CENTRE DE RECHERCHE PUBLIC
Genomics Research Unit

UNIVERSITÉ DU LUXEMBOURG

Interplay between microRNAs, Transcription Factors, and Target Genes: Linking Dynamic Expression Changes to Function

Petr Nazarov

23-07-2013

Nazarov, Reinsbach et al. Nucleic Acids Res. **2013**, 41(5):2817-31

Nazarov PV, Reinsbach SE, Muller A, Nicot N, Philippidou D, Vallar L, Kreis S. *Interplay of microRNAs, transcription factors and target genes: linking dynamic expression changes to function.*

Nucleic Acids Res. 2013; 41(5):2817-31.

MicroRNAs

- Small, non-coding RNAs (~22 nt)
- Negative post-transcriptional regulators

- Pre-miR → 2 **strands**, historically called **guide** and **passenger** strands. However passenger **may be of equivalent importance**.

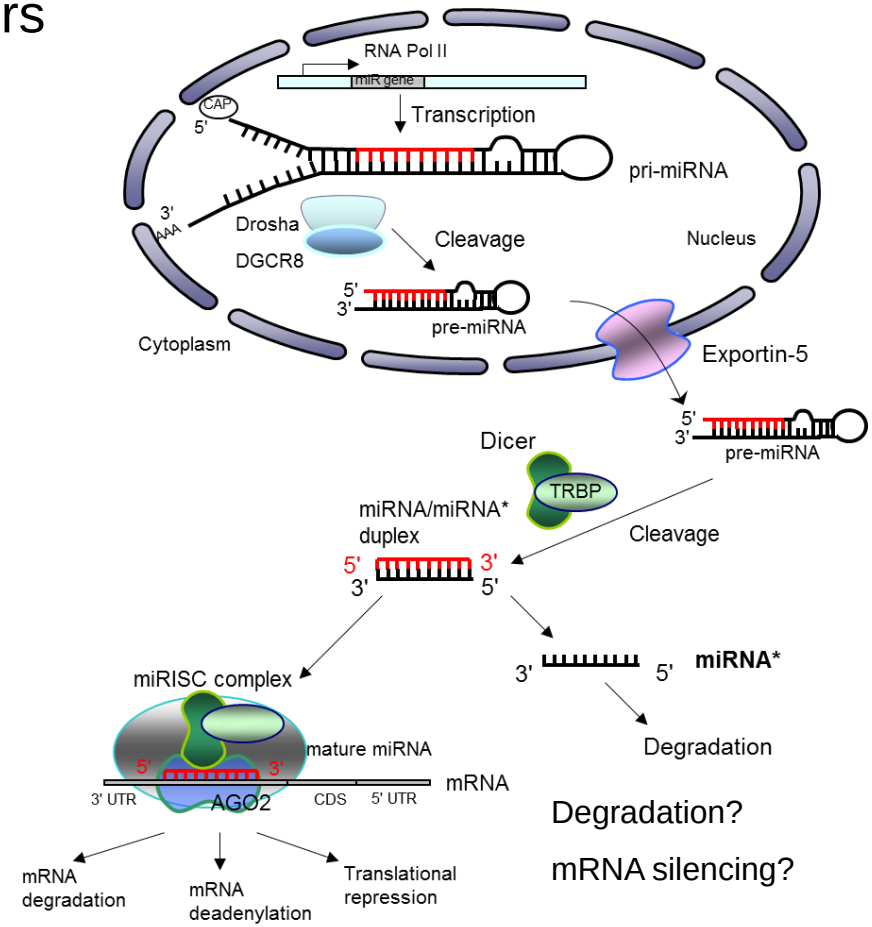
- Involved in many cellular processes and diseases. ~ 60% of genes are regulated by miRNAs.

- Roles in cancer:

→ tumour-suppressing miRNAs (miR-15a)

→ oncogenic miRNAs (miR-21)

Canonical miRNA production.
Adapted from Winter, Nat.Cell Biol. 2009



Reinsbach, et al.(2012) RNA Biology

PROBLEMS

- miRNAs are short and in principle may be not entire complementary to target gene mRNA sequence. This hampers prediction of target gene, and results in a high level of FPs.

- miRNA expression depends on cell type and biological condition.

- Precise mechanisms of miRNA actions are debated.

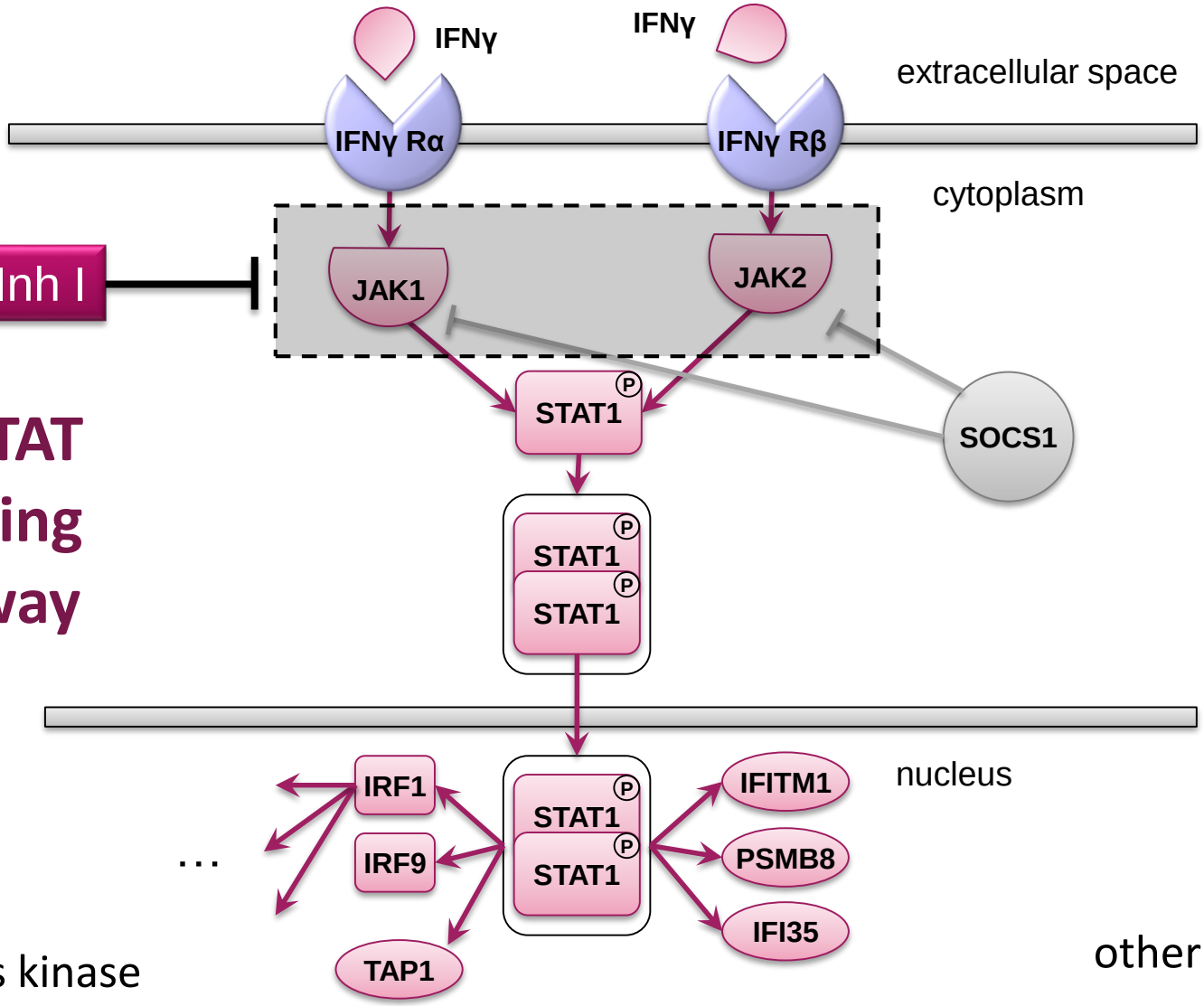
Understanding of mechanisms: cross-sectional studies may not provide enough information (cause-response problem), therefore **dynamic regulation of miRNA expression** changes have to be considered to answer this question.

GOAL and OUTLINE

The aim of the our study was to integrate time-resolved **miRNA**, **mRNA** data and existing **knowledge** about their interaction and functions in order to obtain new information about biological processes.

- Find robust list of miRNA-targeted genes: miR -| mRNA
- Compose/verify a pipeline for analysis of time-resolved MA data
- Investigate dynamics of active biological functions in IFNγ-stimulated melanoma cell line
- Integrate miRNA, mRNA and inferred upstream transcription regulators
- Detect topological motifs in regulation network

