**Homework 4 (Insulin Resistance and Diabetes)**

**This challenging but highly tractable HW IS DEALING WITH ANALYSIS OF GENE EXPRESSION DATA.**

**Required and Recommended Readings:**

**HUMAN DATA FROM THIS PAPER:**

Increased SRF transcriptional activity in human and mouse skeletal muscle is a signature of insulin resistance [3] <http://www.ncbi.nlm.nih.gov/pubmed/21393865>

RECOMMENDED

Diabetes, cancer, and metformin: connections of metabolism and cell proliferation. [2] <http://www.ncbi.nlm.nih.gov/pubmed/22211893>

The target of metformin in type 2 diabetes [1] <http://www.nejm.org/doi/full/10.1056/NEJMcibr1409796>

**HUMAN DATA ANALYSIS ONLY FOR HOMEWORK**

**(due a week for TUESDAY ).**

We will primarily analyze the human data above to gain skills in clustering and dimension reduction.

**Steps:**

1. Plot a heatmap of ALL the human data.
2. Perform a Dimension reduction, i.e. PCA on all the DATA
3. Cluster the patients using k-means OR h-clustering into two or more clusters.
4. Display the BMI status (or any other label you think is interesting) on the patients
5. Are the clusters you are seeing enriched in high BMI or low BMI
6. Do male and female patients cluster?
7. Perform DAVID analysis on the gene clusters – pls check out KEGG pathways and GO (Gene Ontology).
8. Discuss your results.

FIRST STEPS:

Please read the papers first and take a look at the raw data and observe any challenges. We will post a clean version of the datasets as well. If you want to practice your own raw data analysis and compare to the posted version feel free to start raw analysis early (but expect challenges).

Increased SRF transcriptional activity in human and mouse skeletal muscle is a signature of insulin resistance <http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE25462>

Comments on GSEA/DAVID: Now that we have several lists of differentially expressed genes, we can ask which gene sets are overrepresented in these lists. Using your genes of interest, determine which gene sets they represent via DAVID and GSEA (DAVID ANALYSIS IS REQUIRED). Once again, if you are having trouble contact the instructors.

**Data Exploration: Dimensionality Reduction and Principal Component Analysis AND MORE!**

Here, you will do some conventional or novel exploratory analyses on these data. Using the gene expression profiles for these studies, project the data onto PCA space using the 1st and 2nd principal component. PCA can be done in whatever programming language you are most comfortable with.

**Systems Biology of Human Disease Insights:**

You need to be very transparent about what you are doing for your analysis and state your hypothesis as cleanly as possible.

Make sure you specify how you do PCA OR any other analysis.

**References**

Jin, W., Goldfine, A. B., Boes, T., Henry, R. R., Ciaraldi, T. P., Kim, E.-Y., Emecan, M., Fitzpatrick, C., Sen, A., Shah, A., Mun, E., Vokes, M., Schroeder, J., Tatro, E., Jimenezchillaron, J., and Patti, M.-E. Increased SRF transcriptional activity in human and mouse skeletal muscle is a signature of insulin resistance. J Clin Invest 0, 0 (2011), 0.

Viollet, B., Guigas, B., Sanz Garcia, N., Leclerc, J., Foretz, M., and Andreelli, F. Cellular and molecular mechanisms of metformin: an overview. Clinical Science 122, 6 (2011), 253–270.