BIOINFORMATICS APPLICATIONS NOTE Vol. 25 no. 6 2009, pages 832–833
doi:10.1093/bioinformatics/btp059

Systems biology

CORNA: testing gene lists for regulation by microRNAs

X. Wu* and M. Watson

Bioinformatics Group, Institute for Animal Health, Compton, RG20 7NN, UK

Received on November 19, 2008; revised on January 06, 2009; accepted on January 25, 2009

Advance Access publication January 29, 2009

Associate Editor: Ivo Hofacker

ABSTRACT

Motivation: With the increasing use of post-genomics techniques to examine a wide variety of biological systems in laboratories throughout the world, scientists are often presented with lists of genes that they must make sense of. A consistently challenging problem is that of defining co-regulated genes within those gene lists. In recent years, microRNAs have emerged as a mechanism for regulating several cellular processes. In this article, we report on how gene lists and microRNA targets data may be integrated to test for significant associations between gene lists and microRNAs.

Results: We discuss CORNA, a package written in R and released under the GNU GPL, which allows users to test gene lists for significant microRNA–target associations using one of three separate statistical tests, to link microRNA targets to functional annotation and to visualize quantitative data associated with those data.

Availability: CORNA is available as an R package from http://corna.sf.net

Contact: xikun.wu@bbsrc.ac.uk

1 INTRODUCTION

Experiments involving post-genomics technologies such as microarrays, proteomics and systems biology often present scientists with gene lists that they must attempt to make sense of. Several software packages exist that allow scientists to assign functional annotation to gene lists, and to assign statistical significance to those associations. These include tools for associating genes with biological ontologies (e.g. Falcon and Gentleman, 2007) and with biological pathways (e.g. Salomonis *et al*., 2007).

A particular challenge is that of assessing which genes in a given gene list are co-regulated. miRBase (Griffiths-Jones *et al*., 2006), a database of all known microRNAs, has been created and there have been several published software tools that attempt to predict the targets of microRNAs (Brennecke *et al*., 2005). An excel-based tool (Creighton *et al*., 2008) has been produced for linking microarray data to microRNA targets information.

Here we describe CORNA, a package for R that allows scientists to analyse gene lists in the context of microRNA–target predictions. Methods exist to test for significant microRNA–target relationships in gene lists, and to test for significant associations between microRNAs and pathways and GO terms. The software is flexible and can read data from public databases or from a scientists own data files. CORNA is released as open-source under the GNU GPL.

*To whom correspondence should be addressed.

2 FLOW OF INFORMATION

Central to the flow of information through CORNA is the gene list from which the user may test for significant microRNA–target associations. The user may also start with a microRNA, find genes that are associated with that microRNA and then test that gene list for significant associations with KEGG pathways or GO terms. The user may also plot quantitative data associated with the targets of a particular microRNA.

2.1 Inputs

CORNA exclusively uses R vectors and data frames. CORNA includes functions for reading microRNA–target data directly from miRBase and microRNA.org (Betel *et al*., 2008). There are also helper functions to read gene and GO term data using biomaRt (Durinck *et al*., 2005); microarray data directly from GEO (Barrett *et al*., 2008); and pathway data directly from KEGG (Kanehisa *et al*., 2004).

2.2 Methods

CORNA employs three complementary statistical methods for enrichments analysis of relationships within lists of genes. These are the HyperGeometric test, Fisher's exact test and the χ^2-test.

2.3 Outputs

If the user tests a gene list for significant microRNA associations, then the output is an R data frame with one row per microRNA, the observed and expected frequencies from sample and population, and the range of user-selected *P*-values.

Where the user begins with a particular microRNA, the targets information is used to create a gene list and that gene list is tested for enrichment of pathways and GO terms.

There is also a range of plotting functions for plotting quantitative data associated with microRNA targets.

3 EXAMPLE ANALYSIS

3.1 Using CORNA to test for enrichment of microRNA–target relationships in a gene list

The list in this example, *tsam*, consists of 1000 ensembl transcript ids; 940 of these were chosen at random, then 30 predicted targets for two microRNAs were added. The example assumes that the file 'arch.v5.txt.mus_musculus.zip' has been downloaded from miRBase targets.

