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)
```

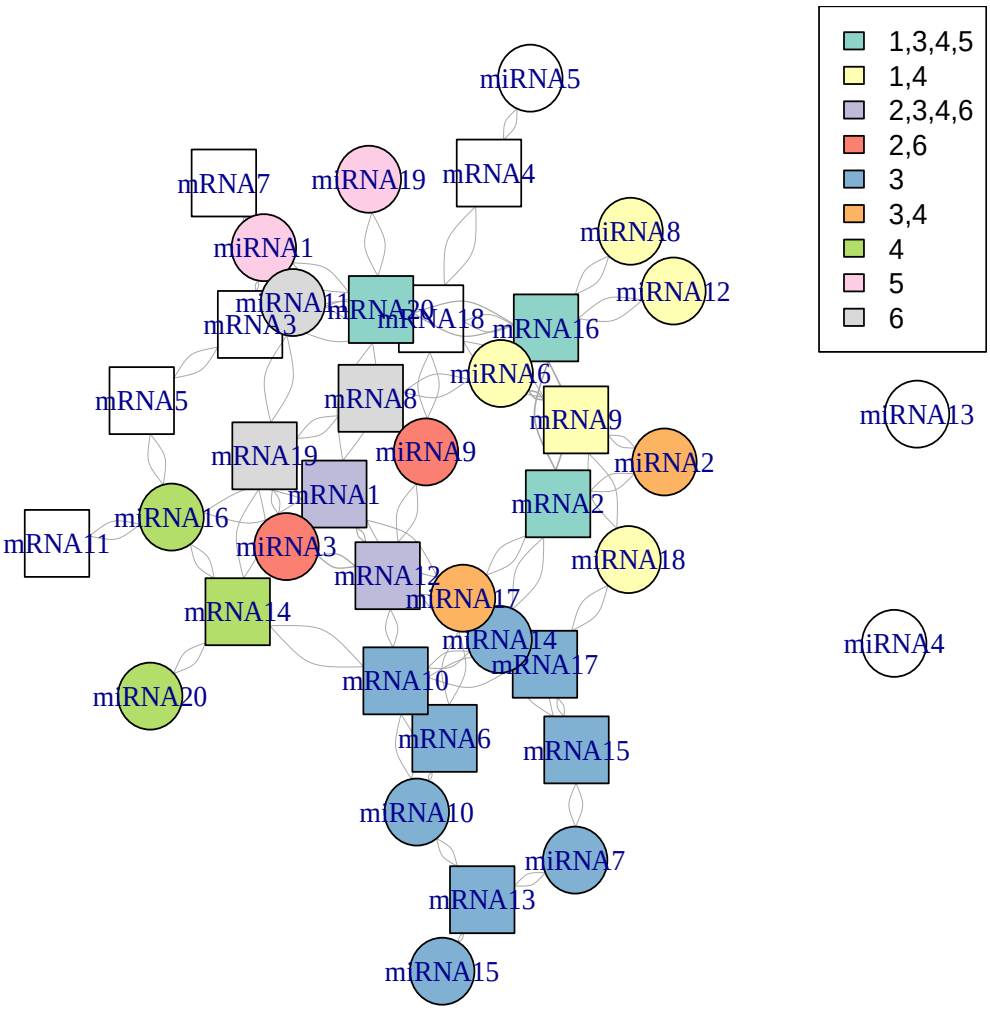

Figure 1: Module assignment on a toy example.

