UNIVERSITY OF MINNESOTA

Duluth Campus

Department of Biomedical Sciences
Medical School Duluth

217 Med
1035 University Drive
Duluth, MN 55812-3031

Office: 218-726-7911
Fax: 218-726-7937
http://www.med.umn.edu/duluth
Email: biomed@d.umn.edu

February 26, 2018

Department of Molecular and Cell Biology and Genetics
Faculty of Science
University of South Bohemia

To whom it may concern,

This document is my review of the written portion of **Sameer Dixit**'s PhD thesis in Molecular and Cell Biology and Genetics. The body of work contained in his thesis overall is quite impressive. There is no doubt that the material it contains is a valuable contribution to the field.

The topic of Mr. Dixit's thesis is RNA editing; specifically, the role of RNA-binding proteins in editing. The format of this thesis is: chapters consisting of published or soon-to-be published works in the subject area of mitochondrial U-indel editing. These include a portion of a review that I myself frequently cite in my grant proposals and manuscripts, including citation of the material written by the candidate included here. Additionally, Mr. Dixit has included a chapter containing preliminary analysis of an tRNA editing enzyme binding to non-tRNA targets. Additional sections are: one describing another unpublished study of effects of depletion of an RNA binding protein, and a proposal for a technical advance building off the published material. These additional sections demonstrate that the breadth of Mr. Dixit's contribution to his field is larger than that revealed by his current publication record. Finally, a thesis introduction places the published material in context.

At the heart of this thesis is a technique called individual-nucleotide-resolution UV-cross-linking and immuno or affinity purification iCLIP (or iCLAP). Use of this technique has allowed superior analysis of targets of RNA binding proteins of interest. Mr. Dixit himself is the investigator who adapted this technique for use in trypanosome mitochondria, with its challenging and unusual transcriptome target, and thus, has moved the field forward. The novelty of the approach and its development, described in Section 6, is the major contributing factor to the scientific excellence of the thesis.

As is typical of a thesis that contains largely published or pre-reviewed material, the use of language in those portions of the manuscript is excellent. Other sections such as the Introduction and unpublished sections reflect less care in both arrangement and language. One helpful lesson from this thesis is the difficulty in keeping pace with science: although the review to which Mr. Dixit contributed the RNA editing section was published in 2015, its sections on RNA processing (not included here, as they are not Mr. Dixit's work) are no longer up-to-date. I suspect that this thesis's introduction relies heavily on this older background material. Consequently, a number of new and influential papers are not cited. Nevertheless, Mr. Dixit is clearly master and owner of the new knowledge presented here.

As Mr. Dixit is now the world expert on his particular topic, I appreciate the opportunity to be here today to ask questions about it. The questions follow the body of this review. In my opinion, Mr. Dixit's thesis is higher in quality than the average PhD thesis of Integrated Biosciences at the University of Minnesota, a multi-campus graduate program of which I am a faculty member. The science is on par with that of typical PhD graduates of the Biochemistry, Molecular, and Cell Biology field of my graduate institution, Cornell University. **Sameer Dixit definitely qualifies for the PhD degree for this work**. He is well prepared to embark on the next step of his training, and I expect him to have a fulfilling and productive experience as a postdoctoral researcher.

Sincerely,

Sara Zimmer Sara Zimmer

Sara Zimmer, Assistant Professor
University of Minnesota

Questions

Introduction (Section 2):

What can more recent studies add to your description of RNA processing contained in this introduction? [helpful may be: Suematsu, T et al, 2016, Mol Cell. 61:364-378; R. Aphasizhev et al, 2016, RNA Biol. 13:1078-1083].

What can more recent studies add to your description of 3' tail addition on mRNAs, and RNA turnover, in trypanosome mitochondria? [helpful may be: Gazestani V et al, 2016 RNA 22:477-486; Gazestani V et al, 2018 Int. J Parasitol. 48:179-189; Zhang L et al. 2017, EMBO J 36:2435-2454].

You say "the transcribed gRNA population contains much more combinations then needed to execute editing, yet the reason for such redundancy remain unknown." There must be theories as to why this exists. Can you cite the study in which you obtained this information, and any hypotheses for this phenomenon?

Section 3, published in mBio:

The thesis describes a huge conglomeration of proteins involved to various extents in mitochondrial U-indel editing, and describes pull-down experiments in which associations between them were identified. Many of these proteins appear to be RNA binding proteins. How do you know that the iCLIP/CLAP methods reveal actual binding associations of the pulled down protein with the identified reads, rather than associations in which other members of the protein complex are the actual RNA binding component? It is possible that this question may be addressed in the formal oral presentation.

