Genome-Wide Association Study of Body Fat Distribution identifies Novel Adiposity Loci and Sex-Specific Genetic Effects

Mathias Rask-Andersen[*1], Torgny Karlsson[1], Weronica E Ek[1], Åsa Johansson[*1]

[1]*Department of Immunology, Genetics and Pathology, Science for Life Laboratory, Uppsala University. Box 256, 751 05, Uppsala, Sweden.*

[*]*Corresponding authors: mathias.rask-andersen@igp.uu.se* and *asa.johansson@igp.uu.se*

Body mass and body fat composition are of clinical interest due to their links to cardiovascular- and metabolic diseases. Fat stored in the trunk has been suggested as more pathogenic compared to fat stored in other compartments of the body. In this study, we performed genome-wide association studies (GWAS) for the proportion of body fat distributed to the arms, legs and trunk estimated from segmental bio-electrical impedance analysis (sBIA) for 362,499 individuals from the UK Biobank. A total of 97 loci, were identified to be associated with body fat distribution, 40 of which have not previously been associated with an anthropometric trait. A high degree of sex-heterogeneity was observed and associations were primarily observed in females, particularly for distribution of fat to the legs or trunk. Our findings also implicate that body fat distribution in females involves mesenchyme derived tissues and cell types, female endocrine tissues a well as several enzymatically active members of the ADAMTS family of metalloproteinases, which are involved in extracellular matrix maintenance and remodeling.

Overweight (body mass index [BMI] >25) and obesity (BMI>30) have reached epidemic proportions globally[1]. Almost 40% of the world's population are now overweight[2] and 10.8% are obese[3]. Obesity is set to become the world's leading preventable risk factor for disease and early death due to the increased risks of developing type 2 diabetes, cardiovascular disease, and cancer[4].

The distribution of adipose tissue to discrete compartments within the human body is associated with differential risk for development of cardiovascular and metabolic disease[5]. Body fat distribution of fat is also well known to differ between sexes. After puberty, women accumulate fat in the trunk and limbs to a proportionally greater extent compared to other parts of the body, while men accumulate a greater extent of fat in the trunk[6]. Accumulation of adipose tissue around the viscera, the internal organs of the body, has been shown to be associated with increased risk of disease in

both men and women[7]. In contrast, the preferential accumulation of adipose tissue in the lower extremities, i.e. the hips and legs, has been suggested as a factor contributing to the lower incidence of myocardial infarction and coronary death observed in women during middle age[8]. The differential distribution of body fat between sexes has been attributed to downstream effects of sex hormone secretion[5]. However, the biological mechanisms that underlie body fat distribution have not been fully elucidated.

BMI is commonly used as a proxy measurement of body adiposity in epidemiological studies and in clinical practice. However, BMI is unable to discriminate between adipose and lean mass, and between fat stored in different compartments of the body. Other proxies that better represent distribution of body fat have also been utilized, such as the waist-to-hip ratio (WHR), waist circumference (WC), and hip circumference (HC). Through genome-wide association studies (GWAS), researchers have identified hundreds of loci to be associated with proximal measurements of body mass and body fat distribution such as BMI[9], WHR[10,11] and hip-, and waist circumference[11]. Sex-stratified analyses have revealed sexual dimorphic effects at twenty WHR-associated loci and 19 of these loci displayed stronger effects in women[12]. Body fat mass has also been studied in GWAS by using bio-electrical impedance analysis (BIA) and dual energy X-ray absorptiometry (DXA)[13,14]. BIA measures the electrical impedance through the human body, which can be used to calculate an estimate of the total amount of adipose tissue. The ‘gold standard’ method for measurements of body fat distribution is computed tomography (CT) or magnetic resonance imaging (MRI). However, these methods are costly. A GWAS has been performed for subcutaneous- and visceral adiposity, measured with

computed tomography scans, albeit in a relatively limited number of individuals (N=10,577)[15].

Developments in BIA technology has now allowed for cost-efficient segmental body composition scans that estimate of the fat content of the trunk, arms and legs[16] (Figure 1a). In this study, we used segmental BIA data on 362,499 participants of the UK Biobank to study the genetic determinants of body fat distribution to the trunk, arms and legs. For this purpose, we performed GWAS on the proportion of body fat distributed to these compartments. We also performed sex-stratified analyses to determine sex-specific effects and performed gene-sex interaction analyses to identify effects that differ between men and women.

Results

The proportions of body fat distributed to the arms – arm fat ratio (AFR), the legs – leg fat ratio (LFR) and the trunk – trunk fat ratio (TFR) were calculated by dividing the fat mass per compartment with the total body fat mass for each participant (Figure 1a). We conducted a two-stage GWAS using data from the interim release of genotype data in UK Biobank as a discovery cohort. Another set of participants, for which genotype data were made available as part of the second release, was used for replication. After removing non-Caucasians, genetic outliers and related individuals, 116,138 and 246,360 participants remained in the discovery and replication cohorts, respectively. Basic characteristics of the discovery and replication cohorts are presented in supplementary Table 1. Women were found to have higher total sBIA-estimated fat mass compared to men in both the discovery and replication cohort, as well as higher amount of fat in the arms and legs. Males had higher average

proportion of body fat located in the trunk compared to females (62.2% vs. 50.3%) and women had a larger proportion of body fat located in the legs (39.7% vs. 28.1%). While the total amount of adipose tissue in the arms was estimated to be higher in women compared to men, the fraction of adipose tissue distributed to the arms were similar. Several smaller differences between the discovery and replication cohorts were present (supplementary Table 1), such as some slight differences in height and age between men and women in the discovery and replication cohorts. These differences most likely represent the 50,000 participants for the UK Biobank Lung Exome Variant Evaluation (UK BiLEVE) project that were included in the first release of genotyping data for ~150,000 participants, which were used as a discovery cohort in this study. Selection for UK BiLEVE was conducted with specific consideration to lung function which may reflect the differences in baseline characteristics for this subset of the cohort. However, these differences are unlikely to affect the results from our analyses.

Genome wide association study for body fat ratio

GWAS was performed for each of the three phenotypes (AFR, LFR and TFR) in the whole discovery cohort (sex-combined) and when stratifying by sex (males and females), adjusting for covariates as described in the method section. A total of 25,472,837 imputed SNPs, with MAF of at least 0.0001, were analyzed in the discovery GWAS. LD score regression intercepts[17] ranged from 1.00 to 1.03 (supplementary Figure 1, supplementary Table 2), and were used to adjust for genomic inflation. We used the -clump function in PLINK[18], in combination with conditioning on the most significant SNP, to identify associations that were independent within each GWAS as well as between GWAS for the three body fat

ratios (AFR, TFR, or LFR) or strata (males, females or sex-combined) (see methods). In total, 133 independent associations at 114 loci were observed in the discovery analyses (Figure 1c, Supplementary Figure 2-4, Supplementary Data 1). For each of the independent associations, the leading SNP (the SNP that was most significant in any of the phenotypes or strata) was taken forward for replication. In total, 97 of the 133 independent associations (Supplementary Data 1) replicated ($P<0.05/133$). Out of these, 30 were associated with AFR, 42 with LFR and 65 with TFR in either the sex-combined or sex-stratified analyses. There was substantial overlap in associated loci between LFR and TFR loci (Figure 2) while associations with AFR overlapped to a small degree. One locus in the vicinity of *ADAMTSL3* was associated with all three phenotypes.

In the sex-stratified GWAS, only six loci were associated (replication P-value < 0.05/133) with body fat ratios in males while 71 loci were observed in females. Sex-heterogenous effects of associated variants were tested for using the GWAMA software. We observed 42 independent body fat ratio-associated variants whose effect differed between females and males ($P < 5.15*10^{-4}$, Table 2). In all but two cases, the effects were stronger, or only present in females. Stronger effects in males were observed for two AFR-associated variants rs3812049 near *SLC12A2*, and rs11289753 near *PLCE1* (Table 2).

LD score regression (LDSC) was used to estimate the fraction of variance of body fat ratios that could be explained by SNPs, i.e. the SNP heritability[17]. SNP heritability was higher in females compared to males for all traits and ranged from ~20% to ~26% in females and from 12% to 16% in males (supplementary Table 2).

Phenotypic and genetic correlation between body fat ratios and anthropometric phenotypes

Phenotypic correlations were assessed, in males and females separately, by calculating squared semi-partial correlation coefficients from the ANOVA table of two nested linear models that were adjusted for age and principal components, while genetic correlations were estimated using cross-trait LD score regression[19] (see methods). Overall, the genetic and phenotypic correlations showed a large degree of similarity (supplementary Table 3 and 4) and the correlation between the anthropometric traits and the ratios were in the same direction for phenotypic and genetic correlations for all correlations that reached the threshold for significance. In females, BMI and WC was strongly correlated with AFR both with regards to phenotypic (BMI - 78.9% and WC - 54.3%; squared semi-partial correlation coefficients, see supplementary Table 3) and genetic (BMI - 79.2% and WC - 51,8%; squared genetic correlation coefficients, see supplementary Table 4) correlations. Height also contributed to a moderate degree in explaining the phenotypic variance in LFR and TFR in females (16.0% and 25.3%) even though the genotypic correlation between height and both LFR and TFR was even higher (42.3% and 60.8% variance explained). In males, anthropometric traits contributed only to a small degree, up to 4.8%, in explaining the phenotypic variance in the ratios. This is also supported by the lower genetic correlations between these traits with the highest correlation seen between BMI and TFR (14.4% variance explained).

