Skoltech

Skolkovo Institute of Science and Technology

Jury Member Report – Doctor of Philosophy thesis.

Name of Candidate: Anna Shiriaeva

PhD Program: Life Sciences

Title of Thesis: Interference and primed adaptation intermediates in type I CRISPR-Cas systems

Supervisor: Professor Konstantin Severinov, Skoltech

Name of the Reviewer:

I confirm the absence of any conflict of interest	**Signature:**  **Date: 09-10-2020**

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.

Below

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

Relevant

- The relevance of the methods used in the dissertation

Relevant

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

High

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

In perspective

- The quality of publications

High

The Doctoral Thesis by Anna Shiriaeva entitled "Interference and primed adaptation intermediates in type I CRISPR-Cas systems" is dedicated to deciphering the exact structure of the prespacers produced during type I-F and type I-E CRISPR-Cas primed adaptation, towards comprehensive understanding of the whole machinery.

Really comprehensive review of the current literature summarizes a huge piece of knowledge by type I CRISPR-Cas system interference, adaptation – including prespacer generation, and possible role of DNA double-strand break repair in the latter process.

The reported results sound persuading, with a number of clear findings that justify the novelty of the work, confirmed by solid publications.

Some questions below I would ask the author to comment:

1) Not quite clear how self-ligation was avoided in FragSeq.

2) In general, not quite clear what was going on in the bioinformatic analysis of FragSeq:

- Were the reads that looked like two or more ligated fragments filtered off? Or "uniquely aligned" barrier worked for this aim? Not clear.
- Was the reads-per-UMI coverage analyzed? Look like it was not high, then how errors in UMIs themselves were accounted for?
- How many UMIs obtained for different strains? (this information becomes only quantitative if sufficient coverage was achieved)
- Were the processed, UMI-clustered data normalized before further mappings, using UMI (that would be rational)?

3) " Only reads with a length 16 to 100 nt uniquely aligned to the genome were further analyzed."

So this is artificial selection, since the library was like up to 500 bp.

Then: "The lengths of fragments mapped to the genome in all strains varied in the range of ≈16-100 nt with a median length of fragments outside the PPS-containing region of 45-47 nt."

Formally sounds strange: since the upper limit was artificially selected, one can say that estimation of median length is a nonsense. In reality, Figure 20 shows that this ok, but should be better discussed, starting from the point why 100 nt was selected as a length threshold.

4) Figure 16.

The text says that:

"No enrichment was detected in *cas1* mutant, *cas3* mutant, and the nontargeting strain (Figure 16B,C)."

In reality, there is enrichment in *cas1* mutant strain, if we look at a wider region, as it should be according to observed in whole genome sequencing, shown in Figure 13.

At the same time, there is a nice gap in the middle of this wide hill, specific for the *cas1* mutant strain. This observation is not discussed, while it could be (if it is not an artefact) a critical piece of the whole puzzle.

Like, in my naïve interpretation: *cas1* stabilizes the fragments produced by *cas3,* otherwise degrared, and this mechanism works in proximity to PPS region, while further DNA degradation beyond +/- 200,000 bp region is mostly driven by other enzymes.

Or more complicated story, but anyway the nice gap in the hill not discussed.

5) Figure 22.

Motifs are still enriched upstream in the *cas3* nuclease mutant. But relative abundance of these molecules is low, according to Figure 16. No comparative analysis here in terms of numbers, hard to interpret what it means.

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*