

RmiR.db package vignette

Francesco Favero*

April 2, 2009

Contents

1	Introduction	1
2	Querying and evaluating a miRNA target database	2
2.1	Find a list of miRNAs or targets	4

1 Introduction

RmiR.db is an R package which includes various databases of miRNA targets:

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

With the package it is possible to evaluate or compare different miRNA target database or also retrieve the targets or the miRNAs, given a list of miRNAs or a list of genes respectively.

*favero.francesco@gmail.com

2 Querying and evaluating a miRNA target database

The miRNA targets databases are included in an SQLite object. We can browse and inspect them directly in an R environment:

```
> library(RmiR.db)
> dbListTables(RmiR_dbconn())

[1] "miranda"      "mirbase"      "mirtarget2"   "pictar"       "tarbase"
[6] "targetscan"
```

We should make a SQL query to have the desired results only:

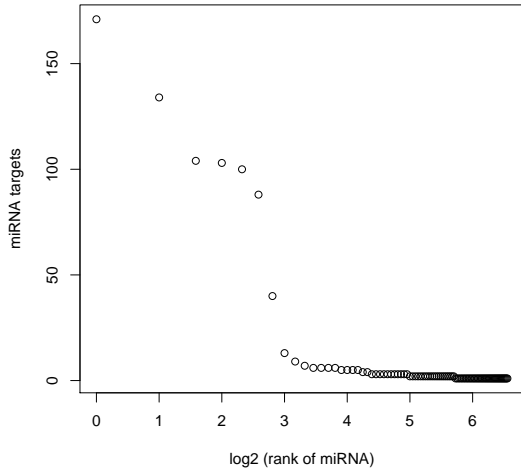
```
> dbGetQuery(RmiR_dbconn(), "SELECT * FROM tarbase WHERE mature_miRNA='hsa-miR-21'")

  mature_miRNA gene_id   pmid
1   hsa-miR-21   7168 17363372
2   hsa-miR-21   7168 17363372
3   hsa-miR-21  27250 18270520
4   hsa-miR-21  27250 17968323
5   hsa-miR-21  27250 18372920
6   hsa-miR-21  27250 17991735
7   hsa-miR-21   5728 17681183
8   hsa-miR-21   5728 16762633
9   hsa-miR-21   5268 18270520
```

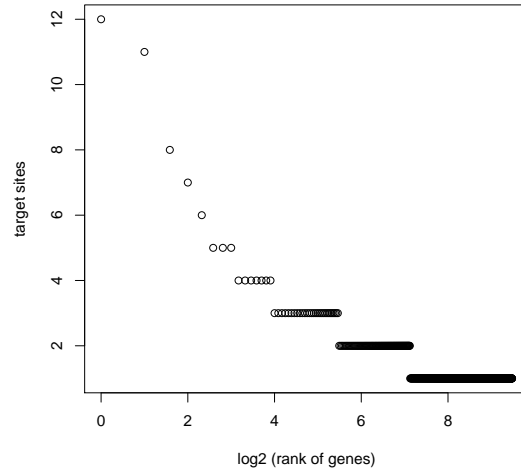
Every query gives a `mature_miRNA` column with the microRNA name and a `gene_id` column with the entrez gene id of the target. There could be also other additional columns useful for further investigation. These columns depend on the database. For example, in TarBase we have the PubMed ID of the article which proves the relation between the miRNA and its target, in TargetScan there are the start and the end point of the miRNA seed in the gene, and so on.

To evaluate the consistency of a database we can visualize two properties of the miRNA/Target relationship; the *multiplicity* and the *cooperativity* :

```
> tarbase <- dbReadTable(RmiR_dbconn(), "tarbase")[, 1:2]
> tarb_mir <- sort(table(tarbase$mature_miRNA), decreasing = T)
> plot(x = log2(c(1:length(tarb_mir))), y = tarb_mir, ylab = "miRNA targets",
+      xlab = "log2 (rank of miRNA)")
> tarb_gene <- sort(table(tarbase$gene_id), decreasing = T)
> plot(x = log2(c(1:length(tarb_gene))), y = tarb_gene, ylab = "target sites",
+      xlab = "log2 (rank of genes)")
```

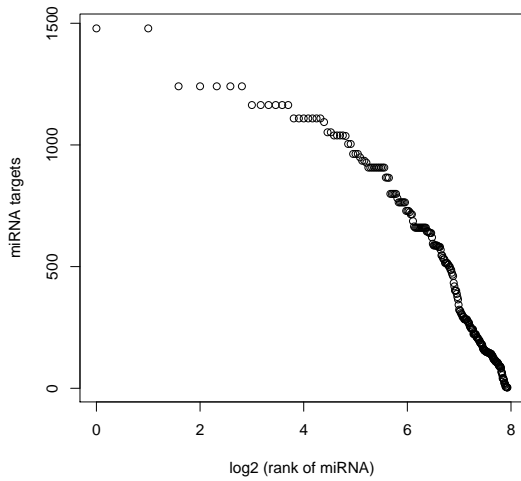


(a) Multiplicity of miRNA in TarBase.

