

¹ Transcripts with high distal heritability mediate genetic effects on
² complex traits

³

⁴ Anna L. Tyler, J. Matthew Mahoney, Mark P. Keller, Candice N. Baker, Margaret Gaca, Anuj Srivastava,
⁵ Isabela Gerdes Gyuricza, Madeleine Braun, Nadia A. Rosenthal, Alan D. Attie, Gary A. Churchill and
⁶ Gregory W. Carter

⁷ **Abstract**

⁸ Although many genes are subject to local regulation, recent evidence suggests that complex distal regulation
⁹ may be more important in mediating phenotypic variability. To assess the role of distal gene regulation in
¹⁰ complex traits, we combined multi-tissue transcriptomes with physiological outcomes to model diet-induced
¹¹ obesity and metabolic disease in a population of 371 Diversity Outbred mice. Using a novel high-dimensional
¹² mediation analysis, we identified a composite transcriptome signature that summarized genetic effects on
¹³ gene expression and explained 30% of the variation across all metabolic traits. The signature was heritable,
¹⁴ interpretable in biological terms, and predicted obesity status from gene expression in an independently
¹⁵ derived mouse cohort and multiple human studies. Transcripts contributing most strongly to this composite
¹⁶ mediator frequently had complex, distal regulation distributed throughout the genome. These results suggest
¹⁷ that trait-relevant variation in transcription is largely distally regulated, but is nonetheless identifiable,
¹⁸ interpretable, and translatable across species.

¹⁹ **Introduction**

²⁰ In the quest to understand the genetic architecture of complex traits, gene expression is an important
²¹ mediator between genotype and phenotype. There is ample evidence from genome-wide association studies
²² (GWAS) that regulation of gene expression accounts for the bulk of the genetic effect on complex traits, as
²³ most trait-associated variants lie in gene regulatory regions^{1–7}. It is widely assumed that these variants
²⁴ influence local transcription, and methods such as transcriptome-wide association studies (TWAS)^{8–11} and
²⁵ summary data-based Mendelian randomization (SMR)¹⁰ capitalize on this idea to identify genes associated

²⁶ with multiple disease traits^{12–15}

²⁷ Despite the great promise of these methods, explaining trait effects with local gene regulation has been more
²⁸ difficult than initially assumed^{16;17}. Although trait-associated variants tend to lie in non-coding, regulatory
²⁹ regions, they often do not have detectable effects on gene expression¹⁶ and tend not to co-localize with
³⁰ expression quantitative trait loci (eQTLs)^{17;18}.

³¹ One possible explanation for these observations is that gene expression is not being measured in the appropriate
³² cell types and thus local eQTLs influencing traits cannot be detected¹⁶. An alternative explanation that has
³³ been discussed in recent years is that effects of these variants are mediated not through local regulation of
³⁴ gene expression, but through distal regulation^{18–20;15}. In this model, a gene's expression is influenced by
³⁵ many variants throughout the genome through their cumulative effects on a broader regulatory network. In
³⁶ other words, the heritable component of the transcriptome is an emergent state arising from the myriad
³⁷ molecular interactions defining and constraining gene expression.

³⁸ To assess the role of wide-spread distal gene regulation on complex traits, we investigated diet-induced
³⁹ obesity and metabolic disease as an archetypal example. Diet-induced obesity and metabolic disease are
⁴⁰ genetically complex with hundreds of variants mapped through GWAS^{21;22}. These variants are known to act
⁴¹ through multiple tissues that interact dynamically with each other^{23;24}, including adipose tissue, pancreatic
⁴² islets, liver, and skeletal muscle. The multi-system etiology of metabolic disease complicates mechanistic
⁴³ dissection of the genetic architecture, requiring large, dedicated data sets that include high-dimensional,
⁴⁴ clinically relevant phenotyping, dense genotyping in a highly recombined population, and transcriptome-wide
⁴⁵ measurements of gene expression in multiple tissues.

⁴⁶ Measuring gene expression in multiple tissues is critical to adequately assess the extent to which local gene
⁴⁷ regulation varies across the tissues and whether such variability might account for previous failed attempts to
⁴⁸ identify trait-relevant local eQTL. Such data sets are extremely difficult to obtain in human populations,
⁴⁹ particularly in the large numbers of subjects required for adequate statistical power. Thus, to further
⁵⁰ investigate the role of local and distal gene regulation on complex traits, we generated two complementary
⁵¹ data sets: A discovery data set in a large population of diversity outbred (DO) mice²⁵, and an independent
⁵² validation data set derived by crossing inbred strains from the Collaborative Cross (CC) mice²⁶ to form CC
⁵³ F1 mice (CC-RIX). Both populations modeled diet-induced obesity and metabolic disease¹².

⁵⁴ The DO population and CC recombinant inbred lines were derived from the same eight inbred founder mouse
⁵⁵ strains, five classical lab strains, and three strains more recently derived from wild mice²⁵. They represent
⁵⁶ three subspecies of mouse *Mus musculus domesticus*, *Mus musculus musculus*, and *Mus musculus castaneus*,

57 and capture 90% of the known variation in laboratory mice²⁷. The DO mice are maintained with a breeding
58 scheme that ensures equal contributions from each founder across the genome thus rendering almost the
59 whole genome visible to genetic inquiry²⁵. The CC mice were initially outcrossed to recombine the genomes
60 from all eight founders, and then inbred for at least 20 generations to generate multiple inbred lines. Because
61 these two populations have common ancestral haplotypes, we could directly and unambiguously compare
62 the local genetic effects on gene expression at the whole-transcriptome level while varying the population
63 structure driving distal regulation.

64 In the DO population, we paired clinically relevant metabolic traits from 371 mice¹², including body weight,
65 plasma levels of insulin, glucose and lipids, with transcriptome-wide gene expression in four tissues related to
66 metabolic disease: adipose tissue, pancreatic islets, liver, and skeletal muscle. We measured similar metabolic
67 traits in a CC-RIX population and gene expression from three of the four tissues used in the DO: adipose
68 tissue, liver, and skeletal muscle. Because the CC-RIX carry the same founder alleles as the DO, local gene
69 regulation is expected to match between the populations, but because the alleles are recombined through
70 the genome, distal effects are expected to vary from those in the DO, allowing us to directly assess the
71 role of local gene regulation in driving trait-associated transcript variation. Together, these data enable a
72 comprehensive view into the genetic architecture of metabolic disease.

73 Results

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

76 Genetic variation contributed to wide phenotypic variation

77 Although the environment was consistent across the DO mice, the genetic diversity present in this population
78 resulted in widely varying distributions across physiological measurements (Fig. 1). For example, body
79 weights of adult individuals varied from less than the average adult C57BL/6J (B6) body weight to several
80 times the body weight of a B6 adult in both sexes (Males: 18.5 - 69.1g, Females: 16.0 - 54.8g) (Fig. 1A).
81 Fasting blood glucose (FBG) also varied considerably (Fig. 1B), although few of the animals had FBG levels
82 that would indicate pre-diabetes (19 animals, 3.8%), or diabetes (7 animals, 1.4%) according to previously
83 developed cutoffs (pre-diabetes: FBG \geq 250 mg/dL, diabetes: FBG \geq 300, mg/dL)²⁸. Males had higher
84 FBG than females on average (Fig. 1C) as has been observed before suggesting either that males were more
85 susceptible to metabolic disease on the high-fat, high-sugar (HFHS) diet, or that males and females may
86 require different thresholds for pre-diabetes and diabetes.

87 Body weight was strongly positively correlated with food consumption (Fig. 1D $R^2 = 0.51$, $p < 2.2 \times 10^{-16}$)
88 and FBG (Fig. 1E, $R^2 = 0.21$, $p < 2.2 \times 10^{-16}$) suggesting a link between behavioral factors and metabolic
89 disease. However, the heritability of this trait and others (Fig. 1F) indicates that genetics contribute
90 substantially to correlates of metabolic disease in this population.

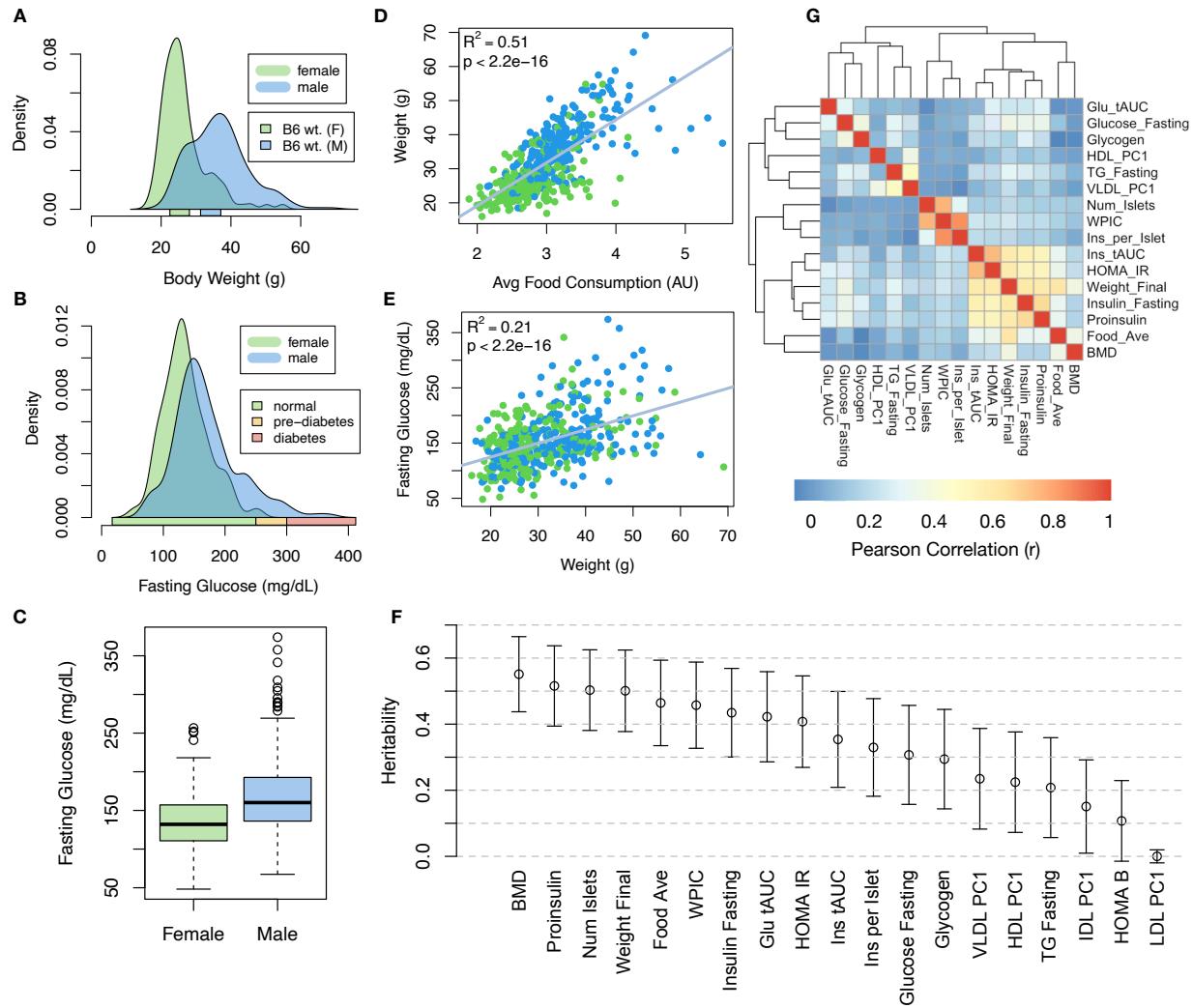


Figure 1: Clinical overview. **A.** Distributions of body weight in the diversity outbred mice. Sex is indicated by color. The average B6 male and female adult weights at 24 weeks of age are indicated by blue and green bars on the x-axis. **B.** The distribution of fasting glucose across the population 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. BMD - bone mineral density, WPIC - whole pancreas insulin content, Glu tAUC - glucose total area under the curve, HOMA IR - homeostatic measurement of insulin resistance, HOMA B - homeostatic measure of beta cell health, VLDL - very low-density lipoprotein, LDL - low-density lipoprotein, IDL - intermediate density lipoprotein, HDL - high-density lipoprotein, TG - triglyceride.

91 **Distal Heritability Correlated with Phenotype Relevance**

92 To comprehensively assess the genetic control of gene expression in metabolic disease we measured overall
93 gene expression via bulk RNA-Seq in adipose, islet, liver, and skeletal muscle in the DO cohort (Supp. Fig.
94 S1A-H). We performed eQTL analysis using R/qltl2²⁹ (Methods) and identified both local and distal eQTLs
95 for transcripts in each of the four tissues (Supp. Fig. S1). Significant local eQTLs far outnumbered distal
96 eQTLs (Supp. Fig. S1F) and tended to be shared across tissues (Supp. Fig. S1G) whereas the few significant
97 distal eQTLs we identified tended to be tissue-specific (Supp. Fig. S1H)

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

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

113 **High-Dimensional Mediation identified a high-heritability composite trait that was mediated
114 by a composite transcript**

115 The above univariate analyses establish the importance of distal heritability for trait-relevant transcripts.
116 However, the number of transcripts dramatically exceeds the number of phenotypes. Thus, we expect the
117 heritable, trait-relevant transcripts to be highly correlated and organized according to coherent, biological
118 processes representing the mediating endophenotypes driving clinical trait variation. To identify these
119 endophenotypes in a theoretically principled way, we developed a novel dimension-reduction technique,
120 high-dimension mediation analytis (HDMA), that uses the theory of causal graphical models to identify a
121 transcriptomic signature that is simultaneously 1) highly heritable, 2) strongly correlated to the measured

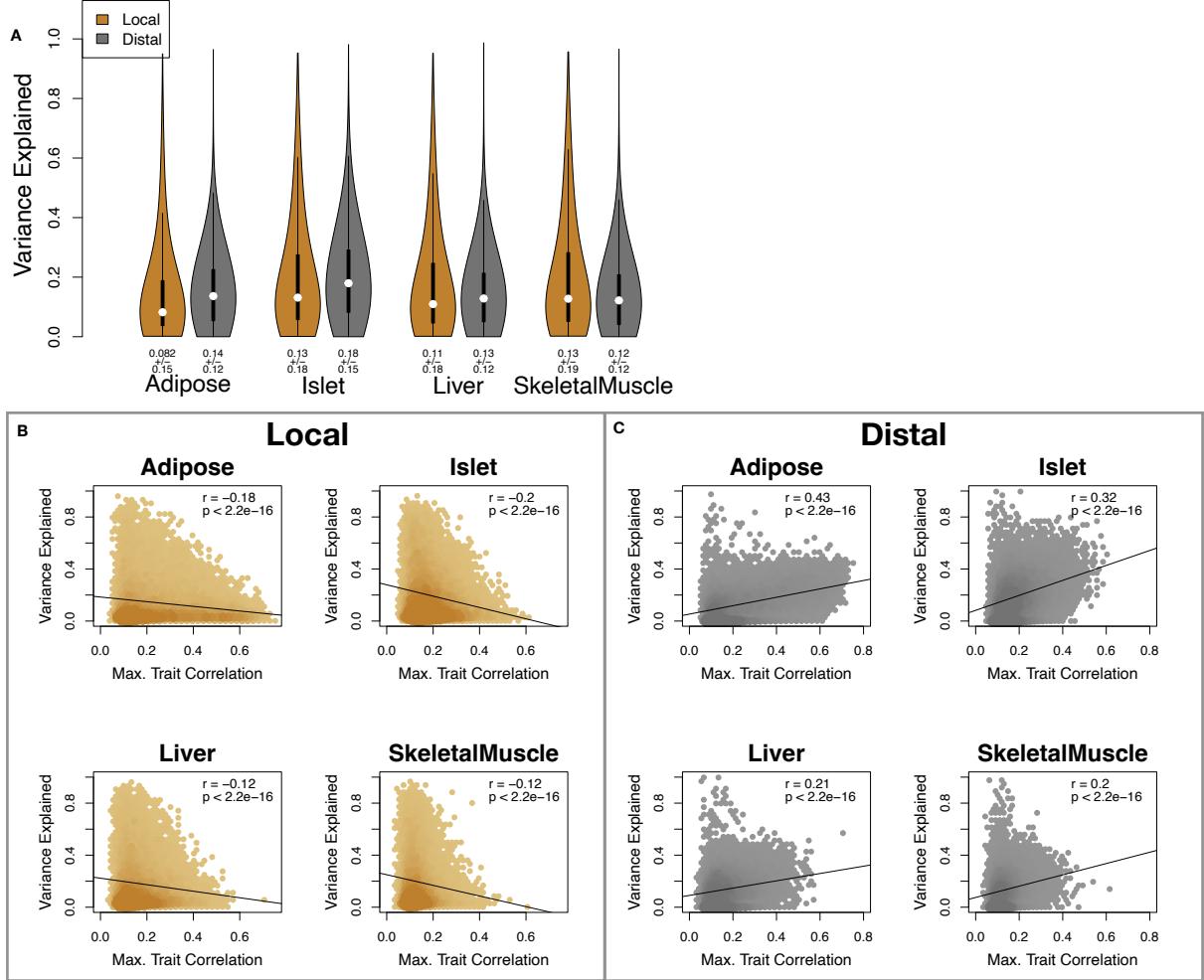


Figure 2: Transcript heritability and trait relevance. **A.** Distributions of distal and local heritability of transcripts across the four tissues. Overall local and distal factors contribute equally to transcript heritability. The relationship between **(B.)** local and **(C.)** distal heritability 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 (r) and p values for each correlation are shown in the upper-right of each panel.

phenotypes, and 3) conforms to the causal mediation hypothesis (Fig. 3). HDMA projects the high-dimensional scores—a composite genome score (G_C), a composite transcriptome score (T_C), and a composite phenome score (P_C)—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

$$\text{Corr}(G_C, P_C) = \text{Corr}(G_C, T_C)\text{Corr}(T_C, P_C)$$

which is equivalent to the partial correlation of G_C and P_C after controlling for T_C being zero. The value $\text{Corr}(G_C, T_C)\text{Corr}(T_C, P_C)$ 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 developed³⁰. A complete mathematical derivation and implementation details for HDMA are available in Supp. Methods.

