Computational Prediction of microRNA Targets in Animals

Xiao Fan and Lukasz Kurgan

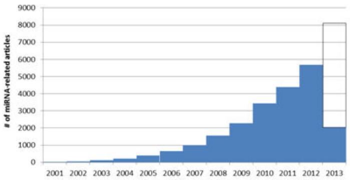
Outline

- Introduction
  - microRNA (miRNA) and miRNA targets
  - overview of state-of-the-art in target prediction
- Review of miRNA target predictors
  - scope
  - usability (popularity/impact), methodologies, and evaluation
- Empirical evaluation
  - setup
  - new consensus
  - results
- Conclusions

Introduction

- Noncoding small RNAs
  - 19-23 nucleotides
  - binding usually results in mRNA degradation or translational repression of proteins

```
hsa-miR-16 : NM_001667
miRNA  3' GCGGUUAUAAAUGCACGACGAU 5'
          |  |   |      ||||||||
RNA    5' CACCCC-AGCCUGCUGCUGCUA 3'
```



- Used
  - to study signal transduction and pathogenesis of various diseases: cancer and genetic, neurodegenerative, and metabolic diseases
  - for target validation and lead optimization in preclinical drug development

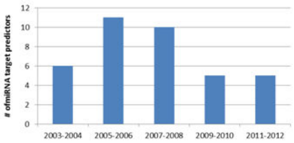
---

Reviews of miRNA targets predictors

review	Analytical description						Empirical evaluation			
	# methods	Latest method	Method-ology	Evalu-ation	Avail-ability	Usability /Impact	Dataset size	Inclusion of TN	# methods	Ensemble
This talk	**37**	**2012**	✓	✓	✓	✓	**236**	✓	**8 (2012)**	✓
Bioinformation, 2012	11	2007	✓							
Curr. Mol. Med. 2011	11	2007	✓							
New Biotechnology 2010	7	2009	✓		✓	O				
Exp. Mol. Med. 2010	30	2008	✓		✓	O				
Mamm. Genome 2010	10	2008	✓		✓	O				
Bioinformatics 2009	10	2007	✓		✓		150	✓	7 (2007)	✓
FEBS J. 2009	8	2008	✓							
Nucleic Acids Res. 2009	14	2008	✓							
Cell 2009	9	2008	✓		✓					
Methods Enzymol. 2007	9	2006	✓		✓	O				
Drug Discov. Today 2007	14	2006	✓		✓					
DNA Cell Biol. 2007	9	2006	✓		✓					
Nat. Methods 2006	20	2006	✓		✓	O	84		5 (2005)	✓
Comput. Biol. Chem. 2006	10	2005	✓		✓	O				
Nat. Genet. 2006	11	2006			✓	O	113	✓	9 (2005)	

Scope of this review

- 37 miRNA target predictors in animals
- Analysis from all key perspectives including methodology, evaluation, usability and popularity/impact



- Empirical, comparative analysis
  - created two benchmark databases at the duplex and gene levels utilizing recent experimental data
  - comprehensive assessment of 8 representative methods including 2 latest predictors
  - developed a new, simple and effective consensus method

Scope of this review

- 37 miRNA target predictors in animals

standalone(s), web server (ws),
pre-computated (p), upon request (ur)

