

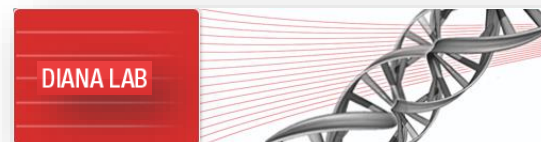
# DIANA Tools and Databases: In silico investigation of miRNA functions

[https://carolina.imis.athena-innovation.gr/diana\\_tools/web/index.php?r=site%2Ftools](https://carolina.imis.athena-innovation.gr/diana_tools/web/index.php?r=site%2Ftools)

Dimitris Grigoriadis, PhD candidate  
Filippos S Kardaras, PhD candidate  
Marios Miliotis, PhD candidate

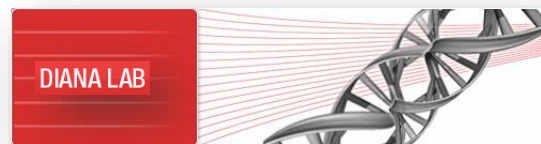
Prof. Artemis G Hatzigeorgiou

DIANA Lab, Department of Computer Science and Biomedical Informatics, University  
of Thessaly, 35131 Lamia, Greece

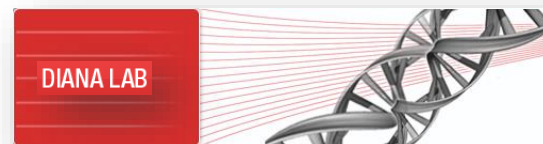


# Sections

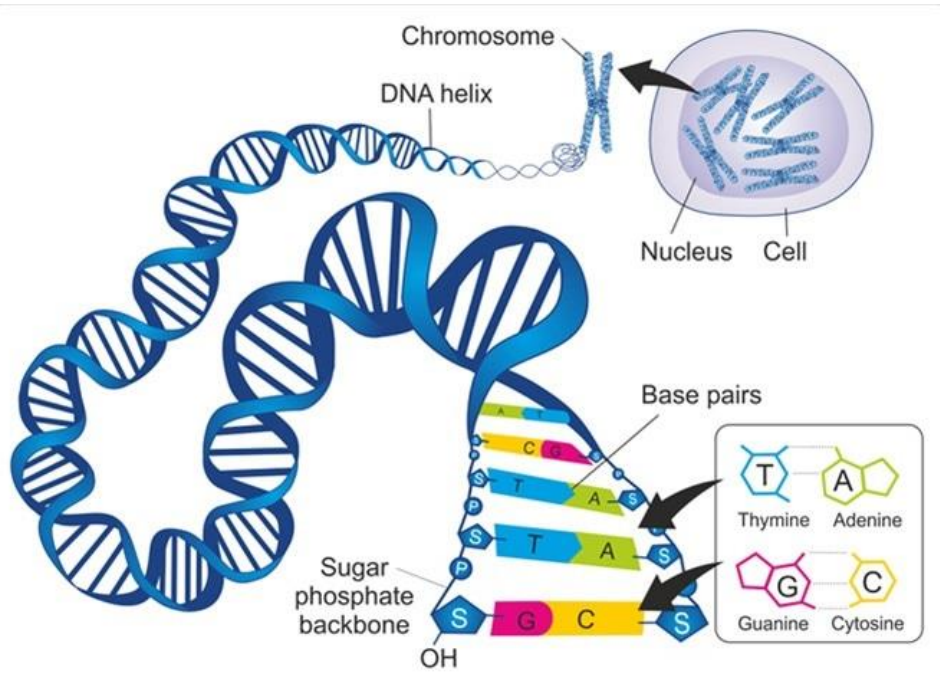
- Introduction to Molecular Biology and miRNAs
- Introduction to Next Generation Sequencing
- **DIANA-miRPath v3:** Target-based analysis of miRNA functions
- **DIANA-miRGen v4:** Indexing promoters for miRNAs using CAGE datasets
- **DIANA-mirExTra v2:** DEA suite for (s)RNA-Seq data revealing important miRNAs and TFs



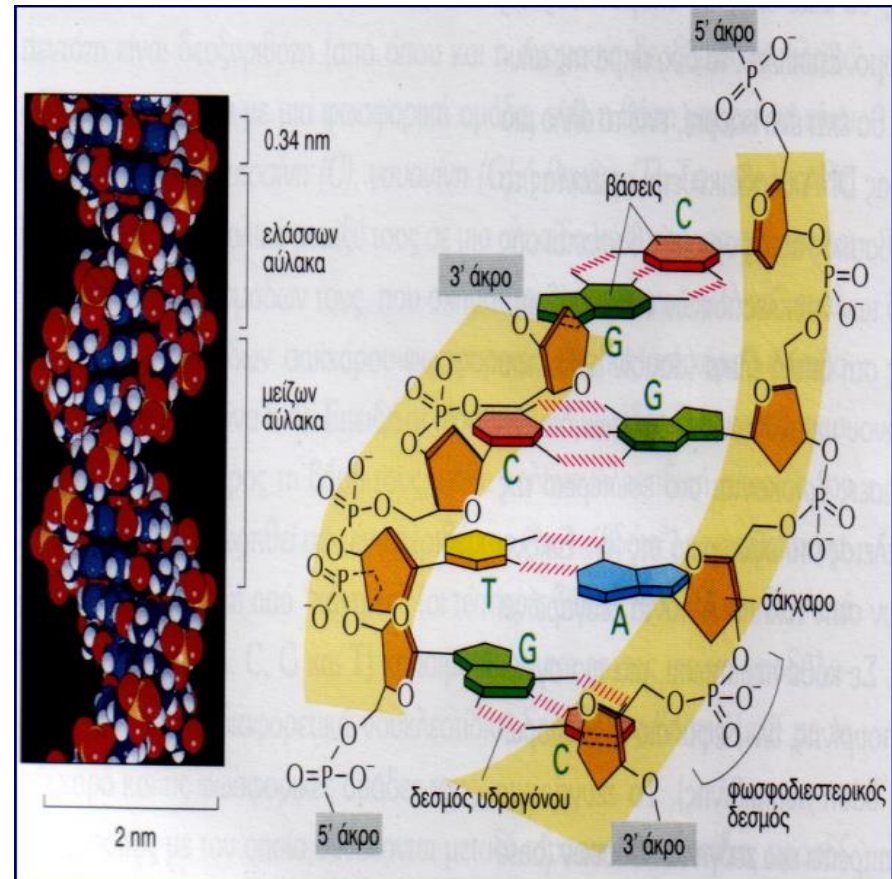
# Introduction to Molecular Biology and miRNAs



# Βιολογικά μακρομόρια: DNA, RNA, πρωτεΐνες



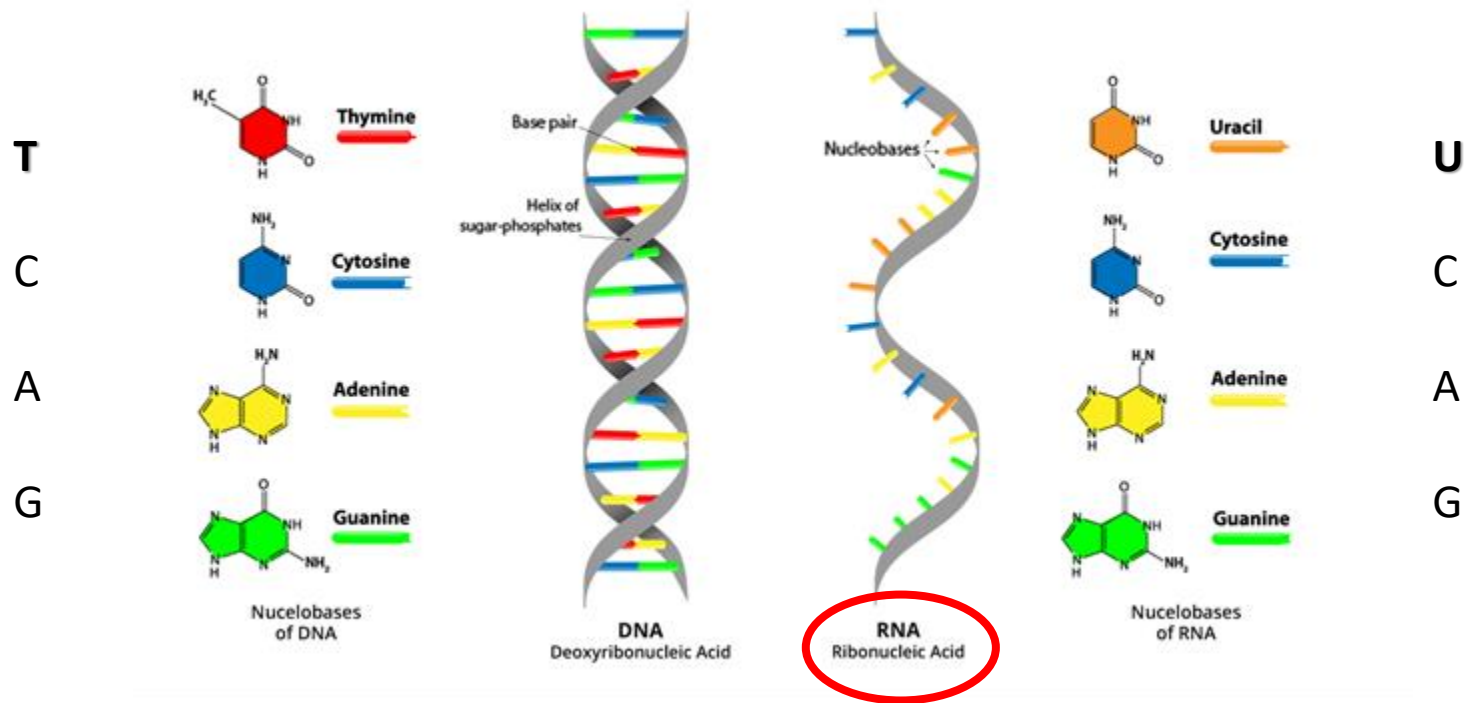
<https://www.news-medical.net/health/What-is-a-DNA-Diet.aspx>



DNA: Δίκλωνο μόριο που περιέχει τις πληροφορίες που καθορίζουν όλα τα χαρακτηριστικά ενός οργανισμού. Οι πληροφορίες είναι αποθηκευμένες στην αλληλουχία των βάσεων A,T,G,C. Οι δύο κλώνοι/αλυσίδες είναι συμπληρωματικές (ζεύγη A-T, G-C).

# Βιολογικά μακρομόρια: DNA, RNA, πρωτεΐνες ...

RNA: προϊόν **μεταγραφής** συγκεκριμένων περιοχών του DNA (γονίδια)

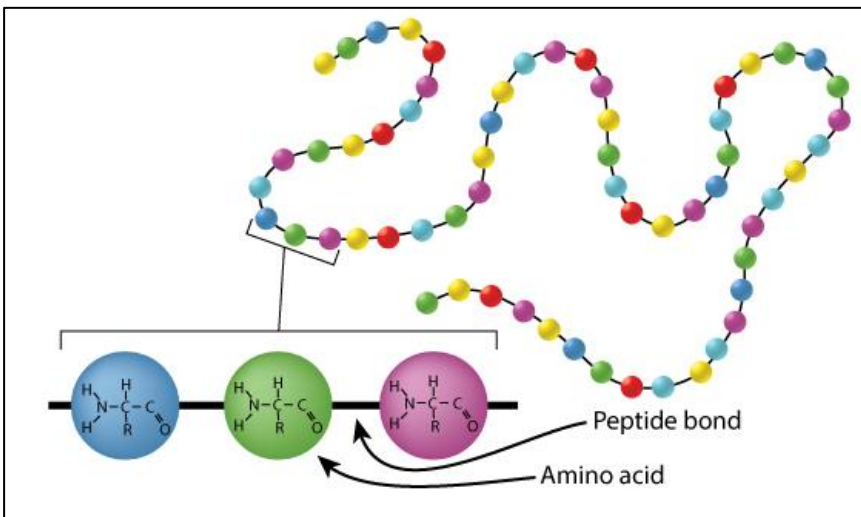


<https://www.99science.com/what-sugar-is-found-in-rna/>

# Τα είδη των RNA και ο ρόλος τους

- **mRNA** (κατά τη *μετάφραση*, το mRNA «διαβάζεται» από τα ριβοσώματα και παράγεται πρωτεΐνη)
- **rRNA** (συστατικό των ριβοσωμάτων)
- **tRNA** (συμμετέχει στη *μετάφραση* )
- **snRNA** (συμμετέχει στην ωρίμανση του mRNA)
- **miRNA** (συνδέεται με mRNA-στόχους, οδηγώντας σε αναστολή της μετάφρασης τους ή αποικοδόμηση τους, “*post-transcriptional regulation*”)
- .....

# Βιολογικά μακρομόρια: DNA, RNA, πρωτεΐνες ...

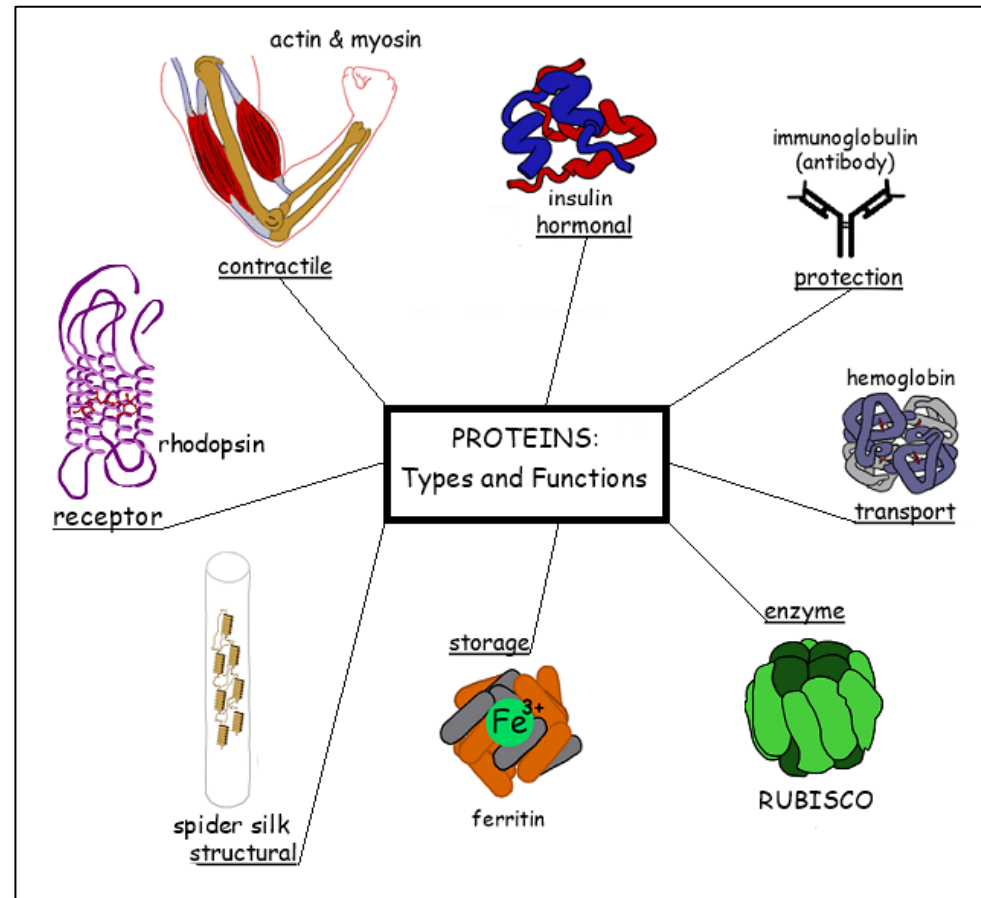


<https://content.byui.edu/file/a236934c-3c60-4fe9-90aa-d343b3e3a640/1/module3/readings/proteins.html>

Οι πρωτεΐνες είναι το προϊόν της **μετάφρασης** των mRNA και αποτελούνται από αμινοξέα.

