

An integrated approach to identify environmental modulators of genetic risk factors for complex traits

Brunilda Balliu,^{1,13,*} Ivan Carcamo-Orive,^{2,13} Michael J. Gloude-mans,³ Daniel C. Nachun,⁴ Matthew G. Durrant,⁵ Steven Gazal,⁶ Chong Y. Park,⁷ David A. Knowles,^{8,9} Martin Wabitsch,¹⁰ Thomas Quertermous,¹¹ Joshua W. Knowles,^{11,13,*} and Stephen B. Montgomery^{12,13}

Summary

Complex traits and diseases can be influenced by both genetics and environment. However, given the large number of environmental stimuli and power challenges for gene-by-environment testing, it remains a critical challenge to identify and prioritize specific disease-relevant environmental exposures. We propose a framework for leveraging signals from transcriptional responses to environmental perturbations to identify disease-relevant perturbations that can modulate genetic risk for complex traits and inform the functions of genetic variants associated with complex traits. We perturbed human skeletal-muscle-, fat-, and liver-relevant cell lines with 21 perturbations affecting insulin resistance, glucose homeostasis, and metabolic regulation in humans and identified thousands of environmentally responsive genes. By combining these data with GWASs from 31 distinct polygenic traits, we show that the heritability of multiple traits is enriched in regions surrounding genes responsive to specific perturbations and, further, that environmentally responsive genes are enriched for associations with specific diseases and phenotypes from the GWAS Catalog. Overall, we demonstrate the advantages of large-scale characterization of transcriptional changes in diversely stimulated and pathologically relevant cells to identify disease-relevant perturbations.

Introduction

Genome-wide association studies (GWASs) have identified thousands of genetic variants associated with complex diseases and traits.¹ The majority of these variants fall into non-coding regions of the genome and, as a result, their mechanism of action remains largely unknown.² In recent years, researchers have gained an increasingly clear picture of which parts of the genome are active in a range of tissues and cell types.^{3–6} Integrating such information with results from GWASs has identified cell types, tissues, and regulatory elements relevant to specific diseases and phenotypes and moved the field toward a mechanistic understanding of GWAS hits.^{7–9} In addition, genomic colocalization and transcriptome-wide association studies combining results from GWASs and expression quantitative trait locus (eQTL) studies have identified candidate causal genes and their mechanisms of action.^{10–12}

Despite these advances, a modest fraction of GWAS-associated variants and eQTLs colocalize for any trait^{13,14} providing the perspective that many disease-relevant effects are modulated by yet-to-be-discovered environmental factors. To address this challenge, multiple studies

have mapped eQTLs *in vitro* that are responsive to the environment.^{15–26} For example, the Immune Variation Project identified eQTLs in human CD4+ T lymphocytes with different effects across distinct immune states.¹⁷ These previously unknown, immune-state-specific eQTLs were enriched for autoimmune-disease-associated variants, underscoring the importance of exploring contexts beyond tissues and cell types to reveal the specificity of genetic associations. Although there is mounting evidence that the environment modulates genetic effects, GWASs and eQTL studies rarely measure and test for genetic interactions with environmental exposures. This is, in part, due to the difficulty of identifying and collecting information on the most relevant environmental exposures in GWAS cohorts and performing eQTL studies in contexts that are relevant for the specific trait or disease.

In this study, we extend the current understanding of inherited variation in complex traits by implementing a framework to model signals from transcriptional responses to environmental perturbations to catalog and prioritize disease-relevant environments that can modulate genetic risk for complex traits and inform the functions of genetic variants and genes associated with complex traits.

¹Department of Computational Medicine, University of California, Los Angeles, Los Angeles, CA 90095, USA; ²Department of Medicine, Division of Cardiovascular Medicine, Cardiovascular Institute and Stanford Diabetes Research Center, Stanford University School of Medicine, Stanford, CA 94305, USA;

³Biomedical Informatics Training Program and Department of Pathology, Stanford University School of Medicine, Stanford, CA 94305, USA; ⁴Department of Pathology, Stanford University School of Medicine, Stanford, CA 94305, USA; ⁵Department of Genetics, Stanford University School of Medicine, Stanford, CA 94305, USA; ⁶Center for Genetic Epidemiology, Keck School of Medicine, University of Southern California, Los Angeles, CA 90033, USA; ⁷Department of Medicine, Division of Cardiovascular Medicine, Stanford University School of Medicine, Stanford, CA 94305, USA; ⁸New York Genome Center, New York, NY 10013, USA; ⁹Department of Computer Science, Columbia University, New York, NY 10027, USA; ¹⁰Department of Pediatrics and Adolescent Medicine, Division of Pediatric Endocrinology, Ulm University, Ulm 89075, Germany; ¹¹Department of Medicine, Division of Cardiology and Cardiovascular Institute, Stanford Diabetes Research Center, Stanford University School of Medicine, Stanford, CA 94305, USA; ¹²Department of Pathology and Department of Genetics, Stanford University School of Medicine, Stanford, CA 94305, USA

¹³These authors contributed equally

*Correspondence: bballiu@ucla.edu (B.B.), knowlej@stanford.edu (J.W.K.)

<https://doi.org/10.1016/j.ajhg.2021.08.014>

© 2021 American Society of Human Genetics.



Specifically, we first assessed environmental effects on gene expression levels in three metabolic human cell lines by performing RNA sequencing (RNA-seq) in muscle-, fat-, and liver-relevant cell lines treated with 21 different environmental perturbations related to aspects of glucose and insulin metabolism, kinase inhibitors, inflammation, fatty acid metabolism, etc. ($n = 234$ samples). We identified thousands of environmentally responsive genes underlying disease-associated response pathways and characterized the specificity and sharing of these effects across perturbations and cell lines. Next, to identify disease-relevant perturbations, we coupled our gene expression data with GWAS summary statistics of 31 complex traits and diseases as well as associations from the GWAS Catalog. We confirmed several well-established environmental-phenotype associations, e.g., the role of TGF- β 1 on asthma,²⁷ and provided additional evidence for recent and less well-understood associations, e.g., the role of leptin on major depressive disorder.²⁸ Last, to further illustrate how perturbation experiments inform the functions of complex trait-associated variants, we integrate our perturbation data with genomic colocalization studies and show that the effects of these perturbations in the relevant tissues identify context-specific molecular mechanisms of GWAS hits for diverse cardiometabolic traits.

This resource characterizes the dynamic transcriptional landscape in metabolic tissues and provides a framework for identification and prioritization of disease-relevant perturbations and disentanglement of the complex gene-environment interactions that determine disease susceptibility, which is particularly relevant for complex traits such as insulin resistance (IR), diabetes, and obesity.

Material and methods

Cell culturing and perturbations

Experiments were conducted with human skeletal muscle^{29,30} (HMCL-7304 myocytes; provided by Institute of Child Health, University College London), fat³¹ (terminally differentiated Simpson-Golabi-Behmel syndrome [SGBS] pre-adipocytes; provided by Dr. Martin Wabitsch, Ulm University, Ulm, Germany), and liver³² (HepG2; ATCC) cell lines. SGBS and HMCL-7304 cells were differentiated as described previously.^{30,31} Each cell line was starved for 6 h in Eagle's minimal essential medium (EMEM) with no fetal bovine serum (FBS) (HepG2), Dulbecco's modified Eagle's medium (DMEM)/F12 supplemented with pan/bio and penicillin/streptomycin but no FBS (SGBS adipocytes), or HMCL growth medium (PromoCell) without supplements. For the glucose condition DMEM, no glucose medium (Thermo) was used. After starvation and washing with PBS, the cells were exposed for 2 h to one of the 21 perturbations shown in Figure 1A in triplicate for each cell-line-perturbation combination. We selected a stimulation window of 2 h to allow enough time for transcriptional changes to occur and, at the same time, to minimize potential secondary responses that are not direct transcriptional effects of the selected perturbations, as previously assayed for insulin in the liver and skeletal muscle cells from mice.³³ In addition, we selected the concentrations of use, shown in Data S1, on the basis of consensus

research of available literature, particularly in the cells of interest. Last, we prepared triplicate control samples for glucose (no glucose medium) and four sets of triplicate control samples (no stimulation) for all other perturbations in each cell line, resulting in a sample size of 234.

RNA isolation and sequencing

After stimulation, the cells were washed with PBS and collected in PureLink RNA extraction lysis buffer supplemented with 1% 2-mercaptoethanol, flash-frozen in dry ice, and stored at -80°C . RNA extraction was performed with the PureLink RNA Mini kit (Thermo). 260/280 and RNA integrity number (RIN) values were assessed before sequencing for sample purity and integrity. Library preparation was performed at Novogene company. Liver samples were sequenced in HiSeq 4000 (Illumina), and fat and muscle were sequenced in Novaseq 6000 (Illumina) at 150 bp paired-end reads.

RNA-seq quality control

Reference genome (hg19) and gene model annotation (GRCh37.p13) files were downloaded from the UCSC Genome Browser website directly. Indexes of the reference genome were built with STAR,³⁴ and paired-end clean reads were aligned to the reference genome via STAR (v.2.6.0; with default option mismatch = 10). Bam files were filtered for uniquely mapped reads, sorted, and indexed via SAMtools³⁵ (v.1.4.1). We used HTSeq³⁶ (v.0.11.0) to count the read numbers mapped to each gene (with option -m union).

