

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

3

4 **Abstract**

5 The transcriptome is increasingly viewed as a bridge between genetic risk factors for complex disease and  
6 their associated pathophysiology. Powerful insights into disease mechanism can be made by linking genetic  
7 variants affecting gene expression (expression quantitative trait loci - eQTLs) to phenotypes.

8 **Introduction**

9 In the quest to understand the genetic architecture of complex traits, gene expression is an important bridge  
10 between genotype and phenotype. By identifying transcripts that mediate the effect of genetic variants on  
11 traits, we get one step closer to elaborating a complete molecular understanding of the genotype-phenotype  
12 map. Moreover, there is ample evidence from genome-wide association studies (GWAS) that regulation  
13 of gene expression accounts for the bulk of the genetic effect on complex traits, as most trait-associated  
14 variants lie in gene regulatory regions<sup>1–7</sup>. It is widely assumed that these variants influence local transcription,  
15 and methods such as transcriptome-wide association studies (TWAS)<sup>8–11</sup>, summary data-based Mendelian  
16 randomization (SMR)<sup>10</sup>, and others capitalize on this idea to identify genes associated with multiple disease  
17 traits<sup>12–15</sup>.

18 Despite the great promise of these methods, however, they have not been as widely successful as it seemed they  
19 could have been, and the vast majority of complex trait heritability remains unexplained [REF], implicating  
20 a subtler genetic architecture than assumed by popular methods. In fact, although trait-associated variants  
21 tend to lie in non-coding, regulatory regions, they often do not have detectable effects on gene expression<sup>16</sup>  
22 and tend not to co-localize with expression quantitative trait loci (eQTLs)<sup>17;18</sup>.

23 One possible explanation for these observations is that gene expression is not being measured in the appropriate  
24 cell types and thus true eQTLs influencing traits cannot be detected<sup>16</sup>. An alternative explanation that has  
25 been discussed in recent years is that effects of these variants are mediated not through local regulation of

26 gene expression, but through distal regulation<sup>18–20;15</sup>.

27 In this model, a gene's expression is influenced by potentially many variants throughout the genome through  
28 their cumulative effects on the regulatory network targeting that gene. In other words, the heritable  
29 component of the transcriptome is an emergent state arising from the myriad molecular interactions defining  
30 and constraining gene expression.

31 To assess the role of wide-spread distal gene regulation on complex traits, we studied diet-induced obesity  
32 and metabolic disease as an archetypal example. Diet-induced obesity and metabolic disease are genetically  
33 complex with hundreds of variants mapped through GWAS [REFS]. These variants are known to act through  
34 multiple tissues [REFS], including adipose tissue, pancreatic islets, liver, and skeletal muscle, that interact  
35 dynamically with each other. The multi-system etiology of metabolic disease complicates mechanistic  
36 dissection of the genetic architecture, requiring large, dedicated data sets that include high-dimensional,  
37 clinically relevant phenotyping, dense genotyping in a highly recombined population, and transcriptome-wide  
38 measurements of gene expression in multiple tissues. Measuring gene expression in multiple tissues is critical  
39 to adequately assess the extent to which local gene regulation varies across multiple tissues and whether such  
40 variability might account for previous failed attempts to identify trait-relevant local eQTL. Such data sets  
41 are extremely difficult to obtain in human populations, particularly in the large numbers of subjects required  
42 for statistical testing. Thus, to investigate further the role of local and distal gene regulation on complex  
43 traits, we have generated an appropriate data set in a large population of diversity outbred (DO) mice<sup>21</sup> in a  
44 population model of diet-induced obesity and metabolic disease<sup>12</sup>.

45 The DO mice were derived from eight inbred founder mouse strains, five classical lab strains, and three  
46 strains more recently derived from wild mice<sup>21</sup>. They represent three subspecies of mouse *Mus musculus*  
47 *domesticus*, *Mus musculus musculus*, and *Mus musculus castaneus*, and capture 90% of the known variation  
48 in laboratory mice<sup>22</sup>. They are maintained with a breeding scheme that ensures equal contributions from  
49 each founder across the genome thus rendering almost the whole genome visible to genetic inquiry<sup>21</sup>. We  
50 measured clinically relevant metabolic traits, including body weight, plasma levels of insulin and glucose,  
51 and plasma lipids in 500 DO mice. We further measured transcriptome-wide gene expression in four tissues  
52 related to metabolic disease: adipose tissue, pancreatic islets, liver, and skeletal muscle. Taken together,  
53 these data enable a comprehensive view into the genetic architecture of metabolic disease.

54 To assess the role of gene regulation in mediating variation in metabolic traits in this population, we propose  
55 a novel high-dimensional mediation analysis (HDMA) that uses the theory of causal modeling to explicitly  
56 model the emergent transcriptomic state mediating the distal effects driving trait-relevant gene expression. In

57 univariate approaches, such as TWAS, SMR, and other Mendelian randomization approaches, each transcript  
58 is tested independently for mediation of a local variant on a trait. This process requires huge numbers of  
59 statistical tests, which is computationally expensive, requires strict corrections for multiple testing, and  
60 assumes independence of genetic variants and transcripts. Such methods are therefore limited to detecting  
61 only the largest statistical effects and are biased toward local gene regulation. In contrast, HDMA assesses  
62 genome-, transcriptome-, and phenome-wide relationships as a whole and, in particular, identifies a highly  
63 heritable trait profile that is perfectly mediated by a transcriptomic signature. We showed these transcriptomic  
64 signatures are tissue-specific and highly interpretable in terms of biological processes as well as cell type  
65 composition. Heritability analysis of the transcripts showed that the strongest transcriptional mediators  
66 of metabolic disease had low local heritability and high distal heritability. Finally, we showed that the  
67 transcriptomic signatures identified in the DO population predicted obesity in an independent population of  
68 Collaborative Cross recombinant inbred (CC-RIX) mice and in human subjects. In contrast, local eQTL were  
69 unable to predict obesity in the CC-RIX mice. Together our results suggest that both the tissue used for gene  
70 expression analysis as well as distal gene regulation are critically important in identifying transcriptional  
71 mediators of the genome on complex traits.

## 72 Results

### 73 Genetic variation contributed to wide phenotypic variation

74 Although the environment was consistent across all animals, the genetic diversity present in this population  
75 resulted in widely varying distributions across physiological measurements (Fig. 1). For example, body  
76 weights of adult individuals varied from less than the average adult B6 body weight to several times the body  
77 weight of a B6 adult in both sexes (Fig. 1A). Fasting blood glucose (FBG) also varied considerably (Fig. 1B)  
78 although few of the animals had FBG levels that would indicate pre-diabetes (19 animals, 3.8%), or diabetes  
79 (7 animals, 1.4%) according to previously developed cutoffs (pre-diabetes:  $\text{FBG} \geq 250 \text{ mg/dL}$ , diabetes:  $\text{FBG}$   
80  $\geq 300 \text{ mg/dL}$ )<sup>23</sup>. Males had higher FBG than females on average (Fig. 1C) as has been observed before  
81 suggesting either that males were more susceptible to metabolic disease on the high-fat diet, or that males  
82 and females may require different thresholds for pre-diabetes and diabetes.

83 Body weight was strongly positively correlated with food consumption (Fig. 1D  $R^2 = 0.51, p = 1.5 \times 10^{-75}$ )  
84 and fasting blood glucose (FBG) (Fig. 1E,  $R^2 = 0.21, p = 1.4 \times 10^{-26}$ ) suggesting a link between behavioral  
85 factors and metabolic disease. However, the heritability of this trait and others (Fig. 1F) indicates that  
86 background genetics contribute substantially to correlates of metabolic disease in this population.

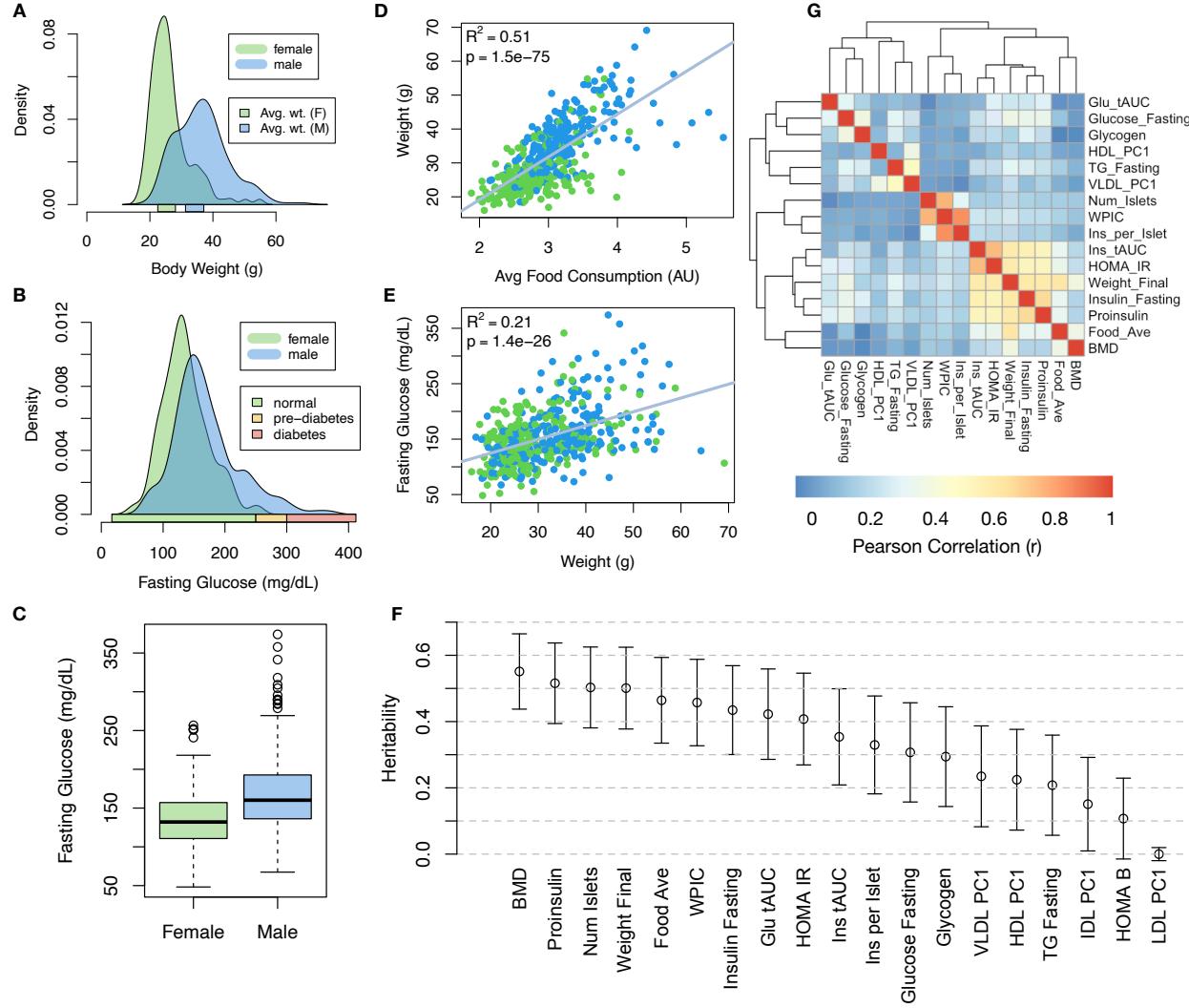


Figure 1: Clinical overview. **A.** Distributions of final body weight in the diversity outbred mice. Sex is indicated by color. The average B6 male and female adult weights at 24 weeks of age are indicated by blue and green bars on the x-axis. **B.** The distribution of final fasting glucose across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

87 The landscape of trait correlations (Fig. 1G) shows that most of the metabolic trait pairs were relatively  
 88 weakly correlated indicating complex relationships among the measured traits. This low level of redundancy  
 89 suggests a broad sampling of multiple heritable aspects of metabolic disease including overall body weight,  
 90 glucose homeostasis, pancreatic composition and liver function.

