs in non-Hispanic white women and men, respectively (Table 5). Including GERA- and GERA+GIANT-identified SNPs slightly increased estimates to 3.5 and 3.2%. The variance explained was similar in Latino groups with 4.1 and 2.8% in women and men, respectively, and slightly less in the other groups with 2.6 and 1.7% in East Asians, and 2.0 and 1.3% in African Americans. We also estimated the variance explained in the independent UKB, using GERA/previous effect sizes. In the UKB, the variance explained by previously reported hits in non-Hispanic whites was 2.7% in women and 3.0% in men, increasing to 3.0% in women and 3.3% in men, with similar attenuations in the other groups as in GERA. Stratifying by age groups in GERA, variance explained decreased by age, which is also

**Table 4 Heritability estimates**

Group	GERA $h^2$ (95% C.I.) (N)	UKB $h^2$ (95% C.I.) (N)
Father–offspring	0.546 (0.436, 0.652) (1134)	0.636 (0.552, 0.716) (1820)
Father–son	0.622 (0.422, 0.816) (324)	0.558 (0.424, 0.688) (749)
Father–daughter	0.468 (0.396, 0.654) (810)	0.678 (0.57, 0.782) (1071)
Mother–offspring	0.618 (0.534, 0.702) (1778)	0.642 (0.588, 0.696) (4134)
Mother–son	0.564 (0.402, 0.720) (514)	0.652 (0.564, 0.738) (1616)
Mother–daughter	0.638 (0.538, 0.736) (1264)	0.65 (0.578, 0.718) (2518)
Sibling	0.704 (0.612, 0.790) (1487)	0.57 (0.546, 0.594) (21,650)
Sibling–male/male	0.708 (0.484, 0.914) (254)	0.612 (0.556, 0.666) (4202)
Sibling–female/female	0.820 (0.684, 0.946) (621)	0.624 (0.582, 0.664) (7659)
Sibling–male/female	0.542 (0.392, 0.684) (612)	0.486 (0.460, 0.534) (1820)
Array–typed PC-Relate	0.211 (0.195, 0.227) (62,791)	—
Array–typed GCTA	0.278 (0.262, 0.294) (62,791)	—
Array–imputed PC-Relate	0.210 (0.196, 0.224) (62,791)	—
Array–imputed GCTA	0.354 (0.331, 0.377) (62,791)	—

The estimates for parent–offspring and sibling heritabilities are twice the correlation estimates (see *Materials and Methods*). GERA, Genetic Epidemiology Research on Adult Health and Aging cohort; PC, principle component; UKB, UK Biobank.

seen in the UKB, although the effect is more pronounced in women in the UKB (Figure 3).

### Tissue eQTL enrichment analysis

For further aggregate biological insight into the role of BMI-associated SNPs, we also utilized the GTEx eQTLs to test for enrichment of all lead previously and newly identified BMI-associated variants. For each tissue, we determined whether the proportion of eQTLs was greater than expected. Expression in the cerebellum was different from the median expression over all tissues ( $P = 0.0002$ ; Figure 4; Bonferroni significance = 0.0011).

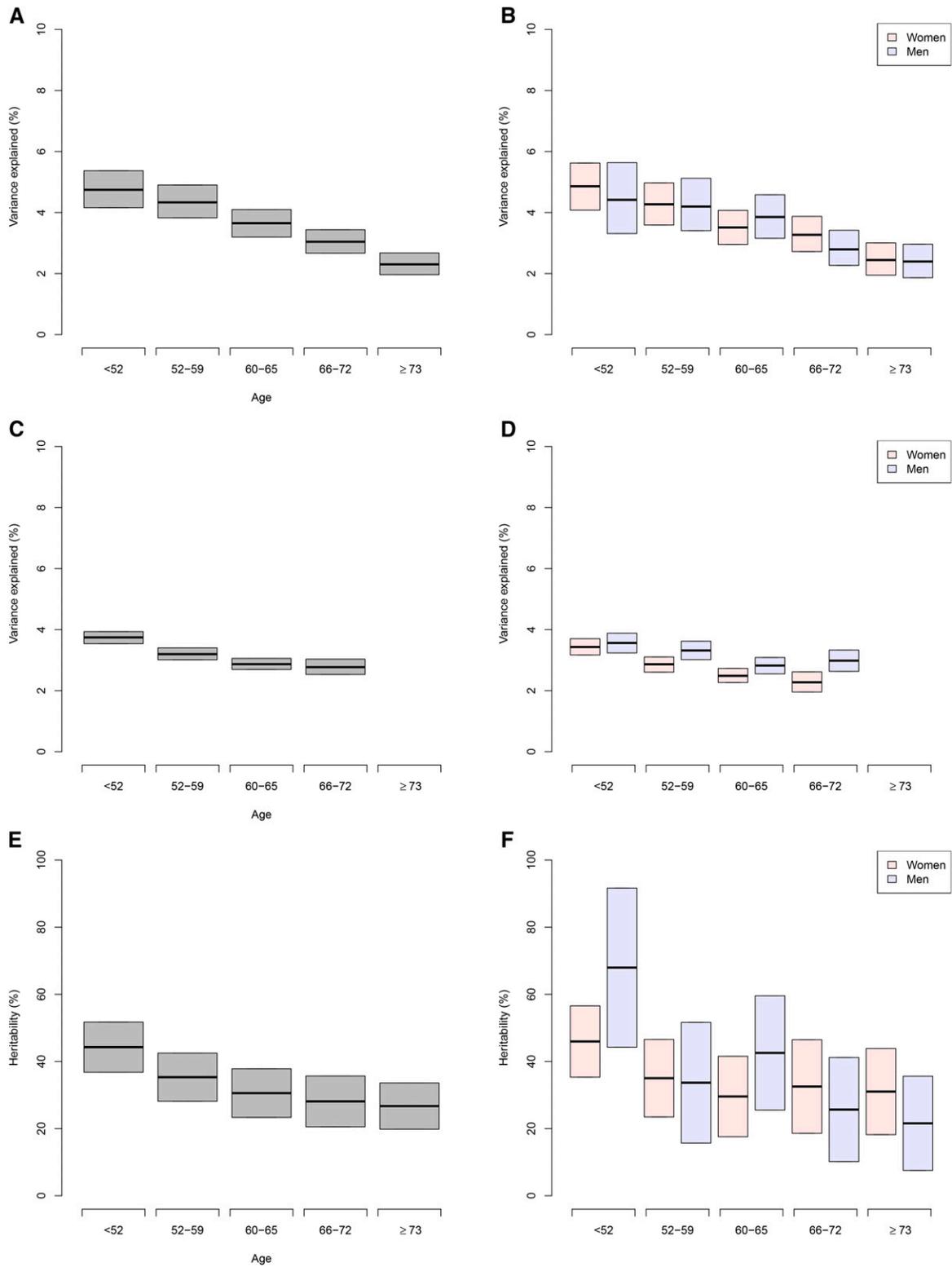
### Discussion

In the large, ethnically diverse GERA cohort with EHR-derived BMI measurements, we noted nonlinear and interaction effects between ancestry PCs, in addition to nationality effects, demonstrating effects of both ancestry and location (environment) on BMI; 30 novel, independent BMI-associated loci (not previously associated with BMI or adiposity-related phenotypes) very slightly increased variance explained; and variance explained decreased with increasing age by both GRS and heritability estimates.