There was also a strong correlation between LFR and TFR in both males and females (>82% of the variance explained in both males and females and for both genetic and

phenotypic correlations, supplementary Table 3 and 4). LFR and TFR were inversely correlated, which agrees well with the large overlap in GWAS results for these phenotypes and the fact that the effect estimates from the GWAS was in the opposite direction for LFR and TFR (Supplementary Data 1). In contrast, AFR appeared to be more independent as only a low amount of phenotypic and genetic correlation was observed between AFR and the other two traits (supplementary Table 3 and 4).

Overlap with findings from previous GWAS

Body fat ratio-associated SNPs were tested for overlap with associations from previous GWAS for anthropometric traits by determining LD with entries from GWAS-catalog[20]. We used a strict cut-off of $R^2 < 0.1$ to distinguish novel findings. In total, we identify 40 body fat ratio-associated signals that have not previously been associated with an anthropometric trait (Table 1). Forty body fat ratio associated signals overlapped with previously identified height-associated loci[21,22] (Supplementary Data 1) and the majority of these signals were associated with TFR and/or LFR (36 out of 40). For AFR, the strongest associations were observed at well-known BMI and adiposity-associated loci such as: *FTO*, *MC4R*, *TMEM18*, *SEC16B* and *TFAP2B* (Supplementary Data 1), which agrees with the strong correlations (genotypic and phenotypic) between AFR and BMI.

We also compared the direction of the effects for overlapping GWAS results by estimating the effects of leading body fat ratio-associated SNPs on the respective overlapping anthropometric traits in the discovery cohort. The effects of TFR-associated SNPs were directionally consistent with effects on height and WHRadjBMI, while the effects were the opposite for LFR. The direction of effects

for AFR-associated SNPs were consistent with effects on BMI-, WC- and WHR (supplementary Table 5), which agrees with the strong genetic and phenotypic correlation between AFR and these phenotypes.

Among the novel loci, five overlapped with traits associated with cardiovascular disease: two loci near *ANKDD1B* and *KAT8* that have been associated with LDL cholesterol[23] and triglycerides[24], respectively; and three loci near *KCNH2*, *XKR6* and *BMP2* that have been associated with QT interval[25–27], thickness of the carotid intima media[28] and QRS complex[29], respectively (supplementary Table 6).

Functional annotation of associated loci

Functional annotation of the GWAS loci was performed by identifying overlap with eQTLs from the Genotype-Tissue Expression (GTEx) project[30] and by identifying potentially deleterious missense variants in LD ($R^2 > 0.8$) with our leading SNPs (see method section). In total, 31 body fat ratio-associated loci overlapped with an eQTL (Supplementary Data 2), and 11 leading SNPs were in LD with a potentially deleterious missense variant (Table 3). The probability for functional effects of missense variants has been predicted by sequence analyses[31,32] and Polyphen and SIFT-scores were used to assess the deleteriousness of the variants. Plausible causal missense variants were found in *ACAN*, *ADAMTS17*, FGFR4 and *ADAMTS10*, where the leading SNPs were predicted to have functional effects (Table 3). Rs315855, within *FGFR4*, has also previously been shown to be associated with progression of cancer[33,34] and to affect insulin secretion *in vitro*[35].

Enrichment analyses

To identify the functional roles and tissue specificity of associated variants, we performed enrichment analyses with DEPICT (Data-driven Expression Prioritized Integration for Complex Traits[36], see method section). In these analyses we used summary statistics from sex-stratified GWAS on the combined cohort (195,043 women and 167,408 men) in order to maximize statistical power. Enrichment was only detected for TFR- and LFR-associated genes in females as well as LFR-associated genes in males (supplementary Data 3). We identified 212 enriched gene sets and there was a substantial overlap of enriched gene sets between TFR- and LFR- in females as well as moderate overlap with LFR-associated gene sets in males (supplementary Figure 5). The large fraction of overlapping gene sets between LFR and TFR in females agrees well with the large overlap in GWAS signals. Gene sets related to bone morphology and skeletal development (abnormal skeleton morphology, short limbs, decreased length of long bones, skeletal system morphogenesis) were among the most strongly associated with both LFR and TFR. We also find the 'TGFβ signaling pathway' gene set to be enriched for genes within the TFR and LFR-associated loci in females, as well as TGFβ downstream mediators: 'SMAD1-', 'SMAD2-', 'SMAD3-' and 'SMAD7 protein-protein interaction subnetworks' (supplementary Data 3).

The results from enrichment analyses were compared with results from previous GWAS for height[21], BMI[9] and WHRadjBMI[12]. Gene-sets that were enriched for LFR and TFR-associated genes were compared to results from previous GWAS for height[21], BMI[9] and WHRadjBMI[12]. Substantial overlap of enriched gene sets was observed between LFR/TFR- with height and WHRadjBMI (Figure 3a). In contrast,

BMI associated biological processes overlapped only to a marginal extent: 2% of all gene sets that enriched for BMI-associated genes were also enriched for LFR- and TFR-associated genes.

Tissue enrichment was observed for LFR and TFR-associated genes in females (Figure 3b-c) in gene sets related to female reproduction, musculoskeletal systems, chondrocytes, mesenchymal stem cells, and fibroblasts. For TFR, DEPICT also revealed enrichment of genes associated with adipose tissue cells as well as endocrine and cardiovascular systems (Figure 3). Tissue enrichment was not seen for the other traits or strata.

Discussion

In this study, we performed GWAS on distribution of body fat to different compartments of the human body and identified and replicated 97 independent associations of which 40 have not been associated with any adiposity related phenotype previously. In contrast to previous studies, we have not addressed the total amount of fat but rather the fraction of the total body fat mass that is located in the arms, legs and trunk. Body fat distribution is well-known to differ between males and females, which we also clearly show in our study. We also show that the genetic effects that influence fat distribution are stronger in females compared to males. These results are consistent with previous GWAS that have revealed sexual dimorphisms in genetic loci for adiposity-related phenotypes, such as waist-circumference and waist-to-hip ratio[10,37,38]. Phenotypic and genetic correlations, as well as results from GWAS and subsequent enrichment analyses, also revealed that the amount of fat stored in the arms in females is highly correlated with BMI and WC.

This suggests that the proportion of fat stored in the arms will generally increase with increased accumulation of body mass and adipose tissue. In contrast, males exhibited moderate- to weak phenotypic and genetic correlations between the distribution of fat to different parts of the body and anthropometric traits, which indicates that the proportions of body fat mass in different compartments of the male body remains more stable as body mass and body adiposity increases. Among the three phenotypes analyzed in this study LFR and TFR were inversely correlated in both males and females. This suggests that LFR and TFR to a large extent describe one trait, i.e. the distribution of adipose tissue between these two compartments which is further supported by the large overlap in GWAS loci between the two phenotypes. In contrast, AFR was only weakly correlated with the other two traits.

Tissue enrichment revealed an important role in body fat distribution in females for mesenchyme derived tissues: i.e. adipose and musculoskeletal tissues; as well as tissues related to female reproduction. This suggests that the distribution of fat to the legs and trunk in females is mainly driven by the effect of female gonadal hormones on mesenchymal progenitors of musculoskeletal and adipose tissues. Enrichment analyses also showed that LFR and TFR have unique features that separates them from other anthropometric measurements, which was indicated by the portion of LFR/TFR-associated gene sets that did not overlap with height-, BMI or WHRadjBMI-associated gene sets. However, there was also an overlap in the functional aspects between these traits with both height and WHRadjBMI. This is indicated by the tissue enrichment profile for LFR/TFR-associated genes, which shares features with tissue enrichments reported for height in previous GWAS[39]: height-associated genes were strongly enriched in musculoskeletal tissue types with

additional enrichment in cardiovascular and endocrine tissue types; as well as WHRadjBMI[12]: associated genes were enriched in adipocytes and adipose tissue subtypes. Of particular note, we did not identify any enrichment of body fat ratio-associated genes in CNS tissue gene sets in contrasts to enrichment analyses in previous GWAS for BMI where CNS have been indicated to play a prominent role in obesity susceptibility[9].

In the GWAS for LFR and TFR in females, we find that several genes that highlight the influence of biological processes related to the interaction between cells and the extracellular matrix (ECM), as well as ECM-maintenance and remodeling. These include *ADAMTS2*, *ADAMTS3*, *ADAMTS10*, *ADAMTS14*, and *ADAMTS17*, which encode extracellular proteases that are involved in enzymatic remodeling of the ECM. Two leading SNPs were in LD with potentially damaging missense mutations in *ADAMTS10* and *ADAMTS17* and two other GWAS signals overlapped with eQTLs for *ADAMTS14* and *ADAMTS3*. In addition, possibly deleterious missense mutations in LD with our leading GWAS SNPs were also found for *VCAN* and *ACAN*. Both *VCAN* and *ACAN* encode chondroitin sulfate proteoglycan core proteins that constitute structural components of the extracellular matrix, particularly in soft tissues[40]. These proteins also serve as major substrates for ADAMTS proteinases[41]. ECM forms the three-dimensional support structure for connective and soft tissue. In fat tissue, the ECM regulates adipocyte expansion and proliferation[42]. Remodeling of the ECM is required to allow for adipose tissue growth and this is achieved through enzymatic processing of extracellular molecules such as proteoglycans, collagen and hyaluronic acid. For example, ADAMTS2, 3 and 14 act as procollagen N-propeptidases that mediate the maturation of triple helical collagen fibrils[43,44]. We

therefore propose that effects of genetic variation in biological systems involved in ECM-remodeling as a factor underlying normal variation in female body fat distribution.

