University of Central Florida

Computational methods for miRNA binding interactions

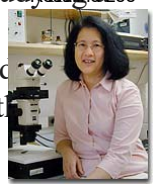
Funded by NSF

Small ncRNAs

- **Two flavors of small non-coding RNA:**
  1. micro RNA (miRNA)
  2. short interfering RNA (siRNA)
- **Properties of small non-coding RNA:**
  - Involved in silencing other mRNA transcripts.
  - Called "small" because they are usually only about 21-24 nucleotides long.
  - Synthesized by first cutting up longer precursor sequences (like the 61nt one that Lee discovered).
  - Silence an mRNA by base pairing with some sequence on the mRNA.

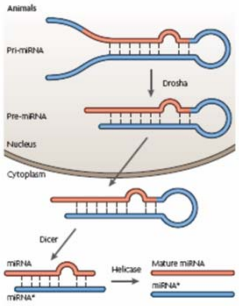
Discovery of small RNAs

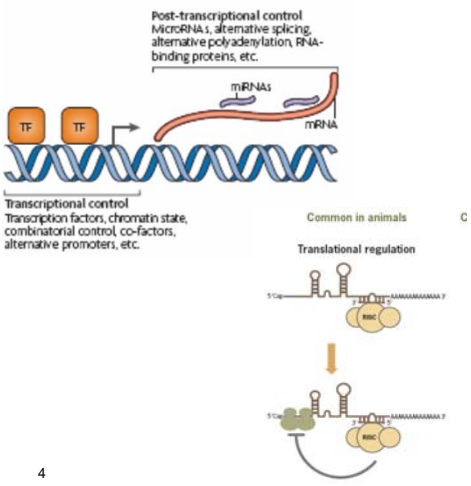
- **The first small RNA:**
  - In 1993 Rosalind Lee (Victor Ambros lab) was studying a non-coding gene in *C. elegans*, lin-4, that was involved in silencing of another gene, lin-14, at the appropriate time in the development of the worm *C. elegans*.



  - Two small transcripts of lin-4 (22nt and 61nt) were found to be complementary to a sequence in the 3' UTR of lin-14.
  - Because lin-4 encoded no protein, she deduced that it must be these transcripts that are causing the silencing by RNA-RNA interactions.
  - The second small RNA wasn't discovered until 2000!

MicroRNA

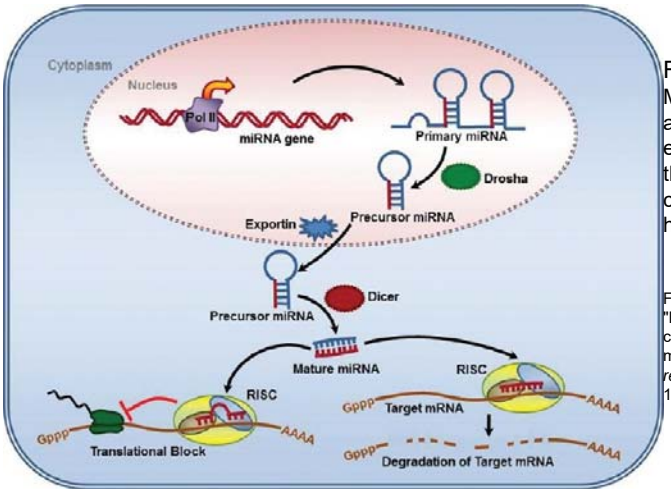




http://cs273a.stanford.edu [Bejerano Fall10/11]

