Shiny SOP

01/06/2020

Log-On to our Remote Windows Desktop *Ash*

1. Click on the Windows Start button located in the bottom left corner of your screen
2. In the search bar, enter and select *Remote Desktop Connection*
3. In the new window, you’re username (PHYS\username, for example PHYS\jblamer) and the name of the remote computer (ash.phys.mcw.edu) should appear. If not, enter that information
4. Enter your usual credentials to log-on

Open the Shiny App

1. Once on the remote server, locate the shiny app on the Z Drive (it’s a file called global.R):
   1. **Z:\Projects\Project Management\Analysis\prafuls\_shinyApp\global.R**
2. The global.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 global.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**

Locate 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. Locate these two files on your computer. For example, here’s the exact location (also called the *path*) of the DRC & TKI data: **Z:\Projects\Project Management\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, but when you save your changes, excel will warn you *“Some features in your workbook will be lost if you save it as a Text file, do you want to keep using this format?”*. Just click “yes”.
      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. Once you locate the data, return to the global.R file
3. Enter the exact location (the *path*) of the raw counts file and the design matrix file into the code using quotation marks. For the counts file, the location should be entered on Line 10. For the design matrix file, this should be done on Line 13.
   1. NOTE: you might need to include the hidden extension when entering the file location. For example, the raw counts file in the folder is listed as “raw\_counts”, while its full name is actually “raw\_counts.counts”. This is also true for the design matrix file (deseq2\_design vs. deseq2\_design.txt)
   2. NOTE: file *paths* will have **back-slashes** to denote the different folders located within one another (C:\Users\jblamer\Documents), however, in the global.R file, these back-slashes must be changed to **forward-slashes** (C:/Users/jblamer/Documents)
4. Finally, there’s a green play button near the middle-top of the RStudio window labeled “Run App”. Click this and the analysis will launch.

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 a number of filters to appear, each one representing a different variable in your Design Matrix.
   1. NOTE: the filters are exclusive, meaning that if you want to remove female samples, you would choose Female. In other words, you’re choosing the samples to remove from the analysis, not the ones to include
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. 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.
   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.
   6. For pathway analysis, we can download the Results Table 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). 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?>