Predictors	Availability	
	type	URL
Stark *et al.*	✗	✗
TargetScan	s, ws	http://www.targetscan.org/
DIANA-microT	ws	http:http://diana.pcbi.upenn.edu/DIANA-microT
RNAhybrid	s, ws	http://bibiserv.techfak.uni-bielefeld.de/rnahybrid/
miRanda	s, ws, p	http://www.microrna.org/microrna/home.do
Rajewsky's	✗	✗
TargetScanS	s, ws	http://genes.mit.edu/tscan/targetscanS2005.html
Robins	✗	✗
Xie *et al.*	✗	✗
PicTar	ws, p	http://pictar.mdc-berlin.de/
MovingTarget	ur	✗
MicroInspector	ws	http://bioinfo.uni-plovdiv.bg/microinspector/
TargetBoost	demo	http://www.interagon.com/demos.html
EMBL	p	http://mirnas.russelllab.org/
miTarget	✗	http://cbit.snu.ac.kr/miTarget
RNA22	ws, p	https://cm.jefferson.edu/rna22v1.0/
MicroTar	s	http://tiger.dbs.nus.edu.sg/microtar/
EIMMo	ws, p	http://www.mirz.unibas.ch/ElMMo3/
STarMir	ws	http://sfold.wadsworth.org/cgi-bin/starmir.pl
Yan *et al.*	✗	http://www.biosino.org/~kanghu/mRTP/mRTP.html
PITA	s, ws	http://genie.weizmann.ac.il/pubs/mir07/index.html
NbmiRTar	ws	http://wotan.wistar.upenn.edu/NBmiRTar
TargetRank	ws	http://hollywood.mit.edu/targetrank/
MirTarget2	ws, p	http://mirdb.org/miRDB/
mirWIP	ws	http://146.189.76.171/query.php
HuMiTar	ur	✗
mirTif	ws	http://bsal.ym.edu.tw/mirtif/
TargetMiner	ws, p	http://www.isical.ac.in/~bioinfo_miu/targetminer20.htm
TargetSpy	s, ws	http://www.targetspy.org/
Mtar	✗	✗
SVMicrO	s	http://compgenomics.utsa.edu/svmicro.html
RepTar	s, ws, p	http://bioinformatics.ekmd.huji.ac.il/reptar/
PACMIT	✗	✗
MultiMiTar	s, ws	http://www.isical.ac.in/~bioinfo_miu/multimitar.htm
miREE	s	http://didattica-online.polito.it/eda/miREE/
miRcode	ws, p	http://www.mircode.org
miRmap	s, ws, p	http://mirmap.ezlab.org/

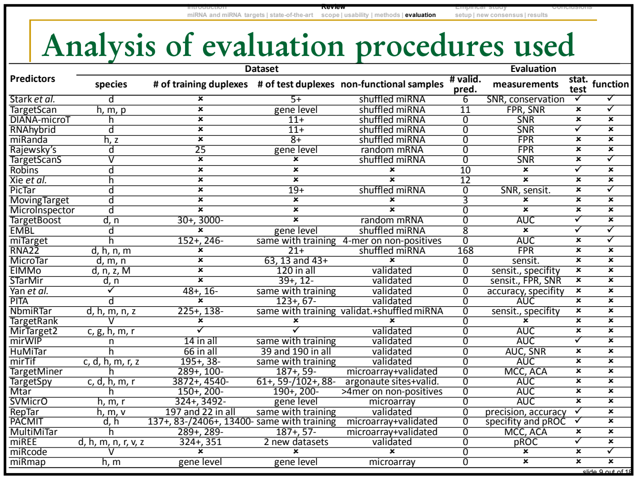
Analysis of usability

anopheles(a), chicken(c),
drosophila(d), fungi(f), dog(g),
human(h), mouse(m), nematode(n),
rat(r), zebra fish(z), primate(P),
mammal(M), vertebrate(V)

