Mirsynergy: detect synergistic miRNA regulatory modules by overlapping neighbourhood expansion

Yue Li
yueli@cs.toronto.edu

February 19, 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)
```

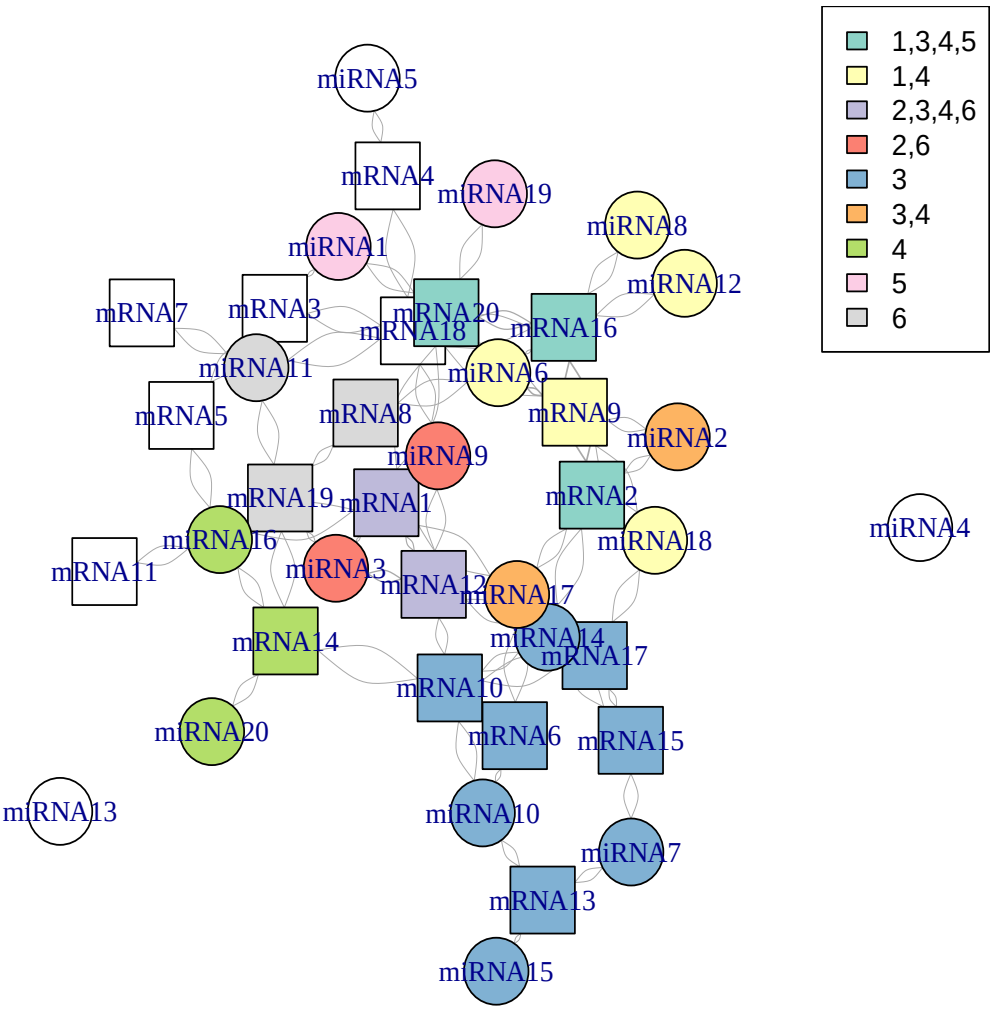

Figure 1: Module assignment on a toy example.

