# Project Notes: Office Hours

**Automated Enrichment Mapping:**

Given Data Set from RNA-seq (in list form):

* Clusters are found based on organizations that classify them.
* Given gene expression data, statistically significant genes are selected.
* The differentially related genes can be taken and annotated (DAVID)
  + Ontology dataset/biological processes/pathways used to categorize the data.

Enrichment Analysis Process:

1. RNA-seq data is obtained and used to form an expression matrix
2. Clusters are found from scored false discovery rates
3. The significant data is found related to subject of study
4. Data is graphed

Enrichment Algorithm = (clustering using enrichment)

* Algorithm indicates how many genes were annotated
* To achieve in program: Reproduce algorithm or use API

Automating this process (data→representation) is extremely useful

* Make biologist’s life easier

Format of program:

* Develop BioJS or other library where you can plug in genes and get ‘bio processes’ or ‘protein networks; then target them.
* Develop an analysis pipeline (Data→Figure) in the form of a webpage.
  + Example: Drug Mapping
    - <http://bioinformaticslab.ucsd.edu/~stan/>

Approach:

* Set up webspace that calls JQuery and BioJs
  + JQuery – webthings that simplify Javascript
* Implement pipeline:

Tools and Assets:

* Languages:
  + HTML 5
    - CANVAS
  + Java
    - JQuery
* Programs:
  + DAVID

Heat Mapping: <http://openscreen.cz/software/inchlib/home/>

* Requires clusters as input (problem)
* Needs 2D array prepared.
* Common data sets that need to be transformed
* Make it easier for input spreadsheet, easier to use
* ChIP-seq binding protein data
  + →Motif visualization

Differential Phosphorylation:

* Ripping two sets of data to find common pathways
  + Common Questions:
    - Is there a correlation between genetic and phosphorylation data?
      * Tools: Kegg →Hedgehog genome
    - Given this problem how would analyze and correlate this?

**Resources:**

<http://david.abcc.ncifcrf.gov/>

<http://www.baderlab.org/Software/EnrichmentMap/DavidTutorial>

[http://homer.saving technical problems with my project and would like more time to fix them.alk.edu/homer/basicTutorial/clustering.html](http://homer.salk.edu/homer/basicTutorial/clustering.html)

<http://openscreen.cz/software/inchlib/home/>

**CLustering:**

<http://stackoverflow.com/questions/2978286/hierarchical-clustering-with-gene-expression-matrix-in-python>

<http://blog.nextgenetics.net/?e=44>