

Jury Member Report – Doctor of Philosophy thesis.

Name of Candidate: Aleksei Mironov

PhD Program: Life Sciences

Title of Thesis: Tissue-specificity and regulation of aberrant alternative splicing

Supervisor: Assistant Professor, Dmitri D. Pervouchine

Name of the Reviewer: Assistant Professor Ekaterina Khrameeva

I confirm the absence of any conflict of interest (Alternatively, Reviewer can formulate a possible conflict)	Date: 06-09-2022
--	-------------------------

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

- **Brief evaluation of the thesis quality and overall structure of the dissertation.**
The thesis presents an excellent work characterizing aberrant splicing based on the analysis of high-throughput sequencing data. It covers tandem alternative splicing sites, their tissue-specific expression, impact on protein structure, evolutionary selection (Chapter 5) and unproductive splicing, including its tissue-specificity and regulation by RBPs (Chapter 6).
- **The relevance of the topic of dissertation work to its actual content.**
The topic of the thesis matches its contents well.
- **The relevance of the methods used in the dissertation.**
Methods used in the thesis are relevant and applied correctly, to my best knowledge, in all presented analyses. The used methods are well described and presented with enough details.
- **The scientific significance of the results obtained and their compliance with the international level and current state of the art.**

The presented research employs advanced data analysis methods and relies on the current state-of-the-art datasets, therefore coping with the international level. Few studies cover aberrant splicing, which might be important in specific physiological conditions. Therefore, rarely expressed aberrant splice isoforms are poorly characterized. This thesis fills the missing gap, describing two types of aberrant splicing events: tandem alternative splicing sites and unproductive splicing events.

- **The relevance of the obtained results to applications (if applicable).**
- **The quality of publications.**
High enough to pass the PhD program requirements.

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

The thesis is clearly written and I have few comments regarding its content and presentation of the results.

The description of used methods is provided with enough detail in general. Yet, I would appreciate if the author could clarify several issues. First, regarding the catalogue of annotated splice sites. It is written that they “were extracted from the comprehensive annotation of the GENCODE database v19 [174] and from UCSC RefSeq database [175]”. But what if these two databases were conflicting each other? Or simply a union of annotated splice sites was taken here? I suggest to add more details on this procedure, to enable reproducibility. Second, regarding the read mapping procedure applied to the GTEx data. Was it a *de novo* mapping, not using information about annotated splice sites? I think this detail is important for the presented analysis and should be specified here. Third, regarding differential expression analysis procedure (page 46). It relies on each tissue against all other tissues comparisons. Potentially, this procedure can be biased if tissues are not uniformly presented in the dataset. For example, if the dataset contains too many samples representing similar brain tissues (e.g., many cortical areas) with similar expression profiles, gene expression differences attributed to these brain tissues might be under-represented by this analysis strategy. I think this potential issue should be discussed in the thesis. Fourth, why marmoset and galago genomes were chosen in the evolutionary analysis (page 51)? Was their genome assembly and annotation quality sufficient? I think it should also be discussed in the thesis.

All results presented in the thesis seem to be solid. Yet, a few of them could be a bit improved. At page 85, the author writes that “significantly expressed miSS preferentially affect disordered protein regions, and tissue-specific miSS are found in disordered regions even more frequently (Fig 5-28).” Is there a way to estimate the statistical significance of the observed increase? At page 89, it is written that “a significant positive selection was detected in maSS and constitutive splice sites (Fig 5-33, A, right).” I suggest to mark this significant observation with a star (or stars) at the Fig. 5-33. It would simplify perceiving of this result. At page 91, it is written that “miSS are considerably weaker (Fig 5-35, A).” Is this decrease significant? If yes, it should be also marked with a star in Fig. 5-35.

The thesis contains very few typos and grammar issues. For example, there is one at page 95, line 5 (“trhousands”). Also, there is a missing comma at page 95, line 18 (after “DLG4”).

Provisional Recommendation

V 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*