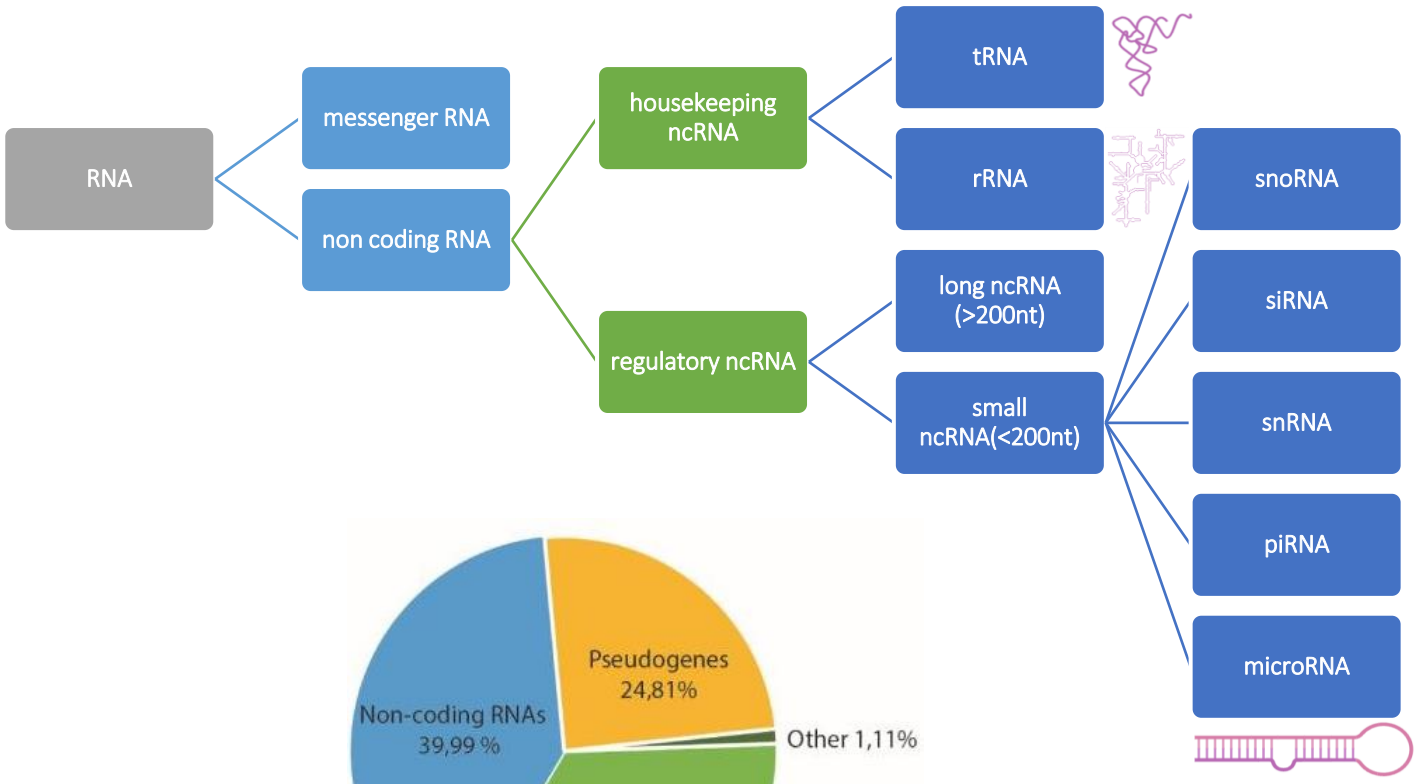
microRNA prediction

Mert Cihan

12.12.2022

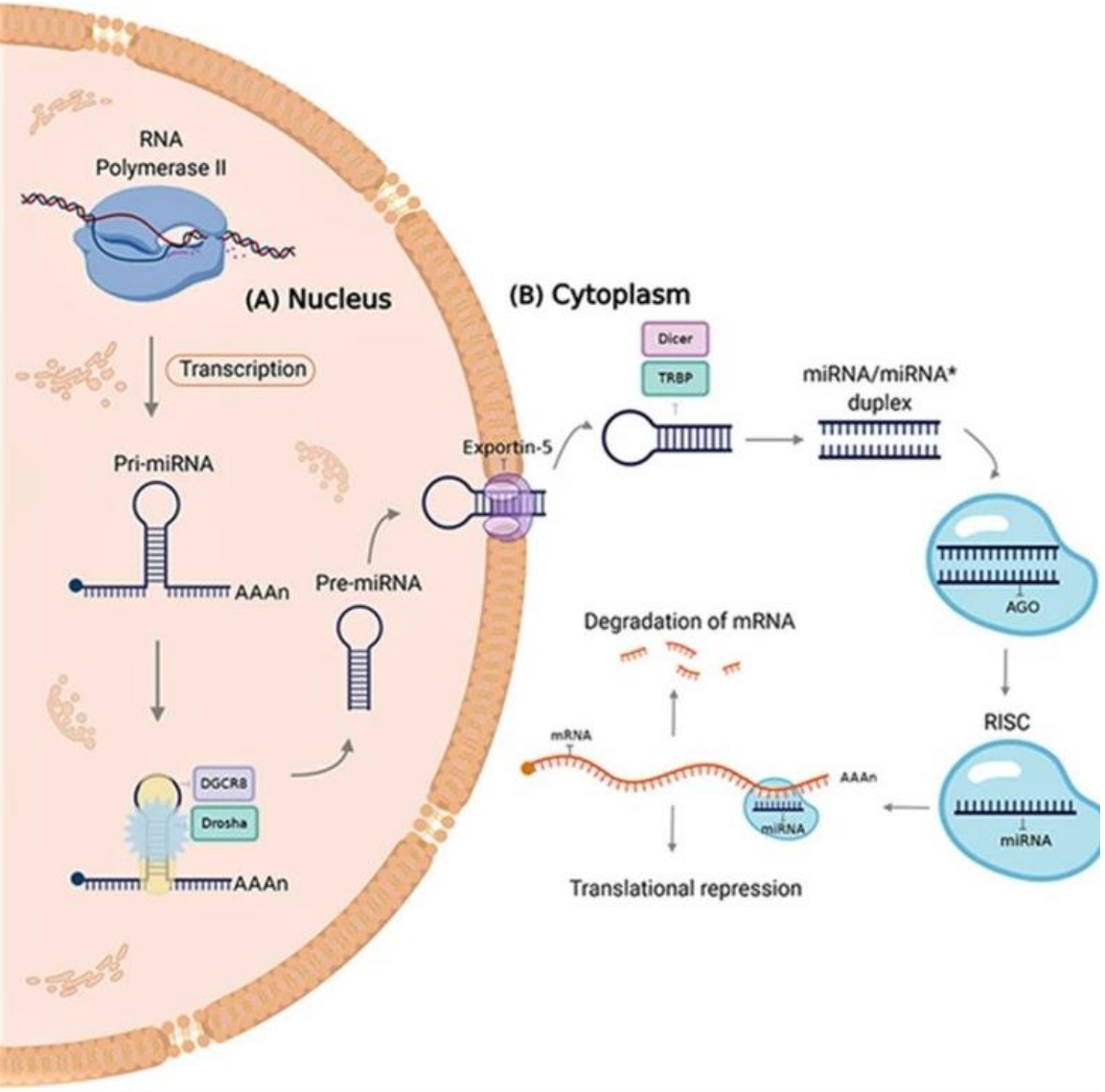
Transcriptome



La Ferlita et al. (2018)

- Most transcripts in the human genome are non-coding
- Non-coding RNA essential for gene regulation at transcriptional and post-transcriptional level
