Protein-coding variants implicate novel genes related to lipid homeostasis contributing to body-fat distribution

Body-fat distribution is a risk factor for adverse cardiovascular health consequences. We analyzed the association of body-fat distribution, assessed by waist-to-hip ratio adjusted for body mass index, with 228,985 predicted coding and splice site variants available on exome arrays in up to 344,369 individuals from five major ancestries (discovery) and 132,177 European-ancestry individuals (validation). We identified 15 common (minor allele frequency, MAF ≥5%) and nine low-frequency or rare (MAF <5%) coding novel variants. Pathway/gene set enrichment analyses identified lipid particle, adiponectin, abnormal white adipose tissue physiology and bone development and morphology as important contributors to fat distribution, while cross-trait associations highlight cardiometabolic traits. In functional follow-up analyses, specifically in Drosophila RNAi-knockdowns, we observed a significant increase in the total body triglyceride levels for two genes (DNAH10 and PLXND1). We implicate novel genes in fat distribution, stressing the importance of interrogating low-frequency and protein-coding variants.

Central body-fat distribution, as assessed by waist-to-hip ratio (WHR), is a heritable and a well-established risk factor for adverse metabolic outcomes. Lower values of WHR are associated with lower risk of cardiometabolic diseases such as type 2 diabetes, or differences in bone structure and gluteal muscle mass. These epidemiological associations are consistent with our previously reported genome-wide association study (GWAS) results of 49 loci associated with WHR (after adjusting for body mass index, WHRadjBMI)1. Notably, genetic predisposition to higher WHRadjBMI is associated with increased risk of type 2 diabetes and coronary heart disease (CHD), which seems to be causal.

Recently, large-scale studies have identified roughly 125 common loci for multiple measures of central obesity, primarily non-coding variants of relatively modest effect. Large-scale interrogation of coding and splice site variants, including both common (minor allele frequency, MAF ≥2%) and low-frequency or rare variants (LF/RV: MAF <5%), may lead to additional insights into the etiology of central obesity. Here, we identify and characterize such variants associated with WHRadjBMI using ExomeChip array genotypes.

Results

Protein-coding and splice site variation associations. We conducted a two-stage fixed-effects meta-analysis testing additive and recessive models to detect protein-coding genetic variants that influence WHRadjBMI (Methods and Fig. 1). Stage 1 included up to 228,985 variants (218,195 low-frequency or rare) in up to 344,369 individuals from 74 studies of European, South and East Asian, African and Hispanic/Latino descent individuals (Supplementary Data 1–3). Stage 2 assessed 70 suggestive (P < 2 × 10⁻⁸) stage 1 variants in two cohorts, UK Biobank (UKBB) and deCODE for a total stage 1 and 2 sample size of 476,546 (88% European). Of the 70 variants considered, two common and five low-frequency or rare were not available in stage 2 (Tables 1 and 2 and Supplementary Data 4–6). Variants are considered novel and statistically significant if they were greater than one megabase (Mb) from a previously identified WHRadjBMI SNP and achieve array-wide significance (P < 2 × 10⁻⁸, stage 1 and 2).

In our primary meta-analysis, including all stage 1 and 2 samples, we identified 48 coding variants (16 novel) across 43 genes, 47 assuming an additive model and one under a recessive model (Table 1 and Supplementary Figs. 1–4). Due to possible heterogeneity, we also performed European-only meta-analysis. Here, four additional coding variants were significant (three novel) assuming an additive model (Table 1 and Supplementary Figs. 5–8). Of these 52 significant variants, 11 were low-frequency or rare and displayed larger effect estimates than many previously reported common variants, including seven novel variants in ACVR1C and DARS2. Variants with MAF ≤1% had effect sizes approximately three times greater than those of common variants (MAF ≥5%). Despite large sample size, we cannot rule out the possibility that additional low-frequency or rare variants with smaller effects exist (see estimated 80% power in Fig. 2). However, in the absence of common variants with similarly large effects, our results point to the importance of investigating LF/RVs.

Given established sex differences in the genetic underpinnings of WHRadjBMI, we also performed sex-stratified analyses. We detected four additional novel variants that exhibit significant sex-specific effects (Psex < 7.14 × 10⁻⁶, Methods) in UGGT2 and MMP14 for men; and DSTYK1 and ANGPTL4 for women (Table 2 and Supplementary Figs. 9–15); including low-frequency or rare in UGGT2 and ANGPTL4 (MAFmen = 0.6% and MAVFwomen = 1.9%, respectively). Additionally, 14 variants from the sex-combined meta-analyses displayed significantly stronger effects in women, including the novel, low-frequency ACVR1C variant (rs55920843, MAF = 1.1%). Overall, 19 of the 56 variants (32%) identified across all meta-analyses (48 from all ancestry, 4 from European-only and 4 from sex-stratified analyses) showed significant sex-specific effects on WHRadjBMI: 16 variants with significantly stronger effects in women and three in men (Fig. 1).

In summary, we identified 56 array-wide significant coding variants (P < 2 × 10⁻⁷); 43 common (14 novel) and 13 low-frequency or rare (9 novel). For the 55 significant variants from the additive model, we examined potential collider bias (Methods, Supplementary Table 1 and Supplementary Note). Overall, 51 of 55 variants were robust to collider bias. Of these, 25 variants were nominally associated with BMI (P < 0.05), yet effect sizes changed little after correction for potential biases (15% change in...
For four of the 55 SNPs (rs141845046, rs1034405, rs3617, rs9469913), attenuation following correction was noted ($P_{\text{corrected}} > 9 \times 10^{-4}$, 0.05/55), including one novel variant, rs1034405 in C3orf18, demonstrating a possible overestimation of these effects in the current analysis.

Using stage 1 results, we then aggregated LF/RVs across genes and tested their joint effect with SKAT and burden tests 19 (Supplementary Table 2 and Methods). None of the five genes that reached array-wide significance ($P < 2.5 \times 10^{-6}$, 0.05/16,222 genes tested: RAPGEF3, ACVR1C, ANGPTL4, DNAI1 and NOP2) remained significant after conditioning on the most significant single variant.

**Conditional analyses.** We next implemented conditional analyses to determine (1) the total number of independent signals identified and (2) whether the 33 variants near known GWAS signals ($<+/-1$ Mb) represent independent novel associations. We used approximate joint conditional analyses to test for independence in
### Table 1 | Association results for combined sexes. Association results based on an additive or recessive model for coding variants that met array-wide significance (P < 2 × 10⁻¹⁴) in the sex-combined meta-analyses

| Locus | Chr/position (GRCh37)⁸ | rsID | EA | OA | Gene¹ | Amino acid change¹ | If locus is known, nearby (<1Mb) published variant(s) | N  | EAF | P² | s.e.m. | P value for sex-heterogeneity⁷ | Other criteria for significance⁶ |
|-------|------------------------|------|----|----|-------|-------------------|---------------------------------|----|-----|-----|-------|--------------------------------|---------------------------------|

#### Variants in novel loci

**All ancestry additive model sex-combined analyses**

| Locus | Chr/position (GRCh37)⁸ | rsID | EA | OA | Gene¹ | Amino acid change¹ | If locus is known, nearby (<1Mb) published variant(s) | N  | EAF | P² | s.e.m. | P value for sex-heterogeneity⁷ | Other criteria for significance⁶ |
|-------|------------------------|------|----|----|-------|-------------------|---------------------------------|----|-----|-----|-------|--------------------------------|---------------------------------|

#### Variants in previously identified loci

**All ancestry additive model sex-combined analyses**

| Locus | Chr/position (GRCh37)⁸ | rsID | EA | OA | Gene¹ | Amino acid change¹ | If locus is known, nearby (<1Mb) published variant(s) | N  | EAF | P² | s.e.m. | P value for sex-heterogeneity⁷ | Other criteria for significance⁶ |
|-------|------------------------|------|----|----|-------|-------------------|---------------------------------|----|-----|-----|-------|--------------------------------|---------------------------------|

Continued
Table 1 | Association results for combined sexes. Association results based on an additive or recessive model for coding variants that met array-wide significance (P < 2 × 10^-7) in the sex-combined meta-analyses (Continued)

| Locus ( +/- 1 Mb of a given variant) | Chr position (GRCh37) | rsID | EA OA | Gene | Amino acid change | If locus is known, nearby (<1 Mb) published variant(s) | N | EAF | β* | s.e.m. | P value | P value for sex heterogeneity | Other criteria for significance |
|-------------------------------------|-----------------------|------|-------|-------|------------------|-------------------------------------------------|-----|------|-----|--------|---------|-------------------------------|-------------------------------|
| 11 | 6:127476516 | rs1892172 | A G | RSPO3 | synonymous | rs1961815, rs72995041, rs1936805 | 476,358 | 0.543 | 0.031 | 0.002 | 2.6E-47 | 7.7E-09 | |
| 12 | 6:127767954 | rs139745911 | A G | KIAA0408 | PS045 | | 391,469 | 0.010 | 0.103 | 0.012 | 6.8E-19 | 2.0E-04 | |
| 13 | 7:73012042 | rs53320626 | G A | MLXIP | A358V | rs6976930 | 451,158 | 0.880 | 0.020 | 0.003 | 1.8E-09 | 1.5E-01 | |
| 14 | 7:73020337 | rs3812316 | C G | Q241H | | 454,738 | 0.871 | 0.021 | 0.003 | 2.0E-10 | 5.8E-02 | |
| 15 | 10:95931087 | rs17471407 | T G | PCE1 | R240L | rs10786152 | 476,475 | 0.173 | 0.083 | 0.003 | 2.5E-11 | 9.5E-01 | |
| 16 | 11:64031241 | rs24516799 | T C | PCE3 | S778L | rs11231693 | 476,457 | 0.061 | 0.034 | 0.004 | 9.1E-15 | 1.3E-04 | |
| 17 | 12:123444507 | rs58843120 | G T | ADB9 | F29L | rs4765219, rs863750 | 466,498 | 0.987 | 0.053 | 0.009 | 1.3E-08 | 3.5E-01 | |
| 18 | 12:124265687 | rs11057353 | T C | DNAH10 | S228P | | 476,360 | 0.373 | 0.018 | 0.002 | 2.1E-16 | 2.7E-08 | |
| 19 | 12:124427301 | rs33491281 | C T | T1785M | | | 476,395 | 0.889 | 0.025 | 0.003 | 2.9E-14 | 3.1E-08 | |

All ancestry recessive model sex-combined analyses

| Locus | Chr position (GRCh37) | rsID | EA OA | Gene | Amino acid change | If locus is known, nearby (<1 Mb) published variant(s) | N | EAF | β* | s.e.m. | P value | P value for sex heterogeneity | Other criteria for significance |
|-------|-----------------------|------|-------|-------|------------------|-------------------------------------------------|-----|------|-----|--------|---------|-------------------------------|-------------------------------|
| 20 | 17:17425631 | rs897453 | C T | PEMT | V58L | | 476,546 | 0.569 | 0.025 | 0.004 | 4.1E-11 | 8.2E-01 | |

European ancestry additive model sex-combined analyses

| Locus | Chr position (GRCh37) | rsID | EA OA | Gene | Amino acid change | If locus is known, nearby (<1 Mb) published variant(s) | N | EAF | β* | s.e.m. | P value | P value for sex heterogeneity | Other criteria for significance |
|-------|-----------------------|------|-------|-------|------------------|-------------------------------------------------|-----|------|-----|--------|---------|-------------------------------|-------------------------------|
| 6 | 3:129293526 | rs2255703 | T C | PLXN1D | M870V | | 420,520 | 0.620 | 0.014 | 0.002 | 3.1E-09 | 1.6E-04 | |

