Name of the student: Ms. Sylvia Ramírez
Thesis: Analysis of the expression of transcripts at imprinted loci across mammalian species during early development

Reviewer's statement

In her Bachelor thesis, Ms. Sylvia Ramírez studies the expression of transcripts at imprinted loci across mammalian species during early development.

Ms. Ramírez shows broad knowledge of the topic of her thesis. Information within the literature review is appropriate for the research topics studied. The aims are clear and fulfilled by the presented results. The results are mostly observational and descriptive. Nevertheless, Ms. Ramírez demonstrates promising expertise in a broad range of bioinformatics and next-generation sequence data analysis, but she didn't present the obtained results in a clear manner in some cases. Her findings contribute to the understanding of the expression of transcripts at imprinted loci across mammalian species during early development. Ms. Ramírez highlights the importance of imprinted regions in mammals, particularly, novel and developmental stage- or species-related regions. Moreover, the development of programming scripts which facilitate the bioinformatic identification pipeline of the imprinted regions. Finally, exploring the role of transposable elements in the imprinted gene expression. The thesis satisfies the formal criteria. Information is correctly referenced, including the figures. Figures are generally well designed but sometimes with unclear legends.

I have comments regarding the text :

The **abstract** without any descriptive information, and the **introduction** as well. While the **Background** section was well written and informative except the **section 3.5.1 "Non-coding RNAs, alternative promoters and transposable elements as important regulators of imprinting"** which was not well-organized, both sections **4-Aims** and **5-Workflow** overview need to improve while the **material and method** section could be made shorter without too much detailed information.
The **results** section was not well written with a lot of language error, which make it hard to understand. Also, the results presented in a non-scientific way, i.e.: vague expressions such as : **"almost all transcripts, many transcripts, For the majority of imprinting, In the majority of cases, In the majority of cases, a substantial proportion of transcripts"** were used without mentioned exact numbers and/or percentages. Also, some results were over-presented using table and figure,I.e.: the results of **"repetitive elements as TSSs in rat and cow"** were presented in **tables 12 and 13**, and showed again in **figures 10 and 11**. Moreover, other tables and figures need to merge, I,e.: **table 7 and 8** and **figures 12-36** in a way allow the readers to notice the differences between mouse, rate, and cow. Finally, the **discussion** section was informative but included a lot of vague expressions.

Additionally, i have few comments regarding the formal side of the thesis, pointing out minor issues that could be improved:

1. There is no need to mention the command line for each program, which makes the material and methods section hard to read. If it is necessary, all command line can be inserted as a supplement or appendix.

2. The python scripts should be deposited at one of the public platforms such as "https://github.com ", which will be easier to evaluate the scripts.

Questions for the author:

1. In section **6.6 Downstream analysis**: **Why you decided to perform the downstream analysis only for rat and cow?**, then **what was the importance of assembly the transcriptome of pig, marmoset, and macaque rhesus?**

2. In sections **7.1.1 Identification and selection of datasets for the analysis:** it was mentioned that the study particularly interested in RNA-seq datasets from the oocytes, embryos, and placenta. But according to **Figure 3**, the **placenta RNA-seq reads** were not found for both **rat and cow** species while the **embryos RNA-seq reads** were not found for the **rat** species.

3. According to **figure 3**, not all developmental stage RNA-seq datasets for all species were found in **the European Nucleotide Archive** (ENA, https://www.ebi.ac.uk). Did you try to search for the missing datasets in **NCBI, Sequence Read Archive (SRA)** (https://www.ncbi.nlm.nih.gov/sra)?

4. Is it imprinted regions or genes? What the difference between region and gene?

5. In **3.5. Imprinted gene clusters in mammals section:** it mentioned that imprinted genes clustered into 28 clusters. In which base?

6. In **python scripts**, why chromosomes and bases specified as a value?

Overall, i appropriate the time and effort of Ms. Sylvia Ramírez to produce such much results in a short time. Therefore, i recommend the thesis for defense with grade 1 (excellent).

In České Budějovice, 20.05.2019

Abdoallah Sharaf, Ph.D.