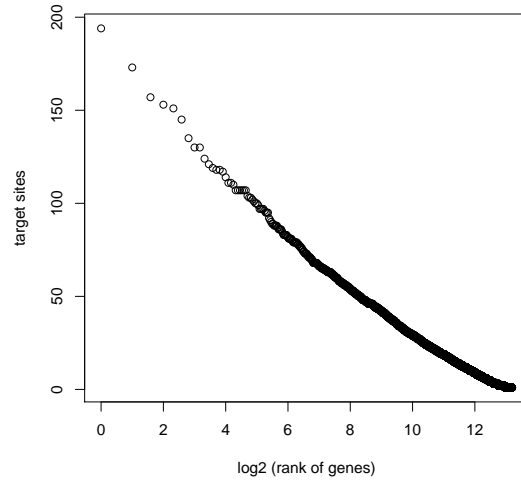


(b) Cooperativity of miRNA in TarBase.

Figure 1: Plot of the multiplicity and cooperativity generated with the TarBase database.



(a) Multiplicity of miRNA in TargetScan.



(b) Cooperativity of miRNA in TargetScan.

Figure 2: Plot of the multiplicity and cooperativity generated with the TargetScan database.

```

> targetscan <- dbReadTable(RmiR_dbconn(), "targetscan")[, 1:2]
> targ_mir <- sort(table(targetscan$mature_miRNA), decreasing = T)
> plot(x = log2(c(1:length(targ_mir))), y = targ_mir, ylab = "miRNA targets",
+      xlab = "log2 (rank of miRNA)")
> targ_gene <- sort(table(targetscan$gene_id), decreasing = T)
> plot(x = log2(c(1:length(targ_gene))), y = targ_gene, ylab = "target sites",
+      xlab = "log2 (rank of genes)")

```

From the graphs we can see that for some miRNAs the number of predicted targets is huge (Fig. 2(a)) compared with the number of experimentally validated targets (Fig. 1(a)).

For a predicted database we note how miRNA have a cooperative control for a lot of gene targets (Fig. 2(b)), when in the TarBase database many gene targets do not have more than four target sites (Fig. 1(b)).

2.1 Find a list of miRNAs or targets

In general the result of an analysis is a list of genes or microRNAs. A nice continuation it is to look for interesting miRNAs or gene targets matching the results.

```

> 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")
> genes <- c("A_23_P171258", "A_23_P150053", "A_23_P150053", "A_23_P150053",
+           "A_23_P202435", "A_24_P90097", "A_23_P127948")

```

We have created a list of miRNA and a list of genes, we use the table of targetscan we created in the previous example, to look for the information we need:

```

> mirs <- targetscan[targetscan$mature_miRNA %in% mirna, ]
> nrow(mirs)

```

```
[1] 5479
```

```
> mirs[1:10, ]
```

	mature_miRNA	gene_id
36870	hsa-miR-148b	57419
36873	hsa-miR-148b	22870
36876	hsa-miR-148b	11176
36879	hsa-miR-148b	8065
36882	hsa-miR-148b	7471
36885	hsa-miR-148b	285527
36888	hsa-miR-148b	79718
36891	hsa-miR-148b	93
36894	hsa-miR-148b	85461
36897	hsa-miR-148b	8556

```
> library(hgug4112a.db)
> targs <- targetscan[targets$gene_id %in% mget(genes, hgug4112aENTREZID),
+ ]
> nrow(targs)
```

```
[1] 34
```

```
> targs[1:10, ]
```

	mature_miRNA	gene_id
24852	hsa-miR-128	59
35204	hsa-miR-143	120
36449	hsa-miR-145	120
36550	hsa-miR-145	120
38093	hsa-miR-152	22
38094	hsa-miR-148b	22
38095	hsa-miR-148a	22
54948	hsa-miR-181d	133
54949	hsa-miR-181c	133
54950	hsa-miR-181b	133