UCLA
UCLA Previously Published Works

Title

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

Permalink

https://escholarship.org/uc/item/30j7c9rq

Journal

Nature metabolism, 1(6)

ISSN

2522-5812

Authors

Garske, Kristina M
Pan, David Z
Miao, Zong
et al.

Publication Date

2019-06-01

DOI

10.1038/s42255-019-0071-6

Peer reviewed

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

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

[1]Department of Human Genetics, David Geffen School of Medicine at UCLA, Los Angeles, CA, USA, 90095

[2]Bioinformatics Interdepartmental Program, UCLA, Los Angeles, CA, USA, 90095

[3]Department of Computer Science, UCLA, Los Angeles, CA, USA, 90095

[4]Vache and Tamar Manoukian Division of Digestive Diseases, UCLA, Los Angeles, CA, USA, 90095

[5]Institute for Human Genetics, Department of Epidemiology and Biostatistics, Department of Bioengineering and Therapeutic Sciences, UCSF, San Francisco, CA, USA, 94143

[6]Department of Genetics, University of North Carolina, Chapel Hill, NC, USA, 27599

[7]Department of Biomathematics, David Geffen School of Medicine at UCLA, Los Angeles, CA, USA, 90095

[8]Institute of Clinical Medicine, Internal Medicine, University of Eastern Finland and Kuopio University Hospital, Kuopio, Finland, FI-70210

Abstract

Identifying gene-environment interactions (GxEs) contributing to human cardiometabolic disorders is challenging. Here we apply a reverse GxE candidate search by deriving candidate variants from promoter-enhancer interactions that respond to dietary fatty acid challenge through altered chromatin accessibility in human primary adipocytes. We then test all variants residing in the lipid-responsive open chromatin sites within adipocyte promoter-enhancer contacts for interaction effects between the genotype and dietary saturated fat intake on body mass index (BMI) in the UK Biobank. We discover 14 novel GxE variants in 12 lipid-responsive promoters,

including well-known lipid genes (*LIPE, CARM1*, and *PLIN2*) and novel genes, such as *LDB3*, for which we provide further functional and integrative genomics evidence. We further identify 24 GxE variants in enhancers, totaling 38 new GxE variants for BMI in the UK Biobank, demonstrating that molecular genomics data produced in physiologically relevant contexts can discover new functional GxE mechanisms in humans.

Introduction

Cardiometabolic disorders develop as a result of genetic predisposition, environmental factors, and their interactions[1,2]. Genome-wide association studies (GWAS) have detected additive genetic effects for these traits, but we still have not identified the biological mechanisms explaining how genetic variation is involved in the increasing prevalence of the obesogenic cardiometabolic disorders. Some examples of gene-environment interactions (GxEs) are emerging, including for instance, the highly replicated BMI risk variant rs9939609 in the intron of the *FTO* gene that exhibits a significant interaction with physical activity on BMI[3]. However, overall there are scarce replicated GxE signals for cardiometabolic disorders in humans[1]. It has remained challenging to identify these signals due to small cohort sizes and poorly standardized human environmental phenotypes. Even with large cohorts such as the UK Biobank[4], the statistical power to detect GxEs using a genome-wide agnostic search is limited due to the small effect sizes of GxEs and heavy multiple testing penalties. Furthermore, once GxE signals have been detected, the mechanisms underlying the associations remain unclear, warranting further fine-mapping studies.

To systematically identify genes involved in GxEs, we set out to quantify molecular genomic responses to saturated and monounsaturated fatty acid challenge in human primary 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 onto the genomic molecular mechanisms relevant for lipid responses in human adipocytes. We hypothesized that these genomic responses then provide targeted regions harboring candidate genetic variants for GxE analysis in the large UK Biobank cohort[4]. This should restrict the multiple testing burden hampering the typical agnostic genome-wide GxE analysis, and expand our knowledge of the true environmental exposures responsible for the GxE 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 that respond to obesogenic cellular context and contribute to cardiometabolic traits in humans.

Results

Adipocyte-accessible chromatin identifies regulatory regions

To obtain human primary adipocytes for the study of lipid effects on chromatin dynamics, we first differentiated human primary white preadipocytes to adipocytes *in vitro* (Figure 1a). We performed ATAC-seq[5] on three biological replicates of the preadipocytes and adipocytes

to identify open chromatin regions that are differentially accessible between the two cell types (Figure 1b; Supplementary Table 1; Supplementary Table 2). The 50,336 ATAC-seq peaks that were more accessible in human primary adipocytes (i.e. "adipocyte-accessible") included the promoters of adipocyte-specific genes *ADIPOQ* and *PPARGC1A* (Figure 1c,d), providing evidence that we successfully differentiated the adipocytes *in vitro*. To explore whether the adipocyte-accessible peaks harbor transcription factor (TF) motifs relevant for adipocyte biology, we performed TF motif enrichment analyses using HOMER[6]. We found that the top enriched motif corresponds to the CCAAT-enhancer binding protein alpha (CEBPA) transcription factor motif (Figure 1b), an important regulator of the later stages of adipogenesis[7]. We then classified the adipocyte-accessible peaks into functional genomic annotations[8], and observed that the adipocyte-accessible peaks fell more often into adipocyte enhancers and less frequently into quiescent regions when compared with the full peak set or preadipocyte-accessible peaks (Supplementary Figure 1). Taken together, these results provide evidence that the *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 harboring 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 (Figure 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 Figure 2). We identified 1,653 ATAC-seq peaks that were differentially accessible in the lipid-challenged human primary adipocytes, referred to as lipid-responsive peaks (Figure 2b; Supplementary Figure 3; Supplementary Table 3; Supplementary Table 4). We cross-referenced these lipid-responsive peaks with the adipocyte-accessible peaks (Figure 1), and found that the 570 adipocyte-accessible, lipid-responsive peaks fell mostly into enhancer and promoter annotations in adipocytes[8], in line with their likely importance in environmental responses and regulation of gene expression in adipocytes (Figure 2b; Supplementary Figure 4).

Noteworthy, 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 (Figure 2b). These context-dependent open chromatin regions fell into a higher percentage of quiescent annotations imputed from data created in unchallenged adipocytes[8] (Supplementary Figure 4). This indicates that genomic regions that are not normally open and accessible for TF binding in untreated, steady-state adipocytes or preadipocytes became 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 SFA treatment effects in context-dependent open chromatin regions were shifted towards increased accessibility (Figure 2c,d). This is in contrast to the MUFA responses, which were generally evenly

distributed between the peaks with decreased and increased accessibility (Figure 2c,d). This suggests that the quality of fatty acids can result in distinct signaling effects on genome-level responses to lipid intake in adipocytes, and particularly SFA intake seems to activate regions of the genome that are normally inactive in untreated adipocytes.

Upon closer examination, we found 14 context-dependent peaks, exhibiting increased accessibility in the SFA-treated adipocytes, falling into a ~50 kb region on chromosome 10 (Figure 2e). The locus contains a total of 15 lipid-responsive peaks, which span the entirety of the glycerol-3-phosphate acyltransferase, mitochondrial (*GPAM*) gene (Figure 2e). *GPAM* prefers saturated fatty acid substrates, and has been associated with serum lipid traits[9–11] [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[12], in previous GWAS studies[13]. These earlier GWAS associations at this locus, combined with the strong genomic response to SFA treatment in adipocytes observed here, suggest that dysregulation of this important lipogenic pathway in adipocytes could contribute to obesogenic cardiometabolic disorders, such as dyslipidemias 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)[14] in the lipid-challenged human adipocytes in two biological replicates per condition (Supplementary Table 5). We identified 264 lipid-responsive ATAC-seq peaks that fall within adipocyte chromosomal interactions. To test whether these interacting, lipid-responsive regions of the genome harbor motifs for TFs that are important for lipid metabolism, we performed TF motif enrichment analysis[6] comparing the lipid-responsive to non-lipid-responsive peaks within the chromosomal interactions. We found that peroxisome proliferator-activated receptor gamma (PPARG), an important TF in adipogenesis and lipid uptake, and its co-factor retinoid X receptor (RXR), are within the top 10 highest enriched motifs (Figure 3a; 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 the lipid uptake.

To identify the target genes of the adipocyte responses to lipid challenge, we first focused on the interacting promoters from the adipocyte pCHi-C (Figure 3b), as promoters are more highly enriched for SNPs that contribute to the heritability of local gene expression[15,16]. The 86 interacting pCHi-C baits represent 154 gene promoters given that the resolution of pCHi-C interaction data depends on the restriction sites in the genome (Figure 3b; Supplementary Table 7). We performed a KEGG pathway enrichment analysis[17] on the set of 154 interacting lipid-responsive target genes, which resulted in 2 significantly enriched pathways of amino acid metabolism (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 +/− 500 kb) using PHASTcons[18], and found that the lipid-responsive

protein-coding genes have higher conservation scores across placental mammals than all other protein-coding genes in the human genome (p=0.020) (Figure 3c).

We further investigated whether these lipid-responsive genes exhibit constraints on genetic mutations, using the probability of the gene being loss-of-function intolerant (pLI), defined in Lek *et al.*[19] as genes that are highly unlikely to exhibit protein-truncating mutations in humans. We found that of the genes for which pLI scores are available (n=104)[19], 27 genes (26.0%) are 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) (Figure 3d). Taken 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 interact with lipid-responsive enhancers exhibit similar characteristics as the genes that have lipid-responsive promoters. We first found that 169 lipid-responsive enhancers interact with 223 promoter-containing *Hin*dIII baits in the adipocyte pCHi-C (Supplementary Figure 5). Because multiple gene promoters can be captured within a single *Hin*dIII fragment, these 223 baits represent 323 gene promoters (Supplementary Figure 5; Supplementary Table 9). When we tested whether these 323 genes are 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 interacts with ~2 promoters (Supplementary Figure 5), thus leading to ambiguities regarding which gene may be the true target of the lipid signaling response.

