

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

<sup>3</sup>

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

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

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

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

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

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

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

31 One possible explanation for these observations is that gene expression is not being measured in the appropriate  
32 cell types and thus local eQTLs influencing traits cannot be detected<sup>16</sup>. An alternative explanation that has  
33 been discussed in recent years is that effects of these variants are mediated not through local regulation of  
34 gene expression, but through distal regulation<sup>18–20;15</sup>.

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

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

47 Measuring gene expression in multiple tissues is critical to adequately assess the extent to which local gene  
48 regulation varies across the tissues and whether such variability might account for previous failed attempts to  
49 identify trait-relevant local eQTL. Such data sets are extremely difficult to obtain in human populations,  
50 particularly in the large numbers of subjects required for adequate statistical power. Thus, to further  
51 investigate the role of local and distal gene regulation on complex traits, we generated two complementary  
52 data sets: A discovery data set in a large population of diversity outbred (DO) mice<sup>21</sup>, and an independent  
53 validation data set derived by crossing inbred strains from the Collaborative Cross (CC) mice<sup>22</sup> to form CC  
54 F1 mice (CC-RIX). Both populations modeled diet-induced obesity and metabolic disease<sup>12</sup>

55 The DO population and CC recombinant inbred lines were derived from the same eight inbred founder mouse  
56 strains, five classical lab strains, and three strains more recently derived from wild mice<sup>21</sup>. They represent

57 three subspecies of mouse *Mus musculus domesticus*, *Mus musculus musculus*, and *Mus musculus castaneus*,  
58 and capture 90% of the known variation in laboratory mice<sup>23</sup>. The DO mice are maintained with a breeding  
59 scheme that ensures equal contributions from each founder across the genome thus rendering almost the  
60 whole genome visible to genetic inquiry<sup>21</sup>. The CC mice were initially outcrossed to recombine the genomes  
61 from all eight founders, and then inbred for at least 20 generations to generate multiple inbred lines. Because  
62 these two populations have common ancestral haplotypes we could directly and unambiguously compare  
63 the local genetic effects on gene expression at the whole-transcriptome level while varying the population  
64 structure driving distal regulation.

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

## 74 Results

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

### 77 Genetic variation contributed to wide phenotypic variation

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

87 pre-diabetes and diabetes.

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

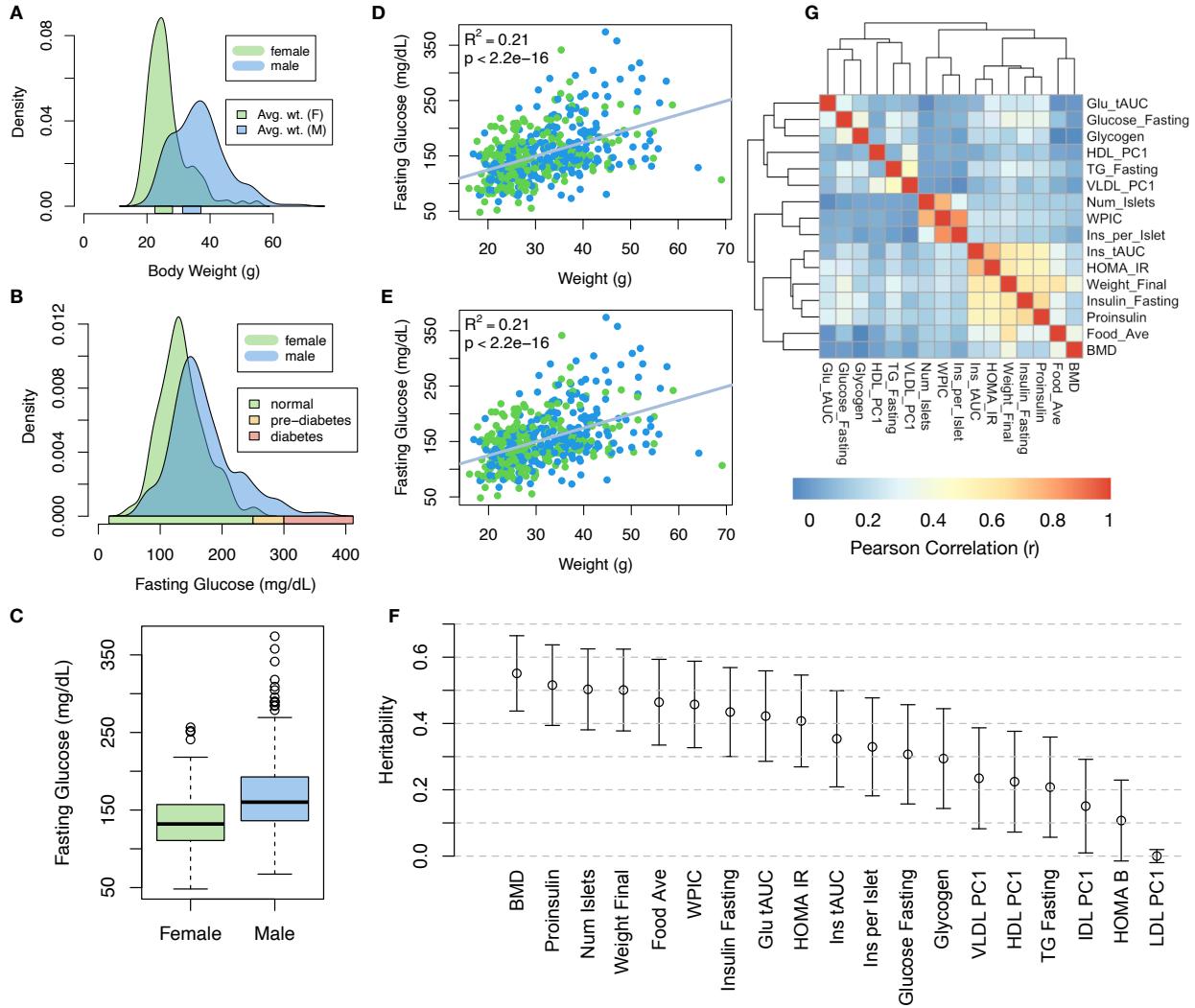


Figure 1: Clinical overview. **A.** Distributions of final body weight in the diversity outbred mice. Sex is indicated by color. The average B6 male and female adult weights at 24 weeks of age are indicated by blue and green bars on the x-axis. **B.** The distribution of final fasting glucose across the population split by sex. Normal, pre-diabetic, and diabetic fasting glucose levels for mice are shown by colored bars along the x-axis. **C.** Males had higher fasting blood glucose on average than females. **D.** The relationship between food consumption and body weight for both sexes. **E.** Relationship between body weight and fasting glucose for both sexes. **F.** Heritability estimates for each physiological trait. Bars show standard error of the estimate. **G.** Correlation structure between physiological traits.

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

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

96 **Distal Heritability Correlated with Phenotype Relevance**

97 To comprehensively assess the genetic control of gene expression in metabolic disease we assayed adipose, islet,  
98 liver, and skeletal muscle gene expression in the DO cohort. We performed eQTL analysis using R/qtl2<sup>25</sup>  
99 (Methods) and identified both local and distal eQTLs for transcripts in each of the four tissues (Supp. Fig  
100 1). Significant local eQTLs far outnumbered distal eQTLs (Supp. Fig. 1F) and tended to be shared across  
101 tissues (Supp. Fig. 1G) whereas the few significant distal eQTLs we identified tended to be tissue-specific  
102 (Supp. Fig. 1H)

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

106 To assess the importance of genetic regulation transcript levels to organism-level traits, we compared the  
107 local and distal heritabilities of transcripts to their trait relevance, defined as the maximum correlation  
108 of a transcript across all traits. The local heritability of transcripts was negatively correlated with their  
109 trait relevance (Fig. 2B), suggesting that the more local genotype influenced transcript abundance, the  
110 less effect this variation had on the measured traits. Conversely, the distal heritability of transcripts was  
111 positively correlated with trait relevance (Fig. 2C). That is, transcripts that were more highly correlated  
112 with the measured traits tended to be distally, rather than locally, heritable. Importantly, this pattern was  
113 consistent across all tissues, strongly suggesting that this is a generic finding. This finding is consistent with  
114 previous observations that low-heritability transcripts explain more expression-mediated disease heritability  
115 than high-heritability transcripts<sup>19</sup>. However, the positive relationship between trait correlation and distal  
116 heritability demonstrated further that there are diffuse genetic effects throughout the genome converging on  
117 trait-related transcripts.

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

120 The above univariate analyses establish the importance of distal heritability for trait-relevant transcripts.  
121 However, the number of transcripts dramatically exceeds the number of phenotypes. Thus, we expect the  
122 heritable, trait-relevant transcripts to be highly correlated and organized according to coherent, emergent

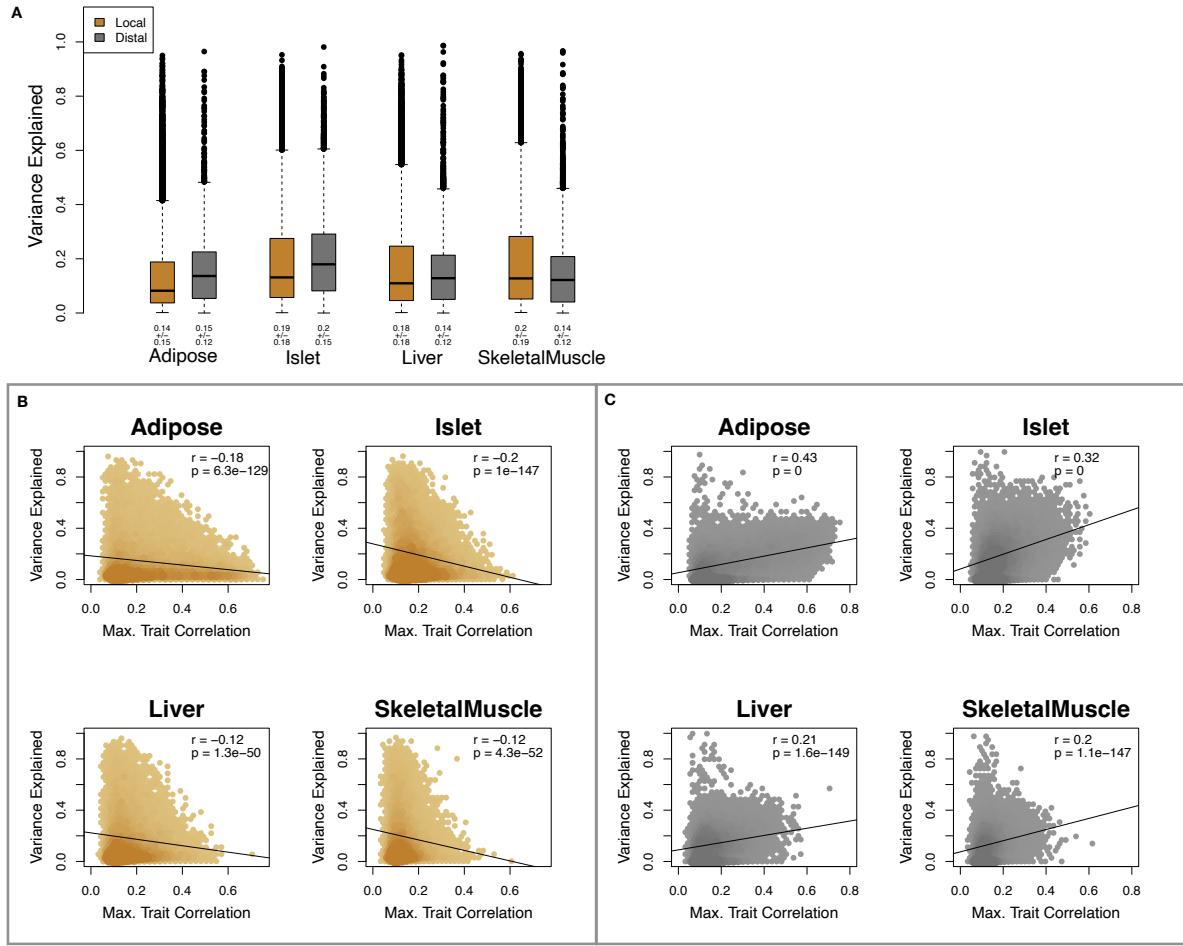


Figure 2: Transcript heritability and trait relevance. **A.** Distributions of distal and local heritability of transcripts across the four tissues. Overall local and distal factors contribute equally to transcript heritability. The relationship between **(B.)** local and **(C.)** distal heritability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