For all subsequent analyses, we focused only on expressed genes, i.e., genes that have median expression counts above 10 in at least one of the conditions (perturbation or control) within each cell line, i.e., 17,660, 17,140, and 16,722 genes for fat, muscle, and liver cells, respectively. As a measure of quality control, we looked at several RNA-seq technical metrics (Figures S1–S3), e.g., RNA integrity number, percent guanine-cytosine (GC) content, percent of uniquely mapped reads, etc. One sample (TGF- β 1 in fat) was dropped because of failing these quality control metrics. All remaining samples had RIN above 8, at least 16 million reads, GC content, an average of 36% of reads marked as PCR duplicates, at least 84% of their reads mapped uniquely, and an average of 95%, 4%, and 98% exon, intron, and transcript overlapping reads, respectively (Figure S1). Moreover, for all samples, their median Spearman expression correlation (D-statistic) with other samples was at least 0.96 (Figure S1). We used principal-component analysis (PCA) to identify gene expression outliers. After removing the low-quality sample mentioned above and adjusting for major components of variability (see below), no outliers are present based on the first two principal components (Figure S3).

Identifying major components of variability in RNA-seq data

We identified and adjust for major components of gene expression variability by computing the percent of gene expression variance for each gene explained by technical and biological variables via the linear mixed models implemented in the R package variance-Partition³⁷ (Figure S2). Prior to computing the percent of variance explained, we variance stabilized and log2-transformed the expression of each gene within cell lines by using the R package DESeq2. Then, we centered and scaled each gene to have zero mean and unit variance. For each gene expressed in each cell line, we used a linear mixed model with the effect of medium,

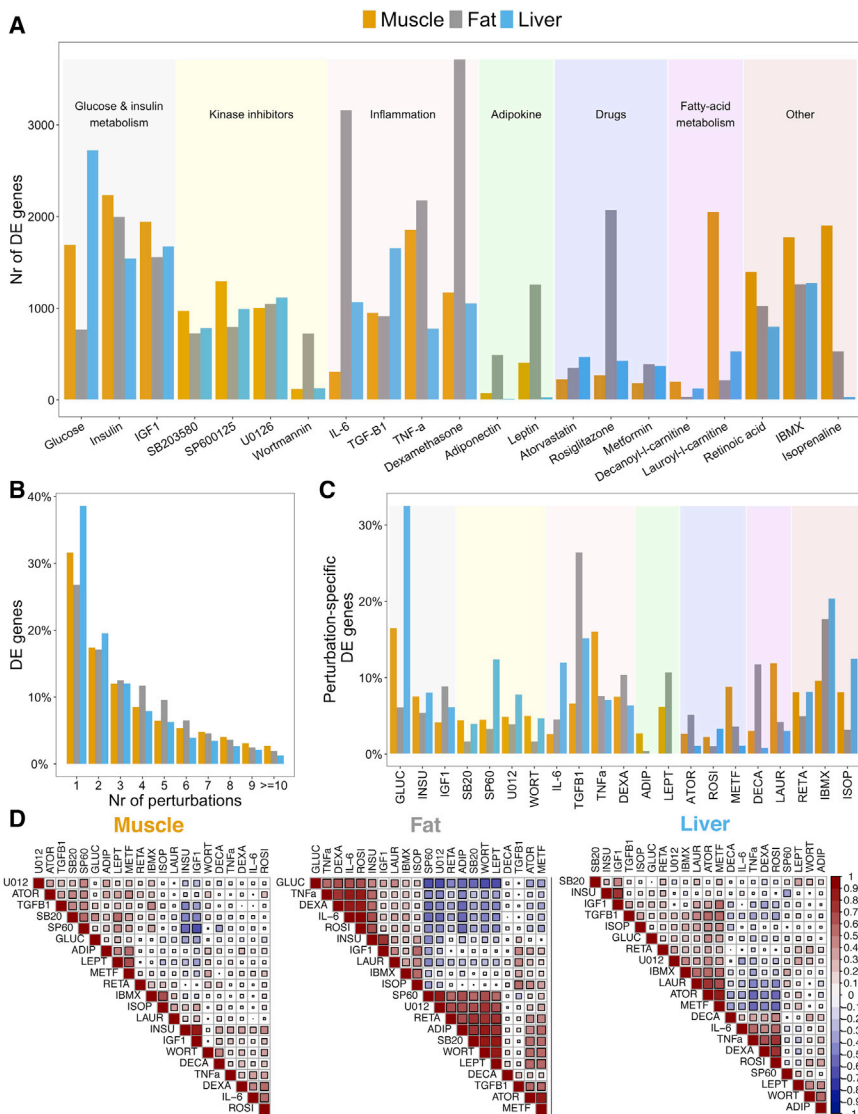


Figure 1. Perturbations induce large-scale changes in gene expression in muscle, fat, and liver

(A) Number of DE genes for each perturbation in each cell line (FDR < 5%). See [Data S2](#), [S3](#), and [S4](#) for extended DE summary statistics.

(B) Proportion of DE genes that change in response to up to ten perturbations in each cell line. See [Data S4](#) for extended results on sharing of DE genes.

(C) Proportion of perturbation-specific DE genes, i.e., genes that change in response to a single perturbation, for each perturbation and within each cell line.

(D) Correlation of DE patterns between different perturbations within each cell line. Each square is Spearman's correlation between the DE test statistic of a pair of perturbations across all genes.

DE, differentially expressed; FDR, false discovery rate.

by differentiation date and, within that, by RNA extraction date, cell collection, and starvation date, and plate number (for most but not all treatments, see [Data S1](#)). To adjust for all other variables, we include them in the model when testing for differential expression by treatment.

Principal-component analysis

To identify gene expression outliers, we run principal-component analysis (PCA) within cell lines ([Figure S3](#)). Prior to applying PCA, we variance stabilized and log₂-transformed the expression of each gene within cell lines by using the R package DESeq2.³⁸ Then, we centered and scaled each gene to have zero mean and

unit variance. We also applied PCA to expression data corrected for all major components of expression variability, as defined in the previous section. After removing the outlier sample mentioned above and after we regress out all major components of expression variability, we do not see any outliers based on the two first principal components.

number of cells plated, plate number, sequencing batch, cell collection, differentiation, RNA extraction, starvation, and treatment date as random and the effects of all other variables as fixed. Sequencing batch and number of cells plated only differed and were modeled for liver samples, while differentiation date only differed and was modeled for fat and muscle samples. We correct all subsequent analyses for all variables that explain, on average, more than 1% of expression variability in either cell line, i.e., percent GC content, percent exon overlapping reads, RNA concentration, percent of reads marked as PCR duplicates, RNA 260/280 ratio, RIN, and percent uniquely mapped reads. For analyses done in liver cells, we also correct for sequencing batch. The number of cells plated, plate number, and cell collection, starvation, differentiation, and RNA extraction date are highly correlated with the treatment and could act as confounders of the treatment effect. To account for this, for liver cells we matched treated and untreated samples by the number of cells plated, collection and starvation date, and plate number (for most but not all treatments, see [Data S1](#)). Because of this matching, we could not correct our analyses for RNA extraction date because it was collinear with treatment status within each treated-untreated pair of samples. For fat and muscle, we matched treated and untreated samples

by differentiation date and, within that, by RNA extraction date, cell collection, and starvation date, and plate number (for most but not all treatments, see [Data S1](#)). To adjust for all other variables, we include them in the model when testing for differential expression by treatment.

Differential expression analyses

We characterize transcriptional responses to each perturbation in each cell line by using the negative binomial models implemented in the R package DESeq2,³⁸ adjusting for major technical components of expression variability identified in the last section. To account for multiple testing across cell lines, perturbations, and genes, we use the hierarchical error control strategies implemented in the R package TreeBH³⁹ with cell line, genes, and treatments in levels 1, 2, and 3, respectively. This hierarchical procedure adjusts for all the tests performed and allows us to make statements about the differential expression at the gene, gene-perturbation, and gene-perturbation-cell-line level. We call a gene perturbation specific within a cell line if the gene is differentially expressed (DE) in that specific perturbation but not in any other perturbation in that cell line (false discovery rate [FDR] < 5% at each level). A gene is

assumed perturbation and cell line specific if the gene is DE in that specific perturbation but not in any other perturbation in that or the other cell lines or that specific perturbation in the other cell lines (FDR < 5% at each level). To assess agreement between our DE results and external studies listed in [Table S1](#), we extracted lists of DE genes from each study and computed the proportion of those DE genes that validated in our study either by using the p_1 statistic⁴⁰ (when the list was large enough to allow this) or by computing the proportion of those DE genes that were also DE in our study at 5% FDR.

Correlation and hierarchical clustering of the transcriptional response to perturbations

We computed correlation and performed hierarchical cluster analysis of the transcriptional response to perturbations in each cell line by using the test statistics from the DE analyses for all genes that were significant (FDR < 5%) in at least one perturbation and cell line. We used the R package `corrplot`⁴¹ to get a graphical display of the correlation matrix and hierarchical clustering of the perturbations.

Enrichment analyses for biological pathways

We performed over-representation analysis⁴² by using the R package `clusterProfiler`⁴³ and pathways from the `ConsensusPathDB` database.⁴⁴ We adjust for multiple testing within each perturbation and cell line by using the Benjamini-Hochberg⁴⁵ procedure.

Linkage disequilibrium score regression analysis

We downloaded the baseline model linkage disequilibrium (LD) scores, regression weights, and allele frequencies from Kundaje et al.³ Annotations for each perturbation and cell line were built with the pipeline described on the LD score regression wiki and according to Finucane et al.⁹ Specifically, for each of the 63 combinations of 21 perturbations and three cell lines, we add 100 kb windows (default LDSCreg threshold) on either side of the transcribed region of each DE gene in that combination to construct a genome annotation corresponding to that perturbation-cell-line combination. Due to its unusual genetic architecture and LD pattern, we excluded the human leukocyte antigen (HLA) region from all analyses. Z scores for the significance of the estimated total heritability for each trait were computed as $h^2/se(h^2)$, where h^2 and $se(h^2)$ are the SNP-based heritability estimated and standard errors from LDSCreg. Z scores and p values for the significance of the partitioned and conditional heritability for each trait-perturbation-cell-type combination were obtained via the option `-h2-cts` flag. We adjust for multiple testing within each trait and cell line by using the Benjamini-Hochberg procedure.

