# The role and functions of miRNA-mediated circuits in the human regulatory network

**TABIS 20/9/2013**

*Michele Caselle – University of Torino and INFN*
*caselle@to.infn.it*

# Plan of the talk

Introduction: **Gene Regulation and Regulatory Networks**

Three examples:

**mixed Feed Forward Loops (FFL)**

**MiRNA mediated self-loops**

**Sponge loops**

# References


- **D. Cora', A. Re, D. Taverna and M. Caselle**
  *"Genome-Wide Survey of MicroRna-Transcription Factor Feed-Forward Regulatory Circuits in Human"*
  Molecular BioSystems. 2009 Aug; **5**(8):854-67.

- **M.Osella, C. Bosia, D. Cora' and M. Caselle**
  *"The role of incoherent microRNA-mediated FFL in noise buffering*
  PloS Computational Biology (2011) **7**(3): e1001101

- **M. El Baroudi, D. Cora', M.Osella, C. Bosia, and M. Caselle**
  *"A curated database of miRNA mediated Feed Forward Loops involving MYC as Master Regulator"*
  PloS ONE (2011) **6**(3):e14742

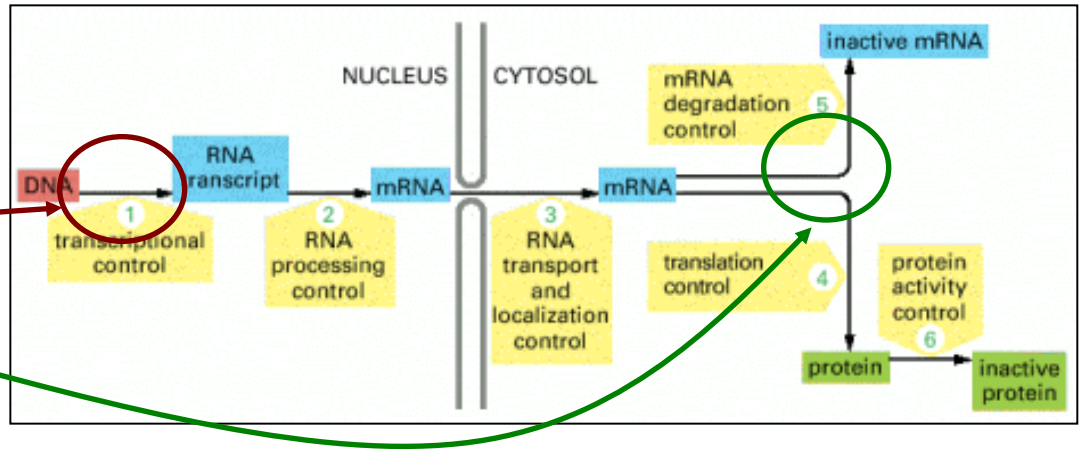
- **C. Bosia, M. Osella, M. El. Baroudi, D. Cora', M. Caselle**
  *"Gene autoregulation via intronic microRNAs and its function"*
  BMC Systems Biology 2012, **6**:131

# Gene Regulation

Gene expression is tightly regulated. All cells in the body carry the full set of genes, but only express about 20% of them at any particular time. Different proteins are expressed in different cells (neurons, muscle cells....) according to the different functions of the cell.

Among the various regulatory steps the most important ones are:

- transcriptional control, by **Transcription Factors.**
- post-transcriptional control, by **microRNAs.**

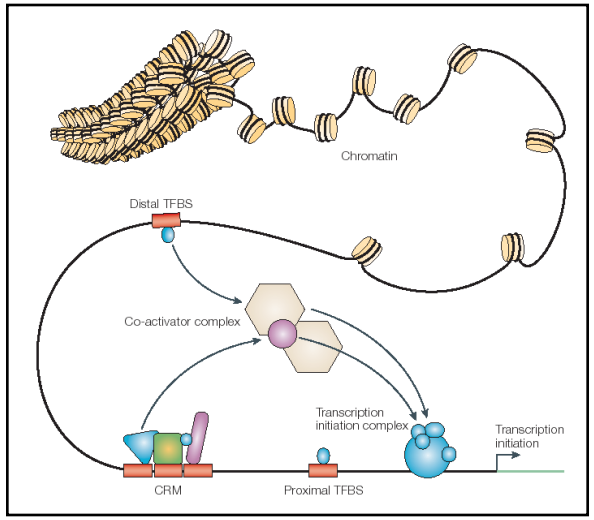



Alberts, *Molecular Biology of the Cell*

# Transcription Factors and miRNAs

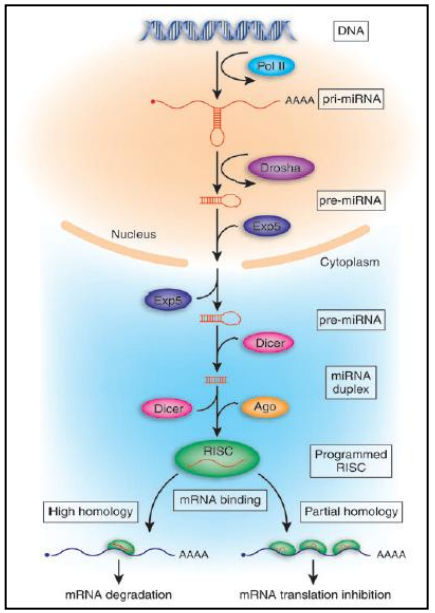
- **Regulation of gene expression** mainly mediated by:

**Transcription Factors (TFs)**: proteins binding to specific recognition **motifs (TFBSs)** usually short (5-10 bp) and located **upstream** of the coding region of the regulated gene.

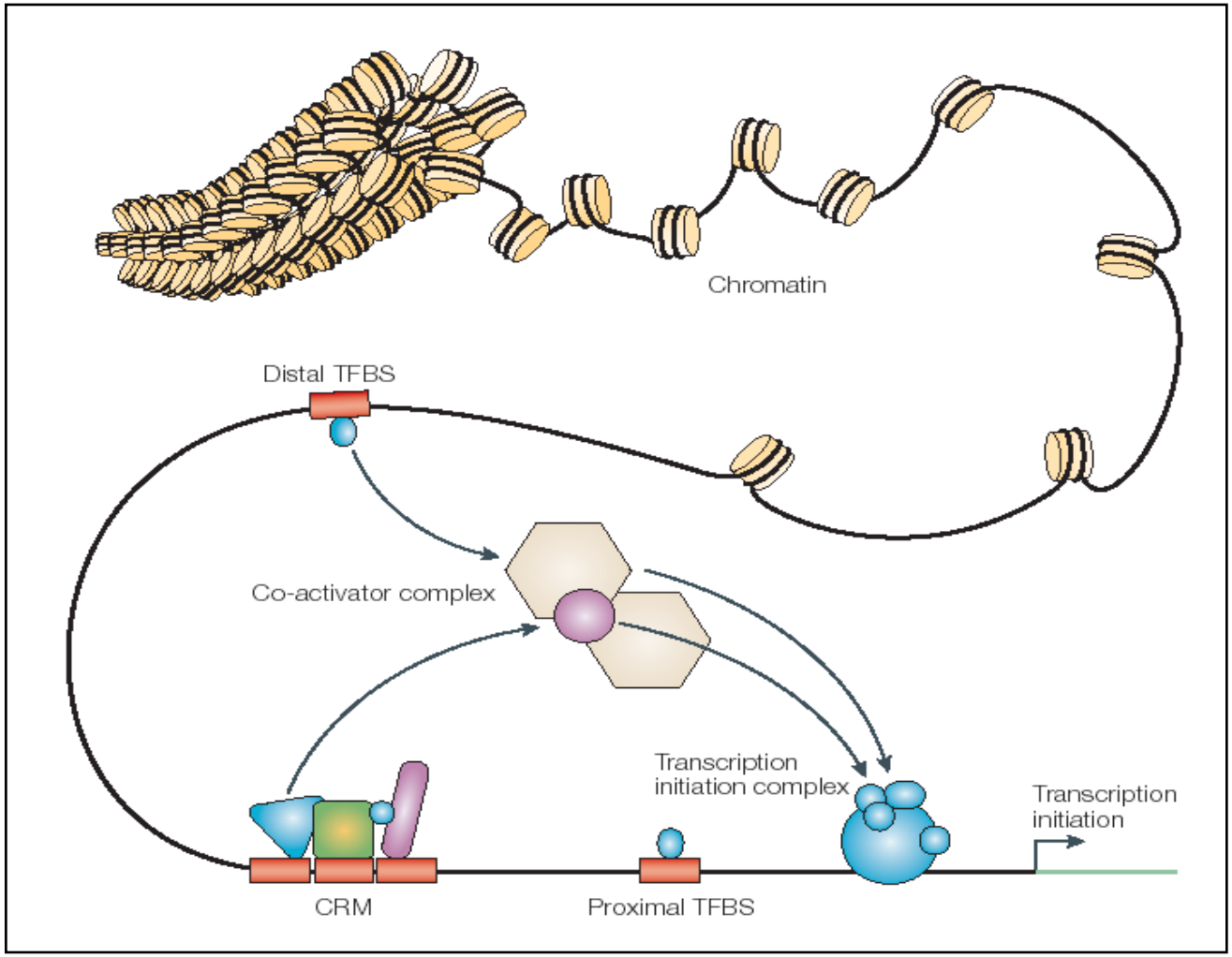



*Wassermann*, Nat. Rev. Genetics

**MicroRNAs (miRNAs)** are a family of small RNAs (typically **21 - 25** nucleotide long) that **negatively regulate gene expression at the posttranscriptional level**, (usually) thanks to the "seed" region in 3'-UTR regions.



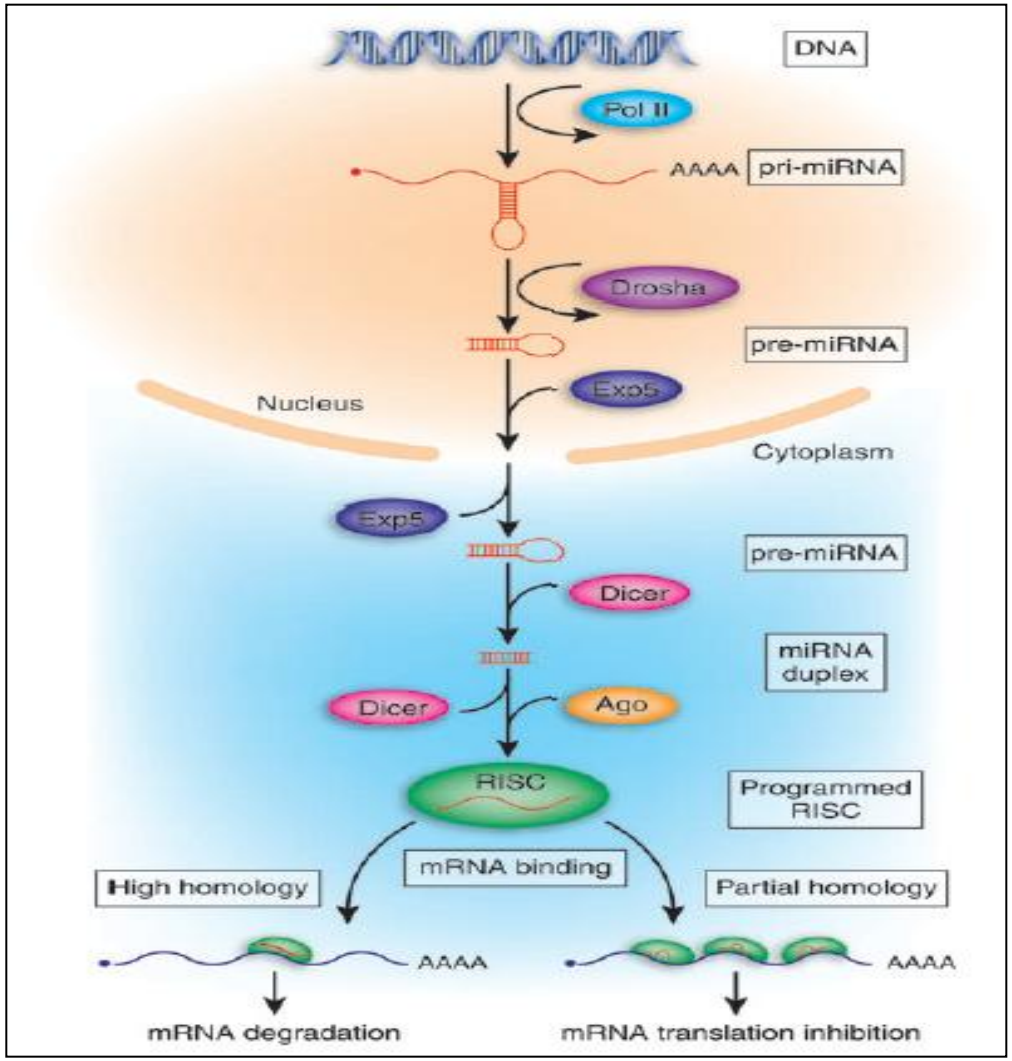
# Transcription Factors




*Wassermann*, Nat. Rev. Genetics

# MicroRNAs

**MicroRNAs (miRNAs)** are a family of small RNAs (typically **21 - 25** nucleotide long) that **negatively regulate gene expression at the post-transcriptional level**, (usually) thanks to the "seed" region in 3'-UTR regions.




*Wassermann*, Nat. Rev. Genetics

# MicroRNAs as regulatory genes

MiRNAs expression is regulated by the same TF which regulate all the other genes

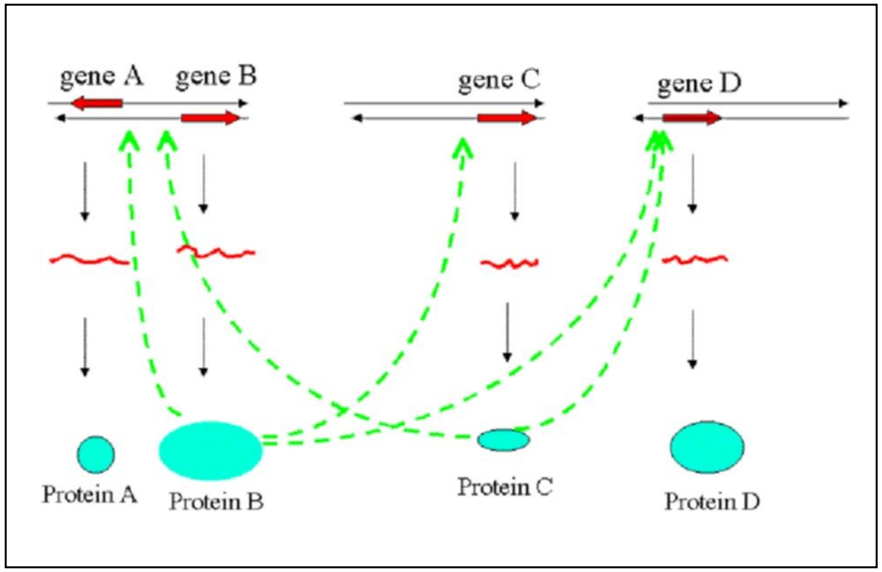
Regulation by miRNAs is a combinatorial process. Each miRNA is expected to control from one to hundreds of targets while a given mRNA can be under control of many different miRNAs. Usually miRNA binding sites are overrepresented in the 3'-utr sequence of target genes.