Using HDMA we identified the major axis of variation in the transcriptome that was consistent with mediating the effects of the genome on metabolic traits (Fig 3). Fig. 3A shows the partial correlations (ρ) between the pairs of these composite vectors. The partial correlation between G_C and T_C was 0.42, and the partial correlation between T_C and P_C was 0.78. However, when the transcriptome was taken into account, the partial correlation between G_C and P_C was effectively zero (0.039). P_C 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. 1F). 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, 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 vectors³¹. 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 G_C and T_C multiplied by the partial correlation of T_C and P_C . This represents the strength of the path from G_C to P_C that is putatively mediated through T_C . The null distribution of the path coefficient is shown in Fig. 3B, 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 ($p < 10^{-16}$). Fig. 3C illustrates this observation in more detail. Although we identified high correlations between G_C and T_C , and modest correlations between T_C and P_C in the null data (Fig 3C), 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 G_C - T_C correlation and the T_C - P_C 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

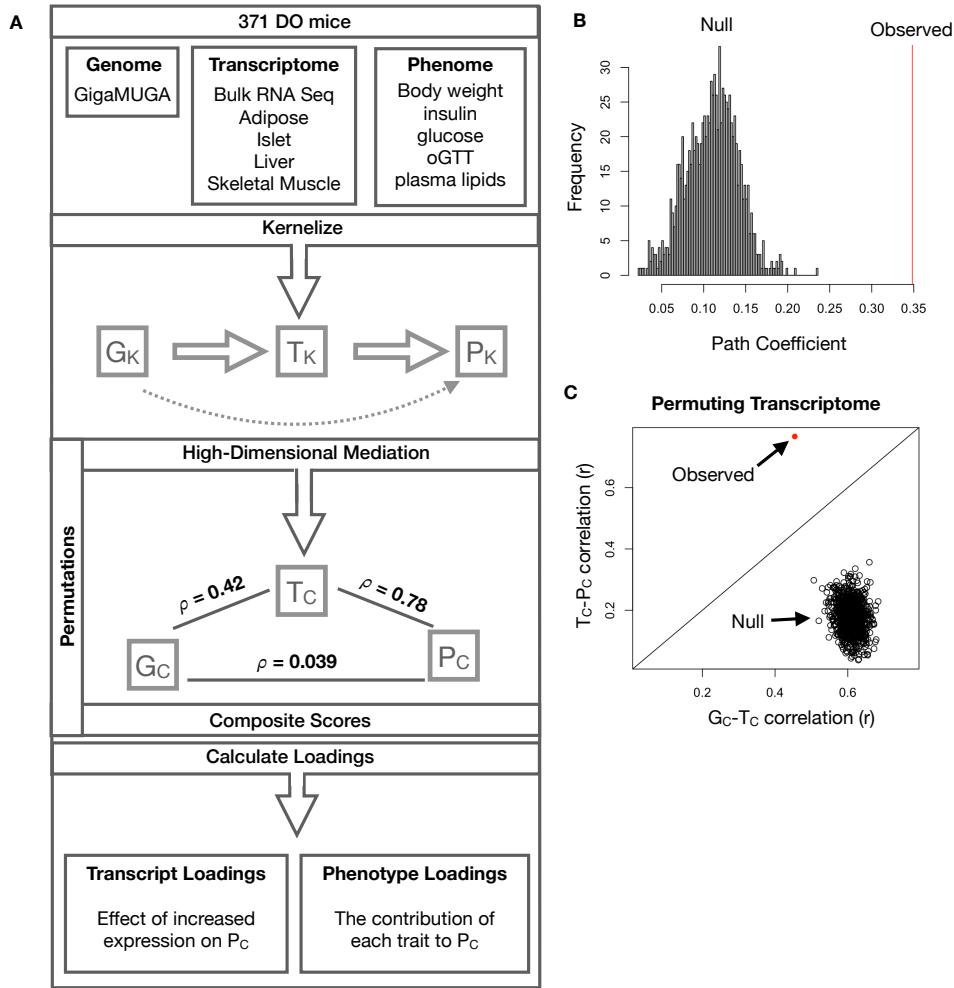


Figure 3: High-dimensional mediation. **A.** Workflow indicating major steps of high-dimensional mediation. The genotype, transcriptome, and phenotype matrices were independently normalized and converted to kernel matrices representing the pairwise relationships between individuals for each data modality (K_G = genome kernel, K_T = transcriptome kernel; K_P = phenotype kernel). High-dimensional mediation was applied to these matrices to maximize the direct path $G \rightarrow T \rightarrow P$, the mediating pathway (arrows), while simultaneously minimizing the direct $G \rightarrow P$ pathway (dotted line). The composite vectors that resulted from high-dimensional mediation were G_c , T_c , and P_c . The partial correlations ρ 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 G_c - T_c correlation vs. the T_c - P_c correlation compared with the observed value (red dot).

160 genetically determined variation in phenotype that is mediated through genetically determined variation in
 161 transcription.

162 **Body weight and insulin resistance were highly represented in the expression-mediated com-**
163 **posite trait**

164 Each composite score is simply a weighted combination of the measured variables and the magnitude and
165 sign of the weights, called loadings, correspond the relative importance and directionality of each variable
166 in the composite score. The loadings of each measured trait onto P_C indicate how much each contributed
167 to the composite phenotype. Body weight contributed the most (Fig. 4), followed by homeostatic insulin
168 resistance (HOMA_IR) and fasting plasma insulin levels (Insulin_Fasting). We can thus interpret P_C as an
169 index of metabolic disease (Fig. 4B). Individuals with high values of P_C have a higher metabolic disease
170 index and greater metabolic disease, including higher body weight and higher insulin resistance. We refer
171 to P_C as the metabolic disease index (MDI) going forward. Traits contributing the least to the MDI were
172 measures of cholesterol and pancreas composition. Thus, when we interpret the transcriptomic signature
173 identified by HDMA, we are explaining primarily the putative transcriptional mediation of body weight and
174 insulin resistance, as opposed to cholesterol measurements.

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

177 We interpreted large loadings onto transcripts as indicating strong mediation of the effect of genetics on MDI.
178 Large positive loadings indicate that higher expression was associated with a higher MDI (i.e. higher risk of
179 obesity and metabolic disease on the HFHS diet) (Fig. 4C). Conversely, large negative loadings indicate that
180 high expression of these transcripts was associated with a lower MDI (i.e. lower risk of obesity and metabolic
181 disease on the HFHS diet) (Fig. 4C). We used gene set enrichment analysis (GSEA)^{32;33} to look for biological
182 processes and pathways that were enriched at the top and bottom of this list (Methods).

183 In adipose tissue, both GO processes and KEGG pathway enrichments pointed to an axis of inflammation
184 and metabolism (Figs. S2 and S3). GO terms and KEGG pathways associated with inflammation were
185 positively associated with MDI, indicating that increased expression in inflammatory pathways was associated
186 with a higher MDI. It is well established that adipose tissue in obese individuals is inflamed and infiltrated
187 by macrophages^{34–38}, and the results here suggest that this may be a dominant heritable component of
188 metabolic disease.

189 The strongest negative enrichments in adipose tissue were related to mitochondrial activity in general, and
190 thermogenesis in particular (Figs. S2 and S2). Genes in the KEGG oxidative phosphorylation pathway in
191 mice were almost universally negatively loaded in adipose tissue, suggesting that increased expression of
192 these genes was associated with reduced MDI (Supp. Fig. S4). Consistent with this observations, it has been

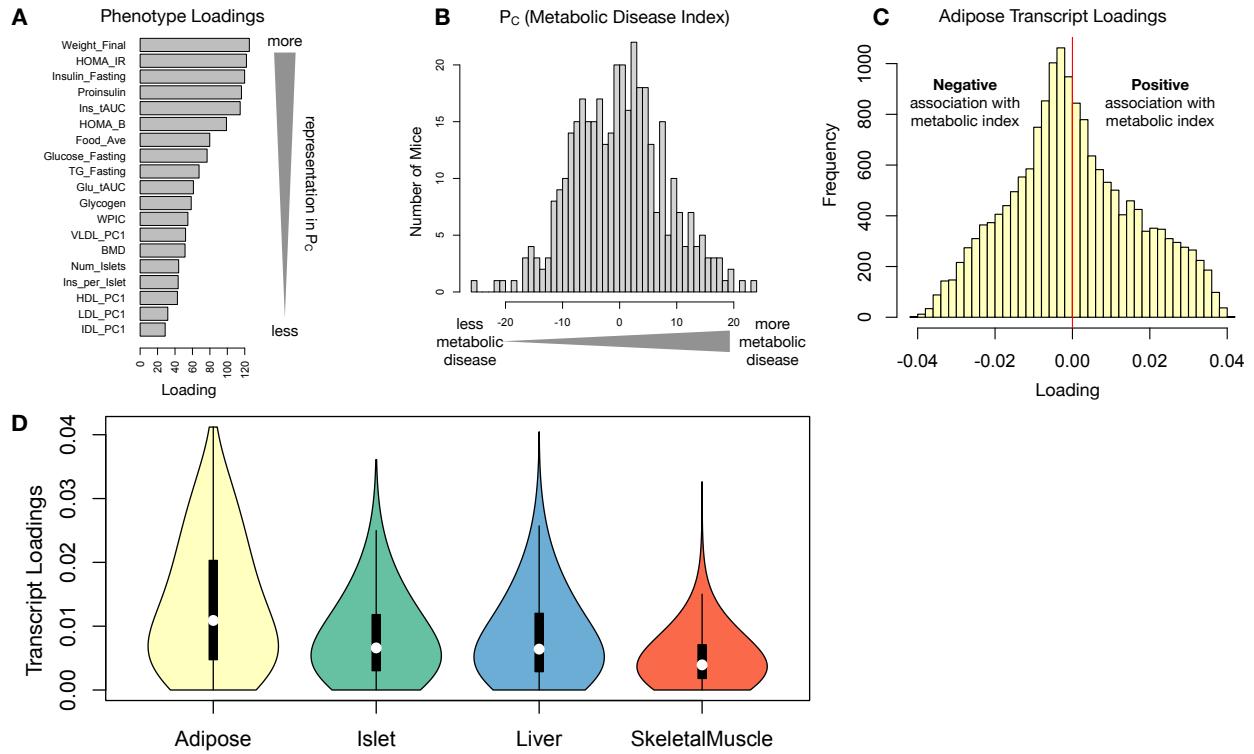


Figure 4: Interpretation of loadings. **A.** Loadings across traits. Body weight and insulin resistance contributed the most to the composite trait. **B.** Phenotype scores across individuals. Individuals with large positive phenotype scores had higher body weight and insulin resistance than average. 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.

193 shown previously that mouse strains with greater thermogenic potential are also less susceptible to obesity
194 on a HFHS diet³⁹.

195 Transcripts associated with the citric acid (TCA) cycle as well as the catabolism of the branched-chain amino
196 acids (BCAA) (valine, leucine, and isoleucine) were strongly enriched with negative loadings in adipose
197 tissue (Supp. Figs. S2, S5 and S6). Expression of genes in both pathways (for which there is some overlap)
198 has been previously associated with insulin sensitivity^{12;40;41}, suggesting that heritable variation in regulation
199 of these pathways may influence risk of insulin resistance.

200 Looking at the 10 most positively and negatively loaded transcripts from each tissue, it is apparent that
201 transcripts in the adipose tissue had the largest loadings, both positive and negative, of all tissues (Fig. 5A
202 bar plot) This suggests that much of the effect of genetics on body weight and insulin resistance is mediated
203 through gene expression in adipose tissue. The strongest loadings in liver and pancreas were comparable,

204 and those in skeletal muscle were the weakest (Fig. 5A), suggesting that less of the genetic effects were
205 mediated through transcription in skeletal muscle. Heritability analysis showed that transcripts with the
206 largest loadings had higher distal heritability than local heritability (Fig. 5A heat map and box plot). This
207 pattern contrasts with transcripts nominated by TWAS (Fig. 5B), which tended to have lower loadings,
208 higher local heritability and lower distal heritability. Transcripts with the highest local heritability in each
209 tissue (Fig. 5C) had the lowest loadings, consistent with our findings above (Fig. 2B).

210 We performed a literature search for the genes in each of these groups along with the terms “diabetes”,
211 “obesity”, and the name of the expressing tissue to determine whether any of these genes had previous
212 associations with metabolic disease in the literature (Methods). Multiple genes in each group had been
213 previously associated with obesity and diabetes (Fig. 5 bolded gene names). Genes with high loadings were
214 most highly enriched for previous literature support. They were 2.4 times more likely than TWAS hits and 3.8
215 times more likely than genes with high local heritability to be previously associated with obesity or diabetes.

216 **Tissue-specific transcriptional programs were associated with metabolic traits**

217 Clustering of transcripts with top loadings in each tissue showed tissue-specific functional modules associated
218 with obesity and insulin resistance (Fig. 6A) (Methods). The clustering highlights the importance of immune
219 activation particularly in adipose tissue. The “mitosis” cluster had large positive loadings in three of the four
220 tissues potentially suggesting system-wide proliferation of immune cells. Otherwise, all clusters were strongly
221 loaded in only one or two tissues. For example, the lipid metabolism cluster was loaded most heavily in liver.
222 The positive loadings suggest that high expression of these genes particularly in the liver was associated with
223 increased metabolic disease. This cluster included the gene *Pparg*, whose primary role is in the adipose tissue
224 where it is considered a master regulator of adipogenesis⁴². Agonists of *Pparg*, such as thiazolidinediones, are
225 FDA-approved to treat type II diabetes, and reduce inflammation and adipose hypertrophy⁴². Consistent
226 with this role, the loading for *Pparg* in adipose tissue was negative, suggesting that higher expression was
227 associated with leaner mice (Fig. 6B). In contrast, *Pparg* had a large positive loading in liver, where it is
228 known to play a role in the development of hepatic steatosis, or fatty liver. Mice that lack *Pparg* specifically
229 in the liver, are protected from developing steatosis and show reduced expression of lipogenic genes^{43;44}.
230 Overexpression of *Pparg* in the livers of mice with a *Ppara* knockout, causes upregulation of genes involved in
231 adipogenesis⁴⁵. In the livers of both mice and humans high *Pparg* expression is associated with hepatocytes
232 that accumulate large lipid droplets and have gene expression profiles similar to that of adipocytes^{46;47}.