We further determined whether the genes that interact with lipid-responsive enhancers exhibit mutational constraints by checking whether the *cis* regions (gene body +/− 500 kb) of the protein-coding genes in this set (n=217) have higher average conservation scores than all other protein-coding genes in the genome. Contrary to what we observed for the genes with lipid-responsive promoters (Figure 3c), we did not observe a significant difference in the conservation scores for genes that interact with lipid-responsive enhancers (Supplementary Figure 5). Interestingly, of the enhancer-interacting genes that have a pLI score (n=207)[19], 50 (24.2%) are LoF-intolerant, which is significantly higher than would be expected by chance alone (p=0.014) (Supplementary Figure 5). Taken together, these results are consistent with more moderate functional significance of the genes that interact with lipid-responsive enhancers when compared to the genes with lipid-responsive ATAC-seq peaks in their promoters.

Lipid responses contribute to serum lipid trait heritability

We hypothesized that the genes we were able to identify through the lipid-responsive promoter-enhancer interactions might highlight important genomic regions that contribute to the heritability of cardiometabolic traits. We first found that 5 lipid-responsive gene promoters and 3 lipid-responsive enhancers within adipocyte chromosomal interactions contain SNPs with genome-wide-significant ($p < 5\times10^{-8}$) associations with serum lipid traits, identified in a meta-GWAS of ~180,000 individuals[11] (Table 1; Supplementary Table

10). One of the lipid-responsive GWAS loci is the well-known nutritional response locus containing the fatty acid desaturase (*FADS1/2/3*) gene cluster on chromosome 11 (Figure 4a), which harbors SNPs that have been associated with multiple cardiometabolic traits[20] and intermediate phenotypes[21,22]. In line with the observed pleiotropy among the serum lipid-associated SNPs, the lipid-responsive peak in the *FADS2* promoter contains GWAS SNPs for all tested serum lipid traits[11] (LDL, HDL, TC, TG), with the strongest signal coming from rs99780 for LDL ($p=2.39\times10^{-21}$; Table 1). Notably, the observed open chromatin peak in *FADS2* is more accessible in the palmitic acid treatment when compared with the oleic acid treatment (Figure 4b; FDR=0.0021), corresponding to the fact that one of the FADS2 substrates is palmitic acid[23].

This response at a serum lipid GWAS locus is reminiscent of the strong SFA response at the *GPAM* gene (Figure 2e); in fact, all five serum lipid GWAS SNPs 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 interact with gene promoters in the 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.

Since sub-genome-wide significant signals likely 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* regions (gene body +/− 500 kb) of all 154 lipid-responsive, interacting promoters, contribute significantly to the heritability of serum lipid levels. For these analyses, we used the LD Score partitioned heritability method[24], and GWAS summary statistics from the high-powered meta-GWAS for serum lipid traits[11]. We found that 2.9% of all variants reside within the *cis* regions of the 154 genes, and these variants contribute 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 reside within the *cis* regions of genes with promoters interacting with lipid-responsive enhancers. These SNPs contribute significantly to the heritability of HDL, but not to the other lipid traits (Supplementary Table 14). This is consistent with a more diffuse overall functional characterization of the genes that interact with lipid-responsive enhancers when compared to the genes with lipid-responsive open chromatin in their promoters. Overall, these results indicate that the adipocyte lipid-responsive, interacting loci are important in modulating serum lipid levels in humans, and provide evidence that variants in these regions may have a role in GxEs in humans.

Lipid responses identify novel GxEs for BMI in UK Biobank

The large, deeply phenotyped UK Biobank[4] cohort can provide a valuable resource for GxE studies, particularly since the participants' environmental phenotypes have been characterized in a systematic manner. Saturated fat intake has known adverse effects on cardiometabolic disorders[25–27], and we present evidence here in human adipocytes for an enhanced effect of human adipocyte genomic responses to SFA intake on cardiometabolic

traits relative to MUFA (Figure 2c–e; Supplementary Table 11; Supplementary Table 12). To maximize the number of individuals for which phenotypes are available, as well as aim for the most relevant environment and cardiometabolic outcome, we used dietary intake of saturated fat (24-hour recall) as the environmental variable, and body mass index (BMI) as the outcome for our GxE analysis.

We first tested whether there are any genome-wide significant signals for GxEs using 167,908 individuals in the UK Biobank. We corrected the BMI measurements for array type, sex (inferred), age (when attended assessment center), age^2, the assessment center ID, and genetic principle components of 1-20, as done previously[28]. We then inverse normal transformed the residuals to account for mean-variance relationships in the phenotype, which has been shown previously to impact the GxE signals[29]. As can be seen in the Q-Q plot from the genome-wide scan for GxEs, there is no evidence of genomic inflation in the GxE linear model (Supplementary Figure 6). Furthermore, the fact that we were not able to detect any genome-wide significant signals in this genome-wide GxE analysis (see Equation 1 in Methods) (Supplementary Figure 6) supports the feasibility of our reverse GxE candidate search approach, which identifies prior functional candidates for GxE analyses from molecular genomics data produced in 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 provide strong candidates for GxEs in humans. The 154 promoters in chromosomal interactions contained 91 lipid-responsive open chromatin sites (Figure 3b), and we determined that 75 of these 91 candidate regions contained variants with a minor allele frequency (MAF) > 0.05 in the set of 167,908 individuals for which we have both dietary saturated fat intake and BMI phenotypes available (Supplementary Table 15).

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

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

Identifying altered chromatin states at the GxE SNP sites

The differential chromatin accessibility in response to lipid challenge in adipocytes likely stems from altered chromatin states, such as TF binding or histone modifications. This is supported by our result that the lipid-responsive regions within chromosomal interactions are enriched for important lipid metabolism TF motifs (Figure 3a; Supplementary Table 6). To determine the predicted allelic effect of the GxE SNPs on these chromatin features, we used the DeepSEA tool[30], which applies a deep learning algorithm to publicly available molecular genomics data to predict chromatin features based on genomic sequence *in silico*. Notably, 11 of the 20 (55%) GxE SNPs in lipid-responsive promoters had a functional significance score of <0.05, and the predicted impact of the GxE SNPs included differential binding of RXRA (Supplementary Table 19). Conversely, only 5 of the 26 (19%) GxE SNPs in lipid-responsive enhancers had functional significance scores of <0.05 (Supplementary Table 20). It is worth noting that the publicly available data used to train the DeepSEA neural net do not include molecular genomics data for adipocytes or adipose tissue. Therefore, the GxE SNPs may fall into cell-type-specific regulatory elements, and thus disrupt chromatin features that cannot be predicted through this tool.

To obtain further evidence for the function of the GxE SNPs in adipose tissue, we checked whether the GxE SNPs affect local gene expression in human adipose tissue as *cis* expression quantitative trait loci (*cis*-eQTLs), and whether the *cis*-eQTL target gene is the same gene that harbors the lipid-responsive promoter and GxE signal. We found that 3 of the 12 genes with lipid-responsive promoters harboring GxE signals are also regulated in *cis* by their GxE SNP at the genome-wide significant level in the subcutaneous adipose RNA-seq data (n=335) of the Finnish METSIM cohort[15,31] (Table 2; Supplementary Table 16). These genes are glioma tumor suppressor candidate region gene 2 protein (*GLTSCR2*), *PLIN2*, and LIM domain binding 3 (*LDB3*). Additionally, 2 of the 27 genes interacting with lipid-responsive enhancers were regulated in *cis* by their interacting GxE SNP at the genome-wide significant level in human subcutaneous adipose tissue (Supplementary Table 18): thrombospondin 2 (*THBS2*) and Cbp/P300 interacting transactivator with glu/asp rich carboxy-terminal domain 4 (*CITED4*).

We further checked whether the imputed *cis* expression values for these 5 *cis*-eQTL target genes (eGenes) are correlated with BMI or other obesogenic cardiometabolic phenotypes, as determined through transcriptome-wide association analysis (TWAS)[32,33]. The *LDB3* adipose expression model was strongly associated with BMI, arm fat percentage (genome-wide significant TWAS scores > 5.0), and other related body fat distribution phenotypes (TWAS scores > 4.0)[29]. Furthermore, the tibial artery expression model for *LDB3* was also significantly associated with high blood pressure and cardiovascular disease (genome-wide significant TWAS scores < −5.0)[29]. None of the other eGene expression models were associated with cardiometabolic phenotypes at the genome-wide significant level.

Since the adipose expression models of *LDB3* were significantly associated with BMI in TWAS[29], we followed up on the most significant GxE SNP in the *LDB3* lipid-responsive peak (Figure 5a), rs10788522, which is also an adipose *cis*-eQTL SNP for *LDB3* in the METSIM cohort (Table 2). As is evidenced by the ATAC-seq reads intersecting with the SNP rs10788522 (Figure 5b), we found, using electrophoretic mobility shift assays

(EMSAs), that adipocyte nuclear protein binds this GxE SNP (Figure 5c; Supplementary Table 21). Whereas *LDB3* is expressed in the subcutaneous adipose tissue from the METSIM cohort, as well as the Gene Tissue Expression (GTEx) consortium[34], we could not reliably detect *LDB3* expression using quantitative PCR (qPCR) in the lipid-challenged adipocytes (data not shown). We therefore checked whether any publicly available datasets analyzed the transcriptome of sorted cells from human adipose tissue. In previously published microarray datasets (GSE80654, GSE100795) from human adipose biopsies in which the adipocytes were collected and the remaining cell types separated by fluorescence-activated cell sorting (FACS), *LDB3* is expressed in human adipocytes at a comparable level to the *LDB3* expression in the other adipose cell types[35,36]. Taken together with our result that the *LDB3* promoter responds to lipid uptake in adipocytes, these data support the conclusion that the GxE SNP rs10788522 regulates *LDB3* adipocyte expression 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 GxE SNP rs10788522, narrowing it to the promoter of *LDB3* that exhibits differential open chromatin in response to lipid challenge in human adipocytes (Figure 5; Figure 6).

Discussion

It is well-established that environment plays a significant role in the development of cardiometabolic disorders. However, GxEs have been challenging to detect due to both the lack of extensive study cohorts with sufficient statistical power to detect the small GxE effects, and complexity of environmental exposures that are difficult to measure in a standardized way in humans[1]. Systematic identification of the effects of defined environmental contributions to cardiometabolic disorders is thus necessary to effectively move toward the promise of precision medicine. Through our integration of context-specific molecular genomics data with human epidemiologic and clinical outcome data in the UK Biobank, we provide much-needed information of how the chromatin landscape of human adipocytes responds to external environmental signals, and identify the molecular basis of novel GxEs in humans (Figure 6).

