

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

³

⁴ Anna L. Tyler, J. Matthew Mahoney, Mark P. Keller, Candice N. Baker, Margaret Gaca, Anuj Srivastava,
⁵ Isabela Gerdes Gyuricza, Madeleine J. Braun, Nadia A. Rosenthal, Alan D. Attie, Gary A. Churchill and
⁶ Gregory W. Carter

⁷ **Abstract**

⁸ Although many genes are subject to local regulation, recent evidence suggests that complex distal regulation
⁹ may be more important in mediating phenotypic variability. To assess the role of distal gene regulation in
¹⁰ complex traits, we combined multi-tissue transcriptomes with physiological outcomes to model diet-induced
¹¹ obesity and metabolic disease in a population of Diversity Outbred mice. Using a novel high-dimensional
¹² mediation analysis, we identified a composite transcriptome signature that summarized genetic effects on
¹³ gene expression and explained 30% of the variation across all metabolic traits. The signature was heritable,
¹⁴ interpretable in biological terms, and predicted obesity status from gene expression in an independently
¹⁵ derived mouse cohort and multiple human studies. Transcripts contributing most strongly to this composite
¹⁶ mediator frequently had complex, distal regulation distributed throughout the genome. These results suggest
¹⁷ that trait-relevant variation in transcription is largely distally regulated, but is nonetheless identifiable,
¹⁸ interpretable, and translatable across species.

¹⁹ **Introduction**

²⁰ Evidence from genome-wide association studies (GWAS) suggests that most heritable variation in complex
²¹ traits is mediated through regulation of gene expression. The majority of trait-associated variants lie in gene
²² regulatory regions^{1–7}, suggesting a relatively simple causal model in which a variant alters the homeostatic
²³ expression level of a nearby gene which, in turn, alters a trait. Statistical methods such as transcriptome-wide
²⁴ association studies (TWAS)^{8–11} and summary data-based Mendelian randomization (SMR)¹⁰ have used this
²⁵ idea to identify genes associated with multiple disease traits^{12–15}. However, despite the great promise of these

26 methods, explaining trait effects with local gene regulation has been more difficult than initially assumed^{16;17}.
27 Although trait-associated variants typically lie in non-coding, regulatory regions, these variants often have no
28 detectable effects on gene expression¹⁶ and tend not to co-localize with expression quantitative trait loci
29 (eQTLs)^{17;18}. These observations suggest that the relationship among genetic variants, gene expression, and
30 organism-level traits is more complex than the simple, local model.

31 In recent years the conversation around the genetic architecture of common disease traits has been addressing
32 this complexity, and there is increased interest in distal effects as potential drivers of trait variation^{18–20;15}.
33 In particular, the omnigenic model posits that trait-driving genes are cumulatively influenced by many distal
34 variants. In this view, the heritable transcriptomic signatures driving clinical traits are an emergent state
35 arising from the myriad molecular interactions defining and constraining gene expression. Consistent with
36 this view, it has been suggested that part of the difficulty in explaining trait variation through local eQTLs
37 may arise in part because gene expression is not measured in the appropriate cell types¹⁶, or cell states²¹,
38 and thus local eQTLs influencing traits cannot be detected in bulk tissue samples. This context dependence
39 emphasizes the essential role of complex regulatory and tissue networks in mediating variant effects. The
40 mechanistic dissection of complex traits in this model is more challenging because it requires addressing
41 network-mediated effects that are weaker and greater in number. However, the comparative importance of
42 distal effects over local effects is currently only conjectured and extremely challenging to address in human
43 populations.

44 To assess the role of wide-spread distal gene regulation in the genetic architecture of complex traits, we used
45 genetically diverse mice as a model system. In mice we can obtain simultaneous measurements of the genome,
46 transcriptome, and phenotype in all individuals. We used diet-induced obesity and metabolic disease as an
47 archetypal example of a complex trait. In humans, these phenotypes are genetically complex with hundreds of
48 variants mapped through GWAS^{22;23} that are known to act through multiple tissues^{24;25}. Likewise in mice,
49 metabolic traits are also genetically complex²⁶ and synteny analysis implicates a high degree of concordance
50 in the genetic architecture between species^{26;12}. Furthermore, in contrast to humans, in mice we have access
51 to multiple disease-relevant tissues in the same individuals with sufficient numbers for adequate statistical
52 power.

53 We generated two complementary data sets: a discovery data set in a large population of Diversity Outbred
54 (DO) mice²⁷, and an independent validation data set derived by crossing inbred strains from the Collaborative
55 Cross (CC) recombinant inbred lines²⁸ to form CC recombinant inbred intercross (CC-RIX) mice. Both
56 populations were maintained on a high-fat, high-sugar diet to model diet-induced obesity and metabolic
57 disease¹².

58 The DO population and CC recombinant inbred lines were derived from the same eight inbred founder
59 strains: five classical lab strains and three strains more recently derived from wild mice²⁷, representing three
60 subspecies and capturing 90% of the known variation in laboratory mice²⁹. The DO mice are maintained
61 with a breeding scheme that ensures equal contributions from each founder across the genome thus rendering
62 almost the whole genome visible to genetic inquiry and maximizing power to detect eQTLs²⁷. The CC mice
63 were initially intercrossed to recombine the genomes from all eight founders, and then inbred for at least 20
64 generations to create recombinant inbred lines^{28;30;29}. Because these two populations have common ancestral
65 haplotypes but highly distinct kinship structure, we could directly and unambiguously compare the local
66 genetic effects on gene expression at the whole-transcriptome level while varying the population structure
67 driving distal regulation.

68 In the DO population, we paired clinically relevant metabolic traits, including body weight and plasma levels
69 of insulin, glucose and lipids¹², with transcriptome-wide gene expression in four tissues related to metabolic
70 disease: adipose tissue, pancreatic islets, liver, and skeletal muscle. We measured similar metabolic traits
71 in a CC-RIX population and gene expression from three of the four tissues used in the DO: adipose tissue,
72 liver, and skeletal muscle. Measuring gene expression in multiple tissues is critical to adequately assess the
73 extent to which local gene regulation varies across the tissues and whether such variability might account for
74 previous failed attempts to identify trait-relevant local eQTLs. The CC-RIX carry the same founder alleles
75 as the DO. Thus, local gene regulation is expected to match between the populations. However, because
76 the alleles are recombined throughout the genome, distal effects are expected to vary from those in the DO,
77 allowing us to directly assess the role of distal gene regulation in driving trait-associated transcript variation.
78 To mechanistically dissect distal effects on metabolic disease, we developed a novel dimension reduction
79 framework called high-dimensional mediation analysis (HDMA) to identify the heritable transcriptomic
80 signatures driving trait variation, which we compared between mouse populations and to human data sets
81 with measured adipose gene expression. Together, these data enable a comprehensive view into the genetic
82 architecture of metabolic disease.

83 Results

84 Genetic variation contributed to wide phenotypic variation

85 Although the environment was consistent across the DO mice, the genetic diversity present in this population
86 resulted in widely varying distributions across physiological measurements (Fig. 1). For example, body
87 weights of adult individuals varied from less than the average adult C57BL/6J (B6) body weight to several
88 times the body weight of a B6 adult in both sexes (Males: 18.5 - 69.1g, Females: 16.0 - 54.8g) (Fig. 1A).

89 Fasting blood glucose (FBG) also varied considerably (Fig. 1B), although few of the animals had FBG levels
90 that would indicate pre-diabetes (19 animals, 3.8%), or diabetes (7 animals, 1.4%) according to previously
91 developed cutoffs (pre-diabetes: $\text{FBG} \geq 250 \text{ mg/dL}$, diabetes: $\text{FBG} \geq 300 \text{ mg/dL}$)³¹. Males had higher
92 FBG than females on average (Fig. 1C) as has been observed before suggesting either that males were more
93 susceptible to metabolic disease on the high-fat, high-sugar (HFHS) diet, or that males and females may
94 require different thresholds for pre-diabetes and diabetes.

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

99 The trait correlations (Fig. 1G) showed that most of the metabolic trait pairs were only modestly correlated,
100 which, in conjunction with the trait decomposition (Supp. Fig. S1), suggests complex relationships among
101 the measured traits and a broad sampling of multiple heritable aspects of metabolic disease including overall
102 body weight, glucose homeostasis, and pancreatic function.

103 **Distal Heritability Correlated with Phenotype Relevance**

104 To comprehensively assess the genetic control of gene expression in metabolic disease we measured overall
105 gene expression via bulk RNA-Seq in adipose, islet, liver, and skeletal muscle in the DO cohort (Supp. Fig.
106 S2). We performed eQTL analysis using R/qltl2³² (Methods) and identified both local and distal eQTLs for
107 transcripts in each of the four tissues (Supp. Fig. S2B-E). Significant local eQTLs far outnumbered distal
108 eQTLs (Supp. Fig. S2F) and tended to be shared across tissues (Supp. Fig. S2G) whereas the few significant
109 distal eQTLs we identified tended to be tissue-specific (Supp. Fig. S2H)

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

114 To assess the importance of genetic regulation of transcript levels to clinical traits, we compared the local
115 and distal heritabilities of transcripts to their trait relevance, defined as the maximum trait correlation for
116 each transcript. The local heritability of transcripts was negatively correlated with their trait relevance (Fig.
117 2B), suggesting that the more local genotype influenced transcript abundance, the less effect this variation
118 had on the measured traits. Conversely, the distal heritability of transcripts was positively correlated with

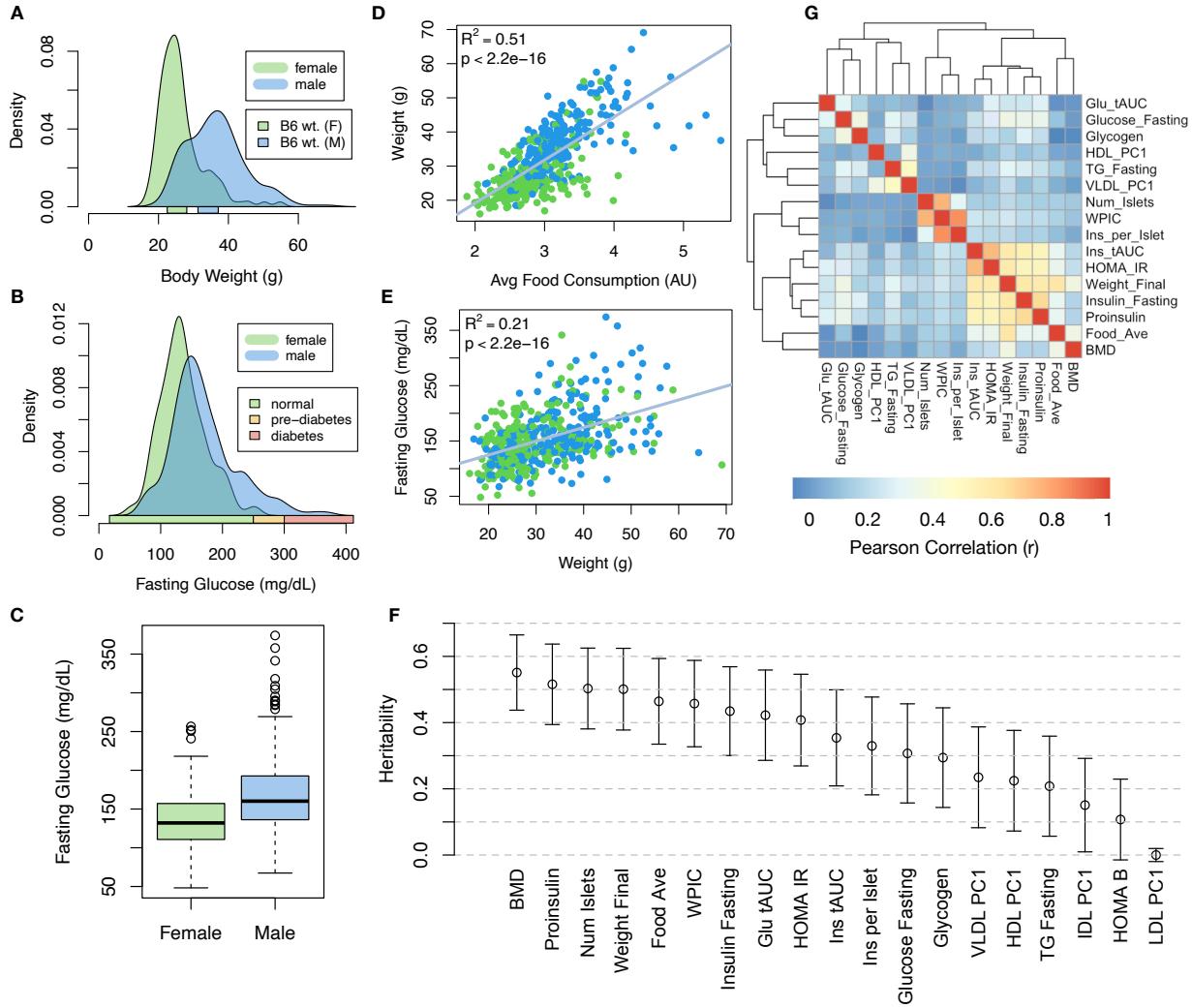


Figure 1: Clinical overview. **A.** Distributions of body weight in the diversity outbred mice. Sex is indicated by color. The average B6 male and female adult weights at 24 weeks of age are indicated by blue and green bars on the x-axis. **B.** The distribution of fasting glucose across the population 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. BMD - bone mineral density, WPIC - whole pancreas insulin content, Glu tAUC - glucose total area under the curve, HOMA IR - homeostatic measurement of insulin resistance, HOMA B - homeostatic measure of beta cell health, VLDL - very low-density lipoprotein, LDL - low-density lipoprotein, IDL - intermediate density lipoprotein, HDL - high-density lipoprotein, TG - triglyceride.

119 trait relevance (Fig. 2C). That is, transcripts that were more highly correlated with the measured traits
 120 tended to be distally, rather than locally, heritable. Importantly, this pattern was consistent across all tissues.
 121 This finding is consistent with previous observations that transcripts with low local heritability explain more
 122 expression-mediated disease heritability than transcripts with high local heritability¹⁹. However, the positive

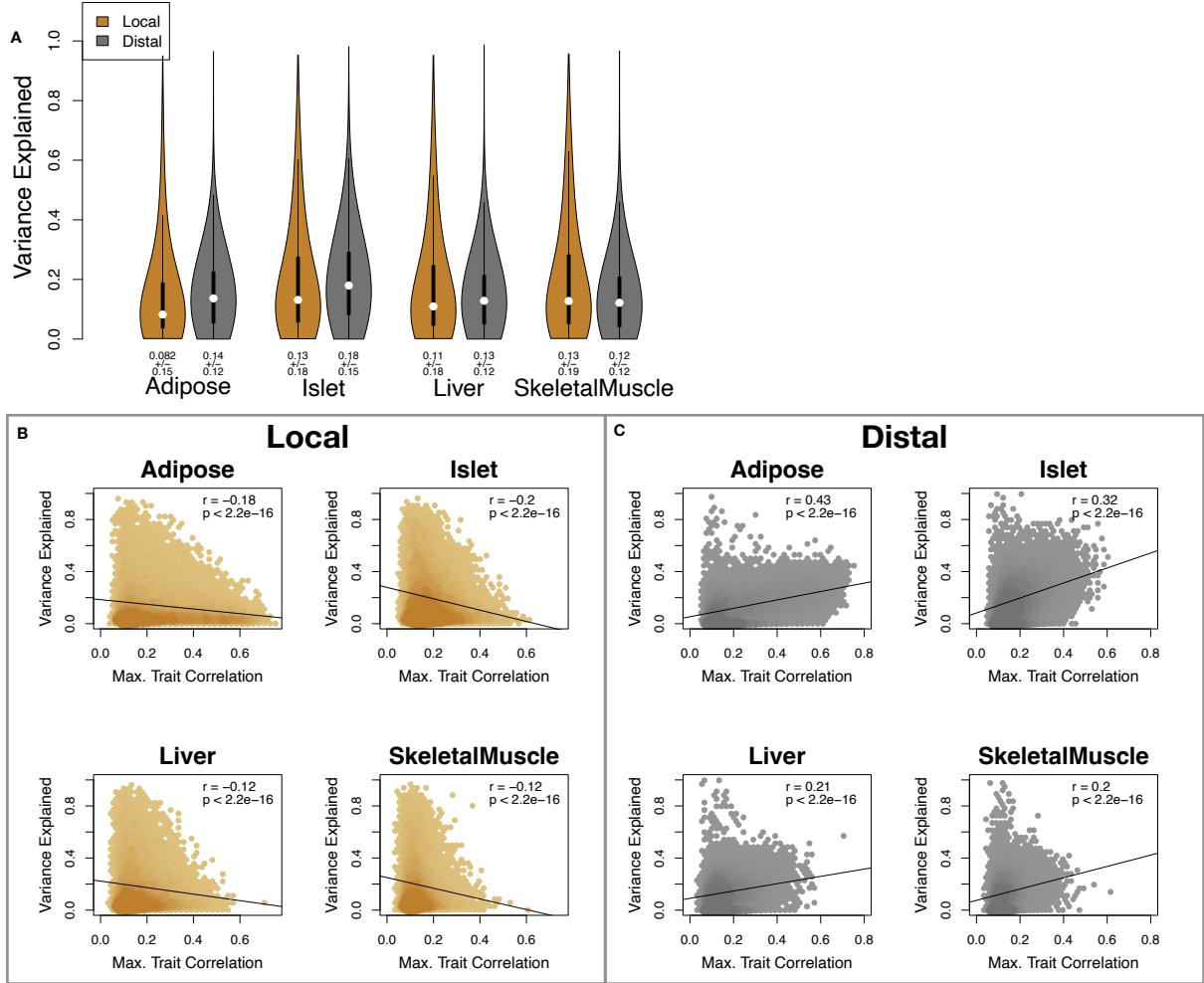


Figure 2: Transcript heritability and trait relevance. **A.** Distributions of distal and local heritability of transcripts across the four tissues. Overall local and distal factors contribute equally to transcript heritability. The relationship between **(B.)** local and **(C.)** distal heritability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

123 relationship between trait correlation and distal heritability demonstrated further that there are diffuse
124 genetic effects throughout the genome converging on trait-related transcripts.