233 The local and distal heritability of *Pparg* is low in adipose tissue suggesting its expression in this tissue is
234 highly constrained in the population (Fig. 6B). However, the distal heritability of *Pparg* in liver is relatively

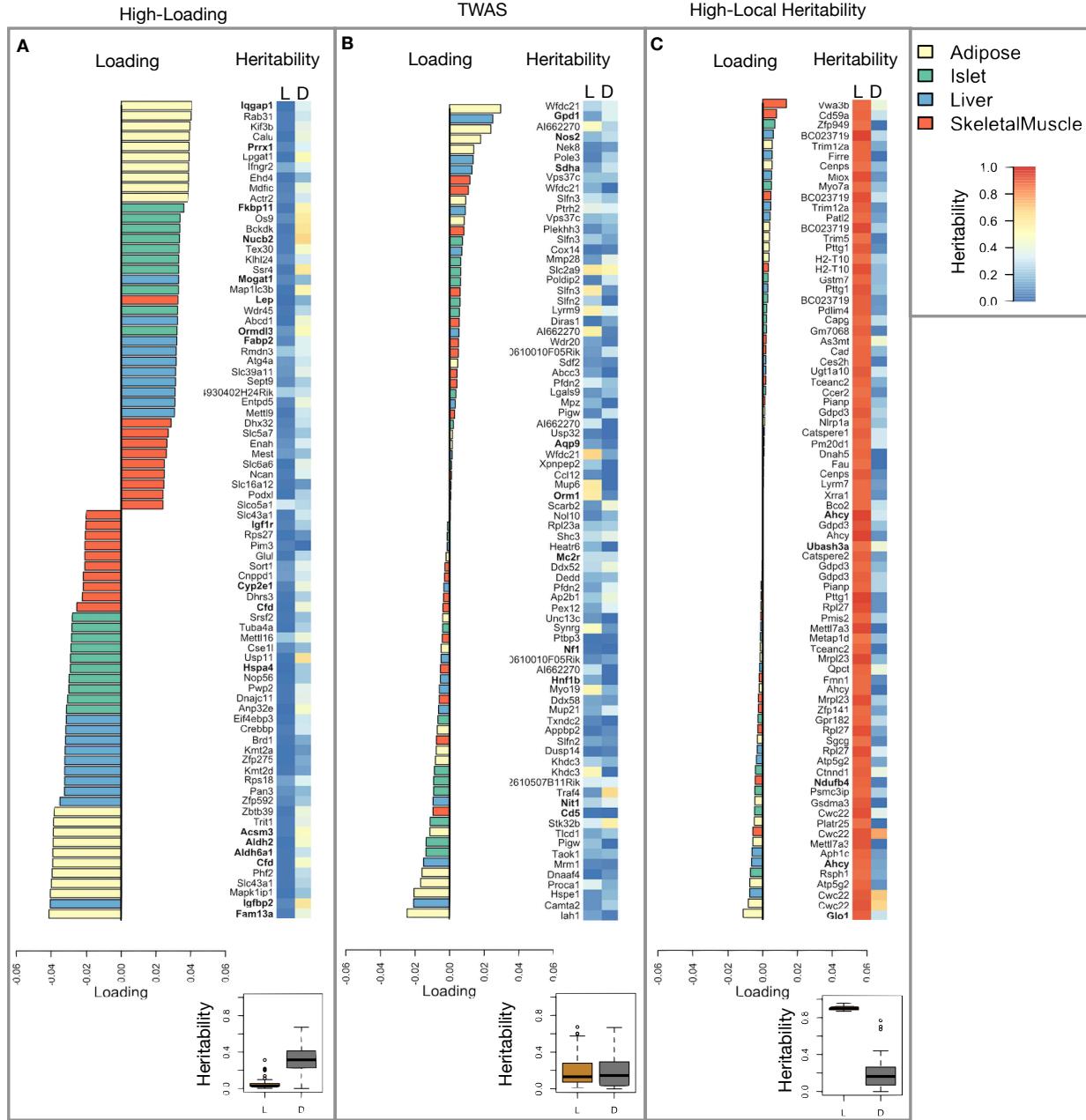


Figure 5: Transcripts with high loadings have high distal heritability and literature support. Each panel has a bar plot showing the loadings of transcripts selected by different criteria. Bar color indicates the tissue of origin. The heatmap shows the local (L - left) and distal (D - right) heritability of each transcript.

A

high suggesting it is complexly regulated and has sufficient variation in this population to drive variation in phenotype. Both local and distal heritability 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 MDI. These results highlight

238 the importance of tissue context when investigating the role of heritable transcript variability in driving
 239 phenotype.

240 Gene lists for all clusters are available in Supp. File 1.

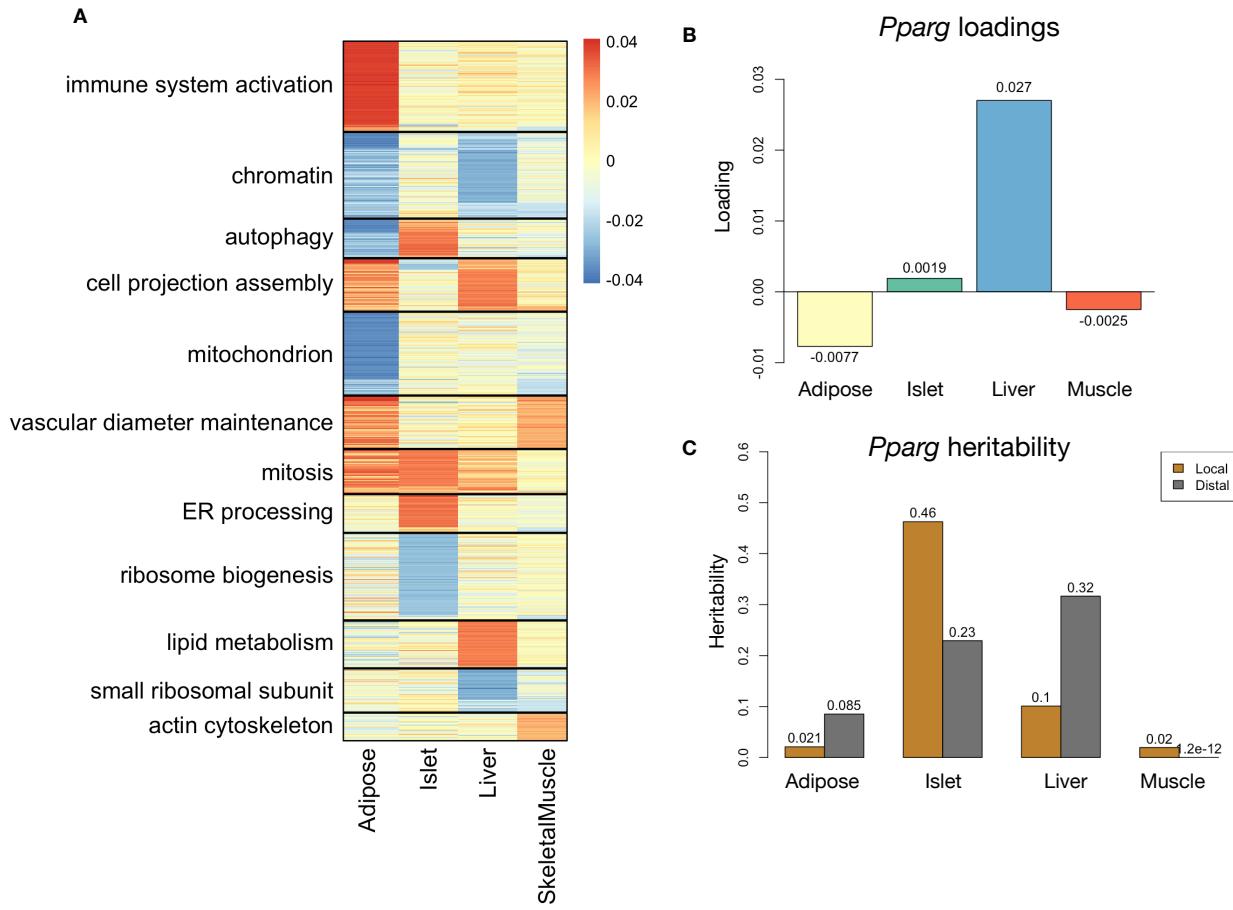


Figure 6: Tissue-specific transcriptional programs were associated with obesity and insulin resistance. **A** Heat map showing the loadings of all transcripts with loadings greater than 2.5 standard deviations from the mean in any tissue. The heat map was clustered using k medoid clustering. Functional 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.

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

242 To test whether the transcript loadings identified in the DO could be translated to another population, we
 243 tested whether they could predict metabolic phenotype in an independent population of CC-RIX mice, which
 244 were F1 mice derived from multiple pairings of Collaborative Cross (CC)^{48–51} strains (Fig. 7) (Methods).
 245 We tested two questions. First, we asked whether the loadings identified in the DO mice were relevant to
 246 the relationship between the transcriptome and the phenotype in the CC-RIX. We predicted body weight
 247 (a surrogate for MDI) 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. 7B 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 MDI. This result also confirms the validity and translatability of the transcript loadings and their relationship to metabolic disease.

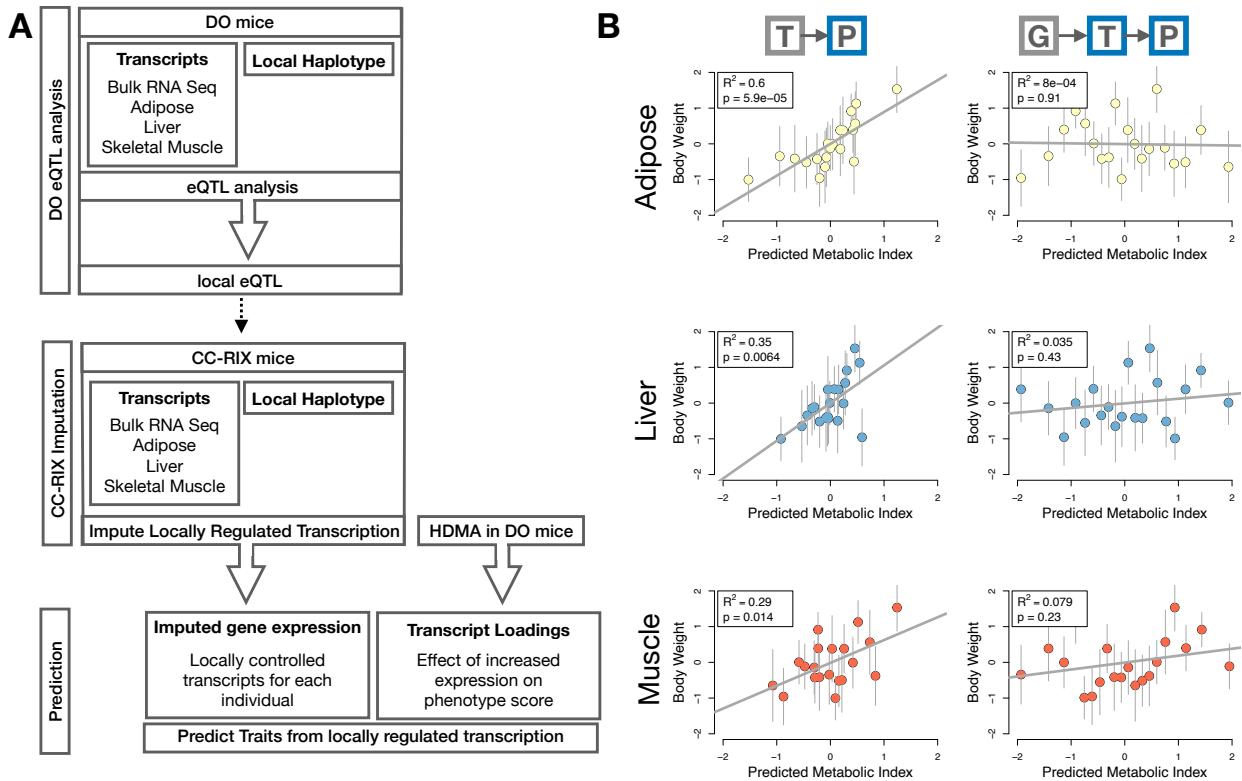


Figure 7: Transcription, but not local genotype, predicts phenotype in the CC-RIX. **A.** Workflow showing procedure for translating HDMA results to an independent population of mice. **B.** Relationships between the predicted metabolic disease index (MDI) and measured body weight. The left column shows 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. 7A). 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. S7). However, these imputed values failed to predict body weight in the CC-RIX when weighted with the transcript loadings from HDMA. (Fig. 7B right column). This result suggests that local regulation of gene expression is

260 not the primary factor driving heritability of complex traits, consistent with our findings in the DO population
261 that distal heritability was a major driver of trait-relevant variation and that high-loading transcripts had
262 comparatively high distal and low local heritability.

263 **Distally heritable transcriptomic signatures reflected variation in composition of adipose tissue**
264 **and islets**

265 The interpretation of global genetic influences on gene expression and phenotype is potentially more challenging
266 than the interpretation and translation of local genetic influences, as genetic effects cannot be localized to
267 individual gene variants or transcripts. However, there are global patterns across the loadings that can
268 inform mechanism. For example, heritable variation in cell type composition can be inferred from transcript
269 loadings. We observed above that immune activation in the adipose tissues was a highly enriched process
270 correlating with obesity in the DO population. For example, in humans, it has been extensively observed
271 that macrophage infiltration in adipose tissue is a marker of obesity and metabolic disease⁵². To determine
272 whether the immune activation reflected a heritable change in cell composition in adipose tissue in DO mice,
273 we compared loadings of cell-type specific genes in adipose tissue (Methods). Consistent with human results,
274 the mean loading of macrophage-specific genes was significantly greater than 0 (Fig. 8A), indicating that
275 obese mice were genetically predisposed to have high levels of macrophage infiltration in adipose tissue in
276 response to the HFHS diet. Loading for marker genes for other cell types were not statistically different from
277 zero, indicating that changes in the abundance of those cell types is not a mediator of MDI.

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

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

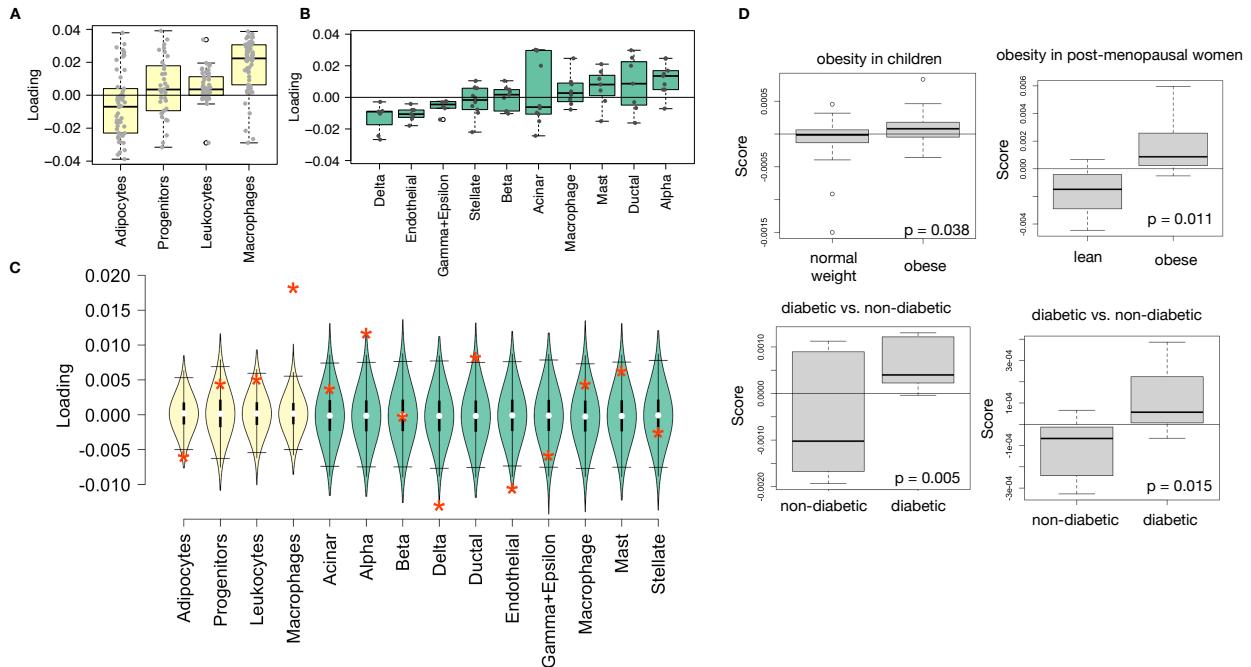


Figure 8: HDMA results translate to humans. **A.** Distribution of loadings for cell-type-specific transcripts in adipose tissue. **B.** Distribution of loadings for cell-type-specific transcripts in pancreatic islets (green). **C.** Null distributions for the mean loading of randomly selected transcripts 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.

290 Heritable transcriptomic signatures translated to human disease

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

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

302 **Existing therapies are predicted to target mediator gene signatures**

303 Another potential application of the transcript loading landscape is in ranking potential drug candidates
304 for the treatment of metabolic disease. Although high-loading transcripts may be good candidates for
305 understanding specific biology related to obesity, the transcriptome overall is highly interconnected and
306 redundant. The ConnectivityMap (CMAP) database⁵³ developed by the Broad Institute allows querying
307 thousands of compounds that reverse or enhance the extreme ends of transcriptomic signatures in multiple
308 different cell types. By identifying drugs that reverse pathogenic transcriptomic signatures, we can potentially
309 identify compounds that have favorable effects on gene expression.

