



Article scientifique

Article

2016

Published version

Open Access

This is the published version of the publication, made available in accordance with the publisher's policy.

Adiposity-Dependent Regulatory Effects on Multi-tissue Transcriptomes

Glastonbury, Craig A.; Viñuela, Ana; Buil Demur, Alfonso Alberto; Halldorsson, Gisli H.; Thorleifsson, Gudmar; Helgason, Hannes; Thorsteinsdottir, Unnur; Stefansson, Kari; Dermitzakis, Emmanouil; Spector, Tim D.; Small, Kerrin S.

How to cite

GLASTONBURY, Craig A. et al. Adiposity-Dependent Regulatory Effects on Multi-tissue Transcriptomes. In: American Journal of Human Genetics, 2016, vol. 99, n° 3, p. 567–579. doi: 10.1016/j.ajhg.2016.07.001

This publication URL: <https://archive-ouverte.unige.ch/unige:99517>

Publication DOI: [10.1016/j.ajhg.2016.07.001](https://doi.org/10.1016/j.ajhg.2016.07.001)

Adiposity-Dependent Regulatory Effects on Multi-tissue Transcriptomes

Craig A. Glastonbury,¹ Ana Viñuela,^{1,8} Alfonso Buil,² Gisli H. Halldorsson,³ Gudmar Thorleifsson,³ Hannes Helgason,^{3,4} Unnur Thorsteinsdottir,^{3,5} Kari Stefansson,^{3,5} Emmanouil T. Dermitzakis,^{2,6,7} Tim D. Spector,¹ and Kerrin S. Small^{1,*}

Obesity is a global epidemic that is causally associated with a range of diseases, including type 2 diabetes and cardiovascular disease, at the population-level. However, there is marked heterogeneity in obesity-related outcomes among individuals. This might reflect genotype-dependent responses to adiposity. Given that adiposity, measured by BMI, is associated with widespread changes in gene expression and regulatory variants mediate the majority of known complex trait loci, we sought to identify gene-by-BMI ($G \times \text{BMI}$) interactions on the regulation of gene expression in a multi-tissue RNA-sequencing (RNA-seq) dataset from the TwinsUK cohort ($n = 856$). At a false discovery rate of 5%, we identified 16 *cis* $G \times \text{BMI}$ interactions (top *cis* interaction: *CHURC1*, rs7143432, $p = 2.0 \times 10^{-12}$) and one variant regulating 53 genes in *trans* (top *trans* interaction: *ZNF423*, rs3851570, $p = 8.2 \times 10^{-13}$), all in adipose tissue. The interactions were adipose-specific and enriched for variants overlapping adipocyte enhancers, and regulated genes were enriched for metabolic and inflammatory processes. We replicated a subset of the interactions in an independent adipose RNA-seq dataset (deCODE genetics, $n = 754$). We also confirmed the interactions with an alternate measure of obesity, dual-energy X-ray absorptiometry (DXA)-derived visceral-fat-volume measurements, in a subset of TwinsUK individuals ($n = 682$). The identified $G \times \text{BMI}$ regulatory effects demonstrate the dynamic nature of gene regulation and reveal a functional mechanism underlying the heterogeneous response to obesity. Additionally, we have provided a web browser allowing interactive exploration of the dataset, including of association between expression, BMI, and $G \times \text{BMI}$ regulatory effects in four tissues.

Introduction

Obesity (MIM: 601665) is a global epidemic that has been robustly associated with a range of co-morbidities such as cardiovascular disease, insulin resistance, type 2 diabetes (T2D [MIM: 125853]), and increased risk of certain types of cancer.^{1,2} However, at the individual level, co-morbidity development among obese individuals is heterogeneous, suggesting that genetics and adiposity might interact to mediate downstream disease and complex trait development.³ The existence of gene-by-adiposity interactions on complex traits is supported by twin studies demonstrating that BMI modifies the heritability of co-morbid traits such as blood pressure and insulin sensitivity.^{4,5}

Obesity manifests as excess adipose tissue and has a systemic effect on bodily function. BMI-associated genetic variants are enriched for hypothalamic processes,⁶ which suggests that the variants that cause obesity exert their effects primarily in the brain. In contrast, variants for many obesity co-morbid traits are primarily thought to regulate genes active in certain peripheral tissues, such as adipose, muscle, and the liver (insulin resistance),⁷ the heart and endothelial cells (QT-interval, which is predictive of cardiovascular disease),⁸ and adipose (body-fat distribution),⁹ for example. Therefore, BMI could influence co-morbidity development by modifying gene expression in relevant

peripheral tissues directly or by interacting with regulatory variants active in those tissues.

Gene-by-environment ($G \times E$) interactions have been identified extensively in model organisms,^{10–13} but identifying $G \times E$ interactions on complex traits in humans has proven to be difficult due to the need to obtain accurately measured environmental exposures and large sample sizes. In contrast, studies of $G \times E$ effects on the regulation of gene expression, often termed context-specific expression quantitative trait loci (eQTL) analyses, have been more successful and can capture environments that act on the cellular, tissue, or organismal level. Regulatory $G \times E$ effects have been discovered for various environments, including age and sex,¹⁴ in vivo and ex vivo treatment response,^{15–18} and tissue and/or cell of residence.^{19,20} Twin studies agnostic to the underlying environment have also identified $G \times E$ regulatory variants.^{21,22} Most genetic variants identified through genome-wide association studies (GWASs) are thought to be regulatory variants; thus, utilizing gene expression to identify gene-by-environment interactions is a promising strategy for identifying factors that interact with disease-relevant regulatory variation.²² Because 60% of disease-associated eQTLs are tissue dependent, it is critical to study the genetic regulation of expression in the appropriate disease-relevant tissues.²³

¹Department of Twin Research and Genetic Epidemiology, King's College London, London SE1 7EH, UK; ²Department of Genetic Medicine and Development, University of Geneva Medical School, Geneva 1211, Switzerland; ³deCODE Genetics, Sturlugata 8, Reykjavik IS-101, Iceland; ⁴School of Engineering and Natural Sciences, University of Iceland, Reykjavik 107, Iceland; ⁵Faculty of Medicine, University of Iceland, Reykjavik 101, Iceland; ⁶Institute for Genetics and Genomics in Geneva (iGE3), University of Geneva, Geneva C1211, Switzerland; ⁷Swiss Institute of Bioinformatics, 1015 Lausanne, Switzerland; ⁸Present address: Department of Genetic Medicine and Development, University of Geneva Medical School, Geneva 1211, Switzerland

*Correspondence: kerrin.small@kcl.ac.uk

<http://dx.doi.org/10.1016/j.ajhg.2016.07.001>

© 2016 The Authors. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

In this study, we considered BMI as a physiological environment and conduct a genome-wide search for gene-by-BMI ($G \times \text{BMI}$) interactions on the regulation of gene expression by utilizing a multi-tissue (adipose, skin, whole blood, and lymphoblastoid cell lines [LCLs]) RNA-sequencing (RNA-seq) dataset including data collected from 856 healthy female twins. We first demonstrated that BMI has a pervasive effect on gene expression in peripheral tissues and that the strongest effects are observed in adipose tissue. We then identified 16 significant *cis* $G \times \text{BMI}$ interactions (false discovery rate [FDR] = 5%) on the regulation of gene expression in adipose tissue and we provide evidence of replicated examples in an independent adipose tissue cohort (deCODE, $n = 754$). The $G \times \text{BMI}$ regulatory interactions are adipose specific and are enriched for metabolic and inflammatory pathways. By extending the analysis in *trans* we were able to identify one *cis* $G \times \text{BMI}$ variant that regulates the expression of 53 genes in *trans* in a BMI-dependent manner. Regulatory interactions such as these will have increasingly important utility in characterizing the functionality and context specificity of genetic variation discovered via traditional GWAS approaches. To enable exploration of our results, we have developed an interactive website that allows researchers to model and plot data in real-time (see the [Web Resources](#)). Using this service, users can investigate genes and SNPs of interest for $G \times \text{BMI}$ effects and explore the relationship between BMI and exon-level expression in four separate tissues.

Material and Methods

Sample Collection

The study included 856 healthy female twins who are a part of the TwinsUK registry and are all of European ancestry. Punch biopsies of subcutaneous adipose tissue from a photo-protected area of the stomach adjacent and inferior to the umbilicus were obtained from consenting individuals. Skin from the punch biopsy was then dissected to separate it from adipose tissue, and both samples were weighed and immediately frozen with liquid nitrogen. Peripheral-blood samples were also collected as part of the study, and LCLs were generated via transformation of the B-lymphocyte fraction with Epstein-Barr virus (EBV). The European Collection of Cell Cultures agency performed the transformation process. All the procedures followed were in accordance with the ethical standards of the St. Thomas Research Ethics Committee (reference 07/H0802/84) at St. Thomas Hospital in London. Volunteers gave informed consent and signed a consent form before the biopsy procedure. Volunteers were supplied with an appropriate detailed information sheet regarding the research project and biopsy procedure by post before attending the biopsy.

Genotyping and Imputation

Genome-wide SNP data for the TwinsUK individuals were generated as previously described.^{6,25,26} In short, TwinsUK samples were genotyped on a combination of platforms (HumanHap300, HumanHap610Q, and 1M-Duo Illumina arrays). Quality control and merging of the array datasets has previously been described

in detail.²⁵ The cleaned data were pre-phased with IMPUTE2 with no reference panel and were then imputed into the 1000 Genomes phase 1 reference panel (interim, data freeze accessed November 10, 2010; the 1000 Genomes Project Consortium 2012).²⁷ Variants with an INFO score >0.8 on all platforms and a MAF $>5\%$ were retained for analysis.

Phenotype Collection

Height, weight, and visceral fat volume were measured at the time of biopsy. Visceral fat volume was measured via dual-energy X-ray absorptiometry (DXA; Hologic QDR 4500 Plus), according to the standard manufacturer's protocol.

RNA-Seq

RNA-seq data were generated as previously described.²² In brief, samples were prepared for sequencing with the unstranded Illumina TruSeq sample preparation kit and sequenced on a HiSeq 2000 machine. The 49 bp paired-end reads were aligned to the UCSC Genome Browser GRCh37 reference genome with the Burrows-Wheeler Aligner.²⁸ GENCODE v.10 was used to annotate genes. Samples were excluded if they failed to have more than 10 million reads map to known exons or if the sequence data did not correspond to actual genotype data.