Predictors	Ease to use					Impact/popularity		
	# of parameters	species	# of targets	form of gene	new miRNA	review highlighted	review considered	# citations per year
Stark *et al.*	~	d	~	~	✗	0	4	35.0
TargetScan	3	d, h, m, n, z	a few	name	✓	3	13	392.5
DIANA-microT	1	d, h, m, n, r, t	a few	name	✓	2	13	41.9
RNAhybrid	8	any	dozens	sequence	✓	2	11	56.9
miRanda	0	d, h, m, r, n	thousands*	none	✗	0	14	48.3
Rajewsky's	~	d	~	~	~	0	2	18.6
TargetScanS	0	d, m, n, other V	hundreds*	name	✗	4	10	392.5
Robins	~	~	~	~	~	0	2	14.8
Xie *et al.*	~	~	~	~	~	0	2	124.9
PicTar	0	d, h, m, n	dozens	name	✗	1	15	25.8
MovingTarget	~	~	~	~	~	0	5	6.1
MicroInspector	2	any	a few	both	✓	0	3	12.4
TargetBoost	~	n	~	~	~	0	7	7.1
EMBL	~	~	hundreds*	~	~	1	5	64.3
miTarget	~	~	~	~	~	0	8	10.4
RNA22	4	d, h, m, n	thousands*	name	✗	0	12	69.6
MicroTar	0	any	~	~	✓	0	3	4.9
EIMMo	0	d, h, m, r, n, z	a few	name	✗	1	5	17.3
STarMir	0	h, m	dozens	both	✓	0	1	27.7
Yan *et al.*	~	~	~	~	~	0	2	3.5
PITA	5	d, h, m, n	a few	both	✓	1	5	75.8
NbmiRTar	3	any	a few	sequence	✓	0	3	0.0
TargetRank	0	h, m	hundreds*	none	✗	0	1	20.7
MirTarget2	0	c, g, h, m, r	a few	name	✗	0	3	23.4
mirWIP	1	n	a few	name	✗	0	3	16.4
HuMiTar	3	h	~	sequence	✓	0	1	2.4
mirTif	0	any	1	sequence	✓	0	1	4.0
TargetMiner	0	h	a few	name	✗	0	1	8.0
TargetSpy	2	c, d, h, m, r	a few	name	✗	0	0	7.0
Mtar	~	~	~	~	~	0	0	3.7
SVMicrO	~	h, m, r	~	~	~	0	0	2.7
RepTar	7	h, m	a few	name	✗	0	0	2.0
PACMIT	~	~	~	~	~	0	0	9.0
MultiMiTar	0	h	a few	both	✓	0	0	1.0
miREE	0	d, h, m, n, r, z	dozens	both	✓	0	0	1.0
miRcode	3	h	a few	name	✗	0	0	6.0
miRmap	4	c, e, h, m, o, r, w, z	a few	both	✓	0	0	0.0

