Mirsynergy: detect synergistic miRNA regulatory modules by overlapping neighbourhood expansion

Yue Li

yueli@cs.toronto.edu

July 15, 2014

1 Introduction

MicroRNAs (miRNAs) are ~22 nucleotide small noncoding RNA that base-pair with mRNA primarily at the $3'$ untranslated region (UTR) to cause mRNA degradation or translational repression [1]. Aberrant miRNA expression is implicated in tumorigenesis [4]. Construction of microRNA regulatory modules (MiRM) will aid deciphering aberrant transcriptional regulatory network in cancer but is computationally challenging. Existing methods are stochastic or require a fixed number of regulatory modules. We propose *Mirsynergy*, a deterministic overlapping clustering algorithm adapted from a recently developed framework. Briefly, Mirsynergy operates in two stages that first forms MiRM based on co-occurring miRNAs and then expand the MiRM by greedily including (excluding) mRNA into (from) the MiRM to maximize the synergy score, which is a function of miRNA-mRNA and gene-gene interactions (manuscript in prep).

2 Demonstration

In the following example, we first simulate 20 mRNA and 20 mRNA and the interactions among them, and then apply `mirsynergy` to the simulated data to produce module assignments. We then visualize the module assignments in Fig.1

```
> library(Mirsynergy)
> load(system.file("extdata/toy_modules.RData", package="Mirsynergy"))
> # run mirsynergy clustering
> V <- mirsynergy(W, H, verbose=FALSE)
> summary_modules(V)

$moduleSummaryInfo
  miRNA mRNA total   synergy     density
1     4    4    12 0.1680051 0.04426190
2     2    2     6 0.1654560 0.09630038
3     6   10    22 0.1870070 0.02471431
```

```
4      8      7     23 0.1821842 0.02318249
5      2      3      7 0.1640842 0.08457176
6      3      4     10 0.1602223 0.04856618
```

```
$miRNA.internal
  modules miRNA
1       2     2
2       1     3
3       1     4
4       1     6
5       1     8
```

```
$mRNA.internal
  modules mRNA
1       1     2
2       1     3
3       2     4
4       1     7
5       1    10
```

Additionally, we can also export the module assignments in a Cytoscape-friendly format as two separate files containing the edges and nodes using the function `tabular_module` (see function manual for details).

3   Real test

In this section, we demonstrate the real utility of *Mirsynergy* in construct miRNA regulatory modules from real breast cancer tumor samples. Specifically, we downloaded the test data in the units of RPKM (read per kilobase of exon per million mapped reads) and RPM (reads per million miRNA mapped) of 13306 mRNA and 710 miRNA for the 15 individuals from TCGA (The Cancer Genome Atlas). We furhter log2-transformed and mean-centred the data. For demonstration purpose, we used 20% of the expression data containing 2661 mRNA and 142 miRNA expression. Moreover, the corresponding sequence-based miRNA-target site matrix $\mathbf{W}$ was downloaded from TargetScanHuman 6.2 database [3] and the gene-gene interaction (GGI) data matrix $\mathbf{H}$ including transcription factor binding sites (TFBS) and protein-protein interaction (PPI) data were processed from TRANSFAC [6] and BioGrid [5], respectively.

```
> load(system.file("extdata/tcga_brca_testdata.RData", package="Mirsynergy"
```

