

**UCLA**

**UCLA Electronic Theses and Dissertations**

**Title**

Investigation of genomic mechanisms regulating adipose tissue function and influencing body mass index and waist-hip-ratio

**Permalink**

<https://escholarship.org/uc/item/3t02h2dq>

**Author**

Pan, David Zhi-Chao

**Publication Date**

2020

**Supplemental Material**

<https://escholarship.org/uc/item/3t02h2dq#supplemental>

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA

Los Angeles

Investigation of genomic mechanisms regulating adipose tissue function and influencing body  
mass index and waist-hip-ratio

A dissertation submitted in partial satisfaction of the  
requirements for the degree Doctor of Philosophy  
in Bioinformatics

by

David Zhi-Chao Pan

2020

© Copyright by  
David Zhi-Chao Pan  
2020

## ABSTRACT OF THE DISSERTATION

Investigation of genomic mechanisms regulating adipose tissue function and influencing body mass index and waist-hip-ratio

by

David Zhi-Chao Pan

Doctor of Philosophy in Bioinformatics

University of California, Los Angeles, 2020

Professor Päivi Pajukanta, Chair

Obesity is a well-established risk factor for multiple common disorders, such as type 2 diabetes (T2D), hypertriglyceridemia, non-alcoholic fatty liver disease (NAFLD), coronary artery disease (CAD), and certain cancers. The rates of obesity-related deaths have risen sharply globally over the last 20 years, with over 70% of adults in the United States now classified as overweight or obese according to the Centers for Disease Control (CDC). Currently, as the world faces one of the worst infectious-disease outbreaks in a century, new data are also emerging showing that obesity is a key risk factor for severe forms of COVID-19. However, the complex underlying mechanisms of obesity, especially the susceptibility genes and their regulatory mechanisms, remain elusive. To address this scientific knowledge gap, we have employed integrative multi-omics approaches on human subcutaneous adipose RNA-sequencing (RNA-seq) data from multiple cohorts; epigenomic data from chromosomal interactions and open chromatin in relevant adipose cell-types; large scale obesity genome-wide association studies (GWAS) for body mass index (BMI) and waist-hip-ratio adjusted for BMI (WHRadjBMI); and one of the largest population cohort to date, the UK Biobank (UKB).

In Chapter 2, we fine-mapped BMI GWAS loci using *cis*-expression quantitative trait loci (eQTLs) from the METabolic Syndrome In Men (METSIM) cohort and chromosomal interactions from adipocyte promoter Capture Hi-C (pChi-C). We discovered that the pChi-C interactions are enriched for central adipogenesis transcription factors (TFs), Peroxisome Proliferator Activated Receptor Gamma (*PPARG*) and CCAAT Enhancer Binding Protein Beta (*CEBPB*), and identified four key GWAS gene examples as well as 38 additional candidate genes with *cis*-eQTLs in chromosomal interactions whose gene expression are strongly associated with obesity-related traits, such as BMI, blood metabolites, and lipids.

In Chapter 3, I discuss my contribution to a study about context-specific changes in open chromatin and pChi-C interactions in human primary adipocytes and the variants in those open chromatin regions that respond to lipid intake. Using these context-specific molecular data, we provide candidate gene-environment interaction (GxE) variants that significantly alter TF motifs in open chromatin regions, which are evolutionarily conserved and have a key role in adipogenesis and the responses to lipid intake. These candidate GxE variants with molecular priors were then tested for interactions with saturated fat intake on obesity in the UKB, resulting in the discovery of novel GxE variants for obesity.

In Chapter 4, we move beyond *cis*-eQTLs, to *trans*-eQTLs and master *trans* regulatory TFs that control adipose co-expression networks important for obesity. To advance the discovery of unknown genetic and molecular mechanisms regulating abdominal adiposity and the sex-specific distribution of body fat, we searched for genetic master *trans* regulators of WHRadjBMI by employing integrative genomics approaches on human adipose RNA-seq data and WHRadjBMI GWAS. We provide novel genomic evidence, verified by our functional

knockdown studies in human primary preadipocytes, for the causal role of the TF, T-Box Transcription Factor 15 (*TBX15*), in controlling accumulation of abdominal fat and adiposity.

All in all, we have combined these omics and phenotype data using computational and functional techniques to identify genes and their regulatory mechanisms affecting obesity. Our studies suggest that by integrating the multi-omics data and elucidating the mechanisms underlying obesity, we can further the understanding of the risks associated with obesity and its comorbidities to move personalized medicine forward.

The dissertation of David Zhi-Chao Pan is approved.

Janet S. Sinsheimer

Roel A. Ophoff

Jason Ernst

Päivi Pajukanta, Committee Chair

University of California, Los Angeles

2020

To my wife, Yi-Ching Hsieh, and my parents, Shih-Hsie Pan and Lan-Jen Tsai, and my sisters,

Margaret Pan and Alice Pan



## TABLE OF CONTENTS

List of Tables	.....	viii
List of Figures	.....	xi
Acknowledgements	.....	xiii
Vita	.....	xv
Glossary	.....	xvi
Chapter 1	Introduction .....	1
	References .....	16
Chapter 2	Integration of human adipocyte chromosomal interactions with adipose gene expression prioritizes obesity-related genes from GWAS .....	21
	References .....	31
Chapter 3	Reverse gene–environment interaction approach to identify variants influencing body-mass index in humans .....	52
	References .....	64
	Supplementary references .....	97
Chapter 4	Identification of <i>TBX15</i> as an adipose master <i>trans</i> regulator of abdominal obesity genes .....	98
	References .....	117
Chapter 5	Discussion and future directions.....	159
	References.....	169

## LIST OF TABLES

Table II-1	Thirteen representative eGenes (9 most significant genes and 4 GWAS loci) that correlate with BMI in METSIM and TwinsUK (for the full data on all 54 genes, see Supplementary Table 6) .....	26
Table II-S1	Parameters used for identification of novel cis-eQTL and looping interactions .....	42
Table II-S2	Histone mark enrichment in looping HindIII fragments in primary HWA .....	43
Table II-S3	Adipocyte chromosomal interactions are enriched for 30 transcription factors (adjusted $p < 0.05$ ) when compared to CD34+ chromosomal interactions .....	44
Table II-S4	LD score enrichments, heritability estimates, and p-values using the published LD Score software .....	45
Table II-S5	LD score enrichments, heritability estimates, and p-values after modification of the LD score software .....	46
Table II-S6	Fifty-four eGenes in METSIM, including the 42 genes replicated for correlation with BMI and effect direction in TwinsUK .....	47
Table II-S7	The 42 replicated BMI-correlated eGenes show significant enrichment for metabolic and inflammatory pathways using KEGG pathway analysis as implemented in WebGestalt .....	48
Table II-S8	DeepSEA analysis of the variants in the MAP2K5 locus supports the functionality of the looping cis-eQTL SNP rs4776984 .....	49
Table II-S9	Significant CHiCAGO interaction and replication scores from a separate HWA Capture Hi-C experiment verify the looping cis-eQTLs for the four identified obesity-related loci .....	50
Table II-S10	DNA oligonucleotides used for electrophoretic mobility shift assay .....	51
Table III-1	Five lipid-responsive ATAC-seq peaks in interacting promoters overlap with GWAS SNPs for serum lipid traits .....	58
Table III-2	Significant G×E interactions affecting BMI from a multivariable linear model for 290 promoter SNPs in lipid-responsive ATAC-seq peaks .....	60
Table III-S1	Significant G×E interactions affecting BMI from a multivariable linear model for 290 promoter SNPs in lipid-responsive ATAC-seq peaks .....	76
Table III-S2	Differentially accessible ATAC-seq peaks between human preadipocytes and adipocytes .....	77
Table III-S3	Sequencing, read processing, and QC metrics for adipocyte lipid-challenge ATAC-seq .....	78
Table III-S4	Differentially accessible ATAC-seq peaks in lipid-challenged human adipocytes .....	79
Table III-S5	Sequencing and read processing metrics for adipocyte lipid-challenge pChI-C .....	80
Table III-S6	The top 10 TF motifs enriched in adipocyte lipid-responsive open chromatin regions in chromosomal interactions .....	81

Table III-S7	154 genes with lipid-responsive promoters in chromosomal interactions in adipocytes .....	82
Table III-S8	KEGG pathway enrichment analysis of 154 genes with lipid-responsive promoters .....	83
Table III-S9	323 gene promoters physically interact with lipid-responsive enhancers in adipocytes .....	84
Table III-S10	Three lipid-responsive ATAC-peaks in interacting enhancers overlap with GWAS SNPs for serum lipid traits .....	85
Table III-S11	Lipid-responsive gene promoters with GWAS SNPs respond to SFA treatment .....	86
Table III-S12	Lipid-responsive enhancers with GWAS SNPs stratified by quality of fatty acid .....	87
Table III-S13	LDSC analysis of SNPs in cis regions of the 154 lipid-responsive promoters .....	88
Table III-S14	LDSC analysis of SNPs in cis regions of genes with lipid-responsive enhancers .....	89
Table III-S15	75 lipid-responsive peaks in gene promoters contain SNPs with MAF > 0.05 in the UK Biobank .....	90
Table III-S16	Significant GxE promoter SNPs with LD proxies .....	91
Table III-S17	142 lipid-responsive peaks in enhancers contain SNPs with MAF > 0.05 in the UK Biobank .....	92
Table III-S18	Significant G×E interactions with BMI from a multivariable linear model for 410 enhancer SNPs .....	93
Table III-S19	DeepSEA analysis of the 20 G×E SNPs in interacting lipid-responsive gene promoters .....	94
Table III-S20	DeepSEA analysis of the 26 G×E SNPs in interacting lipid-responsive enhancers .....	95
Table III-S21	EMSA oligo probes used for analysis of GxE SNP rs10788522 .....	96
Table IV-S1	Characteristics of the genes (n=347) in the WHRadjBMI co-expression network, as reported by WGCNA and ranked by network membership .....	137
Table IV-S2	Characteristics of the obesity GWAS genes in WHRadjBMI co-expression network, as reported by WGCNA and ranked by network membership .....	145
Table IV-S3	Characteristics of the adipocyte marker genes in WHRadjBMI co-expression network, as reported by WGCNA and ranked by network membership .....	147
Table IV-S4	KEGG pathway enrichment results (passing FDR<0.05) from WebGestalt for the WHRadjBMI co-expression network genes .....	148
Table IV-S5	Gene Ontology cellular component enrichment results (passing FDR<0.05) from WebGestalt for the WHRadjBMI co-expression network genes .....	149

Table IV-S6	Stratified LD Score Regression results for WHRadjBMI, T2D, and BMI using the cis variants (+/-500kb from the ends of the gene) of the WHRadjBMI co-expression network genes .....	151
Table IV-S7	Characteristics of the TFs in WHRadjBMI co-expression network, as reported by WGCNA and ranked by network membership .....	152
Table IV-S8	Significant TWAS heritability estimates (p<0.01) for the TFs in the WHRadjBMI co-expression network .....	153
Table IV-S9	TWAS p-values and Z-scores for associations of TFs (with significant TWAS heritability (p<0.01)) with WHRadjBMI .....	154
Table IV-S10	Significantly differentially expressed genes (FDR<0.05) in the WHRadjBMI co-expression network in the <i>TBX15</i> knockdown experiment ranked by p-value .....	155

## LIST OF FIGURES

Figure II-1	Open chromatin sites (DHSs) within adipocyte promoter CHi-C chromosomal interactions show significant enrichment in cis expression .....	24
Figure II-2	Overview of the study design targeted to identify new genes for obesity and related metabolic traits. A schematic illustrating the integration of multi-omics data utilized in this study to elucidate genetics of obesity-related traits .....	25
Figure II-3	Promoter Capture Hi-C enables refinement of the BMI GWAS locus that colocalizes with cis-eQTLs interacting with the target gene promoter of MAP2K5 .....	27
Figure II-4	Predicted TF motifs and electrophoretic mobility shift assay (EMSA) at the rs4776984 site indicate allele-specific binding .....	28
Figure II-S1	Modification to LD Score regression software does not show significant changes when compared with the data obtained using the published version .....	34
Figure II-S2	Overview of the study design targeted to identify causal and reactive BMI-correlated genes .....	35
Figure II-S3	Promoter Capture Hi-C enables refinement of the GWAS loci that colocalizes with cis-eQTLs interacting with the target gene promoter of ORMDL3,LACTB, and ACADS .....	36
Figure II-S4	Two independent replicates for the Electrophoretic mobility shift assay (EMSA) data show increased binding of nuclear protein extracted from primary human white adipocytes (HWA) to the alternate allele when compared to the reference allele of the MAP2K5 cis-eQTL SNP rs4776984 .....	38
Figure II-S5	Three independent replicates for the Electrophoretic mobility shift assay (EMSA) do not show a supershift when using antibody against CTCF and nuclear protein extracted from primary human white adipocytes (HWA) at the MAP2K5 cis-eQTL SNP rs4776984 .....	39
Figure II-S6	The Electrophoretic mobility shift assay (EMSA) does not show a supershift when using a different antibody against CTCF and nuclear protein extracted from primary human white adipocytes (HWA) at the MAP2K5 cis-eQTL SNP rs4776984.....	40
Figure II-S7	Three independent replicates for the Electrophoretic mobility shift assay (EMSA) do not show specific binding using purified CTCF protein at the MAP2K5 cis-eQTL SNP rs4776984.....	41
Figure III-1	ATAC-seq analysis comparing primary human preadipocytes and adipocytes indicates successful adipocyte differentiation and widespread changes in chromatin accessibility .....	54
Figure III-2	Lipid-responsive regions fall within adipocyte accessible regions of the genome, as well as within context-dependent regions that are not present in untreated adipocytes .....	55

Figure III-3	The 154 genes with lipid-responsive promoters within chromosomal interactions exhibit cross-species conservation and constraints on loss-of-function mutations, in line with their potential importance for energy homeostasis and survival .....	57
Figure III-4	A lipid-responsive open chromatin region in human primary adipocytes at the 11q12.2 FADS1–FADS2–FADS3 locus harbours GWAS SNPs for serum lipid traits .....	59
Figure III-5	Fine-mapping of the gene–diet interaction for BMI in the LDB3 promoter region .....	61
Figure III-6	Analytical approach .....	62
Figure III-S1	Adipocyte-accessible peaks fall more into adipocyte enhancers and promoters than the preadipocyte-accessible peaks or the full peak set.....	70
Figure III-S2	Fatty acid lipid challenge in human adipocytes leads to increased storage of lipids in lipid droplets .....	71
Figure III-S3	Violin plots show the distribution of log <sub>2</sub> fold-change (log <sub>2</sub> FC) for all differentially accessible peaks from the lipid challenge in adipocytes ....	72
Figure III-S4	Lipid-responsive peaks in adipocyte-accessible regions fall more into adipocyte enhancers and promoters than lipid-responsive peaks in context-dependent regions .....	73
Figure III-S5	The 323 genes with promoters that interact with lipid-responsive enhancers exhibit constraints on loss-of-function mutations.....	74
Figure III-S6	Testing all SNPs genome-wide for gene-by-saturated fat intake effect on BMI does not show inflation or result in significant GxEs at the genome-wide significance threshold .....	75
Figure IV-1	Schematic overview of the study design (a), discovery of the red WHRadjBMI-associated co-expression network that is enriched for TFs and GWAS genes (b), and enriched for upregulated adipose tissue - specific DE genes when compared to other tissues (c) in GTEx .....	132
Figure IV-2	PRS scores confirm sexual dimorphism of WHRadjBMI and demonstrate the importance of WHRadjBMI co-expression network genes for WHRadjBMI in males .....	133
Figure IV-3	TWAS and FOCUS results in GTEx v8 subcutaneous adipose RNA-seq data implicates TBX15 as the only TF in the WHRadjBMI co-expression network causal for WHRadjBMI .....	134
Figure IV-4	The EMSA results demonstrate increased protein binding at the alternate allele of the variant rs1779445, a context-specific cis regulator of TBX15 and trans regulator of the WHRadjBMI co-expression network (a, b, c) and knockdown of TBX15 in human primary preadipocytes significantly affects 130 genes (FDR<0.05) in the WHRadjBMI co-expression network (d, e) .....	135
Figure IV-S1	WGCNA identifies 2 co-expression networks in the METSIM adipose RNA-seq cohort (n=335), significantly correlated with WHRadjBMI and fasting serum insulin .....	136

## ACKNOWLEDGEMENTS

First and foremost, I would like to thank my advisor, Professor Päivi Pajukanta for her guidance and support throughout all the work presented in this thesis. I have always been inspired by her drive to pursue science and to always examine more deeply into any issue, large or small.

I would like to thank all my committee members, Professors Janet S. Sinsheimer, Roel A. Ophoff, and Jason Ernst, who have helped guide my dissertation and thesis investigations. I would especially like to thank Professor Janet S. Sinsheimer for her close collaboration and invaluable insights into all my manuscripts, which comprise this thesis.

I would also like to thank all my collaborators. Among them, I would specifically like to thank, Kristina M. Garske, Zong Miao, and Marcus Alvarez. It has been a pleasure to collaborate with them on all my projects throughout my Ph.D. and their continuous motivation has always brought projects closer to perfection. I would also like to thank Aldons J. Lusis and the members of his NIH Program Project Grant, for their suggestions and counterpoint views that have always provided new directions to take projects.

Lastly, I would like to thank my family, who have always been supportive throughout my life choices and especially throughout my Ph.D. I would especially like to thank my wife, Yi-Ching Hsieh, who has always been there supporting me throughout these five years and the longer journey to get to where we are now.

This thesis was supported by the NIH-NCI grant T32LM012424 and NIH-NIDDK grant F31 DK118865. The Genotype-Tissue Expression (GTEx) Project was supported by the Common Fund of the Office of the Director of the National Institutes of Health, and by NCI, NHGRI, NHLBI, NIDA, NIMH, and NINDS. The data used for the analyses described in this

manuscript were obtained from dbGaP accession number phs000424.v8.p2 under project 10810. This research has been conducted using the UKBiobank Resource under Application Number 3934.

Chapter 2 is a reprint of “Integration of human adipocyte chromosomal interactions with adipose gene expression prioritizes obesity-related genes from GWAS” by David Z. Pan, Kristina M. Garske, Marcus Alvarez, Yash V. Bhagat, James Boocock, Elina Nikkola, Zong Miao, Chelsea K. Raulerson, Rita M. Cantor, Mete Civelek, Craig A. Glastonbury, Kerrin S. Small, Michael Boehnke, Aldons J. Lusis, Janet S. Sinsheimer, Karen L. Mohlke, Markku Laakso, Päivi Pajukanta, and Arthur Ko.

Chapter 3 is a reprint of “Reverse gene–environment interaction approach to identify variants influencing body-mass index in humans” by Krstina M. Garske, David Z. Pan, Zong Miao, Yash V. Bhagat, Caroline Comenho, Christopher R. Robles, Jihane N. Benhammou, Marcus Alvarez, Arthur Ko, Chun Jimmie Ye, Joseph R. Pisegna, Karen L. Mohlke, Janet S. Sinsheimer, Markku Laakso, and Päivi Pajukanta.

Chapter 4 is a submitted manuscript entitled “Identification of *TBX15* as an adipose master *trans* regulator of abdominal obesity genes” by David Z. Pan, Zong Miao, Caroline Comenho, Sandhya Rajkumar, Amogha Koka, Marcus Alvarez, Dorota Kaminska, Janet S. Sinsheimer, Karen L. Mohlke, Nicholas Mancuso, Kirsi Pietiläinen, Jussi Pihlajamäki, Markku Laakso, Kristina M. Garske, and Päivi Pajukanta.



## VITA

### Education

Doctoral Candidate in Bioinformatics, University of California, Los Angeles

September 2015 – June 2020

Masters of Science in Welding Engineering, The Ohio State University

March 2011 – December 2012

Bachelors of Science in Materials Science and Engineering, University of California, Berkeley

January 2007 – June 2010

### Publications

**Pan DZ**, Miao Z, Comenho C, Rajkumar S, Koka A, Alvarez M, Ko A, Kaminska D, Sinsheimer JS, Mohlke KL, Mancuso N, Pietilainen, K, Pihlajamäki J, Laakso M, Garske KM, Pajukanta P. *Identification of TBX15 as a master trans regulator of abdominal obesity genes*. Submitted (under review).

Miao Z, Garske KM, **Pan DZ**, Koka A, Kaminska D, Sinsheimer JS, Pihlajamäki J, Pajukanta P. *Non-invasive prediction of NAFLD empowers the discovery of 90 novel genetic variants and the establishment of NAFLD PRS and causal role of NAFLD in coronary artery disease*. Submitted.

Garske KM, **Pan DZ**, Miao Z, Bhagat YV, Comenho C, Robles CR, Benhammou JN, Alvarez M, Ko A, Ye CJ, Pisegna JR, Mohlke K, Sinsheimer JS, Laakso M, Pajukanta P. *Reverse gene-environment interaction approach to identify variants influencing body-mass index in humans*. *Nature Metabolism*. 1, 630–642, 2019.

**Pan DZ**, Garske KM, Alvarez M, Bhagat YV, Boocock J, Nikkola E, Miao Z, Raulerson CK, Cantor RM, Civelek M, Glastonbury CA, Small KS, Boehnke M, Lusk AJ, Sinsheimer JS, Mohlke KL, Laakso M, Pajukanta P, Ko A. *Integration of human adipocyte chromosomal interactions with local adipose gene expression identifies obesity-related genes beyond GWAS*. *Nature Communications*. 9, 2018.

Freund MK, Burch KS, Shi H, Mancuso N, Kichaev G, Garske KM, **Pan DZ**, Miao Z, Mohlke KL, Laakso M, Pajukanta P, Pasaniuc B, Arboleda VA. *Phenotype-Specific Enrichment of Mendelian Disorder Genes near GWAS Regions across 62 Complex Traits*. *The American Journal of Human Genetics*. 103(4):535-552, 2018.

## GLOSSARY OF TERMS AND ABBREVIATIONS

3C – Chromosome Confirmation Capture

5C – Chromosome confirmation capture carbon copy

ADIPOQ - Adiponectin

ATAC-seq – Assay for transposase-accessible chromatin using sequencing

BAT – Brown adipose tissue

BMI – Body mass index

CAD – Coronary artery disease

CEBPB – CCAAT Enhancer Binding Protein Beta

ChICAGO – Capture Hi-C Analysis of Genomic Organisation

ChIP-seq – Chromatin immunoprecipitation sequencing

CTCF – CCCTC-Binding Factor

EHR – Electronic health record

EMSA – Electrophoretic mobility shift assay

EP300 - E1A Binding Protein P300

eQTL – Expression quantitative trait locus

FOCUS – Fine mapping of causal sets

GTE<sub>x</sub> – Genotype expression cohort

GWAS – Genome wide association study

GxE – Gene-environment interaction

Hi-C – High throughput chromosome confirmation capture

KLF14 – Kruppel Like Factor 14

KLF15 – Kruppel Like Factor 15

LD – Linkage disequilibrium

LoF – Loss of function

MAF – Minor allele frequency

MAP2K5 - Mitogen-Activated Protein Kinase Kinase 5

METSIM – METabolic Syndrome In Men cohort

NAFLD – Non-alcoholic fatty liver disease

NIH dbGAP – National Institutes of Health database of Genotypes and Phenotypes

PPARA – Peroxisome Proliferator Activated Receptor Alpha

PPARG – Peroxisome Proliferator Activated Receptor Gamma

PC – Principal component

pChI-C – Promoter Capture Hi-C

PRS – Polygenic risk score

PWM – Position weight matrix

q-PCR – Quantitative PCR

RNA-seq – RNA-sequencing

RREB1 – Ras-responsive element binding protein 1

RXRA - Retinoid X Receptor Alpha

SAT – Subcutaneous adipose tissue

SNP – Single nucleotide polymorphisms

T2D – Type 2 diabetes

TBX15 – T-box Transcription Factor 15

TF – Transcription factor

TFBS – Transcription factor binding site

TWAS – Transcriptome wide association study

UKB – UK Biobank

VAT – Visceral adipose tissue

WAT – White adipose tissue

WES – Whole exome sequencing

WGCNA – Weighted gene co-expression network analysis

WGS – Whole genome sequencing

WHRadjBMI – Waist-hip-ratio adjusted for body mass index

ZNF800 – Zinc finger protein 800

# Chapter 1

Introduction

## 1.1 Adipose biology and obesity

As the world faces one of the worst infectious-disease outbreaks in a century, new data are emerging showing that obesity is a key risk factor for severe forms of COVID-19 infection in individuals less than 60 years of age<sup>1,2</sup>. Obesity is clinically diagnosed by a body mass index (BMI) greater than 30 kg/m<sup>2</sup>, while severe obesity is defined as BMI greater than 40 kg/m<sup>2</sup>. One of the key tissues involved in obesity is adipose tissue. Although originally thought of as simply a storage organ, adipose is now recognized as an important endocrine regulator of energy homeostasis in the human body through secretion of hormones, cytokines, and metabolites (known as adipokines)<sup>3</sup>. Over the years, adipokines, such as leptin<sup>4</sup> and adiponectin<sup>5,6</sup>, as well as many others<sup>7-10</sup>, have been shown to be secreted by adipose tissue and have effects on obesity.

Many other organisms and human infants have both white adipose tissue (WAT), for the storage of lipids, and brown adipose tissue (BAT), for thermoregulation; however, adult humans have lost most BAT deposits in the body, except for small deposits near the neck<sup>3</sup>. Further subdividing the types of adipose tissue, WAT has two main types of deposits, subcutaneous WAT (SAT) and visceral WAT (VAT). While VAT is thought to be more metabolically important than SAT, with links to type 2 diabetes (T2D) and insulin resistance<sup>11,12</sup>, SAT displays larger changes in weight<sup>13</sup> and is easier to acquire through less invasive biopsies from living individuals.

To further complicate studies of WAT and obesity, as individuals become obese, WAT undergoes two types of expansion: hypertrophy or hyperplasia, both of which have been shown to be regulated by genetic as well as environmental factors<sup>14</sup>. Hyperplasia represents the healthy increase in the number of adipocytes to store lipids, which drives metabolic health and temporally even metabolically healthy obesity. This process, which includes differentiation of

preadipocytes to mature adipocytes, is controlled by adipogenesis master regulators, such as peroxisome proliferator-activated receptor gamma (*PPARG*) and CCAAT enhancer binding protein (*CEBPB*)<sup>15</sup>. Hypertrophy, on the other hand, is characterized by an increase in adipocyte size instead of number, which leads to large adipocytes storing fat and expanding in size until they burst. When adipocytes burst, this triggers an immune response, initiated by macrophages and then furthered by other immune cells, such as neutrophils and T cells, to clean up the cell debris. This infiltration of immune cells leads to an increase in the release of inflammatory cytokines, resulting in obese individuals' WAT residing in a continuously inflamed state, altering the adipose tissue function and leading to comorbidities of obesity, such as impaired insulin sensitivity<sup>16-19</sup>. The molecular mechanisms underlying the differences between individuals who undergo hypertrophy versus hyperplasia and the gene regulatory mechanisms altered by obesity during adipogenesis are still incompletely understood.

## **1.2 Genomic regulatory landscape and chromosomal interactions**

As the rapid development of next-generation sequencing techniques has decreased the price and increased the efficiency of acquiring gene expression information<sup>20</sup>, it provides the opportunity to increase our understanding of the regulatory mechanisms that control gene expression<sup>21</sup>. When focusing on the genes that have the largest amount of prior information, i.e. the protein coding genes, one of the most common regulatory mechanisms consists of DNA binding proteins, such as transcription factors (TFs), binding to DNA motifs (TF binding sites (TFBSs)), thus activating or repressing gene promoters, which in turn, raises or lowers the gene expression levels, respectively<sup>22</sup>. There are TFs that are known to be master regulators, which control specific key pathways within cells, such as those related to immune response<sup>23</sup>, as well as those that drive cell differentiation, such as *PPARG* and *CEBPB* in adipocytes<sup>15</sup>. Experimental methods of assessing

TF binding and TFBSs, including chromatin immunoprecipitation sequencing (ChIP-seq), involve crosslinking TFs with their bound sequences, cutting the DNA with exonucleases, pulling down the protein of interest with a specific antibody, and finally sequencing the DNA to which the protein is bound to find the particular TFBS sequence<sup>23</sup>. Computational methods have complimented this technology by analyzing the sequences produced by many ChIP-seq experiments, and finding canonical TFBSs<sup>24</sup>. These computationally defined TFBSs consist of matrices (position weight matrix (PWM)), indicating how likely a specific DNA nucleotide is to appear at a certain position in the TFBS, with extensive databases curated of these PWMs (TRANSFEC<sup>25</sup>; JASPAR<sup>26</sup>). However, when scanning the entire human genome using these PWMs, many more statistically significant locations appear across the human genome than are confirmed by ChIP-seq experiments, indicating that TF binding to DNA is pervasive and depends on a multitude of complex factors beyond the recognition of a particular DNA sequence<sup>27</sup>.

TFs usually bind in complexes made up of multiple proteins binding to DNA together, partially explaining the differences between computationally defined TFBSs and experimental findings<sup>21</sup>. This fact also indicates that large regions of the human genome consisting of many TFBSs may also have a role in gene regulation. These large regions are collectively known as regulatory elements, usually classified into two main categories: enhancers and repressors, which increase or decrease gene expression, respectively<sup>28</sup>. Regulatory elements are usually defined by the modifications on the histones, the proteins that affect the organization of DNA into its most basic unit, the nucleosome<sup>29,30</sup>. ChIP-seq experiments, along with identifying TFBSs, can identify regions of DNA that are wrapped around nucleosomes with certain marks, such as H3K4me1 and H3K27ac at active enhancers<sup>29</sup>. Computationally, aggregating these ChIP-seq



experiment data across many cell-types and cell lines can provide a genome-wide picture of regulatory elements, such as those from ChromHMM<sup>31</sup>. These computationally identified regulatory elements can be combined with information about the three dimensional organization of the genome and TFBSs to help identify regions of the genome with enhancing and repressing effects<sup>32</sup>.

The organization of these regulatory elements and the three-dimensional confirmation of DNA is of great importance for gene regulation, and changes to the three-dimensional confirmation of DNA may have direct effects on gene expression<sup>33</sup>. The first modern interaction assay, called chromosome confirmation capture (3C), was invented in 2002<sup>34</sup>, and it assesses the contact frequency between two genomic regions, making up one chromosomal interaction, at a time. The general experimental procedure consists of cross-linking DNA, using a restriction enzyme to cut the DNA, then re-ligating the cut ends, and finally using targeted quantitative PCR (q-PCR) of the re-ligated fragment of DNA of interest. The subsequent technologies, named 4C<sup>35-38</sup>, capturing all other interacting genomic loci interacting with a specific genomic loci of interest; chromosome confirmation capture carbon copy (5C)<sup>39,40</sup>, capturing all pairwise interacting loci across a few mega bases (Mb) of the genome; and finally High throughput chromosome confirmation capture (Hi-C)<sup>41</sup>, capturing all pairwise interacting loci genome-wide, continued the trend of assessing an increasing number of interactions across the genome in conjunction with the use of next-generation sequencing technologies. However, the low resolution of high throughput techniques, such as Hi-C (generally genomic regions of 5kb-25kb in length interacting with each other), and the extremely high number of sequencing reads (>1x10<sup>9</sup> reads) required to identify high resolution interactions<sup>42</sup>, has limited the use of Hi-C and driven more targeted technologies to identify interactions between specific genomic regions. One

such technique, known as promoter Capture-HiC (pCHiC)<sup>43</sup>, adds an additional step to the experimental protocol for Hi-C, using targeted biotinylated RNA probes to target genomic loci containing promoters. This targeted approach, allows for the capture of all interacting loci with known gene promoters, and reduces the number of sequencing reads required by having a lower number of total interactions than Hi-C. Since genes and their promoters are arguably the better understood part of the human genome, identifying chromosomal interactions involving gene promoters, is useful for studies investigating mechanisms of gene regulation in relation to important biological pathways relevant to specific tissue types, such as adipose and adipocytes. Currently, pCHi-C is mostly used to study local, *cis* chromosomal interactions (distance <1Mb), while identification of biologically relevant longer range, *trans* chromosomal interactions (distance >1Mb or inter-chromosomal), has remained challenging. These challenges stem from the fact that in most pCHi-C experiments sequencing depth is restricted by the cost of sequencing. Therefore, *trans* chromosomal interactions have fewer sequencing reads than *cis* chromosomal interactions as they in general occur less frequently<sup>44</sup>. Therefore, they do not pass computationally set thresholds for identifying chromosomal interactions despite lower thresholds for *trans* chromosomal interactions when compared to *cis* chromosomal interactions. In addition, except in model organisms, *trans* interactions have not been proven to exist yet using genome-wide assays<sup>45,46</sup>. Therefore, current computational methods to analyze pCHi-C data treat *trans* chromosomal interactions as a measure of the background noise in pCHi-C experiments, thus warranting further methods developments to identify robust *trans* chromosomal interactions.

### **1.3 Genome-wide association studies and genetic architecture of obesity**

Common DNA variants and their associations with disease, such as obesity, have been captured in genome-wide association studies (GWAS), albeit with a stringent threshold on association ( $p < 5 \times 10^{-8}$ ), to account for the common independent loci in the human genome<sup>47</sup>. This widely used genome-wide significance threshold is based on extrapolations of the number of independent variants in 10 highly genotyped regions of the human genome using European population cohort samples in 2005<sup>48</sup>. While a significant proportion of the phenotypic variation in BMI is attributed to genetic variation (heritability of BMI ~0.4-0.7)<sup>49</sup>, understanding the mechanisms underlying this heritable component has been challenging. The 97 loci identified in a GWAS for BMI in ~340,000 subjects explain only 2.7% of the variance in BMI, and all HapMap phase 3 genetic variants (minor allele frequency (MAF) > 1%) (~1.5M single nucleotide polymorphisms (SNPs)) were estimated to account for ~21% of the variance in BMI in 16,275 unrelated individuals<sup>49</sup>. Thus, it is likely that gene-environment interactions (GxE) and rare variants also explain some of the BMI heritability<sup>50</sup>.

Recently it has been recognized that BMI cannot reliably differentiate fat from lean mass and that the metabolically detrimental abdominal obesity can be more accurately estimated using the waist-hip-ratio (WHR), which even after adjusting for BMI (WHRadjBMI) is still highly heritable (heritability ~0.22-0.61)<sup>51-54</sup>. Previous GWAS have also shown that WHRadjBMI GWAS genes are enriched for adipose-expressed genes with known adipose tissue functions, whereas BMI GWAS genes are enriched for genes expressed primarily in brain<sup>55</sup>. Regardless of the trait, due to linkage disequilibrium (LD) between variants and the presence of multiple genes and variants at each GWAS loci, the causal variant(s) and gene(s) are not immediately apparent from GWAS, hindering our ability to understand the finer details of biological mechanisms by

which GWAS loci contribute to obesity, and thus warranting a detailed fine mapping of each GWAS locus.

#### **1.4 Expression quantitative trait locus analyses in obesogenic tissue cohorts**

Common DNA variants, in addition to being associated with diseases and traits through GWAS, can also be directly linked to gene expression as expression quantitative trait loci (eQTL)<sup>56</sup>. eQTLs are measured by the strength of association between the number of copies of a certain allele of a DNA variant and differences in gene expression<sup>56-58</sup>. The discovery of eQTLs requires collection of population cohorts to capture genetic and gene expression variation, with genotyping arrays to assess DNA variants and RNA-sequencing (RNA-seq) to measure gene expression. As with chromosomal interactions, eQTLs are divided into two types, *cis* ( $\leq 1$ Mb from gene) and *trans* ( $> 1$ Mb from the gene or on a different chromosome)<sup>59</sup>; however, these thresholds for *cis* and *trans* for eQTLs are simply conventions that do not represent any true biological structure. Human gene expression cohorts are gradually increasing in size as sample collection becomes easier and RNA-seq cheaper, and thus, *cis*-eQTL discovery has become feasible; however, widespread identification of *trans*-eQTLs remains challenging, except in model organisms<sup>60,61</sup>. Even with costs decreasing, the size of the current human cohorts makes the discovery of *trans*-eQTLs limited by the extensive number of statistical tests required to be corrected for in the *trans*-eQTL analysis. Further complicating the discovery of relevant *cis*- and *trans*-eQTLs, is the type of tissue available from human cohorts for RNA-seq, since eQTLs and gene expression can also be tissue-specific, as is shown in cohorts with RNA-seq data from multiple tissues, such as GTEx<sup>58</sup>.

While human cohorts with samples from blood are among the largest (meta-analyses  $n_{\text{total}} \sim 30,000$ ) and are just beginning to reach large enough sizes to reliably identify *trans*-eQTLs at the genome-wide level<sup>62</sup>, cohorts with RNA-seq data from obesogenic tissues, such as adipose, liver, and muscle with genotyping data, remain still small, each cohort usually having samples from less than 1,000 individuals<sup>63,64</sup>. This limited size is partly due to the invasiveness of procedures to collect human tissue samples. Therefore, innovative targeted approaches alleviating the sample size requirement are the focus of intense ongoing research to identify *trans* effects relevant for obesity and related cardiometabolic traits and endpoints.

### **1.5 *Trans*-regulators in obesogenic tissue cohorts**

*Trans*-eQTLs regulated by TFs are the most common type of long-range, *trans* gene regulation<sup>59,61</sup>. For obesity-related tissue cohorts, there are only a few published examples of *trans*-eQTLs, including Krüppel Like Factor 14 (*KLF14*)<sup>65</sup> and Zinc finger protein 800 (*ZNF800*)<sup>66</sup>. One recent idea is that *cis*-eQTLs may also be *trans*-eQTLs, with the *trans* effect mediated by a gene whose expression is controlled by the *cis*-eQTL. One mechanism explaining these *cis*-mediated *trans*-eQTLs is that the gene is a TF, which would directly explain how the gene could affect many downstream genes across multiple chromosomes<sup>67</sup>. This idea has been suggested and explored in the GTE<sub>x</sub> cohort<sup>58</sup>, but it still remains unexplored in cohorts with obesogenic tissue samples from non-post mortem samples.

### **1.6 Transcriptome wide association studies of obesity**

Transcriptome wide association study (TWAS) is a method to test for association between gene expression and a trait by weighting the effects of all *cis* variants on gene expression and testing

their weighted association with a GWAS trait<sup>68,69</sup>. Since the gene expression and weights on *cis* variants are unique per gene, TWAS provides evidence for causality as it is the measure between fixed genotypes and traits. Thus, TWAS leverages both power from reference gene expression cohorts with RNA-seq data, such as GTEx (n~600)<sup>58</sup>, to calculate the weights on *cis* variants, and power from GWAS (n~100,000-800,000) for the association between the weighted variants and traits. Previous studies have illuminated the limitations of TWAS<sup>68</sup>, namely that it could identify multiple potential causal genes in a locus. This is due to the fact that nearby genes can have related variants in their *cis* regions due to underlying LD structure. However, this does not detract from the ability of TWAS to identify causal gene sets. Furthermore, follow-up methods, such as fine-mapping of causal gene sets (FOCUS)<sup>70</sup>, circumvent this issue by fine-mapping TWAS results through identification of a gene set containing the causal gene(s) in a locus at a predefined level of credibility, based on their posterior inclusion probability of being the causal gene, while accounting for shared *cis* variation among the genes at a locus. Combined, utilizing summary statistics from GWAS and transcriptomics reference panels, TWAS and FOCUS are an important pipeline for providing statistical evidence of the causality of genes for GWAS traits. One final caveat of TWAS and FOCUS is that they require the transcriptomic reference panels and GWAS to be from cohorts of the same ethnicity. As most cohorts currently contain a majority of European individuals, the use of TWAS and FOCUS in non-European populations has so far been limited by the availability and collection of more diverse cohorts.

## **1.6 Weighted Gene Co-expression Network Analysis for identifying key networks and genes for obesity**

Weighted Gene Co-expression Network Analysis (WGCNA), created in 2005 and placed into an R package in 2008<sup>71</sup>, was designed to find co-expressed genes associated with phenotypic traits. WGCNA forms networks out of genes that are highly co-expressed. It avoids the extensive multiple testing correction that would be necessary for testing each individual gene with traits. Instead, WGCNA summarizes collective gene expression of each network by identifying a network eigengene (i.e. the first principal component (PC) of the expression of all the genes in a network)<sup>71</sup>. As WGCNA needs only modest sizes of cohorts with gene expression data, WGCNA has also been utilized in human cardiometabolic transcriptomics cohorts for the past 10 years<sup>72-74</sup> to discover co-expression networks and genes important for obesity. In addition, it has been suggested the since TFs can control many genes across different chromosomes, they might also be master regulators of co-expression networks<sup>75</sup>. However, these studies rarely have any functional follow-up, proving that the network and master TFs are indeed acting as computationally calculated.

### **1.7 Polygenic risk score development for cardiometabolic disorders**

Polygenic risk scores (PRSs) represent the sum of common variants (MAF>1%) associated with a specific disease or phenotype, weighted by their effect sizes<sup>76</sup>. PRSs are used to assess an individual's risk for a certain disease or phenotype based on their genetics<sup>76</sup>. As the collection of genotype data increases, with large cohorts such as the UK Biobank (UKB)<sup>77</sup> (n~500,000) that have extensive phenotype and genotype data available for study, PRSs are becoming more feasible. PRS studies for cardiometabolic traits were started almost 10 years ago, and by now PRSs have been built for obesity in Europeans using BMI<sup>78,79</sup>, and many other related cardiometabolic traits, such as T2D<sup>80,81</sup>, coronary artery disease (CAD)<sup>82</sup>, and WHRadjBMI<sup>83</sup>.

Although PRSs are commonly built using independent variants across the genome, it is known that variants in certain regions of the genome, such as regulatory elements or promoters, are more disease relevant than others<sup>84</sup>. Therefore, using a targeted set of variants with prior evidence for connection to a phenotype or disease may yield a more localized and accurate PRS estimation.

One limitation of PRSs is that PRSs predict the disease risk independently of family history and other electronic health record (EHR) information<sup>82</sup>. Therefore, both EHR and PRS information are needed to get a more complete risk prediction for an individual. This difference can be explained by the fact that as PRSs capture all common variants, there are still contributions from rare variants to the disease risk. The rare variants often have large effect sizes even though they are present in a very small portion of the population. However, to acquire rare variant information, different and costlier technologies, such as whole exome sequencing (WES) and whole genome sequencing (WGS), are required because rare variants are not part of standard genotyping arrays that capture common variants<sup>85</sup>. Overall, this limitation does not discount the usefulness of PRSs as they can provide a useful biomarker at the individual level with just the collection of genotype information. This information could ultimately be included with an individual's EHR records in a clinical setting. Additionally, recent longitudinal studies have also indicated that in general individuals with a higher polygenic risk have an earlier age of onset of common cardiometabolic diseases, such as T2D, and various types of cancer<sup>86</sup>, further showing the utility of PRSs.

## **1.8 Integrating the current state of knowledge to this thesis**



This chapter has summarized first the current state of knowledge about adipose and adipocyte biology and its relationship to obesity. Building on this, the chapter described our present understanding of gene regulatory mechanisms from large DNA elements, such as enhancers and repressors, to smaller elements, such as TFBSs. In addition, the chapter covered the difference between *cis*-regulatory elements, *cis* gene regulation, and *cis*-eQTLs, when compared to the elusive *trans* gene regulation in human cohorts of obesogenic tissues. Finally, I discussed the current state of relevant genomic approaches, GWAS, TWAS, WGCNA, and PRS, and their applications and limitations.

Moving forward, in Chapter 2, I discuss the fine-mapping of BMI GWAS loci using *cis*-eQTLs from the METabolic Syndrome in Men (METSIM) cohort and human primary adipocyte pChI-C data. As genes make up only a few percent of the human genome, there are many intergenic GWAS variants where the closest gene to the variant is bookmarked for the GWAS locus. However, *cis*-eQTLs show that GWAS variants do not always affect the nearest gene, leading us to use functional evidence, such as chromosomal interactions from pChI-C in human primary adipocytes, to fine map and directly discover the mechanism of action of *cis*-eQTLs. We also characterized the pChI-C interactions to be enriched for adipogenesis TFs, PPAR $\gamma$  and CEBP $\beta$ . This study identified four key examples of obesity GWAS genes associated with BMI, serum metabolites, and lipids, as well as 38 additional candidate genes with *cis*-eQTLs in chromosomal interactions whose expression levels are strongly associated with BMI. These findings are important as they identify novel obesity loci; further show that GWAS variants are not necessarily connected to their nearest gene; and discover that one important regulatory mechanism for the connection between variants and genes is through promoter-enhancer interactions. This work was published in *Nature Communications* in 2018<sup>87</sup>.

Next, in Chapter 3, I discuss a study I contributed to, identifying changes in open chromatin regions within pChI-C chromosomal interactions. In the study, we used assay for transposase-accessible chromatin using sequencing (ATAC-seq) and pChI-C in human primary adipocytes treated with saturated fatty acids or monounsaturated fatty acids to examine changes in chromosomal interactions and open chromatin regions responding to lipid intake. Then we searched for variants residing at these context-specific sites responsive to the lipid intake using the UKB. Our goal was to identify variants interacting with saturated fat intake to influence obesity, i.e. BMI. Specifically, I contributed to the examining the conservation of lipid-responsive open chromatin regions in enhancers and promoters involved in chromosomal interactions. I showed that genes with lipid-responsive open chromatin regions in their promoters were significantly loss of function (LoF) intolerant and had higher conservation scores, meaning that they are evolutionarily constrained, when compared to the remaining protein-coding genes across the genome. Similarly, for enhancers with lipid-responsive open chromatin regions, I showed that they exhibited a higher conservation score than protein-coding genes across the genome, and furthermore that the target genes of the lipid-responsive enhancers via chromosomal interactions were LoF intolerant. In addition, I examined the same open chromatin regions in enhancers and promoters containing gene-environment interaction (GxE) variants for those predicted to significantly affect TF binding using the deep learning tool DeepSEA<sup>88</sup>, and found that 55% of them showed significant alterations of TF motifs, including retinoid X receptor alpha (RXRA), an important TF for adipogenesis and lipolysis<sup>89</sup>. This work was published in *Nature Metabolism* in 2019<sup>90</sup>.

Moving beyond *cis*-eQTLs, in Chapter 4, I discuss the identification of master TFs that control adipose co-expression networks important for obesity and provide functional evidence

for the mechanism of action of the identified master regulator. As it has been shown previously that WHRadjBMI is a better proxy for abdominal, metabolically harmful obesity than BMI, and that BMI cannot distinguish between lean mass from fat mass<sup>53</sup>, we used WHRadjBMI as our phenotype of interest and proxy for the metabolically harmful abdominal obesity. Previous studies have also shown that WHRadjBMI is a sex-dependent trait, reflecting the physiological differences in body fat distribution and muscle mass between males and females, with males in general exhibiting more muscle mass and females more fat mass when matched for BMI and age<sup>91,92</sup>. To advance the discovery of unknown genetic and molecular mechanisms regulating abdominal adiposity and the sex-specific distribution of body fat, we searched for genetic master regulators of WHRadjBMI by employing integrative genomics approaches on human adipose RNA-sequencing (RNA-seq) data (n~1,400), WHRadjBMI GWAS, TWAS, and PRS data from the WHRadjBMI GWAS cohorts and the UKB (n~700,000). We provide novel genomic evidence, verified by our functional studies in human primary preadipocytes, for the causal role of the TF, *TBX15*, in controlling accumulation of abdominal fat and adiposity. This work has been submitted in 2020 (Pan et al. submitted).

## References

1. Simonnet, A. *et al.* High prevalence of obesity in severe acute respiratory syndrome coronavirus-2 (SARS-CoV-2) requiring invasive mechanical ventilation. *Obesity* (2020).
2. Lighter, J. *et al.* Obesity in patients younger than 60 years is a risk factor for Covid-19 hospital admission. *Clin. Infect. Dis.* 1–29 (2020).
3. Choe, S. S., Huh, J. Y., Hwang, I. J., Kim, J. I. & Kim, J. B. Adipose tissue remodeling: Its role in energy metabolism and metabolic disorders. *Front. Endocrinol. (Lausanne)*. **7**, 1–16 (2016).
4. Zhang Y *et al.* Positional cloning of the mouse obese gene and its human homologue. *Nature* **372**, 425–432 (1994).
5. Hu, E., Liang, P. & Spiegelman, B. M. AdipoQ is a novel adipose-specific gene dysregulated in obesity. *J. Biol. Chem.* **271**, 10697–10703 (1996).
6. Scherer, P. E., Williams, S., Fogliano, M., Baldini, G. & Lodish, H. F. A novel serum protein similar to C1q, produced exclusively in adipocytes. *J. Biol. Chem.* **270**, 26746–26749 (1995).
7. Yang, Q. *et al.* Serum retinol binding protein 4 contributes to insulin resistance in obesity and type 2 diabetes. *Nature* **436**, 356–362 (2005).
8. Stepan, C. M. *et al.* The hormone resistin links obesity to diabetes. *Nature* **409**, 307–312 (2001).
9. De Souza Batista, C. M. *et al.* Omentin plasma levels and gene expression are decreased in obesity. *Diabetes* **56**, 1655–1661 (2007).
10. Yang, R. Z. *et al.* Identification of omentin as a novel depot-specific adipokine in human adipose tissue: Possible role in modulating insulin action. *Am. J. Physiol. - Endocrinol. Metab.* **290**, 1253–1261 (2006).
11. Meisinger, C., Döring, A., Thorand, B., Heier, M. & Löwel, H. Body fat distribution and risk of type 2 diabetes in the general population: Are there differences between men and women? The MONICA/KORA Augsburg Cohort Study. *Am. J. Clin. Nutr.* **84**, 483–489 (2006).
12. Gesta, S., Tseng, Y. H. & Kahn, C. R. Developmental Origin of Fat: Tracking Obesity to Its Source. *Cell* **131**, 242–256 (2007).
13. Merlotti, C., Ceriani, V., Morabito, A. & Pontiroli, A. E. Subcutaneous fat loss is greater than visceral fat loss with diet and exercise, weight-loss promoting drugs and bariatric surgery: A critical review and meta-analysis. *Int. J. Obes.* **41**, 672–682 (2017).
14. Spiegelman, B. M. & Flier, J. S. Obesity and the Regulation Review of Energy Balance total fast of approximately 150 days! This impressive energy reserve is due both to the high energy content of triglycerides versus polysaccharides, and the fact. *Cell* **104**, 531–543 (2001).
15. Lefterova, M. I. *et al.* PPAR $\gamma$  and C/EBP factors orchestrate adipocyte biology via adjacent binding on a genome-wide scale. *Genes Dev.* **22**, 2941–2952 (2008).
16. Kim, J. I. *et al.* Lipid-Overloaded Enlarged Adipocytes Provoke Insulin Resistance Independent of Inflammation. *Mol. Cell. Biol.* **35**, 1686–1699 (2015).
17. Klötting, N. *et al.* MicroRNA expression in human omental and subcutaneous adipose tissue. *PLoS One* **4**, 2–7 (2009).
18. Cotillard, A. *et al.* Adipocyte size threshold matters: Link with risk of type 2 diabetes and improved insulin resistance after gastric bypass. *J. Clin. Endocrinol. Metab.* **99**, 1466–1470 (2014).

19. Stern, J. S., Hollander, N., Batchelor, B. R., Cohn, C. K. & Hirsch, J. Adipose-Cell Size and Immunoreactive Insulin Levels in Obese and Normal-Weight Adults. *Lancet* **300**, 948–951 (1972).
20. Stark, R., Grzelak, M. & Hadfield, J. RNA sequencing: the teenage years. *Nat. Rev. Genet.* **20**, 631–656 (2019).
21. Stadhouders, R., Filion, G. J. & Graf, T. Transcription factors and 3D genome conformation in cell-fate decisions. *Nature* **569**, 345–354 (2019).
22. Lambert, S. A. *et al.* The Human Transcription Factors. *Cell* **172**, 650–665 (2018).
23. Singh, H., Khan, A. A. & Dinner, A. R. Gene regulatory networks in the immune system. *Trends Immunol.* **35**, 211–218 (2014).
24. Smith, A. D., Sumazin, P. & Zhang, M. Q. Identifying tissue-selective transcription factor binding sites in vertebrate promoters. *Proc. Natl. Acad. Sci. U. S. A.* **102**, 1560–1565 (2005).
25. Matys, V. *et al.* TRANSFAC®: Transcriptional regulation, from patterns to profiles. *Nucleic Acids Res.* **31**, 374–378 (2003).
26. Mathelier, A. *et al.* JASPAR 2016: A major expansion and update of the open-access database of transcription factor binding profiles. *Nucleic Acids Res.* **44**, D110–D115 (2016).
27. Wasserman, W. W. & Sandelin, A. Applied bioinformatics for the identification of regulatory elements. *Nat. Rev. Genet.* **5**, 276–287 (2004).
28. Schoenfelder, S. & Fraser, P. Long-range enhancer–promoter contacts in gene expression control. *Nat. Rev. Genet.* (2019).
29. Creyghton, M. P. *et al.* Histone H3K27ac separates active from poised enhancers and predicts developmental state. *Proc. Natl. Acad. Sci. U. S. A.* **107**, 21931–21936 (2010).
30. Andersson, R. *et al.* An atlas of active enhancers across human cell types and tissues. *Nature* **507**, 455–461 (2014).
31. Ernst, J. & Kellis, M. Chromatin-state discovery and genome annotation with ChromHMM. *Nat. Protoc.* **12**, 2478–2492 (2017).
32. Jaroszewicz, A. & Ernst, J. An integrative approach for fine-mapping chromatin interactions. *Bioinformatics* **36**, 1704–1711 (2020).
33. Sati, S. & Cavalli, G. Chromosome conformation capture technologies and their impact in understanding genome function. *Chromosoma* **126**, 33–44 (2017).
34. Dekker, J., Rippe, K., Dekker, M. & Kleckner, N. Capturing chromosome conformation. *Science (80-. )*. **295**, 1306–1311 (2002).
35. Würtele, H. & Chartrand, P. Genome-wide scanning of HoxB1-associated loci in mouse ES cells using an open-ended Chromosome Conformation Capture methodology. *Chromosom. Res.* **14**, 477–495 (2006).
36. Simonis, M. *et al.* Nuclear organization of active and inactive chromatin domains uncovered by chromosome conformation capture-on-chip (4C). *Nat. Genet.* **38**, 1348–1354 (2006).
37. Lomvardas, S. *et al.* Interchromosomal Interactions and Olfactory Receptor Choice. *Cell* **126**, 403–413 (2006).
38. Zhao, Z. *et al.* Circular chromosome conformation capture (4C) uncovers extensive networks of epigenetically regulated intra- and interchromosomal interactions. *Nat. Genet.* **38**, 1341–1347 (2006).
39. Dostie, J. *et al.* Chromosome Conformation Capture Carbon Copy (5C): A massively

- parallel solution for mapping interactions between genomic elements. *Genome Res.* 1299–1309 (2006).
40. Ferraiuolo, M. A., Sanyal, A., Naumova, N., Dekker, J. & Dostie, J. From cells to chromatin: Capturing snapshots of genome organization with 5C technology. *Methods* **58**, 255–267 (2012).
  41. Lieberman-aiden, E. *et al.* Comprehensive Mapping of Long-Range Interactions Reveals Folding Principles of the Human Genome. *Science* (80-. ). **326**, 289–294 (2009).
  42. Jin, F. *et al.* A high-resolution map of the three-dimensional chromatin interactome in human cells. *Nature* **503**, 290–294 (2013).
  43. Mifsud, B. *et al.* Mapping long-range promoter contacts in human cells with high-resolution capture Hi-C. *Nat. Genet.* **47**, 598–606 (2015).
  44. Dekker, J. & Misteli, T. Long-range chromatin interactions. *Cold Spring Harb. Perspect. Biol.* **7**, (2015).
  45. Erceg, J. *et al.* The genome-wide multi-layered architecture of chromosome pairing in early *Drosophila* embryos. *Nat. Commun.* **10**, (2019).
  46. Chowdhary, S., Kainth, A. S. & Gross, D. S. Chromosome conformation capture that detects novel cis- and trans-interactions in budding yeast. *Methods* **170**, 4–16 (2020).
  47. Tam, V. *et al.* Benefits and limitations of genome-wide association studies. *Nat. Rev. Genet.* **20**, 467–484 (2019).
  48. Fadista, J., Manning, A. K., Florez, J. C. & Groop, L. The (in)famous GWAS P-value threshold revisited and updated for low-frequency variants. *Eur. J. Hum. Genet.* **24**, 1202–1205 (2016).
  49. Locke, A. E. *et al.* Genetic studies of body mass index yield new insights for obesity biology. *Nature* **518**, 197–206 (2015).
  50. Abadi, A. *et al.* Penetrance of Polygenic Obesity Susceptibility Loci across the Body Mass Index Distribution. *Am. J. Hum. Genet.* **101**, 925–938 (2017).
  51. 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).
  52. Mills, G. W. *et al.* Heritability estimates for beta cell function and features of the insulin resistance syndrome in UK families with an increased susceptibility to Type 2 diabetes. *Diabetologia* **47**, 732–738 (2004).
  53. Shungin, D. *et al.* New genetic loci link adipose and insulin biology to body fat distribution. *Nature* **518**, 187–196 (2015).
  54. Souren, N. Y. *et al.* Anthropometry, carbohydrate and lipid metabolism in the East Flanders Prospective Twin Survey: Heritabilities. *Diabetologia* **50**, 2107–2116 (2007).
  55. Heid, I. M. *et al.* Meta-analysis identifies 13 new loci associated with waist-hip ratio and reveals sexual dimorphism in the genetic basis of fat distribution. *Nat. Genet.* **42**, 949–960 (2010).
  56. Schadt, E. E. *et al.* An integrative genomics approach to infer causal associations between gene expression and disease. *Nat. Genet.* **37**, 710–717 (2005).
  57. Lonsdale, J. *et al.* The Genotype-Tissue Expression (GTEx) project. *Nat. Genet.* **45**, 580–585 (2013).
  58. The Genotype Tissue Expression Consortium. The GTEx Consortium atlas of genetic regulatory effects across human tissues. *bioRxiv.* (2019).
  59. Yao, C. *et al.* Dynamic Role of trans Regulation of Gene Expression in Relation to

- Complex Traits. *Am. J. Hum. Genet.* **100**, 571–580 (2017).
60. Yvert, G. *et al.* Trans-acting regulatory variation in *Saccharomyces cerevisiae* and the role of transcription factors. *Nat. Genet.* **35**, 57–64 (2003).
  61. Albert, F. W., Bloom, J. S., Siegel, J., Day, L. & Kruglyak, L. Genetics of trans-regulatory variation in gene expression. *Elife* **7**, 1–39 (2018).
  62. Vösa, U. *et al.* Unraveling the polygenic architecture of complex traits using blood eQTL meta-analysis. 1–57
  63. Laakso, M. *et al.* The Metabolic Syndrome in Men study: A resource for studies of metabolic & cardiovascular diseases. *J. Lipid Res.* **58**, 481–493 (2017).
  64. Moayyeri, A., Hammond, C. J., Hart, D. J. & Spector, T. D. The UK Adult Twin Registry (TwinsUK Resource). *Twin Res. Hum. Genet.* **16**, 144–149 (2013).
  65. Small, K. S. *et al.* Regulatory variants at KLF14 influence type 2 diabetes risk via a female-specific effect on adipocyte size and body composition. *Nat. Genet.* **50**, 572–580 (2018).
  66. Civelek, M. *et al.* Genetic Regulation of Adipose Gene Expression and Cardio-Metabolic Traits. *Am. J. Hum. Genet.* **100**, 428–443 (2017).
  67. Yang, F., Wang, J., Pierce, B. L. & Chen, L. S. Identifying cis-mediators for trans-eQTLs across many human tissues using genomic mediation analysis. *Genome Res.* **27**, 1859–1871 (2017).
  68. Wainberg, M. *et al.* Opportunities and challenges for transcriptome-wide association studies. *Nat. Genet.* **51**, 592–599 (2019).
  69. Gusev, A. *et al.* Integrative approaches for large-scale transcriptome-wide association studies. *Nat. Genet.* **48**, 245–252 (2016).
  70. Mancuso, N. *et al.* Probabilistic fine-mapping of transcriptome-wide association studies. *Nat. Genet.* **51**, 675–682 (2019).
  71. Langfelder, P. & Horvath, S. WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics* **9**, 559 (2008).
  72. Kogelman, L. J. A. *et al.* Identification of co-expression gene networks, regulatory genes and pathways for obesity based on adipose tissue RNA Sequencing in a porcine model. *BMC Med. Genomics* **7**, 1–16 (2014).
  73. Hao, R. H. *et al.* Gene expression profiles indicate tissue-specific obesity regulation changes and strong obesity relevant tissues. *Int. J. Obes.* **42**, 363–369 (2018).
  74. Haas, B. E. *et al.* Adipose Co-expression networks across Finns and Mexicans identify novel triglyceride-associated genes. *BMC Med. Genomics* **5**, (2012).
  75. Skinkyte-Juskiene, R., Kogelman, L. J. A. & Kadarmideen, H. N. Transcription Factor Co-expression Networks of Adipose RNA-Seq Data Reveal Regulatory Mechanisms of Obesity. *Curr. Genomics* **19**, 289–299 (2018).
  76. Purcell, S. M. *et al.* Common polygenic variation contributes to risk of schizophrenia and bipolar disorder. *Nature* **460**, 748–752 (2009).
  77. Sudlow, C. *et al.* UK Biobank: An Open Access Resource for Identifying the Causes of a Wide Range of Complex Diseases of Middle and Old Age. *PLoS Med.* **12**, 1–10 (2015).
  78. Khera, A. V. *et al.* Polygenic Prediction of Weight and Obesity Trajectories from Birth to Adulthood. *Cell* **177**, 587–596.e9 (2019).
  79. Peterson, R. E. *et al.* Genetic risk sum score comprised of common polygenic variation is associated with body mass index. *Hum. Genet.* **129**, 221–230 (2011).
  80. Vassy, J. L. *et al.* Polygenic type 2 diabetes prediction at the limit of common variant

- detection. *Diabetes* **63**, 2172–2182 (2014).
81. Udler, M. S., McCarthy, M. I., Florez, J. C. & Mahajan, A. Genetic Risk Scores for Diabetes Diagnosis and Precision Medicine. *Endocr. Rev.* **40**, 1500–1520 (2019).
  82. Elliott, J. *et al.* Predictive Accuracy of a Polygenic Risk Score-Enhanced Prediction Model vs a Clinical Risk Score for Coronary Artery Disease. *JAMA - J. Am. Med. Assoc.* **323**, 636–645 (2020).
  83. Emdin, C. A. *et al.* Genetic association of waist-to-hip ratio with cardiometabolic traits, type 2 diabetes, and coronary heart disease. *JAMA - J. Am. Med. Assoc.* **317**, 626–634 (2017).
  84. Li, M. J., Sham, P. C. & Wang, J. Genetic variant representation, annotation and prioritization in the post-GWAS era. *Cell Res.* **22**, 1505–1508 (2012).
  85. Povysil, G. *et al.* Rare-variant collapsing analyses for complex traits: guidelines and applications. *Nat. Rev. Genet.* **20**, 747–759 (2019).
  86. Mars, N. J. *et al.* Polygenic and clinical risk scores and their impact on age at onset of cardiometabolic diseases and common cancers. *bioRxiv.* (2019).
  87. Pan, D. Z. *et al.* Integration of human adipocyte chromosomal interactions with adipose gene expression prioritizes obesity-related genes from GWAS. *Nat. Commun.* **9**, (2018).
  88. Zhou, J. & Troyanskaya, O. G. Predicting effects of noncoding variants with deep learning-based sequence model. *Nat. Methods* **12**, 931–934 (2015).
  89. Imai, T., Jiang, M., Chambon, P. & Metzger, D. Impaired adipogenesis and lipolysis in the mouse upon selective ablation of the retinoid X receptor  $\alpha$  mediated by a tamoxifen-inducible chimeric Cre recombinase (Cre-ERT2) in adipocytes. *Proc. Natl. Acad. Sci. U. S. A.* **98**, 224–228 (2001).
  90. Garske, K. M. *et al.* Reverse gene–environment interaction approach to identify variants influencing body-mass index in humans. *Nat. Metab.* **1**, 630–642 (2019).
  91. Rask-Andersen, M., Karlsson, T., Ek, W. E. & Johansson, Å. Genome-wide association study of body fat distribution identifies adiposity loci and sex-specific genetic effects. *Nat. Commun.* **10**, (2019).
  92. Schorr, M. *et al.* Sex differences in body composition and association with cardiometabolic risk. *Biol. Sex Differ.* **9**, 1–10 (2018).



