Transcripts with high distal heritability mediate genetic effects on complex traits

Anna L Tyler, J Matthew Mahoney, Mark P Keller, Candice N. Baker, Margaret Gaca, Anuj Srivastava, Isabela Gerdes Gyuricza, Madeleine Braun, Nadia A Rosenthal, Alan D Attie, Gary A Churchill and Gregory W Carter

## Abstract

Gene expression is an important mediator of genetic effects on phenotype. Although many genes are subject to simple, local regulation, recent evidence suggests that complex distal regulation may be more important in mediating trait variability. To investigate this possibility, we combined two large, data sets modeling diet-induced obesity and metabolic disease in genetically diverse mice. Using a novel high-dimensional mediation analysis, we identified a heritable composite transcript that explained 30% of the variation across all metabolic traits. The composite transcript was interpretable in terms of enriched biological processes and predicted obesity status in an independent mouse cohort as well as in human cohorts with measured gene expression. Transcripts contributing most strongly to this composite mediator tended to have complex, distal regulation distributed throughout the genome. These results suggest that trait-relevant variation in transcription is largely distally regulated, but is nonetheless identifiable, interpretable, and translatable across species.

## Introduction

In the quest to understand the genetic architecture of complex traits, gene expression is an important mediator between genotype and phenotype. There is ample evidence from genome-wide association studies (GWAS) that regulation of gene expression accounts for the bulk of the genetic effect on complex traits, as most trait-associated variants lie in gene regulatory regions1–7. It is widely assumed that these variants influence local transcription, and methods such as transcriptome-wide association studies (TWAS)8–11, summary data-based Mendelian randomization (SMR)10, and others capitalize on this idea to identify genes associated with multiple disease traits12–15

Despite the great promise of these methods, explaining trait effects with local gene regulation has been more difficult than initially assumed16,17. Although trait-associated variants tend to lie in non-coding, regulatory regions, they often do not have detectable effects on gene expression16 and tend not to co-localize with expression quantitative trait loci (eQTLs)17,18.

One possible explanation for these observations is that gene expression is not being measured in the appropriate cell types and thus local eQTLs influencing traits cannot be detected16. An alternative explanation that has been discussed in recent years is that effects of these variants are mediated not through local regulation of gene expression, but through distal regulation15,18–20.

In this model, a gene’s expression is influenced by many variants throughout the genome through their cumulative effects on a broader regulatory network. In other words, the heritable component of the transcriptome is an emergent state arising from the myriad molecular interactions defining and constraining gene expression.

To assess the role of wide-spread distal gene regulation on complex traits, we investigated diet-induced obesity and metabolic disease as an archetypal example. Diet-induced obesity and metabolic disease are genetically complex with hundreds of variants mapped through GWAS [REFS]. These variants are known to act through multiple tissues that interact dynamically with each other [REFS], including adipose tissue, pancreatic islets, liver, and skeletal muscle. The multi-system etiology of metabolic disease complicates mechanistic dissection of the genetic architecture, requiring large, dedicated data sets that include high-dimensional, clinically relevant phenotyping, dense genotyping in a highly recombined population, and transcriptome-wide measurements of gene expression in multiple tissues.

Measuring gene expression in multiple tissues is critical to adequately assess the extent to which local gene regulation varies across the tissues and whether such variablilty might account for previous failed attempts to identify trait-relevant local eQTL. Such data sets are extremely difficult to obtain in human populations, particularly in the large numbers of subjects required for adequate statistical power. Thus, to further investigate the role of local and distal gene regulation on complex traits, we generated two complementary data sets: A discovery data set in a large population of diversity outbred (DO) mice21, and an independent validation data set derived by crossing inbred strains from the Collaborative Cross (CC) mice22 to form CC F1 mice (CC-RIX). Both populations modeled diet-induced obesity and metabolic disease12

The DO population and CC recombinant inbred lines were derived from the same eight inbred founder mouse strains, five classical lab strains, and three strains more recently derived from wild mice21. They represent three subspecies of mouse *Mus musculus domesticus*, *Mus musculus musculus*, and *Mus musculus casteneus*, and capture 90% of the known variation in laboratory mice23. The DO mice are maintained with a breeding scheme that ensures equal contributions from each founder across the genome thus rendering almost the whole genome visible to genetic inquiry21. The CC mice were initially outcrossed to recombine the genomes from all eight founders, and then inbred for at least 20 generations to generate multiple inbred lines. Because these two populations have common ancestral haplotyped we could directly and unambiguously compare the local genetic effects on gene expression at the whole-transcriptome level while varying the population structure driving distal regulation.

In the DO population, we paired clinically relevant metabolic traits from 500 mice [REF], including body weight, plasma levels of insulin and glucose and plasma lipids, with transcriptome-wide gene expression in four tissues related to metabolic disease: adipose tissue, pancreatic islets, liver, and skeletal muscle. We measured similar metabolic traits in the CC-RIX and gene expression from three of the four tissues used in the DO: adipose tissue, liver, and skeletal muscle. Because the CC-RIX carry the same founder alleles as the DO, local gene regulation is expected to match between the populations, but because the alleles are recombined through the genome, distal effects are expected to vary from those in the DO, allowing us to directly assess the role of local gene regulation in driving trait-associated transcript variation. Together, these data enable a comprehensive view into the genetic architecture of metabolic disease.

## Results

To comprehensively assess the genetic control of gene expression in metabolic disease in mice, we assayed metabolic traits and multi-tissue gene expression in DO mice.

### Genetic variation contributed to wide phenotypic variation

Although the environment was consistent across all animals, the genetic diversity present in the DO population resulted in widely varying distributions across physiological measurements (Fig. [1](#fig:trait_overview)). For example, body weights of adult individuals varied from less than the average adult B6 body weight to several times the body weight of a B6 adult in both sexes (Males: 18.5 - 69.1g, Females: 16.0 - 54.8g) (Fig. [1](#fig:trait_overview)A). Fasting blood glucose (FBG) also varied considerably (Fig. [1](#fig:trait_overview)B), although few of the animals had FBG levels that would indicate pre-diabetes (19 animals, 3.8%), or diabetes (7 animals, 1.4%) according to previously developed cutoffs (pre-diabetes: FBG 250 mg/dL, diabetes: FBG 300, mg/dL)24. Males had higher FBG than females on average (Fig. [1](#fig:trait_overview)C) as has been observed before suggesting either that males were more susceptible to metabolic disease on the high-fat diet, or that males and females may require different thresholds for pre-diabetes and diabetes.

Body weight was strongly positively correlated with food consumption (Fig. [1](#fig:trait_overview)D 0.51, ) and fasting blood glucose (FBG) (Fig. [1](#fig:trait_overview)E, 0.21, ) suggesting a link between behavioral factors and metabolic disease. However, the heritability of this trait and others (Fig. [1](#fig:trait_overview)F) indicates that background genetics contribute substantially to correlates of metabolic disease in this population.

![](data:application/pdf;base64,)

Clinical overview. **A.** Distributions of final body weight in the diversity outbred mice. Sex is indicated by color. The average B6 male and female adult weights at 24 weeks of age are indicated by blue and green bars on the x-axis. **B.** The distribution of final fasting glucose across the population split by sex. Normal, pre-diabetic, and diabetic fasting glucose levels for mice are shown by colored bars along the x-axis. **C.** Males had higher fasting blood glucose on average than females. **D.** The relationship between food consumption and body weight for both sexes. **E.** Relationship between body weight and fasting glucose for both sexes. **F.** Heritability estimates for each physiological trait. Bars show standard error of the estimate. **G.** Correlation structure between pairs of physiological traits.

The trait correlations (Fig. [1](#fig:trait_overview)G) showed that most of the metabolic trait pairs were only modestly correlated indicating complex relationships among the measured traits. This low level of redundancy suggests a broad sampling of multiple heritable aspects of metabolic disease including overall body weight, glucose homeostasis, pancreatic composition and liver function.

