Combining evidence from human genetic and functional screens to identify pathways altering obesity and fat distribution

Nikolas A. Baya[1,2,*], Ilknur Sur Erdem[3,4*], Samvida S. Venkatesh[1,2], Saskia Reibe[1,5], Philip D. Charles[1], Elena Navarro-Guerrero[6], Barney Hill[1,5], Frederik Heymann Lassen[1,2], Melina Claussnitzer[7-10], Duncan S. Palmer[1,5,7,#], Cecilia M. Lindgren[1,2,3,7,#]

1. Big Data Institute, Li Ka Shing Centre for Health Information and Discovery, University of Oxford, Oxford OX3 7LF, United Kingdom.
2. Centre for Human Genetics, Nuffield Department of Medicine, University of Oxford, Oxford OX3 7BN, United Kingdom.
3. Nuffield Department of Women's and Reproductive Health, Medical Sciences Division, University of Oxford, United Kingdom.
4. Chinese Academy for Medical Sciences Oxford Institute, Nuffield Department of Medicine, University of Oxford, Oxford OX3 7FZ, United Kingdom.
5. Nuffield Department of Population Health, Medical Sciences Division, University of Oxford, Oxford, United Kingdom.
6. Target Discovery Institute, Nuffield Department of Medicine, University of Oxford, Oxford OX3 7FZ, United Kingdom.
7. Broad Institute of MIT and Harvard, Cambridge, Massachusetts, United States of America.
8. Broad Institute of MIT and Harvard, Novo Nordisk Foundation Center for Genomic Mechanisms of Disease & Type 2 Diabetes Systems Genomics Initiative, Cambridge, Massachusetts, United States of America.
9. Center for Genomic Medicine and Endocrine Division, Massachusetts General Hospital, Boston, Massachusetts, United States of America.
10. Harvard Medical School, Harvard University, Boston, Massachusetts, United States of America.

*, # shared authorship

Abstract

Overall adiposity and body fat distribution are heritable traits associated with altered risk of cardiometabolic disease and mortality. Performing rare variant (minor allele frequency<1%) association testing using exome-sequencing data from 402,375 participants in the UK Biobank (UKB) for nine overall and tissue-specific fat distribution traits, we identified 19 genes where putatively damaging rare variation associated with at least one trait (Bonferroni-adjusted $P<1.58\times10^{-7}$) and 52 additional genes at FDR≤1% ($P\leq4.37\times10^{-5}$). These 71 genes exhibited higher ($P=3.58\times10^{-18}$) common variant prioritisation scores than genes not significantly enriched for rare putatively damaging variation, with evidence of monotonic allelic series (dose-response relationships) among ultra-rare variants (minor allele count≤10) in 22 genes. Five of the 71 genes have cognate protein UKB Olink data available; all five associated ($P<3.80\times10^{-6}$) with three or more analysed traits. Combining rare and common variation evidence, allelic series and proteomics, we selected 17 genes for CRISPR knockout in human white adipose tissue cell lines. In three previously uncharacterised target genes, knockout increased (two-sided *t*-test $P<0.05$) lipid accumulation, a cellular phenotype relevant for fat mass traits, compared to Cas9-empty negative controls: *COL5A3* (fold change [FC]=1.72, $P=0.0028$), *EXOC7* (FC=1.35, $P=0.0096$), and *TRIP10* (FC=1.39, $P=0.0157$); furthermore, knockout of *SLTM* resulted in reduced lipid accumulation (FC=0.51, $P=1.91\times10^{-4}$). Integrating across population-based genetic and *in vitro* functional evidence, we highlight therapeutic avenues for altering obesity and body fat distribution by modulating lipid accumulation.

Introduction

One in four adults globally are either overweight or obese[1,2]. While higher overall adiposity increases morbidity and mortality,[1,3] disease risk is also informed by the location and distribution of excess fat within particular depots[4,5]. Abnormal distribution of fat is often attributed to lipodystrophy syndromes, which can cause generalised or selective fat mass loss and depot-specific fat growth[6]. Independent of overall body mass index (BMI), individuals with higher central adiposity have increased risk of cardiometabolic diseases, including type 2 diabetes (T2D) and stroke[7,8]; in contrast, individuals with higher hip circumference, an indicator for gluteal adiposity, have lower risk of such outcomes. For example, a standard deviation (SD) increase in hip circumference has been shown to reduce risk of T2D by ~40%[9] or myocardial infarction by ~10%[10]. Previous studies indicate that fat distribution, as assessed by waist-to-hip ratio (WHR), has a strong heritable component independent of BMI, with narrow sense heritability of up to 56% in women and 32% in men[8,11].

BMI-associated genes are enriched in tissues of the central nervous system, notably the hypothalamus which is involved in appetite regulation[12]. Indeed, blockbuster GLP1-receptor agonists that are prescribed for weight loss[13,14] act primarily through the dorsomedial hypothalamus to control food intake[15]. In contrast, genome-wide association studies (GWASs) for WHR adjusted for BMI (WHRadjBMI) indicate enrichment of genes associated with fat

distribution in adipose tissue[16]. However, there are currently few therapeutic avenues to modify both obesity and fat distribution[17].

Understanding the genetic aetiology of body fat and fat distribution may drive new therapies for obesity. Mendelian genetic studies have identified rare variants in genes such as *PPARG*, a master regulator of adipocyte differentiation, and *INSR*, the insulin receptor, associated with lipodystrophies and extreme forms of central body fat distribution[18,19]. In the general population, rare protein-coding variants with large effects can point to genetic and molecular mechanisms underpinning fat distribution. For example, previous work has demonstrated that rare loss-of-function alleles in *GPR75* are protective against obesity[20], and protein-truncating variants in *INHBE* are associated with favourable fat distribution[21].

Here, we integrated exome-sequencing (ES) and common variant GWASs for nine obesity-related and fat distribution traits to nominate genes associated with overall obesity or central adiposity in up to 402,375 participants in UKB. To understand the biological pathways by which these genes may act, we assessed the proteome-wide consequences of these genes and tested 1,463 plasma proteins for associations with body fat distribution in a subset of 42,828 individuals in UKB. Finally, we designed *in vitro* functional assays for lipid accumulation and glucose uptake in human white adipocytes to biologically validate the identified genes using CRISPR-Cas9 knockout (KO) experiments. Taken together, we demonstrate that converging multi-modal evidence from GWAS, ES, proteomics, and *in vitro* KO studies can generate therapeutic targets for obesity or lipodystrophy.

Results

Convergence of rare and common variant evidence for genes associated with obesity and fat distribution

We performed rare variant (minor allele frequency [MAF] <1%) and gene-level testing across 1,827,504 variants in 18,788 genes for association with nine obesity or fat distribution related traits, including overall obesity (BMI and body fat percentage), central fat distribution (WHRadjBMI), and tissue-specific fat component phenotypes derived from dual energy X-ray absorptiometry (DXA) and magnetic resonance imaging (MRI) scans (such as android tissue fat percentage, abdominal fat ratio, and visceral adipose tissue volume), in up to 402,375 participants of European ancestry in the UKB (Supp. Tables 1 and 2). We defined damaging rare variants as those annotated to be high-confidence predicted loss-of-function (pLoF) or damaging missense using a combination of existing effect prediction tools (LOFTEE[22], CADD[23], REVEL[24], and SpliceAI[25]; Methods). To account for potential confounding due to nearby common variants, we conditioned gene-level associations on fine-mapped GWAS loci on the same chromosome (Methods).

We use the optimal sequence kernel association test (SKAT-O) when reporting *P*-values for all

gene-level obesity and fat distribution trait associations. This approach unifies the burden test, which is more powerful when a high density of variants in a gene are causal and have the same effect direction, and the sequence kernel association test, which is more powerful when there is a lower density of causal variants in a gene or when the causal variants have mixed effect directions. However, SKAT-O does not provide effect size estimates, thus all gene-level effect sizes are estimated by burden testing.

We identified 19 unique genes carrying rare damaging variation associated with BMI, body fat percentage, or WHRadjBMI at exome-wide significance (SKAT-O $P<1.58\times10^{-7}$, Bonferroni adjustment for 315,996 unique tests) (Table 1 and Figure 1a). Variant-level testing identified 12 rare missense variants significantly associated ($P<5\times10^{-8}$, canonical single-variant genome-wide Bonferroni adjustment for one million independent genomic regions) with WHRadjBMI (8 variants), BMI (3), or body fat percentage (1) (Supp. Table 3). These variants were located in nine genes, of which three (*MC4R, PDE3B, PLIN1*) also carried a significant burden of rare damaging variation. Two genes (*PDE3B* and *PLIN1*) harboured more than one significantly associated rare missense variant (Supp. Table 3). We found evidence of monotonic allelic series, that is, increasingly large effects of increasingly damaging variants in a dose-response relationship, among ultra-rare variants (minor allele count (MAC)≤10) in *COL5A3, DIDO1, INSR, PLIN1, PTPRG, PPARG* and *SLC12A5* (Figure 1b and Supp. Figure 3).

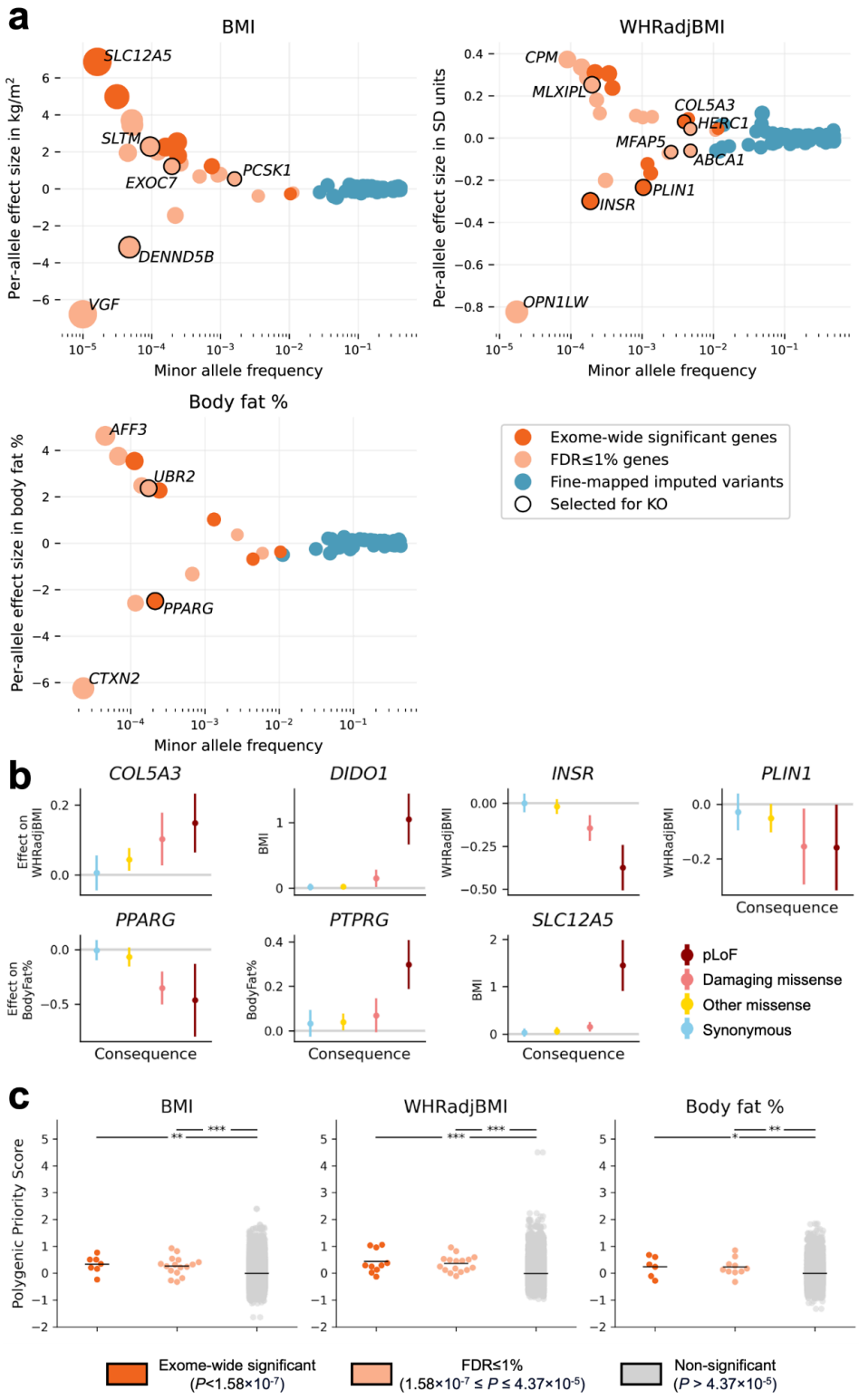

Figure 1. Consequences of rare missense variation on obesity or fat distribution related traits in UKB. a, Gene- and variant-level effects as a function of aggregated minor allele frequency. Only gene-level results which are exome-wide or FDR≤1% significant are shown. Only fine-mapped common (MAF>1%) variants with posterior inclusion probability ≥0.9 are shown. Effect sizes for BMI and body fat

percentage are converted to the kg/m^2 and fat percentage scale, respectively, by multiplying the effect size in SD units by the empirical SD of each trait (4.75 kg/m^2 for BMI, 8.51% for body fat percentage). Genes selected for KO are outlined in black and labeled. Genes with highest and lowest effect size for each trait are also labeled. Point sizes are scaled by effect size. **b,** Allelic series for exome-wide significant genes with monotonic relationships between effect size and consequence severity in ultra-rare (MAC≤10) variant burden tests. Confidence intervals for effect size defined as +/-1.96 standard errors. **c,** Enrichment of Polygenic Priority Scores (PoPS)[26] among exome-wide and FDR≤1% significant genes compared to other genes. Significance of 1-sided *t*-test with alternate hypothesis that a significant gene category has higher average PoPS than the non-significant gene category: ***P<0.001, **P<0.01, *P<0.05. The mean PoPS for each gene category are indicated by black bars.

