Ch 4 – L4.1

micro RNA

Textbook

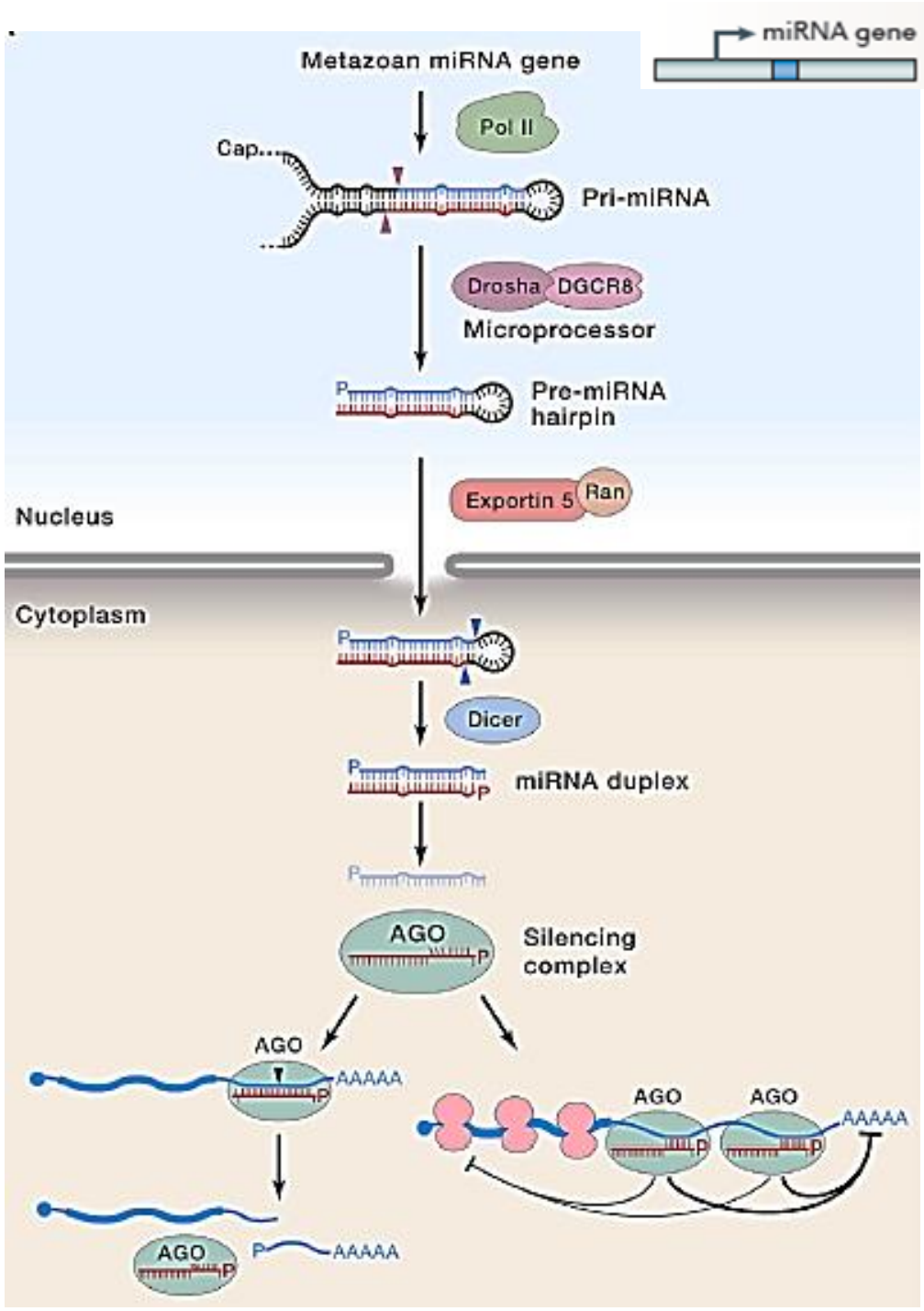
Cell

Leading Edge
Review

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

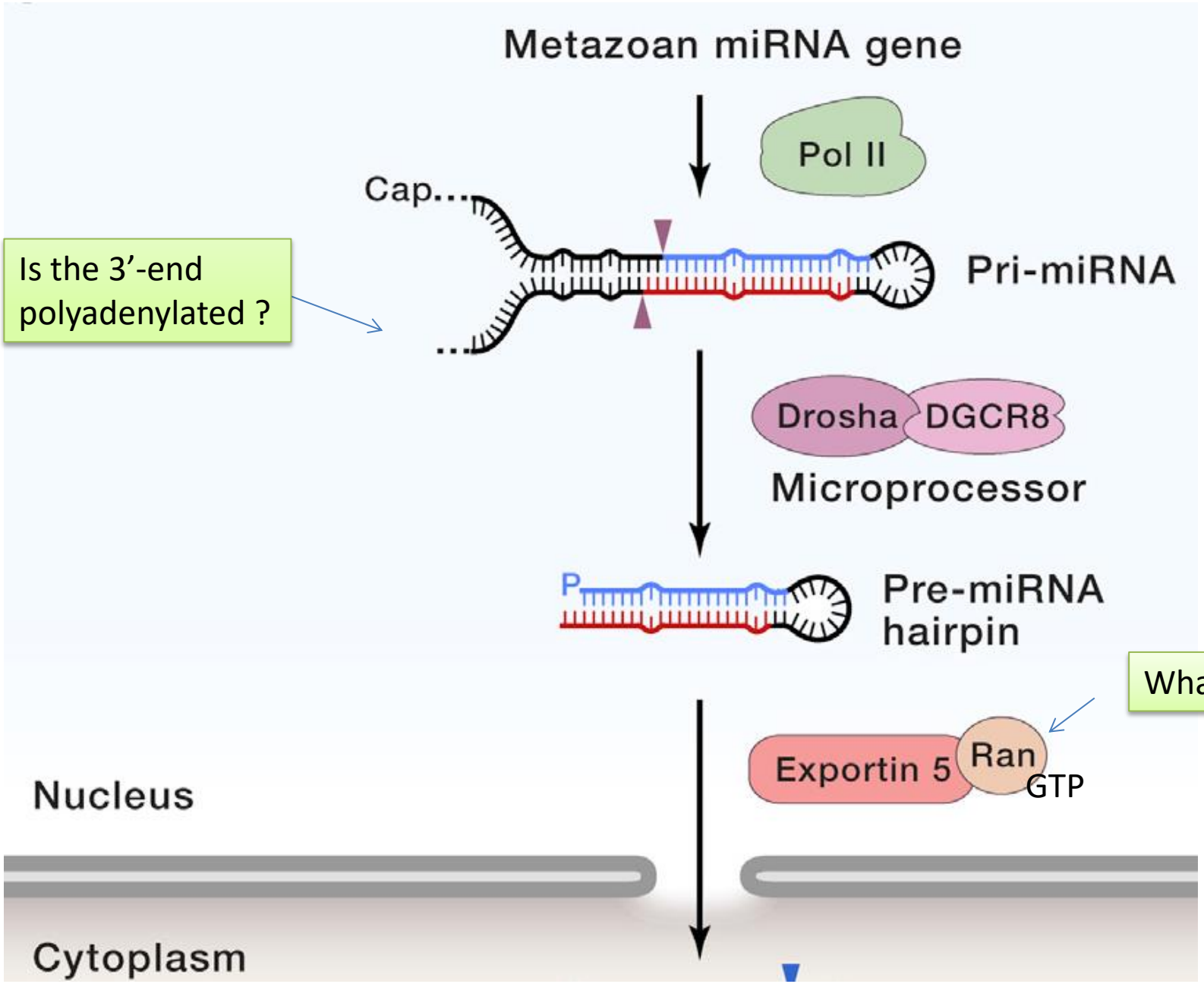
MicroRNAs (miRNAs) are ~22 nt RNAs that direct posttranscriptional repression of mRNA targets in diverse eukaryotic lineages. In humans and other mammals, these small RNAs help sculpt the expression of most mRNAs. This article reviews advances in our understanding of the defining features of metazoan miRNAs and their biogenesis, genomics, and evolution. It then reviews how metazoan miRNAs are regulated, how they recognize and cause repression of their targets, and the biological functions of this repression, with a compilation of knockout phenotypes that shows that important biological functions have been identified for most of the broadly conserved miRNAs of mammals.