We reversed the typical approach of selecting candidate GxEs among GWAS SNPs, by first scanning the genome for molecular responses to controlled environmental stimuli, using a cellular model of saturated or monounsaturated fat intake effects on chromatin dynamics in human primary adipocytes. Through our integration of chromatin accessibility and chromosomal interactions in the lipid-challenged adipocytes, we identified lipid-responsive open chromatin within promoter-enhancer contacts, effectively identifying candidate GxE genes with strong prior 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[4]. This led to the identification of 14 significant, non-redundant GxE SNPs in 12 gene promoter regions and 24 non-redundant GxE SNPs in 20 enhancers, representing novel gene-diet interactions affecting BMI (Figure 6).

We observed that the *LDB3* gene is regulated in *cis* by its promoter GxE SNP in human adipose tissue, and the previous TWAS analyses have shown that the imputed local adipose expression of *LDB3* is significantly associated with BMI and related cardiometabolic phenotypes[32,33]. Notably, individuals with nonsynonymous mutations in exon 6 of *LDB3* have been shown to exhibit autosomal dominant myofibrillar myopathy characterized by fatty degeneration (i.e. steatosis) of the muscle that progresses with age[37,38]. This is suggestive of the role of lipid metabolism in the pathophysiology of these variants. Here we show that the accessibility of the promoter region is increased in response to lipid challenge in human adipocytes, and adipocyte nuclear protein binds to the GxE SNP site. Although we did not detect *LDB3* expression in our cultured adipocytes, we found that the gene is expressed in mature adipocytes isolated from human adipose biopsies, i.e. *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 results provide a functional mechanism for the GxE effect in human adipocytes, adding to our 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 the adipose tissue. These gene regions have a higher conservation score, and the 154 genes are more likely to be loss-of-function (LoF) - intolerant than expected by chance, suggesting that there are evolutionary constraints to maintain their proper function. The LoF-intolerant genes are widely and highly expressed[19], 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[25,27], and a GRS for obesity-related traits was previously shown to interact with saturated fat (SF) intake to affect BMI[26]. However, as the authors of that study note, the underlying mechanisms for the GxEs remain elusive, particularly when the effect is estimated across tens of SNPs[26]. 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 likely specific to saturated fat intake and, when dysregulated at the genetic level, could underlie GWAS and GxE 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 monounsaturated fatty acid, palmitoleic acid (C16:1), by stearoyl-CoA desaturase (SCD). Thus, we do not know whether the effects we see at the DNA level are directly due to the SFA signaling, or whether these effects result from downstream signaling mechanisms of the SCD pathway, affecting MUFA concentrations inside the cell.

Our finding that the SNPs within the *cis* regions of the 154 lipid-responsive gene promoters (gene body +/− 500 kb) contribute 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 5 adipocyte lipid-responsive, interacting gene promoters and 3 lipid-responsive enhancers that harbor genome-wide significant signals for serum lipid traits, including the well-known nutrient-sensing *FADS1/2/3* gene cluster. Specific lipid-associated SNPs within this locus have undergone positive selection in Inuits, thought to be in response to the polyunsaturated fatty acids (PUFAs) in high marine mammalian fat diets[39]. Furthermore, a gene-diet interaction for PUFA intake has been identified in Europeans at the *FADS1/2/3* locus[40], and while many studies of the effects of *FADS1/2* 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, have been less conclusive[41,42]. 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 GxEs through molecular genomic responses to relevant environmental stimuli.

While 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 the corresponding genomic annotations were created in cells at steady or unstimulated baseline states[43–46]. Context-specific molecular genomics studies have mainly been performed in immune cell types[47–49]; however, 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[8], 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 our understanding of environmental epigenomic responses underlying complex traits. We obtained a global assessment of genomic responses of human primary adipocytes to dietary fatty acid uptake, through incorporating open chromatin and chromosomal interaction data that we followed for GxEs in UK Biobank. Overall, our study helped discover candidate functional mechanisms at 38 novel 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.

Materials and Methods

Cell lines and culture reagents

We obtained and cultured the primary human white preadipocyte cells as recommended by PromoCell (PromoCell C-12731, lot 395Z024) for preadipocyte growth and differentiation

into adipocytes. Cell media (PromoCell) was supplemented with 1% penicillin-streptomycin. We maintained the cells at 37ºC in a humidified atmosphere at 5% CO2. For the lipid challenge experiments in the adipocytes, we serum-starved for 16 hours using 0.5% fetal calf serum (FCS) in supplemented adipocyte basal medium (PromoCell), prior to treatment with 200 μM palmitic acid:BSA, 200 μM oleic acid:BSA complex (Sigma Alrich O3008), or 0.23% fatty acid free bovine serum albumin (BSA, Sigma Aldrich A8806) as a vehicle control, in media containing 0.5% FCS for 24 hours prior to performing experiments.

Palmitic acid conjugation to bovine serum albumin

We dissolved 25.6 mg palmitic acid (PA, Sigma Aldrich P5585) into 1 mL 0.15 M NaCl at 70ºC in a shaking heat block to make 100 mM PA. We added the PA solution dropwise into 10% w/v BSA in 0.15 M NaCl at 37ºC to make PA:BSA conjugate at 8 mM stock PA.

Oil red o staining and quantification

We prepared oil red o (ORO) stock by making 0.3% ORO in >99% isopropanol and filtering through a 0.45 μm filter. This was diluted 3:2 ORO:H2O, incubated at room temperature (RT) for 10 minutes, and then filtered through a 0.22 μm filter. We fixed cells for 30-60 minutes in 10% formalin, rinsed with distilled water, and incubated 2-5 minutes at RT with 60% isopropanol. We stained with ORO for 15 minutes, rinsed well with water, and collected images for quantification. Cells were photographed with a Keyence bright-field, light microscope under 10x magnification for publication images and 20x magnification for lipid droplet (LD) quantification. LD numbers were counted for ~20 cells per condition (untreated, BSA, PA, and OA). The total area of ORO staining was quantified using ImageJ[50].

Assay for transposase accessible chromatin (ATAC) –sequencing

We performed the ATAC-seq protocol in the untreated primary human preadipocytes and adipocytes in 300,000 nuclei in three biological replicates per cell type, similarly as developed in Buenrostro et al.[5] Specifically, we lysed cells in ice-cold lysis buffer (10 mM Tris-HCl pH 7.4, 10 mM NaCl, 3 mM MgCl2) plus 0.03-0.1% Tween-20 for 10 minutes on ice. We centrifuged at 4°C, 500 $x\ g$ for 10 minutes, and then resuspended the nuclear pellet in 50 ul of transposition master mix (25 ul 2x TDE1 buffer, 2.5 ul transposase, 22.5 ul nuclease-free water; Illumina FC-121-1030). We incubated at 37ºC for 30 minutes, and then purified the DNA with the Qiagen MinElute kit (Qiagen 28204). Libraries were amplified using 6 cycles and sequenced on the Illumina HiSeq 4000 to produce an average of 23,376,290 (+/− 3,337,206) reads.

For the ATAC-seq in primary human adipocytes that underwent lipid challenge, we performed omni-ATAC as developed in Corces et al.[51] in 300,000 nuclei in three biological replicates per condition. Specifically, we treated the cells with DNase I (Worthington, 200 U/ml) at 37°C for 30 minutes, trypsinized the cells and resuspended in 50 ul of ice-cold lysis buffer (10 mM Tris-HCl pH 7.4, 10 mM NaCl, 3 mM MgCl2, 0.1% Igepal CA-630, 0.1% Tween-20, and 0.01% digitonin) and incubated on ice for 3 minutes. We washed with 1 ml ice-cold lysis quench (10 mM Tris-HCl pH 7.4, 10 mM NaCl, 3 mM MgCl2, 0.1%

Tween-20) and centrifuged at 4°C, 500 $x\ g$ for 10 minutes. The nuclear pellet was resuspended in 50 ul transposition master mix (25 ul 2x TDE1 buffer, and 2.5 ul transposase, 16.5 ul PBS, 0.5 ul 1% digitonin, 0.5 ul 10% Tween-20, and 5 ul nuclease-free water). We incubated at 37ºC for 30 minutes with mixing at 1000 RPM, and then purified the DNA with the Qiagen MinElute kit. Libraries were amplified using 6-7 cycles and sequenced on the Illumina HiSeq 4000 to produce an average of 40,315,572 (+/−14,577,770) reads.

ATAC-seq data processing and peak calling

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

Identification of differentially accessible ATAC-seq peaks

Read alignments from all untreated human preadipocyte and adipocyte (3 biological replicates per cell type) libraries were merged prior to peak calling. Peaks were called using MACS2[54] v2.1.1 (using the BEDPE function) and peaks with an FDR < 0.05 were retained. We filtered out peaks in blacklisted regions and peaks that did not replicate in two out of the three biological replicates in at least one condition. For differential accessibility analyses, we retained peaks with counts per million (cpm) >= 1 in at least 3 libraries. We then input aligned read counts per peak into cqn[55] v1.20.0 and normalized the counts for GC content, peak length, and library size before inputting the counts into edgeR[56] v3.16.5 to detect differentially accessible peaks between the preadipocytes and adipocytes using the generalized linear model (GLM) functionality and quasi-likelihood (QL) F-test and an FDR threshold of 0.05. Libraries for the adipocyte lipid challenge ATAC-seq were processed identically up until the differential accessibility analysis. To detect open chromatin regions that exhibit differential accessibility in the lipid-challenged adipocytes, we performed a one-way analysis of variance (ANOVA) using the GLM functionality and QL F-test functionality in edgeR[56] and an FDR threshold of 0.05. 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.

Transcription factor motif enrichment in ATAC-peaks

We used Hypergeometric Optimization of Motif EnRichment (HOMER, v4.9)[6] to investigate the enrichment of motif sequences in the open chromatin regions. For the enrichment in differentially accessible peaks between untreated preadipocytes and adipocytes, we used the consensus peak set (i.e. all peaks that were called both in 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 also 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 the enrichment in lipid-responsive peaks in adipocyte promoter-enhancer contacts, we used non-differentially accessible peaks within the promoter-enhancer contacts as the background.

