RmiR package vignette

Francesco Favero*

April 15, 2009

Contents

1 Introduction

RmiR is an R package for the analysis of microRNA and gene expression microarrays. The goal of this package is to couple microRNA and gene expression data (coming from the same RNA). We match miRNAs to corresponding gene targets using the criteria applied by different databases. The package uses various databases of microRNA targets from the *RmiR.db* package:

- mirBase

*favero.francesco@gmail.com

- targetScan
- miRanda from microrna.org
- tarBase from Diana Labs
- mirTarget2 from mirDB
- picTar

To use the package a list of miRNA and a list of genes are required, both with the respective expression values. It is more interesting to have two series, one for microRNA expression and another for gene expression, relative to different times or different treatments. In this case *RmiR* is useful to investigate the correlation between miRNA/Target couples.

It is also possible to use the databases to simply retrieve the targets or the miRNAs, given a list of miRNAs or a list of genes respectively.

The package also includes some tools to visualise the results for coupled data and for time series experiments. The *RSVGTipsDevice* package is required which creates plots in SVG format, that one can visualize with a SVG viewer or in a browser like Mozilla Firefox, Safari and others.

For more details about using the databases have a look at *RmiR.db* package vignette.

```
> vignette("RmiR.db")
```

2 Coupling miRNA and gene expression data

An analysis with *RmiR* starts giving to the `read.mir` function the list of miRNAs and the list of genes with the right annotation.

Leaving the default values the function searches in *targetscan* and *pictar* and prints only the couples present in both databases.

```
> genes <- data.frame(genes = c("A_23_P171258", "A_23_P150053",
+     "A_23_P150053", "A_23_P150053", "A_23_P202435", "A_24_P90097",
+     "A_23_P127948"))
> genes$expr <- c(1.21, -1.5, -1.34, -1.45, -2.41, -2.32, -3.03)
> mirna <- data.frame(mirna = c("hsa-miR-148b", "hsa-miR-27b",
+     "hsa-miR-25", "hsa-miR-181a", "hsa-miR-27a", "hsa-miR-7",
+     "hsa-miR-32", "hsa-miR-32", "hsa-miR-7"))
> mirna$expr <- c(1.23, 3.52, -2.42, 5.2, 2.2, -1.42, -1.23, -1.2,
+     -1.37)
```

```
> genes
```

```
         genes  expr
1 A_23_P171258  1.21
2 A_23_P150053 -1.50
3 A_23_P150053 -1.34
4 A_23_P150053 -1.45
5 A_23_P202435 -2.41
6  A_24_P90097 -2.32
7 A_23_P127948 -3.03
```

```
> mirna
```

```
         mirna  expr
1 hsa-miR-148b  1.23
2  hsa-miR-27b  3.52
3   hsa-miR-25 -2.42
4 hsa-miR-181a  5.20
5  hsa-miR-27a  2.20
6    hsa-miR-7 -1.42
7   hsa-miR-32 -1.23
8   hsa-miR-32 -1.20
9    hsa-miR-7 -1.37
```

2.1 Default parameters

The default parameters are set to search the items present in both *TargetScan* and in *PicTar* databases, and the `genes` object is identified by probe name. If it is not specified, the function makes the average of the different probes identifying the same gene, computing also the coefficient of variation (CV). If there is just one result, no average can be done, so the CV will be NA.

```
> library(RmiR)
> read.mir(genes = genes, mirna = mirna, annotation = "hgug4112a.db")

In targetscan database there are 3 genes and 6 microRNA which are in your files.
          -----------------------------------------

In pictar database there are 3 genes and 6 microRNA which are in your files.
          -----------------------------------------
  gene_id mature_miRNA mirExpr      mirCV symbol geneExpr geneCV
1      22 hsa-miR-148b   1.230         NA  ABCB7     1.21     NA
4     133 hsa-miR-181a   5.200         NA    ADM    -3.03     NA
5     133   hsa-miR-25  -2.420         NA    ADM    -3.03     NA
6     133   hsa-miR-32  -1.215 0.01745943    ADM    -3.03     NA
```

