microRNAs y secuenciación masiva

Biocomputación
Grado en Bioquímica

Regulación de la expresión génica

En eucariotas, la expresión génica se encuentra regulada en cualquiera de las etapas que llevan del DNA (gen) a la proteína

- Se modifica la estructura de la cromatina para iniciar la transcripción
- El inicio y los niveles de transcripción se regulan mediante factores de transcripción que se unen a elementos reguladoras en *cis*
- *Splicing*, (adición del *cap* 5') y cola poli(A)
- Proceso del transporte al citoplasma
- Degradación del mRNA: mayor rango de vida media en eucariotas (en procariotas se degradan rápidamente)
- **Regulación post-transcripcional (microRNAs)**
- Se regula la iniciación de la traducción
- Modificaciones postraduccionales pueden cambiar la conformación o llevar a la degradación

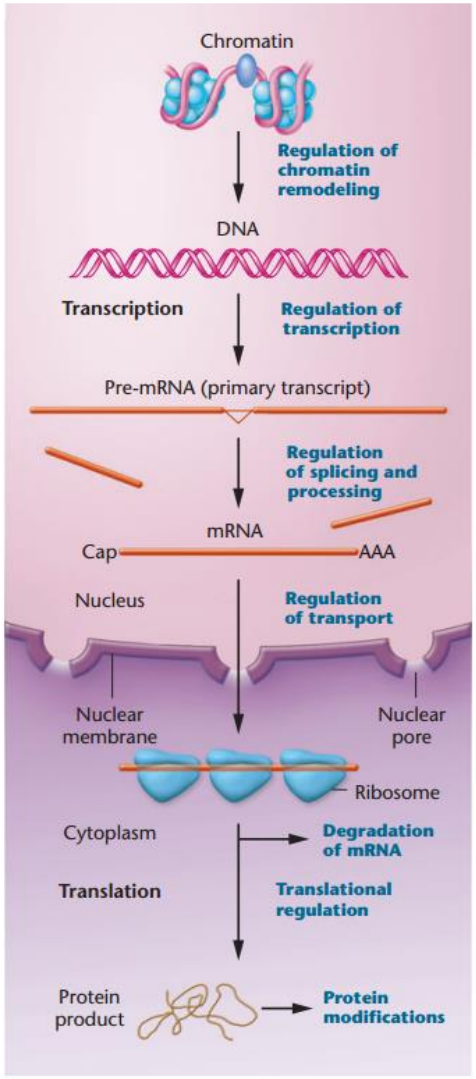
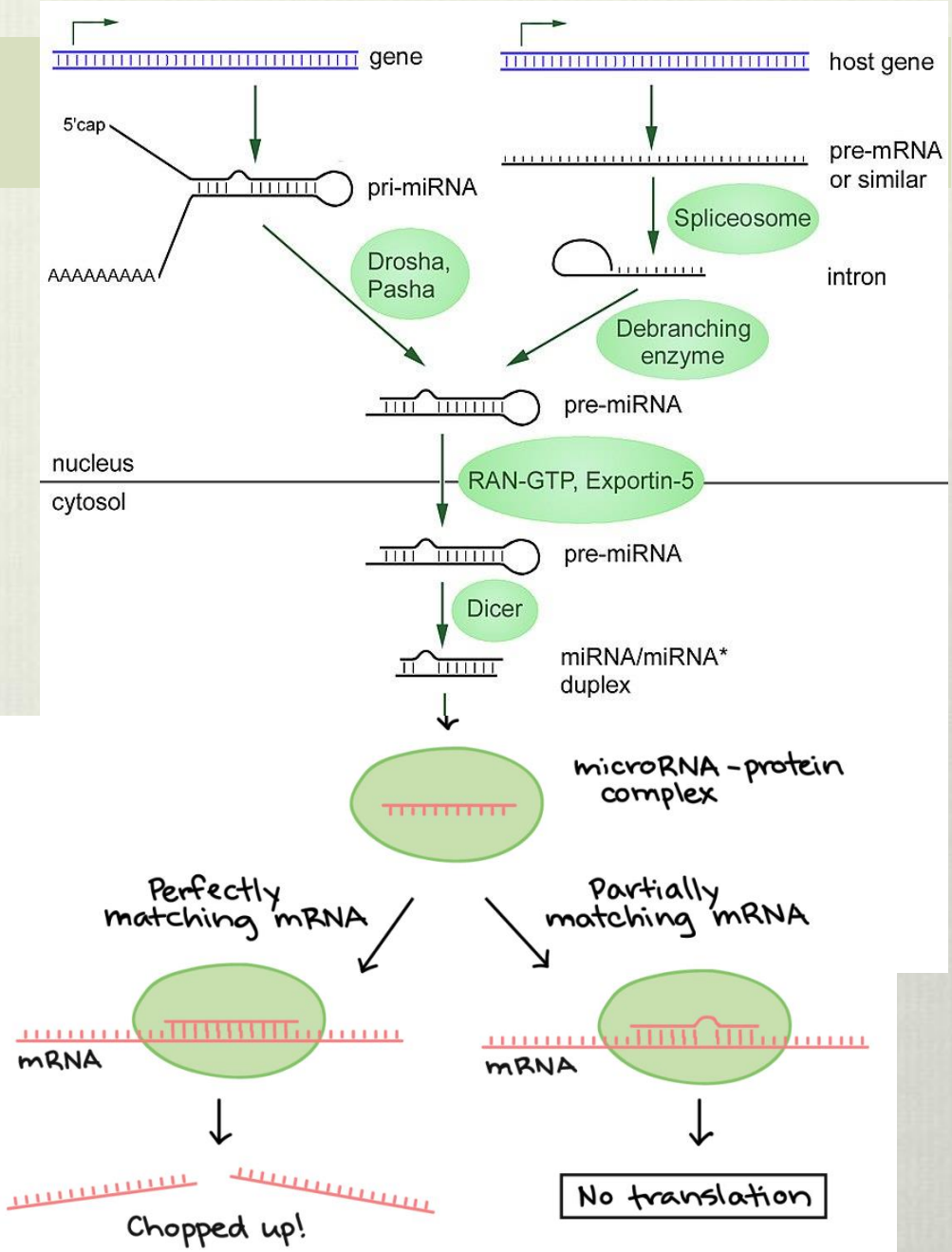


FIGURE 17–1 Regulation can occur at any stage in the expression of genetic material in eukaryotes. All these forms of regulation affect the degree to which a gene is expressed.

Biogénesis

Regulación postranscripcional de la expresión génica

- Los microRNA se transcriben por general mediante Pol-II, pero no se traducen a proteína (genes no-codificantes)
- El pre-microRNA tiene una estructura secundaria característica de 'horquilla'
- Sufren varias pasos de maduración
- Regulan la expresión génica mediante la unión (junto a un complejo proteico llamado RISC) al 3' UTR de un gen



Como empezó todo…

Cell, Vol. 75, 843–854, December 3, 1993, Copyright © 1993 by Cell Press

The C. elegans Heterochronic Gene *lin-4* Encodes Small RNAs with Antisense Complementarity to *lin-14*

Rosalind C. Lee,*† Rhonda L. Feinbaum,*‡ and Victor Ambros†
Harvard University
Department of Cellular and Developmental Biology
Cambridge, Massachusetts 02138

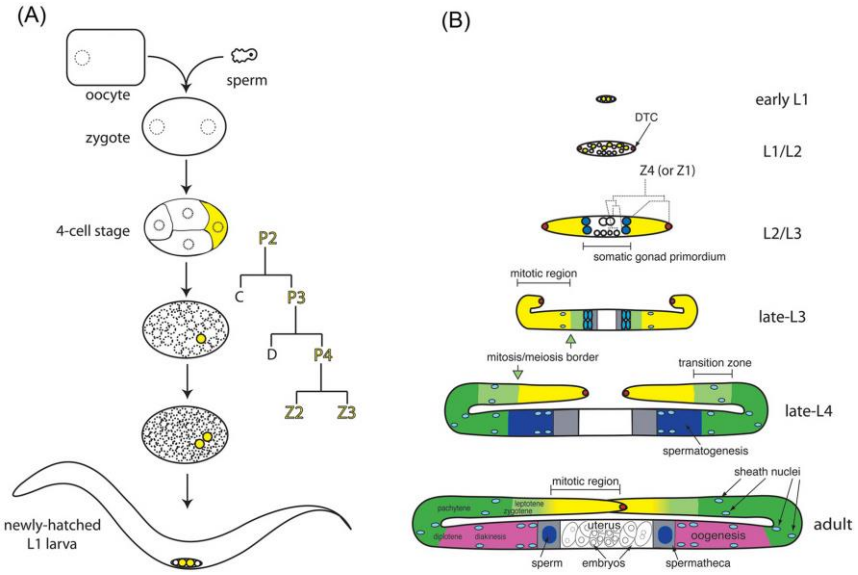
Ambros and Horvitz, 1987). Animals carrying a *lin-4* loss-of-function (*lf*) mutation, *lin-4(e912)*, display reiterations of early fates at inappropriately late developmental stages; cell lineage patterns normally specific for the L1 are reiterated at later stages, and the animals execute extra larval molts (Chalfie et al., 1981). The consequences of these heterochronic developmental patterns include the ab-

Genética directa (Forward genetics)

Estudio de la transición L1-L2 en *C. elegans*

Pérdida de función de Lin-4: *Retraso en el desarrollo*

Pérdida de la función de lin-14: *Expresión temprána de genes específicos de estadios L2, L3 y L4*



Como empezó todo…

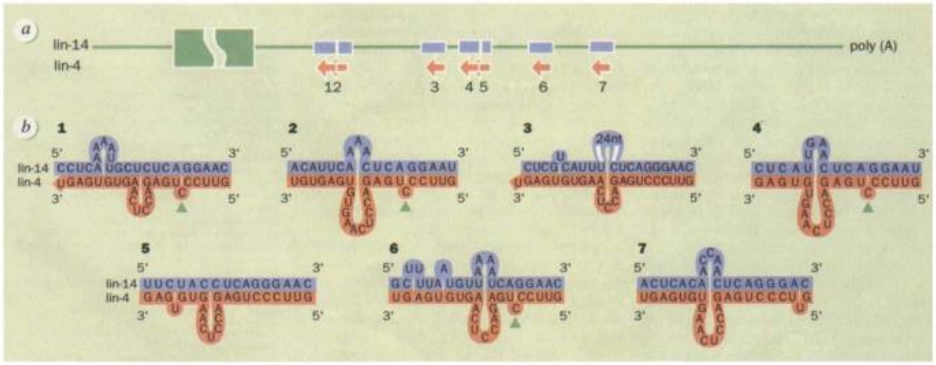
Cell, Vol. 75, 843–854, December 3, 1993, Copyright © 1993 by Cell Press

The C. elegans Heterochronic Gene *lin-4* Encodes Small RNAs with Antisense Complementarity to *lin-14*

Rosalind C. Lee,*† Rhonda L. Feinbaum,*‡ and Victor Ambros†
Harvard University
Department of Cellular and Developmental Biology
Cambridge, Massachusetts 02138

Ambros and Horvitz, 1987). Animals carrying a *lin-4* loss-of-function (*lf*) mutation, *lin-4(e912)*, display reiterations of early fates at inappropriately late developmental stages; cell lineage patterns normally specific for the L1 are reiterated at later stages, and the animals execute extra larval molts (Chalfie et al., 1981). The consequences of these heterochronic developmental patterns include the ab-

Comparison of the *lin-4* genomic sequence from these four species and site-directed mutagenesis of potential open reading frames indicated that *lin-4* does not encode a protein. Two small *lin-4* transcripts of approximately 22 and 61 nt were identified in C. elegans and found to contain sequences complementary to a repeated sequence element in the 3′ untranslated region (UTR) of *lin-14* mRNA, suggesting that *lin-4* regulates *lin-14* translation via an antisense RNA–RNA interaction.



