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. Individual responses to your comments are below. Our responses are 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?

We performed high dimensional mediation on the tissues separately. We compared the variance explained by the model as well as the correlations between the latent variables (figures below). Using the tissues together explained slightly more variance than using the tissues separately, and the correlations between the latent variables were comparable. Because we were interested in explaining as much variance as possible, we used the tissues together.





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. We compared the observed distributions to null transcript loading distributions. The observed distributions extended well beyond the ends of the null distributions. 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. We have also added scatter plots to this figure to show the correlations between the latent variables

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 manuscript. 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.

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 using these data (PMID: 29567659). We have added a short paragraph to the DO section of the methods to describe these measurements. The islets in the DO mice were isolated and insulin per islet was measured. The whole pancreas insulin content was derived 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. 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.

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 better 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 the high degree of linkage disequilibrium in the DO mice compared to human populations, as well as the 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 discuss the comparison to findings in humans.

• 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?

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

• “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?

Minor Comments:

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

We have added text to the introduction to define local and distal and distinguish these terms from cis and trans.

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

Good point. We have changed "calculate" to "estimate."

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.

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.

(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, 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 did not correlate with body weight at all. 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.

Thank you! This is an interesting way to add more information to the plot. We have changed this heat map to show heritability on the diagonal. It changed the clustering in an interesting way.

(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 have changed Fig 2 according to these suggestions. We replaced the violin plots with overlapping density plots. We also replaced the linear models in Fig. 2B and C with splines. We binned the transcripts into centiles based on their variance explained. We then calculated the mean and 95th percentile of their maximum trait correlation. We smoothed these values using splines for better visualization. The top line in each panel shows the 95th percentile and the lower line shows the mean.

(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.

(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 (figure 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 distinguish the distributions. We have also added scatter plots to show correlations between the latent variables across the tissues.



(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 color ramp 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. eQTLs were an incidental piece of this study, as we focused primarily on local and distal heritability. We compared the relative allele coefficients from haplotypes in the three subspecies for local and distal eQTL. On average, the *castaneus* (CAST), and *musculus* (PWK) alleles had stronger allele effects than the *domesticus* alleles (all the rest). There was no difference in contributions between local and distal eQTL.

(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 captured by physical location on the genome (See Box 2 in PMID: 18597885). 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.

(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.

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

(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.

(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.

(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. 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.

(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.