125 **High-Dimensional Mediation Analysis identified a high-heritability composite trait that was
126 mediated by a composite transcript**

127 The above univariate analyses establish the importance of distal heritability for trait-relevant transcripts.
128 However, the number of transcripts dramatically exceeds the number of phenotypes. Thus, we expect the
129 heritable, trait-relevant transcripts to be highly correlated and organized according to coherent, biological

processes representing the mediating endophenotypes driving clinical trait variation. To identify these endophenotypes in a theoretically principled way, we developed a novel dimension-reduction technique, high-dimension mediation analysis (HDMA), that uses the theory of causal graphical models to identify a transcriptomic signature that is simultaneously 1) highly heritable, 2) strongly correlated to the measured phenotypes, and 3) conforms to the causal mediation hypothesis (Fig. 3). HDMA projects the high-dimensional genome, transcriptome, and phenotype data onto one-dimensional scores—a composite genome score (G_C), a composite transcriptome score (T_C), and a composite phenotype score (P_C)—and uses the univariate theory of mediation to constrain these projections to satisfy the hypotheses of perfect mediation, namely that upon controlling for the transcriptomic score, the genome score is uncorrelated to the phenotype score. A complete mathematical derivation and implementation details for HDMA are available in Supp. Methods.

Using HDMA we identified the major axis of variation in the transcriptome that was consistent with mediating the effects of the genome on metabolic traits (Fig 3). Fig. 3A shows the partial correlations (ρ) between the pairs of these composite vectors. The partial correlation between G_C and T_C was 0.42, and the partial correlation between T_C and P_C was 0.78. However, when the transcriptome was taken into account, the partial correlation between G_C and P_C was effectively zero (0.039). P_C captured 30% of the overall trait variance, and its estimated heritability was 0.71 ± 0.084 , which was higher than any of the measured traits (Fig. 1F). Thus, HDMA identified a maximally heritable metabolic composite trait and a highly heritable component of the transcriptome that are correlated as expected in the perfect mediation model.

As discussed in Supp. Methods, HDMA is related to a generalized form of CCA. Standard CCA is prone to over-fitting because in any two large matrices it can be trivial to identify highly correlated composite vectors³³. To assess whether our implementation of HDMA was similarly prone to over-fitting in a high-dimensional space, we performed permutation testing. We permuted the individual labels on the transcriptome matrix 10,000 times and recalculated the path coefficient, which is the correlation of G_C and T_C multiplied by the correlation of T_C and P_C . This represents the strength of the path from G_C to P_C that is putatively mediated through T_C . The null distribution of the path coefficient is shown in Fig. 3B, and the observed path coefficient from the original data is indicated by a red line. The observed path coefficient was well outside the null distribution generated by permutations ($p < 10^{-16}$). Fig. 3C illustrates this observation in more detail. Although we identified high correlations between G_C and T_C , and modest correlations between T_C and P_C in the null data (Fig 3C), these two values could not be maximized simultaneously in the null data. In contrast, the red dot shows that in the real data both the G_C - T_C correlation and the T_C - P_C correlation could be maximized simultaneously suggesting that the path from genotype to phenotype through transcriptome is highly non-trivial and identifiable in this case. These results suggest that these composite vectors represent

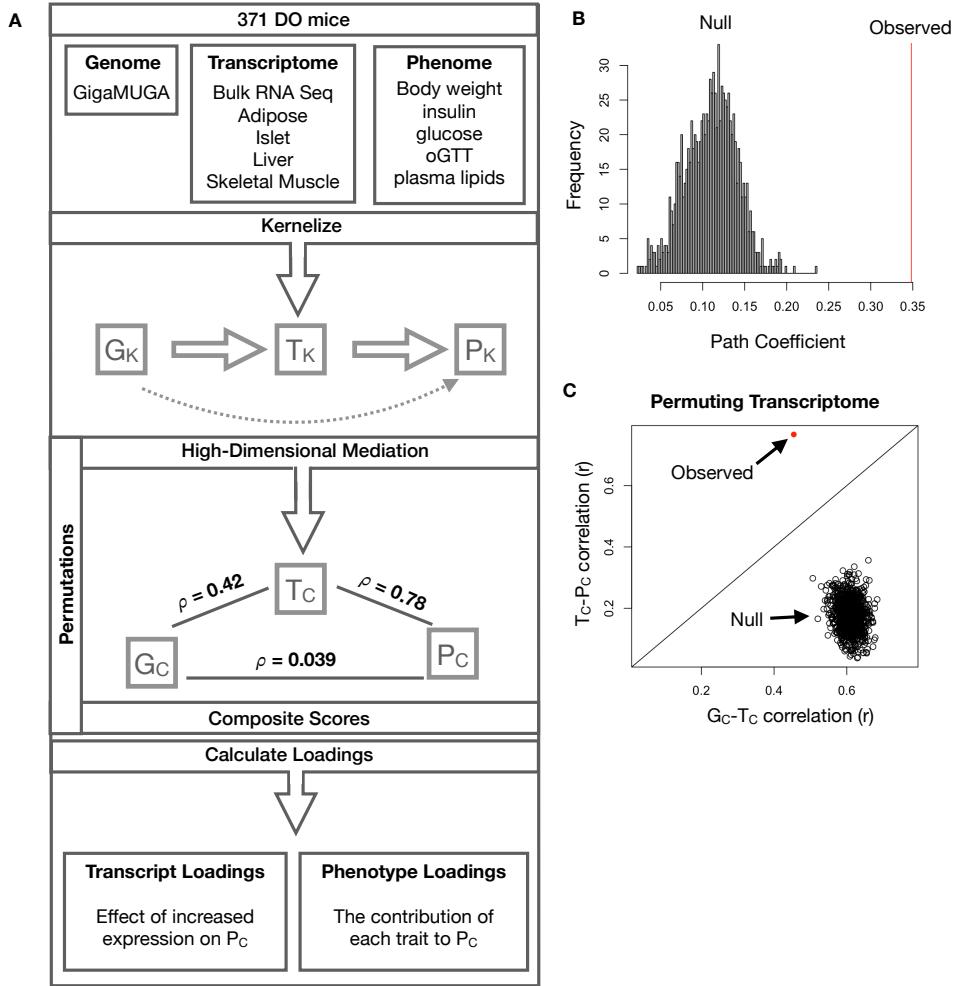


Figure 3: High-dimensional mediation. **A.** Workflow indicating major steps of high-dimensional mediation. The genotype, transcriptome, and phenotype matrices were independently normalized and converted to kernel matrices representing the pairwise relationships between individuals for each data modality (K_G = genome kernel, K_T = transcriptome kernel; K_P = 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).

genetically determined variation in phenotype that is mediated through genetically determined variation in transcription.

164 Body weight and insulin resistance were highly represented in the expression-mediated com-
165 posite trait

166 Each composite score is a weighted combination of the measured variables. The magnitude and sign of the
167 weights, called loadings, correspond to the relative importance and directionality of each variable in the
168 composite score. The loadings of each measured trait onto P_C indicate how much each contributed to the
169 composite phenotype. Body weight contributed the most (Fig. 4), followed by homeostatic insulin resistance
170 (HOMA_IR) and fasting plasma insulin levels (Insulin_Fasting). We can thus interpret P_C as an index
171 of metabolic disease (Fig. 4B). Individuals with high values of P_C have a higher metabolic disease index
172 (MDI) and greater metabolic disease, including higher body weight and higher insulin resistance. We refer to
173 P_C as the MDI going forward. Traits contributing the least to the MDI were measures of cholesterol and
174 pancreas composition. Thus, when we interpret the transcriptomic signature identified by HDMA, we are
175 explaining primarily the putative transcriptional mediation of body weight and insulin resistance, as opposed
176 to cholesterol measurements.

177 High-loading transcripts had low local heritability, high distal heritability, and were linked
178 mechanistically to obesity

179 We interpreted large loadings onto transcripts as indicating strong mediation of the effect of genetics on the
180 MDI. Large positive loadings indicate that higher expression was associated with a higher MDI (i.e. higher
181 risk of obesity and metabolic disease on the HFHS diet) (Fig. 4C). Conversely, large negative loadings
182 indicate that high expression of these transcripts was associated with a lower MDI (i.e. lower risk of obesity
183 and metabolic disease on the HFHS diet) (Fig. 4C). We used gene set enrichment analysis (GSEA)^{34;35} to
184 look for biological processes and pathways that were enriched at the top and bottom of this list (Methods).

185 In adipose tissue, both GO processes and KEGG pathway enrichments pointed to an axis of inflammation and
186 metabolism (Figs. S3 and S4). GO terms and KEGG pathways associated with inflammation were positively
187 associated with the MDI, indicating that increased expression in inflammatory pathways was associated
188 with a higher burden of disease. It is well established that adipose tissue in obese individuals is inflamed
189 and infiltrated by macrophages³⁶⁻⁴⁰, and the results here suggest that this may be a dominant heritable
190 component of metabolic disease.

191 The strongest negative enrichments in adipose tissue were related to mitochondrial activity in general, and
192 thermogenesis in particular (Figs. S3 and S3). Genes in the KEGG oxidative phosphorylation pathway were
193 almost universally negatively loaded in adipose tissue, suggesting that increased expression of these genes was
194 associated with reduced MDI (Supp. Fig. S5). Consistent with this observation, it has been shown previously

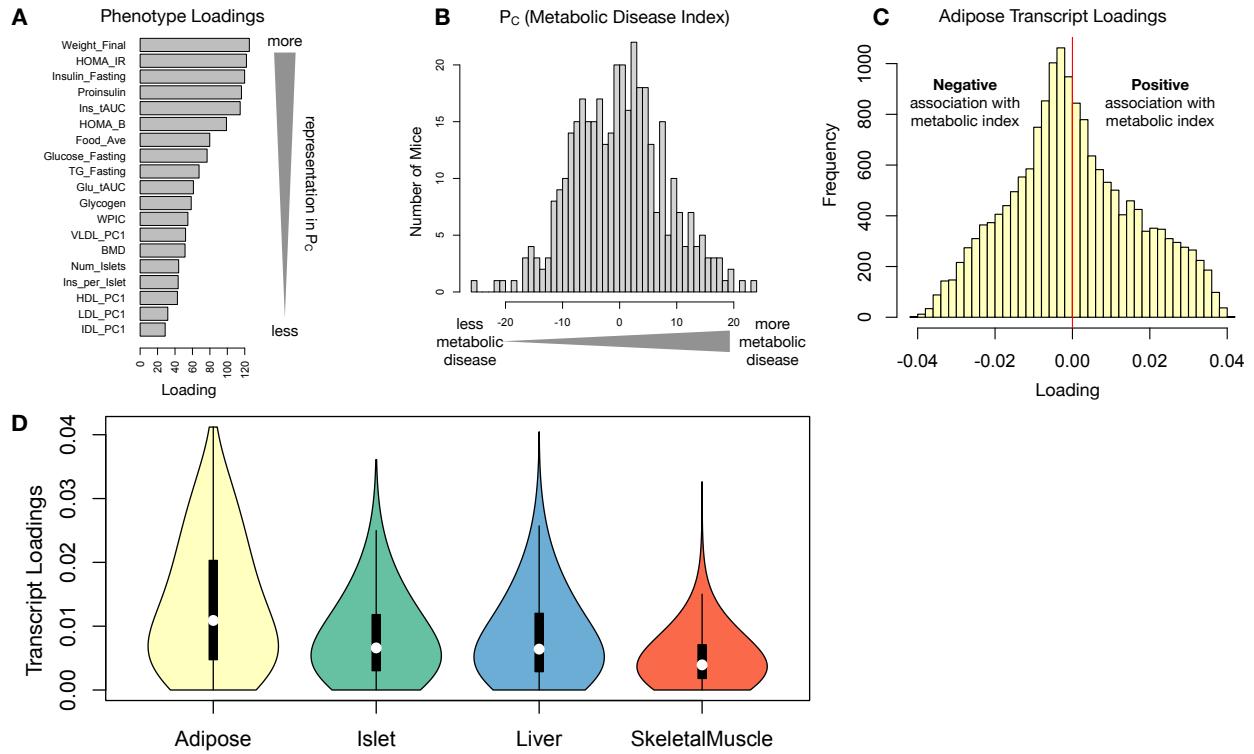


Figure 4: Interpretation of loadings. **A.** Loadings across traits. Body weight and insulin resistance contributed the most to the composite trait. **B.** Phenotype scores across individuals. Individuals with large positive phenotype scores had higher body weight and insulin resistance than average. 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 adipose tissue gene expression was a strong mediator of genotype on body weight and insulin resistance.

195 that mouse strains with greater thermogenic potential are also less susceptible to obesity on an obesigenic
196 diet⁴¹.

197 Transcripts associated with the citric acid cycle as well as the catabolism of the branched-chain amino acids
198 (valine, leucine, and isoleucine) were strongly enriched with negative loadings in adipose tissue (Supp. Figs.
199 S3, S6 and S7). Expression of genes in both pathways (for which there is some overlap) has been previously
200 associated with insulin sensitivity^{12;42;43}, suggesting that heritable variation in regulation of these pathways
201 may influence risk of insulin resistance.

202 Looking at the 10 most positively and negatively loaded transcripts from each tissue, it is apparent that
203 transcripts in the adipose tissue had the largest loadings, both positive and negative (Fig. 5A bar plot). This
204 suggests that much of the effect of genetics on body weight and insulin resistance is mediated through gene
205 expression in adipose tissue. The strongest loadings in liver and pancreas were comparable, and those in

206 skeletal muscle were the weakest (Fig. 5A), suggesting that less of the genetic effects were mediated through
207 transcription in skeletal muscle. As expected, heritability analysis showed that transcripts with the largest
208 loadings had higher distal heritability than local heritability (Fig. 5A heat map and box plot). This pattern
209 contrasts with transcripts nominated by TWAS (Fig. 5B), which tended to have lower loadings, higher local
210 heritability and lower distal heritability. Transcripts with the highest local heritability in each tissue (Fig.
211 5C) had the lowest loadings, consistent with our findings above (Fig. 2B).

212 We performed a literature search for the genes in each of these groups along with the terms “diabetes”,
213 “obesity”, and the name of the expressing tissue to determine whether any of these genes had previous
214 associations with metabolic disease in the literature (Methods). Multiple genes in each group had been
215 previously associated with obesity and diabetes (Fig. 5 bolded gene names). Genes with high loadings were
216 most highly enriched for previous literature support. They were 2.4 times more likely than TWAS hits and 3.8
217 times more likely than genes with high local heritability to be previously associated with obesity or diabetes.