Exon Quantification and Normalization

To quantify exons, all overlapping exons of a gene were merged into one meta-exon. We counted reads as mapping to a given meta-exon if either of its start or end coordinates overlapped a meta-exon boundary. All read-count quantifications were corrected for variation in sequencing depth between samples by normalizing the number of reads to the median number of well-mapped reads. We only used exons that were quantified in more than 90% of the individuals. Exon expression values were rank-based inverse normal transformed for downstream analysis.

Transcriptome-wide Association Analysis

To determine whether expressed exons were associated to a phenotype of interest (BMI/visceral fat), each exon was treated as a quantitative trait in a linear mixed-effects model implemented with the lme4 package.²⁹ Phenotypic data were treated as continuous independent traits. A full model with the phenotype fitted (**Model 1**) was compared to a null model in which the same model was fitted but the phenotype (BMI) was omitted. These models were compared with a one-degree-of-freedom ANOVA. All known technical variables were included as covariates in the model. To model the twin structure in our data, we included two multi-level indicator variables, termed family and zygosity, as random effects in the model. These variables are coded as follows: if individuals i and j are co-twins, we code $\text{family}_i = \text{family}_j$, if individuals i and j are monozygotic co-twins, we code $\text{zygosity}_i = \text{zygosity}_j$, and if they are dizygotic co-twins, we code $\text{zygosity}_i \neq \text{zygosity}_j$. The zygosity term captures the increased genetic relatedness within an MZ twin pair as compared to a DZ twin pair. If individuals i and j are unrelated, $\text{family}_i \neq \text{family}_j$ and $\text{zygosity}_i \neq \text{zygosity}_j$. We estimated a FDR within each tissue by using the package "qvalue"³⁰ to obtain q values that correspond to a FDR of 5%. Model 1 is as follows:

$$y_i \sim X\beta + Zu + \varepsilon, \quad (\text{Model 1})$$

where y_i is the i^{th} non-PEER (probabilistic estimation of expression residuals)-corrected exon expression vector. X is a design matrix of

fixed effects: phenotype (BMI or visceral fat), age, age squared, mean GC content, and insert size mode. Z represents all random effects: primer index, batch (blood only), family, and zygosity. ϵ is a residual error term representing unaccounted for variation in expression.

cis G × BMI Interaction Discovery

Expression residuals corrected for family structure and technical variables and 50 PEER³¹ factors were used to identify G × BMI interactions (Model 2) with all SNPs within 1 Mb of the transcription start site. The interaction test was implemented with the “modellinear_cross” function in the Matrix eQTL R package.³² Model 2 is as follows:

$$y \sim I + \beta A + \beta A^2 + \beta P + \beta SNP + \beta P \times SNP + \epsilon, \quad (\text{Model 2})$$

where y = expression, I = intercept, A = age, A^2 = age squared, P = phenotype (BMI), SNP = allele dosage, and ϵ = residual error.

Correction for Hidden Confounders

A common approach when performing eQTL mapping has been to adjust expression profiles for unknown latent variables. The aim is to capture hidden factors that confound gene expression measurements, such as unmeasured technical variables (e.g., batch effects). However, these methods can also capture biological factors, and thus, depending on the analysis performed, might not be appropriate. We utilized PEER, a Bayesian method similar to surrogate variable analysis, which has been shown to increase the ability to detect genetic and interaction effects threefold.³¹ We calculated latent factors by applying PEER to expression residuals that had been corrected for family structure and technical variables. PEER was run using no additional covariates and accounting for mean expression. In adipose tissue, three of the first five PEER factors were significantly correlated to BMI ($|r_{\max}| = 0.52$, $p = 1.2 \times 10^{-51}$, Figure S1). We used PEER-corrected data for *cis* G × BMI interaction analysis, but not for direct association of BMI with expression levels or for *trans* G × BMI interaction analysis (see below). Removing BMI-associated PEER factors maximizes the ability to discover interactions with BMI by minimizing model co-linearity (Figures S1 and S2). It has also been demonstrated that removing latent factors, even those highly correlated with the environment of interest, results in the discovery of significantly more interactions on gene expression.²⁰

FDR Estimation

The FDR for calling an interaction significant was estimated based on an approximated permutation strategy as previously described and implemented.^{33,34} Expression residuals were calculated in which all main effects (BMI, SNP, and age) were regressed out. Expression values were permuted for each exon while preserving genotype structure. The interaction-term (BMI × SNP) significance was calculated for each exon. All p values were stored to calculate a genome-wide FDR. Because genes vary in the number of exons they are composed of, separate FDRs were calculated by classifying genes based on a similar number of exons, determined from the distribution of exons expressed in each tissue. A 5% FDR was calculated by computing the ratio of permuted test statistics more significant to the observed interactions, divided by the number of permutations performed ($n = 100$). All interactions with a corrected FDR p value <0.05 were classified as significant and taken forward for replication.

trans G × BMI Discovery

Given that *cis* effects are enriched for *trans* effects and we were underpowered to perform a genome-wide *trans*-interaction analysis, we utilized a two-step procedure to identify *trans* G × BMI interactions in adipose tissue. We first repeated the *cis* G × BMI scan as described above, using expression residuals corrected for family structure and known technical variables but not corrected for PEER factors. PEER correction removes broadly acting variance components, including the effects of multi-gene *trans*-regulators, and is therefore not appropriate for *trans* eQTL analysis. The resulting four significant non-PEER-corrected *cis* G × BMI variants (rs1464171, rs3851570, rs113368712 and rs35662778; FDR = 5%) were then tested for *trans* G × BMI effects. *Trans* effects were tested for all genes at a distance greater than 5 MB from the variant or on a different chromosome. *Trans* G × BMI significance was assessed with a strict Bonferroni correction, accounting for testing four variants against 116,643 exons and corresponding to a p value threshold of 1.1×10^{-7} .

deCODE Replication Cohort

The replication cohort consists of 754 RNA-seq subcutaneous adipose samples obtained from Icelandic individuals, all of whom had imputed genotypes, as previously described.³⁵ The sample set is comprised of 333 males and 421 female participants with an age 47 ± 14 (mean ± SD) and BMI 30 ± 6.6 . RNA-seq reads were aligned to *Homo sapiens* build 38 (UCSC Genome Browser) with TopHat³⁶ v.2.0.12 with a supplied set of known transcripts in GTF format (RefSeq, NCBI). TopHat was configured such that it first attempts to align reads to the provided transcriptome, then, for reads that do not map fully to the transcriptome, it attempts to map them onto the genome. Overlapping exons were merged into one meta-exon. Fragments were counted if they mapped to a meta-exon and if either read in the pair had the start or end (aligned) coordinate overlapping a meta-exon boundary. We excluded exons that had zero fragment counts for more than 90% of individuals. Counts were normalized for number of reads mapped and exon lengths and were rank normal transformed. The following covariates were used to correct for technical differences in the RNA experiments: average fragment length, exonic rate, number of genes detected, number of mapped read pairs, number of alternative alignments, and percentage of reads originating from coding bases (PCT), along with 50 inferred hidden factors which were evaluated with PEER. The technical RNA-seq quality metrics were gathered with the CollectRnaSeqMetrics tool in Picard v.1.79 and RNA-SeQC v.1.1.6.³⁷

Replication Analysis

The 16 *cis* G × BMI interactions (FDR = 5%) were taken forward for replication in deCODE. 13 genes were available in the deCODE dataset. The two datasets were called with different genome builds and gene annotation sets, preventing simple mapping of meta-exons. We thus used liftover to map the TwinsUK meta-exon coordinates from GrCh37 to GrCh38 (UCSC Genome Browser) and defined corresponding meta-exons as those with more than a 90% overlap in length in both annotation sets. This strategy identified corresponding meta-exons for eight genes (*CHURC1* [MIM: 608577], *CIDEA* [MIM: 604440], *ZNF117* [MIM: 194624], *PEPD* [MIM: 613230], *ANXA5* [MIM: 131230], *HLA-DQB2* [MIM: 615161], *IFNARI* [MIM: 107450], and *SCFD2*). An additional three genes (*ADH1A* [MIM: 103700], *SPAG17* [MIM: 616554], and

ERV3-1 [MIM: 131170]) had a partial overlap, ranging from 33%–85%, with a deCODE meta-exon Two genes (*PHACTR3* [MIM: 608725] and *CAST* [MIM: 114090]) had minimal overlap of 4% with a deCODE meta-exon, indicating the meta-exon does not represent an analogous quantification. Given that multiple TwinsUK *PHACTR3* meta-exons were significant, and the second most significant TwinsUK meta-exon (chr20: 58,349,298–58,349,545, p value = 1.7×10^{-8} in TwinsUK) exactly overlapped a deCODE meta-exon, we used the meta-exon corresponding to the second TwinsUK signal for the replication analysis of *PHACTR3*. Exact meta-exon coordinates are provided in Table S1. Given the low annotation overlap at *CAST* (4%), we tested all *CAST* exons for replication and corrected for the number of exons tested in deCODE (30). The three genes unavailable for replications were *RP11-71E19.1*, which was not present in the RefSeq gene annotation used by deCODE, *POU6F2* (MIM: 609062), which did not pass quality control in deCODE (low expression), and *SIK1* (MIM: 605705), which was quantified, but the corresponding SNP rs12482956 and proxies were unavailable in the deCODE imputation.

Integration with GWAS

Overlap of the 16 $G \times \text{BMI}$ lead SNPs with GWAS variants was determined by searching the National Human Genome Research Institute (NHGRI) database (accessed June 19, 2015) for each SNP or proxy SNPs ($r^2 > 0.6$). The database was filtered to include only genome-wide significant loci. We tested the overlap identified at *ADH1A* for colocalization by implementing the regulatory trait concordance (RTC) method.⁵³ The RTC score assesses evidence for causality by testing whether the effect of a GWAS SNP abrogates the effect of an eQTL while accounting for the local linkage disequilibrium (LD) structure of the locus. All SNPs with $\text{MAF} > 5\%$ were in the window 100.1–100.6 Mb, a 250 kb window centered on the index SNP tested. For each of the 1,129 variants, we fitted an interaction (Model 3) and main effect (Model 4) model:

$$ADH1A_expression \sim \text{BMI} \times \text{SNP}_N + \text{BMI} \times \text{rs1693457} \quad (\text{Model 3})$$

and

$$ADH1A_expression \sim \text{SNP}_N + \text{BMI} \times \text{rs1693457}, \quad (\text{Model 4})$$

where SNP_N is SNP_1 – $\text{SNP}_{1,129}$.