- Heterochromatin formation, histone modification, DNA methylation

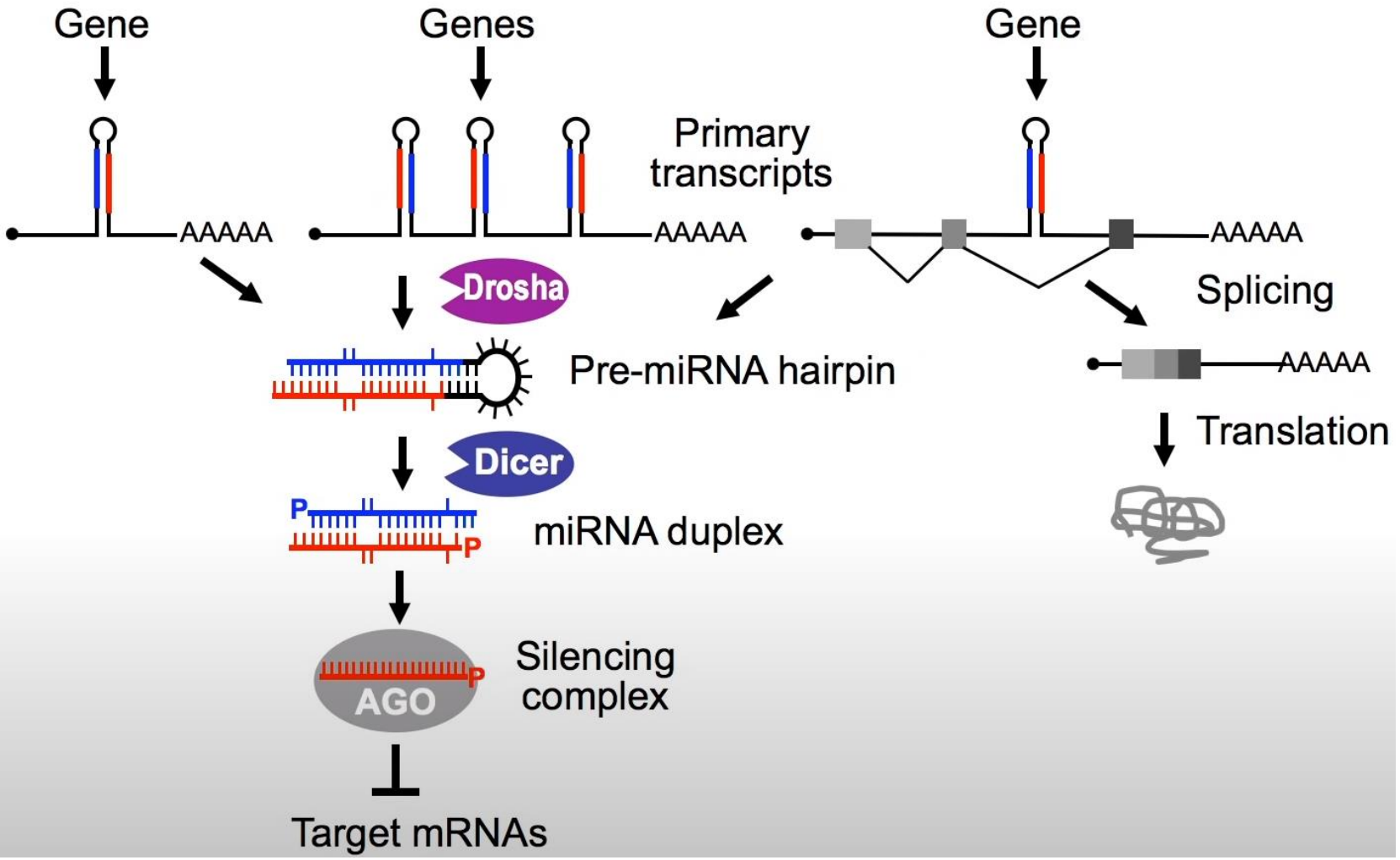
microRNA biogenesis



- 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 degrada tion
- microRNA families often enriched in targets of tr anscription factors (redundant functions)

Paul et al. (2020)

