Author’s response to reviews

Title: MyoMiner: Explore Gene Co-expression in Normal and Pathological Muscle

Authors:

Apostolos Malatras (apmalatras@biol.uoa.gr)

Ioannis Michalopoulos (imichalop@bioacademy.gr)

Stephanie Duguez (s.duguez@ulster.ac.uk)

Gillian Butler-Browne (gillian.butler-browne@upmc.fr)

Simone Spuler (Simone.Spuler@charite.de)

William J Duddy (w.duddy@ulster.ac.uk; billduddy@hotmail.com)

Version: 1 Date: 14 Feb 2020

Author’s response to reviews:

We thank both reviewers for their time and insight, and for the very constructive improvements that they have indicated. Please see our responses in-line below, and please note that we have uploaded a version of this letter with our responses in a different colour to the reviewer’s comments. We have also uploaded a version of our manuscript with tracked changes, for ease of reference.

Reviewer reports:

Samuel Handelman (Reviewer 1): I apologize for the delay in returning these comments.

Malatras et al. report MyoMiner, which serves and compares expression correlations between muscle tissue samples. Key features of MyoMiner include significant batch-correction (using PCA), and stratification of samples along several important characteristics. The authors are to be especially commended for their diligence in addressing metadata across contributing groups. This is useful work and worthy of publication; this manuscript will be of interest both within the muscle domain, and will help to establish best practices for expression analysis in a range of tissues and systems.

We thank the reviewer for their positive comments, and for highlighting our effort to address metadata across studies.

Major questions/comments:
The authors are not using GTEx muscle results (which is RNAseq rather than array-based), or indeed RNAseq results of any kind. This bears considerable discussion.

GTEx uses PEER factors rather than PCA to both detect-for and adjust for confounding. The approach utilized by GTEx will therefore account for inter-sample variability in, for example, post-mortem interval, relative fat content / cell type within the muscle biopsy, and RNA integrity. Are the samples reported here entirely live-tissue biopsies, or are cadaver samples included as well? Summary information on the tissue collection (realizing that the authors employ 74 series.)

A central purpose of MyoMiner is to enable comparison of gene co-expression between different muscle pathologies and treatment conditions in human and murine samples, including both whole muscle and cell studies of myoblasts and myotubes. GTEx includes a muscle dataset but this is limited to whole muscle tissues from healthy human subjects. Similarly, although some RNA-Seq datasets have been generated for muscle studies and are in the public domain, these tend to be limited in sample numbers and in diversity of pathologies/conditions compared to the vast number of microarray samples available.

However, we understand that the reviewer may be concerned about the relative quality of data – that RNA-Seq may have greater accuracy, for example due to its superior detection of low abundance transcripts and avoidance of saturation at high detection levels (the limitations of which for microarrays are already noted in the manuscript’s discussion section) - despite our best efforts to be rigorous in quality control of microarray data – and that this could undermine the overall value of MyoMiner if microarray data were considered to be untrustworthy. For this reason, in this revision we test gene co-expression in the GTEx muscle dataset against the healthy whole muscle sub-set of MyoMiner, by pair-wise comparison across a panel of 20 genes. We show that there is moderately strong agreement in the pair-wise correlation values in MyoMiner compared to those obtained from the GTEx data, with an $r$ value of 0.66 (Figure S1 of the revised manuscript). We add some discussion of this to our existing discussion paragraph, suggesting that differences between microarrays and RNA-Seq technology (i.e. the limitations of microarrays) may be responsible for the relatively strong agreement that we observe with resources (SEEK and MEM) that make greater use of microarrays (please see response to reviewer 2). We thank the reviewer for prompting this important validation, which we hope improves the paper.

Regarding PEER factors, we understand these to relate mainly to the problem of clinical confounding parameters (such as post-mortem interval, relative fat content, etc.) that were important in GTEx. These are not really an issue in MyoMiner as here we already create categories that are homogenous in terms of their clinical characteristics (also see elements of the response to reviewer 2 below).

Minor comments:

For the audience in BMC Genomics, some of the formulas in the methods could be moved into the supplement.

We have now moved equations 3-6 ($\rho$ to $Z$ Fisher transformation and confidence interval equations) to the supplementary file (Table S6).
Qian Zhu (Reviewer 2): This paper describes a web based resource for searching for coexpression genes and networks in muscle related gene expression studies. I have a number of concerns. First the writing should be improved substantially in introduction and methods due to the general weak motivation (please see below). The paper also needs to compare with coexpression genes retrieved by other prevailing coexpression webservers, such as MEM, SEEK. The motivation for doing the differential coexpression analysis is also weak. I have some question regarding how the integration is performed (whether correlations are combined or samples are combined and correlation is computed just once?). The batch effect correction needs to be demonstrated more widely (ie across all datasets in the compendium).

Please see our responses below to each of these specific concerns.

Authors should explain the motivation for coexpression studies better. Why would users do coexpression analysis when they can do differential expression analysis? Some papers need to be cited here.

