

<sup>1</sup> Transcripts with high distal heritability mediate genetic effects on  
<sup>2</sup> complex traits

<sup>3</sup>

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

<sup>7</sup> **Abstract**

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

<sup>19</sup> **Introduction**

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

26 with multiple disease traits<sup>12–15</sup>

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

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 detected<sup>16</sup>. 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<sup>18–20;15</sup>.

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 obesity  
40 and metabolic disease as an archetypal example. Diet-induced obesity and metabolic disease are genetically  
41 complex with hundreds of variants mapped through GWAS [REFS]. These variants are known to act through  
42 multiple tissues that interact dynamically with each other [REFS], 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) mice<sup>21</sup>, and an independent  
53 validation data set derived by crossing inbred strains from the Collaborative Cross (CC) mice<sup>22</sup> to form CC  
54 F1 mice (CC-RIX). Both populations modeled diet-induced obesity and metabolic disease<sup>12</sup>

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 mice<sup>21</sup>. 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 mice<sup>23</sup>. 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 inquiry<sup>21</sup>. 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 mice [REF], including body  
66 weight, plasma levels of insulin and glucose and plasma lipids, with transcriptome-wide gene expression in  
67 four tissues related to metabolic disease: adipose tissue, pancreatic islets, liver, and skeletal muscle. We  
68 measured similar metabolic traits in the CC-RIX and gene expression from three of the four tissues used  
69 in the DO: adipose tissue, liver, and skeletal muscle. Because the CC-RIX carry the same founder alleles  
70 as the DO, local gene regulation is expected to match between the populations, but because the alleles are  
71 recombined through the genome, distal effects are expected to vary from those in the DO, allowing us to  
72 directly assess the role of local gene regulation in driving trait-associated transcript variation. Together, these  
73 data enable a 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)<sup>24</sup>. 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.

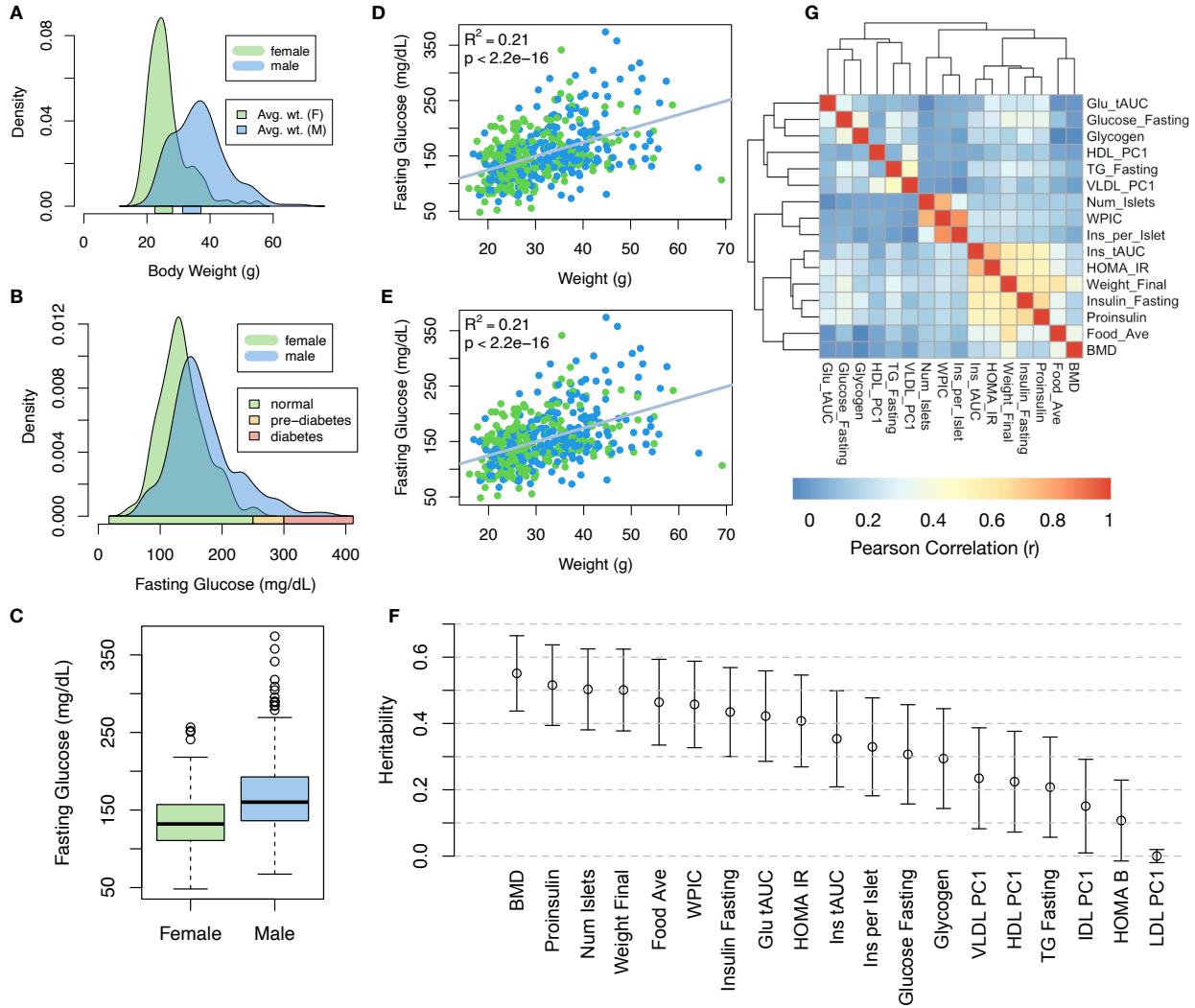


Figure 1: Clinical overview. **A.** Distributions of final body weight in the diversity outbred mice. Sex is indicated by color. The average B6 male and female adult weights at 24 weeks of age are indicated by blue and green bars on the x-axis. **B.** The distribution of final fasting glucose across 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/qltl<sup>25</sup>  
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 transcripts<sup>19</sup>. 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

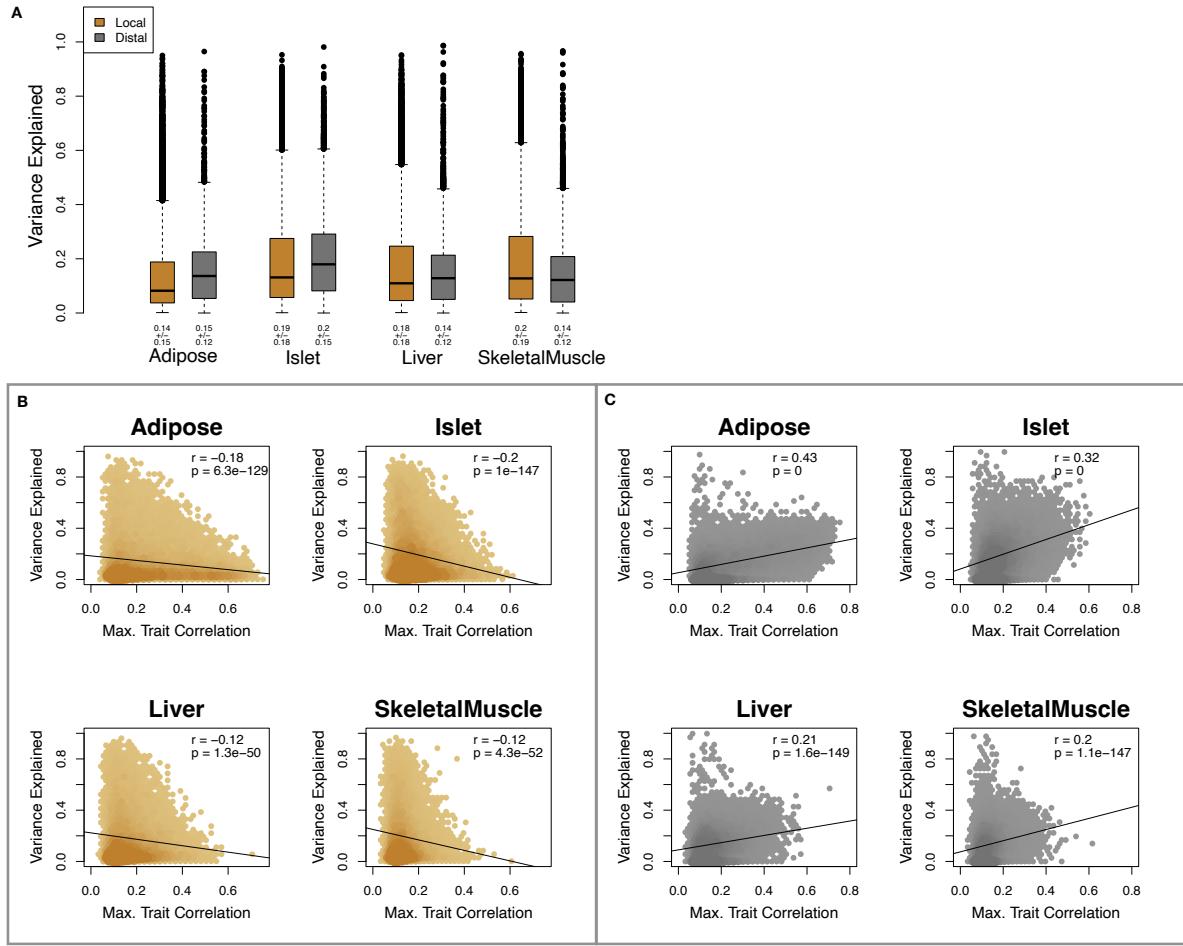


Figure 2: Transcript heritability and trait relevance. **A.** Distributions of distal and local heritability of transcripts across the four tissues. Overall local and distal factors contribute equally to transcript heritability. The relationship between **(B.)** local and **(C.)** distal heritability 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 developed<sup>26</sup>. 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 ( $\rho$ ) 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  $0.71 \pm 0.084$ , 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 XXX, HDMA is related to a generalized form of CCA. Standard CCA is prone to over-fitting because in any two large matrices it can be trivial to identify highly correlated composite vectors [REF]. To assess whether our implementation of HDMA was similarly prone to over-fitting in a high-dimensional space, we performed permutation testing. We permuted the individual labels on the transcriptome matrix 1000 times and recalculated the path coefficient, which is the partial correlation of  $G_C$  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

