

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<sup>1–7</sup>. It is widely assumed  
14 that these variants influence local transcription, and methods such as transcriptome-wide association studies  
15 (TWAS)<sup>8–11</sup>, summary data-based Mendelian randomization (SMR)<sup>10</sup>, and others have capitalized on this  
16 idea to identify genes associated with multiple disease traits<sup>12–15</sup>.

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  
20 effects on gene expression<sup>16</sup> and tend not to co-localize with expression quantitative trait loci (eQTLs)<sup>17;18</sup>.

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<sup>16</sup>. An alternative explanation that has  
23 been discussed in recent years is that effects of these variants are mediated not through local regulation of  
24 gene expression, but through distal regulation<sup>18–20;15</sup>.

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

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

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

To assess the role of gene regulation in mediating variation in metabolic traits in this population, we propose high-dimensional mediation (HDM). In univariate approaches, such as TWAS, SMR, and other Mendelian randomization approaches, each transcript is tested independently for mediation of a local variant on a trait. This process requires huge numbers of statistical tests, which is computationally expensive, requires strict corrections for multiple testing, and assumes independence of genetic variants and transcripts. Such methods are therefore limited to detecting only the largest statistical effects and are biased toward local gene regulation. In contrast, with high-dimensional mediation we assessed broad relationships among the genome, transcriptome, and phenotype as a whole and identified a highly heritable composite trait that was perfectly mediated by a composite transcript. We show that composite transcripts were tissue-specific and highly interpretable in terms of biological processes as well as cell type composition. Heritability analysis of the transcripts showed that the strongest transcriptional mediators of metabolic disease had low local heritability and high distal heritability. Finally, we show that the composite transcripts identified in the DO population predicted obesity in an independent population of Collaborative Cross recombinant inbred (CC-RIX) mice and in human subjects. In contrast, local eQTL were unable to predict obesity in the CC-RIX mice. Together our results suggest that both the tissue used for gene expression analysis as well as distal gene regulation are 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:  $\text{FBG}$   
66  $\geq 300, \text{ mg/dL}$ )<sup>22</sup>. 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<sup>23</sup> (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,

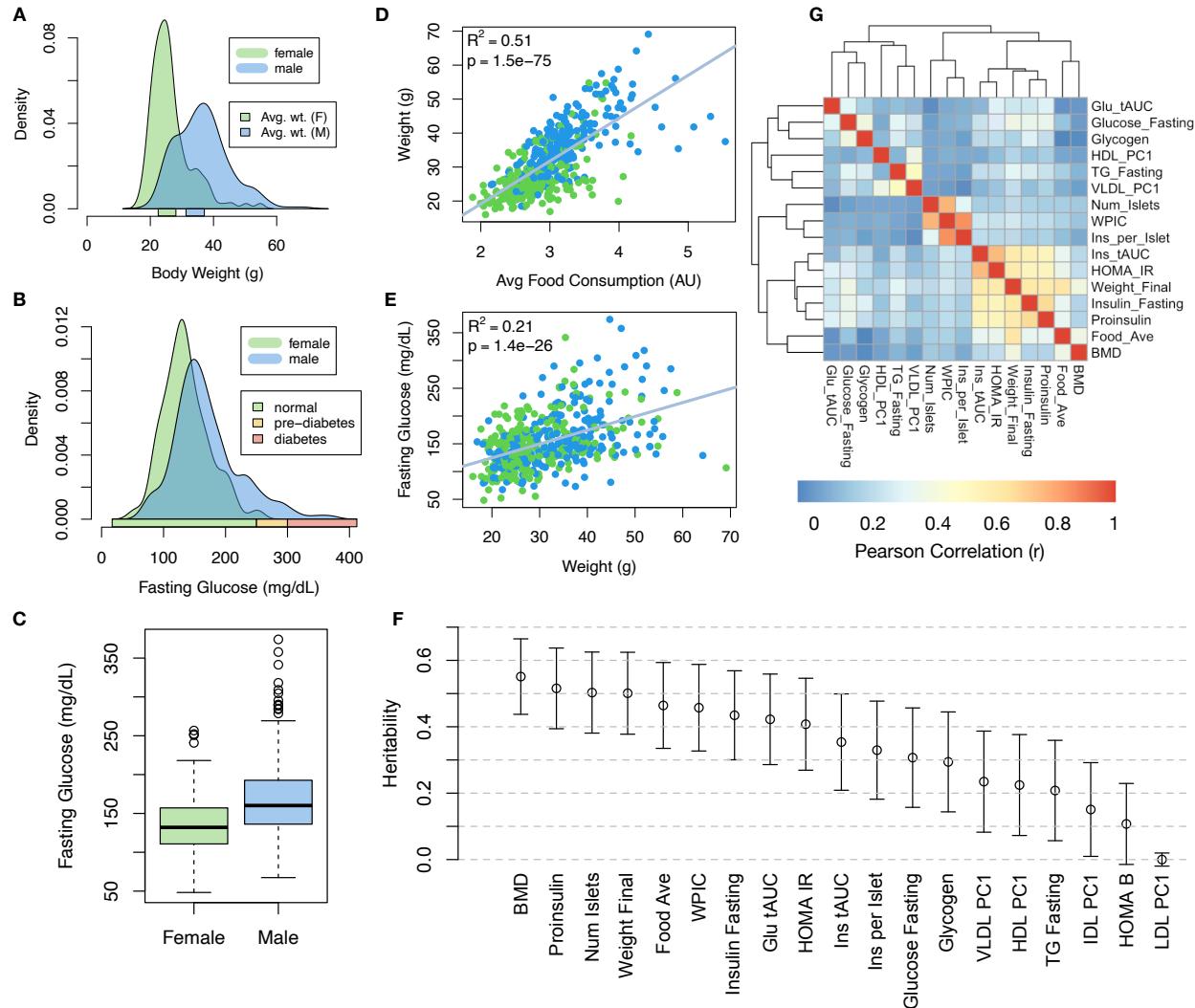


Figure 1: Clinical overview. **A.** Distributions of final body weight in the diversity outbred mice. Sex is indicated by color. The average B6 male and female adult weights at 24 weeks of age are indicated by blue and green bars on the x-axis. **B.** The distribution of final fasting glucose across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

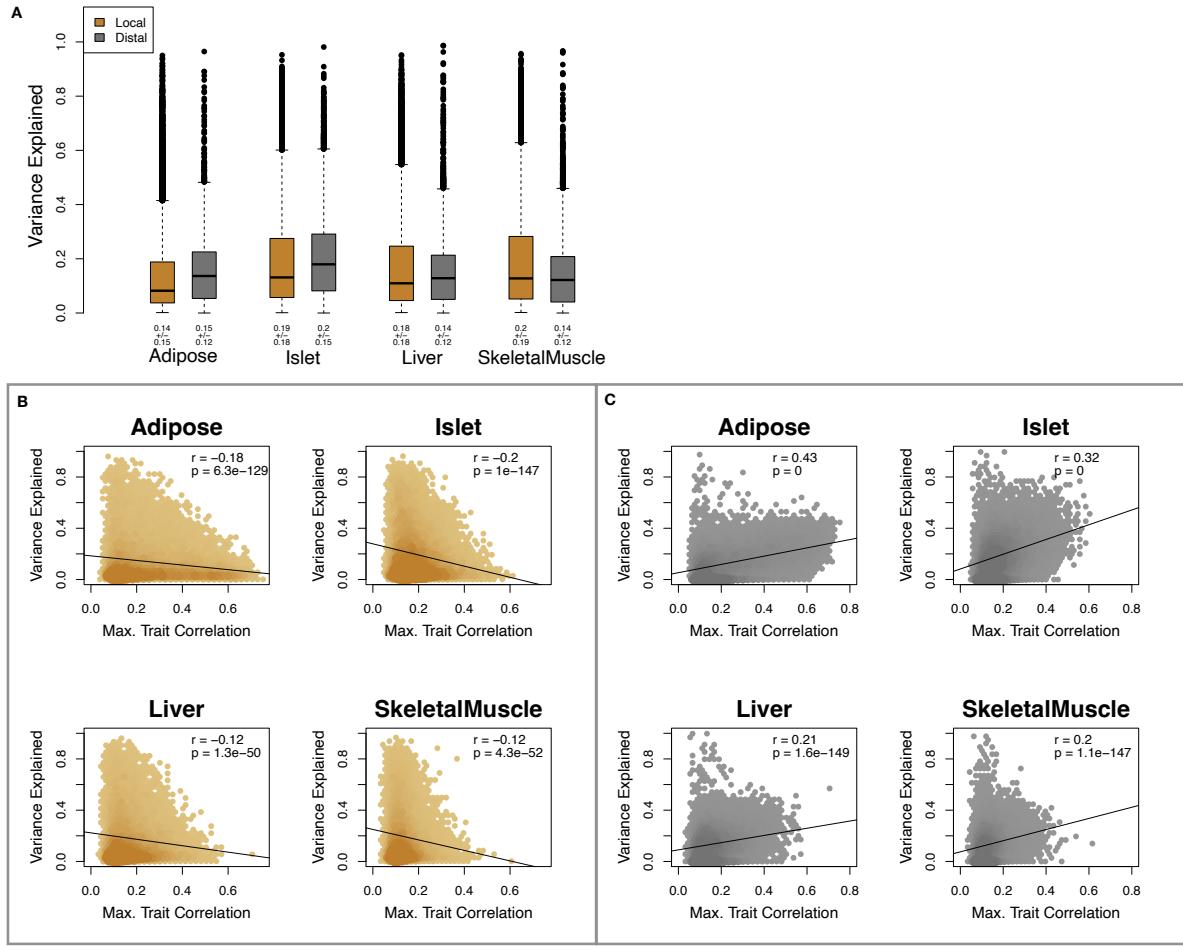


Figure 2: Transcript heritability and trait relevance. **A.** Distributions of distal and local heritability of transcripts across the four tissues. Overall local and distal factors contribute equally to transcript heritability. The relationship between **(B.)** local and **(C.)** distal heritability 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)<sup>24</sup> 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