### Distal Heritability Correlated with Phenotype Relevance

To comprehensively assess the genetic control of gene expression in metabolic disease we assayed adipose, islet, liver, and skeletal muscle gene expression in the DO cohort. We performed eQTL analysis using R/qtl225 (Methods) and identified both local and distal eQTLs for transcripts in each of the four tissues (Supp. Fig 1). Significant local eQTLs far outnumbered distal eQTLs (Supp. Fig. 1F) and tended to be shared across tissues (Supp. Fig. 1G) whereas the few significant distal eQTLs we identified tended to be tissue-specific (Supp. Fig. 1H)

We calculated the heritability of each transcript in terms of local and distal genetic factors (Methods). Overall, local and distal genetic factors contributed approximately equally to transcript abundance. In all tissues, both local and distal factors explained between 8 and 18% of the variance in the median transcript (Fig [2](#fig:motivation)A).

![](data:application/pdf;base64,)

Transcript heritability and trait relevance. **A.** Distributions of distal and local heritability of transcripts across the four tissues. Overall local and distal factors contribute equally to transcript heritability. The relationship between (**B.**) local and (**C.**) distal heritability and trait relevance across all four tissues. Here trait relevance is defined as the maximum correlation between the transcript and all traits. Local heritability was negatively correlated with trait relevance, and distal heritability is positively correlated with trait relevance. Pearson () and values for each correlation are shown in the upper-right of each panel.

To assess the importance of genetic regulation transcript levels to organism-level traits, we compared the local and distal heritabilities of transcripts to their trait relevance, defined as the maximum correlation of a transcript across all traits. The local heritability of transcripts was negatively correlated with their trait relevance (Fig. [2](#fig:motivation)B), suggesting that the more local genotype influenced transcript abundance, the less effect this variation had on the measured traits. Conversely, the distal heritability of transcripts was positively correlated with trait relevance (Fig. [2](#fig:motivation)C). That is, transcripts that were more highly correlated with the measured traits tended to be distally, rather than locally, heritable. Importantly, this pattern was consistent across all tissues, strongly suggesting that this is a generic finding. This finding is consistent with previous observations that low-heritability transcripts explain more expression-mediated disease heritability than high-heritability transcripts19. However, the positive relationship between trait correlation and distal heritability demonstrated further that there are diffuse genetic effects throughout the genome converging on trait-related transcripts.

### High-Dimensional Mediation identified a high-heritability composite trait that was mediated by a composite transcript

The above univariate analyses establish the importance of distal heritability for trait-relevant transcripts. However, the number of transcripts dramatically exceeds the number of phenotypes. Thus, we expect the heritable, trait-relevant transcripts to be highly correlated and organized according to coherent, emergent biological processes representing the mediating endophenotypes driving clinical trait variation. To identify these endophenotypes in a theoretically principled way, we developed a novel dimension-reduction technique, HDMA, that uses the theory of causal graphical models to identify a transcriptomic signature that is simultaneously 1) highly heritable, 2) strongly correlated to the measured phenotypes, and 3) conforms to the causal mediation hypothesis (Fig. [3](#fig:workflow)). HDMA projects the high-dimensional scores–a composite genome score (), a composite transcriptome score (), and a composite phenome score ()–and uses the univariate theory of mediation to constrain these projections to satisfy the hypotheses of perfect mediation, namely that upon controlling for the transcriptomic score, the genome score is uncorrelated to the phenome score. Formally, perfect mediation implies a constraint on the correlation coefficients among scores as

which is equivalent to the partial correlation of and after controlling for T\_C being zero. The value is called the path coefficient of the mediation model. The projections of the high-dimensional data matrices in HDMA are designed to satisfy this constraint, and thus conform to the perfect mediation hypothesis, as closely as possible. We stress, however, that validating any causal assertion requires direct experimentation and, thus, that the output of HDMA are scores that are consistent with causal mediation. Thus, HDMA is a strategy for causal hypothesis generation, where the causal mediator is a complex endophenotype learned from a high-dimensional readout.

Operationally, HDMA is closely related to generalized canonical correlation analysis (CCA), for which provably convergent algorithms have recently been developed26. A complete mathematical derivation and implementation details for HDMA are available in **Supp. Methods XXX**.

![](data:application/pdf;base64,)

High-dimensional mediation. **A.** Workflow indicating major steps of high-dimensional mediation. The genotype, transcriptome, and phenotype matrices were independently normalized and converted to kernel matrices representing the pairwise relationships between individuals for each data modality ( = genome kernel, = transcriptome kernel; = phenome kernel). High-dimensional mediation was applied to these matrices to maximize the direct path , the mediating pathway (arrows), while simultaneously minimizing the direct pathway (dotted line). The composite vectors that resulted from high-dimensional mediation were , , and . The partial correlations between these vectors indicated perfect mediation. Transcript and trait loadings were calculated as described in the methods. **B.** The null distribution of the path coefficient derived from 10,000 permutations compared to the observed path coefficient (red line). **C.** The null distribution of the - correlation vs. the - correlation compared with the observed value (red dot).

We used HDMA to identify the major axis of variation in the transcriptome was consistent with mediating the effects of the genome on metabolic traits (Fig [3](#fig:workflow)). Fig. [3](#fig:workflow)A shows the partial correlations () between the pairs of these composite vectors. The partial correlation between and was 0.42, and the partial correlation between and was 0.78. However, when the transcriptome was taken into account, the partial correlation between and was effectively zero (0.039). captured 30% of the overall trait variance, and its estimated heritability was 0.71 0.084, which was higher than any of the measured traits (Fig. [1](#fig:trait_overview)F). Thus, HDMA identified a maximally heritable metabolic composite trait and a highly heritable component of the transcriptome that are correlated as expected in the perfectly mediated model.

As discussed in Supp. Methods XXX, HDMA is related to a generalized form of CCA. Standard CCA is prone to over-fitting because in any two large matrices it can be trivial to identify highly correlated composite vectors [REF]. To assess whether our implementation of HDMA was similarly prone to over-fitting in a high-dimensional space, we performed permutation testing. We permuted the individual labels on the transcriptome matrix 1000 times and recalculated the path coefficient, which is the partial correlation of and multiplied by the partial correlation of and . This represents the strenght of the path from to that is putatively mediated through . The null distribution of the path coefficient is shown in Fig. [3](#fig:workflow)B, and the observed path coefficient from the original data is indicated by a red line. The observed path coefficient was well outside the null distribution generated by permutations (). Fig. [3](#fig:workflow)C illustrates this observation in more detail. Although we identified high correlations between and , and modest correlations between and in the null data (Fig [3](#fig:workflow)C), these two values could not be maximized simultaneously in the null data. In contrast, the red dot shows that in the real data both the - correlation and the - correlation could be maximized simultaneously suggesting that the path from genotype to phenotype through transcriptome is highly non-trivial and identifiable in this case. These results suggest that these composite vectors represent genetically determined variation in phenotype that is mediated through genetically determined variation in transcription.

### Body weight and insulin resistance were highly represented in the expression-mediated composite trait

Each composite score is simply a weighted combination of the measured variables and the magnitude and sign of the weights, called loadings, correspond the relative importance and directionality of each variable in the composite score. The loadings of each measured trait onto indicate how much each contributed to the composite phenotype. Final body weight contributed the most (Fig. [4](#fig:interpretation)), followed by homeostatic insulin resistance (HOMA\_IR) and fasting plasma insulin levels (Insulin\_Fasting). We can thus interpret as an index of metabolic disease (Fig. [4](#fig:interpretation)B). Individuals with high values of have a higher metabolic index and greater metabolic disease, including higher body weight and higher insulin resistance. We refer to as the metabolic index going forward. Traits contributing the least to the metabolic index were measures of cholesterol and pancreas composition. Thus, when we interpret the transcriptomic signature identified by HDMA, we are explaining primarily the putative transcriptional mediation of body weight and insulin resistance, as opposed to cholesterol measurements.

