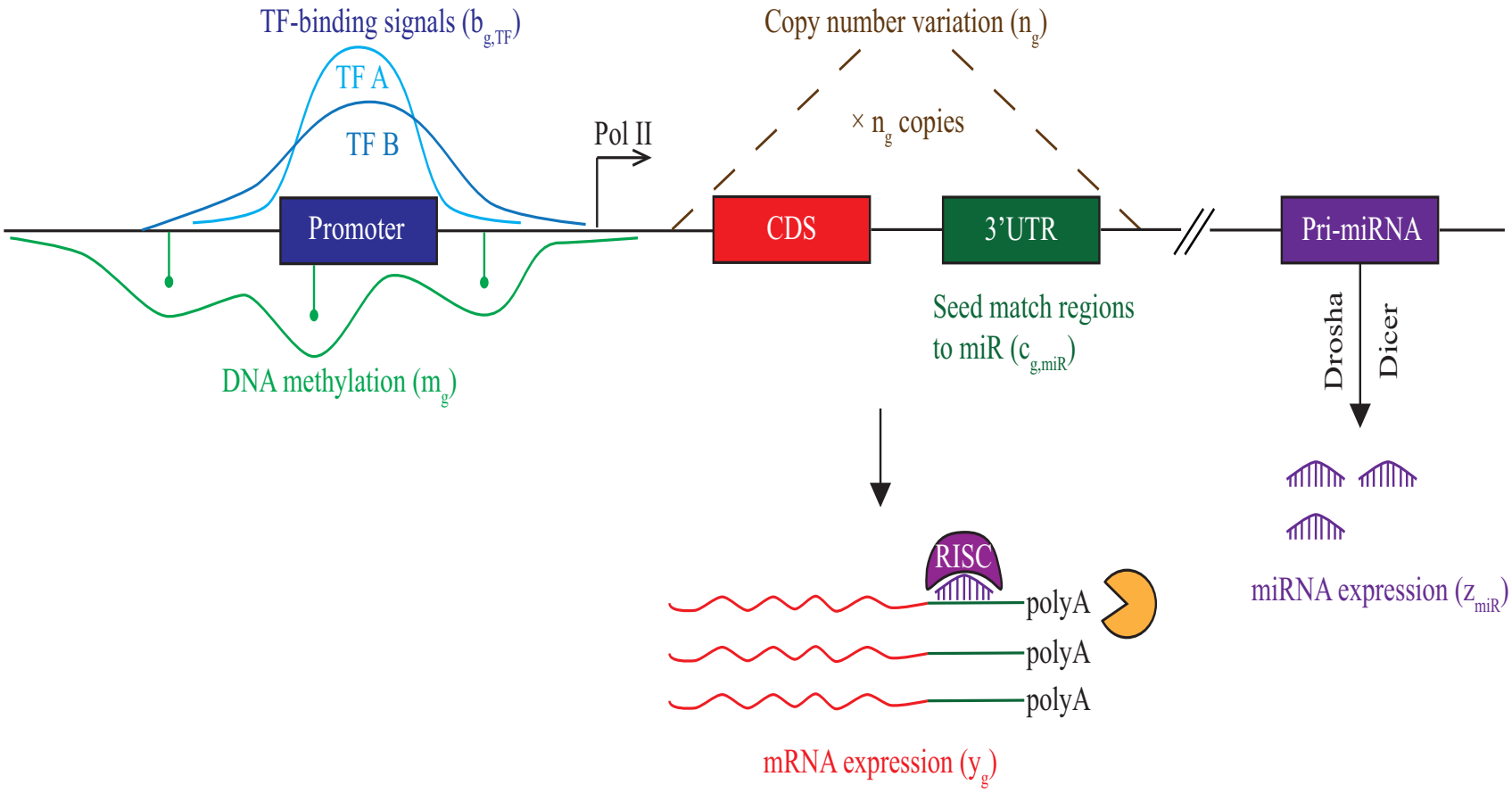
Computational methods to infer context-specific microRNA regulatory networks

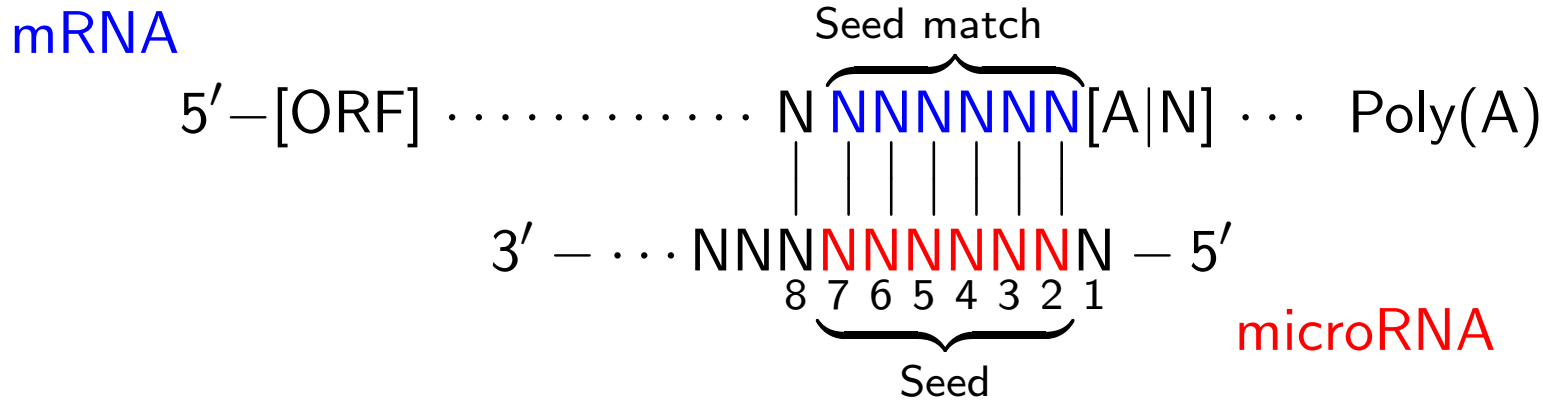
Yue Li
PhD candidate
Computational Biology Group
Department of Computer Science
University of Toronto

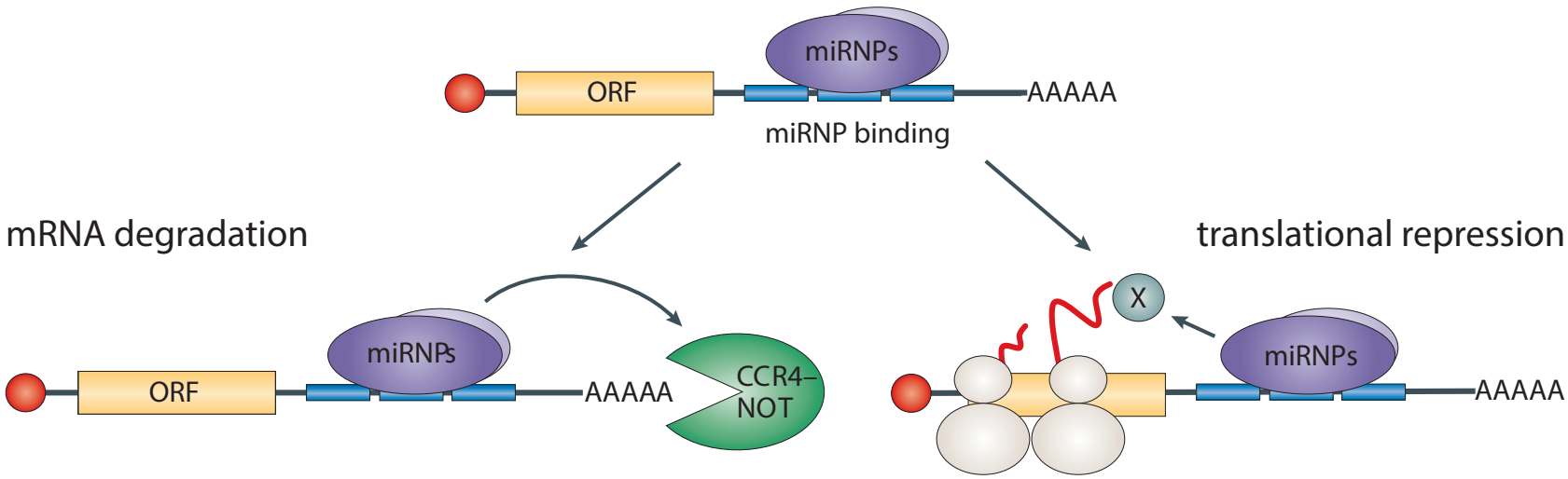
Transcriptional and post-transcriptional regulatory program



MicroRNA

- MicroRNA (miRNA): $\sim$ 22 nucleotide small endogenous RNA;
- Form Watson-Crick basepair with 3′UTR of target mRNA:

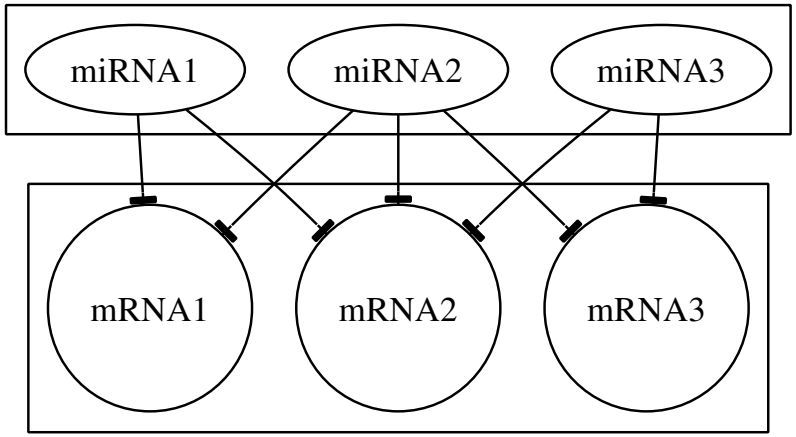




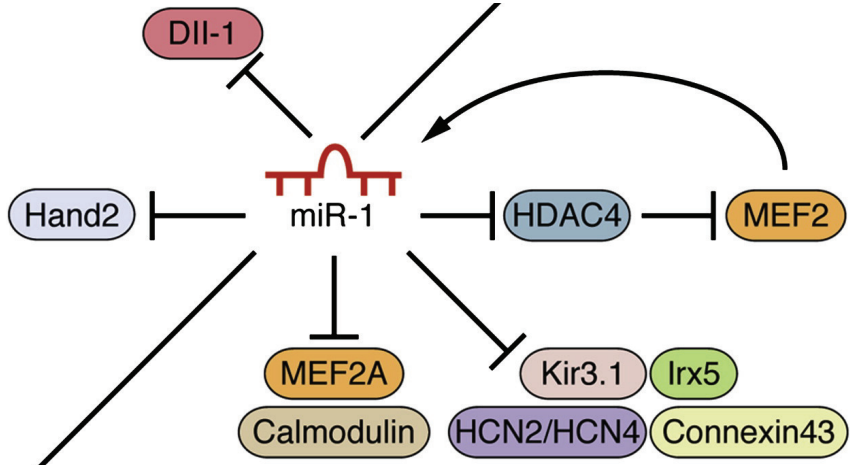
(Filipowicz et al., Nature Reviews, Genetics **9**, 102-114, 2008;
Bartel, Cell, **136**(2), 215-233, 2009)

N-to-*N* relationships between miRNA and mRNA

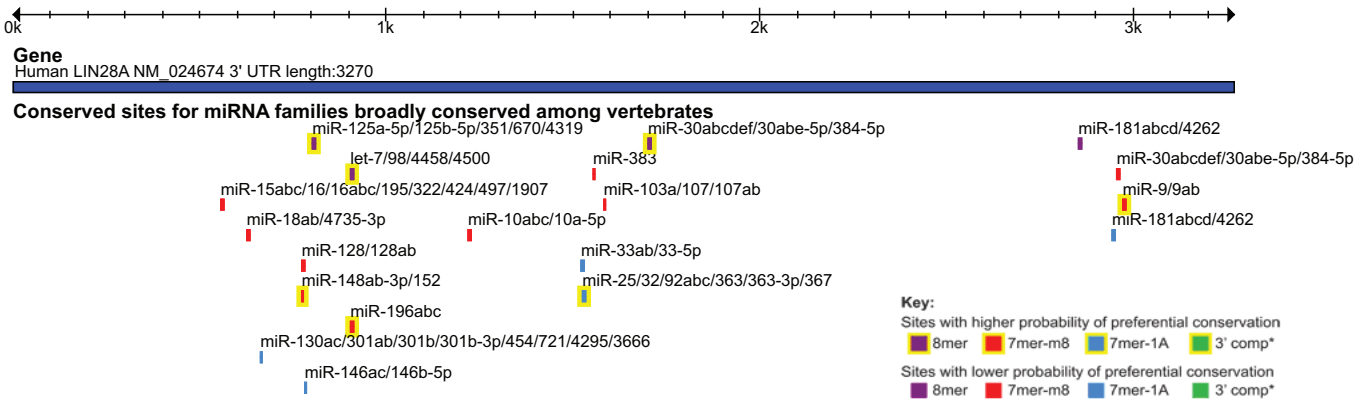
Bipartite graph



miR-1 regulation targets (Liu and Olson, 2010)



Lin28 mRNA (TargetScanHuman 6.2)



MicroRNA target predictions

I. Sequence-based approach:
   - Watson-Crick seed match;
   - Evolutionary conservation;
   - Binding free energy;
   - Lots of false positives: $\sim$ 2:1 signal:noise, $\sim$ 33% specificity;
   - e.g., TargetScan (Lewis et al., 2003), MiRanda (Enright et al., 2004), PicTar (Krek et al., 2005), etc.

II. Expression-based approach:
   - Correlation-based approach:
     - Expression correlation miRNA and mRNA across multiple tissue types;
     - Higher accuracy than sequence-based but require lots of expression profiling data;
     - Not condition/tissue-specific;
     - e.g., GenMiR++ (Huang et al., 2007), HCtarget (Su et al., 2011), GroupMiR (Le et al., 2011)
   - Transfection-based approach (Liu et al., 2010):
     - Expression profiling before and after a miRNA transfection;
     - Promising context-specific targets of the miRNA;
     - Major confounder: off-target effects

High-throughput heterogeneous data

- The dynamic of miRNA expression is implicated in various phenotypic changes including embryonic development and diseases especially in cancers (Spizzo et al. (2009). SnapShot: MicroRNAs in Cancer. *Cell*, 137(3), 586)
- Maturity of high-throughput technologies such microarrays, large/small RNA-seq, ChIP-seq enable large scale profiling across diverse cell types or conditions, for instance:
- TCGA data enable studying cancer-specific miRNA-mRNA interaction with hundreds of expression profiles measured by microarray/RNA-seq for each cancer type
- ENCODE data enable studying the interplay between TF and miRNA regulation using ChIP-seq and RNA-seq
- **Challenges/opportunities**: High-throughput heterogeneous data enable predicting **context-specific functional miRNA regulatory networks** but present unique computational challenges.

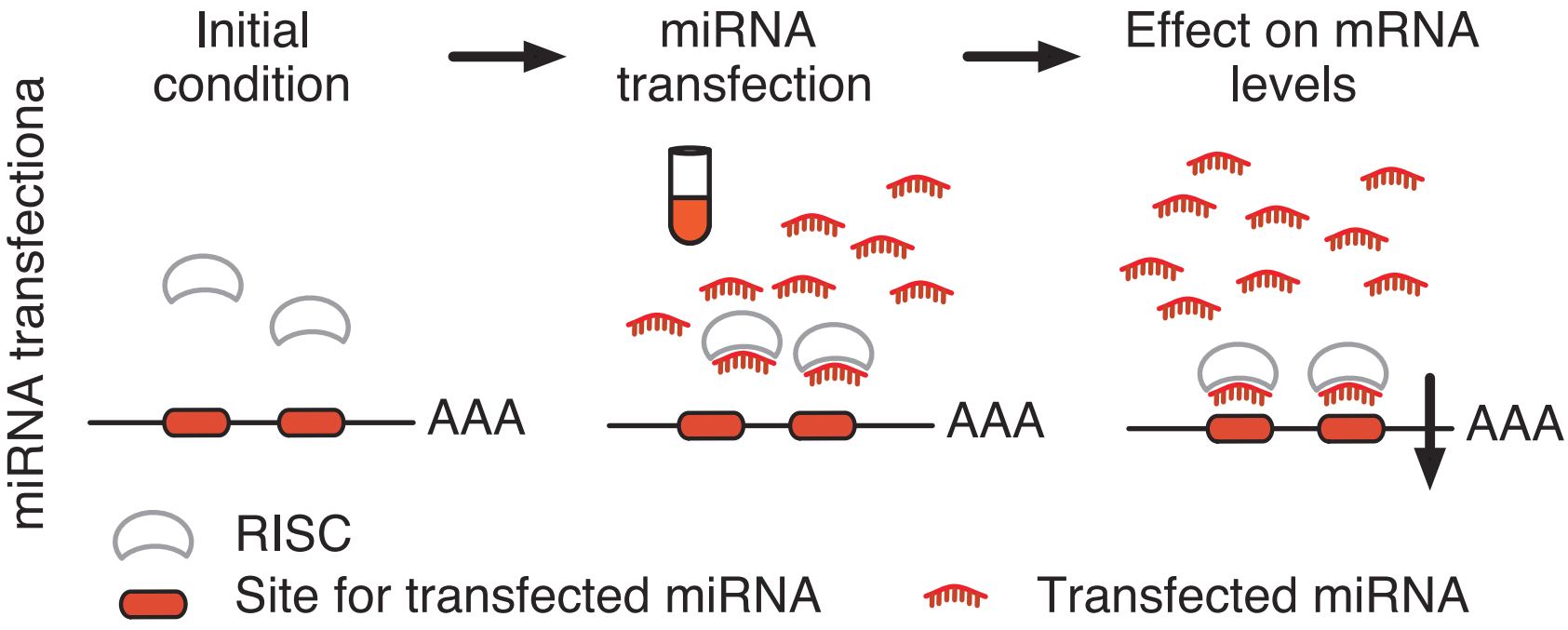
Outline

- Background
- Project 1: A probabilistic approach to explore human miRNA targetome by integrating miRNA-overexpression data and sequence information
- Project 2: Inferring probabilistic miRNA-mRNA interaction signatures in cancers: a role-switch approach
- Project 3: Mirsynergy: detecting miRNA synergistic regulatory network modules by overlapping neighbourhood expansion
- Project 4: Regression analysis of combined gene expression regulation in acute myeloid leukemia

Project 1: A probabilistic approach to explore human miRNA targetome by integrating miRNA-overexpression data and sequence information

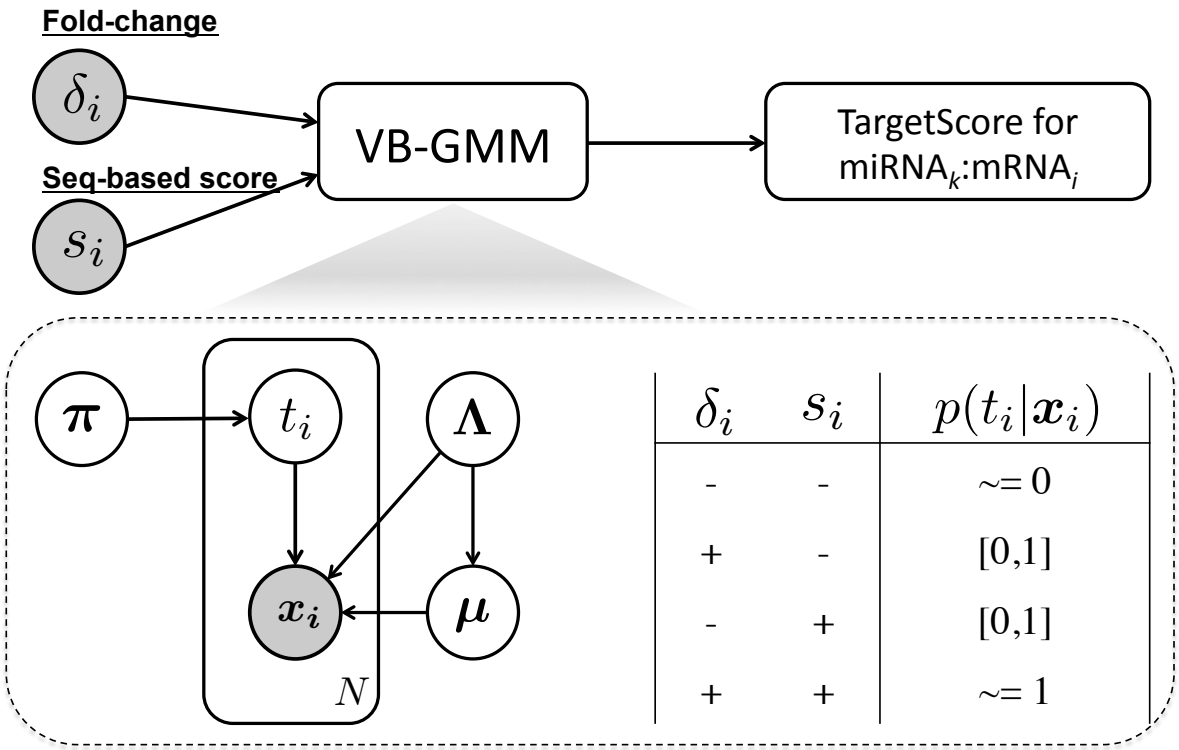
Publication:
Li, Y., Goldenberg, A., Wong, KC., Zhang Z. A probabilistic approach to explore human miRNA targetome by integrating miRNA-overexpression data and sequence information. *Bioinformatics*. (Oxford, England), **30**(5), 621–628. (May 2013)

MicroRNA (miRNA) overexpression



Challenge: Off-target effects

TargetScore model



δ_i	s_i	$p(t_i\|\boldsymbol{x}_i)$
-	-	$\sim= 0$
+	-	[0,1]
-	+	[0,1]
+	+	$\sim= 1$

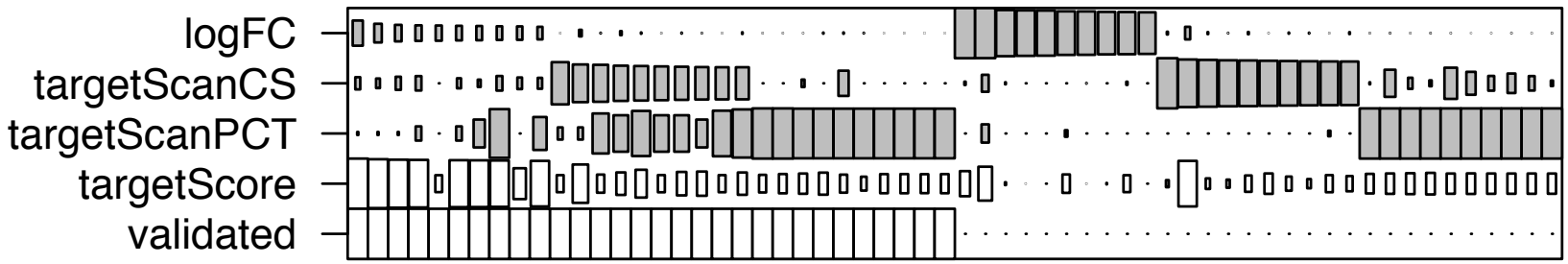
$$\text{TargetScore} = \sigma(-\log FC)\left(\frac{1}{L+1}\sum_{\mathbf{x}_l \in \{\mathbf{x}_f, \mathbf{x}_1, \ldots, \mathbf{x}_L\}} p(t|\mathbf{x}_l)\right)$$

where $\sigma(-\log FC) = \frac{1}{1+\exp(\log FC)}$, $p(t|\mathbf{x}_l)$ is the posterior of true target given the observed score $\mathbf{x}_l$, inferred by variational Bayesian Gaussian mixture model (Ch10, PR & ML, Bishop, 2006).