Due 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.[15] in two biological replicates per condition (BSA, OA, and PA). These methods were adapted by closely following in-nucleus Hi-C methods in Rao et al.[57] and Nagano et al.[58] Specifically, we fixed 7-10 M 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 nM Tris-HCl pH 8.0, 10 mM NaCl, 0.2% Igepal CA-630, and 1X protease inhibitors: cOmplete™, EDTA-free Protease Inhibitor cocktail) on ice for 30 minutes with occasional agitation. We split into 5M nuclei aliquots, and centrifuged at 2500 $x\,g$ for 5 min at 4°C. The nuclear pellets were resuspended in 50 ul of 0.5% SDS in 1X NEBuffer 2 (New England Biolabs) and incubated at 62°C for 10 minutes. We then added 145 ul water and 25 ul 10% Triton X-100 and incubated at 37°C for 15 minutes and digested chromatin by adding 25 ul 10X NEBuffer 2 and 400 U of *Hin*dIII (New England Biolabs), incubating at 37°C overnight with shaking (950 rpm).

The next day, we marked the DNA ends with biotin [1.5 ul 10 mM dATP, 1.5 ul 10 mM dGTP, 1.5 ul 10 mM dTTP, 37.5 ul 0.4 mM biotin-14-dCTP (Invitrogen), and 8 ul 5U/µl Klenow (New England Biolabs)], incubating for 60 min at 37°C; then added 895 ul ligation mix [663 ul of water, 120 ul of 10X NEB T4 DNA ligase buffer, 100 ul of 10% Triton X-100, 12 ul of 10 mg/ml bovine serum albumin, and 5µl of 400 U/µl T4 DNA Ligase (New England Biolabs)]. Ligation was performed at room temperature for 4 hours with slow rotation, then 50 ul of 20 mg/ml proteinase K and 120 µl of 10% SDS were added and incubated at 55°C for 30 minutes. We added 130 ul 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 using the Covaris M220 instrument. Double-size selection was performed using the SPRI select agent (Beckman Coulter) by adding 0.55X volume and then 0.15X volume according to manufacturer's instructions, eluting the final DNA in 300 ul 10 mM Tris pH 8.0.

Biotin pulldown was performed using 150 ul of 10 mg/ml DYNAL™ MyOne™ Dynabeads™ Streptavidin T1 (Invitrogen 65601) per sample. First, the beads were washed twice with 400 ul 1X Tween Washing Buffer (1X TWB: 5mM Tris-HCl (pH 7.5); 0.5mM EDTA; 1M NaCl; 0.05% Tween 20), then they were resuspended in 300 ul of 2X Binding Buffer (2X BB: 10mM Tris-HCl (pH 7.5); 1mM EDTA; 2M NaCl) and added to the 300 ul of sheared and size-selected DNA. We incubated at room temperature for 15 minutes with rotation to bind biotinylated DNA to the streptavidin beads. We washed twice with mixing at 55°C with 600 ul 1X TWB, then washed beads in 100 ul 1X NEB T4 DNA ligase buffer and repaired the ends of the DNA by resuspending beads in 100 ul master mix (88 ul of 1X NEB T4 DNA ligase buffer with 10 mM ATP 2 ul of 25 mM dNTP mix, 5 ul of 10 U/ul NEB T4 PNK, 4 ul of 3 U/ul NEB T4 DNA polymerase I, 1 ul of 5 U/ul NEB DNA polymerase I, Large (Klenow) Fragment) and incubated at room temperature for 30 minutes. We washed twice again with 1X TWB, once in 1X NEBuffer 2, then resuspended 100 ul of dATP

attachment master mix (90 ul of 1X NEBuffer 2, 5 ul of 10 mM dATP, 5 ul of 5 U/ul NEB Klenow exo minus) and incubated for 30 minutes at 37°C. We washed twice in 1X TWB then resuspended beads in 100 ul 1X T4 DNA ligase buffer. We followed the Agilent Sureselect[XT] manufacturer instructions to ligate the paired-end adapters. The beads were then washed twice with 1X TWB, resuspended in 32 ul 1X Tris buffer, and DNA was removed from the streptavidin beads by heating at 98°C for 10 minutes. We then followed the Agilent SureSelect[XT] manufacturer instructions for the pre-capture PCR, using 8 cycles.

Promoter Capture Hi-C library preparation

The RNA baits were designed in Mifsud et al.[14] for capturing *Hin*dIII fragments containing gene promoters (Dr. Cameron Osborne kindly shared the exact design). As described in Mifsud et al.[14], 120-mer RNA baits were designed to target both ends of *Hin*dIII 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 25-65%, contained less than 3 consecutive Ns, and was within 330 bp of HindIII fragment ends. The Hi-C library hybridization to the capture library was performed according to Agilent SureSelect[XT] manufacturer instructions. 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. All six libraries were sequenced together across 2 lanes on the Illumina HiSeq 4000 to produce an average of 127,069,374 (+/−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.[15], using the Hi-C User Pipeline (HiCUP) v0.5.9[59] default settings except the insert size restrictions for the filtering step were set to 200-600 bp. We called significant interactions on each library separately with the Capture Hi-C Analysis of Genome Organization (CHiCAGO) software v1.1.1[60]. We used the default threshold of 5 for calling significant interactions. To create a stringent set of interactions, we merged all pCHi-C final alignments and called interactions using CHiCAGO, again using a threshold of 5; then filtered these interactions to include only those that were called in both biological replicates in at least one condition (BSA, OA, PA).

Cross-species conservation analysis

For the across-species conservation analysis, we used the PhastCons score[18] available on the UCSC ENCODE database. Briefly, the PhastCons score uses a phylo-HMM to predict the 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 +/−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 +/− 500kb. 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[24] 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 +/−500 kb) and used the summary statistics from the lipid GWAS study[11] to estimate the proportion of heritability explained by the 154 lipid-responsive promoters or 323 promoters that interact with lipid-responsive enhancers in adipocytes, for the four lipid GWAS traits: serum triglycerides, HDL cholesterol, LDL cholesterol, and total cholesterol.

Genotype and phenotype data from the UK Biobank cohort

We downloaded the imputed genotype data from the UK Biobank cohort[4]. We identified all of the SNPs in the lipid-responsive gene promoters involved in chromosomal interactions. For the GxE interaction test, we filtered out the SNPs that had genotype missing rate more than 5% or the minor allele frequency (MAF) less than 5%. We used the BMI value collected from the initial UK Biobank assessment visit at which the participants were recruited. The 24-hour recall of saturated fat consumption in diet was collected at 5 different time points including the initial assessment and 4 online-cycle collections. If an individual had the 24-hour recall of the saturated fat collected at multiple time points, we used the value closest to the initial assessment. We then selected the unrelated Caucasian individuals from the UK Biobank participants who had both BMI and saturated fat diet data collected for the GxE analysis. We corrected BMI for the following covariates and performed inverse normal transformation to ensure that BMI is normally distributed: Array type, sex (inferred), age (when attended assessment center), age^2, the assessment center ID, and genetic principle components of 1-20.

Genome-wide GxE scan

To verify that our significant GxE interactions were not caused by an overall inflation, we fitted the linear GxE interaction model on all SNPs across the human genome in the UKBiobank data. We first selected the SNPs that are not in the same LD block ($r^2 < 0.2$) and then used the following linear model to detect the GxE between each LD pruned SNP and saturated fat intake on BMI:

$$Y = \alpha + \beta_G g + \beta_E e + \beta_{GE} ge + \varepsilon \quad \text{Equation 1:}$$

Y is a vector of inverse normalized transformed BMI values and g represents the vector of number of minor alleles in the genotypes of the target SNP for the individuals in the study sample. We use e for the vector of saturated fat intake levels as the environmental covariate and ε represents a vector of random errors, in which each entry is independently and normally distributed. The α and β are the estimated parameters. The test for an interaction is based on the coefficient β_{GE}. A non-zero β_{GE} indicates that there is an interaction between the genotype and environment factor for the outcome phenotype. We constructed a Q-Q plot to compare the p-values of β_{GE} and the expected p-values based on the multiple testing.

Testing for GxE using multi-variable linear model

We used the following multi-variable linear model to detect the GxEs between the SNPs and saturated fat intake for BMI:

$$Y = \alpha + \beta_E e + \sum_{i=1}^{N} (\beta_{Gi} \times g_i) + \sum_{i=1}^{N} (\beta_{GEi} \times g_i e) + \varepsilon \quad \text{Equation 2:}$$

Y is a vector of inverse normalized transformed BMI values and g_i represents a vector of the number of minor alleles in the genotypes of the target SNP i, i=1, …, 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 ε represents a vector of random errors, in which each entry is independently and normally distributed. The α and βs are the estimated parameters. The significance of the interaction is given for the coefficient β_{GEi}. A non-zero β_{GEi} indicates that there is an interaction between the genotype i and environment factor (i.e. the 24-hour saturated fat recall) for the outcome phenotype (i.e. BMI). Individual β_{GEi} are estimated conditioned on the effects of the other genotypes and β_{Gi} from the multi-variable linear model. The p-values given for the individual β_{GEi} are calculated using t-tests.

Electrophoretic mobility shift assay

Nuclear protein was extracted from adipocytes using the nuclear protein extraction kit (Active Motif 40010) following manufacturer instructions. Oligonucleotide probes (15 bp flanking 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 unlabeled probe of the same sequence was added to the reaction mixture at 100X excess. The reaction was incubated for 30 minutes at room temperature, and then loaded on a 6% retardation gel (ThermoFisher Scientific EC6365BOX) that was run in 0.5X TBE buffer. We transferred the contents of the gel to a nylon membrane, and visualized with the chemi-luminescent reagent as recommended.

Cis-expression quantitative trait loci in the METSIM cohort

We obtained the subcutaneous adipose *cis*-eQTL variants identified in the RNA-seq data (n=335) from the Finnish METSIM cohort[15,31].

Reporting summary

Detailed information on experimental design and reagents is also collected on our Reporting Summary, which is published alongside this paper.

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 have been deposited in GEO under accession code GSE129574 and are available upon request from the corresponding author.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements

References

1. Joseph PG, Pare G & Anand SS Exploring Gene-Environment Relationships in Cardiovascular Disease. Canadian Journal of Cardiology 29, 37–45 (2013). [PubMed: 23261319]

2. Heianza Y & Qi L Gene-diet interaction and precision nutrition in obesity. International Journal of Molecular Sciences 18, (2017).

3. Kilpeläinen TO 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, (2011).

4. 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, (2015).

