

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 genetic contributions to complex traits, gene expression is an important bridge
10 between genotype and phenotype. The majority of variants identified in GWAS are in regulatory regions of
11 the genome [cite], suggesting that they influence clinical phenotypes through regulation of gene expression.
12 Consistent with this idea, powerful insights into disease mechanism can be made by linking genetic variants
13 affecting gene expression (expression quantitative trait loci - eQTLs) to phenotypes. In particular, mediation
14 analysis has been used to identify transcripts that mediate the effect of genetic variants on phenotypes. In
15 mice... [1, 2, 3] (bmediatR) [4] In humans... [5, 6]

16 Thus far, the primary focus of expression mediated traits has been on local genetic variation; that is genetic
17 variation that influences the transcription of local genes, thereby causing variation in traits. However, there
18 is evidence that the bulk of disease heritability is mediated by the distal component of gene expression,
19 rather than the local component [7]. Yao et al. [cite] observed that genes with low local heritability explain
20 more expression-mediated disease heritability than genes with high local heritability. We have observed a
21 similar pattern in mice, which we describe here. Thus, identifying heritable components of complex traits
22 that are mediated through distally regulated variation in gene expression may provide important insights
23 into mechanisms regulating complex traits.

24 Identification of distal factors influencing gene expression and traits is challenging, as the multiple testing
25 corrections are much more severe for distal effects [@pmid24013639]. However, systems approaches that

26 consider the entire transcriptome simultaneously and avoid univariate testing provide promising avenues for
27 identification of broad transcriptomic patterns influencing complex traits that provide both biological insight
28 and targets for therapeutics. Here we propose high-dimenaional mediation (HDM) as one such systems
29 approach for identification of the heritable portion of the transcriptome that mediates the effect of the genome
30 on phenome. HDM uses a regularized and generalized canonical correlation analysis (RGCCA) [cite], which
31 is an extension of canonical correlation analysis (CCA) that allows for more than two data sets with an
32 arbitrary relationship among them. Thus, we can identify linear combinations of the genome, transcriptome,
33 and phenome, that describe the mediation of the genetic effects on the phenome through the transcriptome.
34 Because of the central dogma of molecular biology, information flow is directed out of the genome, and not
35 back into it. Thus, the otherwise undirected relationships between genome, transcriptome, and phenome can
36 be inferred as a causal mediation by the transcriptome of the effects of the genome on the phenome.

37 Here we apply HDM

38 Results

39 Genetic variation contributes to wide phenotypic variation

40 A population of 500 diversity outbred mice (XXX male and XXX female), was placed on a high-fat (XXX/%),
41 high-sugar (XXX/%) diet starting at XXX weeks of age as described previously [cite]. Each individual was
42 assessed longitudinally for multiple metabolic measures including fasting glucose levels, glucose tolerance,
43 insulin levels, body weight, and blood lipid levels (Methods).

44 Although the environment was consistent across animals, the genetic diversity present in this population
45 resulted in widely varying distributions across physiological measurements (Fig. 1). For example, body
46 weights of adult individuals varied from less than the average adult B6 body weight to several times the
47 body weight of a B6 adult in both sexes (Fig. 1A). Fasting blood glucose also varied considerably (Fig.
48 1B) although few of the animals had FBG levels that would indicate pre-diabetes (XXX/%), or diabetes
49 (XXX/%) according to previously developed cutoffs (pre-diabetes: $\text{FBG} \geq 250 \text{ mg/dL}$, diabetes: $\text{FBG} \geq 300$
50 mg/dL) [8]. Males had higher FBG than females on average (Fig. 1C) as has been observed before suggesting
51 either that males were more susceptible to metabolic disease on the high-fat diet, or that males and females
52 may require different thresholds for pre-diabetes and diabetes.

53 Body weight was strongly positively correlated with food consumption (Fig. 1D $R^2 = 0.31$) and fasting blood
54 glucose (FBG) (Fig. 1E, $R^2 = 0.21$) suggesting a link between behavioral factors and metabolic disease.
55 However, the heritability of this trait and others (Fig. 1F) indicates that background genetics contribute

56 substantially to correlates of metabolic disease in this population.

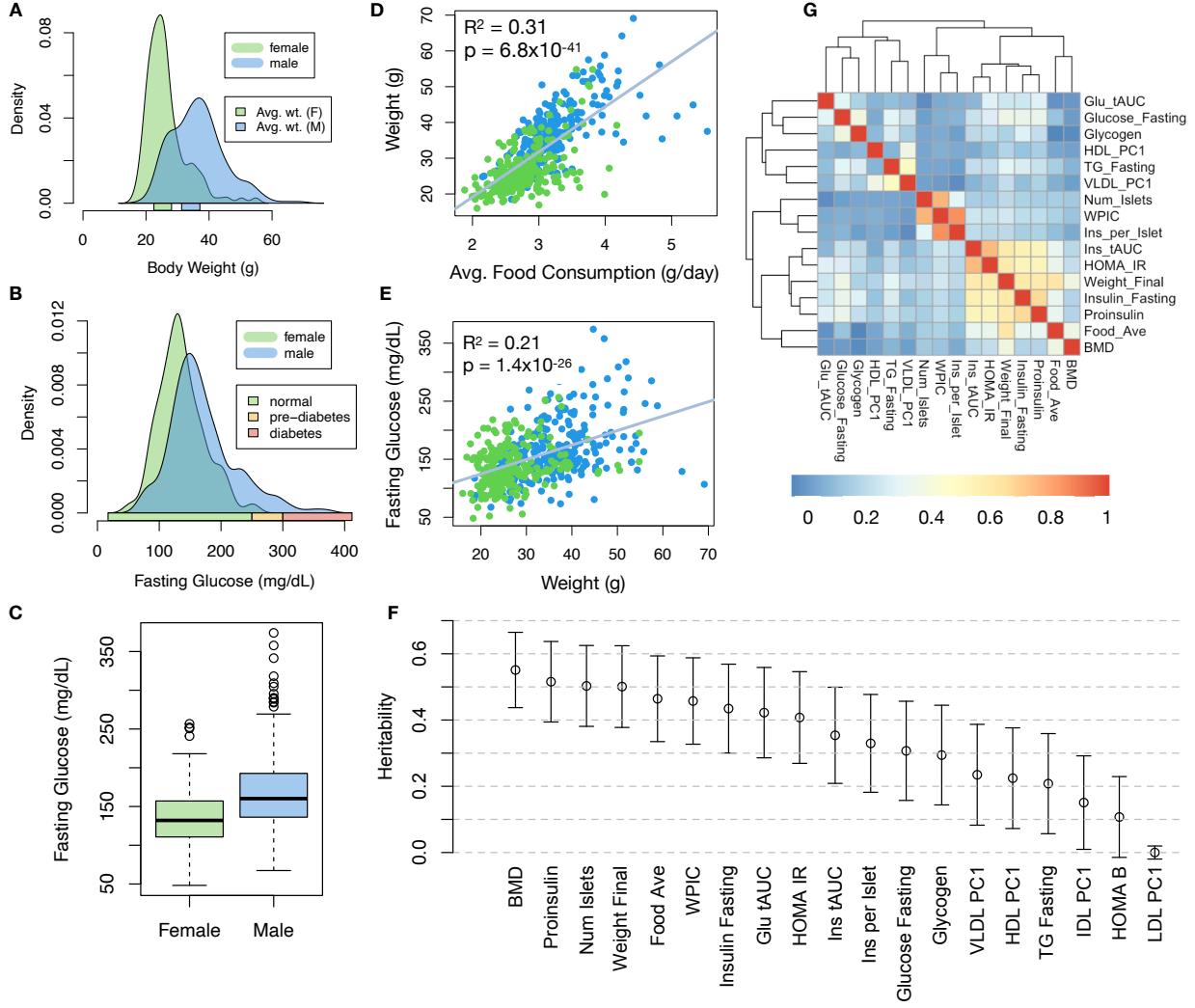


Figure 1: Clinical overview. **A.** Distributions of final body weight in the diversity outbred mice. Sex is indicated by color. The average B6 male and female adult weights at 24 weeks of age are indicated by blue and green bars on the x-axis. **B.** The distribution of final fasting glucose across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