MicroRNA transfection data collected from GEO

Category	Count
GEO Series	84
Platform	6
Cell/Tissue	77
MicroRNA	113

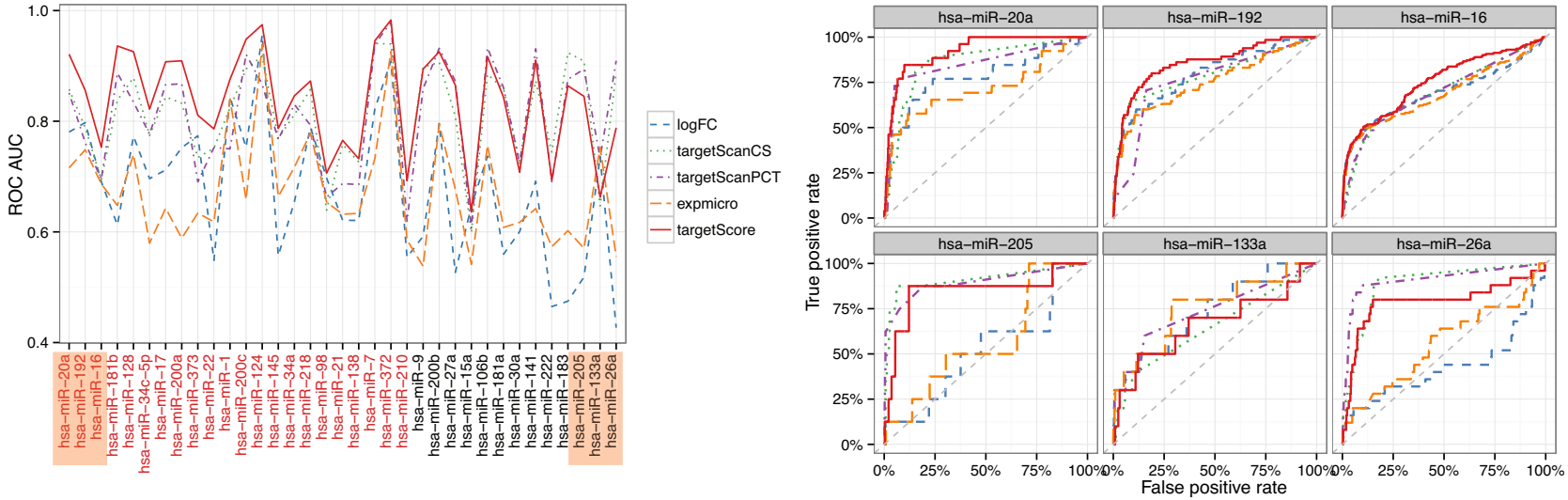
Hinton plot of target feature scores



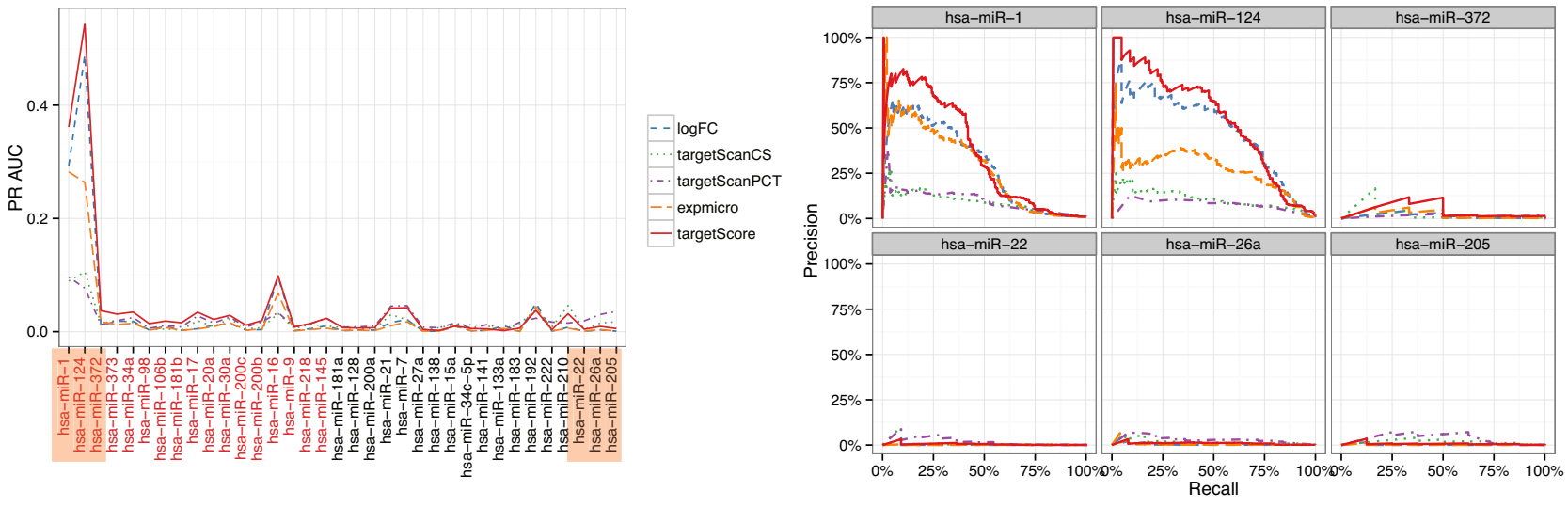
$$\text{TargetScore} = \sigma(-\log FC)\left(\frac{1}{L+1}\sum_{\mathbf{x}_l \in \{\mathbf{x}_f, \mathbf{x}_1, \ldots, \mathbf{x}_L\}} p(t|\mathbf{x}_l)\right)$$

Evaluation of target prediction methods on 35 miRNAs

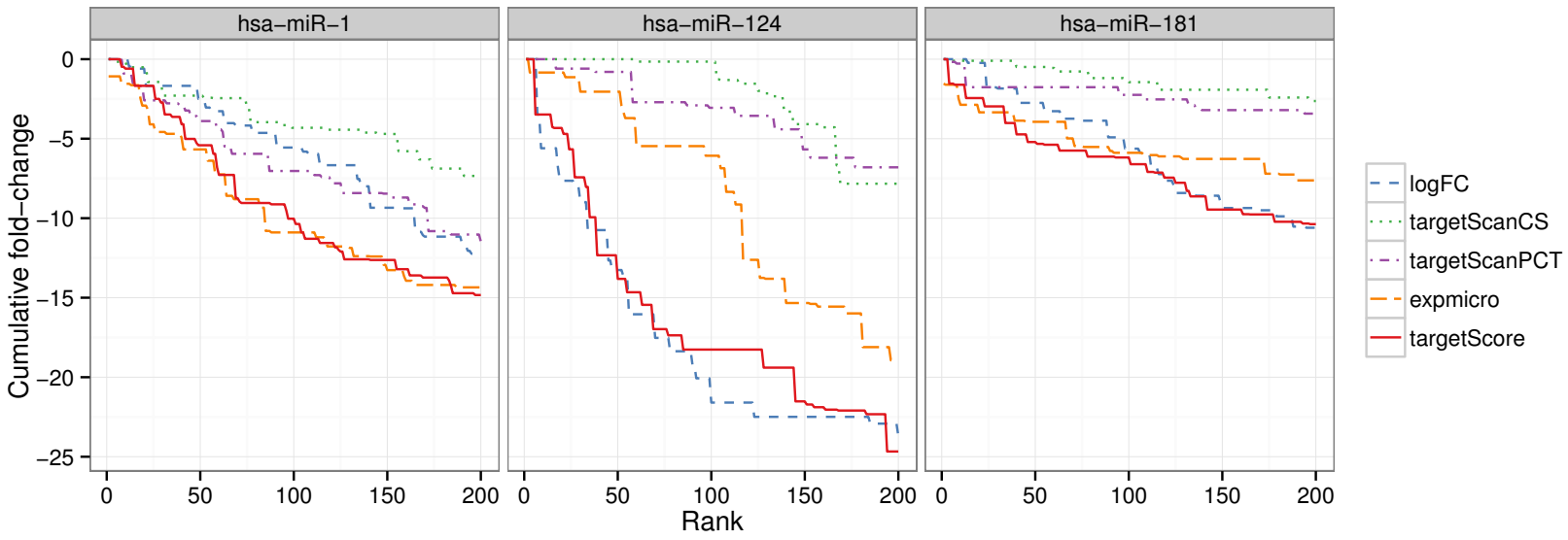
A. ROC



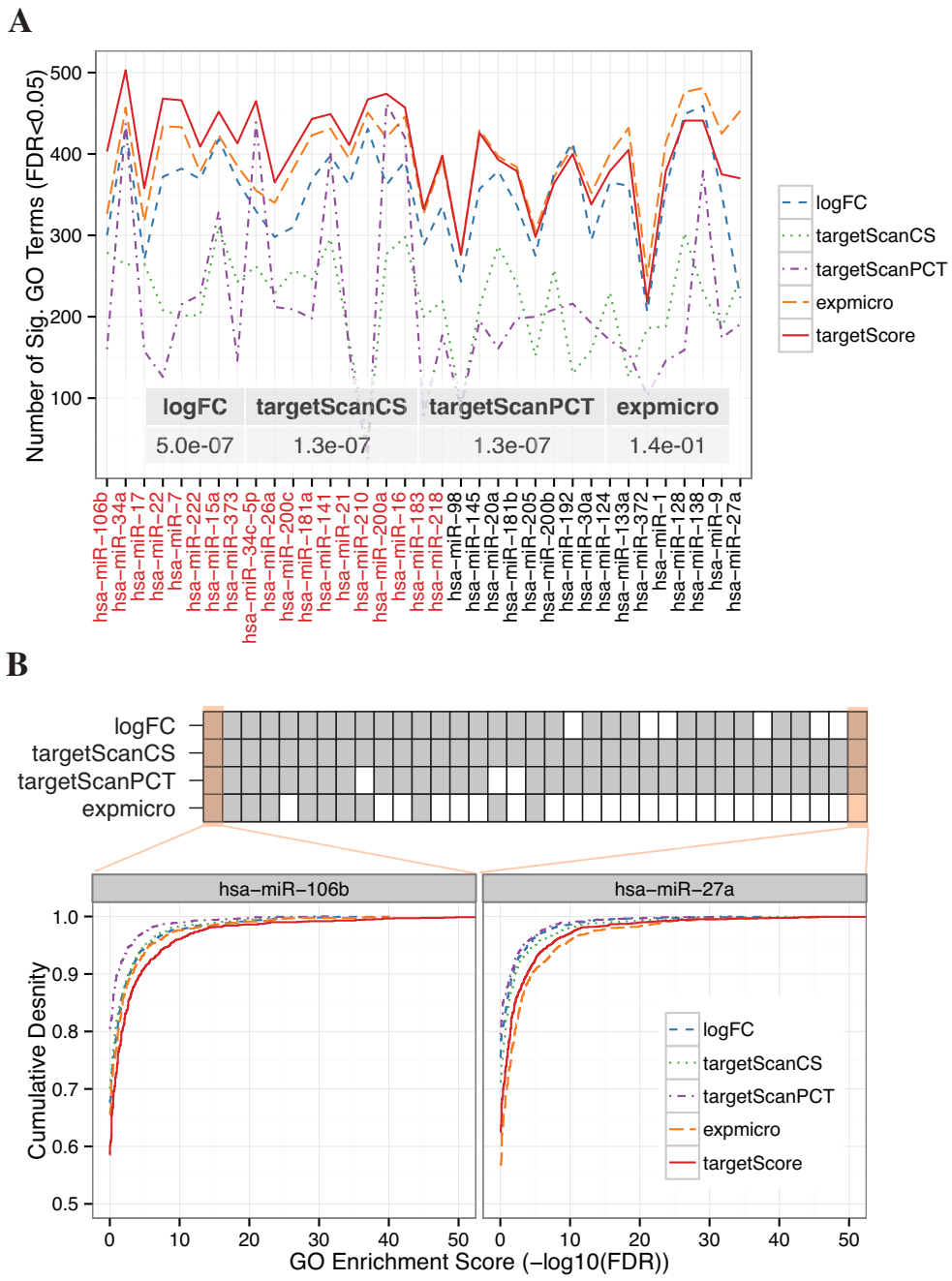
B. Precision-recall



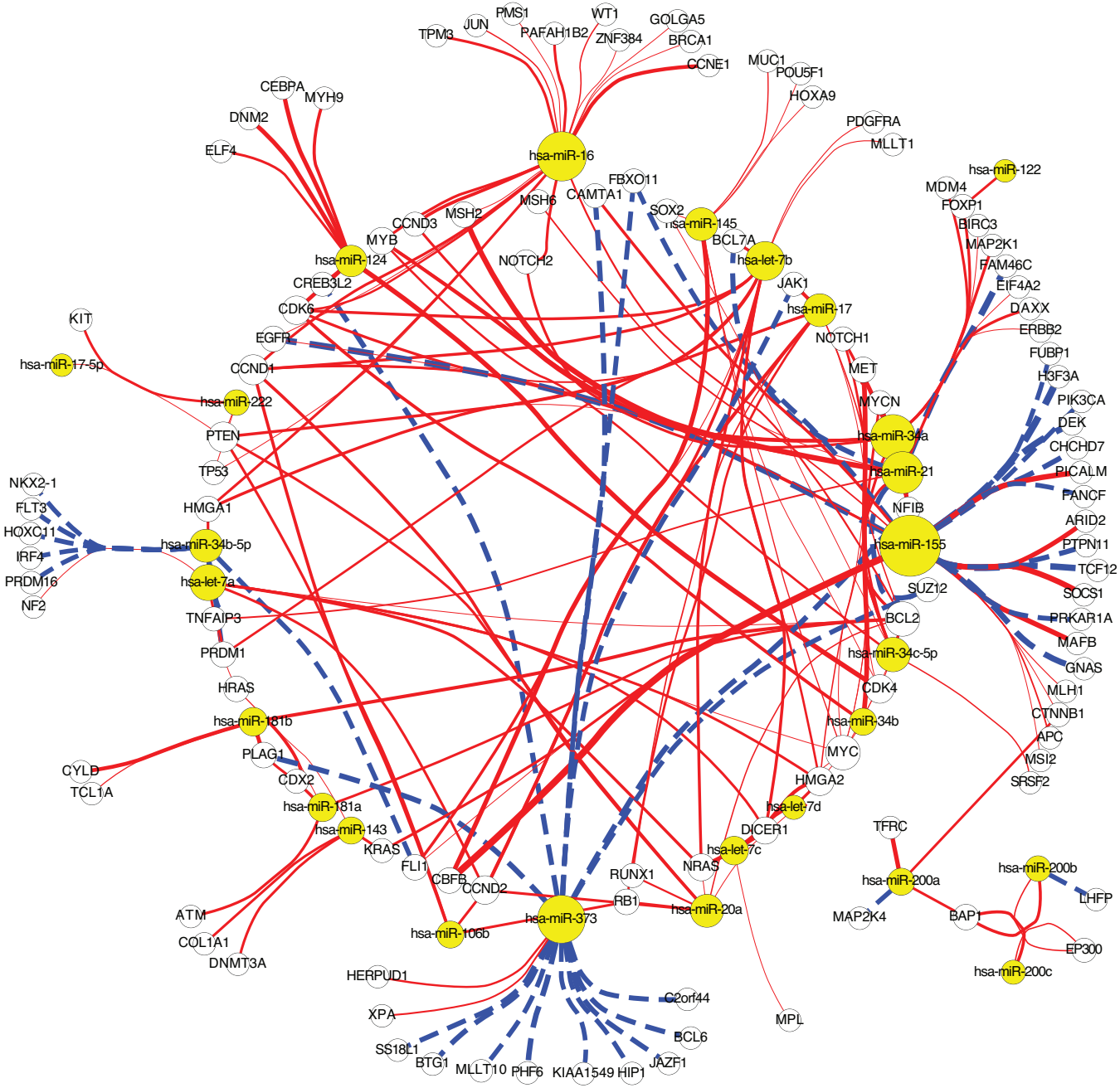
Cumulative sum of protein down-regulation



Comparison of GO enrichments



Oncomir-oncogene network



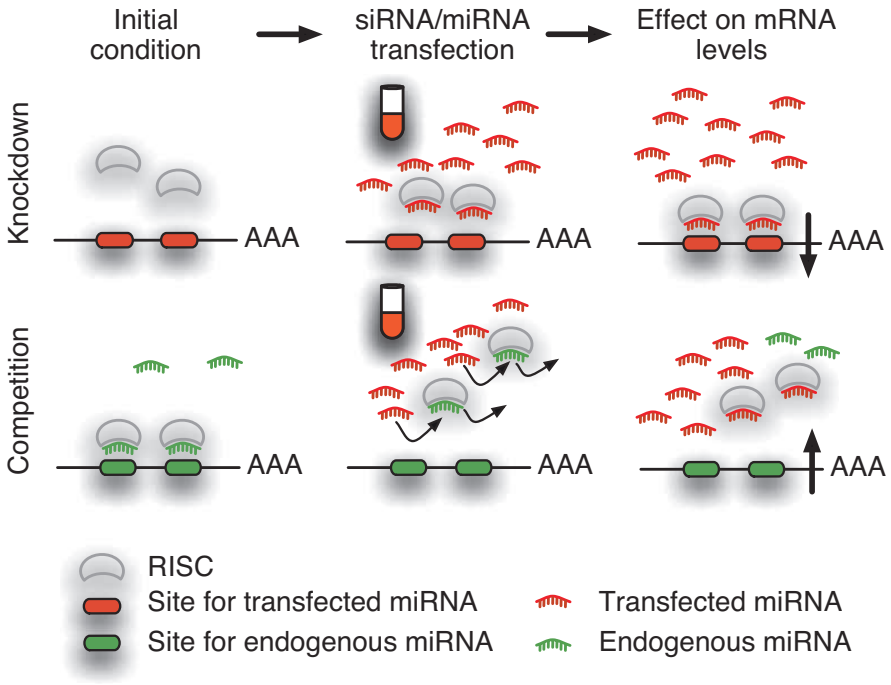
Project 1 summary

- We developed a Bayesian probabilistic scoring strategy called TargetScore using miRNA overexpression and target site information to predict context-specific miRNA-mRNA interactions.
- TargetScore demonstrates superior performance over the existing methods in predicting validated miRNA targets, protein down-regulation, and functional enrichment in majority of the test data.
- TargetScore is implemented as a stand-alone R package and freely available at Bioconductor website.

