

Concordance between male- and female-specific GWAS results helps define underlying genetic architecture of complex traits

Corresponding Author: Dr David Buchner

This file contains all reviewer reports in order by version, followed by all author rebuttals in order by version.

Version 0:

Reviewer comments:

Reviewer #1

(Remarks to the Author)

The manuscript by Miller and colleagues investigates the importance of considering biologic sex in genetic associations studies. Utilizing a previously published mouse study (HMDP) the authors investigate the effect of sex on genetic association results and test for Concordance between sex stratified results—results shared by both sexes. They also compare these results to more traditional models where sex is used as a covariate. These sex-focused studies are supplemented by additional analysis that focus on outliers and normalization. The latter demonstrate the robust nature of the genetic approach. Specifically, that the results are not solely dependent on outlier values or highly variable by normalization techniques. The methods are well described and the importance of biological sex on complex traits is unquestionable. Below are some suggestions and questions in regards to the analysis and conclusions

Questions

- The underlying data is based on a prior publication that used HMDP mice that were F1. Does the use of F1 mice affect the results? Theoretically the genetic effect at many loci are ½ traditional recombinant inbred mice. This should be discussed.
- What is role of effect size on concordance of the QTL/association? Do large effect size replicate more than others? Are those with large effect size likely to have same directionality between sexes?
- Figure 1 shows concordant association results between sexes at a p-value of $p < 0.01$ through $p = 0.95$. Most of these associations are assumed to be modest (given p values) so concordance of directionality is impressive. The authors suggest alternatives to p-values could be considered. Have they considered presenting the data in figure 1 according to betas for the regression model? If direction is similar between sexes how different are the beta's for SNPs between sexes.
- It is understood that the focus of the manuscript is a global analysis and presentation of data. That being said an example of concordant and discordant results at a specific SNPs or locus would aid the overall study
- The presentation of the epistasis data and analysis is difficult to follow. The presentation of the data in figure 2 closely resembles figure 1 (main effects). The text delineates how these analysis and data are different to come to the conclusion that there is not widespread detectable epistasis.

Similar results are shown for BMI and height in UK biobank. Some reference to the effect size of a snp and the likelihood of concordance (or not) is important. This is quite relevant to a critical point the authors make about PRS scores. If the argument is to improve PRS scores by utilizing all SNPs then understanding the overall effect size of the SNPs between sexes is important (in addition to directionality).

Reviewer #2

(Remarks to the Author)

In this manuscript, the authors used Hybrid Mouse Diversity Panel (HMDP) to first examine heritability and genetic correlations of 8 metabolic traits and 3 blood related traits in male and female mice. Then, they performed sex-stratified GWAS to identify loci with main effects on the above traits. Using the same threshold, much more QTLs were detected in male population than female population, which could not be explained by the normality and outliers of the data. Next, by examining the directionality of the main effects, the authors found that although the male-significant QTLs did not reach significance in female population, most of them still share the same directionality. This also holds true in the human BMI and height traits in UK Biobank. However, most of the conclusions from this work were very vague as the authors mostly just listed some descriptive numbers (correlations, number of significant SNPs, percentages of QTLs share the same directionality) rather than performing meaningful statistical testing to draw more assertive conclusions. Both the main figures and tables are not very informative, making the novelty and significance of this work very limited.

Comments to authors:

