

MicroRNAs

Artemis G. Hatzigeorgiou

Assistant Professor

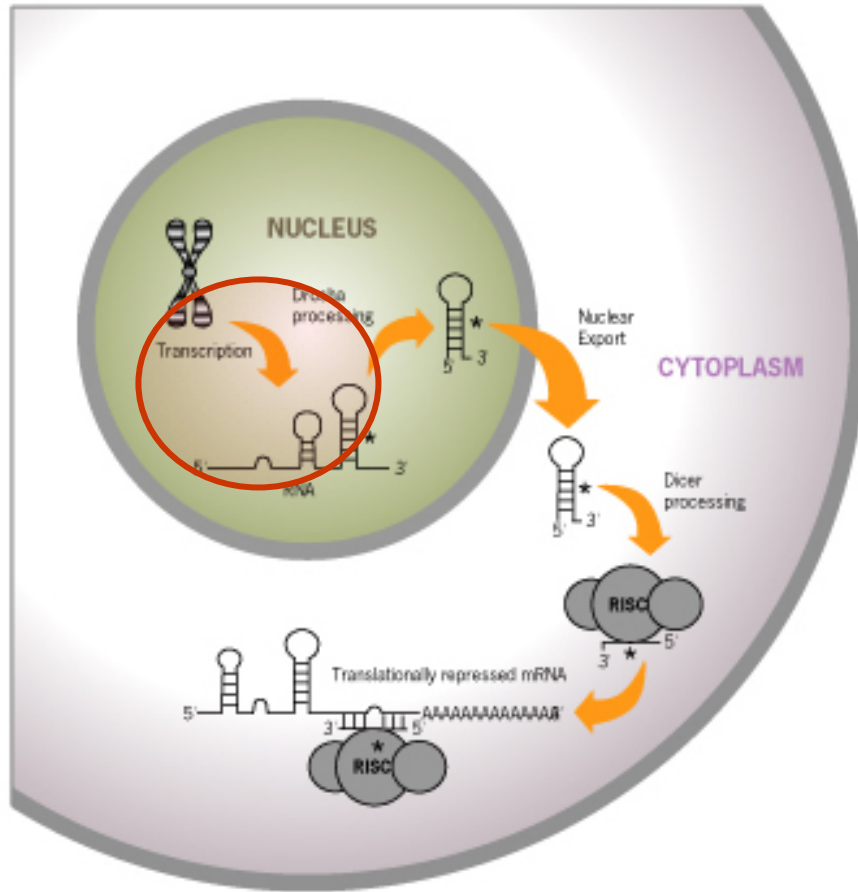
Department of Genetics, Medical School

Department of Computational and Information Science, SEAS

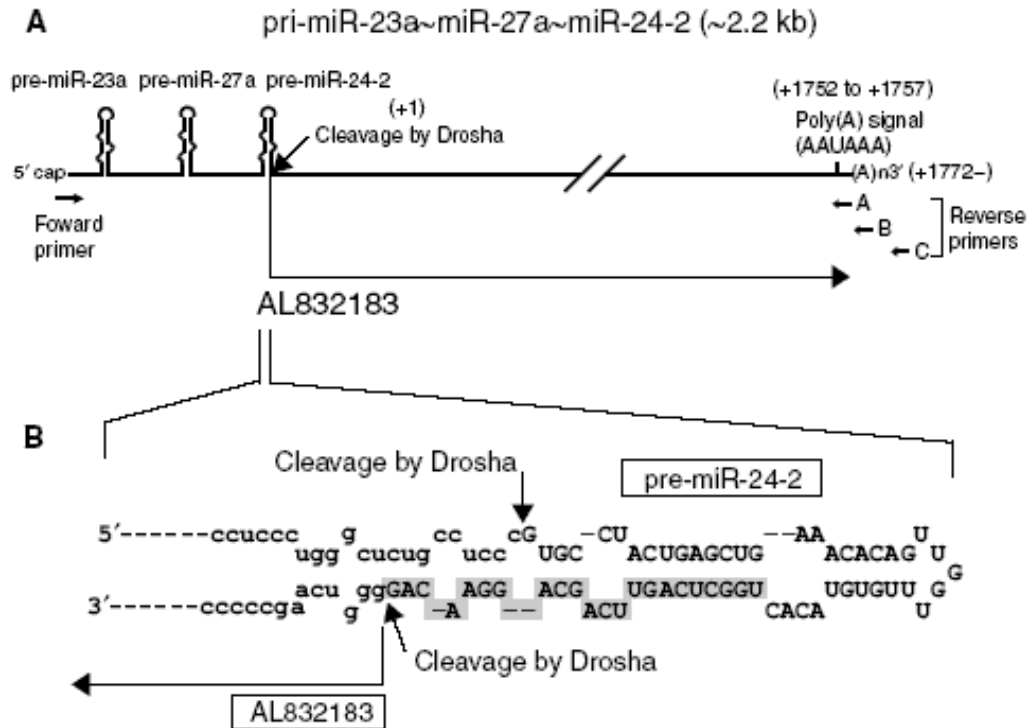
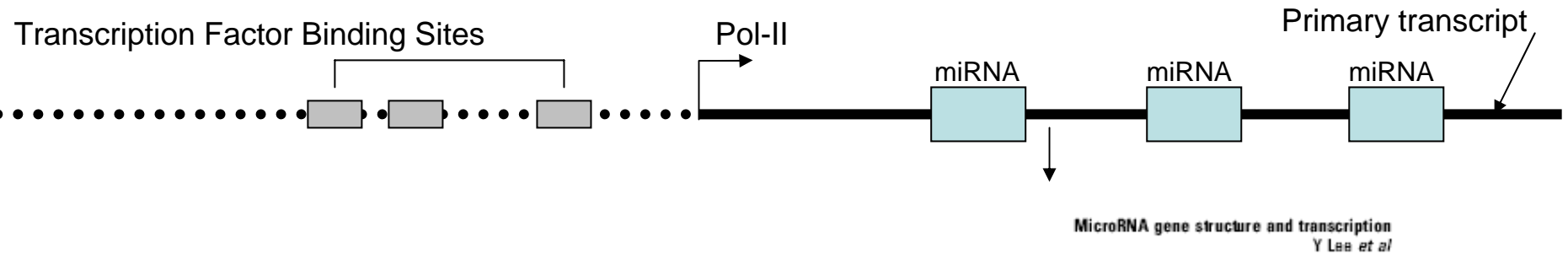
Center for Bioinformatics

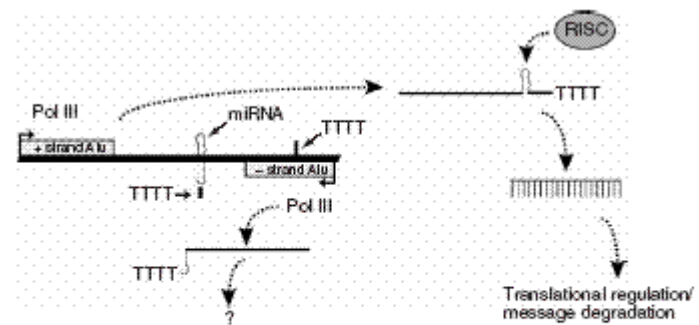
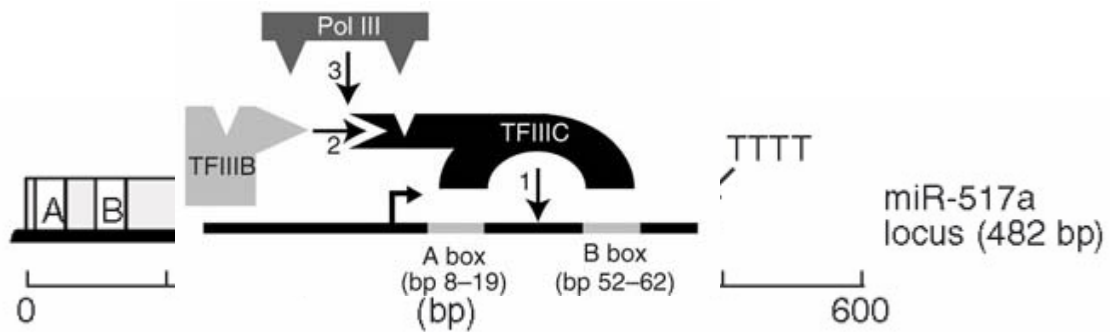
University of Pennsylvania





MicroRNA promoters



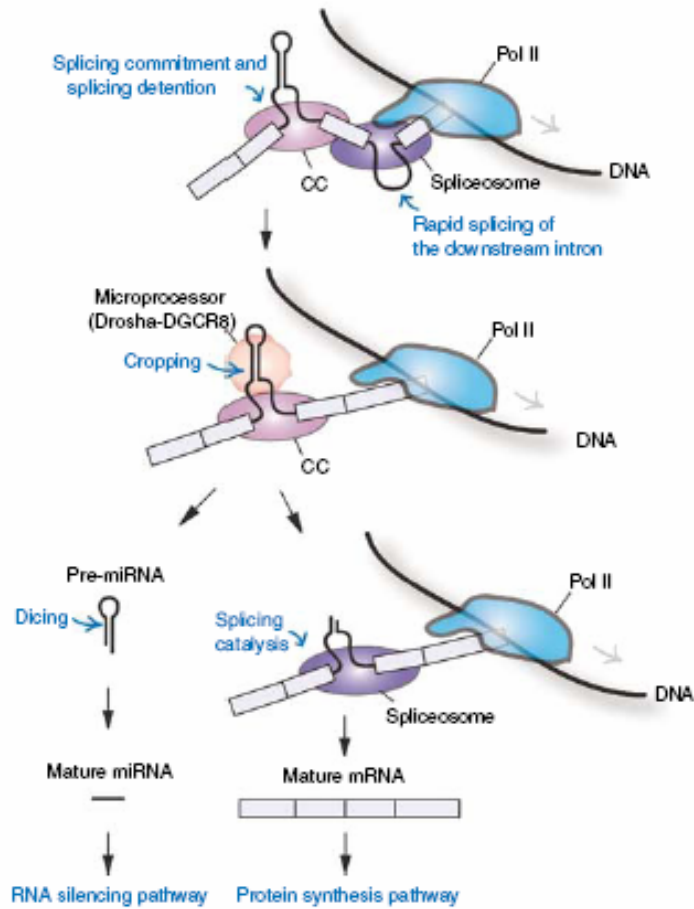


Organism	UCSC Known Genes	Refseq Genes	Genscan Genes
<i>A. gambiae</i>			16/37 (43.2%)
<i>C. elegans</i>		20/116 (17.2%)	
<i>C. familiaris</i>		1/6 (16.7%)	2/6 (33.3%)
<i>D. melanogaster</i>		21/78 (26.9%)	26/78 (33.3%)
<i>G. gallus</i>		10/147 (6.8%)	81/147 (55.1%)
<i>H. sapiens</i>	166/466 (35.6%)	157/466 (33.7%)	237/466 (50.9%)
<i>M. musculus</i>	129/367 (35.1%)	117/367 (31.9%)	203/367 (55.3%)
<i>P. troglodytes</i>		21/65 (32.3%)	35/65 (53.8%)
<i>R. norvegicus</i>	27/228 (11.8%)	33/228 (14.5%)	117/228 (51.3%)
<i>T. nigroviridis</i>			52/143 (36.4%)

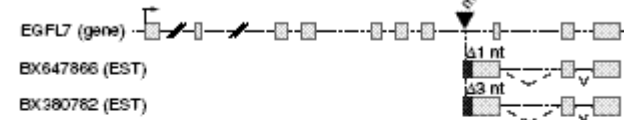
Table 1 displays the proportion of miRNAs in UCSC Known Genes, Refseq Genes, and Genscan Genes for each species.

Processing of intronic microRNAs

Y-K Kim and VN Kim



B miR-126 in EGFL7



A to I ; from A-U to I-U wobble

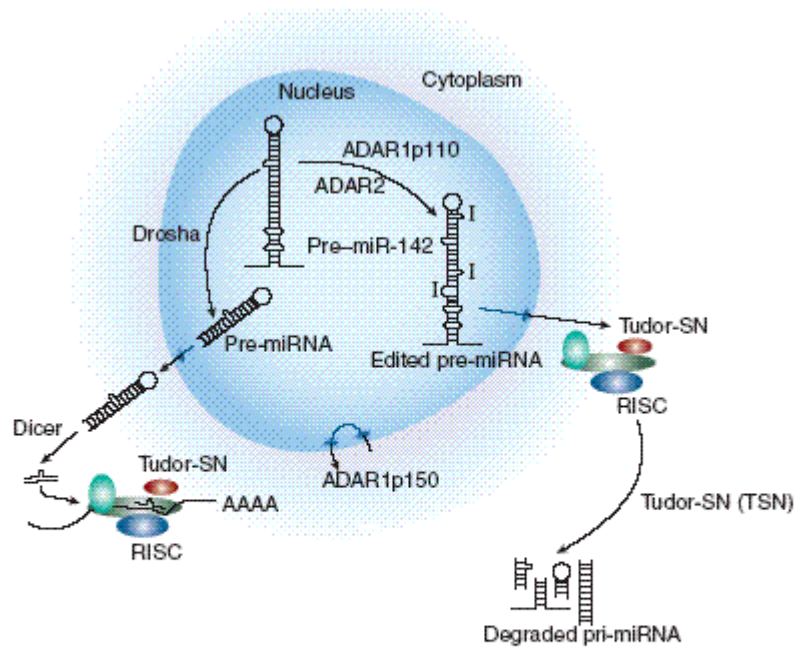


Figure 1 The consequences of editing of pri-miR-142 by the nuclear editing enzymes, ADAR1p110 or ADAR2. Editing introduces inosines that can affect RNA structure, and this interferes with processing by Drosha. The edited pri-miRNA is then thought to be exported to the cytoplasm, where it is degraded by Tudor-SN.