310 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
311 (Supp. Figs S8 and S9). To get more tissue-specific results, we also looked at top results in cell types that
312 most closely resembled our tissues. We looked at results in adipocytes (ASC) as well as pancreatic tumor
313 cells (YAPC) regardless of *p* value (Supp. Figs S10 and S11).

315 Looking across all cell types, the notable top hits from the adipose tissue loadings included mTOR inhibitors
316 and glucocorticoid agonists (Supp. Fig. S8). It is thought that metformin, which is commonly used to
317 improve glycemic control, acts, at least in part, by inhibiting mTOR signaling^{54;55}. However, long-term use
318 of other mTOR inhibitors, such as rapamycin, are known to cause insulin resistance and β -cell toxicity^{55–57}.
319 Glucocorticoids are used to reduce inflammation, which was a prominent signature in the adipose tissues,
320 but these drugs also promote hyperglycemia and diabetes^{58;59}. Accute treatment with glucocorticoids has
321 further been shown to reduce thermogenesis in rodent adipocytes^{60–62}, but increase thermogenesis in human
322 adipocytes^{63;64}. Thus, the pathways identified by CMAP across all cell types were highly related to the
323 transcript loading profiles, but the relationship was not a simple reversal.

324 The top hit for the adipose composite transcript in CMAP adipocytes was a PARP inhibitor (Supp. Fig.
325 S10). PARPs play a role in lipid metabolism and are involved in the development of obesity and diabetes⁶⁵.
326 PARP1 inhibition increases mitochondrial biogenesis⁶⁶. Inhibition of PARP1 activity can further prevent
327 necrosis in favor of the less inflammatory apoptosis⁶⁷, thereby potentially reducing inflammation in stressed
328 adipocytes. Other notable hits among the top 20 were BTK inhibitors, which have been observed to suppress
329 inflammation and improve insulin resistance⁶⁸ as well as to reduce insulin antibodies in type I diabetes⁶⁹.
330 IkappaB kinase (IKK) is an enzyme complex involved in regulating cellular responses to inflammation⁷⁰.
331 Inhibitors of IKK have been shown to improve glucose control in type II diabetes^{71;72}.

332 Among the top most significant hits for the transcript loadings from pancreatic islets (Supp. Fig. S9), was

333 suppression of T cell receptor signaling, which is known to be involved in Type 1 diabetes⁷³, as well as
334 TNFR1, which has been associated with mortality in diabetes patients⁷⁴. Suppression of NOD1/2 signaling
335 was also among the top hits. NOD1 and 2 sense ER stress^{75;76}, which is associated with β -cell death in type
336 1 and type 2 diabetes⁷⁷. This cell death process is dependent on NOD1/2 signaling⁷⁵, although the specifics
337 have not yet been worked out.

338 We also looked specifically at hits in pancreatic tumor cells (YAPC) regardless of significance level to get a
339 transcriptional response more specific to the pancreas (Supp. Fig. S11). Hits in this list included widely used
340 diabetes drugs, such as sulfonylureas, PPAR receptor agonists, and insulin sensitizers. Rosiglitazone is a
341 PPAR- γ agonist and was one of the most prescribed drugs for type 2 diabetes before its use was reduced due
342 to cardiac side-effects⁷⁸. Sulfonylureas are another commonly prescribed drug class for type 2 diabetes, but
343 also have notable side effects including hypoglycemia and accelerated β -cell death⁷⁹.

344 In summary, the high-loading transcripts derived from HDMA in mice prioritized of drugs with demonstrated
345 effectiveness in reducing type 2 diabetes phenotypes in humans in a tissue-specific manner. Drugs identified
346 using the islet loadings are known diabetes drugs that act directly on pancreatic function. Drugs identified
347 by the adipose loadings tended to reduce inflammatory responses and have been shown incidentally to reduce
348 obesity-related morbidity.

349 Discussion

350 Here we investigated the relative contributions of local and distal gene regulation in four tissues to heritable
351 variation in traits related to metabolic disease in genetically diverse mice. We found that distal heritability
352 was positively correlated with trait relatedness, whereas high heritability was negatively correlated with
353 trait relatedness. We used a novel high-dimensional mediation analysis (HDMA) to identify tissue-specific
354 composite transcripts that are predicted to mediate the effect of genetic background on metabolic traits. The
355 adipose-derived composite transcript robustly predicted body weight in an independent cohort of diverse
356 mice with disparate population structure. However, gene expression imputed from local genotype failed to
357 predict body weight in the second population. Taken together, these results highlight the complexity of gene
358 expression regulation in relation to trait heritability and suggest that heritable trait variation is mediated
359 primarily through distal gene regulation.

360 Supplemental Discussion

361 Our result that distal regulation accounted for most trait-related gene expression differences is consistent
362 with a complex model of genetic trait determination. It has frequently been assumed that gene regulation in

363 *cis* is the primary driver of genetically associated trait variation, but attempts to use local gene regulation
364 to explain phenotypic variation have had limited success^{16;17}. In recent years, evidence has mounted that
365 distal gene regulation may be an important mediator of trait heritability^{19;18;80}. It has been observed that
366 transcripts with high local heritability explain less expression-mediated disease heritability than those with
367 low local heritability¹⁹. Consistent with this observation, genes located near GWAS hits tend to be complexly
368 regulated¹⁸. They also tend to be enriched with functional annotations, in contrast to genes with simple
369 local regulation, which tend to be depleted of functional annotations suggesting they are less likely to be
370 directly involved in disease traits¹⁸. These observations are consistent with principles of robustness in complex
371 systems in which simple regulation of important elements leads to fragility of the system⁸¹⁻⁸³. Our results
372 are consistent, instead, with a more complex picture where genes whose expression can drive trait variation
373 are buffered from local genetic variation but are extensively influenced indirectly by genetic variation in the
374 regulatory networks converging on those genes.

375 Our results are consistent with the recently proposed omnigenic model, which posits that complex traits are
376 massively polygenic and that their heritability is spread out across the genome⁸⁴. In the omnigenic model,
377 genes are classified either as “core genes,” which directly impinge on the trait, or “peripheral genes,” which
378 are not directly trait-related, but influence core genes through the complex gene regulatory network. HDMA
379 explicitly models a central proposal of the omnigenic model which posits that once the expression of the
380 core genes (i.e. trait-mediating genes) is accounted for, there should be no residual correlation between the
381 genome and the phenotype. Here, when the composite transcript was taken into account there was no residual
382 correlation between the composite genome and composite phenotype (Fig. 3A).

383 Unlike in the omnigenic model, we did not observe a clear demarcation between the core and peripheral
384 genes in loading magnitude, but we do not necessarily expect a clear separation given the complexity of gene
385 regulation and the genotype-phenotype map⁸⁵.

386 An extension of the omnigenic model proposed that most heritability of complex traits is driven by weak
387 distal eQTLs that are potentially below the detection threshold in studies with feasible sample sizes⁸⁰. This
388 is consistent with what we observed here. For example, *Nucb2*, had a high loading in islets and was also
389 strongly distally regulated (66% distal heritability) (Fig. 5). Although its transcription was highly heritable
390 in islets, that regulation was distributed across the genome, with no clear distal eQTL (Supp. Fig. S12).
391 Thus, although distal regulation of some genes may be strong, this regulation is likely to be highly complex
392 and not easily localized.

393 Individual high-loading transcripts also demonstrated biologically interpretable, tissue-specific patterns. We

394 highlighted *Pparg*, which is known to be protective in adipose tissue⁴² where it was negatively loaded, and
395 harmful in the liver^{43–47}, where it was positively loaded. Such granular patterns may be useful in generating
396 hypotheses for further testing, and prioritizing genes as therapeutic targets. The tissue-specific nature of
397 the loadings also may provide clues to tissue-specific effects, or side effects, of targeting particular genes
398 system-wide.

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

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

411 Data Availability

412 **DO mice:** Genotypes, phenotypes, and pancreatic islet gene expression data were previously published¹².
413 Gene expression for the other tissues can be found at the Gene Expression Omnibus url{<https://www.ncbi.nlm.nih.gov/geo/>} with the following accession numbers: DO adipose tissue - GSE266549; DO liver tissue -
414 GSE266569; DO skeletal muscle - GSE266567. Expression data with calculated eQTLs are available on the
415 Synapse <https://www.synapse.org/>, synIDXXX

416 Genotypes: Sequence data for the DO mice used here are available from the Sequence Read Archive
417 <https://www.ncbi.nlm.nih.gov/sra/> (study number SRP125176). Genotype data for the CC mice are available
418 from University of North Carolina Computational Systems Biology (<http://www.csbio.unc.edu/CCstatus/CCGenomes/>).

419 Gene expression: Data can be found at the Gene Expression Omnibus url{<https://www.ncbi.nlm.nih.gov/geo/>}
420 with the following accession numbers: DO adipose tissue - GSE266549; DO liver tissue - GSE266569; DO
421 skeletal muscle - GSE266567; CC-RIX adipose tissue - GSE237737; CC-RIX liver tissue - GSE237743;

⁴²⁴ CC-RIX skeletal muscle - GSE237747. Quantified pancreatic islet gene expression for the DO mice, along
⁴²⁵ with their genotypes and phenotypes can be found on Dryad <https://datadryad.org/stash/dataset/doi:10.5061/dryad.pj105>.

⁴²⁷ Phenotypes: Metabolic phenotypes for the DO mice along with genotypes and quantified gene expression
⁴²⁸ area available from <https://datadryad.org/stash/dataset/doi:10.5061/dryad.pj105>

⁴²⁹ Metabolic phenotypes for the CC-RIX mice are available from XXX

⁴³⁰ **Acknowledgements**

⁴³¹ Here we thank people and cite funding sources.

432 Supplemental Figures

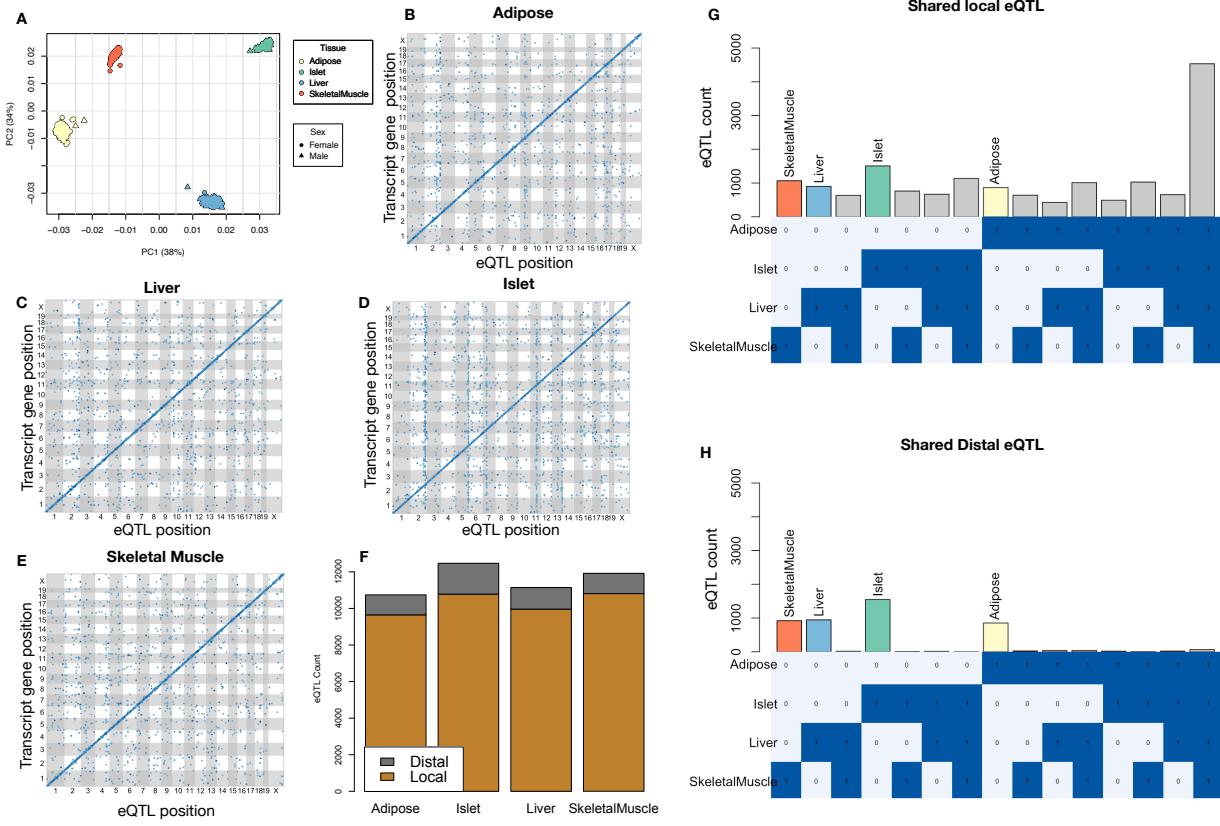


Figure S1: Overview of eQTL analysis in DO mice. **A.** RNA seq samples from the four different tissues clustered by tissue. **B.-E.** eQTL maps are shown for each tissue. The *x*-axis shows the position of the mapped eQTL, and the *y*-axis shows the physical position of the gene encoding each mapped 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.

KEGG pathway enrichments by GSEA

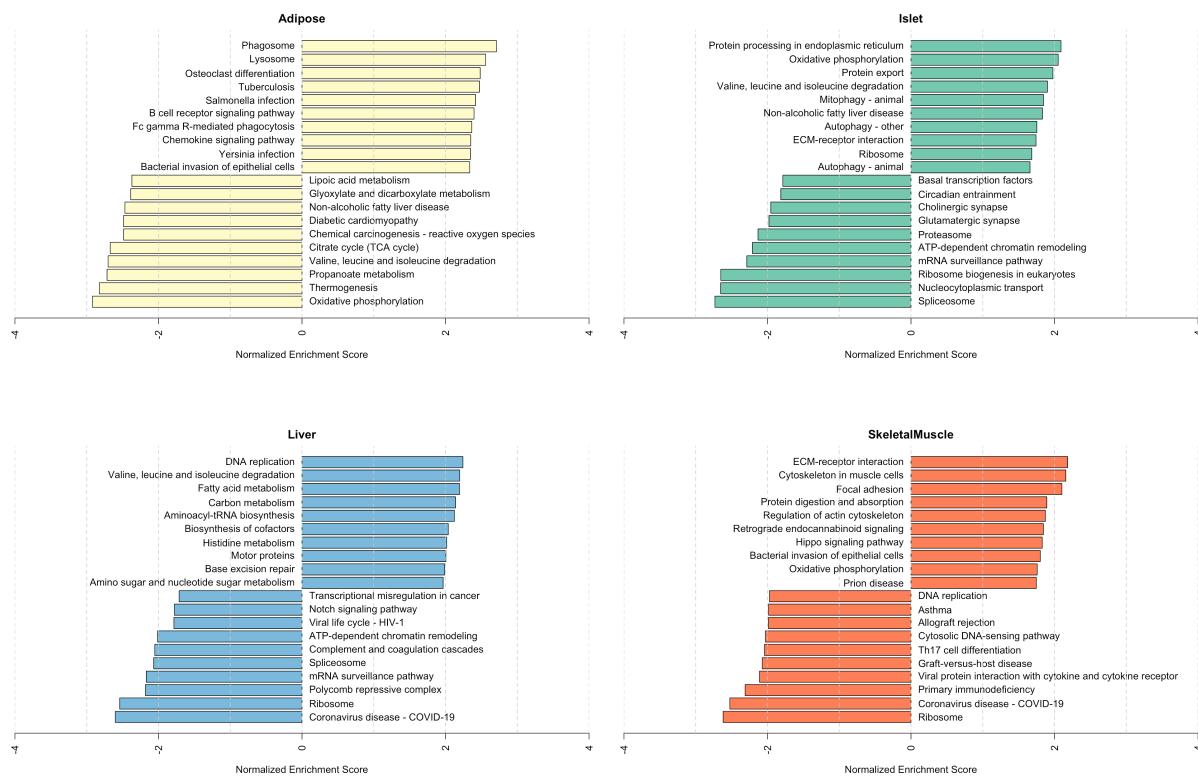


Figure S2: Bar plots showing normalized enrichment scores (NES) for KEGG pathways as determined by fast gene score enrichment analysis (fgsea). Only the top 10 positive and top 10 negative scores are shown. Colors indicate tissue. The name beside each bar shows the name of each enriched KEGG pathway.