Fig. 4C,D are experiments where either RNA binding proteins MRB8170 or MRB4150 are pulled down, the co-precipitated RNA extracted, and qRT-PCR performed for mRNAs of interest, relative to their abundance in the total lysate. I notice that the relative abundance of edited transcripts bound to MRB8170 are 100-fold greater than their pre-edited counterparts. For MRB4160, they are 10,000-10,000,000-fold higher (if my math is correct). This fact is not addressed in the text. Please comment on 1. Whether or not this supports your model of these two proteins "marking" mRNAs for editing, and 2. what it might mean that the homologues' ratios are so different.

Section 4, under review:

You demonstrate that loss of MRB8170/4160 can cause loss of some proteins from positions on mRNA. Can you clarify which of the proteins that are part of MRB1, or other associated proteins or complexes that you study, have RNA binding domains of their own, and could theoretically bind RNA without MRB8170.4160. Explanation was provided as to why TbRGG2 may require co-binding with MRB81670/4160, what about the other players? Finally, according to your model, MRP1/2 is regulating efficiency of editing by differentially crowding out TbRGG2 that facilitates editing progression from MRB8170/4160 binding sites on various transcripts. If so, then what causes MRP1/2 to bind in a superior manner on some transcripts rather than others? Also, is there any evidence to suggest that the regulation isn't the other way around, that TbRGGs isn't the one crowding MRP1/2 out from binding sites?

Section 6: technical proposal:

Although high-throughput sequencing-based CLIP technologies are increasingly common, few researchers are completely familiar with the details of the process. Therefore, technical background and definitions are necessary for readers to make use of your peak-calling proposal. For example, exactly what IS peak calling? When you acquire read data in your experiments, is it possible to tell what portion of the read was directly in contact with the protein (crosslinked), versus merely near the protein that was pulled down? This may be something you will explain in the formal oral presentation.

You propose a new program to generate a more accurate image of mitochondrial mRNA targets bound to RBPs via iCLIP/iCLAP. The thesis describes the problems to be: unmappable reads that are non-canonically edited, differential abundances of fully- to partially- to pre-edited reads in the total population that are unknown, and ambiguities in assignment of a read to a totally pre-edited or fully-edited versus partially edited transcript. Which of these problems can be addressed by your proposed methodology?

Section 7, Unpublished study #1:

The discussion in this section mainly reiterates the results of the study to date and stops there. Do these results suggest anything, or provide insight into the associations or steps of post-transcriptional processing complexes?

Section 8, Unpublished study #2:

Reads from 18 non-tRNA targets were found bound to TbADAT2/3 that normally binds and acts on tRNAs. Are tRNAs also picked up with this method used? If so, what was the ratio of tRNAs to non-tRNAs that were collected? Can you provide evidence of confidence that you are collecting non-tRNA reads at a higher amount than expected for "background noise"? What controls if any would you suggest the next person do, going forward?

Of those 18 non-tRNA targets, how many are included in Table 1/Figure 2, which is a list of gene ontology terms represented by these targets? Given this number, can you justify utilization of a standard GO enrichment analysis to support a model or hypothesis?

The read coverage of the three mRNAs shown in Fig. 1B definitely show a pattern. Have you investigated positioning on any other transcripts? Is this a commonality, or are you showing the subgroup of the population that exhibits this pattern?

End of question section

Brno, 28.2. 2018

Evaluation report on the PhD thesis of Sameer Dixit

I have critically and thoroughly read the Ph.D, thesis "The role of RNA-binding proteins in post-transcriptional gene regulation of *Trypanosoma brucei*, submitted by Mr. Sameer Dixit and supervised by Prof. Julius Lukes. The thesis consists of 11 pages long Introduction, one first author original published paper, one co-author review, one first author unpublished manuscript and three chapters describing preliminary results. The main unifying theme of Sameer's PhD work is using the CLIP methodology to reveal spectrum of RNA targets of several RNA-binding proteins (RBPs) of Trypanosoma brucei and thus help to understand their role in RNA processing and regulation of gene expression. Most of the studied factors include RBPs operating in the Trypanosoma mitochondrion and acting in RNA editing of mitochondrial mRNAs. The topic is timely and highly interesting with a potential to uncover yet unknown molecular mechanisms. The thesis is unfortunately written in a rather reduced manner. It lacks few essential parts and chapters, such as Aims and overall discussion of the whole thesis findings, conclusions and ideally also an outlook, that would demonstrate the applicant's readiness as a future researcher. It would greatly improve the overall impression of the undoubtedly good work that Sameer has done, if the thesis was presented in a way, where he would interconnect logically the individual projects. Considering, that the main methodical aproach used was the CLIP-seq technique, it would help to include a separate chapter with the method description, comparison of several protocols used to date by various groups (e.g. PAR-CLIP, iCLIP. eCLIP, etc) and explain what optimization was needed for Trypanosoma mt RBPs. Moreover, the thesis lacks few essential parts and chapters, such as Aims, list of abbreviations, and overall discussion of the whole thesis findings, conclusions and ideally also an outlook, that would demonstrate the applicant's readiness as a future researcher. The two published works (original paper and the review) underwent already a critical review process therefore my following comments and questions concern mainly the unpublished parts.