EXPERIMENTAL DESIGN

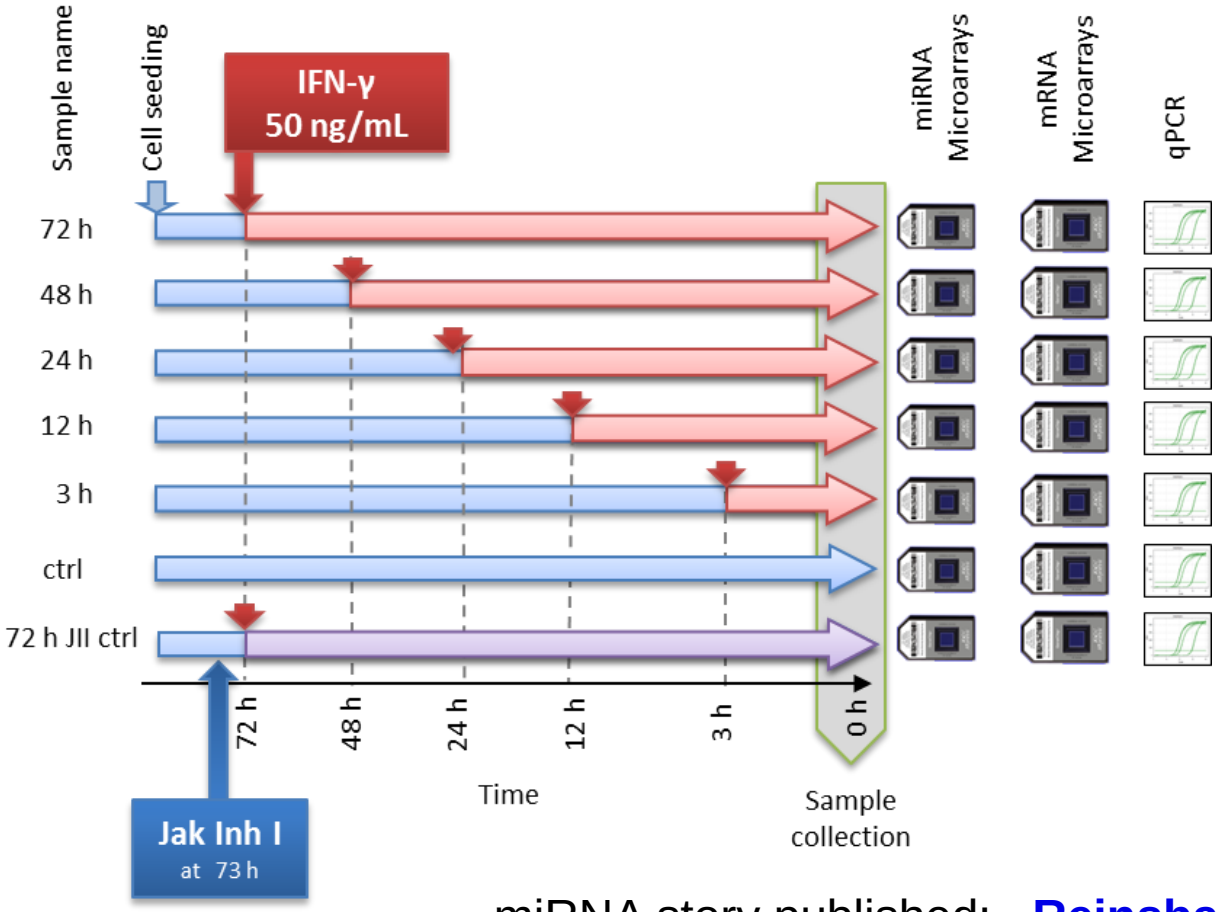


JAK-STAT signaling pathway

JAK = Janus kinase

EXPERIMENTAL DESIGN

Human melanoma **A375** cells were seeded together and cultured until sample collection. Cells were IFNγ stimulated at different time points.



mRNA
Affymetrix
HuGene 1.0 ST arrays

miRNA
Affymetrix
miRNA v2 arrays

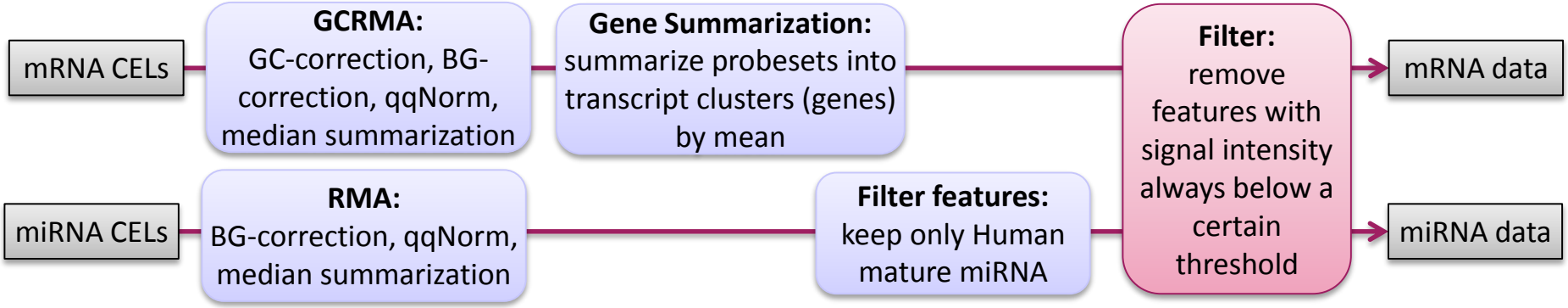
miRBase v.15

1105 human mature miRs

miRNA story published: **Reinsbach, et al (2012) RNA Biology**

PREPROCESSING PIPELINE

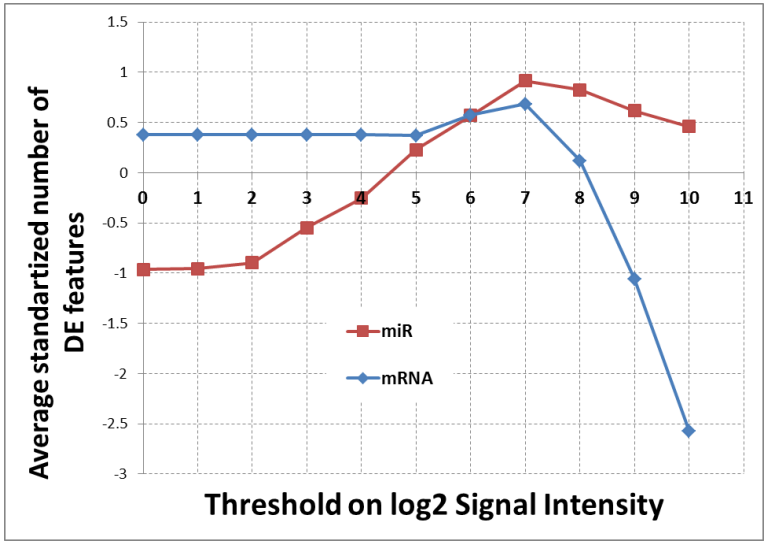
Standard pre-processing scheme was used in **Partek® GS** *(alternatively one can use APT)*



It can be beneficial to remove lowly expressed features

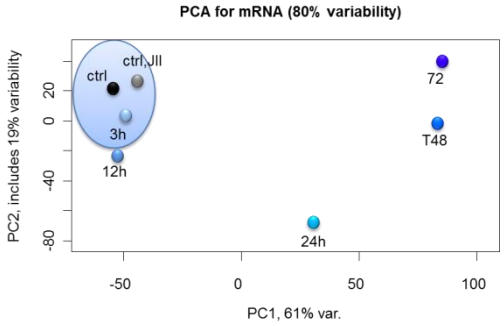
Features with expression (log2 signal intensity) constantly below a certain threshold may be removed.

1. This improves number of significant DEG
2. and improves qPCR validation afterwards.

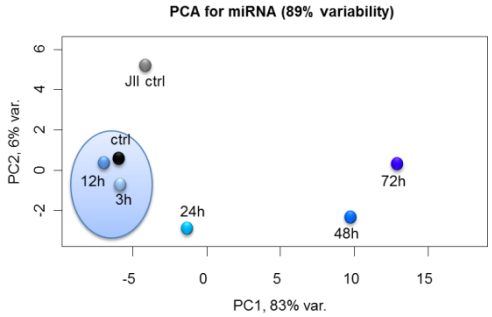


Based on these profiles and our previous experience we selected threshold = 7

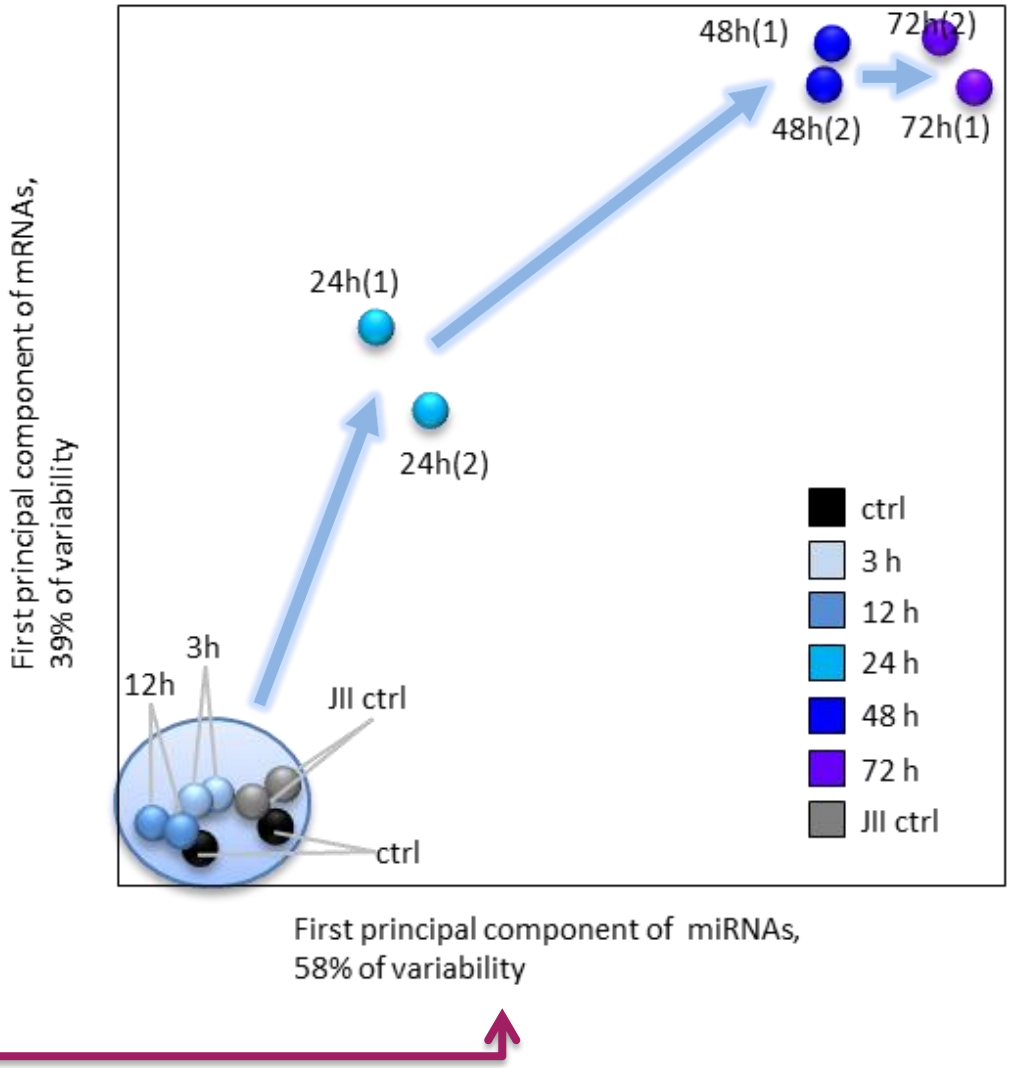
PCA: DATA OVERVIEW



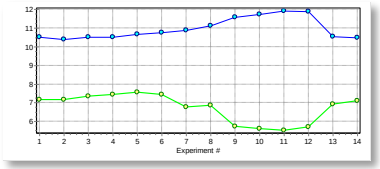
PC_1 of mRNA



PC_1 of miR



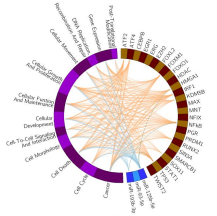
ANALYSIS PIPELINE



Co-expression Analysis (CEA)

Potential miRNA targets

Highly connected clusters



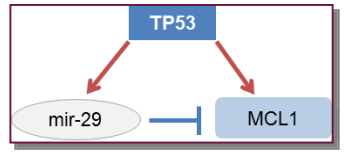
miRNA data

mRNA data

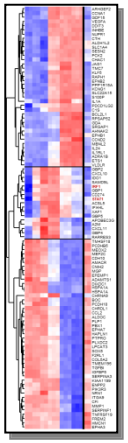
Integrative Analysis and Presentation

CIRCOS

Network motifs



Differential Expression Analysis (DEA)