```
> 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: 0.575 (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=6.23e-02; synergy=2e-01):
hsa-miR-759 hsa-miR-1273d hsa-miR-495 hsa-miR-4252
CACNA1B NKX2-1 GRHL1 CNTN2 RFX4
M2 (density=3.76e-02; synergy=2.08e-01):
hsa-miR-3183 hsa-miR-3174 hsa-miR-764 hsa-miR-495 hsa-miR-519d hsa-miR-4316
RASD2 CACNA1B AIF1L GFOD2 NKX2-1 ZSCAN20 GABBR2 EFR3B CNTN2 RFX4
M3 (density=4.22e-02; synergy=2.41e-01):
hsa-miR-302a hsa-miR-520b hsa-miR-302e hsa-miR-3134
CLP1 TSEN34 FBXO41 SLC2A4 TRPV6 LEFTY2 LRP8 FTSJD1 IDH1 BNC1 DPP3 ZNF473
M4 (density=5.32e-02; synergy=1.99e-01):
hsa-miR-4328 hsa-miR-4309 hsa-miR-548m
SMAD9 POLD3 LMO4 PAPD7 DEPDC1 AGK NUP210 MLL2
M5 (density=5.46e-02; synergy=2.28e-01):
hsa-miR-4311 hsa-miR-424 hsa-miR-1193 hsa-miR-601
WDR43 LRRCC1 SEH1L FAM60A PCDHA7 RIMS2 TAF7L PCDHA6
M6 (density=2.77e-02; synergy=2.12e-01):
hsa-miR-1912 hsa-miR-4284 hsa-miR-181c hsa-miR-1227 hsa-miR-1261 hsa-miR-125
FOXM1 EN1 TGIF2 TMEM194B CD163 ERC2 GALK2 PLEK SLC2A12 KCNJ10 GCNT2
M7 (density=4.34e-02; synergy=1.46e-01):
hsa-miR-891b hsa-miR-2054 hsa-miR-1322
CBFB ZNF644 TRIM33 ITGA2 CSDE1 KCNJ10 RUNX1
M8 (density=5.32e-02; synergy=2.12e-01):
hsa-miR-216a hsa-miR-4262 hsa-miR-181d hsa-miR-3941
EN1 TBPL1 HYOU1 NF1 CELSR3 PCDHA7 SLC7A6OS ABTB2
M9 (density=6.25e-02; synergy=1.85e-01):
hsa-miR-541 hsa-miR-1229 hsa-miR-33a
ZNF423 EBF1 PCDH7 EPHA8
M10 (density=2.81e-02; synergy=1.6e-01):
hsa-miR-30b hsa-miR-4272 hsa-miR-4256 hsa-miR-921
ARL10 RAB27B CTTNBP2NL TXNDC5 ELFN2 PGM3 CAP1 ANP32E UCHL5 C14orf129 ERRFI1
M11 (density=3.03e-02; synergy=2.12e-01):
hsa-miR-374c hsa-miR-3692 hsa-miR-759 hsa-miR-1273d hsa-miR-3665 hsa-miR-495
DCLK2 ZC3H11A CACNA1B NKX2-1 GRHL1 HDAC9 CNTN2 RFX4 ONECUT1
M12 (density=6.53e-02; synergy=1.92e-01):
hsa-miR-519e hsa-miR-4313 hsa-miR-4290
TRAF4 DNAJC11 RGS9BP
M13 (density=3.74e-02; synergy=2.13e-01):
hsa-miR-626 hsa-miR-122 hsa-miR-3658 hsa-miR-762 hsa-miR-4296
ZRANB2 POU5F1 KIAA0947 FREM2 FAM84A CTPS TRA2B EPHB4 MDGA2
M14 (density=4.3e-02; synergy=1.66e-01):
hsa-miR-320e hsa-miR-340 hsa-miR-1297
GIPC2 CYP4V2 ACADSB AGPAT5 ITPR2 PALLD FGF1 SYT1
```

```
M15 (density=8.54e-02; synergy=1.52e-01):
hsa-miR-185 hsa-miR-625
GEMIN8 NFIX
M16 (density=9.62e-02; synergy=1.86e-01):
hsa-miR-4308 hsa-miR-1301
VPS37B BLMH SYNM
M17 (density=4e-02; synergy=1.08e-01):
hsa-miR-3148 hsa-miR-3689b hsa-miR-4276
SLC6A8 SOAT1
M18 (density=8.41e-02; synergy=1.98e-01):
hsa-miR-377 hsa-miR-448
YEATS2 RNGTT MAPK6 MAP3K7 DAAM1
M19 (density=3.01e-02; synergy=1.68e-01):
hsa-miR-320e hsa-miR-340 hsa-miR-610 hsa-miR-3161 hsa-miR-1297
GIPC2 CYP4V2 ACADSB AGPAT5 ITPR2 PALLD FGF1 PUS7 SYT1
M20 (density=2.58e-02; synergy=2.33e-01):
hsa-miR-4311 hsa-miR-216a hsa-miR-424 hsa-miR-1193 hsa-miR-487a hsa-miR-426
WDR43 LRRCC1 SEH1L EN1 FAM60A TBPL1 HYOU1 NF1 CELSR3 PCDHA7 SLC7A6OS ABTB2 I
M21 (density=1.59e-02; synergy=1.93e-01):
hsa-miR-1912 hsa-miR-30b hsa-miR-4284 hsa-miR-181c hsa-miR-4272 hsa-miR-122
RAB27B FOXM1 EN1 CTTNBP2NL TGIF2 TMEM194B ELFN2 CD163 ERC2 PGM3 GALK2 ANP32I
M22 (density=3.01e-02; synergy=1.77e-01):
hsa-miR-320e hsa-miR-340 hsa-miR-4257 hsa-miR-610 hsa-miR-1297
GIPC2 CYP4V2 ACADSB AGPAT5 ITPR2 PALLD FGF1 RSPO4 PUS7 SYT1
```

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

[1] "Time elapsed (LASSO+Mirsynergy): 0.619 (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 3.0.2 (2013-09-25)
Platform: x86_64-apple-darwin10.8.0 (64-bit)

locale:
[1] C/en_CA.UTF-8/en_CA.UTF-8/C/en_CA.UTF-8/en_CA.UTF-8
```

```
> plot_module_summary(V)
```