## Chapter 2

Integration of human adipocyte chromosomal interactions with adipose gene expression  
prioritizes obesity-related genes from GWAS

ARTICLE

DOI: 10.1038/s41467-018-03554-9

OPEN

# Integration of human adipocyte chromosomal interactions with adipose gene expression prioritizes obesity-related genes from GWAS

David Z. Pan<sup>1,2</sup>, Kristina M. Garske<sup>1</sup>, Marcus Alvarez<sup>1</sup>, Yash V. Bhagat<sup>1</sup>, James Boocock<sup>1</sup>, Elina Nikkola<sup>1</sup>, Zong Miao<sup>1,2</sup>, Chelsea K. Raulerson<sup>3</sup>, Rita M. Cantor<sup>1</sup>, Mete Civelek<sup>4</sup>, Craig A. Glastonbury<sup>5</sup>, Kerrin S. Small<sup>6</sup>, Michael Boehnke<sup>7</sup>, Aldons J. Lusis<sup>1</sup>, Janet S. Sinsheimer<sup>1,8</sup>, Karen L. Mohlke<sup>3</sup>, Markku Laakso<sup>9</sup>, Päivi Pajukanta<sup>1,2,10</sup> & Arthur Ko<sup>1,10</sup>

Increased adiposity is a hallmark of obesity and overweight, which affect 2.2 billion people world-wide. Understanding the genetic and molecular mechanisms that underlie obesity-related phenotypes can help to improve treatment options and drug development. Here we perform promoter Capture Hi-C in human adipocytes to investigate interactions between gene promoters and distal elements as a transcription-regulating mechanism contributing to these phenotypes. We find that promoter-interacting elements in human adipocytes are enriched for adipose-related transcription factor motifs, such as PPARG and CEBPB, and contribute to heritability of *cis*-regulated gene expression. We further intersect these data with published genome-wide association studies for BMI and BMI-related metabolic traits to identify the genes that are under genetic *cis* regulation in human adipocytes via chromosomal interactions. This integrative genomics approach identifies four *cis*-eQTL-eGene relationships associated with BMI or obesity-related traits, including rs4776984 and *MAP2K5*, which we further confirm by EMSA, and highlights 38 additional candidate genes.

<sup>1</sup>Department of Human Genetics, David Geffen School of Medicine at UCLA, Los Angeles, CA 90095, USA. <sup>2</sup>Bioinformatics Interdepartmental Program, UCLA, Los Angeles, CA 90095, USA. <sup>3</sup>Department of Genetics, University of North Carolina, Chapel Hill, NC 27599, USA. <sup>4</sup>Department of Biomedical Engineering, University of Virginia, Charlottesville, VA 22904, USA. <sup>5</sup>Big Data Institute, University of Oxford, Oxford OX3 7LF, UK. <sup>6</sup>Department of Twin Research and Genetic Epidemiology, King's College, London, UK. <sup>7</sup>Department of Biostatistics, University of Michigan, Ann Arbor, MI 48109, USA. <sup>8</sup>Department of Biomathematics, David Geffen School of Medicine at UCLA, Los Angeles, CA 90095, USA. <sup>9</sup>Institute of Clinical Medicine, Internal Medicine, University of Eastern Finland and Kuopio University Hospital, FI-70210 Kuopio, Finland. <sup>10</sup>Molecular Biology Institute at UCLA, Los Angeles, CA 90095, USA. These authors contributed equally: David Z. Pan, Kristina M. Garske. Correspondence and requests for materials should be addressed to A.K. (email: a5ko@ucla.edu)

Obesity is a serious health epidemic world-wide. A recent study of 195 countries estimated that 2.2 billion people were overweight or obese in 2015<sup>1</sup>. Clinically, obesity is diagnosed by a body mass index (BMI) greater than 30. While a significant proportion of the phenotypic variation in BMI is attributed to genetic variation (heritability of BMI ~0.4–0.7<sup>2</sup>), understanding the mechanisms underlying this heritable component has been challenging. The 97 loci identified in a genome-wide association study (GWAS) for BMI in ~340,000 subjects explain only 2.7% of the variance in BMI, and all HapMap phase 3 genetic variants (~1.5 M single nucleotide polymorphisms (SNPs)) were estimated to account for ~21% of the variance in BMI in 16,275 unrelated individuals<sup>2</sup>. The causal variants and genes are not immediately apparent from GWAS, hindering our ability to understand the biological mechanisms by which genetics contribute to obesity. To address this knowledge gap, we integrate chromosomal interaction data from primary human white adipocytes (HWA) with adipose gene expression and clinical phenotype data (BMI, waist-hip ratio, fasting insulin, and Matsuda index) to elucidate molecular pathways involved in genetic regulation in *cis*.

Combining genotype and RNA-sequencing (RNA-seq) data enables the detection of expression quantitative trait loci (eQTLs) that regulate transcription of near-by genes (i.e., in *cis*). These *cis*-eQTLs often reside in regulatory elements, including promoters, enhancers, and super-enhancers. However, the mechanism by which *cis*-eQTLs regulate their respective eGene(s) is seldom established because identification of the true regulatory variants among SNPs in tight linkage disequilibrium (LD) has proven challenging<sup>3</sup>. Enhancers modulate target gene expression levels via their interaction with promoters, and disruption or improper looping of enhancer sites can contribute to disease risk<sup>4,5</sup>. Promoter Capture Hi-C (pChi-C) enables detection of promoter interactions at a higher resolution and at lower sequencing depth than that required for Hi-C<sup>6</sup>. Incorporating a chromosomal interaction map constructed from pChi-C and *cis*-eQTL data can help elucidate the functional mechanisms by which the genetic variants affect gene expression. By overlapping these looping *cis*-eQTLs with trait-associated variants identified in independent, large-scale GWAS, we can assess which GWAS variants could affect expression of regional genes via chromosomal interactions.

To search for genes that are functionally important for adipose tissue biology, we performed a *cis*-eQTL analysis using genome-wide SNP data and adipose RNA-seq data from individuals of the Finnish METabolic Syndrome In Men (METSIM) cohort. We identified 42 genes, regulated by *cis*-eQTLs that reside in regions that physically interact with the promoters of genes. Adipose expression of these 42 genes was robustly correlated with BMI, and among them four genes, *MAP2K5*, *LACTB*, *ORMDL3*, and *ACADS*, were regulated by SNPs (or their tight LD proxies) previously identified in GWAS for BMI or a related metabolic trait, located at the regulatory element-promoter interaction sites. These data provide converging evidence for effects of looping *cis*-eQTL variants on gene expression associated with obesity and related metabolic traits. Our results show that these integrative genomics methods involving pChi-C data in primary HWA can identify regulatory circuits comprising both regulatory elements and their target gene(s) that operate in a complex obesity-related metabolic trait.

## Results

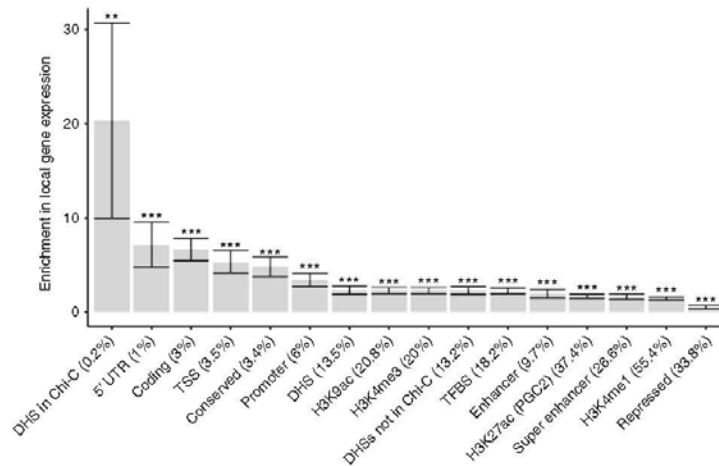
**Characterization of the adipocyte chromosomal interactions.** Adipose tissue is highly heterogeneous, containing adipocytes, preadipocytes, stem cells, and various immune cells. We performed pChi-C in primary HWA with the goal of identifying

physical interactions between adipose *cis*-eQTLs and target gene promoters. We employed the pChi-C protocol as described previously<sup>7</sup>. Briefly, we fixed primary HWA to crosslink proteins to DNA, and after digestion with the *HindIII* restriction endonuclease, we performed in-nucleus ligation and biotinylated RNA bait hybridization to pull down only those *HindIII* fragments with annotated gene promoters<sup>6</sup>. To detect the regions that interact with the promoter-containing *HindIII* fragments, we mapped the reads to the genome, and assigned reads to *HindIII* fragments to allow for fragment-level resolution of those regions interacting with the baited fragments containing gene promoters. The key pChi-C sequencing metrics are shown in Supplementary Table 1.

We first confirmed that the non-promoter regions in adipocyte chromosomal interactions are enriched for enhancer (H3K4me1, H3K4me3, and H3K27ac), repressor (H3K27me3, H3K9me3) histone marks, and DNase I hypersensitive sites (DHSs) (Supplementary Table 2). As there are no publicly available DHS data for adipocytes or adipose tissue, we used the union of DHSs in all cell types from ENCODE and Roadmap rather than DHSs in a single, non-adipocyte cell type<sup>8</sup>. Intersecting the adipocyte and previously published primary CD34<sup>+</sup> cell pChi-C data<sup>6</sup>, we found that 68.0% of adipocyte pChi-C chromosomal interactions were observed in adipocytes but not in CD34<sup>+</sup> cells. In the following, we used the same public DHS data to focus on open chromatin regions as they are more likely to bind transcription factors (TFs) and, thus, be relevant for chromosomal looping interactions within the interacting *HindIII* fragments.

We examined whether the DHSs are enriched for adipose-related TF motifs, using the Hypergeometric Optimization of Motif Enrichment (HOMER) software<sup>9</sup> that calculates the number of times a TF motif is seen in target and background sequences. The proportion of times the TF motif is seen in the target when compared to the background is then tested for enrichment in the target sequences. We found that when compared to DHSs within CD34<sup>+</sup> chromatin interactions, the DHSs within the adipocyte chromatin interactions are enriched for 26 of 332 TF motifs (FDR < 5%) (Supplementary Table 3), including CCAAT/enhancer binding protein beta (CEBPB, *p*-value =  $1.00 \times 10^{-10}$ ) and peroxisome proliferator-activated receptor gamma (PPARG, *p*-value = 0.01), both of which are well-known key players in adipose biology<sup>10</sup>. To address the potential bias of using a different pChi-C dataset as background, we also performed HOMER analysis comparing the DHSs in adipocyte interactions to DHSs in non-interacting, non-promoter regions in the remainder of the genome. The results were similar, and both CEBPB and PPARG were also enriched in the latter analysis (CEBPB, *p*-value =  $1.00 \times 10^{-24}$ ; PPARG, *p*-value =  $1.00 \times 10^{-6}$ ; complete enrichment results not shown). These results suggest that the cell-type based pChi-C interaction data enable the detection of interactions important for that cell type within a heterogeneous human tissue.

**Chromosomal interactions explain expression heritability.** To investigate whether the variants residing within open chromatin of chromosomal looping regions in adipocytes are enriched for SNPs that contribute to the heritability of *cis* expression regulation, we partitioned the heritability of *cis* regulation of human adipose gene expression into 52 functional categories using a modified partitioned LD Score regression method<sup>11</sup> (see Methods). The 52 functional categories are derived from 26 main annotations that include coding regions, untranslated regions (UTRs), promoters, intronic regions, histone marks, DNase I hypersensitivity sites (DHSs), predicted enhancers, conserved regions, and other annotations<sup>11</sup> (Supplementary



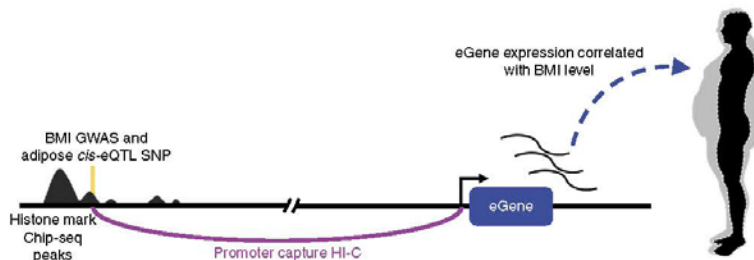
**Fig. 1** Open chromatin sites (DHSs) within adipocyte promoter ChI-C chromosomal interactions show significant enrichment in *cis* expression. Enrichments in *cis* expression with error bars for different categories using LD score regression analysis (see Methods). For the horizontal axis labels, the value in parentheses shows the percentage of SNPs contained within the respective annotation category that contributed to the enrichment calculation. For the significance threshold after Bonferroni correction above each bar, \* indicates a  $p$ -value  $< 0.05$ ; \*\*, a  $p$ -value  $< 0.001$ ; and \*\*\*, a  $p$ -value  $< 0.0001$ , respectively. The  $p$ -values were estimated based on Z scores calculated from the normal distribution. Error bars represent jackknife standard errors around the estimates of enrichment

Figure 1, Supplementary Tables 4–5). The partitioned LD Score regression method<sup>11</sup> utilizes summary association statistics of all variants on gene expression to estimate the degree to which variants in different annotation categories explain the heritability of *cis* and *trans* expression regulation while accounting for the LD among functional annotations. To assess the enrichment of heritability mediated by the variants in the chromosomal interactions detected by pChI-C on a per-gene basis, we further modified the LD score method, as described in detail in the Methods. Importantly, these modifications did not change the 52 baseline enrichments significantly when compared with the data obtained using the unmodified version<sup>11</sup> (Supplementary Figure 1, Supplementary Tables 4–5). These analyses revealed that open chromatin regions (i.e., DHSs) within the adipocyte chromosomal interactions are enriched for sequences that contribute to heritability of gene expression regulation in *cis* (Fig. 1,  $p$ -value  $< 0.002$ , enrichment = 20.3 (SD ± 5.2), average proportion of SNPs = 0.23%). The variants residing within the open chromatin regions within adipocyte chromosomal interactions explain 4.6% of the heritability of adipose tissue gene expression in *cis*, despite only accounting for 0.23% of the SNPs per *cis* gene region on average, indicating the functionality of these SNPs at the DHSs of distal interactions in regulating *cis* expression.

**Identification of genes regulated by looping *cis*-eQTL SNPs.** To identify adipose-expressed genes regulated by SNPs (eGenes), we performed a *cis*-eQTL analysis using 335 individuals from the METSIM cohort with both genome-wide SNP data and adipose RNA-seq data available (Fig. 2; Methods). Using the published adipose *cis*-eQTL data and criteria for significance from GTEx<sup>12</sup> (see Methods), we found 487,679 *cis*-eQTLs for 4,650 eGenes in the METSIM dataset and confirmed these same SNPs as *cis*-eQTLs by look-up in GTEx. 386,068 of the 487,679 (79.0%) *cis*-eQTL SNPs had the same target gene and direction of effect in both cohorts (Supplementary Figure 2). Only the 386,068

*cis*-eQTL SNPs that were replicated for effect direction and target gene (Supplementary Table 1) in the GTEx adipose RNA-seq data were used in our subsequent downstream analyses (Supplementary Figure 2). Overall, 4,332 of 4,650 of *cis*-eQTL-eGene relationships (93.0%) were replicated using the published adipose *cis*-eQTL data and criteria for significance from GTEx<sup>12</sup> (see Methods). To restrict these adipose *cis*-eQTL SNPs to those that likely function through transcription factor (TF) binding at distal regulatory elements, we determined which of these eGene promoters were involved in looping interactions with the *cis*-eQTLs, assayed through pChI-C in primary HWA (Fig. 2; Methods). Of the 4,332 eGenes identified in our *cis*-eQTL analysis, 576 (13.4%), were involved in these looping interactions (permutation  $p$ -value  $< 0.00001$ ) (Fig. 2, Methods, Supplementary Figure 2, and Supplementary Table 1).

We next determined the set of 576 looping eGenes with expression levels that are correlated with BMI in METSIM (Pearson correlation, adjusted  $p < 1.15 \times 10^{-5}$  to correct for the 4,332 replicated eGenes identified in our *cis*-eQTL analysis). We found 54 of 576 (9.40%) BMI-correlated eGenes with promoters involved in looping interactions with their *cis*-eQTL SNP (Supplementary Table 6). In our subsequent second replication analysis, the expression levels of 42 out of 54 genes (replication rate of 77.8%) were correlated with BMI in adipose RNA-seq data from the TwinsUK cohort ( $n = 720$ ) with the same direction of effect on BMI as in METSIM (Bonferroni adjusted  $p < 0.001$ ) (Table 1, Supplementary Table 6). Another four of the 54 genes were not available in the TwinsUK dataset. The effects sizes and  $p$ -values obtained for BMI associations in TwinsUK and METSIM, using a linear regression model in both, show comparable results to those obtained using the Pearson correlations (Table 1, Supplementary Table 6). These 42 BMI-correlated genes are functionally enriched for four pathways with fatty acid metabolism as a top ranking pathway (Supplementary Table 7) based on KEGG pathway enrichment using WebGestalt<sup>13</sup> (Benjamini-Hochberg adjusted  $p < 0.05$ ); however, the small



**Fig. 2** Overview of the study design targeted to identify new genes for obesity and related metabolic traits. A schematic illustrating the integration of multi-omics data utilized in this study to elucidate genetics of obesity-related traits.

number of genes in these pathway analyses warrant verification in future studies. Only these 42 replicated genes were further investigated in our downstream analyses.

#### Adipocyte chromosomal interactions prioritize GWAS genes.

To investigate which of the 42 BMI-correlated eGenes are regulated by GWAS variants previously identified for BMI and related metabolic traits, we determined which interacting *cis*-eQTL variants are GWAS variants (or their LD proxies,  $r^2 > 0.80$ ), using  $p < 5.00 \times 10^{-8}$  as a criterion to select variants. As the goal of the current study was to dissect the molecular contribution of adipose and adipocyte biology to traits that can influence the pathophysiology of obesity, we examined GWAS for BMI and the metabolic traits that have previously been shown to exhibit comorbidities with obesity (e.g., serum lipids and type 2 diabetes) or that are influenced by obesity or correlated with BMI (e.g., metabolites and WHR). We used all GWAS variants ( $p$ -value  $< 5.00 \times 10^{-8}$ ) identified in a previous metabolite GWAS of ~7000 individuals<sup>14</sup>, lipid GWAS of ~180,000 individuals<sup>15</sup>, an extensive BMI GWAS study of ~340,000 individuals<sup>2</sup>, a sequencing-based GWAS for type 2 diabetes<sup>16</sup>, and a waist-hip-ratio (WHR) adjusted for BMI GWAS of ~220,000 individuals<sup>17</sup>. We found a GWAS variant for BMI, regulating mitogen-activated protein kinase kinase 5 (*MAP2K5*); a GWAS variant for high-density lipoprotein cholesterol (HDL-C), regulating orosomucoid like sphingolipid biosynthesis regulator 3 (*ORMDL3*); and two GWAS variants for serum metabolites (succinylcarnitine and butyrylcarnitine), regulating lactamase beta (*LACTB*) and acyl-CoA dehydrogenase, C-2 To C-3 short chain (*ACADS*), among the 42 genes (Fig. 3a, b; Supplementary Figure 3a–f), with the looping interactions spanning 287 kb, 16 kb, 151 kb, and 183 kb, respectively. We found that the interacting *cis*-eQTL-containing *HindIII* fragments for *LACTB* and *MAP2K5* are located within the promoter and intron of other genes. Furthermore, using the integrated pChI-C and *cis*-eQTL data, we found that the SNPs in these regulatory *HindIII* fragments regulate genes that are not their nearest gene for 3 of the 4 BMI-correlated eGenes (Fig. 3a, b, Supplementary Figure 3a–f).

**The looping BMI GWAS SNPs regulate *MAP2K5*.** For *MAP2K5*, the reported BMI GWAS SNP itself is not located within the regulatory, *cis*-eQTL-containing *HindIII* fragment involved in the looping interaction; however, SNPs in tight LD with the GWAS SNP (using a criterion of  $r^2 > 0.80$ ) are in the regulatory *HindIII* fragment that is interacting with the target gene promoter (Fig. 3b). The regulatory *HindIII* fragment contains 16 *cis*-eQTL SNPs that are LD proxies for the BMI GWAS SNP<sup>2</sup> (rs16951275), which has a total of 62 LD proxies in the

METSIM cohort. To prioritize a candidate functional variant within these 16 SNPs within the *HindIII* fragment, we first examined the predicted TF motifs that may be affected by each SNP using the data curated from ChIP-seq by Kheradpour and Kellis<sup>18</sup>. We found that only rs4776984, which is in almost perfect LD with the original BMI GWAS variant rs16951275 ( $r^2 = 0.98$ ), showed a predicted increase in binding of CTCF, which is a TF known to mediate chromosomal interactions (Fig. 4a).

We also used the deep learning-based sequence analyzer (DeepSEA)<sup>19</sup> to examine the allelic effect on protein binding of rs4776984 and the 15 other looping *cis*-eQTLs of *MAP2K5*. Of these 16 looping *cis*-eQTLs, six were potentially functional and of these, two variants passed the functional significance score of  $< 0.05$  using DeepSEA. Of the two, our candidate functional eQTL SNP, rs4776984, resulted in the most significant functional score ( $2.36 \times 10^{-3}$ ) (Supplementary Table 8) and was the only variant passing a functional significance score of  $< 0.01$  among the 16 variants. Thus, the DeepSEA result further supports the differential TF binding at the variant site rs4776984 among all possible looping *cis*-eQTLs at the *MAP2K5* locus (Supplementary Table 8). The looping *cis*-eQTL site also shows a ChIP-seq peak for the histone mark H3K4me1 in ENCODE adipose nuclei ChIP-seq data; however, notably it also shows the presence of the histone marks H3K27me3 and H3K9me3 (Fig. 3b), two marks known to be associated with transcriptional repression. Furthermore, the gene expression of *MAP2K5* is negatively correlated with BMI ( $p$ -value =  $7.83 \times 10^{-6}$ ). These data implicate *MAP2K5* as a gene regulated by the BMI GWAS signal via a repressive chromosomal interaction.

To functionally assess whether there is differential allele-specific binding of proteins at the candidate functional *MAP2K5* eQTL, rs4776984, we performed electrophoretic mobility shift assays (EMSAs) using nuclear protein from primary HWA. The results show reduced protein binding of the reference allele when compared to the alternate allele of rs4776984, consistently in three independent experiments (Fig. 4b, Supplementary Figure 4), in line with the predicted disruption in protein binding for CTCF<sup>18</sup> (Fig. 4a). We performed the supershift experiment using the CTCF antibody and adipocyte nuclear extract, but did not observe a supershift in any of the three replicated experiments (Supplementary Figure 5). We repeated the supershift experiment using a different CTCF antibody (EMD Millipore 07-729), which resulted in the same negative finding (Supplementary Figure 6). To further verify the negative supershift result, we also directly tested the CTCF protein for allele-specific binding at rs4776984 using EMSA in 3 replicate experiments, and did not find evidence of sole CTCF protein binding (Supplementary

**Table 1** Thirteen representative eGenes (9 most significant genes and 4 GWAS loci) that correlate with BMI in METSIM and TwinsUK (for the full data on all 54 genes, see Supplementary Table 6)

Rank <sup>a</sup>	Gene	Chr <sup>f</sup>	Pearson		Linear regression			TwinsUK <sup>e</sup>		
			METSIM <sup>c</sup>		METSIM <sup>d</sup>					
			Effect size (r)	p-value	Effect size (β)	SE	p-value	Effect size (β)	SE	p-value
1	<i>ADH1B</i>	4	-0.45	$7.40 \times 10^{-18}$	-0.21	0.02	$1.68 \times 10^{-20}$	-0.58	0.03	$4.47 \times 10^{-71}$
2	<i>ORMDL3</i> <sup>b</sup>	17	-0.45	$8.57 \times 10^{-18}$	-0.16	0.02	$2.06 \times 10^{-20}$	-0.58	0.03	$2.65 \times 10^{-70}$
3	<i>AKR1C3</i>	10	0.33	$4.78 \times 10^{-10}$	0.13	0.02	$2.95 \times 10^{-11}$	0.49	0.03	$5.19 \times 10^{-54}$
4	<i>CMY3</i>	16	0.41	$4.32 \times 10^{-15}$	0.087	0.01	$3.84 \times 10^{-17}$	0.50	0.03	$6.64 \times 10^{-52}$
5	<i>LPIN1</i>	2	-0.38	$1.49 \times 10^{-13}$	-0.14	0.02	$2.27 \times 10^{-15}$	-0.47	0.03	$2.38 \times 10^{-44}$
6	<i>RNF157</i>	17	-0.29	$5.19 \times 10^{-8}$	-0.096	0.02	$5.87 \times 10^{-9}$	-0.47	0.03	$8.86 \times 10^{-42}$
7	<i>MYOF</i>	10	0.32	$1.07 \times 10^{-9}$	0.086	0.01	$7.37 \times 10^{-11}$	0.46	0.03	$2.59 \times 10^{-40}$
8	<i>NAA40</i>	11	0.28	$1.81 \times 10^{-7}$	0.052	0.009	$2.67 \times 10^{-8}$	0.46	0.03	$4.00 \times 10^{-40}$
9	<i>TMEM165</i>	4	0.33	$2.45 \times 10^{-9}$	0.045	0.007	$1.84 \times 10^{-10}$	0.45	0.03	$3.52 \times 10^{-37}$
10	<i>RFL</i>	11	0.27	$1.02 \times 10^{-6}$	0.035	0.006	$1.84 \times 10^{-7}$	0.43	0.03	$5.67 \times 10^{-37}$
28	<i>ACADS</i> <sup>b</sup>	12	-0.37	$2.91 \times 10^{-12}$	-0.085	0.01	$7.12 \times 10^{-14}$	-0.24	0.03	$6.65 \times 10^{-19}$
31	<i>LACTB</i> <sup>b</sup>	15	0.30	$1.67 \times 10^{-8}$	0.069	0.01	$1.40 \times 10^{-9}$	0.32	0.04	$4.94 \times 10^{-18}$
34	<i>MAP2K5</i> <sup>b</sup>	15	-0.25	$7.83 \times 10^{-6}$	-0.039	0.01	$1.90 \times 10^{-6}$	-0.21	0.03	$3.81 \times 10^{-10}$

<sup>a</sup> Thirteen representative eGenes, including 4 GWAS loci, ranked by their p-value in the TwinsUK cohort dataset  
<sup>b</sup> GWAS gene  
<sup>c</sup> Effect size (r), Pearson rho and p-value calculated from Pearson correlation between gene expression and BMI (see Methods)  
<sup>d</sup> Effect size, standard error (SE), and p-value calculated using a linear regression model with BMI and age, age<sup>2</sup> and the 14 technical factors as covariates when compared to a null model without BMI. These models were compared using an F-test (see Methods)  
<sup>e</sup> Effect size, standard error (SE), and p-value calculated from linear mixed effects model. A full model including BMI was compared to a null model in which the same model was fitted, but with BMI omitted. These models were compared using an F-test (see Methods)  
<sup>f</sup> Chr is an abbreviation for chromosome

Figure 7). However, a supershift experiment may remain negative even in the presence of true TF binding if a complex instead of a single TF alone is required for the TF binding<sup>20</sup>.

**Interacting GWAS SNPs implicate three other genes.** For *ORMDL3*, there is a single lipid GWAS SNP, rs8076131, in the *HindIII* fragment, which is also the only looping *cis*-eQTL SNP interacting with the *ORMDL3* promoter. Variant rs8076131 lies in a region with enhancer histone marks H3K4me1 and H3K27ac in adipose nuclei (Supplementary Figure 3a,b). The expression of *ORMDL3* is negatively correlated with BMI ( $p = 8.57 \times 10^{-18}$ ), in line with its known role as a negative regulator of sphingolipids that are positively correlated with obesity<sup>21,22</sup>.

The regulatory *HindIII* fragment that loops with the *LACTB* promoter contains two reported metabolite GWAS SNPs in tight LD with each other (rs1017546 and rs3784671,  $r^2 = 0.97$ ), both sharing 35 LD proxies ( $r^2 > 0.80$ ) in the METSIM cohort. One of the two index GWAS SNPs within the *HindIII* fragment, rs3784671, resides in a region enriched for the enhancer histone marks H3K4me1 and H3K27ac in adipose nuclei (Supplementary Figure 3c, d). This metabolite GWAS SNP, rs3784671, is associated with succinylcarnitine levels, which have previously been shown to be positively correlated with BMI in KORA ( $p = 1.0 \times 10^{-12}$ ) and TwinsUK ( $p = 5.3 \times 10^{-5}$ )<sup>23</sup>. Accordingly, the expression of *LACTB* is positively correlated with BMI ( $p = 1.19 \times 10^{-8}$ ). Notably, *LACTB* has been implicated as a causal gene for obesity in mice<sup>24</sup>, further supporting our integrated human data that implicates *LACTB* involvement in an obesity-related metabolic trait.

The most significant metabolite GWAS SNP for *ACADS*, rs10774569, is not located within the regulatory, *cis*-eQTL-containing *HindIII* fragment. Instead, a single *cis*-eQTL SNP rs12310161, in perfect LD ( $r^2 = 1.0$ ) with the GWAS SNP rs10774569, is the only *cis*-eQTL SNP located within the regulatory *HindIII* fragment, looping with the fragment

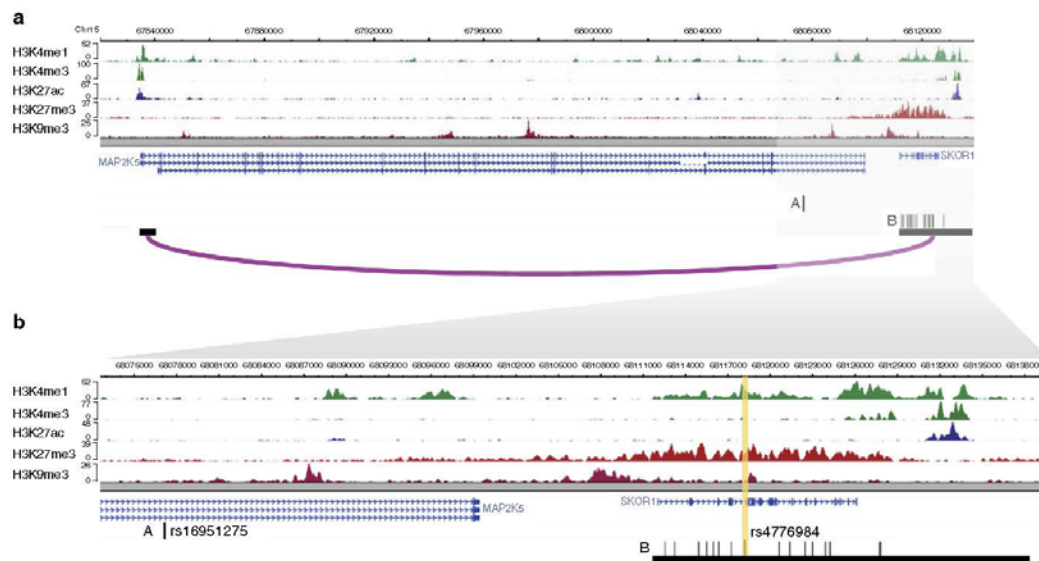
containing the promoter of *ACADS*. This looping *cis*-eQTL SNP also resides in a region enriched for enhancer histone marks H3K4me1 and H3K27ac in adipose nuclei (Supplementary Figure 3e, f). The expression of *ACADS* has a negative correlation with BMI ( $p = 2.91 \times 10^{-12}$ ), and the alternate allele is associated with an increase in expression of *ACADS*, suggesting that this allele has a protective effect against obesity.

Finally, we repeated the pChIP-C experiments in the same HWA cell line in a separate experiment with two replicates and found the same GWAS SNP interactions as in the first experiment (Supplementary Table 9). This validation data thus provides further support for our conclusions and the robustness of interactions we report.

## Discussion

BMI is a highly complex trait caused by the poorly characterized interplay between genetic and environmental factors with upper heritability estimates reaching 70%<sup>2</sup>. Understanding how genome-wide signals with small effect sizes contribute to BMI on a molecular level has proven to be difficult. Delineating the underlying biological mechanisms of these signals is crucial to better understand the development of obesity and its concomitant cardiometabolic disorders. In this study, we performed promoter Capture Hi-C (pChIP-C) in primary human white adipocytes (HWA) to identify BMI-correlated adipose-expressed genes that are under genetic regulation in *cis* by variants that physically interact with gene promoters. Through our method of integrating GWAS, *cis*-eQTL analyses, chromosomal interactions, and robust replication of the data from GTEx and TwinsUK, we were able to identify 42 candidate genes for future obesity research.

In the absence of adipocyte DHS information, we used DHS data from all tissues in the ENCODE and Roadmap Epigenomics project to label open chromatin regions within the adipocyte chromosomal interactions<sup>8</sup>. Despite this methodological compromise, our results demonstrate that variants in these regions



**Fig. 3** Promoter Capture Hi-C enables refinement of the BMI GWAS locus that colocalizes with *cis*-eQTLs interacting with the target gene promoter of *MAP2K5*. Genomic landscape of the BMI locus, *MAP2K5* (panels **a**, **b**), modified from the WashU Genome Browser to show the histone mark calls from ChIP-seq data; gene transcripts; promoter and eQTL *HindIII* fragments that interact in primary human white adipocytes (HWA); and GWAS SNPs (A, the rs number indicated in the magnified box) or their LD proxies (B,  $r^2 > 0.8$ ) located in the interacting *HindIII* fragment. The vertical yellow band highlights the *cis*-eQTL variant (the rs number is indicated in the magnified box). **a** Genomic landscape containing *MAP2K5* and the interacting *cis*-eQTL variant and corresponding BMI GWAS SNP. **b** Magnification of the boxed region in (**a**)

explain a significant portion (4.6%) of the heritability of *cis*-regulated expression in human subcutaneous adipose tissue. Even though the total percentage of variants within the intersection of open chromatin regions and adipocyte chromosomal looping sites is small (0.23%), the enrichment implies that these SNPs are functionally relevant for adipocyte biology and gene regulation in *cis*.

The enrichment of TF binding motifs for CEBPB and PPARG in chromosomal interactions found in adipocyte but not in CD34<sup>+</sup> cells confirms that the regulatory circuits identified here are relevant to adipose biology. These two TFs have previously been shown to occupy shared regulatory sites. Apart from being an enhancer binding protein, which is in concordance with its presence at chromosomal interaction sites, CEBPB has been demonstrated to precede the binding of PPARG at many regulatory sites<sup>25</sup>, suggesting that CEBPB primes the regulatory regions for the binding of the adipose master regulator PPARG.

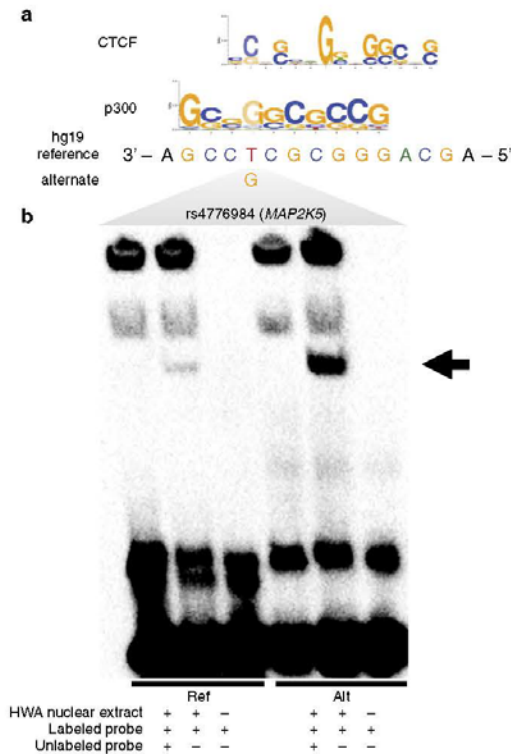
One of our looping *cis*-eQTL variants is a tight LD proxy ( $r^2 = 0.98$ ) for a regional BMI lead GWAS SNP (rs16951275)<sup>2</sup>. Typical fine mapping techniques such as overlaying histone marks, transcription factor motif scans, or eQTL searches do not necessarily reveal the mechanism through which a SNP might function. We refined the GWAS signal from 64 to 16 LD SNPs within a *HindIII* fragment that interacts with the *MAP2K5* promoter by overlaying *cis*-eQTLs, the promoter-enhancer interaction map, and the expression-BMI correlation. The top candidate, rs4776984, increased HWA nuclear protein binding in an allele-specific way in our EMSA experiment and lies within the repressor histone marks H3K27me3 and H3K9me3 in ENCODE adipose nuclei data. Recent studies have suggested that repressor elements function through looping interactions in a similar manner to enhancer elements<sup>6,26</sup>, which would align well with the

negative correlation between expression of *MAP2K5* and BMI level.

The region at the *MAP2K5* locus, exhibiting increased binding for the alternate allele for rs4776984, contains predicted motifs for the looping interaction protein, CTCF, and other TFs (Supplementary Table 8). We did not find evidence of CTCF binding at rs4776984 in our supershift and protein binding EMSA experiments. However, a supershift experiment may remain negative even in the presence of true TF binding if a complex instead of a single TF alone is required for the TF binding<sup>20</sup>. Furthermore, using DeepSEA analysis, we confirmed the potential for differential TF binding at the variant site rs4776984 among all possible looping *cis*-eQTLs at the *MAP2K5* locus. Noteworthy, since DeepSEA identified multiple TFs as potential binders of rs4776984 site in an allele-specific way, future studies testing a larger set of TFs are warranted to identify the actual TF that binds this site. We postulate that TF binding at this interaction site would lead to a repressive looping mechanism, in this case altering *MAP2K5* expression in adipocytes.

*MAP2K5* is a member of the ERK5 MAP kinase signaling cascade, and the importance of ERK5 signaling in adipose was previously demonstrated in *Erk5* knock-out mice, which exhibit increased adiposity<sup>27</sup>. This suggests that changes in ERK5 signaling in adipocytes could be relevant for human obesity. *MAP2K5* is a strong and specific activator of ERK5 in the ERK5 MAP kinase signaling cascade<sup>28</sup>, supporting further study of *MAP2K5* in connection with increased adiposity.

The intronic *ORMDL3* GWAS variant rs8076131 is associated with high-density lipoprotein cholesterol (HDL-C)<sup>15</sup> and is the only *cis*-eQTL SNP in the *HindIII* fragment that interacts with the *ORMDL3* promoter in our adipocyte pChIP-C data. *ORMDL3* is a negative regulator of the synthesis of sphingolipids that are



**Fig. 4** Predicted TF motifs and electrophoretic mobility shift assay (EMSA) at the rs4776984 site indicate allele-specific binding. **a** Predicted TF motifs for CTCF and p300, as well as the hg19 reference genome sequence. **b** Biotinylated (labeled probe) 31-bp oligonucleotide complexes with  $\pm 15$  bp flanking the reference or alternate allele for variant rs4776984 were incubated with nuclear protein extracted from primary HWA and resolved on a 6% polyacrylamide gel. Competitor assays were performed by incubating the reaction with  $\times 100$  excess of unlabeled (no biotin) oligonucleotide complexes with identical sequence. Arrow denotes specific binding of HWA nuclear protein to reference (left) and alternate (right) allele

produced in response to obesity and related metabolic traits, such as inflammation and insulin resistance<sup>21,22</sup>, and that interfere with important signaling pathways associated with these traits<sup>22</sup>. Corroborating this, we show that *ORMDL3* expression is negatively correlated with BMI, and the *cis*-eQTL and risk variant rs8076131 decreases *ORMDL3* expression, potentially through a change in the chromosomal interaction between the enhancer and promoter of *ORMDL3*, as has been shown for this enhancer site previously<sup>29</sup>.

We found that the metabolite GWAS SNP, rs3784671, is a looping *cis*-eQTL variant associated with the expression levels of the *LACTB* gene. Although this variant is a *cis*-eQTL for *LACTB* both in our study and the GTEx adipose cohort, it lies within the promoter for the *APHIB* gene, for which it is not a *cis*-eQTL in our study. Through overlap of adipose *cis*-eQTL data and adipocyte pChI-C data, we established that rs3784671 does not act through the adjacent *APHIB* gene and filtered the 35 *cis*-eQTL variants for *LACTB* down to a single variant, rs3784671. This

variant is negatively associated with the levels of succinylcarnitine, a metabolite positively correlated with BMI in two independent cohorts, KORA and TwinsUK, previously<sup>23</sup>. Succinylcarnitine is a molecule in the butanoate metabolism pathway; butanoate has been implicated in anti-inflammation, protection against obesity, and an increase in leptin levels<sup>30</sup>. Furthermore, as the succinylcarnitine GWAS variant rs3784671 is an eQTL for *LACTB*, associated with an increase in *LACTB* expression, we postulate that *LACTB* expression increases succinylcarnitine. This is in agreement with a mouse study that shows that butanoate metabolism is reduced in *Lactb* transgenic mice<sup>24</sup>. Notably, support for *LACTB* as a causal gene for obesity derives from functional studies using transgenic overexpression of *Lactb* in mice, resulting in an increase in the fat-mass-to-lean-mass ratio<sup>24,31</sup>. Although the function of *LACTB* in adipose has not been fully elucidated, these studies suggest that a reduction in *LACTB* function and, in turn, an increase in butanoate metabolism and decrease of succinylcarnitine levels are beneficial for obesity treatment. Further molecular studies at the protein level are, however, required to determine the function of *ORMDL3* and *LACTB* in connection with obesity.

We identified a perfect LD proxy for a metabolite GWAS SNP that lies within a *HindIII* fragment that regulates the *ACADS* gene and interacts with its promoter. *ACADS* is a mitochondrial protein that catalyzes the first step of the fatty acid beta-oxidation pathway. Proper mitochondrial function is imperative for adipose function and energy homeostasis. In addition to the METSIM and TwinsUK adipose RNA-seq data sets used in our study, a previous study identified *ACADS* when systematically searching for genes over and under-expressed in obese versus lean adipose tissue<sup>32</sup>. Furthermore, all 3 datasets show a consistent negative correlation between *ACADS* expression and BMI, in support of its well-established mitochondrial function. The interacting *cis*-eQTL and GWAS SNP, rs12310161, is located within enhancer histone marks in adipose nuclei and in the HepG2 liver cell line, with the alternate allele exhibiting a positive effect on gene expression, in line with it being a protective allele. Interestingly, this variant falls within a TEA Domain Transcription Factor 4 (*TEAD4*) ChIP-seq peak in the HepG2 cells. *TEAD4* expression is regulated by Peroxisome Proliferator Activated Receptor alpha (*PPAR $\alpha$* )<sup>33</sup>, the major regulator of beta-oxidation of fatty acid pathways in liver and brown adipose tissue. Taken together, these results suggest that the interacting *cis*-eQTL and metabolite GWAS SNP, rs12310161, functions within an enhancer to increase *ACADS* expression and mitochondrial fatty acid beta-oxidation in adipose.

As the pChI-C experiments were performed in primary HWA, we are able to focus on physical chromosomal interactions directly in human adipocytes among all cell types present in adipose tissue. Adipocytes perform central adipose functions, including lipogenesis and lipolysis. Further investigation of the adipose genes, which are under *cis* genetic regulation via chromosomal looping to the promoters and are correlated with BMI, is likely to provide much needed insight into cellular processes contributing to obesity. Our data provide 38 new candidate genes, including some known functionally relevant genes for adiposity such as *LPINI*<sup>34</sup> and *AKRIC*<sup>35</sup>, that have so far not been highlighted by GWAS for BMI or obesity-related metabolic traits. We postulate that identification of some of these 38 candidates as obesity GWAS genes may require much larger GWA studies, while others may represent genes responding to obesity in human adipose tissue. Our analysis of the looping *cis*-eQTLs for other GWAS traits correlated with BMI, such as serum metabolites and lipids, led to the identification three additional obesity-related metabolic GWAS genes. We recognize that brain and other



tissues likely account for some of the BMI GWAS signals and that GWAS variants may act via other mechanisms, such as *trans* regulation and alternative splicing, that warrant future investigation. Although the four looping *cis*-eQTL variants identified at GWAS loci in our study represent either the GWAS tag SNPs (as is the case at the *ORMDL3* and *LACTB* loci) or they are in perfect or almost perfect LD with the GWAS SNP ( $r^2 = 1.0$  at the *ACADS* locus and  $r^2 = 0.98$  at the *MAP2K5* locus), we recognize that the looping variants may not always be the strongest *cis*-eQTL SNPs at these loci and, thus, additional fine mapping is needed to fully elucidate all functional regulatory *cis*-eQTL variants.

The current study uses the integration of multi-level genomic and functional data to enhance the understanding of genome-wide molecular signals underlying obesity. GWAS signals often fall within non-coding regulatory regions of the genome, and the affected gene(s) often remain unclear. Similarly, the local LD structure frequently hinders the identification and functional characterization of the actual eQTL SNP even though the eQTL target gene is known. Through the integration of multi-layer genomics data in a functionally relevant human cell type and tissue and replication in the GTEx and TwinsUK cohorts, we show that the DHSs within the interacting chromosomal regions are enriched for tissue-specific TF motifs and explain a significant proportion of the heritability of gene expression in *cis*. Furthermore, we identified *LACTB*, *ACADS*, *ORMDL3*, and *MAP2K5* as obesity-related genes in humans and provide a set of 38 non-GWAS candidate genes for future studies in obesity.

## Methods

**Cell lines and culture reagents.** We obtained and cultured the primary human white preadipocyte (HWP) cells as recommended by PromoCell (PromoCell C-12731, lot 3952024) for preadipocyte growth and differentiation into adipocytes. Cell media (PromoCell) was supplemented with 1% penicillin-streptomycin. We maintained the cells at 37 °C in a humidified atmosphere at 5% CO<sub>2</sub>. We serum-starved the primary human white adipocytes (HWA) for 16 h using 0.5% fetal calf serum (FCS) in supplemented adipocyte basal medium (PromoCell), prior to treatment with 0.23% fatty acid free bovine serum albumin (BSA, Sigma Aldrich A8806) in media containing 0.5% FCS for 24 h prior to fixation.

**Adipocyte fixation and nuclei collection.** We rinsed 10 M adherent HWA with serum-free media prior to fixation. We fixed the HWA directly in culture plates with 2% formaldehyde (EMD Millipore 344198) in serum-free adipocyte nutrition media (PromoCell). We incubated cells in fixation medium with rocking at room temperature for 1 min, and then quenched with 1 M ice-cold glycine for a final concentration of 125 mM. After 5 min of rocking incubation at room temperature, we rinsed fixed cells twice with ice-cold PBS. Then we incubated the cells with rocking on ice with ice-cold permeabilization buffer (10 mM Tris-HCl pH 8, 10 mM NaCl, 0.2% Igepal CA-630, Complete EDTA-free protease inhibitor cocktail [Roche])<sup>36</sup> for 30 min. We scraped cells from the culture plate and centrifuged at 2500 × g for 5 min at 4 °C to pellet nuclei. The supernatant was discarded and nuclei were flash frozen in liquid nitrogen and put at -80 °C.

**Hi-C library preparation.** We prepared the Hi-C library as described in Rao et al.<sup>7</sup> with modifications. We processed 10 M HWA nuclei in 5 M cell aliquots, closely following Rao et al.<sup>7</sup> protocol I.a.1 except we digested chromatin with 400U of *HindIII* (New England Biolabs R3104) at 37 °C overnight with shaking (950 rpm). After digestion, we pelleted nuclei with centrifugation at 2500 × g for 5 min at 4 °C. We then resuspended nuclei in 265 µl 1 × NEBuffer 2 and removed 10% of the cells and kept on ice for a 3 C control<sup>37</sup>. We followed Rao et al.<sup>7</sup> protocol I.a.1 to end-fill and mark with biotin, perform in-nucleus ligation, degrade protein, and perform ethanol precipitation and purification, except we used biotin-14-dGTP (Invitrogen 19518-018) to incorporate biotin during the end-filling step. After quality control to examine Hi-C marking and ligation efficiency, we sheared 5 µg of DNA to 250–550 bp using the Covaris M220 instrument. We performed double size-selection using Agencourt AMPure XP beads (Beckman Coulter A63881) as described in Rao et al.<sup>7</sup> protocol I.a.1.

We immobilized the fragments containing biotin using DYNAL™ MyOne™ Dynabeads™ Streptavidin T1 (Invitrogen 65601) beads following Rao et al.<sup>7</sup> protocol I.a.1. After end-repair and attachment of dATP, we ligated Illumina paired-end adaptors to the bead-bound library following the SureSelect™ user manual for Illumina Paired-End Multiplexed Sequencing (Agilent Technologies). After washing, we resuspended the Hi-C library in 20 µl of 1 × Tris buffer and

subsequently removed the Streptavidin beads from the DNA by heating at 98 °C for 10 min. We then amplified the adaptor-ligated library using 8 PCR cycles and purified using Agencourt AMPure XP beads, following the SureSelect™ user manual.

**Capture Hi-C.** The RNA baits were designed in Mifsud et al.<sup>6</sup> for capturing *HindIII* fragments containing gene promoters (Dr. Cameron Osborne kindly shared the exact design). As described in Mifsud et al.<sup>6</sup>, 120-mer RNA baits were designed to target both ends of *HindIII* fragments that contain annotated gene promoters (Ensembl promoters of protein-coding, noncoding, antisense, snRNA, miRNA and snoRNA transcripts). The bait sequence was deemed valid if GC content ranged from 25 to 65%, contained <3 consecutive Ns, and was within 330 bp of *HindIII* fragment ends. A total of 550 ng of the Hi-C library was hybridized to the biotinylated RNA baits, captured with DYNAL™ MyOne™ Dynabeads™ Streptavidin T1 beads, and amplified in a post-capture PCR to add indexes, using 12 PCR cycles. The library was sequenced on the Illumina HiSeq 4000 platform.

**Capture Hi-C data processing and interaction calling.** To ensure all downstream analysis was comparable, we first reduced the number of sequencing reads to match the number used in Mifsud et al.<sup>6</sup> analysis of their CHi-C data. We next processed the sequencing reads with the Hi-C User Pipeline (HiCUP) software<sup>38</sup>, aligning reads to the human reference genome (GRCh37/hg19) and using all HiCUP default parameters. We called significant chromosomal interactions with the Capture Hi-C Analysis of Genome Organization (CHiCAGO) software<sup>39</sup>, using default parameters, including the threshold of 5 for calling significant interactions. Briefly, the background is estimated by borrowing information across interactions on two separate components of the data: the interactions with baited fragments in the surrounding genomic region are used to model Brownian collisions, which are distance-dependent interactions, and interchromosomal interactions are used to model technical noise. CHiCAGO then employs a weighted *p*-value based on the expected number of interactions at a range of distances<sup>39</sup>.

**Adipocyte nuclear protein extraction.** Nuclear protein was extracted from adipocytes after centrifugation of cells at 200 × g for 5 min using a nuclear protein extraction kit as recommended (Active Motif 40010). The quantity of protein extracted was measured with BCA protein assay kit (Pierce 23227).

**Electrophoretic mobility shift assay.** Oligonucleotide probes (15 bp flanking SNP site for reference or alternate allele) (Supplementary Table 10) with a biotin tag at the 5' end of the sequence (Integrated DNA Technologies) were incubated with HWA nuclear protein and the working reagent from the Gelshtift Chemiluminescent EMSA kit (Active Motif 37341). For competitor assays, an unlabeled probe of the same sequence was added to the reaction mixture at 100 × excess. The reaction was incubated for 30 min at room temperature, and then loaded on a 6% retardation gel (ThermoFisher Scientific EC6365BOX) that was run in 0.5 × TBE buffer. The contents of the gel were transferred to a nylon membrane, and visualized with the chemi-luminescent reagent as recommended. Similarly, we performed the EMSAs with 1 µg purified CTCF protein (Origene TP720882). Supershift assays were performed with 1 µg anti-CTCF antibodies (Santa Cruz sc-15914 and EMD Millipore 07-729) that were incubated on ice with nuclear protein from HWA for 30 min prior to addition of oligonucleotide probes and run on gel.

**Study cohort.** The study sample consisted of a subset of the participants of the Finnish Metabolic Syndrome in Men (METSIM; *n* = 10,197) cohort, described in detail previously<sup>40,41</sup>. The study was approved by the local ethics committee (Research Ethics Committee, Hospital District of Northern Savo) and all participants gave a written informed consent. The METSIM participants are Finnish males recruited at the University of Eastern Finland and Kuopio University Hospital, Kuopio, Finland. The median age of the METSIM participants is 57 years (range: 45–74 years). The biochemical lipid, glucose, and other clinical and metabolic phenotypes were measured, as described previously<sup>40,41</sup>. A random subset of the METSIM men underwent a subcutaneous abdominal adipose needle biopsy, with 335 unrelated individuals (IBD sharing estimated as <0.2 using a genetic relationship matrix calculated in PLINK<sup>42</sup>) analyzed here using RNA-seq.

**Identification of *cis*-eQTL SNPs.** We processed the METSIM RNA-seq dataset similarly as described in Rodriguez et al.<sup>43</sup>. Briefly, for the METSIM RNA-seq dataset, we isolated total RNA from abdominal subcutaneous adipose needle biopsy using the Qiagen miRNeasy kit. Polyadenylated mRNA was prepared using the Illumina TruSeq RNA Sample Preparation Kit v2 and sequenced using Illumina HiSeq 2000 platform generating paired-end, 50-bp reads. We used STAR<sup>44</sup> to align the reads to the hg19 reference genome, and assembled transcripts using Cufflinks v2.2.1<sup>45</sup>. We filtered genes for those with expression of FPKM > 0 in more than 90% of the samples. Additional details of this dataset have been previously described in Rodriguez et al.<sup>43</sup>. We inverse-normal transformed the FPKMs and adjusted for hidden confounding factors using PEER<sup>46</sup> by removing 22 PEER

factors based on a *cis*-eQTL analysis on chromosome 20 and choosing an optimal number of PEER factors without a loss of statistical power.

To decrease computation time, we prephased the METSIM genotype data, produced using the Illumina HumanOmniExpress BeadChip, by employing SHAPEIT<sup>47</sup> with the phase 1 version 3 reference panel of the 1000 Genomes Project. We performed imputations with the same reference panel and IMPUTE<sup>48</sup> with a cosmopolitan imputation approach, which included all populations from the 1000 Genomes Project, to ensure a high accuracy and maximize the number of imputed SNPs. Imputed data were filtered using the quality control inclusion criteria of info  $\geq 0.8$ , MAF  $\geq 5\%$ , and Hardy-Weinberg equilibrium (HWE)  $p > 0.00001$ . The *cis*-eQTL analysis was performed using Matrix-eQTL<sup>49</sup>. We classified the variants as in *cis* if they were within 1 Mb of either end of a gene. The first three genetic principal components were included as covariates in the *cis*-eQTL analysis to account for population stratification.

**Replication of *cis*-eQTL analysis in GTEx.** To ensure robustness of the results, we filtered the *cis*-eQTL SNPs and their target genes detected in the METSIM cohort so that both the *cis*-eQTL SNP and its predicted target gene were replicated in the *cis*-eQTL data by the GTEx Consortium ( $n = 277$ ) for subcutaneous adipose tissue, filtered using their permutation test for significance, which used the adaptive permutation scheme in FastQTL<sup>50</sup> and a permutation test  $p$ -value threshold equal to the empirical  $p$ -value of the gene closest to the FDR 5% threshold, as reported by GTEx<sup>12</sup>. Only replicated adipose *cis*-eQTLs and their target genes were used in our downstream analyses.

**Heritability of *cis* expression in chromosomal interactions.** To investigate the functional importance of open chromatin regions (i.e., DHSs) within chromosomal interactions in adipocytes to heritability of *cis* expression, we used LD-score regression<sup>51</sup>. More specifically, we generated an annotation for each region within 1 Mb of the TSS of every gene with at least 1 significant promoter interaction. Per gene, this annotation consists of marking the variants within a distal fragment within 1 Mb of the TSS that interact with the fragment containing the promoter of the gene. We further refined these annotations to the open chromatin regions available for TF binding. Accordingly, we only marked those variants located in regions identified in the union of DNase I hypersensitivity sites (DHSs) from all tissues in the ENCODE and Roadmap Epigenomics project<sup>52</sup>. Since these chromosomal interaction annotations change on a per-gene basis, we could not use the genome-wide overlapping matrix in the original software, which treats the annotations as fixed genome-wide. In our analyses, we generated an average overlapping matrix aggregated across all the regions. Importantly, we tested that this weighted overlap matrix does not qualitatively change the overall enrichment of heritability of gene expression for fixed annotations, such as coding regions (Supplementary Figure 1). These changes amount to altering equations 7 and 8 from Liu et al.<sup>11</sup> as follows (Equation 1 and 2).

Equation 1: Modified equation 8 from Liu et al.<sup>11</sup> using a weighted overlap matrix instead of the genome-wide average.

$$\text{prop}_{\text{reg}(C)} = \frac{h_g^2(C)}{h_g^2(\text{total})} = \frac{\sum_C \bar{\tau}_C \bar{M}_C / C}{\sum_C \bar{\tau}_C \bar{M}_C}$$

$$\text{Where } \bar{M} = \frac{\sum_{\text{gene } i} M_i}{NSNP_i}$$

In the equation above,  $C$  is a given annotation category;  $\text{prop}_{\text{reg}(C)}$  is the proportion of heritability for a given category;  $\bar{\tau}_C$  is the regression coefficient for the category;  $\bar{M}$  is the average overlap matrix for each local region;  $M_i$  is the overlap matrix for each gene in the local region; and  $NSNP_i$  is the number of SNPs in each local region.

Equation 2: Modified equation 9 from Liu et al.<sup>11</sup> using the average proportion of SNPs instead of the genome-wide average.

$$\text{enrichment}(C) = \frac{\text{prop}_{\text{reg}(C)}^2}{\text{prop}_{SNPs}(\bar{M})}$$

$$\text{Where } \bar{M} = \frac{\sum_{\text{gene } i} M_i}{NSNP_i}$$

In the equation above, the variables are as in Equation 1, and  $\text{prop}_{\text{SNPs}(\bar{M})}$  is the proportion of SNPs in the overlap matrix for a given category.

**Transcription factor motif enrichment in adipocytes.** We used Hypergeometric Optimization of Motif Enrichment (HOMER, v4.9) to investigate the enrichment of known TFs in the open chromatin regions (i.e., DHSs) within chromosomal interactions in adipocytes<sup>5</sup>. As input data, we used chromosomal interactions in

adipocytes that interacted with a promoter fragment intersected with the union of all DHSs from ENCODE and Epigenomics Roadmap. We chose to use the DHSs in all cell types since there are no publicly available DHS data in adipocytes or adipose. Furthermore, since we were interested in the TF enrichments in adipocytes, we used CD34<sup>+</sup> chromosomal interactions intersected with the union of all DHSs as the background input file<sup>6</sup>. Any regions that were shared between the CD34<sup>+</sup> and adipocyte datasets were not considered in this analysis. We considered significant any TFs that were enriched in the DHSs within chromosomal interactions in adipocytes at an FDR of 5%. To ensure our background input file was not biasing the results, we also performed the same analysis with all DHSs not found in adipocyte chromosomal interactions as the background input.

We also assessed predicted differential TF binding using the tool deep learning-based sequence analyzer (DeepSEA)<sup>19</sup>, which assesses differential histone modification, TF binding, and DHS profiles using a deep learning-based algorithmic approach and gives a functional significance score at the single nucleotide resolution.

**Overlap of *cis*-eQTL SNPs and chromosomal interactions.** To investigate functional *cis*-eQTL SNPs, we overlapped the imputed *cis*-eQTL SNPs and their target genes with Capture Hi-C chromosomal interactions by first overlapping the position other end of the looping interaction with the location of the *cis*-eQTL SNP. These were subsequently designated as regulatory element *cis*-eQTL SNPs. Simultaneously, we examined the identity of the predicted target gene for the *cis*-eQTL SNP and the gene involved in the looping interaction for a match. Only when both these criteria were fulfilled, was the *cis*-eQTL SNP defined as a looping *cis*-eQTL SNP and considered for further analyses.

**Identification of LD proxies of GWAS SNPs.** GWAS variants associated with BMI were obtained from Locke et al.<sup>2</sup>, and with lipids and metabolites from Willer et al.<sup>15</sup> and Shin et al.<sup>14</sup>. We identified the *cis*-eQTL SNPs in tight LD ( $r^2 > 0.80$ ) with GWAS variants in the METSIM adipose RNA-seq dataset using PLINK<sup>42</sup> and used them as the LD proxies for BMI, lipid, and metabolite GWAS SNPs. These sets of *cis*-eQTL SNPs were considered as BMI GWAS SNPs, lipid GWAS SNPs, and metabolite GWAS SNPs, respectively. These set of BMI, lipid, and metabolite GWAS SNPs were then overlapped with the looping *cis*-eQTL SNPs to identify all BMI, lipid, and metabolite GWAS SNPs involved in chromosomal interactions acting through distant regulatory elements.

**Correlation of BMI with adipose gene expression.** The BMI measurements in the METSIM cohort were first adjusted for age, age<sup>2</sup> and then the resulting residuals were inverse normal transformed to reduce the possible outlier effects. Next, we log transformed the FPKM values and then corrected them for 14 technical factors, including the RIN values, batch, percentage of coding reads, 5' to 3' bias, and percentage of uniquely mapped reads using Picard tools. The expression levels were correlated with the BMI measurements using Pearson correlation. The  $p$ -values were corrected for multiple testing for the number of eGenes using the Bonferroni correction (adjusted  $p$ -value  $< 0.05$ ). To directly compare the effects sizes and  $p$ -values obtained for BMI associations in TwinsUK with those in METSIM, we also tested the 42 replicated genes using a linear regression model with BMI and age, age<sup>2</sup> and the 14 technical factors as covariates when compared to a null model without BMI in METSIM (Table 1 and Supplementary Table 6). These models were compared using an F-test.

**Replication of BMI-adipose gene expression correlation.** Association analysis between BMI and adipose expression in the TwinsUK cohort was performed on 720 female twins. RNA-seq was generated as previously described<sup>53</sup> and gene level quantifications were generated to Gencode v19. Association between gene expression level and inverted normalized BMI was tested with a linear mixed effects model (LME) implemented using the lme4 package<sup>53</sup>. A full model including BMI was compared to a null model in which the same model was fitted, but with BMI omitted. These models were compared using an F-test. All known technical variables (batch, GC content, insert size mode, and primer index), age, age<sup>2</sup>, and family structure were included as covariates in the models. All variables were centered and scaled to unit variance. Four genes were not present in the TwinsUK cohort dataset and we were thus unable to test them for replication, resulting in 54 genes tested for replication. Each replicated gene was examined to determine if effect size direction in TwinsUK and METSIM has the same direction. A Bonferroni corrected  $p$ -value (adjusted  $p < 0.001$ ) with the same direction of effect as in METSIM was considered as statistical evidence for replication in the TwinsUK.

