

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

3

4 Anna L Tyler, J Matthew Mahoney, Mark P Keller, Candice N. Baker, Anuj Srivastava, Isabela Gerdes  
5 Gyuricza, [Other Churchill people? Jeff Harder?], Nadia A Rosenthal, Alan D Attie, Gary A Churchill and  
6 Gregory W Carter

7 **Abstract**

8 Variation in gene expression is an important mediator of the effect of genotype on phenotype. Evidence  
9 suggests that local gene regulation may play only a minor role in trait-relevant variation in gene expression.  
10 Here we combined two large, purpose-built data sets to investigate the relative contributions of local and distal  
11 gene regulation in mediating the connection between genotype and phenotype. Using a novel high-dimensional  
12 mediation analysis, we identified a heritable composite transcript that explained 30% of trait variation in a  
13 genetically diverse mouse model of diet-induced obesity and metabolic disease. The composite transcript was  
14 interpretable in terms of enriched biological processes, and predicted obesity status in an independent mouse  
15 cohort as well as human cohorts with measured gene expression. Transcripts contributing most strongly to  
16 this mediator tended to have complex, distal regulation distributed throughout the genome. These results  
17 suggest that the majority of heritable transcript variation mediating trait variation is regulated in a complex  
18 manner, but is nonetheless translatable across species.

19 **Introduction**

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

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

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

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

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

<sup>39</sup> To assess the role of wide-spread distal gene regulation on complex traits, we investigated diet-induced obesity  
<sup>40</sup> and metabolic disease as an archetypal example. Diet-induced obesity and metabolic disease are genetically  
<sup>41</sup> complex with hundreds of variants mapped through GWAS [REFS]. These variants are known to act through  
<sup>42</sup> multiple tissues that interact dynamically with each other [REFS], including adipose tissue, pancreatic  
<sup>43</sup> islets, liver, and skeletal muscle. The multi-system etiology of metabolic disease complicates mechanistic  
<sup>44</sup> dissection of the genetic architecture, requiring large, dedicated data sets that include high-dimensional,  
<sup>45</sup> clinically relevant phenotyping, dense genotyping in a highly recombined population, and transcriptome-wide  
<sup>46</sup> measurements of gene expression in multiple tissues. Measuring gene expression in multiple tissues is critical  
<sup>47</sup> to adequately assess the extent to which local gene regulation varies across multiple tissues and whether such  
<sup>48</sup> variability might account for previous failed attempts to identify trait-relevant local eQTL. Such data sets  
<sup>49</sup> are extremely difficult to obtain in human populations, particularly in the large numbers of subjects required  
<sup>50</sup> for adequate statistical power. Thus, to investigate further the role of local and distal gene regulation on  
<sup>51</sup> complex traits, we have generated an appropriate data set in a large population of diversity outbred (DO)  
<sup>52</sup> mice<sup>21</sup> in a population model of diet-induced obesity and metabolic disease<sup>12</sup>.

<sup>53</sup> The DO mice were derived from eight inbred founder mouse strains, five classical lab strains, and three  
<sup>54</sup> strains more recently derived from wild mice<sup>21</sup>. They represent three subspecies of mouse *Mus musculus*  
<sup>55</sup> *domesticus*, *Mus musculus musculus*, and *Mus musculus castaneus*, and capture 90% of the known variation  
<sup>56</sup> in laboratory mice<sup>22</sup>. They are maintained with a breeding scheme that ensures equal contributions from

57 each founder across the genome thus rendering almost the whole genome visible to genetic inquiry<sup>21</sup>. We  
58 paired clinically relevant metabolic traits from 500 DO mice [REF], including body weight, plasma levels  
59 of insulin and glucose and plasma lipids, with transcriptome-wide gene expression in four tissues related to  
60 metabolic disease: adipose tissue, pancreatic islets, liver, and skeletal muscle. Taken together, these data  
61 enable a comprehensive view into the genetic architecture of metabolic disease.

## 62 Results

### 63 Genetic variation contributed to wide phenotypic variation

64 Although the environment was consistent across all animals, the genetic diversity present in this population  
65 resulted in widely varying distributions across physiological measurements (Fig. 1). For example, body  
66 weights of adult individuals varied from less than the average adult B6 body weight to several times the body  
67 weight of a B6 adult in both sexes (Males: 18.5 - 69.1g, Females: 16.0 - 54.8g) (Fig. 1A). Fasting blood  
68 glucose (FBG) also varied considerably (Fig. 1B) although few of the animals had FBG levels that would  
69 indicate pre-diabetes (19 animals, 3.8%), or diabetes (7 animals, 1.4%) according to previously developed  
70 cutoffs (pre-diabetes:  $\text{FBG} \geq 250 \text{ mg/dL}$ , diabetes:  $\text{FBG} \geq 300 \text{ mg/dL}$ )<sup>23</sup>. Males had higher FBG than  
71 females on average (Fig. 1C) as has been observed before suggesting either that males were more susceptible  
72 to metabolic disease on the high-fat diet, or that males and females may require different thresholds for  
73 pre-diabetes and diabetes.

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

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

### 82 Distal Heritability Correlated with Phenotype Relevance

83 We performed eQTL analysis using R/qtl2<sup>24</sup> (Methods) and identified both local and distal eQTLs for  
84 transcripts in each of the four tissues (Supp. Fig 1). Significant local eQTLs far outnumbered distal eQTLs  
85 (Supp. Fig. 1F) and tended to be shared across tissues (Supp. Fig. 1G) whereas the few significant distal

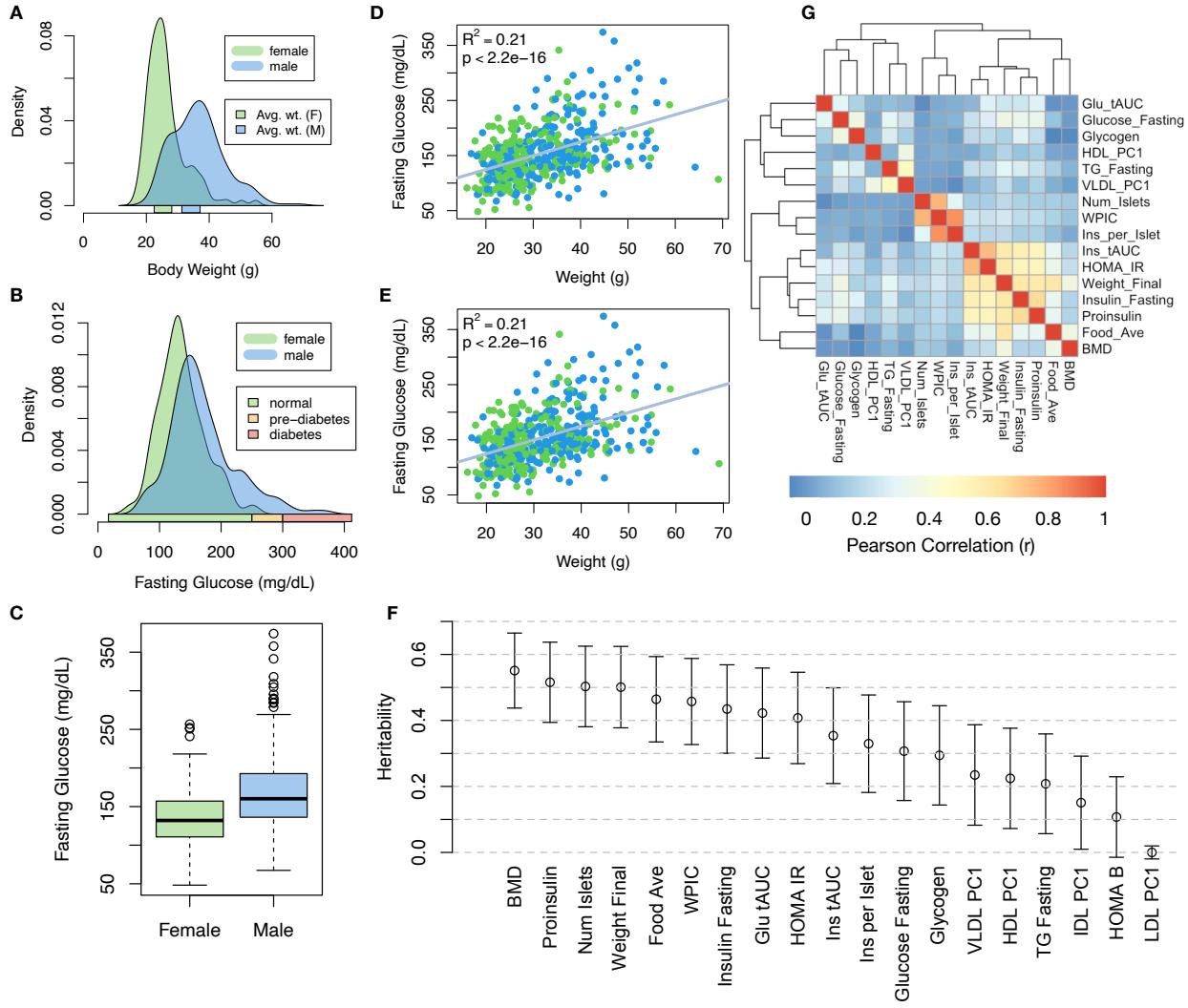


Figure 1: Clinical overview. **A.** Distributions of final body weight in the diversity outbred mice. Sex is indicated by color. The average B6 male and female adult weights at 24 weeks of age are indicated by blue and green bars on the x-axis. **B.** The distribution of final fasting glucose across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

86 eQTLs we identified tended to be tissue-specific (Supp. Fig. 1H)