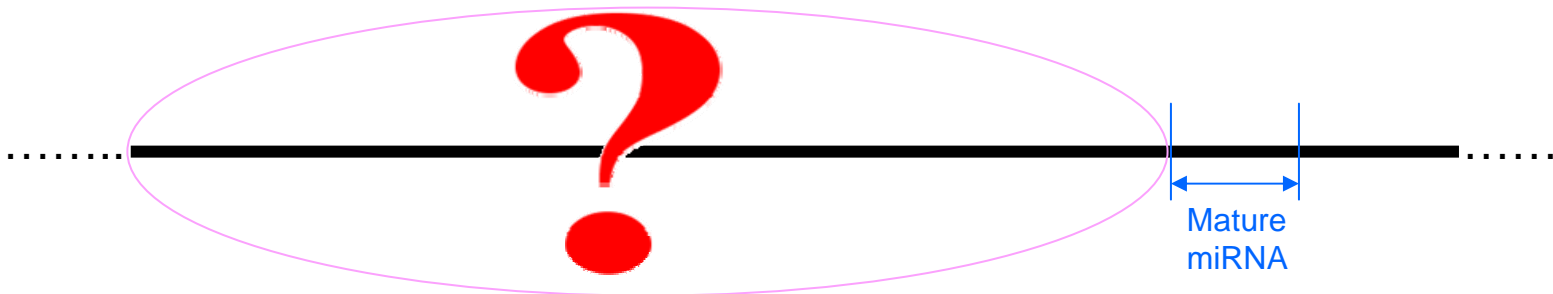
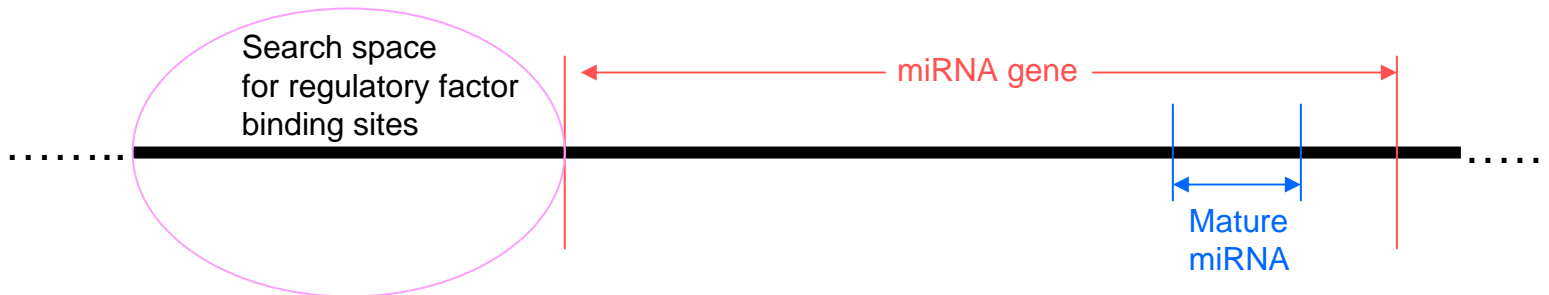
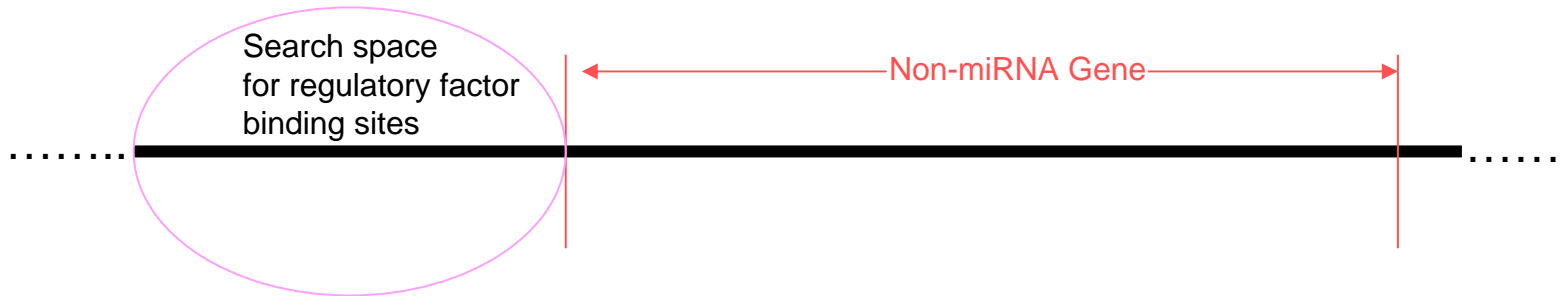
M. O'Connel & L Keegan, Nature Str.& Mol. Biology (2006)

W. Yang, et al , Str.& Mol. Biology (2006)

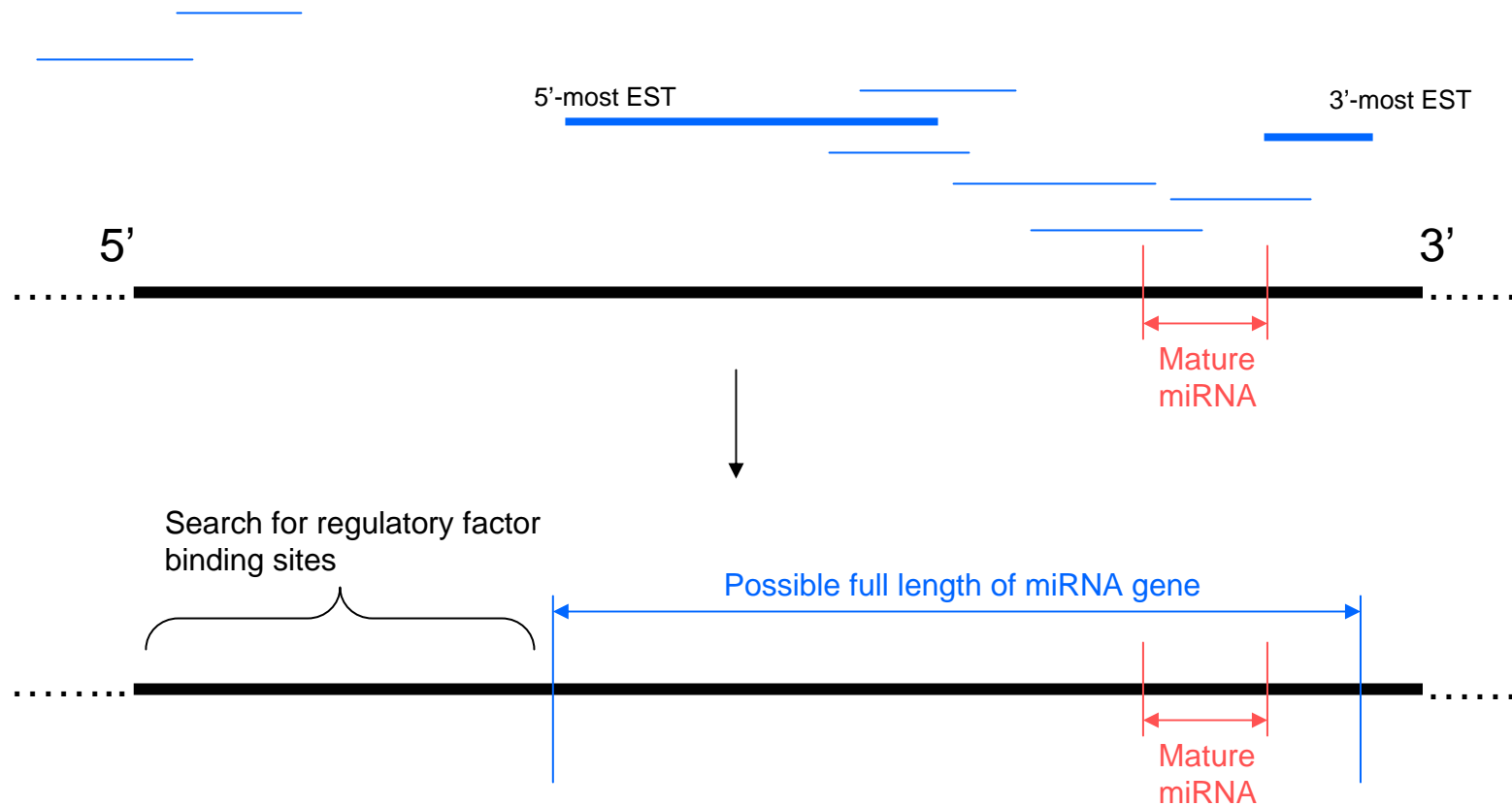
Organism	Cluster Distance 500 nt	Cluster Distance 1 kb	Cluster Distance 5 kb	Cluster Distance 50 kb
<i>H. sapiens</i>	99/466 (21.2%)	142/466 (30.5%)	204/466 (43.8%)	230/466 (49.4%)
<i>M. musculus</i>	107/367 (29.2%)	133/367 (36.2%)	169/367 (46.0%)	200/367 (54.5%)
<i>R. norvegicus</i>	73/228 (32.0%)	91/228 (39.9%)	112/228 (49.1%)	127/228 (55.7%)
<i>G. gallus</i>	39/147 (26.5%)	50/147 (34.0%)	65/147 (44.2%)	79/147 (53.7%)

Table 2 displays the proportion of miRNAs falling into clusters of size two or more for a sample collection of species. Cluster distance is the maximum distance between any two miRNAs considered to be in the same cluster.

- *What is the most biologically meaningful region to search for regulatory factor binding sites?*



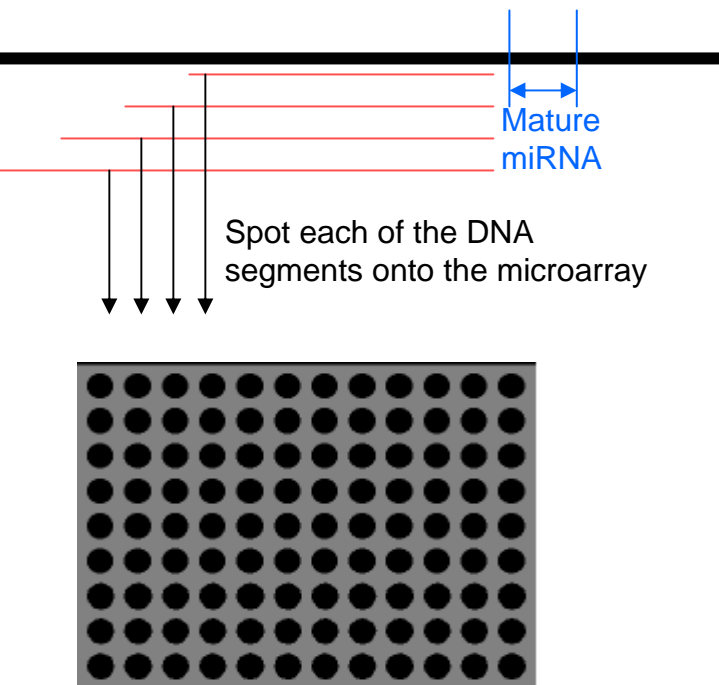
EST-based approximation of human and mouse miRNA primary transcripts



Sethupathy P, Megraw M, Barrasa M, Hatzigeorgiou A. 2005. Computational Identification of Regulatory Factors Involved in MicroRNA Transcription. *Lecture Notes in Computer Science*. 3746:457-468.

➔ High-throughput experimental investigation to define full-length transcripts

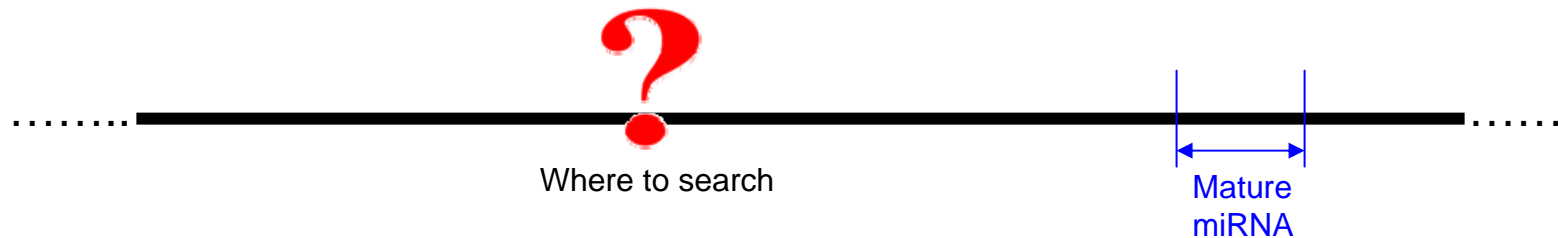
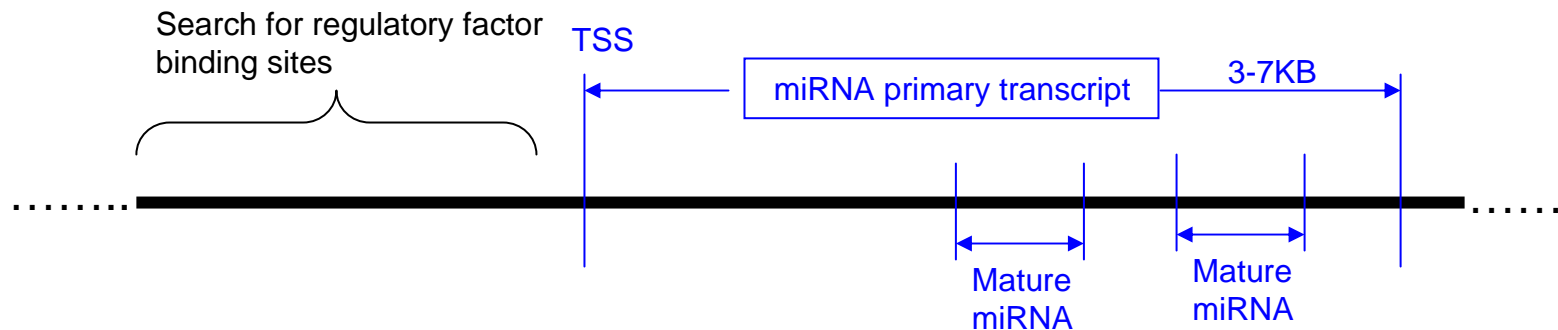
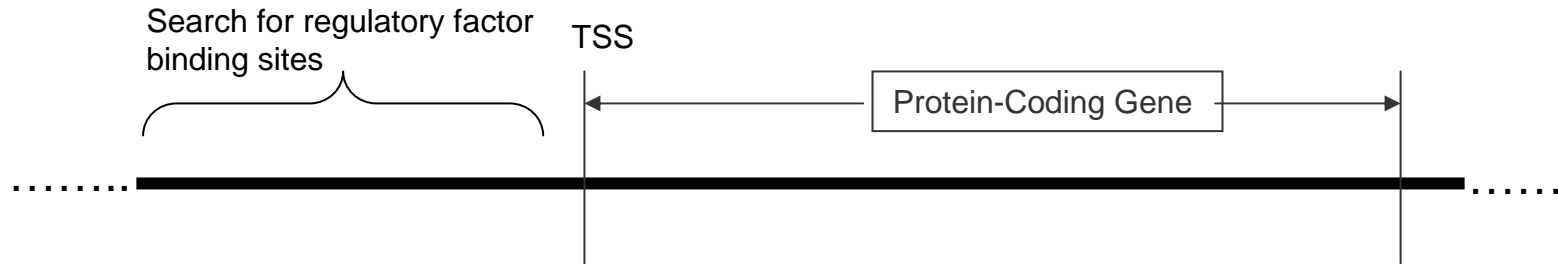
Experimental



What is the longest segment that hybridizes to RNA from Drosha depleted cell lines?