91 **Distal Heritability Correlated with Phenotype Relevance**

92 We performed eQTL analysis using R/qtl2<sup>24</sup> (Methods) and identified both local and distal eQTLs for  
 93 transcripts in each of the four tissues (Supp. Fig 9). Significant local eQTLs far outnumbered distal eQTLs  
 94 (Supp. Fig. 9F) and tended to be shared across tissues (Supp. Fig. 9G) whereas the few significant distal  
 95 eQTLs we identified tended to be tissue-specific (Supp. Fig. 9H)

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

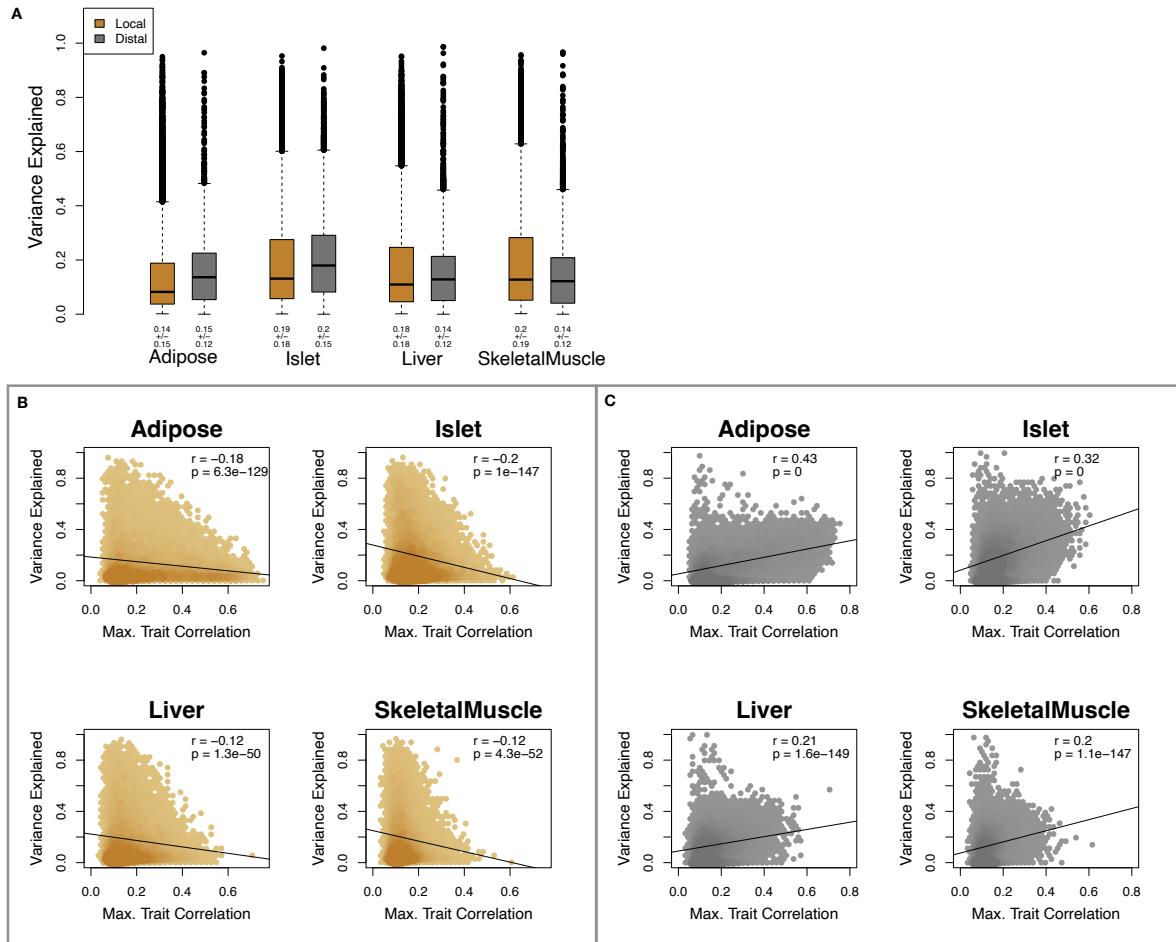


Figure 2: Transcript heritability and trait relevance. **A.** Distributions of distal and local heritability of transcripts across the four tissues. Overall local and distal factors contribute equally to transcript heritability. The relationship between **(B.)** local and **(C.)** distal heritability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

99 The local heritability of transcripts was negatively correlated with their trait relevance, defined as the  
100 maximum correlation of a transcript across all traits (Fig. 2B). This suggests that the more local genotype  
101 influenced transcript abundance, the less effect this variation had on the measured traits. Conversely, the  
102 distal heritability of transcripts was positively correlated with trait relevance (Fig. 2C). That is, transcripts  
103 that were more highly correlated with the measured traits tended to be distally, rather than locally, heritable.  
104 Importantly, this pattern was consistent across all tissues, strongly suggesting that this is a generic finding.  
105 This finding is consistent with previous observations that low-heritability transcripts explain more expression-  
106 mediated disease heritability than high-heritability transcripts<sup>19</sup>. However, the positive relationship between  
107 trait correlation and distal heritability demonstrated further that there are diffuse genetic effects throughout  
108 the genome converging on trait-related transcripts.

109 **High-Dimensional Mediation identified a high-heritability composite trait that was perfectly  
110 mediated by a composite transcript**

111 The above univariate analyses establish the importance of distal heritability for trait-relevant transcripts.  
112 However, the number of transcripts dramatically exceeds the number of degrees of freedom of the phenome.  
113 Thus, we expect the heritable, trait-relevant transcripts to be highly correlated and organized according  
114 to coherent, emergent biological processes representing the mediating endophenotypes driving clinical trait  
115 variation. To identify these endophenotypes in a theoretically principled way, we developed a novel dimension-  
116 reduction technique, HDMA, that uses the theory of causal graphical models to identify a transcriptomic  
117 signature that is simultaneously 1) highly heritable, 2) strongly correlated to the measured phenotypes, and  
118 3) conforms to the causal mediation hypothesis (Fig. 3). HDMA projects the high-dimensional scores—a  
119 composite genome score ( $G_C$ ), a composite transcriptome score ( $T_C$ ), and a composite phenome score  
120 ( $P_C$ )—and uses the univariate theory of mediation to constrain these projections to satisfy the hypotheses of  
121 perfect mediation. Specifically, perfect mediation implies that upon controlling for the transcriptomic score,  
122 the genome score is uncorrelated to the phenome score, which can also be viewed as a constraint on the  
123 correlation coefficients

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

124 which corresponds to the path coefficient in the mediation model [REF]. Operationally, HDMA is closely  
125 related to generalized canonical correlation analysis, for which provably convergent algorithms have recently  
126 been developed<sup>25</sup>. Implementation details for HDMA are available in **Supp. Methods XXX**.

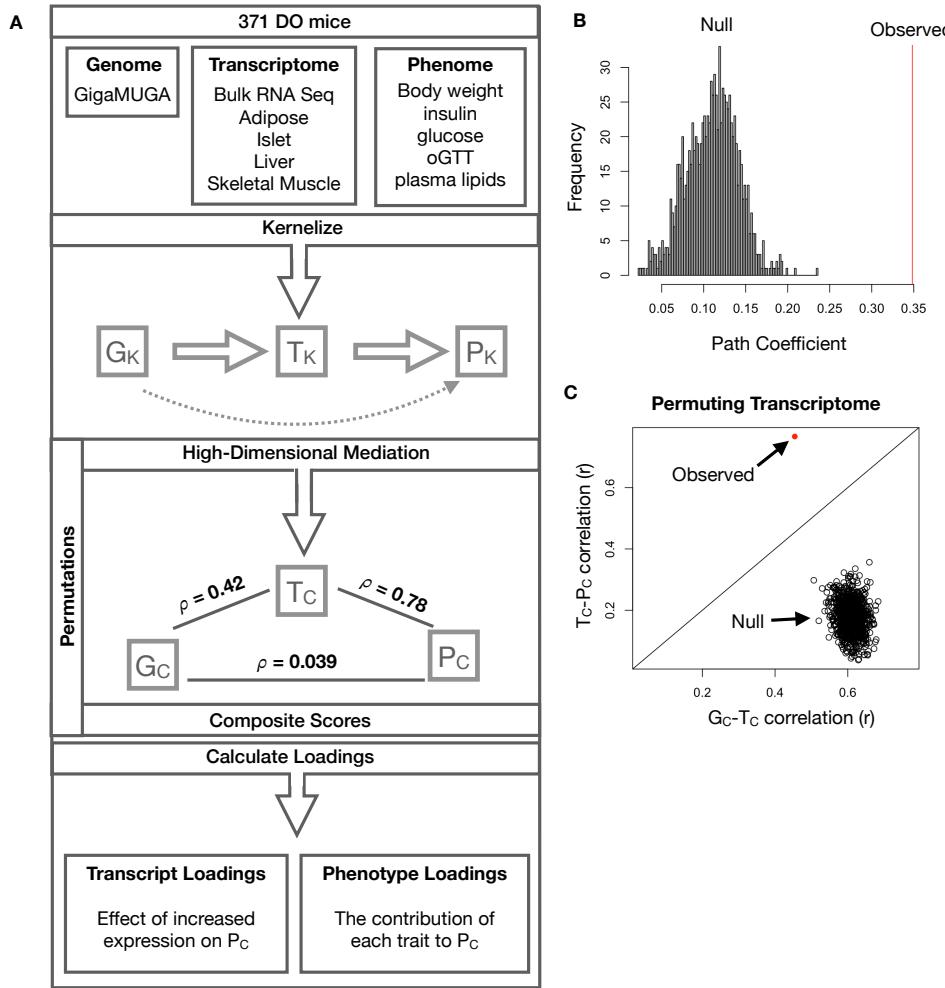


Figure 3: High-dimensional mediation. **A.** Workflow indicating major steps of high-dimensional mediation. The genotype, transcriptome, and phenotype matrices were kernelized to yield single matrices representing the relationships between all individuals for each data modality ( $G_K$  = genome kernel,  $T_K$  = transcriptome kernel;  $P_K$  = phenotype kernel). High-dimensional mediation was applied to these matrices to maximize the direct path  $G \rightarrow T \rightarrow P$ , the mediating pathway (arrows), while simultaneously minimizing the direct  $G \rightarrow P$  pathway (dotted line). The composite vectors that resulted from high-dimensional mediation were  $G_c$ ,  $T_c$ , and  $P_c$ . The partial correlations  $\rho$  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).

127 We used high-dimensional mediation to identify the major axis of variation in the transcriptome that mediated  
 128 the effects of the genome on metabolic traits (Fig 3). Fig. 3A shows the partial correlations ( $\rho$ ) between  
 129 the pairs of these composite vectors. The partial correlation between  $G_c$  and  $T_c$  was 0.42, and the partial  
 130 correlation between  $T_c$  and  $P_c$  was 0.78. However, when the transcriptome was taken into account, the  
 131 partial correlation between  $G_c$  and  $P_c$  was effectively 0 (0.039).  $P_c$  captured 30% of the overall trait variance,  
 132 and its estimated heritability was  $0.71 \pm 0.084\%$ , which was higher than any of the individual traits (Fig.

133 1F). Thus, HDMA identified a maximally heritable metabolic trait that was perfectly mediated by a highly  
134 heritable component of the transcriptome.

135 Standard CCA is prone to over-fitting because in any two large matrices it can be trivial to identify highly  
136 correlated composite vectors [REF]. To assess whether our implementation of HDMA was similarly prone to  
137 over-fitting in a high-dimensional space, we performed permutation testing. We permuted the individual  
138 labels on the transcriptome matrix 1000 times and recalculated the path coefficient, which is the partial  
139 correlation of  $G_C$  and  $T_C$  multiplied by the partial correlation of  $T_C$  and  $P_C$ . This represents the path  
140 from  $G_C$  to  $P_C$  that is mediated through  $T_C$ . The null distribution of the path coefficient is shown in Fig.  
141 3B, and the observed path coefficient from the original data is indicated by the red line. The observed  
142 path coefficient was well outside the null distribution generated by permutations ( $p < 10^{-16}$ ). Fig. 3C  
143 illustrates this observation in more detail. Although we identified high correlations between  $G_C$  and  $T_C$ , and  
144 modest correlations between  $T_C$  and  $P_C$  in the null data (Fig 3C), these two values could not be maximized  
145 simultaneously in the null data. In contrast, the red dot shows that in the real data both the  $G_C$ - $T_C$   
146 correlation and the  $T_C$ - $P_C$  correlation could be maximized simultaneously suggesting that the path from  
147 genotype to phenotype through transcriptome is highly non-trivial and identifiable in this case. These results  
148 suggest that these composite vectors represent genetically determined variation in phenotype that is mediated  
149 through genetically determined variation in transcription.