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

90 The local heritability of transcripts was negatively correlated with their trait relevance, defined as the

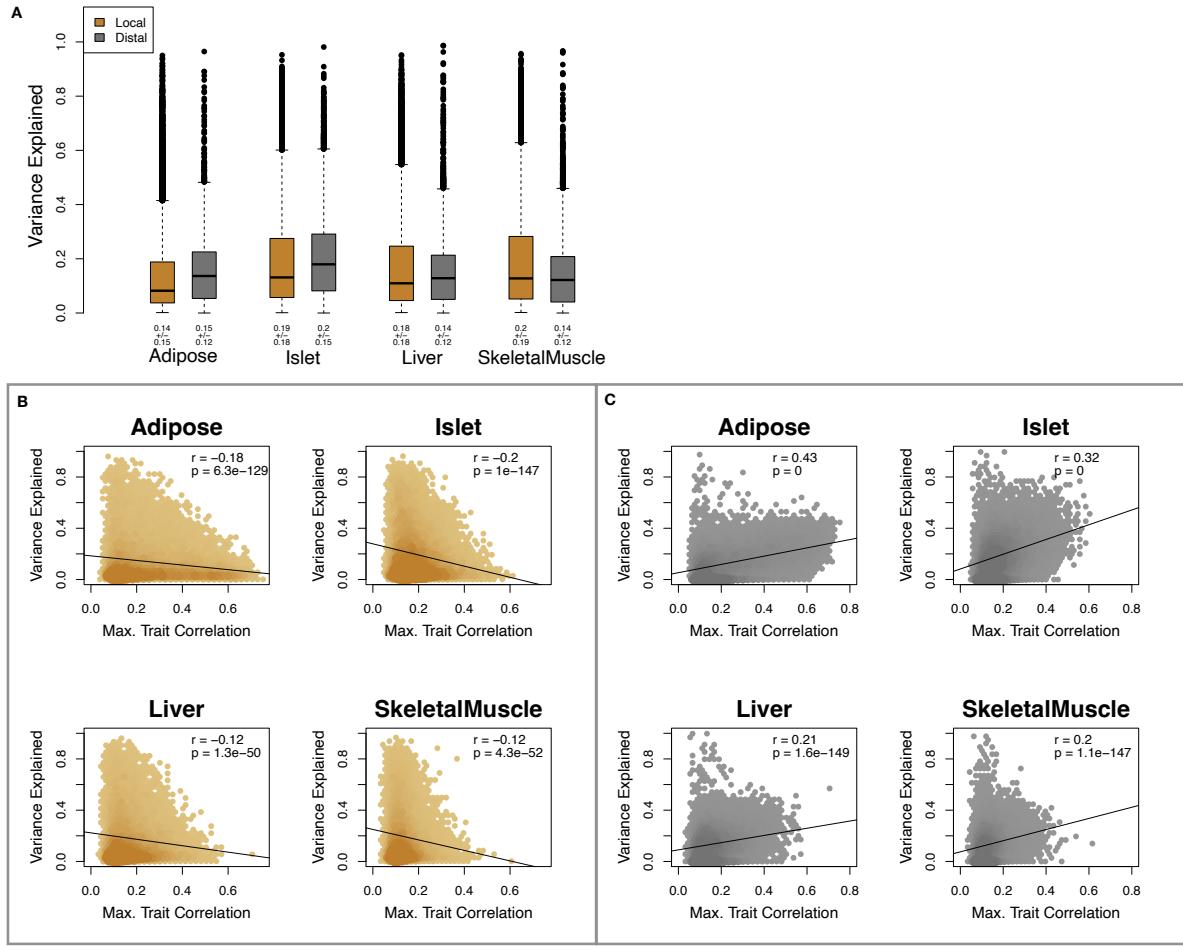


Figure 2: Transcript heritability and trait relevance. **A.** Distributions of distal and local heritability of transcripts across the four tissues. Overall local and distal factors contribute equally to transcript heritability. The relationship between **(B.)** local and **(C.)** distal heritability and trait relevance across all four tissues. Here trait relevance is defined as the maximum correlation between the transcript and all traits. Local heritability was negatively correlated with trait relevance, and distal heritability is positively correlated with trait relevance. Pearson ( $r$ ) and  $p$  values for each correlation are shown in the upper-right of each panel.

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

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

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

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

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

118 We used high-dimensioal mediation to identify the major axis of variation in the transcriptome that mediated  
119 the effects of the genome on metabolic traits (Fig 3). Fig. 3A shows the partial correlations ( $\rho$ ) between  
120 the pairs of these composite vectors. The partial correlation between  $G_C$  and  $T_C$  was 0.42, and the partial  
121 correlation between  $T_C$  and  $P_C$  was 0.78. However, when the transcriptome was taken into account, the partial  
122 correlation between  $G_C$  and  $P_C$  was effectively zero (0.039).  $P_C$  captured 30% of the overall trait variance,  
123 and its estimated heritability was  $0.71 \pm 0.084$ , which was higher than any of the measured traits (Fig. 1F).  
124 Thus, HDMA identified a maximally heritable metabolic composite trait that was perfectly mediated by a  
125 highly heritable component of the transcriptome.

126 Standard CCA is prone to over-fitting because in any two large matrices it can be trivial to identify highly  
127 correlated composite vectors [REF]. To assess whether our implementation of HDMA was similarly prone to

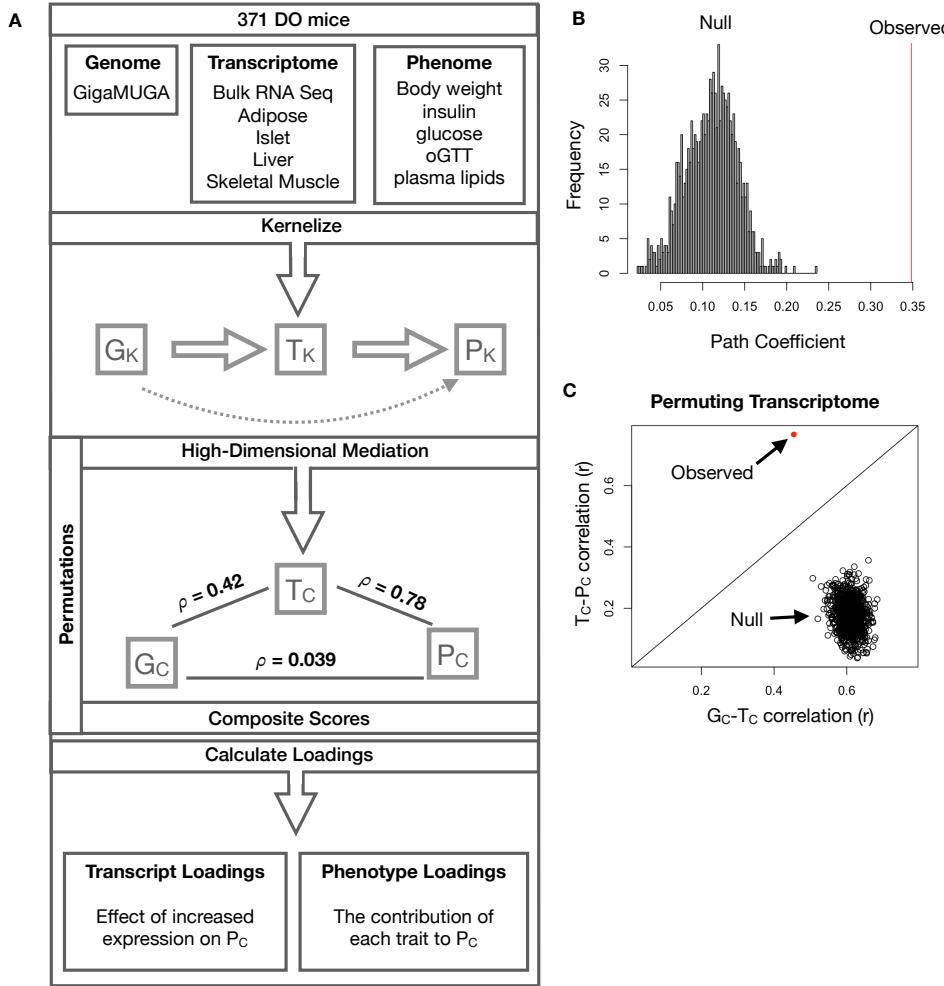


Figure 3: High-dimensional mediation. **A.** Workflow indicating major steps of high-dimensional mediation. The genotype, transcriptome, and phenotype matrices were kernelized to yield single matrices representing the relationships between all individuals for each data modality ( $G_K$  = genome kernel,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

128 over-fitting in a high-dimensional space, we performed permutation testing. We permuted the individual  
 129 labels on the transcriptome matrix 1000 times and recalculated the path coefficient, which is the partial  
 130 correlation of  $G_c$  and  $T_c$  multiplied by the partial correlation of  $T_c$  and  $P_c$ . This represents the path  
 131 from  $G_c$  to  $P_c$  that is mediated through  $T_c$ . The null distribution of the path coefficient is shown in Fig.  
 132 3B, and the observed path coefficient from the original data is indicated by the red line. The observed  
 133 path coefficient was well outside the null distribution generated by permutations ( $p < 10^{-16}$ ). Fig. 3C

134 illustrates this observation in more detail. Although we identified high correlations between  $G_C$  and  $T_C$ , and  
135 modest correlations between  $T_C$  and  $P_C$  in the null data (Fig 3C), these two values could not be maximized  
136 simultaneously in the null data. In contrast, the red dot shows that in the real data both the  $G_C-T_C$   
137 correlation and the  $T_C-P_C$  correlation could be maximized simultaneously suggesting that the path from  
138 genotype to phenotype through transcriptome is highly non-trivial and identifiable in this case. These results  
139 suggest that these composite vectors represent genetically determined variation in phenotype that is mediated  
140 through genetically determined variation in transcription.

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

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

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

156 We interpreted large loadings onto transcripts as indicating strong mediation of the effect of genetics on  
157 metabolic index. Large positive loadings indicate that higher expression was associated with a higher  
158 metabolic index (i.e. higher risk of obesity and metabolic disease on the high-fat diet) (Fig. 4C). Conversely,  
159 large negative loadings indicate that high expression of these transcripts was associated with a lower metabolic  
160 index (i.e. lower risk of obesity and metabolic disease on the high-fat diet) (Fig. 4C). We used gene set  
161 enrichment analysis (GSEA)<sup>26;27</sup> to look for biological processes and pathways that were enriched at the top  
162 and bottom of this list (Methods).

163 In adipose tissue, both GO processes and KEGG pathway enrichments pointed to an axis of inflammation

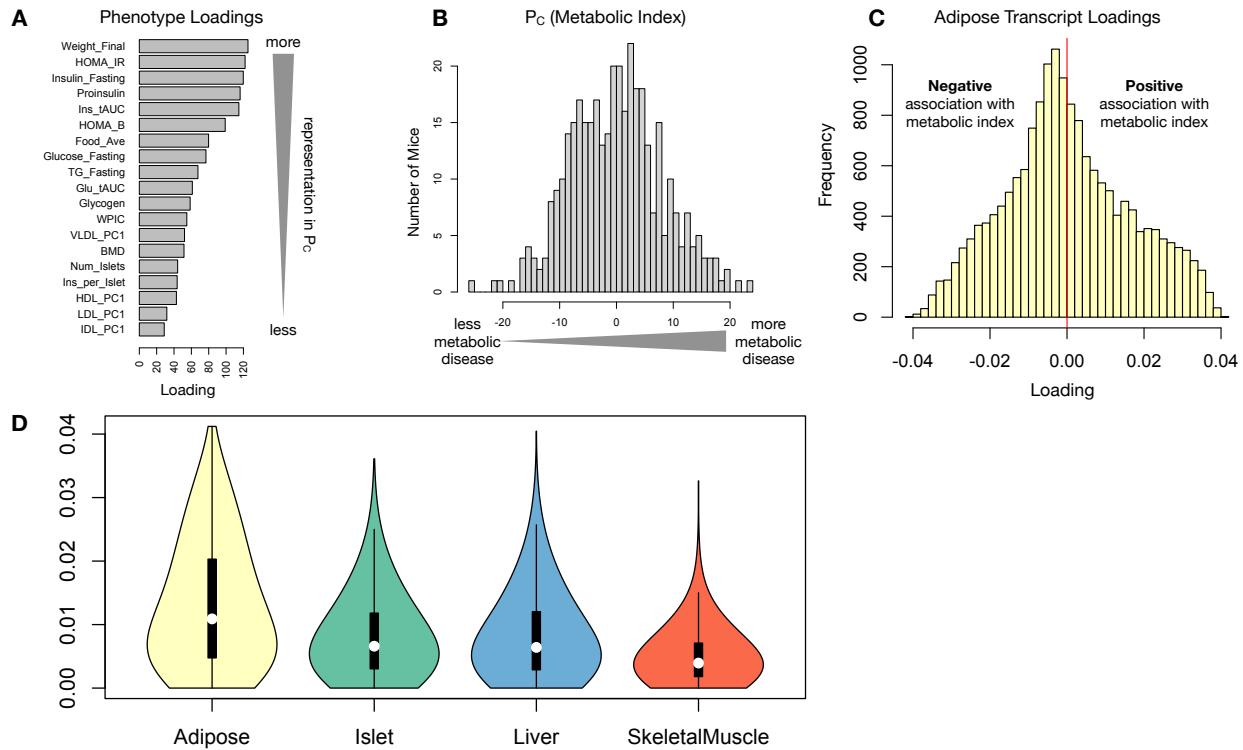


Figure 4: Interpretation of loadings. **A.** Loadings across traits. Body weight and insulin resistance contributed the most to the composite trait. **B.** Phenotype scores across individuals. Individuals with large positive phenotype scores had higher body weight and insulin resistance than average.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