We found notable interactions (non-Hispanic whites and East Asians) and nonlinear effects (East Asians) in the distribution of BMI by ancestry PCs. These ancestry effects may reflect genetic differences or they may reflect environmental/cultural differences. Specifically, the vertical cline in East Asians appears to be determined by intact nationalities, and, e.g., diets may differ among those nationalities. However, for Latinos, the European–Native American ancestry PC does not have the same distinct clusters of nationalities (Banda *et al.* 2015), and we additionally see a nationality effect, with individuals from Mexico and Central–South America overlapping on ancestry PCs, but differing with respect to average BMI, implicating environmental/cultural effects in this difference. African Americans also did not have distinct subgroups.

Novel BMI-associated loci support the important role of signaling pathways linked to adipose cell impairment, including the adipogenesis and insulin signaling pathways (Pradhan *et al.* 2017). In this study, we identified *L3MBTL3* and *AKT3* as novel BMI loci, which replicated at Bonferroni significance; these two loci have been shown to be involved in the insulin signaling pathway. SNPs in *L3MBTL3* have been shown to contribute to increased adult height and birth length (Paternoster *et al.* 2011). Recently, *L3MBTL3* was reported to be associated with insulin resistance and affect adipocyte differentiation (Lotta *et al.* 2017). *AKT3* encodes the AKT serine/threonine kinase 3, and mutations in this gene can cause an overgrowth of the brain, called megalencephaly (Alcantara *et al.* 2017). *AKT3* protein is part of the phosphatidylinositol-3-kinase (PI3K)-AKT-MTOR pathway and has been shown to be stimulated by insulin (Brozinick *et al.* 2003; Medina *et al.* 2005; Xie *et al.* 2016). Thus, mutations in those genes could contribute to impaired Akt-dependent insulin signaling in adipocytes, leading to adipose tissue accumulation and insulin resistance. In this study, we also identified *KDM4C*, which encodes a member of the Jumonji domain 2 (JMJD2) family. *KDM4C* has been shown to be involved in the PPAR $\gamma$  transcriptional activation and regulation of adipogenesis (Lizcano *et al.* 2011). Thus, these findings support an important role of *KDM4C* in the etiology of obesity, and suggest that this gene might be a good therapeutic target to treat obesity. Other genes within the novel loci identified in the current study have a plausible role in biological mechanisms relevant to obesity etiology. For example, *ARNTL* has been reported to contribute to a morningness-associated pathway related to circadian rhythms (Hu *et al.* 2016). Our results support previous work showing a relationship between the genetics of morningness, circadian rhythms, and metabolic traits, including BMI (Lane *et al.* 2017).

In addition, our tissue eQTL enrichment analysis revealed that GWAS BMI-associated SNPs were enriched in the cerebellum. This finding is consistent with previous works showing the importance of brain structure, and especially gray



**Figure 3** Variance explained. Variance explained by previously and newly identified SNPs stratified by age in GERA non-Hispanic whites overall (A) and by sex (B), and in UKB non-Hispanic whites overall (C) and by sex (D). Heritability in GERA non-Hispanic whites overall (E) and by sex (F). For GERA, one random measurement from each available individual was included in each age bin (each GERA individual was used in multiple age bins whenever possible); for UKB, we used only the one measurement available. GERA, Genetic Epidemiology Research on Adult Health and Aging cohort; UKB, UK Biobank.

**Table 5 GRS in GERA and UKB ethnicity groups**

Group	Sex	P (426 SNPs)	P+G+GG (457 SNPs)	P+G+GG	P+G+GG	P+G+GG	P+G+GG
GERA non-Hispanic whites	F	R <sup>2</sup> 0.032	R <sup>2</sup> 0.035	Mean 7.632	SD 0.265	Eff. 0.772	P 10 <sup>-374</sup>
GERA non-Hispanic whites	M	0.030	0.032	6.776	0.248	0.612	10 <sup>-244</sup>
GERA Latinos	F	0.039	0.041	7.604	0.257	0.856	10 <sup>-48</sup>
GERA Latinos	M	0.028	0.028	6.746	0.239	0.600	10 <sup>-22</sup>
GERA East Asians	F	0.024	0.026	7.340	0.243	0.689	10 <sup>-26</sup>
GERA East Asians	M	0.015	0.017	6.536	0.212	0.573	10 <sup>-12</sup>
GERA African Americans	F	0.015	0.020	7.606	0.229	0.666	10 <sup>-9</sup>
GERA African Americans	M	0.012	0.013	6.699	0.212	0.466	10 <sup>-4</sup>
UKB non-Hispanic whites	F	0.027	0.030	8.315	0.287	0.639	10 <sup>-1545</sup>
UKB non-Hispanic whites	M	0.030	0.033	6.017	0.217	0.760	10 <sup>-1451</sup>
UKB mixed/other	F	0.033	0.035	8.346	0.288	0.700	10 <sup>-34</sup>
UKB mixed/other	M	0.029	0.032	5.956	0.217	0.748	10 <sup>-26</sup>
UKB East Asians	F	0.017	0.020	8.184	0.251	0.571	10 <sup>-6</sup>
UKB East Asians	M	0.016	0.014	5.864	0.190	0.599	10 <sup>-3</sup>
UKB African British	F	0.013	0.013	8.554	0.245	0.524	10 <sup>-15</sup>
UKB African British	M	0.018	0.015	5.818	0.179	0.573	10 <sup>-13</sup>
UKB South Asians	F	0.018	0.019	8.289	0.269	0.567	10 <sup>-19</sup>
UKB South Asians	M	0.022	0.021	5.928	0.207	0.624	10 <sup>-24</sup>

Abbreviations: P, previously-identified; G, GERA-identified; GG, GERA+GIANT-identified; R<sup>2</sup>, variance explained from GRS using GERA meta-analysis effect sizes (stratified by sex); Eff., effect; GERA, Genetic Epidemiology Research on Adult Health and Aging cohort; GIANT, Genetic Investigation of Anthropomorphic Traits consortium; UKB, UK Biobank.

