

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
²² in gene regulatory regions^{1–7}, suggesting a relatively simple causal model in which a variant alters the
²³ homeostatic expression level of a nearby (local) gene which, in turn, alters a trait. Statistical methods such
²⁴ as transcriptome-wide association studies (TWAS)^{8–11} and summary data-based Mendelian randomization
²⁵ (SMR)¹⁰ have used this idea to identify genes associated with multiple disease traits^{12–15}. However, despite

26 the great promise of these methods, explaining trait effects with local gene regulation has been more difficult
27 than initially assumed^{16;17}. Although trait-associated variants typically lie in non-coding, regulatory regions,
28 these variants often have no detectable effects on gene expression¹⁶ and tend not to co-localize with expression
29 quantitative trait loci (eQTLs)^{17;18}. These observations suggest that the relationship among genetic variants,
30 gene expression, and 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 more distant (distal) genetic effects as potential drivers
33 of trait variation^{18–20;15;21}. In general, distal effects are defined as being greater than 4 or 5Mb away from
34 the transcription start site of a given gene. We use the terms local and distal rather than *cis* and *trans*
35 because *cis* and *trans* have specific biochemical meanings²², whereas local and distal are defined only by
36 genomic position. The importance of distal genetic effects is proposed in the omnigenic model, which posits
37 that trait-driving genes are cumulatively influenced by many distal variants. In this view, the heritable
38 transcriptomic signatures driving clinical traits are an emergent state arising from the myriad molecular
39 interactions defining and constraining gene expression. Consistent with this view, it has been suggested
40 that part of the difficulty in explaining trait variation through local eQTLs may arise in part because gene
41 expression is not measured in the appropriate cell types¹⁶, or cell states²³, and thus local eQTLs influencing
42 traits cannot be detected in bulk tissue samples. This context dependence emphasizes the essential role of
43 complex regulatory and tissue networks in mediating variant effects. The mechanistic dissection of complex
44 traits in this model is more challenging because it requires addressing network-mediated effects that are
45 weaker and greater in number. However, the comparative importance of distal effects over local effects is
46 currently only conjectured and extremely challenging to address in human populations.

47 To assess the role of wide-spread distal gene regulation in the genetic architecture of complex traits, we used
48 genetically diverse mice as a model system. In mice we can obtain simultaneous measurements of the genome,
49 transcriptome, and phenome in all individuals. We used diet-induced obesity and metabolic disease as an
50 archetypal example of a complex trait. In humans, these phenotypes are genetically complex with hundreds of
51 variants mapped through GWAS^{24;25} that are known to act through multiple tissues^{26;27}. Likewise in mice,
52 metabolic traits are also genetically complex²⁸ and synteny analysis implicates a high degree of concordance
53 in the genetic architecture between species^{28;12}. Furthermore, in contrast to humans, in mice we have access
54 to multiple disease-relevant tissues in the same individuals with sufficient numbers for adequate statistical
55 power.

56 We generated two complementary data sets: a discovery data set in a large population of Diversity Outbred
57 (DO) mice²⁹, and an independent validation data set derived by crossing inbred strains from the Collaborative

58 Cross (CC) recombinant inbred lines³⁰ to form CC recombinant inbred intercross (CC-RIX) mice. Both
59 populations were maintained on a high-fat, high-sugar diet to model diet-induced obesity and metabolic
60 disease¹².

61 The DO population and CC recombinant inbred lines were derived from the same eight inbred founder
62 strains: five classical lab strains and three strains more recently derived from wild mice²⁹, representing three
63 subspecies and capturing 90% of the known variation in laboratory mice³¹. The DO mice are maintained
64 with a breeding scheme that ensures equal contributions from each founder across the genome thus rendering
65 almost the whole genome visible to genetic inquiry and maximizing power to detect eQTLs²⁹. The CC mice
66 were initially intercrossed to recombine the genomes from all eight founders, and then inbred for at least 20
67 generations to create recombinant inbred lines^{30;32;31}. Because these two populations have common ancestral
68 haplotypes but highly distinct kinship structure, we could directly and unambiguously compare the local
69 genetic effects on gene expression at the whole-transcriptome level while varying the population structure
70 driving distal regulation.

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

86 **Results**

87 **Genetic variation contributed to wide phenotypic variation**

88 Although the environment was consistent across the DO mice, the genetic diversity present in this population
89 resulted in widely varying distributions across physiological measurements (Fig. 1). For example, body
90 weights of adult individuals varied from less than the average adult C57BL/6J (B6) body weight to several
91 times the body weight of a B6 adult in both sexes (Males: 18.5 - 69.1g, Females: 16.0 - 54.8g) (Fig. 1A).
92 Fasting blood glucose (FBG) also varied considerably (Fig. 1B), although few of the animals had FBG levels
93 that would indicate pre-diabetes (19 animals, 3.8%), or diabetes (7 animals, 1.4%) according to previously
94 developed cutoffs (pre-diabetes: $\text{FBG} \geq 250 \text{ mg/dL}$, diabetes: $\text{FBG} \geq 300 \text{ mg/dL}$)³³. Males had higher
95 FBG than females on average (Fig. 1C) as has been observed before suggesting either that males were more
96 susceptible to metabolic disease on the high-fat, high-sugar (HFHS) diet, or that males and females may
97 require different thresholds for pre-diabetes and diabetes.

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

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

106 **Distal Heritability Correlated with Phenotype Relevance**

107 To comprehensively assess the genetic control of gene expression in metabolic disease we measured overall
108 gene expression via bulk RNA-Seq in adipose, islet, liver, and skeletal muscle in the DO cohort (Supp. Fig.
109 S2). We performed eQTL analysis using R/qtL2³⁴ (Methods) and identified both local and distal eQTLs for
110 transcripts in each of the four tissues (Supp. Fig. S2B-E). Significant local eQTLs far outnumbered distal
111 eQTLs (Supp. Fig. S2F) and tended to be shared across tissues (Supp. Fig. S2G) whereas the few significant
112 distal eQTLs we identified tended to be tissue-specific (Supp. Fig. S2H)

113 We estimated the heritability of each transcript in terms of local and all non-local (distal) genetic factors
114 (Methods). Overall, local and distal genetic factors contributed approximately equally to transcript abundance.

115 In all tissues, both local and distal factors explained between 8 and 17% of the variance in the median

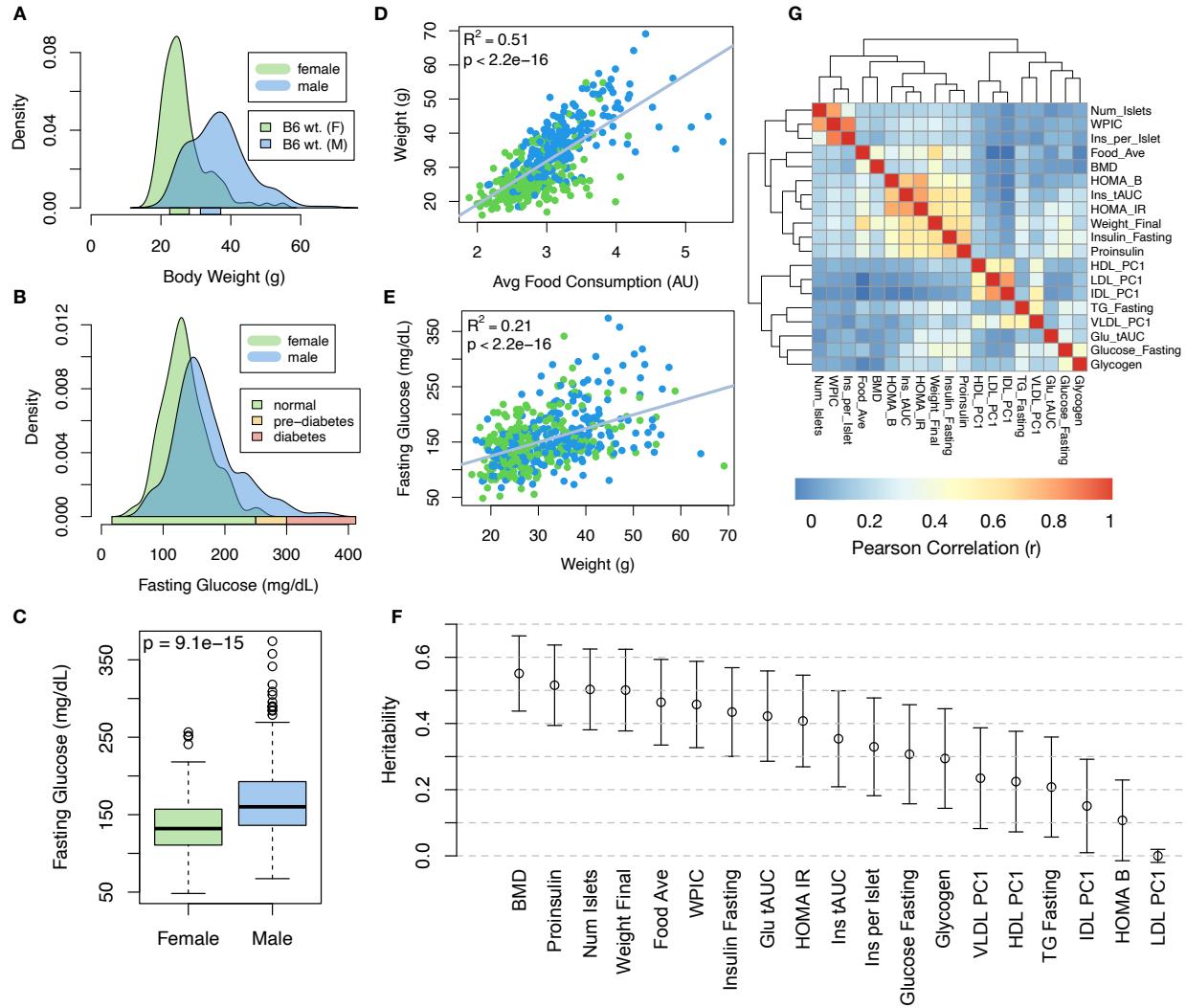


Figure 1: Clinical overview. **A.** Distributions of final body weight in the diversity outbred mice. Sex is indicated by color. The average B6 male and female adult weights at 24 weeks of age are indicated by blue and green bars on the x-axis. **B.** The distribution of final fasting glucose across 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 ($p = 9.1 \times 10^{-15}$). **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.

116 transcript (Fig. 2A).

117 The equal contribution of local and distal genetic variants to the heritability of transcript abundance contrasts
 118 with findings in humans in which local variants have been found to explain 20-30% of total heritability, while

119 distal variants explain the remaining 70-80%^{35;36}. This discrepancy may arise due to the high degree of
 120 linkage disequilibrium in the DO mice compared to human populations. Each genetic marker in the mice
 121 captures information from a larger genomic region than each genetic marker in human populations, and thus
 122 may capture more local regulatory variants than SNPs capture in humans. It has been found that transcripts
 123 with multiple local eQTL have higher local heritability than transcripts with single local eQTL³⁷. Because of
 124 the high diversity in the DO and the high rates of linkage disequilibrium, it is possible that there are more
 125 local variants regulating transcription creating a proportionally larger effect of local regulation.

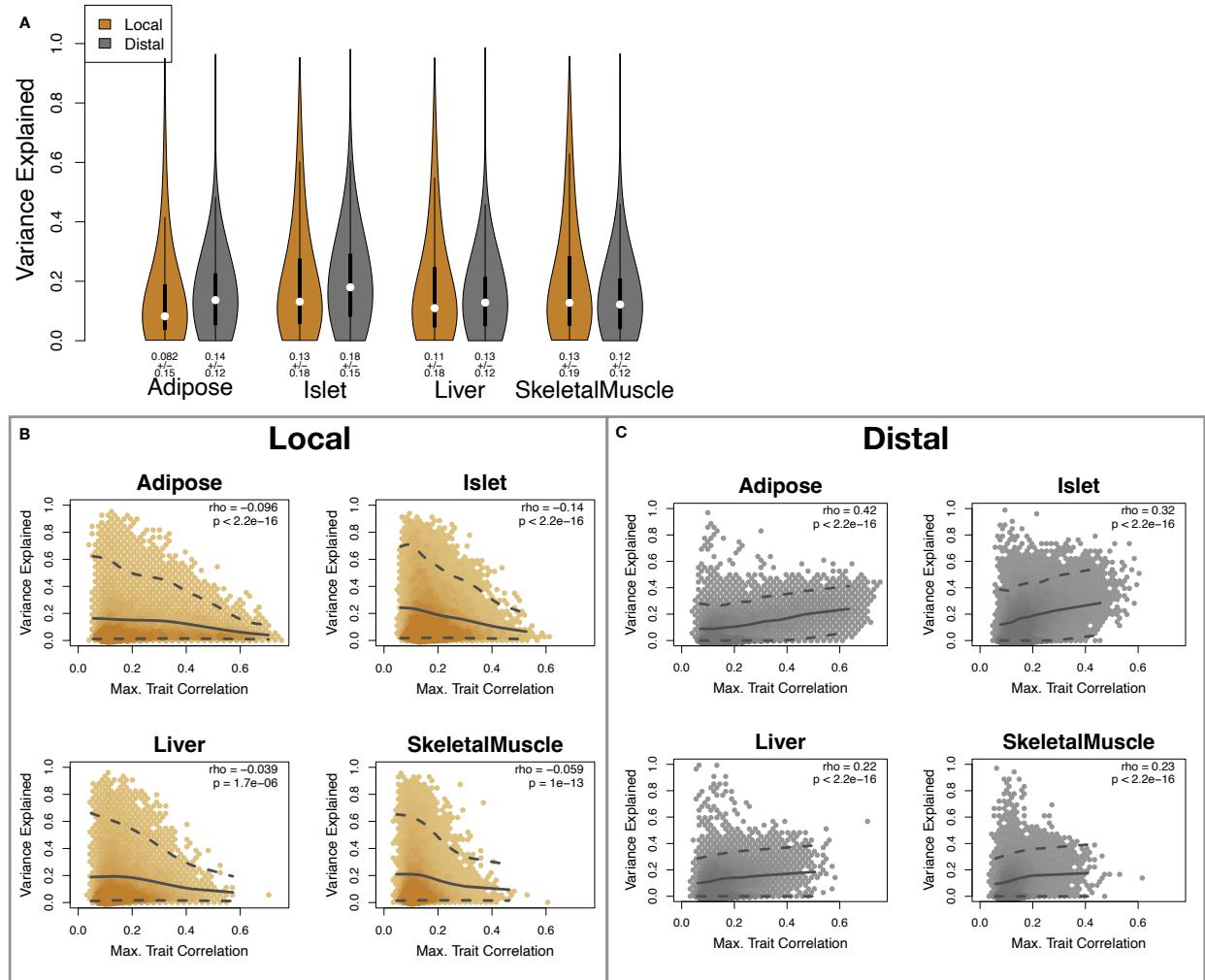


Figure 2: Transcript heritability and trait relevance. **A.** Distributions of local and distal heritability of transcripts across the four tissues. Overall local and distal factors contributed equally to transcript heritability. Arrows indicate the median of each distribution. The relationship between **(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. The upper and lower dashed line in each panel show the 95th and 5th percentile correlation. The solid line shows the mean trait correlation in transcripts with increasing variance explained either locally (B) or distally (C). Transcripts that are highly correlated with traits tended to have low local heritability and high distal heritability.