Είδη πρωτεϊνών:

- Ενζυμα
- Δομικές
- Μεταφορικές
- Κινητήριες
- Αποταμιευτικές
- Σηματοδοτικές
- Υποδοχείς
- Ρυθμιστικές
- Ειδικές

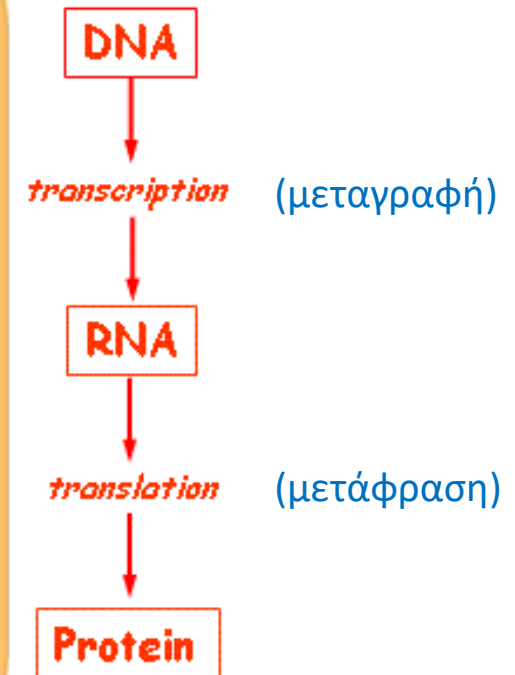
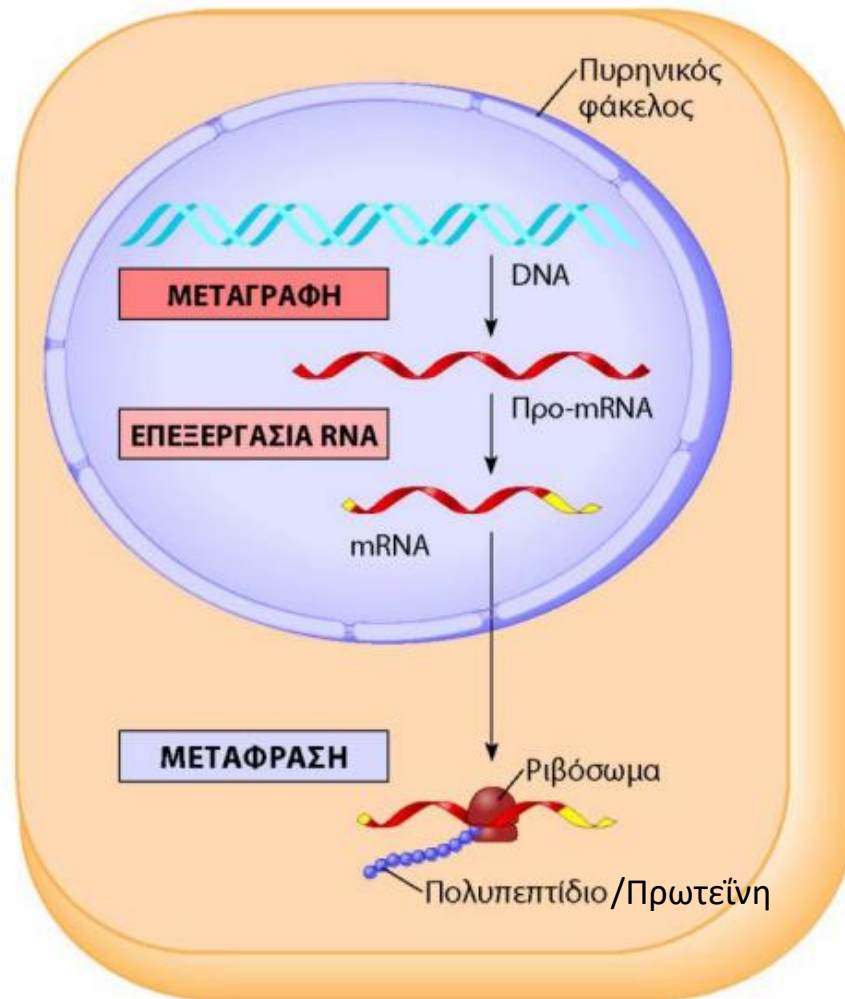


<https://alevelbiology.co.uk/notes/functions-of-proteins/>



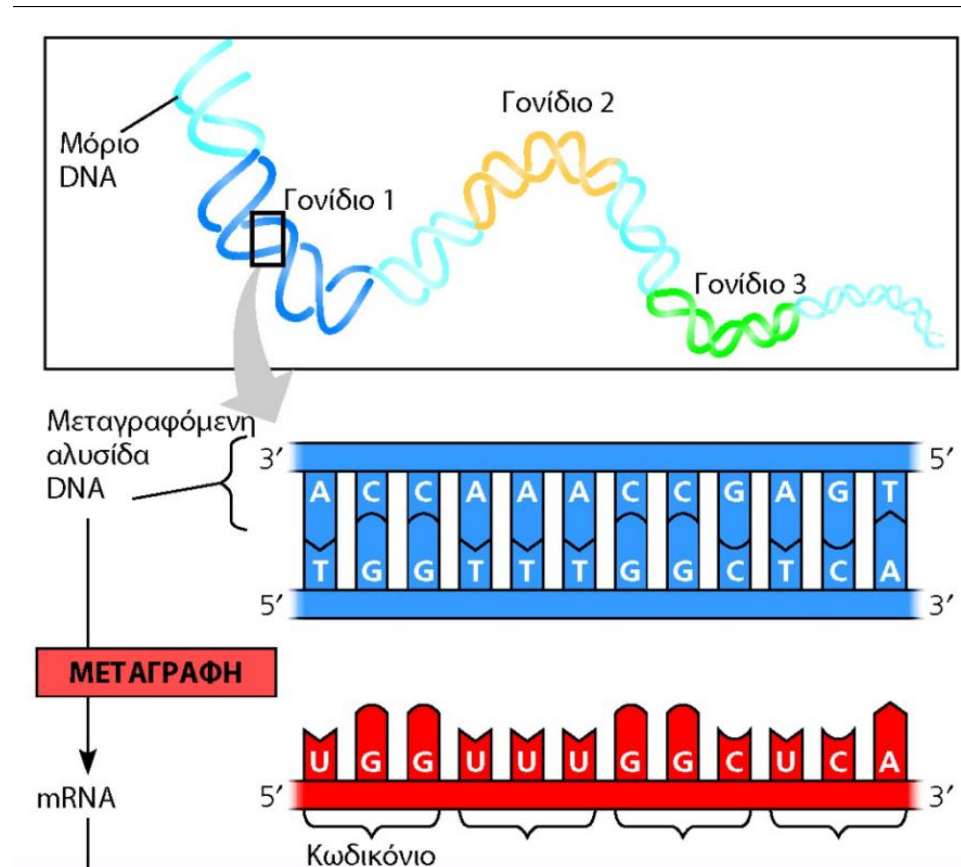
# Βιολογικά μακρομόρια: DNA, RNA, πρωτεΐνες ...

Ευκαρυωτικό κύτταρο



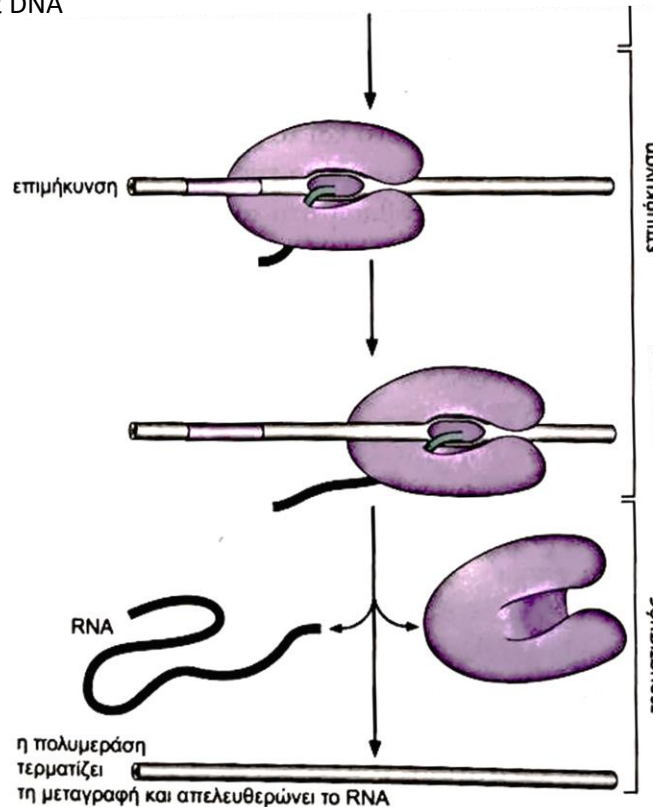
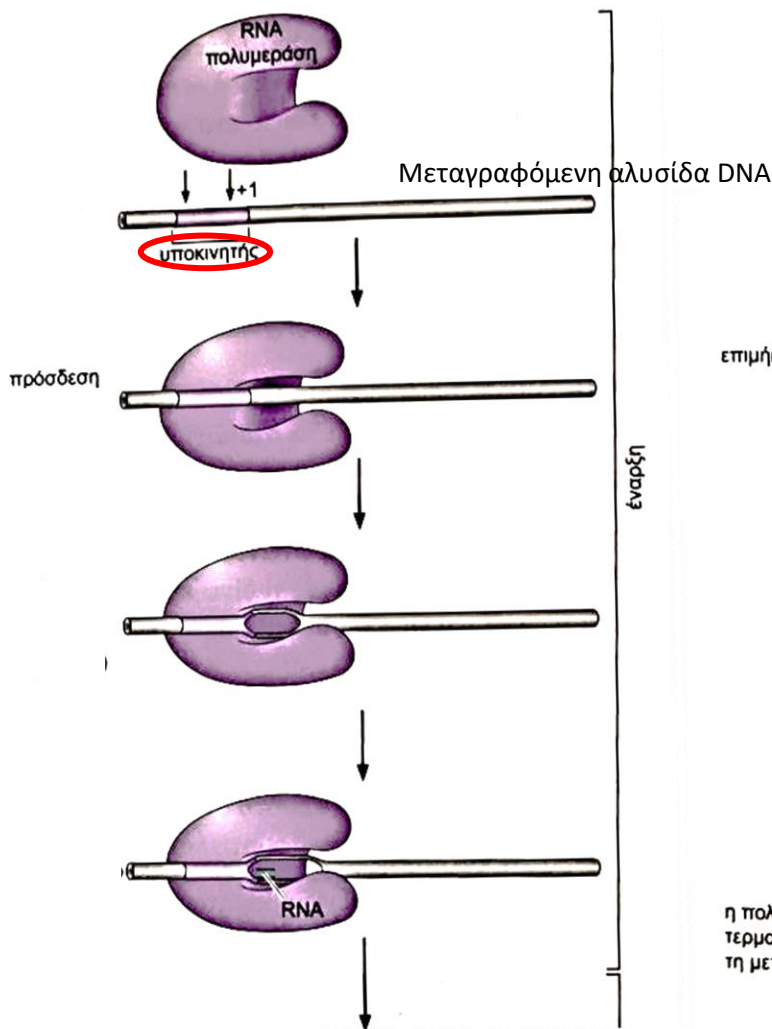


# Μεταγραφή



# Έναρξη, επιμήκυνση και τερματισμός της μεταγραφής

Η θέση του **υποκινητή** σηματοδοτεί τη θέση έναρξης της μεταγραφής (Transcription start site, **TSS**)

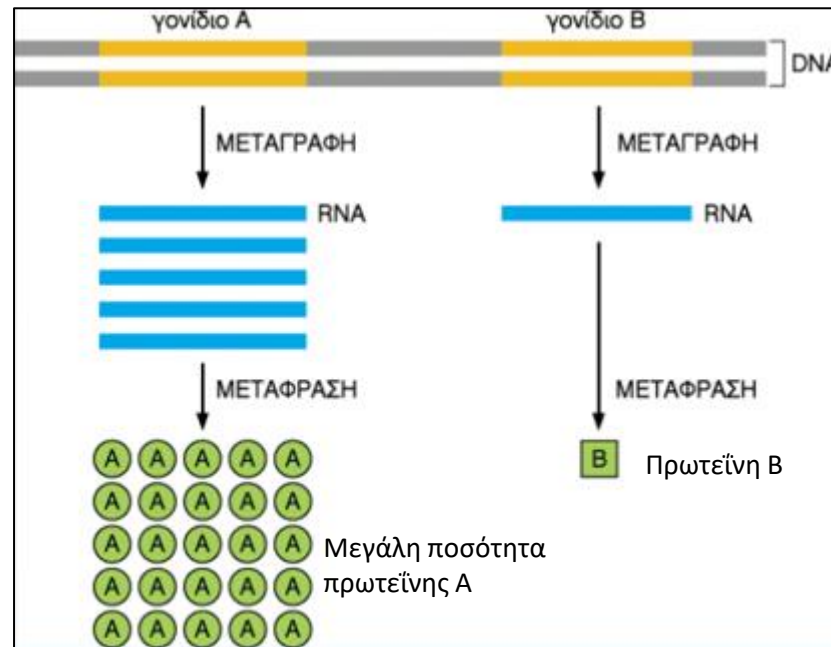


**ΕΙΚΟΝΑ 12-3** Οι φάσεις του κύκλου της μεταγραφής: Έναρξη, επιμήκυνση και τερματισμός. Η εικόνα παρουσιάζει το γενικό σχέδιο του κύκλου της μεταγραφής. Τα χαρακτηριστικά που παρουσιάζονται ισχύουν και στην περίπτωση των βακτηρίων και στην περίπτωση των ευκαρυωτών. Δεν παρουσιάζονται οι άλλοι παράγοντες που απαιτούνται για την έναρξη, την επιμήκυνση και τον τερματισμό.

Το νουκλεοτίδιο του DNA που κωδικοποιεί την αρχή της αλυσίδας του RNA καλείται θέση έναρξης της μεταγραφής και υποδεικνύεται ως θέση "+1".

# Ρύθμιση της γονιδιακής έκφρασης

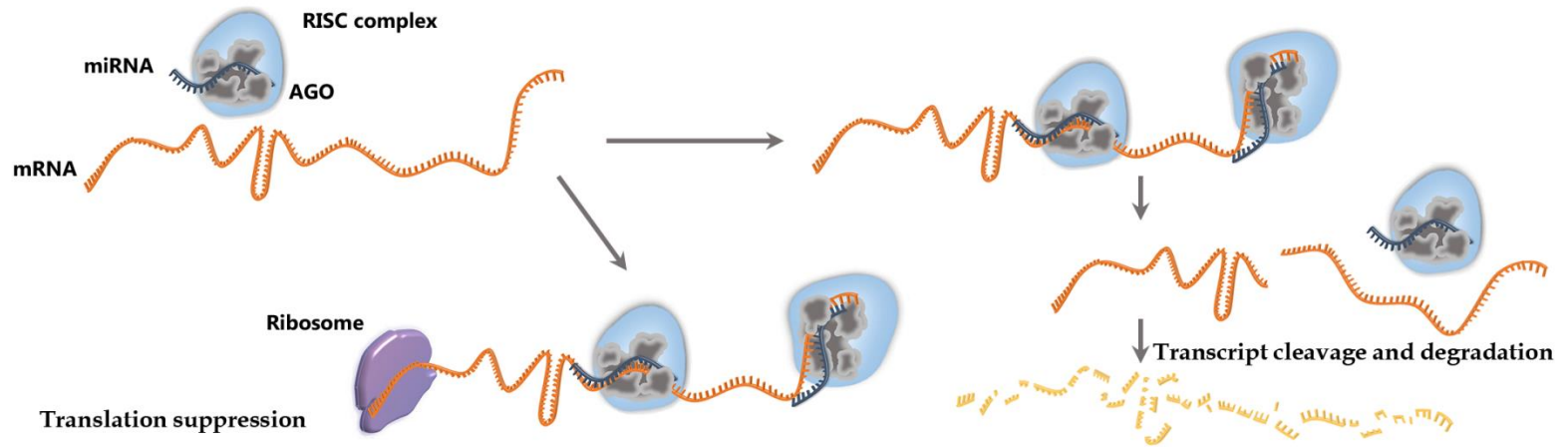
- **Γονιδιακή έκφραση:** μεταγραφή + μετάφραση των γονιδίων
- Όλα τα γονίδια δεν εκφράζονται στον ίδιο βαθμό



# Ρύθμιση της γονιδιακής έκφρασης

1. Στο επίπεδο της μεταγραφής (πχ. Η πρόσδεση μεταγραφικών παραγόντων (TFs) στον υποκινητή κάθε γονιδίου καθορίζει αν θα μεταγραφεί το γονίδιο ή όχι)
2. Στο επίπεδο μετά τη μεταγραφή (“post-transcriptional gene regulation”) (πχ. Δράση των miRNAs)
3. Στο επίπεδο της μετάφρασης
4. Στο επίπεδο μετά τη μετάφραση

