

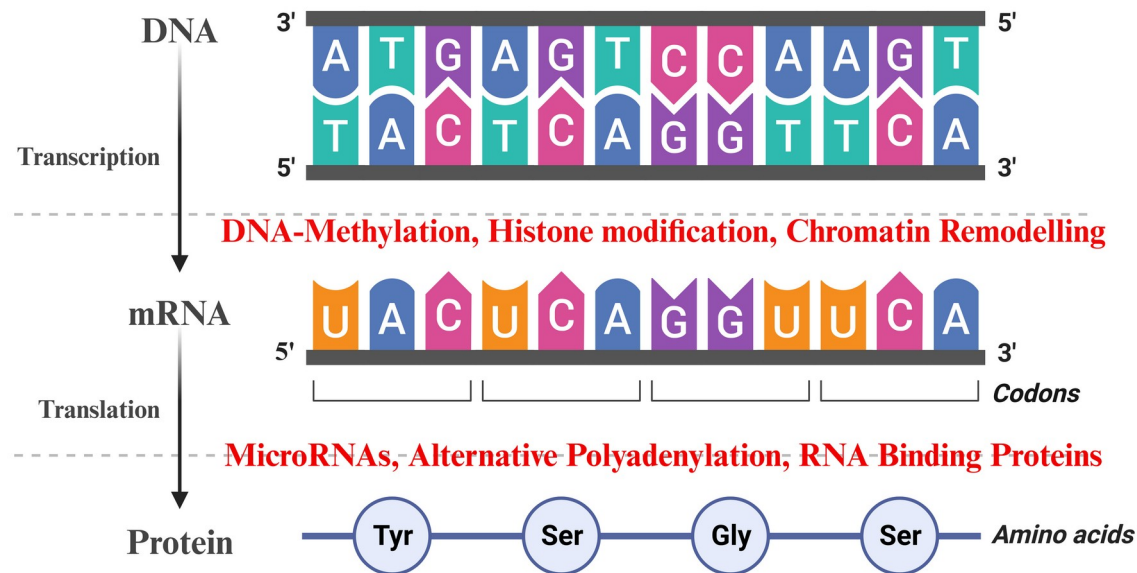
# MicroRNA prediction

Mert Cihan

29.11.2024

Supervisor: Prof. Miguel Andrade

# Gene Expression And Regulation



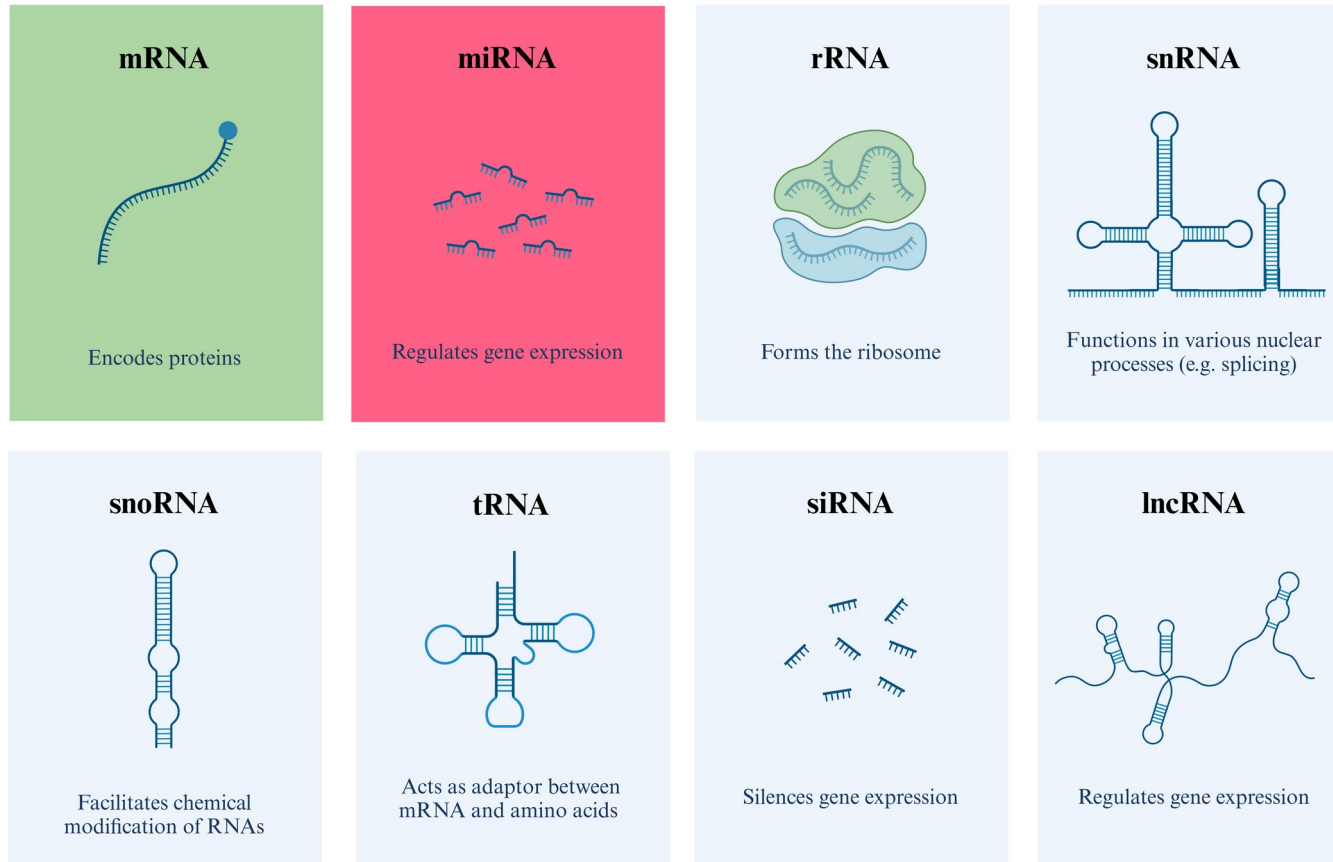
## Pre-Transcriptional

- **DNA Methylation:** Addition of methyl groups to DNA, silencing gene expression.
- **Histone Modifications:** Chemical changes to histone proteins, influencing chromatin structure.
- **Transcription Factor Binding:** Activation or repression of gene transcription by specific proteins.