Conclusions

In summary, GWAS of body fat distribution determined by sBIA reveals a genetic architecture that influences distribution of adipose tissue to the arms, legs and trunk. Genetic associations and effects clearly differ between sexes, in particular for distribution of adipose tissue to the legs and trunk. The distribution of body fat in women has been previously been suggested as a causal factor leading to lower risk of cardiovascular and metabolic disease, as well as cardiovascular mortality for women in middle age[5] and genetic studies have identified SNPs that are associated with a 'favorable' body fat distribution[45], i.e. with higher BMI but lower risk of cardiovascular and metabolic disease. The capacity for peripheral adipose storage has been highlighted as one of the components underlying this phenomenon[45]. Resolving the genetic determinants and mechanisms that lead to a favorable distribution of body fat may help in risk assessment and in identifying novel venues for intervention to prevent or treat obesity-related disease.

Author contributions

MRA, TK, WE and ÅJ conceived of and designed the study. Analysis was performed by MRA and WE under supervision by ÅJ. MRA analyzed the data and wrote the first draft of the manuscript. All authors contributed to the final version of the manuscript.

Acknowledgements

We are grateful to the participants and staff of the UK Biobank. Access to UK Biobank genetic and phenotypic data was granted under application no. 15152. Computations were performed on the computational cluster at the Uppsala Multidisciplinary Center for Advanced Computational Science (UPPMAX) under projects b2016021 and b2017066. The work was supported by grants from the Swedish Society for Medical Research (SSMF), the Kjell and Märta Beijers Foundation, Göran Gustafssons Foundation, the Swedish Medical Research Council (Project Number 2015-03327), the Marcus Borgström Foundation, and the Åke Wiberg Foundation.

online Methods

UK Biobank participants

The first release of imputed genotype data from UK Biobank (N = 152,249) was used as a discovery cohort, and imputed genotype data from an unrelated set of participants from the third genotype release (N = 326,565) as a replication cohort. Participants who self-reported as being of British descent (data field 21000) and were classified as Caucasian by principal component analysis (data field 22006) were included in the analysis. Genetic relatedness pairing was provided by the UK Biobank (Data field 22011). Participants were removed due to relatedness based on kinship data (estimated genetic relationship > 0.044), poor genotyping call rate (<95%), high heterozygosity (Data field 22010), or sex-errors (Data filed 22001). After filtering, 116,138 participants remained in the discovery cohort and 246,361 in the replication cohort and were included in downstream analyses.

Ethics

Ethical approval to collect participant data was given by the North West Multicentre Research Ethics Committee, the National Information Governance Board for Health & Social Care, and the Community Health Index Advisory Group. UK Biobank possesses a generic Research Tissue Bank approval granted by the National Research Ethics Service (http://www.hra.nhs.uk/), which lets applicants conduct research on UK Biobank data without obtaining separate ethical approvals. Access to UK Biobank genetic and phenotypic data was granted under application no. 15152. All participants provided signed consent to participate in UK Biobank [46].

Genotyping, imputations and QC.

Genotyping in the discovery cohort had been performed on two custom-designed microarrays: referred to as UK BiLEVE and Axiom arrays, which genotyped 807,411 and 820,967 SNPs, respectively. Imputation had been performed using UK10K[47] and 1000 genomes phase 3[48] as reference panels. Prior to analysis, we filtered SNPs based on call rate (--geno 0.05), HWE (P-value > 10^{-20}, MAF (--maf 0.0001) and imputation quality (Info > 0.3) resulting in 25,472,837 SNPs in the discovery cohort. The third release of data from the UK Biobank contained genotyped and imputed data for 488,366 participants (partly overlapping with the first release). For our replication analyses, we included an independent subset that did not overlap with the discovery cohort. Genotyping in this subset was performed exclusively on the UK Biobank Axiom Array. This dataset included 47,512,111 SNPs that were filtered based on HWE ($P<10^{-20}$), call rate (--geno 0.05), (Info >0.3) and MAF (--maf 0.0001). All genomic positions are in reference to hg19/build 37.

Phenotypic measurements

The phenotypes used in this study derive from impedance measurements produced by the Tanita BC418MA body composition analyzer. Participants were barefoot, wearing light indoor clothing, and measurements were taken with participants in the standing position. Height and weight were entered manually into the analyzer before measurement. The Tanita BC418MA uses eight electrodes: two for each foot and two for each hand. This allows for five impedance measurements: whole body, right leg, left leg, right arm and left arm (Figure 1a.). Body fat for the whole body and individual body parts had been calculated using a regression formula, that was derived from reference measurements of body composition by DXA (Figure 1b) in Japanese and Western subjects. This formula uses weight, age, height and impedance measurements[49] as input data. Arm, and leg fat masses were averaged over both limbs. Arm, leg, and trunk fat masses were then divided by the total body fat mass to obtain the ratios of fat mass for the arms, legs and trunk, i.e. what proportion of the total fat in the body is distributed to each of these compartments. These variables were analyzed in this study and were named: arm fat ratio (AFR), leg fat ratio (LFR), and trunk fat ratio (TFR).

Assessing the relationship between adipose tissue ratios and anthropometric traits

Phenotypic correlations between fat distribution ratios and anthropometric traits were estimated by calculating semi-partial correlation coefficients for males and females separately , using anova.glm in R. Adipose tissue ratios (AFR, LFR or TFR) were set as the response variable. BMI, waist circumference, waist circumference adjusted for BMI, waist-to-hip ratio, height, or one of the other ratios were included as the last term in a linear model that included, age and principal components as covariates. The reduction in residual deviance, i.e., the reductions in the residual sum of squares as

BMI, waist circumference, waist circumference adjusted for BMI, waist-to-hip ratio, height, or one of the other ratios was added to the model, is presented as percentages of the total deviance of the null model in supplementary Table 3.

Genome wide association study for body fat ratio

A two-stage GWAS was performed using a discovery and a replication cohort. Body fat ratios were adjusted for age, age squared and normalized by rank-transformation separately in males and females using the *rntransform* function included in the GenABEL library[50]. GWAS was performed in PLINK v1.90b3n[18] using linear regression models with AFR, LFR, and TFR as the response variables and the SNPs as predictor variables. A batch variable was used as covariate in the GWAS for the discovery analyses to adjust for genotyping array (Axiom and BiLEVE). We also included age, the first 15 principal components and sex (in the sex-combined analyses) as covariates in the GWAS. LD score regression intercepts (see further information below), calculated using ldsc[17], were used to adjust for genomic inflation, by dividing the square of the t-statistic for each tested SNP with the LD-score regression intercept for that GWAS, and then calculating new p-values based on the adjusted t-statistic. We used a threshold of $P<10^{-7}$, after adjusting for LD score intercept, as threshold for significance in the discovery cohort.

The -clump function in PLINK was used to identify the number of independent signals in each GWAS. This function groups associated SNPs based on the linkage disequilibrium (LD) pattern. The parameters for clumps were set to: -clump-p1 $1*10^{-7}$, clump-p2 $1*10^{-7}$, clump-r2 0.10 and --clump-kb 1000. This function groups SNPs within one million base pairs that were associated with the trait at $p < 1*10^{-7}$. Several

associations were found in more than one of the three body fat ratios (AFR, TFR, or LFR) or strata (males, females or sex-combined) and different leading SNPs were observed for different traits and strata at several loci. To assess whether these represented the same signal, we assessed the LD between overlapping leading SNPs in PLINK. SNPs in low LD (R^2-value < 0.05) were considered to represent independent signals. For leading SNPs where we could not exclude that they were in LD (R^2-value >= 0.05) we performed conditional analysis in PLINK, conditioning on the most significant SNP across all phenotypes and strata. Associations with a P< $1*10^{-7}$ after conditioning for the most significant SNP were considered as being independent signals. For each independent signal, the leading SNP (lowest p-value) was taken forward for replication. Meta analyses of results from the discovery and replication cohorts was performed with the METAL software[51] for all independent associations that were taken forward for replication.

SNP heritability, and genetic correlations

We estimated SNP heritability and genetic correlations using LD score regression (LDSC), implemented in the 'ldsc' software package[17]. LDSC uses LD patterns and summary stats from GWAS as input. For genetic correlations, we performed additional sex-stratified GWAS in the UK biobank (using the same covariates as for the ratios) for standard anthropometric traits, BMI, height, WC, WHR, WCadjBMI and WHRadjBMI, in the discovery cohort. GWAS summary stats were filtered for SNPs included in HapMap3 to reduce likelihood of bias induced by poor imputation quality. After this filtering, 1,164,192 SNPs remained for LDSC analyses. LD scores from the European data of the 1000 Genomes project (including LD patterns for all the HapMap3 SNPs) for use with LDSC were downloaded from the Broad institute at:

https://data.broadinstitute.org/alkesgroup/LDSCORE/eur_w_ld_chr.tar.bz2. Genetic correlations between the three body fat ratios and anthropometric traits were assessed by cross-trait LD score regression.