![](data:application/pdf;base64,)

Interpretation of loadings. **A.** Loadings across traits. Body weight and insulin resistance contributed the most to the composite trait. **B.** Phenotype scores across individuals. Individuals with large positive phenotype scores had higher body weight and insulin resistance than average. Individuals with large negative phenotype scores had lower body weight and insulin resistance than average. **C.** Distribution of transcript loadings in adipose tissue. For transcripts with large positive loadings, higher expression was associated with higher phenotype scores. For transcripts with large negative loadings, higher expression was associated with lower phenotype scores. **D.** Distribution of absolute value of transcript loadings across tissues. Transcripts in adipose tissue had the largest loadings indicating that adipose tissue gene expression was a strong mediator of genotype on body weight and insulin resistance.

### High-loading transcripts have low local heritability, high distal heritability, and were linked mechanistically to obesity

We interpreted large loadings onto transcripts as indicating strong mediation of the effect of genetics on metabolic index. Large positive loadings indicate that higher expression was associated with a higher metabolic index (i.e. higher risk of obesity and metabolic disease on the high-fat diet) (Fig. [4](#fig:interpretation)C). Conversely, large negative loadings indicate that high expression of these transcripts was associated with a lower metabolic index (i.e. lower risk of obesity and metabolic disease on the high-fat diet) (Fig. [4](#fig:interpretation)C). We used gene set enrichment analysis (GSEA)27,28 to look for biological processes and pathways that were enriched at the top and bottom of this list (Methods).

In adipose tissue, both GO processes and KEGG pathway enrichments pointed to an axis of inflammation and metabolism (Supp. Fig. 2 and Fig. [11](#fig:top_enrich_go)). GO terms and KEGG pathways associated with inflammation, particularly macrophage infiltration, were positively associated with metabolic index, indicating that increased expression in inflammatory pathways was associated with a higher metabolic index. It is well established that adipose tissue in obese individuals is inflamed [cite] and infiltrated by macrophages [cite], and the results here suggest that this may be a dominant heritable component of metabolic disease.

The strongest negative enrichments in adipose tissue were related to mitochondial activity in general, and thermogenesis in particular (Supp. Fig. 2 and Fig. [11](#fig:top_enrich_go)). It has been shown mouse strains with greater thermogenic potential are also less susceptible to obesity on a high-fat diet [cite].

Transcripts associated with the citric acid (TCA) cycle as well as the catabolism of the branched-chain amino acids (BCAA) (valine, leuceine, and isoleucine) were strongly enriched with negative loadings in adipose tissue (Supp. Fig. 3). Expression of genes in both pathways (for which there is some overlap) has been previously associated with insulin sensitivity12,29,30, suggesting that heritable variation in regulation of these pathways may influence risk of insulin resistance.

Looking a the 10 most positively and negatively loaded transcripts from each tissue, it is apparent that transcripts in the adipose tissue had the largest loadings, both positive and negative, of all tissues (Fig. [5](#fig:loading_heritability)A bar plot) This suggests that much of the effect of genetics on body weight and insulin reisistance is mediated through gene expression in adipose tissue. The strongest loadings in liver and pancreas were comparable, and those in skeletal muscle were the weakest (Fig. [5](#fig:loading_heritability)A), suggesting that less of the genetic effects were mediated through transcription in skeletal muscle. Heritability analysis showed that transcripts with the largest loadings had higher distal heritability than local heritability (Fig. [5](#fig:loading_heritability)A heat map and box plot). This pattern contrasts with transcripts nominated by TWAS (Fig. [5](#fig:loading_heritability)B), which tended to have lower loadings, higher local heritability and lower distal heritability. Transcripts with the highest local heritability in each tissue (Fig. [5](#fig:loading_heritability)C) had the lowest loadings, consistent with our findings above (Fig. [2](#fig:motivation)B).

We performed a literature search for the genes in each of these groups along with the terms “diabetes”, “obesity”, and the name of the expressing tissue to determine whether any of these genes had previous associations with metabolic disease in the literature (Methods). Multiple genes in each group had been previously associated with obesity and diabetes (Fig. [5](#fig:loading_heritability) bolded gene names). Genes with high loadings were most highly enriched for previous literature support. They were 2.375 more likely than TWAS hits and 3.8 times more likely than genes with high local heritability to be previously associated with obesity or diabetes.

![](data:application/pdf;base64,)

Transcripts with high loadings have high distal heritability and literature support. Each panel has a bar plot showing the loadings of transcripts selected by different criteria. Bar color indicates the tissue of origin. The heat map shows the local (L - left) and distal (D - right) heritability of each transcript. **A.** Loadings for the 10 transcripts with the largest positive loadings and the 10 transcripts with the largest negative loadings for each tissue. **B.** Loadings of TWAS candidates with the 10 largest positive correlations with traits and the largest negative correlations with traits across all four tissues. **C.** The transcripts with the largest local heritability (top 20) across all four tissues.

### Tissue-specific transriptional programs were associated with metabolic traits

Clustering of transcripts with top loadings in each tissue showed tissue-specific functional modules associated with obesity and insulin resistance (Fig. [6](#fig:toa)A) (Methods). The clustering highlights the importance of immune activation particularly in adipose tissue. The “mitosis” cluster had large positive loadings in three of the four tissues potentially suggesting system-wide hyptertrophy. Otherwise, all clusters were strongly loaded in only one or two tissues. For example, the lipid metabolism cluster was loaded most heavily in liver. The positive loadings suggest that high expression of these genes particularly in the liver was associated with increased metabolic disease. This cluster included the gene *Pparg*, whose primary role is in the adipose tissue where it is considered a master regulator of adipogenesis31. Agonists of *Pparg*, such as thiazolidinediones, are FDA-approved to treat type II diabetes, and reduce inflammation and adipose hyptertrophy31. Consistent with this role, the loading for *Pparg* in adipose tissue was negative, suggesting that higher expression was associated with leaner mice (Fig. [6](#fig:toa)B). In contrast, *Pparg* had a large positive loading in liver, where it is known to play a role in the development of hepatic steatosis, or fatty liver. Mice that lack *Pparg* specifically in the liver, are protected from developing steatosis and show reduced expression of lipogenic genes32,33. Overexpression of *Pparg* in the livers of mice with a *Ppara* knockout, causes upregulation of genes involved in adipogenesis34. In the livers of both mice and humans high *Pparg* expression is associated with hepatocytes that accumulate large lipid droplets and have gene expression profiles similar to that of adipocytes35,36.

The local and distal heritability of *Pparg* is low in adipose tissue suggesting its expression in this tissue is highly constrained in the population (Fig. [6](#fig:toa)B). However, the distal heritability of *Pparg* in liver is relatively high suggesting it is complexly regulated and has sufficient variation in this population to drive variation in phenotype. Both local and distal heribatility of *Pparg* in the islet are relatively high, but the loading is low, suggesting that variability of expression in the islet does not drive variation in metabolic index. These results highlight the importance of tissue context when investigating the role of heritable transcript variability in driving phenotype.

Gene lists for all clusters are available in Supplemental File XXX.

![](data:application/pdf;base64,)

Tissue-specific transcriptional programs were associated with obesity and insulin resistance. **A** Heat map showing the loadings of all transcripts with loadings greater than 2.5 standard deviations from the mean in any tissue. The heat map was clustered using k medoid clustering. Functional enrichments of each cluster are indicated along the left margin. **B** Loadings for *Pparg* in different tissues. **C** Local and distal of *Pparg* expression in different tissues.

### Gene expression, but not local eQTLs, predicted body weight in an independent population

To test whether the transcript loadings identified in the DO could be translated to another population, we tested whether they could predict metabolic phenotype in an independent population of CC-RIX mice, which were F1 mice derived from multiple pairings of Collaborative Cross (CC) [cite] strains (Fig. [7](#fig:cc_prediction)) (Methods). We tested two questions. First, we asked whether the loadings identified in the DO mice were relevant to the relationship between the transcriptome and the phenome in the CC-RIX. We predicted body weight (a surrogate for metabolic index) in each CC-RIX individual using measured gene expression in each tissue and the transcript loadings identified in the DO (Methods). The predicted body weight and acutal body weight were highly correlated in all tissues (Fig. [7](#fig:cc_prediction)B left column). The best prediction was achieved for adipose tissue, which supports the observation in the DO that adipose expression was the strongest mediator of the genetic effect on metabolic index. This result also confirms the validity and translatability of the transcript loadings and their relationship to metabolic disease.