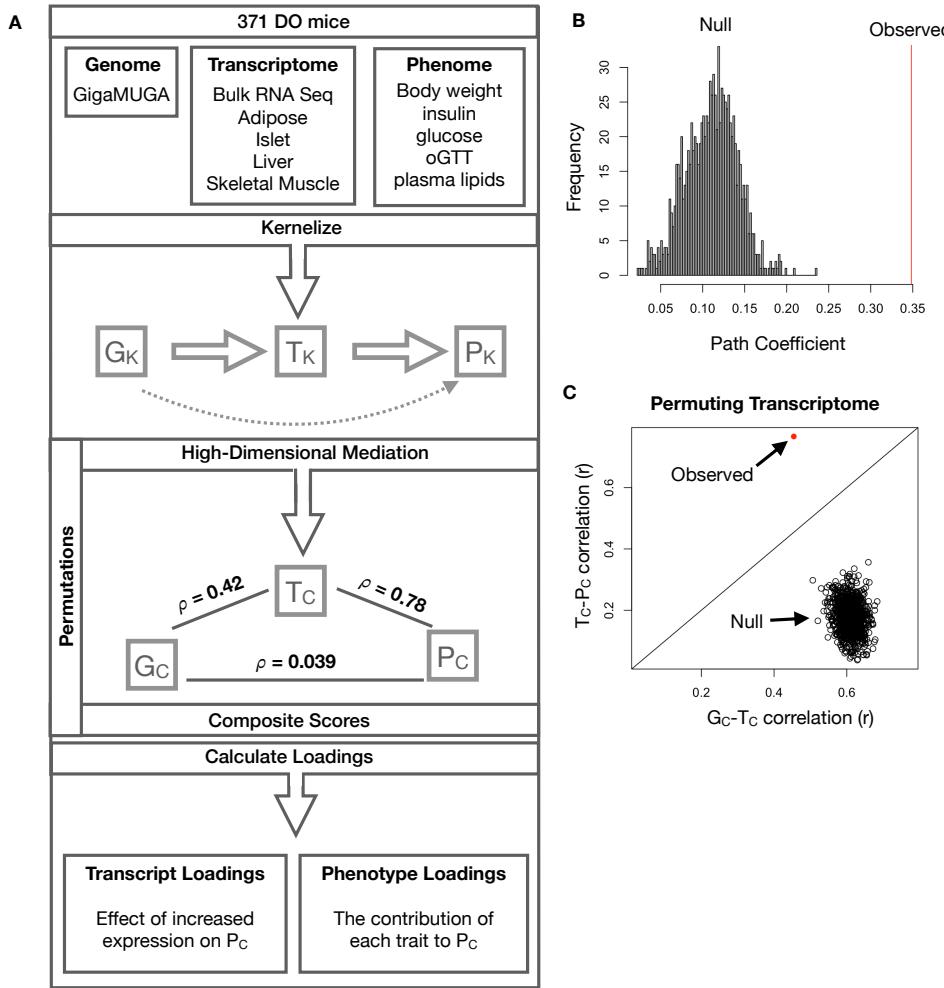


Figure 3: High-dimensional mediation. **A.** Workflow indicating major steps of high-dimensional mediation. The genotype, transcriptome, and phenotype matrices were kernelized to yield single matrices representing the relationships between all individuals for each data modality ( $G_K$  = genome kernel,  $T_K$  = 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)<sup>25;26</sup> to look for biological processes and pathways that were enriched at the top  
138 and bottom of this list (Methods).

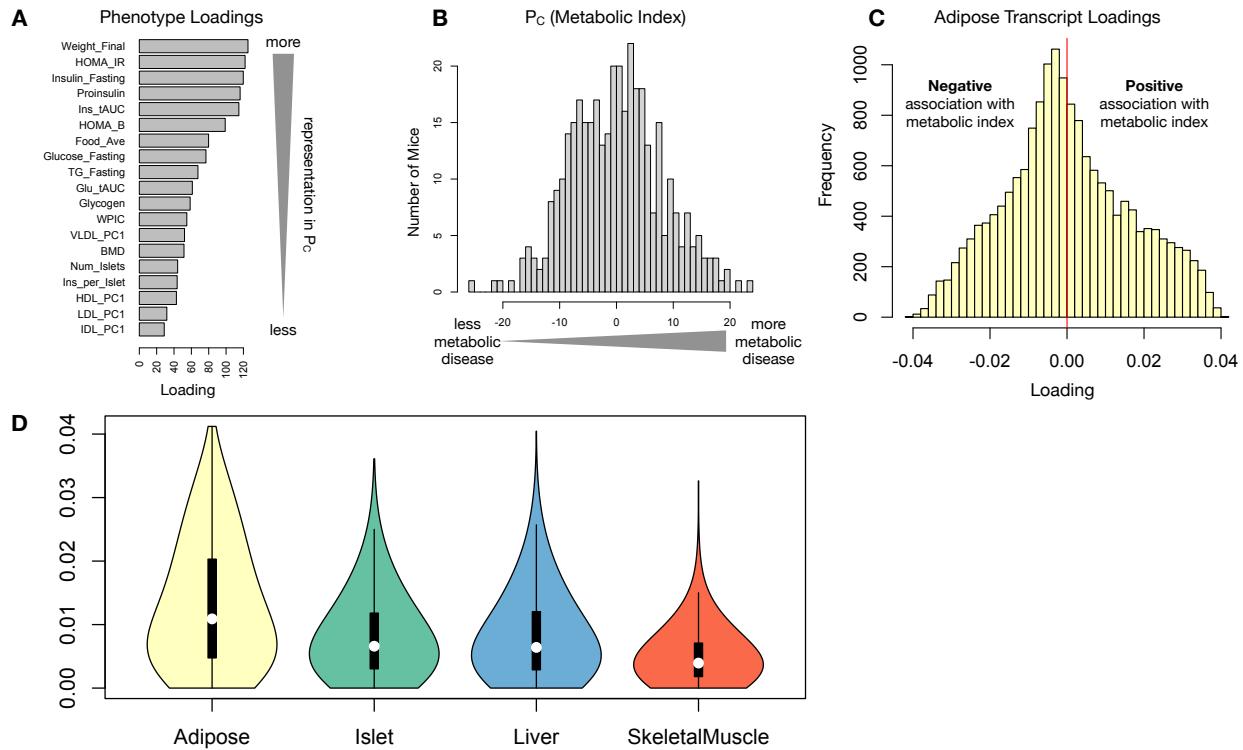


Figure 4: Interpretation of loadings. **A.** Loadings across traits. Body weight and insulin resistance contributed the most to the composite trait. **B.** Phenotype scores across individuals. Individuals with large positive phenotype scores had higher body weight and insulin resistance than average. 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<sup>12;27;28</sup>, suggesting that heritable variation in regulation of these  
152 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<sup>29</sup>. Agonists of *Pparg*, such as Thiazolidinediones, which are FDA-approved to treat type II  
178 diabetes, reduce inflammation and adipose hypertrophy<sup>29</sup>. 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).

