

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 Gregory
⁶ W Carter

⁷ **Abstract**

⁸ Gene expression is an important mediator of genetic effects on phenotype. Although many genes are subject
⁹ to simple, local regulation, recent evidence suggests that complex distal regulation may be more important
¹⁰ in mediating trait variability. To investigate this possibility, we combined two large, data sets modeling
¹¹ diet-induced obesity and metabolic disease in genetically diverse mice. Using a novel high-dimensional
¹² mediation analysis, we identified a heritable composite transcript that explained 30% of the variation across
¹³ all metabolic traits. The composite transcript was interpretable in terms of enriched biological processes
¹⁴ and predicted obesity status in an independent mouse cohort as well as in human cohorts with measured
¹⁵ gene expression. Transcripts contributing most strongly to this composite mediator tended to have complex,
¹⁶ distal regulation distributed throughout the genome. These results suggest that trait-relevant variation in
¹⁷ transcription is largely distally regulated, but is nonetheless identifiable, interpretable, and translatable across
¹⁸ species.

¹⁹ **Introduction**

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

²⁶ 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}.

³⁵ In this model, a gene's expression is influenced by many variants throughout the genome through their
³⁶ cumulative effects on a broader regulatory network. In other words, the heritable component of the
³⁷ transcriptome is an emergent state arising from the myriad molecular interactions defining and constraining
³⁸ gene expression.

³⁹ To assess the role of wide-spread distal gene regulation on complex traits, we investigated diet-induced
⁴⁰ obesity and metabolic disease as an archetypal example. Diet-induced obesity and metabolic disease are
⁴¹ genetically complex with hundreds of variants mapped through GWAS^{21;22}. These variants are known to act
⁴² through multiple tissues that interact dynamically with each other^{23;24}, including adipose tissue, pancreatic
⁴³ islets, liver, and skeletal muscle. The multi-system etiology of metabolic disease complicates mechanistic
⁴⁴ dissection of the genetic architecture, requiring large, dedicated data sets that include high-dimensional,
⁴⁵ clinically relevant phenotyping, dense genotyping in a highly recombined population, and transcriptome-wide
⁴⁶ measurements of gene expression in multiple tissues.

⁴⁷ Measuring gene expression in multiple tissues is critical to adequately assess the extent to which local gene
⁴⁸ regulation varies across the tissues and whether such variability might account for previous failed attempts to
⁴⁹ identify trait-relevant local eQTL. Such data sets are extremely difficult to obtain in human populations,
⁵⁰ particularly in the large numbers of subjects required for adequate statistical power. Thus, to further
⁵¹ investigate the role of local and distal gene regulation on complex traits, we generated two complementary
⁵² data sets: A discovery data set in a large population of diversity outbred (DO) mice²⁵, and an independent
⁵³ validation data set derived by crossing inbred strains from the Collaborative Cross (CC) mice²⁶ to form CC
⁵⁴ F1 mice (CC-RIX). Both populations modeled diet-induced obesity and metabolic disease¹²

⁵⁵ The DO population and CC recombinant inbred lines were derived from the same eight inbred founder mouse
⁵⁶ strains, five classical lab strains, and three strains more recently derived from wild mice²⁵. They represent

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

65 In the DO population, we paired clinically relevant metabolic traits from 500 mice¹², including body weight,
66 plasma levels of insulin and glucose and plasma lipids, with transcriptome-wide gene expression in four
67 tissues related to metabolic disease: adipose tissue, pancreatic islets, liver, and skeletal muscle. We measured
68 similar metabolic traits in the CC-RIX and gene expression from three of the four tissues used in the DO:
69 adipose tissue, liver, and skeletal muscle. Because the CC-RIX carry the same founder alleles as the DO,
70 local gene regulation is expected to match between the populations, but because the alleles are recombinant
71 through the genome, distal effects are expected to vary from those in the DO, allowing us to directly assess
72 the role of local gene regulation in driving trait-associated transcript variation. Together, these data enable a
73 comprehensive view into the genetic architecture of metabolic disease.

74 Results

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

77 Genetic variation contributed to wide phenotypic variation

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

87 pre-diabetes and diabetes.

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

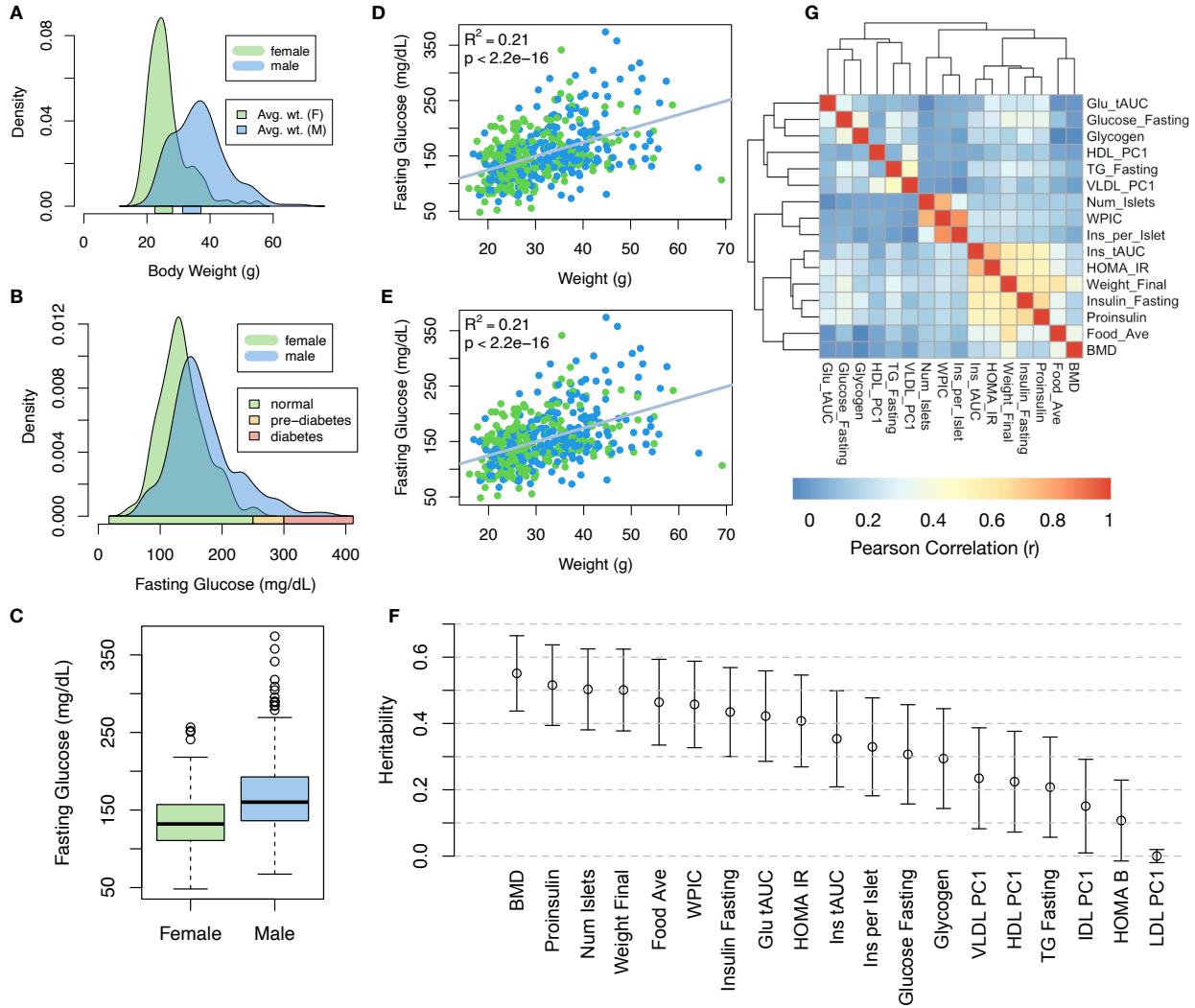


Figure 1: Clinical overview. **A.** Distributions of final body weight in the diversity outbred mice. Sex is indicated by color. The average B6 male and female adult weights at 24 weeks of age are indicated by blue and green bars on the x-axis. **B.** The distribution of final fasting glucose across 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 physiological traits.

92 The trait correlations (Fig. 1G) showed that most of the metabolic trait pairs were only modestly correlated

93 indicating complex relationships among the measured traits. This low level of redundancy suggests a broad
94 sampling of multiple heritable aspects of metabolic disease including overall body weight, glucose homeostasis,
95 pancreatic composition and liver function.

96 **Distal Heritability Correlated with Phenotype Relevance**

97 To comprehensively assess the genetic control of gene expression in metabolic disease we assayed adipose, islet,
98 liver, and skeletal muscle gene expression in the DO cohort. We performed eQTL analysis using R/qltl²⁹
99 (Methods) and identified both local and distal eQTLs for transcripts in each of the four tissues (Fig. S1).
100 Significant local eQTLs far outnumbered distal eQTLs (Fig. S1F) and tended to be shared across tissues
101 (Fig. S1G) whereas the few significant distal eQTLs we identified tended to be tissue-specific (Fig. S1H)
102 We calculated the heritability of each transcript in terms of local and distal genetic factors (Methods). Overall,
103 local and distal genetic factors contributed approximately equally to transcript abundance. In all tissues,
104 both local and distal factors explained between 8 and 18% of the variance in the median transcript (Fig 2A).
105 To assess the importance of genetic regulation transcript levels to organism-level traits, we compared the
106 local and distal heritabilities of transcripts to their trait relevance, defined as the maximum correlation
107 of a transcript across all traits. The local heritability of transcripts was negatively correlated with their
108 trait relevance (Fig. 2B), suggesting that the more local genotype influenced transcript abundance, the
109 less effect this variation had on the measured traits. Conversely, the distal heritability of transcripts was
110 positively correlated with trait relevance (Fig. 2C). That is, transcripts that were more highly correlated
111 with the measured traits tended to be distally, rather than locally, heritable. Importantly, this pattern was
112 consistent across all tissues, strongly suggesting that this is a generic finding. This finding is consistent with
113 previous observations that low-heritability transcripts explain more expression-mediated disease heritability
114 than high-heritability transcripts¹⁹. However, the positive relationship between trait correlation and distal
115 heritability demonstrated further that there are diffuse genetic effects throughout the genome converging on
116 trait-related transcripts.

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

119 The above univariate analyses establish the importance of distal heritability for trait-relevant transcripts.
120 However, the number of transcripts dramatically exceeds the number of phenotypes. Thus, we expect the
121 heritable, trait-relevant transcripts to be highly correlated and organized according to coherent, emergent
122 biological processes representing the mediating endophenotypes driving clinical trait variation. To identify

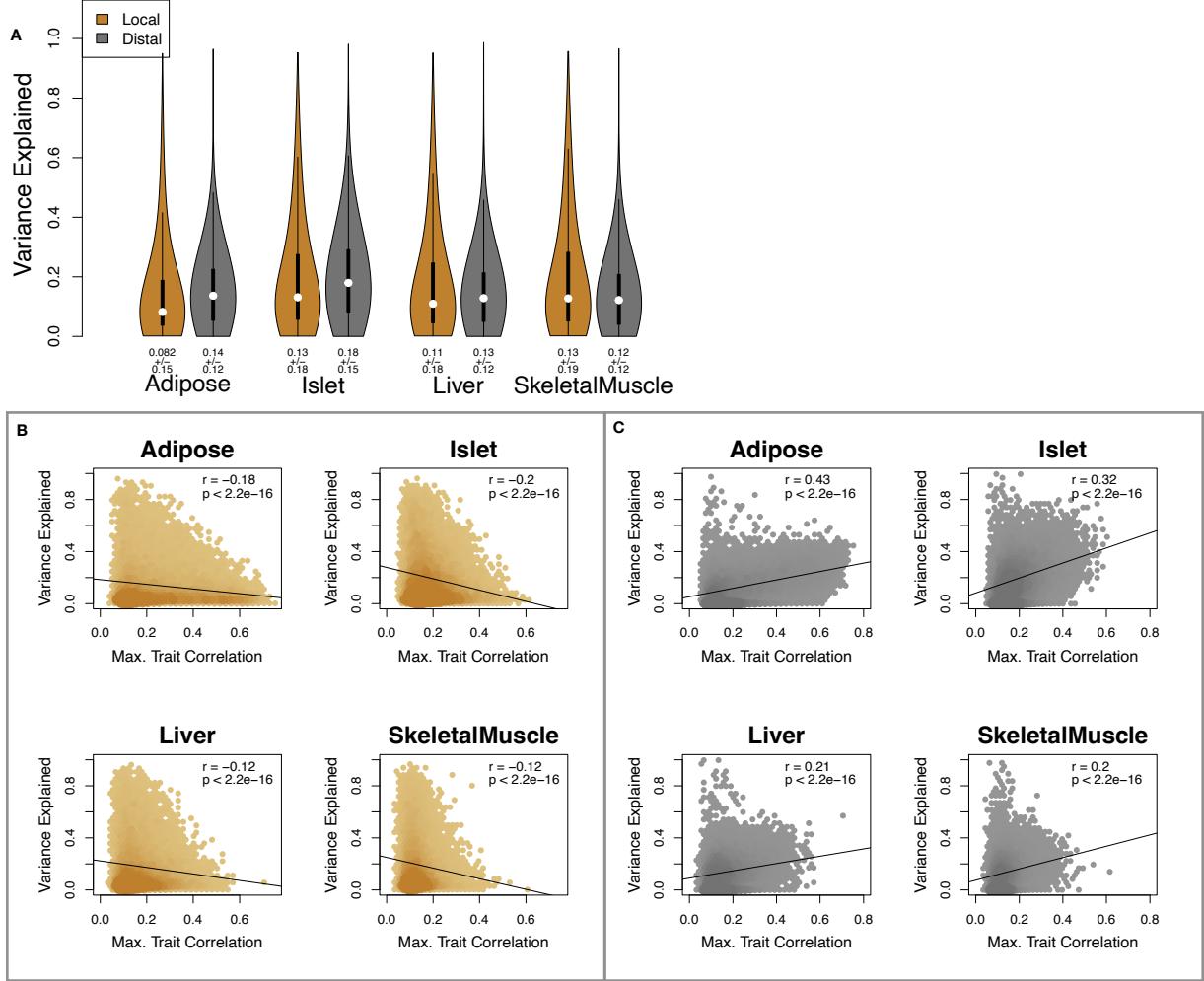


Figure 2: Transcript heritability and trait relevance. **A.** Distributions of distal and local heritability of transcripts across the four tissues. Overall local and distal factors contribute equally to transcript heritability. The relationship between **(B.)** local and **(C.)** distal heritability 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.