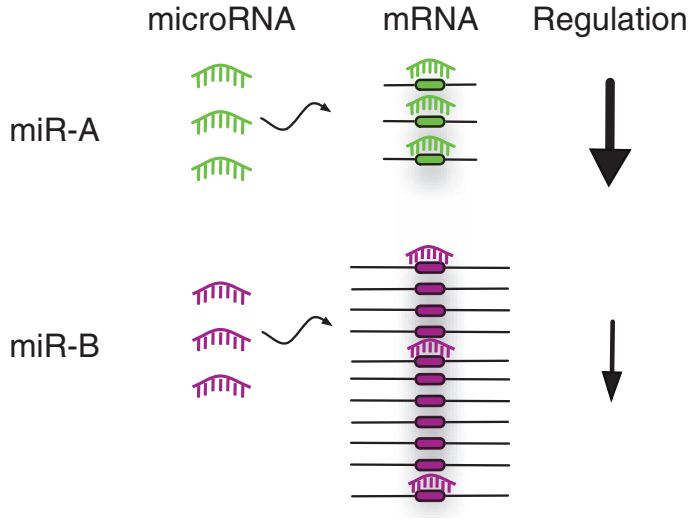
Project 2: Inferring probabilistic miRNA-mRNA interaction signatures in cancers: a role-switch approach

Publication:
Li, Y., Liang, C., Wong, KC, Jin, K., and Zhang, Z. Inferring probabilistic miRNA-mRNA interaction signatures in cancers: a role-switch approach. *Nucleic Acids Research*, **42**(9), e76. (Feb, 2014)

A. miRNAs competes for RISC (Khan, *et al*., 2009)



B. mRNA abundance "dilute" individual miRNA effects on each mRNA (Arvey, *et al*., 2010)

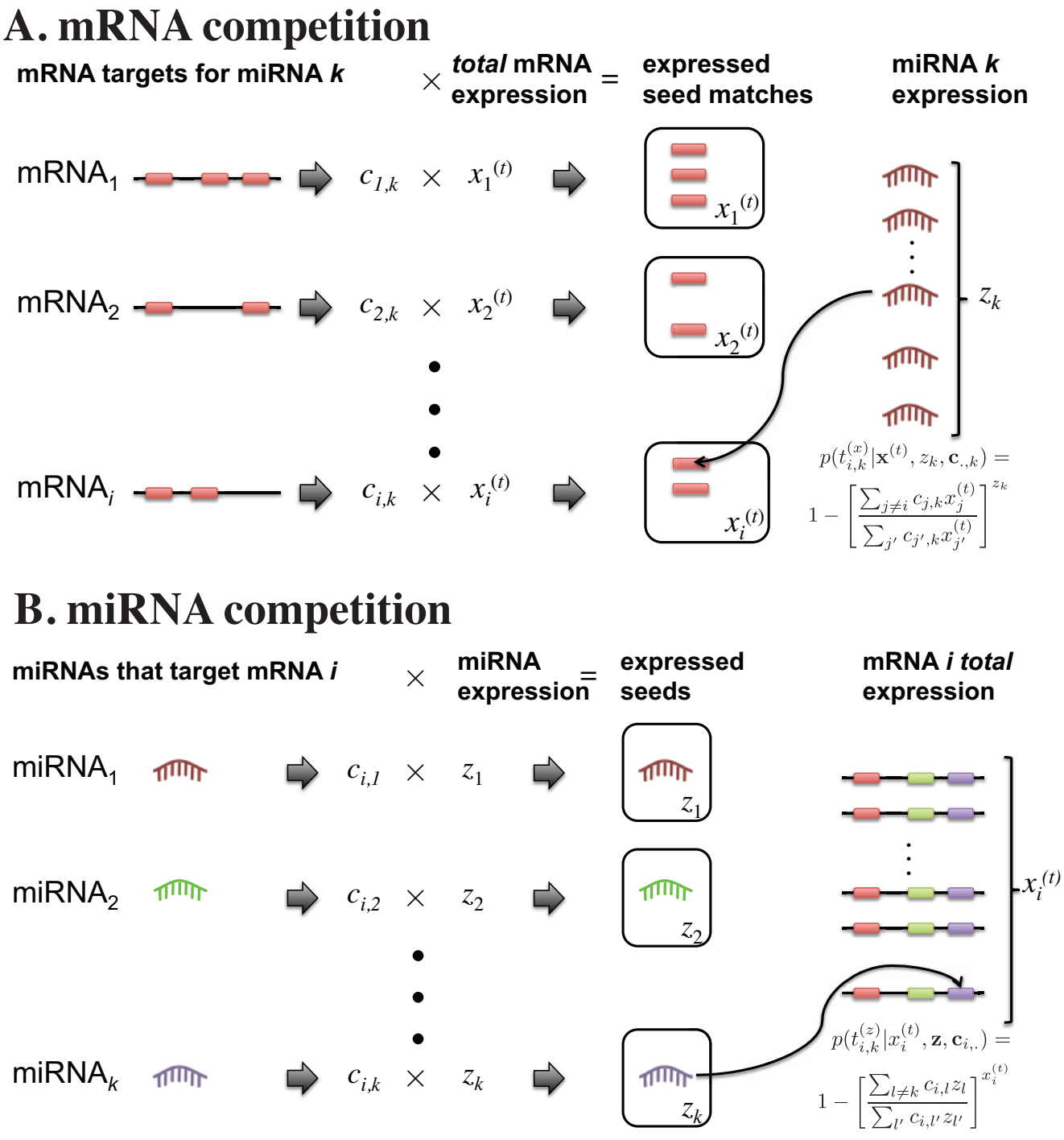


Motivation:

- Correlation/regression-based methods require a large number of samples; and the identified miRNA-mRNA interactions (MMI) are not sample-specific
- mRNA competition is rarely considered but important; e.g., dilution effects (Arvey et al., 2010)

I/O of the proposed model:

- **Input**: paired mRNA-miRNA expression profiles of N mRNA and M miRNA from *a single sample*, $N \times M$ seed-match matrix
- **Output**: inferred $N \times M$ probabilistic miRNA-mRNA interaction signatures (ProMISe)


A. mRNA competition
mRNA targets for miRNA k
total mRNA expression
expressed seed matches
miRNA k expression
mRNA1
mRNA2
mRNAi
c1,k × x1(t)
c2,k × x2(t)
ci,k × xi(t)
x1(t)
x2(t)
xi(t)
zk
p(t(x)i,k|x(t), zk, c.,k) =
1 − [∑j≠i cj,k xj(t) / ∑j' cj',k xj'(t)]^zk
B. miRNA competition
miRNAs that target mRNA i
miRNA expression
expressed seeds
mRNA i total expression
miRNA1
miRNA2
miRNAk
ci,1 × z1
ci,2 × z2
ci,k × zk
z1
z2
zk
xi(t)
p(t(z)i,k|xi(t), z, ci,.) =
1 − [∑l≠k ci,l zl / ∑l' ci,l' zl']^xi(t)

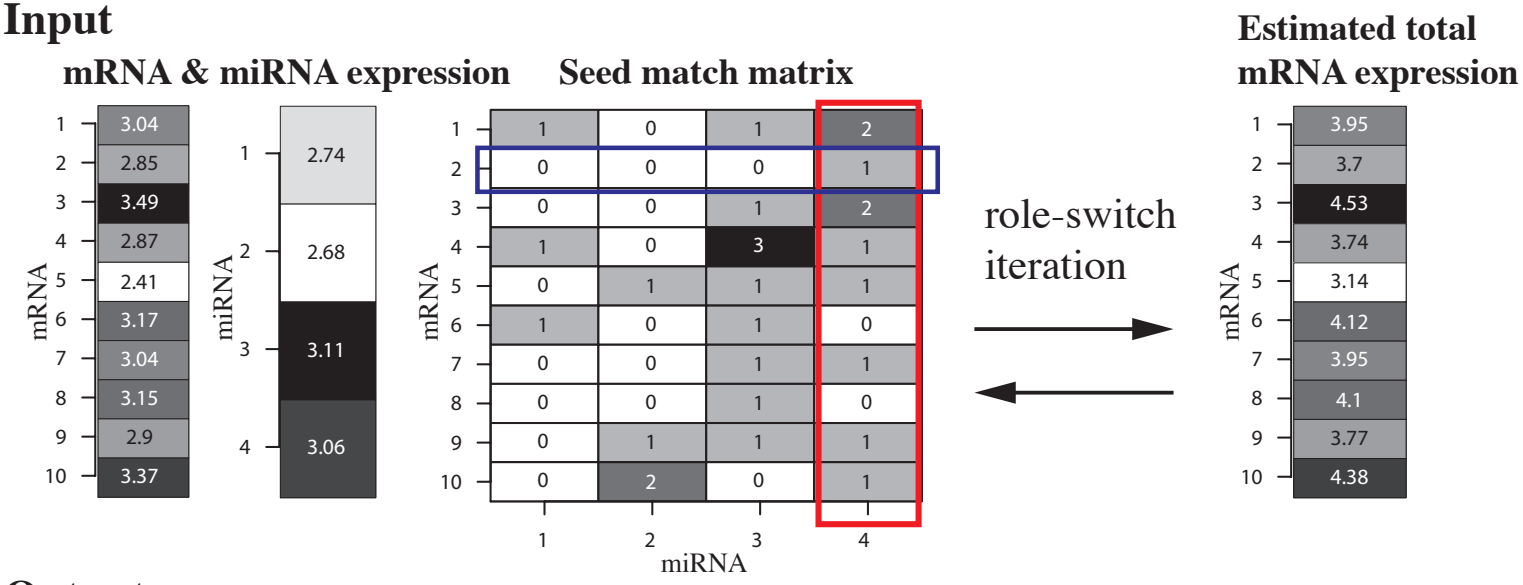
ProMISe model details:

Initialize $\mathbf{x}^{(t)} = \mathbf{x}^{(o)}$

1. $p(t_{i,k}^{(x)}|\mathbf{x}^{(t)}, z_k, \mathbf{c}_{.,k}) = 1 - \left[\frac{\sum_{j\neq i} c_{j,k} x_j^{(t)}}{\sum_{j'} c_{j',k} x_{j'}^{(t)}}\right]^{z_k}$
2. $p(t_{i,k}^{(z)}|x_i^{(t)}, \mathbf{z}, \mathbf{c}_{i,.}) = 1 - \left[\frac{\sum_{l\neq k} c_{i,l} z_l}{\sum_{l'} c_{i,l'} z_{l'}}\right]^{x_i^{(t)}}$
3. $p(t_{i,k}^{(j)}|\mathbf{x}^{(t)}, \mathbf{z}, \mathbf{C}) = p(t_{i,k}^{(x)}|\mathbf{x}^{(t)}, z_k, \mathbf{c}_{.,k}) p(t_{i,k}^{(z)}|x_i^{(t)}, \mathbf{z}, \mathbf{c}_{i,.})$
4. $\Delta x_{i,k} = \eta p(t_{i,k}|\mathbf{x}^{(t)}, z_k, \mathbf{c}_{.,k}) x_i^{(t)}$
5. $x_i^{(t)*} = x_i^{(t)} + \sum_k \Delta x_{i,k}; \quad x_i^{(t)} = \frac{x_i^{(t)*}}{\sum_i x_i^{(t)*}} T$
6. Repeat 1-5 until
   $\max\left[\left|p(t_{i,k}^{(x)}|\mathbf{x}^{(t)}, z_k, \mathbf{c}_{.,k})^t - p(t_{i,k}^{(x)}|\mathbf{x}^{(t)}, z_k, \mathbf{c}_{.,k})^{t-1}\right|\right] < tol,$
   $\max\left[\left|p(t_{i,k}^{(z)}|x_i^{(t)}, \mathbf{z}, \mathbf{c}_{i,.})^t - p(t_{i,k}^{(z)}|x_i^{(t)}, \mathbf{z}, \mathbf{c}_{i,.})^{t-1}\right|\right] < tol$

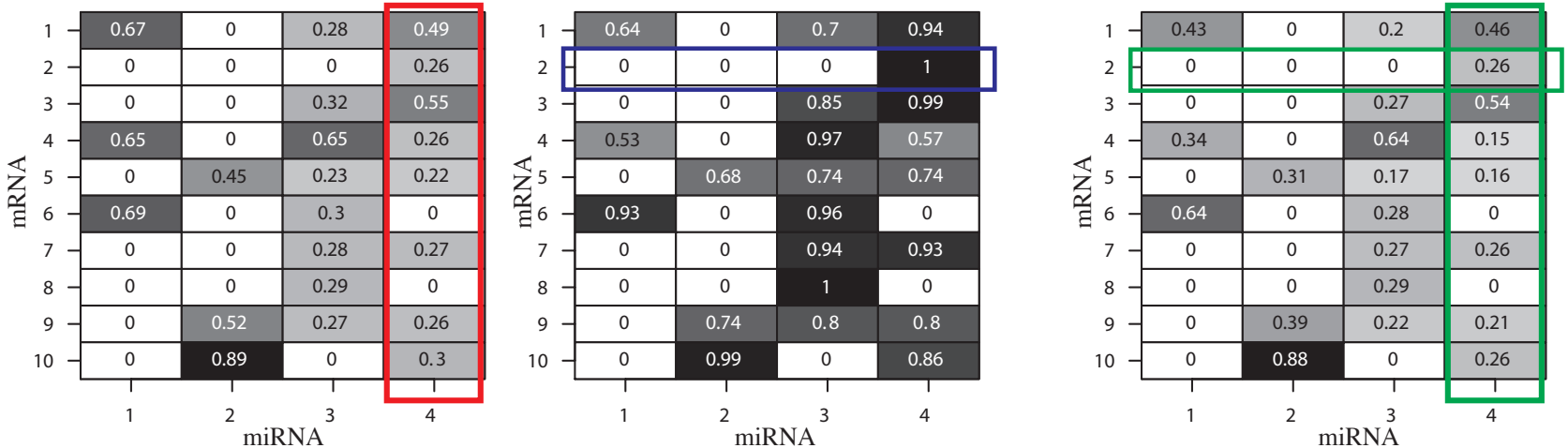
Toy example:

Input

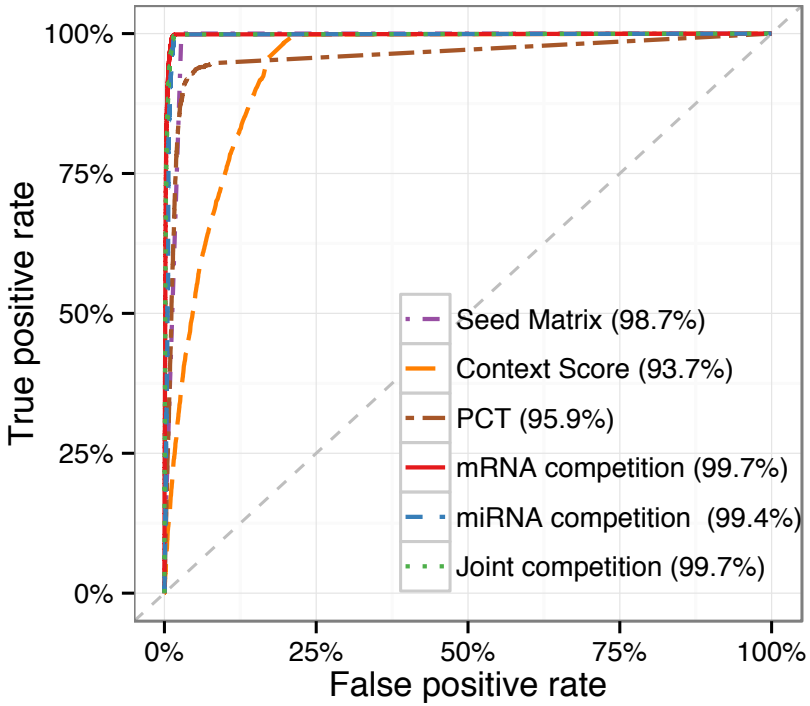


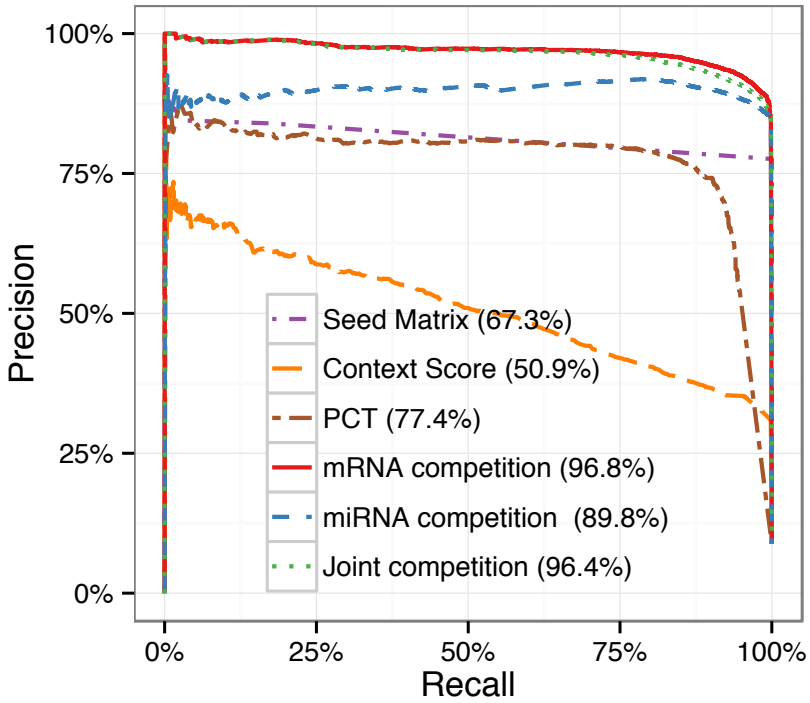
Output

$$p(t_{i,k}^{(x)}|\mathbf{x}^{(t)}, z_k, \mathbf{c}_{.,k}) \quad * \quad p(t_{i,k}^{(z)}|x_i^{(t)}, \mathbf{z}, \mathbf{c}_{i,.}) \quad \xrightarrow{\text{Elementwise Product}} \quad p(t_{i,k}^{(j)}|x_i^{(t)}, \mathbf{z}, \mathbf{c}_{i,.})$$



Identification of confidence targets from PAR-CLIP and miRNA knockdown data in HEK293 (Hafner et al., 2010)

