User's Guide

Using the R-Peridot Graphical User Interface (GUI) in Windows and GNU/Linux Systems

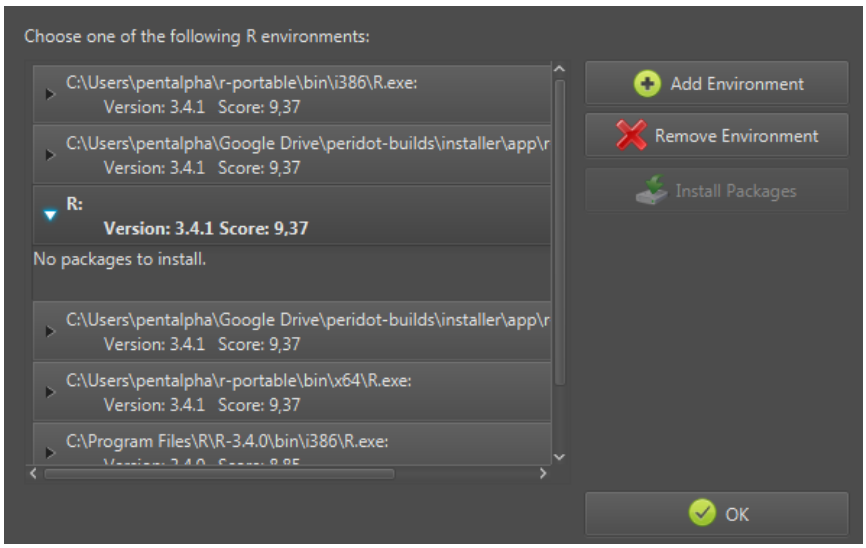
Index

R-Peridot tries to deliver an intuitive user experience, so you don't have to read long instructions to use the software. But in case you need such instructions, read this manual.

To open R-Peridot you can use a desktop/menu shortcut (if you used the Windows installer) or directly open the "r-peridot-gui.jar" file. The first interface you will be presented to is the R Environment Manager.

1. The R Environment Manager

The default modules require several R packages from Bioconductor and CRAN repositories, the R Environment Manager takes care of this installation process. It will open at the first time R-Peridot is being used and you can always open it using the “Tools > R Environments” menu option.



R Environment Manager: Through this interface you can select/add/remove the desired R environment or install missing required packages.

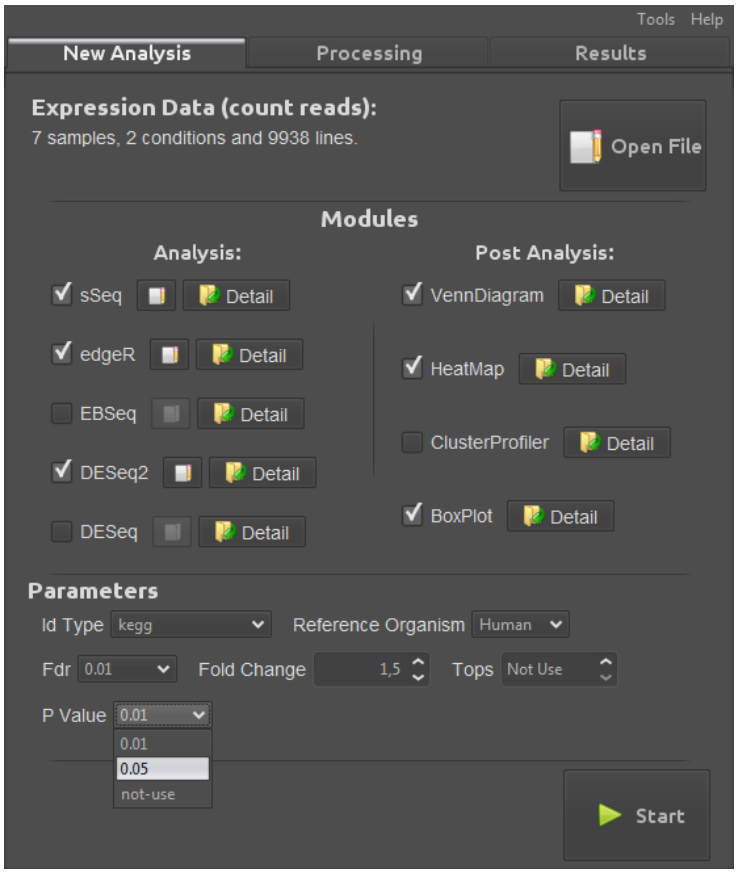
The R Environment Manager scans the system for R environments, but it may not find your R installation if it is in an unusual directory. If no environments are listed, you can click the “**Add Environment**” button and a file chooser dialog will appear, use it to select the R executable you want to use. If you don’t want an environment in the list, you can use the “**Remove environment**” button.