Top GO term enrichments by GSEA

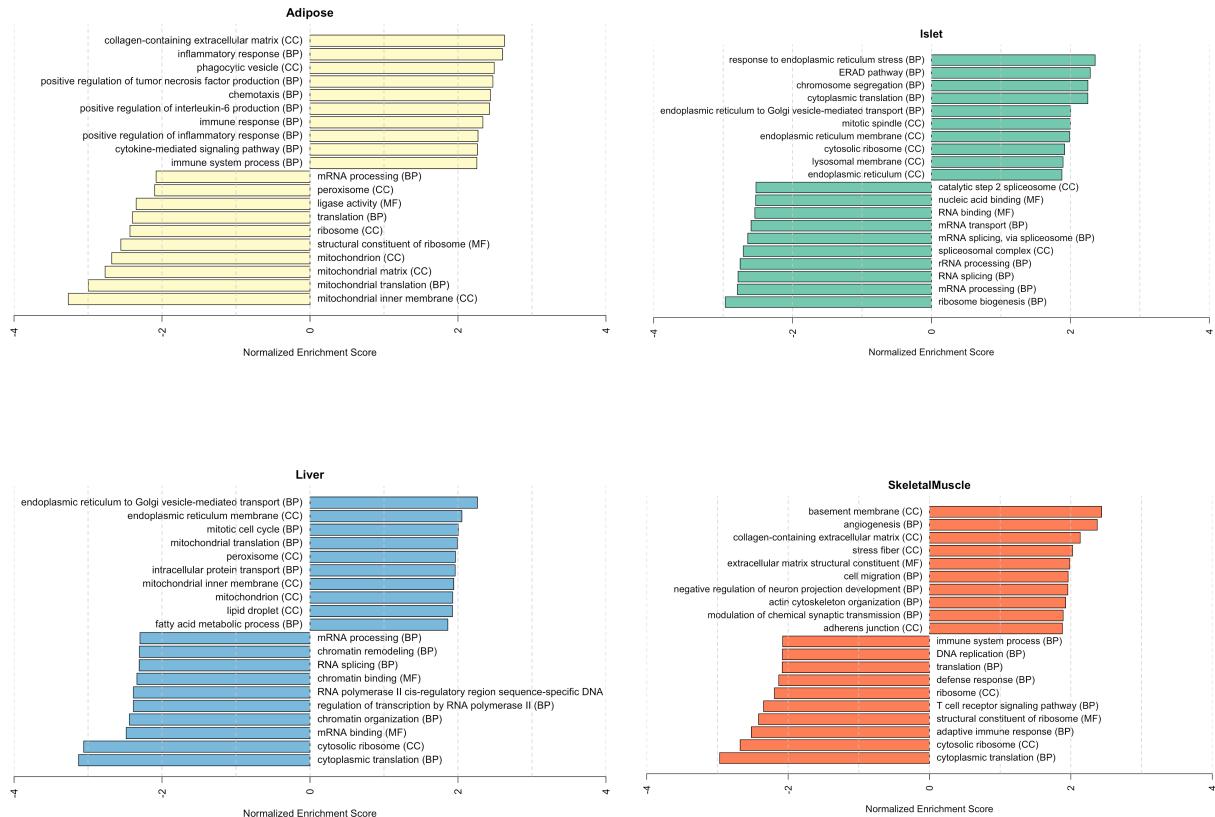


Figure S3: Bar plots showing normalized enrichment scores (NES) for GO terms as determined by fast gene score enrichment analysis (fgsea). Only the top 10 positive and top 10 negative scores are shown. Colors indicate tissue. The name beside each bar shows the name of each enriched GO term. The letters in parentheses indicate whether the term is from the biological process ontology (BP), the molecular function ontology (MF), or the cellular compartment ontology (CC).

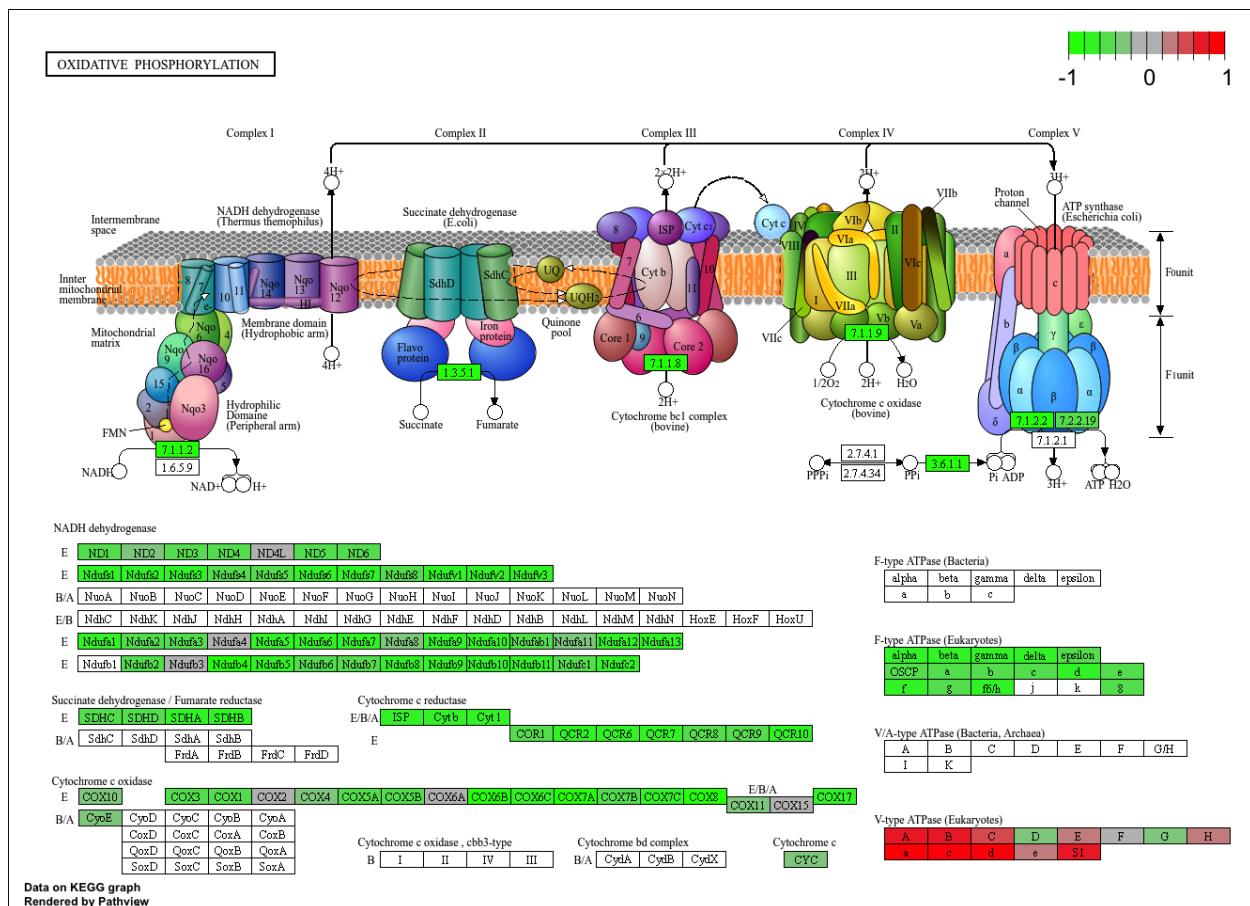


Figure S4: The KEGG pathway for oxidative phosphorylation in mice. Each element is colored based on its HDMA loading from adipose tissue normalized to run from -1 to 1. Genes highlighted in green had negative loadings, and those highlighted in red had positive loadings. Almost the entire pathway was strongly negatively loaded indicating that increased expression of genes involved in oxidative phosphorylation was associated with reduced MDI.

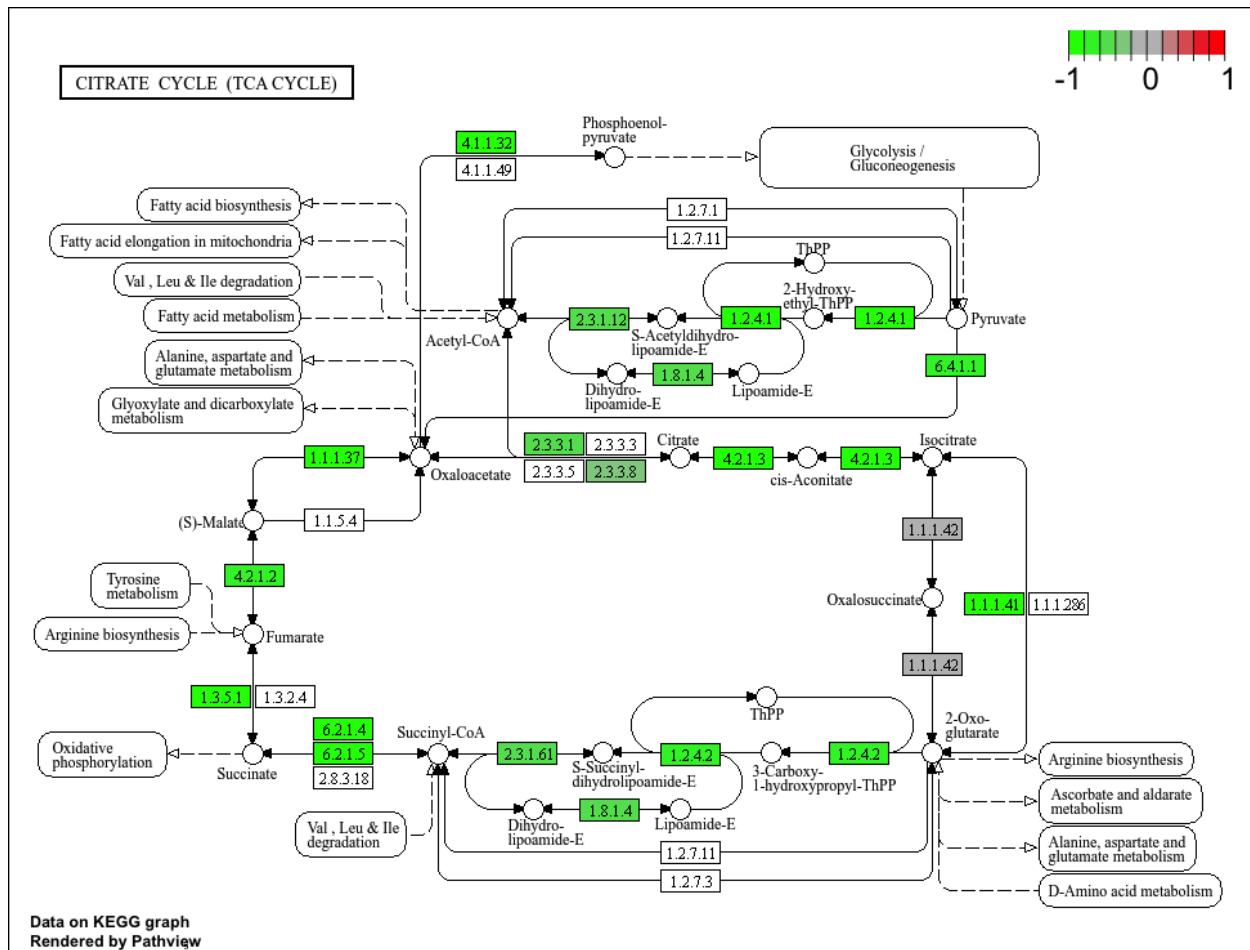


Figure S5: The KEGG pathway for the TCA (citric acid) cycle in mice. Each element is colored based on its HDMA loading from adipose tissue normalized to run from -1 to 1. Genes highlighted in green had negative loadings, and those highlighted in red had positive loadings. Many genes in the cycle were strongly negatively loaded indicating that increased expression of genes involved in branched-chain amino acid degradation was associated with reduced MDI.

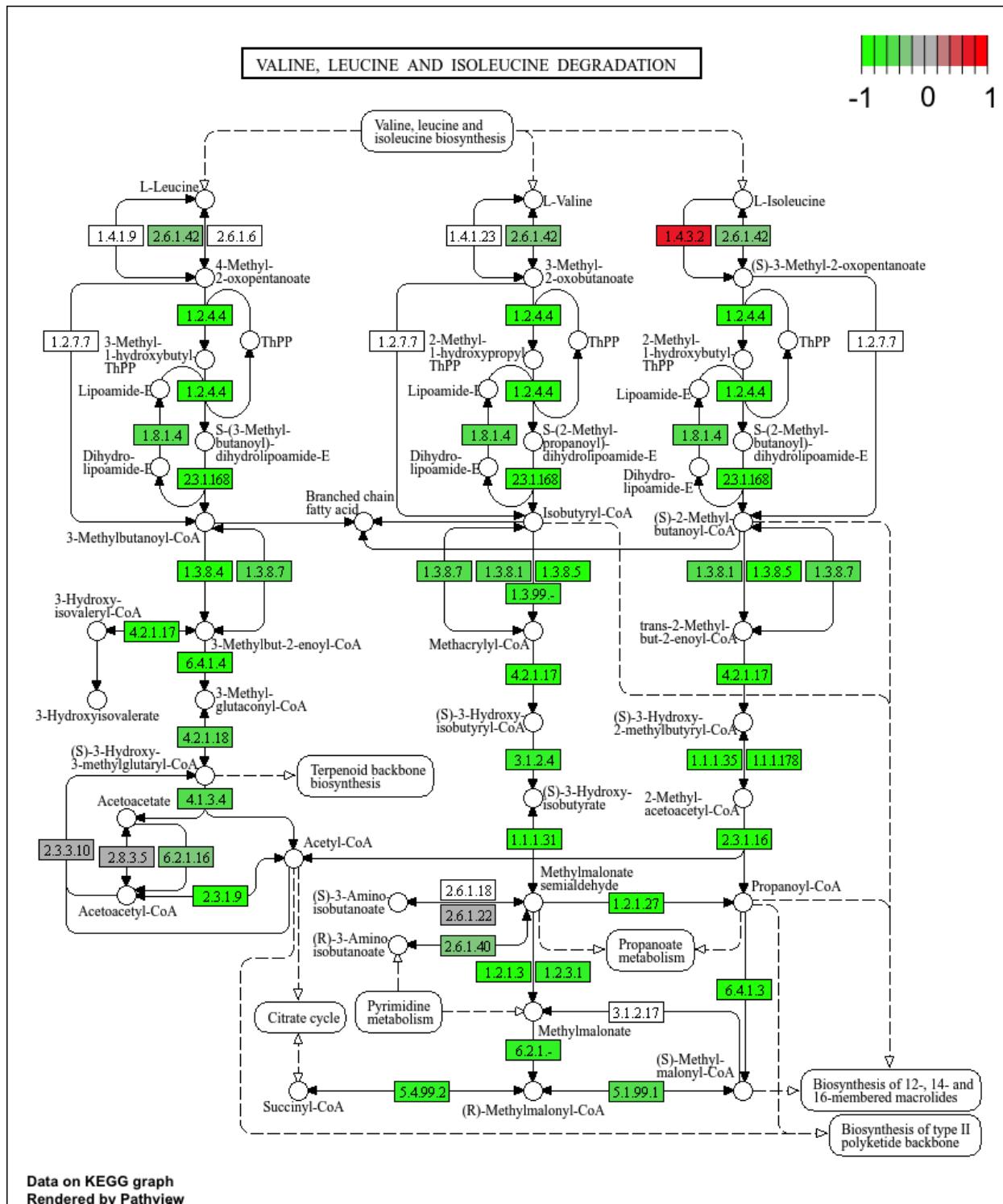


Figure S6: The KEGG pathway for branched-chain amino acid degradation in mice. Each element is colored based on its HDMA loading from adipose tissue normalized to run from -1 to 1. Genes highlighted in green had negative loadings, and those highlighted in red had positive loadings. Almost the entire pathway was strongly negatively loaded indicating that increased expression of genes involved in branched-chain amino acid degradation was associated with reduced MDI.