Abbreviations: GRCh37: human genome assembly build 37; rsID: based on all SNPs; VEP: Ensembl Variant Effect Predictor toolset; GTEx: Genotype-Tissue Expression project; s.d.: standard deviation; s.e.m.: standard error of the mean; N: sample size; EAF: effect allele frequency; EA: effect allele; OA: other allele. *Coding variants refer to variants located in the exons and splicing junction regions. Variant positions are reported according to human genome assembly build 37 and their alleles are coded based on the positive strand. *Gene the variant falls in and amino acid change from the most abundant coding transcript is shown (protein annotation is based on the VEP toolset and transcript abundance from GTEx database). Previously published variants within +/-1 Mb from Graff et al.14 and Ng et al.16. *Effect size is based on the s.d. per effect allele and for sex heterogeneity, testing for difference between women-specific and men-specific beta estimates and standard errors, was calculated using EasyStrata: Winkler et al.11. Bold P values met significance threshold after Bonferroni correction (P < 7.14E-04; that is, 0.05/70 variants). rs1334576 in RREB1 is a new signal in a known locus that is independent from the known signal, rs1294410 (see Supplementary Table 4 and Supplementary Data 7); rs139745911 in KIAA0408 is a secondary signal in a known locus (see Supplementary Table 3). Each flag indicates that a secondary criterion for significance may not be met: G, P > 5 x 10^-4 (GWAS significant); C, Association signal was not robust against collider bias; 5 variant was not available in stage 2 studies for validation of stage 1 association.

stage 1 (Methods and Supplementary Table 3)19. Only the RSPO3-KIAA0408 locus contains two independent variants 291 kilo bases apart, rs1892172 in RSPO3 (MAF = 46.1%, P condition = 4.37 × 10^-23 in the combined sexes and P condition = 2.4 × 10^-20 in women) and rs139745911 in KIAA0408 (MAF = 0.9%, P condition = 3.68 × 10^-11 in combined sexes and P condition = 1.46 × 10^-11 in women, Supplementary Table 3 and Fig. 3a). For the 33 variants within 1 Mb of previously identified WHRadjBMI SNPs, sex-combined conditional analyses identified one coding variant representing a novel independent signal in a known locus (RREB1; stage 1 meta-analysis, rs1334576, MAF = 44%, P condition = 3.06 × 10^-7, Supplementary Table 4 and Fig. 3b; UKBB analysis, P condition = 1.24 × 10^-4, Supplementary Table 7). In summary, we identified 56 WHRadjBMI-associated coding variants in 41 independent association signals, 24 of which are new or independent of known GWAS-identified tag SNPs (either > +/-1 Mb or array-wide significant following conditional analyses) (Fig. 1). Thus, we identified 15 common and 9 low-frequency or rare novel and independent variants following conditional analyses.

Gene set and pathway enrichment analysis. To determine whether significant coding variants highlight novel or previously identified biological pathways, we applied two complementary methods, EC-DEPICT (ExomeChip Data-driven Expression-Prioritized Integration for Complex Traits)20,21 and PASCAL22 (Methods). For PASCAL all variants were used, for EC-DEPICT we examined only 361 variants with suggestive significance (P < 5 × 10^-4)20,21 from the all ancestries combined sexes analysis (which after clumping and filtering became 101 lead variants in 101 genes). We separately analyzed variants that exhibited significant sex-specific effects (P value < 5 × 10^-4).

The sex-combined analyses identified 49 significantly enriched gene sets (false discovery rate, FDR < 0.05) that grouped into 25 meta-gene sets (Supplementary Note and Supplementary Data 8). We noted a cluster of meta-gene sets with direct relevance to metabolic aspects of obesity (‘enhanced lipolysis’, ‘abnormal glucose homeostasis’, ‘increased circulating insulin level’ and ‘decreased susceptibility to diet-induced obesity’); we observed two significant adiponectin-related gene sets in these meta-gene sets. While
Table 2 | Association results for sex-stratified analyses. Association results based on an additive or recessive model for coding variants that met array-wide significance ($P < 2 \times 10^{-8}$) in the sex-specific meta-analyses and reach Bonferroni corrected $P$ value for sex heterogeneity ($P_{\text{sexhet}} < 7.14 \times 10^{-4}$).

| Locus (±/−1 Mb of a given variant) | Chr.position (GRCh37) | rsID | EA | OA | Gene | Amino acid change | In sex-combined analyses | If locus is known, nearby (< 1 Mb) published variant(s) | $P$ value for sex-heterogeneity | N | EAF | s.e.m. | P   | N | EAF | s.e.m. | P  |
|-----------------------------------|------------------------|------|----|----|------|-----------------|------------------------|--------------------------|--------------------------------|----|-----|-------|-----|----|-----|-------|-----|
| **Variants in novel loci**        |                        |      |    |    |      |                 |                        |                          |                                |    |     |       |     |    |     |       |     |
| All ancestry additive model men-only analyses |                        |      |    |    |      |                 |                        |                          |                                |    |     |       |     |    |     |       |     |
| 1                                 | 13:96666597            | rs148108950 | A | G | UGGT2 | P175L | No | − | 1.5E-06 | 203,009 | 0.006 | 0.130 | 0.024 | 6.1E-08 | 221,390 | 0.004 | −0.044 | 0.027 | 1.1E-01 | G |
| 2                                 | 14:23312594            | rs1042704 | A | G | MMP14 | D273N | No | − | 2.6E-04 | 226,646 | 0.202 | 0.021 | 0.004 | 2.6E-08 | 250,018 | 0.197 | 0.002 | 0.004 | 6.1E-01 | 1 |
| **All ancestry additive model women-only analyses** |                        |      |    |    |      |                 |                        |                          |                                |    |     |       |     |    |     |       |     |
| 3                                 | 1:20513041            | rs3851294 | G | A | DTYK | C64IR | No | − | 9.8E-08 | 225,803 | 0.914 | −0.005 | 0.005 | 3.4E-01 | 249,471 | 0.912 | 0.034 | 0.005 | 4.5E-11 | 1 |
| 4                                 | 2:158412701           | rs55920843 | T | G | ACVR1C | N150H | Yes | − | 1.7E-07 | 210,071 | 0.989 | 0.006 | 0.015 | 7.2E-01 | 245,808 | 0.989 | 0.113 | 0.014 | 1.7E-15 | 1 |
| 5                                 | 19:8429323            | rs16843064 | G | A | ANGPT1L | E40K | No | − | 1.3E-07 | 203,098 | 0.981 | −0.017 | 0.011 | 1.4E-01 | 243,351 | 0.981 | 0.064 | 0.011 | 1.2E-09 | 1 |
| **Variants in previously identified loci** |                        |      |    |    |      |                 |                        |                          |                                |    |     |       |     |    |     |       |     |
| All ancestry additive model women-only analyses |                        |      |    |    |      |                 |                        |                          |                                |    |     |       |     |    |     |       |     |
| 1                                 | 1:54987704            | rs141845046 | C | T | ZBTB7 | P190S | Yes | rs905938 | 7.9E-07 | 226,709 | 0.975 | 0.004 | 0.010 | 6.9E-01 | 250,084 | 0.977 | 0.070 | 0.010 | 2.3E-13 | 1 |
| 2                                 | 2:165551201           | rs7607980 | T | C | COBL1 | N941D | Yes | rs128249 | 3.0E-30 | 173,600 | 0.880 | −0.018 | 0.005 | 5.8E-04 | 216,636 | 0.878 | 0.062 | 0.005 | 6.7E-39 | 1 |
| 3                                 | 3:12913788            | rs6266958 | C | T | EFCAB12 | R197H | Yes | rs10804591 | 9.3E-05 | 226,690 | 0.937 | 0.018 | 0.006 | 3.1E-03 | 250,045 | 0.936 | 0.051 | 0.006 | 8.1E-18 | 1 |
| 4                                 | 4:89625427            | rs1804080 | G | C | HERC3 | E946Q | Yes | rs9991328 | 4.1E-06 | 222,556 | 0.839 | 0.008 | 0.004 | 6.6E-02 | 223,877 | 0.837 | 0.034 | 0.004 | 2.1E-16 | 1 |
| 5                                 | 6:12747616            | rs1892172 | T | C | FAM13A | V443I | Yes | rs11961815, rs7299040, rs1936605 | 7.7E-09 | 226,677 | 0.541 | 0.018 | 0.003 | 5.6E-10 | 250,034 | 0.545 | 0.042 | 0.003 | 3.4E-48 | 1 |
| 6                                 | 11:64303124           | rs35169799 | T | C | PLEC3 | S778L | Yes | rs11231693 | 1.3E-04 | 226,713 | 0.061 | 0.016 | 0.006 | 9.6E-03 | 250,097 | 0.061 | 0.049 | 0.006 | 6.7E-16 | 1 |
| 7                                 | 12:42665687           | rs11057353 | T | C | DNAH10 | S228P | Yes | rs4765219 | 2.7E-08 | 226,659 | 0.370 | 0.005 | 0.003 | 8.3E-02 | 250,054 | 0.376 | 0.029 | 0.003 | 3.1E-22 | 1 |
| 8                                 | 12:42330311           | rs34934281 | C | T | TSHB5M | P104S | Yes | rs863750 | 3.1E-08 | 226,682 | 0.891 | 0.006 | 0.005 | 1.9E-01 | 250,066 | 0.887 | 0.043 | 0.005 | 1.4E-20 | 1 |
| 9                                 | 12:42472306           | rs11057401 | T | A | CCDC92 | S53C | Yes | rs13974991 | 5.5E-11 | 223,324 | 0.701 | 0.013 | 0.003 | 4.3E-05 | 244,678 | 0.689 | 0.043 | 0.003 | 1.0E-11 | 1 |

Abbreviations: GRCh37, human genome assembly build 37; rsID, based on dbSNP; VEP, Ensembl Variant Effect Predictor toolset; GTEx, Genotype-Tissue Expression project; N, sample size; EA, effect allele; OA, other allele; EAF, effect allele frequency. Coding variants refer to variants located in the exons and splicing junction regions. Bonferroni corrected $P$ value for the number of SNPs tested for sex-heterogeneity $< 7.14 \times 10^{-4}$; that is, 0.05/70 variants. Variant positions are reported according to human genome assembly build 37 and their alleles are coded based on the positive strand. The gene variant falls in and amino acid change from the most abundant coding transcript is shown (protein annotation is based on the VEP toolset and transcript abundance from GTEx database). Variants were also identified as array-wide significant in the sex-combined analyses. Previously published variants within $±/−1$ Mb are from Shungin et al. Abbreviation for $P$ value for sex heterogeneity; $\times 10^{-4}$ denotes $P > 5 \times 10^{-4}$.
Fig. 2 | Minor allele frequency compared to estimated effect. This scatter plot displays the relationship between MAF and the estimated effect (β) for each significant coding variant in our meta-analyses. All novel WHRadjBMI variants are highlighted in orange, and variants identified only in sex-specific analyses by triangles. Eighty percent power was calculated based on the total sample size in the stage 1 and 2 meta-analysis and only in sex-specific analyses by triangles. Eighty percent power was calculated based on the total sample size in the stage 1 and 2 meta-analysis and P = 2 × 10⁻⁵. Estimated effects are shown in original units (cm cm⁻¹) calculated by using effect sizes in s.d. units times SD of WHR in the ARIC study (sexes combined = 0.067, men = 0.052, women = 0.080).