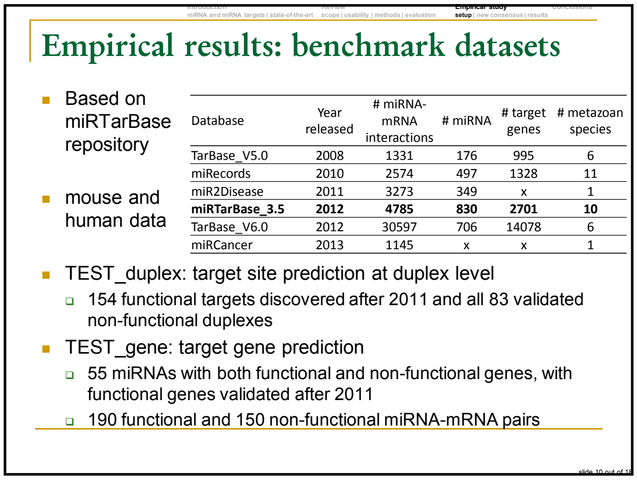
Analysis of methodologies used

Predictors	Model type	Target region	Complementarity		Site accessibility		Conservation	Multiple sites	Feature	
			seed	non seed	free energy	AU %			#	selection
Stark *et al.*	screening	3'UTR	1-8	miRNA size+5	mFold	✗	a, d	✓	✗	✗
TargetScan	score	3'UTR	7mer-m8	to 1st mismatch	RNAeval	✓	m, r, z	✗	✗	✗
DIANA-microT	score	3'UTR	✗	38 nt	✓	✗	m	✓	✗	✗
RNAhybrid	score	3'UTR, CDS	6mer	✗	RNAcalibrate	✗	a, d	✓	✗	✗
miRanda	score	3'UTR, CDS	7mer-m8	✗	✓	✗	f, m, r	✓	✗	✗
Rajewsky's	score	3'UTR	1-8	✗	mFold	✗	d	✗	✗	✗
TargetScanS	score	3'UTR, ORFs	6mer	✗	✗	✗	c, g, h, m, r	✗	✗	✗
Robins	score	3'UTR	2-8	✗	✓	✗	✗	✗	✗	✗
Xie *et al.*	score	promoters and 3'UTR	8mer	✗	✗	✗	g, h, m, r	✓	✗	✗
PicTar	score	3'UTR	7mer-A1, 7mer-m8	remaining	✓	✗	d	✗	✗	✗
MovingTarget	screening	3'UTR	1-8	50 nt	DINAMelt	✗	d	✓	✗	✗
MicroInspector	score	3'UTR	7mer-A1, 7mer-m8	✗	✓	✗	✗	✗	✗	✗
TargetBoost	GP	3'UTR	pattern	30 nt	GPboost	✗	✗	✗	✓	✗
EMBL	score	3'UTR	6mer	10th nt to end	✓	✗	d	✓	✗	✗
miTarget	SVM	3'UTR	2-7	20 nt	✓	✓	✗	✗	15	wrapper
RNA22	score	3'UTR, CDS	✗	pattern	✗	✗	✗	✓	✗	✗
MicroTar	score	3'UTR	7mer-A1, 7mer-m8	✗	Vienna RNA	✗	✗	✓	✗	✗
EIMMo	Bayesian	3'UTR	7mer-A1, 7mer-m8	✗	✗	✗	✓	✗	✓	✗
STarMir	score	3'UTR, CDS, 5'UTR	✗	about miRNA size	sFold	✗	✗	✗	✗	✗
Yan *et al.*	SVM	3'UTR	1-8	✗	Vienna RNA	✗	✗	✗	48	filter
PITA	score	3'UTR	6mer	✗	RNAfold	✗	✗	✓	✗	✗
NbmiRTar	Bayesian	3'UTR	7mer-A1, 7mer-m8	remaining	Vienna RNA	✗	d	✗	57	✗
TargetRank	score	3'UTR	6mer	✗	✗	✓	✓	✓	✗	✗
MirTarget2	SVM	3'UTR	6mer	✗	RNAfold	✗	✗	✗	6	filter
mirWIP	score	3'UTR	8mer	✗	mFold	✗	✗	✓	✗	✗
HuMiTar	score	3'UTR	6mer	9-13, 14-20 nt	✗	✗	✗	✗	✗	✗
mirTif	SVM	3'UTR	1-8	remaining	✗	✗	✗	✗	25	filter
TargetMiner	SVM	3'UTR	6mer	13-16 nt	✓	✓	✓	✓	30	filter
TargetSpy	DS	3'UTR	✗	all	RNAduplex	✗	✗	✗	7	filter
Mtar	ANN	3'UTR, CDS, 5'UTR	6mer	remaining	RNAfold	✗	✗	✗	16	✗
SVMicrO	SVM	3'UTR	5 patterns	remaining	miRNAbind	✓	✓	✗	39	wrapper
RepTar	screening	3'UTR	6mer	remaining	RNAcofold	✗	✓	✓	✗	✗
PACMIT	screening	3'UTR	✗	remaining	RNAfold	✗	✗	✓	✗	✗
MultiMiTar	SVM	3'UTR	6mer	13-16 nt	✓	✓	✗	✓	39	filter
miREE	SVM	3'UTR	1-8	13-16nt, remaing	RNAfold	✓	✗	✗	25	filter
miRcode	screening	3'UTR, CDS, 5'UTR	7mer-A1, 7mer-m8	✗	✗	✗	P, M, other V	✓	✗	✗
miRmap	regression	3'UTR	6mer	remaining	Vienna RNA	✓	M	✓	12	filter