Enrichment for diseases and traits in the GWAS Catalog

We downloaded the entire GWAS Catalog (v.1.0.2) with Experimental Factor Ontology (EFO) annotations, including the parent category of each trait, from the European Bioinformatics Institute (EBI) (see [web resources](#)). We assume the “MAPPED GENE”, i.e., the gene mapped to the strongest SNP as reported in the GWAS Catalog, is the GWAS gene. For the enrichment of groups of GWAS traits, i.e., EFO parent terms, we only keep the unique GWAS genes reported across all traits within each EFO parent term. We excluded all results annotated with “other measurement,” “other disease,” and “other trait” EFO parent terms as well as duplicated entries. For the enrichment of specific traits,

we only test traits with at least 100 reported associated genes. To test for the significance of the enrichment, we used Fisher's exact test. For each perturbation-cell-line combination, we use an equal number of non-DE genes matched for length and median gene expression by using the R package `optmatch`.⁴⁶ We adjust for multiple testing across all (parent) traits, cell lines, and perturbations by using the Benjamini-Hochberg procedure.

Colocalization analysis of GWAS and eQTL effects and combination with DE signal

We performed colocalization analysis by using our custom integration of the FINEMAP⁴⁷ and eCAVIAR¹¹ methods. For each GWAS and eQTL overlap (GWAS and eQTL $p < 5 \times 10^{-8}$ for at least one SNP in each), we narrowed our summary statistics to the set of SNPs tested for association with both the given GWAS trait and the given QTL trait and removed all sites containing less than ten SNPs after this filter. Using the full 1000 Genomes dataset from phase 3 as a reference population,⁴⁸ we estimated LD between every pair of SNPs. We then ran FINEMAP independently on the GWAS and the eQTL summary statistics to obtain posterior probabilities of causality for each of the remaining SNPs. Because the canonical colocalization posterior probability (CLPP) score described in the eCAVIAR method is highly conservative in regions with densely profiled, high-LD SNPs, we use the following LD-modified CLPP score, $CLPP_{mod} = \sum_{i=1}^N \sum_{j=1}^N g_i e_j LD_{ij}$, where N is the total number of variants at the locus, g_i is the probability that the i^{th} variant is the causal variant for the GWAS trait, e_j is the probability that the j^{th} variant is the causal variant for the eQTL trait, and LD_{ij} is the estimated LD (r^2) between the i^{th} and the j^{th} variant in a reference population (we use 1000 Genomes Phase 3 data here). $CLPP_{mod}$ represents the sum of causal probabilities across all pairs of GWAS-eQTL variants at the locus, and each pair's contribution to the final score is weighted by the LD between these two variants. Like the original CLPP score, $CLPP_{mod}$ takes values between 0 and 1, and high values indicate higher colocalization probabilities. Subsequent visual inspection of juxtaposed GWAS and eQTL LocusCompare plots at high- and low- $CLPP_{mod}$ -score loci confirmed that our LD-modified CLPP score detects true colocalized loci but without disproportionately penalizing high-LD loci.

To test whether the genes DE in at least one of our perturbations and cell lines are enriched for candidate causal IR genes for at least one IR-related trait and GTEx tissue, we used Fisher's exact test. Candidate causal IR genes, denoted as “High P(Causal),” are defined as genes with $CLPP_{mod}$ above 40%, which corresponds to the 80th $CLPP_{mod}$ percentile. To test for the significance of the difference in median $CLPP_{mod}$ between DE and non-DE genes for each combination of perturbation and cell line, we used the two-sample Wilcoxon rank-sum test. We adjust for multiple testing by using the Benjamini-Hochberg procedure.

Results

Transcriptome map of 21 perturbations across human skeletal muscle, fat, and liver cell lines

We generated a transcriptome map of multiple chemical and environmental perturbations in well-established human skeletal muscle, fat, and liver cell lines ($n = 234$ samples). Specifically, we studied 21 environmental perturbations covering multiple aspects of glucose and insulin

metabolism, inflammation, and fatty acid metabolism and including both low-density lipoprotein (LDL)-lowering and anti-diabetic drugs (Figure 1 and Data S1). For each perturbation and cell line and matched controls, we conducted assays in triplicate and applied differential expression analysis. We observed that most perturbations induced broad gene expression changes in at least one cell line at FDR < 5% (Figures 1A and S4A, Data S2, S3, and S4). Several perturbations induced broad changes across all cell lines; for example, insulin and IGF1 altered the gene expression of 1,500–2,000 genes in each cell line. Other perturbations had broad changes only in specific cell lines. For example, IL-6, lauroyl-L-carnitine, and glucose had more pronounced effects in fat, muscle, and liver, respectively, impacting the expression of 3,161, 2,051, and 2,724 genes. A limited number of studies have examined the effect of some of the perturbations in cell lines similar to the ones considered here. Despite several differences in study design, e.g., exposure time or concentration of use, we see a high agreement (56.48%–89.69%) in the list of DE genes (Table S1).

Despite the broad effects for each perturbation, multiple DE genes showed perturbation-specific effects within each cell line, highlighting a unique molecular response to each perturbation. We observed 1,883 genes in muscle, 1,813 genes in fat, and 2,231 genes in the liver altered by only a single perturbation in their respective cell lines (Figure 1B and Data S5). The largest proportions of perturbation-specific DE genes were found in glucose-stimulated liver cell lines and TGF- β 1-stimulated fat cell lines. For these perturbations, 32.6% and 26.4% of DE genes were not altered by any of the other 20 perturbations in the same cell line (Figure 1C). By further stratifying across these cell lines, we identified 627, 742, and 808 genes that were both perturbation- and cell line-specific DE genes in muscle, fat, and liver (FDR < 5%; Figure S4B and Data S5). Glucose-stimulated liver cells also provided the largest amount of perturbation- and cell-line-specific DE genes; 9.8% of DE genes were not altered by any of the other 20 perturbations in any cell line or by glucose stimulation in fat or muscle.

To identify the relationships between perturbations on the basis of their overall transcriptional responses, we assessed the correlation of DE genes between each pair of perturbations within the same cell line (Figures 1D). The correlation of the effect of some perturbations was similar across cell lines, e.g., the effects of insulin and IGF1 were positively correlated in all three cell lines, i.e., Spearman's $r = 0.88, 0.76,$ and 0.71 in muscle, fat, and liver, respectively. The relationship of other perturbations, however, was dependent on the cellular context, e.g., while the effects of glucose and wortmannin were moderately correlated in fat (Spearman's $r = -0.63$), their correlation in muscle and liver was low (Spearman's $r = 0.02$ and 0.2 , respectively).

To explore the shared and specific pathways altered by each perturbation, we performed enrichment analysis of

DE genes in annotated pathways from Consensus-PathDB⁴⁴ (Data S6, S7, and S8). Our analysis highlighted multiple shared pathways across perturbations and cell lines related to PI3K-AKT-mTOR, MAPK, adipogenesis, and TGF- β signaling (Figure S4C). We also observed several differences in pathway enrichments; for example, pathways related to FOXA2 and FOXA3 transcription factor networks had greater enrichment across several perturbations in the liver than in muscle and fat, transcriptional regulation by RUNX2 had greater enrichment in muscle than in liver and fat, and chromatin organization and remodeling pathways had greater enrichment in fat than in liver and muscle. In addition, for genes affected by multiple perturbations, we saw strong enrichment pathways related to insulin signaling and resistance.

Combined, our concurrent assessment of multiple metabolically relevant perturbations across cell lines highlights the relationships between complex cell-specific molecular mechanisms and provides a genome-wide map of genes and signaling pathways with potential environmental contributions to complex disease susceptibility.

Prioritizing complex disease-relevant environmental perturbations

To measure the relevance of diverse environmental perturbations in complex diseases, we analyzed our transcriptome data together with GWAS summary statistics for 31 diseases and complex traits broadly related to multiple cardiometabolic, psychiatric, autoimmune, and reproductive traits, as well as hematological measurements (Figures 2 and S5; Data S9). We hypothesized that environmental perturbations impact disease through the same genes that confer susceptibility to the trait. To this end, for each of the 21 perturbations across the three cell lines, we used stratified LD score regression⁹ (LDSCreg) to test whether disease heritability, i.e., the proportion of phenotypic variance determined by genotypic variance, is enriched in regions surrounding DE genes for that perturbation and cell line, adjusting for heritability explained by a baseline model of genetic architecture⁹ and by regions surrounding genes expressed in the specific cell line.

For 26 of the 31 traits tested, the SNP-based heritability estimate was sufficiently large to partition reliably with LDSCreg, i.e., heritability Z score ≥ 7 (Data S9). In 19 of these traits, at least one perturbation in at least one cell line was enriched for heritability (FDR < 10%; Figure 2). Several of the enrichments recapitulate important known biology. For example, among cardiometabolic traits, high-density lipoprotein (HDL) and triglyceride levels were enriched for dexamethasone ($p = 2.10 \times 10^{-3}$ and $p = 6.53 \times 10^{-3}$), a corticosteroid known to induce dyslipidemia,^{49,50} and cardiovascular disease was enriched for rosiglitazone ($p = 6.44 \times 10^{-3}$), an antidiabetic drug shown to increase risk of cardiovascular disease.⁵¹ In addition, these enrichments were often manifested through a single specific relevant cell line. For example, waist-hip ratio (WHR) heritability was enriched for genes whose



Figure 2. Prioritizing complex disease-relevant environmental perturbations via heritability enrichment analysis

Heritability enrichment results for each complex trait. Each point represents a perturbation-cell-line combination that passes the FDR < 10% cut-off. The y axis represents the $-\log_{10}(p)$ value of heritability enrichment, the x axis indicates perturbation, the color of the point indicates cell line, and the shading color within each panel indicates the perturbation category from Figure 1A. Numerical results are reported in Data S9.

expression is modified by perturbations in fat, while triglyceride and HDL level heritability were enriched for genes whose expression is modified by perturbations in the liver.