# microRNAs

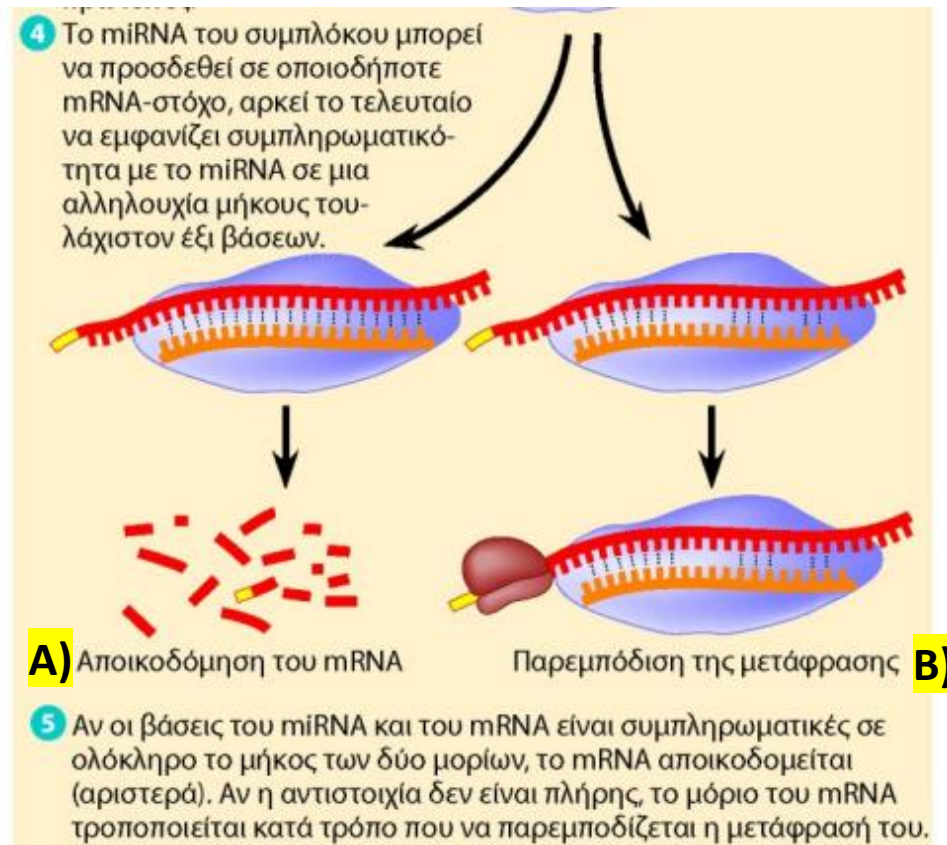


- transcribed in the nucleus
- maturation process in the cytoplasm
- small ncRNAs ~22nt
- central **post-transcriptional regulators**
- >50% human mRNAs are under miRNA regulation
- conserved in species

# Two roles of microRNAs

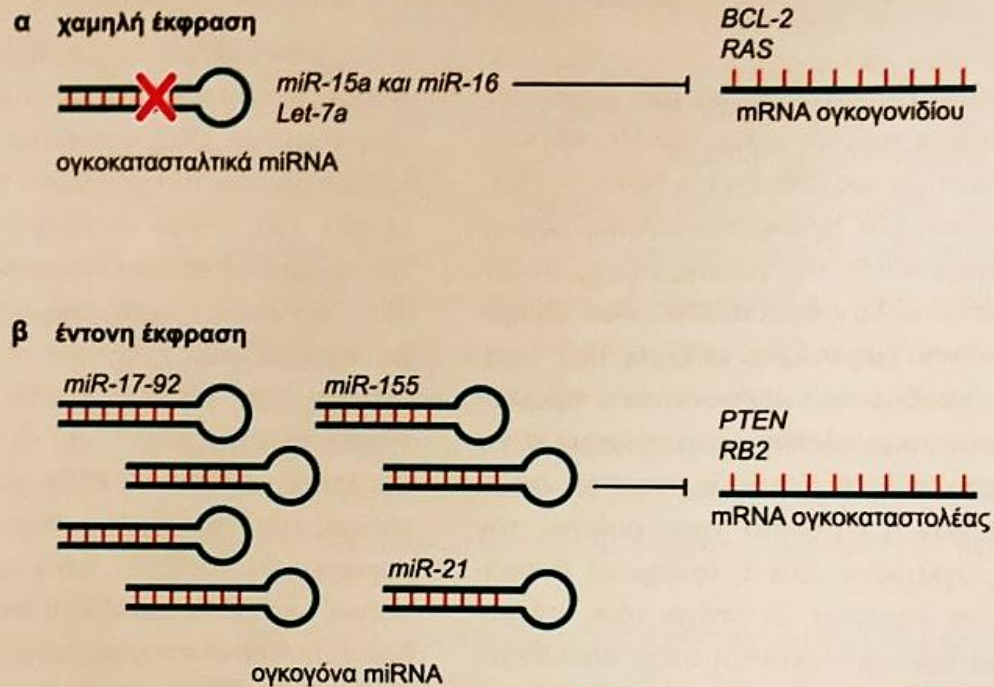
A) Destabilization and degradation of mRNA

B) Translational silencing





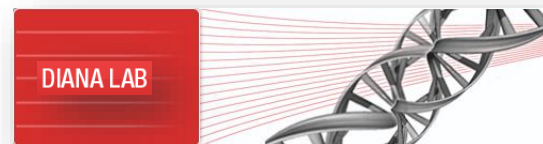
# Παράδειγμα: miRNAs και καρκίνος



**ΠΛΑΙΣΙΟ 18-3 ΕΙΚΟΝΑ 1** Τα μόρια miRNA ως ογκοκαταστολείς των ογκογονιδίων. (α) Σε αυτό το μοντέλο, ένα μόριο miRNA, το οποίο συνήθως μειώνει την έκφραση ενός ογκογονιδίου, μπορεί να δράσει ως ένα ογκοκατασταλτικό γονίδιο. Η απώλεια της δραστηριότητας του miRNA, λόγω μεταλλαγής ή απαλοιφής, για παράδειγμα, θα μπορούσε να καταλήξει σε μία μη φυσιολογική έκφραση του ογκογονιδίου-στόχου, το οποίο ακολούθως θα συνεισέφερε στο σχηματισμό του όγκου. (β) Εδώ, η ενίσχυση ή υπερέκφραση ενός miRNA, το οποίο ελαττώνει την έκφραση ενός ογκοκαταστολέα, ή άλλων σημαντικών γονιδίων που συμμετέχουν στη διαφοροποίηση, θα μπορούσε να συνεισφέρει στο σχηματισμό του όγκου, μέσω διέγερσης του πολλαπλασιασμού, της αγγειογένεσης και της εισβολής (Επανασχεδιασμένο με την άδεια των Garzon R. et al. 2006. *Trends Mol. Med.* 12: 580-587, Εικόνα 2. © Elsevier).



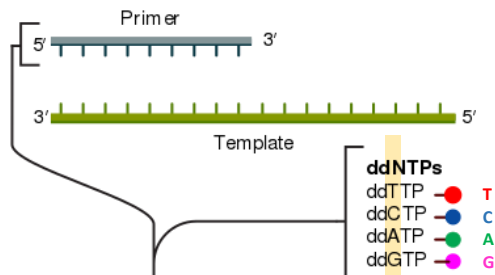
# Introduction to Next Generation Sequencing



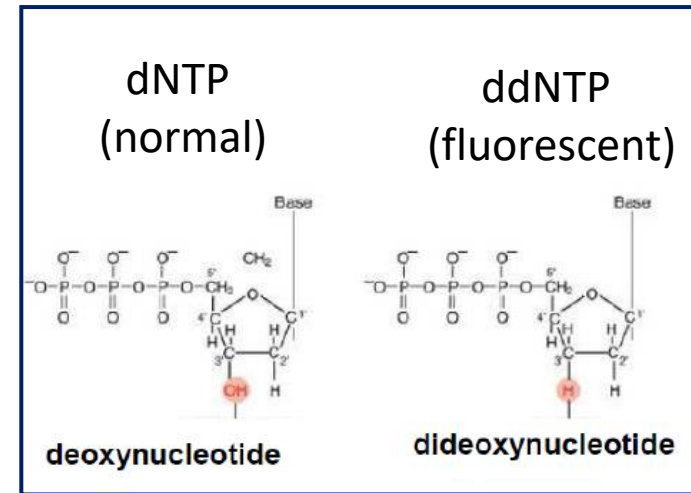
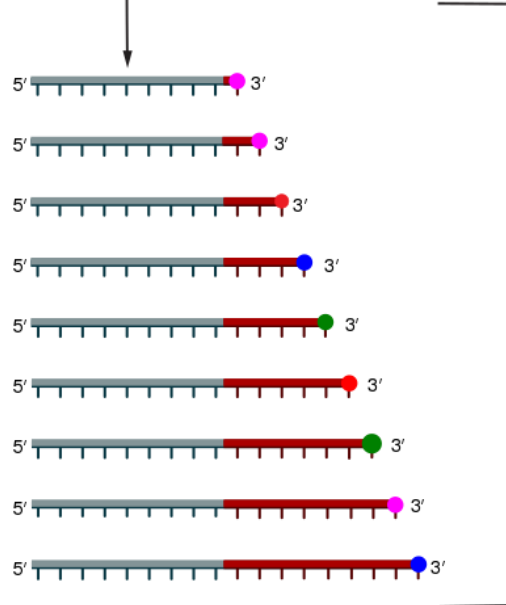
# Sanger sequencing: chain termination method

## ① Reaction mixture

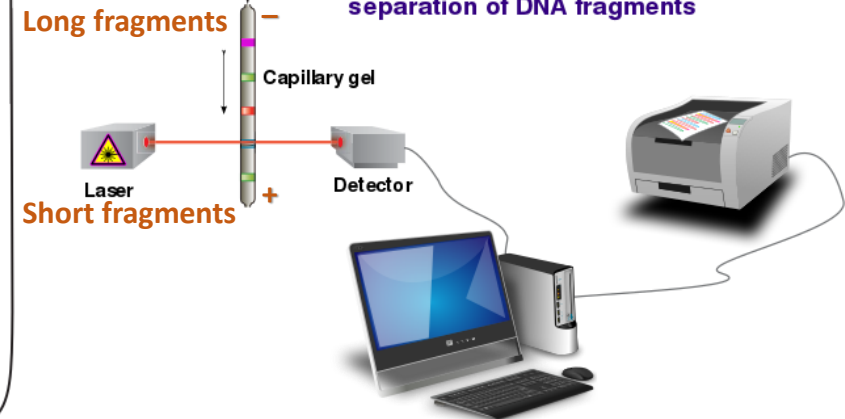
- ▶ Primer and DNA template
- ▶ DNA polymerase
- ▶ ddNTPs with flouochromes
- ▶ dNTPs (dATP, dCTP, dGTP, and dTTP)



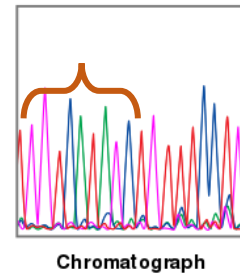
## ② Primer elongation and chain termination



## ③ Capillary gel electrophoresis separation of DNA fragments



## ④ Laser detection of flouochromes and computational sequence analysis



# Next-generation sequencing (NGS)

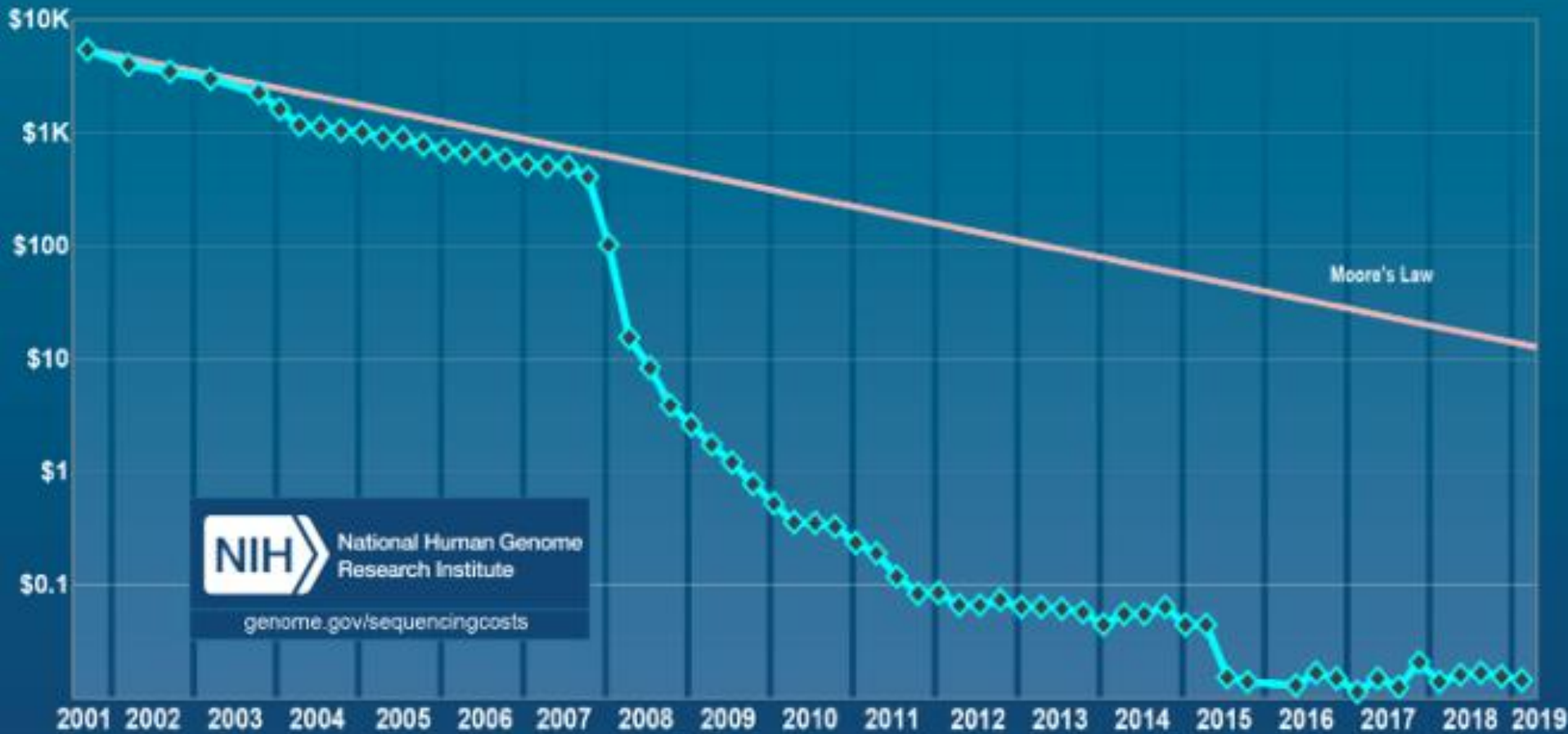
Using **NGS** an entire human genome can be sequenced within a single day.

In contrast, the previous Sanger sequencing technology, used to decipher the human genome, required over a decade to deliver the final draft.