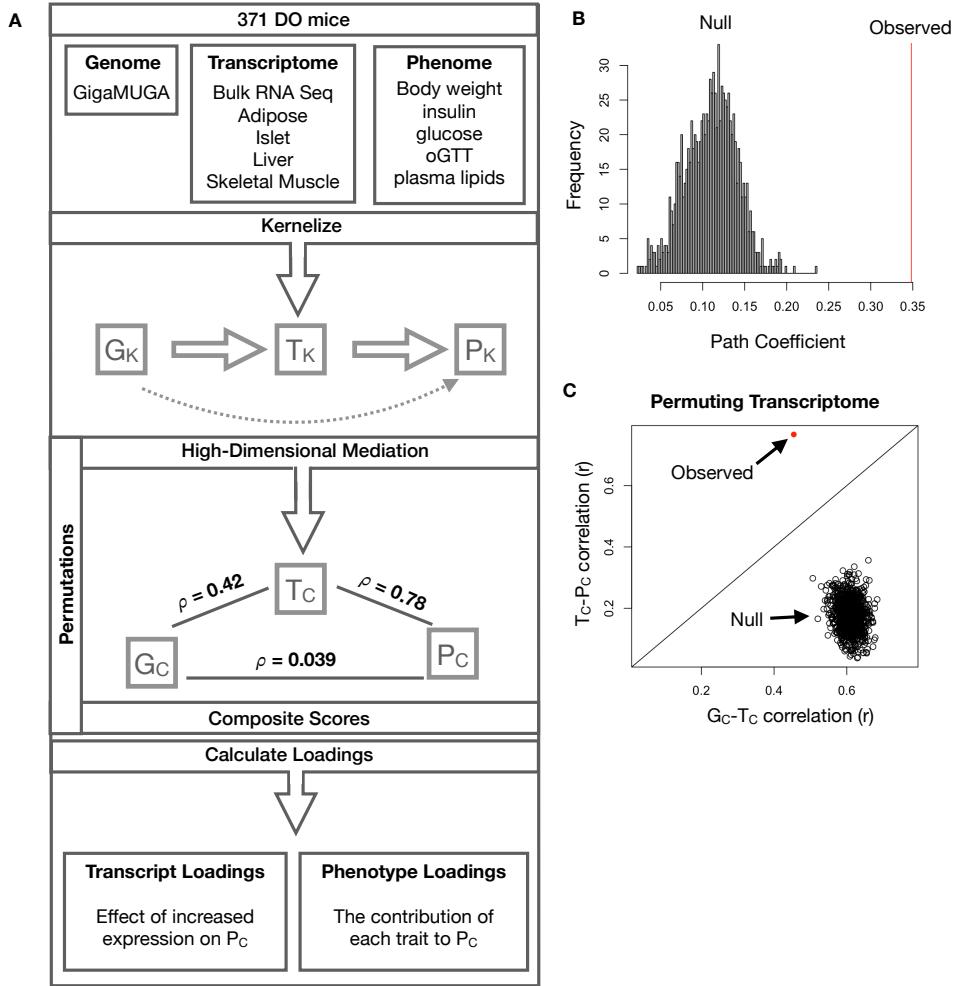


Figure 3: High-dimensional mediation. **A.** Workflow indicating major steps of high-dimensional mediation. The genotype, transcriptome, and phenotype matrices were independently normalized and converted to kernel matrices representing the pairwise relationships between individuals for each data modality ( $K_G$  = 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  $\rho$  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 maximized simultaneously in the null data. In contrast, the red dot shows that in the real data both the  
 161  $G_C-T_C$  correlation and the  $T_C-P_C$  correlation could be maximized simultaneously suggesting that the path  
 162 from genotype to phenotype through transcriptome is highly non-trivial and identifiable in this case. These  
 163 results suggest that these composite vectors represent genetically determined variation in phenotype that is  
 164 mediated through genetically determined variation in 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)<sup>27;28</sup> 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 [cite] and infiltrated by macrophages [cite], and the results here suggest  
192 that 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

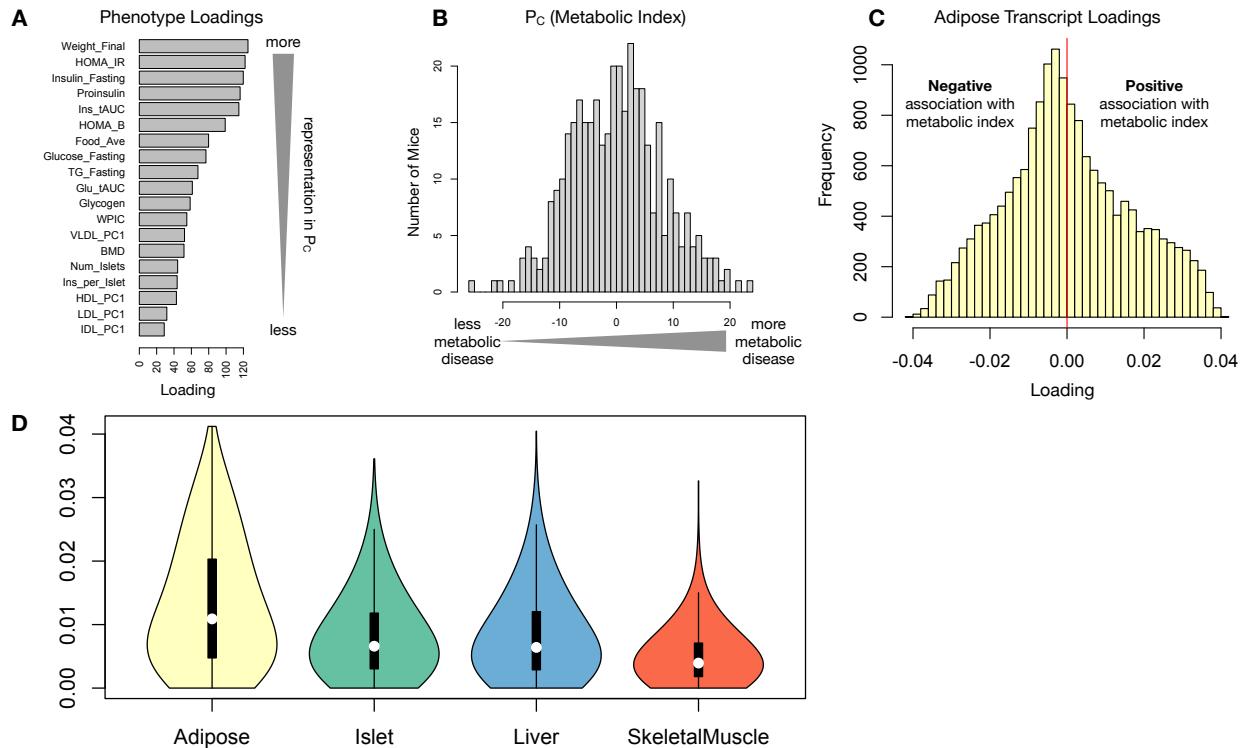


Figure 4: Interpretation of loadings. **A.** Loadings across traits. Body weight and insulin resistance contributed the most to the composite trait. **B.** Phenotype scores across individuals. Individuals with large positive phenotype scores had higher body weight and insulin resistance than average. 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 diet<sup>29</sup>.

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<sup>12;30;31</sup>, 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 2.375 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 adipogenesis<sup>32</sup>. Agonists of *Pparg*, such as thiazolidinediones, are  
229 FDA-approved to treat type II diabetes, and reduce inflammation and adipose hypertrophy<sup>32</sup>. 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<sup>33;34</sup>.  
234 Overexpression of *Pparg* in the livers of mice with a *Ppara* knockout, causes upregulation of genes involved in  
235 adipogenesis<sup>35</sup>. 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<sup>36;37</sup>.

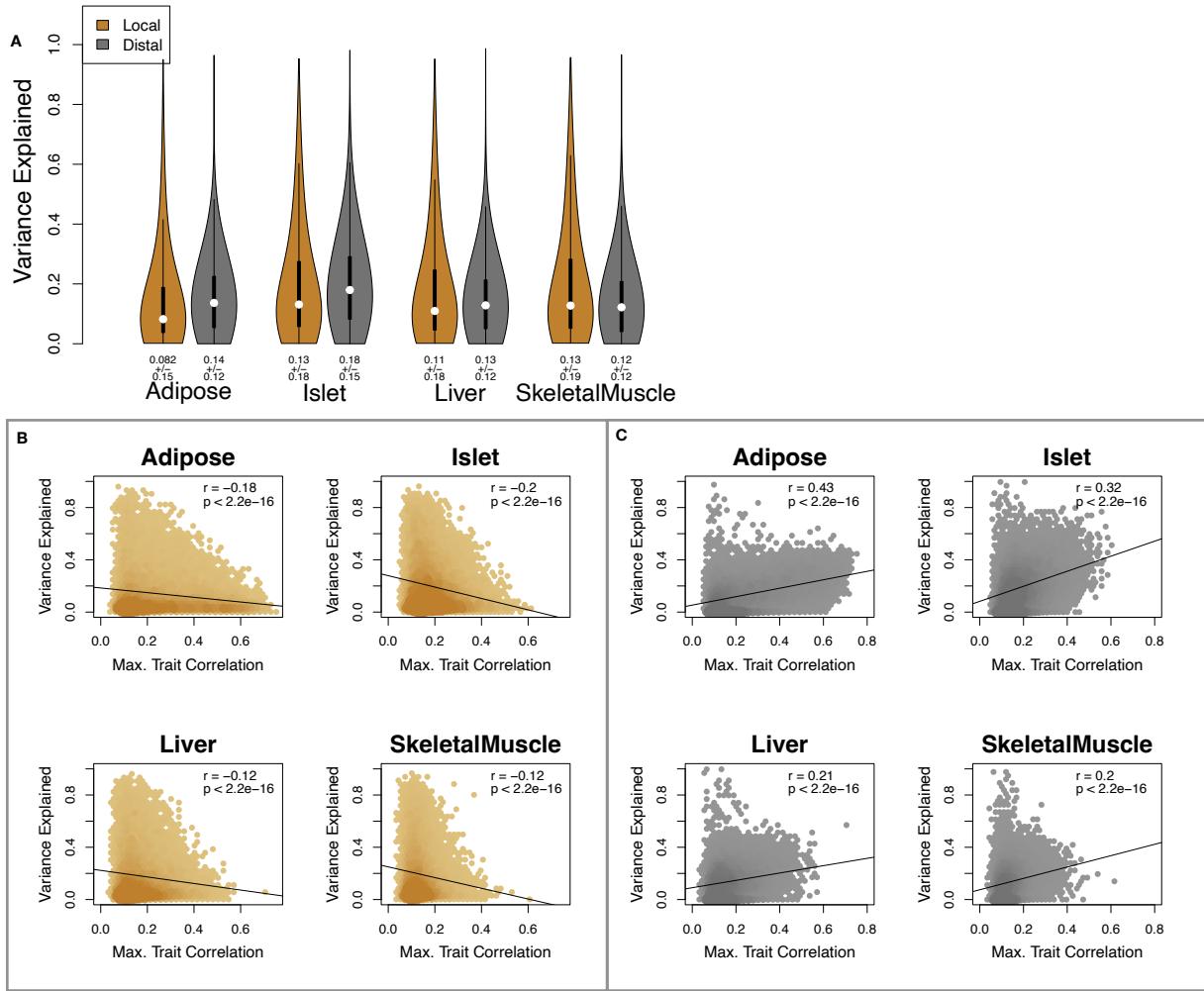


Figure 5: Transcripts with high loadings have high distal heritability and literature support. Each panel has a bar plot showing the loadings of transcripts selected by different criteria. Bar color indicates the tissue of origin. The heat map shows the local (L - left) and distal (D - right) heritability 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  
 240 phenotype. Both local and distal heritability of *Pparg* in the islet are relatively high, but the loading is low,  
 241 suggesting that variability of expression in the islet does not drive variation in metabolic index. These results  
 242 highlight the importance of tissue context when investigating the role of heritable transcript variability in  
 243 driving phenotype.

<sup>244</sup> Gene lists for all clusters are available in Supplemental File XXX.