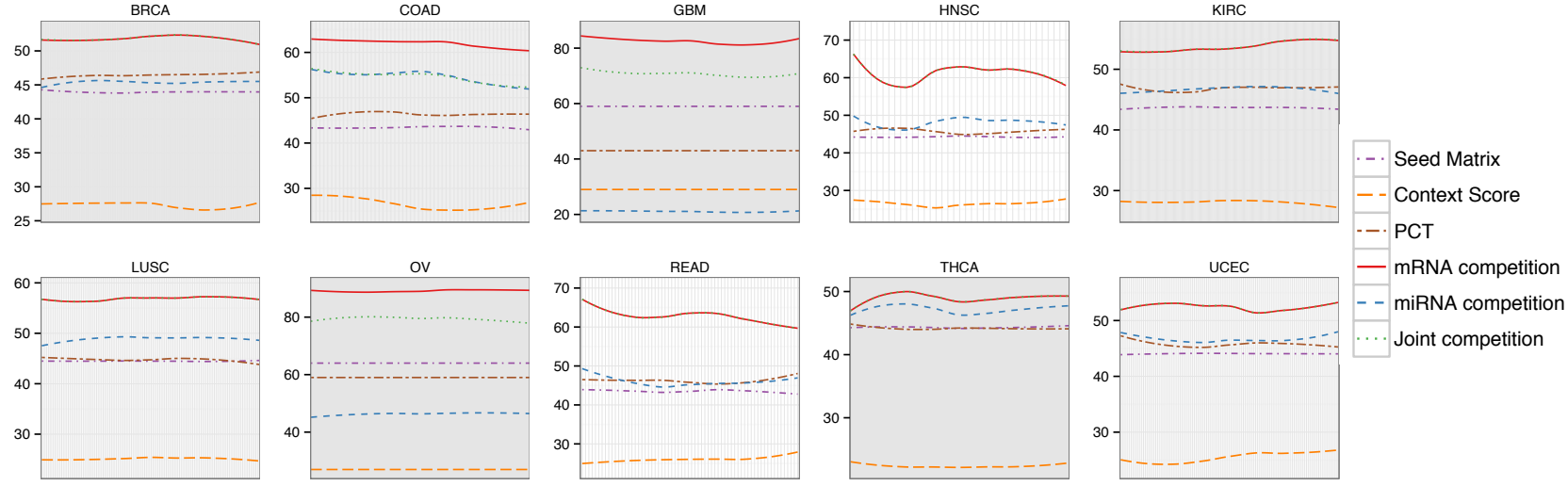


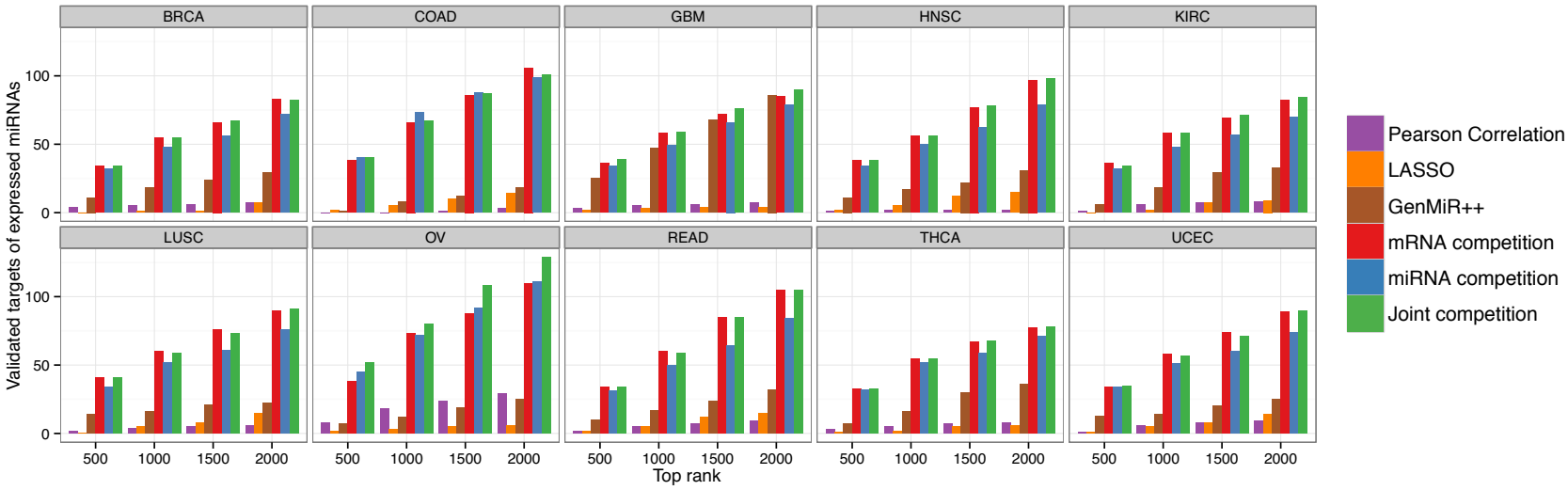


Paired expression data from TCGA

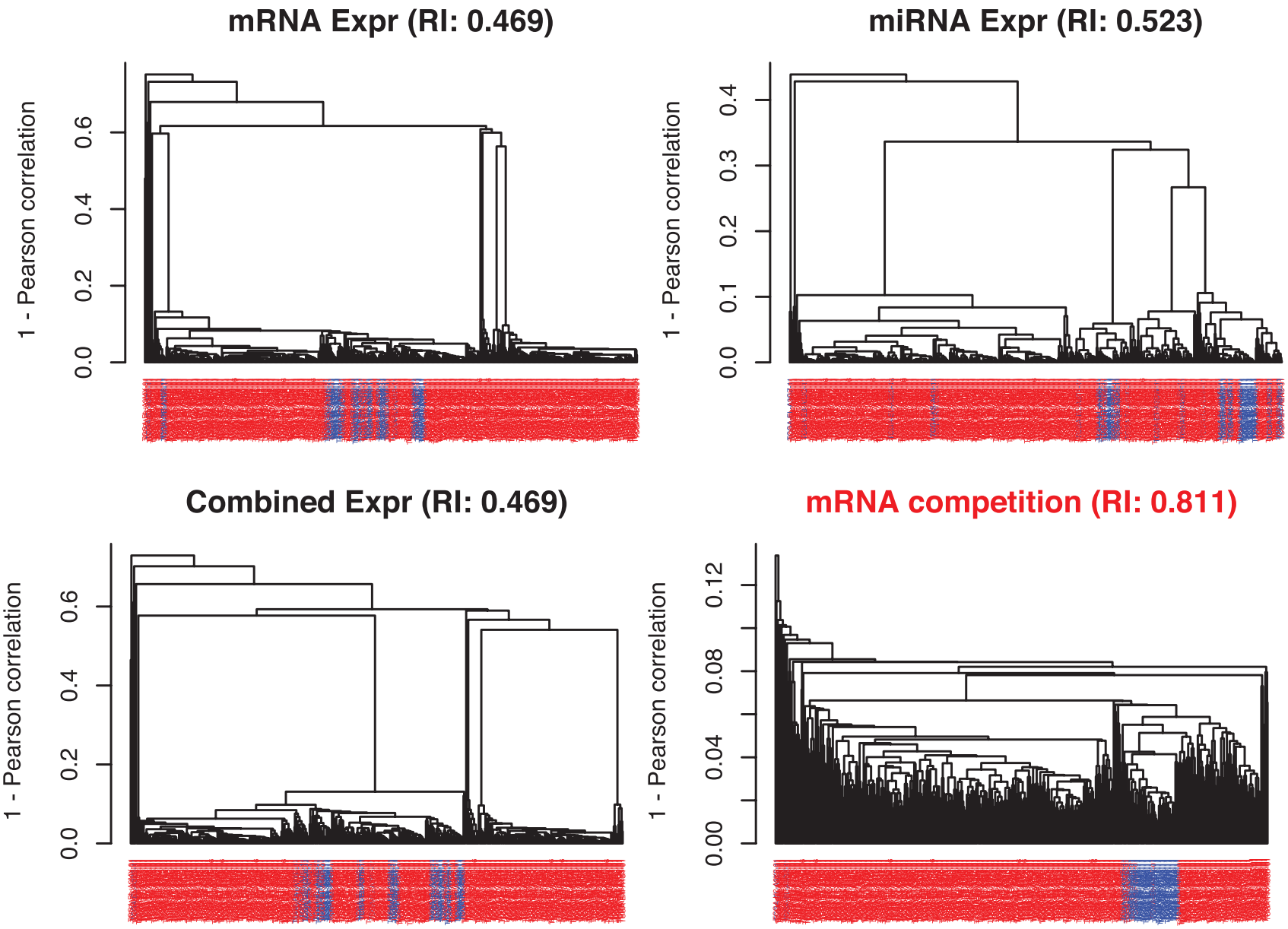
Cancer	mRNA	miRNA	normal	tumor	Total
BRCA	13306	710	14	317	331
COAD	13306	710	0	177	177
GBM	10344	338	10	496	506
HNSC	13306	710	0	37	37
KIRC	13306	710	0	274	274
LUSC	13306	710	0	132	132
OV	8371	542	8	565	573
READ	13306	710	0	66	66
THCA	13306	710	58	485	543
UCEC	13306	710	1	124	125

Identification of validated targets from TCGA data

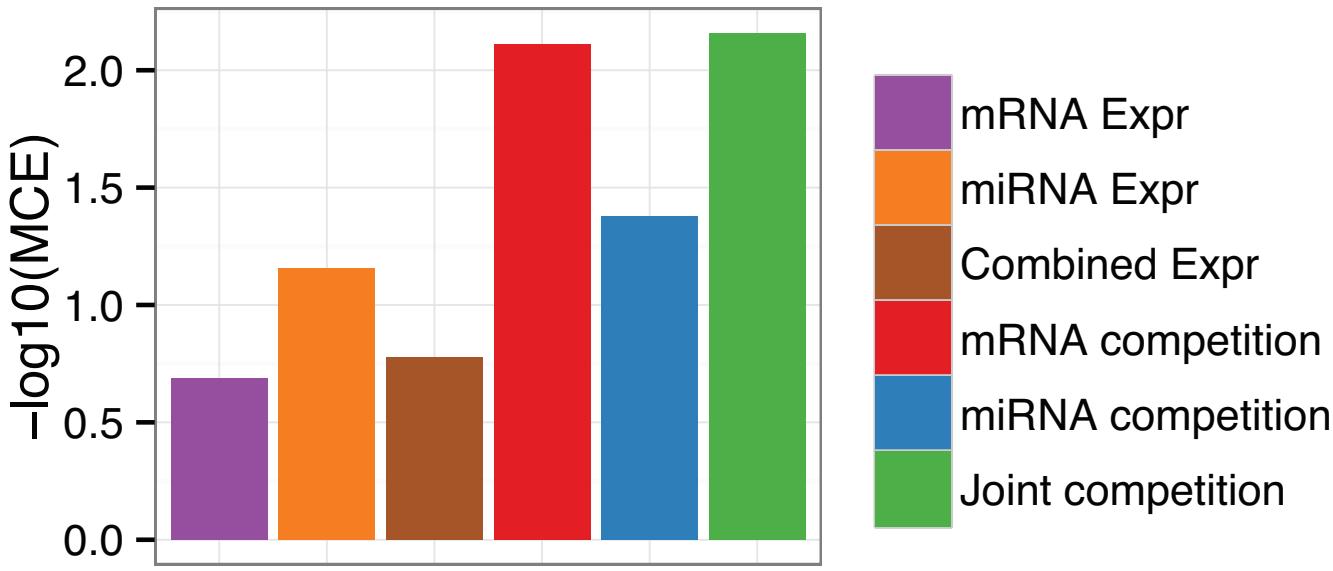




Thyroid cancer tumors exhibit distinct ProMISe signatures (unsupervised hierarchical clustering)

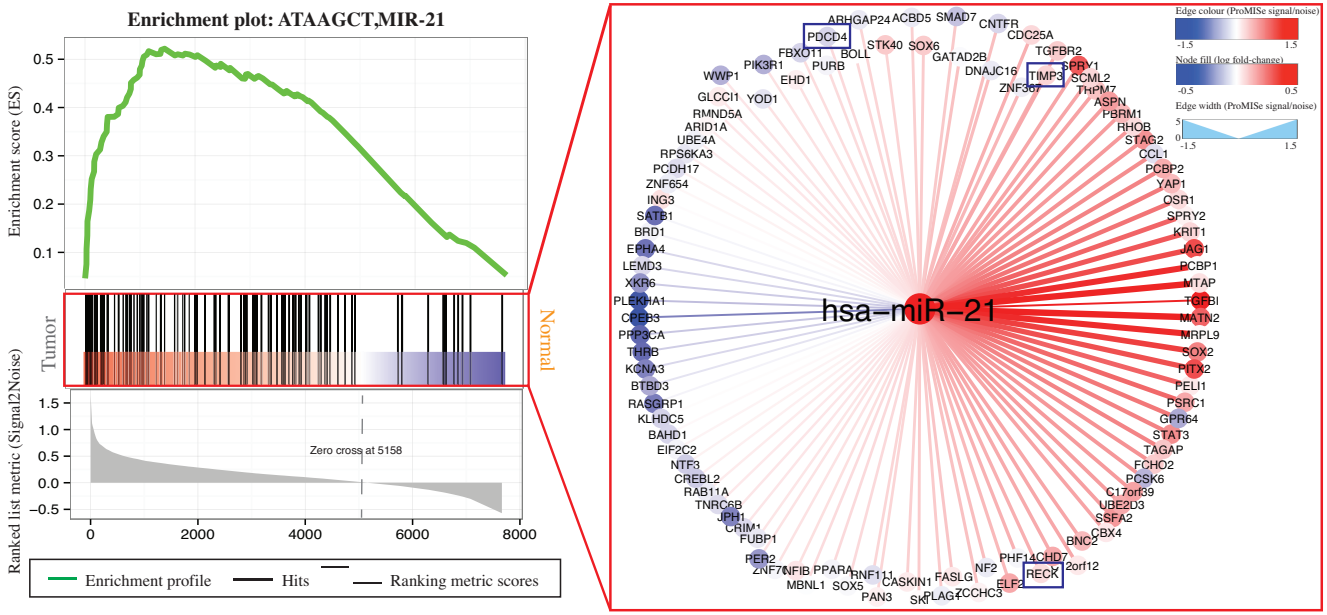


Thyroid cancer tumors exhibit distinct ProMISe signatures (supervised logistic regression with LOOCV)

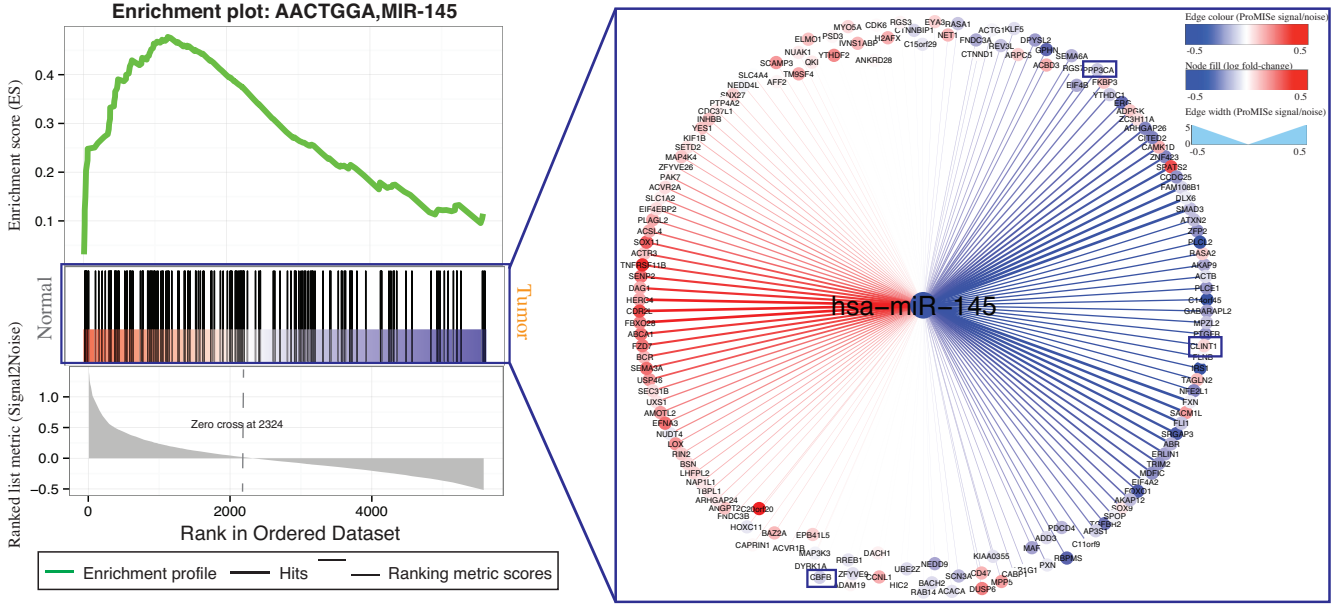


Gene set enrichment analysis (Subramanian et al., 2005)

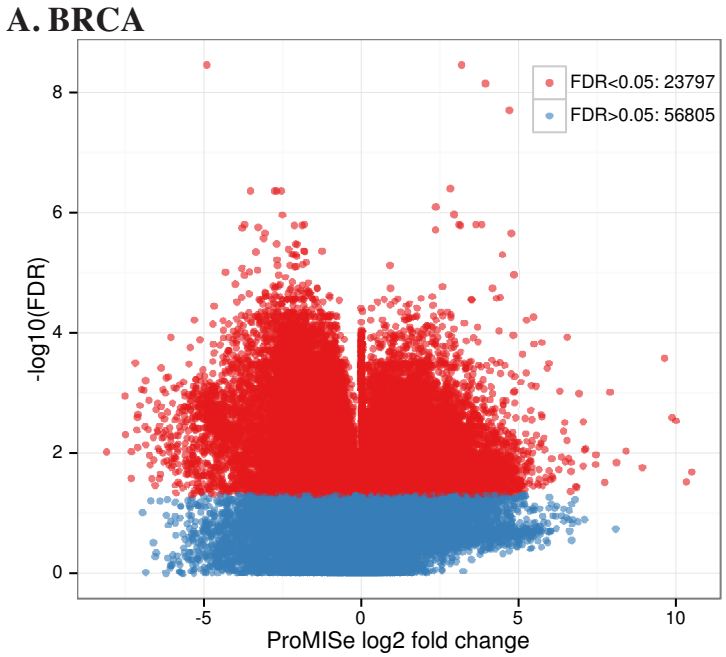
A. miR-21 exhibits higher activity in glioblastoma

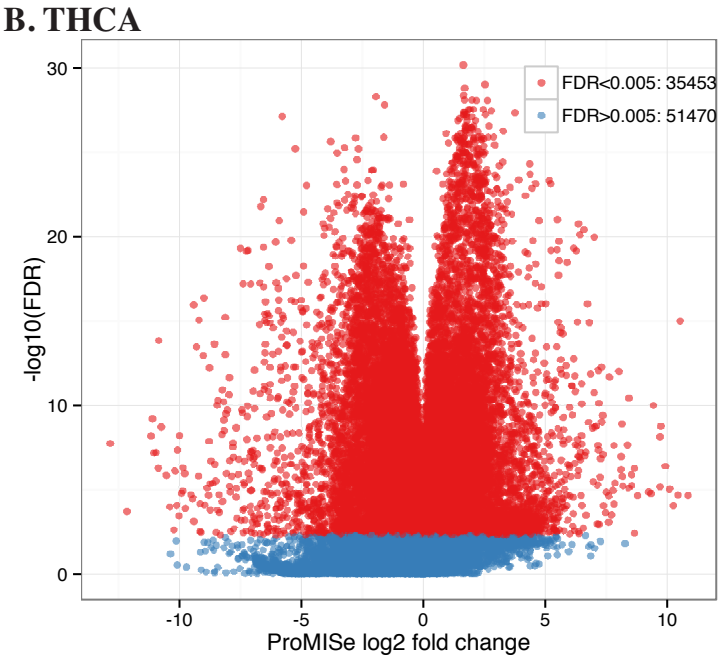


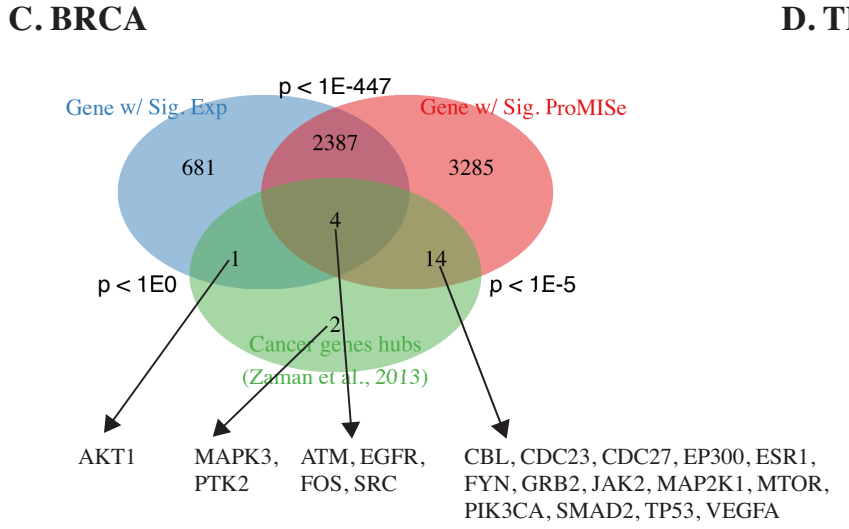
B. miR-145 exhibits lower activity in ovarian cancer

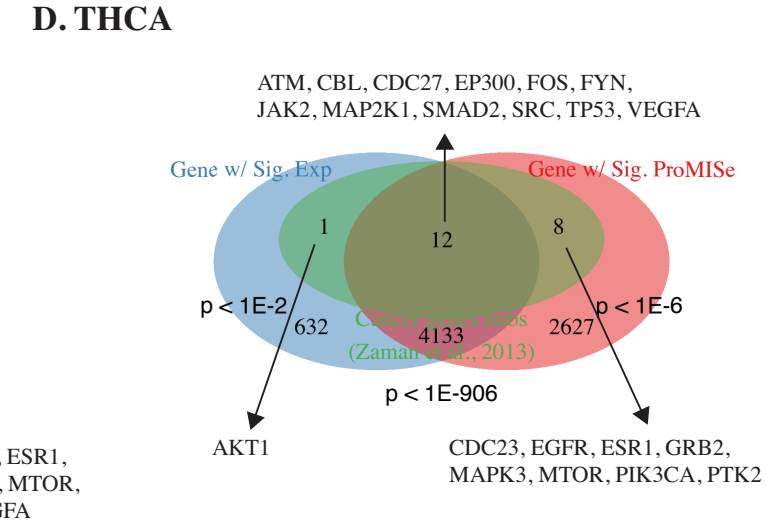


Paired test on 14 or 58 matched tumor/normal in breast cancer (BRCA) or thyroid cancer (THCA)