Fig. 3 | Regional association plots for known loci with novel coding signals identified by conditional analyses. Point color reflects r² calculated from the ARIC dataset. a. There are two independent variants in RSPO3 and KIAA0408, based on results from the stage 1 all ancestry women (N = 180,131 for RSPO3 and 139,056 for KIAA0408). b. We have a variant in RREB1 that is independent of the GWAS variant rs1294410, based on results from the stage 1 all ancestry sex-combined individuals (N = 319,090; GWAS signal tagging variant rs1294421, rs1294410-rs1294421, r² = 0.85, 1000 Genomes Phase 3 version 5 all ancestries reference set).
We also compared the coding pathways to those identified in the total previous GWAS effort (using both coding and regulatory variants) identifying a total of 158 gene sets. Forty-two gene sets were enriched in both analyses, and we found high concordance in the meta-gene set. Annotations for the genes indicate (1) the minor allele frequency of the significant ExomeChip (EC) variant (blue; if multiple variants, the lowest-frequency variant was kept), (2) whether the variant’s P value reached array-wide significance (<2×10\(^{-5}\)) or suggestive significance (<5×10\(^{-5}\)) (shades of purple), (3) whether the variant was novel, overlapping ‘relaxed’ GWAS signals from Shungin et al.\(^{10}\) (GWAS P<5×10\(^{-5}\)), or overlapping ‘stringent’ GWAS signals (GWAS P<5×10\(^{-3}\)) (pink) and (4) whether the gene was included in the gene set enrichment analysis or excluded by filters (shades of brown/orange) (Methods and Supplementary Note). Annotations for the gene sets indicate if the meta-gene set was found significant (shades of green; FDR<0.01, or not significant) in the DEPICT analysis of GWAS results from Shungin et al.\(^{10}\).

Fig. 4 | Heat maps showing DEPICT gene set enrichment results from the stage 1 all ancestry sex-combined individuals (N = 344,369). For any given square, the color indicates how strongly the corresponding gene (x axis) is predicted to belong to the reconstituted gene set (y axis). This value is based on the gene’s z-score for gene set inclusion in DEPICT’s reconstituted gene sets, where red indicates a higher and blue a lower z-score. To visually reduce redundancy and increase clarity, we chose one representative ‘meta-gene set’ for each group of highly correlated gene sets based on

We compared the coding pathways to those identified in the total previous GWAS effort (using both coding and regulatory variants) identifying a total of 158 gene sets. Forty-two gene sets were enriched in both analyses, and we found high concordance in the -log10 (P values) between ExomeChip and GWAS gene set enrichment (Pearson’s r (coding versus regulatory only) = 0.38, P<10\(^{-3}\); Pearson’s r (coding versus coding + regulatory) = 0.51, P<10\(^{-15}\)). Nonetheless, some gene sets were enriched specifically for variants in coding regions (for example, decreased susceptibility to diet-induced obesity, abnormal skeletal morphology) or unique to variants in regulatory regions (for example, transcriptional regulation of white adipocytes) (Supplementary Fig. 17).

The EC-DEPICT and PASCAL results showed a moderate but strongly significant correlation for (EC-DEPICT and the PASCAL max statistic, r = 28, P = 9.8×10\(^{-21}\); for EC-DEPICT and the PASCAL sum statistic, r = 0.287, P = 5.42×10\(^{-22}\)). Common gene sets strongly implicate a role for skeletal biology, glucose homeostasis/insulin signaling and adipocyte biology (Supplementary Fig. 18).

Cross-trait associations. To assess the clinical relevance of our identified variants with cardiometabolic, anthropometric and reproductive traits, we conducted association lookups from existing ExomeChip studies of 15 traits (Supplementary Data 11 and Supplementary Fig. 19)\(^{21,25-29}\). Variants in STAB1 and PLCB3 displayed the greatest number of significant associations with seven different traits (P<9.8×10\(^{-4}\), 0.05/51 variants tested). Also, these two genes cluster together with RSPO3, DnA1H10, MNS1, COBL1, Ccdc92 and ITIH3. The WHR-increasing alleles in this cluster exhibit a previously described pattern of increased cardiometabolic risk (for example, increased fasting insulin, two-hour glucose (TwoHGl)), and triglycerides and decreased high-density lipoprotein cholesterol (HDL)), but also decreased BMI\(^{30-36}\). The impact of central obesity may be causal, as a 1 s.d. increase in genetic risk of central adiposity was previously associated with higher total cholesterol, triglycerides, fasting insulin and TwoHGl and lower HDL\(^{3}\).

We conducted a search in the NHGRI-EBI GWAS Catalog\(^{37,38}\) to determine whether our variants are in high linkage disequilibrium (R\(^2\) > 0.7) with variants associated with traits or diseases not covered by our cross-trait lookups (Supplementary Data 12). We identified several cardiometabolic traits (adiponectin, CHD and so on), diet/behavioral traits potentially related to obesity (carbohydrate, fat intake and so on), behavioral and neurological traits (schizophrenia, bipolar disorder and so on) and inflammatory or autoimmune diseases (Crohn’s Disease, multiple sclerosis and so on).

Given the established correlation between total body-fat percentage and WHR of up to 0.483\(^{39-41}\), we examined the association of our top exome variants with both total body-fat percentage and truncal fat percentage available in a sub-sample of UKBB (N = 118,160) (Supplementary Tables 5 and 6). Seven of the common novel variants were significantly associated (P<0.001, 0.05/48 variants examined) with both body-fat percentage and truncal fat percentage in the sexes-combined analysis (COBLL1, UHRF1BP1, WSCD2, Ccdc92, IFI30, MPV17L2, IZUMO1) and two with truncal fat percentage in women only (EFCAB12, GDF5). Only rs7607980 in COBLL1 is near a known body-fat percentage GWAS locus (rs6736827; R\(^2\) = 0.1989, distance, 6,751 base pairs, with our tag SNP)\(^{42}\). Of the nine SNPs associated with at least one of these two traits, all variants displayed much greater magnitude of effect on truncal fat percentage compared to body-fat percentage (Supplementary Fig. 20).

Previous studies have demonstrated the importance of examining common and LF/RRVs in genes with mutations known to cause monogenic diseases\(^{43,44}\). Thus, we assessed enrichment of WHRADjBMI variants in monogenic lipodystrophy and/or insulin resistance genes\(^{45,44}\) (Supplementary Data 13). No significant enrichment was observed, possibly due in part to the small number
**Box 1 | Genes of biological interest harboring WHR-associated variants**

**PLXND1** (3:129284818, rs2652973, known locus). The major allele of a common non-synonymous variant in Plexin D1 (L1412V, MAF 26.7%) is associated with increased WHRadjBMI ($\beta$ (s.e.m.) = 0.0156 (0.0024), $P = 9.16 \times 10^{-11}$). PLXND1 encodes a semaphorin class 3 and 4 receptor gene, and therefore, is involved in cell to cell signaling and regulation of growth and development for a number of different cell and tissue types, including those in the cardiovascular system, skeleton, kidneys and the central nervous system [72–81]. Mutations in this gene are associated with Moebius syndrome [82–85] and persistent truncus arteriosus [86,87]. PLXND1 is involved in angiogenesis as part of the SEMA and VEGF signaling pathways [88–90]. PLXND1 was implicated in the development of type 2 diabetes through its interaction with SEMA3E in mice. SEMA3E and PLXND1 are upregulated in adipose tissue in response to diet-induced obesity, creating a cascade of adipose inflammation, insulin resistance and diabetes mellitus [91]. PLXND1 is highly expressed in adipose (both subcutaneous and visceral) (GTEx). PLXND1 is highly intolerant of mutations and therefore highly conserved (Supplementary Data 16). Last, our lead variant is predicted as damaging or possibly damaging for all algorithms examined (SIFT, Polyphen2/HDIV, Polyphen2/HVAR, LRT, MutationTaster).

**ACVRL1** (2:158412701, rs55920843, novel locus). The major allele of a low-frequency non-synonymous variant in activin A receptor type 1C (rs55920843, N150H, MAF 1.1%) is associated with increased WHRadjBMI ($\beta$ (s.e.m.) = 0.0652 (0.0105), $P = 4.81 \times 10^{-10}$). ACVRL1, also called Activin receptor-like kinase 7 (ALK7), encodes a type I receptor for TGFβ (Transforming Growth Factor, Beta-1), and is integral for the activation of SMAD transcription factors; therefore, ACVRL1 plays an important role in cellular growth and differentiation [92–94], including adipocytes [95]. Mouse Acvrl1 decreases secretion of insulin and is involved in lipid storage [96,97]. ACVRL1 exhibits the highest expression in adipose tissue, but is also highly expressed in the brain (GTEx) [90–92]. Expression is associated with body fat, carbohydrate metabolism and lipids in both obese and lean individuals [92]. ACVRL1 is moderately tolerant of mutations (ExAC constraint scores: synonymous, −0.86; non-synonymous, 1.25; LoF, 0.04 and Supplementary Data 16). Last, our lead variant is predicted as damaging for two of five algorithms examined (LRT and MutationTaster).

**FGFR2** (10:123279643, rs13815382, novel locus). The minor allele of a rare synonymous variant in Fibroblast Growth Factor Receptor 2 (rs13815382, MAF 0.99%) is associated with increased WHRadjBMI ($\beta$ (s.e.m.) = 0.258 (0.049), $P = 1.38 \times 10^{-10}$). The extracellular portion of the FGFR2 protein binds with fibroblast growth factors, influencing mitogenesis and differentiation. Mutations in this gene have been associated with many rare monogenic disorders, including skeletal deformities, craniosynostosis, eye abnormalities and LADD syndrome, as well as several cancers including breast, lung and gastric cancer. Methylation of FGFR2 is associated with high birth weight percentile [95]. FGFR2 is tolerant of synonymous mutations, but highly intolerant of missense and loss-of-function mutations (ExAC constraint scores: synonymous, −0.9; missense, 2.74; LoF, 1.0 and Supplementary Data 16). Last, this variant is not predicted to be damaging on the basis of any of the five algorithms tested.

**ANGPTL4** (19:8429323, rs116843064, novel locus). The major allele of a non-synonymous low-frequency variant in Angiopoietin Like 4 (rs116843064, E40K, EAF 98.1%) is associated with increased WHRadjBMI ($\beta$ (s.e.m.) = 0.064 (0.011), $P = 1.20 \times 10^{-9}$). ANGPTL4 encodes a glycosylated, secreted protein containing a C-terminal fibrinogen domain. The encoded protein is induced by peroxisome proliferation activators and functions as a serum hormone that regulates glucose homeostasis, triglyceride metabolism [96,97] and insulin sensitivity [98]. Angptl4-deficient mice have hypoglycemia and increased lipoprotein lipase activity, while transgenic mice overexpressing Angptl4 in the liver have higher plasma triglyceride levels and decreased lipoprotein lipase activity [99]. The major allele of rs116843064 has been previously associated with increased risk of CHD and increased triglycerides [100]. ANGPTL4 is moderately tolerant of mutations (ExAC constraint scores: synonymous, 1.18; missense, 0.21; LoF:0.0 and Supplementary Data 16). Last, our lead variant is predicted for damaging for four of five algorithms (SIFT, Polyphen2/HDIV, Polyphen2/HVAR and MutationTaster).

**RREB1** (6:7211818, rs1334576, novel association signal). The major allele of a common non-synonymous variant in the Ras responsive element binding protein 1 (rs1334576, G195R, MAF = 56%) is associated with increased WHRadjBMI ($\beta$ (s.e.m.) = 0.017 (0.002), $P = 3.9 \times 10^{-10}$). This variant is independent of the previously reported GWAS signal in the RREB1 region (rs1294410; 6:66738752). The protein encoded by this gene is a zinc finger transcription factor that binds to RAS-responsive elements (RREs) of gene promoters. It has been shown that the calcitonin gene promoter contains an RRE and that the encoded protein binds there and increases expression of calcitonin, which may be involved in Ras/Raf-mediated cell differentiation [101–103]. The ras-responsive transcription factor RREB1 is a candidate gene for type 2 diabetes associated end-stage kidney disease [104]. This variant is highly intolerant to loss-of-function (ExAC constraint score LoF: 1, Supplementary Data 16).

