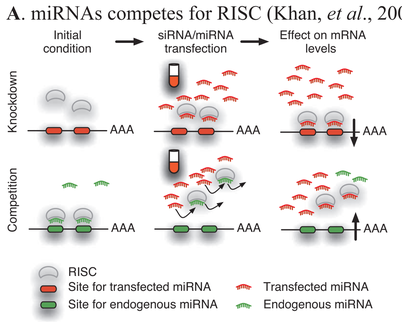
Inferring probabilistic miRNA-mRNA interaction signatures in cancers: a role-switch approach

Yue Li[1,2], Cheng Liang [2], Ka-Chun Wong[1,2], Ke Jin [2,3], and Zhaolei Zhang[1,2,3,4,*]

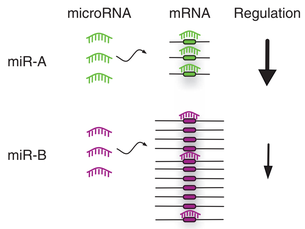
[1]*Department of Computer Science,* [2]*The Donnelly Centre,* [3]*Department of Molecular Genetics,* [4]*Banting and Best Department of Medical Research, University of Toronto*

Background

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

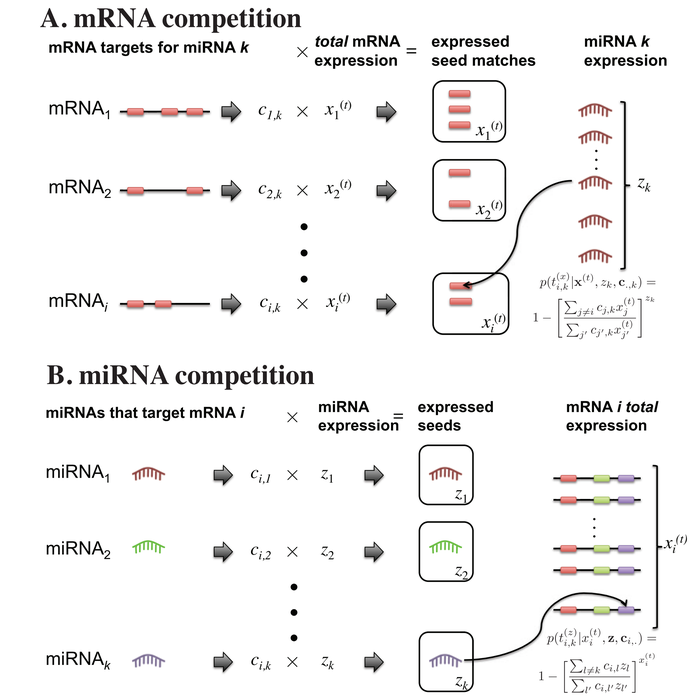


B. mRNA abundance "dilute" individual miRNA effects on each mRNA (Arvey, *et al.*, 2010)

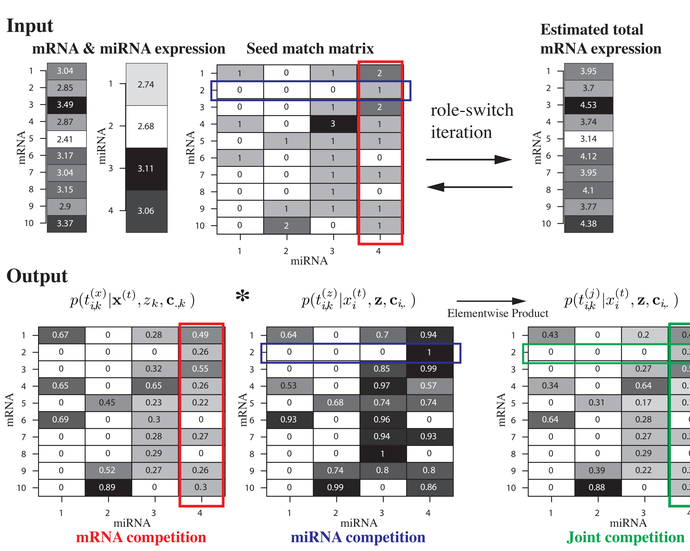


- Existing sequence-based miRNA target prediction methods:
  - Approach: seed match, evolutionary conservation, binding free energy
  - Limitation: Lots of false positives ($\sim$ 2:1 signal:noise); not context-specific
- Existing expression-based miRNA target prediction methods:
  - Approach: Pearson correlation (Tsang *et al.*, 2007); Bayesian inference (Huang *et al.*, 2007); LASSO (Lu *et al.*, 2011)
  - Limitation: Require lots of expression profiling data; mRNA competition is rarely considered but important; Not sample-specific

Probabilistic MiRNA-mRNA Interaction Signature

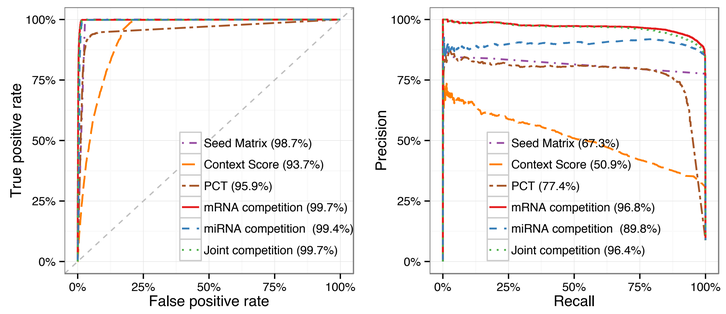


1. Initialize $\mathbf{x}^{(t)} = \mathbf{x}^{(o)}$
2. $p(t_{i,k}^{(x)}|\mathbf{x}^{(t)}, z_k, \mathbf{c}_{.,k}) = 1 - \left[\frac{\sum_{j\neq i} c_{j,k} x_j^{(t)}}{\sum_{j'} c_{j',k} x_{j'}^{(t)}}\right]^{z_k}, \quad p(t_{i,k}^{(z)}|x_i^{(t)}, \mathbf{z}, \mathbf{c}_{i,.}) = 1 - \left[\frac{\sum_{l\neq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



Results

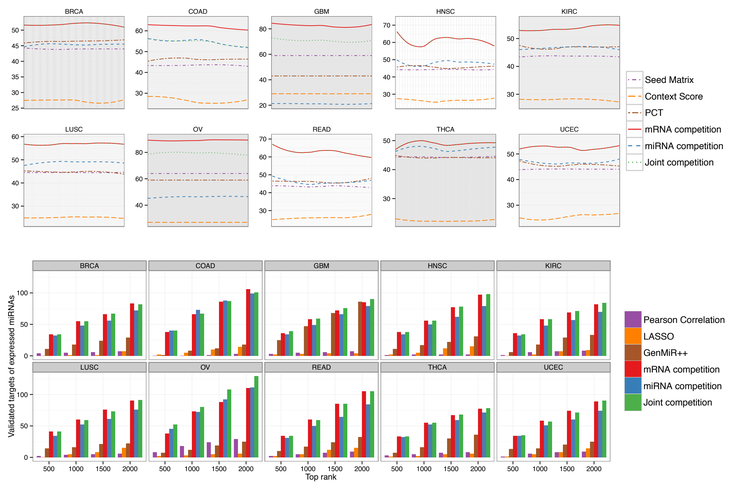
Identification of confidence targets from PAR-CLIP and miRNA knockdown data in HEK293 (Hafner et al., 2010)



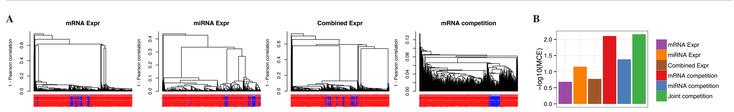
Processed expression data for 10 cancer types from The Cancer Genome Atlas (TCGA)

