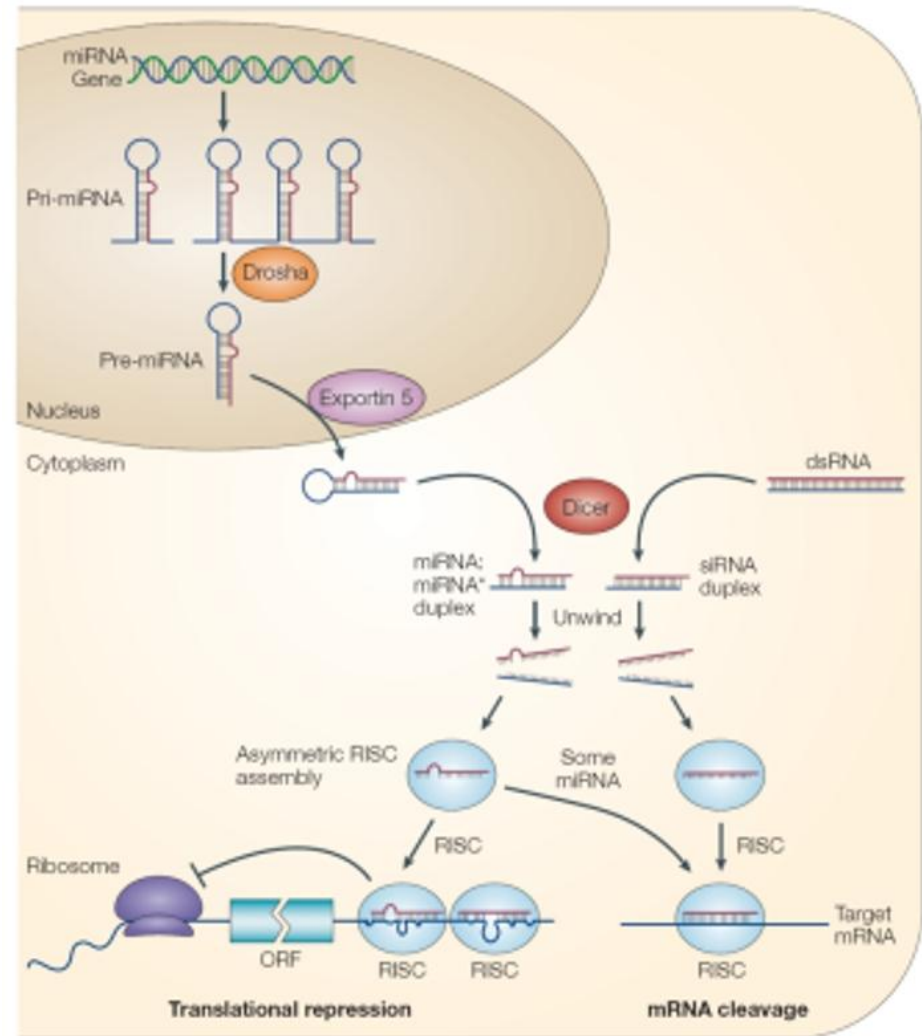


microRNA

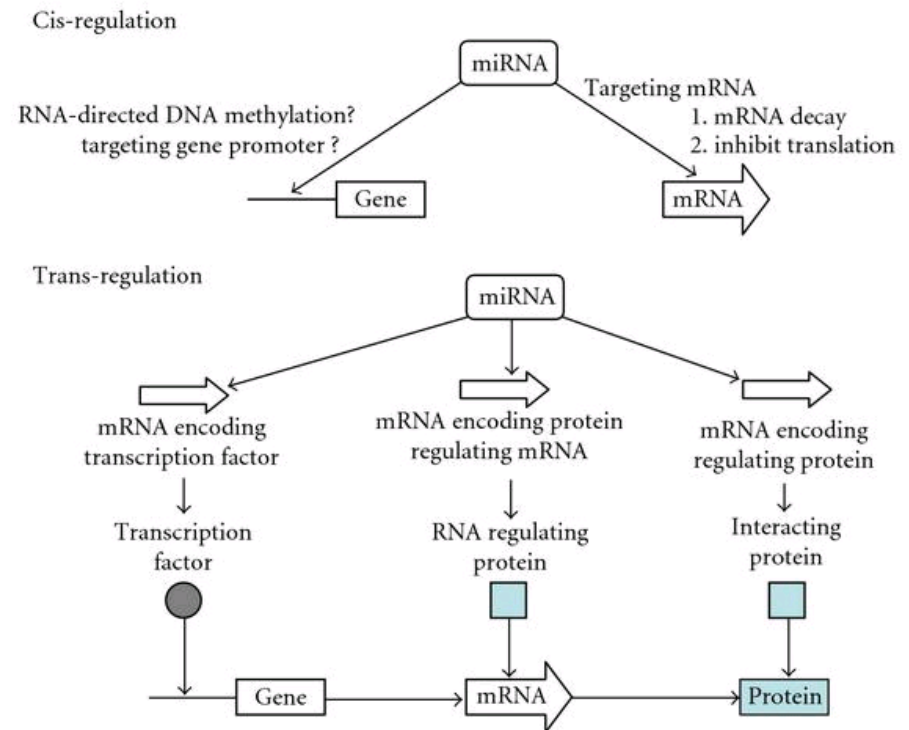
Un microRNA es un ARN monocatenario de entre 19 y 25 nt de longitud. Están implicados en la regulación génica post-transcripcional y probablemente también en la metilación del ADN.