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

61 **Distal Heritability Correlates with Phenotype Relevance**

62 To elaborate the mechanistic details of genetic effects on metabolic phenotypes in the DO population, we
63 also measured gene expression in four tissues known to be involved in metabolic disease: adipose, pancreatic
64 islet, liver, and skeletal muscle. To confirm the heritability of transcript levels, we performed expression QTL
65 analysis using R/qtl2 [cite] (Methods) and identified both local and distal eQTL for transcripts in each tissue
66 (Supp. Fig 9). Significant local eQTLs far outnumbered distal eQTLs (Supp. Fig. 9F) and tended to be
67 shared across tissues (Supp. Fig. 9G) whereas the few significant distal eQTL we identified tended to be
68 tissue-specific (Supp. Fig. 9H)

69 To better compare the relative contribution of local and distal genetics to transcript levels, we performed a
70 heritability analysis for each transcript (Methods). Overall, local and distal factors contributed approximately
71 equally to transcript abundance. In all tissues, both local and distal factors explained between 13 and 19% of
72 the variance in the median transcript (Fig 2A).

73 Local heritability of transcripts was negatively correlated with their trait relevance, defined as the maximum
74 correlation of a transcript across all traits (Fig. 2B). This suggests that the more local genotype influenced
75 transcript abundance, the less effect variation in transcript abundance was related to the measured traits.
76 Conversely, distal heritability of transcripts was positively correlated with trait relevance (Fig. 2C). That is,
77 transcripts that were more highly correlated with the measured traits tended to be distally, rather than locally,
78 heritable. That trait-correlated transcripts have low local heritability is consistent with previous observations
79 that low-heritability transcripts explain more expression-mediated disease heritability than high-heritability
80 transcripts [7]. However, the positive relationship between trait correlation and distal heritability suggests
81 that there are alternative mechanisms through which genetic regulation of transcripts may influence traits.

82 **High-Dimensional Mediation identifies composite transcript that perfectly mediates composite
83 trait**

84 To identify mechanisms through which genetic regulation of transcripts influences heritable traits, we propose
85 high-dimensional mediation (HDM) (Fig. 3). In this process we kernelize each of the genome, transcriptome,
86 and phenotype, and perform regularized and sparse generalized canonical correlation analysis (RGCCA) [cite]
87 in which we explicitly model the mediation by the transcriptome of the effect of the genome on the phenotype
88 (Methods, Fig. 3). RGCCA is an extended form of canonical correlation analysis (CCA) [cite] in which
89 multiple data sets can be analyzed simultaneously with explicit relationships.

90 The result of this process is three vectors representing the composite genome (G_C), composite transcriptome

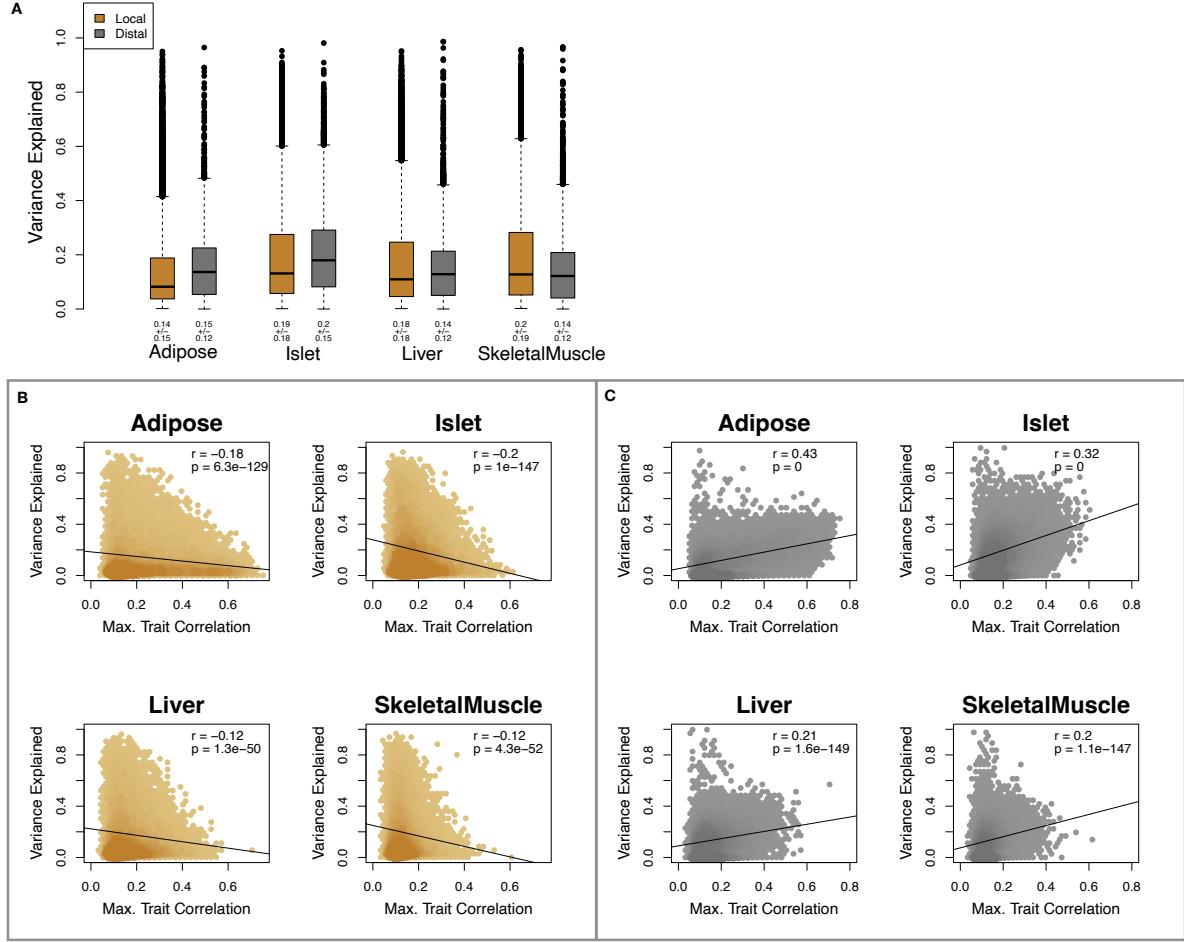


Figure 2: Transcript heritability and trait relevance. **A.** Distributions of distal and local heritability of transcripts across the four tissues. Overall local and distal factors contribute equally to transcript heritability. The relationship between **(B.)** local and **(C.)** distal heritability and trait relevance across all four tissues. Here trait relevance is defined as the maximum correlation between the transcript and all traits. Local heritability is 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 (T_C) and the composite phenotype (P_C) where the composite transcriptome perfectly mediates the effect of the
 92 composite genome on the composite phenotype. Each vector is of length n where n is the number of individual
 93 mice. Fig. 3A shows the partial correlations between all pairs of composite vectors. The partial correlation r
 94 between G_C and T_S was 0.46, and the partial correlation between T_S and P_S was 0.78. However, when the
 95 transcriptome was taken into account, the partial correlation between G_S and P_S was effectively 0 (-0.01).
 96 Standard CCA is prone to over-fitting because in any two large matrices it can be trivial to identify
 97 highly correlated composite vectors. To assess whether RGCCA was similarly prone to over-fitting in a
 98 high-dimensional space, we performed permutation testing. We permuted the individual labels on the

