

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 mediating transcripts, we get one step closer to a molecular  
11 understanding of how genetic variants influence traits. Moreover, there is evidence from genome-wide  
12 association studies (GWAS) that regulation of gene expression accounts for the bulk of the genetic effect  
13 on complex traits, as most trait-associated variants lie in gene regulatory regions [1, 2, 3, 4, 5, 6, 7]. It is  
14 widely assumed that these variants influence local transcription, and methods such as transcriptome-wide  
15 association studies (TWAS) [8, 9, 10, 11], summary data-based Mendelian randomization (SMR) [10], and  
16 others have capitalized on this idea to identify genes associated with multiple disease traits [12, 13, 14, 15]

17 Despite the great promise of these methods, however, they have not been as widely successful as it seemed  
18 they could have been, and the vast majority of complex trait heritability remains unexplained. Although  
19 trait-associated variants tend to lie in non-coding, regulatory regions, they often do not have detectable effects  
20 on gene expression [16] and tend not to co-localize with expression quantitative trait loci (eQTLs) [17, 18].

21 One possible explanation for these observations is that gene expression is not being measured in the appropriate  
22 cell types and thus true eQTLs influencing traits cannot be detected [16]. An alternative explanation that  
23 has been discussed in recent years is that effects of these variants are mediated not through local regulation  
24 of gene expression, but through distal regulation [18, 19, 20, 15].

25 However, assessing the role of wide-spread distal gene regulation on complex traits requires large, dedicated data

26 sets that include high-dimensional, clinically relevant phenotyping, dense genotyping in a highly recombined  
27 population, and transcriptome-wide measurements of gene expression in multiple tissues. Measuring gene  
28 expression in multiple tissues is critical to adequately assess the extent to which local gene regulation varies  
29 across multiple tissues and whether such variability might account for previous failed attempts to identify  
30 trait-relevant local eQTL. Such data sets are extremely difficult to obtain in human populations, particularly  
31 in the large numbers of subjects required for statistical testing. Thus, to investigate further the role of local  
32 and distal gene regulation on complex traits, we have generated an appropriate data set in a large population  
33 of diversity outbred (DO) mice [21] in a population model of diet-induced obesity and metabolic disease [12].

34 The DO mice were derived from eight inbred founder mouse strains, five classical lab strains, and three  
35 strains more recently derived from wild mice [21]. They represent three subspecies of mouse *Mus musculus*  
36 *domesticus*, *Mus musculus musculus*, and *Mus musculus castaneus*, and capture 90% of the known variation  
37 in laboratory mice [cite]. They are maintained with a breeding scheme that ensures equal contributions from  
38 each founder across the genome thus rendering almost the whole genome visible to genetic inquiry [21]. We  
39 measured clinically relevant metabolic traits, including body weight, plasma levels of insulin and glucose,  
40 and plasma lipids in 500 DO mice. We further measured transcriptome-wide gene expression in four tissues  
41 related to metabolic disease: adipose tissue, pancreatic islets, liver, and skeletal muscle.

42 To assess the role of gene regulation in mediating variation in metabolic traits in this population, we propose  
43 high-dimensional mediation (HDM). In univariate approaches, such as TWAS, SMR, and other Mendelian  
44 randomization approaches, each transcript is tested independently for mediation of a local variant on a  
45 trait. This process requires huge numbers of statistical tests, which is computationally expensive, requires  
46 strict corrections for multiple testing, and assumes independence of genetic variants and transcripts. Such  
47 methods are therefore limited to detecting only the largest statistical effects and are biased toward local gene  
48 regulation. In contrast, with high-dimensional mediation we assessed broad relationships among the genome,  
49 transcriptome, and phenotype as a whole and identified a highly heritable composite trait that was perfectly  
50 mediated by a composite transcript. We show that composite transcripts were tissue-specific and highly  
51 interpretable in terms of biological processes as well as cell type composition. Heritability analysis of the  
52 transcripts showed that the strongest transcriptional mediators of metabolic disease had low local heritability  
53 and high distal heritability. Finally, we show that the composite transcripts identified in the DO population  
54 predicted obesity in an independent population of Collaborative Cross recombinant inbred (CC-RIX) mice  
55 and in human subjects. In contrast, local eQTL were unable to predict obesity in the CC-RIX mice. Together  
56 our results suggest that both the tissue used for gene expression analysis as well as distal gene regulation are  
57 critically important in identifying transcriptional mediators of the genome on complex traits.

58 **Results**

59 **Genetic variation contributed to wide phenotypic variation**

60 Although the environment was consistent across all animals, the genetic diversity present in this population  
61 resulted in widely varying distributions across physiological measurements (Fig. 1). For example, body  
62 weights of adult individuals varied from less than the average adult B6 body weight to several times the body  
63 weight of a B6 adult in both sexes (Fig. 1A). Fasting blood glucose (FBG) also varied considerably (Fig. 1B)  
64 although few of the animals had FBG levels that would indicate pre-diabetes ( animals, ), or diabetes (7  
65 animals, 1.4) according to previously developed cutoffs (pre-diabetes:  $\text{FBG} \geq 250 \text{ mg/dL}$ , diabetes:  $\text{FBG} \geq$   
66 300, mg/dL) [22]. Males had higher FBG than females on average (Fig. 1C) as has been observed before  
67 suggesting either that males were more susceptible to metabolic disease on the high-fat diet, or that males  
68 and females may require different thresholds for pre-diabetes and diabetes.

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

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

77 **Distal Heritability Correlated with Phenotype Relevance**

78 We performed eQTL analysis using R/qltl2 [23] (Methods) and identified both local and distal eQTL for  
79 transcripts in each of the four tissues (Supp. Fig 9). Significant local eQTLs far outnumbered distal eQTLs  
80 (Supp. Fig. 9F) and tended to be shared across tissues (Supp. Fig. 9G) whereas the few significant distal  
81 eQTL we identified tended to be tissue-specific (Supp. Fig. 9H)

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

85 Local heritability of transcripts was negatively correlated with their trait relevance, defined as the maximum  
86 correlation of a transcript across all traits (Fig. 2B). This suggests that the more local genotype influenced  
87 transcript abundance, the less effect variation in transcript abundance had on the measured traits. Conversely,

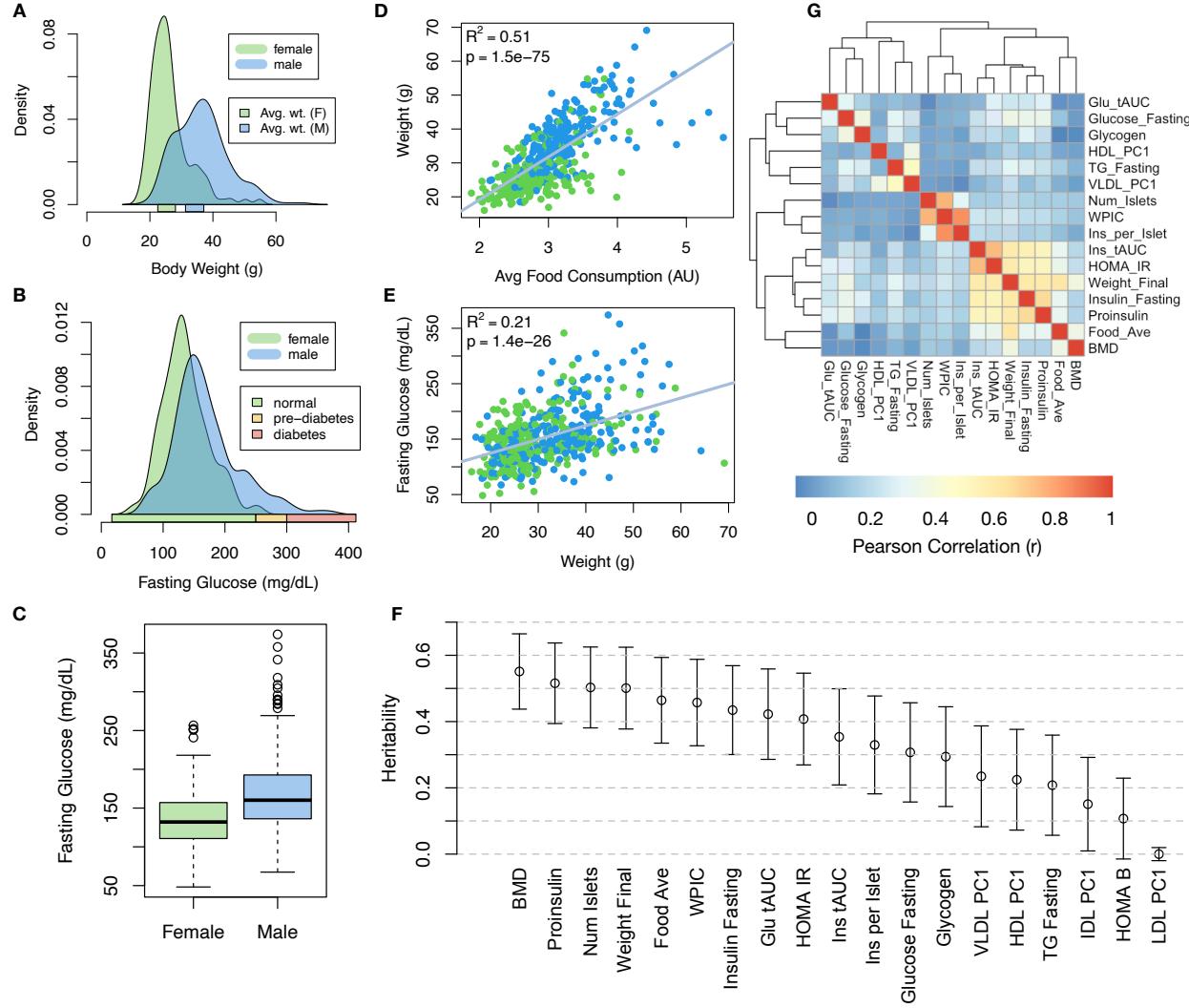


Figure 1: Clinical overview. **A.** Distributions of final body weight in the diversity outbred mice. Sex is indicated by color. The average B6 male and female adult weights at 24 weeks of age are indicated by blue and green bars on the x-axis. **B.** The distribution of final fasting glucose across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

distal heritability of transcripts was positively correlated with trait relevance (Fig. 2C). That is, transcripts that were more highly correlated with the measured traits tended to be distally, rather than locally, heritable. That trait-correlated transcripts have low local heritability is consistent with previous observations that low-heritability transcripts explain more expression-mediated disease heritability than high-heritability transcripts [19]. However, the positive relationship between trait correlation and distal heritability suggests that there are alternative mechanisms through which genetic regulation of transcripts may influence traits.