Given as input the 2661×15 mRNA and 142×15 miRNA expression matrix along with the 2661×142 target site matrix, we first construct an expression-based miRNA-mRNA interaction score (MMIS) matrix using LASSO from *glmnet* by treating mRNA as response and miRNA as input variables [2].

```
> load(system.file("extdata/toy_modules.RData", package="Mirsynergy"))
> plot_modules(V,W,H)
```
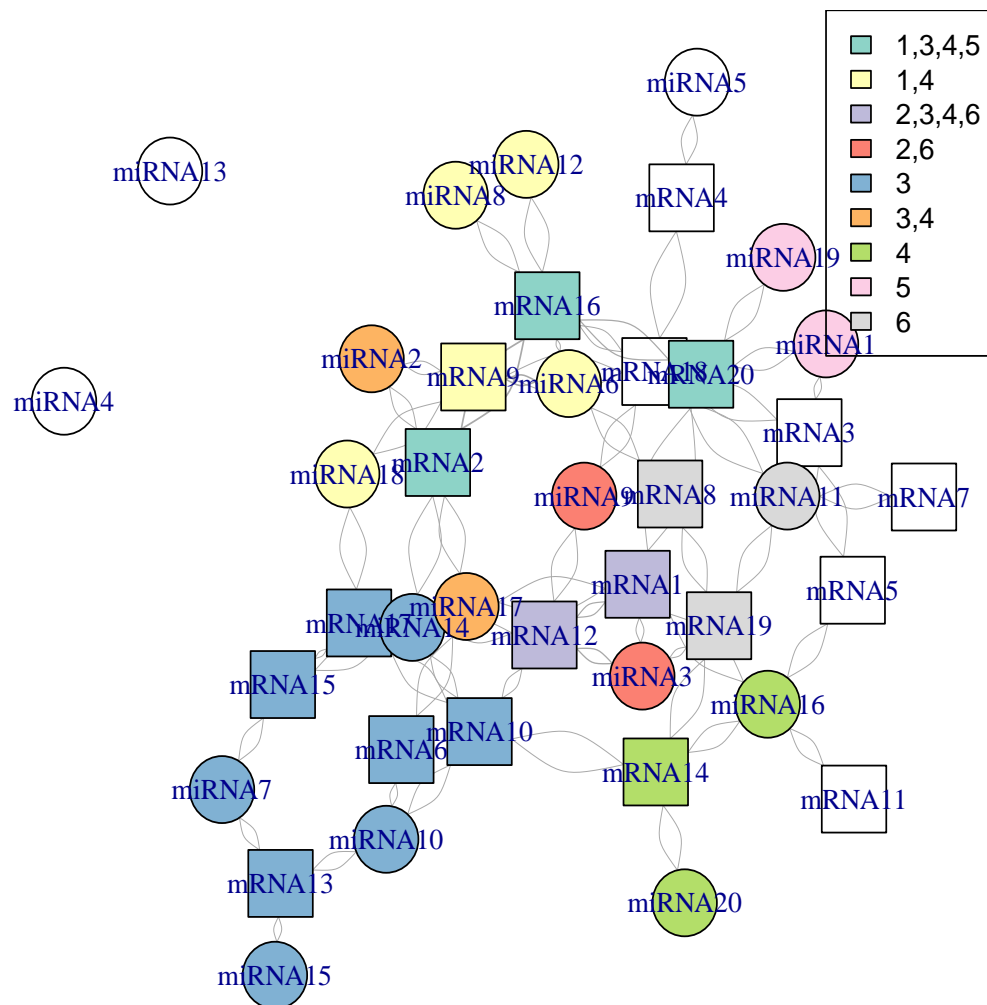

Figure 1: Module assignment on a toy example.

```
> library(glmnet)
> ptm <- proc.time()
> # lasso across all samples
> # X: N x T (input variables)
> #
> obs <- t(Z)   # T x M
> # run LASSO to construct W
> W <- lapply(1:nrow(X), function(i) {
+
+           pred <- matrix(rep(0, nrow(Z)), nrow=1,
+                     dimnames=list(rownames(X)[i], rownames(Z)))
+
+           c_i <- t(matrix(rep(C[i,,drop=FALSE], nrow(obs)), ncol=nrow(obs))
+
+           c_i <- (c_i > 0) + 0 # convert to binary matrix
+
+           inp <- obs * c_i
+
+           # use only miRNA with at least one non-zero entry across T sample.
+           inp <- inp[, apply(abs(inp), 2, max)>0, drop=FALSE]
+
+           if(ncol(inp) >= 2) {
+
+                   # NOTE: negative coef means potential parget (remove inte.
+                   x <- coef(cv.glmnet(inp, X[i,], nfolds=3), s="lambda.min"
+
+                   pred[, match(colnames(inp), colnames(pred))] <- x
+           }
+           pred[pred>0] <- 0
+
+           pred <- abs(pred)
+
+           pred[pred>1] <- 1
+
+           pred
+ })
> W <- do.call("rbind", W)
> dimnames(W) <- dimnames(C)
> print(sprintf("Time elapsed for LASSO: %.3f (min)",
+           (proc.time() - ptm)[3]/60))

[1] "Time elapsed for LASSO: 0.931 (min)"
```