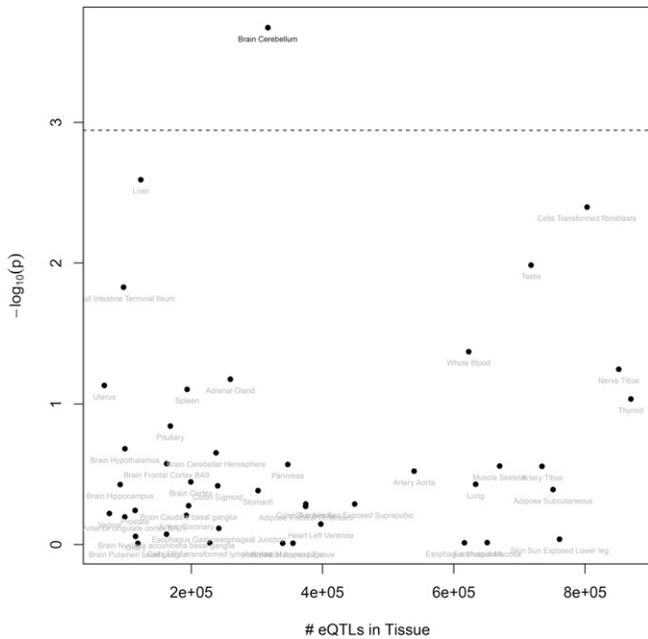
matter volume in the cerebellum, in obesity susceptibility in humans and in rats (Locke *et al.* 2015). In this study, we also identified *ADGRB3* (or *BAI3* for brain-specific angiogenesis inhibitor 3) as a novel BMI locus, which is a cell-adhesion G protein-coupled receptor (GPCR). *BAI3* is highly expressed in Purkinje cells (neurons located in the cerebellar cortex of the brain), and it is involved in *C1ql1* signaling in the mouse cerebellum to mediate normal motor learning (Lanoue *et al.* 2013; Kakegawa *et al.* 2015). Other GPCRs have been previously reported to be associated with obesity in humans and mice (Ichimura *et al.* 2012; Nakajima *et al.* 2016), and are currently being investigated as promising targets for drug discovery to treat metabolic diseases (Hauser *et al.* 2017; Riddy *et al.* 2018; Sloop *et al.* 2018).

As per the variance explained by genetic risk factors, we showed here a decreasing variance explained with increasing age in both GERA and UKB (albeit stronger in women in the UKB). However, we note that it is possible that the decrease in variance explained by age could also reflect a cohort-year effect. In addition we found that GERA- and GIANT-identified SNPs also slightly increased the variance explained in GERA and the UKB, for a total of 3.5% in GERA non-Hispanic white females and 3.2% in males, and 3.0% in UKB non-Hispanic white females and 3.3% in males. Earlier studies of BMI estimated 1.45% of the variance explained by known SNPs (Speliotes *et al.* 2010), with more recent GIANT estimates of 2.7% (but the same discovery cohort was used, which can bias estimates upwards) (Locke *et al.* 2015), and 3.5% from a younger Finnish cohort of a younger mean age of 30 (Horikoshi *et al.* 2015). In addition, our estimate of the variation explained by all variants was 21% with PC-Relate, which adjusts for population substructure in the kinship estimate directly, which is less than the 35% when adjusting for

population substructure as covariates; these are close to recent previous estimates of 27% (Yang *et al.* 2015). These are still about one-half of the family-based estimates of 40%, as has been noted (Yang *et al.* 2015); additional variation may be due to rare variants, which were not well assessed here, or unaccounted for shared environmental effects. Finally, our estimate of gene-age interaction effects of 4.2% (95% C.I. = 0.9–7.7%) was estimated as approximately one-half of previous work (8.1%) (Robinson *et al.* 2017), though our age groups were slightly older. In addition, our results were consistent with previous work (Robinson *et al.* 2017), showing suggestive evidence of a gene-age interaction effect on BMI, with the largest differences existing between the youngest group compared to older groups.

Because our study was conducted in a single, large, diverse discovery cohort, we were able to evaluate effects across ethnicity groups in the same setting. The variance explained by previously reported loci was estimated highest in non-Hispanic white and Latino women in both GERA and UKB, with GERA and UKB East Asian women and males both lower, and with GERA African American women lower, GERA African American men higher, and overall UKB men and women both lower than GERA.

There were several limitations to our study. In addition to replication data on the sex chromosomes being unavailable in GIANT, we note an additional limitation in using GIANT summary statistics expanded from HapMap 22, rather than full 1000 Genomes Project imputed results, as such results are not available. This use of approximated results, in addition to the assumption that all test statistics from GIANT are perfectly imputed, is likely conservative in terms of the true effect at each SNP. Nevertheless, 87% of the SNPs identified in the GERA+GIANT meta-analysis replicated at a strict Bonferroni



**Figure 4** Tissue eQTL enrichment. Enrichment of tissue for proportion of eQTLs that are genome-wide significant by GTEx tissue type (44 tissues, labeled in figure). eQTL, expression QTL; GTEx, Genotype-Tissue Expression consortium.

correction in the UKB, and we could test SNPs on the sex chromosomes in the UKB.

In summary, our results demonstrate the value of conducting genetic studies in large, diverse cohorts, enabled by linking EHRs with genome-wide genotype data, and expand our knowledge of the genetic basis of BMI.

### Acknowledgments

We thank the Kaiser Permanente Northern California members who have generously agreed to participate in the Kaiser Permanente Research Program on Genes, Environment, and Health. This research has been conducted using the UK Biobank Resource. This work was supported by grants R21 AG-046616 and K01 DC-013300 to T.J.H., and R01 EY-027004 to E.J., from the National Institutes of Health (NIH). Support for participant enrollment, survey completion, and biospecimen collection for the RPGEH was provided by the Robert Wood Johnson Foundation, the Wayne and Gladys Valley Foundation, the Ellison Medical Foundation, and Kaiser Permanente national and regional community benefit programs. Genotyping of the GERA cohort was funded by a grant from the National Institute on Aging, the National Institute of Mental Health, and the NIH Common Fund (grant RC2 AG-036607 to C.S. and N.R.). The funders had no role in study design, data collection and analysis, the decision to publish, or preparation of the manuscript. The authors declare that they have no competing interests.

Author contributions: T.J.H., H.C. and E.J. conceived and designed the study. T.J.H. performed the statistical analysis.

T.J.H., H.C., and J.Y. performed in silico analyses. T.J.H., H.C., and E.J. interpreted the results of the analysis, and wrote the initial draft. J.Y., Y.B., M.N.K., M.G., C.S., and N.R. contributed to the critical review of the manuscript.