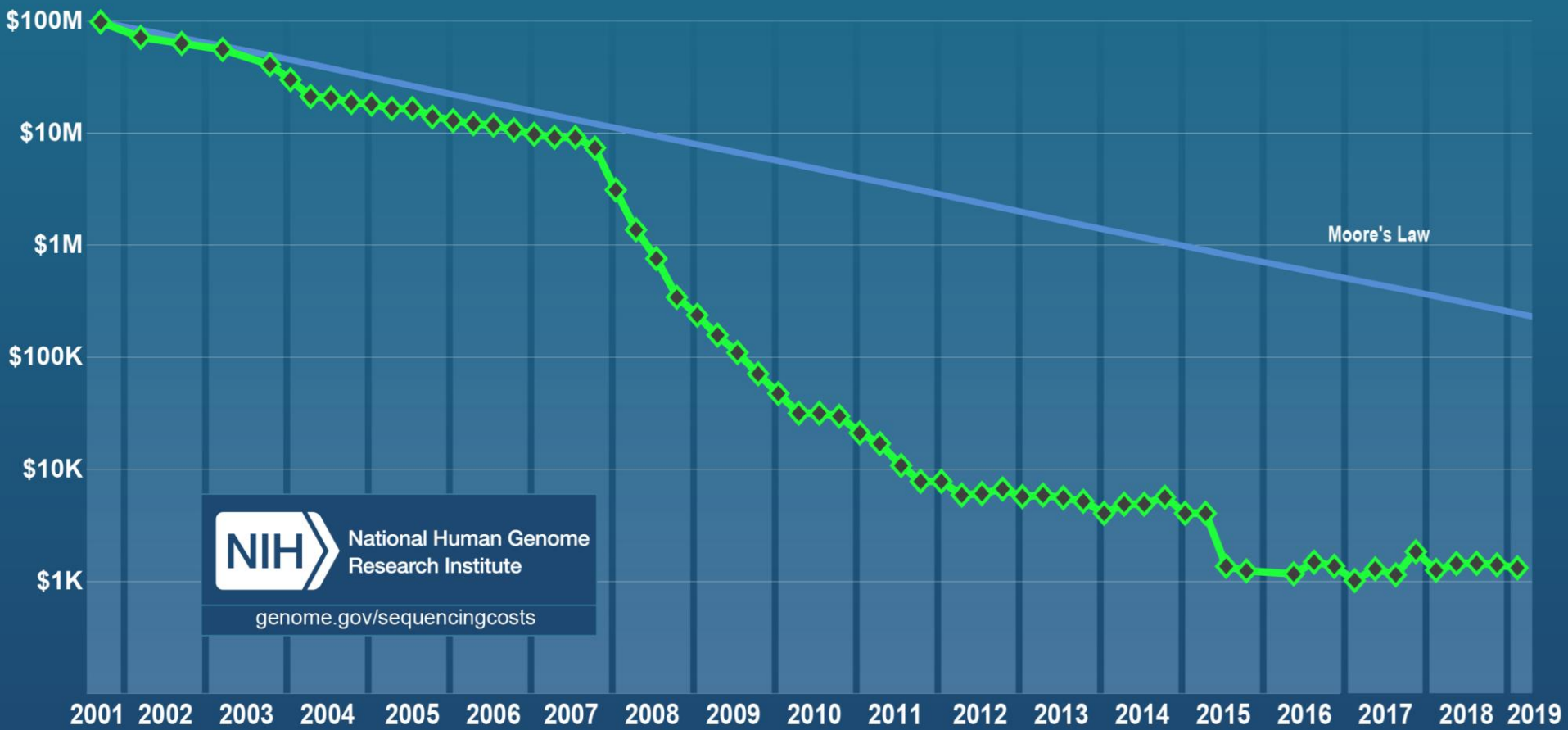
All NGS platforms perform sequencing of millions of small fragments of DNA in parallel. Bioinformatics analyses are used to piece together these fragments by mapping the individual reads to the human reference genome.

**Moore's law** in computing: doubling of computing power every two years. Here we see cost halving each two years.

## Cost per Raw Megabase of DNA Sequence



# Cost per Genome



# Next-gen sequencing technologies

- Six main technologies
- All massively parallel sequencing
  - Sequencing by synthesis
  - Sequencing by ligation
- Mostly produce short reads <400bp
- Read numbers vary from 1 million to ~1 billion per run

# Next-gen sequencing technologies



**Roche GS-FLX**



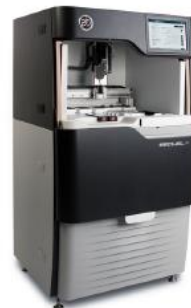
**Life Technologies SOLiD**



**Illumina HiSeq**



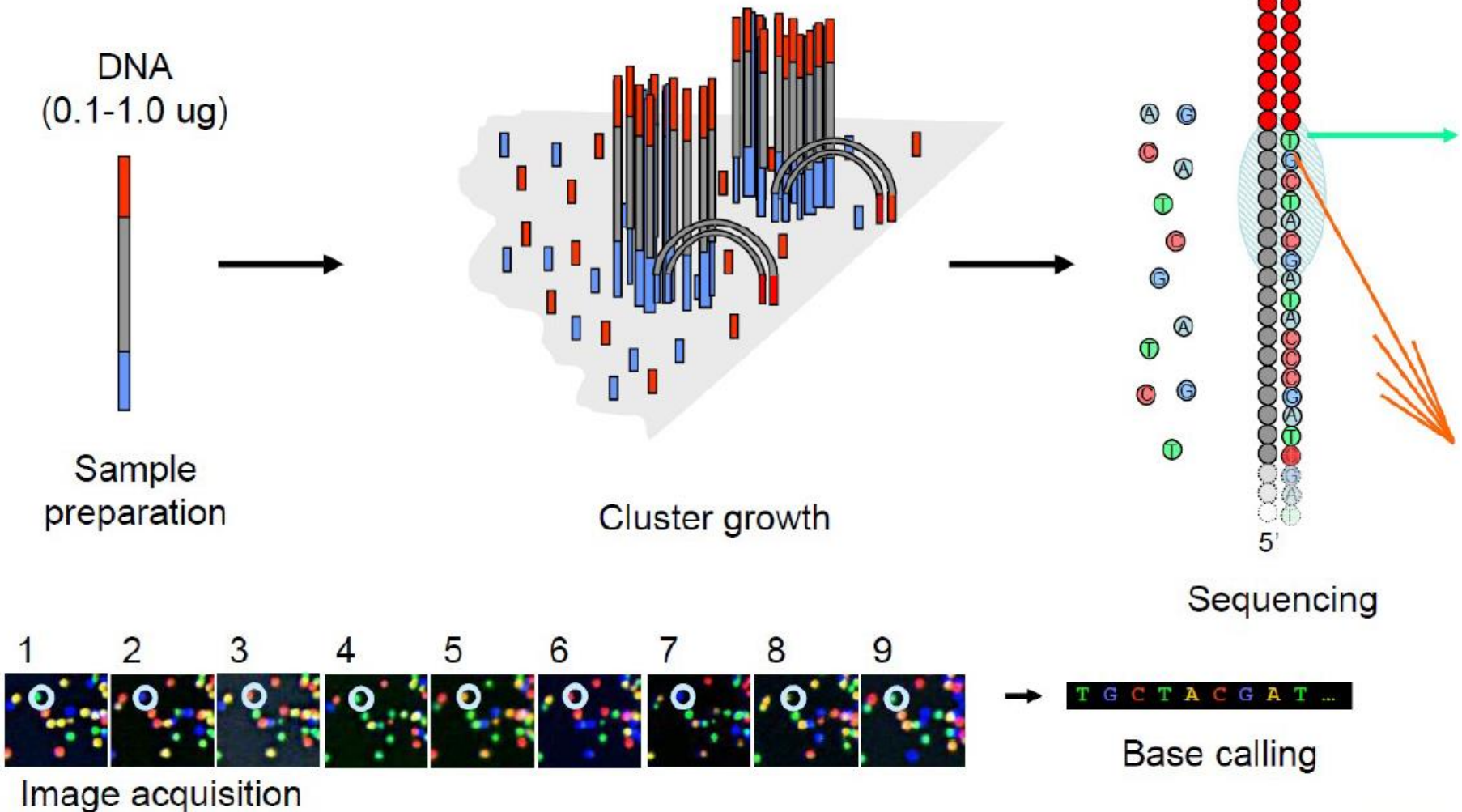
**Life Technologies Ion Torrent/Proton**



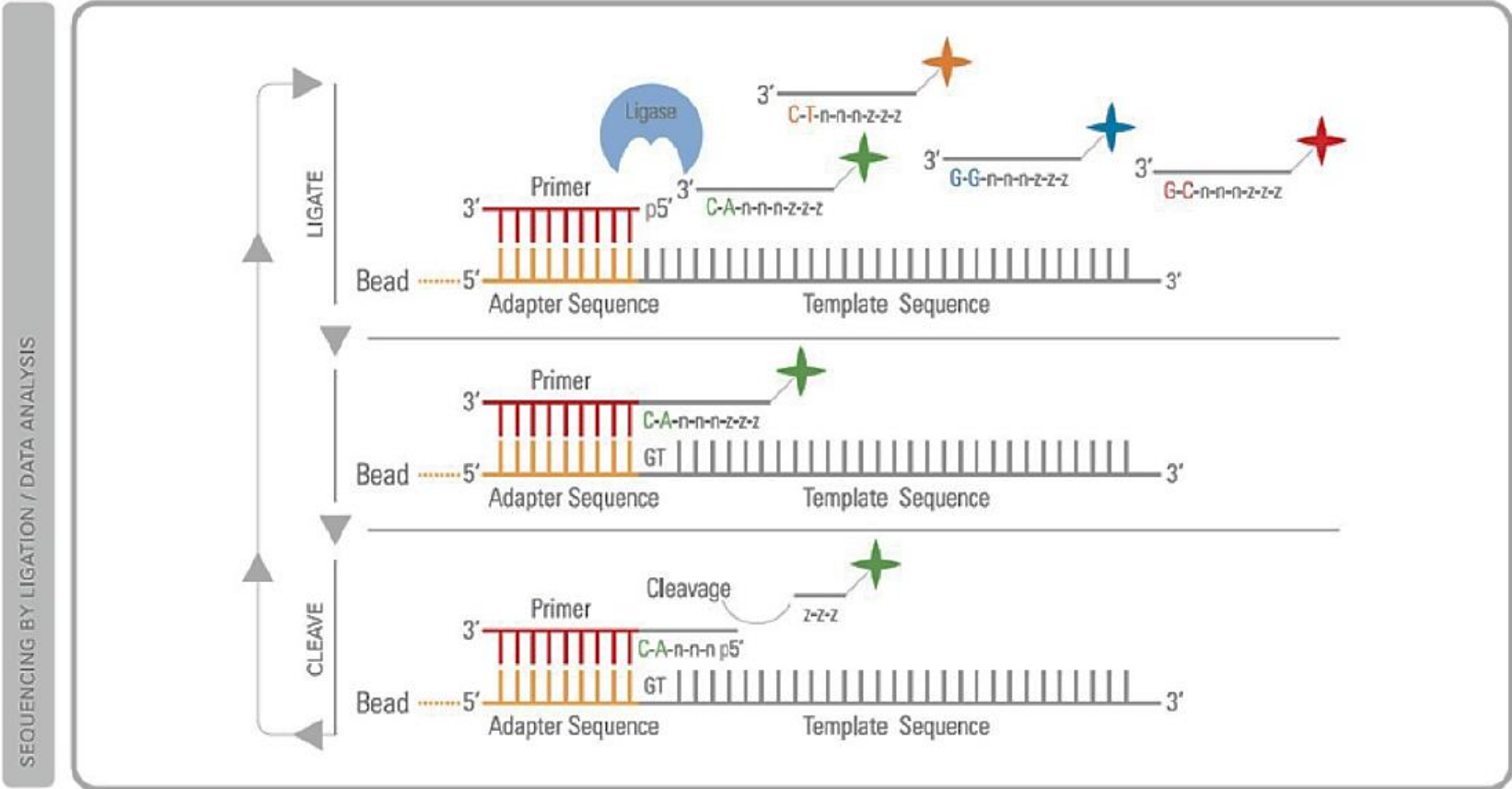


# Sequencing by synthesis

## Illumina Sequencing Technology



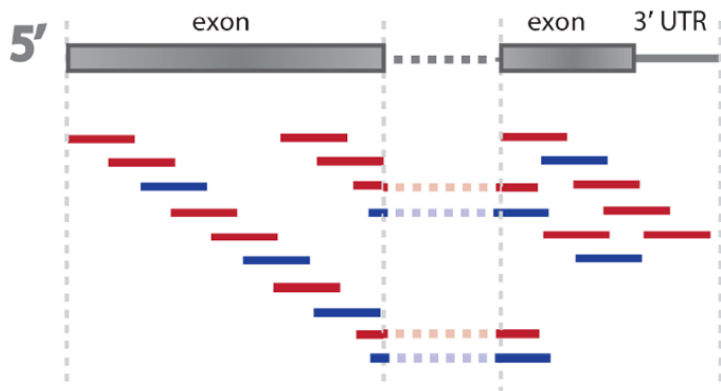
# Sequencing by ligation



# Sequencing options

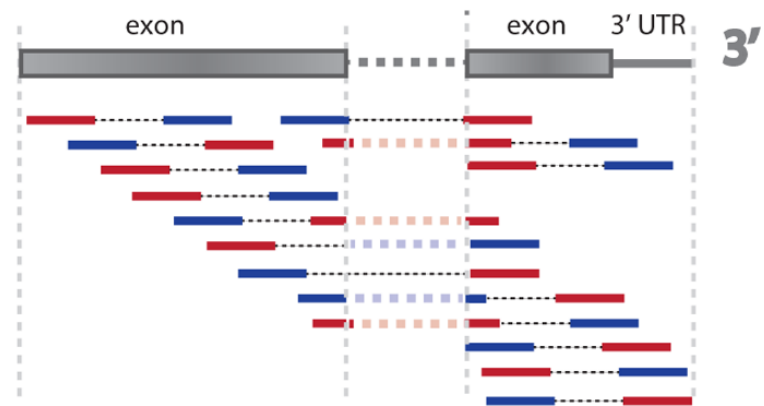
## Single-end

- Cheaper
- Suitable for more general purpose
- analyzes, e.g. Differential Expression



## Paired-end

- More information regarding the length
- and the position of a read
- Useful for spliced junctions, indels etc.



# Output: **fastq** file with sequences

## Fastq file example

```
read id 1 @ERR000589.41 EAS139_45:5:1:2:111/1
sequence 2 CTTTCCTCCCTGCTTTCCTGGCCCCACCATTTCAGGGAACATCTTGTCAT
3 +
4 3IIIIIIIIIIII>1IIIF9BG08E00I%IG+&?(4)%00646.C1#&(
read id 5 @ERR000589.42 EAS139_45:5:1:2:1293/1
sequence 6 AGTTGTTAAAATCCAAGCCAATTAAGATAGTCTTATCTTTTTAAAAGAAAT
7 +
8 IIIIIIGII.AIIII=?I9G-/II=+I=4?761BA2C9I+5A711+&>1$/I
```

### The information provided per read is:

Line 1: identifier

Line 2: **sequence**

Line 3: separator (always “+”)

Line 4: the quality score for each position of the read (same length as line 1)