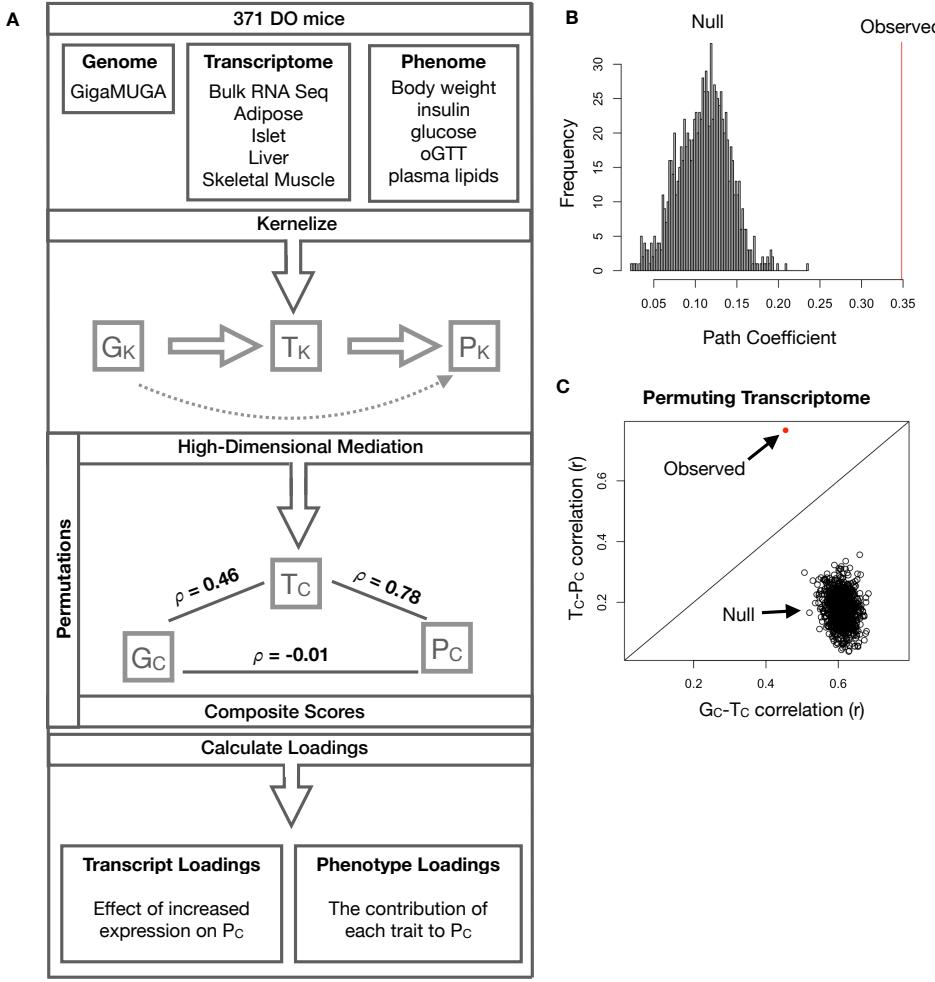


Figure 3: High-dimensional mediation. **A.** Workflow indicating major steps of high-dimensional mediation. The genotype, transcriptome, and phenotype matrices were kernelized to yield single matrices representing the relationships between all individuals for each data modality (G_K = genome kernel,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

99 transcriptome kernel matrix 1000 times and recalculated the path coefficient, which is the partial correlation
 100 of G_C and T_C multiplied by the partial correlation of T_C and P_C . This represents the path from G_C to
 101 P_C that is mediated through T_C . The null distribution of the path coefficient is shown in Fig. 3B, and the
 102 observed path coefficient from the original data is indicated by the red line. The observed path coefficient
 103 was well outside the null distribution generated by permutations. Fig. 3C illustrates this observation in more
 104 detail. Although we identified high correlations between G_C and T_C , and modest correlations between T_C and

105 P_C in the null data (Fig 3C), these two values could not be maximized simultaneously. The red dot shows that
106 in the real data both the G_C - T_C correlation and the T_C - P_C correlation could be maximized simultaneously
107 suggesting that that path from genotype to phenotype through transcriptome is highly non-trivial and
108 identifiable in this case. These results suggest that these composite vectors represent genetically determined
109 variation in phenotype that is mediated through genetically determined variation in transcription.

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

112 The loadings of each measured trait onto P_C indicate how much each contributed to P_C . Final body weight
113 contributed the most to P_C (Fig. 4), followed by homeostatic insulin resistance (HOMA_IR) and fasting
114 plasma insulin levels (Insulin_Fasting). The high loadings of these traits indicate that these are the primary
115 traits mediated by T_C . Traits contributing the least to P_C were measures of cholesterol and pancreas
116 composition. The smaller contributions of these traits indicate a weaker relationship with the heritable
117 transcriptomic signature described by T_C . Thus, when we interpret the transcriptomic signature identified
118 by HDM, we are explaining primarily transcriptional mediation of body weight and insulin resistance, as
119 opposed to cholesterol measurements. Because higher composite trait scores have large, positive contributions
120 from body weight and insulin resistance, larger positive scores for individual mice indicate greater metabolic
121 disease (Fig. 4B)

122 **High-loading transcripts have low local heritability, high distal heritability, and are linked**
123 **mechanistically to obesity**

124 Transcripts that most strongly correlated with T_C were the best mediators of effect of genetics on P_C . Large
125 positive loadings indicate that inheriting higher expression was associated with a higher P_C (higher risk of
126 obesity and metabolic disease on the high-fat diet) (Fig. 4C). Conversely, large negative loadings indicate
127 that inheriting lower expression of these transcripts was associated with a lower P_C (lower risk of obesity and
128 metabolic disease on the high-fat diet) (Fig. 4C). Functional enrichments for the most highly correlated and
129 anti-correlated transcripts are shown in Supp. Fig. 10 and represent known biology of obesity and diabetes.
130 In adipose tissue, for example, the transcripts most strongly correlated with T_C were enriched for immune
131 system signaling and cell motility. It is well established that adipose tissue in obese individuals is highly
132 inflamed [cite] and infiltrated by macrophages [cite]. The transcripts most strongly negatively correlated with
133 T_C were enriched for metabolism of the branched-chain amino acids (BCAA), valine, leucine, and isoleucine.
134 BCAA are used in adipose tissue in lipogenesis, and inhibiting BCAA catabolism inhibits adipogenesis [9].
135 BCAA levels are also related to insulin resistance and are elevated in insulin-resistant obese individuals

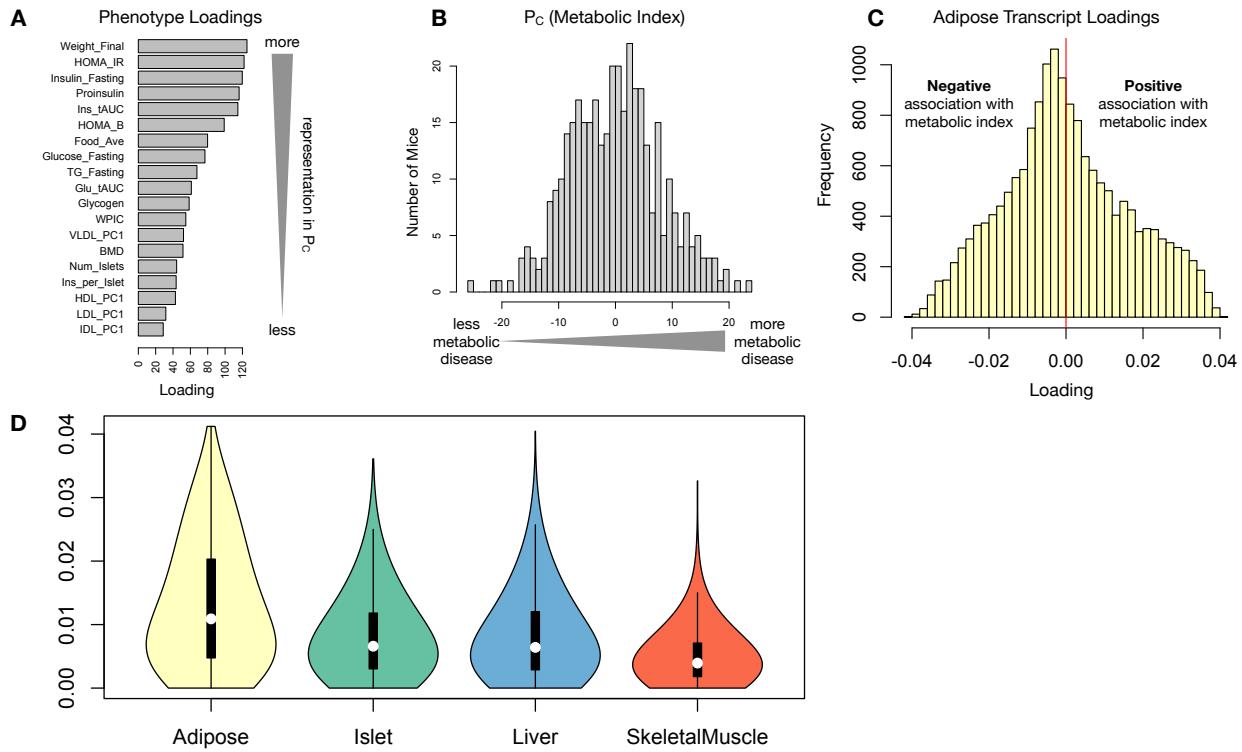


Figure 4: Interpretation of loadings. **A.** Loadings across traits. Body weight and insulin resistance contributed the most to the composite trait. **B.** Phenotype scores across individuals. Individuals with large positive phenotype scores had higher body weight and insulin resistance than average.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