Analysis of evaluation procedures used

Predictors	Dataset					Evaluation		
	species	# of training duplexes	# of test duplexes	non-functional samples	# valid. pred.	measurements	stat. test	function
Stark *et al.*	d	×	5+	shuffled miRNA	6	SNR, conservation	✓	✓
TargetScan	h, m, p	×	gene level	shuffled miRNA	11	FPR, SNR	×	✓
DIANA-microT	h	×	11+	shuffled miRNA	0	SNR	×	×
RNAhybrid	d	×	11+	shuffled miRNA	0	SNR	✓	×
miRanda	h, z	×	8+	shuffled miRNA	0	FPR	×	×
Rajewsky's	d	25	gene level	random mRNA	0	FPR	×	×
TargetScanS	V	×	×	shuffled miRNA	0	SNR	×	✓
Robins	d	×	×	×	10	×	✓	×
Xie *et al.*	h	×	×	×	12	×	×	×
PicTar	d	×	19+	shuffled miRNA	0	SNR, sensit.	×	✓
MovingTarget	d	×	×	×	3	×	×	×
MicroInspector	d	×	×	×	0	×	×	×
TargetBoost	d, n	30+, 3000-	×	random mRNA	0	AUC	✓	×
EMBL	d	×	gene level	shuffled miRNA	8	×	✓	✓
miTarget	h	152+, 246-	same with training	4-mer on non-positives	0	AUC	×	✓
RNA22	d, h, n, m	×	21+	shuffled miRNA	168	FPR	×	×
MicroTar	d, m, n	×	63, 13 and 43+	×	0	sensit.	×	×
EIMMo	d, n, z, M	×	120 in all	validated	0	sensit., specifity	×	×
STarMir	d, n	×	39+, 12-	validated	0	sensit., FPR, SNR	×	×
Yan *et al.*	✓	48+, 16-	same with training	validated	0	accuracy, specifity	×	×
PITA	d	×	123+, 67-	validated	0	AUC	×	×
NbmiRTar	d, h, m, n, z	225+, 138-	same with training	validat.+shuffled miRNA	0	sensit., specifity	×	×
TargetRank	V	×	×	×	0	×	×	×
MirTarget2	c, g, h, m, r	✓	✓	validated	0	AUC	×	×
mirWIP	n	14 in all	same with training	validated	0	AUC	✓	×
HuMiTar	h	66 in all	39 and 190 in all	validated	0	AUC, SNR	×	×
mirTif	c, d, h, m, r, z	195+, 38-	same with training	validated	0	AUC	×	×
TargetMiner	h	289+, 100-	187+, 59-	microarray+validated	0	MCC, ACA	×	×
TargetSpy	c, d, h, m, r	3872+, 4540-	61+, 59-/102+, 88-	argonaute sites+valid.	0	AUC	×	×
Mtar	h	150+, 200-	190+, 200-	>4mer on non-positives	0	AUC	×	×
SVMicrO	h, m, r	324+, 3492-	gene level	microarray	0	AUC	×	×
RepTar	h, m, v	197 and 22 in all	same with training	validated	0	precision, accuracy	✓	×
PACMIT	d, h	137+, 83-/2406+, 13400-	same with training	microarray+validated	0	specifity and pROC	✓	×
MultiMiTar	h	289+, 289-	187+, 57-	microarray+validated	0	MCC, ACA	×	×
miREE	d, h, m, n, r, v, z	324+, 351	2 new datasets	validated	0	pROC	✓	×
miRcode	V	×	×	×	0	×	×	✓
miRmap	h, m	gene level	gene level	microarray	0	×	×	×