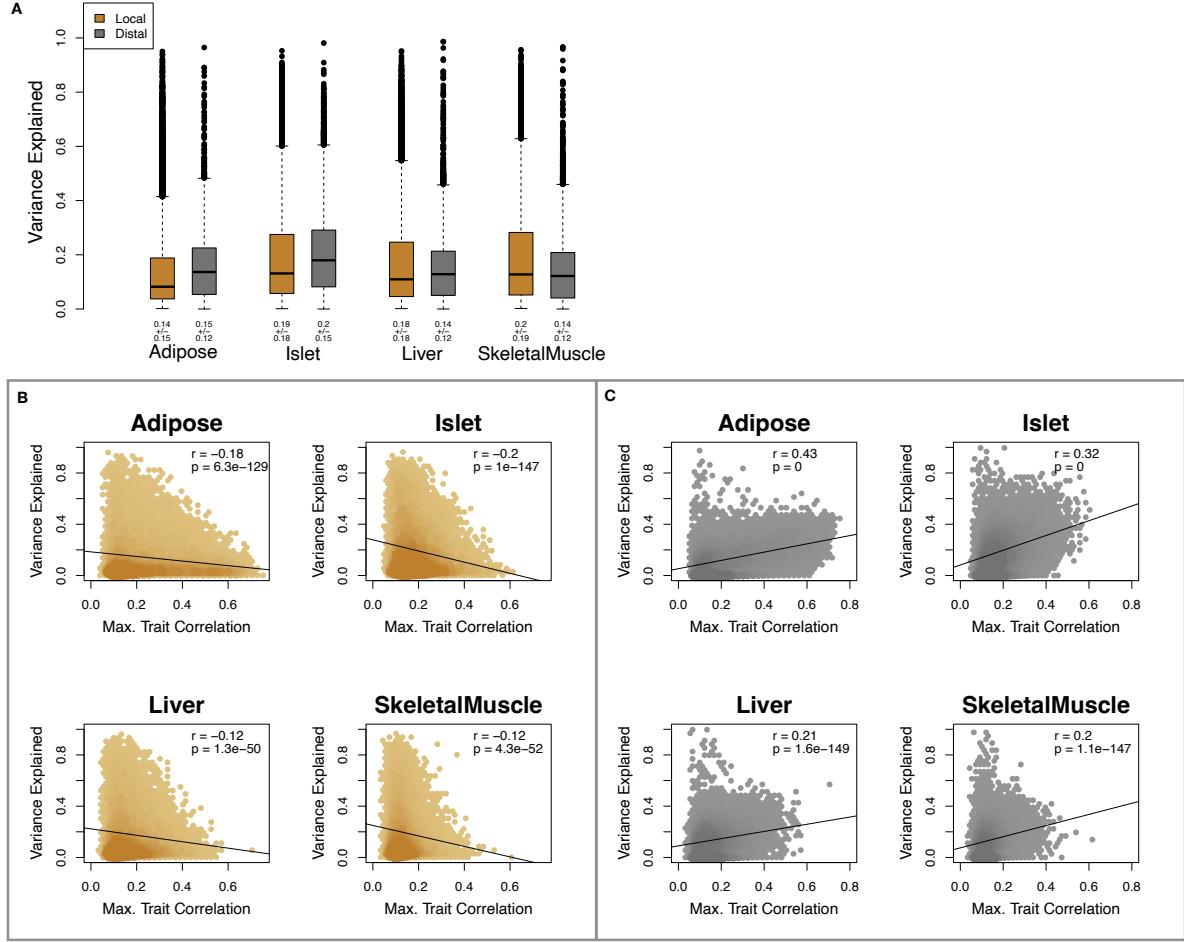


Figure 2: Transcript heritability and trait relevance. **A.** Distributions of distal and local heritability of transcripts across the four tissues. Overall local and distal factors contribute equally to transcript heritability. The relationship between **(B.)** local and **(C.)** distal heritability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

- 94 **High-Dimensional Mediation identified a high-heritability composite trait that was perfectly  
95 mediated by a composite transcript**
- 96 We used high-dimensional mediation to identify the major axis of variation in the transcriptome that mediated  
97 the effects of the genome on metabolic traits (Fig. 3). We kernelized the genome, phenome, and transcriptome  
98 matrices and used generalized canonical correlation analysis (RGCCA) [24] to identify a composite transcript  
99 ( $T_C$ ) that perfectly mediated the effect of the composite genome ( $G_C$ ) on the composite phenome ( $P_C$ ).  
100 Fig. 3A shows the partial correlations ( $\rho$ ) between the pairs of these composite vectors. The partial correlation  
101 between  $G_C$  and  $T_S$  was 0.42, and the partial correlation between  $T_S$  and  $P_S$  was 0.78. However, when the

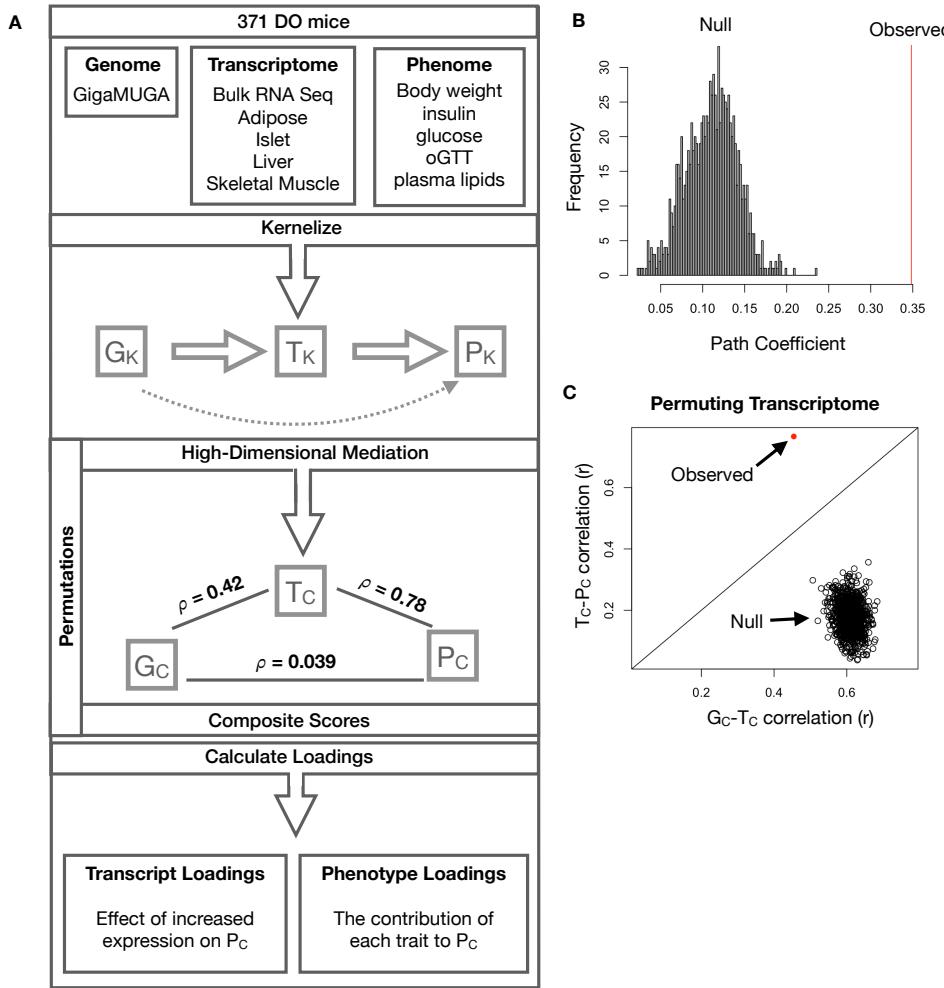


Figure 3: High-dimensional mediation. **A.** Workflow indicating major steps of high-dimensional mediation. The genotype, transcriptome, and phenotype matrices were kernelized to yield single matrices representing the relationships between all individuals for each data modality ( $G_K$  = genome kernel,  $T_K$  = 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).