- Los microRNA se transcriben a partir de genes de ADN pero no se traducen a proteína (genes no-codificantes)
- Son presentes en un amplio rango de especies tanto en plantas como en animales.
- Muchos de ellos son altamente conservados
- La mayoría de los genes de microRNA se ubican en regiones intergénicas y tienen su propio promotor y elementos regulatorios
- Aprox. 40% de los genes de microRNA están ubicado en intrones → se transcriben conjuntamente con el gen hospedador.
- Están involucrados en muchos procesos básicos (metabolismo, desarrollo, sistema inmunológico, etc.)
- Algunos microRNA están implicados en el desarrollo de patológicas como el cáncer

- La mayoría se transcriben mediante polimerasa II (algunos mediante pol III) como largos transcritos primarios (pre-microRNA)
- El pri-miRNA se procesa mediante la proteína Drosha → pre-miRNA
- El pre-miRNA se exporta al citoplasma mediante Exportin 5
- Dicer procesa el pre-miRNA en el citoplasma y genera el microRNA maduro
- El microRNA maduro se asocia con el complejo proteico RISC (RNA-induced silencing complex)
- RISC inicia o la inhibición de la traducción o la degradación del mRNA

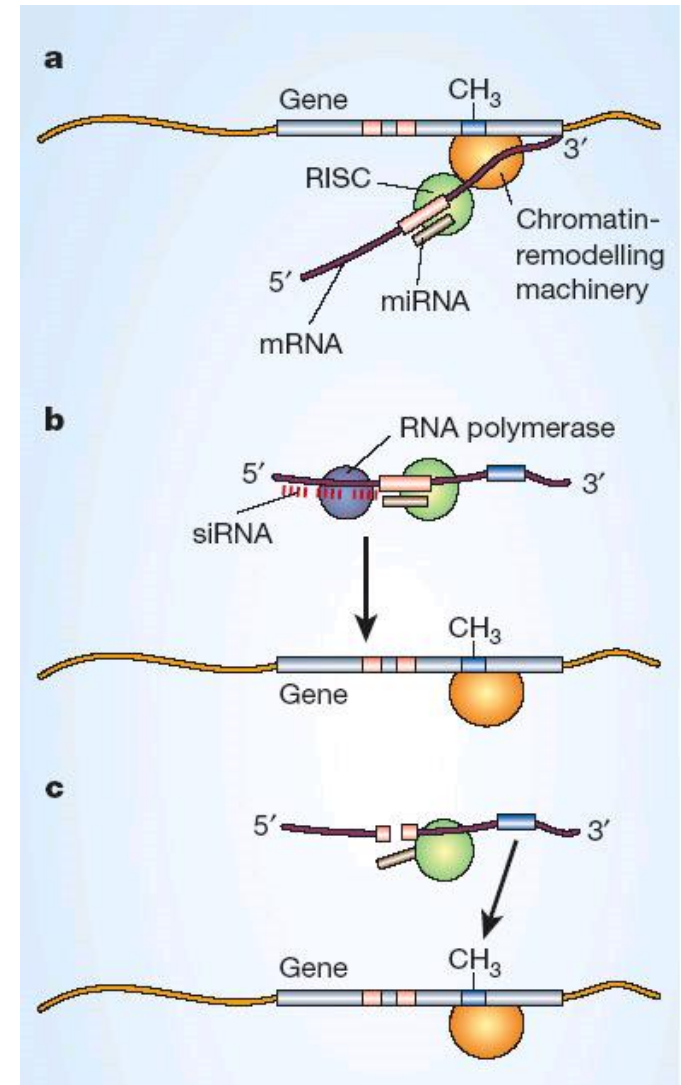


- Se suelen unir a la región 3' UTR donde inician la degradación del mRNA o inhiben la traducción
- Si la función es por inhibición de traducción o por degradación depende posiblemente del grado de complementariedad entre microRNA y 3'UTR
- Hay otras posibles funciones como la metilación del ADN que de momento parece bien establecido en plantas



