

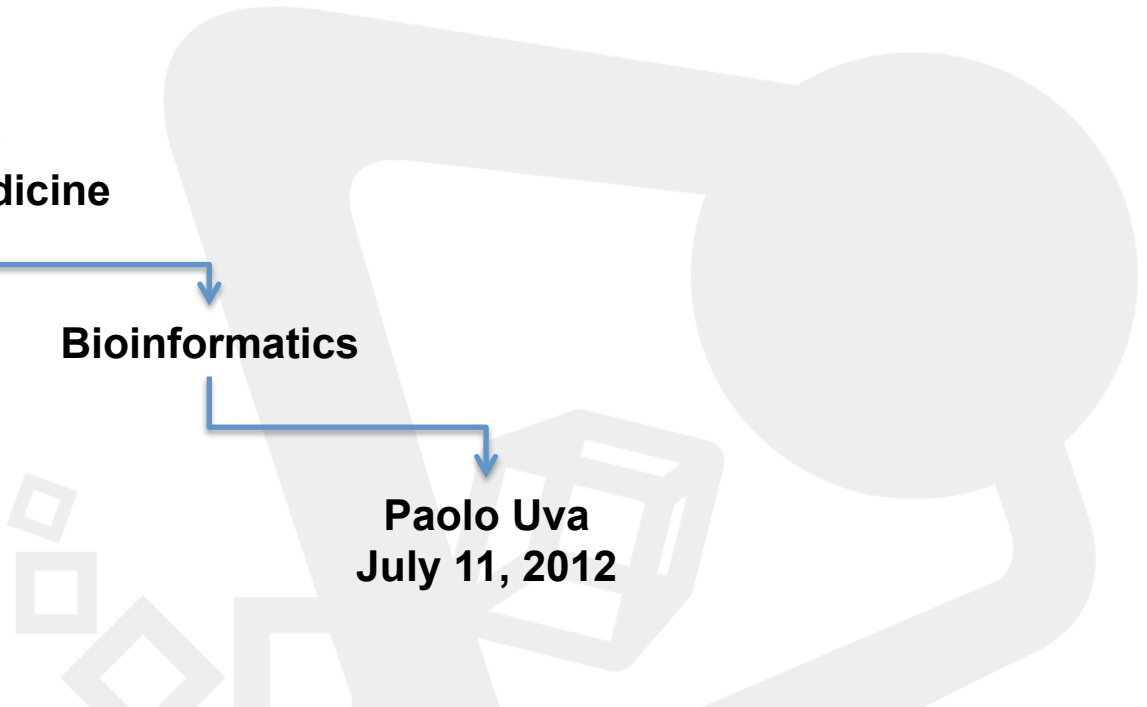
## Inferring the functional role of microRNAs from gene expression data

CRS4

Biomedicine

Bioinformatics

Paolo Uva  
July 11, 2012



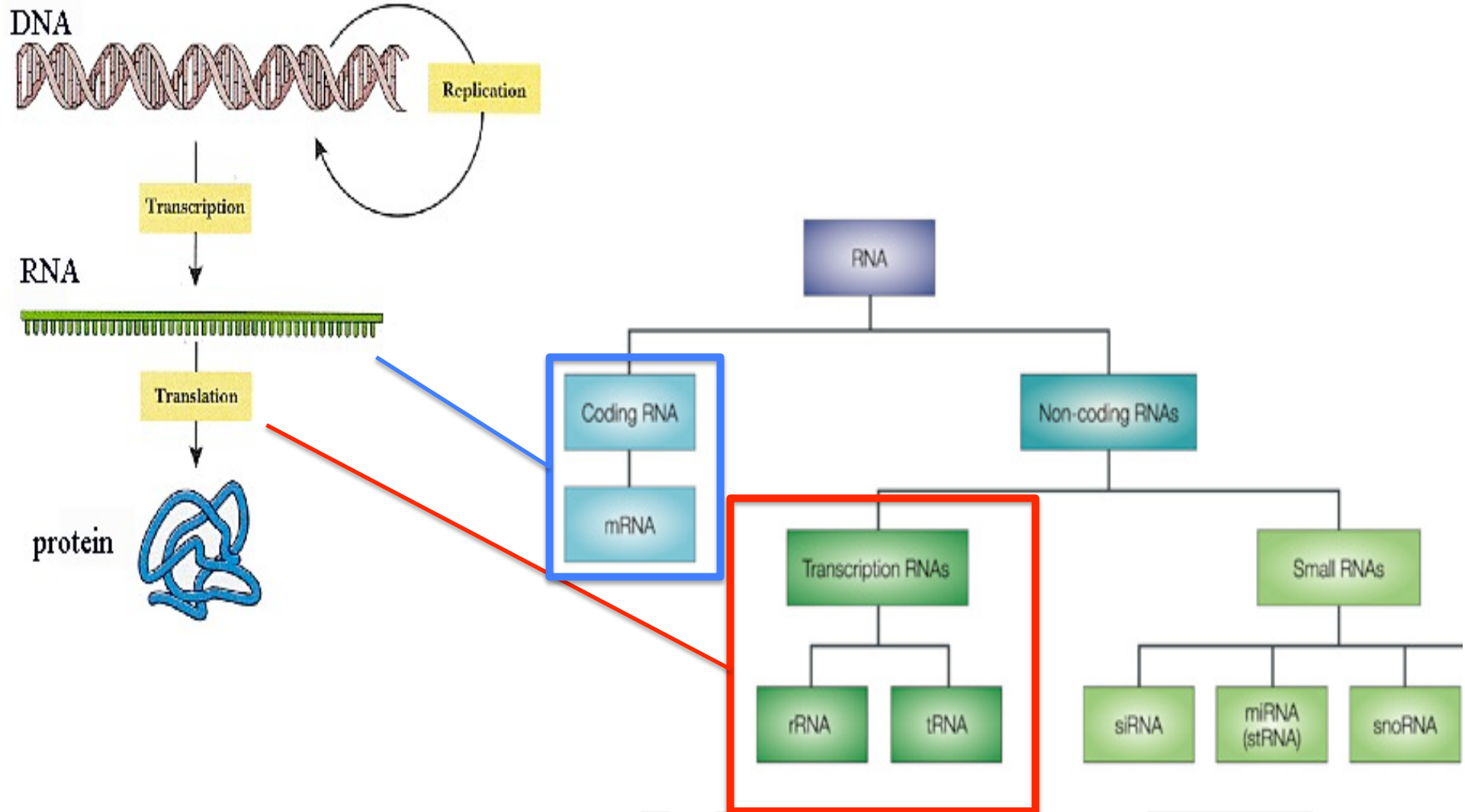
- Pharmaceutical company
- Fondazione San Raffaele, Milan
- IDI-IRCCS, Rome  
(*UO – Ricerca Finalizzata*)
- Humanitas – IRCCS, Milan  
(*UO – Giovani Ricercatori*)