Proposed base-pairing between *lin-4* RNA and elements in the *lin-14* 3′UTR. *a*, *lin-14* mRNA, with the protein-coding region indicated by a large, broken box. Regions of the *lin-14* 3′UTR which potentially can pair with *lin-4* RNA are in blue, with their relative positions drawn to scale. *lin-4* RNA is in red. *b*, Hypothetical base-pairing schemes; base-pairing has not been demonstrated experimentally (see text). The arrow indicates the C in *lin-4* which, when changed to U, prevents repression of *lin-14* mRNA. The termini of *lin-4* RNA have been deduced using DNA probes, rather than by direct analysis of the RNA, and so are imprecise. (Figure derived from refs 1, 2 and 8.)

Como empezó todo…

Cell, Vol. 75, 843–854, December 3, 1993, Copyright © 1993 by Cell Press

The C. elegans Heterochronic Gene *lin-4* Encodes Small RNAs with Antisense Complementarity to *lin-14*

Rosalind C. Lee,*† Rhonda L. Feinbaum,*‡ and Victor Ambros†
Harvard University
Department of Cellular and Developmental Biology
Cambridge, Massachusetts 02138

Ambros and Horvitz, 1987). Animals carrying a *lin-4* loss-of-function (*lf*) mutation, *lin-4(e912)*, display reiterations of early fates at inappropriately late developmental stages; cell lineage patterns normally specific for the L1 are reiterated at later stages, and the animals execute extra larval molts (Chalfie et al., 1981). The consequences of these heterochronic developmental patterns include the ab-

Comparison of the *lin-4* genomic sequence from these four species and site-directed mutagenesis of potential open reading frames indicated that *lin-4* does not encode a protein. Two small *lin-4* transcripts of approximately 22 and 61 nt were identified in C. elegans and found to contain sequences complementary to a repeated sequence element in the 3′ untranslated region (UTR) of *lin-14* mRNA, suggesting that *lin-4* regulates *lin-14* translation via an antisense RNA–RNA interaction.

Tres principios principales:

1. Gen no-codificante con una longitud de 22 nt
2. Regula la expresión génica mediante complementariedad de secuencia
3. Las dianas se ubican el la región 3’ UTR

rimentally (see text).
anged to U, prevents
n-4 RNA have been
t analysis of the RNA,
1, 2 and 8.)

17

Como empezó todo…

Cell, Vol. 75, 843–854, December 3, 1993, Copyright © 1993 by Cell Press

The C. elegans Heterochronic Gene *lin-4* Encodes Small RNAs with Antisense Complementarity to *lin-14*

Rosalind C. Lee,*† Rhonda L. Feinbaum,*‡ and Victor Ambros†
Harvard University
Department of Cellular and Developmental Biology
Cambridge, Massachusetts 02138

Ambros and Horvitz, 1987). Animals carrying a *lin-4* loss-of-function (*lf*) mutation, *lin-4(e912)*, display reiterations of early fates at inappropriately late developmental stages; cell lineage patterns normally specific for the L1 are reiterated at later stages, and the animals execute extra larval molts (Chalfie et al., 1981). The consequences of these heterochronic developmental patterns include the ab-

Comparison of the *lin-4* genomic sequence from these four species and site-directed mutagenesis of potential open reading frames indicated that *lin-4* does not encode a protein. Two small *lin-4* transcripts of approximately 22 and 61 nt were identified in *C. elegans* and found to contain sequences complementary to a repeated sequence element in the 3′ untranslated region (UTR) of *lin-14* mRNA, suggesting that *lin-4* regulates *lin-14* translation via an antisense RNA–RNA interaction.

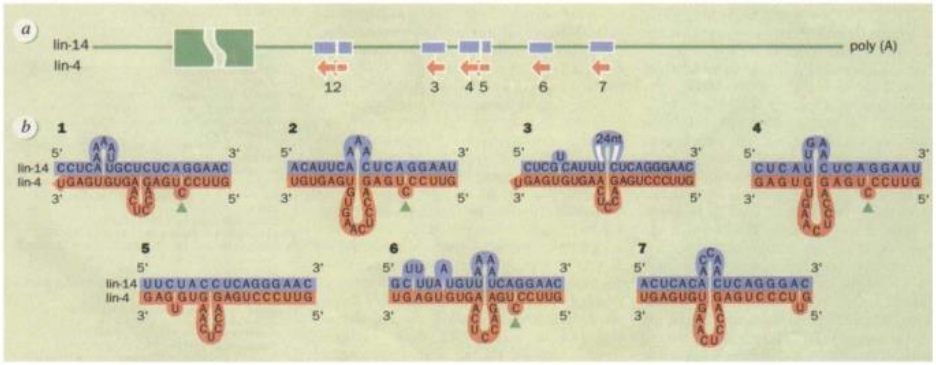
?? ↓ ??

News & Views | Published: 06 January 1994

Deviants — or emissaries

Marvin Wickens & Kathy Takayama

Nature **367**, 17–18 (1994) | Download Citation



Proposed base-pairing between *lin-4* RNA and elements in the *lin-14* 3′UTR. *a*, *lin-14* mRNA, with the protein-coding region indicated by a large, broken box. Regions of the *lin-14* 3′UTR which potentially can pair with *lin-4* RNA are in blue, with their relative positions drawn to scale. *lin-4* RNA is in red. *b*, Hypothetical base-pairing schemes; base-pairing has not been demonstrated experimentally (see text). The arrow indicates the C in *lin-4* which, when changed to U, prevents repression of *lin-14* mRNA. The termini of *lin-4* RNA have been deduced using DNA probes, rather than by direct analysis of the RNA, and so are imprecise. (Figure derived from refs 1, 2 and 8.)

Eran emisarios …

Conservation of the sequence and temporal expression of *let-7* heterochronic regulatory RNA

Amy E. Pasquinelli*†, Brenda J. Reinhart*†, Frank Slack‡, Mark Q. Martindale§, Mitzi I. Kuroda‖, Betsy Maller‡, David C. Hayward¶, Eldon E. Ball¶, Bernard Degnan#, Peter Müller☆, Jürg Spring☆, Ashok Srinivasan**, Mark Fishman**, John Finnerty††, Joseph Corbo‡‡, Michael Levine‡‡, Patrick Leahy§§, Eric Davidson§§ & Gary Ruvkun*

NATURE | VOL 408 | 2 NOVEMBER 2000 | www.nature.com

- Let-7 es altamente conservado entre *C. elegans*, *D. melanogaster* and **H. sapiens**
- Let-7 tiene dianas en la región 3'UTR

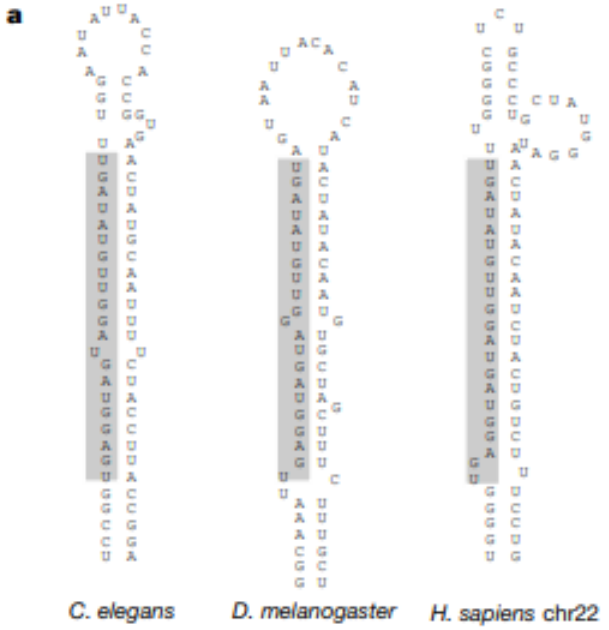


Figure 1 *let-7* gene sequences. **a**, Stem-loop structures of *C. elegans*, *D. melanogaster* and *Homo sapiens let-7* theoretical longer transcripts. The 21-nt *let-7* region is shaded. *let-7* genomic regions from *C. elegans* (Z70203), *D. melanogaster* (AE003659) and *H. sapiens* chromosome 22 (AL049853). The two human *let-7* homologous genes on chromosome 9 are tandemly arranged and separated by 369 base pairs; clustered with the human chromosome 22 *let-7* exact match is an 18/21 match to the *let-7* RNA. **b**, The 3′ UTRs of the *D. melanogaster* (AA3990768) and *Danio rerio lin-41* (AI794385) cDNAs contain *let-7* complementary sites.

Early 2000 – the show begins

Current Biology
Volume 12, Issue 9, 30 April 2002, Pages 735-739

Brief communication

Identification of Tissue-Specific MicroRNAs from Mouse

Mariana Lagos-Quintana, Reinhard Rauhut, Abdullah Yalcin, Jutta Meyer, Winfried Lendeckel, Thomas Tuschl
Show more

https://doi.org/10.1016/S0960-9822(02)00809-6

- microRNAs existen en casi todos los animales y plantas
- microRNAs son altamente conservados

REPORT

Identification of Novel Genes Coding for Small Expressed RNAs

Mariana Lagos-Quintana, Reinhard Rauhut, Winfried Lendeckel, Thomas Tuschl*
+ See all authors and affiliations

Science 26 Oct 2001:
Vol. 294, Issue 5543, pp. 853-858
DOI: 10.1126/science.1064921

REPORT

An Abundant Class of Tiny RNAs with Probable Regulatory Roles in *Caenorhabditis elegans*

Nelson C. Lau, Lee P. Lim, Earl G. Weinstein, David P. Bartel*
+ See all authors and affiliations

Science 26 Oct 2001:
Vol. 294, Issue 5543, pp. 858-862
DOI: 10.1126/science.1065062

REPORT

An Extensive Class of Small RNAs in *Caenorhabditis elegans*

Rosalind C. Lee, Victor Ambros*
+ See all authors and affiliations

Science 26 Oct 2001:
Vol. 294, Issue 5543, pp. 862-864
DOI: 10.1126/science.1065329

Genes Dev. 2002 Jul 1;16(13):1616-26.

MicroRNAs in plants.