Dynamics of functions

Significant miRNA and mRNA

Functional Analysis (IPA)

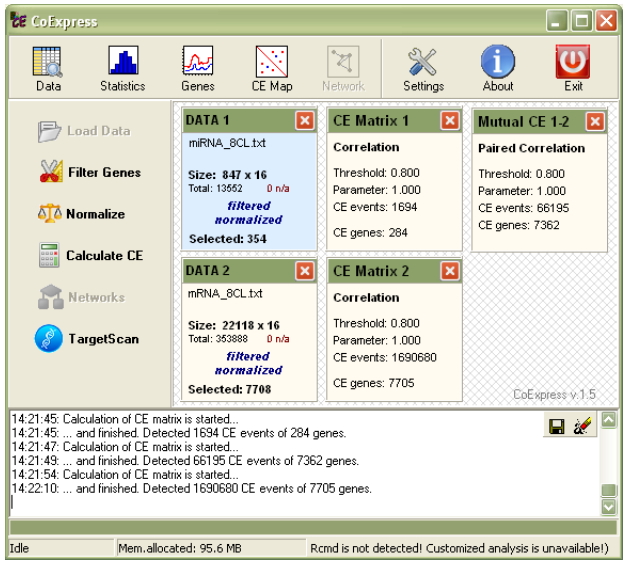
Infering of up-stream regulators (IPA)

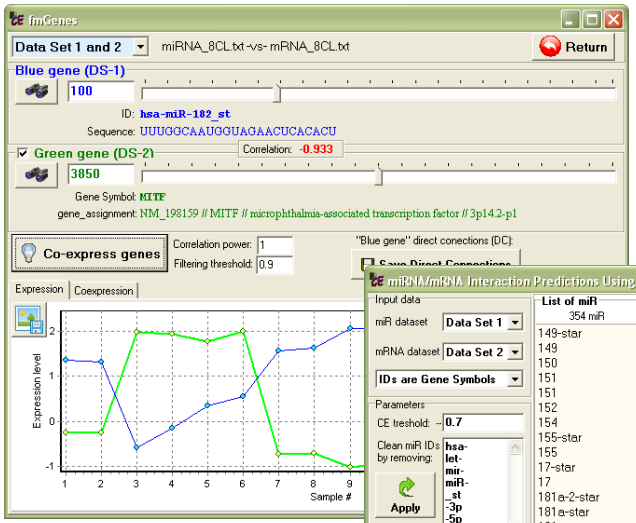
Up-stream transcription regulators

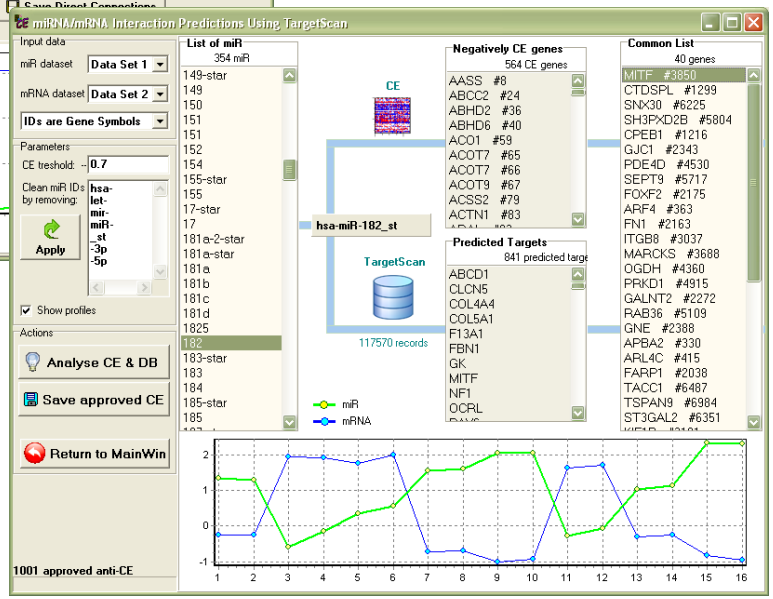
CO-EXPRESSION ANALYSIS

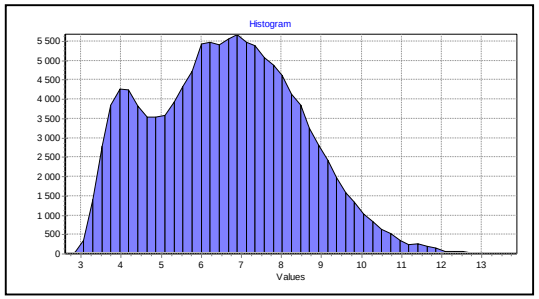
- Analysis was performed using a user-friendly tool **CoExpress**:

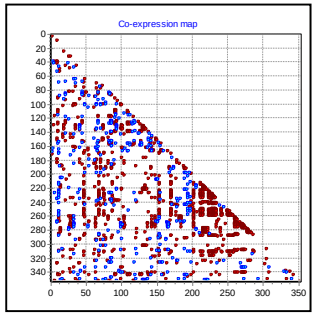
http://www.bioinformatics.lu/CoExpress/









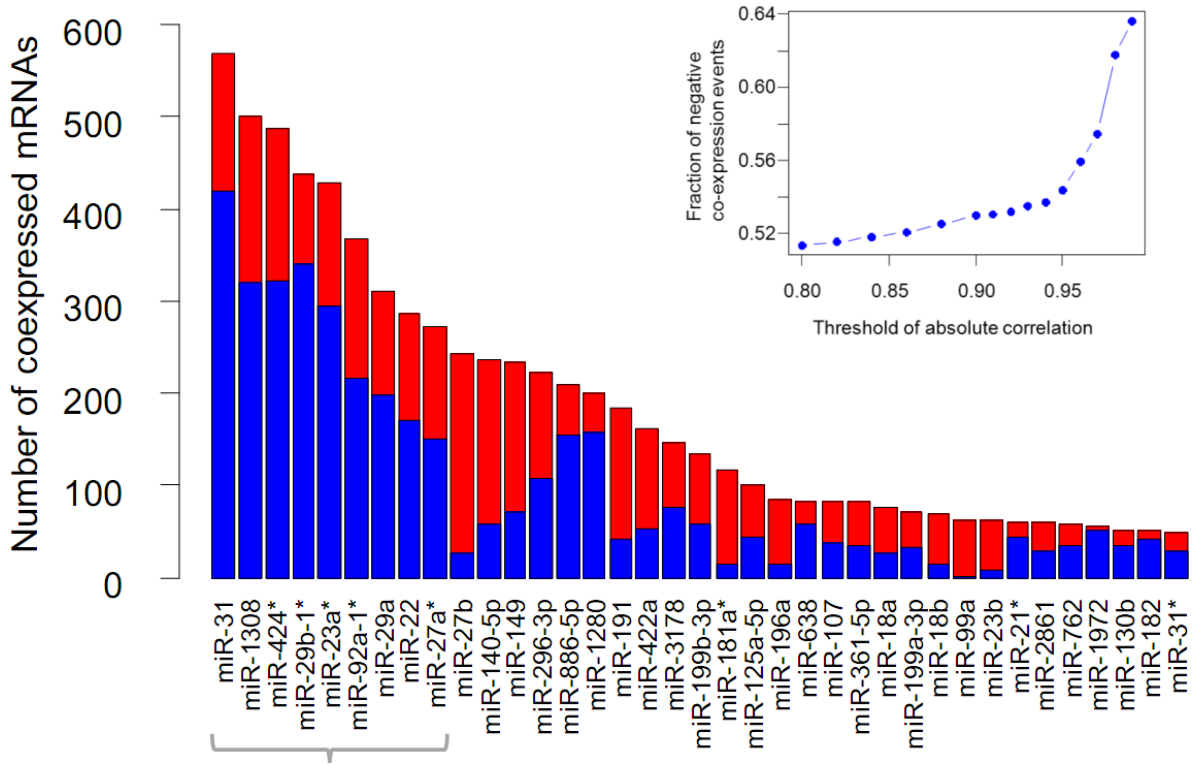


CO-EXPRESSION ANALYSIS

- Analysis was performed using a user-friendly tool **CoExpress**:

http://www.bioinformatics.lu/CoExpress/

Top co-expressed miRNAs:

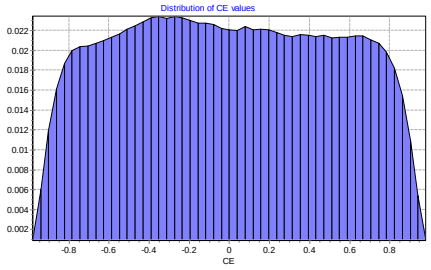


Negative miRNA:mRNA correlations exceed positive ones

CO-EXPRESSION ANALYSIS

http://www.bioinformatics.lu/CoExpress/

Because of the lack of independent points we cannot use co-expression as a discovery tool. **However CE analysis can be combined with existing knowledge:**



TargetScan ,

TarBase,

etc...