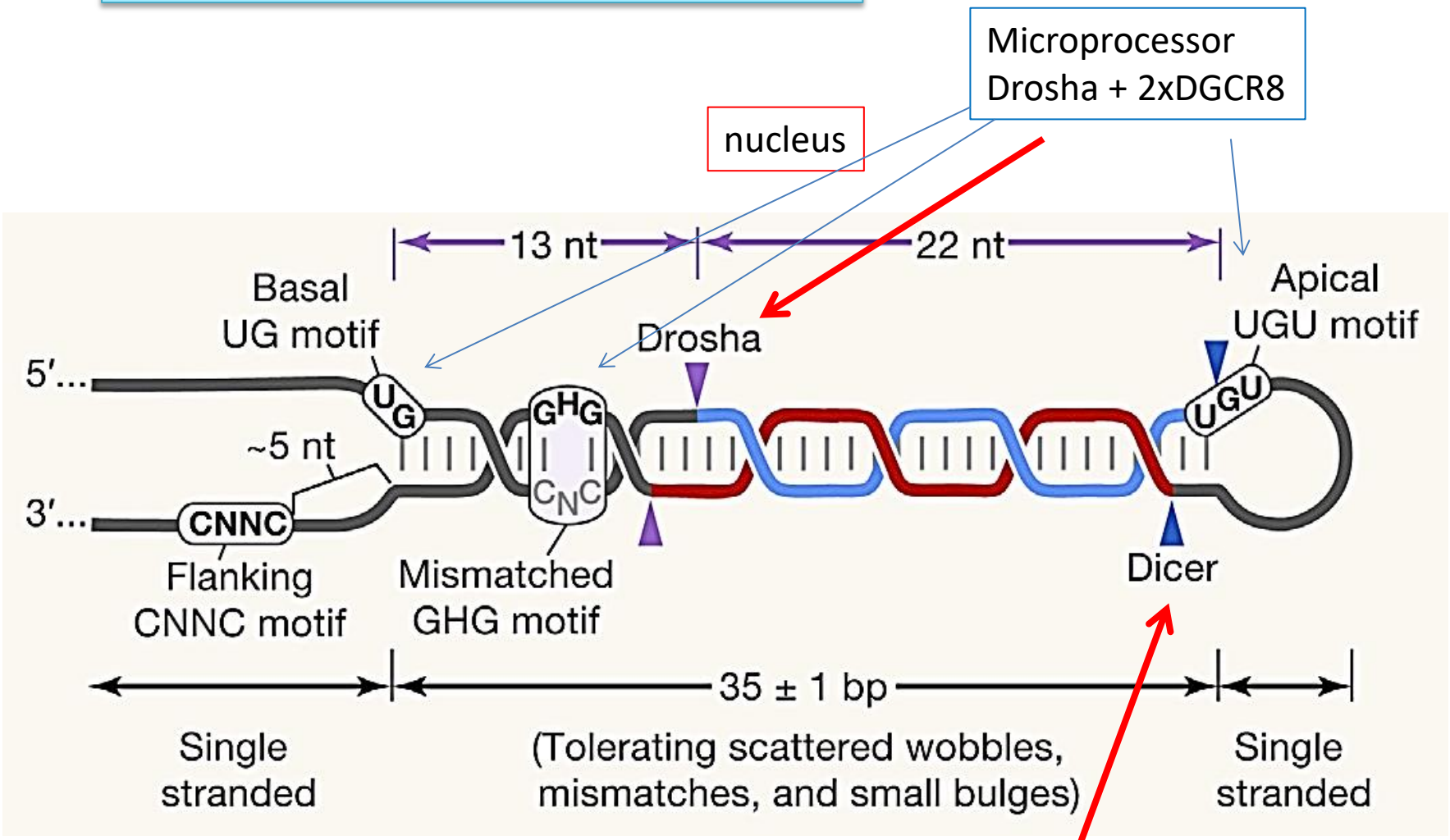
https://www.ensembl.org/index.html

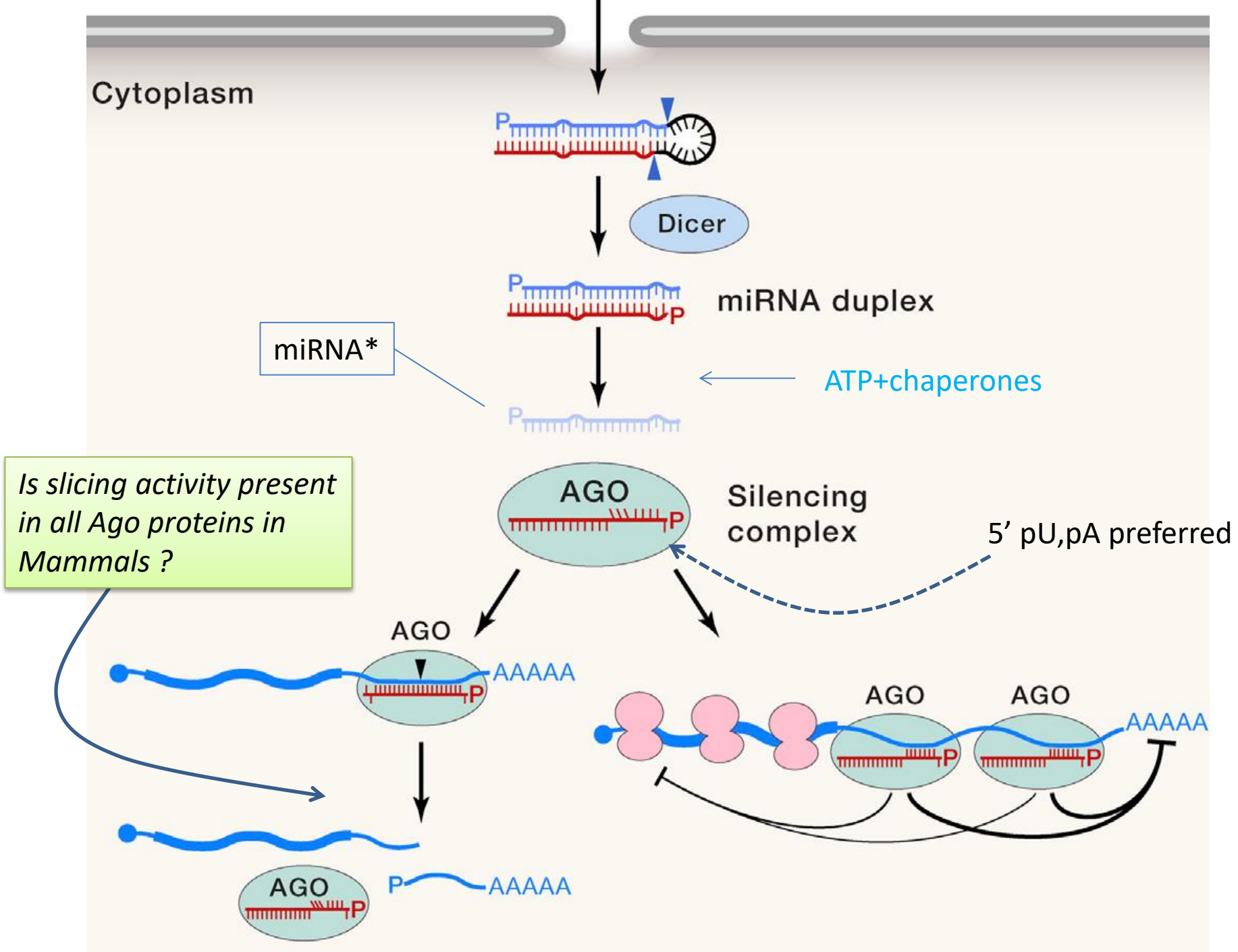
https://www.gencodegenes.org/

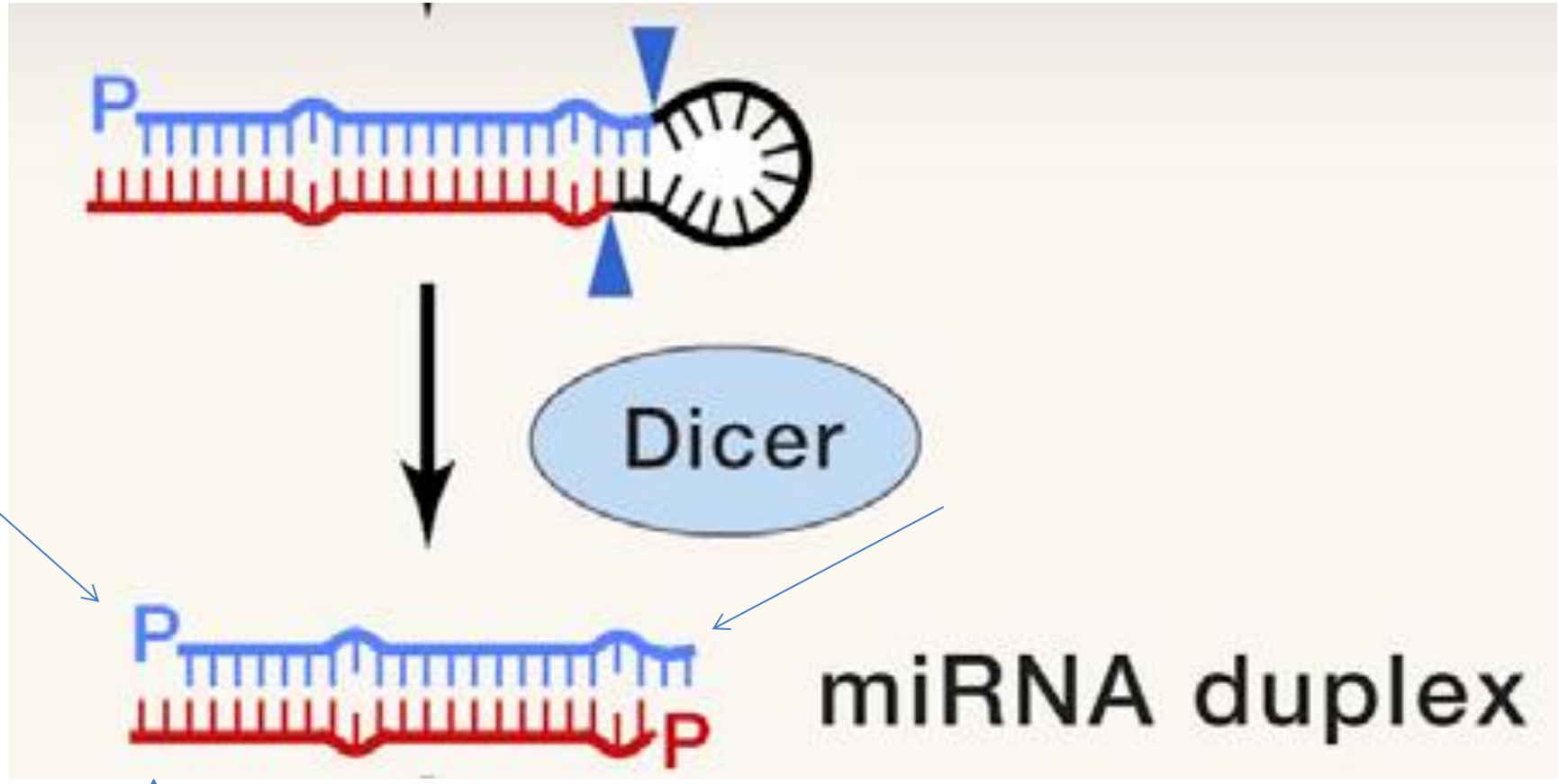
http://www.mirbase.org/index.shtml


Metazoan miRNA gene
Pol II
Cap
Pri-miRNA
Is the 3'-end polyadenylated ?
Drosha
DGCR8
Microprocessor
P
Pre-miRNA hairpin
What is Ran ?
Exportin 5
Ran
GTP
Nucleus
Cytoplasm