The miRNA Promoter Problem

Where do we expect to find regulatory factor binding sites?



miRNA Promoter Element Discovery in Arabidopsis

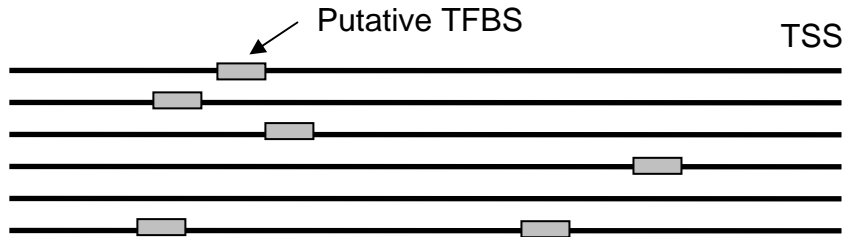
Data

A published set of Transcription Start Sites (TSSs) for 52 miRNA primary transcripts identified in Arabidopsis via 5'-RACE (Carrington, Aug. 2005).

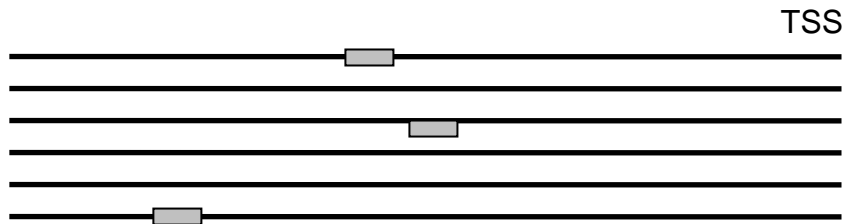
Project Goal

Are there any known Transcription Factor Binding Sites (TFBSs) which appear in a higher proportion of miRNA promoters than protein-coding gene promoters?

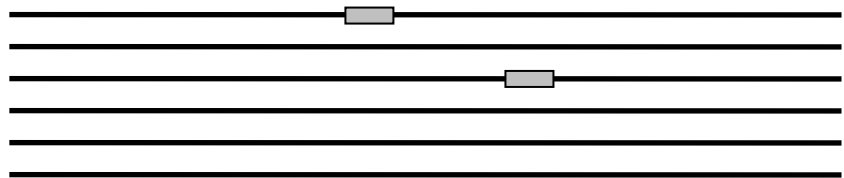
Comparison of binding site frequency



miRNA promoters

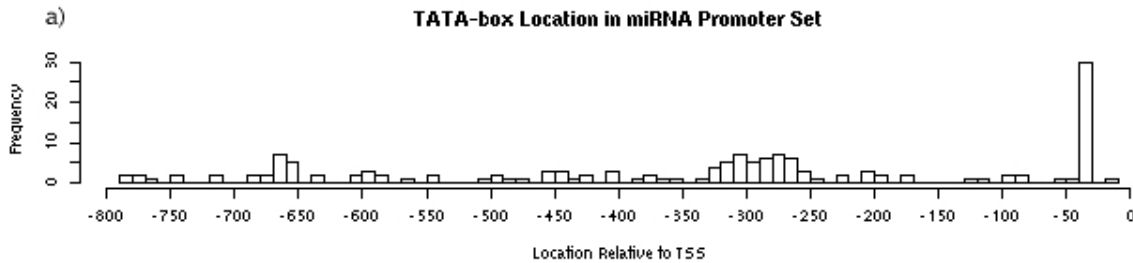


Protein-coding
gene promoters

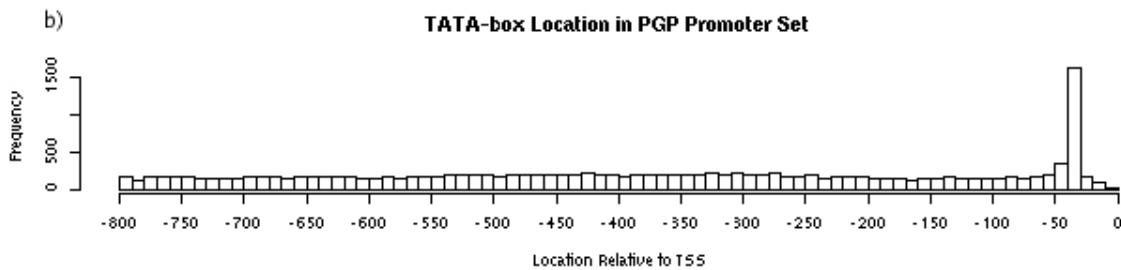


Randomly selected
Arabidopsis sequences

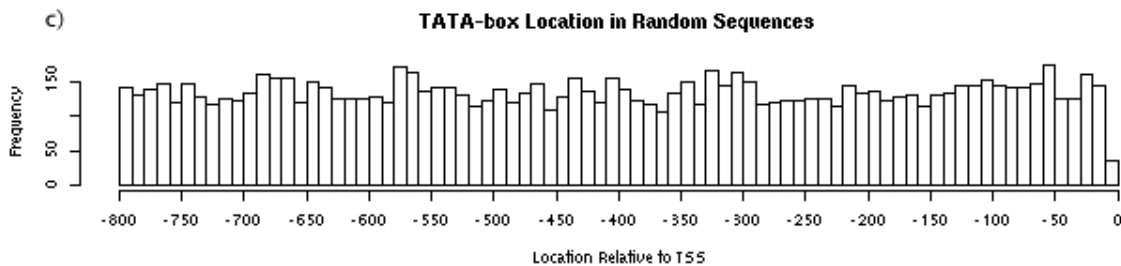
Histograms of TATA-box binding site locations



miRNA promoter set



protein-coding gene promoter (PGP) set



set of randomly selected Arabidopsis genome sequences.

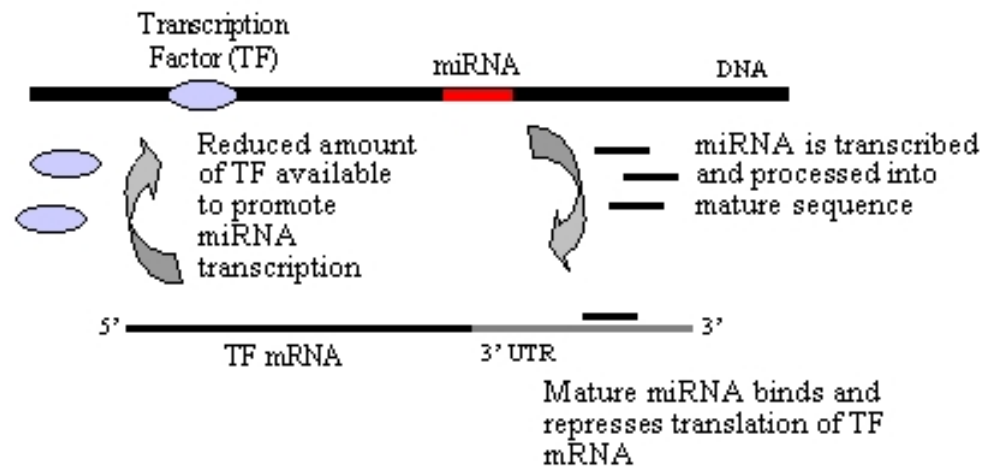
Results

TF Binding Site Motif		Count	Proportion of Sequences			Posterior Prob.	
Name	Consensus	miRNA	miRNA	PGP	Random	miRNA > PGP	miRNA > Random
TATA-box	TATA(A/T)A(T/A)A	42	0.81	0.52	0.39	0.98	1.00
AtMYC2	CACATG	14	0.27	0.17	0.18	0.91	0.90
ARF	TGTCTC	14	0.27	0.17	0.20	0.92	0.81
SORLREP3	TGTATATAT	8	0.15	0.04	0.03	1.00	1.00
LFY	CCA(T/A)TG	24	0.46	0.34	0.44	0.89	0.57

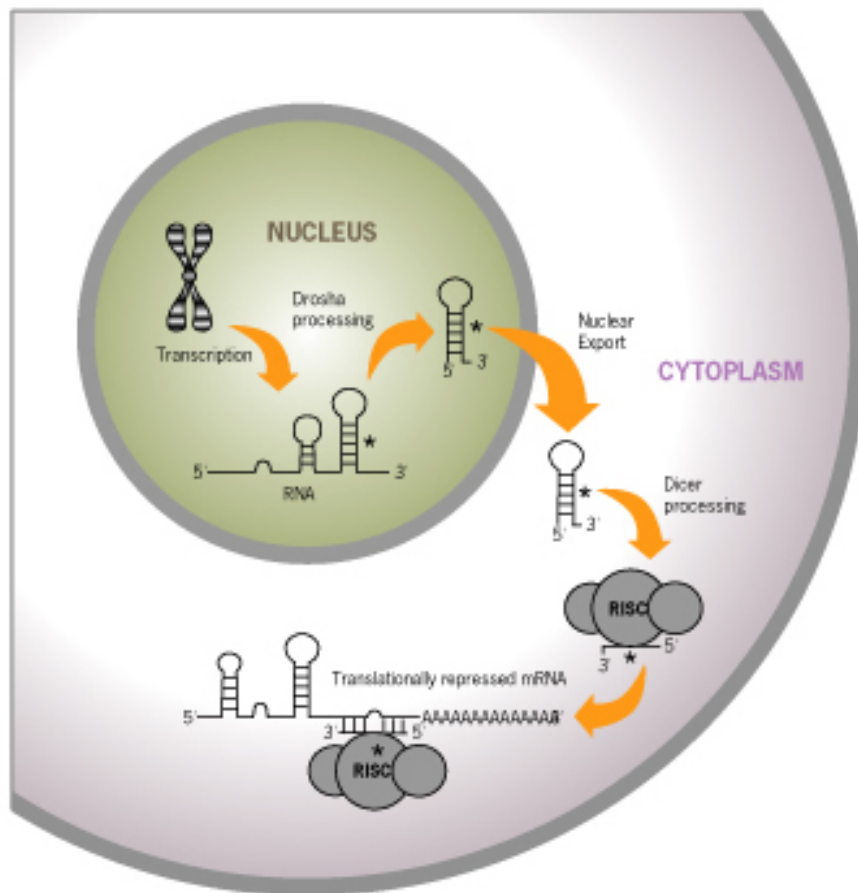
- Chart shows the proportion of miRNA promoters, protein-coding gene promoters, and random sequences which contain at least one observation of the given binding site motif.

Promoter Element Discovery in Arabidopsis

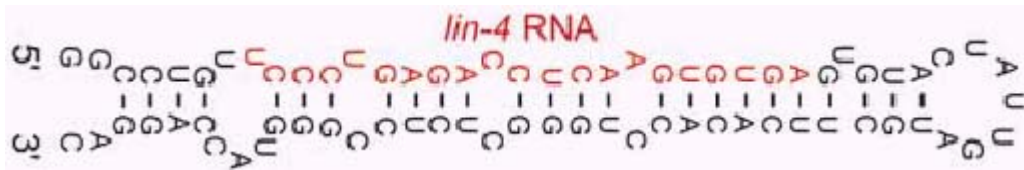
Two miRNAs with putative ARF binding sites upstream, miR-160 and miR-167, have experimentally supported targets belonging to the ARF gene family.



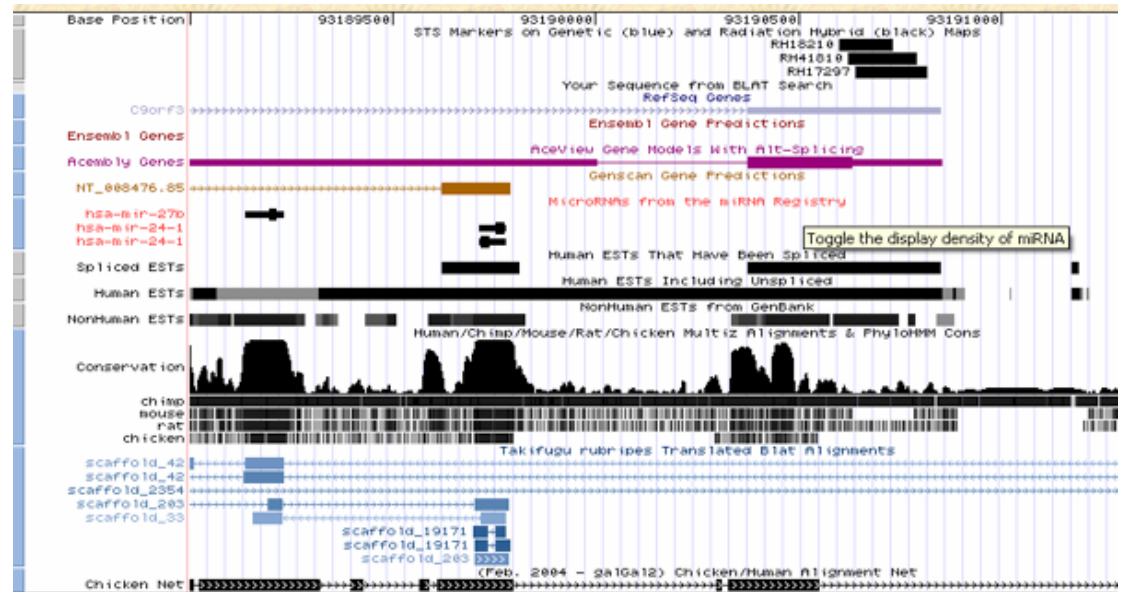
Megraw M, Baev V, Rusinov R, Jensen S, Kalantidis K, Hatzigeorgiou A (2006)
MicroRNA Promoter Element Discovery in Arabidopsis. **RNA**