The results were then ranked by increasing significance of the interaction p value ($\text{BMI} \times \text{rs1693457}$), and RTC scores were calculated as follows:

$$\text{RTC} = \frac{N_{\text{SNPs}} - \text{Rank}_{\text{GWAS_SNP}}}{N_{\text{SNPs}}} \quad (\text{Equation 1})$$

where N_{SNPs} = number of SNPs tested.

$\text{Rank}_{\text{GWAS_SNP}}$ = the rank of the GWAS SNP in the full list of ordered test statistics.

Tissue Specificity and eQTL π_1 Analysis

To determine the tissue specificity of identified $G \times \text{BMI}$ interactions and whether they are enriched for main-effect eQTLs, π_1 analysis was performed.³⁰ π_0 is a measure for estimating the proportion of true null hypotheses. $1 - \pi_0 = \pi_1$ can therefore be used to measure the proportion of significant (true) associations. The significant exon-SNP pairs from adipose tissue were matched to the same exon-SNP pair in each of the

other tissues, and a π_1 value was estimated with the R package “qvalue”³⁰. The same method was used to determine whether the significant $G \times \text{BMI}$ interactions are enriched for main-effect eQTLs.

Gene Set Enrichment

Both *cis* and *trans* gene set enrichment analysis data were analyzed through the use of QIAGEN’s Ingenuity Pathway Analysis (IPA). Input consists of the gene list of interest, and to correctly control for background genes, we used all genes expressed in our adipose tissue expression datasets as a reference to compare against. Benjamini-Hochberg corrected p values were calculated as specified by Ingenuity. The macrophage-enriched metabolic network (MEMN) gene membership was tested for $G \times \text{BMI}$ enrichment via π_1 analysis, as above.

Trans-Network Mediation Analysis

Significant mediation was determined by computing Sobel’s test statistic.³⁸ To calculate the mediation score, the following four models and equations were implemented:

$$y \sim \beta_1 A + \beta_2 A^2 + \beta_3 P + \beta_4 G + \beta_5 P \times G + \epsilon, \quad (\text{Model 5})$$

$$y \sim \beta_1 E + \beta_2 A + \beta_3 A^2 + \beta_4 P + \beta_5 G + \beta_6 P \times G + \epsilon, \quad (\text{Model 6})$$

$$\text{Mediation_score} = \frac{\beta_{M5_5} - \beta_{M6_6}}{\beta_{M5_5}}, \quad (\text{Equation 2})$$

and

$$Z = \frac{\beta_{M5_5} \times \beta_{M6_6}}{\sqrt{\beta_{M6_6}^2 \times S_{M5_5}^2 + \beta_{M5_5}^2 \times S_{M6_6}^2}}, \quad (\text{Equation 3})$$

where y = *trans* gene expression, A = age, A^2 = age squared, E = *cis* gene expression, P = BMI, and G = *cis* genotype. $M5_5$ is the interaction coefficient from Model 5. $M6_6$ is the interaction coefficient from Model 6. S^2 is the SE of each interaction coefficient.

By conditioning on *cis* gene expression (the mediator, E) we can determine whether each individual interaction detected in *trans* is regulated in *cis* or is independent by quantifying $\Delta\beta P \times G$ (Equation 2). We can test the significance of this change by using Sobel’s test statistic (Equation 3).

Roadmap Epigenomics Functional Element Enrichment Analysis

To investigate enrichment of cell-type-specific enhancers, HaploReg v.4 was utilized.³⁹ HaploReg takes a list of SNPs and uses a binomial test for enhancer enrichment in different cell types based on the SNP of interest ($G \times \text{BMI}$ SNP) and any SNPs that are in high LD with the lead SNP ($r^2 > 0.8$). These SNPs are then compared to a background that consists of the common variants (5%) obtained from the 1000 Genomes Project. Enhancer annotation was obtained from the ChromHMM 15-state model.

Interactive Website

The website was designed with RMarkdown, custom CSS, and a shiny backend server for real-time statistical analysis. All models implemented are those described above, with additional options given to users to allow the choice of covariate inclusion and/or PEER correction.

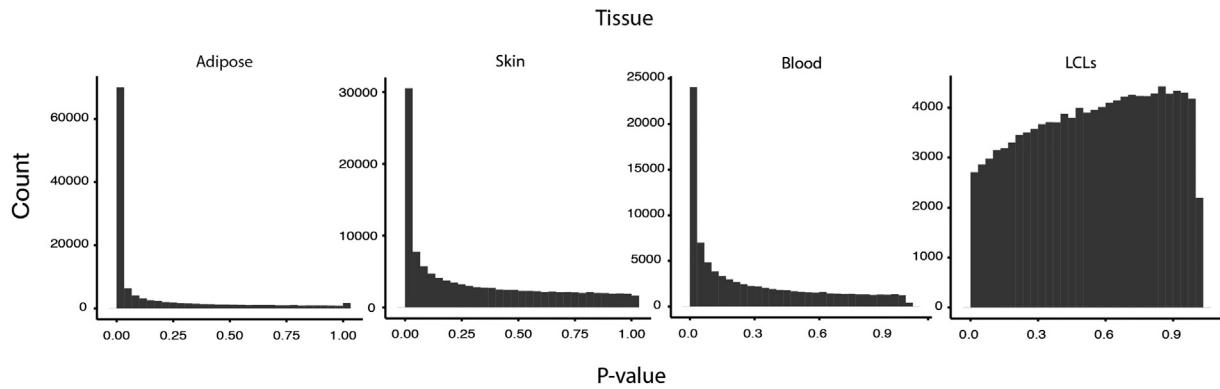


Figure 1. BMI Is Highly Associated with Gene Expression in Multiple Tissues
p value distribution of association between BMI and expression of all measured exons in each tissue.

Results

BMI Has a Pervasive Effect on Gene Expression

To characterize the effect of BMI on the transcriptome, we utilized a multi-tissue RNA-seq dataset including data from 856 healthy female twins in the TwinsUK cohort.²² Participants were 38–84 years old (median = 60) with a BMI range of 16–47 (median = 25) at the time of biopsy (Figure S3). As described previously,²² after quality control, RNA-seq data and imputed genotypes from the 1000 Genomes phase 1 reference panel were available for 720 subcutaneous adipose tissue samples, 672 skin samples, 765 LCLs, and 368 whole-blood samples. All analysis was based on exon-level quantifications with 16,149 to 18,229 genes expressed per tissue (Table S2).

We first assessed the extent of association between BMI and gene expression in all four tissues (adipose, skin, blood, and LCLs). BMI has a strong influence on expression in each of the primary tissues, but little to no effect in LCLs; with a FDR of 5%, 16,818 genes, 9,216 genes, 6,640 genes, and zero genes had at least one exon associated with BMI in adipose, skin, blood, and LCLs respectively (Figure 1). We assessed the tissue specificity of the associations by using π_1 estimates (Figure S4). Approximately half of the associations detected in adipose were observed in the other two primary tissues (skin $\pi_1 = 0.53$, blood $\pi_1 = 0.54$), whereas adipose captured over 75% of the associations detected in skin ($\pi_1 = 0.78$) and blood ($\pi_1 = 0.76$). There was no enrichment for shared effects in LCLs ($\pi_1 = 0$ for all tissues), which is consistent with a lack of LCL expression and whole-body trait associations seen for multiple traits,²¹ suggesting the transformation process or subsequent cell culture has removed the in vivo physiological environmental effects captured by profiling primary tissue. Our estimates of a pervasive and tissue-specific influence of BMI on expression are consistent with previous estimates from microarrays⁴⁰ and confirm that BMI acts as a strong physiological influence on gene expression. Full summary statistics for the association between BMI and exon-level expression in all four tissues can be found at our website (see Web Resources).

Identification of *cis* G × BMI Regulatory Variants

To identify BMI-dependent regulatory effects, we performed a global *cis* scan for G × BMI eQTLs in each tissue by using PEER-corrected expression residuals. Significant G × BMI effects (FDR = 5%) were called with a per-tissue FDR, determined by permutation (see Material and Methods). Our FDR method ensures we penalize against genes with more exons given that, statistically, they are more likely to show association by chance. We identified 16 G × BMI regulatory effects in adipose (Table 1, Table S3, Figure 2) and none in the other three tissues. The 16 genes regulated by a G × BMI effect in adipose are *ADH1A*, *ANXA5*, *CAST*, *CHURC1*, *CIDEA*, *ERV3-1*, *HLA-DQB2*, *IFNAR1*, *PEPD*, *PHACTR3*, *SCFD2*, *SPAG17*, and *ZNF117*. 12 of the 16 G × BMI variants also had a significant main effect (eQTL) on adipose expression when tested without the interaction term (Table 1, Table S3).

We sought replication in an independent dataset of subcutaneous adipose tissue biopsies (Icelandic cohort, deCODE genetics, $n = 754$), which included quantification of 13 of the 16 genes. Because TwinsUK and deCODE used different alignment strategies, genome builds, and gene annotation versions, some exons did not directly map between the two datasets. We identified corresponding exons with a greater than 90% overlap in annotation in both datasets for nine genes (Table S1). Of these nine, three were replicated and eight showed a consistent direction of effect between the studies (Figure 3A). The replicated genes were *PEPD* ($p_{\text{TwinsUK}} = 4.8 \times 10^{-10}$, $p_{\text{deCODE}} = 4.2 \times 10^{-6}$), *PHACTR3* ($p_{\text{TwinsUK}} = 1.6 \times 10^{-8}$, $p_{\text{deCODE}} = 1.1 \times 10^{-4}$), and *CHURC1* ($p_{\text{TwinsUK}} = 2.0 \times 10^{-12}$, $p_{\text{deCODE}} = 8.5 \times 10^{-4}$) (Table 1). Another four genes showed partial overlap with a deCODE exon: *ERV3-1* (85% overlap), *SPAG17* (69% overlap), *ADH1A* (33% overlap), and *CAST* (4% overlap). None of these replicated and only *CAST* had a consistent direction of effect, with $p = 0.053$. Given the low annotation overlap at *CAST* (4%), we examined all other *CAST* exons in deCODE, and the exon in chr5: 96,076,448–96,076,487 was associated in a consistent direction at nominal significance ($p = 0.001$, $p = 0.03$).