- 164 and metabolism (Supp. Fig. 2 and Fig. 11). GO terms and KEGG pathways associated with inflammation,  
 165 particularly macrophage infiltration, were positively associated with metabolic index, indicating that increased  
 166 expression in inflammatory pathways was associated with a higher metabolic index. It is well established that  
 167 adipose tissue in obese individuals is inflamed [cite] and infiltrated by macrophages [cite], and the results  
 168 here suggest that this may be a heritable component of metabolic disease.
- 169 The strongest negative enrichments in adipose tissue were related to mitochondrial activity in general, and  
 170 thermogenesis in particular (Supp. Fig. 2 and Fig. 11). It has been shown mouse strains with greater  
 171 thermogenic potential are also less susceptible to obesity on a high-fat diet [cite].
- 172 Transcripts associated with the citric acid (TCA) cycle as well as the catabolism of the branched-chain amino  
 173 acids (BCAA) (valine, leucine, and isoleucine) were strongly enriched with negative loadings in adipose

<sup>174</sup> tissue (Supp. Fig. 3). Expression of genes in both pathways (for which there is some overlap) has been  
<sup>175</sup> previously associated with insulin sensitivity<sup>12;28;29</sup>, suggesting that heritable variation in regulation of these  
<sup>176</sup> pathways may influence risk of insulin resistance.

<sup>177</sup> Looking at the 10 most positively and negatively loaded transcripts from each tissue, it is apparent that  
<sup>178</sup> transcripts in the adipose tissue had the largest loadings, both positive and negative, of all tissues (Fig. 5A  
<sup>179</sup> bar plot) This suggests that much of the effect of genetics on body weight and insulin resistance is mediated  
<sup>180</sup> through gene expression in adipose tissue. The strongest loadings in liver and pancreas were comparable,  
<sup>181</sup> and those in skeletal muscle were the weakest (Fig. 5A), suggesting that less of the genetic effects were  
<sup>182</sup> mediated through transcription in skeletal muscle. Heritability analysis showed that transcripts with the  
<sup>183</sup> largest loadings had higher distal heritability than local heritability (Fig. 5A heat map and box plot). This  
<sup>184</sup> pattern contrasts with transcripts nominated by TWAS (Fig. 5B), which tended to have lower loadings,  
<sup>185</sup> higher local heritability and lower distal heritability. Transcripts with the highest local heritability in each  
<sup>186</sup> tissue (Fig. 5C) had the lowest loadings.

<sup>187</sup> We performed a literature search for the genes in each of these groups along with the terms “diabetes”,  
<sup>188</sup> “obesity”, and the name of the expressing tissue to determine whether any of these genes had previous  
<sup>189</sup> associations with metabolic disease in the literature (Methods). Multiple genes in each group had been  
<sup>190</sup> previously associated with obesity and diabetes (Fig. 5 bolded gene names). Genes with high loadings were  
<sup>191</sup> most highly enriched for previous literature support. They were 2.375 more likely than TWAS hits and 3.8  
<sup>192</sup> times more likely than genes with high local heritability to be previously associated with obesity or diabetes.

### <sup>193</sup> **Tissue-specific transcriptional programs were associated with metabolic traits**

<sup>194</sup> Clustering of transcripts with top loadings in each tissue showed tissue-specific functional modules associated  
<sup>195</sup> with obesity and insulin resistance (Fig. 6A) (Methods). The clustering highlights the importance of immune  
<sup>196</sup> activation particularly in adipose tissue. Except for the “mitosis” cluster, which had large positive loadings  
<sup>197</sup> in three of the four tissues, all clusters were strongly loaded in only one or two tissues. For example, the lipid  
<sup>198</sup> metabolism cluster was loaded most heavily in liver. The positive loadings suggest that high expression of  
<sup>199</sup> these genes particularly in the liver was associated with increased metabolic disease. This cluster included  
<sup>200</sup> the gene *Pparg*, whose primary role is in the adipose tissue where it is considered a master regulator of  
<sup>201</sup> adipogenesis<sup>30</sup>. Agonists of *Pparg*, such as thiazolidinediones, are FDA-approved to treat type II diabetes,  
<sup>202</sup> and reduce inflammation and adipose hypertrophy<sup>30</sup>. Consistent with this role, the loading for *Pparg* in  
<sup>203</sup> adipose tissue was negative, suggesting that higher expression was associated with leaner mice (Fig. 6B). In  
<sup>204</sup> contrast, *Pparg* had a large positive loading in liver, where it is known to play a role in the development of

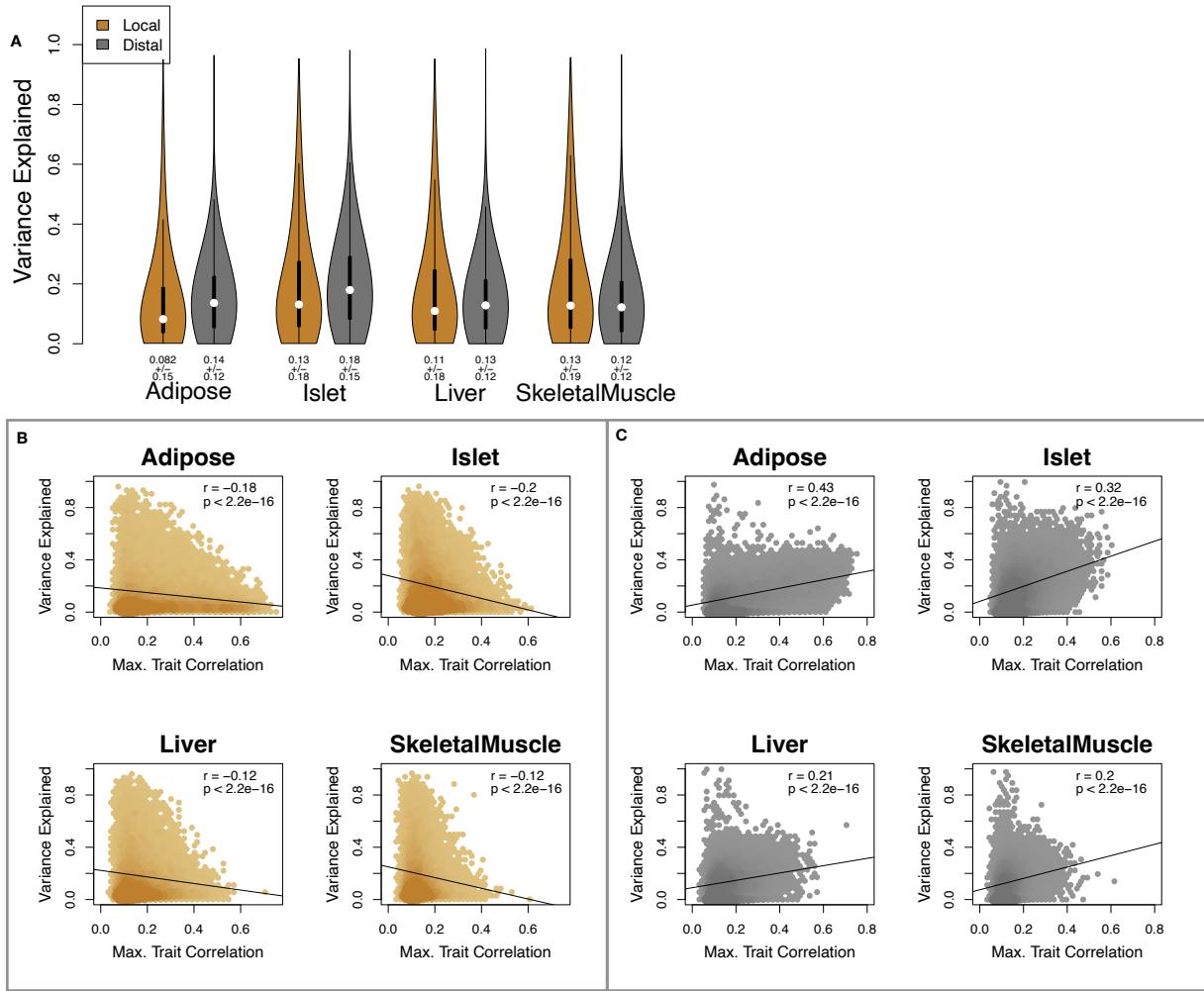


Figure 5: Transcripts with high loadings have high distal heritability and literature support. Each panel has a bar plot showing the loadings of transcripts selected by different criteria. Bar color indicates the tissue of origin. The heat map shows the local (L - left) and distal (D - right) heritability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

- hepatic steatosis, or fatty liver. Mice that lack *Pparg* specifically in the liver, are protected from developing steatosis and show reduced expression of lipogenic genes<sup>31;32</sup>. Overexpression of *Pparg* in the livers of mice with a *Ppara* knockout, causes upregulation of genes involved in adipogenesis<sup>33</sup>. In the livers of both mice and humans high *Pparg* expression is associated with hepatocytes that accumulate large lipid droplets and have gene expression profiles similar to adipocytes<sup>34;35</sup>.
- The local and distal heritability of *Pparg* is low in adipose tissue suggesting its expression in this tissue is highly constrained in the population (Fig. 6B). However, the distal heritability of *Pparg* in liver is relatively

212 high suggesting it is complexly regulated and has sufficient variation in this population to drive variation in  
 213 phenotype. Both local and distal heritability of *Pparg* in the islet are relatively high, but the loading is low,  
 214 suggesting that variability of expression in the islet does not drive variation in metabolic index. These results  
 215 highlight the importance of tissue context when investigating the role of heritable transcript variability in  
 216 driving phenotype.