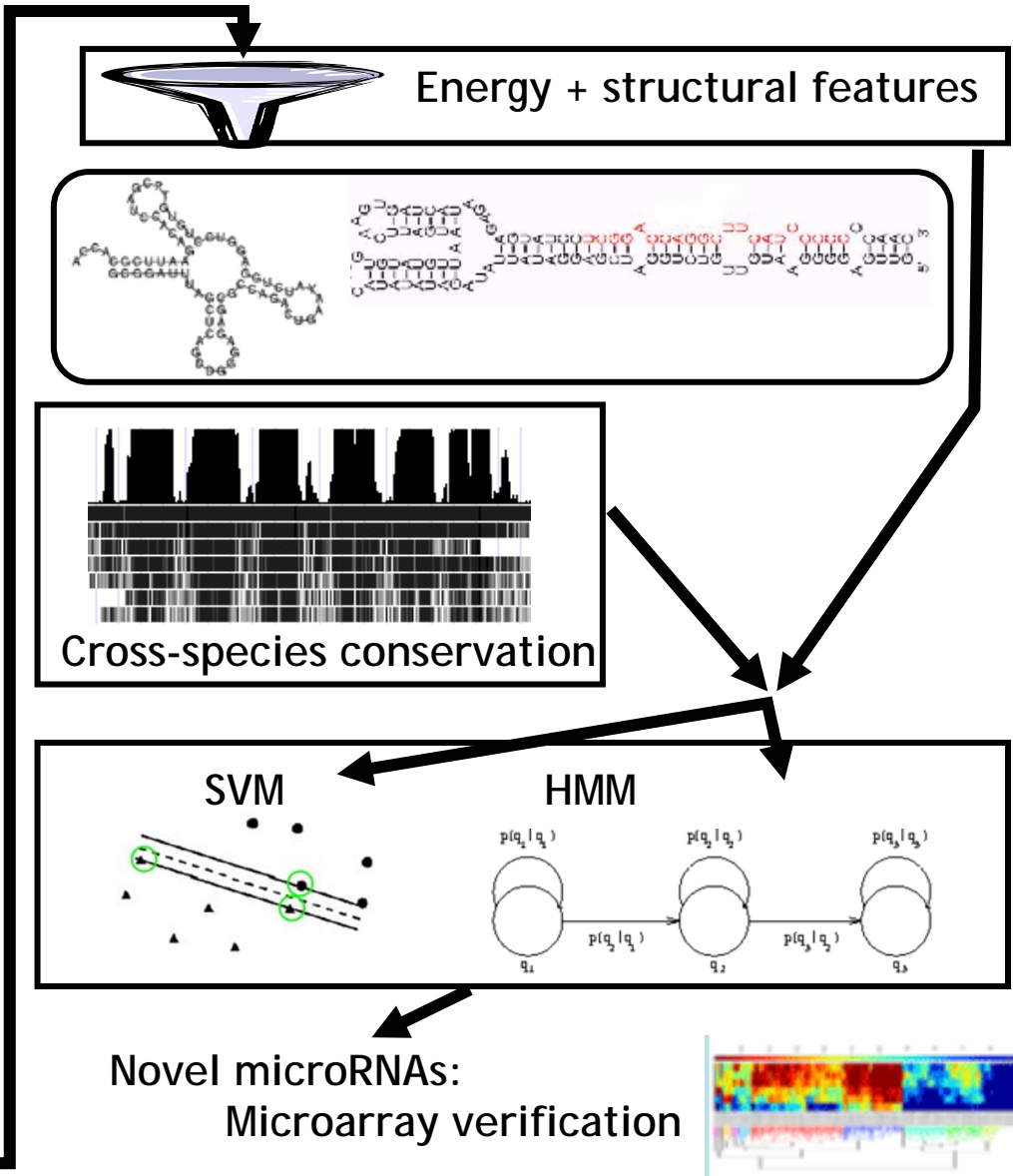
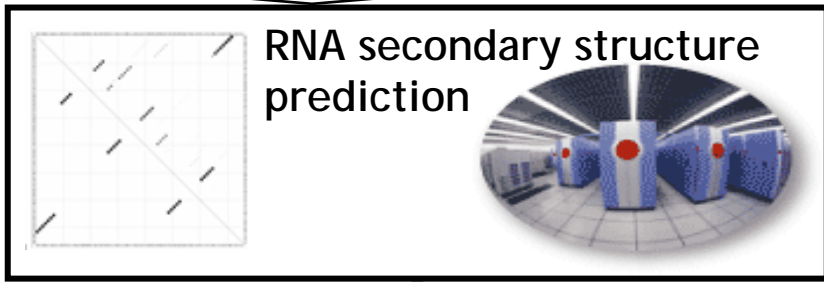
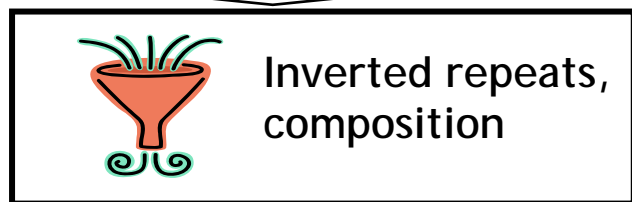
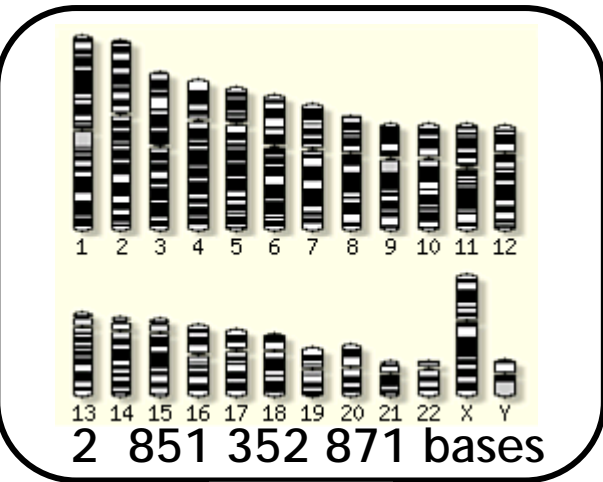
predicted secondary structure



comparative analysis

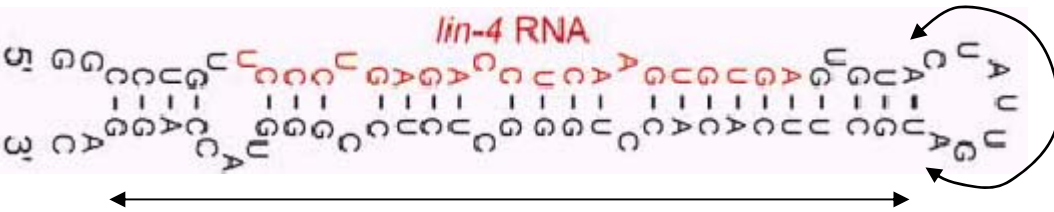


miRNA computational prediction pipeline

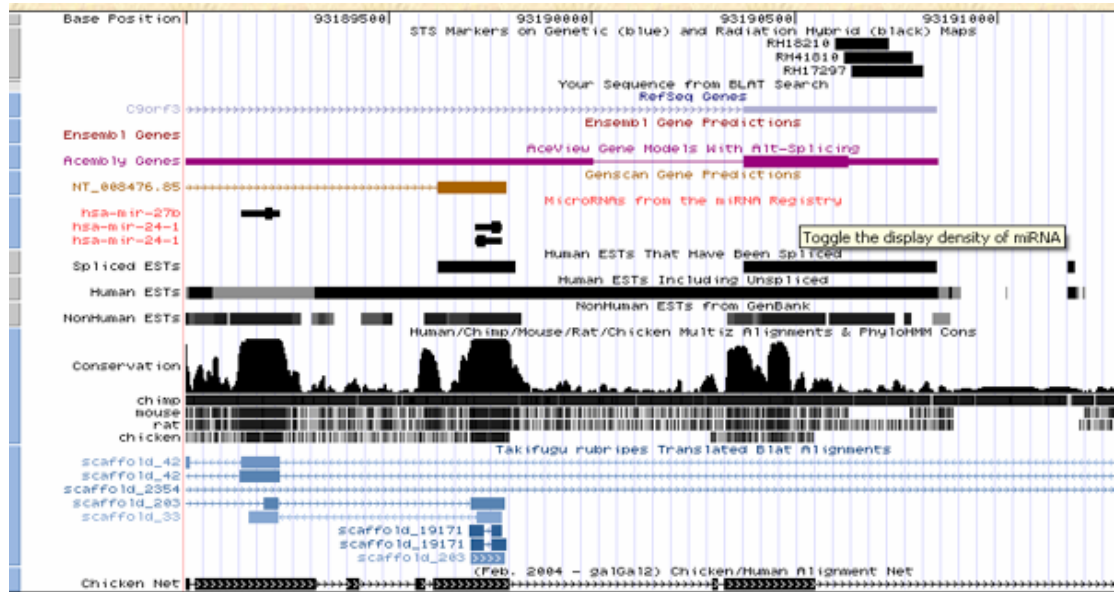


Prediction features

predicted secondary structure



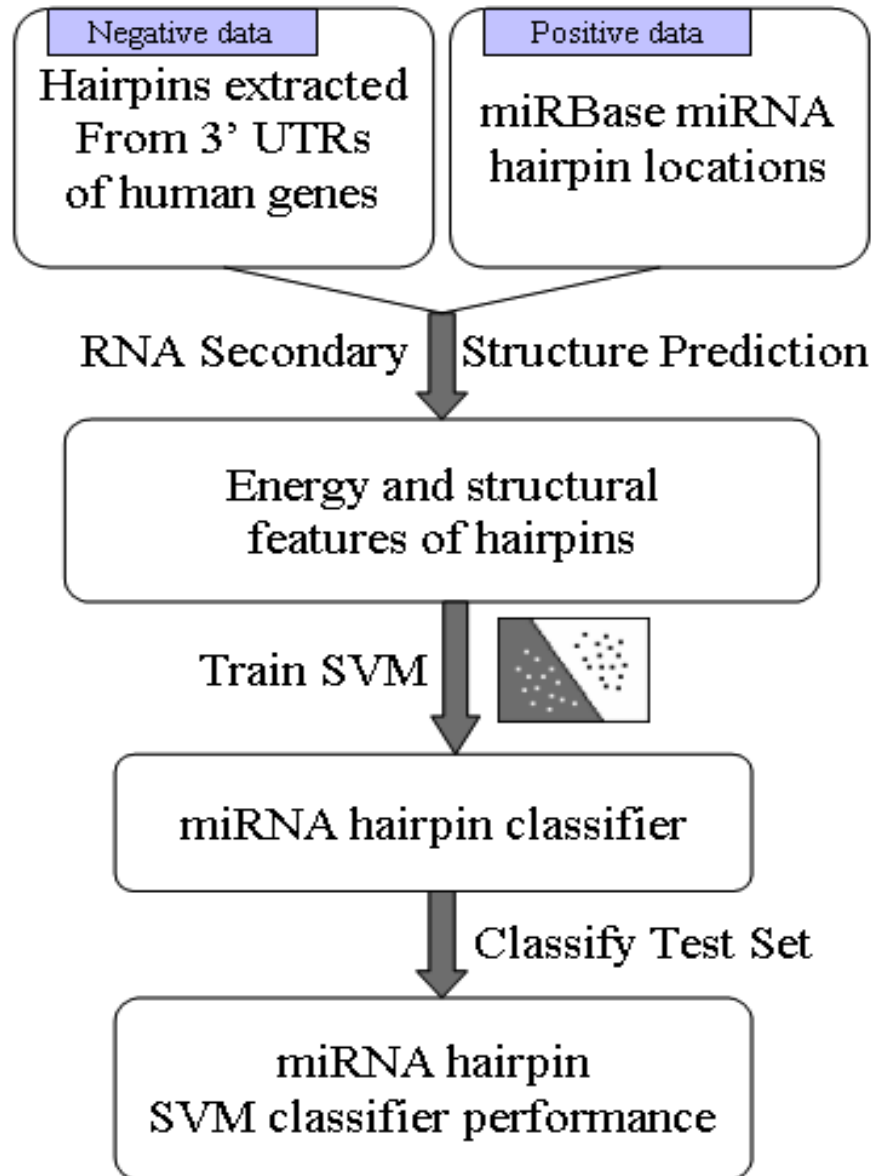
comparative analysis



1. Stem_Length
2. GC_Content
3. Stem_BPs
4. maxLinHelix
5. MatureCons
6. MatureOppositeCons
7. ArmCons
8. SS_Energy
9. MatureBPs
10. MatureEnergyProfile

=> 10 features for SVM classification

Training the Classifier



Combined miRNA gene prediction: DIANA-microG

Windows of 110 nt : calculate secondary structure

Structural features & conservation features

Classifier (SVM)

0.5

Ranking the results:

_____ 0.9
_____ 0.81
_____ 0.72 known miRNA

8,000 cand. ,
verification
with chip
experiment.

Data analysis of experimental verification

Expression data from 8184 predicted miRNAs on 6 human tissues

Expression above 98% negative cutoff in any of 6 tissues:




Probe type : #above / total

rRNA	:	37	/	44	=	84.1 %	} positive control
known ncRNA	:	168	/	397	=	42.3 %	
true miRNA	:	368	/	1782	=	18.0 %	
Our predictions	:	2834	/	8197	=	33.4 %	
Negative random	:	10	/	500	=	2.0 %	} negative control

microRNA IDENTIFICATION in ENCODE region

Computational approach based on machine learning approach (Support Vector Machine) predicts 8,000 new miRNA candidates.

Printing a chip with these predictions (2 X 30nt probes for each put. miRNA).
30% expressed in 6 tissues (thymus, placenta, lung, ovary, liver and brain)

- ✓ **ENCODE MicroRNA PREDICTION**  271
microRNA candidates
- ✓ **MICROARRAY TESTING**  54
microRNA candidates
- ✓ **OVERLAP WITH TILING ARRAY EXPRESSION SIGNAL**  16
microRNA candidates

