

Jury Member Report – Doctor of Philosophy thesis.

Name of Candidate: Mariia Vlasenok

PhD Program: Life Sciences

Title of Thesis: Transcriptomic analysis of the interaction between pre-mRNA splicing and intronic polyadenylation

Supervisor: Associate Professor Dmitri Pervouchine

Name of the Reviewer: Ivan V. Kulakovskiy

I confirm the absence of any conflict of interest (Alternatively, Reviewer can formulate a possible conflict)	Date: 16-11-2023
--	-------------------------

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

(Please see below)

It was a pleasure to review the thesis of Mariia Vlasenok entitled "TRANSCRIPTOMIC ANALYSIS OF THE INTERACTION BETWEEN PRE-MRNA SPLICING AND INTRONIC POLYADENYLATION".

The presented research describes a large-scale systematic analysis of alternative polyadenylation sites, their usage across cell types, and their relation to splicing.

The presented text follows the traditional scheme, including a concise and up-to-date literature review, a well-structured Methods section, followed by Results, Discussion, and Conclusions. The employed methodology is adequate, and the accompanying statistical analysis is fully appropriate as well. I do not doubt the level of the presented research, which conforms to the highest international standards in the field. I consider the results of the applicant to be valuable not only in terms of particular novel scientific findings related to the diversity, abundance, and localization of alternative polyadenylation sites in the human genome but also as a large-scale data source to fuel further studies of polyadenylation of spliced transcripts in different human tissues. As a side note, I am truly impressed by the language of the dissertation with very few mistypes, nicely prepared illustrations, careful formatting, and the overall effort spent on making the text reader-friendly.

Yet, with all the positive impressions, I nonetheless have several critical comments and questions.

(1) The text is overflowed by acronyms and their usage could have been more careful, e.g. the title of section 5.2.4. is a certain pinnacle: "Abundance and tissue-specificity of STE, CTE and SPI iPASCs".

(2) The list of publications (page 4) misses certain metadata of the applicant's papers, such as page numbers and DOI identifiers. Also, most of the thesis text is written from a first-person point of view. I expect that all the research described in the text was indeed fully performed by the applicant, but it would be beneficial to explicitly state the author's contribution to each of the publications.

(3) The Introduction lacks literature references for key claims to better substantiate the setting of the study.

(4) Page 31: it is unclear how particular precision and recall metrics relate to 'the best performance'. Were they selected based on e.g. an optimal F1?

(5) A review of PAS identification methods in Chapter 2 would benefit from a brief discussion of the applications of modern long-read and whole-transcript sequencing (e.g. Nanopore).

(6) Throughout the text, the authors repeatedly refer to the 'expression' of polyadenylation sites, which might be a common jargon but to me seems strongly confusing and incorrect. I would suggest referring to PAS usage instead, as it seems widely accepted in the field.

(7) Page 43: it is unclear why different read mapping tools were used for RNA-Seq and 3'-seq data, as both STAR and HISAT can perform splicing-aware mapping.

(8) Page 44, 1st paragraph: it would be useful to explicitly show the fraction of gene annotations unused due to overlaps. What do you mean by "all bookended" intervals (same paragraph)?

(9) Page 49: 1st paragraph, Qi is not defined; 2nd paragraph: how many shuffles were generated per sequence? Can we consider the ddG estimates to be stable in the case only a single shuffled sequence was used as a control?

(10) Page 52: there are explicitly provided normalization factors, but without extra information or a dedicated clarification it is impossible to judge whether they are adequate or not.

(11) Page 54: if you consider the impact of PCR duplicates, why duplicate filtering (e.g. with Picard) was not performed explicitly?

(12) Page 55: how and when did you filter PAS in adenine-rich regions? How these regions were defined in the first place? These details seem to be missing from Methods.

(13) Page 56: ".. genes .. to be expressed in .. datasets". Genes are probably expressed in cells, not in datasets.

(14) Page 64: by IQR of PAS here you mean the IQR of PAS borders or locations, right?

(15) When switching from a small-scale 'benchmarking' study of PAS found with RNA-Seq versus 3'-seq to a large-scale study with GTEx data, there were more stringent filters introduced, which, for some reason, were not re-tested on a smaller-scale benchmarking data, although it is of interest to know if it could improve the precision/recall of PAS recognition achieved in benchmarking.

All in all, these issues, while requiring minor modification of the thesis text or discussion during the thesis defense, do not affect my overall highly positive impressions from reading the text and the scientific rigor of the presented research. I wholeheartedly recommend the candidate defend the thesis by means of a formal thesis defense.

Sincerely yours,

Ivan V. Kulakovskiy, PhD, DSci, Prof

IPR RAS, Pushchino

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