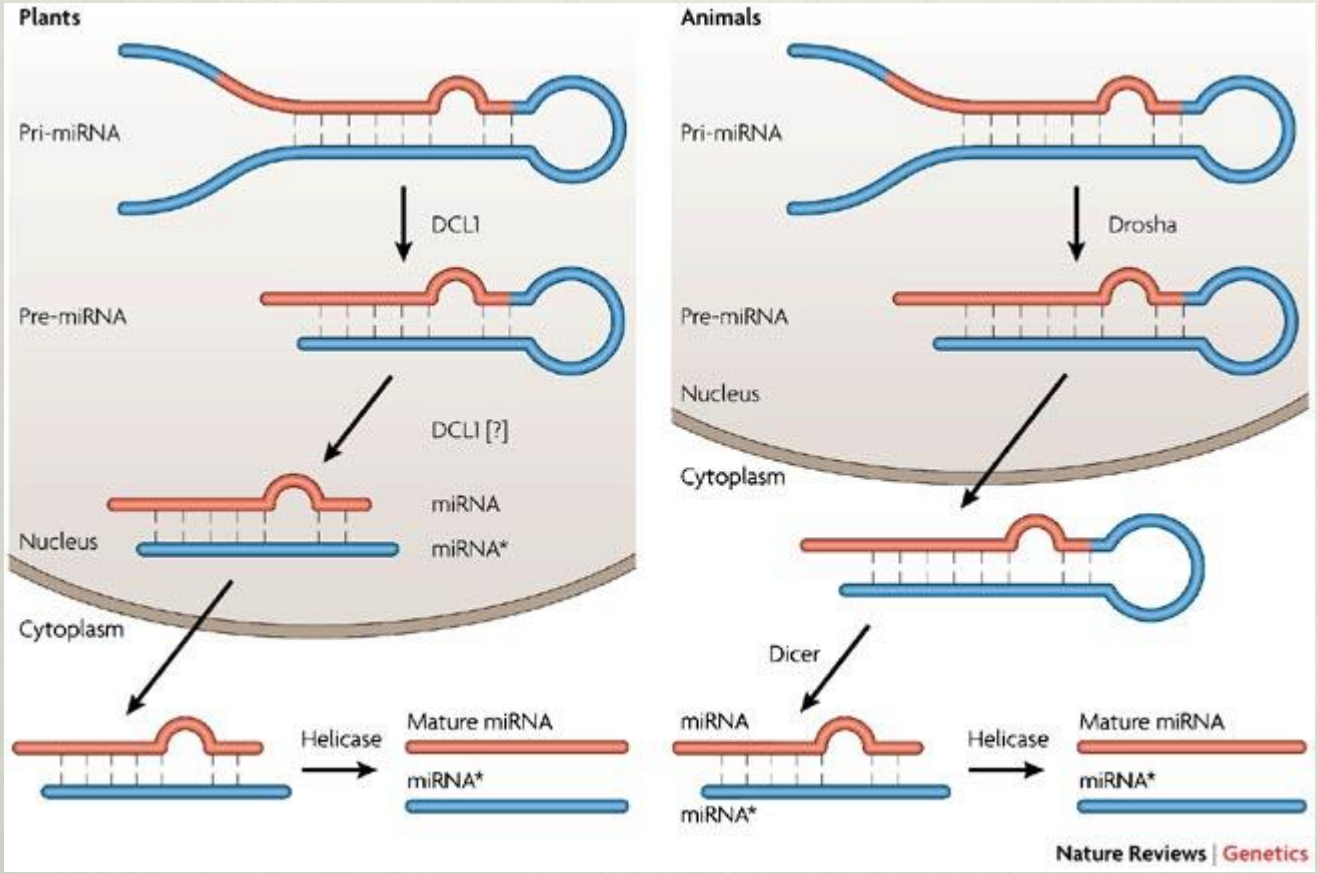
Reinhart BJ[1], Weinstein EG, Rhoades MW, Bartel B, Bartel DP.
Author information

Erratum in
Genes Dev 2002 Sep 1;16(17):2313.

Abstract
MicroRNAs (miRNAs) are an extensive class of ~22-nucleotide noncoding RNAs thought to regulate gene expression in metazoans. We find that miRNAs are also present in plants, indicating that this class of noncoding RNA arose early in eukaryotic evolution. In this paper 16 Arabidopsis miRNAs are described, many of which have differential expression patterns in development. Eight are absolutely conserved in the rice genome. The plant miRNA loci potentially encode stem-loop precursors similar to those processed by Dicer (a ribonuclease III) in animals. Mutation of an Arabidopsis Dicer homolog, CARPEL FACTORY, prevents the accumulation of miRNAs, showing that similar mechanisms direct miRNA processing in plants and animals. The previously described roles of CARPEL FACTORY in the development of Arabidopsis embryos, leaves, and floral meristems suggest that the miRNAs could play regulatory roles in the development of plants as well as animals.

Comparación con plantas

- Se desconoce microRNAs homólogos entre animales y plantas
- En plantas la maduración de la secuencia termina en el núcleo
- Los microRNAs de plantas se metilan en su extremo 3' para protegerlos de 3'-5' exonucleasas de la familia SND



FROM THE FOLLOWING ARTICLE:

The evolution of gene regulation by transcription factors and microRNAs

Kevin Chen & Nikolaus Rajewsky

Nature Reviews Genetics **8**, 93-103 (February 2007)

doi:10.1038/nrg1990

Importancia de los microRNAs

Mas de 45,000 dianas conservadas de microRNAs en las partes 3'UTR de los genes codificantes
60% de los genes codificantes ha sido bajo selección negativa para mantener el apareamiento con algún microRNA

- La conservación de las dianas ya indica funciones importantes
- Los microRNAs estarán implicados en virtualmente todas las funciones moleculares en menor o mayor grado

- microRNAs especifico en mamíferos tienen menos dianas conservadas que microRNAs mas antiguas

Genome Res. 2009 Jan; 19(1): 92–105. PMCID: PMC2612969
doi: 10.1101/gr.082701.108

Most mammalian mRNAs are conserved targets of microRNAs

Robin C. Friedman,[1,2,3] Kyle Kai-How Farh,[1,2,4] Christopher B. Burge,[1,5] and David P. Bartel[1,2,5]

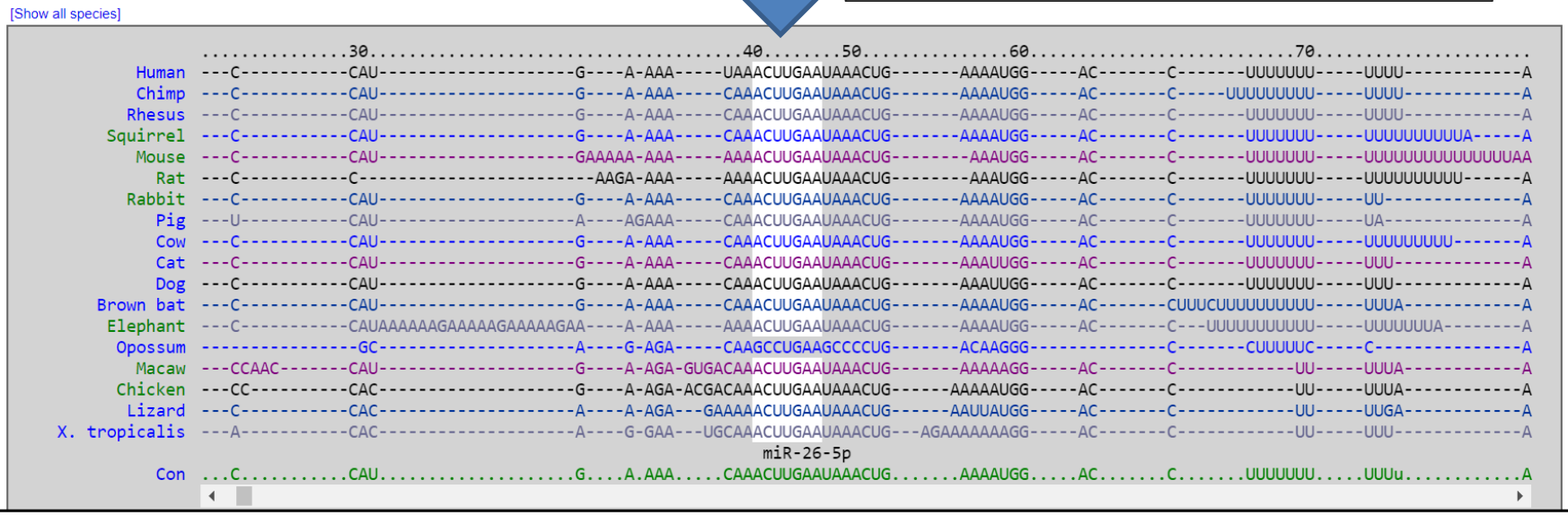
Author information ► Article notes ► Copyright and License information ►

https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2612969/

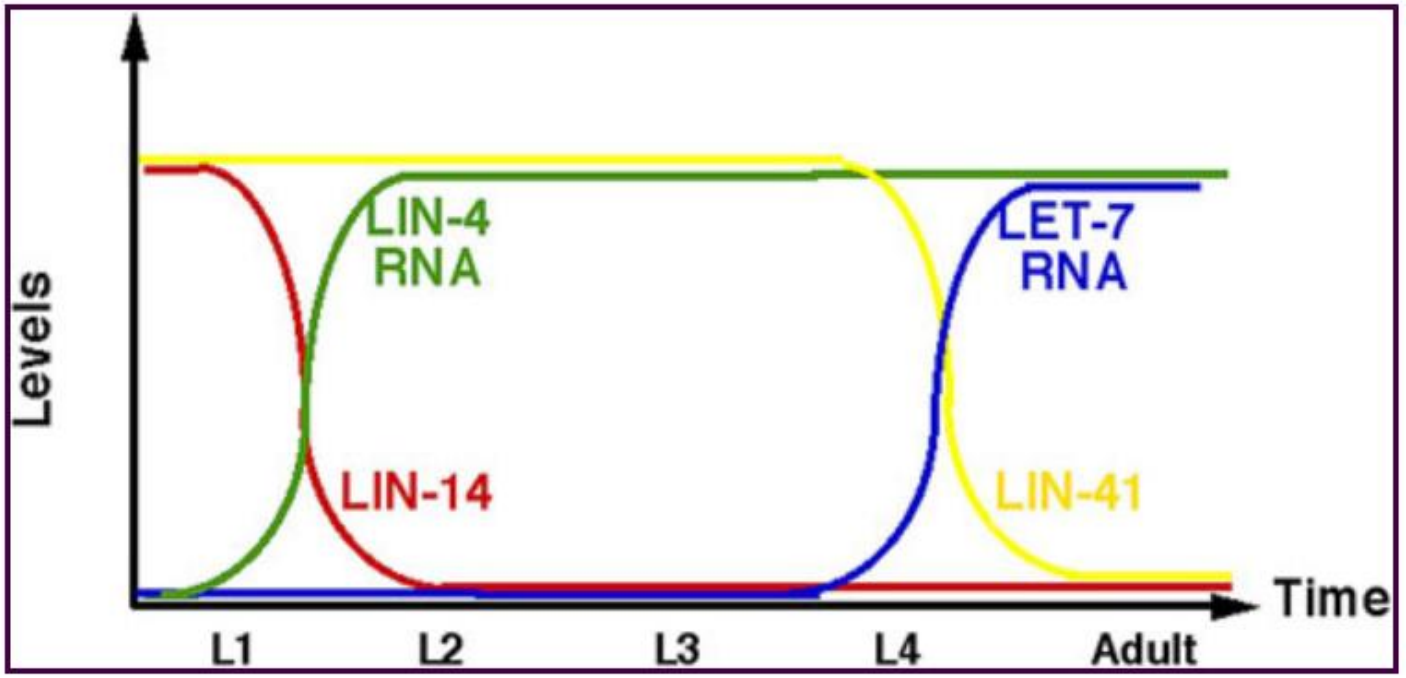
Conservación and función

Mas de 45,000 dianas conservadas de microRNAs en las partes 3’UTR de los genes codificantes

Conservación del la diana, una secuencia complementaria a la del extremo 5’ del microRNA



Expresión temporal de lin-4



http://www.wormbook.org/chapters/www_microRNA/microRNAs.html

… small size and potential to be processed from hairpin precursors, but most were not expressed in a temporal manner. Because they were identified without the help of genetics, their functions were not known—
what was known is that they were small, and so they were called ‘‘microRNAs’’ (Lagos-Quintana

Genética inversa: fenotipos producidos por miRNAs

Review

Metazoan MicroRNAs

David P. Bartel [1, 2]

Show more

https://doi.org/10.1016/ Cell Volume 173, Issue 1, 22 March 2018, Pages 20-51

- **El *knockout* de 69 de 88 miRNA en Drosophila muestran fenotipos severos**
- **En ratones, se observa fenotipos para 20 de los miRNAs mas conservados**
- **En cuando mas conservado, mas probable de producir un fenotipo**

Table 2. Abnormal Phenotypes Observed in Mice after Knocking Out One or More Members of a Broadly Conserved miRNA Family

miRNA family (# of genes)	miRNA(s) removed	Phenotype
miR-15/16/195/322/497 (7)	miR-15a~16-1	Increased proliferation of B cells; development of lymphoproliferative disorders (Klein et al., 2010); arrested maturation of natural killer cells (Sullivan et al., 2015); increased phagocytosis of macrophages; reduced mortality in bacterial sepsis models (Moon et al., 2014)
	miR-15b~16-2	Development of B cell lymphoproliferative disorders (Lovat et al., 2015)
miR-17/20/93/106 (6)	miR-17, miR-20a	Perinatal lethality, incomplete penetrance; vertebral homeotic transformations and other skeletal defects; reduced body weight; reduced pre-B cells (Han et al., 2015)
miR-18 (2)	miR-18a	Reduced body weight (Han et al., 2015)
miR-19 (3)	miR-19a, miR-19b-1	Perinatal lethality, incomplete penetrance; reduced body weight; reduced tumorigenesis in sensitized background (*Myc*) (Han et al., 2015)
miR-21 (1)	miR-21	Reduced growth and increased apoptosis of eosinophil progenitors (Lu et al., 2013b); altered macrophage polarization with more M2 macrophages and fewer M1 macrophages (Wang et al., 2015); increased bone mass with decreased bone resorption (Hu et al., 2017); altered pathology in >20 disease/injury models