1. The results on heritability and genetic correlation of traits are less informative given that only scores and correlation values were listed which is hard to draw clear conclusions from. More quantitative statistical tests should be performed on the results between the two sexes to draw more solid conclusions.
2. For figure 1,2, and 3, how meaningful it is to group main effect loci by their p values (especially when $p > 0.05$ which indicates non-significance. eg. $0.3 < p < 0.4$ vs $0.4 < p < 0.5$)? Grouping them by effect size or confidence intervals might be more meaningful.
3. Given the observation that male-specific and female-specific analysis identified very different QTLs, the authors performed normalization, outlier removal, and randomization of the data. The conclusion was that sex-specific QTLs are not likely to be artifacts due to the nature of the data. Such analysis took a good chunk of the result part, but it seems a bit deviated from the theme of this paper. Also, those results did not result in any main figures. The authors should consider downplay those results as they are less informative and not very relevant to the theme of this paper.
4. The main purpose and the main findings of this work are unclear. The authors made it clear that sex-specific GWAS resulted in more signals that would be concealed in sex-combined analysis, and that was indeed their observations. However, their subsequent work was to prove that many of the male-specific QTLs have the same directionality in females even though they did not reach significance. This is the conclusion of all the three main figures. QTLs that are only significant in one sex indicates 1) they might be part of the genetic architecture of the trait; 2) such relationship might be sex-specific. It is unclear what extra value it can bring by stating that the direction of allelic effects is the same in both sexes even though p value is not significant.
5. The authors hypothesized that "loci that are not associated with a trait should have equal chances of trending in the same direction of effect". The author also stated that "the percentage of loci with effects in the same direction in males and females was far greater than the 50% expected by chance (Fig. 1A, Fig 1B) across nearly all levels of statistical significance". Does it mean all the loci being screened should contribute to the genetic architecture of complex traits (even though p val is greater than 0.8 or 0.9)?
6. Discussion has no limitations listed. For example, the authors proposed a very loose criteria to call loci with main effects (same direction in male and female). However, this will lead to a lot of false positives.

Here are some minor concerns:

1. It is premature to conclude that "the metabolic traits had higher heritability than the blood traits in both sexes" since the number of traits in each category is too limited to draw this conclusion.
2. Are the authors using main effect and marginal effect interchangeably? If yes, try to use one terminology to avoid confusions.
3. Figure S3 is difficult to interpret. Some SNPs with both positive (or both negative) beta in male and female were labeled as circle (opposite direction). Any reason behind this?
4. It was very hard to see the shape of the dots in figure 1C.

Reviewer #3

(Remarks to the Author)

I co-reviewed this manuscript with one of the reviewers who provided the listed reports. This is part of the Nature Communications initiative to facilitate training in peer review and to provide appropriate recognition for Early Career Researchers who co-review manuscripts.

Reviewer #4

(Remarks to the Author)

This study presents an innovative approach to understanding the genetic architecture of complex traits by comparing sex-stratified and sex-combined GWAS results. The authors provide a compelling case for the importance of sex-stratification in uncovering associations that might otherwise be missed and propose an interesting method to assess concordance in effect directions between sexes as evidence of shared genetic influences. The study is well-motivated, and several aspects of the methodology are particularly intriguing. However, some areas could benefit from further clarification and additional validation.

- One key question is why so few significant loci were identified in males and why none were detected in females, despite the high reported heritability. The authors suggest this is due to power limitations, but the contrast between Table 1 and

Table 2 raises some questions about the analytical approach. Expanding the analysis to include additional phenotypes from the UK Biobank, particularly disease traits, could help clarify whether power is the primary factor driving these results.

- The interpretation of Table 3 is another area that could be strengthened. The authors emphasize the importance of sex-specific effects but later suggest that effect directions are generally concordant between sexes. If the observed sex differences in Table 2 are mainly due to limited power, the proportion of concordant loci in Table 3 should be higher than 50%. Further discussion on the rationale behind this analysis and its implications would help solidify the argument.

- The study primarily relies on mouse data, which provides a useful model, but there are occasional contradictions regarding power limitations. While the authors note that certain results lack statistical power, they also draw broad conclusions from this dataset. Clarifying the strengths and limitations of the mouse data, and where possible, incorporating more human data from UK Biobank, would enhance confidence in the findings. Additionally, differences in SNP coverage between the human and mouse datasets could contribute to the observed disparities in significant loci, and addressing this could provide valuable context.

- The phantom pleiotropy analysis is another intriguing aspect of the study, but its interpretation could be refined. The authors initially suggest the presence of interactions, then describe them as artifacts, and finally propose possible epistasis. Providing a clearer and more cohesive discussion of these results, along with potential follow-up analyses, would strengthen this section.

Version 1:

Reviewer comments:

Reviewer #2

(Remarks to the Author)

The reviewer is satisfied with the revision.

Reviewer #3

(Remarks to the Author)

I co-reviewed this manuscript with one of the reviewers who provided the listed reports. This is part of the Nature Communications initiative to facilitate training in peer review and to provide appropriate recognition for Early Career Researchers who co-review manuscripts.