218 **Tissue-specific transcriptional programs were associated with metabolic traits**

219 Clustering of transcripts with top loadings in each tissue showed tissue-specific functional modules associated
220 with obesity and insulin resistance (Fig. 6A) (Methods). The clustering highlights the importance of immune
221 activation particularly in adipose tissue. The “mitosis” cluster had large positive loadings in three of the four
222 tissues potentially suggesting system-wide proliferation of immune cells. Otherwise, all clusters were strongly
223 loaded in only one or two tissues. For example, the lipid metabolism cluster was loaded most heavily in liver.
224 The positive loadings suggest that high expression of these genes, particularly in the liver, was associated with
225 increased metabolic disease. This cluster included the gene *Pparg*, whose primary role is in the adipose tissue
226 where it is considered a master regulator of adipogenesis⁴⁴. Agonists of *Pparg*, such as thiazolidinediones, are
227 FDA-approved to treat type II diabetes, and reduce inflammation and adipose hypertrophy⁴⁴. Consistent
228 with this role, the loading for *Pparg* in adipose tissue was negative, suggesting that higher expression was
229 associated with leaner mice (Fig. 6B). In contrast, *Pparg* had a large positive loading in liver, where it is
230 known to play a role in the development of hepatic steatosis, or fatty liver. Mice that lack *Pparg* specifically
231 in the liver, are protected from developing steatosis and show reduced expression of lipogenic genes^{45;46}.
232 Overexpression of *Pparg* in the livers of mice with a *Ppara* knockout, causes upregulation of genes involved in
233 adipogenesis⁴⁷. In the livers of both mice and humans high *Pparg* expression is associated with hepatocytes
234 that accumulate large lipid droplets and have gene expression profiles similar to that of adipocytes^{48;49}.
235 The local and distal heritability of *Pparg* is low in adipose tissue suggesting its expression in this tissue is
236 highly constrained in the population (Fig. 6B). However, the distal heritability of *Pparg* in liver is relatively

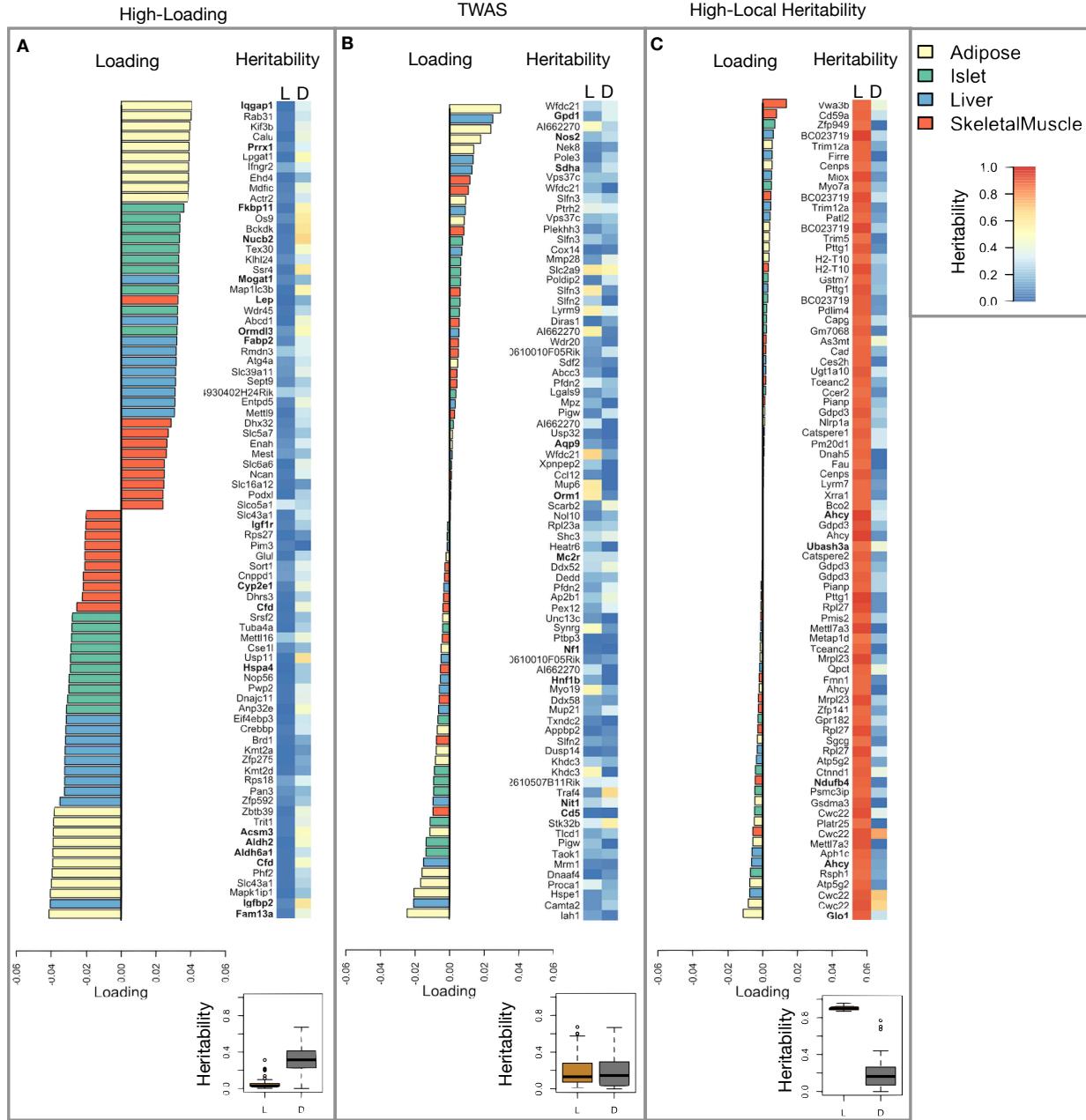


Figure 5: Transcripts with high loadings have high distal heritability and literature support. Each panel has a bar plot showing the loadings of transcripts selected by different criteria. Bar color indicates the tissue of origin. The heat map shows the local (L - left) and distal (D - right) heritability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

237 high suggesting it is complexly regulated and has sufficient variation in this population to drive variation in
 238 phenotype. Both local and distal heritability of *Pparg* in the islet are relatively high, but the loading is low,
 239 suggesting that variability of expression in the islet does not drive variation in MDI. These results highlight

240 the importance of tissue context when investigating the role of heritable transcript variability in driving
 241 phenotype.

242 Gene lists for all clusters are available in Supp. File 1.

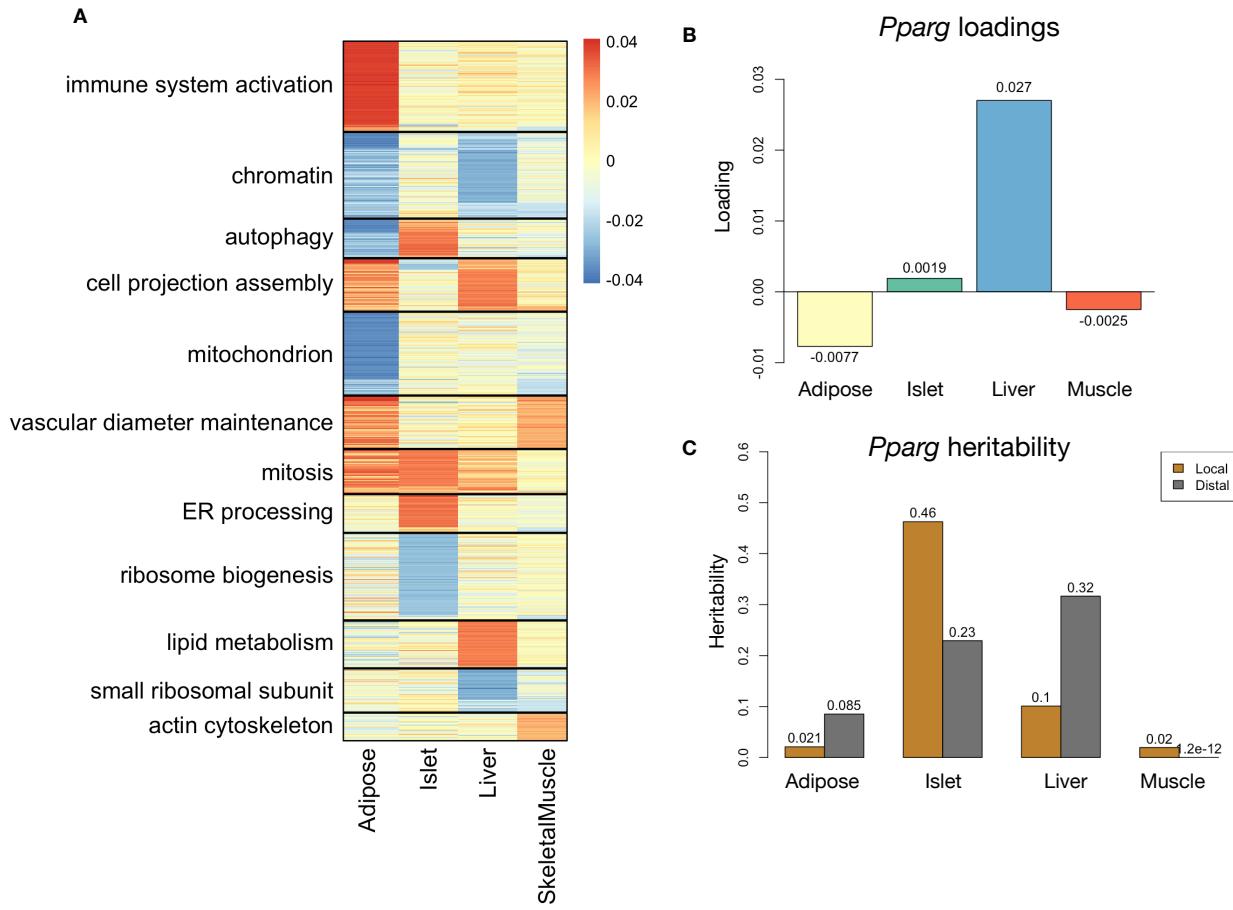


Figure 6: Tissue-specific transcriptional programs were associated with obesity and insulin resistance. **A** Heat map showing the loadings of all transcripts with loadings greater than 2.5 standard deviations from the mean in any tissue. The heat map was clustered using k medoid clustering. Functional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

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

244 To test whether the transcript loadings identified in the DO could be translated to another population, we
 245 tested whether they could predict metabolic phenotype in an independent population of CC-RIX mice, which
 246 were F1 mice derived from multiple pairings of Collaborative Cross (CC)^{50;30;51;52} strains (Fig. 7) (Methods).
 247 We tested two questions. First, we asked whether the loadings identified in the DO mice were relevant to
 248 the relationship between the transcriptome and the phenotype in the CC-RIX. We predicted body weight
 249 (a surrogate for MDI) 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 acutal body weight were highly correlated (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 MDI. This result also confirms the validity and translatability of the transcript loadings and their relationship to metabolic disease.

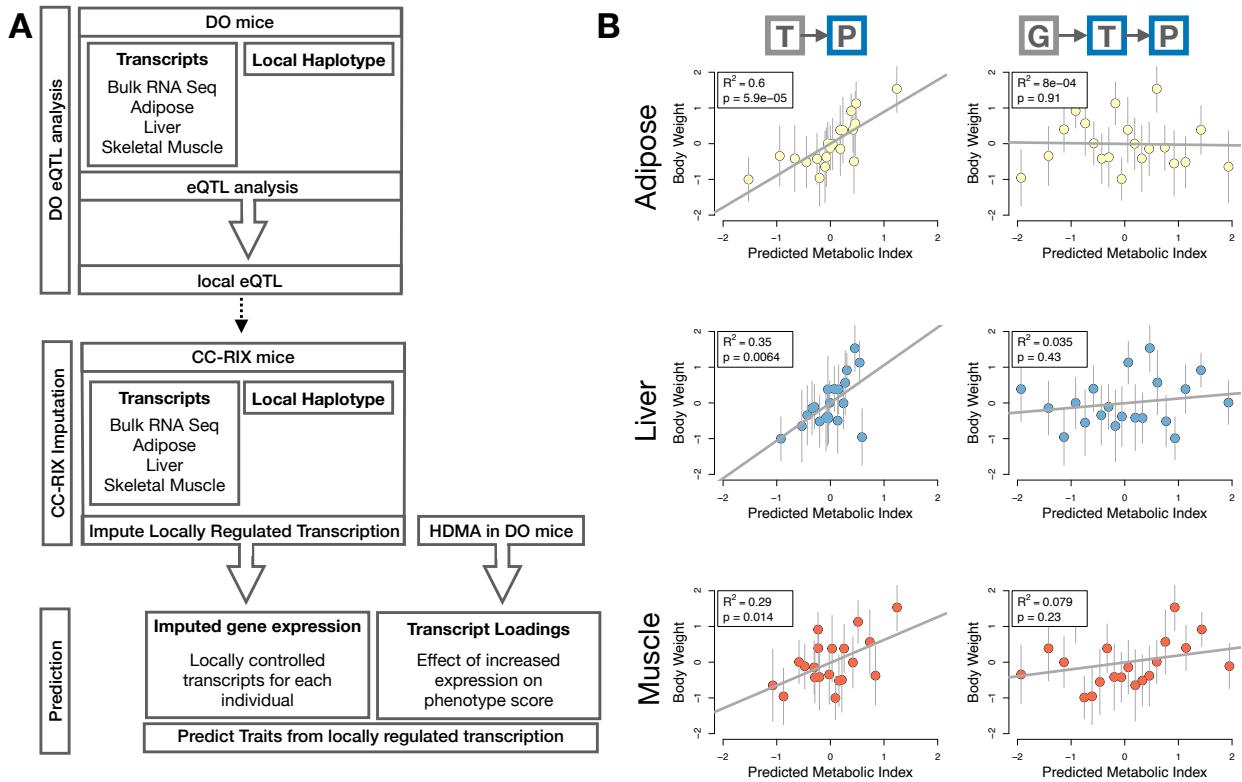


Figure 7: Transcription, but not local genotype, predicts phenotype in the CC-RIX. **A.** Workflow showing procedure for translating HDMA results to an independent population of mice. **B.** Relationships between the predicted metabolic disease index (MDI) and measured body weight. The left column shows 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 trait-relevant gene expression, we should be able to predict phenotype in the CC-RIX using transcripts imputed from local genotype (Fig. 7A). The DO and the CC-RIX were derived from the same eight founder strains and so carry the same alleles throughout the genome. We imputed gene expression in the CC-RIX using local genotype and were able to estimate variation in gene transcription robustly (Supp. Fig. S8). However, these imputed values failed to predict body weight in the CC-RIX when weighted with the loadings from HDMA. (Fig. 7B right column). This result suggests that local regulation of

262 gene expression is not the primary factor driving heritability of complex traits. It is also consistent with our
263 findings in the DO population that distal heritability was a major driver of trait-relevant gene expression and
264 that high-loading transcripts had comparatively high distal and low local heritability.

265 **Distally heritable transcriptomic signatures reflected variation in composition of adipose tissue**
266 **and islets**

267 The interpretation of global genetic influences on gene expression and phenotype is potentially more challenging
268 than the interpretation and translation of local genetic influences, as genetic effects cannot be localized to
269 individual gene variants or transcripts. However, there are global patterns across the loadings that can
270 inform mechanism. For example, heritable variation in cell type composition can be inferred from transcript
271 loadings. We observed above that immune activation in the adipose tissue was a highly enriched process
272 correlating with obesity in the DO population. In humans, it has been extensively observed that macrophage
273 infiltration in adipose tissue is a marker of obesity and metabolic disease⁵³. To determine whether the
274 immune activation reflected a heritable change in cell composition in adipose tissue in DO mice, we compared
275 loadings of cell-type specific genes in adipose tissue (Methods). The mean loading of macrophage-specific
276 genes was significantly greater than 0 (Fig. 8A), indicating that obese mice were genetically predisposed
277 to have high levels of macrophage infiltration in adipose tissue in response to the HFHS diet. Loadings for
278 marker genes for other cell types were not statistically different from zero, indicating that changes in the
279 abundance of those cell types is not a mediator of MDI.

280 We also compared loadings of cell-type specific transcripts in islet (Methods). The mean loadings for alpha-cell
281 specific transcripts were significantly greater than 0, while the mean loadings for delta- and endothelial-cell
282 specific genes were significantly less than 0 (Fig. 8B). These results suggest that mice with higher MDI
283 inherited an altered cell composition that predisposed them to metabolic disease, or that these compositional
284 changes were induced by the HFHS diet in a heritable way. In either case, these results support the hypothesis
285 that alterations in islet composition drive variation in MDI. Notably, the mean loading for pancreatic beta cell
286 marker transcripts was not significantly different from zero. We stress that this is not necessarily reflective of
287 the function of the beta cells in the obese mice, but rather suggests that any variation in the number of beta
288 cells in these mice was unrelated to obesity and insulin resistance, the major contributors to MDI. This is
289 further consistent with the islet composition traits having small loadings in the phenome score (Fig. 4).

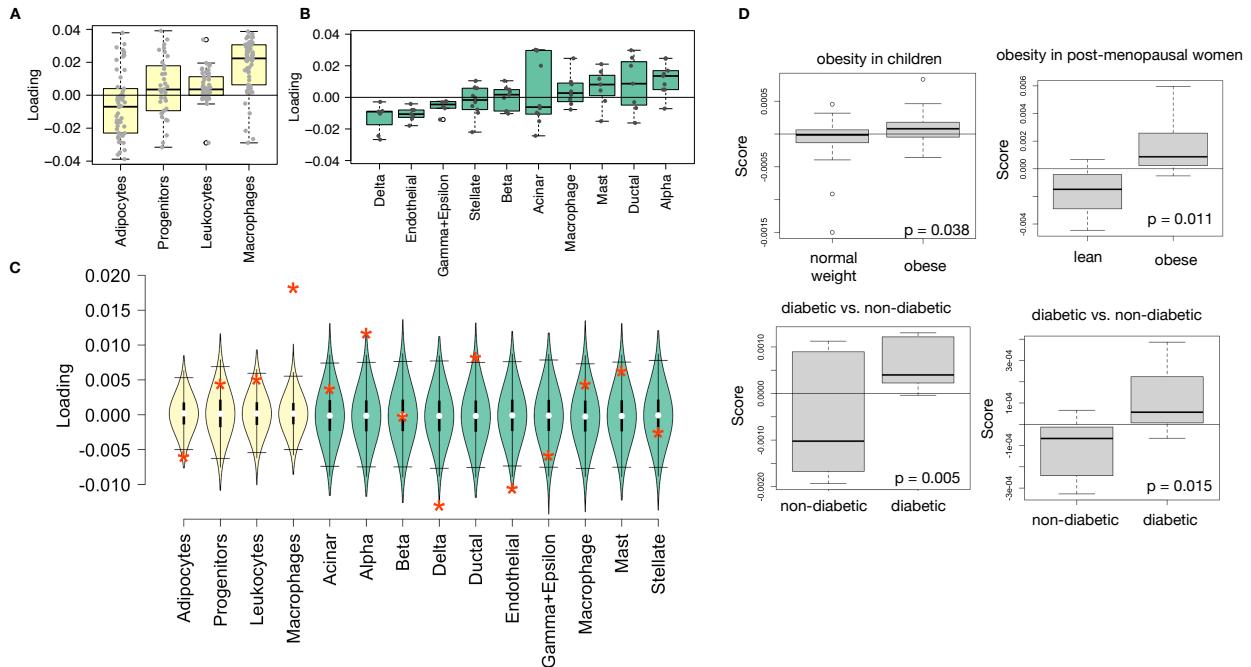


Figure 8: HDMA results translate to humans. **A.** Distribution of loadings for cell-type-specific transcripts in adipose tissue. **B.** Distribution of loadings for cell-type-specific transcripts in pancreatic islets (green). **C.** Null distributions for the mean loading of randomly selected transcripts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

290 Heritable transcriptomic signatures translated to human disease

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

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

302 **Existing therapies are predicted to target mediator gene signatures**

303 Another potential application of the transcript loading landscape is in ranking potential drug candidates
304 for the treatment of metabolic disease. Although high-loading transcripts may be good candidates for
305 understanding specific biology related to obesity, the transcriptome overall is highly interconnected and
306 redundant. The ConnectivityMap (CMAP) database⁵⁴ developed by the Broad Institute allows querying
307 thousands of compounds that reverse or enhance the extreme ends of transcriptomic signatures in multiple
308 different cell types. By identifying drugs that reverse pathogenic transcriptomic signatures, we can potentially
309 identify compounds that have favorable effects on gene expression. To test this hypothesis, we queried the
310 CMAP database through the CLUE online query tool (<https://clue.io/query/>, version 1.1.1.43) (Methods).
311 We identified top anti-correlated hits across all cell types (Supp. Figs S9 and S10). To get more tissue-specific
312 results, we also looked at top results in cell types that most closely resembled our tissues. We looked at
313 results in adipocytes (ASC) as well as pancreatic tumor cells (YAPC) regardless of *p* value (Supp. Figs S11
314 and S12).