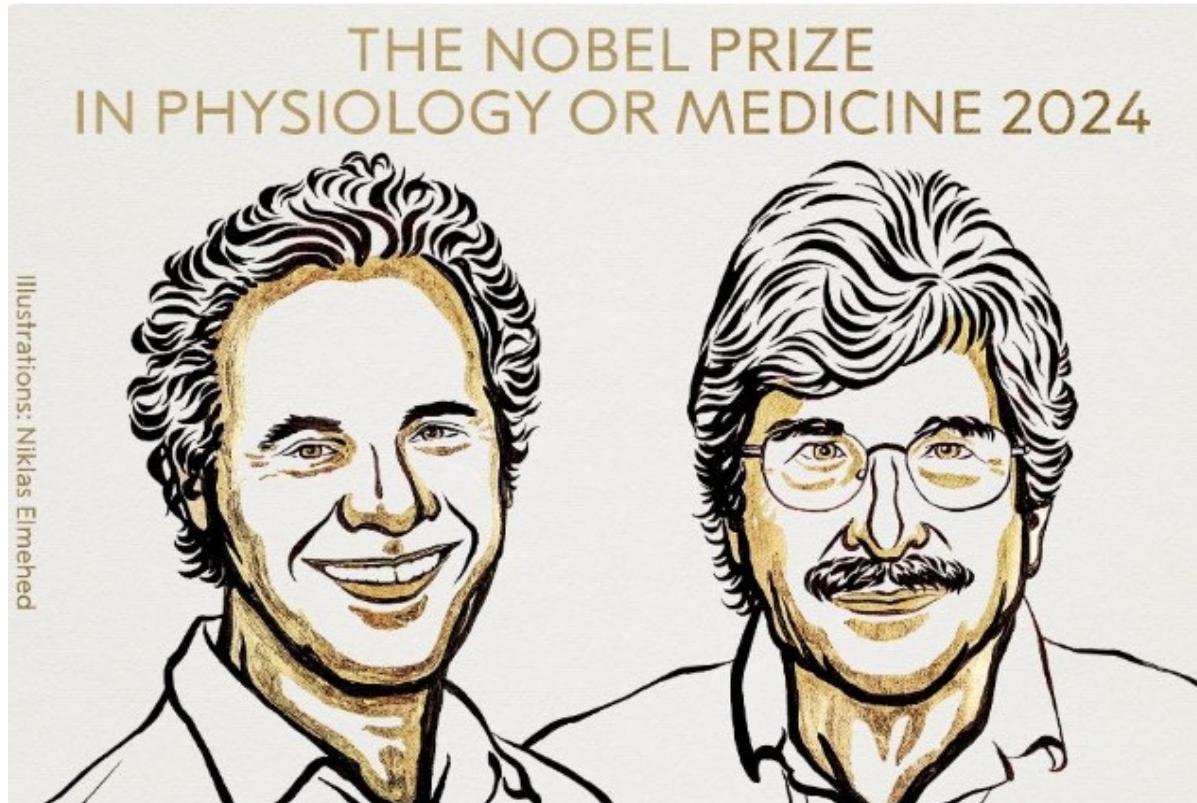
## (Post)Transcriptional

- **RNA Splicing:** Removal of introns and joining of exons to generate diverse mRNA isoforms.
- **mRNA Stability:** Regulation of mRNA half-life influences translation.
- **MicroRNAs (miRNAs):** Small RNAs bind to mRNA 3' UTR, leading to degradation or translation inhibition.

# Transcriptome



- Transcriptome is a vast collection of RNA molecules transcribed from DNA
- Significant proportion of the transcriptome consists of non-coding RNAs (ncRNAs, up to 90 %), which do not code for proteins but play pivotal regulatory roles
- Non-coding RNAs add intricate layers of gene regulation and contribute significantly to the complexity of cellular processes.



'The Nobel Assembly at the Karolinska Institutet has today decided to award

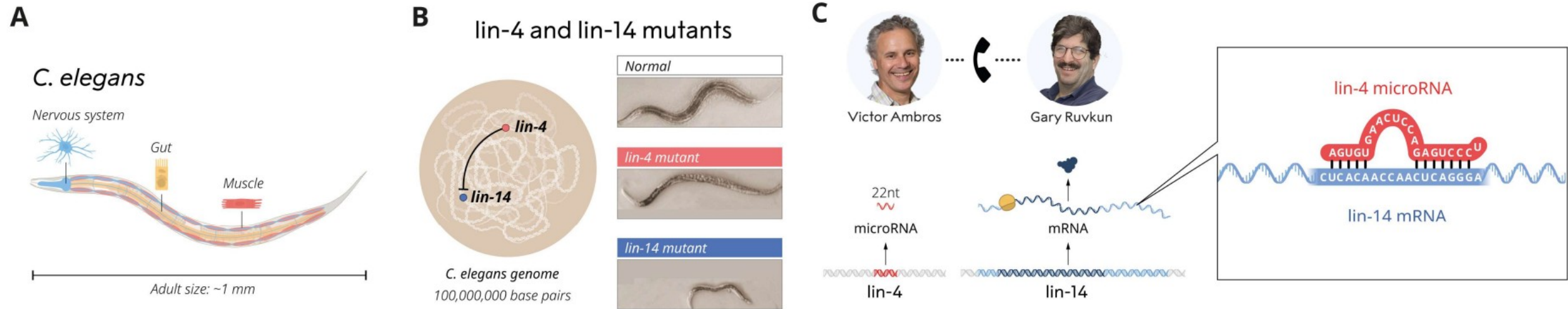
the

2024 Nobel Prize in Physiology or  
Medicine

jointly to **Victor Ambros and Gary  
Ruvkun**

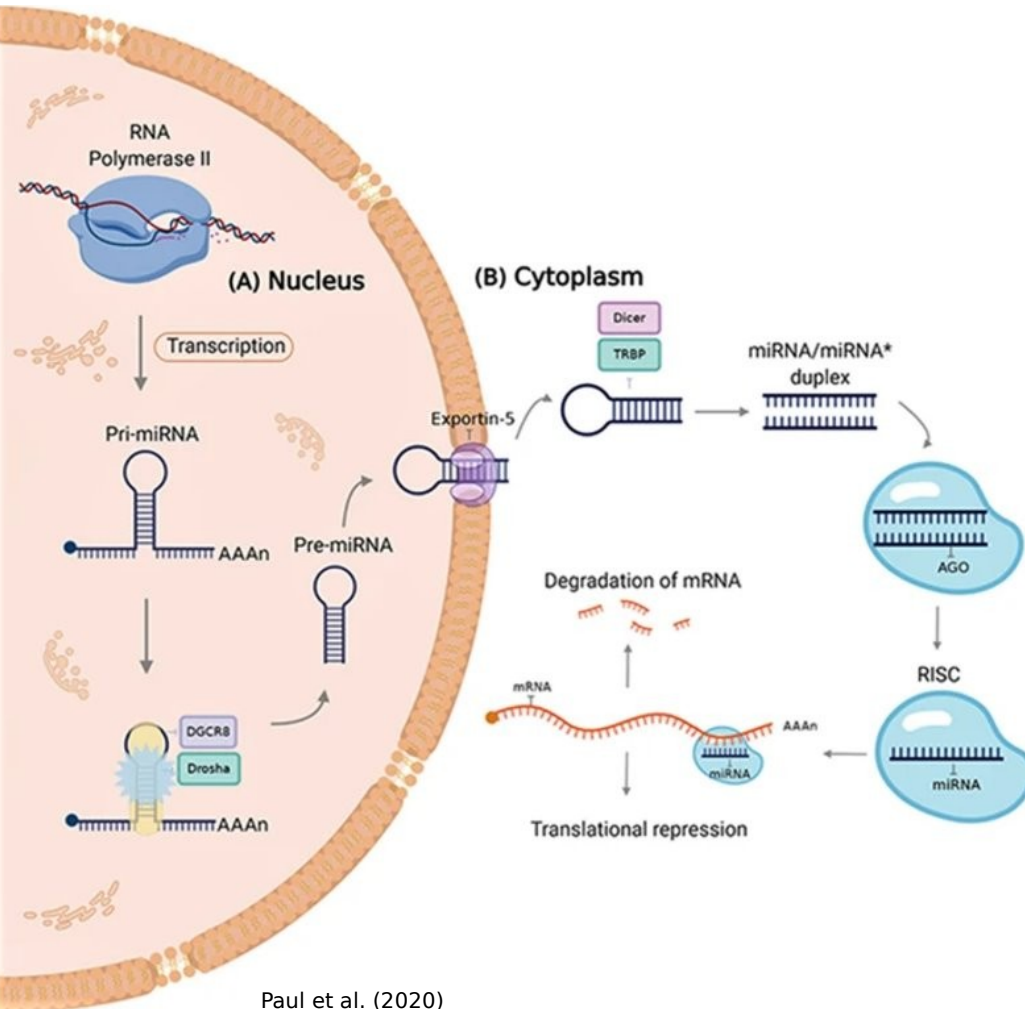
**for the discovery of microRNA and**

# Nobel Price 2024



**Discovery of microRNAs (*lin-4*):** Ambros and Ruvkun found that the *lin-4* gene in *C. elegans* produces a tiny RNA molecule (microRNA) that binds to complementary sequences in the *lin-14* mRNA, blocking protein production and ensuring proper timing of developmental stages.