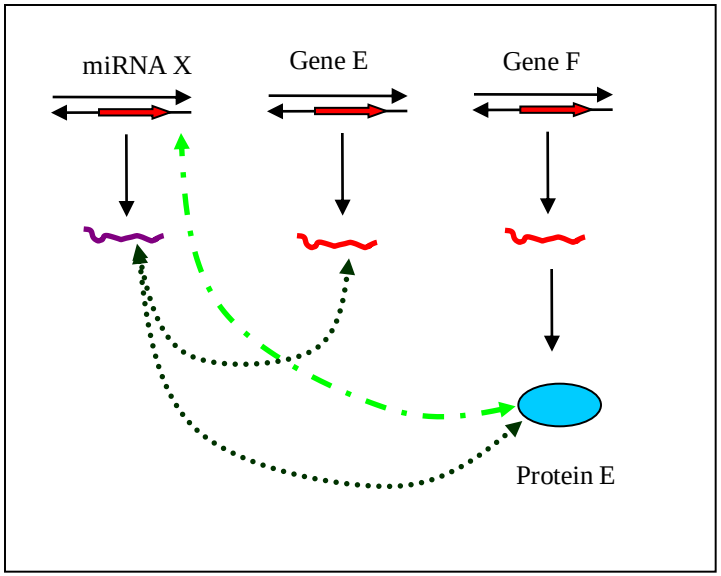
Transcription Factors and miRNAs share very similar regulatory strategies. The main difference between the two is that while TF act as a sort of on/off switch, it seems that the miRNA role is to fine tune the gene expression.

# Regulatory Networks 1

**Key 1 -->** **TFs** are themselves proteins produced by other genes, and they act in a combinatorial way, resulting in a complex network of interactions between genes and their products.
**--> Transcriptional Network**

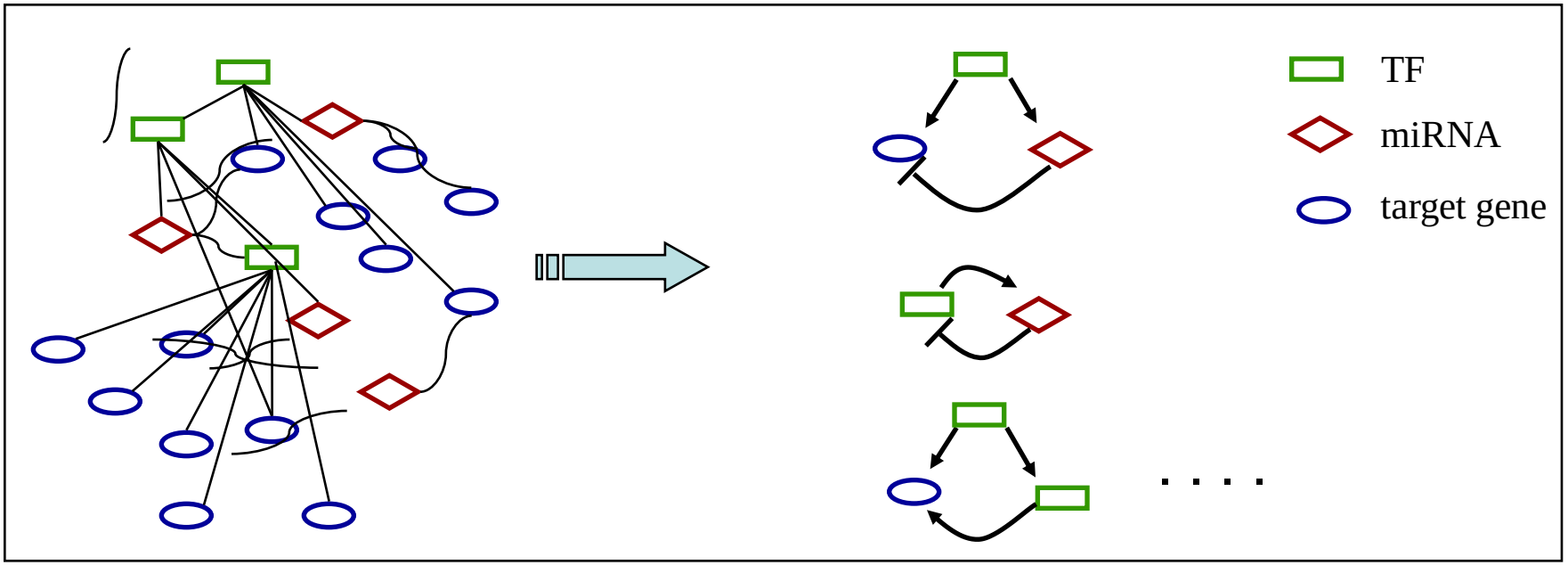
**miRNAs** also act in a combinatorial and one-to-many way, and, moreover, are transcribed from same POL-II promotes of TFs.
**--> Post-Transcriptional Network**



# Regulatory Networks 2

**Key 2 -->** Biological functions are performed by groups of genes which act in an interdependent and synergetic way. A complex network can be divided into simpler, distinct regulatory patterns called **network motifs**, typically composed by 3 or 4 interacting components which are able to perform elementary signal processing functions.



# Take-home Message

Optimal gene regulation can only be achieved combining transcriptional and post-transcriptional regulation.

High level regulatory functions like noise buffering, adaptation, fold change detection, stabilization of targets concentration ratios require the combination of Transcription Factors and microRNA in suitable elementary regulatory circuits: "network motifs"

# CircuitsDB

Several methods exist to study, separately TF-related and microRNA-related regulatory networks, but comparable information is lacking to explicitly connect them.

We have developed a **database** of the most important network motifs (FFLs, selfoops, feedback loops) combining together **TFs**, **miRNAs** and (in the last version: **CircuitsDB2**) also **lincRNAs**

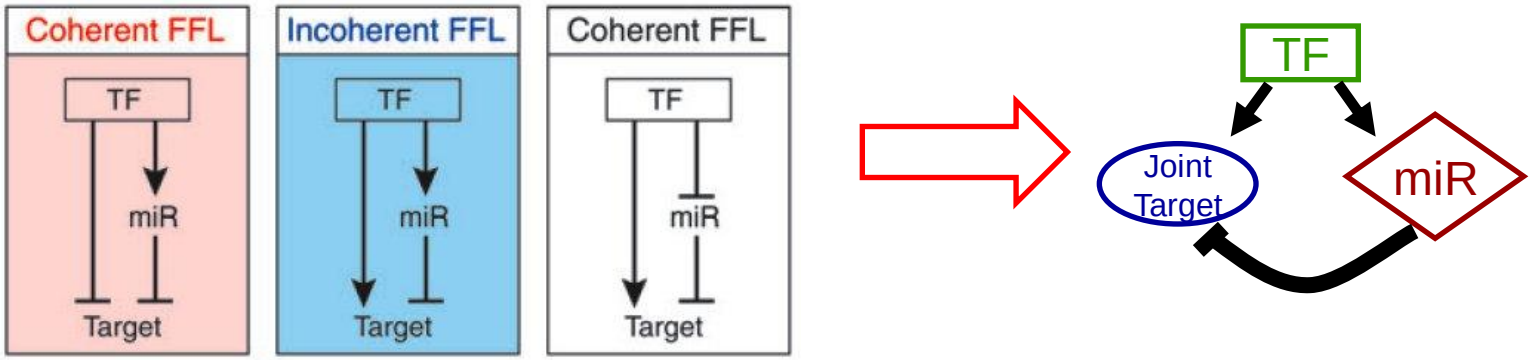
Using:

→ Experimental data (mainly from the ENCODE project)

→ Genome-wide bioinformatic analysis combining sequence over-representation, evolutionary conservation, scanning of PWMs, miRNA seed search…

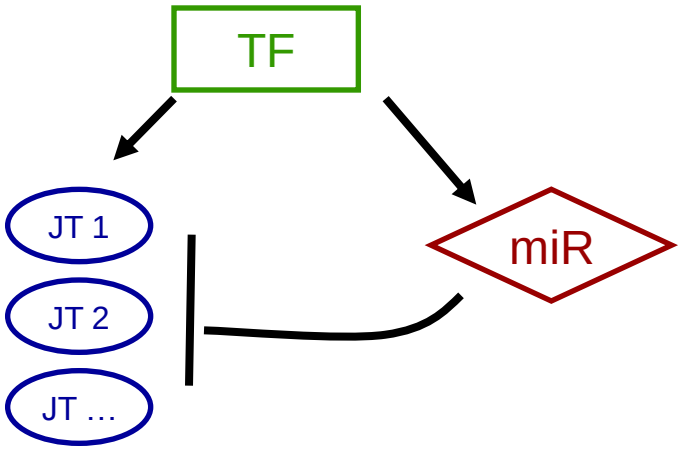
# Example: MiRNA mediated Feed Forward Loops

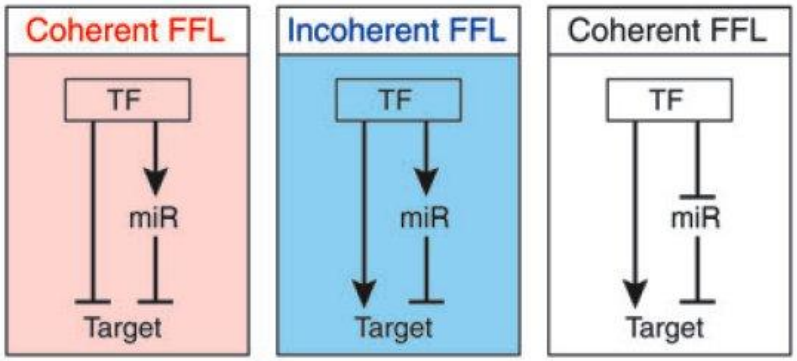
**Mixed Feed-Forward Regulatory Loops** --> network motifs in which a master Transcription Factor (TF) regulates a miRNA and together with it a set of Joint Target coding genes.



# Example: list of MiRNA mediated FFLs in CircuitsDB

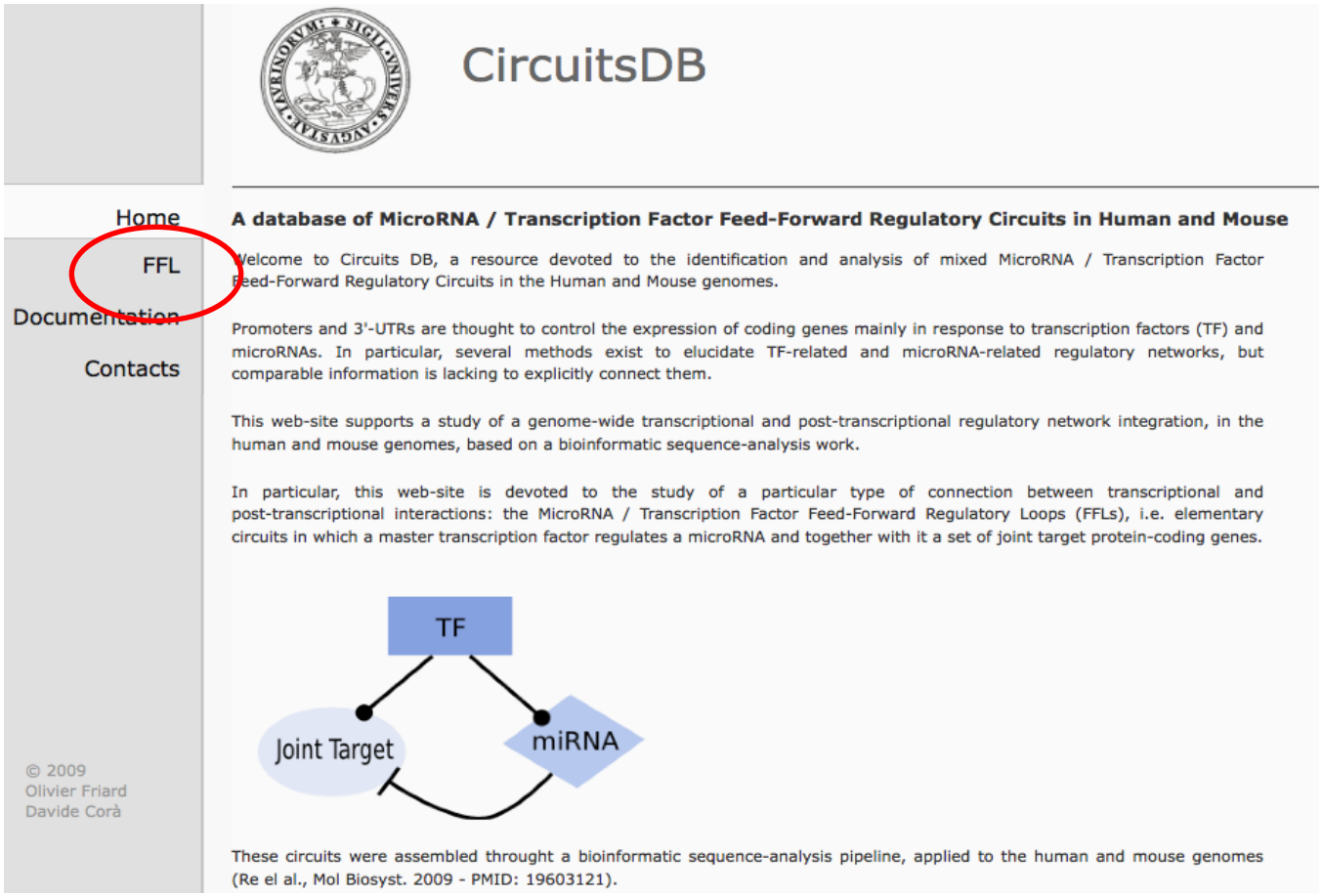
**Human mixed FFLs catalogue -->** The list contains **5030** different "**single target circuits**", corresponding to **638** "**merged circuits**", involving a Total of 2625 joint target genes (JTs), 101 TFs and 133 miRNAs. # of JTs ranged from 1 to 38.



# CircuitsDB

We have developed a **web-based** graphical interface to get free access to the database of mixed FFLs.



# Circuits assessment 1 functional analysis

We analyzed each FFL looking for an **enrichment in Gene Ontology categories in the set of their joint targets**.

To assess this enrichment we used the standard exact Fisher test with a p-value threshold
**$p < 10-4$**

We found a few enriched GO categories involving various aspects of organism differentiation and development

# Circuits assessment 1 functional analysis

| FFL id | JTs | Fisher test pvalue | Gene Ontology characterization |
|---|---|---|---|
| ER\|hsa-miR-135b | GBE1 HCN2 CD99L2 TTC21A BSN RNASE11 ANGPT2 Q49AQ9_HUMAN NP_057628.1 LZTS2 ZNF69 FAM129A FMOD IL11 ISCA1 PR285_HUMAN CITED1 TGM2 MUSK DEFB123 MFSD3 C17orf28 NP_787078.1 PRLR | 4.11e-05 | cellular protein complex assembly (P) |
| HMGIY\|hsa-miR-152 | EDG1 Q86V52_HUMAN DMRTA2 SLC25A32 FGF1 ITGA5 MEOX2 EPAS1 ZNF33A ADAM17 MAPK6 RNF182 | 6.48e-05 | angiogenesis (P) |
| ICSBP\|hsa-miR-223 | ADM GAST PRL GTDC1 FOXO3A | 1.40e-06<br>2.18e-05<br>7.49e-05 | hormone activity (F)<br>reproductive process (P)<br>multicellular organism reproduction (P) |
| IRF1\|hsa-miR-126 | FGFR FGFL7 GOLPH3 BDH2 ZADH2 | 8.01e-05 | regulation of cell migration (P) |
| IRF-7\|hsa-miR-26a | VAX1 GALNT10 CA3 EIF2S1 NDUFA4 ARP19_HUMAN FBXO42 RPIA FBXL19 ALS2CR2 XR_017723.1 GSK3B DBR1 TTC13 NT5DC1 | 8.01e-05<br>6.25e-05 | regulation of cell migration (P)<br>cellular response to stress (P) |
| MYC\|hsa-miR-17-5p | EDD1 TAF5L HIF1A Q6ZR74_HUMAN OSBPL10 E2F1 ACP1 MYNN CENTB5 GDA | 9.40e-05<br>9.56e-05 | cellular metabolic process (P)<br>primary metabolic process (P) |
| MYOD\|hsa-miR-140 | ANK2 TSSK2 EIF2AK1 HMX2 THY1 ALAS2 UROC1 CDKL4 PPARA CYBB PPL CDS2 ZIC3 | 7.20e-06<br>6.61e-05 | hemoglobin metabolic process (P)<br>organ development (P) |
| SRY\|hsa-miR-26a | FANCA GSK3B RPIA Q6ZQV3_HUMAN ALS2CR2 KIF1C RG9MTD2 CDS1 BAG4 PPP2R3C | 2.68e-05<br>5.64e-05 | protein export from nucleus (P)<br>anti-apoptosis (P) |

# Circuits assessment 2: looking for cancer related genes

In these last few years it is becoming increasingly clear that miRNAs play a central role in **cancer development** (e.g. Blattener *Mol Syst. Biol. 2008*).