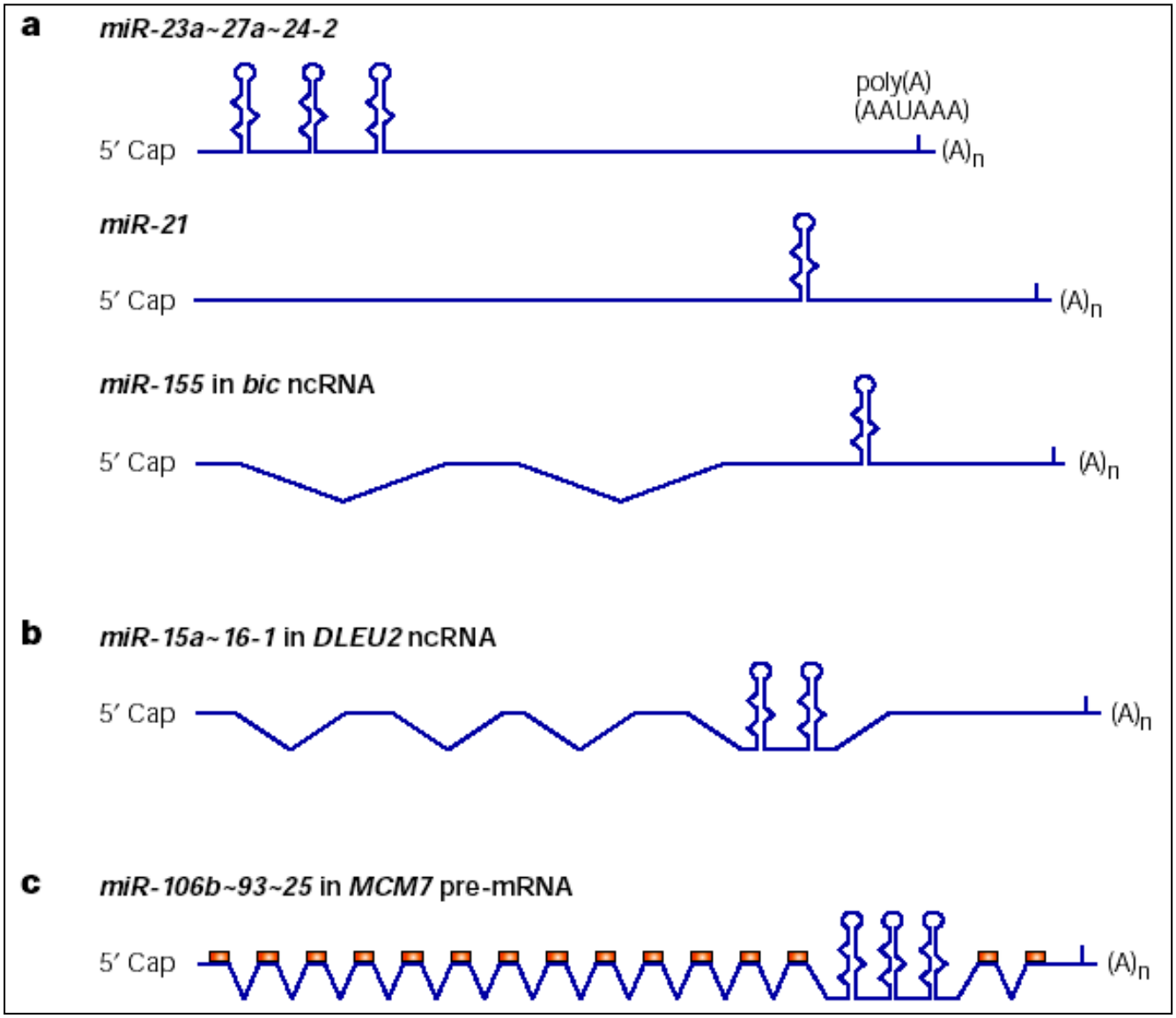
Features defining a canonical pre-miRNA




Cytoplasm
P
Dicer
P
P
miRNA duplex
miRNA*
ATP+chaperones
P
AGO
P
Silencing complex
5’ pU,pA preferred
Is slicing activity present in all Ago proteins in Mammals ?
AGO
P
AAAAA
AGO
AGO
P
P
AAAAA
AGO
P
P
AAAAA


P
Dicer
P
P
miRNA duplex

Biogenesis



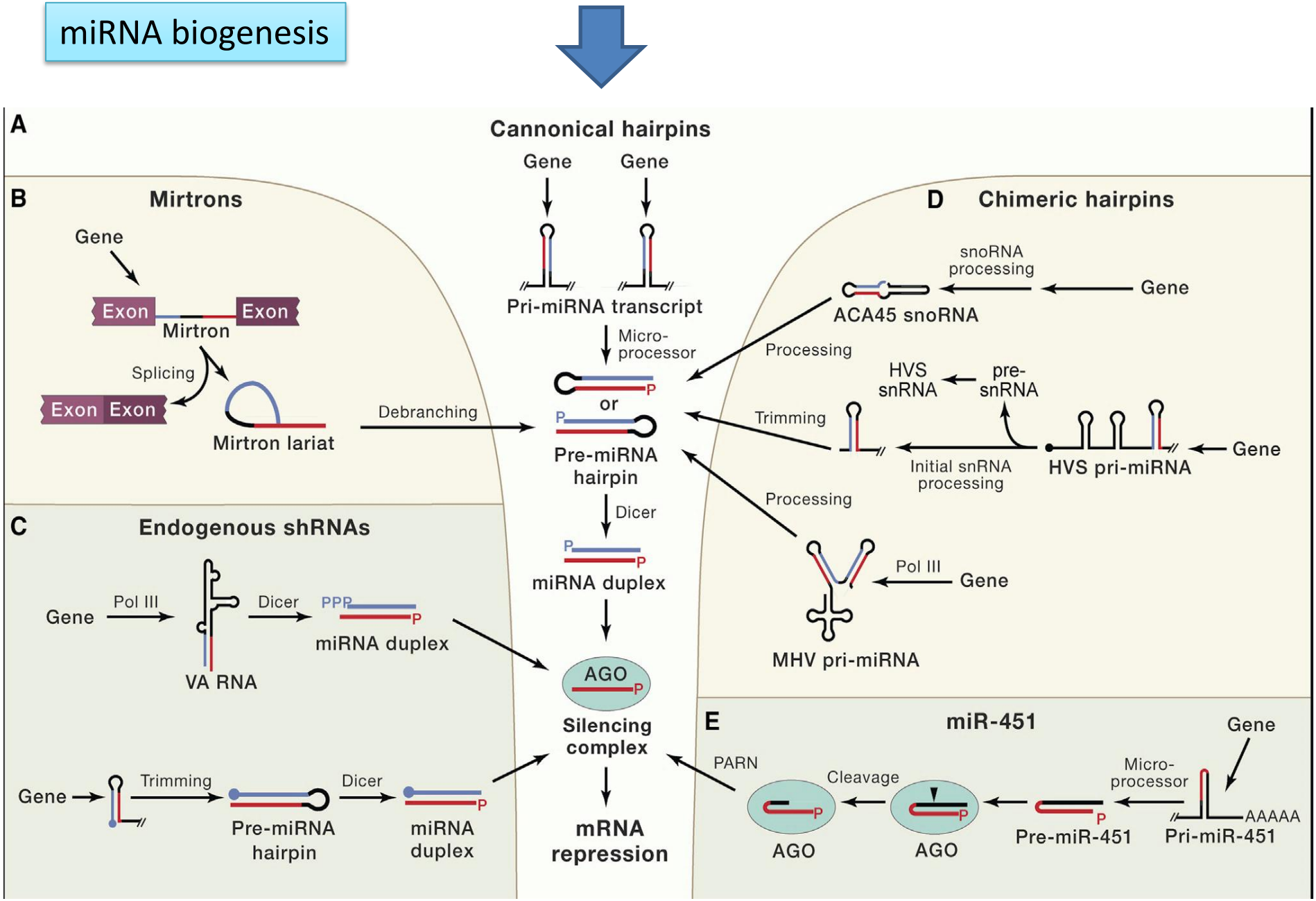
The structure of five pri-miRNAs. (This figure is not to scale**)**

a | Exonic miRNAs in non-coding transcripts

b | Intronic miRNAs in non-coding transcripts.