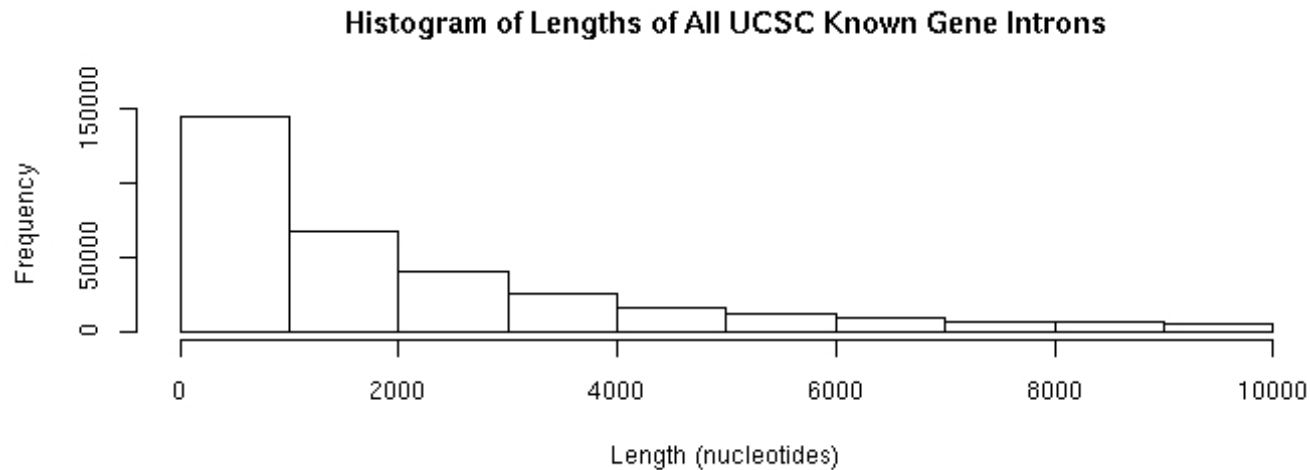
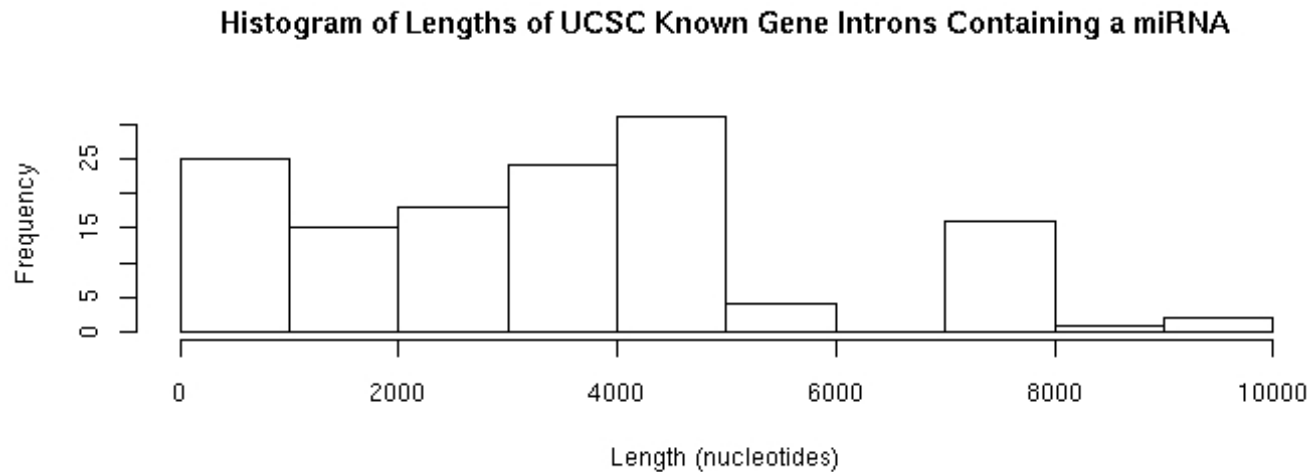
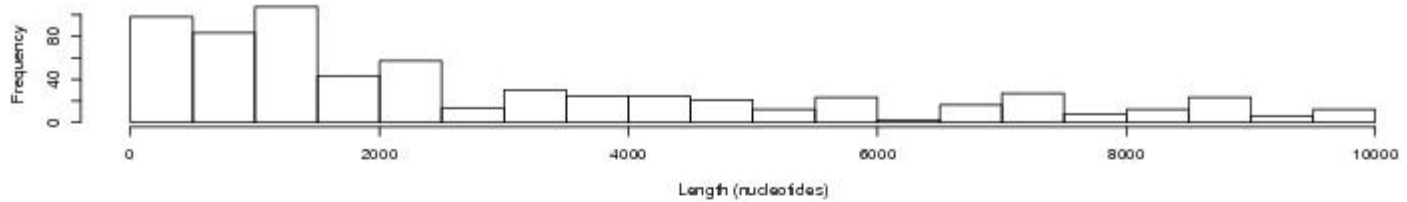


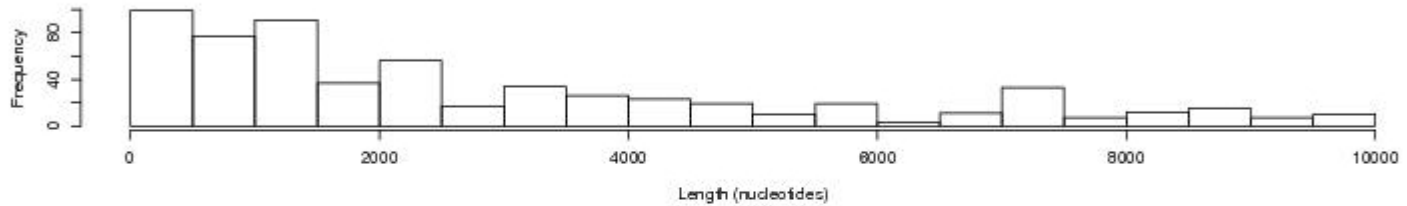
Figure 1: Histogram Comparison of Intron Lengths Up to 10kb

These histograms compare the distribution of lengths of UCSC Known Introns containing a miRNA to the distribution of lengths of all UCSC Known Genes, for intron lengths up to 10kb.

Histogram of Lengths of UCSC Known Gene Introns Containing a miRNA from file expressed.intronLength.txt



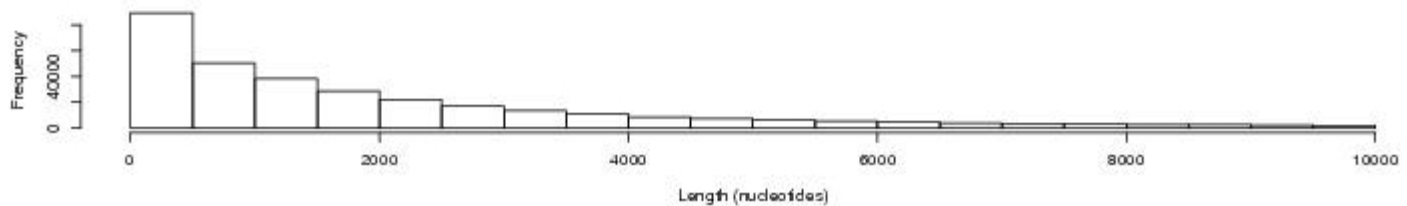
Histogram of Lengths of UCSC Known Gene Introns Containing a miRNA from file expressed_vsn.intronLength.txt

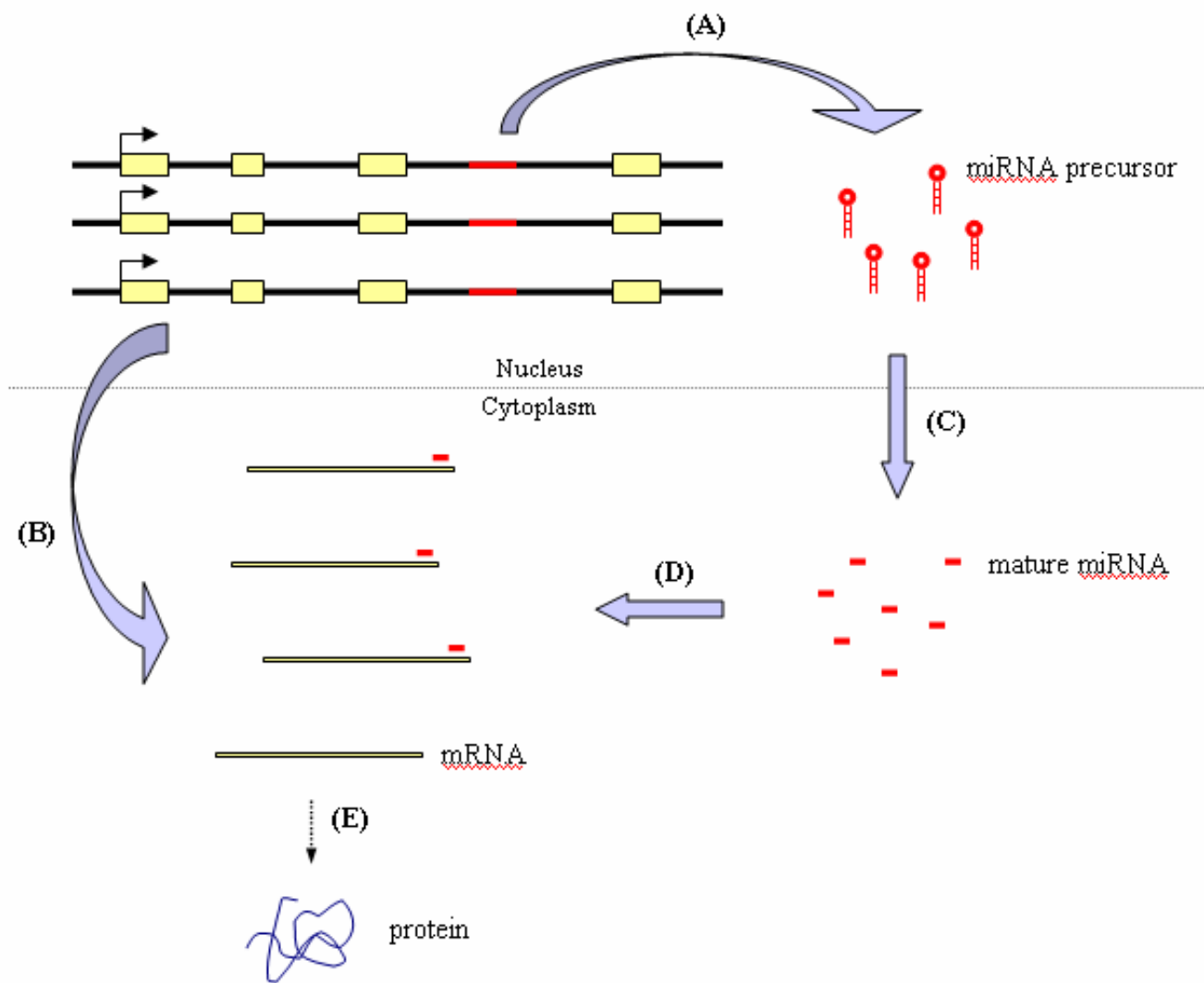


Histogram of Lengths of UCSC Known Gene Introns Containing a miRNA from file belowChipScore.intronLength.txt



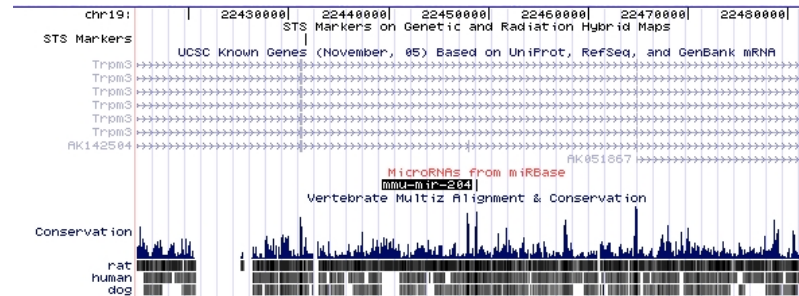
Histogram of Lengths of All UCSC Known Gene Introns



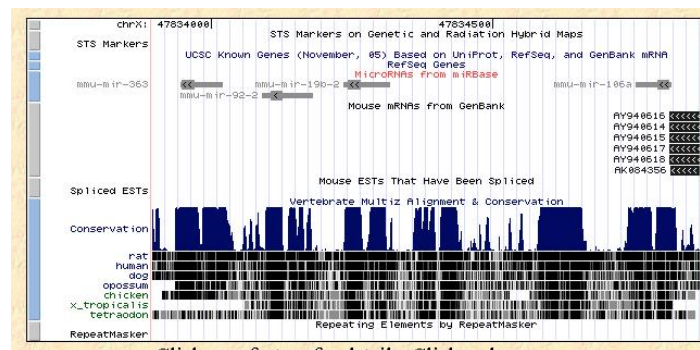


miRGen: A database for the study of animal microRNA genomic organization and function

- Where are the miRNAs with respect to UCSC Known Genes?

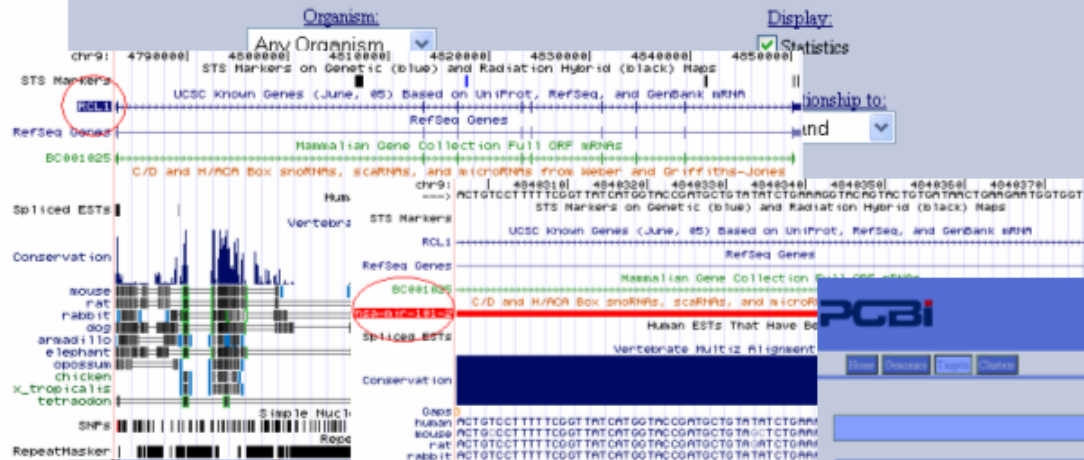


- Which miRNAs fall into clusters at a given inter-miRNA distance?



miRNA Genomics

Query



Human miRNA Targets

Query

Target Type: Experimentally Supported Targets (TarBase)

Search by miRNA: hsa-mir-101-2

OR Gene ID: RefSeq ID

Submit Query

Results

Experimentally Supported Targets from TarBase

miRNA	Gene	Chromosomal Location	miR	Target Site	Target	Source	Class	Year
miR-101	Bax1	chr7:149942122-148819962	Other	3'UTR	in vitro reporter gene assay (Luciferase)	Chao	2	Leung et al, 2003
miR-101	MYCN	chr2:16831280-16807726	Other	3'UTR	in vitro reporter gene assay (Luciferase)	Chao	2	Leung et al, 2003

miRNA Clusters

The Clusters interface provides predicted miRNA clusters at any given inter-miRNA distance, and provides specific functional information on the targets of miRNAs within each cluster.