relative to weight-matched non-insulin resistant individuals [10]. In the DO mice studied here, inheriting reduced expression of genes involved in BCAA catabolism was associated with reduced body weight and insulin resistance.

Transcripts in the adipose tissue had the largest loadings, both positive and negative, of all tissues, suggesting that much of the effect of genetics on body weight and insulin resistance is mediated through gene expression in adipose tissue (Fig. 5A). The loadings in liver and pancreas were comparable, and those in skeletal muscle were the weakest (Fig. 5A), suggesting that less of the genetic effects were mediated through transcription in skeletal muscle. Across all tissues, transcripts with the largest loadings tended to have relatively high distal heritability compared with local heritability (Fig. 5A). Transcripts with the highest local heritability tended to have very weak loadings and were 3.6 times less likely to be associated with diabetes and obesity in the literature than transcripts with high loadings (Fig. fig:loading_heritabilityB, Methods). TWAS-nominated

transcripts also had relatively weak loadings and high local heritability (Fig. 4C). They were half as likely as transcripts with the highest loadings to be associated with diabetes and obesity in the literature (Fig. 4C).

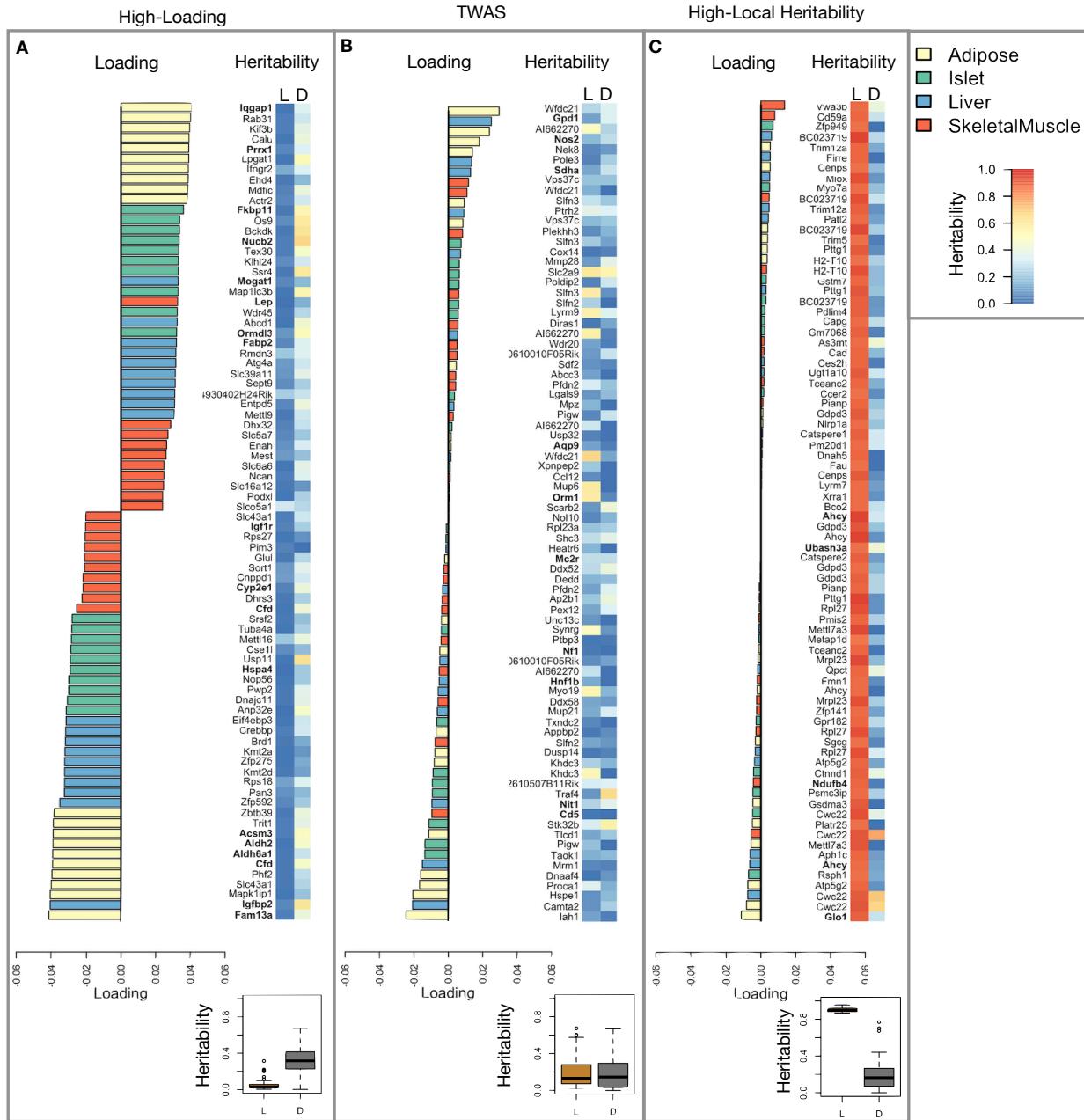


Figure 5: Transcripts with high loadings have high distal heritability and literature support. Each panel has a bar plot showing the loadings of transcripts selected by different criteria. Bar color indicates the tissue of origin. The heat map shows the local (L - left) and distal (D - right) heritability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

149 **Tissue-specific transcriptional programs are associated with metabolic traits**

150 Clustering of transcripts with top loadings in each tissue shows tissue-specific functional modules associated
 151 with obesity and insulin resistance in the DO population (Fig. 6). Many of these modules, such as leptin
 152 signaling in adipose tissue [cite] and skeletal muscle [cite], as well as apelin signaling [cite] have well established
 153 functional roles in diabetes and obesity.