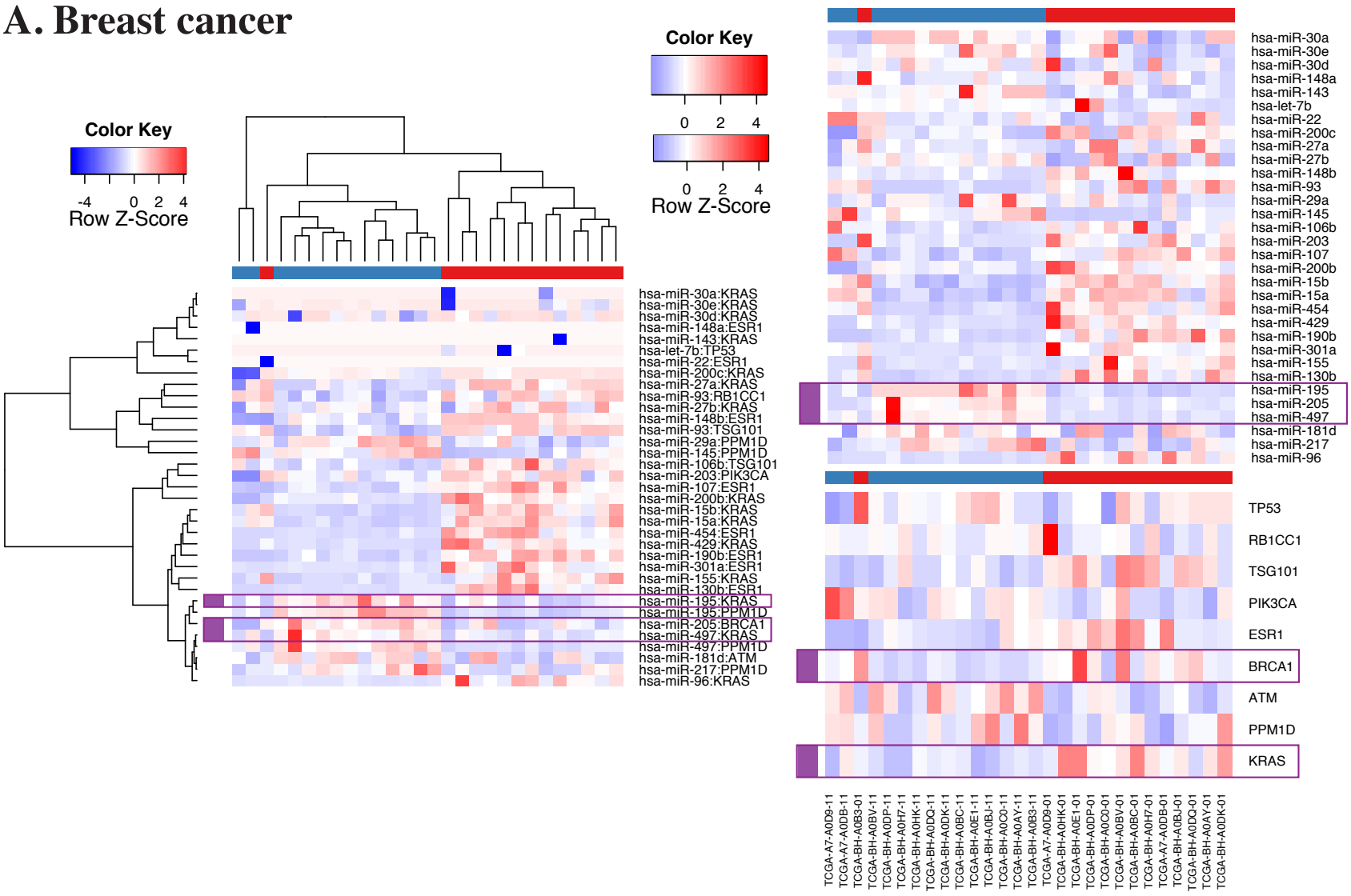




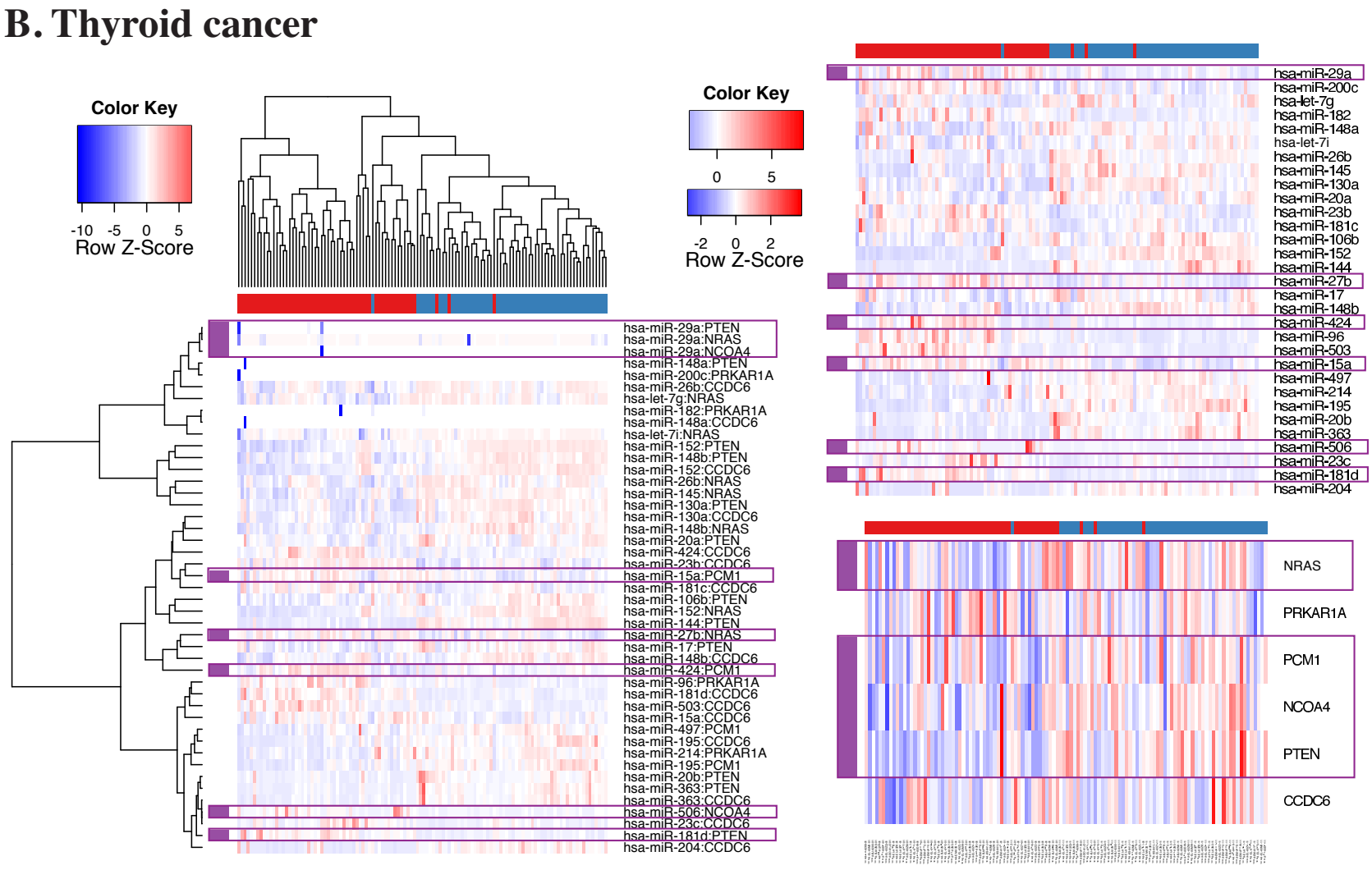




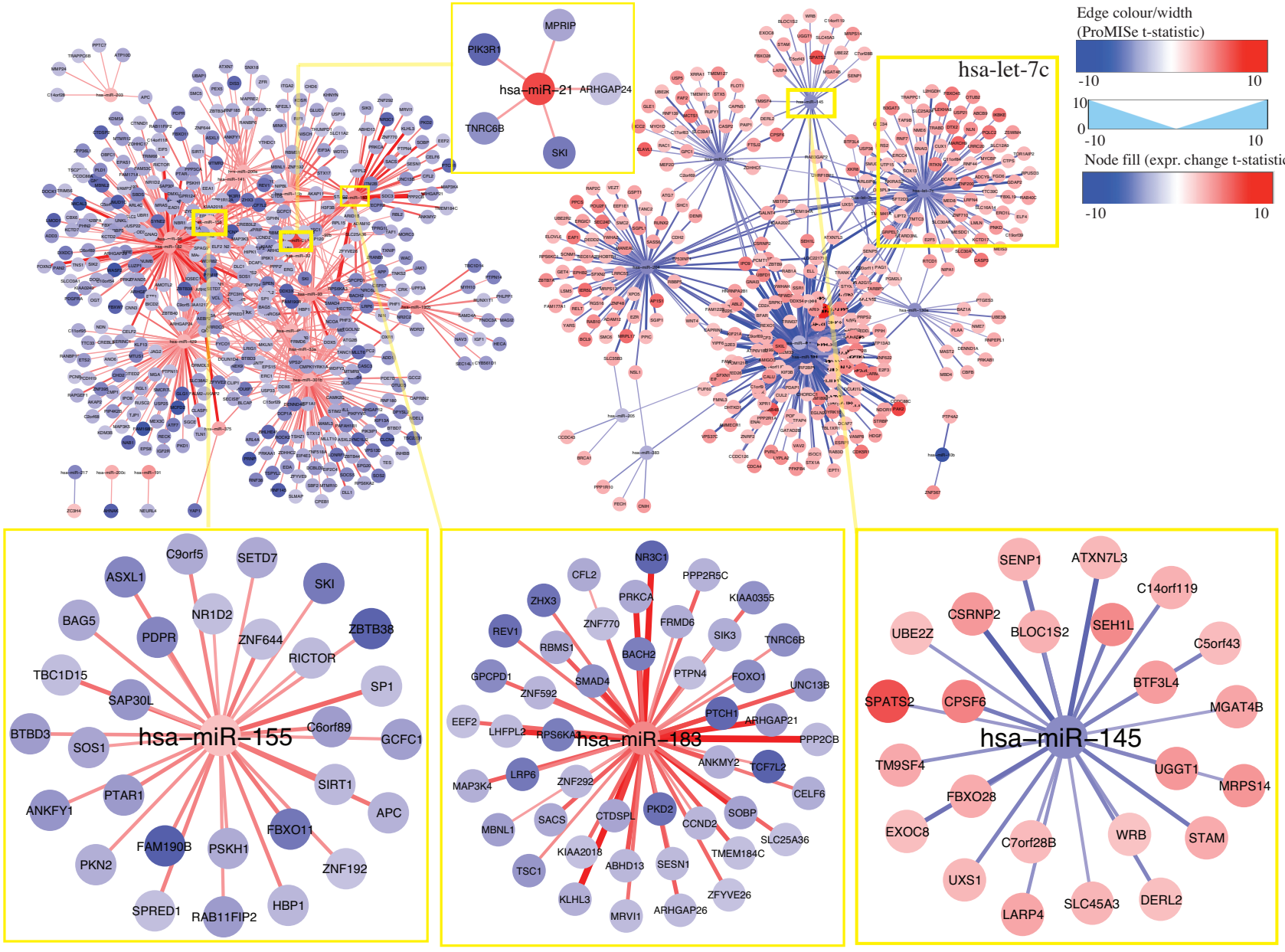
Heatmaps using differential MMI and expression profiles



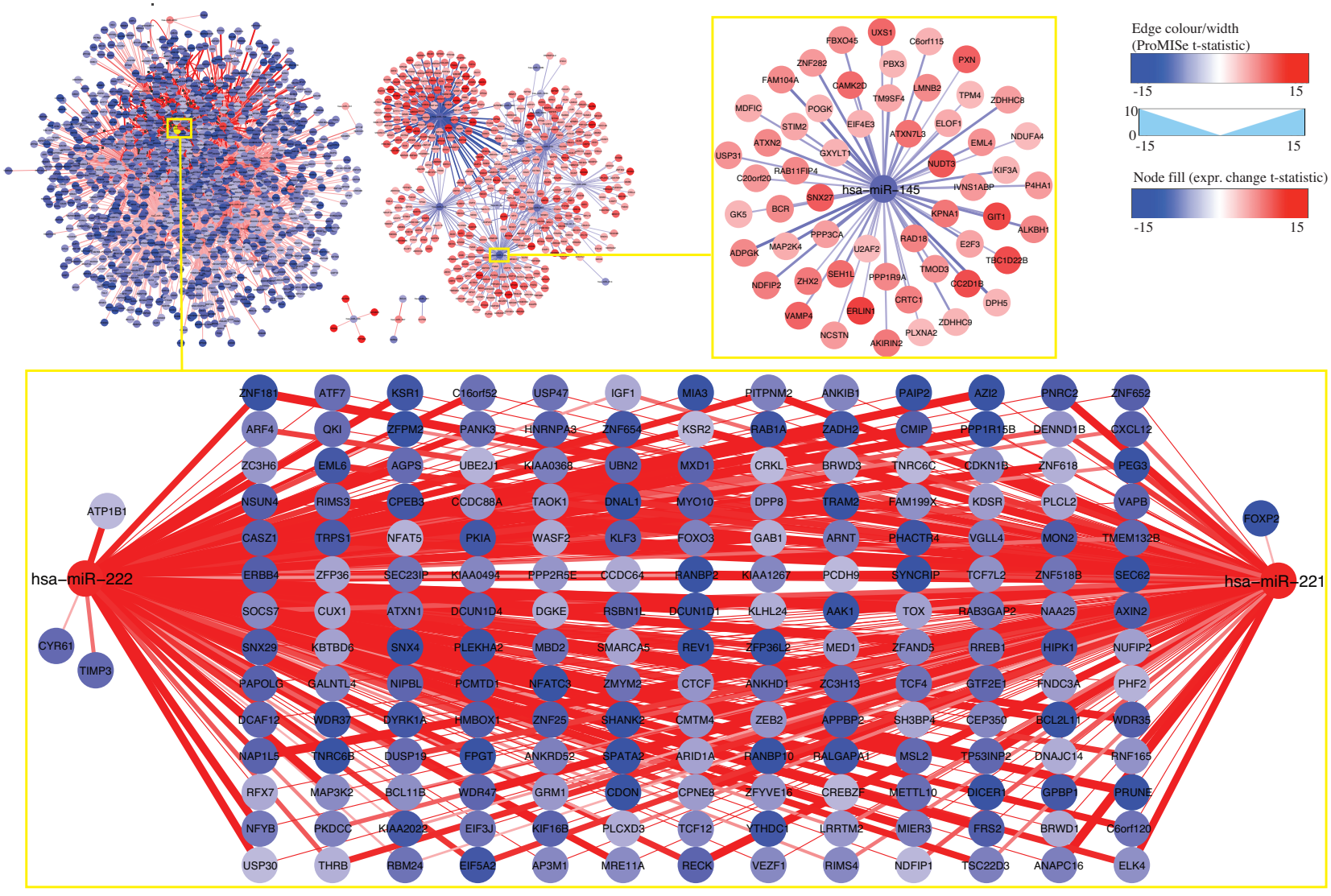
Heatmaps using differential MMI and expression profiles



Network view of the 1257 filtered MMI for breast cancer



Network view of the 3255 filtered MMI for thyroid cancer



Project 2 summary

- We described a novel probabilistic approach to infer sample-specific probabilistic miRNA-mRNA interaction signatures (ProMISe) in cancers using TCGA data
- ProMISe takes into account both miRNA as well as mRNA competition to reflect the sample-specific dynamics
- Comparing with existing methods, ProMISe demonstrates superior performance in identifying known interactions
- We identified several interesting miRNA-mRNA regulatory relationships based on paired comparison of the ProMISe signatures between normal and tumor samples from breast and thyroid cancer patients
- ProMISe is implemented as a stand-alone R package and freely available at Bioconductor website.

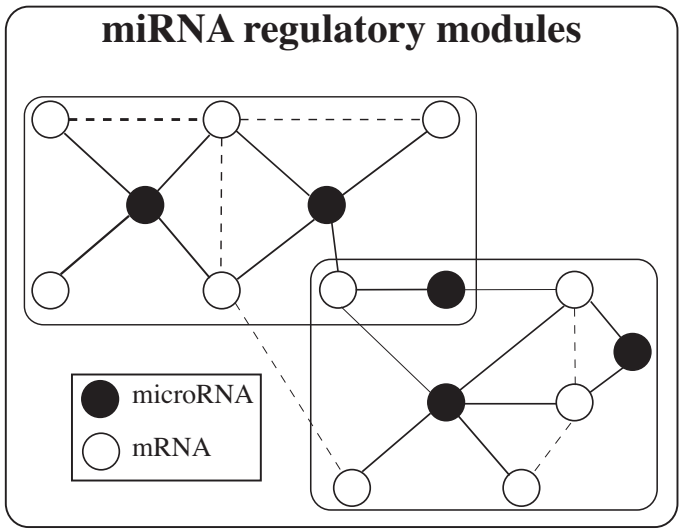
Project 3: Mirsynergy: detecting miRNA synergistic regulatory network modules by overlapping neighbourhood expansion

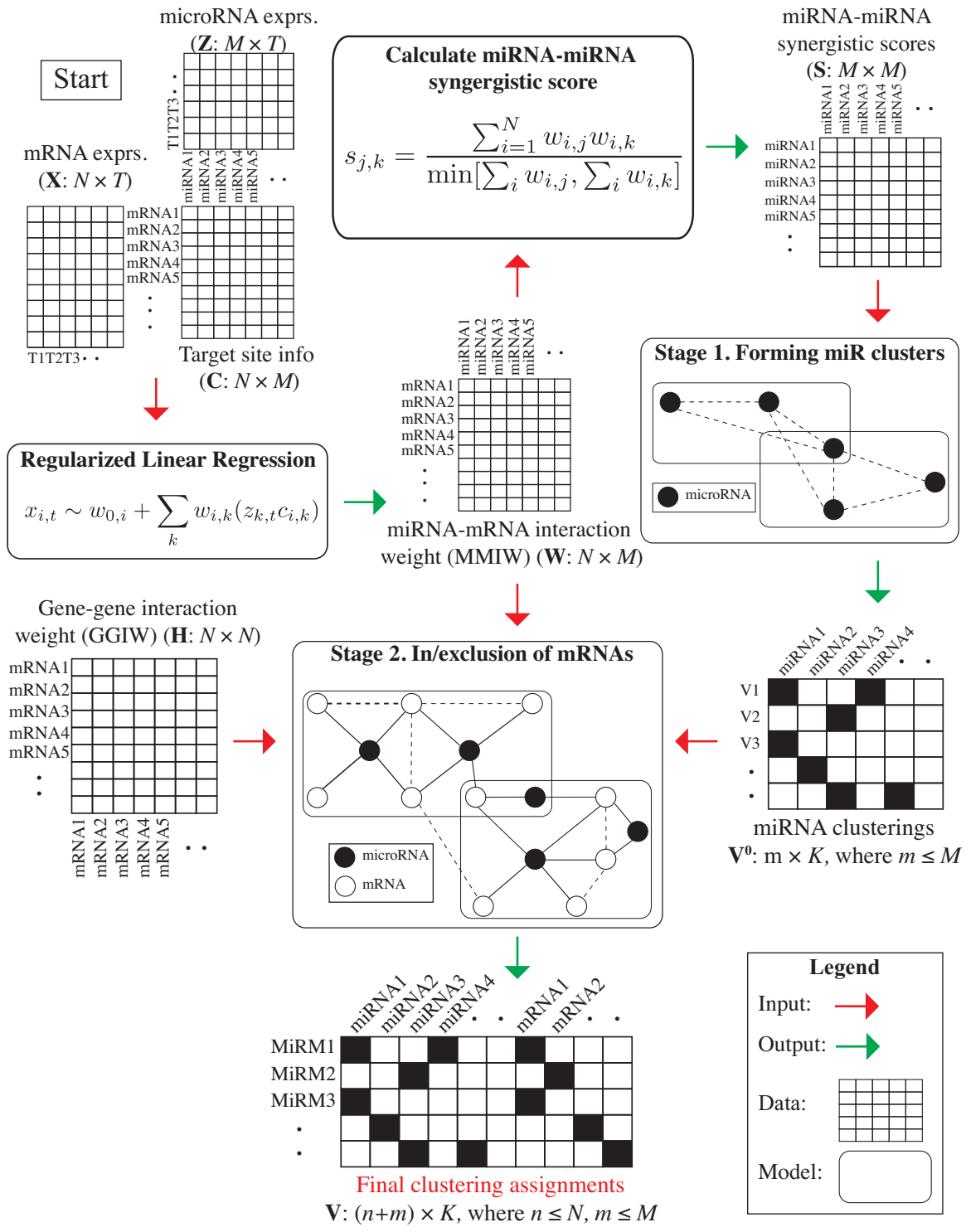
Publication:
Li, Y.*, Liang, C.*, Wong, KC., Luo, J., Zhang Z. Mirsynergy: detecting synergistic miRNA regulatory modules by overlapping neighbourhood expansion. *Bioinformatics* (Oxford, England). (May 2014)

*joint first author

MicroRNA regulatory modules (MiRMs) are plausible

- Single miRNA have rather modest effect on mRNA expression, but synergistic action of multiple miRNAs may confer more potent effects (Krek et al. 2005)
- The possible combinations of 2578 mature miRNAs (miRBase Human) may react to virtually any environmental changes
- If multiple workers (miRNAs) are engaging in overlapping tasks, then disruption of one worker may incur lesser detrimental effects than 1-task-1-worker (Boross et al., 2009)
- Existing methods designed to construct MiRMs usually require fixed number of MiRMs and computationally expensive