### Literature Cited

Ahmad, S., W. Zhao, F. Renström, A. Rasheed, M. Zaidi *et al.*, 2016 A novel interaction between the FLJ33534 locus and smoking in obesity: a genome-wide study of 14 131 Pakistani adults. *Int. J. Obes. (Lond)* 40: 186–190. <https://doi.org/10.1038/ijo.2015.152>

Akiyama, M., Y. Okada, M. Kanai, A. Takahashi, Y. Momozawa *et al.*, 2017 Genome-wide association study identifies 112 new loci for body mass index in the Japanese population. *Nat. Genet.* 49: 1458–1467. <https://doi.org/10.1038/ng.3951>

Alcantara, D., A. E. Timms, K. Gripp, L. Baker, K. Park *et al.*, 2017 Mutations of AKT3 are associated with a wide spectrum of developmental disorders including extreme megalencephaly. *Brain* 140: 2610–2622. <https://doi.org/10.1093/brain/awx203>

Bakshi, A., Z. Zhu, A. A. E. Vinkhuyzen, W. D. Hill, A. F. McRae *et al.*, 2016 Fast set-based association analysis using summary data from GWAS identifies novel gene loci for human complex traits. *Sci. Rep.* 6: 32894. <https://doi.org/10.1038/srep32894>

Banda, Y., M. N. Kvale, T. J. Hoffmann, S. E. Hesselton, D. Ranatunga *et al.*, 2015 Characterizing race/ethnicity and genetic ancestry for 100,000 subjects in the genetic epidemiology research on adult health and aging (GERA) cohort. *Genetics* 200: 1285–1295. <https://doi.org/10.1534/genetics.115.178616>

Benjamin, E. J., M. J. Blaha, S. E. Chiuve, M. Cushman, S. R. Das *et al.*, 2017 Heart disease and stroke statistics-2017 update: a report from the American heart association. *Circulation* 135: e146–e603 (erratum: *Circulation* 135: e646); (erratum: *Circulation* 136: e196). <https://doi.org/10.1161/CIR.0000000000000485>

Berndt, S. I., S. Gustafsson, R. Mägi, A. Ganna, E. Wheeler *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. <https://doi.org/10.1038/ng.2606>

Bhaskaran, K., I. Douglas, H. Forbes, I. dos-Santos-Silva, D. A. Leon *et al.*, 2014 Body-mass index and risk of 22 specific cancers: a population-based cohort study of 5.24 million UK adults. *Lancet* 384: 755–765. [https://doi.org/10.1016/S0140-6736\(14\)60892-8](https://doi.org/10.1016/S0140-6736(14)60892-8)

Boyle, A. P., C. L. Araya, C. Brdlik, P. Cayting, C. Cheng *et al.*, 2014 Comparative analysis of regulatory information and circuits across distant species. *Nature* 512: 453–456. <https://doi.org/10.1038/nature13668>

Bradfield, J. P., H. R. Taal, N. J. Timpson, A. Scherag, C. Lecoeur *et al.*, 2012 A genome-wide association meta-analysis identifies new childhood obesity loci. *Nat. Genet.* 44: 526–531. <https://doi.org/10.1038/ng.2247>

Brozinick, J. T., B. R. Roberts, and G. L. Dohm, 2003 Defective signaling through Akt-2 and -3 but not Akt-1 in insulin-resistant human skeletal muscle: potential role in insulin resistance. *Diabetes* 52: 935–941. <https://doi.org/10.2337/diabetes.52.4.935>

Bulik-Sullivan, B. K., P.-R. Loh, H. K. Finucane, S. Ripke, J. Yang *et al.*, 2015 LD Score regression distinguishes confounding from polygenicity in genome-wide association studies. *Nat. Genet.* 47: 291–295. <https://doi.org/10.1038/ng.3211>

Chambers, J. C., P. Elliott, D. Zabaneh, W. Zhang, Y. Li *et al.*, 2008 Common genetic variation near MC4R is associated with waist circumference and insulin resistance. *Nat. Genet.* 40: 716–718. <https://doi.org/10.1038/ng.156>

Chang, C. C., C. C. Chow, L. C. Tellier, S. Vattikuti, S. M. Purcell *et al.*, 2015 Second-generation PLINK: rising to the challenge