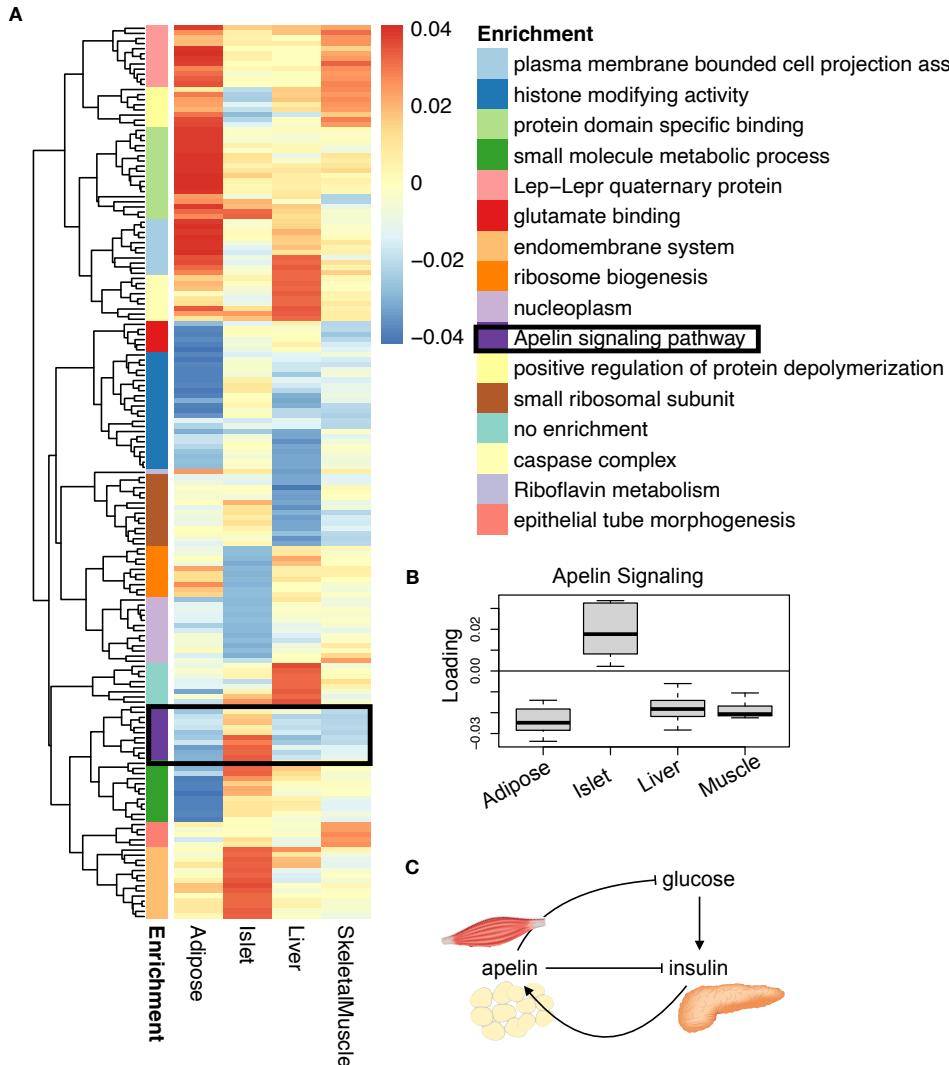


Figure 6: Tissue-specific transcriptional programs are associated with obesity and insulin resistance. **A.** Heat map showing the loadings of all transcripts with loadings greater than 2.5 standard deviations from the mean in any tissue. The heat map is hierarchically clustered. Functional enrichment of each cluster is indicated by color. An example cluster enriched for apelin signaling is highlighted. **B.** The distribution of the loadings of the apelin cluster across all tissues. These transcripts have positive loadings in the islet, and negative loadings in all other tissues. **C.** Model showing the interactions between glucose-stimulated insulin secretion by the pancreas and apelin signaling in skeletal and adipose tissue.

154 **Gene expression, but not local eQTLs predict body weight in an independent population**

155 The loading of each transcript indicates how inherited expression levels influence metabolic phenotypes.
 156 If local regulation is the predominant factor influencing gene expression, we should be able to predict an
 157 individual's phenotype based on their genotypes across all local eQTLs. We tested this hypothesis in an
 158 independent population of F1 mice generated through multiple pairings of Collaborative Cross (CC) [cite]
 159 strains (Fig. 7A) (Methods).

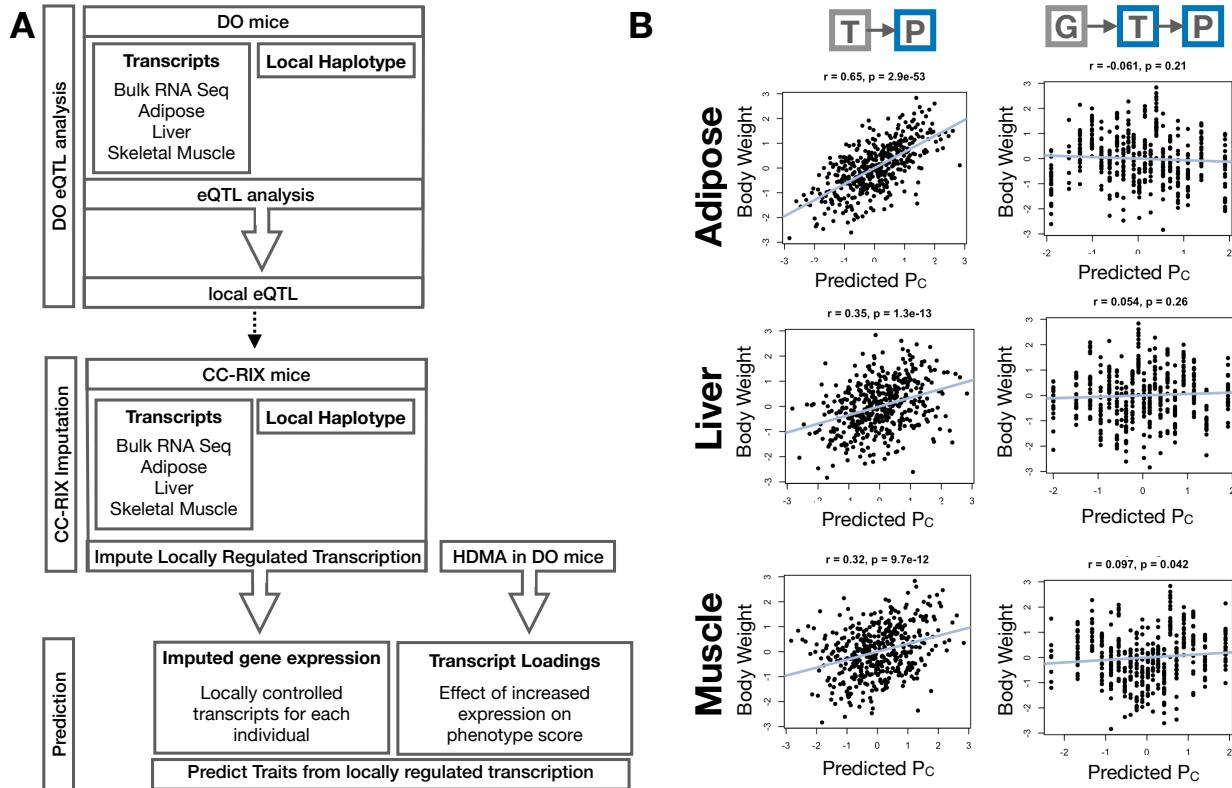


Figure 7: Transcription, but not local genotype, predicts phenotype in the CC-RIX. **A.** Workflow showing procedure for translating HDM results to an independent population of mice. **B.** Relationships between the predicted metabolic index and measured body weight. The left column shows the predictions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

160 We first tested whether the transcript loadings derived from HDM in the DO were relevant to the relationship
 161 between the transcriptome and the phenotype in the CC-RIX. To do this, we multiplied the transcript loadings
 162 derived from HDM in the DO mice by transcript measurements in the CC-RIX standardized across individuals.
 163 This created a transcript vector weighted by importance to metabolic disease as determined in the DO.
 164 The mean of this vector was the predicted metabolic index for the animal based on its transcription in
 165 either adipose tissue, liver, or skeletal muscle. Across all three tissues, weighted transcription values were

166 significantly correlated with metabolic index in the CC-RIX population measured as body weight (Fig. 7B left
167 column). Adipose tissue transcription yielded the most accurate prediction (stats). This result confirms the
168 validity and translatability of the transcript loadings determined in the DO population and their relationship
169 to metabolic disease. It also supports the observation that transcription in adipose tissue is the strongest
170 mediator of genetic effects on metabolic index.

171 We then tested whether this mediation signal was encoded by local genotype. To do this, we imputed gene
172 expression in the CC-RIX using local genotype. We were able to estimate variation in gene transcription
173 robustly. The correlation between measured gene expression and imputed gene expression across all tissues
174 was close to $R = 0.5$, and the variance explained by local genotype was comparable in the DO and CC-RIX
175 (Supp. Fig. 11). However, when weighted with the loadings derived from HDM in the DO population, these
176 imputed transcripts across all tissues failed to predict metabolic index in the CC-RIX (Fig. 7B right column).
177 Taken together, these results support the hypothesis that distal, rather than local genetic factors are primarily
178 driving complex-trait related variation in gene expression.

