Thanks for the suggestions from reviewer 3, which we all implemented in the submitted minor revision as detailed in the following.

Reviewer 3

Basic reporting

Well written throughout, with no major errors.

Thank you.

Experimental design

The aim of the manuscript is well described, and the research gap is pointed out, i.e. that genotypes from large sequencing datasets and large phenotype datasets have not been well integrated.
The authors have addressed or at least acknowledged previous concerns about significance and power, so that they are now clear to the reader.

Thank you.

Validity of the findings

Again, the authors have addressed my concerns abut the validity of the findings, and pointed out the caveats in a way that is clear to the reader.
There is still some question about the novelty of the biological findings (as they are not validated), but I consider this to be secondary to the main point of the paper: the authors have produced an easily used code for looking for variants between strains that they can expand in the future, and shown that it gives useful outputs.

Thank you.

Comments for the author

Dear Authors,
I feel that you have addressed the review comments well, in that you have made clear to the reader what the short-comings are, and how to minimize false discoveries.

Thank you.

Smaller, minor suggestions are below:

Lines 62-63: The preprint Ashbrook et al was recently published, PMID: 33472028 https://t.co/gDeKS50ANt?amp=1

We updated the reference.

Lines 69-70: I think it would read better if the sentence was “Several of the aforementioned resources allow the user to interactively query genotypes for 70 user-selected inbred mouse strains for input genes or genetic regions”.

We updated the sentence as suggested.

Lines 70-72: The Variant browser in GeneNetwork has some of this functionality (http://www.genenetwork.org/snp_browser and http://gn1.genenetwork.org/snpbrowser.html), however it does not seem to be programmatic accessible, limiting its usefulness compared to MouseFM.

We added a corresponding sentence as follows:
Moreover, the variant browser in GeneNetwork allows also comparison of genotypes between strains, however, data can only be extracted gene- or region-wise and is not accessible programmatically.

Line 181: Should read “we used” rather than “we use”.

Corrected.

Expression quantitative trait loci section: It would be interesting to know how many of the novel (i.e. not found in Mostafavi) variants are possibly cis (e.g. with ~5-10Mb of the gene)? I would expect many of the these large effect variants to be in cis. This might be another quality metric, or another way to determine the best candidate gene for expression of that gene.

We agree. In the current experimental setting though, we only aim to identify cis eQTLs, and thus assessed the same region as in Mostafavi *et al*., i.e. variants within 1 MB of the transcription start site. We made this clearer by adding the sentence:
In our experimental setting, we also use a 1 MB cutoff and aim to detect cis eQTLs with MouseFM.

Lines 263-272: Looking at the Zimmerman et al 2019 paper, it seems that they don't give confidence intervals (e.g. 1.5 LOD drop), but rather just gave the peak position. Eyeballing their QTL maps (without having the data to reanalyze it), it seems like your Chr7, and certainly your Chr11 SNPs overlap with the QTL detected in Zimmerman et al.
To me, it would be a plausible suggestion that you have identified variants underlying those two QTL. It is also possible that the two QTL that you do not see in this study are driven by the other strains that you did not use.

We agree, and thanks a lot for investigating this closer. So far, we considered the signals not to overlap, because their peak position is not contained within the MouseFM reported loci.

It is true though, that according to Fig. 4 of Zimmerman *et al*. the QTL regions are much larger and likely include the MouseFM loci on chromosome 7 and 11. We updated the sentence accordingly as suggested:
None of the loci identified by us overlaps with the markers of peak LOD score reported by Zimmerman et al., but according to visual inspection, two of their four QTL regions overlap with regions reported by MouseFM, one on chromosome 7 and one on chromosome 11. MouseFM may thus have identified variants underlying those two QTLs. The two other loci reported by Zimmerman et al. may have been missed by MouseFM, because they are driven by strains not used here.
Accordingly, we updated lines 317-319, which now read:
Re-analyzing a study on interfrontal bone formation (IF) resulted in MouseFM loci that did not overlap the markers of peak LOD score reported in the original study, but according to visual inspection, two of the corresponding QTLs.

Lines 289-290: It seems odd that you detect the same QTL region on chr7 for Dystrophic cardiac calcification and Interfrontal bone. Is this due to chance? Biological overlap (e.g. something in the calcification/bone pathway)? Or similar strains, and therefore regions that are divergent between the groups appear often?

Yes, we were also wondering whether this is driven by a shared biological mechanism, strain similarity, or by chance. Shared biology is plausible, as well as strain group similarity. However, additional lab and in silico experiments are needed to investigate this closer.

Line 301: “a setting” is not needed.

We removed it.

Line 310-311: If the data for the eQTL mapping was available, you could redo it, using only your subset of strains, to test this hypothesis.

This is true, we updated the sentence accordingly:
Secondly, previously undetected eQTLs may occur in this smaller set, which could be tested in future work by repeating the Mostafavi et al. analysis for the exact same strains used by MouseFM.

Line 379-380: the expression datasets are also an example of quantitative traits.

Yes, we thus updated the sentence to:
MouseFM can also be performed on quantitative traits as we showed for expression data and in the interfrontal bone example.