**Data availability.** The human primary adipocyte Capture Hi-C data are available at GEO (Accession ID: GSE110619)

Received: 12 September 2017 Accepted: 22 February 2018

Published online: 17 April 2018

## References

- Gregg, E. W. & Shaw, J. E. Health effects of overweight and obesity in 195 countries over 25 years. *N. Engl. J. Med.* **377**, 13–27 (2017).
- Locke, A., Kahali, B., Berndt, S., Justice, A. & Pers, T. Genetic studies of body mass index yield new insights for obesity biology. *Nature* **518**, 197–206 (2015).
- Claussnitzer, M. et al. FTO obesity variant circuitry and adipocyte browning in humans. *N. Engl. J. Med.* **373**, 895–907 (2015).
- Hnisz, D. et al. Activation of proto-oncogenes by disruption of chromosome neighborhoods. *Science* **351**, 1454–1458 (2016).
- Lupianez, D. G. et al. Disruptions of topological chromatin domains cause pathogenic rewiring of gene-enhancer interactions. *Cell* **161**, 1012–1025 (2015).
- Mifaud, B. et al. Mapping long-range promoter contacts in human cells with high-resolution capture Hi-C. *Nat. Genet.* **47**, 598–606 (2015).
- Rao, S. S. P. et al. A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. *Cell* **159**, 1665–1680 (2014).
- Trynka, G. & Raychaudhuri, S. Using chromatin marks to interpret and localize genetic associations to complex human traits and diseases. *Curr. Opin. Genet. Dev.* **23**, 635–641 (2013).
- Heinz, S. et al. Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. *Mol. Cell* **38**, 576–589 (2010).
- Moseti, D., Regassa, A. & Kim, W. K. Molecular regulation of adipogenesis and potential anti-adipogenic bioactive molecules. *Int. J. Mol. Sci.* **17**, 124 (2016).
- Liu, X. et al. Functional architectures of local and distal regulation of gene expression in multiple human tissues. *Am. J. Hum. Genet.* **100**, 605–616 (2017).
- Ardlie, K. G. et al. The Genotype-Tissue Expression (GTEx) pilot analysis: multitissue gene regulation in humans. *Science* **348**, 648–660 (2015).
- Wang, J., Duncan, D., Shi, Z. & Zhang, B. WEB-based Gene Set Analysis Toolkit (WebGestalt): update 2013. *Nucleic Acids Res.* **41**, W77–W83 (2013).
- Shin, S.-Y. et al. An atlas of genetic influences on human blood metabolites. *Nat. Genet.* **46**, 543–550 (2014).
- Willer, C. J. et al. Discovery and refinement of loci associated with lipid levels. *Nat. Genet.* **45**, 1274–1283 (2013).
- Fuchsberger, C. et al. The genetic architecture of type 2 diabetes. *Nature* **536**, 41–47 (2016).
- Shungin, D. et al. New genetic link local adipose and insulin biology to body fat distribution. *Nature* **518**, 187–196 (2015).
- Kheradpour, P. & Kelis, M. Systematic discovery and characterization of regulatory motifs in ENCODE TF binding experiments. *Nucleic Acids Res.* **42**, 2976–2987 (2014).
- Zhou, J. & Troyanskaya, O. G. Predicting effects of noncoding variants with deep learning-based sequence model. *Nat. Methods* **12**, 931–934 (2015).
- Roman, T. S. et al. A type 2 diabetes-associated functional regulatory variant in a pancreatic islet enhancer at the ADCY5 locus. *Diabetes* **66**, 2521–2530 (2017).
- Russo, S. B., Ross, J. S. & Cowart, L. A. Sphingolipids in obesity, type 2 diabetes, and metabolic disease. *Handb. Exp. Pharmacol.* **216**, 373–401 (2013).
- Kang, S. C., Kim, B. R., Lee, S. Y. & Park, T. S. Sphingolipid metabolism and obesity-induced inflammation. *Front. Endocrinol.* **4**, 67 (2013).
- Suhre, K. et al. Human metabolic individuality in biomedical and pharmaceutical research. *Nature* **477**, 714–715 (2011).
- Yang, X. et al. Validation of candidate causal genes for obesity that affect shared metabolic pathways and networks. *Nat. Genet.* **41**, 415–423 (2009).
- Lefterova, M., Zhang, Y. & Steger, D. PPARgamma and C/EBP factors orchestrate adipocyte biology via adjacent binding on a genome-wide scale. *Genes Dev.* **22**, 2941–2952 (2008).
- Jin, F. et al. A high-resolution map of the three-dimensional chromatin interactome in human cells. *Nature* **503**, 290–294 (2013).
- Zhu, H. et al. Role of extracellular signal-regulated kinase 5 in adipocyte signaling. *J. Biol. Chem.* **289**, 6311–6322 (2014).
- Kato, Y. et al. BMK1/ERK5 regulates serum-induced early gene expression through transcription factor MEF2C. *EMBO J.* **16**, 7054–7066 (1997).
- Schmiedel, B. J. et al. 17q21 asthma-risk variants switch CTCF binding and regulate IL-2 production by T cells. *Nat. Commun.* **7**, 13426 (2016).
- Chakraborti, C. K. New-found link between microbiota and obesity. *World J. Gastrointest. Pathophysiol.* **6**, 110–119 (2015).
- Chen, Y. et al. Variations in DNA elucidate molecular networks that cause disease. *Nature* **452**, 429–435 (2008).
- Henegar, C. et al. Adipose tissue transcriptomic signature highlights the pathological relevance of extracellular matrix in human obesity. *Genome Biol.* **9**, R14 (2008).
- Kaneko, K. J. & DePamphilis, M. L. TEAD4 establishes the energy homeostasis essential for blastocoele formation. *Development* **140**, 3680–3690 (2013).
- Phan, J. & Reue, K. Lipin, a lipodystrophy and obesity gene. *Cell Metab.* **1**, 73–83 (2005).
- O'Reilly, M. W. et al. AKR1C3-mediated adipose androgen generation drives lipotoxicity in women with polycystic ovary syndrome. *J. Clin. Endocrinol. Metab.* **102**, 3327–3339 (2017).
- Nagano, T. et al. Comparison of Hi-C results using in-solution versus in-nucleus ligation. *Genome Biol.* **16**, 175 (2015).
- Lieberman-Aiden, E. et al. Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Sci. (80-.)* **326**, 289–293 (2009).
- Wingett, S. W. et al. HiCUP: pipeline for mapping and processing Hi-C data. *F1000Research* **4**, (2015).
- Cairns, J. et al. CHiCAGO: robust detection of DNA looping interactions in Capture Hi-C data. *Genome Biol.* **17**, 127 (2016).
- Stančáková, A. et al. Hyperglycemia and a common variant of GCKR are associated with the levels of eight amino acids in 9,369 finnish men. *Diabetes* **61**, 1895–1902 (2012).
- Laakso, M. et al. METabolic Syndrome In Men (METSIM) Study: a resource for studies of metabolic and cardiovascular diseases. *J. Lipid Res.* **58**, 481–493 (2017).
- Purcell, S. et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.* **81**, 559–575 (2007).
- Rodríguez, A. et al. Molecular characterization of the lipid genome-wide association study signal on chromosome 18q11.2 implicates HNF4A-mediated regulation of the TMEM241 gene. *Arterioscler. Thromb. Vasc. Biol.* **36**, 1350–1355 (2016).
- Dobin, A. et al. STAR: Ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21 (2013).
- Trapnell, C. et al. Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nat. Biotechnol.* **28**, 511–515 (2010).
- Stegle, O., Parts, L., Piipari, M., Winn, J. & Durbin, R. Using probabilistic estimation of expression residuals (PEER) to obtain increased power and interpretability of gene expression analyses. *Nat. Protoc.* **7**, 500–507 (2012).
- Delaneau, O. et al. Integrating sequence and array data to create an improved 1000 Genomes Project haplotype reference panel. *Nat. Commun.* **5**, 3934 (2014).
- Howie, B. N., Donnelly, P. & Marchini, J. A flexible and accurate genotype imputation method for the next generation of genome-wide association studies. *PLoS Genet.* **5**, e1000529 (2009).
- Shabalina, A. A. Matrix eQTL: ultra fast eQTL analysis via large matrix operations. *Bioinformatics* **28**, 1353–1358 (2012).
- Ongen, H., Buil, A., Brown, A. A., Dermizakis, E. T. & Delaneau, O. Fast and efficient QTL mapper for thousands of molecular phenotypes. *Bioinformatics* **32**, 1479–1485 (2016).
- Finucane, H. K. et al. Partitioning heritability by functional annotation using genome-wide association summary statistics. *Nat. Genet.* **47**, 1228–1235 (2015).
- Buil, A. et al. Gene-gene and gene-environment interactions detected by transcriptome sequence analysis in twins. *Nat. Genet.* **47**, 88–91 (2014).
- Bates, D., Mächler, M., Bolker, B. & Walker, S. Fitting linear mixed-effects models using lme4. *J. Stat. Softw.* **67**, 1–48 (2015).

## Acknowledgements

We thank the individuals who participated in the METSIM and GTEx studies. We also thank the sequencing core at UCLA for performing the RNA sequencing. In addition, we thank Cameron Osborne for his advice with the CHi-C protocol. We thank Xuanyao Liu for his assistance with the LD Score software. Francis Collins is thanked for providing the METSIM genotype data. The Genotype-Tissue Expression (GTEx) Project was supported by the Common Fund of the Office of the Director of the National Institutes of Health, and by NCI, NHGRI, NHLBI, NIDA, NIMH, and NINDS. The data used for the analyses described in this manuscript were obtained from: the GTEx Portal on 03/23/17. This study was funded by National Institutes of Health (NIH) grants HL-095056, HL-28481, U01 DK105561, R01 HL121172, and DK093757. D.Z.P. was supported by the NIH-NCI National Cancer Institute grant T32LM012424, M.A. was supported by the NIH grant T32HG002536, and A.K. by NIH grant F31HL127921. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the article. Genotyping for the METSIM cohort were supported by NIH grants DK072193, DK093757, DK062370, and Z01HG000024 and provided by the Center for Inherited Disease Research (CIDR). CIDR is fully funded through a federal contract from the NIH to The Johns Hopkins University, contract number HHSN268201200008I.

## Author contributions

D.Z.P., K.M.G., P.P., and A.K. designed the study. D.Z.P., K.M.G., J.B., P.P., A.K., R.M.C., J.S.S. and Z.M. performed methods development and statistical analysis. D.Z.P., K.M.G., M.A., Z.M., and J.B. performed computation analysis of the data. K.M.G. and Y.V.B. performed the experiments. A.K., E.N., M.A., K.L.M., C.R., and P.P. performed RNA-sequencing and quality control. M.L. performed phenotyping. M.C., A.J.L., M.L., E.N., K.

L.M., M.B., and P.P. performed data collection and METSIM genotyping. C.A.G. and K.S. S performed the replication analysis (TwinsUK). D.Z.P., K.M.G., A.K., and P.P. wrote the manuscript and all authors read, reviewed, and/or edited the manuscript.

### Additional information

**Supplementary Information** accompanies this paper at <https://doi.org/10.1038/s41467-018-03554-9>.

**Competing interests:** The authors declare no competing interests.

**Reprints and permission** information is available online at <http://npg.nature.com/reprintsandpermissions/>

**Publisher's note:** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

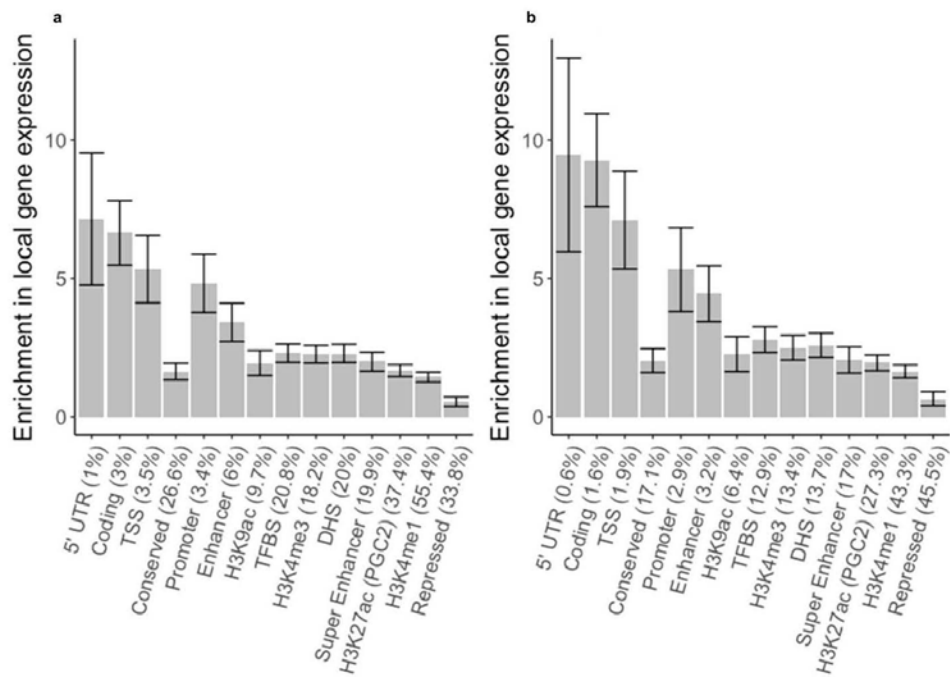


**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2018

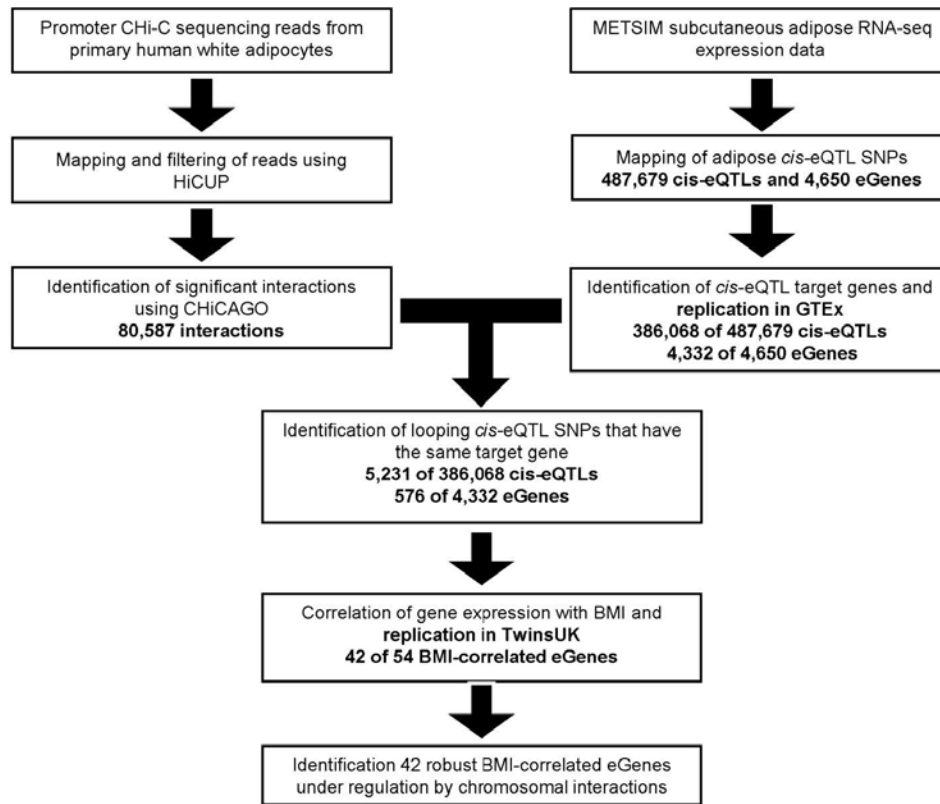
Integration of human adipocyte chromosomal interactions with adipose gene expression  
prioritizes obesity-related genes from GWAS

Pan et al.



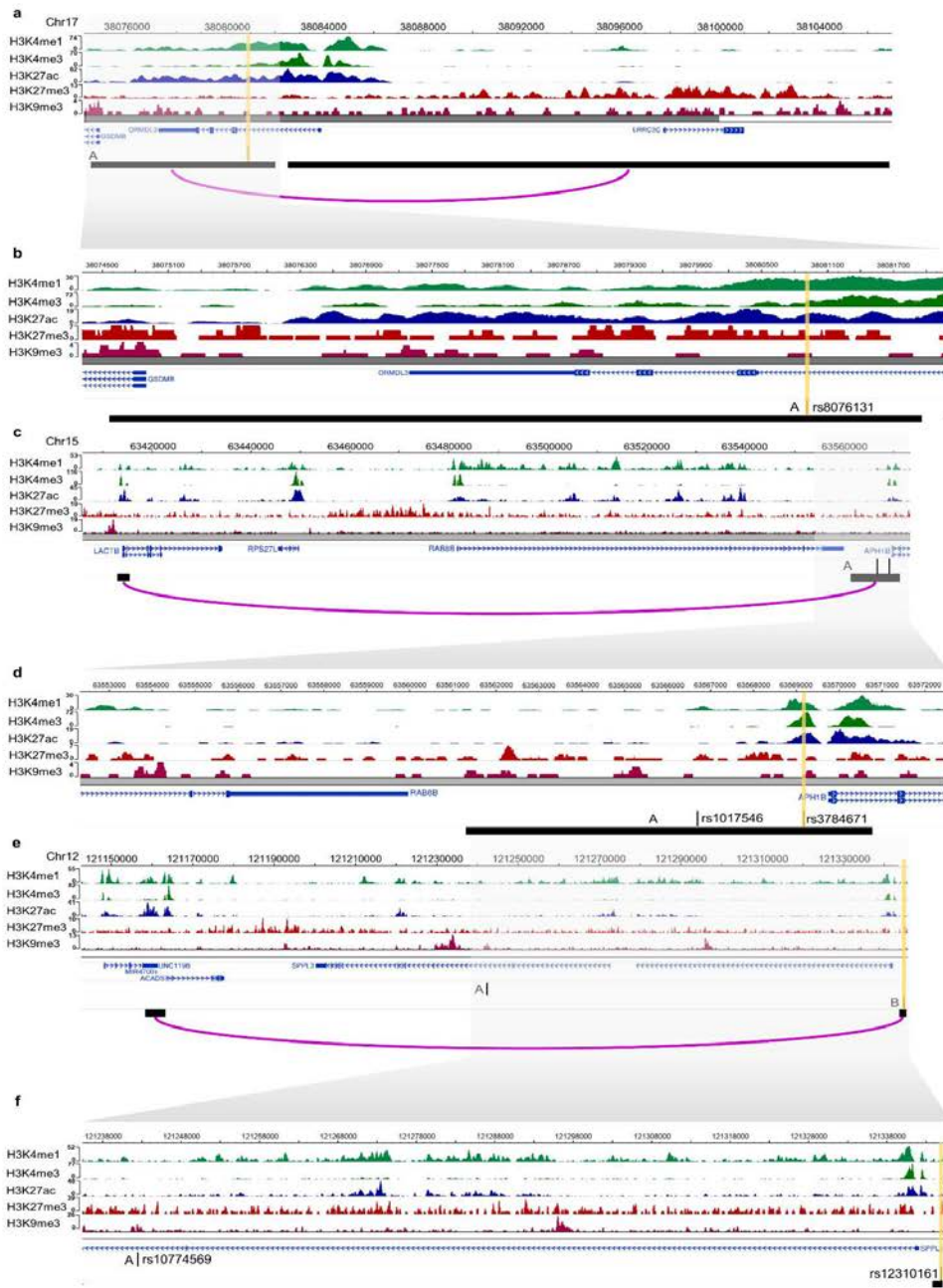
**Supplementary Figure 1. Modification to LD Score regression software does not show significant changes when compared with the data obtained using the published version.**

Enrichments in local gene expression with error bars for different categories using the LD score regression analysis. For the horizontal axis labels, the value in parentheses shows the percentage of SNPs contained within the respective annotation category that contributed to the enrichment calculation (for the full data on all 52 baseline annotation categories, see Supplementary Table 3-4). Error bars represent jackknife standard errors around the estimates of enrichment. (a) Enrichment in local gene expression for the modified LD Score regression software. (b) Enrichment in local gene expression for the original, unmodified LD Score regression software.



**Supplementary Figure 2. Overview of the study design targeted to identify causal and reactive BMI-correlated genes.**

Flow chart showing the data processing and analysis pipeline of the promoter Capture Hi-C in primary human white adipocytes (HWA) (the left side); adipose RNA-sequencing followed by *cis*-eQTL mapping (the right side); and the integration of these genomics data (in the middle) to identify eGenes correlated with BMI.

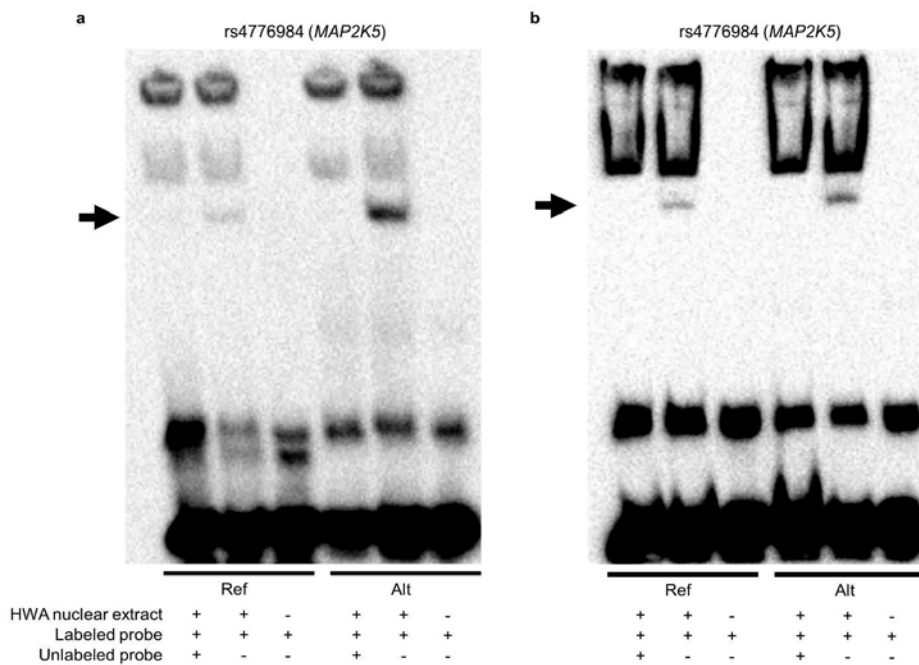




**Supplementary Figure 3. Promoter Capture Hi-C enables refinement of the GWAS loci that colocalizes with *cis*-eQTLs interacting with the target gene promoter of *ORMDL3*, *LACTB*, and *ACADS*.**

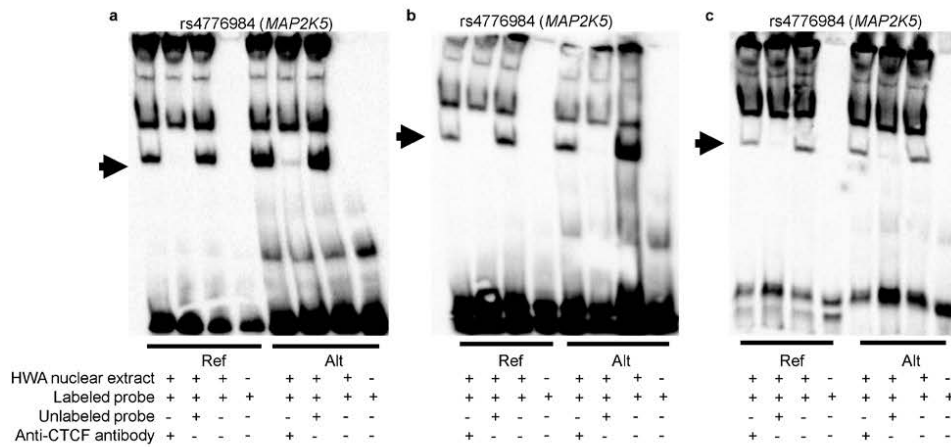
Genomic landscape of the lipid GWAS locus, *ORMDL3* (panels a, b), metabolite GWAS locus, *LACTB* (panels c, d), and metabolite GWAS locus, *ACADS* (panels e, f), modified from the WashU Genome Browser to show the histone mark calls from ChIP-seq data; gene transcripts; promoter and eQTL *Hind*III fragments that interact in primary human white adipocytes (HWA); and GWAS SNP (A, the rs number indicated in the magnified box) or their LD proxies if applicable (B,  $r^2 > 0.80$ ) located in the interacting *Hind*III fragment. The vertical yellow band highlights the significantly influential variant (the rs number is indicated in the magnified box).

(a) Genomic landscape containing *ORMDL3* and the interacting lipid GWAS SNP. (b) Magnification of the boxed region in (a). (c) Genomic landscape containing *LACTB* and the interacting metabolite GWAS SNPs. (d) Magnification of the boxed region in (c). (e) Genomic landscape containing *ACADS* and the interacting *cis*-eQTLs and corresponding metabolite GWAS SNP. (f) Magnification of the boxed region in (e).



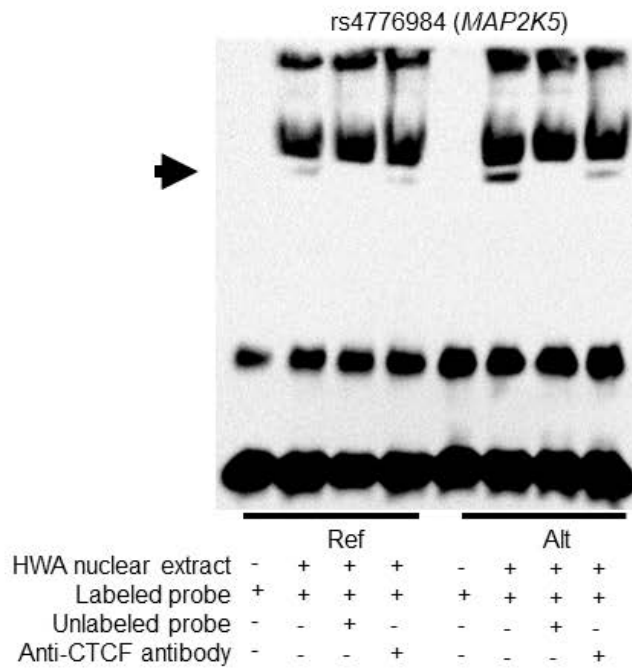
**Supplementary Figure 4. Two independent replicates for the Electrophoretic mobility shift assay (EMSA) data show increased binding of nuclear protein extracted from primary human white adipocytes (HWA) to the alternate allele when compared to the reference allele of the *MAP2K5* cis-eQTL SNP rs4776984.**

Biotinylated (labeled probe) 31-bp oligonucleotide complexes with +/-15 bp flanking the reference or alternate allele for variant rs4776984 were incubated with nuclear protein extracted from primary HWA and resolved on a 6% polyacrylamide gel. Competitor assays were performed by incubating the reaction with 100X excess of unlabeled (no biotin) oligonucleotide complexes with identical sequence. Arrow denotes specific binding of HWA nuclear protein to reference (left) and alternate (right) allele. (a) First replicate of the EMSA for rs4776984. (b) Second replicate of the EMSA for rs4776984.



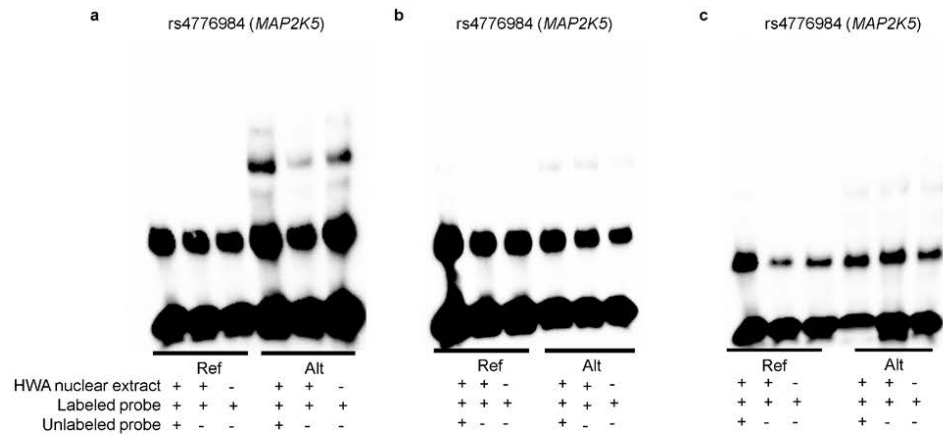
**Supplementary Figure 5. Three independent replicates for the Electrophoretic mobility shift assay (EMSA) do not show a supershift when using antibody against CTCF and nuclear protein extracted from primary human white adipocytes (HWA) at the *MAP2K5* *cis*-eQTL SNP rs4776984.**

Biotinylated (labeled probe) 31-bp oligonucleotide complexes with +/-15 bp flanking the reference or alternate allele for variant rs4776984 were incubated with nuclear protein extracted from primary HWA and resolved on a 6% polyacrylamide gel. Competitor assays were performed by incubating the reaction with 100X excess of unlabeled (no biotin) oligonucleotide complexes with identical sequence. Arrow denotes specific binding of HWA nuclear protein to reference (left) and alternate (right) allele. Supershift assays were performed with 1 µg anti-CTCF antibodies (Santa Cruz sc-15914). (a) First replicate of the supershift EMSA for rs4776984. (b) Second replicate of the supershift EMSA for rs4776984. (c) Third replicate of the supershift EMSA for rs4776984.



**Supplementary Figure 6. The Electrophoretic mobility shift assay (EMSA) does not show a supershift when using a different antibody against CTCF and nuclear protein extracted from primary human white adipocytes (HWA) at the *MAP2K5* cis-eQTL SNP rs4776984.**

Biotinylated (labeled probe) 31-bp oligonucleotide complexes with +/-15 bp flanking the reference or alternate allele for variant rs4776984 were incubated with nuclear protein extracted from primary HWA and resolved on a 6% polyacrylamide gel. Competitor assays were performed by incubating the reaction with 100X excess of unlabeled (no biotin) oligonucleotide complexes with identical sequence. Arrow denotes specific binding of HWA nuclear protein to reference (left) and alternate (right) allele. Supershift assays were performed with 1µg anti-CTCF antibodies (EMD Millipore 07-729).



**Supplementary Figure 7. Three independent replicates for the Electrophoretic mobility shift assay (EMSA) do not show specific binding using purified CTCF protein at the *MAP2K5* cis-eQTL SNP rs4776984.**

Biotinylated (labeled probe) 31-bp oligonucleotide complexes with +/-15 bp flanking the reference or alternate allele for variant rs4776984 were incubated with purified CTCF protein (Origene TP720882) and resolved on a 6% polyacrylamide gel in our EMSA experiment.

Competitor assays were performed by incubating the reaction with 100X excess of unlabeled (no biotin) oligonucleotide complexes with identical sequence. The reference allele is on the left and alternate allele on the right. (a) First replicate of the EMSA for rs4776984. (b) Second replicate of the EMSA for rs4776984. (c) Third replicate of the EMSA for rs4776984.

**Supplementary Table 1. Parameters used for identification of novel *cis*-eQTL and looping interactions**

<i>cis</i> -eQTL discovery	METSIM (n=335)
Type of genetic data	Illumina Omni Express
# <i>cis</i> -eQTL SNPs with the same target gene and beta direction replicated in subcutaneous adipose GTEx data	396,069
# PEER factors corrected	22
# Genetic principal components corrected	3
Minor allele frequency (MAF)	> 5%
Type of expression data	RNA-seq
Normalization technique	Inverse normal transformation of FPKMs
FDR significance threshold for <i>cis</i> -eQTL SNPs	< 5%
# of <i>cis</i> -eQTL target genes with looping interactions	4,332
Promoter capture Hi-C	Primary human white adipocytes
# reads from sequencing	138,217,259
# uniquely aligned paired reads	101,187,918
# valid pairs of reads after capture Hi-C specific filtering by HiCUP	88,583,089
# significant looping interaction pairs identified from CHICAGO	80,567
# METSIM genes in looping interaction pairs	10,083

Supplementary Table 2. Histone mark enrichment in looping *HindIII* fragments in primary HWA

Histone mark	Base pairs of feature enrichment in looping <i>HindIII</i> fragments	Base pairs of feature enrichment in random <i>HindIII</i> fragments*	Standard deviation	<i>p</i> -value <sup>†</sup>
H3K4me1	42181	39278.45	56.94	<2.2x10 <sup>-16</sup>
H3K4me3	42347	39502.39	55.86	<2.2x10 <sup>-16</sup>
H3K27ac	42095	39597.19	49.24	<2.2x10 <sup>-16</sup>
H3K27me3	42813	40529.33	43.78	<2.2x10 <sup>-16</sup>
H3K9me3	41222	39408.19	54.07	<2.2x10 <sup>-16</sup>
DHS	35578	30547.74	89.82	<2.2x10 <sup>-16</sup>

\*Random *HindIII* fragments were controlled for distance away from the target promoter when selected.

<sup>†</sup>*p*-value computed from Pearson's chi-squared test.

Supplementary Table 3. Adipocyte chromosomal interactions are enriched for 30 transcription factors (adjusted  $p < 0.05$ ) when compared to CD34+ chromosomal interactions

Motif logo	Motif name	p-value	Adjusted p-value	Number of target sequences with motif (of 189013)	Percent of target sequences with motif	Number of background sequences with motif (of 173261)	Percent of background sequences with motif
	CTCF(Zf)/CD34+ CTCF-ChIP-Seq(Sarski_et_al)/Homer	$1.00 \times 10^{-102}$	0	5746	3.04%	3915.7	2.26%
	BORIS(Zf)/K562-CTCF-ChIP-Seq(GSE32465)/Homer	$1.00 \times 10^{-63}$	0	6338	3.35%	4760.1	2.75%
	CEBP(iz2IP)/ThioMac-CEBPb-ChIP-Seq(GSE21512)/Homer	$1.00 \times 10^{-10}$	0	19025	10.01%	16547.8	9.56%
	Sp5(Zf)/mES-Sp5 Flag-ChIP-Seq(GSE72989)/Homer	$1.00 \times 10^{-7}$	0	18995	9.99%	14950.8	8.64%
	Elk4(ETS)/Hela-Elk4-ChIP-Seq(GSE31477)/Homer	$1.00 \times 10^{-9}$	0.00010	10491	5.55%	9180.3	5.31%
	YY1(Zf)/Promoter/Homer	$1.00 \times 10^{-9}$	0.00030	917	0.43%	638.7	0.37%
	NRF1(NRF)/MCF7-NRF1-ChIP-Seq(Unpublished)/Homer	$1.00 \times 10^{-6}$	0.00040	1163	0.61%	935.2	0.54%
	TEAD2(TEA)/Py21-Tead2-ChIP-Seq(GSE55709)/Homer	$1.00 \times 10^{-4}$	0.00050	12965	6.89%	11440.4	6.61%
	E2F3(E2F)/MEF-E2F3-ChIP-Seq(GSE71379)/Homer	$1.00 \times 10^{-4}$	0.00070	7993	4.23%	6996.9	4.04%
	Era(NR)/HepG2-Era-ChIP-Seq(GSE31477)/Homer	$1.00 \times 10^{-4}$	0.0023	40662	21.50%	36580.6	21.14%
	TEAD(TEA)/Fibroblast-PU.1-ChIP-Seq(Unpublished)/Homer	$1.00 \times 10^{-4}$	0.0028	17798	9.41%	15648.7	9.16%
	Elk1(ETS)/Hela-Elk1-ChIP-Seq(GSE31477)/Homer	$1.00 \times 10^{-4}$	0.0028	10671	5.64%	9422.4	5.45%
	TEAD4(TEA)/Tropoblast-Tead4-ChIP-Seq(GSE37350)/Homer	$1.00 \times 10^{-4}$	0.0028	20464	10.82%	18265.1	10.56%
	GFY(7)/Promoter/Homer	$1.00 \times 10^{-3}$	0.0031	1159	0.61%	950.2	0.55%
	E2F4(E2F)/K562-E2F4-ChIP-Seq(GSE31477)/Homer	$1.00 \times 10^{-3}$	0.0069	5131	2.71%	4475.5	2.59%
	Sp1(Zf)/Promoter/Homer	$1.00 \times 10^{-3}$	0.0094	3622	1.92%	3133.5	1.81%
	NFY(CCAAT)/Promoter/Homer	$1.00 \times 10^{-3}$	0.0094	14237	7.53%	12676.1	7.33%
	Ronin(THAP)/ES-Thap11-ChIP-Seq(GSE51522)/Homer	$1.00 \times 10^{-3}$	0.012	445	0.23%	346.4	0.20%
	Olig2(BHLH)/Neuron-Olig2-ChIP-Seq(GSE30892)/Homer	$1.00 \times 10^{-3}$	0.014	45873	24.26%	41429.3	23.94%
	E2F6(E2F)/Hela-E2F6-ChIP-Seq(GSE31477)/Homer	$1.00 \times 10^{-3}$	0.022	7047	3.73%	6221.3	3.60%
	ZNF143(STAF/Zf)/CUTLL-ZNF143-ChIP-Seq(GSE29600)/Homer	$1.00 \times 10^{-2}$	0.024	5580	2.95%	4906.1	2.84%
	DUX4(Homobox)/Myoblasts-DUX4-V5-ChIP-Seq(GSE75791)/Homer	$1.00 \times 10^{-2}$	0.029	936	0.49%	777.7	0.45%
	NRF1(NRF)/Promoter/Homer	$1.00 \times 10^{-2}$	0.036	1622	0.86%	1382.4	0.80%
	PPARE(NR)/DR1G3L1-Pparg-ChIP-Seq(GSE13511)/Homer	$1.00 \times 10^{-2}$	0.039	17385	9.19%	15586.6	9.01%
	Pax7(Paired/Homobox)/longest/Myoblast-Pax7-ChIP-Seq(GSE25094)/Homer	$1.00 \times 10^{-2}$	0.040	513	0.27%	415	0.24%
	NFAT1(RHD)/Jurkat-NFAT1-ChIP-Seq(Jolma_et_al)/Homer	$1.00 \times 10^{-2}$	0.045	20039	10.60%	18002.5	10.40%



Supplementary Table 4. LD score enrichments, heritability estimates, and p-values using the published LD Score software<sup>9</sup>

Category	Prop. of SNPs	Prop. of h <sup>2</sup>	Enrichment	SE	p-value
Coding_UCSC	0.02	0.15	9.26	0.9	6.66x10 <sup>-25</sup>
Coding_UCSC.extend.500	0.07	0.28	4.21	0.3	1.27x10 <sup>-42</sup>
Conserved_LindbladToh	0.03	0.15	5.32	0.8	1.26x10 <sup>-53</sup>
Conserved_LindbladToh.extend.500	0.34	0.56	1.64	0.1	4.45x10 <sup>-7</sup>
CTCF_Hoffman	0.02	0.06	2.52	0.5	4.92x10 <sup>-3</sup>
CTCF_Hoffman.extend.500	0.07	0.11	1.58	0.3	2.59x10 <sup>-2</sup>
DGF_ENCODE	0.14	0.34	2.46	0.2	6.87x10 <sup>-5</sup>
DGF_ENCODE.extend.500	0.54	0.77	1.42	0.1	2.60x10 <sup>-5</sup>
DHS_peaks_Trynka	0.11	0.27	2.37	0.3	3.92x10 <sup>-5</sup>
DHS_Trynka	0.17	0.35	2.05	0.2	8.32x10 <sup>-5</sup>
DHS_Trynka.extend.500	0.50	0.72	1.44	0.1	3.56x10 <sup>-5</sup>
Enhancer_Andersson	0.00	0.01	2.60	2.0	4.30x10 <sup>-1</sup>
Enhancer_Andersson.extend.500	0.02	0.03	1.70	0.7	2.87x10 <sup>-1</sup>
Enhancer_Hoffman	0.06	0.15	2.27	0.3	6.34x10 <sup>-7</sup>
Enhancer_Hoffman.extend.500	0.16	0.30	1.93	0.2	2.51x10 <sup>-7</sup>
FetalDHS_Trynka	0.09	0.28	3.25	0.3	3.78x10 <sup>-12</sup>
FetalDHS_Trynka.extend.500	0.29	0.48	1.69	0.2	2.28x10 <sup>-5</sup>
H3K27ac_Hnisz	0.39	0.65	1.64	0.1	2.77x10 <sup>-5</sup>
H3K27ac_Hnisz.extend.500	0.43	0.68	1.58	0.1	9.46x10 <sup>-5</sup>
H3K27ac_PGC2	0.27	0.53	1.95	0.1	2.57x10 <sup>-11</sup>
H3K27ac_PGC2.extend.500	0.34	0.61	1.80	0.1	8.44x10 <sup>-1</sup>
H3K4me1_peaks_Trynka	0.18	0.35	1.98	0.2	2.26x10 <sup>-5</sup>
H3K4me1_Trynka	0.43	0.71	1.64	0.1	5.61x10 <sup>-5</sup>
H3K4me1_Trynka.extend.500	0.61	0.86	1.41	0.1	6.55x10 <sup>-5</sup>
H3K4me3_peaks_Trynka	0.04	0.11	2.55	0.6	1.35x10 <sup>-3</sup>
H3K4me3_Trynka	0.14	0.35	2.59	0.2	5.48x10 <sup>-13</sup>
H3K4me3_Trynka.extend.500	0.26	0.49	1.88	0.1	4.52x10 <sup>-10</sup>
H3K9ac_peaks_Trynka	0.04	0.12	3.03	0.5	1.03x10 <sup>-5</sup>
H3K9ac_Trynka	0.13	0.36	2.79	0.2	1.07x10 <sup>-14</sup>
H3K9ac_Trynka.extend.500	0.23	0.50	2.15	0.2	1.18x10 <sup>-14</sup>
Intron_UCSC	0.39	0.40	1.00	0.1	9.63x10 <sup>-1</sup>
Intron_UCSC.extend.500	0.40	0.51	1.27	0.1	3.63x10 <sup>-5</sup>
PromoterFlanking_Hoffman	0.01	0.04	4.43	1.1	1.82x10 <sup>-2</sup>
PromoterFlanking_Hoffman.extend.500	0.03	0.13	3.70	0.4	3.39x10 <sup>-12</sup>
Promoter_UCSC	0.03	0.14	4.45	0.5	6.18x10 <sup>-12</sup>
Promoter_UCSC.extend.500	0.04	0.18	4.51	0.4	1.01x10 <sup>-10</sup>
Repressed_Hoffman	0.45	0.30	0.65	0.1	5.99x10 <sup>-10</sup>
Repressed_Hoffman.extend.500	0.71	0.47	0.65	0.1	8.43x10 <sup>-5</sup>
SuperEnhancer_Hnisz	0.17	0.35	2.03	0.2	1.79x10 <sup>-5</sup>
SuperEnhancer_Hnisz.extend.500	0.17	0.34	1.97	0.2	1.57x10 <sup>-5</sup>
TFBS_ENCODE	0.13	0.34	2.50	0.2	2.61x10 <sup>-11</sup>
TFBS_ENCODE.extend.500	0.35	0.49	1.42	0.1	6.98x10 <sup>-4</sup>
Transcribed_Hoffman	0.35	0.38	1.07	0.1	5.34x10 <sup>-1</sup>
Transcribed_Hoffman.extend.500	0.77	0.74	0.97	0.1	6.53x10 <sup>-1</sup>
TSS_Hoffman	0.02	0.13	7.10	0.9	5.15x10 <sup>-12</sup>
TSS_Hoffman.extend.500	0.04	0.19	5.23	0.5	6.73x10 <sup>-10</sup>
UTR_3_UCSC	0.01	0.08	6.88	0.9	4.11x10 <sup>-10</sup>
UTR_3_UCSC.extend.500	0.03	0.13	4.48	0.6	4.99x10 <sup>-10</sup>
UTR_5_UCSC	0.01	0.05	9.46	1.8	1.36x10 <sup>-5</sup>
UTR_5_UCSC.extend.500	0.03	0.15	5.10	0.5	1.19x10 <sup>-14</sup>
WeakEnhancer_Hoffman	0.02	0.04	1.92	0.6	1.02x10 <sup>-1</sup>
WeakEnhancer_Hoffman.extend.500	0.09	0.17	1.92	0.2	1.58x10 <sup>-4</sup>

Supplementary Table 5. LD score enrichments, heritability estimates, and p-values after modification of the LD score software

Category	Prop. of SNPs	Prop. of h <sup>2</sup>	Enrichment	SE	p-value
Coding_UCSC	0.03	0.20	6.64	0.6	3.44x10 <sup>-22</sup>
Coding_UCSC.extend.500	0.12	0.39	3.19	0.2	5.65x10 <sup>-42</sup>
Conserved_LindbladToh	0.03	0.16	4.82	0.5	3.25x10 <sup>-13</sup>
Conserved_LindbladToh.extend.500	0.40	0.63	1.59	0.1	3.09x10 <sup>-10</sup>
CTCF_Hoffman	0.03	0.07	2.18	0.4	1.54x10 <sup>-3</sup>
CTCF_Hoffman.extend.500	0.10	0.14	1.47	0.2	7.92x10 <sup>-3</sup>
DGF_ENCODE	0.18	0.40	2.19	0.2	3.73x10 <sup>-4</sup>
DGF_ENCODE.extend.500	0.64	0.83	1.29	0.1	1.23x10 <sup>-3</sup>
DHS_peaks_Trynka	0.14	0.31	2.32	0.2	2.94x10 <sup>-3</sup>
DHS_Trynka	0.20	0.40	1.99	0.2	7.84x10 <sup>-3</sup>
DHS_Trynka.extend.500	0.56	0.76	1.34	0.1	2.98x10 <sup>-3</sup>
Enhancer_Andersson	0.01	0.01	2.21	1.4	3.80x10 <sup>-1</sup>
Enhancer_Andersson.extend.500	0.03	0.04	1.48	0.5	2.89x10 <sup>-3</sup>
Enhancer_Hoffman	0.10	0.19	1.95	0.2	2.96x10 <sup>-3</sup>
Enhancer_Hoffman.extend.500	0.23	0.38	1.66	0.1	5.12x10 <sup>-3</sup>
FetalDHS_Trynka	0.11	0.31	2.92	0.2	4.30x10 <sup>-3</sup>
FetalDHS_Trynka.extend.500	0.34	0.53	1.58	0.1	1.68x10 <sup>-3</sup>
H3K27ac_Hnisz	0.54	0.75	1.39	0.1	7.15x10 <sup>-3</sup>
H3K27ac_Hnisz.extend.500	0.57	0.77	1.35	0.1	2.25x10 <sup>-3</sup>
H3K27ac_PGC2	0.37	0.62	1.67	0.1	7.02x10 <sup>-3</sup>
H3K27ac_PGC2.extend.500	0.46	0.71	1.54	0.1	6.89x10 <sup>-3</sup>
H3K4me1_peaks_Trynka	0.24	0.41	1.74	0.2	4.82x10 <sup>-3</sup>
H3K4me1_Trynka	0.55	0.80	1.44	0.1	9.82x10 <sup>-3</sup>
H3K4me1_Trynka.extend.500	0.74	0.92	1.25	0.1	3.15x10 <sup>-3</sup>
H3K4me3_peaks_Trynka	0.06	0.15	2.36	0.3	6.40x10 <sup>-3</sup>
H3K4me3_Trynka	0.20	0.46	2.29	0.2	4.79x10 <sup>-3</sup>
H3K4me3_Trynka.extend.500	0.35	0.60	1.71	0.1	2.44x10 <sup>-3</sup>
H3K9ac_peaks_Trynka	0.07	0.17	2.50	0.3	2.10x10 <sup>-3</sup>
H3K9ac_Trynka	0.21	0.48	2.30	0.2	5.70x10 <sup>-3</sup>
H3K9ac_Trynka.extend.500	0.36	0.64	1.77	0.1	4.00x10 <sup>-3</sup>
Intron_UCSC	0.47	0.43	0.91	0.1	2.60x10 <sup>-3</sup>
Intron_UCSC.extend.500	0.49	0.60	1.21	0.1	4.05x10 <sup>-3</sup>
PromoterFlanking_Hoffman	0.01	0.04	3.30	0.8	2.11x10 <sup>-3</sup>
PromoterFlanking_Hoffman.extend.500	0.05	0.15	2.91	0.3	3.11x10 <sup>-3</sup>
Promoter_UCSC	0.06	0.20	3.41	0.4	2.98x10 <sup>-3</sup>
Promoter_UCSC.extend.500	0.07	0.25	3.43	0.3	2.08x10 <sup>-3</sup>
Repressed_Hoffman	0.34	0.18	0.55	0.1	2.20x10 <sup>-3</sup>
Repressed_Hoffman.extend.500	0.56	0.32	0.57	0.1	1.10x10 <sup>-3</sup>
SuperEnhancer_Hnisz	0.27	0.43	1.63	0.2	1.92x10 <sup>-3</sup>
SuperEnhancer_Hnisz.extend.500	0.27	0.43	1.59	0.1	2.07x10 <sup>-3</sup>
TFBS_ENCODE	0.18	0.41	2.26	0.2	2.21x10 <sup>-3</sup>
TFBS_ENCODE.extend.500	0.43	0.60	1.38	0.1	7.13x10 <sup>-3</sup>
Transcribed_Hoffman	0.43	0.43	1.02	0.1	7.99x10 <sup>-3</sup>
Transcribed_Hoffman.extend.500	0.76	0.72	0.94	0.1	3.15x10 <sup>-3</sup>
TSS_Hoffman	0.03	0.19	5.34	0.8	9.97x10 <sup>-3</sup>
TSS_Hoffman.extend.500	0.06	0.26	3.97	0.4	1.84x10 <sup>-3</sup>
UTR_3_UCSC	0.02	0.10	5.08	0.6	2.30x10 <sup>-3</sup>
UTR_3_UCSC.extend.500	0.05	0.16	3.41	0.4	2.80x10 <sup>-3</sup>
UTR_5_UCSC	0.01	0.07	7.15	1.2	2.72x10 <sup>-3</sup>
UTR_5_UCSC.extend.500	0.05	0.20	3.99	0.4	4.37x10 <sup>-3</sup>
WeakEnhancer_Hoffman	0.03	0.05	1.67	0.4	7.95x10 <sup>-3</sup>
WeakEnhancer_Hoffman.extend.500	0.13	0.22	1.62	0.2	3.10x10 <sup>-3</sup>

Supplementary Table 6. Fifty-four eGenes in METSIM, including the 42 genes replicated for correlation with BMI and effect direction in TwinsUK

Gene	Chr <sup>a</sup>	Pearson			Linear regression				
		Effect size (r)	METSIM <sup>†</sup>		METSIM <sup>‡</sup>		TwinsUK <sup>§</sup>		
			p-value	Effect size (β)	SE	p-value	Effect size (β)	SE	p-value
ADH1B	4	-0.45	7.40x10 <sup>-18</sup>	-0.21	0.02	1.88x10 <sup>-20</sup>	-0.58	0.03	4.47x10 <sup>-71</sup>
ORMDL3 <sup>*</sup>	17	-0.45	8.57x10 <sup>-18</sup>	-0.16	0.02	2.06x10 <sup>-20</sup>	-0.58	0.03	2.65x10 <sup>-70</sup>
AKR1C3	10	0.33	4.78x10 <sup>-10</sup>	0.13	0.02	2.95x10 <sup>-11</sup>	0.49	0.03	5.19x10 <sup>-24</sup>
CMTM3	16	0.41	4.32x10 <sup>-15</sup>	0.087	0.01	3.84x10 <sup>-17</sup>	0.50	0.03	6.64x10 <sup>-52</sup>
LPIN1	2	-0.38	1.49x10 <sup>-13</sup>	-0.14	0.02	2.27x10 <sup>-15</sup>	-0.47	0.03	2.38x10 <sup>-44</sup>
RNF157	17	-0.29	5.19x10 <sup>-9</sup>	-0.096	0.02	5.87x10 <sup>-9</sup>	-0.47	0.03	8.86x10 <sup>-42</sup>
MYOF	10	0.32	1.07x10 <sup>-9</sup>	0.086	0.01	7.37x10 <sup>-11</sup>	0.46	0.03	2.59x10 <sup>-41</sup>
NAA40	11	0.28	1.81x10 <sup>-7</sup>	0.052	0.009	2.67x10 <sup>-8</sup>	0.46	0.03	4.00x10 <sup>-40</sup>
TMEM165	4	0.33	2.45x10 <sup>-9</sup>	0.045	0.007	1.84x10 <sup>-10</sup>	0.45	0.03	3.52x10 <sup>-37</sup>
RFFL	11	0.27	1.02x10 <sup>-9</sup>	0.035	0.006	1.84x10 <sup>-9</sup>	0.43	0.03	5.67x10 <sup>-37</sup>
TMCO6	5	-0.28	9.23x10 <sup>-9</sup>	-0.060	0.01	1.18x10 <sup>-9</sup>	-0.44	0.03	5.04x10 <sup>-36</sup>
SCRN2	17	-0.38	2.23x10 <sup>-13</sup>	-0.10	0.01	3.79x10 <sup>-15</sup>	-0.38	0.03	5.32x10 <sup>-35</sup>
CSGALNACT1	8	0.24	1.00x10 <sup>-5</sup>	0.047	0.01	2.04x10 <sup>-5</sup>	0.42	0.03	1.41x10 <sup>-31</sup>
TAPBP	6	0.25	6.71x10 <sup>-6</sup>	0.047	0.02	1.60x10 <sup>-5</sup>	0.32	0.03	1.52x10 <sup>-29</sup>
CLN8	8	0.32	4.50x10 <sup>-9</sup>	0.044	0.007	3.67x10 <sup>-10</sup>	0.36	0.03	4.41x10 <sup>-29</sup>
DRAM1	12	0.30	1.87x10 <sup>-9</sup>	0.050	0.008	1.80x10 <sup>-9</sup>	0.40	0.03	5.94x10 <sup>-29</sup>
WNT2B	1	0.25	2.44x10 <sup>-9</sup>	0.026	0.005	4.90x10 <sup>-7</sup>	0.38	0.03	1.41x10 <sup>-27</sup>
S100A1	1	-0.27	2.52x10 <sup>-7</sup>	-0.20	0.04	3.59x10 <sup>-11</sup>	-0.38	0.03	3.69x10 <sup>-25</sup>
RPS6KL1	14	0.26	2.54x10 <sup>-6</sup>	0.060	0.01	5.25x10 <sup>-7</sup>	0.34	0.03	3.27x10 <sup>-25</sup>
SLC16A7	12	-0.26	3.47x10 <sup>-6</sup>	-0.068	0.01	7.60x10 <sup>-7</sup>	-0.30	0.03	2.08x10 <sup>-23</sup>
ZNF592	15	-0.27	8.26x10 <sup>-7</sup>	-0.037	0.007	1.40x10 <sup>-7</sup>	-0.33	0.03	2.10x10 <sup>-23</sup>
MFSO1	3	0.31	8.31x10 <sup>-9</sup>	0.069	0.01	6.70x10 <sup>-10</sup>	0.35	0.04	2.82x10 <sup>-22</sup>
HYI	1	-0.31	6.45x10 <sup>-9</sup>	-0.11	0.02	5.52x10 <sup>-10</sup>	-0.29	0.03	5.95x10 <sup>-22</sup>
ANXA4	2	0.24	1.04x10 <sup>-5</sup>	0.045	0.009	2.52x10 <sup>-6</sup>	0.35	0.04	1.20x10 <sup>-21</sup>
RAB30	11	0.24	8.19x10 <sup>-6</sup>	0.040	0.008	1.98x10 <sup>-6</sup>	0.31	0.03	1.16x10 <sup>-21</sup>
PLD1	3	-0.28	2.28x10 <sup>-7</sup>	-0.050	0.009	3.24x10 <sup>-8</sup>	-0.32	0.03	7.95x10 <sup>-21</sup>
MYO5A	15	0.30	3.20x10 <sup>-9</sup>	0.049	0.008	3.41x10 <sup>-9</sup>	0.32	0.04	4.61x10 <sup>-19</sup>
ACADS <sup>*</sup>	12	-0.37	2.91x10 <sup>-12</sup>	-0.085	0.01	7.12x10 <sup>-14</sup>	-0.24	0.03	6.65x10 <sup>-19</sup>
SCAI	9	-0.28	1.81x10 <sup>-7</sup>	-0.034	0.006	2.50x10 <sup>-8</sup>	-0.27	0.03	1.42x10 <sup>-18</sup>
HLA-DRB1	6	0.25	3.53x10 <sup>-6</sup>	0.14	0.03	7.83x10 <sup>-7</sup>	0.31	0.03	2.09x10 <sup>-18</sup>
LACTB <sup>*</sup>	15	0.30	1.67x10 <sup>-8</sup>	0.069	0.01	1.40x10 <sup>-9</sup>	0.32	0.04	4.94x10 <sup>-18</sup>
GPHN	14	-0.43	7.51x10 <sup>-17</sup>	-0.11	0.01	3.20x10 <sup>-19</sup>	-0.29	0.03	4.28x10 <sup>-17</sup>
MPHOSPH8	13	-0.24	8.25x10 <sup>-6</sup>	-0.033	0.007	2.02x10 <sup>-6</sup>	-0.23	0.04	3.97x10 <sup>-11</sup>
MAP2K5 <sup>*</sup>	15	-0.25	7.83x10 <sup>-6</sup>	-0.039	0.008	1.90x10 <sup>-6</sup>	-0.21	0.03	3.81x10 <sup>-10</sup>
RRNAD1	1	-0.24	1.05x10 <sup>-5</sup>	-0.032	0.007	2.30x10 <sup>-6</sup>	-0.19	0.03	3.14x10 <sup>-9</sup>
CCDC50	3	-0.33	1.16x10 <sup>-9</sup>	-0.059	0.009	7.24x10 <sup>-11</sup>	-0.18	0.03	9.93x10 <sup>-9</sup>
RAD54L2	3	-0.25	2.32x10 <sup>-6</sup>	-0.030	0.006	4.70x10 <sup>-7</sup>	-0.20	0.04	2.79x10 <sup>-9</sup>
SCMH1	1	-0.32	1.11x10 <sup>-9</sup>	-0.047	0.007	7.55x10 <sup>-11</sup>	-0.19	0.03	3.85x10 <sup>-9</sup>
ATP7B	13	-0.26	6.30x10 <sup>-7</sup>	-0.040	0.007	1.11x10 <sup>-7</sup>	-0.20	0.04	7.22x10 <sup>-9</sup>
CYP7B1	8	0.24	6.90x10 <sup>-6</sup>	0.047	0.01	1.64x10 <sup>-6</sup>	0.19	0.03	1.07x10 <sup>-7</sup>
RERE	1	-0.24	1.02x10 <sup>-5</sup>	-0.031	0.006	2.61x10 <sup>-6</sup>	-0.17	0.04	5.39x10 <sup>-6</sup>
RPAP1	15	-0.35	1.86x10 <sup>-10</sup>	-0.042	0.006	8.82x10 <sup>-12</sup>	-0.14	0.03	9.58x10 <sup>-6</sup>
ARHGEF7	13	-0.36	4.12x10 <sup>-11</sup>	-0.050	0.007	1.60x10 <sup>-12</sup>	0.022	0.04	NS <sup>§</sup>
NCKIPSD	3	-0.34	3.28x10 <sup>-10</sup>	-0.067	0.01	1.51x10 <sup>-11</sup>	-0.042	0.03	NS <sup>§</sup>
NDUFS2	1	-0.24	9.38x10 <sup>-6</sup>	-0.029	0.006	2.17x10 <sup>-6</sup>	-0.048	0.03	NS <sup>§</sup>
REEP1	2	-0.24	6.38x10 <sup>-6</sup>	-0.033	0.007	1.45x10 <sup>-6</sup>	0.022	0.04	NS <sup>§</sup>
RGCC	13	-0.25	2.81x10 <sup>-6</sup>	-0.076	0.01	4.91x10 <sup>-7</sup>	0.087	0.03	NS <sup>§</sup>
SETD6	16	-0.27	3.99x10 <sup>-7</sup>	-0.041	0.007	6.30x10 <sup>-8</sup>	-0.047	0.04	NS <sup>§</sup>
SLC35A3	12	-0.26	1.13x10 <sup>-6</sup>	-0.024	0.004	2.15x10 <sup>-7</sup>	0.043	0.03	NS <sup>§</sup>
SPAG7	17	-0.26	1.21x10 <sup>-6</sup>	-0.035	0.007	2.27x10 <sup>-7</sup>	-0.076	0.03	NS <sup>§</sup>
NUDCD3	7	-0.34	7.00x10 <sup>-10</sup>	-0.032	0.005	4.17x10 <sup>-11</sup>	NA <sup>‡</sup>	NA <sup>‡</sup>	NA <sup>‡</sup>
RP11-387H17.4	17	-0.40	4.40x10 <sup>-14</sup>	-0.26	0.03	4.74x10 <sup>-16</sup>	NA <sup>‡</sup>	NA <sup>‡</sup>	NA <sup>‡</sup>
RSBN1L-AS1	7	-0.36	1.65x10 <sup>-11</sup>	-0.056	0.007	5.73x10 <sup>-13</sup>	NA <sup>‡</sup>	NA <sup>‡</sup>	NA <sup>‡</sup>
TUBB2B	6	0.34	1.01x10 <sup>-10</sup>	0.14	0.02	4.51x10 <sup>-12</sup>	NA <sup>‡</sup>	NA <sup>‡</sup>	NA <sup>‡</sup>

<sup>a</sup>GWAS gene.

<sup>†</sup>Effect size (r, Pearson rho) and p-value calculated from Pearson correlation between gene expression and BMI (see Methods).

<sup>‡</sup>Effect size, standard error (SE), and p-value calculated using a linear regression model with BMI and age, age<sup>2</sup> and the 14 technical factors as covariates when compared to a null model without BMI. These models were compared using an F-test (see Methods).

<sup>§</sup>Effect size, standard error (SE), and p-value calculated from linear mixed effects model. A full model including BMI was compared to a null model in which the same model was fitted, but with the phenotype (BMI) omitted. These models were compared using an F-test (see Methods).

<sup>||</sup>Adjusted p-value > 9.26x10<sup>-4</sup>.

<sup>¶</sup>Value not applicable due to inability to test for replication in TwinsUK cohort.

<sup>a</sup>Chr indicates chromosome.

Supplementary Table 7. The 42 replicated BMI-correlated eGenes show significant enrichment for metabolic and inflammatory pathways using KEGG pathway analysis as implemented in WebGestalt<sup>13</sup>

KEGG Pathway Name	Ratio of Enrichment	Number of Genes	Genes in Pathway	p-value	Adjusted p-value*
Fatty acid metabolism	18.76	2	<i>ACADS</i> <i>ADH1B</i>	0.0051	0.010
Metabolism of xenobiotics by cytochrome P450	21.78	2	<i>AKR1C3</i> <i>ADH1B</i>	0.0038	0.010
Steroid hormone biosynthesis	30.69	2	<i>AKR1C3</i> <i>CYP7B1</i>	0.0019	0.010
Antigen processing and presentation	11.85	2	<i>HLA-DRB1</i> <i>TAPBP</i>	0.012	0.019

\*p-value adjusted using Benjamini-Hochberg correction for multiple testing.

Supplementary Table 8. DeepSEA analysis of the variants in the MAP2K5 locus supports the functionality of the looping *cis*-eQTL SNP rs4776984.

SNP ID	Chr	Position	Ref	Alt	DeepSEA score
rs4776984	chr15	68118194	A	C	$2.36 \times 10^{-3}$
rs4776982	chr15	68114974	A	G	$3.90 \times 10^{-2}$
rs4492996	chr15	68113240	A	G	$7.16 \times 10^{-2}$
rs4776990	chr15	68137364	C	T	$1.09 \times 10^{-1}$
rs28742003	chr15	68127769	C	T	$1.30 \times 10^{-1}$
rs28427879	chr15	68124256	G	T	$1.98 \times 10^{-1}$

**Supplementary Table 9. Significant CHICAGO interaction and replication scores from a separate HWA Capture Hi-C experiment verify the looping *cis*-eQTLs for the four identified obesity-related loci.**

Other End	Baited Fragment	Target Gene	Looping <i>cis</i> -eQTL	CHICAGO score	Replication score
chr15,67834655,67840760	chr15,68111739,68138337	MAP2K5	rs4476984	5.05	6.15
chr17,38082534,38106859	chr17,38074576,38081958	ORMDL3	rs8076131	6.35	6.73
chr15,63413071,63415370	chr15,63561331,63570763	LACTB	rs3784671	6.65	13.92
chr12,121158545,121162946	chr12,121343847,121345146	ACADS	rs10774569	5.29	6.62

**Supplementary Table 10. DNA oligonucleotides used for electrophoretic mobility shift assay.**

DNA oligonucleotides	Sequence (5' -> 3') for positive and negative strand
Reference allele – A (positive) biotinylated probe*	GCGCGCCCAACTCGGAGCGCCCTGCTGGGCG
Reference allele – A (negative) biotinylated probe	CGCCCAGCAGGGCGCTCCGAGTTGGGCGCGC
Alternate allele – C (positive) biotinylated probe*	GCGCGCCCAACTCGGCGCGCCCTGCTGGGCG
Alternate allele – C (negative) biotinylated probe	CGCCCAGCAGGGCGCGCCGAGTTGGGCGCGC

\*Biotinylated probes were created by adding biotin to the 5' end of positive strand probes.