→We filtered our results looking for FFLs containing at least *two* cancer related miRNA or target gene.

Sources: oncomiRs reported in

- *Esquela-Kerscher and FJ Slack, Nat Rev Cancer 2006*
- *Zhang et al, Dev Biol, 2007*

cancer genes reported in

- *Cancer Gene Census database.*

| | | | |
|---|---|---|---|
| AP-1\|hsa-miR-142-3p | | hsa-miR-142-3p | DDIT3 |
| ATF-1\|hsa-miR-199a* | | hsa-miR-199a* | MTCP1 |
| ATF6\|hsa-miR-199a* | | hsa-miR-199a* | MTCP1 |
| ER\|hsa-miR-375 | | | TPR, USP6 |
| HIF-1\|hsa-miR-199a* | | hsa-miR-199a* | MTCP1 |
| HNF-3\|hsa-let-7a | | hsa-let-7a | CCND2 |
| HNF-3\|hsa-let-7f | | hsa-let-7f | CCND2 |
| HNF-3\|hsa-miR-30a-5p | | | MYH11, BCL9 |
| HNF-3\|hsa-miR-30c | | | MYH11, BCL9 |
| HSF2\|hsa-let-7a | | hsa-let-7a | MYCN |
| HSF2\|hsa-let-7f | | hsa-let-7f | MYCN |
| HSF2\|hsa-miR-199a* | | hsa-miR-199a* | MYCN |
| IRF\|hsa-miR-125b | | hsa-miR-125b | BCL2 |
| IY\|hsa-miR-296 | | | RPL22, BCL2 |
| MYC\|hsa-miR-17-5p | MYC | hsa-miR-17-5p | |
| MYC\|hsa-miR-19a | MYC | hsa-miR-19a | |
| MYC\|hsa-miR-20a | MYC | hsa-miR-20a | |
| NF-Y\|hsa-miR-223 | | | APC, ATF1 |
| OCTAMER\|hsa-miR-125b | | hsa-miR-125b | IRF4 |
| PAX-4\|hsa-miR-125b | | hsa-miR-125b | IRF4 |
| SOX-5\|hsa-miR-125b | | hsa-miR-125b | SS18 |
| SOX-5\|hsa-miR-29a | | | EXT1,COL1A1 |
| SRY\|hsa-miR-221 | | hsa-miR-221 | CCND2 |
| SRY\|hsa-miR-412 | | | BRAF, ATIC |

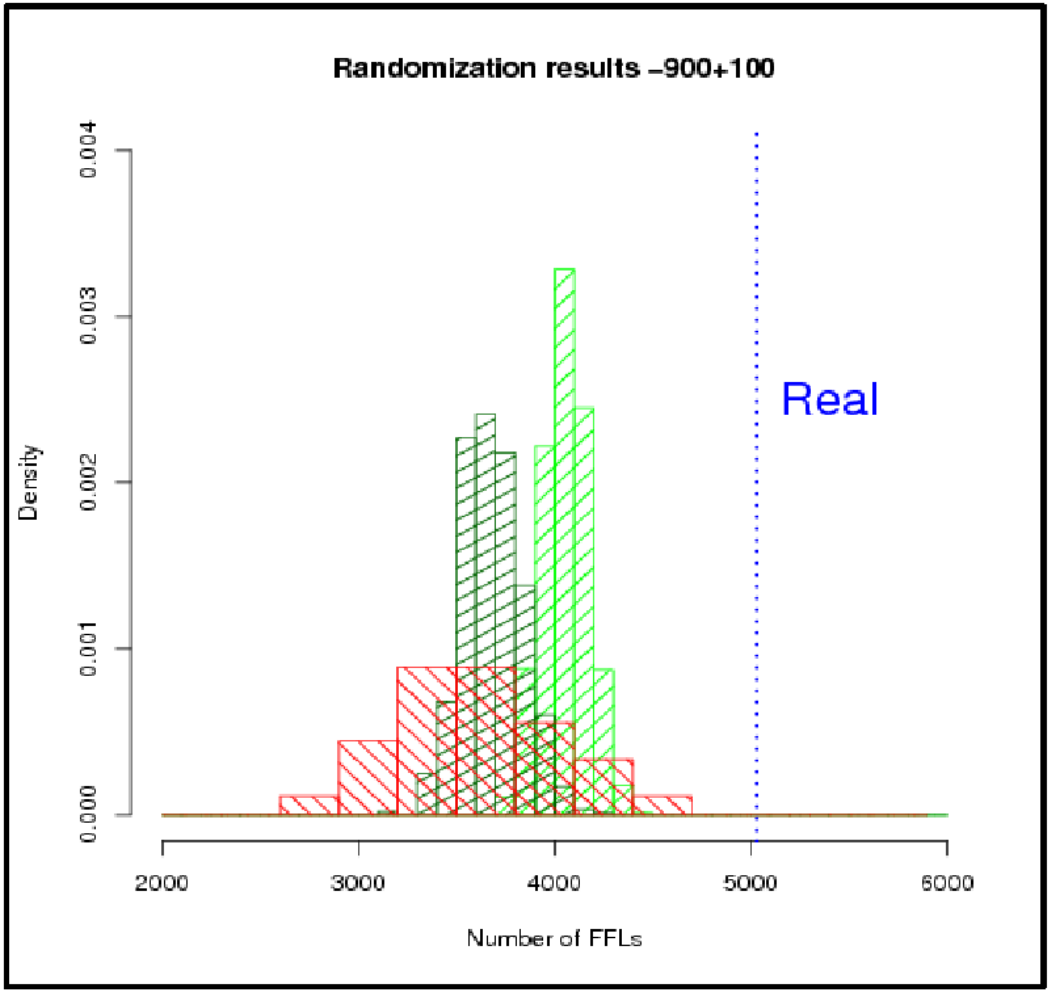
# Circuits assessment 3: mixed FFLs as network motifs

Elementary regulatory circuits (the so called ”network motifs”) were shown
to be over-represented in transcriptional networks.
(Milo et a., *Science 2002*, Shen-Orr et., *Nat Genetics 2002*)

In order to quantify the overrepresentation of our mixed FFLs we perfomed various **randomization tests.**

- Complete node replacement, *Z = 9.2*
- Random reshuffling of miRNA promoters and seeds, *Z = 8.1*
- Edge Switching, *Z = 8.4*

# FFLs are over-represented



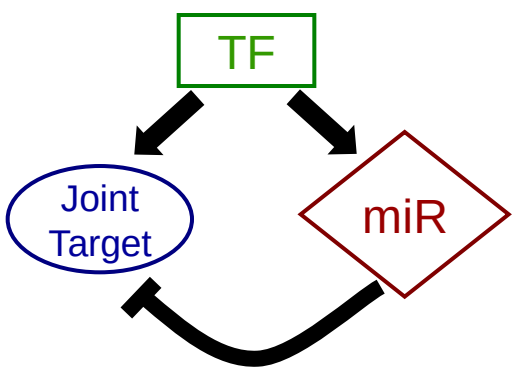
# Functional role of mixed FFLs

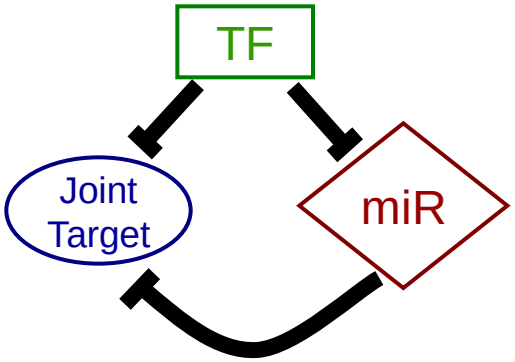
Depending on the type of transcriptional regulation (excitatory or inhibitory) exerted by the master TF on the miRNA and on the targets, FFLs may be classified as

- incoherent ("type I" FFLs), or
- coherent ("type II" FFLs).

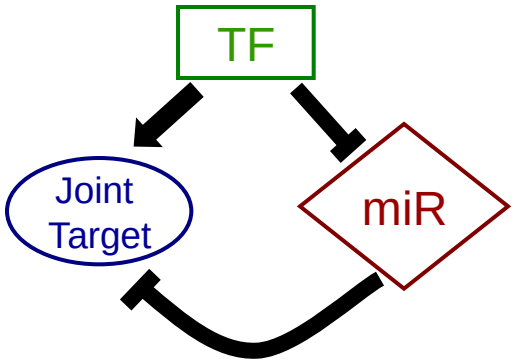
# Type I and II FFLs

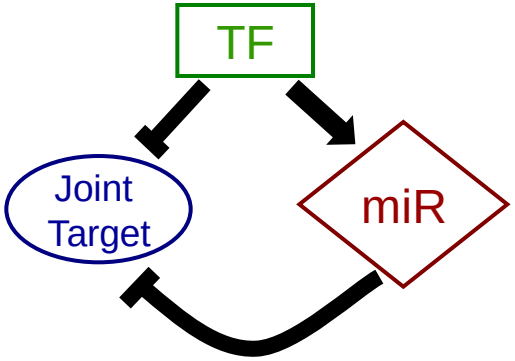
Possible biological role for mixed TF/miRNA network motifs:




type I circuits




type II circuits

- Type II (coherent) circuits lead to a reinforcement of the transcriptional regulation at the post-transcriptional level and might be important to eliminate the already transcribed mRNAs when the transcription of a target gene is switched off.

- Type I (incoherent) circuits allow for a fine tuning of gene expression, setting the optimal functional value of the protein through a miRNA repression

# Additional role: noise damping

**Fine tuning is useless without a tight control of cell to cell fluctuations.**

Type I (incoherent) FFLs can also **stabilize the steady state production of the target protein by damping translational and transcriptional fluctuations**.

In a simple TF-target interaction any fluctuation of master TF could induce a non-linear increase in the amount of its target products. The presence, among the targets, of a miRNA which downregulates the other targets might represent a simple and effective way to control these fluctuations.

# Study of protein fluctuations via stochastic equations

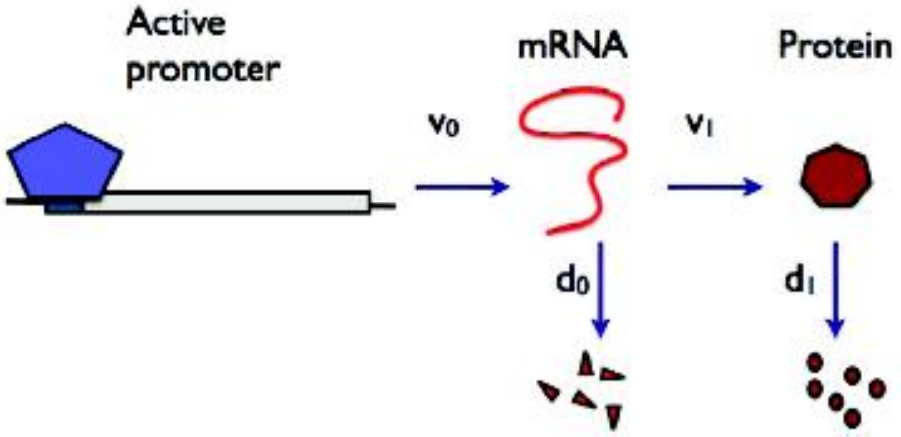
The only way to address this issue is to describe the FFLs in terms of stochastic equations and to compare the results with those obtained with that of a standard transcription + translation process

In both cases fluctuations are proportional to the mean number of proteins produced by a single mRNA. This number is a function of the miRNA-mRNA affinity.

# Stochastic equations for gene expression: two steps model.

(*Shaharezaei V, Swain PS PNAS (2008) 105, 17256)*

This model assumes that the promoter is always active and so has only two stochastic variables: the number of mRNAs and the number of proteins



The probability of having *m* mRNAs and *n* proteins at time *t* satisfies the master equation:

$$\frac{\partial P_{m,n}}{\partial t} = v_0(P_{m-1,n} - P_{m,n}) + v_1 m(P_{m,n-1} - P_{m,n})$$
$$+ d_0[(m+1)P_{m+1,n} - mP_{m,n}]$$
$$+ d_1[(n+1)P_{m,n+1} - nP_{m,n}]$$

The master equation can be rewritten as a differential equation using the generating function:

$$F(z',z) = \sum_{m,n} z'^m z^n P_{m,n},$$

Setting: $$a = \nu_0/d_1, b = \nu_1/d_0, \gamma = d_0/d_1,$$

$$u = z' - 1 \text{ and } v = z - 1.$$

and $\tau = d_1 t,$ we find:

$$\frac{\partial F}{\partial v} - \gamma \left[ b(1+u) - \frac{u}{v} \right] \frac{\partial F}{\partial u} + \frac{1}{v} \frac{\partial F}{\partial \tau} = a \frac{u}{v} F,$$

If we assume that the protein lifetime is much longer than that of the mRNA then the equation simplifies (the mRNA is at steady state) and can be solved exactly:

$$F(z,\tau) = \left[\frac{1 - b(z-1)\mathrm{e}^{-\tau}}{1 + b - bz}\right]^{a}$$

leading to an exact expression for the probability distribution:

$$P_n(\tau) = \frac{\Gamma(a+n)}{\Gamma(n+1)\Gamma(a)} \left(\frac{b}{1+b}\right)^n \left(\frac{1+b\mathrm{e}^{-\tau}}{1+b}\right)^a \times {}_2F_1\left(-n, -a, 1-a-n; \frac{1+b}{\mathrm{e}^{\tau}+b}\right)$$

which at steady state becomes the well known negative binomial distribution:

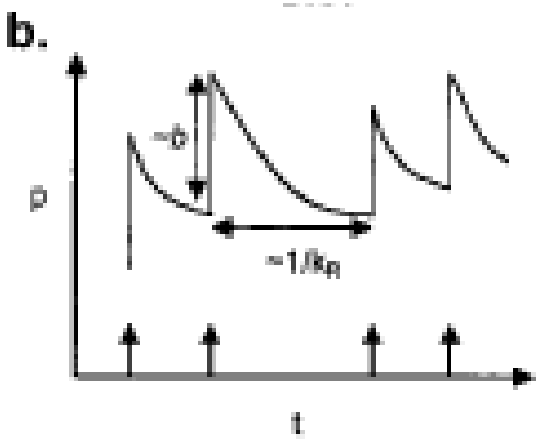
$$P_n = \frac{\Gamma(a+n)}{\Gamma(n+1)\Gamma(a)} \left(\frac{b}{1+b}\right)^n \left(1 - \frac{b}{1+b}\right)^a$$

The corresponding mean value and fluctuations of the number of proteins are:

$$\langle n \rangle = ab(1 - e^{-\tau}),$$

$$\langle n^2 \rangle - \langle n \rangle^2 = \langle n \rangle (1 + b + be^{-\tau})$$

Where b is the mean number of proteins produced by a single mRNA (burst parameter). Fluctuations strongly depend on the burst parameter b.