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

152 Each composite score is simply a weighted combination of the measured variables and the magnitude and  
153 sign of the weights, called loadings, correspond the relative importance and directionality of each variable in  
154 the composite score. The loadings of each measured trait onto  $P_C$  indicate how much each contributed to  
155 the composite phenotype. Final body weight contributed the most (Fig. 4), followed by homeostatic insulin  
156 resistance (HOMA\_IR) and fasting plasma insulin levels (Insulin\_Fasting). We can thus interpret  $P_C$  as  
157 an index of metabolic disease (Fig. 4B). Individuals with high values of  $P_C$  have a higher metabolic index  
158 and greater metabolic disease, including higher body weight and higher insulin resistance. We refer to  $P_C$  as  
159 the metabolic index going forward. Traits contributing the least to the metabolic index were measures of  
160 cholesterol and pancreas composition. Thus, when we interpret the transcriptomic signature identified by  
161 HDMA, we are explaining primarily the transcriptional mediation of body weight and insulin resistance, as  
162 opposed to cholesterol measurements.

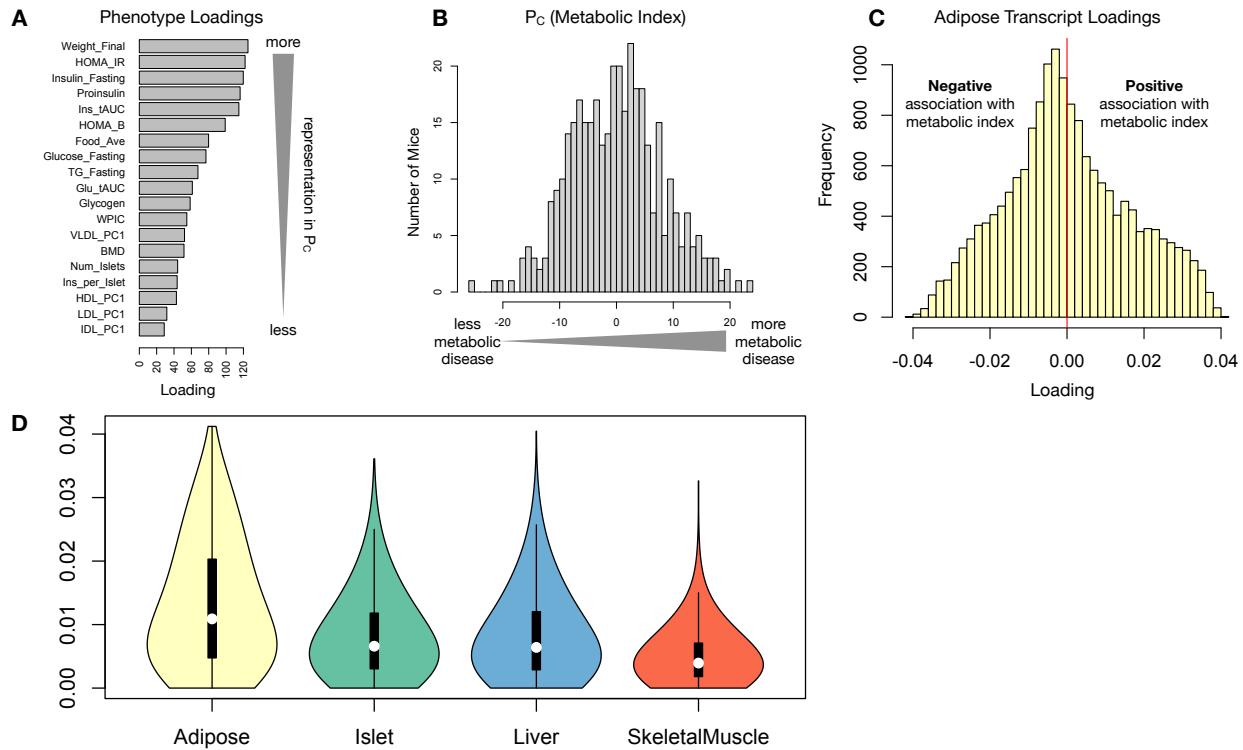


Figure 4: Interpretation of loadings. **A.** Loadings across traits. Body weight and insulin resistance contributed the most to the composite trait. **B.** Phenotype scores across individuals. Individuals with large positive phenotype scores had higher body weight and insulin resistance than average. 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 transcripts in adipose tissue were the best mediators of the genetic effects on body weight and insulin resistance.

- 163 **High-loading transcripts have low local heritability, high distal heritability, and were linked**  
 164 **mechanistically to obesity**
- 165 We interpreted large loadings onto transcripts as indicating strong mediation of the effect of genetics on  
 166 metabolic index. Large positive loadings indicate that higher expression was associated with a higher  
 167 metabolic index (i.e. higher risk of obesity and metabolic disease on the high-fat diet) (Fig. 4C). Conversely,  
 168 large negative loadings indicate that high expression of these transcripts was associated with a lower metabolic  
 169 index (i.e. lower risk of obesity and metabolic disease on the high-fat diet) (Fig. 4C). We used gene set  
 170 enrichment analysis (GSEA)<sup>26;27</sup> to look for biological processes and pathways that were enriched at the top  
 171 and bottom of this list (Methods).
- 172 In adipose tissue, both GO processes and KEGG pathway enrichments pointed to an axis of inflammation

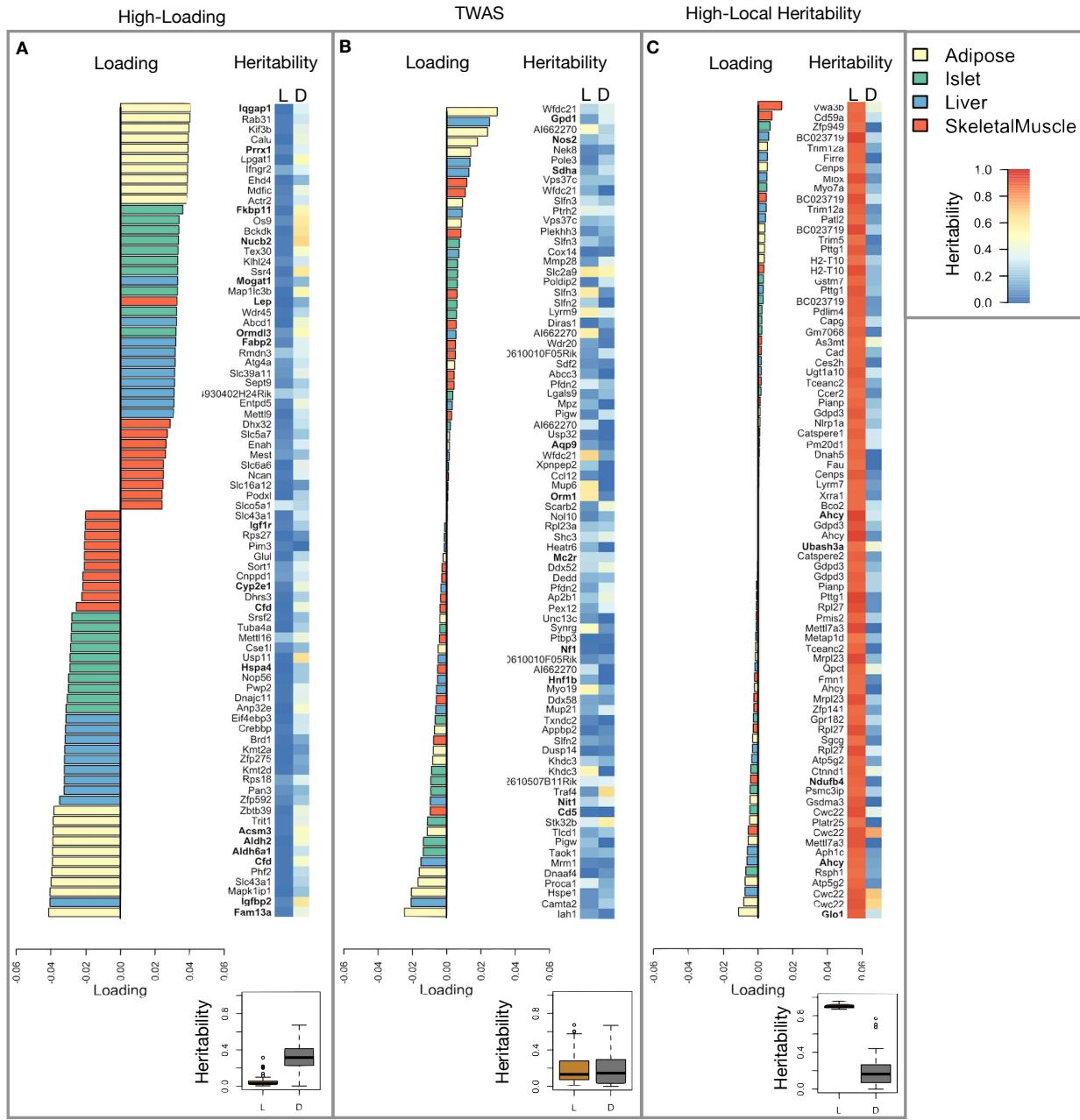
173 and metabolism (Supp. Fig. 10 and 11). GO terms and KEGG pathways associated with inflammation,  
174 particularly macrophage infiltration, were positively associated with metabolic index, indicating that increased  
175 expression in inflammatory pathways was associated with a higher metabolic index. It is well established  
176 that adipose tissue in obese individuals is highly inflamed [cite] and infiltrated by macrophages [cite], and the  
177 results here suggest that this may be a heritable component of metabolic disease.

178 The strongest negative enrichments in adipose tissue were related to mitochondrial activity in general, and  
179 thermogenesis in particular (Supp. Fig. 10 and 11). It has been shown mouse strains with greater thermogenic  
180 potential are also less susceptible to obesity on a high-fat diet [cite].

181 Transcripts associated with the citric acid (TCA) cycle as well as the catabolism of the branched-chain amino  
182 acids (BCAA) (valine, leucine, and isoleucine) were strongly enriched with negative loadings in adipose  
183 tissue (Supp. Fig. XXX). Expression of genes in both pathways (for which there is some overlap) has been  
184 previously associated with insulin sensitivity<sup>12;28;29</sup>, suggesting that heritable variation in regulation of these  
185 pathways may influence risk of insulin resistance.

186 Looking at the 10 most positively and negatively loaded transcripts from each tissue, it is apparent that  
187 transcripts in the adipose tissue had the largest loadings, both positive and negative, of all tissues (Fig. 5A  
188 bar plot). This suggests that much of the effect of genetics on body weight and insulin resistance is mediated  
189 through gene expression in adipose tissue. The strongest loadings in liver and pancreas were comparable,  
190 and those in skeletal muscle were the weakest (Fig. 5A), suggesting that less of the genetic effects were  
191 mediated through transcription in skeletal muscle. Heritability analysis showed that transcripts with the  
192 largest loadings tended to have relatively high distal heritability compared with local heritability (Fig. 5A  
193 heat map and box plot). This pattern contrasts with transcripts nominated by TWAS (Fig. 5B), which  
194 tended to have lower loadings, higher local heritability and lower distal heritability. Transcripts with the  
195 highest local heritability in each tissue (Fig. 5C) had the lowest loadings.

196 We performed a literature search for the genes in each of these groups along with the terms “diabetes”,  
197 “obesity”, and the name of the expressing tissue to determine whether any of these genes had previous  
198 associations with metabolic disease in the literature (Methods). Multiple genes in each group had been  
199 previously associated with obesity and diabetes (Fig. 5 bolded gene names). Genes with high loadings were  
200 most highly enriched for previous literature support. They were 2.25 more likely than TWAS hits and 3.6  
201 times more likely than genes with high local heritability to be previously associated with obesity or diabetes.



activation particularly in adipose tissue. Except for the “mitosis” cluster, which had large positive loadings in three of the four tissues, all clusters were strongly loaded in only one or two tissues. For example, the lipid metabolism cluster was loaded most heavily in liver. The positive loadings suggest that high expression of these genes particularly in the liver was associated with increased metabolic disease. This cluster included the gene *Pparg*, whose primary role is in the adipose tissue where it is considered a master regulator of adipogenesis<sup>30</sup>. Agonists of *Pparg*, such as thiazolidinediones, which are FDA-approved to treat type II diabetes, reduce inflammation and adipose hypertrophy<sup>30</sup>. Consistent with this role, the loading for *Pparg* in adipose tissue was negative, suggesting that higher expression was associated with leaner mice (Fig. 6B). In contrast, *Pparg* had a large positive loading in liver, where it is known to play a role in the development of hepatic steatosis, or fatty liver. Mice that lack *Pparg* specifically in the liver, are protected from developing steatosis and show reduced expression of lipogenic genes<sup>31;32</sup>. Overexpression of *Pparg* in the livers of mice with a *Ppara* knockout, causes upregulation of genes involved in adipogenesis<sup>33</sup>. In the livers of both mice and humans high *Pparg* expression is associated with hepatocytes that accumulate large lipid droplets and have gene expression profiles similar to adipocytes<sup>34;35</sup>.