© 2009 The Author(s)
This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (http://creativecommons.org/licenses/by-nc/2.0/uk/) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

```
targets <- miRBase2df.fun(
             file="arch.v5.txt.mus_musculus.zip")
             data(CORNA.DATA)
res <- corna.test.fun(
          x=tsam,
          y=unique(targets$tran),
          z=targets,
          p.adjust="BH")
```

The only two microRNAs with a significant adjusted *P*-values are those used to bias the transcript list. The user may work with genes simply by converting the transcript list to microRNA–gene relationships using the *BioMart2df.fun* and *corna.map.fun* functions.

3.2 Using CORNA to test for KEGG pathways associated with a microRNA list

The microRNA used in this example is 'mmu-mir-155', and we use the predicted targets from miRBase to test for enrichment of KEGG pathways.

```
tran2gene <- BioMart2df.fun(
               biomart="ensembl",
               dataset="mmusculus_gene_ensembl",
               col.old=c("ensembl_transcript_id",
                         "ensembl_gene_id"),
               col.new=c("tran", "gene"))
mir2gene <- corna.map.fun(targets, tran2gene,
              "gene",
              "mir")
gvec <- corna.map.fun(mir2gene,
                      "mmu-mir-155",
                      "mir",
                      "gene")
gene2path <- KEGG2df.fun(org="mmu")
gvec <- intersect(gvec, unique(gene2path$gene))
test <- corna.test.fun(
            gvec,
            unique(gene2path$gene),
            gene2path,
            hypergeometric=T,
            fisher=T,
            fisher.alternative="greater",
            min.pop=10,
            sort="fisher")
```

We first convert the microRNA–transcript relationship to a microRNA–gene relationship using the *BioMart2df.fun* and *corna.map.fun* functions. We then find those genes predicted to be targets of mmu-mir-155. The next stage is to use the *KEGG2df.fun* function to obtain links between genes and pathways from KEGG for *Mus musculus*. Finally, we set the sample to be only those genes targeted by mmu-mir-155 that have a pathway link, and perform hypergeometric and Fisher's exact tests for the KEGG pathways involved. The top five pathways can be seen in Table 1.

Table 1. Top five significant pathways from CORNA for mmu-mir-155

ID	Description	Expected	Observation	*P*-value
00190	Oxidative phosphorylation	5	12	0.002
00400	Phenylalanine etc. biosynthesis	0	3	0.003
00500	Starch and sucrose metabolism	2	6	0.012
05020	Parkinson's disease	5	10	0.016
04010	MAPK signaling pathway	9	15	0.044

4 SUMMARY

With increasing use of large-scale post-genomics techniques, scientists are often presented with lists of genes. MicroRNAs have emerged as an important regulator of gene function. In this article, we have shown that CORNA can be used to test for significant associations between genes, microRNAs, pathways and GO terms. CORNA can also be used to plot quantitative data associated with microRNA targets. CORNA is flexible and can read data from public databases or from a user's own files. CORNA has been tested on both Microsoft Windows and Red Hat Linux. CORNA is released under the GNU GPL and is available from http://corna.sf.net.

Funding: BBSRC (core strategic grant of the Institute for Animal Health).

Conflict of Interest: none declared.

REFERENCES

Barrett,T. *et al*. (2008) NCBI GEO: archive for high-throughput functional genomic data. *Nucleic Acids Res.*, **37**, D5–D15.

Betel,D. *et al*. (2008) The microRNA.org resource: targets and expression. *Nucleic Acids Res.*, **36**, D149–D153.

Brennecke,J. *et al*. (2005) Principles of microRNA-target recognition. *PLoS Biol.*, **3**, e85.

Creighton,C.J. *et al*. (2008) A bioinformatics tool for linking gene expression profiling results with public databases of microRNA target predictions. *RNA*, **14**, 2290–2296.

Durinck,S. *et al*. (2005) BioMart and Bioconductor: a powerful link between biological databases and microarray data analysis. *Bioinformatics*, **21**, 3439–3440.

Falcon,S. and Gentleman,R. (2007) Using GOstats to test gene lists for GO term association. *Bioinformatics*, **23**, 257–258.

Griffiths-Jones,S. *et al*. (2006) miRBase: microRNA sequences, targets and gene nomenclature. *Nucleic Acids Res.*, **34**, D140–D144.

Kanehisa,M. *et al*. (2004) The KEGG resource for deciphering the genome. *Nucleic Acids Res.*, **32**, D277–D280.

Salomonis,N. *et al*. (2007) GenMAPP 2: new features and resources for pathway analysis. *BMC Bioinformatics*, **8**, 217.