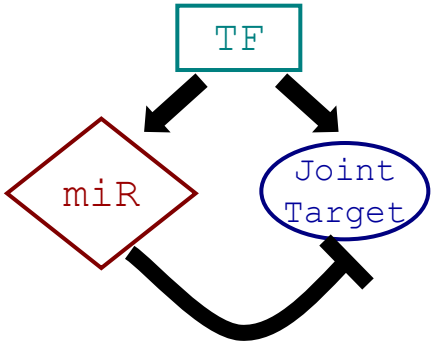
The same analysis can be performed in the case of the inchoerent FFL, leading to a relevant reduction of noise

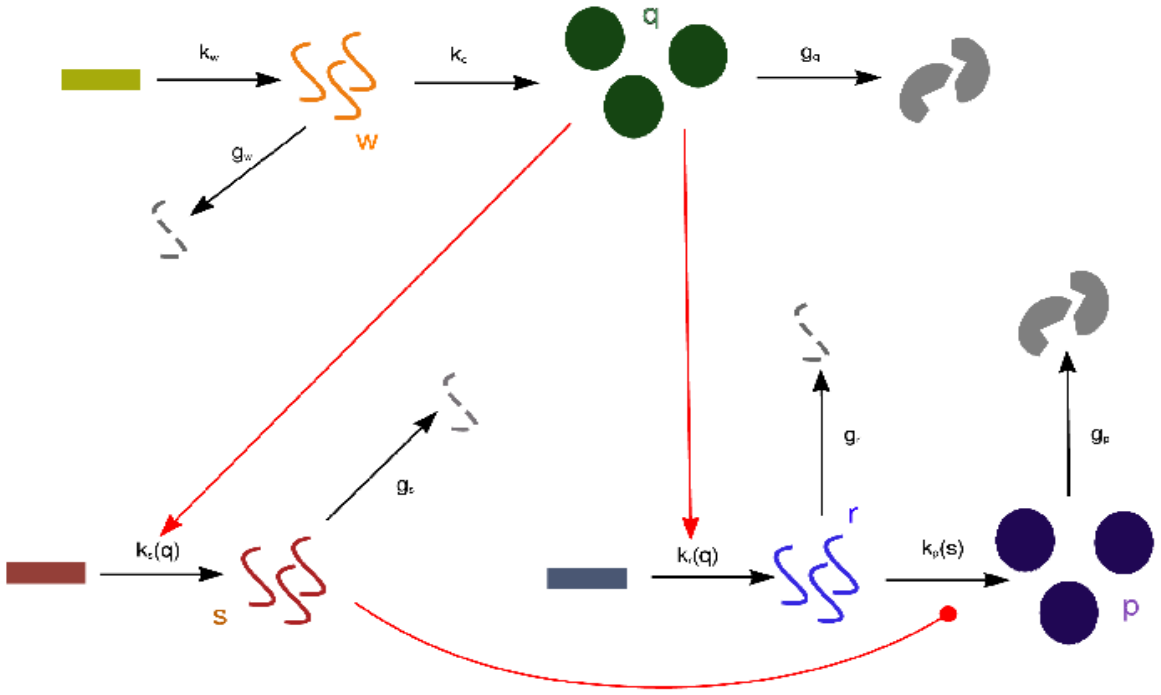
The noise reduction can be traced back to two parallel mechanisms:

- The different efficiency of the mRNA translation in the two cases:
noise reduction is a function of the miRNA-mRNA affinity

-The correlated fluctuations of miRNA and target under fluctuations in the transcriptional efficiency of the master Transcription Factor

# Master equation for the incoherent FFL



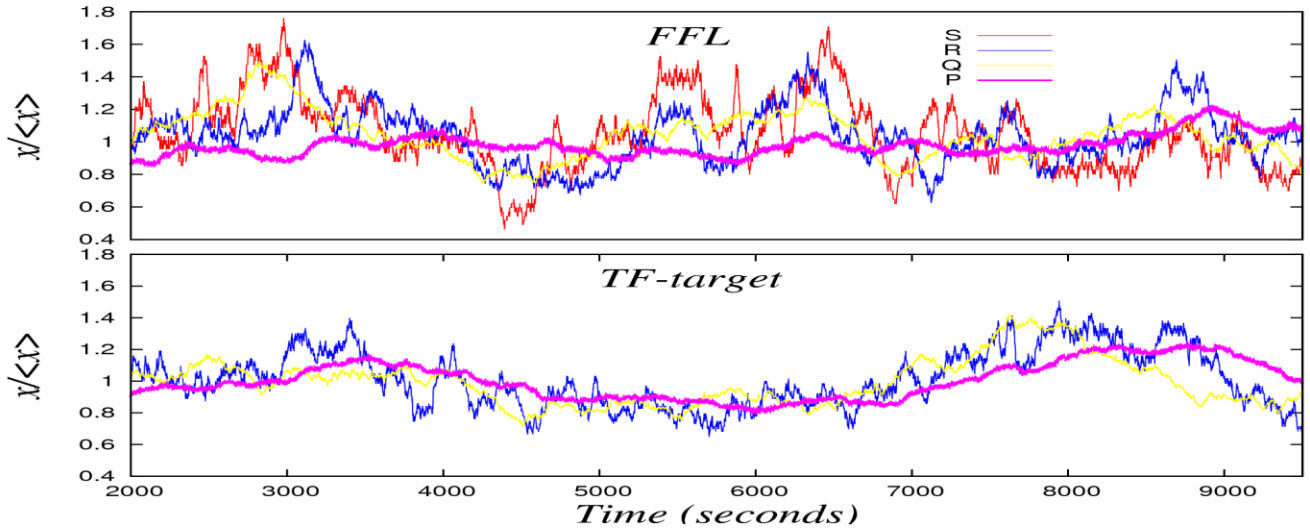



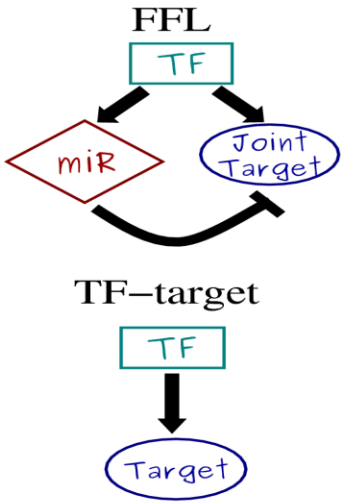

## Master Equation

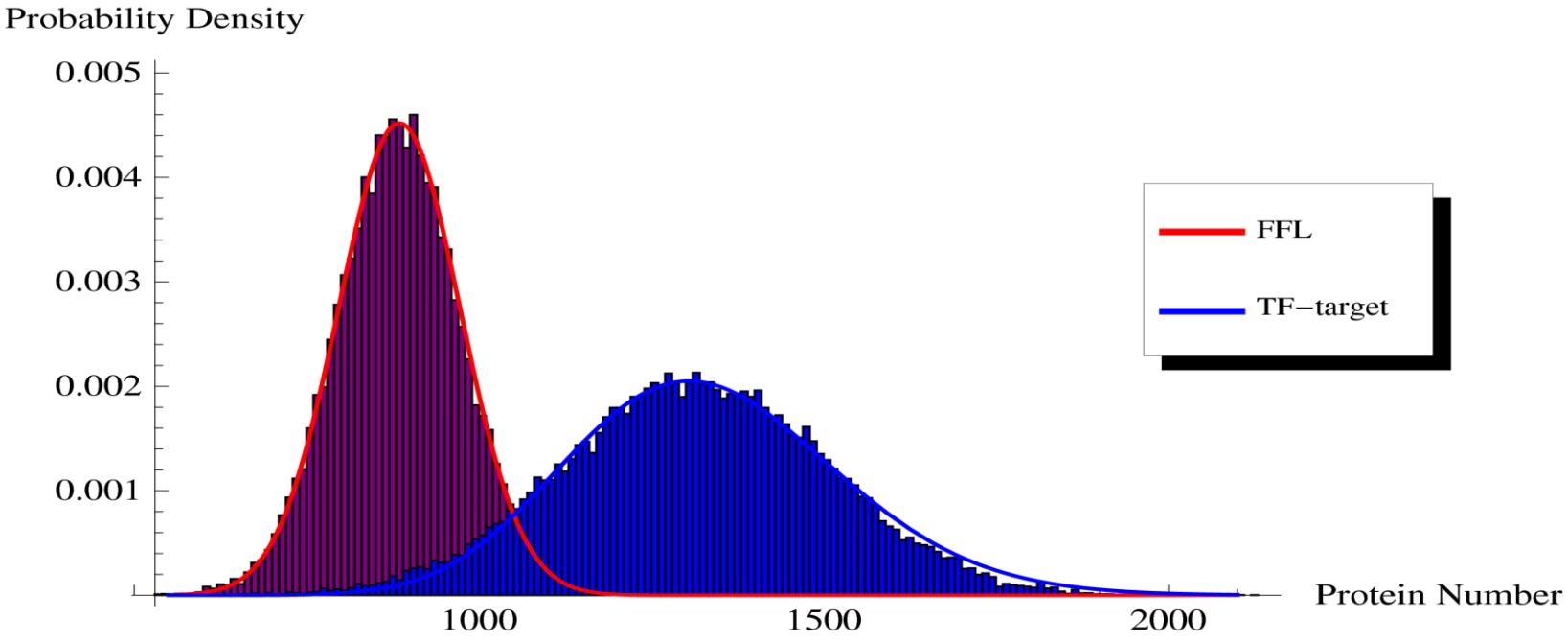
$$\frac{\partial P_{w,q,s,r,p}}{\partial t} = k_w(P_{w-1,q,s,r,p} - P_{w,q,s,r,p}) + k_q w(P_{w,q-1,s,r,p} - P_{w,q,s,r,p})$$
$$+k_r(q)(P_{w,q,s,r-1,p} - P_{w,q,s,r,p}) + k_p(s)r(P_{w,q,s,r,p-1} - P_{w,q,s,r,p})$$
$$+k_s(q)(P_{w,q,s-1,r,p} - P_{w,q,s,r,p}) + g_w\Big[(w+1)P_{w+1,q,s,r,p} - wP_{w,q,s,r,p}\Big]$$
$$+g_q\Big[(q+1)P_{w,q+1,s,r,p} - qP_{w,q,s,r,p}\Big] + g_r\Big[(r+1)P_{w,q,s,r+1,p} - rP_{w,q,s,r,p}\Big]$$
$$+g_s\Big[(s+1)P_{w,q,s+1,r,p} - sP_{w,q,s,r,p}\Big] + g_p\Big[(p+1)P_{w,q,s,r,p+1} - pP_{w,q,s,r,p}\Big]$$

- The first two moments can be calculated with the moment generating function method.
- Non linear model