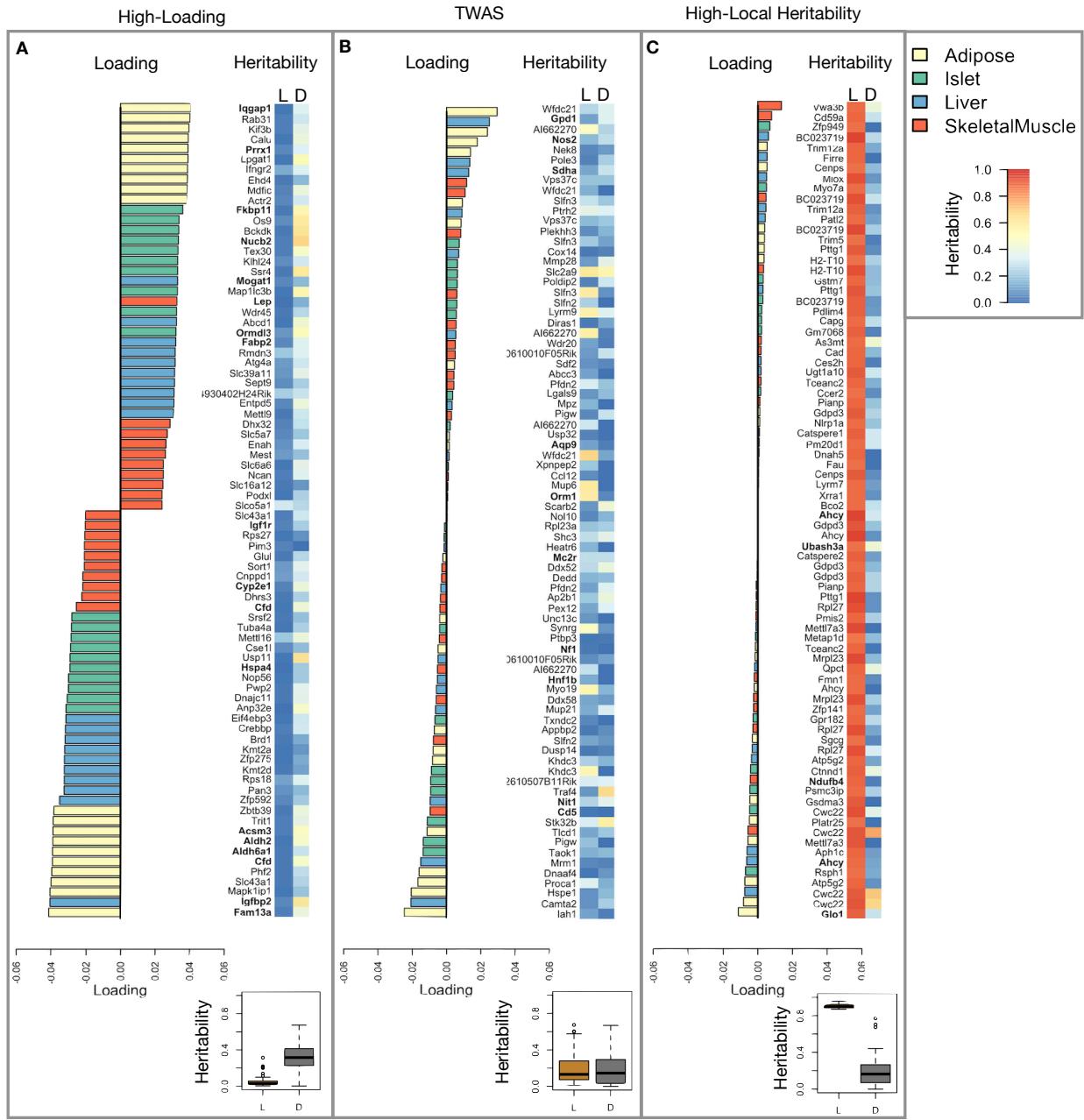


Figure 5: Transcripts with high loadings have high distal heritability and literature support. Each panel has a bar plot showing the loadings of transcripts selected by different criteria. Bar color indicates the tissue of origin. The heat map shows the local (L - left) and distal (D - right) heritability of each transcript. **A.** Loadings for the 10 transcripts with the largest positive loadings and the 10 transcripts with the largest negative loadings for each tissue. **B.** Loadings of TWAS candidates with the 10 largest positive correlations with traits and the largest negative correlations with traits across all four tissues. **C.** The transcripts with the largest local heritability (top 20) across all four tissues.

180 In contrast, *Pparg* had a large positive loading in liver, where it is known to play a role in the development of  
 181 hepatic steatosis, or fatty liver. Mice that lack *Pparg* specifically in the liver, are protected from developing  
 182 steatosis and show reduced expression of lipogenic genes<sup>30;31</sup>. Overexpression of *Pparg* in the livers of mice

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

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.

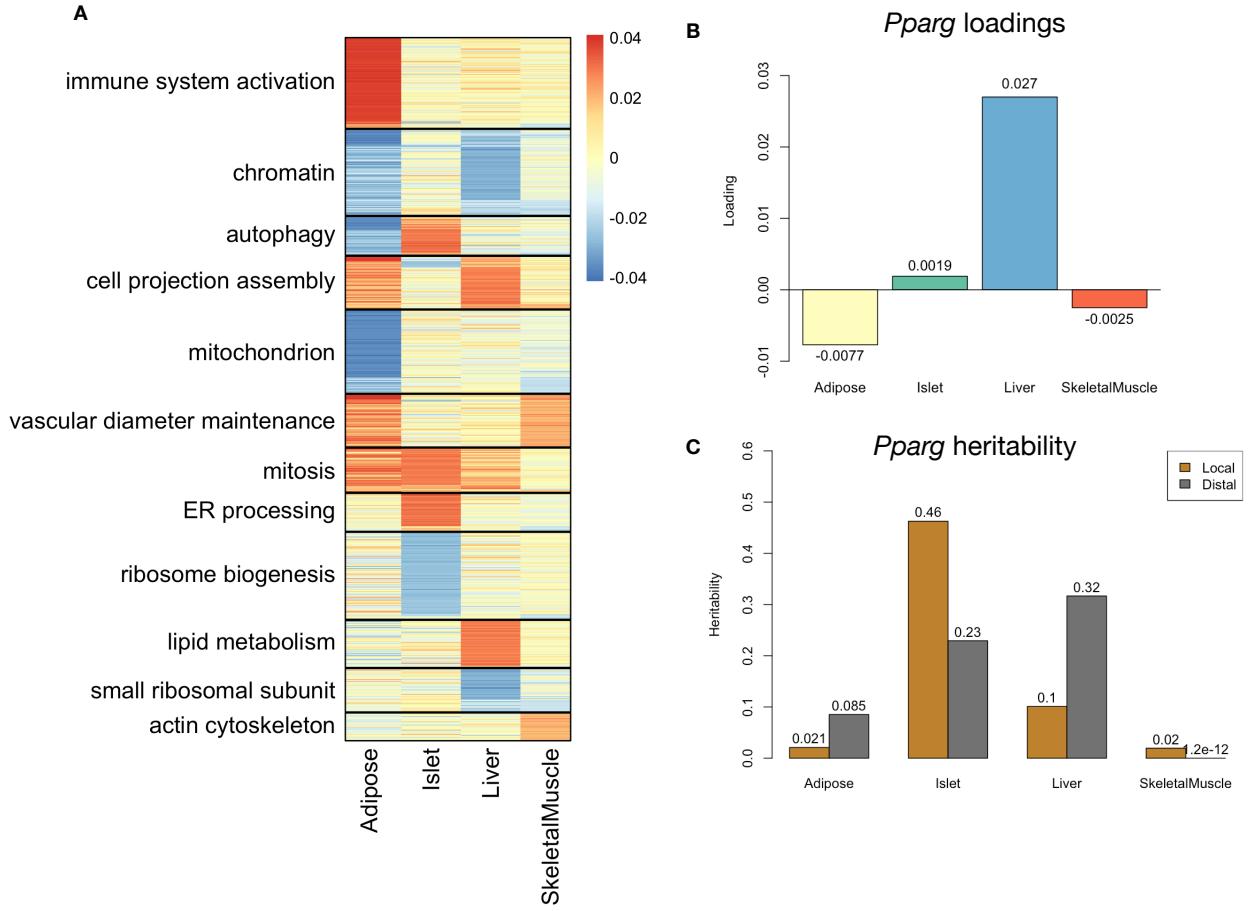


Figure 6: Tissue-specific transcriptional programs were associated with obesity and insulin resistance. **A** Heat map showing the loadings of all transcripts with loadings greater than 2.5 standard deviations from the mean in any tissue. The heat map was clustered using k medoid clustering. Functional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

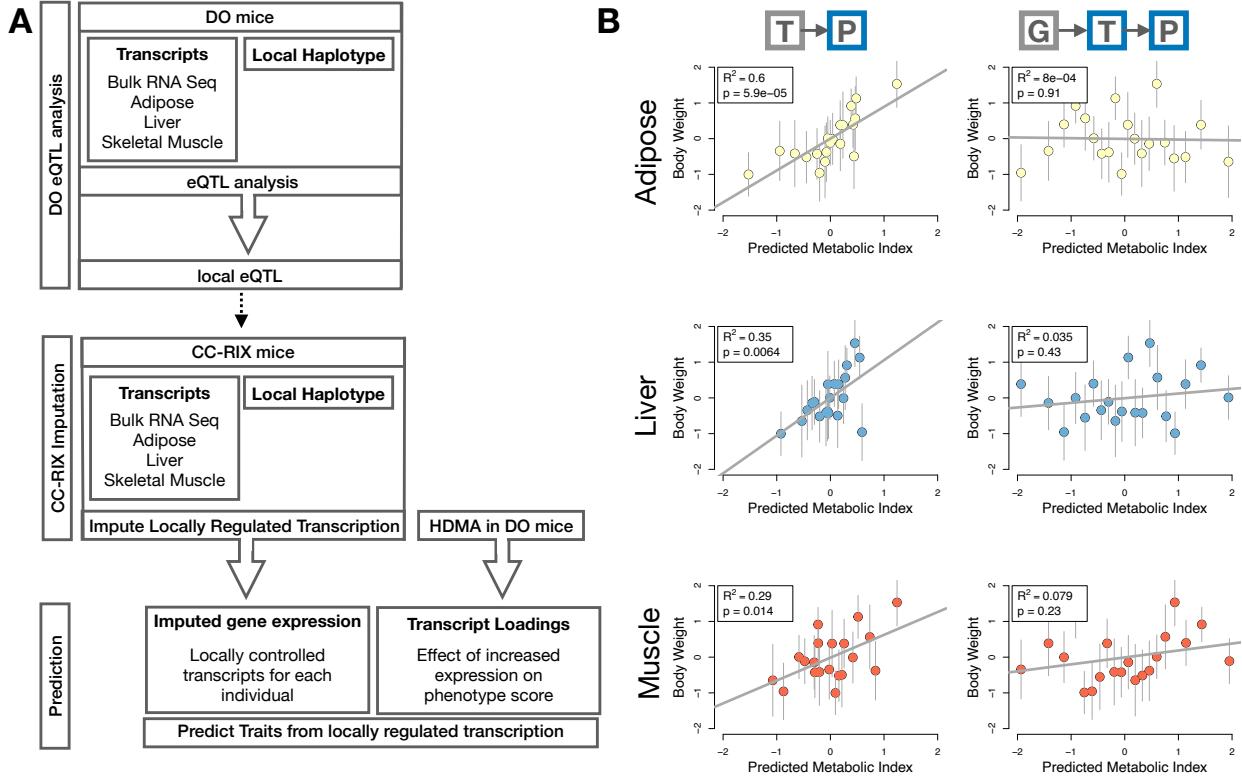


Figure 7: Transcription, but not local genotype, predicts phenotype in the CC-RIX. **A.** Workflow showing procedure for translating HDM results to an independent population of mice. **B.** Relationships between the predicted metabolic index and measured body weight. The left column shows the predictions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