After selecting the desired environment, R-Peridot will show which packages are missing. If there are missing packages, click on the “**Install Packages**” button. It will try to install the packages from R-Peridot’s package repository, which contains the more compatible version of each package. If it cannot install from our repository, the installer will try to use CRAN or Bioconductor directly. In case errors occur, try to install the packages directly within the desired R environment (the Installation Guide has detailed information on how to do this).

Each listed environment has a score to help you choose the better option. The score ranges from 0.0 to 10.0 and is calculated based on the R version (the recommended version is R 3.4.1) and the currently installed packages. A score above 8.0 is just fine. You can also try to run R-Peridot with environments with low scores, but in this case we cannot assure you the modules will function properly. After selecting the environment you want, click on the “**OK**” button.

2. “New Analysis” Tab

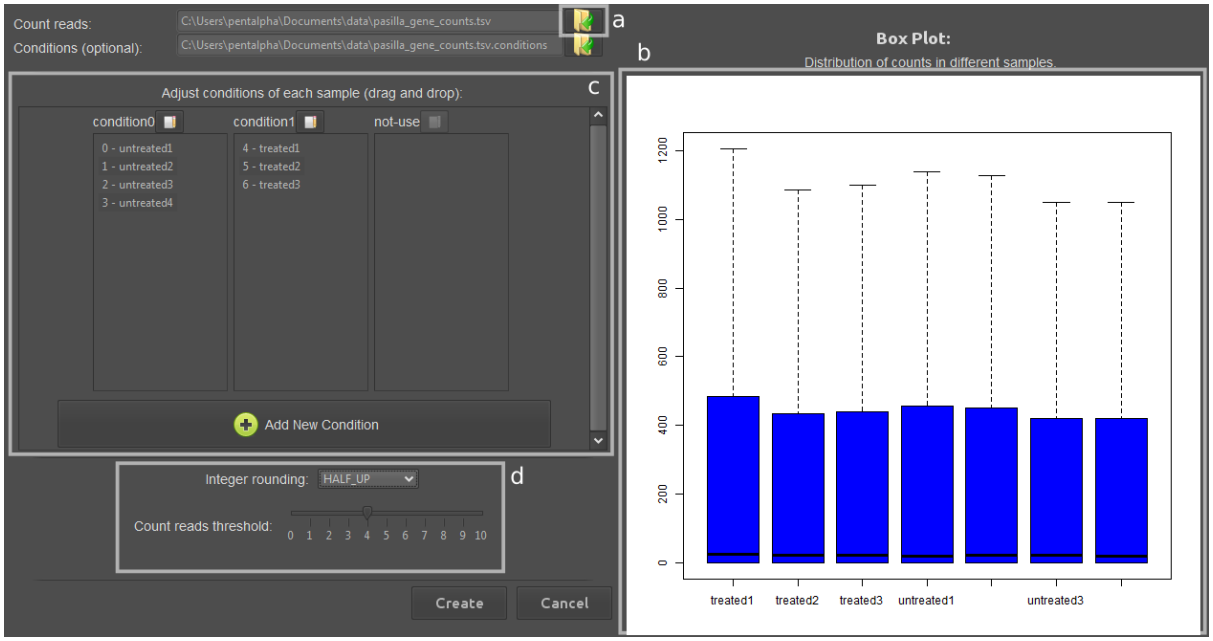
Here you can define the input data, the desired modules and the parameter values. The interface begins only with the option to define the input data. When it is defined, the list of checkboxes to choose the modules is revealed. Once modules are chosen, the parameters are revealed.



New Analysis Tab: When the process of defining the analysis is being completed, the interface looks like this.