![](data:application/pdf;base64,)

Transcription, but not local genotype, predicts phenotype in the CC-RIX. **A.** Workflow showing procedure for translating HDMA results to an independent population of mice. **B.** Relationships between the predicted metabolic index and measured body weight. The left column shows the predictions using measured transcripts. The right column shows the prediction using transcript levels imputed from local genotype. Gray boxes indicate measured quantities, and blue boxes indicate calculated quantities. The dots in each panel represent individual CC-RIX strains. The gray lines show the standard deviation on body weight for the strain.

The second question related to the source of the relevant variation in gene expression. If local regulation was the predominant factor influencing gene expression, we should be able to predict phenotype in the CC-RIX using transcripts imputed from local genotype (Fig. [7](#fig:cc_prediction)A). The DO and the CC-RIX were derived from the same eight founder strains and so carry the same alleles throughout the genome. We imputed gene expression in the CC-RIX using local genotype and were able to estimate variation in gene transcription robustly (Supp. Fig. 4). However, these imputed values failed to predict body weight in the CC-RIX when weighted with the loadings from HDMA. (Fig. [7](#fig:cc_prediction)B right column). This result suggests that local regulation of gene expression is not the primary factor driving heritability of complex traits, consistent with our findings in the DO population that distal heritability was a major driver of trait-relevant variation and that high-loading transcripts had comparatively high distal and low local heritability.

### Distally heritable transcriptomic signatures reflected variation in composition of adipose tissue and islets

The interpretation of global genetic influences on gene expression and phenotype is potentially more challenging than the interpretation and translation of local genetic influences, as genetic effects cannot be localized to individual gene variants or transcripts. However, there are global patterns across the loadings that can inform mechanism. For example, heritable variation in cell type composition can be inferred from transcript loadings. We observed above that immune activation in the adipose tissues was a highly enriched process correlating with obesity in the DO population. For example, in humans, it has been extensively observed that macrophage infiltration in adipose tissue is a marker of obesity and metabolic disease [REF]. To determine whether the immune activation reflected a heritable change in cell composition in adipose tissue in DO mice, we compared loadings of cell-type specific genes in adipose tissue (Methods). Consistent with human results, the mean loading of macrophage-specific genes was significantly greater than 0 (Fig. [8](#fig:human_translation)A), indicating that obese mice were genetically predisposed to have high levels of macrophage infiltration in adipose tissue in response to the high-fat, high-sugar diet. Loading for marker genes for other cell types were not statistically different from zero, indicating that changes in the abundance of those cell types is not a mediator of metabolic index.

![](data:application/pdf;base64,)

HDMA results translate to humans. **A.** Distribution of loadings for cell-type-specific transcripts in adipose tissue. **B.** Distribution of loadings for cell-type-specific transcripts in pancreatic islets (green). **C.** Null distributions for the mean loading of randomly selected transcripts in each cell type compared with the observed mean loading of each group of transcripts (red asterisk). **D.** Predictions of metabolic phenotypes in four adipose transcription data sets downloaded from GEO. In each study the obese/diabetic patients were predicted to have greater metabolic disease than the lean/non-diabetic patients based on the HDMA results from DO mice.

We also compared loadings of cell-type specific transcripts in islet (Methods). The mean loadings for alpha-cell specific transcripts were significantly greater than 0, while the mean loadings for delta- and endothelial-cell specific genes were significantly less than 0 (Fig. [8](#fig:human_translation)B). These results suggest either that mice with higher metabolic index had inherited a higher proportions of alpha cells, and lower proportions of endothelial and delta cells in their pancreatic islets, that such compositional changes were induced by the HFHS diet in a heritable way, or both. In either case, these results support the hypothesis that alterations in islet composition drive variation in metabolic index.

Notably, the loadings for pancreatic beta cell-type specific loadings was not significantly different from zero. We stress that this is not necessarily reflective of the function of the beta cells in the obese mice, but rather suggests that any variation in the number of beta cells in these mice was unrelated to obesity and insulin resistance, the major contributors to metabolic index. This is further consistent with the islet composition traits having small loadings in the phenome score (Fig. [4](#fig:interpretation)).

### Heritable transcriptomic signatures translated to human disease

Ultimately, the heritable transcriptomic signatures that we identified in DO mice will be useful if they inform pathogenicity and treatment of human disease. To investigate the potential for translation of the gene signatures identified in DO mice, we compared them to transcriptional profiles in obese and non-obese human subjects (Methods). We limited our analysis to adipose tissue because the adipose tissue signature had the strongest relationship to obesity and insulin resistance in the DO.

We calculated a predicted obesity score for each individual in the human studies based on their adipose tissue gene expression (Methods) and compared the predicted scores for obese and non-obese groups as well as diabetic and non-diabetic groups. In all cases, the predicted obesity scores were higher on average for individuals in the obese and diabetic groups compared with the lean and non-diabetic groups (Fig. [8](#fig:human_translation)D). This indicates that the distally heritable signature of metabolic index identified in DO mice is relevant to obesity and diabetes in human subjects.

### Existing therapies are predicted to target mediator gene signatures

Another potential application of the transcript loading landscape is in ranking potential drug candidates for the treatment of metabolic disease. Although high-loading transcripts may be good candidates for understanding specific biology related to obesity, the transcriptome overall is highly interconnected and redundant, and focusing on individual transcripts for treatment may be less effective than using broader transcriptomic signatures that capture the emergent biology [cite or remove]. The ConnectivityMap (CMAP) database37 developed by the Broad Institute allows querying thousands of compounds that reverse or enhance the extreme ends of transcriptomic signatures in multiple different cell types. By identifying drugs that reverse pathogenic transcriptomic signatures, we can potentially identify compounds that have favorable effects on gene expression.

To test this hypothesis, we queried the CMAP database through the CLUE online query tool (<https://clue.io/query/>, version 1.1.1.43) (Methods). We identified top anti-correlated hits across all cell types. To get more tissue-specific results, we also looked at top results in cell types that most closely resembled our tissues. We looked at results in adipocytes (ASC) as well as pancreatic tumor cells (YAPC) regardless of value (Supplemental Figure XXX and XXX).

Looking across all cell types, the notable top hits from the adipose tissue loadings included mTOR inhibitors and glucocorticoid agonists (Supplemental Figure XXX). It is thought that metformin, which is commonly used to improve glycemic control, acts, at least in part, by inhibiting mTOR signaling38,39. However, long-term use of other mTOR inhibitors, such as rapamycin, are known to cause insulin resistance and -cell toxicity39–41. Glucocorticoids are used to reduce inflammation, which was a prominent siganture in the adipose tissues, but these drugs also promote hyperglycemia and diabetes42,43. Accute treatment with glucocorticoids has further been shown to reduce thermogenesis in rodent adipocytes44–46, but increase thermogenesis in human adipocytes47,48. Thus, the pathways identified by CMAP across all cell types were highly related to the transcript loading profiles, but the relationship was not a simple reversal.

The top hit for the adipose composite transcript in CMAP adipocytes was a PARP inhibitor (Supplemental Figure XXXB). PARPs play a role in lipid metabolism and are involved in the development of obesity and diabetes49. PARP1 inhibition increases mitochondrial biogenesis50. Inihibition of PARP1 activity can further prevent necrosis in favor of the less inflammatory apoptosis51, thereby potentially reducing inflammation in stressed adipocytes. Other notable hits among the top 20 were BTK inhibitors, which have been observed to suppress inflammation and improve insulin resistance52 as well as to reduce insulin antibodies in type I diabetes53. IKK inhibitors have been shown to improve glucose control in type II diabetes54,55.