The local and distal heritability of *Pparg* is low in adipose tissue suggesting its expression in this tissue is highly constrained in the population (Fig. 6B). However, the distal heritability of *Pparg* in liver is relatively high suggesting it is complexly regulated and has sufficient variation in this population to drive variation in phenotype. Both local and distal heritability of *Pparg* in the islet are relatively high, but the loading is low, suggesting that variability of expression in the islet does not drive variation in metabolic index. These results highlight the importance of tissue context when investigating the role of heritable transcript variability in driving phenotype.

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

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

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

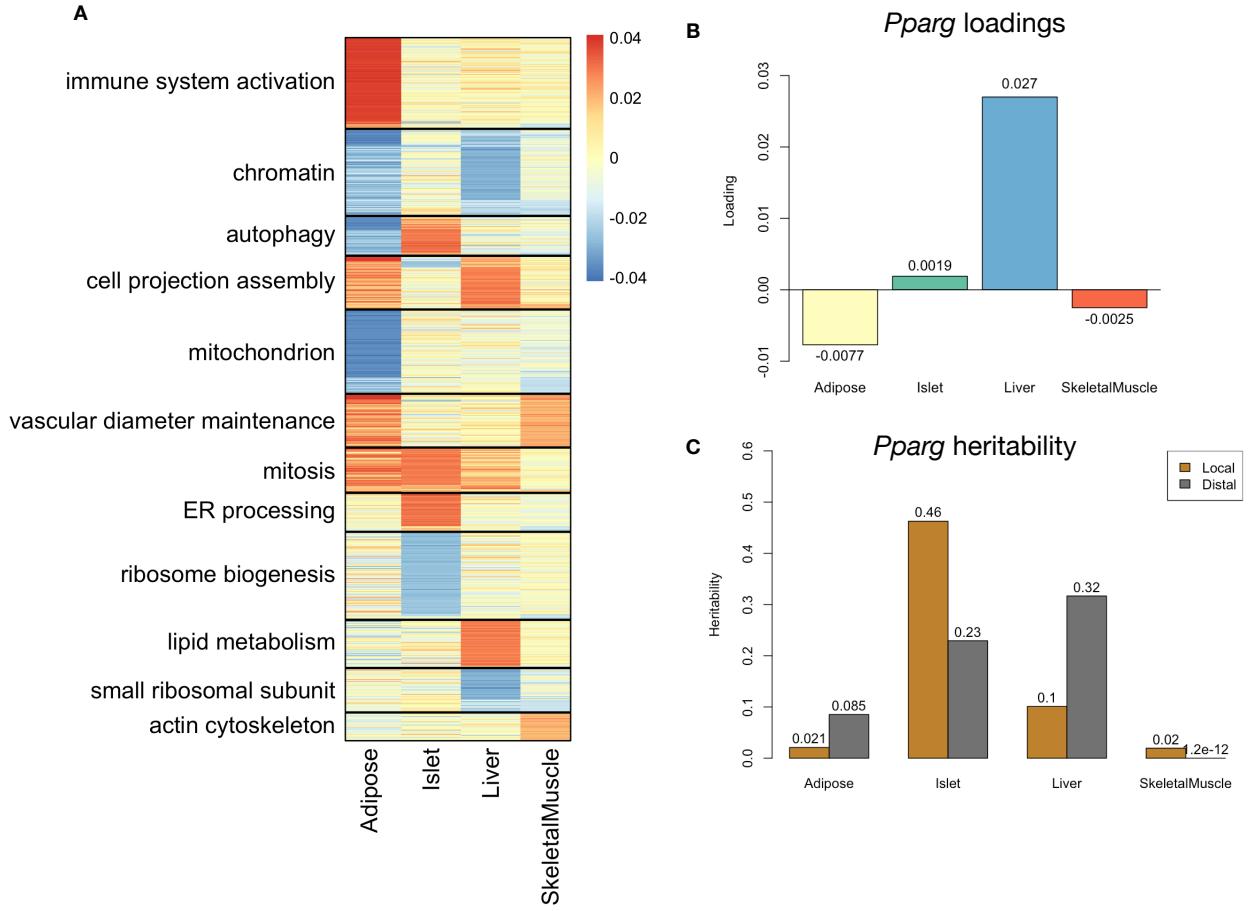


Figure 6: Tissue-specific transcriptional programs were associated with obesity and insulin resistance. **A** Heat map showing the loadings of all transcripts with loadings greater than 2.5 standard deviations from the mean in any tissue. The heat map was clustered using k medoid clustering. Functional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

tissue, which supports the observation in the DO that adipose expression was the strongest mediator of the genetic effect on metabolic index. This result also confirms the validity and translatability of the transcript loadings and their relationship to metabolic disease.

The second question related to the source of the relevant variation in gene expression. If local regulation was the predominant factor influencing gene expression, we should be able to predict phenotype in the CC-RIX using transcripts imputed from local genotype (Fig. 7A). The DO and the CC-RIX were derived from the same eight founder strains and so carry the same alleles throughout the genome. We imputed gene expression in the CC-RIX using local genotype and were able to estimate variation in gene transcription robustly (Supp. Fig. 12). However, these imputed values failed to predict body weight in the CC-RIX when weighted with the loadings from HDMA. (Fig. 7B right column). This result suggests that local regulation of gene expression is

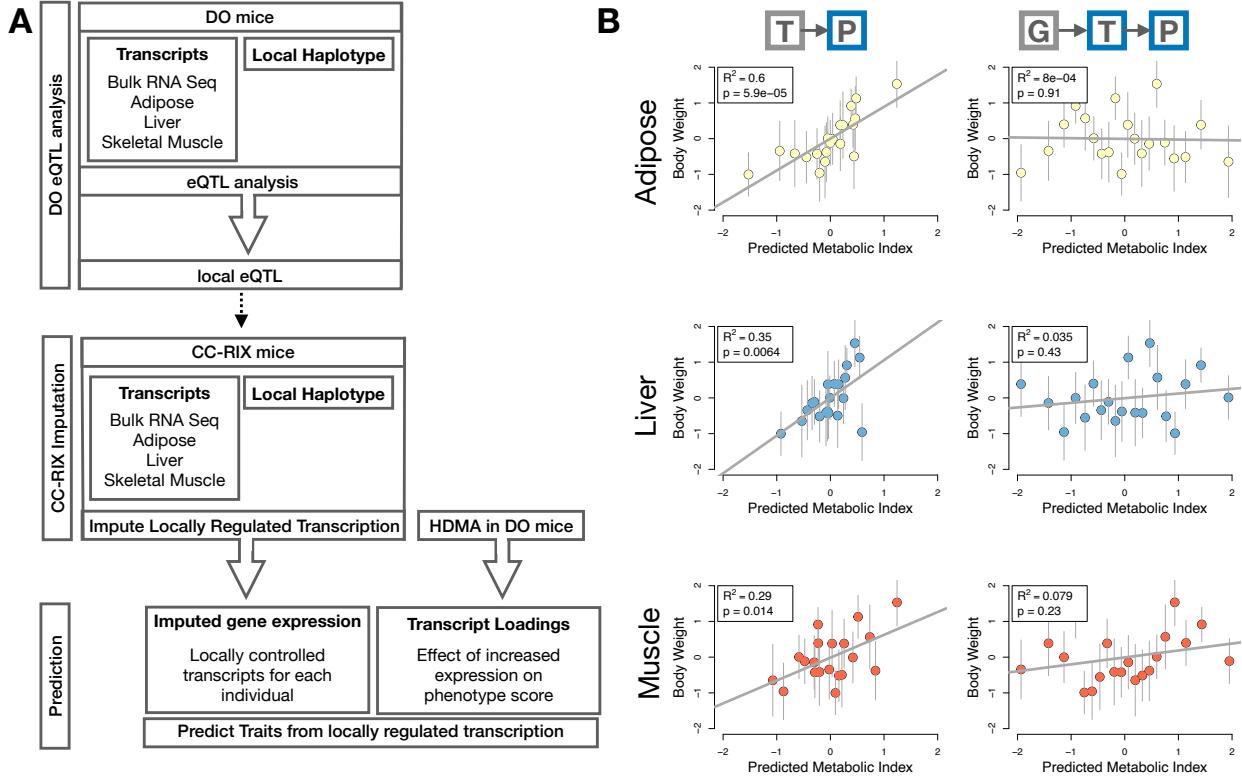


Figure 7: Transcription, but not local genotype, predicts phenotype in the CC-RIX. **A.** Workflow showing procedure for translating HDMA results to an independent population of mice. **B.** Relationships between the predicted metabolic index and measured body weight. The left column shows the predictions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

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

249 **Distally heritable transcriptomic signatures reflected variation in composition of adipose tissue  
 250 and islets**

251 The interpretation of global genetic influences on gene expression and phenotype is potentially more challenging  
 252 than the interpretation and translation of local genetic influences, as genetic effects cannot be localized to  
 253 individual gene variants or transcripts. However, there are global patterns across the loadings that can  
 254 inform mechanism. For example, heritable variation in cell type composition can be derived from transcript  
 255 loadings. We observed above that immune activation in the adipose tissues was an important driver of obesity  
 256 in the DO population. To determine whether this is reflected as an increase in macrophages in adipose

tissue, we compared loadings of cell-type specific genes in adipose tissue (Methods). The mean loading of macrophage-specific genes was significantly greater than 0 (Fig. 8A), indicating that obese mice were genetically predisposed to have high levels of macrophage infiltration in adipose tissue in response to the high-fat, high-sugar diet.

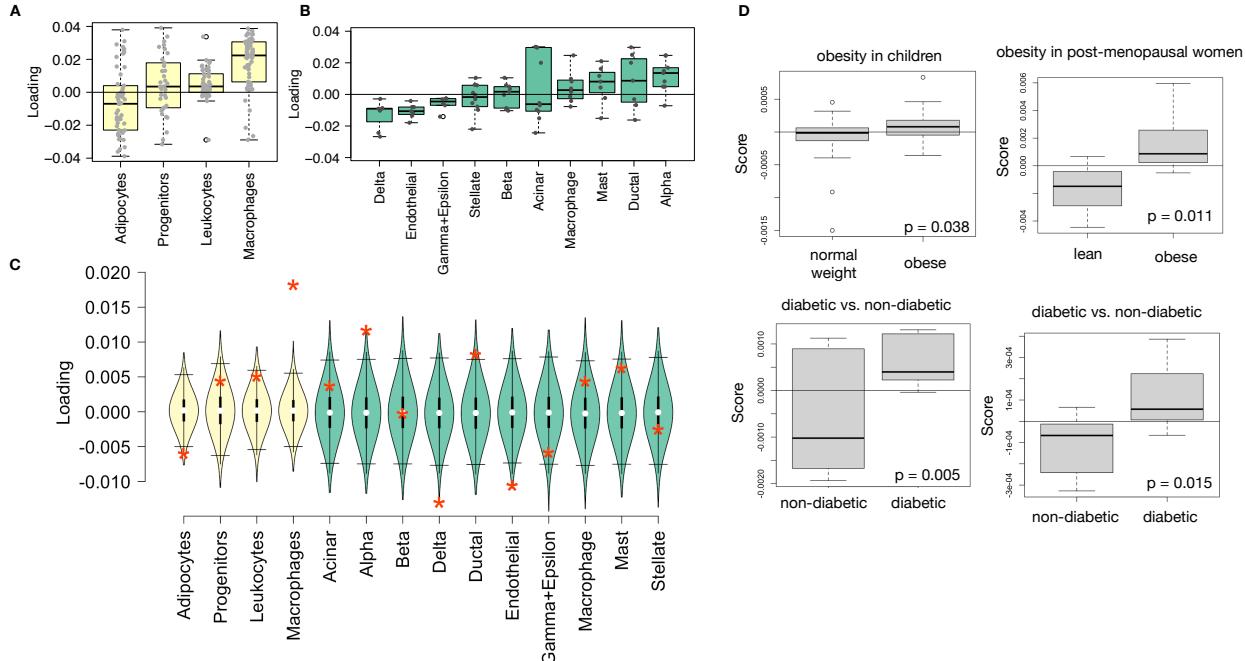


Figure 8: HDMA results translate to humans. **A.** Distribution of loadings for cell-type-specific transcripts in adipose tissue. **B.** Distribution of loadings for cell-type-specific transcripts in pancreatic islets (green). **C.** Null distributions for the mean loading of randomly selected transcripts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

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

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

271 is further consistent with the islet composition traits having small loadings in the genome score (Fig. 4).