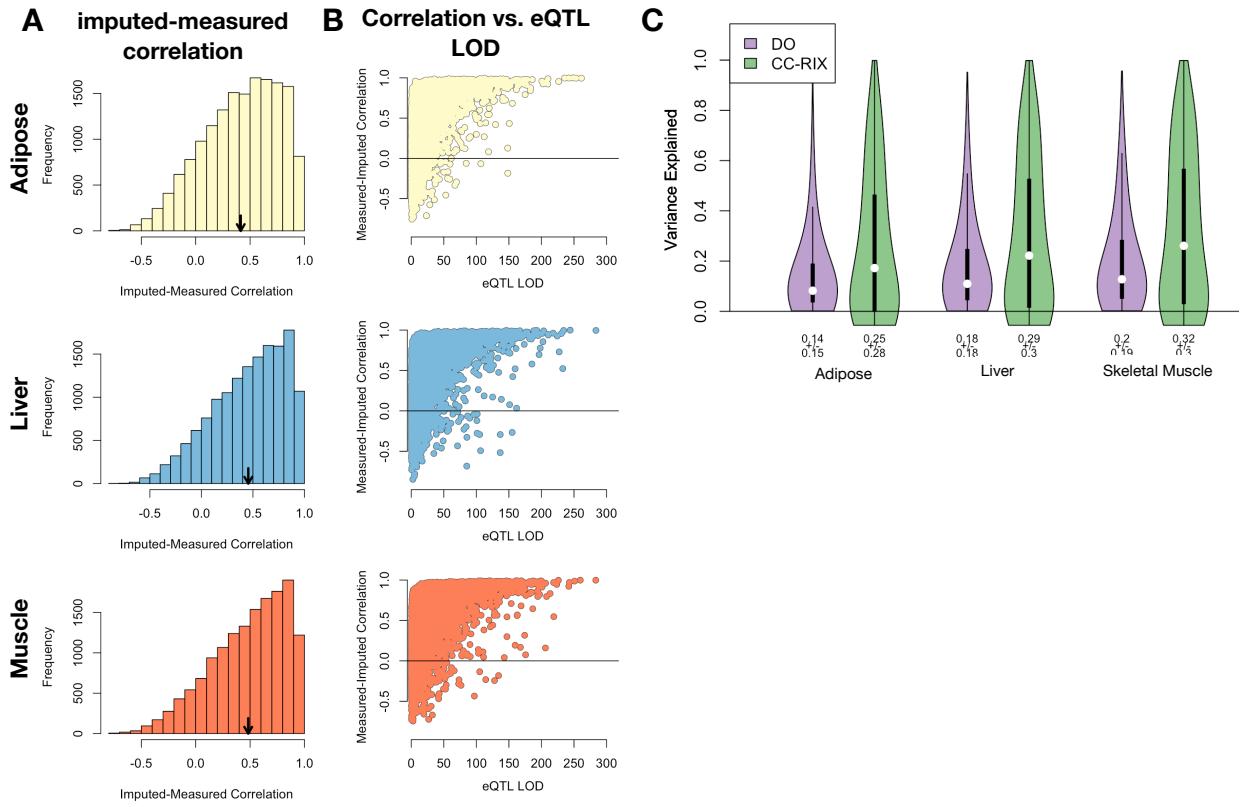


Figure S7: Validation of transcript imputation in the CC-RIX. **A.** Distributions of correlations between imputed and measured transcripts in the CC-RIX. The mean of each distribution is shown by the red line. All distributions were skewed toward positive correlations and had positive means near a Pearson 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.

| id | norm_gs | cell_iname | pert_type | raw_cs▲ | fdr_q_nlog10 | set_type | src_set_id |
|----|---------|------------|------------|---------|--------------|-------------|-----------------------------------|
| | | HA1E | TRT_CP | -0.97 | 15.65 | PCL | CP_PROTEIN_SYNTHESIS_INHIBITOR |
| | | PC3 | TRT_SH.CGS | -0.90 | 15.65 | PATHWAY_SET | BIOCARTA_EIF4_PATHWAY |
| | | A375 | TRT_CP | -0.87 | 15.65 | MOA_CLASS | RAF_INHIBITOR |
| | | HCC515 | TRT_CP | -0.84 | 15.65 | PCL | CP_TOPOISOMERASE_INHIBITOR |
| | | HEPG2 | TRT_SH.CGS | -0.82 | 15.65 | PATHWAY_SET | BIOCARTA_BCR_PATHWAY |
| | | PC3 | TRT_CP | -0.77 | 15.65 | MOA_CLASS | MTOR_INHIBITOR |
| | | HCC515 | TRT_CP | -0.76 | 15.65 | PCL | CP_GLUCOCORTICOID_RECEPTORAGONIST |
| | | HCC515 | TRT_CP | -0.76 | 15.65 | MOA_CLASS | GLUCOCORTICOID_RECEPTORAGONIST |
| | | A375 | TRT_CP | -0.72 | 15.65 | MOA_CLASS | MTOR_INHIBITOR |
| | | -666 | TRT_CP | -0.70 | 15.65 | PCL | CP_PROTEIN_SYNTHESIS_INHIBITOR |
| | | -666 | TRT_CP | -0.68 | 15.65 | PCL | CP_JAK_INHIBITOR |
| | | A549 | TRT_CP | -0.67 | 15.65 | PCL | CP_GLUCOCORTICOID_RECEPTORAGONIST |
| | | A549 | TRT_CP | -0.67 | 15.65 | MOA_CLASS | GLUCOCORTICOID_RECEPTORAGONIST |
| | | -666 | TRT_CP | -0.57 | 15.65 | PCL | CP_MTOR_INHIBITOR |
| | | -666 | TRT_CP | -0.55 | 15.65 | MOA_CLASS | MTOR_INHIBITOR |
| | | -666 | TRT_CP | -0.55 | 15.65 | PCL | CP_PI3K_INHIBITOR |
| | | -666 | TRT_CP | 0.85 | 15.65 | MOA_CLASS | PKC_ACTIVATOR |

Figure S8: CMAP results using the adipose tissue composite transcript as an input. All query results with a $-\log_{10}(q) > 15$ across all cell types are shown.

| id | norm_gs | cell_iname | pert_type | raw_cs▲ | fdr_q_nlog10 | set_type | src_set_id |
|----|---------|------------|------------|---------|--------------|-------------|--|
| | | VCAP | TRT_SH.CGS | -0.99 | 15.65 | PATHWAY_SET | REACTOME_DOWNSTREAM_TCR_SIGNALING |
| | | VCAP | TRT_SH.CGS | -0.99 | 15.65 | PATHWAY_SET | REACTOME_NOD1_2_SIGNALING_PATHWAY |
| | | A549 | TRT_SH.CGS | -0.92 | 15.65 | PATHWAY_SET | BIOCARTA_TNFR1_PATHWAY |
| | | VCAP | TRT_SH.CGS | -0.92 | 15.65 | PATHWAY_SET | HALLMARK_WNT_BETA_CATENIN_SIGNALING |
| | | HT29 | TRT_CP | -0.92 | 15.65 | PCL | CP_TUBULIN_INHIBITOR |
| | | -666 | TRT_OE | -0.88 | 15.65 | PCL | OE_CELL_CYCLE_INHIBITION |
| | | VCAP | TRT_SH.CGS | -0.87 | 15.65 | PATHWAY_SET | REACTOME_P75_NTR_RECECTOR_MEDiated_SIGNALLING |
| | | HT29 | TRT_CP | -0.86 | 15.65 | MOA_CLASS | TUBULIN_INHIBITOR |
| | | MCF7 | TRT_CP | -0.85 | 15.65 | PCL | CP_TUBULIN_INHIBITOR |
| | | -666 | TRT_CP | -0.81 | 15.65 | PCL | CP_PROTEASOME_INHIBITOR |
| | | -666 | TRT_SH.CGS | -0.80 | 15.65 | PATHWAY_SET | REACTOME_DOWNREGULATION_OF_ERBB2_ERBB3_SIGNALING |
| | | HCC515 | TRT_CP | -0.80 | 15.65 | PCL | CP_GLUCOCORTICOID_RECEPTORAGONIST |
| | | HCC515 | TRT_CP | -0.80 | 15.65 | MOA_CLASS | GLUCOCORTICOID_RECEPTORAGONIST |
| | | A549 | TRT_OE | -0.78 | 15.65 | PATHWAY_SET | REACTOME_RAF_MAP_KINASE CASCADE |
| | | A549 | TRT_OE | -0.78 | 15.65 | PATHWAY_SET | PID_RAS_PATHWAY |
| | | -666 | TRT_SH.CGS | -0.78 | 15.65 | PCL | KD_RIBOSOMAL_40S_SUBUNIT |
| | | A549 | TRT_OE | -0.76 | 15.65 | PATHWAY_SET | REACTOME_SIGNALLING_TO_P38_VIA_RIT_AND_RIN |
| | | A549 | TRT_OE | -0.76 | 15.65 | PATHWAY_SET | REACTOME_PROLONGED_ERK_ACTIVATION_EVENTS |
| | | A549 | TRT_OE | -0.73 | 15.65 | PATHWAY_SET | PID_TCR_RAS_PATHWAY |
| | | HA1E | TRT_OE | -0.73 | 15.65 | PATHWAY_SET | REACTOME_SHC RELATED EVENTS |
| | | HA1E | TRT_OE | -0.71 | 15.65 | PATHWAY_SET | PID_EPHB_FWD_PATHWAY |
| | | -666 | TRT_CP | -0.70 | 15.65 | MOA_CLASS | GLYCOGEN_SYNTHASE_KINASE_INHIBITOR |
| | | HA1E | TRT_OE | -0.70 | 15.65 | PATHWAY_SET | PID_GMCSF_PATHWAY |
| | | A549 | TRT_OE | -0.69 | 15.65 | PATHWAY_SET | REACTOME_SIGNALLING_TO_ERKS |
| | | -666 | TRT_LIG | -0.69 | 15.65 | PATHWAY_SET | PID_ERBB_NETWORK_PATHWAY |
| | | -666 | TRT_CP | -0.67 | 15.65 | MOA_CLASS | PROTEASOME_INHIBITOR |
| | | -666 | TRT_CP | -0.66 | 15.65 | PCL | CP_GLYCOGEN_SYNTHASE_KINASE_INHIBITOR |
| | | -666 | TRT_CP | 0.73 | 15.65 | MOA_CLASS | MTOR_INHIBITOR |

Figure S9: CMAP results using the pancreatic islet composite transcript as an input. All query results with a $-\log_{10}(q) > 15$ across all cell types are shown.

| id | norm_CS | cell_iname | pert_type | raw_CS ▲ | fdr_q_nlog10 | set_type | src_set_id |
|----|---------|------------|-----------|----------|--------------|-----------|---|
| | | ASC | TRT_CP | -0.94 | 0.79 | PCL | CP_PARP_INHIBITOR |
| | | ASC | TRT_CP | -0.94 | 0.79 | MOA_CLASS | PROTEIN_TYROSINE_KINASE_INHIBITOR |
| | | ASC | TRT_CP | -0.84 | 0.45 | MOA_CLASS | BTK_INHIBITOR |
| | | ASC | TRT_CP | -0.81 | 0.39 | MOA_CLASS | LEUCINE_RICH_REPEAT_KINASE_INHIBITOR |
| | | ASC | TRT_CP | -0.81 | 0.79 | PCL | CP_HSP_INHIBITOR |
| | | ASC | TRT_CP | -0.80 | 0.93 | PCL | CP_EGFR_INHIBITOR |
| | | ASC | TRT_CP | -0.79 | 0.32 | MOA_CLASS | T-TYPE_CALCIUM_CHANNEL_BLOCKER |
| | | ASC | TRT_CP | -0.79 | 1.09 | PCL | CP_MTOR_INHIBITOR |
| | | ASC | TRT_CP | -0.76 | 0.97 | PCL | CP_PI3K_INHIBITOR |
| | | ASC | TRT_CP | -0.75 | 0.20 | MOA_CLASS | HISTONE_DEMETHYLASE_INHIBITOR |
| | | ASC | TRT_CP | -0.74 | 0.42 | PCL | CP_IKK_INHIBITOR |
| | | ASC | TRT_CP | -0.74 | 0.83 | PCL | CP_AURORA_KINASE_INHIBITOR |
| | | ASC | TRT_CP | -0.74 | 0.17 | PCL | CP_LEUCINE_RICH_REPEAT_KINASE_INHIBITOR |
| | | ASC | TRT_CP | -0.72 | 0.36 | PCL | CP_BROMODOMAIN_INHIBITOR |
| | | ASC | TRT_CP | -0.71 | 1.09 | MOA_CLASS | TYROSINE_KINASE_INHIBITOR |
| | | ASC | TRT_CP | -0.70 | 0.82 | PCL | CP_PROTEIN_SYNTHESIS_INHIBITOR |
| | | ASC | TRT_CP | -0.67 | 0.69 | PCL | CP_SRC_INHIBITOR |
| | | ASC | TRT_CP | -0.67 | 0.81 | MOA_CLASS | AURORA_KINASE_INHIBITOR |
| | | ASC | TRT_CP | -0.65 | 0.89 | MOA_CLASS | FLT3_INHIBITOR |
| | | ASC | TRT_CP | -0.62 | 0.40 | MOA_CLASS | FGFR_INHIBITOR |
| | | ASC | TRT_CP | -0.59 | 0.66 | MOA_CLASS | MEK_INHIBITOR |
| | | ASC | TRT_CP | -0.59 | 0.13 | MOA_CLASS | SYK_INHIBITOR |
| | | ASC | TRT_CP | -0.58 | 0.01 | PCL | CP_PKC_INHIBITOR |
| | | ASC | TRT_CP | -0.58 | 0.65 | PCL | CP_HDAC_INHIBITOR |
| | | ASC | TRT_CP | -0.58 | 0.65 | PCL | CP_ATPASE_INHIBITOR |
| | | ASC | TRT_CP | -0.53 | 0.09 | PCL | CP_FLT3_INHIBITOR |
| | | ASC | TRT_CP | -0.53 | 0.42 | PCL | CP_P38_MAPK_INHIBITOR |
| | | ASC | TRT_CP | -0.53 | 0.22 | MOA_CLASS | IKK_INHIBITOR |
| | | ASC | TRT_CP | -0.52 | 0.58 | PCL | CP_VEGFR_INHIBITOR |
| | | ASC | TRT_CP | -0.51 | -0.00 | PCL | CP_T_TYPE_CALCIUM_CHANNEL_BLOCKER |

Figure S10: CMAP results using the adipose tissue composite transcript as an input. Query results are limited to the 30 most negatively correlated signals from normal adipocytes.

| norm_CS | | | | | | |
|---------|-----------|-----------|--------|--------------|-------------|---------------------------------------|
| id | cell_name | pert_type | raw_CS | fdr_q_nlog10 | set_type | src_set_id |
| | YAPC | TRT_CP | -1.00 | 0.67 | MOA_CLASS | ABL_KINASE_INHIBITOR |
| | YAPC | TRT_CP | -0.99 | 0.66 | PCL | CP_CDK_INHIBITOR |
| | YAPC | TRT_CP | -0.97 | 1.41 | PCL | CP_TOPOISOMERASE_INHIBITOR |
| | YAPC | TRT_CP | -0.95 | 0.70 | MOA_CLASS | THYMIDYLATE_SYNTHASE_INHIBITOR |
| | YAPC | TRT_CP | -0.95 | 0.62 | MOA_CLASS | ADRENERGIC_INHIBITOR |
| | YAPC | TRT_CP | -0.94 | 0.50 | MOA_CLASS | BENZODIAZEPINE_RECECTOR_ANTAGONIST |
| | YAPC | TRT_CP | -0.89 | 0.63 | PCL | CP_RIBONUCLEOTIDE_REDUCTASE_INHIBITOR |
| | YAPC | TRT_CP | -0.88 | 0.52 | MOA_CLASS | VASOPRESSIN_RECECTOR_ANTAGONIST |
| | YAPC | TRT_CP | -0.85 | 0.63 | MOA_CLASS | ANGIOTENSIN_RECECTOR_ANTAGONIST |
| | YAPC | TRT_CP | -0.85 | 0.33 | PCL | CP_CANNABINOID_RECECTORAGONIST |
| | YAPC | TRT_CP | -0.84 | 0.30 | PCL | CP_RETINOID_RECECTORAGONIST |
| | YAPC | TRT_CP | -0.83 | 1.19 | MOA_CLASS | NFKB_PATHWAY_INHIBITOR |
| | YAPC | TRT_CP | -0.83 | 0.54 | MOA_CLASS | DNA_ALKYLATING_DRUG |
| | YAPC | TRT_CP | -0.80 | 0.50 | MOA_CLASS | CHOLESTEROL_INHIBITOR |
| | YAPC | TRT_CP | -0.79 | 0.15 | MOA_CLASS | SULFONYLUREA |
| | YAPC | TRT_CP | -0.78 | 0.52 | MOA_CLASS | HIV_INTEGRASE_INHIBITOR |
| | YAPC | TRT_CP | -0.78 | 0.13 | MOA_CLASS | LEUKOTRIENE_INHIBITOR |
| | YAPC | TRT_CP | -0.78 | 0.45 | PCL | CP_PPAR_RECECTORAGONIST |
| | YAPC | TRT_CP | -0.78 | 0.54 | MOA_CLASS | INSULIN_SENSITIZER |
| | YAPC | TRT_CP | -0.77 | 0.51 | MOA_CLASS | ESTROGEN_RECECTORANTAGONIST |
| | YAPC | TRT_CP | -0.77 | 0.76 | MOA_CLASS | DNA_SYNTHESIS_INHIBITOR |
| | YAPC | TRT_XPR | -0.77 | 0.67 | PATHWAY_SET | BIOCARTA_PARKIN_PATHWAY |
| | YAPC | TRT_CP | -0.77 | 0.51 | PCL | CP_VEGFR_INHIBITOR |
| | YAPC | TRT_CP | -0.75 | 0.39 | MOA_CLASS | RNA_SYNTHESIS_INHIBITOR |
| | YAPC | TRT_CP | -0.72 | 0.60 | MOA_CLASS | BCR-ABL_KINASE_INHIBITOR |
| | YAPC | TRT_XPR | -0.71 | 0.66 | PATHWAY_SET | BIOCARTA_EIF_PATHWAY |
| | YAPC | TRT_XPR | -0.69 | 0.54 | PATHWAY_SET | PID_CIRCADIAN_PATHWAY |
| | YAPC | TRT_CP | -0.68 | 0.77 | MOA_CLASS | TOPOISOMERASE_INHIBITOR |
| | YAPC | TRT_XPR | -0.64 | 0.49 | PATHWAY_SET | BIOCARTA_CBL_PATHWAY |
| | YAPC | TRT_CP | -0.64 | 0.53 | MOA_CLASS | TUBULIN_INHIBITOR |

Figure S11: CMAP results using the pancreatic islet composite transcript as an input. Query results are limited to the 30 most negatively correlated signals from YAPC cells, which were derived from a pancreatic carcinoma cells.