4

miRNA-mediated post-transcriptional regulation

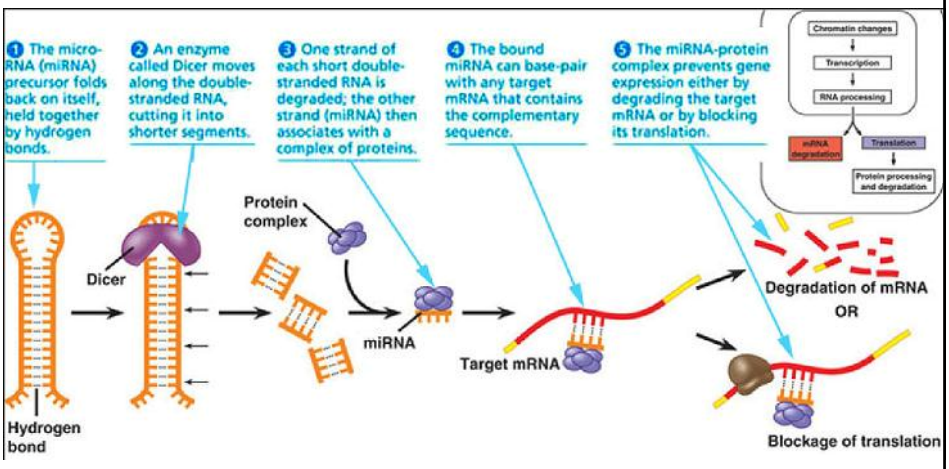


Facts:
MicroRNAs (miRNAs) appear to regulate the expression of more than 60% of Protein coding genes of the human genome

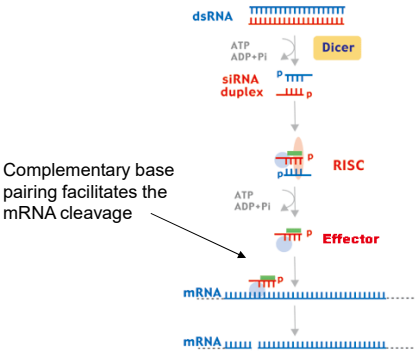
Friedman, Robin C., et al. "Most mammalian mRNAs are conserved targets of microRNAs." *Genome research* 19.1 (2009): 92-105.S

Joshi, Sachindra R., et al. "MicroRNAs-control of essential genes: Implications for pulmonary vascular disease." Pulmonary circulation 1.3 (2011): 357.

miRNA Pathway Illustration



siRNA Pathway Illustration



Complementary base pairing facilitates the mRNA cleavage

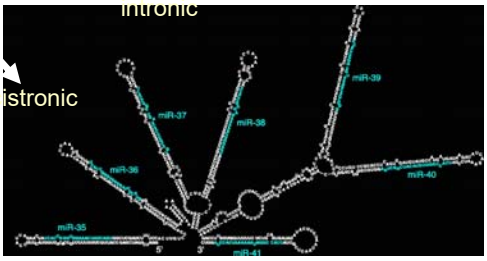
known miRNAs in human

intergenic

intronic

polycistronic

monocistronic



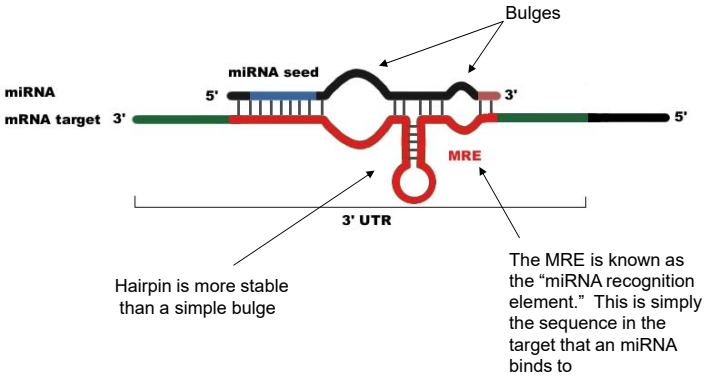
Features of miRNAs

- **Hundreds miRNA genes are already identified in human genome.**
- **Most miRNAs start with a U**
- **The second 7-mer on the 5' end is known as the "seed."**
  - When an miRNAs bind to their targets, the seed sequence has perfect or near-perfect alignment to some part of the target sequence.
  - **Example**: UGAGCUUAGCAG . . .

