Review of the bachelor thesis
Age-associated changes of transposable elements activity in the oocytes and differences in their content across mammalian genomes
by Bădici Marco-Constantin

The bachelor thesis deals with the topic of transposable elements in mammalian genomes and their activity in relations to DNA methylation in mouse oocytes. The thesis is 157 pages long and divided into classical chapters such as Abstract, Introduction, Aims, Methods, etc. The text is well written with minimum grammar mistakes and typos, however, some of them change the meaning significantly (e.g. kbp x bp, 6 x 7 column).

The introduction provides a concise summary of topics related to transposable elements in mammalian genomes and their role in ageing and oocytes. The chapter reads well although there are some parts where the sentences do not follow in logical and smooth order. The aims of the study are well defined.

Materials and methods are described in many details; however, some crucial information is still missing, e.g. a number of replicas of the single-cell RNAseq. This chapter is supplemented with R and Python scripts, which are extensively commented. The code could have been more polished and sometimes there are easier ways around, but the student always found a working solution for the tasks. I would just suggest loading fewer packages into R, mainly because the tidyverse package set already contains ggplot2, tidyr, etc. so it is not necessary to use these libraries separately. I would also suggest the use of function pivot_longer (from tidyverse) instead of melt (from reshape2) as it is more user friendly in my opinion. This chapter does not contain an explanation/ meaning for the DNA methylation was measured (%of methylated CGs), moreover, while allocating the DNA methylation levels into4 categories, as there are gaps in between the used intervals (e.g. one category ends with DNA methylation 24.9 and the following one begins with 25). The Repeat Explorer Archive output also contains multiple tabular formatted files, which could have been used instead of the html for easies data manipulation. In my opinion, the increase in sample size for guinea pig does not make sense, as the sample size is for reads even in pair-end mode

The results are well supported by multiple tables and figures as well as statistical analysis and they are described well. However, there is a discrepancy between table 2 and its description in the text, as the table shows higher values for aged oocyte but the methylation is described as decreasing. The plots would need larger axis text and axis titles, as it is sometimes barely readable. In addition, the legends to some plots would need more care; legend titles such as “variable” are not informative. Moreover, some axis titles are more plot titles and the meaning of the corresponding axis is only described in the figure

legend. The mentioned correlation between DNA methylation and TEs expression in three TE groups, however, no correlation coefficients are provided.

The discussion is again well written and all main results are set into the context of published data. Some inconsistencies between results from different approaches are well discussed including hypothetical explanations.

Questions:

1) What was the coverage of data used in Repeat Explorer?
2) Based on the violin plot (figure 3) it does not seem any of the data has normal distribution. However, the Shapiro-Wilk test indicated this type of distribution in some cases, can you explain that?
3) Was the correlation between DNA methylation and transcription tested also in absolute values, not changes? If not, would you expect the same results as in the analysis in the present thesis?
4) What oocyte stage was used of single-cell sequencing? How this stage relates to DNA methylation erasure during oogenesis?

Despite the minor issues listed above, I consider this thesis to be very good. The author proved to be able to work with bioinformatical data, draw reasonable conclusions and present supporting evidence. In my opinion, the present thesis fulfils all requirements and I recommend it for a successful defence.

May 27th, 2021, České Budějovice

Martina Dalíková