2.2 Change the `at.least` parameter

We can select the result present in at least one database:

```
> read.mir(genes = genes, mirna = mirna, annotation = "hgug4112a.db",
+     at.least = 1)

In targetscan database there are 3 genes and 6 microRNA which are in your files.
          -----------------------------------------
```

```
In pictar database there are 3 genes and 6 microRNA which are in your files.
          ----------------------------------------
   gene_id mature_miRNA mirExpr      mirCV symbol geneExpr     geneCV
1       22 hsa-miR-148b   1.230         NA  ABCB7    1.210         NA
2       59  hsa-miR-27a   2.200         NA   ADD3   -1.450 0.05724023
3       59  hsa-miR-27b   3.520         NA   ADD3   -1.450 0.05724023
4      133 hsa-miR-181a   5.200         NA    ADM   -3.030         NA
5      133   hsa-miR-25  -2.420         NA    ADM   -3.030         NA
6      133   hsa-miR-32  -1.215 0.01745943    ADM   -3.030         NA
23     120  hsa-miR-27a   2.200         NA  ACTA2   -2.365 0.02690893
61     120  hsa-miR-27b   3.520         NA  ACTA2   -2.365 0.02690893
```

Searching with `at.least` equal to 1 basically gives the union of the results from the selected databases.

2.3 Different genes identification.

If the result is annotated by another identification than the platform probe ID, we can specify the annotation identifiers with the parameter `id`.

The possible values are *"probes"*, *"genes"*, *"alias"*, *"ensembl"* and *"unigene"*. An example with entrez gene id identifiers:

```
> genes.e <- genes
> genes.e$gene_id <- c(22, 59, 59, 59, 120, 120, 133)
> genes.e <- genes.e[, c("gene_id", "expr")]
> genes.e

  gene_id  expr
1      22  1.21
2      59 -1.50
3      59 -1.34
4      59 -1.45
5     120 -2.41
6     120 -2.32
7     133 -3.03

> read.mir(genes = genes.e, mirna = mirna, annotation = "hgug4112a.db",
+     id = "genes")

In targetscan database there are 3 genes and 6 microRNA which are in your files.
          ----------------------------------------
```

```
In pictar database there are 3 genes and 6 microRNA which are in your files.
          ------------------------------------------
  gene_id mature_miRNA mirExpr      mirCV symbol geneExpr geneCV
1      22 hsa-miR-148b   1.230         NA  ABCB7     1.21     NA
4     133 hsa-miR-181a   5.200         NA    ADM    -3.03     NA
5     133   hsa-miR-25  -2.420         NA    ADM    -3.03     NA
6     133   hsa-miR-32  -1.215 0.01745943    ADM    -3.03     NA
```

Another example mixing official HGNC symbols and others aliases:

```
> genes.a <- genes
> genes.a$alias <- c("ABCB7", "ADD3", "ADDL", "ADD3", "AAT6", "ACTA2",
+     "ADM")
> genes.a <- genes.a[, c("alias", "expr")]
> genes.a

  alias  expr
1 ABCB7  1.21
2  ADD3 -1.50
3  ADDL -1.34
4  ADD3 -1.45
5  AAT6 -2.41
6 ACTA2 -2.32
7   ADM -3.03

> read.mir(genes = genes.a, mirna = mirna, annotation = "hgug4112a.db",
+     id = "alias")

In targetscan database there are 3 genes and 6 microRNA which are in your files.
          ------------------------------------------


In pictar database there are 3 genes and 6 microRNA which are in your files.
          ------------------------------------------
  gene_id mature_miRNA mirExpr      mirCV symbol geneExpr geneCV
1      22 hsa-miR-148b   1.230         NA  ABCB7     1.21     NA
4     133 hsa-miR-181a   5.200         NA    ADM    -3.03     NA
5     133   hsa-miR-25  -2.420         NA    ADM    -3.03     NA
6     133   hsa-miR-32  -1.215 0.01745943    ADM    -3.03     NA
```