1. There are numerous typos and several imprecise statements in the thesis. For instance, already the Annotation on the second page (misspelled as Anotation) includes the following statement that I find difficult to understand: 'Further, this thesis provides evidence where the combinatorial interplay of RBPs might fine-tune the levels of edited mRNA.' or the part "Summary", page, third sentence:

Trypanosomatids are best studied model organism from the early branching kinetoplastids. Which ones are "late branching kinetoplastids"?

2. I noticed frequent improper referencing, often using the review from Read et al., 2016 or reviews from the supervisor (Lukes) lab instead of referring to the original works. Here are few examples;

Introduction, page 6: The reference to Read et al., 2016 and Verner et al, 2015 for the sentence The entire process of U- insertion/deletion is called RNA editing." is incorrect. I believe, the term RNA editing has been coined by other team(s) much earlier. ...or the last sentence on page 6, the absence of mitochondrially encoded tRNAs and their import has been uncovered much earlier than in 2011..and studied by Steve Hayduk and Larry Simpson 1990s.

page 7, Chapter 2.3: states that that transcription initiation at minor strand of maxicircles has not been determined till 1992 (as it cites Michelotti et al. from 1992) - so what is the current knowledge status after almost 30 years?

page 8: Lukes et al 2011 is not the correct reference for the finding that transcription of maxicircles involves several RNA processing steps before translation...

page 12: "Editing starts from the 3'end of a pre-edited mRNA..." was not discovered by Hashimi, nor Read.

etc.

3. Chapter 2. The submitted manuscript contains results of the iCLIP-based RNA-binding studies of MRP1, a protein involved in early steps of mt mRNA editing. The author performed two replicate iCLIP experiments. There are strange numbers of reads resulting from each replicate sequencing: 123,8207 and 589,9639. I suppose, there was a mistake in the comma. Nevertheless, how does author explain such a dramatic difference in number of reads in the individual experiments? The data are subsequently compared to the bidning of other two RBP editing factors. One dataset is from the first chapter of the thesis, the CLIP results of DRBPD18 were published by a different group. The text does not describe, whether the DRBPD18 was also an iCLIP experiment.

4. In addition, the bioinformatics analyses are often a bottleneck of the high-throughput sequencing studies. The thesis contains a rather brief chapter on page 81 - 84 without any logical integration within the whole text. I would welcome a much more elaborate chapter describing what scripts and

bioinformatics tools have been tested, modified in order to analyze such complicated data as the reads containing unencoded sequences.

5. The last two chapters of the iCLIP studies of MERS1 on page 85 -89 and of tbADAT2 on pages 90 - 97 stand completely detached from the first two thirds of the thesis.

6. I find the work in Chapter numbered as 4 very interesting, but also highly speculative. It includes rather preliminary results on the *T. brucei* ADAT2 protein that is a homologue of the tRNA deaminase. Why were the iCLIP-seq results are mapped primarily to mRNAs? Were any tRNA targets identified in this experiment? Concerning the bound and annotated reads as mRNAs, the work found binding of TbADAT2 mostly at the 3' mRNA termini. Since ADAT is an RNA deaminase, did the authors check for A to I editing in the bound reads or mRNAs? Did they check for predictions of secondary structure formation - for instance, if a tRNA-like structures could be formed. This feature has been previously demonstrated for several human mRNA 3' ends.

Additional questions from a non-parasitology person:
1. Are really all genes in all trypanosomes transcribed in polycistronic fassion as stated in the Introduction?
2. What is the difference in the six Trypanosoma mt mRNAs, that do not undergo RNA editing?

Despite the many critical comments, to my opinion, Mr. Sameer Dixit proved to be an independent young researcher and therefore I fully recommend this PhD thesis for a defense and support his graduation. Although, a better presentation of his whole PhD work would significantly improve the quality of the thesis.

Doc. Stepanka Vanacova, Ph.D.