Gene (genomic location)	Phenotype	Consequence	Variants in mask	Total MAF (MAC)	SKAT-O *P*-value	Burden *P*-value	Effect size [95% CI]	Effect size, unweighted
ANKRD12 (18:9136228)	WHRadjBMI	pLoF	127	0.00038 (308)	1.71×10^{-9}	6.67×10^{-8}	0.010 [0.006, 0.013]	0.238
APBA1 (9:69427532)	BMI	pLoF	59	0.00016 (128)	6.25×10^{-8}	5.18×10^{-8}	0.019 [0.012, 0.026]	0.475
BLTP1 (4:122152331)	BMI	pLoF	282	0.00075 (600)	2.15×10^{-10}	9.19×10^{-11}	0.011 [0.007, 0.014]	0.258
COL5A3 (19:9959561)	WHRadjBMI	pLoF plus damaging missense	362	0.00388 (3,111)	5.23×10^{-10}	2.98×10^{-9}	0.003 [0.002, 0.004]	0.078
DIDO1 (20:62877738)	BMI	pLoF	22	0.00003 (25)	1.14×10^{-7}	1.14×10^{-7}	0.042 [0.026, 0.057]	1.050
GPR151 (5:146513144)	Body fat %	pLoF	37	0.01039 (8,209)	6.60×10^{-8}	2.51×10^{-7}	-0.002 [-0.003, -0.001]	-0.043
	BMI	pLoF	37	0.01036 (8,313)	1.35×10^{-7}	2.02×10^{-7}	-0.003 [-0.004, -0.002]	-0.058
INHBE (12:57452323)	WHRadjBMI	pLoF	23	0.00119 (954)	3.27×10^{-9}	1.37×10^{-7}	-0.005 [-0.007, -0.003]	-0.123
INSR (19:7112255)	WHRadjBMI	pLoF	55	0.00019 (151)	6.38×10^{-8}	4.20×10^{-7}	-0.012 [-0.017, -0.007]	-0.299
KEAP1 (19:10486125)	WHRadjBMI	pLoF plus damaging missense	88	0.00034 (273)	1.76×10^{-11}	2.52×10^{-12}	0.012 [0.009, 0.016]	0.305
MC4R (18:60371062)	BMI	pLoF	23	0.00023 (187)	9.30×10^{-13}	1.87×10^{-13}	0.021 [0.016, 0.027]	0.532
PDE3B (11:14643804)	WHRadjBMI	pLoF	68	0.00132 (1,057)	5.91×10^{-10}	7.09×10^{-10}	-0.007 [-0.009, -0.005]	-0.166
	Body fat %	pLoF	66	0.00132 (1,046)	1.34×10^{-7}	2.09×10^{-7}	0.005 [0.003, 0.007]	0.102
PKD1 (16:2088708)	WHRadjBMI	pLoF	97	0.00022 (176)	7.68×10^{-11}	8.50×10^{-7}	0.012 [0.007, 0.017]	0.311
PLD1 (3:171600404)	WHRadjBMI	pLoF plus damaging	240	0.01163 (9,324)	4.69×10^{-10}	2.88×10^{-9}	0.002 [0.001, 0.003]	0.044

		missense						
PLIN1 (15:89664367)	WHRadjBMI	pLoF	48	0.00104 (836)	8.94×10^{-23}	8.98×10^{-21}	-0.009 [-0.011, -0.007]	-0.233
PLIN4 (19:4502192)	WHRadjBMI	pLoF	121	0.00444 (3,560)	1.00×10^{-13}	3.25×10^{-14}	0.004 [0.003, 0.005]	0.091
	Body fat %	pLoF	120	0.00444 (3,506)	1.23×10^{-10}	2.11×10^{-10}	-0.003 [-0.004, -0.002]	-0.080
PPARG (3:12287368)	Body fat %	pLoF plus damaging missense	66	0.00021 (168)	2.90×10^{-8}	4.34×10^{-7}	-0.012 [-0.016, -0.007]	-0.292
PTPRG (3:61561569)	BMI	pLoF	120	0.00024 (194)	1.37×10^{-8}	8.31×10^{-8}	0.015 [0.01, 0.021]	0.377
	Body fat %	pLoF	120	0.00024 (192)	1.35×10^{-7}	8.25×10^{-7}	0.011 [0.006, 0.015]	0.267
RIF1 (2:151409883)	Body fat %	pLoF	66	0.00011 (89)	1.41×10^{-7}	1.41×10^{-7}	0.017 [0.01, 0.023]	0.416
SLC12A5 (20:46021690)	BMI	pLoF	13	0.00002 (13)	1.21×10^{-7}	1.21×10^{-7}	0.058 [0.036, 0.079]	1.446

Table 1. Exome-wide significant gene-level associations. Gene-level exome-wide significance threshold: SKAT-O $P<1.58\times10^{-7}$. Genomic location indicates the chromosome and base pair coordinates for the start of the gene in Genome Reference Consortium Human Build 38. Variants in the mask are those which are included in the SKAT-O test for a given gene. They include all variants in the gene with MAF<1% and consequence annotation matching the specific mask. Total MAF is the sum of MAF for all variants in the mask. In parentheses, MAC is the total minor allele count for the gene among individuals included in the association test. The effect size of a gene is estimated using burden testing and is in units of the phenotype's SD. The 95% confidence interval is defined by an interval centred on the effect size point estimate, +/-1.96 standard errors. The unweighted effect size is a version of the burden effect size that does not weight variants by Beta(MAF; 1, 25), the default weighting used by SAIGE. Using the unweighted effect sizes puts the gene-level effect on the same scale as variant-level effects. See Supp. Note 1 for calculation of unweighted effect size. WHRadjBMI, waist-to-hip ratio adjusted for BMI; BMI, body mass index; pLoF, predicted loss-of-function.

Genes associated with obesity and fat distribution in rare variant gene-level tests overlapped with those highlighted by common variant associations, demonstrated by a significant enrichment of the Polygenic Priority Score (PoPS)[26] (one-sided *t*-test $P<0.05$, Figure 1c) among genes harbouring exome-wide significant damaging rare variant association signals. Moreover, 52 rare-variant gene-level associations which do not reach exome-wide significance but pass a less stringent significance threshold (SKAT-O $P\leq4.37\times10^{-5}$, using the Benjamini-Hochberg procedure[27] for FDR≤1% on minimum SKAT-O *P*-value across traits and two variant masks; Methods) are also significantly enriched for PoPS scores (one-sided *t*-test $P<0.01$, Figure 1c), indicating value in examining these additional genes as potential therapeutic targets for obesity and lipodystrophy (Supp. Table 4). Five additional gene-level associations also passed the FDR≤1% significance threshold, but were supported by minor allele counts less than five (Supp. Note 2).

Of the 31 genes known to cause severe monogenic obesity or lipodystrophy[6,28], three (*PPARG*, *PLIN1*, and *MC4R*) are associated with an obesity or fat distribution trait at the exome-wide level ($P<1.58\times10^{-7}$) in our rare-variant gene-level association testing, representing a 125-fold enrichment over random chance (Fisher's exact test $P=3.87\times10^{-6}$). Of the 71 FDR-significant genes, 23 have previously been reported to have rare variant gene-level associations with traits related to obesity or fat distribution in UKB participants, across a variety of variant annotation and gene-level association testing methods (Supp. Table 11)[20,21,29,30].

Although we did not find any genes associated with phenotypes derived from tissue-specific fat components at exome-wide significance (Bonferroni $P<1.58\times10^{-7}$), nine gene-level associations for these tissue-specific fat component phenotypes were significant at the FDR≤1% threshold ($P\leq4.37\times10^{-5}$), including six genes associated with the ratio of android to gynoid tissue fat percentage, two with visceral adipose tissue volume and one with android fat percentage (Supp. Table 4).

To evaluate the phenome-wide effects of genes related to obesity and fat distribution, we scanned 4,529 phenotypes in Genebass[31] summary statistics for associations with the 71 FDR-significant genes. We found 549 significant associations (Genebass SKAT-O $P\leq9.69\times10^{-6}$, FDR≤1%) across 211 traits for 42/71 genes (Supp. Figure 5, Supp. Table 5). The most significant Genebass associations were among blood biochemistry lipid phenotypes (Supp. Figure 5a, Supp. Table 5). We also observed shared associations across FDR-significant genes for phenotypes significantly genetically correlated to BMI and body fat percentage, such as fat mass and distribution traits. For example, eight FDR-significant genes (*BLTP1, DIDO1, GPR151, MC4R, PDE3B, PLIN4, RIF1, UBR2*) are associated with arm fat mass, leg fat mass, arm fat percentage, whole body fat mass, leg fat percentage, and hip circumference (all traits have common variant LD Score regression[32] $r_g>0.80$, $P<1\times10^{-300}$ with BMI and body fat percentage, Supp. Table 5)[33]. We see some evidence for pleiotropy, as defined by associations with phenotypes genetically uncorrelated with BMI or body fat percentage, among results with common variant genetic correlations available: *PDE3B* and *STAB1* are associated with platelet distribution width (BMI $r_g=0.025$, $P=0.274$; body fat percentage $r_g=9.7\times10^{-4}$, $P=0.958$) and *ANKRD12* is associated with neutrophil percentage (BMI $r_g=-0.029$, $P=0.117$; body fat percentage $r_g=-8.8\times10^{-3}$, $P=0.639$) (Supp. Table 5).

Sex- and age-specific effects of genes associated with obesity and fat distribution

We performed all association tests in sex-specific strata, identifying three genes (*INSR, PDE3B, PLIN4*) with significant differences (sex-difference[34] $P<2.67\times10^{-6}$, Bonferroni adjusted for 18,737 genes tested for sex-differential effects) in burden effect sizes. We observed opposing sex-specific effects for *INSR* on WHRadjBMI (female burden beta (SE)=-0.00458 (9.44×10^{-4}), male beta=0.00195 (0.00100), sex-difference $P=1.66\times10^{-6}$). We found that female-specific effects may be driving the reported sex-combined associations between WHRadjBMI and *PDE3B* (female beta=-0.0169 (0.00169), male beta=-0.00288 (0.00179), sex-difference $P=8.02\times10^{-9}$) and *PLIN4* (female beta=0.00822 (9.15×10^{-4}), male beta=0.00108 (9.86×10^{-4}),

sex-difference P=8.19×10^{-8}) (Supp. Figure 6a, Supp. Table 6). Within the sex-specific analysis, we find two genes with female-specific associations (P<2.67×10^{-6}, Bonferroni adjusted for 18,737 genes tested for sex-specific effects) of pLoF variation on obesity and fat distribution (*CASQ1* with body fat percentage, female SKAT-O P=1.12×10^{-6}; *PLXND1* with WHRadjBMI, female P=9.69×10^{-10}), and two genes with male-specific associations, both for the ratio of android to gynoid tissue fat percentage (*ZNF841* male P=1.04×10^{-6}; *TINF2* male P=4.84×10^{-7}) (Supp. Figure 6b, Supp. Table 6), which do not reach significance at the FDR≤1% threshold (P≤4.37×10^{-5}) in sex-combined analyses.

We leveraged the age at diagnosis of obesity in longitudinal health records linked to UKB participants to identify that a burden of rare missense variation (MAF<1%) in five genes was associated (P<2.35×10^{-4}, Bonferroni adjusted for 213 gene-consequence mask pairs) with elevated lifetime risk of obesity (Supp. Figure 7, Supp. Table 7). Individuals carrying rare missense variants in *MC4R* (Cox proportional hazard ratio [HR] (SE)=1.46 (0.0843), P=7.75×10^{-6}), men with pLoF variants in *SLTM* (HR=5.37 (0.354), P=2.05×10^{-6}) and damaging missense variants in *PCSK1* (HR=1.86 (0.165), P=1.63×10^{-4}), and women with pLoF variants in *DIDO1* (HR=11.2 (0.578), P=2.85×10^{-5}) and *SLC12A5* (HR=14.8 (0.708), P=1.45×10^{-4}) were at risk of earlier age at onset of obesity (Supp. Figure 7, Supp. Table 7).

Proteome-wide consequences of obesity and fat distribution traits and associated genes

We assessed the relationship between nine obesity and fat distribution traits (Supp. Figure 8, Supp. Table 2) and 1,463 plasma proteins profiled in a subset of 42,828 individuals in the UKB Olink Plasma Proteomics Project[35]. To do this, we regressed the obesity and fat distribution traits on each protein level independently, and adjusted for *sex*, *age* and *age*^{2}, *age-by-sex* interactions, UKB assessment centre, and the first 21 genetic principal components (PCs).

We found that 1,345 proteins (92% of the proteins tested) were significantly associated with at least one obesity or fat distribution trait (P<3.80×10^{-6}, Bonferroni adjustment for 13,167 tests, Figure 2, Supp. Table 8, Supp. Figures 8 and 9), adding to previous reports that obesity is accompanied by a sweeping alteration of the entire plasma proteome[36,37]. 407 proteins were significantly associated with all nine body fat traits (Supp. Figure 9), of which 331 (81%) were positively correlated with all traits, and 71 (17%) were negatively correlated with all traits.

To identify significantly enriched pathways among associated proteins, we applied GSEAPreranked[38] to a list of proteins ordered by the *z*-score of the most significant association across the nine traits. 27 pathways were enriched (FDR<25% GSEA default threshold) among proteins whose levels were positively correlated with obesity or fat distribution traits, including: organic acid binding, known to play a role in metabolic disease[39] (42 proteins, FDR=7.2%, FWER P=0.0671), interleukin-10 signalling, linked to insulin resistance[40] (30 proteins, FDR=9.0%, FWER P=0.293), and cellular response to salt, possibly reflecting the association between dietary sodium intake and obesity[41] (20 proteins, FDR=9.1%, FWER P=0.161) (Supp.

Table 9). We did not find any pathways markedly enriched (FDR<25%) among proteins negatively correlated with obesity or fat distribution traits.

Across the proteome, only five proteins (CCN5, PTPRF, SLITRK2, THY1, and XG) displayed heterogeneity in the direction of their effects on obesity and fat distribution traits, all of which were positively associated with BMI, body fat percentage, and all tissue-specific fat components, but negatively associated with WHRadjBMI (all assocations $P<3.80\times10^{-6}$, Supp. Figure 10 and Supp. Table 8). Overexpression of *CCN5* improves insulin sensitivity in obese mice[42], and overexpression of *THY1* activity blocks adipocyte formation through inhibition of PPARG in mouse cell lines[43]. These results indicate potential therapeutic targets to alter fat distribution independent of overall obesity.

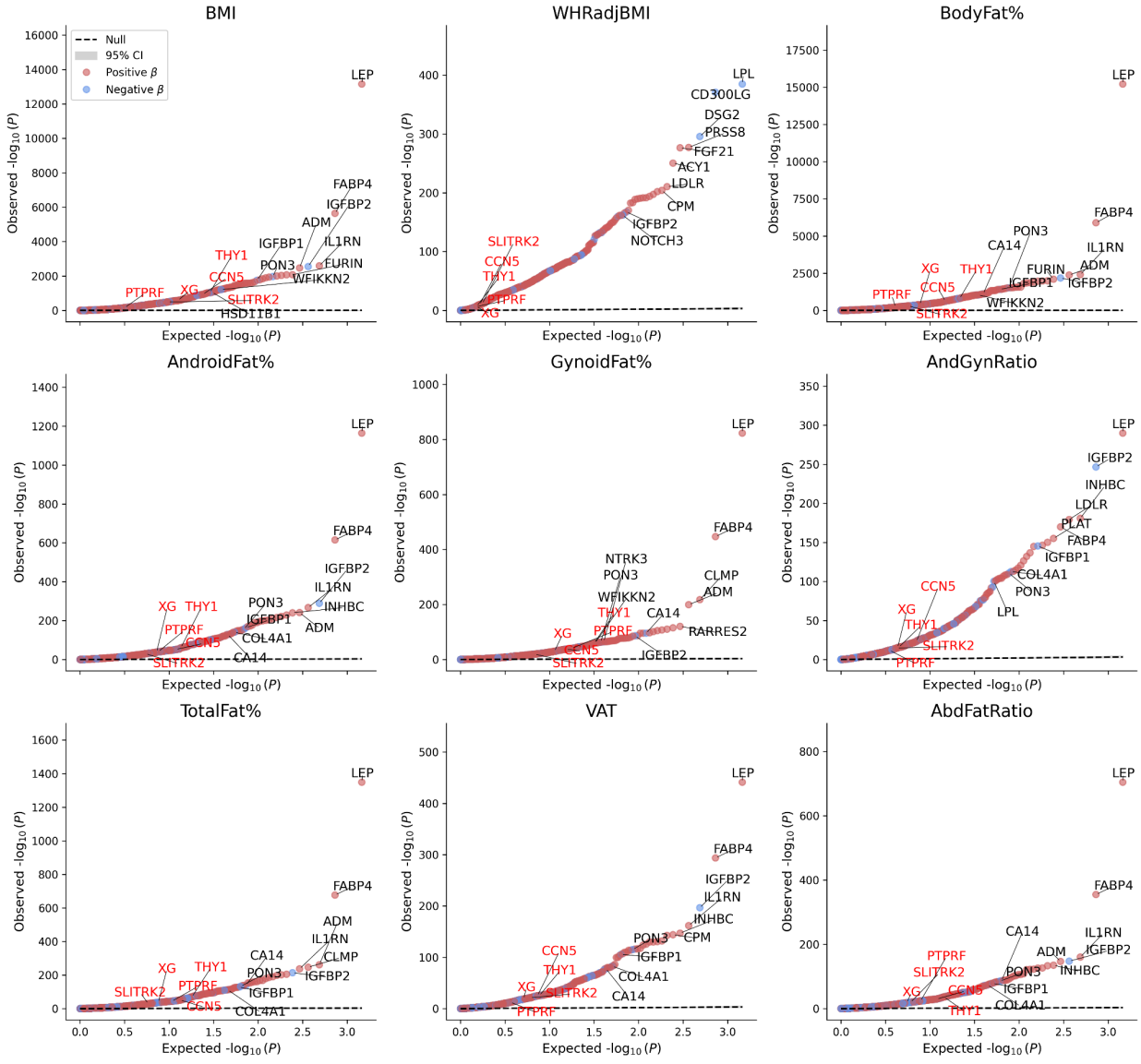

Figure 2. Relationship of the plasma proteome with obesity and fat distribution related traits in UKB. Quantile-quantile (QQ) plots of the observed *P*-value against expected *P*-value (on the log scale)

for proteome-wide associations with each trait. A red point indicates a protein has a positive effect on the trait, a blue point indicates a negative effect. The null expectation (observed equal to expected) and 95% CI for the null are indicated by the black dashed line and grey shading. The ten most significant proteins are labelled in black text. The five proteins with mixed direction of effect across traits are labelled in red text. BMI, body mass index; WHRadjBMI, waist-to-hip ratio adjusted for BMI, BodyFat%, body fat percentage; AndroidFat%, android tissue fat percentage; GynoidFat%, gynoid tissue fat percentage, AndGynRatio, ratio of android to gynoid fat percentage; TotalFat%, total body fat percentage; VAT, visceral adipose tissue volume; AbdFatRatio, abdominal fat ratio.