2.4 Results by platform probe ID

If we are using the object `genes` annotated by microarray probes, sometimes we do not need the average of the results, for example when we would like to test each probe separately. In this case, it is preferable to have just the result as it is, annotated by probe. This will obviously cause much redundancy:

```
> read.mir(genes = genes, mirna = mirna, annotation = "hgug4112a.db",
+     at.least = 1, id.out = "probes")

In targetscan database there are 3 genes and 6 microRNA which are in your files.
          ----------------------------------------

In pictar database there are 3 genes and 6 microRNA which are in your files.
          ----------------------------------------
```

	gene_id	mature_miRNA	mirExpr	mirCV	probe_id	geneExpr	geneCV
1	22	hsa-miR-148b	1.230	NA	A_23_P171258	1.21	NA
2	59	hsa-miR-27a	2.200	NA	A_23_P150053	-1.50	NA
3	59	hsa-miR-27a	2.200	NA	A_23_P150053	-1.34	NA
4	59	hsa-miR-27a	2.200	NA	A_23_P150053	-1.45	NA
5	59	hsa-miR-27b	3.520	NA	A_23_P150053	-1.50	NA
6	59	hsa-miR-27b	3.520	NA	A_23_P150053	-1.34	NA
7	59	hsa-miR-27b	3.520	NA	A_23_P150053	-1.45	NA
8	133	hsa-miR-181a	5.200	NA	A_23_P127948	-3.03	NA
9	133	hsa-miR-25	-2.420	NA	A_23_P127948	-3.03	NA
10	133	hsa-miR-32	-1.215	0.01745943	A_23_P127948	-3.03	NA
21	120	hsa-miR-27a	2.200	NA	A_23_P202435	-2.41	NA
31	120	hsa-miR-27a	2.200	NA	A_24_P90097	-2.32	NA
101	120	hsa-miR-27b	3.520	NA	A_23_P202435	-2.41	NA
11	120	hsa-miR-27b	3.520	NA	A_24_P90097	-2.32	NA

3 Correlation between series

3.1 Expected results

The control of target genes by microRNAs occurs at *post-trascriptional* level. The miRNA binds to its target gene and inhibits *translation*.

In some cases the mRNA is degraded by the miRNA annealing, but this is not necessary to stop the *translation* process. In the first case it is possible to see a decrease of the gene expression, while in the other cases we shouldn't observe any particular trend or even an increase of the expression value.

With a time series of microarray data it is possible to see if a target of one or more miRNAs gradually changes its expression with the equal or opposite change of the relative miRNA. Looking at the correlation between the trend of the different miRNA/target couples, we can obtain the *correlated* and the *anti-correlated* couples. This does not mean that there is a sure biological relevance, but it could give some hints for further inverstigation.

3.2 The function RmiRtc

To use the function RmiRtc we need two or more objects created with the function read.mir.

```
> data(RmiR)
> res1 <- read.mir(gene = gene1, mirna = mir1, annotation = "hgug4112a.db")

In targetscan database there are 13 genes and 35 microRNA which are in your files.
          ----------------------------------------

In pictar database there are 7 genes and 27 microRNA which are in your files.
          ----------------------------------------

> res2 <- read.mir(gene = gene2, mirna = mir2, annotation = "hgug4112a.db")

In targetscan database there are 12 genes and 23 microRNA which are in your files.
          ----------------------------------------

In pictar database there are 6 genes and 24 microRNA which are in your files.
          ----------------------------------------

> res3 <- read.mir(gene = gene3, mirna = mir3, annotation = "hgug4112a.db")

In targetscan database there are 13 genes and 35 microRNA which are in your files.
          ----------------------------------------

In pictar database there are 7 genes and 27 microRNA which are in your files.
          ----------------------------------------

> res_tc <- RmiRtc(timeline = c("res1", "res2", "res3"), timevalue = c(12,
+     24, 48))
```

We can decide to filter the object by a correlation and/or a gene expression threshold:

```
> res_fil <- readRmiRtc(res_tc, correlation = -0.9, exprLev = 1,
+     annotation = "hgug4112a.db")
> res_fil$reps

  symbol miRNAs gene_id
2    APP      3     351
3  VLDLR      3    7436
1  CENPV      1  201161
```

The function `readRmiRtc` filter the genes by the absolute expression value set in the arguments, and returns a list with the genes ranked by the number of miRNAs satisfying the correlation threshold.

To see in details the expression of miRNA and genes it is possible to plot the various trends, or print the desired results:

```
> cbind(res_fil$couples, res_fil$geneExpr, res_fil$mirExpr)[res_fil$couples$gene_id
+     351 & res_fil$cor <= -0.9, ]
```

```
   gene_id mature_miRNA   12    24    48 12.1 24.1 48.1
22     351  hsa-miR-20a 0.71 -0.95 -1.67 0.32 1.73 2.12
17     351  hsa-miR-20b 0.71 -0.95 -1.67 0.06 1.10 1.61
19     351   hsa-miR-93 0.71 -0.95 -1.67 0.30 1.25 1.19
```

4 Plot and visualization of the results

4.1 SVG output

In order to present the results, the pachage uses a tool to visualize the data.