- 102 transcriptome was taken into account, the partial correlation between  $G_S$  and  $P_S$  was effectively 0 (0.039).
- 103 The estimated heritability of the composite phenotype was heritability of  $0.71 \pm 0.084$ , which was higher than
- 104 any of the individual traits (Fig. 1F). Thus, we have identified a maximally heritable metabolic trait that is
- 105 perfectly mediated by a heritable component of the transcriptome.
- 106 Standard CCA is prone to over-fitting because in any two large matrices it can be trivial to identify
- 107 highly correlated composite vectors. To assess whether RGCCA was similarly prone to over-fitting in

108 a high-dimensional space, we performed permutation testing. We permuted the individual labels on the  
109 transcriptome kernel matrix 1000 times and recalculated the path coefficient, which is the partial correlation of  
110  $G_C$  and  $T_C$  multiplied by the partial correlation of  $T_C$  and  $P_C$ . This represents the path from  $G_C$  to  $P_C$  that is  
111 mediated through  $T_C$ . The null distribution of the path coefficient is shown in Fig. 3B, and the observed path  
112 coefficient from the original data is indicated by the red line. The observed path coefficient was well outside  
113 the null distribution generated by permutations. Fig. 3C illustrates this observation in more detail. Although  
114 we identified high correlations between  $G_C$  and  $T_C$ , and modest correlations between  $T_C$  and  $P_C$  in the null  
115 data (Fig 3C), these two values could not be maximized simultaneously. The red dot shows that in the real  
116 data both the  $G_C$ - $T_C$  correlation and the  $T_C$ - $P_C$  correlation could be maximized simultaneously suggesting  
117 that the path from genotype to phenotype through transcriptome is highly non-trivial and identifiable in  
118 this case. These results suggest that these composite vectors represent genetically determined variation in  
119 phenotype that is mediated through genetically determined variation in transcription.

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

122 The loadings of each measured trait onto  $P_C$  indicate how much each contributed to the composite phenotype.  
123 Final body weight contributed the most (Fig. 4), followed by homeostatic insulin resistance (HOMA\_IR) and  
124 fasting plasma insulin levels (Insulin\_Fasting). We can thus interpret  $P_C$  as an index of metabolic disease (Fig.  
125 4B). Individuals with high values of  $P_C$  have a higher metabolic index and greater metabolic disease, including  
126 higher body weight and higher insulin resistance. We refer to  $P_C$  as the metabolic index going forward. Traits  
127 contributing the least to the metabolic index were measures of cholesterol and pancreas composition. Thus,  
128 when we interpret the transcriptomic signature identified by HDM, we are explaining primarily transcriptional  
129 mediation of body weight and insulin resistance, as opposed to cholesterol measurements.

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

132 We interpreted large loadings onto transcripts as indicating strong mediation of the effect of genetics on  
133 metabolic index. Large positive loadings indicate that higher expression was associated with a higher  
134 metabolic index (i.e. higher risk of obesity and metabolic disease on the high-fat diet) (Fig. 4C). Conversely,  
135 large negative loadings indicate that high expression of these transcripts was associated with a lower metabolic  
136 index (i.e. lower risk of obesity and metabolic disease on the high-fat diet) (Fig. 4C). We used gene set  
137 enrichment analysis (GSEA) [25, 26] to look for biological processes and pathways that were enriched at the  
138 top and bottom of this list (Methods).

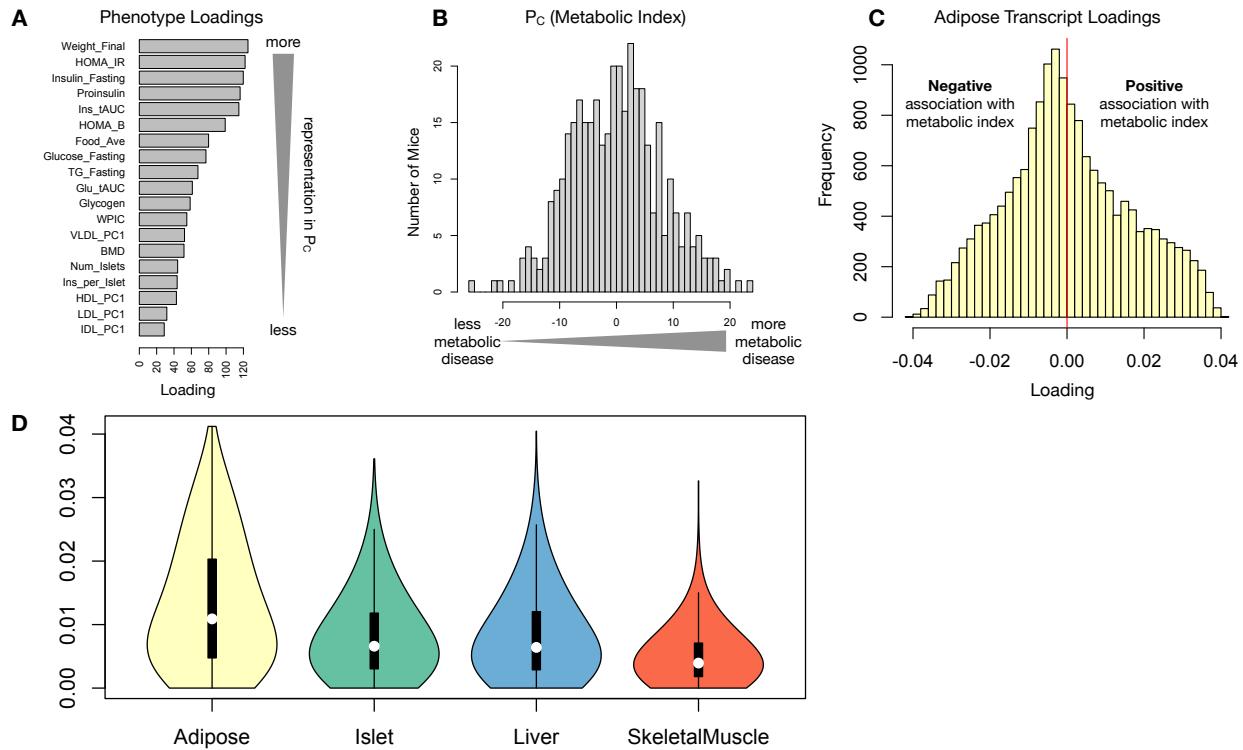


Figure 4: Interpretation of loadings. **A.** Loadings across traits. Body weight and insulin resistance contributed the most to the composite trait. **B.** Phenotype scores across individuals. Individuals with large positive phenotype scores had higher body weight and insulin resistance than average.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

- 139 In adipose tissue, both GO processes and KEGG pathway enrichments pointed to an axis of inflammation  
 140 and metabolism (Supp. Fig. 10 and 11). GP terms and KEGG pathways associated with inflammation,  
 141 particularly macrophage infiltration, were positively associated with metabolic index, indicating that increased  
 142 expression in inflammatory pathways was associated with a higher metabolic index. It is well established  
 143 that adipose tissue in obese individuals is highly inflamed [cite] and infiltrated by macrophages [cite], and the  
 144 results here suggest that this may be a heritable component of metabolic disease.
- 145 The strongest negative enrichments in adipose tissue were related to mitochondrial activity in general, and  
 146 thermogenesis in particular (Supp. Fig. 10 and 11). It has been shown mouse strains with greater thermogenic  
 147 potential are also less susceptible to obesity on a high-fat diet [cite].
- 148 Transcripts associated with the citric acid (TCA) cycle as well as the catabolism of branched-chain amino

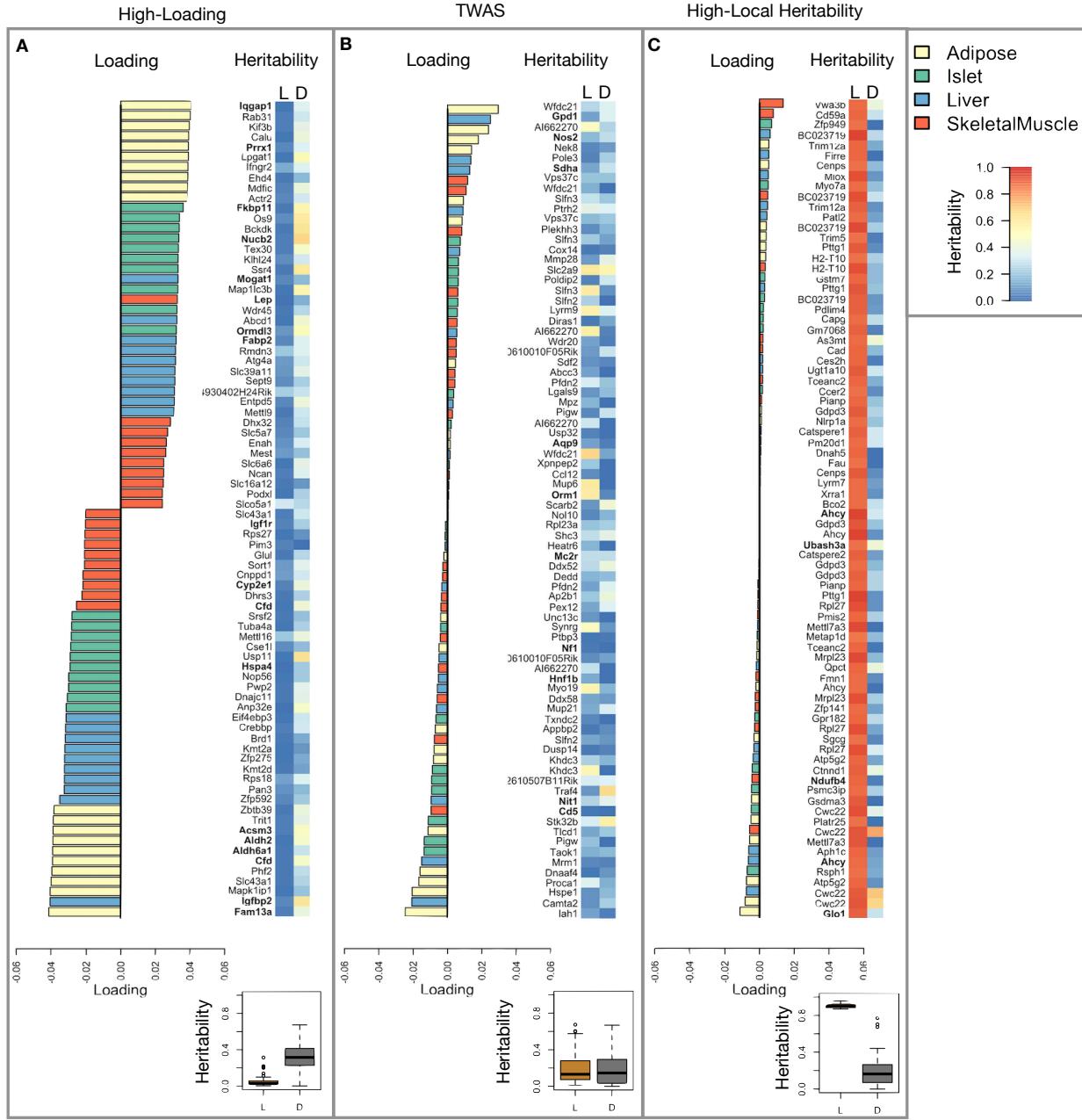
149 acids (BCAA), valine, leuceine, and isoleucine were strongly enriched with negative loadings in adipose  
150 tissue (Supp. Fig. XXX). Expression of genes in both pathways (for which there is some overlap) has been  
151 previously associated with insulin sensitivity [12, 27, 28], suggesting that heritable variation in regulation of  
152 these pathways may influence risk of insulin resistance.