Among the top most significant hits for the transcript loadings from pancreatic islets (Fig. XXX), was suppression of T cell receptor signaling, which is known to be involved in Type 1 diabetes56, as well as TNFR1, which has been associated with mortality in diabetes patients57. Suppression of NOD1/2 signaling was also among the top hits. NOD1 and 2 sense ER stress58,59, which is associated with -cell death in type 1 and type 2 diabetes60. This cell death process is dependent on NOD1/2 signaling58, although the specifics have not yet been worked out.

We also looked specifically at hits in pancreatic tumor cells (YAPC) regardless of significance level to get a transcriptional response more specific to the pancreas. Hits in this list included widely used diabetes drugs, such as sulfonylureas, PPAR receptor agonists, and insulin sensitizers. Rosiglitazone is a PPAR- agonist and was one of the most prescribed drugs for type 2 diabetes before its use was reduced due to cardiac side-effects61. Sulfonylureas are another commonly prescribed drug class for type 2 diabetes, but also have notable side effects including hypoglycemia and accellerated -cell death62.

## Discussion

Here we used a novel high-dimensional mediation analysis (HDMA) to investigate the relative contributions of local and distal gene regulation to heritable trait variation in genetically diverse mouse models of diet-induced obesity and metabolic disease. We identified tissue-specific composite transcripts that are predicted to mediate the effect of genetic background on metabolic traits. Transcripts contributing most strongly to these composite transcripts were distally, but not locally heritable, and composite transcripts were able to predict obesity in a large, independent mouse population with divergent population structure, whereas models using local eQTL only could not. Moreover, the composite transcript from mouse adipose tissue translated to predict obesity and diabetes status in human cohorts with measured adipose gene expression. Taken together, these results support the hypothesis that gene expression mediating the effect of genetic background on metabolic phenotypes is primarily distally regulated, and that that the heritable endophenotypes defined by gene expression signatures translate between mice and humans. We speculate that the central importance of distal heritability found in this study is likely to a generic finding for complex common diseases and could have significant consequences for the development of therapies for these diseases.

Genetics is indispensable for the dissection of disease mechanisms because it is one of the only data modalities that supports causal inferences about molecules and disease outcomes [REF]. It has frequently been assumed that gene regulation in *cis* is the primary driver of genetically associated trait variation, but attempts to use local gene regulation to explain phenotypic variation have had limited success16,17. In recent years, evidence has mounted that distal gene regulation may be an important mediator of trait heritability18,19,63. It has been observed that transcripts with high local heritability explain less expression-mediated disease heritability than transcripts with low local heritability19. Consistent with this observation, genes located near GWAS hits tend to be complexly regulated18. They also tend to be enriched with functional annotations, in contrast to genes with simple local regulation, which tend to be depleted of functional annotations suggesting they are less likely to be directly involved in disease traits18. These observations are consistent with principles of robustness in complex systems64–66. If a transcript were both important to a trait and subject to strong local regulation, a population would be susceptible to extremes in phenotype that might frequently cross the threshold to disease. Indeed, strong disruption of highly trait-relevant genes is the cause of Mendelian disease. Our results are consistent, instead, with a more complex picture where genes whose expression can drive trait variation are buffered from local genetic variation but are extensively influenced indirectly by genetic variation in the regulatory networks converging on those genes.

Recently, the omnigenic model of complex traits has been proposed, which posits that complex traits are massively polygenic and that their heritability is spread out across the genome67. In the omnigenic model, genes are classified either as “core genes,” which directly impinge on the trait, or “peripheral genes,” which are not directly trait-related, but influence core genes through the complex gene regulatory network. HDMA explicitly models a central proposal of the omnigenic model which posits that once the expression of the core genes (i.e. trait-mediating genes) is accounted for, there should be no residual correlation between the genome and the phenome. Here, when the composite transcript was taken into account there was no residual correlation between the composite genome and composite phenome (Fig. [3](#fig:workflow)A).

Thus, the transcript loadings can be interpreted as indicating higher “core-ness” of a transcript. Unlike in the omnigenic model, we did not observe a clear demarcation between the core and peripheral genes in loading magnitude, but we do not necessarily expect a clear separation given the complexity of gene regulation and the genotype-phenotype map68.

An extension of the omnigenic model proposed that most heritability of complex traits is driven by weak distal eQTLs that are potentially below detection threshold in feasible sample sizes63. This is consistent with what we observed here. The transcripts with the largest loadings were strongly distally regulated and only weakly locally regulated, suggesting that distal gene regulation plays a primary role in driving heritable trait variation. We saw further that the patterns of distal heritability were complex and spread across the genome, even for transcripts whose expression was strongly regulated by distal factors, rather than localizing to a few detectable distal eQTLs. For example, *Nucb2*, had a high loading in islets and was also strongly distally regulated (66% distal heritability) (Fig. [5](#fig:loading_heritability)). This gene is expressed in pancreatic cells and is involved in insulin and glucagon release69–71. Although its transcription was highly heritable in islets, that regulation was distributed across the genome, with no clear distal eQTL (Supp. Fig. 5). Thus, although distal regulation of some genes may be strong, this regulation is likely to be highly complex and not easily localized.

We stress that HDMA is a method for causal hypothesis generation. As with any causal inference approach, the output of HDMA can only be said to be consistent with causal mediation but does not prove it, which requires experimentation with direct control over the mediating variable [REF]. The issue of experimentation, however, is subtle. The dimension-reduction in HDMA is distinguished by the fact that the putative causal intermediates can be emergent states defined by the expression of thousands of genes. This is a strength, because the mediating variable can be a higher-order process such as “macrophage activation and infiltration”, but, in contrast to univariate hypotheses at the level of individual transcripts, the relevant validation experiment may be technologically infeasible, unknowable a priori, or both. Nevertheless, downstream analyses of the composite transcripts strongly supports a causal interpretation. Indeed, the composite transcripts identified by HDMA are richly interpretable in both tissue- and gene-specific manners. The transcripts with the strongest loadings were enriched in biological functions previously known to be involved in the pathogenesis of metabolic disease, such as inflammation in adipose tissue. That these processes were identified in this analysis suggests additionally that they have a heritable component, and that some individuals are genetically susceptible to greater adipose inflammation on a high-fat, high-sugar diet.

Individual high-loading transcripts also demonstrated biologically interpretable, tissue-specific patterns. We highlighted *Pparg*, which is known to be protective in adipose tissue31 where it was negatively loaded, and harmful in the liver32–36, where it was positively loaded. Such granular patterns may be useful in generating hypotheses for further testing, and prioritizing genes as therapeutic targets. The tissue-specific nature of the loadings also may provide clues to tissue-specific effects, or side effects, of targeting particular genes system-wide.

We showed further that these composite transcripts can be used as weighted vectors in multiple types of analysis, such as drug prioritization using gene set enrichment analysis (GSEA) and the CMAP database. In particular, the CMAP analysis identified drugs which have been demonstrated to reverse insulin resistance and other aspects of metabolic disease. This finding supports the causal role of these full gene signatures in pathogenesis of metabolic disease and thus their utility in prioritizing drugs and gene targets as therapeutics.

Another useful application of the composite transcripts is to pair them with cell-type specific genes to generate causal hypotheses about changes in cell composition in individual tissues. Combining the multi-tissue, transcriptome-wide weighted vectors with public databases and data sets thus provides a path for generating a wide range of testable hypotheses. Moreover, each data set presented here was derived from human tissues or cell lines, thus demonstrating the translatability of these results. That the mouse-derived adipose composite transcript was able to classify human adipose gene expression in terms of obesity and diabetes status further supports the direct translatablility of these findings, the utility of HDMA, and the continued importance of mouse models of human disease in which it is possible to obtain complete transriptomes in mutliple tissues across large numbers of individuals.