c | Intronic miRNAs in protein-coding transcripts.

miRNA biogenesis



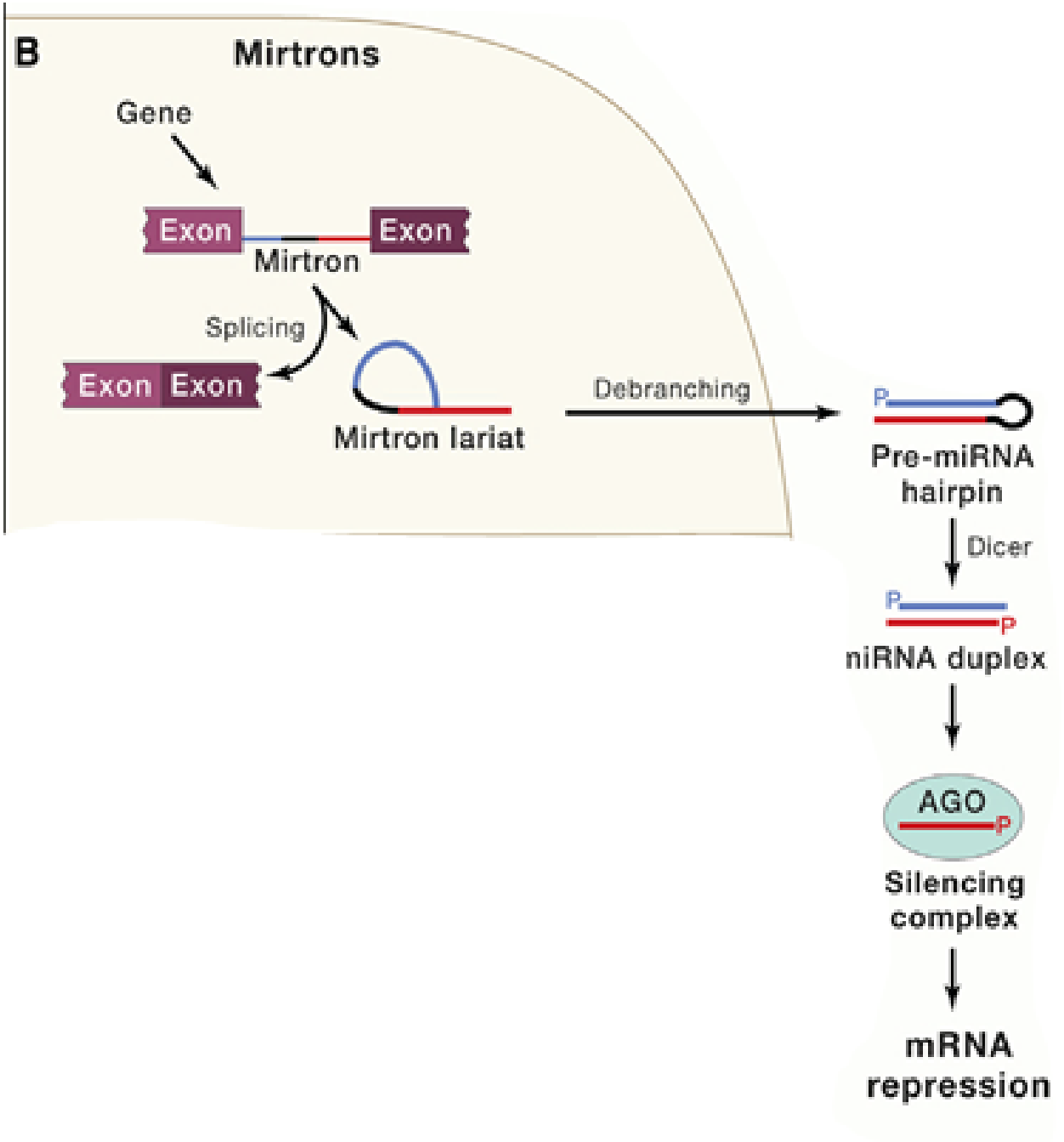
Genomic organization and transcription of miRNA genes

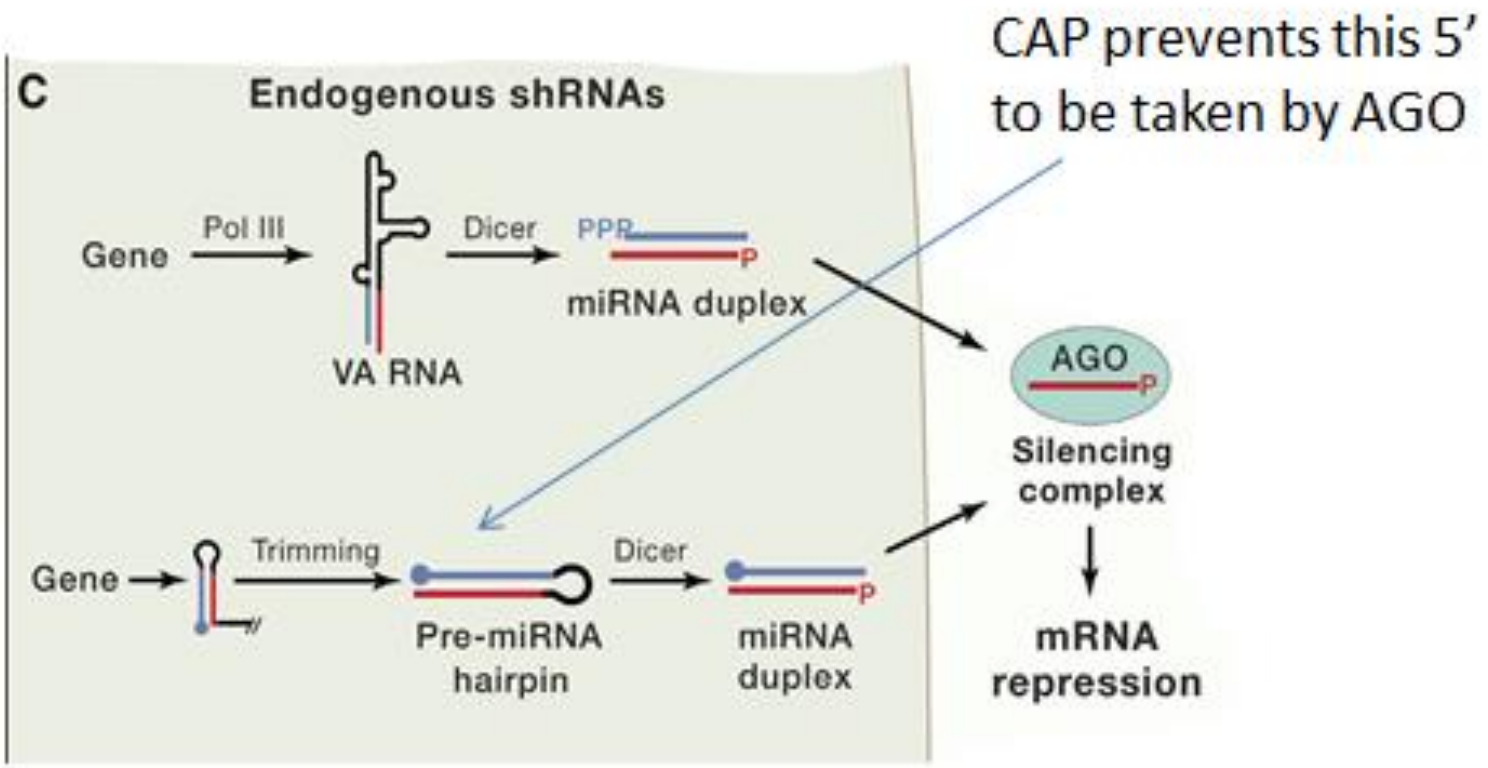
miRNAs are encoded in diverse regions of the genome including both protein coding and non-coding transcription units.

Approximately 50% of miRNAs are derived from noncoding RNA transcripts, while an additional ~40% are located within the introns of protein coding genes.

The majority of miRNAs are transcribed by RNA polymerase (RNA pol) II and bear a 7-methyl guanylate cap at the 5' end and poly (A) tail at the 3' end.