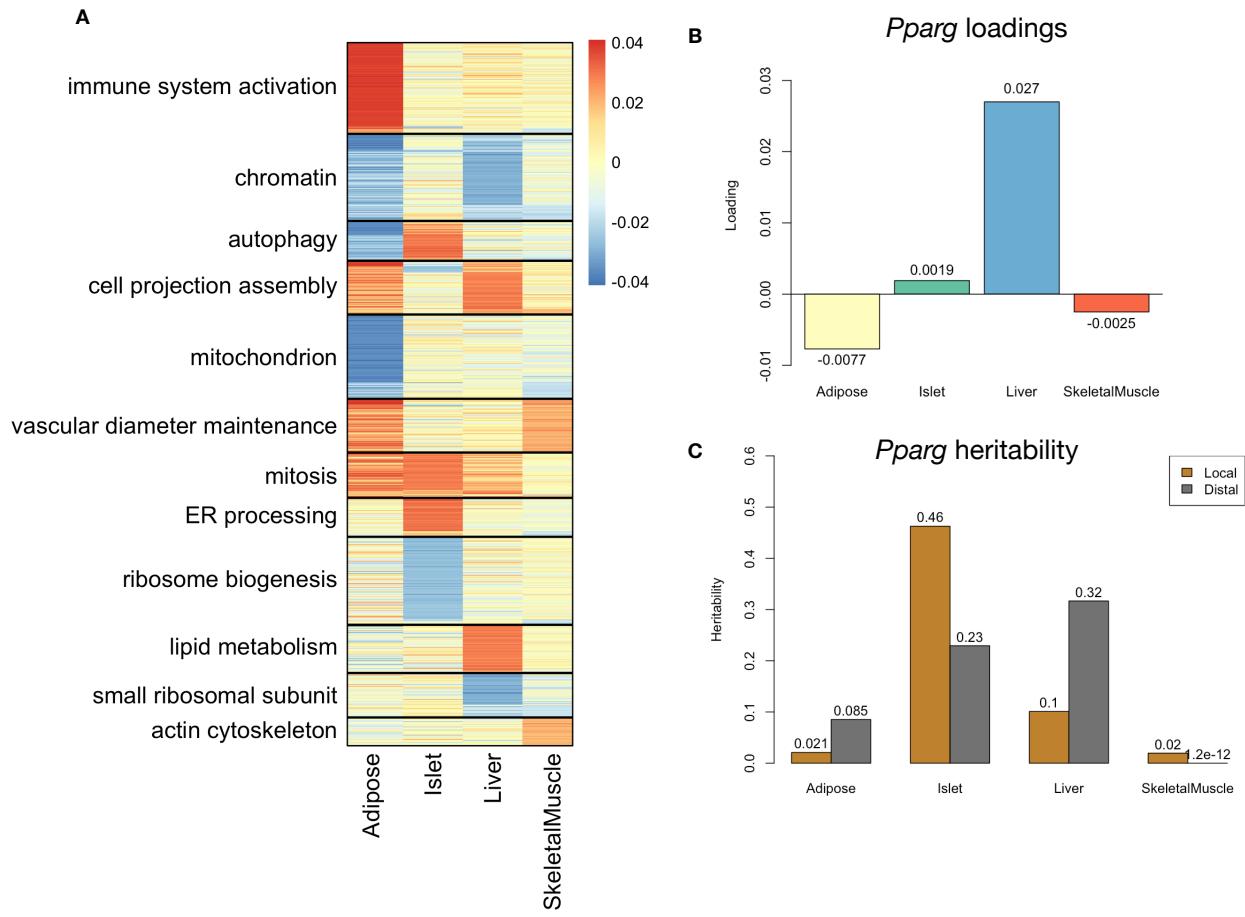


Figure 6: Tissue-specific transcriptional programs were associated with obesity and insulin resistance. **A** Heat map showing the loadings of all transcripts with loadings greater than 2.5 standard deviations from the mean in any tissue. The heat map was clustered using k medoid clustering. Functional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

#### <sup>245</sup> Gene expression, but not local eQTLs, predicted body weight in an independent population

<sup>246</sup> To test whether the transcript loadings identified in the DO could be translated to another population, we  
<sup>247</sup> tested whether they could predict metabolic phenotype in an independent population of CC-RIX mice, which  
<sup>248</sup> were F1 mice derived from multiple pairings of Collaborative Cross (CC) [cite] strains (Fig. 7) (Methods).  
<sup>249</sup> We tested two questions. First, we asked whether the loadings identified in the DO mice were relevant to  
<sup>250</sup> the relationship between the transcriptome and the phenotype in the CC-RIX. We predicted body weight (a  
<sup>251</sup> surrogate for metabolic index) in each CC-RIX individual using measured gene expression in each tissue and  
<sup>252</sup> the transcript loadings identified in the DO (Methods). The predicted body weight and actual body weight  
<sup>253</sup> 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.

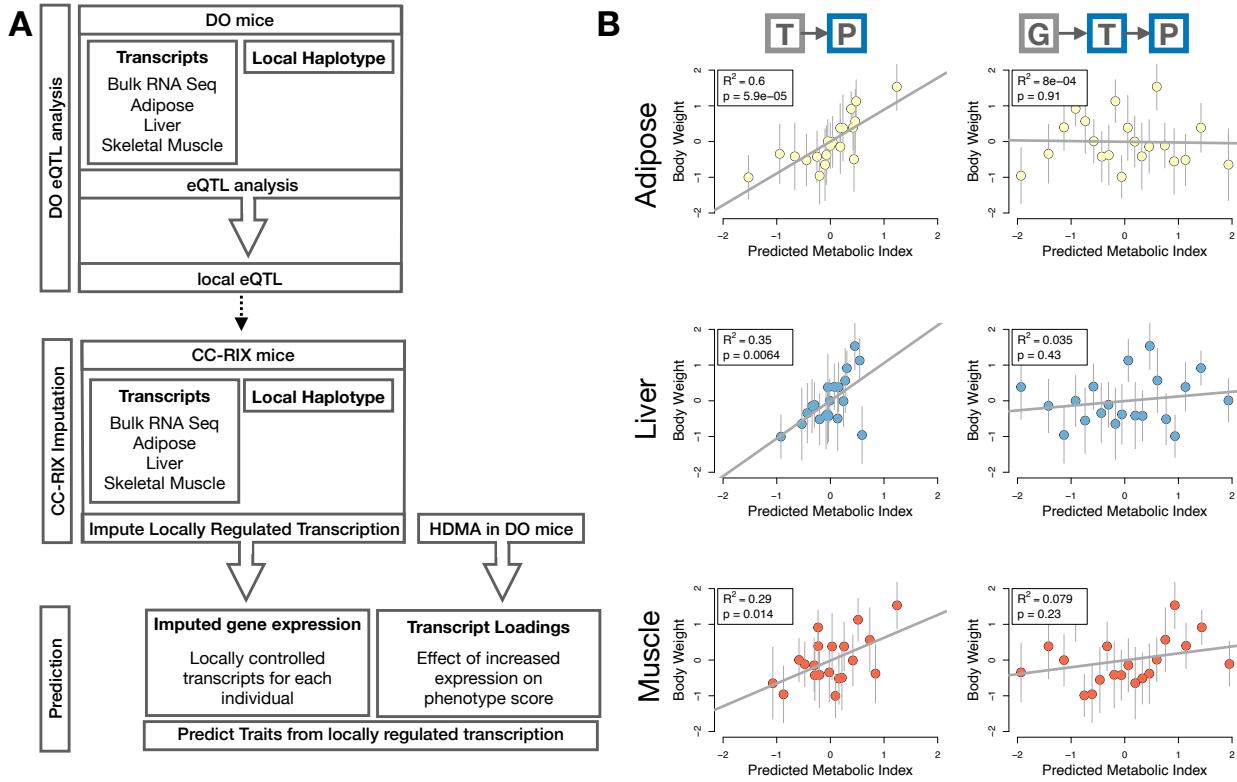


Figure 7: Transcription, but not local genotype, predicts phenotype in the CC-RIX. **A.** Workflow showing procedure for translating HDMA results to an independent population of mice. **B.** Relationships between the predicted metabolic index and measured body weight. The left column shows the predictions 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. Fig. S7). However, these imputed values failed to predict body weight in the CC-RIX when weighted with the loadings from HDMA. (Fig. 7B right column). This result suggests that local regulation of gene expression is not the primary factor driving heritability of complex traits, consistent with our findings in the DO population that distal heritability was a major driver of trait-relevant variation and that high-loading transcripts had

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 that  
275 macrophage infiltration in adipose tissue is a marker of obesity and metabolic disease [REF]. 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  
293 resistance, the major contributors to metabolic index. This is further consistent with the islet composition  
294 traits having small loadings in the phenome score (Fig. 4).

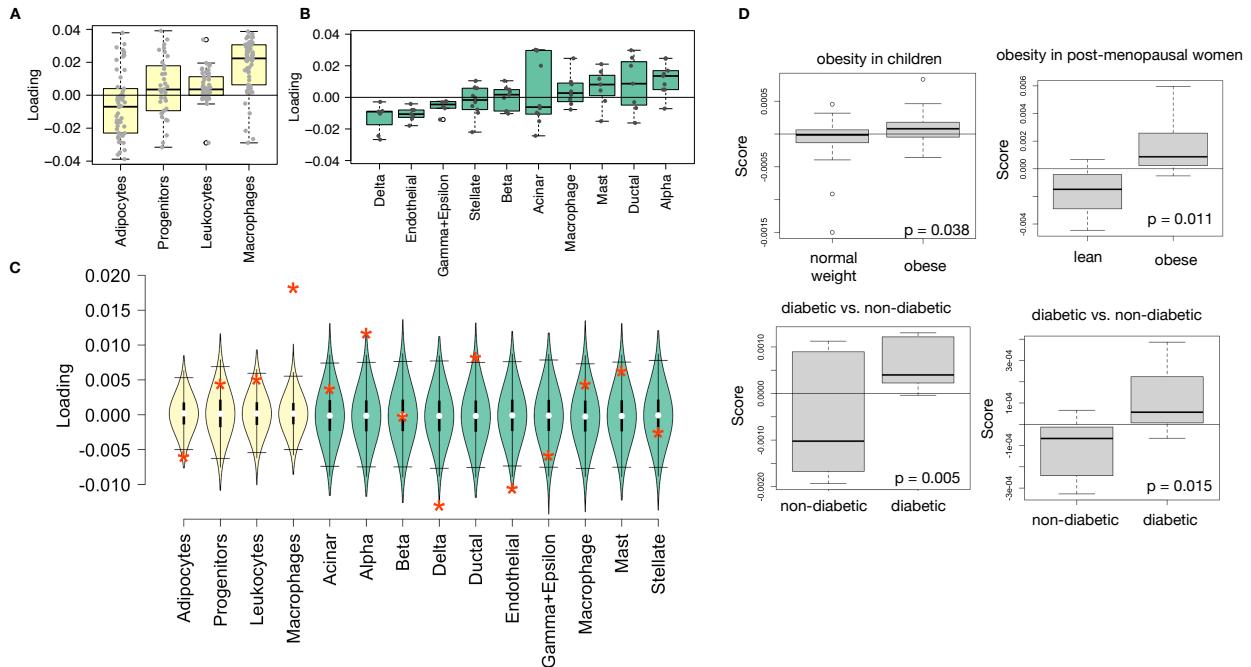


Figure 8: HDMA results translate to humans. **A.** Distribution of loadings for cell-type-specific transcripts in adipose tissue. **B.** Distribution of loadings for cell-type-specific transcripts in pancreatic islets (green). **C.** Null distributions for the mean loading of randomly selected transcripts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

#### 295 Heritable transcriptomic signatures translated to human disease

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

301 We calculated a predicted obesity score for each individual in the human studies based on their adipose  
 302 tissue gene expression (Methods) and compared the predicted scores for obese and non-obese groups as well  
 303 as diabetic and non-diabetic groups. In all cases, the predicted obesity scores were higher on average for  
 304 individuals in the obese and diabetic groups compared with the lean and non-diabetic groups (Fig. 8D). This  
 305 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, and focusing on individual transcripts for treatment may be less effective than using broader  
312 transcriptomic signatures that capture the emergent biology [cite or remove]. The ConnectivityMap (CMAP)  
313 database<sup>38</sup> developed by the Broad Institute allows querying thousands of compounds that reverse or enhance  
314 the extreme ends of transcriptomic signatures in multiple different cell types. By identifying drugs that  
315 reverse pathogenic transcriptomic signatures, we can potentially identify compounds that have favorable  
316 effects on gene expression.