**DAGLB** (7:6449496, rs2303361, novel locus). The minor allele of a common non-synonymous variant (rs2303361, Q664R, MAF 22%) in DAGLB (Diacylglycerol lipase beta) is associated with increased WHRadjBMI ($\beta$ (s.e.m.) = 0.0136 (0.0025), $P = 6.24 \times 10^{-14}$). DAGLB encodes a diacylglycerol (DAG) lipase that catalyzes the hydrolysis of DAG to 2-arachidonoyl-glycerol, the most abundant endocannabinoid in tissues. In the brain, DAGL activity is required for axonal growth during development and for retrograde synaptic signaling at mature synapses (2-AG) [105]. The DAGLB variant, rs702485 (7:64494272, $r^2 = 0.306$ and $D^2 = 1$ with rs2303361) has been previously associated with HDL. Pathway analyses indicate a role in the triglyceride lipase activity pathway [106]. DAGLB is tolerant of synonymous mutations, but intolerant of missense and loss-of-function mutations (ExAC constraint scores: synonymous, −0.76; missense, 1.07; LoF: 0.94 and Supplementary Data 16). Last, this variant is not predicted to be damaging by any of the algorithms tested.

**MLXIP** (7:73012042, rs35332062 and 7:73020337, rs3812316, known locus). The major alleles of two common non-synonymous variants (A358V, MAF = 12%; Q241H, MAF = 12%) in MLXIP (MLX interacting protein like) are associated with increased WHRadjBMI ($\beta$ (s.e.m.) = 0.02 (0.0033), $P = 1.78 \times 10^{-4}$; $\beta$ (s.e.m.) = 0.0213 (0.0034), $P = 1.98 \times 10^{-10}$). These variants are in strong linkage disequilibrium ($r^2 = 1.00, D^2 = 1.00, 1000 Genomes CEU$). This gene encodes a basic helix-loop-helix leucine zipper transcription factor of the Myc/Max/Mad superfamily. This protein forms a heterodimeric complex and binds and activates carbohydrate response element (ChoRE) motifs in the promoters of triglyceride synthesis genes in a glucose-dependent manner [74,75]. This gene is possibly involved in the growth hormone signaling pathway and lipid metabolism. The WHRadjBMI-associated
variant rs3812316 in this gene has been associated with the risk of non-alcoholic fatty liver disease and coronary artery disease. Furthermore, Williams–Beuren syndrome (an autosomal dominant disorder characterized by short stature, abnormal weight gain, various cardiovascular defects and mental retardation) is caused by a deletion of about 26 genes from the long arm of chromosome 7 including MLXIPL. MLXIPL is generally intolerant to variation, and therefore conserved (ExAC constraint scores: synonymous, 0.48, missense, 1.16, LoF, 0.68, Supplementary Data 16). Last, both variants reported here are predicted as possible or probably damaging by one of the algorithms tested (PolyPhen).

**RAPGEF3** (12:48143315, rs145878042, novel locus). The major allele of a low-frequency non-synonymous variant in Rap Guanine-Nucleotide-Exchange Factor (GEF) 3 (rs145878042, L300P, MAF = 1.1%) is associated with increased WHRadjBMI ($\beta$ (s.e.m) = 0.085 (0.010), $P = 7.15 \times 10^{-10}$). RAPGEF3 codes for an intracellular cAMP sensor, also known as Epac (the Exchange Protein directly Activated by Cyclic AMP). Among its many known functions, RAPGEF3 regulates the ATP sensitivity of the KATP channel involved in insulin secretion, may be important in regulating adipocyte differentiation, and plays an important role in regulating adiposity and energy balance. RAPGEF3 is tolerant of mutations (ExAC constraint scores: synonymous, −0.47; non-synonymous, 0.32; LoF, 0 and Supplementary Data 16). Last, our lead variant is predicted as damaging or possibly damaging for all five algorithms examined (SIFT, Polyphen2/HDIV, Polyphen2/HVAR, LRT, MutationTaster).

**Drosophila knockdown.** Considering the genetic evidence of adipose and insulin biology in determining body-fat distribution, and the lipid signature of the variants described herein, we examined whole-body triglyceride levels in adult *Drosophila*, a model organism in which the fat body is an organ functionally analogous to mammalian liver and adipose tissue as triglycerides are the major source of fat storage. Of the 51 genes harboring our 56 significantly associated variants, we identified 27 *Drosophila* orthologs for functional follow-up analyses. We selected genes with large changes in triglyceride levels (>20% increase or >40% decrease, as chance alone is not a probable cause for changes of this magnitude) from an existing large-scale screen with ≤2 replicates per knockdown strain. Two orthologs, for PLXN1 and DNAH10, met these criteria and were subjected to additional knockdown experiments with ≥5 replicates using tissue-specific drivers (fat body (cg-Gal4) and neuronal (elav-Gal4) specific RNAi-knockdowns) (Supplementary Table 9). A significant ($P < 0.025, 0.05/2$ orthologs) increase in the total body triglyceride levels was observed in DNAH10 ortholog knockdown strains for both the fat body and neuronal drivers. Only the neuronal driver knockdown for PLXN1 produced a significant change in triglyceride storage. DNAH10 and PLXN1 both lie in previous GWAS-identified regions. Adjacent genes have been highlighted as probable candidates for the DNAH10 association region, including *CCDC92* and *ZNMT64* based on expression quantitative trait locus (eQTL) evidence. Of note, rs110573535 in DNAH10 showed suggestive significance after conditioning on the known GWAS variants in nearby *CCDC92* (sex-combined $P_{\text{conditional}} = 7.56 \times 10^{-15}$; women-only rs110573535 $P_{\text{conditional}} = 5.86 \times 10^{-15}$, Supplementary Table 4) thus providing some evidence of multiple causal variants/gees underlying this signal. Further analyses are needed to determine whether the implicated coding variants from the current analysis are the putatively functional variants.

**eQTL lookups.** We examined the cis-association of variants with expression level of nearby genes in subcutaneous and visceral omental adipose, skeletal muscle and pancreas tissue from the Genotype-Tissue Expression (GTEx) project, and assessed whether exome and eQTL associations implicated the same signal (Methods and Supplementary Data 14–15). The lead exome variant of implicated genes and the relatively small number of variants in monogenic disease-causing genes (Supplementary Fig. 21).

**Genetic architecture of WHRadjBMI coding variants.** We used summary statistics from our stage 1 primary meta-analysis results to estimate the phenotypic variance explained by subsets of variants across various significance thresholds ($P < 2 \times 10^{-7}$ to 0.2) and conservatively using only independent SNPs (Supplementary Table 7, Methods and Supplementary Fig. 22). For only independent coding variants that reached suggestive significance in stage 1 ($P < 2 \times 10^{-7}$), 33 SNPs explain 0.38% of the variance. The 1,786 independent SNPs with a liberal threshold of $P < 0.02$ explain 13 times more variation (5.12%). However, these large SNPs with a liberal threshold (Supplementary Table 8 and Supplementary Fig. 23).
was associated with expression level of the gene itself for DAGLB, MLXIPL, CCD92, MAPKBP1, LRR3C6 and UQCC1. However, for MLXIPL, MAPKBP1 and LRR3C6, the lead variant is also associated with expression of nearby genes. At three additional loci, the lead exome variant is only associated with expression level of nearby genes (HEMK1 at C3orf18; NT5DC2, SMIM4 and TMEM110 at STAR1/ITH3 and C6orf106 at UHRF1BP1). Thus, although detected with a missense variant, these results are also consistent with a regulatory mechanism of effect, and the association signal may well be due to linkage disequilibrium with nearby regulatory variants.

Some of the coding genes implicated by eQTL analyses are known to be involved in adipocyte differentiation or insulin sensitivity: for example, for MLXIPL, the encoded carbohydrate responsive element binding protein is a transcription factor, regulating glucose-mediated induction of de novo lipogenesis in adipose tissue and expression of its beta isoform in adipose tissue is positively correlated with adipose insulin sensitivity. For CCD92, the reduced adipocyte lipid accumulation on knockdown confirmed the involvement of its encoded protein in adipose differentiation.

**Biological curation.** To investigate the possible functional role of the identified variants, we conducted thorough searches of the literature and publicly available bioinformatics databases (Supplementary Data 16–17, Box 1 and Methods). Many of our novel LF/RVs are in genes that are intolerant of non-synonymous mutations (for example, ACVR1C, DARS2, FGFR2, ExAC Constraint Scores >0.5). Other coding variants lie in genes that are involved in glucose homeostasis (for example, ACVR1C, UGGT2, ANGPTL4a), angiogenesis (RASIP1), adipogenesis (RAPGEF3) and lipid biology (ANGPTL4, DAGLB).

**Discussion**

Our analysis of coding variants from ExomeChip data in up to 476,546 individuals identified a total of 56 array-wide significant WHRadjBMI-associated variants in 41 independent association signals, including 24 newly identified (23 novel and one independent of known GWAS signals). Nine of these variants were low-frequency or rare, indicating an important role for such variants in the polygenic architecture of fat distribution. While, due to their rarity, these coding variants explain a small proportion of the trait variance at a population level, they may be more functionally tractable than non-coding variants and have a critical impact at the individual level. For instance, the association between a LF/RV (rs11209026; R381Q; MAF <5% in ExAC) located in the IL23R gene and multiple inflammatory diseases led to development of new therapies targeting IL23 and IL12 in the same pathway. Thus, we are encouraged that our LF/RV’s display large effect sizes: all but one of the nine novel LF/RV’s display larger effects than the 49 SNPs reported in Shungin et al., and some of these effects were up to seven-fold larger than those previously reported for GWAS. This finding mirrors results for other cardiometabolic traits, and suggests variants of possible clinical significance with even larger effect and rarer variants will probably be detected with greater sample sizes.

We continue to observe sexual dimorphism in the genetic architecture of WHRadjBMI. We identified 19 coding variants with significant sex differences, of which 16 (84%) display larger effects in women compared to men. Of the variants outside GWAS loci, we reported three (two LF/RVs) that show a significantly stronger effect in women and two (one LF/RV) that show a stronger effect in men. Genetic variants continue to explain a higher proportion of the phenotypic variation in body-fat distribution in women compared to men. Of the novel female (DSTYK and ANGPTL4a) and male (UGGT2 and MMP14) specific signals, only ANGPTL4 implicated fat distribution related biology associated with both lipid biology and cardiovascular traits (Box 1). Sexual dimorphism in fat distribution is apparent and at sexually dimorphic loci, hormones with different levels in men and women may interact with genomic and epigenomic factors to regulate gene activity, although this remains to be tested. Dissecting the underlying molecular mechanisms of the sexual dimorphism in body-fat distribution and how it is correlated with—and causing—important comorbidities such as cardiometabolic diseases will be crucial for improved understanding of disease pathogenesis.

Overall, we observe fewer significant associations, pathways and cross-trait associations between WHRadjBMI and coding variants on the ExomeChip than Turcott et al. for BMI. One reason for this may be smaller sample size (N_{WHRadjBMI} = 476,546, N_{BMI} = 718,639), and thus, lower statistical power. Power is probably not the only contributing factor, as trait architecture, heritability (possibly overestimated in some phenotypes) and phenotype precision all probably contribute to our study’s capacity to identify LF/RVs with large effects. Further, it is possible that the comparative lack of significant findings for WHRadjBMI may be a result of higher selective pressure against genetic predisposition to cardiometabolic phenotypes, thus rarer risk variants.

The ExomeChip is limited by the variants present on the chip, which was largely dictated by sequencing studies in European-ancestry populations and MAF detection criteria of ~0.012%. It is probable that through increased sample size, use of chips designed to detect variation across a range of continental ancestries, and high quality, deep imputation with large reference samples, future studies will detect additional variation from the entire allele frequency range that contributes to fat distribution.

The collected genetic and epidemiologic evidence has demonstrated that increased central adiposity is correlated with risk of type 2 diabetes and CVD, and that this association is probably causal with potential mediation through blood pressure, triglyceride-rich lipoproteins, glucose and insulin. This observation yields an immediate follow-up question: which mechanisms regulate depot-specific fat accumulation and are risks for disease driven by increased visceral and/or decreased subcutaneous adipose tissue mass? Pathway analysis identified several novel pathways and gene sets related to metabolism and adipose regulation, bone growth and development and adiponectin, a hormone that has been linked to ‘healthy’ expansion of adipose tissue and insulin sensitivity. Similarly, expression/eQTL results support the relevance of adipogenesis, adipocyte biology, and insulin signaling, supporting our previous findings for WHRadjBMI. We also provide evidence suggesting known biological functions and pathways contributing to body-fat distribution (for example, diet-induced obesity, angiogenesis, bone growth/morphology and lipid synthesis).