315 Looking across all cell types, the notable top hits from the adipose tissue loadings included mTOR inhibitors
316 and glucocorticoid agonists (Supp. Fig. S9). It is thought that metformin, which is commonly used to
317 improve glycemic control, acts, at least in part, by inhibiting mTOR signaling^{55;56}. However, long-term use
318 of other mTOR inhibitors, such as rapamycin, are known to cause insulin resistance and β -cell toxicity^{56–58}.
319 Glucocorticoids are used to reduce inflammation, which was a prominent signature in the adipose tissue,
320 but these drugs also promote hyperglycemia and diabetes^{59;60}. Accute treatment with glucocorticoids has
321 further been shown to reduce thermogenesis in rodent adipocytes^{61–63}, but increase thermogenesis in human
322 adipocytes^{64;65}. Thus, the pathways identified by CMAP across all cell types were highly related to the
323 transcript loading profiles, but the relationship was not a simple reversal.

324 The top hit for the adipose composite transcript in CMAP adipocytes was a PARP inhibitor (Supp. Fig.
325 S11). PARPs play a role in lipid metabolism and are involved in the development of obesity and diabetes⁶⁶.
326 PARP1 inhibition increases mitochondrial biogenesis⁶⁷. Inhibition of PARP1 activity can further prevent
327 necrosis in favor of the less inflammatory apoptosis⁶⁸, thereby potentially reducing inflammation in stressed
328 adipocytes. Other notable hits among the top 20 were BTK inhibitors, which have been observed to suppress
329 inflammation and improve insulin resistance⁶⁹ as well as to reduce insulin antibodies in type I diabetes⁷⁰.
330 IkappaB kinase (IKK) is an enzyme complex involved in regulating cellular responses to inflammation⁷¹.
331 Inhibitors of IKK have been shown to improve glucose control in type II diabetes^{72;73}.

332 Among the top most significant hits for the transcript loadings from pancreatic islets (Supp. Fig. S10),

333 was suppression of T cell receptor signaling, which is known to be involved in Type 1 diabetes⁷⁴, as well as
334 TNFR1, which has been associated with mortality in diabetes patients⁷⁵. Suppression of NOD1/2 signaling
335 was also among the top hits. NOD1 and 2 sense ER stress^{76;77}, which is associated with β -cell death in type
336 1 and type 2 diabetes⁷⁸. This cell death process is dependent on NOD1/2 signaling⁷⁶, although the specifics
337 have not yet been worked out.

338 We also looked specifically at hits in pancreatic tumor cells (YAPC) regardless of significance level to get a
339 transcriptional response more specific to the pancreas (Supp. Fig. S12). Hits in this list included widely used
340 diabetes drugs, such as sulfonylureas, PPAR receptor agonists, and insulin sensitizers. Rosiglitazone is a
341 PPAR- γ agonist and was one of the most prescribed drugs for type 2 diabetes before its use was reduced due
342 to cardiac side-effects⁷⁹. Sulfonylureas are another commonly prescribed drug class for type 2 diabetes, but
343 also have notable side effects including hypoglycemia and accelerated β -cell death⁸⁰.

344 In summary, the high-loading transcripts derived from HDMA in mice prioritized of drugs with demonstrated
345 effectiveness in reducing type 2 diabetes phenotypes in humans in a tissue-specific manner. Drugs identified
346 using the islet loadings are known diabetes drugs that act directly on pancreatic function. Drugs identified
347 by the adipose loadings tended to reduce inflammatory responses and have been shown incidentally to reduce
348 obesity-related morbidity.

349 Discussion

350 Here we investigated the relative contributions of local and distal gene regulation in four tissues to heritable
351 variation in traits related to metabolic disease in genetically diverse mice. We found that distal heritability
352 was positively correlated with trait relatedness, whereas high heritability was negatively correlated with
353 trait relatedness. We used a novel high-dimensional mediation analysis (HDMA) to identify tissue-specific
354 composite transcripts that are predicted to mediate the effect of genetic background on metabolic traits. The
355 adipose-derived composite transcript robustly predicted body weight in an independent cohort of diverse
356 mice with disparate population structure, as well as to humans. However, gene expression imputed from
357 local genotype failed to predict body weight in the second population. Taken together, these results highlight
358 the complexity of gene expression regulation in relation to trait heritability and suggest that heritable trait
359 variation is mediated primarily through distal gene regulation.

360 Supplemental Discussion

361 Our result that distal regulation accounted for most trait-related gene expression differences is consistent
362 with a complex model of genetic trait determination. It has frequently been assumed that gene regulation in

363 *cis* is the primary driver of genetically associated trait variation, but attempts to use local gene regulation
364 to explain phenotypic variation have had limited success^{16;17}. In recent years, evidence has mounted that
365 distal gene regulation may be an important mediator of trait heritability^{19;18;81}. It has been observed that
366 transcripts with high local heritability explain less expression-mediated disease heritability than those with
367 low local heritability¹⁹. Consistent with this observation, genes located near GWAS hits tend to be complexly
368 regulated¹⁸. They also tend to be enriched with functional annotations, in contrast to genes with simple
369 local regulation, which tend to be depleted of functional annotations suggesting they are less likely to be
370 directly involved in disease traits¹⁸. These observations are consistent with principles of robustness in complex
371 systems in which simple regulation of important elements leads to fragility of the system⁸²⁻⁸⁴. Our results
372 are consistent, instead, with a more complex picture where genes whose expression can drive trait variation
373 are buffered from local genetic variation but are extensively influenced indirectly by genetic variation in the
374 regulatory networks converging on those genes.

375 Our results are also consistent with the recently proposed omnigenic model, which posits that complex traits
376 are massively polygenic and that their heritability is spread out across the genome⁸⁵. In the omnigenic model,
377 genes are classified either as “core genes,” which directly impinge on the trait, or “peripheral genes,” which
378 are not directly trait-related, but influence core genes through the complex gene regulatory network. HDMA
379 explicitly models a central proposal of the omnigenic model which posits that once the expression of the
380 core genes (i.e. trait-mediating genes) is accounted for, there should be no residual correlation between the
381 genome and the phenotype. Here, we were able to fit this model and identified a composite transcript that,
382 when taken into account, left no residual correlation between the composite genome and composite phenotype
383 scores (Fig. 3A).

384 Unlike in the omnigenic model, we did not observe a clear demarcation between the core and peripheral
385 genes in loading magnitude, but we do not necessarily expect a clear separation given the complexity of gene
386 regulation and the genotype-phenotype map⁸⁶.

387 An extension of the omnigenic model proposed that most heritability of complex traits is driven by weak
388 distal eQTLs that are potentially below the detection threshold in studies with feasible sample sizes⁸¹. This
389 is consistent with what we observed here. For example, *Nucb2*, had a high loading in islets and was also
390 strongly distally regulated (66% distal heritability) (Fig. 5). This gene is expressed in pancreatic β cells and
391 is involved in insulin and glucagon release⁸⁷⁻⁸⁹. Although its transcription was highly heritable in islets, that
392 regulation was distributed across the genome, with no clear distal eQTL (Supp. Fig. S13). Thus, although
393 distal regulation of some genes may be strong, this regulation is likely to be highly complex and not easily
394 localized.

395 Individual high-loading transcripts also demonstrated biologically interpretable, tissue-specific patterns. We
396 highlighted *Pparg*, which is known to be protective in adipose tissue⁴⁴ where it was negatively loaded, and
397 harmful in the liver^{45–49}, where it was positively loaded. Such granular patterns may be useful in generating
398 hypotheses for further testing, and prioritizing genes as therapeutic targets. The tissue-specific nature of
399 the loadings also may provide clues to tissue-specific effects, or side effects, of targeting particular genes
400 system-wide.

401 In addition to identifying individual transcripts of interest, the composite transcripts can be used as weighted
402 vectors in multiple types of analysis, such as drug prioritization using gene set enrichment analysis (GSEA)
403 and the CMAP database. In particular, the CMAP analysis identified drugs which have been demonstrated
404 to reverse insulin resistance and other aspects of metabolic disease. This finding supports the causal role of
405 these full gene signatures in pathogenesis of metabolic disease and thus their utility in prioritizing drugs and
406 gene targets as therapeutics.

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

413 Data and Code Availability

414 **DO mice:** Genotypes, phenotypes, and pancreatic islet gene expression data were previously published¹².
415 Gene expression for the other tissues can be found at the Gene Expression Omnibus <https://www.ncbi.nlm.nih.gov/geo/> with the following accession numbers: DO adipose tissue - GSE266549; DO liver tissue
416 - GSE266569; DO skeletal muscle - GSE266567. Expression data with calculated eQTLs are available at
417 Figshare <https://figshare.com/> DOI: 10.6084/m9.figshare.27066979

419 **CC-RIX mice:** Gene expression can be found at the Gene Expression Omnibus <https://www.ncbi.nlm.nih.gov/geo/> with the following accession numbers: CC-RIX adipose tissue - GSE237737; CC-RIX liver tissue -
420 GSE237743; CC-RIX skeletal muscle - GSE237747. Count matrices and phenotype data can be found at
421 Figshare <https://figshare.com/> DOI: 10.6084/m9.figshare.27066979

423 **Code:** All code used to run the analyses reported here are available at Figshare: <https://figshare.com/> DOI:
424 10.6084/m9.figshare.27066979

425 **Acknowledgements**

426 This project was supported by The Jackson Laboratory Cube Initiative, as well as grants from the National
427 Institutes of Health (grant numbers.: R01DK101573, R01DK102948, and RC2DK125961) (to A. D. A.) and
428 by the University of Wisconsin-Madison, Department of Biochemistry and Office of the Vice Chancellor for
429 Research and Graduate Education with funding from the Wisconsin Alumni Research Foundation (to M. P.
430 K.).

431 We thank the following scientific services at The Jackson Laboratory: Genome Technologies for the RNA
432 sequencing, necropsy services for the tissue harvests, and the Center for Biometric Analysis for metabolic
433 phenotyping.

434 **Supplemental Figures**

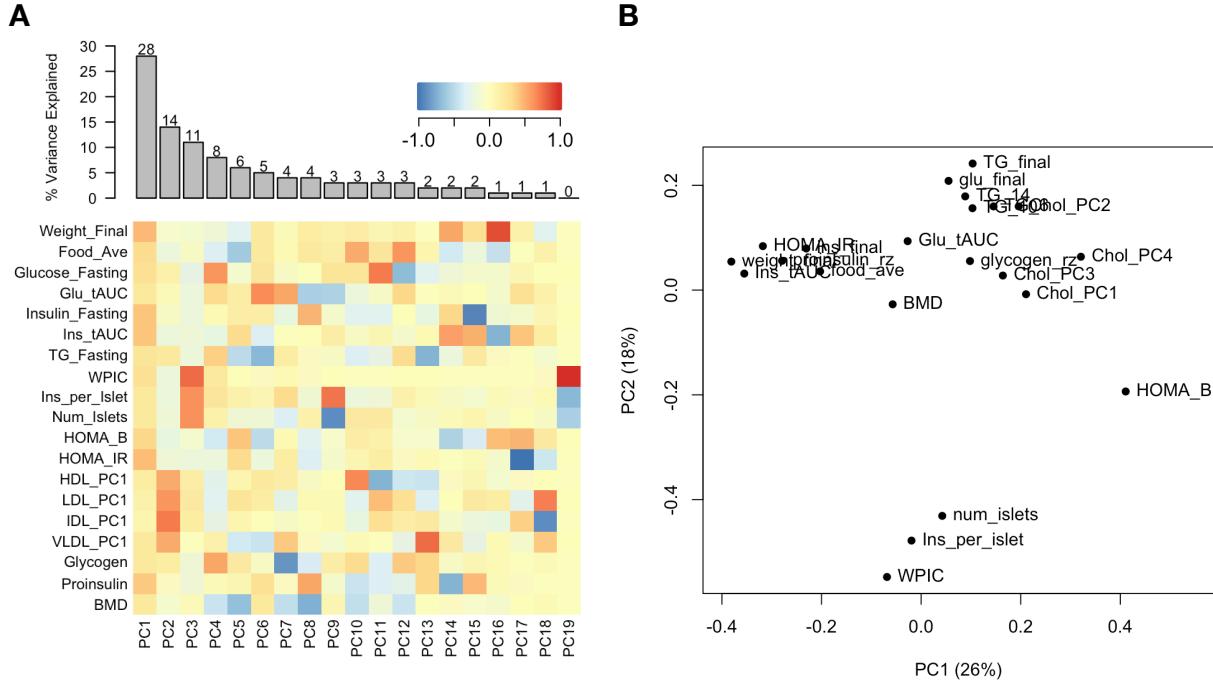


Figure S1: Trait matrix decomposition. **A** The heat map shows the loadings of each trait onto each principal component of the trait matrix. The bars at the top show the percent variance explained for each principal component. **B** Traits plotted by the first and second principal components of the trait matrix. This view shows clustering of traits into insulin- and weight-related traits, lipid-related traits, and ex-vivo pancreatic measurements.

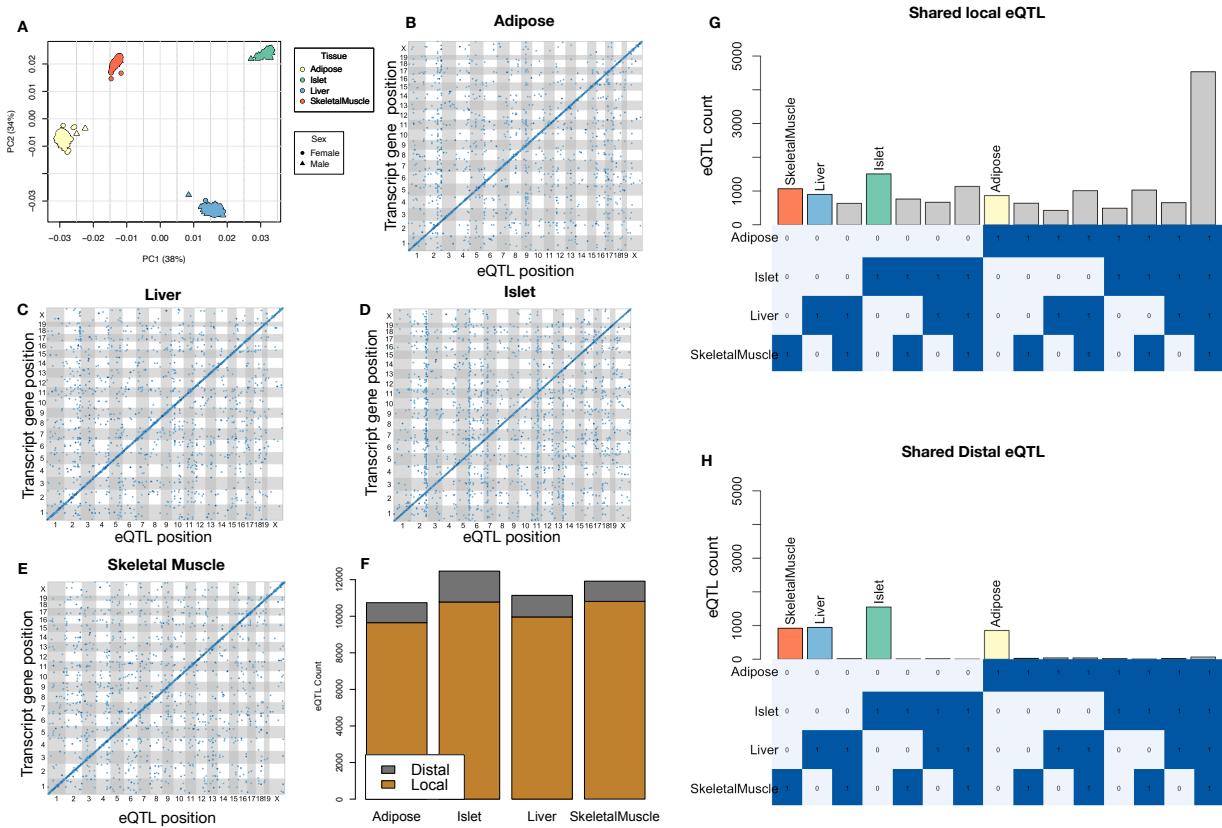


Figure S2: Overview of eQTL analysis in DO mice. **A.** RNA seq samples from the four different tissues clustered by tissue. **B.-E.** eQTL maps are shown for each tissue. The *x*-axis shows the position of the mapped eQTL, and the *y*-axis shows the physical position of the gene encoding each mapped 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 eQTLs are much more numerous than distal eQTL. **G.** Counts of transcripts with local eQTL shared across multiple tissues. The majority of local eQTLs were shared across all four tissues. **H.** Counts of transcripts with distal eQTL shared across multiple tissues. The majority of distal eQTL were tissue-specific and not shared across multiple tissues. For both G and H, eQTL for a given transcript were considered shared in two tissues if they were within 4Mb of each other. Colored bars indicate the counts for individual tissues for easy of visualization.

