**THP1 Computational Methods README**

Rochelle Lai

Objective: Conduct computational analysis for fifteen THP1 RNA-sequencing samples. The samples are labeled as follows:

* Controls: CTL\_1…CTL\_5
* PMA treated: PMA\_6…PMA\_10
* MOTSc + PMA treated: MOTScPMA\_11… MOTScPMA \_15

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 phenoTest.
4. Create bubble plot from GSEA results.
5. Create scatter plot.

**Kallisto**

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

*> Kallisto\_Mapping*

THP1\_kallisto.sh

This script maps THP1 FASTQ files (not in folder) to the Human genome (Human.GRCh38.p12.Ens96.cdna\_2019-5-13.txt) using Terminal (Mac).

* Kallisto 0.43.1

*> Parse\_Files*

THP1\_parse\_kallisto\_mappings.pl

This script parses the results from THP1\_kallisto.sh (abundance.tsv files) into one counts file (2019-05-13\_THP1\_kallisto\_mappings.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 THP1\_DESeq2.R

THP1\_DESeq2.R

This script takes in the THP1 Kallisto-mapped gene counts file (2019-05-13\_THP1\_kallisto\_mappings.txt) and processes it through DESeq2 modeling to find differentially expressed genes between control, PMA-treated, and MOTSc + PMA-treated samples (FDR 5%). Output tables for the comparison between control vs. PMA-treated, control vs. MOTSc + PMA-treated, and PMA-treated vs. MOTSc + PMA-treated samples will be produced. Using DESeq2 results, the script will also create an annotated volcano plot (for PMA-treated vs. MOTSc + PMA-treated samples), a multidimensional scaling analysis (for all samples), and clustering plot (for all samples).

* R Version 3.4.1 (2017-06-30)
* DESeq2 1.16.1
* pvclust 2.0.0

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)

**GSEA**

This folder contains the input and output for THP1\_GSEA.R

THP1\_GSEA.R

This script conducts gene set enrichment analysis (GSEA) using DESeq2 all gene statistics output files (2019-05-14\_PMA\_vs\_MOTScPMA\_THP1\_RNAseq\_all\_genes\_statistics.txt, 2019-05-14\_CTL\_vs\_PMA\_THP1\_RNAseq\_all\_genes\_statistics.txt, 2019-05-14\_CTL\_vs\_MOTScPMA\_THP1\_RNAseq\_all\_genes\_statistics.txt) and GeneIDs from pathways within the [C5 (GO) gene set collection from Molecular Signature Database](http://software.broadinstitute.org/gsea/msigdb/genesets.jsp?collection=C5) (c5.all.v6.2.symbols.gmt).

* R Version 3.4.1 (2017-06-30)
* phenoTest 1.24.0
* qusage 2.10.0

Molecular Signature Database Citation

Subramanian, Tamayo, et al. (2005), PNAS 102, 15545-15550, [http://www.broad.mit.edu/gsea](http://www.broad.mit.edu/gsea/)

* PMID: [16199517](https://www.ncbi.nlm.nih.gov/pubmed/16199517)

phenoTest GSEA Citation

Planet E (2019). *phenoTest: Tools to test association between gene expression and phenotype in a way that is efficient, structured, fast and scalable. We also provide tools to do GSEA (Gene set enrichment analysis) and copy number variation.*. R package version 1.32.0.

**GSEA\_Bubble\_Plot**

This folder contains the input and output for THP1\_GSEA\_Bubble\_Plot.R

THP1\_GSEA\_Bubble\_Plot.R

This script generates a bubble plot and accompanying txt file from THP1 GSEA statistics (FDR 5%) for GO pathways.

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

**Scatter\_Plot**

This folder contains the input and output for THP1\_Scatter\_Plot.R

THP1\_Scatter\_Plot.R

This script takes in all gene statistics files from DESeq2 (2019-05-14\_CTL\_vs\_PMA\_THP1\_RNAseq\_all\_genes\_statistics.txt, 2019-05-14\_PMA\_vs\_MOTScPMA\_THP1\_RNAseq\_all\_genes\_statistics) and creates a scatter plot that illustrates correlation and highlights the twenty most significant genes that change only with MOTSc + PMA treatment with a standard deviation greater than one.

* R Version 3.4.1 (2017-06-30)
* ggplot2 3.1.0
* bitops 1.0.6
* data.table 1.11.8