

<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 J. Braun, Nadia A. Rosenthal, Alan D. Attie, Gary A. Churchill and  
<sup>6</sup> Gregory W. Carter

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

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

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

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

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

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

<sup>31</sup> One possible explanation for these observations is that gene expression is not being measured in the appropriate  
<sup>32</sup> cell types<sup>16</sup>, or cell state<sup>19</sup> and thus local eQTLs influencing traits cannot be detected. An alternative  
<sup>33</sup> explanation that has been discussed in recent years is that effects of these variants are mediated not through  
<sup>34</sup> local regulation of gene expression, but through distal regulation<sup>18;20;21;15</sup>. In this model, a gene's expression  
<sup>35</sup> is influenced by many variants throughout the genome through their cumulative effects on a broader regulatory  
<sup>36</sup> network. In other words, the heritable component of the transcriptome is an emergent state arising from the  
<sup>37</sup> myriad molecular interactions defining and constraining gene expression.

<sup>38</sup> To assess the role of wide-spread distal gene regulation on complex traits, we investigated diet-induced  
<sup>39</sup> obesity and metabolic disease as an archetypal example. Diet-induced obesity and metabolic disease are  
<sup>40</sup> genetically complex with hundreds of variants mapped through GWAS<sup>22;23</sup>. These variants are known to act  
<sup>41</sup> through multiple tissues that interact dynamically with each other<sup>24;25</sup>, including adipose tissue, pancreatic  
<sup>42</sup> islets, liver, and skeletal muscle. The multi-system etiology of metabolic disease complicates mechanistic  
<sup>43</sup> dissection of the genetic architecture, requiring large, dedicated data sets that include high-dimensional,  
<sup>44</sup> clinically relevant phenotyping, dense genotyping in a highly recombined population, and transcriptome-wide  
<sup>45</sup> measurements of gene expression in multiple tissues.

<sup>46</sup> Measuring gene expression in multiple tissues is critical to adequately assess the extent to which local gene  
<sup>47</sup> regulation varies across the tissues and whether such variability might account for previous failed attempts to  
<sup>48</sup> identify trait-relevant local eQTL. Such data sets are extremely difficult to obtain in human populations,  
<sup>49</sup> particularly in the large numbers of subjects required for adequate statistical power. Thus, to further  
<sup>50</sup> investigate the role of local and distal gene regulation on complex traits, we generated two complementary  
<sup>51</sup> data sets: A discovery data set in a large population of diversity outbred (DO) mice<sup>26</sup>, and an independent  
<sup>52</sup> validation data set derived by crossing inbred strains from the Collaborative Cross (CC) mice<sup>27</sup> to form CC  
<sup>53</sup> F1 mice (CC-RIX). Both populations modeled diet-induced obesity and metabolic disease<sup>12</sup>.

<sup>54</sup> The DO population and CC recombinant inbred lines were derived from the same eight inbred founder mouse  
<sup>55</sup> strains, five classical lab strains, and three strains more recently derived from wild mice<sup>26</sup>. They represent  
<sup>56</sup> three subspecies of mouse *Mus musculus domesticus*, *Mus musculus musculus*, and *Mus musculus castaneus*,

57 and capture 90% of the known variation in laboratory mice<sup>28</sup>. The DO mice are maintained with a breeding  
58 scheme that ensures equal contributions from each founder across the genome thus rendering almost the  
59 whole genome visible to genetic inquiry<sup>26</sup>. The CC mice were initially outcrossed to recombine the genomes  
60 from all eight founders, and then inbred for at least 20 generations to generate multiple inbred lines. Because  
61 these two populations have common ancestral haplotypes, we could directly and unambiguously compare  
62 the local genetic effects on gene expression at the whole-transcriptome level while varying the population  
63 structure driving distal regulation.

64 In the DO population, we paired clinically relevant metabolic traits from 371 mice<sup>12</sup>, including body weight,  
65 plasma levels of insulin, glucose and lipids, with transcriptome-wide gene expression in four tissues related to  
66 metabolic disease: adipose tissue, pancreatic islets, liver, and skeletal muscle. We measured similar metabolic  
67 traits in a CC-RIX population and gene expression from three of the four tissues used in the DO: adipose  
68 tissue, liver, and skeletal muscle. Because the CC-RIX carry the same founder alleles as the DO, local gene  
69 regulation is expected to match between the populations, but because the alleles are recombined through  
70 the genome, distal effects are expected to vary from those in the DO, allowing us to directly assess the  
71 role of local gene regulation in driving trait-associated transcript variation. Together, these data enable a  
72 comprehensive view into the genetic architecture of metabolic disease.

## 73 Results

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

### 76 Genetic variation contributed to wide phenotypic variation

77 Although the environment was consistent across the DO mice, the genetic diversity present in this population  
78 resulted in widely varying distributions across physiological measurements (Fig. 1). For example, body  
79 weights of adult individuals varied from less than the average adult C57BL/6J (B6) body weight to several  
80 times the body weight of a B6 adult in both sexes (Males: 18.5 - 69.1g, Females: 16.0 - 54.8g) (Fig. 1A).  
81 Fasting blood glucose (FBG) also varied considerably (Fig. 1B), although few of the animals had FBG levels  
82 that would indicate pre-diabetes (19 animals, 3.8%), or diabetes (7 animals, 1.4%) according to previously  
83 developed cutoffs (pre-diabetes: FBG  $\geq$  250 mg/dL, diabetes: FBG  $\geq$  300, mg/dL)<sup>29</sup>. Males had higher  
84 FBG than females on average (Fig. 1C) as has been observed before suggesting either that males were more  
85 susceptible to metabolic disease on the high-fat, high-sugar (HFHS) diet, or that males and females may  
86 require different thresholds for pre-diabetes and diabetes.

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

91 The trait correlations (Fig. 1G) showed that most of the metabolic trait pairs were only modestly correlated,  
92 which combined with the trait decomposition (Supp. Fig. S1), suggest complex relationships among the  
93 measured traits and a broad sampling of multiple heritable aspects of metabolic disease including overall  
94 body weight, glucose homeostasis, and pancreatic function.

#### 95 **Distal Heritability Correlated with Phenotype Relevance**

96 To comprehensively assess the genetic control of gene expression in metabolic disease we measured overall  
97 gene expression via bulk RNA-Seq in adipose, islet, liver, and skeletal muscle in the DO cohort (Supp. Fig.  
98 S2A-H). We performed eQTL analysis using R/qltl2<sup>30</sup> (Methods) and identified both local and distal eQTLs  
99 for transcripts in each of the four tissues (Supp. Fig. S2). Significant local eQTLs far outnumbered distal  
100 eQTLs (Supp. Fig. S2F) and tended to be shared across tissues (Supp. Fig. S2G) whereas the few significant  
101 distal eQTLs we identified tended to be tissue-specific (Supp. Fig. S2H)

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>20</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.

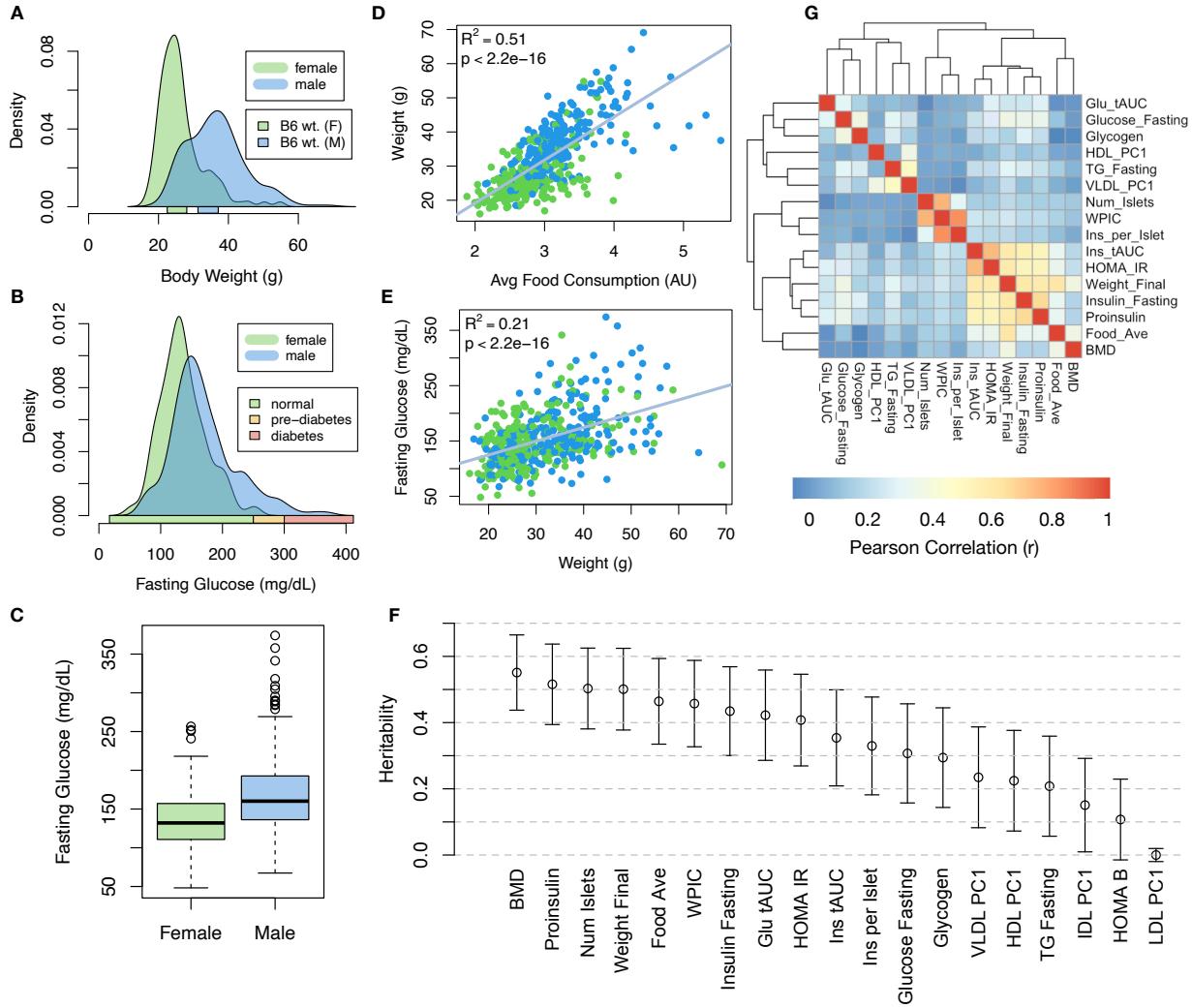


Figure 1: Clinical overview. **A.** Distributions of body weight in the diversity outbred mice. Sex is indicated by color. The average B6 male and female adult weights at 24 weeks of age are indicated by blue and green bars on the x-axis. **B.** The distribution of fasting glucose across the population split by sex. Normal, pre-diabetic, and diabetic fasting glucose levels for mice are shown by colored bars along the x-axis. **C.** Males had higher fasting blood glucose on average than females. **D.** The relationship between food consumption and body weight for both sexes. **E.** Relationship between body weight and fasting glucose for both sexes. **F.** Heritability estimates for each physiological trait. Bars show standard error of the estimate. **G.** Correlation structure between pairs of physiological traits. BMD - bone mineral density, WPIC - whole pancreas insulin content, Glu tAUC - glucose total area under the curve, HOMA IR - homeostatic measurement of insulin resistance, HOMA B - homeostatic measure of beta cell health, VLDL - very low-density lipoprotein, LDL - low-density lipoprotein, IDL - intermediate density lipoprotein, HDL - high-density lipoprotein, TG - triglyceride.

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

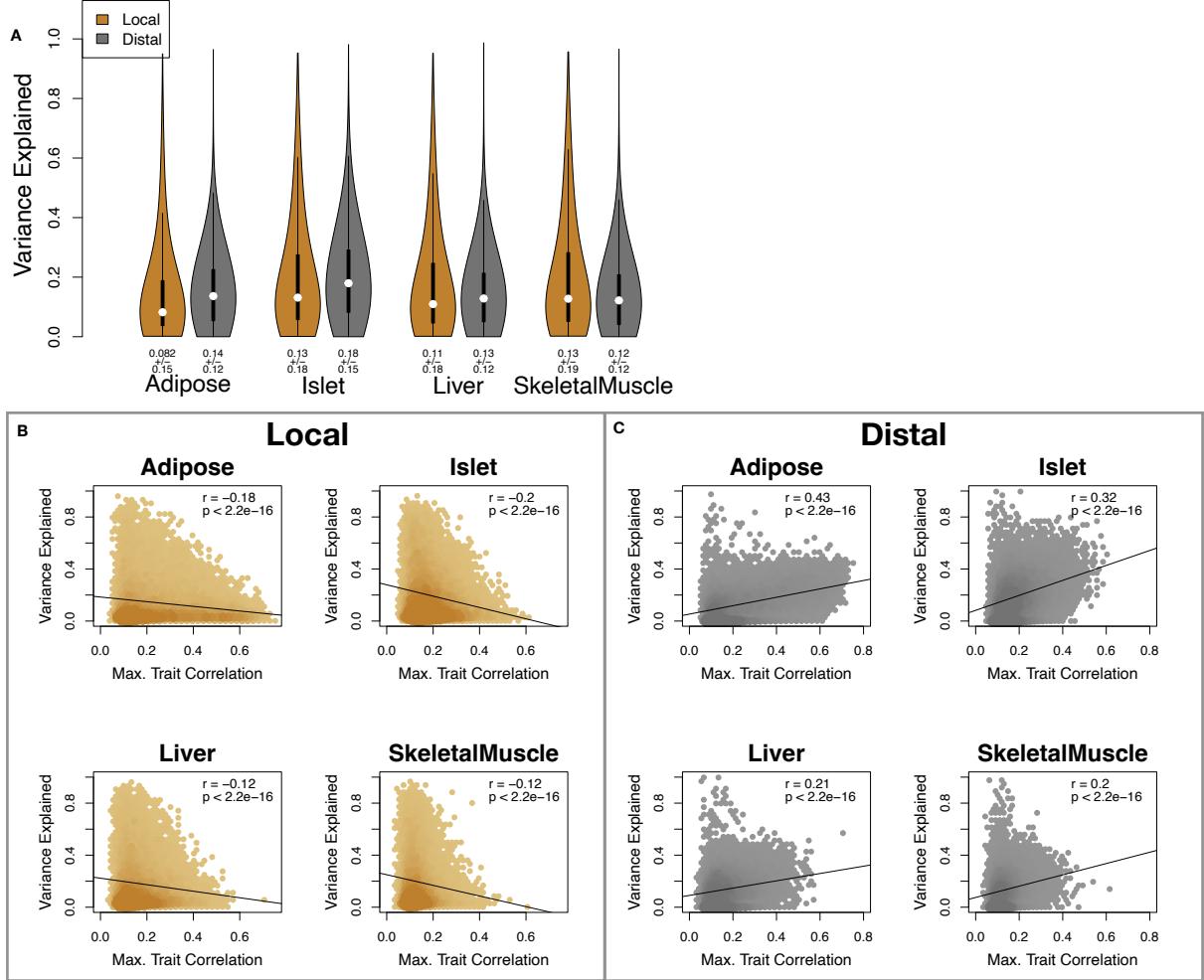


Figure 2: Transcript heritability and trait relevance. **A.** Distributions of distal and local heritability of transcripts across the four tissues. Overall local and distal factors contribute equally to transcript heritability. The relationship between **(B.)** local and **(C.)** distal heritability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