KEGG pathway enrichments by GSEA

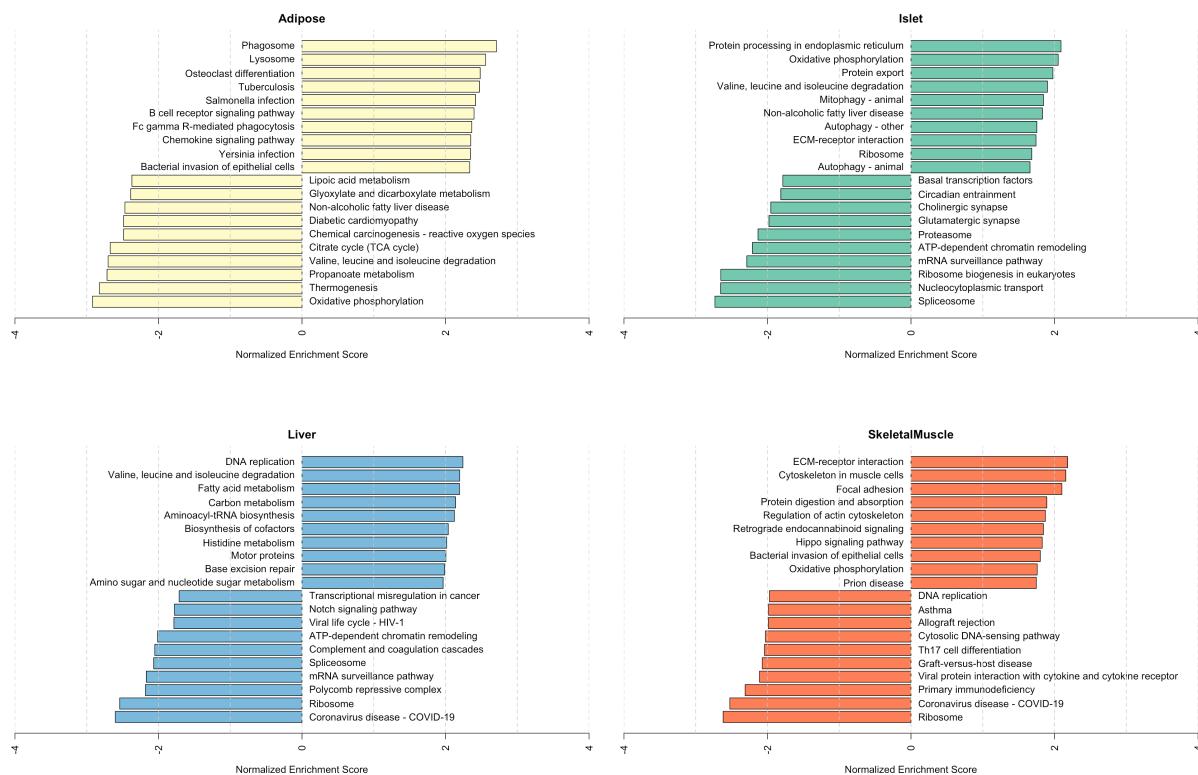


Figure S3: Bar plots showing normalized enrichment scores (NES) for KEGG pathways as determined by fast gene score enrichment analysis (fgsea). Only the top 10 positive and top 10 negative scores are shown. Colors indicate tissue. The name beside each bar shows the name of each enriched KEGG pathway.

Top GO term enrichments by GSEA

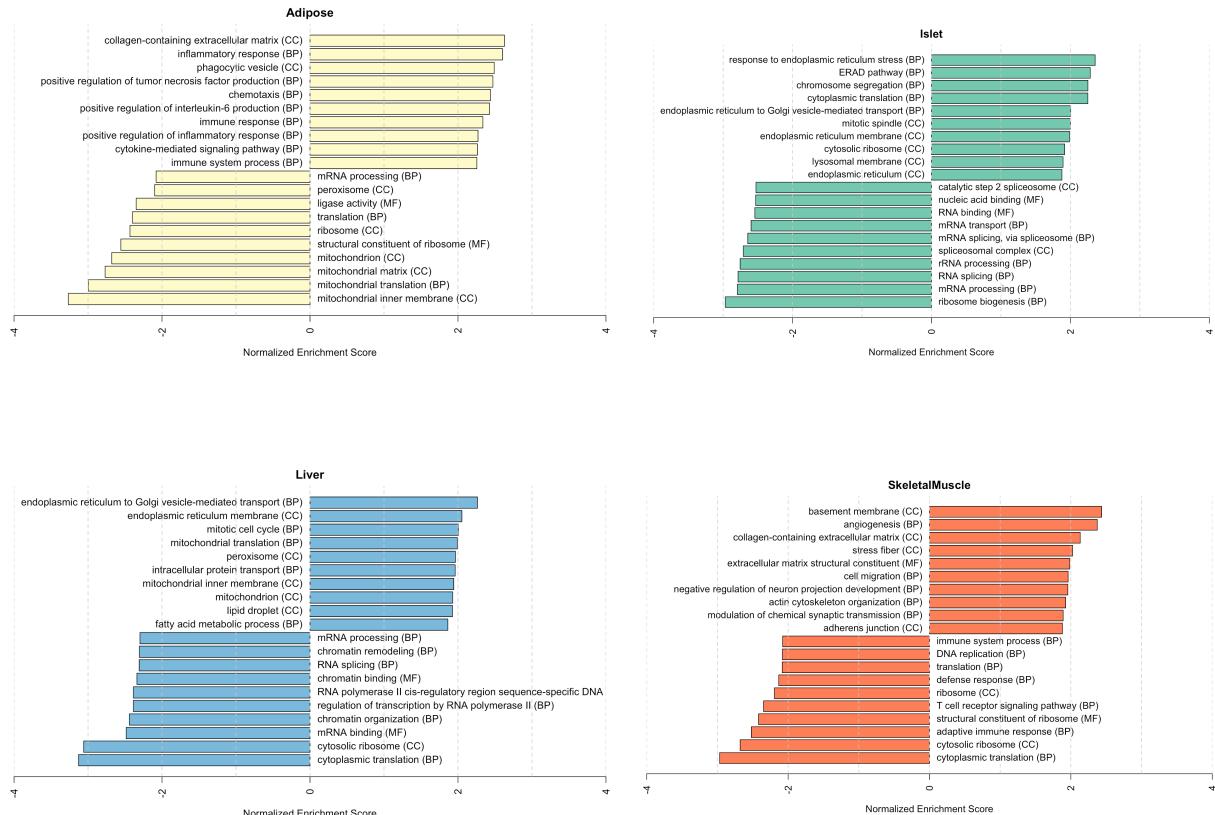


Figure S4: Bar plots showing normalized enrichment scores (NES) for GO terms as determined by fast gene score enrichment analysis (fgsea). Only the top 10 positive and top 10 negative scores are shown. Colors indicate tissue. The name beside each bar shows the name of each enriched GO term. The letters in parentheses indicate whether the term is from the biological process ontology (BP), the molecular function ontology (MF), or the cellular compartment ontology (CC).

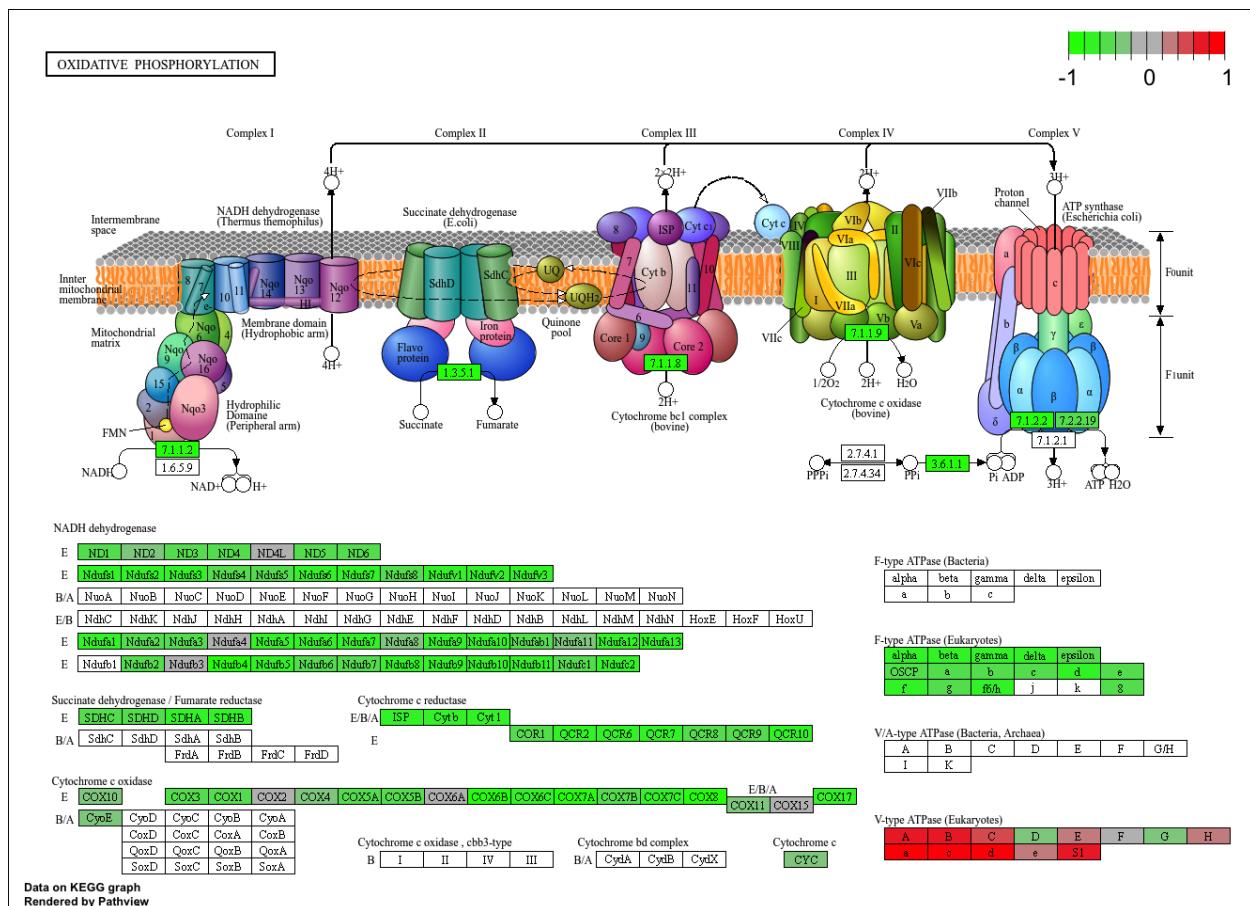


Figure S5: The KEGG pathway for oxidative phosphorylation in mice. Each element is colored based on its HDMA loading from adipose tissue normalized to run from -1 to 1. Genes highlighted in green had negative loadings, and those highlighted in red had positive loadings. Almost the entire pathway was strongly negatively loaded indicating that increased expression of genes involved in oxidative phosphorylation was associated with reduced MDI.

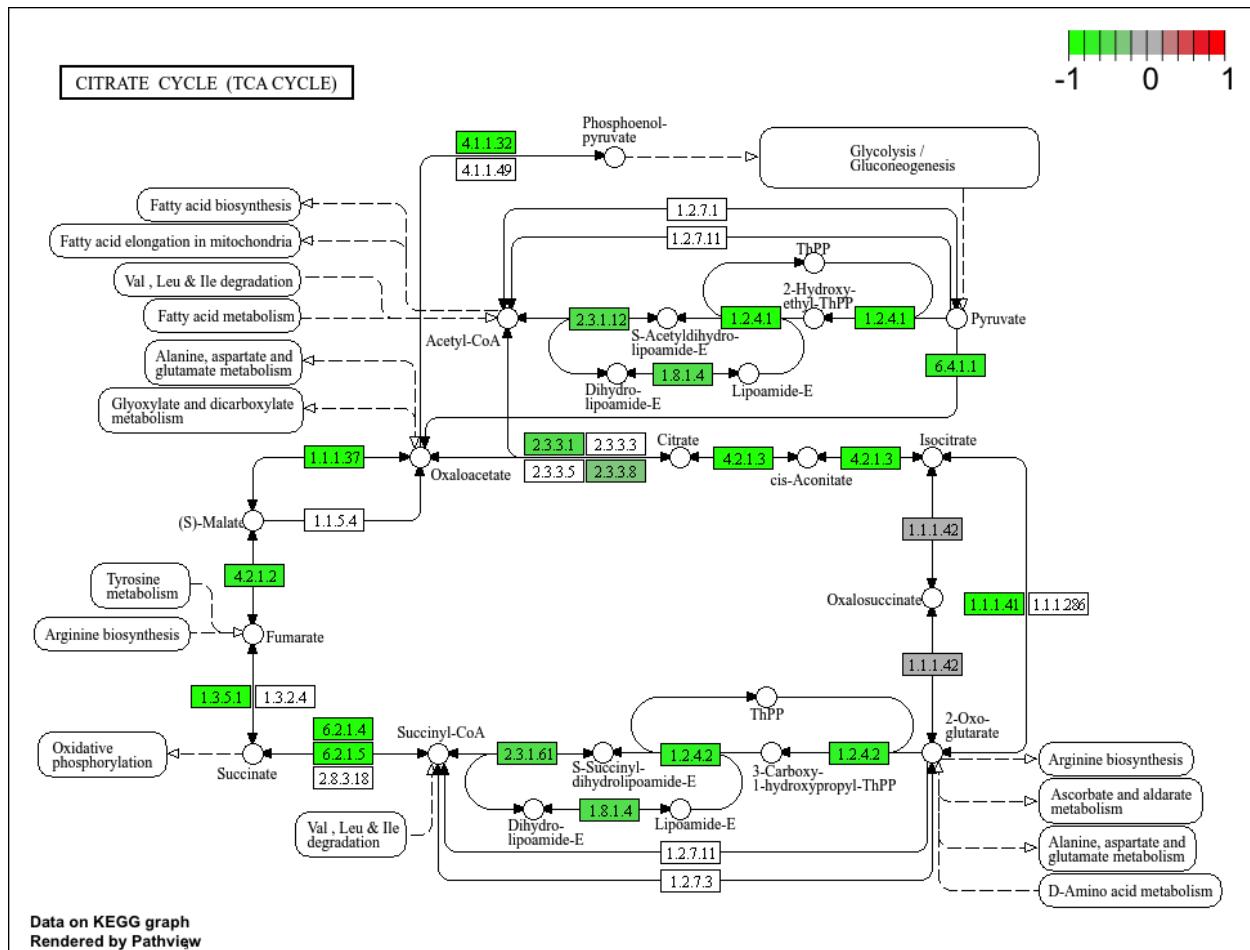


Figure S6: The KEGG pathway for the TCA (citric acid) cycle in mice. Each element is colored based on its HDMA loading from adipose tissue normalized to run from -1 to 1. Genes highlighted in green had negative loadings, and those highlighted in red had positive loadings. Many genes in the cycle were strongly negatively loaded indicating that increased expression of genes involved in branched-chain amino acid degradation was associated with reduced MDI.