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

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

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

284 **Targeting gene signatures**

285 Another global view of the transcript loading landscape is in ranking potential drug candidates for the  
286 treatment of metabolic disease. Although high-loading transcripts may be good candidates for understanding  
287 specific biology related to obesity, the transcriptome overall is highly interconnected and redundant, and  
288 focusing on individual transcripts for treatment may be less effective than using broader transcriptomic  
289 signatures that capture the emergent biology. The ConnectivityMap (CMAP) database [cite] developed by  
290 the Broad Institute allows us to query thousands of compounds that reverse or enhance the extreme ends  
291 of transcriptomic signatures in multiple different cell types. By identifying drugs that reverse pathogenic  
292 transcriptomic signatures, we can potentially identify compounds that have favorable effects on gene expression.

293 To test this hypothesis, we queried the CMAP database through the CLUE online query tool [cite] (Methods).  
294 We identified top anti-correlated hits both across all cell types, as well as in adipocytes and pancreatic tumor  
295 cells (Supplemental Figure XXX and XXX).

296 Looking broadly across cell types, the notable top hits from the adipose tissue loadings included mTOR  
297 inhibitors and glucocorticoid agonists (Supplemental Figure XXX). It is thought that metformin, which  
298 is commonly used to improve glycemic control, acts, at least in part, by inhibiting mTOR signaling<sup>36;37</sup>.  
299 However, long-term use of other mTOR inhibitors, such as rapamycin, are known to cause insulin resistance  
300 and  $\beta$ -cell toxicity<sup>37-39</sup>. Glucocorticoids are used to reduce inflammation, which was a prominent signature

301 in the adipose tissues, but these drugs also promote hyperglycemia and diabetes<sup>40;41</sup>. Accute treatment  
302 with glucocorticoids has further been shown to reduce thermogenesis in rodent adipocytes<sup>42–44</sup>, but increase  
303 thermogenesis in human adipocytes<sup>45;46</sup>. Thus, the pathways identified by CMAP across all cell types were  
304 highly related to the transcript loading profiles, but the relationship was not a simple reversal.

305 The top hit in adipocytes was a PARP inhibitor (Supplemental Figure XXXB). PARPs play a role in lipid  
306 metabolism and are involved in the development of obesity and diabetes<sup>47</sup>. PARP1 inhibition increases  
307 mitochondrial biogenesis<sup>48</sup>. Inhibition of PARP1 activity can further prevent necrosis in favor of the less  
308 inflammatory apoptosis<sup>49</sup>, thereby potentially reducing inflammation in stressed adipocytes. Other notable  
309 hits in the top 20 were BTK inhibitors, which have been observed to suppress inflammation and improve  
310 insulin resistance<sup>50</sup> as well as to reduce insulin antibodies in type I diabetes<sup>51</sup>, and IKK inhibitors have been  
311 shown to improve glucose control in type II diabetes<sup>52;53</sup>.

312 The CMAP database includes assays in multiple cell types. Among the top hits for the query with transcript  
313 loadings from pancreatic islets (Fig. XXX), was suppression of T cell receptor signaling, which is known to  
314 be involved in Type 1 diabetes<sup>54</sup>, as well as TNFR1, which has been associated with mortality in diabetes  
315 patients<sup>55</sup>. Suppression of NOD1/2 signaling was also among the top hits. NOD1 and 2 sense ER stress<sup>56;57</sup>,  
316 which is associated with  $\beta$ -cell death in type 1 and type 2 diabetes<sup>58</sup>. This cell death process is dependent  
317 on NOD1/2 signaling<sup>56</sup>, although the specifics have not yet been worked out.

318 Among the top hits in pancreatic tumor cells were known diabetes drugs, including sulfonylureas, PPAR  
319 receptor agonists, and insulin sensitizers. Rosiglitazone is a PPAR- $\gamma$  agonist and was one of the most  
320 prescribed drugs for type 2 diabetes before its use was reduced due to cardiac side-effects<sup>59</sup>. Sulfonylureas  
321 are another commonly prescribed drug class for type 2 diabetes, but also have notable side effects including  
322 hypoglycemia and accellerated  $\beta$ -cell death<sup>60</sup>.

## 323 Discussion

324 It is thought that the bulk of the effect of genomic variation on complex traits is mediated through regulation  
325 of gene expression. It has widely been assumed that this regulation is largely in *cis*, but attempts to use  
326 local gene regulation to explain phenotypic variation have yet to explain much trait heritability. In recent  
327 years, the discussion has turned to distal gene regulation. Although, distal gene regulation is more complex  
328 to identify, evidence suggests that it is an important component of trait heritability.

329 Yao *et al.*<sup>19</sup> observed that in humans, transcripts with low local heritability explained more expression-  
330 mediated disease heritability than transcripts with high local heritability. We observed the same trend here

331 in mice in a well-powered study of all major tissues of action simultaneously. This pattern is consistent  
332 with principles of robustness in complex systems<sup>61–63</sup>. If a transcript were both important to a trait and  
333 subject to strong local regulation, a population would be susceptible to extremes in phenotype that might  
334 frequently cross the threshold to disease. Indeed, strong disruption of highly trait-relevant genes is the cause  
335 of Mendelian disease.

336 Rather, studies suggest that genes that are near GWAS hits and have obvious functional relevance to a trait  
337 tend to have highly complex regulatory landscapes under strong selection pressures<sup>18</sup>. In contrast, genes  
338 with strong local regulation tend to be depleted of functional annotations and are under looser selection  
339 constraints<sup>18</sup>. These observations and others led Liu et al.<sup>64</sup> to suggest that most heritability of complex  
340 traits is driven by weak distal eQTLs. They proposed a framework of understanding heritability of complex  
341 traits in which massive polygenicity is distributed across common variants in both functional “core genes”, as  
342 well as more peripheral genes that may not seem obviously related to the trait.

343 Here, we used a large, comprehensive, and purpose-built data set to investigate the genetic architecture of  
344 complex traits related to metabolic disease in mice as well as the roles of local and distal gene regulation  
345 in mediating these traits. We presented a systems-level method called high-dimension mediation analysis  
346 (HDMA) to specifically identify the distally regulated transcriptomic signature mediating the effect of the  
347 genotype on phenotype. This approach contrasts with traditional univariate approaches in several important  
348 respects. First, in contrast to univariate approaches, which assume independence of genetic variants and  
349 transcripts, HDMA allows for arbitrarily complex gene regulation, as well as the interconnectedness and  
350 redundancy of the transcriptome. Second, rather than assuming a single, large genetic effect as univariate  
351 approaches do, HDMA assumes that traits are highly polygenic, and that genetic effects are weak and are  
352 distributed across the genome. HDMA does not use statistical thresholds to identify true positive effects, but  
353 generates a weighted vector of transcripts that can be analyzed as a whole, or dissected to identify transcripts  
354 with stronger and weaker effects. This method explicitly models a central proposal of the omnigenic model  
355 which posits that once the expression of the core genes (i.e. trait-mediating genes) is accounted for, there  
356 should be no residual correlation between the genome and the phenome.

357 Using HDMA, we identified a highly heritable composite trait (0.71% heritable) that explained 30% of overall  
358 trait variance, and was perfectly mediated by a composite transcript that included expression from four tissues  
359 known to be involved in metabolic disease. Gene expression in adipose tissue was the strongest mediator of  
360 genetic effects on metabolic disease. Further analysis of the loadings onto transcripts in each tissue revealed  
361 that the mediating signatures were tissue-specific transcriptional programs, many of which were previously  
362 known to be involved in the pathogenesis of metabolic disease. We showed here that regulation of these

363 programs is heritable and mediated a large proportion of disease risk.

364 The transcripts with the highest loadings are similar to the core genes of the omnigenic model. These were  
365 transcripts of moderate local heritability that were highly functionally related to the traits. Transcripts with  
366 small loadings are more peripheral to the traits measured in this experiment. There was no clear demarcation  
367 between the core and peripheral genes as far as loading, but a clear separation should not be expected given  
368 the complexity of gene regulation and the genotype-phenotype map<sup>65</sup>.

369 The strength of mediation (transcript loading) was negatively correlated with local heritability and positively  
370 correlated with distal heritability, suggesting that distal gene regulation was the dominant mode through which  
371 gene expression mediated the effect of genotype on phenotype. We saw further that the distal heritability  
372 was weak and spread across the genome, consistent with the prediction by Liu *et al.*<sup>64</sup> that trait heritability  
373 is mediated through weak distal eQTLs. Most strongly mediating transcripts had modest distal heritability,  
374 and even for those whose expression was strongly regulated by distal factors, these factors were multiple  
375 and spread across the genome. For example, *Nucb2*, was a strongly mediating transcript in islet and was  
376 also strongly distally regulated (66% distal heritability). This gene is expressed in pancreatic  $\beta$  cells and is  
377 involved in insulin and glucagon release<sup>66–68</sup>. Although its transcription was highly heritable in islets, that  
378 regulation was distributed across the genome, with no clear distal eQTL (Supp. Fig. 13). Thus, although  
379 distal regulation of some genes may be strong, this regulation is likely to be highly complex and not easily  
380 localized.

381 The high complexity of gene regulation combined with a systems-level analysis yields continuous results  
382 that do not necessarily implicate individual transcripts or genetic loci in disease pathogenesis. Most studies  
383 have focused on pinpointing individual loci whose mechanistic roles can be clearly dissected through further  
384 experiments and exploited as therapeutic targets. In this analysis, too, it is possible to focus on individual  
385 genes and their context in both tissues and pathways. For example, the loadings on *Pparg* were tissue-specific  
386 in a way that comports with known biology, i.e. it is known to be protective in adipose tissue where it was  
387 negatively loaded, and harmful in the liver, where it was positively loaded. However, whole transcriptomic  
388 signatures can also be informative in their own right. Combined with increasing amounts of high-dimensional  
389 data in public databases, weighted vectors can be useful for generating hypotheses and potential drug  
390 treatments. We showed that weighted vectors of genes can be analyzed for enriched biological functions  
391 and pathways using GSEA. These vectors can also be paired with data about cell-type specific genes to  
392 generate hypotheses about cell composition in individual tissues. Gene expression derived from patient  
393 biopsies confirmed that the transcriptional signatures we identified in mice predict obesity status in humans,  
394 further supporting the translatability of these results. Finally, we used the CMAP database to show that the

395 transcriptomic signatures we identified in mice could be translated into human drug targets, as currently  
396 used diabetes drugs and targets were among the top hits for reversing the disease-associated signatures. That  
397 these drugs are known to reverse diabetes pathogenesis supports the causal role of these gene signatures in  
398 disease risk as modeled by high-dimensional mediation.

399 In conclusion, we have shown that both tissue specificity and distal gene regulation are critically important to  
400 understanding the genetic architecture of complex traits. We identified important genes and gene signatures  
401 that were heritable, causal of disease, and translatable to other mouse populations and to humans. Finally,  
402 we have shown that by directly acknowledging the complexity of both gene regulation and the genotype-to-  
403 phenotype map, we can gain a new perspective on disease pathogenesis and develop actionable hypotheses  
404 about pathogenic mechanisms and potential treatments.