Altogether, our results have shown that both tissue specificity and distal gene regulation are critically important to understanding the genetic architecture of complex traits. We identified important genes and gene signatures that were heritable, plausibly causal of disease, and translatable to other mouse populations and to humans. Finally, we have shown that by directly acknowledging the complexity of both gene regulation and the genotype-to-phenotype map, we can gain a new perspective on disease pathogenesis and develop actionable hypotheses about pathogenic mechanisms and potential treatments.

### Old discussion

Here we used a novel high-dimensional mediation analysis (HDMA) to investigate the relative contributions of local and distal gene regulation to heritable trait variation in a genetically diverse mouse model of diet-induced obesity and metabolic disease. We identified tissue-specific composite transcripts mediating the effect of genetic background on metabolic traits. Transcripts contributing most strongly to these composite transcripts were distally heritable. These composite transcripts, but not local eQTL, were able to predict obesity in an independent mouse population with divergent allelic structure. Moreover, the composite transcript from adipose tissue predicted obesity and diabetes status in human cohorts with measured adipose gene expression. Taken together, these results support the hypothesis that gene expression mediating the effect of genetic background on phenotype is primarily distally regulated, and that the gene regulatory networks influencing metabolic disease are conserved across mice and humans.

It has frequently been assumed that gene regulation in *cis* is the primary driver of genetically associated trait variation, but attempts to use local gene regulation to explain phenotypic variation have had limited success16,17. In recent years, evidence has mounted that distal gene regulation may be an important mediator of trait heritability18,19,63. It has been observed that transcripts with high local heritability explain less expression-mediated disease heritability than transcripts with low local heritability19. Consistent with this observation, genes located near GWAS hits tend to be complexly regulated18. They also tend to be enriched with functional annotations, in contrast to genes with simple local regulation, which tend to be depleted of functional annotations suggesting they are less likely to be directly involved in disease traits18. These observations are consistent with principles of robustness in complex systems64–66. If a transcript were both important to a trait and subject to strong local regulation, a population would be susceptible to extremes in phenotype that might frequently cross the threshold to disease. Indeed, strong disruption of highly trait-relevant genes is the cause of Mendelian disease.

The composite transcripts we identified here supported the hypothesis that distally regulated gene expression is the dominant mediator of trait variation. Transcript loadings (the degree to which they contributed to the composite transcript) were negatively correlated with local heritability and positively correlated with distal heritability. The most strongly loaded transcripts were enriched for functional annotations associated with metabolic disease. These distally regulated composite transcripts were highly heritable and explained a high proportion of disease risk, further supporting their role as mediators. The composite transcripts were moreover able to predict obesity in an independent cohort of mice whereas models using local eQTL only could not. Together these observations suggest that distal gene regulation was the dominant mode through which gene expression mediated the effect of genetic background on complex metabolic traits.

Identification of this distally heritable signature depended on the high-dimensional approach we used. Because HDMA uses a kinship matrix rather than genotypes at individual loci, it allows for arbitrarily complex gene regulation, as well as the interconnectedness and redundancy of the transcriptome. This feature also means that HDMA assumes that traits are highly polygenic and distributed across the genome. In contrast, one-dimensional, univariate approaches assume a large, localized genetic effect. Thus, the HDMA approach is consistent with the omnigenic model of complex traits which posits that complex traits are massively polygenic and that their heritability is spread out across the genome67. In the omnigenic model, genes are classified either as “core genes,” which directly impinge on the trait, or “peripheral genes,” which are not directly trait-related, but influence core genes through the complex gene regulatory network. HDMA explicitly models a central proposal of the omnigenic model which posits that once the expression of the core genes (i.e. trait-mediating genes) is accounted for, there should be no residual correlation between the genome and the phenome. Here, when the composite transcript was taken into account there was no residual correlation between the composite genome and composite phenome (Fig. [3](#fig:workflow)A).

Thus, the composite transcript is essentially a weighted vector with larger weights (loadings) indicating higher “core-ness” of a transcript. There was no clear demarcation between the core and peripheral genes in loading magnitude, but we do not necessarily expect a clear separation given the complexity of gene regulation and the genotype-phenotype map68. Still, the transcripts with the largest loadings had high distal heritability, low local heritability, and were enriched for biological processes related to metabolic traits, as we would predict for core genes.

An extension of the omnigenic model proposed that most heritability of complex traits is driven by weak distal eQTLs63. This is consistent with what we observed here. The transcripts with the largest loadings were strongly distally regulated and only weakly locally regulated, suggesting that distal gene regulation plays a primary role in driving heritable trait variation. We saw further that the patterns of distal heritability were complex spread across the genome. Even for transcripts whose expression was strongly regulated by distal factors, these factors were multiple and spread across the genome. For example, *Nucb2*, was a strongly mediating transcript in islet and was also strongly distally regulated (66% distal heritability) (Fig. [5](#fig:loading_heritability)). This gene is expressed in pancreatic cells and is involved in insulin and glucagon release69–71. Although its transcription was highly heritable in islets, that regulation was distributed across the genome, with no clear distal eQTL (Supp. Fig. 5). Thus, although distal regulation of some genes may be strong, this regulation is likely to be highly complex and not easily localized.

The composite transcripts identified by HDMA are richly interpretable in both tissue- and gene-specific manners. The transcripts with the strongest loadings were enriched in biological functions previously known to be involved in the pathogenesis of metabolic disease, such as inflammation in adipose tissue. That these processes were identified in this analysis suggests that they have a heritable component, and that some individuals are genetically susceptible to greater adipose inflammation on a high-fat, high-sugar diet.

Individual transcripts also demonstrated biologically interpretable, tissue-specific patterns. We highlighted *Pparg*, which is known to be protective in adipose tissue31 where it was negatively loaded, and harmful in the liver32–36, where it was positively loaded. Such granular patterns may be useful in generating hypotheses for further testing, and prioritizing genes as therapeutic targets. The tissue-specific nature of the loadings also may provide clues to tissue-specific effects, or side effects, of targeting particular genes system-wide, since antagonists of *Pparg* may reduce fatty liver disease, but exacerbate adipose tissue inflammation.

We showed further that these composite transcripts can be used as weighted vectors in multiple types of analysis, such as drug prioritization using gene set enrichment analysis (GSEA) and the CMAP database. In particular, the CMAP analysis identified drugs which have been demonstrated to reverse insulin resistance and other aspects of metabolic disease. This finding supports the causal role of these gene signatures in pathogenesis of metabolic disease and thus their utility in prioritizing drugs and gene targets as therapeutics.

Another useful application of the composite transcripts is to pair them with cell-type specific genes to generate hypotheses about cell composition in individual tissues. Combining the multi-tissue, transcriptome-wide weighted vectors with public databases and data sets thus provides a path for generating a wide range of testable hypotheses. Moreover, each data set presented here was derived from human tissues or cell lines, thus demonstrating the translatability of these results. That the mouse-derived adipose composite transcript was able to classify human adipose gene expression in terms of obesity and diabetes status further supports the direct translatablility of these findings, the utility of HDMA, and the continued importance of mouse models of human disease in which it is possible to obtain complete transriptomes in mutliple tissues across large numbers of individuals.

In conclusion, we have shown that both tissue specificity and distal gene regulation are critically important to understanding the genetic architecture of complex traits. We identified important genes and gene signatures that were heritable, causal of disease, and translatable to other mouse populations and to humans. Finally, we have shown that by directly acknowledging the complexity of both gene regulation and the genotype-to-phenotype map, we can gain a new perspective on disease pathogenesis and develop actionable hypotheses about pathogenic mechanisms and potential treatments.