```
> 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: 1.071 (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=1.94e-02; synergy=2.1e-01):
hsa-miR-302a hsa-miR-3201 hsa-miR-520b hsa-miR-1193 hsa-miR-548n hsa-miR-92
PLAU LRP2 EPHA2 TSEN34 EBF1 PKNOX1 FBXO41 SLC2A4 TRPV6 KPNA3 FTSJD1 PCDH7 I
M2 (density=3.03e-02; synergy=1.91e-01):
hsa-miR-98 hsa-miR-4284 hsa-miR-492 hsa-miR-1227 hsa-miR-1261 hsa-miR-3941
TBX5 FOXM1 ATP7B TGIF2 EFNB3 RNF165 SLITRK3 SLC2A12 COL11A1 C5orf62
M3 (density=2.12e-02; synergy=2.12e-01):
hsa-miR-320e hsa-miR-513b hsa-miR-30b hsa-miR-340 hsa-miR-2054 hsa-miR-1297
GIPC2 ATF1 ID2 RAB27B C6orf170 RBL2 GPR126 ACADSB PTGS2 AGPAT5 ELFN2 CELF2 I
M4 (density=2.8e-02; synergy=2.57e-01):
hsa-miR-4328 hsa-miR-621 hsa-miR-4309 hsa-miR-939 hsa-miR-143
BAZ1B RAI14 ARPC4 USP15 ZRANB2 POLD3 RRP1B YTHDC1 MYO1C LMO4 ITSN1 PAPD7 SA
M5 (density=8.28e-02; synergy=2.54e-01):
hsa-miR-4311 hsa-miR-601
WDR43 LRRCC1 SEH1L FAM60A RIMS2 TAF7L
M6 (density=1.1e-01; synergy=1.61e-01):
hsa-miR-759 hsa-miR-605
D4S234E
M7 (density=1.93e-02; synergy=2.33e-01):
hsa-miR-4328 hsa-miR-621 hsa-miR-181c hsa-miR-4309 hsa-miR-216a hsa-miR-939
EN1 ARPC4 DPP6 POLD3 RAB35 USP6NL PROX1 TBPL1 CD163 TRANK1 HYOU1 PLEK ANXA1
M8 (density=1.03e-01; synergy=1.69e-01):
hsa-miR-891b hsa-miR-1322
CBFB LAT RUNX1
M9 (density=7.26e-02; synergy=2e-01):
hsa-miR-626 hsa-miR-3155 hsa-let-7e
FREM2 SLC1A4 CTPS MDGA2 ZNF473
M10 (density=1.38e-02; synergy=2.11e-01):
hsa-miR-302a hsa-miR-3201 hsa-miR-3183 hsa-miR-520b hsa-miR-1193 hsa-miR-75
ZC3HAV1L EPHA2 AIF1L TSEN34 EBF1 GFOD2 NKX2-1 GRHL1 D4S234E FBXO41 SLC2A4 T
M11 (density=5.98e-02; synergy=1.32e-01):
hsa-miR-185 hsa-miR-1254
STX1B GEMIN8 NFIX
M12 (density=2.32e-02; synergy=1.4e-01):
hsa-miR-3201 hsa-miR-759 hsa-miR-548n hsa-miR-921 hsa-miR-605 hsa-miR-33a hs
VCAN EBF1 D4S234E PCDH7 RNF150
M13 (density=1.14e-02; synergy=2.15e-01):
hsa-miR-320e hsa-miR-626 hsa-miR-513b hsa-miR-30b hsa-miR-340 hsa-miR-1229 h
GIPC2 ATF1 ID2 RAB27B C6orf170 ETNK2 RBL2 RNF170 GPR126 CPEB4 ACADSB PTGS2 I
M14 (density=3.42e-02; synergy=2.02e-01):
hsa-miR-374c hsa-miR-3148 hsa-miR-3692 hsa-miR-4276 hsa-miR-3714
SLC4A10 STAC UBAP2L TXNDC5 SLC16A1 YEATS2 SLC4A7 DUSP15 SOAT1 FECH
```

```
M15 (density=3.11e-02; synergy=1.95e-01):
hsa-miR-98 hsa-miR-4284 hsa-miR-492 hsa-miR-1227 hsa-miR-1261 hsa-miR-3125
TBX5 FOXM1 ATP7B TGIF2 EFNB3 RNF165 SLITRK3 SLC2A12 COL11A1 C5orf62
M16 (density=2.68e-02; synergy=2.04e-01):
hsa-miR-513b hsa-miR-30b hsa-miR-340 hsa-miR-2054 hsa-miR-4296
GIPC2 ATF1 ID2 RAB27B C6orf170 RBL2 GPR126 ACADSB ELFN2 BOLL CDC25A ABCA13 M
M17 (density=4.22e-02; synergy=1.02e-01):
hsa-miR-3165 hsa-miR-3154
RFX5 PPPDE2 C9orf150
M18 (density=1.55e-02; synergy=2.06e-01):
hsa-miR-320e hsa-miR-513b hsa-miR-30b hsa-miR-340 hsa-miR-137 hsa-miR-181d M
GIPC2 ATF1 ID2 RAB27B C6orf170 RBL2 GPR126 ACADSB PTGS2 AGPAT5 ELFN2 CELF2 M
M19 (density=8.63e-02; synergy=1.91e-01):
hsa-miR-665 hsa-miR-661
GNA13 CHMP4A CHMP4C F2R
M20 (density=5.56e-02; synergy=3.32e-01):
hsa-miR-4256 hsa-miR-617
SIAH1 UBE2E2 MDM4 TRIM7 CDKN1A CACYBP TRIM23 RNF165 RNF8 MLL4 MDC1 UBE2D1 UB
M21 (density=2.49e-02; synergy=1.86e-01):
hsa-miR-374c hsa-miR-3148 hsa-miR-3692 hsa-miR-541 hsa-miR-3689b hsa-miR-427
SLC4A10 STAC UBAP2L TXNDC5 SLC16A1 YEATS2 SLC4A7 DUSP15 SOAT1 FECH
M22 (density=4.34e-02; synergy=1.36e-01):
hsa-miR-320e hsa-miR-1301 hsa-miR-1297
AGPAT5 UBE4B FAM118B TMEM184B
M23 (density=4.58e-02; synergy=1.66e-01):
hsa-miR-519e hsa-miR-1910 hsa-miR-4290
RCBTB2 TRAF4 DNAJC11 RGS9BP SIT1
M24 (density=2.38e-02; synergy=5.21e-02):
hsa-miR-548y hsa-miR-3135
DOCK2 PXDN
M25 (density=1.84e-02; synergy=2.01e-01):
hsa-miR-302a hsa-miR-3201 hsa-miR-520b hsa-miR-1193 hsa-miR-759 hsa-miR-548m
EPHA2 TSEN34 EBF1 D4S234E FBXO41 SLC2A4 TRPV6 KPNA3 FTSJD1 PCDH7 IDH1 BNC1 N
```

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

[1] "Time elapsed (LASSO+Mirsynergy): 1.233 (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.).