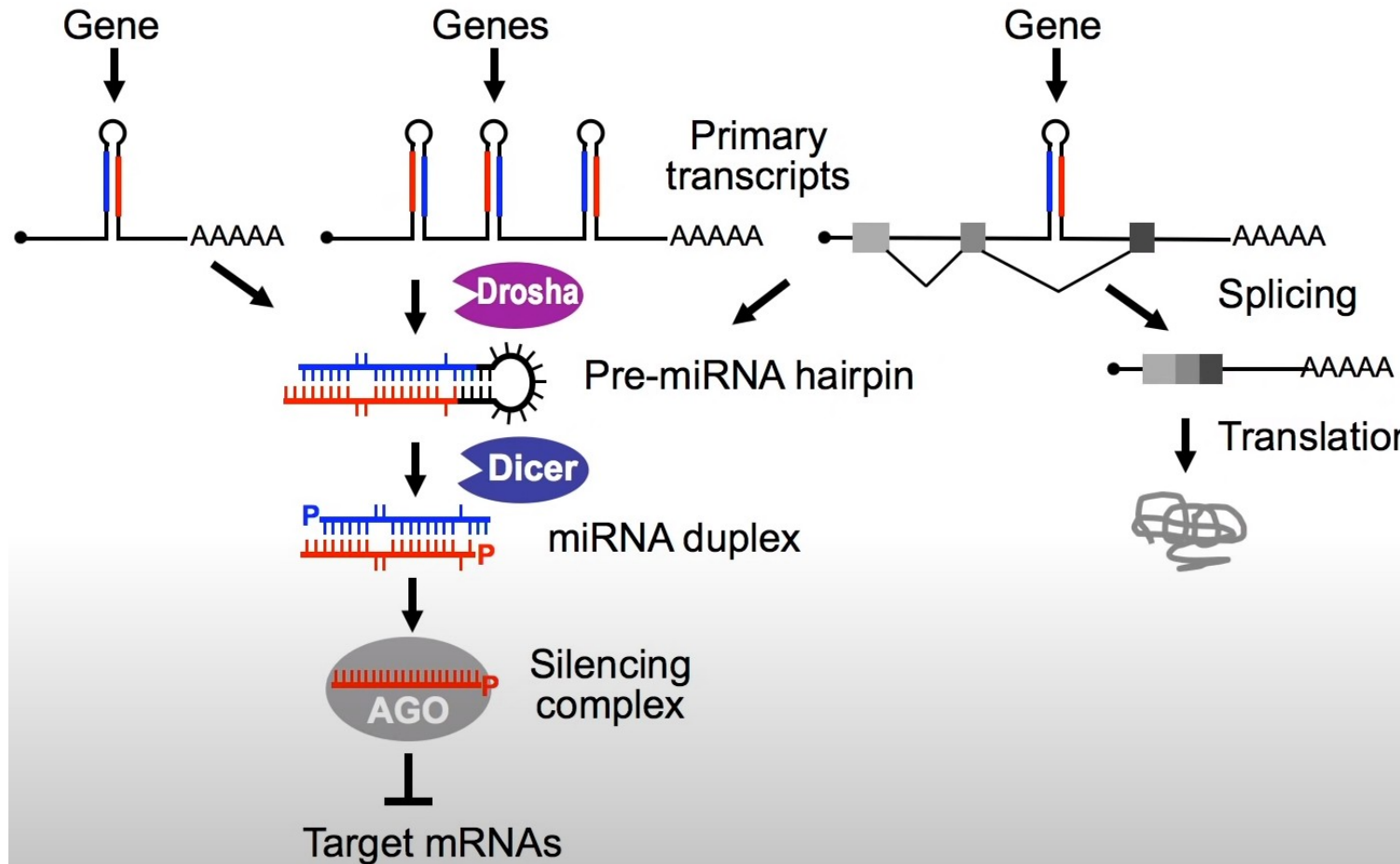
# microRNA regulation



Paul et al. (2020)

- microRNAs are small non-coding RNAs (22nt)
- Approximately 2500 human microRNAs
- Key-proteins: Drosha, Dicer, Argonaute
- **Binding of microRNA seed in silencing complex to complementary 3' untranslated region of mRNA**
- Lead to translational repression/ mRNA degradation
- microRNA families often enriched in targets of transcription factors (redundant functions)

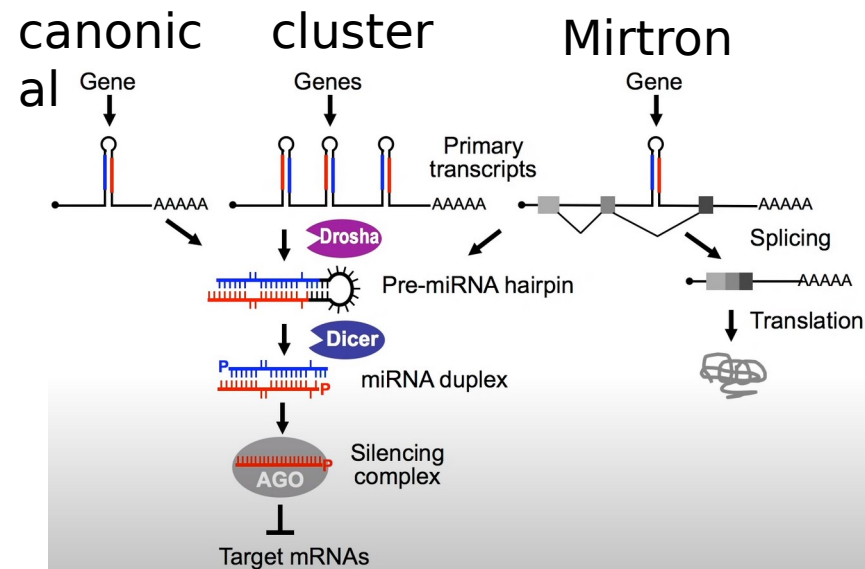
# microRNA regulation



- Approximately 2300 microRNA genes
- Different primary transcripts (up to 6)
- Multiple microRNAs can be within one ORF
- Multifunctional transcripts (MIRTRONS)
- Non-concial biogenesis can avoid Drosha/Dicer dependency

Paul et al. (2020)

# Ex.1 Explore microRNA genes



Use the UCSC genome browser (<https://genome.ucsc.edu/>) to explore human **MIR17HG**, **MIR1224** and **MIR155** and answer the following questions:

Which one is the 'canonical' microRNA gene?

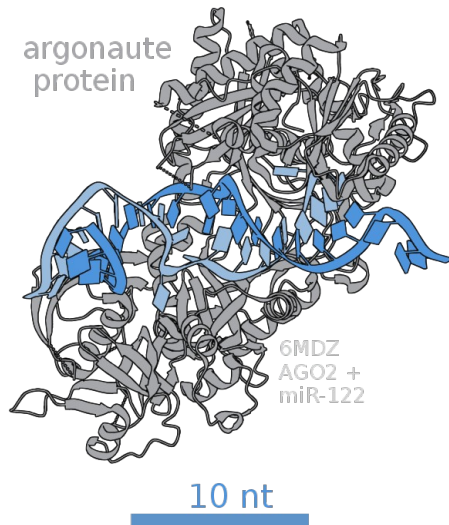
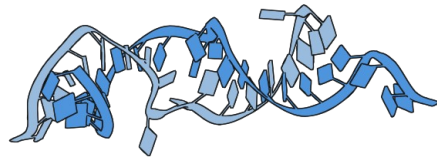
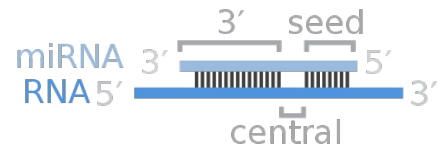
Which one is the Mirtron and what's the host gene?

Which one is the microRNA cluster gene and how many hairpins does it have?