Features of miRNAs

- **Many miRNAs are conserved across species:**
  - For half of known human miRNAs, >18% of all occurrences of one of these miRNA seeds are conserved among human, dog, rat, and mouse.
- **As a rule, the full sequence of miRNAs is almost never completely complementary to the target sequence.**
  - Common to see a loop or bulge after the seed when binding.
  - Loop/bulge is often a hairpin because of stability.

miRNA Binding



Locating miRNA Genes: Experimentally

- ?

Locating miRNA Genes: Experimentally

- **Locating miRNA experimentally is difficult.**

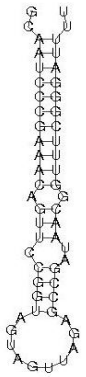
- **Procedure:**
  1. Find a gene that causes down-regulation of another gene.
  2. Determine if no protein is encoded.
  3. Analyze the sequence to determine if it is complementary to its target.

Locating miRNA Genes: Comparative Genomics

- **Idea: Find the seed binding sites.**
  1. Examine well-conserved 3' UTRs among species to find well-conserved 8-mers (A + seed) that might be an miRNA target sequence.
  2. Look for a sequence complementary to this 8-mer to identify a potential miRNA seed. Once found, check flanking sequence to see if any stable hairpin structures can form—these are potentially pre-miRNAs.
  3. To determine the possibility of secondary RNA structure, use RNAfold.

Locating miRNA Genes: Example

- **Suppose you found a well-conserved 8-mer in 3' UTRs (this could be where an miRNA seed binds in its target).**
  - **Example**: AGACTAGG
- **Look elsewhere in genome for complementary sequence (this could be an miRNA seed).**
  - **Example**: TCTGATCC
- **When TCTGATCC is found, check to see (with RNAfold) if the sequences around it could form hairpin; if so, this could be an miRNA gene.**



Finding miRNA Targets: Method 1

- **Now we know of some miRNAs, but where do they attack?**

Finding miRNA Targets: Method 1

- **Now we know of some miRNAs, but where do they attack?**
- **Goal: Find the targets of a set of miRNAs that are shared between human and mouse.**
  - Looking for the miRNA recognition element (MRE), not whole mRNA. This is just the part that the miRNA would bind to.
- **Basic Assumption: Whole miRNA:MRE interactions (binding) are likely to have highly energetically favorable base pairing.**
- **Basic Method: Look through the conserved 3' UTRs—this is where the MREs are most likely to be located—and try to make an alignment that minimizes the binding energy between the miRNA sequence and the UTRs (most favorable).**

Finding miRNA Targets: Method 1

- **Method:**
  - First look at the binding energies of all 38-mers of the mRNA when binding to the miRNA. Subsequently apply several filters to pick alignments that "look" like miRNA binding.
  - Why 38-mers? ~22 nt for the miRNA and the rest to allow for bulges, loops, etc.
- **Algorithm: Use a modified dynamic programming sequence alignment algorithm to calculate the binding energies for each 38-mer.**
- **Modifications: Scoring and speedup**

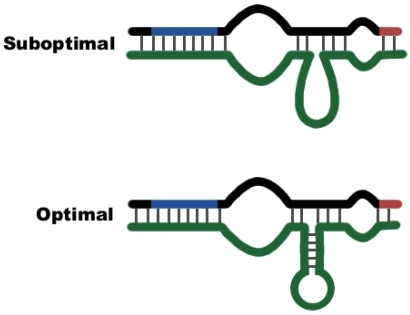
Finding miRNA Targets: Method 1

- **Scoring:**
  - Mismatches and indels allowed.
  - Matrix based on RNA-RNA binding energies.
    - Use known binding energies of Watson-Crick pairing and wobble (G-U) pairing.
  - Binding energy (score) calculated for every two adjacent pairings (unlike the standard alignment algorithm which just takes into account the "score" for one pair at a time).
    - Adds dimensions to scoring matrix.
    - Adds complexity to recurrence relation.