Several notable examples were also observed for other tested traits. For psychiatric disorders, leptin, a hormone produced and secreted by white adipose tissue that is associated with antidepressant-like actions,^{28,52,53} was enriched for heritability of major depressive disorder (MIM: 608516) via its effect in fat cell lines ($p = 2.11 \times 10^{-3}$). In addition,

adiponectin, plasma levels of which appear to be altered in neurological disorders with metabolic and inflammatory components,^{54–56} was enriched for heritability of schizophrenia (MIM: 181500) ($p = 5.81 \times 10^{-3}$). For tested autoimmune diseases, TGF- β 1, an immune-suppressive cytokine dysregulated in the intestines of inflammatory bowel disease (MIM: 266600) affected individuals,⁵⁷ was enriched for heritability of Crohn disease ($p = 1.18 \times 10^{-2}$), as well as heritability of allergy (MIM: 607154), eczema (MIM: 603165), and asthma (MIM: 600807)^{27,58–60} ($p = 1.08 \times$

10^{-4}), three diseases with shared genetic origin.⁶¹ Several perturbations were also enriched for the heritability of hematological measurements; for example, dexamethasone, a synthetic glucocorticoid known to deplete peripheral blood lymphocytes and impact immune response,⁶² was enriched for heritability of lymphocyte count ($p = 2.88 \times 10^{-4}$). Lastly, for reproductive traits, glucose was enriched for heritability of age at menarche (MIM: 610873)—older age at menarche is associated with reduced risk of glucose metabolism disorder⁶³—while IGF1, whose serum levels rapidly decrease after menopause,⁶⁴ was enriched for heritability of age at menopause (MIM: 300488) ($p = 2.21 \times 10^{-3}$).

Identifying environmental perturbations impacting GWAS-significant loci

Beyond the broad polygenic impact of the tested perturbations and to analyze a larger number of traits, we sought to prioritize the subset of perturbations that were enriched for impact on GWAS-significant loci in specific complex diseases. We tested for enrichment of DE genes for *cis*-SNPs associated with diseases and phenotypes in the GWAS Catalog.⁶⁵ Because many traits had a small number of associations, we first tested for enrichment within groups of similar traits, as defined in the GWAS Catalog (Figure 3, Data S10).

We observed significant enrichment for at least one perturbation and cell line across all 14 groups of complex diseases and traits tested (FDR < 10%). For example, genes responsive to the effect of rosiglitazone, an insulin sensitizer known to affect plasma lipid levels,⁶⁶ were enriched within GWAS-significant hits for lipid or lipoprotein measurements (odds ratio [OR] = 2.00 and $p = 5.98 \times 10^{-3}$). In addition, genes responsive to the effect of retinoic acid, a metabolite of vitamin A that is synthesized in the liver and whose signaling dysregulation contributes to hepatic disease,⁶⁷ were enriched within GWAS hits for liver enzyme measurements (OR_{Muscle} = 2.71 and $p_{\text{Muscle}} = 9.06 \times 10^{-4}$; OR_{Liver} = 2.60 and $p_{\text{Liver}} = 1.8 \times 10^{-2}$). Moreover, atorvastatin and metformin, two perturbations with highly correlated DE signals (Figure 1D) known to reduce cardiovascular morbidity,^{68–72} were both enriched within GWAS hits for cardiovascular measurements (OR_{ATOR-Liver} = 1.75 and $p_{\text{ATOR-Liver}} = 1.58 \times 10^{-2}$; OR_{ATOR-Muscle} = 2.53 and $p_{\text{ATOR-Muscle}} = 1.52 \times 10^{-2}$; OR_{METF-Liver} = 1.86 and $p_{\text{METF-Liver}} = 1.56 \times 10^{-2}$). In line with the LDSC regression-based enrichment for Crohn disease, we observed that genes responsive to TGF- β 1 were enriched within GWAS-significant hits for digestive system disorders (OR = 2.7 and $p = 3.96 \times 10^{-9}$).

More generally, we observed that GWAS hits for immune system disorders or inflammatory measurements were enriched in genes responsive to the effect of inflammatory perturbations, e.g., OR_{TNF α} = 1.92 and 1.77 and $p_{\text{TNF}\alpha}$ = 3.84×10^{-8} and 2.54×10^{-8} for immune system disorders and inflammatory measurements, respectively. Neurological disorders were also enriched for inflammatory pertur-

bations, although to a lesser extent (e.g., OR_{TNF α} = 1.35 and $p_{\text{TNF}\alpha}$ = 8.60×10^{-4}). Associations with lipid or lipoprotein measurements and drug metabolism traits were enriched in genes responsive to several perturbations via the liver, where most drug metabolism occurs,⁷³ and associations with body measurements were enriched via muscle.

For traits with a large number of GWAS hits, i.e., traits with at least 100 reported associated loci, we tested enrichments directly (Figure S5). In 14 of the 152 complex diseases and traits tested, we observed significant enrichment for at least one perturbation and cell line (FDR < 10%). For example, genes in muscle cells that were responsive to the effect of isoprenaline, a beta-adrenergic agonist with effects on cardiac muscle,⁷⁴ were enriched within GWAS-significant hits for cardiovascular disease (OR = 2.24 and $p = 2.56 \times 10^{-4}$). In addition, consistent with the LDSCreg-based enrichment of dexamethasone for HDL heritability in the liver, genes responsive to dexamethasone in the liver were enriched within GWAS-significant associations for total cholesterol levels (OR = 3.08 and $p = 1.94 \times 10^{-4}$). Lastly, genes responsive to IGF1 in the liver were enriched within significant associations for birth weight (OR = 3.86 and $p = 5.48 \times 10^{-5}$), consistent with prior observations of negative correlation between IGF1 levels and birth weight.^{75,76}

Environmental perturbations harbor causal genes and help inform their functions

A major challenge with GWAS data in isolation is identifying causal disease genes. Here, we assessed whether combining GWASs with relevant environmental perturbations helped to identify or reinforce causal disease genes and to inform on their molecular functions. Because many of our perturbations were related to cardiometabolic traits, including IR, obesity, and type 2 diabetes (T2D [MIM: 125853]), we tested whether genes affected by our panel of perturbations harbored candidate causal genes underlying loci for seven cardiometabolic traits. To assess this, we integrated our perturbation data with results from genomic colocalization analyses of GWAS loci for these seven traits and GTEx eQTLs in visceral and subcutaneous adipose, skeletal muscle, and liver tissues.⁵ We observed that genes with a transcriptional response to at least one of our environmental perturbations are enriched among the candidate causal genes, i.e., genes with high LD-modified posterior colocalization probability (CLPP_{mod}), for cardiometabolic traits (Figure 4A; OR = 1.40, Fisher's exact test p value = 5.33×10^{-4}). We next assessed whether DE genes for specific perturbation-cell-line combinations were more likely to be causal compared with non-DE genes (Figure 4B). Genes responsive to isoprenaline, SP600125 (a c-Jun N-terminal kinase inhibitor that plays an essential role in TLR-mediated inflammatory responses), and TNF α in fat had significantly higher median CLPPs_{mod} (FDR < 10%) compared to non-DE genes (Wilcoxon test; $p_{\text{ISOP}} = 8.7 \times 10^{-4}$, $p_{\text{SP60}} = 6.0 \times 10^{-3}$, and $p_{\text{TNF}\alpha} = 6.0 \times 10^{-3}$).

