# Total Functional Score of Enhancer Elements (TFSEE) for Breast Cancer Subtypes

This directory contains the scripts for identification of key

breast cancer subtype-specific TFs that act at transcribed enhancers to

dictate gene expression patterns determining growth outcomes, using TFSEE.

## Dependencies

This code requires python 2.7+ to run.

The pythons scripts require the following python packages:

- biopython-1.70

- pandas-0.20.1

- numpy-1.12.1

- scikit-learn-0.18.1

- matplotlib-2.0.2

- seaborn-0.8.1

- scipy-0.19.0

Install the dependencies.

```sh

pip install -r requirements.txt

```

#### Data Requirements

This pipeline requires data from the following analysis to be used as input.

1. De novo identification of enhancers using GRO-seq and [groHMM](http://www.bioconductor.org/packages/release/bioc/html/groHMM.html)

2. Normalize Enhancer Expression using GRO-seq: For each cell line, quantify the GRO-seq reads, RPKM, that fall within a 1 kb region around the center of the overlap for paired enhancer transcripts or from the 5′ end of unpaired enhancer transcripts

3. Normalize Enhancer Expression using ChIP-seq: For each cell line, quantify the ChIP-seq reads, RPKM, from H3K4me1, H3K27ac, and input for each enhancer within the universe of GRO-seq-defined enhancers

4. Motif Predictions: De novo motif analyses on a 1 kb region of expressed enhancers for each cell line using [MEME](http://meme-suite.org/) and  matched to known motifs using TOMTOM and  [JASPAR](http://jaspar.genereg.net/)

5. Normalize Transcription Factor Expression using RNA-seq:  For each cell line, quantify the RNA-seq reads, FPKM, for each  transcription factor that is a binding target for the motifs

### Scripts

- Compute TFSEE to identify cognate transcription factors: tfsee\_analysis.py