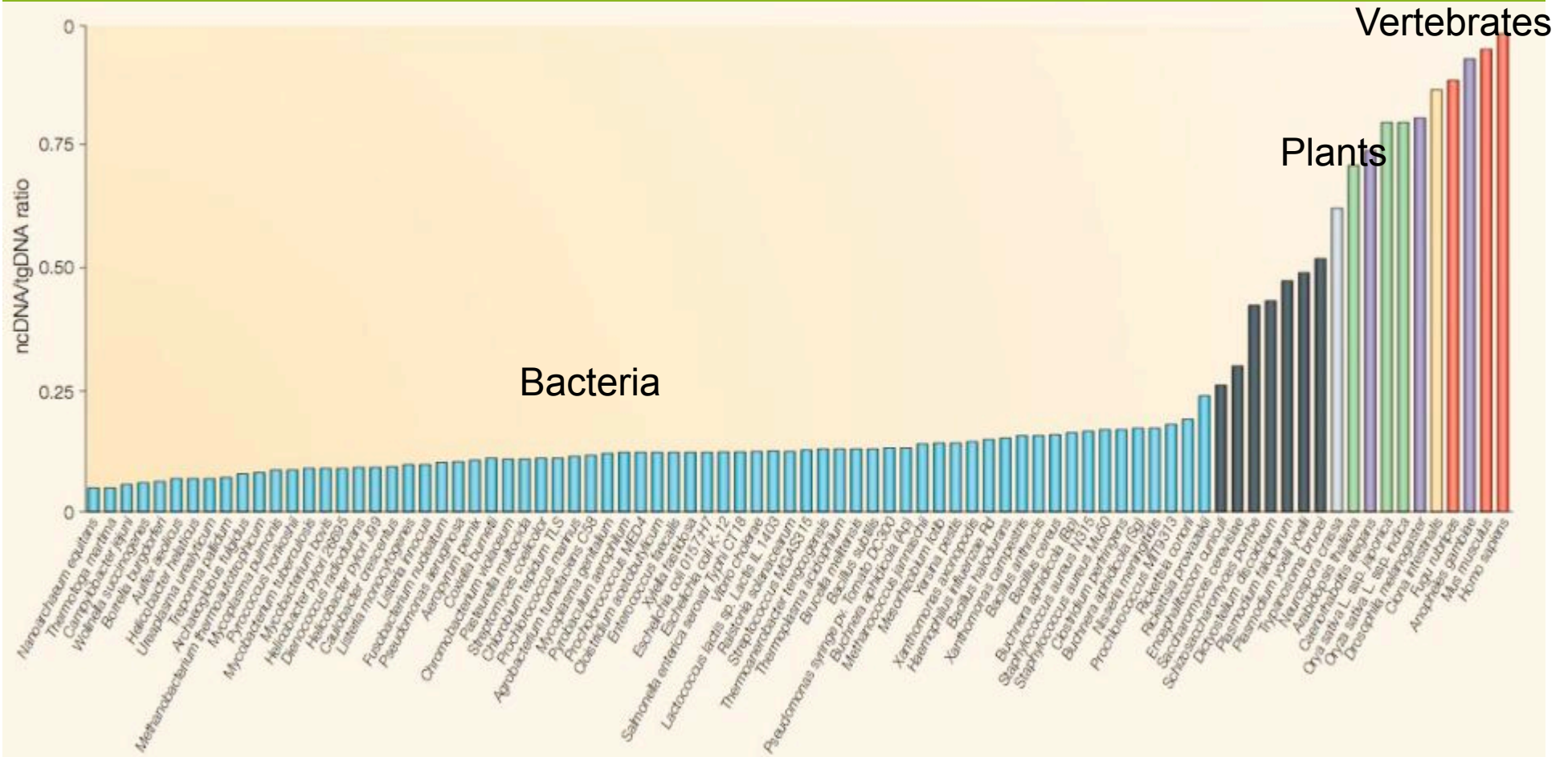
ISTITUTO CLINICO  
HUMANITAS



- First part is about “what is a microRNA?”
- Second part is about discovery of microRNAs and their “targets”
- Third part is about a tool developed “in-house” applied to a real case study



# Ratio of Non-Protein-Coding DNA to Protein-Coding DNA During Evolution





## microRNAs play important roles

- "I think that virtually every gene, at some point in the growth and development of the animal, is going to be found to be influenced by microRNAs."

David P. Bartel (*1k citations/year in last 10y*)



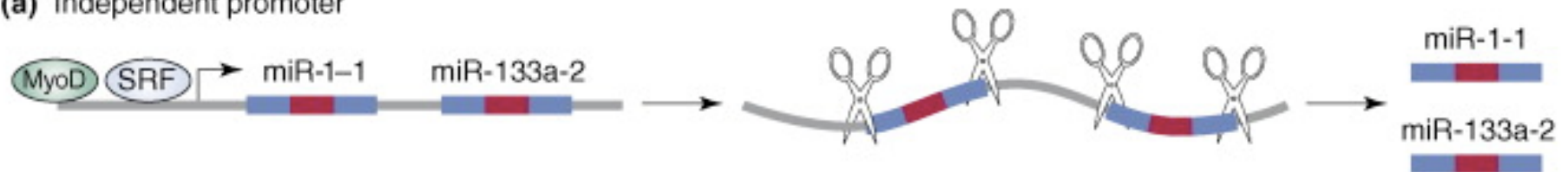
**1962.** The Nobel Prize in Physiology or Medicine to Watson, Crick and Wilkins for their discovery of **the molecular structure of DNA**

**1993.** discovery of first non-coding gene.  
Mechanism not understood

**2006.** The Nobel Prize in Physiology or Medicine to **Andrew Z. Fire** and **Craig C. Mello** for their discovery of **"RNA interference – gene silencing by double-stranded RNA"**

# Genomic Organization of miRNA Genes

(a) Independent promoter



(b) Intronic

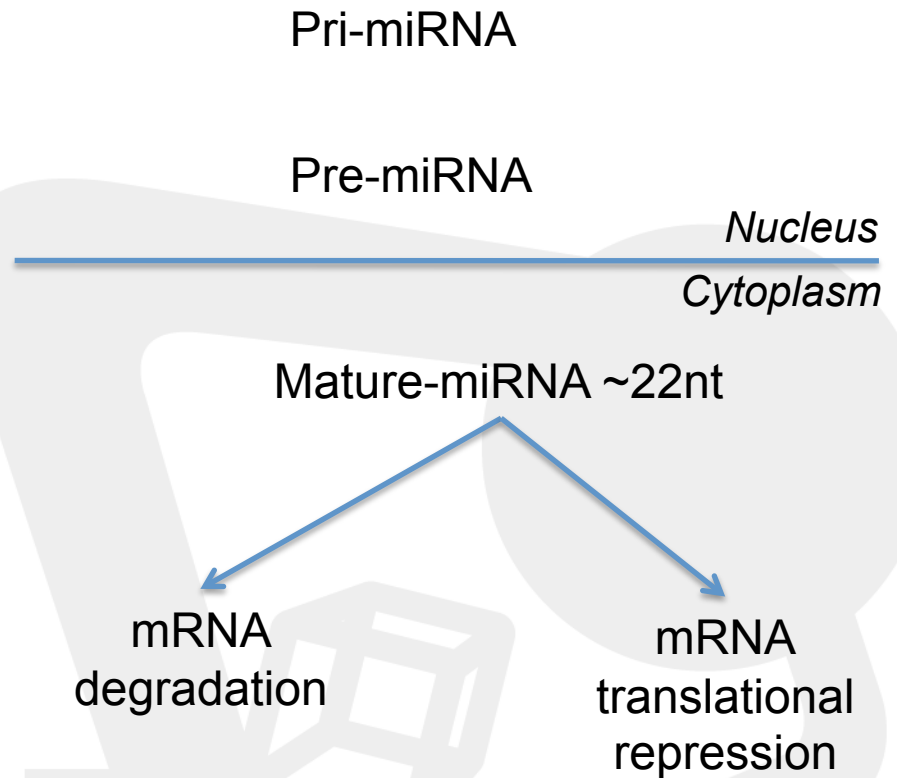
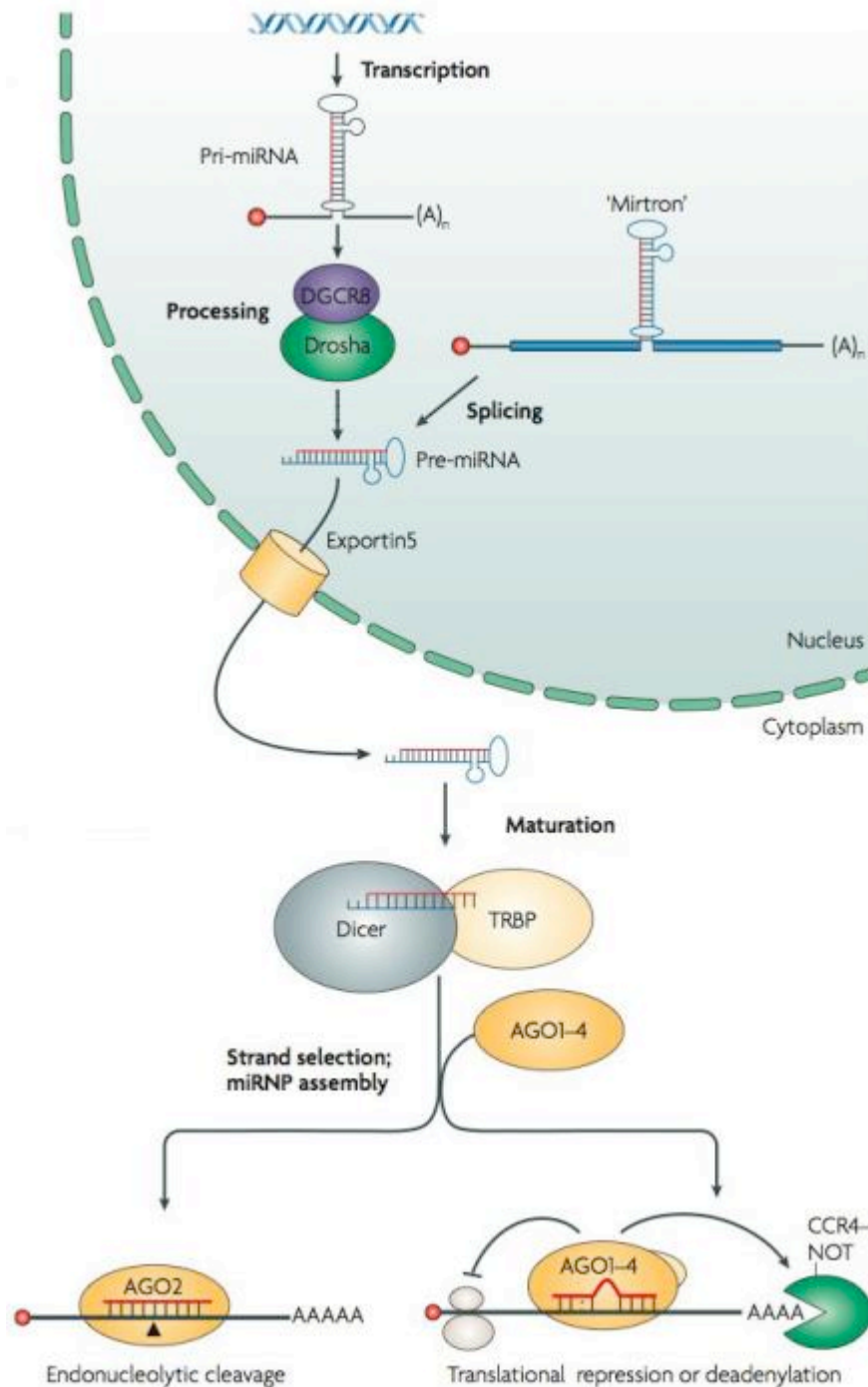