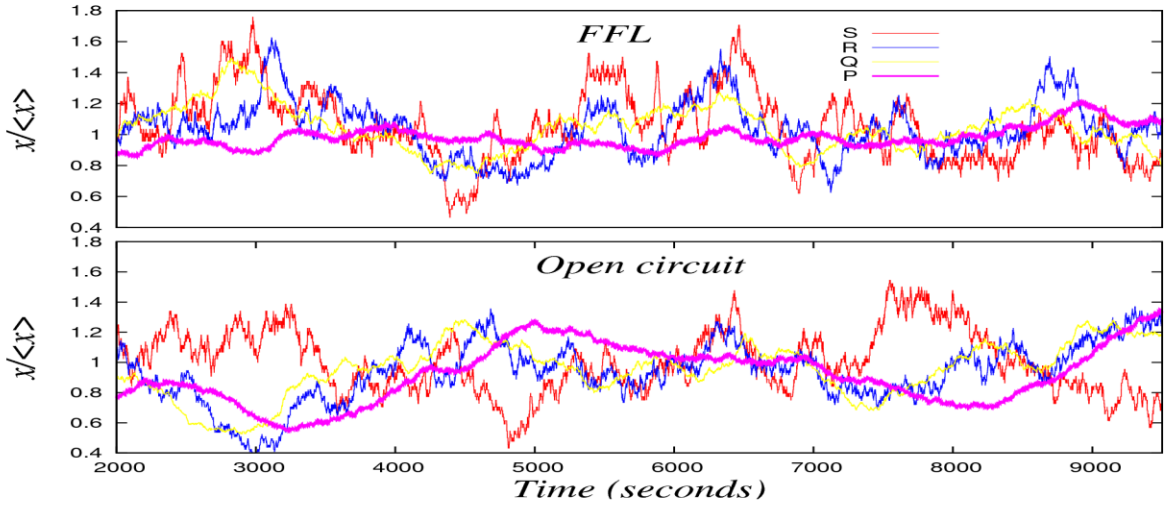
# Noise Buffering I

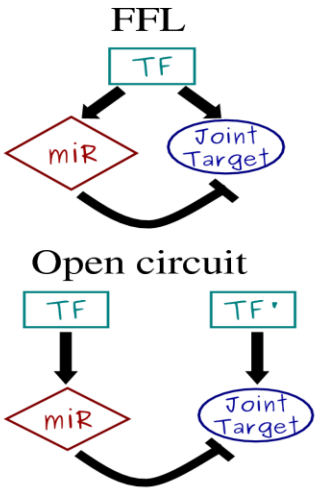


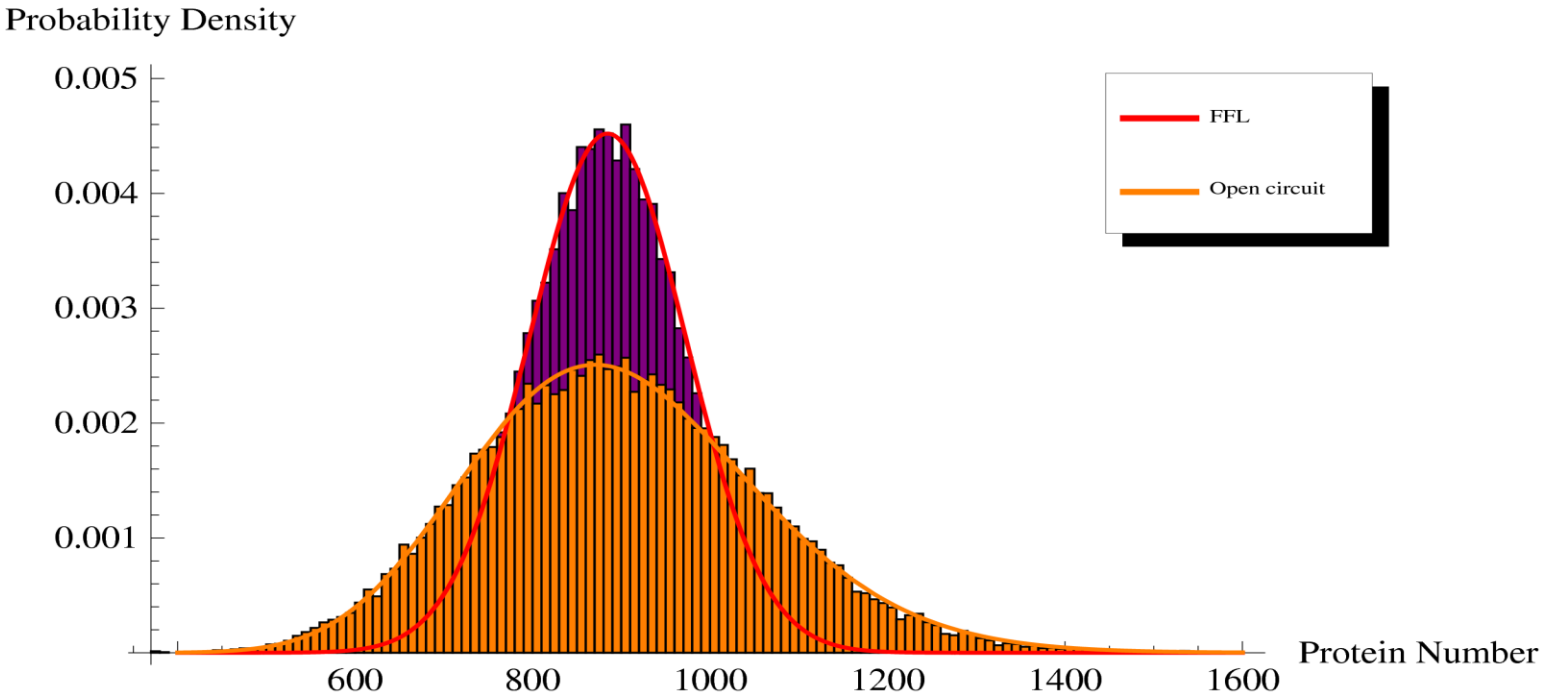



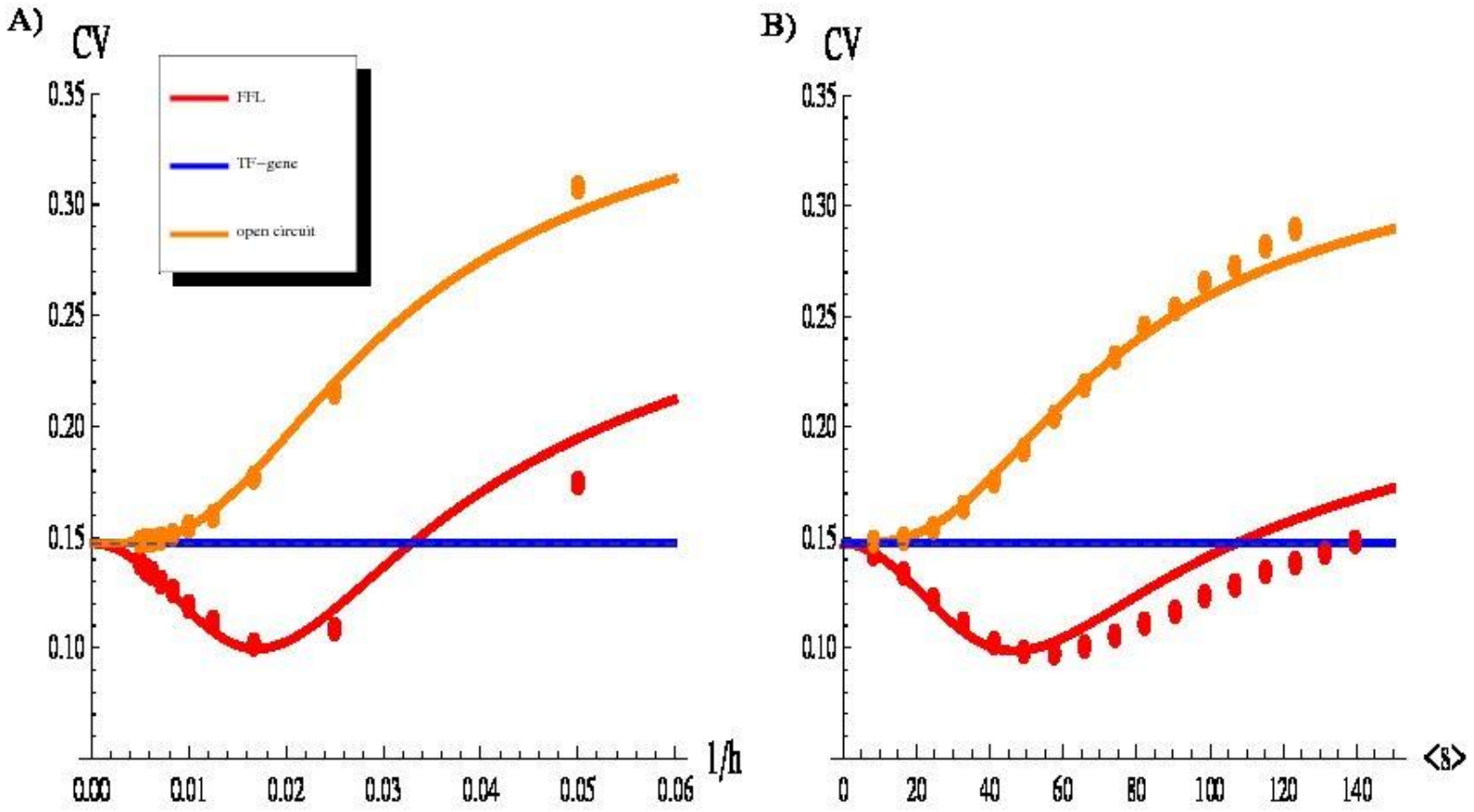
# Noise Buffering II



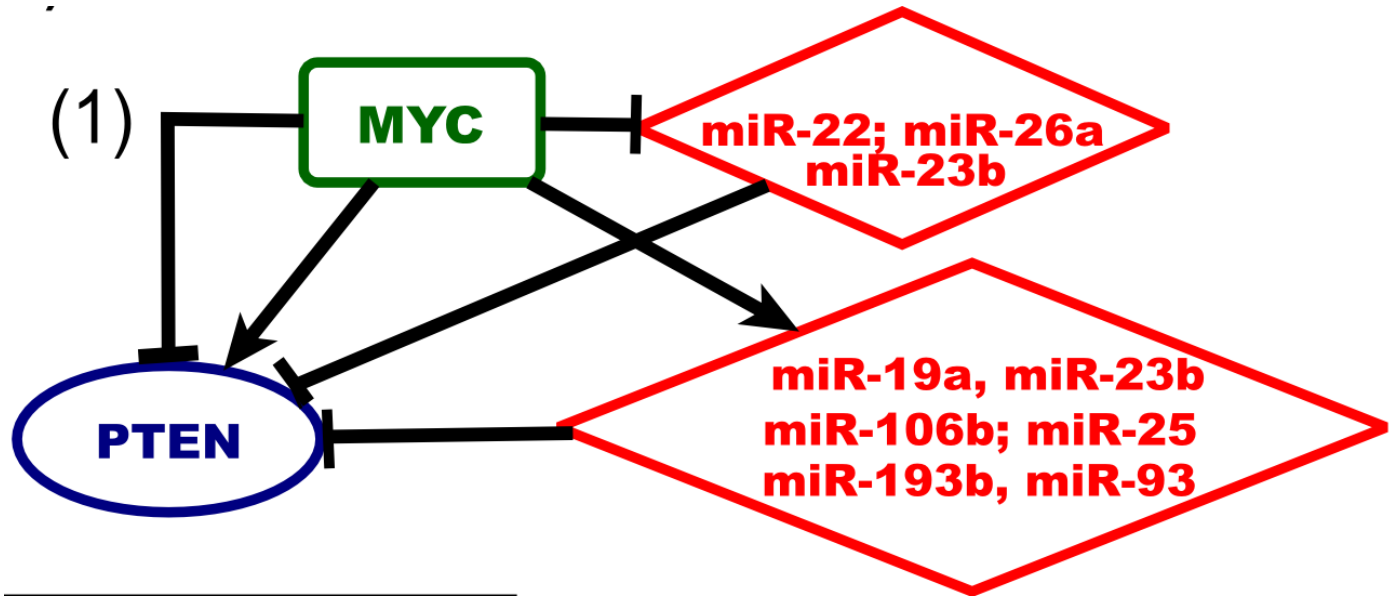



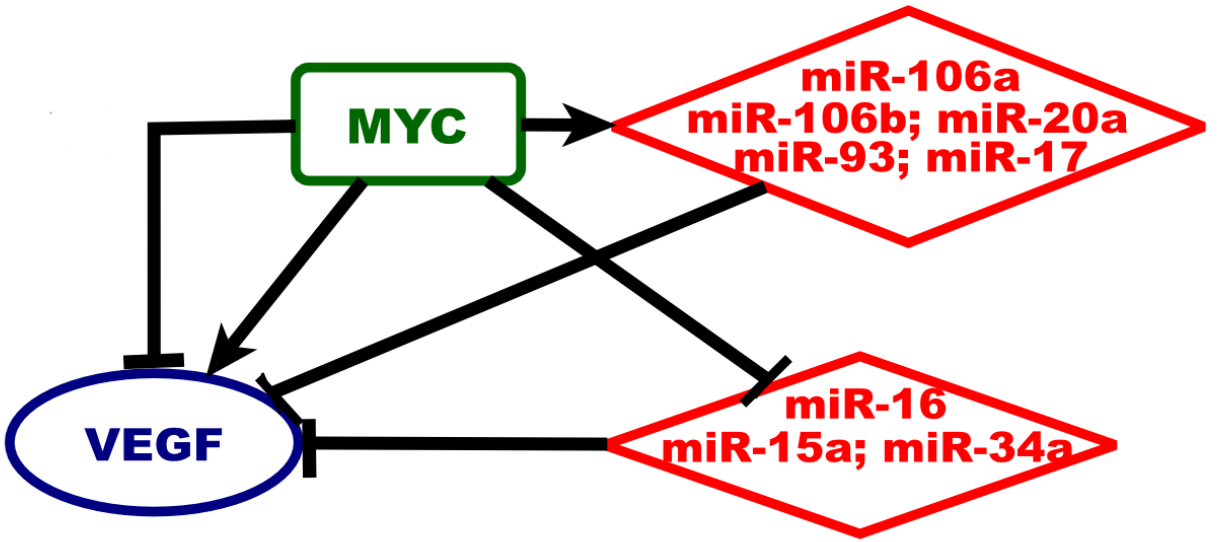
# Optimal noise reduction for intermediate values of miRNA/mRNA affinity



# Example : regulation of PTEN

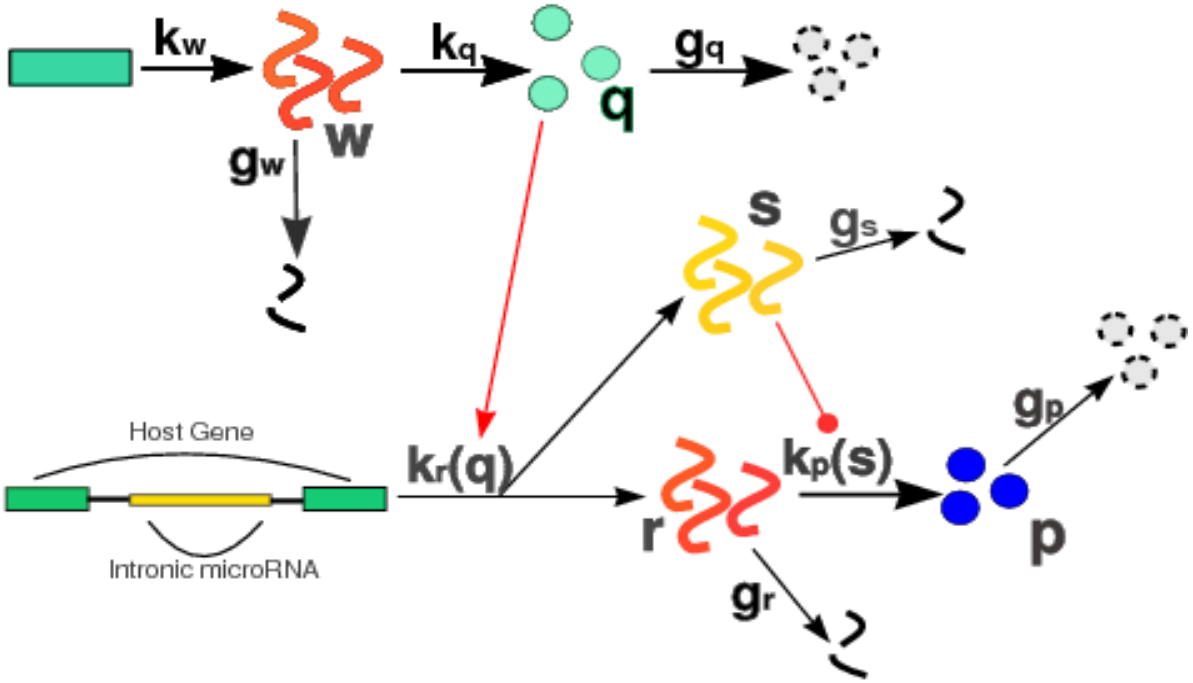


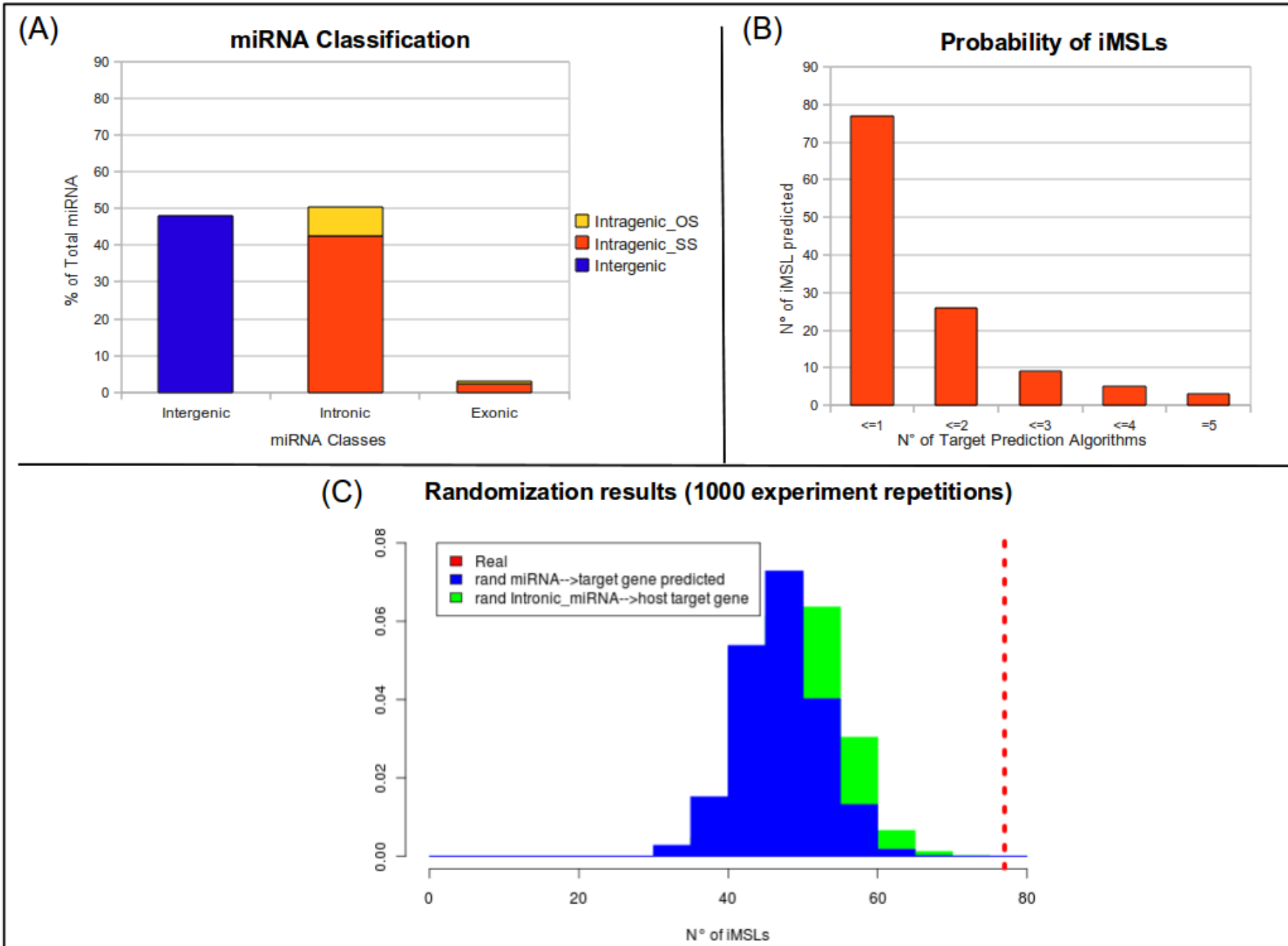
# Example 2: regulation of VEGF



# MiRNA mediated selfloop

The same analysis can be extended to the simplest possible mixed circuit: the miRNA mediated selfloop in which an intronic miRNA regulates its host gene.