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

322 Looking across all cell types, the notable top hits from the adipose tissue loadings included mTOR inhibitors  
323 and glucocorticoid agonists (Supp. Fig. S8). It is thought that metformin, which is commonly used to  
324 improve glycemic control, acts, at least in part, by inhibiting mTOR signaling<sup>39;40</sup>. However, long-term use  
325 of other mTOR inhibitors, such as rapamycin, are known to cause insulin resistance and  $\beta$ -cell toxicity<sup>40–42</sup>.  
326 Glucocorticoids are used to reduce inflammation, which was a prominent signature in the adipose tissues,  
327 but these drugs also promote hyperglycemia and diabetes<sup>43;44</sup>. Accute treatment with glucocorticoids has  
328 further been shown to reduce thermogenesis in rodent adipocytes<sup>45–47</sup>, but increase thermogenesis in human  
329 adipocytes<sup>48;49</sup>. Thus, the pathways identified by CMAP across all cell types were highly related to the  
330 transcript loading profiles, but the relationship was not a simple reversal.

331 The top hit for the adipose composite transcript in CMAP adipocytes was a PARP inhibitor (Supp. Fig.  
332 S10). PARPs play a role in lipid metabolism and are involved in the development of obesity and diabetes<sup>50</sup>.  
333 PARP1 inhibition increases mitochondrial biogenesis<sup>51</sup>. Inhibition of PARP1 activity can further prevent  
334 necrosis in favor of the less inflammatory apoptosis<sup>52</sup>, thereby potentially reducing inflammation in stressed  
335 adipocytes. Other notable hits among the top 20 were BTK inhibitors, which have been observed to suppress  
336 inflammation and improve insulin resistance<sup>53</sup> as well as to reduce insulin antibodies in type I diabetes<sup>54</sup>.  
337 IKK inhibitors have been shown to improve glucose control in type II diabetes<sup>55;56</sup>.

338 Among the top most significant hits for the transcript loadings from pancreatic islets (Supp. Fig. S9), was  
339 suppression of T cell receptor signaling, which is known to be involved in Type 1 diabetes<sup>57</sup>, as well as  
340 TNFR1, which has been associated with mortality in diabetes patients<sup>58</sup>. Suppression of NOD1/2 signaling  
341 was also among the top hits. NOD1 and 2 sense ER stress<sup>59;60</sup>, which is associated with  $\beta$ -cell death in type  
342 1 and type 2 diabetes<sup>61</sup>. This cell death process is dependent on NOD1/2 signaling<sup>59</sup>, although the specifics  
343 have not yet been worked out.

344 We also looked specifically at hits in pancreatic tumor cells (YAPC) regardless of significance level to get a  
345 transcriptional response more specific to the pancreas (Supp. Fig. S11). Hits in this list included widely used  
346 diabetes drugs, such as sulfonylureas, PPAR receptor agonists, and insulin sensitizers. Rosiglitazone is a  
347 PPAR- $\gamma$  agonist and was one of the most prescribed drugs for type 2 diabetes before its use was reduced due  
348 to cardiac side-effects<sup>62</sup>. Sulfonylureas are another commonly prescribed drug class for type 2 diabetes, but  
349 also have notable side effects including hypoglycemia and accelerated  $\beta$ -cell death<sup>63</sup>.

## 350 Discussion

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

364 Genetics is indispensable for the dissection of disease mechanisms because it is one of the only data modalities  
365 that supports causal inferences about molecules and disease outcomes [REF]. It has frequently been assumed  
366 that gene regulation in *cis* is the primary driver of genetically associated trait variation, but attempts to use  
367 local gene regulation to explain phenotypic variation have had limited success<sup>16;17</sup>. In recent years, evidence  
368 has mounted that distal gene regulation may be an important mediator of trait heritability<sup>19;18;64</sup>. It has

369 been observed that transcripts with high local heritability explain less expression-mediated disease heritability  
370 than those with low local heritability<sup>19</sup>. Consistent with this observation, genes located near GWAS hits  
371 tend to be complexly regulated<sup>18</sup>. They also tend to be enriched with functional annotations, in contrast  
372 to genes with simple local regulation, which tend to be depleted of functional annotations suggesting they  
373 are less likely to be directly involved in disease traits<sup>18</sup>. These observations are consistent with principles  
374 of robustness in complex systems in which simple regulation of important elements leads to fragility of the  
375 system<sup>65–67</sup>. Our results are consistent, instead, with a more complex picture where genes whose expression  
376 can drive trait variation are buffered from local genetic variation but are extensively influenced indirectly by  
377 genetic variation in the regulatory networks converging on those genes.

378 Recently, the omnigenic model of complex traits has been proposed, which posits that complex traits are  
379 massively polygenic and that their heritability is spread out across the genome<sup>68</sup>. In the omnigenic model,  
380 genes are classified either as “core genes,” which directly impinge on the trait, or “peripheral genes,” which  
381 are not directly trait-related, but influence core genes through the complex gene regulatory network. HDMA  
382 explicitly models a central proposal of the omnigenic model which posits that once the expression of the  
383 core genes (i.e. trait-mediating genes) is accounted for, there should be no residual correlation between the  
384 genome and the phenotype. Here, when the composite transcript was taken into account there was no residual  
385 correlation between the composite genome and composite phenotype (Fig. 3A).

386 Thus, the transcript loadings can be interpreted as indicating higher “core-ness” of a transcript. Unlike in the  
387 omnigenic model, we did not observe a clear demarcation between the core and peripheral genes in loading  
388 magnitude, but we do not necessarily expect a clear separation given the complexity of gene regulation and  
389 the genotype-phenotype map<sup>69</sup>.

390 An extension of the omnigenic model proposed that most heritability of complex traits is driven by weak  
391 distal eQTLs that are potentially below the detection threshold in studies with feasible sample sizes<sup>64</sup>. This  
392 is consistent with what we observed here. The transcripts with the largest loadings were strongly distally  
393 regulated and only weakly locally regulated, suggesting that distal gene regulation plays a primary role in  
394 driving heritable trait variation. We saw further that the patterns of distal heritability were not localized to  
395 detectable distal eQTL, but rather were complex and spread across the genome, even for transcripts whose  
396 expression was strongly regulated by distal factors. For example, *Nucb2*, had a high loading in islets and  
397 was also strongly distally regulated (66% distal heritability) (Fig. 5). This gene is expressed in pancreatic  $\beta$   
398 cells and is involved in insulin and glucagon release<sup>70–72</sup>. Although its transcription was highly heritable in  
399 islets, that regulation was distributed across the genome, with no clear distal eQTL (Supp. Fig. S12). Thus,  
400 although distal regulation of some genes may be strong, this regulation is likely to be highly complex and not

401 easily localized.

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

416 Individual high-loading transcripts also demonstrated biologically interpretable, tissue-specific patterns. We  
417 highlighted *Pparg*, which is known to be protective in adipose tissue<sup>32</sup> where it was negatively loaded, and  
418 harmful in the liver<sup>33–37</sup>, where it was positively loaded. Such granular patterns may be useful in generating  
419 hypotheses for further testing, and prioritizing genes as therapeutic targets. The tissue-specific nature of  
420 the loadings also may provide clues to tissue-specific effects, or side effects, of targeting particular genes  
421 system-wide.

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

428 Another useful application of the composite transcripts is to pair them with cell-type specific genes to generate  
429 causal hypotheses about changes in cell composition in individual tissues. Combining the multi-tissue,  
430 transcriptome-wide weighted vectors with public databases and data sets thus provides a path for generating a  
431 wide range of testable hypotheses. Moreover, each data set presented here was derived from human tissues or

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

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

443 **Data Availability**

444 Here we tell people where to find the data

445 **Acknowledgements**

446 Here we thank people

447 **Supplemental Figures**

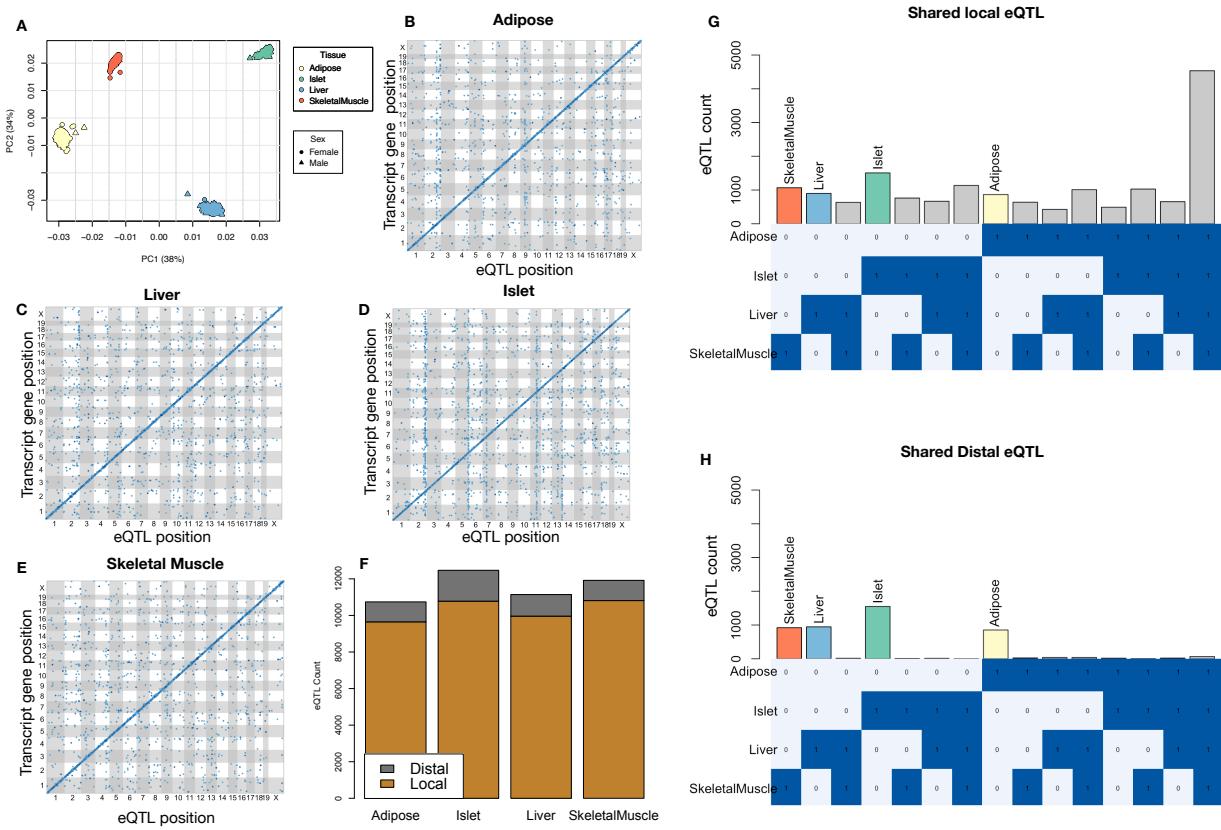


Figure S1: Overview of eQTL analysis in DO mice. **A.** RNA seq samples from the four different tissues clustered by tissue. **B.-E.** eQTL maps are shown for each tissue. The *x*-axis shows the position of the mapped eQTL, and the *y*-axis shows the physical position of the gene encoding each mapped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

448 **References**

- [1] M. T. Maurano, R. Humbert, E. Rynes, R. E. Thurman, E. Haugen, H. Wang, A. P. Reynolds, R. Sandstrom, H. Qu, J. Brody, A. Shafer, F. Neri, K. Lee, T. Kutyavin, S. Stehling-Sun, A. K. Johnson, T. K. Canfield, E. Giste, M. Diegel, D. Bates, R. S. Hansen, S. Neph, P. J. Sabo, S. Heimfeld, A. Raubitschek, S. Ziegler, C. Cotsapas, N. Sotoodehnia, I. Glass, S. R. Sunyaev, R. Kaul, and J. A.