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). HDMA projects the high-dimensional scores—a composite genome score (G_C), a composite transcriptome score (T_C), and a composite phenotype 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 phenotype 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 XXX**.

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. 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,

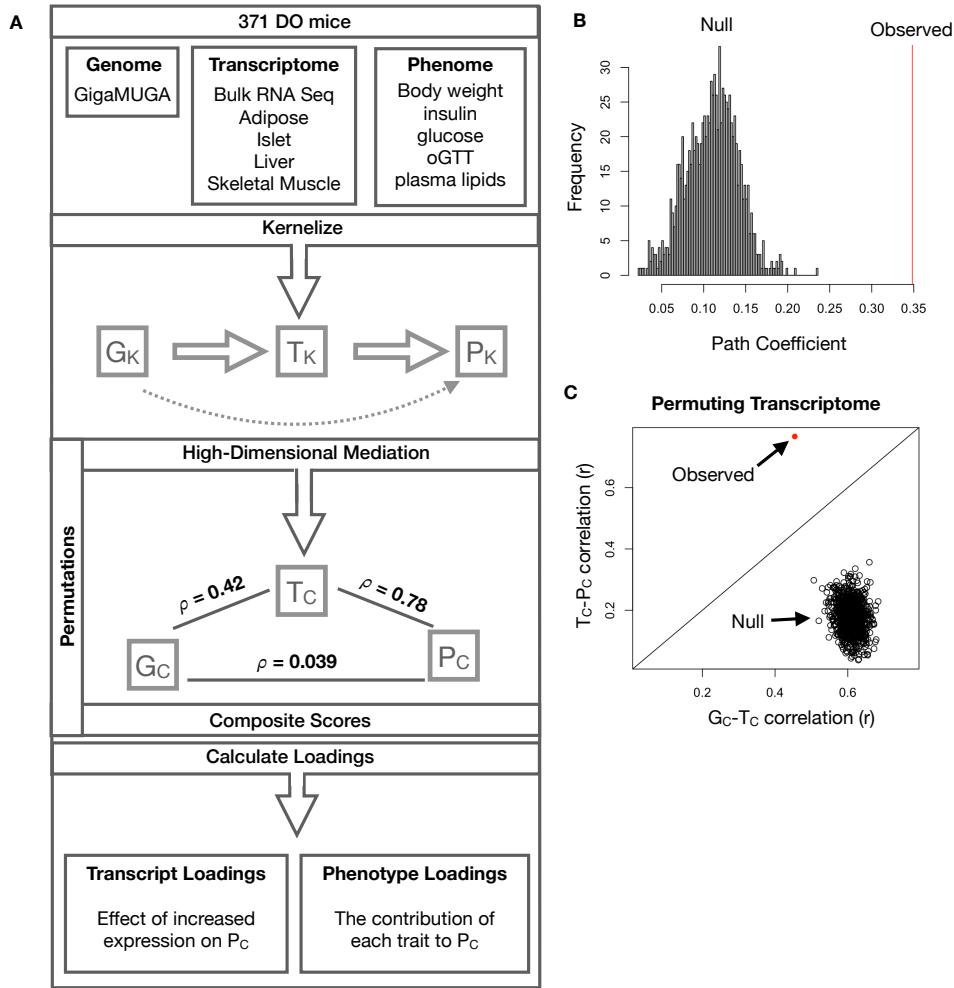


Figure 3: High-dimensional mediation. **A.** Workflow indicating major steps of high-dimensional mediation. The genotype, transcriptome, and phenotype matrices were independently normalized and converted to kernel matrices representing the pairwise relationships between individuals for each data modality (K_G = genome kernel, K_T = transcriptome kernel; K_P = phenome 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 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
 161 maximized simultaneously suggesting that the path from genotype to phenotype through transcriptome is
 162 highly non-trivial and identifiable in this case. These results suggest that these composite vectors represent
 163 genetically determined variation in phenotype that is mediated through genetically determined variation in
 164 transcription.

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

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

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

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

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

193 The strongest negative enrichments in adipose tissue were related to mitochondrial activity in general, and
194 thermogenesis in particular (Figs. S2 and S2). Genes in the KEGG oxidative phosphorylation pathway in
195 mice were almost universally negatively loaded in adipose tissue, suggesting that increased expression of these

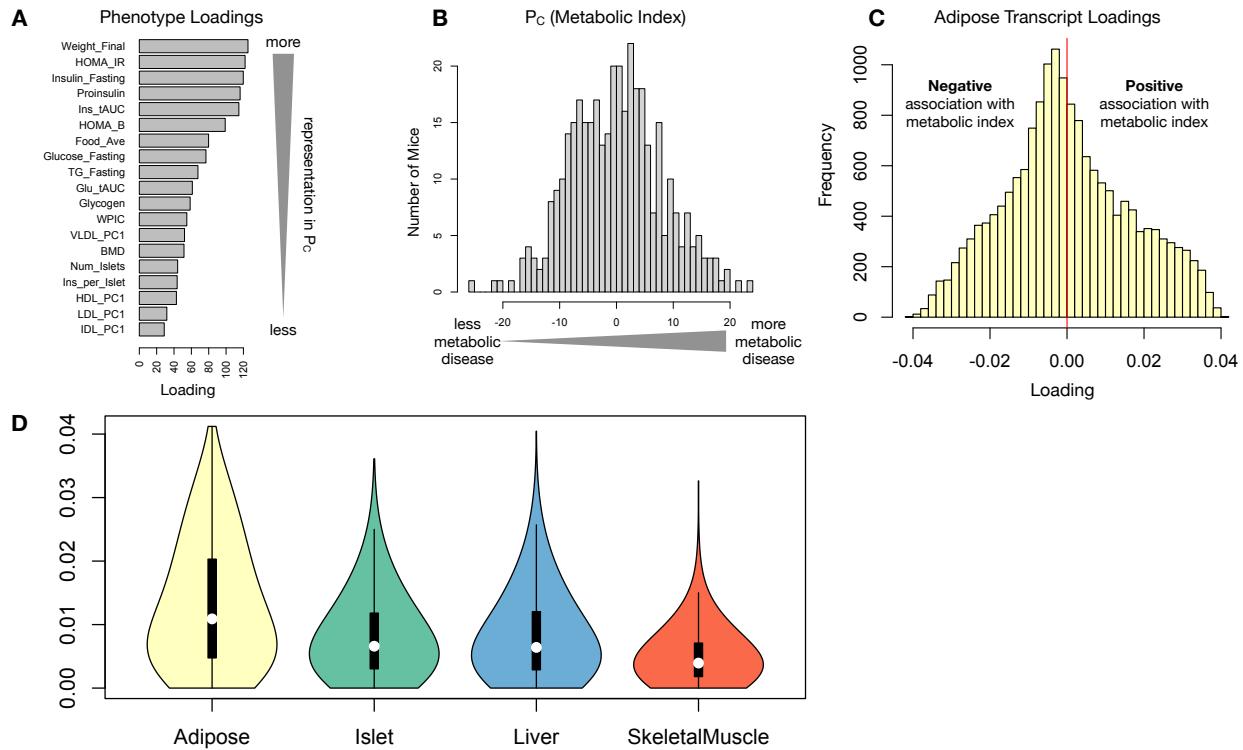


Figure 4: Interpretation of loadings. **A.** Loadings across traits. Body weight and insulin resistance contributed the most to the composite trait. **B.** Phenotype scores across individuals. Individuals with large positive phenotype scores had higher body weight and insulin resistance than average.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

196 genes was associated with reduced metabolic index (Supp. Fig. S4). Consistent with this observations, it
 197 has been shown previously that mouse strains with greater thermogenic potential are also less susceptible to
 198 obesity on a high-fat diet³⁹.

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

204 Looking a the 10 most positively and negatively loaded transcripts from each tissue, it is apparent that
 205 transcripts in the adipose tissue had the largest loadings, both positive and negative, of all tissues (Fig. 5A

206 bar plot) This suggests that much of the effect of genetics on body weight and insulin resistance is mediated
207 through gene expression in adipose tissue. The strongest loadings in liver and pancreas were comparable,
208 and those in skeletal muscle were the weakest (Fig. 5A), suggesting that less of the genetic effects were
209 mediated through transcription in skeletal muscle. Heritability analysis showed that transcripts with the
210 largest loadings had higher distal heritability than local heritability (Fig. 5A heat map and box plot). This
211 pattern contrasts with transcripts nominated by TWAS (Fig. 5B), which tended to have lower loadings,
212 higher local heritability and lower distal heritability. Transcripts with the highest local heritability in each
213 tissue (Fig. 5C) had the lowest loadings, consistent with our findings above (Fig. 2B).

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

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

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

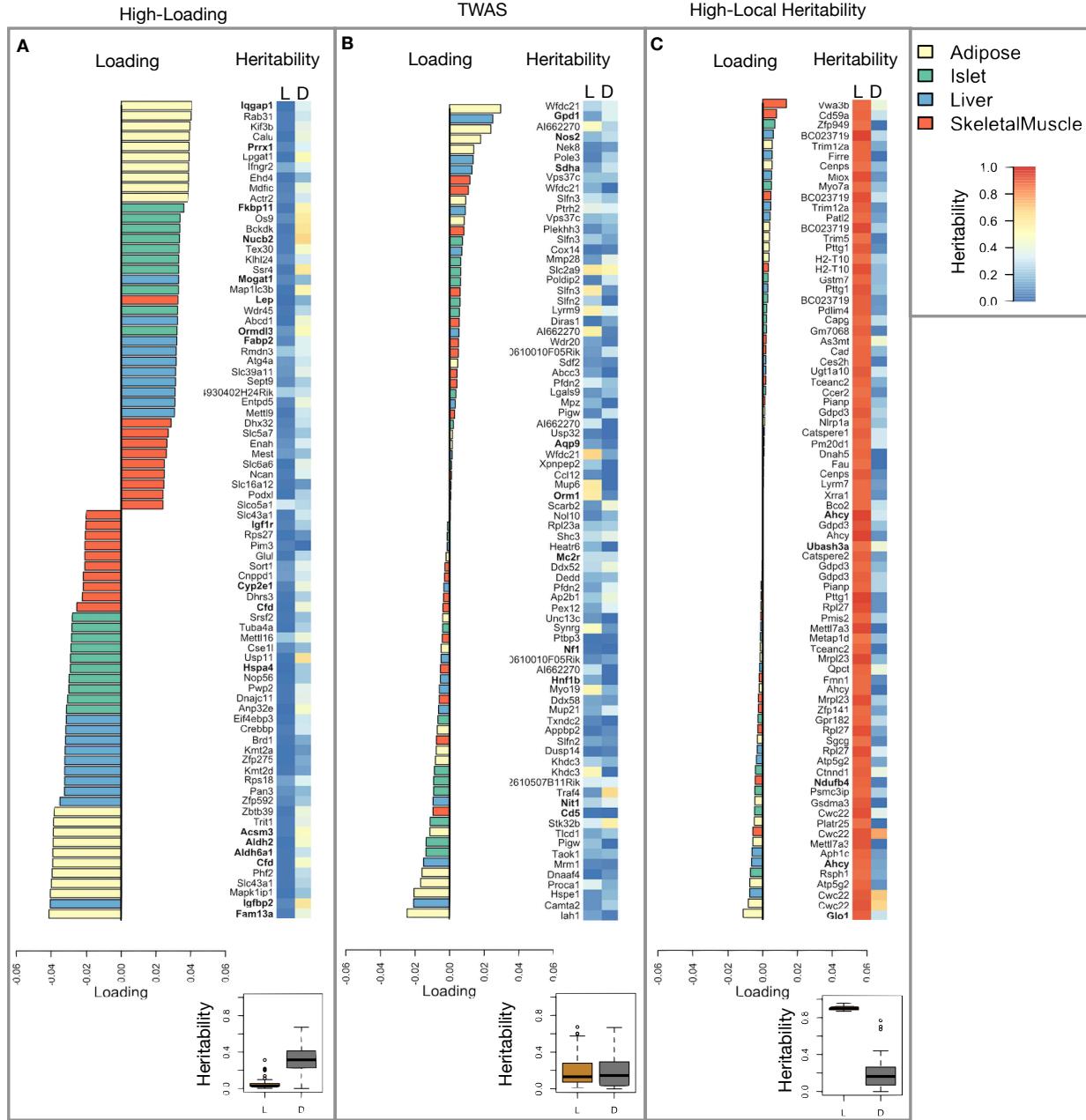


Figure 5: Transcripts with high loadings have high distal heritability and literature support. Each panel has a bar plot showing the loadings of transcripts selected by different criteria. Bar color indicates the tissue of origin. The heat map shows the local (L - left) and distal (D - right) heritability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

237 The local and distal heritability of *Pparg* is low in adipose tissue suggesting its expression in this tissue is
 238 highly constrained in the population (Fig. 6B). However, the distal heritability of *Pparg* in liver is relatively
 239 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 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.