Reviewer #4

(Remarks to the Author)

The authors have thoroughly addressed previous concerns with thoughtful clarifications and new analyses, including additional UK Biobank phenotypes, downsampling experiments, and effect size correlations. These revisions strengthen the argument that limited power, not biological absence, explains sex-specific differences in detected loci. With improved clarity around phantom pleiotropy and a well-articulated discussion of limitations, the manuscript now presents a compelling and balanced contribution to understanding sex-stratified genetic architecture.

Open Access This Peer Review File is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

In cases where reviewers are anonymous, credit should be given to 'Anonymous Referee' and the source.

The images or other third party material in this Peer Review File are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.

To view a copy of this license, visit <https://creativecommons.org/licenses/by/4.0/>

We thank all of the reviewers for their very helpful comments regarding our previous submission and have addressed all of their comments both below and in the manuscript. We feel that these changes significantly improved the manuscript and thus hope that you will agree that it is now suitable for publication in *Nature Communications*. Our point-by-point responses to the reviewer comments are below:

Reviewer comments:

1. Reviewer #1 comment: *The underlying data is based on a prior publication that used HMDP mice that where F1. Does the use of F1 mice affect the results? Theoretically the genetic effect at many loci are ½ traditional recombinant inbred mice. This should be discussed.*

Response: The reviewer is correct, the data was from F1 mice. We have now clarified this in the introduction. We also added a statement (see below) to the discussion as to how this may affect results and interpretations by limiting the analyses to only variants in the heterozygous state which may in fact underestimate the number of loci regulating each trait.

"It is important to consider also that the mice analyzed were all F1 mice from crosses between C57BL/6J and the HMDP strains. This somewhat limits the variation present, as alleles within the HMDP strains were all fixed in the heterozygous state, thus it is expected that the loci identified underestimate the potential number of loci regulating each trait given that HMDP alleles were not studied in the homozygous state."

2. Reviewer #1 comment: *What is role of effect size on concordance of the QTL/association? Do large effect size replicate more than others? Are those with large effect size likely to have same directionality between sexes?*

Response: This is an excellent point. We have looked at this explicitly now and added results for the role of effect sizes in terms of replicating results across sexes. This is presented as a new figure for the HMDP analyses (Fig. 2) and a new figure panel for the UK Biobank analyses (Fig. 4B). In addition, we added an explicit correlation analysis showing that the effect size correlation across sexes is significant for all traits. This is included as a new table (Table 4) and the correlation plots for all quantitative traits are shown in Supplemental Figures 3, 4, and 6.

3. Reviewer #1 comment: *Figure 1 shows concordant association results between sexes at a p-value of $p < 0.01$ through $p = 0.95$. Most of these associations are assumed to be modest (given p values) so concordance of directionality is impressive. The authors suggest alternatives to p-values could be considered. Have they considered presenting the data in figure 1 according to betas for the regression model? If direction is similar between sexes how different are the beta's for SNPs between sexes.*

Response: As noted in the response above (point #2) we now present these results suggested by the reviewer. Overall, the relationships sorted by beta values recapitulate those sorted by p values.

4. Reviewer #1 comment: *It is understood that the focus of the manuscript is a global analysis and presentation of data. That being said an example of concordant and discordant results at a specific SNPs or locus would aid the overall study.*

Response: The original manuscript included an example of a concordant SNP in the results section (rs48316748). We have now added a new example of a discordant SNP to facilitate a better understanding of the methods and results. The relevant section of the text now reads as follows:

“For example, SNP rs48316748 was the most significant SNP associated with body weight in male mice in the HMDP ($p < 1.27E-11$, $Beta = 2.94$), with the “A” allele being associated with increased body weight. The “A” allele at this SNP was also associated with an increase in body weight in the analysis of body weight in females, although the p -value was not statistically significant ($p = 0.11$, $Beta = 0.74$) (Fig. S3). In contrast, SNP rs4831674 was associated with a positive beta value of 0.16 in males ($p=0.67$) but was associated with a negative beta value of -0.40 ($p=0.30$) in females. As such, SNP rs48316748 had a concordant direction of effect in males and females, whereas SNP rs4831674 was not concordant for direction of effect between the sexes.”