Bao *et al.*⁴ find that methylation (the addition of CH₃ groups) occurs in the *PHB* and *PHV* genes; they suggest that this must involve microRNAs (miRNAs), as mutations that disrupt the genes' miRNA-binding sites prevent methylation. How might methylation occur? In the proposed models, mature miRNA is first produced from a precursor and exported from the nucleus. a, The mature miRNA might return to the nucleus with an RNA-induced silencing complex (RISC), and base-pair to a matching sequence (red) in the messenger RNA (mRNA) being produced from *PHB/PHV*. The RISC might then, via the mRNA, recruit 'chromatin-remodelling' machinery to the DNA to achieve methylation. b, Small interfering RNAs (siRNAs) might be produced by an RNA-dependent RNA polymerase, directed by miRNAs or by miRNA-induced mRNA cleavage. The siRNAs might then directly or indirectly (via 'spreading') target the gene's blue region for methylation. c, The mRNA could be modified by an miRNA-induced 'mark' (such as cleavage), and then guided to the DNA to induce methylation.

http://www.nature.com/nature/journal/v433/n7025/fig_tab/433472a_F1.html



- miRBase es una base de datos que contiene las secuencias de microRNA, tanto el microRNA maduro como el pre-microRNA y una serie de anotaciones. La base de datos cuenta también con un navegador y motor de búsqueda.
- miRBase también mantiene su propia predicción de dianas de microRNAs llamado microCosm. Pero también asigna las dianas de TargetScan y Pictar

Stem-loop sequence MI0000078	
Accession	MI0000078
ID	hsa-mir-22
Symbol	HGNC:MIR22
Description	Homo sapiens miR-22 stem-loop
Stem-loop	<pre> u cc - a u ccu ggc gag gcag<u>u</u>gu<u>u</u>cu<u>u</u>cag <u>u</u>ggca gcu<u>u</u>ua gu g a ccg cuc cgu<u>u</u>gu<u>u</u>ca<u>u</u>ga<u>u</u>g<u>u</u> accgu cga<u>u</u>au cg c u -c g - - acc </pre> Get sequence
Genome context	<i>Coordinates (GRCh37)</i> 17: 1617197-1617281 [-]
	<i>Overlapping transcripts</i> sense OTTHUMT00000255250 ; C17orf91-001; exon 3 ENST00000397273 ; C17orf91-201; exon 2 ENST00000334146 ; C17orf91-001; exon 3
View flanking features	
Database links	EMBL: AF480525
	EMBL: AJ421742
	HGNC: 31599; MIR22
	ENTREZGENE: 407004; MIR22
Gene family	MIPF0000053; mir-22

- Los microRNA en miRBase suelen tener nombres como hsa-mir-22. Las primeras tres letras indican la especie. Por ejemplo, hsa para humano, mmu para ratón, rno para rata, etc.
- Un nombre de microRNA en minúscula (hsa-mir-22) hace referencia al gen de microRNA o al pre-microRNA. Al microRNA maduro se refiere con 'miR' (hsa-miR-22)
- El número que lleva el nombre del microRNA se asigna de forma secuencial.
- A veces se detectan dos microRNA maduros que provienen del mismo pre-microRNA. En estos casos, al menos frecuente se asigna un asterisco, por ejemplo hsa-miR-19 (microRNA predominante) y hsa-miR-19*
- Si los datos experimentales no permiten determinar que microRNA es la predominante, se asignan nombres como miR-142-5p (del brazo 5') and miR-142-3p (del brazo 3').
- Excepciones a estas reglas se mantienen por motivos históricos en las familias let-7 y lin-4
- Mas información acerca de la nomenclatura se puede encontrar en el siguiente artículo: Victor Ambros, Bonnie Bartel, David P. Bartel, Christopher B. Burge, James C. Carrington, Xuemei Chen, Gideon Dreyfuss, Sean R. Eddy, Sam Griffiths-Jones, Mhairi Marshall, Marjori Matzke, Gary Ruvkun, and Thomas Tuschl. [A uniform system for microRNA annotation.](#) *RNA* 2003 9(3):277-279.

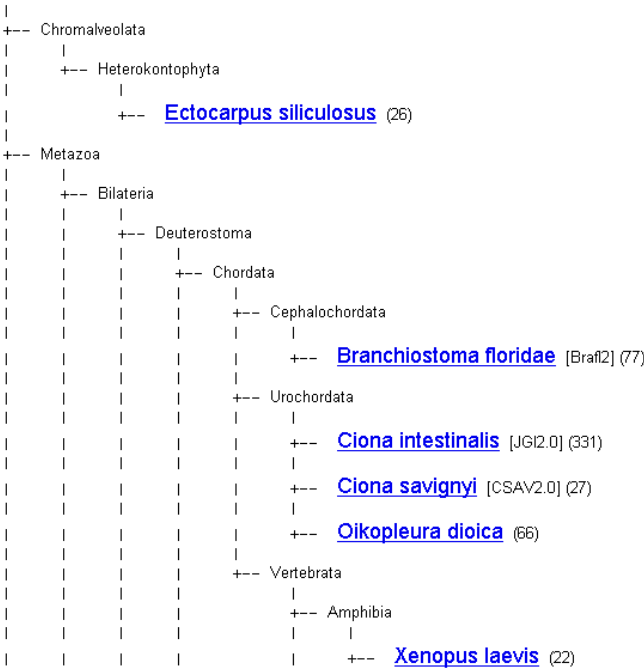
Para extraer datos de una especie

- 1) Seguir el enlace 'Browse'
- 2) Escoger una especie



Browse miRBase by species

Key: Species [assembly version] (miRNA count)

