



Integrate Salmon to MultiQC

CSE – 549 COMPUTATIONAL BIOLOGY

NOVEMBER 06, 2017

Alok Thatikunta | 111498776
Pratik Sushil Zambani | 111500248
Saahil Khurana | 111446700
Sweta Kumari | 111497926

PURPOSE

MultiQC is a reporting tool that parses summary statistics from results and log files generated by other bioinformatics tools. The purpose is to integrate Salmon, a transcript quantification and mapping tools for RNA sequences to MultiQC.

SCOPE

- Plot GC bias of Salmon runs on MultiQC
- Parser for sequence bias
- Plot 5' and 3' sequence biases for each nucleotide of Salmon runs on MultiQC
- Generate Heatmap of similarity between samples

APPROACH

GC-BIAS

- Parse binary input files of observed and expected GC bias of Salmon runs
- Compute the weighted linear combination for observed and expected values
- Plot the ratio of observed values to expected values for each sample using the MultiQC inbuilt line graph module

SEQUENCE BIAS

- Implement parser module to read sequence bias binary data file
- Plot the ratio of observed values to expected values for 5' and 3' for each of the 4 nucleotides (A, C, G, T) where each curve represents the marginalized o^{th} - order probabilities of the context

HEATMAP

- Calculate similarity between Salmon samples using Jensen-Shannon divergence

$$JSD(P \parallel Q) = \frac{1}{2} D(P \parallel M) + \frac{1}{2} D(Q \parallel M)$$

where $M = \frac{1}{2}(P + Q)$

P is Probability of first sample

Q is Probability of second sample

D is Kullback–Leibler divergence

- In heatmap, the divergence of 0 indicates more similarity, while divergence of 1 indicates that distribution is different.
- Similarity is calculated for each feature, viz. GC bias, each nucleotide for 3' sequence bias and 5' sequence bias
- Generate heatmap by taking average of similarities of all the features
- If a sample doesn't contain some feature, the similarity function excludes that feature while calculating the distance

REFERENCES

- <http://salmon.readthedocs.io/en/latest/>
- <http://multiqc.info/docs/>
- https://en.wikipedia.org/wiki/Jensen-Shannon_divergence
- <https://stackoverflow.com/questions/15880133/jensen-shannon-divergence>