5. Reviewer #1 comment: The presentation of the epistasis data and analysis is difficult to follow. The presentation of the data in figure 2 closely resembles figure 1 (main effects). The text delineates how these analysis and data are different to come to the conclusion that there is not widespread detectable epistasis.

Response: We have tried to rewrite the section on epistasis to make it clearer. Our primary message is that in real data we detect evidence of common effect directions for interactions (epistasis). However, we now also show via simulations that many of these interactions are likely to be data artifacts and not real epistasis. Hence, we conclude that detected statistical interactions may not reflect real epistasis, a result that we find surprising.

6. Reviewer #1 comment: Similar results are shown for BMI and height in UK biobank. Some reference to the effect size of a snp and the likelihood of concordance (or not) is important. This is quite relevant to a critical point the authors make about PRS scores. If the argument is to improve PRS scores by utilizing all SNPs then understanding the overall effect size of the SNPs between sexes is important (in addition to directionality).

Response: As described above (point #2), we have now added a complete analysis section based on effect sizes for all quantitative traits as noted above. We also address the reviewers’ important points in the discussion regarding how this may affect the role of PRS.

7. Reviewer #2 comment: The results on heritability and genetic correlation of traits are less informative given that only scores and correlation values were listed which is hard to draw clear conclusions from. More quantitative statistical tests should be performed on the results between the two sexes to draw more solid conclusions.

Response: We thank the reviewer for suggesting additional analyses as we think these strengthen our conclusions. Specifically, we have now added more analyses regarding specifics of correlations between sexes, including results on effects sizes and replicability across sexes and explicitly examining effect size correlations. We show that correlations are significant for all traits and tests. We think that these new analyses make our points clearer and, in fact, stronger, further supporting our main conclusions of widespread concordance between the sexes.

With respect to the results on heritability and genetic correlations, and as also noted in minor comment #13 below, statistical testing is not really possible across trait classes. Therefore, we have added a statement suggesting caution when comparing heritability or genetic correlations between traits or between the sexes. The following statement was added to the results section:

“That said, it is important to exercise caution when comparing heritability estimates, or genetic correlations as discussed below, between traits or sexes given the relatively small sample sizes involved.”

8. Reviewer #2 comment: For figure 1,2, and 3, how meaningful it is to group main effect loci by their p values (especially when $p > 0.05$ which indicates non-significance. eg. $0.3 < p < 0.4$ vs $0.4 < p < 0.5$)? Grouping them by effect size or confidence intervals might be more meaningful.

Response: We agree that effect sizes are as important as p values and have added analyses to show similar results for effect sizes, as described above. However, we do think analyses based on p values have merit as in most genetic analyses these are the bases for testing for replication.

9. Reviewer #2 comment: Given the observation that male-specific and female-specific analysis identified very different QTLs, the authors performed normalization, outlier removal, and randomization of the data. The conclusion was that sex-specific QTLs are not likely to be artifacts due to the nature of the data. Such analysis took a good chunk of the result part, but it seems a bit deviated from the theme of this paper. Also, those results did not result in any main figures. The authors should consider downplay those results as they are less informative and not very relevant to the theme of this paper.

Response: We certainly agree that these analyses are not the main focus of the paper. However, we feel strongly that it is important to demonstrate the rigor of our findings, and that these analyses should not be minimized. In addition, given the comments made by reviewer #1, which we have quoted below, we think they are essential to showing the robustness of the results. We have therefore chosen to leave these in the paper as they were, as they demonstrate that our results stand up to scrutiny.

Relevant comments from reviewer #1: “These sex-focused studies are supplemented by additional analyses that focus on outliers and normalization. The latter demonstrate the robust nature of the genetic approach. Specifically, the results are not solely dependent on outlier values or highly variable by normalization techniques. The methods are well described and confirm the importance of biological sex on complex traits.”

