Dear Editors and Reviewers,

Thank you so much for your careful reading of our manuscript. We appreciate all the time you put into this and the feedback you have provided. We have addressed each of your comments, which has greatly improved the clarity of the manuscript.

Changes to the manuscript include the following:

1) Color changes for the tissues to avoid overlap with heat maps. These can be changed if not suitable.

2) Shortening of results section relating to CMAP analysis.

3) Alteration of Figure 2B and C to use non-parametric statistics and lines

4) Alteration of Figure 4D to show null distributions and numbers of transcripts with large loadings.

5) Multiple additions of clarifying language in the main text and the methods section.

6) Made sure all numeric results in the manuscript were programmatically reported rather than entered by hand. Some of the numbers changed slightly but not enough to affect any conclusions.

Individual responses to comments are below in blue text. We have also provided a manuscript documents with tracked changes for ease of comparison.

Thank you again for all your efforts.

Best,

Gregory W. Carter

Reviewer #1 (Remarks to the Author):

Tyler et al. have used multiple metabolic phenotypes, genotypes and transcriptome data from four metabolically active tissues from the diversity outbred mice to assess the role of distal vs local gene expression on a composite metabolic phenotype (mainly comprising body weight and insulin sensitivity). They initially found that local eQTL negatively correlated with trait relevance, while distal eQTL positively associated with trait relevance. They then used high dimensional mediation analysis (HDMA) to identify transcriptome signatures that are heritable, correlate with the composite metabolic phenotype, and impact the phenotype in a casual fashion (e.g., when taken into account, the signal between genotype and phenotype disappears). They found that transcripts that contributed most strongly to the phenotype were under distal, as opposed to local genetic control and were able to identify relevant biological pathways. They then used the transcript loadings to validate these findings in CC-RIX mice and human data. The manuscript convincingly demonstrates the importance of distal eQTL for metabolic phenotypes, a finding supported by recent publications. By comparing to TWAS and genes with high local heritability, this work emphasizes the complexity of the genetic architecture underlying complex traits and argues against a focus on local eQTL. The use of HDMA is a strength and represents a tool that will likely be used by others for other models and phenotypes. The work does not highlight any new biological pathways - everything they found has previously been reported in the literature. It is not surprising that transcriptomic signatures involved in inflammation or mitochondrial function play a role in obesity, and therefore one would expect these to be replicated in the CC-RIX and human cohorts. Demonstrating tissue-specificity is also not new. Despite lack of novel biological pathways, this work clearly demonstrates the ability of HDMA to identify appropriate transcriptomic signatures as well as the importance of distal eQTL for complex traits.

Minor comments:

1) In the results section, Fig. 3a, is the composite transcriptome score for all tissues combined? Would partial correlations differ if tissues were looked at individually?

Yes, the numbers presented in Fig. 3a are for all tissues combined. To answer your question, we performed high dimensional mediation on the tissues separately to compare the results. We compared the variance explained by the model as well as the correlations between the latent variables (figures below). Combining the tissues explained slightly more variance than using the tissues separately (upper figure), and the correlations between the latent variables were comparable (lower figure). Because we were interested in explaining as much variance as possible, we used the tissues together. We feel that adding this information to the manuscript would be more confusing than helpful, so we present these results only in this response.





2) Results: In HDMA, transcripts are given loadings and these loadings are used to perform GSEA. This allows you to determine which biological pathways are altered in the composite phenotype. You do not mention the number of genes that have high or low loadings for each tissue. In other words, how many genes fall into the transcriptomic signatures that satisfy HDMA criteria (heritable, correlate with trait and fits a causal model). It would be worthwhile to include this information as a supplementary table.

We agree that this information is important. To address this, we compared the observed distributions to null transcript loading distributions. The observed distributions extended well beyond the ends of the null distributions (now shown in Fig. 4). We have modified Figure 4 to show the transcript loadings for each tissue separately compared to a null distribution. We have marked the number of genes that have loadings beyond the tails of the null distribution for each tissue. The new panels are shown below.