"Indeed, loss-of-function studies disrupting miRNA genes in mice have revealed diverse phenotypes, including defects in the **development** of the skeleton, teeth, brain, eyes, neurons, muscle, heart, lungs, kidneys, vasculature, liver, pancreas, intestine, skin, fat, breast, ovaries, testes, placenta, thymus, and each hematopoietic lineage, as well as **cellular, physiological, and behavioral defects**. Many of these developmental and physiological defects **affect embryonic or postnatal viability** or cause other severe conditions, such as **epilepsy, deafness, retinal degeneration, infertility, immune disorders, or cancer**. In addition, some miRNA- knockout strains have altered susceptibility to infections, and many have differential responses to mouse models of diseases or injuries."

microRNAs y función

- **Conservación de miRNAs y dianas**
- **Muchos miRNAs producen fenotipos en estudios de knock-down**

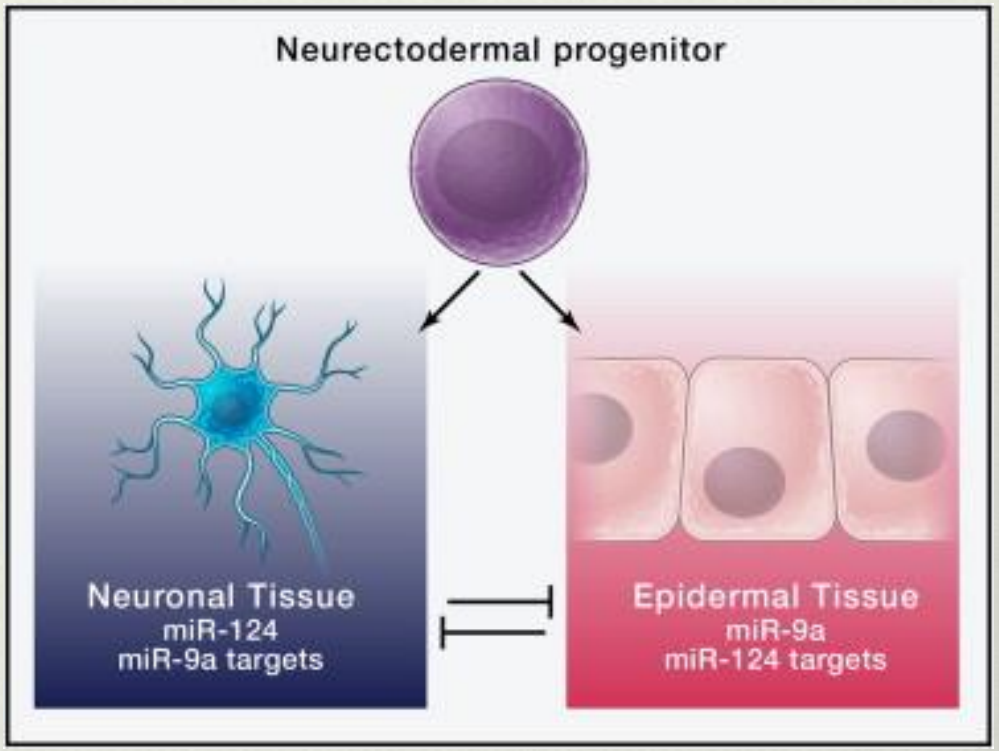
microRNAs desempeñan funciones importantes en muchos procesos biológicos

How are these functions carried out, how can we classify or define them?

Importancia de los microRNAs

- Las células progenitoras neuroectodérmicas expresan miR-124 cuando se diferencian en neuronas
- Los genes se expresan durante esa transición tienden a no tener dianas para este microRNA

Este microRNA estabiliza la transición eliminado los transcritos 'erróneos'



- miR-1 es específico de musculo
- La isoforma especifica de musculo de la tropomisoina-1 no tiene dianas para miR-1
- Otras isoformas, que no deben estar en el musculo, si que tienen dianas

Cell
CellPress
Volume 149, Issue 3, 27 April 2012, Pages 515-524

Review

Roles for MicroRNAs in Conferring Robustness to Biological Processes

Margaret S. Ebert [1, 2, 3], Phillip A. Sharp [1, 2]

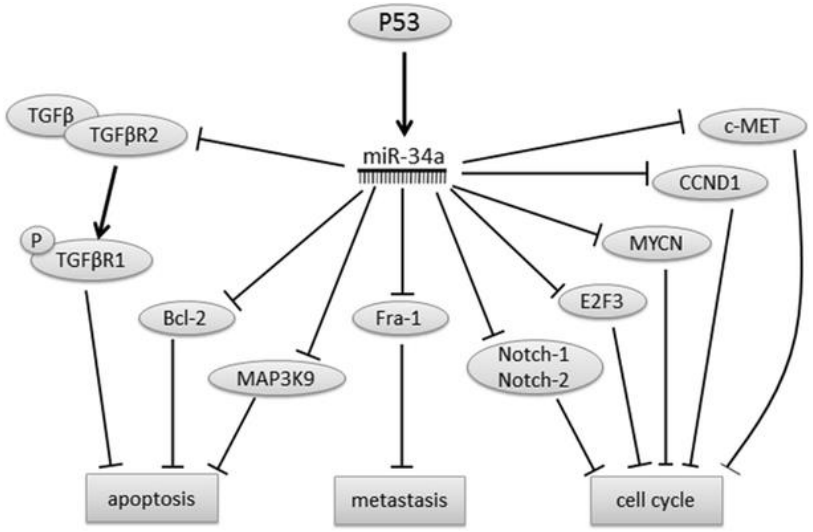
Target recognition

Tumor Biology
April 2015, Volume 36, Issue 4, pp 2481–2490 | Cite as

MicroRNA-34a inhibits the proliferation and promotes the apoptosis of non-small cell lung cancer H1299 cell line by targeting TGFβR2

Authors | Authors and affiliations

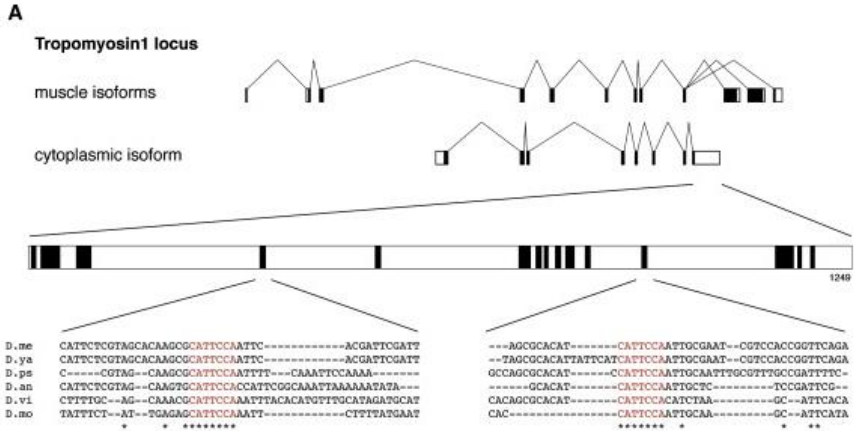
Zhong-Liang Ma, Pin-Pin Hou, Yan-Li Li, De-Tao Wang, Tian-Wei Yuan, Jia-Li Wei, Bo-Tao Zhao, Jia-Tao Lou, Xin-Tai Zhao, Yan Jin, You-Xin Jin



https://www.cell.com/cell/fulltext/S0092-8674(05)01272-9

Animal MicroRNAs Confer Robustness to Gene Expression and Have a Significant Impact on 3′UTR Evolution

Alexander Stark,[1,2,3] Julius Brennecke,[1,2] Natascha Bushati,[1] Robert B. Russell,[1] and Stephen M. Cohen[1,*]
[1]European Molecular Biology Laboratory, Meyerhofstrasse 1, 69117 Heidelberg, Germany
[2]These authors contributed equally to this work.
[3]Present address: Broad Institute of MIT and Harvard, Cambridge, MA 02141, USA and Computer Science and Artificial Intelligence Laboratory, Massachusetts Institute of Technology, Cambridge, MA 02139, USA.
*Contact: cohen@embl.de
DOI 10.1016/j.cell.2005.11.023



Conferring robustness to gene expression

‘Fine-tuning’ gene expression in complex networks

Both ‘ways’ imply negative regulation of gene expression through sequence complementarity

microRNAs y cáncer

Developmental Cell
CellPress
Volume 11, Issue 4, October 2006, Pages 441-450

Review

The Diverse Functions of MicroRNAs in Animal Development and Disease

Wigard P. Kloosterman [1], Ronald H.A. Plasterk [1, 2]

Mutaciones en los microRNAs o dianas pueden afectar la función – observable en muchos tipos de cáncer

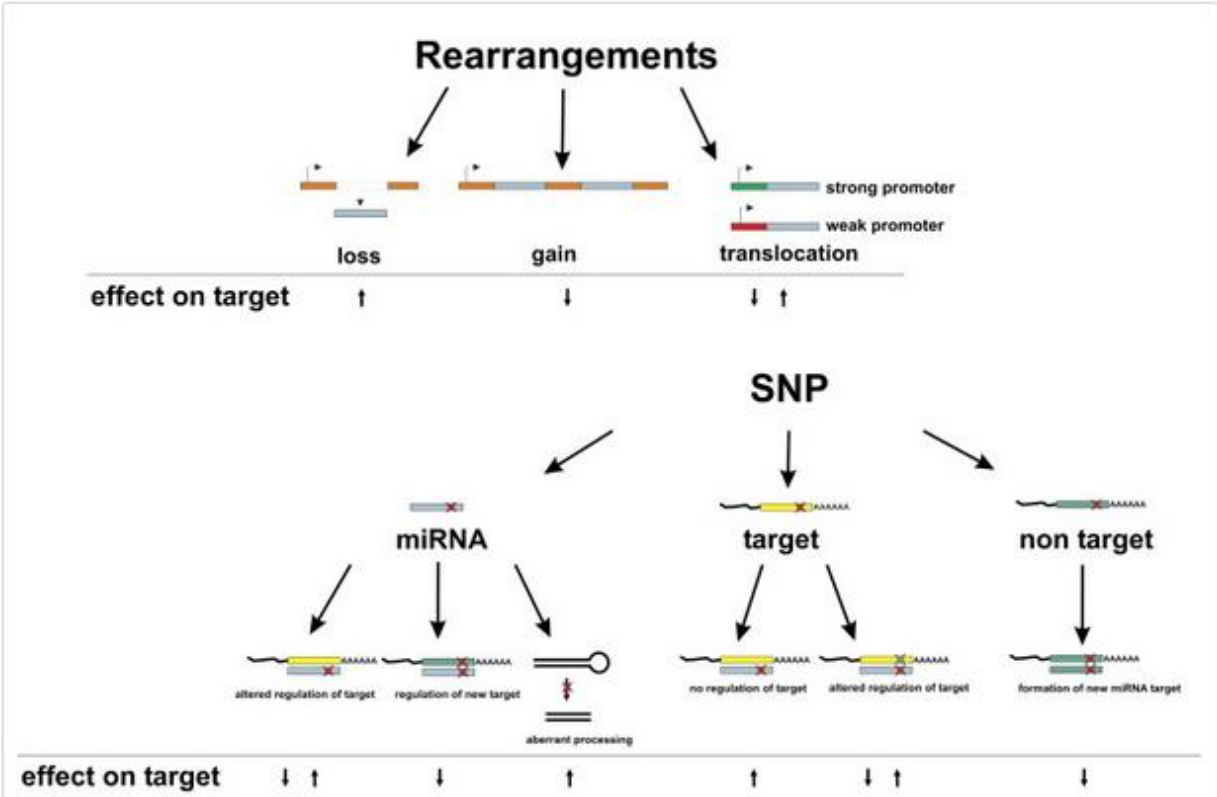


Figure 3.