179 **Distally heritable transcriptomic signatures reflect variation in composition of adipose tissue
180 and islets**

181 Interpretation of global distal genetic influences on gene expression and phenotype is potentially more
182 challenging than interpretation and translation of local genetic influences. Effects can not be located to
183 individual gene variants or transcripts, but because we have a measure of importance across all transcripts in
184 multiple tissues, we can look at global patterns. We noted earlier that functional enrichments of transcripts
185 with large positive loadings in the adipose tissue, suggested that the obese mice in the population had a
186 genetic predisposition toward elevated macrophage infiltration into the adipose tissue. This suggests heritabl
187 variability in cell-type composition of the adipose tissue. We investigated this further bioinformatically
188 by comparing the loadings of cell-type-specific transcripts (Methods). For adipose tissue we used a list of
189 cell-type specific genes identified in human adipose tissue

190 In adipose tissue, the mean loading of macrophage-specific genes was substantially greater than 0 (Fig. 8A),
191 indicating that obese mice were genetically predisposed to have high levels of macrophage infiltration in
192 adipose tissue in response to the high-fat, high-sugar diet.

193 In islet, the mean loadings for alpha-cell specific transcripts were significantly positive, while the mean
194 loadings for delta- and endothelial-cell specific genes were significantly negative (Fig. 8B). These results
195 suggest that obese mice had inherited higher proportions of alpha cells, and lower proportions of endothelial

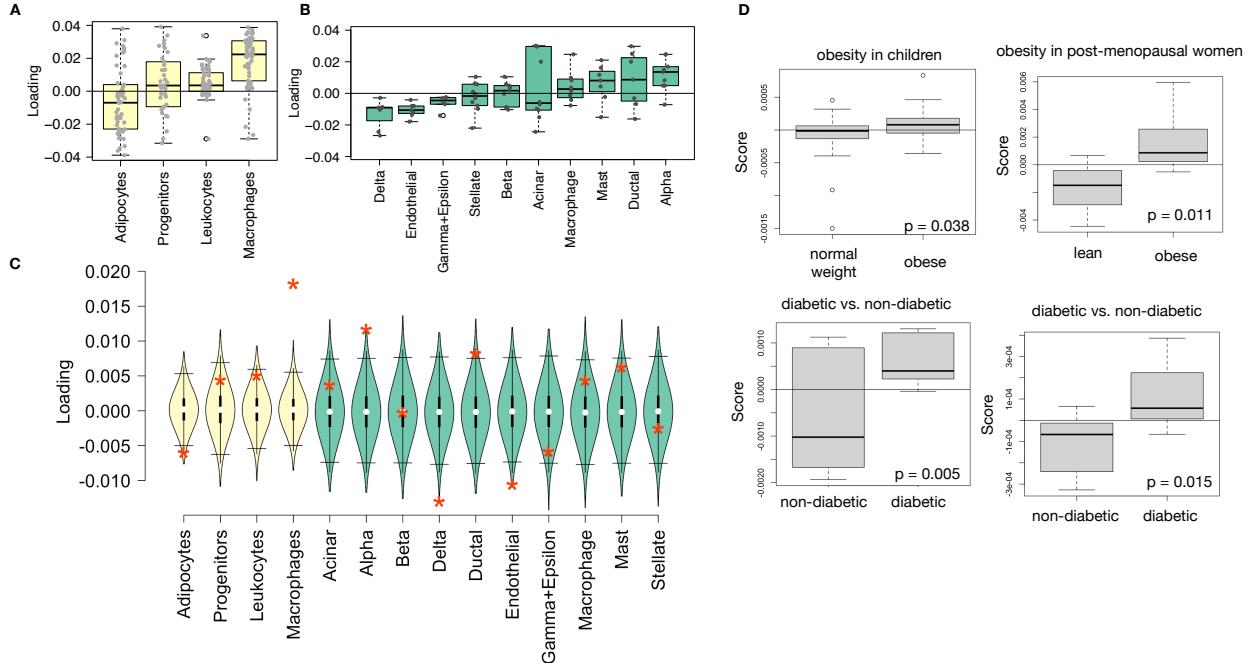


Figure 8: HDM results translate to humans. **A.** Distribution of loadings for cell-type-specific transcripts in adipose tissue. **B.** Distribution of loadings for cell-type-specific transcripts in pancreatic islets (green). **C.** Null distributions for the mean loading of randomly selected transcripts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

196 and delta cells in their pancreatic islets.

197 The loadings for pancreatic beta cell-type specific loadings was not significantly different from zero. This
198 does not reflect on the function of the beta cells in the obese mice, but rather suggests that mice prone to
199 obesity were not obese because they inherited fewer beta cells than non-obese mice.

200 Biological interpretation of alpha, endothelial, delta cells??

201 Distally heritable transcriptomic signatures translate to human disease

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

207 We calculated a predicted obesity score for each individual in the human studies based on their adipose

208 tissue gene expression (Methods) and compared the predicted scores for obese and non-obese groups as well
209 as diabetic and non-diabetic groups. In all cases, the predicted obesity scores were higher on average for
210 individuals in the obese and diabetic groups compared with the lean and non-diabetic groups, indicating that
211 the distally heritable signature of obesity identified in DO mice is relevant to obesity and diabetes in human
212 subjects.

213 **Targeting gene signatures**

214 Although high-loading transcripts are likely good candidates for understanding specific biology related to
215 obesity, we emphasize that the transcriptome overall is highly interconnected and redundant, and that
216 focusing on individual transcripts for treatment may be less effective than using the transcriptomic signature
217 as a whole. The ConnectivityMap (CMAP) database [cite] developed by the Broad Institute allows us to
218 query thousands of compounds that reverse or enhance transcriptomic signatures as a whole in multiple
219 different cell types. By identifying drugs that reverse pathogenic transcriptomic signatures as a whole rather
220 than targeting individual genes, we can potentially increase efficacy of tested compounds.

221 We thus queried the CMAP database through the CLUE online query tool developed by The Broad Institute
222 [cite] (Methods).

223 Alternatively, we can target the gene signature as a whole using CMAP. Identifying drugs to target gene
224 signatures is possible through CMAP. We put our loadings from islet into CMAP. The top hit was PPAR
225 receptor agonist. Rosiglitazone, a widely used diabetes drug, is a PPAR receptor agonist. Another class of
226 drugs on the list was sulfonylureas, which are another major class of drugs for type 2 diabetes.