Table 1. Significant *cis* G × BMI Regulatory Interactions in Adipose, FDR = 5%

Gene	SNP	EA	EAF	TwinsUK β	TwinsUK p Value	deCODE β	deCODE p Value	Main-effect eQTL (FDR = 1%)	Enhancer
<i>CHURC1</i>	rs7143432	A	0.78	0.026	2.0×10^{-12}	0.011	8.5×10^{-4}	adipose, skin, blood, LCLs	–
<i>CAST</i>	rs13160562	G	0.69	–0.032	3.9×10^{-12}	–0.004	0.053 ^a	adipose, skin, blood, LCLs	adipocyte
<i>CIDEA</i>	rs7505859	C	0.62	–0.028	3.1×10^{-11}	–0.004	0.21	adipose, skin	adipocyte
<i>ZNF117</i>	rs6948760	T	0.40	0.039	4.4×10^{-11}	0.0009	0.79	adipose, skin, blood, LCLs	–
<i>ADH1A</i>	rs1693457	C	0.18	0.034	5.9×10^{-11}	–0.0015	0.85 ^a	adipose	adipocyte
<i>RP11-71E19.1</i>	rs1980140	A	0.79	–0.058	6.1×10^{-11}	NA	NA	adipose	adipocyte
<i>PEPD</i>	rs10415555	A	0.81	–0.044	4.8×10^{-10}	–0.014	4.2×10^{-6}	adipose, skin	adipocyte
<i>ANXA5</i>	rs2306420	G	0.71	0.022	1.4×10^{-9}	0.001	0.52	adipose, skin, blood, LCLs	adipocyte
<i>SIK1</i>	rs12482956	A	0.71	0.058	3.0×10^{-9}	NA	NA	–	blood
<i>HLA-DQB2</i>	rs114370295	T	0.27	–0.050	3.5×10^{-9}	–0.004	0.45	adipose	–
<i>ERV3-1</i>	rs11979998	C	0.52	0.032	8.4×10^{-9}	–0.0008	0.84 ^a	adipose, skin, blood, LCLs	blood
<i>POU6F2</i>	rs34792397	G	0.75	–0.041	9.9×10^{-9}	NA	NA	adipose	–
<i>IFNAR1</i>	rs2834098	C	0.78	–0.047	1.4×10^{-8}	0.002	0.58	–	stem cells
<i>SCFD2</i>	rs7687982	A	0.75	–0.059	1.5×10^{-8}	–0.006	0.26	–	aMSC
<i>PHACTR3</i>	rs6070866	G	0.51	–0.044	1.7×10^{-8}	–0.020	1.1×10^{-4}	adipose	brain
<i>SPAG17</i>	rs9661038	G	0.64	0.043	2.8×10^{-8}	–0.004	0.083 ^a	–	–

Exact exon coordinates are listed in Tables S1 and S3. The main-effect column lists the TwinsUK tissues where the listed gene-SNP pair has a main-effect *cis* eQTL when tested without the interaction effect (a dash denotes no eQTL in any tissue). All interaction effects were adipose specific. If a SNP or its proxy ($r^2 > 0.8$) falls within an enhancer as defined by ENCODE, the cell type that enhancer is primarily active in is listed in the Enhancer column. EA, effect allele; EAF, effect-allele frequency in the TwinsUK discovery sample.

^adeCODE exons with less than 90% overlap with the corresponding deCODE exon (see methods, Table S1).

corrected for 30 *CAST* exons). Overall, the replication demonstrates the robustness of our findings and the discovery and independent replication of regulatory variants whose effects are influenced by adiposity.

Tissue Specificity of G × BMI Regulatory Effects and Expression

The G × BMI regulatory effects are highly tissue specific. None of the 16 G × BMI interaction effects are present in the other TwinsUK tissues (Table S3). Although 11 G × BMI regulated genes showed multi-tissue expression, five genes (*PHACTR3*, *ADH1A*, *RP11-71E19.1*, *POU6F2*, and *SPAG17*) have adipose-specific expression within TwinsUK but no detectable expression in blood, skin, or LCLs. The G × BMI variants are enriched for tissue-specific regulatory potential; we tested the 16 variants for enrichment in enhancer activity in the 127 Roadmap Epigenomics cell types and found an enrichment only in purified adipocytes (mesenchymal-stem-cell-derived adipocytes, $p = 0.028$). Nine G × BMI variants disrupt specific or multiple transcription factor binding motifs, and eight SNPs or their proxies ($r^2 > 0.8$) have a significant evolutionary GERP conservation score (Table S3). The tissue specificity of G × BMI effects, expression of the genes they act on, and regulatory annotation of the G × BMI variants highlights the importance of studying gene regulation in the appropriate tissue for a disease of interest.

G × BMI Regulatory Variants and Genes Link to Related Traits

We next investigated links between the G × BMI genes and variants and related traits. *CAST* (calpastatin) is a ubiquitously expressed endogenous inhibitor of the calpains (calcium-dependent cysteine proteases). The calpastatin-calpain system has been extensively implicated in cardiac remodeling and heart failure and has been proposed as a potential therapeutic target for heart disease.⁴¹ *CAST* is implicated in modulating the immune response in certain cell types,⁴² including inhibition of macrophage hyperactivation under inflammatory conditions,⁴³ which could be linked to the varying degrees of systemic inflammation seen in obese individuals.

Another intriguing signal is at *PEPD* (peptidase D). *PEPD*'s function is the recycling of proline, and it has also been shown to be essential for collagen production. Variants intronic to *PEPD* are associated with T2D, adiponectin, triglyceride levels, and fasting insulin;^{44–47} however, the lead GWAS SNPs are in low LD with the G × BMI regulatory variant ($r^2 = 0.145$, $D' = 1$). *PEPD*'s link to T2D and adiponectin, although not fully understood, is interesting in the light of BMI's relationship with T2D development and adiponectin regulation.

CIDEA (cell death activator) function has been studied extensively in model organisms, and *CIDEA*-knockout mice show higher basal metabolic rates, lipolysis, and

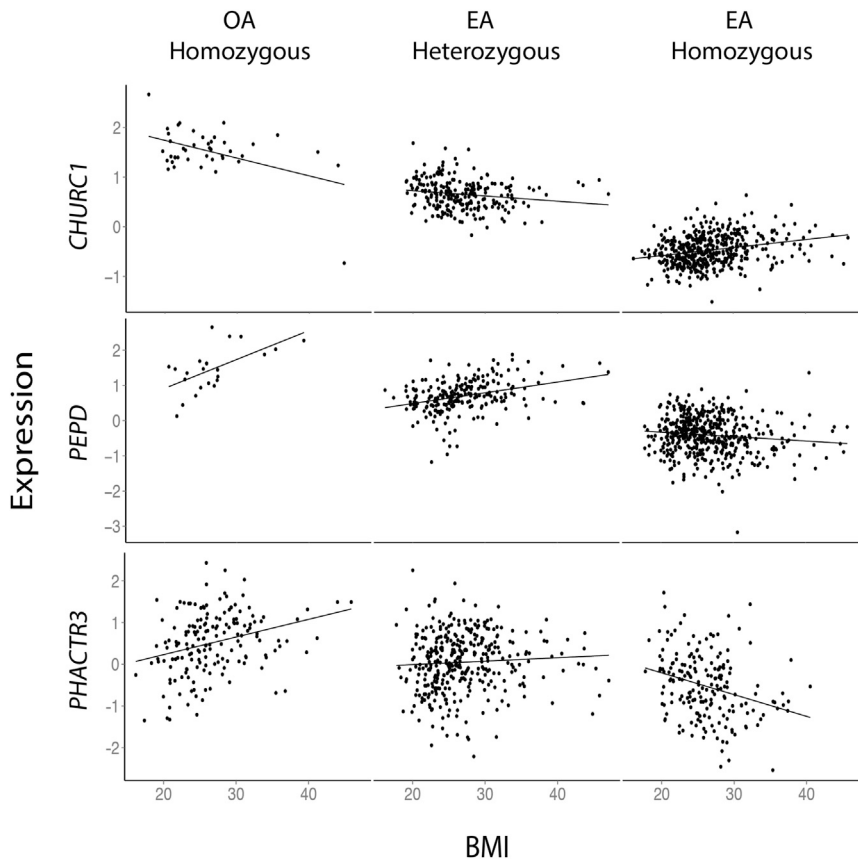


Figure 2. Example of Three FDR = 5% Significant Adipose *cis* G × BMI Regulatory Interactions

The vertical axis represents expression of a given gene and the horizontal axis represents BMI. Each point represents an individual. Each plot is split on allele dosage to show how the relationship between expression and BMI is dependent on genotype (OA, other allele; EA, effect allele; corresponding variants and alleles are listed in Table 1). A change in the slope of the association between BMI and expression association across genotypic classes indicates a genotype-dependent response to BMI, or G × BMI interaction. For example, *CHURC1* expression decreases with BMI in individuals homozygous for the effect allele, but increases with BMI in other allele homozygotes regardless of the mean change in gene expression (main-effect eQTL).

higher core body temperatures.⁴⁸ Mice without functional copies of *CIDEA* are resistant to both diet-induced obesity and T2D.^{49,50} Similar evidence has been observed in humans.⁵¹ In our data, *CIDEA* expression has a strong negative correlation with BMI (p value = 8.1×10^{-54}). However, individuals with the minor allele of rs7505859 show the opposite relationship (*CIDEA* is positively correlated with BMI). Again, this is an interesting relationship in the light of BMI's contribution to T2D development.

We integrated the 16 G × BMI variants with the NHGRI database of GWASs to determine whether the G × BMI regulatory effects are linked to common traits or disease. A proxy SNP (rs1229977, $r^2 = 0.63$) for the G × BMI variant regulating *ADH1A* (alcohol dehydrogenase 1A), rs1693457, is associated with esophageal cancer (MIM: 133239) in a large GWAS.⁵² The GWAS association at this locus covers several ADH family genes; however, we show that, in addition to the G × BMI effect, rs1693457 is a main-effect eQTL ($\beta = -0.89$, $p = 2.9 \times 10^{-41}$) for *ADH1A*. To formally test for colocalization between the regulatory variant and the GWAS signal, we implemented the regulatory trait concordance method (RTC).⁵³ The esophageal cancer GWAS SNP (rs1229977) had an RTC score of 0.98 when tested with either the interaction or main-effect model, indicating that the regulatory and GWAS signals are tagging the same underlying variant. There is prior evidence for a G × E interaction at rs1229977—the association to esophageal cancer is modi-

fied by alcohol consumption.⁵² Given the complicated links between BMI, alcohol intake, and smoking, it is possible that BMI is acting as a proxy for correlated external environmental factors at this locus, which could also explain the lack of replication of the *ADH1A* G × BMI effect in the Icelandic data, given that external environments can vary between countries.