Mechanisms that Link miRNAs to Disease

There are two possible scenarios: (1) either the expression level of the miRNA changes due to genomic rearrangements; or (2) there is a gain or loss of an miRNA-target interaction due to a mutation in a 3′UTR or a mutation in the miRNA.

microRNAs y cáncer

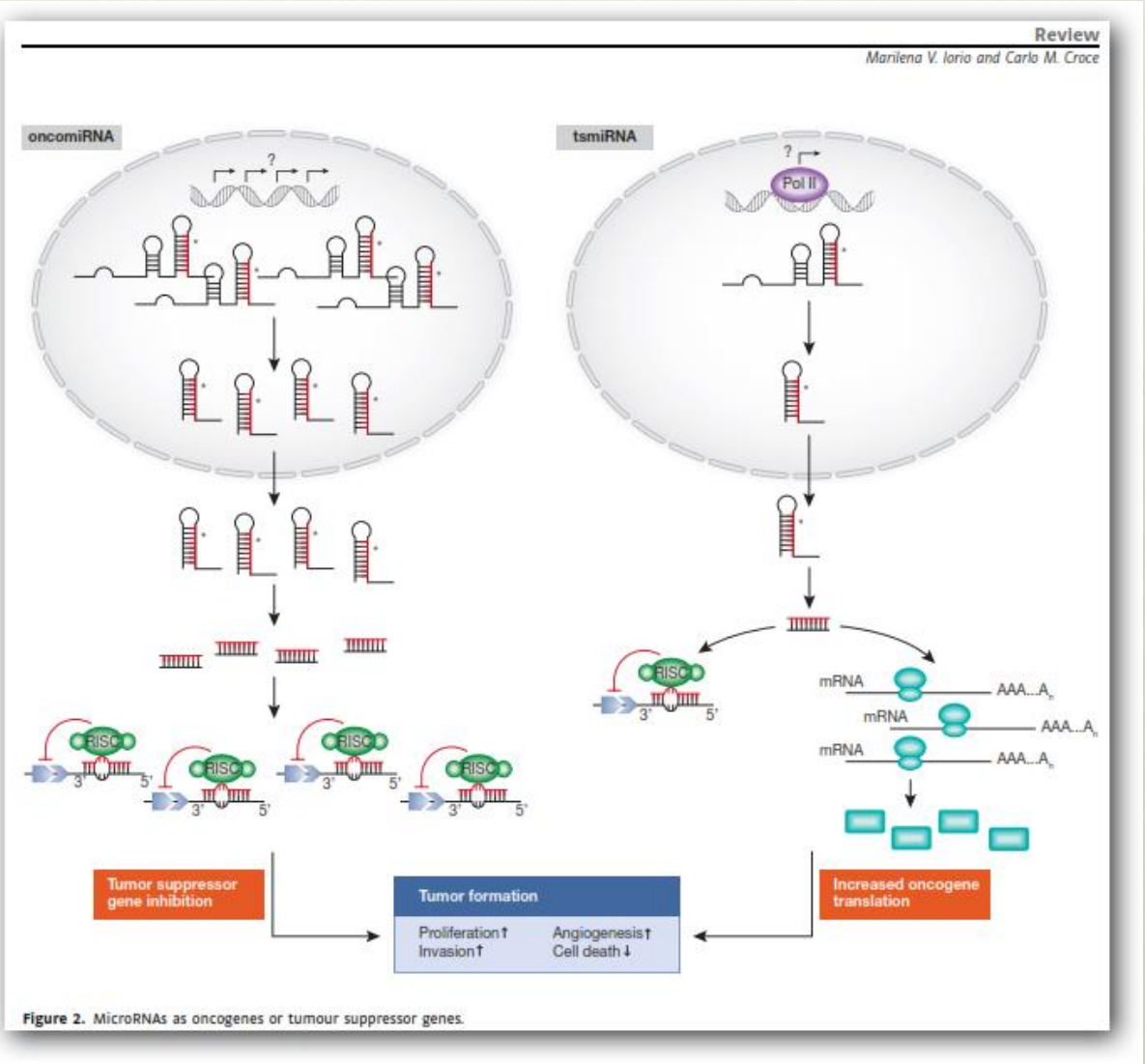
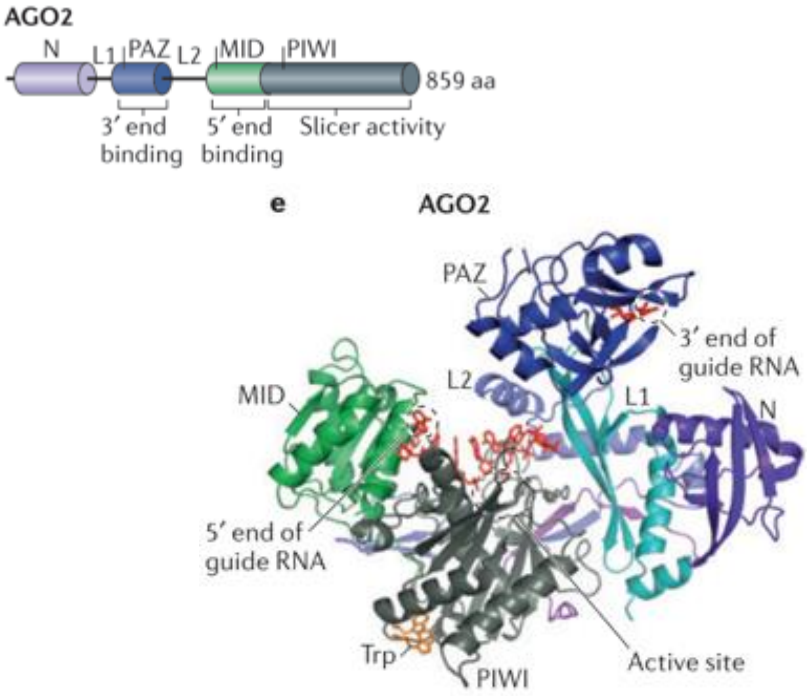


Figure 2. MicroRNAs as oncogenes or tumour suppressor genes.

How are the targets recognized?



- AGO are key proteins
- AGO2 has catalytic activity
- 5′ base of the guide RNA with a preference for U or A binding
- AGO2-deficient mice are embryonic lethal

Review Article | Published: 16 July 2014

Regulation of microRNA biogenesis

Minju Ha & V. Narry Kim

Nature Reviews Molecular Cell Biology **15**, 509–524 (2014) | Download Citation

Metazoan MicroRNAs

David P. Bartel[1,2,*]
[1]Howard Hughes Medical Institute and Whitehead Institute for Biomedical Research, Cambridge, MA 02142, USA
[2]Department of Biology, Massachusetts Institute of Technology, Cambridge, MA 02139, USA
*Correspondence: dbartel@wi.mit.edu
https://doi.org/10.1016/j.cell.2018.03.006

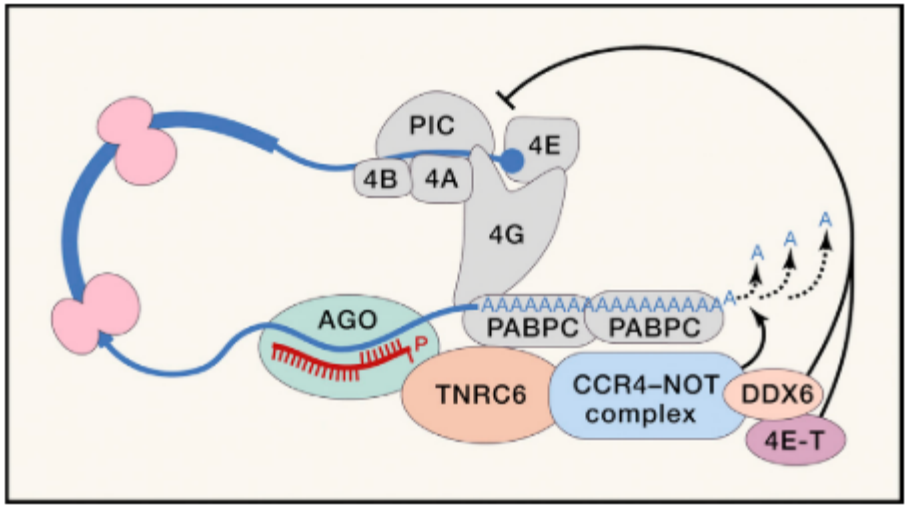


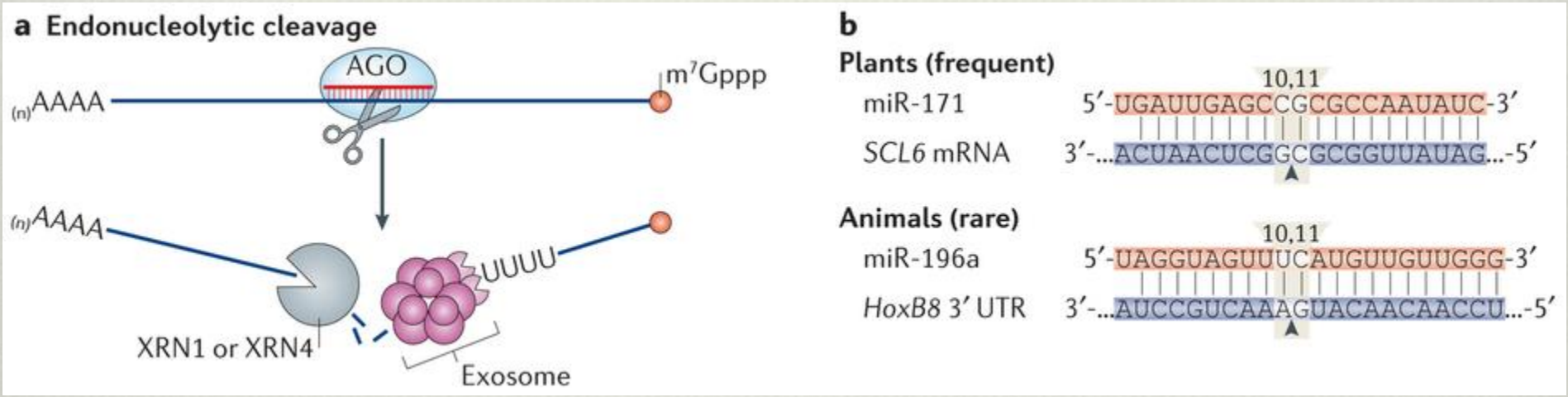
Figure 4. The Dominant Mechanisms of miRNA-Guided Repression in Bilaterian Animals

Guided by the miRNA, the silencing complex associates with the mRNA and recruits TNRC6, which interacts with PABPC and recruits either the PAN2–PAN3 deadenylase complex (not shown) or the CCR4–NOT deadenylase complex, either of which shortens the mRNA poly(A) tail. Alternative downstream consequences of poly(A)-tail shortening, which are not depicted in this figure, consummate this major mode of TNRC6-mediated repression; in early embryos, tail shortening reduces translation initiation with little effect on mRNA stability, whereas in most other developmental contexts, tail shortening hastens decapping and degradation of the mRNA with relatively little effect on translation initiation. Although not through tail shortening, recruitment of TNRC6 can nonetheless repress translation initiation in post-embryonic cells through a parallel mechanism that involves CCR4–NOT-mediated recruitment of DDX6 and 4E-T. This translation initiation normally involves the recruitment of the 43S preinitiation complex (PIC) through the action of initiation factors (4A, 4B, 4E, 4G).

¿Como detectan sus dianas?

