Mirsynergy: detect synergistic miRNA regulatory modules by overlapping neighbourhood expansion

Yue Li
yueli@cs.toronto.edu

October 13, 2015

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)
```

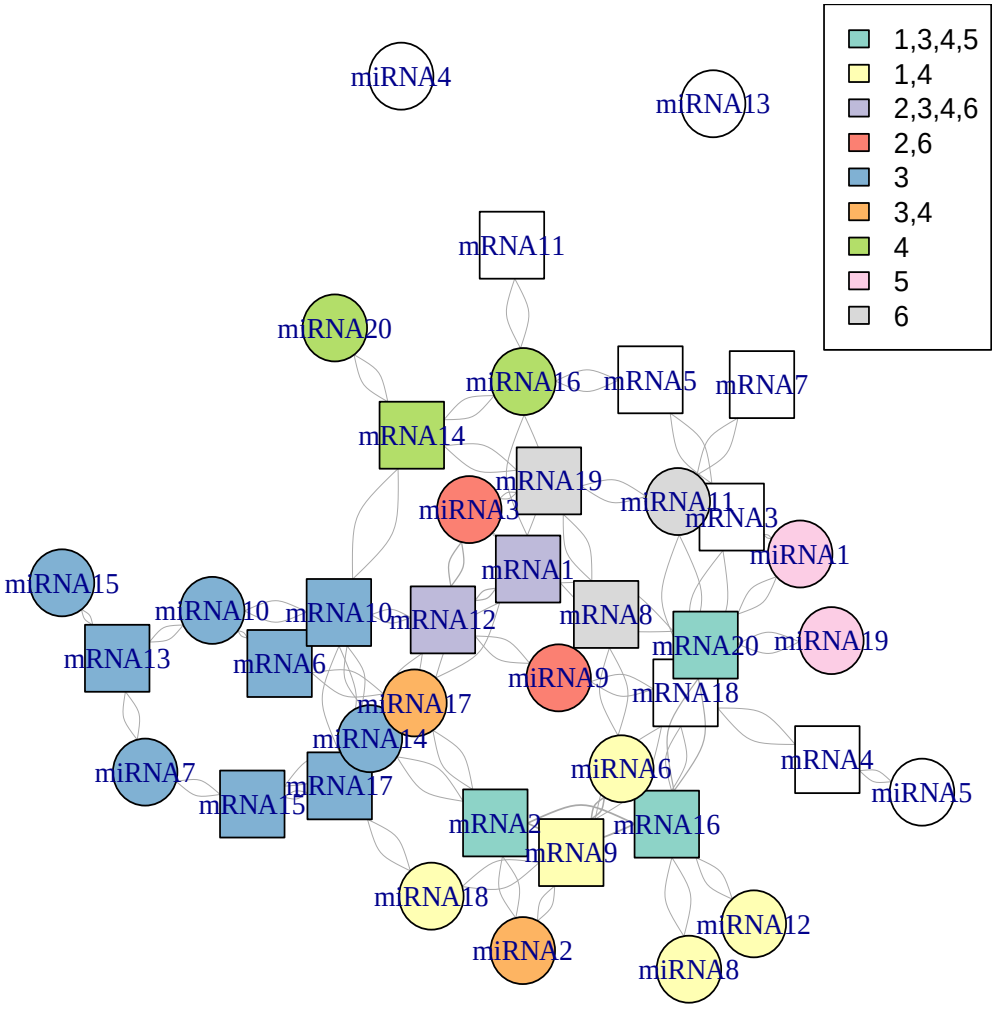

Figure 1: Module assignment on a toy example.