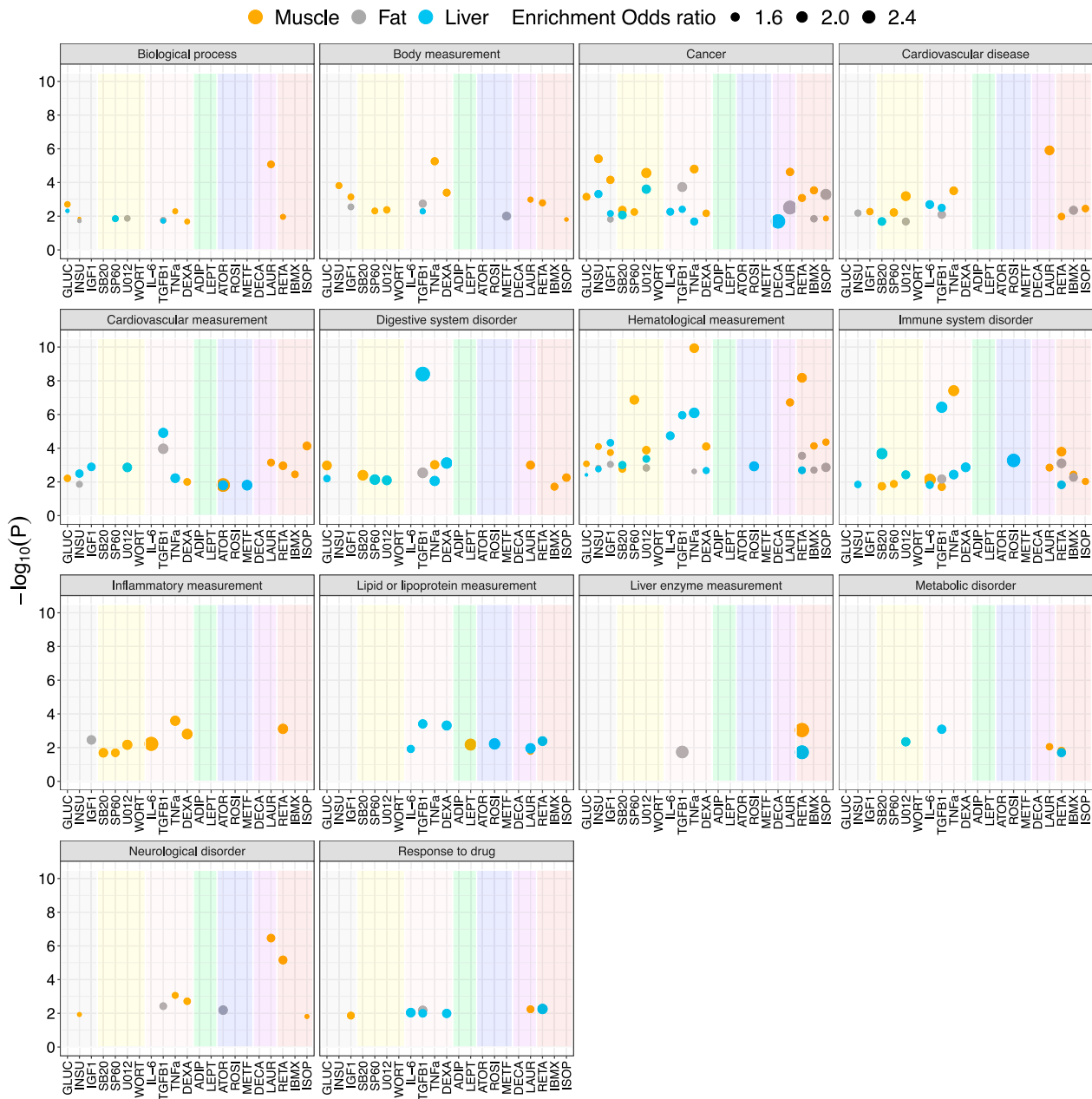


Figure 3. Identifying environmental perturbations impacting significant GWAS loci

GWAS enrichment results for each group of complex traits from the GWAS Catalog. Each point represents a perturbation-cell-line combination that passes the $FDR < 10\%$ cut-off; the color of the point indicates the cell line and the shading color within each panel indicates the perturbation category from Figure 1A. The y axis represents the $-\log_{10}(p)$ value of Fisher's exact test and the size indicates the odds ratio for the enrichment of GWAS hits of each group of traits from the GWAS Catalog. Numerical results are reported in Data S10. Results for specific traits, rather than groups of traits, are displayed in Figure S5.

To explore how perturbation experiments can inform the function of candidate causal genes underlying cardio-metabolic loci, we intersected the DE patterns in each cell line and perturbation with the colocalization patterns in the matched tissue. We illustrate four such examples (Figures 4C–4E): three loci in which a single gene showed high $CLPP_{mod}$ and one locus with more complex colocalization patterns (five out of seven genes in the locus showing high $CLPP_{mod}$).

Results from the colocalization analysis associated *FAM13A* (MIM: 613299) genetic variants in subcutaneous fat with several traits of interest, i.e., HDL, T2D, triglycerides, WHR, and fasting insulin (Figure 4C, locus 1). We recently described the role of *FAM13A* in adipocyte differentiation and the contribution to body fat distribution.⁷⁷ The DE patterns of *FAM13A* in our perturbation experiment (Figure 4D, locus 1) not only reinforce the role of *FAM13A* in adipose tissue but also suggest an additional

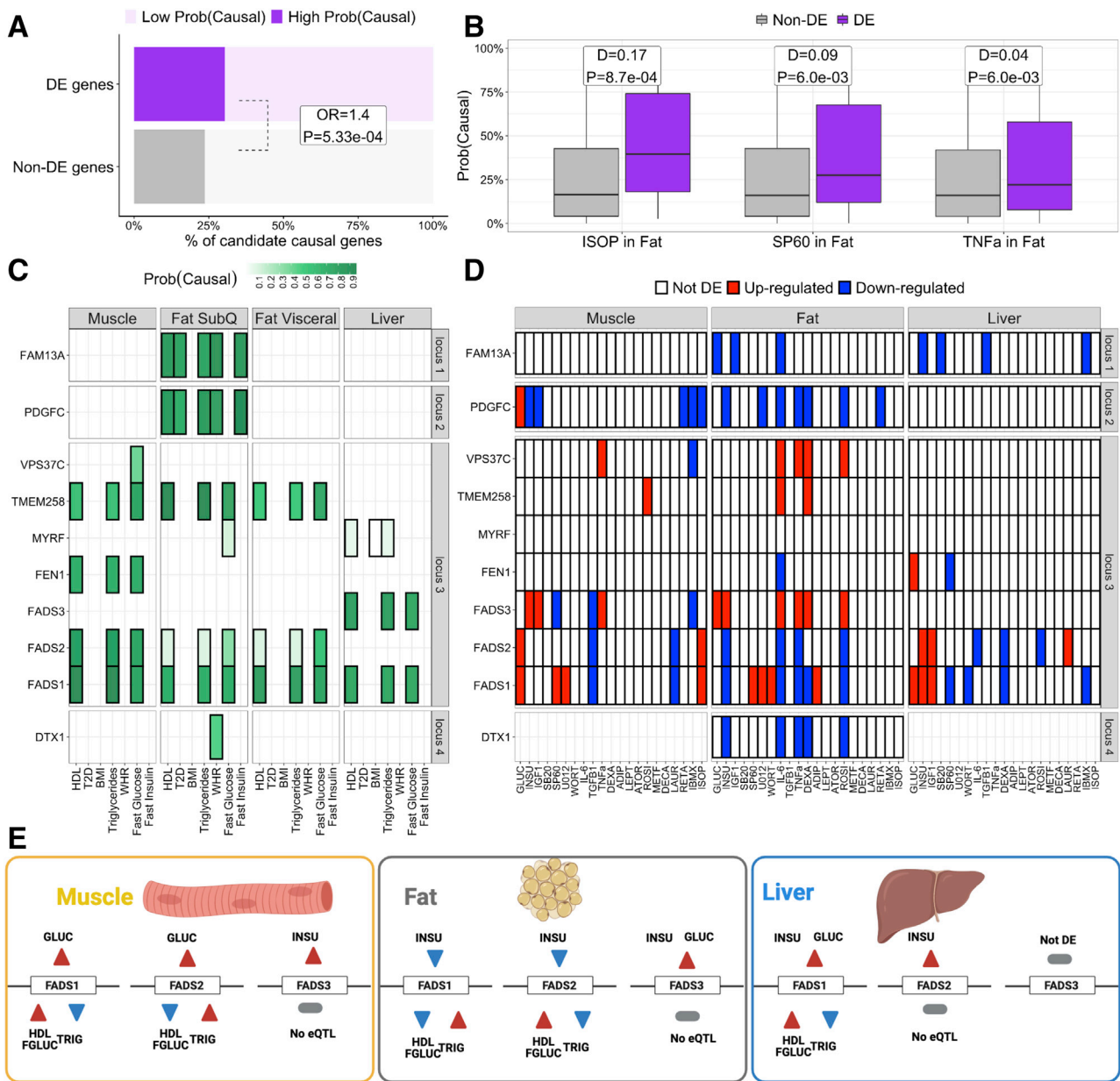


Figure 4. Environmental perturbations can help inform the functionality of causal genes underlying cardiometabolic trait loci
 (A) Percent of causal genes (High Prob(Causal)) underlying cardiometabolic traits loci that are DE (purple) or not DE (gray) in at least one perturbation and cell line. OR/P: odds ratio and Fisher's exact test p value for the enrichment of DE genes among causal genes compared to non-DE genes.
 (B) Perturbation-cell-line combinations with a significant (FDR < 10%) difference in median (D) $CLPP_{mod}$ between DE and non-DE genes, according to the two-samples Wilcoxon rank-sum test.
 (C and D) Examples of loci for which intersecting the effects of perturbations (C) with the colocalization results (D) helps inform the functionality of candidate causal genes. The color indicates $CLPP_{mod}$ (C) or DE direction (D). White boxes with crosses indicate that the gene was not tested for colocalization or DE.
 (E) Effect of glucose and insulin in the expression of the three FADS genes and the effect of the expression of these genes on HDL, fasting glucose (FGLUC), and triglycerides (TRIG), the three traits for which FADS genes colocalize. The color of the triangles indicates either the effect of the perturbation on the gene (red, upregulation; blue, downregulation) and the effect that up-/downregulation of the gene has on the phenotype (red, increased phenotype, blue, decreased phenotype). DE, differential expression; $CLPP_{mod}$, LD-adjusted colocalization posterior probability; FDR, false discovery rate.

metabolic function in the liver not captured by the colocalization results. The role of *FAM13A* in the regulation of hepatic glucose and lipid metabolism was recently confirmed by Lin et al.⁷⁸ Another candidate gene, *PDGFC*

(MIM: 608452), shows an identical colocalization pattern to *FAM13A* (Figure 4C, locus 2) and the perturbation data also support its importance in the adipose tissue (Figure 4D, locus 2). However, the perturbation data

identify an additional role of *PDGFC* in skeletal muscle in contrast with the role of *FAM13A* in the liver.