123 biological processes representing the mediating endophenotypes driving clinical trait variation. To identify  
 124 these endophenotypes in a theoretically principled way, we developed a novel dimension-reduction technique,  
 125 HDMA, that uses the theory of causal graphical models to identify a transcriptomic signature that is  
 126 simultaneously 1) highly heritable, 2) strongly correlated to the measured phenotypes, and 3) conforms to the  
 127 causal mediation hypothesis (Fig. 3). HDMA projects the high-dimensional scores—a composite genome score  
 128 ( $G_C$ ), a composite transcriptome score ( $T_C$ ), and a composite phenotype score ( $P_C$ )—and uses the univariate  
 129 theory of mediation to constrain these projections to satisfy the hypotheses of perfect mediation, namely  
 130 that upon controlling for the transcriptomic score, the genome score is uncorrelated to the phenotype score.  
 131 Formally, perfect mediation implies a constraint on the correlation coefficients among scores as

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

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

Operationally, HDMA is closely related to generalized canonical correlation analysis (CCA), for which provably convergent algorithms have recently been developed<sup>26</sup>. A complete mathematical derivation and implementation details for HDMA are available in **Supp. Methods XXX**.

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

As discussed in Supp. Methods XXX, HDMA is related to a generalized form of CCA. Standard CCA is prone to over-fitting because in any two large matrices it can be trivial to identify highly correlated composite vectors [REF]. To assess whether our implementation of HDMA was similarly prone to over-fitting in a high-dimensional space, we performed permutation testing. We permuted the individual labels on the transcriptome matrix 1000 times and recalculated the path coefficient, which is the partial correlation of  $G_C$  and  $T_C$  multiplied by the partial correlation of  $T_C$  and  $P_C$ . This represents the strength of the path from  $G_C$  to  $P_C$  that is putatively mediated through  $T_C$ . The null distribution of the path coefficient is shown in Fig. 3B, and the observed path coefficient from the original data is indicated by a red line. The observed path coefficient was well outside the null distribution generated by permutations ( $p < 10^{-16}$ ). Fig. 3C illustrates this observation in more detail. Although we identified high correlations between  $G_C$  and  $T_C$ , and modest correlations between  $T_C$  and  $P_C$  in the null data (Fig 3C), these two values could not be

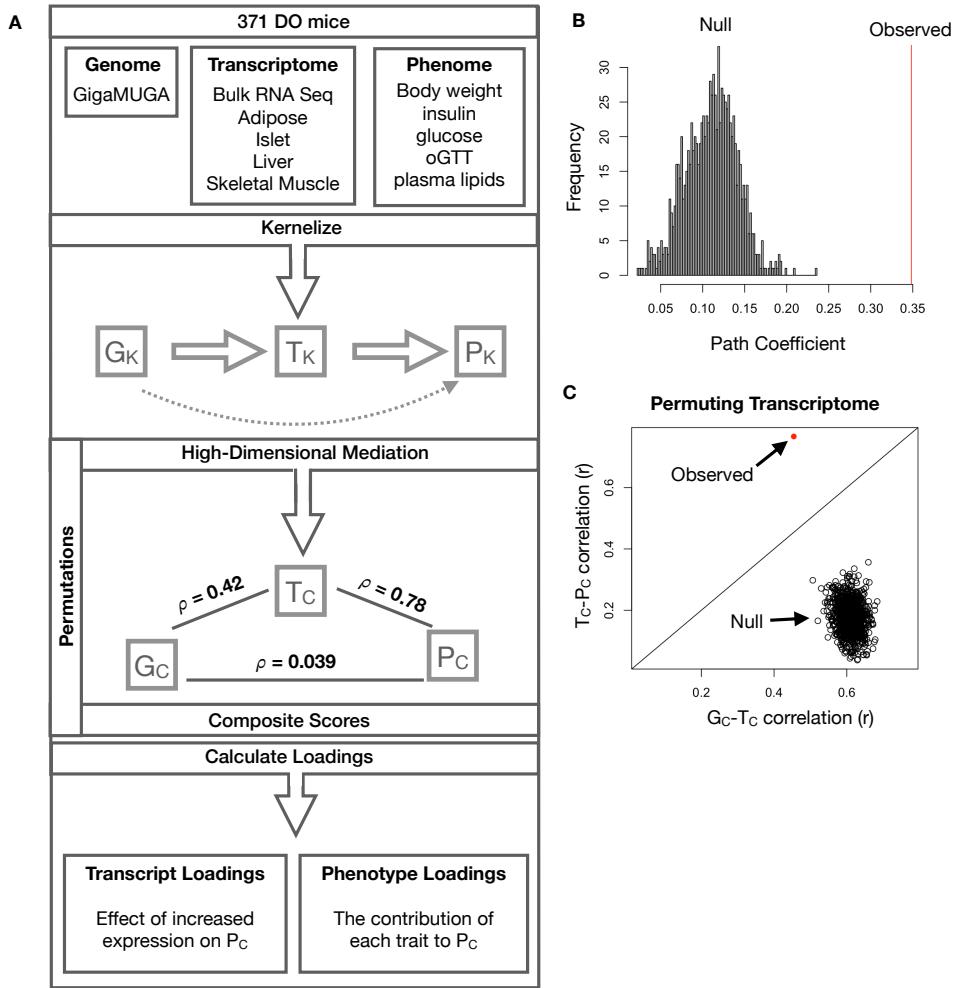


Figure 3: High-dimensional mediation. **A.** Workflow indicating major steps of high-dimensional mediation. The genotype, transcriptome, and phenotype matrices were independently normalized and converted to kernel matrices representing the pairwise relationships between individuals for each data modality ( $K_G$  = genome kernel,  $K_T$  = transcriptome kernel;  $K_P$  = phenome kernel). High-dimensional mediation was applied to these matrices to maximize the direct path  $G \rightarrow T \rightarrow P$ , the mediating pathway (arrows), while simultaneously minimizing the direct  $G \rightarrow P$  pathway (dotted line). The composite vectors that resulted from high-dimensional mediation were  $G_C$ ,  $T_C$ , and  $P_C$ . The partial correlations  $\rho$  between these vectors indicated perfect mediation. Transcript and trait loadings were calculated as described in the methods. **B.** The null distribution of the path coefficient derived from 10,000 permutations compared to the observed path coefficient (red line). **C.** The null distribution of the  $G_C-T_C$  correlation vs. the  $T_C-P_C$  correlation compared with the observed value (red dot).

161 maximized simultaneously in the null data. In contrast, the red dot shows that in the real data both the  
 162  $G_C-T_C$  correlation and the  $T_C-P_C$  correlation could be maximized simultaneously suggesting that the path  
 163 from genotype to phenotype through transcriptome is highly non-trivial and identifiable in this case. These  
 164 results suggest that these composite vectors represent genetically determined variation in phenotype that is  
 165 mediated through genetically determined variation in 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 in  
170 the composite score. The loadings of each measured trait onto  $P_C$  indicate how much each contributed to  
171 the composite phenotype. Final 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  
173 an index of metabolic disease (Fig. 4B). Individuals with high values of  $P_C$  have a higher metabolic index  
174 and greater metabolic disease, including higher body weight and higher insulin resistance. We refer to  $P_C$   
175 as the metabolic index going forward. Traits contributing the least to the metabolic index were measures  
176 of cholesterol and pancreas composition. Thus, when we interpret the transcriptomic signature identified  
177 by HDMA, we are explaining primarily the putative transcriptional mediation of body weight and insulin  
178 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  
182 metabolic index. Large positive loadings indicate that higher expression was associated with a higher  
183 metabolic index (i.e. higher risk of obesity and metabolic disease on the high-fat diet) (Fig. 4C). Conversely,  
184 large negative loadings indicate that high expression of these transcripts was associated with a lower metabolic  
185 index (i.e. lower risk of obesity and metabolic disease on the high-fat diet) (Fig. 4C). We used gene set  
186 enrichment analysis (GSEA)<sup>27;28</sup> to look for biological processes and pathways that were enriched at the top  
187 and bottom of this list (Methods).

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

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

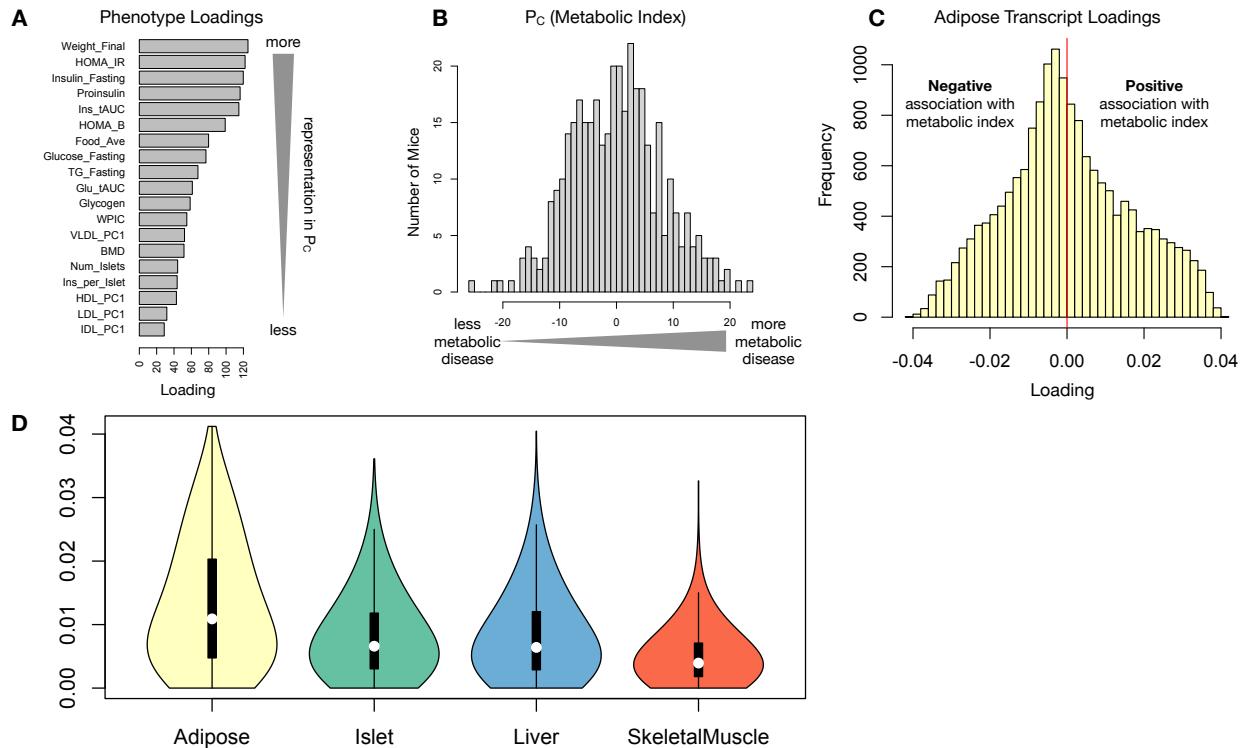


Figure 4: Interpretation of loadings. **A.** Loadings across traits. Body weight and insulin resistance contributed the most to the composite trait. **B.** Phenotype scores across individuals. Individuals with large positive phenotype scores had higher body weight and insulin resistance than average.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