5. Buenrostro JD, Giresi PG, Zaba LC, Chang HY & Greenleaf WJ 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). [PubMed: 24097267]

6. 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). [PubMed: 20513432]

7. Lefterova MI & Lazar MA New developments in adipogenesis. Trends in Endocrinology and Metabolism 20, 107–114 (2009). [PubMed: 19269847]

8. Ernst J & Kellis M Large-scale imputation of epigenomic datasets for systematic annotation of diverse human tissues. Nat. Biotechnol 33, 364–76 (2015). [PubMed: 25690853]

9. Surakka I et al. The impact of low-frequency and rare variants on lipid levels. Nat. Genet 47, 589–597 (2015). [PubMed: 25961943]

10. Teslovich TM et al. Biological, clinical and population relevance of 95 loci for blood lipids. Nature 466, 707–713 (2010). [PubMed: 20686565]

11. Willer CJ et al. Discovery and refinement of loci associated with lipid levels. Nat. Genet 45, 1274–1283 (2013). [PubMed: 24097068]

12. 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). [PubMed: 29403010]

13. MacArthur J et al. The new NHGRI-EBI Catalog of published genome-wide association studies (GWAS Catalog). Nucleic Acids Res. 45, D896–D901 (2017). [PubMed: 27899670]

14. Mifsud B et al. Mapping long-range promoter contacts in human cells with high-resolution capture Hi-C. Nat. Genet 47, 598–606 (2015). [PubMed: 25938943]

15. Pan DZ et al. Integration of human adipocyte chromosomal interactions with adipose gene expression prioritizes obesity-related genes from GWAS. Nat. Commun 9, 1512 (2018). [PubMed: 29666371]

16. 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). [PubMed: 28343628]

17. Wang J, Duncan D, Shi Z & Zhang B WEB-based GEne SeT AnaLysis Toolkit (WebGestalt): update 2013. Nucleic Acids Res. 41, W77–W83 (2013). [PubMed: 23703215]

18. Siepel A et al. Evolutionarily conserved elements in vertebrate, insect, worm, and yeast genomes. Genome Res. 15, 1034–1050 (2005). [PubMed: 16024819]

19. Lek M et al. Analysis of protein-coding genetic variation in 60,706 humans. Nature 536, 285–291 (2016). [PubMed: 27535533]

20. Kathiresan S et al. Common variants at 30 loci contribute to polygenic dyslipidemia. Nat. Genet 41, 56–65 (2009). [PubMed: 19060906]

21. Shin S-Y et al. An atlas of genetic influences on human blood metabolites. Nat. Genet 46, 543–550 (2014). [PubMed: 24816252]

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). [PubMed: 27005778]

23. Vaittinen M et al. FADS2 genotype regulates delta-6 desaturase activity and inflammation in human adipose tissue. J. Lipid Res 57, 56–65 (2016). [PubMed: 26609056]

24. Finucane HK et al. Partitioning heritability by functional annotation using genome-wide association summary statistics. Nat. Genet 47, 1228–1235 (2015). [PubMed: 26414678]

25. Nettleton JA, Brouwer IA, Geleijnse JM & 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). [PubMed: 24794412]

27. Luukkonen PK et al. Saturated Fat Is More Metabolically Harmful for the Human Liver Than Unsaturated Fat or Simple Sugars. Diabetes Care 41, 1732–39 (2018). [PubMed: 29844096]

28. Bycroft C et al. The UK Biobank resource with deep phenotyping and genomic data. Nature 562, 203–209 (2018). [PubMed: 30305743]

29. Robinson MR et al. Genotype-covariate interaction effects and the heritability of adult body mass index. Nat. Genet 49, 1174–1181 (2017). [PubMed: 28692066]

30. Zhou J & Troyanskaya OG Predicting effects of noncoding variants with deep learning–based sequence model. Nat. Methods 12, 931–934 (2015). [PubMed: 26301843]

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). [PubMed: 28119442]

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). [PubMed: 28238358]

33. Gusev A et al. Integrative approaches for large-scale transcriptome-wide association studies. Nat. Genet 48, 245–252 (2016). [PubMed: 26854917]

34. Ardlie KG et al. The Genotype-Tissue Expression (GTEx) pilot analysis: Multitissue gene regulation in humans. Science 348, 648–660 (2015). [PubMed: 25954001]

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). [PubMed: 29087381]

36. Acosta JR et al. Single cell transcriptomics suggest that human adipocyte progenitor cells constitute a homogeneous cell population. Stem Cell Res. Ther 8, 250 (2017). [PubMed: 29116032]

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). [PubMed: 27546599]

38. Griggs R et al. Zaspopathy in a large classic late-onset distal myopathy family. Brain 130, 1477–1484 (2007). [PubMed: 17337483]

39. Fumagalli M et al. Greenlandic Inuit show genetic signatures of diet and climate adaptation. Science 349, 1343–1347 (2015). [PubMed: 26383953]

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). [PubMed: 20484448]

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). [PubMed: 30157936]

42. M V & D A 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). [PubMed: 22955616]

44. Bujold D et al. The International Human Epigenome Consortium Data Portal. Cell Syst. 3, 496–499 (2016). [PubMed: 27863956]

45. Roadmap Epigenomics Consortium et al. Integrative analysis of 111 reference human epigenomes. Nature 518, 317–330 (2015). [PubMed: 25693563]

46. Bernstein BE et al. The NIH roadmap epigenomics mapping consortium. Nature Biotechnology 28, 1045–1048 (2010).

47. Simeonov DR et al. Discovery of stimulation-responsive immune enhancers with CRISPR activation. Nature 549, 111–115 (2017). [PubMed: 28854172]

48. Gate RE et al. Genetic determinants of co-accessible chromatin regions in activated T cells across humans. Nat. Genet 50, 1140–1150 (2018). [PubMed: 29988122]

49. Phan AT, Goldrath AW & Glass CK Metabolic and Epigenetic Coordination of T Cell and Macrophage Immunity. Immunity 46, 714–729 (2017). [PubMed: 28514673]

50. Schneider CA, Rasband WS & Eliceiri KW NIH Image to ImageJ: 25 years of image analysis. Nature Methods 9, 671–675 (2012). [PubMed: 22930834]

51. Corces MR et al. An improved ATAC-seq protocol reduces background and enables interrogation of frozen tissues. Nat. Methods 14, 959–962 (2017). [PubMed: 28846090]

52. Langmead B & Salzberg SL Fast gapped-read alignment with Bowtie 2. Nat Methods 9, 357–359 (2012). [PubMed: 22388286]

53. Li H et al. The Sequence Alignment/Map format and SAMtools. Bioinformatics 25, 2078–2079 (2009). [PubMed: 19505943]

54. Zhang Y et al. Model-based Analysis of ChIP-Seq (MACS). Genome Biol. 9, R137 (2008). [PubMed: 18798982]

55. Hansen KD, Irizarry RA & Wu Z Removing technical variability in RNA-seq data using conditional quantile normalization. Biostatistics 13, 204–216 (2012). [PubMed: 22285995]

56. Robinson MD, McCarthy DJ & Smyth GK edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics 26, 139–40 (2010). [PubMed: 19910308]

57. Rao SSP et al. A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. Cell 159, 1665–1680 (2014). [PubMed: 25497547]

58. Nagano T et al. Comparison of Hi-C results using in-solution versus in-nucleus ligation. Genome Biol. 16, 175 (2015). [PubMed: 26306623]

59. Wingett SW et al. HiCUP: pipeline for mapping and processing Hi-C data. F1000Research 4, 1310 (2015). [PubMed: 26835000]

60. Cairns J et al. CHiCAGO: robust detection of DNA looping interactions in Capture Hi-C data. Genome Biol. 17, 127 (2016). [PubMed: 27306882]
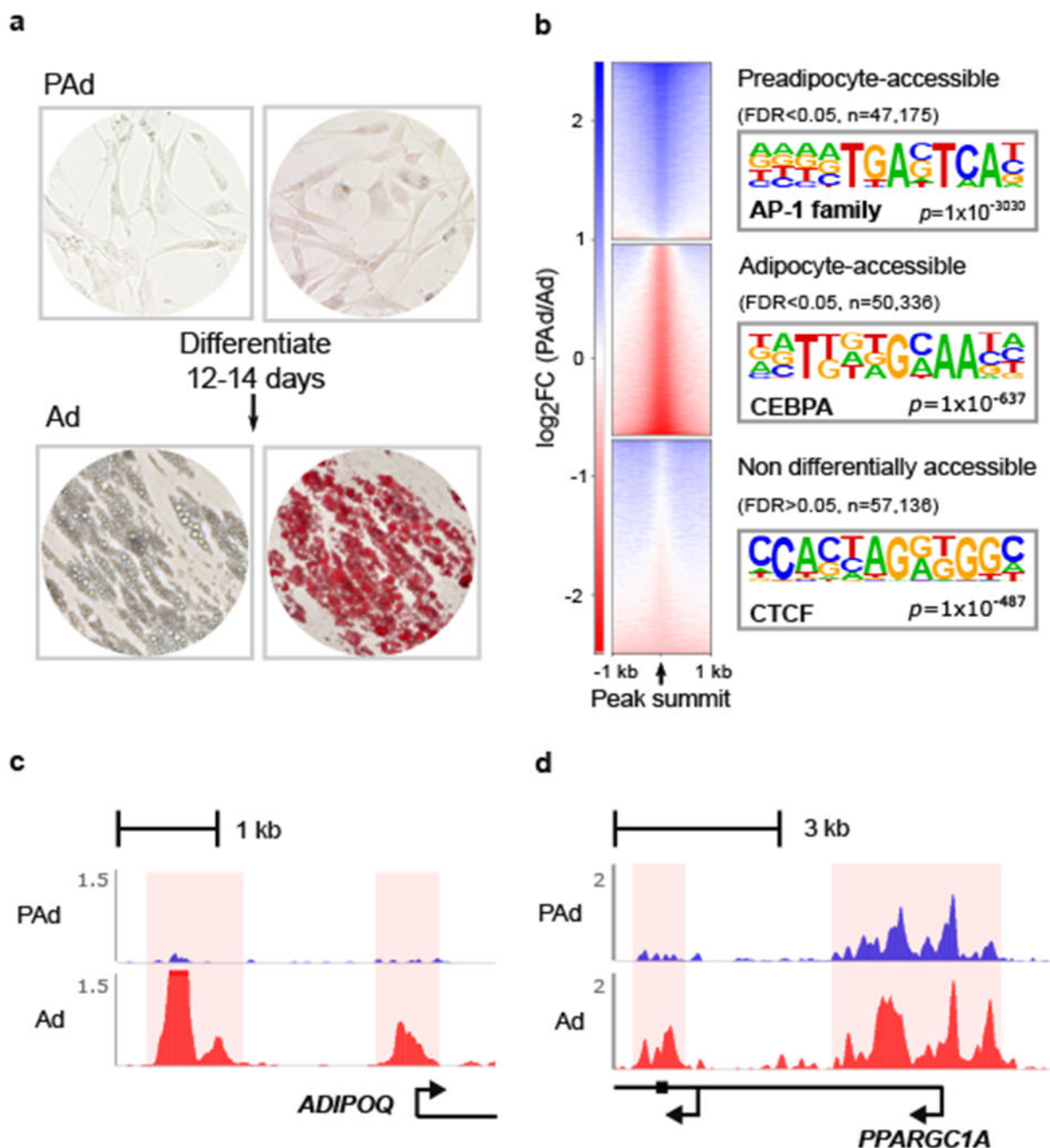
Figure 1. Differential ATAC-seq analysis between human primary preadipocytes and adipocytes indicates successful adipocyte differentiation and wide-spread changes in chromatin accessibility.
(**a**) Bright-field images of preadipocytes (top) and *in vitro* differentiated adipocytes (bottom) for unstained cells (left) and cells stained with oil red o (right). Images are representative examples from 2 independent experiments. (**b**) Heatmaps show $\log_2$ fold-change ($\log_2$FC) of bins per million mapped reads (BPM) for preadipocytes 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 QL F-test (see Methods) for differential accessibility between