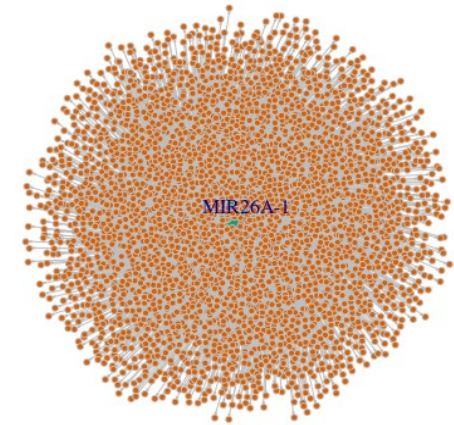
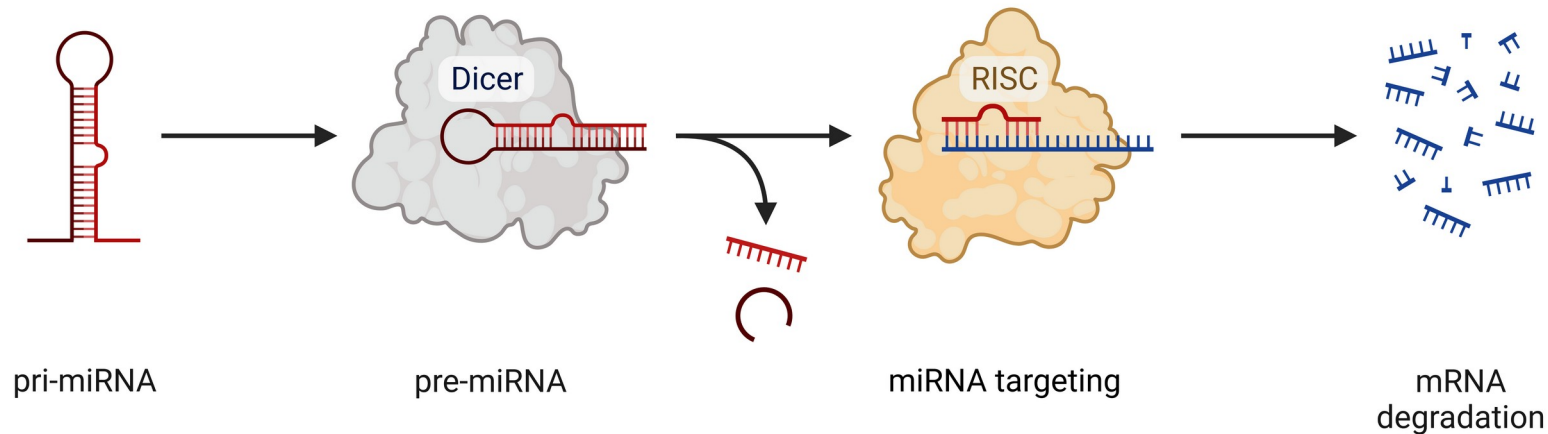


# microRNA regulation



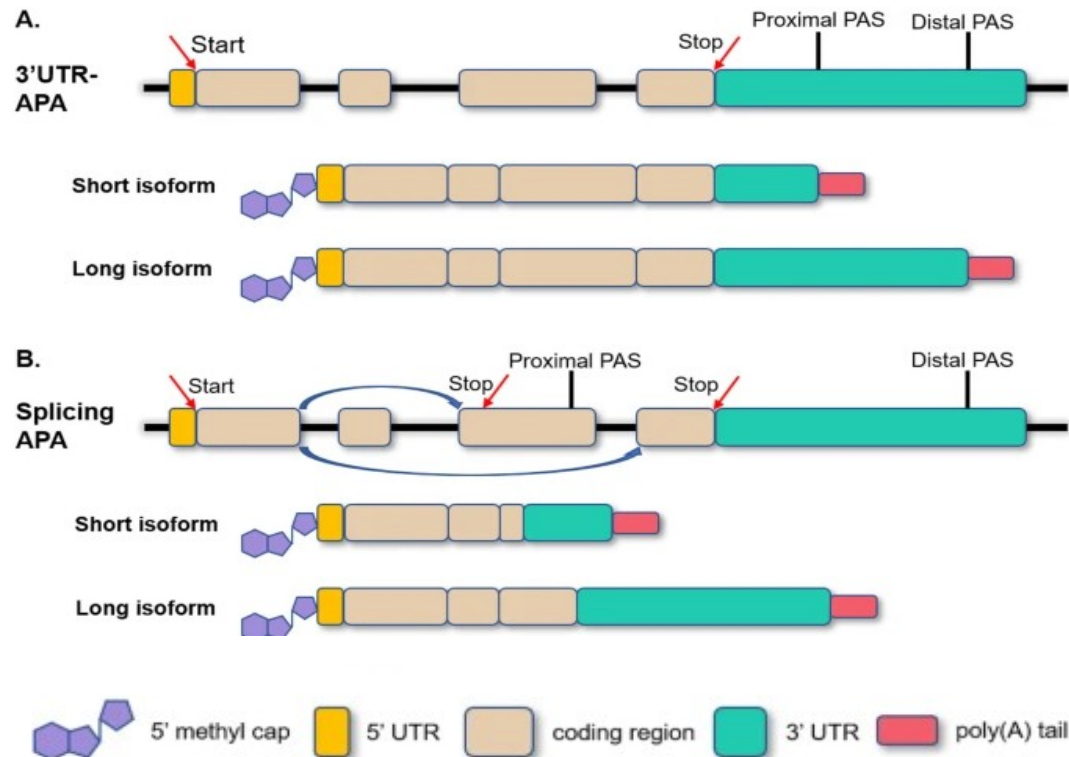
- Mature microRNA is loaded into argonaute protein
- **microRNA seed (nt 2-8 from 5' end in mature microRNA)** binds to complementary 3'UTR sequence of the gene
- 3'UTRs of mRNAs can be really long; on average 800 nt, but up to 10,000 nt
- Searching only complementary sequences leads to many false positives
- Integrate further information for scoring of miRNA-mRNA interactions

# microRNA Regulation



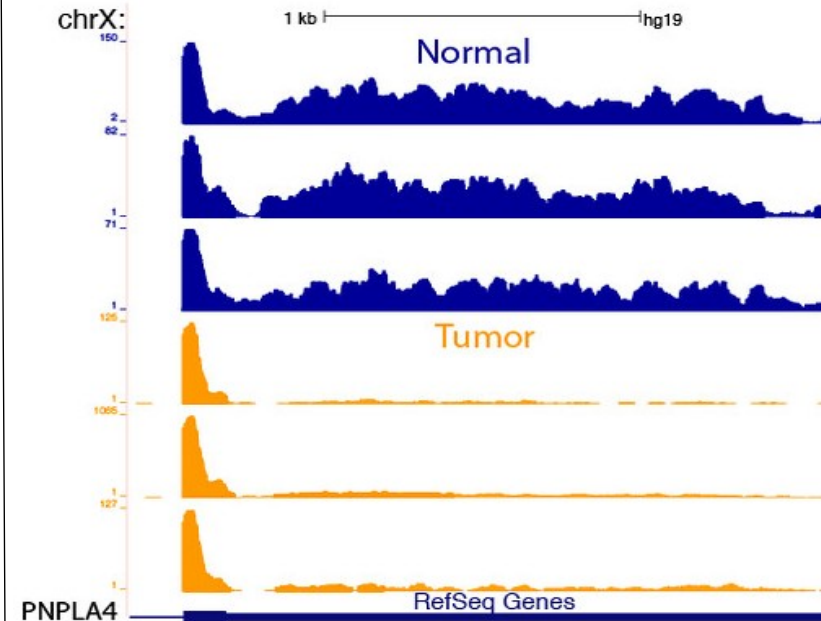
- **Diverse Binding:** MiRNAs can target multiple mRNAs, and one mRNA can have multiple miRNA binding sites due to imperfect complementarity.
- **Complex Prediction:** Predicting miRNA binding sites is challenging because it involves various factors and is not solely based on sequence matching.
- **Experimental Validation:** Accurate validation of miRNA-mRNA interactions often requires experimental techniques in addition to computational predictions.

