Dr hab. Agata Starosta, prof. IBB PAN
Instytut Biochemii i Biofizyki Polskiej Akademii Nauk
Pawińskiego 5a, 02-106 Warszawa
agata.starosta@ibb.waw.pl
tel. 22 592 3341

Warszawa, 15.03.2025 r.

Review of the doctoral dissertation of Monika Kwiatkowska

Doctoral dissertation entitled "Genomic characterization of long-noncoding RNAs in the zebrafish genome" was prepared by Monika Kwiatkowska under the supervision of professor Barbara Uszczyńska-Ratajczak and dr. Sílvia Carbonell Sala. The work was submitted to the Scientific Council of the Institute of Bioorganic Chemistry PAS in Poznań.

The main goal of the work presented in the thesis was to unveil previously missed long-noncoding RNAs (lncRNAs) in a vertebrate model organism – zebrafish. Results presented in this thesis are novel.

The dissertation is written in English. The content of the dissertation is consistent with the subject indicated in the title. The layout of the thesis is appropriate and correct for this format. The thesis is divided into a number of chapters: *Abstract, Introduction, Objectives, Materials and Methods, Results, Discussion, Conclusions* and *Perspectives*. The author used an extensive bibliography referring to 158 scientific publications. The author uses language and terminology appropriate for this format. The table of contents is prepared correctly. The work is clear and factual. The text is provided with 14 figures in the *Introduction* and 45 figures in the *Results* section, 12 supplemental figures and 4 tables? The figures are prepared carefully, the descriptions on them are clear. The list of all figures is missing. The bibliography was used appropriately in the text to introduce the subject and to discuss the results.

The author begins with an *Introduction* with a concise description of the current knowledge on lncRNAs. She describes what lncRNAs are in a RNA fraction of the cell, classes of lncRNAs, features of their architecture, possible functions, their evolution, including evolution of functions, conservation of lncRNAs across the species, including sequence, structure, function conservation and positional conservation (localization in the genome) which is later expanded in the studies by Monika Kwiatkowska. PhD candidate explains the feature of zebrafish which makes it a great model organism for vertebrates. The author summarizes current knowledge gap regarding identification and annotation of lncRNAs. She describes available approaches to genome annotations, which she is also applying in her work. The PhD candidate finished her *Introduction* with a description of the experimental approaches which she is using, including CapTrap-seq which allows to enrich for full length RNAs with a 5'cap and a polyA tail, and a template switching oligo (TSO) protocol which enrich a polyA-tailed RNAs.

In the next chapter, the author describes *Objectives* focusing on acquiring accurate, comprehensive annotations of lncRNAs in zebrafish, including tasks related to (1) identification of the orthologous lncRNAs between zebrafish, mouse and human; (2) improvement of lncRNAs annotations using long

reads NGS protocols and Nanopore platform; (3) improvement of the detection of lncRNAs with low abundance to build an extra layer of information of the RNA fraction of the zebrafish model.

In the *Materials* and *Methods* section, the author provides lists of reagents and equipment used in the experiments. The PhD candidate describes protocols she used to carry out her research tasks. She describes protocols for library preparations for approaches used in this work, and for data analyses.

Results are divided into three sections: (A) optimization of full-length long-reads RNA sequencing, (B) optimization of isolation and sequencing of the low abundant RNAs, (C) attempt to characterize the function of the identified RNAs.

In the first part of the Results (A), the author begins with the description of sample preparation, library preparation and sequencing using the Oxford Nanopore Technologies platform. She presents raw reads distribution of the RNAs prepared either in CapTrap-seq or TSO approach, across all tested conditions, including 2-4 cell (early developmental cell), 28hpf (cell 28 hrs post fertilization) from embryos and heart and testis from adult zebrafish. It remains unclear to the reviewer if this was a single experiment or if the biological replicates were performed. The PhD candidate is demonstrating differences between the output of the CapTrap-seq and TSO approach, including error rate, gene biotype detection, detection of polyadenylated reads, and detection of full length transcripts. Here, I would like to ask the PhD candidate to explain the detection of non-PolyA RNAs since the polyA-enrichment was applied for library preparations. She then applies a size-selection cutoff of 500 bp to increase longer cDNA molecules and applies CapTrap-seq and TSO approach. She again analyses the detection of polyadenylated reads, the detection of full-length transcripts, transcript length distribution, annotations of transcripts, novel loci identifications, and novel transcript models (alternative transcripts). The results reveal a significant number of novel loci and novel transcript splice forms. This data is increasing vastly the current knowledge of RNA molecules present in zebrafish, which makes it an even more compelling model organism to study.

In this part, the step of size cutoff is not clear to me and I would like to ask if the PhD candidate could try to explain what fraction of lncRNAs in a length range of 200-500 nt could be missed from the datasets?

I am missing here an analyses comparing lncRNAs and mRNAs. It would be beneficial if data was spilt into separate datasets and reanalyzed in a way described above. I would also be interesting to see what is the overlap of detected lncRNAs and mRNAs between different approaches of libraries preparations , as well as analyses between different tissues, developmental stages (for example: Venn diagrams, Volcano plots, heat maps showing patters for different tissue/organs, PCA).