preadipocytes and adipocytes using ATAC-seq libraries from n=3 replicates per cell type. The top enriched transcription factor motif for the indicated peak set is listed (right). Enrichment p-values are derived from the hypergeometric enrichment test of the proportion of the given top *de novo* identified[6] TF motifs in the 3 indicated peak sets compared with the background set (see Methods). (**c,d**) Read coverage (BPM) in one representative (out of n=3 replicates per cell type) preadipocyte (blue) and adipocyte (red) ATAC-seq library at the adipocyte-accessible ATAC-peaks in the promoters of (**c**) the adipocyte hormone gene, adiponectin (*ADIPOQ*) and (**d**) the adipocyte-specific marker, peroxisome proliferator-activated receptor gamma coactivator 1-alpha (*PPARGC1A*). PAd indicates preadipocytes; and Ad, adipocytes.
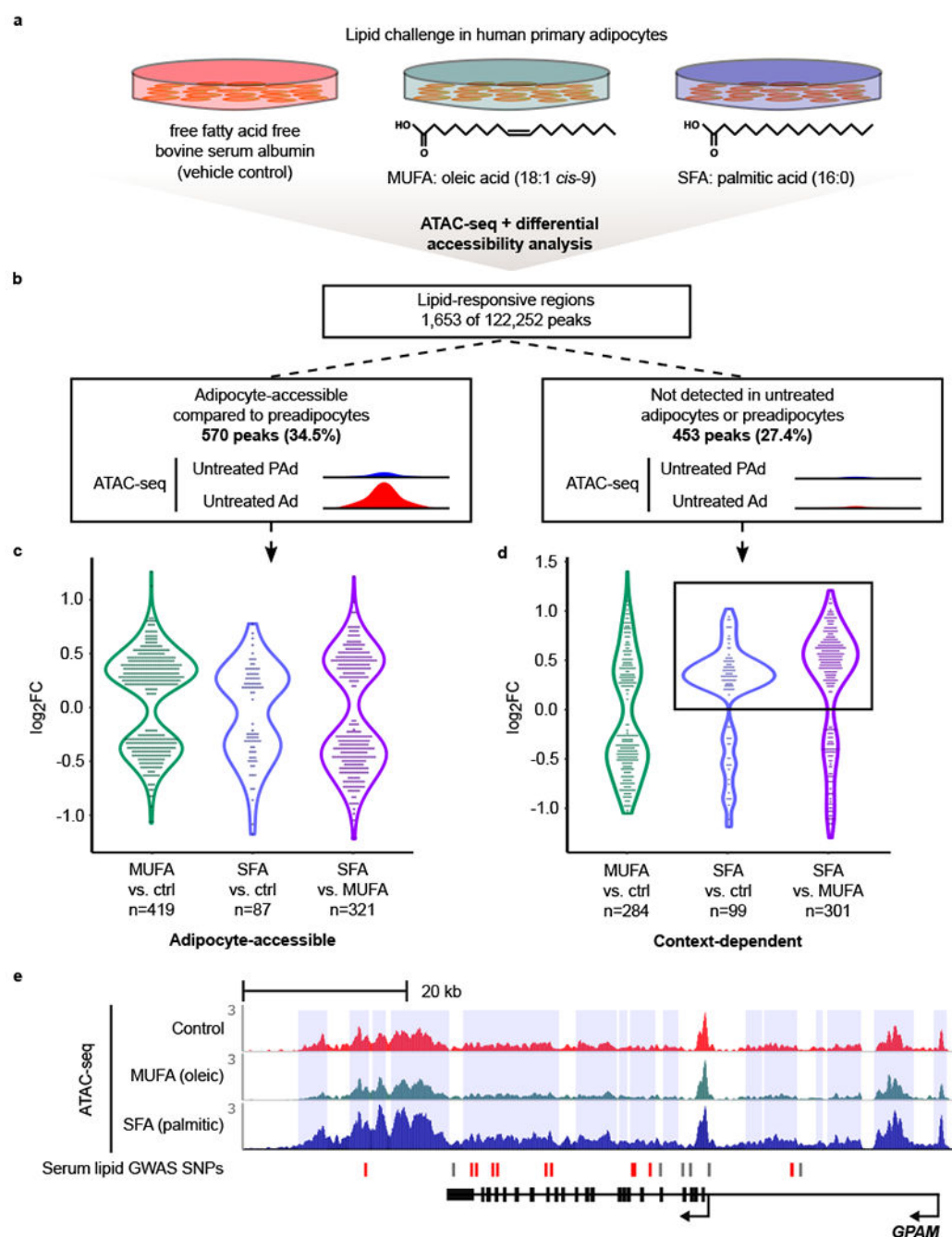
Figure 2. Lipid-responsive regions fall within adipocyte-accessible regions of the genome, as well as within context-dependent regions that are not present in untreated adipocytes.

(**a**) Schematic overview of the lipid challenge experiment in human primary adipocytes. The treatments were performed in 3 replicates per condition. (**b**) Schematic overview indicates the two categories of lipid-responsive peaks used for all downstream analyses. 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. Significant lipid-responsive peaks categorized as adipocyte-

accessible (n=570) or "context-dependent" (i.e. not identified in untreated preadipocytes or adipocytes; n=453) were used in all downstream analyses. (**c,d**) Violin plots show the log$_2$FC of the differentially accessible peaks in the indicated comparisons, stratified by whether the peak was (**c**) adipocyte-accessible, or (**d**) context-dependent. (**c**) The violin plot characteristics are as follows: MUFA vs. ctrl (n=419) range: −1.07 – 1.25; median: 0.26; 25th percentile: −0.34; and 75th percentile: 0.42. SFA vs. ctrl (n=87) range: −1.17 – 0.77; median: 0.15; 25th percentile: −0.34; and 75th percentile: 0.27. SFA vs. MUFA (n=321) range: −1.22 – 1.21; median: −0.26; 25th percentile: −0.48; and 75th percentile: 0.43. (**d**) The violin plot characteristics are as follows: MUFA vs. ctrl (n=284) range: −1.05 – 1.40; median: −0.33; 25th percentile: −0.53; and 75th percentile: 0.37. SFA vs. ctrl (n=99) range: −1.19 – 1.02; median: 0.34; 25th percentile: 0.20; and 75th percentile: 0.47. SFA vs. MUFA (n=301) range: −1.30 – 1.21; median: 0.49; 25th percentile: −0.19; and 75th percentile: 0.65. Box indicates shift toward increased accessibility in SFA-treated cells, specifically observed in the context-dependent peaks. (**e**) Human genome browser snapshot (WashU) of the *GPAM* locus, which harbors 15 lipid-responsive peaks in a ~50-kb region (highlighted in blue vertical rectangles). Fourteen of these peaks are SFA-responsive. Read coverage (BPM) in one representative (of n=3 replicates per condition) control (red), MUFA (green), and SFA (blue) ATAC-seq library. The serum lipid GWAS SNPs are categorized as being in the lipid-responsive peak (red) or outside of the lipid-responsive peak (grey). MUFA indicates monounsaturated fatty acids; SFA, saturated fatty acids; PAd, preadipocytes; and Ad, adipocytes.
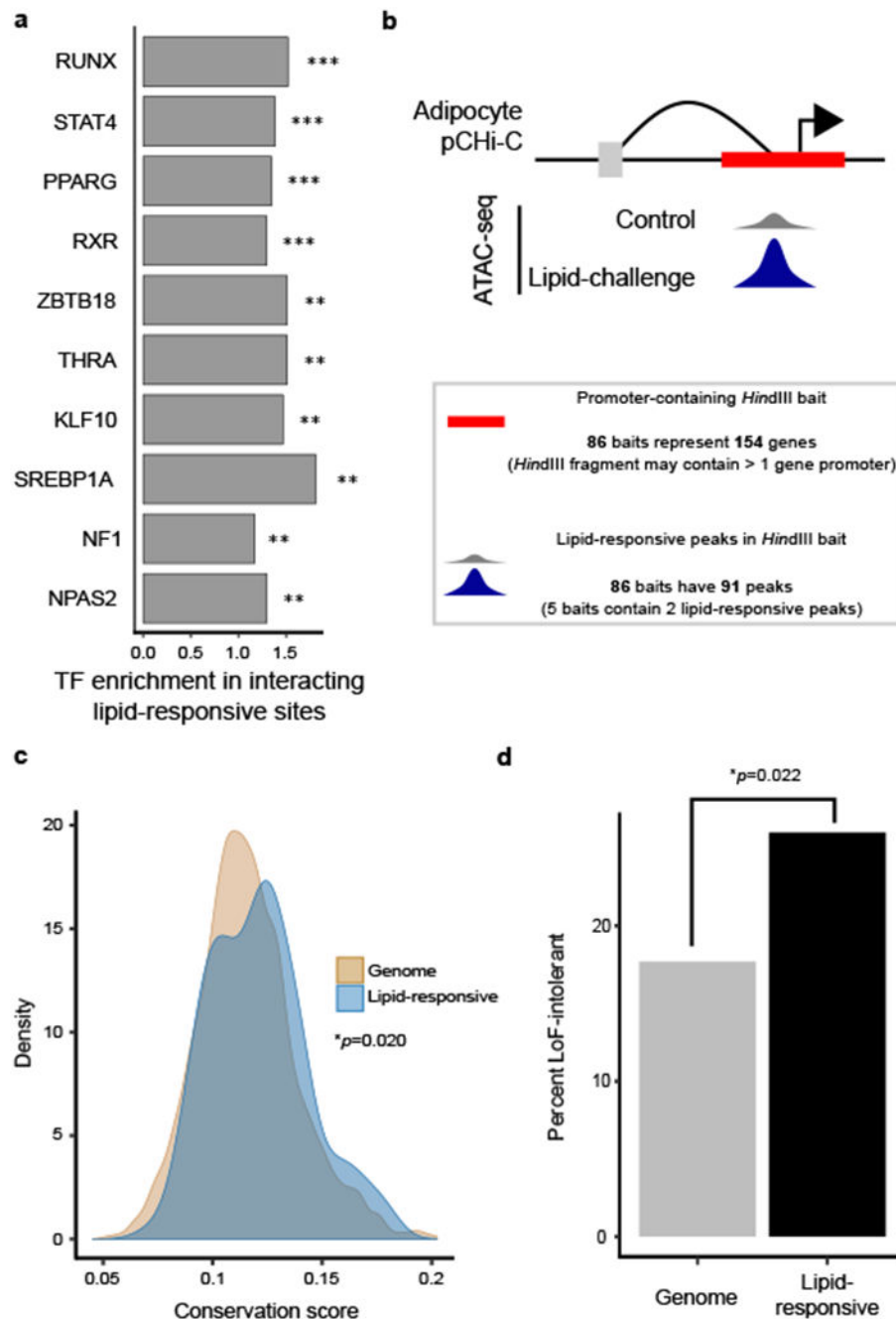
Figure 3. The 154 genes with lipid-responsive promoters within chromosomal interactions exhibit cross-species conservation and constraints on loss-of-function mutations, in line with their potential importance for energy homeostasis and survival.