## Data Availability

Here we tell people where to find the data

## Acknowledgements

Here we thank people

## Supplemental Figures

![](data:application/pdf;base64,)

Overview of eQTL analysis in DO mice. **A.** RNA seq samples from the four different tissues clustered by tissue. **B.-E.** eQTL maps are shown for each tissue. The -axis shows the position of the mapped eQTL, and the -axis shows the physical position of the gene encoding each mapped transcript. Each dot represents an eQTL with a minimum LOD score of 8. The dots on the diagonal are locally regulated eQTL for which the mapped eQTL is at the within 4Mb of the encoding gene. Dots off the diagonal are distally regulated eQTL for which the mapped eQTL is distant from the gene encoding the transcript. **F.** Comparison of the total number of local and distal eQTL with a minimum LOD score of 8 in each tissue. All tissues have comparable numbers of eQTL. Local eQTL are much more numerous than distal eQTL. **G.** Counts of transcripts with local eQTL shared across multiple tissues. The majority of local eQTL were shared across all four tissues. **H.** Counts of transcripts with distal eQTL shared across multiple tissues. The majority of distal eQTL were tissue-specific and not shared across multiple tissues. For both G and H, eQTL for a given transcript were considered shared in two tissues if they were within 4Mb of each other. Colored bars indicate the counts for individual tissues for easy of visualization.

![](data:application/pdf;base64,)

Bar plots showing normalized enrichment scores (NES) for KEGG pathways as determined by fast gene score enrichment analysis (fgsea). Only the top 10 positive and top 10 negative scores are shown. Colors indicate tissue. The name beside each bar shows the name of each enriched KEGG pathway.

![](data:application/pdf;base64,)

Bar plots showing normalized enrichment scores (NES) for GO terms as determined by fast gene score enrichment analysis (fgsea). Only the top 10 positive and top 10 negative scores are shown. Colors indicate tissue. The name beside each bar shows the name of each enriched GO term. The letters in parentheses indicate whether the term is from the biological process ontology (BP), the molecular function ontology (MF), or the cellular compartment ontology (CC).

![](data:application/pdf;base64,)

Validation of transcript imputation in the CC-RIX. **A.** Distributions of correlations between imputed and measured transcripts in the CC-RIX. The mean of each distribution is shown by the red line. All distributions were skewed toward positive correlations and had positive means near a Pearson correlation (r) of 0.5. **B.** The relationship between the correlation between measured and imputed expression in the CC-RIX (x-axis) and eQTL LOD score. As expected, imputations are more accurate for transcripts with strong local eQTL. **C.** Variance explained by local genotype in the DO and CC-RIX.

![](data:application/pdf;base64,)

Regulation of *Nucb2* expression in islet. *Nucb2* is encoded on mouse chromosome 7 at 116.5 Mb (red line). In islets the heritability of *Nucb2* expression levels is 69% heritable. This LOD score trace shows that there is no local eQTL at that position, nor any strong distal eQTL anywhere else in the genome.

1. Maurano, M. T. *et al.* Systematic localization of common disease-associated variation in regulatory DNA. *Science* **337**, 1190–1195 (2012).

2. Farh, K. K. *et al.* Genetic and epigenetic fine mapping of causal autoimmune disease variants. *Nature* **518**, 337–343 (2015).

3. Pennisi, E. The Biology of Genomes. Disease risk links to gene regulation. *Science* **332**, 1031 (2011).

4. Hindorff, L. A. *et al.* Potential etiologic and functional implications of genome-wide association loci for human diseases and traits. *Proc Natl Acad Sci* **106**, 9362–9367 (2009).

5. Pickrell, J. K. Joint analysis of functional genomic data and genome-wide association studies of 18 human traits. *Am J Hum Genet* **94**, 559–573 (2014).

6. Welter, D. *et al.* The NHGRI GWAS Catalog, a curated resource of SNP-trait associations. *Nucleic Acids Res* **42**, D1001–1006 (2014).

7. Li, Y. I. *et al.* RNA splicing is a primary link between genetic variation and disease. *Science* **352**, 600–604 (2016).

8. Zhou, D. *et al.* A unified framework for joint-tissue transcriptome-wide association and Mendelian randomization analysis. *Nat Genet* **52**, 1239–1246 (2020).

9. Gamazon, E. R. *et al.* A gene-based association method for mapping traits using reference transcriptome data. *Nat Genet* **47**, 1091–1098 (2015).

10. Zhu, Z. *et al.* Integration of summary data from GWAS and eQTL studies predicts complex trait gene targets. *Nat Genet* **48**, 481–487 (2016).

11. Gusev, A. *et al.* Integrative approaches for large-scale transcriptome-wide association studies. *Nat Genet* **48**, 245–252 (2016).

12. Keller, M. P. *et al.* Genetic Drivers of Pancreatic Islet Function. *Genetics* **209**, 335–356 (2018).

13. Crouse, W. L., Keele, G. R., Gastonguay, M. S., Churchill, G. A. & Valdar, W. A Bayesian model selection approach to mediation analysis. *PLoS Genet* **18**, e1010184 (2022).

14. Chick, J. M. *et al.* Defining the consequences of genetic variation on a proteome-wide scale. *Nature* **534**, 500–505 (2016).

15. Wheeler, H. E. *et al.* Imputed gene associations identify replicable trans-acting genes enriched in transcription pathways and complex traits. *Genet Epidemiol* **43**, 596–608 (2019).

16. Umans, B. D., Battle, A. & Gilad, Y. Where Are the Disease-Associated eQTLs? *Trends Genet* **37**, 109–124 (2021).

17. Connally, N. J. *et al.* The missing link between genetic association and regulatory function. *Elife* **11**, (2022).

18. Mostafavi, H., Spence, J. P., Naqvi, S. & Pritchard, J. K. Systematic differences in discovery of genetic effects on gene expression and complex traits. *Nat Genet* **55**, 1866–1875 (2023).

19. Yao, D. W., O’Connor, L. J., Price, A. L. & Gusev, A. Quantifying genetic effects on disease mediated by assayed gene expression levels. *Nat Genet* **52**, 626–633 (2020).

20. Liu, X. *et al.* GBAT: a gene-based association test for robust detection of trans-gene regulation. *Genome Biol* **21**, 211 (2020).

21. Churchill, G. A., Gatti, D. M., Munger, S. C. & Svenson, K. L. The Diversity Outbred mouse population. *Mamm Genome* **23**, 713–718 (2012).

22. Chesler, E. J. *et al.* The collaborative cross at oak ridge national laboratory: Developing a powerful resource for systems genetics. *Mammalian Genome* **19**, 382–389 (2008).

23. Saul, M. C., Philip, V. M., Reinholdt, L. G. & Chesler, E. J. High-diversity mouse populations for complex traits. *Trends in Genetics* **35**, 501–514 (2019).

24. Clee, S. M. & Attie, A. D. The genetic landscape of type 2 diabetes in mice. *Endocr Rev* **28**, 48–83 (2007).

25. Broman, K. W. *et al.* R/qtl2: Software for Mapping Quantitative Trait Loci with High-Dimensional Data and Multiparent Populations. *Genetics* **211**, 495–502 (2019).

26. Girka, F. *et al.* [*RGCCA: Regularized and Sparse Generalized Canonical Correlation Analysis for Multiblock Data*](https://CRAN.R-project.org/package=RGCCA). (2023).

27. Korotkevich, G., Sukhov, V. & Sergushichev, A. Fast gene set enrichment analysis. *bioRxiv* (2019) doi:[10.1101/060012](https://doi.org/10.1101/060012).

28. Subramanian, A. *et al.* Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci U S A* **102**, 15545–15550 (2005).

29. Newgard, C. B. Interplay between lipids and branched-chain amino acids in development of insulin resistance. *Cell Metab* **15**, 606–614 (2012).

30. Sears, D. D. *et al.* Mechanisms of human insulin resistance and thiazolidinedione-mediated insulin sensitization. *Proc Natl Acad Sci U S A* **106**, 18745–18750 (2009).

31. Stienstra, R., Duval, C., ller, M. & Kersten, S. PPARs, Obesity, and Inflammation. *PPAR Res* **2007**, 95974 (2007).

32. Gavrilova, O. *et al.* Liver peroxisome proliferator-activated receptor gamma contributes to hepatic steatosis, triglyceride clearance, and regulation of body fat mass. *J Biol Chem* **278**, 34268–34276 (2003).