With the package *RSVGTipsDevice* we can plot the data for a single miRNA and gene coupled experiment or for a series of experiments. The resulting image is in *svg* format with the properties to have dynamic tips and hyperlinks.

4.2 Plot of single experiment

```
> plotRmiRtc(res1[res1$gene_id == 351, ], svgname = "res1_351.svg",
+     svgTips = T)
```

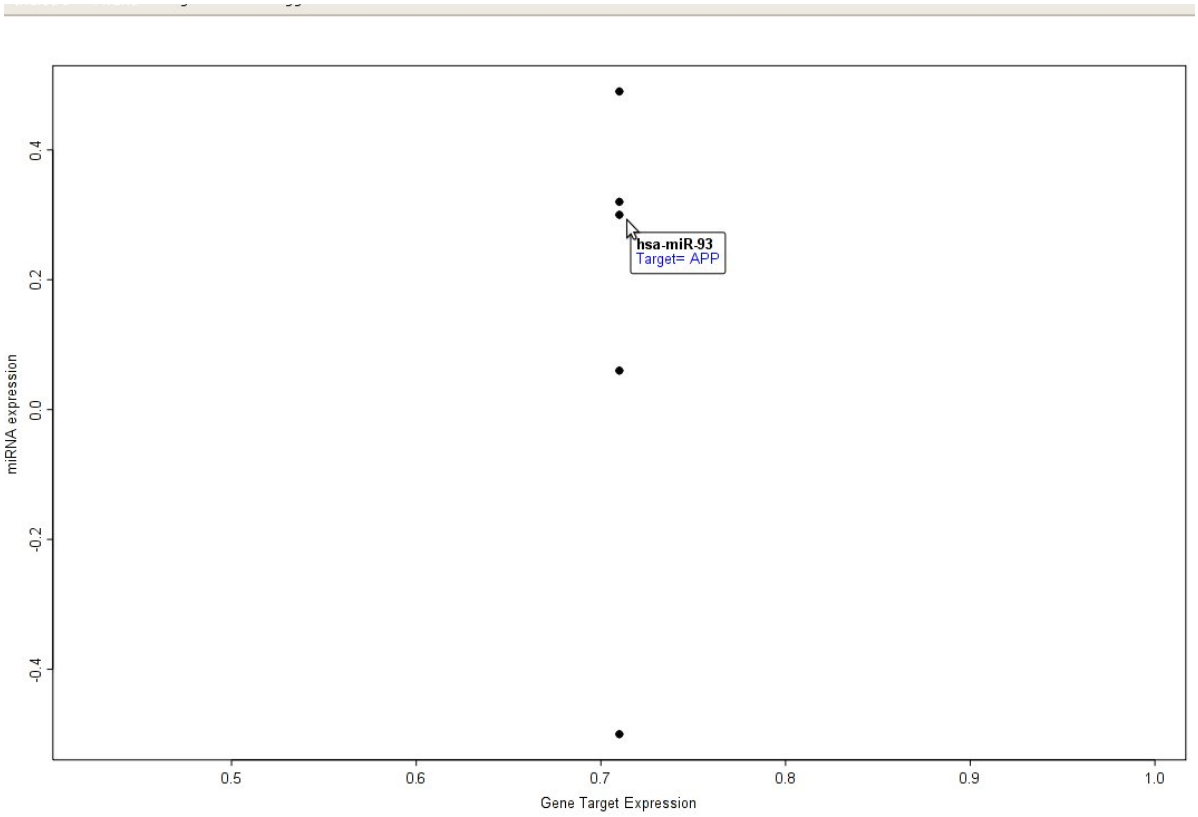
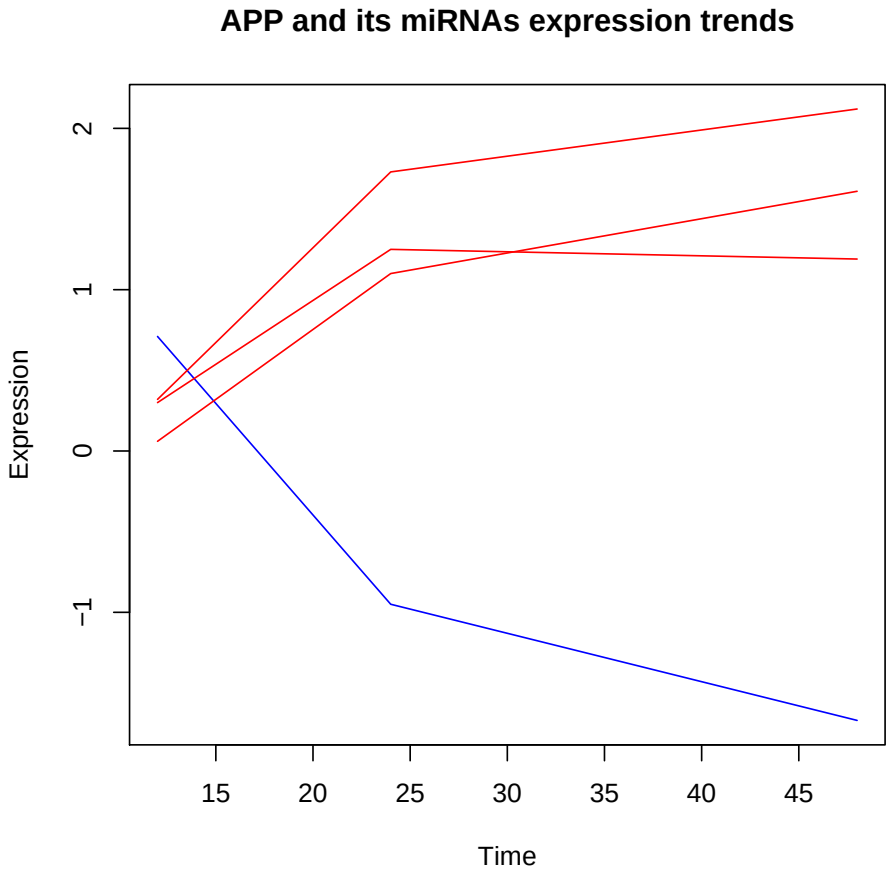