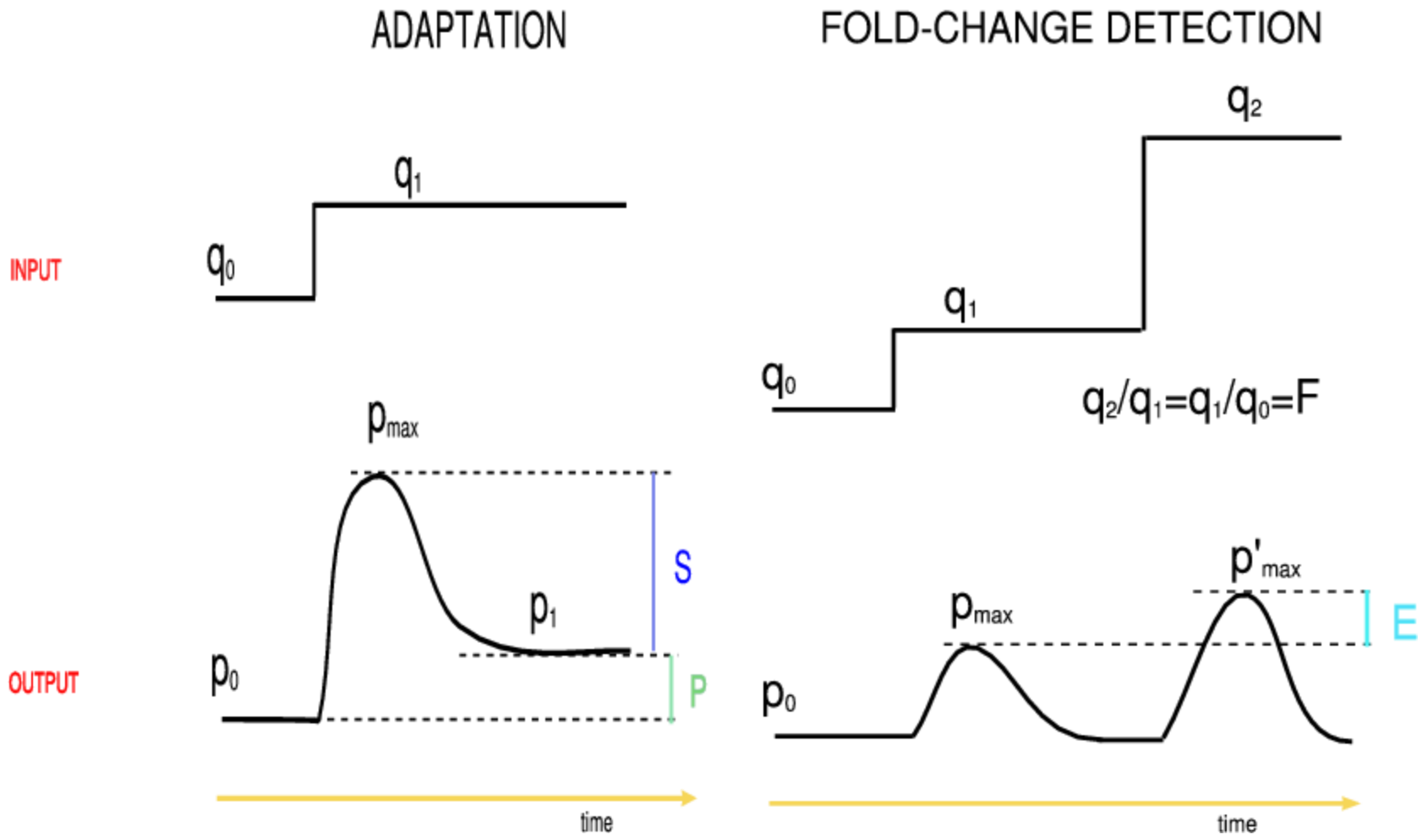

(A)
miRNA Classification
% of Total miRNA
90
80
70
60
50
40
30
20
10
0
Intergenic
Intronic
Exonic
miRNA Classes
Intragenic_OS
Intragenic_SS
Intergenic
(B)
Probability of iMSLs
N° of iMSL predicted
90
80
70
60
50
40
30
20
10
0
<=1
<=2
<=3
<=4
=5
N° of Target Prediction Algorithms
(C)
Randomization results (1000 experiment repetitions)
Real
rand miRNA-->target gene predicted
rand Intronic_miRNA-->host target gene
0.08
0.06
0.04
0.02
0.00
0
20
40
60
80
N° of iMSLs

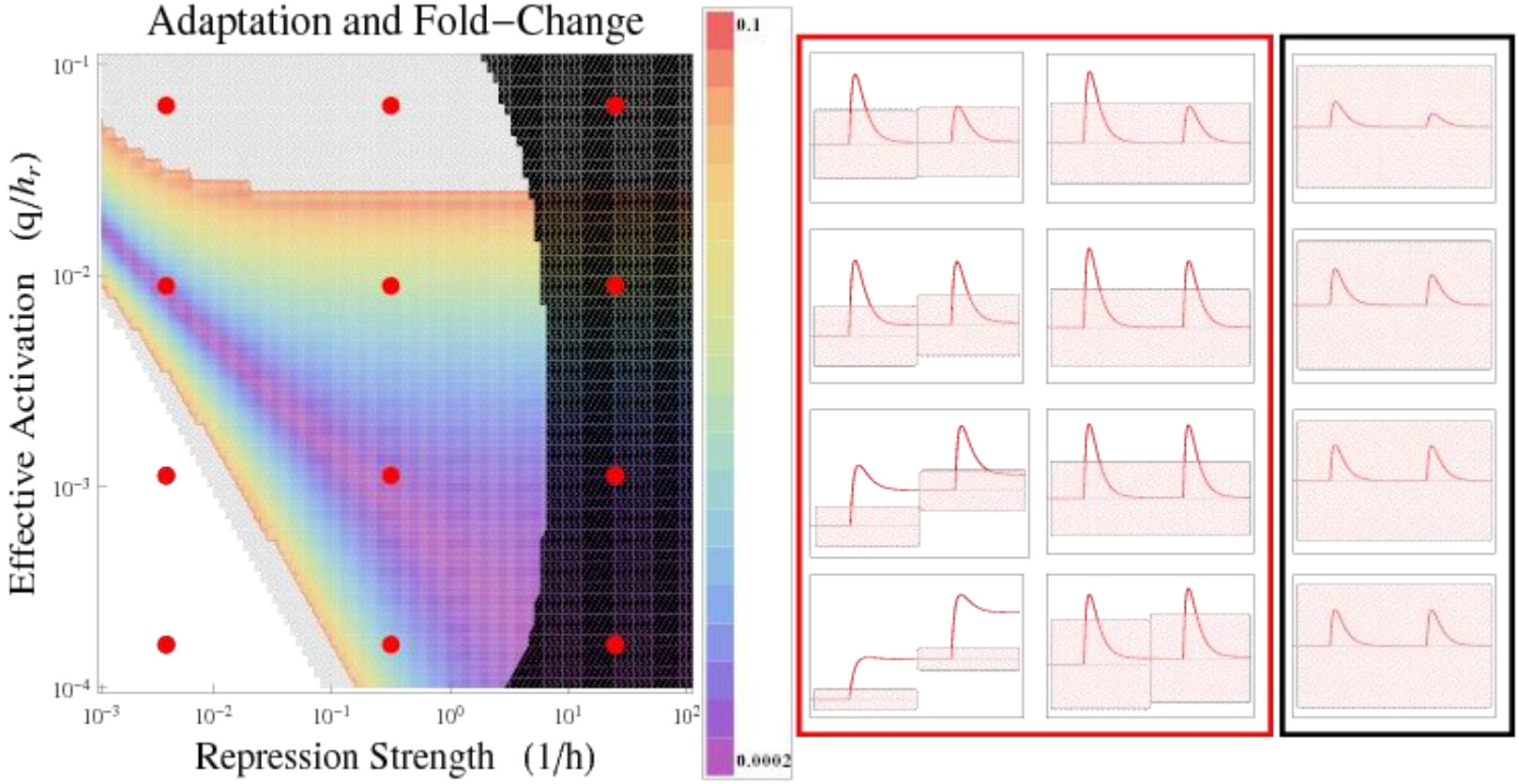
# MiRNA mediated Selfloops 2

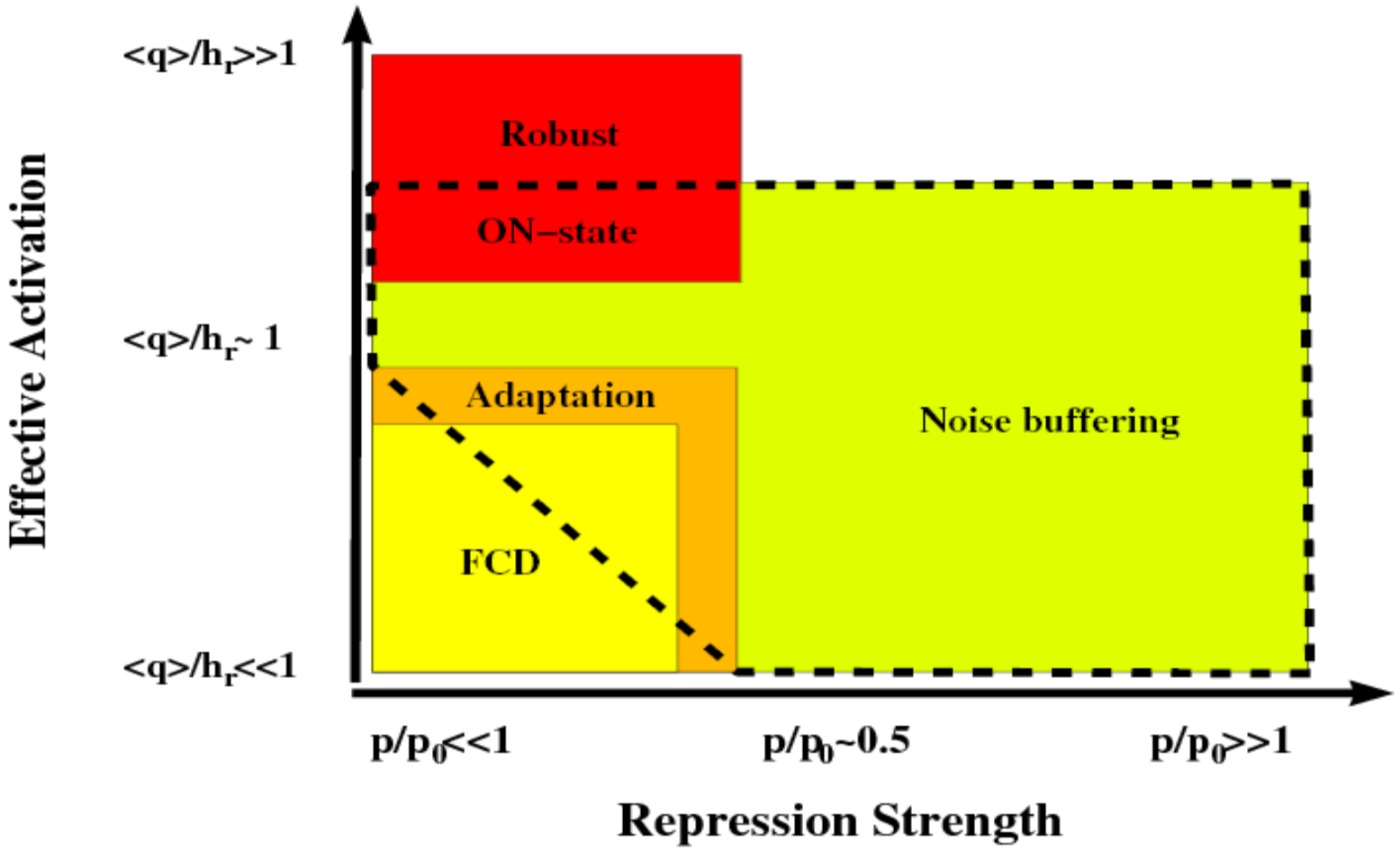
Also in this case we found fine tuning and noise reduction properties. Moreover this circuit, depending on the values of the parameters, is able to perform:

**adaptation** and **fold-change detection**

The parameter space can be summarized using only two quantities: the "effective activation" **<q>/h_r** where <q> is the mean concentration of the activating TF and h_r the corresponding dissociation constant and the miRNA repression strength **1/h**


ADAPTATION
FOLD-CHANGE DETECTION
INPUT
OUTPUT
q0
q1
q2
p0
pmax
p1
S
P
p'max
E
q2/q1=q1/q0=F
time
time

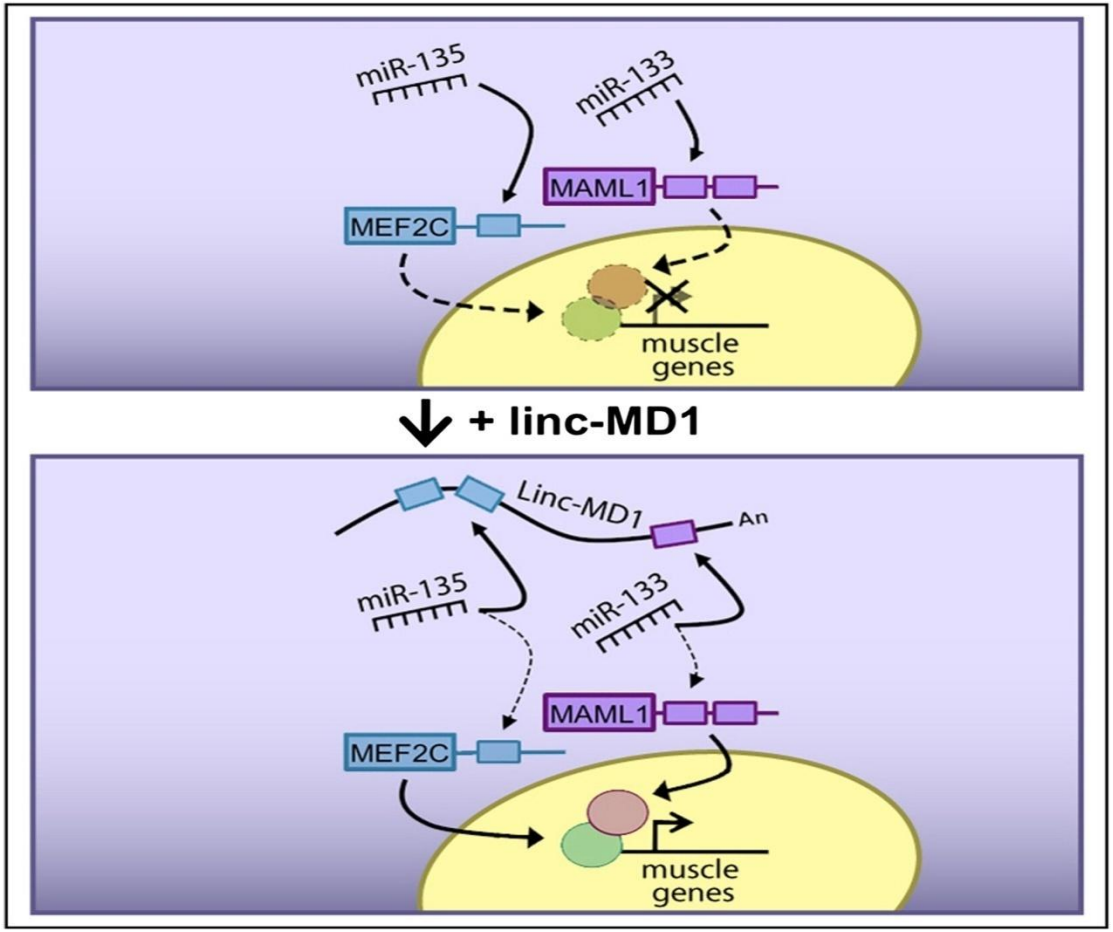

Adaptation and Fold−Change
Effective Activation (q/$h_r$)
$10^{-1}$
$10^{-2}$
$10^{-3}$
$10^{-4}$
$10^{-3}$
$10^{-2}$
$10^{-1}$
$10^{0}$
$10^{1}$
$10^{2}$
Repression Strength (1/h)
0.1
0.0002