Given the **W** and **H**, we can now apply `mirsynergy` to obtain MiRM assignments.

```
> V <- mirsynergy(W, H, verbose=FALSE)
> print_modules2(V)

M1 (density=3.25e-02; synergy=1.78e-01):
hsa-miR-4271 hsa-miR-181c hsa-miR-1193 hsa-miR-3672 hsa-miR-609
TUB TRANK1 GALK2 PLEK SMG5 KCNJ10 RAI1 ANP32A PEG3 ABTB2
M2 (density=3.87e-02; synergy=2.38e-01):
hsa-miR-302a hsa-miR-494 hsa-miR-302e hsa-miR-3125 hsa-let-7e hsa-miR-3134
CLP1 NSF TSEN34 RELN MYCN SLC2A4 TRPV6 LEFTY2 LRP8 IDH1 ZNF473
M3 (density=2.72e-02; synergy=1.61e-01):
hsa-miR-320e hsa-miR-340 hsa-miR-552 hsa-miR-610 hsa-miR-1271 hsa-miR-1297
CYP4V2 ACADSB AGPAT5 GLIS2 ITPR2 PALLD FGF1 SYT1
M4 (density=7.71e-02; synergy=2.07e-01):
hsa-miR-759 hsa-miR-1273d hsa-miR-495
CACNA1B NKX2-1 D4S234E GABBR2 RFX4
M5 (density=2.05e-02; synergy=1.89e-01):
hsa-miR-513b hsa-miR-30b hsa-miR-340 hsa-miR-620 hsa-miR-610 hsa-miR-921 hsa
ATF1 CYP4V2 C6orf170 STAC GPR126 ACADSB BOLL ITPR2 CDC25A PALLD ABCA13 TEAD1
M6 (density=4.05e-02; synergy=2.02e-01):
hsa-miR-626 hsa-miR-621 hsa-miR-122 hsa-miR-3658 hsa-miR-762
SNX16 PCNT KIAA0947 FAM84A CTPS CCDC25 MDGA2
M7 (density=4.63e-02; synergy=1.97e-01):
hsa-miR-1912 hsa-miR-4284 hsa-miR-555 hsa-miR-617
FOXM1 TGIF2 TMEM194B XPO5 IPO9 SASS6
M8 (density=3.8e-02; synergy=2.36e-01):
hsa-miR-3183 hsa-miR-4308 hsa-miR-759 hsa-miR-1273d hsa-miR-495 hsa-miR-519d
VPS37B ZC3HAV1L CACNA1B AIF1L GFOD2 NKX2-1 D4S234E GABBR2 SYNM RFX4 PCDHA11
M9 (density=5.86e-02; synergy=1.66e-01):
hsa-miR-541 hsa-miR-1229 hsa-miR-33a
EBF1 PCDH7 EPHA8 ZNF746
M10 (density=1.17e-02; synergy=1.83e-01):
hsa-miR-320e hsa-miR-93 hsa-miR-513b hsa-miR-30b hsa-miR-340 hsa-miR-424 hsa
ATF1 RAB27B NUAK1 CYP4V2 C6orf170 STAC GPR126 ACADSB AGPAT5 SLC40A1 BOLL ITI
M11 (density=2.37e-02; synergy=1.63e-01):
hsa-miR-4271 hsa-miR-181c hsa-miR-1193 hsa-miR-3672 hsa-miR-4293 hsa-miR-609
TUB TRANK1 GALK2 PLEK SMG5 KCNJ10 RAI1 ANP32A PEG3 ABTB2
M12 (density=5.73e-02; synergy=1.99e-01):
hsa-miR-4328 hsa-miR-548m
POLD3 ANP32E LMO4 UCHL5 ITSN1 PAPD7 DEPDC1 AGK KIF1B RAB3IP
M13 (density=8.47e-02; synergy=2.11e-01):
hsa-miR-4311 hsa-miR-601
WDR43 SEH1L PPM1L FAM60A RIMS2
M14 (density=7.89e-02; synergy=1.45e-01):
hsa-miR-891b hsa-miR-1322
CBFB ZNF644 CSDE1
```

```
M15 (density=4.73e-02; synergy=1.85e-01):
hsa-miR-98 hsa-miR-665 hsa-miR-661
TBX5 ANAPC7 MDC1 DUSP4 COL11A1 CHMP4C C5orf62 GATA4 PLEKHG6
M16 (density=3.96e-02; synergy=1.35e-01):
hsa-miR-185 hsa-miR-625 hsa-miR-4276
MFRP HAUS5 NFIX SYNGAP1
M17 (density=7.2e-02; synergy=2.01e-01):
hsa-miR-548n hsa-miR-629
SLC25A3 ZRANB2 PPP2R4 PSIP1 TRA2B HDGF DEK
M18 (density=4.08e-02; synergy=1.11e-01):
hsa-miR-520b hsa-miR-137 hsa-miR-372
ACSL6 LEFTY2 TP63 BNC1 FOXF2
M19 (density=7.35e-02; synergy=1.91e-01):
hsa-miR-377 hsa-miR-448
PPM1L UBAP2L YEATS2 PPP5C EPHA8 ZNF746 MAP3K7 DAAM1
M20 (density=5.8e-02; synergy=2.09e-01):
hsa-miR-4311 hsa-miR-4290 hsa-miR-601
WDR43 SEH1L PPM1L FAM60A RGS9BP RIMS2
M21 (density=1.63e-02; synergy=1.92e-01):
hsa-miR-320e hsa-miR-513b hsa-miR-30b hsa-miR-340 hsa-miR-4309 hsa-miR-620
TBX5 ATF1 CYP4V2 C6orf170 STAC GPR126 ACADSB AGPAT5 BOLL ITPR2 ANAPC7 MDC1
M22 (density=1.4e-02; synergy=1.85e-01):
hsa-miR-320e hsa-miR-513b hsa-miR-3201 hsa-miR-30b hsa-miR-340 hsa-miR-620
ATF1 RAB27B CYP4V2 C6orf170 STAC GPR126 ACADSB AGPAT5 PTPRZ1 BOLL ITPR2 CDC2

> print(sprintf("Time elapsed (LASSO+Mirsynergy): %.3f (min)",
+   (proc.time() - ptm)[3]/60))

[1] "Time elapsed (LASSO+Mirsynergy): 1.079 (min)"
```