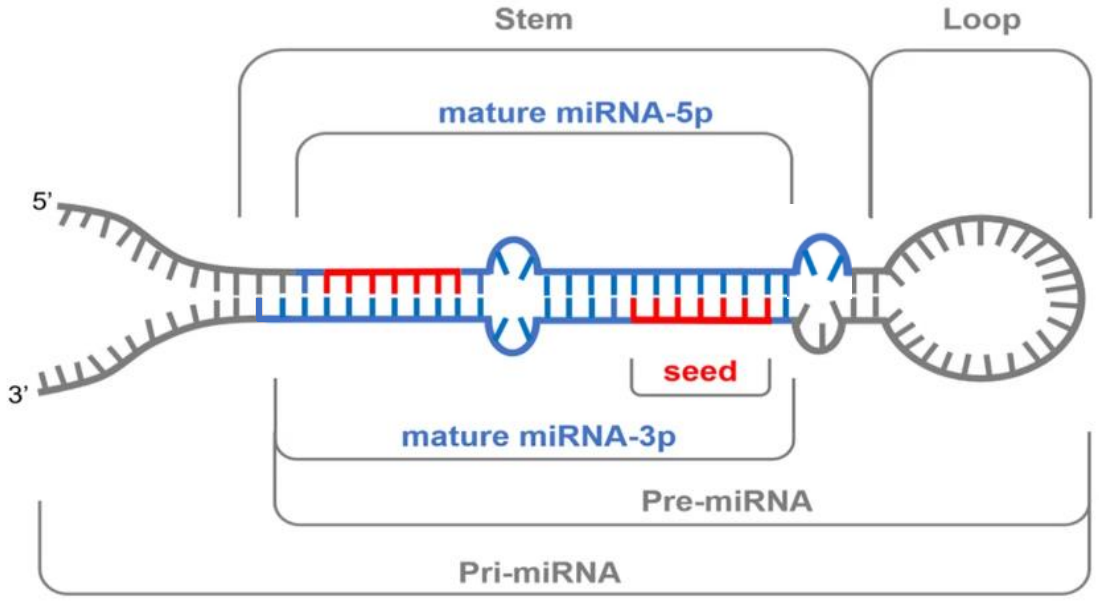
microRNA genes

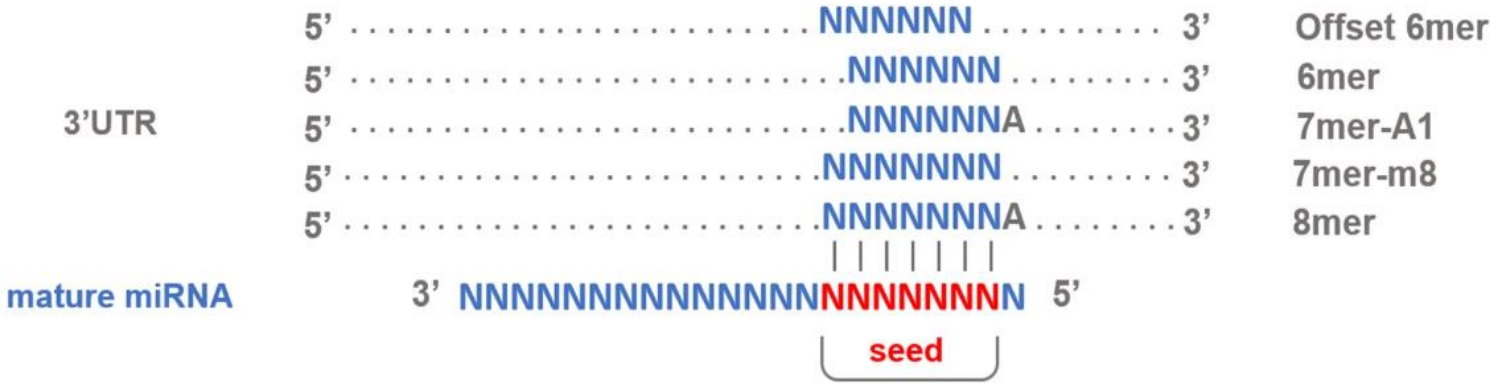


Adapted from Kim et al. (2005)

- 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

microRNA-mRNA





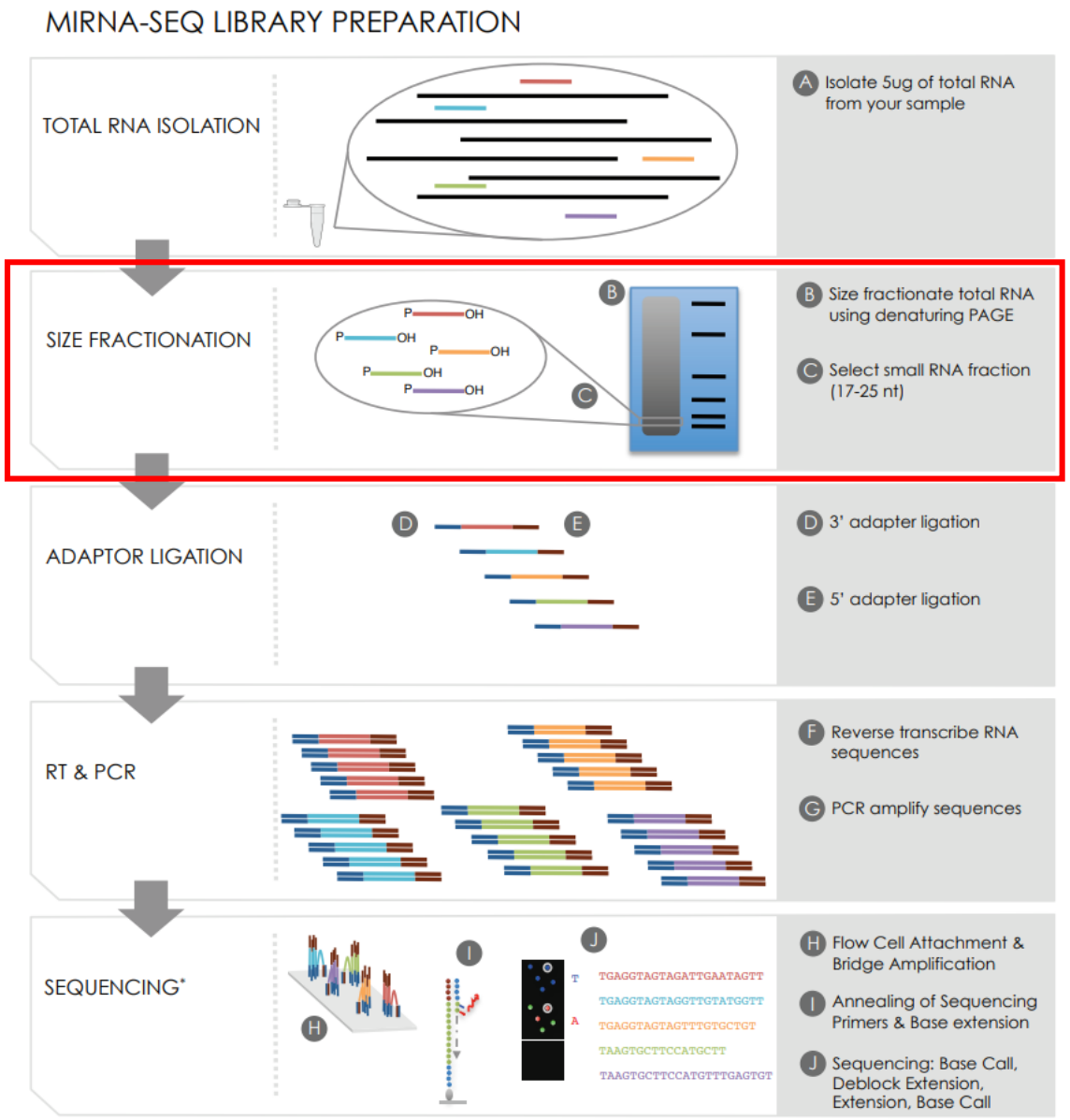
- **Drosha 3' cut leaves a 2 nt overlap and Dicer 5' cut leaves a 2nt overlap**
- Strong propensity for one of both mature strands
- Many microRNA binding sites in 3'UTRs are broadly conserved
- Multiple microRNA binding sites for the same microRNA family
- >60% of protein coding genes are targeted