## KEGG pathway enrichments by GSEA

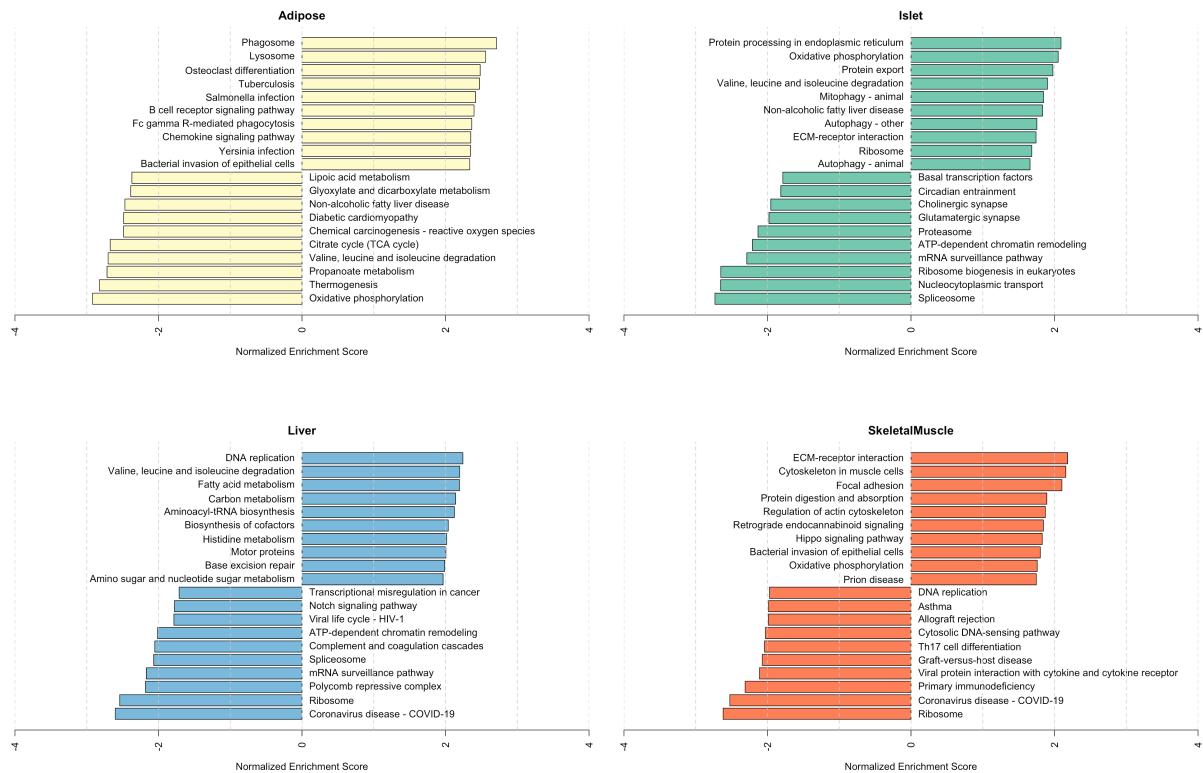


Figure S2: Bar plots showing normalized enrichment scores (NES) for KEGG pathways as determined by fast gene score enrichment analysis (fgsea). Only the top 10 positive and top 10 negative scores are shown. Colors indicate tissue. The name beside each bar shows the name of each enriched KEGG pathway.

- 453 Stamatoyanopoulos. Systematic localization of common disease-associated variation in regulatory DNA.  
 454 *Science*, 337(6099):1190–1195, Sep 2012.
- 455 [2] K. K. Farh, A. Marson, J. Zhu, M. Kleinewietfeld, W. J. Housley, S. Beik, N. Shores, H. Whitton, R. J.  
 456 Ryan, A. A. Shishkin, M. Hatan, M. J. Carrasco-Alfonso, D. Mayer, C. J. Luckey, N. A. Patsopoulos,  
 457 P. L. De Jager, V. K. Kuchroo, C. B. Epstein, M. J. Daly, D. A. Hafler, and B. E. Bernstein. Genetic  
 458 and epigenetic fine mapping of causal autoimmune disease variants. *Nature*, 518(7539):337–343, Feb  
 459 2015.
- 460 [3] E. Pennisi. The Biology of Genomes. Disease risk links to gene regulation. *Science*, 332(6033):1031, May  
 461 2011.
- 462 [4] L. A. Hindorff, P. Sethupathy, H. A. Junkins, E. M. Ramos, J. P. Mehta, F. S. Collins, and T. A. Manolio.

## Top GO term enrichments by GSEA

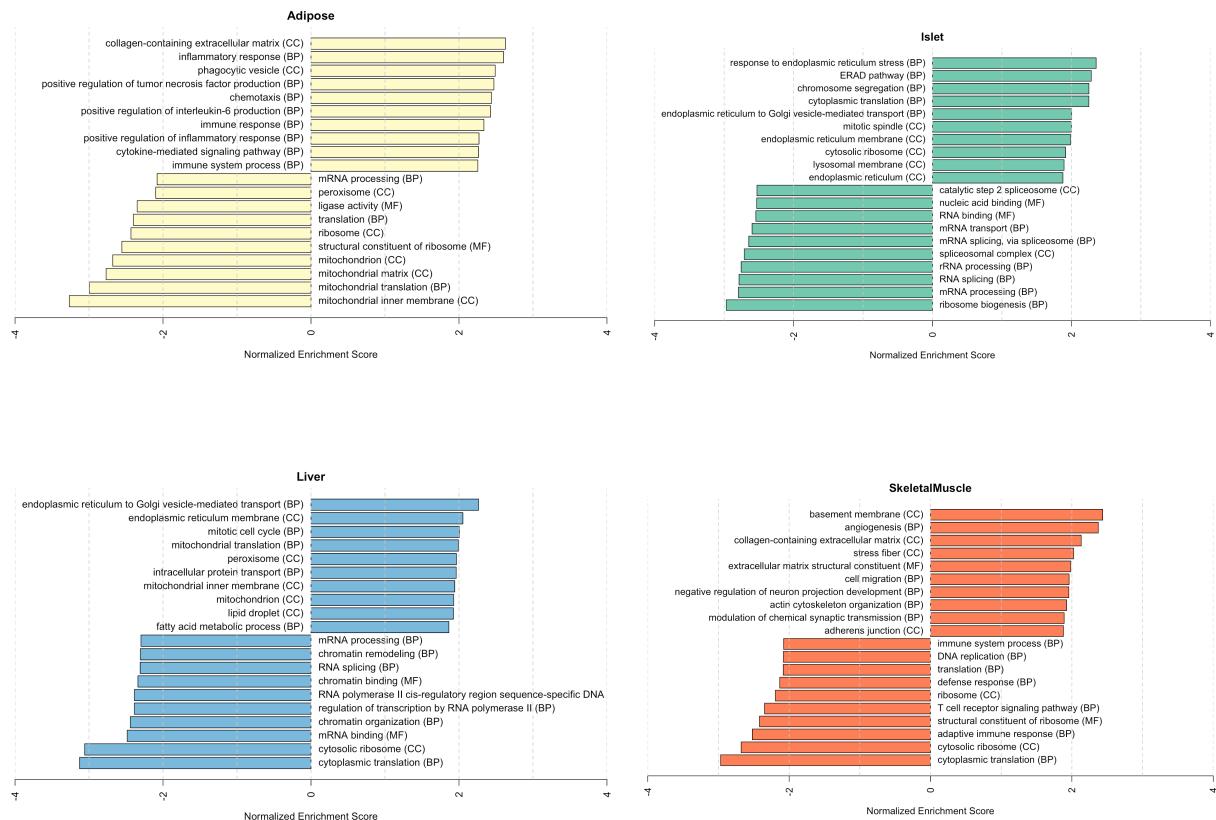


Figure S3: Bar plots showing normalized enrichment scores (NES) for GO terms as determined by fast gene score enrichment analysis (fgsea). Only the top 10 positive and top 10 negative scores are shown. Colors indicate tissue. The name beside each bar shows the name of each enriched GO term. The letters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

- 463      Potential etiologic and functional implications of genome-wide association loci for human diseases and  
464      traits. *Proc Natl Acad Sci*, 106(23):9362–9367, Jun 2009.
- 465      [5] J. K. Pickrell. Joint analysis of functional genomic data and genome-wide association studies of 18  
466      human traits. *Am J Hum Genet*, 94(4):559–573, Apr 2014.
- 467      [6] D. Welter, J. MacArthur, J. Morales, T. Burdett, P. Hall, H. Junkins, A. Klemm, P. Flicek, T. Manolio,  
468      L. Hindorff, and H. Parkinson. The NHGRI GWAS Catalog, a curated resource of SNP-trait associations.  
469      *Nucleic Acids Res*, 42(Database issue):D1001–1006, Jan 2014.
- 470      [7] Y. I. Li, B. van de Geijn, A. Raj, D. A. Knowles, A. A. Petti, D. Golan, Y. Gilad, and J. K. Pritchard.  
471      RNA splicing is a primary link between genetic variation and disease. *Science*, 352(6285):600–604, Apr

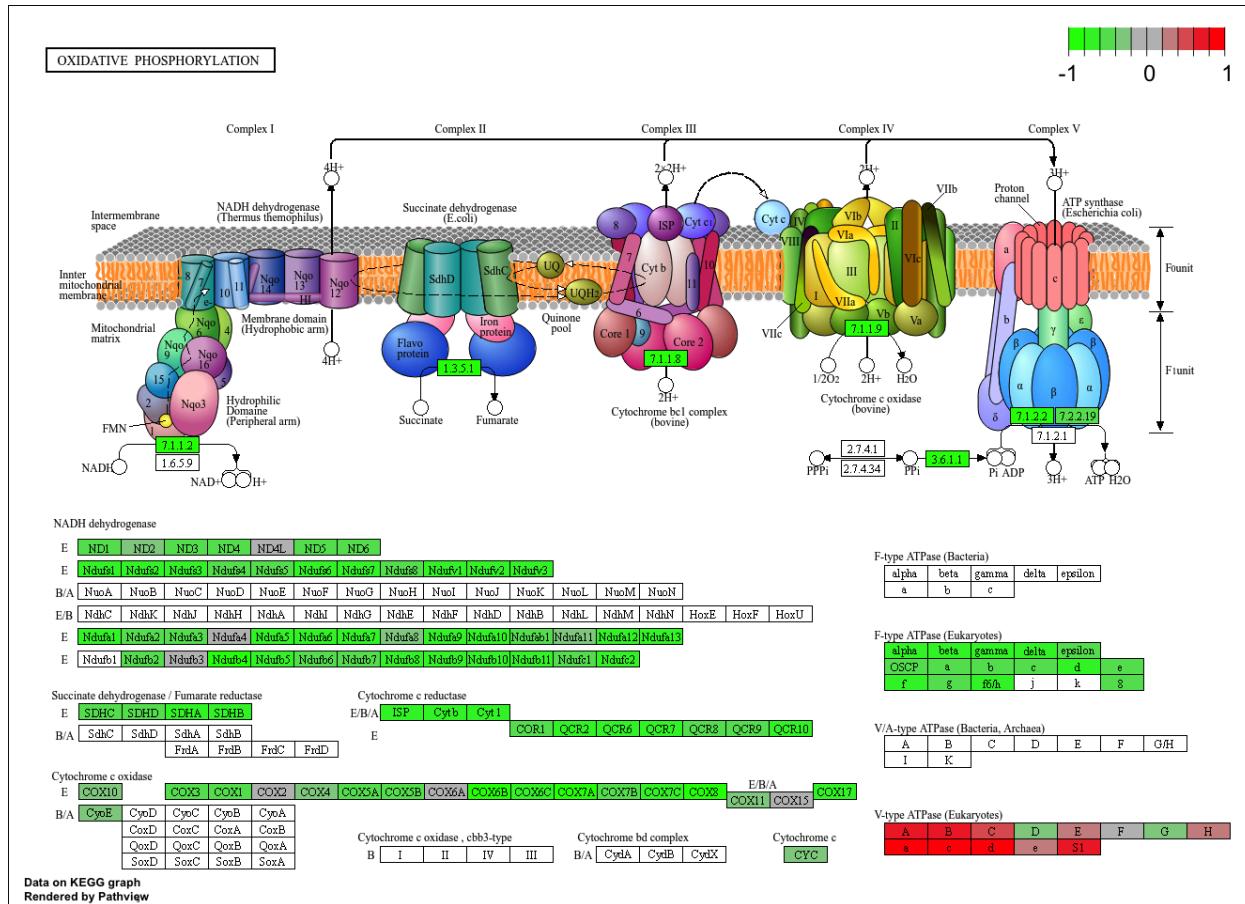


Figure S4: The KEGG pathway for oxidative phosphorylation in mice. Each element is colored based on its HDMA loading from adipose tissue normalized to run from -1 to 1. Genes highlighted in green had negative loadings, and those highlighted in red had positive loadings. Almost the entire pathway was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

472                   2016.