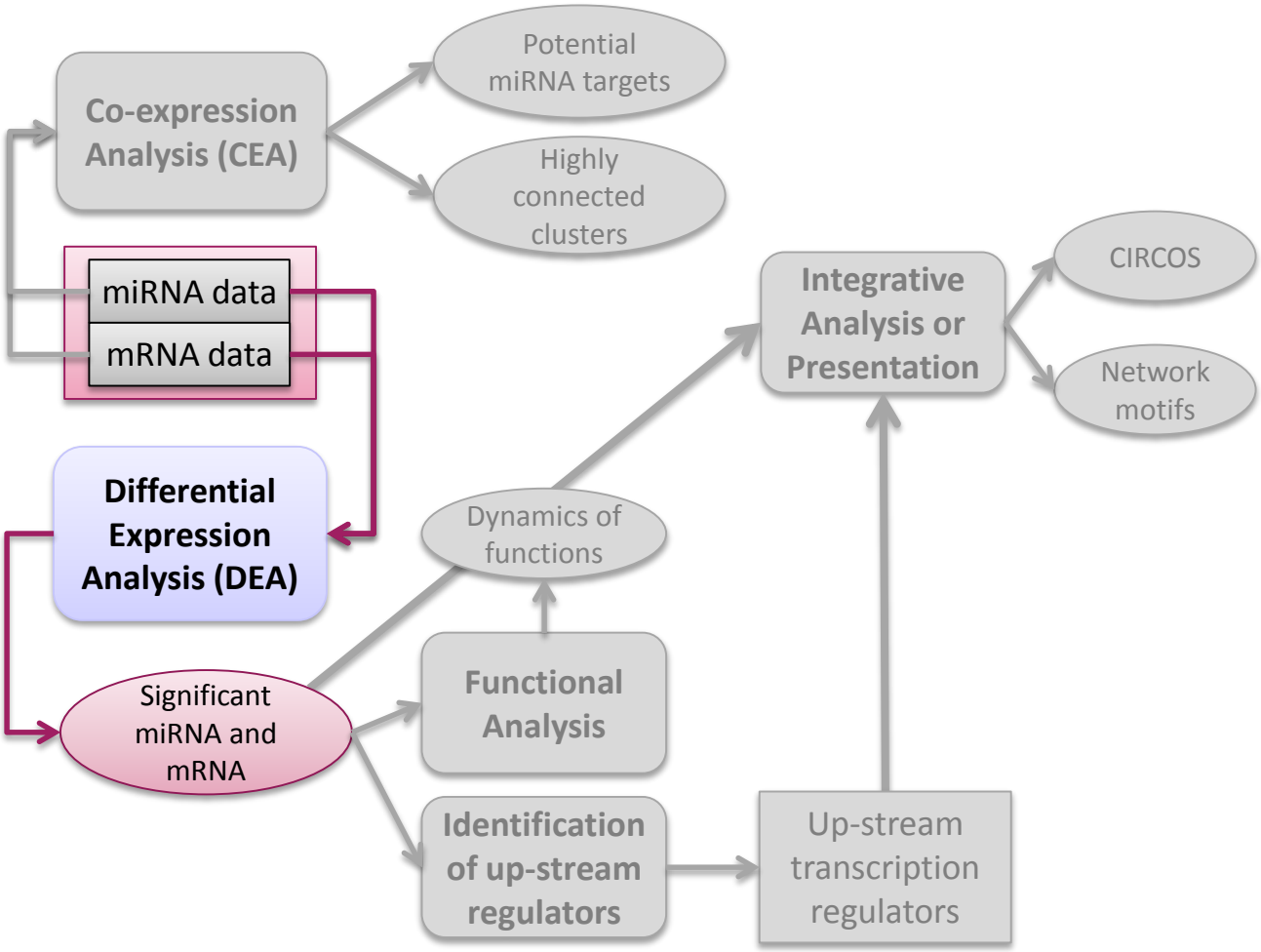
Use known or predicted interactions to filter observed negatively correlated miRNA:mRNA pairs

List of high quality miRNA-targets

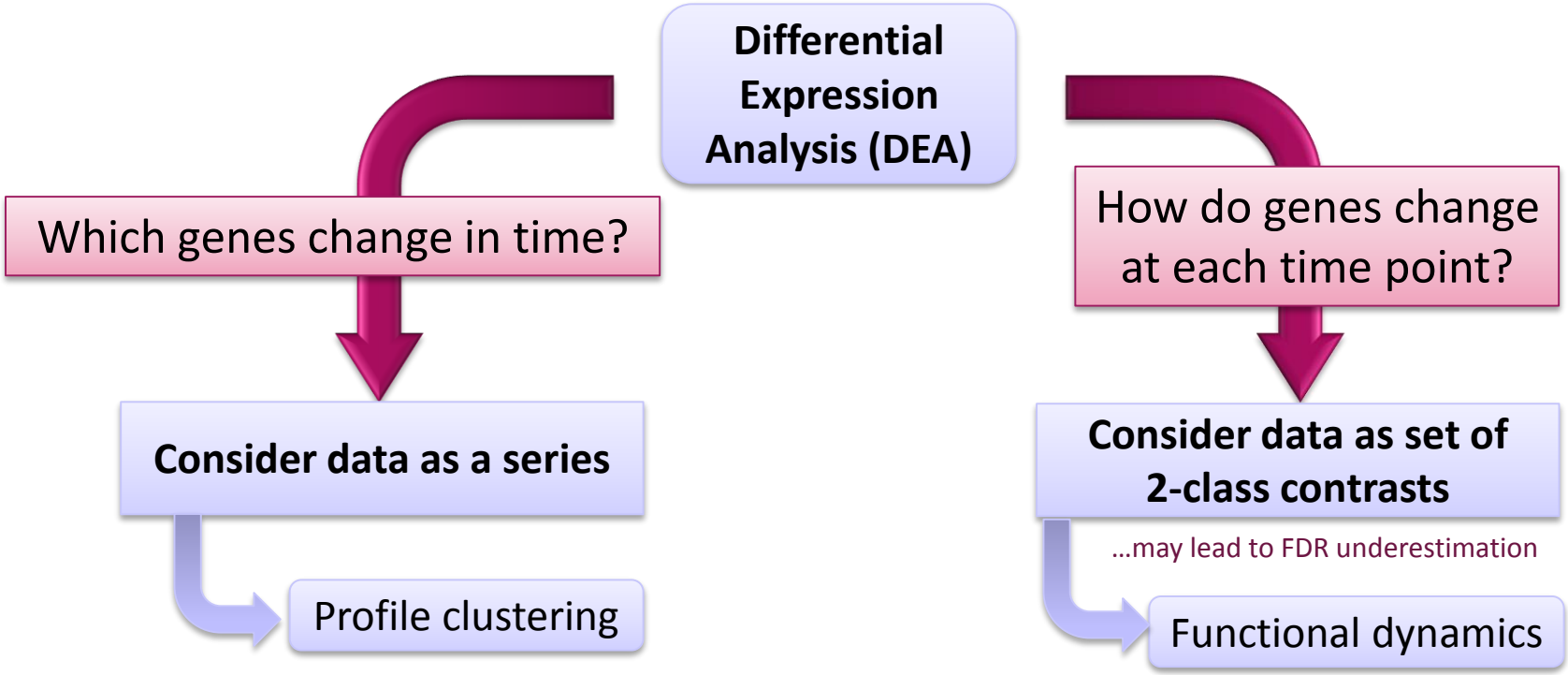
TargetScan v6: we observed **398** predicted miRNA –| mRNA interactions (with r < -0.9)

TarBase v6: we observed **17** known miRNA –| mRNA interactions (with r < -0.9)

ANALYSIS PIPELINE



DIFFERENTIAL EXPRESSION ANALYSIS



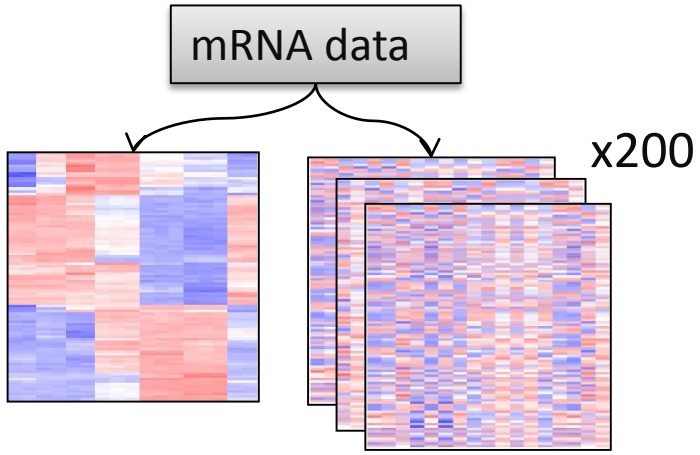
We tested several DEA methods, applicable for time course MA data

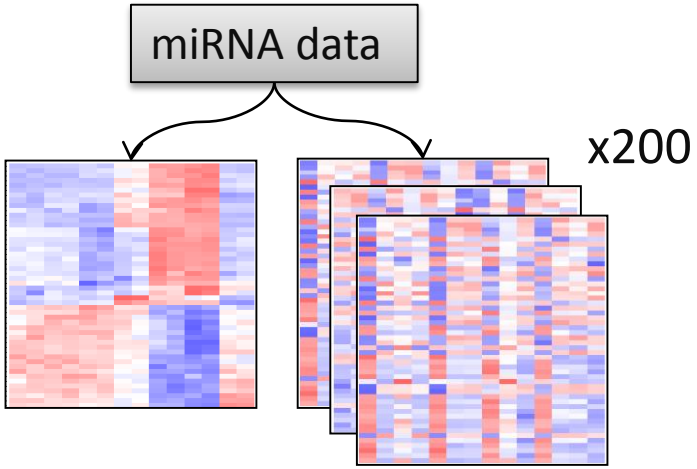
- **Limma** (linear models for microarray data) → Smyth, GK (2005). Limma: linear models for microarray data
- **BETR** (Bayesian Estimation of Temporal Regulation) → Aryee, MJ (2009) BMC Bioinformatics
- **timecourse** R-pkg (multivariate empirical Bayes models) → Tai YC, Speed TP (2006) Annals of Statistics

DIFFERENTIAL EXPRESSION ANALYSIS

Data for Benchmarking

Datasets + 200 permutations:

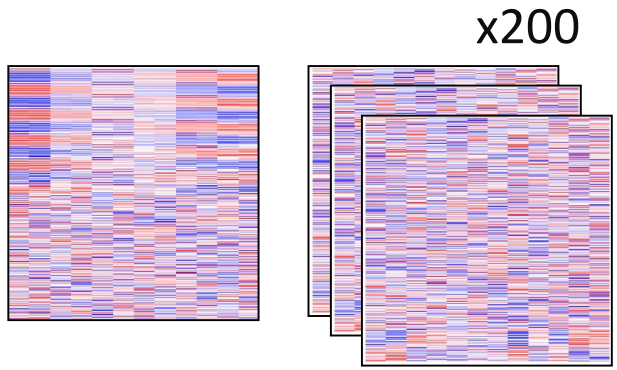




Synthetic data

Synthetic data set consisted of 700 features.

- 350 of those represented **various profiles** (peak, smooth peak, steep regulation, smooth regulation, linear) mixed with normal noise in different proportions, so that average r2 of profiles and noisy data varied between 0.999 and 0.5.
- Another 350 features were **completely random**.