(c) Exonic

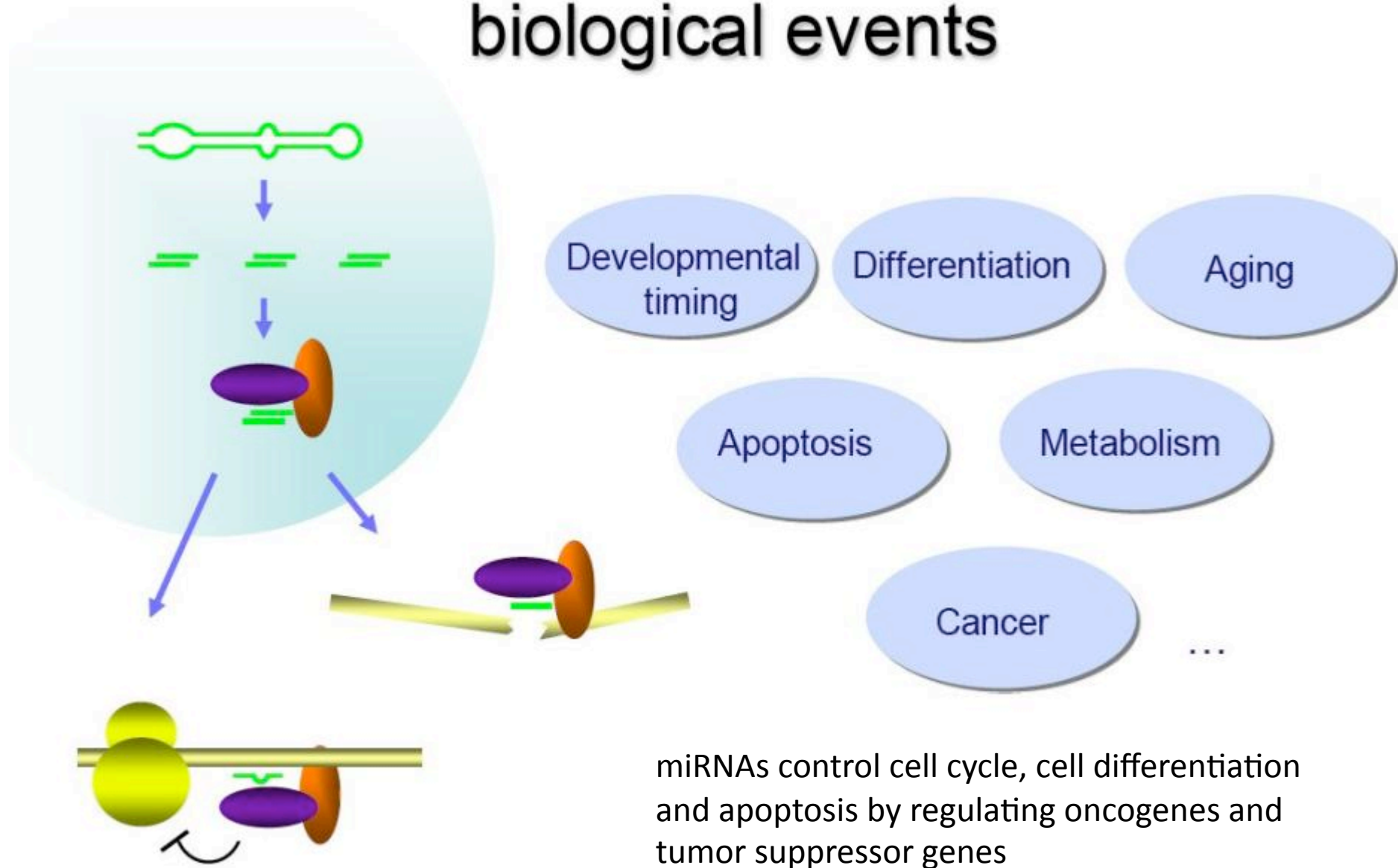




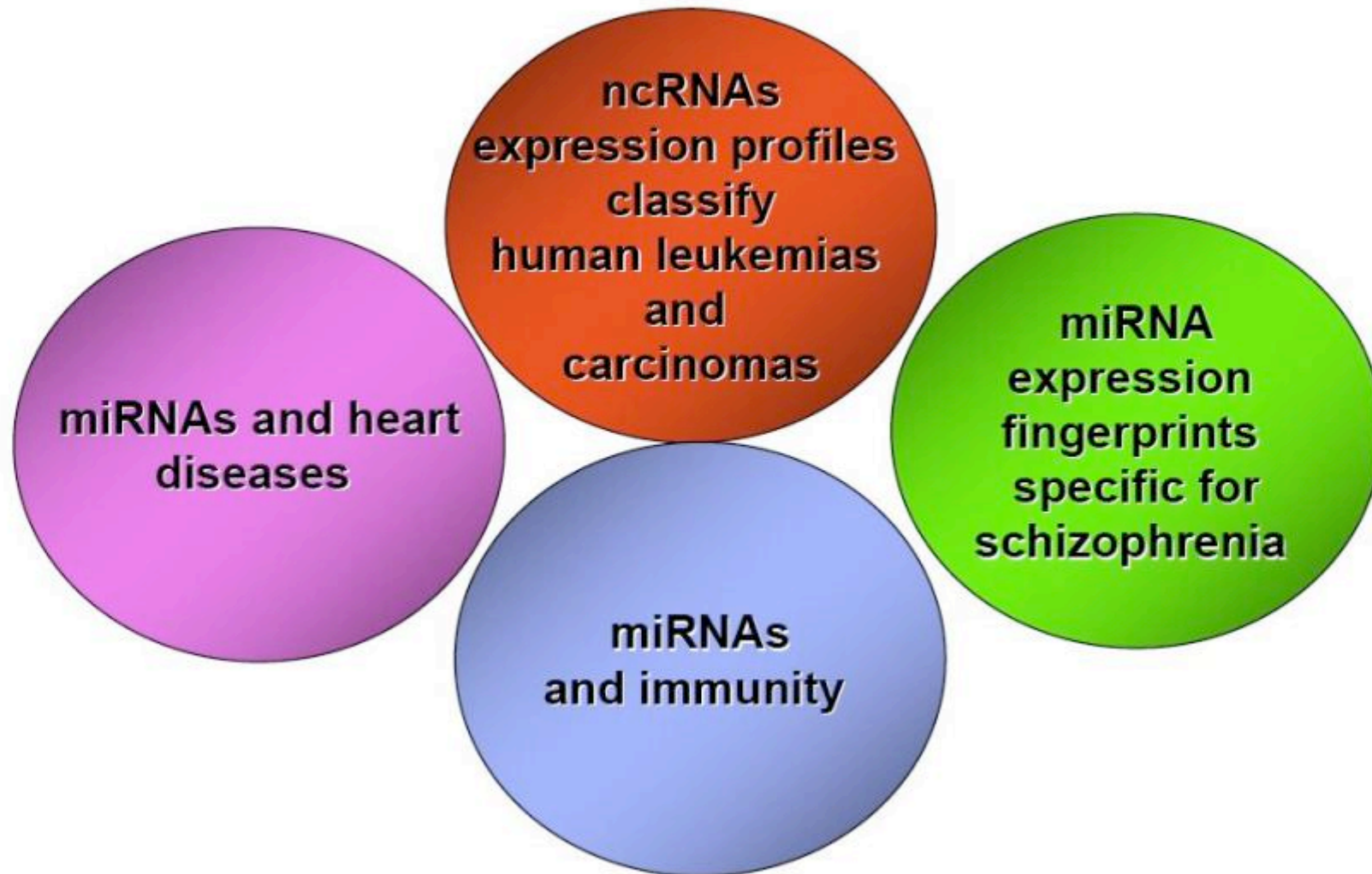
# Biogenesis of microRNAs



# Thousands of microRNAs act in multiple biological events

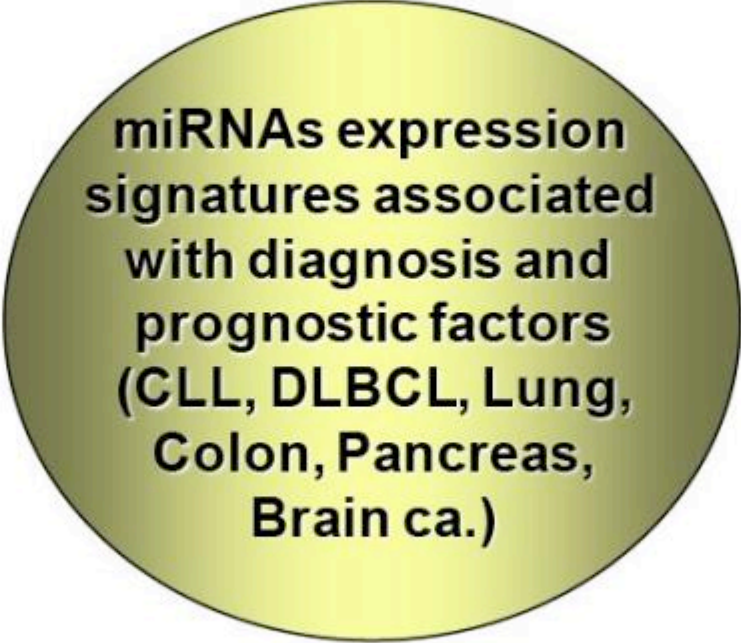


***ALTERATIONS OF NONCODING RNAS ARE FOUND IN EVERY TYPE OF HUMAN DISEASE***



(Calin et al, PNAS 2002; Lu et al, Nature, 2005; Landgraf et al, Cell 2007; Perkins et al Genome Biol 2007; Hansen et al PLoS ONE, 2007; Beveridge et al, Hum Molec Genet 2008, Baltimore D, Nat Immunol 2008; van Rooij, Trends Genet, 2008 )

# ***MIRNA PROFILING AS A NEW DIAGNOSTIC & PROGNOSTIC TOOL FOR CANCER PATIENTS***



**miRNAs expression  
signatures associated  
with diagnosis and  