# Alternative Polyadenylation regulates microRNAs



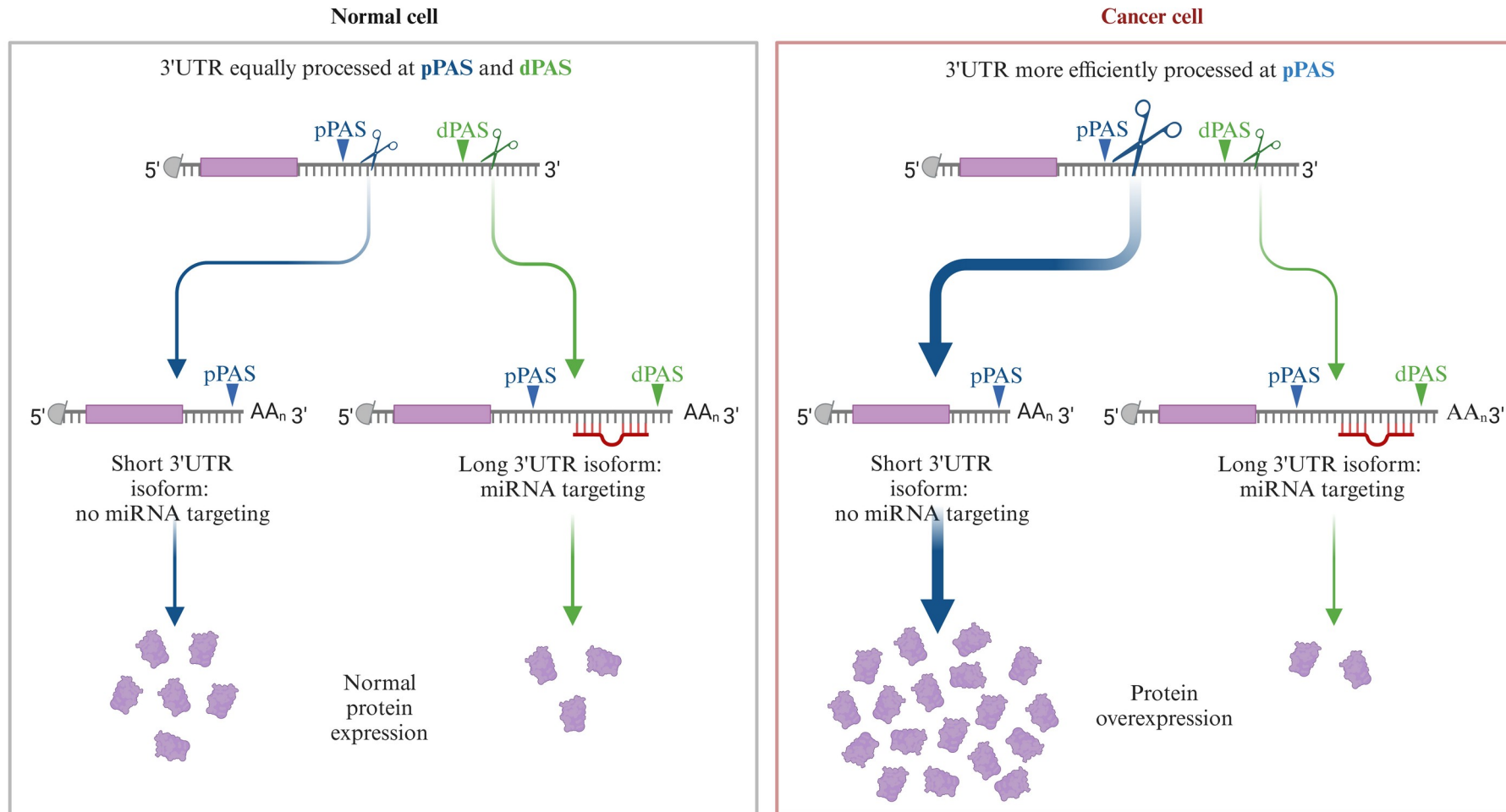
adapted from Zhang et al. (2021)

## DaPars Algorithm / TC3A database



Xia et al. (2014)

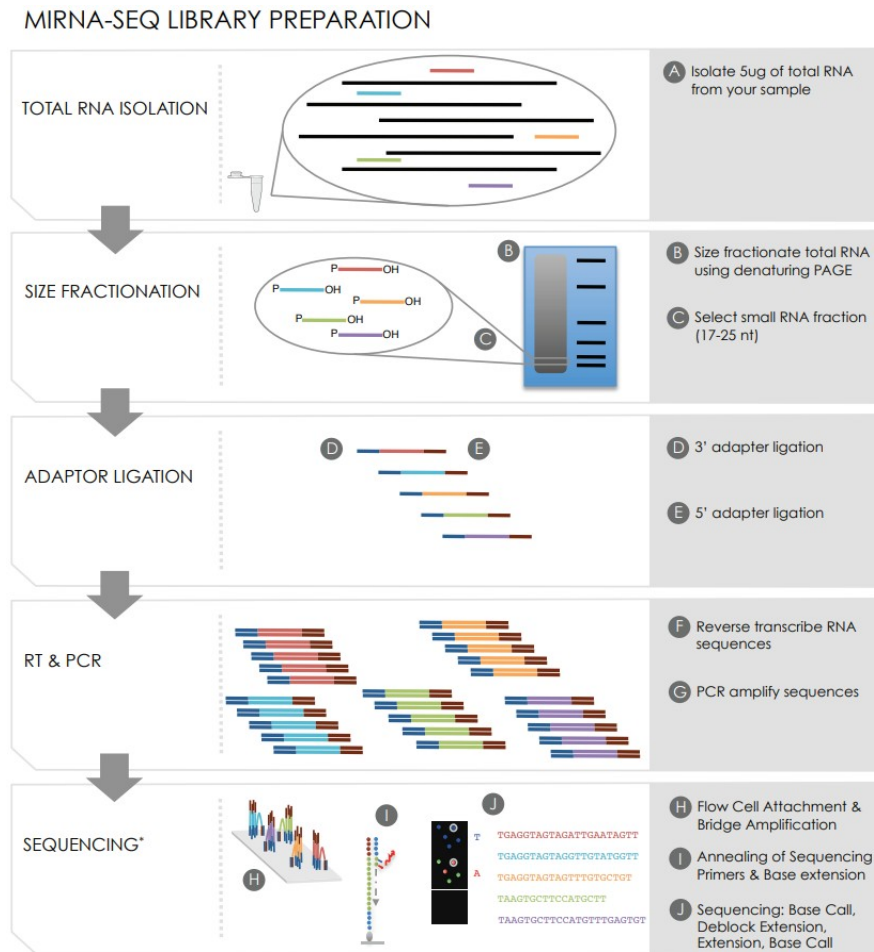
# Alternative Polyadenylation



# Experimental microRNA detection

<b>METHOD</b>	<b>DISADVANTAGES</b>	<b>ADVANTAGES</b>
Northern blotting	Low-throughput-Low sensitivity-Laborious and very time consuming	High specificity-Readily available and easy-to-perform
In situ hybridization	Low-throughput/Semi-quantitative	Monitor cellular and sub-cellular distributions/spatiotemporal expression profile
Reverse transcription( RT-qPCR)	Cannot identify novel miRNAs	High sensitivity and specificity - Can be used for absolute quantification
Microarray	Low-sensitivity and specificity/ Cannot identify novel	Comparing the relative abundance of specific miRNAs/low cost
Next generation sequencing	miRNAs Substantial computational work required	Very high sensitivity-High accuracy in distinguishing variants of miRNAs

# Experimental microRNA detection



\*Illumina sequencing method depicted however other sequencing platforms can also be used.