Another complementary example is illustrated by the colocalization for *DTX1* (MIM: 602582), which is specifically associated with WHR and subcutaneous fat (Figure 4C, locus 4) and whose expression is regulated by insulin, IL-6, TNF- α , dexamethasone, and rosiglitazone in mature human adipocytes (Figure 4D, locus 4).

Finally, genetic variants in the *FADS* locus have been associated with HDL cholesterol, triglyceride levels, fasting glucose, and T2D^{79–81} and our colocalization analysis was consistent with these observations (Figure 4C, locus 3). However, the high amount of LD, the gene density, and the pleiotropy of *FADS* genes have challenged the dissection of individual gene effects. Particularly informative is the case of *FADS1* (MIM: 606148), *FADS2* (MIM: 606149), and *FADS3* (MIM: 606150), for which the DE patterns for glucose and insulin (Figure 4D, locus 3) point, among others, to a fine-tuned cell-line- and perturbation-specific regulation of the *FADS* locus in the context of metabolic homeostasis (Figure 4E).

Together these results highlight the importance of perturbation experiments to contextualize GWAS associations and results from genomic colocalization analyses.

Discussion

We have profiled transcriptional responses to multiple environmental perturbations to identify disease-relevant perturbations modulating genetic risk for complex traits and to inform the functionality of causal genes. By combining gene expression data with GWAS summary statistics of complex traits, we show that the heritability of multiple complex traits is enriched in regions surrounding genes responsive to particular sets of perturbation-cell-line combinations. We confirmed several well-established associations, e.g., the role of TGF- β 1 on asthma, and provided additional evidence for recent and less-well-understood associations, e.g., the role of leptin on major depressive disorder. In addition, beyond the broad polygenic impacts of the tested perturbations, we were able to prioritize the subset of perturbations that are enriched for their impact on GWAS-significant loci in specific groups of complex diseases. We observed that environmentally responsive genes are enriched for *cis*-SNPs associated with a broad spectrum of diseases and phenotypes from the GWAS Catalog. Further, by integrating gene expression data with information from genomic colocalization studies, we showed that environmentally responsive genes are enriched for candidate causal genes for cardiometabolic traits and that the effects of these perturbations in the relevant tissues further suggest context-specific molecular mechanisms of GWAS hits for cardiometabolic traits.

Our approach interrogated multiple cell lines and perturbations, but comparable applications will be limited by the specific cell lines and environmental perturbations assayed and the exposure time and concentrations selected.

Here, we chose the concentrations of use on the basis of an extensive literature search and the exposure time to reduce the likelihood of secondary regulatory mechanisms. Our data demonstrate that, for most perturbations, there is a notable transcriptional response after a 2 h stimulation window. This suggests that longer stimulation times should be avoided, e.g., the typical 24 h stimulation period, if the purpose of the experimental setting is to describe genes that are likely to be directly regulated by the perturbations assayed. Further, the use of cell lines provides the opportunity for conducting well-controlled perturbation experiments; however, it is unknown the degree to which all findings would generalize to primary cells. Perturbation experiments on well-defined primary cells from multiple individuals might better mimic the disease-associated environment, but, until such experiments become more feasible, measuring expression in cell lines with tightly controlled cellular environments provides a more tractable setting to study gene-environment interactions.^{20,26} For some of the diseases we considered, the studied cell lines might not represent the cell type or tissue through which disease is manifested. However, because we observed that cell lines can share transcriptional responses (Figure 1D), our study design has shown that we can identify important perturbations without the causal cell type being examined.

To perform heritability enrichment analyses, we build annotations for each perturbation by using a fixed (100 kb size) window around each DE gene, which might lead to wide estimates of regulatory territories and decrease power to fine-map disease-relevant contexts. As perturbed chromatin immunoprecipitation sequencing (ChIP-seq) experiments become available for these perturbations and cell lines they can help further narrow down the regulatory regions for these environments. In addition, because fine-mapping summary statistics are not available for many of the traits deposited in the GWAS Catalog, we used the closest genes as proxy of the causal gene. While the closest gene is the most likely causal gene in only about 50% of examined loci, we do not expect this to disproportionately affect DE versus non-DE genes and upwardly bias our enrichment results. Furthermore, because most perturbations affect a small number of genes, the LDSC regression analysis for the enrichment of DE genes for trait heritability is based on partitions that cover a very small part of the genome. All these factors can substantially reduce power to identify disease-relevant contexts, and we thus chose to use a 10% FDR threshold and highlight several positive control examples. We illustrate the impact of the FDR threshold on the enrichment for GWAS-significant loci in Figures S6 and S7. Last, our genomic colocalization analyses are based on eQTLs mapped in GTEx tissues from healthy donors and might miss the effect of regulatory variants that manifest uniquely in the presence of the perturbation environment. Although we do not expect this to upwardly bias the enrichment of DE genes for likely causal genes, it might mask the effects of eQTLs for perturbations that are less likely to be present

in healthy donors, e.g., dexamethasone, and reduce the power to detect enrichment of DE genes in candidate causal genes responsive to these perturbations.

In conclusion, we demonstrate the advantages of large-scale characterization of transcriptional changes in diversely stimulated and pathologically relevant cells to identify disease-relevant perturbations that modulate genetic risk for complex traits. We also provide a broad resource of the dynamic transcriptional landscape in metabolic tissues. To our knowledge, this is the largest and most complete study of transcriptional effects of metabolically relevant perturbations in human fat, liver, and skeletal muscle cell lines. In addition, we show that integrating GWAS and eQTL results with perturbation experiments can inform the function of candidate causal genes and prioritize genes and environmental stimuli for follow-up experiments. Combined, this work demonstrates how integrating differential expression, eQTL, and GWAS data can inform genetic and environmental components of complex disease mechanisms.

Data and code availability

The data discussed in this publication have been deposited in NCBI's Gene Expression Omnibus (GEO) and are accessible through GEO series accession number GEO: GSE179347. The code and processed data used for these analyses are available at <https://github.com/BrunildaBalliu/PerturbationGxE>.

Supplemental information

Supplemental information can be found online at <https://doi.org/10.1016/j.ajhg.2021.08.014>.

Acknowledgments

We thank Erik Ingelsson, Noah Zaitlen, Paivi Pajukanta, and Aldon J. Lusis for helpful comments and suggestions that helped to improve the quality of our manuscript. J.K. is funded by NIH R01 DK116750, R01 DK120565, R01 DK106236, R01 DK107437, P30DK116074, and ADA 1-19-JDF-108. S.B.M. is supported by NIH grants R01AG066490, U01HG009431, R01HL142015, R01HG008150, and U01HG009080. T.Q. is supported by NIH grants R01 HL139478, R01 HL145708, R01 HL134817, R01 HL151535, and R01 HL156846 and a Human Cell Atlas grant from the Chan Zuckerberg Foundation. M.G.D. is supported by the National Science Foundation Graduate Research Fellowship.

Declaration of interests

S.B.M. is on the SAB for Myome.

