Shiny SOP

10/19/2023

Open the Shiny App

1. Locate the shiny app on the Z Drive:
   1. **Z:\Projects\Project Management\Analysis\prafuls\_shinyApp\server.R**
2. The **server.R** file should have an “R” in a blue circle as its icon. If so, simply double-click the file and a program called RStudio should open. If not, right-click **server.R** and select the “Open with…” option. Choose the following application to open the file with the RStudio program:
   1. **C:\Program Files\RStudio\bin\RStudio**
3. Once open in RStudio, click the “Run App” button in the top right corner of the screen



Upload the Data

1. Each shiny analysis requires two excel-like files which are specific to the dataset that you want to analyze: a raw counts file and a design matrix file. For example, here’s the location of the first two flowcells of DRC and TKI data: **Z:\Projects\ProjectManagement\Cytotoxicity\Data\TKI\_UWGS\run1\_2\_combined**. This folder contains two files:
   1. TKI/DRC Raw Counts File (raw\_counts\_run1\_run2\_combined\_genenames.counts)
   2. TKI/DRC Design Matrix File (deseq2\_design.txt)
      1. NOTE: these are “Text” files, but can be opened and edited in Excel. It’s not advised to make changes to the counts file. However, you may want to modify the design matrix (for example, to add an additional variable). This can easily be done in Excel
      2. Also, if you don’t want to edit the design matrix because you’re worried about “ruining” the file, you can always make a copy of it (copy/paste), then work on the copy, thereby saving the original.
2. In the left panel of the shiny app under the “Upload Design Matrix File”, select “Browse…” and go to your design matrix
3. In the left panel of the shiny app under the “Upload Raw Counts File”, select “Browse…” and go to your raw counts
4. NOTE: sometimes you’ll see red text in the left panel indicating an error when uploading these files, re-uploading them or restarting the shiny app will often resolve this

Perform the Analysis

1. View the data by clicking the “View Design Matrix” tab.
2. Use the “Filter Data” checkbox to subset your data. Clicking on this checkbox will cause several filters to appear, each one representing a different variable in your Design Matrix. As you set filters, the changes should be reflected real-time in the design matrix
3. QC Plots
   1. Once you’ve filtered the data, go to the first tab (QC Plots) and click the “Update Plots” action button. This will produce the Heatmaps, PCAs, Normalization graphs, and Dendrograms.
   2. In the left panel, you can modulate a few different parameters
      1. **Minimum Median Read Depth**: some genes are very lowly expressed across all samples. The min median depth calculates the median count for each gene across all samples, and removes genes with a median less than the parameter setting (default set at 10 reads)
      2. **Number of Highly Variable Genes for Analysis**: some gene counts vary significantly across samples, whereas other genes don’t show any expression change at all. The number of highly variable genes ranks the genes from most-to-least variable, then selects the top-most variable genes (default set at 500) to create the PCA and dendrogram plots.
      3. **PCA, Dendrogram, and Heatmap Annotation**: this setting annotates the samples in the PCA, Dendrogram, and Heatmaps (by color and/or shape) based the chosen variable
      4. **PC Rotation for Download**: Principle Components (PCs) attempt to capture large sources of variation in our data. For example, say we have two clusters in our PCA: one cluster located on the left of the graph (ie. a very negative value for PC1) and a second cluster located on the right of the graph (ie. a very positive value for PC1). We can download the PC1 rotation to obtain a ranked list of variables (in our case Genes) that’s contributing to this split in our data along PC1
4. QC Plots with a Custom Gene Set
   1. To re-create the above plots (PCA, dendrogram, heatmap) with a set of custom genes, go to the “Custom Gene Set” tab
   2. All analysis settings remain the same, however you can enter your own gene set (space, tab, comma, or \n delimited should all work) in the “Enter Custom Gene Set” area.
   3. Alternatively, you can check the “Chose from a List of PreDefined Gene Sets” box and a drop-down list of pre-defined gene sets will appear
5. Analysis of cluster robustness with pvclust
   1. The “Run pvclust” tab is currently not working, for more info see <https://github.com/shimo-lab/pvclust>
6. Differential Expression Analysis
   1. Again, use the “Filter Data” checkbox to subset your data. Once you have the desired data, go to the second tab (DE Analysis)
   2. Choose the variable that represents your feature of interest, then choose the specific groups (also called “levels”) that you’re interested in.
      1. Example 1: if interested in the differences between Male and Female, your feature of interest would be Gender, and your levels Male and Female.
      2. Example 2: If interested in the difference between two different TKIs, your feature of interest would be Condition, and your levels would be the two TKIs that you want to compare
   3. Enter any control variables. These are variables that we want to “control” for when we run our analysis. For example, we often want to see the differences between two conditions, but Cell Line differences tend to play a larger role and can obfuscate our analysis. Here, we can “control” for Cell Line, thereby focusing on Condition.
      1. NOTE: we cannot “control” for variables if they are confounded with our feature of interest. For example, if we have 6 male stimulated samples and 6 female unstimulated samples, we cannot run stim vs unstim analysis and control for gender, since all stim samples are male and all unstims are female. Instead, you’ll get an error that the design matrix is not of full rank
      2. <https://hbctraining.github.io/DGE_workshop_salmon_online/lessons/04a_design_formulas.html> for more information on design formulas
   4. Click the “Perform/Update DE Analysis” action button
   5. This will produce a normalization plot, results summary, results table, a list top differentially expressed genes, and an expression plot. The results table is the most important piece of information here and can be downloaded by clicking the “Download Results Table” button in the bottom left (the Results Table is also a “Text” file, but can be opened in excel for easier analysis).
   6. We can import this table directly into IPA.
      1. Open IPA
      2. Go to “New” >> “Core Analysis…” >> “Upload”
      3. Browse to find your results table
      4. Retain Columns 1, 2, 3, and 7
         1. Column 1 (“ID”) will be the gene names
         2. Columns 2, 3, and 7 (“Observation 1”) will be the Expression Counts, Fold Change in Expression, and p-value (FDR) respectively
      5. Click “save” and choose a name for the dataset
      6. Set your IPA parameters then click “Run Analysis”
      7. Qiagen’s online webinars/tutorials may be helpful for IPA: <https://tv.qiagenbioinformatics.com/tag/IPA%20webinar?>