The importance and main purpose of gene co-expression analysis is in discovering the mechanistic links between genes. As such, gene co-expression is another form of functional association, alongside other types of functional associations such as protein-protein interactions determined by immunoprecipitation experiments, or protein cellular co-localization as determined by immunostaining. Since functional association data can be regarded as a graph structure, gene co-expression can be used for network biology or network medicine types of analysis. As such, the study of gene co-expression can be used to understand better the mechanisms of molecular interaction within a cell. This may be the whole cell, in which case it can be called interactomics, or it can be focused on a specific gene, function, or pathway (an example from muscle research is the dystrophin-associated protein complex, the DAPC, comprising a number of largely co-expressed genes, many of which are implicated in neuromuscular pathologies). A major purpose of MyoMiner is to facilitate and enhance the study of changes to molecular interaction networks, both at the whole cell level and for specific functions (e.g. DAPC), between different pathologies and conditions. This is quite different from the main purpose of differential expression analysis. Differential expression analysis per se tells us about the differences in individual gene transcript levels between conditions, and does not tell us about the mechanistic interactions between genes. It may be combined with functional enrichment testing to detect changes across gene sets (representing functions, pathways, and cellular components), but even then it cannot tell us about mechanistic relationships between genes within a given gene set. Furthermore, the most direct application of differential expression analysis is to compare disease or treated conditions against healthy or untreated controls, and the results of such differential expression analyses have already been reported for the majority of datasets that are present in MyoMiner – these are reported in the publications emerging from each of the studies that are included in MyoMiner. We would suggest that these comparisons are often best carried out within each study, and that there would be minimal novelty in revisiting these previous analyses. In contrast, estimates of gene co-expression are improved considerably by meta-analysis across datasets, and gene co-expression in muscle has not previously been addressed.
We apologize to the reviewer if this rationale was not satisfactorily explained in the original manuscript. In the revision, we have added text to the introduction summarizing the points above, and have added several additional references.

It seems that authors have already well annotated the muscle samples into 142 categories, including tissue origin, age, gender, anatomic part, experimental condition. Given the diverse annotations, why would a user still do a coexpression analysis when differential analysis is more natural to addressing specific questions? Again better motivation is needed.

Please see the rationale above explaining the difference in purpose between gene co-expression analysis and differential expression analysis. A systematic analysis of differential expression across 142 categories would indeed be an interesting study, but it would be a massive undertaking and quite separate in purpose from the present work. Again, we have added text to the introduction to better explain the rationale.

Since this paper creates a web-server for querying coexpressed genes, it seems that there are some key citations missing: MEM (P Adler et al 2009, Genome Biology), SEEK (Zhu et al 2015, Nature Methods). These should be included, and a discussion comparing them to your tool should be briefly mentioned in introduction. Last time I checked, the Refine Search function of SEEK also can enable coexpression queries in a tissue specific way (by limiting to muscle tissue). How does MyoMiner compare to this Refined Search function of SEEK? I would like to see some comparison of coexpressed genes between MyoMiner, SEEK, and MEM, in order to validate that the coexpressed genes you found are accurate.

We apologize to have omitted these important references, which are now included early in the introduction as key examples of tools that facilitate the exploration and analysis of gene co-expression across tissues, common technological platforms, or conditions. The introduction then narrows in on tissue-specific co-expression resources, explaining that MyoMiner includes many pathologic and other conditions of muscle, not just healthy human samples (please see response to reviewer 1), and that these are separated by category.

We thank the reviewer for the suggestion to use a comparison with existing servers to validate MyoMiner. To do so, we have calculated all 190 pair-wise Pearson correlation values across a panel of 20 muscle-relevant genes for the healthy human whole muscle category of MyoMiner (Human|Both genders|All Ages|Skeletal muscle|Normal), and compared these against similar correlation values for the same pairs of genes given by MEM and SEEK for the closest relevant datasets that we could identify in the MEM and SEEK databases. Pearson correlation could be obtained directly from SEEK, whereas MEM returns a p-value for the strength of correlation that is not directly comparable to Pearson, so for MEM we ranked the 190 pair-wise correlations and compared rankings between the two tools. A text search for “Skeletal Muscle” in the MEM tool enabled extraction of correlation values for a dataset that combined 25 muscle-relevant studies on the Affymetrix HG U133 Plus 2.0 array. For SEEK, the ‘Muscle (Non-cancer)’ dataset was chosen, which combines 87 data series, most (but not all – e.g. there are some endothelial studies) of which are muscle-related. We observed strong agreement in correlation values for MyoMiner with SEEK (Pearson r = 0.87) and with MEM (Spearman rho = 0.74). These results
have now been added to the manuscript (Figures 4 and S1), and we thank the reviewer once more for this suggestion, which we hope strengthens the paper.

Authors again go into length calculating differential coexpression network and genes between conditions. The motivation for doing so is really unclear when it is much more apparent, and easier to do a differential expression analysis instead (both analyses require human annotations of expression samples. If these annotations have been provided, I would think that a DE analysis is more natural). What are the advantages to doing a differential coexpression analysis? Quantitative comparisons would be needed.

Please see our responses above. The differences between co-expression analysis and differential expression analysis are qualitative rather than quantitative, as they serve different purposes.