In the second part of the Results (B), the author applies a Capture Long-Read Sequencing (CLS), a targeted RNA sequencing technique. The PhD candidate designs specific probes for CLS approach and combines a CLS protocol with the CapTrap-seq approach. She achieved a significant enrichment of RNAs from all tested samples for further analyses. However, can this type of analyses be quantified? What does the increased detection rate of transcript tell about the sample? Similarly as above, I am missing the analyses comparing the exact RNAs between pre and post approach. Are you detecting the same transcripts or are there new ones? What is the length distribution of identified lncRNAs and mRNAs (as separate sets).

In the third and the last part of Results (C), attempts to characterize identified lncRNAs functionally. This chapter is confusing as it implies the analyses of lncRNA, while the PhD candidate is showing data and imaging for various small non-coding RNAs. I am missing some more explanations in this chapter. Moreover, showing localization of the RNA or pattern of expression does not reveal the function of it.

In *Discussion*, the author summarizes her achievements. In *Summary*, the author shows approaches that allow for the identification of novel lncRNAs loci and transcript splice forms. She builds the first comprehensive database of lncRNAs in zebrafish and focuses on positionally conserved genes.

Critical Remarks:

- It would be interesting to see how data from CapTrap-seq and TSO, both using a polyA enrichement approach, compares to rRNA depleted samples. Can you comment what could be expected? What is the expected percentage of polyA-tailed and circular lncRNAs? Would Illumina seq benefit to be included as a control?
- Can the data be used in a quantitative way or only qualitative? The approached used here, would favor some but not all RNAs.
- What is the length distribution of the lncRNAs in zebrafish? What could be the fraction of lncRNAs in a length range of 200-500 nt?
- Are the identities of lncRNAs between different protocols are similar? Venn diagrams / Volcano plots between data sets would be interesting to see. If there are more introns detected in the TSO approach, are the introns-absent, mature lncRNAs at similar levels as for CapTrap-seq? is this approach giving an extra insight into RNA processing? Could rRNA-depletion approach and sequencing of all RNAs give and extra insight?
- Is the data prepared for a single experiment only?
- If some lncRNAs are coding short peptides, why are they still considered noncoding?
- Why are some lncRNAs from Figures 4.10, 4.17, non-polyA if the polyA-enrichment was used as a selection method?
- Please explain hg38 and mm10 from Figure 4.23, hg38, mm10, danRer11 from Figure 4.24. What are they? Why they were selected? Description is missing as well as explanation of the rationale behind choosing this particular genes.
- Why dapB is a negative control? The choice of controls is not explained.

Despite the critical remarks, I evaluate this PhD thesis *positively*. One can see the enormous amount of work put into the research, even if not all the intended goals were achieved. These results are a significant contribution to the characterization of the long non-coding RNAs of the zebrafish model organism, fill a gap in knowledge and will be the basis for many future studies. The application of long reads sequencing greatly expands the transcriptome of the zebrafish. Not only new lncRNAs were uncovered but also new splice-forms of protein coding RNAs. This achievement is of great importance for the field and future work with zebrafish.

———

Niniejszą pracę oceniam *pozytywnie*. Widać ogrom pracy włożony w badania, nawet jeśli nie wszystkie zamierzone cele zostały osiągnięte. Wyniki stanowią znaczący wkład w charakteryzację długich niekodujących RNA organizmu modelowego danio pręgowanego, uzupełniają lukę w wiedzy i będą podstawą wielu przyszłych badań. Odkryto nie tylko nowe lncRNA, ale także nowe formy splicingu RNA kodującego białka. Osiągnięcie to ma ogromne znaczenie dla dyscypliny i przyszłych pracy z danio pręgowanym.

Przedstawiona do recenzji rozprawa doktorska spełnia warunki określone w Ustawie z dnia 20 lipca 2018 roku prawo o szkolnictwie wyższym i nauce (Dz. U. z 2018 r. poz. 1668 ze zm.) oraz w Sposobie postępowania w sprawie nadania stopnia doktora w Instytucie Chemii Bioorganicznej PAN w Poznaniu (uchwała Rady Naukowej ICHB PAN nr 28/2024/Internet z dnia 20 marca 2024 r.) i wnioskuję do Rady Naukowej Instytutu Chemii Bioorganicznej PAN i dopuszczenie mgr inż. Moniki Kwiatkowskiej do dalszych etapów postępowania o nadanie stopnia doktora.

Dr hab. Agata Starosta, prof. IBB PAN

INSTYTUT BIOCHEMII I BIOFIZYKI
POLSKIEJ AKADEMII NAUK
ul. Pawińskiego 5A, 02-106 Warszawa, Polska

www.ibb.edu.pl secretariate@ibb.waw.pl
Tel.: +48 22 592 21 45, Faks: +48 22 592 21 90
NIP: 5261039742, REGON: 000325819