- of larger and richer datasets. *Gigascience* 4: 7. <https://doi.org/10.1186/s13742-015-0047-8>
- Chen, G.-B., 2014 Estimating heritability of complex traits from genome-wide association studies using IBS-based Haseman-Elston regression. *Stat. Genet. Methodol.* 5: 107. <https://doi.org/10.3389/fgene.2014.00107>
- Chen, W., B. R. Larrabee, I. G. Ovsyannikova, R. B. Kennedy, I. H. Haralambieva *et al.*, 2015 Fine mapping causal variants with an approximate Bayesian method using marginal test statistics. *Genetics* 200: 719–736. <https://doi.org/10.1534/genetics.115.176107>
- Chu, A. Y., X. Deng, V. A. Fisher, A. Drong, Y. Zhang *et al.*, 2017 Multiethnic genome-wide meta-analysis of ectopic fat depots identifies loci associated with adipocyte development and differentiation. *Nat. Genet.* 49: 125–130. <https://doi.org/10.1038/ng.3738>
- Comuzzie, A. G., S. A. Cole, S. L. Laston, V. S. Voruganti, K. Haack *et al.*, 2012 Novel genetic loci identified for the pathophysiology of childhood obesity in the Hispanic population. *PLoS One* 7: e51954. <https://doi.org/10.1371/journal.pone.0051954>
- Conomos, M. P., A. P. Reiner, B. S. Weir, and T. A. Thornton, 2016 Model-free estimation of recent genetic relatedness. *Am. J. Hum. Genet.* 98: 127–148. <https://doi.org/10.1016/j.ajhg.2015.11.022>
- Cotsapas, C., E. K. Speliotes, I. J. Hatoum, D. M. Greenawalt, R. Dobrin *et al.*, 2009 Common body mass index-associated variants confer risk of extreme obesity. *Hum. Mol. Genet.* 18: 3502–3507. <https://doi.org/10.1093/hmg/ddp292>
- Das, S., L. Forer, S. Schönherr, C. Sidore, A. E. Locke *et al.*, 2016 Next-generation genotype imputation service and methods. *Nat. Genet.* 48: 1284–1287. <https://doi.org/10.1038/ng.3656>
- Delaneau, O., J. Marchini, and J. F. Zagury, 2011 A linear complexity phasing method for thousands of genomes. *Nat. Methods* 9: 179–181. <https://doi.org/10.1038/nmeth.1785>
- ENCODE Project Consortium, 2012 An integrated encyclopedia of DNA elements in the human genome. *Nature* 489: 57–74. <https://doi.org/10.1038/nature11247>
- Felix, J. F., J. P. Bradfield, C. Monnereau, R. J. van der Valk, E. Stergiakouli *et al.*, 2016 Genome-wide association analysis identifies three new susceptibility loci for childhood body mass index. *Hum. Mol. Genet.* 25: 389–403. <https://doi.org/10.1093/hmg/ddv472>
- Flegal, K. M., M. D. Carroll, B. K. Kit, and C. L. Ogden, 2012 Prevalence of obesity and trends in the distribution of body mass index among US adults, 1999–2010. *JAMA* 307: 491–497. <https://doi.org/10.1001/jama.2012.39>
- Gong, J., F. Schumacher, U. Lim, L. A. Hindorff, J. Haessler *et al.*, 2013 Fine mapping and identification of BMI loci in African Americans. *Am. J. Hum. Genet.* 93: 661–671. <https://doi.org/10.1016/j.ajhg.2013.08.012>
- Graff, M., J. S. Ngwa, T. Workalemahu, G. Homuth, S. Schipf *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. <https://doi.org/10.1093/hmg/ddt205>
- Graff, M., R. A. Scott, A. E. Justice, K. L. Young, M. F. Feitosa *et al.*, 2017 Genome-wide physical activity interactions in adiposity — a meta-analysis of 200,452 adults. *PLoS Genet.* 13: e1006528 (erratum: *PLoS Genet.* 13: e1006972). <https://doi.org/10.1371/journal.pgen.1006528>
- GTEX Consortium, 2015 Human genomics. The genotype-tissue expression (GTEx) pilot analysis: multitissue gene regulation in humans. *Science* 348: 648–660. <https://doi.org/10.1126/science.1262110>
- Hägg, S., A. Ganna, S. W. Van Der Laan, T. Esko, T. H. Pers *et al.*, 2015 Gene-based meta-analysis of genome-wide association studies implicates new loci involved in obesity. *Hum. Mol. Genet.* 24: 6849–6860. <https://doi.org/10.1093/hmg/ddv379>
- Han, B., and E. Eskin, 2011 Random-effects model aimed at discovering associations in meta-analysis of genome-wide association studies. *Am. J. Hum. Genet.* 88: 586–598. <https://doi.org/10.1016/j.ajhg.2011.04.014>
- Hauser, A. S., M. M. Attwood, M. Rask-Andersen, H. B. Schiöth, and D. E. Gloriam, 2017 Trends in GPCR drug discovery: new agents, targets and indications. *Nat. Rev. Drug Discov.* 16: 829–842. <https://doi.org/10.1038/nrd.2017.178>
- Heard-Costa, N. L., M. C. Zillikens, K. L. Monda, A. Johansson, T. B. Harris *et al.*, 2009 NRXN3 is a novel locus for waist circumference: a genome-wide association study from the CHARGE Consortium. *PLoS Genet.* 5: e1000539. <https://doi.org/10.1371/journal.pgen.1000539>
- Heid, I. M., A. U. Jackson, J. C. Randall, T. W. Winkler, L. Qi *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 [corrigenda: *Nat. Genet.* 43: 1164 (2011)]. <https://doi.org/10.1038/ng.685>
- Hemani, G., J. Yang, A. Vinkhuyzen, J. E. Powell, G. Willemsen *et al.*, 2013 Inference of the genetic architecture underlying BMI and height with the use of 20,240 sibling pairs. *Am. J. Hum. Genet.* 93: 865–875. <https://doi.org/10.1016/j.ajhg.2013.10.005>
- Hoffmann, T. J., M. N. Kvale, S. E. Hesselson, Y. Zhan, C. Aquino *et al.*, 2011a Next generation genome-wide association tool: design and coverage of a high-throughput European-optimized SNP array. *Genomics* 98: 79–89. <https://doi.org/10.1016/j.ygeno.2011.04.005>
- Hoffmann, T. J., Y. Zhan, M. N. Kvale, S. E. Hesselson, J. Gollub *et al.*, 2011b Design and coverage of high throughput genotyping arrays optimized for individuals of East Asian, African American, and Latino race/ethnicity using imputation and a novel hybrid SNP selection algorithm. *Genomics* 98: 422–430. <https://doi.org/10.1016/j.ygeno.2011.08.007>
- Hoffmann, T. J., G. B. Ehret, P. Nandakumar, D. Ranatunga, C. Schaefer *et al.*, 2017 Genome-wide association analyses using electronic health records identify new loci influencing blood pressure variation. *Nat. Genet.* 49: 54–64. <https://doi.org/10.1038/ng.3715>
- Hoffmann, T. J., E. Theusch, T. Haldar, D. K. Ranatunga, E. Jorgenson *et al.*, 2018 A large electronic-health-record-based genome-wide study of serum lipids. *Nat. Genet.* 50: 401–413. <https://doi.org/10.1038/s41588-018-0064-5>
- Horikoshi, M., R. Mägi, M. van de Bunt, I. Surakka, A.-P. Sarin *et al.*, 2015 Discovery and fine-mapping of glycaemic and obesity-related trait loci using high-density imputation. *PLoS Genet.* 11: e1005230. <https://doi.org/10.1371/journal.pgen.1005230>
- Howe, B., J. Marchini, and M. Stephens, 2011 Genotype imputation with thousands of genomes. *G3 (Bethesda)* 1: 457–470. <https://doi.org/10.1534/g3.111.001198>
- Howe, B., C. Fuchsberger, M. Stephens, J. Marchini, and G. R. Abecasis, 2012 Fast and accurate genotype imputation in genome-wide association studies through pre-phasing. *Nat. Genet.* 44: 955–959. <https://doi.org/10.1038/ng.2354>
- Howe, B. N., P. Donnelly, and J. Marchini, 2009 A flexible and accurate genotype imputation method for the next generation of genome-wide association studies. *PLoS Genet.* 5: e1000529. <https://doi.org/10.1371/journal.pgen.1000529>
- Hu, H., C. D. Huff, Y. Yamamura, X. Wu, and S. S. Strom, 2015 The relationship between native American ancestry, body mass index and diabetes risk among Mexican-Americans. *PLoS One* 10: e0141260. <https://doi.org/10.1371/journal.pone.0141260>
- Hu, Y., A. Shmygelska, D. Tran, N. Eriksson, J. Y. Tung *et al.*, 2016 GWAS of 89,283 individuals identifies genetic variants associated with self-reporting of being a morning person. *Nat. Commun.* 7: 10448. <https://doi.org/10.1038/ncomms10448>