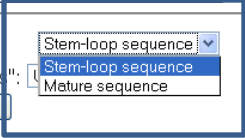


hsa-mir-4314	MI0015846	17	7991374	7991465	+	☐
hsa-mir-4315-1	MI0015844	17	43552729	43552801	-	☐
hsa-mir-4315-2	MI0015983	17	62818148	62818220	-	☐
hsa-mir-4316	MI0015845	17	75393066	75393136	-	☐
hsa-mir-4317	MI0015850	18	6374360	6374424	-	☐
hsa-mir-4318	MI0015847	18	35237098	35237178	+	☐
hsa-mir-4319	MI0015848	18	42550047	42550131	-	☐
hsa-mir-4320	MI0015849	18	47652869	47652933	-	☐
hsa-mir-4321	MI0015852	19	2250638	2250717	+	☐
hsa-mir-4322	MI0015851	19	10341089	10341161	+	☐
hsa-mir-4323	MI0015853	19	42637597	42637665	-	☐
hsa-mir-4324	MI0015854	19	49812054	49812125	-	☐
hsa-mir-4325	MI0015865	20	55896558	55896647	-	☐
hsa-mir-4326	MI0015866	20	61918160	61918218	+	☐
hsa-mir-4327	MI0015867	21	31747612	31747696	-	☐
hsa-mir-4328	MI0015904	X	78156691	78156746	-	☐
hsa-mir-4329	MI0015901	X	112023946	112024016	-	☐
hsa-mir-4330	MI0015902	X	150336694	150336798	+	☐

Get selected sequences:

Select sequences and output type, then click "Fetch Sequence":

Fetch Sequences Select all Reset



- 3) Escoger microRNA maduro o pre-microRNA
- 4) Con 'Fetch Sequences' se obtiene las secuencias

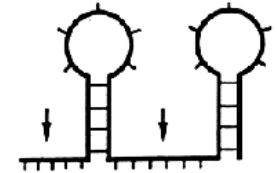
Definiciones

- 1) Duplex: región con complementariedad perfecta – todas las bases se emparejan
- 2) Bulge: en una de las dos hebras hay bases que no tienen pareja
- 3) Bucles internos: en las dos hebras hay bases que no tienen parejas
- 4) Hairpin: la estructura secundaria forma una horquilla

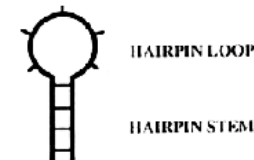
a. DUPLEXES



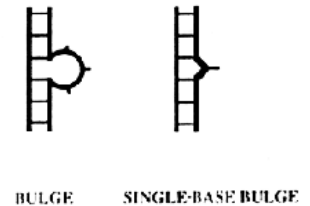
b. SINGLE STRANDED REGIONS



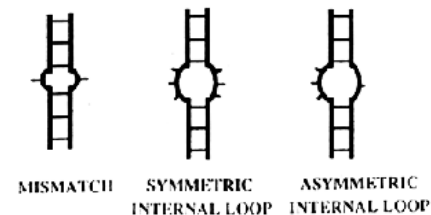
c. HAIRPINS



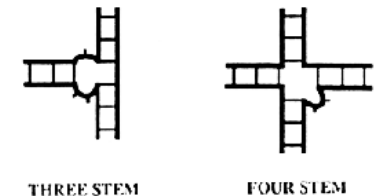
d. BULGES

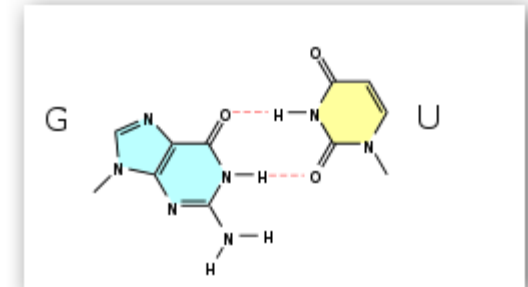
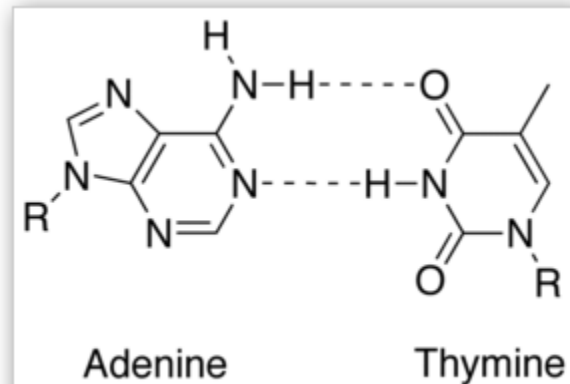
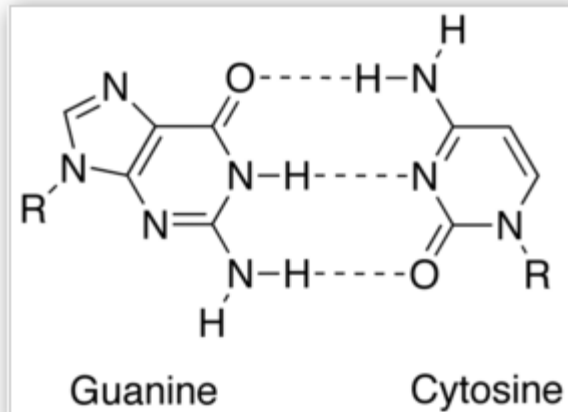