Overlap with findings from previous GWAS

Leading SNPs from all independent signals in our analyses were cross referenced with the NHGRI-EBI catalog of published genome-wide association studies (GWAS Catalog – data downloaded on 23 April 2018) [20] to determine if body fat ratio-associated signals overlapped with previously identified anthropometric associations from previous GWAS. We used a cutoff of $R^2 < 0.1$ between SNPs from our analyses and anthropometric trait-associated SNPs ($P < 5*10^{-8}$) from GWAS catalog to determine if an association was novel. LD between data in the GWAS catalogue and our leading SNPs were calculated using PLINK v1.90b3n[18]. In addition, leading SNPs at Body fat ratio-associated loci that potentially overlapped ($R^2 > 0.1$) with signals from previous GWAS were tested for association with standard anthropometric traits (BMI, height, WC, WHR, WCadjBMI and WHRadjBMI) in the UK biobank discovery cohort using PLINK v1.90b3n[18] through linear regression modelling and including sex, age a batch variable and 15 principal components as covariates. Here, a P-value of $P < 1e-7$ was considered significant.

Functional annotation of associated loci

Associated loci were investigated for overlap with eQTLs from the GTEx project [30]. The threshold for significance for the eQTLs was set to $2.3*10^{-9}$ in agreement with previous studies[52]. The strongest associated SNP for each tissue and gene in the GTEx dataset was identified. We then estimated the LD between the top eQTL SNPs and the

leading SNPs for each independent association from our analysis. If a SNPs from our analyses were in LD ($R^2 > 0.8$) with a leading eQTL SNP the two signals were considered overlapping.

Leading SNPs, and all SNPs in LD ($R^2 > 0.8$) with a leading SNPs from our analyses (LD determined in the UK biobank cohort in PLINK) were cross-referenced with dbSNP (human 9606 b150) in order to identify potentially deleterious intragenic variants in LD ($R^2 > 0.8$) with body fat ratio-associated variants. Polyphen and SIFT-scores for the missense variants (extracted from Ensembl - www.ensembl.org) were used to assess the deleteriousness of the variants.

Enrichment analysis

To identify the functional roles and tissue specificity of associated variants, we performed tissue and gene-set enrichment analyses using DEPICT[36]. For the gene-set enrichment in DEPICT, gene expression data from 77,840 samples have been used to predict gene function for all genes in the genome based on similarities in gene expression. In comparison to standard enrichment tools that apply a binary definition to define membership in a set of genes that have been associated with a biological pathway or functional category (genes are either included or not included), in DEPICT, the probability of a gene being a member of a gene set has instead been estimated based on correlation in gene expression. This membership probability to each gene set has been estimated for all genes in the human genome and the membership probabilities for each gene have been designated 'reconstituted' gene sets. A total of 14,461 reconstituted gene sets have been generated which represent a wide set of biological annotations (Gene Ontology [GO], KEGG, REACTOME,

Mammalian Phenotype [MP], etc.). For tissue enrichment in DEPCIT, microarray data from 37,427 human tissues have been used to identify genes with high expression in different cells and tissues. When performing the tissue enrichment analyses in DEPICT, the most significant GWAS hits are scanned for enrichment of genes that are highly expressed in any cells or tissues.

For the enrichment analyses, we performed sex-stratified GWAS for AFR, LFR and TFR on a combined set of the UK Biobank cohort, in order to achieve high power for enrichment analyses, including 195,043 females and 167,408 males. The 'clump' functionality in PLINK is used to determine associated loci. The p-value cut off for association for 'clump' was set at $P < 10^{-7}$. In the enrichment analyses, DEPICT assesses whether the reconstituted gene sets are enriched for genes within trait-associated loci[36]. The False Discovery Rate (FDR)[53] was used to adjust for multiple testing. Twelve analyses were run in total (tissue enrichment and gene-set enrichment) and FDR<0.05/12 was considered significant.

Interaction between SNPs and sex

We used the GWAMA software[54] to test for heterogenous effects of associated SNPs between sexes. In GWAMA, fixed-effect estimates of sex-specific and sex-combined beta coefficients and standard errors are calculated from GWAS summary statistics to test for heterogeneous allelic effects between females and males. GWAMA obtains a test-statistic by subtracting the sex-combined squared t-statistic from the sum of the two sex-specific squared t-statistics. This test statistic is asymptotically chi-square-distributed and equivalent to a normal z-test of the difference in allelic effects between sexes. SNPs that were significant in the replication analyses in any of the

strata were tested for heterogeneity between sexes in the replication cohort. Bonferroni correction was used to correct for multiple testing and p-values < 0.05/97 were considered to be significant.

Data availability

The data that support the findings of this study are available from UK Biobank (http://www.ukbiobank.ac.uk/about-biobank-uk/). Restrictions apply to the availability of these data, which were used under license for the current study (project no. 15152). Data is available for bona fide researchers upon application to the UK Biobank. Summary statistics from all association tests are available for download at: https://myfiles.uu.se/ssf/s/readFile/share/3993/1270878243748486898/publicLink/GWAS_summary_stats_ratios.zip.

References

1. Afshin, A. *et al*. Health Effects of Overweight and Obesity in 195 Countries over 25 Years The. *N. Engl. J. Med.* **377,** 13–27 (2017).
2. Ng, M. *et al*. Global, regional, and national prevalence of overweight and obesity in children and adults during 1980-2013: A systematic analysis for the Global Burden of Disease Study 2013. *Lancet* **384,** 766–781 (2014).
3. Di Cesare, M. *et al*. Trends in adult body-mass index in 200 countries from 1975 to 2014: A pooled analysis of 1698 population-based measurement studies with 19.2 million participants. *Lancet* **387,** 1377–1396 (2016).
4. Pi-Sunyer, F. X. *et al*. *Clinical guidelines on the identification, evaluation, and treatment of overweight and obesity in adults: The Evidence Report*. (1998).

5. Link, J. C. & Reue, K. The Genetic Basis for Sex Differences in Obesity and Lipid Metabolism. *Annu. Rev. Nutr*. **37,** annurev-nutr-071816-064827 (2017).

6. Bouchard, C., Despreis, J. P. & Maurieige, P. Genetic and nongenetic determinants of regional fat distribution. *Endocr. Rev.* **14,** 72–93 (1993).

7. Wajchenburg, B. L. Subcutaneous and Visceral Adipose Tissue□: Their Relation to the Metabolic Syndrome. *Endocr. Rev.* **21,** 697–738 (2014).

8. Tunstall-Pedoe, H. Myth and paradox of coronary risk and the menopause. *Lancet* **351,** 1425–1427 (1998).

9. Locke, A. E. *et al*. Genetic studies of body mass index yield new insights for obesity biology. *Nature* **518,** 197–206 (2015).

10. Heid, I. M. *et al.* Meta-analysis identifies 13 new loci associated with waist-hip ratio and reveals sexual dimorphism in the genetic basis of fat distribution. *Nat. Genet.* **42,** 949–960 (2010).

11. 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).

12. Shungin, D. *et al.* New genetic loci link adipose and insulin biology to body fat distribution. *Nature* **518,** 187–196 (2015).

13. Kilpeläinen, T. O. *et al*. Genetic variation near IRS1 associates with reduced adiposity and an impaired metabolic profile. *Nat. Genet.* **43,** 753–760 (2011).

14. Lu, Y. *et al.* New loci for body fat percentage reveal link between adiposity and cardiometabolic disease risk. *Nat. Commun.* **7,** (2016).

15. Fox, C. S. *et al*. Genome-wide association for abdominal subcutaneous and visceral adipose reveals a novel locus for visceral fat in women. *PLoS Genet.* **8,** (2012).

16. Mally, K., Trentmann, J., Heller, M. & Dittmar, M. Reliability and accuracy of

segmental bioelectrical impedance analysis for assessing muscle and fat mass in older Europeans: A comparison with dual-energy X-ray absorptiometry. *Eur. J. Appl. Physiol.* **111,** 1879–1887 (2011).

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

18. Weeks, J. P. plink: An R Package for Linking Mixed-Format Tests Using IRT-Based Methods. *J. Stat. Softw*. **35,** 1–33 (2010).

19. Bulik-Sullivan, B. *et al.* An atlas of genetic correlations across human diseases and traits. *Nat. Genet.* **47,** 1236–1241 (2015).

20. MacArthur, J. *et al.* The new NHGRI-EBI Catalog of published genome-wide association studies (GWAS Catalog). *Nucleic Acids Res*. **45,** D896–D901 (2017).

21. Wood, A. R. *et al.* Defining the role of common variation in the genomic and biological architecture of adult human height. *Nat. Genet.* **46,** 1173–86 (2014).

22. He, M. *et al.* Meta-analysis of genome-wide association studies of adult height in East Asians identi fi es 17 novel loci. *Hum. Mol. Genet.* **24,** 1791–1800 (2017).

23. Kathiresan, S. *et al.* Six new loci associated with blood low-density lipoprotein cholesterol, high-density lipoprotein cholesterol or triglycerides in humans. *Nat. Genet.* **40,** 189–197 (2008).