33. Matsusue, K. *et al.* Liver-specific disruption of PPARgamma in leptin-deficient mice improves fatty liver but aggravates diabetic phenotypes. *J Clin Invest* **111**, 737–747 (2003).

34. Patsouris, D., Reddy, J. K., ller, M. & Kersten, S. Peroxisome proliferator-activated receptor alpha mediates the effects of high-fat diet on hepatic gene expression. *Endocrinology* **147**, 1508–1516 (2006).

35. Schadinger, S. E., Bucher, N. L., Schreiber, B. M. & Farmer, S. R. PPARgamma2 regulates lipogenesis and lipid accumulation in steatotic hepatocytes. *Am J Physiol Endocrinol Metab* **288**, E1195–1205 (2005).

36. Motomura, W. *et al.* Up-regulation of ADRP in fatty liver in human and liver steatosis in mice fed with high fat diet. *Biochem Biophys Res Commun* **340**, 1111–1118 (2006).

37. Lamb, J. *et al.* The Connectivity Map: using gene-expression signatures to connect small molecules, genes, and disease. *Science* **313**, 1929–1935 (2006).

38. Amin, S., Lux, A. & O’Callaghan, F. The journey of metformin from glycaemic control to mTOR inhibition and the suppression of tumour growth. *Br J Clin Pharmacol* **85**, 37–46 (2019).

39. Kezic, A., Popovic, L. & Lalic, K. mTOR Inhibitor Therapy and Metabolic Consequences: Where Do We Stand? *Oxid Med Cell Longev* **2018**, 2640342 (2018).

40. Barlow, A. D., Nicholson, M. L. & Herbert, T. P. -cells and a review of the underlying molecular mechanisms. *Diabetes* **62**, 2674–2682 (2013).

41. Gu, Y., Lindner, J., Kumar, A., Yuan, W. & Magnuson, M. A. Rictor/mTORC2 is essential for maintaining a balance between beta-cell proliferation and cell size. *Diabetes* **60**, 827–837 (2011).

42. Geer, E. B., Islam, J. & Buettner, C. Mechanisms of glucocorticoid-induced insulin resistance: focus on adipose tissue function and lipid metabolism. *Endocrinol Metab Clin North Am* **43**, 75–102 (2014).

43. Li, J. X. & Cummins, C. L. Fresh insights into glucocorticoid-induced diabetes mellitus and new therapeutic directions. *Nat Rev Endocrinol* **18**, 540–557 (2022).

44. Lee, R. A., Harris, C. A. & Wang, J. C. Glucocorticoid Receptor and Adipocyte Biology. *Nucl Receptor Res* **5**, (2018).

45. Viengchareun, S., Penfornis, P., Zennaro, M. C. & s, M. Mineralocorticoid and glucocorticoid receptors inhibit UCP expression and function in brown adipocytes. *Am J Physiol Endocrinol Metab* **280**, E640–649 (2001).

46. Liu, J. *et al.* -HSD1 in regulating brown adipocyte function. *J Mol Endocrinol* **50**, 103–113 (2013).

47. Ramage, L. E. *et al.* Glucocorticoids Acutely Increase Brown Adipose Tissue Activity in Humans, Revealing Species-Specific Differences in UCP-1 Regulation. *Cell Metab* **24**, 130–141 (2016).

48. Barclay, J. L. *et al.* Effects of glucocorticoids on human brown adipocytes. *J Endocrinol* **224**, 139–147 (2015).

49. ó, M., Gupte, R., Kraus, W. L., Pacher, P. & Bai, P. PARPs in lipid metabolism and related diseases. *Prog Lipid Res* **84**, 101117 (2021).

50. Bai, P. *et al.* PARP-1 inhibition increases mitochondrial metabolism through SIRT1 activation. *Cell Metab* **13**, 461–468 (2011).

51. Chiarugi, A. & Moskowitz, M. A. Cell biology. PARP-1–a perpetrator of apoptotic cell death? *Science* **297**, 200–201 (2002).

52. Althubiti, M. *et al.* BTK targeting suppresses inflammatory genes and ameliorates insulin resistance. *Eur Cytokine Netw* **31**, 168–179 (2020).

53. Skrabs, C., Pickl, W. F., Perkmann, T., ger, U. & Gessl, A. Rapid decline in insulin antibodies and glutamic acid decarboxylase autoantibodies with ibrutinib therapy of chronic lymphocytic leukaemia. *J Clin Pharm Ther* **43**, 145–149 (2018).

54. Oral, E. A. *et al.* and TBK1 Improves Glucose Control in a Subset of Patients with Type 2 Diabetes. *Cell Metab* **26**, 157–170 (2017).

55. Arkan, M. C. *et al.* IKK-beta links inflammation to obesity-induced insulin resistance. *Nat Med* **11**, 191–198 (2005).

56. Clark, M., Kroger, C. J., Ke, Q. & Tisch, R. M. The Role of T Cell Receptor Signaling in the Development of Type 1 Diabetes. *Front Immunol* **11**, 615371 (2020).

57. ndell, P. *et al.* TNFR1 is associated with short-term mortality in patients with diabetes and acute dyspnea seeking care at the emergency department. *Acta Diabetol* **57**, 1145–1150 (2020).

58. Keestra-Gounder, A. M. *et al.* NOD1 and NOD2 signalling links ER stress with inflammation. *Nature* **532**, 394–397 (2016).

59. Keestra-Gounder, A. M. & Tsolis, R. M. NOD1 and NOD2: Beyond Peptidoglycan Sensing. *Trends Immunol* **38**, 758–767 (2017).

60. Montane, J., Cadavez, L. & Novials, A. Stress and the inflammatory process: a major cause of pancreatic cell death in type 2 diabetes. *Diabetes Metab Syndr Obes* **7**, 25–34 (2014).

61. Kahn, B. B. & McGraw, T. E., and type 2 diabetes. *N Engl J Med* **363**, 2667–2669 (2010).

62. Del Prato, S. & Pulizzi, N. The place of sulfonylureas in the therapy for type 2 diabetes mellitus. *Metabolism* **55**, S20–27 (2006).

63. Liu, X., Li, Y. I. & Pritchard, J. K. Trans Effects on Gene Expression Can Drive Omnigenic Inheritance. *Cell* **177**, 1022–1034 (2019).

64. Hallgrimsson, B. *et al.* The developmental-genetics of canalization. *Semin Cell Dev Biol* **88**, 67–79 (2019).

65. Siegal, M. L. & Bergman, A. Waddington’s canalization revisited: developmental stability and evolution. *Proc Natl Acad Sci U S A* **99**, 10528–10532 (2002).

66. Paaby, A. B. & Gibson, G. Cryptic Genetic Variation in Evolutionary Developmental Genetics. *Biology (Basel)* **5**, (2016).

67. Boyle, E. A., Li, Y. I. & Pritchard, J. K. An Expanded View of Complex Traits: From Polygenic to Omnigenic. *Cell* **169**, 1177–1186 (2017).

68. Wray, N. R., Wijmenga, C., Sullivan, P. F., Yang, J. & Visscher, P. M. Common disease is more complex than implied by the core gene omnigenic model. *Cell* **173**, 1573–1580 (2018).

69. Riva, M. *et al.* Nesfatin-1 stimulates glucagon and insulin secretion and beta cell NUCB2 is reduced in human type 2 diabetic subjects. *Cell Tissue Res* **346**, 393–405 (2011).

70. Nakata, M. & Yada, T. Role of NUCB2/nesfatin-1 in glucose control: diverse functions in islets, adipocytes and brain. *Curr Pharm Des* **19**, 6960–6965 (2013).

71. Shimizu, H. & Osaki, A. Nesfatin/Nucleobindin-2 (NUCB2) and Glucose Homeostasis. *Curr Hypertens Rev* Nesfatin/Nucleobindin–2 (NUCB2) and Glucose Homeostasis. (2014).