prognostic factors  
(CLL, DLBCL, Lung,  
Colon, Pancreas,  
Brain ca.)**

miRNAs are mis-expressed in cancer and are therefore  
excellent diagnostic/prognostic markers

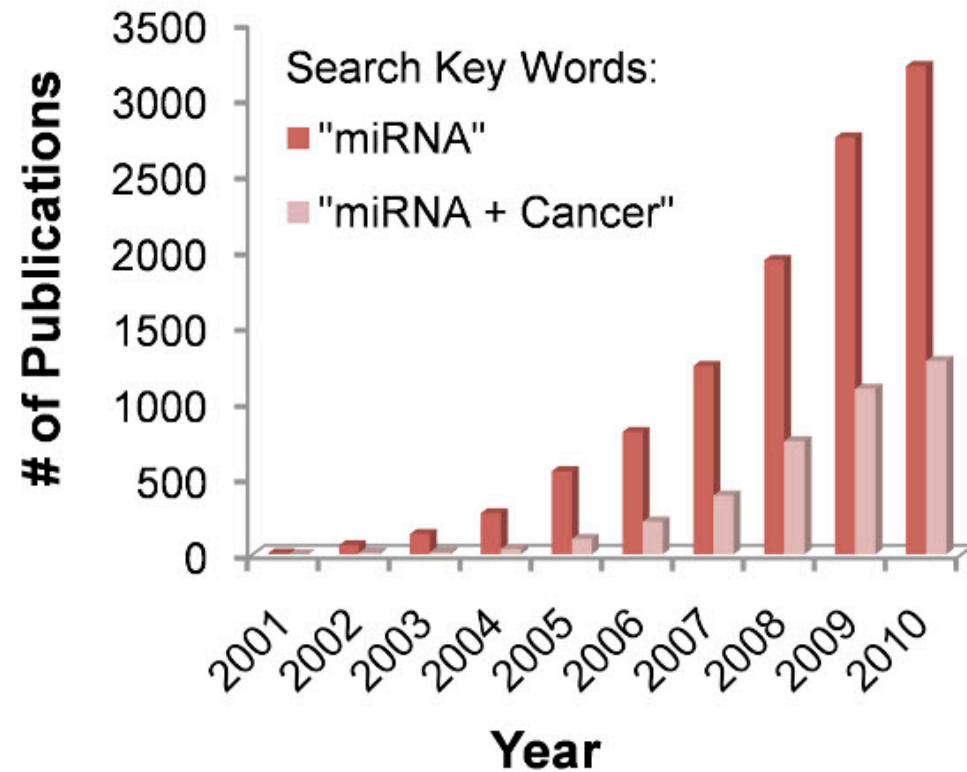
(Michael et al, Molec Cancer Res 2003; Lu et al, Nature, 2005; Eis et al, PNAS, 2005  
Lui et al, Cancer Res 2007, Bloomston et al, JAMA 2007; Mi et al, PNAS, 2007; Garzon et al, Blood in press 2008)

© [George Calin](#)

Some miRNAs e.g. *mir-155*, can cause cancer and oncogenic miRNAs may be therapeutic targets in cancer

Other miRNAs like *let-7*, may prevent cancer and may be therapeutic molecules themselves.

MicroRNAs could augment current cancer therapies.



