RBiomirGS: MicroRNA gene set analysis [User Manual]

Hanane Hadj-Moussa and Jing Zhang, 2017.

*Note: This techbull will refer you various websites and to the RBioFS webpage that is hosted on our lab website (http://kenstoreylab.com/?page_id=2865).

I. Installing R and RStudio
To run RBiomirGS you must first install R and RStudio on your computer. RStudio is an IDE (interactive development environment) that allows you to interact with R in a more user-friendly way. RStudio will be the main interface for R – even though you could launch R console alone, we almost never do so.

Links to both of these programs are on the lab website (www.kensotreylab.com) → Research → Research Tools → RBiomirGS (http://kenstoreylab.com/?page_id=2865). Both R and RStudio are open-source software, meaning they are completely free. Additionally, both programs are cross platform, ie they will run on Windows, Mac and Linux computers.

A. Installing "R"
1) Install '**R**' by visiting (https://www.r-project.org) and then selecting your preferred CRAN mirror to download the installation file. We will use the University of Toronto's (http://cran.utstat.utoronto.ca/).
2) Download the version of R that corresponds with your operating system. For example, if you are working on a Windows computer click **Download R for Windows** → **install R for the first time** → **Download the newest version of R**
3) Save the installation .exe file → open the file → follow the R Setup Wizard's simple installation instructions.

B. Installing RStudio
1) To install **RStudio** visit (https://www.rstudio.com/). Click **RStudio** → **Desktop** → **Download RStudio Desktop**
2) Download the RStudio version that corresponds with your operating system. For example, if you are working on a Windows computer click **RStudio (latest version)**
3) Save the installation .exe file → open the file → follow the RStudio Setup Wizard's simple installation instructions.

II. Setting up RBiomirGS

Open the RStudio application, this will be the platform you will use to run RBiomirGS. The picture below is a breakdown of the different RStudio quadrants and panels. ****EVERYTHING IN RSTUDIO IS CASE-SENSITIVE****
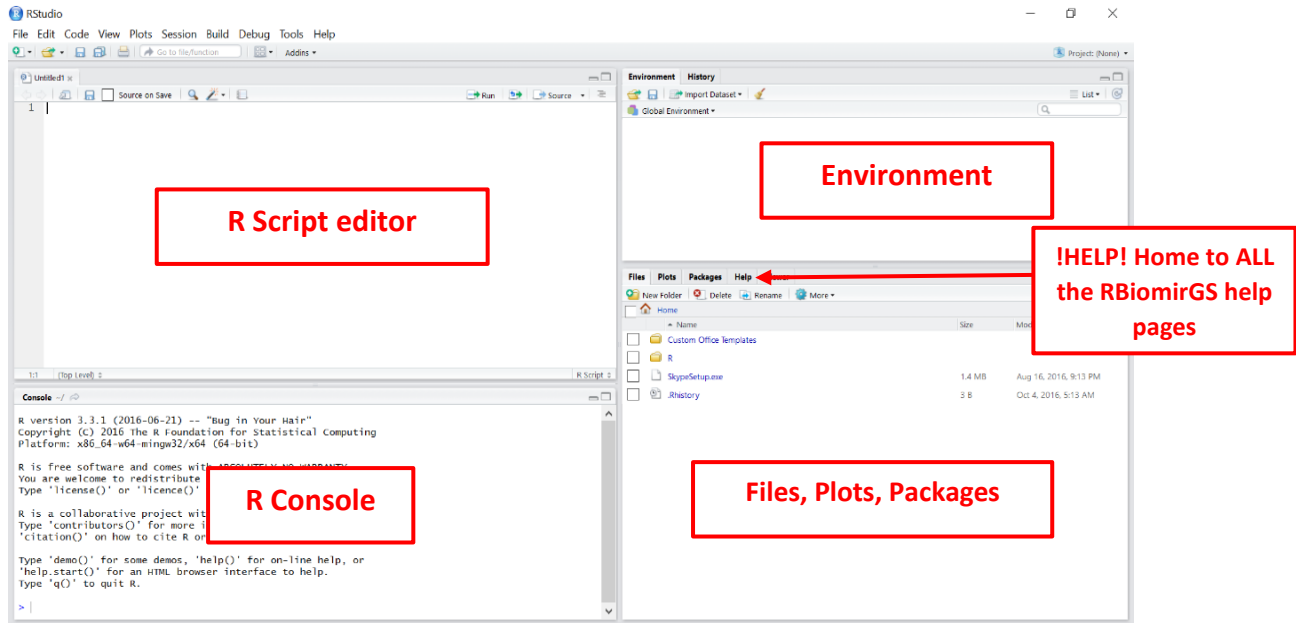
1) To setup the R Script text editor go to **File → New File → R Script (make sure to select R Script, otherwise you will not have the option to run your commands)**. The R Script editor is the main window in which you will interact with R. You should prepare all your R commands before running your commands in the console. Also, the R script editor allows you to run multiple commands at the same time (as long as you highlight all the commands you want to run).

