**C2C12 Computational Methods README**

Rochelle Lai

Objective: Conduct computational analysis for twelve C2C12 RNA-sequencing samples. The samples are labeled as follows:

* Controls: CTL\_1…CTL\_6
* MOTSc-treated: MOTSc\_1…MOTSc\_6

Pipeline

1. Map trimmed FASTQ files to *Mus musculus* genome (from Ensembl) using Kallisto.
2. Conduct differential gene expression analysis using DESeq2.
3. Conduct Gene Set Enrichment Analysis (GSEA) using clusterProfiler.
4. Create bubble plot from GSEA results.

**Kallisto**

This folder contains two subfolders, “Kallisto\_Mapping” and “Parse\_Files.”

*> Kallisto\_Mapping*

C2C12\_kallisto.sh

This script maps C2C12 FASTQ files (not in folder) to the *Mus musculus* genome (Mus\_musculus.GRCm38.p6.Ens94.cdna\_2018-12-19.fa) using Terminal (Mac).

* Kallisto 0.43.0 for CTL\_1 & 3
* Kallisto 0.43.1 for CTL\_2-6, MOTSc\_1-6

*> Parse\_Files*

C2C12\_parse\_kallisto\_mappings.pl

This script parses the results from C2C12\_kallisto.sh (abundance.tsv files) into one counts file (2019-04-25\_C2C12\_RNAseq\_kallisto\_mapping.txt) using Terminal (Mac).

Kallisto Citation

Bray, N. L., Pimentel, H., Melsted, P. & Pachter, L. Near-optimal probabilistic RNA-seq quantification, Nature Biotechnology 34, 525-527(2016), doi:10.1038/nbt.3519

* PMID: [27043002](https://www.ncbi.nlm.nih.gov/pubmed/27043002)

**DESeq2**

This folder contains the input and output for C2C12\_DESeq2.R

C2C12\_DESeq2.R

This script takes in the C2C12 kallisto-mapped gene counts file (2019-04-25\_C2C12\_RNAseq\_kallisto\_mapping.txt) and processes it through DESeq2 modeling to find differentially expressed genes between control and MOTSc-treated samples (FDR 5%). Using DESeq2 results, the script will also create a heat map.

* R Version 3.4.1 (2017-06-30)
* DESeq2 1.16.1
* pheatmap 1.0.10

DESeq2 Citation

Love MI, Huber W, Anders S (2014). “Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2.” *Genome Biology*, 15, 550. doi: 10.1186/s13059-014-0550-8

* PMID: [25516281](https://www.ncbi.nlm.nih.gov/pubmed/25516281)

**PCA**

This folder contains the input and output for C2C12\_PCA.R

C2C12\_PCA.R

This script generates a PCA plot of PC1 and PC2 of gene expression (transcriptomics) for C2C12 normalized counts (2019-04-25\_C2C12\_RNAseq\_log2\_counts\_matrix.txt).

* R Version 3.4.1 (2017-06-30)
* pvclust 2.0.0

**ClusterProfiler\_GSEA**

This folder contains two subfolders, “C2C12” and “Muscle.”

*> C2C12*

C2C12\_ClusterProfiler\_GSEA.R

This script perform GSEA at FDR 15% using the clusterProfiler package for C2C12 DESeq2 all gene statistics (2019-04-25\_C2C12\_RNAseq\_all\_genes\_statistics.txt).

* R Version 3.5.0 (2018-04-23)
* clusterProfiler 3.10.1
* org.Mm.eg.db 3.7.0

*> Muscle*

Muscle\_ClusterProfiler\_GSEA.R

This script perform GSEA at FDR 15% using the clusterProfiler package for Muscle DESeq2 all gene statistics (2019-03-27\_Muscle\_MOTSc\_global\_\_all\_genes\_statistics\_sva.txt).

* R Version 3.5.0 (2018-04-23)
* clusterProfiler 3.10.1
* org.Mm.eg.db 3.7.0

clusterProfiler Citation

Yu G, Wang L, Han Y, He Q (2012). “clusterProfiler: an R package for comparing biological themes among gene clusters.” *OMICS: A Journal of Integrative Biology*, 16(5), 284-287. doi: 10.1089/omi.2011.0118.

* PMID: [22455463](https://www.ncbi.nlm.nih.gov/pubmed/22455463)

**GSEA\_Bubble\_Plot**

This folder contains three subfolders, “C2C12”, “C2C12\_and\_Muscle” and “Muscle.”

*> C2C12*

C2C12\_ GSEA\_ClusterProfiler\_Bubble\_Plot\_.R

This script generates bubble plots from C2C12 GSEA statistics (FDR 15%) for KEGG and GO-BP pathways.

* R Version 3.4.1 (2017-06-30)
* pheatmap 1.0.10
* ggplot2 3.1.0
* scales 1.0.0

*> C2C12\_and\_Muscle*

C2C12\_and\_Muscle\_GSEA\_ClusterProfiler\_Bubble\_Plot\_.R

This script generates bubble plots from C2C12 and Muscle GSEA statistics (FDR 15%) for KEGG and GO-BP pathways.

* R Version 3.4.1 (2017-06-30)
* pheatmap 1.0.10
* ggplot2 3.1.0
* scales 1.0.0

*> Muscle*

Muscle\_GSEA\_ClusterProfiler\_Bubble\_Plot\_.R

This script generates bubble plots from Muscle GSEA statistics (FDR 15%) for KEGG and GO-BP pathways.

* R Version 3.4.1 (2017-06-30)
* pheatmap 1.0.10
* ggplot2 3.1.0
* scales 1.0.0