e. INTERNAL LOOPS



f. JUNCTIONS

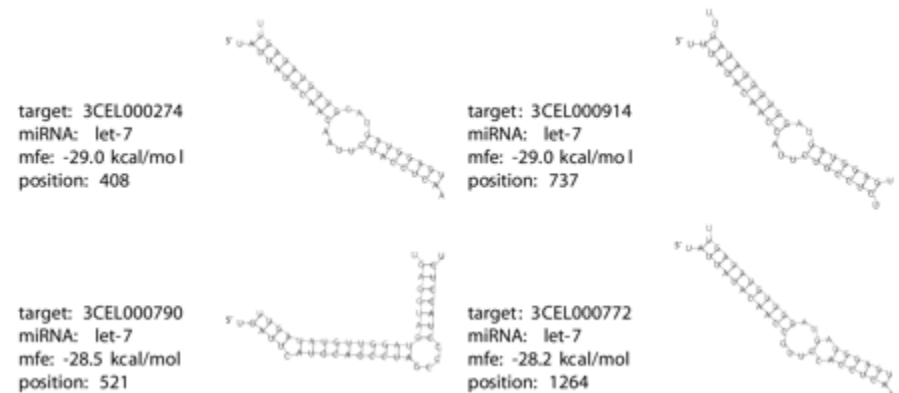
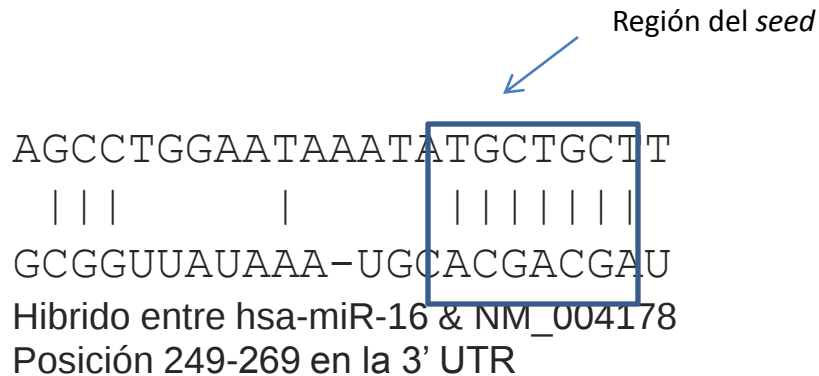




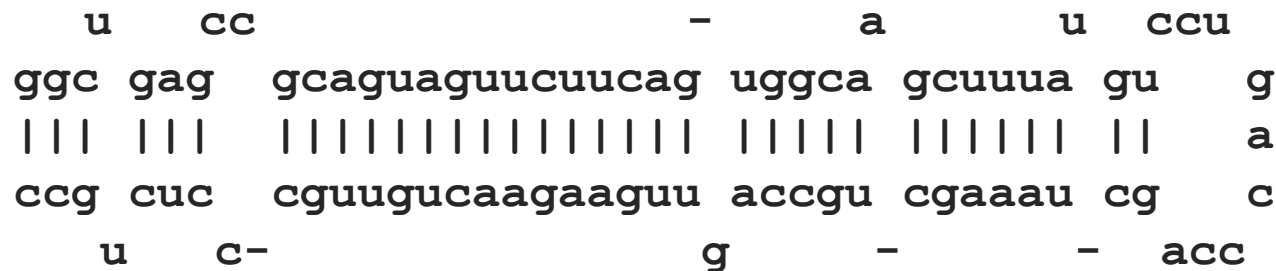
Stacking Energies for base pairs						
	A/U	C/G	G/C	U/A	G/U	U/G
A/U	-0.9	-1.8	-2.3	-1.1	-1.1	-0.8
C/G	-1.7	-2.9	-3.4	-2.3	-2.1	-1.4
G/C	-2.1	-2.0	-2.9	-1.8	-1.9	-1.2
U/A	-0.9	-1.7	-2.1	-0.9	-1.0	-0.5
G/U	-0.5	-1.2	-1.4	-0.8	-0.4	-0.2
U/G	-1.0	-1.9	-2.1	-1.1	-1.5	-0.4

Destabilizing Energies for Loops					
Number of Bases	1	5	10	20	30
Internal	--	5.3	6.6	7.0	7.4
Bulge	3.9	4.8	5.5	6.3	6.7
Hairpin	--	4.4	5.3	6.1	6.5