```
> 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
+                 x <- as.numeric(coef(glmnet(inp, X[i,]), s=0.1)[-1])
+                 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: 0.248 (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=3.31e-02; synergy=2.14e-01):
hsa-miR-302a hsa-miR-520b hsa-miR-302e hsa-miR-3134
MYBL1 LHX8 CLP1 GLS ACVR1 BAMBI TRHDE TSEN34 PRR16 FBXO41 MYCN TRPV6 LEFTY2
M2 (density=5.03e-02; synergy=2.02e-01):
hsa-miR-4311 hsa-miR-424 hsa-miR-1193
WDR43 LRRCC1 SEH1L PPM1L FAM60A SLC2A14 AKAP6 SIX3 PCDHA7 OTX1 TAF7L PCDHA6
M3 (density=3.53e-02; synergy=2.05e-01):
hsa-miR-3183 hsa-miR-3174 hsa-miR-764 hsa-miR-1273d hsa-miR-495 hsa-miR-519
RASD2 ZC3HAV1L CACNA1B AQP4 AIF1L NKX2-1 GRHL1 RAP1GAP2 ZSCAN20 LPAR3 GABBR
M4 (density=2.94e-02; synergy=1.67e-01):
hsa-miR-4271 hsa-miR-181c hsa-miR-4313 hsa-miR-609 hsa-miR-518e
PTPRU EN1 ANKRD17 NOL4 TUB CD163 TRANK1 GALK2 GATA6 PLEK SMG5 KPNA3 KCNJ10
M5 (density=3.01e-02; synergy=2.07e-01):
hsa-miR-676 hsa-miR-4311 hsa-miR-424 hsa-miR-1193 hsa-miR-608 hsa-miR-3161
WDR43 AMOTL1 LRRCC1 SEH1L PPM1L FAM107A FAM60A KCNQ4 SLC2A14 AKAP6 SIX3 PCD
M6 (density=1.66e-02; synergy=2.3e-01):
hsa-miR-93 hsa-miR-676 hsa-miR-4311 hsa-miR-625 hsa-miR-424 hsa-miR-1193 hsa
WDR43 AMOTL1 LRRCC1 SEH1L FAIM2 PPM1L KCNK10 FAM107A FAM60A GK NOL4 SLC40A1
M7 (density=3.94e-02; synergy=1.76e-01):
hsa-miR-3201 hsa-miR-18b hsa-miR-335 hsa-let-7e
C10orf140 ATF1 HDLBP ODZ4 LOR ZNF641 SLC1A4 IGF2BP2 MYCN NTN1 TEAD1 FAM46A
M8 (density=3.1e-02; synergy=1.73e-01):
hsa-miR-3692 hsa-miR-4284 hsa-miR-122 hsa-miR-3915 hsa-miR-548s hsa-miR-448
CELF5 FOXM1 SLC4A10 TGIF2 TMEM194B RNF165 ABLIM3 SLC2A12 PIP5K1A RNGTT CMTM4
M9 (density=6.71e-02; synergy=1.77e-01):
hsa-miR-891b hsa-miR-1322
NMNAT2 CBFB ZNF644 ITGA2 KCNJ10 PDS5B
M10 (density=5.79e-02; synergy=1.92e-01):
hsa-miR-586 hsa-miR-3165 hsa-miR-4276
ZNF384 ZNF879 ZFP1 SOAT1 ADAT2
M11 (density=3.31e-02; synergy=2.06e-01):
hsa-miR-208b hsa-miR-216a hsa-miR-4262
AFG3L2 EDAR EN1 DPP6 USP6NL CNKSR3 L1CAM ATP11C CD163 GATA6 HAND2 CELSR3 RYR
M12 (density=5.43e-02; synergy=1.62e-01):
hsa-miR-541 hsa-miR-1229 hsa-miR-33a
VCAN CCL5 EMILIN3 SDC1 PCDH7 EPHA8
M13 (density=2.53e-02; synergy=2.06e-01):
hsa-miR-302a hsa-miR-520b hsa-miR-4293 hsa-miR-302e hsa-miR-106a hsa-miR-31
MYBL1 LHX8 CLP1 GLS ACVR1 BAMBI TRHDE TSEN34 FRZB PRR16 FBXO41 MYCN TRPV6 LE
M14 (density=2.49e-02; synergy=2.24e-01):
hsa-miR-4311 hsa-miR-625 hsa-miR-424 hsa-miR-1193 hsa-miR-4257 hsa-miR-552
WDR43 LRRCC1 SEH1L PPM1L SCN3A FAM60A GK NOL4 GPD1 ABCG8 RELN AKAP6 SIX3 FOX
```

```
M15 (density=5.08e-02; synergy=2.25e-01):
hsa-miR-513b hsa-miR-1234
KIAA1161 C6orf170 GPR126 PAK6 BOLL ABCA13 NUPL1 INSM1 SEMA3B DMD FIGF KIF26A
M16 (density=4.01e-02; synergy=1.72e-01):
hsa-miR-185 hsa-miR-1254 hsa-miR-661
STX1B CADM4 MFRP GEMIN8 GJB1 NFIX TET2 NPAS4 PLEKHG6 OCLN
M17 (density=1.58e-02; synergy=1.92e-01):
hsa-miR-4328 hsa-miR-3148 hsa-miR-208b hsa-miR-216a hsa-miR-4262 hsa-miR-605
AFG3L2 EDAR SARM1 FKBP1A EN1 DPP6 AGPAT9 POLD3 USP6NL CNKSR3 L1CAM ATP11C CD
M18 (density=1.68e-02; synergy=2.09e-01):
hsa-miR-676 hsa-miR-4311 hsa-miR-519e hsa-miR-424 hsa-miR-1193 hsa-miR-608 hs
WDR43 AMOTL1 LRRCC1 SEH1L PPM1L FAM107A FAM60A KCNQ4 GDAP1 SLC2A14 AKAP6 SIX
M19 (density=2.06e-02; synergy=1.95e-01):
hsa-miR-208b hsa-miR-216a hsa-miR-3658 hsa-miR-4262 hsa-miR-601
AFG3L2 EDAR FKBP1A EN1 SNX16 DPP6 KIAA0947 USP6NL CNKSR3 L1CAM ATP11C CD163
```

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