121 heritable, trait-relevant transcripts to be highly correlated and organized according to coherent, biological  
 122 processes representing the mediating endophenotypes driving clinical trait variation. To identify these  
 123 endophenotypes in a theoretically principled way, we developed a novel dimension-reduction technique,  
 124 high-dimension mediation analysis (HDMA), that uses the theory of causal graphical models to identify a  
 125 transcriptomic signature that is simultaneously 1) highly heritable, 2) strongly correlated to the measured  
 126 phenotypes, and 3) conforms to the causal mediation hypothesis (Fig. 3). HDMA projects the high-dimensional  
 127 scores—a composite genome score ( $G_C$ ), a composite transcriptome score ( $T_C$ ), and a composite phenome score  
 128 ( $P_C$ )—and uses the univariate theory of mediation to constrain these projections to satisfy the hypotheses of

129 perfect mediation, namely that upon controlling for the transcriptomic score, the genome score is uncorrelated  
130 to the phenotype score. Formally, perfect mediation implies a constraint on the correlation coefficients among  
131 scores as

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

132 which is equivalent to the partial correlation of  $G_C$  and  $P_C$  after controlling for  $T_C$  being zero. The value  
133  $\text{Corr}(G_C, T_C)\text{Corr}(T_C, P_C)$  is called the path coefficient of the mediation model. The projections of the  
134 high-dimensional data matrices in HDMA are designed to satisfy this constraint, and thus conform to the  
135 perfect mediation hypothesis, as closely as possible. We stress, however, that validating any causal assertion  
136 requires direct experimentation and, thus, that the output of HDMA are scores that are consistent with  
137 causal mediation. Thus, HDMA is a strategy for causal hypothesis generation, where the causal mediator is a  
138 complex endophenotype learned from a high-dimensional readout.

139 Operationally, HDMA is closely related to generalized canonical correlation analysis (CCA), for which  
140 provably convergent algorithms have recently been developed<sup>31</sup>. A complete mathematical derivation and  
141 implementation details for HDMA are available in Supp. Methods.

142 Using HDMA we identified the major axis of variation in the transcriptome that was consistent with mediating  
143 the effects of the genome on metabolic traits (Fig 3). Fig. 3A shows the partial correlations ( $\rho$ ) between  
144 the pairs of these composite vectors. The partial correlation between  $G_C$  and  $T_C$  was 0.42, and the partial  
145 correlation between  $T_C$  and  $P_C$  was 0.78. However, when the transcriptome was taken into account, the  
146 partial correlation between  $G_C$  and  $P_C$  was effectively zero (0.039).  $P_C$  captured 30% of the overall trait  
147 variance, and its estimated heritability was  $0.71 \pm 0.084$ , which was higher than any of the measured traits  
148 (Fig. 1F). Thus, HDMA identified a maximally heritable metabolic composite trait and a highly heritable  
149 component of the transcriptome that are correlated as expected in the perfectly mediated model.

150 As discussed in Supp. Methods, HDMA is related to a generalized form of CCA. Standard CCA is prone to  
151 over-fitting because in any two large matrices it can be trivial to identify highly correlated composite vectors<sup>32</sup>.  
152 To assess whether our implementation of HDMA was similarly prone to over-fitting in a high-dimensional  
153 space, we performed permutation testing. We permuted the individual labels on the transcriptome matrix 1000  
154 times and recalculated the path coefficient, which is the partial correlation of  $G_C$  and  $T_C$  multiplied by the  
155 partial correlation of  $T_C$  and  $P_C$ . This represents the strength of the path from  $G_C$  to  $P_C$  that is putatively  
156 mediated through  $T_C$ . The null distribution of the path coefficient is shown in Fig. 3B, and the observed path  
157 coefficient from the original data is indicated by a red line. The observed path coefficient was well outside the

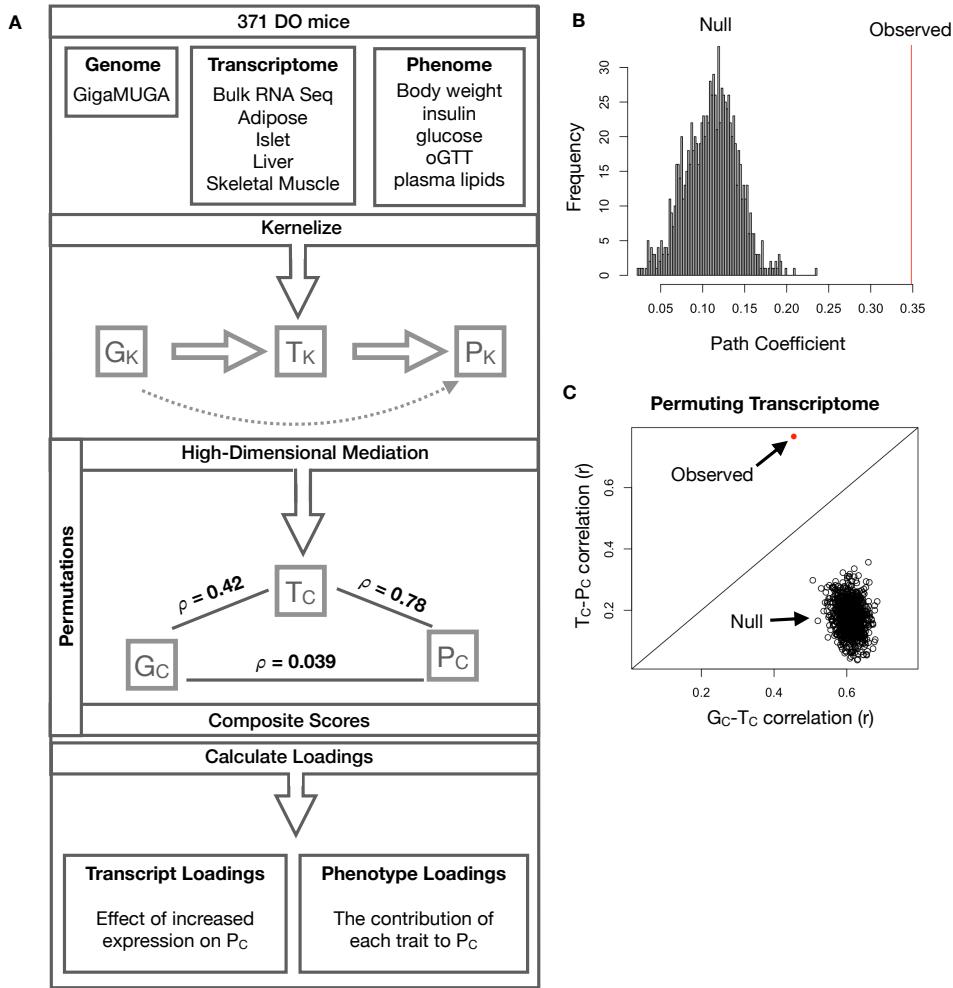


Figure 3: High-dimensional mediation. **A.** Workflow indicating major steps of high-dimensional mediation. The genotype, transcriptome, and phenotype matrices were independently normalized and converted to kernel matrices representing the pairwise relationships between individuals for each data modality ( $K_G$  = 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).

158 null distribution generated by permutations ( $p < 10^{-16}$ ). Fig. 3C illustrates this observation in more detail.  
 159 Although we identified high correlations between  $G_c$  and  $T_c$ , and modest correlations between  $T_c$  and  $P_c$  in  
 160 the null data (Fig 3C), these two values could not be maximized simultaneously in the null data. In contrast,  
 161 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  
 162 maximized simultaneously suggesting that the path from genotype to phenotype through transcriptome is

163 highly non-trivial and identifiable in this case. These results suggest that these composite vectors represent  
164 genetically determined variation in phenotype that is mediated through genetically determined variation in  
165 transcription.

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

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

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

181 We interpreted large loadings onto transcripts as indicating strong mediation of the effect of genetics on MDI.  
182 Large positive loadings indicate that higher expression was associated with a higher MDI (i.e. higher risk of  
183 obesity and metabolic disease on the HFHS diet) (Fig. 4C). Conversely, large negative loadings indicate that  
184 high expression of these transcripts was associated with a lower MDI (i.e. lower risk of obesity and metabolic  
185 disease on the HFHS diet) (Fig. 4C). We used gene set enrichment analysis (GSEA)<sup>33;34</sup> to look for biological  
186 processes and pathways that were enriched at the top and bottom of this list (Methods).

187 In adipose tissue, both GO processes and KEGG pathway enrichments pointed to an axis of inflammation  
188 and metabolism (Figs. S3 and S4). GO terms and KEGG pathways associated with inflammation were  
189 positively associated with MDI, indicating that increased expression in inflammatory pathways was associated  
190 with a higher MDI. It is well established that adipose tissue in obese individuals is inflamed and infiltrated  
191 by macrophages<sup>35-39</sup>, and the results here suggest that this may be a dominant heritable component of  
192 metabolic disease.

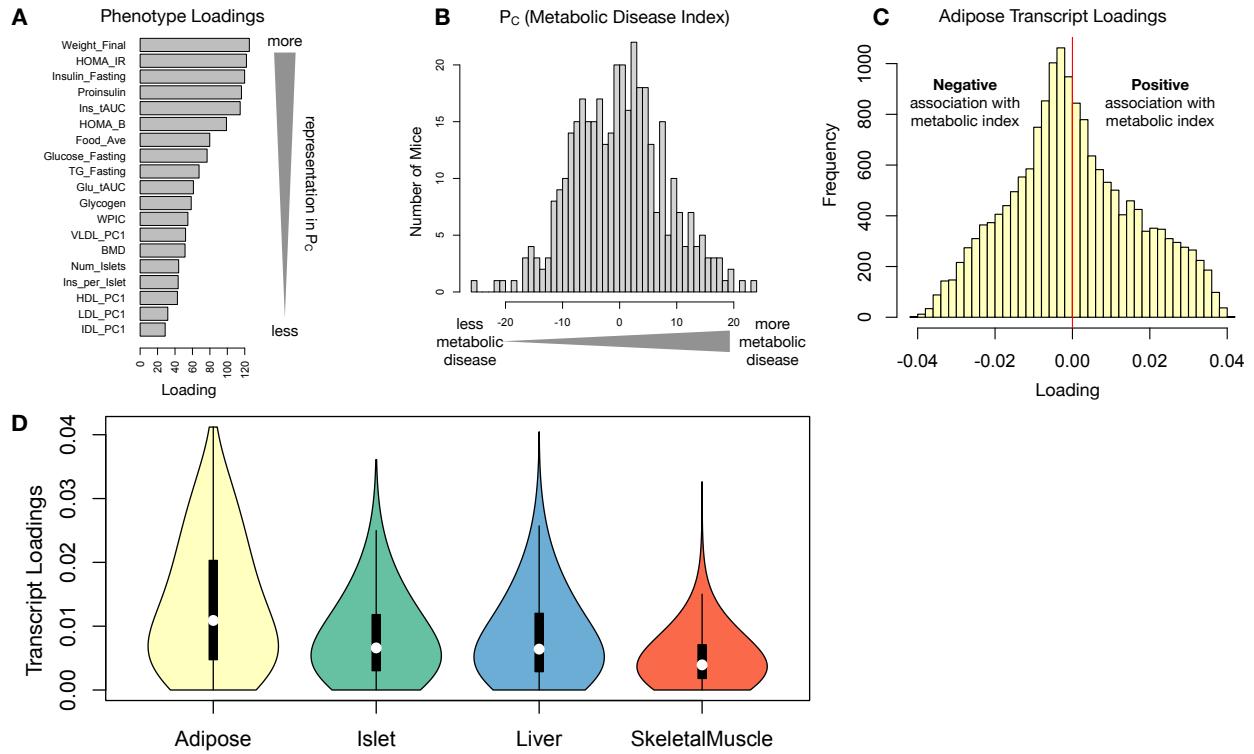


Figure 4: Interpretation of loadings. **A.** Loadings across traits. Body weight and insulin resistance contributed the most to the composite trait. **B.** Phenotype scores across individuals. Individuals with large positive phenotype scores had higher body weight and insulin resistance than average. Individuals with large negative phenotype scores had lower body weight and insulin resistance than average. **C.** Distribution of transcript loadings in adipose tissue. For transcripts with large positive loadings, higher expression was associated with higher phenotype scores. For transcripts with large negative loadings, higher expression was associated with lower phenotype scores. **D.** Distribution of absolute value of transcript loadings across tissues. Transcripts in adipose tissue had the largest loadings indicating that adipose tissue gene expression was a strong mediator of genotype on body weight and insulin resistance.

193 The strongest negative enrichments in adipose tissue were related to mitochondrial activity in general, and  
 194 thermogenesis in particular (Figs. S3 and S3). Genes in the KEGG oxidative phosphorylation pathway in  
 195 mice were almost universally negatively loaded in adipose tissue, suggesting that increased expression of  
 196 these genes was associated with reduced MDI (Supp. Fig. S5). Consistent with this observations, it has been  
 197 shown previously that mouse strains with greater thermogenic potential are also less susceptible to obesity  
 198 on a HFHS diet<sup>40</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. S3, S6 and S7). Expression of genes in both pathways (for which there is some overlap)  
 202 has been previously associated with insulin sensitivity<sup>12;41;42</sup>, suggesting that heritable variation in regulation  
 203 of these pathways may influence risk of insulin resistance.