- La predicción se basa en los híbridos que se forman entre el microRNA y el mRNA
- Los híbridos se calculan mediante modelos termodinámicos – *minimum free energy algorithm* (RNAfold, RNAhybrid, etc.)
- Alunas propiedades importantes: energía libre, existencia de una región semilla (*seed region*), numero de desemparejamientos, numero de bucles
- Otras propiedades importantes pero menos entendidos son la “accesibilidad” de la estructura secundaria y la interacción entre varias dianas en la misma 3’ UTR
- La predicción de dianas de microRNA se basa fuertemente en la presencia de un “seed” (emparejamiento perfecto de los 7 primeros nucleótidos entre el extremo 5’ del microRNA y la región 3’ UTR) y la señal filogenética
- Se estima que aprox. El 40% de todas las dianas no tienen “seed” → tienen regiones compensatorias , es decir muchos emparejamientos entre el extremo 3’ del microRNA y la UTR 3’
- Algunos de los algoritmos mas usados son: TargetScanS, PicTar, miRanda, RNAhybrid, TargetSpy



- La predicción de los genes de microRNA se basa en las propiedades de la secuencia y de la estructura secundaria
- La estructura secundaria de los microRNA forma siempre una “horquilla” (hairpin structure)
- Algunas propiedades importantes son: presencia de una horquilla, longitud de la secuencia, numero de enlaces, suma de energía de enlace, tamaño de bucles, etc.
- Se suele deslizar una ventana a lo largo de la secuencia objeto comprobando en cada punto si existe una horquilla → los candidatos que forman una horquilla se analiza mas detenidamente mediante modelos de aprendizaje automatizado
- Para reducir el número de falsos positivos se emplea frecuentemente la señal filogenética



Con los parámetros por defecto, el programa predice 42 genes de microRNA en una secuencia de aprox. 100 kb



An miRNA scanning Tool by:
Center for Computational Biology and Bioinformatics
Jawaharlal Nehru University
New Delhi 110067
E-mails: ajlok0200@mail.jnu.ac.in, st0645@students.jnu.ac.in
Telephone: +91-11-26717630

Total precursor miRNAs found: **42**.

[FASTA format \(sequences only\)](#) [FASTA format \(structures only\)](#) [FASTA format \(both sequences and structures\)](#)

[illegible]

La secuenciación masiva brinda nuevas posibilidades a la predicción de microRNAs ya que reduce drásticamente el número de candidatos (secuencias transcritas)

Introducción

Existen protocolos para medir la expresión de RNA corto. Para ello, se extrae primero el RNA total de las células. Mediante electroforesis en gel se puede extraer la banda que corresponde a longitudes entre 17 y 30 nt. Esta banda se amplifica y se secuencia mediante secuenciación masiva.



La salida de los secuenciadores se suele dar en formato fastq

```
@SRR037876.8543926 GSM522374_1:1:148:931:861
TAGTTCTACAGTCCGACGATCTCGTATGCCGTCTTC
+
BB@+?0:4@B@-@/A<3A7@-=@<1=@87=?<==9#
@SRR037876.8543927 GSM522374_1:1:148:931:517
AGCTACATTGTCTGCTGGGTTTCTCGTATGCCGTCT
+
BBC@3<1=872661.@C;@A93+?;.:2.?386<;>
@SRR037876.8543928 GSM522374_1:1:148:931:648
TGAGGTAGTAGGTTGTGTGGTTAATCGTATGCCGCT
+
5@059)@6?':9>0<@@)@=BA8)99@3258?####
@SRR037876.8543929 GSM522374_1:1:148:931:770
GCTACATTGTCTGCTGGGTTTCTCGTATGCCGTCTT
+
```

Identificador

Secuencia del 'read'

Calidad del 'read'

La calidad

- Los caracteres se puede convertir en un score (Q, Phred score)
- $Q = \text{ASCII}(\text{carácter}) - 64$

Interpretación del Phred score

Phred quality scores are logarithmically linked to error probabilities

Phred Quality Score	Probability of incorrect base call	Base call accuracy
10	1 in 10	90 %
20	1 in 100	99 %
30	1 in 1000	99.9 %
40	1 in 10000	99.99 %
50	1 in 100000	99.999 %

El programa tiene las siguientes características:

- Detecta todos los microRNAs conocidos
- Mapea los reads a librerías de otros elementos transcritos (RFAM, mRNA, etc)
- Predice nuevos microRNAs

Para poder usar miRanalyzer tenemos que convertir el formato fastq en otro llamado RC (read-count)

sequence	count
GCTATGACGGTTACACTCTCCGGTCG	2.0
TAGGTCAAGGTGTAGCCCATGAGGTG	14.0
AAAGGGATTTTTGGAGCAGGGAGATG	2.0
GGCTGCCTGCGGATGAAGTCGTATGG	1.0

Para convertir fastq a formato RC podemos usar un script en Perl
<http://bioinfo2.ugr.es/miRanalyzer/DB/groupReads.zip>