Line 5

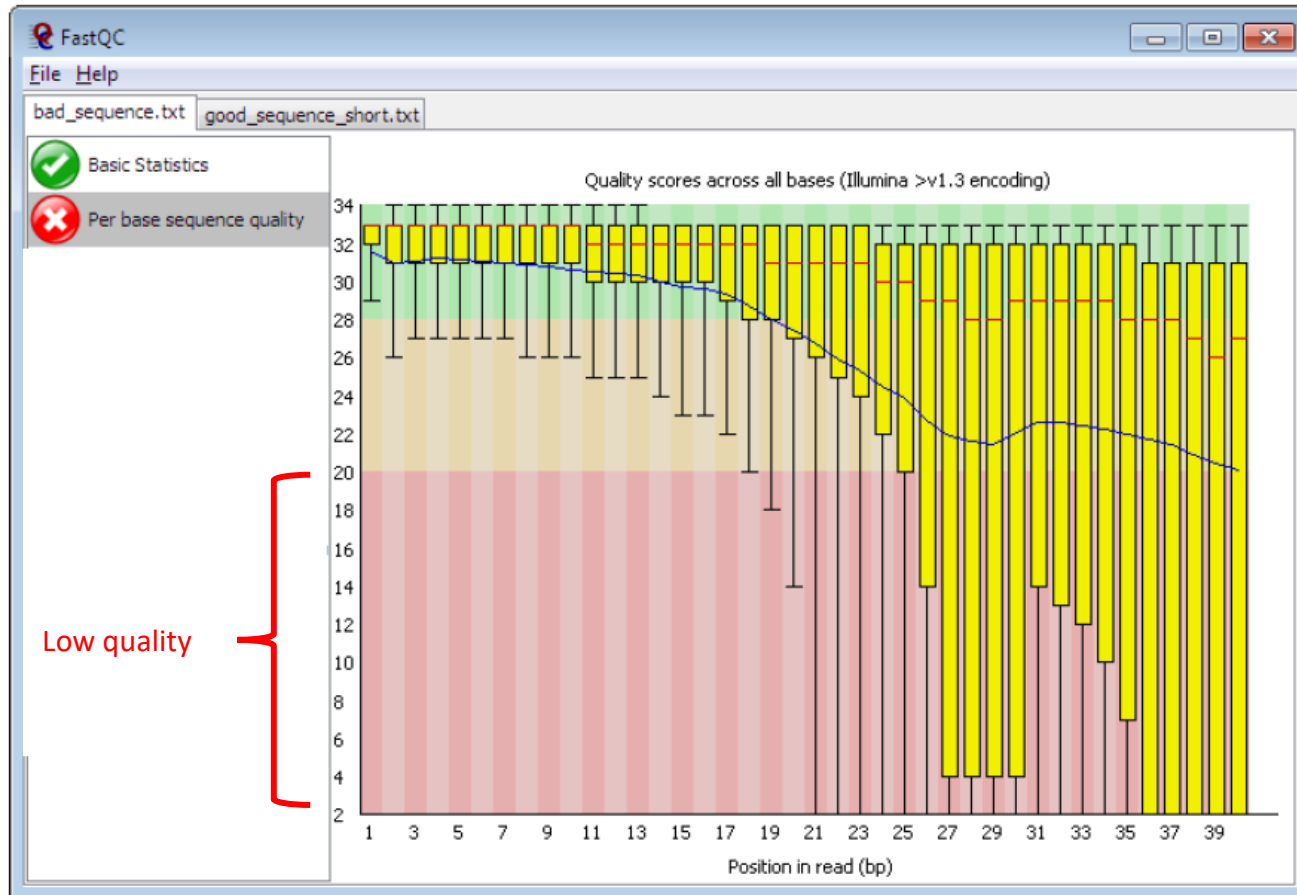
Line 6

Line 7

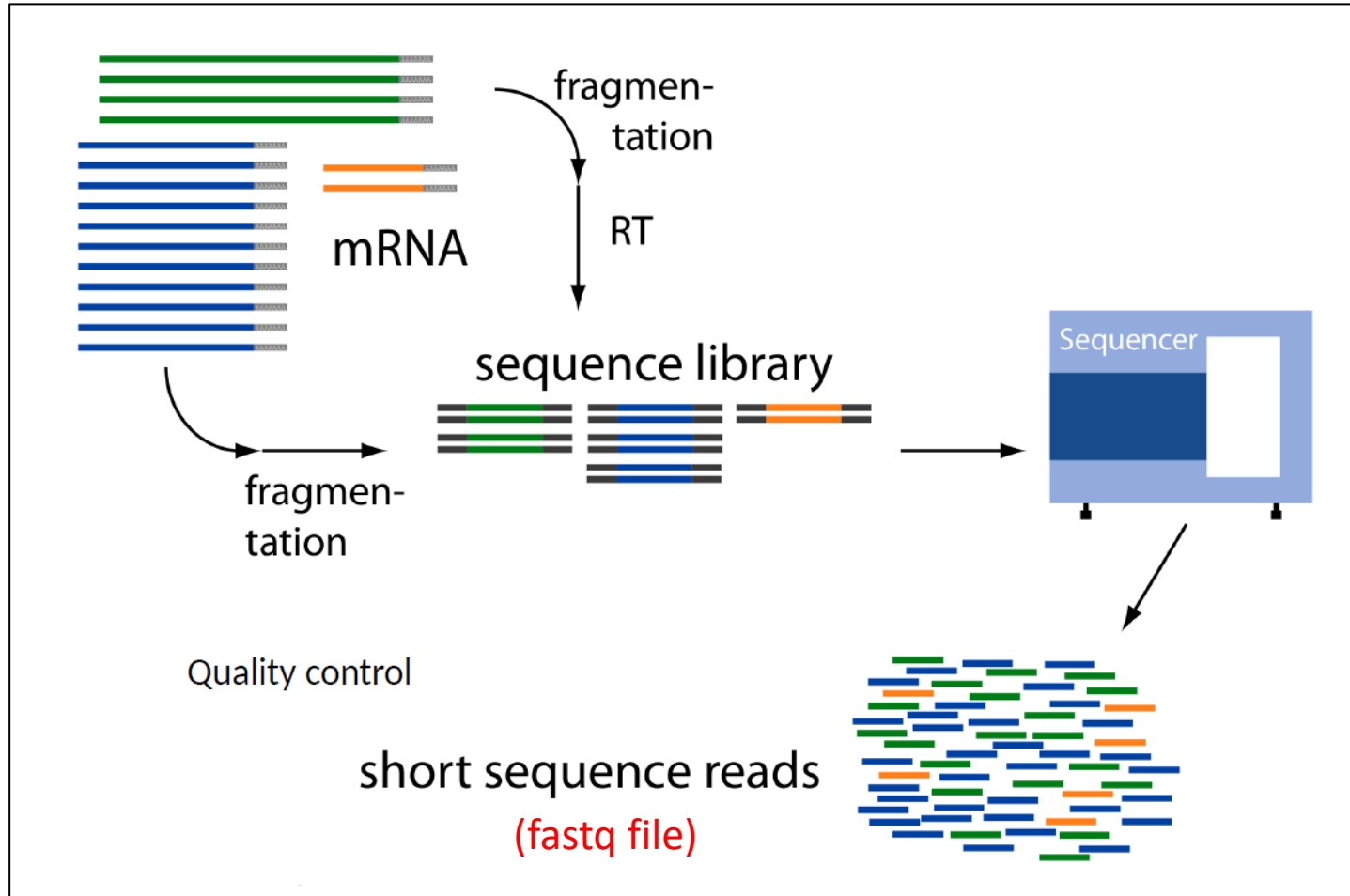
Line 8 .....

# Quality check

Trim or remove from fastq file reads with low quality, etc. ...



# Overview of RNA-Seq





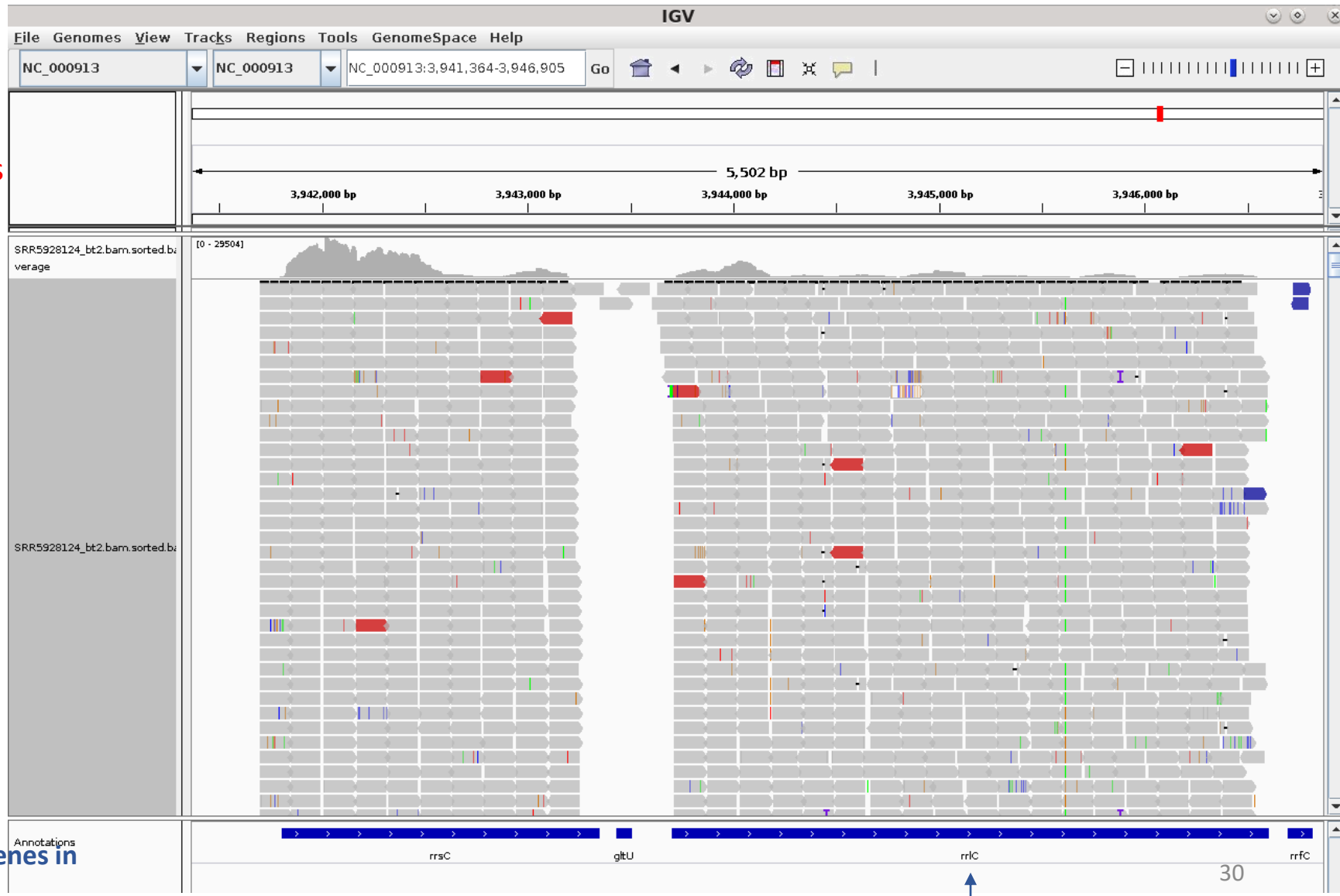


# Map/ align RNA-Seq reads to genome (IGV, Integrative Genomics Viewer)

Genome positions

reads

Known genes in genome



↑ gene

# DIANA miRPath v3

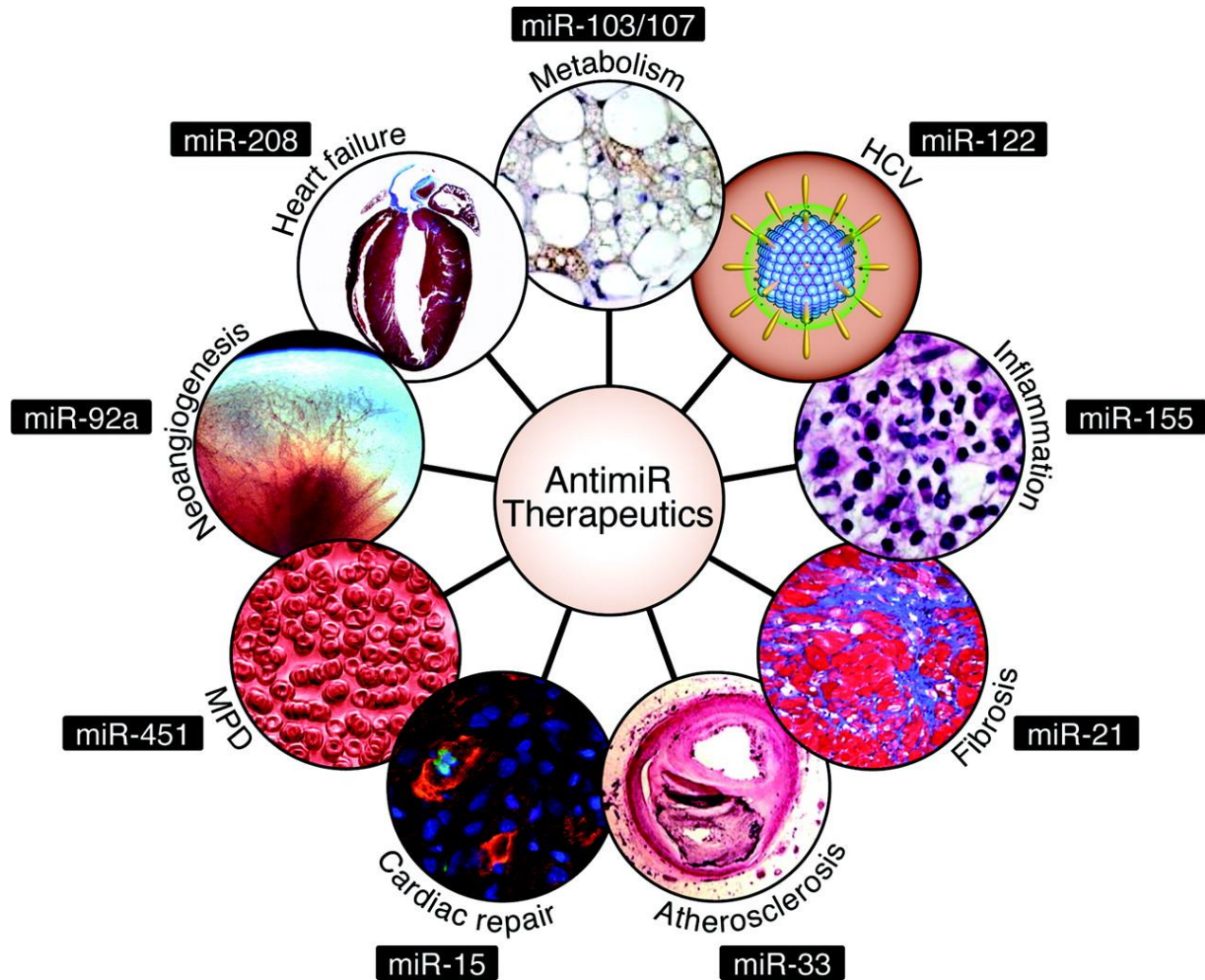
Target-based analysis of miRNA functions

Marios Miliotis, PhD Candidate

Prof. Artemis G Hatzigeorgiou



# Specific miRNAs that are currently being pursued as clinical candidates



*Eva van Rooij et al. Circ Res. 2012;110:496-507*

# miRPath purpose

What are the **functions/pathways** in which one or more **miRNAs** are **involved**?

- Need 3 things (at least) to answer:
  1. miRNA
  2. Target genes
  3. Biologic process/pathway of target genes

# Pathway/Gene ontology definition

- A **pathway** is

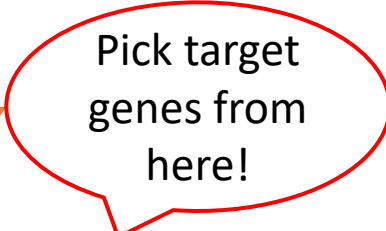
A series of interactions between molecules that lead to a product or a change in a cell

- A **gene ontology** is


a representation of a certain term in a cell. It has a name and can be:

- a cell compartment (e.g. nucleus, mitochondria)
- a molecular function (e.g. binding, catalysis)
- a biological process (e.g. cell cycle)


# Target gene resource



Pick target genes from here!

- 
- **DIANA-TarBase/LncBase:** databases with experimentally supported miRNA-gene pairs
  - **DIANA-microT-CDS:** a web-server of *in silico* predicted miRNA-gene interactions
  - **DIANA-miRPath:** Investigations of the combinatorial effects of miRNAs using semantic annotation sources

# Target gene resource



Or from here!

- **DIANA-TarBase/LncBase:** databases with experimentally supported miRNA-gene pairs
- **DIANA-microT-CDS:** a web-server of *in silico* predicted miRNA-gene interactions
- **DIANA-miRPath:** Investigations of the combinatorial effects of miRNAs using semantic annotation sources



# Inferring biological meaning from gene sets

Two widely utilized resource databases

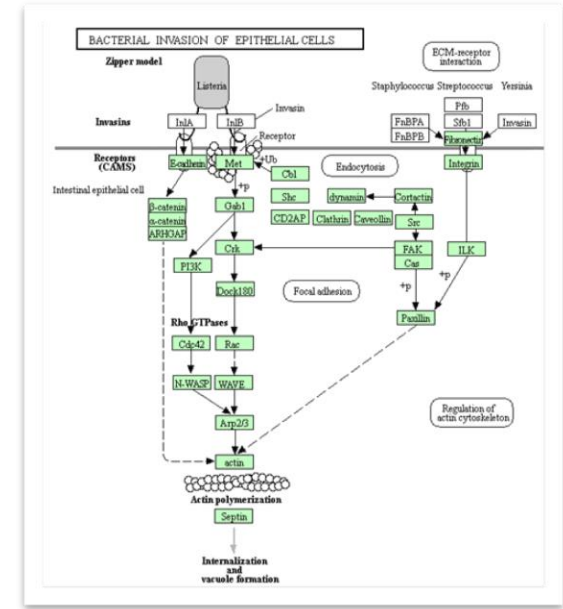
- **KEGG pathways** <http://www.kegg.jp/>

- ✓ Pathway maps
- ✓ Disease-related lists
- ✓ Biochemical reactions

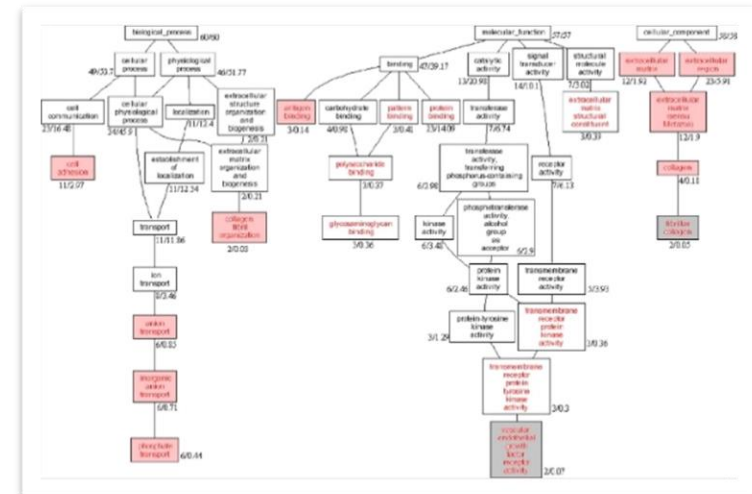
Pick biologic processes from here!