Figure 1: Visualization in a browser with SVG support of the results of the read.mir function selecting a single gene target and its miRNA

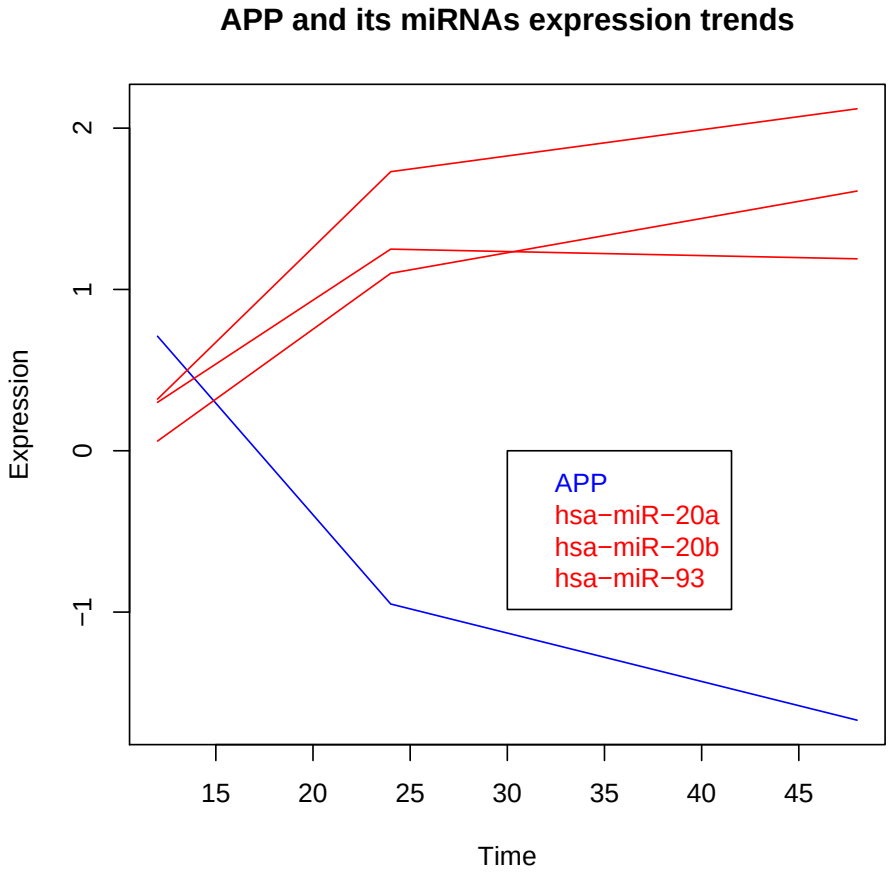
4.3 Plot a time course experiment

```
> plotRmiRtc(res_fil, gene_id = 351)
```