153 Looking at the 10 strongest positive and negative loaded transcripts from each tissue, it is apparent that  
154 transcripts in the adipose tissue had the largest loadings, both positive and negative, of all tissues (Fig.  
155 5A bar plot) This suggesting that much of the effect of genetics on body weight and insulin resistance  
156 is mediated through gene expression in adipose tissue. The strongest loadings in liver and pancreas were  
157 comparable, and those in skeletal muscle were the weakest (Fig. 5A), suggesting that less of the genetic  
158 effects were mediated through transcription in skeletal muscle. Heritability analysis showed that transcripts  
159 with the largest loadings tended to have relatively high distal heritability compared with local heritability  
160 (Fig. 5A heat map and box plot). This pattern contrasts with transcripts nominated by TWAS (Fig. 5B),  
161 which tended to have lower loadings, higher local heritability and lower distal heritability. Transcripts with  
162 the highest local heritability in each tissue (Fig. 5C) had the lowest loadings.

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

#### 169 **Tissue-specific transcriptional programs were associated with metabolic traits**

170 Clustering of transcripts with top loadings in each tissue showed tissue-specific functional modules associated  
171 with obesity and insulin resistance (Fig. 6A) (Methods). The clustering highlights the importance of immune  
172 activation particularly in adipose tissue. Except for the “mitosis” cluster, which had large positive loadings in  
173 three of the four tissues, all clusters were strongly loaded in only one or two tissues. For example, the lipid  
174 metabolism cluster was loaded most heavily in liver. The positive loadings suggest that high expression of  
175 these genes particularly in the liver was associated with increased metabolic disease. This cluster included  
176 the gene *Pparg*, whose primary role is in the adipose tissue where it is considered a master regulator of  
177 adipogenesis [29]. Agonists of *Pparg*, such as Thiazolidinediones, which are FDA-approved to treat type II  
178 diabetes, reduce inflammation and adipose hypertrophy [29]. Consistent with this role, the loading for *Pparg*  
179 in adipose tissue was negative, suggesting that higher expression was associated with leaner mice (Fig. 6B).



183 mice with a *Ppara* knockout, causes upregulation of genes involved in adipogenesis [32]. In the livers of both  
184 mice and humans high *Pparg* expression is associated with hepatocytes that accumulate large lipid droplets  
185 and have gene expression profiles similar to adipocytes [33, 34].

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

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

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

195 To test whether the transcript loadings identified in the DO could be translated to another population, we  
196 tested whether they could predict metabolic a phenotype in an independent population of CC-RIX mice, which  
197 were F1 mice derived from multiple pairings of Collaborative Cross (CC) [cite] strains (Fig. 7) (Methods).  
198 We tested two questions. First, we asked whether the loadings identified in the DO mice were relevant to the  
199 relationship between the transcriptome and the genome in the CC-RIX. We predicted body weight in each  
200 CC-RIX individual using measured gene expression in each tissue and the transcript loadings identified in the  
201 DO (Methods). The predicted body weight and actual body weight were highly correlated in all tissues (Fig.  
202 7B left column). The best prediction was achieved for adipose tissue, which supports the observation in the  
203 DO that adipose expression was the strongest mediator of the genetic effect on metabolic index. This result  
204 also confirms the validity and translatability of the transcript loadings and their relationship to metabolic  
205 disease.

206 The second question related to the source of the relevant variation in gene expression. If local regulation was  
207 the predominant factor influencing gene expression, we should be able to predict phenotype in the CC-RIX  
208 using transcripts imputed from local genotype (Fig. 7A). The DO and the CC-RIX were derived from the  
209 same eight founder strains and so carry the same alleles throughout the genome. We imputed gene expression  
210 in the CC-RIX using local genotype and were able to estimate variation in gene transcription robustly (Supp.  
211 Fig. 12). However, these imputed values failed to predict body weight in the CC-RIX when weighted with  
212 the loadings from HDM. (Fig. 7B right column). This result suggests that local regulation of gene expression  
213 is not the primary factor driving heritability of complex traits.

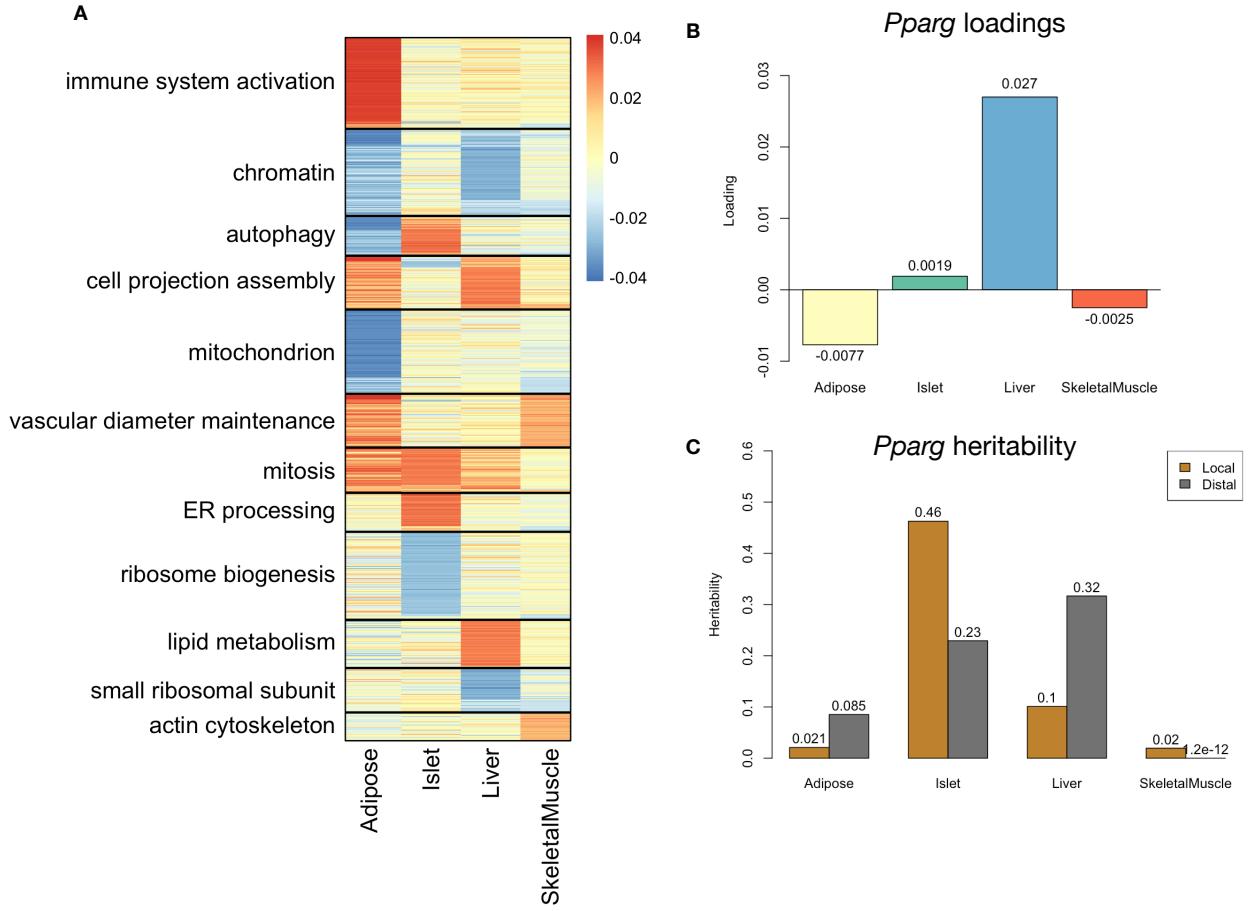


Figure 6: Tissue-specific transcriptional programs were associated with obesity and insulin resistance. **A** Heat map showing the loadings of all transcripts with loadings greater than 2.5 standard deviations from the mean in any tissue. The heat map was clustered using k medoid clustering. Functional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