There are several convenience functions implemented in the package to generate summary information such as Fig.2. In particular, the plot depicts the m/miRNA distribution across modules (upper panels) as well as the synergy distribution by itself and as a function of the number of miRNA (bottom panels).

For more details, please refer to our paper (manuscript in prep.).

4   Session Info

```
> sessionInfo()

R version 3.1.1 (2014-07-10)
Platform: i386-w64-mingw32/i386 (32-bit)

locale:
[1] LC_COLLATE=C
```
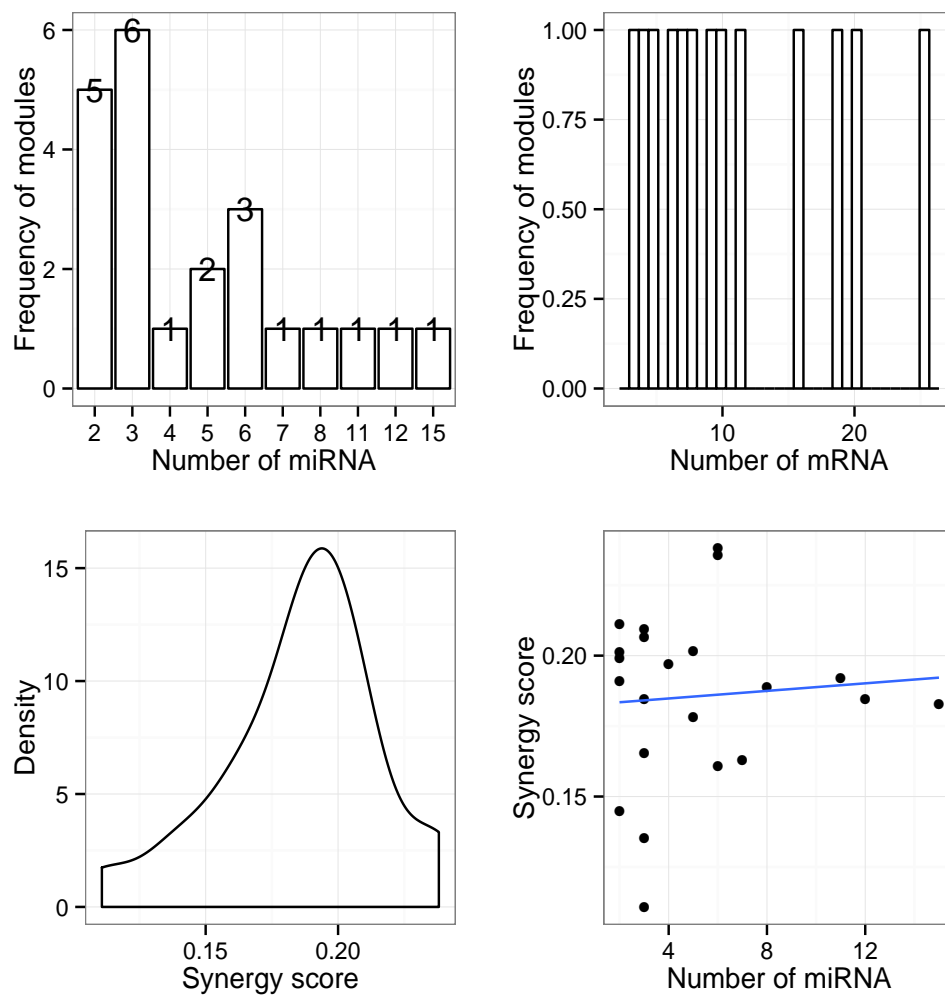
```
> plot_module_summary(V)
```