197 Transcripts associated with the citric acid (TCA) cycle as well as the catabolism of the branched-chain amino  
 198 acids (BCAA) (valine, leucine, and isoleucine) were strongly enriched with negative loadings in adipose  
 199 tissue (Supp. Fig. 3). Expression of genes in both pathways (for which there is some overlap) has been  
 200 previously associated with insulin sensitivity<sup>12;29;30</sup>, suggesting that heritable variation in regulation of these  
 201 pathways may influence risk of insulin resistance.  
 202 Looking at the 10 most positively and negatively loaded transcripts from each tissue, it is apparent that  
 203 transcripts in the adipose tissue had the largest loadings, both positive and negative, of all tissues (Fig. 5A  
 204 bar plot) This suggests that much of the effect of genetics on body weight and insulin resistance is mediated  
 205 through gene expression in adipose tissue. The strongest loadings in liver and pancreas were comparable,  
 206 and those in skeletal muscle were the weakest (Fig. 5A), suggesting that less of the genetic effects were  
 207 mediated through transcription in skeletal muscle. Heritability analysis showed that transcripts with the

208 largest loadings had higher distal heritability than local heritability (Fig. 5A heat map and box plot). This  
209 pattern contrasts with transcripts nominated by TWAS (Fig. 5B), which tended to have lower loadings,  
210 higher local heritability and lower distal heritability. Transcripts with the highest local heritability in each  
211 tissue (Fig. 5C) had the lowest loadings, consistent with our findings above (Fig. 2B).

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

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

219 Clustering of transcripts with top loadings in each tissue showed tissue-specific functional modules associated  
220 with obesity and insulin resistance (Fig. 6A) (Methods). The clustering highlights the importance of immune  
221 activation particularly in adipose tissue. The “mitosis” cluster had large positive loadings in three of the  
222 four tissues potentially suggesting system-wide hypertrophy. Otherwise, all clusters were strongly loaded in  
223 only one or two tissues. For example, the lipid metabolism cluster was loaded most heavily in liver. The  
224 positive loadings suggest that high expression of these genes particularly in the liver was associated with  
225 increased metabolic disease. This cluster included the gene *Pparg*, whose primary role is in the adipose tissue  
226 where it is considered a master regulator of adipogenesis<sup>31</sup>. Agonists of *Pparg*, such as thiazolidinediones, are  
227 FDA-approved to treat type II diabetes, and reduce inflammation and adipose hypertrophy<sup>31</sup>. Consistent  
228 with this role, the loading for *Pparg* in adipose tissue was negative, suggesting that higher expression was  
229 associated with leaner mice (Fig. 6B). In contrast, *Pparg* had a large positive loading in liver, where it is  
230 known to play a role in the development of hepatic steatosis, or fatty liver. Mice that lack *Pparg* specifically  
231 in the liver, are protected from developing steatosis and show reduced expression of lipogenic genes<sup>32;33</sup>.  
232 Overexpression of *Pparg* in the livers of mice with a *Ppara* knockout, causes upregulation of genes involved in  
233 adipogenesis<sup>34</sup>. In the livers of both mice and humans high *Pparg* expression is associated with hepatocytes  
234 that accumulate large lipid droplets and have gene expression profiles similar to that of adipocytes<sup>35;36</sup>.

235 The local and distal heritability of *Pparg* is low in adipose tissue suggesting its expression in this tissue is  
236 highly constrained in the population (Fig. 6B). However, the distal heritability of *Pparg* in liver is relatively  
237 high suggesting it is complexly regulated and has sufficient variation in this population to drive variation in  
238 phenotype. Both local and distal heritability of *Pparg* in the islet are relatively high, but the loading is low,

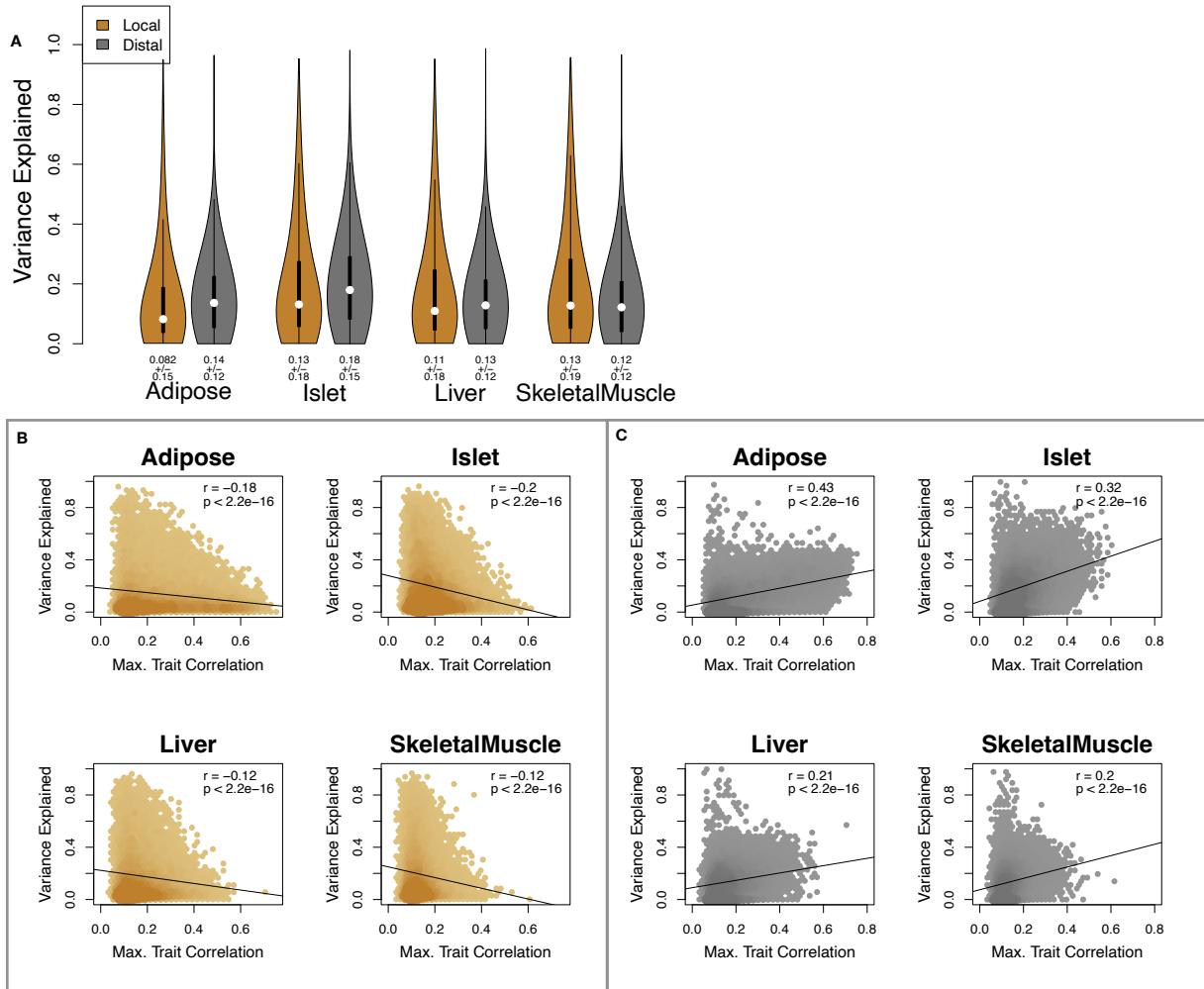


Figure 5: Transcripts with high loadings have high distal heritability and literature support. Each panel has a bar plot showing the loadings of transcripts selected by different criteria. Bar color indicates the tissue of origin. The heat map shows the local (L - left) and distal (D - right) heritability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

239 suggesting that variability of expression in the islet does not drive variation in metabolic index. These results  
 240 highlight the importance of tissue context when investigating the role of heritable transcript variability in  
 241 driving phenotype.