DIFFERENTIAL EXPRESSION ANALYSIS

Benchmarking Results

Data + 200 permutations: miRNA data | mRNA data | Synthetic data*

☺

limma	1. miRNA data set		2. mRNA dataset		3. Synthetic dataset		
Desired FDR	Significant	Permut. FDR	Significant	Permut. FDR	Significant	Permut. FDR	Real FDR
0.001	65	0.001	4681	0.000	218	0	0.000
0.01	92	0.003	7935	0.000	261	0	0.004
0.05	115	0.005	10766	0.001	305	0	0.046

☹

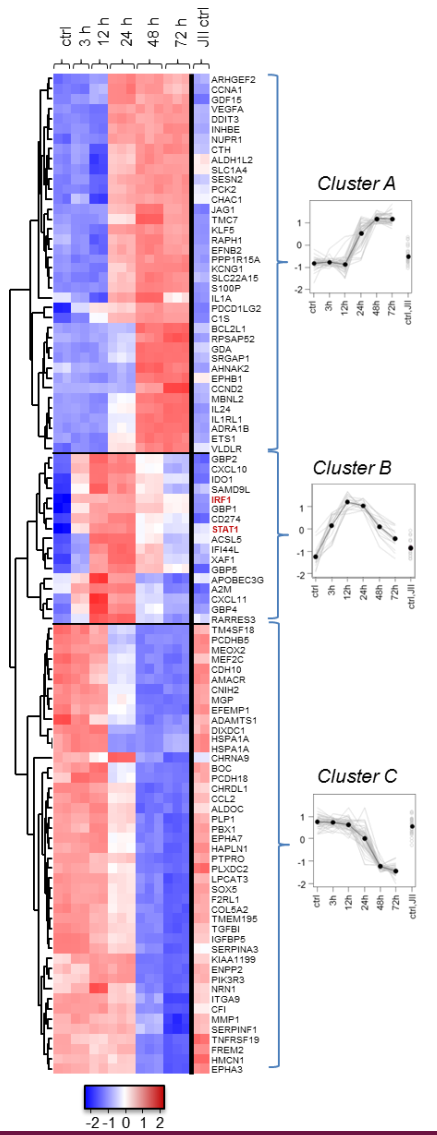
betr	1. miRNA data set		2. mRNA dataset		3. Synthetic dataset		
Desired FDR	Significant	Permut. FDR	Significant	Permut. FDR	Significant	Permut. FDR	Real FDR
0.001	98	0.106	11485	0.078	365	0.023	0.142
0.01	104	0.136	12533	0.105	392	0.033	0.184
0.05	109	0.164	13370	0.133	432	0.042	0.229

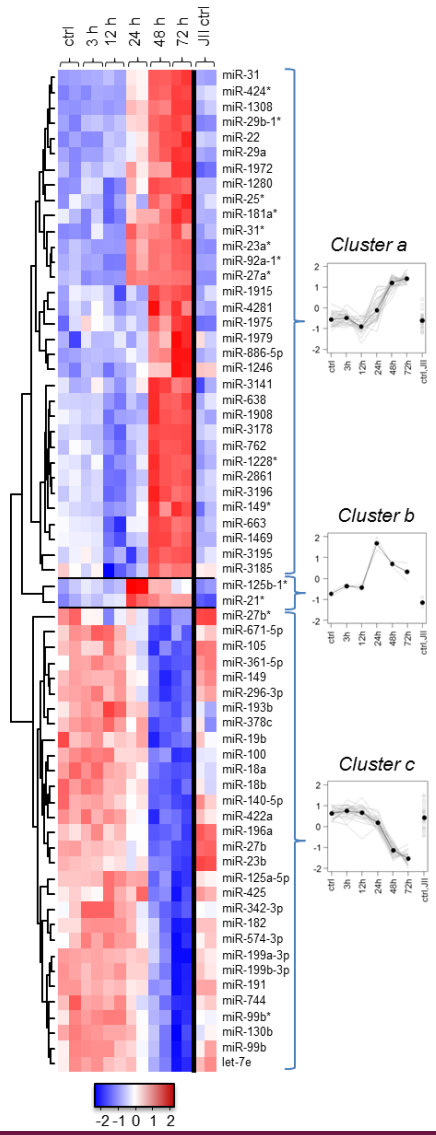
?

timecourse	1. miRNA data set		2. mRNA dataset		3. Synthetic dataset		
Desired FDR	Significant	Permut. FDR	Significant	Permut. FDR	Significant	Permut. FDR	Real FDR
0.001	25	0.001	461	0.001	520	0.001	0.352
0.01	38	0.009	1778	0.010	562	0.010	0.391
0.05	68	0.050	8022	0.050	606	0.049	0.426

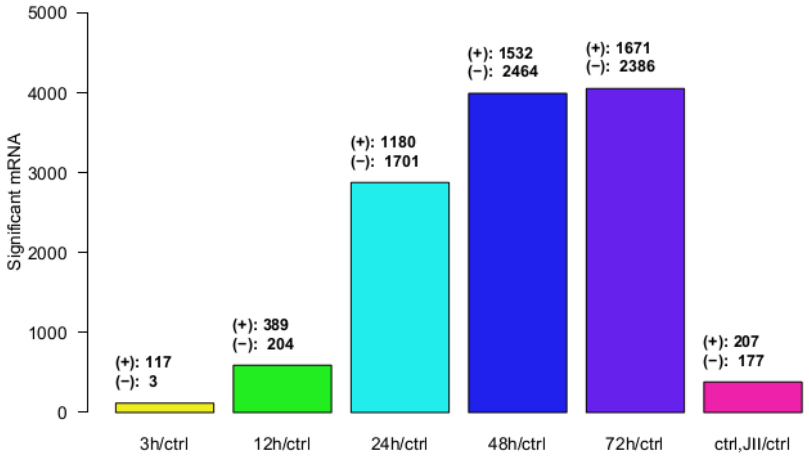
(*) Synthetic data set consisted of 700 features. 350 of those represented various profiles (peak, smooth peak, steep regulation, smooth regulation, linear) mixed with normal noise in different proportions, so that average r2 of profiles and noisy data varied between 0.999 and 0.5. Another 350 features were completely random.