The ultimate aim of genetic investigations of obesity-related traits is to identify dysregulated genomic pathways leading to obesity pathogenesis that may result in a myriad of downstream illnesses. Thus, our findings may enhance the understanding of central obesity and identify new molecular targets to avert its negative health consequences. Significant cross-trait associations are consistent with expected direction of effect for several traits; that is, the WHR-increasing allele is associated with higher values of triglycerides, DBP, fasting insulin, total cholesterol, LDL and type 2 diabetes across many significant variants. However, it is worth noting that there are some exceptions. For example, rs9469913-A in UHRF1BP1 is associated with both increased WHRadjBMI and increased HDL. Also, we identified two variants in MLXIPL (rs3812316 and rs35332062), a well-known lipids-associated locus, in which the WHRadjBMI-increasing allele also increases all lipid levels, risk for hypertriglyceridemia, SBP and DBP. However, our findings show a significant and negative association with HbA1C, and nominally significant and negative associations with two-hour glucose, fasting glucose, and Type 2 diabetes, and potential negative associations with biomarkers for liver disease (for example, gamma glutamyl transpeptidase). Other notable exceptions include ITH3
(negatively associated with BMI, HbA1C, LDL and SBP), DAGLB (positively associated with HDL), and STAB1 (negatively associated with total cholesterol, LDL and SBP). Therefore, caution in selecting pathways for therapeutic targets is warranted; we must look beyond the effects on central adiposity to the potential cascading effects of related diseases.

A major finding from this study is the importance of lipid metabolism for body-fat distribution. In fact, pathway analyses that highlight enhanced lipolysis, cross-trait associations with circulating lipid levels, existing biological evidence from the literature, and knockdown experiments in Drosophila, point to novel candidate genes (ANGPTL4, ACVR1C, DAGLB, MGA, RASIP1 and IZUMO1) and new candidates in known regions (DNAIN10 and MLXIPL) related to lipid biology and their role in fat storage. ACVR1C, MLXIPL and ANGPTL4, all of which are involved in lipid homeostasis, are all excellent candidate genes for central adiposity. Carriers of inactivating mutations in ANGPTL4 (Angiopoietin Like 4), for example, display low triglycerides and low risk of coronary artery disease. ACVR1C encodes the activin receptor-like kinase 7 protein (ALK7), a receptor for the transcription factor TGFβ-1, well-known for its central role in general growth and development and adipocyte development in particular. ACVR1C exhibits the highest expression in adipose tissue, but is also highly expressed in the brain. In mice, decreased activity of ACVR1C upregulates PPARγ and C/EBPα pathways and increases lipolysis in adipocytes, thus decreasing weight and diabetes. Such activity suggests a role for ALK7 in adipose tissue signaling and a possible therapeutic target. MLXIPL, also important for lipid metabolism and postnatal cellular growth, encodes a transcription factor that is highly conserved, most probably damaging and associated with reduced MLXIPL expression in adipose tissue. Furthermore, in a recent longitudinal, in vitro transcriptome analysis of adipogenesis in human adipose-derived stromal cells, gene expression of MLXIPL was upregulated during the maturation of adipocytes, suggesting a critical role in the regulation of adipocyte size and accumulation. However, given our cross-trait associations with variants in MLXIPL and diabetes-related traits, development of therapeutic targets must be approached cautiously.

Our 24 novel variants for WHRadjBMI highlight the importance of lipid metabolism in the genetic underpinnings of body-fat distribution. We continue to demonstrate the critical role of adipocyte biology and insulin resistance for central obesity and offer support for potentially causal genes underlying previously identified fat distribution loci. Notably, our findings offer potential new therapeutic targets for intervention in the risks associated with abdominal fat accumulation and represents a major advance in our understanding of the underlying biology and genetic architecture of central adiposity.

**References**

1. Pschon, T. et al. General and abdominal adiposity and risk of death in Europe. *N. Engl. J. Med.* 359, 2105–2120 (2008).
2. Wang, Y., Rimm, E. B., Stampfer, M. J., Willett, W. C. & Hu, F. B. Comparison of abdominal adiposity and overall obesity in predicting risk of type 2 diabetes among men. *Am. J. Clin. Nutr.* 81, 555–563 (2005).
3. Canoy, D. Distribution of body fat and risk of coronary heart disease in men and women. *Carr. Opin. Cardiol.* 23, 591–598 (2008).
4. Snijder, M. B. et al. Associations of hip and thigh circumferences independent of waist circumference with the incidence of type 2 diabetes: the Hoorn Study. *Am. J. Clin. Nutr.* 77, 1192–1197 (2003).
5. Yusuf, S. et al. Obesity and the risk of myocardial infarction in 27,000 participants from 52 countries: a case-control study. *Lancet* 366, 1640–1649 (2005).
6. Mason, C., Craig, C. L. & Katzmarzyk, P. T. Influence of central and extremity circumferences on all-cause mortality in men and women. *Obesity* 16, 2690–2695 (2008).
7. Karpe, F. & Pinnick, K. E. Biology of upper-body and lower-body adipose tissue—link to whole-body phenotypes. *Nat. Rev. Endocrinol.* 11, 90–100 (2015).
8. Manolopoulos, K. N., Karpe, F. & Frayn, K. N. Glutamchomoronal body fat as a determinant of metabolic health. *Int. J. Obes.* 34, 949–959 (2010).
9. Emdin, C. A. et al. Genetic association of waist-to-hip ratio with cardio-metabolic traits, type 2 diabetes, and coronary heart disease. *JAMA* 317, 626–634 (2017).
10. Shungin, D. et al. New genetic loci link adipose and insulin biology to body fat distribution. *Nature* 518, 187–196 (2015).
11. Winkler, T. W. et al. 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 (2015).
12. Wen, W. et al. Genome-wide association studies in east asians identify new loci for waist-hip ratio and waist circumference. *Sci. Rep.* 6, 17958 (2016).
13. Gao, C. et al. A comprehensive analysis of common and rare variants to identify adiposity loci in hispanic americans: the iras family study (IRASFS). *PLoS ONE* 10, e0134649 (2015).
14. Graff, M. et al. Genome-wide physical activity interactions in adiposity—meta-analysis of 200,452 adults. *PLoS Genet.* 13, e1006528 (2017).
15. Justice, A. E. et al. Genome-wide meta-analysis of 241,258 adults accounting for smoking behaviour identifies novel loci for obesity traits. *Nat. Commun.* 8, 14977 (2017).
16. Ng, M. C. Y. et al. 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 (2017).
17. Aschard, H., Viljianmalsson, B. J., Joshi, A. D., Price, A. L. & Kraft, P. Adjusting for heritable covariates can bias effect estimates in genome-wide association studies. *Am. J. Hum. Genet.* 96, 329–339 (2015).
18. Day, F. R., Loh, P. R., Scott, R. A., Ong, K. K. & Perry, J. R. A robust example of collider bias in a genetic association study. *Am. J. Hum. Genet.* 98, 392–393 (2016).
19. Feng, S., Liu, D., Zhan, X., Wing, M. K. & Abecasis, G. R. RAREMETAL: fast and powerful meta-analysis for rare variants. *Bioinformatics* 30, 2828–2828 (2014).
20. Pers, T. H. et al. Biological interpretation of genome-wide association studies using predicted gene functions. *Nat. Commun.* 6, 5890 (2015).
21. Marouli, E. et al. Rare and low-frequency coding variants alter adult human adult height. *Nature* 542, 186–190 (2017).
22. Lamparter, D., Marbach, D., Rueedi, R., Katalik, Z. & Bergmann, S. Fast and rigorous computation of gene and pathway scores from SNP-based summary statistics. *PLoS Comput. Biol.* 12, e1004714 (2016).
23. Locke, A. E. et al. Genetic studies of body mass index yield new insights for obesity biology. *Nature* 518, 197–206 (2015).
24. Kawai, M., de Paula, F. J. & Rosen, C. J. New insights into osteoporosis: the bone-fat connection. *J. Intern. Med.* 272, 317–329 (2012).
25. Turcot, V. et al. Proximal arterial disease and obesity. *Circ. Cardiovasc. Genet.* 3, 272829 (2010).
26.滿, D. J. et al. Exome-wide association study of plasma lipids in >300,000 individuals. *Nat. Genet.* 49, 1756–1766 (2017).
27. Kraja, A. T. et al. New blood pressure-associated loci identified in meta-analyses of 475,000 individuals. *Circ. Cardiovasc. Genet.* 10, e004676 (2015).
28. Mahajan, A. et al. Identification and functional characterization of G6PC2 coding variants influencing glycomic traits define an effector transcript at the G6PC2-ABC1B1 locus. *PLoS Genet.* 11, e1004676 (2015).
29. Manning, A. et al. A low-frequency inactivating AKT2 variant enriched in the finnish population is associated with fasting insulin levels and type 2 diabetes risk. *Diabetes* 66, 2019–2032 (2017).
Acknowledgements

This work was primarily supported through funding from the National Institute of Health (NIH): 1K99HL130580, R01-DK089256, 2R01HD057194, U10HG007416, R01DK1101855, T32 HL070655, KL2TR001109; and the American Heart Association (AHA). S.S. and P.M. received European Molecular Biology Organization (EMBO) Starting Investigator grants (STIF1650011 and 161669017). Co-author Y. Jia recently passed away while this work was in process. This study was completed as part of the Genomic Investigation of Antriglandar Traits (GANT) Consortium. This research has been conducted using the UK Biobank resource. A full list of acknowledgements is provided in the Supplementary Data 18.

Author contributions

Writing Group: L.A.C., R.S.F., T.M.E., M. Graff, H.M.H., J.N.H., A.E.J., T.K., Z.K., C.M.L., R.J.F.L., Y.L., K.E.N., V.T., K.L.Y.

Data preparation group: T.A., I.B.B., T.E., S.F., M. Graff, H.M.H., A.E.J., T.K., D.L.I., K.S.L., A.E.J., R.F.L., Y.L., E. Maroudi, N.G.D.M., C.M., G.P., M.C.Y.N., M.A.R., S.S., C.R.F., T.P., V.T., S.V., M.T.W., T.W.W., K.C., X.Z.

WHR meta-analyses: P.L.A., H.M.H., A.E.J., T.K., M. Graff, C.M.L., R.J.F.L., K.E.N., V.T., K.L.Y.

Pleiotropy working group: M.A., G. Mohankee, J.P.C., F.D., J.C.F., H.M.H., H.K., H.M.H., A.E.J., C.M.L., D.L.I., R.F.L., A. Mahajan, E. Maroudi, G.M., M.I.M., P.B.M., G.M.P., J.R.B., K.S.R., X.S., S.W., J.N., C.W.

Phenome-wide association studies: L. Battarache, J.C.D., A.G. A. Mabaia, M.I.M.

Gen-set enrichment analyses: S.B., R.S.F., J.N.H., Z.K., D.L., T.H.P., T.E.V.

eQTL analyses: C.R., Y.L., K.L.M.

Monogenic and syndrome genetic enrichment analyses: H.M.H., A.K.M.

M.D. Obesity Screwel: A.L.C.