```
[1] "Time elapsed (LASSO+Mirsynergy): 0.491 (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 3.2.2 (2015-08-14)
Platform: x86_64-pc-linux-gnu (64-bit)
Running under: Ubuntu 14.04.3 LTS

locale:
 [1] LC_CTYPE=en_US.UTF-8       LC_NUMERIC=C
 [3] LC_TIME=en_US.UTF-8        LC_COLLATE=C
 [5] LC_MONETARY=en_US.UTF-8    LC_MESSAGES=en_US.UTF-8
 [7] LC_PAPER=en_US.UTF-8       LC_NAME=C
 [9] LC_ADDRESS=C               LC_TELEPHONE=C
[11] LC_MEASUREMENT=en_US.UTF-8 LC_IDENTIFICATION=C

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

```
> plot_module_summary(V)
```

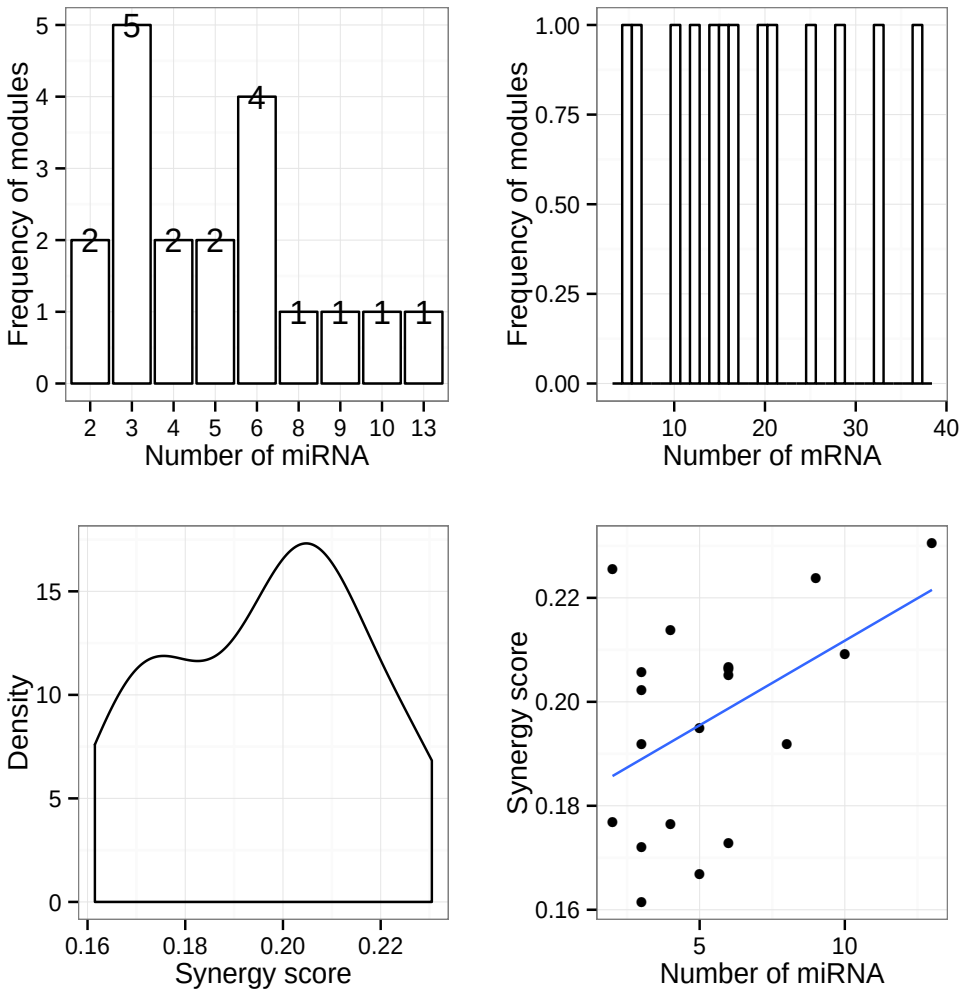

Figure 2: Summary information on MiRM using test data from TCGA-BRCA. Top panels: m/miRNA distribution across modulesas; Bottom panels: the synergy distribution by itself and as a function of the number of miRNA.

```
other attached packages:
[1] glmnet_2.0-2     foreach_1.4.3    Matrix_1.2-2     Mirsynergy_1.6.0
[5] ggplot2_1.0.1    igraph_1.0.1

loaded via a namespace (and not attached):
 [1] Rcpp_0.12.1         knitr_1.11          magrittr_1.5        MASS_7.3-44
 [5] munsell_0.4.2       colorspace_1.2-6    lattice_0.20-33     stringr_1.0.0
 [9] plyr_1.8.3          tools_3.2.2         parallel_3.2.2      grid_3.2.2
[13] gtable_0.1.2        iterators_1.0.8     digest_0.6.8        gridExtra_2.0
[17] RColorBrewer_1.1-2 reshape2_1.4.1      codetools_0.2-14    labeling_0.3
[21] stringi_0.5-5       scales_0.3.0        reshape_0.8.5       proto_0.3-10
```

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.