EWCE

Jake Herb September 2020

# Part 1: Folders/organization

1. EWCE\_main\_folder
   1. EWCE.Rproj
   2. EWCE
      1. EWCE\_CTD\_examples
         1. 4 folders that relate the to generate of CTDs from 4 different reference datasets.
            1. NOTE: in the case of “Zeisel\_2018” the source data can be accessed publicly by following the link posted in the associated Rmarkdown file and included in the text file called “READ\_ME\_Zeisel\_2018\_source\_data”
      2. EWCE\_analysis\_examples
         1. 3 folders that relate to running the EWCE\_analysis function on ALS data using CTDs from 3 different reference datasets
      3. EWCE\_pipeline
         1. A folder that contains the scripts and Rmarkdown files required to run the 3 steps of the EWCE pipeline. ONE) ctd generation TWO)ewce\_analysis THREE) data\_visualization
      4. RNAseq\_datasets\_to\_analyze
         1. A consolidation of the ALS datasets currently available to be analyzed with EWCE
            1. ALS\_model2\_allALSvControl\_limma\_res.RData <- entire ALS dataset
            2. Model\_3b\_limma\_res.RData <- sporadic ALS only
            3. Model\_4\_limma\_res.RData <- data from a single hospital
   3. READ\_ME\_How\_to\_use\_EWCE

# Part 2: How to use this pipeline

EWCE is a tool to help identify if a target gene list is enriched within a particular cell type. I recommend familiarizing yourself with the example walkthrough provided by the package’s author Nathan Skene, which can be found here -> <https://nathanskene.github.io/EWCE/articles/EWCE.html>

The citation for EWCE method can be found here also be found here

[*Skene, et al. Identification of Vulnerable Cell Types in Major Brain Disorders Using Single Cell Transcriptomes and Expression Weighted Cell Type Enrichment. Front. Neurosci, 2016.*](https://www.frontiersin.org/articles/10.3389/fnins.2016.00016/full)

What I have contributed here, is a couple of functions that wrap around EWCE’s function to make it quick and easy to analyze your favorite bulk-tissue RNA-sequencing dataset using EWCE. To begin, you will need 2 things — a bulk-tissue RNAseq dataset and a single-cell RNAseq reference dataset from either mice or humans

The pipeline has 3 steps

1. Generate a Cell Type Data Object (CTD) from the reference dataset
2. Run EWCE analysis on your bulk tissue RNA dataset (after previously performing the differential gene expression analysis)
3. Visualize the resulting EWCE output data with a heat map

I have included some examples of how the code works in the EWCE main folder under CTD/analysis examples. If you are bringing new data to this pipeline, be aware that there can be a substantial amount of preprocessing necessary to shape your data into the format needed for EWCE

The CTD\_generator function requires a counts matrix with unique CellID’s as column names and unique gene names as row names. The CTD\_generator function also requires a metadata file with unique CellID’s as row names and cell-type annotations for each CellID.

The CTD output of the CTD\_generator function is then fed into the EWCE\_analysis function with the differential gene expression data from a bulk-tissue RNAseq experiment. The bulk-tissue differential gene expression data is used to generate 2 target gene lists (one for the most upregulated genes, and one for the most downregulated genes). EWCE asks whether those 2 gene lists are enriched within the any of the cell-type transcriptomes characterized by the reference CTD.

The EWCE analysis output data can then be fed into the EWCE decasualization function, which shows the enrichment of the target gene lists by each cell type.