Effective Activation
<q>/h_r>>1
<q>/h_r~ 1
<q>/h_r<<1
Robust
ON−state
Adaptation
FCD
Noise buffering
p/p_0<<1
p/p_0~0.5
p/p_0>>1
Repression Strength

# "Sponge-like" interactions

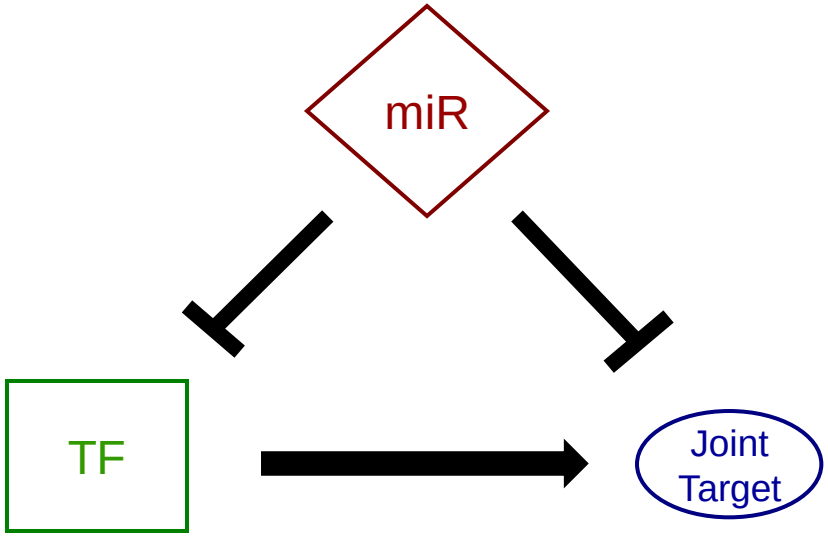
RNA transcripts can cross-talk by competing for common microRNAs. These transcripts act as "sponges" for the miRNAs thus inducing an indirect regulatory interaction on their partners

Sumazin et al. Cell 2011
Cesana et al. Cell 2011
Tay et al. Cell 2011




Example of a sponge like interaction
(from Cesana et al Cell 2011)

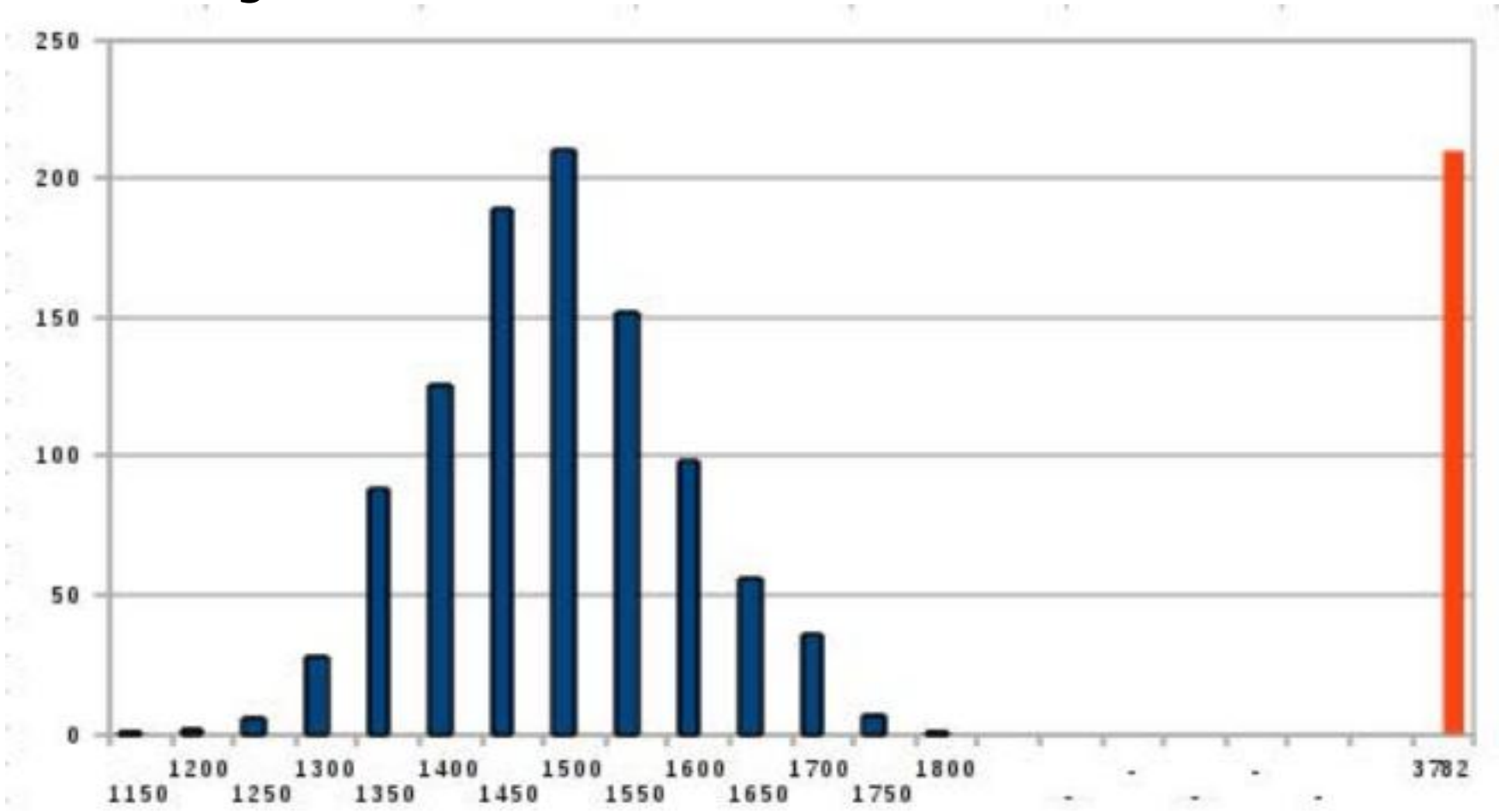
# "Sponge" loops



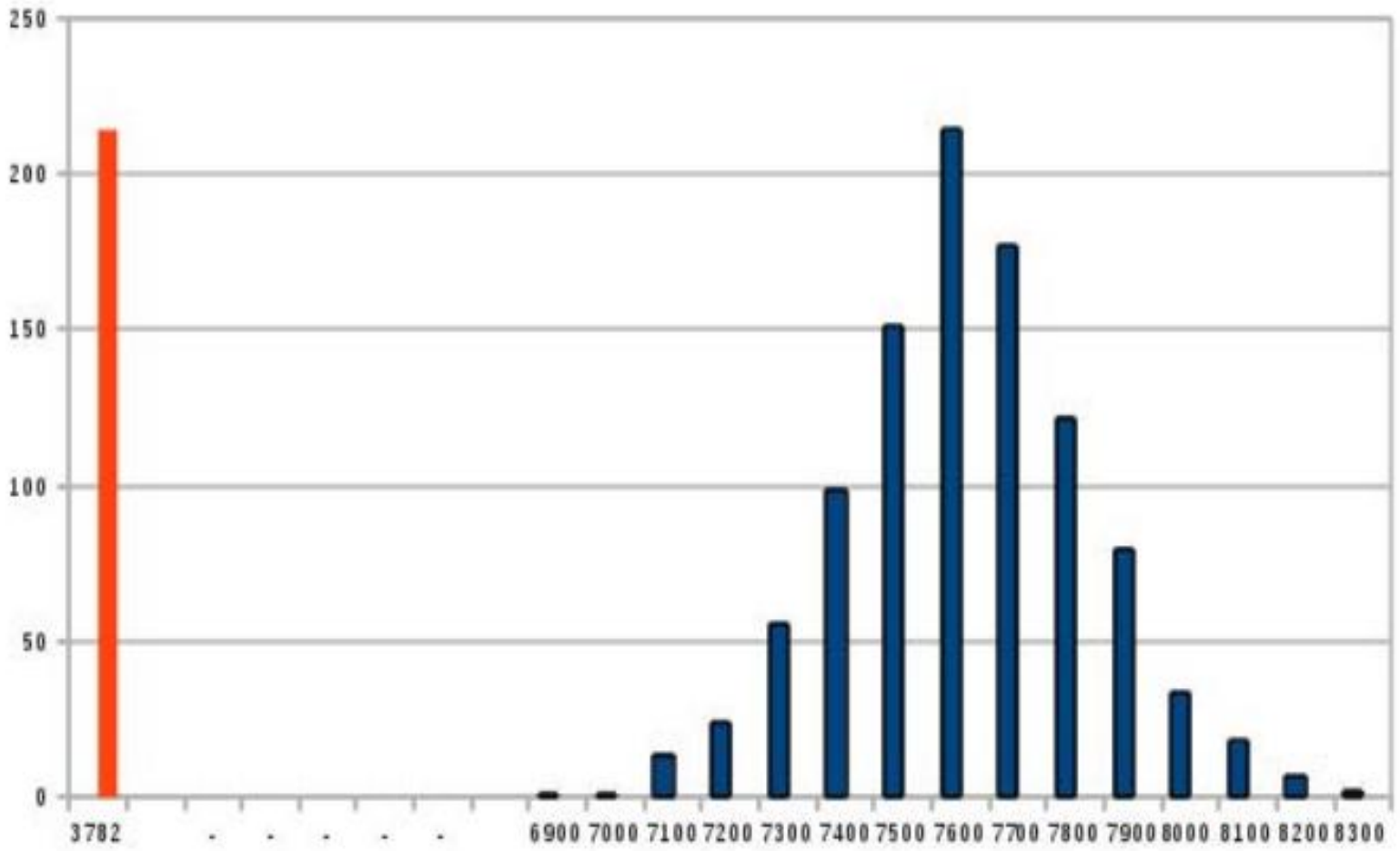

Two main functions:

- Enhance and speed up target protein production
- Correlate target-TF fluctuations: **homeostatic effect**

These loops show a very peculiar enrichment pattern:
They are strongly enriched under random reshuffling of miRNA-target links



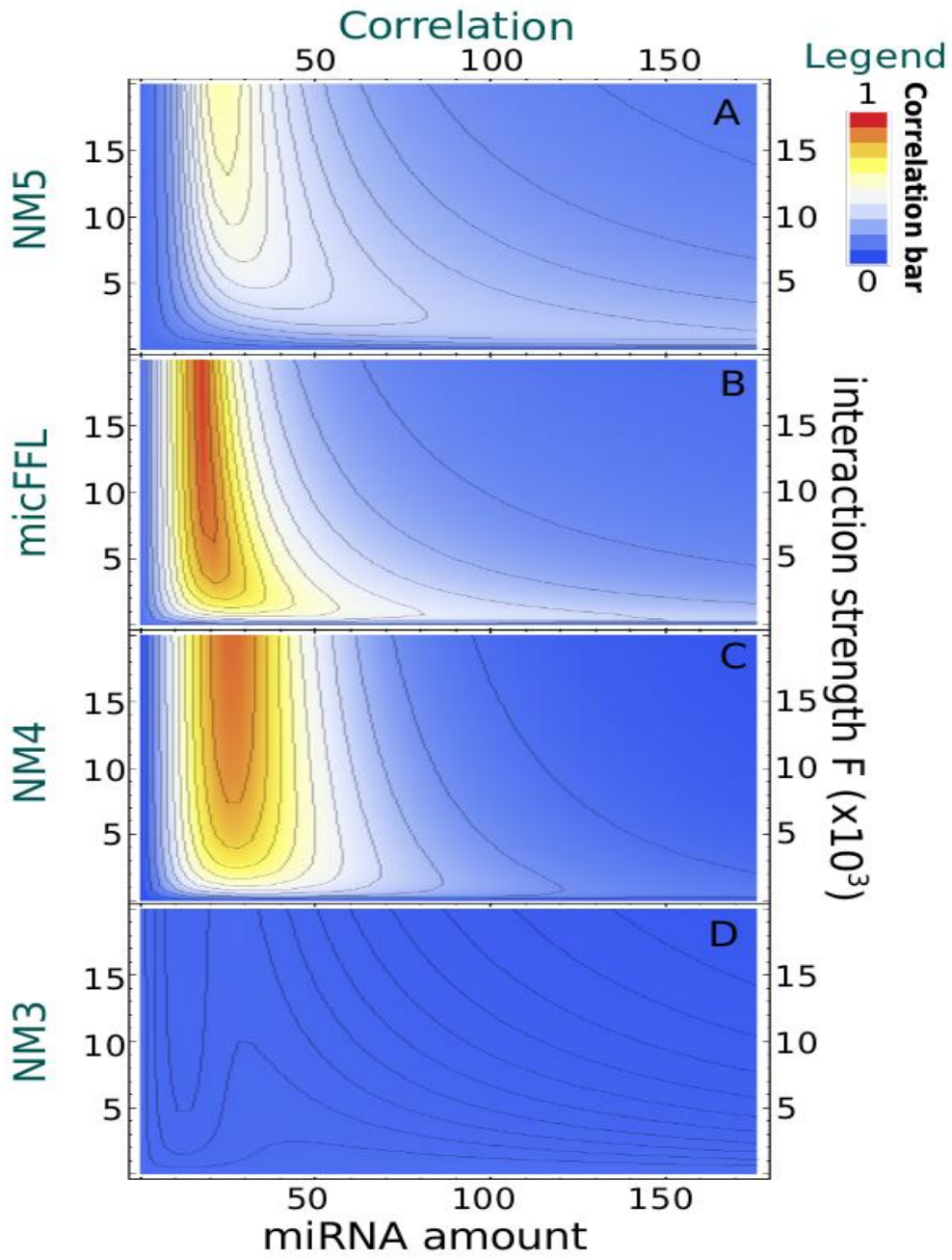
But at the same time they are strongly depleted under random reshuffling of TF-target links.