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

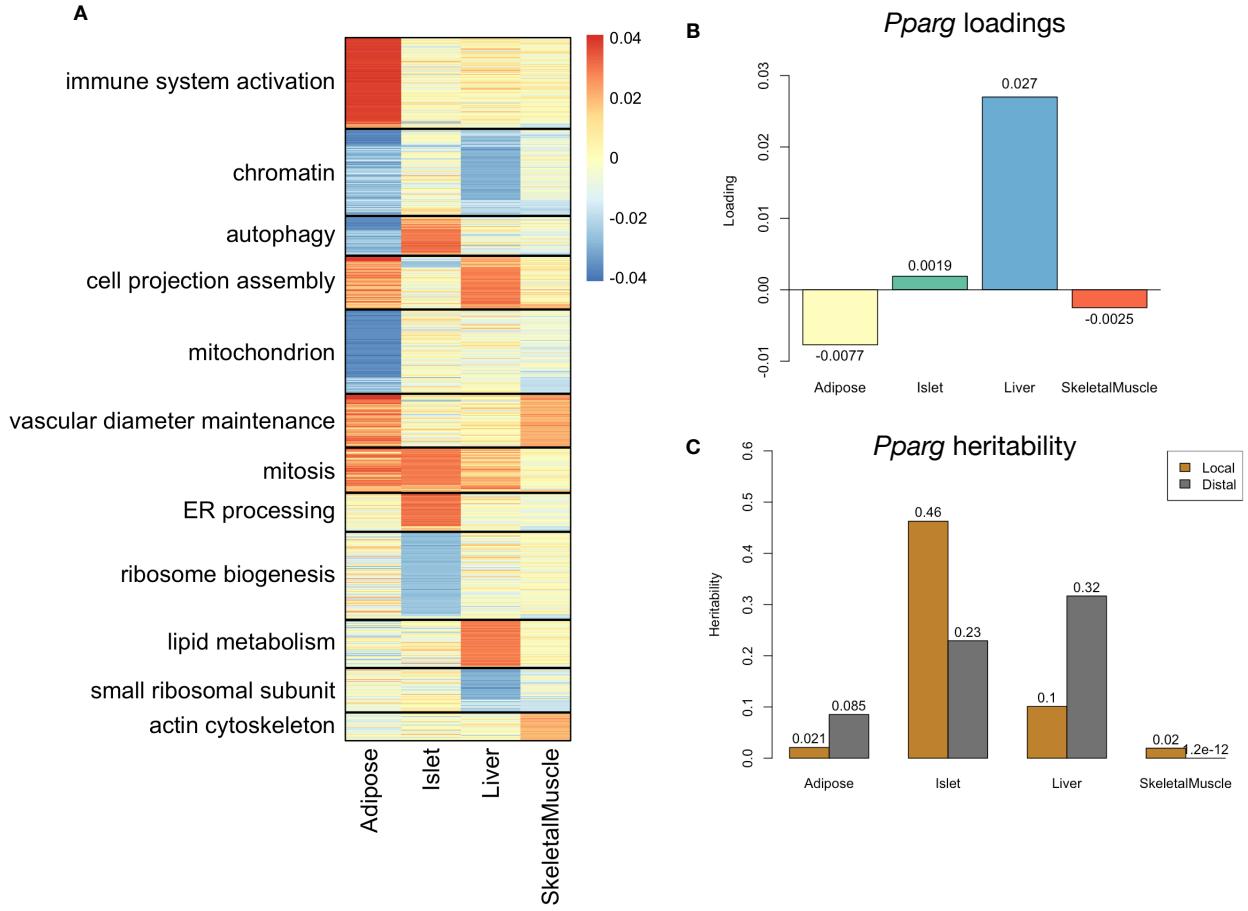


Figure 6: Tissue-specific transcriptional programs were associated with obesity and insulin resistance. **A** Heat map showing the loadings of all transcripts with loadings greater than 2.5 standard deviations from the mean in any tissue. The heat map was clustered using k medoid clustering. Functional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

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

244 To test whether the transcript loadings identified in the DO could be translated to another population, we  
 245 tested whether they could predict metabolic phenotype in an independent population of CC-RIX mice, which  
 246 were F1 mice derived from multiple pairings of Collaborative Cross (CC) [cite] strains (Fig. 7) (Methods).  
 247 We tested two questions. First, we asked whether the loadings identified in the DO mice were relevant to  
 248 the relationship between the transcriptome and the phenotype in the CC-RIX. We predicted body weight (a  
 249 surrogate for metabolic index) in each CC-RIX individual using measured gene expression in each tissue and  
 250 the transcript loadings identified in the DO (Methods). The predicted body weight and acutal body weight  
 251 were highly correlated in all tissues (Fig. 7B left column). The best prediction was achieved for adipose  
 252 tissue, which supports the observation in the DO that adipose expression was the strongest mediator of the

253 genetic effect on metabolic index. This result also confirms the validity and translatability of the transcript  
 254 loadings and their relationship to metabolic disease.

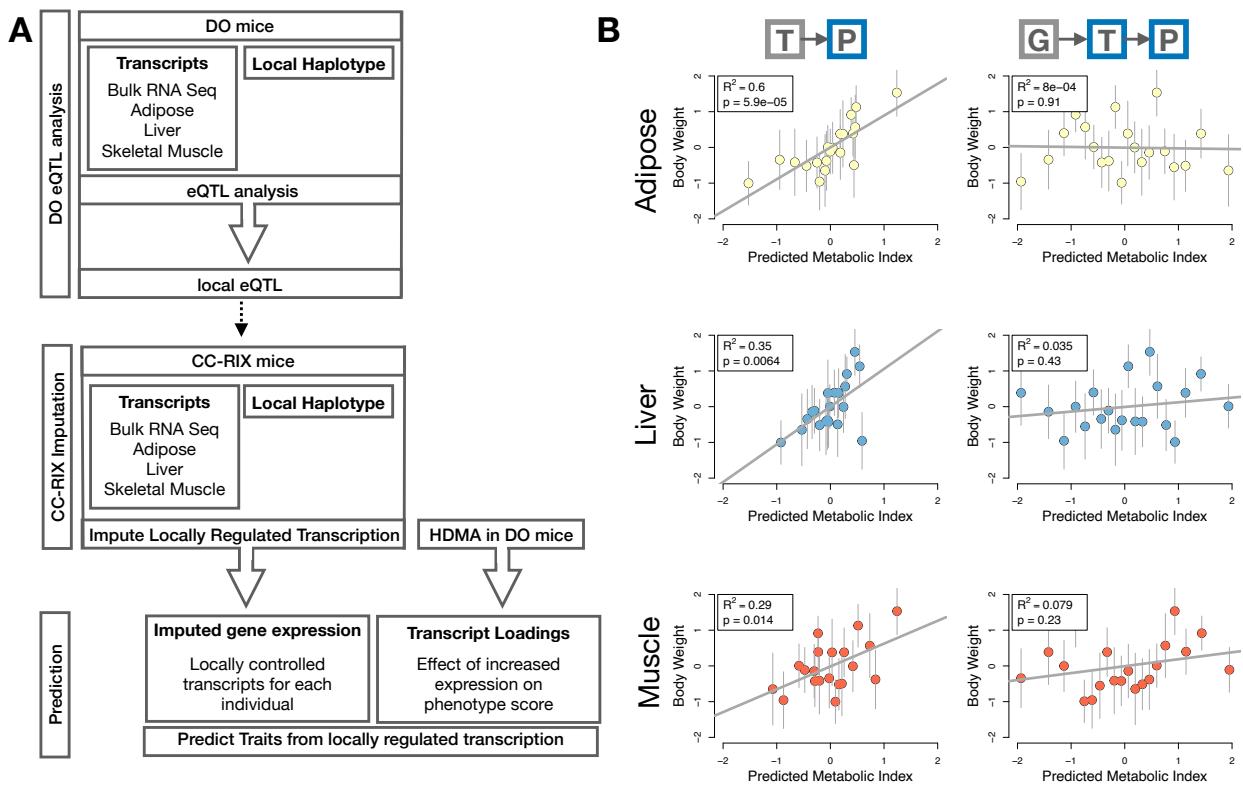


Figure 7: Transcription, but not local genotype, predicts phenotype in the CC-RIX. **A.** Workflow showing procedure for translating HDMA results to an independent population of mice. **B.** Relationships between the predicted metabolic index and measured body weight. The left column shows the predictions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

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

265 **Distally heritable transcriptomic signatures reflected variation in composition of adipose tissue  
266 and islets**

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

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

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

293 **Heritable transcriptomic signatures translated to human disease**

294 Ultimately, the heritable transcriptomic signatures that we identified in DO mice will be useful if they inform  
295 pathogenicity and treatment of human disease. To investigate the potential for translation of the gene

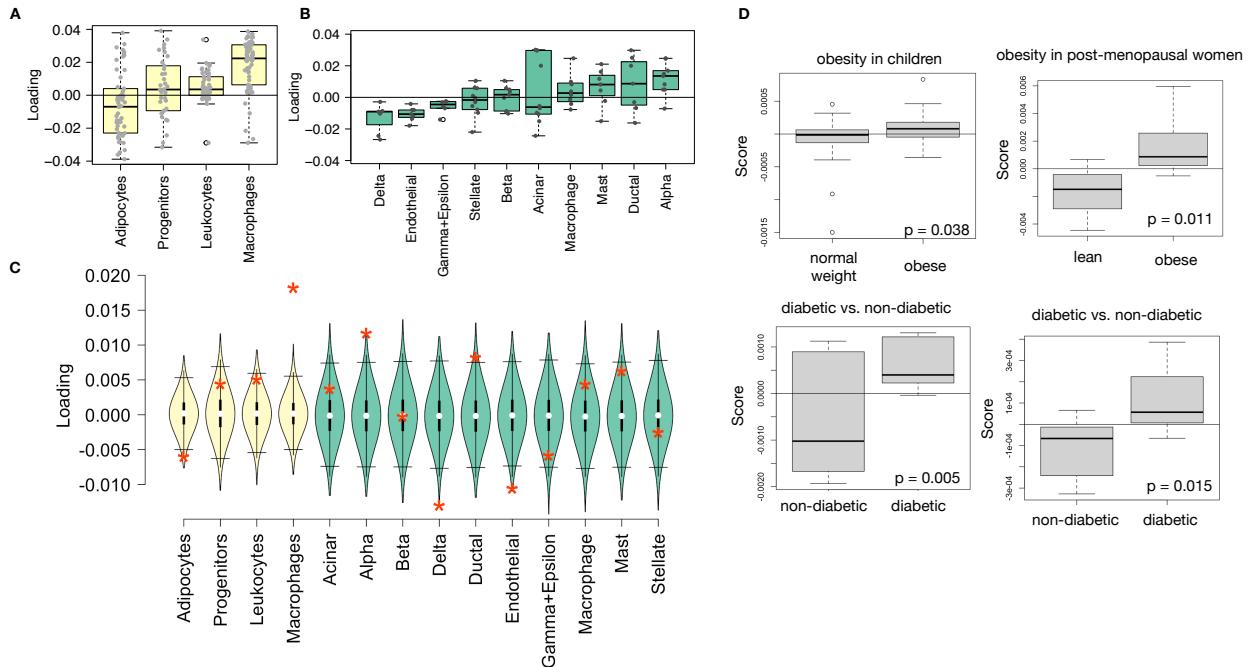


Figure 8: HDMA results translate to humans. **A.** Distribution of loadings for cell-type-specific transcripts in adipose tissue. **B.** Distribution of loadings for cell-type-specific transcripts in pancreatic islets (green). **C.** Null distributions for the mean loading of randomly selected transcripts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

296 signatures identified in DO mice, we compared them to transcriptional profiles in obese and non-obese human  
 297 subjects (Methods). We limited our analysis to adipose tissue because the adipose tissue signature had the  
 298 strongest relationship to obesity and insulin resistance in the DO.

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

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

306 Another potential application of the transcript loading landscape is in ranking potential drug candidates  
 307 for the treatment of metabolic disease. Although high-loading transcripts may be good candidates for  
 308 understanding specific biology related to obesity, the transcriptome overall is highly interconnected and

309 redundant, and focusing on individual transcripts for treatment may be less effective than using broader  
310 transcriptomic signatures that capture the emergent biology [cite or remove]. The ConnectivityMap (CMAP)  
311 database<sup>37</sup> developed by the Broad Institute allows querying thousands of compounds that reverse or enhance  
312 the extreme ends of transcriptomic signatures in multiple different cell types. By identifying drugs that  
313 reverse pathogenic transcriptomic signatures, we can potentially identify compounds that have favorable  
314 effects on gene expression.