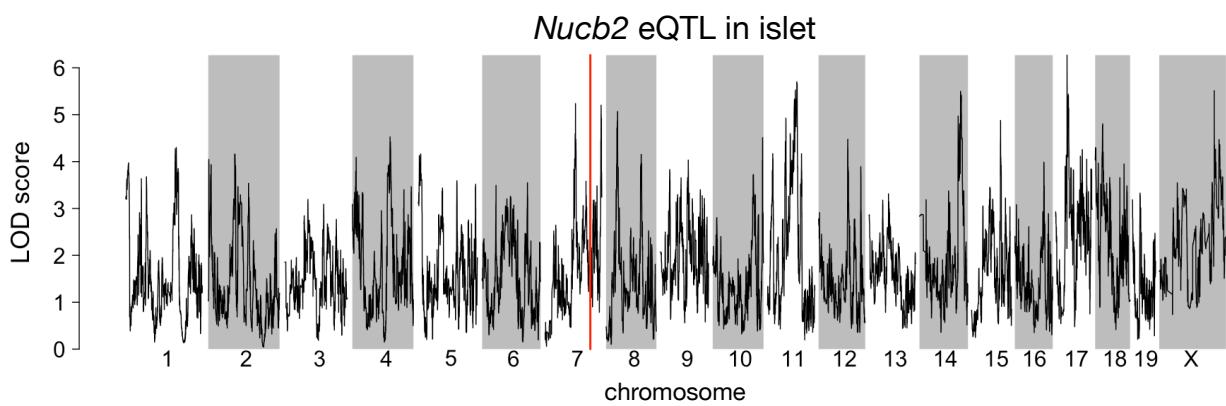


Figure S12: Regulation of *Nucb2* expression in islet. *Nucb2* is encoded on mouse chromosome 7 at 116.5 Mb (red line). In islets the heritability of *Nucb2* expression levels is 69% heritable. This LOD score trace shows that there is no local eQTL at that position, nor any strong distal eQTL anywhere else in the genome.

433 **References**

- 434 [1] M. T. Maurano, R. Humbert, E. Rynes, R. E. Thurman, E. Haugen, H. Wang, A. P. Reynolds,
435 R. Sandstrom, H. Qu, J. Brody, A. Shafer, F. Neri, K. Lee, T. Kutyavin, S. Stehling-Sun, A. K.
436 Johnson, T. K. Canfield, E. Giste, M. Diegel, D. Bates, R. S. Hansen, S. Neph, P. J. Sabo, S. Heimfeld,
437 A. Raubitschek, S. Ziegler, C. Cotsapas, N. Sotoodehnia, I. Glass, S. R. Sunyaev, R. Kaul, and J. A.
438 Stamatoyannopoulos. Systematic localization of common disease-associated variation in regulatory DNA.
439 *Science*, 337(6099):1190–1195, Sep 2012.
- 440 [2] K. K. Farh, A. Marson, J. Zhu, M. Kleinewietfeld, W. J. Housley, S. Beik, N. Shores, H. Whitton, R. J.
441 Ryan, A. A. Shishkin, M. Hatan, M. J. Carrasco-Alfonso, D. Mayer, C. J. Luckey, N. A. Patsopoulos,
442 P. L. De Jager, V. K. Kuchroo, C. B. Epstein, M. J. Daly, D. A. Hafler, and B. E. Bernstein. Genetic
443 and epigenetic fine mapping of causal autoimmune disease variants. *Nature*, 518(7539):337–343, Feb
444 2015.
- 445 [3] E. Pennisi. The Biology of Genomes. Disease risk links to gene regulation. *Science*, 332(6033):1031, May
446 2011.
- 447 [4] L. A. Hindorff, P. Sethupathy, H. A. Junkins, E. M. Ramos, J. P. Mehta, F. S. Collins, and T. A. Manolio.
448 Potential etiologic and functional implications of genome-wide association loci for human diseases and
449 traits. *Proc Natl Acad Sci*, 106(23):9362–9367, Jun 2009.
- 450 [5] J. K. Pickrell. Joint analysis of functional genomic data and genome-wide association studies of 18
451 human traits. *Am J Hum Genet*, 94(4):559–573, Apr 2014.
- 452 [6] D. Welter, J. MacArthur, J. Morales, T. Burdett, P. Hall, H. Junkins, A. Klemm, P. Flieck, T. Manolio,
453 L. Hindorff, and H. Parkinson. The NHGRI GWAS Catalog, a curated resource of SNP-trait associations.
454 *Nucleic Acids Res*, 42(Database issue):D1001–1006, Jan 2014.
- 455 [7] Y. I. Li, B. van de Geijn, A. Raj, D. A. Knowles, A. A. Petti, D. Golan, Y. Gilad, and J. K. Pritchard.
456 RNA splicing is a primary link between genetic variation and disease. *Science*, 352(6285):600–604, Apr
457 2016.
- 458 [8] D. Zhou, Y. Jiang, X. Zhong, N. J. Cox, C. Liu, and E. R. Gamazon. A unified framework for joint-tissue
459 transcriptome-wide association and Mendelian randomization analysis. *Nat Genet*, 52(11):1239–1246,
460 Nov 2020.
- 461 [9] E. R. Gamazon, H. E. Wheeler, K. P. Shah, S. V. Mozaffari, K. Aquino-Michaels, R. J. Carroll, A. E.

- 462 Eyler, J. C. Denny, D. L. Nicolae, N. J. Cox, and H. K. Im. A gene-based association method for
463 mapping traits using reference transcriptome data. *Nat Genet*, 47(9):1091–1098, Sep 2015.
- 464 [10] Z. Zhu, F. Zhang, H. Hu, A. Bakshi, M. R. Robinson, J. E. Powell, G. W. Montgomery, M. E. Goddard,
465 N. R. Wray, P. M. Visscher, and J. Yang. Integration of summary data from GWAS and eQTL studies
466 predicts complex trait gene targets. *Nat Genet*, 48(5):481–487, May 2016.
- 467 [11] A. Gusev, A. Ko, H. Shi, G. Bhatia, W. Chung, B. W. Penninx, R. Jansen, E. J. de Geus, D. I. Boomsma,
468 F. A. Wright, P. F. Sullivan, E. Nikkola, M. Alvarez, M. Civelek, A. J. Lusis, T. ki, E. Raitoharju,
469 M. nen, I. ä, O. T. Raitakari, J. Kuusisto, M. Laakso, A. L. Price, P. Pajukanta, and B. Pasaniuc.
470 Integrative approaches for large-scale transcriptome-wide association studies. *Nat Genet*, 48(3):245–252,
471 Mar 2016.
- 472 [12] M. P. Keller, D. M. Gatti, K. L. Schueler, M. E. Rabaglia, D. S. Stapleton, P. Simecek, M. Vincent,
473 S. Allen, A. T. Broman, R. Bacher, C. Kendzierski, K. W. Broman, B. S. Yandell, G. A. Churchill, and
474 A. D. Attie. Genetic Drivers of Pancreatic Islet Function. *Genetics*, 209(1):335–356, May 2018.
- 475 [13] W. L. Crouse, G. R. Keele, M. S. Gastonguay, G. A. Churchill, and W. Valdar. A Bayesian model
476 selection approach to mediation analysis. *PLoS Genet*, 18(5):e1010184, May 2022.
- 477 [14] J. M. Chick, S. C. Munger, P. Simecek, E. L. Huttlin, K. Choi, D. M. Gatti, N. Raghupathy, K. L. Svenson,
478 G. A. Churchill, and S. P. Gygi. Defining the consequences of genetic variation on a proteome-wide scale.
479 *Nature*, 534(7608):500–505, Jun 2016.
- 480 [15] H. E. Wheeler, S. Ploch, A. N. Barbeira, R. Bonazzola, A. Andaleon, A. Fotuhi Siahpirani, A. Saha,
481 A. Battle, S. Roy, and H. K. Im. Imputed gene associations identify replicable trans-acting genes enriched
482 in transcription pathways and complex traits. *Genet Epidemiol*, 43(6):596–608, Sep 2019.
- 483 [16] B. D. Umans, A. Battle, and Y. Gilad. Where Are the Disease-Associated eQTLs? *Trends Genet*,
484 37(2):109–124, Feb 2021.
- 485 [17] N. J. Connally, S. Nazeen, D. Lee, H. Shi, J. Stamatoyannopoulos, S. Chun, C. Cotsapas, C. A. Cassa,
486 and S. R. Sunyaev. The missing link between genetic association and regulatory function. *Elife*, 11, Dec
487 2022.
- 488 [18] H. Mostafavi, J. P. Spence, S. Naqvi, and J. K. Pritchard. Systematic differences in discovery of genetic
489 effects on gene expression and complex traits. *Nat Genet*, 55(11):1866–1875, Nov 2023.
- 490 [19] D. W. Yao, L. J. O'Connor, A. L. Price, and A. Gusev. Quantifying genetic effects on disease mediated
491 by assayed gene expression levels. *Nat Genet*, 52(6):626–633, Jun 2020.

- 492 [20] X. Liu, J. A. Mefford, A. Dahl, Y. He, M. Subramaniam, A. Battle, A. L. Price, and N. Zaitlen. GBAT:
493 a gene-based association test for robust detection of trans-gene regulation. *Genome Biol*, 21(1):211, Aug
494 2020.
- 495 [21] E. Sollis, A. Mosaku, A. Abid, A. Buniello, M. Cerezo, L. Gil, T. Groza, O. §, P. Hall, J. Hayhurst,
496 A. Ibrahim, Y. Ji, S. John, E. Lewis, J. A. L. MacArthur, A. McMahon, D. Osumi-Sutherland,
497 K. Panoutsopoulou, Z. Pendlington, S. Ramachandran, R. Stefancsik, J. Stewart, P. Whetzel, R. Wilson,
498 L. Hindorff, F. Cunningham, S. A. Lambert, M. Inouye, H. Parkinson, and L. W. Harris. The NHGRI-EBI
499 GWAS Catalog: knowledgebase and deposition resource. *Nucleic Acids Res*, 51(D1):D977–D985, Jan
500 2023.
- 501 [22] R. J. F. Loos and G. S. H. Yeo. The genetics of obesity: from discovery to biology. *Nat Rev Genet*,
502 23(2):120–133, Feb 2022.
- 503 [23] R. K. Singh, P. Kumar, and K. Mahalingam. Molecular genetics of human obesity: A comprehensive
504 review. *C R Biol*, 340(2):87–108, Feb 2017.
- 505 [24] P. Arner. Obesity—a genetic disease of adipose tissue? *Br J Nutr*, 83 Suppl 1:9–16, Mar 2000.
- 506 [25] G. A. Churchill, D. M. Gatti, S. C. Munger, and K. L. Svenson. The Diversity Outbred mouse population.
507 *Mamm Genome*, 23(9-10):713–718, Oct 2012.
- 508 [26] Elissa J Chesler, Darla R Miller, Lisa R Branstetter, Leslie D Galloway, Barbara L Jackson, Vivek M
509 Philip, Brynn H Voy, Cymbeline T Culiat, David W Threadgill, Robert W Williams, et al. The
510 collaborative cross at oak ridge national laboratory: developing a powerful resource for systems genetics.
511 *Mammalian Genome*, 19:382–389, 2008.
- 512 [27] Michael C Saul, Vivek M Philip, Laura G Reinholdt, and Elissa J Chesler. High-diversity mouse
513 populations for complex traits. *Trends in Genetics*, 35(7):501–514, 2019.
- 514 [28] S. M. Clee and A. D. Attie. The genetic landscape of type 2 diabetes in mice. *Endocr Rev*, 28(1):48–83,
515 Feb 2007.
- 516 [29] K. W. Broman, D. M. Gatti, P. Simecek, N. A. Furlotte, P. Prins, Š. Sen, B. S. Yandell, and G. A.
517 Churchill. R/qtl2: Software for Mapping Quantitative Trait Loci with High-Dimensional Data and
518 Multiparent Populations. *Genetics*, 211(2):495–502, Feb 2019.
- 519 [30] Fabien Girka, Etienne Camenen, Caroline Peltier, Arnaud Gloaguen, Vincent Guillemot, Laurent Le
520 Brusquet, and Arthur Tenenhaus. *RGCCA: Regularized and Sparse Generalized Canonical Correlation
521 Analysis for Multiblock Data*, 2023. R package version 3.0.3.

- 522 [31] M. Helmer, S. Warrington, A. R. Mohammadi-Nejad, J. L. Ji, A. Howell, B. Rosand, A. Anticevic,
523 S. N. Sotiropoulos, and J. D. Murray. On the stability of canonical correlation analysis and partial least
524 squares with application to brain-behavior associations. *Commun Biol*, 7(1):217, Feb 2024.
- 525 [32] Gennady Korotkevich, Vladimir Sukhov, and Alexey Sergushichev. Fast gene set enrichment analysis.
526 *bioRxiv*, 2019.
- 527 [33] A. Subramanian, P. Tamayo, V. K. Mootha, S. Mukherjee, B. L. Ebert, M. A. Gillette, A. Paulovich,
528 S. L. Pomeroy, T. R. Golub, E. S. Lander, and J. P. Mesirov. Gene set enrichment analysis: a
529 knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci U S A*,
530 102(43):15545–15550, Oct 2005.
- 531 [34] S. Subramanian and A. Chait. The effect of dietary cholesterol on macrophage accumulation in adipose
532 tissue: implications for systemic inflammation and atherosclerosis. *Curr Opin Lipidol*, 20(1):39–44, Feb
533 2009.
- 534 [35] I. Akoumianakis, N. Akawi, and C. Antoniades. Exploring the Crosstalk between Adipose Tissue and
535 the Cardiovascular System. *Korean Circ J*, 47(5):670–685, Sep 2017.
- 536 [36] I. S. Stafeev, A. V. Vorotnikov, E. I. Ratner, M. Y. Menshikov, and Y. V. Parfyonova. Latent Inflammation
537 and Insulin Resistance in Adipose Tissue. *Int J Endocrinol*, 2017:5076732, 2017.
- 538 [37] I. P. Fischer, M. Irmler, C. W. Meyer, S. J. Sachs, F. Neff, M. de Angelis, J. Beckers, M. H. p, S. M.
539 Hofmann, and S. Ussar. A history of obesity leaves an inflammatory fingerprint in liver and adipose
540 tissue. *Int J Obes (Lond)*, 42(3):507–517, Mar 2018.
- 541 [38] S. Chung, H. Cuffe, S. M. Marshall, A. L. McDaniel, J. H. Ha, K. Kavanagh, C. Hong, P. Tontonoz,
542 R. E. Temel, and J. S. Parks. Dietary cholesterol promotes adipocyte hypertrophy and adipose tissue
543 inflammation in visceral, but not in subcutaneous, fat in monkeys. *Arterioscler Thromb Vasc Biol*,
544 34(9):1880–1887, Sep 2014.
- 545 [39] V. Kus, T. Prazak, P. Brauner, M. Hensler, O. Kuda, P. Flachs, P. Janovska, D. Medrikova, M. Rossmeisl,
546 Z. Jilkova, B. Stefl, E. Pastalkova, Z. Drahota, J. Houstek, and J. Kopecky. Induction of muscle
547 thermogenesis by high-fat diet in mice: association with obesity-resistance. *Am J Physiol Endocrinol
548 Metab*, 295(2):E356–367, Aug 2008.
- 549 [40] C. B. Newgard. Interplay between lipids and branched-chain amino acids in development of insulin
550 resistance. *Cell Metab*, 15(5):606–614, May 2012.
- 551 [41] D. D. Sears, G. Hsiao, A. Hsiao, J. G. Yu, C. H. Courtney, J. M. Ofrecio, J. Chapman, and S. Subramaniam.