3) Methods: CC-RIX, you mention that a sub-set of mice were treated with metformin, but there is no mention of how these data are used in the results. Similarly, CC-RIX mice were euthanized at two time-points (6 months and 12 months) – what time-point was used in the current analysis?

Thank you for catching this error! The cohort that was given metformin will be published in later analyses but was not included in this study. We have deleted the sentences referring to metformin. Both the 6- and 12-month cohorts were used in the CC-RIX analysis. We used age as a covariate in all analyses, and we have added language to the methods to indicate this (lined 51-53 in the Methods with tracked changes).

4) Methods: you state that pancreas was used for whole pancreas insulin content. If this is the case, was a different data-set used to isolate islets for RNAseq? There is also no mention of islets under the RNAseq methods section.

Thank you for catching this omission. The methods regarding the pancreatic islets are described more in detail in the original publication describing these data (PMID: 29567659, ref 12). We have added a short paragraph to the DO section of the methods to describe these measurements (lines 13-16 in the Methods with tracked changes). The islets in the DO mice were isolated and insulin per islet was measured. The whole pancreas insulin content was calculated from insulin per islet and the total number of islets in each pancreas. RNA was then isolated from the islets and sent to The Jackson Laboratory for RNA-Seq.

5) Brain is extremely important tissue for metabolic traits such as obesity. Data here demonstrate that adipose tissue is the most relevant for obesity from the four tissues that were looked at, but brain is likely to have more relevance than adipose tissue. This may be worth a mention.

We agree that brain is likely to have gene expression variation relevant to obesity. We have added a sentence to the results to remind readers that brain is an important tissue in regulation of obesity but was not included in this study (lines 251-253 in manuscript with tracked changes). Thus, we cannot speak to the relative importance of this tissue in this context.

6) Line 148: CCA is not defined

Thank you for catching this. We have added the definition before the first mention of the term (line 170 in the manuscript with tracked changes).

Reviewer #2 (Remarks to the Author):

Tyler et al. conducted an interesting study using a population of Diversity Outbred mice, uncovering a stronger association between distal genetic elements and traits. This partly explains why local genetic variants contribute less to traits via transcripts. The findings also support a previously observed phenomenon in human genetics: “transcripts with low local heritability explain more expression-mediated disease heritability than transcripts with high local heritability.” Below are some comments:

• “Overall, local and distal genetic factors contributed approximately equally to transcript abundance”: Is this similar in human genetics? Do the GTEx data support this conclusion?

We have done a more thorough search of the literature to investigate this question. Multiple human studies have found that the ratio of heritability explained by local and distal regulation is approximately 1:4. In contrast, we found a ratio of about 1:1, that is local and distal genetic factors explained roughly equal amounts of heritability. We suspect that this could be due to several reasons. First, there is a higher degree of genetic diversity in the DO than in human populations. Thus, there may be an accumulation of variants that affect local transcription. Distal effects on transcription, however, are more likely to be buffered by the overall transcriptional network, weakening their observed effects. Second, there is a high degree of linkage disequilibrium in the DO mice compared to human populations, as well as high degree of confidence with which we can estimate ancestral haplotypes. Any given genetic marker in the DO mice captures haplotype information from a relatively large genomic region, thus encompassing many possible nearby regulatory variants. In humans, the information SNPs capture is much more localized and more difficult to assign to ancestral haplotypes. Thus, in mice there may be more information about local regulatory variants captured in the haplotype and therefore more local heritability explained. We have added text to the results section where these data are presented to speculate on the discrepancy between the findings in mice and humans. (Lines 118-127 in manuscript with tracked changes.)

• If I understand correctly, in the high-dimensional mediation analysis, the composite transcriptome score includes a non-genetic component. Could the framework capture the condition that transcripts are influenced by traits?