204 Looking at 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 times 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 four  
224 tissues potentially suggesting system-wide proliferation of immune cells. Otherwise, all clusters were strongly  
225 loaded in only one or two tissues. For example, the lipid metabolism cluster was loaded most heavily in liver.  
226 The 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>43</sup>. Agonists of *Pparg*, such as thiazolidinediones, are  
229 FDA-approved to treat type II diabetes, and reduce inflammation and adipose hypertrophy<sup>43</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>44;45</sup>.  
234 Overexpression of *Pparg* in the livers of mice with a *Ppara* knockout, causes upregulation of genes involved in

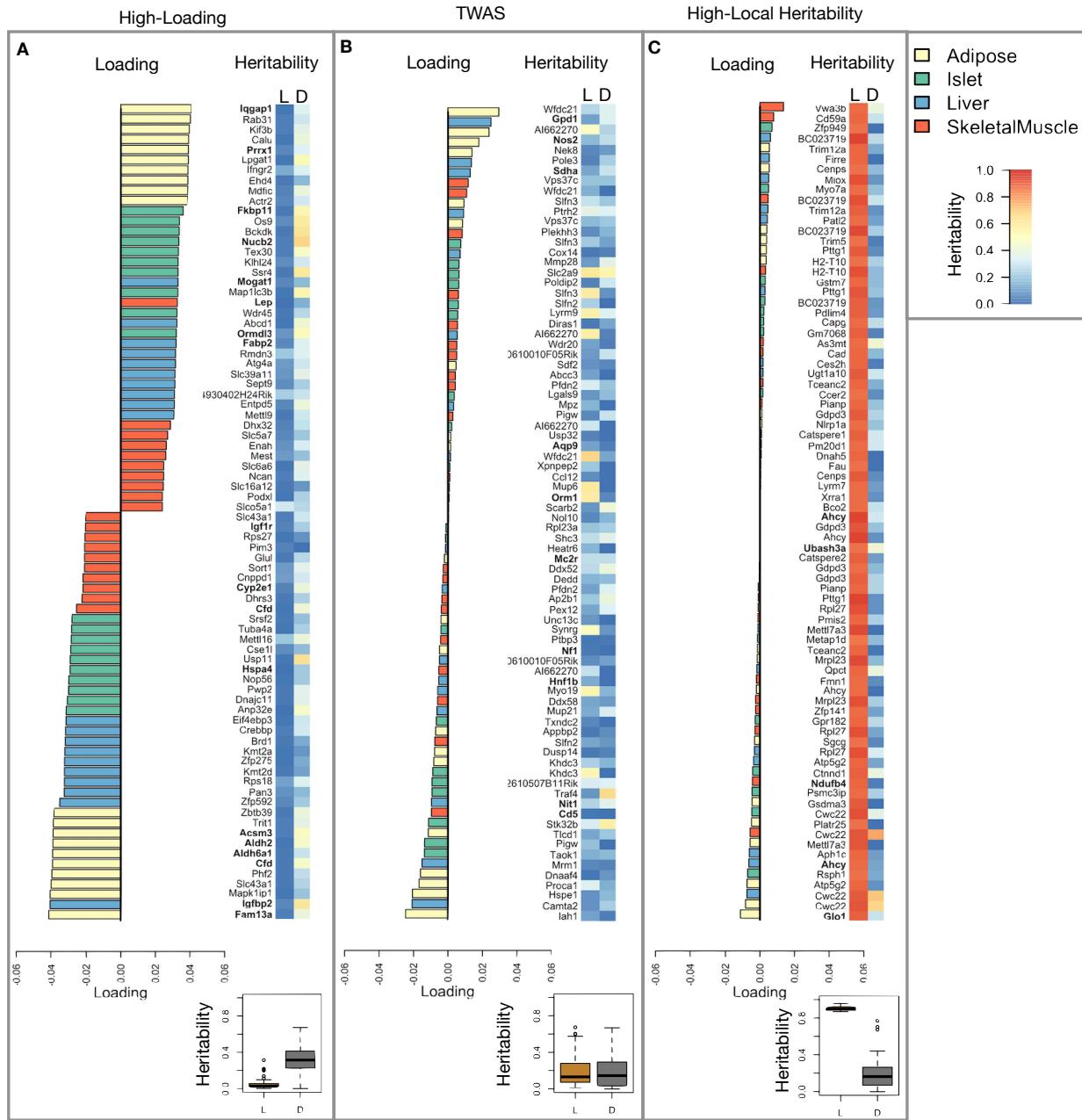


Figure 5: Transcripts with high loadings have high distal heritability and literature support. Each panel has a bar plot showing the loadings of transcripts selected by different criteria. Bar color indicates the tissue of origin. The heat map shows the local (L - left) and distal (D - right) heritability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

235 adipogenesis<sup>46</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>47;48</sup>.

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 MDI. These results highlight  
 242 the importance of tissue context when investigating the role of heritable transcript variability in driving  
 243 phenotype.

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

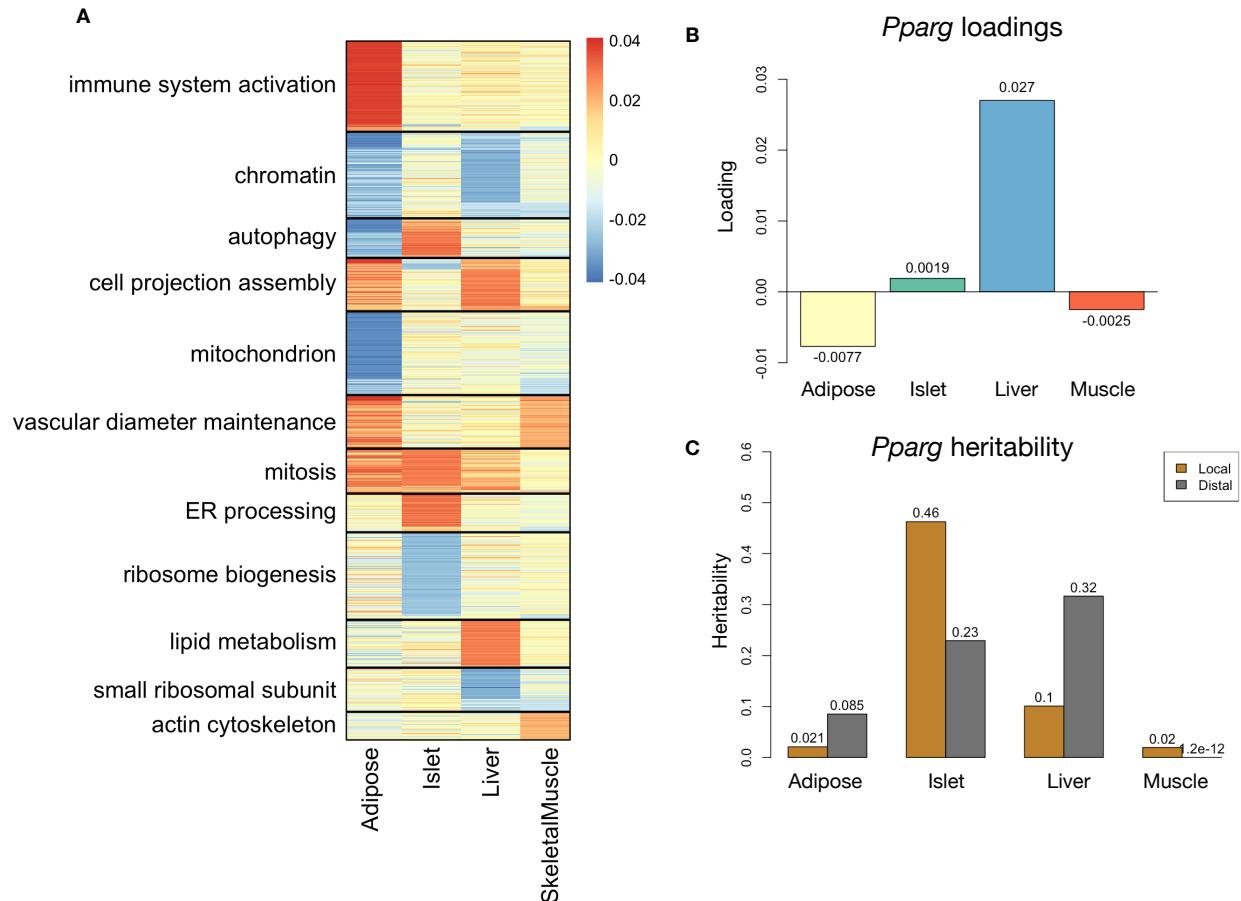


Figure 6: Tissue-specific transcriptional programs were associated with obesity and insulin resistance. **A** Heat map showing the loadings of all transcripts with loadings greater than 2.5 standard deviations from the mean in any tissue. The heat map was clustered using k medoid clustering. Functional enrichments of each cluster are indicated along the left margin. **B** Loadings for *Pparg* in different tissues. **C** Local and distal of *Pparg* expression in different tissues.

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

246 To test whether the transcript loadings identified in the DO could be translated to another population, we  
 247 tested whether they could predict metabolic phenotype in an independent population of CC-RIX mice, which  
 248 were F1 mice derived from multiple pairings of Collaborative Cross (CC)<sup>49–52</sup> strains (Fig. 7) (Methods).  
 249 We tested two questions. First, we asked whether the loadings identified in the DO mice were relevant to  
 250 the relationship between the transcriptome and the phenotype in the CC-RIX. We predicted body weight  
 251 (a surrogate for MDI) in each CC-RIX individual using measured gene expression in each tissue and the  
 252 transcript loadings identified in the DO (Methods). The predicted body weight and actual body weight were  
 253 highly correlated in all tissues (Fig. 7B left column). The best prediction was achieved for adipose tissue,  
 254 which supports the observation in the DO that adipose expression was the strongest mediator of the genetic  
 255 effect on MDI. This result also confirms the validity and translatability of the transcript loadings and their  
 256 relationship to metabolic disease.

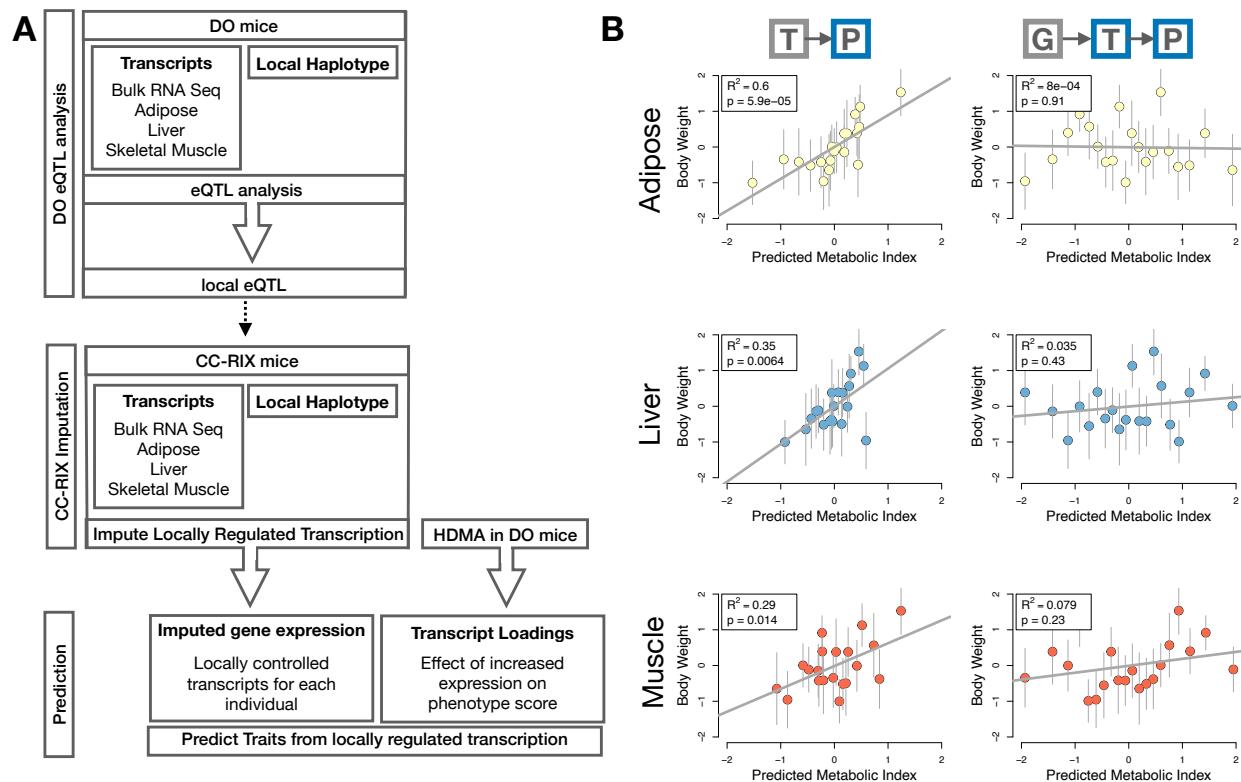


Figure 7: Transcription, but not local genotype, predicts phenotype in the CC-RIX. **A.** Workflow showing procedure for translating HDMA results to an independent population of mice. **B.** Relationships between the predicted metabolic disease index (MDI) and measured body weight. The left column shows the predictions using measured transcripts. The right column shows the prediction using transcript levels imputed from local genotype. Gray boxes indicate measured quantities, and blue boxes indicate calculated quantities. The dots in each panel represent individual CC-RIX strains. The gray lines show the standard deviation on body weight for the strain.

257 The second question related to the source of the relevant variation in gene expression. If local regulation was  
258 the predominant factor influencing gene expression, we should be able to predict phenotype in the CC-RIX  
259 using transcripts imputed from local genotype (Fig. 7A). The DO and the CC-RIX were derived from the  
260 same eight founder strains and so carry the same alleles throughout the genome. We imputed gene expression  
261 in the CC-RIX using local genotype and were able to estimate variation in gene transcription robustly (Supp.  
262 Fig. S8). 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 disease<sup>53</sup>. 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 HFHS diet. Loading for marker genes for other cell types were not statistically different from  
281 zero, indicating that changes in the abundance of those cell types is not a mediator of MDI.