- 473 [8] D. Zhou, Y. Jiang, X. Zhong, N. J. Cox, C. Liu, and E. R. Gamazon. A unified framework for joint-tissue  
474 transcriptome-wide association and Mendelian randomization analysis. *Nat Genet*, 52(11):1239–1246,  
475 Nov 2020.
- 476 [9] E. R. Gamazon, H. E. Wheeler, K. P. Shah, S. V. Mozaffari, K. Aquino-Michaels, R. J. Carroll, A. E.  
477 Eyler, J. C. Denny, D. L. Nicolae, N. J. Cox, and H. K. Im. A gene-based association method for  
478 mapping traits using reference transcriptome data. *Nat Genet*, 47(9):1091–1098, Sep 2015.
- 479 [10] Z. Zhu, F. Zhang, H. Hu, A. Bakshi, M. R. Robinson, J. E. Powell, G. W. Montgomery, M. E. Goddard,  
480 N. R. Wray, P. M. Visscher, and J. Yang. Integration of summary data from GWAS and eQTL studies  
481 predicts complex trait gene targets. *Nat Genet*, 48(5):481–487, May 2016.

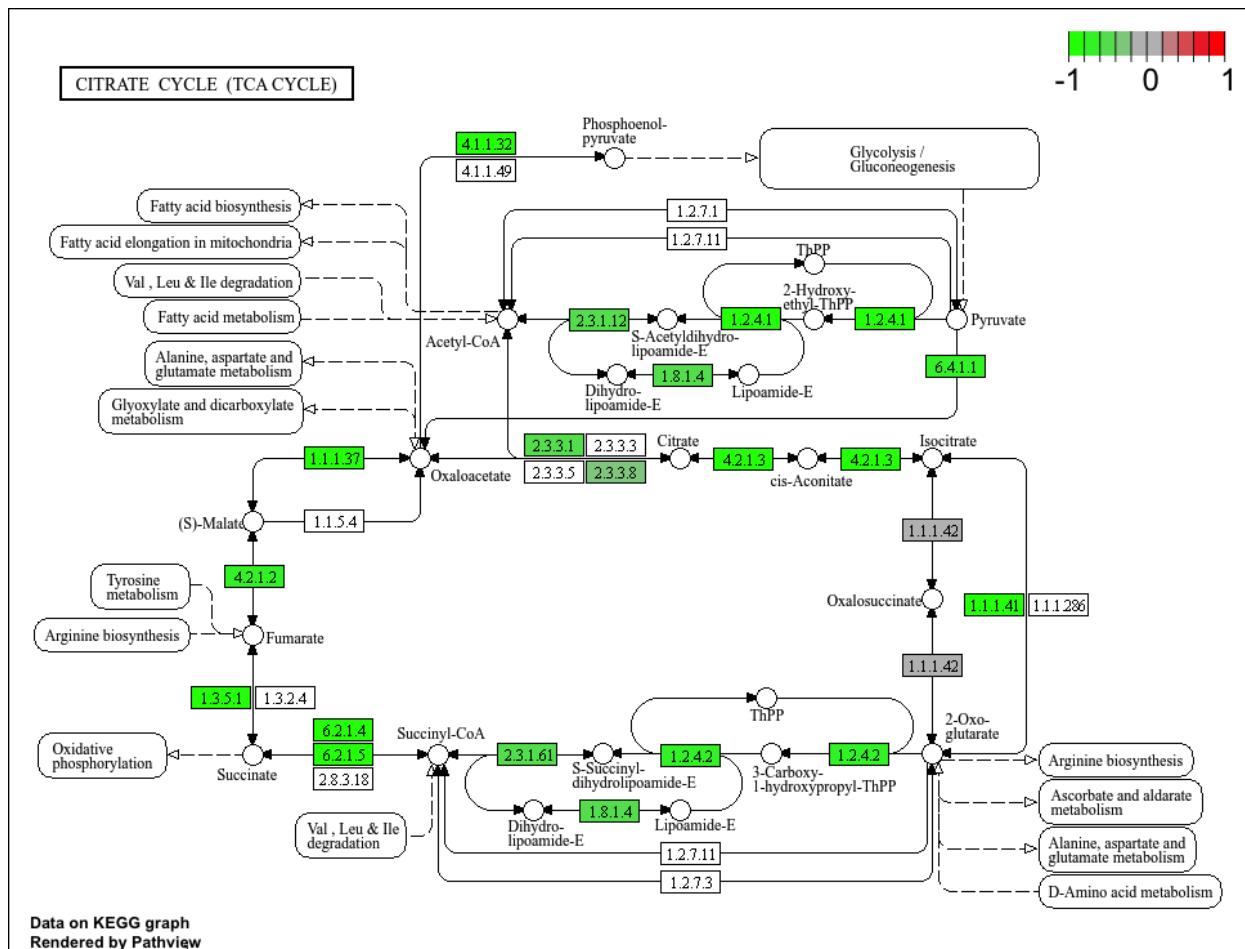


Figure S5: The KEGG pathway for the TCA (citric acid) cycle in mice. Each element is colored based on its HDMA loading from adipose tissue normalized to run from -1 to 1. Genes highlighted in green had negative loadings, and those highlighted in red had positive loadings. Many genes in the cycle were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

- 482 [11] A. Gusev, A. Ko, H. Shi, G. Bhatia, W. Chung, B. W. Penninx, R. Jansen, E. J. de Geus, D. I. Boomsma,  
 483 F. A. Wright, P. F. Sullivan, E. Nikkola, M. Alvarez, M. Civelek, A. J. Lusis, T. ki, E. Raitoharju,  
 484 M. nen, I. ä, O. T. Raitakari, J. Kuusisto, M. Laakso, A. L. Price, P. Pajukanta, and B. Pasaniuc.  
 485 Integrative approaches for large-scale transcriptome-wide association studies. *Nat Genet*, 48(3):245–252,  
 486 Mar 2016.
- 487 [12] M. P. Keller, D. M. Gatti, K. L. Schueler, M. E. Rabaglia, D. S. Stapleton, P. Simecek, M. Vincent,  
 488 S. Allen, A. T. Broman, R. Bacher, C. Kendziora, K. W. Broman, B. S. Yandell, G. A. Churchill, and  
 489 A. D. Attie. Genetic Drivers of Pancreatic Islet Function. *Genetics*, 209(1):335–356, May 2018.
- 490 [13] W. L. Crouse, G. R. Keele, M. S. Gastonguay, G. A. Churchill, and W. Valdar. A Bayesian model  
 491 selection approach to mediation analysis. *PLoS Genet*, 18(5):e1010184, May 2022.

- 492 [14] J. M. Chick, S. C. Munger, P. Simecek, E. L. Huttlin, K. Choi, D. M. Gatti, N. Raghupathy, K. L. Svenson,  
493 G. A. Churchill, and S. P. Gygi. Defining the consequences of genetic variation on a proteome-wide scale.  
494 *Nature*, 534(7608):500–505, Jun 2016.
- 495 [15] H. E. Wheeler, S. Ploch, A. N. Barbeira, R. Bonazzola, A. Andaleon, A. Fotuhi Siahpirani, A. Saha,  
496 A. Battle, S. Roy, and H. K. Im. Imputed gene associations identify replicable trans-acting genes enriched  
497 in transcription pathways and complex traits. *Genet Epidemiol*, 43(6):596–608, Sep 2019.
- 498 [16] B. D. Umans, A. Battle, and Y. Gilad. Where Are the Disease-Associated eQTLs? *Trends Genet*,  
499 37(2):109–124, Feb 2021.
- 500 [17] N. J. Connally, S. Nazeen, D. Lee, H. Shi, J. Stamatoyannopoulos, S. Chun, C. Cotsapas, C. A. Cassa,  
501 and S. R. Sunyaev. The missing link between genetic association and regulatory function. *Elife*, 11, Dec  
502 2022.
- 503 [18] H. Mostafavi, J. P. Spence, S. Naqvi, and J. K. Pritchard. Systematic differences in discovery of genetic  
504 effects on gene expression and complex traits. *Nat Genet*, 55(11):1866–1875, Nov 2023.
- 505 [19] D. W. Yao, L. J. O’Connor, A. L. Price, and A. Gusev. Quantifying genetic effects on disease mediated  
506 by assayed gene expression levels. *Nat Genet*, 52(6):626–633, Jun 2020.
- 507 [20] X. Liu, J. A. Mefford, A. Dahl, Y. He, M. Subramaniam, A. Battle, A. L. Price, and N. Zaitlen. GBAT:  
508 a gene-based association test for robust detection of trans-gene regulation. *Genome Biol*, 21(1):211, Aug  
509 2020.
- 510 [21] G. A. Churchill, D. M. Gatti, S. C. Munger, and K. L. Svenson. The Diversity Outbred mouse population.  
511 *Mamm Genome*, 23(9-10):713–718, Oct 2012.
- 512 [22] Elissa J Chesler, Darla R Miller, Lisa R Branstetter, Leslie D Galloway, Barbara L Jackson, Vivek M  
513 Philip, Brynn H Voy, Cymbeline T Culiat, David W Threadgill, Robert W Williams, et al. The  
514 collaborative cross at oak ridge national laboratory: developing a powerful resource for systems genetics.  
515 *Mammalian Genome*, 19:382–389, 2008.
- 516 [23] Michael C Saul, Vivek M Philip, Laura G Reinholdt, and Elissa J Chesler. High-diversity mouse  
517 populations for complex traits. *Trends in Genetics*, 35(7):501–514, 2019.
- 518 [24] S. M. Clee and A. D. Attie. The genetic landscape of type 2 diabetes in mice. *Endocr Rev*, 28(1):48–83,  
519 Feb 2007.
- 520 [25] K. W. Broman, D. M. Gatti, P. Simecek, N. A. Furlotte, P. Prins, Š. Sen, B. S. Yandell, and G. A.