- Ichimura, A., A. Hirasawa, O. Poulain-Godefroy, A. Bonnefond, T. Hara *et al.*, 2012 Dysfunction of lipid sensor GPR120 leads to obesity in both mouse and human. *Nature* 483: 350–354. <https://doi.org/10.1038/nature10798>
- Jiao, H., P. Arner, J. Hoffstedt, D. Brodin, B. Dubern *et al.*, 2011 Genome wide association study identifies KCNMA1 contributing to human obesity. *BMC Med. Genomics* 4: 51. <https://doi.org/10.1186/1755-8794-4-51>
- Justice, A. E., T. W. Winkler, M. F. Feitosa, M. Graff, V. A. Fisher *et al.*, 2017 Genome-wide meta-analysis of 241,258 adults accounting for smoking behaviour identifies novel loci for obesity traits. *Nat. Commun.* 8: 14977. <https://doi.org/10.1038/ncomms14977>
- Kakegawa, W., N. Mitakidis, E. Miura, M. Abe, K. Matsuda *et al.*, 2015 Anterograde C1ql1 signaling is required in order to determine and maintain a single-winner climbing fiber in the mouse cerebellum. *Neuron* 85: 316–329. <https://doi.org/10.1016/j.neuron.2014.12.020>
- Kilpeläinen, T. O., M. C. Zillikens, A. Stančáková, F. M. Finucane, J. S. Ried *et al.*, 2011 Genetic variation near IRS1 associates with reduced adiposity and an impaired metabolic profile. *Nat. Genet.* 43: 753–760. <https://doi.org/10.1038/ng.866>
- Kim, Y. J., M. J. Go, C. Hu, C. B. Hong, Y. K. Kim *et al.*, 2011 Large-scale genome-wide association studies in east Asians identify new genetic loci influencing metabolic traits. *Nat. Genet.* 43: 990–995. <https://doi.org/10.1038/ng.939>
- Kraja, A. T., D. Vaidya, J. S. Pankow, M. O. Goodarzi, T. L. Assimes *et al.*, 2011 A bivariate genome-wide approach to metabolic syndrome STAMPEED consortium. *Diabetes* 60: 1329–1339. <https://doi.org/10.2337/db10-1011>
- Kvale, M. N., S. Hesselson, T. J. Hoffmann, Y. Cao, D. Chan *et al.*, 2015 Genotyping informatics and quality control for 100,000 subjects in the genetic epidemiology research on adult health and aging (GERA) cohort. *Genetics* 200: 1051–1060. <https://doi.org/10.1534/genetics.115.178905>
- Lane, J. M., J. Liang, I. Vlasac, S. G. Anderson, D. A. Bechtold *et al.*, 2017 Genome-wide association analyses of sleep disturbance traits identify new loci and highlight shared genetics with neuropsychiatric and metabolic traits. *Nat. Genet.* 49: 274–281. <https://doi.org/10.1038/ng.3749>
- Lanoue, V., A. Usardi, S. M. Sigoillot, M. Talleur, K. Iyer *et al.*, 2013 The adhesion-GPCR BAI3, a gene linked to psychiatric disorders, regulates dendrite morphogenesis in neurons. *Mol. Psychiatry* 18: 943–950. <https://doi.org/10.1038/mp.2013.46>
- Lee, S. H., N. R. Wray, M. E. Goddard, and P. M. Visscher, 2011 Estimating missing heritability for disease from genome-wide association studies. *Am. J. Hum. Genet.* 88: 294–305. <https://doi.org/10.1016/j.ajhg.2011.02.002>
- Lindgren, C. M., I. M. Heid, J. C. Randall, C. Lamina, V. Steinthorsdottir *et al.*, 2009 Genome-wide association scan meta-analysis identifies three loci influencing adiposity and fat distribution. *PLoS Genet.* 5: e1000508 (erratum: *PLoS Genet.* 5). <https://doi.org/10.1371/journal.pgen.1000508>
- Liu, C.-T., K. L. Monda, K. C. Taylor, L. Lange, E. W. Demerath *et al.*, 2013 Genome-wide association of body fat distribution in African ancestry populations suggests new loci. *PLoS Genet.* 9: e1003681. <https://doi.org/10.1371/journal.pgen.1003681>
- Liu, Y.-J., X.-G. Liu, L. Wang, C. Dina, H. Yan *et al.*, 2008 Genome-wide association scans identified CTNBL1 as a novel gene for obesity. *Hum. Mol. Genet.* 17: 1803–1813. <https://doi.org/10.1093/hmg/ddn072>
- Lizcano, F., C. Romero, and D. Vargas, 2011 Regulation of adipogenesis by nuclear receptor PPAR $\gamma$  is modulated by the histone demethylase JMJD2C. *Genet. Mol. Biol.* 34: 19–24. <https://doi.org/10.1590/S1415-47572010005000105>
- Locke, A. E., B. Kahali, S. I. Berndt, A. E. Justice, T. H. Pers *et al.*, 2015 Genetic studies of body mass index yield new insights for obesity biology. *Nature* 518: 197–206. <https://doi.org/10.1038/nature14177>
- Loh, P.-R., G. Tucker, B. K. Bulik-Sullivan, B. J. Vilhjálmsson, H. K. Finucane *et al.*, 2015 Efficient Bayesian mixed-model analysis increases association power in large cohorts. *Nat. Genet.* 47: 284–290. <https://doi.org/10.1038/ng.3190>
- Loh, P.-R., P. Danecek, P. F. Palamara, C. Fuchsberger, Y. A. Reshef *et al.*, 2016 Reference-based phasing using the haplotype reference consortium panel. *Nat. Genet.* 48: 1443–1448. <https://doi.org/10.1038/ng.3679>
- Lotta, L. A., P. Gulati, F. R. Day, F. Payne, H. Ongen *et al.*, 2017 Integrative genomic analysis implicates limited peripheral adipose storage capacity in the pathogenesis of human insulin resistance. *Nat. Genet.* 49: 17–26. <https://doi.org/10.1038/ng.3714>
- Marchini, J., and B. Howie, 2010 Genotype imputation for genome-wide association studies. *Nat. Rev. Genet.* 11: 499–511. <https://doi.org/10.1038/nrg2796>
- Medina, E. A., R. R. Afsari, T. Ravid, S. S. Castillo, K. L. Erickson *et al.*, 2005 Tumor necrosis factor- $\alpha$  decreases Akt protein levels in 3T3-L1 adipocytes via the caspase-dependent ubiquitination of Akt. *Endocrinology* 146: 2726–2735. <https://doi.org/10.1210/en.2004-1074>
- Melka, M. G., M. Bernard, A. Mahboubi, M. Abrahamowicz, A. D. Paterson *et al.*, 2012 Genome-wide scan for loci of adolescent obesity and their relationship with blood pressure. *J. Clin. Endocrinol. Metab.* 97: E145–E150. <https://doi.org/10.1210/jc.2011-1801>
- Meyre, D., J. Delplanque, J.-C. Chèvre, C. Lecoeur, S. Lobbens *et al.*, 2009 Genome-wide association study for early-onset and morbid adult obesity identifies three new risk loci in European populations. *Nat. Genet.* 41: 157–159. <https://doi.org/10.1038/ng.301>
- Minster, R. L., N. L. Hawley, C.-T. Su, G. Sun, E. E. Kershaw *et al.*, 2016 A thrifty variant in CREBRF strongly influences body mass index in Samoans. *Nat. Genet.* 48: 1049–1054. <https://doi.org/10.1038/ng.3620>
- Monda, K. L., G. K. Chen, K. C. Taylor, C. Palmer, T. L. Edwards *et al.*, 2013 A meta-analysis identifies new loci associated with body mass index in individuals of African ancestry. *Nat. Genet.* 45: 690–696. <https://doi.org/10.1038/ng.2608>
- Nagy, R., T. S. Boutin, J. Marten, J. E. Huffman, S. M. Kerr *et al.*, 2017 Exploration of haplotype research consortium imputation for genome-wide association studies in 20,032 Generation Scotland participants. *Genome Med.* 9: 23. <https://doi.org/10.1186/s13073-017-0414-4>
- Nakajima, K., Z. Cui, C. Li, J. Meister, Y. Cui *et al.*, 2016 Gs-coupled GPCR signalling in AgRP neurons triggers sustained increase in food intake. *Nat. Commun.* 7: 10268. <https://doi.org/10.1038/ncomms10268>
- Namjou, B., M. Keddache, K. Marsolo, M. Wagner, T. Lingren *et al.*, 2013 EMR-linked GWAS study: investigation of variation landscape of loci for body mass index in children. *Front. Genet.* 4: 268. <https://doi.org/10.3389/fgene.2013.00268>
- Ng, M. C. Y., J. M. Hester, M. R. Wing, J. Li, J. Xu *et al.*, 2012 Genome-wide association of BMI in African Americans. *Obesity (Silver Spring)* 20: 622–627. <https://doi.org/10.1038/oby.2011.154>
- Ng, M. C. Y., M. Graff, Y. Lu, A. E. Justice, P. Mudgal *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 Genet.* 13: e1006719. <https://doi.org/10.1371/journal.pgen.1006719>
- Okada, Y., M. Kubo, H. Ohmiya, A. Takahashi, N. Kumasaka *et al.*, 2012 Common variants at CDKAL1 and KLF9 are associated with body mass index in east Asian populations. *Nat. Genet.* 44: 302–306. <https://doi.org/10.1038/ng.1086>