## Chapter 3

Reverse gene–environment interaction approach to identify variants influencing body-mass index in humans



# Reverse gene–environment interaction approach to identify variants influencing body-mass index in humans

Kristina M. Garske<sup>1</sup>, David Z. Pan<sup>1,2</sup>, Zong Miao<sup>1,2</sup>, Yash V. Bhagat<sup>3</sup>, Caroline Comenho<sup>1</sup>, Christopher R. Robles<sup>3</sup>, Jihane N. Benhammou<sup>1,4</sup>, Marcus Alvarez<sup>1</sup>, Arthur Ko<sup>5</sup>, Chun Jimmie Ye<sup>6</sup>, Joseph R. Pisegna<sup>1,4</sup>, Karen L. Mohlke<sup>7</sup>, Janet S. Sinsheimer<sup>1,8</sup>, Markku Laakso<sup>9</sup> and Päivi Pajukanta<sup>1,2,9\*</sup>

**Identifying gene–environment (G×E) interactions contributing to human cardiometabolic disorders is challenging. Here we apply a reverse G×E candidate search by deriving candidate variants from promoter–enhancer interactions that respond to dietary fatty acid challenge through altered chromatin accessibility in primary human adipocytes. We then test all variants residing in lipid-responsive open chromatin sites in adipocyte promoter–enhancer contacts for interaction effects between genotype and dietary saturated fat intake on body-mass index (BMI) in the UK Biobank. We discover 14 new G×E variants in 12 lipid-responsive promoters, including in well-known lipid-related genes (*LIPE*, *CARM1* and *PLIN2*) and newly associated genes, such as *LDB3*, for which we provide further functional and integrative genomic evidence. We further identify 24 G×E variants in enhancers, for a total of 38 new G×E variants for BMI in the UK Biobank, demonstrating that molecular genomics data produced in physiologically relevant contexts can be applied to discover new functional G×E mechanisms in humans.**

Cardiometabolic disorders develop as a result of genetic predisposition, environmental factors and their interactions<sup>1,2</sup>. Genome-wide association studies (GWAS) have detected additive genetic effects for these traits, but the biological mechanisms explaining how genetic variation is involved in the increasing prevalence of obesogenic cardiometabolic disorders have yet to be identified. Some examples of G×E interactions are emerging, including, for instance, the highly replicated BMI risk variant rs9939609 in an intron of *FTO* that exhibits a significant interaction with physical activity for effect on BMI<sup>3</sup>. However, overall, there are few replicated G×E signals for cardiometabolic disorders in humans<sup>4</sup>. It has remained challenging to identify these signals, owing to small cohort sizes and poorly standardized definitions for human environmental phenotypes. Even with large cohorts such as the UK Biobank<sup>5</sup>, the statistical power to detect G×E interactions by using a genome-wide agnostic search is limited owing to the small effect sizes of G×E interactions and heavy multiple-testing penalties. Furthermore, once G×E signals have been detected, the mechanisms underlying the associations remain unclear, warranting further fine-mapping studies.

To systematically identify genes involved in G×E interactions, we set out to quantify molecular genomic responses to saturated and monounsaturated fatty acid challenge in primary human adipocytes, as a cellular model of dietary fat intake in this key adipose tissue cell type. We measured differences in chromatin accessibility

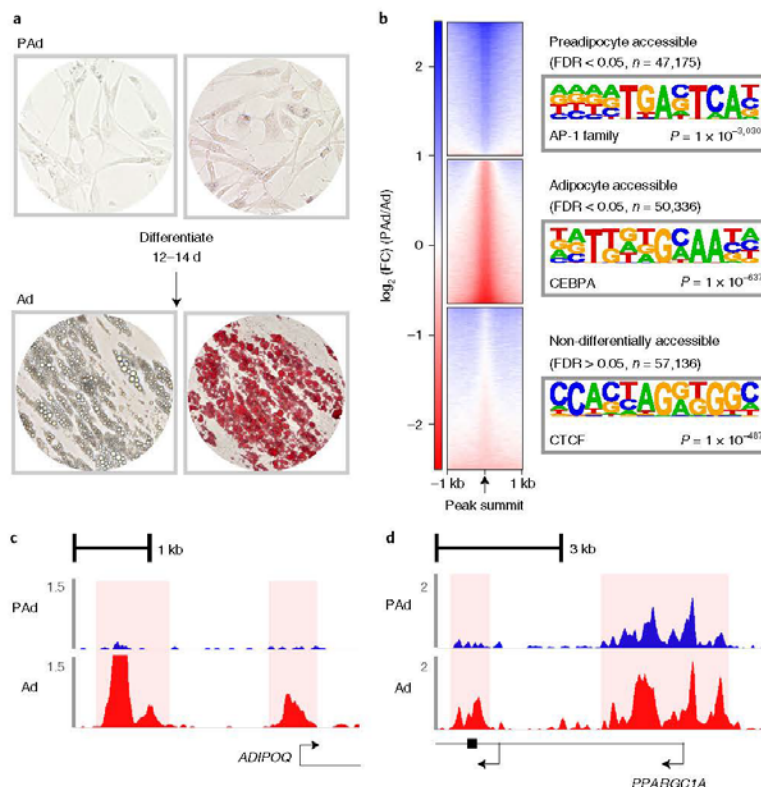
and searched the whole genome for chromosomal interactions between lipid-responsive gene promoters and enhancers to shed new light on the genomic molecular mechanisms relevant for lipid responses in human adipocytes. We hypothesized that these genomic responses would provide targeted regions harbouring candidate genetic variants for analysis of G×E interactions in the large UK Biobank cohort<sup>5</sup>. Using these targeted regions should restrict the multiple-testing burden hampering the typical agnostic genome-wide G×E analysis and expand knowledge of the true environmental exposures responsible for G×E signals, thereby revealing the underlying functional mechanisms. Thus, integrating context-specific molecular genomics with environmental phenotypes and clinical outcome data in the UK Biobank should help elucidate molecular mechanisms occurring in response to obesogenic cellular context that contribute to cardiometabolic traits in humans.

## Results

**Adipocyte accessible chromatin identifies regulatory regions.** To obtain primary human adipocytes for study of the effects of lipids on chromatin dynamics, we first differentiated primary human white preadipocytes into adipocytes in vitro (Fig. 1a). We performed assay for transposase-accessible chromatin using sequencing (ATAC-seq)<sup>6</sup> on three biological replicates of the preadipocytes and adipocytes to identify open chromatin regions that were differentially accessible in the two cell types (Fig. 1b and Supplementary Tables 1 and 2).

<sup>1</sup>Department of Human Genetics, David Geffen School of Medicine at UCLA, Los Angeles, CA, USA. <sup>2</sup>Bioinformatics Interdepartmental Program, UCLA, Los Angeles, CA, USA. <sup>3</sup>Department of Computer Science, UCLA, Los Angeles, CA, USA. <sup>4</sup>Vache and Tamar Manoukian Division of Digestive Diseases, UCLA, Los Angeles, CA, USA. <sup>5</sup>Department of Medicine, David Geffen School of Medicine at UCLA, Los Angeles, CA, USA. <sup>6</sup>Institute for Human Genetics, Department of Epidemiology and Biostatistics and Department of Bioengineering and Therapeutic Sciences, UCSF, San Francisco, CA, USA. <sup>7</sup>Department of Genetics, University of North Carolina, Chapel Hill, NC, USA. <sup>8</sup>Department of Biomathematics, David Geffen School of Medicine at UCLA, Los Angeles, CA, USA. <sup>9</sup>Internal Medicine, Institute of Clinical Medicine, University of Eastern Finland and Kuopio University Hospital, Kuopio, Finland.

\*e-mail: ppajukanta@mednet.ucla.edu

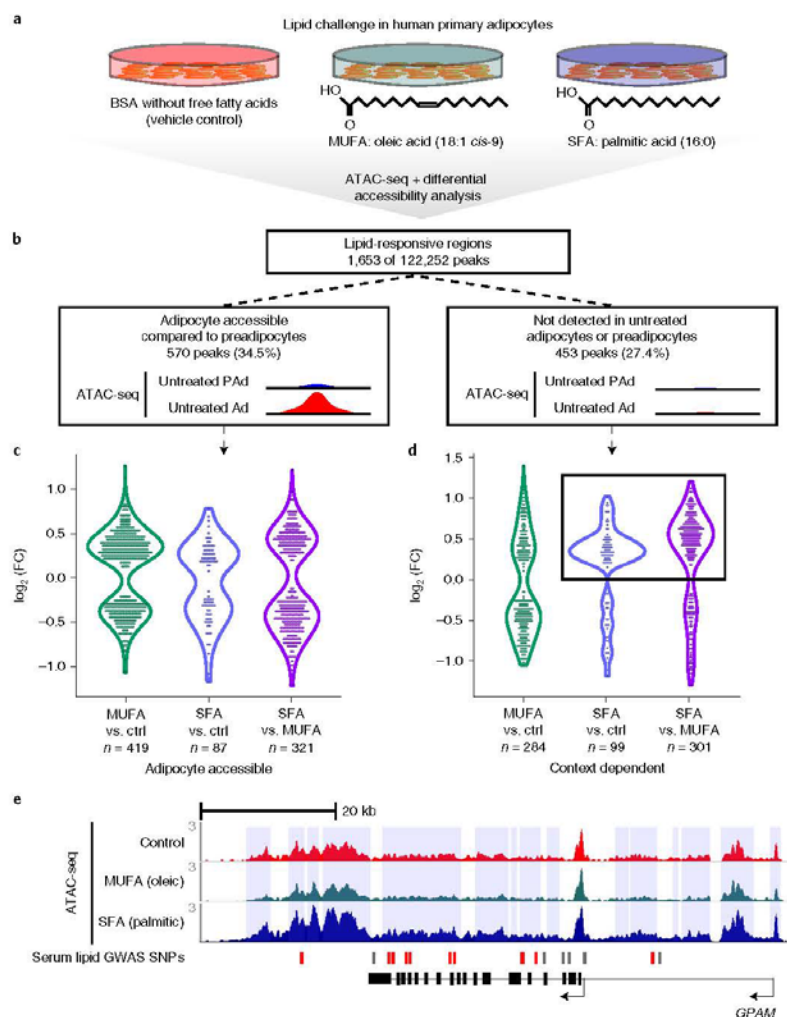


**Fig. 1 | ATAC-seq analysis comparing primary human preadipocytes and adipocytes indicates successful adipocyte differentiation and widespread changes in chromatin accessibility.** **a**, Bright-field images of preadipocytes (top; PAd) and in vitro-differentiated adipocytes (bottom; Ad) for unstained cells (left) and cells stained with Oil Red O (right). Images are representative examples from two independent experiments. **b**, Heat maps showing  $\log_2$ -transformed fold change ( $\log_2$ (FC)) in bins per million mapped reads (BPM) for preadipocytes as compared to adipocytes in the three indicated peak sets. FDR was calculated (adjusting for  $n=154,647$  ATAC-seq peaks) from the  $P$  values of the quasi-likelihood (QL)  $F$  test (see Methods) for differential accessibility between preadipocytes and adipocytes using ATAC-seq libraries from  $n=3$  replicates per cell type. The most enriched TF motif for the indicated peak set is listed to the right. Enrichment  $P$  values were derived from the hypergeometric enrichment test of the proportion of the given top de novo-identified<sup>6</sup> TF motifs in the three indicated peak sets as compared with the background set (see Methods). **c,d**, Read coverage (BPM) in one representative ( $n=3$  replicates per cell type) preadipocyte (blue) and adipocyte (red) ATAC-seq library at the adipocyte accessible ATAC peaks in the promoters of the adipocyte hormone gene adiponectin (*ADIPOQ*) (**c**) and the adipocyte-specific marker peroxisome proliferator-activated receptor gamma coactivator 1-alpha (*PPARGC1A*) (**d**).

The 50,336 ATAC-seq peaks that were more accessible in primary human adipocytes ('adipocyte accessible') included the promoters of the *ADIPOQ* and *PPARGC1A* genes with known adipocyte-specific expression (Fig. 1c,d), providing evidence that we successfully differentiated adipocytes in vitro. To explore whether the adipocyte accessible peaks harboured transcription factor (TF) motifs relevant for adipocyte biology, we performed TF motif enrichment analyses with HOMER<sup>7</sup>. We found that the most enriched motif corresponded to the motif for the CCAAT-enhancer-binding protein alpha (C/EBP $\alpha$ ) TF (Fig. 1b), an important regulator of the later stages of adipogenesis<sup>8</sup>. We then classified the adipocyte accessible peaks into functional genomic annotations<sup>9</sup> and observed that the adipocyte accessible peaks fell more often in adipocyte enhancers and less frequently in quiescent regions when compared with the full peak set or preadipocyte accessible peaks (Supplementary Fig. 1). Taken together, these results provide evidence that in vitro

differentiation of adipocytes leads to an increase in chromatin accessibility in regions important for genomic regulation in adipocytes.

**Genomic responses to dietary lipids in human adipocytes.** We next searched for genomic regions harbouring regulatory elements that mediate adipocyte responses to the intake of different dietary lipids, by treating the adipocytes with the saturated fatty acid (SFA) palmitic acid (C16:0) or the monounsaturated fatty acid (MUFA) oleic acid (C18:1 *cis*-9) and then performing ATAC-seq on three biological replicates per condition (Fig. 2a). We found that treatment with either of these fatty acids resulted in increased staining with Oil Red O, which incorporates into neutral lipids, indicating that the lipid challenge resulted in increased storage of fatty acids in the lipid droplets of the cells (Supplementary Fig. 2). We identified 1,653 ATAC-seq peaks that were differentially accessible in the lipid-challenged primary human adipocytes in



**Fig. 2 | Lipid-responsive regions fall within adipocyte accessible regions of the genome, as well as within context-dependent regions that are not present in untreated adipocytes.** **a**, Schematic overview of the lipid challenge experiment in human primary adipocytes. Treatments were performed in three replicates per condition. **b**, Schematic overview indicating the two categories of lipid-responsive peaks used for all downstream analyses. Peaks were considered differentially accessible at a cutoff of  $\text{FDR} < 0.05$ . FDR was calculated (adjusting for  $n=122,252$  ATAC-seq peaks) from the  $P$  values of the QL  $F$  test (see Methods) in one-way analysis of variance (ANOVA). Significant lipid-responsive peaks categorized as adipocyte accessible ( $n=570$ ) or context dependent (not identified in untreated preadipocytes or adipocytes;  $n=453$ ) were used in all downstream analyses. **c, d**, Violin plots showing the  $\log_2$ -transformed fold change in differentially accessible peaks in the indicated comparisons, stratified by whether the peak was adipocyte accessible (**c**) or context dependent (**d**). In **c**, the violin plot characteristics are as follows: MUFA versus control (ctrl) ( $n=419$ ): range,  $-1.07$  to  $1.25$ ; median,  $0.26$ ; 25th percentile,  $-0.34$ ; 75th percentile,  $0.42$ ; SFA versus control ( $n=87$ ): range,  $-1.17$  to  $0.77$ ; median,  $0.15$ ; 25th percentile,  $-0.34$ ; 75th percentile,  $0.27$ ; SFA versus MUFA ( $n=321$ ): range,  $-1.22$  to  $1.21$ ; median,  $-0.26$ ; 25th percentile,  $-0.48$ ; 75th percentile,  $0.43$ . In **d**, the violin plot characteristics are as follows: MUFA versus control ( $n=284$ ): range,  $-1.05$  to  $1.40$ ; median,  $-0.33$ ; 25th percentile,  $-0.53$ ; 75th percentile,  $0.37$ ; SFA versus control ( $n=99$ ): range,  $-1.19$  to  $1.02$ ; median,  $0.34$ ; 25th percentile,  $0.20$ ; 75th percentile,  $0.47$ ; SFA versus MUFA ( $n=301$ ): range,  $-1.30$  to  $1.21$ ; median,  $0.49$ ; 25th percentile,  $-0.19$ ; 75th percentile,  $0.65$ . The box in **d** indicates a shift towards increased accessibility in SFA-treated cells, observed especially in the context-dependent peaks. **e**, Human genome browser snapshot (WashU) of the GPAM locus, which harbours 15 lipid-responsive peaks in a  $\sim 50$ -kb region ( $n=99$ ): range,  $-1.19$  to  $1.02$ ; median,  $0.34$ ; 25th percentile,  $0.20$ ; 75th percentile,  $0.47$ ; SFA versus MUFA ( $n=301$ ): range,  $-1.30$  to  $1.21$ ; median,  $0.49$ ; 25th percentile,  $-0.19$ ; 75th percentile,  $0.65$ ). Fourteen of these peaks are SFA responsive. Read coverage (BPM) is shown from one representative ATAC-seq library ( $n=3$  replicates per condition) for control (red), MUFA (green) and SFA (blue) treatment. GWAS SNPs for serum lipid traits are categorized as being in a lipid-responsive peak (red) or outside of a lipid-responsive peak (grey).

comparison to control adipocytes, referred to as lipid-responsive peaks (Fig. 2b, Supplementary Fig. 3 and Supplementary Tables 3 and 4). We cross-referenced these lipid-responsive peaks against the adipocyte accessible peaks (Fig. 1) and found that the 570 adipocyte accessible, lipid-responsive peaks fell mostly into enhancer and promoter annotations in adipocytes<sup>6</sup>, in line with their likely importance in environmental responses and regulation of gene expression in adipocytes (Fig. 2b and Supplementary Fig. 4).

Notably, we found that 453 of the 1,653 lipid-responsive peaks were not detected as open chromatin in the initial ATAC-seq data created in untreated adipocytes and preadipocytes (Fig. 2b). When compared to the adipocyte accessible open chromatin, these context-dependent open chromatin regions fell into a higher percentage of quiescent annotations imputed from data created in unchallenged adipocytes<sup>6</sup> (Supplementary Fig. 4). This indicates that genomic regions that are not open and accessible for TF binding in untreated, steady-state adipocytes or preadipocytes become activated in adipocytes under lipid challenge conditions.

When we stratified the SFA and MUFA treatment responses by adipocyte accessible or context-dependent open chromatin regions, we found that the effects of SFA treatment in context-dependent open chromatin regions were shifted towards increased accessibility (Fig. 2c,d). This was in contrast to the MUFA responses, which were generally evenly distributed between peaks with decreased and increased accessibility (Fig. 2c,d). This suggests that different fatty acids can result in distinct signalling effects on genome-level responses to lipid intake in adipocytes, and, in particular, SFA intake seems to activate regions of the genome that are normally inactive in untreated adipocytes.

On closer examination, we found 14 context-dependent peaks, exhibiting increased accessibility in SFA-treated adipocytes, which fell into a ~50-kb region on chromosome 10 (Fig. 2e). The locus contained a total of 15 lipid-responsive peaks, which spanned the entirety of the gene encoding glycerol-3-phosphate acyltransferase, mitochondrial (GPAM) (Fig. 2e). The GPAM enzyme prefers saturated fatty acid substrates, and the GPAM locus has been associated with serum lipid traits<sup>9–11</sup> (triglycerides (TGs), high-density lipoprotein cholesterol (HDL), low-density lipoprotein cholesterol (LDL) and total cholesterol (TC)) and serum alanine aminotransferase (ALT), which is a biomarker for liver health<sup>12</sup>, in previous GWAS<sup>13</sup>. The earlier GWAS associations at this locus, in combination with the strong genomic response to SFA treatment in adipocytes observed here, suggest that dysregulation of the important lipogenic pathway mediated by GPAM in adipocytes could contribute to obesogenic cardiometabolic disorders such as dyslipidaemias and non-alcoholic fatty liver disease (NAFLD).

#### Lipid-responsive gene promoters in chromosomal interactions.

To identify genes under transcriptional regulation via chromosomal interactions, we performed promoter capture Hi-C (pChI-C)<sup>14</sup> on the lipid-challenged human adipocytes with two biological replicates per condition (Supplementary Table 5). We identified 264 lipid-responsive ATAC-seq peaks that fell within adipocyte chromosomal interactions. To test whether these interacting, lipid-responsive regions of the genome harbour motifs for TFs that are important for lipid metabolism, we performed TF motif enrichment analysis<sup>6</sup> comparing the lipid-responsive peaks to non-lipid-responsive peaks within the chromosomal interactions. We found that motifs for peroxisome proliferator-activated receptor gamma (PPAR $\gamma$ ), an important TF in adipogenesis and lipid uptake, and its cofactor retinoid X receptor (RXR) were among the ten most enriched motifs (Fig. 3a and Supplementary Table 6). This indicates that the lipid-responsive sites in adipocyte promoter–enhancer contacts represent genomic regions that are important for mediating cellular responses to lipid uptake.

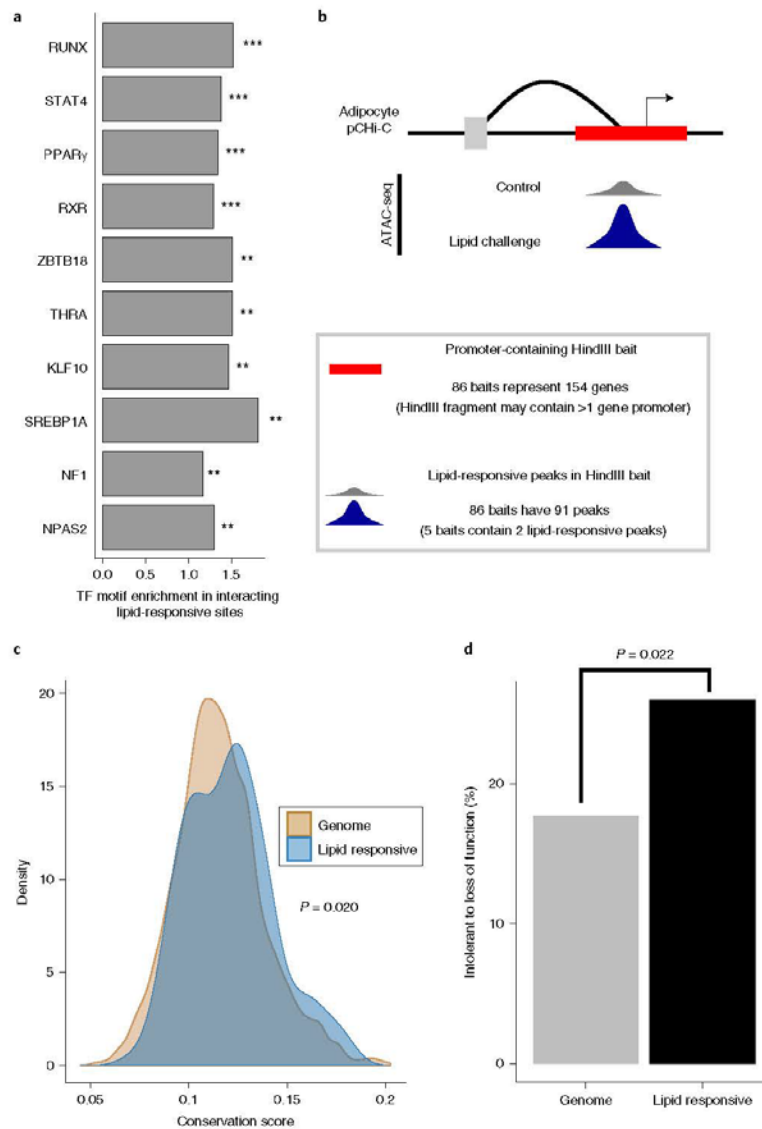
To identify the target genes of the adipocyte responses to lipid challenge, we first focused on the interacting promoters from adipocyte pChI-C (Fig. 3b), as promoters are more highly enriched for single-nucleotide polymorphisms (SNPs) that contribute to the heritability of local gene expression than enhancers<sup>15,16</sup>. The 86 interacting pChI-C baits represented 154 gene promoters, given that the resolution of pChI-C interaction data depends on the frequency of restriction sites in the genome (Fig. 3b and Supplementary Table 7). We performed a Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway enrichment analysis<sup>17</sup> on the set of 154 interacting, lipid-responsive target genes, which identified two significantly enriched pathways for amino acid metabolism (false-discovery rate (FDR) < 0.05) (Supplementary Table 8).

As energy homeostasis is important for survival, we hypothesized that the 154 gene regions responsible for mediating the effects of lipid uptake in adipocytes might exhibit differences in the level of conservation when compared to other genes in the genome. We therefore obtained an average conservation score for the 114 protein-coding genes among the 154 genes (gene body  $\pm$  500 kb) by using PhastCons<sup>18</sup> and found that the lipid-responsive protein-coding genes had higher conservation scores across placental mammals than all other protein-coding genes in the human genome ( $P=0.020$ ) (Fig. 3c).

We further investigated whether these lipid-responsive genes exhibited constraints on genetic mutation, by using the probability of each gene being intolerant to loss-of-function mutation (pLI), defined in Lek et al.<sup>19</sup> as a high unlikelihood of protein-truncating mutations in humans. We found that, of the genes for which pLI scores were available ( $n=104$ )<sup>19</sup>, 27 genes (26.0%) were considered LoF intolerant. Given that 17.7% of all genes are considered LoF intolerant, the pLI for lipid-responsive genes is higher than expected by chance alone ( $P=0.022$ ) (Fig. 3d). Taking these findings together, we identified 154 genes with lipid-responsive promoters in chromosomal interactions that are less tolerant of LoF variants and reside within genomic regions that are more conserved than expected by chance alone.

**Genes that interact with lipid-responsive enhancers.** We next tested whether the genes that interacted with lipid-responsive enhancers exhibited similar characteristics to those of the genes that had lipid-responsive promoters. We first found that 169 lipid-responsive enhancers interacted with 223 promoter-containing HindIII baits in the adipocyte pChI-C analysis (Supplementary Fig. 5). Given that multiple gene promoters can be captured within a single HindIII fragment, these 223 baits represented 323 gene promoters (Supplementary Fig. 5 and Supplementary Table 9). When we tested whether these 323 genes were enriched in any KEGG pathways, we did not find any functional pathways passing multiple-testing correction. This may be due to the fact that, on average, each lipid-responsive enhancer interacted with approximately two promoters (Supplementary Fig. 5), thus leading to ambiguities regarding which gene might be the true target of the lipid signalling response.

We further determined whether the genes that interacted with lipid-responsive enhancers exhibited mutational constraints by determining whether the *cis* regions (gene body  $\pm$  500 kb) of the protein-coding genes in this set ( $n=217$ ) had higher average conservation scores than all other protein-coding genes in the genome. In contrast to what we observed for genes with lipid-responsive promoters (Fig. 3c), we did not observe a significant difference in the conservation scores for genes that interacted with lipid-responsive enhancers (Supplementary Fig. 5). Interestingly, of the enhancer-interacting genes that had a pLI score ( $n=207$ )<sup>19</sup>, 50 (24.2%) were LoF intolerant, which is significantly higher than would be expected by chance alone ( $P=0.014$ ) (Supplementary Fig. 5). Taken together, these results are consistent with more moderate functional signifi-



**Fig. 3 | The 154 genes with lipid-responsive promoters within chromosomal interactions exhibit cross-species conservation and constraints on loss-of-function mutations, in line with their potential importance for energy homeostasis and survival. a**, The top ten most enriched TF motifs in lipid-responsive open chromatin regions in adipocyte chromosomal interactions ( $n=264$ ) include motifs for key TFs in lipid metabolism, such as the cofactors PPAR $\gamma$  and RXR. \*\*FDR < 0.001, \*\*\*FDR < 0.0001. Enrichment  $P$  values were derived from a hypergeometric enrichment test of the proportion of the given TF motif in the peak set as compared with the proportion in the background set of peaks (all non-lipid-responsive peaks in adipocyte chromosomal interactions), adjusted by the Benjamini–Hochberg method for the number of known motifs tested ( $n=364$ )<sup>6</sup>. **b**, Schematic of the data we integrated to identify 154 genes with lipid-responsive promoters in adipocyte chromosomal interactions. **c**, Density plot showing the distribution of per-gene average conservation scores across placental mammals<sup>18</sup> for all protein-coding genes in the genome ( $n=19,316$ ) as compared with all protein-coding genes in the set of 154 genes with lipid-responsive promoters ( $n=114$ ). The  $P$  value was obtained from a two-sided Wilcoxon signed-rank test. **d**, Bar graph showing the proportion of protein-coding genes that are intolerant to loss-of-function mutation (unlikely to have protein-truncating variants in humans)<sup>19</sup> in the whole genome ( $n=3,204/18,122$ ; 17.7%) as compared to the protein-coding genes among the 154 genes ( $n=27/104$ ; 26.0%). The  $P$  value was obtained from a hypergeometric enrichment test.

**Table 1** | Five lipid-responsive ATAC-seq peaks in interacting promoters overlap with GWAS SNPs for serum lipid traits

Peak chr.	Peak start	Peak end	Gene <sup>a</sup>	SNP(s) in peak	MAF <sup>b</sup>	Associated trait <sup>c</sup> (from ref. <sup>11</sup> )	P value <sup>c</sup> (from ref. <sup>11</sup> )	Index SNP (from ref. <sup>11</sup> )	LD with index SNP <sup>d</sup> (r <sup>2</sup> )
11	61,594,652	61,596,828	<i>FADS2-FADS1</i>	rs99780	0.37	TG/HDL/LDL/TC	$2.32 \times 10^{-16}/5.52 \times 10^{-9}$ <b><math>2.39 \times 10^{-21}/8.62 \times 10^{-18}</math></b>	rs174546	0.93
				rs968567	0.16		$3.4 \times 10^{-9}/NS/8.91 \times 10^{-11}$ $/2.27 \times 10^{-9}$		0.35
				rs191508698	0		$1 \times 10^{-6}/2.37 \times 10^{-8}/$ $1.24 \times 10^{-20}/31 \times 10^{-17}$		-
7	73,036,880	73,038,991	<i>MLXIPL</i>	rs55747707	0.2	TG	$3.55 \times 10^{-44}$	rs17145738	0.47
				rs34060476	0.13		$9.73 \times 10^{-46}$		0.77
2	27,432,323	27,432,971	<i>SLCSA6-ATRAID</i>	rs2580759	0.22	TG	$2.18 \times 10^{-17}$	rs1260326	-
				rs1275530	0.72		$1.88 \times 10^{-17}$		-
16	68,115,758	68,116,375	<i>NFATC3</i>	rs2107269	0.015	HDL	$5.07 \times 10^{-45}$	rs16942887	0.66
19	10,981,139	10,983,631	<i>CARM1</i>	rs12460421	0.44	LDL	$4.32 \times 10^{-41}$	rs6511720	-

Lipid-responsive ATAC-seq peaks that fell within promoters in adipocyte chromosomal interactions ( $n=91$ ) were assessed for whether they contained GWAS SNPs for serum lipid traits from the meta-GWAS performed in Willer et al.<sup>11</sup> Chr., chromosome; NS, not significant. <sup>a</sup>The gene listed corresponds to the promoter in the baited HindIII fragment with a lipid-responsive ATAC-seq peak. <sup>b</sup>MAF is the European frequency from the 1000 Genomes Project. <sup>c</sup>The most significant association is in bold when a SNP is associated with more than one serum lipid trait. <sup>d</sup>LD was calculated on the basis of Europeans in the 1000 Genomes Project; LD calculations > 0.2 are reported.

cance of the genes that interact with lipid-responsive enhancers when compared to genes with lipid-responsive ATAC-seq peaks in their promoters.

**Lipid responses contribute to heritability of serum lipid traits.** We hypothesized that the genes we were able to identify through lipid-responsive promoter-enhancer interactions might highlight important genomic regions that contribute to the heritability of cardiometabolic traits. We found that five lipid-responsive gene promoters and three lipid-responsive enhancers within adipocyte chromosomal interactions contained SNPs with genome-wide-significant ( $P < 5 \times 10^{-8}$ ) associations with serum lipid traits, identified in a meta-GWAS of ~180,000 individuals<sup>11</sup> (Table 1 and Supplementary Table 10). One of the lipid-responsive GWAS loci was the well-known nutritional response locus containing the fatty acid desaturase (*FADS1-FADS2-FADS3*) gene cluster on chromosome 11 (Fig. 4a), which harbours SNPs that have been associated with multiple cardiometabolic traits<sup>20</sup> and intermediate phenotypes<sup>21,22</sup>. In line with the observed pleiotropy among the serum-lipid-associated SNPs, the lipid-responsive peak in the *FADS2* promoter contained GWAS SNPs for all tested serum lipid traits<sup>11</sup> (LDL, HDL, TC and TG), with the strongest signal coming from rs99780 for LDL ( $P = 2.39 \times 10^{-21}$ ; Table 1). Notably, the observed open chromatin peak in *FADS2* was more accessible with palmitic acid treatment than with oleic acid treatment (Fig. 4b; FDR = 0.0021), corresponding to the fact that one of the substrates of *FADS2* is palmitic acid<sup>23</sup>.

This response at a GWAS locus for serum lipids was reminiscent of the strong SFA response at the *GPAM* locus (Fig. 2e); in fact, all five GWAS SNPs for serum lipids in lipid-responsive gene promoters within chromosomal interactions exhibited increased chromatin accessibility specifically in response to SFA intake in adipocytes (Supplementary Table 11). The lipid-responsive enhancers that interacted with gene promoters in adipocyte pChI-C exhibited a similar trend (Supplementary Table 12). These results suggest that environmental responses, particularly to saturated fat intake, explain functional mechanisms at these lipid GWAS loci.

Because signals that do not reach genome-wide significance probably also contribute to the heritability of cardiometabolic disorders, particularly with the added effect of relevant environmental stimuli, we wanted to test the combined effect of all

variants in our lipid-responsive regions while still accounting for the linkage disequilibrium (LD) between them. We therefore tested whether genetic variants in the *cis* region (gene body  $\pm$  500 kb) of all 154 lipid-responsive, interacting promoters contributed significantly to the heritability of serum lipid levels. For these analyses, we used the LD score-partitioned heritability method<sup>24</sup> and GWAS summary statistics from the high-powered meta-GWAS for serum lipid traits<sup>11</sup>. We found that 2.9% of all variants resided within the *cis* regions of the 154 genes, and these variants contributed significantly to the heritability of the four lipid traits ( $0.0088 \leq P \leq 0.045$ , with an average enrichment of 2.915; Supplementary Table 13). In contrast, 5.5% of all variants resided within the *cis* regions of genes with promoters interacting with lipid-responsive enhancers. These SNPs contributed significantly to the heritability of HDL, but not to that of the other lipid traits (Supplementary Table 14). This is consistent with the more diffuse overall functional characterization of the genes that interacted with lipid-responsive enhancers when compared to the genes with lipid-responsive open chromatin in their promoters. Overall, these results indicate that adipocyte lipid-responsive, interacting loci are important in modulating serum lipid levels in humans and provide evidence that variants in these regions might have a role in G×E interactions in humans.

**Lipid responses identify new G×E interactions for BMI in UK Biobank.** The large, deeply phenotyped UK Biobank<sup>4</sup> cohort can provide a valuable resource for G×E studies, particularly because the participants' environmental phenotypes have been characterized in a systematic manner. Saturated fat intake has known adverse effects in the context of cardiometabolic disorders<sup>25-27</sup>, and we present evidence here in human adipocytes of an enhanced effect of human adipocyte genomic responses to SFA intake on cardiometabolic traits in comparison to genomic responses from MUFA intake (Fig. 2c-e and Supplementary Tables 11 and 12). To maximize the number of individuals for whom phenotypes were available, as well as to aim for the most relevant environment and cardiometabolic outcome, we used dietary intake of saturated fat (24-h recall) as the environmental variable and BMI as the outcome for our G×E analysis.

We first tested whether there were any genome-wide-significant signals for G×E interactions by using 167,908 individuals in the UK Biobank. We corrected the BMI measurements for array type,

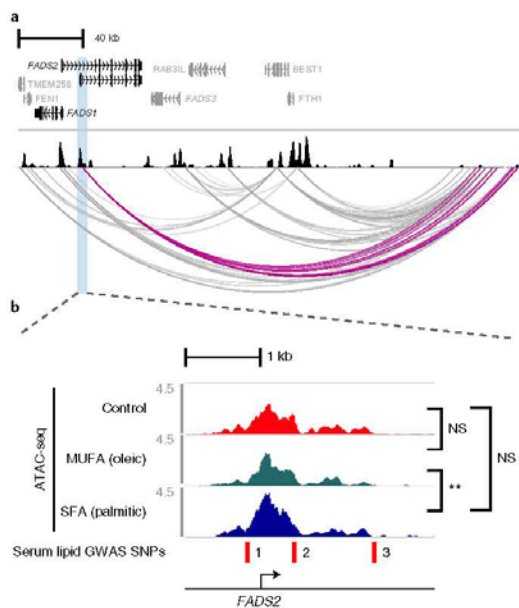
sex (inferred), age (when the participant attended an assessment centre), age<sup>2</sup>, the assessment centre ID and genetic principal components 1–20, as done previously<sup>28</sup>. We then inverse normal transformed the residuals to account for mean–variance relationships in the phenotype, which have been shown previously to impact G×E signals<sup>29</sup>. In the quantile–quantile plot from the genome-wide scan for G×E interactions, there was no evidence of genomic inflation in the G×E linear model (Supplementary Fig. 6). Furthermore, the fact that we were not able to detect any genome-wide significant signals in this genome-wide G×E analysis (see equation (1) in the Methods; Supplementary Fig. 6) supports the feasibility of our reverse G×E candidate search approach, which identifies functional candidates for G×E analyses from molecular genomics data produced under biologically relevant conditions.

We have provided evidence that lipid challenge in adipocytes highlights important regions of the genome that respond to environmental cues and contribute to the heritability of cardiometabolic traits. Thus, these regions represent strong candidates for G×E interactions in humans. The 154 promoters in chromosomal interactions contained 91 lipid-responsive open chromatin sites (Fig. 3b), and we determined that 75 of these 91 candidate regions contained variants with minor allele frequency (MAF) > 0.05 in the set of 167,908 individuals for whom we had both dietary saturated fat intake and BMI phenotypes available (Supplementary Table 15).

We performed G×E analysis by incorporating all SNPs residing in the open chromatin, lipid-responsive promoter regions ( $n = 290$ ; Supplementary Table 15) into a multivariable linear model (see equation (2) in the Methods). This resulted in the identification of 14 significant nonredundant G×E SNPs ( $LD r^2 < 0.2$ ) in 12 interacting promoters, including new G×E SNPs in the promoters of the well-known lipid-associated genes encoding hormone-sensitive lipase (*LIPE*), coactivator-associated arginine methyltransferase 1 (*CARM1*) and perilipin 2 (*PLIN2*) (Table 2 and Supplementary Table 16).

We next performed a similar G×E analysis on all SNPs at the lipid-responsive enhancers that interacted with gene promoters in human adipocytes. Of the 173 lipid-responsive regions within the interacting enhancers (Supplementary Fig. 5), 142 contained SNPs with MAF > 0.05 in the 167,908 individuals in the UK Biobank (Supplementary Table 17). We used the same multivariable linear model approach to test these SNPs ( $n = 410$ ) for an interaction with the effect of saturated fat intake on BMI and found 24 nonredundant ( $LD r^2 < 0.2$ ) significant G×E SNPs (Supplementary Table 18). Given that enhancer fragments can interact with more than one promoter-containing pChI-C bait, these 24 nonredundant SNPs interacted with a total of 27 promoter baits in human adipocytes (Supplementary Table 18).

**Identifying altered chromatin states at G×E SNP sites.** The differential chromatin accessibility in response to lipid challenge in adipocytes probably stems from altered chromatin states, such as TF binding or histone modifications. This idea is supported by our finding that the lipid-responsive regions within chromosomal interactions are enriched for the motifs of TFs important in lipid metabolism (Fig. 3a and Supplementary Table 6). To determine the predicted allelic effect of the G×E SNPs on chromatin features, we used the DeepSEA tool<sup>30</sup>, which applies a deep learning algorithm to publicly available molecular genomics data to predict chromatin features on the basis of genomic sequence *in silico*. Notably, 11 of the 20 (55%) G×E SNPs in lipid-responsive promoters had a functional significance score of less than 0.05, and the predicted impacts of the G×E SNPs included differential binding of RXRA (Supplementary Table 19). Conversely, only 5 of the 26 (19%) G×E SNPs in lipid-responsive enhancers had functional significance scores of less than 0.05 (Supplementary Table 20). It is worth noting that the publicly available data used to train the DeepSEA neural network do not



**Fig. 4 | A lipid-responsive open chromatin region in human primary adipocytes at the 11q12.2 *FADS1-FADS2-FADS3* locus harbours GWAS SNPs for serum lipid traits. a**, Genome browser snapshot showing the *FADS1-FADS2-FADS3* locus with data from adipocyte baseline ATAC-seq (one representative example from  $n = 3$  vehicle control (BSA) ATAC-seq libraries) and pChI-C (interactions identified in at least one condition from the adipocyte lipid challenge pChI-C analysis were included; see Methods). Chromosomal interactions of the *FADS2* promoter are highlighted in magenta. **b**, Read coverage (BPM) in one representative ATAC-seq library ( $n = 3$  replicates per condition) from vehicle control (red), MUFA (green) and SFA (blue) treatment. The lipid-responsive peak in one of the *FADS2* promoters is more accessible in SFA-treated than in MUFA-treated adipocytes and contains three independent GWAS SNPs for serum lipid traits (1, rs191508698; 2, rs968567; 3, rs99780). FDR was calculated (adjusting for  $n = 122,252$  ATAC-seq peaks) from the  $P$  values of the QL  $F$  test (see Methods) in one-way ANOVA. For the post hoc test to determine which comparison was significant after one-way ANOVA (MUFA vs. control, SFA vs. control or SFA vs. MUFA), we determined the least significant difference; \*\*FDR = 0.0021; NS, not significant.

include molecular genomics data for adipocytes or adipose tissue. Therefore, it is possible that the G×E SNPs fall into cell-type-specific regulatory elements and disrupt chromatin features that cannot be predicted with this tool.

To obtain further evidence for the function of the G×E SNPs in adipose tissue, we examined whether the G×E SNPs affected local gene expression in human adipose tissue as *cis* expression quantitative trait loci (*cis*-eQTLs) and whether the *cis*-eQTL target gene was the same gene that harboured the lipid-responsive promoter and G×E signal. We found that 3 of the 12 genes with lipid-responsive promoters harbouring G×E signals were also regulated in *cis* by their G×E SNP at the genome-wide significance level in subcutaneous adipose RNA-seq data ( $n = 335$ ) from the Finnish METSIM cohort<sup>15,31</sup> (Table 2 and Supplementary Table 16). These genes were *GLTSCR2* (encoding glioma tumour-suppressor candidate region gene 2 protein), *PLIN2* and *LDB3* (LIM domain binding 3). Additionally, 2 of the 27 genes interacting with lipid-responsive

**Table 2 | Significant G×E interactions affecting BMI from a multivariable linear model for 290 promoter SNPs in lipid-responsive ATAC-seq peaks**

SNP	$P_g$	$P_{gE}$	$\beta_g$	$\beta_{gE}$	Genes in bait	Cis-eQTL FDR <sup>a</sup> (from ref. <sup>15</sup> )	Target gene (from ref. <sup>15</sup> )	$\log_2$ (FC) (from ref. <sup>15</sup> )
rs1974817 <sup>b</sup>	0.0089	0.0010	2.3	-0.089	GLTSCR2- SNORD23	$2.4 \times 10^{-31}$	SEPWI	0.73
rs58631862	0.032	0.0031	0.085	-0.0035	RGMB	-	-	-
rs74249860	0.0013	0.0043	0.081	-0.0021	SH3GL3	0.021	GOLGA6L4	0.90
rs112438892 <sup>b</sup>	0.0017	0.0050	-0.50	0.012	CARM1	$2.7 \times 10^{-6}$	SMARCA4 ICAM4	-0.28 0.38
rs17625418	0.015	0.0054	-2.2	0.073	GLTSCR2- SNORD23	$1.3 \times 10^{-19}$ 0.0038	SEPWI GLTSCR2	-0.76 -0.3
rs3848589	0.045	0.014	0.073	-0.0027	HOOK2-JUNB	0.0058	CACNA1A	-0.6
rs882881	0.034	0.016	-0.051	0.0017	PLIN2	-	-	-
rs35213231	0.56	0.020	0.66	-0.038	RNU2-10P	0.019	PTPRG	0.29
rs41322049	0.29	0.021	0.032	-0.0021	BLVRB- SFTBN4	-	-	-
rs35678764 <sup>b</sup>	0.028	0.025	-5.7	0.19	RDH8-COL5A3	-	-	-
rs10788522 <sup>b</sup>	0.013	0.027	3.9	-0.14	LDB3	$6.1 \times 10^{-4}$	LDB3	0.42
rs10422283	0.045	0.029	-0.072	0.0023	LIPE-LIPE-AS1	-	-	-
rs867773 <sup>b</sup>	0.31	0.033	-2.0	0.11	PLIN2	0.048	PLIN2	-0.39
1:12245360_CCTTTTT_C	0.047	0.034	0.64	-0.023	TNFRSF18- MIR4632	-	-	-

The reported  $P$  values are from the  $\beta$  values in the multivariable linear model (see equation (2) in the Methods), where  $g$  is the number of minor alleles of the genotype and  $e$  is saturated fat intake. Here  $P_g$  indicates the  $P$  value for the genotype effect and  $P_{gE}$  indicates the  $P$  value for the G×E effect.  $\beta$  values follow the same notation. For the multivariable linear model, there were a total of 290 SNPs and 38,394 individuals with no missing data available for study. <sup>a</sup>Cis-eQTLs were identified in adipose tissue from the METSIM cohort<sup>15,31</sup>. <sup>b</sup>When more than one non-independent SNP (LD  $r^2 > 0.2$ ) has a significant G×E  $P$  value for the lipid-responsive region, only the top SNP is reported; the SNPs in LD with the top SNP are listed in Supplementary Table 16.

enhancers were regulated in *cis* by their interacting G×E SNP at the genome-wide significance level in human subcutaneous adipose tissue (Supplementary Table 18): *THBS2* (thrombospondin 2) and *CITED4* (Cbp/p300-interacting transactivator with Glu/Asp-rich C-terminal domain 4).

We further examined whether the imputed *cis* expression values for these five *cis*-eQTL target genes (eGenes) were correlated with BMI or other obesogenic cardiometabolic phenotypes, as determined through transcriptome-wide association analysis (TWAS)<sup>32,33</sup>. The *LDB3* adipose expression model was strongly associated with BMI, arm fat percentage (genome-wide significant TWAS score > 5.0) and other related body fat distribution phenotypes (TWAS score > 4.0)<sup>32</sup>. Furthermore, the tibial artery expression model for *LDB3* was also significantly associated with high blood pressure and cardiovascular disease (genome-wide significant TWAS score < -5.0)<sup>32</sup>. None of the other eGene expression models was associated with cardiometabolic phenotypes at the genome-wide significance level.

Because the adipose expression models of *LDB3* were significantly associated with BMI in TWAS<sup>32</sup>, we followed up on the most significant G×E SNP in the *LDB3* lipid-responsive peak (Fig. 5a), rs10788522, which was also an adipose *cis*-eQTL for *LDB3* in the METSIM cohort (Table 2). As evidenced by the ATAC-seq reads intersecting with SNP rs10788522 (Fig. 5b), we found, by electrophoretic mobility shift assay (EMSA), that adipocyte nuclear protein bound this G×E SNP (Fig. 5c and Supplementary Table 21). Whereas *LDB3* was expressed in subcutaneous adipose tissue from the METSIM cohort, as well as the Gene-Tissue Expression (GTEx) Project<sup>34</sup>, we could not reliably detect *LDB3* expression by quantitative PCR (qPCR) in the lipid-challenged adipocytes (data not shown). We therefore determined whether any publicly available datasets analysed the transcriptome of sorted cells from human adipose tissue. In previously published microarray datasets

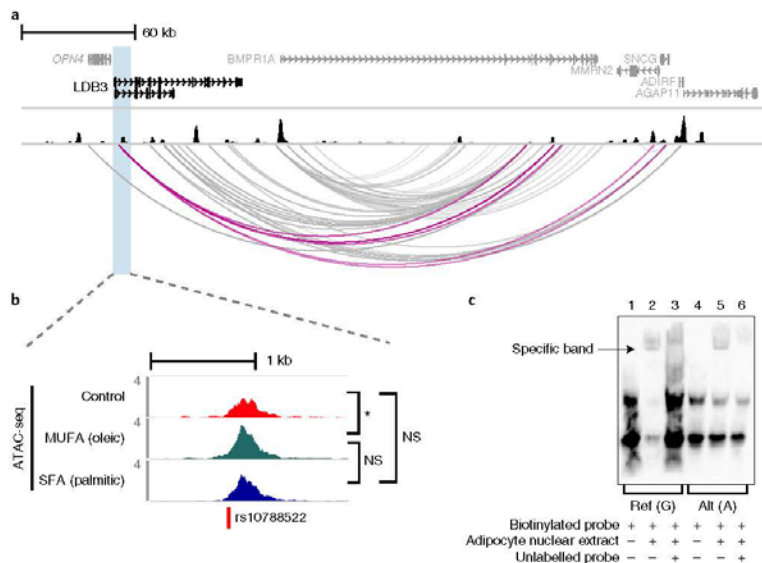
(GSE80654 and GSE100795) from human adipose biopsies in which adipocytes were collected and the remaining cell types were separated by fluorescence-activated cell sorting (FACS), *LDB3* was expressed in human adipocytes at a level comparable to that in the other adipose cell types<sup>35,36</sup>. Taken together with our finding that the *LDB3* promoter responds to lipid uptake in adipocytes, these data support the conclusion that the G×E SNP rs10788522 regulates *LDB3* expression in adipocytes in response to dietary saturated fat and that the interaction has a protective (BMI-lowering) effect (Table 2). Altogether, we provide a mechanistic interpretation and fine-mapping of a causal G×E SNP, rs10788522, narrowing it to the promoter of *LDB3* that exhibits differential open chromatin in response to lipid challenge in human adipocytes (Figs. 5 and 6).

## Discussion

It is well established that environment has a major role in the development of cardiometabolic disorders. However, G×E interactions have been challenging to detect owing to both the lack of extensive study cohorts with sufficient statistical power to detect the small G×E effects and the complexity of environmental exposures that are difficult to measure in a standardized way in humans<sup>1</sup>. Systematic identification of the effects of defined environmental contributions to cardiometabolic disorders is thus necessary to effectively move towards the promise of precision medicine. Through our integration of context-specific molecular genomics data with human epidemiological and clinical outcome data in the UK Biobank, we provide much-needed information on how the chromatin landscape of human adipocytes responds to external environmental signals and identify the molecular basis of new G×E interactions in humans (Fig. 6).

We reversed the typical approach of selecting candidate G×E interactions among GWAS SNPs, by first scanning the genome for molecular responses to controlled environmental stimuli, apply-





**Fig. 5 | Fine-mapping of the gene-diet interaction for BMI in the *LDB3* promoter region. a**, Genome browser snapshot of the *LDB3* locus with adipocyte baseline ATAC-seq (one representative example from  $n = 3$  vehicle control (BSA) ATAC-seq libraries) and pChI-C data (interactions identified in at least one condition from the adipocyte lipid challenge pChI-C analysis are included; see Methods). *LDB3* promoter interactions are highlighted in magenta. **b**, Read coverage (BPM) in one representative ATAC-seq library ( $n = 3$  replicates per condition) for control (red), MUFA (green) and SFA (blue) treatment. The lipid-responsive ATAC-seq peak harbouring the significant G $\times$ E SNP rs10788522 resides in the *LDB3* gene. FDR was calculated (adjusting for  $n = 122,252$  ATAC-seq peaks) from the  $P$  values of the QL  $F$  test (see Methods) in one-way ANOVA. For the post hoc test to determine which comparison was significant after one-way ANOVA (MUFA vs. control, SFA vs. control or SFA vs. MUFA), we determined the least significant difference; \*FDR = 0.024. The  $P$  value for  $\beta_{G \times E}$  at rs10788522 was derived from the multivariable linear model testing for G $\times$ E interactions in the UK Biobank;  $P = 0.027$ . **c**, EMSA showing binding of adipocyte nuclear protein to the G $\times$ E SNP rs10788522. The specific band is competed away for both alleles (lanes 3 and 6). Ref, reference allele; alt, alternate allele.

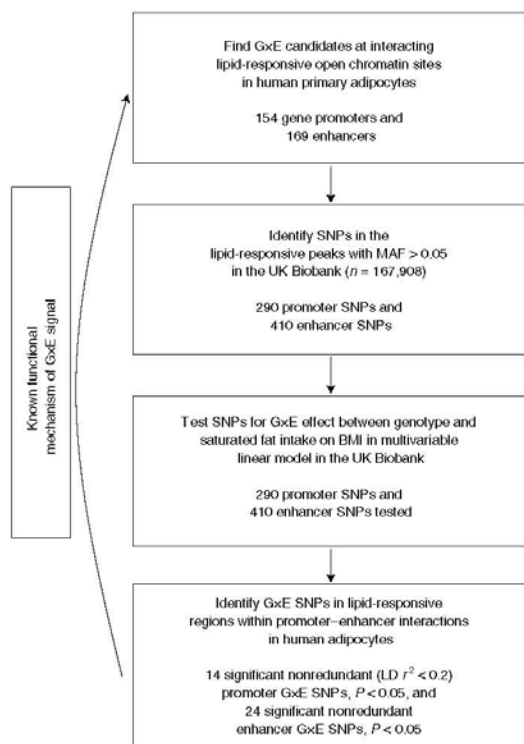
ing a cellular model of the effects of saturated or monounsaturated fat intake on chromatin dynamics in primary human adipocytes. Through our integration of chromatin accessibility and chromosomal interactions in lipid-challenged adipocytes, we identified lipid-responsive open chromatin within promoter-enhancer contacts, effectively identifying candidate G $\times$ E interaction genes with strong evidence of genomic regulation in response to fatty acid uptake and processing in human adipocytes. This systematic approach culminated in testing a total of 700 SNPs in the accessible, lipid-responsive chromatin regions for interactions with dietary saturated fat intake affecting BMI in the UK Biobank<sup>4</sup>. This led to the identification of 14 significant, nonredundant G $\times$ E SNPs in 12 gene promoter regions and 24 nonredundant G $\times$ E SNPs in 20 enhancers, representing new gene-diet interactions affecting BMI (Fig. 6).

We observed that the *LDB3* gene is regulated in *cis* by its promoter G $\times$ E SNP in human adipose tissue, and previous TWAS analyses have shown that the imputed local adipose expression of *LDB3* is significantly associated with BMI and related cardiometabolic phenotypes<sup>32,35</sup>. Notably, individuals with nonsynonymous mutations in exon 6 of *LDB3* have been shown to exhibit autosomal dominant myofibrillar myopathy characterized by fatty degeneration (steatosis) of the muscle that progresses with age<sup>37,38</sup>. This is suggestive of the role of lipid metabolism in the pathophysiology of these variants. Here we show that accessibility of the promoter region is increased in response to lipid challenge in human adipocytes, and adipocyte nuclear protein binds to the G $\times$ E SNP site. Although we did not detect *LDB3* expression in our cultured adipocytes, we

found that the gene was expressed in mature adipocytes isolated from human adipose biopsies, in vivo, suggesting that future studies to understand the role of *LDB3* in adipocytes may require in vivo mouse models. In line with this, *LDB3* is known to bind  $\alpha$ -actinin isoforms that are not muscle specific, and actin cytoskeleton organization is critically important in maintaining proper tissue functions. Taken together with the adipose *cis*-eQTL and TWAS results, our lipid challenge findings provide a functional mechanism for the G $\times$ E effect in human adipocytes, adding to the knowledge of environmental response to diet and the consequent effects on genetic predisposition to cardiometabolic traits in humans.

The 154 lipid-responsive promoters within adipocyte chromosomal interactions provide a set of biologically important genes for studies in adipose tissue. These gene regions have a higher conservation score and the 154 genes are more likely to be intolerant to LoF mutation than expected by chance, suggesting that there are evolutionary constraints to maintain their proper function. The LoF-intolerant genes are widely and highly expressed<sup>19</sup>, and they may exhibit pleiotropy. Nonetheless, the lipid-responsive mechanism of genomic regulation identified for the LoF-intolerant genes in this study provides evidence that these genes may be important in maintaining energy homeostasis, which is critical for survival.

It is known that dietary saturated fat intake is correlated with various adverse cardiometabolic outcomes<sup>26,27</sup>, and a genetic risk score (GRS) for obesity-related traits was previously shown to interact with saturated fat intake to affect BMI<sup>26</sup>. However, as the authors of this study note, the underlying mechanisms for the G $\times$ E



**Fig. 6 | Analytical approach.** Flowchart of our approach to integrate molecular genomics data created in human adipocytes in physiologically relevant contexts that, when combined with human cohort molecular and phenotype data, enable the detection of G×E signals.

interactions remain elusive, particularly when the effect is estimated across tens of SNPs<sup>26</sup>. Here we bridge this knowledge gap and show through a genome-wide scan of regulatory open chromatin responses to saturated versus monounsaturated fat uptake that a subset of these responses are probably specific to saturated fat intake and, when dysregulated at the genetic level, could underlie GWAS and G×E signals for cardiometabolic traits. We note that we assessed the effects of exogenous fatty acids and did not quantify the consequent cellular fatty acid processing, which could include desaturation of palmitic acid (C16:0) to the MUFA palmitoleic acid (C16:1), by stearoyl-CoA desaturase (SCD). Thus, we do not know whether the effects we observed at the DNA level were directly due to SFA signalling or resulted from downstream signalling mechanisms of the SCD pathway, affecting MUFA concentrations inside the cell.

Our finding that SNPs within the *cis* regions of the 154 lipid-responsive gene promoters (gene body ± 500 kb) contributed significantly to the heritability of serum lipid levels suggests that responses to lipid uptake in adipocytes are associated with cellular programs that can modulate serum lipid levels. Correspondingly, we identified five adipocyte lipid-responsive, interacting gene promoters and three lipid-responsive enhancers that harboured genome-wide-significant signals for serum lipid traits, including in the well-known *FADS1–FADS2–FADS3* gene cluster involved in nutrient sensing. Specific lipid-associated SNPs within this locus have undergone positive selection in Inuits, which is thought to

have been in response to the polyunsaturated fatty acids (PUFAs) in diets with high amounts of marine mammalian fat<sup>39</sup>. Furthermore, a gene–diet interaction for PUFA intake has been identified in Europeans at the *FADS1–FADS2–FADS3* locus<sup>40</sup>, and, while many studies of the effects of *FADS1–FADS2* polymorphisms on PUFA metabolism have shown a clear role for this locus in modulating serum fatty acid levels, an understanding of the mechanistic effects of these SNPs, associated with cardiometabolic disorders, has been less conclusive<sup>41,42</sup>. Our results suggest that the underlying mechanisms may derive from the effects of saturated fats at this locus. This additional role for a well-established, yet-inconclusive locus supports the applicability of our approach to identify G×E interactions through characterization of molecular genomic responses to relevant environmental stimuli.

Although there has been a strong international effort by the scientific community to characterize genomic regulatory elements in various cell types and tissues, many of the publicly available datasets and corresponding genomic annotations have been created in cells at steady state or under unstimulated baseline conditions<sup>43–46</sup>. Context-specific molecular genomics studies have mainly been performed in immune cell types<sup>47–49</sup>, while similar studies in other cell types are scarce. In line with the importance of studying molecular genomic phenotypes under physiologically relevant conditions, we found here that, whereas most lipid-responsive regions in human adipocytes reside in known adipocyte enhancers, a subset emerge from regions that were not identified as open chromatin and were annotated as quiescent<sup>5</sup> in unchallenged adipocytes, indicating that quiescent regions of the human genome are activated under specific environmental contexts. Thus, this genome-wide scan for response to fatty acid uptake in adipocytes adds to the currently incomplete understanding of genomic regulation in contexts that are expected to confer complex cardiometabolic disease states.

In conclusion, this study highlights the value of performing genome-wide functional genomics experiments in a context-specific manner to advance understanding of environmental epigenomic responses underlying complex traits. We performed a global assessment of the genomic responses of primary human adipocytes to dietary fatty acid uptake, through incorporating open chromatin and chromosomal interaction data that we followed for G×E interactions in UK Biobank. Overall, our study helped discover candidate functional mechanisms at 38 new gene–diet interactions on BMI, identified over 100 genes important for lipid uptake that may contribute to variance in cardiometabolic traits and uncovered a new set of interacting open chromatin elements responding to lipid challenge in a primary human cell type relevant for lipid synthesis and storage.

## Methods

**Cell lines and culture reagents.** We obtained and cultured primary human white preadipocyte cells as recommended by PromoCell (PromoCell, C-12731, lot 395Z024) for preadipocyte growth and differentiation into adipocytes. Cell medium (PromoCell) was supplemented with 1% penicillin–streptomycin. We maintained the cells at 37 °C in a humidified atmosphere at 5% CO<sub>2</sub>. For the lipid challenge experiments in adipocytes, we serum starved cells for 16 h with 0.5% FCS in supplemented adipocyte basal medium (PromoCell), before treatment with 200 μM palmitic acid:BSA complex, 200 μM oleic acid:BSA complex (Sigma-Aldrich, O3008) or 0.23% fatty acid-free BSA (Sigma-Aldrich, A8806) as a vehicle control, in medium containing 0.5% FCS for 24 h before performing experiments.

**Palmitic acid conjugation to BSA.** We dissolved 25.6 mg of palmitic acid (Sigma-Aldrich, P5585) into 1 ml of 0.15 M NaCl at 70 °C in a shaking heat block to make 100 mM palmitic acid solution. We added the palmitic acid solution dropwise into 10% (wt/vol) BSA in 0.15 M NaCl at 37 °C to generate palmitic acid:BSA conjugate at 8 mM stock for palmitic solution.

**Oil Red O staining and quantification.** We prepared Oil Red O stock by making 0.3% Oil Red O solution in >99% isopropanol and filtering through a 0.45-μm filter. This solution was diluted 3:2 in water, incubated at room temperature for 10 min and then filtered through a 0.22-μm filter. We fixed cells for 30–60 min in 10%

formalin, rinsed with distilled water and incubated for 2–5 min at room temperature with 60% isopropanol. We stained with Oil Red O for 15 min, rinsed well with water and collected images for quantification. Cells were photographed with a Keyence bright-field light microscope under  $\times 10$  magnification for publication images and  $\times 20$  magnification for lipid droplet quantification. Lipid droplet numbers were determined for ~20 cells per condition (untreated, BSA, palmitic acid and oleic acid). The total area of Oil Red O staining was quantified with ImageJ<sup>50</sup>.

**ATAC-seq.** We performed the ATAC-seq protocol in untreated primary human preadipocytes and adipocytes for 300,000 nuclei in three biological replicates per cell type, similarly to the protocol developed in Buenrostro et al.<sup>5</sup>. Specifically, we lysed cells in ice-cold lysis buffer (10 mM Tris-HCl pH 7.4, 10 mM NaCl, 3 mM MgCl<sub>2</sub> plus 0.03–0.1% Tween-20 for 10 min on ice. We centrifuged at 500g for 10 min at 4°C and then resuspended the nuclear pellet in 50  $\mu$ l of transposition master mix (25  $\mu$ l of 2 $\times$  TDE1 buffer, 2.5  $\mu$ l of transposase, 22.5  $\mu$ l of nuclease-free water, Illumina, FC-121-1030). We incubated samples at 37°C for 30 min and then purified the DNA with the Qiagen MinElute kit (Qiagen, 28204). Libraries were amplified for six cycles and sequenced on an Illumina HiSeq 4000 to produce an average of 23,376,290 ( $\pm 3,337,206$ ) reads.

For the ATAC-seq analysis in primary human adipocytes that underwent lipid challenge, we performed omni-ATAC as developed in Corces et al.<sup>51</sup> for 300,000 nuclei in three biological replicates per condition. Specifically, we treated the cells with DNase I (Worthington; 200 U ml<sup>-1</sup>) at 37°C for 30 min, rinsed the cells with ice-cold PBS, scraped the cells gently to the side of the plate, resuspended them in 50  $\mu$ l of ice-cold lysis buffer (10 mM Tris-HCl pH 7.4, 10 mM NaCl, 3 mM MgCl<sub>2</sub>, 0.1% Igepal CA-630, 0.1% Tween-20, 0.01% digitonin) and incubated them on ice for 3 min. We washed with 1 ml of ice-cold lysis quench (10 mM Tris-HCl pH 7.4, 10 mM NaCl, 3 mM MgCl<sub>2</sub>, 0.1% Tween-20) and centrifuged at 500g for 10 min at 4°C. The nuclear pellet was resuspended in 50  $\mu$ l of transposition master mix (25  $\mu$ l of 2 $\times$  TDE1 buffer and 2.5  $\mu$ l of transposase, 16.5  $\mu$ l of PBS, 0.5  $\mu$ l of 1% digitonin, 0.5  $\mu$ l of 10% Tween-20, 5  $\mu$ l of nuclease-free water). We incubated samples at 37°C for 30 min with mixing at 1,000 r.p.m. and then purified the DNA with the Qiagen MinElute kit. Libraries were amplified for 6–7 cycles and sequenced on an Illumina HiSeq 4000 to produce an average of 40,315,572 ( $\pm 14,577,770$ ) reads.