24. Teslovich, T. M. *et al.* Biological, clinical and population relevance of 95 loci for blood lipids. *Nature* **466,** 707–713 (2010).

25. Newton-Cheh, C. *et al.* Common variants at ten loci influence QT interval duration in the QTGEN Study. *Nat. Genet.* **41,** 399–406 (2009).

26. Holm, H. *et al.* Several common variants modulate heart rate, PR interval and QRS duration. *Nat. Genet.* **42,** 117–122 (2010).

27. Arking, D. E. *et al.* Genetic association study of QT interval highlights role for calcium signaling pathways in myocardial repolarization. *Nat. Genet.* **46,** 826–836 (2014).

28. Bis, J. C. *et al.* Meta-analysis of genome-wide association studies from the CHARGE consortium identifies common variants associated with carotid intima media thickness and plaque. *Nat. Genet.* **43,** 940–948 (2011).

29. van der Harst, P. *et al.* 52 Genetic Loci Influencing Myocardial Mass. *J. Am. Coll. Cardiol.* **68,** 1435–1448 (2016).

30. Aguet, F. *et al.* Genetic effects on gene expression across human tissues. *Nature* **550,** 204–213 (2017).

31. Ng, P. C. & Henikoff, S. SIFT□: predicting amino acid changes that affect protein function. **31,** 3812–3814 (2003).

32. Adzhubei, I., Jordan, D. M. & Sunyaev, S. R. *Predicting functional effect of human missense mutations using PolyPhen-2*. *Current Protocols in Human Genetics* (2013). doi:10.1002/0471142905.hg0720s76

33. Allele, A. *et al.* Cancer Progression and Tumor Cell Motility Are Associated with the FGFR4. 840–847 (2002).

34. Frullanti, E. *et al.* Meta and pooled analyses of FGFR4 Gly388Arg polymorphism as a cancer prognostic factor. *Eur. J. Cancer Prev.* **20,** 340–347 (2011).

35. Ezzat, S. *et al.* The Cancer-associated FGFR-G388R Polymorphism Enhances Pancreatic Insulin Secretion and Modifies the Risk of Diabetes. *Cell Metab.* 929–940 (2013). doi:10.1016/j.cmet.2013.05.002

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

37. Randall, J. C. *et al.* Sex-stratified Genome-wide Association Studies Including 270,000 Individuals Show Sexual Dimorphism in Genetic Loci for Anthropometric Traits. *PLoS Genet.* **9,** (2013).

38. 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 genetics* **11,** (2015).

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

40. Wight, T. N. Versican: A versatile extracellular matrix proteoglycan in cell biology. *Curr. Opin. Cell Biol.* **14,** 617–623 (2002).

41. Binder, M. J., McCoombe, S., Williams, E. D., McCulloch, D. R. & Ward, A. C. The extracellular matrix in cancer progression: Role of hyalectan proteoglycans and ADAMTS enzymes. *Cancer Lett.* **385,** 55–64 (2017).

42. Lin, D., Chun, T. & Kang, L. Adipose extracellular matrix remodelling in obesity and insulin. *Biochem. Pharmacol.* **119,** 8–16 (2016).

43. Bekhouche, M. *et al.* Determination of the substrate repertoire of ADAMTS2, 3, and 14 significantly broadens their functions and identifies extracellular matrix organization and TGF-β signaling as primary targets. *FASEB J.* **30,** 1741–1756 (2016).

44. Kelwick, R., Desanlis, I., Wheeler, G. N. & Edwards, D. R. The ADAMTS (A Disintegrin and Metalloproteinase with Thrombospondin motifs) family. (2015). doi:10.1186/s13059-015-0676-3

45. Yaghootkar, H. *et al*. Genetic Evidence for a Link Between Favorable

Adiposity and Lower Risk of Type 2 Diabetes , Hypertension , and Heart Disease. **65,** 2448–2460 (2016).

46. UK Biobank. UK Biobank ethics and governance framework version 3.0. **0,** 3–18 (2007).

47. Walter, K. *et al.* The UK10K project identifies rare variants in health and disease. *Nature* **526,** 82–89 (2015).

48. Auton, A. *et al.* A global reference for human genetic variation. *Nature* **526,** 68–74 (2015).

49. LTD., T. U. *Body Composition Analyzer BC-418MA Instruction Manual*. (2009).

50. Aulchenko, Y. S., Ripke, S., Isaacs, A. & van Duijn, C. M. GenABEL: an R library for genome-wide association analysis. *Bioinformatics* **23,** 1294–1296 (2007).

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

52. Ferreira, M. A. *et al.* Shared genetic origin of asthma, hay fever and eczema elucidates allergic disease biology. *Nat. Genet.* **49,** 1752–1757 (2017).

53. Benjamini, Y. & Hochberg, Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society* **57,** 289–300 (1995).

54. Magi, R., Lindgren, C. M. & Ã, A. P. M. Meta-Analysis of Sex-Specific Genome-Wide Association Studies. **853,** 846–853 (2010).

55. Wood, A. R. *et al.* Defining the role of common variation in the genomic and biological architecture of adult human height. *Nat. Genet.* **46,** 1173–1186

(2014).

56. Mägi, R. & Morris, A. P. GWAMA: Software for genome-wide association meta-analysis. *BMC Bioinformatics* **11,** (2010).

Figure legends

Figure 1. Segmental body impedance analyses use bioelectrical impedance to estimate body composition: fat mass, muscle mass, etc. Adipose tissue mass, in this study, had been estimated using the Tanita BC-418MA body composition analyzer (**a**). This machine uses an eight-electrode method, which allows for five measurements of impedance. Current is supplied to the front of both feet and the fingertips of both hands. Voltage is measured on either heel or thenar portion of the palms. Body composition is derived from a regression formula for each body part. The formula is derived from regression analysis using height, weight, age and impedance for each body part as predictors for composition of each body part as assessed by DXA (**b**). GWAS for AFR, LFR and TFR were conducted in the UK Biobank cohort and revealed associations with loci that have not previously been associated with standard anthropometric traits. (**c**) shows a manhattan plot with combined results for association studies of body fat ratios in combined and sex-stratified analyses. Novel loci are highlighted in red.

Figure 2. Body fat ratio associated loci and the overlap of associations between AFR, LFR and TFR. Loci are denoted by the nearest gene or by the most likely target gene (see methods section).

Figure 3. Enrichment analyses of genes at LFR and TFR associated loci. (a)

Reconstituted gene sets found to be enriched for TFR- and LFR-associated genes (in males and females) were compared with results from previous GWAS on WHRadjBMI[12], BMI[9] and height[55]. Tissue and cell type enrichment of (**b**) TFR- and (**c**) LFR-associated genes. Red bars denote tissues gene sets that were significantly enriched for LFR- and TFR-associated genes at FDR < 0.05/12.

Supplementary Figure and Tables

Supplementary Table 1. Basic characteristics of UK Biobank participants included in the analyses.

Supplementary Table 2. SNP-heritability (*h2*) and LD score regression intercept estimates of body fat ratios.

Supplementary Table 3. Phenotypic correlation between body fat ratios and anthropometric traits.

Supplementary Table 4. Genetic correlations between body fat ratios and anthropometric traits.

Supplementary Table 5. Direction of effect for body fat ratio–associated leading SNPs for standard anthropometric traits.

Supplementary Table 6. Overlap between body fat ratio-associated loci with non-anthropometric trait-associations from previous GWAS.

Supplementary Figure 1. Quantile-quantile (QQ) plots of the distribution of p-values from GWAS of body fat ratios in the UK Bibank sex-combined, female and male discovery cohorts.

Supplementary Figure 2. Manhattan plots for GWAS results for arm fat ratio in the combined discovery cohort and sex-stratified analyses.

Supplementary Figure 3. Manhattan plots for GWAS results for leg fat ratio in the combined discovery cohort and sex-stratified analyses.

Supplementary Figure 4. Manhattan plots for GWAS results for trunk fat ratio in the combined discovery cohort and sex-stratified analyses.

Supplementary Figure 5. Overlap of enriched gene sets from DEPICT for body fat ratios.

Supplementary data sets

Supplementary Data 1. Independent signals for association with body fat ratios.

Supplementary Data 2. Expression quantitative loci (eQTLS) in LD with leading SNPs at body fat ratio-associated loci.

Supplementary Data Set 3. Results from enrichment analyses by DEPICT.

Tables

Table 1. Novel body fat ratio-associated loci. Entries represent loci that have not previously been associated with an anthropometric trait.