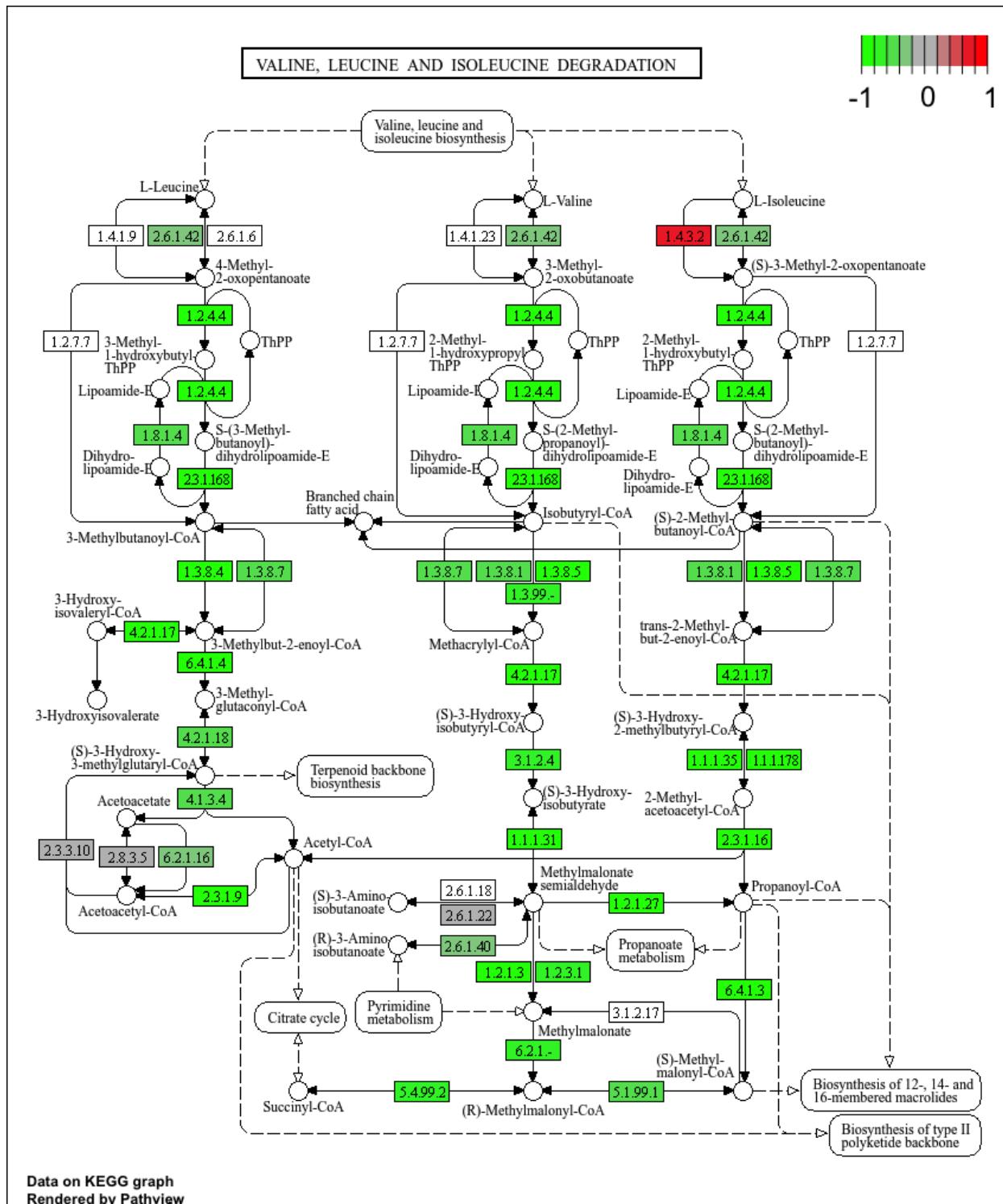


Figure S7: The KEGG pathway for branched-chain amino acid degradation in mice. Each element is colored based on its HDMA loading from adipose tissue normalized to run from -1 to 1. Genes highlighted in green had negative loadings, and those highlighted in red had positive loadings. Almost the entire pathway was strongly negatively loaded indicating that increased expression of genes involved in branched-chain amino acid degradation was associated with reduced MDI.

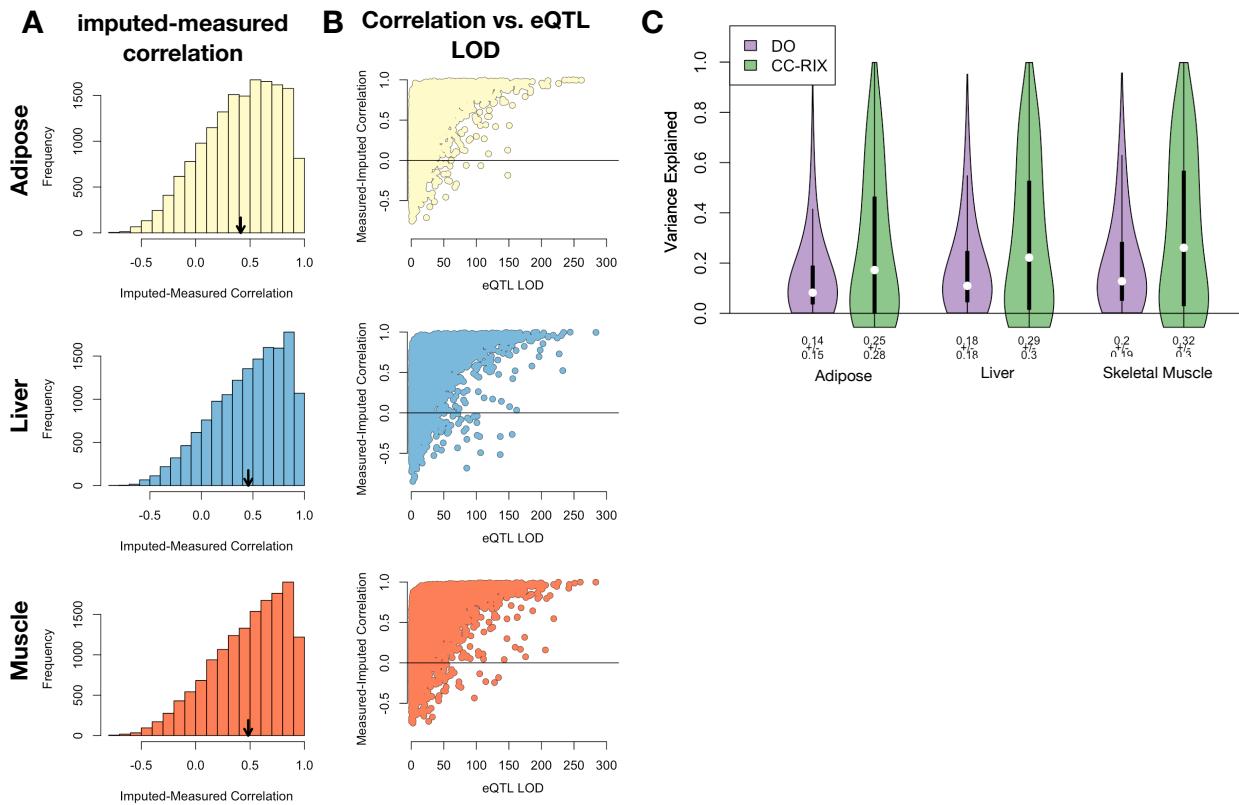


Figure S8: Validation of transcript imputation in the CC-RIX. **A.** Distributions of correlations between imputed and measured transcripts in the CC-RIX. The mean of each distribution is shown by the red line. All distributions were skewed toward positive correlations and had positive means near a Pearson 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 eQTLs. **C.** Variance explained by local genotype in the DO and CC-RIX.

id	norm_ss	cell_iname	pert_type	raw_ss▲	fdr_q_nlog10	set_type	src_set_id
		HA1E	TRT_CP	-0.97	15.65	PCL	CP_PROTEIN_SYNTHESIS_INHIBITOR
		PC3	TRT_SH.CGS	-0.90	15.65	PATHWAY_SET	BIOCARTA_EIF4_PATHWAY
		A375	TRT_CP	-0.87	15.65	MOA_CLASS	RAF_INHIBITOR
		HCC515	TRT_CP	-0.84	15.65	PCL	CP_TOPOISOMERASE_INHIBITOR
		HEPG2	TRT_SH.CGS	-0.82	15.65	PATHWAY_SET	BIOCARTA_BCR_PATHWAY
		PC3	TRT_CP	-0.77	15.65	MOA_CLASS	MTOR_INHIBITOR
		HCC515	TRT_CP	-0.76	15.65	PCL	CP_GLUCOCORTICOID_RECECTORAGONIST
		HCC515	TRT_CP	-0.76	15.65	MOA_CLASS	GLUCOCORTICOID_RECECTORAGONIST
		A375	TRT_CP	-0.72	15.65	MOA_CLASS	MTOR_INHIBITOR
		-666	TRT_CP	-0.70	15.65	PCL	CP_PROTEIN_SYNTHESIS_INHIBITOR
		-666	TRT_CP	-0.68	15.65	PCL	CP_JAK_INHIBITOR
		A549	TRT_CP	-0.67	15.65	PCL	CP_GLUCOCORTICOID_RECECTORAGONIST
		A549	TRT_CP	-0.67	15.65	MOA_CLASS	GLUCOCORTICOID_RECECTORAGONIST
		-666	TRT_CP	-0.57	15.65	PCL	CP_MTOR_INHIBITOR
		-666	TRT_CP	-0.55	15.65	MOA_CLASS	MTOR_INHIBITOR
		-666	TRT_CP	-0.55	15.65	PCL	CP_PI3K_INHIBITOR
		-666	TRT_CP	0.85	15.65	MOA_CLASS	PKC_ACTIVATOR

Figure S9: CMAP results using the *adipose* tissue composite transcript as an input. Table includes results from *all cell types* sorted with a $-\log_{10}(q) > 15$. The results are sorted by the correlation of the query to the input with the most negative results at the top.

id	norm_CS	cell_iname	pert_type	raw_CS▲	fdr_q_nlog10	set_type	src_set_id
		VCAP	TRT_SH.CGS	-0.99	15.65	PATHWAY_SET REACTOME_DOWNSTREAM_TCR_SIGNALING	
		VCAP	TRT_SH.CGS	-0.99	15.65	PATHWAY_SET REACTOME_NOD1_2_SIGNALING_PATHWAY	
		A549	TRT_SH.CGS	-0.92	15.65	PATHWAY_SET BIOCARTA_TNFR1_PATHWAY	
		VCAP	TRT_SH.CGS	-0.92	15.65	PATHWAY_SET HALLMARK_WNT_BETA_CATENIN_SIGNALING	
		HT29	TRT_CP	-0.92	15.65	PCL CP_TUBULIN_INHIBITOR	
-666			TRT_OE	-0.88	15.65	PCL OE_CELL_CYCLE_INHIBITION	
		VCAP	TRT_SH.CGS	-0.87	15.65	PATHWAY_SET REACTOME_P75_NTR_RECECTOR_MEDIATED_SIGNALLING	
		HT29	TRT_CP	-0.86	15.65	MOA_CLASS TUBULIN_INHIBITOR	
		MCF7	TRT_CP	-0.85	15.65	PCL CP_TUBULIN_INHIBITOR	
-666			TRT_CP	-0.81	15.65	PCL CP_PROTEASOME_INHIBITOR	
-666			TRT_SH.CGS	-0.80	15.65	PATHWAY_SET REACTOME_DOWNREGULATION_OF_ERBB2_ERBB3_SIGNALING	
		HCC515	TRT_CP	-0.80	15.65	PCL CP_GLUCOCORTICOID_RECECTORAGONIST	
		HCC515	TRT_CP	-0.80	15.65	MOA_CLASS GLUCOCORTICOID_RECECTORAGONIST	
		A549	TRT_OE	-0.78	15.65	PATHWAY_SET REACTOME_RAF_MAP_KINASE CASCADE	
		A549	TRT_OE	-0.78	15.65	PATHWAY_SET PID_RAS_PATHWAY	
-666			TRT_SH.CGS	-0.78	15.65	PCL KD_RIBOSOMAL_40S_SUBUNIT	
		A549	TRT_OE	-0.76	15.65	PATHWAY_SET REACTOME_SIGNALLING_TO_P38_VIA_RIT_AND_RIN	
		A549	TRT_OE	-0.76	15.65	PATHWAY_SET REACTOME_PROLONGED_ERK_ACTIVATION_EVENTS	
		A549	TRT_OE	-0.73	15.65	PATHWAY_SET PID_TCR_RAS_PATHWAY	
		HA1E	TRT_OE	-0.73	15.65	PATHWAY_SET REACTOME_SHC RELATED_EVENTS	
		HA1E	TRT_OE	-0.71	15.65	PATHWAY_SET PID_EPHB_FWD_PATHWAY	
-666			TRT_CP	-0.70	15.65	MOA_CLASS GLYCOGEN_SYNTHASE_KINASE_INHIBITOR	
		HA1E	TRT_OE	-0.70	15.65	PATHWAY_SET PID_GMCSF_PATHWAY	
		A549	TRT_OE	-0.69	15.65	PATHWAY_SET REACTOME_SIGNALLING_TO_ERKS	
-666			TRT_LIG	-0.69	15.65	PATHWAY_SET PID_ERBB_NETWORK_PATHWAY	
-666			TRT_CP	-0.67	15.65	MOA_CLASS PROTEASOME_INHIBITOR	
-666			TRT_CP	-0.66	15.65	PCL CP_GLYCOGEN_SYNTHASE_KINASE_INHIBITOR	
-666			TRT_CP	0.73	15.65	MOA_CLASS MTOR_INHIBITOR	

Figure S10: CMAP results using the *pancreatic islet* tissue composite transcript as an input. Table includes results from *all cell types* sorted with a $-\log_{10}(q) > 15$. The results are sorted by the correlation of the query to the input with the most negative results at the top.

id	norm_ss	cell_iname	pert_type	raw_ss ▲	fdr_q_nlog10	set_type	src_set_id
		ASC	TRT_CP	-0.94	0.79	PCL	CP_PARP_INHIBITOR
		ASC	TRT_CP	-0.94	0.79	MOA_CLASS	PROTEIN_TYROSINE_KINASE_INHIBITOR
		ASC	TRT_CP	-0.84	0.45	MOA_CLASS	BTK_INHIBITOR
		ASC	TRT_CP	-0.81	0.39	MOA_CLASS	LEUCINE_RICH_REPEAT_KINASE_INHIBITOR
		ASC	TRT_CP	-0.81	0.79	PCL	CP_HSP_INHIBITOR
		ASC	TRT_CP	-0.80	0.93	PCL	CP_EGFR_INHIBITOR
		ASC	TRT_CP	-0.79	0.32	MOA_CLASS	T-TYPE_CALCIUM_CHANNEL_BLOCKER
		ASC	TRT_CP	-0.79	1.09	PCL	CP_MTOR_INHIBITOR
		ASC	TRT_CP	-0.76	0.97	PCL	CP_PI3K_INHIBITOR
		ASC	TRT_CP	-0.75	0.20	MOA_CLASS	HISTONE_DEMETHYLASE_INHIBITOR
		ASC	TRT_CP	-0.74	0.42	PCL	CP_IKK_INHIBITOR
		ASC	TRT_CP	-0.74	0.83	PCL	CP_AURORA_KINASE_INHIBITOR
		ASC	TRT_CP	-0.74	0.17	PCL	CP_LEUCINE_RICH_REPEAT_KINASE_INHIBITOR
		ASC	TRT_CP	-0.72	0.36	PCL	CP_BROMODOMAIN_INHIBITOR
		ASC	TRT_CP	-0.71	1.09	MOA_CLASS	TYROSINE_KINASE_INHIBITOR
		ASC	TRT_CP	-0.70	0.82	PCL	CP_PROTEIN_SYNTHESIS_INHIBITOR
		ASC	TRT_CP	-0.67	0.69	PCL	CP_SRC_INHIBITOR
		ASC	TRT_CP	-0.67	0.81	MOA_CLASS	AURORA_KINASE_INHIBITOR
		ASC	TRT_CP	-0.65	0.89	MOA_CLASS	FLT3_INHIBITOR
		ASC	TRT_CP	-0.62	0.40	MOA_CLASS	FGFR_INHIBITOR
		ASC	TRT_CP	-0.59	0.66	MOA_CLASS	MEK_INHIBITOR
		ASC	TRT_CP	-0.59	0.13	MOA_CLASS	SYK_INHIBITOR
		ASC	TRT_CP	-0.58	0.01	PCL	CP_PKC_INHIBITOR
		ASC	TRT_CP	-0.58	0.65	PCL	CP_HDAC_INHIBITOR
		ASC	TRT_CP	-0.58	0.65	PCL	CP_ATPASE_INHIBITOR
		ASC	TRT_CP	-0.53	0.09	PCL	CP_FLT3_INHIBITOR
		ASC	TRT_CP	-0.53	0.42	PCL	CP_P38_MAPK_INHIBITOR
		ASC	TRT_CP	-0.53	0.22	MOA_CLASS	IKK_INHIBITOR
		ASC	TRT_CP	-0.52	0.58	PCL	CP_VEGFR_INHIBITOR
		ASC	TRT_CP	-0.51	-0.00	PCL	CP_T-TYPE_CALCIUM_CHANNEL_BLOCKER

Figure S11: CMAP results using the *adipose* tissue composite transcript as an input. Table includes the top 30 results derived *only from normal adipocytes* (ASC) regardless of significance. The results are sorted by the correlation of the query to the input with the most negative results at the top.

id	norm_CS	cell_iname	pert_type	raw_CS ▲	fdr_q_nlog10	set_type	src_set_id
		YAPC	TRT_CP	-1.00	0.67	MOA_CLASS	ABL_KINASE_INHIBITOR
		YAPC	TRT_CP	-0.99	0.66	PCL	CP_CDK_INHIBITOR
		YAPC	TRT_CP	-0.97	1.41	PCL	CP_TOPOISOMERASE_INHIBITOR
		YAPC	TRT_CP	-0.95	0.70	MOA_CLASS	THYMIDYLATE_SYNTHASE_INHIBITOR
		YAPC	TRT_CP	-0.95	0.62	MOA_CLASS	ADRENERGIC_INHIBITOR
		YAPC	TRT_CP	-0.94	0.50	MOA_CLASS	BENZODIAZEPINE_RECECTOR_ANTAGONIST
		YAPC	TRT_CP	-0.89	0.63	PCL	CP_RIBONUCLEOTIDE_REDUCTASE_INHIBITOR
		YAPC	TRT_CP	-0.88	0.52	MOA_CLASS	VASOPRESSIN_RECECTOR_ANTAGONIST
		YAPC	TRT_CP	-0.85	0.63	MOA_CLASS	ANGIOTENSIN_RECECTOR_ANTAGONIST
		YAPC	TRT_CP	-0.85	0.33	PCL	CP_CANNABINOID_RECECTORAGONIST
		YAPC	TRT_CP	-0.84	0.30	PCL	CP_RETINOID_RECECTORAGONIST
		YAPC	TRT_CP	-0.83	1.19	MOA_CLASS	NFKB_PATHWAY_INHIBITOR
		YAPC	TRT_CP	-0.83	0.54	MOA_CLASS	DNA_ALKYLATING_DRUG
		YAPC	TRT_CP	-0.80	0.50	MOA_CLASS	CHOLESTEROL_INHIBITOR
		YAPC	TRT_CP	-0.79	0.15	MOA_CLASS	SULFONYLUREA
		YAPC	TRT_CP	-0.78	0.52	MOA_CLASS	HIV_INTEGRASE_INHIBITOR
		YAPC	TRT_CP	-0.78	0.13	MOA_CLASS	LEUKOTRIENE_INHIBITOR
		YAPC	TRT_CP	-0.78	0.45	PCL	CP_PPAR_RECECTORAGONIST
		YAPC	TRT_CP	-0.78	0.54	MOA_CLASS	INSULIN_SENSITIZER
		YAPC	TRT_CP	-0.77	0.51	MOA_CLASS	ESTROGEN_RECECTOR_ANTAGONIST
		YAPC	TRT_CP	-0.77	0.76	MOA_CLASS	DNA_SYNTHESIS_INHIBITOR
		YAPC	TRT_XPR	-0.77	0.67	PATHWAY_SET	BIOCARTA_PARKIN_PATHWAY
		YAPC	TRT_CP	-0.77	0.51	PCL	CP_VEGFR_INHIBITOR
		YAPC	TRT_CP	-0.75	0.39	MOA_CLASS	RNA_SYNTHESIS_INHIBITOR
		YAPC	TRT_CP	-0.72	0.60	MOA_CLASS	BCR-ABL_KINASE_INHIBITOR
		YAPC	TRT_XPR	-0.71	0.66	PATHWAY_SET	BIOCARTA_EIF_PATHWAY
		YAPC	TRT_XPR	-0.69	0.54	PATHWAY_SET	PID_CIRCADIAN_PATHWAY
		YAPC	TRT_CP	-0.68	0.77	MOA_CLASS	TOPOISOMERASE_INHIBITOR
		YAPC	TRT_XPR	-0.64	0.49	PATHWAY_SET	BIOCARTA_CBL_PATHWAY
		YAPC	TRT_CP	-0.64	0.53	MOA_CLASS	TUBULIN_INHIBITOR

Figure S12: CMAP results using the *pancreatic islet* composite transcript as an input. Table includes the top 30 results derived *only from YAPC cells*, which are derived from pancreatic carcinoma cells. Results are shown regardless of significance and are sorted by the correlation of the query to the input with the most negative results at the top.