Start
microRNA exprs.
(Z: M × T)
mRNA exprs.
(X: N × T)
Target site info
(C: N × M)
Regularized Linear Regression
$x_{i,t} \sim w_{0,i} + \sum_k w_{i,k}(z_{k,t}c_{i,k})$
Calculate miRNA-miRNA syngergistic score
$s_{j,k} = \frac{\sum_{i=1}^{N} w_{i,j}w_{i,k}}{\min[\sum_i w_{i,j}, \sum_i w_{i,k}]}$
miRNA-miRNA synergistic scores
(S: M × M)
miRNA-mRNA interaction weight (MMIW) (W: N × M)
Stage 1. Forming miR clusters
microRNA
Gene-gene interaction weight (GGIW) (H: N × N)
Stage 2. In/exclusion of mRNAs
microRNA
mRNA
miRNA clusterings
V⁰: m × K, where m ≤ M
MiRM1
MiRM2
MiRM3
Final clustering assignments
V: (n+m) × K, where n ≤ N, m ≤ M
Legend
Input:
Output:
Data:
Model:

Mirsynergy model details:

- Edge weights $w_{i,k}$ between mRNA i and miRNA k:

$$x_{i,t} \sim w_{i,0} + \sum_k w_{i,k}(z_{k,t}c_{i,k}) \tag{1}$$

- miRNA-miRNA synergistic scores b/w miRNA j & k ($j \neq k$):

$$s_{j,k} = \frac{\sum_{i=1}^{N} w_{i,j}w_{i,k}}{\min[\sum_i w_{i,j}, \sum_i w_{i,k}]} \tag{2}$$

- Synergy score (objective function) $s(V_c)$ for MiRM V_c (Nepusz et al., 2012):

$$s(V_c) = \frac{w^{in}(V_c)}{w^{in}(V_c) + w^{bound}(V_c) + \alpha(V_c)} \tag{3}$$

- Overlap & density score:

$$\omega(V_c, V_{c'}) = \frac{|V_c \cap V_{c'}|^2}{|V_c||V_{c'}|} \tag{4}$$

$$d_1(V_c) = \frac{2w^{in}(V_c)}{m(m-1)}, \quad d_2(V_c) = \frac{w^{in}(V_c)}{n(m+n-1)} \tag{5}$$

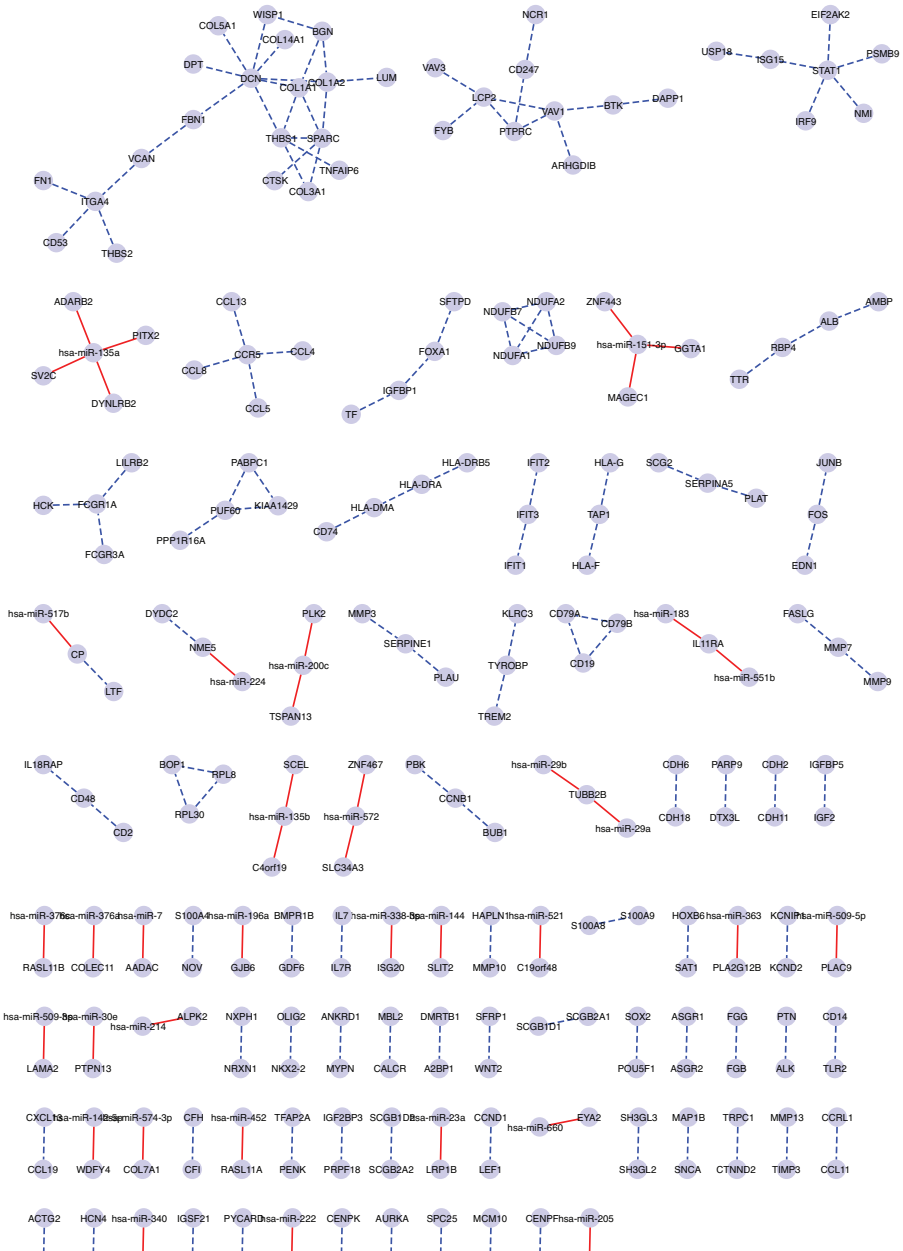
Performance summary of Mirsynergy, SNMNMF (Zhang et al., 2011), and PIMiM (Le and Bar-Joseph, 2013)

Cancer	Method	M#	$\overline{miR}$	$\overline{mR}$	GOES	MiMEC	Time
OV	Mirsynergy	**84**	**4.76**	7.57	**15.64**	**-0.05**	**1**
	SNMNMF	49	4.12	81.37	7.51	0.07	24+
	PIMiM†	40	4.7	67.80	NA	-0.013	NA
BRCA	Mirsynergy	**53**	**5.77**	24.15	**8.74**	**-0.08**	**1.5**
	SNMNMF	39	2.62	71.56	5.56	-0.04	24+
THCA	Mirsynergy	**50**	**7.60**	32.26	**8.04**	**-0.08**	**2**
	SNMNMF	39	2.23	74.82	6.73	-0.04	24+

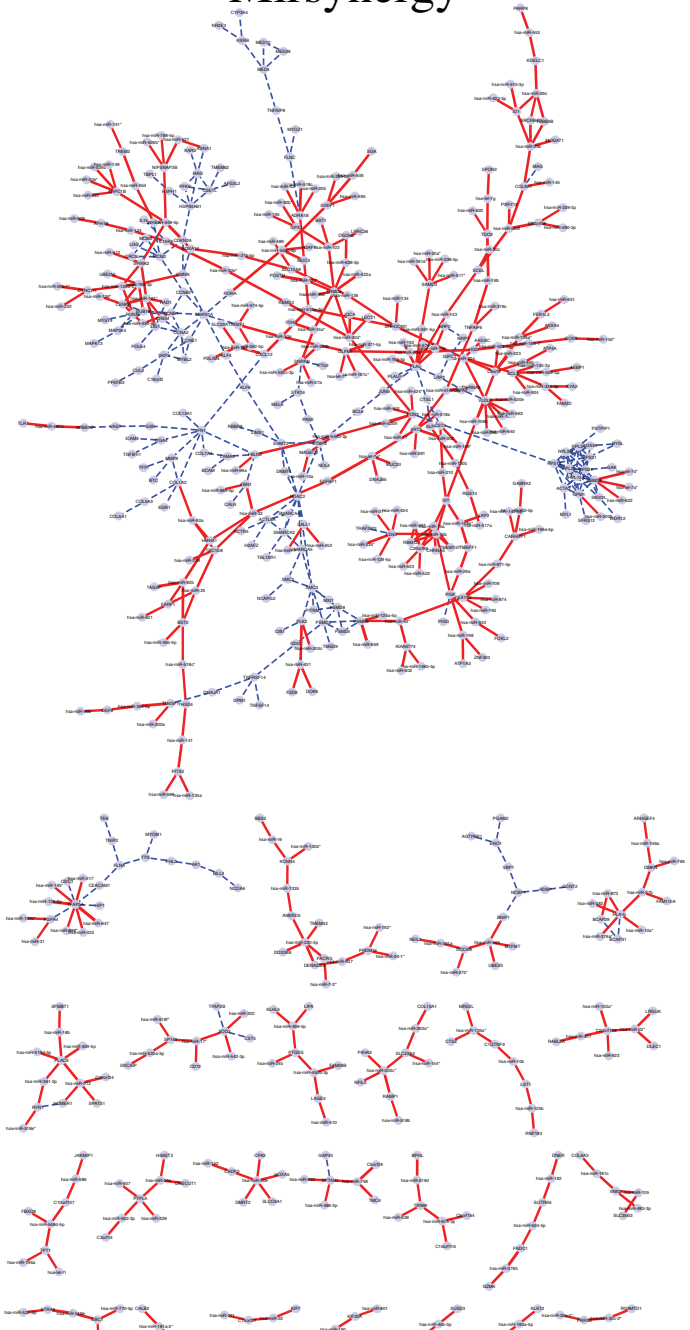
Time complexity: $\mathbf{O(M(N+M))}$ (Mirsynergy); $O(K(T+N+M)^2)$ (SNMNMF or PIMiM)

Network overview using ovarian dataset

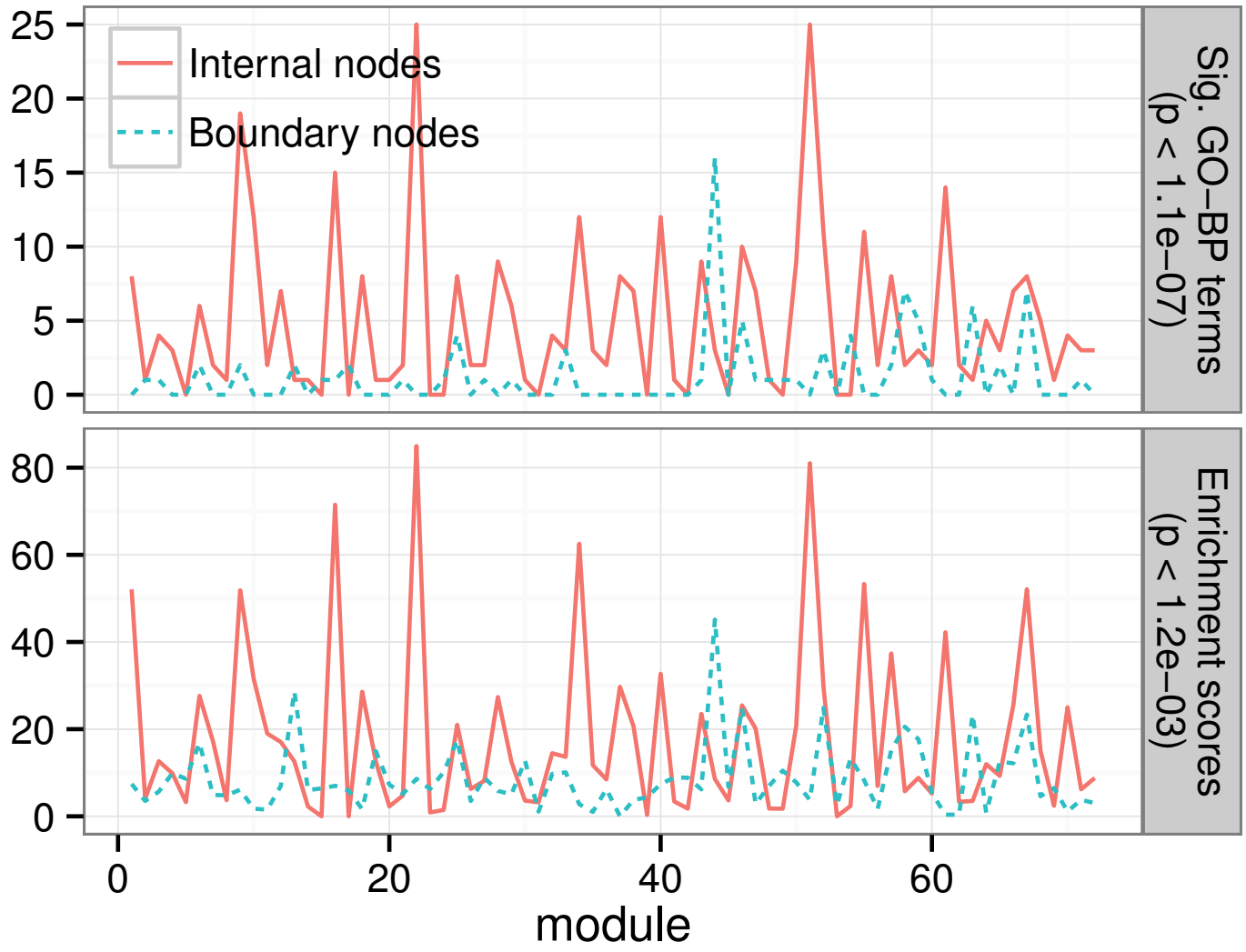
SNMNMF



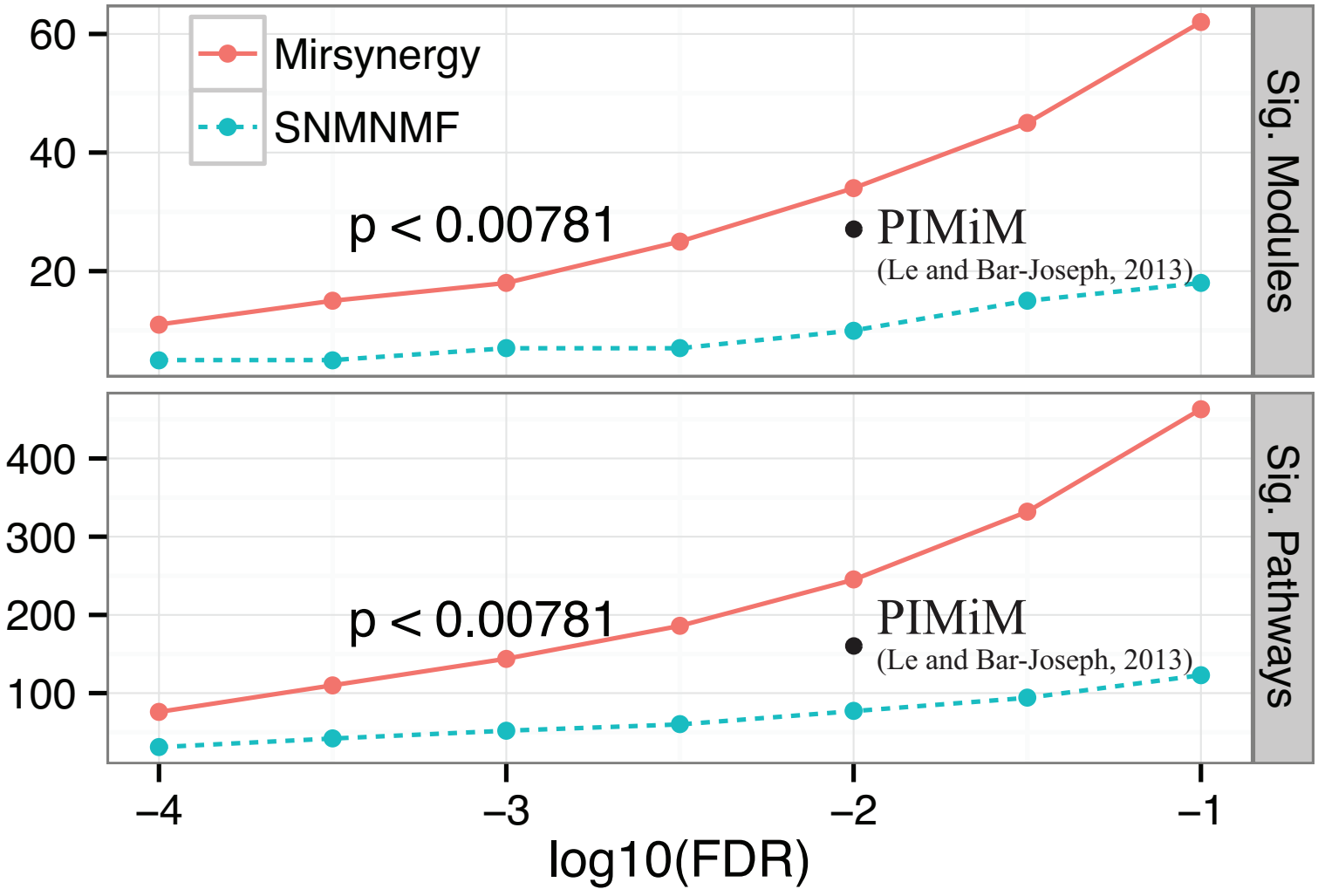
Mirsynergy



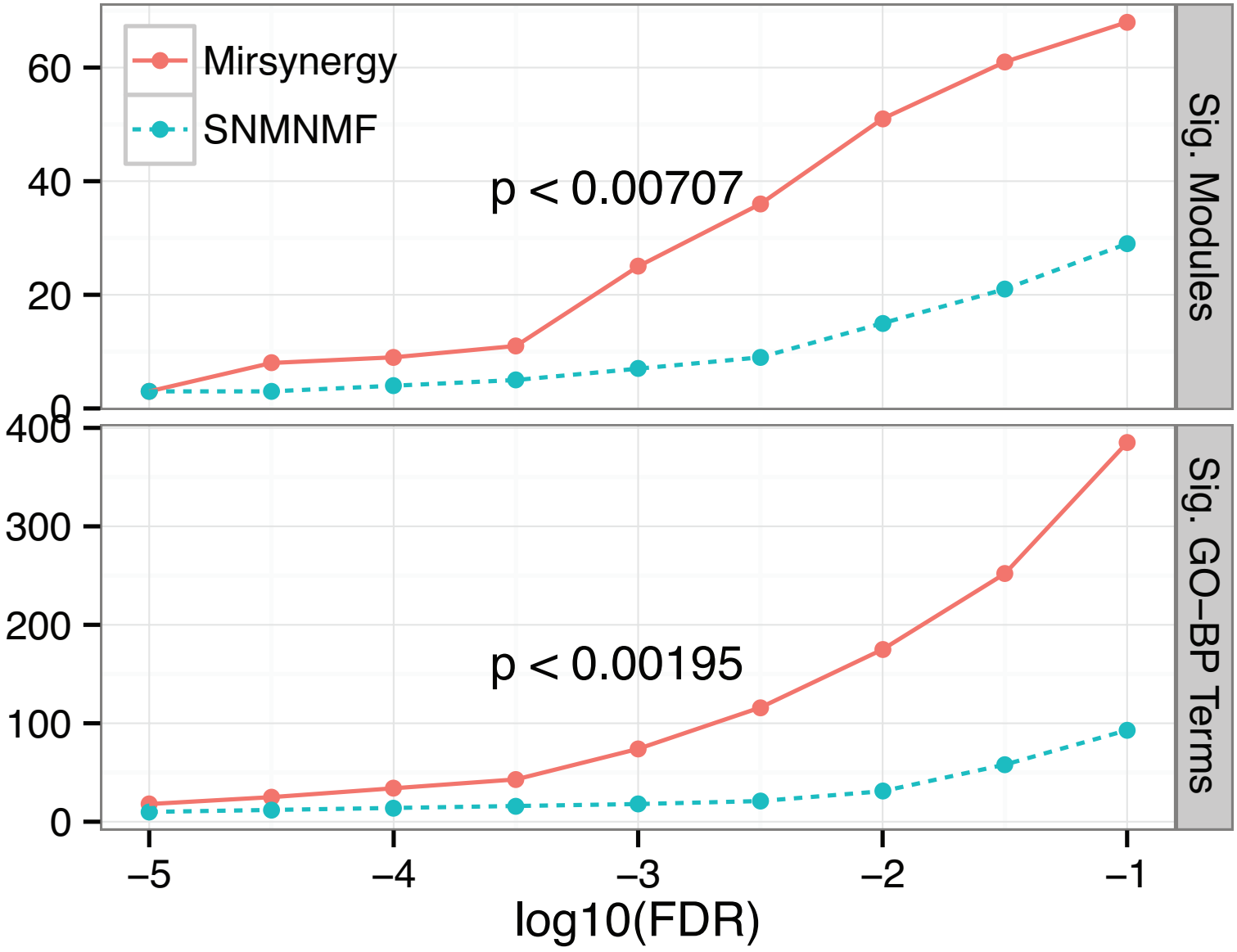
Comparison of internal and boundary nodes of each MiRM