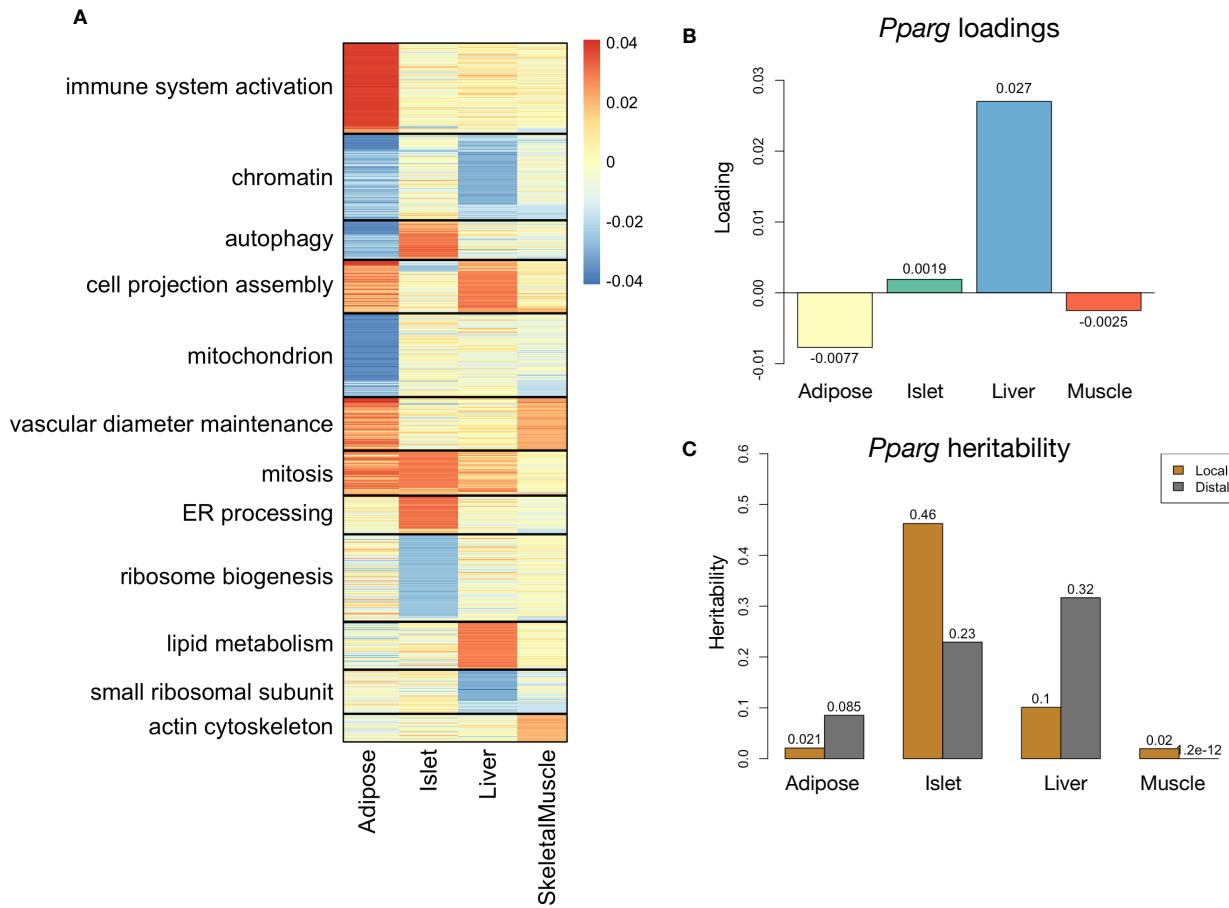


Figure 6: Tissue-specific transcriptional programs were associated with obesity and insulin resistance. **A** Heat map showing the loadings of all transcripts with loadings greater than 2.5 standard deviations from the mean in any tissue. The heat map was clustered using k medoid clustering. Functional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

245 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)^{48–51} strains (Fig. 7) (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 phenotype 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 actual 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 metabolic index. This result also confirms the validity and translatability of the transcript loadings and their relationship to metabolic disease.

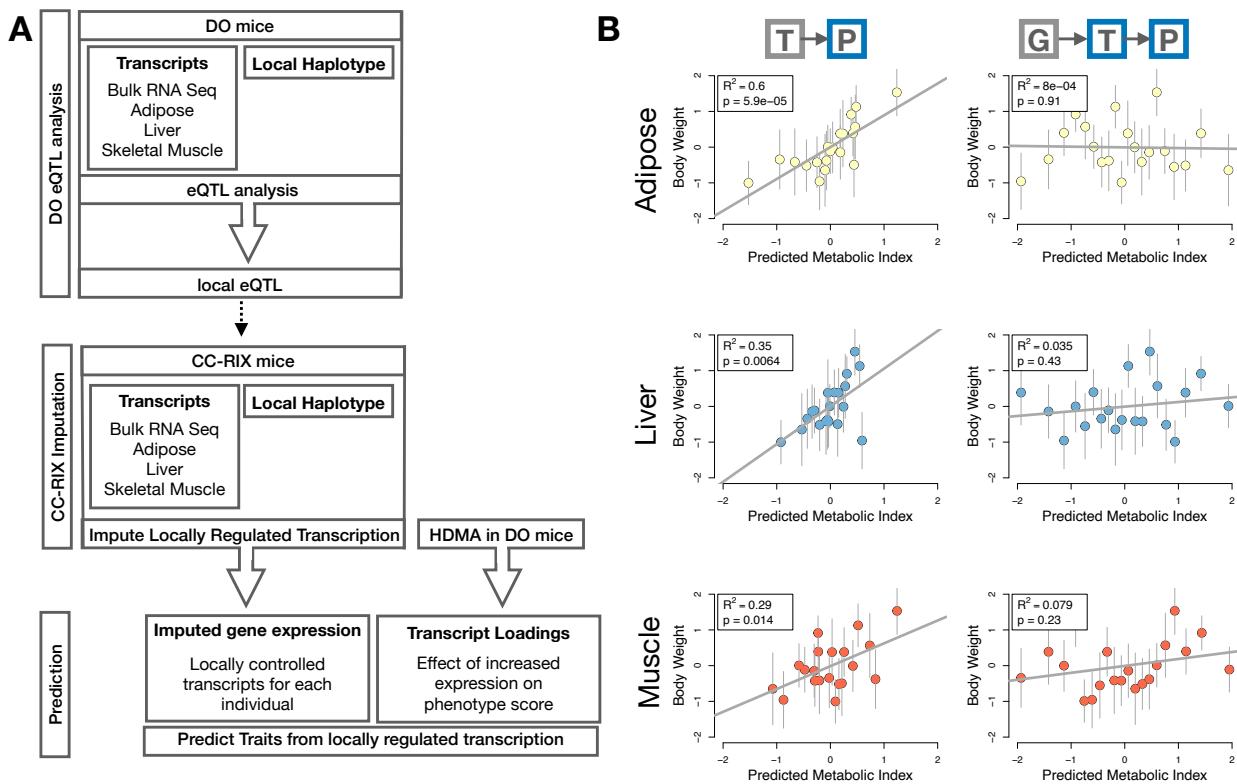


Figure 7: Transcription, but not local genotype, predicts phenotype in the CC-RIX. **A.** Workflow showing procedure for translating HDMA results to an independent population of mice. **B.** Relationships between the predicted metabolic index and measured body weight. The left column shows the predictions 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. 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.

262 Fig. S7). However, these imputed values failed to predict body weight in the CC-RIX when weighted with the
263 loadings from HDMA. (Fig. 7B right column). This result suggests that local regulation of gene expression is
264 not the primary factor driving heritability of complex traits, consistent with our findings in the DO population
265 that distal heritability was a major driver of trait-relevant variation and that high-loading transcripts had
266 comparatively high distal and low local heritability.

267 **Distally heritable transcriptomic signatures reflected variation in composition of adipose tissue
268 and islets**

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

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

290 Notably, the loadings for pancreatic beta cell-type specific loadings was not significantly different from zero.
291 We stress that this is not necessarily reflective of the function of the beta cells in the obese mice, but rather
292 suggests that any variation in the number of beta cells in these mice was unrelated to obesity and insulin

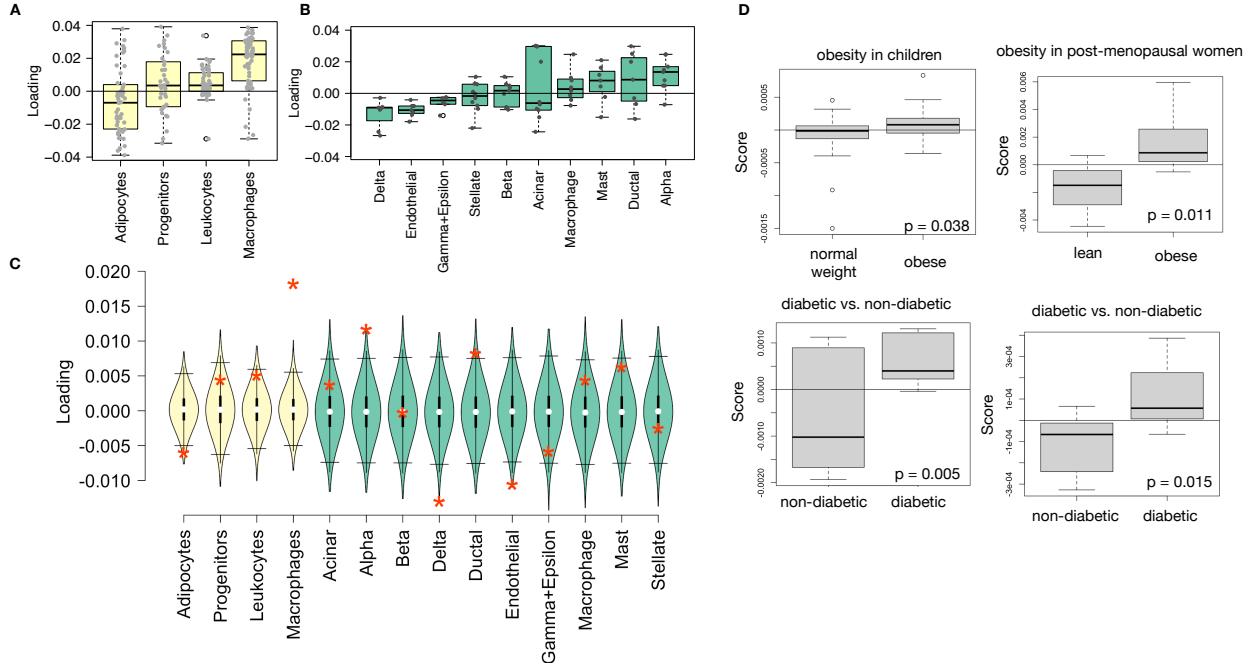


Figure 8: HDMA results translate to humans. **A.** Distribution of loadings for cell-type-specific transcripts in adipose tissue. **B.** Distribution of loadings for cell-type-specific transcripts in pancreatic islets (green). **C.** Null distributions for the mean loading of randomly selected transcripts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

resistance, the major contributors to metabolic index. This is further consistent with the islet composition traits having small loadings in the phenotype score (Fig. 4).

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. 8D). This indicates that the distally heritable signature of metabolic index identified in DO mice is relevant to obesity

306 and diabetes in human subjects.

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

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

315 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
316 (Supp. Figs S8 and S9). To get more tissue-specific results, we also looked at top results in cell types that
317 most closely resembled our tissues. We looked at results in adipocytes (ASC) as well as pancreatic tumor
318 cells (YAPC) regardless of *p* value (Supp. Figs S10 and S11).

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

329 The top hit for the adipose composite transcript in CMAP adipocytes was a PARP inhibitor (Supp. Fig.
330 S10). PARPs play a role in lipid metabolism and are involved in the development of obesity and diabetes⁶⁵.
331 PARP1 inhibition increases mitochondrial biogenesis⁶⁶. Inhibition of PARP1 activity can further prevent
332 necrosis in favor of the less inflammatory apoptosis⁶⁷, thereby potentially reducing inflammation in stressed
333 adipocytes. Other notable hits among the top 20 were BTK inhibitors, which have been observed to suppress
334 inflammation and improve insulin resistance⁶⁸ as well as to reduce insulin antibodies in type I diabetes⁶⁹.
335 IKK inhibitors have been shown to improve glucose control in type II diabetes^{70;71}.