Combining datasets: it is very unclear what approach is taken by the authors. Do authors construct a larger matrix with muscle samples across laboratories, then calculate gene correlations on this larger matrix? Or do authors calculate gene correlations on individual matrices made of samples within same laboratory, then combine correlation values across laboratories? This distinction should be made very clearly. My critiques would depend on what kind of approach authors take, since there are pros/cons associated with each. I would appreciate some discussion of why the authors chose the approach and not the other.

We apologize that our approach to combining datasets was not well detailed in the original manuscript, and that this has led to some confusion. Our approach was to define categories of similar samples (e.g. one category would be “Human|Male|Child|Skeletal Muscle|DMD”). We defined as many such categories as possible, providing that sufficient sample numbers were available (&gt;=12 samples, based on criteria recommended by the WGCA team at the time of analysis). In the revision, we now include two master meta-data files (Table S4; one for human, one murine) as supplemental tables, which will help make clear the organization into these categories.

The methods section now includes more details on the steps taken to combine datasets. Briefly: For each category, we created a single matrix file including all the samples of that category, and on which was carried out the calculation of gene co-expression. This approach made it possible to create a large number of categories while still maintaining high power for each category by satisfying the N&gt;=12 WGCA criteria, which would not have been possible by analyzing each series separately then combining by r-to-z transformation (e.g. a study of 10 disease and 10 control samples has &gt;=12 samples for neither of its 2 categories, but &gt;=12 samples may be obtained by combining it with other data series for each category). We now explain this more clearly and in greater detail in the Methods section.

Batch effect correction: I agree with the effort for doing a batch effect correction, even though the effect in many cases is small. The caveat is that in most cases, batches are not indicated with the samples. The scan date of samples are one way to get batch information but it is not ideal. Even when such batch information is indicated, samples per batch are not balanced across studies or some batch information is present in one study, missing in others. Furthermore, recent
studies have shown that batch effect correction removes meaningful biological signals in the datasets, calling into the question the whole practice of correcting batch effect.

We hope that the improved explanation (discussed above) of how datasets were combined will help to clarify any confusion regarding our approach to batch effects, when considered alongside the ‘batch effects evaluation’ section of the Methods.

The need (or not) for batch correction was established separately for each category, by visual inspection of a 3D PCA plot, colored first by series number, then by scan date. Batch effect correction was carried out only if there was evidence of clustering based on the colors (covariates). It was necessary to test series number for covariance due to the approach described above for combining samples into categories. Based on our experience, scan date is often an important covariate, and it was visibly clustered on a number of the MyoMiner categories.

While I do not want to get into philosphical debate, I would like to ask:

1) can authors quantify the extent of batch effect in the GSE series? It is known batch effect affects only a subset of genes. How many genes per series have significantly changed expression before and after batch effect correction?

We have now tested for differential expression for each gene in each category, comparing before batch correction to after batch correction. None of the 142 categories has any significant changes (T-test \( p<0.05 \)) in expression level of any genes when comparing before batch correction to after batch correction following multiple testing (FDR). Without multiple testing correction, only 2 of the 142 categories had a significant change in expression level of any genes (Human|Skeletal muscle|Normal|All|All: 92 genes; Quadriceps|Normal|All|All: 406 genes). This is now noted in the revised version.

2) what is extent of batch effect in the whole compendium? You have demonstrated batch effect in 3 series in your table 2, but it is anecdotal and not convincing. Can you do PCA on all series (81 human, and 198 mouse, separate for each organism), so that one can see what are the sources of variation in the WHOLE compendium? Then once you have applied batch effect correction on the WHOLE compendium, how much effect does the correction have on the sources of variations in the data? That is, If you do PCA after batch effect correction, what are the sources of variations that remained in the samples?

We hope that the improved explanation (as discussed above) clarifies this – we did not carry out batch effect correction across the whole dataset, but rather by category.

3) how do authors account for missing batch information? For example if series X has categories Age, Strain, Sex annotated, but series Y only has Age and Sex, but does not have strain information. How do authors correct for such situation?

Since we approached batch correction by category rather than across the whole dataset, the large majority of categories were homogenous in terms of age, strain, sex, etc. For several types of category (those including either ‘SkeletalMuscle|Human’, ‘SkeletalMuscle|Mouse’, ‘Heart|Human’, ‘Heart|Mouse’, or ‘Mouse|All strains’) we sought to be as inclusive as possible, and included samples for which either muscle groups, heart tissue type, or mouse strain, were
missing. In these cases, batch corrections were still carried out across data series, but we did not carry out a correction across each of the descriptors for which metadata were not available for all samples. We have now added some text to the results section noting that correction was not possible for these categories due to the fact that the relevant meta-data were not available for some samples.

Other specific questions:
Are the methods appropriate and well described?
No. Method of data integration is not well described (see above).

Please see responses above

Are the conclusions drawn adequately supported by the data shown?
Not sure. Comparison with other prevailing web-servers is needed to know the accuracy of the coexpressed genes returned by authors' method.

Please see responses above.

We thank both reviewers once more for their very constructive and useful insights.