

Jury Member Report – Doctor of Philosophy thesis.


Name of Candidate: Song Guo

PhD Program: Life Sciences

Title of Thesis: Using RNA expression as a quantitative molecular phenotype to study human and vertebrate evolution

Supervisor: Professor Philipp Khaitovich

Name of the Reviewer: Michael Lachmann

<p>I confirm the absence of any conflict of interest</p>  <p>(Alternatively, Reviewer can formulate a possible conflict)</p>	<p>Date: 5/11/2021</p>
---	-------------------------------

The purpose of this report is to obtain an independent review from the members of PhD defense Jury before the thesis defense. The members of PhD defense Jury are asked to submit signed copy of the report at least 30 days prior the thesis defense. The Reviewers are asked to bring a copy of the completed report to the thesis defense and to discuss the contents of each report with each other before the thesis defense.

If the reviewers have any queries about the thesis which they wish to raise in advance, please contact the Chair of the Jury.

Reviewer's Report

Reviewers report should contain the following items:

- Brief evaluation of the thesis quality and overall structure of the dissertation.
- The relevance of the topic of dissertation work to its actual content
- The relevance of the methods used in the dissertation
- The scientific significance of the results obtained and their compliance with the international level and current state of the art
- The relevance of the obtained results to applications (if applicable)
- The quality of publications

The summary of issues to be addressed before/during the thesis defense

Brief evaluation of the thesis quality and overall structure of the dissertation.

The thesis is of high quality, its structure is appropriate for the subject, assessing the evolution of RNA expression within and between species. It is divided into two main part, between and within species evolution.

The relevance of the topic of dissertation work to its actual content

The content and topic of the dissertation exactly align.

The relevance of the methods used in the dissertation

This dissertation uses high quality experimental and analytical methods to address the questions raised

The scientific significance of the results obtained and their compliance with the international level and current state of the art

The results obtained and presented in the thesis are of high relevance to the current state of the art. The thesis will advance our understanding of the evolution of gene expression.

The relevance of the obtained results to applications (if applicable)

Not applicable

The quality of publications

The thesis has high quality. Its subject is interesting, the experimental methods are well chosen to address the question, and the analysis is well done.

I have a few minor comments, which do not have to be addressed before defense of the thesis.

Chapter 3

The samples used for the study are very well chosen, coming from the same location and treated equally. There is still a possibility that population grouping of samples correlates with

socioeconomic status, cultural upbringing and others. It would have been nice to stress this influence a bit more.

Page 45: "The results showed population identity and baby's sex are two major factors." Since there is very little difference in % explained variance between baby's sex, first pregnancy, mothers age, bmi, delivery type, I would say "the two largest factors", or maybe say population identity was the major factor, with 7 other factors having a lesser effect, though still above 3% of the variance."

Page 47, Figure 4. Dots are too big. Is it possible to fix that?

Chapter 4

Page 64. "the largest variance (>30%) in gene expression can be explained by different developmental stages (Figure 18)." It is unclear what in figure 18 relates to this statement.

Page 67. Figure 19, and also in text. I think that calling genes that change their expression in development "develop-related" genes is a bit confusing. That label implies that they control development, not that they are controlled by development. Maybe "development responsive" genes?

Page 84: "most (85%) precursors are in intronic or intergenic regions" it would be nice to state if this number is consistent with expectations, maybe based on relative size of genomic regions"

Page 87: This is quite a complicated question, with a complicated setup and analysis. I had a hard time keeping track for each analysis which miRNA are used, organ-specific, how many developmental stages, how many species, etc. Maybe a schematic overview that also shows that different analyses done would be beneficial.

Page 88: It would be nice to expand upon the statement: "This observation is consistent with the notion suggested by Chen and Rajewsky, postulating that most organ-specific miRNAs might function to minimize deleterious off-targeting effects and to allow natural selection to eliminate slightly deleterious targets over evolution"

Page 91, Figure 44: make sure that each of the histograms has 5 bars, even when proportion is 0. (unless some species are missing stages for some tissues, in which case this should be emphasized more.)

Page 94, figure 47: Why isn't this displayed in terms of fractions, since the text on page 93 refers to fractions?

Page 95. "we used non-organ-specific miRNAs with dynamically changed at least one organ", I think it should be "we used non-organ-specific miRNAs **which** dynamically changed at least one organ"

Page 95: The lower part of this paragraph is very unclear. Too much information compressed into too little text. Break it up, explain more.

Page 98: “whether miRNAs with the same seed, which has the same regulatory functions as usually assumed,” I didn’t understand this sentence

Page 98: Figure 53. Explain this figure more. What is the difference between horizontal and vertical, what is being compared, what can the reader see?

Page 99, 100: text on page 99 says 55A is increasing, but figure legend says A is decreasing. Page 100, Figure 54 is unclear. What do the arc colors indicate? What kind of overlap is plotted?

Page 101, Figure 55. Remove top trees. Two species always show the same tree. There should be more explanation of what the reader is to infer from the figure. Top is a bit darker, does that say anything? Columns are more similar? What do trees say?

Again, increasing, decreasing, organ consistent, etc, are very confusing. It might be good to also have an overall schematic of the analyses.

Provisional Recommendation

I recommend that the candidate should defend the thesis by means of a formal thesis defense

I recommend that the candidate should defend the thesis by means of a formal thesis defense only after appropriate changes would be introduced in candidate’s thesis according to the recommendations of the present report

The thesis is not acceptable and I recommend that the candidate be exempt from the formal thesis defense