336 Among the top most significant hits for the transcript loadings from pancreatic islets (Supp. Fig. S9), was
337 suppression of T cell receptor signaling, which is known to be involved in Type 1 diabetes⁷², as well as
338 TNFR1, which has been associated with mortality in diabetes patients⁷³. Suppression of NOD1/2 signaling
339 was also among the top hits. NOD1 and 2 sense ER stress^{74;75}, which is associated with β -cell death in type
340 1 and type 2 diabetes⁷⁶. This cell death process is dependent on NOD1/2 signaling⁷⁴, although the specifics
341 have not yet been worked out.

342 We also looked specifically at hits in pancreatic tumor cells (YAPC) regardless of significance level to get a
343 transcriptional response more specific to the pancreas (Supp. Fig. S11). Hits in this list included widely used
344 diabetes drugs, such as sulfonylureas, PPAR receptor agonists, and insulin sensitizers. Rosiglitazone is a
345 PPAR- γ agonist and was one of the most prescribed drugs for type 2 diabetes before its use was reduced due
346 to cardiac side-effects⁷⁷. Sulfonylureas are another commonly prescribed drug class for type 2 diabetes, but
347 also have notable side effects including hypoglycemia and accelerated β -cell death⁷⁸.

348 Discussion

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

362 Genetics is indispensable for the dissection of disease mechanisms because it is one of the only data modalities
363 that supports causal inferences about molecules and disease outcomes^{79;80}. It has frequently been assumed
364 that gene regulation in *cis* is the primary driver of genetically associated trait variation, but attempts to use
365 local gene regulation to explain phenotypic variation have had limited success^{16;17}. In recent years, evidence
366 has mounted that distal gene regulation may be an important mediator of trait heritability^{19;18;81}. It has

367 been observed that transcripts with high local heritability explain less expression-mediated disease heritability
368 than those with low local heritability¹⁹. Consistent with this observation, genes located near GWAS hits
369 tend to be complexly regulated¹⁸. They also tend to be enriched with functional annotations, in contrast
370 to genes with simple local regulation, which tend to be depleted of functional annotations suggesting they
371 are less likely to be directly involved in disease traits¹⁸. These observations are consistent with principles
372 of robustness in complex systems in which simple regulation of important elements leads to fragility of the
373 system^{82–84}. Our results are consistent, instead, with a more complex picture where genes whose expression
374 can drive trait variation are buffered from local genetic variation but are extensively influenced indirectly by
375 genetic variation in the regulatory networks converging on those genes.

376 Recently, the omnigenic model of complex traits has been proposed, which posits that complex traits are
377 massively polygenic and that their heritability is spread out across the genome⁸⁵. In the omnigenic model,
378 genes are classified either as “core genes,” which directly impinge on the trait, or “peripheral genes,” which
379 are not directly trait-related, but influence core genes through the complex gene regulatory network. HDMA
380 explicitly models a central proposal of the omnigenic model which posits that once the expression of the
381 core genes (i.e. trait-mediating genes) is accounted for, there should be no residual correlation between the
382 genome and the phenotype. Here, when the composite transcript was taken into account there was no residual
383 correlation between the composite genome and composite phenotype (Fig. 3A).

384 Thus, the transcript loadings can be interpreted as indicating higher “core-ness” of a transcript. Unlike in the
385 omnigenic model, we did not observe a clear demarcation between the core and peripheral genes in loading
386 magnitude, but we do not necessarily expect a clear separation given the complexity of gene regulation and
387 the genotype-phenotype map⁸⁶.

388 An extension of the omnigenic model proposed that most heritability of complex traits is driven by weak
389 distal eQTLs that are potentially below the detection threshold in studies with feasible sample sizes⁸¹. This
390 is consistent with what we observed here. The transcripts with the largest loadings were strongly distally
391 regulated and only weakly locally regulated, suggesting that distal gene regulation plays a primary role in
392 driving heritable trait variation. We saw further that the patterns of distal heritability were not localized to
393 detectable distal eQTL, but rather were complex and spread across the genome, even for transcripts whose
394 expression was strongly regulated by distal factors. For example, *Nucb2*, had a high loading in islets and
395 was also strongly distally regulated (66% distal heritability) (Fig. 5). This gene is expressed in pancreatic β
396 cells and is involved in insulin and glucagon release^{87–89}. Although its transcription was highly heritable in
397 islets, that regulation was distributed across the genome, with no clear distal eQTL (Supp. Fig. S12). Thus,
398 although distal regulation of some genes may be strong, this regulation is likely to be highly complex and not

399 easily localized.

400 We stress that HDMA is a method for causal hypothesis generation. As with any causal inference approach,
401 the output of HDMA can only be said to be consistent with causal mediation but does not prove it.
402 Proving causality requires experimentation with direct control over the mediating variable⁹⁰. The issue
403 of experimentation, however, is subtle. The dimension-reduction in HDMA is distinguished by the fact
404 that the putative causal intermediates can be emergent states defined by the expression of thousands of
405 genes. This is a strength, because the mediating variable can be a higher-order process such as “macrophage
406 activation and infiltration”, but, in contrast to univariate hypotheses at the level of individual transcripts, the
407 relevant validation experiment may be technologically infeasible, unknowable *a priori*, or both. Nevertheless,
408 downstream analyses of the composite transcripts strongly supports a causal interpretation. Indeed, the
409 composite transcripts identified by HDMA are richly interpretable in both tissue- and gene-specific manners.
410 The transcripts with the strongest loadings were enriched in biological functions previously known to be
411 involved in the pathogenesis of metabolic disease, such as inflammation in adipose tissue. That these processes
412 were identified in this analysis suggests additionally that they have a heritable component, and that some
413 individuals are genetically susceptible to greater adipose inflammation on a high-fat, high-sugar diet.

414 Individual high-loading transcripts also demonstrated biologically interpretable, tissue-specific patterns. We
415 highlighted *Pparg*, which is known to be protective in adipose tissue⁴² where it was negatively loaded, and
416 harmful in the liver^{43–47}, where it was positively loaded. Such granular patterns may be useful in generating
417 hypotheses for further testing, and prioritizing genes as therapeutic targets. The tissue-specific nature of
418 the loadings also may provide clues to tissue-specific effects, or side effects, of targeting particular genes
419 system-wide.

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

426 Another useful application of the composite transcripts is to pair them with cell-type specific genes to generate
427 causal hypotheses about changes in cell composition in individual tissues. Combining the multi-tissue,
428 transcriptome-wide weighted vectors with public databases and data sets thus provides a path for generating
429 a wide range of testable hypotheses. Moreover, each publicly available data set we used for interpretation of

430 the HDMA results was derived from human tissues or cell lines, thus demonstrating the translatability of the
431 HDMA results to humans. That the mouse-derived adipose composite transcript was able to classify human
432 adipose gene expression in terms of obesity and diabetes status further supports the direct translatablility of
433 these findings, the utility of HDMA, and the continued importance of mouse models of human disease in
434 which it is possible to obtain complete transcriptomes in mutliple tissues across large numbers of individuals.

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

441 **Data Availability**

442 Here we tell people where to find the data

443 **Acknowledgements**

444 Here we thank people

445 Supplemental Figures

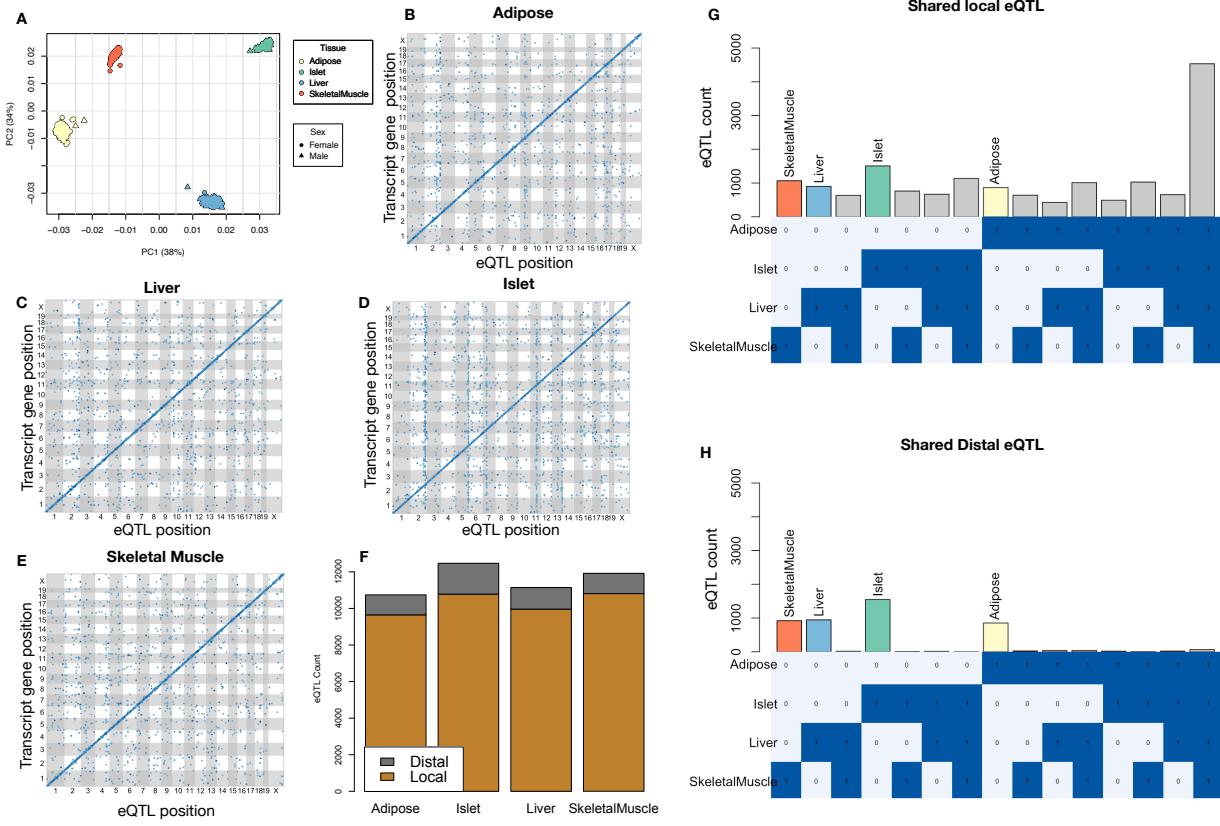


Figure S1: Overview of eQTL analysis in DO mice. **A.** RNA seq samples from the four different tissues clustered by tissue. **B.-E.** eQTL maps are shown for each tissue. The *x*-axis shows the position of the mapped eQTL, and the *y*-axis shows the physical position of the gene encoding each mapped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

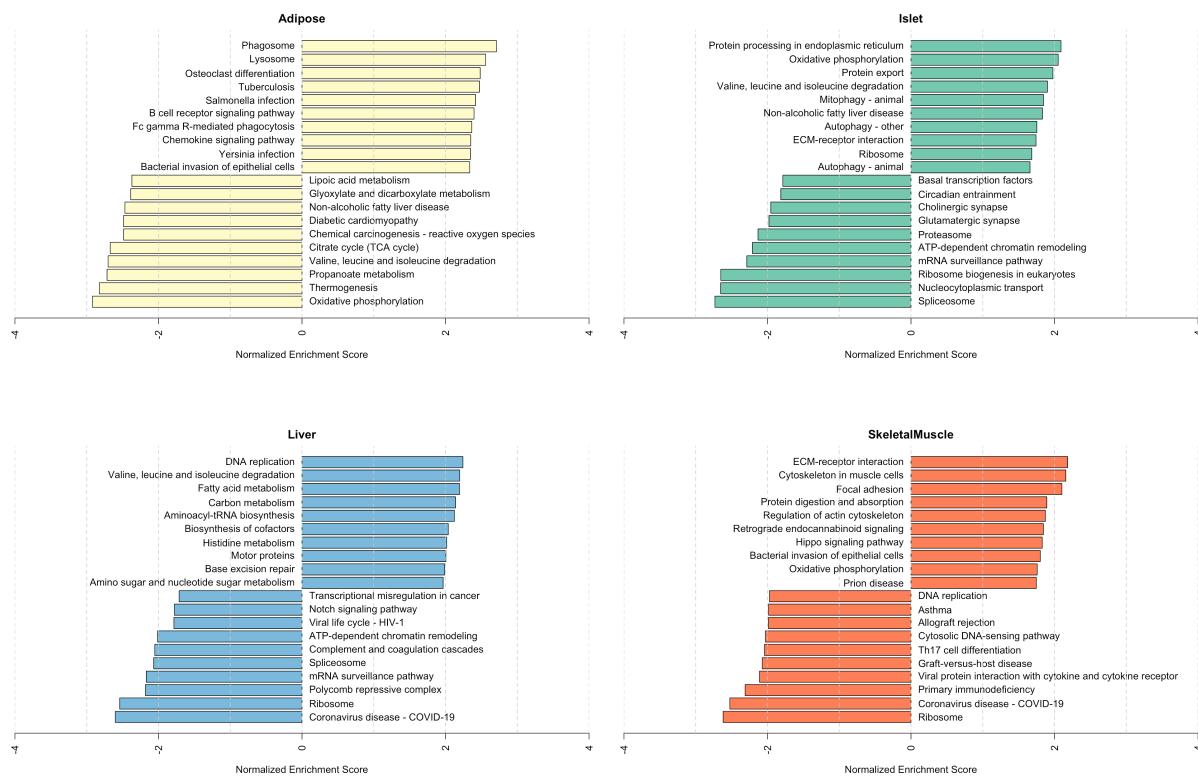


Figure S2: Bar plots showing normalized enrichment scores (NES) for KEGG pathways as determined by fast gene score enrichment analysis (fgsea). Only the top 10 positive and top 10 negative scores are shown. Colors indicate tissue. The name beside each bar shows the name of each enriched KEGG pathway.