Detection of *trans* G × BMI Effects in Adipose Tissue

Evidence from model organisms suggests that *trans* (distal) regulatory variants are more likely to mediate the effects of the environment and are thus strong candidates for G × E interaction effects.¹⁰ In contrast to *cis* eQTLs, *trans* eQTLs have been difficult to identify in humans because of smaller effect sizes and the increased burden of multiple testing in a genome-wide scan. Given that *cis* eQTLs are enriched for *trans*-eQTL effects,⁵⁴ we utilized a two-step strategy to identify *trans* G × BMI interactions in adipose tissue, first identifying *cis* G × BMI interactions and then testing the identified *cis* G × BMI variants for *trans* G × BMI effects with all adipose-expressed exons. For *trans* discovery, both steps were performed with non-PEER-corrected expression residuals because correction for latent factors can remove broadly acting *trans* effects. This strategy identified four *cis* G × BMI variants (Table S4), one of which, rs3851570 (*cis* G × BMI interaction on *ALG9* expression [MIM: 606941], p value = 2.0×10^{-8}), is associated with 53 genes in *trans* (Bonferroni corrected p value threshold of $p < 1.1 \times 10^{-7}$) (Figure 4A, Table S5, Figure S5).

IPA of the 53 genes in the *ALG9* G × BMI *trans*-network revealed enrichment for inhibition of matrix metalloproteases (Benjamini-Hochberg corrected $p = 3.6 \times 10^{-8}$), oxidative phosphorylation (Benjamini-Hochberg corrected

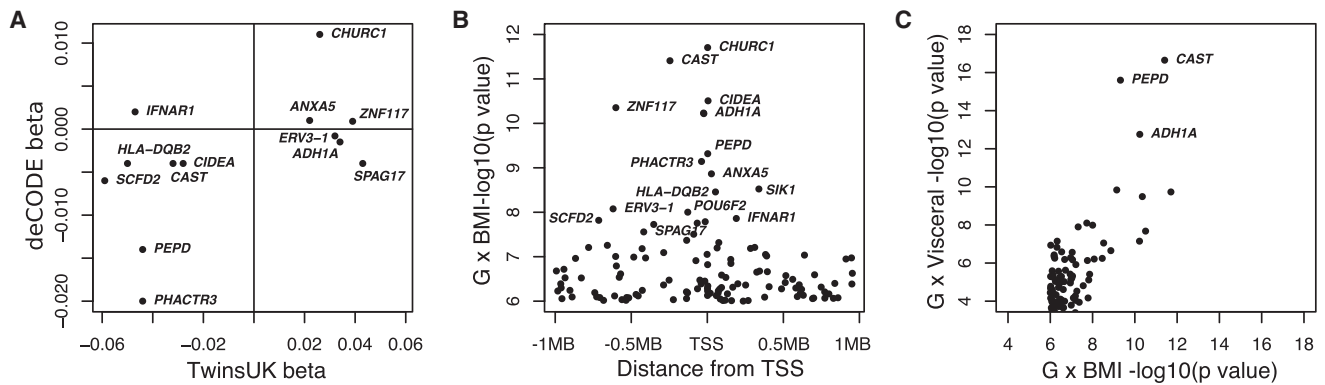


Figure 3. Replication, Location, and Validation of *cis* G × BMI Interactions

(A) Comparison of G × BMI interaction coefficients (FDR = 5%) in the TwinsUK discovery (horizontal axis) and deCODE replication dataset (vertical axis). 9 out of 13 interactions show a consistent direction of effect across cohorts.

(B) The lead G × BMI SNPs all cluster at the TSS. Location of the lead SNPs for the 127 G × BMI interactions with $p < 10^{-6}$ are plotted with respect to the TSS of the corresponding gene. Lead SNPs for the 26 significant G × BMI interactions (FDR = 5%) are labeled with the corresponding gene name.

(C) Comparison of p values of G × BMI interaction and G × visceral-fat interaction for the same SNP gene pairs. All 127 interactions with G × BMI $p < 10^{-6}$ are plotted. All 16 G × BMI (FDR = 5%) show a G × visceral-fat interaction, three of which increased 2–5 orders of magnitude in despite the smaller sample size in the visceral fat analysis ($n = 682$).

$p = 3.1 \times 10^{-4}$), and gene membership of a cardiovascular disease network (Figure 4B), indicating that BMI-dependent regulatory effects at rs3851570 have a wide-ranging role in metabolism and structural remodeling of adipose tissue. *ALG9* itself catalyzes lipid-linked oligosaccharide assembly in the N-glycan biosynthesis pathway. Statistical mediation analysis supports a mediator role of *ALG9* expression in regulation of the *trans* G × BMI network (Sobel’s mediation p value ≤ 0.001 , Table S6) (see Material and Methods). Given that *ALG9* has no known regulatory role on transcription, the regulation of the *trans* genes most likely functions via regulation of signaling cascades or other complex indirect processes. We note that the most significant *trans* G × BMI gene is the transcription factor *ZNF423* ($p = 8.2 \times 10^{-13}$) (Figure 4C). *ZNF423* regulates pre-adipocyte determination and expression of *PPARG* (MIM: 601487), a master regulator of adipocyte differentiation,⁵⁵ and could be a candidate for mediating the widespread *trans* effect, although this would require further investigation. We investigated the *trans* G × BMI effect in a second population by testing rs3851570 against all measured exon expression amounts ($n = 168,951$) in the deCODE dataset. Although the 53 *trans* genes did not replicate in the deCODE dataset, transcriptome-wide we see a significant enrichment for low p values ($\pi_1 = 0.14$), demonstrating that rs3851570 has broad effects on adipose gene expression in a BMI-dependent manner in multiple populations. This is consistent with previous studies, wherein transcriptome-wide regulatory behavior of the multi-gene *trans* variants replicate across cohorts but the top genes associated with such *trans* networks do not.²³

General Properties of *cis* G × BMI Regulatory Effects and Variants

To investigate the common properties of G × BMI regulatory effects and the genes they act upon, we next focused

on the 127 adipose G × BMI *cis* effects that passed a relaxed threshold of $p < 1.0 \times 10^{-6}$ (Table S7). All 127 G × BMI interactions show an opposite direction of effect between expression and BMI in the two-homozygote classes (for example, *PHACTR3*, Figure 2). This high prevalence of opposing direction of effect most likely reflects our power to detect opposing effects rather than the true distribution. Consistent with main-effect eQTL variants, G × BMI variants are enriched for proximity to the transcription start site, most noticeably when looking at genome-wide significant G × BMI variants (median distance, 38 kb) (Figure 3B). The 127 G × BMI variants show significant adipose main effects when tested without an interaction with BMI (number of eQTLs (FDR = 1%): adipose = 20, blood = 8, LCLs = 10, and skin = 10) (Figure S6). Whereas main-effect eQTL enrichment was observed in other tissues, the G × BMI effect is adipose-tissue specific, with $\pi_1 = 0.038$ in skin and $\pi_1 = 0$ in blood and LCLs. G × BMI variants are not directly associated with BMI, and no significant enrichment was seen in a BMI GWAS of 339,224 individuals⁶ ($\pi_1 = 0$), suggesting the G × BMI effects are not the result of indirectly measured gene-by-gene effects.

G × BMI Are Enriched for Key Metabolic Processes

To elucidate the biological consequences of BMI-dependent regulation, we investigated the 127 genes regulated by a G × BMI interaction at $p < 1.0 \times 10^{-6}$ for functional enrichment by using IPA. The 127 genes are enriched for key metabolic processes, including LXL and RXR activation ($p = 6.1 \times 10^{-3}$) and uptake of cholesterol ($p = 9.1 \times 10^{-5}$) (Table S8). *LXL* and *RXR* are part of the insulin-signaling pathway and are the target of several widely used T2D medications.⁵⁶ Additionally, the 127 genes are enriched for the antigen-presentation pathway ($p = 2.0 \times 10^{-4}$) and quantity of macrophages ($p = 5.8 \times 10^{-3}$). We do not see evidence for