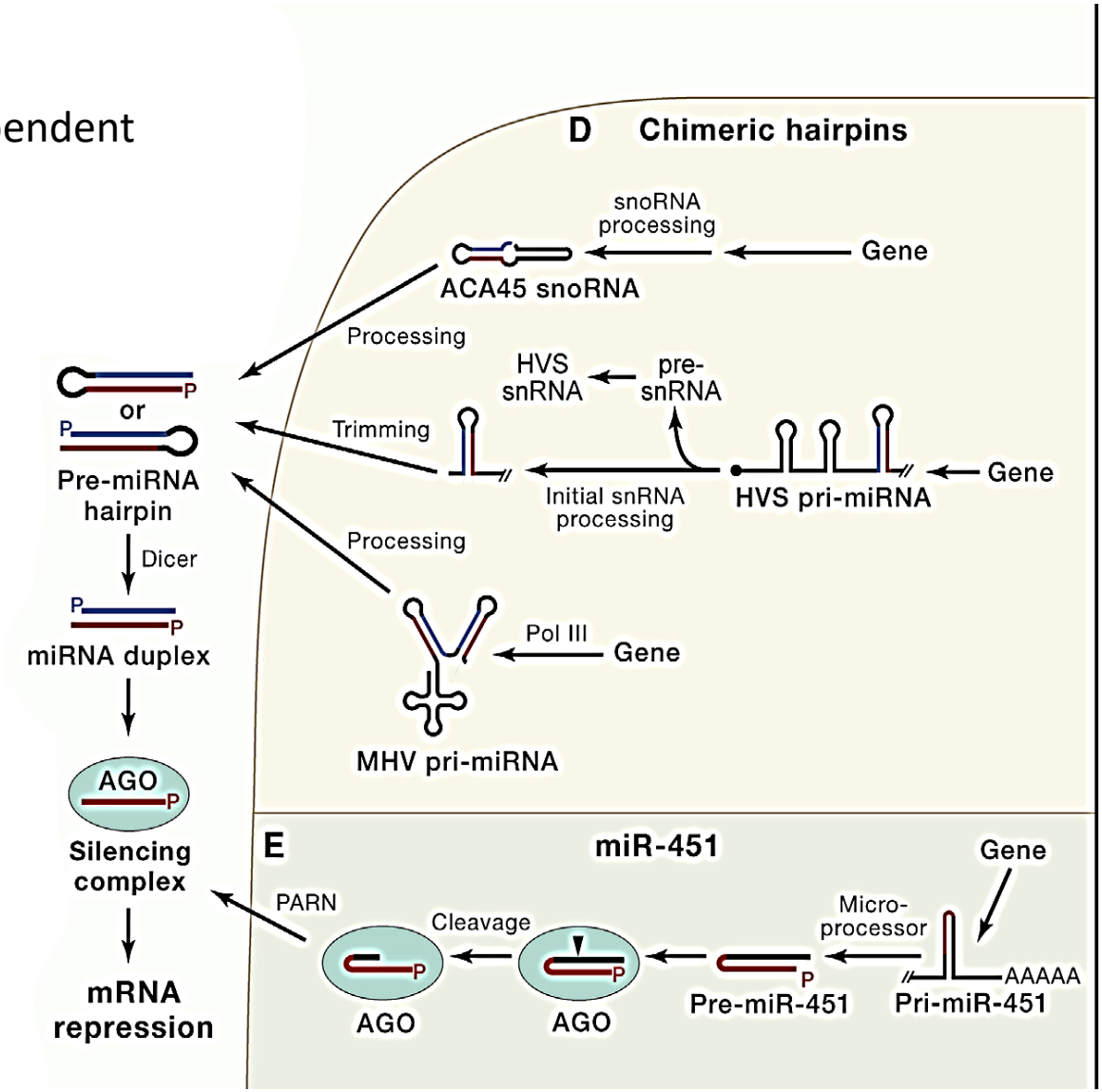
RNA pol III has also been demonstrated to generate the transcripts of a subset of miRNAs.


B
Mirtrons
Gene
Exon
Exon
Mirtron
Splicing
Exon Exon
Mirtron lariat
Debranching
P
Pre-miRNA hairpin
Dicer
P
P
niRNA duplex
AGO
P
Silencing complex
mRNA repression



Drosha-independent

Drosha-independent

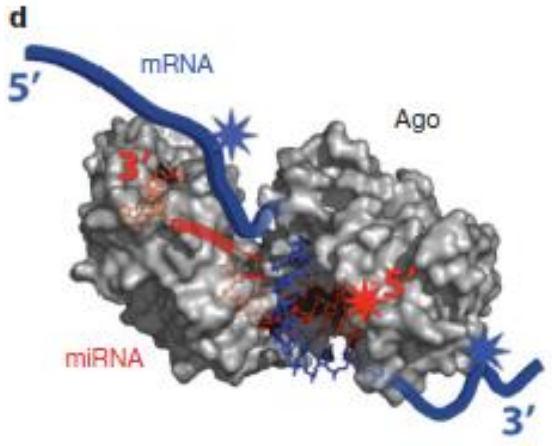


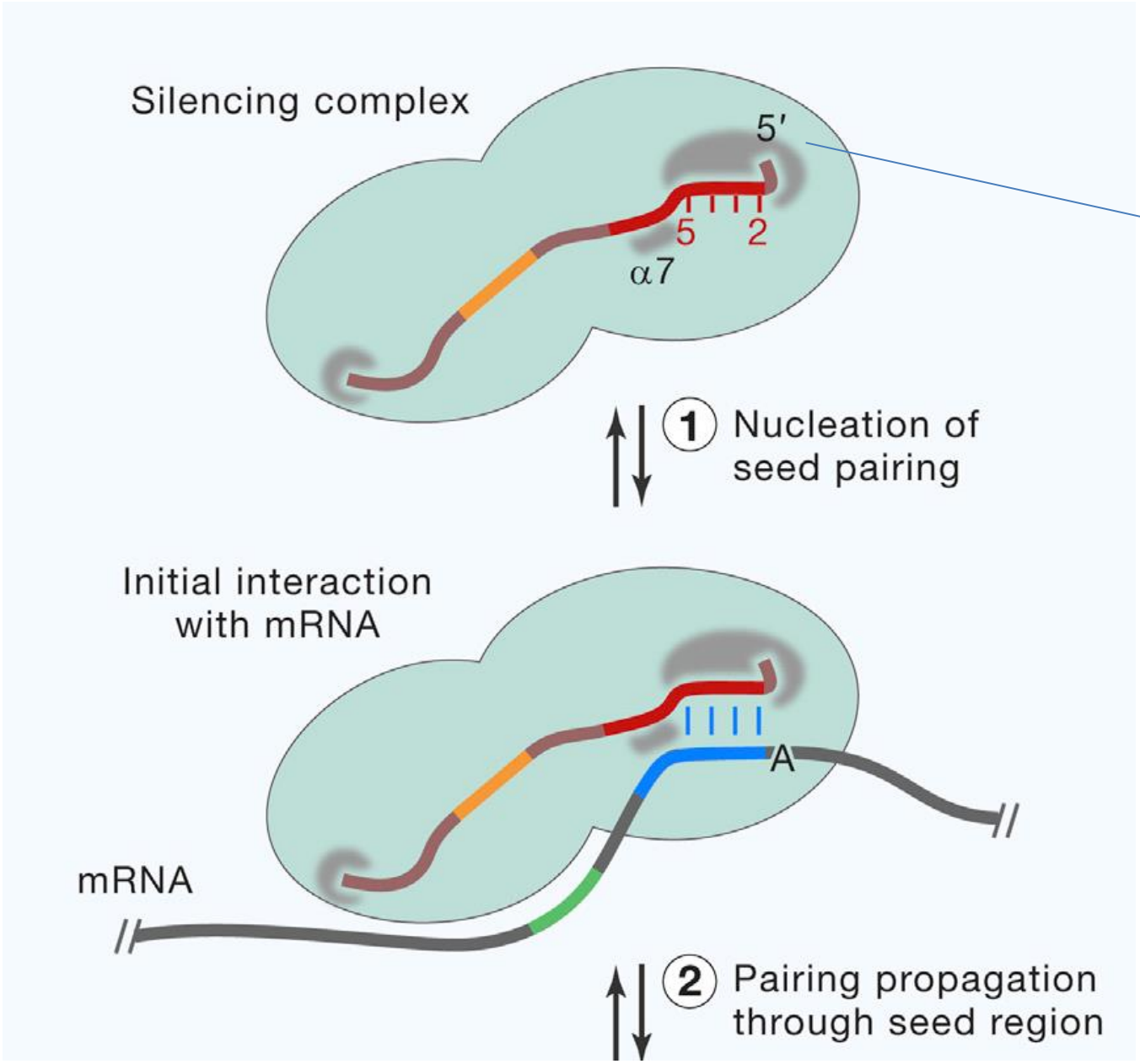
How is the micro-RNA recognized by AGO proteins ?

Which strand is choosen as **guide** ?

Apparently depends on two factors:

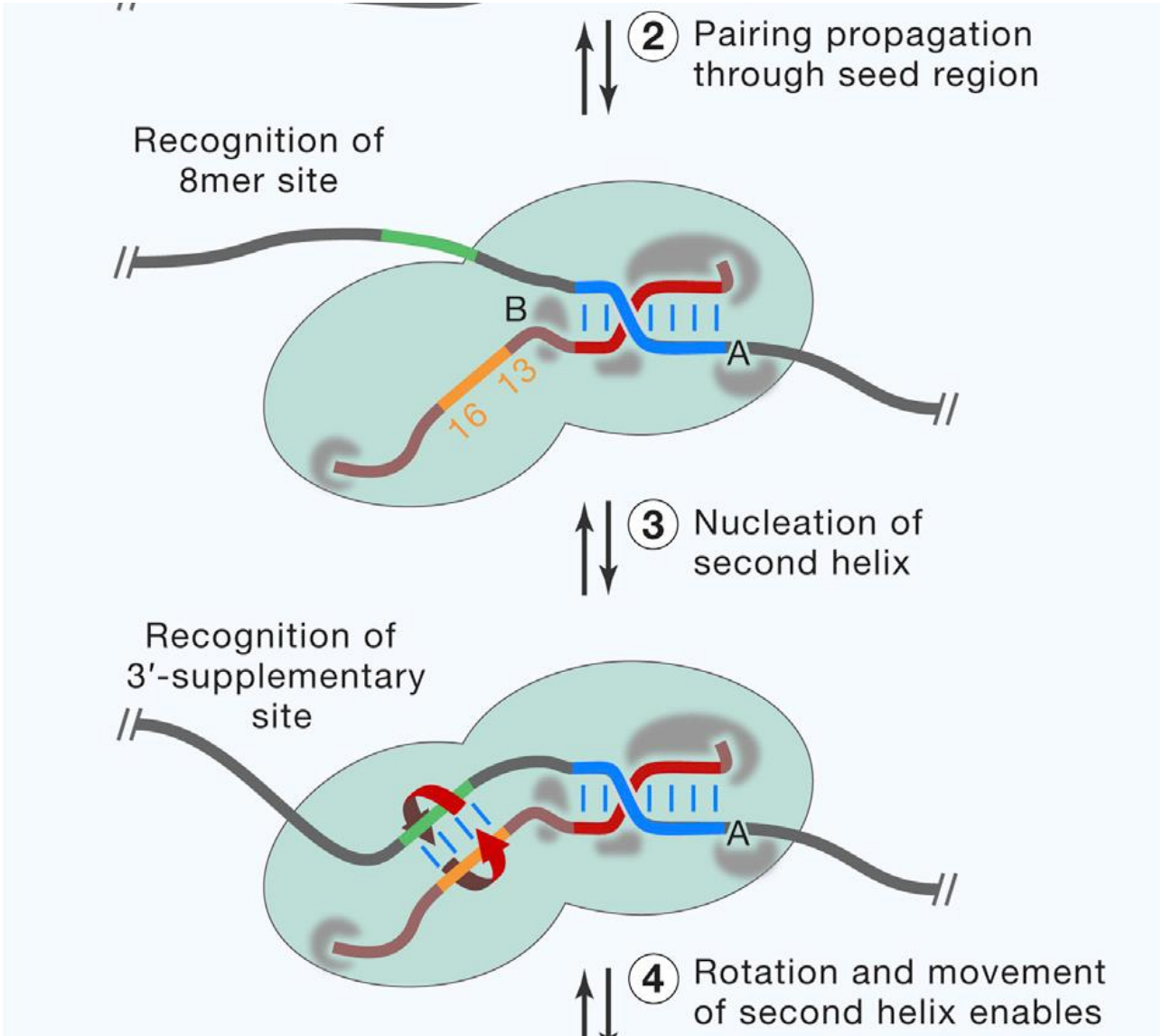
1. an «A» or «U» is preferred as first base (5’-phospho-)
2. the 5’-end with the lowest thermodynamic stability preferred (pA,pU)

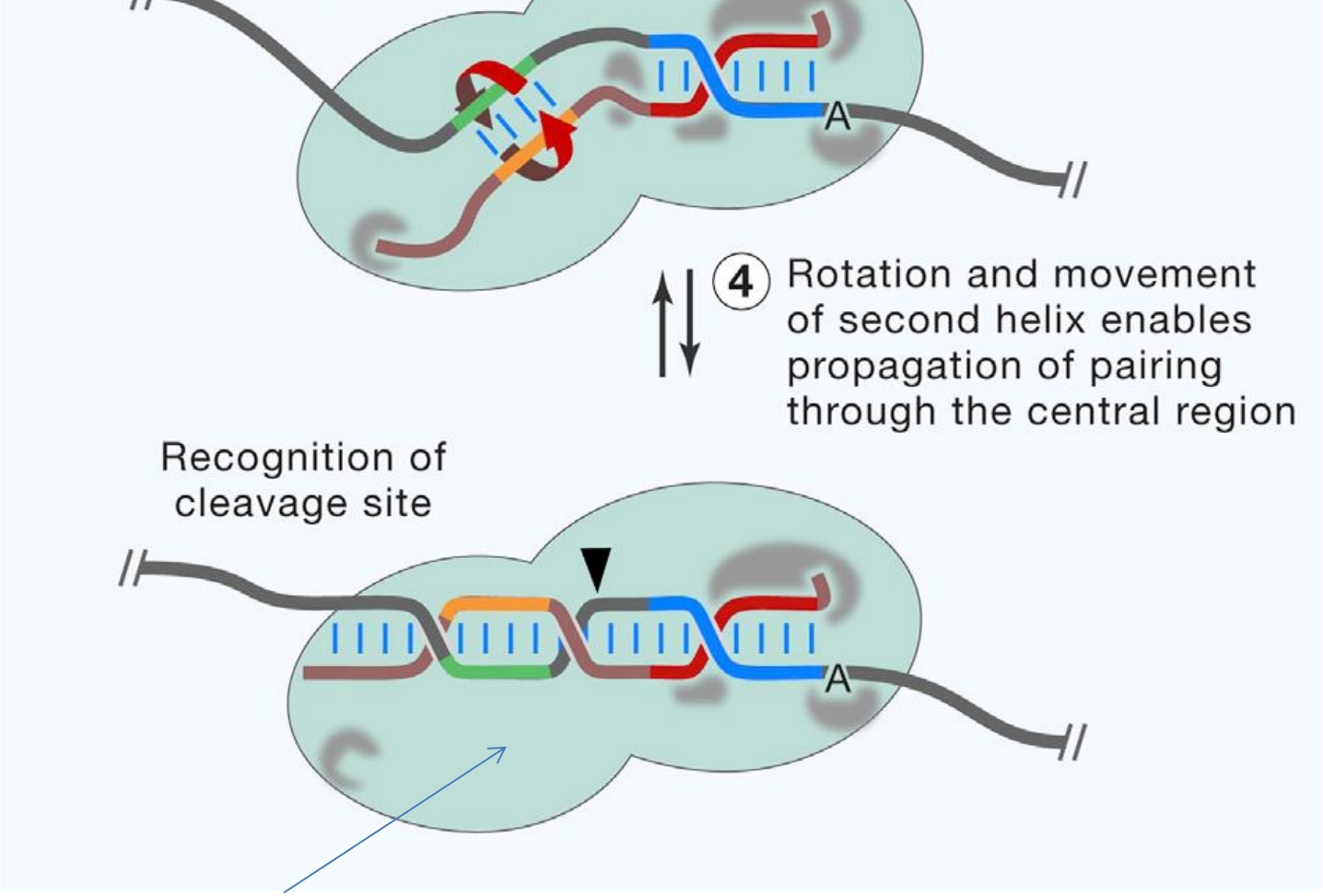




Phospho-required «A» or «U» preferred

First nucleotide bound by AGO


2 Pairing propagation through seed region
Recognition of 8mer site
B
A
16
13
3 Nucleation of second helix
Recognition of 3′-supplementary site
A
4 Rotation and movement of second helix enables



AGO2

How do miRNAs direct target recognition ?

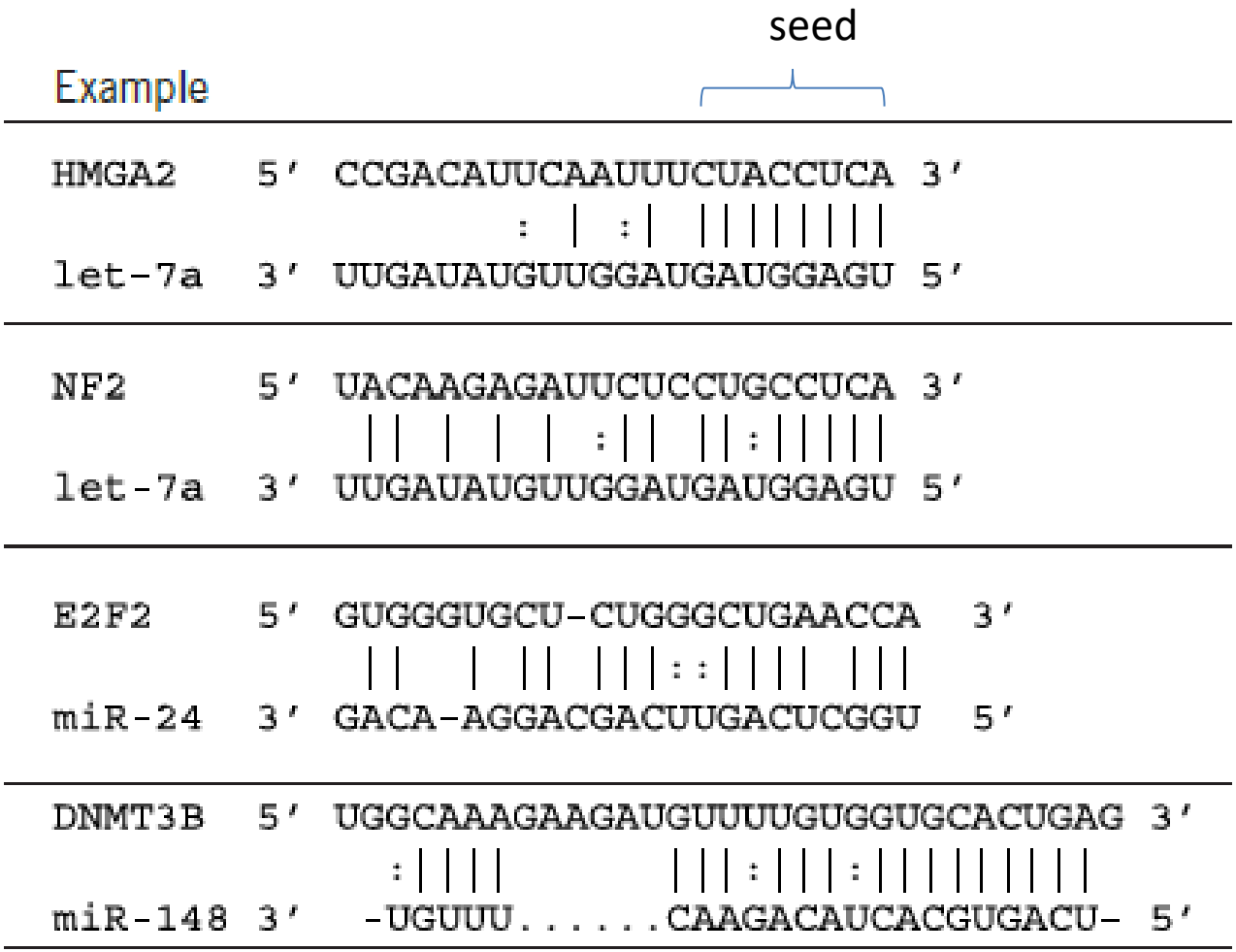
How do micro-RNAs direct target recognition ?