Query

Organism: Cluster Distance:

Link to:

Advanced Query Options:

- All miRNAs contained within the same gene are always considered a separate cluster
- No cluster should have a gene located between miRNA members of the cluster

Use Gene Set:

Results

Cluster Color Coding (with respect to *Known Genes*):

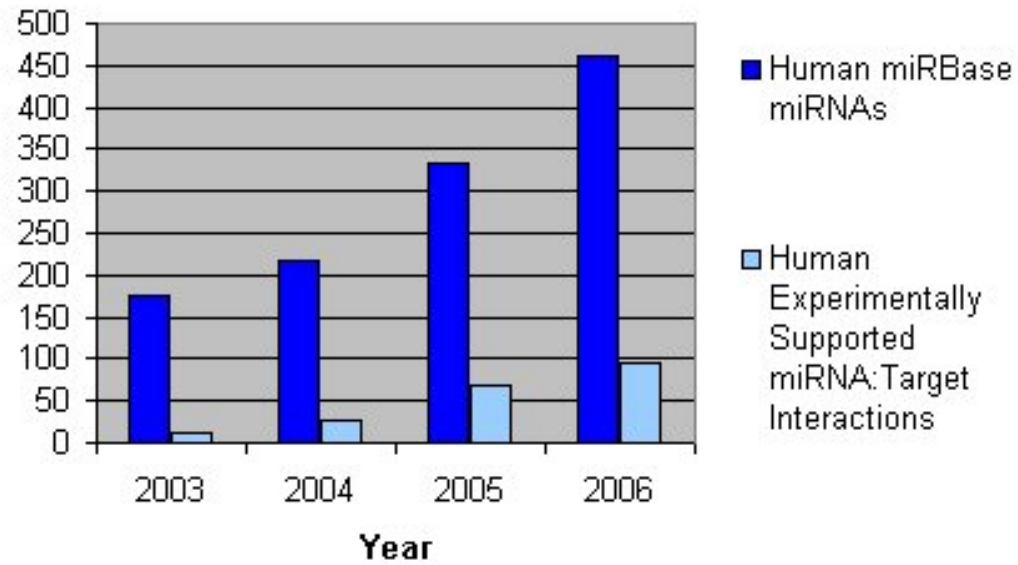
- GENIC
- NON-GENIC

Chr X (+): *hsa-mir-532* *hsa-mir-188* *hsa-mir-500* *hsa-mir-362* *hsa-mir-501* *hsa-mir-660* *hsa-mir-502*
 Chr X (-): *hsa-mir-221* *hsa-mir-222* | *hsa-mir-98* *hsa-let-7f-2* | *hsa-mir-545* *hsa-mir-374* | *hsa-mir-363* *hsa-mir-18b* *hsa-mir-106a* | *hsa-mir-450-1* *hsa-mir-450-2* *hsa-mir-542* *hsa-mir-503* *hsa-mir-424* *hsa-mir-514-1* *hsa-mir-514-2* *hsa-mir-514-3* *hsa-mir-514-3* | *hsa-mir-224* *hsa-mir-300b* *hsa-mir-300a* *hsa-mir-479* | *hsa-mir-30c* *hsa-mir-30e-1*



miRGen ID	RefSeq ID	Associated Gene	miRGen ID	RefSeq ID	Description	GO Terms
hsa-mir-221	hsa-mir-221	SPDY	hsa-mir-222	hsa-mir-222	SPDY (protein-coding)	regulation of transcription, initiation of transcription, transcription
hsa-mir-98	hsa-mir-98	SPDY	hsa-mir-374	hsa-mir-374	SPDY (protein-coding)	regulation of transcription, transcription
hsa-mir-18b	hsa-mir-18b	SPDY	hsa-mir-514-1	hsa-mir-514-1	SPDY (protein-coding)	regulation of transcription, transcription
hsa-mir-514-2	hsa-mir-514-2	SPDY	hsa-mir-514-3	hsa-mir-514-3	SPDY (protein-coding)	regulation of transcription, transcription

Megraw M, Sethupathy P, Corda B, Hatzigeorgiou A. (2007) miRGen: A database for the study of animal microRNA genomic organization and function. NAR



TarBase: Database of experimentally supported miRNA targets

Address <http://www.diana.pcbi.upenn.edu/cgi-bin/search.cgi?mode=Trans> Go Links >>

Google Search ABC Check AutoLink AutoFill Options

PCBI Penn Center for Bioinformatics **Tarbase** Penn

Home Search Download Submit Animal Targets Submit Plant Targets Submission Instructions : Animals Submission Instructions : Plants

Translationally Repressed targets for Human

miRNA	Gene	MRE	Single Site Sufficiency	Indirect Support	Direct Support	Paper	Binding
miR-124	Mtpn	1 site	Unknown	in vivo overexpression of miRNA	in vitro reporter gene assay (Luciferase) AND immunoblotting	Krek et al, 2005	Binding Pictures
miR-375	Mtpn	1 site	Yes	in vitro 2'-O-me inhibition of miRNA AND in vitro overexpression of miRNA	in vitro reporter gene assay (Luciferase) AND immunoblotting	Poy et al, 2004	Binding Pictures
let-7b	Mtpn	Not Given	Unknown	in vivo overexpression of miRNA	in vitro reporter gene assay (Luciferase)	Krek et al, 2005	Binding Pictures
let-7b	Lin28	1 site	Yes		in vitro reporter gene assay (Luciferase)	Kiriakidou et al, 2004	Binding Pictures
miR-141	Clock	1 site	Yes		in vitro reporter gene assay (Luciferase)	Kiriakidou et al, 2004	Binding Pictures
miR-24	MAPK14	1 site	Yes		in vitro reporter gene assay (Luciferase)	Kiriakidou et al, 2004	Binding Pictures
miR-145	FLJ21308	1 site	Yes		in vitro reporter gene assay (Luciferase)	Kiriakidou et al, 2004	Binding Pictures
miR-23a	FLJ13158	1 site	Yes		in vitro reporter gene assay (Luciferase)	Kiriakidou et al, 2004	Binding Pictures
let-7e	SMC1L1	1 site	Yes		in vitro reporter gene assay (Luciferase)	Kiriakidou et al, 2004	Binding Pictures

Sethupathy, P., Corda, B., and Hatzigeorgiou, A.G (2006) *RNA*, 12:192-197

<http://www.diana.pcbi.upenn.edu/tarbase.html>

Current mammalian target prediction programs

- Widely used mammalian target prediction programs
 1. *TargetScan* (MIT, late 2003)
 2. *DIANA-microT* (UPENN, early 2004)
 3. *MiRanda* (Sloan-Kettering, 2004)
 4. *TargetScanS* (MIT, 2005)
 5. *PicTar* (NYU, 2005)
- Each program applies a slightly different set of “rules” that are thought to govern miRNA:target interactions
- How do these programs compare? What are the relative advantages of each? What are the limitations?

Evaluate mammalian target prediction programs

- Sensitivity
 - ~85 human/mouse miRNA:gene interactions that have direct experimental support for at least one target site (includes 32 different miRNAs)
- Specificity
 - How many total predictions are made by each program?

and some heuristics ...

- Compute for each miRNA a list of targets sorted after their minimum free energy
- Identification of targets conserved in human / mouse orthologs.

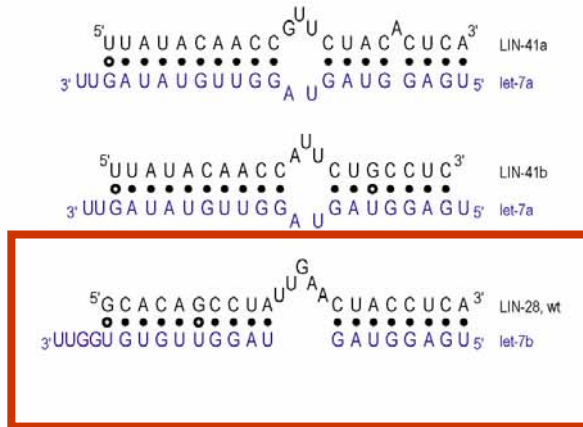


- Selection of 13 targets.

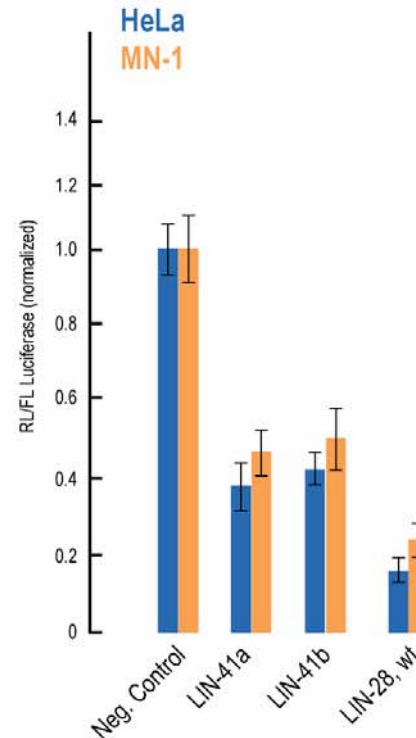
the wet experiment



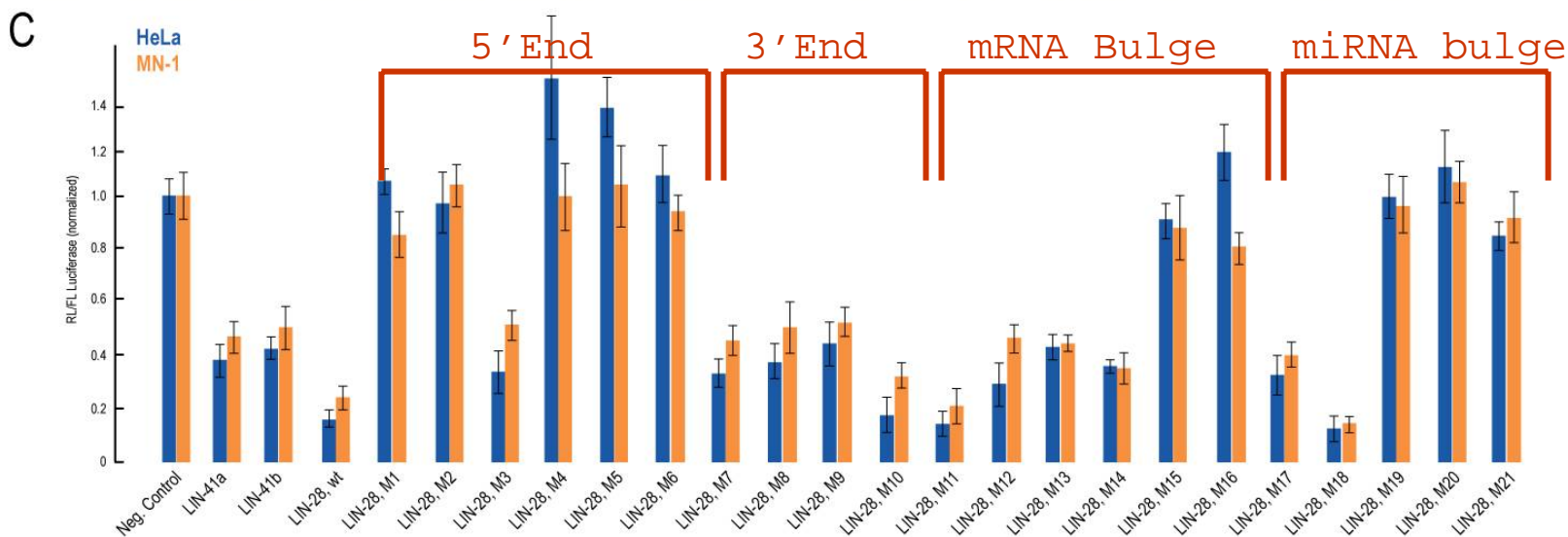
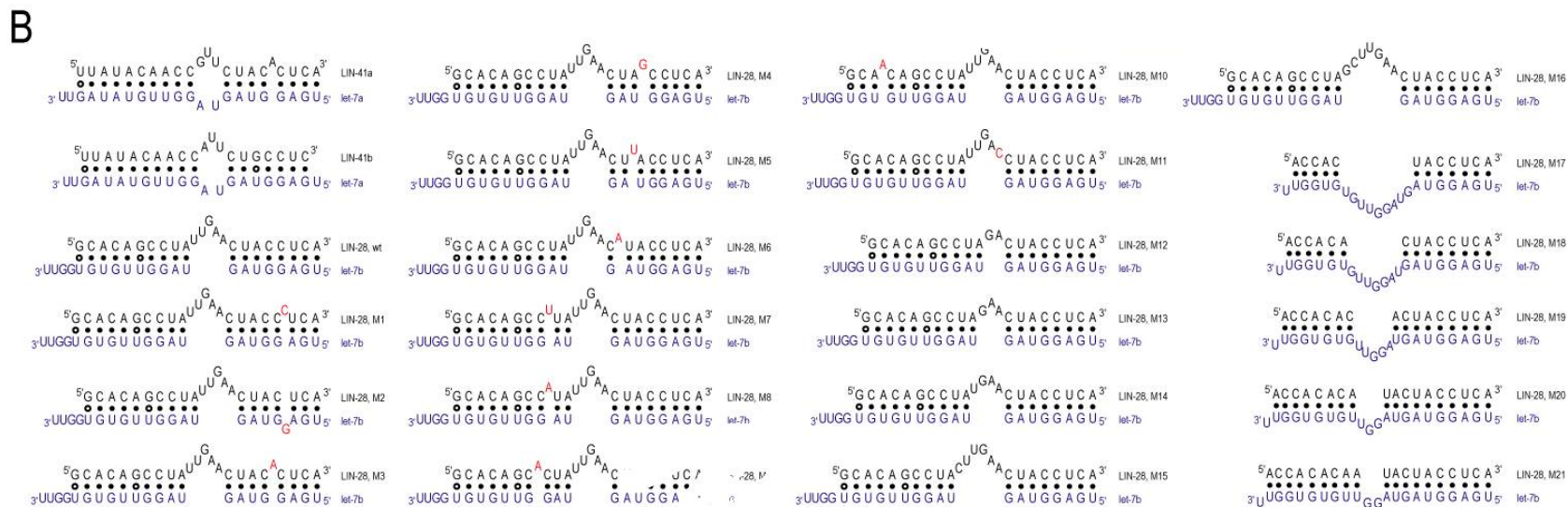
B



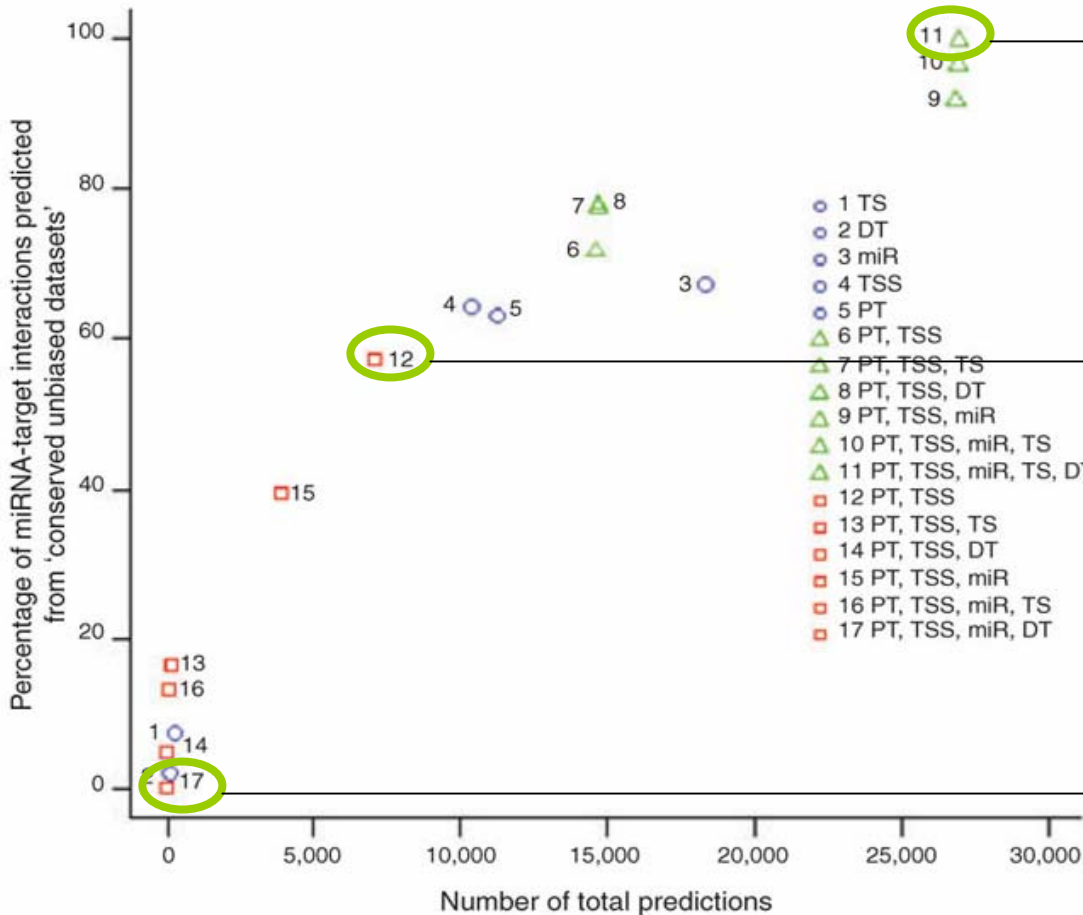
C



Kiriakidou M., Nelson P.T., Kouranov A., Fitziev P., Bouyioukos C., Mourelatos Z. and A. Hatzigeorgiou (2004) A combined computational-experimental approach predicts human microRNA targets. *Genes & Development*, 18(10):1165-78.