Top GO term enrichments by GSEA

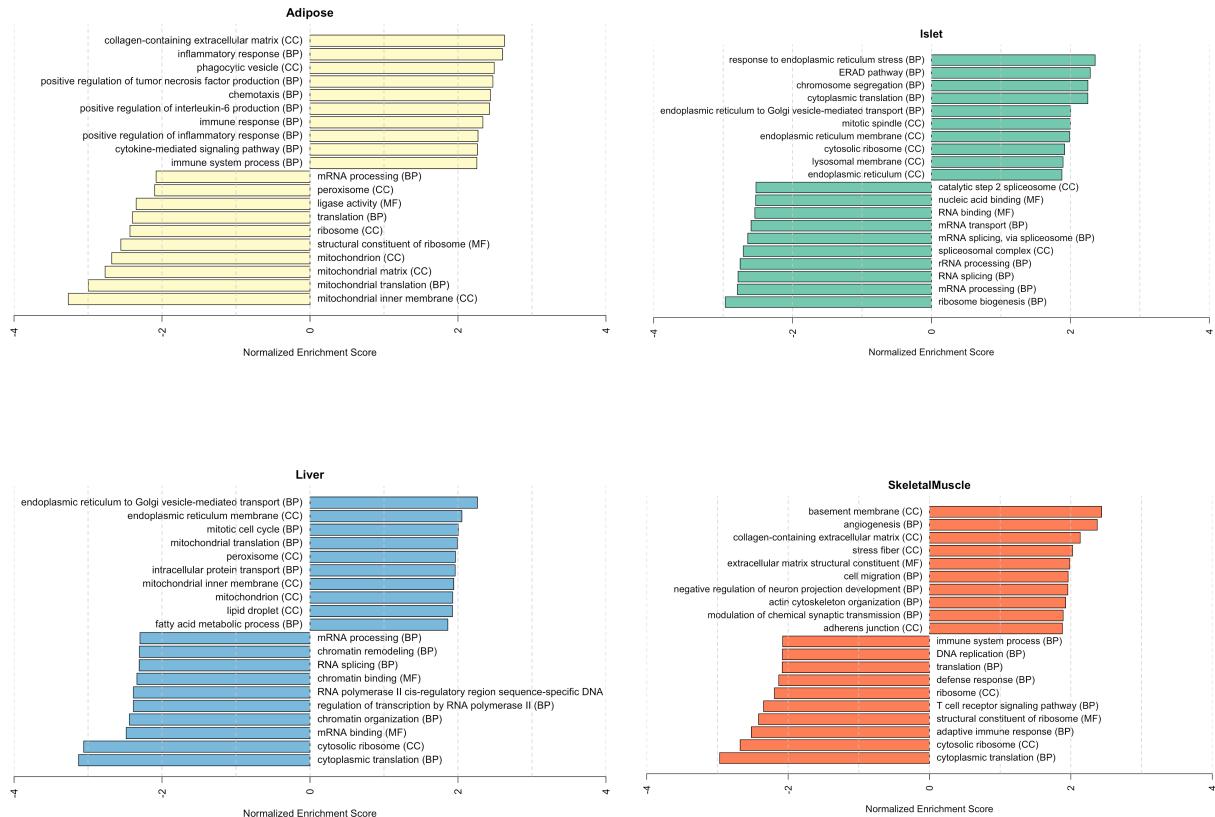


Figure S3: Bar plots showing normalized enrichment scores (NES) for GO terms as determined by fast gene score enrichment analysis (fgsea). Only the top 10 positive and top 10 negative scores are shown. Colors indicate tissue. The name beside each bar shows the name of each enriched GO term. The letters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

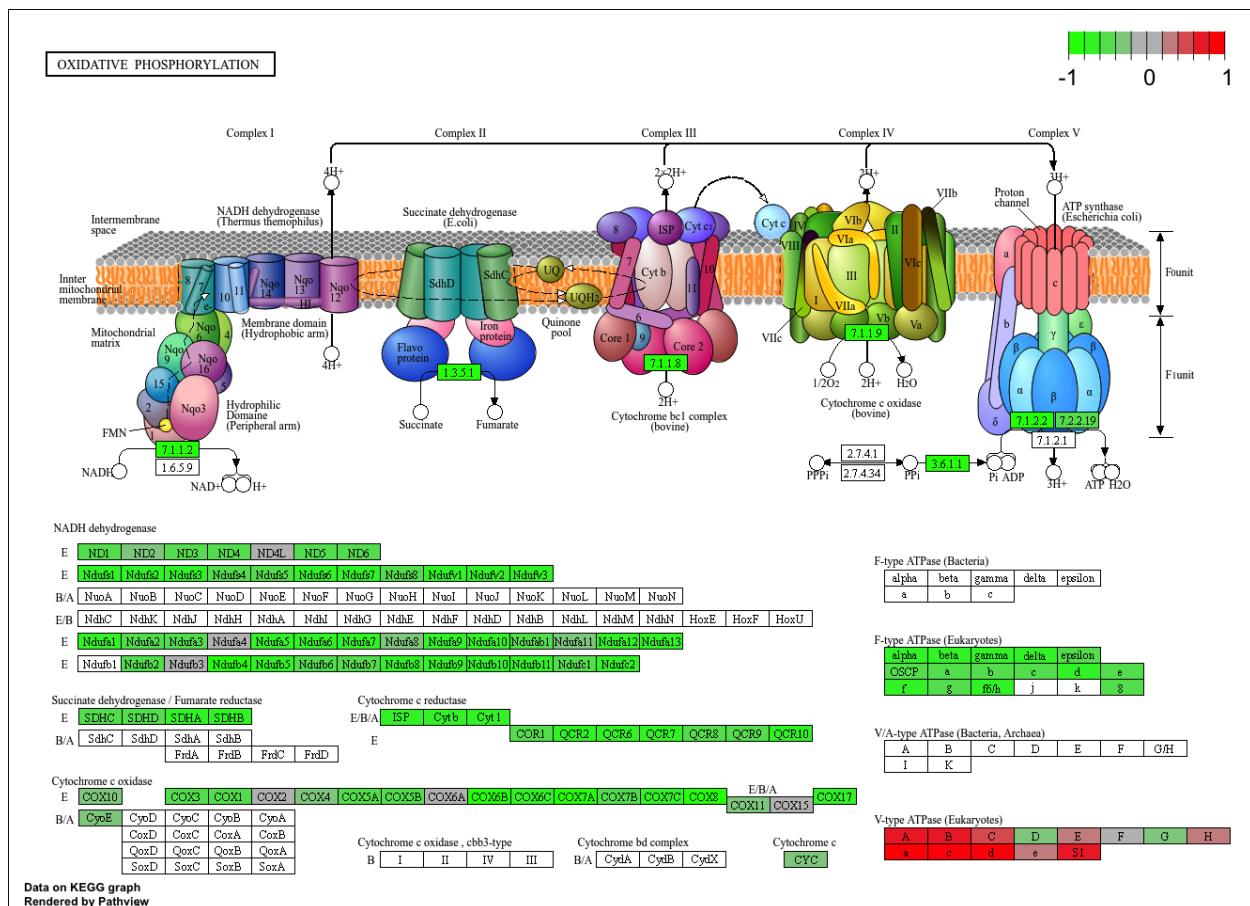


Figure S4: The KEGG pathway for oxidative phosphorylation in mice. Each element is colored based on its HDMA loading from adipose tissue normalized to run from -1 to 1. Genes highlighted in green had negative loadings, and those highlighted in red had positive loadings. Almost the entire pathway was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

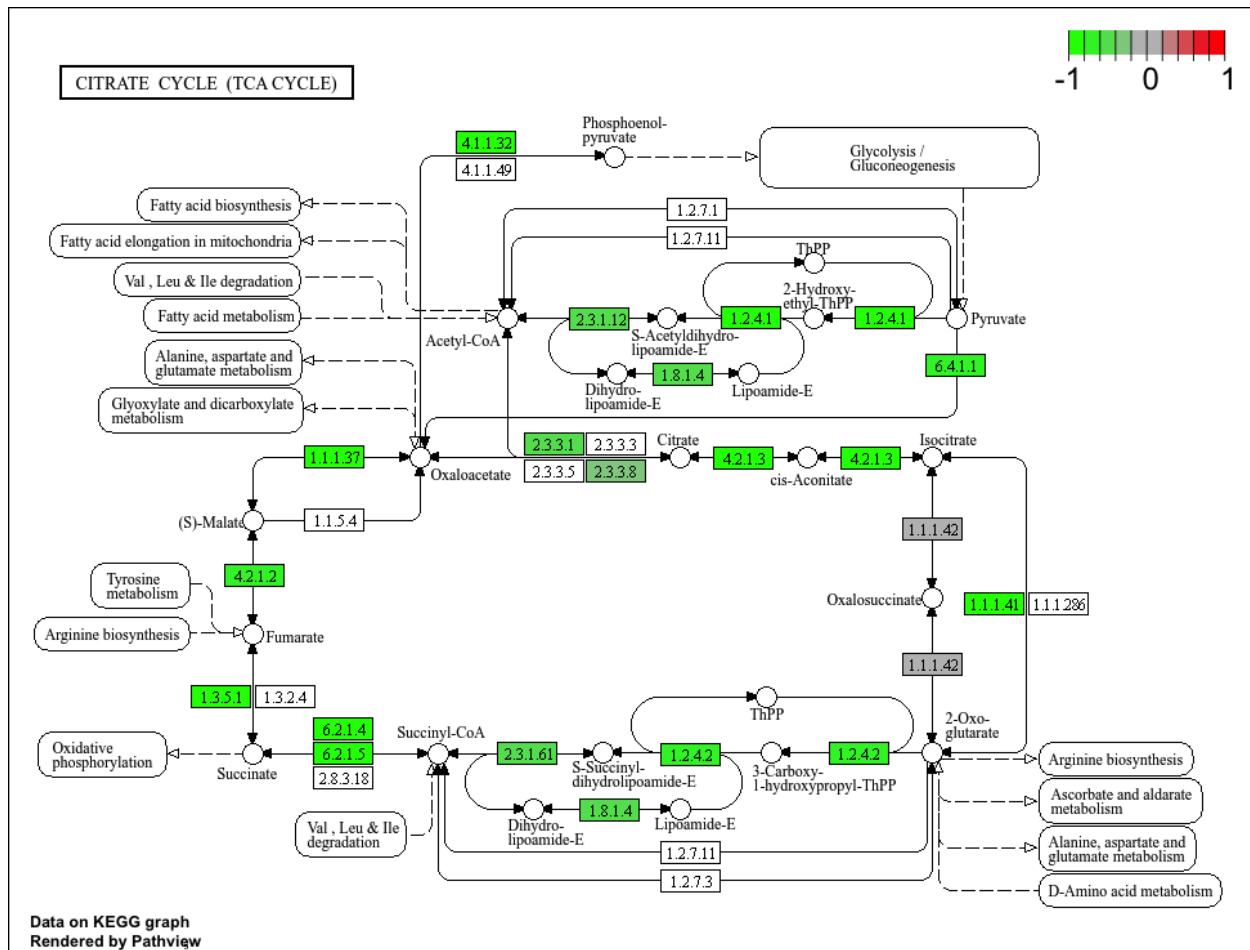


Figure S5: The KEGG pathway for the TCA (citric acid) cycle in mice. Each element is colored based on its HDMA loading from adipose tissue normalized to run from -1 to 1. Genes highlighted in green had negative loadings, and those highlighted in red had positive loadings. Many genes in the cycle were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