126 To assess the importance of genetic regulation of transcript levels to clinical traits, we compared the local
127 and distal heritabilities of transcripts to their trait relevance. We defined trait relevance for a transcript as
128 its maximum absolute Spearman correlation coefficient (ρ) across all traits (Methods). The local heritability
129 of transcripts was negatively associated with their trait relevance (Fig. 2B), suggesting that the more
130 local genotype influenced transcript abundance, the less effect this variation had on the measured traits.
131 Conversely, the distal heritability of transcripts was positively associated with trait relevance (Fig. 2C). That
132 is, transcripts that were more highly correlated with the measured traits tended to be distally, rather than
133 locally, heritable. Importantly, this pattern was consistent across all tissues. This finding is consistent with
134 previous observations that transcripts with low local heritability explain more expression-mediated disease
135 heritability than transcripts with high local heritability¹⁹. However, the positive relationship between trait
136 correlation and distal heritability demonstrated further that there are diffuse genetic effects throughout the
137 genome converging on trait-related transcripts.

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

140 The above univariate analyses establish the importance of distal heritability for trait-relevant transcripts.
141 However, the number of transcripts dramatically exceeds the number of phenotypes. Thus, we expect the
142 heritable, trait-relevant transcripts to be highly correlated and organized according to coherent, biological
143 processes representing the mediating endophenotypes driving clinical trait variation. To identify these
144 endophenotypes in a theoretically principled way, we developed a novel dimension-reduction technique,
145 high-dimension mediation analysis (HDMA), that uses the theory of causal graphical models to identify a
146 transcriptomic signature that is simultaneously 1) highly heritable, 2) strongly correlated to the measured
147 phenotypes, and 3) conforms to the causal mediation hypothesis (Fig. 3). In HDMA, we first use a
148 linear mapping called kernelization to dimension-reduce the genome, transcriptome, and phenotype to kernel
149 matrices G_K , T_K and P_K respectively, which each have the dimensions n by n where n is the number
150 of individuals (Methods). These kernel matrices describe the relationships among the individual mice in
151 genome space, transcriptome space, and phenotype space and ensure that these three omic spaces have
152 the same dimensions, and thus the same weight in the analysis. If not dimension-reduced, the transcriptome
153 would outweigh the phenotype in the model. We then projected these $n \times n$ -dimensional kernel matrices
154 onto one-dimensional scores—a composite genome score (G_C), a composite transcriptome score (T_C), and a
155 composite phenotype score (P_C)—and used the univariate theory of mediation to constrain these projections to
156 satisfy the hypotheses of perfect mediation, namely that upon controlling for the transcriptomic score, the
157 genome score is uncorrelated to the phenotype score. A complete mathematical derivation and implementation

158 details for HDMA are available in the Methods.

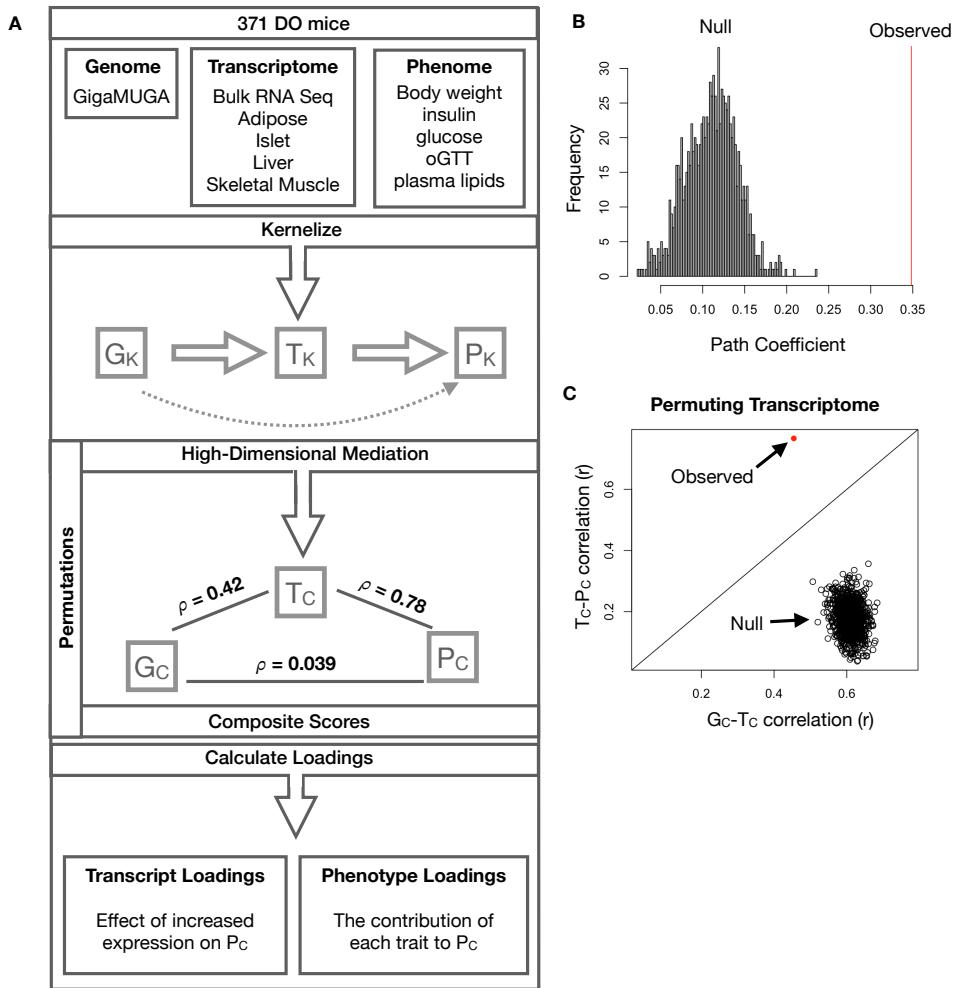


Figure 3: High-dimensional mediation. **A.** Workflow indicating major steps of high-dimensional mediation. The genotype, transcriptome, and phenotype matrices were independently normalized and converted to kernel matrices representing the pairwise relationships between individuals for each data modality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

159 Using HDMA we identified the major axis of variation in the transcriptome that was consistent with mediating
 160 the effects of the genome on metabolic traits (Fig 3). Fig. 3A shows the partial correlations (ρ) between
 161 the pairs of these composite vectors. The partial correlation between G_C and T_C was 0.42, and the partial
 162 correlation between T_C and P_C was 0.78. However, when the transcriptome was taken into account, the

163 partial correlation between G_C and P_C was effectively zero (0.039). P_C captured 30% of the overall trait
164 variance, and its estimated heritability was 0.71 ± 0.084 , which was higher than any of the measured traits
165 (Fig. 1F). Thus, HDMA identified a maximally heritable metabolic composite trait and a highly heritable
166 component of the transcriptome that are correlated as expected in the perfect mediation model.

167 As discussed in the Methods, HDMA is related to a generalized form of canonical correlation analysis (CCA).
168 Standard CCA is prone to over-fitting because in any two large matrices it can be trivial to identify highly
169 correlated composite vectors³⁸. To assess whether our implementation of HDMA was similarly prone to
170 over-fitting in a high-dimensional space, we performed permutation testing. We permuted the individual
171 labels on the transcriptome matrix 10,000 times and recalculated the path coefficient, which is the correlation
172 of G_C and T_C multiplied by the correlation of T_C and P_C . This represents the strength of the path from
173 G_C to P_C that is putatively mediated through T_C . The permutations preserved the correlation between the
174 genome and phenotype, but broke the correlations between the genome and the transcriptome, as well as
175 between the transcriptome and the phenotype. We could thus test whether, given a random transcriptome,
176 HDMA would overfit and identify apparently mediating transcriptomic signatures in random data. The null
177 distribution of the path coefficient is shown in Fig. 3B, and the observed path coefficient from the original
178 data is indicated by a red line. The observed path coefficient was well outside the null distribution generated
179 by permutations ($p < 10^{-16}$). Fig. 3C illustrates this observation in more detail. Although we identified
180 high correlations between G_C and T_C , and modest correlations between T_C and P_C in the null data (Fig 3C),
181 these two values could not be maximized simultaneously in the null data. In contrast, the red dot shows that
182 in the real data both the G_C - T_C correlation and the T_C - P_C correlation could be maximized simultaneously
183 suggesting that the path from genotype to phenotype through the transcriptome is highly non-trivial and
184 identifiable in this case. These results suggest that these composite vectors represent genetically determined
185 variation in phenotype that is mediated through genetically determined variation in transcription.

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

188 Each composite score is a weighted combination of the measured variables. The magnitude and sign of
189 the weights, called loadings, correspond to the relative importance and directionality of each variable in
190 the composite score. The loadings of each measured trait onto P_C indicate how much each contributed
191 to the composite phenotype. Body weight contributed the most (Fig. 4), followed by homeostatic insulin
192 resistance (HOMA_IR) and fasting plasma insulin levels (Insulin_Fasting). We can thus interpret P_C as an
193 index of metabolic disease (Fig. 4B). Individuals with high values of P_C have a higher metabolic disease

194 index (MDI) and greater metabolic disease, including higher body weight and higher insulin resistance. We
 195 refer to P_C as MDI going forward. Traits contributing the least to MDI were measures of cholesterol and
 196 pancreas composition. Thus, when we interpret the transcriptomic signature identified by HDMA, we are
 197 explaining primarily the putative transcriptional mediation of body weight and insulin resistance, as opposed
 198 to cholesterol measurements.

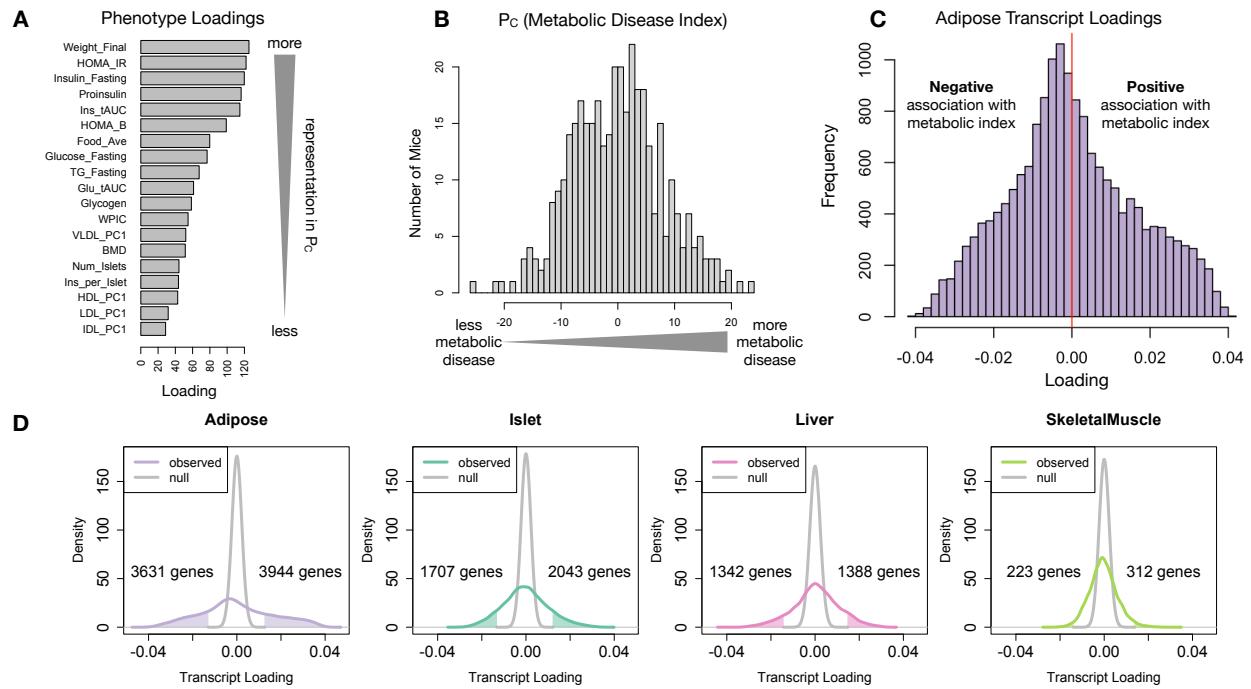


Figure 4: Interpretation of loadings. **A.** Loadings across traits. Body weight and insulin resistance contributed the most to the composite trait. **B.** Phenotype scores (MDI) across individuals. Individuals with large positive MDI had higher body weight and insulin resistance than average. Individuals with large negative MDI 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.** Distributions of loadings across tissues compared to null distributions. Shaded areas represent loadings that were more extreme than the null distribution. Numbers indicate how many transcripts had loadings above and below the extremes of the null. Transcripts in adipose tissue had the most extreme loadings indicating that transcripts in adipose tissue were the best mediators of the genetic effects on body weight and insulin resistance.

**199 High-loading transcripts had low local heritability, high distal heritability, and were linked
200 mechanistically to obesity**

201 We interpreted large loadings onto transcripts as indicating strong mediation of the effect of genetics on MDI.
202 Large positive loadings indicate that higher expression was associated with a higher MDI (i.e. higher risk of
203 obesity and metabolic disease on the HFHS diet) (Fig. 4C). Conversely, large negative loadings indicate that
204 high expression of these transcripts was associated with a lower MDI (i.e. lower risk of obesity and metabolic
205 disease on the HFHS diet) (Fig. 4C). We used gene set enrichment analysis (GSEA)^{39;40} to look for biological
206 processes and pathways that were enriched at the top and bottom of this list (Methods).

207 In adipose tissue, both GO processes and KEGG pathway enrichments pointed to an axis of inflammation
208 and metabolism (Figs. S4 and S5). GO terms and KEGG pathways associated with inflammation were
209 positively associated with MDI, indicating that increased expression in inflammatory pathways was associated
210 with a higher burden of disease. It is well established that adipose tissue in obese individuals is inflamed
211 and infiltrated by macrophages^{41–45}, and the results here suggest that this may be a dominant heritable
212 component of metabolic disease.

213 The strongest negative enrichments in adipose tissue were related to mitochondrial activity in general, and
214 thermogenesis in particular (Figs. S4 and S5). Genes in the KEGG oxidative phosphorylation pathway were
215 almost universally negatively loaded in adipose tissue, suggesting that increased expression of these genes was
216 associated with reduced MDI (Supp. Fig. S6). Consistent with this observation, it has been shown previously
217 that mouse strains with greater thermogenic potential are also less susceptible to obesity on an obesigenic
218 diet⁴⁶.