- 552 Mechanisms of human insulin resistance and thiazolidinedione-mediated insulin sensitization. *Proc Natl*
553 *Acad Sci U S A*, 106(44):18745–18750, Nov 2009.
- 554 [42] R. Stienstra, C. Duval, M. ller, and S. Kersten. PPARs, Obesity, and Inflammation. *PPAR Res*,
555 2007:95974, 2007.
- 556 [43] O. Gavrilova, M. Haluzik, K. Matsusue, J. J. Cutson, L. Johnson, K. R. Dietz, C. J. Nicol, C. Vinson,
557 F. J. Gonzalez, and M. L. Reitman. Liver peroxisome proliferator-activated receptor gamma contributes
558 to hepatic steatosis, triglyceride clearance, and regulation of body fat mass. *J Biol Chem*, 278(36):34268–
559 34276, Sep 2003.
- 560 [44] K. Matsusue, M. Haluzik, G. Lambert, S. H. Yim, O. Gavrilova, J. M. Ward, B. Brewer, M. L. Reitman,
561 and F. J. Gonzalez. Liver-specific disruption of PPARgamma in leptin-deficient mice improves fatty
562 liver but aggravates diabetic phenotypes. *J Clin Invest*, 111(5):737–747, Mar 2003.
- 563 [45] D. Patsouris, J. K. Reddy, M. ller, and S. Kersten. Peroxisome proliferator-activated receptor alpha
564 mediates the effects of high-fat diet on hepatic gene expression. *Endocrinology*, 147(3):1508–1516, Mar
565 2006.
- 566 [46] S. E. Schadinger, N. L. Bucher, B. M. Schreiber, and S. R. Farmer. PPARgamma2 regulates lipogenesis
567 and lipid accumulation in steatotic hepatocytes. *Am J Physiol Endocrinol Metab*, 288(6):E1195–1205,
568 Jun 2005.
- 569 [47] W. Motomura, M. Inoue, T. Ohtake, N. Takahashi, M. Nagamine, S. Tanno, Y. Kohgo, and T. Okumura.
570 Up-regulation of ADRP in fatty liver in human and liver steatosis in mice fed with high fat diet. *Biochem*
571 *Biophys Res Commun*, 340(4):1111–1118, Feb 2006.
- 572 [48] A. Srivastava, A. P. Morgan, M. L. Najarian, V. K. Sarsani, J. S. Sigmon, J. R. Shorter, A. Kashfeen,
573 R. C. McMullan, L. H. Williams, P. guez, M. T. Ferris, P. Sullivan, P. Hock, D. R. Miller, T. A. Bell,
574 L. McMillan, G. A. Churchill, and F. P. de Villena. Genomes of the Mouse Collaborative Cross. *Genetics*,
575 206(2):537–556, Jun 2017.
- 576 [49] D. W. Threadgill, D. R. Miller, G. A. Churchill, and F. P. de Villena. The collaborative cross: a
577 recombinant inbred mouse population for the systems genetic era. *ILAR J*, 52(1):24–31, 2011.
- 578 [50] A. Roberts, F. Pardo-Manuel de Villena, W. Wang, L. McMillan, and D. W. Threadgill. The poly-
579 morphism architecture of mouse genetic resources elucidated using genome-wide resequencing data:
580 implications for QTL discovery and systems genetics. *Mamm Genome*, 18(6-7):473–481, Jul 2007.
- 581 [51] G. A. Churchill, D. C. Airey, H. Allayee, J. M. Angel, A. D. Attie, J. Beatty, W. D. Beavis, J. K.

- 582 Belknap, B. Bennett, W. Berrettini, A. Bleich, M. Bogue, K. W. Broman, K. J. Buck, E. Buckler,
583 M. Burmeister, E. J. Chesler, J. M. Cheverud, S. Clapcote, M. N. Cook, R. D. Cox, J. C. Crabbe,
584 W. E. Crusio, A. Darvasi, C. F. Deschepper, R. W. Doerge, C. R. Farber, J. Forejt, D. Gaile, S. J.
585 Garlow, H. Geiger, H. Gershenfeld, T. Gordon, J. Gu, W. Gu, G. de Haan, N. L. Hayes, C. Heller,
586 H. Himmelbauer, R. Hitzemann, K. Hunter, H. C. Hsu, F. A. Iraqi, B. Ivandic, H. J. Jacob, R. C. Jansen,
587 K. J. Jepsen, D. K. Johnson, T. E. Johnson, G. Kempermann, C. Kendzierski, M. Kotb, R. F. Kooy,
588 B. Llamas, F. Lammert, J. M. Lassalle, P. R. Lowenstein, L. Lu, A. Lusis, K. F. Manly, R. Marcucio,
589 D. Matthews, J. F. Medrano, D. R. Miller, G. Mittelman, B. A. Mock, J. S. Mogil, X. Montagutelli,
590 G. Morahan, D. G. Morris, R. Mott, J. H. Nadeau, H. Nagase, R. S. Nowakowski, B. F. O'Hara, A. V.
591 Osadchuk, G. P. Page, B. Paigen, K. Paigen, A. A. Palmer, H. J. Pan, L. Peltonen-Palotie, J. Peirce,
592 D. Pomp, M. Pravenec, D. R. Prows, Z. Qi, R. H. Reeves, J. Roder, G. D. Rosen, E. E. Schadt, L. C.
593 Schalkwyk, Z. Seltzer, K. Shimomura, S. Shou, M. J. ä, L. D. Siracusa, H. W. Snoeck, J. L. Spearow,
594 K. Svenson, L. M. Tarantino, D. Threadgill, L. A. Toth, W. Valdar, F. P. de Villena, C. Warden,
595 S. Whatley, R. W. Williams, T. Wiltshire, N. Yi, D. Zhang, M. Zhang, and F. Zou. The Collaborative
596 Cross, a community resource for the genetic analysis of complex traits. *Nat Genet*, 36(11):1133–1137,
597 Nov 2004.
- 598 [52] J. Y. Huh, Y. J. Park, M. Ham, and J. B. Kim. Crosstalk between adipocytes and immune cells in
599 adipose tissue inflammation and metabolic dysregulation in obesity. *Mol Cells*, 37(5):365–371, May 2014.
- 600 [53] J. Lamb, E. D. Crawford, D. Peck, J. W. Modell, I. C. Blat, M. J. Wrobel, J. Lerner, J. P. Brunet,
601 A. Subramanian, K. N. Ross, M. Reich, H. Hieronymus, G. Wei, S. A. Armstrong, S. J. Haggarty,
602 P. A. Clemons, R. Wei, S. A. Carr, E. S. Lander, and T. R. Golub. The Connectivity Map: using
603 gene-expression signatures to connect small molecules, genes, and disease. *Science*, 313(5795):1929–1935,
604 Sep 2006.
- 605 [54] S. Amin, A. Lux, and F. O'Callaghan. The journey of metformin from glycaemic control to mTOR
606 inhibition and the suppression of tumour growth. *Br J Clin Pharmacol*, 85(1):37–46, Jan 2019.
- 607 [55] A. Kezic, L. Popovic, and K. Lalic. mTOR Inhibitor Therapy and Metabolic Consequences: Where Do
608 We Stand? *Oxid Med Cell Longev*, 2018:2640342, 2018.
- 609 [56] A. D. Barlow, M. L. Nicholson, and T. P. Herbert. -cells and a review of the underlying molecular
610 mechanisms. *Diabetes*, 62(8):2674–2682, Aug 2013.
- 611 [57] Y. Gu, J. Lindner, A. Kumar, W. Yuan, and M. A. Magnuson. Rictor/mTORC2 is essential for
612 maintaining a balance between beta-cell proliferation and cell size. *Diabetes*, 60(3):827–837, Mar 2011.

- 613 [58] E. B. Geer, J. Islam, and C. Buettner. Mechanisms of glucocorticoid-induced insulin resistance: focus on
614 adipose tissue function and lipid metabolism. *Endocrinol Metab Clin North Am*, 43(1):75–102, Mar 2014.
- 615 [59] J. X. Li and C. L. Cummins. Fresh insights into glucocorticoid-induced diabetes mellitus and new
616 therapeutic directions. *Nat Rev Endocrinol*, 18(9):540–557, Sep 2022.
- 617 [60] R. A. Lee, C. A. Harris, and J. C. Wang. Glucocorticoid Receptor and Adipocyte Biology. *Nucl Receptor
618 Res*, 5, 2018.
- 619 [61] S. Viengchareun, P. Penfornis, M. C. Zennaro, and M. s. Mineralocorticoid and glucocorticoid receptors
620 inhibit UCP expression and function in brown adipocytes. *Am J Physiol Endocrinol Metab*, 280(4):E640–
621 649, Apr 2001.
- 622 [62] J. Liu, X. Kong, L. Wang, H. Qi, W. Di, X. Zhang, L. Wu, X. Chen, J. Yu, J. Zha, S. Lv, A. Zhang,
623 P. Cheng, M. Hu, Y. Li, J. Bi, Y. Li, F. Hu, Y. Zhong, Y. Xu, and G. Ding. -HSD1 in regulating brown
624 adipocyte function. *J Mol Endocrinol*, 50(1):103–113, Feb 2013.
- 625 [63] L. E. Ramage, M. Akyol, A. M. Fletcher, J. Forsythe, M. Nixon, R. N. Carter, E. J. van Beek,
626 N. M. Morton, B. R. Walker, and R. H. Stimson. Glucocorticoids Acutely Increase Brown Adipose
627 Tissue Activity in Humans, Revealing Species-Specific Differences in UCP-1 Regulation. *Cell Metab*,
628 24(1):130–141, Jul 2016.
- 629 [64] J. L. Barclay, H. Agada, C. Jang, M. Ward, N. Wetzig, and K. K. Ho. Effects of glucocorticoids on
630 human brown adipocytes. *J Endocrinol*, 224(2):139–147, Feb 2015.
- 631 [65] M. ó, R. Gupte, W. L. Kraus, P. Pacher, and P. Bai. PARPs in lipid metabolism and related diseases.
632 *Prog Lipid Res*, 84:101117, Nov 2021.
- 633 [66] P. Bai, C. ó, H. Oudart, A. nszki, Y. Cen, C. Thomas, H. Yamamoto, A. Huber, B. Kiss, R. H.
634 Houtkooper, K. Schoonjans, V. Schreiber, A. A. Sauve, J. Menissier-de Murcia, and J. Auwerx. PARP-1
635 inhibition increases mitochondrial metabolism through SIRT1 activation. *Cell Metab*, 13(4):461–468,
636 Apr 2011.
- 637 [67] A. Chiarugi and M. A. Moskowitz. Cell biology. PARP-1—a perpetrator of apoptotic cell death? *Science*,
638 297(5579):200–201, Jul 2002.
- 639 [68] M. Althubiti, R. Almaimani, S. Y. Eid, M. Elzubaier, B. Refaat, S. Idris, T. A. Alqurashi, and M. Z.
640 El-Readi. BTK targeting suppresses inflammatory genes and ameliorates insulin resistance. *Eur Cytokine
641 Netw*, 31(4):168–179, Dec 2020.

- 642 [69] C. Skrabs, W. F. Pickl, T. Perkmann, U. ger, and A. Gessl. Rapid decline in insulin antibodies and
643 glutamic acid decarboxylase autoantibodies with ibrutinib therapy of chronic lymphocytic leukaemia. *J*
644 *Clin Pharm Ther*, 43(1):145–149, Feb 2018.
- 645 [70] Hans Hacker and Michael Karin. Regulation and function of ikk and ikk-related kinases. *Science's*
646 *STKE*, 2006(357):re13–re13, 2006.
- 647 [71] E. A. Oral, S. M. Reilly, A. V. Gomez, R. Meral, L. Butz, N. Ajluni, T. L. Chenevert, E. Korytnaya,
648 A. H. Neidert, R. Hench, D. Rus, J. F. Horowitz, B. Poirier, P. Zhao, K. Lehmann, M. Jain, R. Yu,
649 C. Liddle, M. Ahmadian, M. Downes, R. M. Evans, and A. R. Saltiel. and TBK1 Improves Glucose
650 Control in a Subset of Patients with Type 2 Diabetes. *Cell Metab*, 26(1):157–170, Jul 2017.
- 651 [72] M. C. Arkan, A. L. Hevener, F. R. Greten, S. Maeda, Z. W. Li, J. M. Long, A. Wynshaw-Boris, G. Poli,
652 J. Olefsky, and M. Karin. IKK-beta links inflammation to obesity-induced insulin resistance. *Nat Med*,
653 11(2):191–198, Feb 2005.
- 654 [73] M. Clark, C. J. Kroger, Q. Ke, and R. M. Tisch. The Role of T Cell Receptor Signaling in the
655 Development of Type 1 Diabetes. *Front Immunol*, 11:615371, 2020.
- 656 [74] P. ndell, A. C. Carlsson, A. Larsson, O. Melander, T. Wessman, J. v, and T. Ruge. TNFR1 is associated
657 with short-term mortality in patients with diabetes and acute dyspnea seeking care at the emergency
658 department. *Acta Diabetol*, 57(10):1145–1150, Oct 2020.
- 659 [75] A. M. Keestra-Gounder, M. X. Byndloss, N. Seyffert, B. M. Young, A. vez Arroyo, A. Y. Tsai, S. A.
660 Cevallos, M. G. Winter, O. H. Pham, C. R. Tiffany, M. F. de Jong, T. Kerrinnes, R. Ravindran, P. A.
661 Luciw, S. J. McSorley, A. J. umler, and R. M. Tsolis. NOD1 and NOD2 signalling links ER stress with
662 inflammation. *Nature*, 532(7599):394–397, Apr 2016.
- 663 [76] A. M. Keestra-Gounder and R. M. Tsolis. NOD1 and NOD2: Beyond Peptidoglycan Sensing. *Trends*
664 *Immunol*, 38(10):758–767, Oct 2017.
- 665 [77] J. Montane, L. Cadavez, and A. Novials. Stress and the inflammatory process: a major cause of
666 pancreatic cell death in type 2 diabetes. *Diabetes Metab Syndr Obes*, 7:25–34, 2014.
- 667 [78] B. B. Kahn and T. E. McGraw. , and type 2 diabetes. *N Engl J Med*, 363(27):2667–2669, Dec 2010.
- 668 [79] S. Del Prato and N. Pulizzi. The place of sulfonylureas in the therapy for type 2 diabetes mellitus.
669 *Metabolism*, 55(5 Suppl 1):S20–27, May 2006.

- 670 [80] X. Liu, Y. I. Li, and J. K. Pritchard. Trans Effects on Gene Expression Can Drive Omnipotent Inheritance.
671 *Cell*, 177(4):1022–1034, May 2019.
- 672 [81] B. Hallgrímsson, R. M. Green, D. C. Katz, J. L. Fish, F. P. Bernier, C. C. Roseman, N. M. Young,
673 J. M. Cheverud, and R. S. Marcucio. The developmental-genetics of canalization. *Semin Cell Dev Biol*,
674 88:67–79, Apr 2019.
- 675 [82] M. L. Siegal and A. Bergman. Waddington’s canalization revisited: developmental stability and evolution.
676 *Proc Natl Acad Sci U S A*, 99(16):10528–10532, Aug 2002.
- 677 [83] A. B. Paaby and G. Gibson. Cryptic Genetic Variation in Evolutionary Developmental Genetics. *Biology*
678 (*Basel*), 5(2), Jun 2016.
- 679 [84] E. A. Boyle, Y. I. Li, and J. K. Pritchard. An Expanded View of Complex Traits: From Polygenic to
680 Omnipotent. *Cell*, 169(7):1177–1186, Jun 2017.
- 681 [85] Naomi R Wray, Cisca Wijmenga, Patrick F Sullivan, Jian Yang, and Peter M Visscher. Common disease
682 is more complex than implied by the core gene omnipotent model. *Cell*, 173(7):1573–1580, 2018.