10. Reviewer #2 comment: *The main purpose and the main findings of this work are unclear. The authors made it clear that sex-specific GWAS resulted in more signals that would be concealed in sex-combined analysis, and that was indeed their observations. However, their subsequent work was to prove that many of the male-specific QTLs have the same directionality in females even though they did not reach significance. This is the conclusion of all the three main figures. QTLs that are only significant in one sex indicates 1) they might be part of the genetic architecture of the trait; 2) such relationship might be sex-specific. It is unclear what extra value it can bring by stating that the direction of allelic effects is the same in both sexes even though p value is not significant.*

Response: We are sorry that the main message appears to be unclear. Our primary argument is that the many SNPs that are not statistically significant in one sex should have random direction of effect in the other sex. That they do not is crucial as it indicates that these variants are still likely to underlie part of the architecture of the assessed complex traits. Simply, the findings are inconsistent with the null hypotheses that nominally non-significant SNPs are random in terms of effects across sexes. We have tried to restate this more clearly in the revised text.

11. Reviewer #2 comment: *The authors hypothesized that "loci that are not associated with a trait should have equal chances of trending in the same direction of effect". The author also stated that "the percentage of loci with effects in the same direction in males and females was far greater than the 50% expected by chance (Fig. 1A, Fig 1B) across nearly all levels of statistical significance". Does it mean all the loci being screened should contribute to the genetic architecture of complex traits (even though p val is greater than 0.8 or 0.9)?*

Response: The reviewer is interpreting our results correctly – our findings indicate that most variants associate with trait variation. This may be due to SNPs we analyzed being causal or in LD with causal SNPs. Based on the simulations, the latter might be closer to the truth. However, we found that there are many more causal loci than canonical significance thresholds would reveal, indicating that the traits examined are closer to being omnigenic than generally assumed. Therefore, results need to be interpreted in light of many very small effects that may not be detectable as significant in finite samples but are nonetheless associated with trait variation. One needs to understand this in terms of p values representing probabilities and not definitive statements about association. We have tried to edit the text to make this clearer. We have added text in the discussion to try and explain this better.

12. Reviewer #2 comment: *Discussion has no limitations listed. For example, the authors proposed a very loose criteria to call loci with main effects (same direction in male and female). However, this will lead to a lot of false positives.*

Response: We have now added a limitations section. It reads as follows:

"Our study had a few limitations. First, we recognize the limited diversity in the HMDP, although it is still better than most comparable mouse studies. Also, as noted above, all of the mice were F1s crossed with C57BL/6J, which may limit the generalizability. However, the human studies did not suffer from this limitation and the overall results were still quite similar. For both the mouse and human studies we only examined a limited number of traits that were either metabolic or blood related. Hence, we cannot extrapolate with a high degree of confidence to unrelated traits. Finally, while we still think that the findings are noteworthy, we cannot explain the substantial difference in significant association in males and females. Although it is well-known that male mice respond more strongly to obesogenic environments⁵³, we recognize that our results warrant further explicit examination."

13. Reviewer #2 minor concern: It is premature to conclude that "the metabolic traits had higher heritability than the blood traits in both sexes" since the number of traits in each category is too limited to draw this conclusion.

Response: We completely agree and have changed the text to reflect hits.

14. Reviewer #2 minor concern: Are the authors using main effect and marginal effect interchangeably? If yes, try to use one terminology to avoid confusions.

Response: We have changed the text to use only the term main effects now.

15. Reviewer #2 minor concern: Figure S3 is difficult to interpret. Some SNPs with both positive (or both negative) beta in male and female were labeled as circle (opposite direction). Any reason behind this?

Response: We thank the reviewer for catching this typo. The circle symbol was mistakenly used in Figure S3B. These datapoints have been changed to triangles to indicate concordant directionality.

16. Reviewer #2 minor concern: It was very hard to see the shape of the dots in figure 1C.

Response: To clarify, the symbols in Fig. 1C are all circles. Nonetheless, we agree they are small and have increased the size of each circle in both Fig. 1 and Fig. 3 by 50% to improve readability.

17. Reviewer #4 comment: One key question is why so few significant loci were identified in males and why none were detected in females, despite the high reported heritability. The authors suggest this is due to power limitations, but the contrast between Table 1 and Table 2 raises some questions about the analytical approach. Expanding the analysis to include additional phenotypes from the UK Biobank, particularly disease traits, could help clarify whether power is the primary factor driving these results.