**ATAC-seq data processing and peak calling.** We processed the sequencing reads and performed quality control by using the ENCODE ATAC-seq Data Standards and Prototype Processing Pipeline. Briefly, we aligned reads to the human reference genome (GRCh37/hg19) with Bowtie2 (ref.<sup>52</sup>) v2.2.9 (with parameters -k 4 -X 2000 -local) and filtered out unpaired mapped reads and reads with MAPQ < 30 (SAMtools<sup>53</sup>) as well as duplicates (marked with Picard Tools). Only reads from the autosomes and X chromosome were retained for downstream analyses.

**Identification of differentially accessible ATAC-seq peaks.** Read alignments from all untreated human preadipocyte and adipocyte libraries (three biological replicates per cell type) were merged before peak calling. Peaks were called with MACS2 (ref.<sup>54</sup>) v2.1.1 (by using the BEDPE function), and peaks with FDR < 0.05 were retained. We filtered out peaks in blacklisted regions and peaks that did not replicate in two of the three biological replicates in at least one condition. For differential accessibility analyses, we retained peaks with counts per million (c.p.m.)  $\geq 1$  in at least three libraries. We then input aligned read counts for each peak into *cpm*<sup>55</sup> v1.20.0 and normalized the counts for G+C content, peak length and library size, before inputting the counts into *edgeR*<sup>56</sup> v3.16.5 to detect differentially accessible peaks between preadipocytes and adipocytes with the generalized linear model (GLM) functionality and QL *P* test, applying an FDR threshold of 0.05. Libraries for adipocyte lipid challenge ATAC-seq were processed identically until the differential accessibility analysis. To detect open chromatin regions that exhibited differential accessibility in lipid-challenged adipocytes, we performed one-way ANOVA with the GLM functionality and QL *P* test functionality in *edgeR*<sup>56</sup>, applying an FDR threshold of 0.05. For the post hoc test to determine which comparisons were significant after one-way ANOVA (oleic acid versus BSA, palmitic acid versus BSA or oleic acid versus palmitic acid), we determined the least significant difference.

**Transcription factor motif enrichment in ATAC peaks.** We used HOMER (v4.9)<sup>57</sup> to investigate the enrichment of motif sequences in open chromatin regions. For enrichment in differentially accessible peaks between untreated preadipocytes and adipocytes, we used the consensus peak set (all peaks that were called in both the preadipocyte and adipocyte data) as the background. We utilized the *de novo* motif enrichment functionality. To ensure that our background input file was not biasing the results, we performed the same analysis with the genome as the background input, which produced largely the same results, with smaller *P* values (data not shown). For enrichment in lipid-responsive peaks in adipocyte promoter–enhancer contacts, we used non-differentially accessible peaks within the promoter–enhancer contacts as the background. Owing to the small number of peaks (*n* = 264), we used the known TF motif enrichment functionality.

**Hi-C library preparation.** We prepared the Hi-C libraries for the primary human adipocyte lipid challenge experiment as described in Pan et al.<sup>15</sup> in two biological

replicates per condition (BSA, oleic acid and palmitic acid). These methods were adapted by closely following the in-nucleus Hi-C methods in Rao et al.<sup>57</sup> and Nagano et al.<sup>58</sup>. Specifically, we fixed 7–10 million adherent cells directly in the culture plate in 2% formaldehyde and quenched with glycine to a final concentration of 125 mM. Cells were lysed in ice-cold lysis buffer (10 mM Tris-HCl pH 8.0, 10 mM NaCl, 0.2% Igepal CA-630, 1 $\times$  protease inhibitors: cOmplete, EDTA-free Protease Inhibitor cocktail) on ice for 30 min with occasional agitation. We split lysates into aliquots with 5 million nuclei and centrifuged at 2,500g for 5 min at 4°C. The nuclear pellets were resuspended in 50  $\mu$ l of 0.5% SDS in 1 $\times$  NEBuffer 2 (New England Biolabs) and incubated at 62°C for 10 min. We then added 145  $\mu$ l of water and 25  $\mu$ l of 10% Triton X-100 and incubated at 37°C for 15 min. We digested chromatin by adding 25  $\mu$ l of 10 $\times$  NEBuffer 2 and 400 U of HindIII (New England Biolabs), incubating at 37°C overnight with shaking (950 r.p.m.).

The next day, we marked the DNA ends with biotin (1.5  $\mu$ l of 10 mM dATP, 1.5  $\mu$ l of 10 mM dGTP, 1.5  $\mu$ l of 10 mM dTTP, 37.5  $\mu$ l of 0.4 mM biotin-14-dCTP (Invitrogen), 8  $\mu$ l of 5 U  $\mu$ l<sup>-1</sup> Klenow (New England Biolabs)), incubating for 60 min at 37°C; we then added 895  $\mu$ l of ligation mix (663  $\mu$ l of water, 120  $\mu$ l of 10 $\times$  NEB T4 DNA ligase buffer, 100  $\mu$ l of 10% Triton X-100, 12  $\mu$ l of 10 mg ml<sup>-1</sup> BSA, 5  $\mu$ l of 400 U  $\mu$ l<sup>-1</sup> T4 DNA ligase (New England Biolabs)). Ligation was performed at room temperature for 4 h with slow rotation, and 50  $\mu$ l of 20 mg ml<sup>-1</sup> proteinase K and 120  $\mu$ l of 10% SDS were added with incubation at 55°C for 30 min. We added 130  $\mu$ l of 5 M NaCl and incubated at 68°C overnight. We then performed an ethanol precipitation and sheared the purified DNA to 250–550 bp in size with a Covaris M220 instrument. Double size selection was performed with SPRI select agent (Beckman Coulter) by adding 0.55 volumes and then 0.15 volumes according to the manufacturer's instructions, eluting the final DNA in 300  $\mu$ l of 10 mM Tris pH 8.0.

Biotin pulldown was performed with 150  $\mu$ l of 10 mg ml<sup>-1</sup> DYNAL MyOne Dynabeads Streptavidin T1 (Invitrogen, 65601) per sample. First, the beads were washed twice with 400  $\mu$ l of 1 $\times$  Tween wash buffer (1 $\times$  TWB: 5 mM Tris-HCl pH 7.5, 0.5 mM EDTA, 1 M NaCl, 0.05% Tween-20) and resuspended in 300  $\mu$ l of 2 $\times$  binding buffer (2 $\times$  BB: 10 mM Tris-HCl pH 7.5, 1 mM EDTA, 2 M NaCl). Beads were then added to 300  $\mu$ l of sheared and size-selected DNA. We incubated at room temperature for 15 min with rotation to bind biotinylated DNA to the streptavidin beads. We washed twice with mixing at 55°C by adding 600  $\mu$ l of 1 $\times$  TWB and then washed beads in 100  $\mu$ l of 1 $\times$  NEB T4 DNA ligase buffer. We repaired the ends of the DNA by resuspending beads in 100  $\mu$ l of master mix (88  $\mu$ l of 1 $\times$  NEB T4 DNA ligase buffer with 10 mM ATR, 2  $\mu$ l of 25 mM dNTP mix, 5  $\mu$ l of 10 U  $\mu$ l<sup>-1</sup> NEB T4 PNK, 4  $\mu$ l of 3 U  $\mu$ l<sup>-1</sup> NEB T4 DNA polymerase I, 1  $\mu$ l of 5 U  $\mu$ l<sup>-1</sup> NEB DNA polymerase I, large (Klenow) fragment) and incubated samples at room temperature for 30 min. We washed twice with 1 $\times$  TWB, washed once with 1 $\times$  NEBuffer 2 and then resuspended samples in 100  $\mu$ l of dATP attachment master mix (90  $\mu$ l of 1 $\times$  NEBuffer 2, 5  $\mu$ l of 10 mM dATP, 5  $\mu$ l of 5 U  $\mu$ l<sup>-1</sup> NEB Klenow exo minus) and incubated samples for 30 min at 37°C. We washed twice with 1 $\times$  TWB and resuspended beads in 100  $\mu$ l of 1 $\times$  T4 DNA ligase buffer. We followed the manufacturer's instructions for the Agilent SureSelect to ligate the paired-end adaptors. The beads were then washed twice with 1 $\times$  TWB and resuspended in 32  $\mu$ l of 1 $\times$  Tris buffer. DNA was removed from the streptavidin beads by heating at 98°C for 10 min. We followed the manufacturer's instructions for the Agilent SureSelect for precapture PCR, carried out for eight cycles.

**Promoter capture Hi-C library preparation.** RNA baits were designed in Mifsud et al.<sup>14</sup> for capturing HindIII fragments containing gene promoters (C. Osborne (Department of Medical & Molecular Genetics, King's College London, London, UK) kindly shared the exact design). As described in Mifsud et al.<sup>14</sup>, 120-mer RNA baits were designed to target both ends of HindIII fragments that contained annotated gene promoters (Ensembl promoters of protein-coding, noncoding, antisense, snRNA, miRNA and snoRNA transcripts). A bait sequence was deemed valid if the sequence had a G+C content of 25–65%, contained fewer than three consecutive Ns and was within 330 bp of the ends of the HindIII fragment. Hi-C library hybridization to the capture library was performed according to the manufacturer's instructions for the Agilent SureSelect. A total of 550 ng of the Hi-C library was hybridized to the biotinylated RNA baits, captured with DYNAL MyOne Dynabeads Streptavidin T1 and amplified in post-capture PCR to add index sequences, for 12 PCR cycles. The library was sequenced on the Illumina HiSeq 4000 platform. All six libraries were sequenced together across two lanes of the Illumina HiSeq 4000 to produce an average of 127,069,374 ( $\pm 16,855,586$ ) sequencing reads per library.

**Capture Hi-C data processing and interaction calling.** We processed the sequencing reads as described in Pan et al.<sup>15</sup>, by using the Hi-C User Pipeline (HiCUP) v0.5.9 (ref.<sup>59</sup>) with default settings except that the insert size restrictions for the filtering step were set to 200–600 bp. We called significant interactions for each library separately with CHICAGO software v1.1.1 (ref.<sup>60</sup>). We used the default threshold of 5 for calling significant interactions. To create a stringent set of interactions, we merged all pHi-C final alignments and called interactions with CHICAGO, again by using a threshold of 5; we then filtered these interactions to include only those that were called in both biological replicates in at least one condition (BSA, oleic acid or palmitic acid).

**Cross-species conservation analysis.** For the cross-species conservation analysis, we used the PhastCons score<sup>31</sup> available on the UCSC ENCODE database. Briefly, the PhastCons score uses a phylo-HMM to predict per-base conservation across species. We used the PhastCons scores for placental mammal alignment and calculated the mean score for each of the lipid-responsive regions for the protein-coding lipid responsive genes on autosomes (gene body  $\pm$  500 kb). To create a background set for this comparison, we calculated the mean score for all other protein-coding genes annotated by Ensembl in GRCh37 and their surrounding regions of  $\pm$  500 kb. We performed a non-parametric two-sided Wilcoxon signed-rank test to compare the lipid-responsive regions to the background set.

**LD score analysis.** We used LD score regression<sup>32</sup> to estimate the heritability explained by the lipid-responsive regions. More specifically, we generated an annotation for each lipid-responsive region consisting of the lipid-responsive gene (gene body  $\pm$  500 kb) and used the summary statistics from a lipid GWAS<sup>31</sup> to estimate the proportion of heritability explained by the 154 lipid-responsive promoters or 323 promoters that interacted with lipid-responsive enhancers in adipocytes, for the four lipid GWAS traits: serum TG, HDL, LDL and TC.

**Genotype and phenotype data from the UK Biobank cohort.** We downloaded imputed genotype data from the UK Biobank cohort<sup>3</sup>. We identified all SNPs in the lipid-responsive gene promoters involved in chromosomal interactions. For the G $\times$ E interaction test, we filtered out SNPs that had a genotype missing rate of greater than 5% or a MAF of less than 5%. We used the BMI value collected from the initial UK Biobank assessment visit at which participants were recruited. The data for 24-h recall of saturated fat consumption in diet was collected at five different time points, including during the initial assessment and four online cycle collections. If an individual had 24-h recall of saturated fat consumption collected at multiple time points, we used the value closest to the initial assessment. We then selected unrelated individuals of European ancestry from the UK Biobank participants who had data on both BMI and saturated fat diet collected for the G $\times$ E analysis. We corrected BMI for the following covariates and performed inverse normal transformation to ensure that BMI was normally distributed: array type, sex (inferred), age (when the participant attended the assessment centre), age<sup>2</sup>, the assessment centre ID and genetic principal components 1–20.

**Genome-wide G $\times$ E scan.** To verify that our significant G $\times$ E interactions were not caused by overall inflation, we fitted the linear G $\times$ E interaction model on all SNPs across the human genome in the UK Biobank data. We first selected SNPs that were not in the same LD block ( $r^2 < 0.2$ ) and then used the following linear model to detect the G $\times$ E interaction between each LD-pruned SNP and saturated fat intake on BMI

$$Y = \alpha + \beta_g g + \beta_e e + \beta_{ge} ge + \epsilon \quad (1)$$

where  $Y$  is a vector of inverse normal transformed BMI values and  $g$  represents the vector of the number of minor alleles in the genotypes of the target SNP for the individuals in the study sample. We used  $e$  for the vector of saturated fat intake levels as the environmental covariate, and  $\epsilon$  represents a vector of random errors, in which each entry is independently and normally distributed.  $\alpha$  and  $\beta$  are the estimated parameters. The test for an interaction is based on the coefficient  $\beta_{ge}$ . A non-zero  $\beta_{ge}$  value indicates that there is an interaction between the genotype and environmental factor for the outcome phenotype. We constructed a quantile-quantile plot to compare the  $P$  values of  $\beta_{ge}$  and the expected  $P$  values based on multiple testing.

**Testing for G $\times$ E interaction by multivariable linear model.** We used the following multivariable linear model to detect the G $\times$ E interactions between SNPs and saturated fat intake for BMI

$$Y = \alpha + \beta_e e + \sum_{i=1}^N (\beta_{gi} \times g_i) + \sum_{i=1}^N (\beta_{gei} \times g_i e) + \epsilon \quad (2)$$

where  $Y$  is a vector of inverse normal transformed BMI values and  $g$  represents a vector of the number of minor alleles in the genotypes of the target SNP  $i$ , where  $i = 1, \dots, N$  SNPs, for the individuals in the study sample. We use  $e$  for the vector of saturated fat intake levels as the environmental covariate, and  $\epsilon$  represents a vector of random errors, in which each entry is independently and normally distributed.  $\alpha$  and  $\beta$  are the estimated parameters. The significance of the interaction is given for the coefficient  $\beta_{gei}$ . A non-zero  $\beta_{gei}$  value indicates that there is an interaction between the genotype  $i$  and the environmental factor (24-h saturated fat recall) for the outcome phenotype (BMI). Individual  $\beta_{gei}$  values are estimated with conditioning on the effects of the other genotypes and  $\beta_{gi}$  values from the multivariable linear model. The  $P$  values given for the individual  $\beta_{gei}$  values are calculated by  $t$  test.

**Electrophoretic mobility shift assays.** Nuclear protein was extracted from adipocytes with a nuclear protein extraction kit (Active Motif, 40010) following the manufacturer's instructions. Oligonucleotide probes (corresponding to the

15bp flanking a SNP site for the reference or alternate allele; Supplementary Table 21) with a biotin tag at the 5' end of the forward sequence (Integrated DNA Technologies) were incubated with human adipocyte nuclear protein and the working reagent from the Gelshift Chemiluminescent EMSA kit (Active Motif, 37341). For competitor assays, an unlabelled probe of the same sequence was added to the reaction mixture at 100 $\times$  excess. The reaction was incubated for 30 min at room temperature and then loaded on a 6% retardation gel (Thermo Fisher Scientific, EC6365BOX) that was run in 0.5 $\times$  TBE buffer. We transferred the contents of the gel to a nylon membrane and visualized signal with chemiluminescent reagent as recommended.

***cis*-eQTLs in the METSIM cohort.** We obtained subcutaneous adipose *cis*-eQTL variants identified in RNA-seq data ( $n = 335$ ) from the Finnish METSIM cohort<sup>33,34</sup>.

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

### Data availability

The ATAC-seq data for primary human preadipocytes and adipocytes (untreated and lipid-challenged cells) and the pChI-C data for primary human adipocytes under lipid-challenge conditions have been deposited in the Gene Expression Omnibus under accession GSE129574 and are available upon request from the corresponding author.

Received: 29 November 2018; Accepted: 30 April 2019;

Published online: 14 June 2019

### References

- Joseph, P. G., Pare, G. & Anand, S. S. Exploring gene-environment relationships in cardiovascular disease. *Can. J. Cardiol.* **29**, 37–45 (2013).
- Hetanza, Y. & Qi, L. Gene-diet interaction and precision nutrition in obesity. *Int. J. Mol. Sci.* **18**, E787 (2017).
- Kilpeläinen, T. O. et al. Physical activity attenuates the influence of *FTO* variants on obesity risk: a meta-analysis of 218,166 adults and 19,268 children. *PLoS Med.* **8**, e100116 (2011).
- Sudlow, C. et al. UK Biobank: an open access resource for identifying the causes of a wide range of complex diseases of middle and old age. *PLoS Med.* **12**, e1001779 (2015).
- Buenrostro, J. D., Giresi, P. G., Zaba, L. C., Chang, H. Y. & Greenleaf, W. J. Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. *Nat. Methods* **10**, 1213–1218 (2013).
- Heinz, S. et al. Simple combinations of lineage-determining transcription factors prime *cis*-regulatory elements required for macrophage and B cell identities. *Mol. Cell* **38**, 576–589 (2010).
- Lefterova, M. I. & Lazar, M. A. New developments in adipogenesis. *Trends Endocrinol. Metab.* **20**, 107–114 (2009).
- Ernst, J. & Kellis, M. Large-scale imputation of epigenomic datasets for systematic annotation of diverse human tissues. *Nat. Biotechnol.* **33**, 364–376 (2015).
- Surakka, I. et al. The impact of low-frequency and rare variants on lipid levels. *Nat. Genet.* **47**, 589–597 (2015).
- Teslovich, T. M. et al. Biological, clinical and population relevance of 95 loci for blood lipids. *Nature* **466**, 707–713 (2010).
- Willer, C. J. et al. Discovery and refinement of loci associated with lipid levels. *Nat. Genet.* **45**, 1274–1283 (2013).
- Kanai, M. et al. Genetic analysis of quantitative traits in the Japanese population links cell types to complex human diseases. *Nat. Genet.* **50**, 390–400 (2018).
- MacArthur, J. et al. The new NHGRI-EBI Catalog of published genome-wide association studies (GWAS Catalog). *Nucleic Acids Res.* **45**, D896–D901 (2017).
- Mifsud, B. et al. Mapping long-range promoter contacts in human cells with high-resolution capture Hi-C. *Nat. Genet.* **47**, 598–606 (2015).
- Pan, D. Z. et al. Integration of human adipocyte chromosomal interactions with adipose gene expression prioritizes obesity-related genes from GWAS. *Nat. Commun.* **9**, 1512 (2018).
- Liu, X. et al. Functional architectures of local and distal regulation of gene expression in multiple human tissues. *Am. J. Hum. Genet.* **100**, 605–616 (2017).
- Wang, J., Duncan, D., Shi, Z. & Zhang, B. WEB-based GENE SeT Analysis Toolkit (WebGestalt): update 2013. *Nucleic Acids Res.* **41**, W77–W83 (2013).
- Siepel, A. et al. Evolutionarily conserved elements in vertebrate, insect, worm, and yeast genomes. *Genome Res.* **15**, 1034–1050 (2005).
- Lek, M. et al. Analysis of protein-coding genetic variation in 60,706 humans. *Nature* **536**, 285–291 (2016).
- Kathiresan, S. et al. Common variants at 30 loci contribute to polygenic dyslipidemia. *Nat. Genet.* **41**, 56–65 (2009).

21. Shin, S.-Y. et al. An atlas of genetic influences on human blood metabolites. *Nat. Genet.* **46**, 543–550 (2014).
22. Kettunen, J. et al. Genome-wide study for circulating metabolites identifies 62 loci and reveals novel systemic effects of LPA. *Nat. Commun.* **7**, 11122 (2016).
23. Vaitinen, M. et al. FADS2 genotype regulates delta-6 desaturase activity and inflammation in human adipose tissue. *J. Lipid Res.* **57**, 56–65 (2016).
24. Finucane, H. K. et al. Partitioning heritability by functional annotation using genome-wide association summary statistics. *Nat. Genet.* **47**, 1228–1235 (2015).
25. Nettleton, J. A., Brouwer, I. A., Geleijnse, J. M. & Hornstra, G. Saturated fat consumption and risk of coronary heart disease and ischemic stroke: a science update. *Ann. Nutr. Metab.* **70**, 26–33 (2017).
26. Casas-Agustench, P. et al. Saturated fat intake modulates the association between an obesity genetic risk score and body mass index in two US populations. *J. Acad. Nutr. Diet.* **114**, 1954–1966 (2014).
27. Luukkonen, P. K. et al. Saturated fat is more metabolically harmful for the human liver than unsaturated fat or simple sugars. *Diabetes Care* **41**, 1732–1739 (2018).
28. Bycroft, C. et al. The UK Biobank resource with deep phenotyping and genomic data. *Nature* **562**, 203–209 (2018).
29. Robinson, M. R. et al. Genotype-covariate interaction effects and the heritability of adult body mass index. *Nat. Genet.* **49**, 1174–1181 (2017).
30. Zhou, J. & Troyanskaya, O. G. Predicting effects of noncoding variants with deep learning-based sequence model. *Nat. Methods* **12**, 931–934 (2015).
31. Laakso, M. et al. METabolic Syndrome In Men (METSIM) Study: a resource for studies of metabolic and cardiovascular diseases. *J. Lipid Res.* **58**, 481–493 (2017).
32. Mancuso, N. et al. Integrating gene expression with summary association statistics to identify genes associated with 30 complex traits. *Am. J. Hum. Genet.* **100**, 473–487 (2017).
33. Gusev, A. et al. Integrative approaches for large-scale transcriptome-wide association studies. *Nat. Genet.* **48**, 245–252 (2016).
34. Ardlie, K. G. et al. The Genotype-Tissue Expression (GTEx) pilot analysis: multitissue gene regulation in humans. *Science* **348**, 648–660 (2015).
35. Ehrlund, A. et al. The cell-type specific transcriptome in human adipose tissue and influence of obesity on adipocyte progenitors. *Sci. Data* **4**, 170164 (2017).
36. Acosta, J. R. et al. Single cell transcriptomics suggest that human adipocyte progenitor cells constitute a homogeneous cell population. *Stem Cell Res. Ther.* **8**, 250 (2017).
37. Zheng, J., Chen, S., Chen, Y., Zhu, M. & Hong, D. A novel mutation in the PDZ-like motif of ZASP causes distal ZASP-related myofibrillar myopathy. *Neuropathology* **37**, 45–51 (2017).
38. Griggs, R. et al. Zaspopathy in a large classic late-onset distal myopathy family. *Brain* **130**, 1477–1484 (2007).
39. Furnagalli, M. et al. Greenlandic Inuit show genetic signatures of diet and climate adaptation. *Science* **349**, 1343–1347 (2015).
40. Lu, Y. et al. Dietary n-3 and n-6 polyunsaturated fatty acid intake interacts with FADS1 genetic variation to affect total and HDL-cholesterol concentrations in the Doetinchem Cohort Study. *Am. J. Clin. Nutr.* **92**, 258–265 (2010).
41. He, Z. et al. FADS1–FADS2 genetic polymorphisms are associated with fatty acid metabolism through changes in DNA methylation and gene expression. *Clin. Epigenetics* **10**, 113 (2018).
42. Vernekar, M. & Amarapurkar, D. Diet-gene interplay: an insight into the association of diet and FADS gene polymorphisms. *J. Nutr. Food Sci.* **6**, 503 (2016).
43. Dunham, I. et al. An integrated encyclopedia of DNA elements in the human genome. *Nature* **489**, 57–74 (2012).
44. Bujold, D. et al. The International Human Epigenome Consortium Data Portal. *Cell Syst.* **3**, 496–499 (2016).
45. Roadmap Epigenomics Consortium. Integrative analysis of 111 reference human epigenomes. *Nature* **518**, 317–330 (2015).
46. Bernstein, B. E. et al. The NIH Roadmap Epigenomics Mapping Consortium. *Nat. Biotechnol.* **28**, 1045–1048 (2010).
47. Simeonov, D. R. et al. Discovery of stimulation-responsive immune enhancers with CRISPR activation. *Nature* **549**, 111–115 (2017).
48. Gate, R. E. et al. Genetic determinants of co-accessible chromatin regions in activated T cells across humans. *Nat. Genet.* **50**, 1140–1150 (2018).
49. Phan, A. T., Goldrath, A. W. & Glass, C. K. Metabolic and epigenetic coordination of T cell and macrophage immunity. *Immunity* **46**, 714–729 (2017).
50. Schneider, C. A., Rasband, W. S. & Eliceiri, K. W. NIH Image to ImageJ: 25 years of image analysis. *Nat. Methods* **9**, 671–675 (2012).
51. Corces, M. R. et al. An improved ATAC-seq protocol reduces background and enables interrogation of frozen tissues. *Nat. Methods* **14**, 959–962 (2017).
52. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9**, 357–359 (2012).
53. Li, H. et al. The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078–2079 (2009).
54. Zhang, Y. et al. Model-based analysis of ChIP-Seq (MACS). *Genome Biol.* **9**, R137 (2008).
55. Hansen, K. D., Irizarry, R. A. & Wu, Z. Removing technical variability in RNA-seq data using conditional quantile normalization. *Biostatistics* **13**, 204–216 (2012).
56. Robinson, M. D., McCarthy, D. J. & Smyth, G. K. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **26**, 139–140 (2010).
57. Rao, S. S. P. et al. A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. *Cell* **159**, 1665–1680 (2014).
58. Nagano, T. et al. Comparison of Hi-C results using in-solution versus in-nucleus ligation. *Genome Biol.* **16**, 175 (2015).
59. Wingett, S. W. et al. HiCUP: pipeline for mapping and processing Hi-C data. *Bioinformatics* **31**, 1310 (2015).
60. Cairns, J. et al. CHICAGO: robust detection of DNA looping interactions in Capture Hi-C data. *Genome Biol.* **17**, 127 (2016).

#### Acknowledgements

This research has been conducted using the UK Biobank Resource under application number 33934. We thank the individuals who participated in the METSIM and UK Biobank studies. We also thank the UNGC sequencing core at UCLA for performing the DNA and RNA sequencing. This study was funded by National Institutes of Health (NIH) grants HL-095056, HL-28481 and U01DK105561. K.M.G. was supported by NIH-NHLBI grant 1F31HL142180, M.A. was supported by an HHMI Gilliam grant, D.Z.P. was supported by NIH-NCI grant T32LM012424 and NIH-NIDDK grant F31DK118865, and A.K. was supported by NIH grant F31HL127921. The funders had no role in study design, data collection and analysis, decision to publish or preparation of the article.

#### Author contributions

K.M.G. and P.P. designed the study. K.M.G., D.Z.P., Z.M., J.R.P., C.J.Y., J.S.S. and P.P. performed methods development and statistical analysis. K.M.G., D.Z.P., Z.M., M.A. and C.R.R. performed computational analysis of the data. K.M.G., Y.V.B., M.A., C.C. and J.N.B. performed the experiments. M.L., K.M. and P.P. produced the METSIM RNA-seq data. A.K. performed quality control of the METSIM RNA-seq data. K.M.G. and P.P. wrote the manuscript and all authors read, reviewed and/or edited the manuscript.

#### Competing interests

The authors declare no competing interests.

#### Additional information

Supplementary information is available for this paper at <https://doi.org/10.1038/s42255-019-0071-6>.

Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints).

Correspondence and requests for materials should be addressed to P.P.

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

© The Author(s), under exclusive licence to Springer Nature Limited 2019

## Reporting Summary

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

### Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a Confirmed

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

*Our web collection on [statistics for biologists](#) contains articles on many of the points above.*

### Software and code

Policy information about [availability of computer code](#)

Data collection

N/A

Data analysis

As described in the methods, ATAC-seq reads were aligned using Bowtie2 v2.2.9. Peaks were then called using the MACS2 v2.1.1 pipeline. Transcription factor motif enrichment in the ATAC-seq peaks was performed using HOMER v4.9. Promoter capture Hi-C chromosomal interactions were called using publicly available software HICUP v0.5.9 and CHICAGO v1.2.0. Differential interaction analysis was performed using edgeR. To estimate heritability for lipid traits, partitioned LD score regression (LDSC v1.0.0) was used. All other statistical analyses were performed using R.

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

### Data

Policy information about [availability of data](#)

All manuscripts must include a [data availability statement](#). This statement should provide the following information, where applicable:

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

The human primary preadipocyte and adipocyte (untreated and lipid-challenged) ATAC-seq data; and the human primary adipocyte promoter Capture Hi-C data in the lipid-challenged conditions are available at GEO (Accession ID GSE129574) and upon request from the corresponding author.

## Field-specific reporting

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

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

For a reference copy of the document with all sections, see [nature.com/documents/nr-reporting-summary-flat.pdf](https://www.nature.com/documents/nr-reporting-summary-flat.pdf)

## Life sciences study design

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

Sample size	For ATAC-seq, three biological replicates were performed for the human primary preadipocytes and human primary adipocytes, and for the human primary adipocytes in lipid-challenge conditions. For promoter Capture Hi-C, two biological replicates were performed for each lipid-challenge condition. The gene-environment interaction analysis was conducted using the UK Biobank Resource. The UK Biobank consists of ~500,000 individuals with genotypes and phenotypes. The maximum number of individuals available was included to increase the power of discover of GxE interactions.
Data exclusions	For the ATAC-seq, unpaired mapped reads, reads with MAPQ < 30, and duplicated reads were filtered out. Only reads from the autosomes and the X chromosome were retained for downstream analyses. Peaks called using MACS2 were filtered for FDR < 0.05 and those in blacklisted regions. Peaks with less than 1 count per million in at least 3 libraries were also discarded. To reduce genetic heterogeneity in the UK Biobank analysis, related, non-Caucasian individuals were excluded.
Replication	N/A
Randomization	N/A. This is an observational study, so no randomization was performed.
Blinding	N/A

## Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems		Methods	
n/a	Included in the study	n/a	Included in the study
<input checked="" type="checkbox"/>	<input type="checkbox"/> Antibodies	<input checked="" type="checkbox"/>	<input type="checkbox"/> ChIP-seq
<input type="checkbox"/>	<input checked="" type="checkbox"/> Eukaryotic cell lines	<input checked="" type="checkbox"/>	<input type="checkbox"/> Flow cytometry
<input checked="" type="checkbox"/>	<input type="checkbox"/> Palaeontology	<input checked="" type="checkbox"/>	<input type="checkbox"/> MRI-based neuroimaging
<input checked="" type="checkbox"/>	<input type="checkbox"/> Animals and other organisms		
<input type="checkbox"/>	<input checked="" type="checkbox"/> Human research participants		
<input checked="" type="checkbox"/>	<input type="checkbox"/> Clinical data		

## Eukaryotic cell lines

Policy information about [cell lines](#)

Cell line source(s)	PromoCell C-12731, lot 3952024
Authentication	The human primary preadipocytes were obtained from the established and reputable commercial vendor, PromoCell, and passed standard viability and differentiation capacity requirements prior to being made available.
Mycoplasma contamination	The human primary preadipocytes were tested by PromoCell for common microbiological contaminants and infectious viruses, including Mycoplasma, and found to produce negative results prior to the cells being made available. As these are primary cells, they have not been passaged extensively, and are unlikely to survive Mycoplasma contamination. Furthermore, we observed no evidence of contamination during the culture period.
Commonly misidentified lines (See <a href="#">ICLAC</a> register)	N/A

## Human research participants

Policy information about [studies involving human research participants](#)

Population characteristics	This research was conducted using the UK Biobank Resource. The UK Biobank consists of ~500,000 individuals with genotypes and phenotypes and includes males and females from a broad range of ages. The maximum number of individuals was included to increase the power of discover of GxE interactions.
Recruitment	N/A
Ethics oversight	UK Biobank

Note that full information on the approval of the study protocol must also be provided in the manuscript.

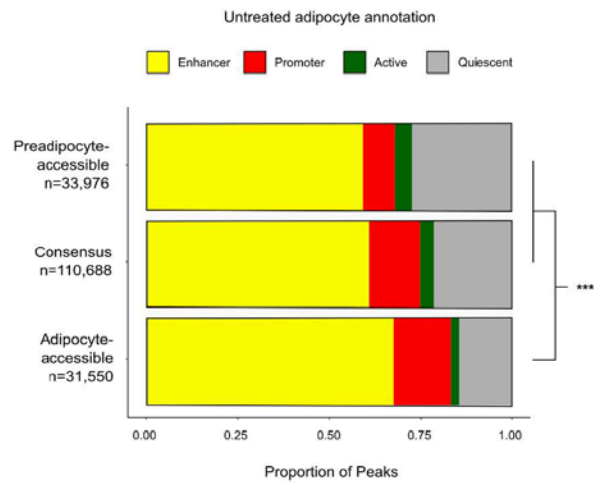


In the format provided by the authors and unedited.

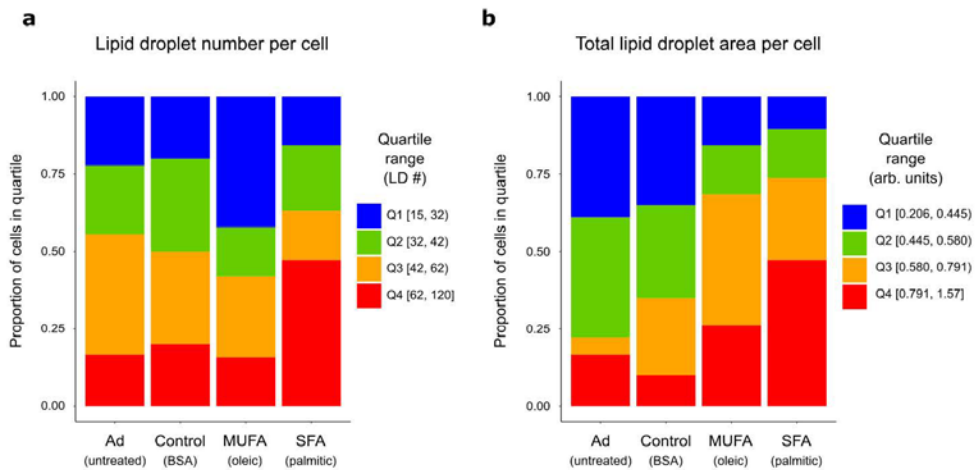
## Reverse gene–environment interaction approach to identify variants influencing body-mass index in humans

Kristina M. Garske<sup>1</sup>, David Z. Pan<sup>1,2</sup>, Zong Miao<sup>1,2</sup>, Yash V. Bhagat<sup>1</sup>, Caroline Comenho<sup>1</sup>, Christopher R. Robles<sup>3</sup>, Jihane N. Benhammou<sup>1,4</sup>, Marcus Alvarez<sup>1</sup>, Arthur Ko<sup>5</sup>, Chun Jimmie Ye<sup>6</sup>, Joseph R. Pisegna<sup>1,4</sup>, Karen L. Mohlke<sup>7</sup>, Janet S. Sinsheimer<sup>1,8</sup>, Markku Laakso<sup>9</sup> and Päivi Pajukanta<sup>1,2,9\*</sup>

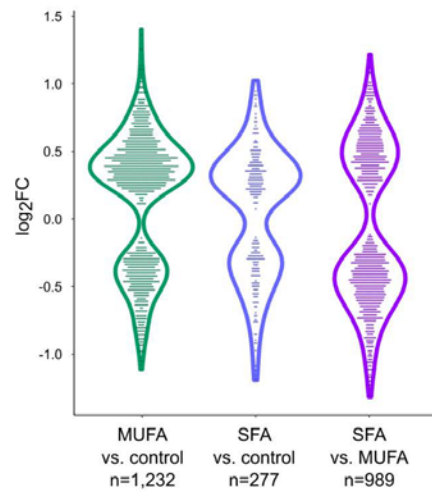
<sup>1</sup>Department of Human Genetics, David Geffen School of Medicine at UCLA, Los Angeles, CA, USA. <sup>2</sup>Bioinformatics Interdepartmental Program, UCLA, Los Angeles, CA, USA. <sup>3</sup>Department of Computer Science, UCLA, Los Angeles, CA, USA. <sup>4</sup>Vache and Tamar Manoukian Division of Digestive Diseases, UCLA, Los Angeles, CA, USA. <sup>5</sup>Department of Medicine, David Geffen School of Medicine at UCLA, Los Angeles, CA, USA. <sup>6</sup>Institute for Human Genetics, Department of Epidemiology and Biostatistics and Department of Bioengineering and Therapeutic Sciences, UCSF, San Francisco, CA, USA. <sup>7</sup>Department of Genetics, University of North Carolina, Chapel Hill, NC, USA. <sup>8</sup>Department of Biomathematics, David Geffen School of Medicine at UCLA, Los Angeles, CA, USA. <sup>9</sup>Internal Medicine, Institute of Clinical Medicine, University of Eastern Finland and Kuopio University Hospital, Kuopio, Finland.  
\*e-mail: ppajukanta@mednet.ucla.edu



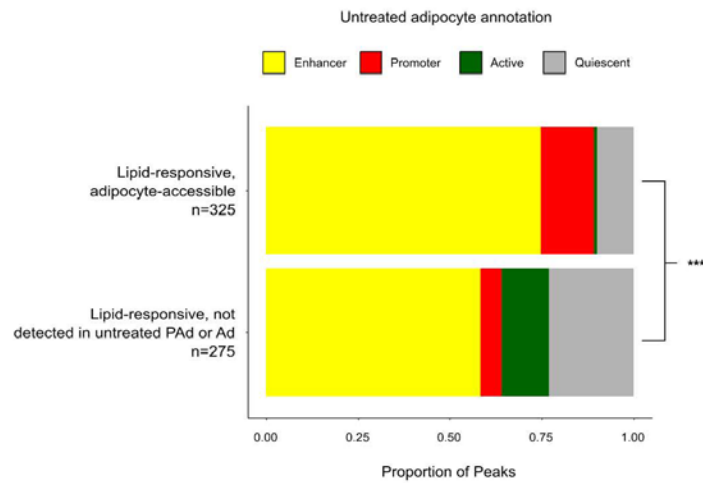
**Supplementary Figure 1. Adipocyte-accessible peaks fall more into adipocyte enhancers and promoters than the preadipocyte-accessible peaks or the full peak set.** ATAC-seq peaks from the indicated peak sets on the y-axis are distributed among four subsets of functional annotations from the 25-state imputed chromHMM<sup>1</sup> annotations from mesenchymal stem cell derived cultured adipocytes. Note that not all peaks were categorized into one of these 4 categories due to minimum peak proportion overlap (>50%) requirement not being met. \*\*\*depicts the *p*-value ( $p < 1 \times 10^{-5}$ ) for the chi-square test for independence between the distributions of peaks in the indicated annotations. Related to Figure 1.



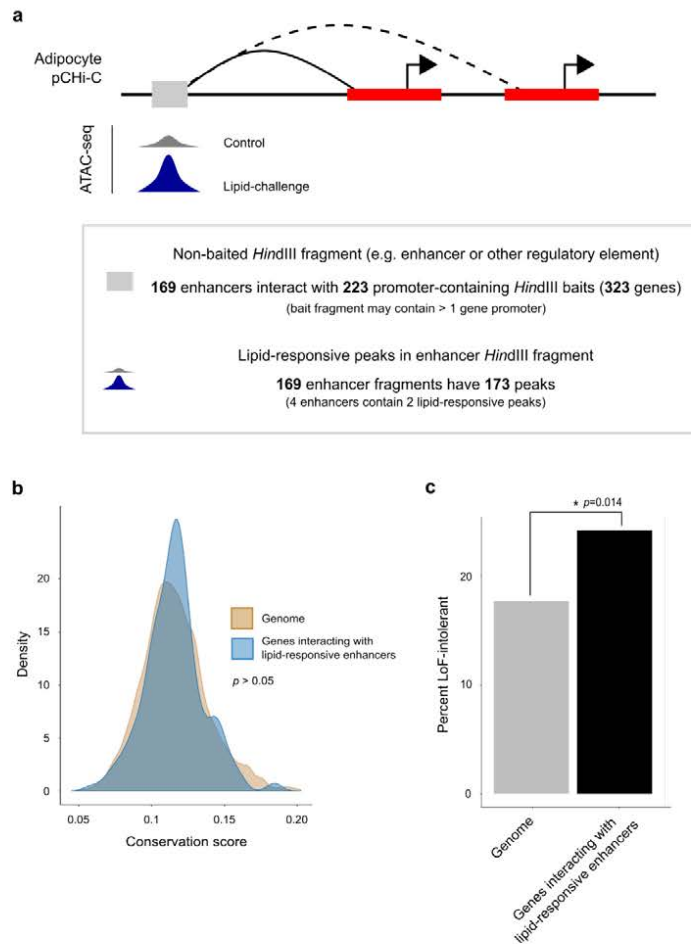
**Supplementary Figure 2. Fatty acid lipid challenge in human adipocytes leads to increased storage of lipids in lipid droplets.** (a,b) The proportion of cells in each of the indicated quartiles are reported for (a) lipid droplet (LD) number per cell, and (b) total LD area per cell, quantified from oil red o staining. Treatment with monounsaturated fatty acid (MUFA) leads to increased total area of LD but fewer total LDs (e.g. large LDs). Treatment with saturated fatty acid (SFA) leads to increased LD number and size. Data presented are from one representative experiment out of two independent experiments with similar results. Related to Figure 2.



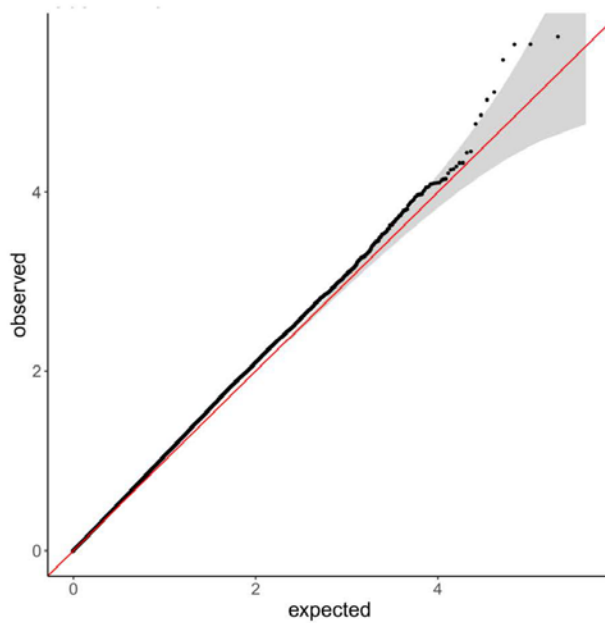
**Supplementary Figure 3. Violin plots show the distribution of log<sub>2</sub> fold-change (log<sub>2</sub>FC) for all differentially accessible peaks from the lipid challenge in adipocytes.** Peaks were considered differentially accessible at a cutoff of FDR < 0.05. FDR was calculated (adjusting for n=122,252 ATAC-seq peaks) from the *p*-values of the QL F-test (see Methods) in the one-way ANOVA. For the *post hoc* test to determine which comparison was significant after the one-way ANOVA (OA vs. BSA, PA vs. BSA, or OA vs. PA), we determined the least significant difference. The violin plot characteristics are as follows: MUFA vs. ctrl (n=1,232) range: -1.11 – 1.40; median: 0.32; 25<sup>th</sup> percentile: -0.32; and 75<sup>th</sup> percentile: 0.47. SFA vs. ctrl (n=277) range: -1.19 – 1.02; median: 0.22; 25<sup>th</sup> percentile: -0.30; and 75<sup>th</sup> percentile: 0.37. SFA vs. MUFA (n=989) range: -1.31 – 1.21; median: -0.27; 25<sup>th</sup> percentile: -0.51; and 75<sup>th</sup> percentile: 0.47. MUFA indicates monounsaturated fatty acid; SFA, saturated fatty acid. Related to Figure 2.



**Supplementary Figure 4. Lipid-responsive peaks in adipocyte-accessible regions fall more into adipocyte enhancers and promoters than lipid-responsive peaks in context-dependent regions.** Lipid-responsive ATAC-seq peaks from the indicated peak sets on the y-axis are distributed among four subsets of functional annotations from the 25-state imputed chromHMM<sup>1</sup> annotations from mesenchymal stem cell derived culture adipocytes. Note that not all peaks were categorized into one of these 4 categories due to minimum peak proportion overlap (>50%) requirement not being met. \*\*\*depicts the *p*-value ( $p < 1 \times 10^{-5}$ ) for the chi-square test for independence between the distributions of peaks in the indicated annotations. Related to Figure 2.



**Supplementary Figure 5. The 323 genes with promoters that interact with lipid-responsive enhancers exhibit constraints on loss-of-function mutations.** (a) Schematic overview of the lipid-responsive sites in non-baited *Hind*III fragments from the adipocyte pChI-C interactions. These data were integrated to identify the 323 gene promoters that interact with lipid-responsive enhancers in adipocytes. (b) Density plot shows the distribution of per-gene average conservation scores across placental mammals<sup>2</sup> for all protein-coding genes in the genome compared to all protein-coding genes in the set of 323 genes whose promoters interact with lipid-responsive enhancers. The two-sided Wilcoxon signed-rank test returned a non-significant  $p$ -value  $> 0.05$ . (c) Bar graph shows the proportion of protein-coding genes that are loss-of-function intolerant (i.e. are unlikely to have protein-truncating variants in humans)<sup>3</sup> in the whole genome ( $n=3,204/18,122$ ; 17.7%) compared to the protein-coding genes among the 323 genes ( $n=50/207$ ; 24.2%). LoF indicates loss-of-function; \*depicts the  $p$ -value for the hypergeometric enrichment test. Compare with Figure 3 results for 154 genes with lipid-responsive promoters.



**Supplementary Figure 6. Testing all SNPs genome-wide for gene-by-saturated fat intake effect on BMI does not show inflation or result in significant GxEs at the genome-wide significance threshold.** We tested the SNPs that are not in the same LD block ( $r^2 < 0.2$ ) genome-wide for a GxE between each SNP and saturated fat intake effect on BMI. There were a total of 211,187 SNPs and 167,908 individuals with no missing data available for study (see Equation 1 in the Methods). The Q-Q plot shows the observed  $p$ -values of  $\beta_{GxE}$ , the expected  $p$ -values (red line) based on the multiple testing, and the 95% confidence interval (shaded area).

**Supplementary Table 1. Sequencing, read processing, and QC metrics for untreated preadipocyte and adipocyte ATAC-seq.**

	Reads	Uniquely aligned	Fraction uniquely aligned	Paired and filtered	De-duplicated	Fraction duplicates	Final Reads	Fraction mtDNA	Fraction reads in peaks
PAd rep1	23,240,008	16,960,464	0.73	20,403,898	16,707,418	0.18	15,810,810	0.019	0.48
PAd rep2	25,705,628	18,552,651	0.72	22,516,812	17,829,509	0.21	16,938,123	0.016	0.46
PAd rep3	27,548,038	19,008,502	0.69	23,815,866	19,142,102	0.20	17,984,628	0.025	0.47
Ad rep1	24,407,273	16,893,473	0.69	21,049,924	16,045,835	0.24	15,038,768	0.033	0.58
Ad rep2	18,190,931	12,109,272	0.67	15,537,582	11,344,211	0.27	10,459,715	0.049	0.53
Ad rep3	21,165,864	13,976,310	0.66	17,956,825	12,512,132	0.30	11,528,741	0.050	0.70

PAd indicates preadipocyte; Ad, adipocyte; mtDNA, mitochondrial DNA.



**Supplementary Table 2. Differentially accessible ATAC-seq peaks between human preadipocytes and adipocytes.** Peaks were considered differentially accessible at a cutoff of  $FDR < 0.05$ . FDR was calculated (adjusting for  $n = 154,647$  ATAC-seq peaks) from the  $P$  values of the QL  $F$  test for differential accessibility between preadipocytes and adipocytes by using ATAC-seq libraries from three replicates per cell type. Related to Fig. 1. See supplementary materials.

**Supplementary Table 3. Sequencing, read processing, and QC metrics for adipocyte lipid-challenge ATAC-seq.**

	Reads	Uniquely aligned	Fraction uniquely aligned	Paired and filtered	De-duplicated	Fraction duplicates	Final Reads	Fraction mtDNA	Fraction reads in peaks
BSA rep1	64,706,941	48,768,335	0.76	56,012,019	45,623,026	0.185	44,927,691	0.012	0.66
BSA rep2	24,133,180	18,644,383	0.77	21,149,557	17,972,138	0.150	17,655,287	0.015	0.66
BSA rep3	23,981,457	17,668,336	0.74	20,677,446	17,553,162	0.151	17,214,846	0.016	0.59
OA rep1	46,775,100	35,874,412	0.77	40,975,870	29,562,701	0.279	29,007,522	0.016	0.70
OA rep2	57,372,688	44,971,629	0.79	51,073,123	39,246,428	0.232	38,742,718	0.010	0.71
OA rep3	33,118,575	25,516,115	0.77	29,143,404	23,401,224	0.197	22,995,249	0.014	0.67
PA rep1	27,688,462	21,034,204	0.76	24,089,236	18,199,958	0.244	17,822,393	0.018	0.64
PA rep2	45,904,086	35,705,224	0.78	40,386,681	31,318,153	0.225	30,874,420	0.011	0.68
PA rep3	39,159,661	29,810,602	0.76	34,234,012	26,610,078	0.223	26,074,118	0.017	0.68

BSA indicates bovine serum albumin; OA, oleic acid; PA, palmitic acid; mtDNA, mitochondrial DNA.











**Supplementary Table 4. Differentially accessible ATAC-seq peaks in lipid-challenged human adipocytes.** The table lists the significant differential ATAC-seq peaks in human primary adipocytes that were treated with saturated (palmitic) or monounsaturated (oleic) fatty acids or vehicle (BSA) control. Peaks were considered differentially accessible at a cutoff of  $FDR < 0.05$ . FDR was calculated (adjusting for  $n = 122,252$  ATAC-seq peaks) from the  $P$  values of the QL  $F$  test in one-way ANOVA. For the post hoc test to determine which comparison was significant after one-way ANOVA (OA vs. BSA, PA vs. BSA or OA vs. PA), we determined the least significant difference. Related to Fig. 2. See supplementary materials.

**Supplementary Table 5. Sequencing and read processing metrics for adipocyte lipid-challenge pCHI-C.**

	Reads	Uniquely mapped and paired	Unique ditags	Cis-close	Cis-far	Trans
BSA rep1	156,781,294	115,665,713	69,664,503	10,560,429	49,905,501	9,198,573
BSA rep2	126,772,704	94,717,631	51,631,011	8,335,003	36,631,083	6,664,925
OA rep1	120,985,559	90,005,610	54,155,183	7,477,013	38,909,160	7,769,010
OA rep2	118,632,574	89,101,812	49,035,437	8,495,017	34,445,736	6,094,684
PA rep1	132,242,479	98,448,728	54,891,280	8,392,648	39,146,001	7,352,631
PA rep2	107,001,633	80,596,383	44,405,325	6,636,844	31,576,330	6,192,151

BSA indicates bovine serum albumin; OA, oleic acid; PA, palmitic acid.

Supplementary Table 6. The top 10 TF motifs enriched in adipocyte lipid-responsive open chromatin regions in chromosomal interactions.

Motif logo	Motif name	<i>p</i> -value	Adjusted <i>p</i> -value	Number of target sequences with motif (of 264)	Percent of target sequences with motif	Number of background sequences with motif (of 30,704)	Percent of background sequences with motif
	RUNX-AML(Runt)/CD4+-PollI-ChIP-Seq(Barski_et_al)/Homer	1.0x10 <sup>-8</sup>	0	134.0	50.76%	10267.5	33.44%
	STAT4(Stat)/CD4-Stat4-ChIP-Seq(GSE22104)/Homer	1.0x10 <sup>-8</sup>	0	166.0	62.88%	13952.6	45.44%
	PPARE(NR),DR1/3T3L1-Pparg-ChIP-Seq(GSE13511)/Homer	1.0x10 <sup>-6</sup>	0	162.0	61.36%	14012.1	45.64%
	RXR(NR),DR1/3T3L1-RXR-ChIP-Seq(GSE13511)/Homer	1.0x10 <sup>-6</sup>	0	173.0	65.53%	15537.1	50.60%
	ZBTB18(Zf)/HEK293-ZBTB18.GFP-ChIP-Seq(GSE58341)/Homer	1.0x10 <sup>-6</sup>	1.0x10 <sup>-4</sup>	108.0	40.91%	8317.5	27.09%
	THRa(NR)/C17.2-THRa-ChIP-Seq(GSE38347)/Homer	1.0x10 <sup>-5</sup>	1.0x10 <sup>-4</sup>	105.0	39.77%	8090.0	26.35%
	KLF10(Zf)/HEK293-KLF10.GFP-ChIP-Seq(GSE58341)/Homer	1.0x10 <sup>-5</sup>	1.0x10 <sup>-4</sup>	108.0	40.91%	8567.6	27.90%
	Srebp1a(bHLH)/HepG2-Srebp1a-ChIP-Seq(GSE31477)/Homer	1.0x10 <sup>-5</sup>	1.0x10 <sup>-4</sup>	59.0	22.35%	3792.7	12.35%
	NF1-halfsite(CTF)/LNCaP-NF1-ChIP-Seq(Unpublished)/Homer	1.0x10 <sup>-4</sup>	3.0x10 <sup>-4</sup>	215.0	81.44%	21406.9	69.72%
	NPAS2(bHLH)/Liver-NPAS2-ChIP-Seq(GSE39860)/Homer	1.0x10 <sup>-4</sup>	3.0x10 <sup>-4</sup>	152.0	57.58%	13632.6	44.40%

Enrichment *p*-values were derived from the hypergeometric enrichment test of proportion of the given TF motif in the peak set [lipid-responsive open chromatin regions in adipocyte chromosomal interactions (n=264)] compared with the background set of peaks [all non-lipid-responsive peaks in adipocyte chromosomal interactions (n=30,704)], adjusted (Benjamini-Hochberg) for the number of known motifs tested (n=364)<sup>4</sup>. The top 10 enriched TF motifs include key TFs in lipid metabolism, such as the co-factors PPAR and RXR. Related to Figure 3.

**Supplementary Table 7. 154 genes with lipid-responsive promoters in chromosomal interactions in adipocytes.** The table lists the Ensembl ID and gene symbol for genes with promoters in interactions in adipocyte promoter-capture Hi-C that also had lipid-responsive ATAC-seq peaks. Related to Fig. 3. See supplementary materials.

Supplementary Table 8. KEGG pathway enrichment analysis of 154 genes with lipid-responsive promoters.

KEGG pathway	Ratio of enrichment	Number of genes	Genes in pathway	FDR
Glycine, serine and threonine metabolism	13.96	5	<i>AGXT2</i> <i>AOC2</i> <i>AOC3</i> <i>GLYCK</i> <i>MAOA</i>	0.0072
Phenylalanine metabolism	23.46	3	<i>AOC2</i> <i>AOC3</i> <i>MAOA</i>	0.036

The 154 genes with lipid-responsive promoters in adipocyte chromosomal interactions were tested for KEGG pathway enrichment using WebGetstalt<sup>5</sup>, using all genes that were involved in adipocyte chromosomal interactions (n=17,052) as the background set. The FDR is calculated from the *p*-values of the hypergeometric test, adjusted for the number of pathways tested through WebGestalt. Related to Figure 3.

**Supplementary Table 9. 323 gene promoters physically interact with lipid-responsive enhancers in adipocytes.** The table lists the Ensembl ID and gene symbol for genes with promoters that interact with enhancers that contained lipid-responsive ATAC-seq peaks. Related to Supplementary Fig. 5. See supplementary materials.



**Supplementary Table 10. Three lipid-responsive ATAC-peaks in interacting enhancers overlap with GWAS SNPs<sup>6</sup> for serum lipid traits.**

peakChr	peakStart	peakEnd	intBaitGene <sup>f</sup>	SNP in peak	MAF <sup>g</sup>	Associated trait <sup>h</sup>	p-value <sup>6</sup>	Index SNP <sup>6</sup>	LD with index SNP <sup>6</sup> (r <sup>2</sup> )
7	73015109	73016308	<u>Bait1:</u> <i>WBSCR22</i> <u>Bait2:</u> <i>STX1A</i>	rs34346326	0.2	TG	1.31e-44	rs17145738	0.5
15	58591111	58592050	<u>Bait1:</u> <i>ADAM10</i> <u>Bait2:</u> <i>RP11-30K9.7/</i> <i>U3.10</i>	rs12899879	0.14	HDL	3.55e-09	rs1532085	-
10	113902081	113908608	<i>ADRA2A</i>	rs2792744	0.28	TC	2.73e-09	rs2255141	0.77

<sup>f</sup>The genes listed are the promoters in the baited *HindIII* fragment with which the lipid-responsive enhancers interact. More than one bait is listed when the lipid-responsive enhancer is interacting with more than one bait in the adipocyte pChI-C; <sup>g</sup>Minor allele frequency (MAF) is the European frequency from the 1000 Genomes Project. Linkage disequilibrium (LD) is calculated based on Europeans in the 1000 Genomes Project; LD calculations > 0.2 are reported. Lipid-responsive ATAC-seq peaks that land in enhancers within adipocyte chromosomal interactions (n=173) were assessed for whether they contain GWAS SNPs for serum lipid traits from the meta-GWAS performed in Willer et al.<sup>6</sup>

Supplementary Table 11. Lipid-responsive gene promoters with GWAS SNPs respond to SFA treatment.

peakChr	peakStart	peakEnd	gene-pChi-C bait	log <sub>2</sub> FC (MUFA/ctrl)	log <sub>2</sub> FC (SFA/ctrl)	log <sub>2</sub> FC (SFA/MUFA)
11	61594652	61596828	<i>FADS2/ FADS1</i>	n.s	n.s	<b>0.37</b>
7	73036880	73038991	<i>MLXIPL</i>	n.s.	n.s	<b>0.29</b>
2	27432323	27432971	<i>SLC5A6/ ATRAID</i>	n.s	n.s	<b>0.33</b>
16	68115758	68116375	<i>NFATC3</i>	n.s	n.s	<b>0.48</b>
19	10981139	10983631	<i>CARM1</i>	n.s	<b>0.24</b>	n.s

MUFA indicates monounsaturated fatty acid; SFA, saturated fatty acid; ctrl, control; n.s., non-significant (based on the *post hoc* test of the one-way ANOVA, see below). Lipid-responsive ATAC-seq peaks that land in promoters within adipocyte chromosomal interactions (n=91) were assessed for whether they contain GWAS SNPs for serum lipid traits from the meta-GWAS performed in Willer et al.<sup>6</sup> The direction of the ATAC-seq differential accessibility effect was then assessed based on the quality (e.g. SFA or MUFA) of the fatty acid. Differential accessibility was evaluated at an FDR cutoff of 0.05. FDR was calculated (adjusting for n=122,252 ATAC-seq peaks) from the *p*-values of the QL F-test (see Methods) in the one-way ANOVA. For the *post hoc* test to determine which comparison was significant after the one-way ANOVA (MUFA vs. ctrl, SFA vs. ctrl, or SFA vs. MUFA), we determined the least significant difference. The lipid-responsive gene promoters in chromosomal interactions that contain GWAS SNPs exhibit increased accessibility in palmitic acid (saturated fatty acid) lipid challenge. Related to Table 1.

Supplementary Table 12. Lipid-responsive enhancers with GWAS SNPs stratified by quality of fatty acid.

peakChr	peakStart	peakEnd	intBaitGene <sup>†</sup>	log <sub>2</sub> FC (MUFA/ctrl)	log <sub>2</sub> FC (SFA/ctrl)	log <sub>2</sub> FC (SFA/MUFA)
7	73015109	73016308	<u>Bait1:</u> <i>WBSCR22</i> <u>Bait2:</u> <i>STX1A</i>	n.s.	<b>0.45</b>	<b>0.64</b>
15	58591111	58592050	<u>Bait1:</u> <i>ADAM10</i> <u>Bait2:</u> <i>RP11-30K9.7/</i> <i>U3.101</i>	<b>0.45</b>	n.s.	<b>-0.28</b>
10	113902081	113908608	<i>ADRA2A</i>	n.s.	<b>0.50</b>	n.s.

<sup>†</sup>The genes listed are the promoters in the baited *Hind*III fragment with which the lipid-responsive enhancers interact. More than one bait is listed when the lipid-responsive enhancer is interacting with more than one bait in the adipocyte pChI-C; MUFA indicates monounsaturated fatty acid; SFA, saturated fatty acid; ctrl, control; n.s., non-significant (based on the *post hoc* test of the one-way ANOVA, see below). Lipid-responsive ATAC-seq peaks that land in enhancers within adipocyte chromosomal interactions (n=173) were assessed for whether they contain GWAS SNPs for serum lipid traits from the meta-GWAS performed in Willer et al.<sup>6</sup> The direction of the effect was then assessed based on the quality (e.g. SFA or MUFA) of the fatty acid. Differential accessibility was evaluated at an FDR cutoff of 0.05. FDR was calculated (adjusting for n=122,252 ATAC-seq peaks) from the *p*-values of the QL F-test (see Methods) in the one-way ANOVA. For the *post hoc* test to determine which comparison was significant after the one-way ANOVA (MUFA vs. ctrl, SFA vs. ctrl, or SFA vs. MUFA), we determined the least significant difference. The lipid-responsive enhancers in chromosomal interactions that contain GWAS SNPs are more often differentially accessible in palmitic acid (saturated fatty acid) lipid challenge. Related to Supplementary Table 10.

Supplementary Table 13. LDSC analysis<sup>7</sup> of SNPs in *cis* regions of the 154 lipid-responsive promoters.

Category	Proportion of SNPs	Proportion of $h^2$	Proportion of $h^2$ SE	Enrichment	Enrichment SE	Enrichment $p$ -value
TC	0.029	0.086	0.021	2.95	0.74	0.0088
LDL-C	0.029	0.084	0.026	2.91	0.91	0.038
HDL-C	0.029	0.083	0.027	2.87	0.92	0.042
Serum TG	0.029	0.085	0.026	2.93	0.90	0.045

SE indicates standard error;  $h^2$ , heritability; TC, total cholesterol; LDL-C, low-density lipoprotein cholesterol; HDL-C, high-density lipoprotein cholesterol; TG, triglycerides. LD score regression (LDSC)<sup>7</sup> was performed using the SNPs in the *cis* regions (gene body +/- 500 kb) of the 154 genes with lipid-responsive promoters in adipocyte chromosomal interactions; and the serum lipid trait summary statistics from the meta-GWAS performed in Willer et al.<sup>6</sup> The enrichment  $p$ -value and the SE for the proportion of  $h^2$  and enrichment were calculated from block jackknife resampling used in the LDSC method. The  $p$ -value reported is not adjusted for multiple tests as these serum lipid traits are highly correlated. The *cis* regions of the 154 genes with lipid-responsive promoters in adipocyte chromosomal interactions contribute significantly to the heritability of serum lipid traits in humans.

Supplementary Table 14. LDSC analysis<sup>7</sup> of SNPs in *cis* regions of genes with lipid-responsive enhancers.