- Ortega, F. B., C. J. Lavie, and S. N. Blair, 2016 Obesity and cardiovascular disease. *Circ. Res.* 118: 1752–1770. <https://doi.org/10.1161/CIRCRESAHA.115.306883>
- Pasaniuc, B., N. Zaitlen, H. Shi, G. Bhatia, A. Gusev *et al.*, 2014 Fast and accurate imputation of summary statistics enhances evidence of functional enrichment. *Bioinformatics* 30: 2906–2914. <https://doi.org/10.1093/bioinformatics/btu416>
- Paternoster, L., L. D. Howe, K. Tilling, M. N. Weedon, R. M. Freathy *et al.*, 2011 Adult height variants affect birth length and growth rate in children. *Hum. Mol. Genet.* 20: 4069–4075. <https://doi.org/10.1093/hmg/ddr309>
- Pei, Y.-F., L. Zhang, Y. Liu, J. Li, H. Shen *et al.*, 2014 Meta-analysis of genome-wide association data identifies novel susceptibility loci for obesity. *Hum. Mol. Genet.* 23: 820–830. <https://doi.org/10.1093/hmg/ddt464>
- Pei, Y.-F., H.-G. Ren, L. Liu, X. Li, C. Fang *et al.*, 2017 Genomic variants at 20p11 associated with body fat mass in the European population. *Obesity (Silver Spring)* 25: 757–764. <https://doi.org/10.1002/oby.21775>
- Pradhan, R. N., M. Zachara, and B. Deplancke, 2017 A systems perspective on brown adipogenesis and metabolic activation. *Obes. Rev. Off. J. Int. Assoc. Study Obes.* 18: 65–81. <https://doi.org/10.1111/obr.12512>
- Riddy, D. M., P. Delerive, R. J. Summers, P. M. Sexton, and C. J. Langmead, 2018 G protein-coupled receptors targeting insulin resistance, obesity, and type 2 diabetes mellitus. *Pharmacol. Rev.* 70: 39–67. <https://doi.org/10.1124/pr.117.014373>
- Ried, J. S., M. J. Jeff, A. Y. Chu, J. L. Bragg-Gresham, and J. van Dongen *et al.*, 2016 A principal component meta-analysis on multiple anthropometric traits identifies novel loci for body shape. *Nat. Commun.* 7: 13357. <https://doi.org/10.1038/ncomms13357>
- Roadmap Epigenomics Consortium, Kundaje, A., W. Meuleman, J. Ernst, M. Bilenky, *et al.*, 2015 Integrative analysis of 111 reference human epigenomes. *Nature* 518: 317–330. <https://doi.org/10.1038/nature14248>
- Robert, S. A., and E. N. Reither, 2004 A multilevel analysis of race, community disadvantage, and body mass index among adults in the US. *Soc. Sci. Med.* 59: 2421–2434. <https://doi.org/10.1016/j.socscimed.2004.03.034>
- Robinson, M. R., G. English, G. Moser, L. R. Lloyd-Jones, M. A. Triplett *et al.*, 2017 Genotype-covariate interaction effects and the heritability of adult body mass index. *Nat. Genet.* 49: 1174–1181. <https://doi.org/10.1038/ng.3912>
- Salinas, Y. D., L. Wang, and A. T. DeWan, 2016 Multiethnic genome-wide association study identifies ethnic-specific associations with body mass index in Hispanics and African Americans. *BMC Genet.* 17: 78. <https://doi.org/10.1186/s12863-016-0387-0>
- Scannell Bryan, M., M. Argos, B. Pierce, L. Tong, M. Rakibuz-Zaman *et al.*, 2014 Genome-wide association studies and heritability estimates of body mass index related phenotypes in Bangladeshi adults. *PLoS One* 9: e105062. <https://doi.org/10.1371/journal.pone.0105062>
- Scherag, A., C. Dina, A. Hinney, V. Vatin, S. Scherag *et al.*, 2010 Two new loci for body-weight regulation identified in a joint analysis of genome-wide association studies for early-onset extreme obesity in French and German study groups. *PLoS Genet.* 6: e1000916. <https://doi.org/10.1371/journal.pgen.1000916>
- Scuteri, A., S. Sanna, W.-M. Chen, M. Uda, G. Albai *et al.*, 2007 Genome-wide association scan shows genetic variants in the FTO gene are associated with obesity-related traits. *PLoS Genet.* 3: e115. <https://doi.org/10.1371/journal.pgen.0030115>
- Shungin, D., T. W. Winkler, D. C. Croteau-Chonka, T. Ferreira, A. E. Locke *et al.*, 2015 New genetic loci link adipose and insulin biology to body fat distribution. *Nature* 518: 187–196. <https://doi.org/10.1038/nature14132>
- Sloop, K. W., P. J. Emmerson, M. A. Statnick, and F. S. Willard, 2018 The current state of GPCR-based drug discovery to treat metabolic disease. *Br. J. Pharmacol.* <https://doi.org/10.1111/bph.14157>
- Southam, L., A. Gilly, D. Süveges, A.-E. Farmaki, J. Schwartzentruber *et al.*, 2017 Whole genome sequencing and imputation in isolated populations identify genetic associations with medically-relevant complex traits. *Nat. Commun.* 8: 15606. <https://doi.org/10.1038/ncomms15606>
- Speliotes, E. K., C. J. Willer, S. I. Berndt, K. L. Monda, G. Thorleifsson *et al.*, 2010 Association analyses of 249,796 individuals reveal 18 new loci associated with body mass index. *Nat. Genet.* 42: 937–948. <https://doi.org/10.1038/ng.686>
- Sudlow, C., J. Gallacher, N. Allen, V. Beral, P. Burton *et al.*, 2015 UK biobank: an open access resource for identifying the causes of a wide range of complex diseases of middle and old age. *PLoS Med.* 12: e1001779. <https://doi.org/10.1371/journal.pmed.1001779>
- Sung, Y. J., L. Pérusse, M. A. Sarzynski, M. Fornage, S. Sidney *et al.*, 2016 Genome-wide association studies suggest sex-specific loci associated with abdominal and visceral fat. *Int. J. Obes.* 40: 662–674. <https://doi.org/10.1038/ijo.2015.217>
- Tachmazidou, I., D. Süveges, J. L. Min, G. R. S. Ritchie, J. Steinberg *et al.*, 2017 Whole-genome sequencing coupled to imputation discovers genetic signals for anthropometric traits. *Am. J. Hum. Genet.* 100: 865–884. <https://doi.org/10.1016/j.ajhg.2017.04.014>
- Thorleifsson, G., G. B. Walters, D. F. Gudbjartsson, V. Steinthorsdottir, P. Sulem *et al.*, 2009 Genome-wide association yields new sequence variants at seven loci that associate with measures of obesity. *Nat. Genet.* 41: 18–24. <https://doi.org/10.1038/ng.274>
- Turcot, V., Y. Lu, H. M. Highland, C. Schurmann, A. E. Justice *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 (erratum: *Nat. Genet.* 50: 765–766); (erratum: *Nat. Genet.* 50: 766–767). <https://doi.org/10.1038/s41588-017-0011-x>
- Wagner, D. R., and V. H. Heyward, 2000 Measures of body composition in blacks and whites: a comparative review. *Am. J. Clin. Nutr.* 71: 1392–1402. <https://doi.org/10.1093/ajcn/71.6.1392>
- Wang, H.-J., A. Hinney, J.-Y. Song, A. Scherag, X.-R. Meng *et al.*, 2016 Association of common variants identified by recent genome-wide association studies with obesity in Chinese children: a case-control study. *BMC Med. Genet.* 17: 7. <https://doi.org/10.1186/s12881-016-0268-4>
- Wang, K., W.-D. Li, C. K. Zhang, Z. Wang, J. T. Glessner *et al.*, 2011 A genome-wide association study on obesity and obesity-related traits. *PLoS One* 6: e18939 [corrigenda: *PLoS One* 7 (2012)]. <https://doi.org/10.1371/journal.pone.0018939>
- Wen, W., Y.-S. Cho, W. Zheng, R. Dorajoo, N. Kato *et al.*, 2012 Meta-analysis identifies common variants associated with body mass index in east Asians. *Nat. Genet.* 44: 307–311. <https://doi.org/10.1038/ng.1087>
- Wen, W., W. Zheng, Y. Okada, F. Takeuchi, Y. Tabara *et al.*, 2014 Meta-analysis of genome-wide association studies in East Asian-ancestry populations identifies four new loci for body mass index. *Hum. Mol. Genet.* 23: 5492–5504. <https://doi.org/10.1093/hmg/ddu248>
- Wen, W., N. Kato, J.-Y. Hwang, X. Guo, Y. Tabara *et al.*, 2016 Genome-wide association studies in East Asians identify new loci for waist-hip ratio and waist circumference. *Sci. Rep.* 6: 17958. <https://doi.org/10.1038/srep17958>
- Wheeler, E., N. Huang, E. G. Bochukova, J. M. Keogh, S. Lindsay *et al.*, 2013 Genome-wide SNP and CNV analysis identifies common and low-frequency variants associated with severe early-onset obesity. *Nat. Genet.* 45: 513–517. <https://doi.org/10.1038/ng.2607>