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

320 Looking across all cell types, the notable top hits from the adipose tissue loadings included mTOR inhibitors  
321 and glucocorticoid agonists (Supplemental Figure XXX). It is thought that metformin, which is commonly  
322 used to improve glycemic control, acts, at least in part, by inhibiting mTOR signaling<sup>38;39</sup>. However,  
323 long-term use of other mTOR inhibitors, such as rapamycin, are known to cause insulin resistance and  
324  $\beta$ -cell toxicity<sup>39–41</sup>. Glucocorticoids are used to reduce inflammation, which was a prominent signature in  
325 the adipose tissues, but these drugs also promote hyperglycemia and diabetes<sup>42;43</sup>. Acute treatment with  
326 glucocorticoids has further been shown to reduce thermogenesis in rodent adipocytes<sup>44–46</sup>, but increase  
327 thermogenesis in human adipocytes<sup>47;48</sup>. Thus, the pathways identified by CMAP across all cell types were  
328 highly related to the transcript loading profiles, but the relationship was not a simple reversal.

329 The top hit for the adipose composite transcript in CMAP adipocytes was a PARP inhibitor (Supplemental  
330 Figure XXXB). PARPs play a role in lipid metabolism and are involved in the development of obesity and  
331 diabetes<sup>49</sup>. PARP1 inhibition increases mitochondrial biogenesis<sup>50</sup>. Inhibition of PARP1 activity can further  
332 prevent necrosis in favor of the less inflammatory apoptosis<sup>51</sup>, thereby potentially reducing inflammation in  
333 stressed adipocytes. Other notable hits among the top 20 were BTK inhibitors, which have been observed  
334 to suppress inflammation and improve insulin resistance<sup>52</sup> as well as to reduce insulin antibodies in type I  
335 diabetes<sup>53</sup>. IKK inhibitors have been shown to improve glucose control in type II diabetes<sup>54;55</sup>.

336 Among the top most significant hits for the transcript loadings from pancreatic islets (Fig. XXX), was  
337 suppression of T cell receptor signaling, which is known to be involved in Type 1 diabetes<sup>56</sup>, as well as  
338 TNFR1, which has been associated with mortality in diabetes patients<sup>57</sup>. Suppression of NOD1/2 signaling  
339 was also among the top hits. NOD1 and 2 sense ER stress<sup>58;59</sup>, which is associated with  $\beta$ -cell death in type

340 1 and type 2 diabetes<sup>60</sup>. This cell death process is dependent on NOD1/2 signaling<sup>58</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. Hits in this list included widely used diabetes drugs,  
344 such as sulfonylureas, PPAR receptor agonists, and insulin sensitizers. Rosiglitazone is a PPAR- $\gamma$  agonist  
345 and was one of the most prescribed drugs for type 2 diabetes before its use was reduced due to cardiac  
346 side-effects<sup>61</sup>. Sulfonylureas are another commonly prescribed drug class for type 2 diabetes, but also have  
347 notable side effects including hypoglycemia and accelerated  $\beta$ -cell death<sup>62</sup>.

## 348 Discussion

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

362 Genetics is indispensable for the dissection of disease mechanisms because it is one of the only data modalities  
363 that supports causal inferences about molecules and disease outcomes [REF]. It has frequently been assumed  
364 that gene regulation in *cis* is the primary driver of genetically associated trait variation, but attempts to use  
365 local gene regulation to explain phenotypic variation have had limited success<sup>16;17</sup>. In recent years, evidence  
366 has mounted that distal gene regulation may be an important mediator of trait heritability<sup>19;18;63</sup>. It has  
367 been observed that transcripts with high local heritability explain less expression-mediated disease heritability  
368 than those with low local heritability<sup>19</sup>. Consistent with this observation, genes located near GWAS hits  
369 tend to be complexly regulated<sup>18</sup>. They also tend to be enriched with functional annotations, in contrast  
370 to genes with simple local regulation, which tend to be depleted of functional annotations suggesting they

371 are less likely to be directly involved in disease traits<sup>18</sup>. These observations are consistent with principles  
372 of robustness in complex systems in which simple regulation of imoprtant elements leads to fragility of the  
373 system<sup>64–66</sup>. Our results are consistent, instead, with a more complex picture where genes whose expression  
374 can drive trait variation are buffered from local genetic variation but are extensively influenced indirectly by  
375 genetic variation in the regulatory networks converging on those genes.

376 Recently, the omnigenic model of complex traits has been proposed, which posits that complex traits are  
377 massively polygenic and that their heritability is spread out across the genome<sup>67</sup>. In the omnigenic model,  
378 genes are classified either as “core genes,” which directly impinge on the trait, or “peripheral genes,” which  
379 are not directly trait-related, but influence core genes through the complex gene regulatory network. HDMA  
380 explicitly models a central proposal of the omnigenic model which posits that once the expression of the  
381 core genes (i.e. trait-mediating genes) is accounted for, there should be no residual correlation between the  
382 genome and the phenotype. Here, when the composite transcript was taken into account there was no residual  
383 correlation between the composite genome and composite phenotype (Fig. 3A).

384 Thus, the transcript loadings can be interpreted as indicating higher “core-ness” of a transcript. Unlike in the  
385 omnigenic model, we did not observe a clear demarcation between the core and peripheral genes in loading  
386 magnitude, but we do not necessarily expect a clear separation given the complexity of gene regulation and  
387 the genotype-phenotype map<sup>68</sup>.

388 An extension of the omnigenic model proposed that most heritability of complex traits is driven by weak  
389 distal eQTLs that are potentially below the detection threshold in studies with feasible sample sizes<sup>63</sup>. This  
390 is consistent with what we observed here. The transcripts with the largest loadings were strongly distally  
391 regulated and only weakly locally regulated, suggesting that distal gene regulation plays a primary role in  
392 driving heritable trait variation. We saw further that the patterns of distal heritability were not localized to  
393 detectable distal eQTL, but rather were complex and spread across the genome, even for transcripts whose  
394 expression was strongly regulated by distal factors. For example, *Nucb2*, had a high loading in islets and was  
395 also strongly distally regulated (66% distal heritability) (Fig. 5). This gene is expressed in pancreatic  $\beta$  cells  
396 and is involved in insulin and glucagon release<sup>69–71</sup>. Although its transcription was highly heritable in islets,  
397 that regulation was distributed across the genome, with no clear distal eQTL (Supp. Fig. 5). Thus, although  
398 distal regulation of some genes may be strong, this regulation is likely to be highly complex and not easily  
399 localized.

400 We stress that HDMA is a method for causal hypothesis generation. As with any causal inference approach,  
401 the output of HDMA can only be said to be consistent with causal mediation but does not prove it. Proving

402 causality requires experimentation with direct control over the mediating variable [REF]. The issue of  
403 experimentation, however, is subtle. The dimension-reduction in HDMA is distinguished by the fact that  
404 the putative causal intermediates can be emergent states defined by the expression of thousands of genes.  
405 This is a strength, because the mediating variable can be a higher-order process such as “macrophage  
406 activation and infiltration”, but, in contrast to univariate hypotheses at the level of individual transcripts, the  
407 relevant validation experiment may be technologically infeasible, unknowable a priori, or both. Nevertheless,  
408 downstream analyses of the composite transcripts strongly supports a causal interpretation. Indeed, the  
409 composite transcripts identified by HDMA are richly interpretable in both tissue- and gene-specific manners.  
410 The transcripts with the strongest loadings were enriched in biological functions previously known to be  
411 involved in the pathogenesis of metabolic disease, such as inflammation in adipose tissue. That these processes  
412 were identified in this analysis suggests additionally that they have a heritable component, and that some  
413 individuals are genetically susceptible to greater adipose inflammation on a high-fat, high-sugar diet.

414 Individual high-loading transcripts also demonstrated biologically interpretable, tissue-specific patterns. We  
415 highlighted *Pparg*, which is known to be protective in adipose tissue<sup>31</sup> where it was negatively loaded, and  
416 harmful in the liver<sup>32–36</sup>, where it was positively loaded. Such granular patterns may be useful in generating  
417 hypotheses for further testing, and prioritizing genes as therapeutic targets. The tissue-specific nature of  
418 the loadings also may provide clues to tissue-specific effects, or side effects, of targeting particular genes  
419 system-wide.

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

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

<sup>434</sup> across large numbers of individuals.

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

## <sup>441</sup> **Data Availability**

<sup>442</sup> Here we tell people where to find the data

## <sup>443</sup> **Acknowledgements**

<sup>444</sup> Here we thank people

445 **Supplemental Figures**

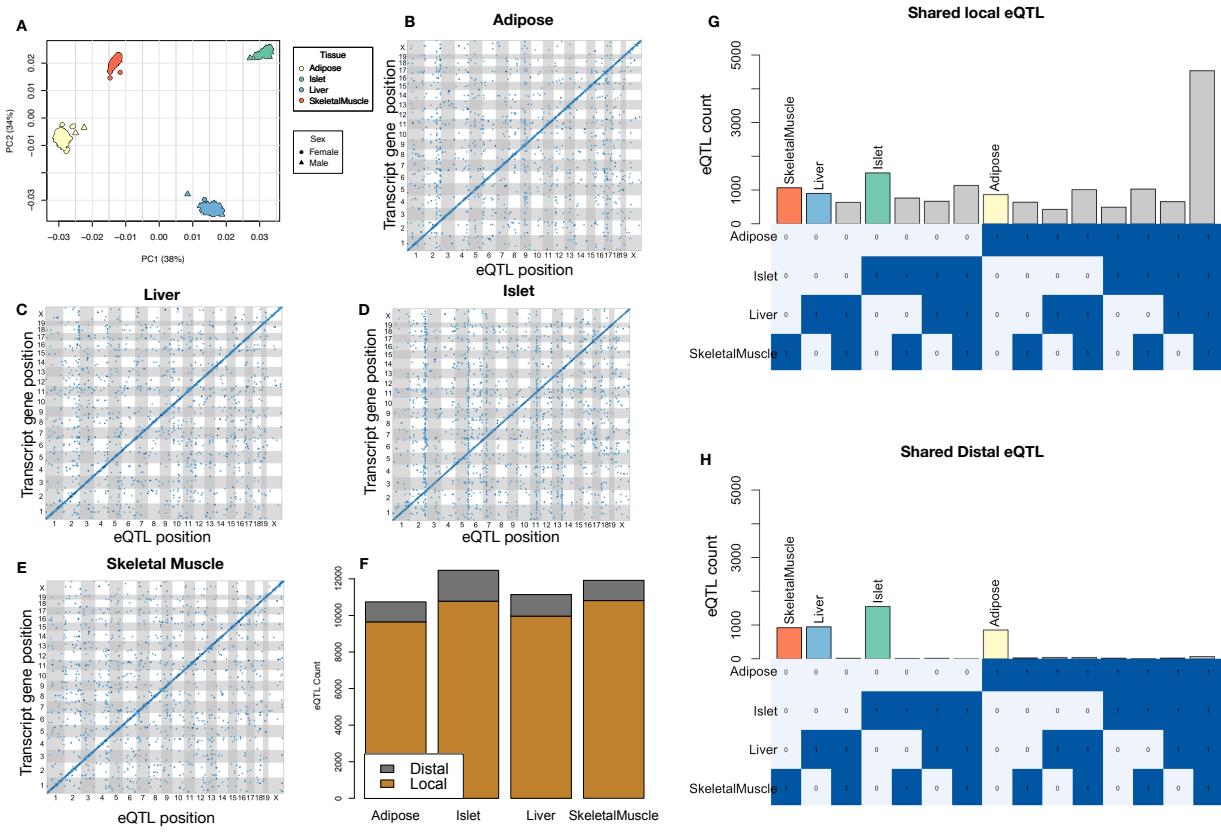


Figure 9: Overview of eQTL analysis in DO mice. **A.** RNA seq samples from the four different tissues clustered by tissue. **B.-E.** eQTL maps are shown for each tissue. The *x*-axis shows the position of the mapped eQTL, and the *y*-axis shows the physical position of the gene encoding each mapped 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.