227 • **Supplemental Table** results from CMAP

228 **Discussion**

229 • distal heritability correlates with phenotype relevance

230 **Data Availability**

231 Here we tell people where to find the data

232 **Acknowledgements**

233 Here we thank people

234 **Supplemental Figures**

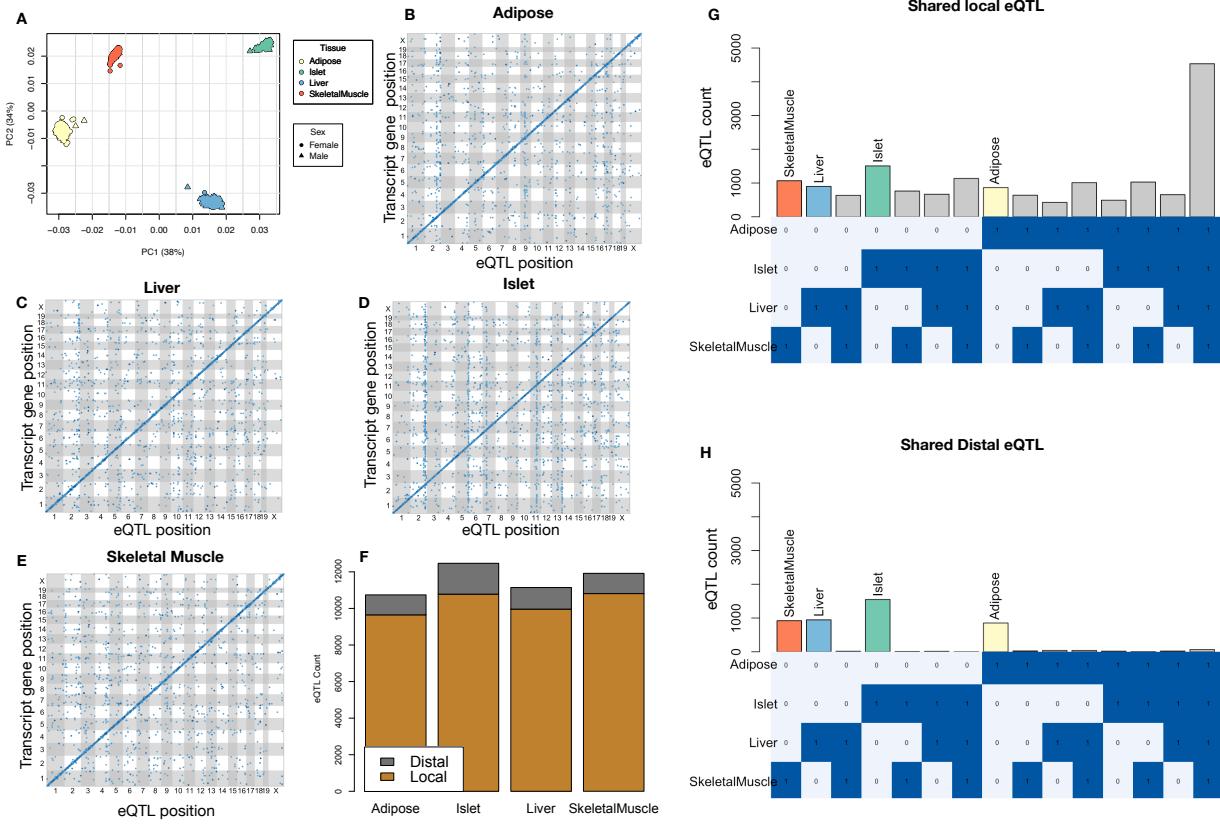


Figure 9: Overview of eQTL analysis in DO mice. **A.** RNA seq samples from the four different tissues clustered by tissue. **B.-E.** eQTL maps are shown for each tissue. The *x*-axis shows the position of the mapped eQTL, and the *y*-axis shows the physical position of the gene encoding each mapped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

Positive Loadings						Negative Loadings							
	term	N-term	N-query	overlap	p_value	domain		term	N-term	N-query	overlap	p_value	domain
Adipose	cell migration	1500	140	43	1.2e-16	GO:BP	KEGG	branched-chain amino acid catabolic process	19	140	8	6.3e-11	GO:BP
	cell motility	1800	140	43	1.7e-14	GO:BP		branched-chain amino acid metabolic process	22	140	8	2.6e-10	GO:BP
	immune system process	2800	140	49	3.9e-11	GO:BP		valine, leucine and isoleucine degradation	56	67	9	1.4e-08	KEGG
	intracellular signal transduction	2600	140	47	4.5e-11	GO:BP		organic acid catabolic process	260	140	15	1.8e-08	GO:BP
	locomotion	1300	140	33	4.7e-11	GO:BP		carboxylic acid catabolic process	260	140	15	1.8e-08	GO:BP
	regulation of cellular component organization	2500	140	46	8.3e-11	GO:BP		small molecule metabolic process	1800	140	35	3.1e-08	GO:BP
	cell surface receptor signaling pathway	2800	140	48	8.6e-11	GO:BP		mitochondrion	1900	140	34	9.3e-08	GO:CC
	positive regulation of response to stimulus	2400	140	44	1.2e-10	GO:BP		carboxylic acid metabolic process	940	140	24	2.9e-07	GO:BP
	vesicle-mediated transport	1500	140	35	2e-10	GO:BP		small molecule catalytic process	390	140	16	4.3e-07	GO:BP
	regulation of cell motility	1000	140	29	4.2e-10	GO:BP		oxoacid metabolic process	960	140	24	4.5e-07	GO:BP
Islet	term	N-term	N-query	overlap	p_value	domain	GO:BP	ribosome biogenesis	320	150	23	5e-16	GO:BP
	mitotic cell cycle process	720	140	31	2.1e-16	GO:BP		small-subunit processome	73	150	10	2.2e-09	GO:CC
	mitotic cell cycle	850	140	31	2.7e-14	GO:BP		preribosome	110	150	11	4.3e-09	GO:CC
	cell cycle process	1300	140	34	4.1e-12	GO:BP		nuclear protein-containing complex	2400	150	41	4.4e-09	GO:CC
	regulation of cell cycle process	730	140	26	2.2e-11	GO:BP		rRNA metabolic process	250	150	15	1.4e-08	GO:BP
	mitotic nuclear division	270	140	17	1.3e-10	GO:BP		rRNA processing	210	150	14	2.3e-08	GO:BP
	regulation of cell cycle	1000	140	29	3.9e-10	GO:BP		nucleocytoplasmic transport	360	150	17	2.4e-08	GO:BP
	organelle fission	500	140	21	4.7e-10	GO:BP		nuclear transport	360	150	17	2.4e-08	GO:BP
	nuclear division	440	140	20	4.7e-10	GO:BP		fibrillar center	150	150	10	3.5e-06	GO:CC
	cell cycle	1800	140	37	1.1e-09	GO:BP		ncRNA metabolic process	640	150	19	3.6e-06	GO:CC
Liver	term	N-term	N-query	overlap	p_value	domain	GO:BP	positive regulation of nitrogen compound metabolic...	2900	150	65	2.1e-21	GO:BP
	small molecule metabolic process	1800	140	39	1e-10	GO:BP		positive regulation of nucleobase-containing compo...	2100	150	55	8.3e-21	GO:BP
	monocarboxylic acid metabolic process	700	140	24	8.5e-10	GO:BP		transcription by RNA polymerase II	2600	150	59	1.2e-19	GO:BP
	oxoacid metabolic process	960	140	27	3.1e-09	GO:BP		regulation of transcription by RNA polymerase II	2500	150	57	9.1e-19	GO:BP
	organic acid metabolic process	970	140	27	3.6e-09	GO:BP		positive regulation of RNA metabolic process	1900	150	50	1.3e-18	GO:BP
	organophosphate metabolic process	1000	140	27	1e-08	GO:BP		positive regulation of DNA-templated transcription	1700	150	48	1.9e-18	GO:BP
	carboxylic acid metabolic process	940	140	26	1.2e-08	GO:BP		positive regulation of RNA biosynthetic process	1700	150	48	2e-18	GO:BP
	catabolic process	2500	140	42	2e-08	GO:BP		positive regulation of macromolecule biosyntheti...	2700	150	57	4.7e-17	GO:BP
	carbohydrate derivative metabolic process	1100	140	27	5.2e-08	GO:BP		protein-DNA complex	860	150	33	1.4e-16	GO:CC
	organic substance catabolic process	2100	140	37	9.4e-08	GO:BP		positive regulation of cellular biosynthetic pro...	2800	150	57	2.7e-16	GO:BP
Muscle	term	N-term	N-query	overlap	p_value	domain	GO:BP	cytosolic ribosome	130	150	16	4.8e-15	GO:CC
	anatomical structure formation involved in morphog...	1200	150	34	6.3e-12	GO:BP		cytoplasmic translation	160	150	16	6.8e-13	GO:BP
	anatomical structure morphogenesis	2800	150	50	1.7e-11	GO:BP		ribosomal subunit	200	150	16	5.6e-12	GO:CC
	circulatory system development	1200	150	33	3.6e-11	GO:BP		cytosolic small ribosomal subunit	45	150	10	1.2e-11	GO:CC
	tube development	1200	150	30	3.5e-09	GO:BP		ribosome	170	85	16	1.6e-10	KEGG
	tube morphogenesis	930	150	26	1.5e-08	GO:BP		small ribosomal subunit	81	150	11	2e-10	GO:CC
	blood vessel development	760	150	23	6.2e-08	GO:BP		translation at presynapse	51	150	10	2.8e-10	GO:BP
	vasculature development	790	150	23	1.4e-07	GO:BP		translation at synapse	52	150	10	3.5e-10	GO:BP
	intracellular signal transduction	2600	150	42	2.1e-07	GO:BP		translation at postsynapse	52	150	10	3.5e-10	GO:BP
	cell migration	1500	150	31	4.9e-07	GO:BP		coronavirus disease - covid-19	240	85	18	3.5e-10	KEGG

Figure 10: Tables showing top 10 functional enrichments for the 150 transcripts with the largest positive and negative loadings across all four tissues.