Overseeing of contributing studies and consortia: (1958 Birth Cohort) P.D.; (Airwave) P.E.; (AMC PAS) G.K.H.; (Amish) J.R.O.; (ARIC) E.B.; (Aric, Add Health) K.E.N.; (BRAVE) E.D.A., R.C.; (BRIGHT) P.B.M.; (CARDIA) M.I.M., P.B.M.; (Cebu Chronic Longitudinal and Nutrition Survey) K.L.M.; (CHD Exome + Consortium) A.S.B.; J.M.M.H., D.F.R., I.D.; (CHES) R.V.; (Clear/EMERGE (Seattle)) G.P.J.; (CROATIA, Korcula) V.V., O. Polasek, L.(deCODE) K. Stefansson, U.T., (DHS) D.W.B.; (DIACORE) C.A.B.; (DPS) I.T., J. Lindström, M.U.; (DSEXTA) T.A.L., H.R.; (E3COACH) A.T.H., T.M.E.; (EGCUT) T.E.; (eMERGE (Seattle)) E.B.L.; (EPIC-Potsdam) M.B.S., H.B.; (EpiHealth) E.L., P.W.F. (EXTEND) A.T.H., T.M.E.; (Family Heart Study) I.B., M.R.; (Framingham Heart Study) I.B., M.R.; (FVG) P.G.; (Generation Scotland) M.P.D.; (Genetic Epidemiology Network of Arteriovasculopathy (GENOA)) S.L.R.K.; (GRAPHIC) N.J.S.; (GSK-STATIM) D.M.W, L.W., H.D.W.; (Health) A. Linneberg; (HELIC) M.H.H.; (HOLY) E.Z., G.D.; (HOLIC) P.M., E.Z., G.D.; (HUNT) K.E.N., C.W.; (I99) T.H., J.I.; (IRASPs) E.L., K.S. (Jackson Heart Study (JHS)) J.G.W. (KORA S4) K. Strauch, M.R.; J.H., M. Blumberg, P.K., P.K.; (LOLipop-Protein Exome) J.C.C.; J.K.S.; (LOLipop-Omni) J.C.C.; J.K.S.; (MARA) I.I.R., X.G.; (METSIM) J.K.; (MONICA-Brienz) G.C.; (Montreal Heart Institute Biobank (MHHB)) M.P.D., G.L.; (Morgantown) D.M.W.; (Morgan) K.E.N.; (Morgan) A.K.M.; (Morgan) S.B., R.S.F., J.N.H., Z.K., D.L.; (Morgan) T.D.S.; (Morgan) S.H.R.; (Morgan) W.S.W.; (Morgan) Y.F.S.; (Morgan) D.M.W.; (Morgan) D.M.W.; (Morgan) D.M.W.

Global Lipids Genetics, C. et al. Discovery and refinement of loci associated with lipid levels. Nat. Genet. 45, 1274–1283 (2013).

Kooner, J. S. et al. Genome-wide scan identifies variation in MLXIPL 3q13.1 that is associated with plasma triglycerides. Nat. Genet. 28, 149–151 (2008).

Kwon, J. et al. Forced expression of the cAMP sensor Epac as a determinant of KATP channel ATP sensitivity in human pancreatic beta-cells and rat INS-1 cells. J. Physiol. 586, 1307–1319 (2008).

Ji, Z., Mei, F. C. & Cheng, X. Epac, not PKA catalytic subunit, is required for 3T3-L1 preadipocyte differentiation. Front Biosci. 2, 392–398 (2010).

Martini, C. N., Plaza, M. V. & Vila Mdel, C. PKA-dependent and independent signaling in 3T3-L1 fibroblast differentiation. Mol. Cell. Endocrinol. 298, 42–49 (2007).

Petersen, T. & et al. Cyclic AMP (cAMP)-mediated stimulation of adipocyte differentiation requires the synergistic action of Epac and cAMP-dependent protein kinase-dependent processes. Mol. Cell. Biol. 28, 3834–3843 (2008).

Yan, J., et al. Enhanced leptin sensitivity, reduced adiposity, and improved glucose homeostasis in mice lacking exchange protein directly activated by cyclic AMP isoform I. Mol. Cell. Biol. 33, 918–926 (2013).

Gesta, S. et al. Evidence for a role of developmental genes in the origin of obesity and body fat distribution. Proc. Natl Acad. Sci. USA 103, 6657–6661 (2006).

Gesta, S. et al. Mesodermal development gene Tbx15 impairs adipocyte differentiation and mitochondrial respiration. Proc. Natl Acad. Sci. USA 108, 2771–2776 (2011).

Lee, K. Y. et al. Tbx15 controls skeletal muscle fibre-type determination and muscle metabolism. Nat. Commun. 6, 8054 (2015).
Germany. 50 The Novo Nordisk Foundation Center for Basic Metabolic Research, Faculty of Health and Medical Sciences, University of Copenhagen, Copenhagen, Denmark. 50 Department of Biochemistry, Wake Forest School of Medicine, Winston-Salem, NC, USA. 50 Department of Clinical Biochemistry, Lillebaelt Hospital, Vejle, Denmark. 50 Institute of Regional Health Research, University of Southern Denmark, Odense, Denmark. 50 Department of Medicine (Medical Genetics), University of Washington, Seattle, WA, USA. 50 MRC/BHF Cardiovascular Epidemiology Unit, Department of Public Health and Primary Care, University of Cambridge, Cambridge, UK. 50 NIHR Blood and Transplant Research Unit in Donor Health and Genomics, Department of Public Health and Primary Care, University of Cambridge, Cambridge, UK. 50 NIHR Barts Cardiovascular Research Centre, Barts and The London School of Medicine & Dentistry, Queen Mary University of London, London, UK. 50 Research Centre on Public Health, University of Milano-Bicocca, Monza, Italy. 50 Lee Kong Chian School of Medicine, Nanyang Technological University, Singapore, Singapore. 50 Department of Cardiology, London North West Healthcare NHS Trust, Ealing Hospital, Middlesex, UK. 50 Department of Epidemiology and Biostatistics, School of Public Health, Imperial College London, London, UK. 50 Imperial College Healthcare NHS Trust, London, UK. 50 MRC-PHE Centre for Environment and Health, Imperial College London, London, UK. 50 Division of Genetics, Brigham and Women's Hospital and Harvard Medical School, Boston, MA, USA. 50 Division of Preventive Medicine, Brigham and Women's and Harvard Medical School, Boston, MA, USA. 50 Medical department, Lillebaelt Hospital, Vejle, Denmark. 50 Medical Genomics and Metabolic Genetics Branch, National Human Genome Research Institute, National Institutes of Health, Bethesda, MD, USA. 50 Department of Biostatistics, University of Liverpool, Liverpool, UK. 50 Menzies Health Institute Queensland, Griffith University, Southport, Queensland, Australia. 50 Department of Biomedical Informatics and Medical Education, University of Washington, Seattle, WA, USA. 50 Wellcome Trust Sanger Institute, Hinxton, UK. 50 British Heart Foundation Cambridge Centre of Excellence, Department of Medicine, University of Cambridge, Cambridge, UK. 50 Julius Center for Health Sciences and Primary Care, University Medical Center Utrecht, Utrecht, The Netherlands. 50 Department of Genetics, Center for Molecular Medicine, University Medical Center Utrecht, Utrecht, The Netherlands. 50 Faculty of Pharmacy, Universite de Montreal, Montreal, Quebec, Canada. 50 Department of Clinical Epidemiology, Leiden University Medical Center, Leiden, The Netherlands. 50 Department of Nutrition and Dietetics, School of Health Science and Education, Harokopio University, Athens, Greece. 50 Division of Epidemiology & Community Health, School of Public Health, University of Minnesota, Minneapolis, MN, USA. 50 Centre for Cancer Genetic Epidemiology, Department of Public Health and Primary Care, University of Cambridge, Cambridge, UK. 50 Department of Internal Medicine B, University Medicine Greifswald, Greifswald, Germany. 50 DZHUK (German Centre for Cardiovascular Research), partner site Greifswald, Greifswald, Germany. 50 Institute of Cardiovascular Science, University College London, London, UK. 50 MRC Integrative Epidemiology Unit, School of Social and Community Medicine, University of Bristol, Bristol, UK. 50 Department of Life Sciences, Brunel University London, Uxbridge, UK. 50 Department of Medicine, Faculty of Medicine, Universite de Montreal, Montreal, Quebec, Canada. 50 Centre for Cancer Genetic Epidemiology, Department of Oncology, University of Cambridge, Cambridge, UK. 50 Department of Epidemiology and Biostatistics, MRC-PHE Centre for Environment and Health, School of Public Health, Imperial College London, London, UK. 50 Department of Hygiene and Epidemiology, University of Ioannina Medical School, Ioannina, Greece. 50 CNR Institute of Clinical Physiology, Pisa, Italy. 50 Department of Clinical & Experimental Medicine, University of Pisa, Pisa, Italy. 50 Toulouse University School of Medicine, Toulouse, France. 50 Institute of Molecular Medicine, The University of Texas Health Science Center at Houston, Houston, TX, USA. 50 Department of Clinical Sciences, Genetic and Molecular Epidemiology Unit, Lund University, Malmo, Sweden. 50 Department of Nutrition, Harvard School of Public Health, Boston, MA, USA. 50 Department of Public Health and Clinical Medicine, Unit of Medicine, Umeå University, Umeå, Sweden. 50 Institute of Clinical Chemistry and Laboratory Medicine, University Medicine Greifswald, Greifswald, Germany. 50 Iliara Gandin, Research Unit, AREA Science Park, Trieste, Italy. 50 Department of Medical Sciences, University of Trieste, Trieste, Italy. 50 Institute for Maternal and Child Health-IRCCS Burlo Garofolo, Trieste, Italy. 50 Geriatrics, Department of Public Health, Uppsala University, Uppsala, Sweden. 50 German Center for Diabetes Research, München-Neuherberg, Germany. 50 Institute of Epidemiology II, Helmholtz Zentrum München—German Research Center for Environmental Health, Neuhberg, Germany. 50 Research Unit of Molecular Epidemiology, Helmholtz Zentrum München—German Research Center for Environmental Health, Neuhberg, Germany. 50 Department of Medical Sciences, Molecular Epidemiology and Science for Life Laboratory, Uppsala University, Uppsala, Sweden. 50 Division of Public Health Sciences, Fred Hutchinson Cancer Research Center, Seattle WA, USA. 50 University of Exeter Medical School, University of Exeter, Exeter, UK. 50 MRC Human Genetics Unit, Institute of Genetics and Molecular Medicine, University of Edinburgh, Edinburgh, UK. 50 Institute of Genetic Epidemiology, Helmholtz Zentrum München—German Research Center for Environmental Health, Neuhberg, Germany. 50 K.G. Jebsen Center for Genetic Epidemiology, Department of Public Health, NTNU, Norwegian University of Science and Technology, Trondheim, Norway. 50 Department of Vascular Medicine, AMC, Amsterdam, The Netherlands. 50 CAS Key Laboratory of Nutrition, Metabolism and Food Safety, Shanghai Institute of Nutrition and Health, Shanghai Institutes for Biological Sciences, University of Chinese Academy of Sciences, Chinese Academy of Sciences, Shanghai, China. 50 Division of Endocrinology and Metabolism, Department of Internal Medicine, Tri-Service General Hospital Songshan Branch, Taipei, Taiwan. 50 School of Medicine, National Defense Medical Center, Taipei, Taiwan. 50 HUNT Research Center, Department of Public Health, Norwegian University of Science and Technology, Levanger, Norway. 50 Department of Neurology, Erasmus Medical Center, Rotterdam, The Netherlands. 50 Department of Radiology, Erasmus Medical Center, Rotterdam, The Netherlands. 50 Stanford Cardiovascular Institute, Stanford University, Stanford, CA, USA. 50 Department of Genome Sciences, University of Washington, Seattle, WA, USA. 50 Faculty of Medicine, Aalborg University, Aalborg, Denmark. 50 Department of Public Health, Faculty of Health and Medical Sciences, University of Copenhagen, Copenhagen, Denmark. 50 Research Center for Prevention and Health, Capital Region of Denmark, Glostrup, Denmark. 50 National Institute for Health and Welfare, Helsinki, Finland. 50 Department of Computational Medicine and Bioinformatics, University of Michigan, Ann Arbor, MI, USA. 50 Department of Internal Medicine, University of Michigan, Ann Arbor, MI, USA. 50 Division of Gastroenterology, University of Michigan, Ann Arbor, MI, USA. 50 Centre for Brain Research, Indian Institute of Science, Bangalore, India. 50 Echininos Medical Centre, Echinios, Greece. 50 Oxford Centre for Diabetes, Endocrinology and Metabolism, Radcliffe Department of Medicine, University of Oxford, Oxford, UK. 50 Oxford NIHR Biomedical Research Centre, Oxford University Hospitals Trust, Oxford, UK. 50 UKCRC Centre of Excellence for Public Health Research, Queens University Belfast, Belfast, UK. 50 Foundation for Research in Health Exercise and Nutrition, Kuopio Research Institute of Exercise Medicine, Kuopio, Finland. 50 National Heart and Lung Institute, Imperial College London, Hammersmith Hospital Campus, London, UK. 50 University Medical Centre Mannheim, 5th Medical Department, University of Heidelberg, Mannheim, Germany. 50 Institute of Clinical Medicine, Internal Medicine, University of Eastern Finland and Kuopio University, Kuopio, Finland. 50 Institute of Biomedicine, School of Medicine, University of Eastern Finland, Kuopio Campus, Finland. 50 Department of Clinical Physiology and Nuclear Medicine, Kuopio University Hospital, Kuopio, Finland. 50 Verge Genomics, San Francisco, CA, USA. 50 Division of Biomedical and Personalized Medicine, Department of Medicine, University of Colorado-Denver, Aurora, CO, USA. 50 Kaiser Permanente Washington Health Research Institute, Seattle, WA, USA. 50 Department of Health Sciences, University of Washington, Seattle, WA, USA. 50 Department of Anthropology, Sociology, and History, University of San Carlos, Cebu City, Philippines. 50 USC-Office of Population Studies Foundation Inc., University of San Carlos, Cebu City, Philippines. 50 Department of Medical Research, Taichung Veterans General Hospital, Taichung, Taiwan. 50 Department of Social Work, Tunghai University, Taichung, Taiwan. 50 Department of Clinical Chemistry, Fimlab Laboratories, Tampere, Finland. 50 Department of Clinical Chemistry, Finnish Cardiovascular Research Center—Tampere, Faculty of Medicine and Life Sciences, University of Tampere, Tampere, Finland. 50 Division of Preventive Medicine, University of Alabama at Birmingham, Birmingham, AL, USA. 50 Department of Medicine, Division of Cardiovascular Medicine, Stanford University School of Medicine, Palo Alto, CA, USA. 50 Uppsala University, Uppsala, Sweden. 50 Center for Clinical Research and Prevention, Bispebjerg and Frederiksberg Hospital, Frederiksberg, Denmark. 50 Department of Clinical Medicine, Faculty of Health and Medical Sciences, University of Copenhagen, Copenhagen, Denmark. 50 Department of
Methods