El microRNA reconoce su diana mediante la complementariedad de secuencia

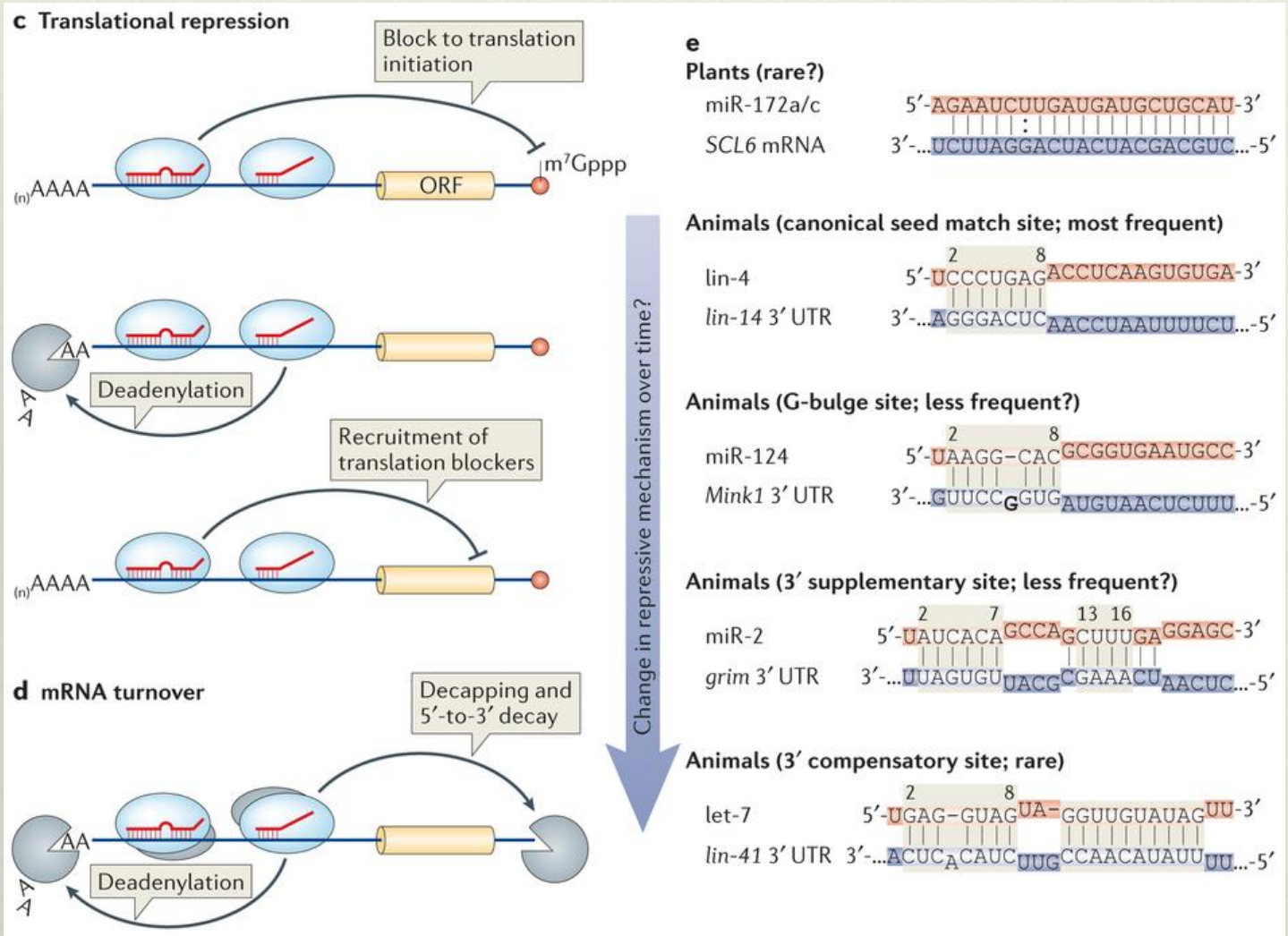
- En plantas la complementariedad suele ser casi 'perfecta'
- En animales, las primeras 8 bases se llaman *seed* (semilla) y son especialmente importante para reconocer la diana



Complementariedad (casi) perfecta induce la digestión endonucleolítica

From
Diversifying microRNA sequence and function
Stefan L. Ameres & Phillip D. Zamore
Nature Reviews Molecular Cell Biology **14**, 475–488 (2013) | doi:10.1038/nrm3611

¿Como detectan sus dianas?

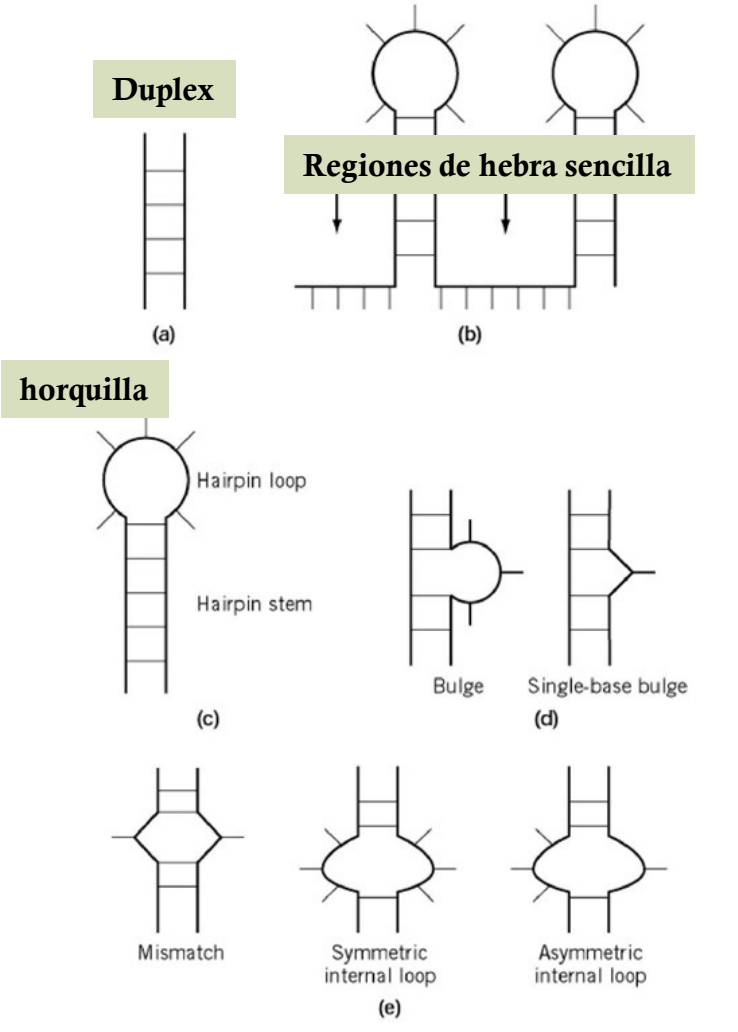


From Diversifying microRNA sequence and function
Stefan L. Ameres & Phillip D. Zamore
Nature Reviews Molecular Cell Biology 14, 475–488 (2013) | doi:10.1038/nrm3611

Complementariedad parcial en plantas
→ Represión de la traducción

Predicción de las dianas

Modelos termodinámicos para calcular estructuras secundarias



Tendencia de biomoléculas en solución acuosa es la de minimizar la energía libre del sistema

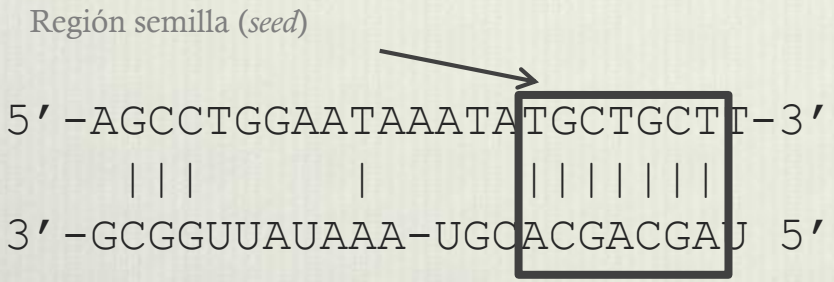
Energía libre de Gibbs

Se define el cambio de la energía libre de un proceso químico (como el plegamiento de una molécula de RNA)

$$\Delta G = \Delta H - T\Delta S$$

http://what-when-how.com/molecular-biology/rna-structure-molecular-biology/

Predicción de las dianas



Complementariedad perfecta entre la posición 2-8 en 5' del microRNA

Hibrido entre hsa-miR-16 (abajo) & NM_004178 (arriba) - Posición 249-269 en la 3' UTR

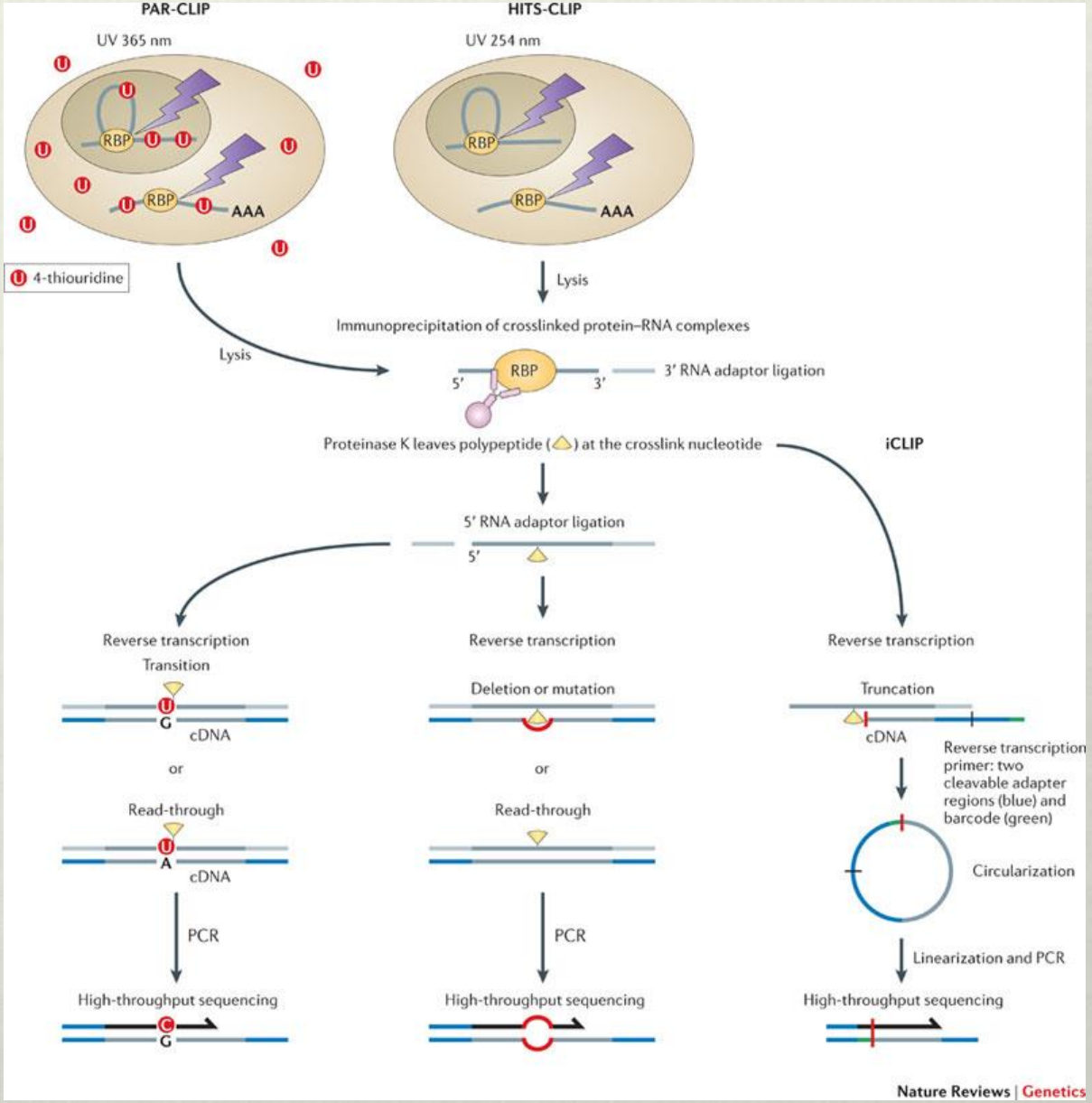
Los híbridos (dúplexes) se calculan mediante modelos termodinámicos – *minimum free energy algorithm*