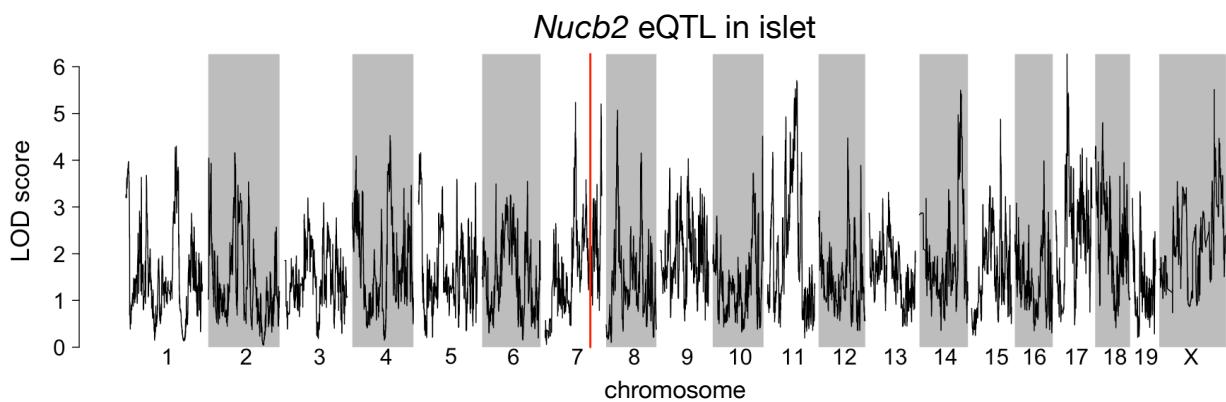


Figure S13: Regulation of *Nucb2* expression in islet. *Nucb2* is encoded on mouse chromosome 7 at 116.5 Mb (red line). In islets the heritability of *Nucb2* expression levels is 69% heritable. This LOD score trace shows that there is no local eQTLs at the position of the gene, nor any strong distal eQTL anywhere else in the genome.

435 **References**

- 436 [1] M. T. Maurano, R. Humbert, E. Rynes, R. E. Thurman, E. Haugen, H. Wang, A. P. Reynolds,
437 R. Sandstrom, H. Qu, J. Brody, A. Shafer, F. Neri, K. Lee, T. Kutyavin, S. Stehling-Sun, A. K.
438 Johnson, T. K. Canfield, E. Giste, M. Diegel, D. Bates, R. S. Hansen, S. Neph, P. J. Sabo, S. Heimfeld,
439 A. Raubitschek, S. Ziegler, C. Cotsapas, N. Sotoodehnia, I. Glass, S. R. Sunyaev, R. Kaul, and J. A.
440 Stamatoyannopoulos. Systematic localization of common disease-associated variation in regulatory DNA.
441 *Science*, 337(6099):1190–1195, Sep 2012.
- 442 [2] K. K. Farh, A. Marson, J. Zhu, M. Kleinewietfeld, W. J. Housley, S. Beik, N. Shores, H. Whitton, R. J.
443 Ryan, A. A. Shishkin, M. Hatan, M. J. Carrasco-Alfonso, D. Mayer, C. J. Luckey, N. A. Patsopoulos,
444 P. L. De Jager, V. K. Kuchroo, C. B. Epstein, M. J. Daly, D. A. Hafler, and B. E. Bernstein. Genetic
445 and epigenetic fine mapping of causal autoimmune disease variants. *Nature*, 518(7539):337–343, Feb
446 2015.
- 447 [3] E. Pennisi. The Biology of Genomes. Disease risk links to gene regulation. *Science*, 332(6033):1031, May
448 2011.
- 449 [4] L. A. Hindorff, P. Sethupathy, H. A. Junkins, E. M. Ramos, J. P. Mehta, F. S. Collins, and T. A. Manolio.
450 Potential etiologic and functional implications of genome-wide association loci for human diseases and
451 traits. *Proc Natl Acad Sci*, 106(23):9362–9367, Jun 2009.
- 452 [5] J. K. Pickrell. Joint analysis of functional genomic data and genome-wide association studies of 18
453 human traits. *Am J Hum Genet*, 94(4):559–573, Apr 2014.
- 454 [6] D. Welter, J. MacArthur, J. Morales, T. Burdett, P. Hall, H. Junkins, A. Klemm, P. Flieck, T. Manolio,
455 L. Hindorff, and H. Parkinson. The NHGRI GWAS Catalog, a curated resource of SNP-trait associations.
456 *Nucleic Acids Res*, 42(Database issue):D1001–1006, Jan 2014.
- 457 [7] Y. I. Li, B. van de Geijn, A. Raj, D. A. Knowles, A. A. Petti, D. Golan, Y. Gilad, and J. K. Pritchard.
458 RNA splicing is a primary link between genetic variation and disease. *Science*, 352(6285):600–604, Apr
459 2016.
- 460 [8] D. Zhou, Y. Jiang, X. Zhong, N. J. Cox, C. Liu, and E. R. Gamazon. A unified framework for joint-tissue
461 transcriptome-wide association and Mendelian randomization analysis. *Nat Genet*, 52(11):1239–1246,
462 Nov 2020.
- 463 [9] E. R. Gamazon, H. E. Wheeler, K. P. Shah, S. V. Mozaffari, K. Aquino-Michaels, R. J. Carroll, A. E.

- 464 Eyler, J. C. Denny, D. L. Nicolae, N. J. Cox, and H. K. Im. A gene-based association method for
465 mapping traits using reference transcriptome data. *Nat Genet*, 47(9):1091–1098, Sep 2015.
- 466 [10] Z. Zhu, F. Zhang, H. Hu, A. Bakshi, M. R. Robinson, J. E. Powell, G. W. Montgomery, M. E. Goddard,
467 N. R. Wray, P. M. Visscher, and J. Yang. Integration of summary data from GWAS and eQTL studies
468 predicts complex trait gene targets. *Nat Genet*, 48(5):481–487, May 2016.
- 469 [11] A. Gusev, A. Ko, H. Shi, G. Bhatia, W. Chung, B. W. Penninx, R. Jansen, E. J. de Geus, D. I. Boomsma,
470 F. A. Wright, P. F. Sullivan, E. Nikkola, M. Alvarez, M. Civelek, A. J. Lusis, T. ki, E. Raitoharju,
471 M. nen, I. ä, O. T. Raitakari, J. Kuusisto, M. Laakso, A. L. Price, P. Pajukanta, and B. Pasaniuc.
472 Integrative approaches for large-scale transcriptome-wide association studies. *Nat Genet*, 48(3):245–252,
473 Mar 2016.
- 474 [12] M. P. Keller, D. M. Gatti, K. L. Schueler, M. E. Rabaglia, D. S. Stapleton, P. Simecek, M. Vincent,
475 S. Allen, A. T. Broman, R. Bacher, C. Kendzierski, K. W. Broman, B. S. Yandell, G. A. Churchill, and
476 A. D. Attie. Genetic Drivers of Pancreatic Islet Function. *Genetics*, 209(1):335–356, May 2018.
- 477 [13] W. L. Crouse, G. R. Keele, M. S. Gastonguay, G. A. Churchill, and W. Valdar. A Bayesian model
478 selection approach to mediation analysis. *PLoS Genet*, 18(5):e1010184, May 2022.
- 479 [14] J. M. Chick, S. C. Munger, P. Simecek, E. L. Huttlin, K. Choi, D. M. Gatti, N. Raghupathy, K. L. Svenson,
480 G. A. Churchill, and S. P. Gygi. Defining the consequences of genetic variation on a proteome-wide scale.
481 *Nature*, 534(7608):500–505, Jun 2016.
- 482 [15] H. E. Wheeler, S. Ploch, A. N. Barbeira, R. Bonazzola, A. Andaleon, A. Fotuhi Siahpirani, A. Saha,
483 A. Battle, S. Roy, and H. K. Im. Imputed gene associations identify replicable trans-acting genes enriched
484 in transcription pathways and complex traits. *Genet Epidemiol*, 43(6):596–608, Sep 2019.
- 485 [16] B. D. Umans, A. Battle, and Y. Gilad. Where Are the Disease-Associated eQTLs? *Trends Genet*,
486 37(2):109–124, Feb 2021.
- 487 [17] N. J. Connally, S. Nazeen, D. Lee, H. Shi, J. Stamatoyannopoulos, S. Chun, C. Cotsapas, C. A. Cassa,
488 and S. R. Sunyaev. The missing link between genetic association and regulatory function. *Elife*, 11, Dec
489 2022.
- 490 [18] H. Mostafavi, J. P. Spence, S. Naqvi, and J. K. Pritchard. Systematic differences in discovery of genetic
491 effects on gene expression and complex traits. *Nat Genet*, 55(11):1866–1875, Nov 2023.
- 492 [19] D. W. Yao, L. J. O'Connor, A. L. Price, and A. Gusev. Quantifying genetic effects on disease mediated
493 by assayed gene expression levels. *Nat Genet*, 52(6):626–633, Jun 2020.

- 494 [20] X. Liu, J. A. Mefford, A. Dahl, Y. He, M. Subramaniam, A. Battle, A. L. Price, and N. Zaitlen. GBAT:
495 a gene-based association test for robust detection of trans-gene regulation. *Genome Biol*, 21(1):211, Aug
496 2020.
- 497 [21] Aparna Nathan, Samira Asgari, Kazuyoshi Ishigaki, Cristian Valencia, Tiffany Amariuta, Yang Luo,
498 Jessica I Beynor, Yuriy Baglaenko, Sara Suliman, Alkes L Price, et al. Single-cell eqtl models reveal
499 dynamic t cell state dependence of disease loci. *Nature*, 606(7912):120–128, 2022.
- 500 [22] E. Sollis, A. Mosaku, A. Abid, A. Buniello, M. Cerezo, L. Gil, T. Groza, O. §, P. Hall, J. Hayhurst,
501 A. Ibrahim, Y. Ji, S. John, E. Lewis, J. A. L. MacArthur, A. McMahon, D. Osumi-Sutherland,
502 K. Panoutsopoulou, Z. Pendlington, S. Ramachandran, R. Stefancsik, J. Stewart, P. Whetzel, R. Wilson,
503 L. Hindorff, F. Cunningham, S. A. Lambert, M. Inouye, H. Parkinson, and L. W. Harris. The NHGRI-EBI
504 GWAS Catalog: knowledgebase and deposition resource. *Nucleic Acids Res*, 51(D1):D977–D985, Jan
505 2023.
- 506 [23] R. J. F. Loos and G. S. H. Yeo. The genetics of obesity: from discovery to biology. *Nat Rev Genet*,
507 23(2):120–133, Feb 2022.
- 508 [24] R. K. Singh, P. Kumar, and K. Mahalingam. Molecular genetics of human obesity: A comprehensive
509 review. *C R Biol*, 340(2):87–108, Feb 2017.
- 510 [25] P. Arner. Obesity—a genetic disease of adipose tissue? *Br J Nutr*, 83 Suppl 1:9–16, Mar 2000.
- 511 [26] Mark P Keller, Mary E Rabaglia, Kathryn L Schueler, Donnie S Stapleton, Daniel M Gatti, Matthew
512 Vincent, Kelly A Mitok, Ziyue Wang, Takanao Ishimura, Shane P Simonett, et al. Gene loci associated
513 with insulin secretion in islets from nondiabetic mice. *The Journal of Clinical Investigation*, 129(10):4419–
514 4432, 2019.
- 515 [27] G. A. Churchill, D. M. Gatti, S. C. Munger, and K. L. Svenson. The Diversity Outbred mouse population.
516 *Mamm Genome*, 23(9-10):713–718, Oct 2012.
- 517 [28] Elissa J Chesler, Darla R Miller, Lisa R Branstetter, Leslie D Galloway, Barbara L Jackson, Vivek M
518 Philip, Brynn H Voy, Cymbeline T Culiat, David W Threadgill, Robert W Williams, et al. The
519 collaborative cross at oak ridge national laboratory: developing a powerful resource for systems genetics.
520 *Mammalian Genome*, 19:382–389, 2008.
- 521 [29] Michael C Saul, Vivek M Philip, Laura G Reinholdt, and Elissa J Chesler. High-diversity mouse
522 populations for complex traits. *Trends in Genetics*, 35(7):501–514, 2019.

- 523 [30] D. W. Threadgill, D. R. Miller, G. A. Churchill, and F. P. de Villena. The collaborative cross: a
524 recombinant inbred mouse population for the systems genetic era. *ILAR J*, 52(1):24–31, 2011.
- 525 [31] S. M. Clee and A. D. Attie. The genetic landscape of type 2 diabetes in mice. *Endocr Rev*, 28(1):48–83,
526 Feb 2007.
- 527 [32] K. W. Broman, D. M. Gatti, P. Simecek, N. A. Furlotte, P. Prins, Š. Sen, B. S. Yandell, and G. A.
528 Churchill. R/qt12: Software for Mapping Quantitative Trait Loci with High-Dimensional Data and
529 Multiparent Populations. *Genetics*, 211(2):495–502, Feb 2019.
- 530 [33] M. Helmer, S. Warrington, A. R. Mohammadi-Nejad, J. L. Ji, A. Howell, B. Rosand, A. Anticevic,
531 S. N. Sotiropoulos, and J. D. Murray. On the stability of canonical correlation analysis and partial least
532 squares with application to brain-behavior associations. *Commun Biol*, 7(1):217, Feb 2024.
- 533 [34] Gennady Korotkevich, Vladimir Sukhov, and Alexey Sergushichev. Fast gene set enrichment analysis.
534 *bioRxiv*, 2019.
- 535 [35] A. Subramanian, P. Tamayo, V. K. Mootha, S. Mukherjee, B. L. Ebert, M. A. Gillette, A. Paulovich,
536 S. L. Pomeroy, T. R. Golub, E. S. Lander, and J. P. Mesirov. Gene set enrichment analysis: a
537 knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci U S A*,
538 102(43):15545–15550, Oct 2005.
- 539 [36] S. Subramanian and A. Chait. The effect of dietary cholesterol on macrophage accumulation in adipose
540 tissue: implications for systemic inflammation and atherosclerosis. *Curr Opin Lipidol*, 20(1):39–44, Feb
541 2009.
- 542 [37] I. Akoumianakis, N. Akawi, and C. Antoniades. Exploring the Crosstalk between Adipose Tissue and
543 the Cardiovascular System. *Korean Circ J*, 47(5):670–685, Sep 2017.
- 544 [38] I. S. Stafeev, A. V. Vorotnikov, E. I. Ratner, M. Y. Menshikov, and Y. V. Parfyonova. Latent Inflammation
545 and Insulin Resistance in Adipose Tissue. *Int J Endocrinol*, 2017:5076732, 2017.
- 546 [39] I. P. Fischer, M. Irmler, C. W. Meyer, S. J. Sachs, F. Neff, M. de Angelis, J. Beckers, M. H. p, S. M.
547 Hofmann, and S. Ussar. A history of obesity leaves an inflammatory fingerprint in liver and adipose
548 tissue. *Int J Obes (Lond)*, 42(3):507–517, Mar 2018.
- 549 [40] S. Chung, H. Cuffe, S. M. Marshall, A. L. McDaniel, J. H. Ha, K. Kavanagh, C. Hong, P. Tontonoz,
550 R. E. Temel, and J. S. Parks. Dietary cholesterol promotes adipocyte hypertrophy and adipose tissue
551 inflammation in visceral, but not in subcutaneous, fat in monkeys. *Arterioscler Thromb Vasc Biol*,
552 34(9):1880–1887, Sep 2014.