- **Gene Ontology** <http://geneontology.org/>

- ✓ **Molecular Function**  
*one step*
- ✓ **Biological Process**  
*a series-of-steps*
- ✓ **Cellular Component**  
*the space*



Kanehisa M, et al. *Nucleic Acids Res.* (2017)



Ashburner M, et al. *Nat. Genetics.* (2000)

# Analyze using miRPath v3

Find  
miRNA-biologic process  
associations!

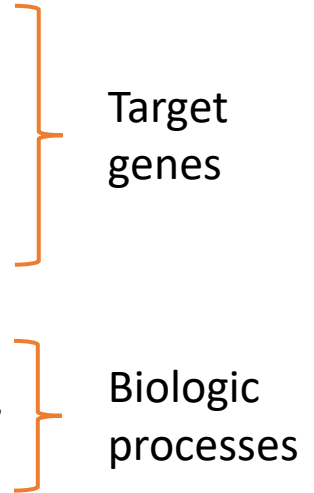
- **DIANA-TarBase/LncBase:** databases with experimentally supported miRNA-gene pairs
- **DIANA-microT-CDS:** a web-server of *in silico* predicted miRNA-gene interactions
- **DIANA-miRPath:** Investigations of the combinatorial effects of miRNAs using semantic annotation sources

# Inferring biological meaning from gene sets

Terms interrogated as lists of components

- A list under study
  - ✓ Differentially expressed genes (RNA-Seq/microarrays)
  - ✓ DNA targeted by transcription factors (ChIP-Seq)
  - ✓ *Genes regulated by microRNAs (CLIP-Seq, prediction algorithms etc.)*
- The resource
  - ✓ One set of interacting molecules per term (e.g. all genes in a KEGG pathway, GO-term or Disease)
- The question (translated in Statistics)
 

Does the list under study contain more components for a specific term than would be expected by chance?
- Term enrichment analyses apply Fisher's Exact Test
  - ✓  $p_{(H0)}$ : *Term-related components in list under study are no-more than would be expected by chance*



# DIANA-miRPath v3.0

## Web-server details

mirPath v3.0 enables:

- ✓ the functional analysis of a selected miRNA
- ✓ exploring the combined effect of multiple miRNAs
- miRNA target resources:
  - ✓ TarBase v7.0 (experimentally supported targets)
  - ✓ microT-CDS & TargetScan (*in silico* predicted targets)
- Available modules:
  - ✓ KEGG pathways
  - ✓ Gene Ontologies
- Reverse search:
  - ✓ Query Pathways/GO-terms to find controlling miRNAs
- Flexible statistics engine extends functionality

# DIANA-miRPath v3.0

Web-server details

4 available modules

## Genes Union

- ✓ targets of all miRNAs under query

*Which pathways are enriched by the sum of targeted genes?*

## Genes Intersection

- ✓ genes commonly targeted by  $>n$  miRNAs

*Which pathways are jointly controlled by more than one miRNAs?*

## Pathways/Categories Union

- ✓ significance calculated for each miRNA-pathway separately
- ✓ merged p-value extracted for each pathway (Fisher's meta-analysis method)

*How much does each miRNA contribute in pathway enrichment?*

*What's the combinatory significance of enrichment?*

## Pathways/Categories Intersection

- ✓ Individual and merged p-values calculated as in Pathways Union
- ✓ Only pathways enriched by all queried miRNAs are included

*Which pathways are enriched by all individual miRNAs ?*

*At what combinatory significance?*

# Querying DIANA-mirPath

*Paraskevopoulou, M. D., Vlachos, I. S. & Hatzigeorgiou, A. G. DIANA-TarBase and DIANA Suite Tools: Studying Experimentally Supported microRNA Targets. Current Protocols in Bioinformatics, 12.14. 11-12.14. 18 (2016).*

<http://www.microrna.gr/miRPathv3/>



## Performing functional characterization of microRNAs with miRPath (I)

1. Navigate to (<http://www.microrna.gr/miRPathv3/>).

2. In the query box , add the following miRNAs one at a time and press the “+” button:

*hsa-let-7a-5p*

*hsa-let-7a-3p*

*hsa-let-7b-5p*

*hsa-let-7b-3p*

*hsa-let-7e-5p*

*hsa-let-7e-3p*

Alternatively, you can create a text file, paste each miRNA inside, followed by the source of your choice (e.g. hsa-miR-34-5p|Tarbase) and select “upload a file”.

3. The default mode (Genes union) will perform enrichment analysis of KEGG pathways using all genes targeted by any of the defined miRNAs. Select “Pathways union” to examine each miRNA individually and combine the group’s function with meta-analysis statistics.



## Performing functional characterization of microRNAs with miRPath (II)

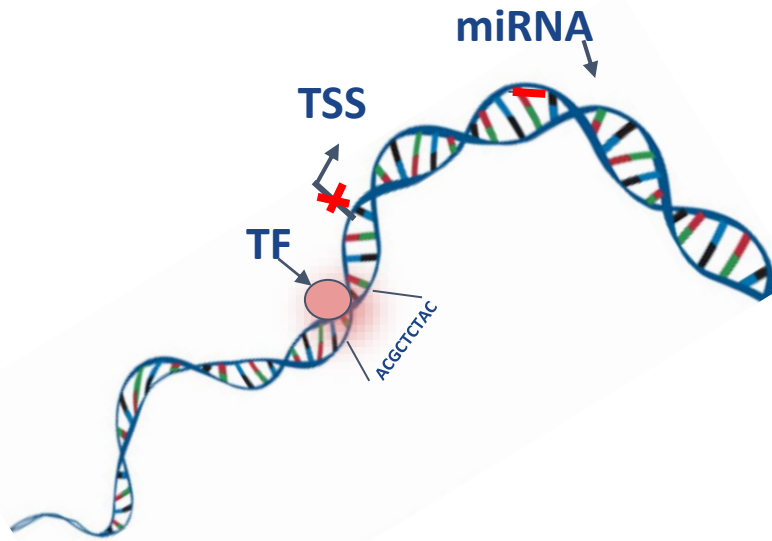
4. In the performed analysis, miRNA targets with experimental validation are retrieved from TarBase v7.0. To use *in silico* predicted targets instead, select “microT-CDS” in every miRNA entry field. Results are updated in real time.
5. Press “details” in “Signaling pathways regulating pluripotency of stem cells” pathway. You can view the 5 miRNAs exerting regulatory roles in this pathway, the targeted genes (“see genes”) and their individual significance levels.
6. Press the link on the pathway name to examine the targeted genes in the pathway. Yellow nodes are targeted by one miRNA and orange ones are targeted by more than one miRNAs.
7. Explore the created graph. miRNAs are clustered based on the pathways/terms they target. Pathways/terms are also clustered on the basis of shared miRNAs. Significance levels are displayed with white-to-red gradient (red – the most significant), allowing to estimate the effect size.
8. Examine the clusters of miRNAs in relation to the significance of their exerted control.

## Performing functional characterization of microRNAs with miRPath (III)

9. Disable the *hsa-let-7a-3p*, *hsa-let-7b-3p* and *hsa-let-7e-3p* and revisit the Heatmap graph.
10. Re-enable all microRNAs. Select “Show microRNA/Pathway Clusters” to only produce dendrograms of miRNA and Pathway clustering analysis.
11. Select “Targeted Pathways Clusters/Heatmap”. Press “Show Heatmap”. Examine the graph. Now the presence/absence of statistical significance is displayed (red/white color) instead of the effect size of the miRNA regulation.
12. Press “Download Results” in the main analysis page to store the results locally in .csv format.



UNIVERSITY OF  
THESSALY

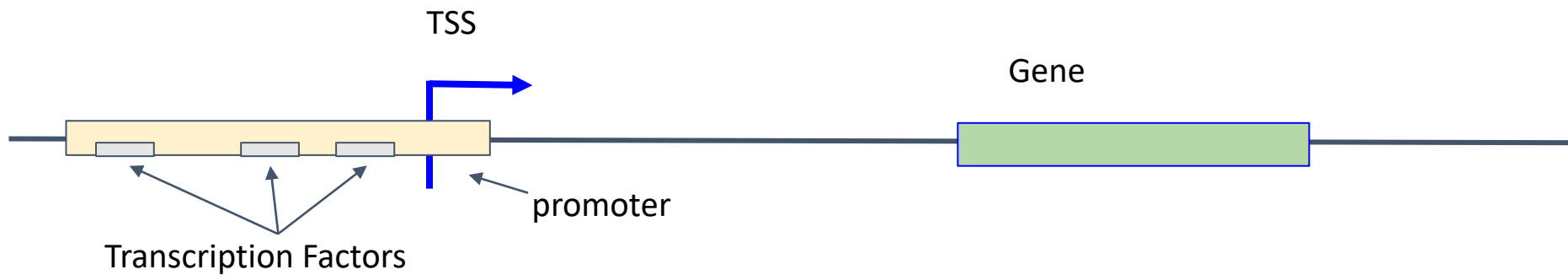


# Indexing promoters for miRNAs using CAGE datasets

Dimitris Grigoriadis



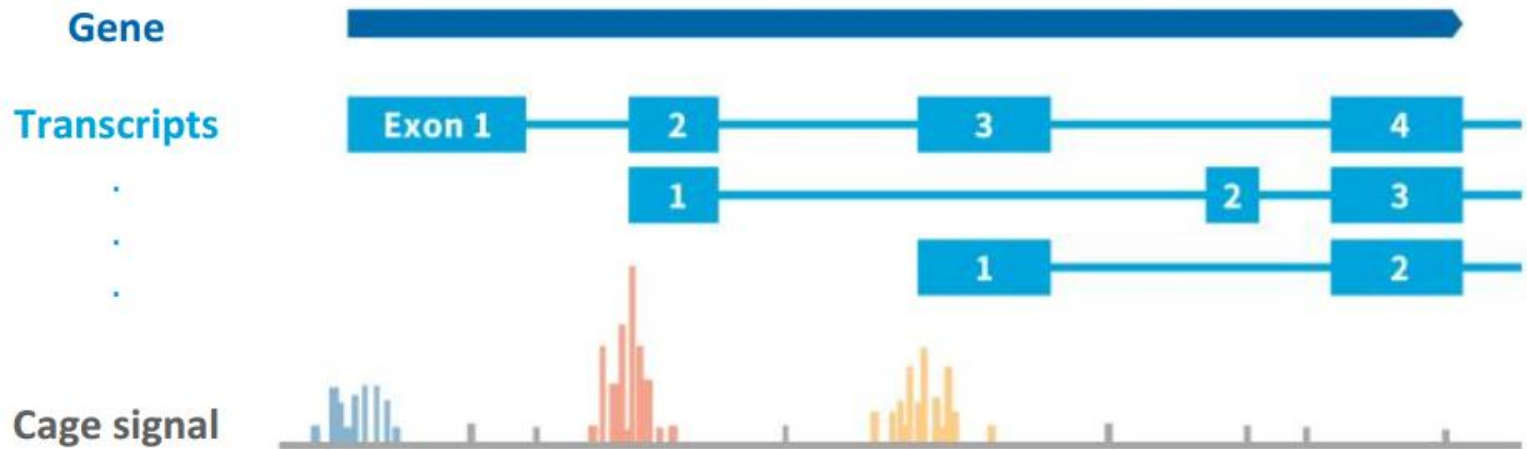
# Annotation



- **Transcription** is the first step of gene expression.
- **Promoter** is a sequence of DNA to which proteins bind and initiate transcription of an RNA molecule.
- **Transcription factor** is a protein that controls the rate of transcription.

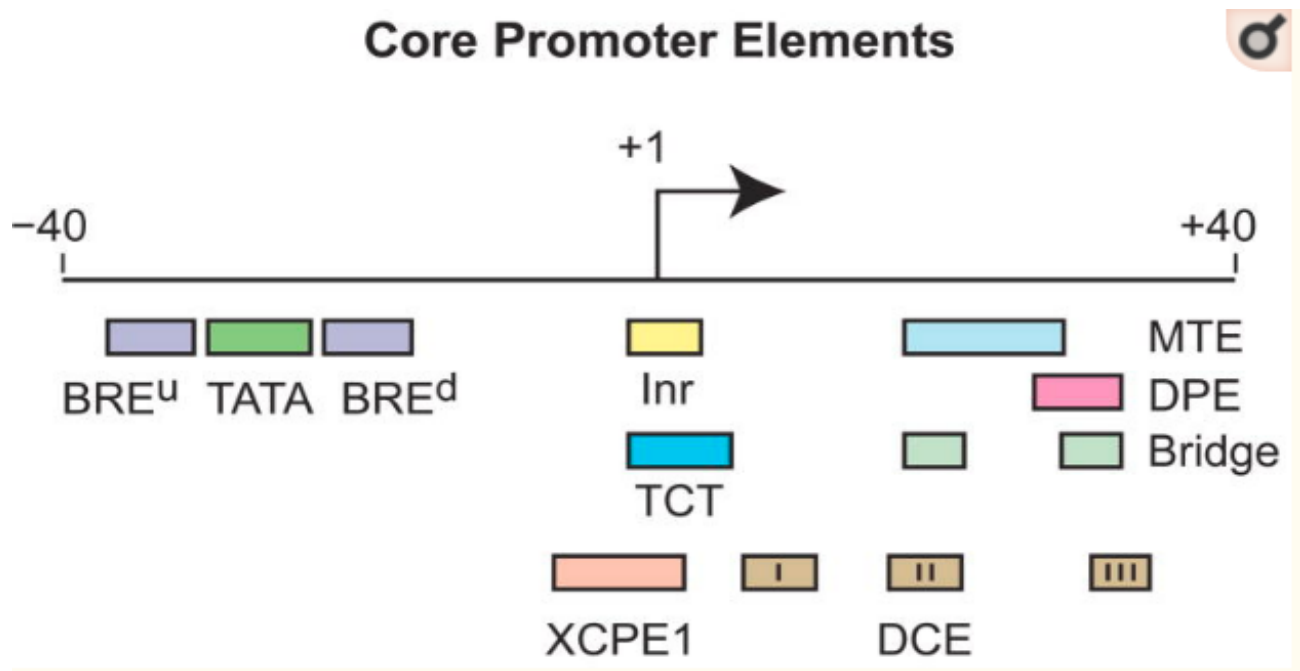
# CAGE-seq

- TSS with single nucleotide resolution
- Expression



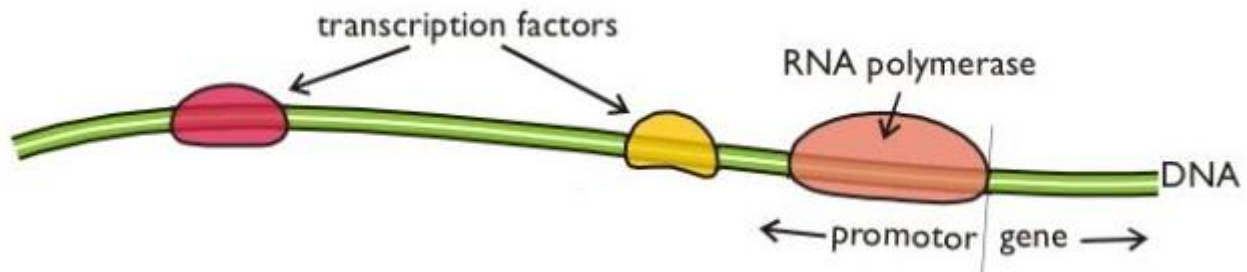
# ADAPT-CAGE

- Machine Learning framework
- Identify Transcription start sites



# Transcription factors

- Control rate of transcription
- Binding to a specific DNA sequence.
- Form a complex in promoter region





# DIANA-miRGen v4

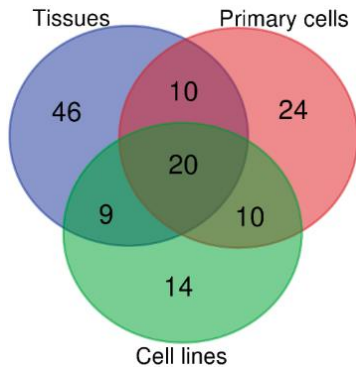
**DIANA-miRGen v4: indexing promoters and regulators for more than 1500 microRNAs**

Nucleic Acids Research, November 2020

Nikos Perdikopanis, Georgios K Georgakilas, Dimitris Grigoriadis, Vasilis Pierros, Ioannis Kavakiotis, Panagiotis Alexiou and Artemis Hatzigeorgiou