Marmol-Sanchez et al. (2020)

Experimental microRNA detection

METHOD	*DISADVANTAGES*	*ADVANTAGES*
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 miRNAs	Comparing the relative abundance of specific miRNAs/low cost
Next generation sequencing	Substantial computational work required	Very high sensitivity-High accuracy in distinguishing variants of miRNAs

Experimental microRNA detection



- 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 degradated RNA fragments

Ex.1 Annotate microRNAs

Exercise 1: Find the (hsa-) microRNA sequence using blast

>seq1

ACATTTACCTAGCAGAAGAAAAATCGTGTTTACGAAGGTGGTTTTCGCAGGGCGAAGCTAATTCGTGCAACTTCCCCAAATGTG
GGAAGCTCGACTGCATAATTTGTGGTAGTGGGAGACTGCGTTCGCTCTTTTCCCCCG

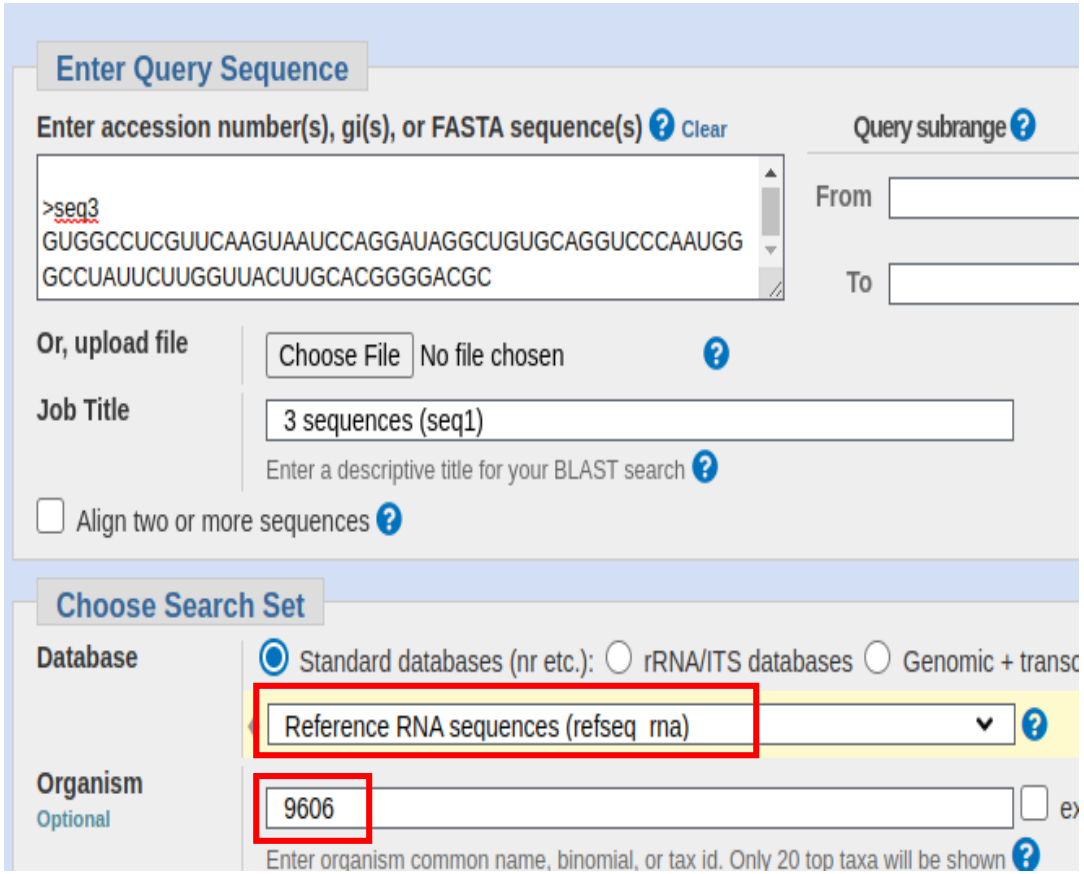
>seq2

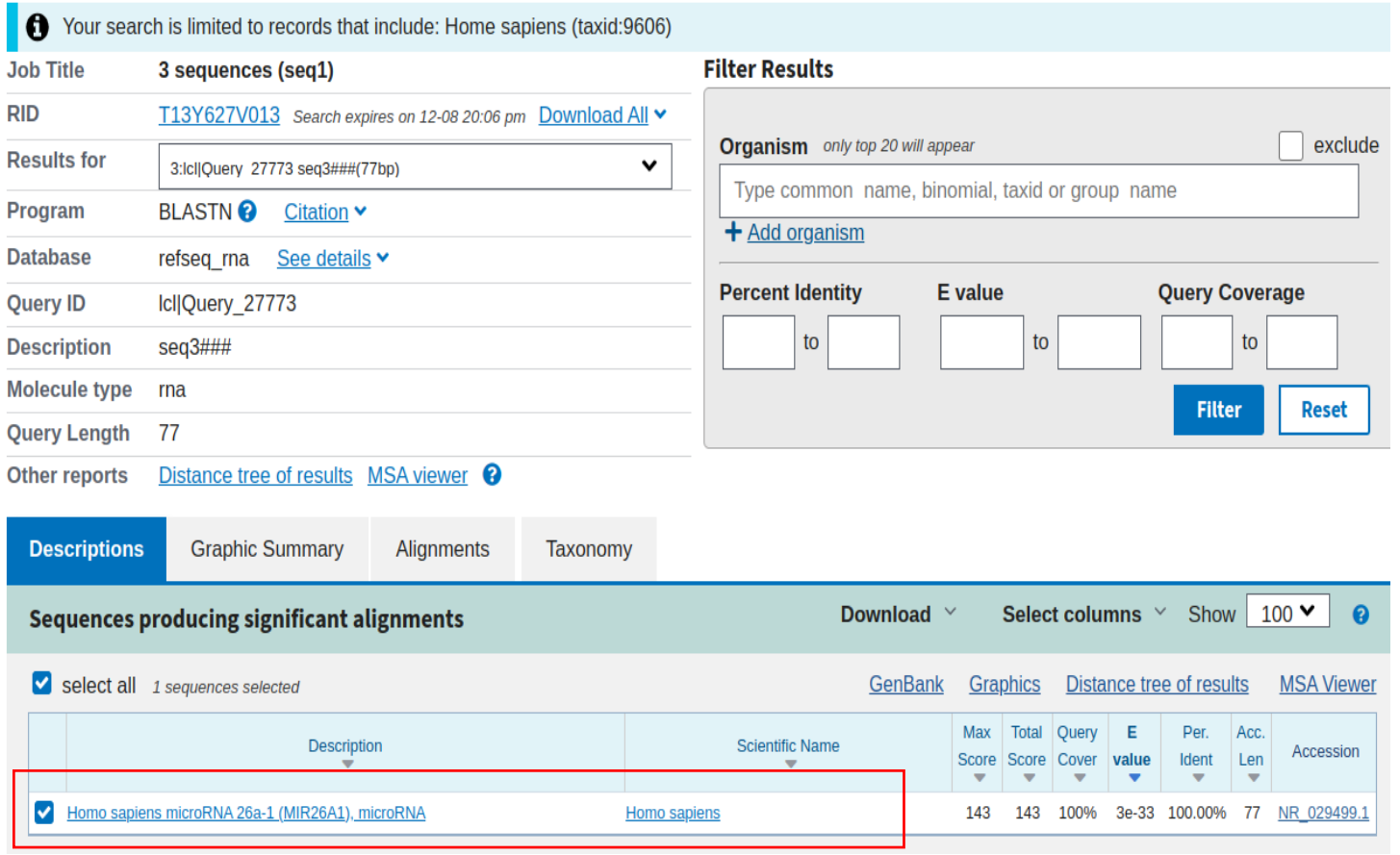
TCCAAACAGACACTGATGGCACCTTCTGCCATTTAGGAATTTGTTTTAAAACAGACATTTGTCTAGATATTTCCTTTGTGGCCTCC
TCCCCATCAAAAGTCAATCAAACATCG

>seq3

GUGGCCUCGUUCAAGUAAUCCAGGAUAGGCUGUGCAGGUCCCAAUGGGCCUAUUCUUGGUUACUUGCACGGGGACGC

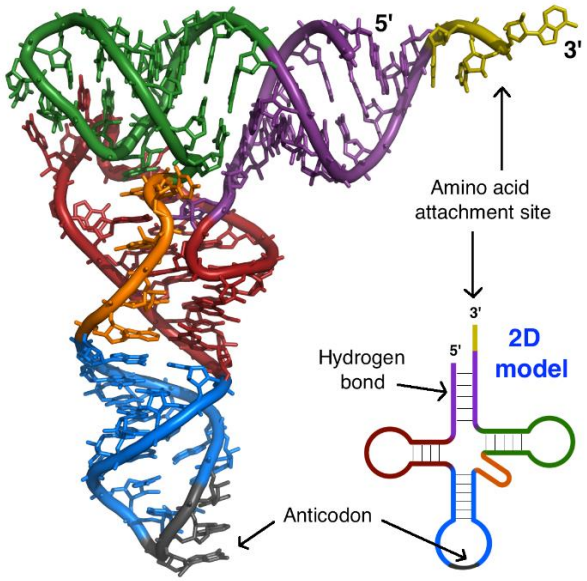
Sol. 1 Annotate microRNAs



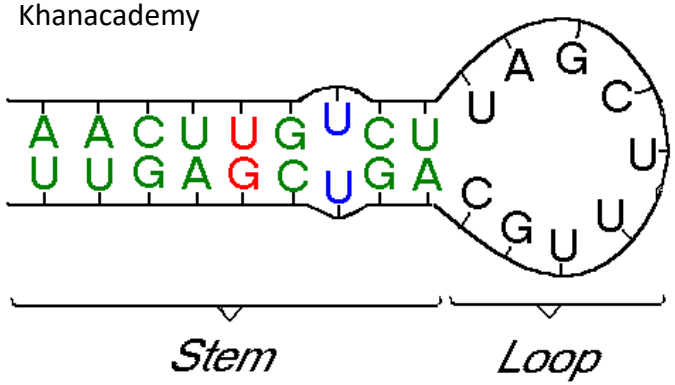


RNA secondary structure

3D-strcture of tRNA



Khanacademy



Watson-Crick pairs
UG pairs
Mismatch

- 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
- Mismatches created wobbles (distinct hydrogen bonds)
- Mismatches can have similar stability