Yes, absolutely. The results presented here are consistent with a mediation model, but do not prove mediation. We used the results from the CMAP investigation as supportive of the mediation model. That is, drugs that act on the transcriptional pattern also act on the traits, thereby supporting a causal role of the transcriptional profile in the pathogenesis of metabolic disease. This point was not made clear in the original manuscript. We have added text to the discussion to say more explicitly that HDMA alone does not prove mediation but is only consistent with it. We further explain that the results from CMAP help support the mediation model (lines 384-389 in manuscript with tracked changes).

• The authors used the cMap data for the analysis. Are these cell-level data technically replicated? If so, are the technical replicate results consistent?

The CMAP data do include technical replicates with a high degree of consistency across replicates (PMID: 29195078, ref 60 in the manuscript). The results that we present are from the CMAP GSEA query, which takes into account shared modes of action and targeted pathways across individual drugs and cell types. These aggregated statistics take into account fluctuations in drugs in each class as well as across replicates.

• “To assess the importance of genetic regulation of transcript levels to clinical traits, we compared the local and distal heritabilities of transcripts to their trait relevance, defined as the maximum trait correlation for each transcript.” Was transcript level standardized during this process? How exactly is this “maximum trait correlation” defined?

Thank you for pointing out this lack of clarity. We regressed sex, DO wave, and DO generation out of the rank normalized transcript values. We also regressed these covariates out of the phenotypes. We then performed Spearman rank correlation between all pairs of traits and transcripts. The trait relevance of a given transcript was the maximum absolute value of its correlation coefficients across all traits. We have clarified this point in both in the methods and the main text.

Minor Comments:

1. Please define “local” and “distal” prominently.

We have added text to the introduction to define local and distal and to distinguish these terms from cis and trans (lines 33-37 in manuscript with tracked changes).

2. In line 110: “We calculated the heritability of each transcript” — the word “calculate” should be replaced with “estimate.”

We have changed "calculate" to "estimate." (line 114 in manuscript with tracked changes)

3. In line 148, when CCA first appears, please provide its full name.

Thank you for catching this. We have added "canonical correlation analysis" to the first instance of CCA (line 170 of manuscript with tracked changes).

Reviewer #3 (Remarks to the Author):

This study uses a cohort of diversity outcross mice to perform a novel mediation analysis of obesity-related traits. The two main findings are that (I) distal eQTLs (ie trans eQTLs) seem to play a larger role in controlling transcripts relevant to obesity traits than do local (cis) eQTLs, and (ii) it is possible to define composite measures of genotype and transcriptome and phenotype and perform a mediation analysis on the composite measures, thereby establishing causality.

Both of these findings are interesting and noteworthy and would be of interest to the readership of Nature Comms. However, I also had a large number of queries about the manuscript. Whilst most of these are related to improving the presentation, there are a two more substantive queries which need to be dealt with satisfactorily.

Major points:

(i) Is the permutation procedure (Fig 3B,C ) valid? It ignores relationships between individuals across omic levels (ie assumes all mice are exchangeable) and so might inflate the apparent significance. If I have understood correctly, the permutation procedure is performed such that it destroys any correlation between the three omic levels – including the existence of any QTLs or eQTLs, and this is likely too harsh a null hypothesis. It might be better to use multivariate generative models to simulate sets of genotypes, transcripts and phenotypes and evaluate performance on those rather than using permutation. An alternative might be to ask how unusual are the correlations cor(P\_C , T\_C) etc for the optimal choices of weights compared to randomly sampled weights using the unpermuted data, possibly augmented with a distribution fitted to the random correlations. In any event, that authors should justify their choice of permutation strategy and explain why it supports their thesis.

We apologize for the lack of clarity on this critical point. The permutation test was performed by only permuting the transcriptomes. The correlations between genome and phenome were preserved in this test. Thus, we preserved the effects of any QTLs while breaking the association between eQTLs (and all other genetic effect on transcription) with QTLs. The essential goal of this procedure was to determine whether it was possible within a random transcriptome to identify a spurious transcriptomic profile that appeared to mediate the true genotype-phenotype correlations as well as our optimized transcriptomic profile in the true data. Our permutation test definitively rules this possibility out; the true transcriptome is more highly aligned with the genotype-phenotype map than expected at random. We stress that this is nontrivial, given that canonical correlation analysis (CCA)-based approaches have problems overfitting because they involved estimating the inverse of a covariance matrix and, hence, can detect such spurious components. We have added clarification of this point, which can be seen in lines 176-179 in the manuscript with tracked changes.