					Discovery		Replication		Meta-analysis	
Locus	**Lead SNP**	**MAF**	**Most likely target gene**	**Strongest associated trait**	***β***	**P-value**	***β***	**P-value**	**P-value**	**Direction**
chr1:10,273,179 - 10,370,712	rs4846204	12.7%	*KIF1B*[α]	TFR - combined	0.035	$2.87*10^{-8}$	0.026	$2.25*10^{-9}$	$7.03*10^{-16}$	++
chr1:155,143,768 - 155,144,300*	rs4971091	37.7%	*KRTCAP2*[α]	AFR - combined	-0.024	$8.23*10^{-8}$	-0.013	$1.35*10^{-5}$	$3.62*10^{-11}$	--
chr1:154,848,581 - 156,027,203	rs180921974	2.3%	*PKLR*[α]	TFR - combined	-0.102	$1.05*10^{-12}$	-0.097	$9.45*10^{-24}$	$7.52*10^{-35}$	--
chr2:57,961,602	rs13011472	48.9%	*VRK2*[α]	AFR - combined	-0.024	$4.81*10^{-8}$	-0.013	$6.96*10^{-6}$	$1.17*10^{-11}$	--
chr2:198,540,352 - 198,729,207	rs148812496	49.7%	*RFTN*[α]	AFR - combined	-0.030	$2.41*10^{-8}$	-0.015	$4.19*10^{-5}$	$8.07*10^{-11}$	--
chr3:48,939,080 - 49,137,904	rs4521268	33.3%	*WDR6*[β]	TFR - combined	-0.025	$4.30*10^{-8}$	-0.021	$5.00*10^{-12}$	$1.46*10^{-18}$	--
chr3:50,034,637 - 50,034,637	rs1986599	11.3%	*RBM6*[α]	TFR - combined	-0.037	$4.28*10^{-8}$	-0.021	$1.47*10^{-6}$	$1.70*10^{-12}$	--
chr3:52,718,183	rs565817433	43.3%	*NEK4*[β]	TFR - combined	-0.023	$6.83*10^{-8}$	-0.013	$5.92*10^{-6}$	$1.13*10^{-11}$	--
chr4:8,599,467 - 8,604,153	rs2241069	46.1%	*CPZ*[α]	TFR - females	0.033	$1.56*10^{-8}$	0.019	$7.86*10^{-7}$	$3.97*10^{-13}$	++
chr5:74,689,249 - 74,934,009	rs34341	42.3%	*ANKDD1B*[β]	AFR - females	-0.036	$1.47*10^{-9}$	-0.025	$1.20*10^{-10}$	$2.74*10^{-18}$	--
chr5:127,263,443 - 127,562,880	rs3812049	25.6%	*SLC12A2*[α]	AFR - males	0.061	$9.24*10^{-18}$	0.021	$1.99*10^{-6}$	$2.40*10^{-6}$	++
chr5:157,952,404	rs1317415	30.7%	*RP11-32D16.1*[α]	LFR - females	-0.034	$6.11*10^{-8}$	-0.018	$1.81*10^{-5}$	$4.71*10^{-11}$	--
chr5:176,677,563 - 176,732,485	rs34022431	2.6%	*NSD1*[α]	AFR - combined	0.071	$8.55*10^{-8}$	0.040	$7.12*10^{-6}$	$1.67*10^{-11}$	++
chr5:178,470,304 - 178,547,313	rs888762	33.0%	*ADAMTS2*[α]	TFR - females	0.037	$5.48*10^{-9}$	0.027	$1.11*10^{-10}$	$7.50*10^{-18}$	++
chr6:34,539,910 - 34,825,662	rs2492863	13.8%	*C6orf106*[α]	LFR - females	-0.053	$3.48*10^{-10}$	-0.046	$5.26*10^{-16}$	$1.47*10^{-24}$	--
chr7:20,374,338 - 20,465,963	rs3823974	41.2%	*ITGB8*[α]	TFR - females	-0.039	$6.56*10^{-11}$	-0.028	$2.43*10^{-12}$	$3.09*10^{-21}$	--
chr7:150,604,406 -	rs56282717	24.3%	*KCNH2*[α]	AFR - combined	-0.027	$5.00*10^{-8}$	-0.020	$5.59*10^{-9}$	$3.03*10^{-}$	--

150,657,209									15	
chr8:8,907,950 - 9,054,817	rs2044387	42.5%	*ERI1*[α]	AFR - females	0.035	$1.14*10^{-8}$	0.024	$1.03*10^{-9}$	$1.69*10^{-16}$	++
chr8:10,802,146	rs12546366	45.5%	*XKR6*[α]	AFR - females	0.031	$9.29*10^{-8}$	0.021	$6.50*10^{-8}$	$8.23*10^{-14}$	++
chr9:16,846,111	rs10962638	14.2%	*BNC2*[α]	TFR - combined	-0.037	$4.22*10^{-9}$	-0.014	$3.90*10^{-4}$	$4.50*10^{-10}$	--
chr9:86,582,923 - 86,639,999	rs7039458	24.8%	*RMI1*[β, δ]	TFR - females	0.038	$2.07*10^{-8}$	0.028	$3.32*10^{-10}$	$7.20*10^{-17}$	++
chr9:129,672,095 - 129,945,847	rs3780327	22.0%	*RALGPS1*[α]	LFR - combined	-0.031	$1.95*10^{-9}$	-0.016	$2.50*10^{-6}$	$3.40*10^{-13}$	--
chr10:72,412,776 - 72,435,363	rs34821335	27.3%	*ADAMTS14*[β]	TFR - females	0.039	$3.40*10^{-9}$	0.023	$2.22*10^{-7}$	$2.89*10^{-14}$	++
chr10:96,009,182 - 96,039,597	rs11289753	43.2%	*PLCE1*[α]	AFR - combined	0.030	$2.41*10^{-12}$	0.021	$1.60*10^{-12}$	$1.22*10^{-22}$	++
chr11:825,110 - 825,777	rs1138714	43.2%	*AP006621.1*[α]	AFR - combined	-0.024	$5.47*10^{-8}$	-0.012	$2.50*10^{-5}$	$6.65*10^{-11}$	--
chr11:65,715,204 - 65,961,498	rs71455793	4.6%	*TSGA10IP*[δ]	TFR - females	-0.082	$4.45*10^{-9}$	-0.039	$1.85*10^{-5}$	$8.29*10^{-12}$	--
chr11:69,449,076 - 69,508,159	rs1789166	35.2%	*ORAOV1*[α]	AFR - combined	-0.027	$1.85*10^{-9}$	-0.013	$2.05*10^{-5}$	$4.92*10^{-12}$	--
chr12:50,468,643 - 51,220,897	rs11614785	34.1%	*LARP4*[α]	TFR - females	0.040	$1.37*10^{-10}$	0.026	$2.93*10^{-10}$	$1.15*10^{-18}$	++
chr13:91,991,743 - 92,024,715	rs6492538	22.1%	*GPC5*[α]	TFR - females	0.039	$3.16*10^{-8}$	0.021	$8.33*10^{-6}$	$1.11*10^{-11}$	++
chr14:50,923,249 - 51,154,000	rs71420186	6.6%	*MAP4K5*[α]	LFR - combined	-0.050	$8.25*10^{-9}$	-0.038	$3.20*10^{-11}$	$2.47*10^{-18}$	--
chr15:67,457,698	rs35874463	5.9%	*SMAD3*[δ]	TFR - females	0.069	$3.58*10^{-8}$	0.064	$1.44*10^{-14}$	$3.10*10^{-21}$	++
chr15:89,349,708 - 89,363,866	rs8026676	47.2%	*RP11-343B18.2*[β]	TFR - females	0.034	$7.24*10^{-9}$	0.020	$2.71*10^{-7}$	$6.37*10^{-14}$	++
chr15:100,692,953 - 100,692,953	rs72755233	11.3%	*ADAMTS17*[δ]	TFR - females	-0.074	$5.09*10^{-16}$	-0.080	$9.32*10^{-39}$	$4.91*10^{-53}$	--
chr16:31,075,175 - 31,107,689	rs8050894	37.7%	*KAT8*[β]	AFR - combined	-0.024	$7.02*10^{-8}$	-0.017	$4.76*10^{-9}$	$3.33*10^{-15}$	--
chr16:69,547,741 - 69,887,720	rs8057620	46.0%	*RP11-419C5.2*[β]	AFR - females	0.035	$2.27*10^{-9}$	0.024	$4.40*10^{-10}$	$1.79*10^{-17}$	++
chr16:90,062,323	rs10584116	9.7%	*C16orf3*[β]	TFR - females	-0.052	$8.99*10^{-8}$	-0.033	$6.97*10^{-7}$	$1.20*10^{-12}$	--

chr17:42,278,916 - 42,290,015	rs2071167	23.5%	*ASB16*[β]	LFR - combined	-0.028	$1.84*10^{-8}$	-0.023	$1.05*10^{-6}$	$1.58*10^{-10}$	--
chr17:47,357,969 - 47,454,515	rs28394864	46.2%	*ZNF652*[α]	TFR - females	-0.036	$8.89*10^{-10}$	-0.024	$1.70*10^{-9}$	$3.51*10^{-17}$	--
chr19:8,637,832 - 8,670,147	rs62621197	2.9%	*ADAMTS10*[δ]	TFR - females	-0.105	$2.17*10^{-9}$	-0.147	$3.45*10^{-39}$	$8.33*10^{-46}$	--
chr20:6,408,920	rs6085551	47.7%	*RP11-199O14.1*[α]	AFR - males	0.035	$2.18*10^{-8}$	0.018	$4.71*10^{-6}$	$7.77*10^{-6}$	++

Leading SNP denotes the strongest associated SNP at each locus. The Most likely target gene -column denotes a gene related to the associated locus either by proximity to the leading SNP(*α*), LD (R^2>0.8) with a leading eQTL SNP for the gene (*β*) or LD (R^2>0.8) with a missense variant within the gene (*δ*). Independent signals within the same locus are denoted by *. *β* - effect size estimate per allele. Meta analyses were performed with METAL[51]. Direction - summary of effect direction for the discovery and replication cohorts, with one '+' or '-' per cohort.