# DIANA-miRGen v4

Over 1000 samples and 1500 miRNAs



**Unique and common tissue among tissues, primary cell and cell line samples**

| Counts for 1027 analyzed CAGE samples according type and sample tissue |                   |                   |                 |                 |
|--|-------------------|-------------------|-----------------|-----------------|
| Type   | Number of Samples | Number of Tissues | Healthy Samples | Disease Samples |
| Primary Cell   | 543               | 64                | 521             | 22              |
| Tissue   | 178               | 85                | 178             | 0               |
| Cell Line  | 306               | 53                | 62              | 244             |

# TF: miRNA



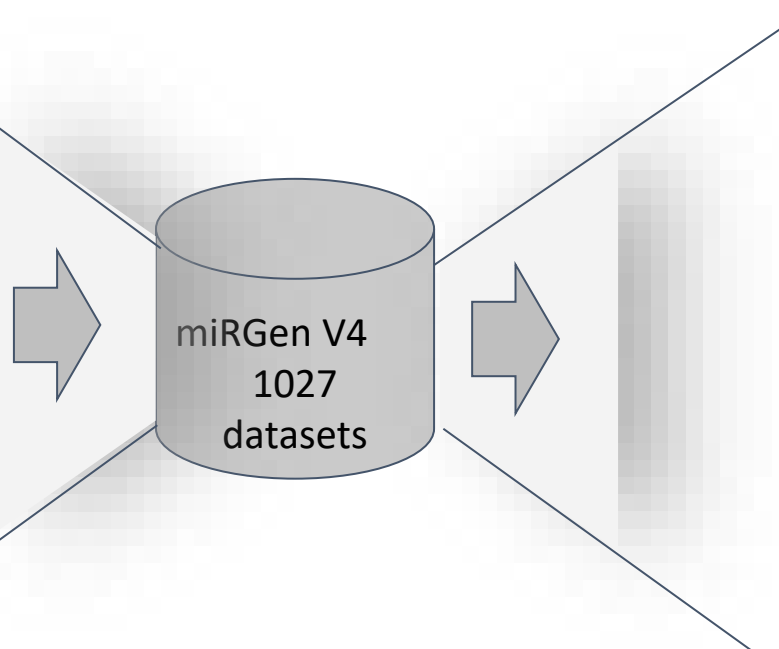
## Annotation of TSS

- Characterization of miRNA promoter regions
- TFBS annotation in promoters using experimental and computational methods

# miRGen v4 Database (2019)

## Input

- miRNA info
- miRNA TSS
- TFs info
- Clusters of TSS
- TFs expression
- TFs : miRNA
- TFBS positions
- TFBS motifs



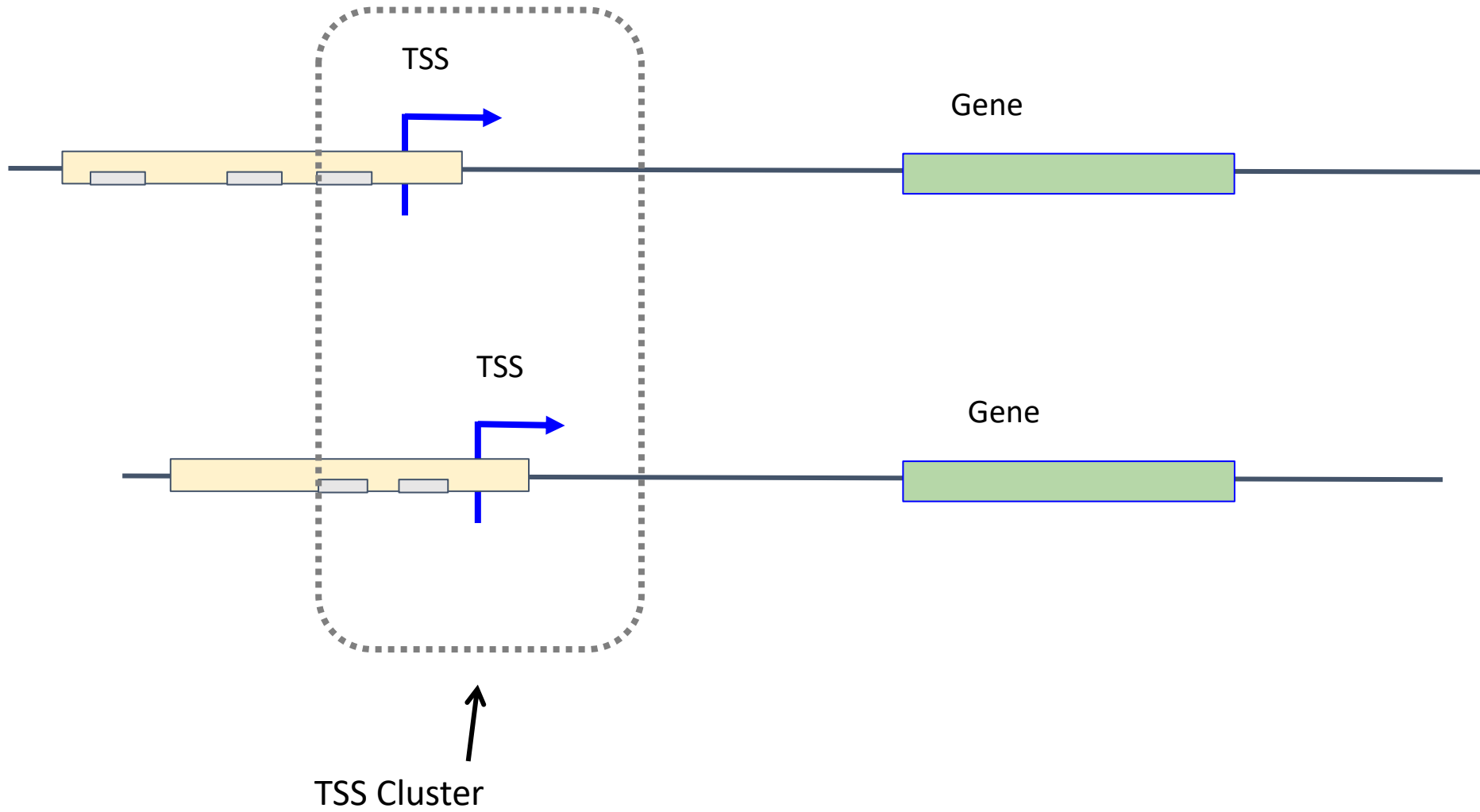
## Interface

The screenshot shows the miRGen v4 web interface with several sections labeled (A) through (G):

- (A)** Search input fields for miRNA (hsa-let-7a-1) and Transcription Factor.
- (B)** Filter options for Types, Tissues, and Samples.
- (C)** Search mode selection: Sample oriented mode, **miRNA oriented mode** (highlighted with a red box), and Execute Search.
- (D)** Search results table with columns for Cluster Code and Coordinates.
- (E)** Detailed search results for hsa-let-7a-1, showing coordinates and TF binding sites.
- (F)** Motif logo for SP1 and E2F4, with expression in TPM and Ensembl Gene IDs.
- (G)** Table of overlaps between selected cluster and search criteria.

<https://diana.e-ce.uth.gr/mirgenv4>

# TSS Cluster



# Annotation



# Index page



Human Mouse | All Healthy Disease | All Only Closest TSS

miRNA

Transcription Factor

Types: **3**

Select Sample Types ▾

Tissues: **135**

Select Sample Tissues ▾

Samples: **1042**

Select Samples ▾

Sample oriented mode miRNA oriented mode | Execute Search | Download

Search by inserting in the corresponding search fields:  
» miRNA names (miRBase v22) e.g., hsa-let-7a-1 » Transcription factor names e.g., KLF5 » combinations of the previous

Run our Example ⚙



# Index page



Human Mouse | All Healthy Disease | All Only Closest TSS

miRNA

Transcription Factor

Types: **3**

Select Sample Types ▾

Tissues: **135**

Select Sample Tissues ▾

Samples: **1042**

Select Samples ▾

Sample oriented mode miRNA oriented mode Execute Search | Download

Search by inserting in the corresponding search fields:

» miRNA names (miRBase v22) e.g., hsa-let-7a-1 » Transcription factor names e.g., KLF5 » combinations of the previous

Run our Example ⚙

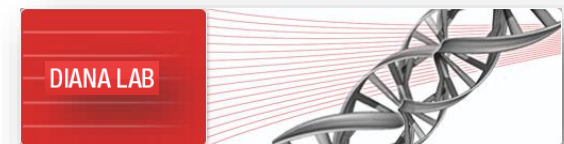
# Queries

1. Find in which cell lines/tissues miRNA hsa-mir-32 is expressed.
2. Filter out disease samples.
3. Find correlations between hsa-let-7a-1 and SP2 (Transcription factor) in adrenal gland tissue.
4. Find samples having same miRNA TSS with hsa-let-7a-1 for Type cell line & adrenal gland Tissue & neuroblastoma (NB-1) Sample.
5. Find SP1 TF overlapping with TSS region with hsa-let-7a-1 for Type cell line & adrenal gland Tissue & neuroblastoma (NB-1) Sample.
6. Explore DIANA Links:
  - i. MicroT-CDS
  - ii. TarBase
  - iii. LncBase
  - iv. mirPath
7. View in UCSC browser
8. Find all clusters miRNA hsa-let-7a-1 is present.
9. Find samples that have the same miRNA cluster with hsa-let-7a-1 for (type) Cell line and (tissue) liver.
10. Find TFs overlapping with miRNA hsa-let-7a-1 cluster for (type) Cell line and (tissue) adrenal gland, sample neuroblastoma NB-1.

## DIANA-mirExTra: DEA suite for (s)RNA-Seq data revealing important miRNAs and TFs

<http://www.microrna.gr/mirextrav2>

*Vlachos, I. S. et al. DIANA-mirExTra v2. 0: Uncovering microRNAs and transcription factors with crucial roles in NGS expression data. Nucleic acids research, gkw455 (2016)*



# DIANA-mirExTra v2.0

DIANA-mirExTra v2.0 (<http://www.microrna.gr/mirextrav2>) is an **online web server** dedicated to uncovering **miRNAs and TFs** playing important regulatory roles between **two investigated states**

mirExTra offers:

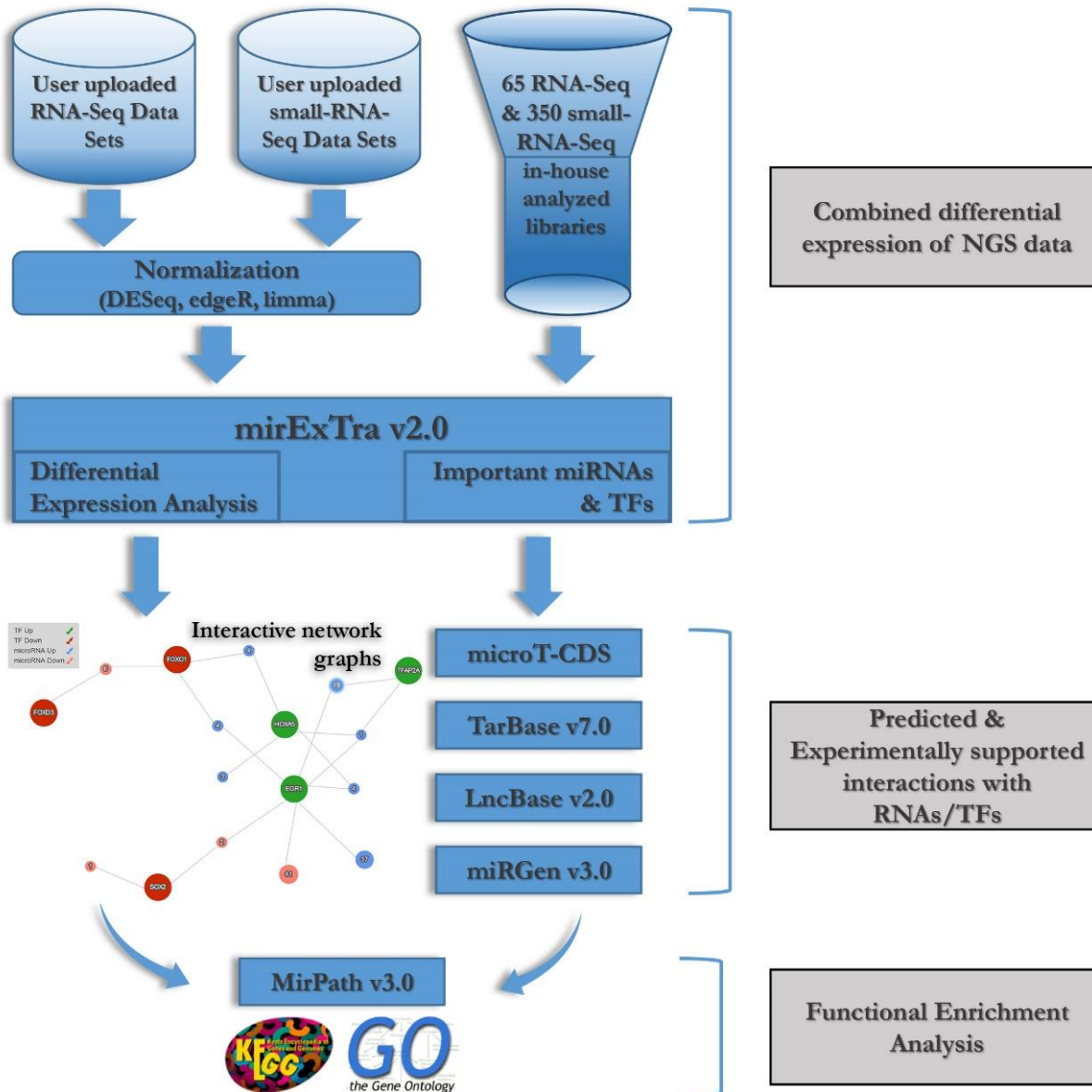
- **An extensive suite for differential expression analysis (DEA)** of RNA-Seq and small-RNA-Seq read count data
- Advanced methodologies for quality checking of NGS expression data and replicate homogeneity tests, including PCA and sample clustering
- **A comprehensive database of ready-to-use NGS expression experiments**
  - ✓ **350 small-RNA-Seq and 65 RNA-Seq** in-house analyzed libraries
  - ✓ covering more than **70 different tissues, 100 cell types** and **90 conditions** in **4 species**

*(Homo sapiens, Mus musculus, Drosophila melanogaster, Caenorhabditis elegans)*

# DIANA-mirExTra v2.0

- **>450,000 experimentally supported miRNA:gene interactions** from DIANA-TarBase 7.0
- **>2 million TF binding sites (TFBSs)** derived from miRGen v3.0 and OregAnno v3.0
- **Predicted miRNA:gene interactions** from DIANA-microT-CDS
- Advanced visualizations, including heat-maps and interactive network graphs
- Exported results enhanced with **rich metadata**
  - ✓ normalized read counts for each sample and group,
  - ✓ fold changes, P-values, FDR levels,
  - ✓ database identifiers (i.e. miRBase and Ensembl accessions),
  - ✓ genomic locations
  - ✓ gene descriptions
- **Intuitive user-friendly interface** available without any login requirements

*DIANA-mirExTra v2.0 can be utilized as a one-stop-shop for differential expression, functional or investigative analyses*



# DIANA-mirExTra v2.0 interface

**2 Modules!**



DIANA TOOLS

mirExTra 2.0

Data Sources  Help 

Select the type of analysis that you want to perform

Differential expression analysis

Find miRNAs and TFs with crucial roles

**Discover microRNAs with central roles between conditions and analyze microRNA/mRNA expression data from NGS experiments with DIANA-miRextra v2.0!** The server enables sophisticated analyses from an easy to use online interface, without requiring bioinformatics expertise of HPC infrastructure.

**Differential Expression Analysis Module (DEA):** Select and compare groups of microRNA or mRNA expression samples (e.g. Normal Stomach samples vs Stomach Adenocarcinoma) from the extensive DIANA expression database. **Hundreds of RNA-Seq and mRNA-Seq libraries with billions of reads** have already been analyzed in-house and uploaded! **Upload your own** expression data from sequencing experiments (RNA-Seq or small-RNA-Seq) and analyze them online. The interface enables powerful analyses and visualizations including differential expression analysis with DESeq, Limma and edgeR, heatmaps, dimensionality reduction (PCA), clustering and more!