Finding miRNA targets: Method 2

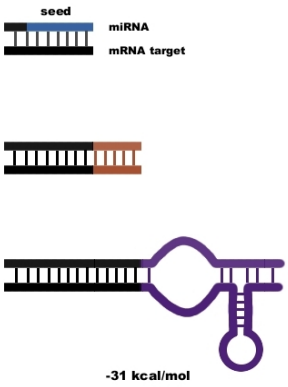
- **Goal: Find the set of miRNA targets for miRNAs shared across multiple species**
  - Trying to identify which genes have 3' UTRs are attacked by miRNAs
- **Basic Assumptions:**
  1. There is perfect binding to the miRNA seed.
  2. Any leftover sequence wants to achieve optimal RNA secondary structure.
- **Basic Method: For each species' set of 3' UTRs, find sites where there is perfect binding of the miRNA seed and "optimal folding" nearby. Look for agreement among all the species.**

Method 2: Example



Method 2: Steps

1. Find a perfect match to the miRNA seed.
2. Extend the matching region if possible.
3. Find the optimal folding for the remaining sequences.
4. Calculate the energy of this interaction.



Method 2: Details

- **Input: A set of miRNAs conserved among species and a set of 3' UTR sequences for those species.**

- **Method: For each organism:**
  1. Find all occurrences in the UTR sequences that match the miRNA seed exactly.
  2. Extend this region with perfect or wobble pairings.
  3. With the remaining sequence of the miRNA, use the program RNAfold to find optimal folding with the next 35 bases of

Method 2: Details

- **Method Cont.:**
  5. Sum up the scores of all interactions for each UTR.
  6. Rank all the organism's gene's UTRs by this score (sum of all interactions in that UTR).
  7. Repeat the above steps for each organism.
  8. Create a cutoff score and a cutoff rank for the UTRs.
  9. Select the set of genes where the orthologous genes across all the sampled species have UTR's that score and rank above this cutoff.

Method 2: Details

- **Verification:**
  - Find the number of predicted binding sites per miRNA.
  - Compare it to number of binding sites for a randomly generated miRNA.
  - The result is much higher.

Analysis of the Two Methods

- **Method 1:**
  - Good at identifying very strong, highly complementary miRNA targets.
  - Found gene targets with one miRNA binding site, failed to identify genes with multiple weaker binding sites.
- **Method 2:**
  - Good at identifying gene targets that have many weaker interactions.
  - Fails to identify single-site genes.

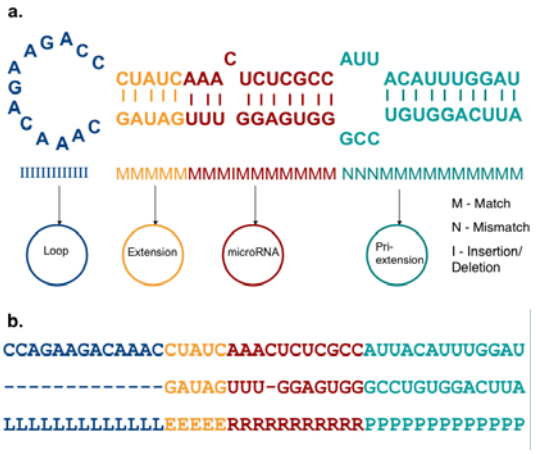
Analysis of the Two Methods

- **Both Methods:**
  - Speed is an issue.
  - Won't find targets that aren't in the 3' UTR of a gene.
- **We need more species sequenced!**
  - Conserved sequences are used to discover small RNAs.
  - Conserved small RNAs are used to discover targets.
  - Confidence in prediction of small RNAs and targets.
  - Allows for broader scope with different combinations of species.