405 Potentially comment on: 1) Mediating on other endophenotypes 2) Limitation that we only find signatures  
406 “consistent with” mediation and that our approach is hypothesis generating 3) Is the kinship matrix the  
407 sum of all local QTLs, or is it something more inclusive, including distal regulation, development, etc. 4)  
408 endophenotypes don’t need to be gene expression. Can be anything you think is causally related to phenotype  
409 and can be manipulated

## 410 **Data Availability**

411 Here we tell people where to find the data

## 412 **Acknowledgements**

413 Here we thank people

414 **Supplemental Figures**

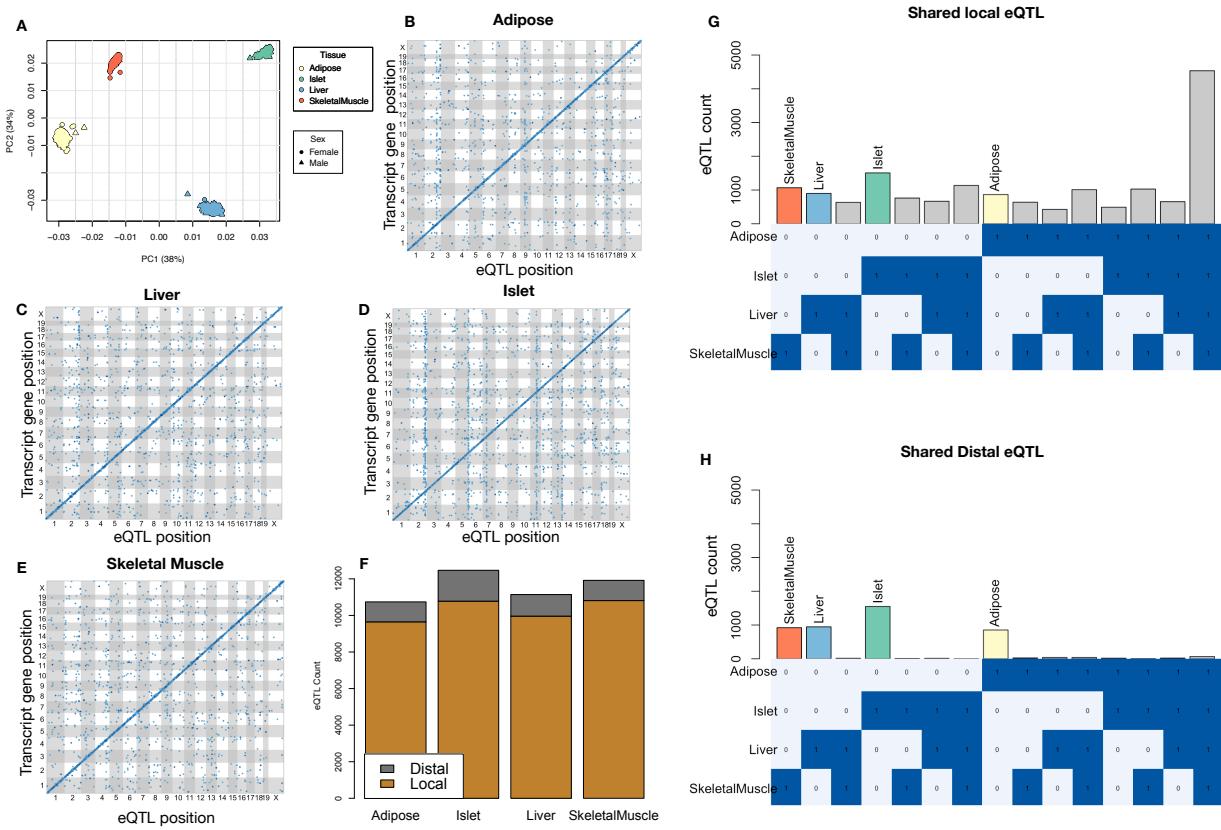


Figure 9: Overview of eQTL analysis in DO mice. **A.** RNA seq samples from the four different tissues clustered by tissue. **B.-E.** eQTL maps are shown for each tissue. The *x*-axis shows the position of the mapped eQTL, and the *y*-axis shows the physical position of the gene encoding each mapped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

415 **References**

- 416 [1] M. T. Maurano, R. Humbert, E. Rynes, R. E. Thurman, E. Haugen, H. Wang, A. P. Reynolds,  
417 R. Sandstrom, H. Qu, J. Brody, A. Shafer, F. Neri, K. Lee, T. Kutyavin, S. Stehling-Sun, A. K.  
418 Johnson, T. K. Canfield, E. Giste, M. Diegel, D. Bates, R. S. Hansen, S. Neph, P. J. Sabo, S. Heimfeld,  
419 A. Raubitschek, S. Ziegler, C. Cotsapas, N. Sotoodehnia, I. Glass, S. R. Sunyaev, R. Kaul, and J. A.

## KEGG pathway enrichments by GSEA

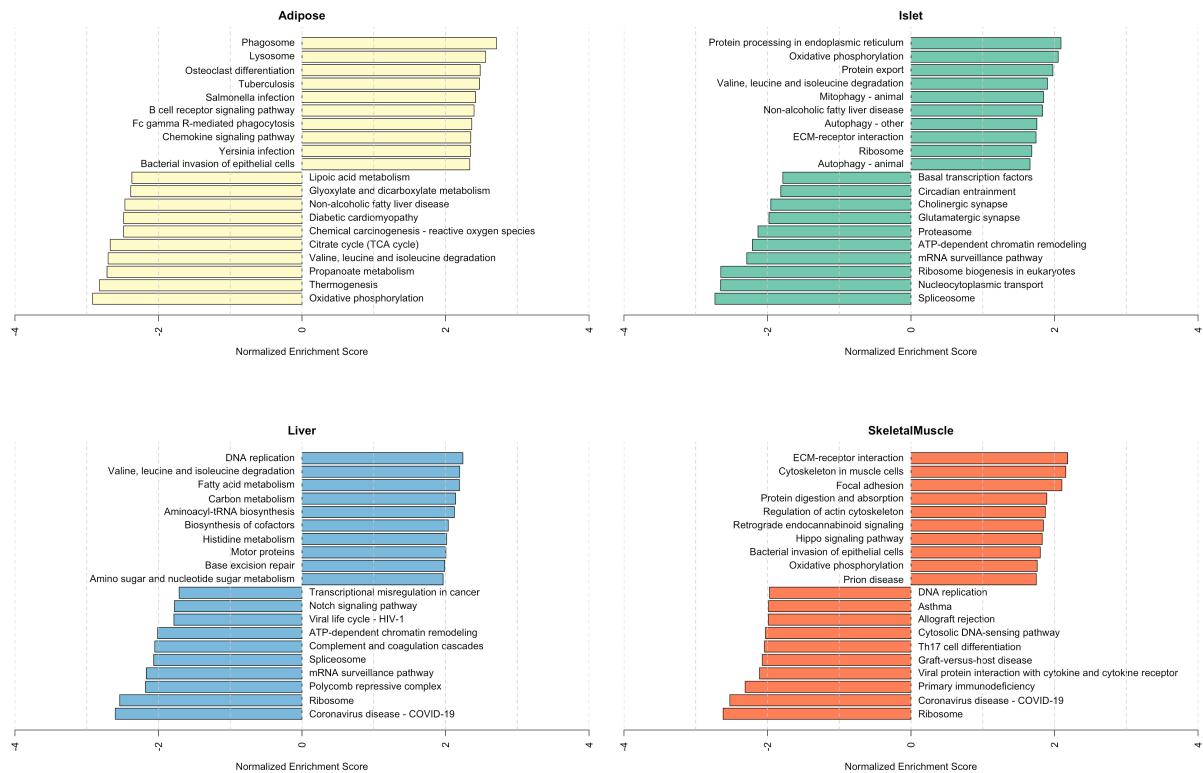


Figure 10: Bar plots showing normalized enrichment scores (NES) for KEGG pathways as determined by fast gene score enrichment analysis (fgsea). Only the top 10 positive and top 10 negative scores are shown. Colors indicate tissue. The name beside each bar shows the name of each enriched KEGG pathway.

- 420 Stamatoyanopoulos. Systematic localization of common disease-associated variation in regulatory DNA.  
 421 *Science*, 337(6099):1190–1195, Sep 2012.
- 422 [2] K. K. Farh, A. Marson, J. Zhu, M. Kleinewietfeld, W. J. Housley, S. Beik, N. Shores, H. Whitton, R. J.  
 423 Ryan, A. A. Shishkin, M. Hatan, M. J. Carrasco-Alfonso, D. Mayer, C. J. Luckey, N. A. Patsopoulos,  
 424 P. L. De Jager, V. K. Kuchroo, C. B. Epstein, M. J. Daly, D. A. Hafler, and B. E. Bernstein. Genetic  
 425 and epigenetic fine mapping of causal autoimmune disease variants. *Nature*, 518(7539):337–343, Feb  
 426 2015.
- 427 [3] E. Pennisi. The Biology of Genomes. Disease risk links to gene regulation. *Science*, 332(6033):1031, May  
 428 2011.
- 429 [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

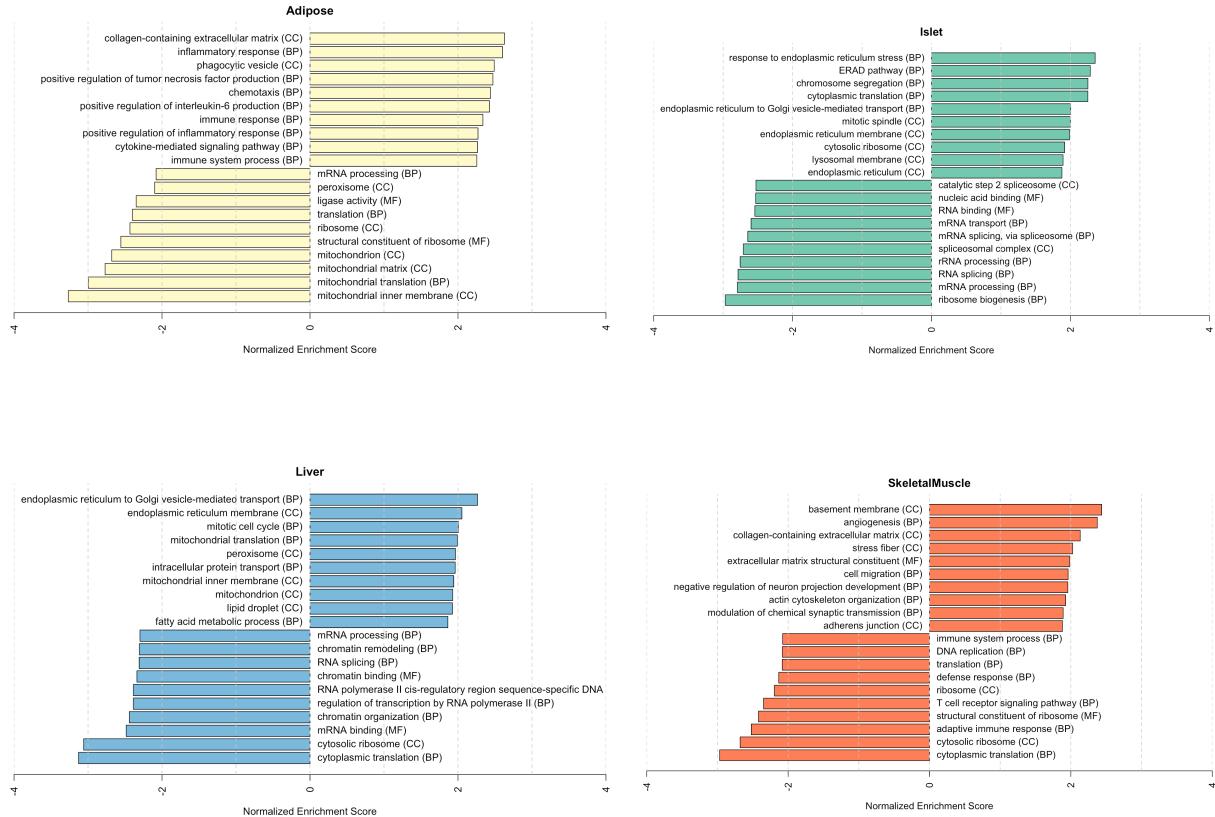


Figure 11: Bar plots showing normalized enrichment scores (NES) for GO terms as determined by fast gene score enrichment analysis (fgsea). Only the top 10 positive and top 10 negative scores are shown. Colors indicate tissue. The name beside each bar shows the name of each enriched GO term. The letters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