BRCA, COAD, GBM, HNSC, KIRC, LUSC, OV, READ, THCA, UCEC

Identification of validated targets for 10 cancer types from TCGA data

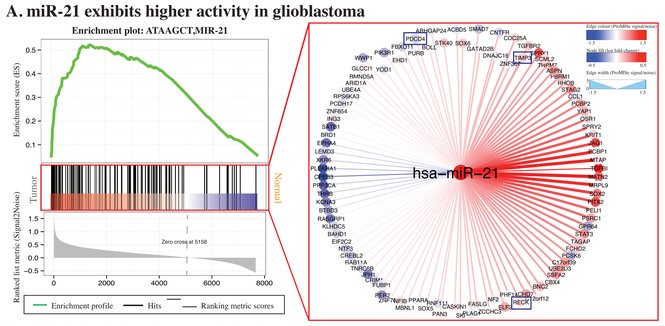


Thyroid cancer tumors exhibit distinct ProMISe signatures

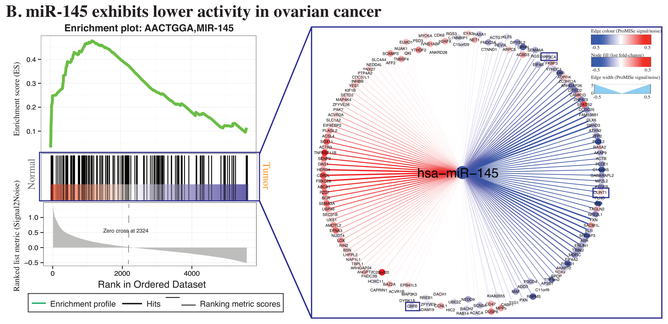


Gene set enrichment analysis using averaged miRNA-mRNA interactions per gene

A. miR-21 exhibits higher activity in glioblastoma

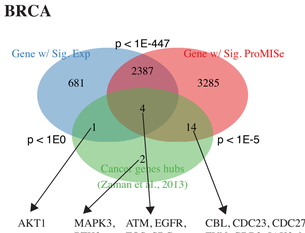


B. miR-145 exhibits lower activity in ovarian cancer

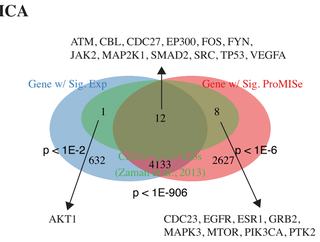


Paired t-test on 14 or 58 matched breast cancer (BRCA) or thyroid cancer (THCA) reveal cancer gene hubs significantly enriched in the predicted miRNA dysregulated genes