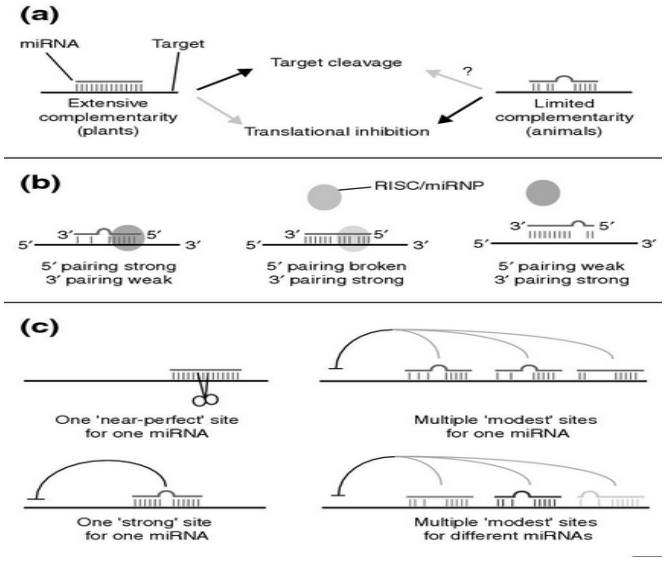
Results

- **Predicted a large portion of already known targets and provided direction for identifying undiscovered targets.**
- **Found that it is more common that genes are regulated by multiple small RNAs.**
- **Found that many small RNAs have multiple targets.**

HHMMiR: hierarchical HMMs for miR (Kadri et al 2009)



Limited understanding of microRNA targeting



Lai, Eric C. "Predicting and validating microRNA targets." *Genome biology* 5 (2004): 2004-5.

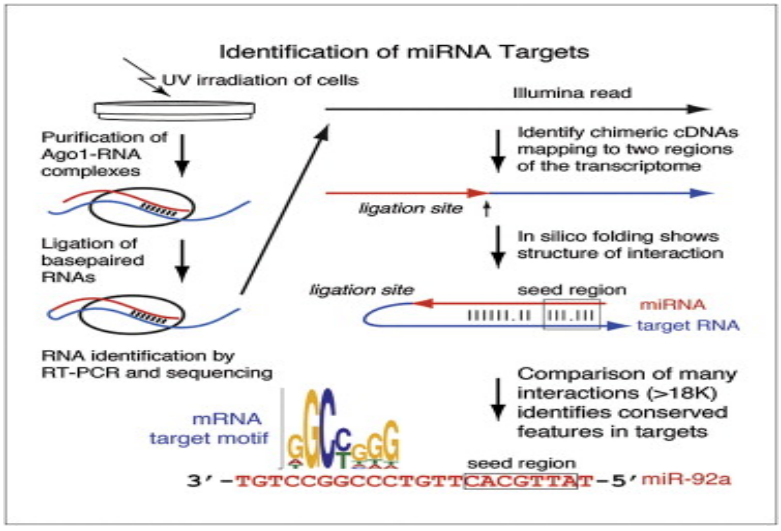
miRNA binds to 3'UTR to regulate the mRNA expression ?

A few years ago:
We think:
May be it's like this

Now:
NO! miRNA not only
Binds to 3'UTR
but also CDS region

Hausser, Jean, et al.
"Analysis of CDS-located miRNA target sites suggests that they can effectively inhibit translation."
Genome research 23.4 (2013): 604-615.

CLASH technique



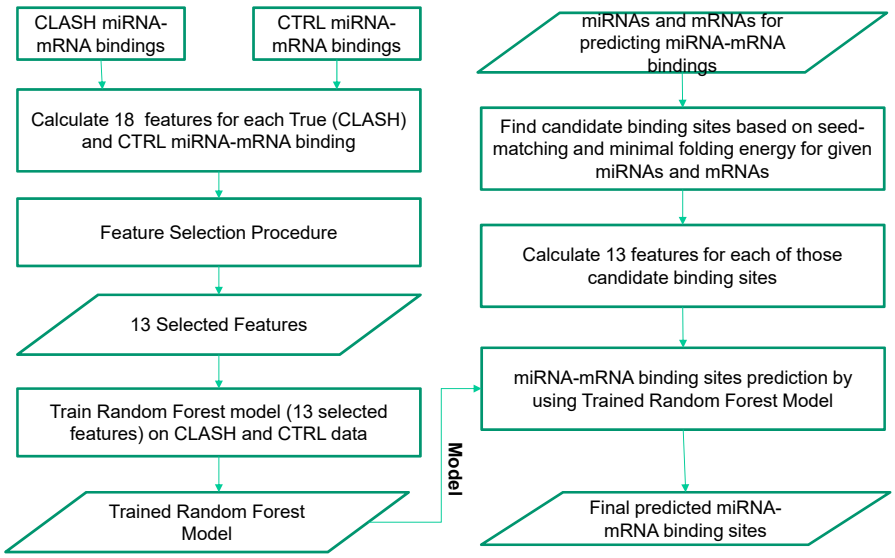
Helwak, Aleksandra, et al. "Mapping the human miRNA interactome by CLASH reveals frequent noncanonical binding." *Cell* 153.3 (2013): 654-665.