- Willer, C. J., E. K. Speliotes, R. J. F. Loos, S. Li, C. M. Lindgren *et al.*, 2009 Six new loci associated with body mass index highlight a neuronal influence on body weight regulation. *Nat. Genet.* 41: 25–34. <https://doi.org/10.1038/ng.287>
- Winkler, T. W., A. E. Justice, M. Graff, L. Barata, M. F. Feitosa *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 [corrigenda: *PLoS Genet.* 12: e1006166 (2016)]. <https://doi.org/10.1371/journal.pgen.1005378>
- Xie, D., A. P. Boyle, L. Wu, J. Zhai, T. Kawli *et al.*, 2013 Dynamic trans-acting factor colocalization in human cells. *Cell* 155: 713–724. <https://doi.org/10.1016/j.cell.2013.09.043>
- Xie, X., Z. Yi, S. Sinha, M. Madan, B. P. Bowen *et al.*, 2016 Proteomics analyses of subcutaneous adipocytes reveal novel abnormalities in human insulin resistance. *Obesity (Silver Spring)* 24: 1506–1514. <https://doi.org/10.1002/oby.21528>
- Yang, F., X. D. Chen, L. J. Tan, J. Shen, D. Y. Li *et al.*, 2014 Genome wide association study: searching for genes underlying body mass index in the Chinese. *Biomed. Environ. Sci. BES* 27: 360–370. <https://doi.org/10.3967/bes2014.061>
- Yang, J., S. H. Lee, M. E. Goddard, and P. M. Visscher, 2011a GCTA: a tool for genome-wide complex trait analysis. *Am. J. Hum. Genet.* 88: 76–82. <https://doi.org/10.1016/j.ajhg.2010.11.011>
- Yang, J., M. N. Weedon, S. Purcell, G. Lettre, K. Estrada *et al.*, 2011b Genomic inflation factors under polygenic inheritance. *Eur. J. Hum. Genet.* 19: 807–812. <https://doi.org/10.1038/ejhg.2011.39>
- Yang, J., R. J. F. Loos, J. E. Powell, S. E. Medland, E. K. Speliotes *et al.*, 2012 FTO genotype is associated with phenotypic variability of body mass index. *Nature* 490: 267–272. <https://doi.org/10.1038/nature11401>
- Yang, J., A. Bakshi, Z. Zhu, G. Hemani, A. A. E. Vinkhuyzen *et al.*, 2015 Genetic variance estimation with imputed variants finds negligible missing heritability for human height and body mass index. *Nat. Genet.* 47: 1114–1120. <https://doi.org/10.1038/ng.3390>
- Zheng, J., Y. Li, G. R. Abecasis, and P. Scheet, 2011 A comparison of approaches to account for uncertainty in analysis of imputed genotypes. *Genet. Epidemiol.* 35: 102–110. <https://doi.org/10.1002/gepi.20552>

*Communicating editor: N. Wray*