DIFFERENTIAL EXPRESSION ANALYSIS

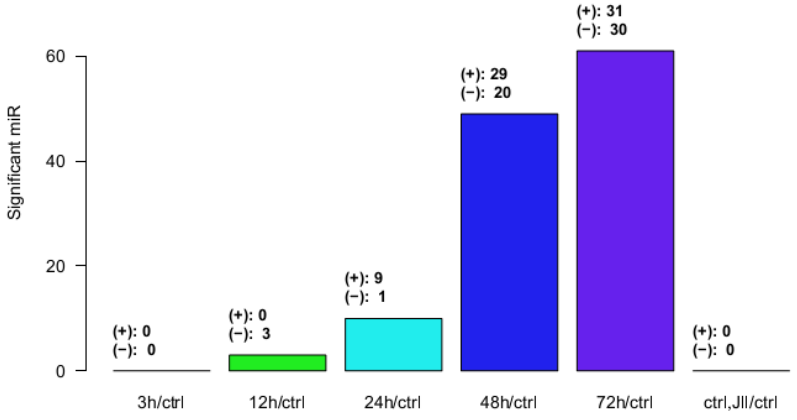




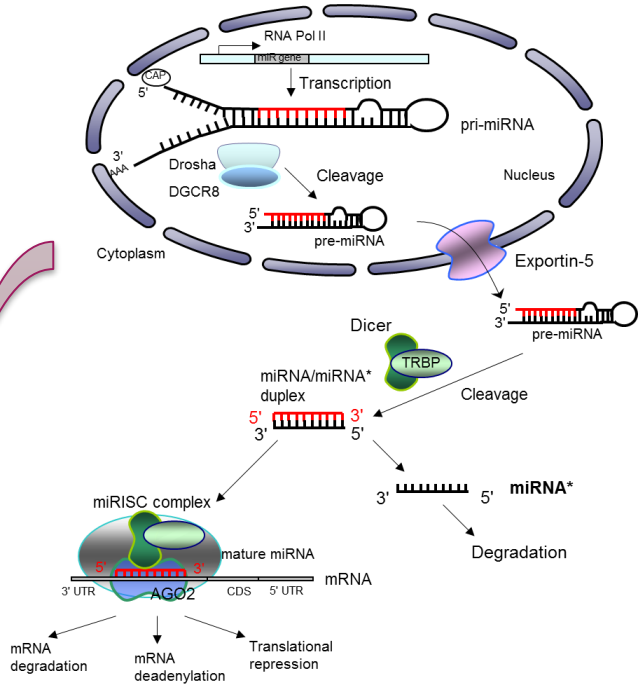
mRNA

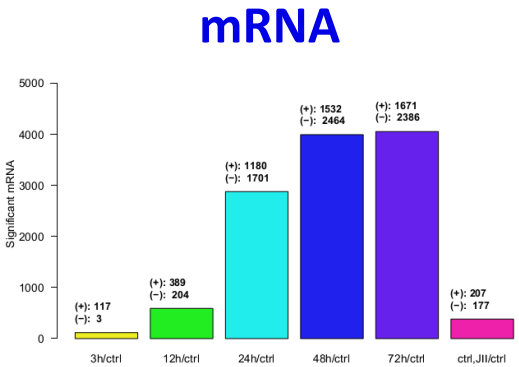


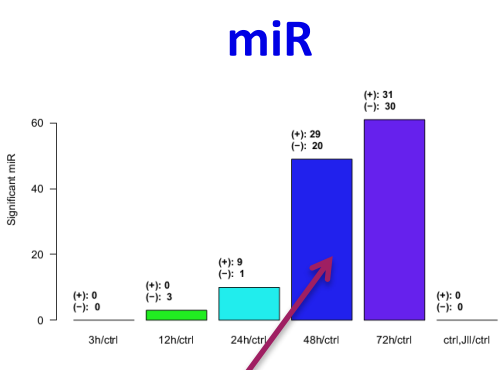
miR



PROFILES

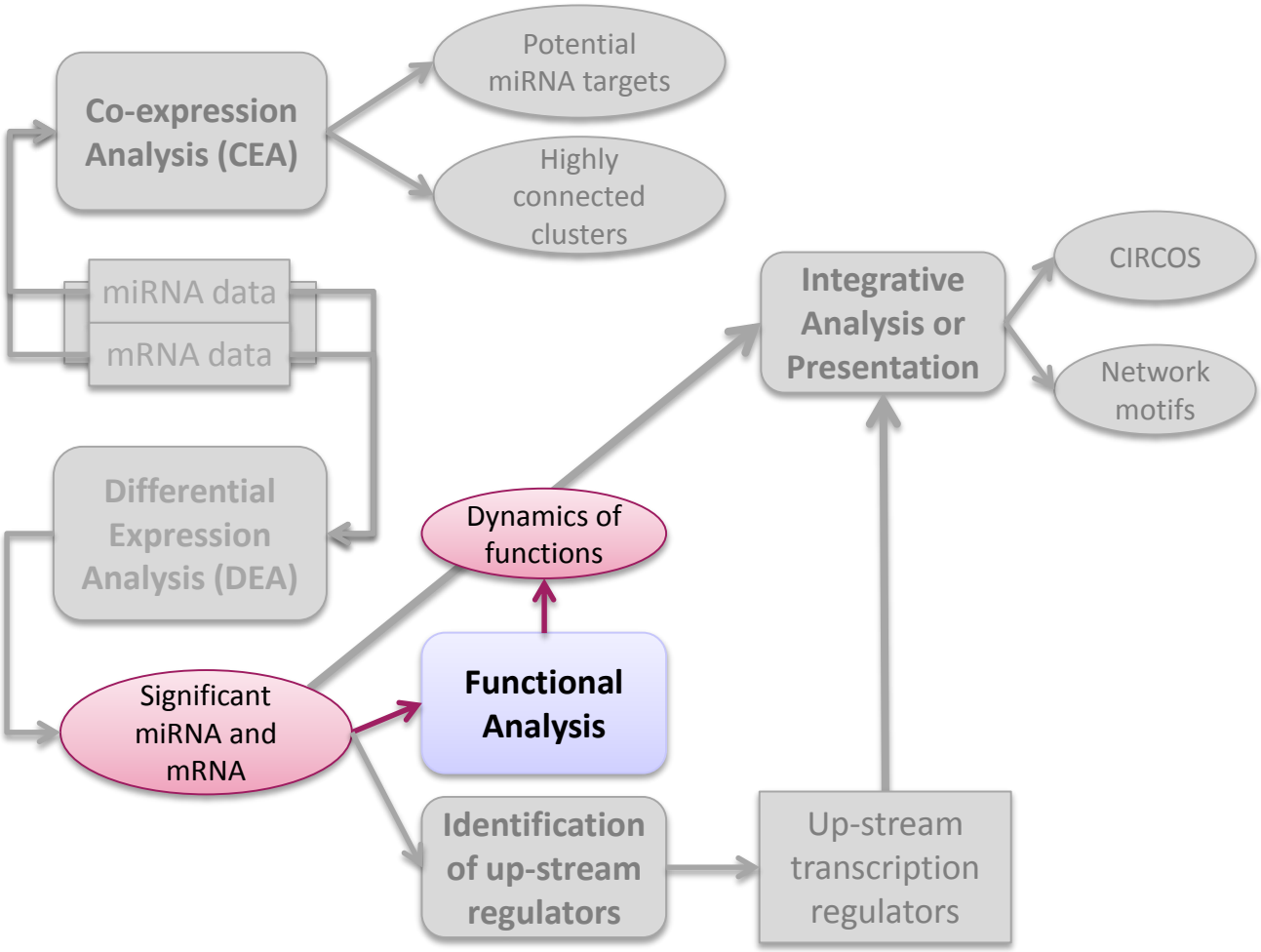






Gene	Complex	3h	12h	24h	48h	72h	<Expr>	FDR
DICER1	RISC	0.078	-0.209	-0.179	0.645	0.880	10.4	6.0E-08
DGCR8	Drosha	-0.056	-0.366	-0.065	0.520	0.604	8.4	8.6E-07
EIF2C2 (Ago2)	RISC	-0.157	-0.351	0.162	0.562	0.569	9.0	5.3E-06
RNASEN	Drosha	-0.009	-0.159	-0.239	-0.227	-0.254	10.3	8.2E-03

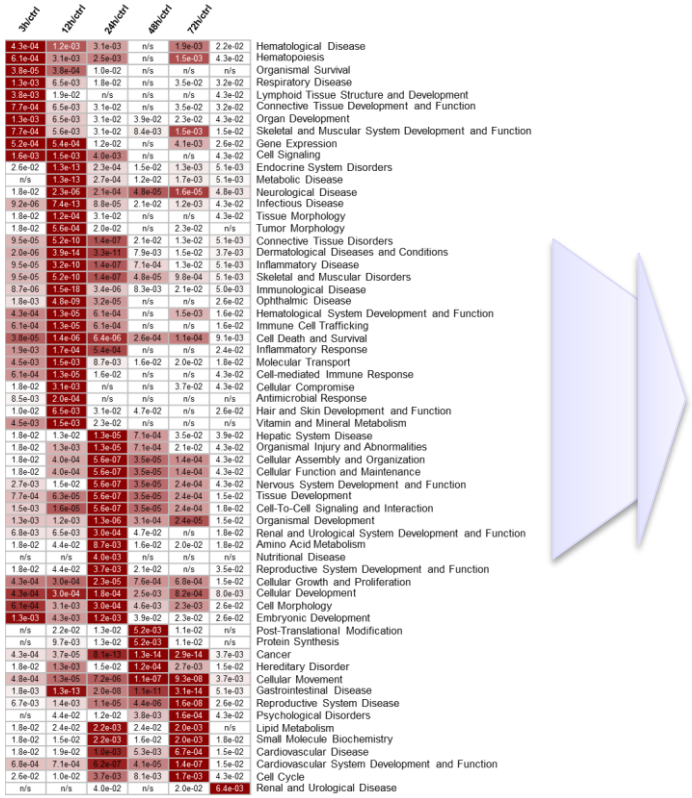
ANALYSIS PIPELINE



FUNCTIONAL ANALYSIS

DEA miRNA:mRNA → Ingenuity® PA functional annotation → Enriched functions with FDR < 0.01 → Combined into categories

Fisher's exact test,
Benjamini-Hochberg correction



3h/ctrl	12h/ctrl	24h/ctrl	48h/ctrl	72h/ctrl	JII ctrl/ctrl	
5.2e-4	5.4e-4	1.2e-2	n/s	4.1e-3	2.6e-2	Gene Expression
1.7e-3	1.5e-3	4.0e-3	n/s	n/s	4.3e-2	Cell Signalling
6.1e-4	1.3e-5	1.6e-2	n/s	n/s	4.3e-2	Cell-mediated Immune Response
6.1e-4	1.3e-5	6.1e-4	n/s	n/s	1.6e-2	Immune Cell Trafficking
4.5e-3	1.5e-3	8.7e-3	1.6e-2	2.0e-2	1.8e-2	Molecular Transport
9.2e-6	7.4e-13	8.8e-5	2.1e-2	1.1e-3	4.3e-2	Infectious Disease
8.7e-6	1.5e-18	3.4e-6	8.3e-3	2.1e-2	5.0e-3	Immunological Disease
9.5e-5	3.2e-10	1.4e-7	7.1e-4	1.3e-2	5.1e-3	Inflammatory Disease
1.9e-3	1.7e-4	5.4e-4	n/s	n/s	2.4e-2	Inflammatory Response
2.0e-6	3.9e-14	3.3e-11	7.9e-3	1.5e-2	3.7e-3	Dermatological Diseases and Conditions
1.0e-2	6.5e-3	3.1e-2	4.7e-2	n/s	2.6e-2	Hair and Skin Development and Function
1.8e-2	5.6e-4	2.0e-2	n/s	2.3e-2	n/s	Tumor Morphology
n/s	1.3e-13	2.7e-4	1.2e-2	1.7e-3	5.1e-3	Metabolic Disease
3.8e-5	1.4e-6	6.4e-6	2.6e-4	1.1e-4	9.1e-3	Cell Death and Survival
6.1e-4	3.1e-3	3.0e-4	4.6e-3	2.3e-3	2.6e-2	Cell Morphology
4.3e-4	3.0e-4	1.8e-4	2.5e-3	8.3e-4	7.9e-3	Cellular Development
4.3e-4	3.0e-4	2.2e-5	7.7e-4	6.8e-4	1.5e-2	Cellular Growth and Proliferation
1.5e-3	1.6e-5	5.6e-7	3.4e-5	2.4e-4	1.8e-2	Cell-To-Cell Signalling and Interaction
7.7e-4	6.3e-5	5.6e-7	3.4e-5	2.4e-4	1.5e-2	Tissue Development
1.8e-2	4.0e-4	5.6e-7	3.4e-5	1.4e-4	4.3e-2	Cellular Assembly and Organisation
n/s	2.2e-2	1.4e-2	5.2e-3	1.1e-2	n/s	Post-Translational Modification
n/s	9.7e-3	1.4e-2	5.2e-3	1.1e-2	n/s	Protein Synthesis
4.3e-4	3.7e-5	8.1e-13	1.3e-14	2.9e-14	3.7e-3	Cancer
4.8e-4	1.3e-5	7.2e-6	1.1e-7	9.4e-8	3.7e-3	Cellular Movement
2.6e-2	1.0e-2	3.8e-3	8.1e-3	1.7e-3	4.3e-2	Cell Cycle

FUNCTIONAL ANALYSIS

Using P-value and Z-score (only Top 50)