**Central microRNA Discovery Module (CmD):** Combine microRNA and mRNA expression data, in order to **identify functional microRNAs** responsible for changes in mRNA expression. You can also upload your own differential expression results or import them from a previous miRExtra analysis! CmD performs a **state-of-the-art** overrepresentation analysis and identifies important microRNAs in your data. It utilizes **in silico** predicted interactions from **DIANA-microT-CDS**, as well as more than 600.000 **experimentally validated interactions from the DIANA-TarBase v7.0**. mRNA and microRNA Differential Expression results are concurrently analyzed and important regulators are found based on functional analysis of their targets.



# DIANA-mirExTra v2.0 interface- DEA module

**1. File Import** → Your miRNAs expr. file (upload icon) or **2. Expression DB** use stored

Your genes expr. file (upload icon) or use stored

**4. Start Analysis** → Differential expression analysis

**12. Example** → [Run example](#)

**3. Analysis Settings** → Species: Human, Statistical method: LIMMA, miRNAs FDR threshold: 0.05, Genes FDR threshold: 0.05, Create comparison graphs: , Create group homogeneity graphs: , Results format: text (.csv)

**9. Functional Analysis** → Find miRNAs controlling mRNAs, Find TFs controlling mRNAs, Find TFs controlling miRNAs

**10. KEGG/GO Analysis** → Run mirPath for Top-20 UP-regulated miRNAs, Run mirPath for Top-20 DOWN-regulated miRNAs, Run mirPath for Top-20 miRNAs

**7. Download** → Download icons for miRNAs and Genes tables

**8. Graphs** → Graphs icons for miRNAs and Genes tables

**5. Result Pane** → miRNAs and Genes tables

**6. Tools & Data** → Information icon (i) in the Genes table

**13. Back & Reset** → Back and Refresh icons

**14. Home** → Home icon

**15. Data Sources** → Data Sources icon

**11. Help** → Help icon

| miRNAs          |          |           |             |         |        | Genes    |        |          |             |         |        |
|-----------------|----------|-----------|-------------|---------|--------|----------|--------|----------|-------------|---------|--------|
| Name            | Group1   | Group2    | Fold Change | P-value | FDR    | Name     | Group1 | Group2   | Fold Change | P-value | FDR    |
| hsa-miR-653-3p  | 3,952.80 | 21,166.40 | 5.354       | 1.3e-10 | 2.5e-7 | C6orf132 | 8.80   | 1,336.60 | 140.00      | 8.1e-12 | 5.7e-8 |
| hsa-miR-30d-3p  | 309.40   | 3,032.80  | 9.774       | 1.3e-9  | 1.2e-6 | AKR1D1   | 0.40   | 160.40   | 120.00      | 8.2e-12 | 5.7e-8 |
| hsa-miR-3168    | 1,683.00 | 38.20     | 0.023       | 9.9e-9  | 6.3e-6 | OR2W3    | 94.60  | 829.60   | 8.70        | 7.1e-11 | 2.5e-7 |
| hsa-miR-383-5p  | 3,842.40 | 10,054.20 | 2.616       | 1.6e-8  | 7.7e-6 | CBLB     | 1.40   | 305.20   | 70.00       | 7.1e-11 | 2.5e-7 |
| hsa-miR-4677-3p | 4.60     | 171.60    | 30.821      | 5.4e-8  | 1.8e-5 | AGO1     | 590.00 | 2,962.00 | 5.00        | 9.3e-11 | 2.6e-7 |
| hsa-miR-3914    | 37.60    | 507.40    | 13.171      | 5.6e-8  | 1.8e-5 | RNF165   | 10.40  | 0.0e+0   | 0.09        | 1.4e-10 | 3.3e-7 |
| hsa-miR-3116    | 80.00    | 1.40      | 0.030       | 8.3e-8  | 2.1e-5 | PGBD3    | 76.60  | 996.80   | 13.00       | 4.7e-10 | 9.3e-7 |

# DEA module – File setups

## miRNA Expression Files

**1st line:** Setup and Group names for each sample separated by tabs.

```
Setup Patient Patient Patient Patient Patient Healthy Healthy Healthy
Healthy
```

**2nd line:** ID and Sample names separated by tabs. The number of groups above and sample names have to be equal

```
ID P1 P2 P3 P4 P5 H1 H2 H3 H4 H5
```

**3rd line till end of the file:** miRBase MIMAT or miRNA name (miRBase 18+ nomenclature (-3p / -5p) and read count (raw read count / not normalized or RPKM). One number for each sample, separated by tabs.

```
MIMAT0000680 15 25 14 12 50 2 1 4 6 10
MIMAT0000104 200 210 250 230 260 30 32 41 56 42
MIMAT0000253 1000 1050 1200 1500 1090 8099 8015 8150 818
9 8400
MIMAT0005824 0 1 0 0 0 100 110 310 120 201
```

# DEA module – File setups

## Gene Expression Files

**1st line:** Setup and Group names for each sample separated by tabs.

Setup Patient Patient Patient Patient Patient Healthy Healthy Healthy Healthy

**2nd line:** ID and Sample names separated by tabs. The number of groups above and sample names have to be equal.

ID P1 P2 P3 P4 P5 H1 H2 H3 H4 H5

**3rd line till end of the file:** Ensemble Gene ID and read count (raw read count / not normalized or RPKM). One number for each sample, separated by tabs.

|                  |      |      |      |      |      |      |      |      |      |      |  |  |  |  |
|------------------|------|------|------|------|------|------|------|------|------|------|--|--|--|--|
| ENSG000000000003 | 15   | 25   | 14   | 12   | 50   | 2    | 1    | 4    | 6    | 10   |  |  |  |  |
| ENSG000000000005 | 200  | 210  | 250  | 230  | 260  | 30   | 32   | 41   | 56   | 42   |  |  |  |  |
| ENSG000000000419 | 1000 | 1050 | 1200 | 1500 | 1090 | 8099 | 8015 | 8150 | 8189 | 8400 |  |  |  |  |
| ENSG000000000457 | 0    | 1    | 0    | 0    | 0    | 100  | 110  | 310  | 120  | 201  |  |  |  |  |

# DEA module – Selecting groups from mirExTra Expression DB

CLOSE X

Please select or import the groups to compare.  
Individual groups can be user-uploaded or selected from the mirExTra expression database.  
Compatible file formats for upload can be found in the Help section

**Species**  
Human ← 1. Species

**Group A**  
Type here to search Groups... ← 2. DB Group A

or  No file selected. ← 3. Import Group A

**Group B**  
Type here to search Groups... ← 4. DB Group B

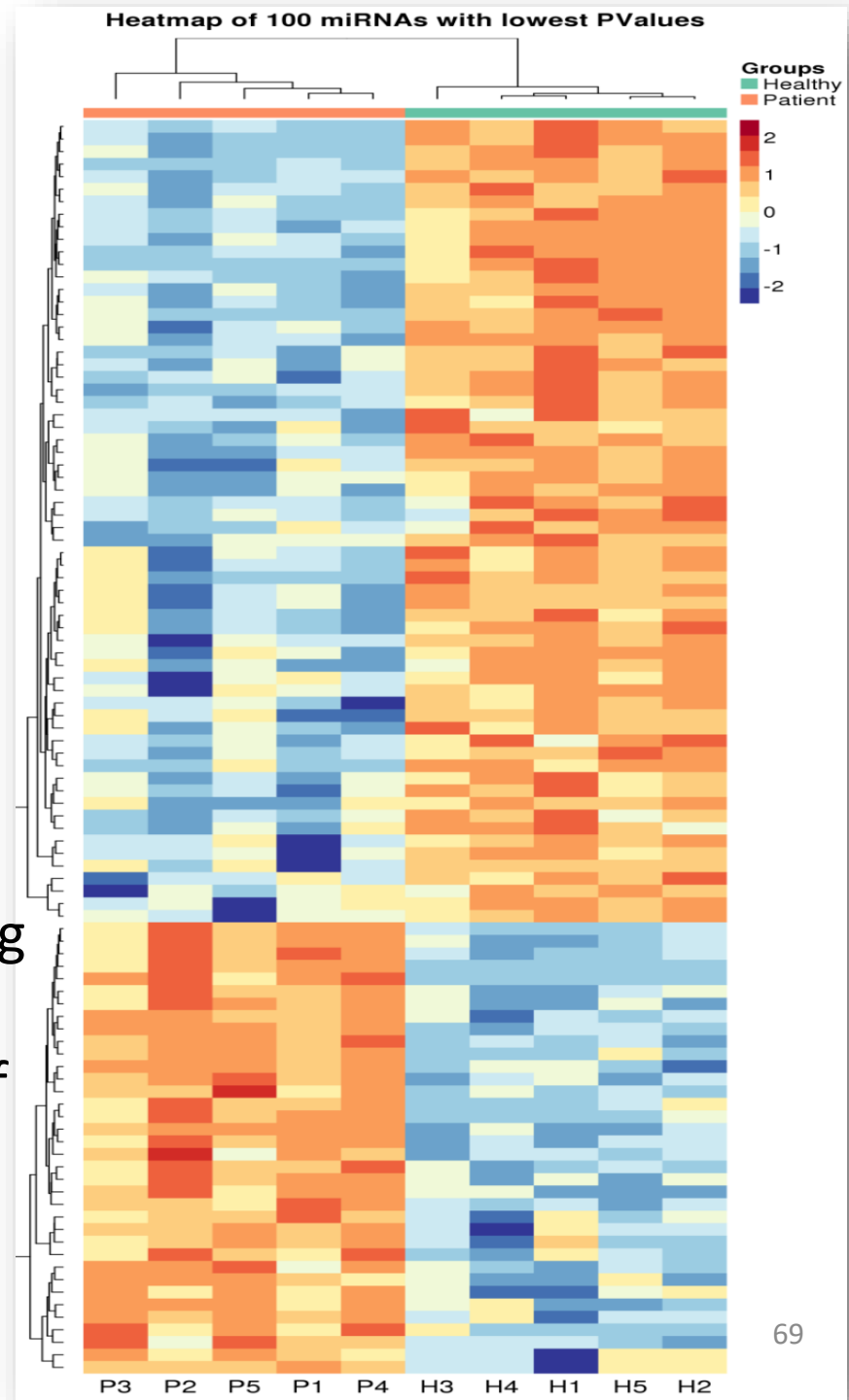
or  No file selected. ← 5. Import Group A

← 6. Finalize Selection

# Graphs (expression heatmap)

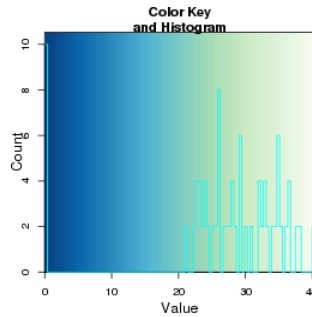
Annotated expression heat map for 100 entries with the lowest p-values.

- Each line = 1 gene/miRNA
- Each column = 1 sample
- Color = level of expression of gene/miRNA in the sample (compared to mean expression of gene)
- Horizontal dendrogram = clustering of samples
- Vertical dendrogram = clustering of genes/miRNAs



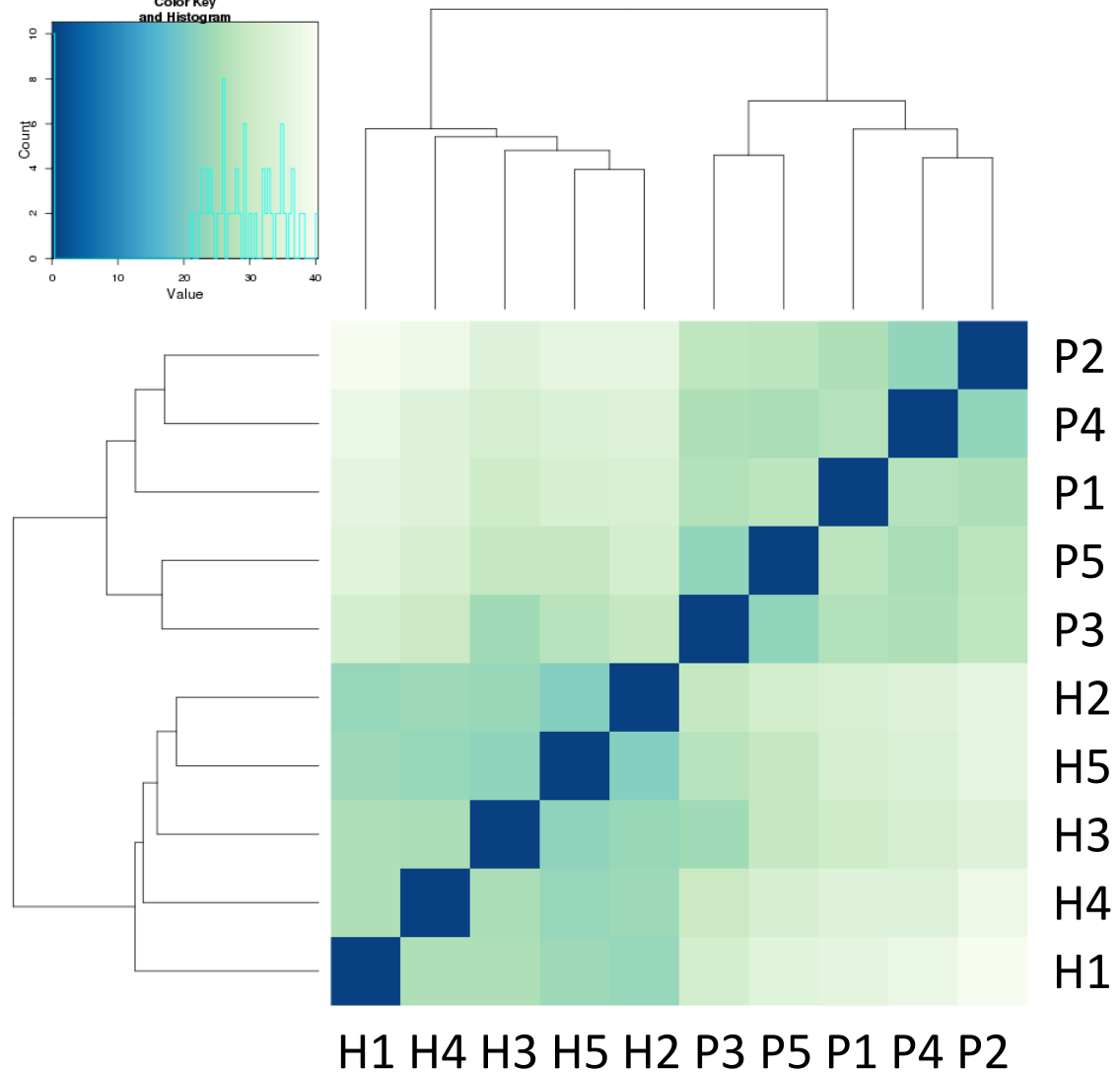
# Graphs (group homogeneity heatmap)

Group homogeneity heatmap, based on mRNA or miRNA expression.



Samples with similar mRNA or miRNA expression patterns are clustered together.

This figure can be used to identify distinct subgroups within the data or samples having dissimilar behaviors.



# Querying DIANA-mirExTra

*Vlachos, I. & Hatzigeorgiou, A. Functional Analysis of miRNAs Using the DIANA Tools Online Suite. Methods in molecular biology (Clifton, NJ) **1517**, 25 (2017)*

# Performing Differential Expression Analysis for microRNAs (I)

1. Navigate to mirExTra v2.0 page (<http://www.microrna.gr/mirextrav2/>).

2. Select “Differential Expression Analysis”. In this protocol, we are using pre-analyzed expression data integrated from the extensive mirExTra database. Select “Use stored” in the miRNA expression file menu.

3. In the pop-up window, choose:

Species: *“Human”*

**Group A: “Healthy Stomach Tissue from Cancer Patients”**

**Group B: “Stomach Adenocarcinoma”**



## Performing Differential Expression Analysis for microRNAs (II)

4. Press “Go!” and proceed to DE settings:

Species: “*Human*”

Statistical Method: “**EDGER**”

miRNA FDR threshold: *0.05*

Gene FDR Threshold: *0.05*

Create comparison graphs: **tick**

Create group homogeneity graphs: **tick**

Results format: **.xls**

5. Select “Differential expression analysis”. When the analysis is completed, the list of de-regulated miRNAs will appear; Significantly up-regulated and down-regulated miRNAs are marked with **green** and **red** respectively.

6. Press the “Graphs” button to examine the created graphs. To save the results file, select “Download” next to “Graphs” button.

7. Use the (i) button next to the miRNA name on the top de-regulated miRNA. Access its relevant entries in (a) microT-CDS, (b) TarBase, (c) LncBase Experimental/Predicted, or (d) miRPath to view *in silico* predicted and experimentally validated targets of the selected miRNA, its experimentally supported/predicted interactions with long-noncoding RNAs and perform functional analysis with miRPath.