New discoveries from CLASH experiment

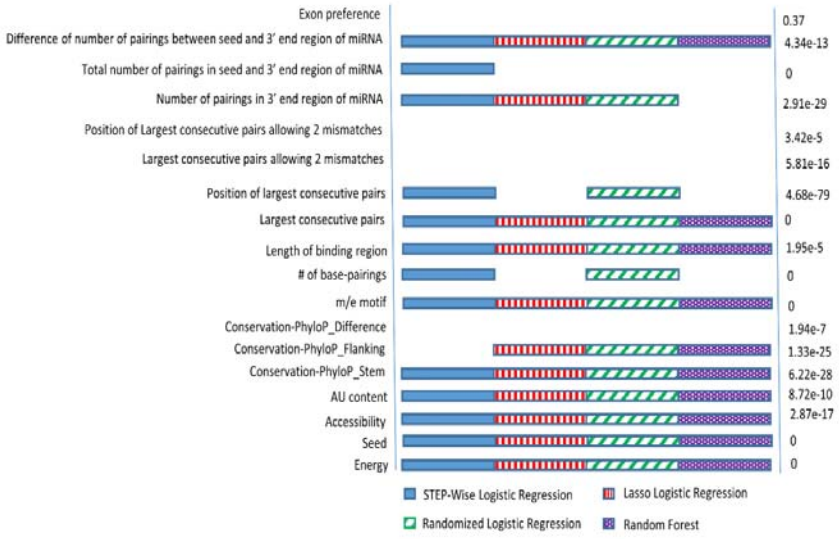
Many new discoveries from the CLASH experiments:

- miRNAs show frequent non-canonical targeting of mRNAs
- Non-seed interactions are common and functional
- And more?

TarPmiR



Feature selection based on CLASH data



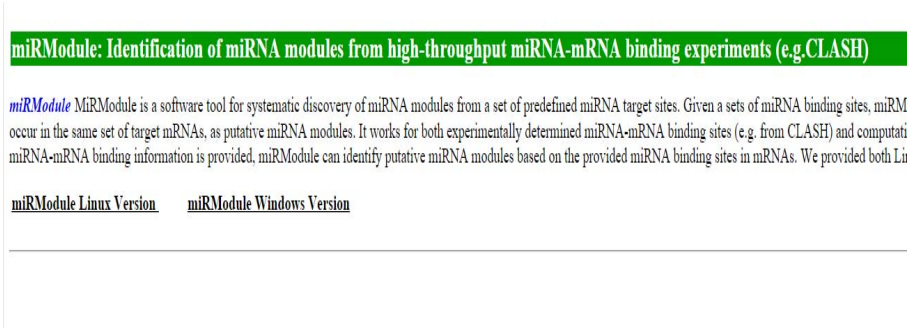
Performance-CLASH dataset

Method	TP	FN	FP	Recall TP/(TP+FN)	Precision TP/(TP+FP)	F-score
TarPmiR	4695	3819	19950	0.551	0.191	0.284
miRanda	3852	4662	51849	0.452	0.069	0.120
TargetScan	1164	7350	10281	0.136	0.101	0.116
Mirmap	1821	6693	30746	0.214	0.056	0.089