(**a**) The top 10 enriched TF motifs in the lipid-responsive open chromatin regions in adipocyte chromosomal interactions (n=264) include key TFs in lipid metabolism, such as the co-factors PPARG, and RXR. **FDR < 0.001; ***FDR < 0.0001. Enrichment *p*-values were derived from the hypergeometric enrichment test of proportion of the given TF motif in the peak set compared with the background set of peaks (all non-lipid-responsive peaks in adipocyte chromosomal interactions), adjusted (Benjamini-Hochberg) for the number of

known motifs tested (n=364)[6]. (**b**) Schematic of the data we integrated to identify the 154 genes with lipid-responsive promoters in adipocyte chromosomal interactions. (**c**) Density plot shows the distribution of per-gene average conservation scores across placental mammals[18] for all protein-coding genes in the genome (n=19,316) compared to all protein-coding genes in the set of 154 genes with lipid-responsive promoters (n=114). *depicts the *p*-value for two-sided Wilcoxon signed-rank test. (**d**) 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)[19] in the whole genome (n=3,204/18,122; 17.7%) compared to the protein-coding genes among the 154 genes (n=27/104; 26.0%). LoF indicates loss-of-function; *depicts the *p*-value for the hypergeometric enrichment test.
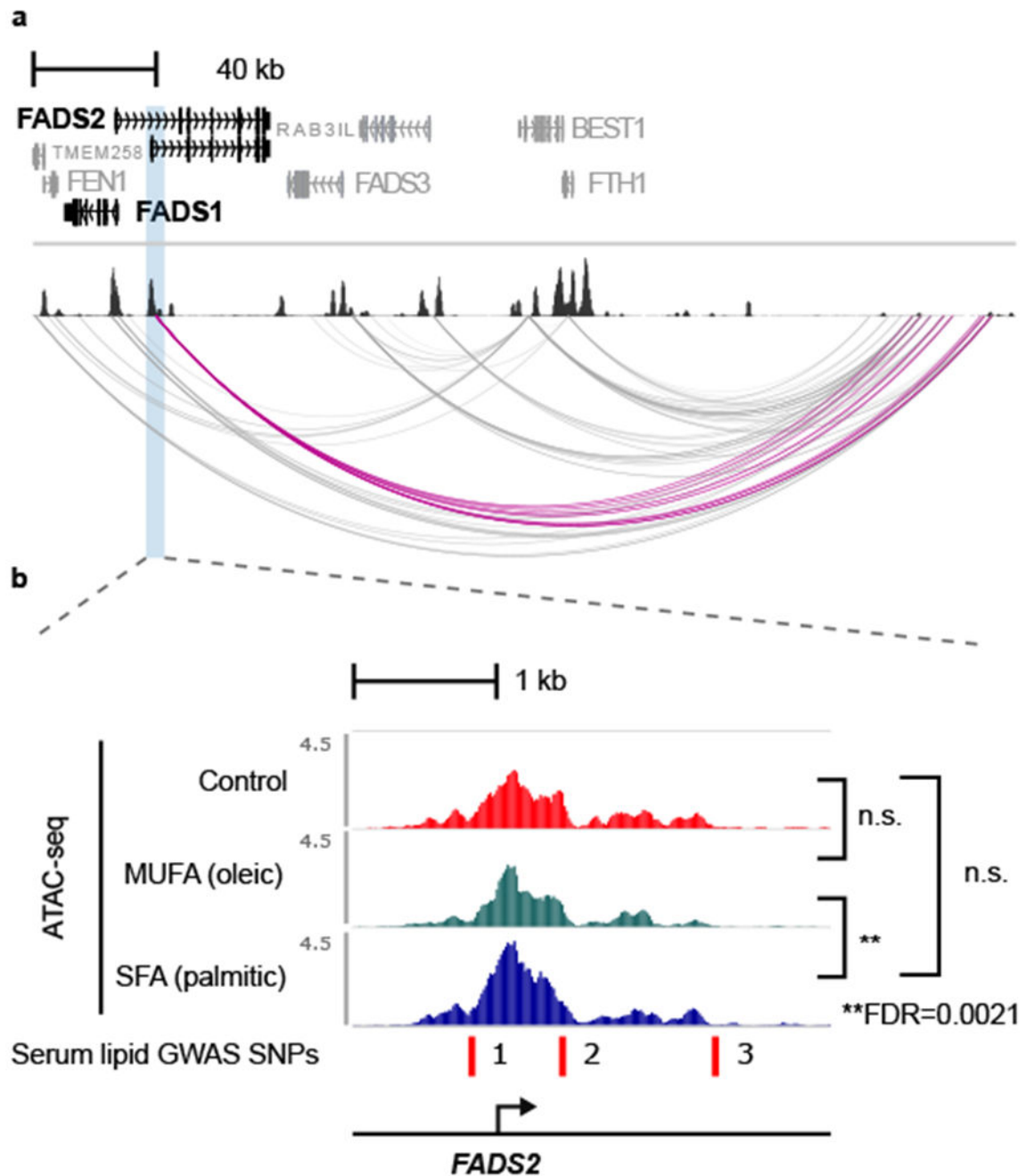
Figure 4. A lipid-responsive open chromatin region in human primary adipocytes at chr11q12.2 *FADS1/2/3* locus harbors GWAS SNPs for serum lipid traits.
(**a**) Genome browser snapshot shows the *FADS1/2/3* locus with adipocyte baseline ATAC-seq [one representative example from n=3 vehicle control (BSA) ATAC-seq libraries] and pCHi-C data (interactions identified in at least one condition from the adipocyte lipid challenge pCHi-C are included; see Methods). The *FADS2* promoter chromosomal interactions are highlighted in magenta. (**b**) Read coverage (BPM) in one representative (out of n=3 replicates per condition) vehicle control (red), MUFA (green), and SFA (blue) ATAC-