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

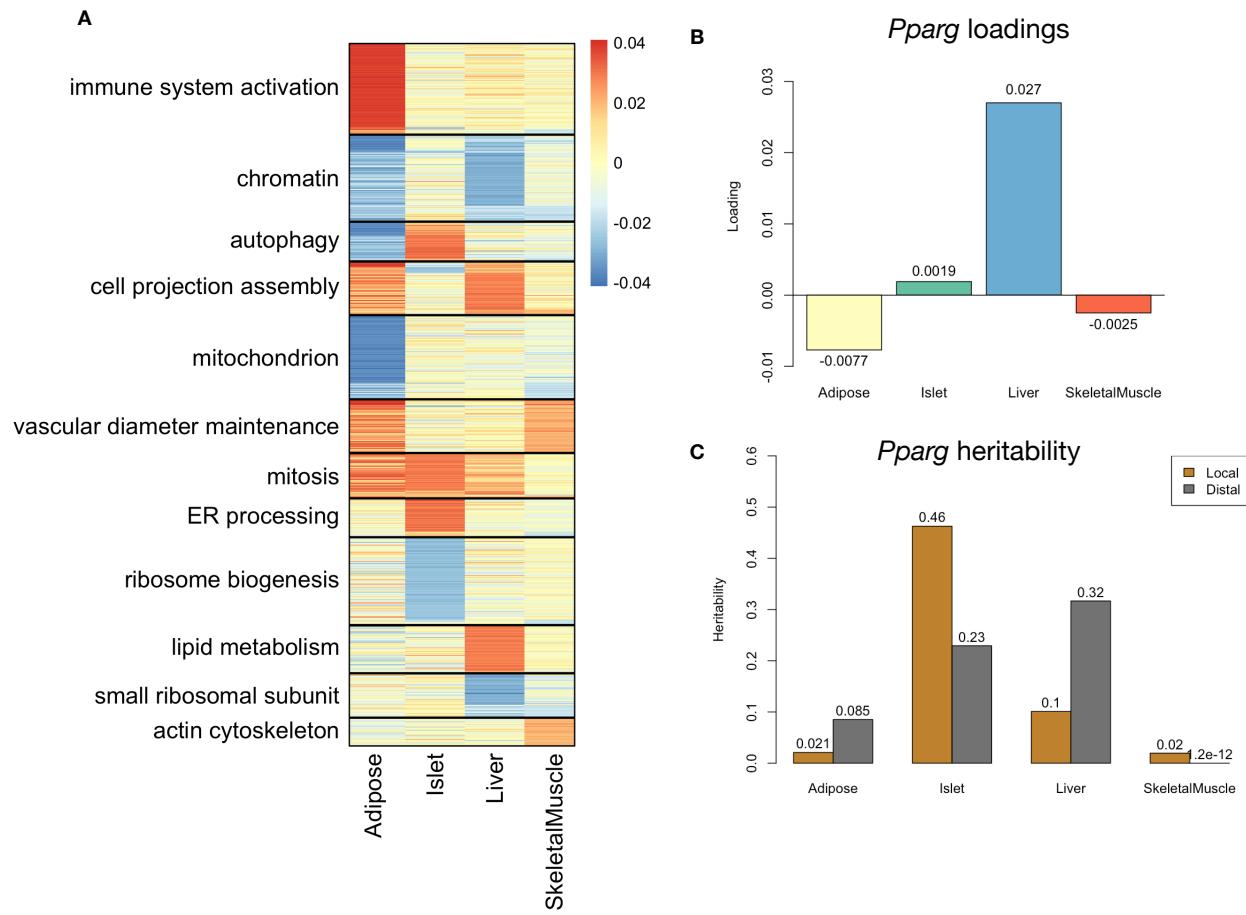


Figure 6: Tissue-specific transcriptional programs were associated with obesity and insulin resistance. **A** Heat map showing the loadings of all transcripts with loadings greater than 2.5 standard deviations from the mean in any tissue. The heat map was clustered using k medoid clustering. Functional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

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

219 To test whether the transcript loadings identified in the DO could be translated to another population, we  
 220 tested whether they could predict metabolic phenotype in an independent population of CC-RIX mice, which

were F1 mice derived from multiple pairings of Collaborative Cross (CC) [cite] strains (Fig. 7) (Methods). We tested two questions. First, we asked whether the loadings identified in the DO mice were relevant to the relationship between the transcriptome and the phenotype in the CC-RIX. We predicted body weight (a surrogate for metabolic index) in each CC-RIX individual using measured gene expression in each tissue and the transcript loadings identified in the DO (Methods). The predicted body weight and actual body weight were highly correlated in all tissues (Fig. 7B left column). The best prediction was achieved for adipose tissue, which supports the observation in the DO that adipose expression was the strongest mediator of the genetic effect on metabolic index. This result also confirms the validity and translatability of the transcript loadings and their relationship to metabolic disease.

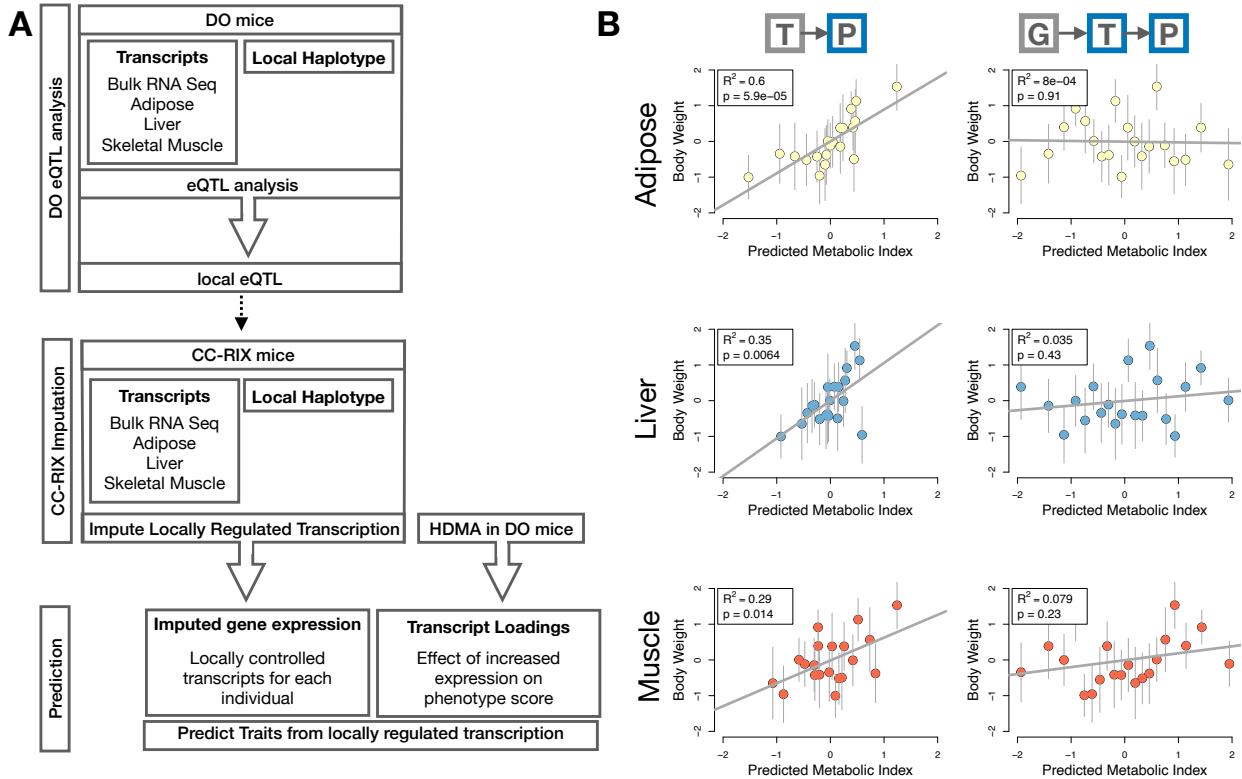


Figure 7: Transcription, but not local genotype, predicts phenotype in the CC-RIX. **A.** Workflow showing procedure for translating HDMA results to an independent population of mice. **B.** Relationships between the predicted metabolic index and measured body weight. The left column shows the predictions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

The second question related to the source of the relevant variation in gene expression. If local regulation was the predominant factor influencing gene expression, we should be able to predict phenotype in the CC-RIX using transcripts imputed from local genotype (Fig. 7A). The DO and the CC-RIX were derived from the

233 same eight founder strains and so carry the same alleles throughout the genome. We imputed gene expression  
234 in the CC-RIX using local genotype and were able to estimate variation in gene transcription robustly (Supp.  
235 Fig. 4). However, these imputed values failed to predict body weight in the CC-RIX when weighted with the  
236 loadings from HDMA. (Fig. 7B right column). This result suggests that local regulation of gene expression is  
237 not the primary factor driving heritability of complex traits, consistent with our findings in the DO population  
238 that distal heritability was a major driver of trait-relevant variation and that high-loading transcripts had  
239 comparatively high distal and low local heritability.

240 **Distally heritable transcriptomic signatures reflected variation in composition of adipose tissue  
241 and islets**

242 The interpretation of global genetic influences on gene expression and phenotype is potentially more challenging  
243 than the interpretation and translation of local genetic influences, as genetic effects cannot be localized to  
244 individual gene variants or transcripts. However, there are global patterns across the loadings that can  
245 inform mechanism. For example, heritable variation in cell type composition can be derived from transcript  
246 loadings. We observed above that immune activation in the adipose tissues was an important driver of obesity  
247 in the DO population. To determine whether this is reflected as an increase in macrophages in adipose  
248 tissue, we compared loadings of cell-type specific genes in adipose tissue (Methods). The mean loading  
249 of macrophage-specific genes was significantly greater than 0 (Fig. 8A), indicating that obese mice were  
250 genetically predisposed to have high levels of macrophage infiltration in adipose tissue in response to the  
251 high-fat, high-sugar diet.

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

259 Notably, the loadings for pancreatic beta cell-type specific loadings was not significantly different from zero.  
260 This is not necessarily reflective of the function of the beta cells in the obese mice, but rather suggests that  
261 any variation in the number of beta cells in these mice was unrelated to obesity and insulin resistance. This  
262 is further consistent with the islet composition traits having small loadings in the phenome score (Fig. 4).

