GCBA815/MCGB815/BMI815
Tools and Algorithms in Bioinformatics
Week-9
Fall 2017

Class Exercise: Gene Set Enrichment Analysis

1. Installing the Software
    a. Go to GSEA web page by typing the web address as
       http://software.broadinstitute.org/gsea/index.jsp
    b. Go to Downloads tab (or type the web address:
       http://software.broadinstitute.org/gsea/downloads.jsp)
    c. There are multiple versions of GSEA. Depending on the Operating System (32 or
       64 bit), download the javaGSEA Desktop Application. For today's class we will be
       downloading "1GB (for 32 or 64-bit Java)" version.
    d. Select the "1GB (for 32 or 64-bit Java)" version from the pull-down menu and
       click "launch". This will download "gsea_XXXXm.jnlp" file.
    e. Click open the downloaded application file. If a message window pops up with
       "Java out of date" error, then first update the Java to the required version (Java
       8).
2. Performing Analysis using GSEA
    a. Data: For this tutorial, download the example p53 dataset available at GSEA web
       site. Go to http://software.broadinstitute.org/gsea/datasets.jsp and download
       three files for p53 dataset.
          i. P53_hgu95av2.gct
         ii. P53_collapsed.gct
        iii. P53.cls
    b. Loading expression data in GSEA
          i. Click on Load data tab from left side menu panel.
         ii. The expression data file can be in gct, res, pcl or txt (tab-delim text
             format. For this tutorial, we are using file in "gct" format.
        iii. Expression data can be loaded in three ways:
                1. Method1: Browse the files and upload the data
                2. Method2: Load the data from GSEA history (or cache). The
                   window "Recently used files" contain all files that have been
                   previously loaded in GSEA.
                3. Method3: Drag and drop files to GSEA
                4. Either drag or drop all three files that you downloaded in step 2.a
                   or browse the location where you saved the files and upload them
                   in GSEA.
    c. Running GSEA
          i. Click on Run GSEA tab from left side menu panel.
         ii. The first section is required fields.
                1. Expression dataset: from the drop down menu, select
                   "P53_hgu95av2"
                2. Gene sets database: Click on the dotted line next to the text box.
                   Select "c2.all.v6.0.symbols.gmt".

3. Number of permutations: 1000 (default)
4. Phenotype labels: "MUT_versus_WT"
5. Chip platform: "HG_U95Av2.chip"
6. Rest of the parameters are default.
 iii. In the Basic fields section, you can provide your analysis name (optional)
 iv. You can also change the folder location to save results (optional).
 v. Execute GSEA by clicking "Run" button in the bottom panel
 vi. You can keep an eye on the status of your run in the left panel.
 vii. Once the analysis is finished, the status will change from "running" to "success".
 viii. Click on the "success" status link. It will open the GSEA analysis report in the browser.
d. GSEA result interpretation
 i. GSEA report has two main sections:
  1. Enrichment in phenotype: "MUT" : provides gene sets enriched in MUT phenotype
  2. Enrichment in phenotype: "WT" : provides gene sets enriched in WT phenotype