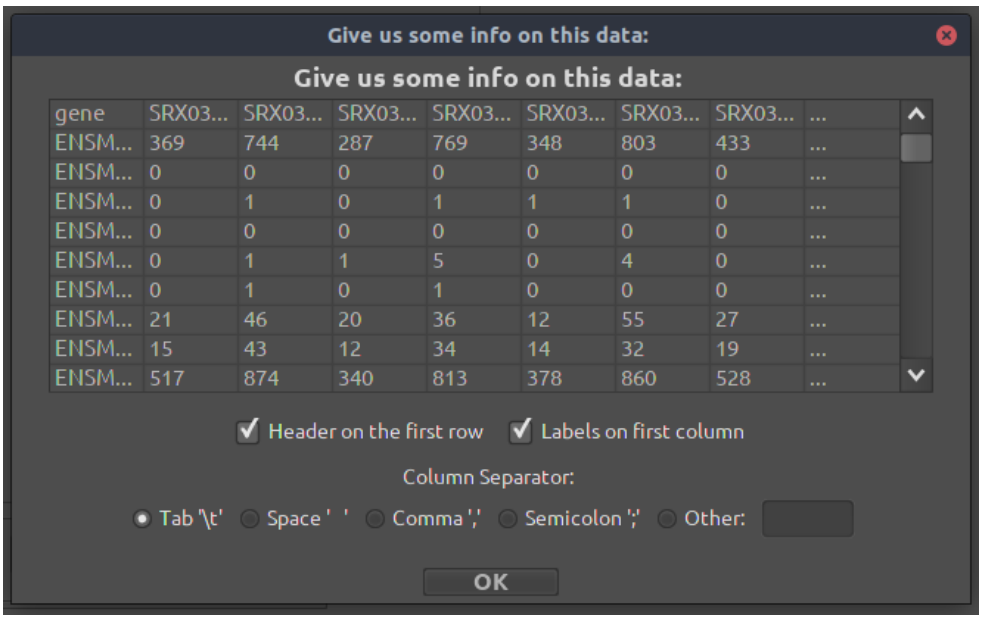
2.1. The “Define Expression Data” Dialog

First, you need to click on “Open File”. This will bring up the “Define Expression Data” dialog:



Define Expression Dialog: The full interface including sample groups and box plot. The a, b, c and d highlighted regions are explained below.

a. First, click in the first button in the top of the dialog, to open the count reads file (for more information about it, read Chapter 9), a spreadsheet text file. After you have chosen a text file, R-Peridot show the following dialog to get some information about the file:



There is a short version of the original file and some questions are asked: There are headers on the first row? Is the first column made of labels? What is the column separator character?

Once all the information is correct and the visualization of the data makes sense, click on the “OK” button.

b. Once a count reads file has been chosen, R-Peridot will display a box plot of the counts in each sample (column) of the file. This is useful to determine what samples you would like to use in the analysis.

c. The next step is to organize the samples in groups (conditions). Create new conditions using the "Add New Condition" button. Move samples between conditions using drag and drop. If there are many samples, try selecting many conditions holding "SHIFT" or "CTRL".

The samples inside the "not-use" condition will be ignored.

To erase a condition, just leave it empty.

d. The differential expression packages expect integer counts by default, so R-Peridot rounds the non-integer values and you can choose the rounding method.

The RNA-Seq technologies are very precise, but the data may have some noise. Some packages may take these values in consideration while looking differentially expressed genes. That's why we created a threshold selector for the reads: read values below the threshold are discarded. To improve performance, lines only with 0's and discarded values are discarded. The default threshold value is 1, that means by default no threshold is being used.

Once you have organized the samples in their respective conditions and everything is ready, click on the "Create" button. That will bring you back to the "New Analysis" tab.

2.2. Choosing the Modules

Now that you have defined the input, the desired modules can be selected. Please note that depending on your input, some analysis modules may be disabled. Most of them don't support more than two conditions or conditions with only one sample (no replicates).

To create a better consensus of the results, with less false positives, we recommend using at least 3 analysis modules.

To select post analysis modules, at least one analysis module have to be selected.

The "*VennDiagram"* module is permanently selected, because it's the responsible for creating the consensus of all results from analysis modules.

Use the "Detail…" button to read about the inputs, results, parameters and the description of each module. You can also read about them in the Chapter 7: Default Modules.