- Cercopithecoidea
- Hominidae
  - [Gorilla gorilla](#) (85)
  - [Homo sapiens](#) [GRCh37] (1527) ←
  - [Pan paniscus](#) (88)
  - [Pongo pygmaeus](#) [PPYG2] (581)
  - [Pan troglodytes](#) [PanTro2.1] (600)
  - [Symphalangus syndactylus](#) (11)
- Lemuridae
- Prototheria
- Rodentia
  - [Cricetulus griseus](#) (2)
  - [Mus musculus](#) [NCBIM37] (741) ←
  - [Rattus norvegicus](#) [RGSC3.4] (408)

Differences can be due to:

- biological reasons
- **lack of experimental evidences**

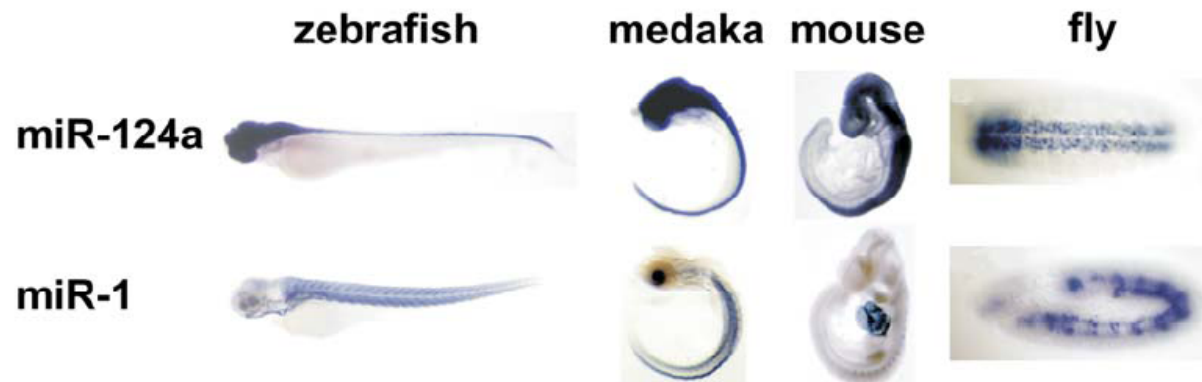
- ...what miRNAs are
- ...that miRNAs have important roles
- ...that miRNAs regulate gene expression by binding to the mRNA of the target genes

***Next: How to find miRNAs and their targets***



# How to find miRNAs?

- Experimental methods



- Computationally:
  - look for evolutionarily conserved sequences
  - check if some of these fold well into the stem-loop structure (“hairpins”) associated with miRNAs



lin-4 family

UCCCUAGAGA . . . CCUAAACUUGGA Hs miR-125b-1
UCCCUAGAGA . . . CCUAAACUUGGA Hs miR-125b-2
UCCCUAGAGA . . . CCUAAACUUGGA Ce lin-4
UCCCUAGAGA . . . CCUAAACUUGGA Ce miR-237

let-7 family

AGAGGUAGUAGGUGGCAUAGU . . . Hs let-7d
UGAGGUAGUAGGUGGCAUAGU . . . Hs let-7e
UGAGGUAGUAGGUGGCAUAGU . . . Hs let-7a-1
UGAGGUAGUAGGUGGCAUAGU . . . Hs let-7a-2
UGAGGUAGUAGGUGGCAUAGU . . . Hs let-7a-3
UGAGGUAGUAGGUGGCAUAGU . . . Hs let-7a-4
UGAGGUAGUAGGUGGCAUAGU . . . Ce let-7
UGAGGUAGUAGGUGGCAUAGU . . . Hs let-7f-1
UGAGGUAGUAGGUGGCAUAGU . . . Hs let-7f-2
UGAGGUAGUAGGUGGCAUAGU . . . Hs miR-98
UGAGGUAGUAGGUGGCAUAGU . . . Hs let-7g
UGAGGUAGUAGGUGGCAUAGU . . . Hs let-7i
UGAGGUAGUAGGUGGCAUAGU . . . Hs let-7b
UGAGGUAGUAGGUGGCAUAGU . . . Hs let-7c
UGAGGUAGUAGGUGGCAUAGU . . . Hs miR-196-1
UGAGGUAGUAGGUGGCAUAGU . . . Hs miR-196-2
UGAGGUAGUAGGUGGCAUAGU . . . Ce miR-84
UGAGGUAGUAGGUGGCAUAGU . . . Ce miR-48
UGAGGUAGUAGGUGGCAUAGU . . . Ce miR-241

mir-1 family

UGGAAUUGUAAAGAAAGUAGUUA Hs miR-1b
UGGAAUUGUAAAGAAAGUAGUUA Hs miR-1d
UGGAAUUGUAAAGAAAGUAGUUA Ce miR-1
UGGAAUUGUAAAGAAAGUAGUUA Hs miR-206

mir-9 family

UCUUUUUUUUUU . . . CUAGCCUG . . . UAUGA Hs miR-9-1
UCUUUUUUUUUU . . . CUAGCCUG . . . UAUGA Hs miR-9-2
UCUUUUUUUUUU . . . CUAGCCUG . . . UAUGA Ce miR-244

mir-10 family

AACCC . . . UAGAUCCGAAACU . . . UUUU . . . Hs miR-100-1
AACCC . . . UAGAUCCGAAACU . . . UUUU . . . Hs miR-100-2
AACCC . . . UAGAUCCGAAACU . . . UUUU . . . Hs miR-99b
AACCC . . . UAGAUCCGAAACU . . . UUUU . . . Ce miR-57
AACCC . . . UAGAUCCGAAACU . . . UUUU . . . Hs miR-10a
AACCC . . . UAGAUCCGAAACU . . . UUUU . . . Hs miR-10b
AACCC . . . UAGAUCCGAAACU . . . UUUU . . . Hs miR-99a
AACCC . . . UAGAUCCGAAACU . . . UUUU . . . Ce miR-51

mir-19 family

UUUGCAAUUC . . . CU . . . UCAAAAACUGA . . . Hs miR-19a
UUUGCAAUUC . . . CU . . . UCAAAAACUGA . . . Hs miR-19b-1
UUUGCAAUUC . . . CU . . . UCAAAAACUGA . . . Hs miR-19b-2
UUUGCAAUUC . . . CU . . . UCAAAAACUGA . . . Ce miR-254

mir-25 family

UAUUUCACUUUU . . . CCG . . . CUU . . . Hs miR-92-1
UAUUUCACUUUU . . . CCG . . . CUU . . . Hs miR-92-2
UAUUUCACUUUU . . . CCG . . . CUU . . . Ce miR-235
UAUUUCACUUUU . . . CCG . . . CUU . . . Hs miR-25-1
UAUUUCACUUUU . . . CCG . . . CUU . . . Hs miR-25-2
UAUUUCACUUUU . . . CCG . . . CUU . . . Hs miR-32

mir-29 family

UAGCACC AUUUUAAAUCAGU . . . Hs miR-29b-1
UAGCACC AUUUUAAAUCAGU . . . Hs miR-29b-2
UAGCACC AUUUUAAAUCAGU . . . Hs miR-29b-3
UAGCACC AUUUUAAAUCAGU . . . Hs miR-29c
UAGCACC AUUUUAAAUCAGU . . . Hs miR-29a-1
UAGCACC AUUUUAAAUCAGU . . . Hs miR-29a-2
UAGCACC AUUUUAAAUCAGU . . . Ce miR-83

mir-31 family

AAGCAAGAUGUUGGCA . . . U . . . AGC . . . Ce miR-72
AAGCAAGAUGUUGGCA . . . U . . . AGC . . . Hs miR-31
AAGCAAGAUGUUGGCA . . . U . . . AGC . . . Ce miR-73

mir-34 family

AGGCAGUUGU . . . UUA . . . UCUGGUUU . . . Ce miR-34
AGGCAGUUGU . . . UUA . . . UCUGGUUU . . . Hs miR-34
AGGCAGUUGU . . . UUA . . . UCUGGUUU . . . Hs miR-122a

mir-50 family

UGAUUUGUAAUUCU . . . UCUUUACAG . . . Ce miR-62
UGAUUUGUAAUUCU . . . UCUUUACAG . . . Ce miR-50
UGAUUUGUAAUUCU . . . UCUUUACAG . . . Hs miR-190
UGAUUUGUAAUUCU . . . UCUUUACAG . . . Ce miR-90

mir-74 family

UGG . . . A . . . AGAA . . . AGGCAGUUC . . . Hs miR-185
UGG . . . A . . . AGAA . . . AGGCAGUUC . . . Ce miR-74

mir-76 family

UCGU . . . UUUU . . . AU . . . UAAAGCCUUGA Ce miR-76
UCGU . . . UUUU . . . AU . . . UAAAGCCUUGA Hs miR-187

mir-79 family

AUAAAGCUUAC . . . U . . . UACCAAAACU . . . Ce miR-79
AUAAAGCUUAC . . . U . . . UACCAAAACU . . . Hs miR-131
AUAAAGCUUAC . . . U . . . UACCAAAACU . . . Ce miR-75