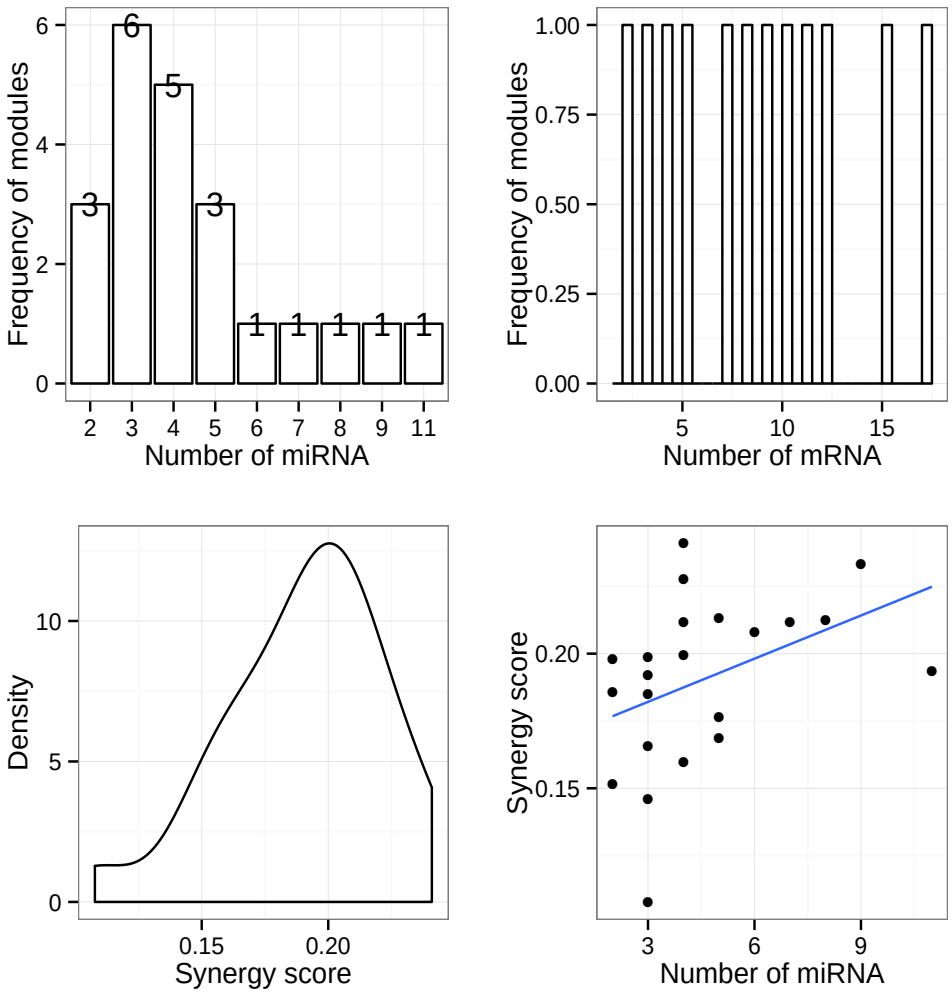

Figure 2: Summary information on MiRM using test data from TCGA-BRCA. Top panels: m/miRNA distribution across modulesas; Bottom panels: the synergy distribution by itself and as a function of the number of miRNA.

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

other attached packages:
[1] glmnet_1.9-5      Matrix_1.0-14     lattice_0.20-24   Mirsynergy_0.99.2
[5] ggplot2_0.9.3.1   igraph_0.6.5-2

loaded via a namespace (and not attached):
 [1] MASS_7.3-29         RColorBrewer_1.0-5 colorspace_1.2-4   dichromat_2.0-
 [5] digest_0.6.3        evaluate_0.5.1     formatR_0.10       grid_3.0.2
 [9] gridExtra_0.9.1     gtable_0.1.2       knitr_1.5          labeling_0.2
[13] munsell_0.4.2       parallel_3.0.2     plyr_1.8           proto_0.3-10
[17] reshape_0.8.4       reshape2_1.2.2     scales_0.2.3       stringr_0.6.2
[21] tools_3.0.2
```

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.