# Reviews and Comments on Paper 67

Submission details
Add comment
Add new review
Request review
Revise review 3
Edit note

## Paper information

**Paper:** Yao Yu, Tao Xu, Xuan Li, Yongtao Yu, Yixue Li and Pei Hao. Association of tissue lineage and gene expression: conservatively and differentially expressed genes define common and special functions of tissues
**Current decision:** **REJECT** (reject)

## Summary of received reviews and comments

Reviews superseded by other reviews are shown in the grey color in the table. All times are GMT.

| | date | PC member | subreviewer | score | confidence |
|---|---|---|---|---|---|
| Review 1 | Feb 19 | Wei Wu | | **-1** | 3 |
| Review 2 | Feb 19 | Uwe Ohler | Stoyan Georgiev | **-2** | 2 |
| Review 3 | Feb 21 | Josh Stuart | | **-2** | 3 |
| Comment 1 | Feb 23 | Eric Xing | | | |
| Comment 2 | Feb 23 | Josh Stuart | | | |

### Review 1

**PC member:** Wei Wu

**Overall rating:** **-1** (weak reject)

**Confidence:** **3** (high)

In this manuscript, the authors investigated tissue lineage and their gene expression patterns in human and mouse. The results presented are interesting, however, I have some concerns over this work.

In order to estimate gene and tissue expression divergence, the authors invented several metrics for calculating gene divergence, tissue expression distance, etc. Since these metrics are arbitrarily defined, the authors should first verify and present evidence to show that they are valid; without doing so, it is difficult to assess how trustworthy their results are. Another way to do this is to verify some of the findings after they present the data from computational analysis.

In my opinion, some of the metrics invented by the authors are too arbitrarily defined. For example, they defined genes which were

| | |
|---|---|
| Review: | expressed as the top 5% in all the tissues as ubiquitously expressed housekeeping genes. In order to identify 'housekeeping' genes, they should first specify what they meant by housekeeping genes since there are different versions of the definition of housekeeping genes in the literature, then they should examine expression values of all genes expressed in all the tissues and see what the top 5% of such genes look like and whether they can be justified as 'housekeeping' genes.<br><br>Same process should be done to justify their selection criteria of conservative genes.<br><br>Some metrics are inconsistently defined. E.g., they used a ranked-based method to estimate the divergence of genes between a pair of tissues and the ranks of genes are obtained based on the magnitude of their expression values. However, when they estimated tissue expression distance, they used expression values of genes directly in the calculation. I would suggest that they define these metrics more consistently to make their results more comparable.<br><br>In section 2.3.1, the authors said the tissue expression distance "D values in all tissue pairs in human and mouse follow a normal distribution (see in Supplementary)" The p values of the GO modules were also calculated based on this assumption or observation. However, I have not found this data in the Supplement Data. There are plots in Figures 1 and 2 in the Supplement Data, but none of them follow normal distributions.<br><br>As one of the major findings of the manuscript, the authors claimed "Tissues from the same segment on fate map share more similar expression pattern than those from different origins." However, this is not supported by their data. Fig. 1 shows that for mesoderm tissues, it is the other way around.<br><br>Figures are hard to read. |
| PC only: | |
| Time: | Feb 19, 06:43 |

**Review 2**

| | |
|---|---|
| PC member: | Uwe Ohler |
| Reviewer: | Stoyan Georgiev |
| Overall rating: | **-2** (reject) |

| | |
|---|---|
| **Confidence:** | **2** (medium) |
| **Review:** | The goal of this paper is to study mammalian tissue development (defined in GO & KEGG) at the molecular level as quantified by gene expression data in multiple tissue types. For this task, they considered 24 orthologous adult tissues, and ~5,000 orthologous gene pairs between human and mouse. Different statistics on the gene and gene set level are applied to look at divergence and similarity between particular tissues; for instance, tissues from the same segment in the embryogenesis fate map share more similarly expressed genes/gene sets than tissues from distant segments, and are more pronounced in ectoderm than endoderm or mesoderm.<br><br>The authors define multiple gene expression based tissue distance/similarity metrics and use them in seemingly arbitrary fashion (with arbitrary cutoffs). The manuscript is poorly organized, with lots of undefined references, e.g. for the distance statistic (page 2). It lacks a coherent description of the methodology, which would place the proposed quantitative techniques in a common framework to be used in the data analysis. The authors refer to multiple tables in supplementary information which is missing, probably because ISMB does not allow for supplements. However, there was enough room to incorporate some of these quite crucial tables in the main paper.<br><br>Overall, while some of the findings sound very interesting, they need more work to be backed up. The ortholog set is quite small, and different technologies used to create the expression panel may confound the results. Extensions to eg more species could provide more confidence. Altogether, there is no assessment of the performance of the approach, against some well-defined baseline or related approaches, which makes it hard to judge the overall contribution. |
| **PC only:** | |
| **Time:** | Feb 19, 18:44 |