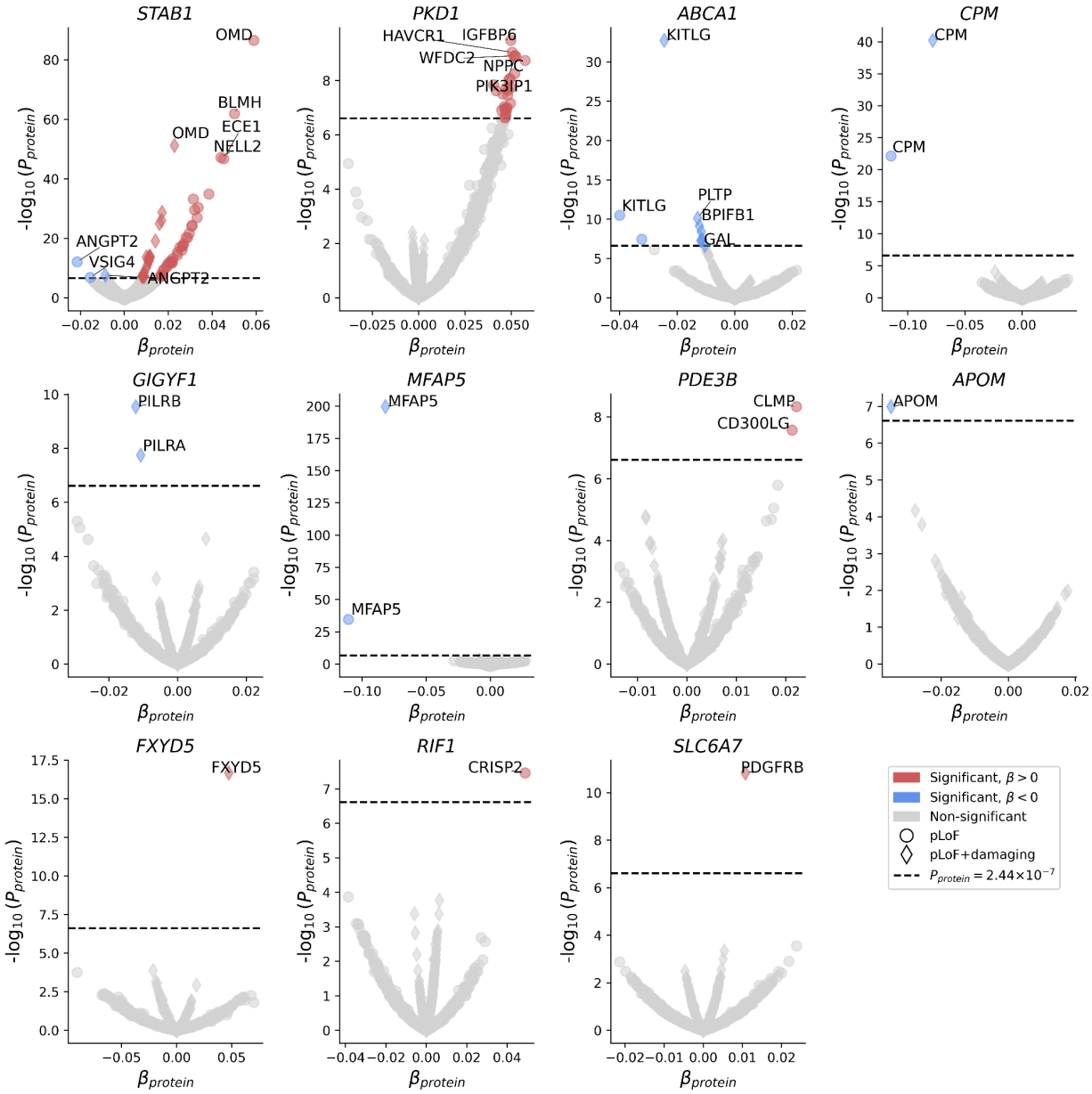

Figure 3. Relationship of the plasma proteome with genes associated with obesity and fat distribution in UKB. Volcano plots of gene burden effect on the proteome for 11 genes significantly associated with obesity and fat distribution traits (SKAT-O $P \leq 4.37 \times 10^{-5}$). Red and blue points indicate proteins with significant gene-level associations (burden $P < 2.44 \times 10^{-7}$, Bonferroni adjustment for 204,743 unique tests; significance threshold indicated by dashed line). Red points indicate positive gene burden

effect and blue points indicate negative gene burden effect. Grey points indicate non-significant associations. The consequence mask of each result is indicated by the shape of the point: Circles are the result of a pLoF-only burden mask, diamonds are the results of a pLoF and damaging missense mask. Effect size is in units of protein standard deviations.

As proteins can link genotype to phenotype, we assessed the proteomic consequences of rare damaging variation in the 71 genes associated with obesity that were identified in the previous section. We performed this gene burden scan for 1,456 proteins in the UKB Olink proteomics data (Methods). Of the five genes which were FDR≤1% significant and had corresponding measurements of cognate protein levels, three (*APOM*, *CPM*, and *MFAP5*) show the expected direction of effect, i.e. a burden of rare damaging variation in the gene significantly reduces protein levels (burden $P<2.44\times10^{-7}$, Bonferroni adjustment for 204,743 unique tests, Figure 3 and Supp. Table 10). However, damaging variation in *FXYD5* is associated with significantly higher levels of plasma FXYD5 (burden beta (SE)=0.0474 (0.00559), $P=2.14\times10^{-17}$). Association of damaging variation in *PLIN1* with plasma PLIN1 did not reach Bonferonni-adjusted significance.

Seven obesity or fat distribution associated genes were significantly associated (burden $P<2.31\times10^{-7}$) with at least one non-cognate plasma protein (Figure 3 and Supp. Table 10). The burden of pLoF variation in *PDE3B*, associated with lower WHRadjBMI (beta (SE)= -0.00676 (0.001096), burden $P= 7.09\times10^{-10}$), is also associated with increased levels of CD300LG (beta (SE)=0.0213 (0.00382), burden $P=2.67\times10^{-8}$) and CLMP (beta (SE)=0.0221 (0.00378), burden $P=4.63\times10^{-9}$). CLMP is an adipocyte-adhesion molecule associated with adipocyte differentiation, and its overexpression reduces weight gain in mice[44]. We also found an association between damaging variation in *GIGYF1*, for which burden of pLoF variation is associated with increased WHRadjBMI (beta (SE) = 0.01145 (0.00243), burden $P=2.60\times10^{-6}$), and lower plasma levels of the paired receptors PILRA (beta (SE)= -0.0108 (0.00191), burden $P=1.79\times10^{-8}$) and PILRB (beta (SE)= -0.0122 (0.00194), burden $P=2.85\times10^{-10}$). PILRA drives macrophage infiltration into adipose and liver tissue and induces obesity when knocked out in mice[45]. Finally, we found multiple plasma proteins significantly associated with damaging variation in obesity and fat distribution genes *ABCA1* (8 proteins), *PKD1* (26), and *STAB1* (37) (Figure 3 and Supp. Table 10). The latter encodes a scavenger receptor and is a known pleiotropic locus[46].

Nominating target genes for *in vitro* functional characterisation through CRISPR knockout

As adipose tissue is enriched for the expression of fat distribution genes identified from common variant GWASs[16], we assessed the mRNA counts of overall and tissue-specific fat distribution-associated genes from our exome-wide gene-level association testing in differentiated human white adipose tissue derived pre-adipocytes (hWAT) *in vitro* (Figure 4a). Genes with FDR-significant results in gene-level tests had higher expression in hWAT 8 and 24 days after differentiation compared to non-significant genes (one-sided *t*-test $P=5.29\times10^{-42}$, comparing mRNA normalised counts). From the 56 of 71 FDR-significant genes sufficiently

expressed in differentiated hWAT (mean normalised mRNA count >10 at 8 and 24 days after differentiation), we selected 14 genes for functional characterisation using CRISPR KO which: (1) have been well characterised for their role in adipogenesis or adipocyte lipid and glucose metabolism (*PPARG*[47–52], *PLIN1*[53–57], *INSR*[58–61]), (2) demonstrated monotonic allelic series indicating a dose-response relationship with gene dosage (*ABCA1*, *COL5A3, DENND5B, INSR, PCSK1, PLIN1, PPARG*) (Supp. Figure 3), (3) significantly altered protein levels ($P<2.31\times10^{-7}$) due to obesity or fat distribution associated genetic variation in UKB (*MFAP5*) (Supp. Table 8), (4) associated with obesity age-of-onset (*SLTM, PCSK1*) (Supp. Figure 7, Supp. Table 7), or (5) are implicated in lipid or glucose metabolism pathways by mouse whole-body or adipocyte-specific gene perturbation (Supp. Table 11). We considered the first set of these (*PPARG*, *PLIN1*, and *INSR*) as positive controls, and included *IRS1*, *IRS2*, and *TBC1D4* as additional positive controls based on previous evidence for their effects on insulin-mediated glucose uptake (Supp. Table 11)[61–72].

Using the Toronto KnockOut version 3.0 library[73], we selected four guide RNAs per gene target to generate 17 KO hWAT cell lines and two negative control cell lines, wild-type and Cas9 empty vector (Figure 4c and Supp. Table 12). We confirmed gene KO through significantly decreased ($P<0.05$, one-sided Wald test) mRNA expression relative to expression in Cas9-empty cell line (Figure 4d) and performed western blots for KOs which were not confirmed by mRNA expression (Figure 4e, Supp. Figure 12). We could not confirm KOs of *PCSK1* or *PLIN1* with mRNA expression or western blots. Transcriptome-wide changes observed from RNA-seq and lower protein expression observed from western blots suggest *PPARG* was at the least substantially knocked down.

Abnormal lipid accumulation and insulin resistance are hallmarks of adipose tissue dysfunction[74]. Molecular pathways associated with lipid droplet formation, lipid metabolism and lipid biosynthesis and insulin sensitivity were differentially regulated with body fat distribution in our UKB proteomics analysis (Supp. Table 9). We therefore measured lipid accumulation using BODIPY staining[75], and assayed insulin sensitivity through glucose uptake, measured using a fluorescent glucose analog[76], under basal and insulin-stimulated conditions in hWAT KO cell lines (Methods). The lipid accumulation assay was conducted for six replicates of each KO cell line and the glucose uptake assays were conducted for three replicates under basal and insulin-stimulated conditions each.

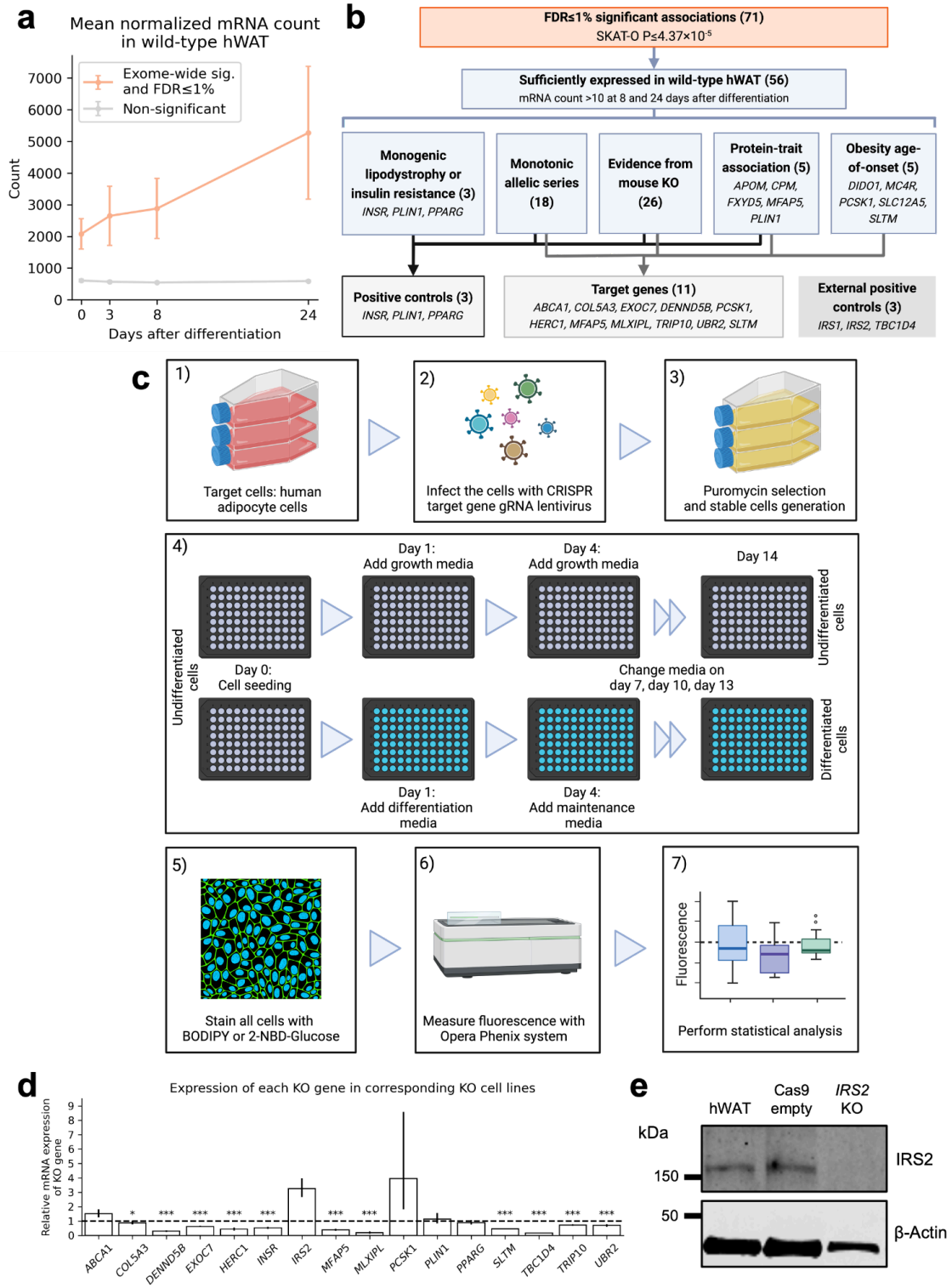

Figure 4. CRISPR KO strategy. a, Mean normalised mRNA counts across four time points (days after differentiation: 0, 3, 8, 24) in human white adipocyte tissue for 71 genes passing FDR≤1% significance (SKAT-O *P*≤4.37×10^{-5}) versus all non-significant genes (n=60,605). Error bars indicate +/-1.96 standard errors of the mean normalised mRNA count. **b,** Selection strategy for genes to include in conditional KO

experiment. Numbers in parentheses indicate the number of genes in a given group. **c,** Experimental design for CRISPR KO, adipocyte cell culture and imaging. **d,** Confirmation of KO efficacy using relative mRNA expression. Error bars indicate +/-1.96 standard errors. One-sided Wald test comparing mRNA count in KO cell line relative to count in Cas9 empty, with the alternative hypothesis that the count is lower in the KO: ***P<0.001, **P<0.01, *P<0.05. *IRS1* KO samples were excluded from RNA sequencing due to low RNA concentrations. **e,** Example of confirmation of KO efficacy using western blots for *IRS2*. For western blots of all other KOs not confirmed by RNA sequencing (*ABCA1, IRS1, PCSK1, PLIN1, PPARG*) see Supp. Figure 12.