The suggestion to consider the performance of HDMA under generative model hypotheses is an important one, and is the subject of ongoing research, but is beyond the scope of this paper. In this study, we use HDMA in a purely descriptive mode.

We randomly sampled the unpermuted weights 10,000 times and recalculated the correlations cor(G\_C, T\_C) and cor(T\_C, P\_C) as shown in Figure 3C. The results are shown below. We think it would add more confusion than information to add this to the manuscript figure, but shows clearly that the correlations identified by HDMA are well

outside the null distribution. The original permutated values from the manuscript in Figure 3 are shown in gray, and the observed correlations are shown in red.



(ii) I’m a bit puzzled by the use of the CC-RIX mice as a validation set. It appears that actual body weights were not measured in the CC-RIX (why not? – weight is a standard phenotype) so they were imputed from local transcript data, which complicates their use for validation and makes it far less convincing. I really don’t see what they add to the study. Why not simply keep back a random 10% of the DO mice, train the models on the 90% and test the predictions into the 10%, for many random samplings? (ie the standard machine learning cross validation procedure).

We apologize for the lack of clarity on this point. Body weight was measured in the CC-RIX. We use measured body weight as the ground truth validation in the CC-RIX population. We estimated the metabolic disease index (MDI), which is largely based on body weight from the DO and compared it to the actual body weight as the validation (Figure 7B). To estimate MDI, we either used the measured transcriptome in the CC-RIX, or a predicted transcriptome based on local genotype. While the MDI based on the measured transcriptome correlated well with body weight, the MDI based on the locally imputed transcriptome was not correlated with body weight. We interpreted this result as support that genetic effects on genotype are not mediated through local gene regulation. It was important to use the CC-RIX data as opposed to validating within the DO data because the point of local and distal regulation is particularly critical when translating results between populations with different allele structure. We suggest that there is a failure of human TWAS results to translate between human populations because genetic effects are mediated through distal gene regulation, which are dramatically different across populations with different allele structures. In this mouse experiment, the two populations shared all ancestral haplotypes, but had dramatically different allele structure, thus allowing us to tease apart the effects of local and distal gene regulation on phenotypic effects.

Minor Points:

(i) Figure 1 is very good, except Figure 1G could be improved if the upper triangle of the heatmap displayed genetic correlations and the main diagonal heritabilities.

This is an interesting way to add more information to the plot. We tried it (shown below), but then realized that it might be confusing to have two different types of data in a single heat map. Because most of the cells show Pearson correlation, but the diagonal has very low values, we think this might cause more confusion, especially since the heritabilities are shown in Fig 1F.



(ii) Figure 2 could be significantly improved by replacing the violin plots in Fig2A with overlapping distributions as in Figure 1A,B. Fig 2B is informative but I think the use of linear regression is not the best way of showing the shapes of the distributions for Local and Distal are different. Clearly a straight line does not fit any of the data very well. Can the authors think of another measure which quantifies the fact that strong local eQTL are more likely to have small trait correlations than strong distal eQTL?

We tried using overlapping distributions (below), but it became difficult to compare the distributions across the tissues. We would like Figure 2A to show a direct comparison of the variance explained by local and distal eQTLs across tissues, so we reverted to the violin plots.



We changed Fig2 B and C to avoid the linear test. We binned the transcripts into centiles based on their variance explained. We then calculated the mean and 95th and 5th percentiles of their maximum trait correlation. We smoothed these values using splines. The top and bottom lines in each panel shows the 95th and 5th percentile respectively and the middle line shows the mean. We reported statistics from Spearman rank correlation tests.

A collage of graphs and diagrams

AI-generated content may be incorrect.

(iii) Fig 3 is hard to follow, particularly since it mentions Kernelization which is otherwise not mentioned in the main paper (it’s mentioned in the Methods) Not sure what G\_K, T\_K , P\_K signify.