- 430 Potential etiologic and functional implications of genome-wide association loci for human diseases and  
 431 traits. *Proc Natl Acad Sci*, 106(23):9362–9367, Jun 2009.
- 432 [5] J. K. Pickrell. Joint analysis of functional genomic data and genome-wide association studies of 18  
 433 human traits. *Am J Hum Genet*, 94(4):559–573, Apr 2014.
- 434 [6] D. Welter, J. MacArthur, J. Morales, T. Burdett, P. Hall, H. Junkins, A. Klemm, P. Flicek, T. Manolio,  
 435 L. Hindorff, and H. Parkinson. The NHGRI GWAS Catalog, a curated resource of SNP-trait associations.  
 436 *Nucleic Acids Res*, 42(Database issue):D1001–1006, Jan 2014.
- 437 [7] Y. I. Li, B. van de Geijn, A. Raj, D. A. Knowles, A. A. Petti, D. Golan, Y. Gilad, and J. K. Pritchard.  
 438 RNA splicing is a primary link between genetic variation and disease. *Science*, 352(6285):600–604, Apr

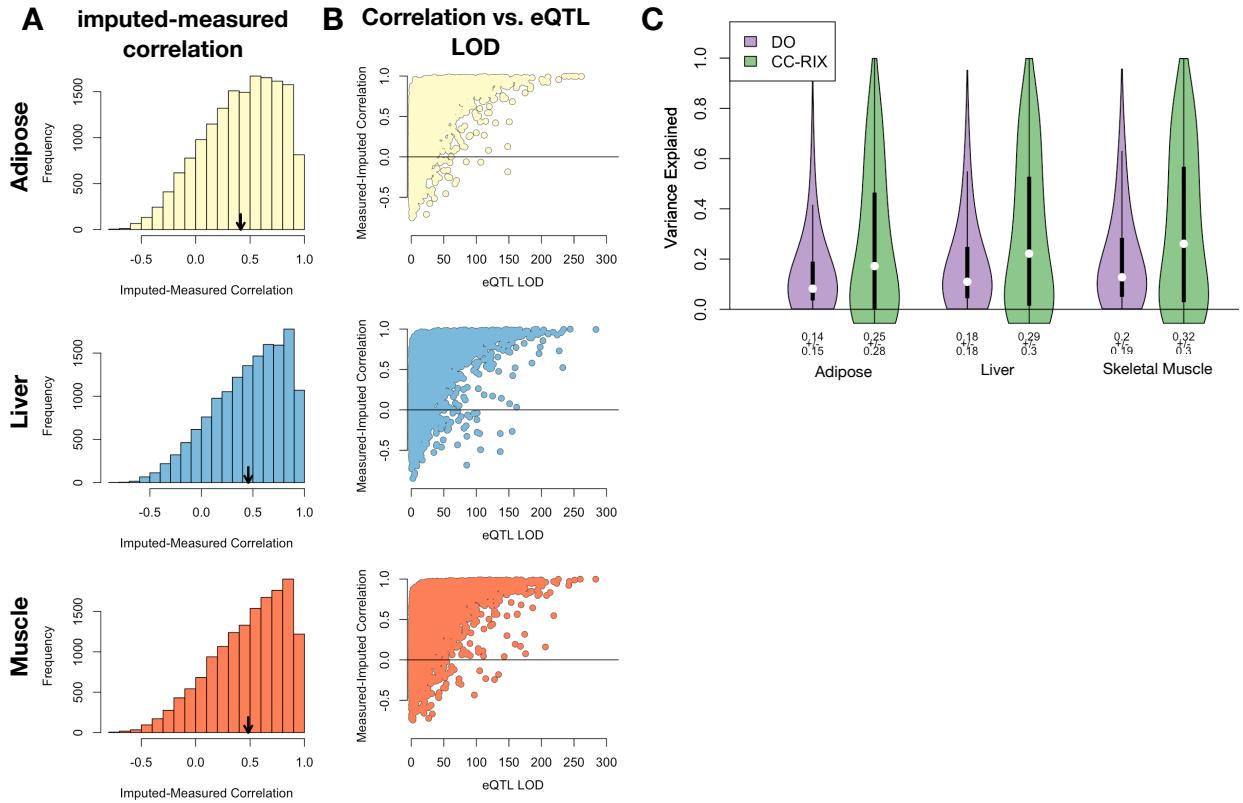


Figure 12: Validation of transcript imputation in the CC-RIX. **A.** Distributions of correlations between imputed and measured transcripts in the CC-RIX. The mean of each distribution is shown by the red line. All distributions were skewed toward positive correlations and had positive means near a Pearson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

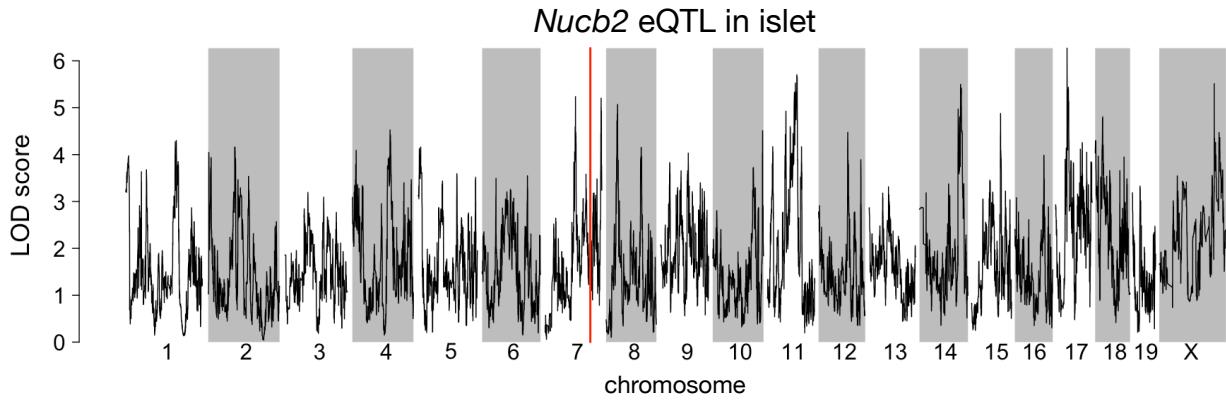


Figure 13: Regulation of *Nucb2* expression in islet. *Nucb2* is encoded on mouse chromosome 7 at 116.5 Mb (red line). In islets the heritability of *Nucb2* expression levels is 69% heritable. This LOD score trace shows that there is no local eQTL at that position, nor any strong distal eQTL anywhere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

- 439 2016.
- 440 [8] D. Zhou, Y. Jiang, X. Zhong, N. J. Cox, C. Liu, and E. R. Gamazon. A unified framework for joint-tissue  
441 transcriptome-wide association and Mendelian randomization analysis. *Nat Genet*, 52(11):1239–1246,  
442 Nov 2020.
- 443 [9] E. R. Gamazon, H. E. Wheeler, K. P. Shah, S. V. Mozaffari, K. Aquino-Michaels, R. J. Carroll, A. E.  
444 Eyler, J. C. Denny, D. L. Nicolae, N. J. Cox, and H. K. Im. A gene-based association method for  
445 mapping traits using reference transcriptome data. *Nat Genet*, 47(9):1091–1098, Sep 2015.
- 446 [10] Z. Zhu, F. Zhang, H. Hu, A. Bakshi, M. R. Robinson, J. E. Powell, G. W. Montgomery, M. E. Goddard,  
447 N. R. Wray, P. M. Visscher, and J. Yang. Integration of summary data from GWAS and eQTL studies  
448 predicts complex trait gene targets. *Nat Genet*, 48(5):481–487, May 2016.
- 449 [11] A. Gusev, A. Ko, H. Shi, G. Bhatia, W. Chung, B. W. Penninx, R. Jansen, E. J. de Geus, D. I. Boomsma,  
450 F. A. Wright, P. F. Sullivan, E. Nikkola, M. Alvarez, M. Civelek, A. J. Lusis, T. ki, E. Raitoharju,  
451 M. nen, I. ä, O. T. Raitakari, J. Kuusisto, M. Laakso, A. L. Price, P. Pajukanta, and B. Pasaniuc.  
452 Integrative approaches for large-scale transcriptome-wide association studies. *Nat Genet*, 48(3):245–252,  
453 Mar 2016.
- 454 [12] M. P. Keller, D. M. Gatti, K. L. Schueler, M. E. Rabaglia, D. S. Stapleton, P. Simecek, M. Vincent,  
455 S. Allen, A. T. Broman, R. Bacher, C. Kendziorski, K. W. Broman, B. S. Yandell, G. A. Churchill, and  
456 A. D. Attie. Genetic Drivers of Pancreatic Islet Function. *Genetics*, 209(1):335–356, May 2018.
- 457 [13] W. L. Crouse, G. R. Keele, M. S. Gastonguay, G. A. Churchill, and W. Valdar. A Bayesian model  
458 selection approach to mediation analysis. *PLoS Genet*, 18(5):e1010184, May 2022.
- 459 [14] J. M. Chick, S. C. Munger, P. Simecek, E. L. Huttlin, K. Choi, D. M. Gatti, N. Raghupathy, K. L. Svenson,  
460 G. A. Churchill, and S. P. Gygi. Defining the consequences of genetic variation on a proteome-wide scale.  
461 *Nature*, 534(7608):500–505, Jun 2016.
- 462 [15] H. E. Wheeler, S. Ploch, A. N. Barbeira, R. Bonazzola, A. Andaleon, A. Fotuhi Siahpirani, A. Saha,  
463 A. Battle, S. Roy, and H. K. Im. Imputed gene associations identify replicable trans-acting genes enriched  
464 in transcription pathways and complex traits. *Genet Epidemiol*, 43(6):596–608, Sep 2019.
- 465 [16] B. D. Umans, A. Battle, and Y. Gilad. Where Are the Disease-Associated eQTLs? *Trends Genet*,  
466 37(2):109–124, Feb 2021.
- 467 [17] N. J. Connally, S. Nazeen, D. Lee, H. Shi, J. Stamatoyannopoulos, S. Chun, C. Cotsapas, C. A. Cassa,

- 468 and S. R. Sunyaev. The missing link between genetic association and regulatory function. *Elife*, 11, Dec  
469 2022.
- 470 [18] H. Mostafavi, J. P. Spence, S. Naqvi, and J. K. Pritchard. Systematic differences in discovery of genetic  
471 effects on gene expression and complex traits. *Nat Genet*, 55(11):1866–1875, Nov 2023.
- 472 [19] D. W. Yao, L. J. O’Connor, A. L. Price, and A. Gusev. Quantifying genetic effects on disease mediated  
473 by assayed gene expression levels. *Nat Genet*, 52(6):626–633, Jun 2020.
- 474 [20] X. Liu, J. A. Mefford, A. Dahl, Y. He, M. Subramaniam, A. Battle, A. L. Price, and N. Zaitlen. GBAT:  
475 a gene-based association test for robust detection of trans-gene regulation. *Genome Biol*, 21(1):211, Aug  
476 2020.
- 477 [21] G. A. Churchill, D. M. Gatti, S. C. Munger, and K. L. Svenson. The Diversity Outbred mouse population.  
478 *Mamm Genome*, 23(9-10):713–718, Oct 2012.
- 479 [22] Michael C Saul, Vivek M Philip, Laura G Reinholdt, and Elissa J Chesler. High-diversity mouse  
480 populations for complex traits. *Trends in Genetics*, 35(7):501–514, 2019.
- 481 [23] S. M. Clee and A. D. Attie. The genetic landscape of type 2 diabetes in mice. *Endocr Rev*, 28(1):48–83,  
482 Feb 2007.
- 483 [24] K. W. Broman, D. M. Gatti, P. Simecek, N. A. Furlotte, P. Prins, Š. Sen, B. S. Yandell, and G. A.  
484 Churchill. R/qt12: Software for Mapping Quantitative Trait Loci with High-Dimensional Data and  
485 Multiparent Populations. *Genetics*, 211(2):495–502, Feb 2019.
- 486 [25] Fabien Girka, Etienne Camenen, Caroline Peltier, Arnaud Gloaguen, Vincent Guillemot, Laurent Le  
487 Brusquet, and Arthur Tenenhaus. *RGCCA: Regularized and Sparse Generalized Canonical Correlation  
488 Analysis for Multiblock Data*, 2023. R package version 3.0.3.
- 489 [26] Gennady Korotkevich, Vladimir Sukhov, and Alexey Sergushichev. Fast gene set enrichment analysis.  
490 *bioRxiv*, 2019.
- 491 [27] A. Subramanian, P. Tamayo, V. K. Mootha, S. Mukherjee, B. L. Ebert, M. A. Gillette, A. Paulovich,  
492 S. L. Pomeroy, T. R. Golub, E. S. Lander, and J. P. Mesirov. Gene set enrichment analysis: a  
493 knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci U S A*,  
494 102(43):15545–15550, Oct 2005.
- 495 [28] C. B. Newgard. Interplay between lipids and branched-chain amino acids in development of insulin  
496 resistance. *Cell Metab*, 15(5):606–614, May 2012.

