

BINF-402 Project

Differential expression analysis of micro-RNA transcriptome between pancreas, prostate and gastrocnemius medialis tissues

Léopold Guyot

November 30, 2023

1 Introduction

This analysis investigates microRNA expression variations across three distinct tissues; prostate gland, pancreas body, and gastrocnemius medialis. For this a simple workflow will be used, consisting of quality control of the reads, followed by a filtering, then a mapping to finish we a classic differential expression analysis. The goal is to unveil tissue-specific expression patterns of miRNA. Tissue selection is strategic, anticipating closer miRNA expression patterns between prostate and pancreas, both glandular, and unique signatures in gastrocnemius medialis, a muscle tissue.

2 Methods

All the data processing was done using the language R (R Core Team, 2023) and several packages.

2.1 Data Retrieval (retrieve_data.R)

All the data sets used in this project have been retrieve from the ENCODE database (Luo et al., 2020). Three tissues have been selectionned; pancreas body, prostate gland and the gastrocnemius medialis tissue. For each tissue, data used was coming from two distinct experiments, each comprising two replicates, thereby totaling four replicates per tissue (cf. "data/sample_table_links.csv" for file access numbers).

The original UCSC hg38 genome was used as reference for the mapping. The NCBI acces-

sion for this genome is GCA_000001405.26.

2.2 Read Quality Control

present QC before and after filtering

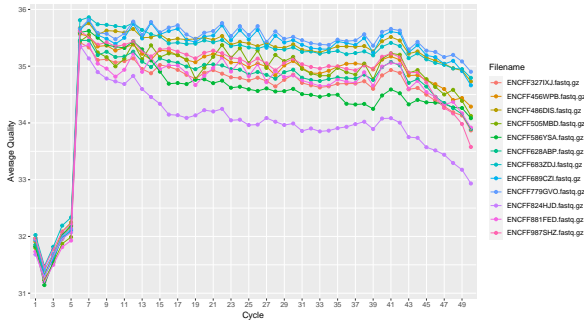


Figure 1: Caption of the figure.

2.3 Mapping

present mapping stats

2.4 Differential Expression Analysis

3 Results

Present the overall results => Differential Expression Analysis

4 Discussion

???? maybe not include it => things I could Improve

5 Conclusion

References

- Y. Luo, B. C. Hitz, I. Gabdank, J. A. Hilton, M. S. Kagda, B. Lam, Z. Myers, P. Sud, J. Jou, K. Lin, et al. New developments on the encyclopedia of dna elements (encode) data portal. *Nucleic acids research*, 48(D1): D882–D889, 2020.
- R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2023. URL <https://www.R-project.org/>.