446 **References**

- 447 [1] M. T. Maurano, R. Humbert, E. Rynes, R. E. Thurman, E. Haugen, H. Wang, A. P. Reynolds,  
448 R. Sandstrom, H. Qu, J. Brody, A. Shafer, F. Neri, K. Lee, T. Kutyavin, S. Stehling-Sun, A. K.  
449 Johnson, T. K. Canfield, E. Giste, M. Diegel, D. Bates, R. S. Hansen, S. Neph, P. J. Sabo, S. Heimfeld,  
450 A. Raubitschek, S. Ziegler, C. Cotsapas, N. Sotoodehnia, I. Glass, S. R. Sunyaev, R. Kaul, and J. A.

## KEGG pathway enrichments by GSEA

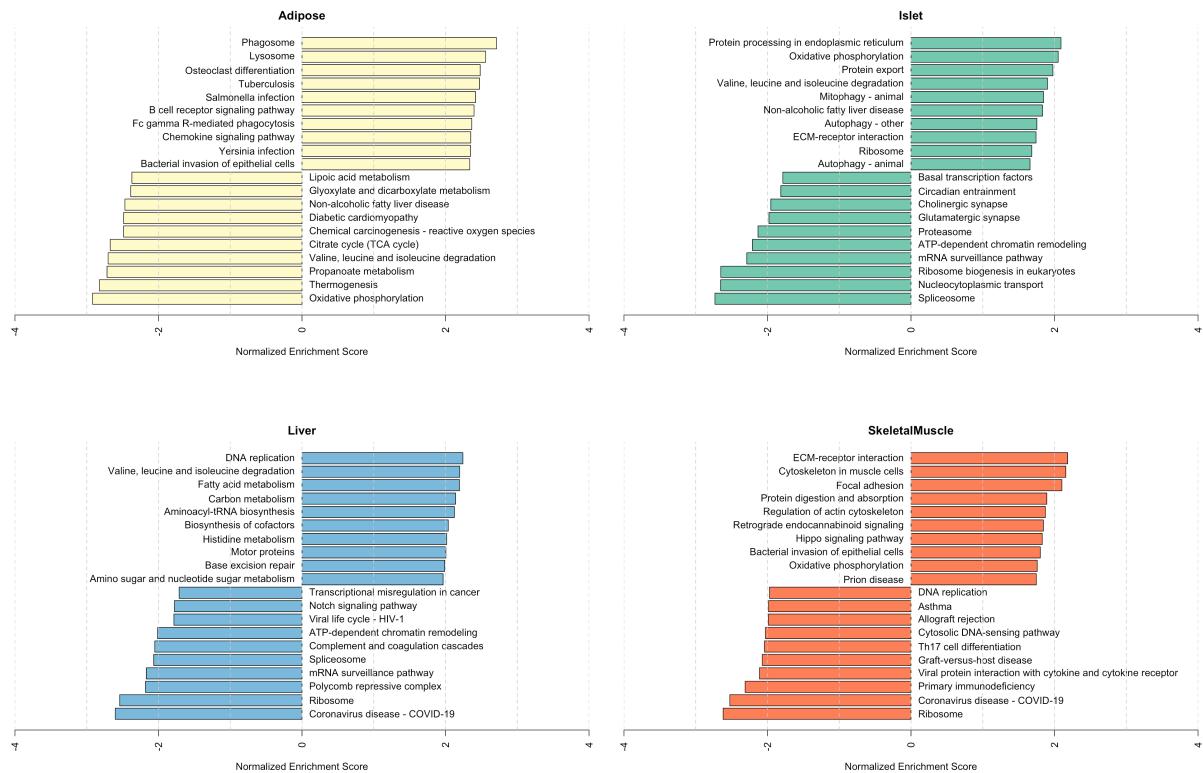


Figure 10: Bar plots showing normalized enrichment scores (NES) for KEGG pathways as determined by fast gene score enrichment analysis (fgsea). Only the top 10 positive and top 10 negative scores are shown. Colors indicate tissue. The name beside each bar shows the name of each enriched KEGG pathway.

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

## Top GO term enrichments by GSEA

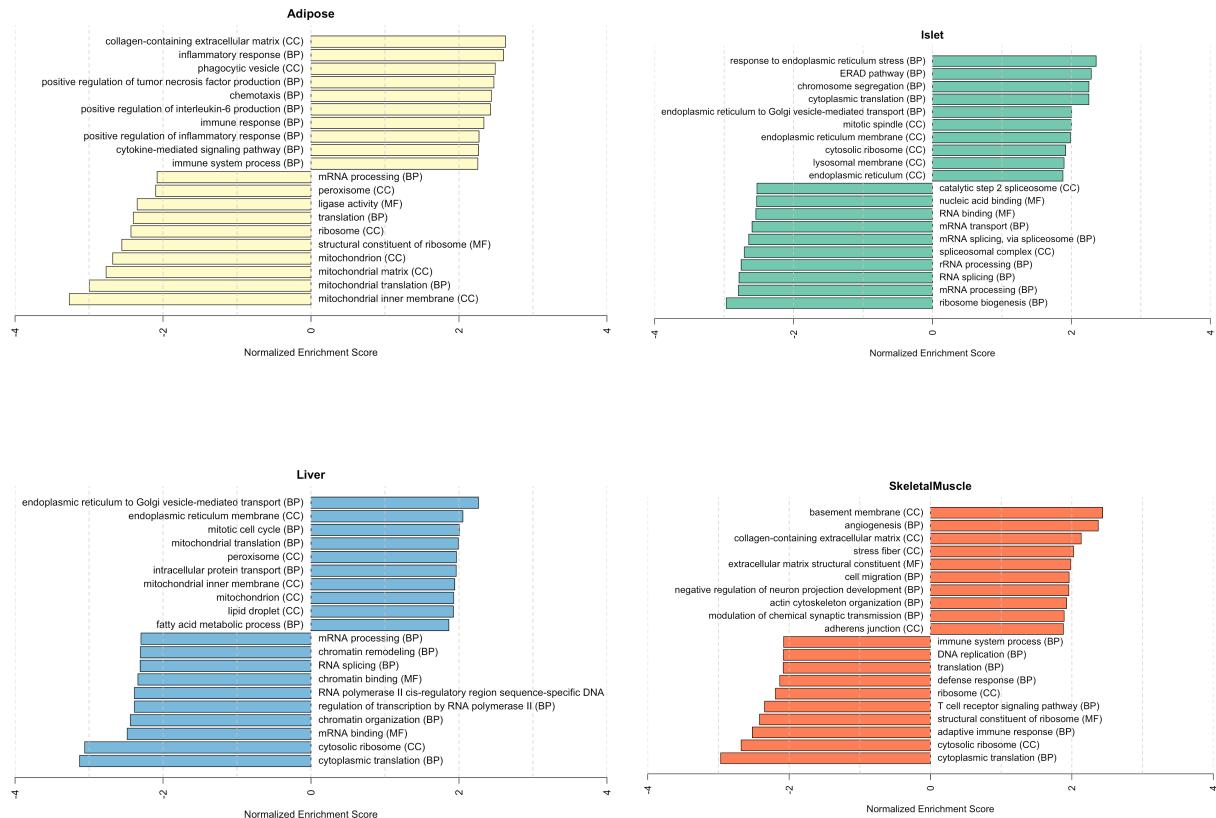


Figure 11: Bar plots showing normalized enrichment scores (NES) for GO terms as determined by fast gene score enrichment analysis (fgsea). Only the top 10 positive and top 10 negative scores are shown. Colors indicate tissue. The name beside each bar shows the name of each enriched GO term. The letters in parentheses indicate whether the term is from the biological process ontology (BP), the molecular function ontology (MF), or the cellular compartment ontology (CC).

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

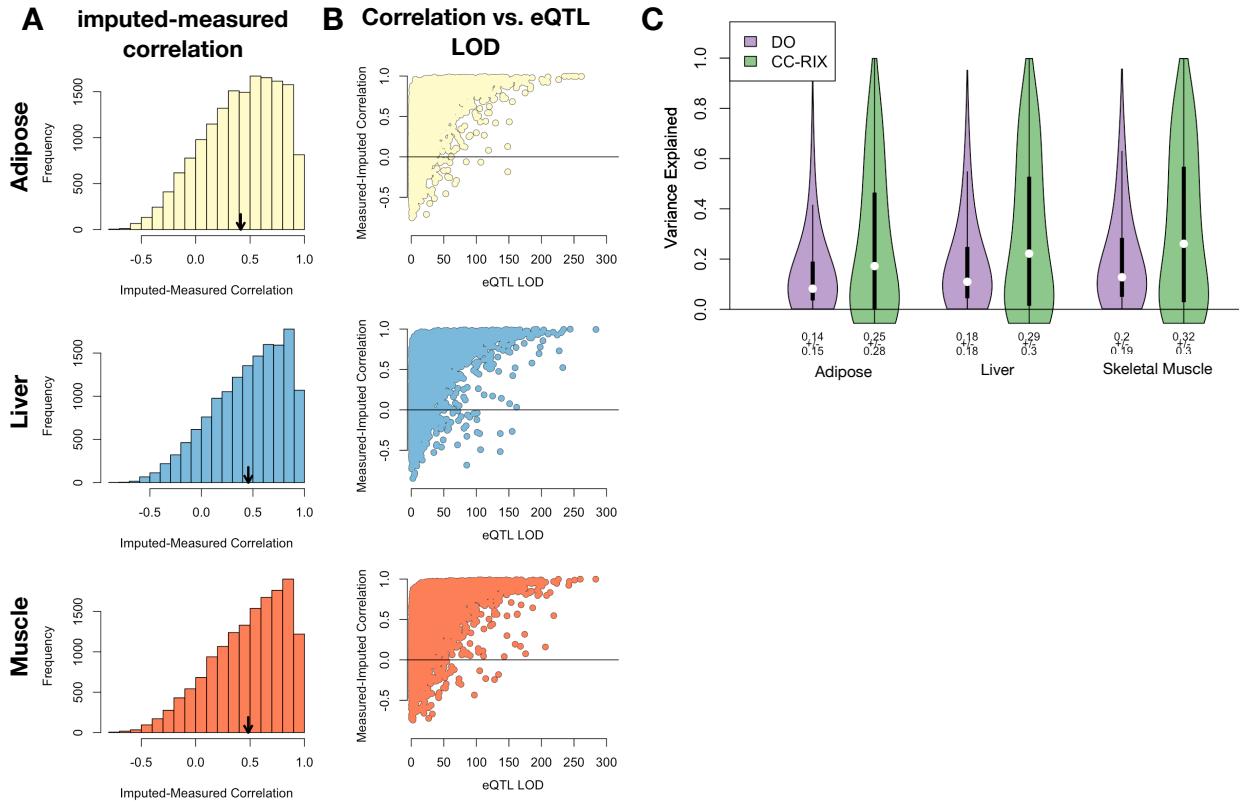


Figure 12: Validation of transcript imputation in the CC-RIX. **A.** Distributions of correlations between imputed and measured transcripts in the CC-RIX. The mean of each distribution is shown by the red line. All distributions were skewed toward positive correlations and had positive means near a Pearson 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.