214 **Distally heritable transcriptomic signatures reflected variation in composition of adipose tissue  
215 and islets**

216 Interpretation of global genetic influences on gene expression and phenotype is potentially more challenging  
217 than interpretation and translation of local genetic influences, as genetic effects cannot be localized to  
218 individual gene variants or transcripts. However, there are global patterns across the loadings that can  
219 inform mechanism. For example, heritable variation in cell type composition can be derived from transcript  
220 loadings. We noted earlier that immune activation in the adipose tissues was an important driver of obesity  
221 in the DO population. To determine whether this is reflected as an increase in macrophages in adipose  
222 tissue, we compared loadings of cell-type specific genes in adipose tissue (Methods). The mean loading  
223 of macrophage-specific genes was substantially greater than 0 (Fig. 8A), indicating that obese mice were

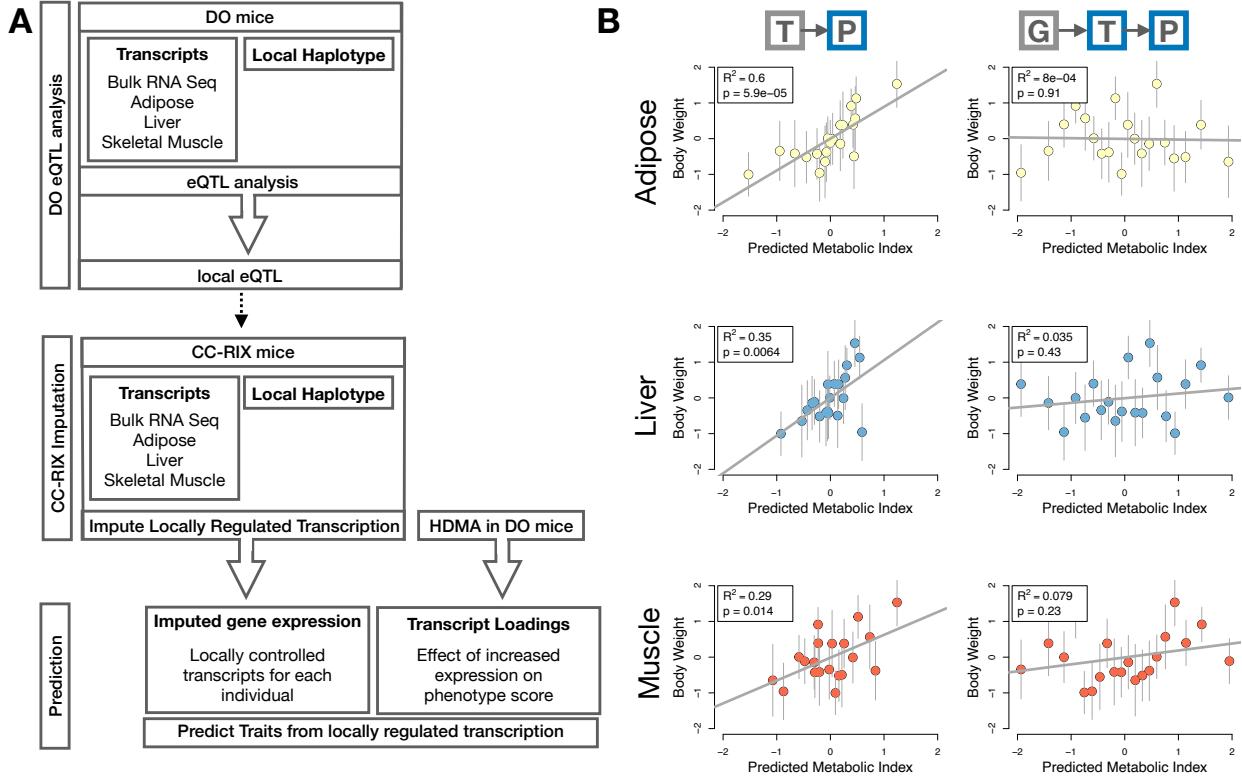


Figure 7: Transcription, but not local genotype, predicts phenotype in the CC-RIX. **A.** Workflow showing procedure for translating HDM results to an independent population of mice. **B.** Relationships between the predicted metabolic index and measured body weight. The left column shows the predictions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

- 224 genetically predisposed to have high levels of macrophage infiltration in adipose tissue in response to the  
 225 high-fat, high-sugar diet.
- 226 We also compared loadings of cell-type specific transcripts in islet (Methods). The mean loadings for alpha-cell  
 227 specific transcripts were significantly greater than 0, while the mean loadings for delta- and endothelial-cell  
 228 specific genes were significantly less than 0 (Fig. 8B). These results suggest that obese mice had inherited  
 229 higher proportions of alpha cells, and lower proportions of endothelial and delta cells in their pancreatic islets.
- 230 The loadings for pancreatic beta cell-type specific loadings was not significantly different from zero. This is  
 231 not necessarily reflective of the function of the beta cells in the obese mice, but rather suggests that any  
 232 variation in the number of beta cells in these mice was unrelated to obesity and insulin resistance.

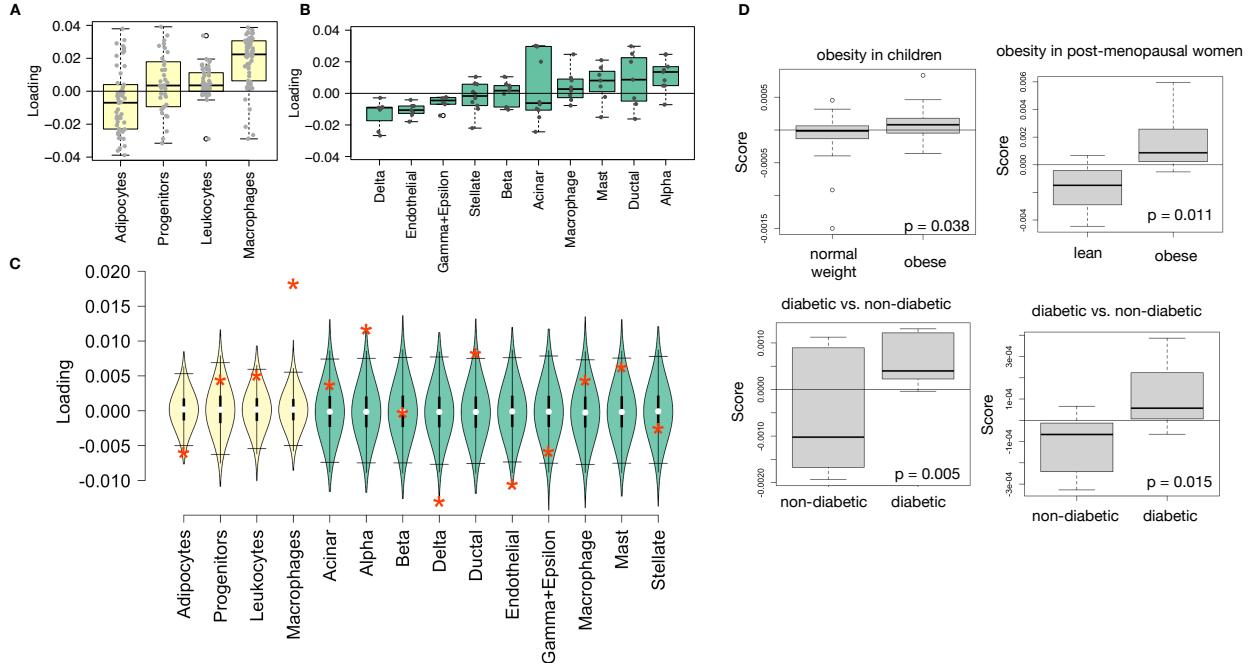


Figure 8: HDM results translate to humans. **A.** Distribution of loadings for cell-type-specific transcripts in adipose tissue. **B.** Distribution of loadings for cell-type-specific transcripts in pancreatic islets (green). **C.** Null distributions for the mean loading of randomly selected transcripts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

### 233 Heritable transcriptomic signatures translated to human disease

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

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

245 **Targeting gene signatures**

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

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

257 • **Supplemental Table** results from CMAP

258 **Discussion**

259 Yao *et al.* [19] observed that genes with low local heritability explained more expression-mediated disease  
260 heritability than genes with high local heritability. This observation is consistent with principles of robustness  
261 in complex systems. If a transcript were both important to a trait and subject to strong local regulation,  
262 a population would be susceptible to extremes in phenotype that might frequently cross the threshold to  
263 disease. Indeed, strong disruption of highly trait-relevant genes is the cause of Mendelian disease.

264 Rather, observations suggest that genes near GWAS hits and have obvious functional relevance to a trait  
265 tend to have highly complex regulatory landscapes under strong selection pressures [18]. In contrast, genes  
266 with strong local regulation tend to be depleted of functional annotations and are under looser selection  
267 constraints [18]. These observations and others led Liu *et al.* [35] to suggest that most heritability of complex  
268 traits is driven by weak trans-eQTLs. They proposed a framework of understanding heritability of complex  
269 traits in which massive polygenicity is distributed across common variants in both functional “core genes”,  
270 as well as more peripheral genes that may not seem obviously related to the trait.