Received: April 6, 2021

Accepted: August 27, 2021

Published: September 27, 2021

Web resources

EBI, <https://www.ebi.ac.uk/gwas/docs/file-downloads>

References

1. MacArthur, J., Bowler, E., Cerezo, M., Gil, L., Hall, P., Hastings, E., Junkins, H., McMahon, A., Milano, A., Morales, J., et al. (2017). The new NHGRI-EBI Catalog of published genome-wide association studies (GWAS Catalog). *Nucleic Acids Res.* 45 (D1), D896–D901.
2. Maurano, M.T., Humbert, R., Rynes, E., Thurman, R.E., Haugen, E., Wang, H., Reynolds, A.P., Sandstrom, R., Qu, H., Brody, J., et al. (2012). Systematic localization of common disease-associated variation in regulatory DNA. *Science* 337, 1190–1195.
3. Kundaje, A., Meuleman, W., Ernst, J., Bilenky, M., Yen, A., Heravi-Moussavi, A., Kheradpour, P., Zhang, Z., Wang, J., Ziller, M.J., et al. (2015). Integrative analysis of 111 reference human epigenomes. *Nature* 518, 317–330.
4. Moore, J.E., Purcaro, M.J., Pratt, H.E., Epstein, C.B., Shores, N., Adrian, J., Kawli, T., Davis, C.A., Dobin, A., Kaul, R., et al. (2020). Expanded encyclopaedias of DNA elements in the human and mouse genomes. *Nature* 583, 699–710.
5. GTEx Consortium (2020). The GTEx Consortium atlas of genetic regulatory effects across human tissues. *Science* 369, 1318–1330.
6. Boix, C.A., James, B.T., Park, Y.P., Meuleman, W., and Kellis, M. (2021). Regulatory genomic circuitry of human disease loci by integrative epigenomics. *Nature* 590, 300–307.
7. Gusev, A., Lee, S.H., Trynka, G., Finucane, H., Vilhjálmsson, B.J., Xu, H., Zang, C., Ripke, S., Bulik-Sullivan, B., Stahl, E., et al. (2014). Partitioning heritability of regulatory and cell-type-specific variants across 11 common diseases. *Am. J. Hum. Genet.* 95, 535–552.
8. Finucane, H.K., Bulik-Sullivan, B., Gusev, A., Trynka, G., Reshef, Y., Loh, P.R., Anttila, V., Xu, H., Zang, C., Farh, K., et al. (2015). Partitioning heritability by functional annotation using genome-wide association summary statistics. *Nat. Genet.* 47, 1228–1235.
9. Finucane, H.K., Reshef, Y.A., Anttila, V., Slowikowski, K., Gusev, A., Byrnes, A., Gazal, S., Loh, P.R., Lareau, C., Shores, N., et al. (2018). Heritability enrichment of specifically expressed genes identifies disease-relevant tissues and cell types. *Nat. Genet.* 50, 621–629.
10. Giambartolomei, C., Vukcevic, D., Schadt, E.E., Franke, L., Hingorani, A.D., Wallace, C., and Plagnol, V. (2014). Bayesian test for colocalisation between pairs of genetic association studies using summary statistics. *PLoS Genet.* 10, e1004383.
11. Hormozdiari, F., van de Bunt, M., Segrè, A.V., Li, X., Joo, J.W.J., Bilow, M., Sul, J.H., Sankararaman, S., Pasaniuc, B., and Eskin, E. (2016). Colocalization of GWAS and eQTL Signals Detects Target Genes. *Am. J. Hum. Genet.* 99, 1245–1260.
12. Gusev, A., Ko, A., Shi, H., Bhatia, G., Chung, W., Penninx, B.W., Jansen, R., de Geus, E.J., Boomsma, D.I., Wright, F.A., et al. (2016). Integrative approaches for large-scale transcriptome-wide association studies. *Nat. Genet.* 48, 245–252.
13. Chun, S., Casparino, A., Patsopoulos, N.A., Croteau-Chonka, D.C., Raby, B.A., De Jager, P.L., Sunyaev, S.R., and Cotsapas, C. (2017). Limited statistical evidence for shared genetic effects of eQTLs and autoimmune-disease-associated loci in three major immune-cell types. *Nat. Genet.* 49, 600–605.
14. Umans, B.D., Battle, A., and Gilad, Y. (2021). Where Are the Disease-Associated eQTLs? *Trends Genet.* 37, 109–124.
15. 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.
16. Lee, M.N., Ye, C., Villani, A.C., Raj, T., Li, W., Eisenhaure, T.M., Imboywa, S.H., Chipendo, P.I., Ran, F.A., Slowikowski, K., et al. (2014). Common genetic variants modulate pathogen-sensing responses in human dendritic cells. *Science* 343, 1246980.
 17. Ye, C.J., Feng, T., Kwon, H.K., Raj, T., Wilson, M.T., Asinovski, N., McCabe, C., Lee, M.H., Frohlich, I., Paik, H.I., et al. (2014). Intersection of population variation and autoimmunity genetics in human T cell activation. *Science* 345, 1254665.
 18. 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.
 19. Quach, H., Rotival, M., Pothlichet, J., Loh, Y.E., Dannemann, M., Zidane, N., Laval, G., Patin, E., Harmant, C., Lopez, M., et al. (2016). Genetic Adaptation and Neandertal Admixture Shaped the Immune System of Human Populations. *Cell* 167, 643–656.e17.
 20. Moyerbrailean, G.A., Richards, A.L., Kurtz, D., Kalita, C.A., Davis, G.O., Harvey, C.T., Alazizi, A., Watza, D., Sorokin, Y., Hauff, N., et al. (2016). High-throughput allele-specific expression across 250 environmental conditions. *Genome Res.* 26, 1627–1638. <https://doi.org/10.1101/gr.209759.116>.
 21. Nédélec, Y., Sanz, J., Baharian, G., Szpiech, Z.A., Pacis, A., Dumaine, A., Grenier, J.C., Freiman, A., Sams, A.J., Hebert, S., et al. (2016). Genetic Ancestry and Natural Selection Drive Population Differences in Immune Responses to Pathogens. *Cell* 167, 657–669.e21.
 22. Knowles, D.A., Davis, J.R., Edgington, H., Raj, A., Favé, M.J., Zhu, X., Potash, J.B., Weissman, M.M., Shi, J., Levinson, D.F., et al. (2017). Allele-specific expression reveals interactions between genetic variation and environment. *Nat. Methods* 14, 699–702.
 23. Kim-Hellmuth, S., Bechheim, M., Pütz, B., Mohammadi, P., Nédélec, Y., Giangreco, N., Becker, J., Kaiser, V., Fricker, N., Beier, E., et al. (2017). Genetic regulatory effects modified by immune activation contribute to autoimmune disease associations. *Nat. Commun.* 8, 266.
 24. Calderon, D., Nguyen, M.L.T., Mezger, A., Kathiria, A., Müller, F., Nguyen, V., Lescano, N., Wu, B., Trombetta, J., Ribado, J.V., et al. (2019). Landscape of stimulation-responsive chromatin across diverse human immune cells. *Nat. Genet.* 51, 1494–1505.
 25. Alasoo, K., Rodrigues, J., Danesh, J., Freitag, D.F., Paul, D.S., and Gaffney, D.J. (2019). Genetic effects on promoter usage are highly context-specific and contribute to complex traits. *eLife* 8, e41673.
 26. Findley, A.S., Monziani, A., Richards, A.L., Rhodes, K., Ward, M.C., Kalita, S.A., Alazizi, A., Pazokitoroudi, A., Sankararaman, S., Wen, X., et al. (2021). Functional dynamic genetic effects on gene regulation are specific to particular cell types and environmental conditions. *eLife* 10, e67077.
 27. Li, H., Romieu, I., Wu, H., Sienna-Monge, J.J., Ramírez-Aguilar, M., del Río-Navarro, B.E., del Lara-Sánchez, I.C., Kistner, E.O., Gjessing, H.K., and London, S.J. (2007). Genetic polymorphisms in transforming growth factor beta-1 (TGFB1) and childhood asthma and atopy. *Hum. Genet.* 121, 529–538.
 28. Ge, T., Fan, J., Yang, W., Cui, R., and Li, B. (2018). Leptin in depression: a potential therapeutic target. *Cell Death Dis.* 9, 1096.
 29. Rokach, O., Ullrich, N.D., Rausch, M., Mouly, V., Zhou, H., Muntoni, F., Zorzato, F., and Treves, S. (2013). Establishment of a human skeletal muscle-derived cell line: biochemical, cellular and electrophysiological characterization. *Biochem. J.* 455, 169–177.
 30. Carcamo-Orive, I., Henrion, M.Y.R., Zhu, K., Beckmann, N.D., Cundiff, P., Moein, S., Zhang, Z., Alamprese, M., D'Souza, S.L., Wabitsch, M., et al. (2020). Predictive network modeling in human induced pluripotent stem cells identifies key driver genes for insulin responsiveness. *PLoS Comput. Biol.* 16, e1008491.
 31. Fischer-Posovszky, P., Newell, F.S., Wabitsch, M., and Tornqvist, H.E. (2008). Human SGBS cells - a unique tool for studies of human fat cell biology. *Obes. Facts* 1, 184–189.
 32. Aden, D.P., Fogel, A., Plotkin, S., Damjanov, I., and Knowles, B.B. (1979). Controlled synthesis of HBsAg in a differentiated human liver carcinoma-derived cell line. *Nature* 282, 615–616.
 33. Batista, T.M., Garcia-Martin, R., Cai, W., Konishi, M., O'Neill, B.T., Sakaguchi, M., Kim, J.H., Jung, D.Y., Kim, J.K., and Kahn, C.R. (2019). Multi-dimensional Transcriptional Remodeling by Physiological Insulin In Vivo. *Cell Rep.* 26, 3429–3443.e3.
 34. Dobin, A., Davis, C.A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., and Gingeras, T.R. (2013). STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* 29, 15–21.
 35. Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R.; and 1000 Genome Project Data Processing Subgroup (2009). The Sequence Alignment/Map format and SAMtools. *Bioinformatics* 25, 2078–2079.
 36. Anders, S., Pyl, P.T., and Huber, W. (2015). HTSeq—a Python framework to work with high-throughput sequencing data. *Bioinformatics* 31, 166–169.
 37. Hoffman, G.E., and Schadt, E.E. (2016). variancePartition: interpreting drivers of variation in complex gene expression studies. *BMC Bioinformatics* 17, 483.
 38. Love, M.I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* 15, 550.
 39. Bogomolov, W., Peterson, C.B., Benjamini, Y., and Sabatti, C. (2020). Hypotheses on a tree: new error rates and testing strategies. *Biometrika*. <https://doi.org/10.1093/biomet/asaa086>.
 40. Storey, J.D., and Tibshirani, R. (2003). Statistical significance for genomewide studies. *Proc. Natl. Acad. Sci. USA* 100, 9440–9445.
 41. Wei, T., and Simko, V. (2021). R package 'corrplot': Visualization of a Correlation Matrix. (Version 0.90). <https://github.com/taiyun/corrplot>.
 42. Boyle, E.I., Weng, S., Gollub, J., Jin, H., Botstein, D., Cherry, J.M., and Sherlock, G. (2004). GO:TermFinder—open source software for accessing Gene Ontology information and finding significantly enriched Gene Ontology terms associated with a list of genes. *Bioinformatics* 20, 3710–3715.
 43. Yu, G., Wang, L.-G., Han, Y., and He, Q.-Y. (2012). clusterProfiler: an R package for comparing biological themes among gene clusters. *OMICS* 16, 284–287.
 44. Kamburov, A., Wierling, C., Lehrach, H., and Herwig, R. (2009). ConsensusPathDB—a database for integrating human