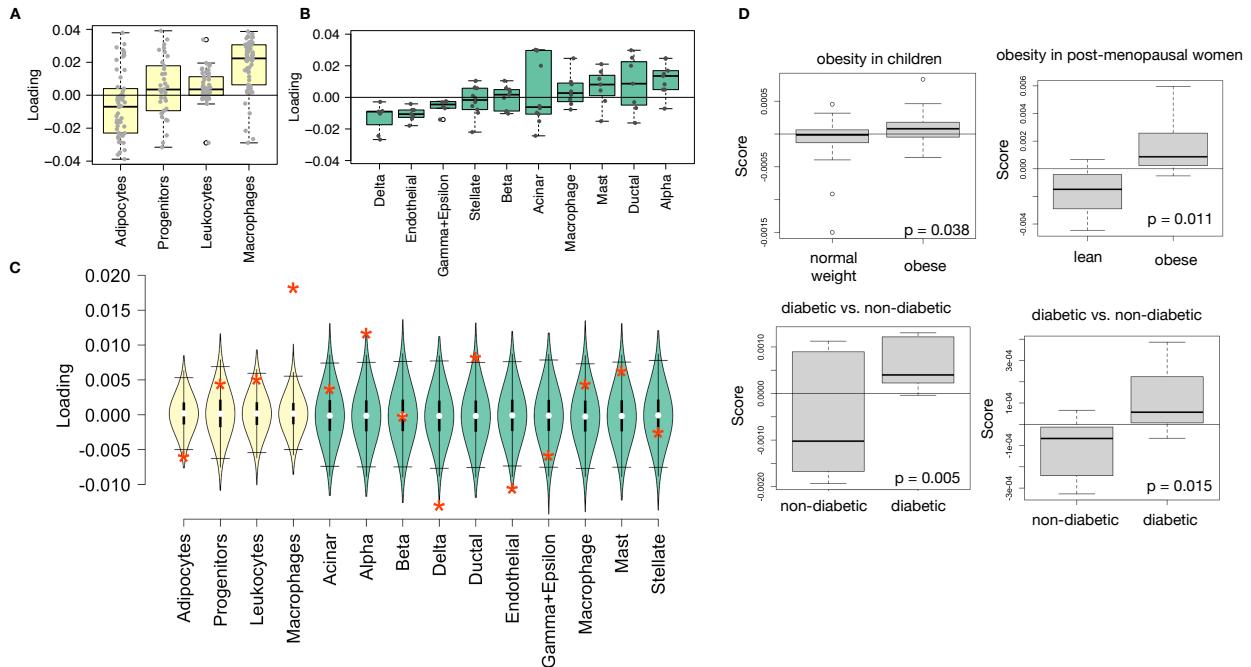


Figure 8: HDMA results translate to humans. **A.** Distribution of loadings for cell-type-specific transcripts in adipose tissue. **B.** Distribution of loadings for cell-type-specific transcripts in pancreatic islets (green). **C.** Null distributions for the mean loading of randomly selected transcripts in each cell type compared with the observed mean loading of each group of transcripts (red asterisk). **D.** Predictions of metabolic phenotypes in four adipose transcription data sets downloaded from GEO. In each study the obese/diabetic patients were predicted to have greater metabolic disease than the lean/non-diabetic patients based on the HDMA results from DO mice.

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

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

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

275 **Targeting gene signatures**

276 Another potential application of the transcript loading landscape is in ranking potential drug candidates  
277 for the treatment of metabolic disease. Although high-loading transcripts may be good candidates for  
278 understanding specific biology related to obesity, the transcriptome overall is highly interconnected and  
279 redundant, and focusing on individual transcripts for treatment may be less effective than using broader  
280 transcriptomic signatures that capture the emergent biology [cite or remove]. The ConnectivityMap (CMAP)  
281 database<sup>36</sup> developed by the Broad Institute allows us to query thousands of compounds that reverse or  
282 enhance the extreme ends of transcriptomic signatures in multiple different cell types. By identifying drugs  
283 that reverse pathogenic transcriptomic signatures, we can potentially identify compounds that have favorable  
284 effects on gene expression.

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

290 Looking broadly across cell types, the notable top hits from the adipose tissue loadings included mTOR  
291 inhibitors and glucocorticoid agonists (Supplemental Figure XXX). It is thought that metformin, which  
292 is commonly used to improve glycemic control, acts, at least in part, by inhibiting mTOR signaling<sup>37;38</sup>.  
293 However, long-term use of other mTOR inhibitors, such as rapamycin, are known to cause insulin resistance  
294 and  $\beta$ -cell toxicity<sup>38–40</sup>. Glucocorticoids are used to reduce inflammation, which was a prominent signature  
295 in the adipose tissues, but these drugs also promote hyperglycemia and diabetes<sup>41;42</sup>. Accute treatment  
296 with glucocorticoids has further been shown to reduce thermogenesis in rodent adipocytes<sup>43–45</sup>, but increase  
297 thermogenesis in human adipocytes<sup>46;47</sup>. Thus, the pathways identified by CMAP across all cell types were  
298 highly related to the transcript loading profiles, but the relationship was not a simple reversal.

299 The top hit for the adipose composite transcript in CMAP adipocytes was a PARP inhibitor (Supplemental  
300 Figure XXXB). PARPs play a role in lipid metabolism and are involved in the development of obesity and  
301 diabetes<sup>48</sup>. PARP1 inhibition increases mitochondrial biogenesis<sup>49</sup>. Inhibition of PARP1 activity can further  
302 prevent necrosis in favor of the less inflammatory apoptosis<sup>50</sup>, thereby potentially reducing inflammation in  
303 stressed adipocytes. Other notable hits among the top 20 were BTK inhibitors, which have been observed  
304 to suppress inflammation and improve insulin resistance<sup>51</sup> as well as to reduce insulin antibodies in type I  
305 diabetes<sup>52</sup>. IKK inhibitors have been shown to improve glucose control in type II diabetes<sup>53;54</sup>.

306 Among the top most significant hits for the transcript loadings from pancreatic islets (Fig. XXX), was  
307 suppression of T cell receptor signaling, which is known to be involved in Type 1 diabetes<sup>55</sup>, as well as  
308 TNFR1, which has been associated with mortality in diabetes patients<sup>56</sup>. Suppression of NOD1/2 signaling  
309 was also among the top hits. NOD1 and 2 sense ER stress<sup>57;58</sup>, which is associated with  $\beta$ -cell death in type  
310 1 and type 2 diabetes<sup>59</sup>. This cell death process is dependent on NOD1/2 signaling<sup>57</sup>, although the specifics  
311 have not yet been worked out.

312 We also looked specifically at hits in pancreatic tumor cells (YAPC) regardless of significance level. Hits  
313 in this list included widely used diabetes drugs, such as sulfonylureas, PPAR receptor agonists, and insulin  
314 sensitizers. Rosiglitazone is a PPAR- $\gamma$  agonist and was one of the most prescribed drugs for type 2 diabetes  
315 before its use was reduced due to cardiac side-effects<sup>60</sup>. Sulfonylureas are another commonly prescribed drug  
316 class for type 2 diabetes, but also have notable side effects including hypoglycemia and accelerated  $\beta$ -cell  
317 death<sup>61</sup>.

## 318 Discussion

319 Here we used a novel high-dimensional mediation analysis (HDMA) to investigate the relative contributions of  
320 local and distal gene regulation to heritable trait variation in a genetically diverse mouse model of diet-induced  
321 obesity and metabolic disease. We identified tissue-specific composite transcripts mediating the effect of  
322 genetic background on metabolic traits. Transcripts contributing most strongly to these composite transcripts  
323 were distally heritable. These composite transcripts, but not local eQTL, were able to predict obesity in  
324 an independent mouse population with divergent allelic structure. Moreover, the composite transcript from  
325 adipose tissue predicted obesity and diabetes status in human cohorts with measured adipose gene expression.  
326 Taken together, these results support the hypothesis that gene expression mediating the effect of genetic  
327 background on phenotype are primarily distally regulated, and that the gene regulatory networks influencing  
328 metabolic disease are conserved across mice and humans.

329 It has frequently been assumed that gene regulation in *cis* is the primary driver of genetically associated  
330 trait variation, but attempts to use local gene regulation to explain phenotypic variation have had limited  
331 success<sup>16;17</sup>. In recent years, evidence has mounted that distal gene regulation may be an important mediator  
332 of trait heritability<sup>19;18;62</sup>. It has been observed that transcripts with high local heritability explain less  
333 expression-mediated disease heritability than transcripts with low local heritability<sup>19</sup>. Consistent with this  
334 observation, genes located near GWAS hits tend to be complexly regulated<sup>18</sup>. They also tend to be enriched  
335 with functional annotations, in contrast to genes with simple local regulation, which tend to be depleted of  
336 functional annotations suggesting they are less likely to be directly involved in disease traits<sup>18</sup>. These ideas

337 are consistent with principles of robustness in complex systems<sup>63–65</sup>. If a transcript were both important to a  
338 trait and subject to strong local regulation, a population would be susceptible to extremes in phenotype that  
339 might frequently cross the threshold to disease. Indeed, strong disruption of highly trait-relevant genes is the  
340 cause of Mendelian disease.

341 Our findings were consistent with previous observations. The composite transcripts we identified were highly  
342 heritable and explained a high proportion of disease risk. Transcript loadings (the degree to which they  
343 contributed to the composite transcript) were negatively correlated with local heritability and positively  
344 correlated with distal heritability. These strongly loaded transcripts were also enriched for functional  
345 annotations associated with metabolic disease. The composite transcripts were moreover able to predict  
346 obesity in an independent cohort of mice whereas models using local eQTL only could not. Together these  
347 observations suggest that distal gene regulation was the dominant mode through which gene expression  
348 mediated the effect of genetic background on complex traits.

349 Identification of this distally heritable signature depended on the high-dimensional approach we used. Because  
350 HDMA uses a kinship matrix rather than genotypes at individual loci, it allows for arbitrarily complex gene  
351 regulation, as well as the interconnectedness and redundancy of the transcriptome. This feature also means  
352 that HDMA assumes that traits are highly polygenic. It also assumes that genetic effects are distributed  
353 across the genome. In contrast, one-dimensional, univariate approaches assume a large, localized genetic  
354 effect. Thus, the HDMA approach is consistent with the omnigenic model of complex traits which posits  
355 that complex traits are massively polygenic and that their heritability is spread out across the genome<sup>66</sup>.  
356 In the omnigenic model, genes are classified either as “core genes,” which directly impinge on the trait, or  
357 “peripheral genes,” which are not directly trait-related, but influence core genes through the complex gene  
358 regulatory network. HDMA explicitly models a central proposal of the omnigenic model which posits that  
359 once the expression of the core genes (i.e. trait-mediating genes) is accounted for, there should be no residual  
360 correlation between the genome and the phenotype. Here, when the composite transcript was taken into  
361 account there was no residual correlation between the composite genome and composite phenotype (Fig. 3A).

362 The composite transcript is essentially a weighted vector with larger weights (loadings) indicating higher  
363 “core-ness” of a transcript. There was no clear demarcation between the core and peripheral genes in loading  
364 magnitude, but we do not necessarily expect a clear separation given the complexity of gene regulation and  
365 the genotype-phenotype map<sup>67</sup>. Still, the transcripts with the largest loadings had high distal heritability,  
366 low local heritability, and were enriched for biological processes related to metabolic traits, as we would  
367 predict for core genes.

368 An extension of the omnigenic model proposed that most heritability of complex traits is driven by weak  
369 distal eQTLs<sup>62</sup>. This is consistent with what we observed here. The transcripts with the largest loadings  
370 were strongly distally regulated and only weakly locally regulated, suggesting that distal gene regulation  
371 plays a primary role in driving heritable trait variation. We saw further that the patterns of distal heritability  
372 were complex spread across the genome. Even for transcripts whose expression was strongly regulated by  
373 distal factors, these factors were multiple and spread across the genome. For example, *Nucb2*, was a strongly  
374 mediating transcript in islet and was also strongly distally regulated (66% distal heritability) (Fig. 5). This  
375 gene is expressed in pancreatic  $\beta$  cells and is involved in insulin and glucagon release<sup>68–70</sup>. Although its  
376 transcription was highly heritable in islets, that regulation was distributed across the genome, with no clear  
377 distal eQTL (Supp. Fig. 5). Thus, although distal regulation of some genes may be strong, this regulation is  
378 likely to be highly complex and not easily localized.