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

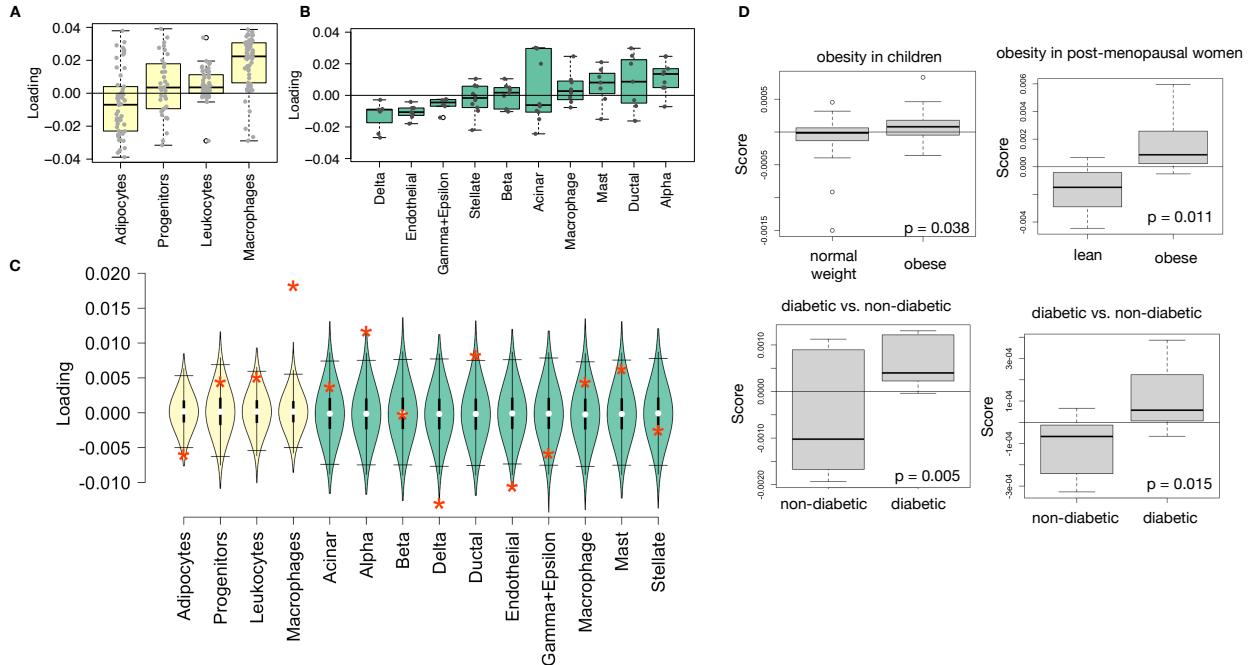


Figure 8: HDMA results translate to humans. **A.** Distribution of loadings for cell-type-specific transcripts in adipose tissue. **B.** Distribution of loadings for cell-type-specific transcripts in pancreatic islets (green). **C.** Null distributions for the mean loading of randomly selected transcripts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

288 MDI.

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

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

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

300 We calculated a predicted obesity score for each individual in the human studies based on their adipose

301 tissue gene expression (Methods) and compared the predicted scores for obese and non-obese groups as well  
302 as diabetic and non-diabetic groups. In all cases, the predicted obesity scores were higher on average for  
303 individuals in the obese and diabetic groups compared with the lean and non-diabetic groups (Fig. 8D).  
304 This indicates that the distally heritable signature of MDI identified in DO mice is relevant to obesity and  
305 diabetes in human subjects.

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

307 Another potential application of the transcript loading landscape is in ranking potential drug candidates  
308 for the treatment of metabolic disease. Although high-loading transcripts may be good candidates for  
309 understanding specific biology related to obesity, the transcriptome overall is highly interconnected and  
310 redundant. The ConnectivityMap (CMAP) database<sup>54</sup> developed by the Broad Institute allows querying  
311 thousands of compounds that reverse or enhance the extreme ends of transcriptomic signatures in multiple  
312 different cell types. By identifying drugs that reverse pathogenic transcriptomic signatures, we can potentially  
313 identify compounds that have favorable effects on gene expression.

314 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  
315 (Supp. Figs S9 and S10). To get more tissue-specific results, we also looked at top results in cell types that  
316 most closely resembled our tissues. We looked at results in adipocytes (ASC) as well as pancreatic tumor  
317 cells (YAPC) regardless of *p* value (Supp. Figs S11 and S12).

318 Looking across all cell types, the notable top hits from the adipose tissue loadings included mTOR inhibitors  
319 and glucocorticoid agonists (Supp. Fig. S9). It is thought that metformin, which is commonly used to  
320 improve glycemic control, acts, at least in part, by inhibiting mTOR signaling<sup>55;56</sup>. However, long-term use  
321 of other mTOR inhibitors, such as rapamycin, are known to cause insulin resistance and  $\beta$ -cell toxicity<sup>56–58</sup>.  
322 Glucocorticoids are used to reduce inflammation, which was a prominent signature in the adipose tissues,  
323 but these drugs also promote hyperglycemia and diabetes<sup>59;60</sup>. Accute treatment with glucocorticoids has  
324 further been shown to reduce thermogenesis in rodent adipocytes<sup>61–63</sup>, but increase thermogenesis in human  
325 adipocytes<sup>64;65</sup>. Thus, the pathways identified by CMAP across all cell types were highly related to the  
326 transcript loading profiles, but the relationship was not a simple reversal.

327 The top hit for the adipose composite transcript in CMAP adipocytes was a PARP inhibitor (Supp. Fig.  
328 S11). PARPs play a role in lipid metabolism and are involved in the development of obesity and diabetes<sup>66</sup>.  
329 PARP1 inhibition increases mitochondrial biogenesis<sup>67</sup>. Inhibition of PARP1 activity can further prevent  
330 necrosis in favor of the less inflammatory apoptosis<sup>68</sup>, thereby potentially reducing inflammation in stressed

332 adipocytes. Other notable hits among the top 20 were BTK inhibitors, which have been observed to suppress  
333 inflammation and improve insulin resistance<sup>69</sup> as well as to reduce insulin antibodies in type I diabetes<sup>70</sup>.  
334 IkappaB kinase (IKK) is an enzyme complex involved in regulating cellular responses to inflammation<sup>71</sup>.  
335 Inhibitors of IKK have been shown to improve glucose control in type II diabetes<sup>72;73</sup>.

336 Among the top most significant hits for the transcript loadings from pancreatic islets (Supp. Fig. S10),  
337 was suppression of T cell receptor signaling, which is known to be involved in Type 1 diabetes<sup>74</sup>, as well as  
338 TNFR1, which has been associated with mortality in diabetes patients<sup>75</sup>. Suppression of NOD1/2 signaling  
339 was also among the top hits. NOD1 and 2 sense ER stress<sup>76;77</sup>, which is associated with  $\beta$ -cell death in type  
340 1 and type 2 diabetes<sup>78</sup>. This cell death process is dependent on NOD1/2 signaling<sup>76</sup>, 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. S12). Hits in this list included widely used  
344 diabetes drugs, such as sulfonylureas, PPAR receptor agonists, and insulin sensitizers. Rosiglitazone is a  
345 PPAR- $\gamma$  agonist and was one of the most prescribed drugs for type 2 diabetes before its use was reduced due  
346 to cardiac side-effects<sup>79</sup>. Sulfonylureas are another commonly prescribed drug class for type 2 diabetes, but  
347 also have notable side effects including hypoglycemia and accelerated  $\beta$ -cell death<sup>80</sup>.

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

## 353 Discussion

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

362 expression regulation in relation to trait heritability and suggest that heritable trait variation is mediated  
363 primarily through distal gene regulation.

## 364 Supplemental Discussion

365 Our result that distal regulation accounted for most trait-related gene expression differences is consistent  
366 with a complex model of genetic trait determination. It has frequently been assumed that gene regulation in  
367 *cis* is the primary driver of genetically associated trait variation, but attempts to use local gene regulation  
368 to explain phenotypic variation have had limited success<sup>16;17</sup>. In recent years, evidence has mounted that  
369 distal gene regulation may be an important mediator of trait heritability<sup>20;18;81</sup>. It has been observed that  
370 transcripts with high local heritability explain less expression-mediated disease heritability than those with  
371 low local heritability<sup>20</sup>. Consistent with this observation, genes located near GWAS hits tend to be complexly  
372 regulated<sup>18</sup>. They also tend to be enriched with functional annotations, in contrast to genes with simple  
373 local regulation, which tend to be depleted of functional annotations suggesting they are less likely to be  
374 directly involved in disease traits<sup>18</sup>. These observations are consistent with principles of robustness in complex  
375 systems in which simple regulation of important elements leads to fragility of the system<sup>82–84</sup>. Our results  
376 are consistent, instead, with a more complex picture where genes whose expression can drive trait variation  
377 are buffered from local genetic variation but are extensively influenced indirectly by genetic variation in the  
378 regulatory networks converging on those genes.

379 Our results are consistent with the recently proposed omnigenic model, which posits that complex traits are  
380 massively polygenic and that their heritability is spread out across the genome<sup>85</sup>. In the omnigenic model,  
381 genes are classified either as “core genes,” which directly impinge on the trait, or “peripheral genes,” which  
382 are not directly trait-related, but influence core genes through the complex gene regulatory network. HDMA  
383 explicitly models a central proposal of the omnigenic model which posits that once the expression of the  
384 core genes (i.e. trait-mediating genes) is accounted for, there should be no residual correlation between the  
385 genome and the phenotype. Here, when the composite transcript was taken into account there was no residual  
386 correlation between the composite genome and composite phenotype (Fig. 3A).

387 Unlike in the omnigenic model, we did not observe a clear demarcation between the core and peripheral  
388 genes in loading magnitude, but we do not necessarily expect a clear separation given the complexity of gene  
389 regulation and the genotype-phenotype map<sup>86</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>81</sup>. This

392 is consistent with what we observed here. For example, *Nucb2*, had a high loading in islets and was also  
393 strongly distally regulated (66% distal heritability) (Fig. 5). Although its transcription was highly heritable  
394 in islets, that regulation was distributed across the genome, with no clear distal eQTL (Supp. Fig. S13).  
395 Thus, although distal regulation of some genes may be strong, this regulation is likely to be highly complex  
396 and not easily localized.

397 Individual high-loading transcripts also demonstrated biologically interpretable, tissue-specific patterns. We  
398 highlighted *Ppary*, which is known to be protective in adipose tissue<sup>43</sup> where it was negatively loaded, and  
399 harmful in the liver<sup>44–48</sup>, where it was positively loaded. Such granular patterns may be useful in generating  
400 hypotheses for further testing, and prioritizing genes as therapeutic targets. The tissue-specific nature of  
401 the loadings also may provide clues to tissue-specific effects, or side effects, of targeting particular genes  
402 system-wide.

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

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

## 415 Data and Code Availability

416 **DO mice:** Genotypes, phenotypes, and pancreatic islet gene expression data were previously published<sup>12</sup>.  
417 Gene expression for the other tissues can be found at the Gene Expression Omnibus <https://www.ncbi.nlm.nih.gov/geo/> with the following accession numbers: DO adipose tissue - GSE266549; DO liver tissue  
418 - GSE266569; DO skeletal muscle - GSE266567. Expression data with calculated eQTLs are available at  
419 Figshare <https://figshare.com/> DOI: 10.6084/m9.figshare.27066979

421 **CC-RIX mice:** Gene expression can be found at the Gene Expression Omnibus <https://www.ncbi.nlm.nih.gov/geo/>

<sup>422</sup> .gov/geo/ with the following accession numbers: CC-RIX adipose tissue - GSE237737; CC-RIX liver tissue -  
<sup>423</sup> GSE237743; CC-RIX skeletal muscle - GSE237747. Count matrices and phenotype data can be found at  
<sup>424</sup> Figshare <https://figshare.com/> DOI: 10.6084/m9.figshare.27066979

<sup>425</sup> **Code:** All code used to run the analyses reported here are available at Figshare: <https://figshare.com/> DOI:  
<sup>426</sup> 10.6084/m9.figshare.27066979

## <sup>427</sup> Acknowledgements

<sup>428</sup> This project was supported by The Jackson Laboratory Cube Initiative, as well as grants from the National  
<sup>429</sup> Institutes of Health (grant numbers.: R01DK101573, R01DK102948, and RC2DK125961) (to A. D. A.) and  
<sup>430</sup> by the University of Wisconsin-Madison, Department of Biochemistry and Office of the Vice Chancellor for  
<sup>431</sup> Research and Graduate Education with funding from the Wisconsin Alumni Research Foundation (to M. P.  
<sup>432</sup> K.).

<sup>433</sup> We thank the following scientific services at The Jackson Laboratory: Genome Technologies for the RNA  
<sup>434</sup> sequencing, necropsy services for the tissue harvests, and the Center for Biometric Analysis for metabolic  
<sup>435</sup> phenotyping.

436 **Supplemental Figures**

**Different ways to view the trait matrix**

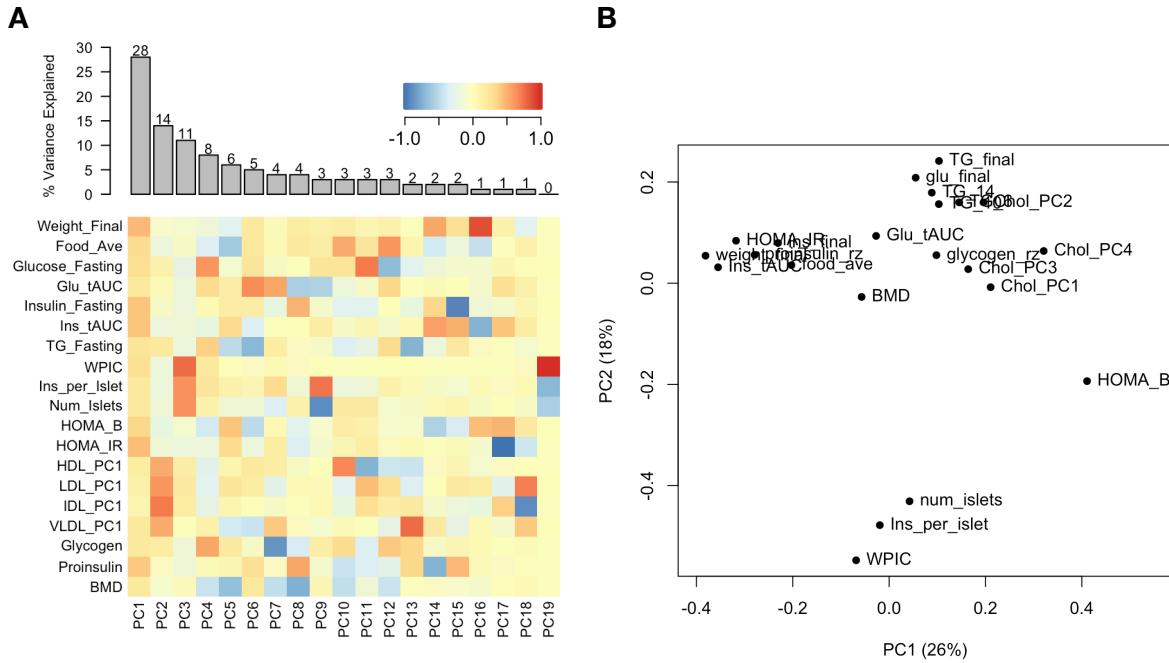


Figure S1: Trait matrix decomposition. **A** Loadings on principal components. The heat map shows the loadings of each trait onto each principal component of the trait matrix. The bars at the top show the percent variance explained for each principal component. **B** Traits plotted by the first and second principal components of the trait matrix. This view shows clustering of traits into insulin- and weight-related traits, lipid-related traits, and ex-vivo pancreatic measurements.