Lanzar el programa sin parámetros produce la salida del imagen: perl groupReadsV2.pl

Solo hay dos parámetros obligatorios: el nombre del fichero de entrada (fastq) y el nombre de salida

```
perl groupReadsV2.pl input=fastq_1.fastq
output=fastq_1.rc
```

```
#####
### All input options are given in this format: option=value ###
### The following options do exist: ###
### Mandatory input parameter: ###
### input=file1,file2,file3,... ###
### Input Files: The script accepts fastq format ###
### output=name of the output file ###
### Optional input parameter: ###
### minReadCount=numeric value: the minimum read count a read must ###
### must have to be written out (default 1) ###
### Quality values: Two different ways to control the quality ###
### of the reads are implemented: 1) all bases in a read must ###
### be better than a given threshold (Phred like Illumina scores ###
### --> value=10 --> one out of ten bases is incorrect, value=20 ###
### --> one out of hundred bases is incorrect etc ###
### or 2) the mean quality score must be above a given value ###
### The value range might be between 0 and 30 ###
### the quality is set like this: ###
### qual=min:value OR ###
### qual=mean:value ###
### (default: qual=min:0) ###
### phred=solid: the default is to rest 64 from the ASCII quality ###
### character (Illumina). Just set this option if the ###
### if the input is from solid. ###
### scarf=true: set this option if the input is in SCARF format ###
### maxLen=value: the maximal length of the reads. If set to ###
### higher values --> more unique reads will be ###
### generated (default: 26) ###
### eachLane=true: set this option if multiple input files are ###
### used. The option forces that a unique read ###
### must be present in all samples (default: false) ###
#####
```

Queing & Execution

Analysis completed

You can bookmark [this page](#)

Download all results in plain text [here](#)

Parameters

Species:	hsa	Assembly:	hg18
Input:	GSM274171-9598_hsa.t...	Mismatches (known):	1
Mismatches (libraries):	1	Mismatches (genome):	1
Score threshold:	0.9	Min. positives:	3
Type	Full analysis	Solid	no

Brief Summary

unique reads:	351607	read count:	834616
filtered unique reads:	40888	filtered read count:	57197
No. known microRNA:	287	No. known microRNA*:	49
No. microRNA* (not miRBase):	22	No. new microRNAs:	166
unique reads (after known):	294759	read count (after known):	647382
unique reads (after lib):	287050	read count (after lib):	635935
unique reads matched:	249179	read count matched:	517838
unique reads not-matched:	37871	read count not-matched:	118097

Mapping to known microRNA (miRBase 16)

Library/ Parameters	mature	ambiguous mature	mature-star	ambiguous mature-star	unobs. mature-star	ambiguos mature-star	unobs. mature-star	hairpin	ambiguous hairpin
No. microRNA	267	61	49	1	22	0	170	28	
fraction (number) of known microRNAs	25.9% (1032)	---	0.0% (191)	---	2.6% (833)	---	16.2% (1048)	---	
unique reads	13092	1074	283	2	56	0	1332	125	
fraction of unique reads	4.2%	0.346%	0.091%	0.001%	0.018%	0.000%	0.429%	0.040%	
read count	123819	2382	633	2	85	0	2818	300	
fraction of read count	15.9%	0.306%	0.081%	0.000%	0.011%	0.000%	0.362%	0.039%	
links to detail pages	details	details	details	details	details	no results	details	details	

Alignment to other transcribed entities

Library/ Parameters	RefSeq_genes	Rfam
number of unique reads	39023	36855
fraction of unique reads	12.56%	11.86%
number of unique reads	79925	40782
fraction of unique reads	10.28%	5.25%
Links	details	details

Predicted candidate microRNAs

No. of read clusters:	310110
No. of checked candidates:	56849
No. new microRNAs:	166 Unique reads (read count): 1135 (5134) details
No. new microRNAs (trans filtered):	127 Unique reads (read count): 915 (4785) details

Predicted microRNAs

[Go Back](#)