mir-80 family

UGAGAUCAUC . . . CU . . . UAAAGCCUUGA Ce miR-81
UGAGAUCAUC . . . CU . . . UAAAGCCUUGA Ce miR-82
UGAGAUCAUC . . . CU . . . UAAAGCCUUGA Ce miR-80
UGAGAUCAUC . . . CU . . . UAAAGCCUUGA Hs miR-143

mir-105 family

UCAAAUUC . . . UCA . . . GACUCCUUU . . . Hs miR-105-1
UCAAAUUC . . . UCA . . . GACUCCUUU . . . Hs miR-105-2
UCAAAUUC . . . UCA . . . GACUCCUUU . . . Ce miR-232

mir-124 family

UAAAGCCACCG . . . CU . . . UAAAGCCUUGA . . . Hs miR-124a
UAAAGCCACCG . . . CU . . . UAAAGCCUUGA . . . Hs miR-124a
UAAAGCCACCG . . . CU . . . UAAAGCCUUGA . . . Hs miR-124a
UAAAGCCACCG . . . CU . . . UAAAGCCUUGA . . . Ce miR-124
UAAAGCCACCG . . . CU . . . UAAAGCCUUGA . . . Ce miR-228
UAAAGCCACCG . . . CU . . . UAAAGCCUUGA . . . Hs miR-183

mir-133 family

UUUUUUUUUU . . . UCA . . . ACCAGCCUUGA Hs miR-133a-1
UUUUUUUUUU . . . UCA . . . ACCAGCCUUGA Hs miR-133a-2
UUUUUUUUUU . . . UCA . . . ACCAGCCUUGA Hs miR-133b
UUUUUUUUUU . . . UCA . . . ACCAGCCUUGA Ce miR-245

mir-137 family

UAUUUCUU . . . CU . . . UCAAAAACUGA . . . Ce miR-234
UAUUUCUU . . . CU . . . UCAAAAACUGA . . . Hs miR-137

mir-141 family

UAUUUCUU . . . CU . . . UCAAAAACUGA . . . Ce miR-236
UAUUUCUU . . . CU . . . UCAAAAACUGA . . . Hs miR-141

mir-193 family

UAUUUCUU . . . CU . . . UCAAAAACUGA . . . Ce miR-240
UAUUUCUU . . . CU . . . UCAAAAACUGA . . . Hs miR-193

mir-220 family

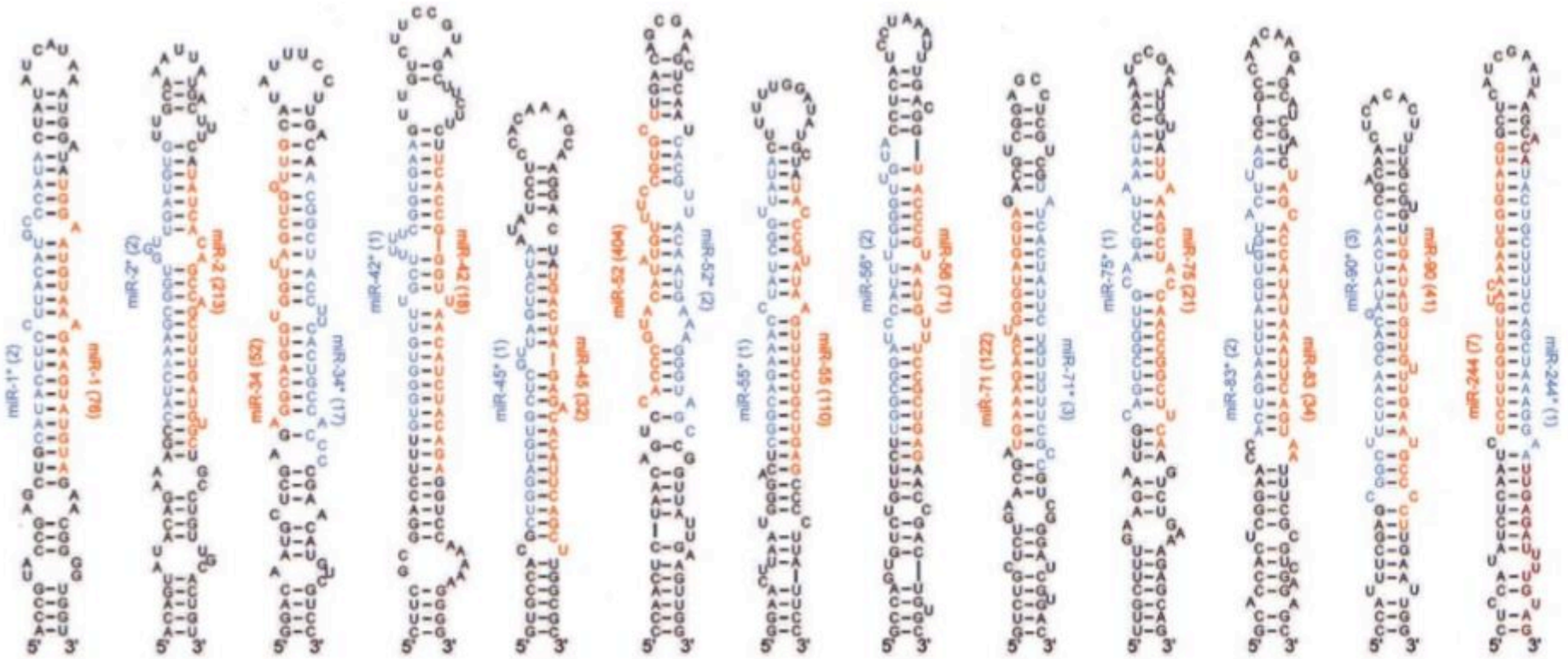
UAUUUCUU . . . CU . . . UCAAAAACUGA . . . Ce miR-253
UAUUUCUU . . . CU . . . UCAAAAACUGA . . . Hs miR-220



Sequence conservation between C. elegans and Homo sapiens miRNAs



# miRNA precursors form stem-loop structure



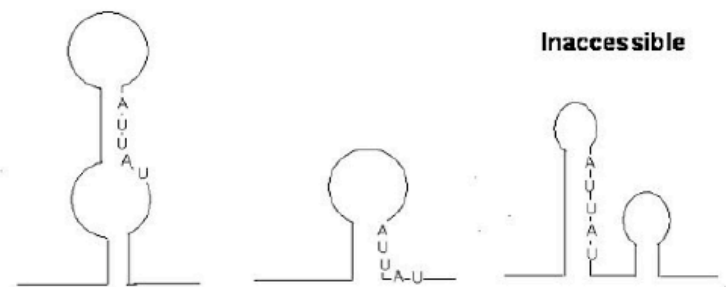
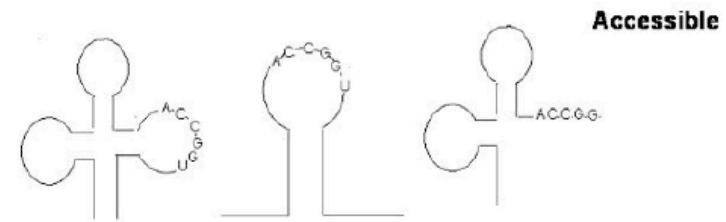
# How to find microRNA targets?

- MicroRNA target sites are located in 3' UTRs
- The target site is short (7-8nt)
- If we allow G-U pairs, mismatches, gaps (bulges), we will find a lot false positives
  - How to remove the false positives?