seq library. The lipid-responsive peak in one of the *FADS2* promoters is more accessible in SFA- compared to MUFA-treated adipocytes and contains 3 independent GWAS SNPs for serum lipid traits (1 indicates rs191508698; 2, rs968567; and 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 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; n.s. indicates non-significant (based on the *post hoc* test of the one-way ANOVA); MUFA, monounsaturated fatty acids; and SFA, saturated fatty acids.
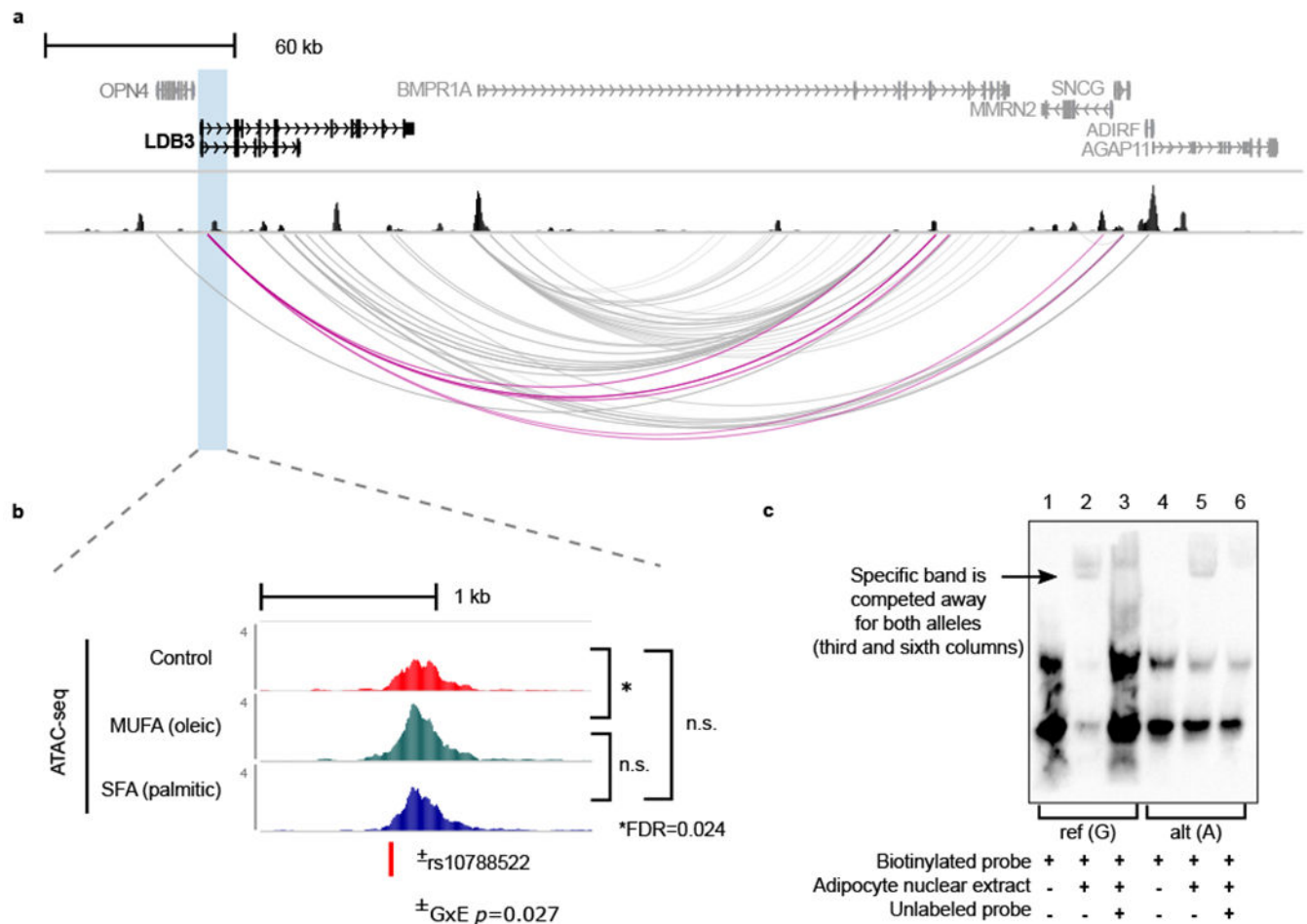
Figure 5. Fine-mapping of the gene-diet interaction for BMI at the *LDB3* promoter region.
(**a**) Genome-browser snapshot of the *LDB3* locus with adipocyte baseline ATAC-seq [one representative example from n=3 vehicle control (BSA) ATAC-seq libraries] and pCHi-C data (interactions identified in at least one condition from the adipocyte lipid challenge pCHi-C are included; see Methods). The *LDB3* promoter interactions are highlighted in magenta. (**b**) Read coverage (BPM) in one representative (out of n=3 replicates per condition) control (red), MUFA (green), and SFA (blue) ATAC-seq library. The lipid-responsive ATAC-seq peak harboring the significant GxE SNP rs10788522 resides in the *LDB3* gene. (**c**) Electrophoretic mobility shift assay (EMSA) shows binding of adipocyte nuclear protein to the GxE SNP rs10788522. *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; n.s. indicates non-significant (based on the *post hoc* test of the one-way ANOVA); and $\pm$depicts the *p*-value for the β_{GEi} for rs10788522 from the multi-variable linear model testing for GxEs in the UK Biobank.
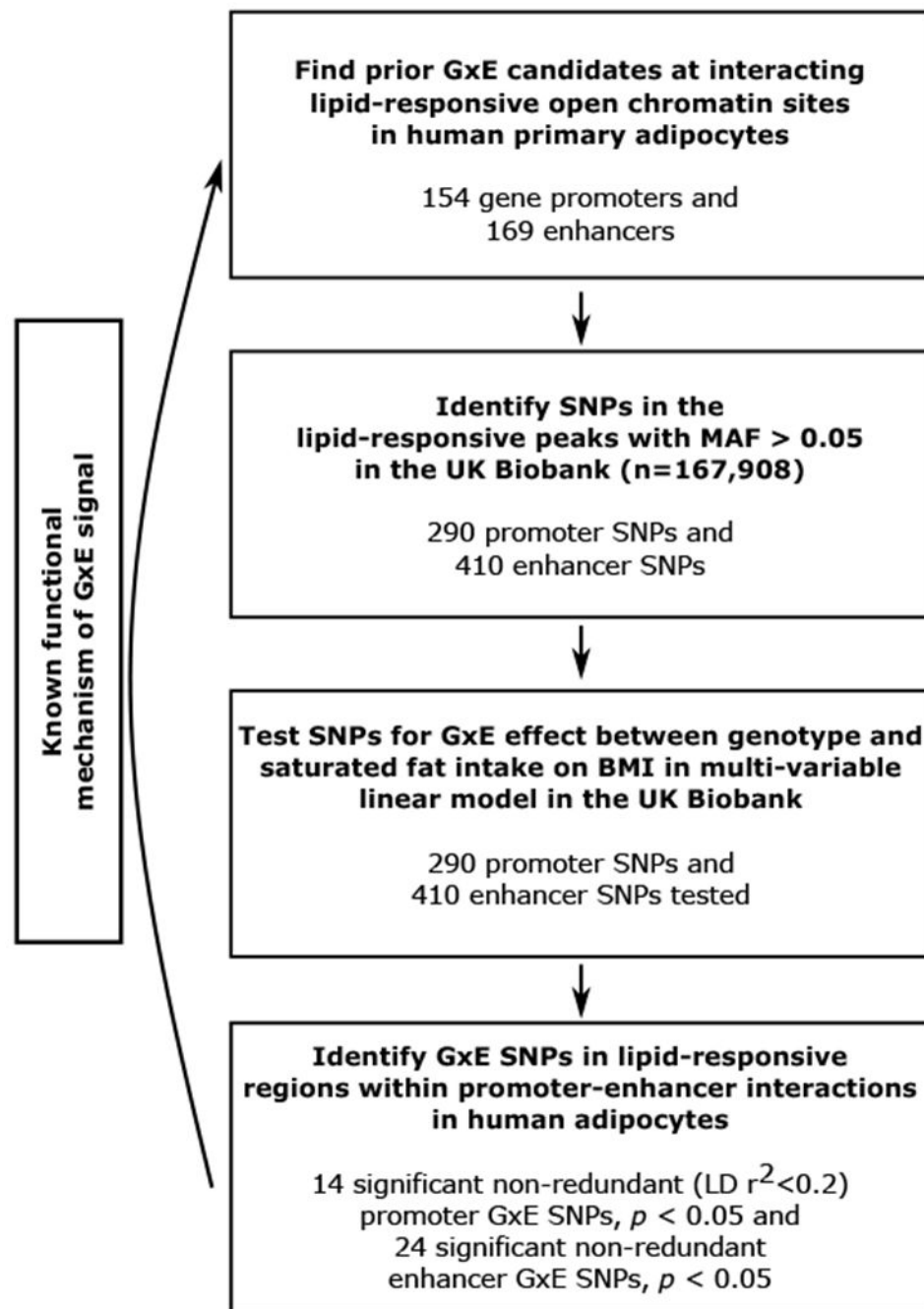
Figure 6. Analytical approach.
Flowchart of our approach to integrate molecular genomics data created in human adipocytes in physiologically relevant contexts, that, when combined with human cohort molecular and phenotype data, enables the detection of GxE signals.

Table 1.

Five lipid-responsive ATAC-peaks in interacting promoters overlap with GWAS SNPs[11] for serum lipid traits.

Peak chr	Peak start	Peak end	Gene[¶]	SNP(s) in peak	MAF[†]	Associated trait[11§]	p-value[11§]	Index SNP[11]	LD with index SNP[†] (r²)
11	61594652	61596828	FADS2/FADS1	rs99780 rs968567 rs191508698	0.37 0.16 0[†]	TG/HDL/**LDL**/TC	2.32e-16/5.52e-09/**2.39e-21**/8.62e-18 3.4e-09/n.s./**8.91e-11**/2.27e-09 1e-16/2.37e-08/**1.24e-20**/3.1e-17	rs174546	0.93 0.35 -
7	73036880	73038991	MLXIPL	rs55747707 rs34060476	0.2 0.13	TG	3.55e-44 9.73e-46	rs17145738	0.47 0.77
2	27432323	27432971	SLC5A6/ATRAID	rs2580759 rs1275530	0.22 0.72	TG	2.18e-17 1.88e-17	rs1260326	- -
16	68115758	68116375	NFATC3	rs2107269	0.015	HDL	5.07e-45	rs16942887	0.66
19	10981139	10983631	CARM1	rs12460421	0.44	LDL	4.32e-11	rs6511720	-

[¶]The gene listed is the promoter in the baited *Hind*III fragment with a lipid-responsive ATAC-seq peak;

[†]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;

[§]Most significant association is in bold when SNP is associated with more than one serum lipid trait. 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.[11]

Table 2.

Significant GxEs affecting BMI from multi-variable linear model for 290 promoter SNPs in lipid-responsive ATAC-seq peaks.

SNP	p-g	p-g*e	β-g	β-g*e	Genes in Bait	cis-eQTL FDR[15][†]	Target Gene[15]	log$_2$FC[15]
rs1974817[§]	0.0089	0.0010	2.3	−0.089	GLTSCR2/SNORD23	2.4E-31	SEPW1	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[§]	0.0017	0.0050	−0.50	0.012	CARM1	2.7E-05/0.015	SMARCA4/ICAM4	−0.28/0.38
rs17625418	0.015	0.0054	−2.2	0.073	GLTSCR2/SNORD23	1.3E-13 0.0038	SEPW1 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/SPTBN4	-	-	-
rs35678764[§]	0.028	0.025	−5.7	0.19	RDH8/COL5A3	-	-	-
rs10788522[§]	0.013	0.027	3.9	−0.14	LDB3	6.1E-04	LDB3	0.42
rs10422283	0.045	0.029	−0.072	0.0023	LIPE/LIPE-AS1	-	-	-
rs867773[§]	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	TNFRSF1B/MIR4632	-	-	-

[†]The cis-eQTLs were identified in the adipose tissue from the METSIM cohort[15,31].

[§]When more than one non-independent SNP (LD r^2 > 0.2) has a significant GxE 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. 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.