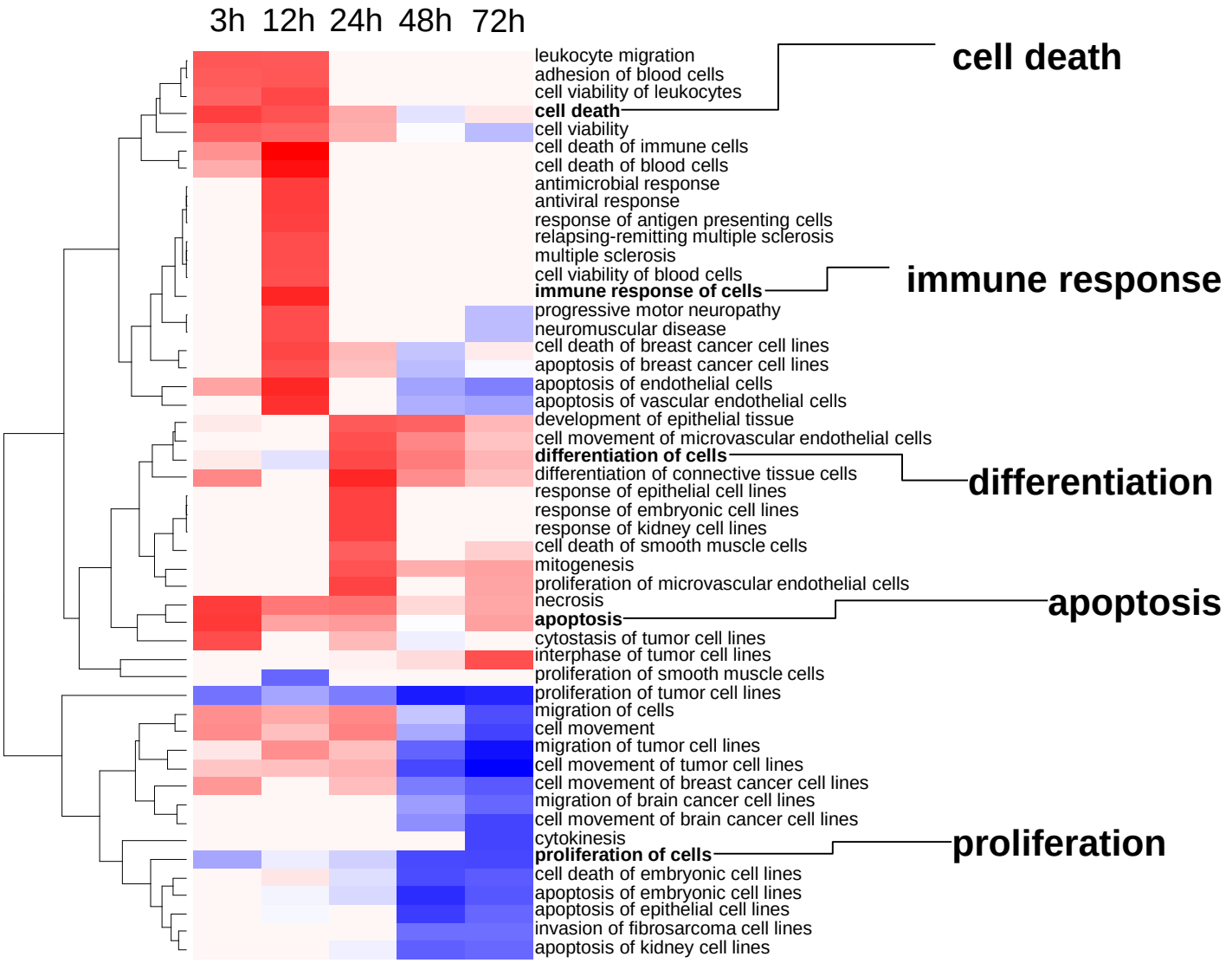
DEA miRNA:mRNA

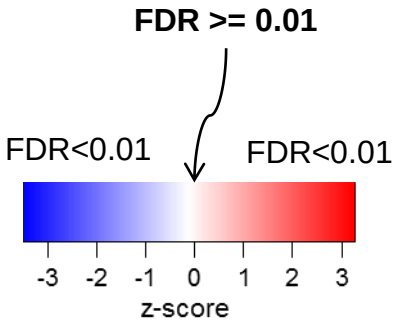
↓

Ingenuity® PA functional annotation

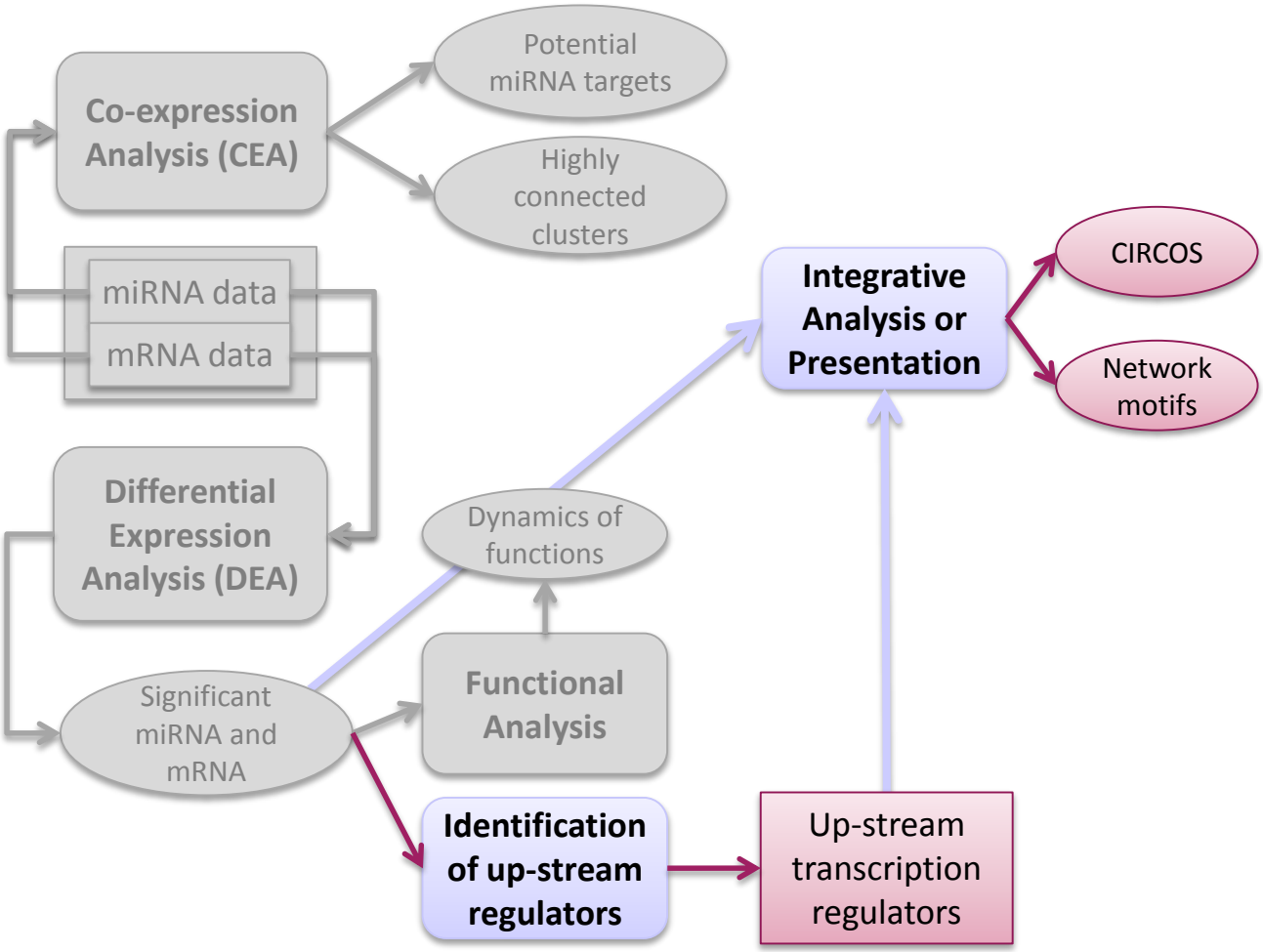
↓

Enriched functions with min FDR < 0.01, max |z|>2



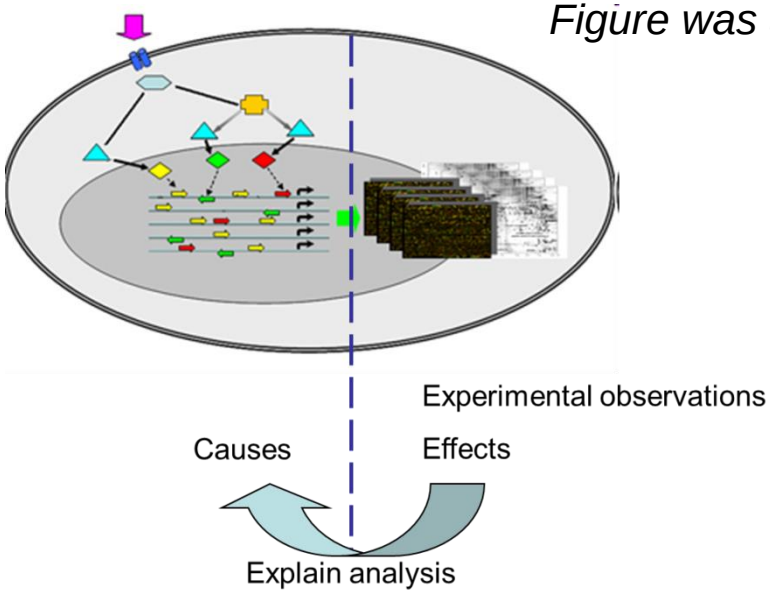


ANALYSIS PIPELINE



UPSTREAM REGULATORS

Figure was adapted from Explain Biobase



Significantly regulated
- miRNAs
- mRNAs

→ **Ingenuity® PA (IPA)** inference about active transcription regulators →

Transcription Regulators:

Overlap **p-value** (significance), and **z-score** (activation or inhibition)

Importantly, inferred upstream regulators include: **miRNAs, TFs, kinases**, chemical, etc. Therefore we included protein level into our analysis.

VISUALISATION OF INTEGRATED DATA

1. Significantly expressed genes are combined into functional categories

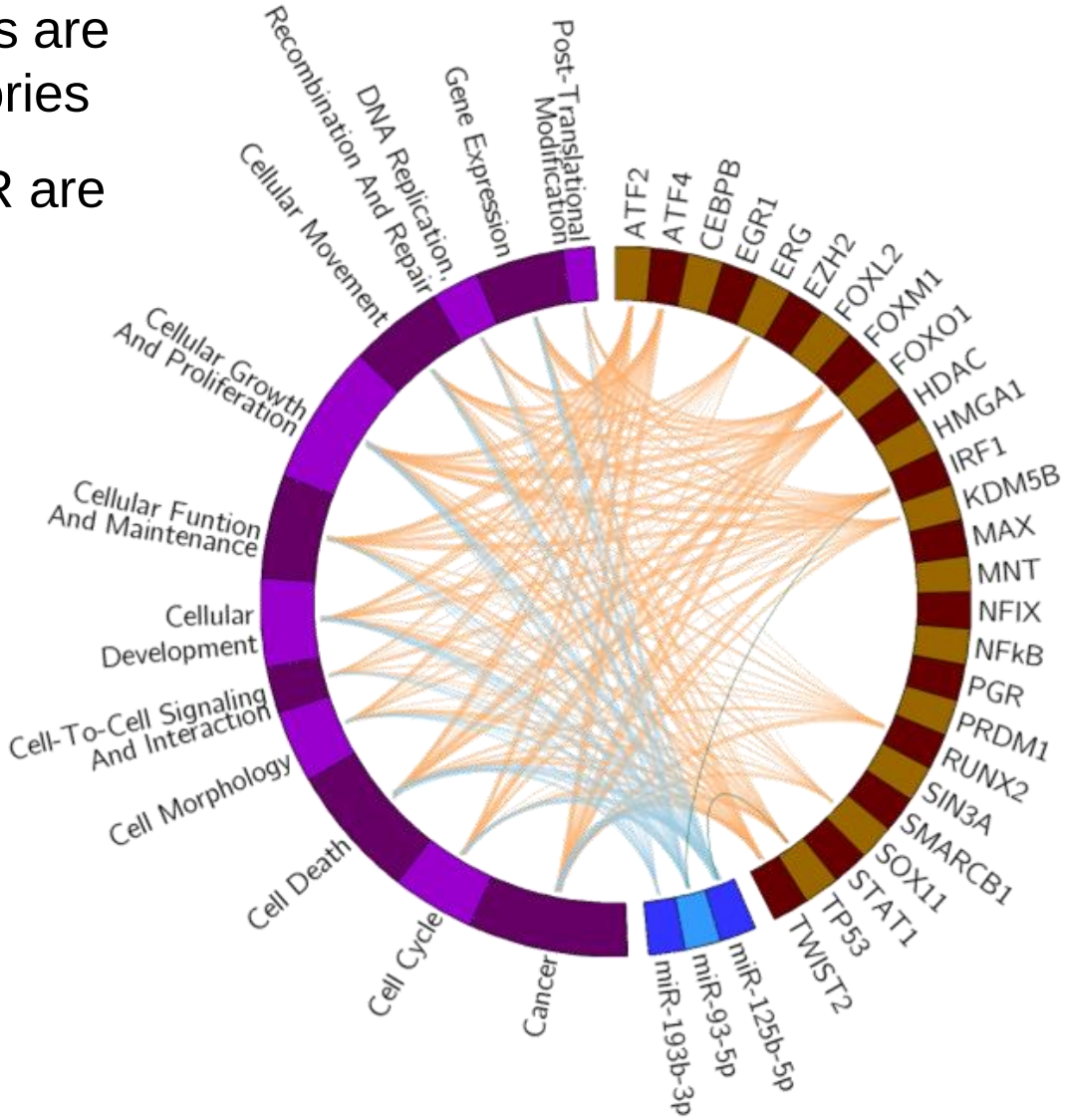
2. miRNA targeting activated TR are selected