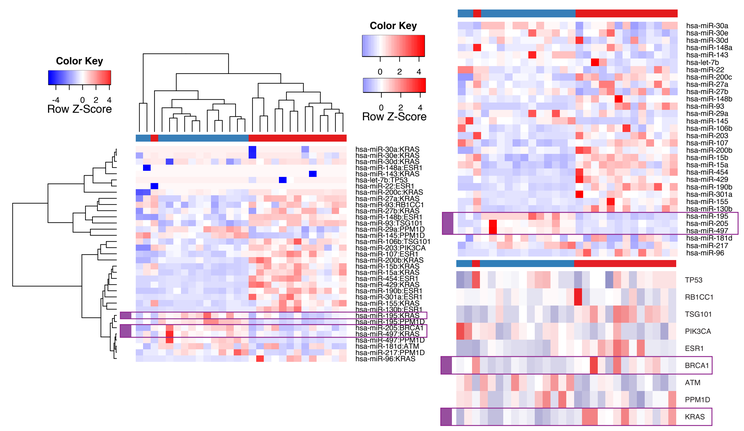
A. BRCA



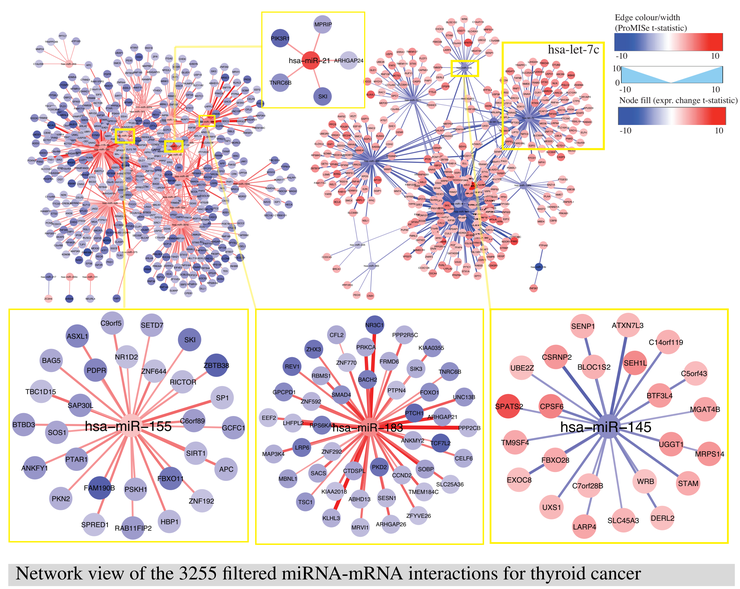
B. THCA



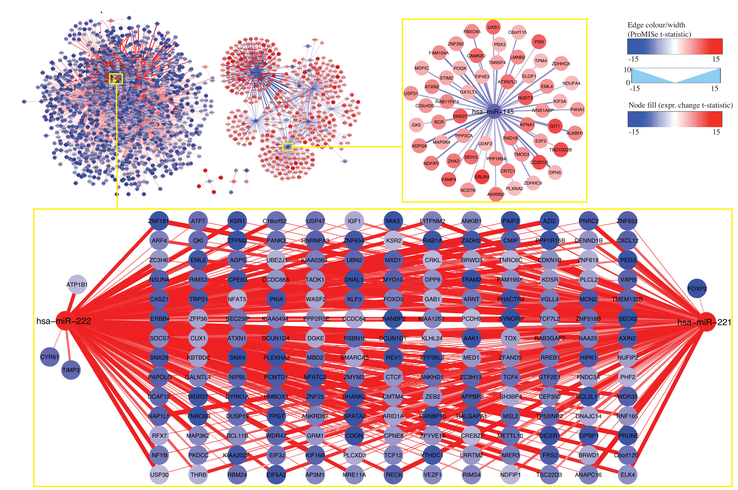
Heatmaps of BRCA-related genes/miRNA involved in signifiant interactions and expression changes; purple boxes highlight interactions with coherent expression changes



Network view of the 1257 filtered miRNA-mRNA interactions for breast cancer



Network view of the 3255 filtered miRNA-mRNA interactions for thyroid cancer



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 available at Bioconductor website

References

Khan, A. A., *et al.*(2009). Transfection of small RNAs globally perturbs gene regulation by endogenous microRNAs. *Nature Biotechnology*, **27**(6), 549-555.

Arvey, A., *et al.*(2010). Target mRNA abundance dilutes microRNA and siRNA activity. *Molecular systems biology*, **6**, 1-7.

Tsang, J., *et al.*(2007). MicroRNA-mediated feedback and feedforward loops are recurrent network motifs in mammals. *Molecular Cell*, **26**(5), 753-767.

Huang, J. C., *et al.*(2007). Using expression profiling data to identify human microRNA targets. *Nature Methods*, **4**(12), 1045-1049.

Lu, Y., *et al.*(2011). A Lasso regression model for the construction of microRNA-target regulatory networks. *Bioinformatics*, **27**(17), 2406-2413.

Hafner, M., *et al.*(2010). Transcriptome-wide Identification of RNA-Binding Protein and MicroRNA Target Sites by PAR-CLIP. *Cell*, **141**(1), 129-141.

Cancer Genome Atlas Research Network. (2008). Comprehensive genomic characterization defines human glioblastoma genes and core pathways. *Nature*, **455**(7216), 1061-1068.

Li, Y., *et al.*(2014). Inferring probabilistic miRNA-mRNA interaction signatures in cancers: a role-switch approach. ***Nucleic Acids Research*****, 42(9), e76.**