379 The composite transcripts identified by HDMA are richly interpretable in both tissue- and gene-specific  
380 manners. The transcripts with the strongest loadings were enriched in biological functions previously known  
381 to be involved in the pathogenesis of metabolic disease, such as inflammation in adipose tissue. That these  
382 processes were identified in this analysis suggests that they have a heritable component, and that some  
383 individuals are genetically susceptible to greater adipose inflammation on a high-fat, high-sugar diet.

384 Individual transcripts also demonstrated biologically interpretable, tissue-specific patterns. We highlighted  
385 *Pparg*, which is known to be protective in adipose tissue<sup>30</sup> where it was negatively loaded, and harmful in the  
386 liver<sup>31–35</sup>, where it was positively loaded. Such granular patterns may be useful in generating hypotheses for  
387 further testing, and prioritizing genes as therapeutic targets. The tissue-specific nature of the loadings also  
388 may provide clues to tissue-specific effects, or side effects, of targeting particular genes system-wide, since  
389 antagonists of *Pparg* may reduce fatty liver disease, but exacerbate adipose tissue inflammation.

390 We showed further that these composite transcripts can be used as weighted vectors in multiple types of  
391 analysis, such as drug prioritization using gene set enrichment analysis (GSEA) and the CMAP database. In  
392 particular, the CMAP analysis identified drugs which have been demonstrated to reverse insulin resistance  
393 and other aspects of metabolic disease. This finding supports the causal role of these gene signatures in  
394 pathogenesis of metabolic disease and thus their utility in prioritizing drugs and gene targets as therapeutics.

395 Another useful application of the composite transcripts is to pair them with cell-type specific genes to generate  
396 hypotheses about cell composition in individual tissues. Combining the multi-tissue, transcriptome-wide  
397 weighted vectors with public databases and data sets thus provides a path for generating a wide range of  
398 testable hypotheses. Moreover, each data set presented here was derived from human tissues or cell lines,

399 thus demonstrating the translatability of these results. That the mouse-derived adipose composite transcript  
400 was able to classify human adipose tissue in terms of obesity and diabetes status further supports the direct  
401 translatability of these findings, utility of HDMA, and the continued importance of mouse models of human  
402 disease in which it is possible to obtain complete transcriptomes in multiple tissues across large numbers of  
403 individuals.

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

410 **Data Availability**

411 Here we tell people where to find the data

412 **Acknowledgements**

413 Here we thank people

414 **Supplemental Figures**

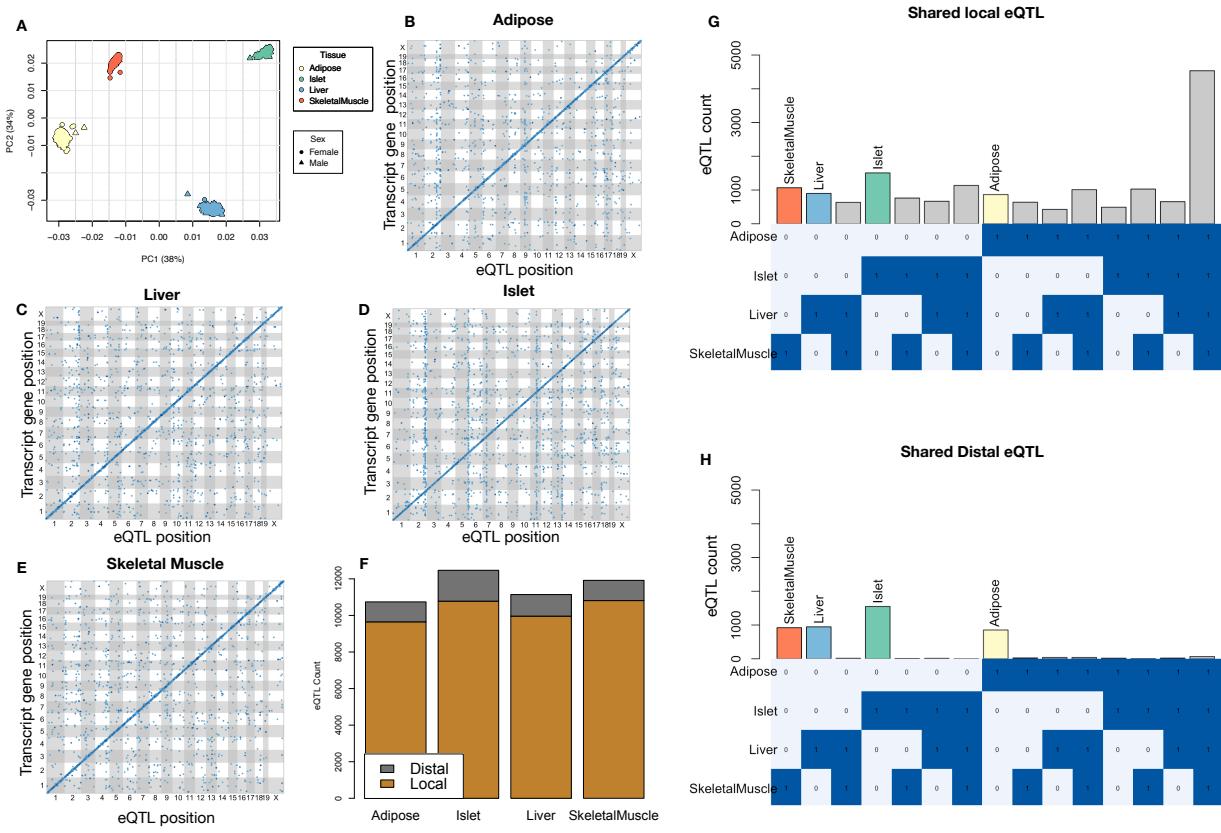


Figure 9: Overview of eQTL analysis in DO mice. **A.** RNA seq samples from the four different tissues clustered by tissue. **B.-E.** eQTL maps are shown for each tissue. The *x*-axis shows the position of the mapped eQTL, and the *y*-axis shows the physical position of the gene encoding each mapped transcript. Each dot represents an eQTL with a minimum LOD score of 8. The dots on the diagonal are locally regulated eQTL for which the mapped eQTL is at the within 4Mb of the encoding gene. Dots off the diagonal are distally regulated eQTL for which the mapped eQTL is distant from the gene encoding the transcript. **F.** Comparison of the total number of local and distal eQTL with a minimum LOD score of 8 in each tissue. All tissues have comparable numbers of eQTL. Local eQTL are much more numerous than distal eQTL. **G.** Counts of transcripts with local eQTL shared across multiple tissues. The majority of local eQTL were shared across all four tissues. **H.** Counts of transcripts with distal eQTL shared across multiple tissues. The majority of distal eQTL were tissue-specific and not shared across multiple tissues. For both G and H, eQTL for a given transcript were considered shared in two tissues if they were within 4Mb of each other. Colored bars indicate the counts for individual tissues for easy of visualization.

415 **References**

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

## KEGG pathway enrichments by GSEA

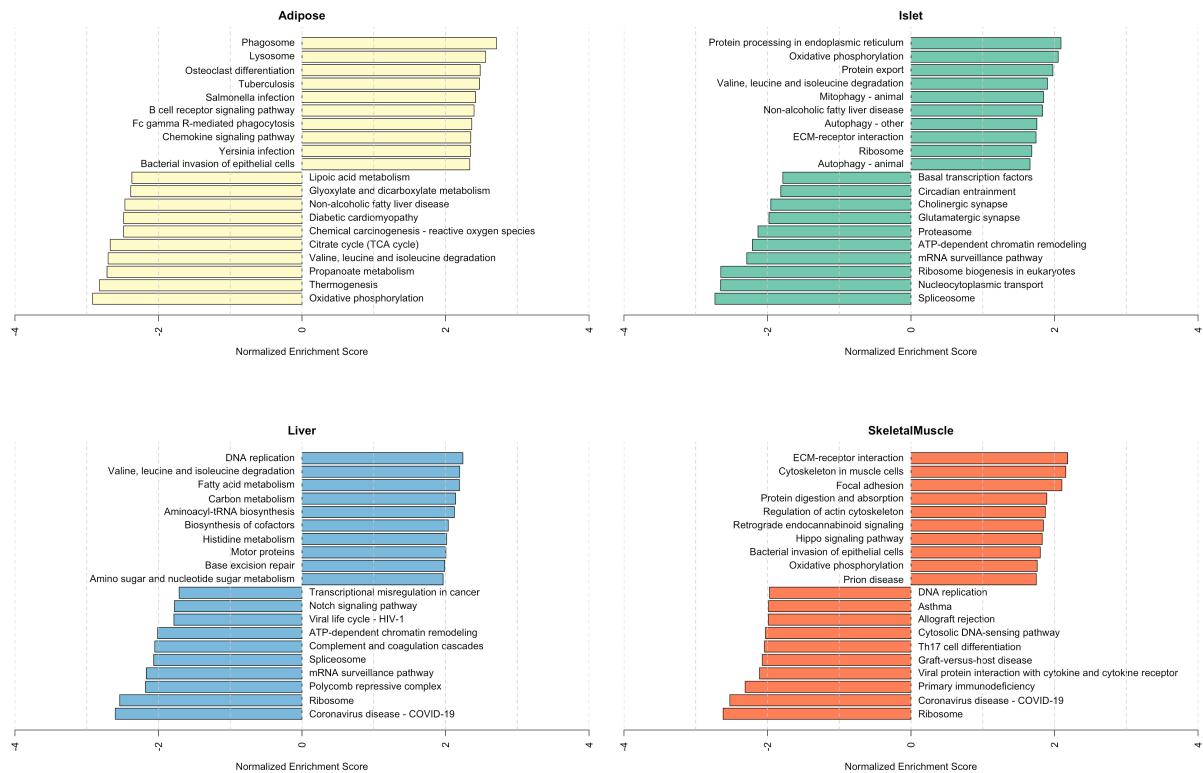


Figure 10: Bar plots showing normalized enrichment scores (NES) for KEGG pathways as determined by fast gene score enrichment analysis (fgsea). Only the top 10 positive and top 10 negative scores are shown. Colors indicate tissue. The name beside each bar shows the name of each enriched KEGG pathway.

- 420 Stamatoyanopoulos. Systematic localization of common disease-associated variation in regulatory DNA.  
 421 *Science*, 337(6099):1190–1195, Sep 2012.
- 422 [2] K. K. Farh, A. Marson, J. Zhu, M. Kleinewietfeld, W. J. Housley, S. Beik, N. Shores, H. Whitton, R. J.  
 423 Ryan, A. A. Shishkin, M. Hatan, M. J. Carrasco-Alfonso, D. Mayer, C. J. Luckey, N. A. Patsopoulos,  
 424 P. L. De Jager, V. K. Kuchroo, C. B. Epstein, M. J. Daly, D. A. Hafler, and B. E. Bernstein. Genetic  
 425 and epigenetic fine mapping of causal autoimmune disease variants. *Nature*, 518(7539):337–343, Feb  
 426 2015.
- 427 [3] E. Pennisi. The Biology of Genomes. Disease risk links to gene regulation. *Science*, 332(6033):1031, May  
 428 2011.
- 429 [4] L. A. Hindorff, P. Sethupathy, H. A. Junkins, E. M. Ramos, J. P. Mehta, F. S. Collins, and T. A. Manolio.