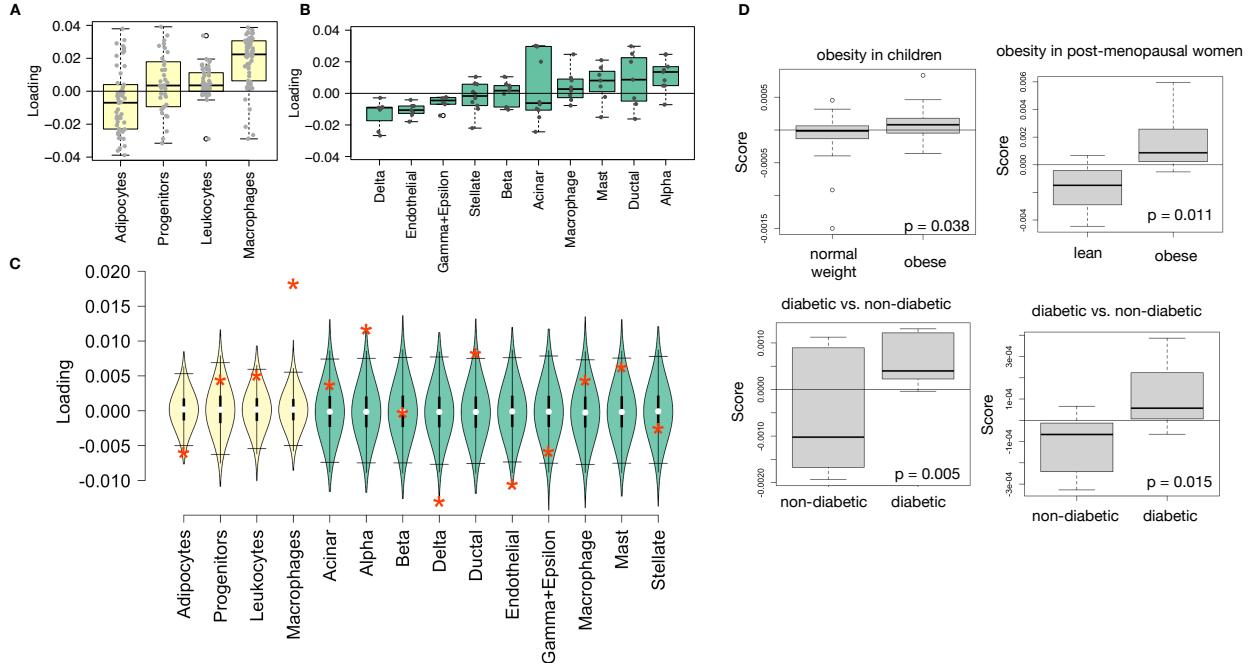


Figure 8: HDM results translate to humans. **A.** Distribution of loadings for cell-type-specific transcripts in adipose tissue. **B.** Distribution of loadings for cell-type-specific transcripts in pancreatic islets (green). **C.** Null distributions for the mean loading of randomly selected transcripts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

<sup>245</sup> **Targeting gene signatures**

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

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

<sup>257</sup> Looking broadly across cell types, the notable top hits from the adipose tissue loadings included mTOR  
<sup>258</sup> inhibitors and glucocorticoid agonists (Supplemental Figure XXX). It is thought that metformin, which  
<sup>259</sup> is commonly used to improve glycemic control, acts, at least in part, by inhibiting mTOR signaling<sup>35;36</sup>.  
<sup>260</sup> However, long-term use of other mTOR inhibitors, such as rapamycin, are known to cause insulin resistance  
<sup>261</sup> and  $\beta$ -cell toxicity<sup>36–38</sup>. Glucocorticoids are used to reduce inflammation, which was a prominent signature  
<sup>262</sup> in the adipose tissues, but these drugs also promote hyperglycemia and diabetes<sup>39;40</sup>. Accute treatment  
<sup>263</sup> with glucocorticoids has further been shown to reduce thermogenesis in rodent adipocytes<sup>41–43</sup>, but increase  
<sup>264</sup> thermogenesis in human adipocytes<sup>44;45</sup>. Thus, the pathways identified by CMAP across all cell types were  
<sup>265</sup> highly related to the transcript loading profiles, but the relationship was not a simple reversal.

<sup>266</sup> The top hit in adipocytes was a PARP inhibitor (Supplemental Figure XXXB). PARPs play a role in lipid  
<sup>267</sup> metabolism and are involved in the development of obesity and diabetes<sup>46</sup>. PARP1 inhibition increases  
<sup>268</sup> mitochondrial biogenesis<sup>47</sup>. Inhibition of PARP1 activity can further prevent necrosis in favor of the less  
<sup>269</sup> inflammatory apoptosis<sup>48</sup>, thereby potentially reducing inflammation in stressed adipocytes. Other notable  
<sup>270</sup> hits in the top 20 were BTK inhibitors, which have been observed to suppress inflammation and improve  
<sup>271</sup> insulin resistance<sup>49</sup> as well as to reduce insulin antibodies in type I diabetes<sup>50</sup>. Similarly, IKK has been  
<sup>272</sup> shown to be associated with insulin resistance<sup>51</sup>, and inhibitors have been shown to improve glucose control  
<sup>273</sup> in type II diabetes<sup>52</sup>.

<sup>274</sup> Among the top hits for the query with transcript loadings from pancreatic islets (Fig. XXX), was suppression  
<sup>275</sup> of T cell receptor signaling, which is known to be involved in Type 1 diabetes<sup>53</sup>, as well as TNFR1, which

276 has been associated with mortality in diabetes patients<sup>54</sup>. Suppression of NOD1/2 signaling was also among  
277 the top hits. NOD1 and 2 sense ER stress<sup>55;56</sup>, which is associated with  $\beta$ -cell death in type 1 and type 2  
278 diabetes<sup>57</sup>. This cell death process is dependent on NOD1/2 signaling<sup>55</sup>, although the specifics have not yet  
279 been worked out.

280 Among the top hits in pancreatic tumor cells were known diabetes drugs, including sulfonylureas, PPAR  
281 receptor agonists, and insulin sensitizers. Rosiglitazone is a PPAR- $\gamma$  agonist and was one of the most  
282 prescribed drugs for type 2 diabetes before its use was reduced due to cardiac side-effects<sup>58</sup>. Sulfonylureas  
283 are another commonly prescribed drug class for type 2 diabetes, but also have notable side effects including  
284 hypoglycemia and accelerated  $\beta$ -cell death<sup>59</sup>.

## 285 Discussion

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

291 Yao *et al.*<sup>19</sup> observed that in humans, transcripts with low local heritability explained more expression-  
292 mediated disease heritability than transcripts with high local heritability. We observed the same trend here  
293 in mice. This pattern is consistent with principles of robustness in complex systems<sup>60–62</sup>. If a transcript  
294 were both important to a trait and subject to strong local regulation, a population would be susceptible to  
295 extremes in phenotype that might frequently cross the threshold to disease. Indeed, strong disruption of  
296 highly trait-relevant genes is the cause of Mendelian disease.

297 Rather, studies suggest that genes that are near GWAS hits and have obvious functional relevance to a trait  
298 tend to have highly complex regulatory landscapes under strong selection pressures<sup>18</sup>. In contrast, genes  
299 with strong local regulation tend to be depleted of functional annotations and are under looser selection  
300 constraints<sup>18</sup>. These observations and others led Liu *et al.*<sup>63</sup> to suggest that most heritability of complex  
301 traits is driven by weak distal eQTLs. They proposed a framework of understanding heritability of complex  
302 traits in which massive polygenicity is distributed across common variants in both functional “core genes”, as  
303 well as more peripheral genes that may not seem obviously related to the trait.

304 Here, we used a large, comprehensive, and purpose-built data set to investigate the genetic architecture of  
305 complex traits related to metabolic disease in mice as well as the roles of local and distal gene regulation in

306 mediating these traits. We presented a systems-level method called high-dimension mediation (HDM). This  
307 approach contrasts with traditional univariate approaches in several important respects. First, in contrast  
308 to univariate approaches, which assume independence of genetic variants and transcripts, HDM allows for  
309 arbitrarily complex gene regulation, as well as the interconnectedness and redundancy of the transcriptome.  
310 Second, rather than assuming a single, large genetic effect as univariate approaches do, HDM assumes that  
311 traits are highly polygenic, and that genetic effects are weak and are distributed across the genome. HDM  
312 does not use statistical thresholds to identify true positive effects, but generates a weighted vector of transcripts  
313 that can be analyzed as a whole, or dissected to identify transcripts with stronger and weaker effects. This  
314 method explicitly models a central proposal of the omnigenic model which posits that once the expression of  
315 the core genes (i.e. trait-mediating genes) is accounted for, there should be no residual correlation between  
316 the genome and the phenotype.

317 Using HDM, we identified a highly heritable composite trait (71% heritable) that was perfectly mediated by  
318 a composite transcript that included expression from four tissues known to be involved in metabolic disease.  
319 Gene expression in adipose tissue was the strongest mediator of genetic effects on metabolic disease. Further  
320 analysis of the loadings onto transcripts in each tissue revealed that the mediating signatures were tissue-  
321 specific transcriptional programs, many of which were previously known to be involved in the pathogenesis  
322 of metabolic disease. We showed here that regulation of these programs is heritable and mediated a large  
323 proportion of disease risk.

324 The transcripts with the highest loadings are similar to the core genes of the omnigenic model. These were  
325 transcripts of moderate heritability that were highly functionally related to the traits. Transcripts with small  
326 loadings are more peripheral to the traits measured in this experiment. There was no clear demarcation  
327 between the core and peripheral genes as far as loading, but a clear separation should not be expected given  
328 the complexity of gene regulation and the genotype-phenotype map.

329 The strength of mediation (transcript loading) was negatively correlated with local heritability and positively  
330 correlated with distal heritability, suggesting that distal gene regulation was the dominant mode through which  
331 gene expression mediated the effect of genotype on phenotype. We saw further that the distal heritability  
332 was weak and spread across the genome, consistent with the prediction by Liu *et al.*<sup>63</sup> that trait heritability  
333 is mediated through weak distal eQTLs. Most strongly mediating transcripts had modest distal heritability,  
334 and even for those whose expression was strongly regulated by distal factors, these factors were multiple  
335 and spread across the genome. For example, Nucb2, was a strongly mediating transcript in islet and was  
336 also strongly distally regulated (66% distal heritability). This gene is expressed in pancreatic  $\beta$  cells and is  
337 involved in insulin and glucagon release<sup>64–66</sup>. Although its transcription was highly heritable in islets, that

338 regulation was distributed across the genome, with no clear distal eQTL (Supp. Fig. 13). Thus, although  
339 distal regulation of some genes may be strong, this regulation is likely to be highly complex and not easily  
340 localized.