- A profile of all small RNAs and miRNAs in the transcriptome
- Small RNA targets are enriched through size selection using size-exclusion gels or commercially developed kits
- Follow RNA-seq procedure
- Bias because of degraded RNA fragments

# Ex.2 Annotate microRNAs

**Exercise 1: Find the (hsa-) microRNA sequence from the sequences\_mirna.fa-file using blast** (<https://blast.ncbi.nlm.nih.gov/Blast.cgi>)

>sequence\_1

```
ACATTTACCTAGCAGAAGAAAAATCGTGTTTACGAAGGTGGTTTTTCGCAGGGGCGAAGCTAATTCGTGCAACTTCCCCAAATGTGG  
GAAGCTCGACTGCATAATTTGTGGTAGTGGGAGACTGCGTTCGCTCTTTTCCCCCG
```

>sequence\_2

```
TCCAAACAGACACTGATGGCACCTTCTGCCATTTAGGAATTTGTTTTAAAACAGACATTTGTCTAGATATTTCTTTGTGGCCTCCT  
CCCCATCAAAAGTCAATCAAACATCG
```

>sequence\_3

```
GTAGAGGAGATGGCGCAGGGGACACGGGCAAAGACTTGGGGGTTCTGTTGGGACCCTCAGACGTGTGTCCTCTTCTCCCTCCTCC  
CAG
```

and so on..



# Sol. 2 Annotate microRNAs

**Enter Query Sequence**

Enter accession number(s), gi(s), or FASTA sequence(s) [?](#) [Clear](#)

Query subrange [?](#)

From

To

```
>seq3
GUGGCCUCGUUCAAGUAAUCCAGGAUAGGCUGUGCAGGUCCAAUGG
GCCUAUUCUUGGUACUUGCACGGGGACGC
```

Or, upload file  No file chosen [?](#)

Job Title

Enter a descriptive title for your BLAST search [?](#)

Align two or more sequences [?](#)

**Choose Search Set**

Database  Standard databases (nr etc.):  rRNA/ITS databases  Genomic + trans

[?](#)

Organism  [?](#)

Optional  exclude

Enter organism common name, binomial, or tax id. Only 20 top taxa will be shown [?](#)

**i** Your search is limited to records that include: Home sapiens (taxid:9606)

Job Title **3 sequences (seq1)**

RID [T13Y627V013](#) Search expires on 12-08 20:06 pm [Download All](#)

Results for

Program BLASTN [?](#) [Citation](#)

Database refseq\_rna [See details](#)

Query ID lcl|Query\_27773

Description seq3###

Molecule type rna

Query Length 77

Other reports [Distance tree of results](#) [MSA viewer](#) [?](#)

**Filter Results**

Organism *only top 20 will appear*  exclude

[+ Add organism](#)

Percent Identity  to

E value  to

Query Coverage  to

**Descriptions** | Graphic Summary | Alignments | Taxonomy

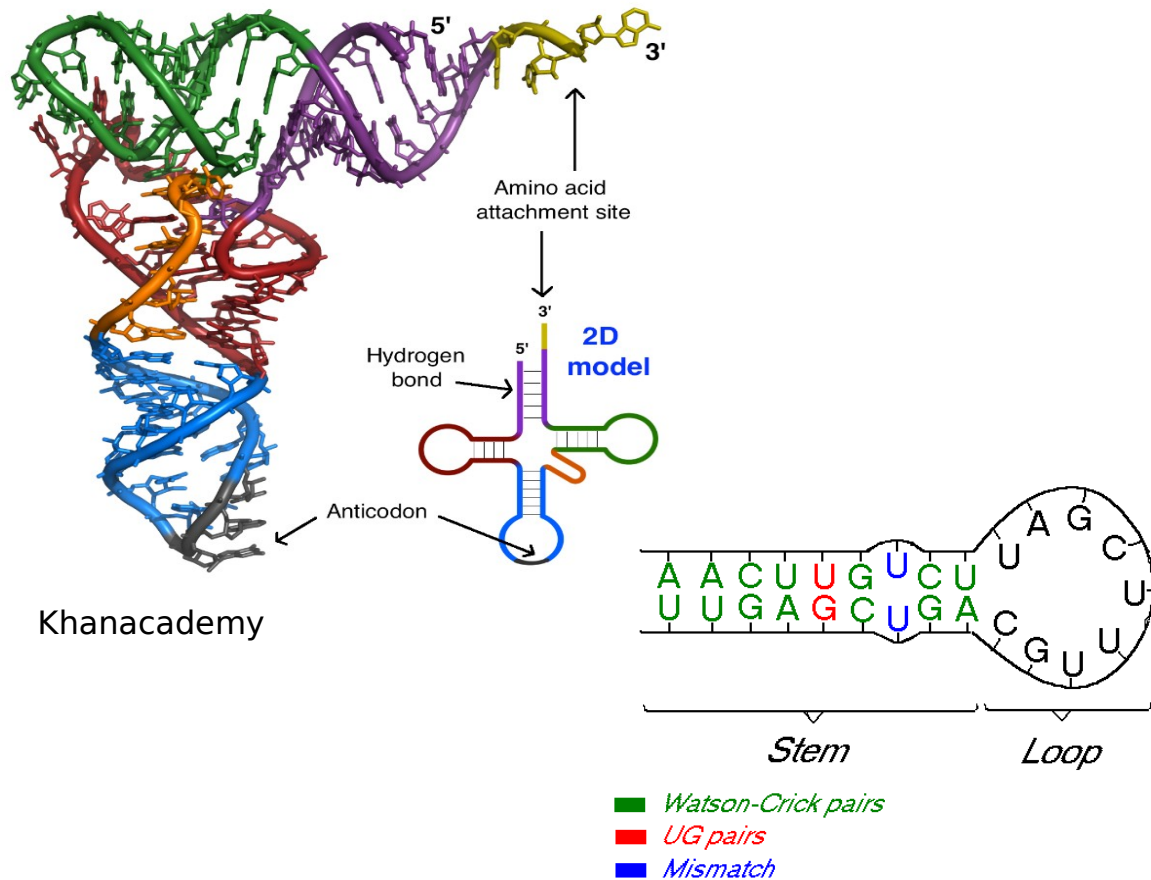
**Sequences producing significant alignments** [Download](#)  [Select columns](#)  Show  [?](#)

select all 1 sequences selected [GenBank](#) [Graphics](#) [Distance tree of results](#) [MSA Viewer](#)

	Description	Scientific Name	Max Score	Total Score	Query Cover	E value	Per. Ident	Acc. Len	Accession
<input checked="" type="checkbox"/>	<a href="#">Homo sapiens microRNA 26a-1 (MIR26A1), microRNA</a>	<a href="#">Homo sapiens</a>	143	143	100%	3e-33	100.00%	77	<a href="#">NR_029499.1</a>

# RNA Secondary Structure

## 3D-structure of tRNA



Khanacademy

- Not only proteins and DNA have secondary structures
- Structured RNA is involved in all aspects of gene expression
- mRNAs are often structured in terminal regions
- Watson-Crick-pairing is preferred, but not as prominent as in DNA
- GU pairing creates wobbles (distinct hydrogen bonds)
- Mismatches can have similar stability

# RNA Secondary Structure

Example:

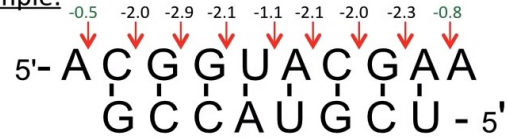
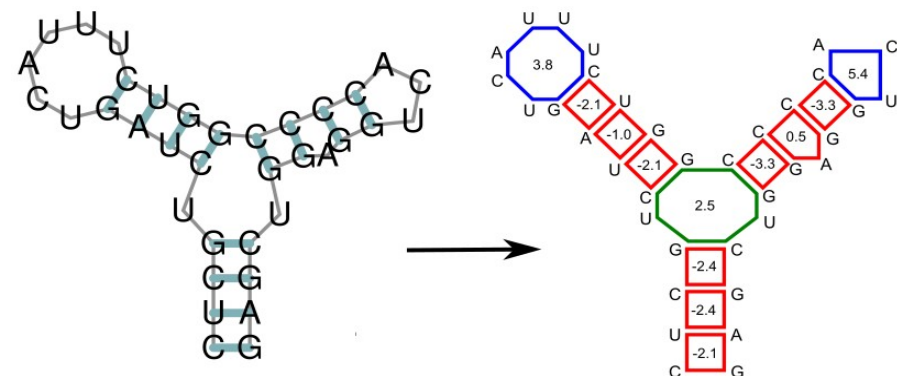
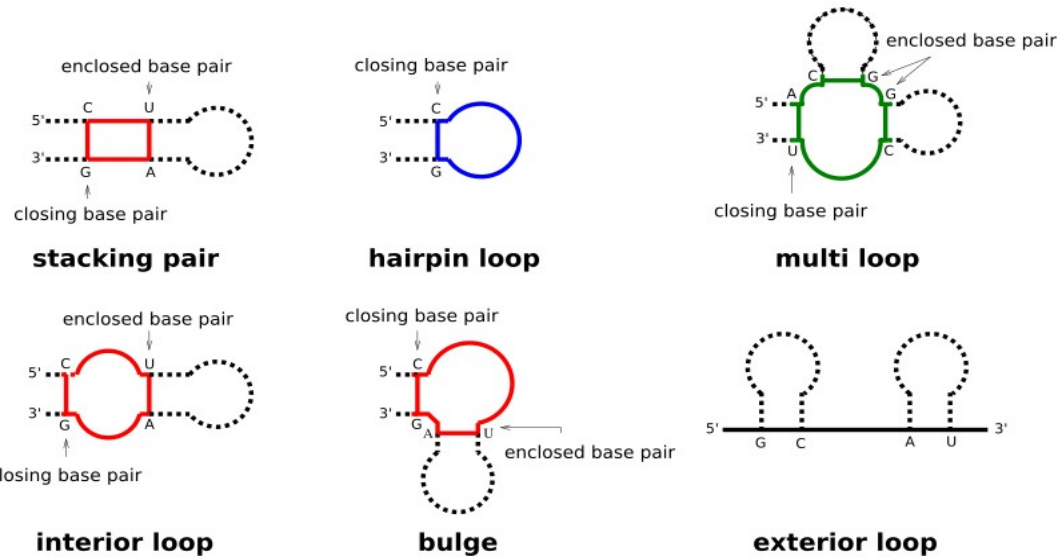


Table 2. Thermodynamic parameters for RNA helix initiation and propagation in 1 M NaCl

Propagation sequence	$\Delta H^\circ$ , kcal/mol	$\Delta S^\circ$ , eu	$\Delta G^\circ$ , kcal/mol
AA ↓ UU	-6.6	-18.4	-0.9
AU ↓ UA	-5.7	-15.5	-0.9
UA ↓ AU	-8.1	-22.6	-1.1
CA ↓ GU	-10.5	-27.8	-1.8
CU ↓ GA	-7.6	-19.2	-1.7
GA ↓ CU	-13.3	-35.5	-2.3
GU ↓ CA	-10.2	-26.2	-2.1
CG ↓ GC	-8.0	-19.4	-2.0
GC ↓ CG	-14.2	-34.9	-3.4
GG ↓ CC	-12.2	-29.7	-2.9
Initiation	(0)	-10.8	3.4
Symmetry correction (self-complementary)	0	-1.4	0.4
Symmetry correction (non-self-complementary)	0	0	0

Freier et al. (1986)



- Turner 1999 RNA parameters
- Mathews 1999 DNA parameters
- Andronescu 2007 RNA parameters
- Mathews 2004 DNA parameters

# Ex.3 microRNA folding

1. Use the *RNAfold webserver* to compute the secondary structure of prior annotated microRNA sequence. **microRNA 26a-1**
2. Find the 22 nt mature microRNA-3p and -5p sequences. (*Tip: Dicer 5' cuts closing basepairs of loop*)

**RNAfold WebServer** 1 Enter Input Parameters 2 View Results

[Home][New job][Help]

The **RNAfold web server** will predict secondary structures of single stranded RNA or DNA sequences. Current limits are 7,500 nt for partition function calculations and 10,000 nt for minimum free energy only predictions.

Simply paste or upload your sequence below and click Proceed. To get more information on the meaning of the options click the symbols. You can test the server using [this sample sequence](#).

Paste or type your sequence here:  [clear]

Show constraint folding

Or upload a file in FASTA format:  No file chosen

**Fold algorithms and basic options**

- minimum free energy (MFE) and partition function
- minimum free energy (MFE) only
- no GU pairs at the end of helices
- avoid isolated base pairs