- 521 Churchill. R/qt12: Software for Mapping Quantitative Trait Loci with High-Dimensional Data and  
522 Multiparent Populations. *Genetics*, 211(2):495–502, Feb 2019.
- 523 [26] Fabien Girka, Etienne Camenen, Caroline Peltier, Arnaud Gloaguen, Vincent Guillemot, Laurent Le  
524 Brusquet, and Arthur Tenenhaus. *RGCCA: Regularized and Sparse Generalized Canonical Correlation*  
525 *Analysis for Multiblock Data*, 2023. R package version 3.0.3.
- 526 [27] Gennady Korotkevich, Vladimir Sukhov, and Alexey Sergushichev. Fast gene set enrichment analysis.  
527 *bioRxiv*, 2019.
- 528 [28] A. Subramanian, P. Tamayo, V. K. Mootha, S. Mukherjee, B. L. Ebert, M. A. Gillette, A. Paulovich,  
529 S. L. Pomeroy, T. R. Golub, E. S. Lander, and J. P. Mesirov. Gene set enrichment analysis: a  
530 knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci U S A*,  
531 102(43):15545–15550, Oct 2005.
- 532 [29] V. Kus, T. Prazak, P. Brauner, M. Hensler, O. Kuda, P. Flachs, P. Janovska, D. Medrikova, M. Rossmeisl,  
533 Z. Jilkova, B. Stefl, E. Pastalkova, Z. Drahota, J. Houstek, and J. Kopecky. Induction of muscle  
534 thermogenesis by high-fat diet in mice: association with obesity-resistance. *Am J Physiol Endocrinol  
535 Metab*, 295(2):E356–367, Aug 2008.
- 536 [30] C. B. Newgard. Interplay between lipids and branched-chain amino acids in development of insulin  
537 resistance. *Cell Metab*, 15(5):606–614, May 2012.
- 538 [31] D. D. Sears, G. Hsiao, A. Hsiao, J. G. Yu, C. H. Courtney, J. M. Ofrecio, J. Chapman, and S. Subramaniam.  
539 Mechanisms of human insulin resistance and thiazolidinedione-mediated insulin sensitization. *Proc Natl  
540 Acad Sci U S A*, 106(44):18745–18750, Nov 2009.
- 541 [32] R. Stienstra, C. Duval, M. ller, and S. Kersten. PPARs, Obesity, and Inflammation. *PPAR Res*,  
542 2007:95974, 2007.
- 543 [33] O. Gavrilova, M. Haluzik, K. Matsusue, J. J. Cutson, L. Johnson, K. R. Dietz, C. J. Nicol, C. Vinson,  
544 F. J. Gonzalez, and M. L. Reitman. Liver peroxisome proliferator-activated receptor gamma contributes  
545 to hepatic steatosis, triglyceride clearance, and regulation of body fat mass. *J Biol Chem*, 278(36):34268–  
546 34276, Sep 2003.
- 547 [34] K. Matsusue, M. Haluzik, G. Lambert, S. H. Yim, O. Gavrilova, J. M. Ward, B. Brewer, M. L. Reitman,  
548 and F. J. Gonzalez. Liver-specific disruption of PPARgamma in leptin-deficient mice improves fatty  
549 liver but aggravates diabetic phenotypes. *J Clin Invest*, 111(5):737–747, Mar 2003.
- 550 [35] D. Patsouris, J. K. Reddy, M. ller, and S. Kersten. Peroxisome proliferator-activated receptor alpha

- 551 mediates the effects of high-fat diet on hepatic gene expression. *Endocrinology*, 147(3):1508–1516, Mar  
552 2006.
- 553 [36] S. E. Schadinger, N. L. Bucher, B. M. Schreiber, and S. R. Farmer. PPARgamma2 regulates lipogenesis  
554 and lipid accumulation in steatotic hepatocytes. *Am J Physiol Endocrinol Metab*, 288(6):E1195–1205,  
555 Jun 2005.
- 556 [37] W. Motomura, M. Inoue, T. Ohtake, N. Takahashi, M. Nagamine, S. Tanno, Y. Kohgo, and T. Okumura.  
557 Up-regulation of ADRP in fatty liver in human and liver steatosis in mice fed with high fat diet. *Biochem  
558 Biophys Res Commun*, 340(4):1111–1118, Feb 2006.
- 559 [38] J. Lamb, E. D. Crawford, D. Peck, J. W. Modell, I. C. Blat, M. J. Wrobel, J. Lerner, J. P. Brunet,  
560 A. Subramanian, K. N. Ross, M. Reich, H. Hieronymus, G. Wei, S. A. Armstrong, S. J. Haggarty,  
561 P. A. Clemons, R. Wei, S. A. Carr, E. S. Lander, and T. R. Golub. The Connectivity Map: using  
562 gene-expression signatures to connect small molecules, genes, and disease. *Science*, 313(5795):1929–1935,  
563 Sep 2006.
- 564 [39] S. Amin, A. Lux, and F. O’Callaghan. The journey of metformin from glycaemic control to mTOR  
565 inhibition and the suppression of tumour growth. *Br J Clin Pharmacol*, 85(1):37–46, Jan 2019.
- 566 [40] A. Kezic, L. Popovic, and K. Lalic. mTOR Inhibitor Therapy and Metabolic Consequences: Where Do  
567 We Stand? *Oxid Med Cell Longev*, 2018:2640342, 2018.
- 568 [41] A. D. Barlow, M. L. Nicholson, and T. P. Herbert. -cells and a review of the underlying molecular  
569 mechanisms. *Diabetes*, 62(8):2674–2682, Aug 2013.
- 570 [42] Y. Gu, J. Lindner, A. Kumar, W. Yuan, and M. A. Magnuson. Rictor/mTORC2 is essential for  
571 maintaining a balance between beta-cell proliferation and cell size. *Diabetes*, 60(3):827–837, Mar 2011.
- 572 [43] E. B. Geer, J. Islam, and C. Buettner. Mechanisms of glucocorticoid-induced insulin resistance: focus on  
573 adipose tissue function and lipid metabolism. *Endocrinol Metab Clin North Am*, 43(1):75–102, Mar 2014.
- 574 [44] J. X. Li and C. L. Cummins. Fresh insights into glucocorticoid-induced diabetes mellitus and new  
575 therapeutic directions. *Nat Rev Endocrinol*, 18(9):540–557, Sep 2022.
- 576 [45] R. A. Lee, C. A. Harris, and J. C. Wang. Glucocorticoid Receptor and Adipocyte Biology. *Nucl Receptor  
577 Res*, 5, 2018.
- 578 [46] S. Viengchareun, P. Penfornis, M. C. Zennaro, and M. s. Mineralocorticoid and glucocorticoid receptors

- 579 inhibit UCP expression and function in brown adipocytes. *Am J Physiol Endocrinol Metab*, 280(4):E640–  
580 649, Apr 2001.
- 581 [47] J. Liu, X. Kong, L. Wang, H. Qi, W. Di, X. Zhang, L. Wu, X. Chen, J. Yu, J. Zha, S. Lv, A. Zhang,  
582 P. Cheng, M. Hu, Y. Li, J. Bi, Y. Li, F. Hu, Y. Zhong, Y. Xu, and G. Ding. -HSD1 in regulating brown  
583 adipocyte function. *J Mol Endocrinol*, 50(1):103–113, Feb 2013.
- 584 [48] L. E. Ramage, M. Akyol, A. M. Fletcher, J. Forsythe, M. Nixon, R. N. Carter, E. J. van Beek,  
585 N. M. Morton, B. R. Walker, and R. H. Stimson. Glucocorticoids Acutely Increase Brown Adipose  
586 Tissue Activity in Humans, Revealing Species-Specific Differences in UCP-1 Regulation. *Cell Metab*,  
587 24(1):130–141, Jul 2016.
- 588 [49] J. L. Barclay, H. Agada, C. Jang, M. Ward, N. Wetzig, and K. K. Ho. Effects of glucocorticoids on  
589 human brown adipocytes. *J Endocrinol*, 224(2):139–147, Feb 2015.
- 590 [50] M. ó, R. Gupte, W. L. Kraus, P. Pacher, and P. Bai. PARPs in lipid metabolism and related diseases.  
591 *Prog Lipid Res*, 84:101117, Nov 2021.
- 592 [51] P. Bai, C. ó, H. Oudart, A. nszki, Y. Cen, C. Thomas, H. Yamamoto, A. Huber, B. Kiss, R. H.  
593 Houtkooper, K. Schoonjans, V. Schreiber, A. A. Sauve, J. Menissier-de Murcia, and J. Auwerx. PARP-1  
594 inhibition increases mitochondrial metabolism through SIRT1 activation. *Cell Metab*, 13(4):461–468,  
595 Apr 2011.
- 596 [52] A. Chiarugi and M. A. Moskowitz. Cell biology. PARP-1—a perpetrator of apoptotic cell death? *Science*,  
597 297(5579):200–201, Jul 2002.
- 598 [53] M. Althubiti, R. Almaimani, S. Y. Eid, M. Elzubaier, B. Refaat, S. Idris, T. A. Alqurashi, and M. Z.  
599 El-Readi. BTK targeting suppresses inflammatory genes and ameliorates insulin resistance. *Eur Cytokine  
600 Netw*, 31(4):168–179, Dec 2020.
- 601 [54] C. Skrabs, W. F. Pickl, T. Perkmann, U. ger, and A. Gessl. Rapid decline in insulin antibodies and  
602 glutamic acid decarboxylase autoantibodies with ibrutinib therapy of chronic lymphocytic leukaemia. *J  
603 Clin Pharm Ther*, 43(1):145–149, Feb 2018.
- 604 [55] E. A. Oral, S. M. Reilly, A. V. Gomez, R. Meral, L. Butz, N. Ajluni, T. L. Chenevert, E. Korytnaya,  
605 A. H. Neider, R. Hench, D. Rus, J. F. Horowitz, B. Poirier, P. Zhao, K. Lehmann, M. Jain, R. Yu,  
606 C. Liddle, M. Ahmadian, M. Downes, R. M. Evans, and A. R. Saltiel. and TBK1 Improves Glucose  
607 Control in a Subset of Patients with Type 2 Diabetes. *Cell Metab*, 26(1):157–170, Jul 2017.
- 608 [56] M. C. Arkan, A. L. Hevener, F. R. Greten, S. Maeda, Z. W. Li, J. M. Long, A. Wynshaw-Boris, G. Poli,

- 609 J. Olefsky, and M. Karin. IKK-beta links inflammation to obesity-induced insulin resistance. *Nat Med*,  
610 11(2):191–198, Feb 2005.
- 611 [57] M. Clark, C. J. Kroger, Q. Ke, and R. M. Tisch. The Role of T Cell Receptor Signaling in the  
612 Development of Type 1 Diabetes. *Front Immunol*, 11:615371, 2020.
- 613 [58] P. Endell, A. C. Carlsson, A. Larsson, O. Melander, T. Wessman, J. v, and T. Ruge. TNFR1 is associated  
614 with short-term mortality in patients with diabetes and acute dyspnea seeking care at the emergency  
615 department. *Acta Diabetol*, 57(10):1145–1150, Oct 2020.
- 616 [59] A. M. Keestra-Gounder, M. X. Byndloss, N. Seyffert, B. M. Young, A. vez Arroyo, A. Y. Tsai, S. A.  
617 Cevallos, M. G. Winter, O. H. Pham, C. R. Tiffany, M. F. de Jong, T. Kerrinnes, R. Ravindran, P. A.  
618 Luciw, S. J. McSorley, A. J. umler, and R. M. Tsolis. NOD1 and NOD2 signalling links ER stress with  
619 inflammation. *Nature*, 532(7599):394–397, Apr 2016.
- 620 [60] A. M. Keestra-Gounder and R. M. Tsolis. NOD1 and NOD2: Beyond Peptidoglycan Sensing. *Trends  
621 Immunol*, 38(10):758–767, Oct 2017.
- 622 [61] J. Montane, L. Cadavez, and A. Novials. Stress and the inflammatory process: a major cause of  
623 pancreatic cell death in type 2 diabetes. *Diabetes Metab Syndr Obes*, 7:25–34, 2014.
- 624 [62] B. B. Kahn and T. E. McGraw. , and type 2 diabetes. *N Engl J Med*, 363(27):2667–2669, Dec 2010.
- 625 [63] S. Del Prato and N. Pulizzi. The place of sulfonylureas in the therapy for type 2 diabetes mellitus.  
626 *Metabolism*, 55(5 Suppl 1):S20–27, May 2006.
- 627 [64] X. Liu, Y. I. Li, and J. K. Pritchard. Trans Effects on Gene Expression Can Drive Omnipotent Inheritance.  
628 *Cell*, 177(4):1022–1034, May 2019.
- 629 [65] B. Hallgrímsson, R. M. Green, D. C. Katz, J. L. Fish, F. P. Bernier, C. C. Roseman, N. M. Young,  
630 J. M. Cheverud, and R. S. Marcucio. The developmental-genetics of canalization. *Semin Cell Dev Biol*,  
631 88:67–79, Apr 2019.
- 632 [66] M. L. Siegal and A. Bergman. Waddington's canalization revisited: developmental stability and evolution.  
633 *Proc Natl Acad Sci U S A*, 99(16):10528–10532, Aug 2002.
- 634 [67] A. B. Paaby and G. Gibson. Cryptic Genetic Variation in Evolutionary Developmental Genetics. *Biology  
635 (Basel)*, 5(2), Jun 2016.
- 636 [68] E. A. Boyle, Y. I. Li, and J. K. Pritchard. An Expanded View of Complex Traits: From Polygenic to  
637 Omnipotent. *Cell*, 169(7):1177–1186, Jun 2017.