Knockout of obesity and fat distribution associated genes alters *in vitro* lipid accumulation and insulin response

We observed significantly reduced (P<0.05 in two-sided t-test of mean fold change [FC] in fluorescence between KO and control Cas9-empty cell line) lipid accumulation and insulin-stimulated glucose uptake in four of six positive control KOs: *PPARG* (lipid accumulation FC=0.245, P=5.53×10^{-7}; glucose uptake FC=0.622, P=0.0186), *IRS2* (lipid FC=0.370, P=9.66×10^{-6}; glucose FC=0.622, P=0.0150), *IRS1* (lipid FC=0.507, P=5.58×10^{-5}; glucose FC=0.747, P=0.0477), and *TBC1D4* (lipid FC=0.741, P=0.0264; glucose FC=0.796, P=0.0469) (Figures 5a and 5c, Supp. Table 13). Three KOs displayed reduced glucose uptake even in the basal state: *PPARG* (FC=0.656, P=0.0203), *IRS1* (FC=0.667, P=0.0242), *IRS2* (FC=0.685, P=0.0335) (Figure 5c, Supp. Table 13). Inactivation of *PPARG* through KO has previously been shown to lower lipid accumulation[49,51,52], while KO or knockdown of *IRS1*, *IRS2*, and *TBC1D4* have previously been shown to reduce insulin-stimulated glucose uptake[61,62,66,68–70], providing validation of the assays used here.

We observed significantly increased lipid accumulation relative to Cas9-empty cells in KOs of *EXOC7* (FC=1.35, P=0.0096), *TRIP10* (FC=1.39, P=0.0157), and *COL5A3* (FC=1.72, P=0.0028) (Figure 5a, Supp. Table 13). The products of these genes are all variously involved in lipid uptake, synthesis, and adipogenesis. EXOC7 is a component of the exocyst complex, which regulates the uptake of free fatty acids by adipocytes[77]. TRIP10 is an interactor of the thyroid hormone receptor (TR-β1)[78], which regulates *de novo* fatty acid synthesis[79]. TR-β1 has been proposed as a target to treat dyslipidaemia[80], and the TR-β1 agonist resmetirom is currently in Phase III clinical trials for treating non-alcoholic fatty liver disease[81]. We also observed increased lipid accumulation in KOs of *COL5A3* (FC=1.72, P=2.78×10^{-3}) (Figure 5a, Supp. Table 13) – as a ubiquitous component of the extracellular matrix, type V collagen informs the proper differentiation and development of adipocytes[82].

Finally, we noted reduced lipid accumulation in KOs of the novel target *SLTM* (FC=0.514, P=1.91×10^{-4}) (Figure 5a, Supp. Table 13). *SLTM* encodes the SAFB-like transcription modulator, which regulates the GLI family of transcription factors in mice[83]. These GLI transcription factors in turn control expression of lipid metabolic genes, including critical adipogenesis transcription factors *PPARG* and *C/EBP* (α, β, γ, and δ)[84]. While this suggests a novel therapeutic target for altering fat deposition, more work is needed to understand why individuals carrying rare pLoF variants in *SLTM* have higher BMI (burden beta (SE) = 0.0191 (0.00455), SKAT-O P=2.65×10^{-6}).

None of the 13 target genes that were not demarcated as positive controls displayed effects on basal or insulin-stimulated glucose uptake in KO cell lines (Figure 5c, Supp. Table 13).

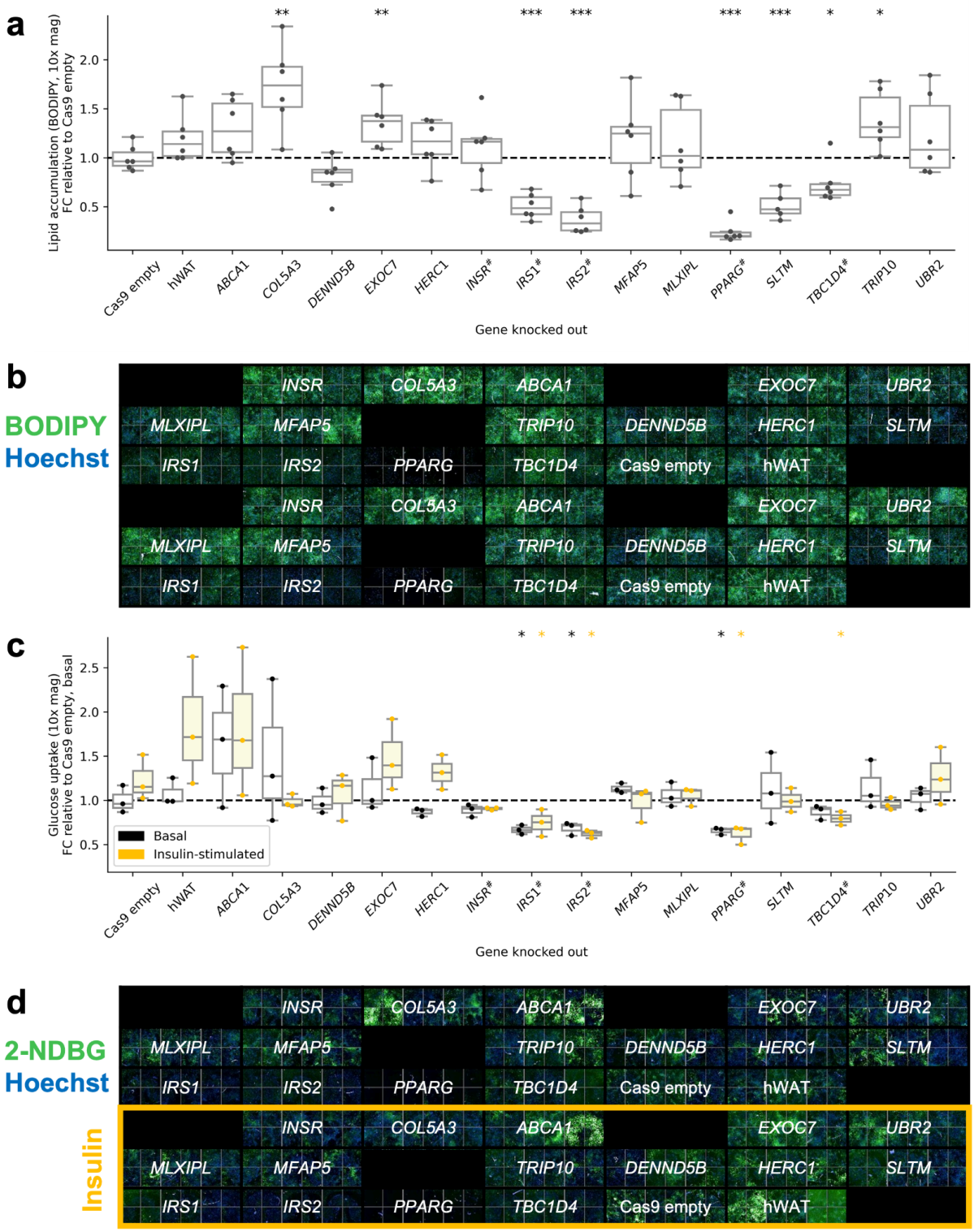

Figure 5. Effect of CRISPR single-gene KO on lipid accumulation and glucose uptake in human adipocytes. a, Effect of single-gene KO on lipid accumulation measured by mean cytoplasm intensity of BODIPY dye fluorescence. **b,** Differentiated human adipocyte KOs stained with BODIPY (green) and Hoechst (blue), imaged at 10x magnification. **c,** Effect of single-gene KO on glucose uptake under basal (black points) and insulin-stimulated conditions (yellow) measured by mean cytoplasm intensity of

2-NBDG fluorescent glucose analog. **d,** Differentiated human adipocyte KOs stained with 2-NDBG (green) and Hoechst (blue), imaged at 10x magnification. Rows 1-3 show adipocytes under basal conditions, rows 4-6 show adipocytes under insulin-stimulated conditions. Fold changes of all basal and insulin-stimulated glucose uptake results are calculated relative to Cas9-empty cells under basal conditions. When testing the effect of KO on insulin-stimulated glucose uptake, each KO is compared to the Cas9-empty cells under insulin-stimulated conditions. Two-sided *t*-test of mean fold change relative to Cas9-empty cells: ***P<0.001, **P<0.01, *P<0.05. #Genes selected as positive controls for lipid accumulation or glucose uptake. FC, fold change.

Transcriptome-wide effects of gene knockouts

Obesity-gene KOs had wide-ranging effects on the hWAT transcriptome, with between eight and 3,253 genes differentially expressed ($|\log_2(\text{fold change})|>1$, FDR-adjusted $P<0.05$) in KO cell lines as compared to Cas9-empty negative controls (Figure 6 and Supp. Table 14). We examined the molecular effects of these KOs through pathway enrichment analyses for pathways in the KEGG[85], REACTOME[86], and GO:Biological Process and GO:Molecular Function[87] databases.

KO of *PPARG*, a master regulator of adipogenesis, significantly alters 95 pathways (Family-wise error rate [FWER] $P<0.05$), including: (1) transcriptional regulation through PRC2-driven DNA methylation (genes=59, $P<0.0001$), histone acetylation (genes=138, $P<0.0001$), and small RNAs (genes=102, $P=0.0001$), (2) protein synthesis by modulating rRNA expression through SIRT1 (genes=63, $P<0.0001$), ERCC6 (CSB) and EHMT2 (G9a) (genes=71, $P<0.0001$) and the B-WICH chromatin remodelling complex (genes=86, $P=0.0001$), and (3) functions critical to cell division such as assembly of the ORC complex at the origin of replication (genes=64, $P<0.0001$), condensation of prophase chromosomes (genes=69, $P<0.0001$), and cell cycle checkpoints (genes=286, $P<0.0001$) (Supp. Table 15). As expected, *PPARG* KO also alters pathways important for adipogenesis, such as the transcription of androgen-receptor regulated genes (genes=62, $P<0.0001$), RUNX1 regulation[88] (genes=89, $P=0.0001$), formation of the beta-catenin TCF trans-activating complex[89] (genes=87, $P=0.0001$), and pre-NOTCH expression and processing[90] (genes=104, $P=0.0003$) (Supp. Table 15). KOs of other positive control genes also significantly affected (FWER $P<0.05$) cell metabolic pathways, including linoleic acid metabolism (*IRS2* KO; genes=24, $P=0.0004$), fatty acid metabolism and adipogenesis (*TBC1D4* KO; genes=84, $P=0.0013$), and ketone metabolism (*TBC1D4* KO; genes=125, $P=0.0037$) (Supp. Table 15), indicating molecular mechanisms by which these genes regulate adipocyte lipid accumulation.

We observed enrichment of several mitotic, transcription, and translation pathways in KOs of lipid accumulation-altering genes *EXOC7* and *TRIP10* (Supp. Table 15), but did not find mechanisms directly implicated in lipid accumulation. On the other hand, while *MLXIPL* KO did not affect lipid accumulation or glucose uptake in our assays, we nevertheless observed enrichment (FWER $P<0.05$) of pathways involved in adipogenesis, such as interferon α/β signalling[91] (genes=66, $P=0.0016$) and lipopolysaccharide binding[92] (genes=32, $P=0.0284$). *MLXIPL* encodes the carbohydrate-responsive element-binding protein (ChREBP), which is thought to regulate gene expression in response to glucose and reduces adipose tissue mass in

mouse KO models[93]. While this may be a potential therapeutic target for fat distribution, mice deficient in MLXIPL/ChREBP do not tolerate fructose in their diet and develop severe diarrhoea and irritable bowel syndrome[94].

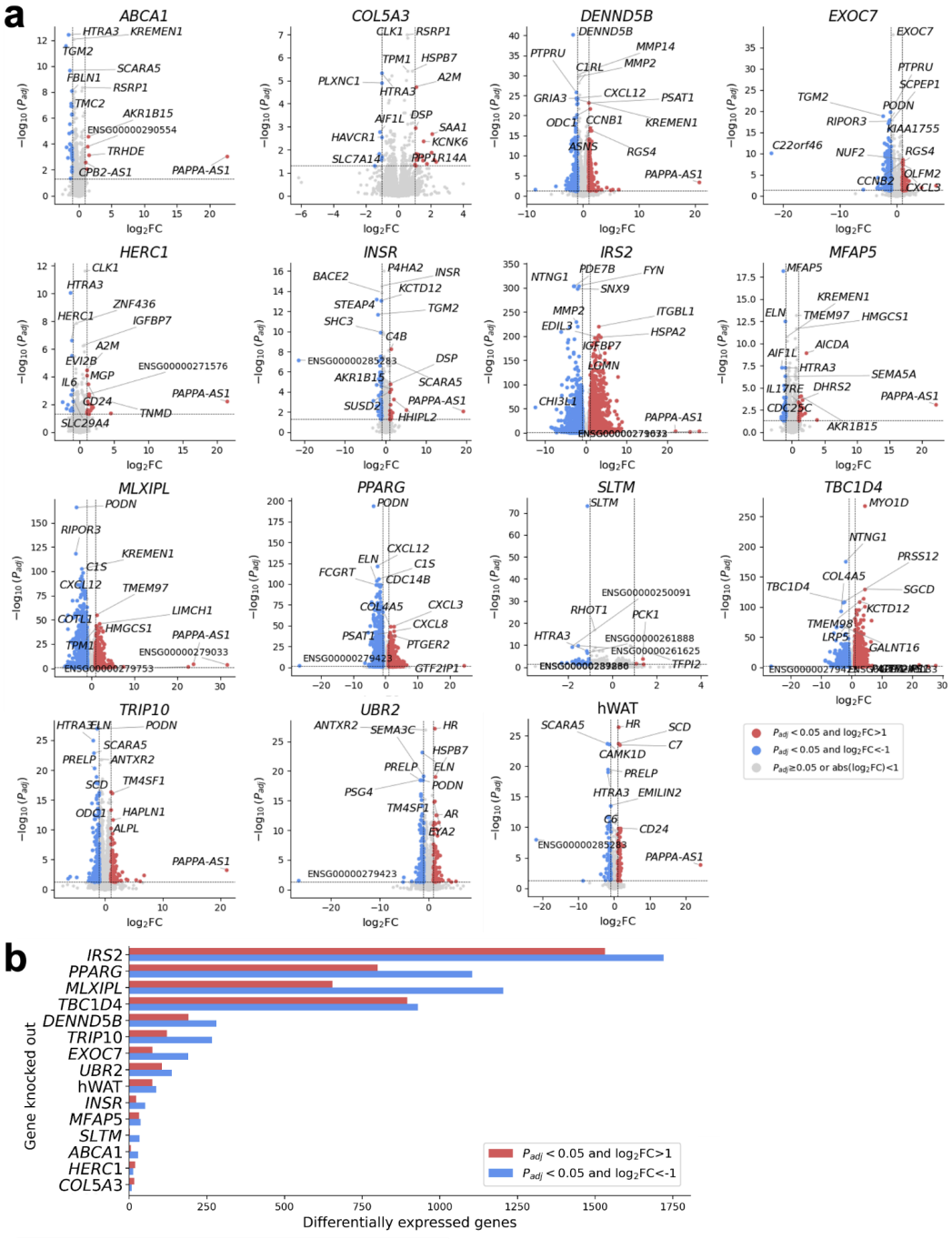

Figure 6. Effect of gene KO on differential gene expression in human adipocytes. a, Gene differential expression results from RNA sequencing data performed on hWAT wild-type and single-gene KOs. Gene differential expression analysis was performed with DESeq2[95], comparing each cell line (wild-type hWAT or KO) to the Cas9-empty cell line. All significance thresholds use P_{adj}, a *P*-value corrected for multiple testing calculated in DESeq2. Blue points indicate genes which are significantly downregulated (P_{adj}<0.05, $\log_2$FC<1). Red points indicate genes which are significantly upregulated