219 Transcripts associated with the citric acid cycle as well as the catabolism of the branched-chain amino acids
220 (valine, leucine, and isoleucine) were strongly enriched with negative loadings in adipose tissue (Supp. Figs.
221 S4, S7 and S8). Expression of genes in both pathways (for which there is some overlap) has been previously
222 associated with insulin sensitivity^{12;47;48}, suggesting that heritable variation in regulation of these pathways
223 may influence risk of insulin resistance.

224 Looking at the 10 most positively and negatively loaded transcripts from each tissue, it is apparent that
225 transcripts in the adipose tissue had the largest loadings, both positive and negative (Fig. 5A bar plot). This
226 suggests that much of the effect of genetics on body weight and insulin resistance is mediated through gene
227 expression in adipose tissue. This finding does not speak to the relative importance of tissues not included
228 in this study, such as brain, in which transcriptional variation may mediate a large portion of the genetic
229 effect on obesity. The strongest 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. As expected, heritability analysis showed that transcripts with the largest loadings had higher distal heritability than local heritability (Fig. 5A heat map and box plot). We also performed TWAS in this population by imputing transcript levels for each gene based on local genotype only and correlating the imputed transcript levels with each trait. In contrast to HDMA, the TWAS procedure tended to nominate transcripts with lower loadings (Fig. 5B), higher local heritability and lower distal heritability. Finally, we focused on transcripts with the highest local heritability in each tissue (Fig. 5C). This procedure selected transcripts with low loadings on average, consistent with our findings above (Fig. 2B).

We performed a literature search for the genes in each of these groups along with the terms “diabetes”, “obesity”, and the name of the expressing tissue to determine whether any of these genes had previous associations with metabolic disease in the literature (Methods). Multiple genes in each group had been previously associated with obesity and diabetes (Fig. 5 bolded gene names). Genes with high loadings were most highly enriched for previous literature support. They were 2.1 times more likely than TWAS hits and 3.8 times more likely than genes with high local heritability to be previously associated with obesity or diabetes.

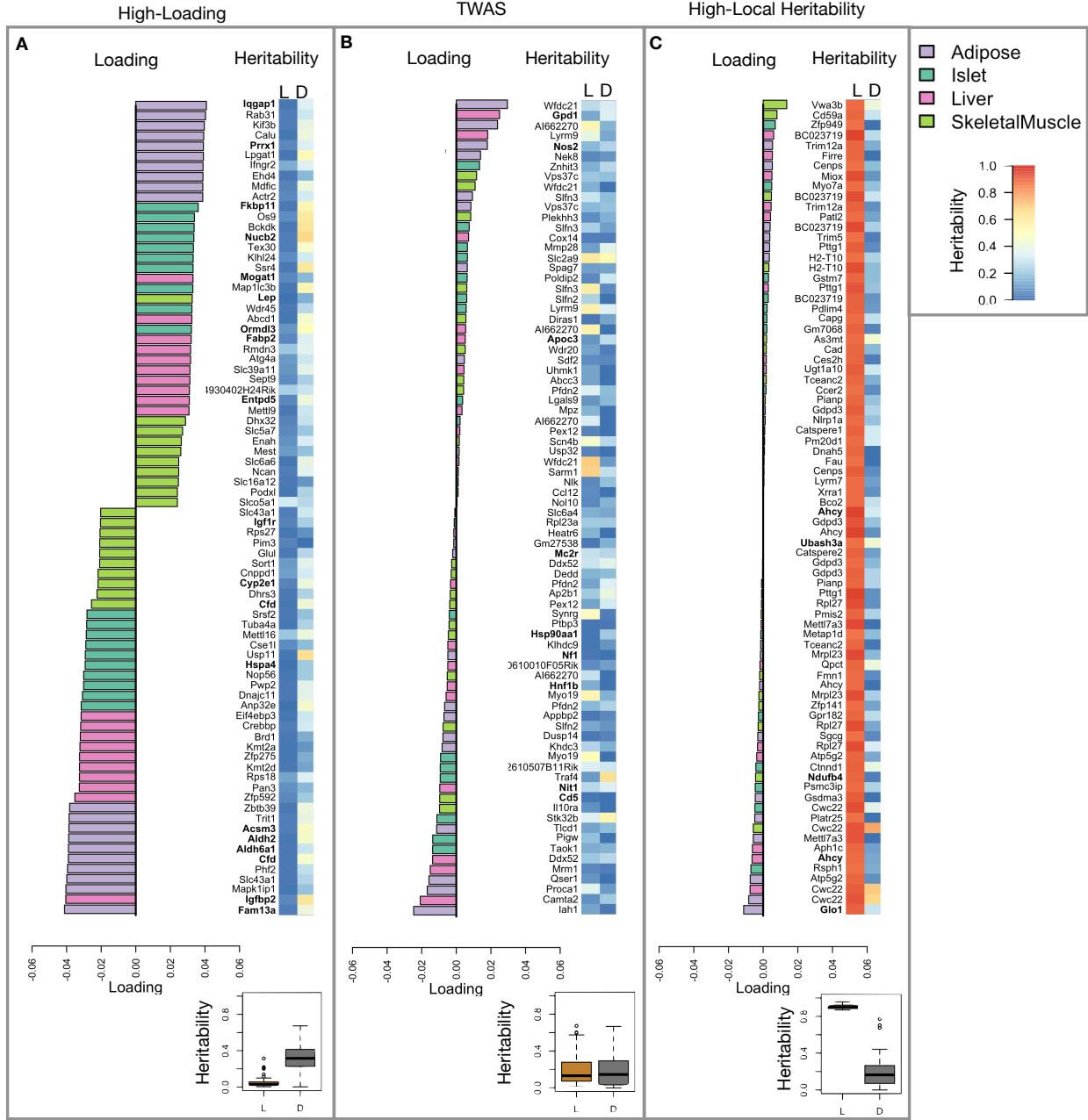


Figure 5: Transcripts with high loadings have high distal heritability and literature support (bolded gene names). Each panel has a bar plot showing the loadings of transcripts selected by different criteria. Bar color indicates the tissue of origin. The heat map shows the local (L - left) and distal (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. Distal heritability was significantly higher than local heritability (t-test $p < 2.2^{-16}$). **B.** Loadings of TWAS candidates with the 10 largest positive correlations with traits and the largest negative correlations with traits across all four tissues. Local and distal heritability were not significantly different for this group (t-test $p = 0.77$). **C.** The transcripts with the largest local heritability (top 20) across all four tissues. Local heritability was significantly higher than distal heritability of these genes (t-test $p < 2.2^{-16}$)

244 Tissue-specific transcriptional programs were associated with metabolic traits

245 Clustering of transcripts with top loadings in each tissue showed tissue-specific functional modules associated
246 with obesity and insulin resistance (Fig. 6A) (Methods). The clustering highlights the importance of immune
247 activation particularly in adipose tissue. The “mitosis” cluster had large positive loadings in three of the four
248 tissues potentially suggesting system-wide proliferation of immune cells. Otherwise, all clusters were strongly
249 loaded in only one or two tissues. For example, the lipid metabolism cluster was loaded most heavily in liver.
250 The positive loadings suggest that high expression of these genes, particularly in the liver, was associated with
251 increased metabolic disease. This cluster included the gene *Pparg*, whose primary role is in the adipose tissue
252 where it is considered a master regulator of adipogenesis⁴⁹. Agonists of *Pparg*, such as thiazolidinediones, are
253 FDA-approved to treat type II diabetes, and reduce inflammation and adipose hypertrophy⁴⁹. Consistent
254 with this role, the loading for *Pparg* in adipose tissue was negative, suggesting that higher expression was
255 associated with leaner mice (Fig. 6B). In contrast, *Pparg* had a large positive loading in liver, where it is
256 known to play a role in the development of hepatic steatosis, or fatty liver. Mice that lack *Pparg* specifically
257 in the liver, are protected from developing steatosis and show reduced expression of lipogenic genes^{50;51}.
258 Overexpression of *Pparg* in the livers of mice with a *Ppara* knockout, causes upregulation of genes involved in
259 adipogenesis⁵². In the livers of both mice and humans high *Pparg* expression is associated with hepatocytes
260 that accumulate large lipid droplets and have gene expression profiles similar to that of adipocytes^{53;54}.
261 The local and distal heritability of *Pparg* is low in adipose tissue suggesting its expression in this tissue is
262 highly constrained in the population (Fig. 6B). However, the distal heritability of *Pparg* in liver is relatively
263 high suggesting it is complexly regulated and has sufficient variation in this population to drive variation in
264 phenotype. Both local and distal heritability of *Pparg* in the islet are relatively high, but the loading is low,
265 suggesting that variability of expression in the islet does not drive variation in MDI. These results highlight
266 the importance of tissue context when investigating the role of heritable transcript variability in driving
267 phenotype. Gene lists for all clusters are available in Supp. File 1.

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

269 To test whether the transcript loadings identified in the DO could be translated to another population, we
270 tested whether they could predict metabolic phenotypes in an independent population of CC-RIX mice, which
271 were F1 mice derived from multiple pairings of Collaborative Cross (CC)^{55;32;56;57} strains (Fig. 7) (Methods).
272 We tested two questions. First, we asked whether the loadings identified in the DO mice were relevant to
273 the relationship between the transcriptome and the phenotype in the CC-RIX. We predicted body weight
274 (a surrogate for MDI) in each CC-RIX individual using measured gene expression in each tissue and the

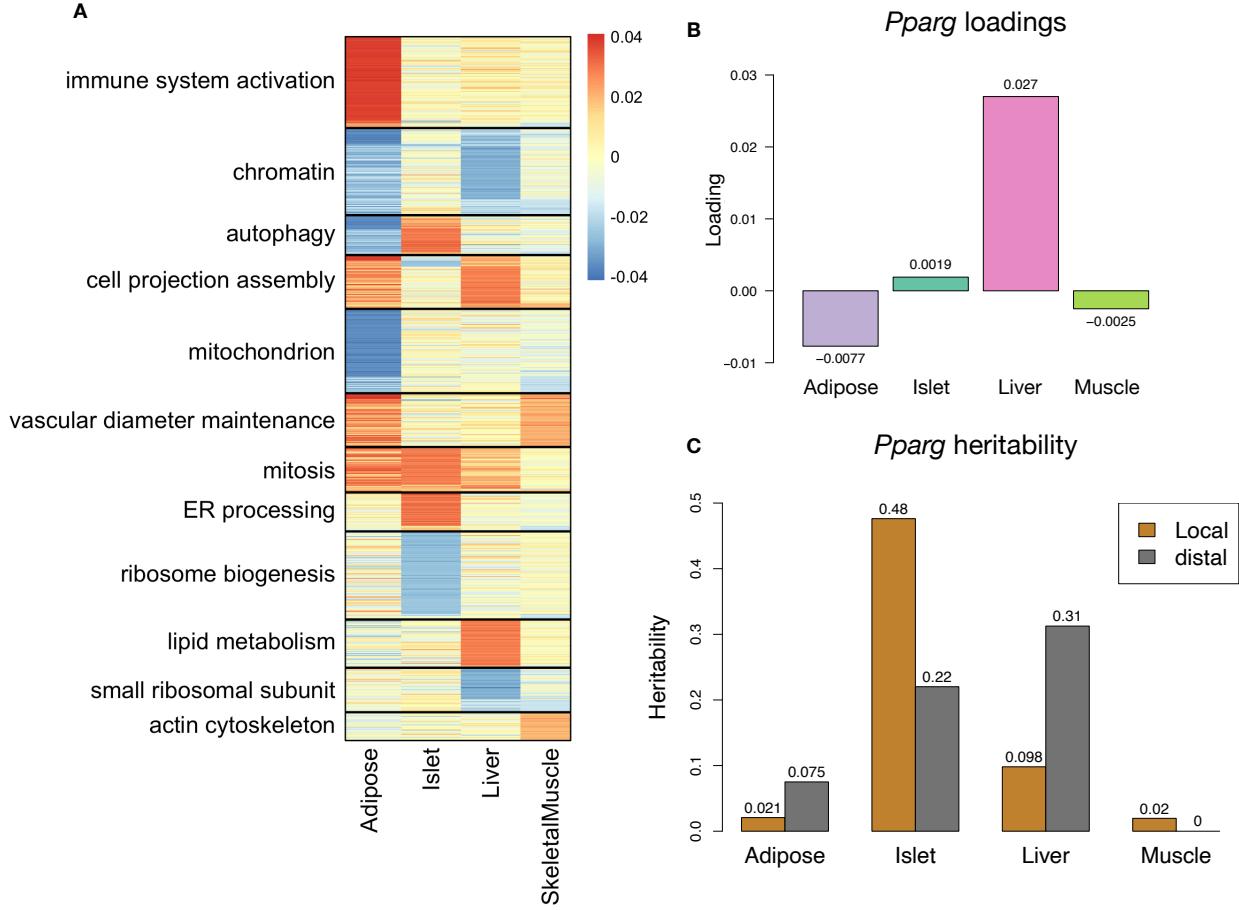


Figure 6: Tissue-specific transcriptional programs were associated with obesity and insulin resistance. **A** Heat map showing the loadings of all transcripts with loadings greater than 2.5 standard deviations from the mean in any tissue. The heat map was clustered using k medoid clustering. Functional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

275 transcript loadings identified in the DO (Methods). The predicted body weight and acutal body weight were
 276 highly correlated (Fig. 7B left column). The best prediction was achieved for adipose tissue, which supports
 277 the observation in the DO that adipose expression was the strongest mediator of the genetic effect on MDI.
 278 This result also confirms the validity and translatability of the transcript loadings and their relationship to
 279 metabolic disease.