We have added a short description of the kernelization terminology to the main text where we describe high-dimensional mediation. The new text appears in lines 150-156 of the manuscript with tracked changes.

(iv) Figure 4 D – Please include scatter plots of P\_C vs T\_C, G\_C vs T\_C, G\_C vs P\_C - this is surely key to understanding. Please replace violin plots with overlapping distributions as in Figure 1A,B

We tried using overlapping histograms to compare these distributions, but the distributions are so similar that the figure was difficult to read (see below). Instead, we replaced the distributions with separate panels showing the observed loading distribution for each tissue compared with the null. We shaded the area of the distributions that were more extreme than the null distribution and noted the number of genes in each extreme group. This adds a little more information to help distinguish the distributions.

The scatter plots took up a lot of room in Figure 3 without adding much information, so we have created a new Supplemental Figure S3 to show the details of these correlations.



(v) Fig 5 is good. It would be helpful to define exactly what is meant by TWAS in this study. I suggest a completely different color ramp is used to indicate tissue type from that used to indicate heritability – it’s a bit confusing. It would also help to report the p-values of the t-tests for comparing the heritability distributions for distal vs local (in the three inset boxplots in the Fig 5)

We have added a short description of our TWAS procedure to the results section that discusses Figure 5. We have also changed the colors for the tissues to select

colors that do not overlap with the heat maps or the local/distal color scheme. We added t test p values to the legend of Figure 5 for each of the box plots.

(vi) Given the diverse ancestry of the DO, involving alleles from three different murine subspecies, it would be interesting to know if the distal or local eQTLs more often involved alleles segregating between subspecies.

This is an interesting question and has been addressed extensively elsewhere (PMIDs 25230953, 21406540, 34849860, 34425882, 33264334, 31961859). Because eQTLs were only an incidental piece of this manuscript (presented only in supplemental figures) and because this issue has been examined previously, we performed a limited analysis that is presented only in this response letter. We collected local and distal allele coefficients for all haplotypes. A comparison of the local and distal eQTL coefficients shows that the *castaneus* (green) and *musculus* (red) haplotypes tend to have larger local coefficients than the *domesticus* strains (gray). An ANOVA with Tukey's Honestly Significant Difference showed that the *castaneus*, *musculus*, and the wild-derived *domesticus* strain (WSB) had significantly stronger local eQTL coefficients than the other *domesticus* strains (top figures). For distal coefficients (bottom figures), both *castaneus*, and *musculus* strains had significantly higher coefficients than all *domesticus* strains.

A screenshot of a graph

AI-generated content may be incorrect.

(vii) Is there a reason for preferring the nomenclature “Local vs Distal” instead of the more usual “cis vs trans”?

We avoid the usage of cis and trans terminology because these terms have specific biochemical definitions that are not completely captured by physical location on the genome (See Box 2 in PMID: 18597885, ref 22 in the ms). Because we are not evaluating the biochemical nature of variants in this study and are classifying variants only by genomic position, we use the terms local and distal. We have added text to the introduction to define these terms and discuss the difference between cis/trans and local/distal terminology.

(viii) The statements around line 280: The mean loadings for alpha-cell specific transcripts were significantly greater than 0, while the mean loadings for delta- and endothelial-cell specific genes were significantly less than 0 (Fig. 8B)

The study does not give p-values for these statements – Fig 8B appears to show boxplots which don’t give an indication of significance. There are several places in the MS where boxplots are used as evidence of “significance” without a formal p-value being provided.

We apologize for this omission. We have added p values to the text for each of these significance claims.

(ix) What were the exact criteria for calling eQTL (local and distal) and pQTL? Were different significance thresholds applied for local vs distal? The methods are vague – why does a LOD score threshold of 8 equate to a p-value of 0.05 (and what does this p-value mean – it is genome wide significance?) Surely it would be better to use an FDR-based threshold.