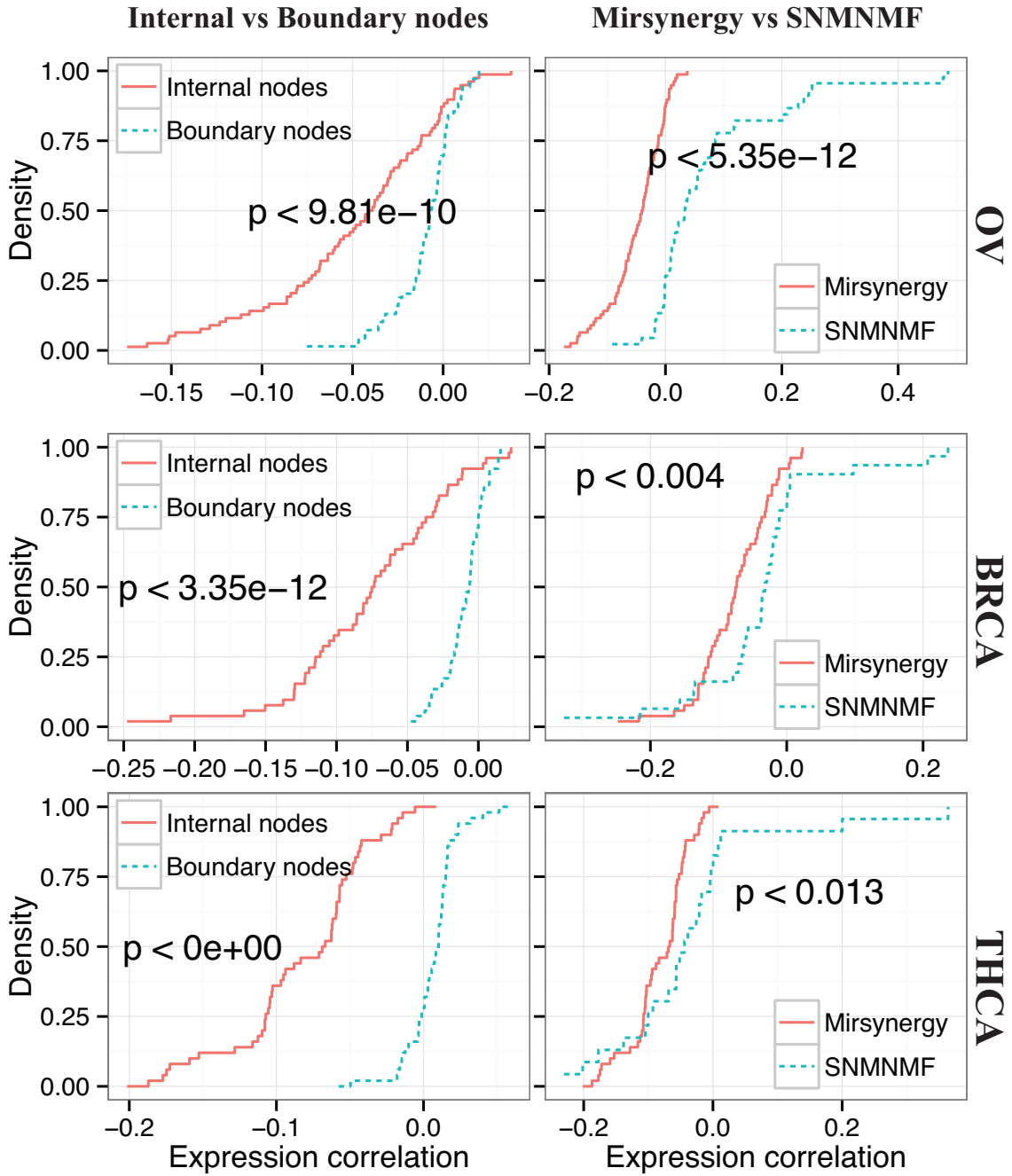
Canonical pathway enrichment comparison



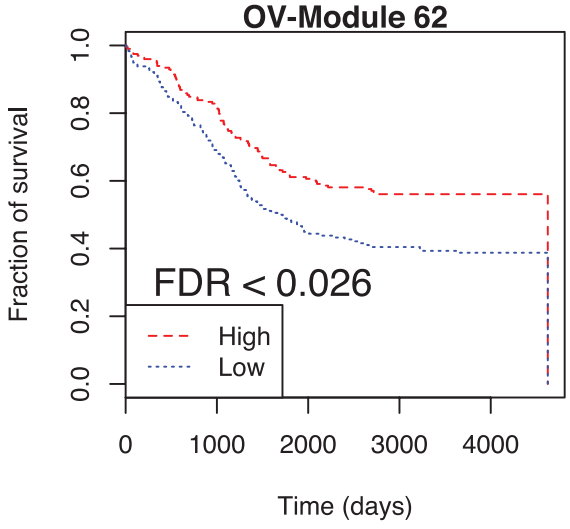
GO enrichment comparison

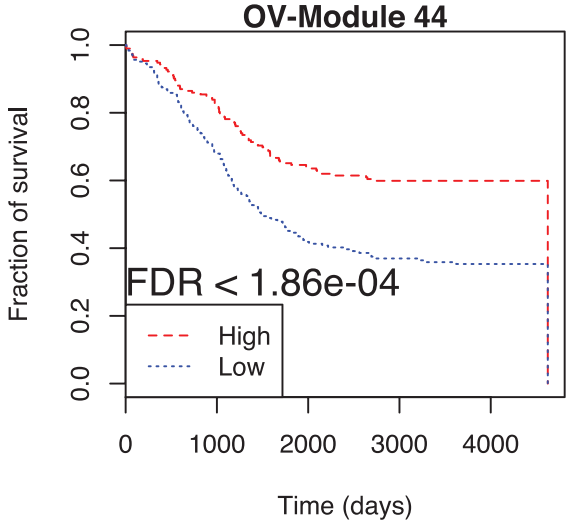


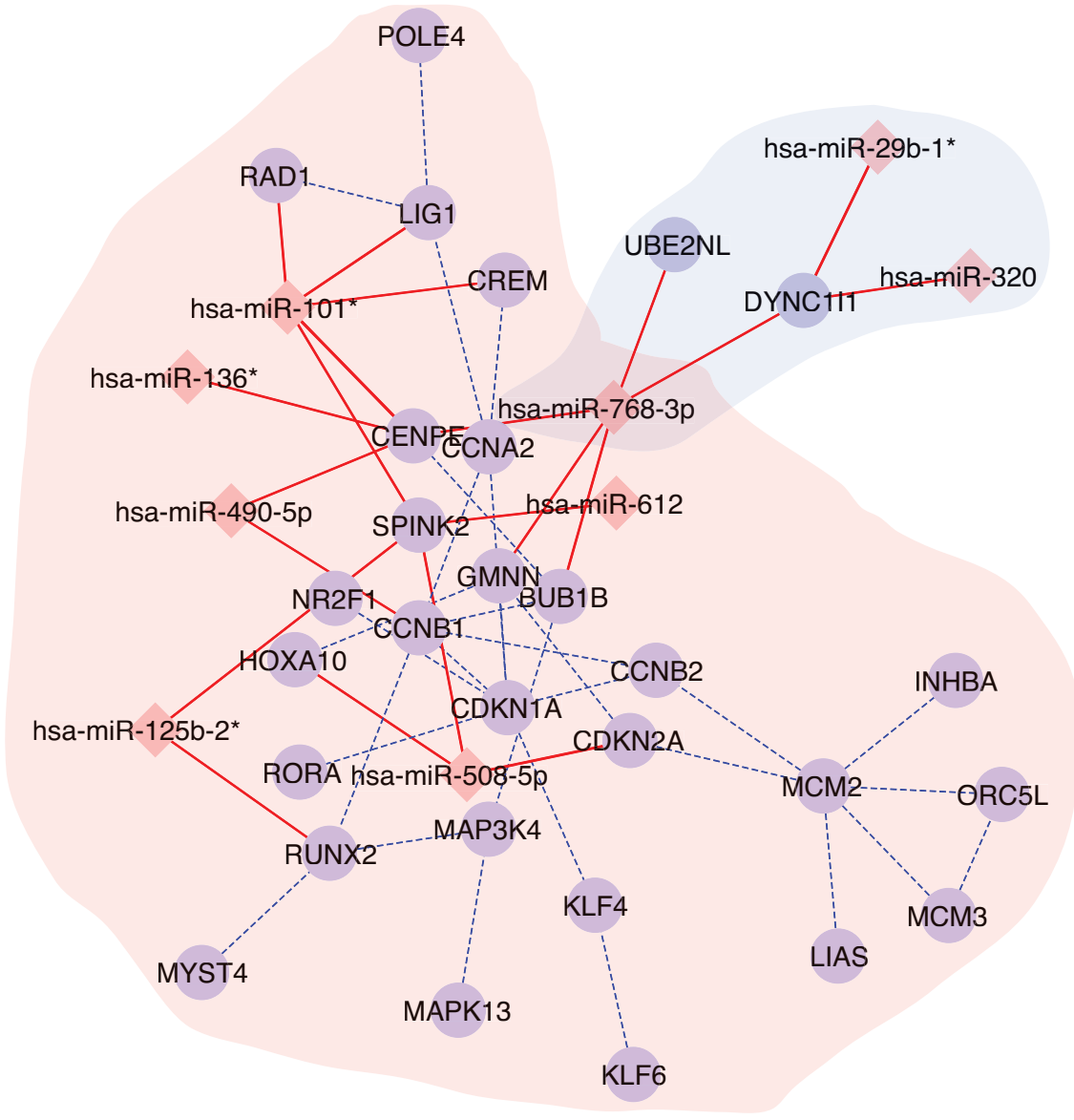
CDF of expression correlation b/w miRNA and mRNA



Survival analysis using MiRMs in ovarian cancer







Project 3 summary

- We described a two-stage clustering deterministic model called Mirsynergy to elucidate miRNA regulatory modules (MiRMs);
- Comparing with existing methods, the main salient features of Mirsynergy including automatic determination of module number while allowing sharing module memberships of both miRNAs and mRNAs, and explicit optimization of synergy score as a function of miRNA-mRNA and gene-gene interaction scores.
- MiRMs detected by Mirsynergy exhibit higher functional enrichments and more coherent expression pattern than modules constructed by other methods.
- Mirsynergy is implemented as a stand-alone R package and freely available at Bioconductor website.

Project 4: Regression analysis of combined gene expression regulation in acute myeloid leukemia

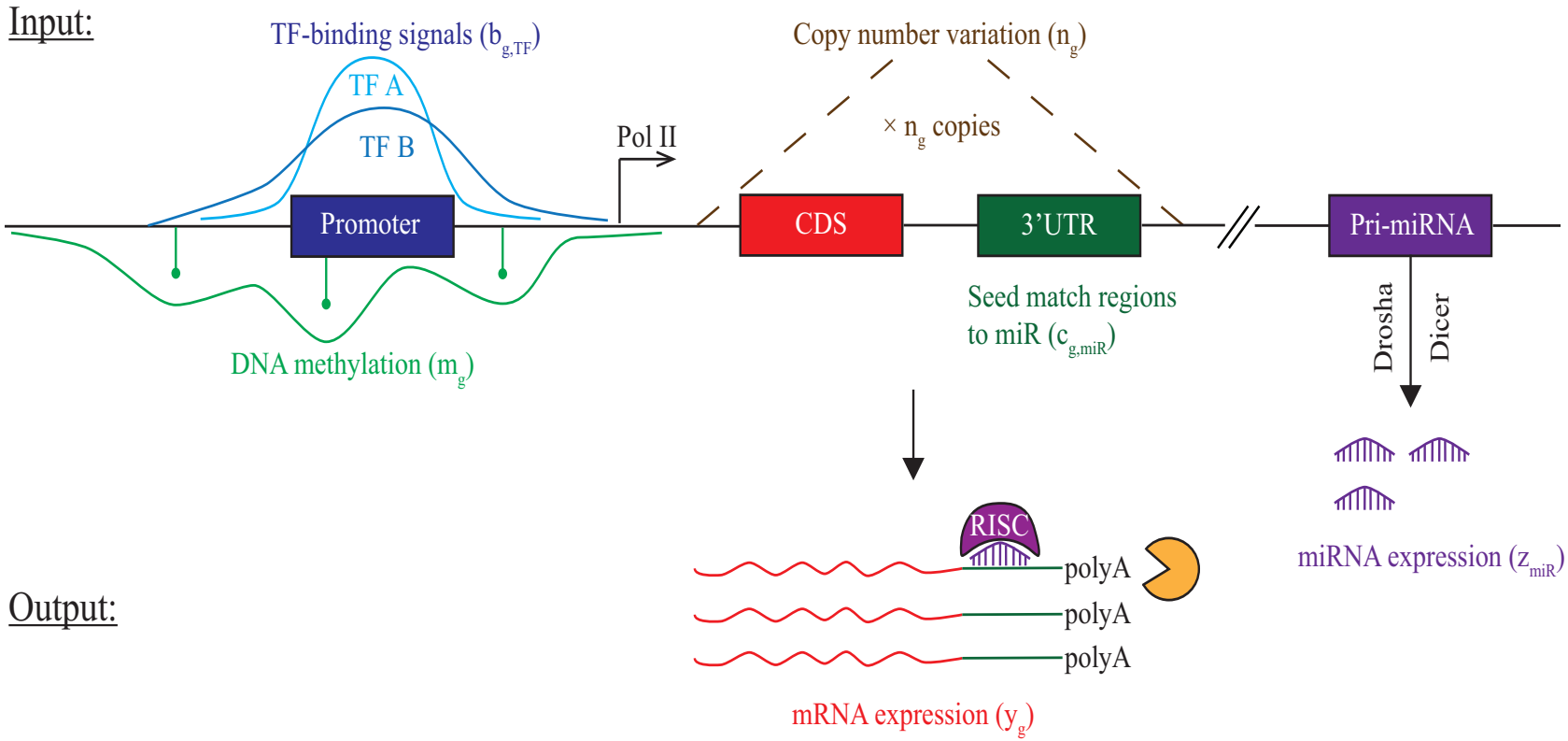
Publication:
Li, Y., Liang, M., Zhang Z. Regression Analysis of Combined Gene Expression Regulation in Acute Myeloid Leukemia. *PloS Computational Biology* (in revision)

Motivation

- Recent studies from TCGA showed that most Acute Myeloid Leukemia (AML) patients lack DNA mutations (TCGA Research Network, 2013; *New England Journal of Medicine*, 368(22), 2059-2074)
- AML tumorigenesis may be explained better by aberrant molecular signatures at the transcriptional and epigenetic levels
- However, there is a lack of an integrative model to take full advantage of the available rich yet heterogeneous data by combining TCGA and ENCODE data

RACER: Regression Analysis of Combined Expression Regulation

A



B

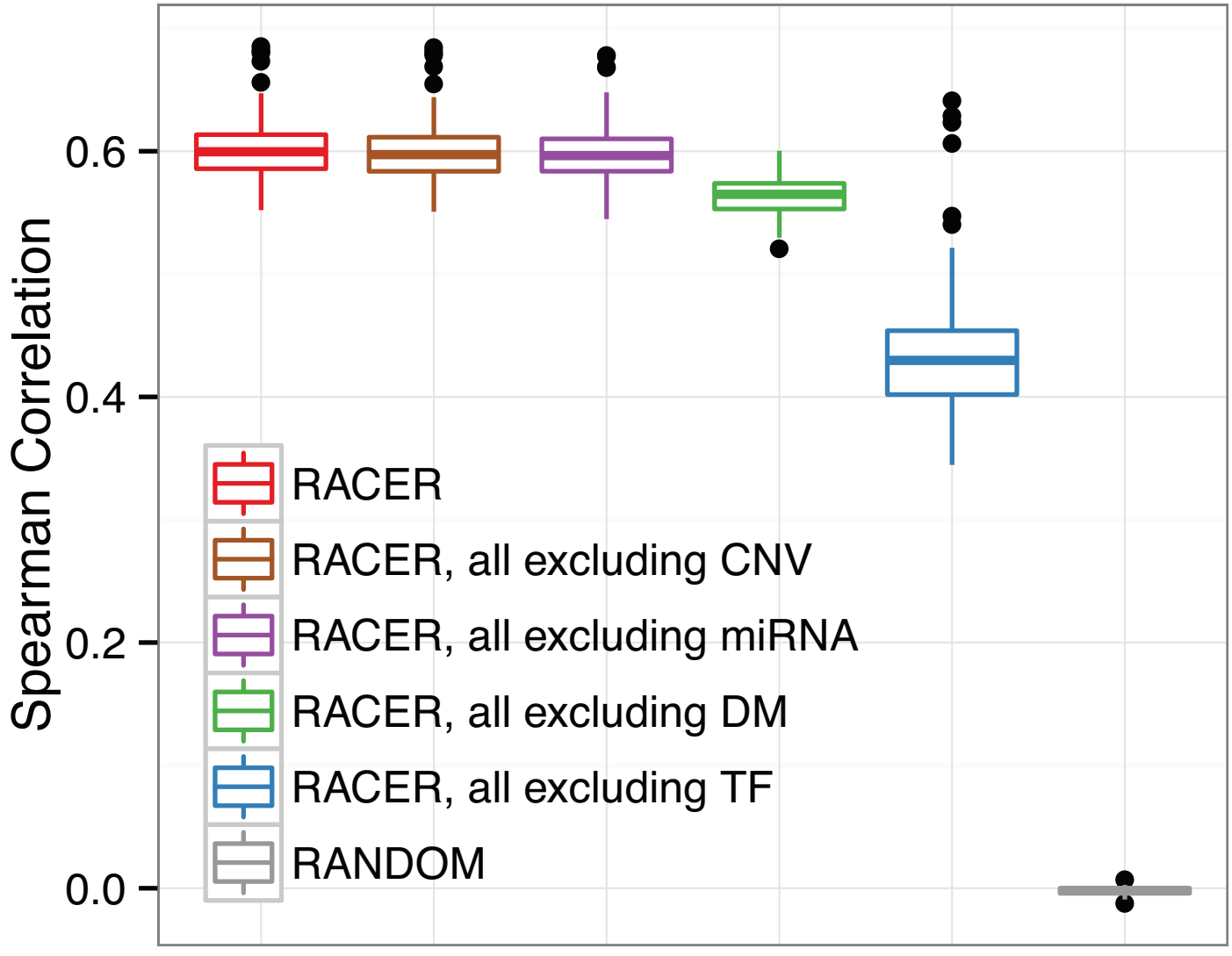
Stage 1: Estimate sample-specific TF and miR activities ($\alpha_{TF,t}$, $\alpha_{miR,t}$) in sample t:

$$[y_{g,t}]_{N\times1} \approx \alpha_0 + \alpha_{CNV,t}[n_{g,t}]_{N\times1} + \alpha_{DM,t}[m_{g,t}]_{N\times1} + [b_{g,TF}]_{N\times K} \times [\alpha_{TF,t}]_{K\times1} + [c_{g,miR}]_{N\times M} \times ([\alpha_{miR,t}]_{M\times1}[z_{miR,t}]_{M\times1})$$

Stage 2: Estimate TF-gene and miRNA-mRNA interactions ($W_{TF,g}$, $W_{g,miR}$) for gene g across all samples:

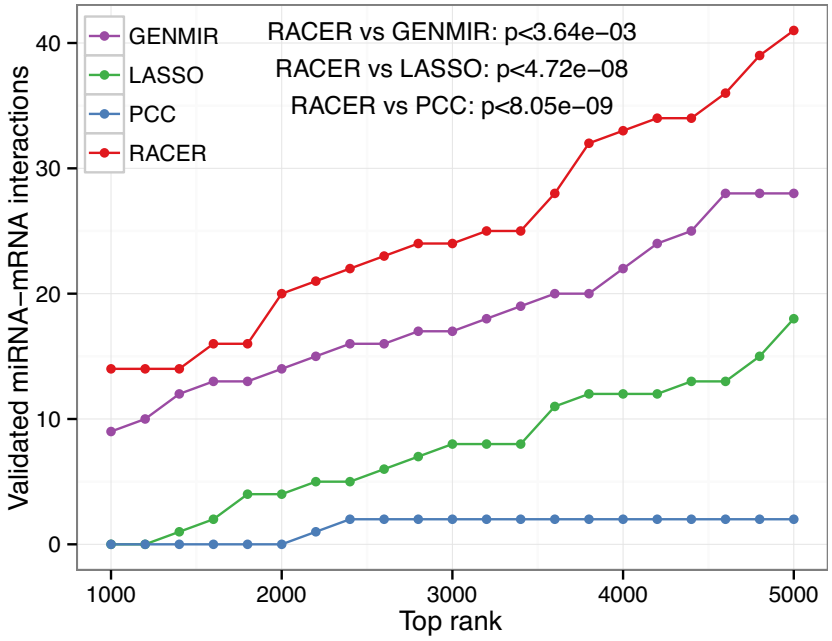
$$[y_{g,t}]_{1\times T} \approx w_0 + w_{g,CNV}[n_{g,t}]_{1\times T} + w_{g,DM}[m_{g,t}]_{1\times T} + [w_{g,TF}]_{1\times K^*} \times [\alpha_{TF,t}]_{K^*\times T} + [w_{g,miR}]_{1\times M^*} \times [\alpha_{miR,t}]_{M^*\times T}$$

TF regulation and DNA methylation confer the highest explanatory power of mRNA expression in AML