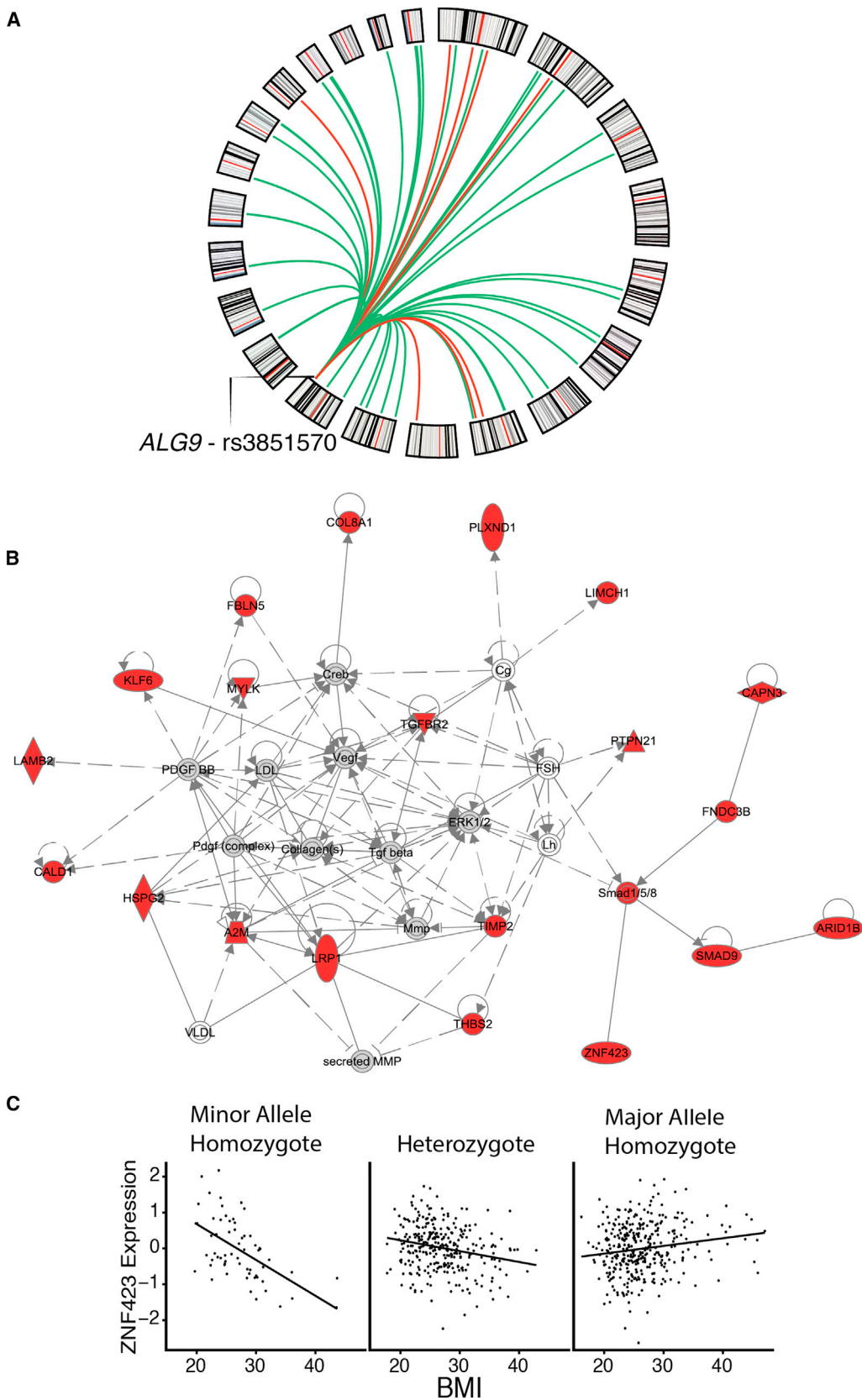


Figure 4. rs3851570 Has a *trans* G × BMI Effect on 53 Genes

(A) Circos plot displaying the location of rs3851570 and *trans* G × BMI associations. Each line represents a *trans* association originating from rs3851570. Green lines indicate a positive beta, and red lines represent a negative beta. The outer circle delineates chromosomes, including cytogenetic bands.

(legend continued on next page)

G × BMI effects in the BMI-associated MEMN of co-expressed adipose genes.^{40,57} The 553 MEMN genes show no enrichment for G × BMI effects ($\pi_1 = 0.021$), and there is not a significant overlap between the 127 G × BMI genes and MEMN genes (binomial $p = 0.78$). However, we do note that of the four G × BMI genes in the MEMN (*ICAM3* [MIM: 146631], *TBX3* [MIM: 601621], *CIDEA*, and *ADH1A*), two do have genome-wide-significant G × BMI effects (*CIDEA* and *ADH1A*).

BMI Accurately Captures the Effects of Adiposity

BMI is an easily measured anthropometric trait that is commonly used as a surrogate for overall adiposity. It has been noted that BMI can potentially misclassify an individual as obese, for example, when a subject has a large lean-muscle mass (typical of athletes). Additionally, BMI measurements do not capture differences in body-fat distribution. Fat accumulation in the abdominal region, particularly of visceral fat, is predictive of multiple adverse health outcomes independent of overall BMI.⁵⁸ We thus sought to validate our G × BMI findings by performing the same analysis on a subset of the TwinsUK cohort that had DXA-measured abdominal visceral fat volume. Within this subset ($n = 682$), visceral fat ranges from 78 to 1,542 g (mean = 627 g) and, as expected, broadly correlates with BMI (Figure S7). All 16 G × BMI interactions (FDR = 5%) showed a similar effect in the G × visceral fat analysis, and despite the drop in sample size, some interactions increased in significance by four orders of magnitude (*CAST* $p = 3.0 \times 10^{-16}$, *PEPD* $p = 3.0 \times 10^{-15}$) (Figure 3C), potentially due to the increased sensitivity gained from an accurate machine-measured phenotype. This provides evidence that G × BMI can capture real adiposity effects and suggests that, although more difficult to obtain, power to detect interactions increases dramatically with phenotypic measurement accuracy, such as in measurements obtained by DXA.

Discussion

Here, we describe the pervasive effect BMI has on peripheral-tissue gene expression and identify robust examples of BMI-dependent regulatory variants. We characterize the properties of G × BMI regulatory variants, showing that they typically have strong main effects and are highly tissue specific. Additionally, we identify a G × BMI interaction that regulates an adipose-specific *trans*-network of 53 genes. Although twin and family studies estimate that ~60% of

expression heritability acts in *trans*,^{23,24} *trans* effects have been difficult to identify in humans. Given the high context specificity of *trans* effects in model organisms,¹⁰ *trans* eQTL discovery efforts in humans might be more fruitful when accounting for environmental interactions.

These findings suggest that identifying G × E interactions on gene expression is possible with significantly smaller sample sizes, compared to complex traits. However, we note that the current study is underpowered, and we expect more G × BMI interactions to be identified with larger sample sizes. Extrapolating from empirical investigation of the power to detect cell-type-specific interactions in whole-blood transcriptomes,²⁰ we expect that power to detect G × BMI effects on expression should scale linearly with sample size. Increasing sample size will most likely necessitate combining data across populations, which can confound G × E analysis if the environment differs across study populations. Using a physiological variable such as BMI should mitigate this caveat, however, we acknowledge that BMI could also be acting as a surrogate for a highly correlated cofactor, such as diet, which might vary across populations and thus complicate replication or meta-analysis efforts. Body-fat traits have been shown to be sexually dimorphic. It is therefore important to point out that although we demonstrated replicated examples of G × BMI effects in a separate adipose tissue cohort, the discovery sample was an all-female cohort, whereas 44% of the replication cohort were men, potentially reducing our replicative power.

It is well documented that increasing BMI induces changes in the cell-type composition and inflammation of adipose tissue.^{59–63} Changes in cell-type composition could be the underlying mechanism of some of our identified G × BMI interactions. This is an intriguing possibility, and we are actively investigating methods to deconvolve cell-type expression data from whole-tissue expression profiles to address this. Enrichment for G × BMI genes in immune response pathways could also represent a change in the activation state of cells already present in adipose tissue or in the genetic control of inflammation susceptibility that takes place under weight gain, during which some individuals, due to their genotype, could undergo less inflammation in an obese state. Given the enrichment for genes involved in metabolic processes, other potential mechanisms include BMI-driven changes in metabolism,⁶⁴ adipocyte size⁶⁵ (a consequence of having to store more lipid droplets in an obese state),⁶⁶ increased vascularization as a result of hypoxia,⁶⁷ and changes in overall energy expenditure.⁶⁸

(B) *Trans* genes are enriched for membership in a cardiovascular disease network (IPA). Plot displays all genes in the cardiovascular disease network. Genes highlighted in red are present in the *trans*-network, and non-*trans* network genes are shown in gray. Lines between genes represent their regulatory relationship. Solid lines represent direct evidence of interaction, while dashed lines indicate indirect interaction evidence. The shape of each node denotes the gene's primary known function (protein kinase, transcription factor, complex, etc).

(C) A signal plot of the most significant G × BMI *trans*-association to rs3851570, at the gene *ZNF423*. The vertical axis represents expression of *ZNF423* and the horizontal-axis is BMI. Each point represents an individual; individuals are grouped into the three boxes according to their allele dosage for SNP rs3851570.

Discovery of $G \times \text{BMI}$ effects on gene expression has broad applicability. Characterization of $G \times \text{BMI}$ regulatory variants and the genes they act upon will help elucidate the downstream consequences of obesity and the molecular pathways leading to associated diseases. Similar to the widespread use of main-effect eQTLs to interpret GWAS loci, context-specific eQTLs can be used both to identify the molecular mechanism of GWAS loci and to identify relevant, interacting environments or risk factors. To facilitate this, we have made the full results of this study freely available in an interactive web service. Finally, identifying individuals whose genotypes predispose them to BMI-specific outcomes could enable targeted interventions for those most likely to respond and could improve accuracy in assessing genotypic risk of obesity-related diseases.

Accession Numbers

The accession number for the RNA-seq data reported in the paper is EGA: EGAS00001000805.

Supplemental Data

Supplemental Data include seven figures and eight tables and can be found with this article online at <http://dx.doi.org/10.1016/j.ajhg.2016.07.001>.

Acknowledgments

The authors would like to thank Julia El-Sayed Moustafa and Alexander Da Silva Couto Alves for feedback on the manuscript as well as on the website. We thank the twins for their contribution to this project. K.S.S. is supported by the Medical Research Council (grants L01999X/1 and MR/J019642/1). The RNA-seq data generation was funded by the European Union 7th Framework Programme grant EuroBATS (259749), which also supported A.V., A.B., T.D.S., and E.T.D. Biopsy collection was funded by a Wellcome Trust program grant (081917/Z/07/Z). T.D.S. is a National Institute for Health Research (NIHR) senior investigator. The TwinsUK study was funded by the Wellcome Trust and European Community's 7th Framework Programme (FP7/2007-2013). The study also receives support from the NIHR-funded BioResource Clinical Research Facility and Biomedical Research Centre based at Guy's and St. Thomas' NHS Foundation Trust in partnership with King's College London. SNP genotyping was performed by the Wellcome Trust Sanger Institute and National Eye Institute via the NIH and Center for Inherited Disease Research. G.H.H., G.T., H.H., U.T., and K.S. are employees of deCODE Genetics (Amgen).