271 We developed high-dimension mediation to test the omnigenic model with a more holistic approach. This  
272 model posits that once the expression of the core genes (i.e. trait-mediating genes) is accounted for, there  
273 should be no residual correlation between the genome and the phenotype. This hypothesis lends itself well to  
274 systems approaches that can account for arbitrarily complex gene regulation, as well as the interconnectedness

275 and redundancy of the transcriptome without explicitly modeling them. The HDM approach we propose here  
276 tests the hypothesis of the omnigenic model

277 Discussion about biological relevance of high-loading transcripts in adipose tissue. TCA cycle and BCAA  
278 catabolism: Selective PPAR $\gamma$  modulation by insulin-sensitizing thiazolidinedione drugs has further been  
279 shown to influence both inflammation and BCAA metabolism in obese rats suggesting a relationship between  
280 these pathways and insulin resistance [36]. BCAA levels are also related to insulin resistance in human  
281 subjects and are elevated in insulin-resistant obese individuals relative to weight-matched non-insulin resistant  
282 individuals [37]. In the DO mice studied here, inheriting increased expression of genes involved in BCAA  
283 catabolism was associated with reduced body weight and insulin resistance.

- 284 • distal heritability correlates with phenotype relevance  
285 • others who use local eQTL to associate genotype with traits often say “we nominated this gene” even  
286 though other nearby genes have higher eQTL LOD scores (27019110, 31465442) Our method supports  
287 the idea that the transcripts with the strongest local regulation are less likely to be functionally related  
288 to the trait

## 289 **Data Availability**

290 Here we tell people where to find the data

## 291 **Acknowledgements**

292 Here we thank people

293 **Supplemental Figures**

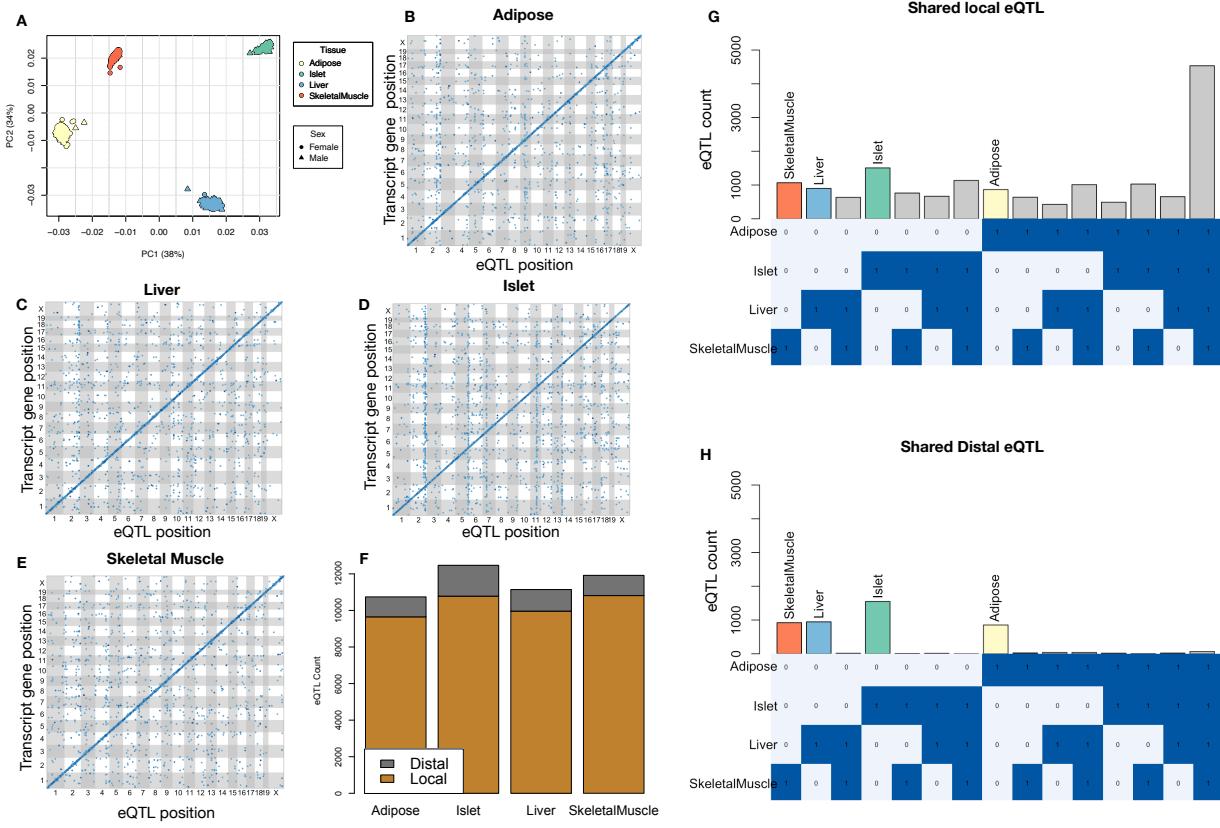


Figure 9: Overview of eQTL analysis in DO mice. **A.** RNA seq samples from the four different tissues clustered by tissue. **B.-E.** eQTL maps are shown for each tissue. The *x*-axis shows the position of the mapped eQTL, and the *y*-axis shows the physical position of the gene encoding each mapped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

## KEGG pathway enrichments by GSEA

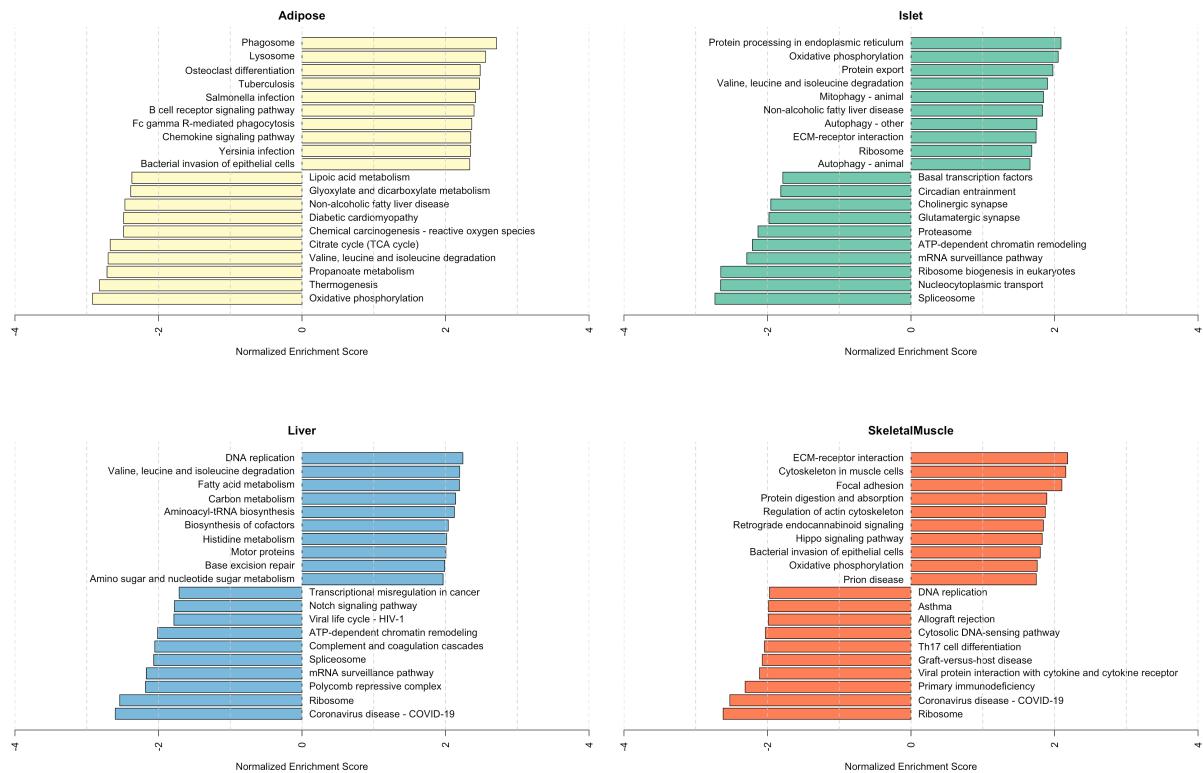


Figure 10: Bar plots showing normalized enrichment scores (NES) for KEGG pathways as determined by fast gene score enrichment analysis (fgsea). Only the top 10 positive and top 10 negative scores are shown. Colors indicate tissue. The name beside each bar shows the name of each enriched KEGG pathway.