Table 2. Sex heterogeneous effects of body fat ratio-associated SNPs was assessed with GWAMA[56] for all replicated SNPs (N=97). P-values denote the results from tests for heterogeneity between sexes. Bonferroni correction was used to correct for multiple testing and P-values < 0.05/97 ($5.15*10^{-4}$) were considered significant.

SNP	Locus	Effect observed in females	Effect observed in males	P_{AFR}	P_{LFR}	P_{TFR}
rs55750792	*MFAP2*[γ]	+		-	-	$3.14*10^{-4}$
rs2273368	*WNT2B*[γ]	+		-	$1.59*10^{-7}$	$1.82*10^{-8}$
rs11205303	*SF3B4*[γ]	++	+	-	$1.61*10^{-5}$	$3.84*10^{-8}$
rs5779197	*TSEN15, GLT25D2*[γ]	+		-	$8.76*10^{-8}$	$2.48*10^{-8}$
rs991967	*TGFB2*[γ]	+		-	$1.83*10^{-4}$	$5.61*10^{-5}$
rs2820443	*LYPLAL1*[γ]	+		-	$3.00*10^{-6}$	$2.11*10^{-5}$
rs10916174	*ZNF678*[γ]	+		-	-	$6.24*10^{-8}$
rs754537	*RBJ, DNAJC27*[γ]	+		$1.58*10^{-14}$	-	-
rs3791679	*EFEMP1*[γ]	+		-	$5.76*10^{-7}$	$4.44*10^{-11}$
rs9853018	*ZBTB38*[γ]	+		-	$1.31*10^{-16}$	$2.29*10^{-26}$
rs4694504	*ADAMTS3*[γ,β]	++	+	-	-	$4.14*10^{-4}$
rs994014	*PRKG2*[γ]	+		-	$1.05*10^{-4}$	$8.29*10^{-6}$
rs7680661	*HHIP*[γ]	+		-	$5.36*10^{-10}$	$8.59*10^{-14}$
rs34341	*ANKDD1B*[γ]	+		$4.91*10^{-4}$	-	-
rs115912456	*VCAN*[γ]	+		-	$4.58*10^{-5}$	$2.23*10^{-4}$
rs3812049	*SLC12A2*[α]	+	++	$2.92*10^{-7}$	$4.65*10^{-4}$	-
rs41271299	*ID4*[γ]	+		-	$8.13*10^{-10}$	$3.18*10^{-13}$
rs9358913	*HIST1H2BE*[γ]	+		-	$1.23*10^{-13}$	$1.29*10^{-12}$
rs9469762	*HMGA1*[γ]	+		-	$3.67*10^{-6}$	-
rs2492863	*C6orf106*[α]	+		-	$3.83*10^{-8}$	-
rs6570507	*GPR126*[γ]	++	+	-	-	$6.10*10^{-5}$
rs798491	*AMZ1*[γ]	+		-	$2.90*10^{-4}$	$4.11*10^{-7}$
rs3823974	*ITGB8*[α]	+		-	$1.84*10^{-3}$	$1.73*10^{-4}$
rs481806	*JAZF1*[γ]	+		-	-	$2.48*10^{-6}$
rs527582137	*CDK6*[γ]	+		-	-	$1.60*10^{-4}$
rs35344761	*PCSK5*[γ]	+		-	$3.07*10^{-8}$	$5.70*10^{-9}$
rs11289753	*PLCE1*[α]		+	$6.94*10^{-5}$	-	-
rs552846225	*KDM2A*[γ]	+		-	$3.66*10^{-6}$	$1.71*10^{-8}$
rs11049566	*CCDC91*[γ]	+		-	$3.06*10^{-3}$	$5.13*10^{-4}$
rs11614785	*LARP4*[α]	+		-	$6.27*10^{-5}$	$4.01*10^{-6}$
rs12905253	*PML*[γ]	+		-	$6.56*10^{-4}$	$1.96*10^{-4}$
rs11856122	*ADAMTSL3*[γ]	+		0.52	$1.08*10^{-18}$	$2.46*10^{-19}$
rs28584580	*ACAN*[γ]	+		-	$2.96*10^{-11}$	$7.31*10^{-15}$
rs3817428	*ACAN*[γ]	+		-	$5.99*10^{-10}$	$3.02*10^{-9}$
rs72755233	*ADAMTS17*[γ]	+		-	$4.19*10^{-8}$	$8.87*10^{-11}$
rs4988781	*ADAMTS17*[γ]	+		-	-	$3.12*10^{-4}$
rs8050894	*KAT8*[γ]	+		$3.40*10^{-4}$	-	-
rs55872725	*FTO*[γ]	++	+	$6.96*10^{-6}$	-	-

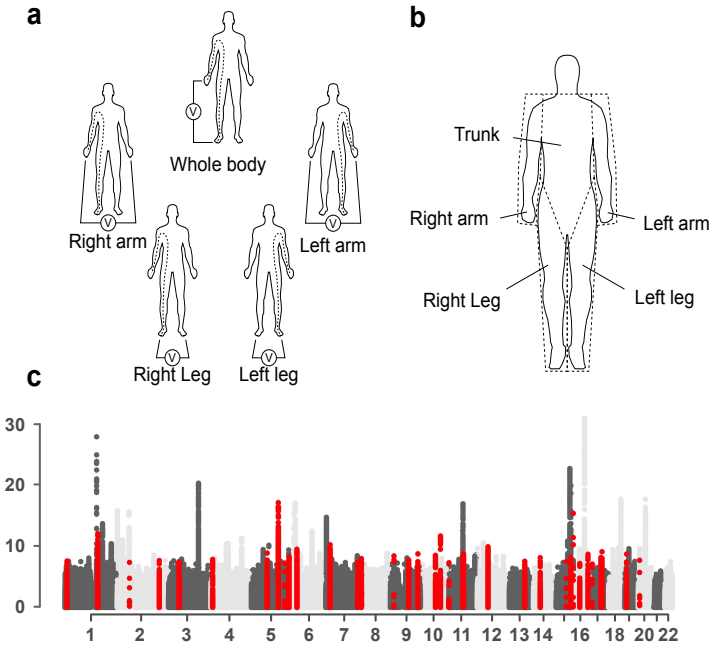
rs4800148	*RBBP8, CABLES1, C18orf45*[γ]	+	-	$3.76*10^{-10}$	$1.14*10^{-10}$
rs62621197	*ADAMTS10*[δ]	+	-	-	$3.54*10^{-11}$
rs2145270	*BMP2*[γ]	+	$2.87*10^{-4}$	-	-
rs143384	*GDF5*[γ]	+	-	$7.21*10^{-6}$	$4.40*10^{-11}$

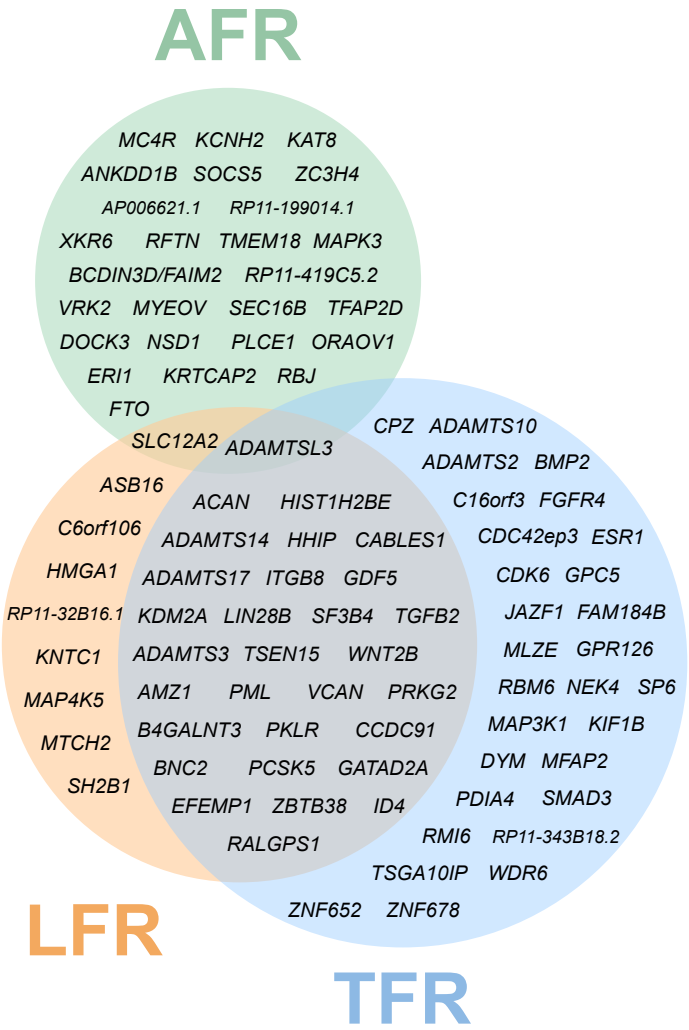
+ denotes an observed effect in one of the sexes, i.e. the SNP is associated ($p < 1e-7$) with one of the body fat ratios in that sex, ++ denotes that the effect was observed to be stronger in one sex when an effect was observed in both sexes. The effects were directionally consistent between sexes in all cases when SNP-effects could be observed in both sexes. The Locus column denotes a gene related to the associated locus either in annotation from previous GWAS(*γ*), LD (R^2>0.8) with a leading eQTL SNP for the gene (*β*) or LD (R^2>0.8) with a missense variant within the gene (*δ*).