- La predicción de dianas de microRNA se basa fuertemente en la presencia de la región semilla
  → La región semilla es muy pequeña → predicción no es muy especifica → **señal filogenética**

- Otras propiedades importantes: energía libre, numero de desemparejamientos, numero de bucles, "accesibilidad" de la estructura secundaria efectos combinatorios la interacción entre varias dianas en la misma 3' UTR

Verificación experimental de las dianas

- Inducir enlaces covalentes entre RNA y proteínas
- Inmunoprecipitación con anticuerpo específico frente a la proteína de interés
- Purificar el RNA
- Generación de librería de cDNA
- Secuenciación & análisis bioinformático



Progress

Nature Reviews Genetics **13**, 77-83 (February 2012) | doi:10.1038/nrg3141
Corrected online: 31 January 2012

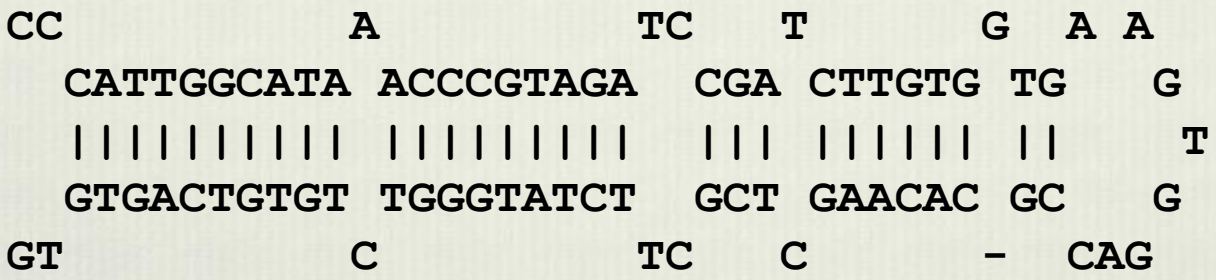
There is an Erratum (1 March 2012) associated with this article.

ARTICLE SERIES: Applications of next-generation sequencing

Protein–RNA interactions: new genomic technologies and perspectives

Julian König[1], Kathi Zarnack[2], Nicholas M. Luscombe[1,2,3] & Jernej Ule[1] About the authors

Detección de genes de microRNA



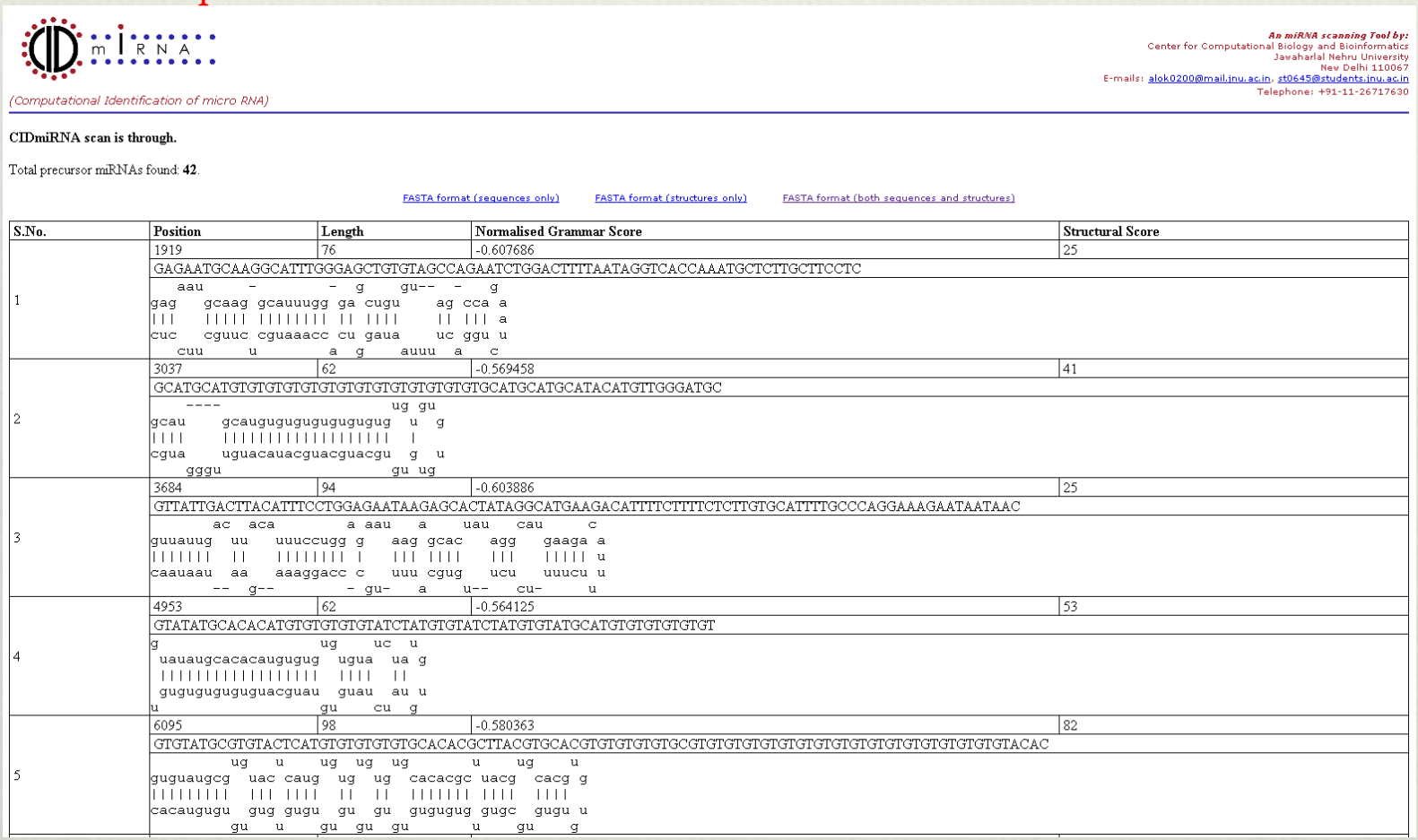
Predicción basada en las propiedades de la estructura secundaria

- Minimum Free Energy
- Number of bindings
- Loop Length
- Number of bulges, bulbs
- Compositional features (G+C content, dinucleotide frequencies)

In the human genome exist approx. 11 million hairpin structures
→ High rate of false positive predictions

Detección de genes de microRNA

Salida del programa CID-miRNA:
Con los parámetros por defecto, el programa predice 42 genes de microRNA en una secuencia de aprox. 100 kb



(Computational Identification of micro RNA)

An miRNA scanning Tool by:
Center for Computational Biology and Bioinformatics
Jawaharlal Nehru University
New Delhi 110067
E-mails: alok0200@mail.jnu.ac.in, st0645@students.jnu.ac.in
Telephone: +91-11-26717630

CIDmiRNA scan is through.

Total precursor miRNAs found: **42**.

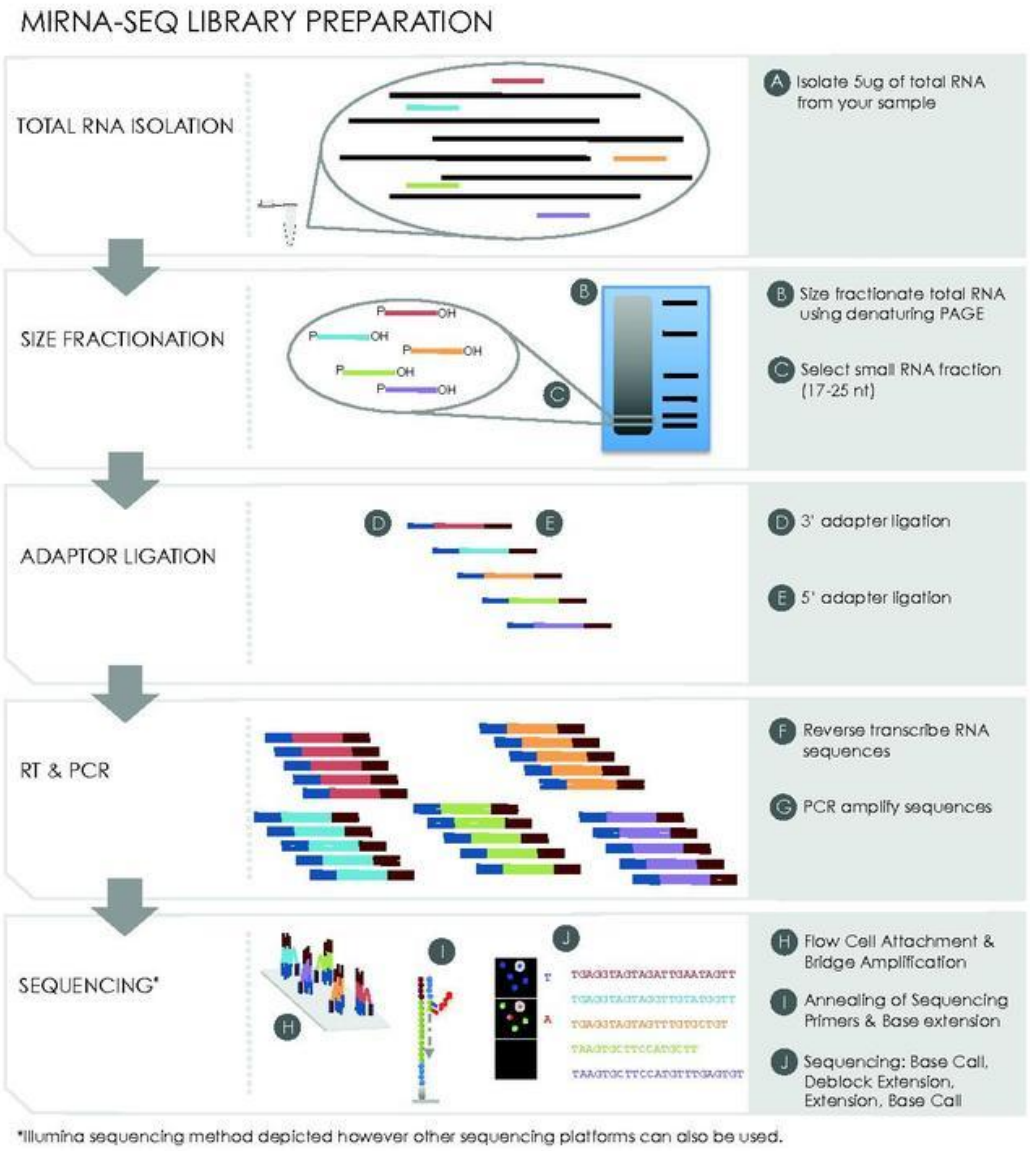
FASTA format (sequences only) FASTA format (structures only) FASTA format (both sequences and structures)