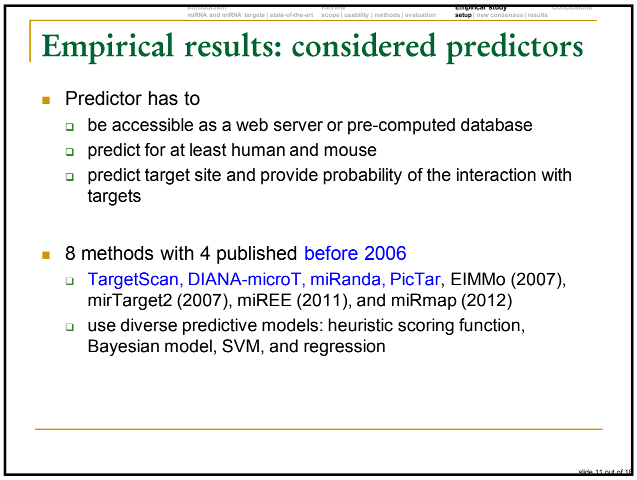


Empirical results: benchmark datasets

- Based on miRTarBase repository
- mouse and human data

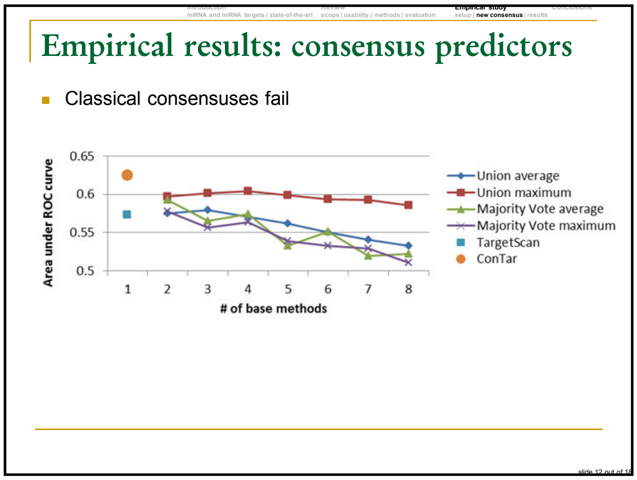
Database	Year released	# miRNA-mRNA interactions	# miRNA	# target genes	# metazoan species
TarBase_V5.0	2008	1331	176	995	6
miRecords	2010	2574	497	1328	11
miR2Disease	2011	3273	349	x	1
miRTarBase_3.5	**2012**	**4785**	**830**	**2701**	**10**
TarBase_V6.0	2012	30597	706	14078	6
miRCancer	2013	1145	x	x	1

- TEST_duplex: target site prediction at duplex level
  - 154 functional targets discovered after 2011 and all 83 validated non-functional duplexes
- TEST_gene: target gene prediction
  - 55 miRNAs with both functional and non-functional genes, with functional genes validated after 2011
  - 190 functional and 150 non-functional miRNA-mRNA pairs



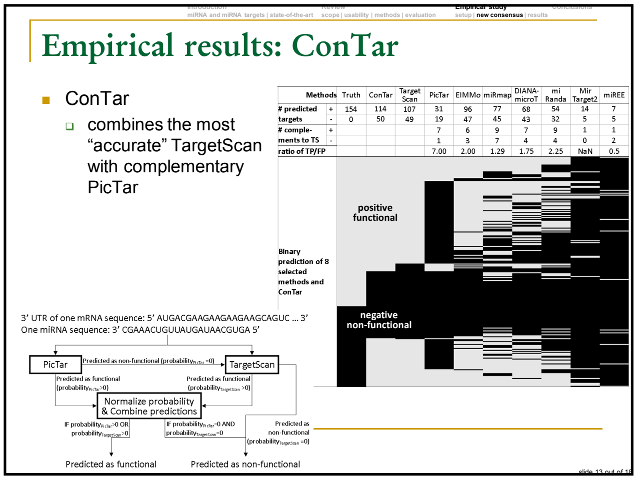
Empirical results: considered predictors

- Predictor has to
  - be accessible as a web server or pre-computed database
  - predict for at least human and mouse
  - predict target site and provide probability of the interaction with targets
- 8 methods with 4 published before 2006
  - TargetScan, DIANA-microT, miRanda, PicTar, EIMMo (2007), mirTarget2 (2007), miREE (2011), and miRmap (2012)
  - use diverse predictive models: heuristic scoring function, Bayesian model, SVM, and regression