280 The second question related to the source of the relevant variation in gene expression. If local regulation was
 281 the predominant factor influencing trait-relevant gene expression, we should be able to predict phenotype in
 282 the CC-RIX using transcripts imputed from local genotype (Fig. 7A). The DO and the CC-RIX were derived
 283 from the same eight founder strains and so carry the same alleles throughout the genome. We imputed gene
 284 expression in the CC-RIX using local genotype and were able to estimate variation in gene transcription

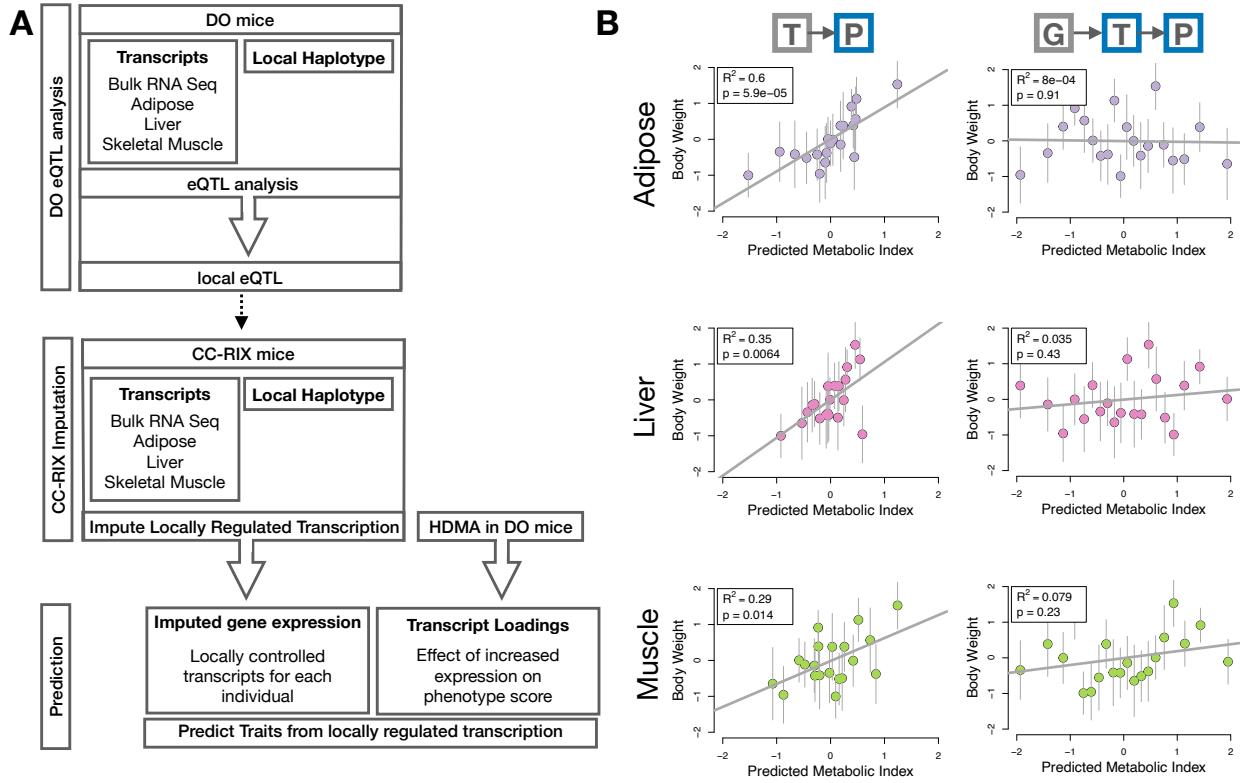


Figure 7: Transcription, but not local genotype, predicts phenotype in the CC-RIX. **A.** Workflow showing procedure for translating HDMA results to an independent population of mice. **B.** Relationships between the predicted metabolic disease index (MDI) and measured body weight in the CC-RIX. The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

robustly (Supp. Fig. S9). 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 gene expression is not the primary factor driving heritability of complex traits. It is also consistent with our findings in the DO population that distal heritability was a major driver of trait-relevant gene expression and that high-loading transcripts had comparatively high distal and low local heritability.

290 Distally heritable transcriptomic signatures reflected variation in composition of adipose tissue 291 and islets

The interpretation of global genetic influences on gene expression and phenotype is potentially more challenging than the interpretation and translation of local genetic influences, as genetic effects cannot be localized to individual gene variants or transcripts. However, there are global patterns across the loadings that can inform mechanism. For example, heritable variation in cell type composition can be inferred from transcript loadings.

296 We observed above that immune activation in the adipose tissue was a highly enriched process correlating
 297 with obesity in the DO population. In humans, it has been extensively observed that macrophage infiltration
 298 in adipose tissue is a marker of obesity and metabolic disease⁵⁸. To determine whether the immune activation
 299 reflected a heritable change in cell composition in adipose tissue in DO mice, we compared loadings of
 300 cell-type specific genes in adipose tissue (Methods). The mean loading of macrophage-specific genes was
 301 significantly greater than 0 (Holm-adjusted two-sided empirical $p < 2 \times 10^{-16}$) (Fig. 8A), indicating that
 302 obese mice were genetically predisposed to have high levels of macrophage infiltration in adipose tissue in
 303 response to the HFHS diet. Loadings for marker genes for other cell types were not statistically different
 304 from zero (Adipocytes: $p = 0.08$, Progenitors: $p = 0.58$, Leukocytes: $p = 0.28$; all Holm-adjusted two-sided
 305 empirical p), indicating that changes in the abundance of those cell types was not a mediator of MDI.

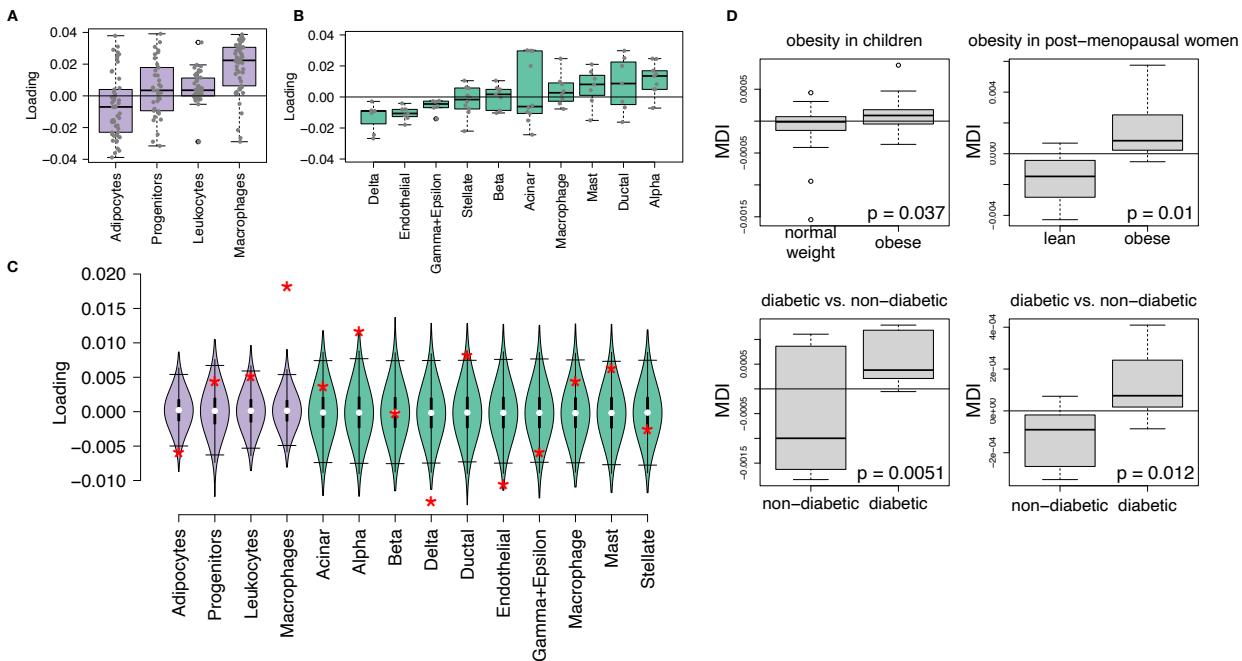


Figure 8: HDMA results translate to humans. **A.** Distribution of loadings for cell-type-specific transcripts in adipose tissue. **B.** Distribution of loadings for cell-type-specific transcripts in pancreatic islets. **C.** Null distributions for the mean loading of randomly selected transcripts in 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 MDI than the lean/non-diabetic patients based on the HDMA results from DO mice.

306 We also compared loadings of cell-type specific transcripts in islet (Methods). The mean loadings for alpha-cell
 307 specific transcripts were significantly greater than 0 ($p = 0.002$), while the mean loadings for delta- (Holm-
 308 adjusted two-sided empirical $p < 2 \times 10^{-16}$) and endothelial-cell (Holm-adjusted two-sided empirical $p = 0.01$)
 309 specific genes were significantly less than 0 (Fig. 8B). These results suggest that mice with higher MDI

310 inherited an altered cell composition that predisposed them to metabolic disease, or that these compositional
311 changes were induced by the HFHS diet in a heritable way. In either case, these results support the hypothesis
312 that alterations in islet composition drive variation in MDI. Notably, the mean loading for pancreatic beta
313 cell marker transcripts was not significantly different from zero (Holm-adjusted two-sided empirical $p = 0.95$).
314 We stress that this is not necessarily reflective of the function of the beta cells in the obese mice, but rather
315 suggests that any variation in the number of beta cells in these mice was unrelated to obesity and insulin
316 resistance, the major contributors to MDI. This is further consistent with the islet composition traits having
317 small loadings in the phenome score (Fig. 4).

318 **Heritable transcriptomic signatures translated to human disease**

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

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

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

331 Another application of the transcript loading landscape is in ranking potential drug candidates for the
332 treatment of metabolic disease. Although high-loading transcripts may be good candidates for understanding
333 specific biology related to obesity, the transcriptome overall is highly interconnected and redundant. The
334 ConnectivityMap (CMAP) database^{59;60} developed by the Broad Institute allows querying thousands of
335 compounds that reverse or enhance the extreme ends of transcriptomic signatures in multiple different cell
336 types. By identifying drugs that reverse pathogenic transcriptomic signatures, we can potentially identify
337 compounds that have favorable effects on gene expression. To test this hypothesis, we queried the CMAP
338 database through the CLUE online query tool (<https://clue.io/query/>, version 1.1.1.43) (Methods). We
339 identified top anti-correlated hits across all cell types (Supp. Figs S10 and S11). To get more tissue-specific

340 results, we also looked at top results in cell types that most closely resembled our tissues. We looked at
341 results in adipocytes (ASC) as well as pancreatic tumor cells (YAPC) regardless of *p* value (Supp. Figs S12
342 and S13).

343 The CMAP database identified both known diabetes drugs (e.g. sulfonylureas), as well as drugs that target
344 pathways known to be involved in diabetes pathogenesis (e.g. mTOR inhibitors). These findings help
345 support the mediation model we fit here. Although the composite variables we identified here are consistent
346 with mediation, they do not prove causality. However, the results from CMAP suggest that reversing the
347 transcriptomic signatures we found also reverses metabolic disease phenotypes, which supports a causal role
348 of the transcript levels in driving pathogenesis of metabolic disease. These results thus support the mediation
349 model we identified here and its translation to therapies in human disease.

350 Discussion

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

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;61;62}. It has been observed
366 that transcripts with high local heritability explain less expression-mediated disease heritability than those
367 with low local heritability¹⁹. Consistent with this observation, genes located near GWAS hits tend to be
368 complexly regulated¹⁸. They also tend to be enriched with functional annotations, in contrast to genes with
369 simple local regulation, which tend to be depleted of functional annotations suggesting they are less likely
370 to be directly involved in disease traits¹⁸. These observations are consistent with principles of robustness

371 in complex systems in which simple regulation of important elements leads to fragility of the system^{63–65}.
372 Our results are consistent, instead, with a more complex picture where genes whose expression can drive
373 trait variation are buffered from local genetic variation but are extensively influenced indirectly by genetic
374 variation in the 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 genome⁶⁶. 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, Supp. Fig S3).

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 map⁶⁷.

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 sizes⁶¹. 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 release^{68–70}. Although its transcription was highly heritable in islets, that
392 regulation was distributed across the genome, with no clear distal eQTL (Supp. Fig. S14). 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 tissue⁴⁹ where it was negatively loaded, and
397 harmful in the liver^{50–54}, 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 hypothesis
405 that HDMA identified transcripts that truly mediate genetic effects on traits. On its own, HDMA identifies
406 transcriptional patterns that are consistent with a mediation model, but alone does not prove mediation.
407 However, the finding that these drugs act both on the transcriptional patterns and on the desired traits
408 support the mediation model and the hypothesis that these transcripts have a causal role in pathogenesis of
409 metabolic disease.

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

416 Data and Code Availability

417 **DO mice:** Genotypes, phenotypes, and pancreatic islet gene expression data were previously published¹².
418 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
419 - GSE266569; DO skeletal muscle - GSE266567. Expression data with calculated eQTLs are available at
420 Figshare https://figshare.com/articles/dataset/Data_and_code_for_High-Dimensional_Mediation_Analysis_HDMA_in_diversity_outbred_mice/27066979 DOI: 10.6084/m9.figshare.27066979
421
422 10.6084/m9.figshare.27066979.v1

424 **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 -
425 GSE237743; CC-RIX skeletal muscle - GSE237747. Count matrices and phenotype data can be found at
426 Figshare https://figshare.com/articles/dataset/Data_and_code_for_High-Dimensional_Mediation_Analysis_HDMA_in_diversity_outbred_mice/27066979 DOI: 10.6084/m9.figshare.27066979
427
428

429 **Code:** All code used to run the analyses reported here are available at Figshare: https://figshare.com/articles/dataset/Data_and_code_for_High-Dimensional_Mediation_Analysis_HDMA_in_diversity_outbred_mice/27066979 DOI: 10.6084/m9.figshare.27066979
430
431

⁴³² **Acknowledgements**

⁴³³ This project was supported by The Jackson Laboratory Cube Initiative, as well as grants from the National
⁴³⁴ Institutes of Health (grant numbers.: R01DK101573, R01DK102948, and RC2DK125961) (to A. D. A.) and
⁴³⁵ by the University of Wisconsin-Madison, Department of Biochemistry and Office of the Vice Chancellor for
⁴³⁶ Research and Graduate Education with funding from the Wisconsin Alumni Research Foundation (to M. P.
⁴³⁷ K.).

⁴³⁸ We thank the following scientific services at The Jackson Laboratory: Genome Technologies for the RNA
⁴³⁹ sequencing, necropsy services for the tissue harvests, and the Center for Biometric Analysis for metabolic
⁴⁴⁰ phenotyping.