2) To use the latest version of RBiomirGS we need to first install the package **devtools.** This package is the main interface by which RBiomirGS can be installed. Both devtools and RBiomirGS are a one-time installation, unless there are recommended updates that are released. To install 'devtools' type the command below in the R Script editor.

```
install.packages("devtools")
```

*Note: **You must be connected to the internet**. You should type all the commands written in this techbull and not just copy them into the text editor. Microsoft word has altered the formatting and syntax of the code written in this techbull. So if you copy/paste this into RStudio, your command will not run, because R Script quotation marks are uniquely annotated. If you copy/paste them then make sure you rewrite your quotation marks.

3) To run your R commands, you have two options:
   a. **R Script editor**, (preferred): On Windows PC, highlight the command you want to run and press Ctrl+Enter. Or place your text cursor on the command you want to run and press Ctrl+Enter. If you separate your command into different lines then make sure all the lines are highlighted before pressing Ctrl+Enter, otherwise your full command will not run.
   *Note: on Mac, replace Ctrl key with Command key.

b. **Console**: Copy your command from your text editor to the Console and press Enter.

4) Once your command has been successfully installed you should receive a message that reads something like this:
```
The downloaded binary packages are in
C:\Users\Hanane\AppData\Local\Temp\Rtmp4gUWDW\downloaded_packages
```

5) Run the following command to install **Bioconductor** (needed to run RBiomirGS), otherwise skip this step:
```
source("https://bioconductor.org/biocLite.R")
biocLite()
```

For help with installing Bioconductor, visit (https://www.bioconductor.org/install/)

6) Now you are ready to install the RBiomirGS package. Run the following command:
```
devtools::install_github("jzhangc/git_RBiomirGS/RBiomirGS", repos =
BiocInstaller::biocinstallRepos())
```

7) You may run into a few errors as the current version of **devtools** does not contain all the dependencies that RBioFS needs, which means you need to manually install the missing packages.

Here are a few examples of errors and solutions, in this first one I was missing the **scales** package so I installed in manually:

```
Error in loadNamespace(i, c(lib.loc, .libPaths()), versionCheck = vI[[i]])
: there is no package called 'scales'
   install.packages("scales")
```

After I installed the missing package (**scales)** I re-ran the RBiomirGS installation command in step 6.

```
Error in loadNamespace(j <- i[[1L]], c(lib.loc, .libPaths()), versionCheck
= vI[[j]]) : there is no package called 'sandwich'
```

Now, I am missing the package (**sandwich)**, so I installed it with the command below and then re-ran the RBiomirGS installation command from step 6. Repeat this process until you have installed all the missing packages.
```
   install.packages("sandwich")
```

Once RBiomirGS has successfully been installed you will get this message:
```
 * DONE (RBiomirGS)
```

8) The next step is to setup your working directory, this is where you will place all your input data files and where all your plots, stats, and graphs will be exported to. In this example my working directory file is on my desktop and it's called "RBiomirGS stuffs". To setup your working directory run the following command and replace "C:\Users\Hanane\Desktop" with your folder address:
For Windows: `setwd("C:\\Users\\Hanane\\Desktop\\RBiomirGS stuffs")`
For Mac and Linux: `setwd("C:\Users\Hanane\Desktop\RBiomirGS stuffs")`
*Note: You will be required to reset your working domain once you exit and re-enter RStudio.