# How to make more accurate predictions

- Incorporating mRNA UTR structure to predict site accessibility
- Conservation of the target site - *if a region has a functional role should be conserved*

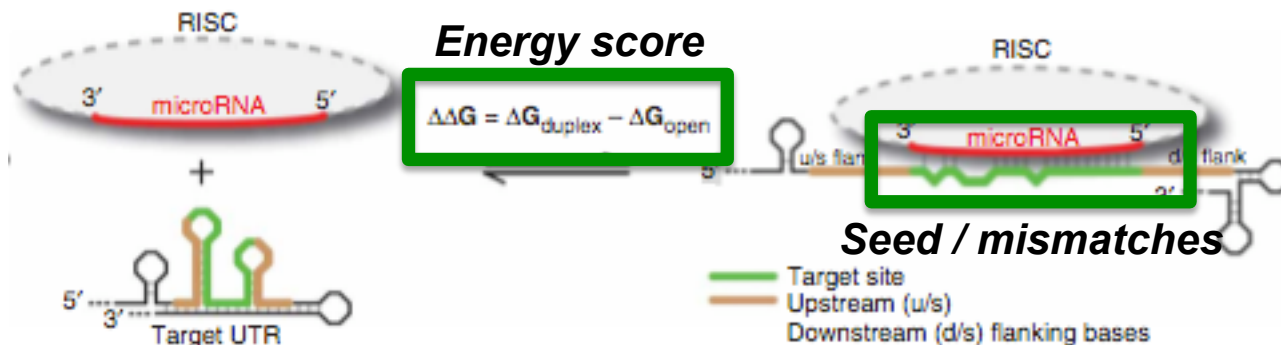
## mRNA UTR



- Learned how to identify miRNAs
- and how the miRNA target genes can be predicted computationally

***Next: A tool for mRNA-microRNA analysis***

# Finding microRNA targets



**A modified version of the original PITA script (Nat Gen 2007):**

- + run locally
- + includes site accessibility
- slow (perl)

Check whether the predicted binding site is conserved

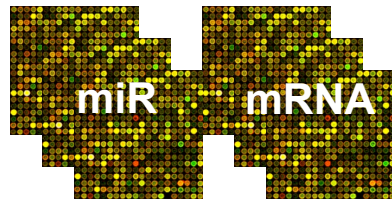
**Filter for target site conservation**

We run the script (embarrassingly parallel) at each release of miRBase (twice/year). The last run generated ~1.5 million of potential miRNA-microRNA interactions

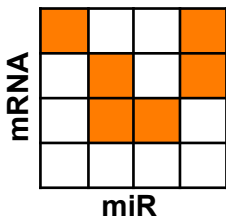
Only a few hundreds of interactions are currently known (from experiments)

# Improving target predictions using expression data

## Expression data

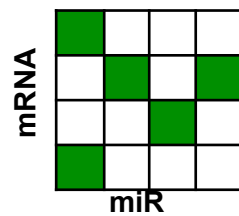


## Correlation matrix

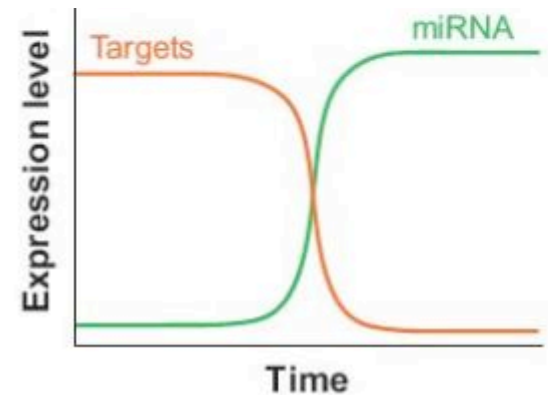
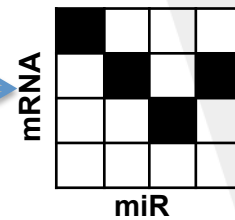


+

## Target predictions



## mRNA-miR adjac matrix



The methodology has been implemented as R package



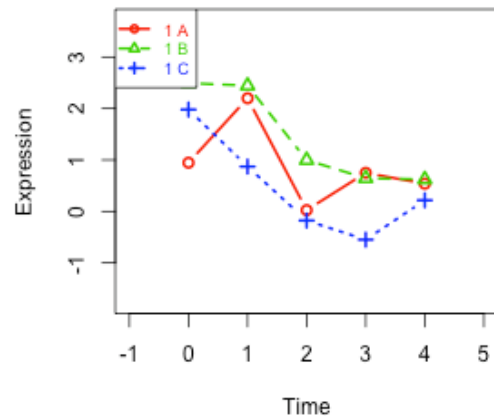


1

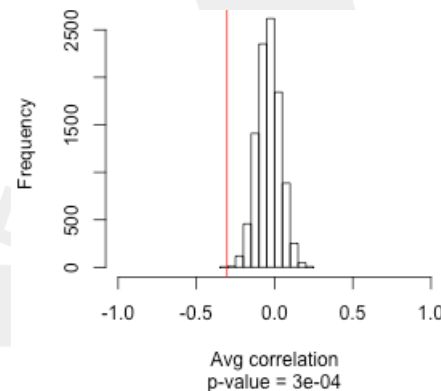
Rank	miR	# targets	# DE targets	P-value
1	<u>mmu-miR-26a</u>	618	69	3e-04
2	<u>mmu-miR-142-3p</u>	225	11	7e-04
3	<u>mmu-miR-101a*</u>	661	26	0.0045
...	...	...	...	...

## Ex: mmu-miR-26a

Expression of individual replicates



Does miR affect target gene expression ?



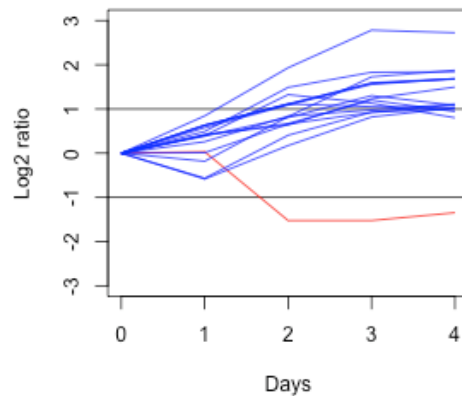
P-value is based on the number of random permutations with avg miR-target correlation lower than the observed value

1

Rank	miR	# targets	# DE targets	P-value
1	<u>mmu-miR-26a</u>	618	69	3e-04
2	<u>mmu-miR-142-3p</u>	225	11	7e-04
3	<u>mmu-miR-101a*</u>	661	26	0.0045
...				

2

mmu-miR-26a  
Cor < -0.7 - 15 DE targets



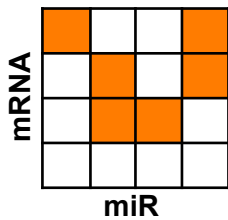
Probeset	EntrezID	GeneSymbol	miR-PCC
10435714	67846	Tmem39a	-0.95
10423577	67154	Mtdh	-0.94
10443527	18712	Pim1	-0.93
10416290	72549	Reep4	-0.92
10459481	70361	Lman1	-0.91
...	...	...	...