Empirical results: consensus predictors

- Classical consensuses fail

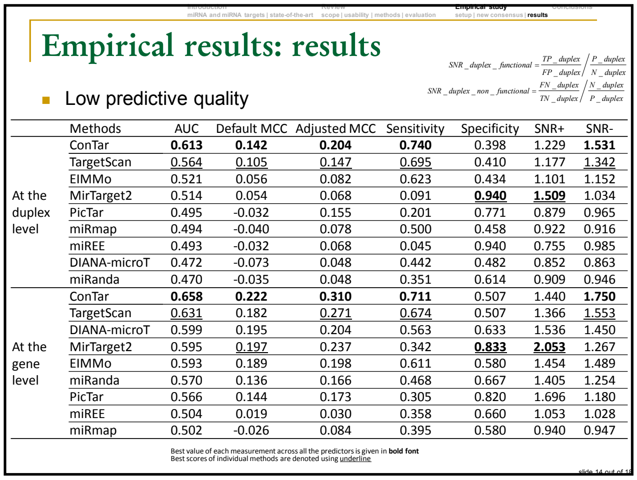


Empirical results: ConTar

- ConTar
  - combines the most "accurate" TargetScan with complementary PicTar

Methods		Truth	ConTar	Target Scan	PicTar	EIMMo	miRmap	DIANA-microT	mi Randa	Mir Target2	miREE
# predicted targets	+	154	114	107	31	96	77	68	54	14	7
	-	0	50	49	19	47	45	43	32	5	5
# complements to TS	+				7	6	9	7	9	1	1
	-				1	3	7	4	4	0	2
ratio of TP/FP					7.00	2.00	1.29	1.75	2.25	NaN	0.5

3' UTR of one mRNA sequence: 5' AUGACGAAGAAGAAGAAGCAGUC … 3'
One miRNA sequence: 3' CGAAACUGUUAUGAUAACGUGA 5'



Empirical results: results

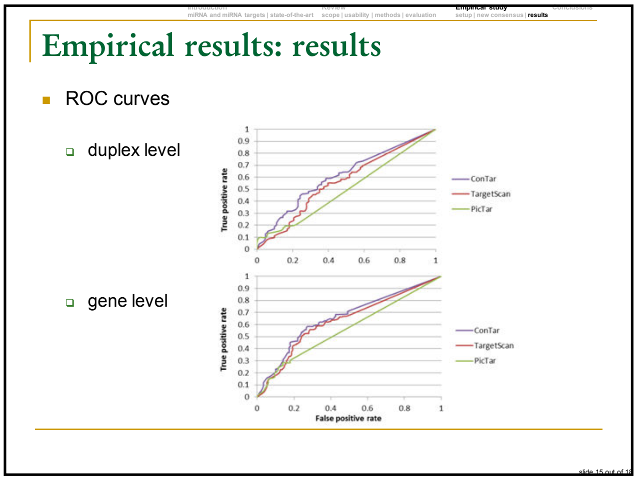
$$SNR\_duplex\_functional = \frac{TP\_duplex}{FP\_duplex} \Big/ \frac{P\_duplex}{N\_duplex}$$

$$SNR\_duplex\_non\_functional = \frac{FN\_duplex}{TN\_duplex} \Big/ \frac{N\_duplex}{P\_duplex}$$