9) In the 'Files, Plots, and Packages' quadrant of the graph you should select 'Packages' and then find RBiomirGS in the list. Once you click on RBiomirGS you will be redirected to a page with the RBiomirGS

documentation and all the help files for the different functions. If you get stuck using one of the commands use these help files to trouble shoot.

10) If your RBiomirGS stops working or there is an error that you have spent the last week troubleshooting, then you should email the developer jzhangcad@gmail.com and ask him nicely for help.

RBiomirGS QUICK START GUIDE:

INPUT *file layout-*

Format: csv

Name: liver.csv

	A	B	C
1	miRNA	FC	pvalue
2	mmu-miR-let-7f-5p	0.58	0.034
3	mmu-miR-1a-5p	0.32	0.037
4	mmu-miR-1b-5p	0.5	0.001
5	mmu-miR-7a-5p	0.74	0.142
6	mmu-miR-9-3-5p	0.59	0.07
7	mmu-miR-10b-5p	0.57	0.006
8	mmu-miR-15a-5p	0.977	0.862
9	mmu-miR-15b-5p	0.906	0.575
10	mmu-miR-16-3p	0.79	0.048
11	mmu-miR-17-5p	0.84	0.339
12	mmu-miR-18a-3p	0.39	0.012
13	mmu-miR-20a-5p	0.4	0.009
14	mmu-miR-21a-3p	0.37	0.028
15	mmu-miR-22-5p	0.77	0.032
16	mmu-miR-23a-5p	0.44	0.029
17	mmu-miR-26a-5p	0.52	0.11
18	mmu-miR-27a-5p	0.45	0.005
19	mmu-miR-28-5p	0.98	0.93
20	mmu-miR-29a-5p	0.43	0.005
21	mmu-miR-30a-5p	0.86	0.433
22	mmu-miR-30c-5p	0.89	0.423
23	mmu-miR-30d-5p	1.25	0.093
24	mmu-miR-30e-5p	0.68	0.081
25	mmu-miR-32-5p	0.6	0.071
26	mmu-miR-33a-5p	0.62	0.196
27	mmu-miR-34a-5p	0.62	0.014
28	mmu-miR-34c-5p	0.5	0.022
29	mmu-miR-96-5p	0.97	0.637

Load the input file into R

```
require('RBiomirGS')
```

```
raw <- read.csv(file = "liver.csv", header = TRUE, stringsAsFactors =
FALSE)
```

Target mRNA mapping

```
rbiomirgs_mrnascan(objTitle = "mmu_liver_predicted", mir = raw$miRNA,
sp = "mmu", addhsaEntrez = TRUE, queryType = "predicted",
parallelComputing = TRUE, clusterType = "PSOCK")
```

Gene set enrichment

```
rbiomirgs_logistic(objTitle = "mirna_mrna_iwls", mirna_DE = raw,
var_mirnaName = "miRNA", var_mirnaFC = "FC", var_mirnaP = "pvalue",
mrnalist = mmu_liver_predicted_mrna_hsa_entrez_list, mrna_Weight =
NULL, gs_file = "kegg.v5.2.entrez.gmt", optim_method = "IWLS", p.adj =
"fdr", parallelComputing = FALSE, clusterType = "PSOCK")
```

Histogram of all gene sets

```
rbiomirgs_histogram(gsadfm = mirna_mrna_iwls_GS, n = "all",
y.rightside = TRUE, yTxtSize = 8, plotWidth = 300, plotHeight = 200,
xLabel = "Gene set", yLabel = "model coefficient")
```

Histogram of top 50 ranked KEGG pathways based on absolute value of the coefficient (with significant adjusted p values < 0.05)

```
rbiomirgs_histogram(gsadfm = mirna_mrna_iwls_GS, gs.name = TRUE, n =
50, y.rightside = FALSE, yTxtSize = 8, plotWidth = 200, plotHeight =
300, xLabel = "Log odss ratio", signif_only = TRUE)
```

Volcano plot of all gene sets

```
rbiomirgs_volcano(gsadfm = mirna_mrna_iwls_GS, topgsLabel = TRUE, n =
15, gsLabelSize = 3, sigColour = "blue", plotWidth = 250, plotHeight =
220, xLabel = "model coefficient")
```

NOTE: Argument 'n' is for the number of pathways to be labelled on the figure