(P_{adj}<0.05, log_2FC>1). Grey points are all other genes. Genes are labelled if they are in the top five most significant (lowest P_{adj}) red or top five most significant blue points, in the top five most significant regardless of point colour, or have abs(log_2FC)>10. The horizontal and vertical dashed lines indicate the significance cutoffs of P_{adj}<0.05 and abs(log_2FC)>1, respectively. **b,** Counts of significantly downregulated (blue, P_{adj}<0.05 and log_2FC<-1) and upregulated (red, P_{adj}<0.05 and log_2FC>1) genes for each single-gene KO and hWAT. The y-axis is sorted by the total number of differentially expressed genes for each KO condition. FC, fold change.

Discussion

Through a series of population-level genetic and *in vitro* functional genomics investigations, we demonstrate a model to map the biological mechanisms of obesity and body fat distribution, from variant to molecular function to systemic phenotype. We utilised complementary rare variant gene-based analyses and common variant prioritisation to nominate 71 genes with robust associations to BMI, WHRadjBMI, body fat percentage, and six tissue-derived fat components in up to 402,375 participants in UKB. By evaluating the sex-specific and phenome-wide associations of these genes, we built a comprehensive understanding of their systemic effects. Combining multiple lines of genomic, transcriptomic, proteomic, and prior functional evidence, we selected 17 genes for functional characterisation by CRISPR KO in human white adipose tissue (hWAT) cell lines. Among these 17 genes were six “positive controls”, of which three replicated known effects of KO on lipid accumulation and glucose uptake in adipocytes. We also observed reduced lipid accumulation in *SLTM* KO, and increased lipid accumulation in KOs of *COL5A3, EXOC7* and *TRIP10*; many of these are potential therapeutic targets to regulate body fat mass and/or distribution. Finally, we observed transcriptomic changes consistent with altered adipogenesis in *MLXIPL* KO cell lines. Taken together, our population-based and *in vitro* genetics investigations highlight molecular mechanisms and therapeutic avenues for fat distribution.

The recent growth in sample sizes of whole-exome sequenced participants in biobanks has accelerated the development of gene-level analyses[96–98]. By annotating rare variants with their putative functional consequences and then collapsing these across each gene, gene-level testing has enabled the discovery of novel genes associated with complex human traits[31,99]. While gene-level testing is a powerful strategy for gene discovery, it remains limited by the requirement of high-quality functional annotations of rare variants. Here, we combined evidence from CADD, REVEL, and SpliceAI annotations[23–25] to generate high-confidence sets of variants, and integrated evidence from common variant studies to prioritise putative risk genes for functional follow up. In the future, improved rare variant masks, such as those incorporating non-coding rare variation outside exonic regions, may help identify additional therapeutic targets to alter fat distribution.

Genome-wide common variant and exome-wide rare variant analyses can provide complementary lines of evidence to nominate genes associated with human traits[99,100]. We found that genes implicated for obesity and body fat distribution by a burden of rare damaging variation were: (1) associated with rare Mendelian forms of obesity and lipodystrophies, and (2) also prioritised by common variant prioritisation scores[26] for BMI, WRHadjBMI and body fat

percentage. Converging evidence from across the allele frequency and phenotypic severity spectra can therefore identify high-confidence genes for a range of human traits. Assessing the sex-specific and phenome-wide consequences of these genes can provide evidence for therapeutic potential, for example by revealing possible side-effects on platelet distribution or neutrophil percentage.

Our evaluation of the proteome-wide consequences of obesity and fat distribution associated genetic variation revealed putative novel mechanisms and therapeutic targets for fat distribution. For example, individuals carrying rare variants in *PDE3B* have higher levels of the adipocyte-adhesion molecule CLMP and lower WHRadjBMI; the overexpression of *CLMP* reduces weight gain in mice[44]. While *PDE3B* is a well-known regulator of energy homeostasis[101], it may additionally act on adipogenesis through its effect on CLMP, which we propose as a therapeutic modulator of obesity. We also found an association between damaging variation in *GIGYF1* and increased WHRadjBMI, accompanied by lower plasma levels of the paired receptors PILRA and PILRB. Loss of function in *GIGYF1* has previously been suggested to impact metabolic health through mosaic loss of the Y chromosome in leukocytes[102]; we suggest that it may additionally impact obesity by lowering expression of *PILRA*, a gene that drives macrophage infiltration into adipose and liver tissue and induces obesity when knocked out in mice[45].

We measured the effects of CRISPR gene KOs on lipid accumulation and glucose uptake (in basal and insulin-stimulated conditions) in hWAT cell lines. To ensure our *in vitro* assays were robust and replicable we performed each assay with six (lipid accumulation) or three (glucose uptake under basal or insulin-stimulated conditions) replicates, and included a set of "positive control" genes that have previously been characterised to affect the readouts of interest. However, we were unable to fully confirm the KO effect in two genes selected as positive controls, *PPARG* and *PLIN1.* The KO of *PPARG* shows a combination of transcriptome-wide disruption in the differential expression analysis and reduction in protein expression from the western blot, suggesting at least a partial knockout. Lipid accumulation and glucose uptake results were also consistent with expected results of a successful *PPARG* KO. Since we could not confirm complete KO of *PPARG*, we avoid reference to this KO as a positive control for the lipid/glucose assay results. We do observe clear evidence of disruption, and therefore retain the KO for downstream analyses with the caveat that interpretation should be tempered by the possibility of incomplete knockout. The KO of *PLIN1* could not be confirmed by mRNA differential expression or by protein expression measured by western blot. The difficulty in observing protein expression with a western blot may be due to low expression of *PLIN1* in the undifferentiated adipocytes used for western blots, as indicated by our longitudinal mRNA analysis of wild-type hWAT (Supp. Figure 11).

Three of four positive controls with confirmed KOs (*IRS1, IRS2, TBC1D4*) affected lipid accumulation or insulin-stimulated glucose uptake in hWAT KO cell lines. Two positive controls with confirmed KOs (*IRS1, IRS2*) also affected basal glucose uptake, however this may be due to the KO affecting adipocyte differentiation rather than affecting a glucose uptake pathway independent of insulin. The only positive control with confirmed KO and which did not

significantly affect lipid accumulation or glucose uptake was *INSR*. *INSR* encodes the insulin receptor gene, whose KO is expected to decrease insulin-stimulated glucose uptake[59]. Indeed, when we pool replicates from basal and insulin-stimulated states, we observe a significant decrease in glucose uptake in *INSR* KO cell lines, but with a smaller effect size than other positive control KOs (Supp. Note 3). We were likely under-powered to identify the effect of *INSR* KO on glucose uptake with only three replicates per cell line.

Of the two genes carrying rare missense variation associated with increased BMI and with confirmed KOs, *EXOC7* KO increased lipid accumulation in hWAT cell lines while *SLTM* KO decreased lipid accumulation. Rare missense variation in *PPARG* is associated with decreased body fat percentage and the disruption of *PPARG* in hWAT cell lines resulted in decreased lipid accumulation. Similarly, of the two genes affecting central obesity (WHRadjBMI or ratio of android to gynoid tissue fat percentage) with confirmed KOs significantly increasing lipid accumulation, *COL5A3* was correspondingly associated with increased central obesity, whereas *TRIP10* was associated with decreased central obesity. Our results demonstrate the importance of systematically characterising the functional effects of gene perturbation at the level of cell lines and model organisms, which may not follow the expected direction of effect based on missense variation in humans. A limitation of our CRISPR screen was the introduction of complete gene KOs; more sensitive study designs in the future may seek to mimic the effects of specific missense or damaging variation.

In summary, we demonstrate that the convergence of evidence across rare and common genetic variation, combined with large-scale proteomics, can help identify high-confidence target genes for overall adiposity and body fat distribution. We provided in-depth functional characterisation through CRISPR KOs in human adipocytes, allowing us to nominate candidate genes for therapeutic targets. Through the functional readouts and transcriptomic analyses, we also highlighted several novel molecular mechanisms by which genetic variation may impact obesity. Our results provide a model by which future work can integrate genetic and functional evidence to identify, design, and validate potential drug targets to alter overall and tissue-specific fat distribution.

Methods

Imputed data quality control

Sample quality control and assigning population labels

Beginning with the 487,409 individuals with phased and imputed genotype data, we restricted to unrelated individuals with low autosomal missingness rates used for PCs by Bycroft *et al*.[103]. We then used genotyping array data, subset to LD-pruned autosomal variants, from these samples to project into the PC space defined by the 1000 Genomes dataset[104], ensuring that we correctly account for shrinkage bias in the projection[105]. Next, we used the ‘super-population’ labels (AFR=Africans, AMR=Admixed Americans, EAS=East Asians, Europeans=EUR, South

Asians=SAS) of the 1000 Genomes dataset to train a RF classifier, using the randomForest (4.6) library in R[106], and predicted the super-population for each of the UKB samples. Samples with classification probability >0.99 for the European super-population, were retained for downstream analysis.

Following our ancestry based filtering regime, we remove samples who withdrew from UKB participation as well as those individuals who were omitted from phasing and imputation. For GWAS, we further subset individuals passing imputed data quality control to those who also passed exome sequencing quality control (described below), resulting in a set of 397,315 individuals.

Variant quality control

Over 92 million imputed variants across the autosomes and chromosome X are available for analysis. As a starting point for our initial collection of GWAS, we subset to variants with MAF >0.1% in the subset of individuals defined in the sample QC procedure and an INFO score >0.8 from the UKB SNP manifest file. Following this collection of filtering steps, 16.7 million variants were retained for common variant GWAS. After performing GWAS, two additional filters were performed, retaining variants with Hardy-Weinberg $P>1\times10^{-10}$, calculated on the whole QCed subset of individuals, and retaining variant-level summary statistics if MAF >0.1%, calculated for each GWAS. This results in 13,117,850 variants for downstream common variant analysis.

Exome sequencing data quality control

Exome sequencing data quality control summary

Using the UKB Research Access Platform we accessed the gnomAD VCFs containing GATK-called genotypes for 454,671 individuals. We followed an identical sample QC protocol to Karczewski *et al.*[31], filtering individuals on covariate-regressed individual-level metrics using median absolute deviations. We removed variants flagged by the gnomAD QC team for failing one or more of their filters (*AC0*, *RF*, *MonoAllelic*, *InbreedingCoeff*). Genotype calls were set to missing if they failed filters for genotype quality, depth and alelle balance (described in Karczewski *et al.*[31]). We filter to European ancestry using the super-population ancestry labels assigned with an RF classifier, described above for imputed data quality control.

Variant and sample-level QC

We defined ‘high quality’ variants as those MAF>0.1% and call rate ≥0.99 falling within the UKB capture intervals plus 50 bp padding. These variants were used to evaluate sample-level metrics of mean call rate and depth and retained samples satisfying all of the following:

- Genetic sex inferred as XX or XY (specifically, genetic sex is defined).
- Mean call rate ≥0.99 among high quality variants.
- Mean coverage ≥20x among high quality variants.
- Not withdrawn.

Next, we removed variants satisfying at least one of the following criteria:

- The variant lies outside the UKB capture plus 50 bp padding.
- The variant lies within a low complexity region.
- The variant lies within a segmental duplication.

Among this (sample, variant) set, we ran Hail's `sample_qc()`[107] to remove samples lying outside the median±4 median absolute deviations (MADs) within each super-population (see above section on imputed data quality control). The QC protocol was split by UKB ES tranche (50k, 200k, 450k) to guard against batch effects, as tranches were sequenced in separate runs. The following metrics were used for QC:

- Number of deletions (`n_deletion`).
- Number of insertions (`n_insertion`).
- Number of SNPs (`n_snp`).
- Ratio of insertions to deletions (`r_insertion_deletion`).
- Ratio of transitions to transversions (`r_ti_tv`).
- Ratio of heterozygous variants to homozygous alternate variants (`r_het_hom_var`).

Following MAD filtering (Supp. Figure 1, Supp. Table 1), 402,375 European samples were retained for analysis. For each sample, we excluded non-passing sites as described in Karzcewski *et al.*[31]. Briefly, an RF classifier was trained to distinguish true positives from false positive variants using a collection of allele and site annotations. Variants were assigned 'PASS' to maximise sensitivity and specificity across a series of readouts in trio data and precision-recall in two truth samples, after which samples with excess heterozygosity (defined as inbreeding coefficient <-0.3) were removed. Next, we removed low quality genotypes by filtering to the subset of genotypes with depth ≥10 (5 among haploid calls), genotype quality ≥20, and minor allele balance >0.2 for all alternate alleles for heterozygous genotypes. Following this filter, we remove variants that were not called as "high quality" among any sample. The resulting high quality European call set consisted of 402,375 samples and 25,229,669 variants.

Variant consequence annotation

We annotated exome-sequencing variants using Variants Effect Predictor (VEP) v105 (corresponding to GENCODE v39)[108] with the LOFTEE v1.04_GRCh38[22] and dbNSFP[109] plugins, annotating variants with CADD v1.6[23], and REVEL using dbNSFP4.3[24] and loss-of-function confidence using LOFTEE. We provide code and instructions for this step in our VEP_105_LOFTEE repository[110], which contains a Docker/Singularity container for reproducibility of annotations. Next, we ran SpliceAI v1.3[25] using the GENCODE v39 gene annotation file to ensure alignment between VEP and SpliceAI transcript annotations. For variant-specific annotations we use 'canonical' transcripts. We separated variants by transcript

using `bcftools +split-vep` and filtered to MANE Select[111] protein-coding transcripts. If the gene lacked a MANE Select transcript we selected the canonical transcript defined by GENCODE v39. Using this collection of missense, pLoF, and splice metrics, and annotations of variant consequence on the canonical transcript, we then determine a set of variant categories for gene-based testing.