S.No.	Position	Length	Normalised Grammar Score	Structural Score
1	1919	76	-0.607686	25
	GAGAATGCAAGGCATTTGGGAGCTGTGTAGCCAGAATCTGGACTTTTAATAGGTCACCAAATGCTCTTGCTTCCTC			
	aau - - g gu-- - g gag gcaag gcauuugg ga cugu ag cca a ||| |||| ||||||| || |||| || ||| a cuc cguuc cguaaacc cu gaua uc ggu u cuu u a g auuu a c			
2	3037	62	-0.569458	41
	GCATGCATGTGTGTGTGTGTGTGTGTGTGTGTGTGCATGCATGCATACATGTTGGGATGC			
	---- ug gu gcau gcaugugugugugugugug u g |||| ||||||||||||||| | cgua uguacauacguacguacgu g u gggu gu ug			
3	3684	94	-0.603886	25
	GTTATTGACTTACATTTCCTGGAGAATAAGAGCACTATAGGCATGAAGACATTTTCTTTTCTCTTGTGCATTTTGCCCAGGAAAGAATAATAAC			
	ac aca a aau a uau cau c guuauug uu uuuccugg g aag gcac agg gaaga a ||||||| || ||||||| | ||| |||| ||| ||||| u caauaau aa aaaggacc c uuu cgug ucu uuucu u -- g-- - gu- a u-- cu- u			
4	4953	62	-0.564125	53
	GTATATGCACACATGTGTGTGTATCTATGTGTATCTATGTGTATGCATGTGTGTGTGTGT			
	g ug uc u uauaugcacacaugugug ugua ua g |||||||||||||||| |||| || gugugugugugugguau guau au u u gu cu g			
5	6095	98	-0.580363	82
	GTGTATGCGTGTACTCATGTGTGTGTGTGCACACGCTTACGTGCACGTGTGTGTGTGCGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTACAC			
	ug u ug ug ug u ug u guguaugcg uac caug ug ug cacacgc uacg cacg g ||||||||| ||| |||| || || ||||||| |||| |||| cacaugugu gug gugu gu gu gugugug gugc gugu u gu u gu gu gu u gu g			

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

Secuenciación masiva & genes de microRNAs

- Extracción del RNA total
- Ligar adaptadores
- Seleccionar por longitud
- Generar una librería de cDNA
- Secuenciar
- Análisis bioinformático



Source:
http://en.wikipedia.org/wiki/MicroRNA_Sequencing

Análisis bioinformático

El punto de partida

Las secuencias de las lecturas en formato fastq

Cada lectura está representada por 4 líneas

```
@SRR037876 GSM522374_1:1:148:931:861
TAGTTCTACAGTCCGACGATCTCGTATGCCGTCTTC
+
BB@+?0:4@B@-@/A<3A7@-=@<1=@87=?<==9#
```

Secuencia de la lectura

Calidad de la lectura

La longitud de las lecturas depende del número de ciclos en las secuenciación, frecuentemente entre 36 y 50 nucleótidos

Flujo del trabajo

1) Control de calidad

- Eliminar lecturas con baja calidad
  → **Muy importante** para la detección de variación de secuencia
  → **Potencialmente importante en la detección de isomiRs**

2) Detectar y eliminar el adaptador

- Se secuencia parte del adaptador si la molécula es mas corta que la lecturas (número de ciclos)
- El adaptador no alinea frente al genoma
- Hay que eliminarlo para poder mapear la lectura y detectar isomiRs

3) Colapsar las lecturas

- Unir lecturas únicas en una entrada única que consiste en
  - Secuencia & conteo (read count) (las veces una secuencia fue observada en un experimento)

4) Alinear las lecturas únicas frente a una referencia

Detectar el adaptador

Mature microRNAs are around 21 nt long --- in the example below, the reads are 50 nt long

→(part of) the adapter is sequenced as well

```
@SRR518946.15 DA19881:1:30F8JAAXX:8:1:8:919 length=50
TAGCTTATCAGACTGATGTTGACTCGTATGCCGTCTTCTGCTTGTGTGTT
+SRR518946.15 DA19881:1:30F8JAAXX:8:1:8:919 length=50
BBBB<BBBABCABBABB@@B>@?=B@7:@6A?=8>B>7?##########
```

Align the adapter sequence to the read

```
                3' RNA Adapter
5' P-UCGUAUGCCGUCUUCUGCUUGU
```

Parameters:

- **Minimum length of detected adapter sequence (10 nt)**
- **Number of allowed mismatches**

```
@SRR518946.15 DA19881:1:30F8JAAXX:8:1:8:919 length=50
TAGCTTATCAGACTGATGTTGACTCGTATGCCGTCTTCTGCTTGTGTGTT
+SRR518946.15 DA19881:1:30F8JAAXX:8:1:8:919 length=50
BBBB<BBBABCABBABB@@B>@?=B@7:@6A?=8>B>7?##########
```

Detectar el adaptador

```
@
TAAGTGGGAGGCCCTCGTATGCCGTCTTCTGCTTGTAAAAAAAAAAAATA
+
BCA6<>>BBAB?AC@AABACBAB?'>A:A>?@A@################
@
TGAGGTAGTAGATTGTATAGTTTCGTATGCCGTCTTCTGCTTGATTATGT
+
BC@2ABCC?BBA?=BABB??ABB@B@=A@?@B?AB;##############
@
TCGTATGCCGTCTTCTGCTTGAAAAAAAAAAAATAATTTTTTTTTTTTTT
+
B@;>?BBB?3?BBBB@9@A?0<AA############################
@
TTCAAGTAATCCAGGATAGGCTTCGTATGCCGTCTTCTGCTTTAATTTTT
+
>CBCBBA?@@CBB@?BA@7:?A7=>8/:7<>=29################
@
TAATACTGCCTGGTAATGATGACTCGTATGCCGTCTTCTGCTTGTTGTGG
+
BCCCCCBCCCCCC?>ACCCBCCBAB>>@@BBAB=;;?=B@##########
@
TGAGGTAGTAGATTGTATAGTTTCGTATGCCGTCTTCTGCTTGATTTTTT
+
BCAAA?BC:6<AABA>@:98=B:AA@A9>>@;??################
@
TAGCTTATCAGACTGATGTTGACTCGTATGCCGTCTTCTGCTTGTGTGTT
+
BBBB<BBBABCABBABB@@B>@?=B@7:@6A?=8>B>7?###########
```

Detectar el adaptador

```
@
TAAGTGGGAGGCCC
+
BCA6<>>BBAB?AC
@
TGAGGTAGTAGATTGTATAGTT
+
BC@2ABCC?BBA?=BABB??AB
@

+

@
TTCAAGTAATCCAGGATAGGCTTCGTATGCCGTCTTCTGCTTTAATTTTT
+
>CBCBBA?@@CBB@?BA@7:?A7=>8/:7<>=29###############
@
TAATACTGCCTGGTAATGATGAC
+
BCCCCCBCCCCCC?>ACCCBCCB
@
TGAGGTAGTAGATTGTATAGTT
+
BCAAA?BC:6<AABA>@:98=B
@
TAGCTTATCAGACTGATGTTGAC
+
BBBB<BBBABCABBABB@@B>@?
```

14 nt

22 nt

0 nt

Not trimmed (50 nt)

22 nt

22 nt

23 nt

Colapsar las lecturas

Convert adapter trimmed fastq file into read/count format

→ **Collapse "redundant reads" into unique reads & read count**

Ficheros de entrada

Fasta:

```
>1999420#1462
TGAGATGAAGCACTGTAGAAAA
>443945#1281
GGGAGCATCTCTCGGTCTATGCTGT
>633088#562
TAGATGAAGCACTGTAGCTCTT
>255762#230
GCATTGGTGGTAGAATTCTCGCC
>516042#97
TCAGATGAAGCACTGTAGCTCTT
>1582566#86
TAGATGAATCACTGTAGCTC
>1462753#79
TGGAATTATGGAAAATGACAGATGGC
>625879#40
GTTAAGATATCCCGGACGAGCCC
>517214#8
TCCTTTGGTATAGTGGTGAGTATCCC
>626077#2
TGACTTTGACCTGAGAGAAGAAGGC
```

Read/count

```
TGAGATGAAGCACTGTAGAAAA   1462
GGGAGCATCTCTCGGTCTATGCTGT      1281
TAGATGAAGCACTGTAGCTCTT    562
GCATTGGTGGTAGAATTCTCGCC   230
TCAGATGAAGCACTGTAGCTCTT     97
TAGATGAATCACTGTAGCTC        86
TGGAATTATGGAAAATGACAGATGGC       79
GTTAAGATATCCCGGACGAGCCC    40
TCCTTTGGTATAGTGGTGAGTATCCC         8
TGACTTTGACCTGAGAGAAGAAGGC          2
```

Asignar lecturas

```
>1999420#1462
TGAGATGAAGCACTGTAGAAAA
>443945#1281
GGGAGCATCTCTCGGTCTATGCTGT
>633088#562
TAGATGAAGCACTGTAGCTCTT
>255762#230
GCATTGGTGGTAGAATTCTCGCC
>516042#97
TCAGATGAAGCACTGTAGCTCTT
>1582566#86
TAGATGAATCACTGTAGCTC
```

From which RNAs are these reads derived?

↓

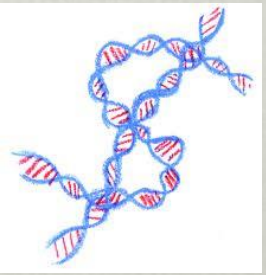
Assign reads to a reference library

1

Map reads to a set of known small RNA sequences

2

Map to the genome & genome annotations



Asignar lecturas

1

Input: read/count

```
>1999420#12682
TGAGGTAGTAGGTTGTGTGGTT
>633088#5692
TGAGATGAAGCACTGTAGCTC
>255762#2630
TGAGGTAGTAGGTTGTATAGTT
>516042#1297
TCAGATGAAGCACTGTAGCTCTT
>443945#181
TGAGGTAGTAGGTTGTGTGGT
>1582566#86
TGAGGTAGTAGGTTGTATAGT
>1462753#79
TGGAATTATGGAAAATGACAGATGGC
>625879#40
GTTAAGATATCCCGGACGAGCCC
>517214#8
TACCCTGTAGAACCGAATTTGTG
>626077#2
TGACTTTGACCTGAGAGAAGAAGGC
```

Alignment
Bowtie, Blast, etc

Sequence library (miRBase)

```
>hsa-let-7b-5p
TGAGGTAGTAGGTTGTGTGGTT

> hsa-miR-143-3p
TGAGATGAAGCACTGTAGCTC

> hsa-let-7a-5p
TGAGGTAGTAGGTTGTATAGTT

> hsa-miR-509-3p
TGATTGGTACGTCTGTGGGTAG

> hsa-miR-10b-5p
TACCCTGTAGAACCGAATTTGTG
```

Asignar lecturas

1

Input: read/count

>1999420#**12682 → read count**
TGAGGTAGTAGGTTGTGTGGTT
>633088#5692
TGAGATGAAGCACTGTAGCTC
>255762#2630
TGAGGTAGTAGGTTGTATAGTT
>516042#1297
TCAGATGAAGCACTGTAGCTCTT
>443945#181
TGAGGTAGTAGGTTGTGTGGT
>1582566#86
TGAGGTAGTAGGTTGTATAGT
>1462753#79
TGGAATTATGGAAAATGACAGATGGC
>625879#40
GTTAAGATATCCCGGACGAGCCC
>517214#8
TACCCTGTAGAACCGAATTTGTG
>626077#2
TGACTTTGACCTGAGAGAAGAAGGC

Alignment
Bowtie, Blast, etc

Sequence library (miRBase)

>hsa-let-7b-5p
TGAGGTAGTAGGTTGTGTGGTT

> hsa-miR-143-3p
TGAGATGAAGCACTGTAGCTC

> hsa-let-7a-5p
TGAGGTAGTAGGTTGTATAGTT

> hsa-miR-509-3p
TGATTGGTACGTCTGTGGGTAG

> hsa-miR-10b-5p
TACCCTGTAGAACCGAATTTGTG

Sum the read count of all mapped reads

Name	Read Count
hsa-let-7b-5p	12863
hsa-miR-143-3p	5692
hsa-let-7a-5p	2716
hsa-miR-10b-5p	8

Detectar nuevos microRNAs

- Drosha/Dicer (DCL in plants) processing patterns can be detected
- Both mature microRNAs (both arms) are represented in the sample?
- 5' end of the mature microRNA shows less fluctuation
- Virtually all reads are organized in one or two clusters