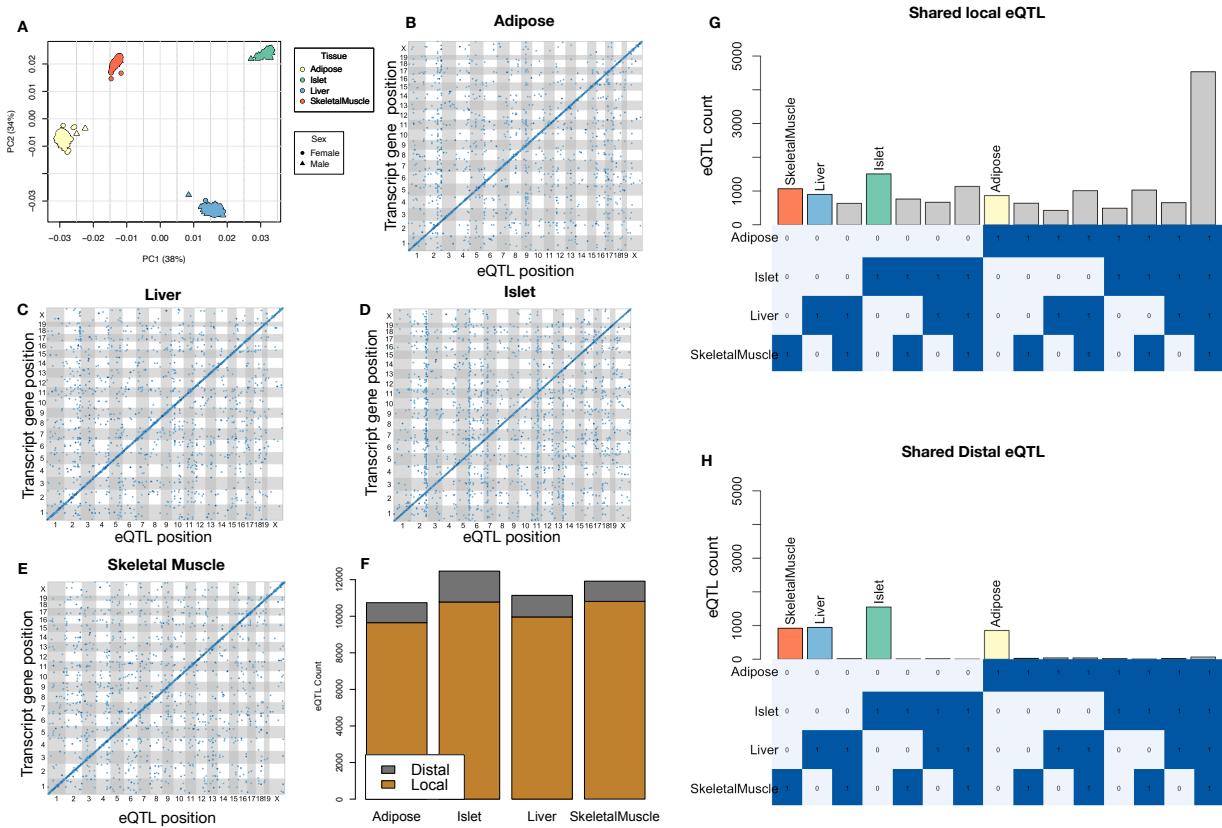


Figure S2: Overview of eQTL analysis in DO mice. **A.** RNA seq samples from the four different tissues clustered by tissue. **B.-E.** eQTL maps are shown for each tissue. The *x*-axis shows the position of the mapped eQTL, and the *y*-axis shows the physical position of the gene encoding each mapped 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

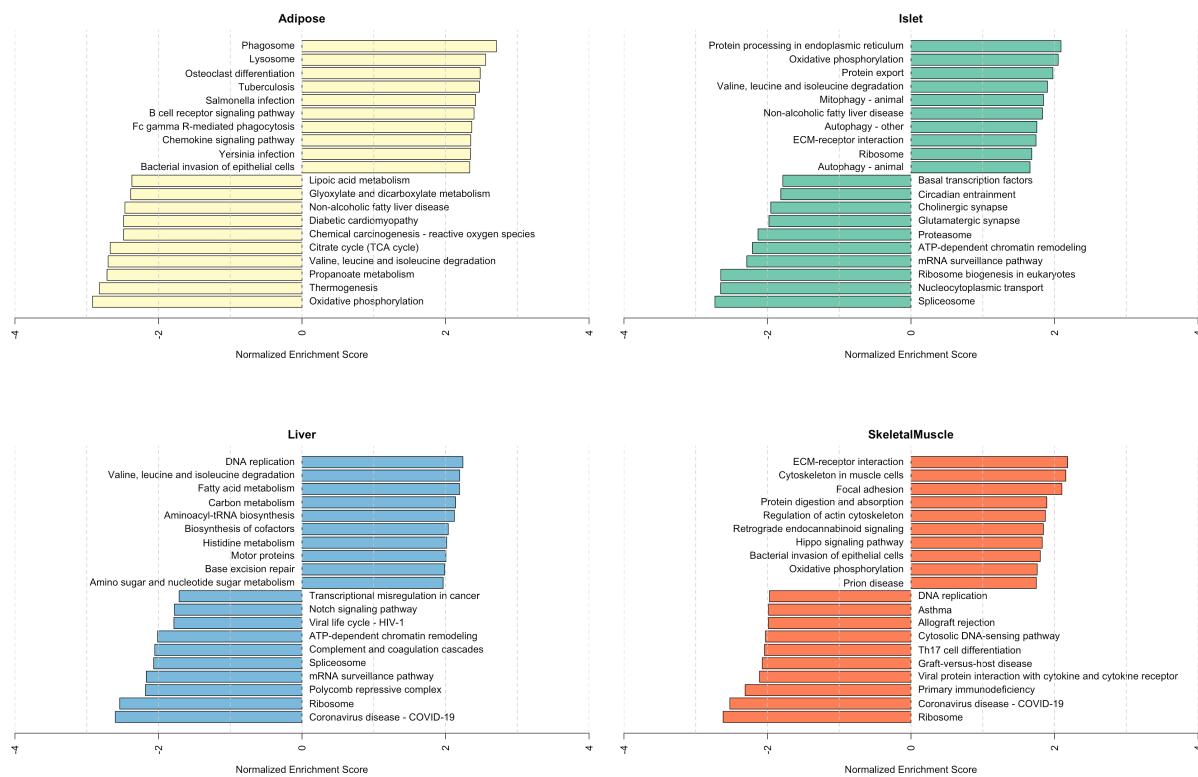


Figure S3: Bar plots showing normalized enrichment scores (NES) for KEGG pathways as determined by fast gene score enrichment analysis (fgsea). Only the top 10 positive and top 10 negative scores are shown. Colors indicate tissue. The name beside each bar shows the name of each enriched KEGG pathway.

## Top GO term enrichments by GSEA

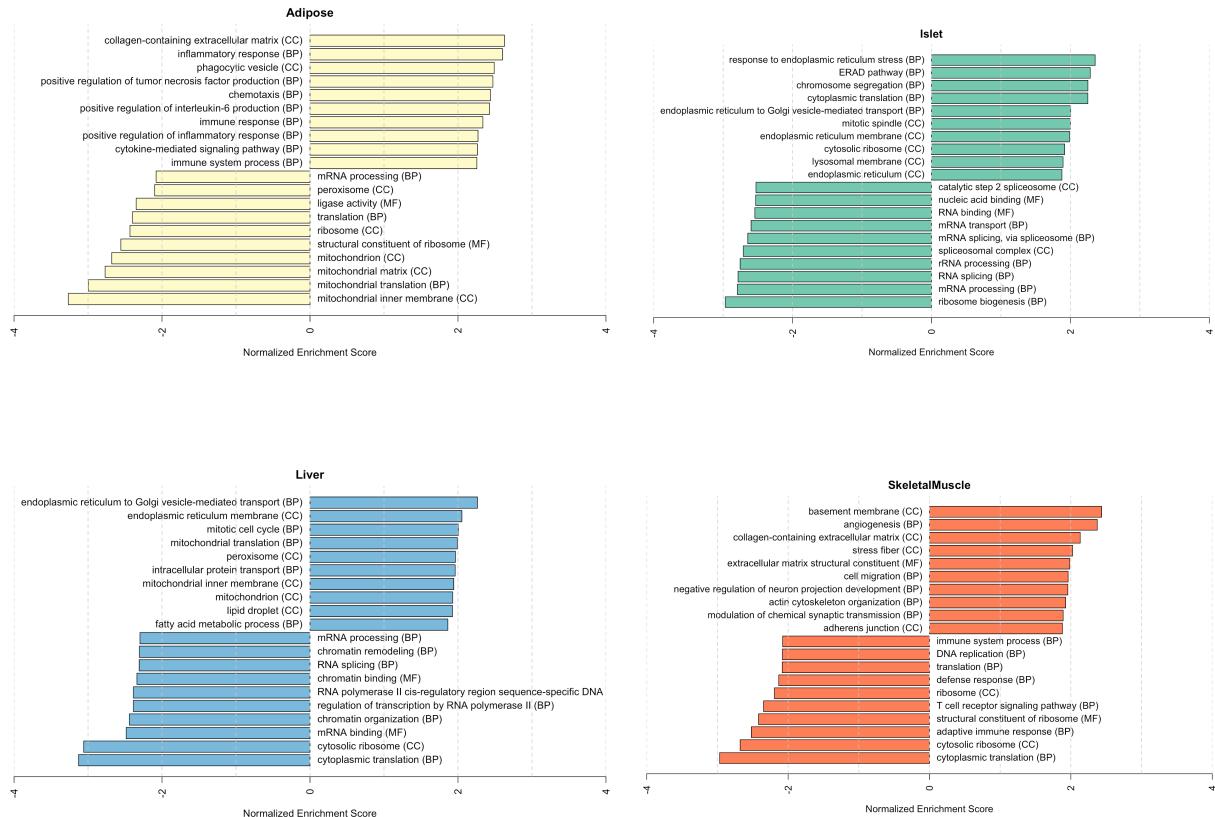


Figure S4: Bar plots showing normalized enrichment scores (NES) for GO terms as determined by fast gene score enrichment analysis (fgsea). Only the top 10 positive and top 10 negative scores are shown. Colors indicate tissue. The name beside each bar shows the name of each enriched GO term. The letters in parentheses indicate whether the term is from the biological process ontology (BP), the molecular function ontology (MF), or the cellular compartment ontology (CC).

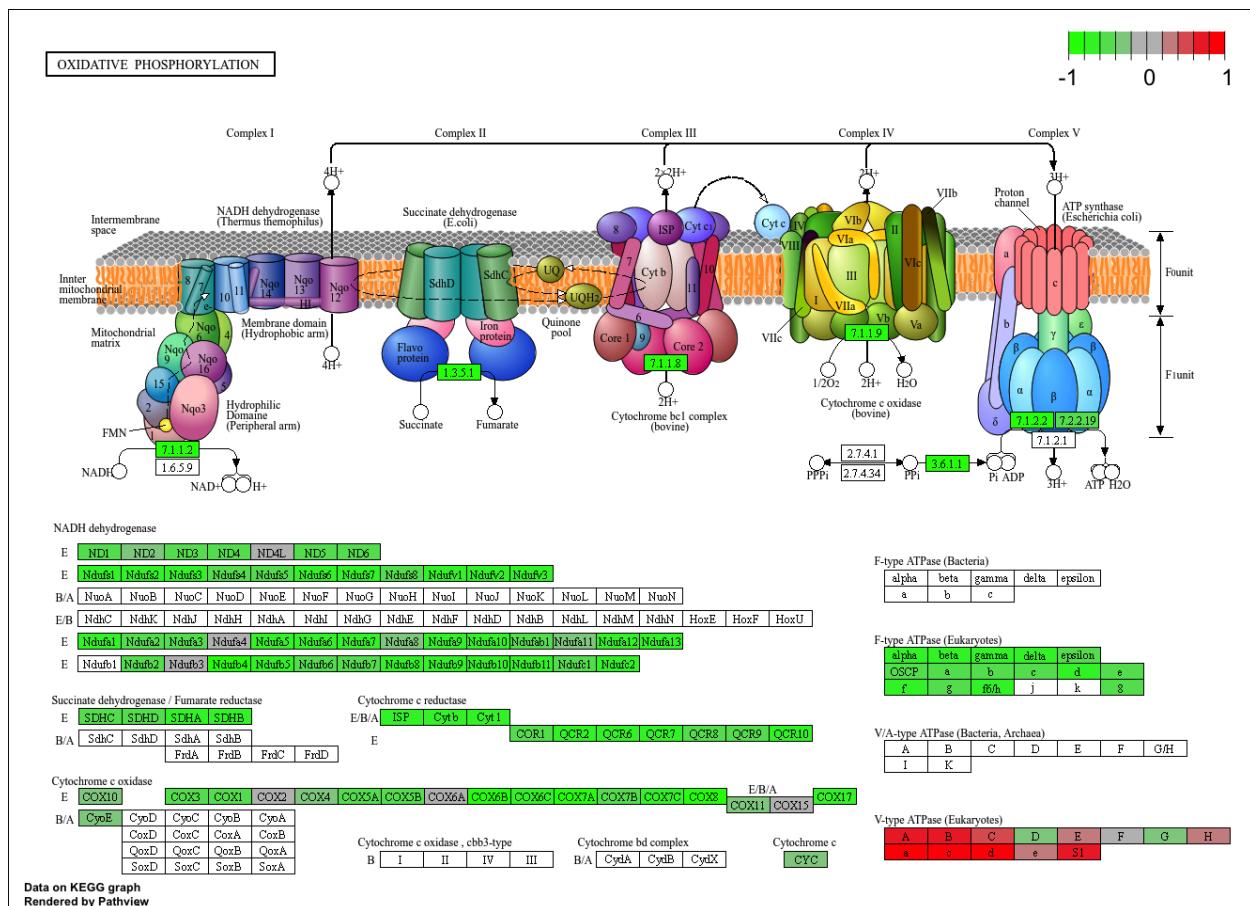


Figure S5: The KEGG pathway for oxidative phosphorylation in mice. Each element is colored based on its HDMA loading from adipose tissue normalized to run from -1 to 1. Genes highlighted in green had negative loadings, and those highlighted in red had positive loadings. Almost the entire pathway was strongly negatively loaded indicating that increased expression of genes involved in oxidative phosphorylation was associated with reduced MDI.