Variant consequence categories

1. **High confidence pLoF**: High-confidence LoF variants, as defined by LOFTEE[22] (LOFTEE HC).
2. **Damaging missense/protein-altering**: At least one of:
    a. Variant annotated as missense/start-loss/stop-loss/in-frame indel and (REVEL≥0.773 or CADD≥28.1 or both).
    b. Any variant with SpliceAI delta score (DS)≥0.2 where SpliceAI DS the maximum of the set {DS_AG, DS_AL, DS_DG, DS_DL} for each annotated variant (where DS_AG, DS_AL, DS_DG and DS_DL are delta score (acceptor gain), delta score (acceptor loss), delta score (donor gain), and delta score (donor loss), respectively).
    c. Low-confidence LoF variants, as defined by LOFTEE (LOFTEE LC)
3. **Other missense/protein-altering**: Missense/start-loss/stop-loss/in-frame indel not categorised in (2).
4. **Synonymous**: Synonymous variants with SpliceAI DS<0.2 in the gene.

REVEL and CADD score cut-offs are chosen to reflect the supporting level for pathogenicity (PP3) from the American College of Medical Genetics and Genomics and the Association for Molecular Pathology (ACMG/AMP) criteria[112].

Phenotype curation

We used the following nine phenotypes, either directly measured by UKB or derived from UKB phenotypes: body mass index (UKB code: 21001), waist-to-hip ratio (derived from waist circumference [UKB code: 48] and hip circumference [UKB code: 49]) with BMI regressed out, body fat percentage (UKB code: 23099), android tissue fat percentage (UKB code: 23247), gynoid tissue fat percentage (UKB code: 23264), ratio of android to gynoid tissue fat percentage (derived phenotype), total tissue fat percentage (UKB code: 23281), visceral adipose tissue volume (UKB code: 22407), abdominal fat ratio (UKB code: 22434).

Genetic association testing

Genetic association testing summary

All variant- and gene-level associations were performed in the European-ancestry subset of the UKB using the Scalable and Accurate Implementation of GEneralised mixed model (SAIGE)[96], a mixed model framework that accounts for sample relatedness.

In SAIGE step 0, we constructed a genetic relatedness matrix (GRM) using the UKB genotyping array data. The genotyped data is LD pruned using PLINK (`--indep-pairwise 50 5 0.05`) [113], and the sparse GRM is calculated using the `createSparseGRM.R` function within SAIGE, using 5,000 randomly selected markers, with relatedness cutoff of 0.05,

To generate a variance ratio file for subsequent steps in SAIGE, we selected 2000 variants from the genotyping array data to define a PLINK dataset. For testing common variants in imputed data, we extracted 2000 variants with MAC≥20. For testing the exome-sequencing data, we extracted two sets of 1000 variants with 10≤MAC<20 and MAC≥20, and combined these sets of markers.

In SAIGE step 1 for each trait, the null model is fit using the curated phenotype data and sparse GRM, with no genetic contribution. Default parameters were used in SAIGE, except `--relatednessCutoff 0.05`, `--useSparseGRMtoFitNULL TRUE` and `--isCateVarianceRatio TRUE`. In the sex-combined analyses, we account for *age, sex, age*^{2}*, age* × *sex, age*^{2} × *sex*, and the first 21 PCs as fixed effects; and *age, age*^{2}, and the first 21 PCs in sex-specific analyses. All continuous traits were inverse rank normalised using the `--invNormalize` flag in SAIGE. For SAIGE step 2, we always use the flag `--LOCO FALSE` and `--is_fastTest TRUE`.

Common variant association testing in imputed data

We performed single-variant genetic association testing of the UKB imputed data using SAIGE version 1.1.6.3 on the Oxford Biomedical Research Computing cluster. For consistency across analyses, only individuals who passed sample-level quality control in both the imputed and exome-sequencing data were included.

Following null model fitting, we carried out variant testing in SAIGE step 2 with default parameters, except for `--is_Firth_beta TRUE` and `--pCutoffforFirth=0.1`.

Finemapping

Using summary statistics from the common variant association tests, finemapping loci are identified for each trait. Starting with the most significant variant in the pool of genome-wide significant variants, a 1 Mb window centred around the variant is created. All genome-wide variants falling in this window are considered part of the locus defined by the most significant variant and removed from the pool of genome-wide significant variants. Then we proceed to the next most significant variant in the pool of genome-wide significant variants and repeat until the pool of genome-wide significant variants is empty. Loci with overlapping windows are merged.

LD matrices were calculated with LDstore v2.0[114], using the same set of individuals used in the GWAS for each trait. Finemapping was performed with FINEMAP v1.4[114], using shotgun stochastic search and allowing a maximum of 10 causal SNPs per locus.

Rare variant and gene-level testing in exome sequencing data

We carried out rare variant and gene-level genetic association testing in the European-ancestry subset of the UKB exome sequencing data using SAIGE. All analyses involving the exome sequencing data was carried out on the UKB Research Analysis Platform (RAP) using SAIGE version wzhou88/saige:1.1.9[96].

Following null model fitting, we carried out variant and gene-level testing in SAIGE step 2 using the variant categories described above, with the `--is_single_in_groupTest TRUE` flag. All other parameters were set to default, except `--maxMAF_in_groupTest=0.0001,0.001,0.01`, although we only report results for maximum MAF of 0.01. We included the following collection of group tests, using the annotations defined above (see 'Variant consequence annotation' methods section):

- High confidence pLoF
- High confidence pLoF or damaging missense/protein-altering

For genes with no damaging missense/protein altering variants, only the pLoF variant consequence mask was considered when performing Bonferroni multiple-testing correction.

Obesity and fat distribution trait gene-level association tests

For each gene and for each obesity and fat distribution trait we tested the SKAT-O association of rare variant (MAF<1%) variation. Exome-wide statistical significance ($P < 1.58\times10^{-7}$) for gene-level tests was defined using Bonferroni correction for 315,996 unique phenotype-gene-consequence mask combinations.

To expand the set of obesity and fat distribution associated genes to consider for functional screening, we followed an FDR-control process similar to Zhou *et al.*[115]. We selected the gene-level association result which had the lowest SKAT-O *P*-value (across all phenotypes and both variant masks) and applied the Benjamini-Hochberg procedure to identify 83 significant genes when controlling FDR to 1% (corresponding to a *P*-value of 4.37×10^{-5}). If we had performed the FDR-controlled selection using the full set of results instead of taking the result with the lowest *P*-value this would be equivalent to an FDR of 12.0%.

We re-ran the gene-level association tests for the 83 FDR-significant genes, conditional on the top finemapped common (MAF<0.1%) variants for finemapped loci (described above). For each trait, we identified finemapped common variants for each trait which have the highest Bayes factor for being a causal variant (i.e. strongest evidence for causality) in their respective loci. Variants tied for the highest Bayes factor are all selected. There are a total of 915 finemapped trait-variant pairs (846 unique variants) on which to condition. We used the --condition flag in SAIGE to condition on the selected common variants when performing gene-level tests for the corresponding phenotype. We perform the gene-level association tests for each chromosome independently and only conditioned on common variants from the same chromosome as the genes being tested.

After conditioning, seven genes (*ACVR1C, PRPH, PYGM, SEC16B, TNFRSF6B, TRIM40, TUBE1*) were no longer significant at the predefined SKAT-O $P \leq 4.37 \times 10^{-5}$ threshold, leaving 76 genes. Of these 76, five genes (*DEFB112*, *CHMP4B*, *FEZF2*, *GLP1R*, *PCBD2*) were flagged as having very low minor allele count (MAC<5), leaving 71 genes to be considered for functional screening.

Olink proteomics

For burden testing across the proteome in the 71 FDR-significant genes, we used 1,456 proteins available in the UKB Olink proteomics data, excluding those which had >10% sample missingness (TACSTD2, CTSS, PCOLCE, NPM1) and zero variance component in SAIGE Step 1 (ENPP5, FOLR3, PM20D1). Statistical significance ($P<2.44 \times 10^{-7}$) for gene burden was defined using Bonferroni correction for 204,743 unique phenotype-gene-consequence combinations.

Common variant gene prioritisation scores

We used Polygenic Priority Scores calculated by Weeks *et al.*[26] for UKB BMI, WHRadjBMI and body fat percentage. Scores were downloaded from https://www.finucanelab.org/data.

Age at diagnosis analyses

We curated age at diagnosis of obesity from the UKB linked primary care and hospital record data using mapping tables generated by Kuan et al.[116]. Any codes related to “history of obesity”, for which accurate age at diagnosis could not be extracted, were excluded. We left-truncated observations at the age of the first record (of any code) in either primary care or hospital data, and right-censored at the age of the last record. For each of the 71 FDR-significant genes identified from ES analyses, we performed Cox-proportional hazards modelling[117] to estimate differences in lifetime risk of developing obesity between carriers of pLoF, damaging missense, other missense, synonymous, or non-coding variants (at MAF<1%) and wild-type individuals (reference group). All effects were adjusted for sex (in sex-combined analyses), the first 10 genetic PCs, birth cohort (in ten-year intervals from 1930 to 1970), and UKB assessment centre. Cox modelling was performed using the R package survival v3.3.1[118]. We visualised age at onset probabilities using Kaplan-Meier survival curves with the R package survminer v0.4.9[119].

Genebass gene-level summary statistics

Gene-level association test summary statistics were calculated by Karczewski *et al.*[31] and downloaded as a Hail Matrix Table from https://app.genebass.org/downloads. Only pLoF-associated results corresponding to our 71 FDR≤1% genes were used. Associations were significant if Genebass SKAT-O $P \leq 9.69 \times 10^{-6}$, controlled for FDR≤1% using the Benjamini-Hochberg procedure[27].

Genetic correlation estimates

Genetic correlation estimates were accessed from the Neale lab's UKB genetic correlation browser: https://ukbb-rg.hail.is/rg_browser/. These genetic correlations were estimated using LD Score regression[32], using summary statistics from common (MAF>0.1%) imputed variant GWAS.

Trait-protein association tests

To assess the effect of plasma protein on each obesity or fat distribution trait, we performed multiple linear regression with each inverse-rank normalised obesity or fat distribution trait as the dependent variable and the plasma protein as the explanatory variable, along with the same covariates used for genetic association testing except sequencing tranche:

$$trait \sim protein + sex + age + age^2 + sex*age + sex*age^2 + PC1 + PC2 + \ldots + PC21 + assessment\ centre.$$

Statistical significance ($P<3.80\times10^{-6}$) for a significant association between protein and trait was defined using Bonferroni correction for a total of 13,167 trait-protein combinations. We used the Python package UpSetPlot[120] to generate UpSet plots.

Pathway enrichment

Trait-protein associations

Using GSEAPreRanked[38] we tested for pathway enrichment in a list of proteins ranked by *z*-score of the most significant trait association across the nine obesity and fat distribution traits. Pathways with positive normalised enrichment score (NES) are those enriched in proteins with strong positive direction of effect (high magnitude positive *z*-score) in obesity and fat distribution traits. Pathways with negative NES are enriched in proteins with strong negative direction of effect (high magnitude negative *z*-score).

CRISPR RNA sequencing

We used Gene Set Enrichment Analysis (GSEA)[38] to test for pathway enrichment in CRISPR adipocytes RNA sequencing data. Regularised log transformed mRNA counts calculated by DESeq2 are provided to GSEA. The pathway enrichment for each KO or wild-type hWAT is compared against the baseline control, the empty Cas9 vector, such that pathways with more positive NES are those enriched in Cas9 empty and those with more negative NES are those enriched in the KO or hWAT.

For both the CRISPR RNA sequencing and trait-protein association results we use pathway sets from KEGG Legacy, KEGG Medicus, REACTOME, GO:Biological Process, and GO:Molecular Function, downloaded from GSEA's Molecular Signatures Database v2023.2.Hs[38]. When running GSEA or GSEAPreRanked we use 10,000 permutations.

Selecting target genes

To choose target genes for CRISPR-Cas9 KO we consider all gene association results with FDR<0.01 which remain significant ($P \leq 4.37 \times 10^{-5}$) after conditioning on common variation significantly associated with the trait (Methods). We then compiled multiple lines of evidence for each gene:

- Sufficiently expressed in wild-type hWAT cells (RNA-seq transcript counts >10 at eight and 24 days after differentiation).
- Responsible for a monogenic form of lipodystrophy[6] or obesity[28].
- Monotonic allelic series indicating a dose-response relationship with gene dosage and a obesity or fat distribution trait.
- Significantly altered protein levels due to obesity or fat distribution associated genetic variation
- Significant obesity age-of-onset association.
- Published functional work involving gene knockout or knockdown implicates the role of the gene in adiposity.

Using these lines of evidence we separated the genes into categories: positive controls (genes previously implicated in obesity pathways by functional work in human adipocytes), plausible candidates (genes which have not been implicated in adiposity pathways by functional work in human adipocytes, but have suggestive evidence from other functional work or from GWAS/burden association results), and impossible candidates (RNA-seq counts are too low or gene seems to be associated with a pathway that we are not testing). From these categories we chose three positive controls (*PLIN1*, *PPARG*, *INSR*) and 11 plausible genes (*COL5A3*, *ABCA1*, *EXOC7*, *MFAP5*, *MLXIPL*, *PCSK1*, *UBR2*, *TRIP10*, *DENND5B*, *HERC1*, *SLTM*). We note that although *PCSK1* is well-characterised as a driver of monogenic obesity, we do not consider it a positive control for our experiment because it primarily acts through neuroendocrine pathways rather than adipose tissue-specific lipid or glucose metabolism[121]. We included three additional positive controls (*IRS1*, *IRS2*, *TBC1D4*) to act as positive controls for the glucose uptake assay due to previous evidence suggesting their role in insulin-stimulated glucose uptake (Supp. Table 11).

Cell Culture and Reagents

Human white adipose tissue cells were cultured in Dulbecco's Modified Eagle's Medium (DMEM, Sigma Aldrich, Cat # D6546) with 10% foetal bovine serum (FBS, ThermoFisher Cat # A5256801) in a 37°C humidified incubator with 5% CO_2[122].

Cell Line generation

Guide RNAs (gRNAs) (four gRNAs per gene target) were commercially synthesised and cloned into the lentiviral vector pLentiCRISPRv2 (GenScript, UK). gRNA sequences are listed in Supp. Table 12. Lentivirus for individual gRNAs was produced. Briefly, CRISPR plasmids were cotransfected in HEK293T with packaging vectors pMD2.G (Addgene, 12259) and psPAX2

(Addgene, 12260) using FuGENE® HD Transfection Reagent (Promega Cat #E2311). Viral supernatant was collected at 48 hours and 72 hours post-transfection. hWAT cells were transduced and selected for 3 days in 1 µg/ml puromycin with media changes as required.

Differentiation of human white adipose tissue cells

The wild-type hWAT and single-gene KO (including the negative control 'KO' with the empty Cas9 vector) cells were cultured in Dulbecco's Modified Eagle Medium (DMEM) containing 10% fetal bovine serum (FBS). To differentiate into adipocytes via addition of differentiation mixture, the cells were induced with 0.5mM 3-isobutyl-1-methylxanthine (IBMX, Sigma Cat # 5879) and 0.1µM dexamethasone (Sigma Cat #1756), 33µM Biotin (Sigma Cat # 4639), 17µM Pantothenate (Sigma Cat #5155), 0.5µM insulin (Sigma Cat # 2643), 2nM 3,3′,5-Triiodo-L-thyronine (T3, Sigma, Cat #6397), and 1µM Rosiglitazone (APE Cat #A4304) for 3 days. Next, culture medium was replaced to DMEM supplemented with 10% FBS and 33µM Biotin, 17µM Pantothenate, 0.5µM insulin, 2 nM 3,3′,5-Triiodo-L-thyronine, and 1µM Rosiglitazone for 3 days. Medium was subsequently changed every 3 days for the following days.

