Mirsynergy: detect synergistic miRNA regulatory modules by overlapping neighbourhood expansion

Yue Li
yueli@cs.toronto.edu

October 13, 2014

1 Introduction

MicroRNAs (miRNAs) are $\sim$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)
```

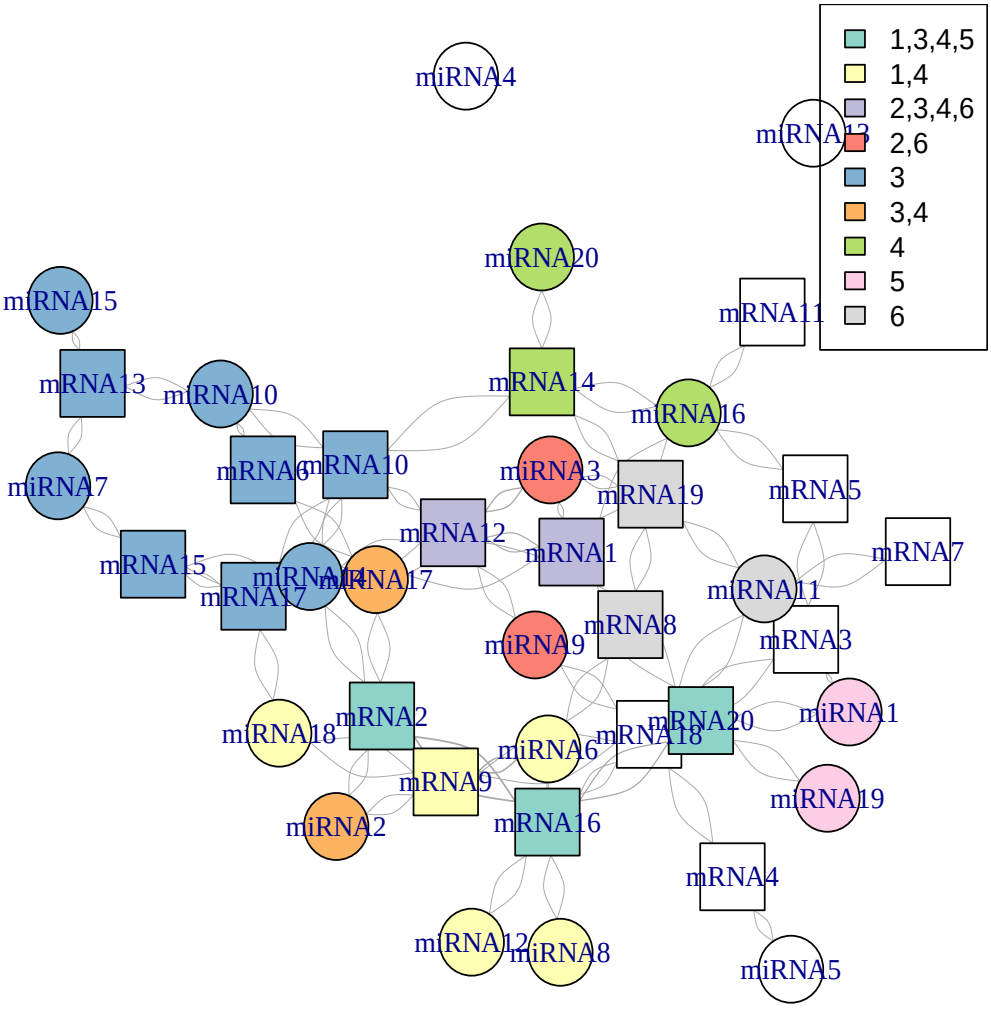

Figure 1: Module assignment on a toy example.