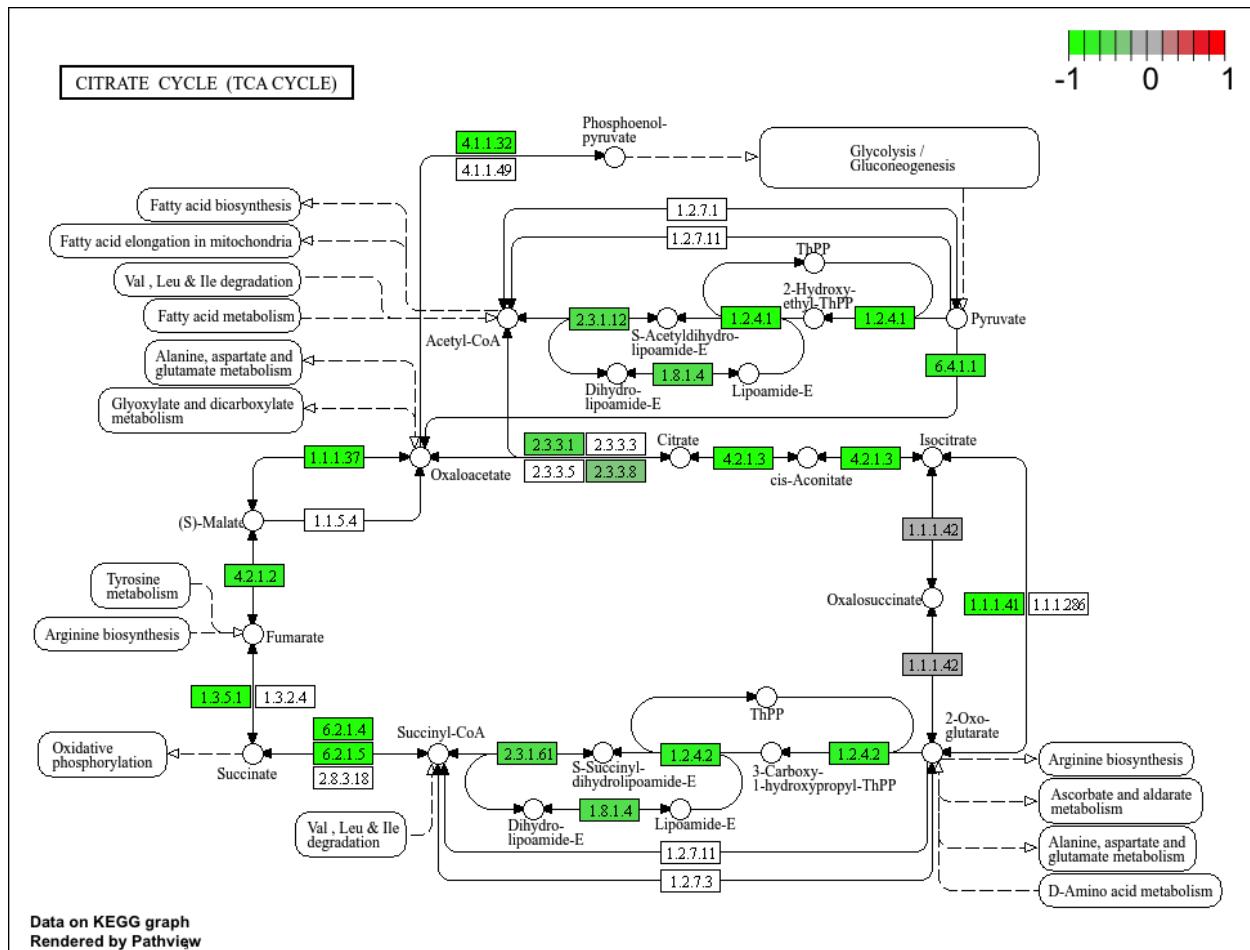


Figure S6: The KEGG pathway for the TCA (citric acid) cycle in mice. Each element is colored based on its HDMA loading from adipose tissue normalized to run from -1 to 1. Genes highlighted in green had negative loadings, and those highlighted in red had positive loadings. Many genes in the cycle were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

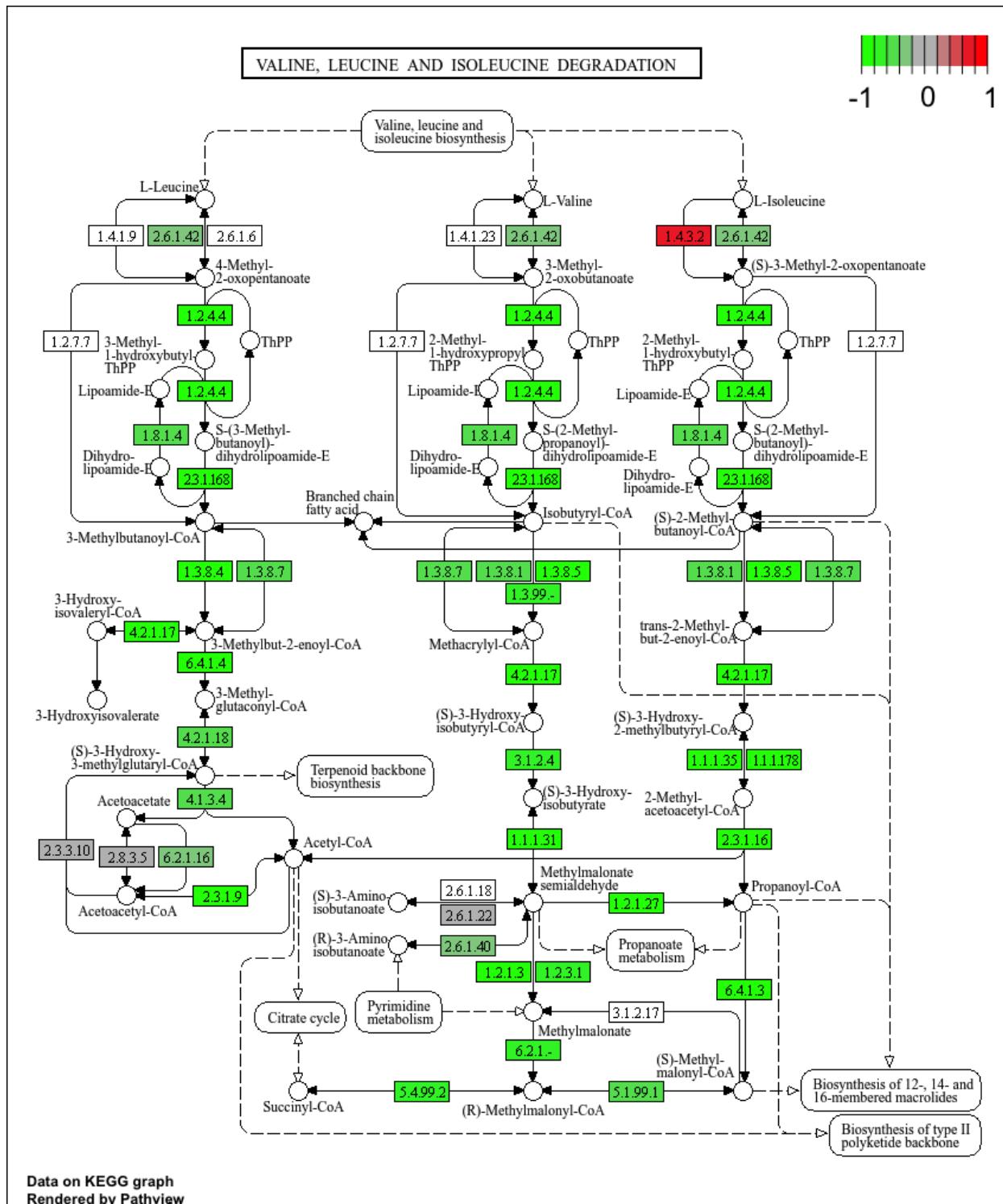


Figure S7: The KEGG pathway for branched-chain amino acid degradation in mice. Each element is colored based on its HDMA loading from adipose tissue normalized to run from -1 to 1. Genes highlighted in green had negative loadings, and those highlighted in red had positive loadings. Almost the entire pathway was strongly negatively loaded indicating that increased expression of genes involved in branched-chain amino acid degradation was associated with reduced MDI.

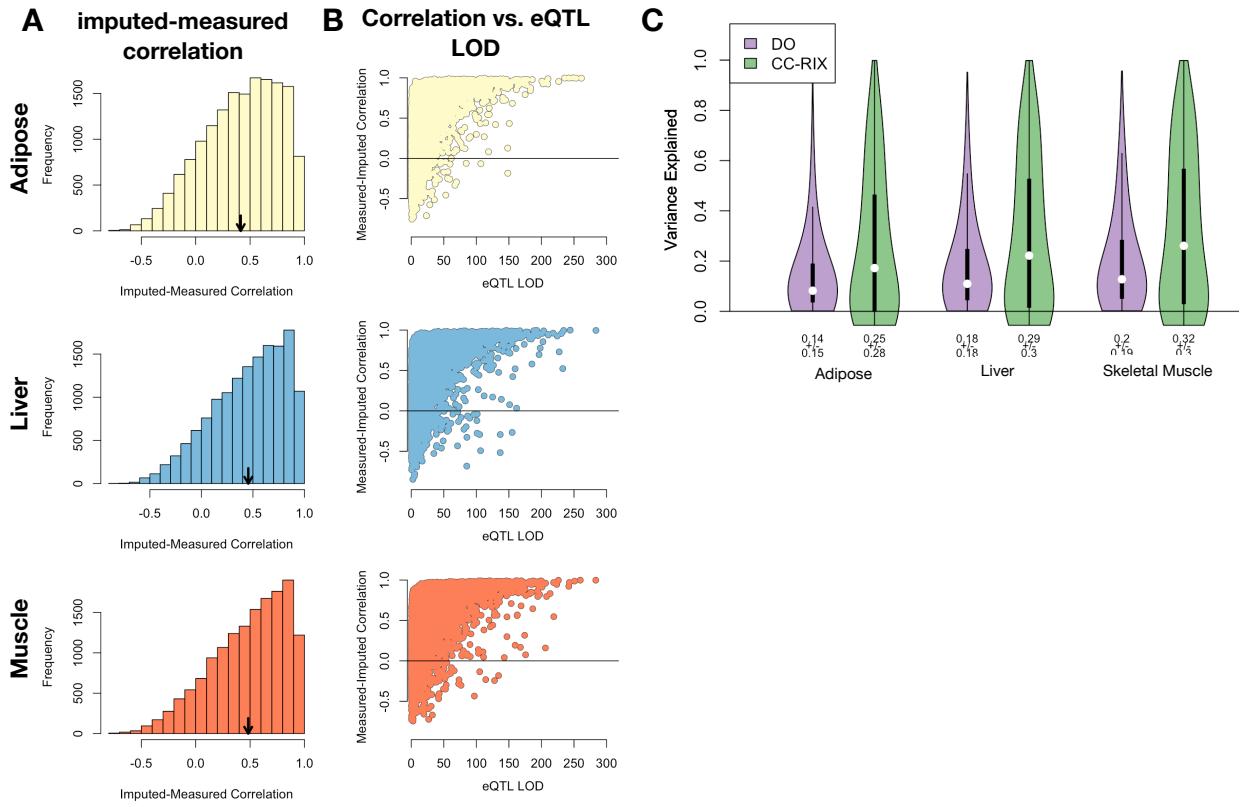


Figure S8: Validation of transcript imputation in the CC-RIX. **A.** Distributions of correlations between imputed and measured transcripts in the CC-RIX. The mean of each distribution is shown by the red line. All distributions were skewed toward positive correlations and had positive means near a Pearson 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.

<b>id</b>	<b>norm_ss</b>	<b>cell_iname</b>	<b>pert_type</b>	<b>raw_ss▲</b>	<b>fdr_q_nlog10</b>	<b>set_type</b>	<b>src_set_id</b>
		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 S9: CMAP results using the *adipose* tissue composite transcript as an input. Table includes results from *all cell types* sorted with a  $-\log_{10}(q) > 15$ . The results are sorted by the correlation of the query to the input with the most negative results at the top.

id	norm_CS	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 S10: CMAP results using the *pancreatic islet* tissue composite transcript as an input. Table includes results from *all cell types* sorted with a  $-\log_{10}(q) > 15$ . The results are sorted by the correlation of the query to the input with the most negative results at the top.

<b>id</b>	<b>norm_ss</b>	<b>cell_iname</b>	<b>pert_type</b>	<b>raw_ss ▲</b>	<b>fdr_q_nlog10</b>	<b>set_type</b>	<b>src_set_id</b>
		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 S11: CMAP results using the *adipose* tissue composite transcript as an input. Table includes the top 30 results derived *only from normal adipocytes* (ASC) regardless of significance. The results are sorted by the correlation of the query to the input with the most negative results at the top.

id	norm_CS	cell_iname	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 S12: CMAP results using the *pancreatic islet* composite transcript as an input. Table includes the top 30 results derived *only from YAPC cells*, which are derived from pancreatic carcinoma cells. Results are shown regardless of significance and are sorted by the correlation of the query to the input with the most negative results at the top.

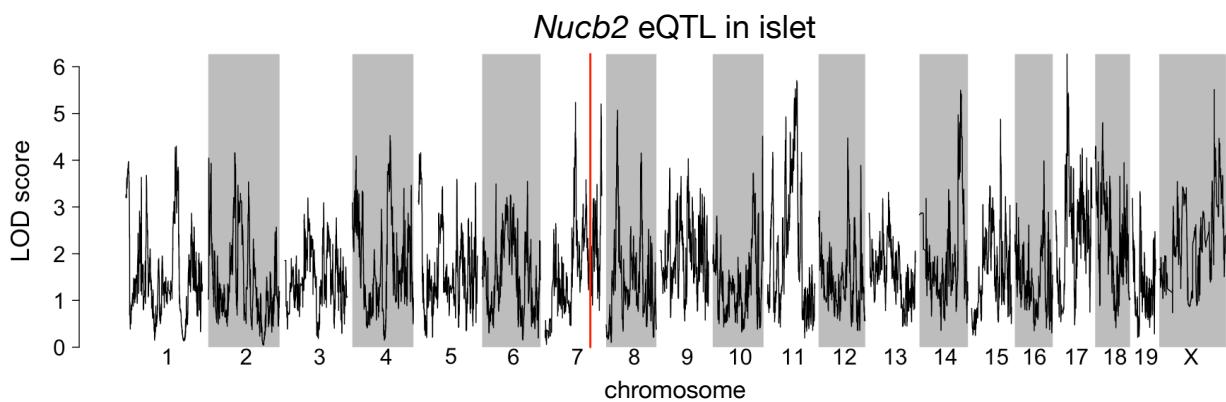


Figure S13: Regulation of *Nucb2* expression in islet. *Nucb2* is encoded on mouse chromosome 7 at 116.5 Mb (red line). In islets the heritability of *Nucb2* expression levels is 69% heritable. This LOD score trace shows that there is no local eQTL at the position of the gene, nor any strong distal eQTL anywhere else in the genome.