**Review 3**

| | |
|---|---|
| **PC member:** | Josh Stuart |
| **Overall rating:** | **-2** (reject) |
| **Confidence:** | **3** (high) |
| | The article provides a survey of comparing expression profiles in different tissues. The authors introduce both a gene-based and a gene set-based measure of similarity. |

Review:

Major Comments

While the idea of comparing tissues based on gene expression profiles is not new, it is still an interesting area. Indeed, the ability to reconstruct the ontological relationships of the tissues from molecular profiles would be fascinating. However, the article suffers in three main areas: 1) lack of methodological development; 2) lack of biological motivation and interpretation; and 3) lack of clarity.

The article describes the application of fairly straightforward distance, overlap, and statistical tests to measure differences in gene expression profiles in different pairs of tissues. No methodology is advanced in this work. The authors use a mixture of sum-of-squares differences, Normal theory, KS tests, Pearson correlations, rank ratios, and hypergeometric distributional overlap tests to assess the pairwise similarity between expression changes in different tissues. In general, their methodology is standard to bioinformatics analysis and does not require the level of desciption that they provide. While the development of novel methodology is not required for publication in ISMB, application of methodology should be more than calculating and reporting statistics. As such, the article would be better suited for a more biology-oriented journal.

The motivation for why the authors embark on comparing all pairs of tissues is not clear. It would be interesting if the authors assessed how easy it is to predict a tissue from its expression profile or reconstruct the known relationships among the tissues. The overall conclusions drawn from all of the computed comparisons are rather superficial. I get the impression that the authors handpick anecdotal results to discuss (e.g. CALM2 and Camk2g are discussed as neural-related but no mention is made of how many other genes like this are, or are not, found). The authors have applied basic bioinformatics to pre-existing datasets, without revealing any novel insights. Their conclusion that "tissues from the same segment on [the] fate map share more similar expression patterns than those from different origins of embryogenesis" is an obvious one that has been explored and utilized to much greater and more conclusive ends. Their conclusions about the tissue specific differences and similarities between mouse and human are highly speculative, and would require substantially more analysis across different species to be substantiated.

Most of the text is hard to follow and needs proofing for correct grammar. I feel the article may have faired better if it had passed through one or two rounds of editing. The paper has numerous

spelling and grammar errors throughout. These errors cause the paper to be difficult and confusing to read. The figures in the paper also could benefit from additional work.

Minor Comments

Figs 1 and 2. The legends are too small to read.

Table 1. A more informative caption would help.

Fig. 3. The x-axis labels should be explained.

Their GO analysis is flawed because it fails to take into account the underlying structure of Gene Ontology, which is not really biological at all. Rather, Gene Ontology is a set of artificially create gene list, with a lot of redundancy and overlap. When viewed as such, conclusions that they make based on the numbers of correlated GO categories such as "gene expression in non-neural tissues is more divergent than in neural related tissues" are highly suspect.

**PC only:**
**Time:** Feb 21, 04:54

**Comment 1**

**By:** Eric Xing
**Comment:** Looks like we have a consensus here. Are we fine with a rejection?
**Time:** Feb 23, 22:07

**Comment 2**

**By:** Josh Stuart
**Comment:** Yes, fine with me to reject.
**Time:** Feb 23, 22:33

## Add comment

**Please type your comments in the text area below. You comments will only be visible to PC members having access to this paper. It will not be sent to the authors.**

Add Comment