341 The high complexity of gene regulation combined with a systems-level analysis yields continuous results  
342 that do not necessarily implicate individual transcripts or genetic loci in disease pathogenesis. Most studies  
343 have focused on pinpointing individual loci whose mechanistic roles can be clearly dissected through further  
344 experiments. In this analysis, too, it is possible to focus on individual genes and their context in both tissues  
345 and pathways. For example, we showed that the loadings on *Pparg* were tissue-specific in a way that comports  
346 with known biology, i.e. it is known to be protective in adipose tissue where it was negatively loaded, and  
347 harmful in the liver, where it was positively loaded. However, continuous results can also be quite informative  
348 in their own right. Combined with increasing amounts of high-dimensional data in public databases, weighted  
349 vectors can be useful for generating hypotheses and potential drug treatments. We showed that weighted  
350 vectors of genes can be analyzed for enriched biological functions and pathways using GSEA. These vectors  
351 can also be paired with data about cell-type specific genes to generate hypotheses about cell composition in  
352 individual tissues. Gene expression derived from patient biopsies confirmed that the transcriptional signatures  
353 we identified in mice predict obesity status in humans, further supporting the translatability of these results.  
354 Finally, we used the CMAP database to show that the transcriptomic signatures we identified in mice could  
355 be translated into human drug targets, as currently used diabetes drugs were among the top hits for reversing  
356 the disease-associated signatures. That these drugs are known to reverse diabetes pathogenesis supports the  
357 causal role of these gene signatures in disease risk as modeled by high-dimensional mediation.

358 In conclusion, we have shown that both tissue specificity and distal gene regulation are critically important  
359 to understanding the genetic architecture of complex traits. Although our systems approach does not  
360 identify individual genetic loci conferring risk of metabolic disease, we identified important genes and gene  
361 signatures that were heritable, causal of disease, and translatable to other mouse populations and to humans.  
362 Finally, we have shown that by directly acknowledging the complexity of both gene regulation and the  
363 genotype-to-phenotype map, we can gain a new perspective on disease pathogenesis and develop actionable  
364 hypotheses about pathogenic mechanisms and potential treatments.

## 365 Data Availability

366 Here we tell people where to find the data

<sup>367</sup> **Acknowledgements**

<sup>368</sup> Here we thank people

369 **Supplemental Figures**

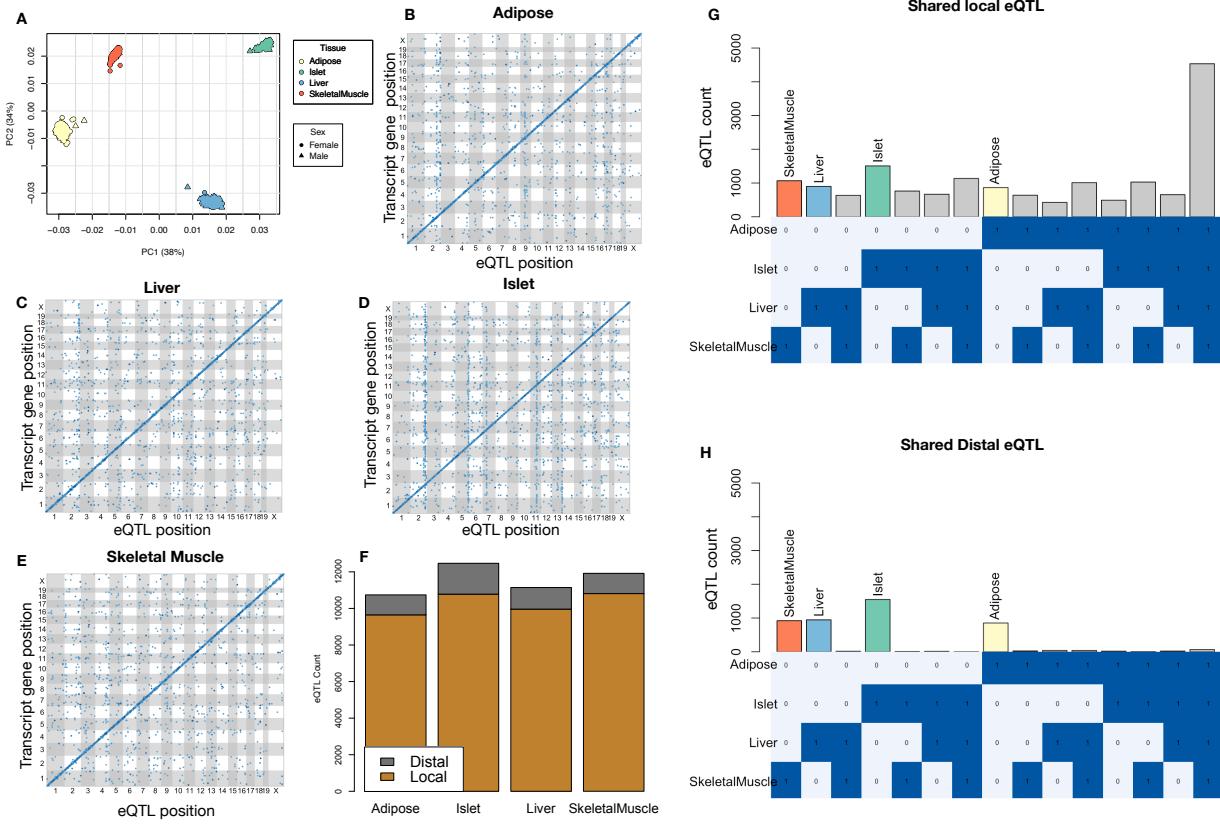


Figure 9: Overview of eQTL analysis in DO mice. **A.** RNA seq samples from the four different tissues clustered by tissue. **B.-E.** eQTL maps are shown for each tissue. The *x*-axis shows the position of the mapped eQTL, and the *y*-axis shows the physical position of the gene encoding each mapped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

370 **References**

- 371 [1] M. T. Maurano, R. Humbert, E. Rynes, R. E. Thurman, E. Haugen, H. Wang, A. P. Reynolds,  
372 R. Sandstrom, H. Qu, J. Brody, A. Shafer, F. Neri, K. Lee, T. Kutyavin, S. Stehling-Sun, A. K.  
373 Johnson, T. K. Canfield, E. Giste, M. Diegel, D. Bates, R. S. Hansen, S. Neph, P. J. Sabo, S. Heimfeld,  
374 A. Raubitschek, S. Ziegler, C. Cotsapas, N. Sotoodehnia, I. Glass, S. R. Sunyaev, R. Kaul, and J. A.

## KEGG pathway enrichments by GSEA

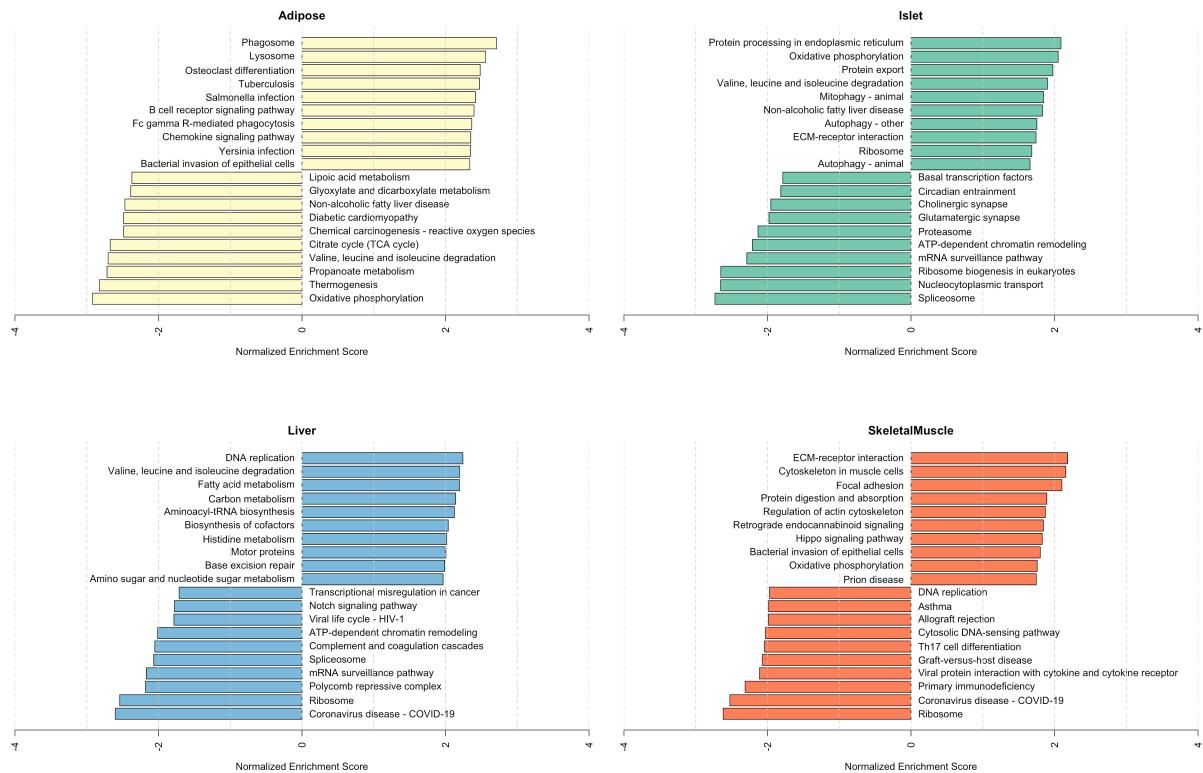


Figure 10: Bar plots showing normalized enrichment scores (NES) for KEGG pathways as determined by fast gene score enrichment analysis (fgsea). Only the top 10 positive and top 10 negative scores are shown. Colors indicate tissue. The name beside each bar shows the name of each enriched KEGG pathway.