Setting the coordinate of the legend:

```
> plotRmiRtc(res_fil, gene_id = 351, legend.y = 0, legend.x = 30)
```




```
> plotRmiRtc(res_fil, gene_id = 351, legend.y = 0, legend.x = 30,
+     svgTips = T)
```

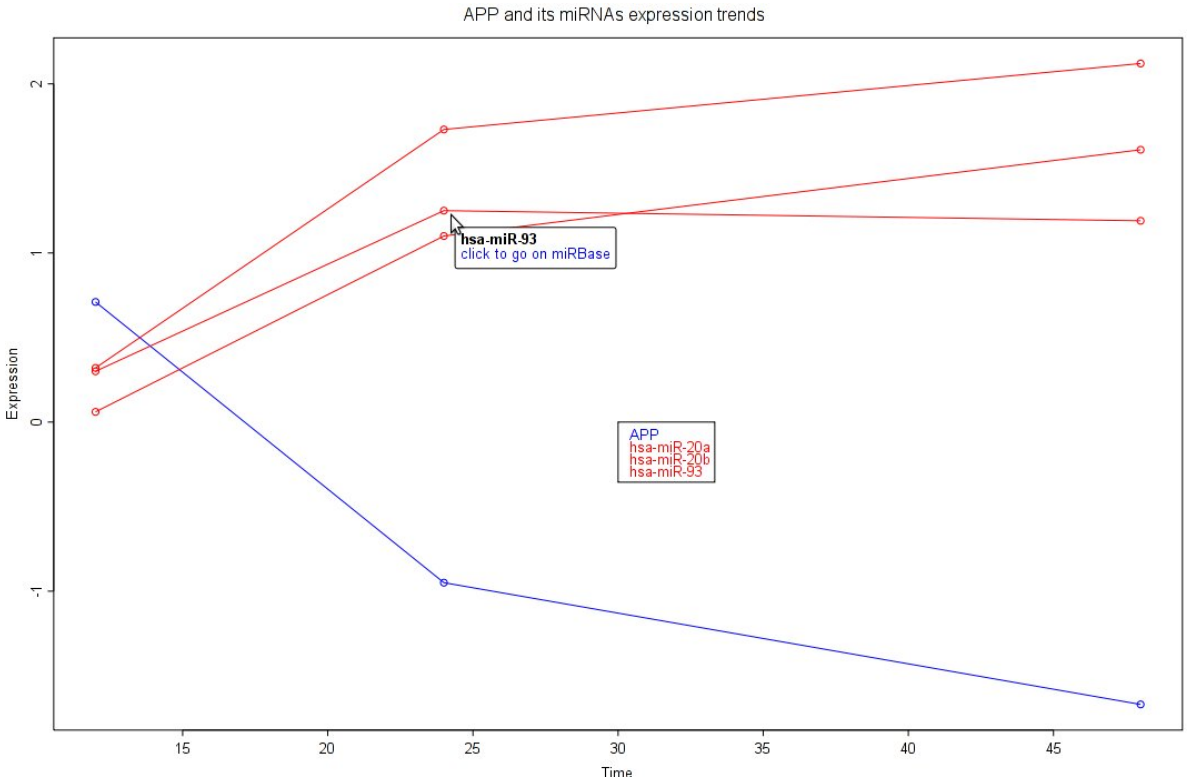


Figure 2: Visualization in a browser with SVG support of a time course experiment of the selected gene with hyperlinks and tips