Category	Proportion of SNPs	Proportion of $h^2$	Proportion of $h^2$ SE	Enrichment	Enrichment SE	Enrichment $p$ -value
TC	0.055	0.078	0.015	1.41	0.28	0.12
LDL-C	0.055	0.057	0.0089	1.03	0.16	0.86
HDL-C	0.055	0.11	0.020	1.94	0.36	0.011
Serum TG	0.055	0.12	0.041	2.10	0.74	0.15

SE indicates standard error;  $h^2$ , heritability; TC, total cholesterol; LDL-C, low-density lipoprotein cholesterol; HDL-C, high-density lipoprotein cholesterol; TG, triglycerides. LD score regression (LDSC)<sup>7</sup> was performed using the SNPs in the *cis* regions (gene body +/- 500 kb) of the 323 genes with promoters that interact with lipid-responsive enhancers in adipocyte chromosomal interactions; and serum lipid trait summary statistics from the meta-GWAS performed in Willer et al.<sup>6</sup> The enrichment  $p$ -value and the SE for the proportion of  $h^2$  and enrichment were calculated from block jackknife resampling used in the LDSC method. The  $p$ -value reported is not adjusted for multiple tests as these serum lipid traits are highly correlated. The *cis* regions of the 323 genes that interact with lipid-responsive enhancers in adipocytes contribute significantly to the heritability of HDL, but not the other serum lipid traits, in humans.

**Supplementary Table 15. 75 lipid-responsive peaks in gene promoters contain SNPs with MAF > 0.05 in the UK Biobank.** The table lists the lipid-responsive ATAC-seq peaks within gene promoters involved in adipocyte chromosomal interactions that contain SNPs with MAF > 0.05 in the UK Biobank ( $n = 75/91$  peaks). The SNPs in these regions were tested for gene–environment interactions in the UK Biobank. Related to Fig. 3 and Table 2. See supplementary materials.

Supplementary Table 16. Significant GxE promoter SNPs with LD proxies.

SNP	<i>p</i> -g	<i>p</i> -g*e	β-g	β-g*e	Genes in Bait	<i>cis</i> -eQTL FDR <sup>§§</sup>	Target Gene <sup>§</sup>	log <sub>2</sub> FC <sup>§</sup>
rs1974817	0.0089	0.0010	2.3	-0.089	<i>GLTSCR2</i> / <i>SNORD23</i>	2.4E-31	<i>SEPW1</i>	0.73
rs2334290	0.22	0.0011	-1.65	0.088		1.2E-08	<i>SEPW1</i>	0.46
rs112438892 <sup>¶</sup>	0.0017	0.0050	-0.50	0.012	<i>CARM1</i>	2.7E-05/ 0.015	<i>SMARCA4</i> / <i>ICAM4</i>	-0.28/ 0.38
rs117569851 <sup>¶</sup>	0.0037	0.016	0.461	-0.011		0.0014/ 0.046	<i>SMARCA4</i> / <i>KRI1</i>	-0.21/ 0.18
rs35678764	0.028	0.025	-5.7	0.19	<i>RDH8</i> / <i>COL5A3</i>	0.012	<i>OLFM2</i>	-0.67
rs66516040 <sup>†</sup>	0.034	0.026	0.19	-0.0057		0.047	<i>PIN1</i>	-0.24
rs9797822	0.044	0.039	5.2	-0.17				
rs10788522	0.013	0.027	3.9	-0.14	<i>LDB3</i>	6.1E-04	<i>LDB3</i>	0.42
rs2354358	0.033	0.042	-3.5	0.14		6.1E-04	<i>LDB3</i>	0.42
rs867773	0.31	0.033	-2.0	0.11	<i>PLIN2</i>	0.048	<i>PLIN2</i>	-0.39
rs12379376	0.19	0.034	2.1	-0.11		0.048	<i>PLIN2</i>	-0.39

<sup>†</sup>The *cis*-eQTLs were identified in the adipose tissue from the METSIM cohort<sup>89</sup>. <sup>‡</sup>This SNP is the only genome-wide significant *cis*-eQTL from the set of GxE SNPs with LD  $r^2 > 0.2$  in the lipid-responsive peak. <sup>¶</sup>These GxE SNPs are *cis*-eQTLs for more than one gene. For 5 of the significant promoter GxE SNPs listed in Table 2, SNPs with LD  $r^2 > 0.2$  in the lipid-responsive region that also exhibited a significant GxE effect of saturated fat intake on BMI are listed. Redundant SNPs are listed together in order of more to less significant. The reported *p*-values are from the βs in the multi-variable linear model (see Equation 2 in the Methods), where *g* is the number of minor alleles of the genotype and *e* is saturated fat intake. Here *p*-g indicates the *p*-value for the genotype effect; *p*-g\*e, the *p*-value for the GxE effect; beta values follow the same notation. For the multi-variable linear model, there were a total of 290 SNPs and 38,394 individuals with no missing data available for study. Related to Table 2.

**Supplementary Table 17. 142 lipid-responsive peaks in enhancers contain SNPs with MAF > 0.05 in the UK Biobank.** The table lists the lipid-responsive ATAC-seq peaks within enhancers involved in adipocyte chromosomal interactions that contain SNPs with MAF > 0.05 in the UK Biobank ( $n = 142/169$  peaks). The SNPs in these regions were tested for gene–environment interactions in the UK Biobank. Related to Supplementary Fig. 5 and Supplementary Table 18. See supplementary materials.



**Supplementary Table 18. Significant G×E interactions with BMI from a multivariable linear model for 410 enhancer SNPs.** <sup>†</sup>The *cis*-eQTLs were identified in adipose tissue from the METSIM cohort; <sup>§</sup>When more than one non-independent SNP (LD  $r^2 > 0.2$ ) has a significant G×E *P* value for the lipid-responsive region, both SNPs are listed together in order of more to less significant. <sup>‡</sup>Genes in separate promoter-containing baits are marked when a lipid-responsive enhancer with a G×E SNP is interacting with more than one bait in adipocyte pChI-C. The reported *P* values are from the multivariable linear model (see equation (2) in the Methods), where *g* is the number of minor alleles of the genotype and *e* is saturated fat intake. Here *p*-*g* indicates the *P* value for the genotype effect and *p*-*g*\**e* indicates the *P* value for the G×E effect; beta values follow the same notation. For the multivariable linear model, there were a total of 410 SNPs and 18,318 individuals with no missing data available for study. See supplementary materials.

**Supplementary Table 19. DeepSEA analysis of the 20 G×E SNPs in interacting lipid-responsive gene promoters.** The table lists the predicted functional impact of promoter G×E SNPs on chromatin features such as transcription factor binding and histone marks. Related to Table 2. See supplementary materials.

**Supplementary Table 20. DeepSEA analysis of the 26 G×E SNPs in interacting lipid-responsive enhancers.** The table lists the predicted functional impact of enhancer G×E SNPs on chromatin features such as transcription factor binding and histone marks. Related to Supplementary Table 18. See supplementary materials.

**Supplementary Table 21. EMSA oligo probes used for analysis of GxE SNP rs10788522.**

<b>Probe Name</b>	<b>Probe Sequence</b>
rs10788522 FWD labeled	biotin - 5'- TCTGGGGAGAGGAAG <u>G/A</u> GGGACAGGCTGAGAC - 3'
rs10788522 FWD unlabeled	5'- TCTGGGGAGAGGAAG <u>G/A</u> GGGACAGGCTGAGAC - 3'
rs10788522 REV unlabeled	5' - GTCTCAGCCTGTCCC <u>C/T</u> TTCCTCTCCCCAGA - 3'

Oligonucleotides were designed to target the GxE SNP rs10788522 in the *LDB3* promoter *HindIII* fragment (+/- 15 bp).

### Supplementary References

1. Ernst, J. & Kellis, M. Large-scale imputation of epigenomic datasets for systematic annotation of diverse human tissues. *Nat. Biotechnol.* **33**, 364–76 (2015).
2. Siepel, A. *et al.* Evolutionarily conserved elements in vertebrate, insect, worm, and yeast genomes. *Genome Res.* **15**, 1034–1050 (2005).
3. Lek, M. *et al.* Analysis of protein-coding genetic variation in 60,706 humans. *Nature* **536**, 285–291 (2016).
4. Heinz, S. *et al.* Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. *Mol. Cell* **38**, 576–589 (2010).
5. Wang, J., Duncan, D., Shi, Z. & Zhang, B. WEB-based GEne SeT AnaLysis Toolkit (WebGestalt): update 2013. *Nucleic Acids Res.* **41**, W77–W83 (2013).
6. Willer, C. J. *et al.* Discovery and refinement of loci associated with lipid levels. *Nat. Genet.* **45**, 1274–1283 (2013).
7. Finucane, H. K. *et al.* Partitioning heritability by functional annotation using genome-wide association summary statistics. *Nat. Genet.* **47**, 1228–1235 (2015).
8. Pan, D. Z. *et al.* Integration of human adipocyte chromosomal interactions with adipose gene expression prioritizes obesity-related genes from GWAS. *Nat. Commun.* **9**, 1512 (2018).
9. Laakso, M. *et al.* METabolic Syndrome In Men (METSIM) Study: a resource for studies of metabolic and cardiovascular diseases. *J. Lipid Res.* **58**, 481–493 (2017).

## Chapter 4

Identification of *TBX15* as an adipose master *trans* regulator of abdominal obesity genes

## Identification of *TBX15* as an adipose master *trans* regulator of abdominal obesity genes

David Z. Pan<sup>1,2</sup>, Zong Miao<sup>1,2</sup>, Caroline Comenho<sup>1</sup>, Sandhya Rajkumar<sup>1,3</sup>, Amogha Koka<sup>1</sup>, Marcus Alvarez<sup>1</sup>, Dorota Kaminska<sup>1,4</sup>, Janet S. Sinsheimer<sup>1,5</sup>, Karen L. Mohlke<sup>6</sup>, Nicholas Mancuso<sup>7</sup>, Kirsi Pietiläinen<sup>8,9</sup>, Jussi Pihlajamäki<sup>4,10</sup>, Markku Laakso<sup>11</sup>, Kristina M. Garske<sup>1</sup>, Päivi Pajukanta<sup>1,2,12\*</sup>

<sup>1</sup>Department of Human Genetics, David Geffen School of Medicine at UCLA, Los Angeles, USA.

<sup>2</sup>Bioinformatics Interdepartmental Program, UCLA, Los Angeles, USA.

<sup>3</sup>Computational and Systems Biology Interdepartmental Program, UCLA, Los Angeles, USA.

<sup>4</sup>Institute of Public Health and Clinical Nutrition, University of Eastern Finland, Kuopio, Finland.

<sup>5</sup>Department of Computational Medicine, David Geffen School of Medicine at UCLA, Los Angeles, USA.

<sup>6</sup>Department of Genetics, University of North Carolina at Chapel Hill, North Carolina, USA.

<sup>7</sup>Center for Genetic Epidemiology, Department of Preventative Medicine, Keck School of Medicine, University of Southern California, Los Angeles, USA

<sup>8</sup>Obesity Research Unit, Research Programs Unit, Diabetes and Obesity, University of Helsinki, Biomedicum Helsinki, Helsinki, Finland

<sup>9</sup>Obesity Center, Endocrinology, Abdominal Center, Helsinki University Central Hospital and University of Helsinki, Helsinki, Finland

<sup>10</sup>Department of Medicine, Endocrinology and Clinical Nutrition, Kuopio University Hospital, Kuopio, Finland.

<sup>11</sup>Department of Medicine, University of Eastern Finland and Kuopio University Hospital, Kuopio, Finland.

<sup>12</sup>Institute for Precision Health at UCLA, Los Angeles, USA.

\* Correspondence should be addressed to Päivi Pajukanta, MD, PhD, Professor of Human Genetics at UCLA (ppajukanta@mednet.ucla.edu)

## Abstract

Waist-Hip-Ratio adjusted for Body Mass Index (WHRadjBMI) is a well-established sex-specific marker for abdominal fat and adiposity, and a predictor of adverse metabolic outcomes, such as type 2 diabetes (T2D). Here, we identified an adipose gene network that contains 35 obesity GWAS genes, and explains a significant amount of polygenic risk for abdominal obesity and T2D in the UK Biobank (n=502,617) in a sex-dependent way. The network is controlled by a novel adipose master transcription factor (TF), *TBX15*, and its weight-loss responsive *cis*-eQTL, rs1779445, a WHRadjBMI GWAS variant that regulates the network in *trans*. When we knocked down *TBX15* in human primary preadipocytes, expression of 130 network genes, including the key adipogenesis TFs, *PPARG* and *KLF15*, were significantly impacted (FDR<0.05), thus functionally verifying the *trans* regulatory effect of *TBX15* on the WHRadjBMI-associated network. Taken together, we discovered a new human adipose master *trans* regulator, *TBX15*, which controls an obesity GWAS gene-enriched network that sex-dependently regulates the accumulation of abdominal fat.



## Introduction

Obesity predisposes individuals to multiple cardiometabolic disorders, including type 2 diabetes (T2D)<sup>1,2</sup>. Furthermore, as the world faces one of the worst infectious-disease outbreaks in a century, new data are emerging showing that obesity and male sex are key risk factors for severe forms of COVID-19 infection in individuals less than 60 years of age<sup>3,4</sup>. However, the underlying genes and regulatory mechanisms orchestrating the sex differences in obesity and body fat distribution are not well understood.

Obesity is clinically diagnosed by a body mass index (BMI) greater than 30 kg/m<sup>2</sup>, while severe obesity is defined as BMI greater than 40 kg/m<sup>2</sup>. However, as BMI cannot reliably differentiate fat from lean mass, the metabolically detrimental abdominal obesity has been more accurately estimated using waist-hip-ratio (WHR), which even after adjusting for BMI (WHRadjBMI) is still highly heritable (heritability~0.22-0.61)<sup>5-8</sup>. WHRadjBMI is a well-established surrogate for abdominal adiposity and body fat distribution, and it has also been correlated with direct imaging assessments of abdominal fat in observational studies<sup>9-11</sup>. It is also recognized as a strong predictor of T2D<sup>12</sup>.

Previous studies have demonstrated that WHRadjBMI is a sexually dimorphic trait, reflecting the physiological differences in body fat and muscle mass, with males in general exhibiting more muscle mass and females more fat mass when matched for BMI and age<sup>13,14</sup>. Furthermore, WHRadjBMI shows large differences in the narrow sense heritability between males (~20%) and females (~50%)<sup>8,15</sup>; yet the biological mechanisms underlying abdominal adiposity and its sex-specific characteristics have remained largely elusive. Previous genome-wide association studies (GWAS) have shown that WHRadjBMI GWAS genes are enriched for adipose-expressed genes with known adipose tissue functions, whereas BMI GWAS genes are enriched for genes expressed primarily in the brain<sup>16</sup>. To advance the discovery of unknown genetic and molecular mechanisms regulating abdominal adiposity and the sex-specific distribution of body fat, we searched for genetic master regulators of WHRadjBMI by employing integrative genomics approaches on human subcutaneous adipose RNA-sequencing (RNA-seq) data (n~1,400) and WHRadjBMI GWAS, transcriptome-wide association studies (TWAS), and polygenic risk score (PRS) data from the WHRadjBMI GWAS cohorts and the UK Biobank (UKB) (n~700,000). Finally, we verified our genomic results using functional studies in a human primary cell type central to adipogenesis.

While local *cis* regulation of genes has been characterized in multiple cohorts and tissues<sup>17,18</sup>, the identification of distal *trans* regulation of adipose gene expression contributing to human obesity has been challenging. There are currently only a few examples of *trans*-eQTL genes, such as Kruppel Like Factor 14 (*KLF14*)<sup>19</sup>, mostly due to the lack of power achievable in the current, relatively small RNA-seq cohorts of relevant obesogenic tissues. One possible *trans* regulatory mechanism of gene expression is transcription factor (TF) binding to the promoters of multiple genes across many chromosomes, which causes them to be co-regulated and co-expressed<sup>20-22</sup>. As with other genes, these TFs themselves are genetically controlled by *cis*-eQTLs, thus indirectly linking the eQTLs to genes regulated by the TFs in *trans*. We hypothesized that adipose co-expression networks can be used to identify novel TFs that *trans* regulate multiple co-expressed target genes important for WHRadjBMI.

Integrating a subcutaneous adipose co-expression network and multi-omics data with extensive human GWAS cohort and the UK Biobank (UKB) data (total n~700,000), we provide novel genomic evidence, verified by our functional studies in human primary preadipocytes, for the causal role of *TBX15* in controlling accumulation of abdominal fat and adiposity. Our study discovers a new key function for the *TBX15* TF in *trans* regulating an adipose network of 347 adipogenesis, mitochondrial, and metabolically important genes, including *PPARG*, *KLF15*, *PPARA*, *ADIPOQ*, and 35 obesity GWAS genes.

## Results

### Discovery of WHRadjBMI-associated co-expression networks in human adipose tissue

In our network analysis, we used waist-hip-ratio adjusted for body mass index (WHRadjBMI) as a surrogate for abdominal adiposity and fat<sup>9-11</sup>, supported by previous GWASs that have demonstrated WHRadjBMI as a more relevant adipose tissue-related obesity trait than BMI<sup>16,23</sup>. To identify co-expression networks correlated with abdominal fat and adiposity, we performed Weighted Gene Co-expression Network Analysis (WGCNA) in the subcutaneous adipose RNA-seq data (n=335) from the Finnish METabolic Syndrome In Men (METSIM) cohort, which has additional measures of adiposity aside from BMI, including WHR. We identified 14 co-expression networks, two of which, red and black (colors assigned to networks by WGCNA arbitrarily), were significantly inversely correlated with WHRadjBMI, WHR, and BMI after adjusting for multiple testing ( $p_{\text{Bonf}} < 8.93 \times 10^{-4}$ ) (Fig. 1, Supplementary Fig. 1). To also examine if the WGCNA co-expression networks are associated with the obesity comorbidity, type 2 diabetes (T2D), we correlated them with fasting serum insulin levels and observed significant inverse correlation of the red and black co-expression networks ( $p_{\text{Bonf}} < 8.93 \times 10^{-4}$ ) (Fig. 1, Supplementary Fig. 1). The red co-expression network, with 347 genes (Supplementary Table 1), contained 35 (10.09%) obesity GWAS genes for BMI, waist circumference (WC), WHR, WHRadjBMI, and WCadjBMI (Fisher's exact test for the red co-expression network GWAS enrichment, odds ratio=5.05,  $p = 2.20 \times 10^{-4}$ ), whereas no such obesity GWAS gene enrichment was observed with the black co-expression network (Fig. 1, Supplementary Table 2).

Since WGCNA co-expression networks may be influenced by different cell types present in heterogeneous tissues such as adipose, we used adipose single-nuclei RNA-seq (snRNA-seq) from Finnish individuals (n=16) to identify marker genes for the key adipose cell types, such as adipocytes, preadipocytes, and macrophages. The red co-expression network was exclusively enriched for adipocyte marker genes ( $p_{\text{hypergeometric}} = 2.20 \times 10^{-10}$ ) (Supplementary Table 3), including the adipocyte secreted adipokine, Adiponectin (*ADIPOQ*), indicating the importance of this co-expression network for adipocyte biology. The red co-expression network was also significantly enriched for key adipose-related metabolic KEGG pathways using WebGestalt<sup>24</sup>, such as PPAR signaling pathway, fatty acid metabolism and degradation, and valine, leucine, and isoleucine degradation (FDR<0.05; Supplementary Table 4), and for GO cellular component mitochondrion-related genes (FDR<0.05; Supplementary Table 5). Furthermore, the red co-

expression network was significantly enriched for genes upregulated in subcutaneous adipose tissue ( $p \sim 1.0 \times 10^{-18}$ ) when compared to the 54 other tissues in Genotype-Tissue Expression (GTEx) v8 cohort<sup>25</sup> in a differential expression (DE) analysis by FUMA<sup>26</sup> (Fig. 1). Due to the significant enrichment of obesity GWAS genes, adipose-related functional pathways, adipocyte cell type marker genes, and adipose tissue-expressed genes, we focused on the red WHRadjBMI co-expression network for subsequent analyses.

### **The WHRadjBMI gene co-expression network is genetically associated with WHRadjBMI and T2D**

To find genetic evidence for the observed link between the co-expression network and WHRadjBMI, we examined whether the 347 network genes contribute significantly to WHRadjBMI trait heritability. We used the stratified LD Score (LDSC) regression method (see Methods) to calculate the WHRadjBMI heritability explained using the WHRadjBMI GWAS summary statistics for all variants in the *cis* regions of the 347 genes ( $\pm 500$ kb from the ends of the gene). These variants will be referred to henceforth as the WHRadjBMI *cis*-variant set. We found that these *cis* regions are significantly enriched for variants explaining the heritability of WHRadjBMI (enrichment=1.61,  $p=4.90 \times 10^{-5}$ ) and T2D (enrichment=1.49,  $p=9.56 \times 10^{-3}$ ) but not significantly enriched for variants explaining the heritability of BMI ( $p > 0.05$ ) (Supplementary Table 6). These summary-level findings indicate that the 347 network genes and their *cis* variants are specifically important in controlling abdominal fat and adiposity and contributing to the clinical metabolic outcome, T2D.

To investigate how the WHRadjBMI co-expression network genes predict individual risk for elevated WHRadjBMI compared to the entire genome, we constructed two separate Polygenic Risk Scores (PRSs) for WHRadjBMI: a genome-wide PRS and a network PRS with just the variants in the WHRadjBMI *cis*-variant set. For these PRS analyses, we used the UK Biobank (UKB) cohort and divided the unrelated Caucasian participants into a test ( $n=130,851$ ) and validation ( $n=261,700$ ) set (see Methods for building the PRS).

To investigate the effectiveness of our genome-wide PRS in predicting WHRadjBMI with the validation set ( $n=261,700$ ) (PRS correlation coefficient with WHRadjBMI=0.206), we divided the individuals into 20 quantiles based on their PRS scores and then by sex. Next, we calculated the odds ratio of being in the top 10<sup>th</sup> percentile of WHRadjBMI, for individuals in

each of the 20 quantiles compared to the lowest quantile. As expected, the PRS predicts WHRadjBMI better in females than males (Females: 6.31-fold increase in risk for elevated WHRadjBMI between the lowest quantile and the 20<sup>th</sup> quantile of the PRS versus Males: 2.96-fold increase in risk for elevated WHRadjBMI) (Fig. 2).

Notably, despite the fact that the network PRS only comprises the variants in the *cis* regions of the 347 network genes, having thus many fewer variants included, the PRS correlation coefficient with WHRadjBMI was 0.110 (compared with the genome-wide PRS correlation coefficient of 0.206, which is less than twice that of the network PRS). Although the PRSs are more predictive of WHRadjBMI in females (Cochran-Mantel-Haenszel test on the 20<sup>th</sup> quantile, genome-wide PRS versus network PRS and males versus females,  $\chi^2_{CMH}=1146.94$ ,  $p_{CMH}=2.07 \times 10^{-251}$ ), the power decrease from using the genome-wide PRS to using the network PRS is much greater for females (20<sup>th</sup> quantile odds ratio: 2.51-fold decrease) when compared to males (20<sup>th</sup> quantile odds ratio: 1.71-fold decrease) (Fig. 2). This suggests that, relative to the genome-wide PRS, the 347 network genes and their *cis* variants constitute a larger percentage of the predicted effect of variants for regulating WHRadjBMI in males when compared to the same PRS predictions in females.

To provide additional evidence that the network PRS is more informative and biologically important in males than females, we tested whether males with the highest genetically predicted WHRadjBMI (based on the network PRS) are more likely to have the clinically relevant metabolic outcome of T2D. Accordingly, we selected individuals with the WHRadjBMI PRS in the highest and lowest deciles (top 10% and lowest 10% network PRS scores), as done previously for BMI in Khera *et al.*<sup>27</sup>, and divided them by sex. We used a logistic regression (see Methods) and when accounting for WHRadjBMI in our model, observed that the network PRS significantly predicted T2D in males ( $\beta=1.12$ ,  $p=9.59 \times 10^{-5}$ ) but not in females ( $p>0.05$ ). These results indicate that the 347 network genes and their *cis* variants significantly contribute to the clinical metabolic outcome, T2D, in males while no such effect was observed in females. In sum, by leveraging subcutaneous adipose RNA-seq data from a cohort with the abdominal adiposity measure, WHR, we identified a WHRadjBMI co-expression network that genetically controls WHRadjBMI and T2D in a sex-dependent manner.

### **The WHRadjBMI network connectivity is sex- and context-dependent**

We hypothesized that the sex-dependent effects we observed with the network PRS for WHRadjBMI and T2D would be reflected in the co-regulation of these genes as well. We therefore tested whether the WHRadjBMI co-expression network connectivity is different between males and females in the independent GTEx v8 subcutaneous adipose RNA-seq data. We performed a network preservation analysis separately in males (n=387) and females (n=194) (see Methods), and found that the network preservation  $Z_{\text{Summary}}$  score was 30 in males versus 22 in females. The  $Z_{\text{Summary}}$  score value is not sensitive to the sample size, and so the relative difference in the number of males and females was not a concern. This lower network preservation in females is in line with the lesser trait prediction observed for WHRadjBMI and T2D with the network PRS in females.

We further tested whether the WHRadjBMI network connectivity is altered context-dependently based on the obesity state. Because the GTEx cohort phenotypes do not include WHRadjBMI, we divided the cohort first by sex and then into the more extreme categories of lean (BMI<25; n<sub>Male</sub>=102, n<sub>Female</sub>=78) and obese (BMI>30; n<sub>Male</sub>=119, n<sub>Female</sub>=41) to increase the chance that there are differences in abdominal adiposity between the sets of individuals. We found that the network preservation  $Z_{\text{Summary}}$  score drastically decreased between lean and obese males ( $Z_{\text{Summary}} - \text{Lean male}=30$  versus  $Z_{\text{Summary}} - \text{Obese male}=19$ ) but remained similar between lean and obese females ( $Z_{\text{Summary}} - \text{Lean female}=20$  versus  $Z_{\text{Summary}} - \text{Obese female}=18$ ). Taken together, the network preservation results suggest that the coordinated expression of the genes in the WHRadjBMI co-expression network is regulated more tightly in males than females, and in a context-specific manner that depends on the obesity state.

### **Identifying candidate master regulators of the WHRadjBMI-associated co-expression network**

To discover transcription factors (TFs) that drive the adipose WHRadjBMI co-expression network in *trans*, we first identified all TFs (n=14) in the network using the PANTHER database<sup>28</sup> (Supplementary Table 7). Next, to test which of these 14 TFs are potentially causal for WHRadjBMI, we performed a targeted Transcriptome-Wide Association Study (TWAS), which is a method to test for association between gene expression and a trait by weighting the effects of *cis* variants on gene expression and testing their weighted association with a GWAS

trait (see Methods). We computed eQTL weights for the variants in the *cis* region ( $\pm$ 500kb from the ends of the gene) around each TF using GTEx v8 cohort data. To accurately estimate the gene expression heritability in TWAS, we used the entire GTEx subcutaneous adipose RNA-seq dataset (n=581). We found that five TFs in the WHRadjBMI co-expression network pass the TWAS heritability thresholds ( $p < 0.01$ ) required for testing the association of the *cis* SNP heritability with phenotypes: T-Box Transcription Factor 15 (*TBX15*), General Transcription Factor IIE Subunit 2 (*GTF2E2*), X-Prolyl Aminopeptidase 3 (*XPNPEP3*), Iroquois Homeobox 1 (*IRX1*), and Zinc Finger Protein 3 (*ZNF3*) (Supplementary Table 8).

We next tested whether these five *cis*-heritable TFs are associated with WHRadjBMI using the computed TWAS weights to impute the TF gene expression and the WHRadjBMI summary statistics from the large UK Biobank (UKB) and GIANT meta-analysis GWAS data (n~700,000). *TBX15*, *XPNPEP3*, and *IRX1* passed the Bonferroni correction for being associated with WHRadjBMI in the TWAS ( $p < 0.017$ ) (Supplementary Table 9), implying that the variants contributing to the *cis*-regulation of these TFs are also important for WHRadjBMI.

The interpretation of TWAS results as evidence of causality can be complicated by other regional genes that may share *cis* variants, LD structure, or co-expression with the putatively causal gene. To better determine if there is statistical support for the TWAS evidence of association between WHRadjBMI and *TBX15*, *XPNPEP3*, and *IRX1*, we used the Fine mapping Of CaUsal Sets (FOCUS) tool, employing the same GTEx v8 cohort and WHRadjBMI GWAS data, and including all genes  $\pm$  3Mb from the ends of our TFs of interest. FOCUS is a fine-mapping approach for TWAS that identifies a gene set containing the causal gene(s) in a locus at a predefined level of credibility, based on their posterior inclusion probability (PIP) of being the causal gene while accounting for shared *cis* variation among genes at a locus (see Methods). The FOCUS analyses showed that *TBX15* and nearby gene Hydroxy-Delta-5-Steroid Dehydrogenase, 3 Beta- And Steroid Delta-Isomerase 2 (*HSD3B2*) were included in the 90% credible set; however, only *TBX15* predicted well in cross-validation (*TBX15*: TWAS cross-validation  $p = 1.54 \times 10^{-7}$ ; *HSD3B2*: TWAS cross-validation  $p > 0.05$ ) and had a higher PIP (*TBX15*: FOCUS PIP > 0.99; *HSD3B2*: FOCUS PIP = 0.908), thus effectively fine-mapping the locus to *TBX15* (Fig. 3). When testing *XPNPEP3* and *IRX1*, FOCUS provided little support for a causal role at current sample sizes (*XPNPEP3*: FOCUS PIP =  $9.90 \times 10^{-5}$ ; *IRX1*: FOCUS PIP = 0.0735). Taken together, the results from TWAS and FOCUS show statistical support for a causal role of the TF, *TBX15*,

in its inverse relationship with WHRadjBMI, and highlight it as a candidate TF driving the WHRadjBMI co-expression network.

### **Identification of a WHRadjBMI co-expression network *trans*-eQTL**

Support for the evidence that *TBX15* is a causal gene in regulating adiposity has been published in mouse knockout studies, where adipose-specific loss of *Tbx15* leads to increased weight gain when mice are put on a high fat diet<sup>29</sup>. This suggests that, in conditions of increased energy intake, a pathological decrease in *TBX15* can drive adiposity. To test for evidence of a similar mechanism in humans, we used subcutaneous adipose RNA-seq data from the Finnish Kuopio OBesity Study (KOBS) bariatric surgery cohort, in which the individuals' average BMI decreased from 43.0 to 34.3 (22.7% decrease) from the time of surgery to the one-year follow-up (n=168 at both time points). A change in WHR could not be assessed in the KOBS cohort as in general it is not possible to reliably measure waist circumference in morbidly obese individuals undergoing bariatric surgery. In these weight loss analyses, we found that *TBX15* showed a significant increase in gene expression in the one-year follow-up ( $\log_2$  fold change (FC)=0.37,  $p=1.48 \times 10^{-6}$ ), in line with context-specific regulation of *TBX15* and its inverse correlation with adiposity.

To identify genetic drivers of the observed context-specific expression of *TBX15* in the extreme obesity state, we searched for *TBX15* *cis*-eQTLs separately in the KOBS bariatric surgery cohort baseline and follow-up adipose RNA-seq data (n=168 at both time points). We discovered that *TBX15* has a *cis*-eQTL, rs1779445 (effect allele: C) ( $\beta_{C \text{ allele}}=0.56$ ,  $p_{\text{Baseline}}=6.7 \times 10^{-6}$ ), exclusive to the extreme obesity state before the bariatric surgery ( $p_{\text{Follow-up}}>0.05$ ). The context-specificity of the *cis*-eQTL effect was also supported by the non-significant *cis* effects in two non-extreme obese cohorts, GTEx and METSIM (n=581, n=335, respectively; FDR>0.05). This context-specific *TBX15* *cis*-eQTL variant, rs1779445, is also a WHRadjBMI GWAS SNP (GIANT, n=224,459) ( $\beta_{C \text{ allele}}=0.032$ ,  $p=1.60 \times 10^{-12}$ )<sup>8</sup>. We recognize that the direct identification of *trans*-eQTLs requires large cohorts; nevertheless, to partially circumvent this, we tested whether rs1779445 regulates the eigengene of the WHRadjBMI network. We found that rs1779445 is a *trans*-eQTL of the network eigengene in the METSIM cohort (n=335) ( $\beta_{C \text{ allele}}=-0.018$ ,  $p=0.044$ ). In Finnish individuals, as confirmed in the 1000 Genomes genotype data, the minor allele frequency (MAF) of rs1779445 (T allele) is lower than



in other European populations ( $MAF_{Finns} \sim 6\%$  versus  $MAF_{Europeans} \sim 20\%$ ) and it exhibits only one LD proxy, rs1623409 ( $r^2=0.82$ ). Since, this LD proxy rs1623409 was not a significant *trans*-eQTL of the eigengene of the WHRadjBMI co-expression network ( $p>0.05$ ), this *trans*-population LD comparison helped identify rs1779445 as the likely sole underlying *trans* variant driving the co-expression network.

### **Electrophoretic mobility shift assay (EMSA) shows increased protein binding at the alternate allele of rs1779445 in males**

To investigate which proteins may bind and regulate gene expression at the context-specific eQTL of *TBX15* (rs1779445), we determined which known TF motifs are altered by the alleles of rs1779445 (T>C) using the PrEdicting Regulatory Functional Effect by Approximate P-value Estimation (PERFECTOS-APE)<sup>30</sup> tool. We found the TF motifs of Ras Responsive Binding Element 1 (RREB1), BRCA1 DNA Repair Associated (BRCA1), Regulatory Factor X1 (RFX1), and Regulatory Factor X2 (RFX2) to be significantly altered, with increased protein binding predicted at the alternate allele. Furthermore, *RREB1* has WHRadjBMI GWAS variants (rs11755724, rs675209) in its promoter, specifically found in European males, further connecting *RREB1* and *TBX15* to WHRadjBMI in males. We performed an electrophoretic mobility shift assay (EMSA) and observed that there was more protein binding at the alternate allele of rs1779445 in males (Fig. 4), in line with the predicted effect of the alternate allele (C) RREB1 binding and known repressive role of RREB1 in transcription<sup>31</sup>. Further studies are warranted to examine and validate the actual TF that is binding at the WHRadjBMI network *trans*-eQTL.

### **Knockdown of *TBX15* in primary human preadipocytes confirms the role of *TBX15* as a master regulator of the WHRadjBMI associated network**

To functionally confirm the role of *TBX15* as a WHRadjBMI co-expression network regulator, we performed knockdown (KD) of *TBX15* via small-interfering RNA (siRNA) in primary human preadipocytes (n=5 isogenic replicates). We successfully performed *TBX15* KD, decreasing its expression by  $\sim 70\%$ , confirmed by RT-qPCR (Fig. 4). Next, we performed RNA-seq to see if the genes in the WHRadjBMI co-expression network are affected by KD of *TBX15*. When comparing to preadipocytes transfected with the negative control siRNA (see Methods), we found that 130 of the 347 network genes (37.46%) are significantly DE ( $FDR<0.05$ ) between the

*TBX15* KD and control, including the well-established key adipogenesis master regulators, *PPARG* and *KLF15* (Supplementary Table 10).

When searching for other TFs affected by *TBX15* KD that may contribute to the widespread *trans* effects of *TBX15*, a total of 8 TFs of the 13 TFs (61.5%) in the WHRadjBMI co-expression network were observed to be significantly DE (FDR<0.05) (*PPARG*, *PPARA*, *KLF15*, *TWIST1*, *XPNPEP3*, *GTF2E2*, *CCNH*, *PER3*) by the *TBX15* KD. This result suggests that *TBX15* affects many additional genes indirectly downstream by regulating other key adipose TFs (Fig. 4).

In summary, these genetic and functional data discover a context-specific *cis*-eQTL, controlling a new human adipose master *trans* regulator, *TBX15*, which in turn controls an obesity GWAS gene-enriched network that sex-dependently modifies the distribution of fat.

## Discussion

WHRadjBMI is a well-established measure of abdominal adiposity, whereas BMI cannot reliably separate fat from lean mass<sup>16</sup>, in line with previous GWAS studies of WHRadjBMI and BMI demonstrating that the WHRadjBMI GWAS loci are more adipose tissue related than the BMI loci in terms of their expression profiles and function<sup>9-11</sup>. Furthermore, while overall obesity measures like BMI do not exhibit sexual dimorphism<sup>8</sup>, WHRadjBMI and fat distribution have clear sex-specific differences that are reflected in differences in heritability<sup>8,15</sup>, GWAS loci<sup>13,23</sup>, and ultimately in risk for disease outcomes such as T2D and cardiovascular disease<sup>14,27</sup>.

However, the underlying biological mechanisms that contribute to the sexual dimorphism of body fat distribution are still poorly understood. Furthermore, the genes behind complex diseases such as obesity are often regulated and dysregulated together, influencing the progression and severity of obesity<sup>32</sup>.

In this work, we used subcutaneous adipose RNA-seq data collected in the METSIM male population cohort, for which we have measures of WHR, to identify a gene co-expression network that is important for regulating WHRadjBMI and exhibits the known sexual dimorphism of this trait at both a genetic and transcriptomic level. We used the UKB to show that the genetic variants in the *cis*-regions of the 347 WHRadjBMI co-expression network genes are significantly enriched for variants that contribute to the heritability of WHRadjBMI and T2D, but not BMI. These variants also have a sex-dependent effect on the ability to predict elevated WHRadjBMI in males when compared to females relative to the entire genome, as shown by the genome-wide and network-specific WHRadjBMI PRSs we constructed. Furthermore, we show that the network PRS significantly predicts the disease outcome, T2D, in males but not in females, even when accounting for the effects from the original trait, WHRadjBMI. These PRS results demonstrate the sex-dependent effects of the 347 WHRadjBMI co-expression network genes and their *cis* variants on both WHRadjBMI and T2D. Finally, we provide evidence for a novel role of the TF, T-Box Transcription Factor 15 (*TBX15*), as a master regulator of this WHRadjBMI network, advancing our understanding of how *trans* regulation of gene expression contributes to normal and obesity-deteriorated adipose tissue function, and the sexually dimorphic accumulation of harmful abdominal fat.

We employed TWAS<sup>33</sup> to discover and FOCUS<sup>34</sup> to fine-map all TFs (n=14) present in the WHRadjBMI network, which resulted in the discovery of the *TBX15* as a master *trans*

regulator for this WHRadjBMI network. *Tbx15* has been directly shown to affect the differentiation of preadipocytes to adipocytes, with reduced expression of key adipogenesis TFs *Cebpa* and *Pparg* in mouse preadipocytes that stably overexpress (OE) *Tbx15*<sup>35</sup>. This study also suggests that even after rescuing the induction of adipogenesis using a PPARG agonist, *Tbx15* OE cells exhibit decreased lipogenesis and increased lipolysis. These results are in line with the inverse relationship of *TBX15* with WHRadjBMI, and also highlight the likelihood of development- and context-specific roles for *TBX15*<sup>35</sup>. Additionally, previous functional studies have shown that *TBX15* affects mitochondria related gene expression and mitochondrial mass in mice<sup>35</sup> and humans<sup>36,37</sup>, in line with the GO cellular-component enrichment of the WHRadjBMI co-expression network genes for mitochondrion-related genes. In addition to mouse knockout studies, where adipose-specific loss of *Tbx15* leads to increased weight gain when mice are put on a high fat diet<sup>29</sup>, these previous studies provide support for our discovery of *TBX15* as a key TF master regulator in human subcutaneous adipose tissue, with adiposity-driven changes in *TBX15* expression affecting its role in maintaining homeostasis of the WHRadjBMI network.

We used the independent subcutaneous adipose RNA-seq data from the GTEx v8 cohort<sup>25</sup> to show that the WHRadjBMI co-expression network is highly preserved in a sex-dependent manner, with males exhibiting a higher network preservation than females. Furthermore, the network preservation is higher in the lean (BMI < 25) state when compared to the obese (BMI > 30) state in males, but is similar between lean and obese females. This apparent breakdown of network connectivity in the obese males supports the idea that aberrant regulation of the network as a whole develops as WHRadjBMI increases. Although the GTEx cohort<sup>25</sup> lacks measurements for WHRadjBMI due to the fact that it consists largely of post-mortem samples, we were able to show the sex- and obesity-dependent effects on this WHRadjBMI network using more extreme BMI cutoffs of lean (BMI < 25) and obese (BMI > 30). However, presently there are no sex-specific guidelines for the BMI cutoffs for the transition between lean, overweight and obese states, let alone WHRadjBMI. To partially circumvent this issue and study the effects of weight differences on *TBX15* expression, we leveraged longitudinal adipose RNA-seq data from the KOBs bariatric surgery cohort, which demonstrated that adipose expression of *TBX15* recovers after dramatic weight loss within an individual. These weight loss results from the KOBs cohort suggest that decreased adipose expression of *TBX15* in obese

individuals contributes to the observed dysregulation of the WHRadjBMI co-expression network.

Although visceral adipose tissue is known to be more strongly linked to metabolic disorders and WHRadjBMI<sup>38,39</sup> than subcutaneous adipose tissue, subcutaneous adipose tissue exhibits larger changes in volume during weight loss or weight gain<sup>40</sup>. Furthermore, subcutaneous adipose biopsies are available through less invasive procedures than visceral adipose tissue biopsies, which require a surgical procedure. Our results from the heritability and PRS analyses; and the context-specificity of the network regulation show that the subcutaneous adipose WHRadjBMI network is both an important driver and responder, respectively, to changes in WHRadjBMI.

To functionally verify that the WHRadjBMI network is driven by *TBX15*, we knocked down *TBX15* via siRNA in primary human preadipocytes, and performed RNA-seq to assess the effects of *TBX15* KD on the expression of all 347 network genes. This experiment showed that knocking down *TBX15* significantly affects the downstream expression of 8 additional TFs, including the key adipogenesis TFs, *PPARG* and *KLF15*, along with 121 other network genes. To the best of our knowledge, our functional study is one of the first examples of experimental validation of a TF *trans* regulating a co-expression network in humans. Furthermore, these DE genes are enriched for the Valine, leucine, and isoleucine degradation KEGG pathway using WebGestalt<sup>24</sup>. This pathway functions in the breakdown of essential branched chain amino acids that humans only obtain in their diet. Previous studies have shown that obese individuals exhibit higher levels of these amino acids in their plasma even when matched for dietary intake or after overnight fasting, most likely due to their impaired degradation<sup>41</sup>. Taken together, these data, along with the recovery of *TBX15* expression after weight loss, indicate that *TBX15* plays an important role in maintaining the homeostasis of this subcutaneous adipose WHRadjBMI co-expression network in the non-obese state.

Lastly, our work helps disentangle the genetics at the *TBX15* locus, where the previous obesity GWAS variants and regional *cis*-eQTLs have been associated with both *TBX15* and nearby gene, *WARS2*<sup>16,42</sup>. WHRadjBMI GWAS variants at the *TBX15*-*WARS2* locus have been shown to exhibit differential effects in males and females, including a male-specific signal within the *TBX15* promoter<sup>8</sup>. Only through the investigation of extreme obesity in the KOBS cohort were we able to establish a *cis*-eQTL function for the WHRadjBMI GWAS SNP<sup>8</sup> rs1779445,

which is a strong *cis*-eQTL of *TBX15* but not of *WARS2* in the extreme obesity state. Furthermore, this *TBX15 cis*-eQTL is a *trans* driver of the overall WHRadjBMI co-expression network expression, defined by the network eigengene, supporting our conclusion that *TBX15* is a master regulator of the WHRadjBMI network. Our functional assay to detect the binding of primary human preadipocyte nuclear protein to this SNP using EMSAs show increased binding at the alternate allele of rs1779445 (C allele), with more binding observed using nuclear extract from male primary preadipocytes than female primary preadipocytes. Based on predictions using position weight matrices (PWMs)<sup>43,30</sup>, we hypothesize that *RREB1* is binding at the alternate allele of rs1779445. *RREB1* itself is a TF with intronic WHRadjBMI GWAS variants in European males<sup>15</sup>, which indicates that upstream regulators of *TBX15* may also act differently between males and females. The confirmation of *RREB1* as an upstream regulator of *TBX15* in key human adipose tissue cell types warrants further investigation.

In summary, we discovered a novel master adipose *trans* regulator, *TBX15*, and its causal effect on WHRadjBMI, with a stronger effect observed in males. We also provide insight into a WHRadjBMI co-expression network containing critical adipose TFs and GWAS genes that *TBX15* regulates, and demonstrate the large contribution of the *cis* variants of these network genes to both WHRadjBMI PRS and T2D PRS in a sex-dependent manner in the UK Biobank. Through our knockdown of *TBX15* in human primary preadipocytes, we provide concrete functional evidence showing that decreasing expression of *TBX15* directly affects expression of 130 genes in the WHRadjBMI co-expression network, including 8 key TFs, thus compounding the downstream effects on metabolically harmful abdominal obesity.

## **Acknowledgements**

We thank the individuals who participated in the METSIM, GTEx, and KOBS studies. We also thank the sequencing core at UCLA for performing the RNA sequencing. The Genotype-Tissue Expression (GTEx) Project was supported by the Common Fund of the Office of the Director of the National Institutes of Health, and by NCI, NHGRI, NHLBI, NIDA, NIMH, and NINDS. The data used for the analyses described in this manuscript were obtained from: the GTEx Portal on 12/05/2019, WashU Epigenome Browser of 03/11/2020. This study was funded by National Institutes of Health (NIH) grants HL-095056, HL-28481, U01 DK105561, R00 HL121172, and DK093757. D. Z. Pan was supported by the NIH-NCI grant T32LM012424 and NIH-NIDDK grant F31 DK118865-02, K. M. Garske by NIH-NHLBI F31-HL142180, Z. Miao was supported by AHA grant 19PRE34430112, C. Comenho was supported by NIH-NIGMS award number R25GM055052, and M. Alvarez was supported by the HHMI Gilliam Fellowship. J. Sinsheimer was partial funded by NIH grants HG-009120 and GM-0532275. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the article. Genotyping for the METSIM cohort were supported by NIH grants DK072193, DK093757, DK062370, and Z01HG000024 and provided by the Center for Inherited Disease Research (CIDR). CIDR is fully funded through a federal contract from the NIH to The Johns Hopkins University, contract number HHSN268201200008I. This research has been conducted using the UKBiobank Resource under Application Number 3934.

## **Author Contributions**

Study design: D.Z.P., K.M.G., P.P. Methods development and statistical analysis: D.Z.P., K.M.G., Z.M., M.A., J.S.S, N.M., and P.P. Computational analysis: D.Z.P., K.M.G., M.A., Z.M., and N.M. Experiments were performed by K.M.G., C.C, S.R., and A.K. RNA-sequencing and quality control: D.Z.P, Z.M., and P.P. METSIM data collection , genotyping, and phenotyping: K.L.M., M.L., and P.P. Data collection from Finnish individuals for snRNA-seq: M.A. P.P, and K.P. KOBS data collection, genotyping, and phenotyping: D.K. and J.P. Manuscript: D.Z.P., K.M.G. and P.P. wrote the manuscript and all authors read, reviewed, and/or edited the manuscript.

**Competing Interests**

The authors declare no competing interests.

**Data availability**

The data that support the findings in this manuscript are available from the UKBiobank. However, restrictions apply to the availability of these data, which were used in this study under UKBiobank Application Number 3934. UKBiobank data are available for bona fide researchers through the application process. The METSIM gene expression data was previously made available in Raulerson *et al.*, Am J Hum Genet, 2019 in GEO under GSE135134.

**Code availability**

All code used for analyses in this study were unaltered from their publicly available sources. Parameters chosen for each analysis are described in detail in Methods.



## References

1. Hales, C. M., Carroll, M. D., Fryar, C. D. & Ogden, C. L. *Prevalence of Obesity Among Adults and Youth: United States, 2015–2016. NCHS data brief, no 288. Hyattsville, MD: National Center for Health Statistics.alth Statistics.* (2017).
2. Benjamin, E. J. *et al.* Heart Disease and Stroke Statistics-2017 Update: A Report from the American Heart Association. *Circulation* **135**, e146–e603 (2017).
3. Lighter, J. *et al.* Obesity in patients younger than 60 years is a risk factor for Covid-19 hospital admission. *Clin. Infect. Dis.* (2020).
4. Simonnet, A. *et al.* High prevalence of obesity in severe acute respiratory syndrome coronavirus-2 (SARS-CoV-2) requiring invasive mechanical ventilation. *Obesity* (2020).
5. 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).
6. Mills, G. W. *et al.* Heritability estimates for beta cell function and features of the insulin resistance syndrome in UK families with an increased susceptibility to Type 2 diabetes. *Diabetologia* **47**, 732–738 (2004).
7. Souren, N. Y. *et al.* Anthropometry, carbohydrate and lipid metabolism in the East Flanders Prospective Twin Survey: Heritabilities. *Diabetologia* **50**, 2107–2116 (2007).
8. Shungin, D. *et al.* New genetic loci link adipose and insulin biology to body fat distribution. *Nature* **518**, 187–196 (2015).
9. Ashwell, M., Cole, T. J. & Dixon, A. K. Obesity: New insight into the anthropometric classification of fat distribution shown by computed tomography. *Br. Med. J. (Clin. Res. Ed).* **290**, 1692 (1985).
10. Seidell, JC; Björntorp, P; Sjöström, L; Sannerstedt, R, Krotkiewski, M; Kvist, H. Regional distribution of muscle and fat mass in men--new insight into the risk of abdominal obesity using computed tomography. *Int. J. Obes.* **13**, 289–303 (1989).
11. Emdin, C. A. *et al.* Genetic association of waist-to-hip ratio with cardiometabolic traits, type 2 diabetes, and coronary heart disease. *JAMA - J. Am. Med. Assoc.* **317**, 626–634 (2017).
12. Meisinger, C., Döring, A., Thorand, B., Heier, M. & Löwel, H. Body fat distribution and risk of type 2 diabetes in the general population: Are there differences between men and women? The MONICA/KORA Augsburg Cohort Study. *Am. J. Clin. Nutr.* **84**, 483–489 (2006).
13. Rask-Andersen, M., Karlsson, T., Ek, W. E. & Johansson, Å. Genome-wide association study of body fat distribution identifies adiposity loci and sex-specific genetic effects. *Nat. Commun.* **10**, (2019).
14. Schorr, M. *et al.* Sex differences in body composition and association with cardiometabolic risk. *Biol. Sex Differ.* **9**, 1–10 (2018).
15. 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).
16. Heid, I. M. *et al.* Meta-analysis identifies 13 new loci associated with waist-hip ratio and reveals sexual dimorphism in the genetic basis of fat distribution. *Nat. Genet.* **42**, 949–960 (2010).
17. Westra, H. J. *et al.* Systematic identification of trans eQTLs as putative drivers of known disease associations. *Nat. Genet.* **45**, 1238–1243 (2013).
18. Aguet, F. *et al.* Genetic effects on gene expression across human tissues. *Nature* **550**, 204–213 (2017).
19. Small, K. S. *et al.* Regulatory variants at KLF14 influence type 2 diabetes risk via a female-specific effect on adipocyte size and body composition. *Nat. Genet.* **50**, 572–580 (2018).
20. Sassone-Corsi, P. & Borrelli, E. Transcriptional regulation by trans-acting factors. *Trends Genet.*

- 2, 215–219 (1986).
21. Borensztein, P. *et al.* cis-Regulatory elements and trans-acting factors directing basal and cAMP-stimulated human renin gene expression in chorionic cells. *Circ. Res.* **74**, 764–773 (1994).
  22. Chen, C. *et al.* The transcription factor POU3F2 regulates a gene coexpression network in brain tissue from patients with psychiatric disorders. *Sci. Transl. Med.* **10**, eaat8178 (2018).
  23. Winkler, T. W. *et al.* A joint view on genetic variants for adiposity differentiates subtypes with distinct metabolic implications. *Nat. Commun.* **9**, (2018).
  24. Liao, Y., Wang, J., Jaehnig, E. J., Shi, Z. & Zhang, B. WebGestalt 2019: gene set analysis toolkit with revamped UIs and APIs. *Nucleic Acids Res.* **47**, W199–W205 (2019).
  25. The Genotype Tissue Expression Consortium. The GTEx Consortium atlas of genetic regulatory effects across human tissues The Genotype Tissue Expression Consortium. *bioRxiv* (2019).
  26. Watanabe, K., Taskesen, E., Van Bochoven, A. & Posthuma, D. Functional mapping and annotation of genetic associations with FUMA. *Nat. Commun.* **8**, (2017).
  27. Khera, A. V. *et al.* Polygenic Prediction of Weight and Obesity Trajectories from Birth to Adulthood. *Cell* **177**, 587–596.e9 (2019).
  28. Thomas, P. D. *et al.* PANTHER: A library of protein families and subfamilies indexed by function. *Genome Res.* **13**, 2129–2141 (2003).
  29. Sun, W. *et al.* Tbx15 is required for adipocyte browning induced by adrenergic signaling pathway. *Mol. Metab.* **28**, 48–57 (2019).
  30. Vorontsov, I. E., Kulakovskiy, I. V., Khimulya, G., Nikolaeva, D. D. & Makeev, V. J. PERFECTOS-APE: Predicting regulatory functional effect of SNPs by approximate P-value estimation. *Bioinforma. 2015 - 6th Int. Conf. Bioinforma. Model. Methods Algorithms, Proceedings; Part 8th Int. Jt. Conf. Biomed. Eng. Syst. Technol. BIOSTEC 2015* 102–108 (2015).
  31. Bonomo, J. A. *et al.* The ras responsive transcription factor RREB1 is a novel candidate gene for type 2 diabetes associated end-stage kidney disease. *Hum. Mol. Genet.* **23**, 6441–6447 (2014).
  32. Emilsson, V. *et al.* Genetics of gene expression and its effect on disease. *Nature* **452**, 423–428 (2008).
  33. Gusev, A. *et al.* Integrative approaches for large-scale transcriptome-wide association studies. *Nat. Genet.* **48**, 245–252 (2016).
  34. Mancuso, N. *et al.* Probabilistic fine-mapping of transcriptome-wide association studies. *Nat. Genet.* **51**, 675–682 (2019).
  35. Gesta, S. *et al.* Mesodermal developmental gene Tbx15 impairs adipocyte differentiation and mitochondrial respiration. *Proc. Natl. Acad. Sci. U. S. A.* **108**, 2771–2776 (2011).
  36. Ejarque, M. *et al.* Adipose tissue mitochondrial dysfunction in human obesity is linked to a specific DNA methylation signature in adipose-derived stem cells. *Int. J. Obes.* **43**, 1256–1268 (2019).
  37. Gburcik, V., Cawthorn, W. P., Nedergaard, J., Timmons, J. A. & Cannon, B. An essential role for tbx15 in the differentiation of brown and ‘brite’ but not white adipocytes. *Am. J. Physiol. - Endocrinol. Metab.* **303**, 1053–1060 (2012).
  38. Gesta, S., Tseng, Y. H. & Kahn, C. R. Developmental Origin of Fat: Tracking Obesity to Its Source. *Cell* **131**, 242–256 (2007).
  39. Fox, C. S. *et al.* Genome-wide association for abdominal subcutaneous and visceral adipose reveals a novel locus for visceral fat in women. *PLoS Genet.* **8**, (2012).
  40. Merlotti, C., Ceriani, V., Morabito, A. & Pontiroli, A. E. Subcutaneous fat loss is greater than visceral fat loss with diet and exercise, weight-loss promoting drugs and bariatric surgery: A critical review and meta-analysis. *Int. J. Obes.* **41**, 672–682 (2017).

41. Siddik, M. A. B. & Shin, A. C. Recent progress on branched-chain amino acids in obesity, diabetes, and beyond. *Endocrinol. Metab.* **34**, 234–246 (2019).
42. Gao, C. *et al.* Genome-Wide Study of Subcutaneous and Visceral Adipose Tissue Reveals Novel Sex-Specific Adiposity Loci in Mexican Americans. *Obesity* **26**, 202–212 (2018).
43. Kheradpour, P. & Kellis, M. Systematic discovery and characterization of regulatory motifs in ENCODE TF binding experiments. *Nucleic Acids Res.* **42**, 2976–2987 (2014).
44. Stančáková, A. *et al.* Hyperglycemia and a common variant of GCKR are associated with the levels of eight amino acids in 9,369 finnish men. *Diabetes* **61**, 1895–1902 (2012).
45. Laakso, M. *et al.* The Metabolic Syndrome in Men study: A resource for studies of metabolic & cardiovascular diseases. *J. Lipid Res.* **58**, 481–493 (2017).
46. Pan, D. Z. *et al.* Integration of human adipocyte chromosomal interactions with adipose gene expression prioritizes obesity-related genes from GWAS. *Nat. Commun.* **9**, (2018).
47. O’Connell, J. *et al.* A General Approach for Haplotype Phasing across the Full Spectrum of Relatedness. *PLoS Genet.* **10**, (2014).
48. Howie, B. N., Donnelly, P. & Marchini, J. A flexible and accurate genotype imputation method for the next generation of genome-wide association studies. *PLoS Genet.* **5**, e1000529 (2009).
49. Chang, C. C. *et al.* Second-generation PLINK: Rising to the challenge of larger and richer datasets. *Gigascience* **4**, 1–16 (2015).
50. Sudlow, C. *et al.* UK Biobank: An Open Access Resource for Identifying the Causes of a Wide Range of Complex Diseases of Middle and Old Age. *PLoS Med.* **12**, 1–10 (2015).
51. Bycroft, C. *et al.* The UK Biobank resource with deep phenotyping and genomic data. *Nature* **562**, 203–209 (2018).
52. Pihlajamäki, J. *et al.* Cholesterol absorption decreases after Roux-en-Y gastric bypass but not after gastric banding. *Metabolism.* **59**, 866–872 (2010).
53. Männistö, V. T. *et al.* Ketone body production is differentially altered in steatosis and non-alcoholic steatohepatitis in obese humans. *Liver Int.* **35**, 1853–1861 (2015).
54. Benhammou, J. N. *et al.* Novel Lipid Long Intervening Noncoding RNA, Oligodendrocyte Maturation-Associated Long Intergenic Noncoding RNA, Regulates the Liver Steatosis Gene Stearoyl-Coenzyme A Desaturase As an Enhancer RNA. *Hepatol. Commun.* **3**, 1356–1372 (2019).
55. Dobin, A. *et al.* STAR: Ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21 (2013).
56. Rodríguez, A. *et al.* Molecular Characterization of the Lipid Genome-Wide Association Study Signal on Chromosome 18q11.2 Implicates HNF4A-Mediated Regulation of the TMEM241 Gene. *Arterioscler. Thromb. Vasc. Biol.* **36**, 1350–1355 (2016).
57. Langfelder, P. & Horvath, S. WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics* **9**, 559 (2008).
58. Langfelder, P., Luo, R., Oldham, M. C. & Horvath, S. Is my network module preserved and reproducible? *PLoS Comput. Biol.* **7**, (2011).
59. Zheng, G. X. Y. *et al.* Massively parallel digital transcriptional profiling of single cells. *Nat. Commun.* **8**, (2017).
60. Frankish, A. *et al.* GENCODE reference annotation for the human and mouse genomes. *Nucleic Acids Res.* **47**, D766–D773 (2019).
61. Stuart, T. *et al.* Comprehensive Integration of Single-Cell Data. *Cell* **177**, 1888–1902.e21 (2019).
62. Alvarez, M. *et al.* Enhancing droplet-based single-nucleus RNA-seq resolution using the semi-supervised machine learning classifier DIEM. *bioRxiv* 786285 (2019).
63. Habib, N. *et al.* Massively parallel single-nucleus RNA-seq with DroNc-seq. *Nat. Methods* **14**, 955–958 (2017).

64. Loh, P. *et al.* Efficient Bayesian mixed-model analysis increases association power in large cohorts. *Nat. Publ. Gr.* **47**, (2015).
65. Bulik-Sullivan, B. *et al.* LD score regression distinguishes confounding from polygenicity in genome-wide association studies. *Nat. Genet.* **47**, 291–295 (2015).
66. Finucane, H. K. *et al.* Partitioning heritability by functional annotation using genome-wide association summary statistics. *Nat. Genet.* **47**, 1228–1235 (2015).
67. Auton, A. *et al.* A global reference for human genetic variation. *Nature* **526**, 68–74 (2015).
68. Vilhjálmsson, B. J. *et al.* Modeling Linkage Disequilibrium Increases Accuracy of Polygenic Risk Scores. *Am. J. Hum. Genet.* **97**, 576–592 (2015).
69. Wainberg, M. *et al.* Opportunities and challenges for transcriptome-wide association studies. *Nat. Genet.* **51**, 592–599 (2019).
70. Liao, Y., Smyth, G. K. & Shi, W. FeatureCounts: An efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* **30**, 923–930 (2014).
71. Robinson, M. D., McCarthy, D. J. & Smyth, G. K. edgeR: A Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **26**, 139–140 (2009).
72. Law, C. W., Chen, Y., Shi, W. & Smyth, G. K. Voom: Precision weights unlock linear model analysis tools for RNA-seq read counts. *Genome Biol.* **15**, 1–17 (2014).
73. Ritchie, M. E. *et al.* Limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res.* **43**, e47 (2015).
74. Stegle, O., Parts, L., Piipari, M., Winn, J. & Durbin, R. Using probabilistic estimation of expression residuals (PEER) to obtain increased power and interpretability of gene expression analyses. *Nat. Protoc.* **7**, 500–507 (2012).
75. Das, S. *et al.* Next-generation genotype imputation service and methods. *Nat. Genet.* **48**, 1284–1287 (2016).
76. Shabalin, A. A. Matrix eQTL: Ultra fast eQTL analysis via large matrix operations. *Bioinformatics* **28**, 1353–1358 (2012).
77. Arribas, J. *et al.* NF- $\kappa$ B mediates the expression of TBX15 in cancer cells. *PLoS One* **11**, 1–14 (2016).
78. Leek, J. T., Johnson, W. E., Parker, H. S., Jaffe, A. E. & Storey, J. D. The SVA package for removing batch effects and other unwanted variation in high-throughput experiments. *Bioinformatics* **28**, 882–883 (2012).
79. Leek, J. T. Svaseq: Removing batch effects and other unwanted noise from sequencing data. *Nucleic Acids Res.* **42**, e161 (2014).
80. Phipson, B., Lee, S., Majewski, I. J., Alexander, W. S. & Smyth, G. Robust Hyperparameter Estimation Protects. *Ann. Appl. Stat.* **10**, 946–963 (2016).

## **Methods**

### **Study Cohorts**

#### **METabolic Syndrome In Men (METSIM)**

The participants in the METabolic Syndrome In Men (METSIM) cohort (n=10,197) are Finnish males recruited at the University of Eastern Finland and Kuopio University Hospital, Kuopio, Finland, as described previously<sup>44-46</sup>. The study was approved by the local ethics committee and all participants gave written informed consent. The median age of the METSIM participants is 57 years (range: 45-74 years). The METSIM participants were genotyped using the OmniExpress (Illumina) genotyping array and phased and imputed using SHAPEIT2 v2.17<sup>47</sup> and IMPUTE2 v2.3.2<sup>48</sup>, respectively. A random subset of the METSIM men underwent an abdominal subcutaneous adipose needle biopsy, with 335 unrelated individuals (IBD<0.2 using a genetic relationship matrix calculated in PLINK v1.9<sup>49</sup>) analyzed here using RNA-seq<sup>46</sup>.

#### **UKBiobank (UKB)**

The UKB is a large cohort (n=502,617) consisting of data from individuals collected across the United Kingdom starting in 2006<sup>27,50</sup>. To avoid hidden confounders from ancestry and relatedness, we used the subset of these individuals who are unrelated and of European ancestry (n=392,551). The genotyping was performed using one of two arrays for over 800,000 different variants<sup>27,51</sup>. The genotypes were then imputed using the Haplotype Reference Consortium (HRC) as well as UK 10K panel and the 1000 Genomes panel<sup>27,51</sup>. The genotypes were filtered for variants with MAF<1% and violation of Hardy-Weinberg Equilibrium ( $p < 1 \times 10^{-6}$ ) before using them for construction of the polygenic risk scores (PRSs) for WHRadjBMI.

#### **Kuopio OBesity Study (KOBS)**

The participants in the longitudinal Kuopio OBesity Study (KOBS) cohort (n=168) consist of Finnish obese individuals undergoing bariatric surgery and participating in a one year follow-up, recruited at the University of Eastern Finland and Kuopio University Hospital, Kuopio, Finland, as described previously<sup>52-54</sup>. The study was approved by the local ethics committee and all participants gave written informed consent. All participants underwent a pre-screening for a detailed medical history, and the inclusion criterion was a pre-surgery BMI of  $\geq 40$  kg/m<sup>2</sup> or 35 kg/m<sup>2</sup> with a significant comorbidity, such as type 2 diabetes (T2D). The biopsy samples were taken from subcutaneous adipose tissue at the time of bariatric surgery and one year after the

surgery. Refined phenotypic measurements and clinical characteristics were also measured at both time points<sup>52,53</sup>.