- 375 Stamatoyanopoulos. Systematic localization of common disease-associated variation in regulatory DNA.  
 376 *Science*, 337(6099):1190–1195, Sep 2012.
- 377 [2] K. K. Farh, A. Marson, J. Zhu, M. Kleinewietfeld, W. J. Housley, S. Beik, N. Shores, H. Whitton, R. J.  
 378 Ryan, A. A. Shishkin, M. Hatan, M. J. Carrasco-Alfonso, D. Mayer, C. J. Luckey, N. A. Patsopoulos,  
 379 P. L. De Jager, V. K. Kuchroo, C. B. Epstein, M. J. Daly, D. A. Hafler, and B. E. Bernstein. Genetic  
 380 and epigenetic fine mapping of causal autoimmune disease variants. *Nature*, 518(7539):337–343, Feb  
 381 2015.
- 382 [3] E. Pennisi. The Biology of Genomes. Disease risk links to gene regulation. *Science*, 332(6033):1031, May  
 383 2011.
- 384 [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

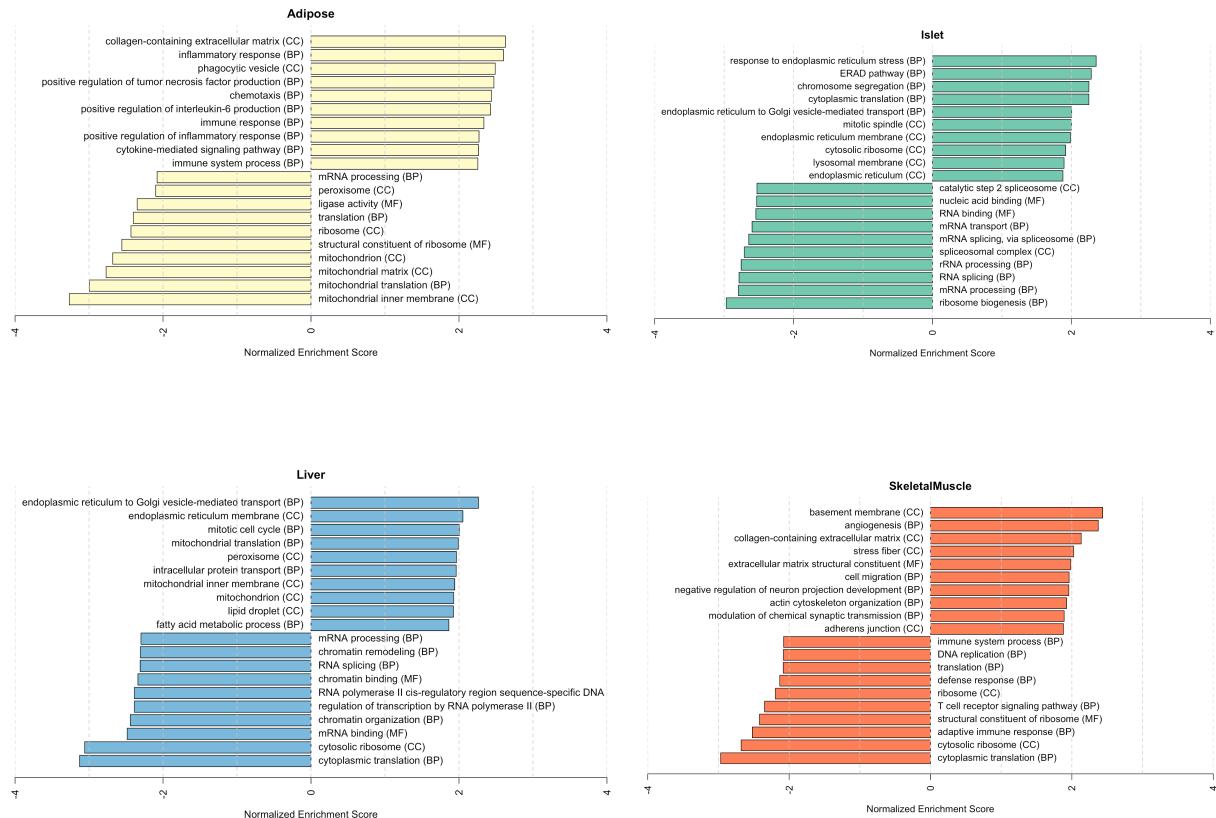


Figure 11: Bar plots showing normalized enrichment scores (NES) for GO terms as determined by fast gene score enrichment analysis (fgsea). Only the top 10 positive and top 10 negative scores are shown. Colors indicate tissue. The name beside each bar shows the name of each enriched GO term. The letters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

- 385 Potential etiologic and functional implications of genome-wide association loci for human diseases and  
 386 traits. *Proc Natl Acad Sci*, 106(23):9362–9367, Jun 2009.
- 387 [5] J. K. Pickrell. Joint analysis of functional genomic data and genome-wide association studies of 18  
 388 human traits. *Am J Hum Genet*, 94(4):559–573, Apr 2014.
- 389 [6] D. Welter, J. MacArthur, J. Morales, T. Burdett, P. Hall, H. Junkins, A. Klemm, P. Flicek, T. Manolio,  
 390 L. Hindorff, and H. Parkinson. The NHGRI GWAS Catalog, a curated resource of SNP-trait associations.  
 391 *Nucleic Acids Res*, 42(Database issue):D1001–1006, Jan 2014.
- 392 [7] Y. I. Li, B. van de Geijn, A. Raj, D. A. Knowles, A. A. Petti, D. Golan, Y. Gilad, and J. K. Pritchard.  
 393 RNA splicing is a primary link between genetic variation and disease. *Science*, 352(6285):600–604, Apr

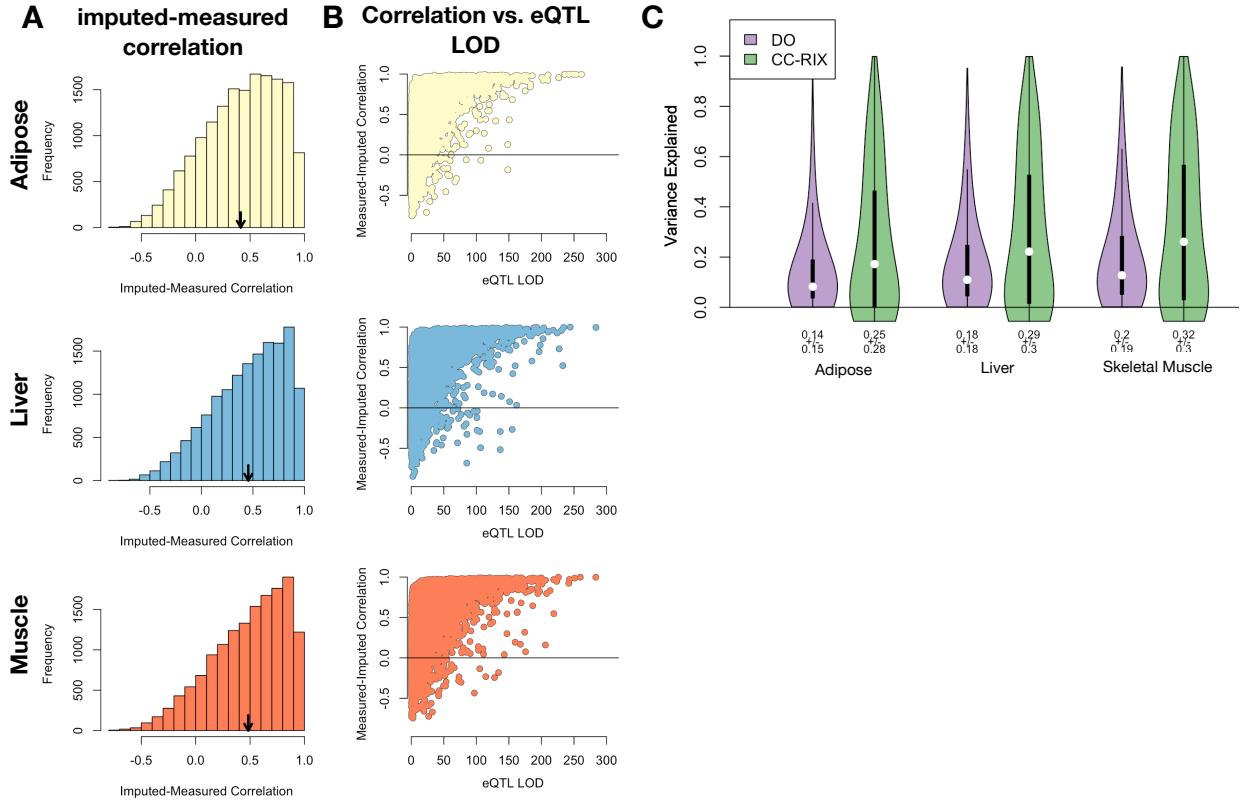


Figure 12: Validation of transcript imputation in the CC-RIX. **A.** Distributions of correlations between imputed and measured transcripts in the CC-RIX. The mean of each distribution is shown by the red line. All distributions were skewed toward positive correlations and had positive means near a Pearson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