Local eQTLs were defined as eQTLs within 4Mb of the transcription start site of the encoding gene. We have added this definition to the methods. We used a nominal p value of 0.05 as a permissive and arbitrary cutoff to compare basic stats of local and distal eQTLs. The eQTLs were a marginal component of the manuscript (only presented in supplemental figures), and changing this threshold did not change the general conclusions about eQTLs that we presented. Namely, local eQTLs outnumber distal eQTLs and local eQTLs tend to be shared across tissues whereas distal eQTLs tend to be tissue specific. We did not look at pQTLs.

(x) Why was heritability computed from called eQTLs rather than from suitably partitioned genetic relationship matrices?

The local heritability of each transcript (regardless of whether it had an eQTL) was calculated using the haplotype probabilities for the marker nearest to the transcription start site of the encoding gene. In typical situations, we would include the kinship matrix as a random effect in this model. However, here we were using the fitted values of the model as the locally encoded transcript level, and we did not want to confound the model by including distal factors. In reviewing these methods we noticed that we mistakenly said that we used the kinship correction. We have now removed this from the text (line 145 in the Methods with tracked changes).

(xi) The human validation material is quite interesting but I am not an expert on this part of the paper. It did seem overly long.

Upon second reading of this section, we agree that it should be shortened. We have reduced the length of this section to highlight the most important results. (Section starts on line 337 in manuscript with tracked changes.)

(xii) The discussion is incredibly short – there was no attempt to place the findings in a wider context.

We initially included a discussion with a "supplemental discussion" to show that we could reduce the length of the manuscript if needed. We have removed the "supplemental discussion" header to include the full-length discussion, which does discuss the work in a broader context.

Methods:

(i) Genotyping (line 25 onwards) Why were haplotypes determined from RNAseq reads and not from the SNP genotypes?

We apologize for the confusing wording of this section. Haplotypes were determined both from GigaMUGA SNPs and by RNA-Seq. Using both methods provides redundancy as a quality control measure. There were several samples in which the two methods disagreed or had poor quality RNA-Seq data. These mice were excluded from the analysis. We have added text to the methods clarifying why both methods were used (lines 34-36 in Methods with tracked changes.)

(ii) It not clear whether the CC-RIX mice were kept in the same animal facility as the DO mice or were from a different experiment. Were they on the same high-fat diet as the DO? - please clarify.

It was not clear in the methods which mice were being described in some sections. We have added "CC-RIX" to one of the headers in the methods to indicate that this section describes the CC-RIX mice (line 40 in Methods with tracked changes). The CC-RIX mice were housed at The Jackson Laboratory. The DO mice were part of a previous experiment and were housed at the University of Wisconsin. DO and CC-RIX mice were maintained on different high-fat, high-sugar diets. DO mice received a HF/HS diet (44.6% kcal fat, 34% carbohydrate, and 17.3% protein) from Envigo Teklad (catalog number TD.08811). The CC-RIX mice received a custom-designed high-fat, high-sugar (HF/HS) diet (Research Diets D19070208).

(iii) It is not clear what the “processed data” (Methods line 31) refer to. Are these the CC-RIX genotypes? If so, what are the gene expression data?

We agree that this heading is confusing. We have changed the heading to "Pre-processed DO data" and added a sentence to the underlying paragraph that these data were part of a previous publication, and we downloaded them directly from Dryad (line 38 in Methods with tracked changes).

(iv) TWAS analysis (methods line 250 onwards. Using just the SNP closest to the TSS for each gene might result in underestimating local genetic effects – it would have been better to have taken the most associated SNP within say 100kb. Not all cis SNPs will be associated with the expression trait, so picking one based solely on location is sub-optimal.

We agree that this method would be problematic in human data. However, the mice used in this experiment have large haplotype blocks, and the markers within 100kb of any given marker will have identical or nearly identical genotype distributions across the animals.

Reviewer #3 (Remarks on code availability):

The link to the code https://figshare.com/DOI:10.6084/m9.figshare.27066979228 does not open

We apologize for this error. We have fixed the URL in the manuscript, and it should now point to the correct page.