### **Alignment of RNA-seq data**

We performed alignment of subcutaneous adipose RNA-seq data (n=335) from the METSIM cohort using STAR v2.5.2<sup>55</sup> with GENCODE v19 annotation of the genome and hg19 version of the human genome, as we described earlier with minor changes<sup>46,56</sup>. Briefly, a 2-pass alignment was performed on 75 base-pair (bp) reads with only uniquely mapped reads counted for gene expression. We discovered that the expression of many genes and technical factors are correlated with the percentage of mitochondrial reads. To avoid the influence of the mitochondrial read number on the data, we excluded the mitochondrial reads from the RNA-seq data when calculating the FPKMs and technical factors. We used FastQC to verify the RNA-seq quality, based on metrics, such as GC content, duplication levels, and sequence quality scores, as well as Picard Tools v2.9.0 to obtain the technical factors from the standard RNA-seq metrics (option CollectRNAseqMetrics), including the median 5' to 3' bias, percentage of intronic reads, and median coverage from the aligned reads.

### **Weighted Gene Co-expression Analysis**

To find co-expression networks in the METSIM adipose RNA-seq cohort, we performed Weighted Gene Co-expression Analysis (WGCNA) v1.68<sup>57</sup> on FPKMs from the subcutaneous adipose RNA-seq data (n=335) from the METSIM cohort. To prevent the influence of technical factors from sequencing and RNA-seq alignment, we included 14 technical factors that were determined by STAR v2.5.2<sup>55</sup> and Picard Tools v2.9.0. The FPKMs were filtered for genes expressed (FPKM>0) in at least 90% of individuals and inverse normal transformed after correcting for technical factors to avoid spurious associations and outlier effects (see above). Phenotypes used for associations with co-expression networks in WGCNA v 1.68<sup>57</sup> were inverse normal transformed after correcting for age, age<sup>2</sup>. The fasting serum insulin levels were corrected for T2D status as well as age and age<sup>2</sup> and then inverse normal transformed. To ensure scale-free network topography, we used a power of 10 for the power function to determine co-expression network membership. All other parameters in WGCNA v 1.68<sup>57</sup> were kept at their default values.

### **Co-expression network preservation**

Using WGCNA v1.68<sup>57,58</sup>, we confirmed the preservation of the co-expression networks from the METSIM subcutaneous adipose RNA-seq (n=335) in the subcutaneous adipose RNA-seq (n=581) from the independent GTEx v8 cohort<sup>25</sup>. We further subdivided the GTEx v8<sup>25</sup> cohort to males (n=387) and females (n=194) and then lean (BMI<25) and obese (BMI>30) individuals of each sex. As the sample sizes for males and females were above the recommended minimum threshold (n=20), the  $Z_{\text{Summary}}$  score value should not be sensitive to the sample size, and so the relative difference in the number of males and females was not a concern. We calculated FPKMs from the RNA-seq data and technical factors from STAR v2.5.2<sup>55</sup> and Picard Tools v2.9.0, as described above. We corrected the expression data for technical factors as well as age, age<sup>2</sup>, sex, race, RIN, sequencing platform, sequencing protocol (PCR-based or PCR-free), and time from death to RNA collection and then inverse normal transformed the data. Default parameters in WGCNA v1.68<sup>57</sup> were used for the co-expression network preservation analysis. Accordingly, a preservation  $10 > Z_{\text{Summary}} > 2$  was considered as weakly to moderately preserved and a  $Z_{\text{Summary}} > 10$  as strongly preserved<sup>57,58</sup>.

### **Single-nucleus RNA-seq (snRNA-seq) of human subcutaneous adipose tissue**

We performed snRNA-seq of frozen adipose from sixteen individuals in order to identify cell types and their gene expression profile. Nuclei were isolated from frozen subcutaneous adipose tissue to input them into the 10X Chromium platform<sup>59</sup>. To isolate nuclei from frozen tissue, the tissue was minced over dry ice and transferred into ice-cold lysis buffer consisting of 0.1% IGEPAL, 10mM Tris-HCl, 10 mM NaCl, and 3 mM MgCl<sub>2</sub>. After a 10-minute incubation period, the lysate was gently homogenized using a dounce homogenizer and filtered through a 70  $\mu\text{m}$  MACS smart strainer (Miltenyi Biotec #130-098-462) to remove debris. Nuclei were centrifuged at 500 g for 5 minutes at 4°C and washed in 1 ml of resuspension buffer (RSB) consisting of 1X PBS, 1.0% BSA, and 0.2 U/ $\mu\text{l}$  RNase inhibitor. We further filtered nuclei using a 40  $\mu\text{m}$  Flowmi cell strainer (Sigma Aldrich # BAH136800040) and centrifuged at 500 g for 5 minutes at 4°C. Pelleted nuclei were re-suspended in wash buffer and immediately processed with the 10X Chromium platform following the Single Cell 3' v2 protocol. After library generation with the 10X Genomics platform, libraries were sequenced on an Illumina NovaSeq S2 at a sequencing depth of 50,000 reads per cell. Reads were aligned to the GRCh38 human

genome reference with Gencode v26 gene annotations<sup>60</sup> using the 10X CellRanger 2.1.1 pipeline. A custom pre-mRNA reference was generated to account for unspliced mRNA by merging all introns and exons of a gene into a single meta-exon.

### **SnRNA-seq data processing and identification of cell type marker genes**

We then clustered the droplets using Seurat v3.1.2<sup>61</sup>. In order to remove droplets contaminated with background RNA, we ran DIEM<sup>62</sup>. After applying filtering, we only considered droplets with at least 200 genes detected<sup>63</sup> to ensure that each droplet had enough information for clustering. The count data were log-normalized using the NormalizeData function in Seurat, using a scaling factor equal to the median of total counts across droplets. The counts for the sixteen adipose tissue samples were merged at this step. The top 2,000 variable genes were then calculated using the FindVariableFeatures function.

Normalized read counts for each gene were scaled to mean 0 and variance 1. We calculated the first 30 PCs to use them as input for clustering. We then ran the Seurat functions FindNeighbors and FindClusters with 30 PCs. In the FindClusters function, we used the default parameters with standard Louvain clustering and a default clustering resolution of 0.8. To identify marker genes for each cluster, we ran a Wilcoxon rank-sum test using the function FindAllMarkers with default parameters and only.pos=TRUE. We corrected for multiple testing using FDR<0.05.

### **T2D GWAS in the UKB**

To identify individuals with T2D in UKB, we selected the individuals who were diagnosed with diabetes (UKB data field 2443) or took medication for diabetes (data field 6153) as T2D cases, while removing the individuals with age of onset of diabetes (data field 25288) <40 years to avoid inclusion of type 1 diabetics in the GWAS analysis. We excluded the individuals with missing information for diagnosis of diabetes (data field 2443) from the GWAS analysis, and then used the individuals who were not diagnosed using these relevant data fields (data fields 2443, 6153, and 25288) as the controls. To account for population stratification, we selected the unrelated, Caucasians (total n after the exclusions=389,738) and used BOLT-LMM<sup>64</sup> to perform the GWAS associations between the genotypes and T2D status. We included age, age<sup>2</sup>, sex, array type, center ID, and 20 genotype PCs as covariates in the GWAS analysis.



### **Stratified LD score regression**

We performed Stratified LD score regression using the LD Score software v1.0.0<sup>65,66</sup>. This analysis was conducted using the GWAS summary statistics from the UKB and GIANT meta-analyses for WHRadjBMI (males, females, and both sexes combined) (n=315,284; n=379,501; n=694,549, respectively) and BMI (n=806,834 both sexes combined)<sup>15</sup> as well as GWAS summary statistics from the UKBiobank for T2D (males, females, and both sexes combined) (n=178,809; n=210,929; n=389,738, respectively). We partitioned the heritability into a category with the *cis*-regions (+/-500kb from the ends of the gene) around the 347 WHRadjBMI co-expression network genes and the 53 standard, overlapping categories used in the LD Score software v1.0.0<sup>65,66</sup>. Briefly, the 53 functional categories are derived from 26 main annotations that include coding regions, untranslated regions (UTRs), promoters, intronic regions, histone marks, DNase I hypersensitivity sites (DHSs), predicted enhancers, conserved regions, and other annotations. The partitioned LD Score regression method utilizes GWAS summary statistics of all variants to estimate how much variants in different annotation categories explain of the heritability of *cis* expression while accounting for the linkage disequilibrium (LD) among variants.

### **Construction of polygenic risk score**

We constructed the polygenic risk scores (PRSs) for WHRadjBMI using the same method for construction of PRSs as outlined for BMI in Khera *et al.*<sup>27</sup>. Briefly, we used the summary statistics from the GIANT GWAS for WHRadjBMI (n=224,459)<sup>8</sup> and a reference panel of the 503 European individuals from the 1000 Genomes phase 3 version 5<sup>67</sup>. We constructed nine candidate scores using the software, LDpred v1.0.6<sup>68</sup>, which adjusts the effect sizes for each variant in the GWAS based on LD structure. Due to the large number of participants, unified recruitment design and phenotypic characterization, the UKB is an ideal cohort for construction and testing of PRSs. Therefore, we tested and validated these candidate scores by dividing the UKB (Unrelated, Caucasian individuals, n=392,551)<sup>50,51</sup> into 2 groups: a testing set consisting of 1/3 of the individuals (n=130,851), and a validation set containing the remaining individuals unused in the testing set (n=261,700). Since the fraction of causal variants is not known *a priori*, we tested a different value of a tuning parameter ( $\rho=1, 0.3, 0.1, 0.03, 0.01, 0.003, 0.001, 0.0003, 0.0001$ ), as suggested by LDpred v1.0.6<sup>68</sup>, in each of our nine candidate scores. We selected the

best score by correlating the PRS with WHRadjBMI using Pearson correlation, which corresponded to  $\rho=0.01$ . We also compared this to five PRS scores constructed using the standard method of PRS construction of LD clumping (LD  $r^2<0.2$ ) and  $p$ -value thresholding ( $p<0.5, 0.1, 0.05, 1\times 10^{-5}, 5\times 10^{-8}$ ), as suggested by LDpred v1.0.6<sup>68</sup>, to confirm that using the tuning parameter constructed a superior PRS. To avoid the influence of technical factors, we corrected WHRadjBMI in the UKB for age, age<sup>2</sup>, sex, array type, center ID, and 20 genotype PCs. To perform statistical tests, we divided the PRS into 20 quantiles and calculated odds ratio of number of individuals in the top 10<sup>th</sup> percentile of WHRadjBMI for males and females separately.

### **Prediction of type 2 diabetes using the WHRadjBMI PRS**

We constructed a linear model to perform logistic regression using the binary T2D status as the outcome in the UKB validation set (n=261,700) that we originally employed to validate the PRSs for WHRadjBMI. We selected the individuals who were diagnosed with diabetes (UKB data field 2443) or took medication for diabetes (data field 6153) as T2D cases, while removing the individuals with age of onset of diabetes (data field 25288) <40 years to avoid inclusion of type 1 diabetics, with remaining individuals identified as controls. To examine individuals in the extremes of the WHRadjBMI spectrum, we selected the UKB participants in the highest (top 10% of network PRS scores) and lowest decile (lowest 10% of network PRS scores) of WHRadjBMI, as determined by the network PRS and divided them by sex. To avoid influence from the original phenotype, WHRadjBMI, as well as any technical factors, our linear model also included WHRadjBMI in addition to the network PRS score, with WHRadjBMI corrected for age, age<sup>2</sup>, sex, array type, center ID, and 20 genotype PCs. We performed a Wald test for the significance of each predictor in the linear model.

### **Transcriptome-Wide Association Studies (TWAS)**

To identify TFs causal for WHRadjBMI, we performed a targeted Transcriptome-wide Association Study (TWAS)<sup>33</sup> using GTEx v8 cohort's subcutaneous (n=581) RNA-seq data<sup>25</sup> to compute the TWAS weights for variants within the *cis*-region (+/-500kb from the ends of the gene) around the 14 TFs in the identified WHRadjBMI co-expression network. As there are not currently TWAS functional weights for genes using GTEx v8 cohort<sup>25</sup> and it has significantly

more samples than the GTEx v7 cohort<sup>18</sup> for adipose tissues, we computed our own weights using the recommended parameters by TWAS<sup>33</sup>. Briefly, to only include variants that will be used in the final association between TWAS and the GWAS trait, variants in the *cis*-region around our 14 TFs were pruned based on the LD reference panel from the TWAS website that was converted by matching variants from GRCh37 to GRCh38 in European individuals from the 1000 Genomes phase 3 version 5<sup>67</sup>. TWAS<sup>33</sup> checks the heritability ( $p < 0.01$ ) and then looks for the best model out of the five standard models to estimate weights for the variants to predict gene expression. To show that the genes computed by TWAS<sup>33</sup> are causal for a WHRadjBMI, we then associated the TWAS model with the weighted variants with WHRadjBMI using the GWAS summary statistics from the UKBiobank and GIANT meta-analysis<sup>15</sup>. The use of these extensive GWASs (total  $n \sim 700,000$  Europeans) should maximize power for association.

### Fine-mapping TWAS results using FOCUS

Recent work<sup>34,69</sup> has shown that TWAS signal at genomic risk regions will be correlated across genes as a result of linkage disequilibrium and prediction weights, which makes distinguishing non-relevant genes from their causal counterparts challenging. To adjust for the correlation in our TWAS test statistics and identify likely causal genes, we applied FOCUS<sup>34</sup>, a recently developed method that models the complete correlation structure within a region to fine-map TWAS signal. FOCUS models the state of genes as “causal” and “non-causal” and performs Bayesian inference over this state variable given the data. Specifically, given  $m$  TWAS z-scores  $\mathbf{z}$  at a genomic risk region, let  $\mathbf{\Sigma} = \mathbf{\Sigma}(\mathbf{W}, \mathbf{V})$  be the correlation structure of predicted expression as a function of the  $m \times p$  prediction weight matrix  $\mathbf{W}$  and the  $p \times p$  LD matrix  $\mathbf{V}$  and let  $\mathbf{c}$  be a binary vector indicating causal status. FOCUS models the likelihood of the calculated z-scores  $\mathbf{z}$  as,

$$\Pr(\mathbf{z} | \mathbf{W}, \mathbf{V}, \mathbf{c}, \sigma_{\alpha}^2) = N(\mathbf{0}, \mathbf{\Sigma} \mathbf{D}_{\mathbf{c}} \mathbf{\Sigma} + \mathbf{\Sigma})$$

where  $\mathbf{D}_{\mathbf{c}} = \text{diag}(\sigma_{\alpha}^2 \cdot \mathbf{c})$  is a diagonal matrix indicating which genes are causal weighted by the variance of their effect sizes. To infer the causal configuration  $\mathbf{c}$ , FOCUS computes the posterior probability as

$$\Pr(\mathbf{c} | \mathbf{z}, \mathbf{W}, \mathbf{V}, \sigma_{\alpha}^2) = \frac{\Pr(\mathbf{z} | \mathbf{W}, \mathbf{V}, \mathbf{c}, \sigma_{\alpha}^2) \Pr(\mathbf{c} | \theta)}{\sum_{\mathbf{c}'} \Pr(\mathbf{z} | \mathbf{W}, \mathbf{V}, \mathbf{c}', \sigma_{\alpha}^2) \Pr(\mathbf{c}' | \theta)}$$

To collapse the probability over configurations  $\mathbf{c}$  to individual genes FOCUS computes the marginal posterior inclusion probability (i.e. PIP) at the  $i$ th gene as  $\Pr(\mathbf{c}_i = 1 | \mathbf{z}, \mathbf{W}, \mathbf{V}, \sigma_{\alpha}^2) =$

$\sum_{c:c_i=1} \Pr(\mathbf{c} | \mathbf{z}, \mathbf{W}, \mathbf{V}, \sigma_{\alpha}^2)$ . Lastly, to reflect the inherent uncertainty of inference, FOCUS computes credible gene sets for a specified credible level. For example, a calibrated 90%-credible gene set contains the causal gene with probability 90%.

### **Differential gene expression analysis in the KOBS cohort**

Using read counts from featureCounts v2.0.0<sup>70</sup>, we performed differential expression (DE) analysis using the edgeR v3.24.3 package<sup>71</sup>. We first performed TMM normalization using the *calcNormFactors* and variance stabilization using voom<sup>72</sup>, and then built a linear model using LIMMA v3.38.3<sup>73</sup> with the blocking factor for the baseline and follow-up measurement time points in KOBS. As with the METSIM data, to avoid the influence of the mitochondrial read number on the data, we excluded the mitochondrial reads from the when obtaining technical factors. Technical factors were determined by STAR v2.5.2<sup>55</sup> and Picard Tools v2.9.0 (option CollectRNAseqMetrics) and included in the linear model in LIMMA v3.38.3<sup>73</sup>, with DE genes passing FDR<0.05 considered as significant.

### **Cis-eQTL analysis in the KOBS cohort**

We performed *cis*-eQTL analyses in the KOBS cohort at two time-points using the subcutaneous adipose RNA-seq data from the surgery and one-year follow-up (n=168 individuals with adipose RNA-seq data at both time points). Given the sample size, we used both sexes combined in the KOBS cohort to maximize power for discovery of context-specific *cis*-eQTLs. We filtered the subcutaneous adipose RNA-seq expression data (FPKMs) to genes expressed (FPKM>0) in greater than 90% of individuals and employed PEER factor<sup>74</sup> analysis to remove hidden confounders. We conducted PEER factor<sup>74</sup> optimization on chromosome 20 to maximize power for discovery for eQTLs, while ensuring hidden confounders were removed, and thus ended up correcting the KOBS expression data for 21 PEER factors. The KOBS cohort was genotyped using the OmniExpress (Illumina) genotyping array. We imputed genotypes using the Michigan Imputation Server<sup>75</sup> and filtered genotypes for variants MAF<5% and those failing Hardy-Weinberg Equilibrium test ( $p > 1 \times 10^{-6}$ ) using PLINK v1.9<sup>49</sup>. We performed *cis*-eQTL analysis using Matrix-eQTL<sup>76</sup>, classifying variants as *cis* if they were within 1Mb of either end of the gene.

### Human primary preadipocyte culture

Human subcutaneous primary white preadipocytes were obtained from Zen-Bio (lot L120116E, female, age 52, BMI 26.5) or PromoCell (lot 403Z001.1, male, age 30, BMI 30). Cells were maintained in a monolayer culture at 37°C and 5% CO<sub>2</sub> using preadipocyte growth medium (PromoCell C-27410) with 1% Gibco Penicillin-Streptomycin (ThermoFisher 15140122) and following PromoCell preadipocyte culturing protocols.

### Electrophoretic mobility shift assay

Nuclear protein was extracted from the human primary preadipocytes using the Nuclear Protein Extract Kit (Active Motif 40010), following manufacturer's protocols. We incubated 250fmol of oligonucleotide probes (15bp flanking SNP site for reference or alternate allele) with a biotin tag at the 5' end of the sequence (Integrated DNA Technologies) with 4 ug of the preadipocyte nuclear protein from the male or female donor, and the working reagent from the Gelshift Chemiluminescent EMSA kit (Active Motif 37341). For competitor assays, we added an unlabeled probe of the same sequence to the reaction mixture at 204x excess. The reaction was incubated for an hour at room temperature, and then loaded on a pre-run 6% retardation gel (ThermoFisher Scientific EC6365BOX) together with the EMSA kit 5X loading buffer. We ran 20µl in the gel with 0.5X TBE buffer at 120V. We then transferred the contents of the gel to a nylon membrane (Invitrogen LC2003) using 20V for 90 minutes and visualized with the chemiluminescent reagent as recommended.

EMSA oligonucleotide probe design for SNP rs1779445.

Oligonucleotide probe	DNA Strand	Sequence (5' -> 3')
Reference (T) allele	Positive	TGACAGTCTCCAACA <b>T</b> AACAGCTCAAAACTA
	Negative	TAGTTTTGAGCTGTTATGTTGGAGACTGTCA
Alternate (C) allele	Positive	TGACAGTCTCCAACA <b>C</b> AACAGCTCAAAACTA
	Negative	TAGTTTTGAGCTGTTGTGTTGGAGACTGTCA

\* Biotinylated probes were created by adding biotin to the 5' end of the positive strand probes.

### Small interfering RNA (si-RNA)-mediated knockdown of *TBX15*

We knocked down *TBX15* in human subcutaneous primary preadipocytes obtained from Zen-Bio (lot L120116E, female, age 52, BMI 26.5). For the siRNA transfection, we used the Dharmacon SMARTpool ON-TARGETplus Human *TBX15* siRNA (L-022116-02) and the Dharmacon siGENOME Non-Targeting siRNA Pool #1 (D-001206-13) as the negative control

(NC). We optimized the siRNA concentration and transfection volumes and then performed two independent siRNA transfection experiments in the human primary white preadipocytes. We used Invitrogen Lipofectamine RNAiMAX (ThermoFisher 13778150) to transfect 50 nM of the *TBX15* or NC siRNAs using reverse transfection. Specifically, we followed the manufacturer's instructions for diluting the siRNA and Lipofectamine RNAiMAX in Gibco Opti-MEM I Reduced Serum Medium (ThermoFisher 31985062) and forming the siRNA-Lipofectamine RNAiMAX complexes. We incubated cell suspensions in the complexes plus serum- and antibiotic-free media (PromoCell C-27417 basal media with supplement kit components minus the fetal calf serum) to a final siRNA concentration of 50 nM. We incubated the transfection reaction at room temperature for 10 minutes before plating 250  $\mu$ l per replicate into 12-well plates, for a total of 5 replicates per siRNA (*TBX15* and NC). After 24 hours of transfection, we added 1 ml of complete preadipocyte growth medium (PromoCell C-27410). 24 hours later, the media was removed and the cells were washed with PBS once prior to being treated with Invitrogen TRIzol reagent (ThermoFisher 15596026). We performed RNA extraction per manufacturer's protocol using the Direct-zol RNA Mini-Prep (Zymo Research R2061).

For the two independent knock-down experiments, we confirmed by RT-qPCR that *TBX15* expression was reduced by an average of >60% for the first experiment and 70% for the second experiment. We synthesized cDNA from 500 ng of RNA using the Applied Biosystems High-Capacity cDNA Reverse Transcription Kit (ThermoFisher Scientific 4368814). We measured relative gene expression by RT-qPCR using an Applied Biosystems QuantStudio 5 detector. To determine the relative percent of *TBX15* expression knockdown in the preadipocytes transfected with the *TBX15* siRNA compared to the NC siRNA, we normalized expression levels to *36B4*. Primers for *TBX15* were obtained from Arribas *et al.*<sup>77</sup>. and validated in-house. Primer sequences are listed below.

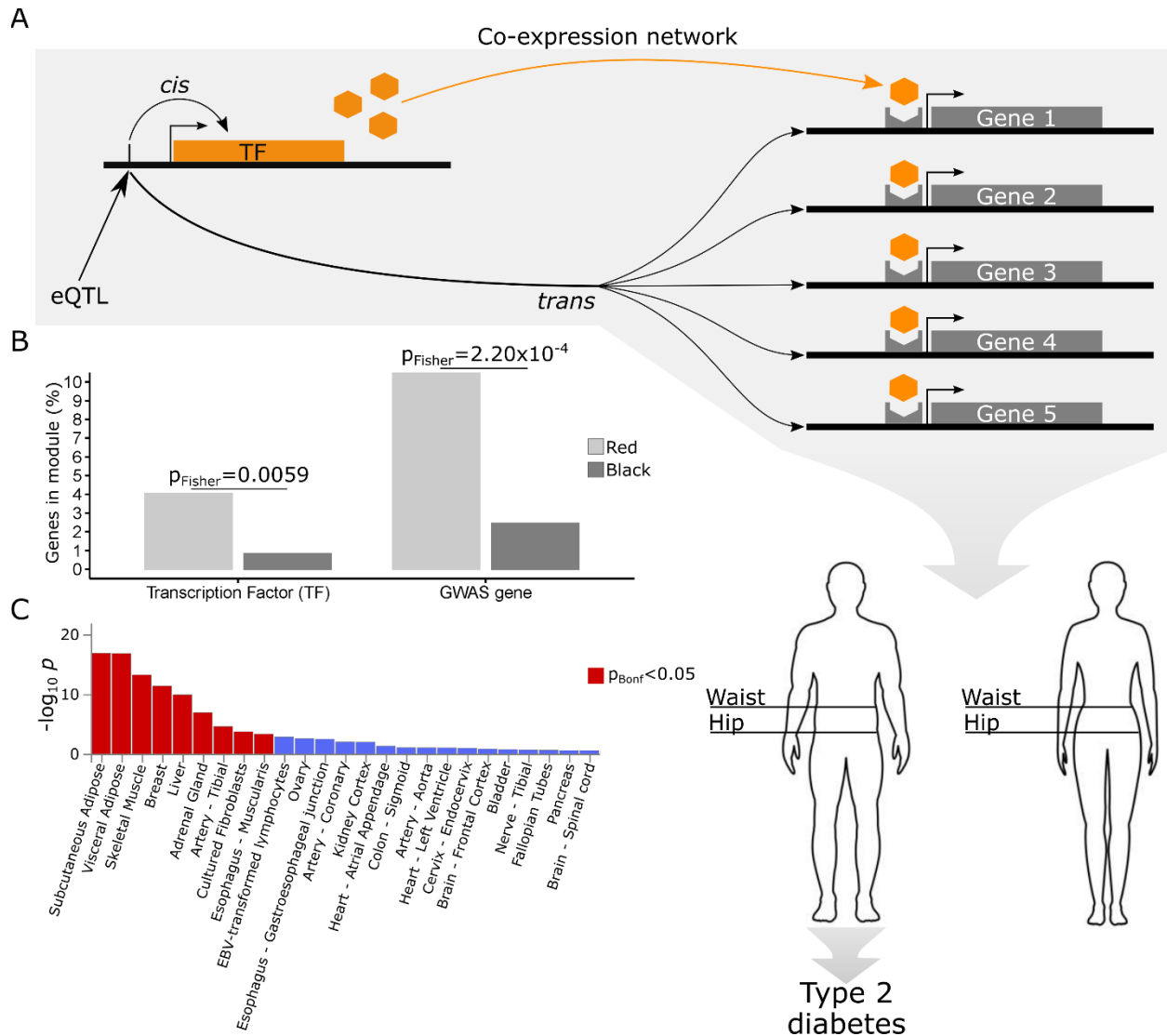
<b>Gene</b>	<b>Primer</b>	<b>Primer Sequence</b>
<i>TBX15</i>	Forward	5'- AAAGCAGGCAGGAGGATGTT-3'
	Reverse	5'- GCACAGGGGAATCAGCATTG-3'
<i>36B4</i>	Forward	5'-CCACGCTGCTGAACATGCT-3'
	Reverse	5'-TCGAACACCTGCTGGATGAC-3'

## **RNA-sequencing and differential expression analysis of si-RNA mediated knockdown of *TBX15***

We submitted the RNA samples from the experiment with an average of 70% knockdown for RNA-sequencing (RNA-seq). Libraries were prepared using the Illumina TruSeq Stranded mRNA kit and sequenced on an Illumina HiSeq 4000 instrument across 2 lanes for an average sequencing depth of 67M reads (+/- 2.5M reads) per sample. Reads were aligned to hg19 with STAR v2.7.0e<sup>55</sup>, using the 2-pass method and the following parameters: --outFilterMultimapNmax 1, --outFilterMismatchNmax 6, --alignIntronMin 20, --alignIntronMax 500000, --chimSegmentMin 15.

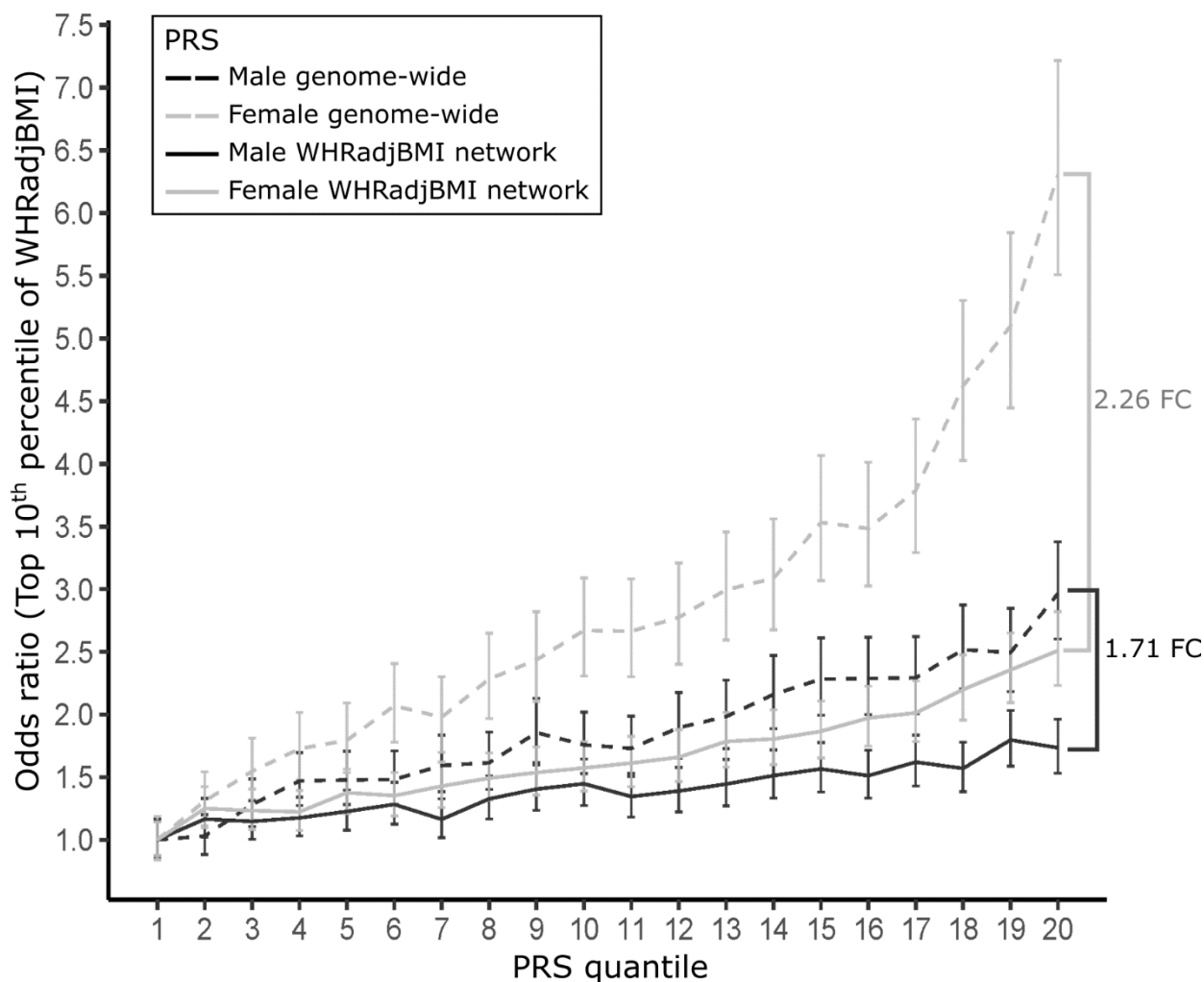
We used the R package sva v3.26.0<sup>78</sup> to estimate surrogate variables for unknown sources of variation in the data. We confirmed that the first surrogate variable (sv1) estimated using the svaseq<sup>79</sup> method is correlated with technical factors known to contribute to variance in RNA-seq data, such as library size, uniquely mapped read percent, and 3' bias, as well as the gene expression first principal component. The various technical factors were obtained from STAR v2.7.0e<sup>55</sup> after sequence alignment (uniquely mapped reads) or from the Picard Tools v2.9.0 (option CollectRnaSeqMetrics). We used the sv1 as a covariate in the differential expression (DE) analysis.

We performed the DE analysis using the R package limma v3.34.9<sup>73,80</sup> and the voom<sup>72</sup> method, including sv1 as a covariate, to identify genes in the WHRadjBMI co-expression network (n=347) that are significantly DE in the *TBX15* knockdown compared to the NC, with FDR<0.05 considered as significant.



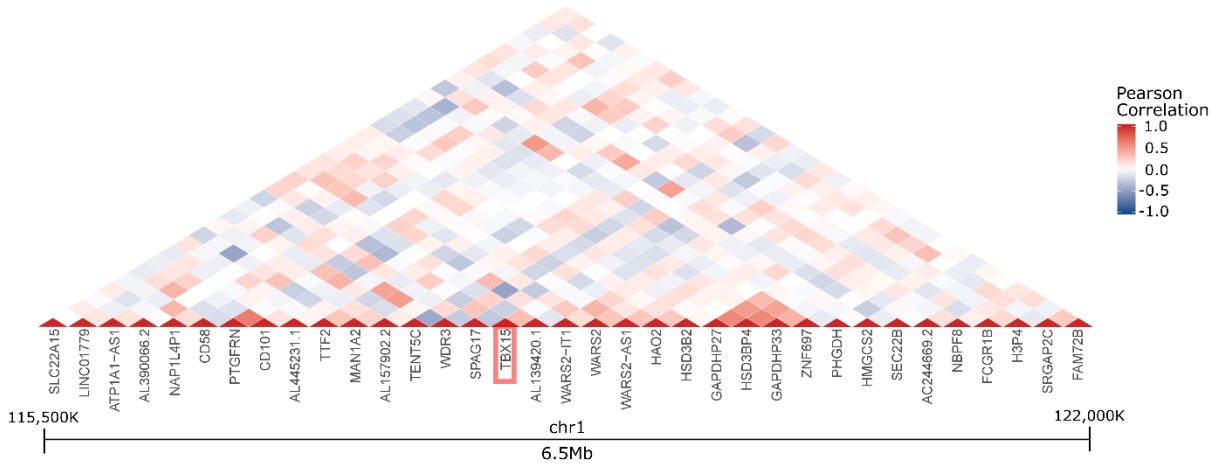
**Fig. 1 | Schematic overview of the study design (a), discovery of the red WHRadjBMI-associated co-expression network that is enriched for TFs and GWAS genes (b), and enriched for upregulated adipose tissue-specific DE genes when compared to other tissues (c) in GTEx<sup>25</sup>.** **a**, Illustrative schematic overview of the current study design, showing an eQTL controlling a TF, i.e. *TBX15*, in *cis*, and co-expression networks in *trans* via the TF, *TBX15*, that ultimately affects WHRadjBMI and clinical metabolic outcome, T2D, in a sex-dependent manner. **b**, Bar plot showing enrichment of TFs and GWAS genes in the red WHRadjBMI-associated co-expression network (light grey) when compared to the black WHRadjBMI-associated co-expression network (dark grey) using the Fisher's exact test. Significance of enrichment using the Fisher's exact test is indicated above each set of bars,  $p_{\text{Fisher}}$ . **c**, Bar plot showing significant enrichment (red) of upregulated adipose tissue-specific DE genes in WHRadjBMI co-expression network using FUMA<sup>26</sup> when compared to the 54 other tissues in the GTEx v8 cohort<sup>25</sup>. GTEx v8 tissues are ranked by enrichment from most enriched to least enriched with the first 25 most enriched tissues shown. The tissue enrichments passing a Bonferroni correction are shown in red, while the non-significant enrichments are shown in blue.



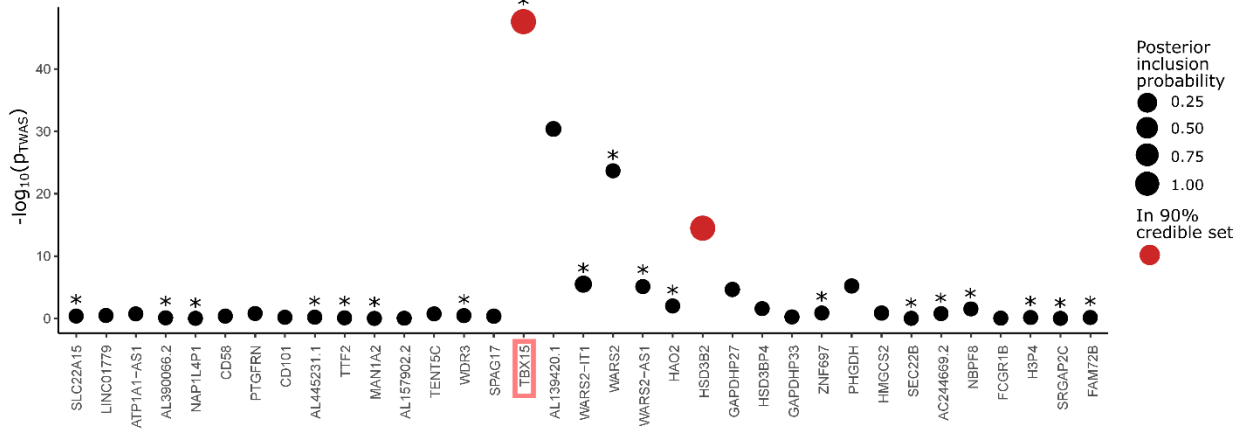


**Fig. 2 | PRS scores confirm sexual dimorphism of WHRadjBMI and demonstrate the importance of WHRadjBMI co-expression network genes for WHRadjBMI in males.** Plot of the PRS for WHRadjBMI in the testing set of the UKBiobank (n=261,700) separated for males (dark grey) and females (light grey) as well as for genome-wide PRS (dashed lines) and WHRadjBMI co-expression network PRS (solid lines; i.e. variants within the *cis* regions of the 347 network genes (+/-500kb from the ends of the gene)). Odds ratio is calculated based on the proportion of individuals in the top 10<sup>th</sup> percentile of WHRadjBMI for males and females in each of the 20 quantiles of the PRS separately. Vertical error bars indicate the 95% CI for the odds ratio. Brackets show a fold change (FC) in the odds ratio for the 20<sup>th</sup> quantile.

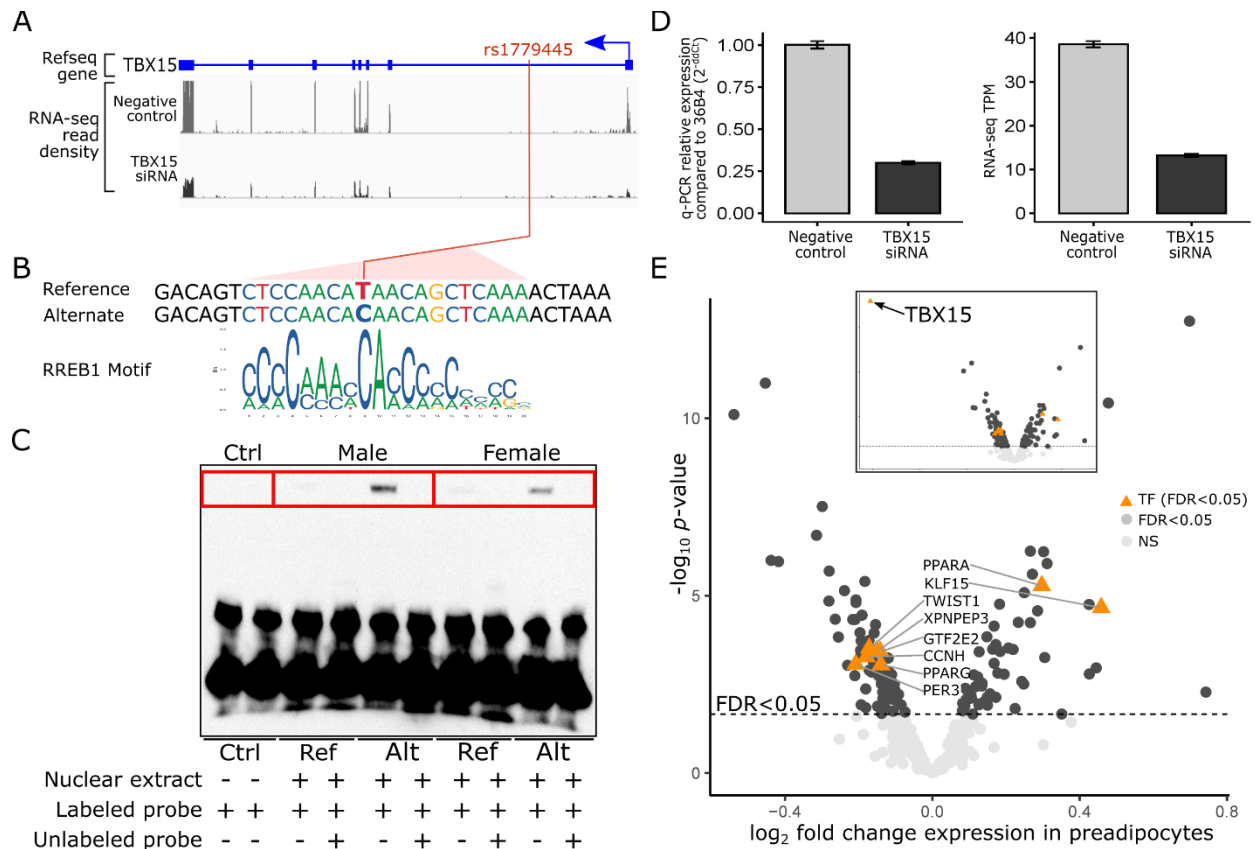
A



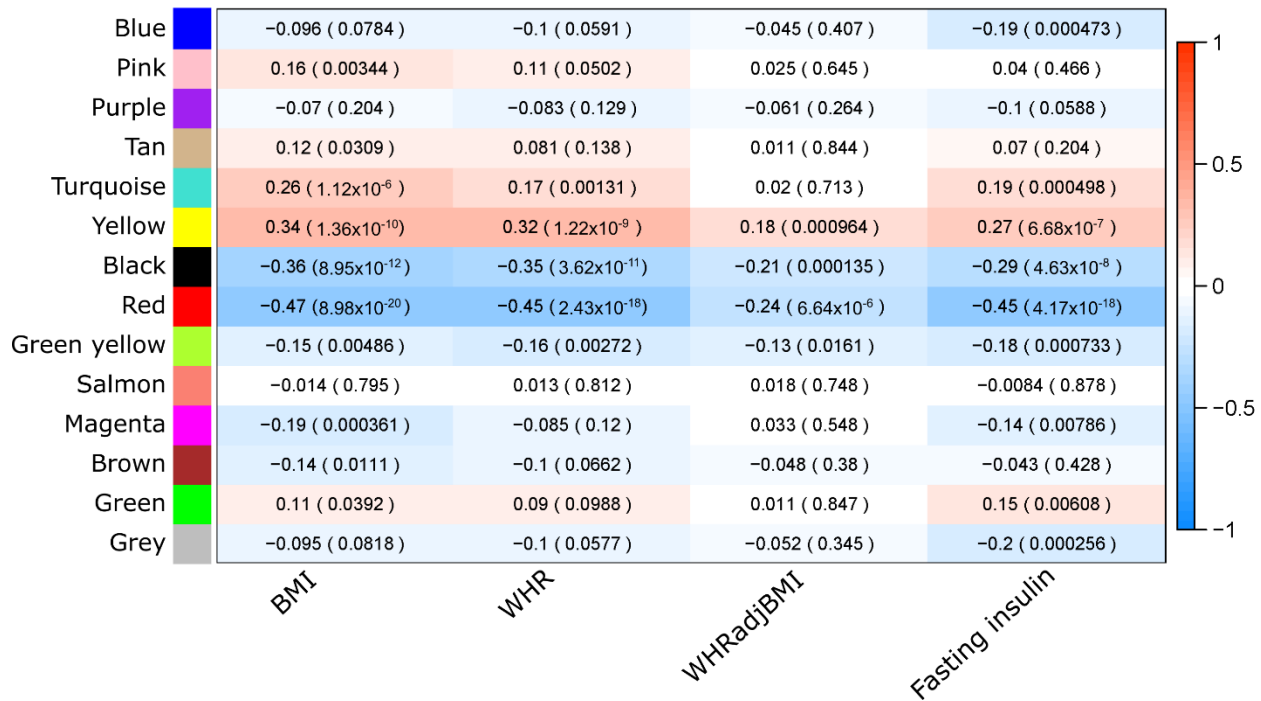
B



**Fig. 3 | TWAS and FOCUS results in GTEx v8 subcutaneous adipose RNA-seq data implicates *TBX15* as the only TF in the WHRadjBMI co-expression network causal for WHRadjBMI. a,** Pairwise Pearson correlation coefficients between all genes in the *TBX15* locus (chr1:115476504-121965583) using the normalized gene expression from the GTEx v8 cohort subcutaneous adipose RNA-seq data (n=581). **b,** Plot of  $-\log_{10} p$ -value for TWAS association with WHRadjBMI for each gene in the *TBX15* locus (chr1:115476504-121965583) with a significant heritability estimate ( $p < 0.01$ ) in the GTEx v8 cohort genotype and subcutaneous adipose RNA-seq data (n=581). Size of the point indicates the magnitude of the FOCUS marginal posterior inclusion probability (PIP). Genes included in the final 90% credible set are marked in red. Stars above points indicate a significant TWAS cross-validation  $p$ -value ( $p < 0.01$ ).



**Fig. 4 | The EMSA results demonstrate increased protein binding at the alternate allele of the variant rs1779445, a context-specific *cis* regulator of *TBX15* and *trans* regulator of the WHRadjBMI co-expression network (a, b, c) and knockdown of *TBX15* in human primary preadipocytes significantly affects 130 genes (FDR<0.05) in the WHRadjBMI co-expression network (d, e). a, Illustration of *TBX15* gene with introns and exons; and the relative RNA-seq read density in the human primary preadipocyte cells transfected with the negative control si-RNA when compared to the cells transfected with the *TBX15* si-RNA. Scales for the read density are equal. b, DNA sequence surrounding the SNP rs1779445 and canonical motif for RREB1 used for predicting binding at rs1779445. c, Representative EMSA (from n=3 independent experiments) shows protein binding at the reference and alternate allele of rs1779445 using nuclear extract from male and female preadipocytes. Control (ctrl); reference allele (ref), and alternate allele (alt); presence (+) and absence (-) of nuclear extract; as well as labeled probe and unlabeled probe (competitor) are indicated below each lane of the EMSA. d, Bar plot showing the qPCR relative expression ( $2^{-\Delta\Delta Ct}$ ) when compared to the housekeeping gene 36B4 and RNA-seq TPMs for *TBX15* in the cells transfected with negative control si-RNA when compared to the cells transfected with *TBX15* si-RNA (n=5). e, Volcano plot of differentially expressed (DE) genes in *TBX15* knockdown experiment, excluding *TBX15*. Significant genes (FDR<0.05) (dark grey), non-significant genes (light grey), and TFs (orange; FDR<0.05) are plotted based on their  $\log_{10} p$ -value and  $\log_2$  fold change in expression. Significantly differentially expressed TFs are labeled. Inlay shows the volcano plot of the *TBX15* DE results with *TBX15* included.**



**Supplementary Fig. 1 | WGCNA<sup>57</sup> identifies 2 co-expression networks in the METSIM adipose RNA-seq cohort (n=335), significantly correlated with WHRadjBMI and fasting serum insulin.** The numbers in the cells represent Pearson correlation results of network eigengenes with BMI, WHR, and WHRadjBMI, and fasting serum insulin (adjusted for T2D status) with correlation coefficients and *p*-values (shown in parenthesis). Associations that pass Bonferroni correction for the number of networks and traits tested ( $p_{\text{Bonf}} < 8.93 \times 10^{-4}$ ) were considered significant.

Supplementary Table 1. Characteristics of the genes (n=347) in the WHRadjBMI co-expression network, as reported by WGCNA<sup>57</sup> and ranked by network membership.

Gene Name	Chr*:start-end (hg19)	Network membership <sup>†</sup>	WHRadjBMI correlation <sup>‡</sup>	Fasting insulin correlation <sup>‡</sup>
HADH	chr4:108910870-108956331	0.929	-0.253	-0.439
ETFPA	chr15:76507696-76603813	0.898	-0.244	-0.454
ALDH6A1	chr14:74523553-74551196	0.895	-0.192	-0.399
HIBADH	chr7:27565061-27702614	0.874	-0.208	-0.367
UQCRC2	chr16:21963981-21994981	0.872	-0.246	-0.357
ACVR1C	chr2:158383279-158485517	0.871	-0.185	-0.402
GPD1L	chr3:32147181-32210205	0.866	-0.237	-0.481
BNIP3	chr10:133781578-133795435	0.855	-0.228	-0.361
MARC2	chr1:220921567-220958150	0.851	-0.192	-0.357
MUT	chr6:49398073-49430904	0.850	-0.166	-0.345
PCCA	chr13:100741269-101182686	0.846	-0.201	-0.356
CCDC50	chr3:191046866-191116459	0.845	-0.194	-0.373
ACSS3	chr12:81331594-81650533	0.840	-0.171	-0.405
DLST	chr14:75348594-75370448	0.839	-0.189	-0.353
PHF13	chr1:6673745-6684093	0.837	-0.218	-0.363
NIPSNAP3B	chr9:107526438-107539738	0.834	-0.217	-0.470
ANO6	chr12:45609770-45834187	0.833	-0.215	-0.366
C1orf43	chr1:154179182-154193104	0.833	-0.210	-0.347
VDAC2	chr10:76969912-76991206	0.830	-0.231	-0.375
NDUFB5	chr3:179322478-179345435	0.829	-0.225	-0.358
AUH	chr9:93976097-94124195	0.825	-0.176	-0.307
ACAT1	chr11:107992243-108018503	0.824	-0.149	-0.351
NAALAD2	chr11:89864683-89926062	0.824	-0.240	-0.408
PEX19	chr1:160246602-160256138	0.824	-0.167	-0.319
PRDX6	chr1:173446405-173457946	0.823	-0.261	-0.395
GBAS	chr7:56019486-56067874	0.820	-0.170	-0.283
HRSP12	chr8:99114572-99129469	0.815	-0.226	-0.376
ADH1B	chr4:100226121-100242558	0.814	-0.160	-0.441
CTH	chr1:70876901-70905534	0.811	-0.174	-0.374
MOCS1	chr6:39867354-39902290	0.809	-0.184	-0.451
SUCLG2	chr3:67410884-67705038	0.809	-0.203	-0.348
CSNK2A2	chr16:58191811-58231824	0.804	-0.256	-0.347
MRPL45	chr17:36452989-36479101	0.804	-0.208	-0.304
PHYH	chr10:13319796-13344412	0.804	-0.188	-0.323
DNAJC19	chr3:180701497-180707562	0.803	-0.131	-0.357
VEGFA	chr6:43737921-43754224	0.801	-0.155	-0.386
MRPL32	chr7:42971799-42988557	0.800	-0.149	-0.348
HSDL2	chr9:115142217-115234690	0.799	-0.167	-0.309
CYB5A	chr18:71920530-71959251	0.798	-0.211	-0.381
TCEB3	chr1:24069645-24088549	0.797	-0.226	-0.332

EIF4EBP2	chr10:72164135-72188374	0.795	-0.160	-0.371
HLF	chr17:53342373-53402426	0.794	-0.195	-0.381
SLC19A2	chr1:169433147-169455241	0.791	-0.184	-0.338
BFAR	chr16:14726672-14763093	0.789	-0.218	-0.336
PDK2	chr17:48172101-48189516	0.787	-0.165	-0.359
RP11-61A14.3	chr16:66923072-66924996	0.787	-0.132	-0.330
MRPS36	chr5:68513587-68525956	0.786	-0.210	-0.331
TWIST1	chr7:19060614-19157295	0.784	-0.160	-0.392
GKAP1	chr9:86354336-86444431	0.783	-0.182	-0.333
PFKFB3	chr10:6186881-6277495	0.783	-0.184	-0.390
TMEM132C	chr12:128751948-129192460	0.783	-0.118	-0.377
ACADM	chr1:76190036-76253260	0.780	-0.195	-0.360
LETMD1	chr12:51441745-51454207	0.779	-0.147	-0.350
MED9	chr17:17380300-17396540	0.779	-0.233	-0.347
GIN3	chr16:58328984-58440048	0.778	-0.231	-0.426
KLF15	chr3:126061478-126076285	0.775	-0.215	-0.300
MRPL10	chr17:45900638-45908900	0.775	-0.190	-0.325
RBPM5-AS1	chr8:30239635-30242917	0.775	-0.210	-0.379
UNG	chr12:109535379-109548797	0.775	-0.243	-0.310
CRLS1	chr20:5986736-6020699	0.774	-0.104	-0.314
TMLHE	chrX:154719776-154899605	0.774	-0.156	-0.288
LRPPRC	chr2:44113647-44223144	0.771	-0.177	-0.280
SLC27A2	chr15:50474393-50528592	0.771	-0.248	-0.407
TMEM100	chr17:53796988-53809482	0.770	-0.229	-0.336
PAIP2B	chr2:71409869-71454213	0.769	-0.237	-0.379
ORMDL3	chr17:38077294-38083854	0.768	-0.202	-0.401
OSBPL1A	chr18:21742008-21977844	0.768	-0.209	-0.343
AASS	chr7:121715701-121784334	0.767	-0.138	-0.393
ATPAF1	chr1:47098409-47139539	0.767	-0.109	-0.283
ASH2L	chr8:37962760-38001594	0.764	-0.104	-0.297
PPARA	chr22:46546424-46639653	0.762	-0.160	-0.362
TXLNG	chrX:16804550-16862642	0.762	-0.160	-0.304
GHR	chr5:42423879-42721979	0.761	-0.100	-0.364
TTL7	chr1:84330711-84464833	0.760	-0.194	-0.360
GMCL1	chr2:70056774-70108528	0.759	-0.208	-0.307
PER3	chr1:7844380-7905237	0.759	-0.186	-0.332
RMND1	chr6:151725989-151773259	0.759	-0.226	-0.289
PJA1	chrX:68380694-68385636	0.755	-0.137	-0.234
ANKRD53	chr2:71205510-71212626	0.754	-0.190	-0.411
NMNAT3	chr3:139279022-139396859	0.753	-0.113	-0.261
GPR146	chr7:1084212-1098897	0.751	-0.182	-0.343
AC003986.6	chr7:19152097-19153894	0.750	-0.193	-0.327
GLUL	chr1:182350839-182361341	0.750	-0.206	-0.371
FAM89A	chr1:231154704-231175992	0.749	-0.153	-0.391
APMAP	chr20:24943561-24973615	0.748	-0.245	-0.310

DHTKD1	chr10:12110971-12165224	0.747	-0.215	-0.250
IMMP2L	chr7:110303110-111202573	0.747	-0.216	-0.302
FDFT1	chr8:11653082-11696818	0.746	-0.160	-0.374
DBT	chr1:100652475-100715390	0.745	-0.127	-0.293
USP30	chr12:109460894-109525831	0.743	-0.159	-0.229
ABHD5	chr3:43731605-43775863	0.741	-0.161	-0.351
SLC35G2	chr3:136537489-136574734	0.741	-0.191	-0.353
TGDS	chr13:95226308-95248511	0.741	-0.174	-0.304
EPB41L4B	chr9:111934255-112083244	0.740	-0.176	-0.353
SDHD	chr11:111957497-111990353	0.740	-0.185	-0.307
GPHN	chr14:66974125-67648520	0.739	-0.204	-0.359
PRKAR2B	chr7:106685094-106802256	0.739	-0.142	-0.372
LACTB2	chr8:71547553-71581409	0.737	-0.229	-0.295
YPEL5	chr2:30369807-30383399	0.737	-0.224	-0.323
MCCC2	chr5:70883115-70954531	0.736	-0.170	-0.216
GRPEL1	chr4:7060633-7069924	0.735	-0.256	-0.356
DLD	chr7:107531415-107572175	0.734	-0.172	-0.281
FBXO27	chr19:39481354-39523425	0.734	-0.129	-0.335
TBX15	chr1:119425669-119532179	0.733	-0.155	-0.311
HSPD1	chr2:198351305-198381461	0.732	-0.203	-0.322
PCBD1	chr10:72642037-72648541	0.732	-0.234	-0.293
PMM1	chr22:41972898-41985894	0.730	-0.208	-0.348
MPDZ	chr9:13105703-13279589	0.729	-0.133	-0.296
SLC41A1	chr1:205758221-205782876	0.729	-0.206	-0.312
STOX1	chr10:70587298-70655188	0.729	-0.216	-0.365
DAPK2	chr15:64199235-64364232	0.728	-0.227	-0.375
MLYCD	chr16:83932731-83949787	0.728	-0.210	-0.293
TARSL2	chr15:102193801-102264807	0.728	-0.205	-0.335
BCKDHB	chr6:80816364-81055987	0.727	-0.145	-0.256
FAM120AOS	chr9:96208776-96215874	0.727	-0.120	-0.292
MLX	chr17:40719086-40725257	0.726	-0.266	-0.302
PMPCB	chr7:102937869-102969958	0.726	-0.140	-0.254
RP11-363E7.4	chr9:19453207-19455171	0.726	-0.109	-0.399
MTHFD1	chr14:64854749-64926722	0.725	-0.215	-0.297
TBC1D20	chr20:416124-443197	0.725	-0.201	-0.357
FAM13A	chr4:89647106-90032549	0.724	-0.140	-0.272
PHLPP1	chr18:60382672-60647666	0.724	-0.116	-0.386
ELP2	chr18:33709407-33757909	0.723	-0.210	-0.350
MCCC1	chr3:182733006-182833863	0.723	-0.135	-0.265
AC108142.1	chr4:182795591-183066402	0.722	-0.181	-0.361
HDDC2	chr6:125541108-125623282	0.721	-0.155	-0.284
MRPL44	chr2:224822121-224832431	0.721	-0.208	-0.280
SRSF4	chr1:29474255-29508499	0.719	-0.175	-0.260
SRP68	chr17:74035184-74068734	0.718	-0.201	-0.323
CCNH	chr5:86687311-86708836	0.716	-0.152	-0.304

PDP2	chr16:66912492-66929657	0.716	-0.098	-0.268
RCL1	chr9:4792869-4885917	0.716	-0.163	-0.299
SLC25A21-AS1	chr14:37641093-37643016	0.716	-0.145	-0.317
THYN1	chr11:134118173-134123264	0.716	-0.179	-0.301
RASL10B	chr17:34058668-34070540	0.715	-0.197	-0.419
MRPL39	chr21:26957968-26979829	0.714	-0.165	-0.282
SIX4	chr14:61176246-61191066	0.712	-0.204	-0.270
MARC1	chr1:220960101-220987735	0.711	-0.224	-0.297
ETFDH	chr4:159593277-159630775	0.710	-0.069	-0.286
GPN3	chr12:110890289-110907073	0.709	-0.166	-0.296
HOMEZ	chr14:23741666-23768656	0.708	-0.272	-0.316
MRPS22	chr3:138724648-139076065	0.707	-0.114	-0.331
BTG3	chr21:18965971-18985265	0.705	-0.136	-0.340
C11orf1	chr11:111749659-111756699	0.705	-0.188	-0.253
IARS2	chr1:220267444-220321380	0.705	-0.159	-0.241
HADHB	chr2:26466038-26513336	0.702	-0.124	-0.252
MRPS27	chr5:71515236-71616473	0.701	-0.110	-0.226
TRHDE-AS1	chr12:72647288-72668687	0.701	-0.135	-0.291
BMP3	chr4:81952119-81978685	0.700	-0.170	-0.425
SLC4A4	chr4:72053003-72437804	0.700	-0.187	-0.324
TTC36	chr11:118398187-118401912	0.700	-0.202	-0.403
ANAPC16	chr10:73975787-73995618	0.699	-0.221	-0.324
SLC19A3	chr2:228549926-228582728	0.698	-0.096	-0.413
RHOT1	chr17:30469473-30580393	0.697	-0.239	-0.287
RP11-61A14.2	chr16:66921918-66922834	0.697	-0.131	-0.283
SDHB	chr1:17345217-17380665	0.697	-0.163	-0.285
LRRC41	chr1:46726868-46769280	0.696	-0.151	-0.285
CECR2	chr22:17840837-18037850	0.695	-0.204	-0.442
LSM6	chr4:147096837-147121152	0.695	-0.201	-0.334
UTS2B	chr3:190984957-191048325	0.694	-0.246	-0.353
GYG2P1	chrY:14475147-14532255	0.691	-0.199	-0.347
NFU1	chr2:69622882-69664760	0.691	-0.173	-0.309
ZNF16	chr8:146155744-146176274	0.691	-0.177	-0.279
C12orf39	chr12:21679241-21690311	0.690	-0.252	-0.449
CFL2	chr14:35179593-35184029	0.690	-0.060	-0.283
PPP3R1	chr2:68405989-68483369	0.690	-0.207	-0.323
ABCB7	chrX:74273115-74376567	0.689	-0.154	-0.251
AKAP1	chr17:55162453-55198710	0.688	-0.138	-0.324
SNRNP27	chr2:70120692-70132707	0.688	-0.141	-0.248
TM7SF2	chr11:64879317-64883856	0.688	-0.149	-0.329
MAP3K5	chr6:136878185-137113656	0.686	-0.228	-0.359
ABHD15	chr17:27887565-27894155	0.684	-0.162	-0.283
AQP7	chr9:33384765-33402643	0.683	-0.153	-0.258
ISCA1	chr9:88879461-88897676	0.681	-0.136	-0.289
ADCK3	chr1:227085237-227175246	0.679	-0.165	-0.350



DNAJA3	chr16:4475806-4506776	0.679	-0.167	-0.267
L2HGDH	chr14:50704281-50779266	0.679	-0.182	-0.292
RP11-789C1.1	chr4:171195070-171204230	0.679	-0.169	-0.386
STRADB	chr2:202252581-202345569	0.679	-0.238	-0.290
ZFYVE21	chr14:104182067-104200005	0.679	-0.120	-0.267
GRSF1	chr4:71681499-71705662	0.678	-0.247	-0.239
RGS17	chr6:153325594-153452384	0.678	-0.167	-0.374
ACADSB	chr10:124768495-124817827	0.677	-0.085	-0.206
PTPN3	chr9:112137746-112260590	0.676	-0.163	-0.369
RDH10	chr8:74206847-74237516	0.676	-0.064	-0.260
ANG	chr14:21152336-21167130	0.674	-0.177	-0.386
C2orf47	chr2:200820040-200873263	0.673	-0.114	-0.247
WDR20	chr14:102605840-102691184	0.672	-0.094	-0.223
CIDEA	chr18:12254318-12277594	0.671	-0.161	-0.392
SETD9	chr5:56205087-56221359	0.670	-0.026	-0.315
HADHA	chr2:26413504-26467594	0.669	-0.177	-0.353
PPARG	chr3:12328867-12475855	0.666	-0.257	-0.308
RP11-387H17.4	chr17:38083995-38095854	0.666	-0.209	-0.405
AFG3L2	chr18:12328943-12377313	0.665	-0.144	-0.218
PDHX	chr11:34937376-35042138	0.665	-0.062	-0.260
GGCT	chr7:30536237-30591095	0.661	-0.189	-0.319
GSDMB	chr17:38060848-38076107	0.660	-0.258	-0.396
ISOC1	chr5:128430444-128449721	0.659	-0.161	-0.256
EIF1	chr17:39845137-39848920	0.658	-0.219	-0.351
SULF1	chr8:70378859-70573150	0.657	-0.185	-0.300
EYS	chr6:64429876-66417118	0.656	-0.140	-0.224
GTF2E2	chr8:30435835-30515768	0.656	-0.165	-0.292
SLC25A27	chr6:46620678-46645930	0.656	-0.123	-0.228
OXCT1	chr5:41730167-41870621	0.655	-0.122	-0.240
XPNPEP3	chr22:41253081-41363838	0.655	-0.118	-0.277
MKNK2	chr19:2037470-2051243	0.654	-0.220	-0.324
SIK2	chr11:111473115-111601577	0.653	-0.038	-0.252
CHCHD3	chr7:132469629-132766848	0.652	-0.194	-0.265
LONRF1	chr8:12579403-12613582	0.652	-0.208	-0.301
ZDHHC4	chr7:6617065-6629005	0.651	-0.139	-0.250
RPAIN	chr17:5322961-5336196	0.650	-0.197	-0.243
ARPC1A	chr7:98923521-98985787	0.648	-0.111	-0.292
CALCRL	chr2:188207856-188313187	0.648	-0.083	-0.286
HMGN3	chr6:79910962-79944406	0.648	-0.146	-0.263
TMEM220	chr17:10602332-10633633	0.647	-0.135	-0.291
BAG4	chr8:38034051-38070819	0.645	-0.116	-0.302
NDFIP2	chr13:80055287-80130210	0.645	-0.089	-0.320
TMEM52	chr1:1849029-1850712	0.644	-0.230	-0.356
GPR180	chr13:95254157-95286899	0.642	-0.186	-0.267
ACO1	chr9:32384618-32454767	0.638	-0.175	-0.250