ID	chr	chrom start	chrom end	strand	unique reads	read count	norm. read count	no. pos. pred.	put. mature
Candidate_1	chr20	19830061	19830169	-	2	2.0	0.000	3	ATGACAAGGTCGGGTGTTCCGAC
Candidate_2	chr20	47368406	47368538	+	2	2.0	0.000	3	GCGCGGCGCGGGCGCGGGAC
Candidate_3	chr20	43997077	43997199	+	1	1.0	0.000	3	CGCGCGGGGCGGGGGGCG
Candidate_4	chr20	53328120	53328258	-	1	1.0	0.000	4	ACTCACAAAAACAACAAAT
Candidate_5	chr10	72400192	72400326	+	3	3.0	0.000	3	TGAGGAAGGAGGCTGCATGG
Candidate_6	chr10	125058885	125059023	+	1	1.0	0.000	3	ACACAACAACAACAAAAAC
Candidate_7	chr10	2279757	2279881	+	1	1.0	0.000	4	GGCTTTGGGCAAGGGGTGTG
Candidate_8	chr11	84872788	84872904	+	11	14.0	0.002	5	GCCCGGAGGATTCAACCCGGTGGCGG
Candidate_9	chr11	75624547	75624653	-	5	9.0	0.001	5	AGATGGCTCGTTGGTCTAGGGGTATG
Candidate_10	chr11	62189428	62189574	+	1	1.0	0.000	5	ACCTCCTTTTCTTGGCGGGATCGG
Candidate_11	chr11	77585758	77585892	+	1	1.0	0.000	4	CGCAGGGAGGAGCACACTGG
Candidate_12	chr11	12451557	12451683	+	2	2.0	0.000	5	TCTTTTCTGAGAGGAGGGAGA
Candidate_13	chr11	63710201	63710343	+	1	1.0	0.000	5	GCGGCGCGTGC GTTGGGAACGCG
Candidate_14	chr21	45675090	45675200	-	1	1.0	0.000	3	GCTTTGGTGGACGAGTGC
Candidate_15	chr12	30394503	30394607	+	1	1.0	0.000	3	GGGGGGGGGCGGGGGGGG
Candidate_16	chr12	51896839	51896939	+	2	2.0	0.000	5	GCGCGCGCGCGCGCGCTGTGGT
Candidate_17	chr12	99103253	99103349	-	4	5.0	0.001	4	GGAGGAAGGAGGGTGTGTGGC
Candidate_18	chr12	51896841	51896939	+	9	11.0	0.001	5	CGCGCGCGCGCGCGCTGTGGT
Candidate_19	chr12	110692003	110692137	+	12	16.0	0.002	4	GGTAAATGGCTGGGTGCAG
Candidate_20	chr12	24778344	24778442	-	24	346.0	0.045	3	TAGAATGGGTGTGATAGGTGGG
Candidate_21	chr12	6925628	6925762	+	1	1.0	0.000	3	TGAGGAAGGAGGCTGTTAGG
Candidate_22	chr12	49444442	49444576	+	1	1.0	0.000	3	GGGGGAGGGGCGCAGCGGCA
Candidate_23	chr22	38044956	38045104	-	5	13.0	0.002	3	CACAGATGATGAATTTATGACGGGCG
Candidate_24	chr22	46058669	46058783	+	1	1.0	0.000	4	GTGAGGCTGTGGACTGGATG
Candidate_25	chr13	40928035	40928167	-	2	2.0	0.000	3	CCCCCTGGTGGTCTAGTGCTTAGGA
Candidate_26	chr13	113606227	113606363	-	1	1.0	0.000	4	CGGGCCGGGGTTAGCTCCAGC
Candidate_27	chr13	113606318	113606434	-	1	1.0	0.000	4	CGGGCCGGGGTTAGCTCCAGC
Candidate_28	chr13	98986575	98986661	-	21	1729.0	0.222	5	GGCTGGTCCGAGTGAGTGGTGTTA
Candidate_29	chr13	113606266	113606402	-	1	1.0	0.000	4	CGGGCCGGGGTTAGCTCCAGC
Candidate_30	chr13	96665612	96665732	-	7	7.0	0.001	3	TGAGCTAGTAGGTCGCACAGTGT
Candidate_31	chr5	88016628	88016750	-	1	1.0	0.000	3	GGGGGCGTGCACCGCAGCTGGGT
Candidate_32	chr5	51849426	51849532	-	4	6.0	0.001	4	GGGGCAACTGTCCCAAGTGGC
Candidate_33	chr5	140275799	140275871	-	8	8.0	0.001	3	TCCCTGTGGACCAAGTGGTA
Candidate_34	chr5	166338495	166338631	+	4	5.0	0.001	3	CGCGCGCGCTGTGTTGTGCG
Candidate_35	chr5	72778607	72778737	-	1	1.0	0.000	4	GGCCCGGCGGCGCGCTC
Candidate_36	chr5	364427	364547	+	2	2.0	0.000	5	TCCAGGGCTGCATCGTCTCGAG
Candidate_37	chr5	132308918	132309014	+	21	37.0	0.005	3	GGCTGGTTCGAAGGTAGTCAGTTATC
Candidate_38	chr5	162848765	162848857	+	1	1.0	0.000	4	GGAGGGGGAGGCTCAGGGGG
Candidate_39	chr5	3541263	3541383	-	1	1.0	0.000	4	GCCCTGGGGCAGACCGTG
Candidate_40	chr5	167525201	167525301	-	2	2.0	0.000	3	GGGGGGGGGGGGTGGGGGGGGG
Candidate_41	chr5	56822222	56822358	+	1	1.0	0.000	5	TCAGAAGATTGCAGGGCAGGA
Candidate_42	chr6	79844360	79844474	+	10	10.0	0.001	3	TGAGCGGTGCGGCGGGGGGGCGG
Candidate_43	chr6	134533794	134533926	-	1	1.0	0.000	3	AGGACAGGACAAAGCGGCTCGGG