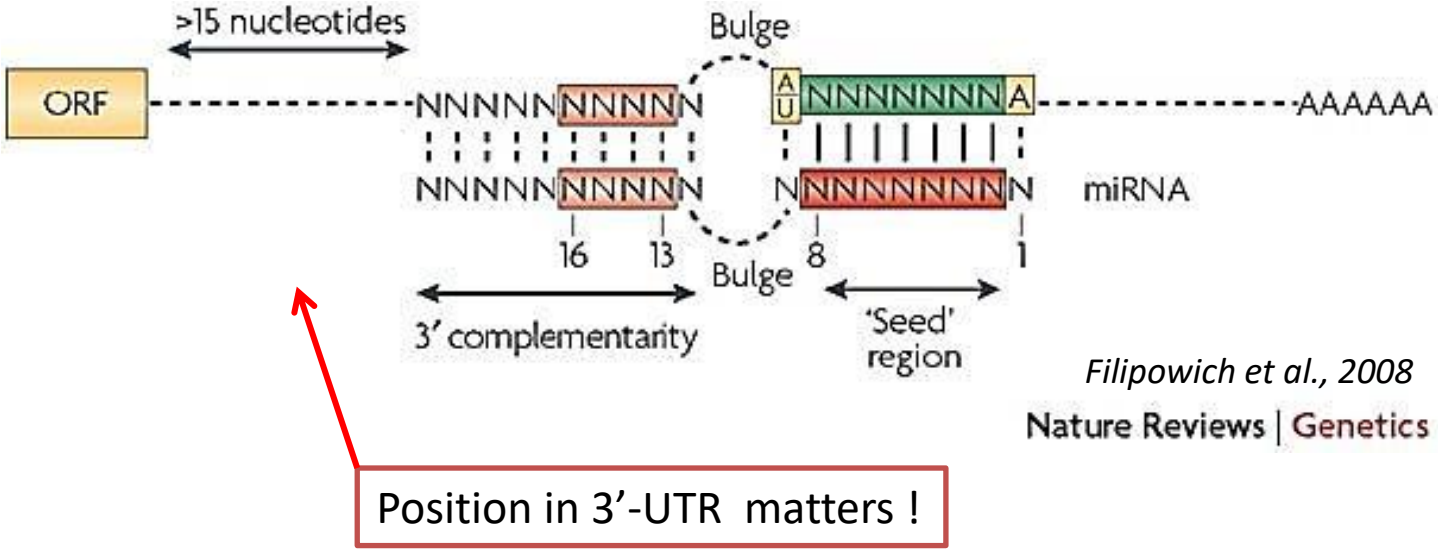
Examples from first discovered miRNA-mRNA pairs.

The 5’-end sequence of miRNA is the main determinant of targeting:

Continuous Watson-Crick pairing is required here (6-8 nt).

Target sequences mainly in mRNA 3’-UTR.



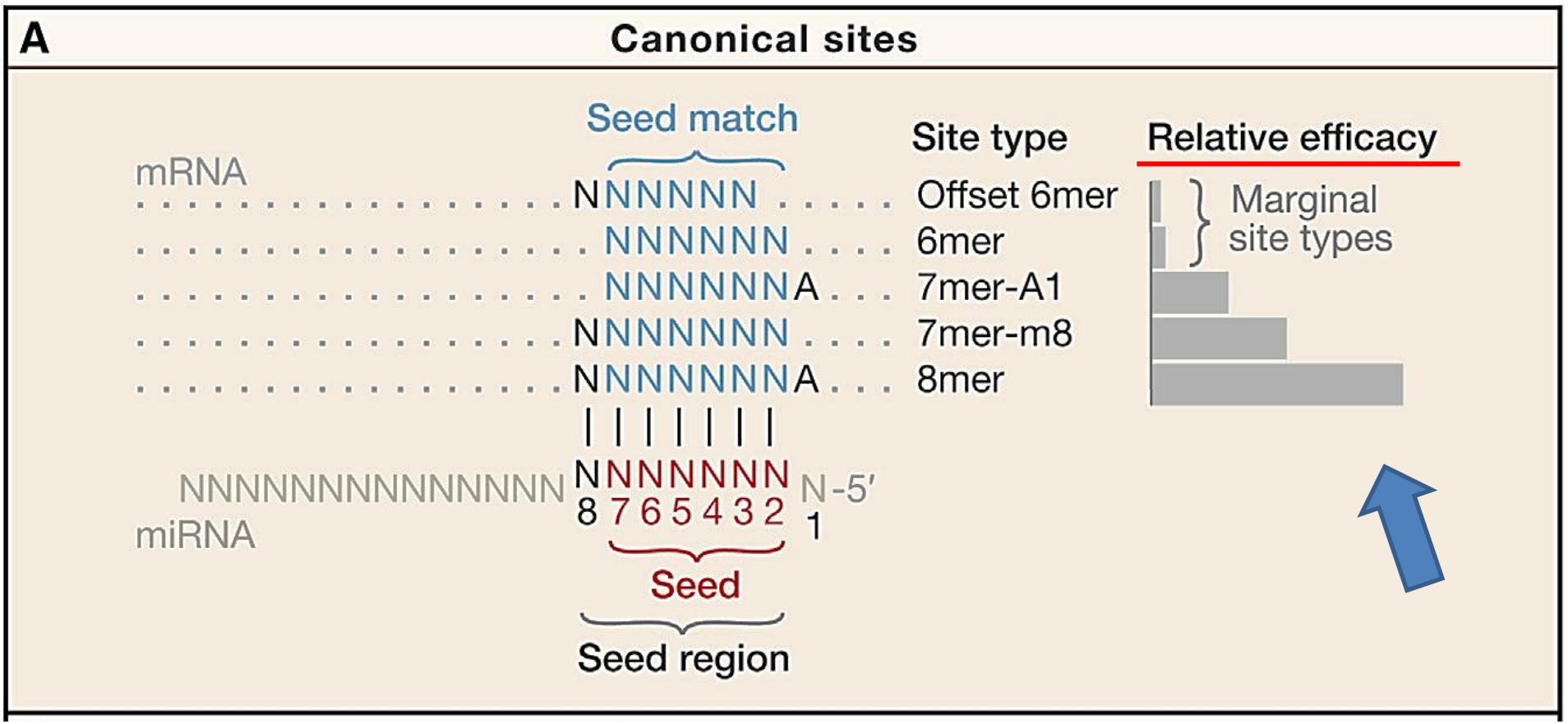


As a rule of thumb, miRNA recognize their targets using 6-8 nt at their 5'-end, followed by variable loop and some additional complementarity towards 3'-end.

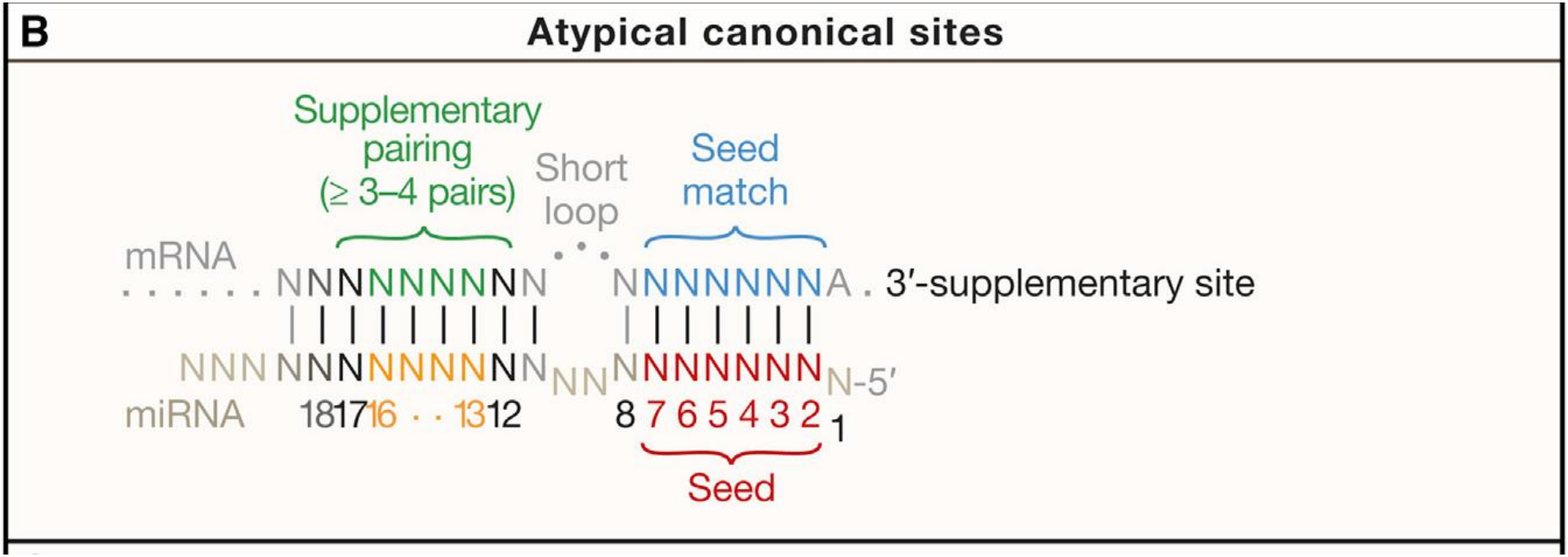
Due to this variable recognition mode, each miRNA recognizes multiple mRNA (or other RNAs) with variable affinity (can calculate from duplex stability).