RNA secondary structure

- **RNA folding has a great relationship with minimum free energy (mfe)**

Example:

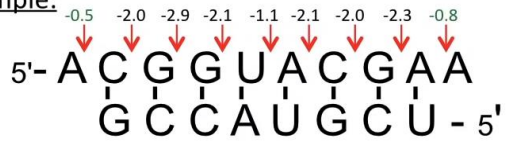
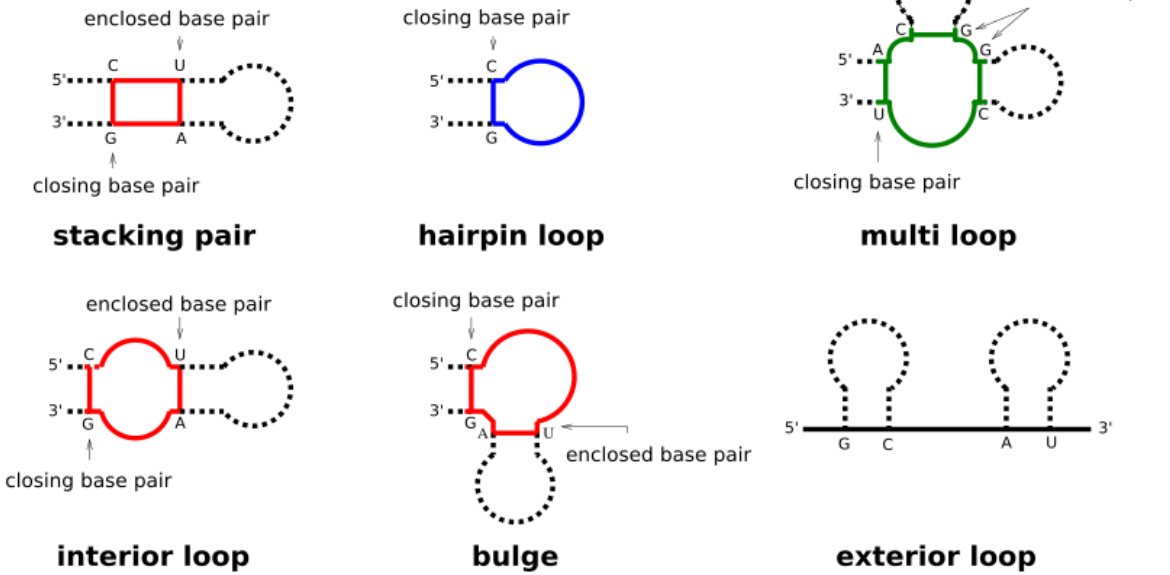
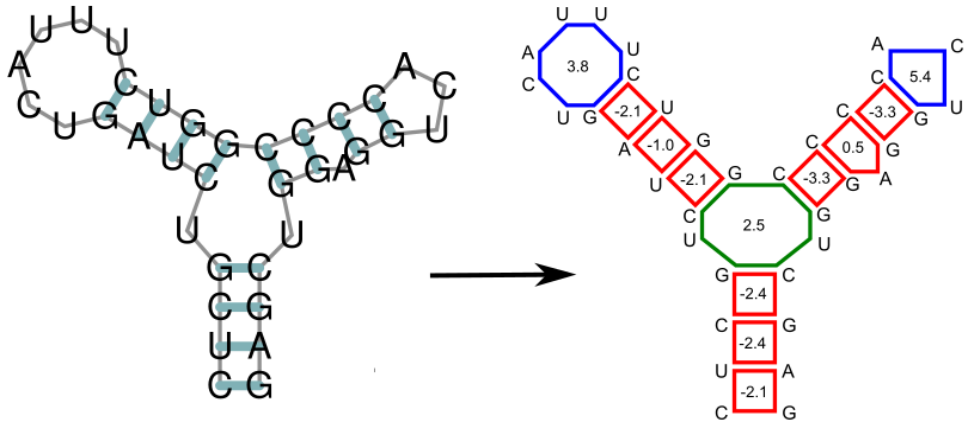


Table 2. Thermodynamic parameters for RNA helix initiation and propagation in 1 M NaCl

Propagation sequence	$\Delta H°$, kcal/mol	$\Delta S°$, eu	$\Delta G°_{37}$, kcal/mol
$\overrightarrow{AA}$ / $\overleftarrow{UU}$	−6.6	−18.4	−0.9
$\overrightarrow{AU}$ / $\overleftarrow{UA}$	−5.7	−15.5	−0.9
$\overrightarrow{UA}$ / $\overleftarrow{AU}$	−8.1	−22.6	−1.1
$\overrightarrow{CA}$ / $\overleftarrow{GU}$	−10.5	−27.8	−1.8
$\overrightarrow{CU}$ / $\overleftarrow{GA}$	−7.6	−19.2	−1.7
$\overrightarrow{GA}$ / $\overleftarrow{CU}$	−13.3	−35.5	−2.3
$\overrightarrow{GU}$ / $\overleftarrow{CA}$	−10.2	−26.2	−2.1
$\overrightarrow{CG}$ / $\overleftarrow{GC}$	−8.0	−19.4	−2.0
$\overrightarrow{GC}$ / $\overleftarrow{CG}$	−14.2	−34.9	−3.4
$\overrightarrow{GG}$ / $\overleftarrow{CC}$	−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)