## References

- 294 [1] M. T. Maurano, R. Humbert, E. Rynes, R. E. Thurman, E. Haugen, H. Wang, A. P. Reynolds,  
 295 R. Sandstrom, H. Qu, J. Brody, A. Shafer, F. Neri, K. Lee, T. Kutyavin, S. Stehling-Sun, A. K.  
 296 Johnson, T. K. Canfield, E. Giste, M. Diegel, D. Bates, R. S. Hansen, S. Neph, P. J. Sabo, S. Heimfeld,  
 297 A. Raubitschek, S. Ziegler, C. Cotsapas, N. Sotoodehnia, I. Glass, S. R. Sunyaev, R. Kaul, and J. A.  
 298 Stamatoyannopoulos. Systematic localization of common disease-associated variation in regulatory DNA.  
 299 *Science*, 337(6099):1190–1195, Sep 2012.
- 300  
 301 [2] K. K. Farh, A. Marson, J. Zhu, M. Kleinewietfeld, W. J. Housley, S. Beik, N. Shores, H. Whitton, R. J.  
 302 Ryan, A. A. Shishkin, M. Hatan, M. J. Carrasco-Alfonso, D. Mayer, C. J. Luckey, N. A. Patsopoulos,  
 303 P. L. De Jager, V. K. Kuchroo, C. B. Epstein, M. J. Daly, D. A. Hafler, and B. E. Bernstein. Genetic

## Top GO term enrichments by GSEA

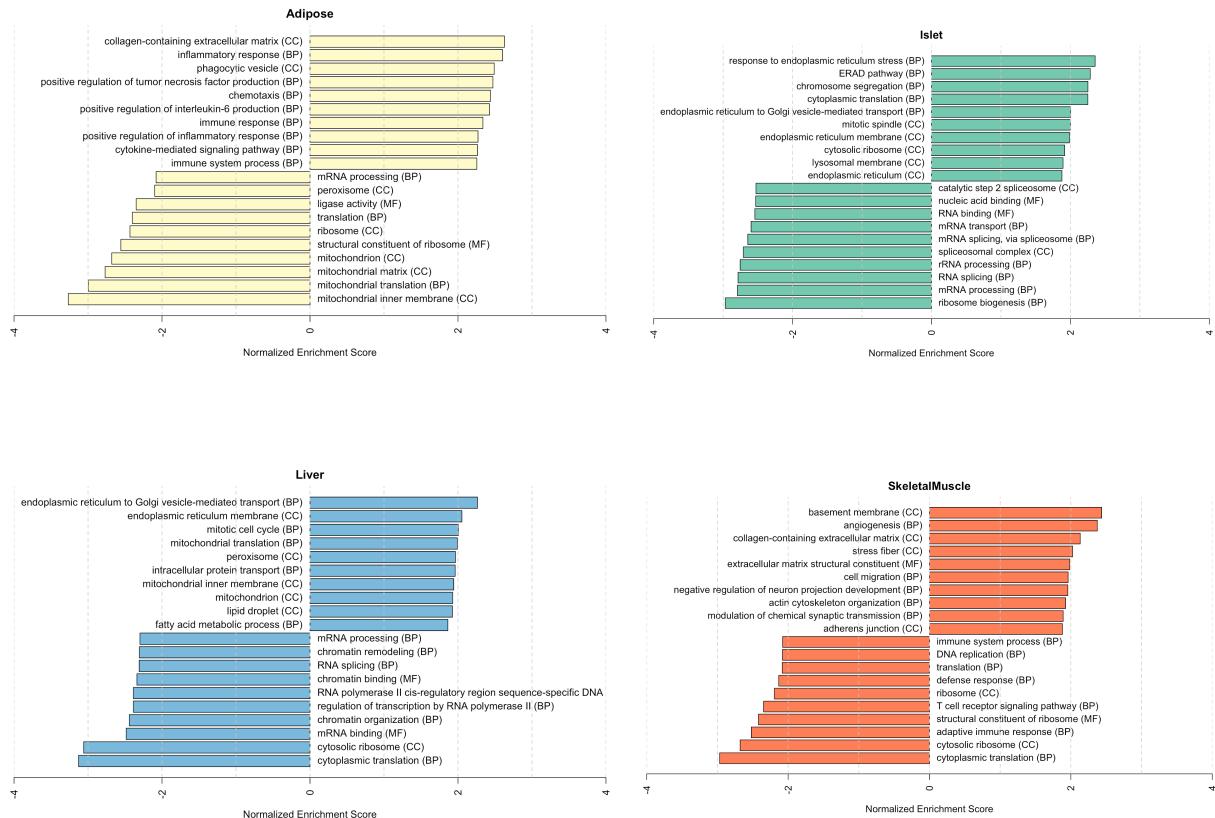


Figure 11: Bar plots showing normalized enrichment scores (NES) for GO terms as determined by fast gene score enrichment analysis (fgsea). Only the top 10 positive and top 10 negative scores are shown. Colors indicate tissue. The name beside each bar shows the name of each enriched GO term. The letters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

- 304 and epigenetic fine mapping of causal autoimmune disease variants. *Nature*, 518(7539):337–343, Feb  
 305 2015.
- 306 [3] E. Pennisi. The Biology of Genomes. Disease risk links to gene regulation. *Science*, 332(6033):1031, May  
 307 2011.
- 308 [4] L. A. Hindorff, P. Sethupathy, H. A. Junkins, E. M. Ramos, J. P. Mehta, F. S. Collins, and T. A. Manolio.  
 309 Potential etiologic and functional implications of genome-wide association loci for human diseases and  
 310 traits. *Proc Natl Acad Sci*, 106(23):9362–9367, Jun 2009.
- 311 [5] J. K. Pickrell. Joint analysis of functional genomic data and genome-wide association studies of 18  
 312 human traits. *Am J Hum Genet*, 94(4):559–573, Apr 2014.

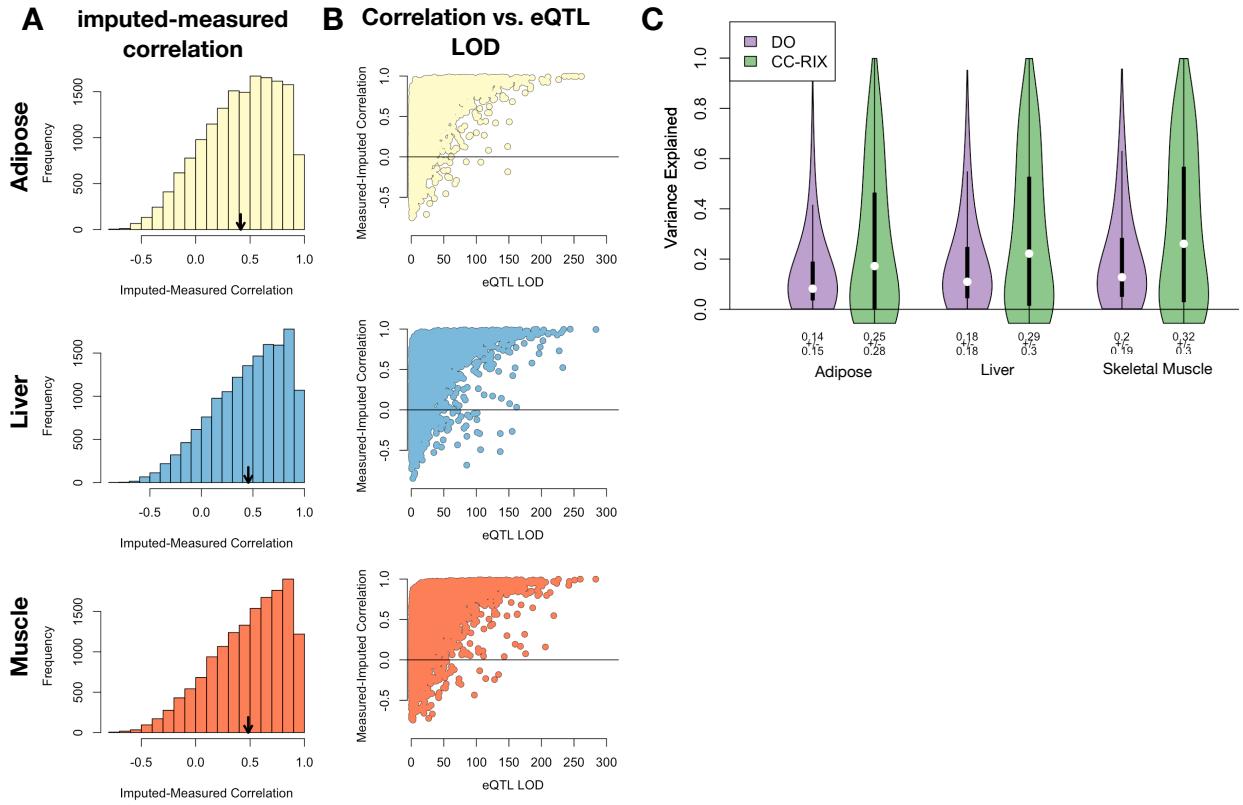


Figure 12: Validation of transcript imputation in the CC-RIX. **A.** Distributions of correlations between imputed and measured transcripts in the CC-RIX. The mean of each distribution is shown by the red line. All distributions were skewed toward positive correlations and had positive means near a Pearson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

- 313 [6] D. Welter, J. MacArthur, J. Morales, T. Burdett, P. Hall, H. Junkins, A. Klemm, P. Flicek, T. Manolio,  
 314 L. Hindorff, and H. Parkinson. The NHGRI GWAS Catalog, a curated resource of SNP-trait associations.  
 315 *Nucleic Acids Res*, 42(Database issue):D1001–1006, Jan 2014.
- 316 [7] Y. I. Li, B. van de Geijn, A. Raj, D. A. Knowles, A. A. Petti, D. Golan, Y. Gilad, and J. K. Pritchard.  
 317 RNA splicing is a primary link between genetic variation and disease. *Science*, 352(6285):600–604, Apr  
 318 2016.
- 319 [8] D. Zhou, Y. Jiang, X. Zhong, N. J. Cox, C. Liu, and E. R. Gamazon. A unified framework for joint-tissue  
 320 transcriptome-wide association and Mendelian randomization analysis. *Nat Genet*, 52(11):1239–1246,  
 321 Nov 2020.
- 322 [9] E. R. Gamazon, H. E. Wheeler, K. P. Shah, S. V. Mozaffari, K. Aquino-Michaels, R. J. Carroll, A. E.  
 323 Eyler, J. C. Denny, D. L. Nicolae, N. J. Cox, and H. K. Im. A gene-based association method for