441 **Supplemental Figures**

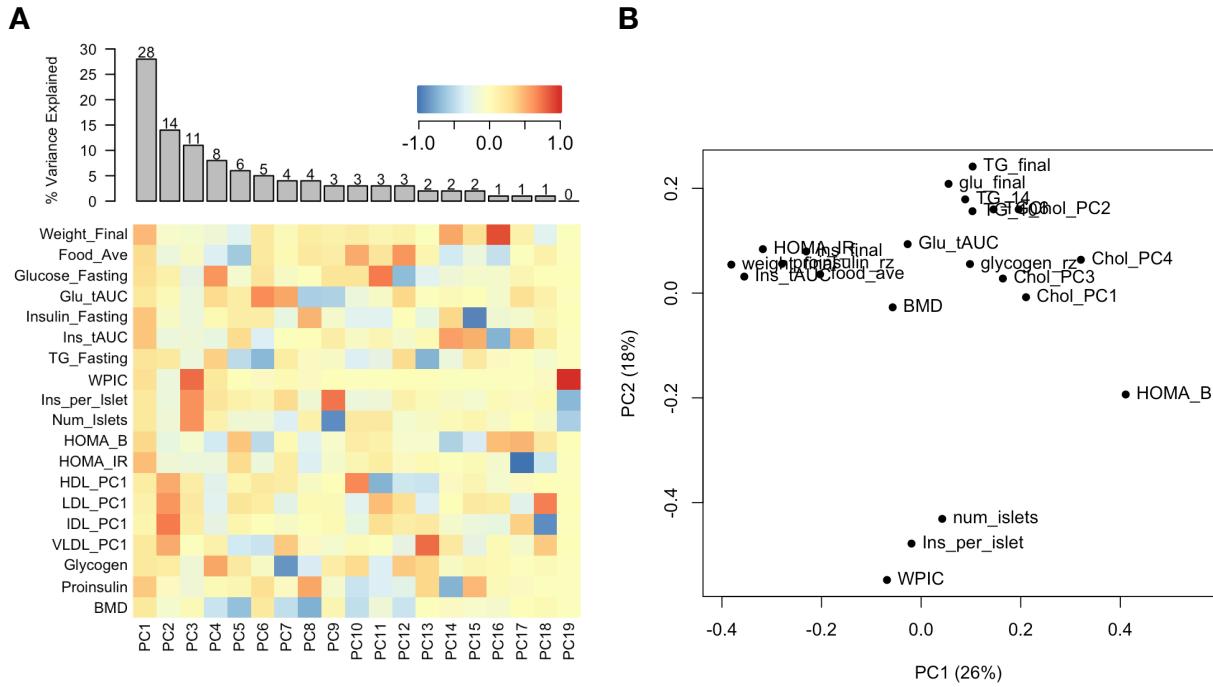


Figure S1: Trait matrix decomposition. **A** The heat map shows the loadings of each trait onto each principal component of the trait matrix. The bars at the top show the percent variance explained for each principal component. **B** Traits plotted by the first and second principal components of the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

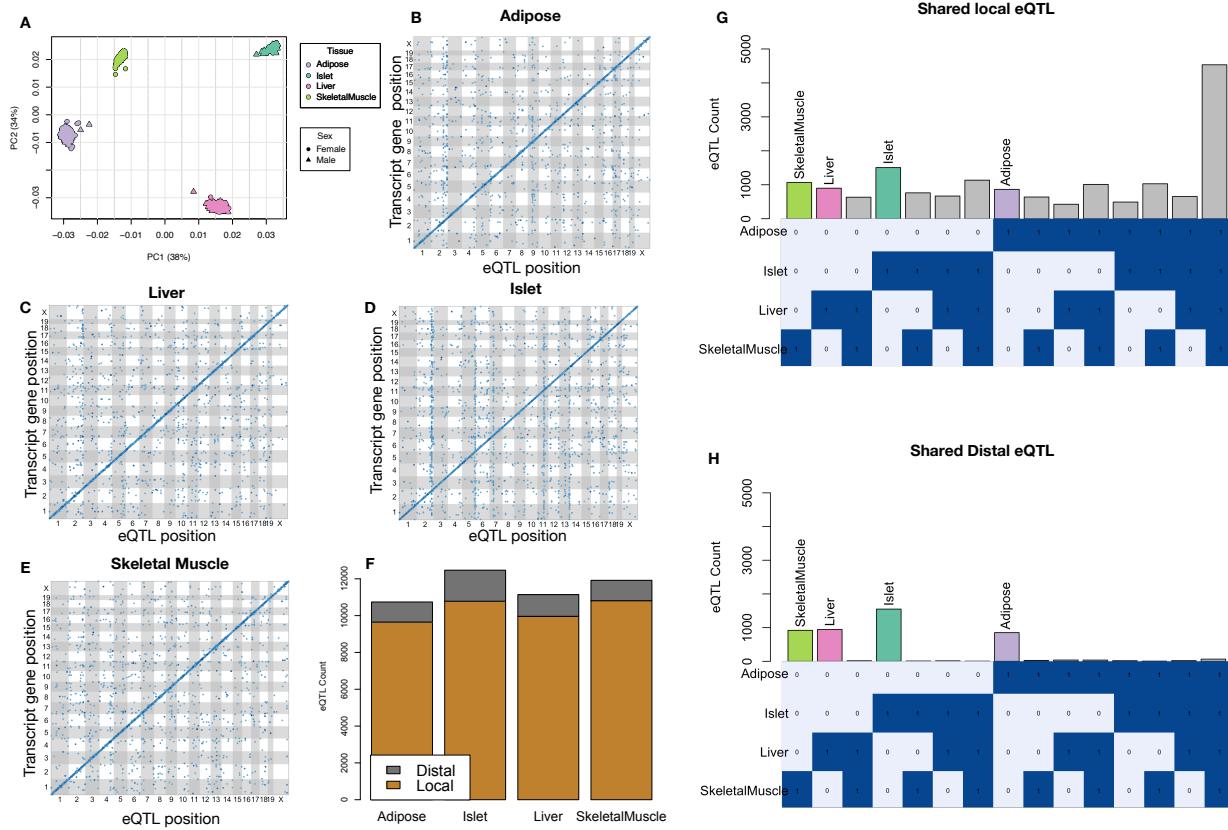


Figure S2: Overview of eQTL analysis in DO mice. **A.** RNA seq samples from the four different tissues clustered by tissue. **B.-E.** eQTL maps are shown for each tissue. The *x*-axis shows the position of the mapped eQTL, and the *y*-axis shows the physical position of the gene encoding each mapped 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.

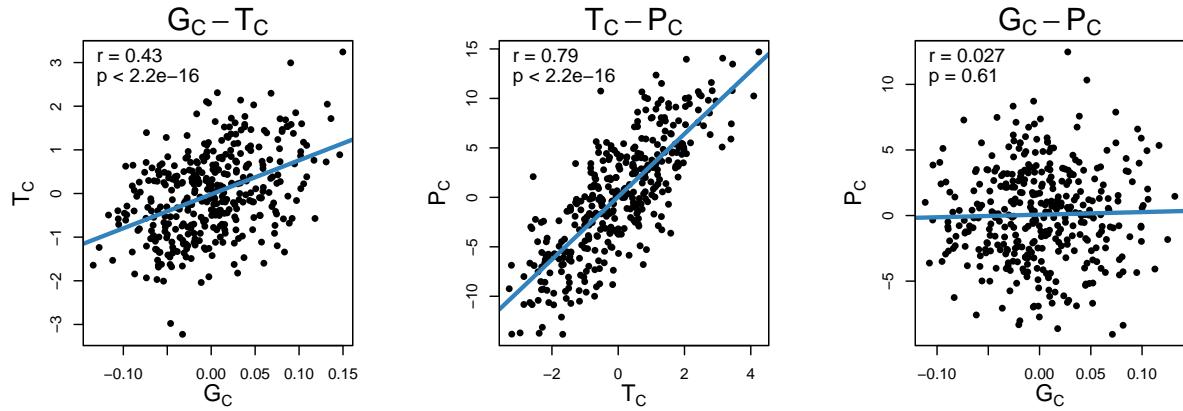


Figure S3: Scatter plots showing correlations between composite vectors for the genome (G_C), the transcriptome (T_C), and the phenotype (P_C). The $G_C - T_C$ correlation is high, the $T_C - P_C$ correlation is high, and there is no significant correlation between G_C and P_C . This correlation structure is consistent with perfect mediation.

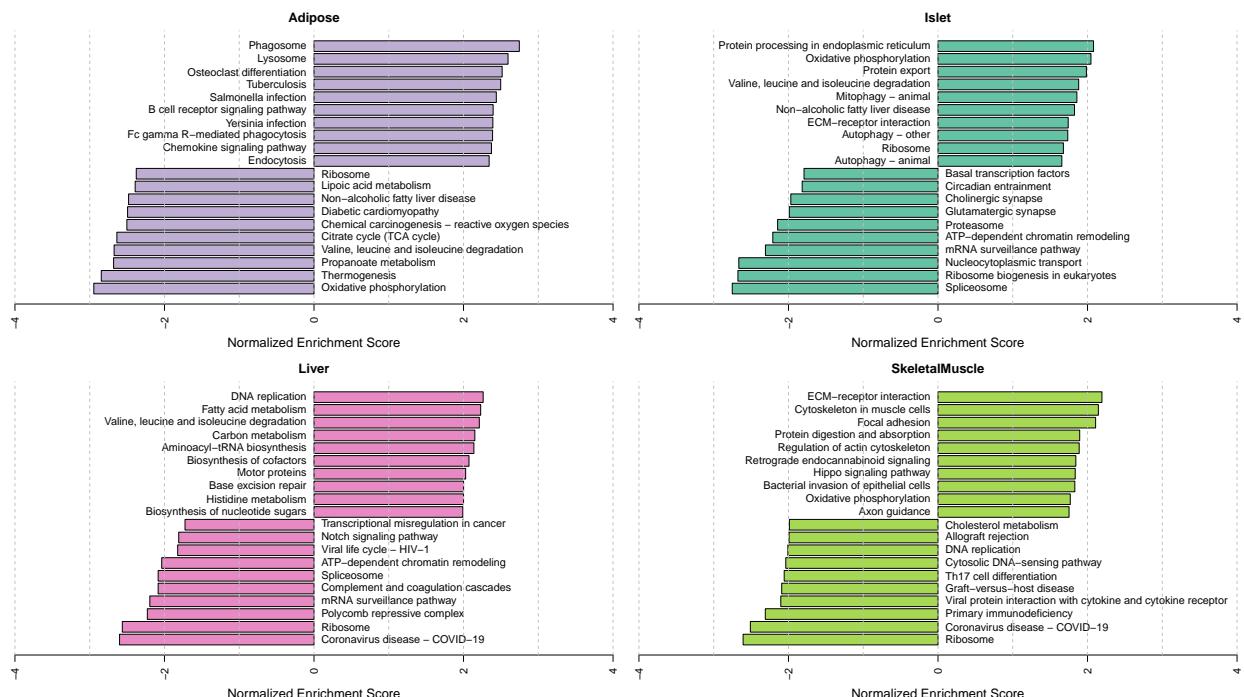


Figure S4: Bar plots showing normalized enrichment scores (NES) for KEGG pathways as determined by fast gene score enrichment analysis (fgsea). Only the top 10 positive and top 10 negative scores are shown. Colors indicate tissue. The name beside each bar shows the name of each enriched KEGG pathway.

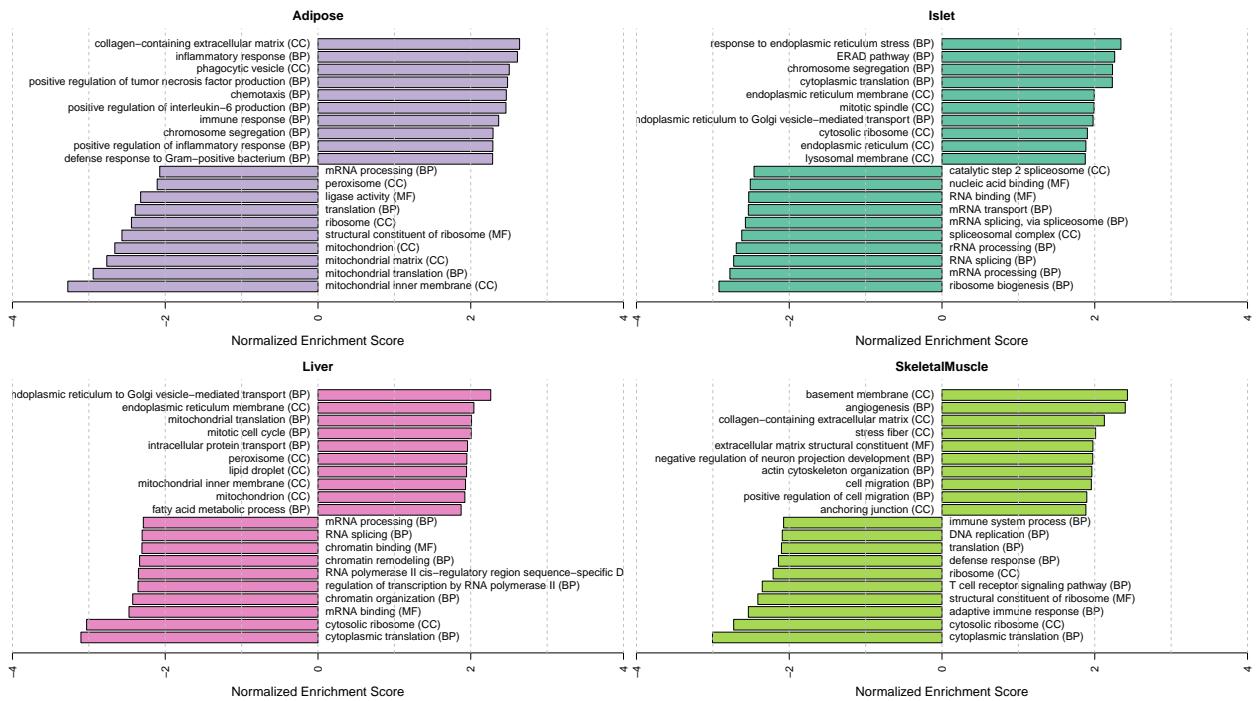


Figure S5: Bar plots showing normalized enrichment scores (NES) for GO terms as determined by fast gene score enrichment analysis (fgsea). Only the top 10 positive and top 10 negative scores are shown. Colors indicate tissue. The name beside each bar shows the name of each enriched GO term. The letters in parentheses indicate whether the term is from the biological process ontology (BP), the molecular function ontology (MF), or the cellular compartment ontology (CC).

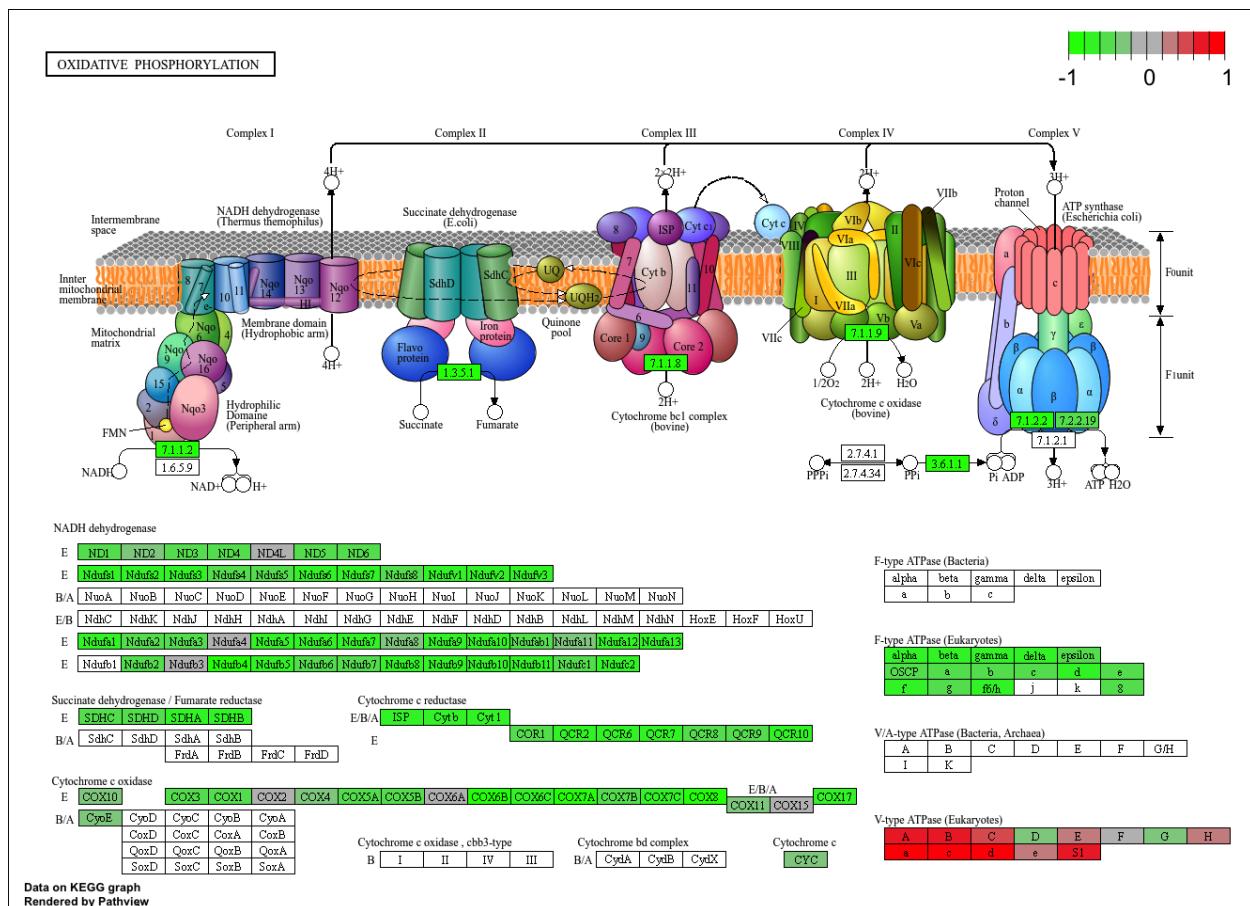


Figure S6: The KEGG pathway for oxidative phosphorylation in mice. Each element is colored based on its HDMA loading from adipose tissue scaled to run from -1 to 1. Genes highlighted in green had negative loadings, and those highlighted in red had positive loadings. Almost the entire pathway was strongly negatively loaded indicating that increased expression of genes involved in oxidative phosphorylation was associated with reduced MDI.