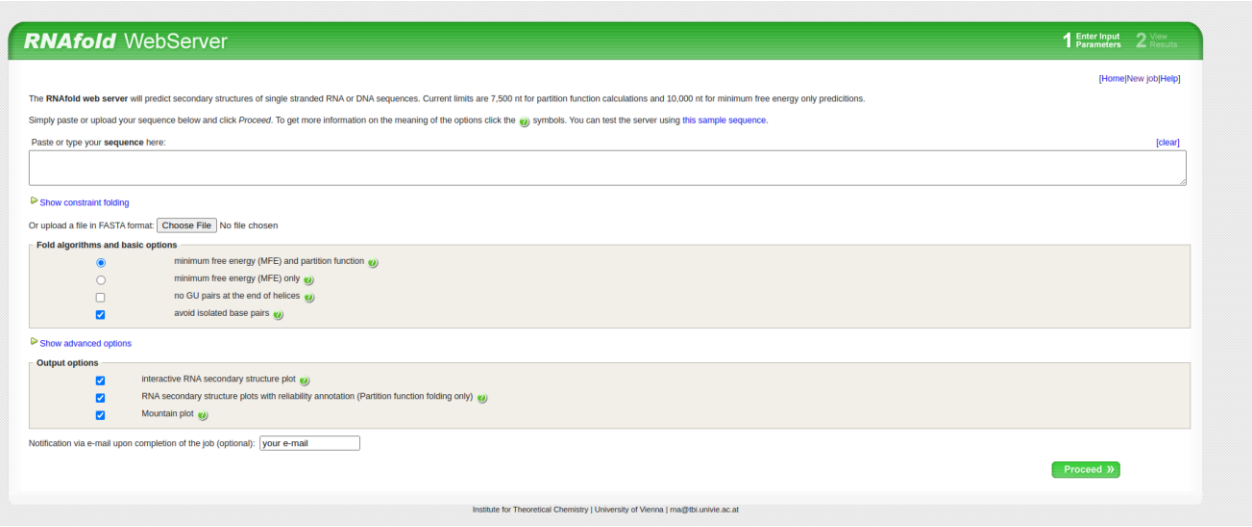


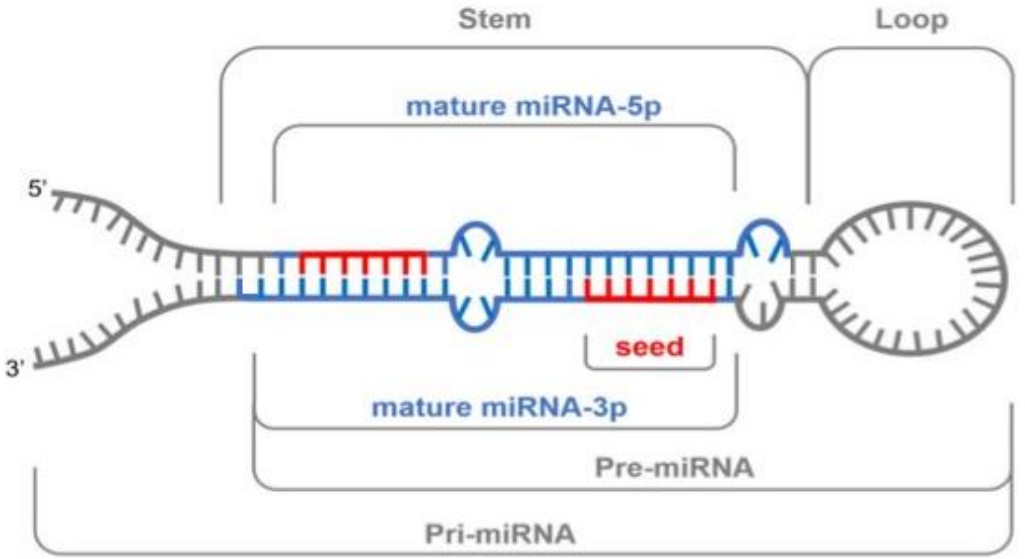


- Turner 1999 RNA parameters
- Mathews 1999 DNA parameters
- Andronescu 2007 RNA parameters
- Mathews 2004 DNA parameters

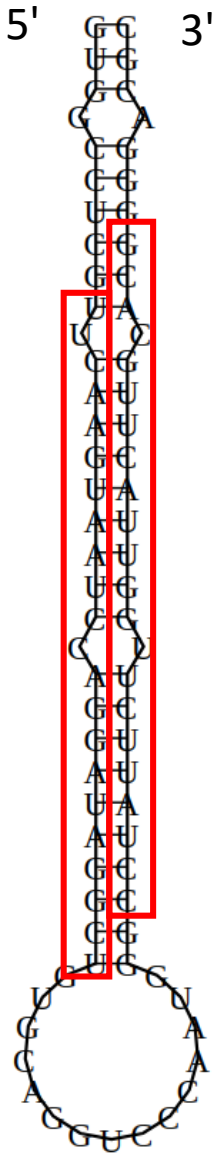
Ex.2 microRNA folding

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





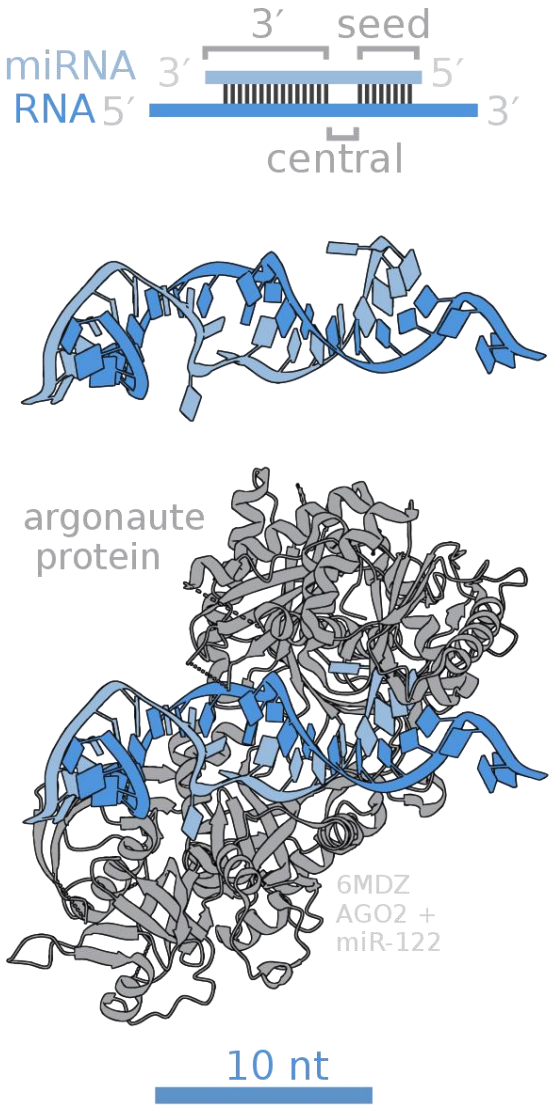
Sol.2 microRNA folding



MFE secondary structure

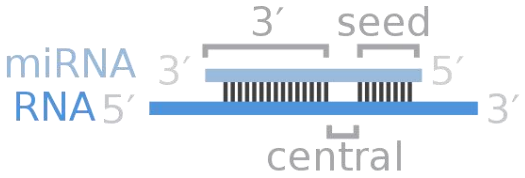
Check further Information on miRBase
Search for hsa-mir-26a-1