Lipid staining with BODIPY

Undifferentiated and differentiated hWAT and single-gene KO cells were stained with 2.9 µM BODIPY 505/515 (ThermoFisher Cat # 4639) and Hoechst (Hoechst 33342, Trihydrochloride, Trihydrate - 10 mg/mL solution in water, ThermoFisher Cat # H3570) for 15 min at 37°C in complete medium, washed with PBS. Cells were analysed for fluorescence intensity using an Opera Phenix™ High-Content Screening System (Revvity, UK) and Harmony® high-content analysis software v4.9 (Revvity, UK) provided by the Cellular High Throughput Screening Group (cmd.ox.ac.uk). For each KO condition, there were six replicates of undifferentiated and differentiated cells.

Glucose uptake assay

Undifferentiated and differentiated hWAT and single-gene KO cells were stained with 2-NBD-Glucose (Cayman Cat# 11046) for 1 hour. After 1 hour, Hoechst (Hoechst 33342, Trihydrochloride, Trihydrate -10mg/mL solution in H_2O, ThermoFisher Cat# H3570) was added for 15 min at 37°C. They were then washed with PBS pH 7.4, 1X (ThermoFisher Cat# 10010-015). Cells were analysed for fluorescence intensity using an Opera Phenix™ High-Content Screening System (Revvity, UK) and Harmony high-content analysis software v4.9 (Revvity, UK) provided by the Cellular High Throughput Screening Group (cmd.ox.ac.uk). For each KO condition, there were six replicates of undifferentiated and differentiated cells, with three replicates each for the basal and insulin-stimulated conditions.

RNA sequencing of human white adipose tissue cells

Wild-type hWAT and single-gene KO cells were seeded in six-well plates and collected after either (1) zero (undifferentiated), 3, 8 or 24 days of differentiation when assessing sufficient expression in wild-type hWAT or (2) 14 days of differentiation when performing RNA-seq for cell

lines involved in the CRISPR KO experiment (wild-type hWAT, and single-gene KO cells, including Cas9-empty cells). Following collection, RNA was isolated using TRIzol (Thermo Fisher Scientific, Carlsband, CA, USA) according to the manufacturer's protocol. Subsequently, RNA samples were purified using the Direct-zol™ RNA Miniprep protocol as per manufacturer's (Zymo Research, Cat # R2050) recommendations. Quantification of RNA was performed using the Nanodrop 1000 Spectrophotometer. All RNA samples were sequenced using Illumina 2x150bp paired-end sequencing. Read alignment was performed with STAR[123]. Transcript quantification estimation was performed with RSEM[124]. Normalised transcript counts were calculated with DESeq2[95]. The single-gene KO for *IRS1* had insufficient RNA for sequencing and therefore was excluded from RNA sequencing and downstream analyses involving RNA-seq data.

Immunoblotting

Cells were washed with ice-cold PBS and lysed with a radioimmunoprecipitation assay (RIPA) buffer containing protease and phosphatase inhibitors. Following 30 minutes of incubation in the lysis buffer, the cells were centrifuged (12000 rpm, 4°C, 15 min). The samples were denatured in 4x sodium dodecyl sulfate (SDS) sample buffer at 95°C for 10 minutes. For immunoblotting, equal amounts of total cell protein were separated by Tris-glycine SDS-PAGE gradient (4-15% acrylamide) gels (#3450123; Bio-Rad). Proteins were then transferred to polyvinylidene difluoride (PVDF) membranes transferred onto nitrocellulose membranes (Millipore). Nonspecific antibody binding was blocked through incubation with blocking solution (5% milk in TBS-T) under constant shaking for 1 hour. After blocking, the membrane was incubated at 4°C with a primary antibody solution overnight. The following primary antibodies were used: ABCA1 (#96292, Cell Signalling Technology), IRS1 (#PA1-1057, Invitrogen), IRS2 (#PA5-119208, Invitrogen), PLIN1 (#9349, Cell Signalling Technology), PCSK1 (#ab220363, Abcam), PPARγ (D69) Antibody (#2430; Cell Signalling Technology), SLTM (#PA5-59154, Invitrogen), and beta-actin monoclonal antibody (#AM4302, Invitrogen). Membranes were washed three times with TBS-T for 15 minutes. The corresponding appropriate secondary antibodies (anti-Rb/M 800, IRDye 800CW) were incubated for 1 hour, and the membrane was washed three times with TBS-T. An Odyssey CLx Infrared imaging system (LI-COR) was used to detect the signal.

Differential expression analysis of CRISPR adipocyte RNA sequencing

Differential expression analysis was performed using DESeq2[95] on RNA sequencing data from CRISPR adipocytes. Genes with raw count less than 10 were removed. The Cas9-empty cell line was used as the baseline condition. Each KO or wild-type hWAT was then separately compared to Cas9 empty.

In the differential expression results of the *IRS2* KO, two genes (*FYN*, *NTNG1*) have P_{adj}=0, which makes it impossible to take the log when plotting. For the purpose of visualisation, we assign a $-\log_{10}(P_{adj})$ value of 303.88 to these genes, which is one greater than the highest $-\log_{10}(P_{adj})$ value observed in the data for nonzero P_{adj}.

Statistical analysis of lipid accumulation and glucose uptake assays

We quantify lipid accumulation and glucose uptake using mean cytoplasm fluorescence of the BODIPY or 2-NBD-Glucose stain, as measured by the Opera Phenix™ High-Content Screening System. For each KO and the wild-type hWAT, we combined replicates with the baseline negative control, Cas9 empty, and regress readout against KO status:

$$readout \sim is\_ko + 1$$

For wild-type hWAT replicates, *is_ko* is True to separate Cas9 empty and the wild-type replicates. After regression, we calculated Cook's distance for each observation. Outliers were removed if Cook's distance exceeded the median of the *F* distribution with degrees of freedom *n* and *n-p*, where *n* is the number of observations included in the regression (lipid accumulation n=12; glucose uptake n=6) and *p* is the number of predictors, including the intercept (p=2)[125]. Only one observation was flagged as an outlier, for the *SLTM* KO: cytoplasm mean fold change relative to Cas9 empty=1.618, Cook's distance=0.865. We removed this outlier and reran the regression for *SLTM* KO lipid accumulation.

We consider the KO to have a significant effect on the readout if P<0.05 for the *is_ko* term from the linear regression. Note that this is the equivalent of a two-sided *t*-test of mean readout between observations from Cas9 empty and a single-gene KO or the wild-type hWAT.

Data and code availability

Summary statistics for all phenotypes will be made available through the GWAS Catalog upon publication. All code used in this study will be made available through GitHub upon publication.

Acknowledgements

We would like to thank Dr. Val Millar and the Ebner group for expert help in the use of PerkinElmer Opera Phenix/High Content Imaging. We also acknowledge wet lab resources from Prof. Benedikt Kessler and Dr. Adan Pinto Fernandez. This research was conducted using the UK Biobank resource under application number 11867. We thank UK Biobank participants for their contribution.

Research ethics statement

This study used data from the UK Biobank. UK Biobank has approval from the North West Multi-centre Research Ethics Committee (MREC) as a Research Tissue Bank (RTB) approval. This approval means that researchers do not require separate ethical clearance and can operate under the RTB approval. This study used a human preadipocyte immortalised clonal cell line derived by Xue *et al.*[122], whose study followed the institutional guidelines of and was

approved by the Human Studies Institutional Review Boards of Beth Israel Deaconess Medical Center and Joslin Diabetes Center.

Funding statement

NAB is supported by the Pembroke College Oxford-Bendich Graduate Scholarship, the Clarendon Fund, and Wellcome Trust Grant Number 224890/Z/21/Z. MC is the Weissman Family MGH Research Scholar Award and supported by the Novo Nordisk Foundation (NNF21SA0072102). CML is supported by the Li Ka Shing Foundation, NIHR Oxford Biomedical Research Centre, Oxford, NIH (1P50HD104224-01), Gates Foundation (INV-024200), and a Wellcome Trust Investigator Award (221782/Z/20/Z). This research was supported by the Wellcome Trust Core Award Grant Number 203141/Z/16/Z with additional support from the NIHR Oxford BRC. The views expressed are those of the authors and not necessarily those of the NHS, the NIHR or the Department of Health.

Competing interests statement

CML reports grants from Bayer AG and Novo Nordisk, has a partner who works at Vertex, is a part-time employee of PHP, and owns equity in PHP and its subsidiaries. The other authors declare no competing interests.

References

1. GBD 2015 Obesity Collaborators *et al.* Health Effects of Overweight and Obesity in 195 Countries over 25 Years. *N. Engl. J. Med.* **377**, 13–27 (2017).
2. WHO | Obesity and overweight. *WHO | Obesity and overweight* http://www.who.int/mediacentre/factsheets/fs311/en/ (2018).
3. Heymsfield, S. B. & Wadden, T. A. Mechanisms, Pathophysiology, and Management of Obesity. *N. Engl. J. Med.* **376**, 254–266 (2017).
4. Jayedi, A., Soltani, S., Zargar, M. S., Khan, T. A. & Shab-Bidar, S. Central fatness and risk of all cause mortality: systematic review and dose-response meta-analysis of 72 prospective cohort studies. *BMJ* **370**, m3324 (2020).
5. Després, J.-P. & Lemieux, I. Abdominal obesity and metabolic syndrome. *Nature* **444**, 881–887 (2006).

6. Zammouri, J. *et al.* Molecular and Cellular Bases of Lipodystrophy Syndromes. *Front. Endocrinol.* **12**, 803189 (2021).
7. 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 men1–3. *Am. J. Clin. Nutr.* **81**, 555–563 (2005).
8. Shungin, D. *et al.* New genetic loci link adipose and insulin biology to body fat distribution. *Nature* **518**, 187–196 (2015).
9. 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).
10. 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).
11. Rose, K. M., Newman, B., Mayer-Davis, E. J. & Selby, J. V. Genetic and behavioral determinants of waist-hip ratio and waist circumference in women twins. *Obes. Res.* **6**, 383–392 (1998).
12. Locke, A. E. *et al.* Genetic studies of body mass index yield new insights for obesity biology. *Nature* **518**, 197–206 (2015).
13. Garvey, W. T. *et al.* Two-year effects of semaglutide in adults with overweight or obesity: the STEP 5 trial. *Nat. Med.* **28**, 2083–2091 (2022).
14. Wadden, T. A. *et al.* Tirzepatide after intensive lifestyle intervention in adults with overweight or obesity: the SURMOUNT-3 phase 3 trial. *Nat. Med.* **29**, 2909–2918 (2023).
15. Kim, K. S. *et al.* GLP-1 increases preingestive satiation via hypothalamic circuits in mice and humans. *Science* eadj2537 (2024).
16. Pulit, S. L. *et al.* Meta-analysis of genome-wide association studies for body fat distribution in 694 649 individuals of European ancestry. *Hum. Mol. Genet.* **28**, 166–174 (2019).
17. Chakhtoura, M. *et al.* Pharmacotherapy of obesity: an update on the available medications

and drugs under investigation. *EClinicalMedicine* **58**, 101882 (2023).

18. Semple, R. K., Savage, D. B., Cochran, E. K., Gorden, P. & O'Rahilly, S. Genetic syndromes of severe insulin resistance. *Endocr. Rev.* **32**, 498–514 (2011).
19. Garg, A. Acquired and inherited lipodystrophies. *N. Engl. J. Med.* **350**, 1220–1234 (2004).
20. Akbari, P. *et al.* Sequencing of 640,000 exomes identifies *GPR75* variants associated with protection from obesity. *Science* **373**, (2021).
21. Akbari, P. *et al.* Multiancestry exome sequencing reveals INHBE mutations associated with favorable fat distribution and protection from diabetes. *Nat. Commun.* **13**, 4844 (2022).
22. Karczewski, K. J. *et al.* The mutational constraint spectrum quantified from variation in 141,456 humans. *Nature* **581**, 434–443 (2020).
23. Rentzsch, P., Witten, D., Cooper, G. M., Shendure, J. & Kircher, M. CADD: predicting the deleteriousness of variants throughout the human genome. *Nucleic Acids Res.* **47**, D886–D894 (2019).
24. Ioannidis, N. M. *et al.* REVEL: An Ensemble Method for Predicting the Pathogenicity of Rare Missense Variants. *Am. J. Hum. Genet.* **99**, 877–885 (2016).
25. Jaganathan, K. *et al.* Predicting Splicing from Primary Sequence with Deep Learning. *Cell* **176**, 535–548.e24 (2019).
26. Weeks, E. M. *et al.* Leveraging polygenic enrichments of gene features to predict genes underlying complex traits and diseases. *Nat. Genet.* **55**, 1267–1276 (2023).
27. Benjamini, Y. & Hochberg, Y. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *J. R. Stat. Soc. Series B Stat. Methodol.* **57**, 289–300 (1995).
28. Loos, R. J. F. & Yeo, G. S. H. The genetics of obesity: from discovery to biology. *Nat. Rev. Genet.* **23**, 120–133 (2022).
29. Zhao, Y. *et al.* Protein-truncating variants in BSN are associated with severe adult-onset obesity, type 2 diabetes and fatty liver disease. *Nat. Genet.* **56**, 579–584 (2024).

30. Kaisinger, L. R. *et al.* Large-scale exome sequence analysis identifies sex- and age-specific determinants of obesity. *Cell Genom* **3**, 100362 (2023).
31. Karczewski, K. J. *et al.* Systematic single-variant and gene-based association testing of thousands of phenotypes in 394,841 UK Biobank exomes. *Cell Genom* **2**, 100168 (2022).
32. Bulik-Sullivan, B. *et al.* An atlas of genetic correlations across human diseases and traits. *Nat. Genet.* **47**, 1236–1241 (2015).
33. UKBB Genetic Correlation. https://ukbb-rg.hail.is/rg_browser/.
34. Martin, J. *et al.* Examining Sex-Differentiated Genetic Effects Across Neuropsychiatric and Behavioral Traits. *Biol. Psychiatry* **89**, 1127–1137 (2021).
35. Sun, B. B. *et al.* Plasma proteomic associations with genetics and health in the UK Biobank. *Nature* **622**, 329–338 (2023).
36. Zaghlool, S. B. *et al.* Revealing the role of the human blood plasma proteome in obesity using genetic drivers. *Nat. Commun.* **12**, 1279 (2021).
37. Goudswaard, L. J. *et al.* Effects of adiposity on the human plasma proteome: observational and Mendelian randomisation estimates. *Int. J. Obes.* **45**, 2221–2229 (2021).
38. Subramanian, A. *et al.* Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl. Acad. Sci. U. S. A.* **102**, 15545–15550 (2005).
39. Furuhashi, M. & Hotamisligil, G. S. Fatty acid-binding proteins: role in metabolic diseases and potential as drug targets. *Nat. Rev. Drug Discov.* **7**, 489–503 (2008).
40. Hong, E.-G. *et al.* Interleukin-10 prevents diet-induced insulin resistance by attenuating macrophage and cytokine response in skeletal muscle. *Diabetes* **58**, 2525–2535 (2009).
41. Lee, J. *et al.* The association between dietary sodium intake and obesity in adults by sodium intake assessment methods: a review of systematic reviews and re-meta-analysis. *Nutr. Res. Pract.* **17**, 175–191 (2023).
42. Grünberg, J. R. *et al.* CCN5/WISP2 and metabolic diseases. *J. Cell Commun. Signal.* **12**,

309–318 (2018).