Studies. Stage 1 included 74 studies (12 case/control, 59 population-based and five family) comprising 344,369 adults of European (N = 288,492), African (N = 13,687), South Asian (N = 29,315), East Asian (N = 6,800) and Hispanic (N = 4,064) descent. Stage 1 meta-analyses were conducted in each ancestry separately and in all ancestries together, for both sex-combined and sex-specific analyses. Follow-up analyses were performed in 132,177 individuals of European ancestry from deCODE and the UK Biobank. Release 1.0 (UKBB) (Supplementary Data 1-3). Informed consent was obtained by the parent study and protocols approved by each study’s institutional review boards.

Phenotypes. For each study, WHR (waist circumference divided by hip circumference) was corrected for age, BMI and genomic principal components (derived from GWAS data, the variants with MAF > 1% on the ExomeChip, and ancestry informative markers available on the ExomeChip), as well as any additional study-specific covariates (for example, recruiting center), in a linear regression model. For studies with unrelated individuals, residuals were calculated separately by sex, whereas for family based studies sex was included as a covariate in models with both men and women. Residuals for case/control studies were calculated separately. Finally, residuals were inverse normal transformed and used as the outcome in association analyses. Phenotype descriptive by study are shown in Supplementary Data 3.

Genotypes and quality control. Most studies followed a standardized protocol and performed genotype calling using the algorithms indicated in Supplementary Data 2, which typically included Illumina. For 10 studies from the Cohorts for Heart and Aging Research in Genomic Epidemiology (CHARGE) Consortium, raw intensity data for samples from seven genotyping centers were combined for joint calling1. Study-specific quality control of the genotyped variants was implemented before association analysis (Supplementary Data 1.2). To assess whether any significant associations with rare and low-frequency variants could be due to allele calling in smaller studies, we performed a general analysis of all studies (>5,000 participants) compared to all studies. We found very high concordance for effect sizes, suggesting that smaller studies do not bias our results (Supplementary Fig. 24).

Study-level statistical analyses. Each cohort performed single-variant analyses for both additive and recessive models in each ancestry, for sexes combined and sex-specific groups, with either RAREMETALWORKER (see URLs) or RVTETests (see URLs), to associate inverse normal transformed WHRadjBMI with genotype accounting for cryptic relatedness (kinship matrix) in a linear mixed model. Both programs perform score- statistic rare-variant association analyses, accommodate unrelated and related individuals and provide single-variant results and variance- covariance matrices. The covariance matrix captures linkage disequilibrium between markers within 1 Mb, which is used for gene-level meta-analyses and conditional analyses10,11,15.

Centralized quality control. Individual cohorts identified ancestry outliers based on 1000 Genomes Phase 1 reference populations. A centralized quality control procedure implemented in EasyQC16 was applied to individual cohort summary statistics to identify cohort-specific problems: (1) possible errors in phenotype residual transformation; (2) strand issues and (3) inflation due to population stratification, cryptic relatedness and genotype biases.

Meta-analyses. Meta-analyses were carried out in parallel by two analysts at two sites using RAREMETAL11. We excluded variants if they had call rate <95%, Hardy–Weinberg equilibrium \( P < 1 \times 10^{-6} \), or large allele frequency deviations from reference populations (>0.6 for all ancestries analyses and >0.3 for ancestry-specific analyses). We also excluded markers not present on the Illumina ExomeChip array 1.0 Y-chromosome and mitochondrial variants, indels, multiallelic markers and problematic variants on the basis of Blat-based sequence alignment. Significance for single-variant analyses was defined at an array-wide significance threshold \( P < 1 \times 10^{-6} \) in two independent studies, UKBB (N = 119,572) and deCODE (N = 12,605), using the same quality control and analytical methodology. Genotyping, study descriptions and phenotype descriptive statistics are provided in Supplementary Data 1-3. Stage 1 and 2 meta-analyses was performed using the inverse-variance weighted fixed-effects method. Significant associations were defined as those nominally significant \( P < 0.05 \) in stage 2 when available in stage 2, and array-wide significance for stage 1 and 2 at \( P < 2 \times 10^{-7} \) (0.05/250,000 246,328 variant tested), Varians are considered novel and statistically significant if they were greater than 1 Mb from a previously identified WHRadjBMI lead SNP and achieved a significant threshold of \( P < 2 \times 10^{-6} \).

Pathway enrichment analyses: EC-DEPICT. We adapted DEPICT, a gene set enrichment analysis method for GWAS data, for use with the ExomeChip (EC-DEPICT) described further in a companion manuscript17. DEPICT uses ‘reconstituted’ gene sets, where different types of gene set (for example, canonical pathways, protein-protein interaction networks, and mouse phenotypes) were extended through large-scale microarray data (see Pers et al.18 for details). EC-DEPICT computes P values based on Swedish ExomeChip data (Malmö Diet and Cancer (MDC), All New Diabetics in Scania (ANDIS) and Scania Diabetes Registry (SDR) cohorts, N = 11,899) and, unlike DEPICT, takes as input only genes directly containing substantial (coding) variants rather than all genes in a specified linkage disequilibrium (Supplementary Note). Two analyses were performed for WHRadjBMI ExomeChip: one with all variants \( P < 5 \times 10^{-6} \) (49 significant gene sets in 25 meta-gene sets, FDR < 0.05) and one with all variants >1 Mb from known GWAS loci (26 significant gene sets in 13 meta-gene sets, FDR < 0.05). Affinity propagation clustering19 was used to group highly relevant gene sets into ‘meta-gene sets’, for each meta-gene set, the model gene set with the best P value was used for visualization (Supplementary Note). EC-DEPICT was written in Python (see URLs).

Pathway enrichment analyses: PASCAL. We also applied PASCAL pathway analysis20 to summary statistics from stage 1 for all coding variants. PASCAL downweights gene-based scores (SUM and MAX) and tests for over-representation of high gene scores in predefined biological pathways. We performed both MAX and SUM estimations for pathway enrichment. MAX is sensitive to gene sets driven by a single signal, while SUM is better for multiple variant associations in the same gene. We used standard pathway libraries from KEGG, REACTOME and BIOCARTA, and also added dichotomized (z > 3) reconstituted gene sets.
from DEPICT\cite{20}. To accurately estimate SNP-by-SNP correlations even for rare variants, we used the UK10k data (TwinUK\cite{21} and ALSpac\cite{22}; \(N=3781\)). To distinguish contributions of regulatory and coding variants, we also applied PASCAL to summary statistics of only regulatory variants (20 kb upstream) and regulatory+coding variants from the Shungin et al.\cite{23} study. In this way, we could investigate what is gained by analyzing coding variants.

**Monogenic obesity enrichment analyses.** We compiled two lists consisting of 31 genes with strong evidence that disruption causes monogenic forms of insulin resistance or diabetes, and eight genes with evidence that disruption causes monogenic forms of lipodystrophy. To test for association enrichment, we conducted simulations by matching each gene with others based on gene length and number of variants tested to create 1,000 matched gene sets and assessed how often the number of variants exceeding set significance thresholds was greater than in our monogenic obesity gene set.

**Variance explained.** We estimated phenotypic variance explained by stage 1 associations in all ancestries for men, women, and combined sexes\cite{24}. For each associated region, we pruned subsets of SNPs within 500 kb of SNPs with the lowest \(P\) value and used varying \(P\) thresholds ranging from \(2\times10^{-10}\) to 0.02) from the combined sexes results. Additionally, we examined all variants and independent variants across a range of MAFs. The variance explained by each subset of SNPs in each stratum was estimated by summing the variance explained by individual top coding variants. To compare variance explained between men and women, we tested for significant differences assuming the weighted sum of \(\chi^2\) distributed variables tended to a Gaussian distribution following Lyapunov’s central limit theorem\cite{25,26}.

**Cross-trait lookups.** To evaluate relationships between WHRadjBMI and related cardiometabolic, anthropometric, and reproductive traits, association results for the 51 WHRadjBMI coding SNPs were requested from seven consortia, including ExomeChip data from GIANT (BMI, height), Global Lipids Genetics Consortium (GLGC) (total cholesterol, triglycerides, HDL-cholesterol, LDL-cholesterol), International Consortium for Blood Pressure (IPCG)\cite{27} (systolic and diastolic blood pressure), Meta-Analyses of Glucose and Insulin-related traits Consortium (MAGIC) (glycemic traits) and DIABETES Genetics Replication And Meta-analysis (DIAGRAM) consortium (type 2 diabetes)\cite{28,29}. For coronary artery disease, we accessed 1000 Genomes Project-imputed GWAS data released by CARDIoGRAMplusC4D\cite{30} and for age at menarche and menopause, we used a combination of ExomeChip and 1000 Genomes Project-imputed GWAS data from ReproGen. Heat maps were generated with gplots (\(R\) v.3.3.2) using Euclidean distance based on \(P\) value, and direction of effect and complete linkage clustering (see URLs).