References

- [1] M. P. Keller, D. M. Gatti, K. L. Schueler, M. E. Rabaglia, D. S. Stapleton, P. Simecek, M. Vincent, S. Allen, A. T. Broman, R. Bacher, C. Kendzierski, K. W. Broman, B. S. Yandell, G. A. Churchill, and A. D. Attie. Genetic Drivers of Pancreatic Islet Function. *Genetics*, 209(1):335–356, May 2018.
- [2] J. H. Kemis, V. Linke, K. L. Barrett, F. J. Boehm, L. L. Traeger, M. P. Keller, M. E. Rabaglia, K. L. Schueler, D. S. Stapleton, D. M. Gatti, G. A. Churchill, D. Amador-Noguez, J. D. Russell, B. S. Yandell, K. W. Broman, J. J. Coon, A. D. Attie, and F. E. Rey. Genetic determinants of gut microbiota composition and bile acid profiles in mice. *PLoS Genet*, 15(8):e1008073, Aug 2019.
- [3] J. M. Chick, S. C. Munger, P. Simecek, E. L. Huttlin, K. Choi, D. M. Gatti, N. Raghupathy, K. L. Svenson, G. A. Churchill, and S. P. Gygi. Author Correction: Defining the consequences of genetic variation on a proteome-wide scale. *Nature*, 606(7915):E16, Jun 2022.

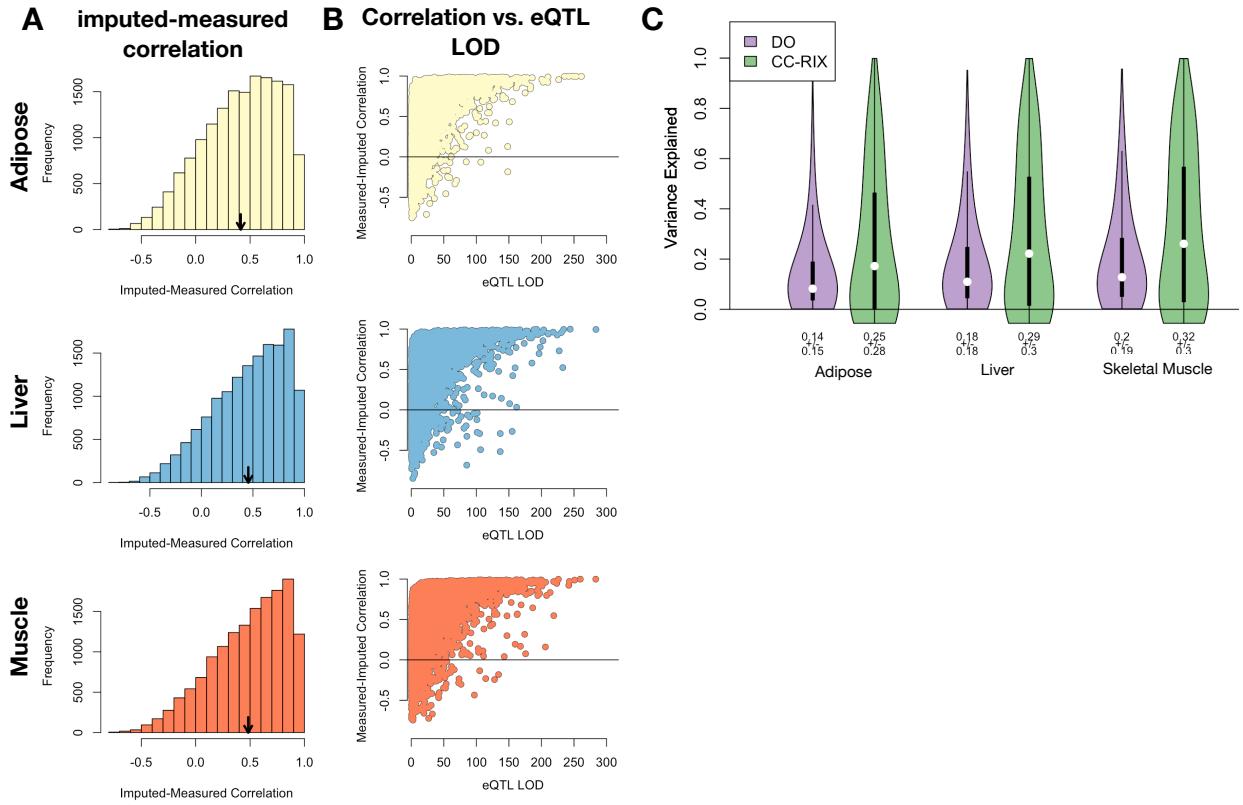


Figure 11: Validation of transcript imputation in the CC-RIX. **A.** Distributions of correlations between imputed and measured transcripts in the CC-RIX. The mean of each distribution is shown by the red line. All distributions were skewed toward positive correlations and had positive means near a Pearson 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.

- 246 [4] W. L. Crouse, G. R. Keele, M. S. Gastonguay, G. A. Churchill, and W. Valdar. A Bayesian model
247 selection approach to mediation analysis. *PLoS Genet*, 18(5):e1010184, May 2022.
- 248 [5] C. Yao, B. H. Chen, R. Joehanes, B. Otlu, X. Zhang, C. Liu, T. Huan, O. Tastan, L. A. Cupples, J. B.
249 Meigs, C. S. Fox, J. E. Freedman, P. Courchesne, C. J. O'Donnell, P. J. Munson, S. Keles, and D. Levy.
250 Integromic analysis of genetic variation and gene expression identifies networks for cardiovascular disease
251 phenotypes. *Circulation*, 131(6):536–549, Feb 2015.
- 252 [6] G. Jiang, A. Chakraborty, Z. Wang, M. Boustani, Y. Liu, T. Skaar, and L. Li. New aQTL SNPs for the
253 CYP2D6 identified by a novel mediation analysis of genome-wide SNP arrays, gene expression arrays,
254 and CYP2D6 activity. *Biomed Res Int*, 2013:493019, 2013.
- 255 [7] D. W. Yao, L. J. O'Connor, A. L. Price, and A. Gusev. Quantifying genetic effects on disease mediated
256 by assayed gene expression levels. *Nat Genet*, 52(6):626–633, Jun 2020.

- 257 [8] S. M. Clee and A. D. Attie. The genetic landscape of type 2 diabetes in mice. *Endocr Rev*, 28(1):48–83,
- 258 Feb 2007.
- 259 [9] C. R. Green, M. Wallace, A. S. Divakaruni, S. A. Phillips, A. N. Murphy, T. P. Ciaraldi, and C. M.
- 260 Metallo. Branched-chain amino acid catabolism fuels adipocyte differentiation and lipogenesis. *Nat Chem Biol*, 12(1):15–21, Jan 2016.
- 261
- 262 [10] D. E. Lackey, C. J. Lynch, K. C. Olson, R. Mostaedi, M. Ali, W. H. Smith, F. Karpe, S. Humphreys,
- 263 D. H. Bedinger, T. N. Dunn, A. P. Thomas, P. J. Oort, D. A. Kieffer, R. Amin, A. Bettaieb, F. G.
- 264 Haj, P. Permana, T. G. Anthony, and S. H. Adams. Regulation of adipose branched-chain amino acid
- 265 catabolism enzyme expression and cross-adipose amino acid flux in human obesity. *Am J Physiol Endocrinol Metab*, 304(11):E1175–1187, Jun 2013.
- 266