Response: We have now added additional phenotypes from UK Biobank, including two dichotomous traits – obesity and type 2 diabetes, as well as a new analysis involving downsampling the number of individuals from the UK Biobank BMI data. With both the dichotomous traits (added to Fig. 4A) and the smaller BMI samples sizes (added as a new Fig. 5), we show that with this reduction in power we fail to detect the signal of concordance between males and females as well as was detected in the larger sample sizes and with the continuous traits, collectively indicating the importance of power.

In addition, with respect to contrast mentioned between Table 1 and Table 2, the high heritability means that the overall genetic effects are large not that any particular SNP will necessarily be large or significant. In fact, our results indicate that the totality of genetic effects are likely made up of many very small effects that are almost completely undetectable using traditional metrics. We now make this point in the discussion.

18. Reviewer #4 comment: The interpretation of Table 3 is another area that could be strengthened. The authors emphasize the importance of sex-specific effects but later suggest that effect directions are generally concordant between sexes. If the observed sex differences in Table 2 are mainly due to limited power, the proportion of concordant loci in Table 3 should be higher than 50%. Further discussion on the rationale behind this analysis and its implications would help solidify the argument.

Response: The proportion of concordant loci in Table 3 are higher than 50% for 2 of the 3 traits. For the third trait, it is also higher than 50%, although not statistically significantly so, likely due to a small sample size.

Nonetheless, this brings up an important point that we have sought to address by adding a new table illustrating the correlation of beta values between the male and female values (Table 4) and new figures showing the correlation data (Fig S3, S4, S6). These data indicated that there is a significant correlation for all traits studied, however the correlations range from 0.2549 to 0.6015, suggesting that there are real differences in beta values between the sexes, potentially explaining why certain loci were identified only in one sex.

19. Reviewer #4 comment: The study primarily relies on mouse data, which provides a useful model, but there are occasional contradictions regarding power limitations. While the authors note that certain results lack statistical power, they also draw broad conclusions from this dataset. Clarifying the strengths and limitations of the mouse data, and where possible, incorporating more human data from UK Biobank, would enhance confidence in the findings. Additionally, differences in SNP coverage between the human and mouse datasets could contribute to the observed disparities in significant loci, and addressing this could provide valuable context.

Response: We thank the reviewer and have added these points to a limitations section now included in the discussion. Also, as noted above we have added two new disease outcomes to the UK Biobank analyses (obesity and type 2 diabetes). In addition, we now call better attention to the fact that the number of SNPs analyzed in the UK Biobank was limited to approximately the same number of SNPs for the mouse HMDP analysis in order to control for the different numbers of SNPs between the mouse and human genomic data.

20. Reviewer #4 comment: The phantom pleiotropy analysis is another intriguing aspect of the study, but its interpretation could be refined. The authors initially suggest the presence of interactions, then describe them as artifacts, and finally propose possible epistasis. Providing a clearer and more cohesive discussion of these results, along with potential follow-up analyses, would strengthen this section.

Response: We agree that these points were not as well explained as possible. We have tried to make our points more transparent. In essence, the evidence for interactions in the analyses is problematic as the simulations were performed without any interaction effects. However, this does not mean that there are no interactions, simply that many of those detected are likely not real.

We thank the reviewers for their time and for their thoughtful comments, which we feel greatly improved the quality and readability of our manuscript. We are glad to hear that they now feel that we have addressed their comments and that the manuscript is now suitable for publication in *Nature Communications*.

Reviewer comments:

Reviewer #2 comment: “The reviewer is satisfied with the revision.”

Reviewer #3 comment: “I co-reviewed this manuscript with one of the reviewers who provided the listed reports. This is part of the Nature Communications initiative to facilitate training in peer review and to provide appropriate recognition for Early Career Researchers who co-review manuscripts.”

3. Reviewer #4 comment: “The authors have thoroughly addressed previous concerns with thoughtful clarifications and new analyses, including additional UK Biobank phenotypes, downsampling experiments, and effect size correlations. These revisions strengthen the argument that limited power, not biological absence, explains sex-specific differences in detected loci. With improved clarity around phantom pleiotropy and a well-articulated discussion of limitations, the manuscript now presents a compelling and balanced contribution to understanding sex-stratified genetic architecture.”