**GWAS catalog lookups.** To determine whether substantial coding variants were associated with any related cardiometabolic or anthropometric traits, we also searched the NHGRI EBI GWAS Catalog for previous associations near our lead SNPs (\(+/-500\) kb). We used PLINK to calculate linkage disequilibrium using \(r^2\). We searched the NHGRI EBI GWAS Catalog for previous associations near our lead SNPs (\(+/-500\) kb). We used PLINK to calculate linkage disequilibrium using \(r^2\). We searched the NHGRI EBI GWAS Catalog for previous associations near our lead SNPs (\(+/-500\) kb). We used PLINK to calculate linkage disequilibrium using \(r^2\).

**eQTLs analysis.** We queried the significant variant (Exome coding SNPs)-gene pairs associated with eGenes across five metabolically relevant tissues (skeletal muscle, subcutaneous adipose, visceral adipose, liver and pancreas) with at least 70 samples in the GTEx database\cite{31}. For each tissue, variants were selected based on the following thresholds: the minor allele was observed in at least 10 samples and MAF \(\geq 1\%\), eGenes, genes with a significant eQTL, are defined on a FDR130 threshold of \(<0.05\). For each eGene, we also listed the most significantly associated variants (eSNPs). Only these exome SNPs with \(r^2 > 0.8\) with eSNPs were considered for biological interpretation (see supplementary eQTL GTEx).

We also performed cis-eQTL analysis in 770 METSIM subcutaneous adipose tissue samples as described in Civelek et al.\cite{32}. A FDR was calculated using all \(P\) values from the cis-eQTL detection in the \(q\)-value package in \(R\). Variables associated with nearby genes at an FDR of less than 1% were considered to be significant (equivalent \(P < 2.46 \times 10^{-5}\)).

For loci with more than one microarray probe set of the same gene associated with the eGene variant, we selected the probe set that provided the strongest linkage disequilibrium \(r^2\) between the eGene variant and the eSNP.

Penetrance analysis. Phenotype and genotype data from UKBB were used for penetrance analysis. Three of 16 rare and low-frequency variants (MAF \(\leq 1\%\)) detected in the final stage 1 and 2 meta-analysis were available in UKBB and had relatively larger effect sizes (\(>0.90\)). Phenotype data for these three variants were stratified by WHR using World Health Organization (WHO) guidelines, which consider women and men with WHR greater than 0.85 and 0.90 as obese, respectively. Genotype and allele counts were used to calculate the number of carriers of the minor allele. The association between obesity versus non-obese carriers for women, men and sexes combined was compared using a \(y^2\) test. Significance was determined using a Bonferroni correction for the number of tests performed (0.05/9 = 5.5 \times 10^{-5}\).

**Reporting Summary.** Further information on experimental design is available in the Nature Research Reporting Summary linked to this article.

**Data availability.** Summary statistics of all analyses are available at https://portals.broadinstitute.org/collaboration/giant/index.php/GIANT_consortium-data-files.

**References.**

112. Sudlow, C. et al. 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 (2015).
113. Liu, D. J. et al. Meta-analysis of gene-level tests for rare variant association. Nat. Genet. 46, 200–204 (2014).
114. Goldstein, J. I. et al. zCall: a rare variant caller for array-based genotyping: genetics and population analysis. Bioinformatics 28, 2543–2545 (2012).
115. Winkler, T. W. et al. Quality control and conduct of genome-wide association meta-analyses. Nat. Protoc. 9, 1192–1212 (2014).
116. Winkler, T. W. et al. EasyStrata: evaluation and visualization of stratified genome-wide association meta-analysis data. Bioinformatics 31, 259–261 (2015).
117. Purcell, S. M. et al. A polygenic burden of rare disruptive mutations in schizophrenia. Nature 506, 185–190 (2014).
118. Yang, J. et al. Genomic inflation factors under polygenic inheritance. Eur. J. Hum. Genet. 19, 807–812 (2011).
119. Yang, J. et al. Conditional and joint multiple-SNP analysis of GWAS summary statistics identifies additional variants influencing complex traits. Nat. Genet. 44, S1–S3 (2012).
120. Marchini, J., Howie, B., Myers, S., McVean, G. & Donnelly, P. A new multipoint method for genome-wide association studies by imputation of genotypes. Nat. Genet. 39, 906–913 (2007).
121. Wellcome Trust Case Control, C. Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. Nature 447, 661–678 (2007).
122. Marchini, J. & Howie, B. Genotype imputation for genome-wide association studies. Nat. Rev. Genet. 11, 499–511 (2010).
123. Frey, B. J. & Dueck, D. Clustering by passing messages between data points. Science 315, 972–976 (2007).
124. Moayyeri, A., Hammond, C. J., Valdes, A. M. & Spector, T. D. Cohort profile: TwinsUK and healthy ageing twin study. Int. J. Epidemiol. 42, 76–85 (2013).
125. Boyd, A. et al. Cohort profile: the ‘children of the 90s’--the index offspring of the Avon Longitudinal Study of parents and children. Int. J. Epidemiol. 42, 111–127 (2013).
126. Kutilik, Z., Whittaker, J., Waterworth, D., Beckmann, J. S. & Bergmann, S. Novel method to estimate the phenotypic variation explained by genome-wide association studies reveals large fraction of the missing heritability. Genet. Epidemiol. 35, 341–349 (2011).
127. Billingsley, P. Probability and Measure (Wiley, New York, 1986).
128. Surendran, P. et al. Trans-ancestry meta-analyses identify rare and common variants associated with blood pressure and hypertension. Nat. Genet. 48, 1151–1161 (2016).
129. Nikpay, M. et al. A comprehensive 1000 Genomes-based genome-wide association meta-analysis of coronary artery disease. Nat. Genet. 47, 1121–1130 (2015).
130. Storey, J. D. & Tibshirani, R. Statistical significance for genomewide studies. Proc. Natl Acad. Sci. USA 100, 9440–9445 (2003).
131. Civelek, M. et al. Genetic regulation of adipose gene expression and cardio-metabolic traits. Am. J. Hum. Genet. 100, 428–443 (2017).
# Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see Authors & Referees and the Editorial Policy Checklist.

## Statistical parameters

When statistical analyses are reported, confirm that the following items are present in the relevant location (e.g. figure legend, table legend, main text, or Methods section).

| Item                                                                 | Confirmed |
|----------------------------------------------------------------------|-----------|
| The exact sample size \(n\) for each experimental group/condition, given as a discrete number and unit of measurement | Yes       |
| An indication of whether measurements were taken from distinct samples or whether the same sample was measured repeatedly | Yes       |
| The statistical test(s) used AND whether they are one- or two-sided | Yes       |
| Only common tests should be described solely by name; describe more complex techniques in the Methods section. |            |
| A description of all covariates tested                             | Yes       |
| A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons | Yes       |
| A full description of the statistics including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals) | Yes       |
| For null hypothesis testing, the test statistic (e.g. \(F\), \(t\), \(r\)) with confidence intervals, effect sizes, degrees of freedom and \(P\) value noted | Yes       |
| Give \(P\) values as exact values whenever suitable.               | Yes       |
| For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings | Yes       |
| For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes | Yes       |
| Estimates of effect sizes (e.g. Cohen's \(d\), Pearson's \(r\)), indicating how they were calculated | Yes       |
| Clearly defined error bars                                          | Yes       |
| State explicitly what error bars represent (e.g. SD, SE, CI)        | Yes       |

Our web collection on statistics for biologists may be useful.

## Software and code

### Policy information about availability of computer code

**Data collection**

| Data collection | None |

**Data analysis**

| RareMetalWorker version 4.13; RVtests version 20140416; rareMETALs R-package version 5.8; EasyQC version 8.6; EasyStrata version 8.6; Eigensoft version 3.0; PLINK version 1.9; R version 3.0.3 |
| SNPRelate version 0.9.19 |

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors/reviewers upon request. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research guidelines for submitting code & software for further information.

## Data

### Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

https://portals.broadinstitute.org/collaboration/giant/index.php/GIANT_consortium_data_files
Field-specific reporting

Please select the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

- Life sciences
- Behavioural & social sciences
- Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/authors/policies/ReportingSummary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size
Sample size in association and expression quantitative trait loci studies were determined from individuals that pass quality control for genotyping and have phenotypes available. For the Drosophila experiment the flies were manually handled and counted and all experiments were performed with one to three biological replicates. We used Quanto version 1.2.4 to determine power given the sample size and allele frequencies. See Figure 2.

Data exclusions
Association studies
Individual cohorts identified ancestry population outliers based on 1000 Genome Project phase 1 ancestry reference populations. A centralized quality-control procedure implemented in EasyQC was applied to individual cohort association summary statistics to identify cohort-specific problems: (1) assessment of possible errors in phenotype residual transformation; (2) comparison of allele frequency alignment against 1000 Genomes Project phase 1 reference data to pinpoint any potential strand issues, and (3) examination of quantile-quantile (QQ) plots per study to identify any inflation arising from population stratification, cryptic relatedness and genotype biases.

Meta-analyses
During the meta-analyses, we excluded variants if they had call rate <95%, Hardy-Weinberg equilibrium P-value <1x10^-7, or large allele frequency deviations from reference populations (>0.6 for all ancestries analyses and >0.3 for ancestry-specific population analyses). We also excluded from downstream analyses markers not present on the Illumina ExomeChip array 1.0, variants on the Y-chromosome or the mitochondrial genome, indels, multiallelic variants, and problematic variants based on the Blat-based sequence alignment analyses.

Study-specific quality control (QC) measures of the genotyped variants were implemented before association analysis (Supplementary Tables 1-2) (this is far to extensive to outline here) - all exclusion criteria and protocols have been implemented and used previously and extensively published on.

Replication
We validated a total of 70 variants from Stage 1 that met P<2x10^-6 in two 1219 independent studies, the UK Biobank (Release 1) and Iceland (deCODE), comprising 119,572 and 12,605 individuals, respectively. The same QC and analytical methodology were used for these studies. For the combined analysis of Stage 1 plus 2, we used the inverse-variance weighted fixed effects meta-analysis method. Significant associations were defined as those nominally significant (P<0.05) in the Stage 2 study and for the combined meta-analysis (Stage 1 plus Stage 2) significance was set at P<2x10^-7 (0.05/250,000 variants). Variants that were not available in Stage 2 studies are noted in text and tables.

Drosophila studies
Flies were investigated pooled in groups of eight, and the screen comprised one to three biological replicates. We followed up each gene with a >0.2 increase or >0.4 decrease in triglyceride content. Thus, orthologues for two genes were brought forward for follow-up, DNAH10 and PLXND1. For both genes, we generated adipose tissue (cg-Gal4) and neuronal (elav-Gal4) specific RNAi-knockdown crosses to knockdown transcripts in a tissue specific manner, leveraging upstream activation sequence (UAS)-inducible short-hairpin knockdown lines, available through the VDRC (Vienna Drosophila Resource Center) to confirm the initial observations. Five-to-seven-day-old males were sorted into groups of 20, and the experiment was carried out in duplicate according to protocol, with one alteration: the samples were cleared of residual particulate debris by centrifugation before absorbance reading. Resulting triglyceride values were normalized to fly weight and larval/predation density. We only highlight the results that are reliably reproduced.

Randomization
In the quantitative trait analysis this is not relevant as there is no allocation into groups.

Blinding
NA

Reporting for specific materials, systems and methods
### Materials & experimental systems

| Involved in the study | n/a | Unique biological materials | Yes | Antibodies | Yes | Eukaryotic cell lines | Yes | Palaeontology | No | Animals and other organisms | Yes | Human research participants | No |

### Methods

| Involved in the study | n/a | ChIP-seq | Yes | Flow cytometry | Yes | MRI-based neuroimaging | Yes |

### Animals and other organisms

Policy information about [studies involving animals; ARRIVE guidelines](#) recommended for reporting animal research

| Laboratory animals | Data are from male progeny of crosses between male UAS-RNAi flies from the Vienna Drosophila Resource Center (VDRC) and Hsp70-GAL4; Tub-GAL8ts virgin females. (Additional details are found within the Online Methods and Supplementary Note). |
| Wild animals | NA |
| Field-collected samples | NA |