Connections show:

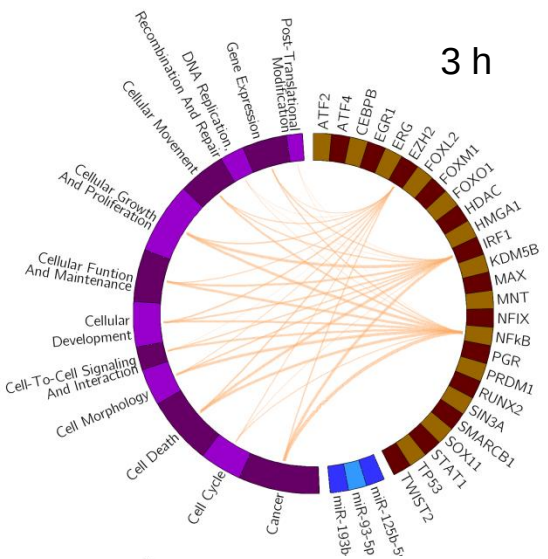
TR -> mRNA

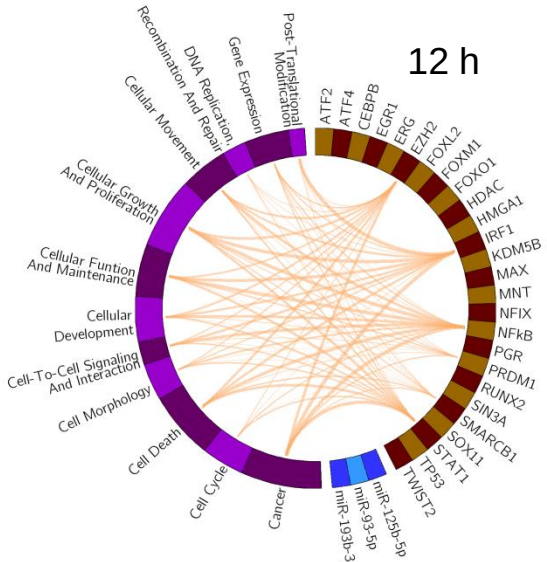
miR -| TR

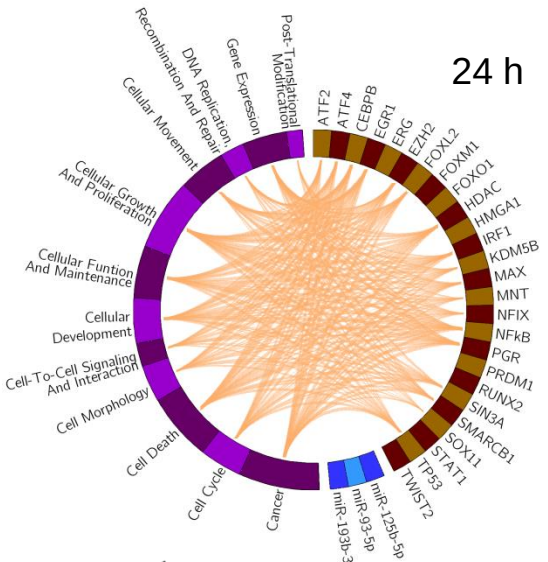
miR -| mRNA

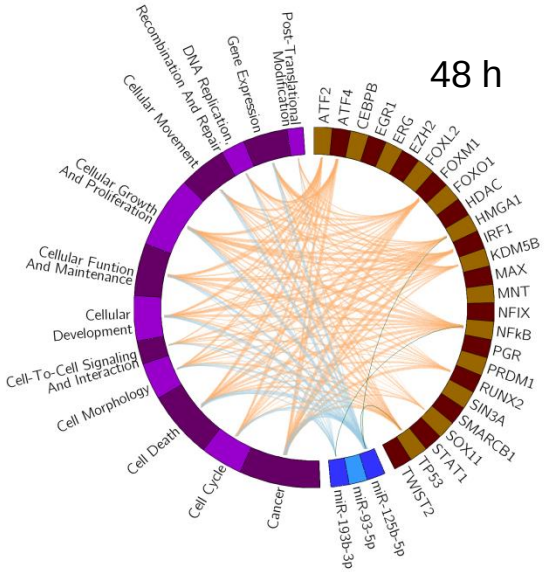


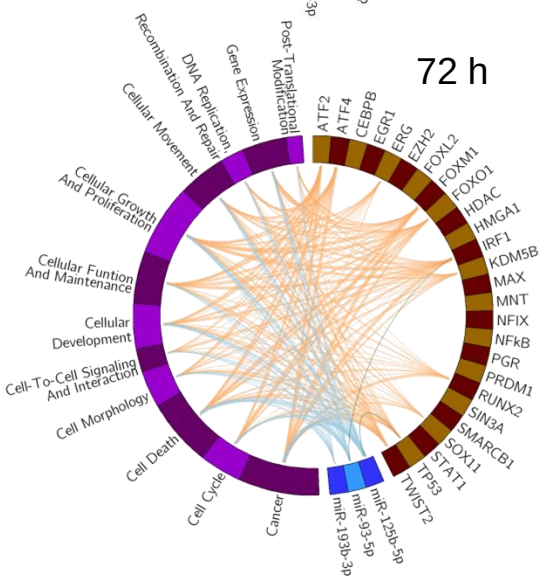
VISUALISATION OF INTEGRATED DATA

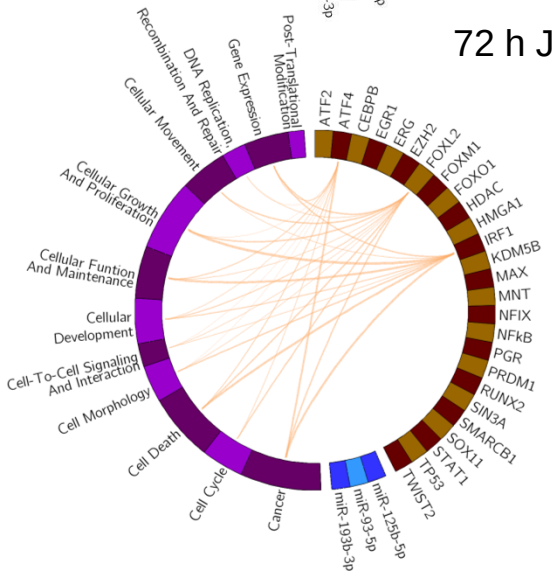






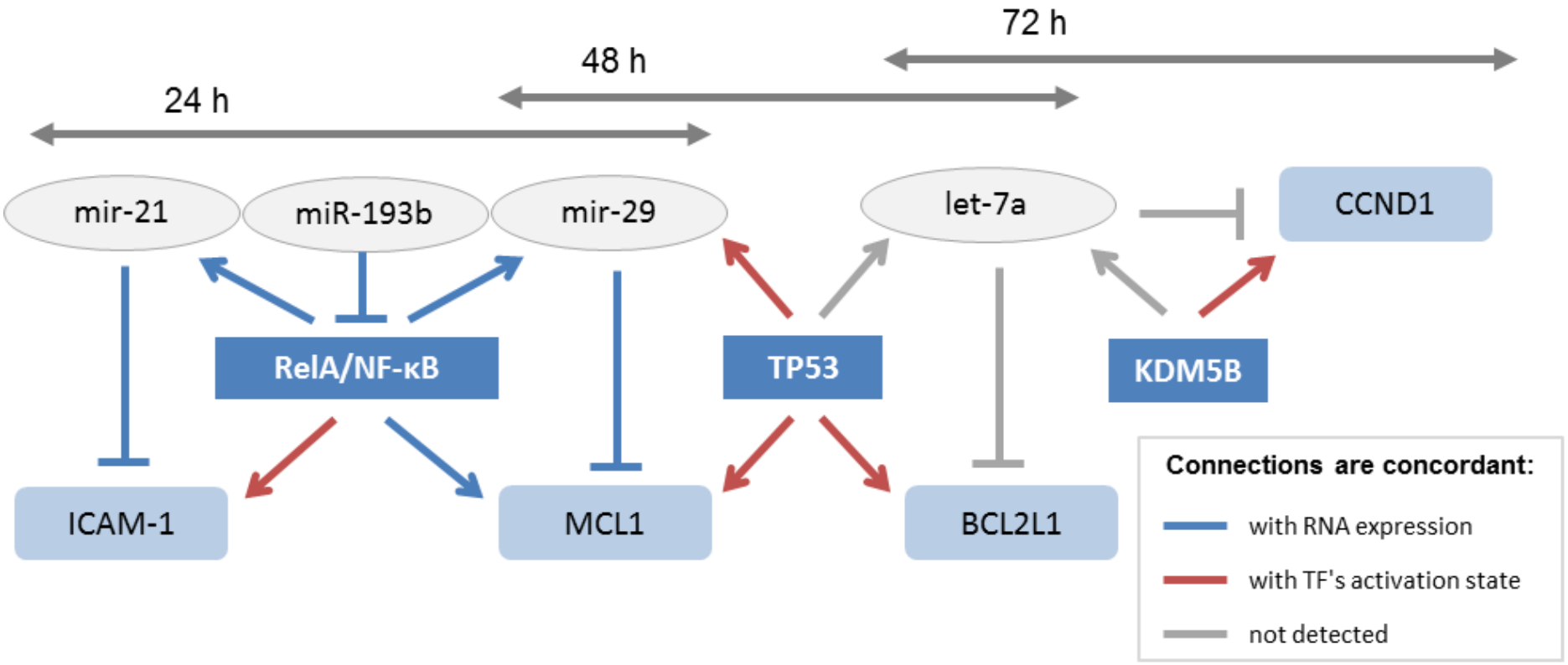






MOTIFS DETECTION

Networks, interconnecting: genes, miRNAs and transcription factors (inferred by Ingenuity), were manually curated. We were able identify several feed forward loops, presented below.



SUMMARY

Biology

- We observed a delayed response of miRNA with respect to mRNA response to IFNγ stimulation

  → miRNA are involved in fine tuning and compensation

- Promising regulation motifs were detected

Analysis

- Pipeline for simultaneous analysis of time-resolved miRNA and mRNA data was validated

- We showed benefits of LIMMA and filtering of not expressed features: lower FP rate and higher number of SDE genes.

ACKNOWLEDGEMENTS

CENTRE DE RECHERCHE PUBLIC SANTÉ

Fonds National de la Recherche Luxembourg

UNIVERSITÉ DU LUXEMBOURG

Arnaud Muller

Susanne Reinsbach

Nathalie Nicot

Demetra Philippidou

Laurent Vallar

Stephanie Kreis