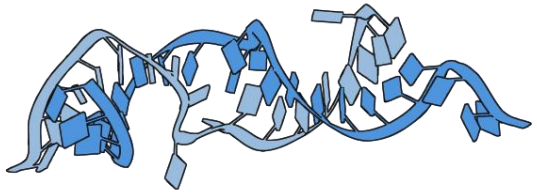
Finding microRNA targets

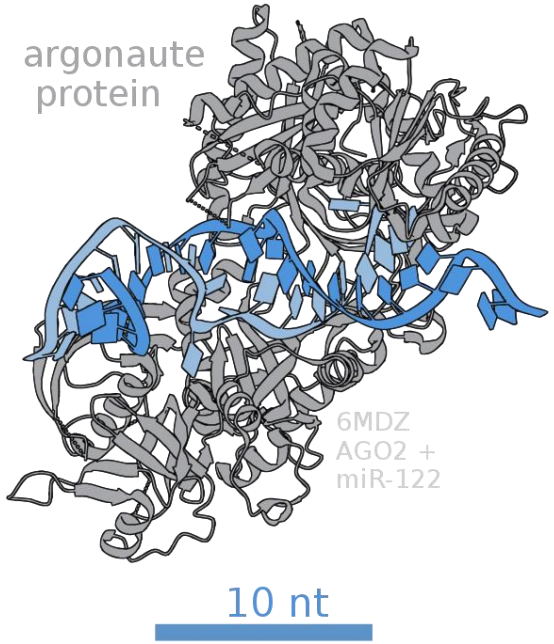


- Mature microRNA is loaded into argonaute protein
- microRNA seed (nt 2-8 from 5' end in mature microRNA) binds to complementary 3'UTR sequence
- 3'UTRs of mRNAs can be really long; on average 800 nt, but up to 10,000 nt
- Searching only complementary sequences leads to many false positives
- Integrate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

Ex. 3 microRNA seed







1. How many human genes are potentially targeted by previously identified mature microRNAs? (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. 3 microRNA seed

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

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_{CT}]

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

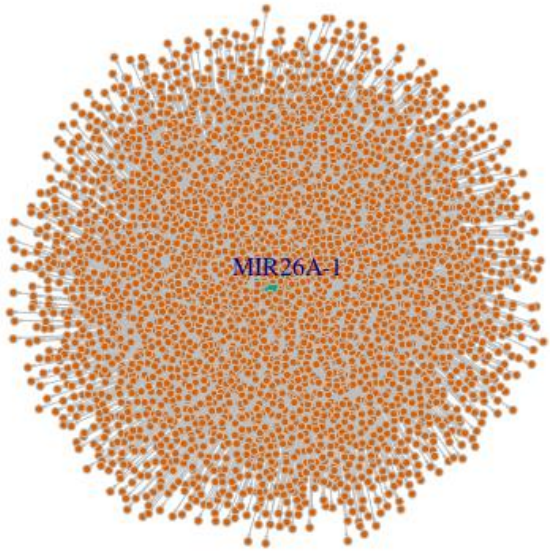
Target gene	Represen-tative transcript ENST0000...	Gene name	Number of 3P-seq tags supporting UTR + 5	Link to sites in UTRs	Conserved sites				Poorly conserved sites				6mer sites	Representative miRNA	Predicted occupancy			Cumulative weighted context++ score	Total context++ score	Aggregate P_{CT}	Previous TargetScan publica-tion(s)
					total	8mer	7mer-m8	7mer-A1	total	8mer	7mer-m8	7mer-A1			mod miRNA	high miRNA	trans-fected miRNA				
		polymerase (RNA) III (DNA directed) polypeptide G																			2009, '11
PTEN	0371953.3	phosphatase and tensin homolog	118	Sites in UTR	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

Position 1261-1268 of PTEN 3' UTR	5' ...ACUGUUAGGGAAUUUUACUUGAA...	8mer
hsa-miR-26a-5p	3' UCGGAUAGGACCUAAUGAACUU	

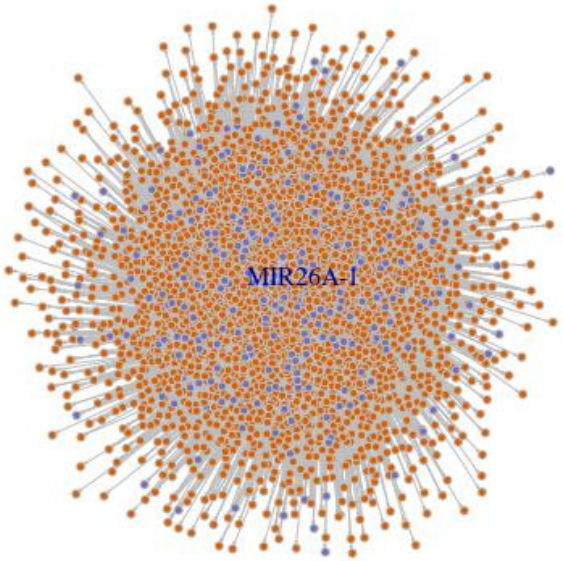
Mature sequence hsa-miR-26a-5p

Accession	MIMAT0000082
Previous IDs	hsa-miR-26a
Sequence	10 - ucaaguaauccaggauaggcu - 31
	Get sequence
Deep sequencing	28542742 reads, 159 experiments

microRNA–mRNA Interactome



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