437 **References**

- 438 [1] M. T. Maurano, R. Humbert, E. Rynes, R. E. Thurman, E. Haugen, H. Wang, A. P. Reynolds,  
439 R. Sandstrom, H. Qu, J. Brody, A. Shafer, F. Neri, K. Lee, T. Kutyavin, S. Stehling-Sun, A. K.  
440 Johnson, T. K. Canfield, E. Giste, M. Diegel, D. Bates, R. S. Hansen, S. Neph, P. J. Sabo, S. Heimfeld,  
441 A. Raubitschek, S. Ziegler, C. Cotsapas, N. Sotoodehnia, I. Glass, S. R. Sunyaev, R. Kaul, and J. A.  
442 Stamatoyannopoulos. Systematic localization of common disease-associated variation in regulatory DNA.  
443 *Science*, 337(6099):1190–1195, Sep 2012.
- 444 [2] K. K. Farh, A. Marson, J. Zhu, M. Kleinewietfeld, W. J. Housley, S. Beik, N. Shores, H. Whitton, R. J.  
445 Ryan, A. A. Shishkin, M. Hatan, M. J. Carrasco-Alfonso, D. Mayer, C. J. Luckey, N. A. Patsopoulos,  
446 P. L. De Jager, V. K. Kuchroo, C. B. Epstein, M. J. Daly, D. A. Hafler, and B. E. Bernstein. Genetic  
447 and epigenetic fine mapping of causal autoimmune disease variants. *Nature*, 518(7539):337–343, Feb  
448 2015.
- 449 [3] E. Pennisi. The Biology of Genomes. Disease risk links to gene regulation. *Science*, 332(6033):1031, May  
450 2011.
- 451 [4] L. A. Hindorff, P. Sethupathy, H. A. Junkins, E. M. Ramos, J. P. Mehta, F. S. Collins, and T. A. Manolio.  
452 Potential etiologic and functional implications of genome-wide association loci for human diseases and  
453 traits. *Proc Natl Acad Sci*, 106(23):9362–9367, Jun 2009.
- 454 [5] J. K. Pickrell. Joint analysis of functional genomic data and genome-wide association studies of 18  
455 human traits. *Am J Hum Genet*, 94(4):559–573, Apr 2014.
- 456 [6] D. Welter, J. MacArthur, J. Morales, T. Burdett, P. Hall, H. Junkins, A. Klemm, P. Flieck, T. Manolio,  
457 L. Hindorff, and H. Parkinson. The NHGRI GWAS Catalog, a curated resource of SNP-trait associations.  
458 *Nucleic Acids Res*, 42(Database issue):D1001–1006, Jan 2014.
- 459 [7] Y. I. Li, B. van de Geijn, A. Raj, D. A. Knowles, A. A. Petti, D. Golan, Y. Gilad, and J. K. Pritchard.  
460 RNA splicing is a primary link between genetic variation and disease. *Science*, 352(6285):600–604, Apr  
461 2016.
- 462 [8] D. Zhou, Y. Jiang, X. Zhong, N. J. Cox, C. Liu, and E. R. Gamazon. A unified framework for joint-tissue  
463 transcriptome-wide association and Mendelian randomization analysis. *Nat Genet*, 52(11):1239–1246,  
464 Nov 2020.
- 465 [9] E. R. Gamazon, H. E. Wheeler, K. P. Shah, S. V. Mozaffari, K. Aquino-Michaels, R. J. Carroll, A. E.

- 466 Eyler, J. C. Denny, D. L. Nicolae, N. J. Cox, and H. K. Im. A gene-based association method for  
467 mapping traits using reference transcriptome data. *Nat Genet*, 47(9):1091–1098, Sep 2015.
- 468 [10] Z. Zhu, F. Zhang, H. Hu, A. Bakshi, M. R. Robinson, J. E. Powell, G. W. Montgomery, M. E. Goddard,  
469 N. R. Wray, P. M. Visscher, and J. Yang. Integration of summary data from GWAS and eQTL studies  
470 predicts complex trait gene targets. *Nat Genet*, 48(5):481–487, May 2016.
- 471 [11] A. Gusev, A. Ko, H. Shi, G. Bhatia, W. Chung, B. W. Penninx, R. Jansen, E. J. de Geus, D. I. Boomsma,  
472 F. A. Wright, P. F. Sullivan, E. Nikkola, M. Alvarez, M. Civelek, A. J. Lusis, T. ki, E. Raitoharju,  
473 M. nen, I. ä, O. T. Raitakari, J. Kuusisto, M. Laakso, A. L. Price, P. Pajukanta, and B. Pasaniuc.  
474 Integrative approaches for large-scale transcriptome-wide association studies. *Nat Genet*, 48(3):245–252,  
475 Mar 2016.
- 476 [12] M. P. Keller, D. M. Gatti, K. L. Schueler, M. E. Rabaglia, D. S. Stapleton, P. Simecek, M. Vincent,  
477 S. Allen, A. T. Broman, R. Bacher, C. Kendzierski, K. W. Broman, B. S. Yandell, G. A. Churchill, and  
478 A. D. Attie. Genetic Drivers of Pancreatic Islet Function. *Genetics*, 209(1):335–356, May 2018.
- 479 [13] W. L. Crouse, G. R. Keele, M. S. Gastonguay, G. A. Churchill, and W. Valdar. A Bayesian model  
480 selection approach to mediation analysis. *PLoS Genet*, 18(5):e1010184, May 2022.
- 481 [14] J. M. Chick, S. C. Munger, P. Simecek, E. L. Huttlin, K. Choi, D. M. Gatti, N. Raghupathy, K. L. Svenson,  
482 G. A. Churchill, and S. P. Gygi. Defining the consequences of genetic variation on a proteome-wide scale.  
483 *Nature*, 534(7608):500–505, Jun 2016.
- 484 [15] H. E. Wheeler, S. Ploch, A. N. Barbeira, R. Bonazzola, A. Andaleon, A. Fotuhi Siahpirani, A. Saha,  
485 A. Battle, S. Roy, and H. K. Im. Imputed gene associations identify replicable trans-acting genes enriched  
486 in transcription pathways and complex traits. *Genet Epidemiol*, 43(6):596–608, Sep 2019.
- 487 [16] B. D. Umans, A. Battle, and Y. Gilad. Where Are the Disease-Associated eQTLs? *Trends Genet*,  
488 37(2):109–124, Feb 2021.
- 489 [17] N. J. Connally, S. Nazeen, D. Lee, H. Shi, J. Stamatoyannopoulos, S. Chun, C. Cotsapas, C. A. Cassa,  
490 and S. R. Sunyaev. The missing link between genetic association and regulatory function. *Elife*, 11, Dec  
491 2022.
- 492 [18] H. Mostafavi, J. P. Spence, S. Naqvi, and J. K. Pritchard. Systematic differences in discovery of genetic  
493 effects on gene expression and complex traits. *Nat Genet*, 55(11):1866–1875, Nov 2023.
- 494 [19] Aparna Nathan, Samira Asgari, Kazuyoshi Ishigaki, Cristian Valencia, Tiffany Amariuta, Yang Luo,

- 495 Jessica I Beynor, Yuriy Baglaenko, Sara Suliman, Alkes L Price, et al. Single-cell eqtl models reveal  
496 dynamic t cell state dependence of disease loci. *Nature*, 606(7912):120–128, 2022.
- 497 [20] D. W. Yao, L. J. O'Connor, A. L. Price, and A. Gusev. Quantifying genetic effects on disease mediated  
498 by assayed gene expression levels. *Nat Genet*, 52(6):626–633, Jun 2020.
- 499 [21] X. Liu, J. A. Mefford, A. Dahl, Y. He, M. Subramaniam, A. Battle, A. L. Price, and N. Zaitlen. GBAT:  
500 a gene-based association test for robust detection of trans-gene regulation. *Genome Biol*, 21(1):211, Aug  
501 2020.
- 502 [22] E. Sollis, A. Mosaku, A. Abid, A. Buniello, M. Cerezo, L. Gil, T. Groza, O. §, P. Hall, J. Hayhurst,  
503 A. Ibrahim, Y. Ji, S. John, E. Lewis, J. A. L. MacArthur, A. McMahon, D. Osumi-Sutherland,  
504 K. Panoutsopoulou, Z. Pendlington, S. Ramachandran, R. Stefancsik, J. Stewart, P. Whetzel, R. Wilson,  
505 L. Hindorff, F. Cunningham, S. A. Lambert, M. Inouye, H. Parkinson, and L. W. Harris. The NHGRI-EBI  
506 GWAS Catalog: knowledgebase and deposition resource. *Nucleic Acids Res*, 51(D1):D977–D985, Jan  
507 2023.
- 508 [23] R. J. F. Loos and G. S. H. Yeo. The genetics of obesity: from discovery to biology. *Nat Rev Genet*,  
509 23(2):120–133, Feb 2022.
- 510 [24] R. K. Singh, P. Kumar, and K. Mahalingam. Molecular genetics of human obesity: A comprehensive  
511 review. *C R Biol*, 340(2):87–108, Feb 2017.
- 512 [25] P. Arner. Obesity—a genetic disease of adipose tissue? *Br J Nutr*, 83 Suppl 1:9–16, Mar 2000.
- 513 [26] G. A. Churchill, D. M. Gatti, S. C. Munger, and K. L. Svenson. The Diversity Outbred mouse population.  
514 *Mamm Genome*, 23(9-10):713–718, Oct 2012.
- 515 [27] Elissa J Chesler, Darla R Miller, Lisa R Branstetter, Leslie D Galloway, Barbara L Jackson, Vivek M  
516 Philip, Brynn H Voy, Cymbeline T Culiat, David W Threadgill, Robert W Williams, et al. The  
517 collaborative cross at oak ridge national laboratory: developing a powerful resource for systems genetics.  
518 *Mammalian Genome*, 19:382–389, 2008.
- 519 [28] Michael C Saul, Vivek M Philip, Laura G Reinholdt, and Elissa J Chesler. High-diversity mouse  
520 populations for complex traits. *Trends in Genetics*, 35(7):501–514, 2019.
- 521 [29] S. M. Clee and A. D. Attie. The genetic landscape of type 2 diabetes in mice. *Endocr Rev*, 28(1):48–83,  
522 Feb 2007.
- 523 [30] K. W. Broman, D. M. Gatti, P. Simecek, N. A. Furlotte, P. Prins, Š. Sen, B. S. Yandell, and G. A.

- 524 Churchill. R/qtL2: Software for Mapping Quantitative Trait Loci with High-Dimensional Data and  
525 Multiparent Populations. *Genetics*, 211(2):495–502, Feb 2019.
- 526 [31] Fabien Girka, Etienne Camenen, Caroline Peltier, Arnaud Gloaguen, Vincent Guillemot, Laurent Le  
527 Brusquet, and Arthur Tenenhaus. *RGCCA: Regularized and Sparse Generalized Canonical Correlation*  
528 *Analysis for Multiblock Data*, 2023. R package version 3.0.3.
- 529 [32] M. Helmer, S. Warrington, A. R. Mohammadi-Nejad, J. L. Ji, A. Howell, B. Rosand, A. Anticevic,  
530 S. N. Sotiropoulos, and J. D. Murray. On the stability of canonical correlation analysis and partial least  
531 squares with application to brain-behavior associations. *Commun Biol*, 7(1):217, Feb 2024.
- 532 [33] Gennady Korotkevich, Vladimir Sukhov, and Alexey Sergushichev. Fast gene set enrichment analysis.  
533 *bioRxiv*, 2019.
- 534 [34] A. Subramanian, P. Tamayo, V. K. Mootha, S. Mukherjee, B. L. Ebert, M. A. Gillette, A. Paulovich,  
535 S. L. Pomeroy, T. R. Golub, E. S. Lander, and J. P. Mesirov. Gene set enrichment analysis: a  
536 knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci U S A*,  
537 102(43):15545–15550, Oct 2005.
- 538 [35] S. Subramanian and A. Chait. The effect of dietary cholesterol on macrophage accumulation in adipose  
539 tissue: implications for systemic inflammation and atherosclerosis. *Curr Opin Lipidol*, 20(1):39–44, Feb  
540 2009.
- 541 [36] I. Akoumianakis, N. Akawi, and C. Antoniades. Exploring the Crosstalk between Adipose Tissue and  
542 the Cardiovascular System. *Korean Circ J*, 47(5):670–685, Sep 2017.
- 543 [37] I. S. Stafeev, A. V. Vorotnikov, E. I. Ratner, M. Y. Menshikov, and Y. V. Parfyonova. Latent Inflammation  
544 and Insulin Resistance in Adipose Tissue. *Int J Endocrinol*, 2017:5076732, 2017.
- 545 [38] I. P. Fischer, M. Irmler, C. W. Meyer, S. J. Sachs, F. Neff, M. de Angelis, J. Beckers, M. H. p, S. M.  
546 Hofmann, and S. Ussar. A history of obesity leaves an inflammatory fingerprint in liver and adipose  
547 tissue. *Int J Obes (Lond)*, 42(3):507–517, Mar 2018.
- 548 [39] S. Chung, H. Cuffe, S. M. Marshall, A. L. McDaniel, J. H. Ha, K. Kavanagh, C. Hong, P. Tontonoz,  
549 R. E. Temel, and J. S. Parks. Dietary cholesterol promotes adipocyte hypertrophy and adipose tissue  
550 inflammation in visceral, but not in subcutaneous, fat in monkeys. *Arterioscler Thromb Vasc Biol*,  
551 34(9):1880–1887, Sep 2014.
- 552 [40] V. Kus, T. Prazak, P. Brauner, M. Hensler, O. Kuda, P. Flachs, P. Janovska, D. Medrikova, M. Rossmeisl,  
553 Z. Jilkova, B. Stefl, E. Pastalkova, Z. Drahota, J. Houstek, and J. Kopecky. Induction of muscle

- 554 thermogenesis by high-fat diet in mice: association with obesity-resistance. *Am J Physiol Endocrinol*  
555 *Metab*, 295(2):E356–367, Aug 2008.
- 556 [41] C. B. Newgard. Interplay between lipids and branched-chain amino acids in development of insulin  
557 resistance. *Cell Metab*, 15(5):606–614, May 2012.
- 558 [42] D. D. Sears, G. Hsiao, A. Hsiao, J. G. Yu, C. H. Courtney, J. M. Ofrecio, J. Chapman, and S. Subramaniam.  
559 Mechanisms of human insulin resistance and thiazolidinedione-mediated insulin sensitization. *Proc Natl  
560 Acad Sci U S A*, 106(44):18745–18750, Nov 2009.
- 561 [43] R. Stienstra, C. Duval, M. ller, and S. Kersten. PPARs, Obesity, and Inflammation. *PPAR Res*,  
562 2007:95974, 2007.
- 563 [44] O. Gavrilova, M. Haluzik, K. Matsusue, J. J. Cutson, L. Johnson, K. R. Dietz, C. J. Nicol, C. Vinson,  
564 F. J. Gonzalez, and M. L. Reitman. Liver peroxisome proliferator-activated receptor gamma contributes  
565 to hepatic steatosis, triglyceride clearance, and regulation of body fat mass. *J Biol Chem*, 278(36):34268–  
566 34276, Sep 2003.
- 567 [45] K. Matsusue, M. Haluzik, G. Lambert, S. H. Yim, O. Gavrilova, J. M. Ward, B. Brewer, M. L. Reitman,  
568 and F. J. Gonzalez. Liver-specific disruption of PPARgamma in leptin-deficient mice improves fatty  
569 liver but aggravates diabetic phenotypes. *J Clin Invest*, 111(5):737–747, Mar 2003.
- 570 [46] D. Patsouris, J. K. Reddy, M. ller, and S. Kersten. Peroxisome proliferator-activated receptor alpha  
571 mediates the effects of high-fat diet on hepatic gene expression. *Endocrinology*, 147(3):1508–1516, Mar  
572 2006.
- 573 [47] S. E. Schadinger, N. L. Bucher, B. M. Schreiber, and S. R. Farmer. PPARgamma2 regulates lipogenesis  
574 and lipid accumulation in steatotic hepatocytes. *Am J Physiol Endocrinol Metab*, 288(6):E1195–1205,  
575 Jun 2005.
- 576 [48] W. Motomura, M. Inoue, T. Ohtake, N. Takahashi, M. Nagamine, S. Tanno, Y. Kohgo, and T. Okumura.  
577 Up-regulation of ADRP in fatty liver in human and liver steatosis in mice fed with high fat diet. *Biochem  
578 Biophys Res Commun*, 340(4):1111–1118, Feb 2006.
- 579 [49] A. Srivastava, A. P. Morgan, M. L. Najarian, V. K. Sarsani, J. S. Sigmon, J. R. Shorter, A. Kashfeen,  
580 R. C. McMullan, L. H. Williams, P. guez, M. T. Ferris, P. Sullivan, P. Hock, D. R. Miller, T. A. Bell,  
581 L. McMillan, G. A. Churchill, and F. P. de Villena. Genomes of the Mouse Collaborative Cross. *Genetics*,  
582 206(2):537–556, Jun 2017.