Performance Spectrum



Union of all programs has 100% sensitivity, but predicts a total of ~30,000 miRNA:target interactions

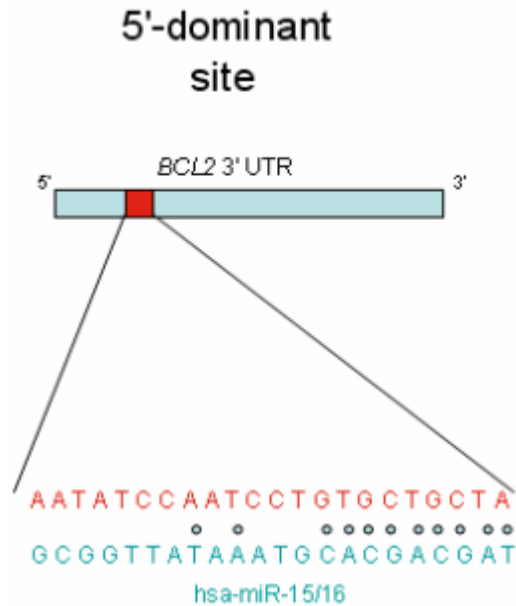
Intersection of TSS and PT has 57% sensitivity, and predicts a total of ~7,000 miRNA:target interactions

Intersection of all programs has 0% sensitivity, but only predicts a total of 2 miRNA:target interactions

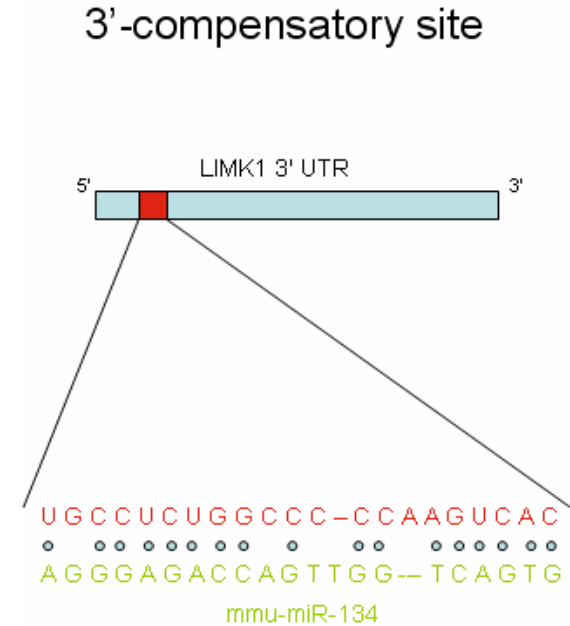
Wide spectrum is largely due to differences in methodology for 3'-compensatory predictions

Sethupathy, P., Megraw, M., and Hatziargiou, A.G. (2006) *Nature Methods*, 3:881-886.

miRNA:target interaction categories



- Perfect base pairing to at least 7 nucleotides starting from the first or second nucleotide at the 5'-end of the miRNA
- Binding to the 3'-end of the miRNA is currently considered irrelevant, but TarBase indicates that it is often extensive.



- Imperfect or shorter stretch of base pairing to at least 7 nucleotides starting from the first or second nucleotide at the 5'-end of the miRNA
- Extensive binding to the 3'-end of the miRNA in order to compensate for the weaker binding to the miRNA 5'-end

Diana-microT 2.0

Diana-microT 2.0 Web Server

microRNA Target Prediction

Enter e-mail address:

Enter miRNA sequence(s)

(check proper format here)

OR

Enter your UTR sequence(s)

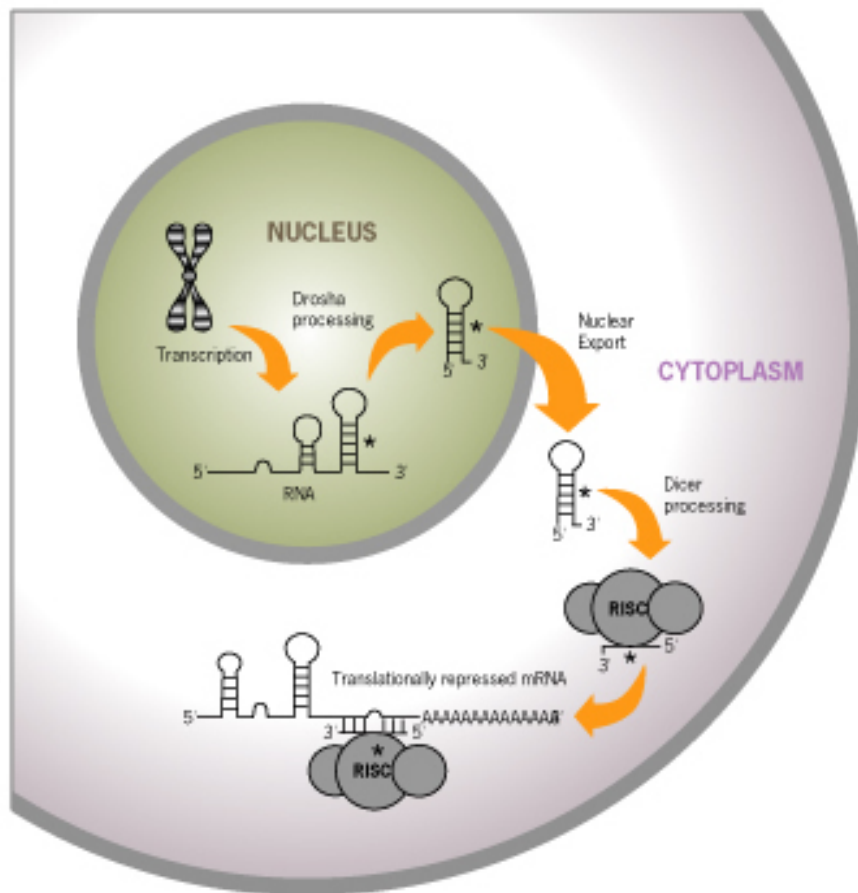
(check proper format here)

OR

Choose mode of conservation:

For a query-able interface to pre-compiled Diana-microT 2.0 target predictions for the human genome, please visit [here](#).

<http://diana.pcbi.upenn.edu/diana-microT2>



Resolving the molecular mechanisms of some polymorphic disease associations

- SNPs that occur in functional miRNA target sites could affect miRNA binding
- Map all annotated SNPs from dbSNP onto all experimentally supported target sites from TarBase
- 2 of the 5 SNPs occur in a region that disrupts the 5'-dominant binding
- 1 of these 2 SNPs is genotyped according to **ALFRED** (Allele Frequency Database)
- Does this SNP impair miR-155 binding and silencing of *AGTR1*?

```

5' UUCACUACCAAAUGAGCCUUAG 3'      Human AGTR1
..|  |  || |||||
3' GGGGAUAGUGCUAAUCCGUAUU 5'      Hsa-miR-155
..|  |  || |||
5' UUCACUACCAAAUGAGCCUUAG 3'      Polymorphic Human AGTR1

```

```

5' GCAGUUUGAAAUUCUGAAUUGCAAAGUACUGA 3'      Human EZH2
|||||  || |||||
3' AGUCA-----UAGUGUCAUGACAU 5'      Hsa-miR-101
|||||  || |||||
5' GCAGUUUGAAAUUCUGAAUUGCAAAGUACUGA 3'      Polymorphic Human EZH2

```

```

5' CCG-CAAGAAAGUGAATCTCACTACUACCUA 3'      Human HOXA7
|||. ||| .|||| |
3' GGGUUGUUG---UACUU-----UGAUGGAU 5'      Hsa-miR-196
|. ||| .|||| |
5' CCG-CAAGAAAGUGAATCTCACTACUACCUA 3'      Polymorphic Human HOXA7

```

```

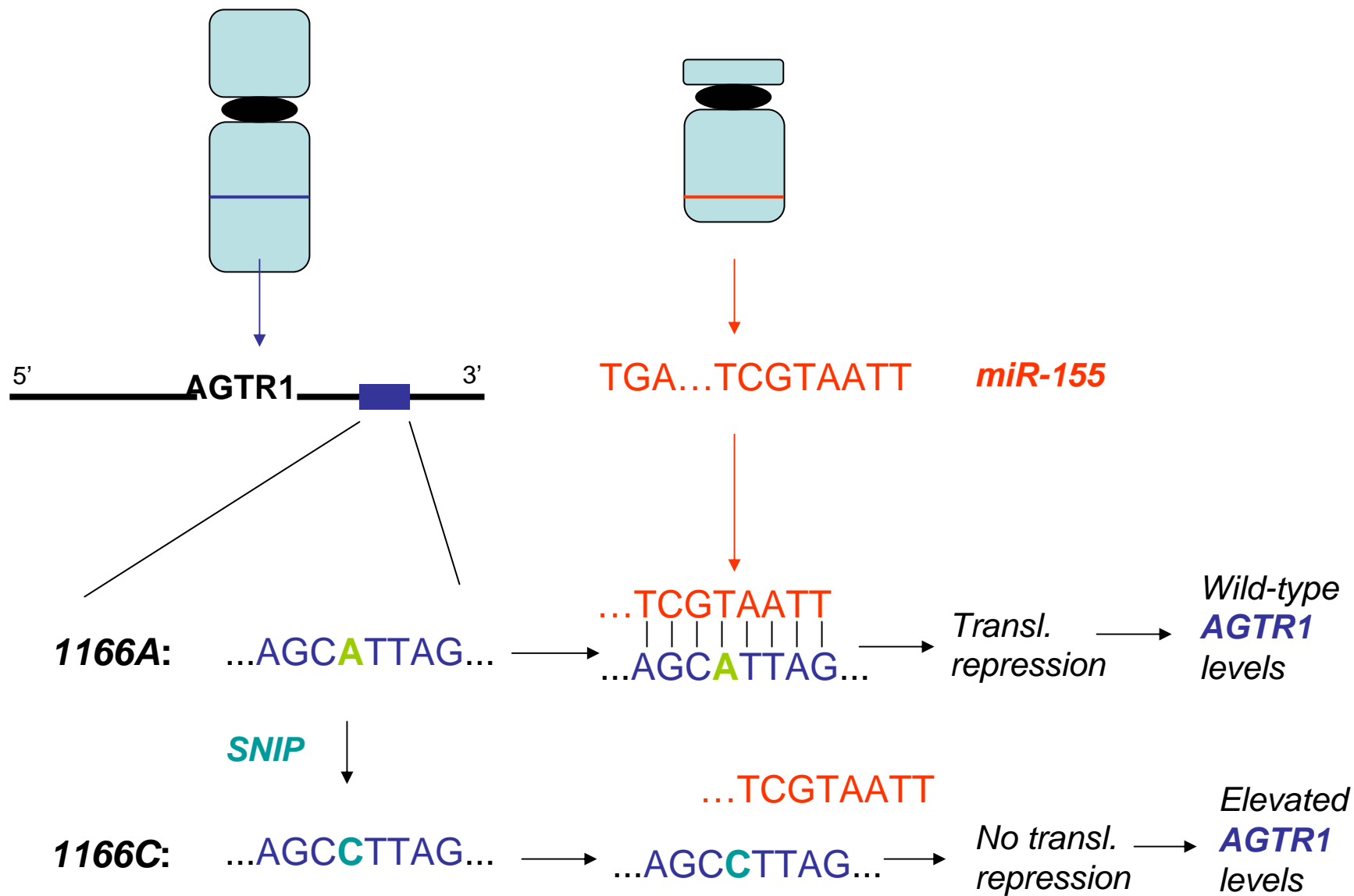
5' UGCCC---CUGGAAAAACUAAAAGAGCCUUGCAUGUACUUGAA 3'      Human SMD1
||||  ||||  |||||
3' UCGGAUAGGACCU-----AAUGAACUU 5'      Hsa-miR-26
|||  ||||  |||||
5' UGCCC---CUGGAAAAACUAAAAGAGCCUUGCAUGUACUUGAA 3'      Polymorphic Human SMD1

```

```

5' CGCGCCGCCUGCGGCACUGCCU 3'      Human DLL1
|.||. ||. |.|||||
3' UGUUGGUCGAUUCUGAGCGGU 5'      Hsa-miR-34
...||. ||. |.|||||
5' CGCGCCGCCUGCGGCACUGCCU 3'      Polymorphic Human DLL1

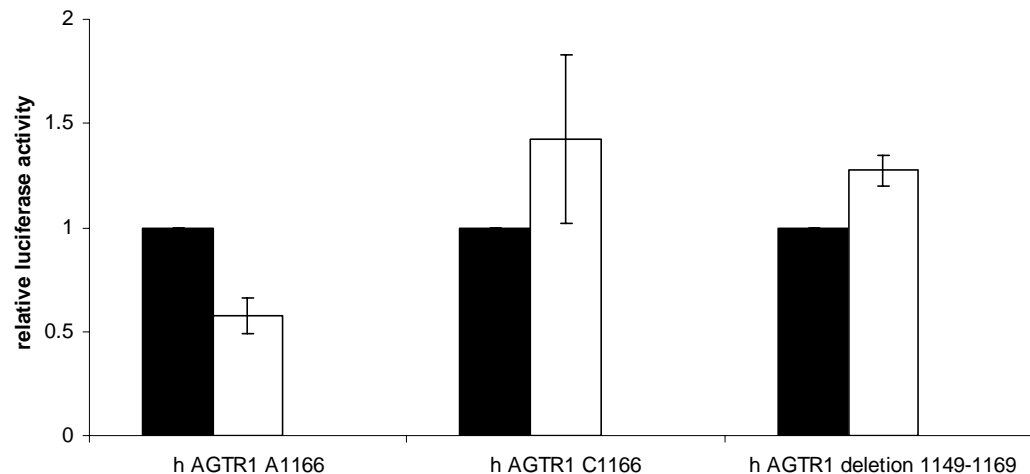
```