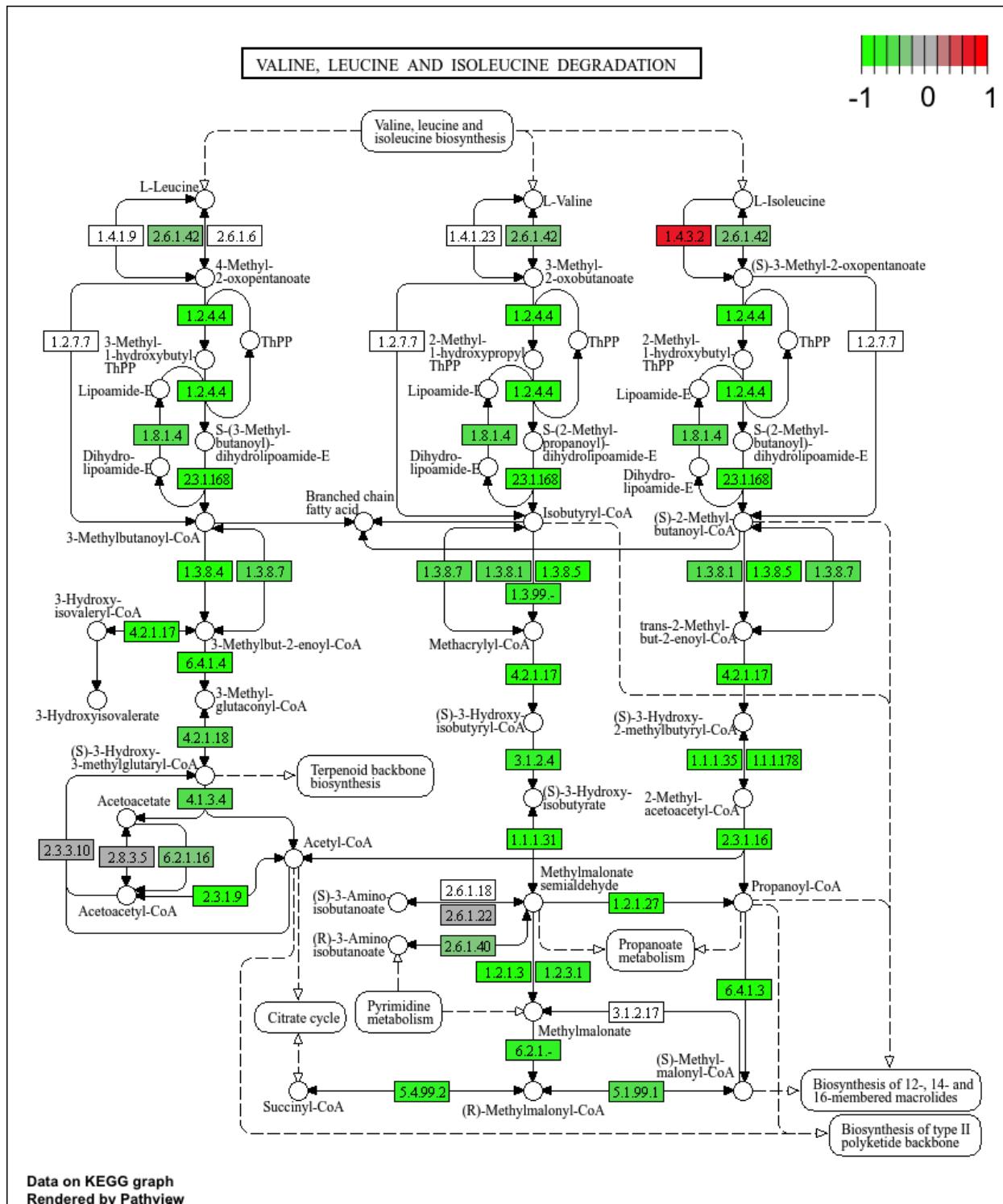


Figure S6: The KEGG pathway for branched-chain amino acid degradation in mice. Each element is colored based on its HDMA loading from adipose tissue normalized to run from -1 to 1. Genes highlighted in green had negative loadings, and those highlighted in red had positive loadings. Almost the entire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

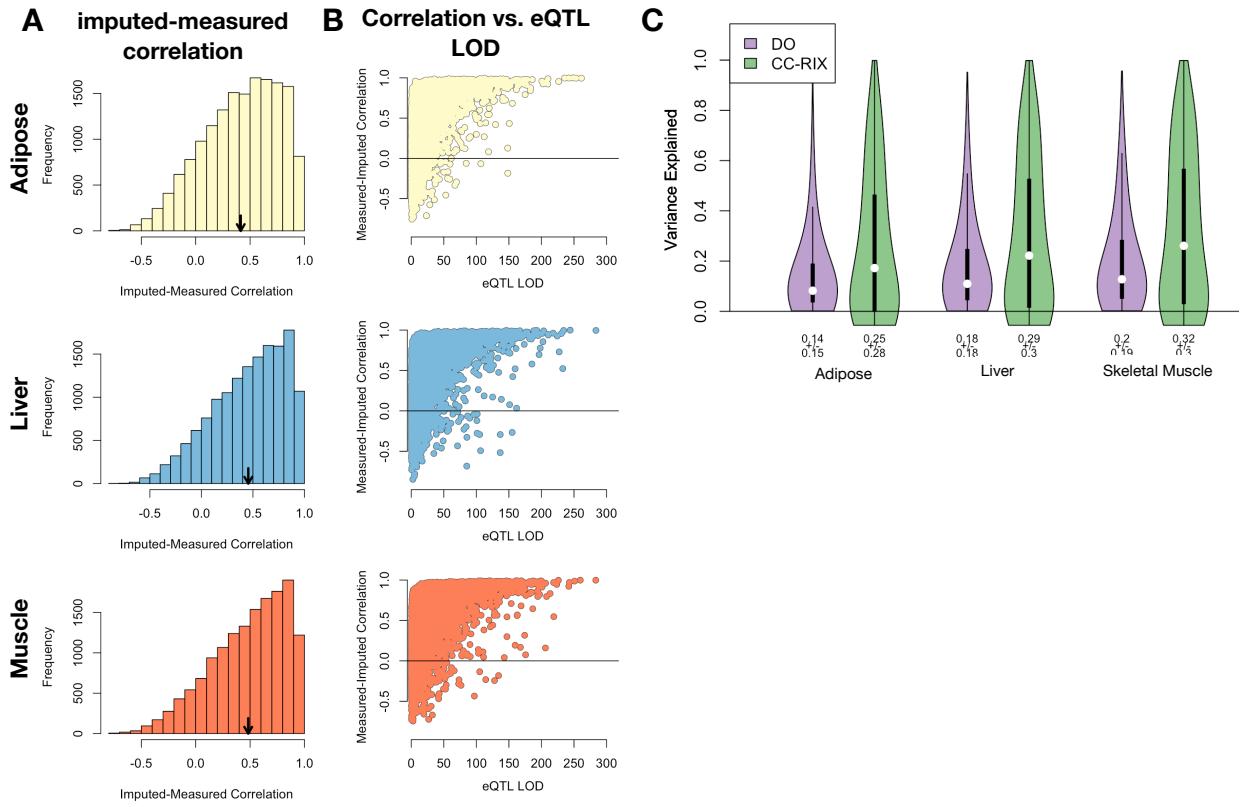


Figure S7: Validation of transcript imputation in the CC-RIX. **A.** Distributions of correlations between imputed and measured transcripts in the CC-RIX. The mean of each distribution is shown by the red line. All distributions were skewed toward positive correlations and had positive means near a Pearson 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: Clue.io results using the adipose tissue composite transcript as an input. All 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: Clue.io results using the pancreatic islet composite transcript as an input. All 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: Clue.io results using the adipose tissue composite transcript as an input. 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_RECECTOR_ANTAGONIST |
| | 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: Clue.io results using the pancreatic islet composite transcript as an input. Results are limited to the 30 most negatively correlated signals from YAPC cells, which were derived from a pancreatic carcinoma cells.

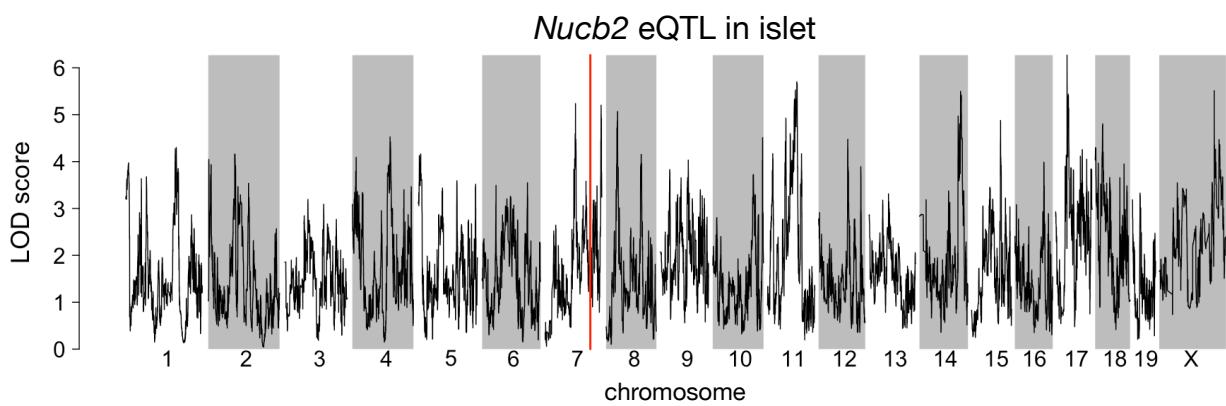


Figure S12: Regulation of *Nucb2* expression in islet. *Nucb2* is encoded on mouse chromosome 7 at 116.5 Mb (red line). In islets the heritability of *Nucb2* expression levels is 69% heritable. This LOD score trace shows that there is no local eQTL at that position, nor any strong distal eQTL anywhere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

446 **References**

- 447 [1] M. T. Maurano, R. Humbert, E. Rynes, R. E. Thurman, E. Haugen, H. Wang, A. P. Reynolds,
448 R. Sandstrom, H. Qu, J. Brody, A. Shafer, F. Neri, K. Lee, T. Kutyavin, S. Stehling-Sun, A. K.
449 Johnson, T. K. Canfield, E. Giste, M. Diegel, D. Bates, R. S. Hansen, S. Neph, P. J. Sabo, S. Heimfeld,
450 A. Raubitschek, S. Ziegler, C. Cotsapas, N. Sotoodehnia, I. Glass, S. R. Sunyaev, R. Kaul, and J. A.
451 Stamatoyannopoulos. Systematic localization of common disease-associated variation in regulatory DNA.
452 *Science*, 337(6099):1190–1195, Sep 2012.
- 453 [2] K. K. Farh, A. Marson, J. Zhu, M. Kleinewietfeld, W. J. Housley, S. Beik, N. Shores, H. Whitton, R. J.
454 Ryan, A. A. Shishkin, M. Hatan, M. J. Carrasco-Alfonso, D. Mayer, C. J. Luckey, N. A. Patsopoulos,
455 P. L. De Jager, V. K. Kuchroo, C. B. Epstein, M. J. Daly, D. A. Hafler, and B. E. Bernstein. Genetic
456 and epigenetic fine mapping of causal autoimmune disease variants. *Nature*, 518(7539):337–343, Feb
457 2015.
- 458 [3] E. Pennisi. The Biology of Genomes. Disease risk links to gene regulation. *Science*, 332(6033):1031, May
459 2011.
- 460 [4] L. A. Hindorff, P. Sethupathy, H. A. Junkins, E. M. Ramos, J. P. Mehta, F. S. Collins, and T. A. Manolio.
461 Potential etiologic and functional implications of genome-wide association loci for human diseases and
462 traits. *Proc Natl Acad Sci*, 106(23):9362–9367, Jun 2009.
- 463 [5] J. K. Pickrell. Joint analysis of functional genomic data and genome-wide association studies of 18
464 human traits. *Am J Hum Genet*, 94(4):559–573, Apr 2014.
- 465 [6] D. Welter, J. MacArthur, J. Morales, T. Burdett, P. Hall, H. Junkins, A. Klemm, P. Flieck, T. Manolio,
466 L. Hindorff, and H. Parkinson. The NHGRI GWAS Catalog, a curated resource of SNP-trait associations.
467 *Nucleic Acids Res*, 42(Database issue):D1001–1006, Jan 2014.
- 468 [7] Y. I. Li, B. van de Geijn, A. Raj, D. A. Knowles, A. A. Petti, D. Golan, Y. Gilad, and J. K. Pritchard.
469 RNA splicing is a primary link between genetic variation and disease. *Science*, 352(6285):600–604, Apr
470 2016.
- 471 [8] D. Zhou, Y. Jiang, X. Zhong, N. J. Cox, C. Liu, and E. R. Gamazon. A unified framework for joint-tissue
472 transcriptome-wide association and Mendelian randomization analysis. *Nat Genet*, 52(11):1239–1246,
473 Nov 2020.
- 474 [9] E. R. Gamazon, H. E. Wheeler, K. P. Shah, S. V. Mozaffari, K. Aquino-Michaels, R. J. Carroll, A. E.