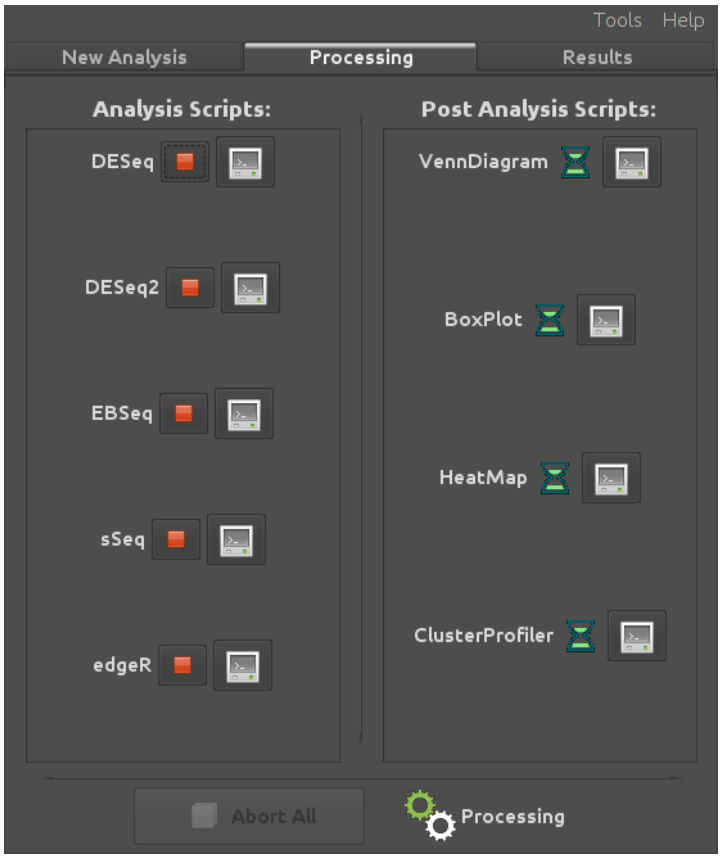
Between the name of the module and the "Detail…" button, there is a small button with a notebook and pencil icon. Use it to specify, for a specific module, parameters different from the ones chosen in the next step.

2.3. Setting Parameters

The parameters are values passed from R-Peridot to the modules. All of them have default values and you don't have to change them. But, if you plan to use the *ClusterProfiler* module, don't forget to choose a Id Type and the correct Reference Organism (only Human, Mouse and Fly are currently supported). Read more about each parameter in the Chapter 8: Default Parameters.

If you have selected the modules you want and the parameters are adequate to your analysis, click on the “Start” button to run the analysis. This will bring you to the “Processing” tab.

3. “Processing” Tab



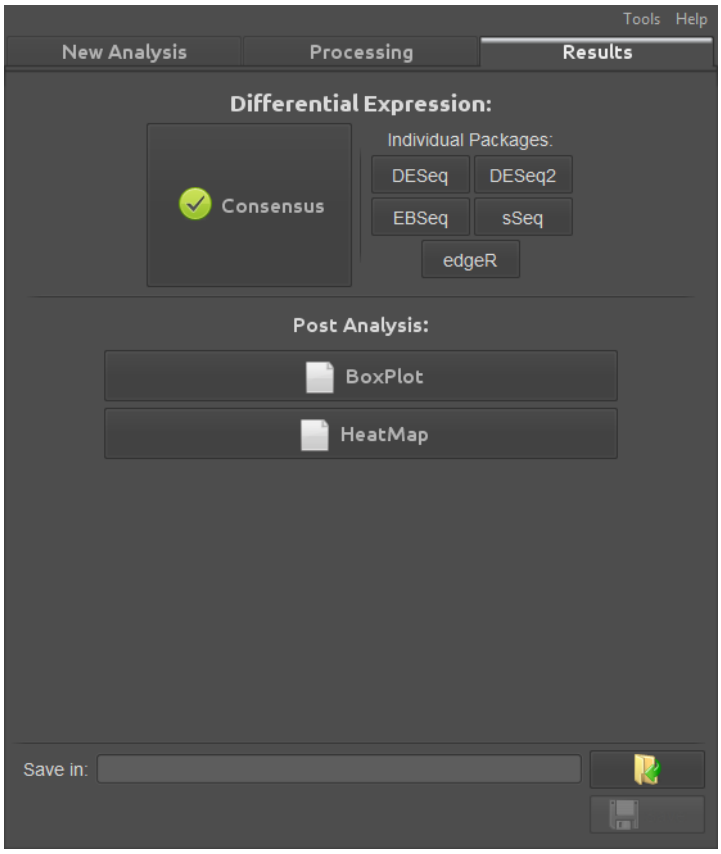
Processing Tab: This interface was made so that the user could watch the progress of each module.

Here, you will have all the selected modules listed. Remember that the modules have dependencies between each other, so some of them will wait (hourglass icon) for others to finish before they can start. The post analysis modules, for instance, always wait for the analysis modules to finish. Once a module starts, the hourglass icon is replaced by a red stop button. If you click that button, the module will stop (and probably fail too). You can also click on the console icon to read the command line output of each module.

The success of a module is determined by the results it produces. If all the mandatory results are created, R-Peridot will consider it as successful. If the opposite happens, the module will have failed. If a module fails, his results won’t appear in the “Results” tab and all the modules that depend on it will fail too.