PPP2R5A	chr1:212458879-212535200	0.638	-0.145	-0.303
RP11-61I13.3	chr6:39849580-39867847	0.638	-0.172	-0.357
MRPS18A	chr6:43639040-43655528	0.636	-0.116	-0.269
VWA8	chr13:42140973-42535256	0.635	-0.146	-0.180
URAHP	chr16:90106169-90114181	0.634	-0.188	-0.325
USP13	chr3:179370543-179507189	0.634	-0.202	-0.307
ACAD8	chr11:134123389-134135749	0.633	-0.130	-0.219
MRPS35	chr12:27863706-27909228	0.633	-0.151	-0.288
ZNF3	chr7:99661656-99680171	0.633	-0.127	-0.296
KCNIP2	chr10:103585731-103603677	0.631	-0.096	-0.271
STXBP1	chr9:130374544-130457460	0.631	-0.089	-0.338
CA3	chr8:86285665-86361269	0.630	-0.233	-0.236
EIF4EBP1	chr8:37887859-37917883	0.630	-0.191	-0.340
NRIP1	chr21:16333556-16437321	0.628	-0.111	-0.287
DNAH9	chr17:11501748-11873065	0.626	-0.093	-0.285
RP11-474O21.5	chr1:12678906-12679250	0.626	-0.171	-0.334
FRMD1	chr6:168456425-168482237	0.625	-0.198	-0.341
EIF1AY	chrY:22737611-22755040	0.622	-0.202	-0.312
KTN1-AS1	chr14:55965996-56046828	0.620	-0.149	-0.274
NDUFS1	chr2:206979541-207024327	0.620	-0.117	-0.209
ADRBK2	chr22:25960816-26125261	0.619	-0.168	-0.249
FHOD3	chr18:33877677-34360018	0.619	-0.231	-0.301
MRPS9	chr2:105654441-105716418	0.619	-0.253	-0.285
UQCC1	chr20:33890369-33999944	0.619	-0.139	-0.241
MPC1	chr6:166778407-166796486	0.618	-0.102	-0.323
CPT2	chr1:53662101-53679869	0.617	-0.118	-0.168
GPATCH11	chr2:37311594-37326387	0.617	-0.119	-0.197
VBPI	chrX:154425284-154468098	0.617	-0.184	-0.294
LRIG1	chr3:66429221-66551687	0.615	-0.117	-0.262
PAXIP1-AS1	chr7:154795158-154797413	0.613	-0.097	-0.221
PRDX3	chr10:120927215-120938345	0.612	-0.151	-0.206
TMEM25	chr11:118401756-118417995	0.612	-0.157	-0.417
ADH1A	chr4:100197524-100212185	0.609	-0.172	-0.344
AK4	chr1:65613232-65697828	0.609	-0.193	-0.283
PHGDH	chr1:120202421-120286838	0.609	-0.225	-0.250
TOMM70A	chr3:100082275-100120242	0.608	-0.124	-0.200
NDUFA5	chr7:123177051-123198309	0.606	-0.104	-0.293
RASSF6	chr4:74437267-74486348	0.603	-0.231	-0.377
IFT46	chr11:118415243-118443685	0.602	-0.128	-0.236
SLC16A7	chr12:59989848-60176395	0.602	-0.185	-0.267
LRRC47	chr1:3696784-3713068	0.601	-0.141	-0.319
RAI2	chrX:17818169-17879457	0.601	-0.060	-0.333
RP11-182I10.3	chr1:65437908-65468159	0.600	-0.050	-0.184
NEDD4L	chr18:55711599-56068772	0.597	-0.170	-0.318
SCO1	chr17:10583654-10601692	0.596	-0.116	-0.247

C17orf53	chr17:42219274-42239844	0.595	-0.184	-0.248
TP73-AS1	chr1:3652548-3663900	0.595	-0.061	-0.262
GLIS1	chr1:53971910-54199877	0.593	-0.208	-0.289
EMC3	chr3:10004221-10052800	0.592	-0.034	-0.304
MAN2A2	chr15:91445448-91465814	0.591	-0.182	-0.278
RP1-266L20.2	chr6:170125187-170125950	0.591	-0.145	-0.336
ANKRD46	chr8:101521980-101572012	0.590	-0.072	-0.247
MRS2	chr6:24403153-24425810	0.589	-0.114	-0.179
NDRG4	chr16:58496750-58547532	0.589	-0.174	-0.368
MRPL35	chr2:86426478-86440917	0.586	-0.124	-0.236
NKIRAS1	chr3:23933151-23988082	0.584	-0.119	-0.188
CENPV	chr17:16245848-16256970	0.582	-0.198	-0.300
TUSC1	chr9:25676396-25678856	0.582	-0.123	-0.316
C1orf50	chr1:43232940-43263968	0.581	-0.199	-0.289
PDHB	chr3:58413357-58419584	0.581	-0.129	-0.256
GFPT1	chr2:69546905-69614382	0.579	-0.096	-0.281
SYAP1	chrX:16737755-16783459	0.579	-0.093	-0.224
RP11-689P11.2	chr4:8483997-8514337	0.576	-0.120	-0.225
VPS72	chr1:151142463-151167797	0.576	-0.178	-0.294
ACO2	chr22:41865129-41924993	0.575	-0.113	-0.203
IRX1	chr5:3596168-3601517	0.574	-0.173	-0.263
ADAMTS9-AS2	chr3:64670585-64997143	0.572	-0.107	-0.284
GLYCTK	chr3:52321105-52329272	0.570	-0.202	-0.283
LONP2	chr16:48278207-48397033	0.567	-0.156	-0.193
SNX3	chr6:108532426-108582464	0.567	-0.091	-0.204
HDDC3	chr15:91474148-91475799	0.564	-0.194	-0.290
MET	chr7:116312444-116438440	0.562	-0.089	-0.310
RGS3	chr9:116207011-116360018	0.561	-0.236	-0.346
SLC43A1	chr11:57252007-57283259	0.561	-0.144	-0.170
CHKA	chr11:67820326-67888911	0.555	-0.214	-0.321
GHITM	chr10:85899196-85913001	0.555	-0.095	-0.214
GBE1	chr3:81538850-81811312	0.550	-0.052	-0.220
RTN3	chr11:63448918-63527363	0.544	-0.136	-0.225
TMEM230	chr20:5080486-5093749	0.540	-0.075	-0.218
ARSEP1	chrY:14460540-14468226	0.539	-0.100	-0.222
SIRT3	chr11:215458-236931	0.539	-0.084	-0.286
ADIPOQ	chr3:186560463-186576252	0.538	-0.071	-0.243
GABARAPL1	chr12:10365057-10375727	0.532	-0.039	-0.272
PRKAG2-AS1	chr7:151574127-151576299	0.532	-0.169	-0.261
ERCC8	chr5:60169658-60240900	0.531	-0.084	-0.228
TMEM42	chr3:44903361-44907162	0.523	-0.111	-0.198
ST6GALNAC6	chr9:130647600-130667687	0.517	-0.107	-0.372
SDHC	chr1:161284047-161332984	0.514	-0.059	-0.163
FBXO9	chr6:52916789-52965671	0.513	-0.119	-0.154
CDKN1C	chr11:2904443-2907111	0.503	-0.156	-0.268

FZD9	chr7:72848109-72850450	0.499	-0.112	-0.235
TSPAN3	chr15:77336359-77376326	0.498	-0.129	-0.136
FBXL5	chr4:15606162-15683302	0.495	-0.132	-0.163
SCOC	chr4:141178440-141306880	0.476	-0.021	-0.213
ATP5F1	chr1:111991486-112005395	0.474	-0.103	-0.159
HSPA9	chr5:137890571-137911133	0.426	-0.078	-0.150
IMMT	chr2:86371055-86422893	0.386	-0.057	-0.126
DSEL	chr18:65173819-65184217	-0.428	0.244	0.253
APBB1IP	chr10:26727132-26856732	-0.483	0.176	0.280
NEK6	chr9:127019885-127115586	-0.505	0.097	0.377
TSSC1	chr2:3192696-3381653	-0.506	0.108	0.259
GNG2	chr14:52292913-52446060	-0.531	0.201	0.334
CORO1C	chr12:109038885-109125372	-0.538	0.256	0.327
SPARC	chr5:151040657-151066726	-0.541	0.290	0.371
LBP	chr20:36974759-37005665	-0.552	0.132	0.384
CAPN1	chr11:64948037-64979477	-0.553	0.140	0.235
FAT2	chr5:150883654-150948505	-0.555	0.166	0.369
AP5S1	chr20:3801178-3805949	-0.576	0.274	0.291
KRT5	chr12:52908359-52914471	-0.576	0.233	0.389
TUBB2A	chr6:3153903-3157760	-0.587	0.208	0.303
NUDT1	chr7:2281857-2290781	-0.592	0.119	0.233
CA11	chr19:49141199-49149569	-0.600	0.169	0.249
TMEM189	chr20:48697663-48770335	-0.602	0.212	0.374
DPP3	chr11:66247484-66277130	-0.621	0.172	0.299
HPD	chr12:122277433-122301502	-0.626	0.220	0.480
PRAF2	chrX:48928813-48931730	-0.626	0.141	0.198
ACTN1	chr14:69340860-69446157	-0.632	0.211	0.379
TMSB10	chr2:85132749-85133795	-0.639	0.149	0.235
GNAI2	chr3:50263724-50296787	-0.640	0.155	0.223
CD248	chr11:66081958-66084515	-0.646	0.226	0.335
HOMER3	chr19:19040010-19052070	-0.647	0.193	0.265
TRPM2	chr21:45770046-45862964	-0.652	0.199	0.354
FLNA	chrX:153576892-153603006	-0.671	0.161	0.310
TMEM104	chr17:72772622-72835918	-0.673	0.165	0.360
ANKDD1A	chr15:65204101-65251042	-0.696	0.212	0.429
MSC	chr8:72753784-72756703	-0.725	0.287	0.361
C9orf16	chr9:130922539-130926207	-0.750	0.171	0.382

\* Abbreviation for chromosome.

† Pearson correlation coefficient with network eigengene as reported by WGCNA<sup>57</sup>.

‡ Pearson correlation coefficient with phenotype as reported by WGCNA<sup>57</sup>.

Supplementary Table 2. Characteristics of the obesity GWAS genes in WHRadjBMI co-expression network, as reported by WGCNA<sup>57</sup> and ranked by network membership.

Gene Name	GWAS trait*	Chr <sup>†</sup> :start-end (hg19)	Network membership <sup>‡</sup>	WHRadjBMI correlation <sup>§</sup>	Fasting insulin correlation <sup>§</sup>
PHF13	BMI	chr1:6673745-6684093	0.837	-0.218	-0.363
ADH1B	BMI	chr4:100226121-100242558	0.814	-0.160	-0.441
VEGFA	BMI WHR WHRadjBMI WCadjBMI	chr6:43737921-43754224	0.801	-0.155	-0.386
MRPL10	BMI	chr17:45900638-45908900	0.775	-0.190	-0.325
TTLL7	BMI	chr1:84330711-84464833	0.760	-0.194	-0.360
EPB41L4B	BMI WHR WHRadjBMI	chr9:111934255-112083244	0.740	-0.176	-0.353
DLD	WHRadjBMI	chr7:107531415-107572175	0.734	-0.172	-0.281
TBX15	BMI WHR WHRadjBMI WCadjBMI	chr1:119425669-119532179	0.733	-0.155	-0.311
BCKDHB	WHR WHRadjBMI WCadjBMI	chr6:80816364-81055987	0.727	-0.145	-0.256
FAM120AOS	BMI	chr9:96208776-96215874	0.727	-0.120	-0.292
FAM13A	WHR WHRadjBMI	chr4:89647106-90032549	0.724	-0.140	-0.272
MRPS22	BMI	chr3:138724648-139076065	0.707	-0.114	-0.331
LRRC41	BMI	chr1:46726868-46769280	0.696	-0.151	-0.285
CECR2	WHRadjBMI	chr22:17840837-18037850	0.695	-0.204	-0.442
ZFYVE21	BMI	chr14:104182067-104200005	0.679	-0.120	-0.267
RGS17	BMI WHR	chr6:153325594-153452384	0.678	-0.167	-0.374
SETD9	WHRadjBMI	chr5:56205087-56221359	0.670	-0.026	-0.315
PPARG	BMI WHR WHRadjBMI WC	chr3:12328867-12475855	0.666	-0.257	-0.308

EYS	BMI	chr6:64429876-66417118	0.656	-0.140	-0.224
ZDHHC4	BMI WHRadjBMI	chr7:6617065-6629005	0.651	-0.139	-0.250
CALCRL	BMI WHR WHRadjBMI	chr2:188207856-188313187	0.648	-0.083	-0.286
TMEM52	BMI	chr1:1849029-1850712	0.644	-0.230	-0.356
MRPS18A	WHRadjBMI	chr6:43639040-43655528	0.636	-0.117	-0.269
NDUFS1	BMI	chr2:206979541-207024327	0.620	-0.117	-0.209
MRPS9	BMI	chr2:105654441-105716418	0.619	-0.253	-0.285
UQCC1	WHR WHRadjBMI	chr20:33890369-33999944	0.619	-0.139	-0.241
LRIG1	BMI	chr3:66429221-66551687	0.615	-0.117	-0.262
ANKRD46	BMI	chr8:101521980-101572012	0.590	-0.072	-0.247
ADAMTS9-AS2	BMI WHR WHRadjBMI WCadjBMI	chr3:64670585-64997143	0.572	-0.107	-0.284
MET	WHRadjBMI	chr7:116312444-116438440	0.562	-0.089	-0.310
GBE1	BMI WC	chr3:81538850-81811312	0.550	-0.052	-0.220
ST6GALNAC6	BMI	chr9:130647600-130667687	0.517	-0.107	-0.372
GNAI2	BMI	chr3:50263724-50296787	-0.640	0.155	0.223
MSC	WHR WHRadjBMI WCadjBMI	chr8:72753784-72756703	-0.725	0.287	0.361
C9orf16	BMI	chr9:130922539-130926207	-0.750	0.171	0.382

\* Abbreviations for GWAS traits: Body Mass Index (BMI), Waist-hip-ratio (WHR), Waist-Circumference (WC), Waist-hip-ratio adjusted for BMI (WHRadjBMI), Waist-Circumference adjusted for BMI (WCadjBMI).

† Abbreviation for chromosome.

‡ Pearson correlation coefficient with network eigengene as reported by WGCNA<sup>57</sup>.

§ Pearson correlation coefficient with phenotype as reported by WGCNA<sup>57</sup>.

Supplementary Table 3. Characteristics of the adipocyte marker genes in WHRadjBMI co-expression network, as reported by WGCNA<sup>57</sup> and ranked by network membership.

Gene name	Chr*:start-end (hg19)	log <sub>2</sub> fold change <sup>†</sup>	Network membership <sup>‡</sup>	WHRadjBMI correlation <sup>§</sup>	Fasting insulin correlation <sup>§</sup>
ANO6	chr12:45609770-45834187	0.210	0.833	-0.215	-0.366
ADH1B	chr4:100226121-100242558	0.285	0.814	-0.160	-0.441
PFKFB3	chr10:6186881-6277495	0.239	0.783	-0.184	-0.390
TMEM132C	chr12:128751948-129192460	0.435	0.783	-0.118	-0.377
GHR	chr5:42423879-42721979	0.846	0.761	-0.100	-0.364
PRKAR2B	chr7:106685094-106802256	0.239	0.739	-0.142	-0.372
DAPK2	chr15:64199235-64364232	0.270	0.728	-0.227	-0.375
FAM13A	chr4:89647106-90032549	0.217	0.724	-0.140	-0.272
MARC1	chr1:220960101-220987735	0.208	0.711	-0.224	-0.297
SLC19A3	chr2:228549926-228582728	0.391	0.698	-0.096	-0.413
AQP7	chr9:33384765-33402643	0.792	0.683	-0.153	-0.258
PPARG	chr3:12328867-12475855	0.398	0.666	-0.257	-0.308
SIK2	chr11:111473115-111601577	0.583	0.653	-0.038	-0.252
KCNIP2	chr10:103585731-103603677	0.228	0.631	-0.096	-0.271
NEDD4L	chr18:55711599-56068772	0.227	0.597	-0.170	-0.318
GBE1	chr3:81538850-81811312	0.354	0.550	-0.052	-0.220
RTN3	chr11:63448918-63527363	0.225	0.544	-0.136	-0.225
ADIPOQ	chr3:186560463-186576252	0.362	0.538	-0.071	-0.243
SPARC	chr5:151040657-151066726	0.254	-0.541	0.290	0.371

\* Abbreviation for chromosome.

† Average log<sub>2</sub> fold change in adipocytes which compared to other cell types (see Methods).

‡ Pearson correlation coefficient with network eigengene as reported by WGCNA<sup>57</sup>.

§ Pearson correlation coefficient with phenotype as reported by WGCNA<sup>57</sup>.

Supplementary Table 4. KEGG pathway enrichment results (passing FDR<0.05) from WebGestalt<sup>24</sup> for the WHRadjBMI co-expression network genes.

Gene set	Description of the pathway	Enrichment Ratio	<i>p</i> -value	FDR
hsa00280	Valine, leucine, and isoleucine degradation	14.91	<2.2x10 <sup>-16</sup>	<2.2x10 <sup>-16</sup>
hsa00640	Propanoate metabolism	15.43	3.19x10 <sup>-12</sup>	5.17x10 <sup>-10</sup>
hsa01200	Carbon metabolism	6.15	1.28x10 <sup>-9</sup>	1.38x10 <sup>-7</sup>
hsa00020	Citrate cycle (TCA cycle)	11.98	2.55x10 <sup>-8</sup>	2.07x10 <sup>-6</sup>
hsa01100	Metabolic pathways	1.94	5.58x10 <sup>-8</sup>	3.62x10 <sup>-6</sup>
hsa00630	Glyoxylate and dicarboxylate metabolism	11.05	3.23x10 <sup>-7</sup>	1.74x10 <sup>-5</sup>
hsa00071	Fatty acid degradation	7.80	1.48x10 <sup>-6</sup>	6.83x10 <sup>-5</sup>
hsa01212	Fatty acid metabolism	5.55	2.23x10 <sup>-4</sup>	9.03x10 <sup>-3</sup>
hsa00650	Butanoate metabolism	7.77	3.66x10 <sup>-4</sup>	0.0132
hsa03320	PPAR signaling pathway	4.28	0.00113	0.0366
hsa04932	Non-alcoholic fatty liver disease (NAFLD)	2.87	0.00149	0.0439



Supplementary Table 5. Gene Ontology cellular component enrichment results (passing FDR<0.05) from WebGestalt<sup>24</sup> for the WHRadjBMI co-expression network genes.

Gene set	Description of the cellular component	Enrichment Ratio	<i>p</i> -value	FDR
GO:005739	Mitochondrion	4.40	<2.2x10 <sup>-16</sup>	<2.2x10 <sup>-16</sup>
GO:0031967	Organelle envelope	3.26	<2.2x10 <sup>-16</sup>	<2.2x10 <sup>-16</sup>
GO:0031975	Envelope	3.26	<2.2x10 <sup>-16</sup>	<2.2x10 <sup>-16</sup>
GO:0044429	Mitochondrial part	5.64	<2.2x10 <sup>-16</sup>	<2.2x10 <sup>-16</sup>
GO:0005740	Mitochondrial envelope	4.57	<2.2x10 <sup>-16</sup>	<2.2x10 <sup>-16</sup>
GO:0031966	Mitochondrial membrane	4.77	<2.2x10 <sup>-16</sup>	<2.2x10 <sup>-16</sup>
GO:0019866	Organelle inner membrane	5.26	<2.2x10 <sup>-16</sup>	<2.2x10 <sup>-16</sup>
GO:0005759	Mitochondrial matrix	8.11	<2.2x10 <sup>-16</sup>	<2.2x10 <sup>-16</sup>
GO:0005743	Mitochondrial inner membrane	5.59	<2.2x10 <sup>-16</sup>	<2.2x10 <sup>-16</sup>
GO:0098798	Mitochondrial protein complex	7.32	<2.2x10 <sup>-16</sup>	<2.2x10 <sup>-16</sup>
GO:1990204	Oxidoreductase complex	9.88	2.22x10 <sup>-13</sup>	2.31x10 <sup>-11</sup>
GO:0044455	Mitochondrial membrane part	6.03	1.64x10 <sup>-12</sup>	1.56x10 <sup>-10</sup>
GO:0045239	Tricarboxylic acid cycle enzyme complex	30.16	8.34x10 <sup>-10</sup>	7.34x10 <sup>-8</sup>
GO:0045240	Dihydrolipoyl dehydrogenase complex	33.61	6.04x10 <sup>-9</sup>	4.94x10 <sup>-7</sup>
GO:0098800	Inner mitochondrial membrane protein complex	6.09	2.43x10 <sup>-7</sup>	1.86x10 <sup>-6</sup>
GO:0000313	Organellar ribosome	7.33	2.78x10 <sup>-7</sup>	1.87x10 <sup>-5</sup>
GO:0005761	Mitochondrial ribosome	7.33	2.78x10 <sup>-7</sup>	1.87x10 <sup>-5</sup>
GO:0043209	Myelin sheath	5.16	5.69x10 <sup>-7</sup>	3.62x10 <sup>-5</sup>
GO:0045252	Oxoglutarate dehydrogenase complex	37.34	1.45x10 <sup>-6</sup>	8.72x10 <sup>-5</sup>
GO:0032592	Integral component of mitochondrial membrane	7.52	2.77x10 <sup>-6</sup>	1.59x10 <sup>-4</sup>
GO:0098573	Intrinsic component of mitochondrial membrane	7.41	3.15x10 <sup>-6</sup>	1.66x10 <sup>-4</sup>
GO:0030062	Mitochondrial tricarboxylic acid cycle enzyme complex	32.01	3.33x10 <sup>-6</sup>	1.66x10 <sup>-4</sup>
GO:0045254	Pyruvate dehydrogenase complex	32.01	3.33x10 <sup>-6</sup>	1.66x10 <sup>-4</sup>
GO:0031304	Intrinsic component of mitochondrial inner membrane	10.32	4.19x10 <sup>-6</sup>	1.92x10 <sup>-4</sup>

GO:0031305	Integral component of mitochondrial inner membrane	10.32	4.19x10 <sup>-6</sup>	1.92x10 <sup>-4</sup>
GO:0000314	Organellar small ribosomal subunit	12.45	6.60x10 <sup>-6</sup>	2.80x10 <sup>-4</sup>
GO:0005763	Mitochondrial small ribosomal subunit	12.45	6.60x10 <sup>-6</sup>	2.80x10 <sup>-4</sup>
GO:0009295	Nucleoid	9.56	7.13x10 <sup>-6</sup>	2.81x10 <sup>-4</sup>
GO:0042645	Mitochondrial nucleoid	9.56	7.13x10 <sup>-6</sup>	2.81x10 <sup>-4</sup>
GO:0031968	Organelle outer membrane	3.77	4.37x10 <sup>-5</sup>	1.67x10 <sup>-3</sup>
GO:0019867	Outer membrane	3.73	4.86x10 <sup>-5</sup>	1.69x10 <sup>-3</sup>
GO:0005947	Mitochondrial alpha-ketoglutarate dehydrogenase complex	33.61	5.48 x10 <sup>-5</sup>	1.69 x10 <sup>-3</sup>
GO:0005749	Mitochondrial respiratory chain complex II, succinate dehydrogenase complex (ubiquinone)	33.61	5.48 x10 <sup>-5</sup>	1.69 x10 <sup>-3</sup>
GO:0045257	Succinate dehydrogenase complex (ubiquinone)	33.61	5.48 x10 <sup>-5</sup>	1.69 x10 <sup>-3</sup>
GO:0045273	Respiratory chain complex II	33.61	5.48 x10 <sup>-5</sup>	1.69 x10 <sup>-3</sup>
GO:0045281	Succinate dehydrogenase complex	33.61	5.48 x10 <sup>-5</sup>	1.69 x10 <sup>-3</sup>
GO:0045283	Fumarate reductase complex	33.61	5.48 x10 <sup>-5</sup>	1.69 x10 <sup>-3</sup>
GO:0005741	Mitochondrial outer membrane	3.89	6.48x10 <sup>-5</sup>	1.95 x10 <sup>-3</sup>
GO:1902494	Catalytic complex	1.83	8.09x10 <sup>-5</sup>	2.37 x10 <sup>-3</sup>
GO:0005777	Peroxisome	4.34	0.00011	2.94 x10 <sup>-3</sup>
GO:0042579	Microbody	4.34	0.00011	2.94 x10 <sup>-3</sup>
GO:0098803	Respiratory chain complex	5.33	0.00013	3.41x10 <sup>-3</sup>
GO:0005840	Ribosome	3.03	0.00065	0.0153
GO:0098796	Membrane protein complex	1.76	0.0015	0.0323

Supplementary Table 6. Stratified LD Score Regression<sup>65,66</sup> results for WHRadjBMI, T2D, and BMI using the *cis* variants (+/-500kb from the ends of the gene) of the WHRadjBMI co-expression network genes.

Trait	N	Prop SNPs*	Prop h <sup>2</sup> †	Prop h <sup>2</sup> SE‡	Enrichment	Enrichment SE	Enrichment p-value
WHRadjBMI – Combined	694,649	0.105	0.169	0.0153	1.61	0.146	4.90x10 <sup>-5</sup>
WHRadjBMI – Male	315,284	0.105	0.154	0.0148	1.46	0.140	1.52x10 <sup>-3</sup>
WHRadjBMI – Female	379,501	0.105	0.178	0.0163	1.69	0.155	2.11x10 <sup>-5</sup>
T2D – Combined	389,738	0.105	0.157	0.0203	1.49	0.193	9.56x10 <sup>-3</sup>
T2D – Male	178,809	0.105	0.161	0.0240	1.53	0.228	0.0177
T2D – Female	210,929	0.105	0.160	0.0304	1.52	0.289	NS
BMI – Combined	806,834	0.105	0.102	0.00560	0.97	0.0532	NS

\* Proportion of SNPs in the *cis*-regions (+/-500kb from the ends of the gene) of the WHRadjBMI co-expression network genes.

† Proportion of heritability explained by the variants in the *cis*-regions (+/-500kb from the ends of the gene) of the WHRadjBMI co-expression network genes.

‡ Standard error of the proportion of heritability.

Supplementary Table 7. Characteristics of the TFs in WHRadjBMI co-expression network, as reported by WGCNA<sup>57</sup> and ranked by network membership.

Gene name	Chr*:start-end (hg19)	Network membership <sup>†</sup>	WHRadjBMI correlation <sup>‡</sup>	Fasting insulin correlation <sup>‡</sup>
HLF	chr17:53342373-53402426	0.794	-0.195	-0.381
TWIST1	chr7:19060614-19157295	0.784	-0.160	-0.392
KLF15	chr3:126061478-126076285	0.775	-0.215	-0.300
PPARA	chr22:46546424-46639653	0.762	-0.160	-0.362
PER3	chr1:7844380-7905237	0.759	-0.186	-0.332
CCNH	chr5:86687311-86708836	0.716	-0.152	-0.304
SIX4	chr14:61176246-61191066	0.712	-0.204	-0.270
TBX15	chr1:119425669-119532179	0.712	-0.204	-0.311
HOMEZ	chr14:23741666-23768656	0.708	-0.272	-0.316
PPARG	chr3:12328867-12475855	0.666	-0.257	-0.308
GTF2E2	chr8:30435835-30515768	0.655	-0.165	-0.292
XPNPEP3	chr22:41253081-41363838	0.655	-0.118	-0.277
ZNF3	chr7:99661656-99680171	0.633	-0.127	-0.296
IRX1	chr5:3596168-3601517	0.574	-0.173	-0.263

\* Abbreviation for chromosome.

<sup>†</sup> Pearson correlation coefficient with network eigengene as reported by WGCNA<sup>57</sup>.

<sup>‡</sup> Pearson correlation coefficient with phenotype as reported by WGCNA<sup>57</sup>.

Supplementary Table 8. Significant TWAS<sup>33</sup> heritability estimates ( $p < 0.01$ ) for the TFs in the WHRadjBMI co-expression network.

Gene Name	Heritability	Heritability standard error	$p$ -value
TBX15	0.101	0.0313	$4.07 \times 10^{-6}$
GTF2E2	0.0703	0.0253	$1.43 \times 10^{-3}$
XPNPEP3	0.302	0.0602	$1.18 \times 10^{-38}$
IRX1	0.0813	0.0237	$1.04 \times 10^{-4}$
ZNF3	0.0821	0.0326	$1.85 \times 10^{-5}$

Supplementary Table 9. TWAS<sup>33</sup> *p*-values and Z-scores for associations of TFs (with significant TWAS<sup>33</sup> heritability (*p*<0.01)) with WHRadjBMI.

Gene Name	TWAS Model <sup>†</sup>	Z-score	<i>p</i> -value
TBX15	Bayesian sparse linear mixed models (bslmm)	15.2	2.11x10 <sup>-52</sup>
GTF2E2	Least absolute shrinkage and selection operator (lasso)	-0.103	NS*
XPNPEP3	Least absolute shrinkage and selection operator (lasso)	-3.05	2.26x10 <sup>-3</sup>
IRX1	Best linear unbiased predictor	4.61	4.03x10 <sup>-6</sup>
ZNF3	Least absolute shrinkage and selection operator (lasso)	2.18	NS*

\* NS indicates a non-significant Bonferroni corrected *p*-value>0.017.

† Best model for expression imputation chosen by TWAS<sup>33</sup>.

Supplementary Table 10. Significantly differentially expressed genes (FDR<0.05) in the WHRadjBMI co-expression network in the *TBX15* knockdown experiment ranked by *p*-value.

Gene name	Chr *:start-end (hg19)	log <sup>2</sup> fold change <sup>†</sup>	<i>p</i> -value	FDR
TBX15	chr1:119425669-119532179	-1.527	1.13x10 <sup>-18</sup>	3.35x10 <sup>-16</sup>
STRADB	chr2:202252581-202345569	0.698	1.81x10 <sup>-13</sup>	2.69x10 <sup>-11</sup>
EIF4EBP2	chr10:72164135-72188374	-0.454	1.02x10 <sup>-11</sup>	1.01x10 <sup>-9</sup>
MET	chr7:116312444-116438440	0.478	3.69x10 <sup>-11</sup>	2.74x10 <sup>-9</sup>
DSEL	chr18:65173819-65184217	-0.539	7.78x10 <sup>-11</sup>	4.62x10 <sup>-9</sup>
VEGFA	chr6:43737921-43754224	-0.299	3.09x10 <sup>-11</sup>	1.53x10 <sup>-6</sup>
STXBP1	chr9:130374544-130457460	-0.315	1.98x10 <sup>-7</sup>	8.42x10 <sup>-6</sup>
PHGDH	chr1:120202421-120286838	0.266	5.61x10 <sup>-7</sup>	1.92 x10 <sup>-5</sup>
ISCA1	chr9:88879461-88897676	0.302	5.82x10 <sup>-7</sup>	1.92 x10 <sup>-5</sup>
SNRNP27	chr2:70120692-70132707	-0.438	1.01x10 <sup>-6</sup>	2.95 x10 <sup>-5</sup>
IMMP2L	chr7:110303110-111202573	-0.418	1.09x10 <sup>-6</sup>	2.95 x10 <sup>-5</sup>
TMEM189	chr20:48697663-48770335	0.312	1.25x10 <sup>-6</sup>	3.09x10 <sup>-5</sup>
THYN1	chr11:134118173-134123264	-0.280	2.01x10 <sup>-6</sup>	4.59x10 <sup>-5</sup>
CFL2	chr14:35179593-35184029	0.271	2.48x10 <sup>-6</sup>	5.26x10 <sup>-5</sup>
UQCRC2	chr16:21963981-21994981	-0.184	3.94x10 <sup>-6</sup>	7.79x10 <sup>-5</sup>
PPARA	chr22:46546424-46639653	0.297	5.24x10 <sup>-6</sup>	9.72x10 <sup>-5</sup>
PCBD1	chr10:72642037-72648541	-0.239	7.22x10 <sup>-6</sup>	1.26 x10 <sup>-4</sup>
FLNA	chrX:153576892-153603006	0.249	8.17x10 <sup>-6</sup>	1.35x10 <sup>-4</sup>
CYB5A	chr18:71920530-71959251	-0.207	1.32x10 <sup>-5</sup>	2.07x10 <sup>-4</sup>
LONRF1	chr8:12579403-12613582	-0.281	1.41x10 <sup>-5</sup>	2.09x10 <sup>-4</sup>
GNG2	chr14:52292913-52446060	-0.207	1.55x10 <sup>-5</sup>	2.19x10 <sup>-4</sup>
APMAP	chr20:24943561-24973615	0.182	1.75x10 <sup>-5</sup>	2.30x10 <sup>-4</sup>
RAI2	chrX:17818169-17879457	0.426	1.78x10 <sup>-5</sup>	2.30x10 <sup>-4</sup>
KLF15	chr3:126061478-126076285	0.459	2.18x10 <sup>-5</sup>	2.69x10 <sup>-4</sup>
ABHD15	chr17:27887565-27894155	0.285	2.68x10 <sup>-5</sup>	3.19x10 <sup>-4</sup>
TXLNG	chrX:16804550-16862642	-0.191	3.54x10 <sup>-5</sup>	4.04x10 <sup>-4</sup>
BCKDHB	chr6:80816364-81055987	-0.264	4.63x10 <sup>-5</sup>	4.78x10 <sup>-4</sup>
HMGN3	chr6:79910962-79944406	-0.216	4.67x10 <sup>-5</sup>	4.78x10 <sup>-4</sup>
ACADSB	chr10:124768495-124817827	-0.210	4.66x10 <sup>-5</sup>	4.78x10 <sup>-4</sup>
TMEM42	chr3:44903361-44907162	0.266	5.81x10 <sup>-5</sup>	5.56x10 <sup>-4</sup>
MAN2A2	chr15:91445448-91465814	0.233	5.70x10 <sup>-5</sup>	5.56x10 <sup>-4</sup>
OSBPL1A	chr18:21742008-21977844	-0.155	6.51x10 <sup>-5</sup>	6.04x10 <sup>-4</sup>

SUCLG2	chr3:67410884-67705038	0.168	7.22x10 <sup>-5</sup>	6.50x10 <sup>-4</sup>
APBB1IP	chr10:26727132-26856732	-0.197	1.10x10 <sup>-4</sup>	9.59x10 <sup>-4</sup>
MKNK2	chr19:2037470-2051243	-0.158	1.13x10 <sup>-4</sup>	9.62x10 <sup>-4</sup>
AUH	chr9:93976097-94124195	-0.255	1.47x10 <sup>-4</sup>	1.18x10 <sup>-3</sup>
SRP68	chr17:74035184-74068734	0.148	1.43x10 <sup>-4</sup>	1.18x10 <sup>-3</sup>
GRPEL1	chr4:7060633-7069924	-0.173	1.83x10 <sup>-4</sup>	1.43x10 <sup>-3</sup>
IFT46	chr11:118415243-118443685	-0.193	1.87x10 <sup>-4</sup>	1.43x10 <sup>-3</sup>
GPD1L	chr3:32147181-32210205	0.173	2.38x10 <sup>-4</sup>	1.77x10 <sup>-3</sup>
TWIST1	chr7:19060614-19157295	-0.171	3.00x10 <sup>-4</sup>	2.12x10 <sup>-3</sup>
DHTKD1	chr10:12110971-12165224	0.207	2.98x10 <sup>-4</sup>	2.12x10 <sup>-3</sup>
PMPCB	chr7:102937869-102969958	-0.152	3.21x10 <sup>-4</sup>	2.16x10 <sup>-3</sup>
TMEM104	chr17:72772622-72835918	0.163	3.20x10 <sup>-4</sup>	2.16x10 <sup>-3</sup>
CTH	chr1:70876901-70905534	0.219	3.30x10 <sup>-4</sup>	2.17x10 <sup>-3</sup>
BTG3	chr21:18965971-18985265	-0.195	3.50x10 <sup>-4</sup>	2.26x10 <sup>-3</sup>
PFKFB3	chr10:6186881-6277495	0.184	3.72x10 <sup>-4</sup>	2.27x10 <sup>-3</sup>
XPNPEP3	chr22:41253081-41363838	-0.151	3.61x10 <sup>-4</sup>	2.27x10 <sup>-3</sup>
GTF2E2	chr8:30435835-30515768	-0.142	3.74x10 <sup>-4</sup>	2.27x10 <sup>-3</sup>
LRRC41	chr1:46726868-46769280	0.127	3.82x10 <sup>-4</sup>	2.27x10 <sup>-3</sup>
DPP3	chr11:66247484-66277130	-0.163	4.19x10 <sup>-4</sup>	2.44x10 <sup>-3</sup>
TARSL2	chr15:102193801-102264807	-0.195	4.50x10 <sup>-4</sup>	2.57x10 <sup>-3</sup>
ATPAF1	chr1:47098409-47139539	-0.188	4.90x10 <sup>-4</sup>	2.74x10 <sup>-3</sup>
CCNH	chr5:86687311-86708836	-0.179	5.28x10 <sup>-4</sup>	2.90x10 <sup>-3</sup>
FAM89A	chr1:231154704-231175992	0.305	5.53x10 <sup>-4</sup>	2.99x10 <sup>-3</sup>
TMEM230	chr20:5080486-5093749	-0.120	5.80x10 <sup>-4</sup>	3.07x10 <sup>-3</sup>
VPS72	chr1:151142463-151167797	-0.161	7.24x10 <sup>-4</sup>	3.77x10 <sup>-3</sup>
PPP2R5A	chr1:212458879-212535200	0.167	8.03x10 <sup>-4</sup>	4.04x10 <sup>-3</sup>
MRS2	chr6:24403153-24425810	-0.156	8.03x10 <sup>-4</sup>	4.04x10 <sup>-3</sup>
PER3	chr1:7844380-7905237	-0.208	8.65x10 <sup>-4</sup>	4.28x10 <sup>-3</sup>
PPARG	chr3:12328867-12475855	-0.140	8.86x10 <sup>-4</sup>	4.31x10 <sup>-3</sup>
PRAF2	chrX:48928813-48931730	-0.231	9.20x10 <sup>-4</sup>	4.41x10 <sup>-3</sup>
SRSF4	chr1:29474255-29508499	-0.159	9.98x10 <sup>-4</sup>	4.67x10 <sup>-3</sup>
SLC41A1	chr1:205758221-205782876	-0.155	1.01x10 <sup>-3</sup>	4.67x10 <sup>-3</sup>
DAPK2	chr15:64199235-64364232	0.445	1.09x10 <sup>-3</sup>	4.96x10 <sup>-3</sup>
ISOC1	chr5:128430444-128449721	-0.171	1.26x10 <sup>-3</sup>	5.65x10 <sup>-3</sup>



GPATCH11	chr2:37311594-37326387	-0.163	1.28 x10 <sup>-3</sup>	5.67x10 <sup>-3</sup>
NDUFB5	chr3:179322478-179345435	-0.123	1.39 x10 <sup>-3</sup>	6.06x10 <sup>-3</sup>
HIBADH	chr7:27565061-27702614	-0.117	1.46 x10 <sup>-3</sup>	6.20x10 <sup>-3</sup>
MRPL32	chr7:42971799-42988557	-0.165	1.44 x10 <sup>-3</sup>	6.20x10 <sup>-3</sup>
PHLPP1	chr18:60382672-60647666	0.177	1.52 x10 <sup>-3</sup>	6.35x10 <sup>-3</sup>
ANKRD53	chr2:71205510-71212626	0.426	1.60 x10 <sup>-3</sup>	6.62x10 <sup>-3</sup>
MTHFD1	chr14:64854749-64926722	0.112	1.69 x10 <sup>-3</sup>	6.68x10 <sup>-3</sup>
SDHC	chr1:161284047-161332984	0.195	1.65 x10 <sup>-3</sup>	6.68x10 <sup>-3</sup>
HSPD1	chr2:198351305-198381461	-0.111	1.67 x10 <sup>-3</sup>	6.68x10 <sup>-3</sup>
L2HGDH	chr14:50704281-50779266	-0.211	1.80 x10 <sup>-3</sup>	6.93x10 <sup>-3</sup>
CHCHD3	chr7:132469629-132766848	-0.147	1.78 x10 <sup>-3</sup>	6.93x10 <sup>-3</sup>
IARS2	chr1:220267444-220321380	0.123	1.94 x10 <sup>-3</sup>	7.39x10 <sup>-3</sup>
HOMER3	chr19:19040010-19052070	-0.139	2.12 x10 <sup>-3</sup>	7.97x10 <sup>-3</sup>
MLX	chr17:40719086-40725257	-0.116	2.16 x10 <sup>-3</sup>	8.03x10 <sup>-3</sup>
TMEM100	chr17:53796988-53809482	0.243	2.64 x10 <sup>-3</sup>	9.67x10 <sup>-3</sup>
EIF4EBP1	chr8:37887859-37917883	0.134	2.94 x10 <sup>-3</sup>	0.0106
ABHD5	chr3:43731605-43775863	-0.139	3.06 x10 <sup>-3</sup>	0.0109
C1orf43	chr1:154179182-154193104	-0.103	3.09 x10 <sup>-3</sup>	0.0109
ZNF16	chr8:146155744-146176274	0.248	3.12 x10 <sup>-3</sup>	0.0109
DLD	chr7:107531415-107572175	-0.109	3.19 x10 <sup>-3</sup>	0.0110
ORMDL3	chr17:38077294-38083854	0.134	3.54 x10 <sup>-3</sup>	0.0121
CAPN1	chr11:64948037-64979477	0.117	3.63 x10 <sup>-3</sup>	0.0123
VWA8	chr13:42140973-42535256	-0.124	4.12 x10 <sup>-3</sup>	0.0135
GLUL	chr1:182350839-182361341	0.0897	4.11 x10 <sup>-3</sup>	0.0135
GBAS	chr7:56019486-56067874	-0.131	4.06 x10 <sup>-3</sup>	0.0135
RP11-61A14.3	chr16:66923072-66924996	-0.181	4.25 x10 <sup>-3</sup>	0.0137
ACTN1	chr14:69340860-69446157	-0.102	4.52 x10 <sup>-3</sup>	0.0144
ARPC1A	chr7:98923521-98985787	0.0967	4.99 x10 <sup>-3</sup>	0.0158
MRPS9	chr2:105654441-105716418	-0.135	5.28 x10 <sup>-3</sup>	0.0163
RASL10B	chr17:34058668-34070540	0.743	5.25 x10 <sup>-3</sup>	0.0163
MRPL39	chr21:26957968-26979829	-0.119	5.84 x10 <sup>-3</sup>	0.0177
MAP3K5	chr6:136878185-137113656	0.157	5.82 x10 <sup>-3</sup>	0.0177
LRPPRC	chr2:44113647-44223144	-0.0919	6.09 x10 <sup>-3</sup>	0.0183
PDHX	chr11:34937376-35042138	-0.121	6.41 x10 <sup>-3</sup>	0.0190

HADH	chr4:108910870-108956331	0.105	6.71 x10 <sup>-3</sup>	0.0197
NUDT1	chr7:2281857-2290781	0.156	7.03 x10 <sup>-3</sup>	0.0205
MOCS1	chr6:39867354-39902290	0.172	7.26 x10 <sup>-3</sup>	0.0209
PHF13	chr1:6673745-6684093	0.149	7.36 x10 <sup>-3</sup>	0.0210
BFAR	chr16:14726672-14763093	0.105	7.67 x10 <sup>-3</sup>	0.0217
NKIRAS1	chr3:23933151-23988082	-0.134	7.76 x10 <sup>-3</sup>	0.0217
HADHB	chr2:26466038-26513336	0.0999	7.96 x10 <sup>-3</sup>	0.0221
EMC3	chr3:10004221-10052800	-0.110	8.25 x10 <sup>-3</sup>	0.0227
PDHB	chr3:58413357-58419584	-0.0998	8.90 x10 <sup>-3</sup>	0.0242
PRDX6	chr1:173446405-173457946	0.0861	9.07 x10 <sup>-3</sup>	0.0245
CHKA	chr11:67820326-67888911	0.139	0.0100	0.0268
CORO1C	chr12:109038885-109125372	0.0825	0.0108	0.0287
CENPV	chr17:16245848-16256970	0.126	0.0112	0.0294
LONP2	chr16:48278207-48397033	-0.0848	0.0120	0.0303
GRSF1	chr4:71681499-71705662	0.0857	0.0118	0.0303
SLC25A27	chr6:46620678-46645930	-0.194	0.0119	0.0303
NRIP1	chr21:16333556-16437321	0.126	0.0119	0.0303
GMCL1	chr2:70056774-70108528	0.150	0.0122	0.0307
SCO1	chr17:10583654-10601692	0.0878	0.0129	0.0323
SULF1	chr8:70378859-70573150	-0.149	0.0130	0.0323
CD248	chr11:66081958-66084515	-0.129	0.0141	0.0345
C9orf16	chr9:130922539-130926207	-0.181	0.0143	0.0348
ANKRD46	chr8:101521980-101572012	0.225	0.0152	0.0368
LSM6	chr4:147096837-147121152	-0.135	0.0161	0.0386
CCDC50	chr3:191046866-191116459	0.0816	0.0175	0.0415
GGCT	chr7:30536237-30591095	-0.107	0.0180	0.0424
MRPS27	chr5:71515236-71616473	-0.0739	0.0191	0.0446
C11orf1	chr11:111749659-111756699	-0.138	0.0208	0.0483
PRDX3	chr10:120927215-120938345	0.109	0.0218	0.0499
ADH1B	chr4:100226121-100242558	0.351	0.0218	0.0499

\* Abbreviation for chromosome.

† Average log<sub>2</sub> fold change in expression in human primary preadipocytes transfected with the *TBX15* siRNA when compared to the cells transfected with the negative control siRNA (see Methods).

# Chapter 5

## Discussion and Future Directions

Since the rise of genotyping arrays and next-generation sequencing and the decrease in price of these technologies in recent years<sup>1,2</sup>, the collection of genotype and gene expression information have increased drastically. However, despite many successes<sup>3-5</sup>, some limitations still remain. For example, even though the GenoType Expression (GTEx) project has collected samples from a total of 54 tissues across the human body and recently reached its final version with 17,382 RNA-sequencing (RNA-seq) samples from a total of 979 individuals<sup>5</sup>, the GTEx samples are mostly (n=948) from post-mortem samples with heterogeneous causes of death and variable times between death and sample collection, which affects the RNA quality and gene expression levels<sup>6</sup>. Human cohorts with genotype, RNA-seq, and deep phenotype data using obesity-related tissue samples from living individuals, such as the METabolic Syndrome In Men (METSIM)<sup>7</sup>, have been extremely useful to push forward the study of obesity and its comorbidities. However, METSIM consists of males and is still relatively limited by its sample size with gene expression data (n=335). The UKBiobank (UKB) cohort (n~500,000) is one of the largest single population cohorts with genotype and extensive phenotype information that is being updated on a tri-annual basis<sup>8</sup>. On the other hand, the UKB currently lacks gene expression information. Finally, the largest genome-wide association study (GWAS) for body mass index (BMI) has reached over 800,000 individuals<sup>9</sup>, giving us power to find more GWAS variants with small effect sizes and relatively small minor allele frequencies (MAF). However, as the effect sizes of novel GWAS variants decrease, the contribution of these loci to disease and phenotypes also decreases, limiting the overall usefulness of continually increasing the size of GWAS<sup>10</sup> even though there undoubtedly is still value in discovering new biology. Furthermore, GWAS only contain genotype and phenotype associations and are confounded by multiple variants in linkage disequilibrium (LD) at many of the GWAS loci, making it difficult to pinpoint the underlying

causal variants and genes in a locus. Despite the limitations, these human cohorts have allowed us to conduct powerful genomic analyses and gain insights into the underlying mechanisms of obesity through integrative and combined approaches to circumvent the limitations of any individual cohort. However, much is still unknown about the key genes and gene-environment interactions (GxEs) contributing to obesity and the context-specific genomic regulatory mechanisms conferring the susceptibility to obesity.

Moving forward, cohorts will most likely not only increase in size but also collect different types of omics data. Recently, tissues, including adipose, have been identified to consist of many more cell types than previously thought<sup>11</sup>. Therefore, single cell omics, including single cell RNA-seq<sup>12</sup>, single cell assay for transposase-accessible chromatin using sequencing (ATAC-seq)<sup>13</sup>, and single cell chromosome capture<sup>14</sup>, are all gaining in popularity. These give insights into the cell populations in a tissue and the differences in gene expression, open chromatin, and chromosomal interactions between different cell types. However, single cell data are still limited by their high cost, which in turn limits the number of samples that can be processed in any single study. Additionally, large scale cohorts in the past, including the vast majority of GWASs, predominantly consist of samples from European individuals<sup>15</sup>. However, the current trend shows an increasing number of studies in non-European cohorts<sup>16,17</sup>, which will allow us to understand the cross-population differences in disease prevalence and risk<sup>18</sup>.

In this thesis, we have employed integrative multi-omics approaches to study human subcutaneous adipose tissue for obesity-related genes and mechanisms. We have characterized two main types of regulatory mechanisms. In chapters 2 and 3, we characterized chromosomal interactions in which DNA form loops bringing distant regulatory elements in physical contact with gene promoters regulating gene expression levels. In Chapter 4 we identified transcription

factors (TFs) binding to gene promoters, resulting in widespread downstream effects on multiple genes across many chromosomes. Furthermore, after characterizing chromosomal interactions and *cis*-eQTLs related to obesity GWAS loci in chapter 2, we examined the effect of gene-environment interacting (GxE) variants at context-specific open chromatin regions on obesity in the UKB in chapter 3. Finally, we identified causal, TF-driven mechanisms for metabolically harmful abdominal obesity and its clinical outcome, type 2 diabetes (T2D) in chapter 4.

More specifically, in Chapter 2, we focused on understanding gene regulatory mechanisms via chromosomal interactions in human adipocytes and elucidating how *cis*-eQTL variants can affect the physical interactions of regulatory elements. We associated these genes by correlating their expression with BMI measurements in the METSIM cohort and identified the gene at 38 novel putative obesity loci whose mechanism of action can be explained through *cis*-eQTLs acting through chromosomal interactions, with deeper investigations into an additional four example loci that are also known GWAS loci for BMI and serum lipids and metabolites.

At the time of the study in Chapter 2, we were limited by the data and technology available in a number of ways. First, we did not have ATAC-seq data available in adipocytes to assess regions of open chromatin and therefore used open chromatin information aggregated across cell types from publicly available sources. ATAC-seq data in adipocytes could have enhanced the resolution of the chromosomal interactions and allowed us to pinpoint important adipocyte-specific regions of open chromatin in the chromosomal interactions. As we have this ATAC-seq data now from the study in chapter 3, further investigation using the ATAC-seq data in adipocytes would be a natural refinement to these previous findings. Second, promoter Capture Hi-C (pCHi-C) was a new and expensive technique at the time this study started. While we were fortunate to collaborate with the original creators for probe and experimental design<sup>19</sup>,

we only had a few replicates of the pChI-C, which may have impacted both the robustness of the chromosomal interactions we could find and our power to computationally detect significant chromosomal interactions. We now have many more replicates of the pChI-C data in adipocytes as a result of the study in Chapter 3 and there are also more publicly available pChI-C data sets in related cell types<sup>20</sup>. Reassessing the robustness of the chromosomal interactions as well as employing the replicates to identify additional chromosomal interactions in adipocytes can now be further investigated in future studies. Third, we used the publicly available Capture Hi-C Analysis of Genomic Organisation (CHiCAGO) R package<sup>21</sup> to detect significant chromosomal interactions; however, since CHiCAGO is the only widely available software for detecting pChI-C interactions to date, it is difficult to assess whether it is actually the best software for detecting interactions. Future investigations into the robustness of CHiCAGO using our current number of replicates of pChI-C as well as other data sets would definitely improve our ability to assess the performance of CHiCAGO and make any necessary modifications. The limitations of CHiCAGO and the replicates we had available also made it impossible to investigate longer range, *trans* chromosomal interactions. *Trans* chromosomal interactions remain an area of active research that has only few examples in the existing literature, all of which are in model organisms<sup>22,23</sup>.

In general, based on the conclusions from the findings in Chapter 2, we could further investigate the identified loci in a few ways. First, we showed at the Mitogen-Activated Protein Kinase Kinase 5 (*MAP2K5*) locus that the BMI GWAS variant in the chromosomal interaction, rs4776984, has differential binding of proteins at its alleles via electrophoretic mobility shift assay (EMSA) and computationally predicted them to be CCCTC-Binding Factor (CTCF) and E1A Binding Protein P300 (EP300), key proteins for chromosomal interactions. However, EMSAs simply show binding of protein using nuclear extract to oligonucleotides, without

indicating which protein is binding. Although we tried some techniques, including EMSA supershift, to assess which proteins were binding to rs4776984, we were unsuccessful, which is not surprising given the well-known challenges of these techniques<sup>24</sup>. Further investigation into the exact proteins or the complex of proteins binding to rs4776984 is warranted using for example mass spec. Finally, we hypothesized that the 38 identified loci regulated by chromosomal interactions contained some genes causal for obesity and some reactive to obesity. Using Transcriptome wide association studies (TWAS)<sup>25</sup> with an independent cohort, such as GTEx<sup>5</sup> and the larger GIANT and UKBiobank BMI GWAS<sup>9</sup>, we could next search for the causal BMI loci among these 38 novel regions.

In Chapter 3, we extended our investigation of chromosomal interactions to those which were responsive to lipid intake, while also focusing on changes in the open chromatin regions, specifically in adipocytes and during adipogenesis. We treated human primary adipocytes with saturated and unsaturated fatty acids and searched for changes in open chromatin regions using ATAC-seq and variants within those responsive open chromatin regions. I focused on transcription factor (TF) binding that changes due to gene-environment interacting (GxE) variants at these evolutionarily constrained open chromatin regions. Since GxE variants are difficult to detect due to their small effect sizes and multiple-testing penalties, using this targeted molecular approach and the large UKB cohort<sup>8</sup>, we identified 14 new GxE variants in lipid-responsive promoters and 24 GxE variants in enhancers that interact with the saturated fat intake on BMI. The underlying detailed molecular mechanisms at each variant site require further investigations.

While we had ATAC-seq data from preadipocytes and adipocytes and assessed the differences between those two stages of adipogenesis, the differentiation of adipocytes takes 14



days and the investigation of changes in open chromatin throughout adipogenesis using other time points, such as 24 and 48 hours after initiation of differentiation, is worth of further investigation. Furthermore, although it was an improvement from the study in chapter 2 to have two biological replicates of the pCHi-C, having even more replicates of pCHi-C, perhaps even up to 5, could further increase the robustness of the chromosomal interactions as well as enable us to identify additional chromosomal interactions with open chromatin regions and GxE variants in enhancers and promoters. As we saw from our data, chromosomal interactions are more dynamic than open chromatin, and therefore additional replicates can improve the identification of these chromosomal interactions. Identification of more chromosomal interactions could lead us to search for genes where both the enhancer and promoter had lipid responsive open chromatin instead of searching for open chromatin in enhancers and promoters separately. Genes with a coordinated opening or closing of chromatin in response to lipid intake could be key genes for response to lipid intake or adipogenesis and the subject of future investigations into the underlying mechanisms. Finally, when we searched for the effect of lipid responsive GxE variants on *cis* genes through a *cis*-eQTL analysis in the METSIM cohort<sup>7</sup>, we only identified five genes with a *cis*-eQTL in either their enhancer or promoter. Replication of these five signals in an independent cohort could further improve the robustness of this finding and a larger cohort, such as GTEx, could improve the power to detect signals. As GTEx<sup>5</sup> is publicly available via the National Institutes of Health database of Genotypes and Phenotypes (NIH dbGAP) and has released its final version with 581 adipose samples, the use of GTEx *cis*-eQTLs will be a natural extension to the findings in Chapter 3.

In Chapter 4, we investigated *trans*-eQTLs and master TFs in adipose tissue employing integrative genomics approaches using human adipose RNA-seq data (n~1,400) and waist-hip-

ratio adjusted for BMI (WHRadjBMI) GWAS<sup>9</sup>. We employed TWAS to provide statistical support for the causal role of the TF, T-box Transcription Factor 15 (*TBX15*), in regulating the accumulation of abdominal fat, measured via the proxy phenotype, WHRadjBMI. We functionally verified in human primary preadipocytes the role of *TBX15* in controlling an adipose network of 347 adipogenesis, mitochondrial, and metabolically important genes, including Peroxisome Proliferator Activated Receptor Gamma (*PPARG*), Krüppel Like Factor 15 (*KLF15*), Peroxisome Proliferator Activated Receptor Alpha (*PPARA*), Adiponectin (*ADIPOQ*), as well as 35 obesity GWAS genes. We built on the current knowledge that *TBX15* is an established GWAS locus<sup>26-28</sup> with previous knockdown studies in mice<sup>29</sup> showing the importance of *TBX15* on adipogenesis and lipolysis.

While we already show that *TBX15* affects this adipose network of genes through small interfering RNA (siRNA) knockdown in human primary preadipocytes, *TBX15* does not have any publicly available ChIP-seq data, let alone in adipocytes, to functionally show where it binds in the human genome. The current binding motif for *TBX15* in the JASPAR database<sup>30</sup> is also very general, with many locations across the genome predicted to harbor this motif, including 300 of the 347 network gene promoters (2kb upstream and 1kb downstream of the TSS) in our study, which is not a significant enrichment when compared to all other promoters across the genome. Producing ChIP-seq data for *TBX15* in relevant cell-types, perhaps also in conjunction with siRNA knockdown, could help dissect the direct targets of *TBX15* among the adipose network genes. As *TBX15* has already be shown to be a GWAS gene in a Mexican-American GWAS for visceral-to-subcutaneous adipose ratio<sup>26</sup>, investigations into *TBX15* and the adipose network genes in other population cohorts is definitely an important next step to extend our

findings beyond Europeans and further show the importance and influence of *TBX15* on abdominal obesity.

In summary we integrated omics data from multiple human cohorts along with genetic and phenotype information as well as functional follow-up data to understand the underlying genes and their regulatory mechanisms contributing to obesity. First, we used pChIP-C, *cis*-eQTLs, and GWAS to identify important obesity genes regulated by chromosomal interactions. We show that chromosomal interactions can effectively fine-map variants in a GWAS locus and identify the genes they directly regulate. We also provided 38 new candidate genes that are causal or reactive for obesity. Second, we identified lipid-intake responsive open chromatin regions via ATAC-seq in human primary adipocytes and obesity GxE variants in these open chromatin regions of enhancers and promoters. We showed that these open chromatin regions are evolutionarily constrained and the GxE variants significantly affect TF binding, including adipogenesis and lipolysis TF, Retinoid X Receptor Alpha (*RXRA*). Finally, we move beyond *cis* regulation of gene expression to TFs and *trans* effects, and identify a master regulator of abdominal obesity, *TBX15*. We also identify the downstream adipose network, which *TBX15* controls, and provide insights into the mechanisms contributing to the sex-dependent accumulation of fat around the abdomen. Our studies suggest that by integrating these multi-omics data and elucidating the mechanisms underlying obesity, we can understand the individual risks associated with obesity and its comorbidities, which will help advance personalized medicine.

## References

1. Stark, R., Grzelak, M. & Hadfield, J. RNA sequencing: the teenage years. *Nat. Rev. Genet.* **20**, 631–656 (2019).
2. Tam, V. *et al.* Benefits and limitations of genome-wide association studies. *Nat. Rev. Genet.* **20**, 467–484 (2019).
3. Schadt, E. E. *et al.* An integrative genomics approach to infer causal associations between gene expression and disease. *Nat. Genet.* **37**, 710–717 (2005).
4. Lonsdale, J. *et al.* The Genotype-Tissue Expression (GTEx) project. *Nat. Genet.* **45**, 580–585 (2013).
5. The Genotype Tissue Expression Consortium. The GTEx Consortium atlas of genetic regulatory effects across human tissues. *bioRxiv.* (2019)
6. Ferreira, P. G. *et al.* The effects of death and post-mortem cold ischemia on human tissue transcriptomes. *Nat. Commun.* **9**, (2018).
7. Laakso, M. *et al.* The Metabolic Syndrome in Men study: A resource for studies of metabolic & cardiovascular diseases. *J. Lipid Res.* **58**, 481–493 (2017).
8. Sudlow, C. *et al.* UK Biobank: An Open Access Resource for Identifying the Causes of a Wide Range of Complex Diseases of Middle and Old Age. *PLoS Med.* **12**, 1–10 (2015).
9. 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).
10. Park, J. H. *et al.* Estimation of effect size distribution from genome-wide association studies and implications for future discoveries. *Nat. Genet.* **42**, 570–575 (2010).
11. Ramirez, A. K. *et al.* Single-cell transcriptional networks in differentiating preadipocytes suggest drivers associated with tissue heterogeneity. *Nat. Commun.* **11**, (2020)
12. Tang, F. *et al.* mRNA-Seq whole-transcriptome analysis of a single cell. *Nat. Methods* **6**, 377–382 (2009).
13. Buenrostro, J. D. *et al.* Single-cell chromatin accessibility reveals principles of regulatory variation. *Nature* **523**, 486–490 (2015).
14. Nagano, T. *et al.* Single-cell Hi-C reveals cell-to-cell variability in chromosome structure. *Nature* **502**, 59–64 (2013).
15. Mills, M. C. & Rahal, C. A scientometric review of genome-wide association studies. *Commun. Biol.* **2**, (2019).
16. Sirugo, G., Williams, S. M. & Tishkoff, S. A. The Missing Diversity in Human Genetic Studies. *Cell* **177**, 26–31 (2019).
17. Rosenberg, N. A. *et al.* Genome-wide association studies in diverse populations. *Nat. Rev. Genet.* **11**, 356–366 (2010).
18. Martin, A. R. *et al.* Clinical use of current polygenic risk scores may exacerbate health disparities. *Nat. Genet.* **51**, 584–591 (2019).
19. Mifsud, B. *et al.* Mapping long-range promoter contacts in human cells with high-resolution capture Hi-C. *Nat. Genet.* **47**, 598–606 (2015).
20. Jung, I. *et al.* A compendium of promoter-centered long-range chromatin interactions in the human genome. *Nat. Genet.* **51**, 1442–1449 (2019).
21. Cairns, J. *et al.* CHiCAGO : robust detection of DNA looping interactions in Capture Hi-C data. *Genome Biol.* **17**, 127 (2016).
22. Erceg, J. *et al.* The genome-wide multi-layered architecture of chromosome pairing in early *Drosophila* embryos. *Nat. Commun.* **10**, (2019).

23. Chowdhary, S., Kainth, A. S. & Gross, D. S. Chromosome conformation capture that detects novel cis- and trans-interactions in budding yeast. *Methods* **170**, 4–16 (2020).
24. Holden, N. S. & Tacon, C. E. Principles and problems of the electrophoretic mobility shift assay. *J. Pharmacol. Toxicol. Methods* **63**, 7–14 (2011).
25. Gusev, A. *et al.* Integrative approaches for large-scale transcriptome-wide association studies. *Nat. Genet.* **48**, 245–252 (2016).
26. Gao, C. *et al.* Americans : The Insulin Resistance Atherosclerosis Family Study. **26**, 202–212 (2018).
27. Heid, I. M. *et al.* Meta-analysis identifies 13 new loci associated with waist-hip ratio and reveals sexual dimorphism in the genetic basis of fat distribution. *Nat. Genet.* **42**, 949–960 (2010).
28. Shungin, D. *et al.* New genetic loci link adipose and insulin biology to body fat distribution. *Nature* **518**, 187–196 (2015).
29. Sun, W. *et al.* Tbx15 is required for adipocyte browning induced by adrenergic signaling pathway. *Mol. Metab.* **28**, 48–57 (2019).
30. Khan, A. *et al.* JASPAR 2018: Update of the open-access database of transcription factor binding profiles and its web framework. *Nucleic Acids Res.* **46**, D260–D266 (2018).
31. Turcot, V. *et al.* Protein-altering variants associated with body mass index implicate pathways that control energy intake and expenditure in obesity. *Nat. Genet.* **50**, 26–35 (2018).