- 475 Eyler, J. C. Denny, D. L. Nicolae, N. J. Cox, and H. K. Im. A gene-based association method for
476 mapping traits using reference transcriptome data. *Nat Genet*, 47(9):1091–1098, Sep 2015.
- 477 [10] Z. Zhu, F. Zhang, H. Hu, A. Bakshi, M. R. Robinson, J. E. Powell, G. W. Montgomery, M. E. Goddard,
478 N. R. Wray, P. M. Visscher, and J. Yang. Integration of summary data from GWAS and eQTL studies
479 predicts complex trait gene targets. *Nat Genet*, 48(5):481–487, May 2016.
- 480 [11] A. Gusev, A. Ko, H. Shi, G. Bhatia, W. Chung, B. W. Penninx, R. Jansen, E. J. de Geus, D. I. Boomsma,
481 F. A. Wright, P. F. Sullivan, E. Nikkola, M. Alvarez, M. Civelek, A. J. Lusis, T. ki, E. Raitoharju,
482 M. nen, I. ä, O. T. Raitakari, J. Kuusisto, M. Laakso, A. L. Price, P. Pajukanta, and B. Pasaniuc.
483 Integrative approaches for large-scale transcriptome-wide association studies. *Nat Genet*, 48(3):245–252,
484 Mar 2016.
- 485 [12] M. P. Keller, D. M. Gatti, K. L. Schueler, M. E. Rabaglia, D. S. Stapleton, P. Simecek, M. Vincent,
486 S. Allen, A. T. Broman, R. Bacher, C. Kendzierski, K. W. Broman, B. S. Yandell, G. A. Churchill, and
487 A. D. Attie. Genetic Drivers of Pancreatic Islet Function. *Genetics*, 209(1):335–356, May 2018.
- 488 [13] W. L. Crouse, G. R. Keele, M. S. Gastonguay, G. A. Churchill, and W. Valdar. A Bayesian model
489 selection approach to mediation analysis. *PLoS Genet*, 18(5):e1010184, May 2022.
- 490 [14] J. M. Chick, S. C. Munger, P. Simecek, E. L. Huttlin, K. Choi, D. M. Gatti, N. Raghupathy, K. L. Svenson,
491 G. A. Churchill, and S. P. Gygi. Defining the consequences of genetic variation on a proteome-wide scale.
492 *Nature*, 534(7608):500–505, Jun 2016.
- 493 [15] H. E. Wheeler, S. Ploch, A. N. Barbeira, R. Bonazzola, A. Andaleon, A. Fotuhi Siahpirani, A. Saha,
494 A. Battle, S. Roy, and H. K. Im. Imputed gene associations identify replicable trans-acting genes enriched
495 in transcription pathways and complex traits. *Genet Epidemiol*, 43(6):596–608, Sep 2019.
- 496 [16] B. D. Umans, A. Battle, and Y. Gilad. Where Are the Disease-Associated eQTLs? *Trends Genet*,
497 37(2):109–124, Feb 2021.
- 498 [17] N. J. Connally, S. Nazeen, D. Lee, H. Shi, J. Stamatoyannopoulos, S. Chun, C. Cotsapas, C. A. Cassa,
499 and S. R. Sunyaev. The missing link between genetic association and regulatory function. *Elife*, 11, Dec
500 2022.
- 501 [18] H. Mostafavi, J. P. Spence, S. Naqvi, and J. K. Pritchard. Systematic differences in discovery of genetic
502 effects on gene expression and complex traits. *Nat Genet*, 55(11):1866–1875, Nov 2023.
- 503 [19] D. W. Yao, L. J. O'Connor, A. L. Price, and A. Gusev. Quantifying genetic effects on disease mediated
504 by assayed gene expression levels. *Nat Genet*, 52(6):626–633, Jun 2020.

- 505 [20] X. Liu, J. A. Mefford, A. Dahl, Y. He, M. Subramaniam, A. Battle, A. L. Price, and N. Zaitlen. GBAT:
506 a gene-based association test for robust detection of trans-gene regulation. *Genome Biol*, 21(1):211, Aug
507 2020.
- 508 [21] E. Sollis, A. Mosaku, A. Abid, A. Buniello, M. Cerezo, L. Gil, T. Groza, O. §, P. Hall, J. Hayhurst,
509 A. Ibrahim, Y. Ji, S. John, E. Lewis, J. A. L. MacArthur, A. McMahon, D. Osumi-Sutherland,
510 K. Panoutsopoulou, Z. Pendlington, S. Ramachandran, R. Stefancsik, J. Stewart, P. Whetzel, R. Wilson,
511 L. Hindorff, F. Cunningham, S. A. Lambert, M. Inouye, H. Parkinson, and L. W. Harris. The NHGRI-EBI
512 GWAS Catalog: knowledgebase and deposition resource. *Nucleic Acids Res*, 51(D1):D977–D985, Jan
513 2023.
- 514 [22] R. J. F. Loos and G. S. H. Yeo. The genetics of obesity: from discovery to biology. *Nat Rev Genet*,
515 23(2):120–133, Feb 2022.
- 516 [23] R. K. Singh, P. Kumar, and K. Mahalingam. Molecular genetics of human obesity: A comprehensive
517 review. *C R Biol*, 340(2):87–108, Feb 2017.
- 518 [24] P. Arner. Obesity—a genetic disease of adipose tissue? *Br J Nutr*, 83 Suppl 1:9–16, Mar 2000.
- 519 [25] G. A. Churchill, D. M. Gatti, S. C. Munger, and K. L. Svenson. The Diversity Outbred mouse population.
520 *Mamm Genome*, 23(9-10):713–718, Oct 2012.
- 521 [26] Elissa J Chesler, Darla R Miller, Lisa R Branstetter, Leslie D Galloway, Barbara L Jackson, Vivek M
522 Philip, Brynn H Voy, Cymbeline T Culiat, David W Threadgill, Robert W Williams, et al. The
523 collaborative cross at oak ridge national laboratory: developing a powerful resource for systems genetics.
524 *Mammalian Genome*, 19:382–389, 2008.
- 525 [27] Michael C Saul, Vivek M Philip, Laura G Reinholdt, and Elissa J Chesler. High-diversity mouse
526 populations for complex traits. *Trends in Genetics*, 35(7):501–514, 2019.
- 527 [28] S. M. Clee and A. D. Attie. The genetic landscape of type 2 diabetes in mice. *Endocr Rev*, 28(1):48–83,
528 Feb 2007.
- 529 [29] K. W. Broman, D. M. Gatti, P. Simecek, N. A. Furlotte, P. Prins, Š. Sen, B. S. Yandell, and G. A.
530 Churchill. R/qtl2: Software for Mapping Quantitative Trait Loci with High-Dimensional Data and
531 Multiparent Populations. *Genetics*, 211(2):495–502, Feb 2019.
- 532 [30] Fabien Girka, Etienne Camenen, Caroline Peltier, Arnaud Gloaguen, Vincent Guillemot, Laurent Le
533 Brusquet, and Arthur Tenenhaus. *RGCCA: Regularized and Sparse Generalized Canonical Correlation
534 Analysis for Multiblock Data*, 2023. R package version 3.0.3.

- 535 [31] M. Helmer, S. Warrington, A. R. Mohammadi-Nejad, J. L. Ji, A. Howell, B. Rosand, A. Anticevic,
536 S. N. Sotiropoulos, and J. D. Murray. On the stability of canonical correlation analysis and partial least
537 squares with application to brain-behavior associations. *Commun Biol*, 7(1):217, Feb 2024.
- 538 [32] Gennady Korotkevich, Vladimir Sukhov, and Alexey Sergushichev. Fast gene set enrichment analysis.
539 *bioRxiv*, 2019.
- 540 [33] A. Subramanian, P. Tamayo, V. K. Mootha, S. Mukherjee, B. L. Ebert, M. A. Gillette, A. Paulovich,
541 S. L. Pomeroy, T. R. Golub, E. S. Lander, and J. P. Mesirov. Gene set enrichment analysis: a
542 knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci U S A*,
543 102(43):15545–15550, Oct 2005.
- 544 [34] S. Subramanian and A. Chait. The effect of dietary cholesterol on macrophage accumulation in adipose
545 tissue: implications for systemic inflammation and atherosclerosis. *Curr Opin Lipidol*, 20(1):39–44, Feb
546 2009.
- 547 [35] I. Akoumianakis, N. Akawi, and C. Antoniades. Exploring the Crosstalk between Adipose Tissue and
548 the Cardiovascular System. *Korean Circ J*, 47(5):670–685, Sep 2017.
- 549 [36] I. S. Stafeev, A. V. Vorotnikov, E. I. Ratner, M. Y. Menshikov, and Y. V. Parfyonova. Latent Inflammation
550 and Insulin Resistance in Adipose Tissue. *Int J Endocrinol*, 2017:5076732, 2017.
- 551 [37] I. P. Fischer, M. Irmler, C. W. Meyer, S. J. Sachs, F. Neff, M. de Angelis, J. Beckers, M. H. p, S. M.
552 Hofmann, and S. Ussar. A history of obesity leaves an inflammatory fingerprint in liver and adipose
553 tissue. *Int J Obes (Lond)*, 42(3):507–517, Mar 2018.
- 554 [38] S. Chung, H. Cuffe, S. M. Marshall, A. L. McDaniel, J. H. Ha, K. Kavanagh, C. Hong, P. Tontonoz,
555 R. E. Temel, and J. S. Parks. Dietary cholesterol promotes adipocyte hypertrophy and adipose tissue
556 inflammation in visceral, but not in subcutaneous, fat in monkeys. *Arterioscler Thromb Vasc Biol*,
557 34(9):1880–1887, Sep 2014.
- 558 [39] V. Kus, T. Prazak, P. Brauner, M. Hensler, O. Kuda, P. Flachs, P. Janovska, D. Medrikova, M. Rossmeisl,
559 Z. Jilkova, B. Stefl, E. Pastalkova, Z. Drahota, J. Houstek, and J. Kopecky. Induction of muscle
560 thermogenesis by high-fat diet in mice: association with obesity-resistance. *Am J Physiol Endocrinol
561 Metab*, 295(2):E356–367, Aug 2008.
- 562 [40] C. B. Newgard. Interplay between lipids and branched-chain amino acids in development of insulin
563 resistance. *Cell Metab*, 15(5):606–614, May 2012.
- 564 [41] D. D. Sears, G. Hsiao, A. Hsiao, J. G. Yu, C. H. Courtney, J. M. Ofrecio, J. Chapman, and S. Subramaniam.

- 565 Mechanisms of human insulin resistance and thiazolidinedione-mediated insulin sensitization. *Proc Natl*
566 *Acad Sci U S A*, 106(44):18745–18750, Nov 2009.
- 567 [42] R. Stienstra, C. Duval, M. ller, and S. Kersten. PPARs, Obesity, and Inflammation. *PPAR Res*,
568 2007:95974, 2007.
- 569 [43] O. Gavrilova, M. Haluzik, K. Matsusue, J. J. Cutson, L. Johnson, K. R. Dietz, C. J. Nicol, C. Vinson,
570 F. J. Gonzalez, and M. L. Reitman. Liver peroxisome proliferator-activated receptor gamma contributes
571 to hepatic steatosis, triglyceride clearance, and regulation of body fat mass. *J Biol Chem*, 278(36):34268–
572 34276, Sep 2003.
- 573 [44] K. Matsusue, M. Haluzik, G. Lambert, S. H. Yim, O. Gavrilova, J. M. Ward, B. Brewer, M. L. Reitman,
574 and F. J. Gonzalez. Liver-specific disruption of PPARgamma in leptin-deficient mice improves fatty
575 liver but aggravates diabetic phenotypes. *J Clin Invest*, 111(5):737–747, Mar 2003.
- 576 [45] D. Patsouris, J. K. Reddy, M. ller, and S. Kersten. Peroxisome proliferator-activated receptor alpha
577 mediates the effects of high-fat diet on hepatic gene expression. *Endocrinology*, 147(3):1508–1516, Mar
578 2006.
- 579 [46] S. E. Schadinger, N. L. Bucher, B. M. Schreiber, and S. R. Farmer. PPARgamma2 regulates lipogenesis
580 and lipid accumulation in steatotic hepatocytes. *Am J Physiol Endocrinol Metab*, 288(6):E1195–1205,
581 Jun 2005.
- 582 [47] W. Motomura, M. Inoue, T. Ohtake, N. Takahashi, M. Nagamine, S. Tanno, Y. Kohgo, and T. Okumura.
583 Up-regulation of ADRP in fatty liver in human and liver steatosis in mice fed with high fat diet. *Biochem*
584 *Biophys Res Commun*, 340(4):1111–1118, Feb 2006.
- 585 [48] A. Srivastava, A. P. Morgan, M. L. Najarian, V. K. Sarsani, J. S. Sigmon, J. R. Shorter, A. Kashfeen,
586 R. C. McMullan, L. H. Williams, P. guez, M. T. Ferris, P. Sullivan, P. Hock, D. R. Miller, T. A. Bell,
587 L. McMillan, G. A. Churchill, and F. P. de Villena. Genomes of the Mouse Collaborative Cross. *Genetics*,
588 206(2):537–556, Jun 2017.
- 589 [49] D. W. Threadgill, D. R. Miller, G. A. Churchill, and F. P. de Villena. The collaborative cross: a
590 recombinant inbred mouse population for the systems genetic era. *ILAR J*, 52(1):24–31, 2011.
- 591 [50] A. Roberts, F. Pardo-Manuel de Villena, W. Wang, L. McMillan, and D. W. Threadgill. The poly-
592 morphism architecture of mouse genetic resources elucidated using genome-wide resequencing data:
593 implications for QTL discovery and systems genetics. *Mamm Genome*, 18(6-7):473–481, Jul 2007.
- 594 [51] G. A. Churchill, D. C. Airey, H. Allayee, J. M. Angel, A. D. Attie, J. Beatty, W. D. Beavis, J. K.