**Statistical evaluation of individual miRNAs is summarized in an interactive HTML report**

## Expression data

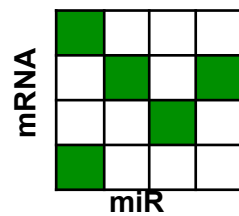


## Correlation matrix

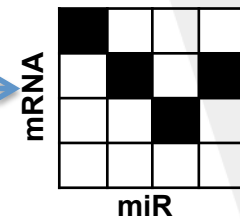


+

## Target predictions



## mRNA-miR adjac matrix



## Questions

**Is my miR 'working'?**

*Report with statistical evaluation of individual miR*

**Which pathways are associated to miR expression?**  
*mirR Functional Maps*

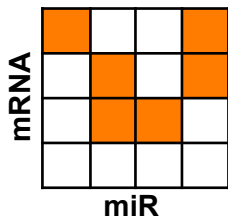


The methodology has been implemented as R package

## Expression data

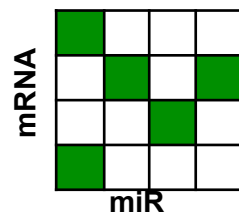


### Correlation matrix

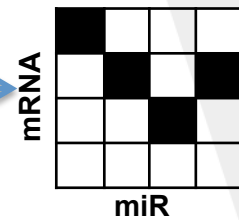


+

### Target predictions



### mRNA-miR adjac matrix



## Questions

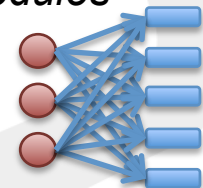
**Is my miR 'working'?**

*Report with statistical evaluation of individual miR*

**Which pathways are associated to miR expression?**  
*miR Functional Maps*



**Do miRs cooperate in regulating gene expression?**  
*miR Regulatory Modules*



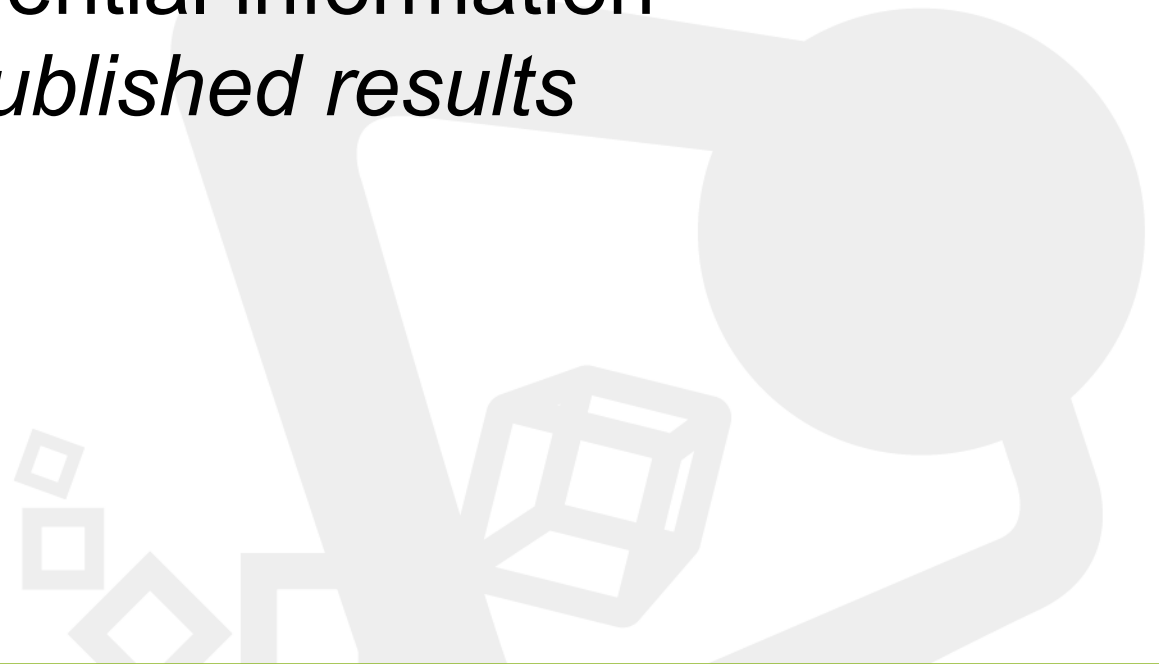
The methodology has been implemented as R package



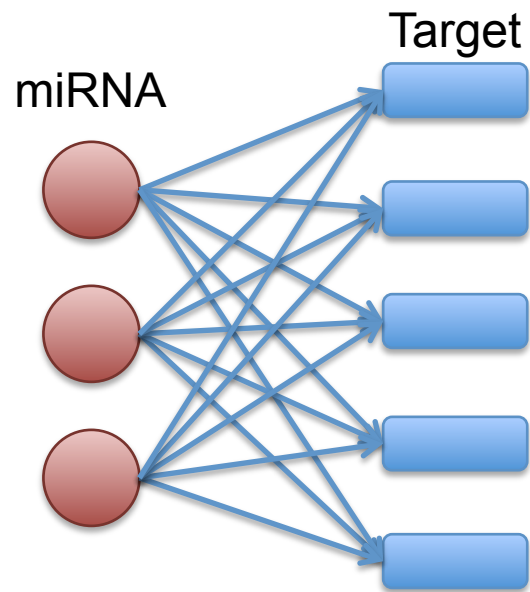


# Case study

Confidential information  
*Unpublished results*



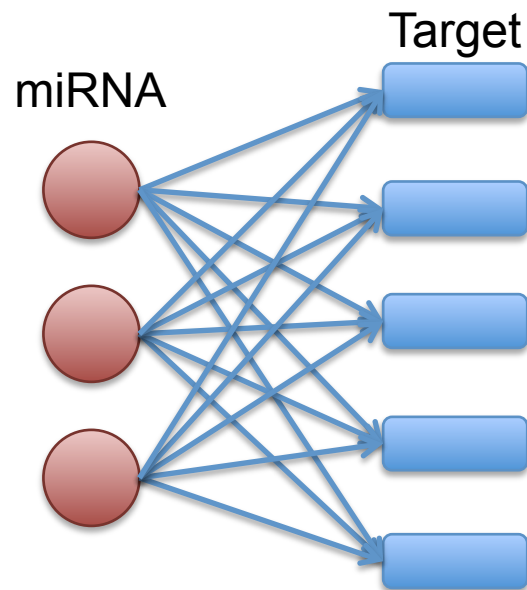
## miRNA-synergism



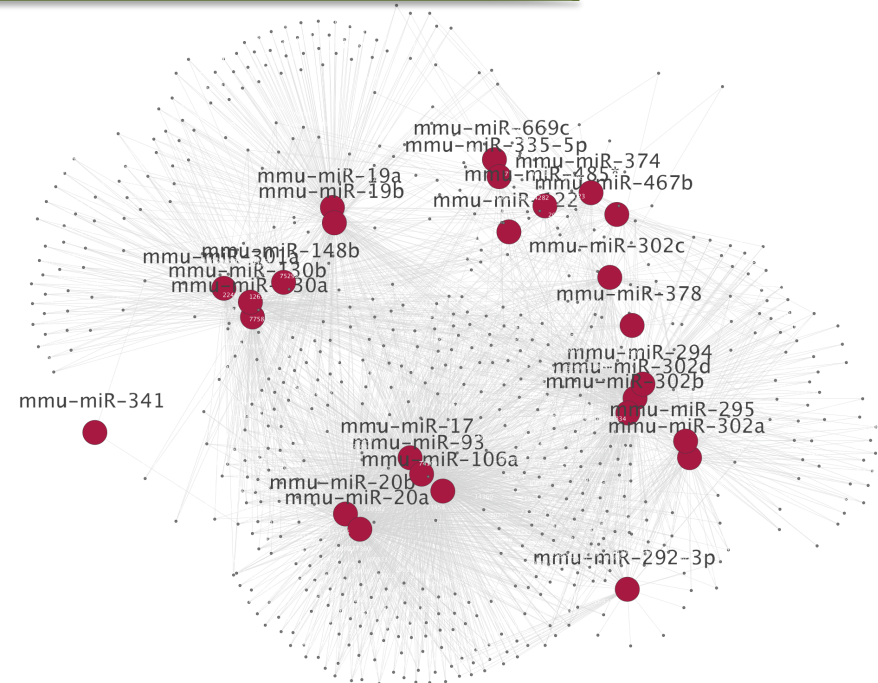
**Biclique**



## miRNA-synergism



**Biclique**



**Community**





# Keeping updated...



**Nobel  
Week  
Dialogue**

*Science & Society*

2012

**The Genetic Revolution  
and its Impact on Society**

9th December  
STOCKHOLM

**151 : 00 : 45**

Days

Hours

Minutes

[Home](#)

[Contact](#)

[➔ Sign up for more info](#)

**SEARCH** ➔

[EVENT INFO](#)

[PROGRAMME](#)

[PARTICIPANTS](#)

[REGISTER](#)

[PARTNERS / SPONSORS](#)

[PRESS](#)

[ABOUT US](#)

## Event Info

[➔ Overview](#)

[Venue](#)

[FAQ](#)

[Contact Us](#)

# The Genetic Revolution and its Impact on Society

*A new feature of Nobel Week in Stockholm, the Nobel Week Dialogue provides an opportunity for discussions that cross the traditional boundaries between science and society.*

## VENUE *Details*

**Stockholm City Conference  
Centre**

Folkets Hus  
Barnhusgatan 12-14  
Stockholm  
Sweden

<http://www.stoccc.se>



**Thank you for your  
attention**