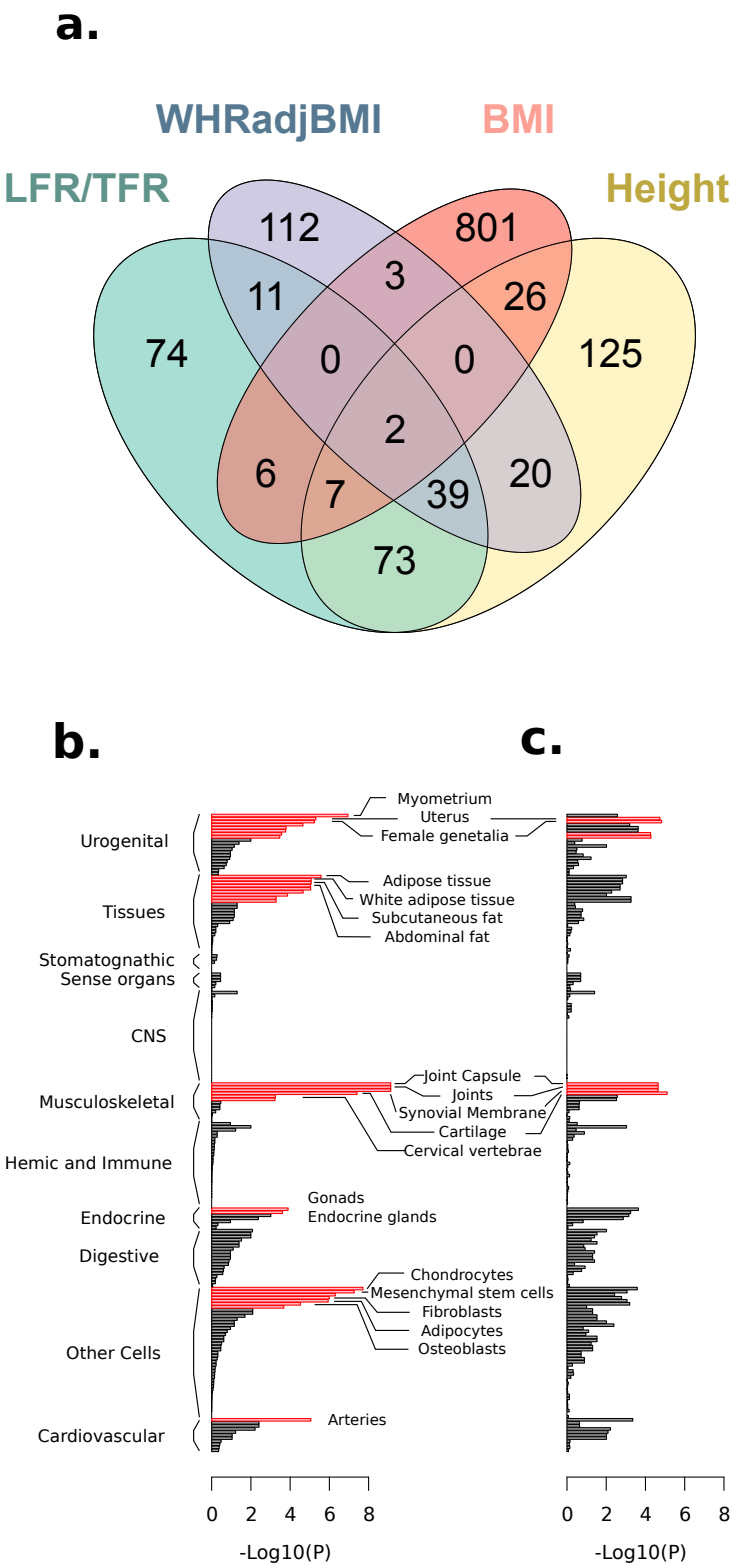
Table 3. Body fat ratio-associated potentially damaging missense variants.

Leading SNP	Chr	Position (bp)	Missense variant	Position (bp)	Distance to leading SNP (bp)	LD w. leading SNP (R2)	Gene	Amino acid substituion	SIFT* score	Polyphen-2* score
rs4733727	8	130,731,484	rs4144738	130,760,850	29,366	0.91	*GSDMC*	Met/Thr	deleterious	probably damaging
rs180921974	1	155,268,131	rs72704117	155,175,089	93,042	0.95	*THBS3*	Ala/Thr	deleterious	possibly damaging
rs3817428	15	89,415,247	leading SNP				*ACAN*	Asp/Glu	deleterious	benign - possibly damaging
rs72755233	15	100,692,953	leading SNP				*ADAMTS17*	Thr/Ile	deleterious	benign
rs115912456	5	82,815,158	rs61749613	82,815,170	12	1.00	*VCAN*	Lys/Glu	tolerated - deleterious	possibly damaging
rs9358913	6	26,239,404	rs7766641	26,184,102	55,302	0.82	*HIST1H2BE*	Gly/Ser	tolerated	probably damaging
rs351855	5	176,520,243	leading SNP				*FGFR4*	Gly/Arg	tolerated	possibly damaging
rs71420186	14	50,960,918	rs12881869	50,923,249	37,669	0.96	*MAP4K5*	Ala/Thr	tolerated	possibly damaging
rs5779197	1	184,009,826	rs2274432	184,020,945	11,119	0.96	*TSEN15*	Gly/Asp	tolerated	benign - possibly damaging
rs143384	20	34,025,756	rs224331	34,022,387	3,369	0.73	*GDF5; GDF5OS*	Ala/Ser; Ala/Glu	tolerated	benign - possibly damaging
rs62621197	19	8,670,147	leading SNP				*ADAMTS10*	Arg/Gln	tolerated	benign - possibly damaging

*Polyphen-2[32] and SIFT[31] scores were extracted from Ensembl (www.ensembl.org). SIFT utilizes protein sequence data to predict the effects of a missense variant while Polyphen-2 also considers protein structure and evolutionary conservation. Polyphen-2 generates a probability that a missense mutation is damaging, while SIFT generates a probability that the amino acid change is tolerated. Polyphen-2 uses a tree-tiered prediction score: 0-0.15 is predicted to be benign, 0.15 to 0.85 possibly damaging and 0.85 -1.0 more confidently predicted to be damaging (i.e., probably damaging). A SIFT score < 0.05 is predicted to be deleterious[31]. Polyphen-2 and SIFT scores are presented as ranges in the cases where they differ due to being part of different transcript variants.


a
Whole body
Right arm
Left arm
Right Leg
Left leg
b
Trunk
Right arm
Left arm
Right Leg
Left leg
c
30
20
10
0
1
2
3
4
5
6
7
8
9
10
11
12
13
14
16
18
20
22


AFR
MC4R KCNH2 KAT8
ANKDD1B SOCS5 ZC3H4
AP006621.1 RP11-199014.1
XKR6 RFTN TMEM18 MAPK3
BCDIN3D/FAIM2 RP11-419C5.2
VRK2 MYEOV SEC16B TFAP2D
DOCK3 NSD1 PLCE1 ORAOV1
ERI1 KRTCAP2 RBJ
FTO
SLC12A2
ADAMTSL3
ASB16
ACAN HIST1H2BE
C6orf106
ADAMTS14 HHIP CABLES1
HMGA1
ADAMTS17 ITGB8 GDF5
RP11-32B16.1
KDM2A LIN28B SF3B4 TGFB2
KNTC1
ADAMTS3 TSEN15 WNT2B
MAP4K5
AMZ1 PML VCAN PRKG2
MTCH2
B4GALNT3 PKLR CCDC91
SH2B1
BNC2 PCSK5 GATAD2A
EFEMP1 ZBTB38 ID4
RALGPS1
CPZ ADAMTS10
ADAMTS2 BMP2
C16orf3 FGFR4
CDC42ep3 ESR1
CDK6 GPC5
JAZF1 FAM184B
MLZE GPR126
RBM6 NEK4 SP6
MAP3K1 KIF1B
DYM MFAP2
PDIA4 SMAD3
RMI6 RP11-343B18.2
TSGA10IP WDR6
ZNF652 ZNF678
LFR
TFR


a.
WHRadjBMI
BMI
LFR/TFR
Height
112
801
3
11
26
74
0
0
125
2
6
7
39
20
73
b.
c.
Urogenital
Tissues
Stomatognathic
Sense organs
CNS
Musculoskeletal
Hemic and Immune
Endocrine
Digestive
Other Cells
Cardiovascular
Myometrium
Uterus
Female genetalia
Adipose tissue
White adipose tissue
Subcutaneous fat
Abdominal fat
Joint Capsule
Joints
Synovial Membrane
Cartilage
Cervical vertebrae
Gonads
Endocrine glands
Chondrocytes
Mesenchymal stem cells
Fibroblasts
Adipocytes
Osteoblasts
Arteries
0 2 4 6 8
-Log10(P)
0 2 4 6 8
-Log10(P)