- 595 Belknap, B. Bennett, W. Berrettini, A. Bleich, M. Bogue, K. W. Broman, K. J. Buck, E. Buckler,
596 M. Burmeister, E. J. Chesler, J. M. Cheverud, S. Clapcote, M. N. Cook, R. D. Cox, J. C. Crabbe,
597 W. E. Crusio, A. Darvasi, C. F. Deschepper, R. W. Doerge, C. R. Farber, J. Forejt, D. Gaile, S. J.
598 Garlow, H. Geiger, H. Gershenfeld, T. Gordon, J. Gu, W. Gu, G. de Haan, N. L. Hayes, C. Heller,
599 H. Himmelbauer, R. Hitzemann, K. Hunter, H. C. Hsu, F. A. Iraqi, B. Ivandic, H. J. Jacob, R. C. Jansen,
600 K. J. Jepsen, D. K. Johnson, T. E. Johnson, G. Kempermann, C. Kendzierski, M. Kotb, R. F. Kooy,
601 B. Llamas, F. Lammert, J. M. Lassalle, P. R. Lowenstein, L. Lu, A. Lusis, K. F. Manly, R. Marcucio,
602 D. Matthews, J. F. Medrano, D. R. Miller, G. Mittelman, B. A. Mock, J. S. Mogil, X. Montagutelli,
603 G. Morahan, D. G. Morris, R. Mott, J. H. Nadeau, H. Nagase, R. S. Nowakowski, B. F. O'Hara, A. V.
604 Osadchuk, G. P. Page, B. Paigen, K. Paigen, A. A. Palmer, H. J. Pan, L. Peltonen-Palotie, J. Peirce,
605 D. Pomp, M. Pravenec, D. R. Prows, Z. Qi, R. H. Reeves, J. Roder, G. D. Rosen, E. E. Schadt, L. C.
606 Schalkwyk, Z. Seltzer, K. Shimomura, S. Shou, M. J. ä, L. D. Siracusa, H. W. Snoeck, J. L. Spearow,
607 K. Svenson, L. M. Tarantino, D. Threadgill, L. A. Toth, W. Valdar, F. P. de Villena, C. Warden,
608 S. Whatley, R. W. Williams, T. Wiltshire, N. Yi, D. Zhang, M. Zhang, and F. Zou. The Collaborative
609 Cross, a community resource for the genetic analysis of complex traits. *Nat Genet*, 36(11):1133–1137,
610 Nov 2004.
- 611 [52] J. Y. Huh, Y. J. Park, M. Ham, and J. B. Kim. Crosstalk between adipocytes and immune cells in
612 adipose tissue inflammation and metabolic dysregulation in obesity. *Mol Cells*, 37(5):365–371, May 2014.
- 613 [53] J. Lamb, E. D. Crawford, D. Peck, J. W. Modell, I. C. Blat, M. J. Wrobel, J. Lerner, J. P. Brunet,
614 A. Subramanian, K. N. Ross, M. Reich, H. Hieronymus, G. Wei, S. A. Armstrong, S. J. Haggarty,
615 P. A. Clemons, R. Wei, S. A. Carr, E. S. Lander, and T. R. Golub. The Connectivity Map: using
616 gene-expression signatures to connect small molecules, genes, and disease. *Science*, 313(5795):1929–1935,
617 Sep 2006.
- 618 [54] S. Amin, A. Lux, and F. O'Callaghan. The journey of metformin from glycaemic control to mTOR
619 inhibition and the suppression of tumour growth. *Br J Clin Pharmacol*, 85(1):37–46, Jan 2019.
- 620 [55] A. Kezic, L. Popovic, and K. Lalic. mTOR Inhibitor Therapy and Metabolic Consequences: Where Do
621 We Stand? *Oxid Med Cell Longev*, 2018:2640342, 2018.
- 622 [56] A. D. Barlow, M. L. Nicholson, and T. P. Herbert. -cells and a review of the underlying molecular
623 mechanisms. *Diabetes*, 62(8):2674–2682, Aug 2013.
- 624 [57] Y. Gu, J. Lindner, A. Kumar, W. Yuan, and M. A. Magnuson. Rictor/mTORC2 is essential for
625 maintaining a balance between beta-cell proliferation and cell size. *Diabetes*, 60(3):827–837, Mar 2011.

- 626 [58] E. B. Geer, J. Islam, and C. Buettner. Mechanisms of glucocorticoid-induced insulin resistance: focus on
627 adipose tissue function and lipid metabolism. *Endocrinol Metab Clin North Am*, 43(1):75–102, Mar 2014.
- 628 [59] J. X. Li and C. L. Cummins. Fresh insights into glucocorticoid-induced diabetes mellitus and new
629 therapeutic directions. *Nat Rev Endocrinol*, 18(9):540–557, Sep 2022.
- 630 [60] R. A. Lee, C. A. Harris, and J. C. Wang. Glucocorticoid Receptor and Adipocyte Biology. *Nucl Receptor
631 Res*, 5, 2018.
- 632 [61] S. Viengchareun, P. Penfornis, M. C. Zennaro, and M. s. Mineralocorticoid and glucocorticoid receptors
633 inhibit UCP expression and function in brown adipocytes. *Am J Physiol Endocrinol Metab*, 280(4):E640–
634 649, Apr 2001.
- 635 [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,
636 P. Cheng, M. Hu, Y. Li, J. Bi, Y. Li, F. Hu, Y. Zhong, Y. Xu, and G. Ding. -HSD1 in regulating brown
637 adipocyte function. *J Mol Endocrinol*, 50(1):103–113, Feb 2013.
- 638 [63] L. E. Ramage, M. Akyol, A. M. Fletcher, J. Forsythe, M. Nixon, R. N. Carter, E. J. van Beek,
639 N. M. Morton, B. R. Walker, and R. H. Stimson. Glucocorticoids Acutely Increase Brown Adipose
640 Tissue Activity in Humans, Revealing Species-Specific Differences in UCP-1 Regulation. *Cell Metab*,
641 24(1):130–141, Jul 2016.
- 642 [64] J. L. Barclay, H. Agada, C. Jang, M. Ward, N. Wetzig, and K. K. Ho. Effects of glucocorticoids on
643 human brown adipocytes. *J Endocrinol*, 224(2):139–147, Feb 2015.
- 644 [65] M. ó, R. Gupte, W. L. Kraus, P. Pacher, and P. Bai. PARPs in lipid metabolism and related diseases.
645 *Prog Lipid Res*, 84:101117, Nov 2021.
- 646 [66] P. Bai, C. ó, H. Oudart, A. nszki, Y. Cen, C. Thomas, H. Yamamoto, A. Huber, B. Kiss, R. H.
647 Houtkooper, K. Schoonjans, V. Schreiber, A. A. Sauve, J. Menissier-de Murcia, and J. Auwerx. PARP-1
648 inhibition increases mitochondrial metabolism through SIRT1 activation. *Cell Metab*, 13(4):461–468,
649 Apr 2011.
- 650 [67] A. Chiarugi and M. A. Moskowitz. Cell biology. PARP-1—a perpetrator of apoptotic cell death? *Science*,
651 297(5579):200–201, Jul 2002.
- 652 [68] M. Althubiti, R. Almaimani, S. Y. Eid, M. Elzubaier, B. Refaat, S. Idris, T. A. Alqurashi, and M. Z.
653 El-Readi. BTK targeting suppresses inflammatory genes and ameliorates insulin resistance. *Eur Cytokine
654 Netw*, 31(4):168–179, Dec 2020.

- 655 [69] C. Skrabs, W. F. Pickl, T. Perkmann, U. ger, and A. Gessl. Rapid decline in insulin antibodies and
656 glutamic acid decarboxylase autoantibodies with ibrutinib therapy of chronic lymphocytic leukaemia. *J*
657 *Clin Pharm Ther*, 43(1):145–149, Feb 2018.
- 658 [70] E. A. Oral, S. M. Reilly, A. V. Gomez, R. Meral, L. Butz, N. Ajluni, T. L. Chenevert, E. Korytnaya,
659 A. H. Neidert, R. Hench, D. Rus, J. F. Horowitz, B. Poirier, P. Zhao, K. Lehmann, M. Jain, R. Yu,
660 C. Liddle, M. Ahmadian, M. Downes, R. M. Evans, and A. R. Saltiel. and TBK1 Improves Glucose
661 Control in a Subset of Patients with Type 2 Diabetes. *Cell Metab*, 26(1):157–170, Jul 2017.
- 662 [71] M. C. Arkan, A. L. Hevener, F. R. Greten, S. Maeda, Z. W. Li, J. M. Long, A. Wynshaw-Boris, G. Poli,
663 J. Olefsky, and M. Karin. IKK-beta links inflammation to obesity-induced insulin resistance. *Nat Med*,
664 11(2):191–198, Feb 2005.
- 665 [72] M. Clark, C. J. Kroger, Q. Ke, and R. M. Tisch. The Role of T Cell Receptor Signaling in the
666 Development of Type 1 Diabetes. *Front Immunol*, 11:615371, 2020.
- 667 [73] P. ndell, A. C. Carlsson, A. Larsson, O. Melander, T. Wessman, J. v, and T. Ruge. TNFR1 is associated
668 with short-term mortality in patients with diabetes and acute dyspnea seeking care at the emergency
669 department. *Acta Diabetol*, 57(10):1145–1150, Oct 2020.
- 670 [74] A. M. Keestra-Gounder, M. X. Byndloss, N. Seyffert, B. M. Young, A. vez Arroyo, A. Y. Tsai, S. A.
671 Cevallos, M. G. Winter, O. H. Pham, C. R. Tiffany, M. F. de Jong, T. Kerrinnes, R. Ravindran, P. A.
672 Luciw, S. J. McSorley, A. J. umler, and R. M. Tsolis. NOD1 and NOD2 signalling links ER stress with
673 inflammation. *Nature*, 532(7599):394–397, Apr 2016.
- 674 [75] A. M. Keestra-Gounder and R. M. Tsolis. NOD1 and NOD2: Beyond Peptidoglycan Sensing. *Trends*
675 *Immunol*, 38(10):758–767, Oct 2017.
- 676 [76] J. Montane, L. Cadavez, and A. Novials. Stress and the inflammatory process: a major cause of
677 pancreatic cell death in type 2 diabetes. *Diabetes Metab Syndr Obes*, 7:25–34, 2014.
- 678 [77] B. B. Kahn and T. E. McGraw. , and type 2 diabetes. *N Engl J Med*, 363(27):2667–2669, Dec 2010.
- 679 [78] S. Del Prato and N. Pulizzi. The place of sulfonylureas in the therapy for type 2 diabetes mellitus.
680 *Metabolism*, 55(5 Suppl 1):S20–27, May 2006.
- 681 [79] E. V. Minikel, J. L. Painter, C. C. Dong, and M. R. Nelson. Refining the impact of genetic evidence on
682 clinical success. *Nature*, 629(8012):624–629, May 2024.
- 683 [80] N. Katsanis. The continuum of causality in human genetic disorders. *Genome Biol*, 17(1):233, Nov 2016.

- 684 [81] X. Liu, Y. I. Li, and J. K. Pritchard. Trans Effects on Gene Expression Can Drive Omnipotent Inheritance.
685 *Cell*, 177(4):1022–1034, May 2019.
- 686 [82] B. Hallgrímsson, R. M. Green, D. C. Katz, J. L. Fish, F. P. Bernier, C. C. Roseman, N. M. Young,
687 J. M. Cheverud, and R. S. Marcucio. The developmental-genetics of canalization. *Semin Cell Dev Biol*,
688 88:67–79, Apr 2019.
- 689 [83] M. L. Siegal and A. Bergman. Waddington’s canalization revisited: developmental stability and evolution.
690 *Proc Natl Acad Sci U S A*, 99(16):10528–10532, Aug 2002.
- 691 [84] A. B. Paaby and G. Gibson. Cryptic Genetic Variation in Evolutionary Developmental Genetics. *Biology*
692 (*Basel*), 5(2), Jun 2016.
- 693 [85] E. A. Boyle, Y. I. Li, and J. K. Pritchard. An Expanded View of Complex Traits: From Polygenic to
694 Omnipotent. *Cell*, 169(7):1177–1186, Jun 2017.
- 695 [86] Naomi R Wray, Cisca Wijmenga, Patrick F Sullivan, Jian Yang, and Peter M Visscher. Common disease
696 is more complex than implied by the core gene omnipotent model. *Cell*, 173(7):1573–1580, 2018.
- 697 [87] M. Riva, M. D. Nitert, U. Voss, R. Sathanoori, A. Lindqvist, C. Ling, and N. Wierup. Nesfatin-1
698 stimulates glucagon and insulin secretion and beta cell NUCB2 is reduced in human type 2 diabetic
699 subjects. *Cell Tissue Res*, 346(3):393–405, Dec 2011.
- 700 [88] M. Nakata and T. Yada. Role of NUCB2/nestin-1 in glucose control: diverse functions in islets,
701 adipocytes and brain. *Curr Pharm Des*, 19(39):6960–6965, 2013.
- 702 [89] H. Shimizu and A. Osaki. Nesfatin/Nucleobindin-2 (NUCB2) and Glucose Homeostasis. *Curr Hypertens*
703 *Rev*, pages Nesfatin/Nucleobindin-2 (NUCB2) and Glucose Homeostasis., Jul 2014.
- 704 [90] Judea Pearl. *Causality*. Cambridge University Press, 2009.