43. Woeller, C. F. *et al.* Thy1 (CD90) controls adipogenesis by regulating activity of the Src family kinase, Fyn. *FASEB J.* **29**, 920–931 (2015).
44. Murakami, K. *et al.* Antiobesity Action of ACAM by Modulating the Dynamics of Cell Adhesion and Actin Polymerization in Adipocytes. *Diabetes* **65**, (2016).
45. Kohyama, M. *et al.* Monocyte infiltration into obese and fibrilized tissues is regulated by PILRα. *Eur. J. Immunol.* **46**, 1214–1223 (2016).
46. Dhindsa, R. S. *et al.* Rare variant associations with plasma protein levels in the UK Biobank. *Nature* **622**, 339–347 (2023).
47. Rosen, E. D. *et al.* PPARγ Is Required for the Differentiation of Adipose Tissue In Vivo and In Vitro. *Mol. Cell* **4**, 611–617 (1999).
48. He, W. *et al.* Adipose-specific peroxisome proliferator-activated receptor γ knockout causes insulin resistance in fat and liver but not in muscle. *Proceedings of the National Academy of Sciences* **100**, 15712–15717 (2003).
49. Jones, J. R. *et al.* Deletion of PPARγ in adipose tissues of mice protects against high fat diet-induced obesity and insulin resistance. *Proceedings of the National Academy of Sciences* **102**, 6207–6212 (2005).
50. Liao, W. *et al.* Suppression of PPAR-γ attenuates insulin-stimulated glucose uptake by affecting both GLUT1 and GLUT4 in 3T3-L1 adipocytes. *American Journal of Physiology-Endocrinology and Metabolism* (2007) doi:10.1152/ajpendo.00695.2006.
51. Kamble, P. G. *et al.* Proof-of-concept for CRISPR/Cas9 gene editing in human preadipocytes: Deletion of FKBP5 and PPARG and effects on adipocyte differentiation and metabolism. *Sci. Rep.* **10**, 10565 (2020).
52. Laber, S. *et al.* Discovering cellular programs of intrinsic and extrinsic drivers of metabolic traits using LipocyteProfiler. *Cell Genom* **3**, 100346 (2023).
53. Tansey, J. T. *et al.* Perilipin ablation results in a lean mouse with aberrant adipocyte

lipolysis, enhanced leptin production, and resistance to diet-induced obesity. *Proc. Natl. Acad. Sci. U. S. A.* **98**, (2001).

54. Sohn, J. H. *et al.* Perilipin 1 (Plin1) deficiency promotes inflammatory responses in lean adipose tissue through lipid dysregulation. *J. Biol. Chem.* **293**, 13974–13988 (2018).
55. Martinez-Botas, J. *et al.* Absence of perilipin results in leanness and reverses obesity in Leprdb/db mice. *Nat. Genet.* **26**, 474–479 (2000).
56. Lyu, Y. *et al.* Defective differentiation of adipose precursor cells from lipodystrophic mice lacking perilipin 1. *PLoS One* **10**, e0117536 (2015).
57. Liu, S. *et al.* Development of hypertrophic cardiomyopathy in perilipin-1 null mice with adipose tissue dysfunction. *Cardiovasc. Res.* **105**, 20–30 (2014).
58. Sakaguchi, M. *et al.* Adipocyte Dynamics and Reversible Metabolic Syndrome in Mice with an Inducible Adipocyte-Specific Deletion of the Insulin Receptor. *Cell Metab.* **25**, 448–462 (2017).
59. Blüher, M. *et al.* Adipose Tissue Selective Insulin Receptor Knockout Protects against Obesity and Obesity-Related Glucose Intolerance. *Dev. Cell* **3**, 25–38 (2002).
60. Insulin receptor-mediated signaling regulates pluripotency markers and lineage differentiation. *Molecular Metabolism* **18**, 153–163 (2018).
61. Groeneveld, M. P., Brierley, G. V., Rocha, N. M., Siddle, K. & Semple, R. K. Acute knockdown of the insulin receptor or its substrates Irs1 and 2 in 3T3-L1 adipocytes suppresses adiponectin production. *Sci. Rep.* **6**, 21105 (2016).
62. Araki, E. *et al.* Alternative pathway of insulin signalling in mice with targeted disruption of the IRS-1 gene. *Nature* **372**, 186–190 (1994).
63. Miki, H. *et al.* Essential Role of Insulin Receptor Substrate 1 (IRS-1) and IRS-2 in Adipocyte Differentiation. *Mol. Cell. Biol.* (2001) doi:10.1128/MCB.21.7.2521-2532.2001.
64. Kubota, N. *et al.* Differential hepatic distribution of insulin receptor substrates causes selective insulin resistance in diabetes and obesity. *Nat. Commun.* **7**, 1–16 (2016).

65. Withers, D. J. *et al.* Disruption of IRS-2 causes type 2 diabetes in mice. *Nature* **391**, 900–904 (1998).

66. Eguez, L. *et al.* Full intracellular retention of GLUT4 requires AS160 Rab GTPase activating protein. *Cell Metab.* **2**, 263–272 (2005).

67. Kramer, H. F. *et al.* Distinct Signals Regulate AS160 Phosphorylation in Response to Insulin, AICAR, and Contraction in Mouse Skeletal Muscle. *Diabetes* **55**, 2067–2076 (2006).

68. Lansey, M. N., Walker, N. N., Hargett, S. R., Stevens, J. R. & Keller, S. R. Deletion of Rab GAP AS160 modifies glucose uptake and GLUT4 translocation in primary skeletal muscles and adipocytes and impairs glucose homeostasis. *American Journal of Physiology-Endocrinology and Metabolism* (2012) doi:10.1152/ajpendo.00316.2012.

69. Wang, H. Y. *et al.* AS160 deficiency causes whole-body insulin resistance via composite effects in multiple tissues. *Biochem. J* **449**, 479–489 (2013).

70. Chadt, A. *et al.* Deletion of Both Rab-GTPase–Activating Proteins TBC1D1 and TBC1D4 in Mice Eliminates Insulin- and AICAR-Stimulated Glucose Transport. *Diabetes* **64**, 746–759 (2014).

71. Mikłosz, A. *et al.* RabGAP AS160/TBC1D4 deficiency increases long-chain fatty acid transport but has little additional effect on obesity and metabolic syndrome in ADMSCs-derived adipocytes of morbidly obese women. *Front. Mol. Biosci.* **10**, 1232159 (2023).

72. Moltke, I. *et al.* A common Greenlandic TBC1D4 variant confers muscle insulin resistance and type 2 diabetes. *Nature* **512**, 190–193 (2014).

73. Hart, T. *et al.* Evaluation and Design of Genome-Wide CRISPR/SpCas9 Knockout Screens. *G3 Genes|Genomes|Genetics* **7**, 2719–2727 (2017).

74. Richard, A. J., White, U., Elks, C. M. & Stephens, J. M. *Adipose Tissue: Physiology to Metabolic Dysfunction*. (MDText.com, Inc., 2020).

75. Qiu, B. & Simon, M. C. BODIPY 493/503 Staining of Neutral Lipid Droplets for Microscopy and Quantification by Flow Cytometry. *Bio Protoc* **6**, (2016).
76. Yamada, K. *et al.* Measurement of glucose uptake and intracellular calcium concentration in single, living pancreatic beta-cells. *J. Biol. Chem.* **275**, (2000).
77. Inoue, M., Akama, T., Jiang, Y. & Chun, T.-H. The Exocyst Complex Regulates Free Fatty Acid Uptake by Adipocytes. *PLoS One* **10**, e0120289 (2015).
78. Lee, J. W., Choi, H. S., Gyuris, J., Brent, R. & Moore, D. D. Two classes of proteins dependent on either the presence or absence of thyroid hormone for interaction with the thyroid hormone receptor. *Mol. Endocrinol.* **9**, (1995).
79. Ma, Y. *et al.* Adipocyte Thyroid Hormone β Receptor–Mediated Hormone Action Fine-tunes Intracellular Glucose and Lipid Metabolism and Systemic Homeostasis. *Diabetes* **72**, 562–574 (2023).
80. Pramfalk, C., Pedrelli, M. & Parini, P. Role of thyroid receptor β in lipid metabolism. *Biochim. Biophys. Acta* **1812**, (2011).
81. Harrison, S. A. *et al.* Resmetirom for nonalcoholic fatty liver disease: a randomized, double-blind, placebo-controlled phase 3 trial. *Nat. Med.* **29**, 2919–2928 (2023).
82. Huang, G. *et al.* α3(V) collagen is critical for glucose homeostasis in mice due to effects in pancreatic islets and peripheral tissues. *J. Clin. Invest.* **121**, (2011).
83. Zhang, Z., Zhan, X., Kim, B. & Wu, J. A proteomic approach identifies SAFB-like transcription modulator (SLTM) as a bidirectional regulator of GLI family zinc finger transcription factors. *J. Biol. Chem.* **294**, 5549–5561 (2019).
84. Pospisilik, J. A. *et al.* Drosophila genome-wide obesity screen reveals hedgehog as a determinant of brown versus white adipose cell fate. *Cell* **140**, 148–160 (2010).
85. Kanehisa, M. & Goto, S. KEGG: Kyoto Encyclopedia of Genes and Genomes. *Nucleic Acids Res.* **28**, 27–30 (2000).
86. Jassal, B. *et al.* The reactome pathway knowledgebase. *Nucleic Acids Res.* **48**, (2020).

87. Aleksander, S. A. *et al.* The Gene Ontology knowledgebase in 2023. *Genetics* **224**, iyad031 (2023).

88. Luo, Y. *et al.* Runx1 regulates osteogenic differentiation of BMSCs by inhibiting adipogenesis through Wnt/β-catenin pathway. *Arch. Oral Biol.* **97**, (2019).

89. de Winter, T. J. J. & Nusse, R. Running Against the Wnt: How Wnt/β-Catenin Suppresses Adipogenesis. *Frontiers in Cell and Developmental Biology* **9**, (2021).

90. Nueda, M.-L. *et al.* DLK proteins modulate NOTCH signaling to influence a brown or white 3T3-L1 adipocyte fate. *Sci. Rep.* **8**, 1–16 (2018).

91. Huang, L.-Y., Chiu, C.-J., Hsing, C.-H. & Hsu, Y.-H. Interferon Family Cytokines in Obesity and Insulin Sensitivity. *Cells* **11**, (2022).

92. Zhao, M. & Chen, X. Endocrine and Metabolic Dysfunction during Aging and Senescence: Effect of lipopolysaccharides on adipogenic potential and premature senescence of adipocyte progenitors. *American Journal of Physiology - Endocrinology and Metabolism* **309**, E334 (2015).

93. Iizuka, K., Bruick, R. K., Liang, G., Horton, J. D. & Uyeda, K. Deficiency of carbohydrate response element-binding protein (ChREBP) reduces lipogenesis as well as glycolysis. *Proc. Natl. Acad. Sci. U. S. A.* **101**, 7281–7286 (2004).

94. Oh, A. R. *et al.* ChREBP deficiency leads to diarrhea-predominant irritable bowel syndrome. *Metabolism* **85**, 286–297 (2018).

95. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, 1–21 (2014).

96. Zhou, W. *et al.* Efficiently controlling for case-control imbalance and sample relatedness in large-scale genetic association studies. *Nat. Genet.* **50**, 1335–1341 (2018).

97. Zhou, W. *et al.* SAIGE-GENE+ improves the efficiency and accuracy of set-based rare variant association tests. *Nat. Genet.* **54**, 1466–1469 (2022).

98. Mbatchou, J. *et al.* Computationally efficient whole-genome regression for quantitative and

binary traits. *Nat. Genet.* **53**, 1097–1103 (2021).

99. Wang, Q. *et al.* Rare variant contribution to human disease in 281,104 UK Biobank exomes. *Nature* **597**, 527–532 (2021).
100. Cirulli, E. T. & Goldstein, D. B. Uncovering the roles of rare variants in common disease through whole-genome sequencing. *Nat. Rev. Genet.* **11**, 415–425 (2010).
101. Degerman, E. *et al.* From PDE3B to the regulation of energy homeostasis. *Curr. Opin. Pharmacol.* **11**, 676 (2011).
102. Zhao, Y. *et al.* GIGYF1 loss of function is associated with clonal mosaicism and adverse metabolic health. *Nat. Commun.* **12**, 1–6 (2021).
103. Bycroft, C. *et al.* The UK Biobank resource with deep phenotyping and genomic data. *Nature* **562**, 203–209 (2018).
104. 1000 Genomes Project Consortium *et al.* A global reference for human genetic variation. *Nature* **526**, 68–74 (2015).
105. Asymptotic properties of principal component analysis and shrinkage-bias adjustment under the generalized spiked population model. *J. Multivar. Anal.* **173**, 145–164 (2019).
106. [No title]. https://cran.r-project.org/web/packages/randomForest/randomForest.pdf.
107. hail 0.2 documentation. https://hail.is/docs/0.2/.
108. McLaren, W. *et al.* The Ensembl Variant Effect Predictor. *Genome Biol.* **17**, 122 (2016).
109. Liu, X., Li, C., Mou, C., Dong, Y. & Tu, Y. dbNSFP v4: a comprehensive database of transcript-specific functional predictions and annotations for human nonsynonymous and splice-site SNVs. *Genome Med.* **12**, 103 (2020).
110. *vep105_loftee*. (Github).
111. Morales, J. *et al.* A joint NCBI and EMBL-EBI transcript set for clinical genomics and research. *Nature* **604**, 310–315 (2022).
112. Richards, S. *et al.* Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and

Genomics and the Association for Molecular Pathology. *Genet. Med.* **17**, 405–424 (2015).

113. Chang, C. C. *et al.* Second-generation PLINK: rising to the challenge of larger and richer datasets. *Gigascience* **4**, 7 (2015).
114. Benner, C. *et al.* Prospects of Fine-Mapping Trait-Associated Genomic Regions by Using Summary Statistics from Genome-wide Association Studies. *Am. J. Hum. Genet.* **101**, 539–551 (2017).
115. Zhou, S. *et al.* Converging evidence from exome sequencing and common variants implicates target genes for osteoporosis. *Nat. Genet.* **55**, 1277–1287 (2023).
116. Kuan, V. *et al.* A chronological map of 308 physical and mental health conditions from 4 million individuals in the English National Health Service. *Lancet Digit Health* **1**, e63–e77 (2019).
117. Kaplan, E. L. & Meier, P. Nonparametric Estimation from Incomplete Observations. *J. Am. Stat. Assoc.* **53**, 457–481 (1958).
118. Therneau, T. M. survival: Survival Analysis. *CRAN: Contributed Packages* The R Foundation https://doi.org/10.32614/cran.package.survival (2001).
119. Kassambara, A., Kosinski, M. & Biecek, P. Drawing Survival Curves using 'ggplot2' [R package survminer version 0.4.8]. (2020).
120. Lex, A., Gehlenborg, N., Strobelt, H., Vuillemot, R. & Pfister, H. UpSet: Visualization of Intersecting Sets. https://doi.org/10.1109/TVCG.2014.2346248.
121. Ramos-Molina, B., Martin, M. G. & Lindberg, I. PCSK1 Variants and Human Obesity. *Prog. Mol. Biol. Transl. Sci.* **140**, 47 (2016).
122. Xue, R. *et al.* Clonal analyses and gene profiling identify genetic biomarkers of the thermogenic potential of human brown and white preadipocytes. *Nat. Med.* **21**, (2015).
123. Dobin, A. *et al.* STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15 (2013).
124. Li, B. & Dewey, C. N. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinformatics* **12**, 1–16 (2011).

125. Fox, J. & Scott Long, J. *Modern Methods of Data Analysis*. (SAGE Publications, Incorporated, 1990).