- 583 [50] D. W. Threadgill, D. R. Miller, G. A. Churchill, and F. P. de Villena. The collaborative cross: a  
584 recombinant inbred mouse population for the systems genetic era. *ILAR J*, 52(1):24–31, 2011.
- 585 [51] A. Roberts, F. Pardo-Manuel de Villena, W. Wang, L. McMillan, and D. W. Threadgill. The poly-  
586 morphism architecture of mouse genetic resources elucidated using genome-wide resequencing data:  
587 implications for QTL discovery and systems genetics. *Mamm Genome*, 18(6-7):473–481, Jul 2007.
- 588 [52] G. A. Churchill, D. C. Airey, H. Allayee, J. M. Angel, A. D. Attie, J. Beatty, W. D. Beavis, J. K.  
589 Belknap, B. Bennett, W. Berrettini, A. Bleich, M. Bogue, K. W. Broman, K. J. Buck, E. Buckler,  
590 M. Burmeister, E. J. Chesler, J. M. Cheverud, S. Clapcote, M. N. Cook, R. D. Cox, J. C. Crabbe,  
591 W. E. Crusio, A. Darvasi, C. F. Deschepper, R. W. Doerge, C. R. Farber, J. Forejt, D. Gaile, S. J.  
592 Garlow, H. Geiger, H. Gershenson, T. Gordon, J. Gu, W. Gu, G. de Haan, N. L. Hayes, C. Heller,  
593 H. Himmelbauer, R. Hitzemann, K. Hunter, H. C. Hsu, F. A. Iraqi, B. Ivandic, H. J. Jacob, R. C. Jansen,  
594 K. J. Jepsen, D. K. Johnson, T. E. Johnson, G. Kempermann, C. Kendzierski, M. Kotb, R. F. Kooy,  
595 B. Llamas, F. Lammert, J. M. Lassalle, P. R. Lowenstein, L. Lu, A. Lusis, K. F. Manly, R. Marcucio,  
596 D. Matthews, J. F. Medrano, D. R. Miller, G. Mittelman, B. A. Mock, J. S. Mogil, X. Montagutelli,  
597 G. Morahan, D. G. Morris, R. Mott, J. H. Nadeau, H. Nagase, R. S. Nowakowski, B. F. O’Hara, A. V.  
598 Osadchuk, G. P. Page, B. Paigen, K. Paigen, A. A. Palmer, H. J. Pan, L. Peltonen-Palotie, J. Peirce,  
599 D. Pomp, M. Pravenec, D. R. Prows, Z. Qi, R. H. Reeves, J. Roder, G. D. Rosen, E. E. Schadt, L. C.  
600 Schalkwyk, Z. Seltzer, K. Shimomura, S. Shou, M. J. ä, L. D. Siracusa, H. W. Snoeck, J. L. Spearow,  
601 K. Svenson, L. M. Tarantino, D. Threadgill, L. A. Toth, W. Valdar, F. P. de Villena, C. Warden,  
602 S. Whatley, R. W. Williams, T. Wiltshire, N. Yi, D. Zhang, M. Zhang, and F. Zou. The Collaborative  
603 Cross, a community resource for the genetic analysis of complex traits. *Nat Genet*, 36(11):1133–1137,  
604 Nov 2004.
- 605 [53] J. Y. Huh, Y. J. Park, M. Ham, and J. B. Kim. Crosstalk between adipocytes and immune cells in  
606 adipose tissue inflammation and metabolic dysregulation in obesity. *Mol Cells*, 37(5):365–371, May 2014.
- 607 [54] J. Lamb, E. D. Crawford, D. Peck, J. W. Modell, I. C. Blat, M. J. Wrobel, J. Lerner, J. P. Brunet,  
608 A. Subramanian, K. N. Ross, M. Reich, H. Hieronymus, G. Wei, S. A. Armstrong, S. J. Haggarty,  
609 P. A. Clemons, R. Wei, S. A. Carr, E. S. Lander, and T. R. Golub. The Connectivity Map: using  
610 gene-expression signatures to connect small molecules, genes, and disease. *Science*, 313(5795):1929–1935,  
611 Sep 2006.
- 612 [55] S. Amin, A. Lux, and F. O’Callaghan. The journey of metformin from glycaemic control to mTOR  
613 inhibition and the suppression of tumour growth. *Br J Clin Pharmacol*, 85(1):37–46, Jan 2019.

- 614 [56] A. Kezic, L. Popovic, and K. Lalic. mTOR Inhibitor Therapy and Metabolic Consequences: Where Do  
615 We Stand? *Oxid Med Cell Longev*, 2018:2640342, 2018.
- 616 [57] A. D. Barlow, M. L. Nicholson, and T. P. Herbert. -cells and a review of the underlying molecular  
617 mechanisms. *Diabetes*, 62(8):2674–2682, Aug 2013.
- 618 [58] Y. Gu, J. Lindner, A. Kumar, W. Yuan, and M. A. Magnuson. Rictor/mTORC2 is essential for  
619 maintaining a balance between beta-cell proliferation and cell size. *Diabetes*, 60(3):827–837, Mar 2011.
- 620 [59] E. B. Geer, J. Islam, and C. Buettner. Mechanisms of glucocorticoid-induced insulin resistance: focus on  
621 adipose tissue function and lipid metabolism. *Endocrinol Metab Clin North Am*, 43(1):75–102, Mar 2014.
- 622 [60] J. X. Li and C. L. Cummins. Fresh insights into glucocorticoid-induced diabetes mellitus and new  
623 therapeutic directions. *Nat Rev Endocrinol*, 18(9):540–557, Sep 2022.
- 624 [61] R. A. Lee, C. A. Harris, and J. C. Wang. Glucocorticoid Receptor and Adipocyte Biology. *Nucl Receptor*  
625 *Res*, 5, 2018.
- 626 [62] S. Viengchareun, P. Penfornis, M. C. Zennaro, and M. s. Mineralocorticoid and glucocorticoid receptors  
627 inhibit UCP expression and function in brown adipocytes. *Am J Physiol Endocrinol Metab*, 280(4):E640–  
628 649, Apr 2001.
- 629 [63] J. Liu, X. Kong, L. Wang, H. Qi, W. Di, X. Zhang, L. Wu, X. Chen, J. Yu, J. Zha, S. Lv, A. Zhang,  
630 P. Cheng, M. Hu, Y. Li, J. Bi, Y. Li, F. Hu, Y. Zhong, Y. Xu, and G. Ding. -HSD1 in regulating brown  
631 adipocyte function. *J Mol Endocrinol*, 50(1):103–113, Feb 2013.
- 632 [64] L. E. Ramage, M. Akyol, A. M. Fletcher, J. Forsythe, M. Nixon, R. N. Carter, E. J. van Beek,  
633 N. M. Morton, B. R. Walker, and R. H. Stimson. Glucocorticoids Acutely Increase Brown Adipose  
634 Tissue Activity in Humans, Revealing Species-Specific Differences in UCP-1 Regulation. *Cell Metab*,  
635 24(1):130–141, Jul 2016.
- 636 [65] J. L. Barclay, H. Agada, C. Jang, M. Ward, N. Wetzig, and K. K. Ho. Effects of glucocorticoids on  
637 human brown adipocytes. *J Endocrinol*, 224(2):139–147, Feb 2015.
- 638 [66] M. ó, R. Gupte, W. L. Kraus, P. Pacher, and P. Bai. PARPs in lipid metabolism and related diseases.  
639 *Prog Lipid Res*, 84:101117, Nov 2021.
- 640 [67] P. Bai, C. ó, H. Oudart, A. nszki, Y. Cen, C. Thomas, H. Yamamoto, A. Huber, B. Kiss, R. H.  
641 Houtkooper, K. Schoonjans, V. Schreiber, A. A. Sauve, J. Menissier-de Murcia, and J. Auwerx. PARP-1

- 642 inhibition increases mitochondrial metabolism through SIRT1 activation. *Cell Metab*, 13(4):461–468,  
643 Apr 2011.
- 644 [68] A. Chiarugi and M. A. Moskowitz. Cell biology. PARP-1—a perpetrator of apoptotic cell death? *Science*,  
645 297(5579):200–201, Jul 2002.
- 646 [69] M. Althubiti, R. Almaimani, S. Y. Eid, M. Elzubaier, B. Refaat, S. Idris, T. A. Alqurashi, and M. Z.  
647 El-Readi. BTK targeting suppresses inflammatory genes and ameliorates insulin resistance. *Eur Cytokine  
648 Netw*, 31(4):168–179, Dec 2020.
- 649 [70] C. Skrabs, W. F. Pickl, T. Perkmann, U. ger, and A. Gessl. Rapid decline in insulin antibodies and  
650 glutamic acid decarboxylase autoantibodies with ibrutinib therapy of chronic lymphocytic leukaemia. *J  
651 Clin Pharm Ther*, 43(1):145–149, Feb 2018.
- 652 [71] Hans Hacker and Michael Karin. Regulation and function of ikk and ikk-related kinases. *Science's  
653 STKE*, 2006(357):re13–re13, 2006.
- 654 [72] E. A. Oral, S. M. Reilly, A. V. Gomez, R. Meral, L. Butz, N. Ajluni, T. L. Chenevert, E. Korytnaya,  
655 A. H. Neidert, R. Hench, D. Rus, J. F. Horowitz, B. Poirier, P. Zhao, K. Lehmann, M. Jain, R. Yu,  
656 C. Liddle, M. Ahmadian, M. Downes, R. M. Evans, and A. R. Saltiel. and TBK1 Improves Glucose  
657 Control in a Subset of Patients with Type 2 Diabetes. *Cell Metab*, 26(1):157–170, Jul 2017.
- 658 [73] M. C. Arkan, A. L. Hevener, F. R. Greten, S. Maeda, Z. W. Li, J. M. Long, A. Wynshaw-Boris, G. Poli,  
659 J. Olefsky, and M. Karin. IKK-beta links inflammation to obesity-induced insulin resistance. *Nat Med*,  
660 11(2):191–198, Feb 2005.
- 661 [74] M. Clark, C. J. Kroger, Q. Ke, and R. M. Tisch. The Role of T Cell Receptor Signaling in the  
662 Development of Type 1 Diabetes. *Front Immunol*, 11:615371, 2020.
- 663 [75] P. ndell, A. C. Carlsson, A. Larsson, O. Melander, T. Wessman, J. v, and T. Ruge. TNFR1 is associated  
664 with short-term mortality in patients with diabetes and acute dyspnea seeking care at the emergency  
665 department. *Acta Diabetol*, 57(10):1145–1150, Oct 2020.
- 666 [76] A. M. Keestra-Gounder, M. X. Byndloss, N. Seyffert, B. M. Young, A. vez Arroyo, A. Y. Tsai, S. A.  
667 Cevallos, M. G. Winter, O. H. Pham, C. R. Tiffany, M. F. de Jong, T. Kerrinnes, R. Ravindran, P. A.  
668 Luciw, S. J. McSorley, A. J. umler, and R. M. Tsolis. NOD1 and NOD2 signalling links ER stress with  
669 inflammation. *Nature*, 532(7599):394–397, Apr 2016.
- 670 [77] A. M. Keestra-Gounder and R. M. Tsolis. NOD1 and NOD2: Beyond Peptidoglycan Sensing. *Trends  
671 Immunol*, 38(10):758–767, Oct 2017.

- 672 [78] J. Montane, L. Cadavez, and A. Novials. Stress and the inflammatory process: a major cause of  
673 pancreatic cell death in type 2 diabetes. *Diabetes Metab Syndr Obes*, 7:25–34, 2014.
- 674 [79] B. B. Kahn and T. E. McGraw. , and type 2 diabetes. *N Engl J Med*, 363(27):2667–2669, Dec 2010.
- 675 [80] S. Del Prato and N. Pulizzi. The place of sulfonylureas in the therapy for type 2 diabetes mellitus.  
676 *Metabolism*, 55(5 Suppl 1):S20–27, May 2006.
- 677 [81] X. Liu, Y. I. Li, and J. K. Pritchard. Trans Effects on Gene Expression Can Drive Omnigenic Inheritance.  
678 *Cell*, 177(4):1022–1034, May 2019.
- 679 [82] B. Hallgrímsson, R. M. Green, D. C. Katz, J. L. Fish, F. P. Bernier, C. C. Roseman, N. M. Young,  
680 J. M. Cheverud, and R. S. Marcucio. The developmental-genetics of canalization. *Semin Cell Dev Biol*,  
681 88:67–79, Apr 2019.
- 682 [83] M. L. Siegal and A. Bergman. Waddington’s canalization revisited: developmental stability and evolution.  
683 *Proc Natl Acad Sci U S A*, 99(16):10528–10532, Aug 2002.
- 684 [84] A. B. Paaby and G. Gibson. Cryptic Genetic Variation in Evolutionary Developmental Genetics. *Biology*  
685 (*Basel*), 5(2), Jun 2016.
- 686 [85] E. A. Boyle, Y. I. Li, and J. K. Pritchard. An Expanded View of Complex Traits: From Polygenic to  
687 Omnigenic. *Cell*, 169(7):1177–1186, Jun 2017.
- 688 [86] Naomi R Wray, Cisca Wijmenga, Patrick F Sullivan, Jian Yang, and Peter M Visscher. Common disease  
689 is more complex than implied by the core gene omnigenic model. *Cell*, 173(7):1573–1580, 2018.