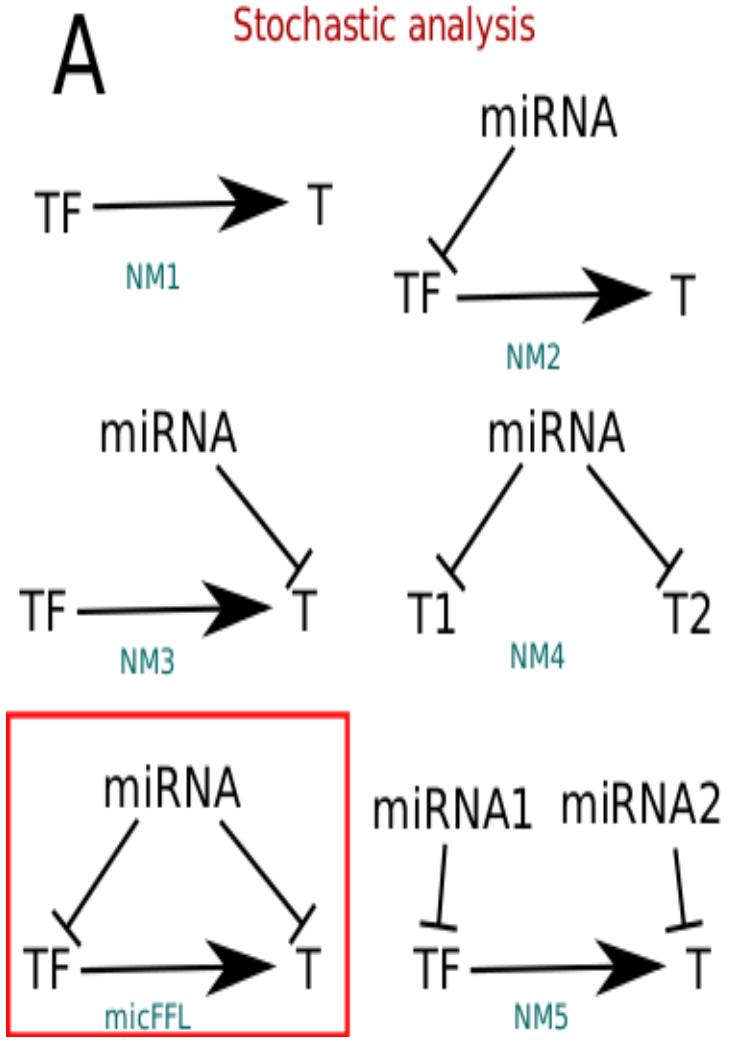


# Sponge loop functions I

This anomalous enrichment pattern is due to the peculiar behaviour of this circuit.
By varying the concentration of the miRNA one can tune the TF/target ratio to any desired value. The particular topology of the loop and the combination of direct transcriptional regulation and indirect sponge interaction is very effective in **controlling the stochastic fluctuations of this ratio.**

This circuit is present in all the situations in which the TF and its target must be kept to a fixed concentration ratio notwithstanding the environmental noise (e.g. when they are part of a complex) but should be avoided in all the other cases

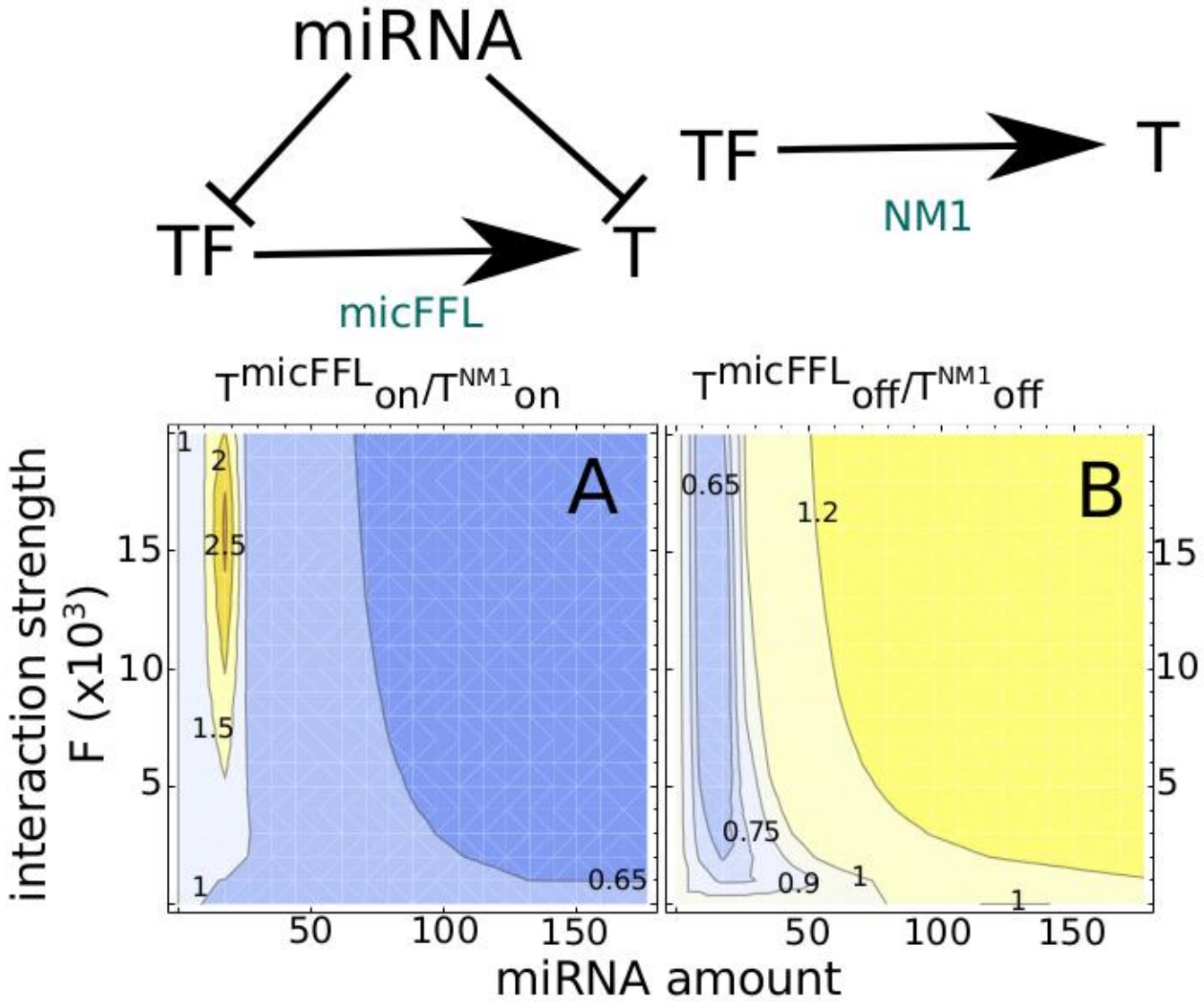

Correlation
50
100
150
Legend
1
Correlation bar
0
A
B
C
D
NM5
micFFL
NM4
NM3
15
10
5
interaction strength F (x10$^3$)
50
100
150
miRNA amount


A
Stochastic analysis
TF
T
NM1
miRNA
TF
T
NM2
miRNA
TF
T
NM3
miRNA
T1
T2
NM4
miRNA
TF
T
micFFL
miRNA1
miRNA2
TF
T
NM5

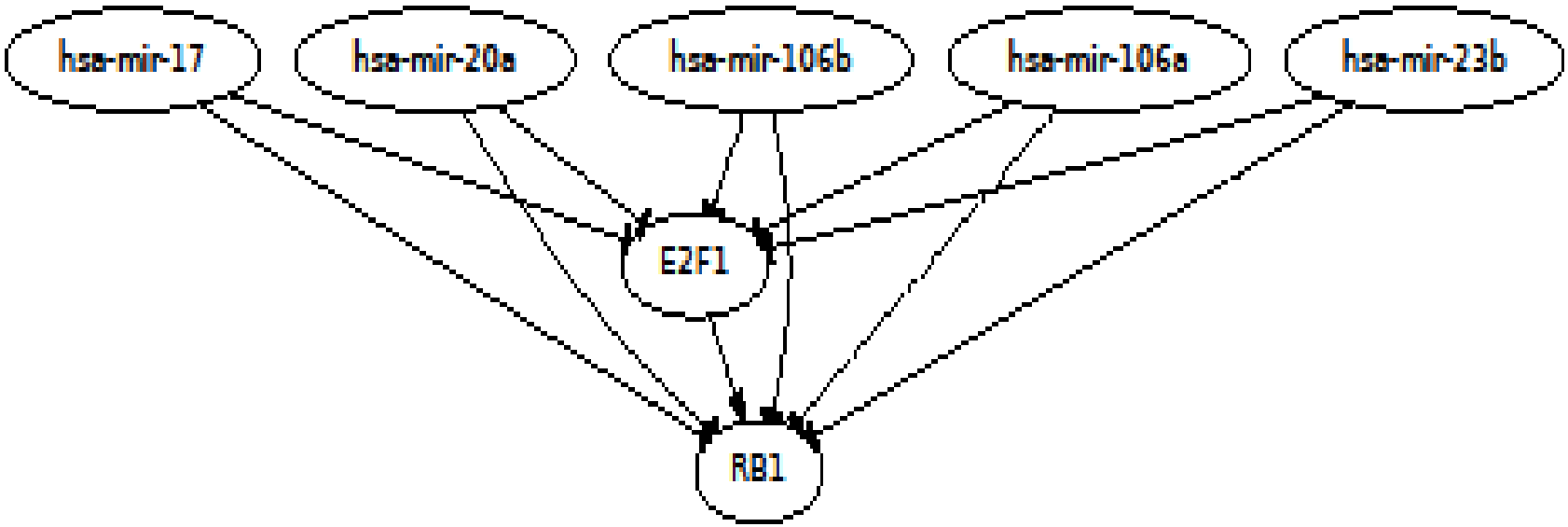
# Sponge loop functions II

The sponge loop is also able to speed up the switch on/switch off dynamics of the target.

→at high miRNA concentration the switch on time decreases
→ at low miRNA concentration the switch off time decreases


miRNA
TF
T
micFFL
TF
T
NM1
T^micFFL_on/T^NM1_on
T^micFFL_off/T^NM1_off
A
B
interaction strength
F (x10³)
15
10
5
50
100
150
miRNA amount
1
2
2.5
1.5
1
0.65
0.65
1.2
0.75
0.9
1
1

# Example : regulation of RB1



# Conclusions

→ The main purpose of our work was to systematically investigate connections between transcriptional and post-transcriptional network interactions, in the human genome.

→ We concentrated in particular on three classes of mixed circuits: miRNA mediated **Feed-Forward Loops**, mixed **selfloops** (mediated by intronic miRNAs) and **sponge loops**

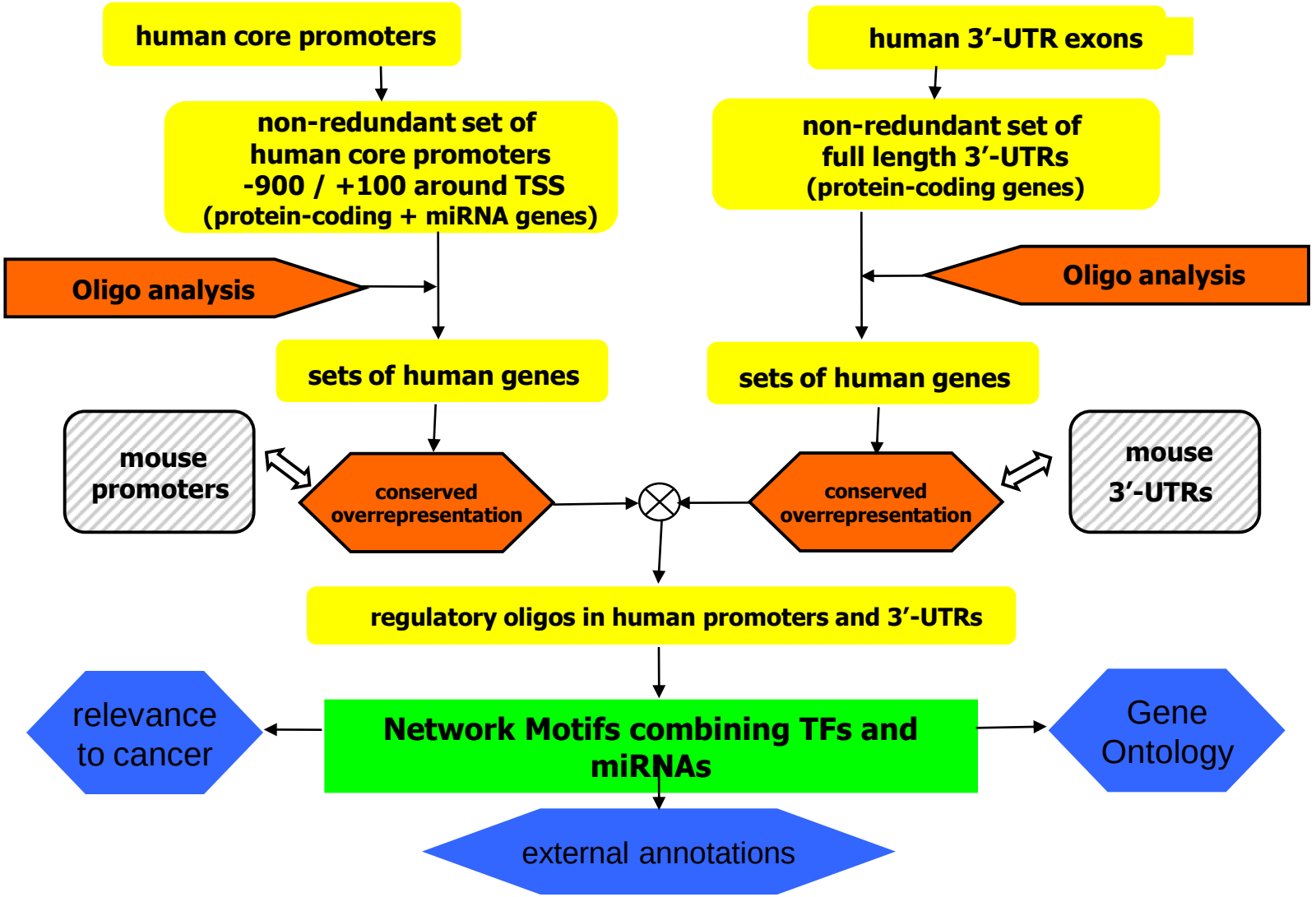
→ We have shown, solving the stochastic equation which describes these circuits that the effect of the interfering miRNA is to damp the intrinsic noise in protein production and more generally to enhance the robustness of the steady state level of the target protein concentrations

We also performed a bioinformatic search of these circuits which is available in a public database:
**http://biocluster.di.unito.it/circuits/**

# Collaborators

| | |
|---|---|
| **A. Riba, M. El Baroudi** | *Dep. of Theoretical Physics* **University of Torino** |
| **M. Osella** | *P. & M. Curie Univ.* **Paris** |
| **C. Bosia** . | *HUGEF* **Torino** |
| **D. Corà** | *IRCC* **Candiolo (TO)** |
| **A. Re** | *CIBIO* **University of Trento** |
| **D. Taverna** | *Dep. of Genetics, Biology and Biochemistry and M.B.C.* **University of Torino** |

# Pipeline


human core promoters
human 3′-UTR exons
non-redundant set of human core promoters -900 / +100 around TSS (protein-coding + miRNA genes)
non-redundant set of full length 3′-UTRs (protein-coding genes)
Oligo analysis
Oligo analysis
sets of human genes
sets of human genes
mouse promoters
conserved overrepresentation
conserved overrepresentation
mouse 3′-UTRs
regulatory oligos in human promoters and 3′-UTRs
relevance to cancer
Network Motifs combining TFs and miRNAs
Gene Ontology
external annotations