- 553 [41] V. Kus, T. Prazak, P. Brauner, M. Hensler, O. Kuda, P. Flachs, P. Janovska, D. Medrikova, M. Rossmeisl,
554 Z. Jilkova, B. Stefl, E. Pastalkova, Z. Drahota, J. Houstek, and J. Kopecky. Induction of muscle
555 thermogenesis by high-fat diet in mice: association with obesity-resistance. *Am J Physiol Endocrinol*
556 *Metab*, 295(2):E356–367, Aug 2008.
- 557 [42] C. B. Newgard. Interplay between lipids and branched-chain amino acids in development of insulin
558 resistance. *Cell Metab*, 15(5):606–614, May 2012.
- 559 [43] D. D. Sears, G. Hsiao, A. Hsiao, J. G. Yu, C. H. Courtney, J. M. Ofrecio, J. Chapman, and S. Subramaniam.
560 Mechanisms of human insulin resistance and thiazolidinedione-mediated insulin sensitization. *Proc Natl*
561 *Acad Sci U S A*, 106(44):18745–18750, Nov 2009.
- 562 [44] R. Stienstra, C. Duval, M. ller, and S. Kersten. PPARs, Obesity, and Inflammation. *PPAR Res*,
563 2007:95974, 2007.
- 564 [45] O. Gavrilova, M. Haluzik, K. Matsusue, J. J. Cutson, L. Johnson, K. R. Dietz, C. J. Nicol, C. Vinson,
565 F. J. Gonzalez, and M. L. Reitman. Liver peroxisome proliferator-activated receptor gamma contributes
566 to hepatic steatosis, triglyceride clearance, and regulation of body fat mass. *J Biol Chem*, 278(36):34268–
567 34276, Sep 2003.
- 568 [46] K. Matsusue, M. Haluzik, G. Lambert, S. H. Yim, O. Gavrilova, J. M. Ward, B. Brewer, M. L. Reitman,
569 and F. J. Gonzalez. Liver-specific disruption of PPARgamma in leptin-deficient mice improves fatty
570 liver but aggravates diabetic phenotypes. *J Clin Invest*, 111(5):737–747, Mar 2003.
- 571 [47] D. Patsouris, J. K. Reddy, M. ller, and S. Kersten. Peroxisome proliferator-activated receptor alpha
572 mediates the effects of high-fat diet on hepatic gene expression. *Endocrinology*, 147(3):1508–1516, Mar
573 2006.
- 574 [48] S. E. Schadinger, N. L. Bucher, B. M. Schreiber, and S. R. Farmer. PPARgamma2 regulates lipogenesis
575 and lipid accumulation in steatotic hepatocytes. *Am J Physiol Endocrinol Metab*, 288(6):E1195–1205,
576 Jun 2005.
- 577 [49] W. Motomura, M. Inoue, T. Ohtake, N. Takahashi, M. Nagamine, S. Tanno, Y. Kohgo, and T. Okumura.
578 Up-regulation of ADRP in fatty liver in human and liver steatosis in mice fed with high fat diet. *Biochem*
579 *Biophys Res Commun*, 340(4):1111–1118, Feb 2006.
- 580 [50] A. Srivastava, A. P. Morgan, M. L. Najarian, V. K. Sarsani, J. S. Sigmon, J. R. Shorter, A. Kashfeen,
581 R. C. McMullan, L. H. Williams, P. guez, M. T. Ferris, P. Sullivan, P. Hock, D. R. Miller, T. A. Bell,

- 582 L. McMillan, G. A. Churchill, and F. P. de Villena. Genomes of the Mouse Collaborative Cross. *Genetics*,
583 206(2):537–556, Jun 2017.
- 584 [51] A. Roberts, F. Pardo-Manuel de Villena, W. Wang, L. McMillan, and D. W. Threadgill. The poly-
585 morphism architecture of mouse genetic resources elucidated using genome-wide resequencing data:
586 implications for QTL discovery and systems genetics. *Mamm Genome*, 18(6-7):473–481, Jul 2007.
- 587 [52] G. A. Churchill, D. C. Airey, H. Allayee, J. M. Angel, A. D. Attie, J. Beatty, W. D. Beavis, J. K.
588 Belknap, B. Bennett, W. Berrettini, A. Bleich, M. Bogue, K. W. Broman, K. J. Buck, E. Buckler,
589 M. Burmeister, E. J. Chesler, J. M. Cheverud, S. Clapcote, M. N. Cook, R. D. Cox, J. C. Crabbe,
590 W. E. Crusio, A. Darvasi, C. F. Deschepper, R. W. Doerge, C. R. Farber, J. Forejt, D. Gaile, S. J.
591 Garlow, H. Geiger, H. Gershenson, T. Gordon, J. Gu, W. Gu, G. de Haan, N. L. Hayes, C. Heller,
592 H. Himmelbauer, R. Hitzemann, K. Hunter, H. C. Hsu, F. A. Iraqi, B. Ivandic, H. J. Jacob, R. C. Jansen,
593 K. J. Jepsen, D. K. Johnson, T. E. Johnson, G. Kempermann, C. Kendziora, M. Kotb, R. F. Kooy,
594 B. Llamas, F. Lammert, J. M. Lassalle, P. R. Lowenstein, L. Lu, A. Lusis, K. F. Manly, R. Marcucio,
595 D. Matthews, J. F. Medrano, D. R. Miller, G. Mittelman, B. A. Mock, J. S. Mogil, X. Montagutelli,
596 G. Morahan, D. G. Morris, R. Mott, J. H. Nadeau, H. Nagase, R. S. Nowakowski, B. F. O’Hara, A. V.
597 Osadchuk, G. P. Page, B. Paigen, K. Paigen, A. A. Palmer, H. J. Pan, L. Peltonen-Palotie, J. Peirce,
598 D. Pomp, M. Pravenec, D. R. Prows, Z. Qi, R. H. Reeves, J. Roder, G. D. Rosen, E. E. Schadt, L. C.
599 Schalkwyk, Z. Seltzer, K. Shimomura, S. Shou, M. J. ä, L. D. Siracusa, H. W. Snoeck, J. L. Spearow,
600 K. Svenson, L. M. Tarantino, D. Threadgill, L. A. Toth, W. Valdar, F. P. de Villena, C. Warden,
601 S. Whatley, R. W. Williams, T. Wiltshire, N. Yi, D. Zhang, M. Zhang, and F. Zou. The Collaborative
602 Cross, a community resource for the genetic analysis of complex traits. *Nat Genet*, 36(11):1133–1137,
603 Nov 2004.
- 604 [53] J. Y. Huh, Y. J. Park, M. Ham, and J. B. Kim. Crosstalk between adipocytes and immune cells in
605 adipose tissue inflammation and metabolic dysregulation in obesity. *Mol Cells*, 37(5):365–371, May 2014.
- 606 [54] J. Lamb, E. D. Crawford, D. Peck, J. W. Modell, I. C. Blat, M. J. Wrobel, J. Lerner, J. P. Brunet,
607 A. Subramanian, K. N. Ross, M. Reich, H. Hieronymus, G. Wei, S. A. Armstrong, S. J. Haggarty,
608 P. A. Clemons, R. Wei, S. A. Carr, E. S. Lander, and T. R. Golub. The Connectivity Map: using
609 gene-expression signatures to connect small molecules, genes, and disease. *Science*, 313(5795):1929–1935,
610 Sep 2006.
- 611 [55] S. Amin, A. Lux, and F. O’Callaghan. The journey of metformin from glycaemic control to mTOR
612 inhibition and the suppression of tumour growth. *Br J Clin Pharmacol*, 85(1):37–46, Jan 2019.

- 613 [56] A. Kezic, L. Popovic, and K. Lalic. mTOR Inhibitor Therapy and Metabolic Consequences: Where Do
614 We Stand? *Oxid Med Cell Longev*, 2018:2640342, 2018.
- 615 [57] A. D. Barlow, M. L. Nicholson, and T. P. Herbert. -cells and a review of the underlying molecular
616 mechanisms. *Diabetes*, 62(8):2674–2682, Aug 2013.
- 617 [58] Y. Gu, J. Lindner, A. Kumar, W. Yuan, and M. A. Magnuson. Rictor/mTORC2 is essential for
618 maintaining a balance between beta-cell proliferation and cell size. *Diabetes*, 60(3):827–837, Mar 2011.
- 619 [59] E. B. Geer, J. Islam, and C. Buettner. Mechanisms of glucocorticoid-induced insulin resistance: focus on
620 adipose tissue function and lipid metabolism. *Endocrinol Metab Clin North Am*, 43(1):75–102, Mar 2014.
- 621 [60] J. X. Li and C. L. Cummins. Fresh insights into glucocorticoid-induced diabetes mellitus and new
622 therapeutic directions. *Nat Rev Endocrinol*, 18(9):540–557, Sep 2022.
- 623 [61] R. A. Lee, C. A. Harris, and J. C. Wang. Glucocorticoid Receptor and Adipocyte Biology. *Nucl Receptor*
624 Res, 5, 2018.
- 625 [62] S. Viengchareun, P. Penfornis, M. C. Zennaro, and M. s. Mineralocorticoid and glucocorticoid receptors
626 inhibit UCP expression and function in brown adipocytes. *Am J Physiol Endocrinol Metab*, 280(4):E640–
627 649, Apr 2001.
- 628 [63] J. Liu, X. Kong, L. Wang, H. Qi, W. Di, X. Zhang, L. Wu, X. Chen, J. Yu, J. Zha, S. Lv, A. Zhang,
629 P. Cheng, M. Hu, Y. Li, J. Bi, Y. Li, F. Hu, Y. Zhong, Y. Xu, and G. Ding. -HSD1 in regulating brown
630 adipocyte function. *J Mol Endocrinol*, 50(1):103–113, Feb 2013.
- 631 [64] L. E. Ramage, M. Akyol, A. M. Fletcher, J. Forsythe, M. Nixon, R. N. Carter, E. J. van Beek,
632 N. M. Morton, B. R. Walker, and R. H. Stimson. Glucocorticoids Acutely Increase Brown Adipose
633 Tissue Activity in Humans, Revealing Species-Specific Differences in UCP-1 Regulation. *Cell Metab*,
634 24(1):130–141, Jul 2016.
- 635 [65] J. L. Barclay, H. Agada, C. Jang, M. Ward, N. Wetzig, and K. K. Ho. Effects of glucocorticoids on
636 human brown adipocytes. *J Endocrinol*, 224(2):139–147, Feb 2015.
- 637 [66] M. ó, R. Gupte, W. L. Kraus, P. Pacher, and P. Bai. PARPs in lipid metabolism and related diseases.
638 *Prog Lipid Res*, 84:101117, Nov 2021.
- 639 [67] P. Bai, C. ó, H. Oudart, A. nszki, Y. Cen, C. Thomas, H. Yamamoto, A. Huber, B. Kiss, R. H.
640 Houtkooper, K. Schoonjans, V. Schreiber, A. A. Sauve, J. Menissier-de Murcia, and J. Auwerx. PARP-1

- 641 inhibition increases mitochondrial metabolism through SIRT1 activation. *Cell Metab*, 13(4):461–468,
642 Apr 2011.
- 643 [68] A. Chiarugi and M. A. Moskowitz. Cell biology. PARP-1—a perpetrator of apoptotic cell death? *Science*,
644 297(5579):200–201, Jul 2002.
- 645 [69] M. Althubiti, R. Almaimani, S. Y. Eid, M. Elzubaier, B. Refaat, S. Idris, T. A. Alqurashi, and M. Z.
646 El-Readi. BTK targeting suppresses inflammatory genes and ameliorates insulin resistance. *Eur Cytokine
647 Netw*, 31(4):168–179, Dec 2020.
- 648 [70] C. Skrabs, W. F. Pickl, T. Perkmann, U. ger, and A. Gessl. Rapid decline in insulin antibodies and
649 glutamic acid decarboxylase autoantibodies with ibrutinib therapy of chronic lymphocytic leukaemia. *J
650 Clin Pharm Ther*, 43(1):145–149, Feb 2018.
- 651 [71] Hans Hacker and Michael Karin. Regulation and function of ikk and ikk-related kinases. *Science's
652 STKE*, 2006(357):re13–re13, 2006.
- 653 [72] E. A. Oral, S. M. Reilly, A. V. Gomez, R. Meral, L. Butz, N. Ajluni, T. L. Chenevert, E. Korytnaya,
654 A. H. Neidert, R. Hench, D. Rus, J. F. Horowitz, B. Poirier, P. Zhao, K. Lehmann, M. Jain, R. Yu,
655 C. Liddle, M. Ahmadian, M. Downes, R. M. Evans, and A. R. Saltiel. and TBK1 Improves Glucose
656 Control in a Subset of Patients with Type 2 Diabetes. *Cell Metab*, 26(1):157–170, Jul 2017.
- 657 [73] M. C. Arkan, A. L. Hevener, F. R. Greten, S. Maeda, Z. W. Li, J. M. Long, A. Wynshaw-Boris, G. Poli,
658 J. Olefsky, and M. Karin. IKK-beta links inflammation to obesity-induced insulin resistance. *Nat Med*,
659 11(2):191–198, Feb 2005.
- 660 [74] M. Clark, C. J. Kroger, Q. Ke, and R. M. Tisch. The Role of T Cell Receptor Signaling in the
661 Development of Type 1 Diabetes. *Front Immunol*, 11:615371, 2020.
- 662 [75] P. ndell, A. C. Carlsson, A. Larsson, O. Melander, T. Wessman, J. v, and T. Ruge. TNFR1 is associated
663 with short-term mortality in patients with diabetes and acute dyspnea seeking care at the emergency
664 department. *Acta Diabetol*, 57(10):1145–1150, Oct 2020.
- 665 [76] A. M. Keestra-Gounder, M. X. Byndloss, N. Seyffert, B. M. Young, A. vez Arroyo, A. Y. Tsai, S. A.
666 Cevallos, M. G. Winter, O. H. Pham, C. R. Tiffany, M. F. de Jong, T. Kerrinnes, R. Ravindran, P. A.
667 Luciw, S. J. McSorley, A. J. umler, and R. M. Tsolis. NOD1 and NOD2 signalling links ER stress with
668 inflammation. *Nature*, 532(7599):394–397, Apr 2016.
- 669 [77] A. M. Keestra-Gounder and R. M. Tsolis. NOD1 and NOD2: Beyond Peptidoglycan Sensing. *Trends
670 Immunol*, 38(10):758–767, Oct 2017.

- 671 [78] J. Montane, L. Cadavez, and A. Novials. Stress and the inflammatory process: a major cause of
672 pancreatic cell death in type 2 diabetes. *Diabetes Metab Syndr Obes*, 7:25–34, 2014.
- 673 [79] B. B. Kahn and T. E. McGraw. , and type 2 diabetes. *N Engl J Med*, 363(27):2667–2669, Dec 2010.
- 674 [80] S. Del Prato and N. Pulizzi. The place of sulfonylureas in the therapy for type 2 diabetes mellitus.
675 *Metabolism*, 55(5 Suppl 1):S20–27, May 2006.
- 676 [81] X. Liu, Y. I. Li, and J. K. Pritchard. Trans Effects on Gene Expression Can Drive Omnipotent Inheritance.
677 *Cell*, 177(4):1022–1034, May 2019.
- 678 [82] B. Hallgrímsson, R. M. Green, D. C. Katz, J. L. Fish, F. P. Bernier, C. C. Roseman, N. M. Young,
679 J. M. Cheverud, and R. S. Marcucio. The developmental-genetics of canalization. *Semin Cell Dev Biol*,
680 88:67–79, Apr 2019.
- 681 [83] M. L. Siegal and A. Bergman. Waddington's canalization revisited: developmental stability and evolution.
682 *Proc Natl Acad Sci U S A*, 99(16):10528–10532, Aug 2002.
- 683 [84] A. B. Paaby and G. Gibson. Cryptic Genetic Variation in Evolutionary Developmental Genetics. *Biology*
684 (*Basel*), 5(2), Jun 2016.
- 685 [85] E. A. Boyle, Y. I. Li, and J. K. Pritchard. An Expanded View of Complex Traits: From Polygenic to
686 Omnipotent. *Cell*, 169(7):1177–1186, Jun 2017.
- 687 [86] Naomi R Wray, Cisca Wijmenga, Patrick F Sullivan, Jian Yang, and Peter M Visscher. Common disease
688 is more complex than implied by the core gene omnipotent model. *Cell*, 173(7):1573–1580, 2018.
- 689 [87] H. Shimizu and A. Osaki. Nesfatin/Nucleobindin-2 (NUCB2) and Glucose Homeostasis. *Curr Hypertens*
690 *Rev*, pages Nesfatin/Nucleobindin-2 (NUCB2) and Glucose Homeostasis., Jul 2014.
- 691 [88] M. Nakata and T. Yada. Role of NUCB2/nesfatin-1 in glucose control: diverse functions in islets,
692 adipocytes and brain. *Curr Pharm Des*, 19(39):6960–6965, 2013.
- 693 [89] M. Riva, M. D. Nitert, U. Voss, R. Sathanoori, A. Lindqvist, C. Ling, and N. Wierup. Nesfatin-1
694 stimulates glucagon and insulin secretion and beta cell NUCB2 is reduced in human type 2 diabetic
695 subjects. *Cell Tissue Res*, 346(3):393–405, Dec 2011.