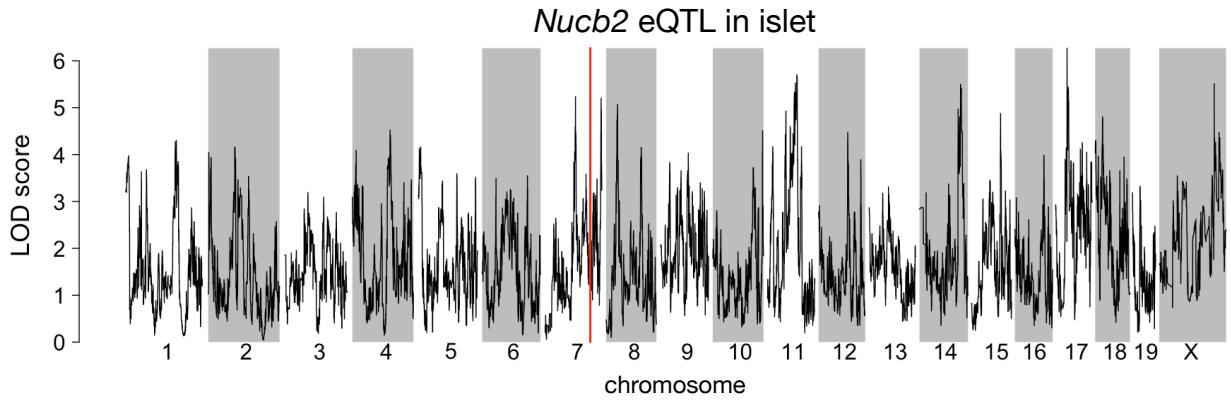


Figure 13: Regulation of *Nucb2* expression in islet. *Nucb2* is encoded on mouse chromosome 7 at 116.5 Mb (red line). In islets the heritability of *Nucb2* expression levels is 69% heritable. This LOD score trace shows that there is no local eQTL at that position, nor any strong distal eQTL anywhere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

- 394 2016.
- 395 [8] D. Zhou, Y. Jiang, X. Zhong, N. J. Cox, C. Liu, and E. R. Gamazon. A unified framework for joint-tissue  
396 transcriptome-wide association and Mendelian randomization analysis. *Nat Genet*, 52(11):1239–1246,  
397 Nov 2020.
- 398 [9] E. R. Gamazon, H. E. Wheeler, K. P. Shah, S. V. Mozaffari, K. Aquino-Michaels, R. J. Carroll, A. E.  
399 Eyler, J. C. Denny, D. L. Nicolae, N. J. Cox, and H. K. Im. A gene-based association method for  
400 mapping traits using reference transcriptome data. *Nat Genet*, 47(9):1091–1098, Sep 2015.
- 401 [10] Z. Zhu, F. Zhang, H. Hu, A. Bakshi, M. R. Robinson, J. E. Powell, G. W. Montgomery, M. E. Goddard,  
402 N. R. Wray, P. M. Visscher, and J. Yang. Integration of summary data from GWAS and eQTL studies  
403 predicts complex trait gene targets. *Nat Genet*, 48(5):481–487, May 2016.
- 404 [11] A. Gusev, A. Ko, H. Shi, G. Bhatia, W. Chung, B. W. Penninx, R. Jansen, E. J. de Geus, D. I. Boomsma,  
405 F. A. Wright, P. F. Sullivan, E. Nikkola, M. Alvarez, M. Civelek, A. J. Lusis, T. ki, E. Raitoharju,  
406 M. nen, I. ä, O. T. Raitakari, J. Kuusisto, M. Laakso, A. L. Price, P. Pajukanta, and B. Pasaniuc.  
407 Integrative approaches for large-scale transcriptome-wide association studies. *Nat Genet*, 48(3):245–252,  
408 Mar 2016.
- 409 [12] M. P. Keller, D. M. Gatti, K. L. Schueler, M. E. Rabaglia, D. S. Stapleton, P. Simecek, M. Vincent,  
410 S. Allen, A. T. Broman, R. Bacher, C. Kendziorski, K. W. Broman, B. S. Yandell, G. A. Churchill, and  
411 A. D. Attie. Genetic Drivers of Pancreatic Islet Function. *Genetics*, 209(1):335–356, May 2018.
- 412 [13] W. L. Crouse, G. R. Keele, M. S. Gastonguay, G. A. Churchill, and W. Valdar. A Bayesian model  
413 selection approach to mediation analysis. *PLoS Genet*, 18(5):e1010184, May 2022.
- 414 [14] J. M. Chick, S. C. Munger, P. Simecek, E. L. Huttlin, K. Choi, D. M. Gatti, N. Raghupathy, K. L. Svenson,  
415 G. A. Churchill, and S. P. Gygi. Defining the consequences of genetic variation on a proteome-wide scale.  
416 *Nature*, 534(7608):500–505, Jun 2016.
- 417 [15] H. E. Wheeler, S. Ploch, A. N. Barbeira, R. Bonazzola, A. Andaleon, A. Fotuhi Siahpirani, A. Saha,  
418 A. Battle, S. Roy, and H. K. Im. Imputed gene associations identify replicable trans-acting genes enriched  
419 in transcription pathways and complex traits. *Genet Epidemiol*, 43(6):596–608, Sep 2019.
- 420 [16] B. D. Umans, A. Battle, and Y. Gilad. Where Are the Disease-Associated eQTLs? *Trends Genet*,  
421 37(2):109–124, Feb 2021.
- 422 [17] N. J. Connally, S. Nazeen, D. Lee, H. Shi, J. Stamatoyannopoulos, S. Chun, C. Cotsapas, C. A. Cassa,

- 423 and S. R. Sunyaev. The missing link between genetic association and regulatory function. *Elife*, 11, Dec  
424 2022.
- 425 [18] H. Mostafavi, J. P. Spence, S. Naqvi, and J. K. Pritchard. Systematic differences in discovery of genetic  
426 effects on gene expression and complex traits. *Nat Genet*, 55(11):1866–1875, Nov 2023.
- 427 [19] D. W. Yao, L. J. O’Connor, A. L. Price, and A. Gusev. Quantifying genetic effects on disease mediated  
428 by assayed gene expression levels. *Nat Genet*, 52(6):626–633, Jun 2020.
- 429 [20] X. Liu, J. A. Mefford, A. Dahl, Y. He, M. Subramaniam, A. Battle, A. L. Price, and N. Zaitlen. GBAT:  
430 a gene-based association test for robust detection of trans-gene regulation. *Genome Biol*, 21(1):211, Aug  
431 2020.
- 432 [21] G. A. Churchill, D. M. Gatti, S. C. Munger, and K. L. Svenson. The Diversity Outbred mouse population.  
433 *Mamm Genome*, 23(9-10):713–718, Oct 2012.
- 434 [22] S. M. Clee and A. D. Attie. The genetic landscape of type 2 diabetes in mice. *Endocr Rev*, 28(1):48–83,  
435 Feb 2007.
- 436 [23] K. W. Broman, D. M. Gatti, P. Simecek, N. A. Furlotte, P. Prins, Š. Sen, B. S. Yandell, and G. A.  
437 Churchill. R/qtL2: Software for Mapping Quantitative Trait Loci with High-Dimensional Data and  
438 Multiparent Populations. *Genetics*, 211(2):495–502, Feb 2019.
- 439 [24] Fabien Girka, Etienne Camenen, Caroline Peltier, Arnaud Gloaguen, Vincent Guillemot, Laurent Le  
440 Brusquet, and Arthur Tenenhaus. *RGCCA: Regularized and Sparse Generalized Canonical Correlation  
441 Analysis for Multiblock Data*, 2023. R package version 3.0.3.
- 442 [25] Gennady Korotkevich, Vladimir Sukhov, and Alexey Sergushichev. Fast gene set enrichment analysis.  
443 *bioRxiv*, 2019.
- 444 [26] A. Subramanian, P. Tamayo, V. K. Mootha, S. Mukherjee, B. L. Ebert, M. A. Gillette, A. Paulovich,  
445 S. L. Pomeroy, T. R. Golub, E. S. Lander, and J. P. Mesirov. Gene set enrichment analysis: a  
446 knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci U S A*,  
447 102(43):15545–15550, Oct 2005.
- 448 [27] C. B. Newgard. Interplay between lipids and branched-chain amino acids in development of insulin  
449 resistance. *Cell Metab*, 15(5):606–614, May 2012.
- 450 [28] D. D. Sears, G. Hsiao, A. Hsiao, J. G. Yu, C. H. Courtney, J. M. Ofrecio, J. Chapman, and S. Subramaniam.

- 451 Mechanisms of human insulin resistance and thiazolidinedione-mediated insulin sensitization. *Proc Natl*  
452 *Acad Sci U S A*, 106(44):18745–18750, Nov 2009.
- 453 [29] R. Stienstra, C. Duval, M. ller, and S. Kersten. PPARs, Obesity, and Inflammation. *PPAR Res*,  
454 2007:95974, 2007.
- 455 [30] O. Gavrilova, M. Haluzik, K. Matsusue, J. J. Cutson, L. Johnson, K. R. Dietz, C. J. Nicol, C. Vinson,  
456 F. J. Gonzalez, and M. L. Reitman. Liver peroxisome proliferator-activated receptor gamma contributes  
457 to hepatic steatosis, triglyceride clearance, and regulation of body fat mass. *J Biol Chem*, 278(36):34268–  
458 34276, Sep 2003.
- 459 [31] K. Matsusue, M. Haluzik, G. Lambert, S. H. Yim, O. Gavrilova, J. M. Ward, B. Brewer, M. L. Reitman,  
460 and F. J. Gonzalez. Liver-specific disruption of PPARgamma in leptin-deficient mice improves fatty  
461 liver but aggravates diabetic phenotypes. *J Clin Invest*, 111(5):737–747, Mar 2003.
- 462 [32] D. Patsouris, J. K. Reddy, M. ller, and S. Kersten. Peroxisome proliferator-activated receptor alpha  
463 mediates the effects of high-fat diet on hepatic gene expression. *Endocrinology*, 147(3):1508–1516, Mar  
464 2006.
- 465 [33] S. E. Schadinger, N. L. Bucher, B. M. Schreiber, and S. R. Farmer. PPARgamma2 regulates lipogenesis  
466 and lipid accumulation in steatotic hepatocytes. *Am J Physiol Endocrinol Metab*, 288(6):E1195–1205,  
467 Jun 2005.
- 468 [34] W. Motomura, M. Inoue, T. Ohtake, N. Takahashi, M. Nagamine, S. Tanno, Y. Kohgo, and T. Okumura.  
469 Up-regulation of ADRP in fatty liver in human and liver steatosis in mice fed with high fat diet. *Biochem*  
470 *Biophys Res Commun*, 340(4):1111–1118, Feb 2006.
- 471 [35] S. Amin, A. Lux, and F. O’Callaghan. The journey of metformin from glycaemic control to mTOR  
472 inhibition and the suppression of tumour growth. *Br J Clin Pharmacol*, 85(1):37–46, Jan 2019.
- 473 [36] A. Kezic, L. Popovic, and K. Lalic. mTOR Inhibitor Therapy and Metabolic Consequences: Where Do  
474 We Stand? *Oxid Med Cell Longev*, 2018:2640342, 2018.
- 475 [37] A. D. Barlow, M. L. Nicholson, and T. P. Herbert. -cells and a review of the underlying molecular  
476 mechanisms. *Diabetes*, 62(8):2674–2682, Aug 2013.
- 477 [38] Y. Gu, J. Lindner, A. Kumar, W. Yuan, and M. A. Magnuson. Rictor/mTORC2 is essential for  
478 maintaining a balance between beta-cell proliferation and cell size. *Diabetes*, 60(3):827–837, Mar 2011.