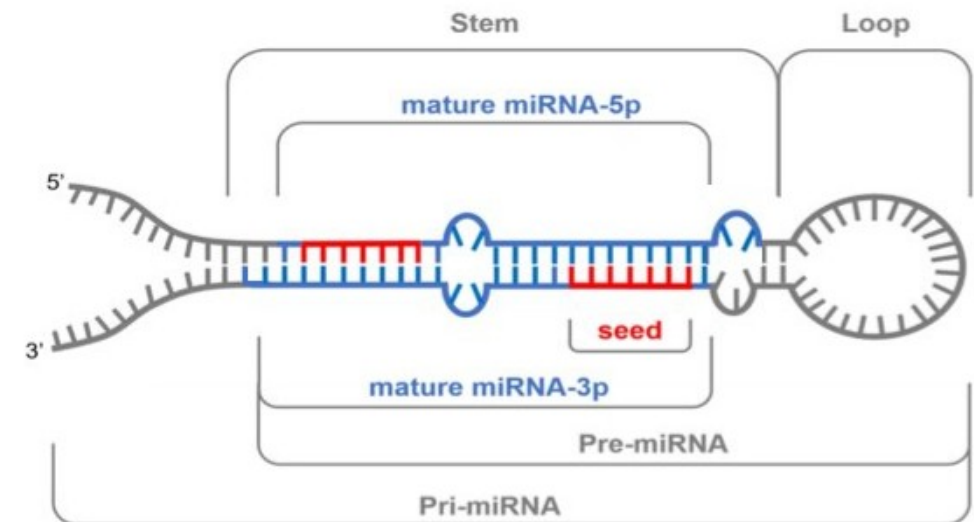
Show advanced options

**Output options**

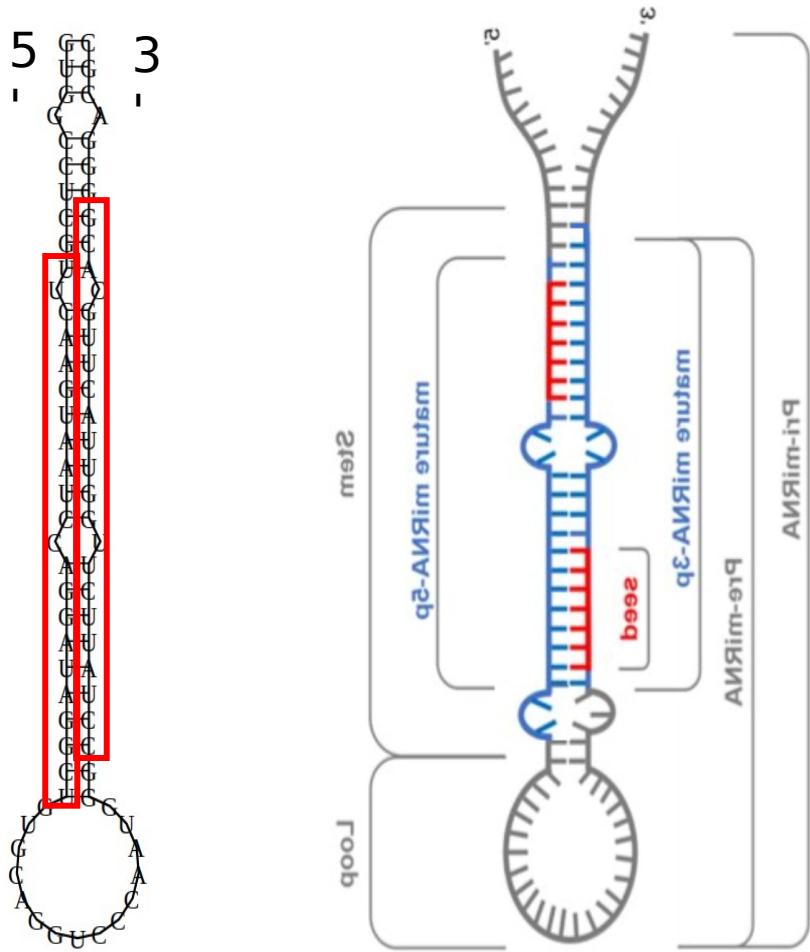
- interactive RNA secondary structure plot
- RNA secondary structure plots with reliability annotation (Partition function folding only)
- Mountain plot

Notification via e-mail upon completion of the job (optional):

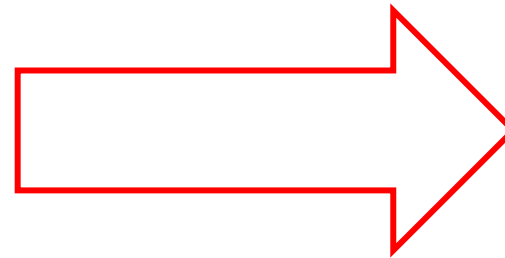
Institute for Theoretical Chemistry | University of Vienna | ma@tbi.univie.ac.at



# Sol.3 microRNA folding

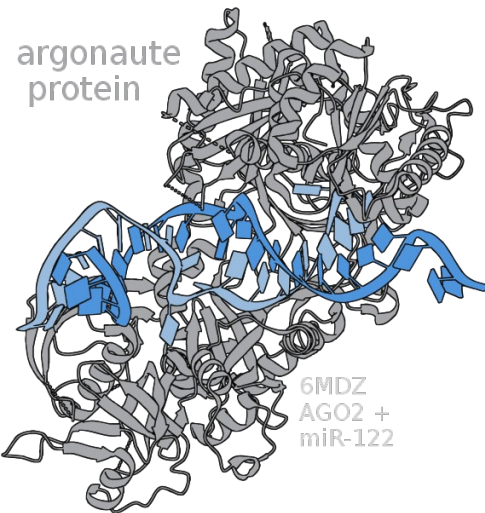
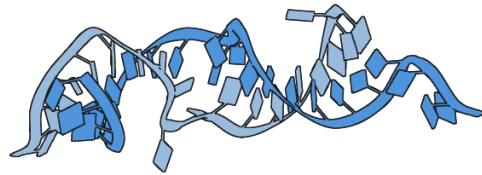
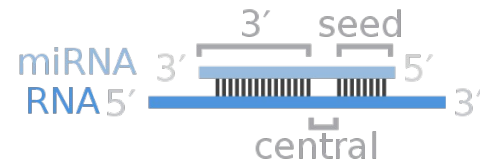


MFE secondary structure



Check further Information on miR  
Search for hsa-mir-26a-5p

# Ex. 4 microRNA seed



10 nt

1. How many human genes are potentially targeted by previously identified mature microRNA (hsa- miR-26-5p)?  
([www.targetscan.org/vert\\_80/](http://www.targetscan.org/vert_80/))
2. Which gene has the most conserved 8-mer binding sites?
3. Find the microRNA seed of previously identified mature microRNA. Does it match the positions in the mature microRNA?

# Sol. 4 microRNA seed

Human miR-26-5p

**1045** transcripts with conserved sites, containing a total of **1209** conserved sites and **579** poorly conserved sites.

Please note that these predicted targets include some false positives. [\[read more\]](#)

Genes with only poorly conserved sites are not shown. [\[View top predicted targets, irrespective of site conservation\]](#)

Table sorted by cumulative weighted context++ score [\[Sort table by predicted occupancy\]](#) [\[Sort table by aggregate P<sub>CT</sub>\]](#)

The table shows at most one transcript per gene, selected for being the most prevalent, based on 3P-seq tags. [\[Download table\]](#)

Target gene	Representative transcript	Gene name	Number of 3P-seq tags supporting UTR + 5	Link to sites in UTRs	Conserved sites				Poorly conserved sites				6mer sites	Representative miRNA	Predicted occupancy			Cumulative weighted context++ score	Total context++ score	Aggregate P <sub>CT</sub>	Previous TargetScan publication(s)
					total	8mer	7mer-m8	7mer-A1	total	8mer	7mer-m8	7mer-A1			mod miRNA	high miRNA	transfected miRNA				
	ENST0000...	polymerase (RNA) III (DNA directed) polypeptide G																			2009, '11
<b>PTEN</b>	0371953.3	phosphatase and tensin homolog	118	<a href="#">Sites in UTR</a>	4	3	0	1	0	0	0	0	0	hsa-miR-26b-5p	0.1153	0.6550	2.1149	-0.55	-0.60	> 0.99	2011, '15

## Mature sequence hsa-miR-26a-5p

Accession [MIMAT0000082](#)

Previous IDs hsa-miR-26a

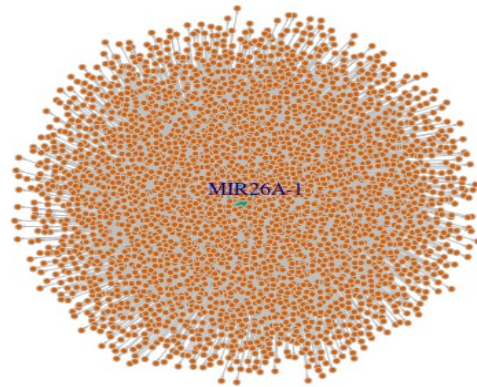
Sequence 10 - **ucaagua**uccaggauaggcu - 31

[Get sequence](#)

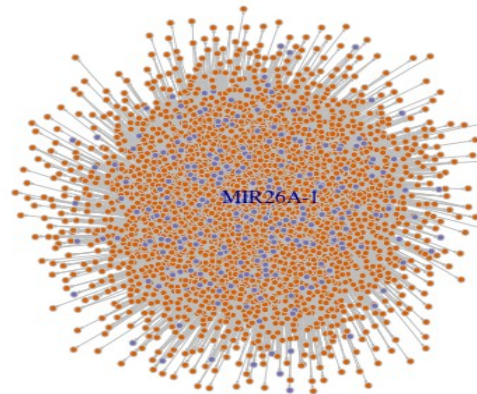
Deep sequencing [28542742](#) reads, 159 experiments

Position 1261-1268 of PTEN 3' UTR	5' ...ACUGUUAGGGAAUUUUACUUGAA...	
hsa-miR-26a-5p	3' UCGGAUAGGACCU <b>AAUGAACU</b>	8mer

# microRNA target prediction



all interactions MIR26A (7897)



conserved sites MIR26A (3196)

- microRNA-mRNA interactions often false-positive
- Experimental validation often indirect from high-throughput
- Integration of further information can help ensuring accuracy:
  - Site type
  - Supplementary pairing
  - Local AU
  - Minimum distance
  - 3' UTR length
  - TA (target site abundance)
  - SPS (seed-pairing stability)
  - Conservation of 3'UTR
  - Conservation of microRNA family
  - Thermodynamic stability of precursor microRNA