Figure 2: Summary information on MiRM using test data from TCGA-BRCA. Top panels: m/miRNA distribution across modulesas; Bottom panels: the synergy distribution by itself and as a function of the number of miRNA.

```
[2] LC_CTYPE=English_United States.1252
[3] LC_MONETARY=English_United States.1252
[4] LC_NUMERIC=C
[5] LC_TIME=English_United States.1252

attached base packages:
[1] stats     graphics  grDevices utils     datasets  methods   base

other attached packages:
[1] glmnet_1.9-8    Matrix_1.1-4    Mirsynergy_1.0.1 ggplot2_1.0.0
[5] igraph_0.7.1

loaded via a namespace (and not attached):
 [1] MASS_7.3-33          RColorBrewer_1.0-5 Rcpp_0.11.2          colorspace_1.
 [5] digest_0.6.4         evaluate_0.5.5     formatR_0.10         grid_3.1.1
 [9] gridExtra_0.9.1      gtable_0.1.2       knitr_1.6            labeling_0.2
[13] lattice_0.20-29      munsell_0.4.2      parallel_3.1.1       plyr_1.8.1
[17] proto_0.3-10         reshape_0.8.5      reshape2_1.4         scales_0.2.4
[21] stringr_0.6.2        tools_3.1.1
```

References

[1] David P Bartel. MicroRNAs: Target Recognition and Regulatory Functions. *Cell*, 136(2):215–233, January 2009.

[2] Jerome Friedman, Trevor Hastie, and Rob Tibshirani. Regularization Paths for Generalized Linear Models via Coordinate Descent. *Journal of statistical software*, 33(1):1–22, 2010.

[3] Robin C Friedman, Kyle Kai-How Farh, Christopher B Burge, and David P Bartel. Most mammalian mRNAs are conserved targets of microRNAs. *Genome Research*, 19(1):92–105, January 2009.

[4] Riccardo Spizzo, Milena S Nicoloso, Carlo M Croce, and George A Calin. SnapShot: MicroRNAs in Cancer. *Cell*, 137(3):586–586.e1, May 2009.

[5] Chris Stark, Bobby-Joe Breitkreutz, Andrew Chatr-Aryamontri, Lorrie Boucher, Rose Oughtred, Michael S Livstone, Julie Nixon, Kimberly Van Auken, Xiaodong Wang, Xiaoqi Shi, Teresa Reguly, Jennifer M Rust, Andrew Winter, Kara Dolinski, and Mike Tyers. The BioGRID Interaction Database: 2011 update. *Nucleic acids research*, 39(Database issue):D698–704, January 2011.

[6] E Wingender, X Chen, R Hehl, H Karas, I Liebich, V Matys, T Meinhardt, M Prüss, I Reuter, and F Schacherer. TRANSFAC: an integrated system for gene expression regulation. *Nucleic acids research*, 28(1):316–319, January 2000.