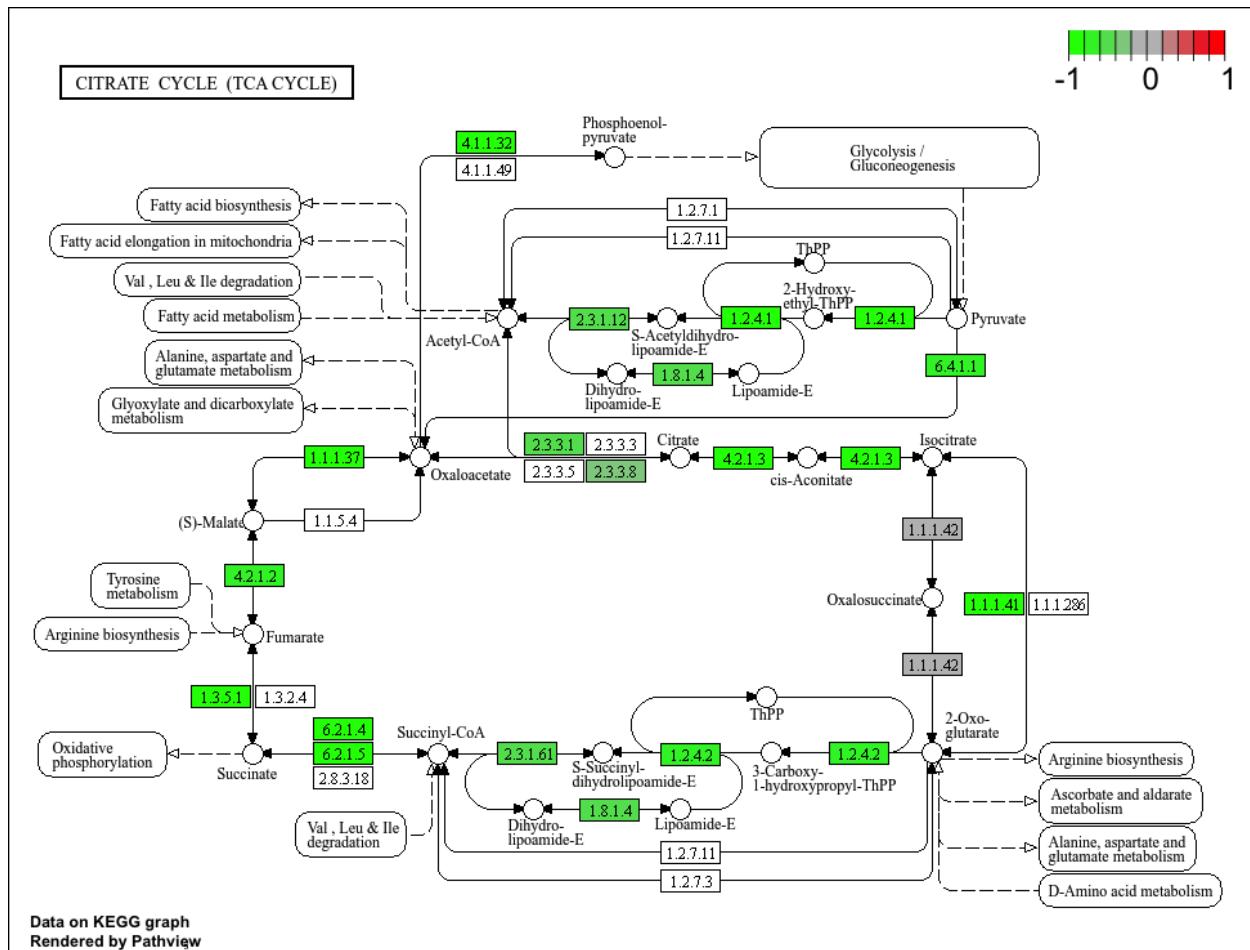


Figure S7: The KEGG pathway for the TCA (citric acid) cycle in mice. Each element is colored based on its HDMA loading from adipose tissue scaled to run from -1 to 1. Genes highlighted in green had negative loadings, and those highlighted in red had positive loadings. Many genes in the cycle were strongly negatively loaded indicating that increased expression of genes involved in the TCA cycle was associated with reduced MDI.

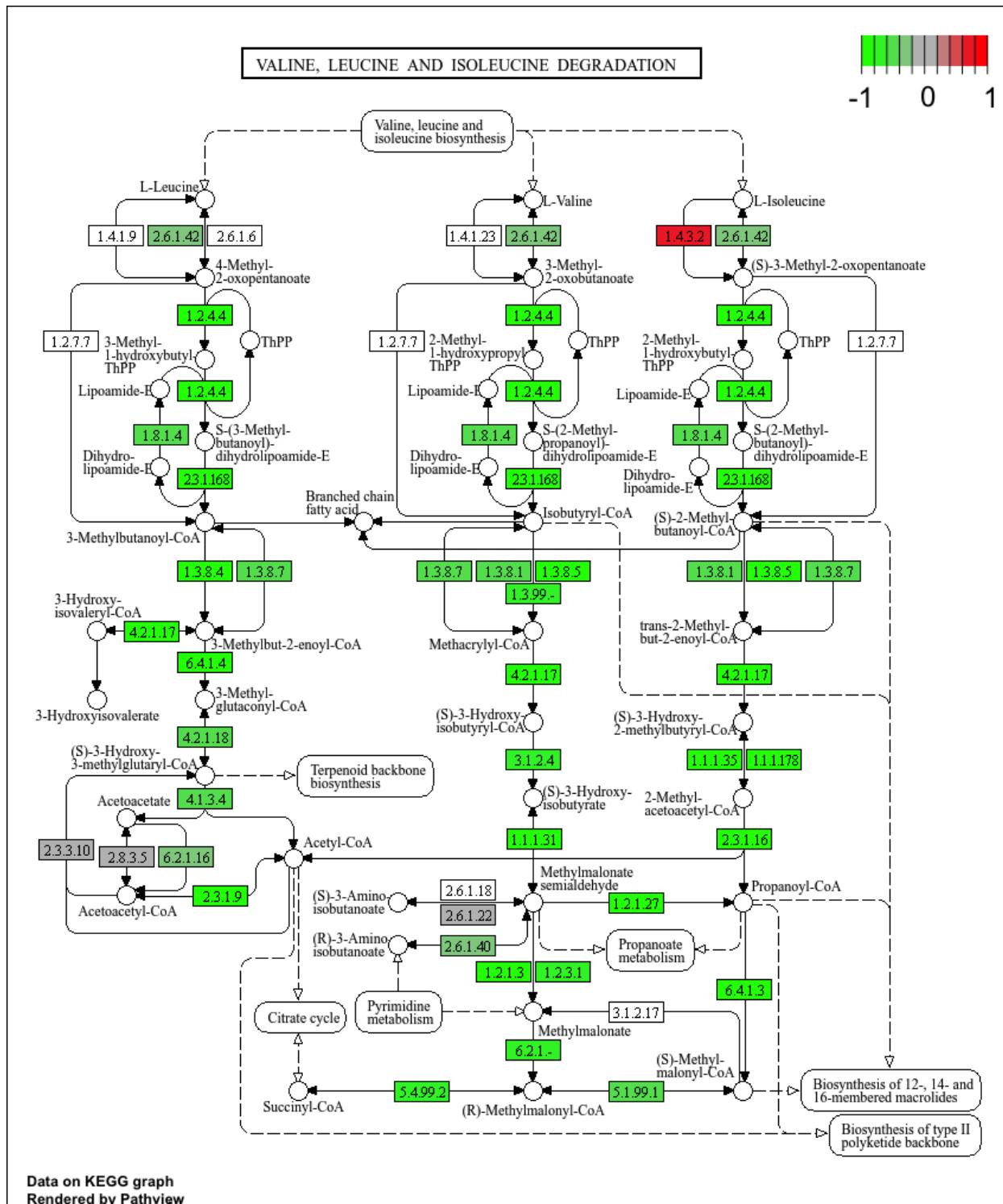


Figure S8: The KEGG pathway for branched-chain amino acid degradation in mice. Each element is colored based on its HDMA loading from adipose tissue scaled to run from -1 to 1. Genes highlighted in green had negative loadings, and those highlighted in red had positive loadings. Almost the entire pathway was strongly negatively loaded indicating that increased expression of genes involved in branched-chain amino acid degradation was associated with reduced MDI.

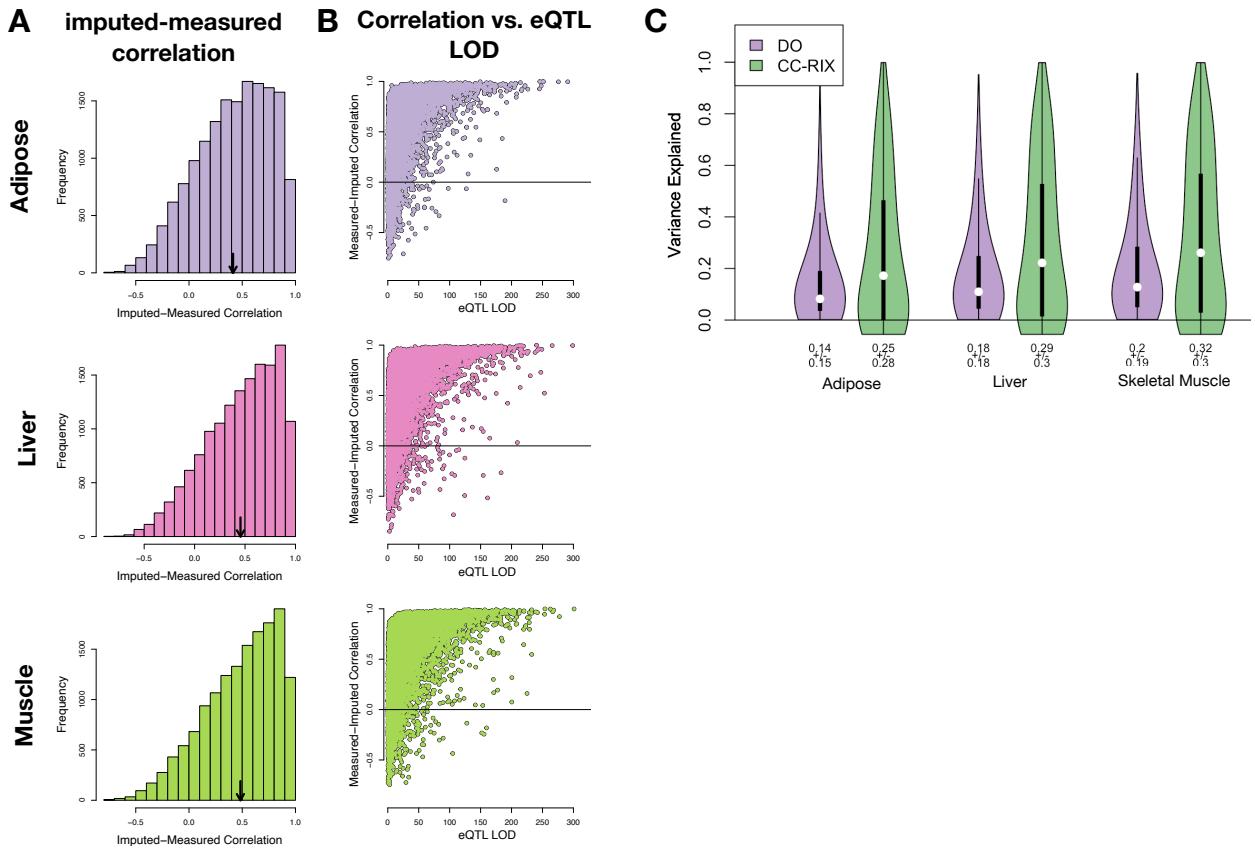


Figure S9: Validation of transcript imputation in the CC-RIX. **A.** Distributions of correlations between imputed and measured transcripts in the CC-RIX. The mean of each distribution is shown by the red line. All distributions were skewed toward positive correlations and had positive means near a Pearson 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.** Distributions of variance explained by local genotype across all transcripts 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 S10: 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 S11: 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 S12: 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 S13: 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.