## Top GO term enrichments by GSEA

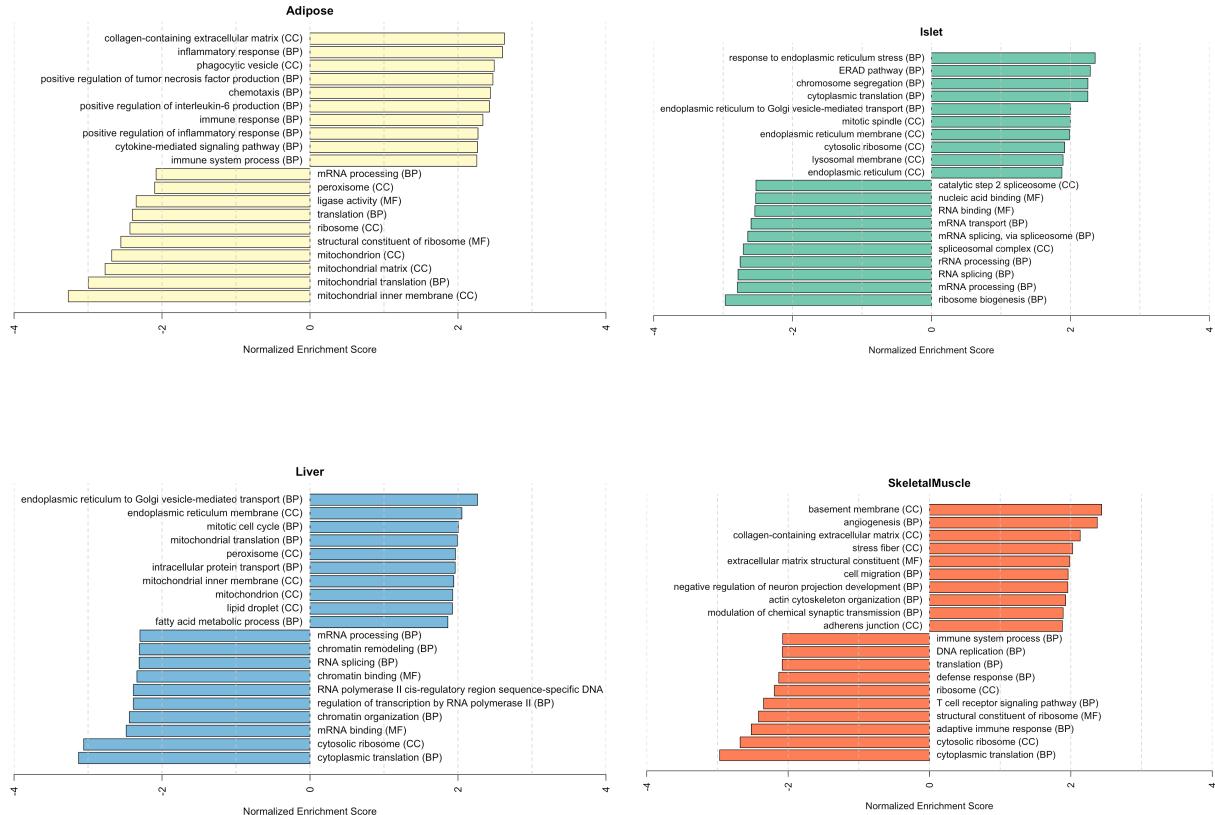


Figure 11: Bar plots showing normalized enrichment scores (NES) for GO terms as determined by fast gene score enrichment analysis (fgsea). Only the top 10 positive and top 10 negative scores are shown. Colors indicate tissue. The name beside each bar shows the name of each enriched GO term. The letters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

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

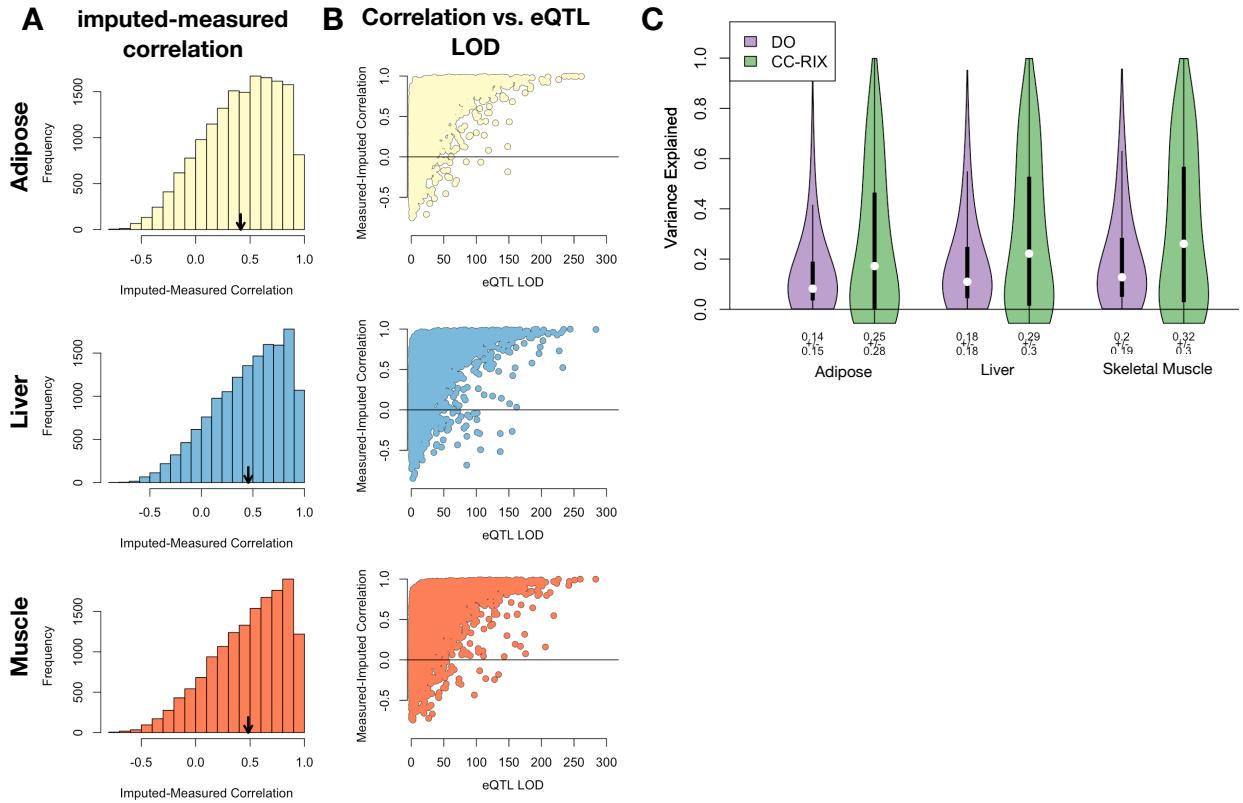


Figure 12: Validation of transcript imputation in the CC-RIX. **A.** Distributions of correlations between imputed and measured transcripts in the CC-RIX. The mean of each distribution is shown by the red line. All distributions were skewed toward positive correlations and had positive means near a Pearson correlation ( $r$ ) of 0.5. **B.** The relationship between the correlation between measured and imputed expression in the CC-RIX (x-axis) and eQTL LOD score. As expected, imputations are more accurate for transcripts with strong local eQTL. **C.** Variance explained by local genotype in the DO and CC-RIX.

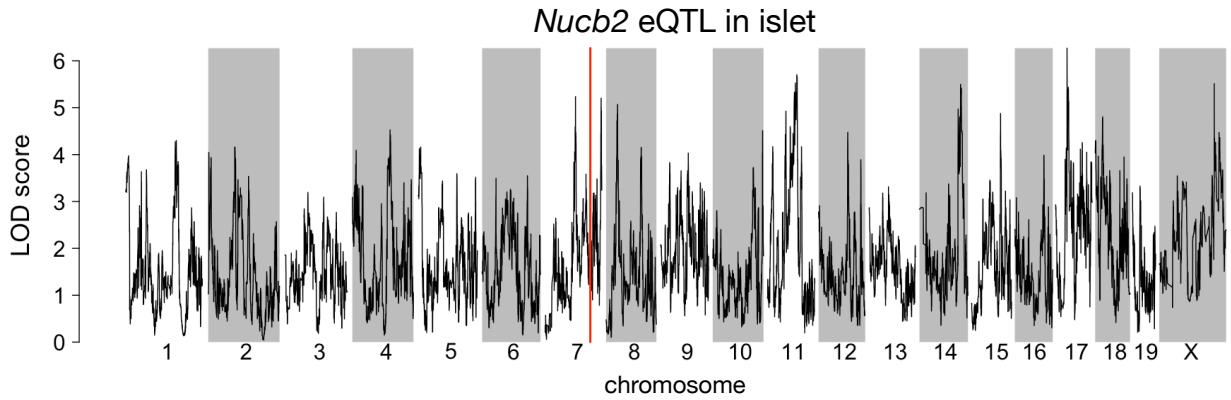


Figure 13: Regulation of *Nucb2* expression in islet. *Nucb2* is encoded on mouse chromosome 7 at 116.5 Mb (red line). In islets the heritability of *Nucb2* expression levels is 69% heritable. This LOD score trace shows that there is no local eQTL at that position, nor any strong distal eQTL anywhere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

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

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

- 497 [29] D. D. Sears, G. Hsiao, A. Hsiao, J. G. Yu, C. H. Courtney, J. M. Ofrecio, J. Chapman, and S. Subramaniam.  
498 Mechanisms of human insulin resistance and thiazolidinedione-mediated insulin sensitization. *Proc Natl  
499 Acad Sci U S A*, 106(44):18745–18750, Nov 2009.
- 500 [30] R. Stienstra, C. Duval, M. ller, and S. Kersten. PPARs, Obesity, and Inflammation. *PPAR Res*,  
501 2007:95974, 2007.
- 502 [31] O. Gavrilova, M. Haluzik, K. Matsusue, J. J. Cutson, L. Johnson, K. R. Dietz, C. J. Nicol, C. Vinson,  
503 F. J. Gonzalez, and M. L. Reitman. Liver peroxisome proliferator-activated receptor gamma contributes  
504 to hepatic steatosis, triglyceride clearance, and regulation of body fat mass. *J Biol Chem*, 278(36):34268–  
505 34276, Sep 2003.
- 506 [32] K. Matsusue, M. Haluzik, G. Lambert, S. H. Yim, O. Gavrilova, J. M. Ward, B. Brewer, M. L. Reitman,  
507 and F. J. Gonzalez. Liver-specific disruption of PPARgamma in leptin-deficient mice improves fatty  
508 liver but aggravates diabetic phenotypes. *J Clin Invest*, 111(5):737–747, Mar 2003.
- 509 [33] D. Patsouris, J. K. Reddy, M. ller, and S. Kersten. Peroxisome proliferator-activated receptor alpha  
510 mediates the effects of high-fat diet on hepatic gene expression. *Endocrinology*, 147(3):1508–1516, Mar  
511 2006.
- 512 [34] S. E. Schadinger, N. L. Bucher, B. M. Schreiber, and S. R. Farmer. PPARgamma2 regulates lipogenesis  
513 and lipid accumulation in steatotic hepatocytes. *Am J Physiol Endocrinol Metab*, 288(6):E1195–1205,  
514 Jun 2005.
- 515 [35] W. Motomura, M. Inoue, T. Ohtake, N. Takahashi, M. Nagamine, S. Tanno, Y. Kohgo, and T. Okumura.  
516 Up-regulation of ADRP in fatty liver in human and liver steatosis in mice fed with high fat diet. *Biochem  
517 Biophys Res Commun*, 340(4):1111–1118, Feb 2006.
- 518 [36] J. Lamb, E. D. Crawford, D. Peck, J. W. Modell, I. C. Blat, M. J. Wrobel, J. Lerner, J. P. Brunet,  
519 A. Subramanian, K. N. Ross, M. Reich, H. Hieronymus, G. Wei, S. A. Armstrong, S. J. Haggarty,  
520 P. A. Clemons, R. Wei, S. A. Carr, E. S. Lander, and T. R. Golub. The Connectivity Map: using  
521 gene-expression signatures to connect small molecules, genes, and disease. *Science*, 313(5795):1929–1935,  
522 Sep 2006.
- 523 [37] S. Amin, A. Lux, and F. O’Callaghan. The journey of metformin from glycaemic control to mTOR  
524 inhibition and the suppression of tumour growth. *Br J Clin Pharmacol*, 85(1):37–46, Jan 2019.
- 525 [38] A. Kezic, L. Popovic, and K. Lalic. mTOR Inhibitor Therapy and Metabolic Consequences: Where Do  
526 We Stand? *Oxid Med Cell Longev*, 2018:2640342, 2018.