```
> 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: 1.250 (min)"
```

Given the $\mathbf{W}$ and $\mathbf{H}$, we can now apply `mirsynergy` to obtain MiRM assignments.

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

M1 (density=3.16e-02; synergy=2.16e-01):
hsa-miR-302a hsa-miR-520b hsa-miR-492 hsa-let-7d hsa-miR-302e hsa-miR-3134
GAB2 ETNK2 MYLK MET NSF BAMBI TRHDE SLC1A4 FBXO41 EGR3 SLC2A4 CRB2 TRPV6 LRI
M2 (density=4.29e-02; synergy=1.8e-01):
hsa-miR-608 hsa-miR-1273d hsa-miR-495 hsa-miR-4293 hsa-miR-4296
KCNQ4 NKX2-1 GABBR2 RTN4R CNTN2 RFX4
M3 (density=6.73e-02; synergy=1.93e-01):
hsa-miR-424 hsa-miR-935 hsa-miR-4252
RHPN2 SLC2A14 RELN PCDHA7 LRP8
M4 (density=4.07e-02; synergy=2.09e-01):
hsa-miR-4328 hsa-miR-621 hsa-miR-4309 hsa-miR-143
POLD3 LMO4 ITSN1 PAPD7 FGF1 AGK SNAP25 KIF1B NUP210 RAB3IP SYT1 STX1A NR2C1
M5 (density=4e-02; synergy=1.89e-01):
hsa-miR-4311 hsa-miR-424 hsa-miR-1193 hsa-miR-759 hsa-miR-601
SEH1L FAM60A SLC2A14 PPP1R8 PCDHA7 TAF7L
M6 (density=2.23e-02; synergy=1.92e-01):
hsa-miR-185 hsa-miR-30b hsa-miR-610 hsa-miR-1273d hsa-miR-495 hsa-miR-1254 h
RAB27B MON2 STAC ETNK2 RNF170 ELFN2 TRHDE MFRP SLC1A4 USP44 NKX2-1 EGR3 CRB2
M7 (density=4.02e-02; synergy=2.04e-01):
hsa-miR-626 hsa-miR-122 hsa-miR-3658 hsa-miR-4327 hsa-miR-762
PCNT KIAA0947 FAM84A CTPS EPHB4 PRKX MDGA2
M8 (density=3.71e-02; synergy=2.17e-01):
hsa-miR-374c hsa-miR-3183 hsa-miR-3692 hsa-miR-4308 hsa-miR-1273d hsa-miR-49
ZC3HAV1L GFOD2 NKX2-1 LPAR3 GABBR2 SYNM CNTN2 RFX4 PCDHA11
M9 (density=8.25e-02; synergy=1.8e-01):
hsa-miR-4272 hsa-miR-921
DLX6 ANP32E UCHL5 PSMB2
M10 (density=4.38e-02; synergy=1.96e-01):
hsa-miR-98 hsa-miR-4284 hsa-miR-1227 hsa-miR-3125
FOXM1 ATP7B TGIF2 EFNB3 TRHDE DUSP4 SLC2A12 COL11A1 PLEKHG6
M11 (density=2.33e-02; synergy=1.71e-01):
hsa-miR-4272 hsa-miR-340 hsa-miR-552 hsa-miR-921 hsa-miR-214 hsa-miR-210
STXBP4 DLX6 ACADSB ANP32E ITPR2 LMO4 UCHL5 ITSN1 PALLD FGF1 PSMB2 SNAP25 MLI
M12 (density=4.84e-02; synergy=2.29e-01):
hsa-miR-181c hsa-miR-3155 hsa-miR-214 hsa-let-7e
CLP1 UBE2E2 RNF8 CD163 SLC1A4 MDC1 PLEK UBE2D1 UBE2D4 ZNRF1 ARIH2 KCNJ10 PEI
M13 (density=1.96e-02; synergy=2.18e-01):
hsa-miR-302a hsa-miR-98 hsa-miR-4284 hsa-miR-520b hsa-miR-492 hsa-miR-1227 h
FOXM1 GAB2 ETNK2 ATP7B SCD TGIF2 MET NSF BAMBI TRHDE SLC1A4 DUSP4 SLC2A12 CO
M14 (density=1.12e-02; synergy=2.16e-01):
hsa-miR-320e hsa-miR-513b hsa-miR-185 hsa-miR-3183 hsa-miR-30b hsa-miR-340 h
RAB27B ZC3HAV1L STAC ETNK2 RNF170 GPR126 ACADSB PTGS2 AGPAT5 ELFN2 CELF2 TRI
```

```
M15 (density=8.63e-02; synergy=1.27e-01):
hsa-miR-548m hsa-miR-617
RCC2
M16 (density=5.18e-02; synergy=1.83e-01):
hsa-miR-4262 hsa-miR-33a hsa-miR-147
VCAN PROX1 TBPL1 HYOU1 RORA PCDH7 ABTB2
M17 (density=1.01e-01; synergy=1.92e-01):
hsa-miR-519e hsa-miR-494
RCBTB2 IGSF10 PNOC
M18 (density=5.85e-02; synergy=1.52e-01):
hsa-miR-1912 hsa-miR-4257 hsa-miR-555
XPO5 IPO9
M19 (density=8.27e-02; synergy=1.55e-01):
hsa-miR-4312 hsa-miR-4256
TCF15 AMD1
M20 (density=7.12e-02; synergy=2.53e-01):
hsa-miR-4271 hsa-miR-335 hsa-miR-4313
PTPRU UBE2E2 RNF8 UBE2D1 UBE2D4 ZNRF1 SMG5 ARIH2 PELI1 RNF150 PJA2
M21 (density=5.28e-02; synergy=1.46e-01):
hsa-miR-548n hsa-miR-629 hsa-miR-1910
SLC25A3 CCNG1 PPP2R4
M22 (density=6.41e-02; synergy=1.87e-01):
hsa-miR-374c hsa-miR-3692 hsa-miR-1273d hsa-miR-495
DUSP4 NKX2-1 GABBR2 CNTN2 RFX4
M23 (density=1.91e-02; synergy=2.3e-01):
hsa-miR-302a hsa-miR-98 hsa-miR-520b hsa-miR-492 hsa-miR-137 hsa-let-7d hsa
SLC25A3 GAB2 ETNK2 ATP7B CCNG1 SCD MYLK MET NSF BAMBI TRHDE PPP2R4 ACSL6 IT
M24 (density=8.67e-02; synergy=1.71e-01):
hsa-miR-377 hsa-miR-448
YEATS2 PPP5C RNGTT MAP3K7 DAAM1
M25 (density=1.84e-02; synergy=1.66e-01):
hsa-miR-4272 hsa-miR-340 hsa-miR-3174 hsa-miR-552 hsa-miR-610 hsa-miR-921 hs
STXBP4 DLX6 ACADSB ANP32E ITPR2 LMO4 UCHL5 ITSN1 PALLD FEN1 FGF1 PSMB2 SNAP2
M26 (density=5.15e-02; synergy=1.36e-01):
hsa-miR-608 hsa-miR-4293 hsa-miR-4296
KCNQ4 RTN4R
M27 (density=7.83e-02; synergy=1.54e-01):
hsa-miR-1267 hsa-miR-31
RPS6KL1 SLA2

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

[1] "Time elapsed (LASSO+Mirsynergy): 1.459 (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).