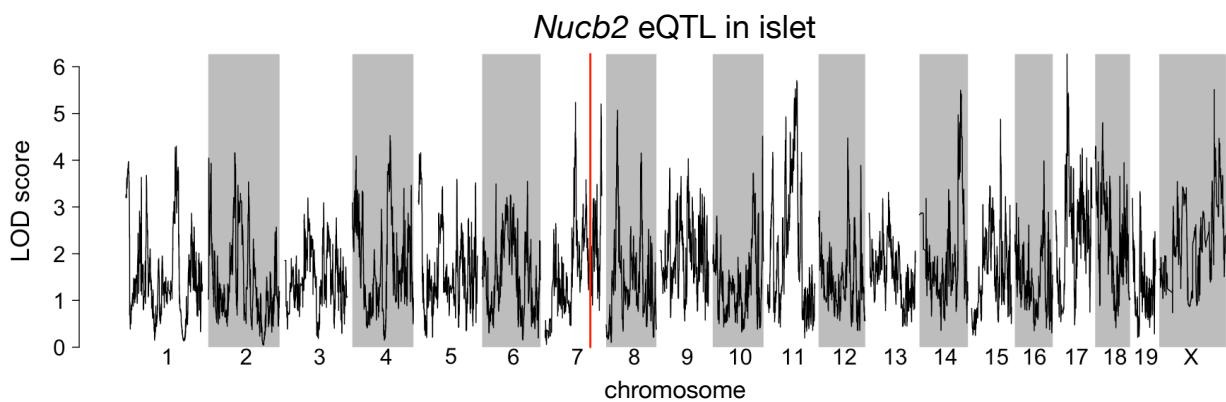


Figure S14: Regulation of *Nucb2* expression in islet. *Nucb2* is encoded on mouse chromosome 7 at 116.5 Mb (red line). In islets the heritability of *Nucb2* expression levels is 69% heritable. This LOD score trace shows that there is no local eQTL at the position of the gene, nor any strong distal eQTLs anywhere else in the genome.

442 **References**

- 443 [1] M. T. Maurano, R. Humbert, E. Rynes, R. E. Thurman, E. Haugen, H. Wang, A. P. Reynolds,
444 R. Sandstrom, H. Qu, J. Brody, A. Shafer, F. Neri, K. Lee, T. Kutyavin, S. Stehling-Sun, A. K.
445 Johnson, T. K. Canfield, E. Giste, M. Diegel, D. Bates, R. S. Hansen, S. Neph, P. J. Sabo, S. Heimfeld,
446 A. Raubitschek, S. Ziegler, C. Cotsapas, N. Sotoodehnia, I. Glass, S. R. Sunyaev, R. Kaul, and J. A.
447 Stamatoyannopoulos. Systematic localization of common disease-associated variation in regulatory DNA.
448 *Science*, 337(6099):1190–1195, Sep 2012.
- 449 [2] K. K. Farh, A. Marson, J. Zhu, M. Kleinewietfeld, W. J. Housley, S. Beik, N. Shores, H. Whitton, R. J.
450 Ryan, A. A. Shishkin, M. Hatan, M. J. Carrasco-Alfonso, D. Mayer, C. J. Luckey, N. A. Patsopoulos,
451 P. L. De Jager, V. K. Kuchroo, C. B. Epstein, M. J. Daly, D. A. Hafler, and B. E. Bernstein. Genetic
452 and epigenetic fine mapping of causal autoimmune disease variants. *Nature*, 518(7539):337–343, Feb
453 2015.
- 454 [3] E. Pennisi. The Biology of Genomes. Disease risk links to gene regulation. *Science*, 332(6033):1031, May
455 2011.
- 456 [4] L. A. Hindorff, P. Sethupathy, H. A. Junkins, E. M. Ramos, J. P. Mehta, F. S. Collins, and T. A. Manolio.
457 Potential etiologic and functional implications of genome-wide association loci for human diseases and
458 traits. *Proc Natl Acad Sci*, 106(23):9362–9367, Jun 2009.
- 459 [5] J. K. Pickrell. Joint analysis of functional genomic data and genome-wide association studies of 18
460 human traits. *Am J Hum Genet*, 94(4):559–573, Apr 2014.
- 461 [6] D. Welter, J. MacArthur, J. Morales, T. Burdett, P. Hall, H. Junkins, A. Klemm, P. Flieck, T. Manolio,
462 L. Hindorff, and H. Parkinson. The NHGRI GWAS Catalog, a curated resource of SNP-trait associations.
463 *Nucleic Acids Res*, 42(Database issue):D1001–1006, Jan 2014.
- 464 [7] Y. I. Li, B. van de Geijn, A. Raj, D. A. Knowles, A. A. Petti, D. Golan, Y. Gilad, and J. K. Pritchard.
465 RNA splicing is a primary link between genetic variation and disease. *Science*, 352(6285):600–604, Apr
466 2016.
- 467 [8] D. Zhou, Y. Jiang, X. Zhong, N. J. Cox, C. Liu, and E. R. Gamazon. A unified framework for joint-tissue
468 transcriptome-wide association and Mendelian randomization analysis. *Nat Genet*, 52(11):1239–1246,
469 Nov 2020.
- 470 [9] E. R. Gamazon, H. E. Wheeler, K. P. Shah, S. V. Mozaffari, K. Aquino-Michaels, R. J. Carroll, A. E.

- 471 Eyler, J. C. Denny, D. L. Nicolae, N. J. Cox, and H. K. Im. A gene-based association method for
472 mapping traits using reference transcriptome data. *Nat Genet*, 47(9):1091–1098, Sep 2015.
- 473 [10] Z. Zhu, F. Zhang, H. Hu, A. Bakshi, M. R. Robinson, J. E. Powell, G. W. Montgomery, M. E. Goddard,
474 N. R. Wray, P. M. Visscher, and J. Yang. Integration of summary data from GWAS and eQTL studies
475 predicts complex trait gene targets. *Nat Genet*, 48(5):481–487, May 2016.
- 476 [11] A. Gusev, A. Ko, H. Shi, G. Bhatia, W. Chung, B. W. Penninx, R. Jansen, E. J. de Geus, D. I. Boomsma,
477 F. A. Wright, P. F. Sullivan, E. Nikkola, M. Alvarez, M. Civelek, A. J. Lusis, T. ki, E. Raitoharju,
478 M. nen, I. ä, O. T. Raitakari, J. Kuusisto, M. Laakso, A. L. Price, P. Pajukanta, and B. Pasaniuc.
479 Integrative approaches for large-scale transcriptome-wide association studies. *Nat Genet*, 48(3):245–252,
480 Mar 2016.
- 481 [12] M. P. Keller, D. M. Gatti, K. L. Schueler, M. E. Rabaglia, D. S. Stapleton, P. Simecek, M. Vincent,
482 S. Allen, A. T. Broman, R. Bacher, C. Kendzierski, K. W. Broman, B. S. Yandell, G. A. Churchill, and
483 A. D. Attie. Genetic Drivers of Pancreatic Islet Function. *Genetics*, 209(1):335–356, May 2018.
- 484 [13] W. L. Crouse, G. R. Keele, M. S. Gastonguay, G. A. Churchill, and W. Valdar. A Bayesian model
485 selection approach to mediation analysis. *PLoS Genet*, 18(5):e1010184, May 2022.
- 486 [14] J. M. Chick, S. C. Munger, P. Simecek, E. L. Huttlin, K. Choi, D. M. Gatti, N. Raghupathy, K. L. Svenson,
487 G. A. Churchill, and S. P. Gygi. Defining the consequences of genetic variation on a proteome-wide scale.
488 *Nature*, 534(7608):500–505, Jun 2016.
- 489 [15] H. E. Wheeler, S. Ploch, A. N. Barbeira, R. Bonazzola, A. Andaleon, A. Fotuhi Siahpirani, A. Saha,
490 A. Battle, S. Roy, and H. K. Im. Imputed gene associations identify replicable trans-acting genes enriched
491 in transcription pathways and complex traits. *Genet Epidemiol*, 43(6):596–608, Sep 2019.
- 492 [16] B. D. Umans, A. Battle, and Y. Gilad. Where Are the Disease-Associated eQTLs? *Trends Genet*,
493 37(2):109–124, Feb 2021.
- 494 [17] N. J. Connally, S. Nazeen, D. Lee, H. Shi, J. Stamatoyannopoulos, S. Chun, C. Cotsapas, C. A. Cassa,
495 and S. R. Sunyaev. The missing link between genetic association and regulatory function. *Elife*, 11, Dec
496 2022.
- 497 [18] H. Mostafavi, J. P. Spence, S. Naqvi, and J. K. Pritchard. Systematic differences in discovery of genetic
498 effects on gene expression and complex traits. *Nat Genet*, 55(11):1866–1875, Nov 2023.
- 499 [19] D. W. Yao, L. J. O’Connor, A. L. Price, and A. Gusev. Quantifying genetic effects on disease mediated
500 by assayed gene expression levels. *Nat Genet*, 52(6):626–633, Jun 2020.

- 501 [20] X. Liu, J. A. Mefford, A. Dahl, Y. He, M. Subramaniam, A. Battle, A. L. Price, and N. Zaitlen. GBAT:
502 a gene-based association test for robust detection of trans-gene regulation. *Genome Biol*, 21(1):211, Aug
503 2020.
- 504 [21] H. J. Westra, M. J. Peters, T. Esko, H. Yaghoobkar, C. Schurmann, J. Kettunen, M. W. Christiansen,
505 B. P. Fairfax, K. Schramm, J. E. Powell, A. Zhernakova, D. V. Zhernakova, J. H. Veldink, L. H. Van den
506 Berg, J. Karjalainen, S. Withoff, A. G. Uitterlinden, A. Hofman, F. Rivadeneira, P. A. C. ' Hoen,
507 E. Reinmaa, K. Fischer, M. Nelis, L. Milani, D. Melzer, L. Ferrucci, A. B. Singleton, D. G. Hernandez,
508 M. A. Nalls, G. Homuth, M. Nauck, D. Radke, U. Iker, M. Perola, V. Salomaa, J. Brody, A. Suchy-Dicey,
509 S. A. Gharib, D. A. Enquobahrie, T. Lumley, G. W. Montgomery, S. Makino, H. Prokisch, C. Herder,
510 M. Roden, H. Grallert, T. Meitinger, K. Strauch, Y. Li, R. C. Jansen, P. M. Visscher, J. C. Knight,
511 B. M. Psaty, S. Ripatti, A. Teumer, T. M. Frayling, A. Metspalu, J. B. J. van Meurs, and L. Franke.
512 Systematic identification of trans eQTLs as putative drivers of known disease associations. *Nat Genet*,
513 45(10):1238–1243, Oct 2013.
- 514 [22] Y. Gilad, S. A. Rifkin, and J. K. Pritchard. Revealing the architecture of gene regulation: the promise
515 of eQTL studies. *Trends Genet*, 24(8):408–415, Aug 2008.
- 516 [23] Aparna Nathan, Samira Asgari, Kazuyoshi Ishigaki, Cristian Valencia, Tiffany Amariuta, Yang Luo,
517 Jessica I Beynor, Yuriy Baglaenko, Sara Suliman, Alkes L Price, et al. Single-cell eqtl models reveal
518 dynamic t cell state dependence of disease loci. *Nature*, 606(7912):120–128, 2022.
- 519 [24] E. Sollis, A. Mosaku, A. Abid, A. Buniello, M. Cerezo, L. Gil, T. Groza, O. §, P. Hall, J. Hayhurst,
520 A. Ibrahim, Y. Ji, S. John, E. Lewis, J. A. L. MacArthur, A. McMahon, D. Osumi-Sutherland,
521 K. Panoutsopoulou, Z. Pendlington, S. Ramachandran, R. Stefancsik, J. Stewart, P. Whetzel, R. Wilson,
522 L. Hindorff, F. Cunningham, S. A. Lambert, M. Inouye, H. Parkinson, and L. W. Harris. The NHGRI-EBI
523 GWAS Catalog: knowledgebase and deposition resource. *Nucleic Acids Res*, 51(D1):D977–D985, Jan
524 2023.
- 525 [25] R. J. F. Loos and G. S. H. Yeo. The genetics of obesity: from discovery to biology. *Nat Rev Genet*,
526 23(2):120–133, Feb 2022.
- 527 [26] R. K. Singh, P. Kumar, and K. Mahalingam. Molecular genetics of human obesity: A comprehensive
528 review. *C R Biol*, 340(2):87–108, Feb 2017.
- 529 [27] P. Arner. Obesity—a genetic disease of adipose tissue? *Br J Nutr*, 83 Suppl 1:9–16, Mar 2000.
- 530 [28] Mark P Keller, Mary E Rabaglia, Kathryn L Schueler, Donnie S Stapleton, Daniel M Gatti, Matthew

- 531 Vincent, Kelly A Mitok, Ziyue Wang, Takanao Ishimura, Shane P Simonett, et al. Gene loci associated
532 with insulin secretion in islets from nondiabetic mice. *The Journal of Clinical Investigation*, 129(10):4419–
533 4432, 2019.
- 534 [29] G. A. Churchill, D. M. Gatti, S. C. Munger, and K. L. Svenson. The Diversity Outbred mouse population.
535 *Mamm Genome*, 23(9-10):713–718, Oct 2012.
- 536 [30] Elissa J Chesler, Darla R Miller, Lisa R Branstetter, Leslie D Galloway, Barbara L Jackson, Vivek M
537 Philip, Brynn H Voy, Cymbeline T Culiat, David W Threadgill, Robert W Williams, et al. The
538 collaborative cross at oak ridge national laboratory: developing a powerful resource for systems genetics.
539 *Mammalian Genome*, 19:382–389, 2008.
- 540 [31] Michael C Saul, Vivek M Philip, Laura G Reinholdt, and Elissa J Chesler. High-diversity mouse
541 populations for complex traits. *Trends in Genetics*, 35(7):501–514, 2019.
- 542 [32] D. W. Threadgill, D. R. Miller, G. A. Churchill, and F. P. de Villena. The collaborative cross: a
543 recombinant inbred mouse population for the systems genetic era. *ILAR J*, 52(1):24–31, 2011.
- 544 [33] S. M. Clee and A. D. Attie. The genetic landscape of type 2 diabetes in mice. *Endocr Rev*, 28(1):48–83,
545 Feb 2007.
- 546 [34] K. W. Broman, D. M. Gatti, P. Simecek, N. A. Furlotte, P. Prins, Š. Sen, B. S. Yandell, and G. A.
547 Churchill. R/qt12: Software for Mapping Quantitative Trait Loci with High-Dimensional Data and
548 Multiparent Populations. *Genetics*, 211(2):495–502, Feb 2019.
- 549 [35] Klaasjan G Ouwens, Rick Jansen, Michel G Nivard, Jenny van Dongen, Maia J Frieser, Jouke-Jan
550 Hottenga, Wibowo Arindrarto, Annique Claringbould, Maarten van Iterson, Hailiang Mei, et al. A
551 characterization of cis-and trans-heritability of rna-seq-based gene expression. *European Journal of
552 Human Genetics*, 28(2):253–263, 2020.
- 553 [36] Alkes L Price, Agnar Helgason, Gudmar Thorleifsson, Steven A McCarroll, Augustine Kong, and Kari
554 Stefansson. Single-tissue and cross-tissue heritability of gene expression via identity-by-descent in related
555 or unrelated individuals. *PLoS genetics*, 7(2):e1001317, 2011.
- 556 [37] Julien Bryois, Alfonso Buil, David M Evans, John P Kemp, Stephen B Montgomery, Donald F Conrad,
557 Karen M Ho, Susan Ring, Matthew Hurles, Panos Deloukas, et al. Cis and trans effects of human
558 genomic variants on gene expression. *PLoS genetics*, 10(7):e1004461, 2014.
- 559 [38] M. Helmer, S. Warrington, A. R. Mohammadi-Nejad, J. L. Ji, A. Howell, B. Rosand, A. Anticevic,