- functional interaction networks. *Nucleic Acids Res.* 37, D623–D628.
45. Benjamini, Y., and Heller, R. (2008). Screening for partial conjunction hypotheses. *Biometrics* 64, 1215–1222.
 46. Hansen, B.B., Fredrickson, M., Buckner, J., Errickson, J., Rauh, R., and Solenberger, P. (2018). optmatch: Functions for Optimal Matching. <https://github.com/markmfredrickson/optmatch>.
 47. Benner, C., Spencer, C.C., Havulinna, A.S., Salomaa, V., Ripatti, S., and Pirinen, M. (2016). FINEMAP: efficient variable selection using summary data from genome-wide association studies. *Bioinformatics* 32, 1493–1501.
 48. Auton, A., Brooks, L.D., Durbin, R.M., Garrison, E.P., Kang, H.M., Korbel, J.O., Marchini, J.L., McCarthy, S., McVean, G.A., Abecasis, G.R.; and 1000 Genomes Project Consortium (2015). A global reference for human genetic variation. *Nature* 526, 68–74.
 49. Krupková, M., Sedová, L., Liska, F., Krenová, D., Kren, V., and Seda, O. (2010). Pharmacogenetic interaction between dexamethasone and Cd36-deficient segment of spontaneously hypertensive rat chromosome 4 affects triacylglycerol and cholesterol distribution into lipoprotein fractions. *Lipids Health Dis.* 9, 38.
 50. Arab Dolatabadi, A., and Mahboubi, M. (2015). A study of the influence of dexamethasone on lipid profile and enzyme lactate dehydrogenase. *J. Med. Life* 8, 72–76.
 51. Wallach, J.D., Wang, K., Zhang, A.D., Cheng, D., Grossetta Nardini, H.K., Lin, H., Bracken, M.B., Desai, M., Krumholz, H.M., and Ross, J.S. (2020). Updating insights into rosiglitazone and cardiovascular risk through shared data: individual patient and summary level meta-analyses. *BMJ* 368, 17078.
 52. Westling, S., Åhrén, B., Träskman-Bendz, L., and Westrin, A. (2004). Low CSF leptin in female suicide attempters with major depression. *J. Affect. Disord.* 81, 41–48.
 53. Milaneschi, Y., Lamers, F., Bot, M., Drent, M.L., and Penninx, B.W.J.H. (2017). Leptin Dysregulation Is Specifically Associated With Major Depression With Atypical Features: Evidence for a Mechanism Connecting Obesity and Depression. *Biol. Psychiatry* 81, 807–814.
 54. Song, X., Fan, X., Song, X., Zhang, J., Zhang, W., Li, X., Gao, J., Harrington, A., Ziedonis, D., and Lv, L. (2013). Elevated levels of adiponectin and other cytokines in drug naïve, first episode schizophrenia patients with normal weight. *Schizophr. Res.* 150, 269–273.
 55. Bartoli, F., Lax, A., Crocamo, C., Clerici, M., and Carrà, G. (2015). Plasma adiponectin levels in schizophrenia and role of second-generation antipsychotics: a meta-analysis. *Psychoneuroendocrinology* 56, 179–189.
 56. Bloemer, J., Pinky, P.D., Govindarajulu, M., Hong, H., Judd, R., Amin, R.H., Moore, T., Dhanasekaran, M., Reed, M.N., and Suppiramaniam, V. (2018). Role of Adiponectin in Central Nervous System Disorders. *Neural Plast.* 2018, 4593530.
 57. Ihara, S., Hirata, Y., and Koike, K. (2017). TGF- β in inflammatory bowel disease: a key regulator of immune cells, epithelium, and the intestinal microbiota. *J. Gastroenterol.* 52, 777–787.
 58. Wortley, M.A., and Bonvini, S.J. (2019). Transforming Growth Factor- β 1: A Novel Cause of Resistance to Bronchodilators in Asthma? *Am. J. Respir. Cell Mol. Biol.* 61, 134–135.
 59. Duvernelle, C., Freund, V., and Frossard, N. (2003). Transforming growth factor-beta and its role in asthma. *Pulm. Pharmacol. Ther.* 16, 181–196.
 60. Yao, Y.S., Chang, W.W., He, L.P., Jin, Y.L., and Li, C.P. (2016). An updated meta-analysis of transforming growth factor- β 1 gene: Three well-characterized polymorphisms with asthma. *Hum. Immunol.* 77, 1291–1299.
 61. Ferreira, M.A., Vonk, J.M., Baurecht, H., Marenholz, I., Tian, C., Hoffman, J.D., Helmer, Q., Tillander, A., Ullemer, V., van Dongen, J., et al. (2017). Shared genetic origin of asthma, hay fever and eczema elucidates allergic disease biology. *Nat. Genet.* 49, 1752–1757.
 62. Parrillo, J.E., and Fauci, A.S. (1978). Mechanisms of corticosteroid action on lymphocyte subpopulations. III. Differential effects of dexamethasone administration on subpopulations of effector cells mediating cellular cytotoxicity in man. *Clin. Exp. Immunol.* 31, 116–125.
 63. Ren, Y., Zou, H., Zhang, D., Han, C., and Hu, D. (2020). Relationship between age at menarche and risk of glucose metabolism disorder: a systematic review and dose-response meta-analysis. *Menopause* 27, 818–826.
 64. Nasu, M., Sugimoto, T., Chihara, M., Hiraumi, M., Kurimoto, F., and Chihara, K. (1997). Effect of natural menopause on serum levels of IGF-I and IGF-binding proteins: relationship with bone mineral density and lipid metabolism in perimenopausal women. *Eur. J. Endocrinol.* 136, 608–616.
 65. Buniello, A., MacArthur, J.A.L., Cerezo, M., Harris, L.W., Hayhurst, J., Malangone, C., McMahon, A., Morales, J., Mountjoy, E., Sollis, E., et al. (2019). The NHGRI-EBI GWAS Catalog of published genome-wide association studies, targeted arrays and summary statistics 2019. *Nucleic Acids Res.* 47 (D1), D1005–D1012.
 66. Shim, W.S., Do, M.Y., Kim, S.K., Kim, H.J., Hur, K.Y., Kang, E.S., Ahn, C.W., Lim, S.K., Lee, H.C., and Cha, B.S. (2006). The long-term effects of rosiglitazone on serum lipid concentrations and body weight. *Clin. Endocrinol. (Oxf.)* 65, 453–459.
 67. Shirakami, Y., Lee, S.-A., Clugston, R.D., and Blaner, W.S. (2012). Hepatic metabolism of retinoids and disease associations. *Biochim. Biophys. Acta* 1821, 124–136.
 68. Arca, M., and Gaspardone, A. (2007). Atorvastatin efficacy in the primary and secondary prevention of cardiovascular events. *Drugs (Suppl 1)*, 29–42.
 69. Collins, R., Reith, C., Emberson, J., Armitage, J., Baigent, C., Blackwell, L., Blumenthal, R., Danesh, J., Smith, G.D., DeMets, D., et al. (2016). Interpretation of the evidence for the efficacy and safety of statin therapy. *Lancet* 388, 2532–2561.
 70. Nesti, L., and Natali, A. (2017). Metformin effects on the heart and the cardiovascular system: A review of experimental and clinical data. *Nutr. Metab. Cardiovasc. Dis.* 27, 657–669.
 71. Han, Y., Xie, H., Liu, Y., Gao, P., Yang, X., and Shen, Z. (2019). Effect of metformin on all-cause and cardiovascular mortality in patients with coronary artery diseases: a systematic review and an updated meta-analysis. *Cardiovasc. Diabetol.* 18, 96.
 72. Orkaby, A.R., Driver, J.A., Ho, Y.L., Lu, B., Costa, L., Honerlaw, J., Galloway, A., Vassy, J.L., Forman, D.E., Gaziano, J.M., et al. (2020). Association of Statin Use With All-Cause and Cardiovascular Mortality in US Veterans 75 Years and Older. *JAMA* 324, 68–78.
 73. Remmer, H. (1970). The role of the liver in drug metabolism. *Am. J. Med.* 49, 617–629.
 74. Szymanski, M.W., and Singh, D.P. (2020). Isoproterenol. In *StatPearls* (StatPearls Publishing).
 75. Fall, C.H., Pandit, A.N., Law, C.M., Yajnik, C.S., Clark, P.M., Breier, B., Osmond, C., Shiell, A.W., Gluckman, P.D., and Barker, D.J. (1995). Size at birth and plasma insulin-like

- growth factor-1 concentrations. *Arch. Dis. Child.* 73, 287–293.
76. de Jong, M., Cranendonk, A., Twisk, J.W.R., and van Weissenbruch, M.M. (2017). IGF-I and relation to growth in infancy and early childhood in very-low-birth-weight infants and term born infants. *PLoS ONE* 12, e0171650.
77. Fathzadeh, M., Li, J., Rao, A., Cook, N., Chennamsetty, I., Seldin, M., Zhou, X., Sangwung, P., Gloudemans, M.J., Keller, M., et al. (2020). FAM13A affects body fat distribution and adipocyte function. *Nat. Commun.* 11, 1465.
78. Lin, X., Liou, Y.H., Li, Y., Gong, L., Li, Y., Hao, Y., Pham, B., Xu, S., Jiang, Z., Li, L., et al. (2020). FAM13A Represses AMPK Activity and Regulates Hepatic Glucose and Lipid Metabolism. *iScience* 23, 100928.
79. 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. (2013). Discovery and refinement of loci associated with lipid levels. *Nat. Genet.* 45, 1274–1283.
80. Dupuis, J., Langenberg, C., Prokopenko, I., Saxena, R., Soranzo, N., Jackson, A.U., Wheeler, E., Glazer, N.L., Bouatia-Naji, N., Gloyn, A.L., et al. (2010). New genetic loci implicated in fasting glucose homeostasis and their impact on type 2 diabetes risk. *Nat. Genet.* 42, 105–116.
81. Goodarzi, M.O., Guo, X., Cui, J., Jones, M.R., Haritunians, T., Xiang, A.H., Chen, Y.D., Taylor, K.D., Buchanan, T.A., Hsueh, W.A., et al. (2013). Systematic evaluation of validated type 2 diabetes and glycaemic trait loci for association with insulin clearance. *Diabetologia* 56, 1282–1290.