```
> plot_module_summary(V)
```

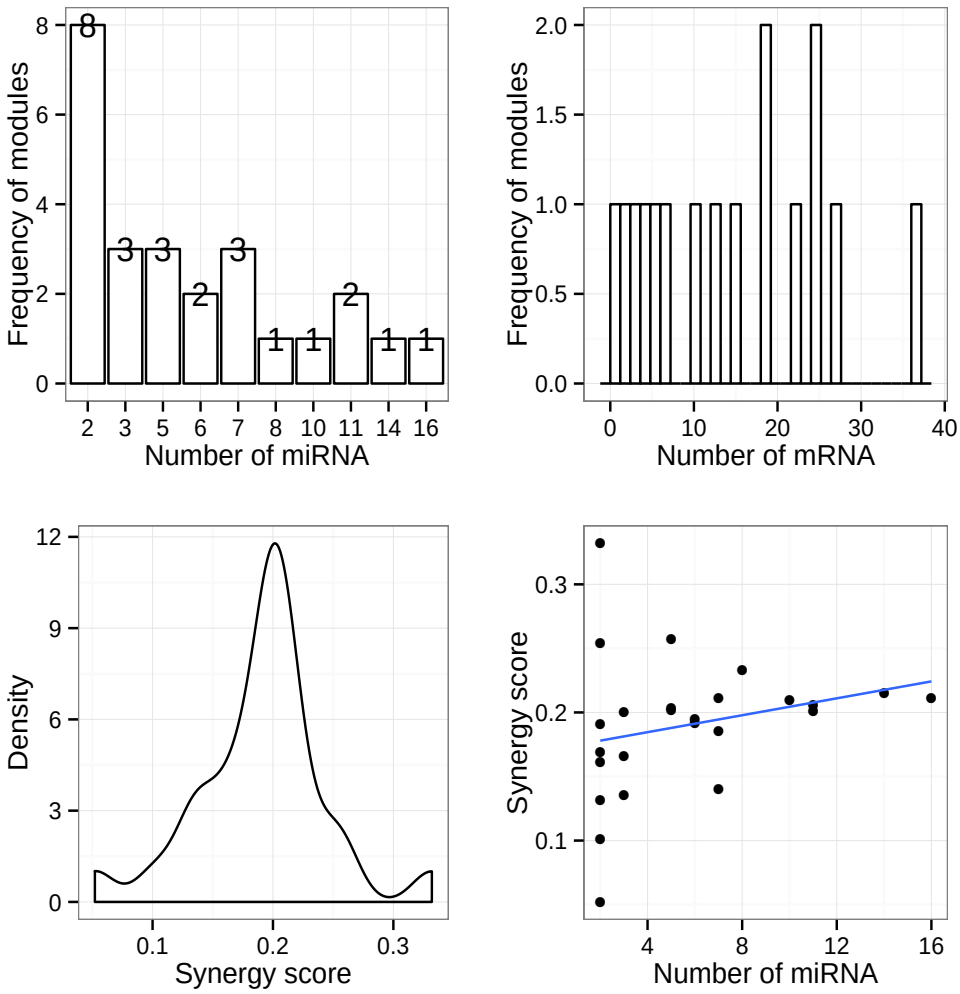

Figure 2: Summary information on MiRM using test data from TCGA-BRCA. Top panels: m/miRNA distribution across modulesas; Bottom panels: the synergy distribution by itself and as a function of the number of miRNA.

4 Session Info

```
> sessionInfo()

R version 3.1.1 Patched (2014-09-25 r66681)
Platform: x86_64-apple-darwin13.1.0 (64-bit)

locale:
[1] C/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] MASS_7.3-35        RColorBrewer_1.0-5 Rcpp_0.11.3        colorspace_1.2
 [5] digest_0.6.4       evaluate_0.5.5     formatR_1.0        grid_3.1.1
 [9] gridExtra_0.9.1    gtable_0.1.2       knitr_1.7          labeling_0.3
[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.