- 638 [69] Naomi R Wray, Cisca Wijmenga, Patrick F Sullivan, Jian Yang, and Peter M Visscher. Common disease  
639 is more complex than implied by the core gene omnigenic model. *Cell*, 173(7):1573–1580, 2018.
- 640 [70] M. Riva, M. D. Nitert, U. Voss, R. Sathanoori, A. Lindqvist, C. Ling, and N. Wierup. Nesfatin-1  
641 stimulates glucagon and insulin secretion and beta cell NUCB2 is reduced in human type 2 diabetic  
642 subjects. *Cell Tissue Res*, 346(3):393–405, Dec 2011.
- 643 [71] M. Nakata and T. Yada. Role of NUCB2/nesfatin-1 in glucose control: diverse functions in islets,  
644 adipocytes and brain. *Curr Pharm Des*, 19(39):6960–6965, 2013.
- 645 [72] H. Shimizu and A. Osaki. Nesfatin/Nucleobindin-2 (NUCB2) and Glucose Homeostasis. *Curr Hypertens  
646 Rev*, pages Nesfatin/Nucleobindin-2 (NUCB2) and Glucose Homeostasis., Jul 2014.

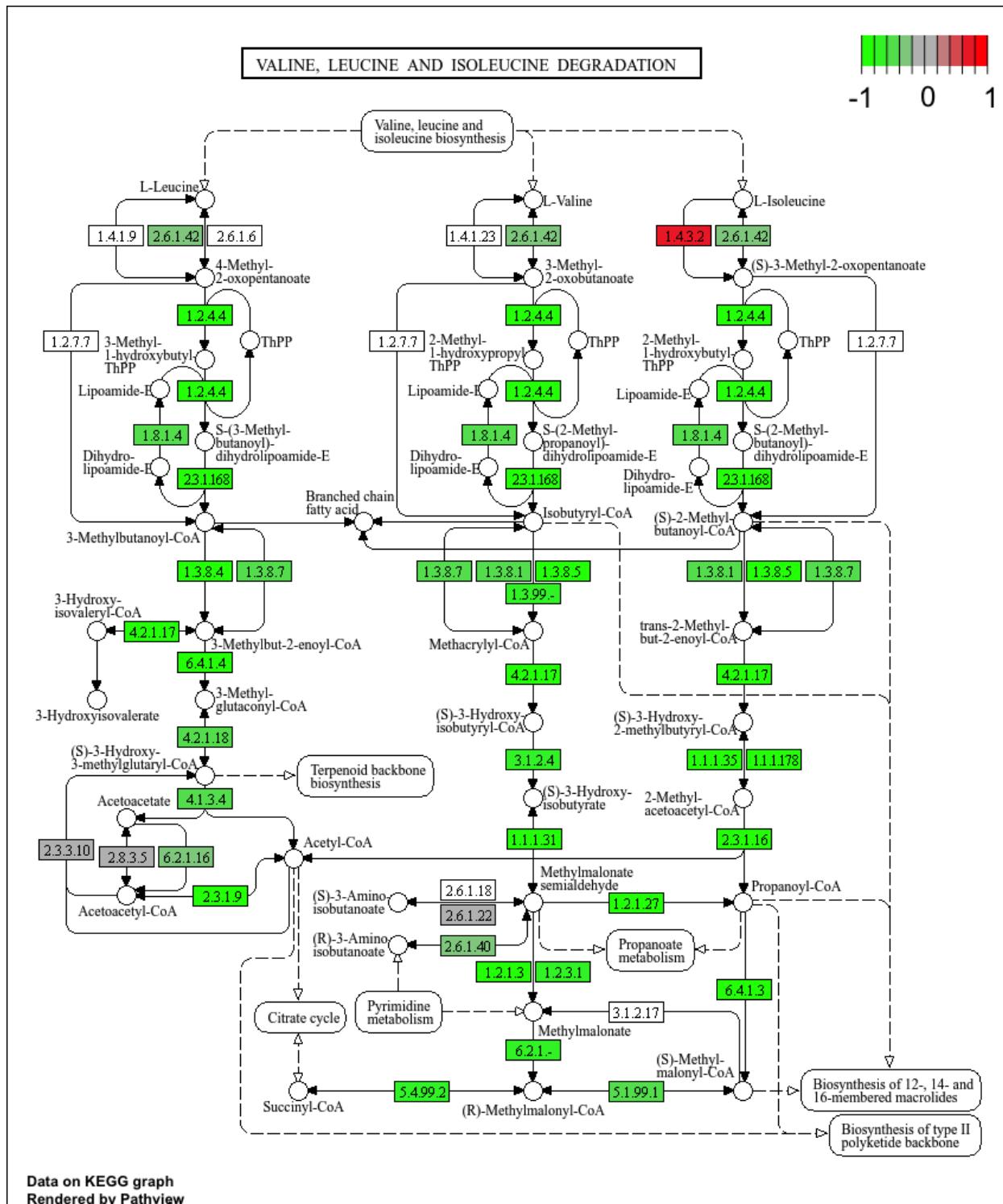


Figure S6: The KEGG pathway for branched-chain amino acid degradation in mice. Each element is colored based on its HDMA loading from adipose tissue normalized to run from -1 to 1. Genes highlighted in green had negative loadings, and those highlighted in red had positive loadings. Almost the entire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

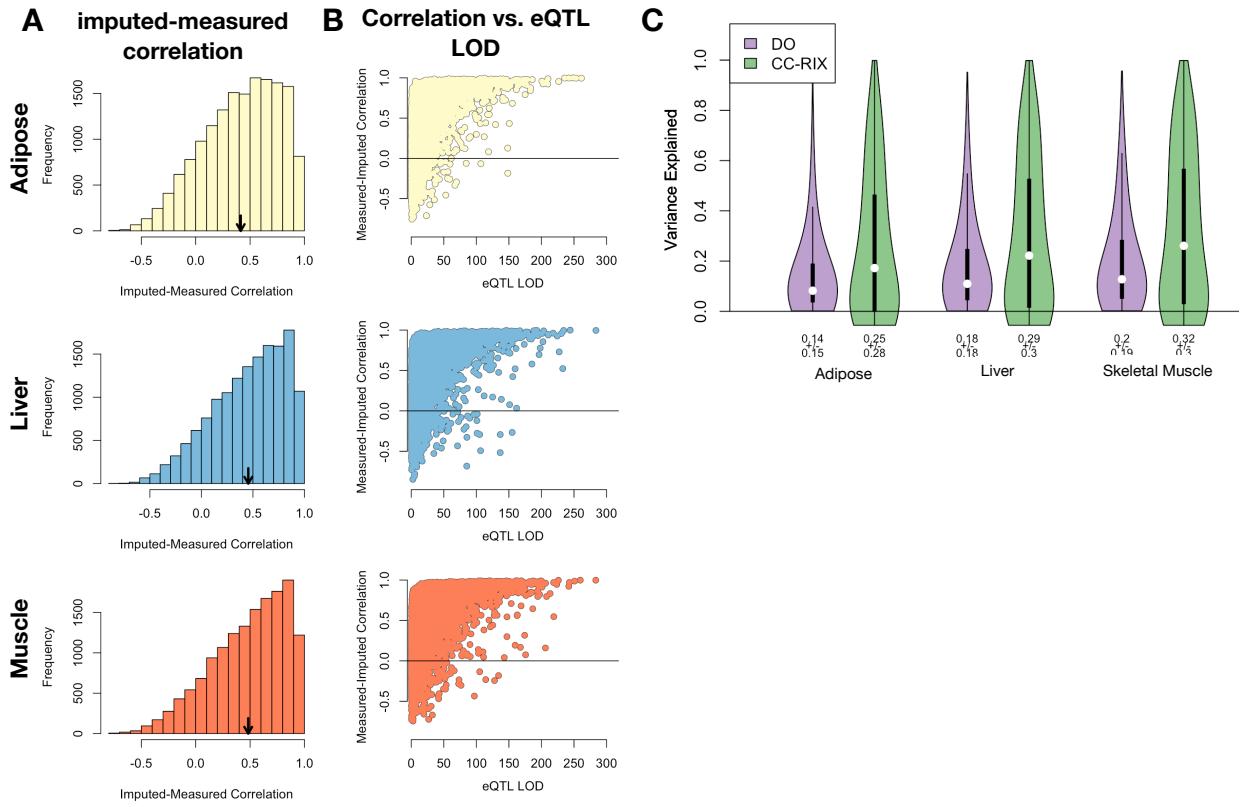


Figure S7: Validation of transcript imputation in the CC-RIX. **A.** Distributions of correlations between imputed and measured transcripts in the CC-RIX. The mean of each distribution is shown by the red line. All distributions were skewed toward positive correlations and had positive means near a Pearson 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_RECECTORAGONIST
		HCC515	TRT_CP	-0.76	15.65	MOA_CLASS	GLUCOCORTICOID_RECECTORAGONIST
		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_RECECTORAGONIST
		A549	TRT_CP	-0.67	15.65	MOA_CLASS	GLUCOCORTICOID_RECECTORAGONIST
		-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_RECECTORAGONIST
		HCC515	TRT_CP	-0.80	15.65	MOA_CLASS	GLUCOCORTICOID_RECECTORAGONIST
		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.

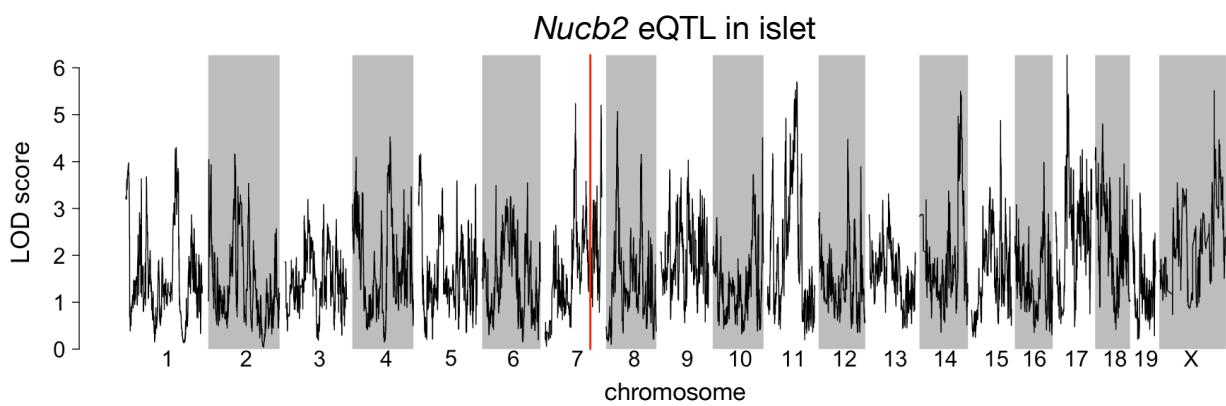


Figure S12: Regulation of *Nucb2* expression in islet. *Nucb2* is encoded on mouse chromosome 7 at 116.5 Mb (red line). In islets the heritability of *Nucb2* expression levels is 69% heritable. This LOD score trace shows that there is no local eQTL at that position, nor any strong distal eQTL anywhere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