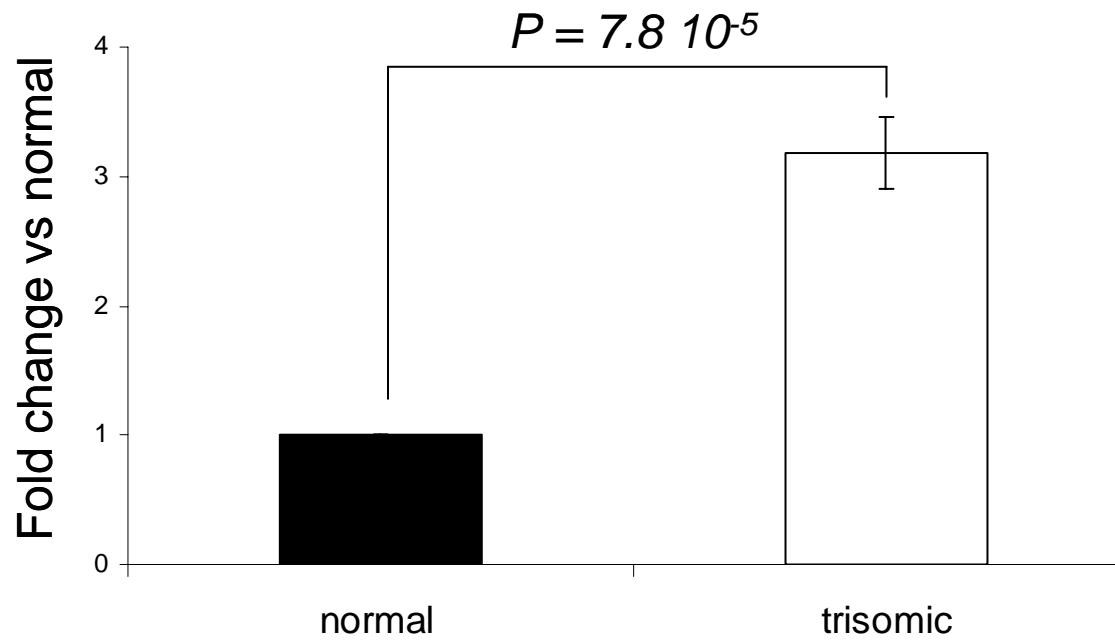
A newly identified role for miR-155 in hypertension

Experimental validation

- In vitro luciferase assay to test the prediction



Sethupathy, P., Borel, C., Gagnebin, M., Grant, G.R, Hatzigeorgiou, A.G, and Antonarakis, S.E. (2006) under review



qRT-PCR for mature miR-155 expression in fibroblast cells from monozygotic twins discordant for trisomy 21.

miRNAs and Cancer Specimens Studied

Human miRNA genes (207) identified from miRNA registry

<http://www.sanger.ac.uk/Software/Rfam/mirna/index.html>

A total of 253 human cancer specimens examined:

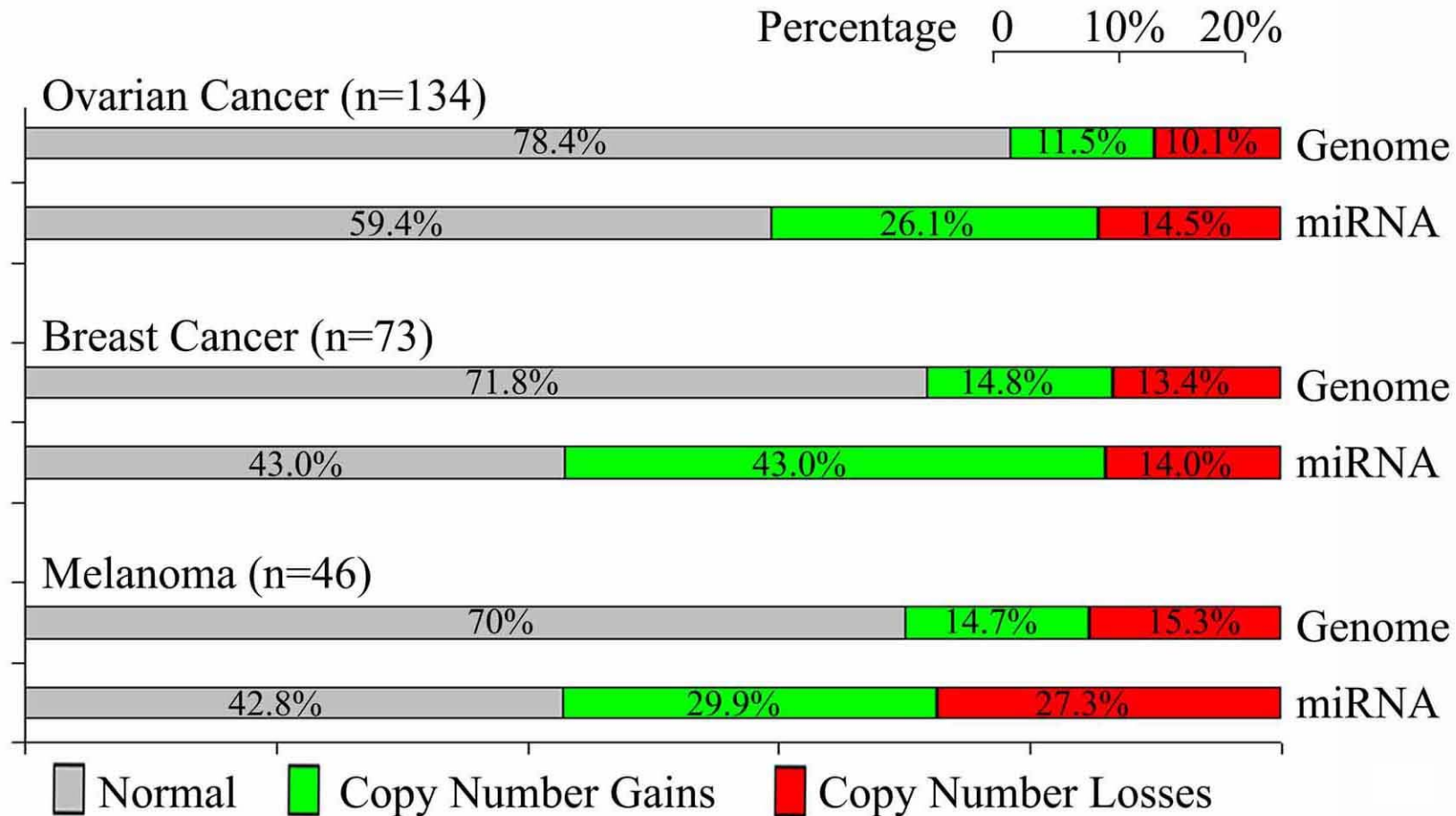
134 ovarian cancer specimens (107 primary tumors and 27 cell lines)

73 breast cancer specimens (55 primary tumors and 18 cell lines)

46 melanoma cell lines



miRNA DNA Copy Number Alterations in Human Cancer

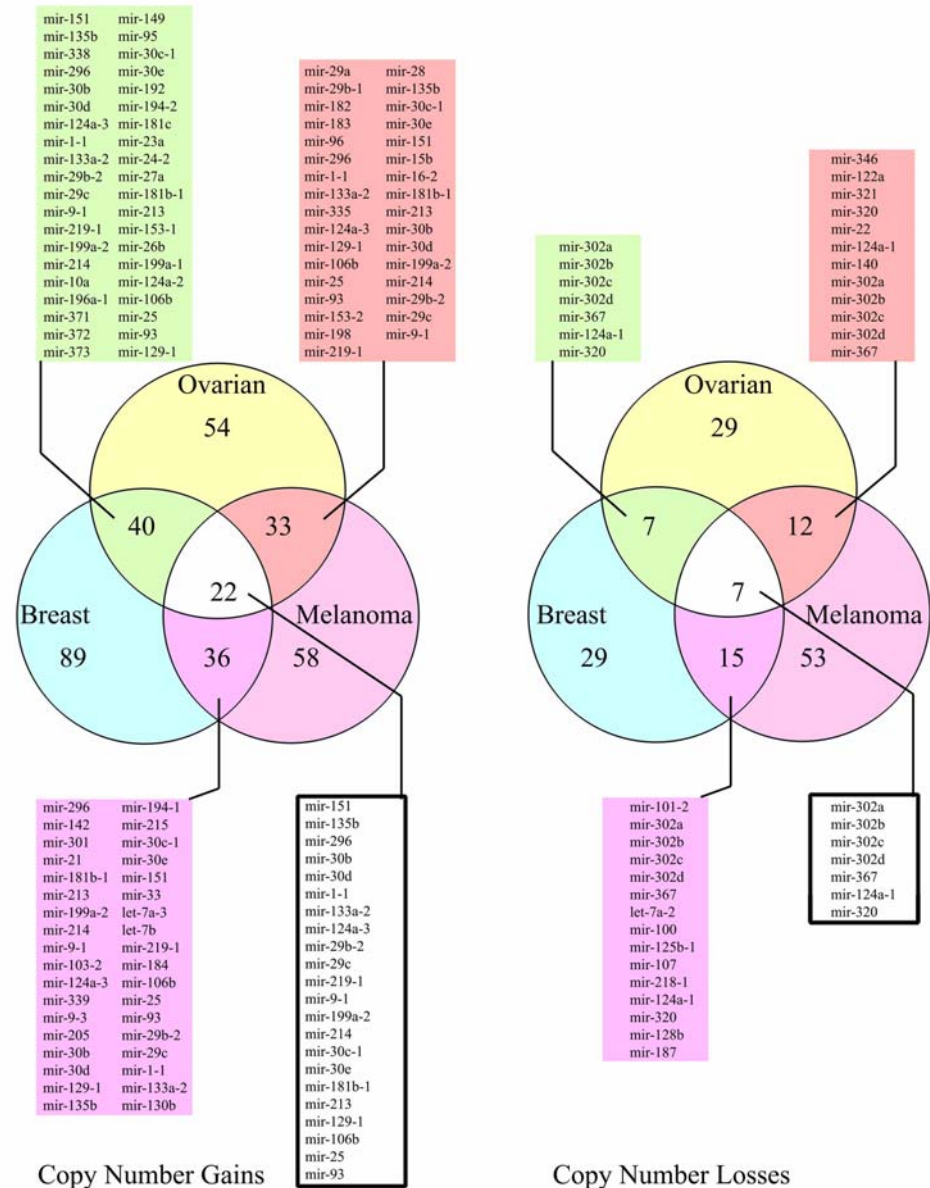


LZhang, J. Huang, N. Yang, J. Greshock, M.S. Megraw, A. Giannakakis, S. Liang, T.L. Naylor, A. Barchetti, M.R. Ward, G. Yao, A. Medina, A. O'brien-Jenkins, D. Katsaros, A. Hatzigeorgiou, P.A. Gimotty, B.L. Weber, and G. Coukos (2006). MicroRNAs exhibit high frequency genomics alterations in human cancers. Proc Natl Acad Sci USA, 103:9136-9141.

miRNAs with Copy Number Changes Shared by 3 Cancer Types

22 miRNA genes with copy number gains and 7 with losses were shared by all three types of cancer.

38 miRNA genes had no copy number change in any of those cancer samples.



DIANA Lab

Molly Megraw, PhD cand.
Praveen Sethupathy, PhD cand.



And. Kouranov, PhD.
P. Fitziev, res. fellow
Benoit Corda, res. Fellow

Support

NSF Career Award Grant
NIH pre-doctoral training grant

Collaborators

Shane Jensen, Dep. of Statistics, Upenn
Krit. Kalandidis, IMBB, FORTH

Martin Reczko, ICS, FORTH
Timothy Hughes, University of Toronto
Aris Economidis, Regeneron Pharmaceutical

St. Antonarakis, University of Geneva
Ch. Borel, University of Geneva

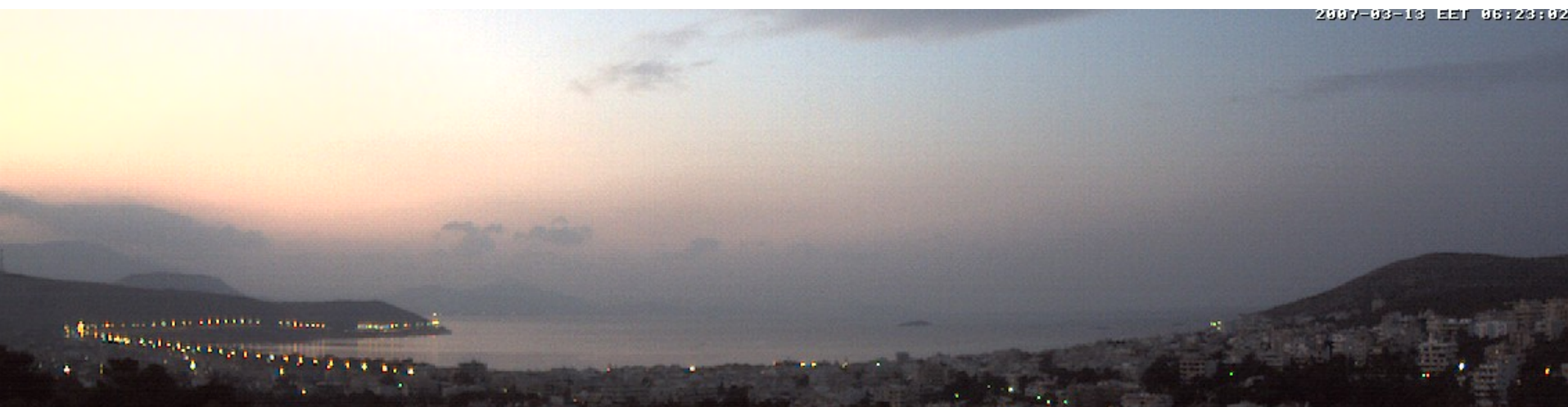
Zis. Mourelatos, Dep. of Pathology, Upenn
Mar. Kiriakidou, Dep. of Pathology, Upenn
Pet. Nielson, Dep. of Pathology, UPenn

Qih. Huang, Wistar Institute, Philadelphia

George Coukos, Dep. of Gyn. Upenn

F. Pereira , A. Bernal

Global Discriminative Training for Higher-Accuracy Computational Gene Prediction" PLoS Computational Biology.(in press)



<http://www.fleming.gr/>

<http://microrna.sanger.ac.uk/>

<http://www.diana.pcbi.upenn.edu/>

<http://pictar.bio.nyu.edu>

<http://genes.mit.edu/targetscan/>

<http://www.microrna.org/>