- 479 [39] E. B. Geer, J. Islam, and C. Buettner. Mechanisms of glucocorticoid-induced insulin resistance: focus on  
480 adipose tissue function and lipid metabolism. *Endocrinol Metab Clin North Am*, 43(1):75–102, Mar 2014.
- 481 [40] J. X. Li and C. L. Cummins. Fresh insights into glucocorticoid-induced diabetes mellitus and new  
482 therapeutic directions. *Nat Rev Endocrinol*, 18(9):540–557, Sep 2022.
- 483 [41] R. A. Lee, C. A. Harris, and J. C. Wang. Glucocorticoid Receptor and Adipocyte Biology. *Nucl Receptor  
484 Res*, 5, 2018.
- 485 [42] S. Viengchareun, P. Penfornis, M. C. Zennaro, and M. s. Mineralocorticoid and glucocorticoid receptors  
486 inhibit UCP expression and function in brown adipocytes. *Am J Physiol Endocrinol Metab*, 280(4):E640–  
487 649, Apr 2001.
- 488 [43] J. Liu, X. Kong, L. Wang, H. Qi, W. Di, X. Zhang, L. Wu, X. Chen, J. Yu, J. Zha, S. Lv, A. Zhang,  
489 P. Cheng, M. Hu, Y. Li, J. Bi, Y. Li, F. Hu, Y. Zhong, Y. Xu, and G. Ding. -HSD1 in regulating brown  
490 adipocyte function. *J Mol Endocrinol*, 50(1):103–113, Feb 2013.
- 491 [44] L. E. Ramage, M. Akyol, A. M. Fletcher, J. Forsythe, M. Nixon, R. N. Carter, E. J. van Beek,  
492 N. M. Morton, B. R. Walker, and R. H. Stimson. Glucocorticoids Acutely Increase Brown Adipose  
493 Tissue Activity in Humans, Revealing Species-Specific Differences in UCP-1 Regulation. *Cell Metab*,  
494 24(1):130–141, Jul 2016.
- 495 [45] J. L. Barclay, H. Agada, C. Jang, M. Ward, N. Wetzig, and K. K. Ho. Effects of glucocorticoids on  
496 human brown adipocytes. *J Endocrinol*, 224(2):139–147, Feb 2015.
- 497 [46] M. ó, R. Gupte, W. L. Kraus, P. Pacher, and P. Bai. PARPs in lipid metabolism and related diseases.  
498 *Prog Lipid Res*, 84:101117, Nov 2021.
- 499 [47] P. Bai, C. ó, H. Oudart, A. nszki, Y. Cen, C. Thomas, H. Yamamoto, A. Huber, B. Kiss, R. H.  
500 Houtkooper, K. Schoonjans, V. Schreiber, A. A. Sauve, J. Menissier-de Murcia, and J. Auwerx. PARP-1  
501 inhibition increases mitochondrial metabolism through SIRT1 activation. *Cell Metab*, 13(4):461–468,  
502 Apr 2011.
- 503 [48] A. Chiarugi and M. A. Moskowitz. Cell biology. PARP-1—a perpetrator of apoptotic cell death? *Science*,  
504 297(5579):200–201, Jul 2002.
- 505 [49] M. Althubiti, R. Almaimani, S. Y. Eid, M. Elzubaier, B. Refaat, S. Idris, T. A. Alqurashi, and M. Z.  
506 El-Readi. BTK targeting suppresses inflammatory genes and ameliorates insulin resistance. *Eur Cytokine  
507 Netw*, 31(4):168–179, Dec 2020.

- 508 [50] C. Skrabs, W. F. Pickl, T. Perkmann, U. ger, and A. Gessl. Rapid decline in insulin antibodies and  
509 glutamic acid decarboxylase autoantibodies with ibrutinib therapy of chronic lymphocytic leukaemia. *J*  
510 *Clin Pharm Ther*, 43(1):145–149, Feb 2018.
- 511 [51] M. C. Arkan, A. L. Hevener, F. R. Greten, S. Maeda, Z. W. Li, J. M. Long, A. Wynshaw-Boris, G. Poli,  
512 J. Olefsky, and M. Karin. IKK-beta links inflammation to obesity-induced insulin resistance. *Nat Med*,  
513 11(2):191–198, Feb 2005.
- 514 [52] E. A. Oral, S. M. Reilly, A. V. Gomez, R. Meral, L. Butz, N. Ajluni, T. L. Chenevert, E. Korytnaya,  
515 A. H. Neidert, R. Hench, D. Rus, J. F. Horowitz, B. Poirier, P. Zhao, K. Lehmann, M. Jain, R. Yu,  
516 C. Liddle, M. Ahmadian, M. Downes, R. M. Evans, and A. R. Saltiel. and TBK1 Improves Glucose  
517 Control in a Subset of Patients with Type 2 Diabetes. *Cell Metab*, 26(1):157–170, Jul 2017.
- 518 [53] M. Clark, C. J. Kroger, Q. Ke, and R. M. Tisch. The Role of T Cell Receptor Signaling in the  
519 Development of Type 1 Diabetes. *Front Immunol*, 11:615371, 2020.
- 520 [54] P. ndell, A. C. Carlsson, A. Larsson, O. Melander, T. Wessman, J. v, and T. Ruge. TNFR1 is associated  
521 with short-term mortality in patients with diabetes and acute dyspnea seeking care at the emergency  
522 department. *Acta Diabetol*, 57(10):1145–1150, Oct 2020.
- 523 [55] A. M. Keestra-Gounder, M. X. Byndloss, N. Seyffert, B. M. Young, A. vez Arroyo, A. Y. Tsai, S. A.  
524 Cevallos, M. G. Winter, O. H. Pham, C. R. Tiffany, M. F. de Jong, T. Kerrinnes, R. Ravindran, P. A.  
525 Luciw, S. J. McSorley, A. J. umler, and R. M. Tsolis. NOD1 and NOD2 signalling links ER stress with  
526 inflammation. *Nature*, 532(7599):394–397, Apr 2016.
- 527 [56] A. M. Keestra-Gounder and R. M. Tsolis. NOD1 and NOD2: Beyond Peptidoglycan Sensing. *Trends*  
528 *Immunol*, 38(10):758–767, Oct 2017.
- 529 [57] J. Montane, L. Cadavez, and A. Novials. Stress and the inflammatory process: a major cause of  
530 pancreatic cell death in type 2 diabetes. *Diabetes Metab Syndr Obes*, 7:25–34, 2014.
- 531 [58] B. B. Kahn and T. E. McGraw. , and type 2 diabetes. *N Engl J Med*, 363(27):2667–2669, Dec 2010.
- 532 [59] S. Del Prato and N. Pulizzi. The place of sulfonylureas in the therapy for type 2 diabetes mellitus.  
533 *Metabolism*, 55(5 Suppl 1):S20–27, May 2006.
- 534 [60] B. Hallgrímsson, R. M. Green, D. C. Katz, J. L. Fish, F. P. Bernier, C. C. Roseman, N. M. Young,  
535 J. M. Cheverud, and R. S. Marcucio. The developmental-genetics of canalization. *Semin Cell Dev Biol*,  
536 88:67–79, Apr 2019.

- 537 [61] M. L. Siegal and A. Bergman. Waddington's canalization revisited: developmental stability and evolution.  
538        *Proc Natl Acad Sci U S A*, 99(16):10528–10532, Aug 2002.
- 539 [62] A. B. Paaby and G. Gibson. Cryptic Genetic Variation in Evolutionary Developmental Genetics. *Biology*  
540        (*Basel*), 5(2), Jun 2016.
- 541 [63] X. Liu, Y. I. Li, and J. K. Pritchard. Trans Effects on Gene Expression Can Drive Omnipotent Inheritance.  
542        *Cell*, 177(4):1022–1034, May 2019.
- 543 [64] M. Riva, M. D. Nitert, U. Voss, R. Sathanoori, A. Lindqvist, C. Ling, and N. Wierup. Nesfatin-1  
544        stimulates glucagon and insulin secretion and beta cell NUCB2 is reduced in human type 2 diabetic  
545        subjects. *Cell Tissue Res*, 346(3):393–405, Dec 2011.
- 546 [65] M. Nakata and T. Yada. Role of NUCB2/nesfatin-1 in glucose control: diverse functions in islets,  
547        adipocytes and brain. *Curr Pharm Des*, 19(39):6960–6965, 2013.
- 548 [66] H. Shimizu and A. Osaki. Nesfatin/Nucleobindin-2 (NUCB2) and Glucose Homeostasis. *Curr Hypertens*  
549        *Rev*, pages Nesfatin/Nucleobindin-2 (NUCB2) and Glucose Homeostasis., Jul 2014.