- 560 S. N. Sotiropoulos, and J. D. Murray. On the stability of canonical correlation analysis and partial least
561 squares with application to brain-behavior associations. *Commun Biol*, 7(1):217, Feb 2024.
- 562 [39] Gennady Korotkevich, Vladimir Sukhov, and Alexey Sergushichev. Fast gene set enrichment analysis.
563 *bioRxiv*, 2019.
- 564 [40] A. Subramanian, P. Tamayo, V. K. Mootha, S. Mukherjee, B. L. Ebert, M. A. Gillette, A. Paulovich,
565 S. L. Pomeroy, T. R. Golub, E. S. Lander, and J. P. Mesirov. Gene set enrichment analysis: a
566 knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci U S A*,
567 102(43):15545–15550, Oct 2005.
- 568 [41] S. Subramanian and A. Chait. The effect of dietary cholesterol on macrophage accumulation in adipose
569 tissue: implications for systemic inflammation and atherosclerosis. *Curr Opin Lipidol*, 20(1):39–44, Feb
570 2009.
- 571 [42] I. Akoumianakis, N. Akawi, and C. Antoniades. Exploring the Crosstalk between Adipose Tissue and
572 the Cardiovascular System. *Korean Circ J*, 47(5):670–685, Sep 2017.
- 573 [43] I. S. Stafeev, A. V. Vorotnikov, E. I. Ratner, M. Y. Menshikov, and Y. V. Parfyonova. Latent Inflammation
574 and Insulin Resistance in Adipose Tissue. *Int J Endocrinol*, 2017:5076732, 2017.
- 575 [44] I. P. Fischer, M. Irmler, C. W. Meyer, S. J. Sachs, F. Neff, M. de Angelis, J. Beckers, M. H. p, S. M.
576 Hofmann, and S. Ussar. A history of obesity leaves an inflammatory fingerprint in liver and adipose
577 tissue. *Int J Obes (Lond)*, 42(3):507–517, Mar 2018.
- 578 [45] S. Chung, H. Cuffe, S. M. Marshall, A. L. McDaniel, J. H. Ha, K. Kavanagh, C. Hong, P. Tontonoz,
579 R. E. Temel, and J. S. Parks. Dietary cholesterol promotes adipocyte hypertrophy and adipose tissue
580 inflammation in visceral, but not in subcutaneous, fat in monkeys. *Arterioscler Thromb Vasc Biol*,
581 34(9):1880–1887, Sep 2014.
- 582 [46] V. Kus, T. Prazak, P. Brauner, M. Hensler, O. Kuda, P. Flachs, P. Janovska, D. Medrikova, M. Rossmeisl,
583 Z. Jilkova, B. Stefl, E. Pastalkova, Z. Drahota, J. Houstek, and J. Kopecky. Induction of muscle
584 thermogenesis by high-fat diet in mice: association with obesity-resistance. *Am J Physiol Endocrinol
585 Metab*, 295(2):E356–367, Aug 2008.
- 586 [47] C. B. Newgard. Interplay between lipids and branched-chain amino acids in development of insulin
587 resistance. *Cell Metab*, 15(5):606–614, May 2012.
- 588 [48] D. D. Sears, G. Hsiao, A. Hsiao, J. G. Yu, C. H. Courtney, J. M. Ofrecio, J. Chapman, and S. Subramaniam.

- 589 Mechanisms of human insulin resistance and thiazolidinedione-mediated insulin sensitization. *Proc Natl*
590 *Acad Sci U S A*, 106(44):18745–18750, Nov 2009.
- 591 [49] R. Stienstra, C. Duval, M. ller, and S. Kersten. PPARs, Obesity, and Inflammation. *PPAR Res*,
592 2007:95974, 2007.
- 593 [50] O. Gavrilova, M. Haluzik, K. Matsusue, J. J. Cutson, L. Johnson, K. R. Dietz, C. J. Nicol, C. Vinson,
594 F. J. Gonzalez, and M. L. Reitman. Liver peroxisome proliferator-activated receptor gamma contributes
595 to hepatic steatosis, triglyceride clearance, and regulation of body fat mass. *J Biol Chem*, 278(36):34268–
596 34276, Sep 2003.
- 597 [51] K. Matsusue, M. Haluzik, G. Lambert, S. H. Yim, O. Gavrilova, J. M. Ward, B. Brewer, M. L. Reitman,
598 and F. J. Gonzalez. Liver-specific disruption of PPARgamma in leptin-deficient mice improves fatty
599 liver but aggravates diabetic phenotypes. *J Clin Invest*, 111(5):737–747, Mar 2003.
- 600 [52] D. Patsouris, J. K. Reddy, M. ller, and S. Kersten. Peroxisome proliferator-activated receptor alpha
601 mediates the effects of high-fat diet on hepatic gene expression. *Endocrinology*, 147(3):1508–1516, Mar
602 2006.
- 603 [53] S. E. Schadinger, N. L. Bucher, B. M. Schreiber, and S. R. Farmer. PPARgamma2 regulates lipogenesis
604 and lipid accumulation in steatotic hepatocytes. *Am J Physiol Endocrinol Metab*, 288(6):E1195–1205,
605 Jun 2005.
- 606 [54] W. Motomura, M. Inoue, T. Ohtake, N. Takahashi, M. Nagamine, S. Tanno, Y. Kohgo, and T. Okumura.
607 Up-regulation of ADRP in fatty liver in human and liver steatosis in mice fed with high fat diet. *Biochem*
608 *Biophys Res Commun*, 340(4):1111–1118, Feb 2006.
- 609 [55] A. Srivastava, A. P. Morgan, M. L. Najarian, V. K. Sarsani, J. S. Sigmon, J. R. Shorter, A. Kashfeen,
610 R. C. McMullan, L. H. Williams, P. guez, M. T. Ferris, P. Sullivan, P. Hock, D. R. Miller, T. A. Bell,
611 L. McMillan, G. A. Churchill, and F. P. de Villena. Genomes of the Mouse Collaborative Cross. *Genetics*,
612 206(2):537–556, Jun 2017.
- 613 [56] A. Roberts, F. Pardo-Manuel de Villena, W. Wang, L. McMillan, and D. W. Threadgill. The poly-
614 morphism architecture of mouse genetic resources elucidated using genome-wide resequencing data:
615 implications for QTL discovery and systems genetics. *Mamm Genome*, 18(6-7):473–481, Jul 2007.
- 616 [57] G. A. Churchill, D. C. Airey, H. Allayee, J. M. Angel, A. D. Attie, J. Beatty, W. D. Beavis, J. K.
617 Belknap, B. Bennett, W. Berrettini, A. Bleich, M. Bogue, K. W. Broman, K. J. Buck, E. Buckler,
618 M. Burmeister, E. J. Chesler, J. M. Cheverud, S. Clapcote, M. N. Cook, R. D. Cox, J. C. Crabbe,

- 619 W. E. Crusio, A. Darvasi, C. F. Deschepper, R. W. Doerge, C. R. Farber, J. Forejt, D. Gaile, S. J.
620 Garlow, H. Geiger, H. Gershenfeld, T. Gordon, J. Gu, W. Gu, G. de Haan, N. L. Hayes, C. Heller,
621 H. Himmelbauer, R. Hitzemann, K. Hunter, H. C. Hsu, F. A. Iraqi, B. Ivandic, H. J. Jacob, R. C. Jansen,
622 K. J. Jepsen, D. K. Johnson, T. E. Johnson, G. Kempermann, C. Kendzierski, M. Kotb, R. F. Kooy,
623 B. Llamas, F. Lammert, J. M. Lassalle, P. R. Lowenstein, L. Lu, A. Lusis, K. F. Manly, R. Marcucio,
624 D. Matthews, J. F. Medrano, D. R. Miller, G. Mittelman, B. A. Mock, J. S. Mogil, X. Montagutelli,
625 G. Morahan, D. G. Morris, R. Mott, J. H. Nadeau, H. Nagase, R. S. Nowakowski, B. F. O'Hara, A. V.
626 Osadchuk, G. P. Page, B. Paigen, K. Paigen, A. A. Palmer, H. J. Pan, L. Peltonen-Palotie, J. Peirce,
627 D. Pomp, M. Praveneč, D. R. Prows, Z. Qi, R. H. Reeves, J. Roder, G. D. Rosen, E. E. Schadt, L. C.
628 Schalkwyk, Z. Seltzer, K. Shimomura, S. Shou, M. J. ä, L. D. Siracusa, H. W. Snoek, J. L. Spearow,
629 K. Svenson, L. M. Tarantino, D. Threadgill, L. A. Toth, W. Valdar, F. P. de Villena, C. Warden,
630 S. Whatley, R. W. Williams, T. Wiltshire, N. Yi, D. Zhang, M. Zhang, and F. Zou. The Collaborative
631 Cross, a community resource for the genetic analysis of complex traits. *Nat Genet*, 36(11):1133–1137,
632 Nov 2004.
- 633 [58] J. Y. Huh, Y. J. Park, M. Ham, and J. B. Kim. Crosstalk between adipocytes and immune cells in
634 adipose tissue inflammation and metabolic dysregulation in obesity. *Mol Cells*, 37(5):365–371, May 2014.
- 635 [59] J. Lamb, E. D. Crawford, D. Peck, J. W. Modell, I. C. Blat, M. J. Wrobel, J. Lerner, J. P. Brunet,
636 A. Subramanian, K. N. Ross, M. Reich, H. Hieronymus, G. Wei, S. A. Armstrong, S. J. Haggarty,
637 P. A. Clemons, R. Wei, S. A. Carr, E. S. Lander, and T. R. Golub. The Connectivity Map: using
638 gene-expression signatures to connect small molecules, genes, and disease. *Science*, 313(5795):1929–1935,
639 Sep 2006.
- 640 [60] Aravind Subramanian, Rajiv Narayan, Steven M Corsello, David D Peck, Ted E Natoli, Xiaodong Lu,
641 Joshua Gould, John F Davis, Andrew A Tubelli, Jacob K Asiedu, et al. A next generation connectivity
642 map: L1000 platform and the first 1,000,000 profiles. *Cell*, 171(6):1437–1452, 2017.
- 643 [61] X. Liu, Y. I. Li, and J. K. Pritchard. Trans Effects on Gene Expression Can Drive Omnipotent Inheritance.
644 *Cell*, 177(4):1022–1034, May 2019.
- 645 [62] U. Vosa, A. Claringbould, H. J. Westra, M. J. Bonder, P. Deelen, B. Zeng, H. Kirsten, A. Saha,
646 R. Kreuzhuber, S. Yazar, H. Brugge, R. Oelen, D. H. de Vries, M. G. P. van der Wijst, S. Kasela,
647 N. Pervjakova, I. Alves, M. J. é, M. Agbessi, M. W. Christiansen, R. Jansen, I. ä, L. Tong, A. Teumer,
648 K. Schramm, G. Hemani, J. Verlouw, H. Yaghootkar, R. nmez Flitman, A. Brown, V. Kukushkina,
649 A. Kalnayenakis, S. eger, E. Porcu, J. Kronberg, J. Kettunen, B. Lee, F. Zhang, T. Qi, J. A. Hernandez,

- 650 W. Arindrarto, F. Beutner, J. Dmitrieva, M. Elansary, B. P. Fairfax, M. Georges, B. T. Heijmans, A. W.
651 Hewitt, M. nen, Y. Kim, J. C. Knight, P. Kovacs, K. Krohn, S. Li, M. Loeffler, U. M. Marigorta, H. Mei,
652 Y. Momozawa, M. ller Nurasyid, M. Nauck, M. G. Nivard, B. W. J. H. Penninx, J. K. Pritchard, O. T.
653 Raitakari, O. Rotzschke, E. P. Slagboom, C. D. A. Stehouwer, M. Stumvoll, P. Sullivan, P. A. C. 't Hoen,
654 J. Thiery, A. njes, J. van Dongen, M. van Iterson, J. H. Veldink, U. lker, R. Warmerdam, C. Wijmenga,
655 M. Swertz, A. Andiappan, G. W. Montgomery, S. Ripatti, M. Perola, Z. Katalik, E. Dermitzakis,
656 S. Bergmann, T. Frayling, J. van Meurs, H. Prokisch, H. Ahsan, B. L. Pierce, T. ki, D. I. Boomsma, B. M.
657 Psaty, S. A. Gharib, P. Awadalla, L. Milani, W. H. Ouwehand, K. Downes, O. Stegle, A. Battle, P. M.
658 Visscher, J. Yang, M. Scholz, J. Powell, G. Gibson, T. Esko, L. Franke, P. A. C. 't Hoen, J. van Meurs,
659 J. van Dongen, M. van Iterson, M. A. Swertz, and M. Jan Bonder. Large-scale cis- and trans-eQTL
660 analyses identify thousands of genetic loci and polygenic scores that regulate blood gene expression. *Nat*
661 *Genet*, 53(9):1300–1310, Sep 2021.
- 662 [63] B. Hallgrímsson, R. M. Green, D. C. Katz, J. L. Fish, F. P. Bernier, C. C. Roseman, N. M. Young,
663 J. M. Cheverud, and R. S. Marcucio. The developmental-genetics of canalization. *Semin Cell Dev Biol*,
664 88:67–79, Apr 2019.
- 665 [64] M. L. Siegal and A. Bergman. Waddington's canalization revisited: developmental stability and evolution.
666 *Proc Natl Acad Sci U S A*, 99(16):10528–10532, Aug 2002.
- 667 [65] A. B. Paaby and G. Gibson. Cryptic Genetic Variation in Evolutionary Developmental Genetics. *Biology*
668 (*Basel*), 5(2), Jun 2016.
- 669 [66] E. A. Boyle, Y. I. Li, and J. K. Pritchard. An Expanded View of Complex Traits: From Polygenic to
670 Omnipigenic. *Cell*, 169(7):1177–1186, Jun 2017.
- 671 [67] Naomi R Wray, Cisca Wijmenga, Patrick F Sullivan, Jian Yang, and Peter M Visscher. Common disease
672 is more complex than implied by the core gene omnipigenic model. *Cell*, 173(7):1573–1580, 2018.
- 673 [68] H. Shimizu and A. Osaki. Nesfatin/Nucleobindin-2 (NUCB2) and Glucose Homeostasis. *Curr Hypertens*
674 *Rev*, pages Nesfatin/Nucleobindin-2 (NUCB2) and Glucose Homeostasis., Jul 2014.
- 675 [69] M. Nakata and T. Yada. Role of NUCB2/nesfatin-1 in glucose control: diverse functions in islets,
676 adipocytes and brain. *Curr Pharm Des*, 19(39):6960–6965, 2013.
- 677 [70] M. Riva, M. D. Nitert, U. Voss, R. Sathanoori, A. Lindqvist, C. Ling, and N. Wierup. Nesfatin-1
678 stimulates glucagon and insulin secretion and beta cell NUCB2 is reduced in human type 2 diabetic
679 subjects. *Cell Tissue Res*, 346(3):393–405, Dec 2011.