Received: February 17, 2016

Accepted: July 1, 2016

Published: September 1, 2016

Web Resources

European Genome-phenome Archive (EGA), <https://www.ebi.ac.uk/ega>

OMIM, <http://www.omim.org/>

PhenoExpress, BMI transcriptome-wide association study and $G \times \text{BMI}$ visualization server, <http://expression.kcl.ac.uk>

RefSeq, <http://www.ncbi.nlm.nih.gov/RefSeq>

UCSC Genome Browser, <http://genome.ucsc.edu>

References

1. Fall, T., Hägg, S., Mägi, R., Ploner, A., Fischer, K., Horikoshi, M., Sarin, A.P., Thorleifsson, G., Ladenvall, C., Kals, M., et al.; European Network for Genetic and Genomic Epidemiology (ENGAGE) consortium (2013). The role of adiposity in cardio-metabolic traits: a Mendelian randomization analysis. *PLoS Med.* 10, e1001474.
2. Haque, R., Van Den Eeden, S.K., Wallner, L.P., Richert-Boe, K., Kallakury, B., Wang, R., and Weinmann, S. (2014). Association of body mass index and prostate cancer mortality. *Obes. Res. Clin. Pract.* 8, e374–e381.
3. McLaughlin, T. (2012). Metabolic heterogeneity of obesity: role of adipose tissue. *Int. J. Obes. Suppl.* 2 (Suppl 1), S8–S10.
4. Simino, J., Shi, G., Weder, A., Boerwinkle, E., Hunt, S.C., and Rao, D.C. (2014). Body mass index modulates blood pressure heritability: the Family Blood Pressure Program. *Am. J. Hypertens.* 27, 610–619.
5. Wang, X., Ding, X., Su, S., Spector, T.D., Mangino, M., Iliadou, A., and Snieder, H. (2009). Heritability of insulin sensitivity and lipid profile depend on BMI: evidence for gene-obesity interaction. *Diabetologia* 52, 2578–2584.
6. Locke, A.E., Kahali, B., Berndt, S.I., Justice, A.E., Pers, T.H., Day, F.R., Powell, C., Vedantam, S., Buchkovich, M.L., Yang, J., et al.; LifeLines Cohort Study; ADIPOGen Consortium; AGEN-BMI Working Group; CARDIOGRAMplusC4D Consortium; CKDGen Consortium; GLGC; ICBP; MAGIC Investigators; MuTHER Consortium; MiGen Consortium; PAGE Consortium; ReproGen Consortium; GENIE Consortium; International Endogene Consortium (2015). Genetic studies of body mass index yield new insights for obesity biology. *Nature* 518, 197–206.
7. Gaulton, K.J., Ferreira, T., Lee, Y., Raimondo, A., Mägi, R., Reschen, M.E., Mahajan, A., Locke, A., Rayner, N.W., Robertson, N., et al.; DIAbetes Genetics Replication And Meta-analysis (DIAGRAM) Consortium (2015). Genetic fine mapping and genomic annotation defines causal mechanisms at type 2 diabetes susceptibility loci. *Nat. Genet.* 47, 1415–1425.
8. Kapoor, A., Sekar, R.B., Hansen, N.F., Fox-Talbot, K., Morley, M., Pihur, V., Chatterjee, S., Brandimarto, J., Moravec, C.S., Putil, S.L., et al.; QT Interval-International GWAS Consortium (2014). An enhancer polymorphism at the cardiomyocyte intercalated disc protein NOS1AP locus is a major regulator of the QT interval. *Am. J. Hum. Genet.* 94, 854–869.
9. Shungin, D., Winkler, T.W., Croteau-Chonka, D.C., Ferreira, T., Locke, A.E., Mägi, R., Strawbridge, R.J., Pers, T.H., Fischer, K., Justice, A.E., et al.; ADIPOGen Consortium; CARDIOGRAMplusC4D Consortium; CKDGen Consortium; GEFOS Consortium; GENIE Consortium; GLGC; ICBP; International Endogene Consortium; LifeLines Cohort Study; MAGIC Investigators; MuTHER Consortium; PAGE Consortium; ReproGen Consortium (2015). New genetic loci link adipose and insulin biology to body fat distribution. *Nature* 518, 187–196.
10. Smith, E.N., and Kruglyak, L. (2008). Gene-environment interaction in yeast gene expression. *PLoS Biol.* 6, e83.
11. Brem, R.B., Yvert, G., Clinton, R., and Kruglyak, L. (2002). Genetic dissection of transcriptional regulation in budding yeast. *Science* 296, 752–755.

12. Eng, K.H., Kvitek, D.J., Keles, S., and Gasch, A.P. (2010). Transient genotype-by-environment interactions following environmental shock provide a source of expression variation for essential genes. *Genetics* 184, 587–593.
13. Grishkevich, V., Ben-Elazar, S., Hashimshony, T., Schott, D.H., Hunter, C.P., and Yanai, I. (2012). A genomic bias for genotype-environment interactions in *C. elegans*. *Mol. Syst. Biol.* 8, 587.
14. Yao, C., Joehanes, R., Johnson, A.D., Huan, T., Esko, T., Ying, S., Freedman, J.E., Murabito, J., Lunetta, K.L., Metspalu, A., et al. (2014). Sex- and age-interacting eQTLs in human complex diseases. *Hum. Mol. Genet.* 23, 1947–1956.
15. Franco, L.M., Bucayas, K.L., Wells, J.M., Niño, D., Wang, X., Zapata, G.E., Arden, N., Renwick, A., Yu, P., Quarles, J.M., et al. (2013). Integrative genomic analysis of the human immune response to influenza vaccination. *eLife* 2, e00299.
16. Fairfax, B.P., Humburg, P., Makino, S., Naranbhai, V., Wong, D., Lau, E., Jostins, L., Plant, K., Andrews, R., McGee, C., and Knight, J.C. (2014). Innate immune activity conditions the effect of regulatory variants upon monocyte gene expression. *Science* 343, 1246949.
17. Maranville, J.C., Luca, F., Richards, A.L., Wen, X., Witonsky, D.B., Baxter, S., Stephens, M., and Di Rienzo, A. (2011). Interactions between glucocorticoid treatment and *cis*-regulatory polymorphisms contribute to cellular response phenotypes. *PLoS Genet.* 7, e1002162.
18. Mangravite, L.M., Engelhardt, B.E., Medina, M.W., Smith, J.D., Brown, C.D., Chasman, D.I., Mecham, B.H., Howie, B., Shim, H., Naidoo, D., et al. (2013). A statin-dependent QTL for GATM expression is associated with statin-induced myopathy. *Nature* 502, 377–380.
19. Francesconi, M., and Lehner, B. (2014). The effects of genetic variation on gene expression dynamics during development. *Nature* 505, 208–211.
20. Westra, H., Arends, D., Esko, T., Peters, M.J., Schurmann, C., Schramm, K., Kettunen, J., Yaghootkar, H., Fairfax, B.P., Andiappan, A.K., et al. (2015). Cell Specific eQTL Analysis without Sorting Cells. *Plos Genetics* 11, UNSP e1005223.
21. Brown, A.A., Buil, A., Viñuela, A., Lappalainen, T., Zheng, H.F., Richards, J.B., Small, K.S., Spector, T.D., Dermitzakis, E.T., and Durbin, R. (2014). Genetic interactions affecting human gene expression identified by variance association mapping. *eLife* 3, e01381.
22. Buil, A., Brown, A.A., Lappalainen, T., Viñuela, A., Davies, M.N., Zheng, H.F., Richards, J.B., Glass, D., Small, K.S., Durbin, R., et al. (2015). Gene-gene and gene-environment interactions detected by transcriptome sequence analysis in twins. *Nat. Genet.* 47, 88–91.
23. Grundberg, E., Small, K.S., Hedman, A.K., Nica, A.C., Buil, A., Keildson, S., Bell, J.T., Yang, T.P., Meduri, E., Barrett, A., et al.; Multiple Tissue Human Expression Resource (MuTHER) Consortium (2012). Mapping *cis*- and *trans*-regulatory effects across multiple tissues in twins. *Nat. Genet.* 44, 1084–1089.
24. Price, A.L., Helgason, A., Thorleifsson, G., McCarroll, S.A., Kong, A., and Stefansson, K. (2011). Single-tissue and cross-tissue heritability of gene expression via identity-by-descent in related or unrelated individuals. *PLoS Genet.* 7, e1001317.
25. Suhre, K., Shin, S.Y., Petersen, A.K., Mohny, R.P., Meredith, D., Wägele, B., Altmaier, E., Deloukas, P., Erdmann, J., Grundberg, E., et al.; CARDIOGRAM (2011). Human metabolic individuality in biomedical and pharmaceutical research. *Nature* 477, 54–60.
26. Wood, A.R., Esko, T., Yang, J., Vedantam, S., Pers, T.H., Gustafsson, S., Chu, A.Y., Estrada, K., Luan, J., Kutalik, Z., et al.; Electronic Medical Records and Genomics (eMEMERGE) Consortium; MIGen Consortium; PAGEGE Consortium; LifeLines Cohort Study (2014). Defining the role of common variation in the genomic and biological architecture of adult human height. *Nat. Genet.* 46, 1173–1186.
27. Zheng, H., Rong, J., Liu, M., Han, F., Zhang, X., Richards, J.B., and Wang, L. (2015). Performance of genotype imputation for low frequency and rare variants from the 1000 Genomes. *PLoS One* 10, e0116487.
28. Li, H., and Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* 25, 1754–1760.
29. Bates, D., Machler, M., Bolker, B., and Walker, S.C. (2014). Fitting linear mixed-effects models using lme4 (ArXiv).
30. Storey, J.D. (2002). A direct approach to false discovery rates. *J. R. Stat. Soc. Series B Stat. Methodol.* 64, 479–498.
31. Stegle, O., Parts, L., Durbin, R., and Winn, J. (2010). A Bayesian framework to account for complex non-genetic factors in gene expression levels greatly increases power in eQTL studies. *PLoS Comput. Biol.* 6, e1000770.
32. Shabalin, A.A. (2012). Matrix eQTL: ultra fast eQTL analysis via large matrix operations. *Bioinformatics* 28, 1353–1358.
33. Anderson, M.J. (2001). Permutation tests for univariate or multivariate analysis of variance and regression. *Can. J. Fish. Aquat. Sci.* 58, 626–639.
34. Gerrits, A., Li, Y., Tesson, B.M., Bystrykh, L.V., Weersing, E., Ausema, A., Dontje, B., Wang, X., Breitling, R., Jansen, R.C., and de Haan, G. (2009). Expression quantitative trait loci are highly sensitive to cellular differentiation state. *PLoS Genet.* 5, e1000692.
35. Gudbjartsson, D.F., Helgason, H., Gudjonsson, S.A., Zink, F., Oddson, A., Gylfason, A., Besenbacher, S., Magnusson, G., Halldorsson, B.V., Hjartarson, E., et al. (2015). Large-scale whole-genome sequencing of the Icelandic population. *Nat. Genet.* 47, 435–444.
36. Trapnell, C., Pachter, L., and Salzberg, S.L. (2009). TopHat: discovering splice junctions with RNA-Seq. *Bioinformatics* 25, 1105–1111.
37. DeLuca, D.S., Levin, J.Z., Sivachenko, A., Fennell, T., Nazaire, M.D., Williams, C., Reich, M., Winckler, W., and Getz, G. (2012). RNA-SeQC: RNA-seq metrics for quality control and process optimization. *Bioinformatics* 28, 1530–1532.
38. Pierce, B.L., Tong, L., Chen, L.S., Rahaman, R., Argos, M., Jasmine, F., Roy, S., Paul-Brutus, R., Westra, H.J., Franke, L., et al. (2014). Mediation analysis demonstrates that *trans*-eQTLs are often explained by *cis*-mediation: a genome-wide analysis among 1,800 South Asians. *PLoS Genet.* 10, e1004818.
39. Ward, L.D., and Kellis, M. (2012). HaploReg: a resource for exploring chromatin states, conservation, and regulatory motif alterations within sets of genetically linked variants. *Nucleic Acids Res.* 40, D930–D934.
40. Emilsson, V., Thorleifsson, G., Zhang, B., Leonardson, A.S., Zink, F., Zhu, J., Carlson, S., Helgason, A., Walters, G.B., Gunnarsdottir, S., et al. (2008). Genetics of gene expression and its effect on disease. *Nature* 452, 423–428.
41. Letavernier, E., Zafrani, L., Perez, J., Letavernier, B., Haymann, J.P., and Baud, L. (2012). The role of calpains in myocardial remodeling and heart failure. *Cardiovasc. Res.* 96, 38–45.
42. Rose, A., Zhi, H., Hoffmann, F., Norton, R., and Hoffmann, P. (2012). Selenoprotein K is a novel target of m-calpain, and