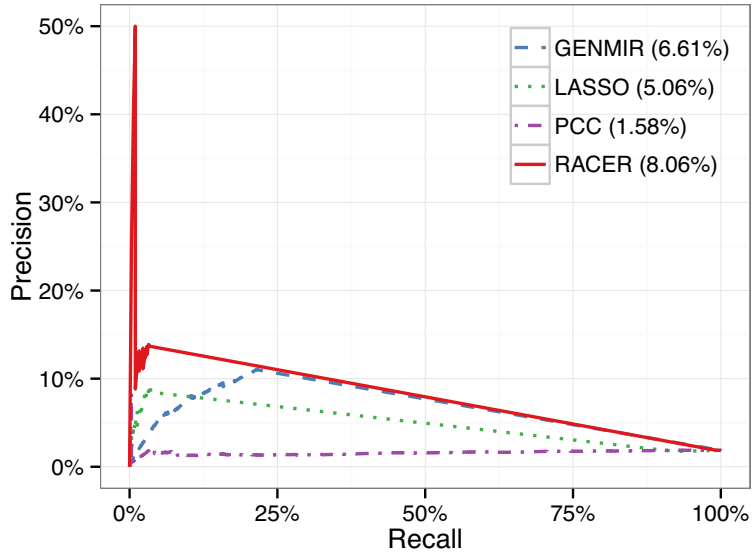


Power analysis of miRNA target predictions

A. Validated pairs from MirTarBase

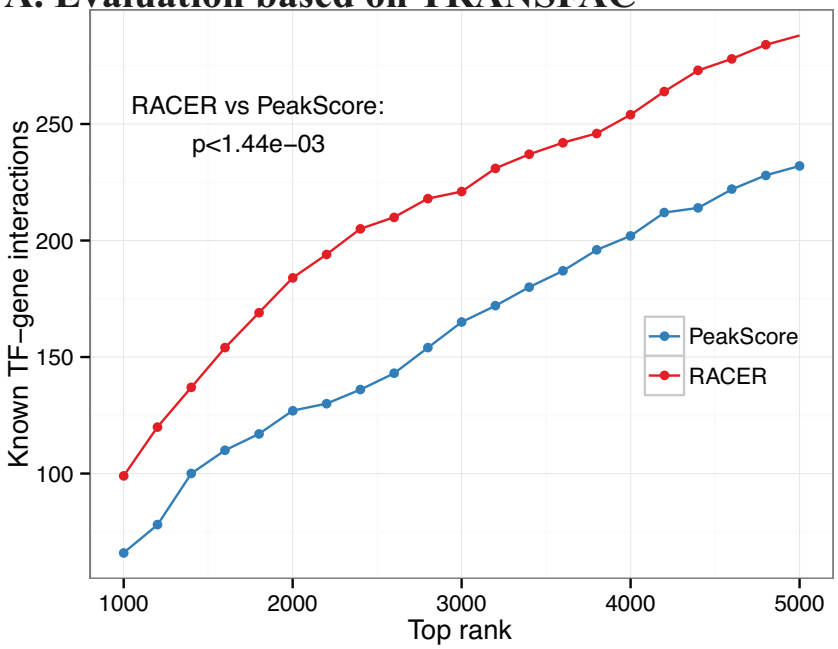


B. Confidence targets of miR-34a in K562-KD

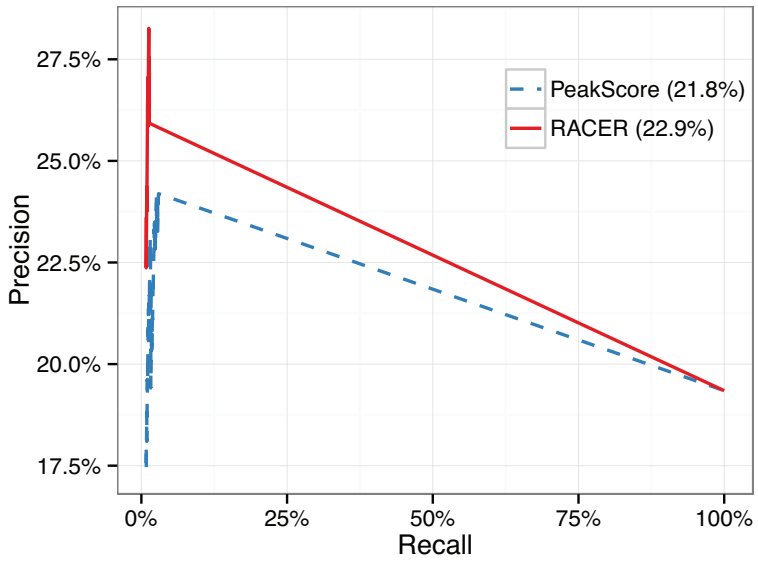


Power analysis of TF target predictions

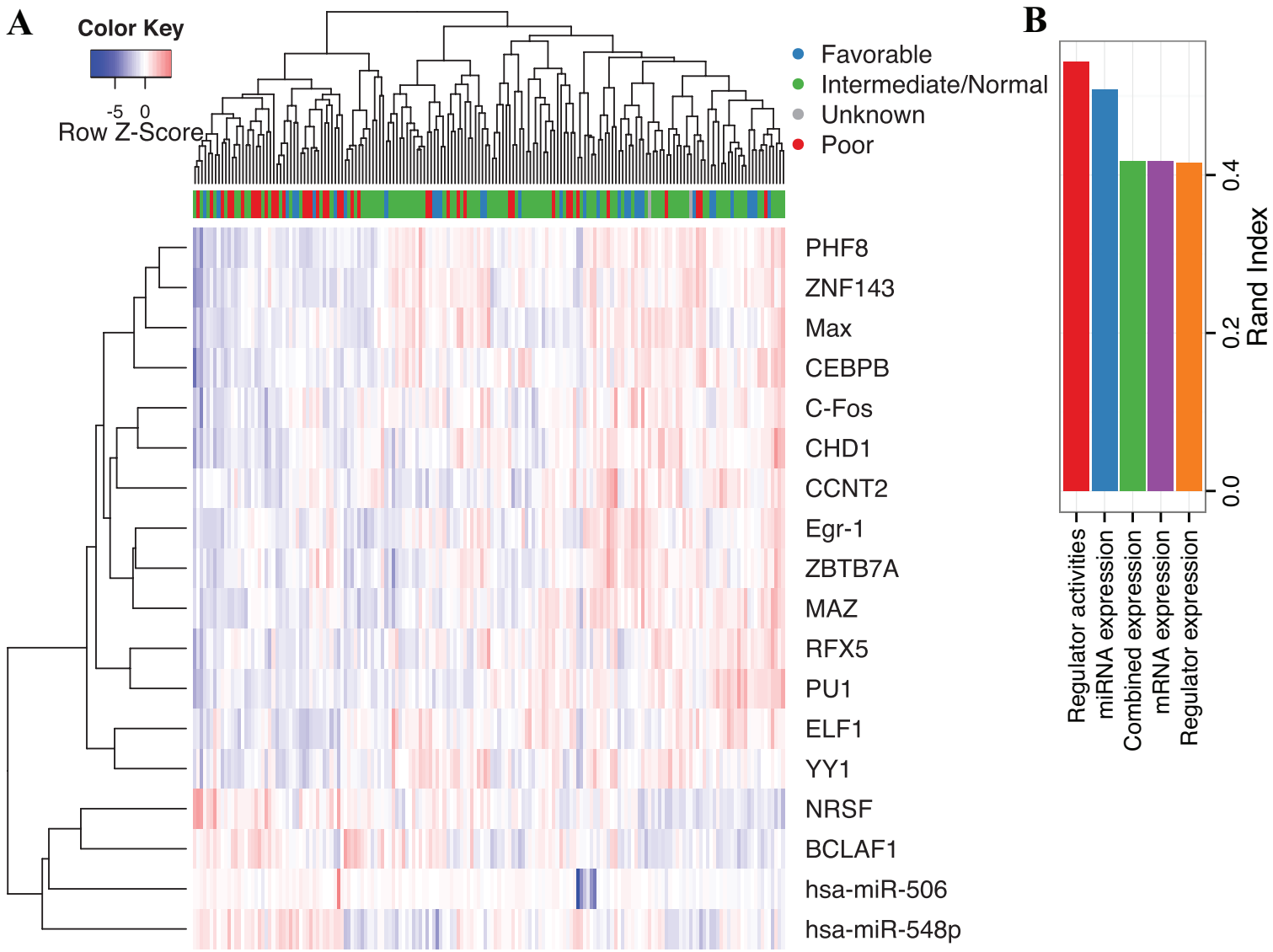
A. Evaluation based on TRANSFAC



B. Confidence targets from GATA2 K562-KD

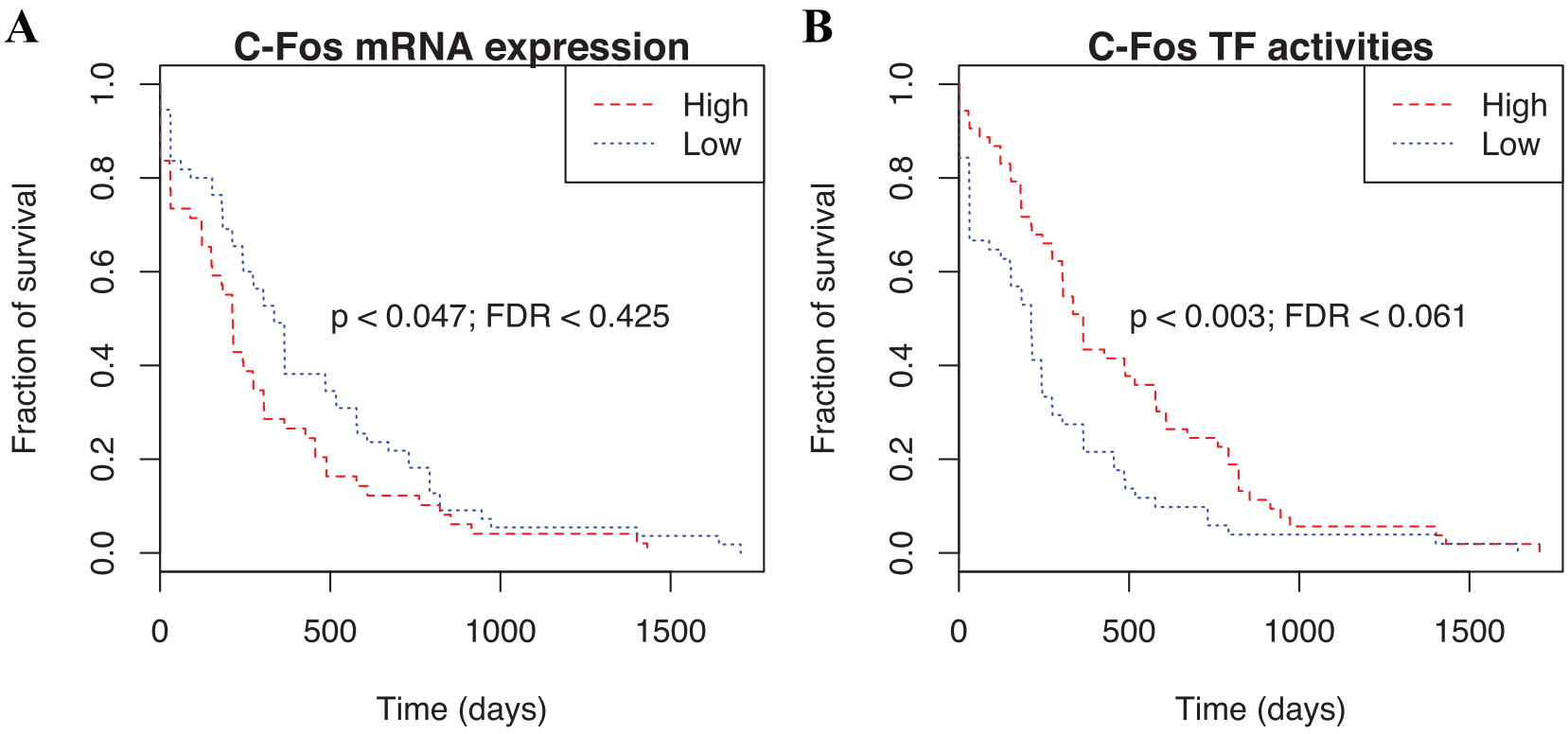


Selected regulator activities cluster by cytogenetic risk

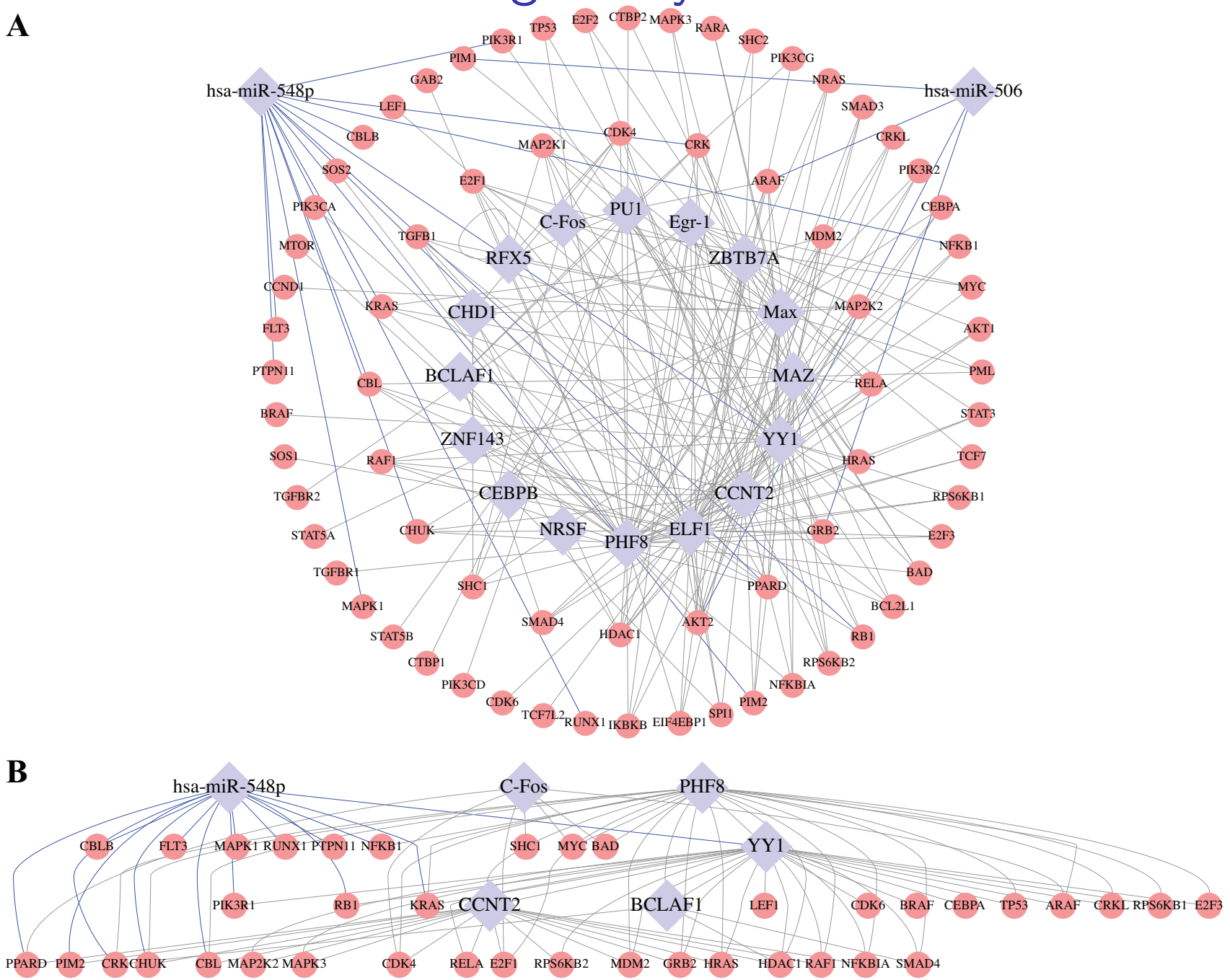


Feature selection: $F(1, N-M-K+1) = \frac{(RSS_{\text{RACER, all excluding regulatorX}} - RSS_{\text{RACER}})}{RSS_{\text{RACER}}/(N-M-K+1)}$

Kaplan-Meier survival analysis



AML regulatory network



Project 4 summary

- Using recently available data from ENCODE and TCGA, we developed a two-stage regression model called RACER to infer the regulatory activities and their target genes in AML
- Based on model comparison, TF and DM contribute the highest explanatory power to mRNA expression
- RACER demonstrates superior performance in identifying known miRNA-mRNA and TF-gene interactions
- Our analysis revealed 18 AML-related regulators, including miR-548p, whose targets are enriched for leukemia-related genes including YY1 as well as c-Fos a potentially novel prognostic marker in AML.

Summary

1. TargetScore: We developed a Bayesian probabilistic scoring strategy using miRNA overexpression and target site information, which demonstrates superior performance over the existing methods in predicting validated miRNA targets in various human cell-lines.
2. ProMISe: We described a novel probabilistic approach to infer sample-specific miRNA-mRNA interaction signatures in cancers using TCGA data, taking into account both miRNA as well as mRNA competition to reflect the sample-specific dynamics.
3. Mirsynergy: We described a two-stage clustering deterministic model to elucidate miRNA regulatory modules with module number systematically determined, module memberships shared among mi/mRNAs, and explicit optimization of synergy score defined as a function of miRNA-mRNA and gene-gene interaction scores.
4. RACER: Using recently available data from ENCODE and TCGA, we developed a two-stage regression model to infer the regulatory activities and their target genes in AML. Our analysis revealed 18 AML-related regulators, including miR-548p, whose targets are enriched for leukemia-related genes including YY1 as well as c-Fos a potentially novel prognostic marker in AML.

Ongoing and future works

1. Predicting recurrent miRNA-mRNA interactions or pan-cancer miRNA regulatory network via joint Bayesian analysis across diverse types of cancer
2. Learning miRNA regulatory elements using AGO-cross-linked miRNA:mRNA ligation data
3. A general statistical model to infer differential N^6-methyladenosine or m6A modification (Wang, Y., **Li, Y.**, et al., 2014)

Select publications:

- **Li, Y.**, Liang, M., Zhang Z. Regression analysis of combined gene expression regulation in acute myeloid leukemia. *PloS Computational Biology* (in revision)
- **Li, Y.***, Liang, C.*, Wong, KC., Luo, J., Zhang Z. Mirsynergy: detecting synergistic miRNA regulatory modules by overlapping neighbourhood expansion. *Bioinformatics*. (Oxford). Advance Access. (May, 2014)
- **Li, Y.**, Liang, C., Wong, KC, Jin, K., and Zhang, Z. Inferring probabilistic miRNA-mRNA interaction signatures in cancers: a role-switch approach. *Nucleic Acids Research*, **42**(9), e76. (Feb, 2014)
- **Li, Y.**, Goldenberg, A., Wong, KC., Zhang Z. A probabilistic approach to explore human miRNA targetome by integrating miRNA-overexpression data and sequence information. *Bioinformatics*. (Oxford), **30**(5), 621–628. (Oct, 2013)

- **Li, Y.***, Liang, C.*, Easterbrook, S., Luo, J., Zhang Z. Investigating functional implication of reinforcing feedback loops in transcriptional regulatory network. *Molecular Biosystems* (under review) (*joint first author)
- Wang, Y., **Li, Y.**, Toth, JI., Petroski, MD., Zhang, Z., and Zhao J. N^6-methyladenosine modification destabilizes developmental regulators in embryonic stem cells. *Nature Cell Biology*, **16**(2), 1-10. (Nov, 2013)
- **Li, Y.**, Zhao, D. Y., Greenblatt, J. F., and Zhang, Z. RIPSeeker: a statistical package for identifying protein-associated transcripts from RIP-seq experiments. *Nucleic Acids Research*, **41**(8), e94. (March, 2013)
- **Li, Y.**, Arsenault, R. J., Trost, B., Slind, J., Griebel, P. J., Napper, S., and Kusalik, A. A Systematic Approach for Analysis of Peptide Array Kinome Data. *Science Signaling*, **5**(220), pl2-pl2. (April, 2012)

*joint first author

Acknowledgements

Zhang Lab:

- Prof. Zhaolei Zhang (PhD advisor)
- Cheng Liang (visiting CS PhD student from Hunan Univ.)
- Ka-Chun Wong (CS PhD student at UT)
- Michael Minggao Ming (Mol.Gen. PhD student at UT)

PhD committee members:

- Prof. Anna Goldenberg (Sick Kid's Hospital & UT)
- Prof. Anthony Bonner (UT)

Collaborators on RNA epigenetic (m6A) project:

- Prof. Crystal Jing Zhao (Sanford-Burnham Medical Research Institute)
- Yang Wang (Postdoc at Zhao lab)

Collaborators on RIP-seq project:

- Prof. Jack Greenblatt
- Dorothy Zhao (PhD student at Greenblatt lab)

Scholarships:

- Natural Sciences and Engineering Research Council (NSERC) of Canada Graduate Scholarship (CGS) Doctoral (2012-15) & Master (2010-2011)
- Ontario Graduate Scholarship (2011-2012); C.C. Gotlieb Fellowship (2011)