- 527 [39] A. D. Barlow, M. L. Nicholson, and T. P. Herbert. -cells and a review of the underlying molecular  
528 mechanisms. *Diabetes*, 62(8):2674–2682, Aug 2013.
- 529 [40] Y. Gu, J. Lindner, A. Kumar, W. Yuan, and M. A. Magnuson. Rictor/mTORC2 is essential for  
530 maintaining a balance between beta-cell proliferation and cell size. *Diabetes*, 60(3):827–837, Mar 2011.
- 531 [41] E. B. Geer, J. Islam, and C. Buettner. Mechanisms of glucocorticoid-induced insulin resistance: focus on  
532 adipose tissue function and lipid metabolism. *Endocrinol Metab Clin North Am*, 43(1):75–102, Mar 2014.
- 533 [42] J. X. Li and C. L. Cummins. Fresh insights into glucocorticoid-induced diabetes mellitus and new  
534 therapeutic directions. *Nat Rev Endocrinol*, 18(9):540–557, Sep 2022.
- 535 [43] R. A. Lee, C. A. Harris, and J. C. Wang. Glucocorticoid Receptor and Adipocyte Biology. *Nucl Receptor*  
536 Res, 5, 2018.
- 537 [44] S. Viengchareun, P. Penfornis, M. C. Zennaro, and M. s. Mineralocorticoid and glucocorticoid receptors  
538 inhibit UCP expression and function in brown adipocytes. *Am J Physiol Endocrinol Metab*, 280(4):E640–  
539 649, Apr 2001.
- 540 [45] J. Liu, X. Kong, L. Wang, H. Qi, W. Di, X. Zhang, L. Wu, X. Chen, J. Yu, J. Zha, S. Lv, A. Zhang,  
541 P. Cheng, M. Hu, Y. Li, J. Bi, Y. Li, F. Hu, Y. Zhong, Y. Xu, and G. Ding. -HSD1 in regulating brown  
542 adipocyte function. *J Mol Endocrinol*, 50(1):103–113, Feb 2013.
- 543 [46] L. E. Ramage, M. Akyol, A. M. Fletcher, J. Forsythe, M. Nixon, R. N. Carter, E. J. van Beek,  
544 N. M. Morton, B. R. Walker, and R. H. Stimson. Glucocorticoids Acutely Increase Brown Adipose  
545 Tissue Activity in Humans, Revealing Species-Specific Differences in UCP-1 Regulation. *Cell Metab*,  
546 24(1):130–141, Jul 2016.
- 547 [47] J. L. Barclay, H. Agada, C. Jang, M. Ward, N. Wetzig, and K. K. Ho. Effects of glucocorticoids on  
548 human brown adipocytes. *J Endocrinol*, 224(2):139–147, Feb 2015.
- 549 [48] M. ó, R. Gupte, W. L. Kraus, P. Pacher, and P. Bai. PARPs in lipid metabolism and related diseases.  
550 *Prog Lipid Res*, 84:101117, Nov 2021.
- 551 [49] P. Bai, C. ó, H. Oudart, A. nszki, Y. Cen, C. Thomas, H. Yamamoto, A. Huber, B. Kiss, R. H.  
552 Houtkooper, K. Schoonjans, V. Schreiber, A. A. Sauve, J. Menissier-de Murcia, and J. Auwerx. PARP-1  
553 inhibition increases mitochondrial metabolism through SIRT1 activation. *Cell Metab*, 13(4):461–468,  
554 Apr 2011.

- 555 [50] A. Chiarugi and M. A. Moskowitz. Cell biology. PARP-1—a perpetrator of apoptotic cell death? *Science*,  
556 297(5579):200–201, Jul 2002.
- 557 [51] M. Althubiti, R. Almaimani, S. Y. Eid, M. Elzubaier, B. Refaat, S. Idris, T. A. Alqurashi, and M. Z.  
558 El-Readi. BTK targeting suppresses inflammatory genes and ameliorates insulin resistance. *Eur Cytokine  
559 Netw*, 31(4):168–179, Dec 2020.
- 560 [52] C. Skrabs, W. F. Pickl, T. Perkmann, U. ger, and A. Gessl. Rapid decline in insulin antibodies and  
561 glutamic acid decarboxylase autoantibodies with ibrutinib therapy of chronic lymphocytic leukaemia. *J  
562 Clin Pharm Ther*, 43(1):145–149, Feb 2018.
- 563 [53] E. A. Oral, S. M. Reilly, A. V. Gomez, R. Meral, L. Butz, N. Ajluni, T. L. Chenevert, E. Korytnaya,  
564 A. H. Neidert, R. Hench, D. Rus, J. F. Horowitz, B. Poirier, P. Zhao, K. Lehmann, M. Jain, R. Yu,  
565 C. Liddle, M. Ahmadian, M. Downes, R. M. Evans, and A. R. Saltiel. and TBK1 Improves Glucose  
566 Control in a Subset of Patients with Type 2 Diabetes. *Cell Metab*, 26(1):157–170, Jul 2017.
- 567 [54] M. C. Arkan, A. L. Hevener, F. R. Greten, S. Maeda, Z. W. Li, J. M. Long, A. Wynshaw-Boris, G. Poli,  
568 J. Olefsky, and M. Karin. IKK-beta links inflammation to obesity-induced insulin resistance. *Nat Med*,  
569 11(2):191–198, Feb 2005.
- 570 [55] M. Clark, C. J. Kroger, Q. Ke, and R. M. Tisch. The Role of T Cell Receptor Signaling in the  
571 Development of Type 1 Diabetes. *Front Immunol*, 11:615371, 2020.
- 572 [56] P. ndell, A. C. Carlsson, A. Larsson, O. Melander, T. Wessman, J. v, and T. Ruge. TNFR1 is associated  
573 with short-term mortality in patients with diabetes and acute dyspnea seeking care at the emergency  
574 department. *Acta Diabetol*, 57(10):1145–1150, Oct 2020.
- 575 [57] A. M. Keestra-Gounder, M. X. Byndloss, N. Seyffert, B. M. Young, A. vez Arroyo, A. Y. Tsai, S. A.  
576 Cevallos, M. G. Winter, O. H. Pham, C. R. Tiffany, M. F. de Jong, T. Kerrinnes, R. Ravindran, P. A.  
577 Luciw, S. J. McSorley, A. J. umler, and R. M. Tsolis. NOD1 and NOD2 signalling links ER stress with  
578 inflammation. *Nature*, 532(7599):394–397, Apr 2016.
- 579 [58] A. M. Keestra-Gounder and R. M. Tsolis. NOD1 and NOD2: Beyond Peptidoglycan Sensing. *Trends  
580 Immunol*, 38(10):758–767, Oct 2017.
- 581 [59] J. Montane, L. Cadavez, and A. Novials. Stress and the inflammatory process: a major cause of  
582 pancreatic cell death in type 2 diabetes. *Diabetes Metab Syndr Obes*, 7:25–34, 2014.
- 583 [60] B. B. Kahn and T. E. McGraw. , and type 2 diabetes. *N Engl J Med*, 363(27):2667–2669, Dec 2010.

- 584 [61] S. Del Prato and N. Pulizzi. The place of sulfonylureas in the therapy for type 2 diabetes mellitus.  
585       *Metabolism*, 55(5 Suppl 1):S20–27, May 2006.
- 586 [62] X. Liu, Y. I. Li, and J. K. Pritchard. Trans Effects on Gene Expression Can Drive Omnipotent Inheritance.  
587       *Cell*, 177(4):1022–1034, May 2019.
- 588 [63] B. Hallgrímsson, R. M. Green, D. C. Katz, J. L. Fish, F. P. Bernier, C. C. Roseman, N. M. Young,  
589       J. M. Cheverud, and R. S. Marcucio. The developmental-genetics of canalization. *Semin Cell Dev Biol*,  
590       88:67–79, Apr 2019.
- 591 [64] M. L. Siegal and A. Bergman. Waddington's canalization revisited: developmental stability and evolution.  
592       *Proc Natl Acad Sci U S A*, 99(16):10528–10532, Aug 2002.
- 593 [65] A. B. Paaby and G. Gibson. Cryptic Genetic Variation in Evolutionary Developmental Genetics. *Biology*  
594       (Basel), 5(2), Jun 2016.
- 595 [66] E. A. Boyle, Y. I. Li, and J. K. Pritchard. An Expanded View of Complex Traits: From Polygenic to  
596       Omnipotent. *Cell*, 169(7):1177–1186, Jun 2017.
- 597 [67] Naomi R Wray, Cisca Wijmenga, Patrick F Sullivan, Jian Yang, and Peter M Visscher. Common disease  
598       is more complex than implied by the core gene omnipotent model. *Cell*, 173(7):1573–1580, 2018.
- 599 [68] M. Riva, M. D. Nitert, U. Voss, R. Sathanoori, A. Lindqvist, C. Ling, and N. Wierup. Nesfatin-1  
600       stimulates glucagon and insulin secretion and beta cell NUCB2 is reduced in human type 2 diabetic  
601       subjects. *Cell Tissue Res*, 346(3):393–405, Dec 2011.
- 602 [69] M. Nakata and T. Yada. Role of NUCB2/nestin-1 in glucose control: diverse functions in islets,  
603       adipocytes and brain. *Curr Pharm Des*, 19(39):6960–6965, 2013.
- 604 [70] H. Shimizu and A. Osaki. Nesfatin/Nucleobindin-2 (NUCB2) and Glucose Homeostasis. *Curr Hypertens*  
605       Rev, pages Nesfatin/Nucleobindin-2 (NUCB2) and Glucose Homeostasis., Jul 2014.