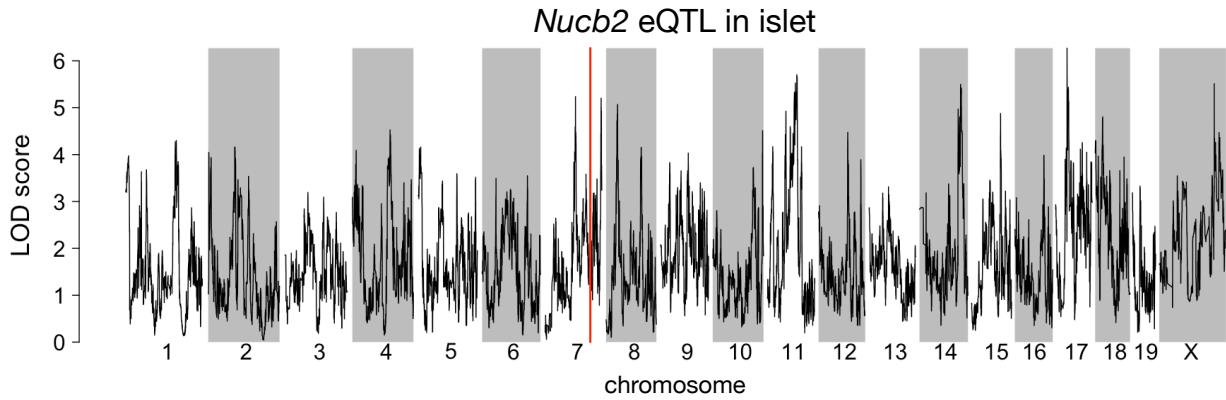


Figure 13: Regulation of *Nucb2* expression in islet. *Nucb2* is encoded on mouse chromosome 7 at 116.5 Mb (red line). In islets the heritability of *Nucb2* expression levels is 69% heritable. This LOD score trace shows that there is no local eQTL at that position, nor any strong distal eQTL anywhere else in the genome.

- 470 2016.
- 471 [8] D. Zhou, Y. Jiang, X. Zhong, N. J. Cox, C. Liu, and E. R. Gamazon. A unified framework for joint-tissue  
472 transcriptome-wide association and Mendelian randomization analysis. *Nat Genet*, 52(11):1239–1246,  
473 Nov 2020.
- 474 [9] E. R. Gamazon, H. E. Wheeler, K. P. Shah, S. V. Mozaffari, K. Aquino-Michaels, R. J. Carroll, A. E.  
475 Eyler, J. C. Denny, D. L. Nicolae, N. J. Cox, and H. K. Im. A gene-based association method for  
476 mapping traits using reference transcriptome data. *Nat Genet*, 47(9):1091–1098, Sep 2015.
- 477 [10] Z. Zhu, F. Zhang, H. Hu, A. Bakshi, M. R. Robinson, J. E. Powell, G. W. Montgomery, M. E. Goddard,  
478 N. R. Wray, P. M. Visscher, and J. Yang. Integration of summary data from GWAS and eQTL studies  
479 predicts complex trait gene targets. *Nat Genet*, 48(5):481–487, May 2016.
- 480 [11] A. Gusev, A. Ko, H. Shi, G. Bhatia, W. Chung, B. W. Penninx, R. Jansen, E. J. de Geus, D. I. Boomsma,  
481 F. A. Wright, P. F. Sullivan, E. Nikkola, M. Alvarez, M. Civelek, A. J. Lusis, T. ki, E. Raitoharju,  
482 M. nen, I. ä, O. T. Raitakari, J. Kuusisto, M. Laakso, A. L. Price, P. Pajukanta, and B. Pasaniuc.  
483 Integrative approaches for large-scale transcriptome-wide association studies. *Nat Genet*, 48(3):245–252,  
484 Mar 2016.
- 485 [12] M. P. Keller, D. M. Gatti, K. L. Schueler, M. E. Rabaglia, D. S. Stapleton, P. Simecek, M. Vincent,  
486 S. Allen, A. T. Broman, R. Bacher, C. Kendziorski, K. W. Broman, B. S. Yandell, G. A. Churchill, and  
487 A. D. Attie. Genetic Drivers of Pancreatic Islet Function. *Genetics*, 209(1):335–356, May 2018.
- 488 [13] W. L. Crouse, G. R. Keele, M. S. Gastonguay, G. A. Churchill, and W. Valdar. A Bayesian model  
489 selection approach to mediation analysis. *PLoS Genet*, 18(5):e1010184, May 2022.
- 490 [14] J. M. Chick, S. C. Munger, P. Simecek, E. L. Huttlin, K. Choi, D. M. Gatti, N. Raghupathy, K. L. Svenson,  
491 G. A. Churchill, and S. P. Gygi. Defining the consequences of genetic variation on a proteome-wide scale.  
492 *Nature*, 534(7608):500–505, Jun 2016.
- 493 [15] H. E. Wheeler, S. Ploch, A. N. Barbeira, R. Bonazzola, A. Andaleon, A. Fotuhi Siahpirani, A. Saha,  
494 A. Battle, S. Roy, and H. K. Im. Imputed gene associations identify replicable trans-acting genes enriched  
495 in transcription pathways and complex traits. *Genet Epidemiol*, 43(6):596–608, Sep 2019.
- 496 [16] B. D. Umans, A. Battle, and Y. Gilad. Where Are the Disease-Associated eQTLs? *Trends Genet*,  
497 37(2):109–124, Feb 2021.
- 498 [17] N. J. Connally, S. Nazeen, D. Lee, H. Shi, J. Stamatoyannopoulos, S. Chun, C. Cotsapas, C. A. Cassa,

- 499 and S. R. Sunyaev. The missing link between genetic association and regulatory function. *Elife*, 11, Dec  
500 2022.
- 501 [18] H. Mostafavi, J. P. Spence, S. Naqvi, and J. K. Pritchard. Systematic differences in discovery of genetic  
502 effects on gene expression and complex traits. *Nat Genet*, 55(11):1866–1875, Nov 2023.
- 503 [19] D. W. Yao, L. J. O’Connor, A. L. Price, and A. Gusev. Quantifying genetic effects on disease mediated  
504 by assayed gene expression levels. *Nat Genet*, 52(6):626–633, Jun 2020.
- 505 [20] X. Liu, J. A. Mefford, A. Dahl, Y. He, M. Subramaniam, A. Battle, A. L. Price, and N. Zaitlen. GBAT:  
506 a gene-based association test for robust detection of trans-gene regulation. *Genome Biol*, 21(1):211, Aug  
507 2020.
- 508 [21] G. A. Churchill, D. M. Gatti, S. C. Munger, and K. L. Svenson. The Diversity Outbred mouse population.  
509 *Mamm Genome*, 23(9-10):713–718, Oct 2012.
- 510 [22] Elissa J Chesler, Darla R Miller, Lisa R Branstetter, Leslie D Galloway, Barbara L Jackson, Vivek M  
511 Philip, Brynn H Voy, Cymbeline T Culiat, David W Threadgill, Robert W Williams, et al. The  
512 collaborative cross at oak ridge national laboratory: developing a powerful resource for systems genetics.  
513 *Mammalian Genome*, 19:382–389, 2008.
- 514 [23] Michael C Saul, Vivek M Philip, Laura G Reinholdt, and Elissa J Chesler. High-diversity mouse  
515 populations for complex traits. *Trends in Genetics*, 35(7):501–514, 2019.
- 516 [24] S. M. Clee and A. D. Attie. The genetic landscape of type 2 diabetes in mice. *Endocr Rev*, 28(1):48–83,  
517 Feb 2007.
- 518 [25] K. W. Broman, D. M. Gatti, P. Simecek, N. A. Furlotte, P. Prins, Š. Sen, B. S. Yandell, and G. A.  
519 Churchill. R/qtl2: Software for Mapping Quantitative Trait Loci with High-Dimensional Data and  
520 Multiparent Populations. *Genetics*, 211(2):495–502, Feb 2019.
- 521 [26] Fabien Girka, Etienne Camenen, Caroline Peltier, Arnaud Gloaguen, Vincent Guillemot, Laurent Le  
522 Brusquet, and Arthur Tenenhaus. *RGCCA: Regularized and Sparse Generalized Canonical Correlation  
523 Analysis for Multiblock Data*, 2023. R package version 3.0.3.
- 524 [27] Gennady Korotkevich, Vladimir Sukhov, and Alexey Sergushichev. Fast gene set enrichment analysis.  
525 *bioRxiv*, 2019.
- 526 [28] A. Subramanian, P. Tamayo, V. K. Mootha, S. Mukherjee, B. L. Ebert, M. A. Gillette, A. Paulovich,  
527 S. L. Pomeroy, T. R. Golub, E. S. Lander, and J. P. Mesirov. Gene set enrichment analysis: a

- knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci U S A*, 102(43):15545–15550, Oct 2005.
- [29] C. B. Newgard. Interplay between lipids and branched-chain amino acids in development of insulin resistance. *Cell Metab*, 15(5):606–614, May 2012.
- [30] D. D. Sears, G. Hsiao, A. Hsiao, J. G. Yu, C. H. Courtney, J. M. Ofrecio, J. Chapman, and S. Subramaniam. Mechanisms of human insulin resistance and thiazolidinedione-mediated insulin sensitization. *Proc Natl Acad Sci U S A*, 106(44):18745–18750, Nov 2009.
- [31] R. Stienstra, C. Duval, M. ller, and S. Kersten. PPARs, Obesity, and Inflammation. *PPAR Res*, 2007:95974, 2007.
- [32] O. Gavrilova, M. Haluzik, K. Matsusue, J. J. Cutson, L. Johnson, K. R. Dietz, C. J. Nicol, C. Vinson, F. J. Gonzalez, and M. L. Reitman. Liver peroxisome proliferator-activated receptor gamma contributes to hepatic steatosis, triglyceride clearance, and regulation of body fat mass. *J Biol Chem*, 278(36):34268–34276, Sep 2003.
- [33] K. Matsusue, M. Haluzik, G. Lambert, S. H. Yim, O. Gavrilova, J. M. Ward, B. Brewer, M. L. Reitman, and F. J. Gonzalez. Liver-specific disruption of PPARgamma in leptin-deficient mice improves fatty liver but aggravates diabetic phenotypes. *J Clin Invest*, 111(5):737–747, Mar 2003.
- [34] D. Patsouris, J. K. Reddy, M. ller, and S. Kersten. Peroxisome proliferator-activated receptor alpha mediates the effects of high-fat diet on hepatic gene expression. *Endocrinology*, 147(3):1508–1516, Mar 2006.
- [35] S. E. Schadinger, N. L. Bucher, B. M. Schreiber, and S. R. Farmer. PPARgamma2 regulates lipogenesis and lipid accumulation in steatotic hepatocytes. *Am J Physiol Endocrinol Metab*, 288(6):E1195–1205, Jun 2005.
- [36] W. Motomura, M. Inoue, T. Ohtake, N. Takahashi, M. Nagamine, S. Tanno, Y. Kohgo, and T. Okumura. Up-regulation of ADRP in fatty liver in human and liver steatosis in mice fed with high fat diet. *Biochem Biophys Res Commun*, 340(4):1111–1118, Feb 2006.
- [37] J. Lamb, E. D. Crawford, D. Peck, J. W. Modell, I. C. Blat, M. J. Wrobel, J. Lerner, J. P. Brunet, A. Subramanian, K. N. Ross, M. Reich, H. Hieronymus, G. Wei, S. A. Armstrong, S. J. Haggarty, P. A. Clemons, R. Wei, S. A. Carr, E. S. Lander, and T. R. Golub. The Connectivity Map: using gene-expression signatures to connect small molecules, genes, and disease. *Science*, 313(5795):1929–1935, Sep 2006.