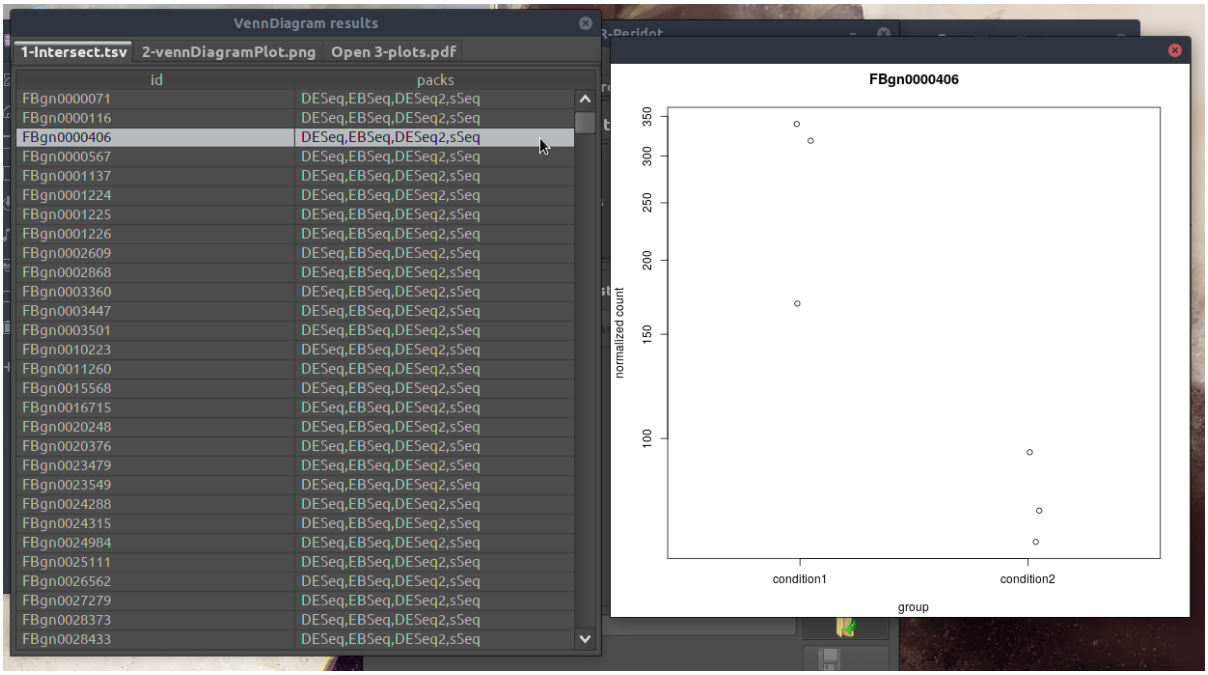
Once all modules have finished being executed, the interface will switch to the “Results” tab.

4. “Results” Tab



Results Tab: As the name suggests, here you can see the results of all successful modules.

At the beginning of this tab, you can access the differential expression analysis results. There is the “Consensus” (resulting from the “VennDiagram” module) and the individual results of each analysis module. After clicking in one of these items, a dialog with a tab for each result will open, including graphics and tables with differentially expressed genes.



The counts plot of a result: If you open one of the tables with a differentially expressed gene per line and click twice on one of these lines, a plot with the counts of this gene across different conditions will appear.

Following the differential expression results are the buttons for the results of each post analysis module, click on one of them to open the results.

These results will only remain stored until you make a new analysis. To save the results, use the “Save in” option at the bottom of tab by choosing a directory to save the results.

5. Tools Menu

The “Tools” menu is the first in the top right corner of the GUI. It includes some tools to manage the modules, R environments and results.

5.1. Modules

#TODO

5.2. R Environments

With this option you can open the “R Environment Manager”, previously discussed in Chapter 1. Please note that any modifications done will restart the “New Analysis” tab.

5.3. Reset User Scripts

This option will reset the modules to their initial state. Default modules previously removed will be reinstalled, any modifications to modules will be lost and new modules will be erased. This is particularly useful in case you did something wrong while playing with the scripts of the modules.

5.4. Refresh Results

This option forces the GUI to reload the results. This is useful in case you made some modification to the files of the results.

6. Help Menu

#TODO

7. Default Modules

#TODO

7.1. Analysis Modules

7.2. Post Analysis Modules

8. Default Parameters

#TODO

9. Input: The Count Reads Table

#TODO

10. Conclusion

#TODO

Pitágoras Alves <alves.pitagoras@gmail.com>
March 28 of 2018
Natal-RN, Brazil