- cleavage is regulated by toll-like receptor-induced calpastatin in macrophages. *J. Immunol.* 188.
43. Huang, Z., Rose, A.H., Hoffmann, F.W., Hashimoto, A.S., Bertino, P., Denk, T., Takano, J., Iwata, N., Saido, T.C., and Hoffmann, P.R. (2013). Calpastatin prevents NF- κ B-mediated hyperactivation of macrophages and attenuates colitis. *J. Immunol.* 191, 3778–3788.
 44. Dastani, Z., Hivert, M.F., Timpson, N., Perry, J.R., Yuan, X., Scott, R.A., Henneman, P., Heid, I.M., Kizer, J.R., Lyytikäinen, L.P., et al.; DIAGRAM+ Consortium; MAGIC Consortium; GLGC Investigators; MuTHER Consortium; DIAGRAM Consortium; GIANT Consortium; Global B Pgen Consortium; Procardis Consortium; MAGIC investigators; GLGC Consortium (2012). Novel loci for adiponectin levels and their influence on type 2 diabetes and metabolic traits: a multi-ethnic meta-analysis of 45,891 individuals. *PLoS Genet.* 8, e1002607.
 45. Manning, A.K., Hivert, M.F., Scott, R.A., Grimsby, J.L., Bouatia-Naji, N., Chen, H., Rybin, D., Liu, C.T., Bielak, L.F., Prokopenko, I., et al.; DIABetes Genetics Replication And Meta-analysis (DIAGRAM) Consortium; Multiple Tissue Human Expression Resource (MUTHER) Consortium (2012). A genome-wide approach accounting for body mass index identifies genetic variants influencing fasting glycemic traits and insulin resistance. *Nat. Genet.* 44, 659–669.
 46. Willer, C.J., Schmidt, E.M., Sengupta, S., Peloso, G.M., Gustafsson, S., Kanoni, S., Ganna, A., Chen, J., Buchkovich, M.L., Mora, S., et al.; Global Lipids Genetics Consortium (2013). Discovery and refinement of loci associated with lipid levels. *Nat. Genet.* 45, 1274–1283.
 47. Cho, Y.S., Chen, C.H., Hu, C., Long, J., Ong, R.T.H., Sim, X., Takeuchi, F., Wu, Y., Go, M.J., Yamauchi, T., et al.; DIAGRAM Consortium; MuTHER Consortium (2011). Meta-analysis of genome-wide association studies identifies eight new loci for type 2 diabetes in east Asians. *Nat. Genet.* 44, 67–72.
 48. Zhou, Z., Yon Toh, S., Chen, Z., Guo, K., Ng, C.P., Ponniah, S., Lin, S.C., Hong, W., and Li, P. (2003). Cidea-deficient mice have lean phenotype and are resistant to obesity. *Nat. Genet.* 35, 49–56.
 49. Wu, L., Zhou, L., Chen, C., Gong, J., Xu, L., Ye, J., Li, D., and Li, P. (2014). Cidea controls lipid droplet fusion and lipid storage in brown and white adipose tissue. *Sci. China Life Sci.* 57, 107–116.
 50. Abreu-Vieira, G., Fischer, A.W., Mattsson, C., de Jong, J.M.A., Shabalina, I.G., Rydén, M., Laurencikiene, J., Arner, P., Cannon, B., Nedergaard, J., and Petrovic, N. (2015). Cidea improves the metabolic profile through expansion of adipose tissue. *Nat. Commun.* 6, 7433.
 51. Puri, V., Ranjit, S., Konda, S., Nicoloso, S.M.C., Straubhaar, J., Chawla, A., Chouinard, M., Lin, C., Burkart, A., Corvera, S., et al. (2008). Cidea is associated with lipid droplets and insulin sensitivity in humans. *Proc. Natl. Acad. Sci. USA* 105, 7833–7838.
 52. Wu, C., Kraft, P., Zhai, K., Chang, J., Wang, Z., Li, Y., Hu, Z., He, Z., Jia, W., Abnet, C.C., et al. (2012). Genome-wide association analyses of esophageal squamous cell carcinoma in Chinese identify multiple susceptibility loci and gene-environment interactions. *Nat. Genet.* 44, 1090–1097.
 53. Nica, A.C., Montgomery, S.B., Dimas, A.S., Stranger, B.E., Beazley, C., Barroso, I., and Dermitzakis, E.T. (2010). Candidate causal regulatory effects by integration of expression QTLs with complex trait genetic associations. *PLoS Genet.* 6, e1000895.
 54. Bryois, J., Buil, A., Evans, D.M., Kemp, J.P., Montgomery, S.B., Conrad, D.F., Ho, K.M., Ring, S., Hurler, M., Deloukas, P., et al. (2014). Cis and trans effects of human genomic variants on gene expression. *PLoS Genet.* 10, e1004461.
 55. Gupta, R.K., Arany, Z., Seale, P., Mepani, R.J., Ye, L., Conroe, H.M., Roby, Y.A., Kulaga, H., Reed, R.R., and Spiegelman, B.M. (2010). Transcriptional control of preadipocyte determination by Zfp423. *Nature* 464, 619–623.
 56. Ding, L., Pang, S., Sun, Y., Tian, Y., Yu, L., and Dang, N. (2014). Coordinated Actions of FXR and LXR in Metabolism: From Pathogenesis to Pharmacological Targets for Type 2 Diabetes. *Int. J. Endocrinol.* 2014, 751859.
 57. Greenawald, D.M., Dobrin, R., Chudin, E., Hatoum, I.J., Suver, C., Beaulaurier, J., Zhang, B., Castro, V., Zhu, J., Sieberts, S.K., et al. (2011). A survey of the genetics of stomach, liver, and adipose gene expression from a morbidly obese cohort. *Genome Res.* 21, 1008–1016.
 58. Björntorp, P. (1990). “Portal” adipose tissue as a generator of risk factors for cardiovascular disease and diabetes. *Arteriosclerosis* 10, 493–496.
 59. Makki, K., Froguel, P., and Wolowczuk, I. (2013). Adipose tissue in obesity-related inflammation and insulin resistance: cells, cytokines, and chemokines. *ISRN Inflamm.* 2013, 139239.
 60. Sun, S., Ji, Y., Kersten, S., and Qi, L. (2012). Mechanisms of inflammatory responses in obese adipose tissue. *Annu. Rev. Nutr.* 32, 261–286.
 61. Xu, H., Barnes, G.T., Yang, Q., Tan, G., Yang, D., Chou, C.J., Sole, J., Nichols, A., Ross, J.S., Tartaglia, L.A., and Chen, H. (2003). Chronic inflammation in fat plays a crucial role in the development of obesity-related insulin resistance. *J. Clin. Invest.* 112, 1821–1830.
 62. Ferrante, A.W., Jr. (2007). Obesity-induced inflammation: a metabolic dialogue in the language of inflammation. *J. Intern. Med.* 262, 408–414.
 63. Ito, A., Suganami, T., and Ogawa, Y. (2009). Inflammatory changes in obese adipose tissue. *Nippon Rinsho* 67, 270–276.
 64. Claussnitzer, M., Dankel, S.N., Kim, K.H., Quon, G., Meuleman, W., Haugen, C., Glunk, V., Sousa, I.S., Beaudry, J.L., Puvion-Dutilleul, V., et al. (2015). FTO Obesity Variant Circuitry and Adipocyte Browning in Humans. *N. Engl. J. Med.* 373, 895–907.
 65. O’Connell, J., Lynch, L., Cawood, T.J., Kwasnik, A., Nolan, N., Geoghegan, J., McCormick, A., O’Farrelly, C., and O’Shea, D. (2010). The relationship of omental and subcutaneous adipocyte size to metabolic disease in severe obesity. *PLoS ONE* 5, e9997.
 66. Borén, J., Taskinen, M.R., Olofsson, S.O., and Levin, M. (2013). Ectopic lipid storage and insulin resistance: a harmful relationship. *J. Intern. Med.* 274, 25–40.
 67. Trayhurn, P. (2013). Hypoxia and adipose tissue function and dysfunction in obesity. *Physiol. Rev.* 93, 1–21.
 68. Whittle, A.J., Jiang, M., Peirce, V., Relat, J., Virtue, S., Ebinuma, H., Fukamachi, I., Yamaguchi, T., Takahashi, M., Murano, T., et al. (2015). Soluble LR11/SorLA represses thermogenesis in adipose tissue and correlates with BMI in humans. *Nat. Commun.* 6, 8951.