- 324 mapping traits using reference transcriptome data. *Nat Genet*, 47(9):1091–1098, Sep 2015.
- 325 [10] Z. Zhu, F. Zhang, H. Hu, A. Bakshi, M. R. Robinson, J. E. Powell, G. W. Montgomery, M. E. Goddard,  
326 N. R. Wray, P. M. Visscher, and J. Yang. Integration of summary data from GWAS and eQTL studies  
327 predicts complex trait gene targets. *Nat Genet*, 48(5):481–487, May 2016.
- 328 [11] A. Gusev, A. Ko, H. Shi, G. Bhatia, W. Chung, B. W. Penninx, R. Jansen, E. J. de Geus, D. I. Boomsma,  
329 F. A. Wright, P. F. Sullivan, E. Nikkola, M. Alvarez, M. Civelek, A. J. Lusis, T. ki, E. Raitoharju,  
330 M. nen, I. ä, O. T. Raitakari, J. Kuusisto, M. Laakso, A. L. Price, P. Pajukanta, and B. Pasaniuc.  
331 Integrative approaches for large-scale transcriptome-wide association studies. *Nat Genet*, 48(3):245–252,  
332 Mar 2016.
- 333 [12] M. P. Keller, D. M. Gatti, K. L. Schueler, M. E. Rabaglia, D. S. Stapleton, P. Simecek, M. Vincent,  
334 S. Allen, A. T. Broman, R. Bacher, C. Kendziora, K. W. Broman, B. S. Yandell, G. A. Churchill, and  
335 A. D. Attie. Genetic Drivers of Pancreatic Islet Function. *Genetics*, 209(1):335–356, May 2018.
- 336 [13] W. L. Crouse, G. R. Keele, M. S. Gastonguay, G. A. Churchill, and W. Valdar. A Bayesian model  
337 selection approach to mediation analysis. *PLoS Genet*, 18(5):e1010184, May 2022.
- 338 [14] J. M. Chick, S. C. Munger, P. Simecek, E. L. Huttlin, K. Choi, D. M. Gatti, N. Raghupathy, K. L. Svenson,  
339 G. A. Churchill, and S. P. Gygi. Defining the consequences of genetic variation on a proteome-wide scale.  
340 *Nature*, 534(7608):500–505, Jun 2016.
- 341 [15] H. E. Wheeler, S. Ploch, A. N. Barbeira, R. Bonazzola, A. Andaleon, A. Fotuhi Siahpirani, A. Saha,  
342 A. Battle, S. Roy, and H. K. Im. Imputed gene associations identify replicable trans-acting genes enriched  
343 in transcription pathways and complex traits. *Genet Epidemiol*, 43(6):596–608, Sep 2019.
- 344 [16] B. D. Umans, A. Battle, and Y. Gilad. Where Are the Disease-Associated eQTLs? *Trends Genet*,  
345 37(2):109–124, Feb 2021.
- 346 [17] N. J. Connally, S. Nazeen, D. Lee, H. Shi, J. Stamatoyannopoulos, S. Chun, C. Cotsapas, C. A. Cassa,  
347 and S. R. Sunyaev. The missing link between genetic association and regulatory function. *Elife*, 11, Dec  
348 2022.
- 349 [18] H. Mostafavi, J. P. Spence, S. Naqvi, and J. K. Pritchard. Systematic differences in discovery of genetic  
350 effects on gene expression and complex traits. *Nat Genet*, 55(11):1866–1875, Nov 2023.
- 351 [19] D. W. Yao, L. J. O’Connor, A. L. Price, and A. Gusev. Quantifying genetic effects on disease mediated  
352 by assayed gene expression levels. *Nat Genet*, 52(6):626–633, Jun 2020.

- 353 [20] X. Liu, J. A. Mefford, A. Dahl, Y. He, M. Subramaniam, A. Battle, A. L. Price, and N. Zaitlen. GBAT:  
354 a gene-based association test for robust detection of trans-gene regulation. *Genome Biol*, 21(1):211, Aug  
355 2020.
- 356 [21] G. A. Churchill, D. M. Gatti, S. C. Munger, and K. L. Svenson. The Diversity Outbred mouse population.  
357 *Mamm Genome*, 23(9-10):713–718, Oct 2012.
- 358 [22] S. M. Clee and A. D. Attie. The genetic landscape of type 2 diabetes in mice. *Endocr Rev*, 28(1):48–83,  
359 Feb 2007.
- 360 [23] K. W. Broman, D. M. Gatti, P. Simecek, N. A. Furlotte, P. Prins, Š. Sen, B. S. Yandell, and G. A.  
361 Churchill. R/qtL2: Software for Mapping Quantitative Trait Loci with High-Dimensional Data and  
362 Multiparent Populations. *Genetics*, 211(2):495–502, Feb 2019.
- 363 [24] Fabien Girka, Etienne Camenen, Caroline Peltier, Arnaud Gloaguen, Vincent Guillemot, Laurent Le  
364 Brusquet, and Arthur Tenenhaus. *RGCCA: Regularized and Sparse Generalized Canonical Correlation*  
365 *Analysis for Multiblock Data*, 2023. R package version 3.0.3.
- 366 [25] Gennady Korotkevich, Vladimir Sukhov, and Alexey Sergushichev. Fast gene set enrichment analysis.  
367 *bioRxiv*, 2019.
- 368 [26] A. Subramanian, P. Tamayo, V. K. Mootha, S. Mukherjee, B. L. Ebert, M. A. Gillette, A. Paulovich,  
369 S. L. Pomeroy, T. R. Golub, E. S. Lander, and J. P. Mesirov. Gene set enrichment analysis: a  
370 knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci U S A*,  
371 102(43):15545–15550, Oct 2005.
- 372 [27] C. B. Newgard. Interplay between lipids and branched-chain amino acids in development of insulin  
373 resistance. *Cell Metab*, 15(5):606–614, May 2012.
- 374 [28] D. D. Sears, G. Hsiao, A. Hsiao, J. G. Yu, C. H. Courtney, J. M. Ofrecio, J. Chapman, and S. Subramaniam.  
375 Mechanisms of human insulin resistance and thiazolidinedione-mediated insulin sensitization. *Proc Natl*  
376 *Acad Sci U S A*, 106(44):18745–18750, Nov 2009.
- 377 [29] R. Stienstra, C. Duval, M. ller, and S. Kersten. PPARs, Obesity, and Inflammation. *PPAR Res*,  
378 2007:95974, 2007.
- 379 [30] O. Gavrilova, M. Haluzik, K. Matsusue, J. J. Cutson, L. Johnson, K. R. Dietz, C. J. Nicol, C. Vinson,  
380 F. J. Gonzalez, and M. L. Reitman. Liver peroxisome proliferator-activated receptor gamma contributes  
381 to hepatic steatosis, triglyceride clearance, and regulation of body fat mass. *J Biol Chem*, 278(36):34268–  
382 34276, Sep 2003.

- 383 [31] K. Matsusue, M. Haluzik, G. Lambert, S. H. Yim, O. Gavrilova, J. M. Ward, B. Brewer, M. L. Reitman,  
384 and F. J. Gonzalez. Liver-specific disruption of PPARgamma in leptin-deficient mice improves fatty  
385 liver but aggravates diabetic phenotypes. *J Clin Invest*, 111(5):737–747, Mar 2003.
- 386 [32] D. Patsouris, J. K. Reddy, M. Iler, and S. Kersten. Peroxisome proliferator-activated receptor alpha  
387 mediates the effects of high-fat diet on hepatic gene expression. *Endocrinology*, 147(3):1508–1516, Mar  
388 2006.
- 389 [33] S. E. Schadinger, N. L. Bucher, B. M. Schreiber, and S. R. Farmer. PPARgamma2 regulates lipogenesis  
390 and lipid accumulation in steatotic hepatocytes. *Am J Physiol Endocrinol Metab*, 288(6):E1195–1205,  
391 Jun 2005.
- 392 [34] W. Motomura, M. Inoue, T. Ohtake, N. Takahashi, M. Nagamine, S. Tanno, Y. Kohgo, and T. Okumura.  
393 Up-regulation of ADRP in fatty liver in human and liver steatosis in mice fed with high fat diet. *Biochem  
394 Biophys Res Commun*, 340(4):1111–1118, Feb 2006.
- 395 [35] X. Liu, Y. I. Li, and J. K. Pritchard. Trans Effects on Gene Expression Can Drive Omnipenic Inheritance.  
396 *Cell*, 177(4):1022–1034, May 2019.
- 397 [36] G. Hsiao, J. Chapman, J. M. Ofrecio, J. Wilkes, J. L. Resnik, D. Thapar, S. Subramaniam, and D. D.  
398 Sears. modulation of insulin sensitivity and metabolic pathways in obese rats. *Am J Physiol Endocrinol  
399 Metab*, 300(1):E164–174, Jan 2011.
- 400 [37] D. E. Lackey, C. J. Lynch, K. C. Olson, R. Mostaedi, M. Ali, W. H. Smith, F. Karpe, S. Humphreys,  
401 D. H. Bedinger, T. N. Dunn, A. P. Thomas, P. J. Oort, D. A. Kieffer, R. Amin, A. Bettaieb, F. G.  
402 Haj, P. Permana, T. G. Anthony, and S. H. Adams. Regulation of adipose branched-chain amino acid  
403 catabolism enzyme expression and cross-adipose amino acid flux in human obesity. *Am J Physiol  
404 Endocrinol Metab*, 304(11):E1175–1187, Jun 2013.