- 497 [29] D. D. Sears, G. Hsiao, A. Hsiao, J. G. Yu, C. H. Courtney, J. M. Ofrecio, J. Chapman, and S. Subramaniam.  
498 Mechanisms of human insulin resistance and thiazolidinedione-mediated insulin sensitization. *Proc Natl  
499 Acad Sci U S A*, 106(44):18745–18750, Nov 2009.
- 500 [30] R. Stienstra, C. Duval, M. ller, and S. Kersten. PPARs, Obesity, and Inflammation. *PPAR Res*,  
501 2007:95974, 2007.
- 502 [31] O. Gavrilova, M. Haluzik, K. Matsusue, J. J. Cutson, L. Johnson, K. R. Dietz, C. J. Nicol, C. Vinson,  
503 F. J. Gonzalez, and M. L. Reitman. Liver peroxisome proliferator-activated receptor gamma contributes  
504 to hepatic steatosis, triglyceride clearance, and regulation of body fat mass. *J Biol Chem*, 278(36):34268–  
505 34276, Sep 2003.
- 506 [32] K. Matsusue, M. Haluzik, G. Lambert, S. H. Yim, O. Gavrilova, J. M. Ward, B. Brewer, M. L. Reitman,  
507 and F. J. Gonzalez. Liver-specific disruption of PPARgamma in leptin-deficient mice improves fatty  
508 liver but aggravates diabetic phenotypes. *J Clin Invest*, 111(5):737–747, Mar 2003.
- 509 [33] D. Patsouris, J. K. Reddy, M. ller, and S. Kersten. Peroxisome proliferator-activated receptor alpha  
510 mediates the effects of high-fat diet on hepatic gene expression. *Endocrinology*, 147(3):1508–1516, Mar  
511 2006.
- 512 [34] S. E. Schadinger, N. L. Bucher, B. M. Schreiber, and S. R. Farmer. PPARgamma2 regulates lipogenesis  
513 and lipid accumulation in steatotic hepatocytes. *Am J Physiol Endocrinol Metab*, 288(6):E1195–1205,  
514 Jun 2005.
- 515 [35] W. Motomura, M. Inoue, T. Ohtake, N. Takahashi, M. Nagamine, S. Tanno, Y. Kohgo, and T. Okumura.  
516 Up-regulation of ADRP in fatty liver in human and liver steatosis in mice fed with high fat diet. *Biochem  
517 Biophys Res Commun*, 340(4):1111–1118, Feb 2006.
- 518 [36] S. Amin, A. Lux, and F. O’Callaghan. The journey of metformin from glycaemic control to mTOR  
519 inhibition and the suppression of tumour growth. *Br J Clin Pharmacol*, 85(1):37–46, Jan 2019.
- 520 [37] A. Kezic, L. Popovic, and K. Lalic. mTOR Inhibitor Therapy and Metabolic Consequences: Where Do  
521 We Stand? *Oxid Med Cell Longev*, 2018:2640342, 2018.
- 522 [38] A. D. Barlow, M. L. Nicholson, and T. P. Herbert. -cells and a review of the underlying molecular  
523 mechanisms. *Diabetes*, 62(8):2674–2682, Aug 2013.
- 524 [39] Y. Gu, J. Lindner, A. Kumar, W. Yuan, and M. A. Magnuson. Rictor/mTORC2 is essential for  
525 maintaining a balance between beta-cell proliferation and cell size. *Diabetes*, 60(3):827–837, Mar 2011.

- 526 [40] E. B. Geer, J. Islam, and C. Buettner. Mechanisms of glucocorticoid-induced insulin resistance: focus on  
527 adipose tissue function and lipid metabolism. *Endocrinol Metab Clin North Am*, 43(1):75–102, Mar 2014.
- 528 [41] J. X. Li and C. L. Cummins. Fresh insights into glucocorticoid-induced diabetes mellitus and new  
529 therapeutic directions. *Nat Rev Endocrinol*, 18(9):540–557, Sep 2022.
- 530 [42] R. A. Lee, C. A. Harris, and J. C. Wang. Glucocorticoid Receptor and Adipocyte Biology. *Nucl Receptor  
531 Res*, 5, 2018.
- 532 [43] S. Viengchareun, P. Penfornis, M. C. Zennaro, and M. s. Mineralocorticoid and glucocorticoid receptors  
533 inhibit UCP expression and function in brown adipocytes. *Am J Physiol Endocrinol Metab*, 280(4):E640–  
534 649, Apr 2001.
- 535 [44] J. Liu, X. Kong, L. Wang, H. Qi, W. Di, X. Zhang, L. Wu, X. Chen, J. Yu, J. Zha, S. Lv, A. Zhang,  
536 P. Cheng, M. Hu, Y. Li, J. Bi, Y. Li, F. Hu, Y. Zhong, Y. Xu, and G. Ding. -HSD1 in regulating brown  
537 adipocyte function. *J Mol Endocrinol*, 50(1):103–113, Feb 2013.
- 538 [45] L. E. Ramage, M. Akyol, A. M. Fletcher, J. Forsythe, M. Nixon, R. N. Carter, E. J. van Beek,  
539 N. M. Morton, B. R. Walker, and R. H. Stimson. Glucocorticoids Acutely Increase Brown Adipose  
540 Tissue Activity in Humans, Revealing Species-Specific Differences in UCP-1 Regulation. *Cell Metab*,  
541 24(1):130–141, Jul 2016.
- 542 [46] J. L. Barclay, H. Agada, C. Jang, M. Ward, N. Wetzig, and K. K. Ho. Effects of glucocorticoids on  
543 human brown adipocytes. *J Endocrinol*, 224(2):139–147, Feb 2015.
- 544 [47] M. ó, R. Gupte, W. L. Kraus, P. Pacher, and P. Bai. PARPs in lipid metabolism and related diseases.  
545 *Prog Lipid Res*, 84:101117, Nov 2021.
- 546 [48] P. Bai, C. ó, H. Oudart, A. nszki, Y. Cen, C. Thomas, H. Yamamoto, A. Huber, B. Kiss, R. H.  
547 Houtkooper, K. Schoonjans, V. Schreiber, A. A. Sauve, J. Menissier-de Murcia, and J. Auwerx. PARP-1  
548 inhibition increases mitochondrial metabolism through SIRT1 activation. *Cell Metab*, 13(4):461–468,  
549 Apr 2011.
- 550 [49] A. Chiarugi and M. A. Moskowitz. Cell biology. PARP-1—a perpetrator of apoptotic cell death? *Science*,  
551 297(5579):200–201, Jul 2002.
- 552 [50] M. Althubiti, R. Almaimani, S. Y. Eid, M. Elzubaier, B. Refaat, S. Idris, T. A. Alqurashi, and M. Z.  
553 El-Readi. BTK targeting suppresses inflammatory genes and ameliorates insulin resistance. *Eur Cytokine  
554 Netw*, 31(4):168–179, Dec 2020.

- 555 [51] C. Skrabs, W. F. Pickl, T. Perkmann, U. ger, and A. Gessl. Rapid decline in insulin antibodies and  
556 glutamic acid decarboxylase autoantibodies with ibrutinib therapy of chronic lymphocytic leukaemia. *J*  
557 *Clin Pharm Ther*, 43(1):145–149, Feb 2018.
- 558 [52] E. A. Oral, S. M. Reilly, A. V. Gomez, R. Meral, L. Butz, N. Ajluni, T. L. Chenevert, E. Korytnaya,  
559 A. H. Neidert, R. Hench, D. Rus, J. F. Horowitz, B. Poirier, P. Zhao, K. Lehmann, M. Jain, R. Yu,  
560 C. Liddle, M. Ahmadian, M. Downes, R. M. Evans, and A. R. Saltiel. and TBK1 Improves Glucose  
561 Control in a Subset of Patients with Type 2 Diabetes. *Cell Metab*, 26(1):157–170, Jul 2017.
- 562 [53] M. C. Arkan, A. L. Hevener, F. R. Greten, S. Maeda, Z. W. Li, J. M. Long, A. Wynshaw-Boris, G. Poli,  
563 J. Olefsky, and M. Karin. IKK-beta links inflammation to obesity-induced insulin resistance. *Nat Med*,  
564 11(2):191–198, Feb 2005.
- 565 [54] M. Clark, C. J. Kroger, Q. Ke, and R. M. Tisch. The Role of T Cell Receptor Signaling in the  
566 Development of Type 1 Diabetes. *Front Immunol*, 11:615371, 2020.
- 567 [55] P. ndell, A. C. Carlsson, A. Larsson, O. Melander, T. Wessman, J. v, and T. Ruge. TNFR1 is associated  
568 with short-term mortality in patients with diabetes and acute dyspnea seeking care at the emergency  
569 department. *Acta Diabetol*, 57(10):1145–1150, Oct 2020.
- 570 [56] A. M. Keestra-Gounder, M. X. Byndloss, N. Seyffert, B. M. Young, A. vez Arroyo, A. Y. Tsai, S. A.  
571 Cevallos, M. G. Winter, O. H. Pham, C. R. Tiffany, M. F. de Jong, T. Kerrinnes, R. Ravindran, P. A.  
572 Luciw, S. J. McSorley, A. J. umler, and R. M. Tsolis. NOD1 and NOD2 signalling links ER stress with  
573 inflammation. *Nature*, 532(7599):394–397, Apr 2016.
- 574 [57] A. M. Keestra-Gounder and R. M. Tsolis. NOD1 and NOD2: Beyond Peptidoglycan Sensing. *Trends*  
575 *Immunol*, 38(10):758–767, Oct 2017.
- 576 [58] J. Montane, L. Cadavez, and A. Novials. Stress and the inflammatory process: a major cause of  
577 pancreatic cell death in type 2 diabetes. *Diabetes Metab Syndr Obes*, 7:25–34, 2014.
- 578 [59] B. B. Kahn and T. E. McGraw. , and type 2 diabetes. *N Engl J Med*, 363(27):2667–2669, Dec 2010.
- 579 [60] S. Del Prato and N. Pulizzi. The place of sulfonylureas in the therapy for type 2 diabetes mellitus.  
580 *Metabolism*, 55(5 Suppl 1):S20–27, May 2006.
- 581 [61] B. Hallgrímsson, R. M. Green, D. C. Katz, J. L. Fish, F. P. Bernier, C. C. Roseman, N. M. Young,  
582 J. M. Cheverud, and R. S. Marcucio. The developmental-genetics of canalization. *Semin Cell Dev Biol*,  
583 88:67–79, Apr 2019.

- 584 [62] M. L. Siegal and A. Bergman. Waddington's canalization revisited: developmental stability and evolution.
- 585       *Proc Natl Acad Sci U S A*, 99(16):10528–10532, Aug 2002.
- 586 [63] A. B. Paaby and G. Gibson. Cryptic Genetic Variation in Evolutionary Developmental Genetics. *Biology*
- 587       *(Basel)*, 5(2), Jun 2016.
- 588 [64] X. Liu, Y. I. Li, and J. K. Pritchard. Trans Effects on Gene Expression Can Drive Omnipotent Inheritance.
- 589       *Cell*, 177(4):1022–1034, May 2019.
- 590 [65] Naomi R Wray, Cisca Wijmenga, Patrick F Sullivan, Jian Yang, and Peter M Visscher. Common disease
- 591       is more complex than implied by the core gene omnipotent model. *Cell*, 173(7):1573–1580, 2018.
- 592 [66] M. Riva, M. D. Nitert, U. Voss, R. Sathanoori, A. Lindqvist, C. Ling, and N. Wierup. Nesfatin-1
- 593       stimulates glucagon and insulin secretion and beta cell NUCB2 is reduced in human type 2 diabetic
- 594       subjects. *Cell Tissue Res*, 346(3):393–405, Dec 2011.
- 595 [67] M. Nakata and T. Yada. Role of NUCB2/nesfatin-1 in glucose control: diverse functions in islets,
- 596       adipocytes and brain. *Curr Pharm Des*, 19(39):6960–6965, 2013.
- 597 [68] H. Shimizu and A. Osaki. Nesfatin/Nucleobindin-2 (NUCB2) and Glucose Homeostasis. *Curr Hypertens*
- 598       *Rev*, pages Nesfatin/Nucleobindin-2 (NUCB2) and Glucose Homeostasis., Jul 2014.