- 558 [38] S. Amin, A. Lux, and F. O'Callaghan. The journey of metformin from glycaemic control to mTOR  
559 inhibition and the suppression of tumour growth. *Br J Clin Pharmacol*, 85(1):37–46, Jan 2019.
- 560 [39] A. Kezic, L. Popovic, and K. Lalic. mTOR Inhibitor Therapy and Metabolic Consequences: Where Do  
561 We Stand? *Oxid Med Cell Longev*, 2018:2640342, 2018.
- 562 [40] A. D. Barlow, M. L. Nicholson, and T. P. Herbert. -cells and a review of the underlying molecular  
563 mechanisms. *Diabetes*, 62(8):2674–2682, Aug 2013.
- 564 [41] Y. Gu, J. Lindner, A. Kumar, W. Yuan, and M. A. Magnuson. Rictor/mTORC2 is essential for  
565 maintaining a balance between beta-cell proliferation and cell size. *Diabetes*, 60(3):827–837, Mar 2011.
- 566 [42] E. B. Geer, J. Islam, and C. Buettner. Mechanisms of glucocorticoid-induced insulin resistance: focus on  
567 adipose tissue function and lipid metabolism. *Endocrinol Metab Clin North Am*, 43(1):75–102, Mar 2014.
- 568 [43] J. X. Li and C. L. Cummins. Fresh insights into glucocorticoid-induced diabetes mellitus and new  
569 therapeutic directions. *Nat Rev Endocrinol*, 18(9):540–557, Sep 2022.
- 570 [44] R. A. Lee, C. A. Harris, and J. C. Wang. Glucocorticoid Receptor and Adipocyte Biology. *Nucl Receptor*  
571 *Res*, 5, 2018.
- 572 [45] S. Viengchareun, P. Penfornis, M. C. Zennaro, and M. s. Mineralocorticoid and glucocorticoid receptors  
573 inhibit UCP expression and function in brown adipocytes. *Am J Physiol Endocrinol Metab*, 280(4):E640–  
574 649, Apr 2001.
- 575 [46] J. Liu, X. Kong, L. Wang, H. Qi, W. Di, X. Zhang, L. Wu, X. Chen, J. Yu, J. Zha, S. Lv, A. Zhang,  
576 P. Cheng, M. Hu, Y. Li, J. Bi, Y. Li, F. Hu, Y. Zhong, Y. Xu, and G. Ding. -HSD1 in regulating brown  
577 adipocyte function. *J Mol Endocrinol*, 50(1):103–113, Feb 2013.
- 578 [47] L. E. Ramage, M. Akyol, A. M. Fletcher, J. Forsythe, M. Nixon, R. N. Carter, E. J. van Beek,  
579 N. M. Morton, B. R. Walker, and R. H. Stimson. Glucocorticoids Acutely Increase Brown Adipose  
580 Tissue Activity in Humans, Revealing Species-Specific Differences in UCP-1 Regulation. *Cell Metab*,  
581 24(1):130–141, Jul 2016.
- 582 [48] J. L. Barclay, H. Agada, C. Jang, M. Ward, N. Wetzig, and K. K. Ho. Effects of glucocorticoids on  
583 human brown adipocytes. *J Endocrinol*, 224(2):139–147, Feb 2015.
- 584 [49] M. ó, R. Gupte, W. L. Kraus, P. Pacher, and P. Bai. PARPs in lipid metabolism and related diseases.  
585 *Prog Lipid Res*, 84:101117, Nov 2021.
- 586 [50] P. Bai, C. ó, H. Oudart, A. nszki, Y. Cen, C. Thomas, H. Yamamoto, A. Huber, B. Kiss, R. H.

- 587 Houtkooper, K. Schoonjans, V. Schreiber, A. A. Sauve, J. Menissier-de Murcia, and J. Auwerx. PARP-1  
588 inhibition increases mitochondrial metabolism through SIRT1 activation. *Cell Metab*, 13(4):461–468,  
589 Apr 2011.
- 590 [51] A. Chiarugi and M. A. Moskowitz. Cell biology. PARP-1—a perpetrator of apoptotic cell death? *Science*,  
591 297(5579):200–201, Jul 2002.
- 592 [52] M. Althubiti, R. Almaini, S. Y. Eid, M. Elzubaier, B. Refaat, S. Idris, T. A. Alqurashi, and M. Z.  
593 El-Readi. BTK targeting suppresses inflammatory genes and ameliorates insulin resistance. *Eur Cytokine  
594 Netw*, 31(4):168–179, Dec 2020.
- 595 [53] C. Skrabs, W. F. Pickl, T. Perkmann, U. ger, and A. Gessl. Rapid decline in insulin antibodies and  
596 glutamic acid decarboxylase autoantibodies with ibrutinib therapy of chronic lymphocytic leukaemia. *J  
597 Clin Pharm Ther*, 43(1):145–149, Feb 2018.
- 598 [54] E. A. Oral, S. M. Reilly, A. V. Gomez, R. Meral, L. Butz, N. Ajluni, T. L. Chenevert, E. Korytnaya,  
599 A. H. Neidert, R. Hench, D. Rus, J. F. Horowitz, B. Poirier, P. Zhao, K. Lehmann, M. Jain, R. Yu,  
600 C. Liddle, M. Ahmadian, M. Downes, R. M. Evans, and A. R. Saltiel. and TBK1 Improves Glucose  
601 Control in a Subset of Patients with Type 2 Diabetes. *Cell Metab*, 26(1):157–170, Jul 2017.
- 602 [55] M. C. Arkan, A. L. Hevener, F. R. Greten, S. Maeda, Z. W. Li, J. M. Long, A. Wynshaw-Boris, G. Poli,  
603 J. Olefsky, and M. Karin. IKK-beta links inflammation to obesity-induced insulin resistance. *Nat Med*,  
604 11(2):191–198, Feb 2005.
- 605 [56] M. Clark, C. J. Kroger, Q. Ke, and R. M. Tisch. The Role of T Cell Receptor Signaling in the  
606 Development of Type 1 Diabetes. *Front Immunol*, 11:615371, 2020.
- 607 [57] P. ndell, A. C. Carlsson, A. Larsson, O. Melander, T. Wessman, J. v, and T. Ruge. TNFR1 is associated  
608 with short-term mortality in patients with diabetes and acute dyspnea seeking care at the emergency  
609 department. *Acta Diabetol*, 57(10):1145–1150, Oct 2020.
- 610 [58] A. M. Keestra-Gounder, M. X. Byndloss, N. Seyffert, B. M. Young, A. vez Arroyo, A. Y. Tsai, S. A.  
611 Cevallos, M. G. Winter, O. H. Pham, C. R. Tiffany, M. F. de Jong, T. Kerrinnes, R. Ravindran, P. A.  
612 Luciw, S. J. McSorley, A. J. umler, and R. M. Tsolis. NOD1 and NOD2 signalling links ER stress with  
613 inflammation. *Nature*, 532(7599):394–397, Apr 2016.
- 614 [59] A. M. Keestra-Gounder and R. M. Tsolis. NOD1 and NOD2: Beyond Peptidoglycan Sensing. *Trends  
615 Immunol*, 38(10):758–767, Oct 2017.

- 616 [60] J. Montane, L. Cadavez, and A. Novials. Stress and the inflammatory process: a major cause of  
617 pancreatic cell death in type 2 diabetes. *Diabetes Metab Syndr Obes*, 7:25–34, 2014.
- 618 [61] B. B. Kahn and T. E. McGraw. , and type 2 diabetes. *N Engl J Med*, 363(27):2667–2669, Dec 2010.
- 619 [62] S. Del Prato and N. Pulizzi. The place of sulfonylureas in the therapy for type 2 diabetes mellitus.  
620 *Metabolism*, 55(5 Suppl 1):S20–27, May 2006.
- 621 [63] X. Liu, Y. I. Li, and J. K. Pritchard. Trans Effects on Gene Expression Can Drive Omnipotent Inheritance.  
622 *Cell*, 177(4):1022–1034, May 2019.
- 623 [64] B. Hallgrímsson, R. M. Green, D. C. Katz, J. L. Fish, F. P. Bernier, C. C. Roseman, N. M. Young,  
624 J. M. Cheverud, and R. S. Marcucio. The developmental-genetics of canalization. *Semin Cell Dev Biol*,  
625 88:67–79, Apr 2019.
- 626 [65] M. L. Siegal and A. Bergman. Waddington's canalization revisited: developmental stability and evolution.  
627 *Proc Natl Acad Sci U S A*, 99(16):10528–10532, Aug 2002.
- 628 [66] A. B. Paaby and G. Gibson. Cryptic Genetic Variation in Evolutionary Developmental Genetics. *Biology*  
629 (*Basel*), 5(2), Jun 2016.
- 630 [67] E. A. Boyle, Y. I. Li, and J. K. Pritchard. An Expanded View of Complex Traits: From Polygenic to  
631 Omnipotent. *Cell*, 169(7):1177–1186, Jun 2017.
- 632 [68] Naomi R Wray, Cisca Wijmenga, Patrick F Sullivan, Jian Yang, and Peter M Visscher. Common disease  
633 is more complex than implied by the core gene omnipotent model. *Cell*, 173(7):1573–1580, 2018.
- 634 [69] M. Riva, M. D. Nitert, U. Voss, R. Sathanoori, A. Lindqvist, C. Ling, and N. Wierup. Nesfatin-1  
635 stimulates glucagon and insulin secretion and beta cell NUCB2 is reduced in human type 2 diabetic  
636 subjects. *Cell Tissue Res*, 346(3):393–405, Dec 2011.
- 637 [70] M. Nakata and T. Yada. Role of NUCB2/nestin-1 in glucose control: diverse functions in islets,  
638 adipocytes and brain. *Curr Pharm Des*, 19(39):6960–6965, 2013.
- 639 [71] H. Shimizu and A. Osaki. Nesfatin/Nucleobindin-2 (NUCB2) and Glucose Homeostasis. *Curr Hypertens*  
640 *Rev*, pages Nesfatin/Nucleobindin-2 (NUCB2) and Glucose Homeostasis., Jul 2014.