- Low predictive quality

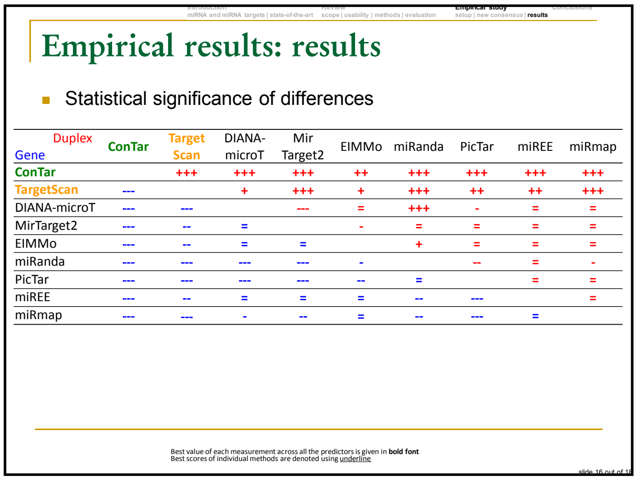
	Methods	AUC	Default MCC	Adjusted MCC	Sensitivity	Specificity	SNR+	SNR-
At the duplex level	ConTar	**0.613**	**0.142**	**0.204**	**0.740**	0.398	1.229	**1.531**
	TargetScan	0.564	0.105	0.147	0.695	0.410	1.177	1.342
	EIMMo	0.521	0.056	0.082	0.623	0.434	1.101	1.152
	MirTarget2	0.514	0.054	0.068	0.091	**0.940**	**1.509**	1.034
	PicTar	0.495	-0.032	0.155	0.201	0.771	0.879	0.965
	miRmap	0.494	-0.040	0.078	0.500	0.458	0.922	0.916
	miREE	0.493	-0.032	0.068	0.045	0.940	0.755	0.985
	DIANA-microT	0.472	-0.073	0.048	0.442	0.482	0.852	0.863
	miRanda	0.470	-0.035	0.048	0.351	0.614	0.909	0.946
At the gene level	ConTar	**0.658**	**0.222**	**0.310**	**0.711**	0.507	1.440	**1.750**
	TargetScan	0.631	0.182	0.271	0.674	0.507	1.366	1.553
	DIANA-microT	0.599	0.195	0.204	0.563	0.633	1.536	1.450
	MirTarget2	0.595	0.197	0.237	0.342	**0.833**	**2.053**	1.267
	EIMMo	0.593	0.189	0.198	0.611	0.580	1.454	1.489
	miRanda	0.570	0.136	0.166	0.468	0.667	1.405	1.254
	PicTar	0.566	0.144	0.173	0.305	0.820	1.696	1.180
	miREE	0.504	0.019	0.030	0.358	0.660	1.053	1.028
	miRmap	0.502	-0.026	0.084	0.395	0.580	0.940	0.947

Best value of each measurement across all the predictors is given in **bold font**
Best scores of individual methods are denoted using underline



Empirical results: results

- ROC curves
  - duplex level
  - gene level



Empirical results: results

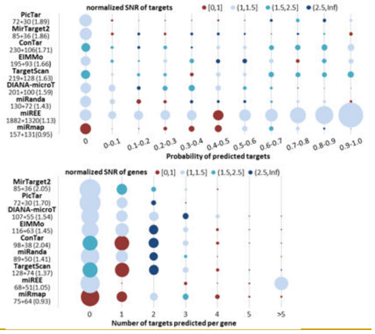
- Statistical significance of differences

Duplex Gene	**ConTar**	**Target Scan**	DIANA-microT	Mir Target2	EIMMo	miRanda	PicTar	miREE	miRmap
ConTar		+++	+++	+++	++	+++	+++	+++	+++
TargetScan	---		+	+++	+	+++	++	++	+++
DIANA-microT	---	---		---	=	+++	-	=	=
MirTarget2	---	--	=		-	=	=	=	=
EIMMo	---	--	=	=		+	=	=	=
miRanda	---	---	---	---	-		--	=	-
PicTar	---	---	---	---	--	=		=	=
miREE	---	--	=	=	=	--	---		=
miRmap	---	---	-	--	=	--	---	=	

Best value of each measurement across all the predictors is given in **bold font**
Best scores of individual methods are denoted using underline

Empirical results: results

- Target gene predictions
  - predictive quality in function of probability
  - predictive quality in function of number of predicted targets



Conclusions

- Low predictive quality
  - likely due to challenging set of nonfunctional duplexes that were hoped to be functional

Seed type	8mer	7-m8	7-A1	6mer	6bp	5bp	4bp	3bp	2bp	Sum
Functional	43	52	13	14	7	2	11	10	2	154
Non-functional	18	26	8	7	8	5	4	2	5	83
ratio	0.30	0.33	0.38	0.33	0.53	0.71	0.27	0.17	0.71	0.35

- No universally best predictor
  - TargetScan excels at the target site prediction, either binary or real-value
  - MirTarget2 has high specifity, useful to find small subset of accurate targets
  - PicTar has high TPR at low FPR, good results with high probability
  - number of functional genes predicted by TargetScan and EIMMo is closest to the native count; they provide good estimate the number of miRNA targets
- New methods that improve predictive quality are needed
  - “standard” consensuses do now work
  - consensuses should consider orthogonality of predictors