```
> plot_module_summary(V)
```

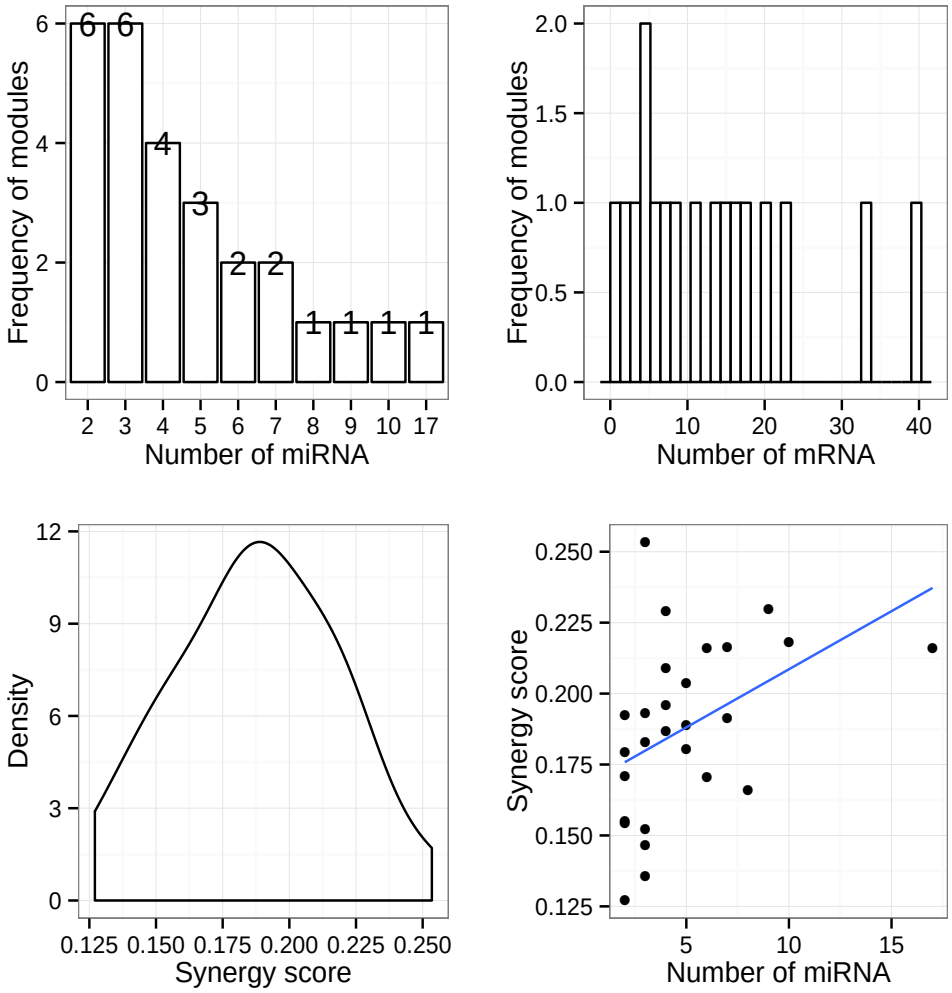

Figure 2: Summary information on MiRM using test data from TCGA-BRCA. Top panels: m/miRNA distribution across modulesas; Bottom panels: the synergy distribution by itself and as a function of the number of miRNA.

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

4 Session Info

```
> sessionInfo()
```

```
R version 3.1.1 Patched (2014-09-25 r66681)
Platform: x86_64-apple-darwin10.8.0 (64-bit)

locale:
[1] en_US.UTF-8/en_US.UTF-8/en_US.UTF-8/C/en_US.UTF-8/en_US.UTF-8

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

other attached packages:
[1] glmnet_1.9-8     Matrix_1.1-4     Mirsynergy_1.2.0 ggplot2_1.0.0
[5] igraph_0.7.1

loaded via a namespace (and not attached):
 [1] colorspace_1.2-4   digest_0.6.4       evaluate_0.5.5     formatR_1.0
 [5] grid_3.1.1         gridExtra_0.9.1    gtable_0.1.2       knitr_1.7
 [9] labeling_0.3       lattice_0.20-29    MASS_7.3-35        munsell_0.4.2
[13] parallel_3.1.1     plyr_1.8.1         proto_0.3-10       RColorBrewer_1
[17] Rcpp_0.11.3        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.