Performance-Independent datasets

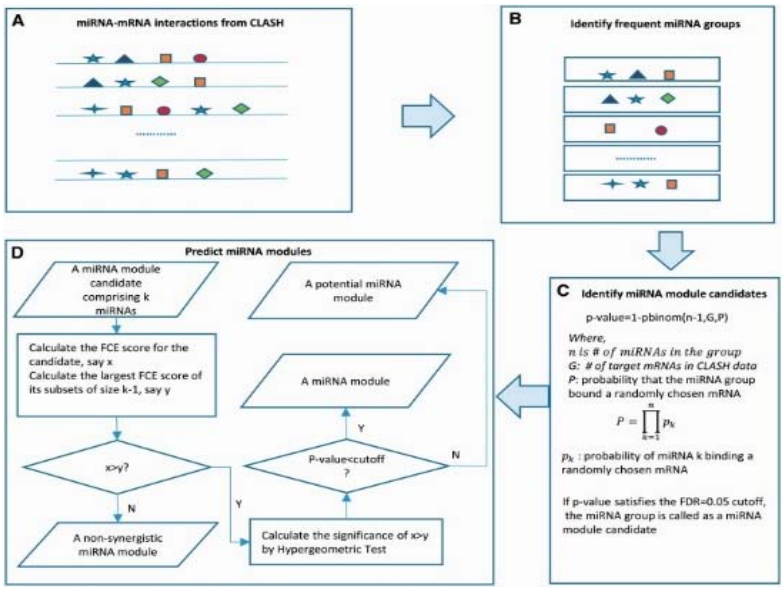
Data set	# of miRNAs input	Performance measurement	TarPmiR	miRanda	TargetScan	miRmap
I	60	# of predictions	240605	246311	219304	504447
		% of correct predictions	11904/16041=74.2%	7061/16041=44.0%	6248/16041=39.0%	7121/16041=44.4%
		Recall	0.742	0.440	0.390	0.444
		Precision	0.0495	0.0287	0.0285	0.014
	120	# of predictions	481135	476827	461280	906654
		% of correct predictions	13846/16041=86.3%	9683/16041=60.4%	8969/16041=55.9%	10342/16041=64.5%
		Recall	0.863	0.604	0.559	0.645
		Precision	0.0288	0.0203	0.0194	0.0114
II	60	# of predictions	469752	453880	437791	971238
		% of correct predictions	34301/43251 =79.3%	20378/43251 =47.1%	17556/43251 =40.6%	20543/43251 =47.5%
		Recall	0.793	0.471	0.406	0.475
		Precision	0.0730	0.0449	0.0401	0.0211
	120	# of predictions	961112	902611	922373	1952258
		% of correct predictions	38821/43251= 89.8%	23762/43251= 54.9%	24578/43251= 56.8%	25667/43251= 59.3%
		Recall	0.898	0.549	0.568	0.593
		Precision	0.0403	0.0263	0.0266	0.0131
III	119	# of predictions	285491	439485	875442	341773
		% of correct predictions	10766/11080=97.2%	9069/11080=81.8%	10084/11080=91.0%	7840/11080=70.8%
		Recall	0.972	0.818	0.910	0.708
		Precision	0.0377	0.0206	0.0115	0.0229
IV	50	# of predicted interactions	102324	87462	69184	113211
		% of correct predictions	5515/15386=35.8%	4335/15386=28.2%	3821/15386=24.8%	4531/15386=29.4%
		Recall	0.358	0.282	0.248	0.294
		Precision	0.0539	0.0496	0.0552	0.0400
	100	# of predicted interactions	412149	337863	286667	413213
		% of correct predictions	8127/15562=52.2%	6442/15562=41.4%	5644/15562=36.3%	5323/15562=34.2%
		Recall	0.522	0.414	0.363	0.342
		Precision	0.0197	0.0191	0.0197	0.0129

mirModule



Jun Ding, Xiaoman Li, and Haiyan Hu. "MicroRNA modules prefer to bind weak and unconventional target sites." *Bioinformatics* (2014): btu833.

miRModule Pipeline



Synergistic miRNAs tend to bind weaker sites