The union of possible RNA targets of a specific miRNA is called «regulome».

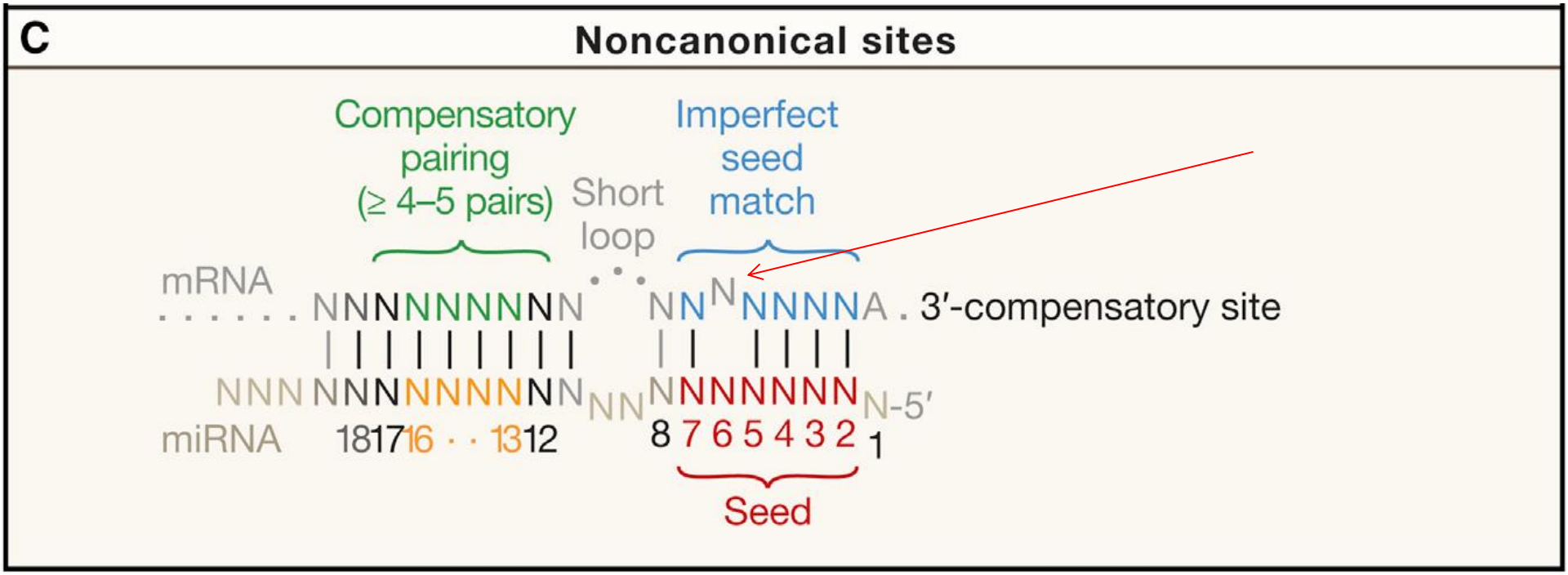
Several algorithms were developed to predict miRNA targets, and are available on websites. However, due to uncertainty of recognition modes, these algorithms work clearly in sub-optimal way.



Canonical sites of mammalian miRNAs. These canonical sites each have 6–7 contiguous Watson–Crick pairs (vertical lines) to the seed region of the miRNA (miRNA positions 2–8). Two of these sites also include an A at position 1. Relative site efficacy in mammalian cells is graphed to the right (log scale). The most effective canonical sites are 7–8 nt sites that include a perfect match to the miRNA seed (positions 2–7, red), whereas the 6 nt sites are the least effective.



The 3’-supplementary site, an atypical type of canonical site.



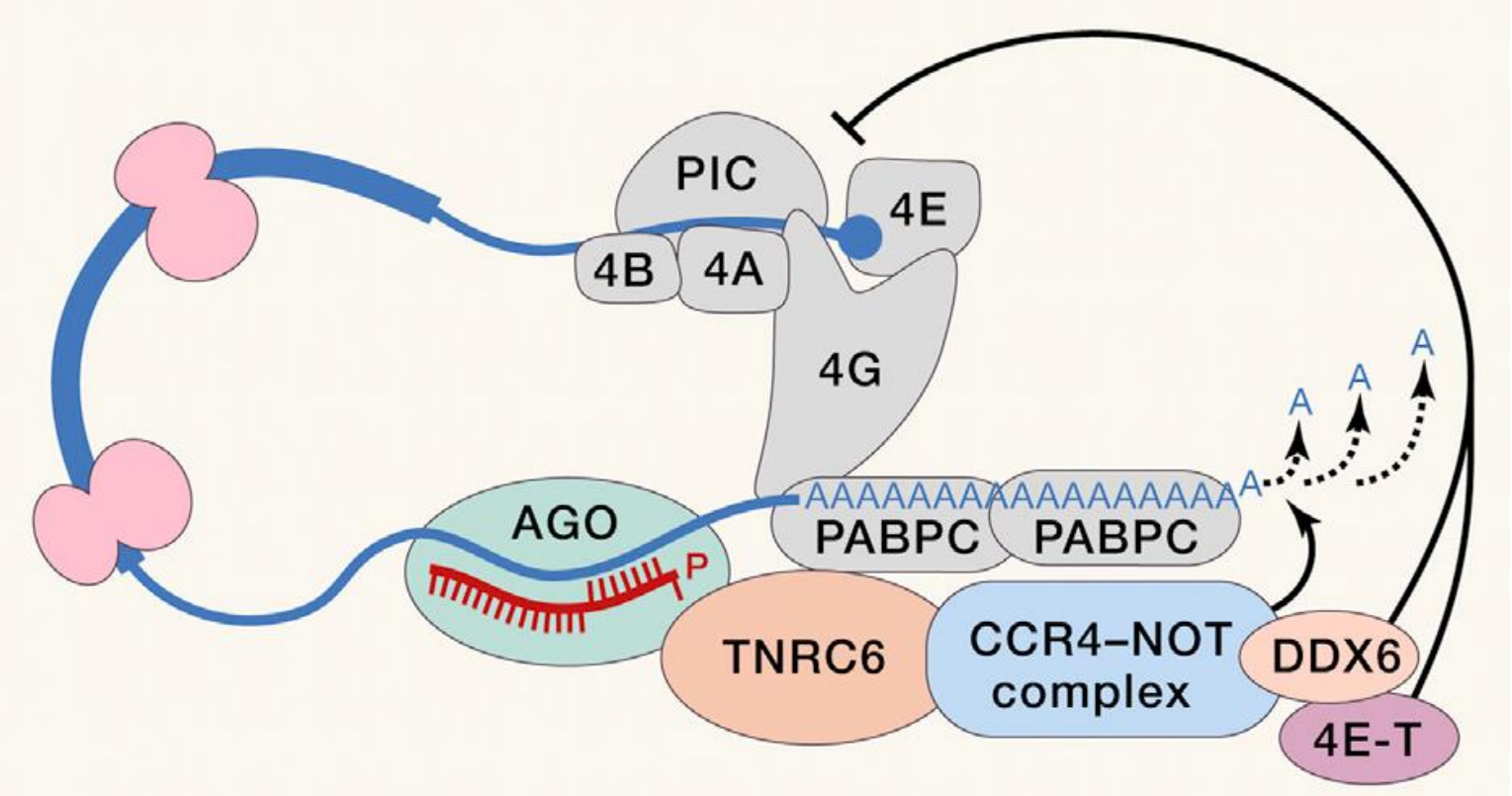
The 3’-compensitory site, a functional type of noncanonical sites

Ch 4 – L4.2

micro RNA

How is post-transcriptional silencing performed ?

The Dominant Mechanisms of miRNA-Guided Repression in Bilaterian Animals



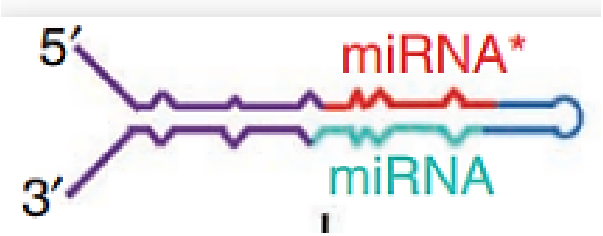
Slicing is determined only when maximal pairing between the micro-RNA and target RNA is present (unusual for miRNA)

This is the condition we see with siRNA (natural or laboratory)

miRNA nomenclature

Names/identifiers in database: hsa-mir-121 (first letter=species; numbering: sequential)
Gene: mir-121
Mature miRNA: miR-121

If there are distinct precursors / genes expressing the same miRNA, add suffix number:
hsa-mir-121-1
hsa-mir-121-2
Mature miRNAs with closely related sequences:
hsa-miR-121a
hsa-miR-121b



In some case, two functional miRNAs derive from the two arms of the same precursor.
If mature forms unbalanced quantitatively: miR-56 the major form (*guide*)
miR-56* the minor form (*passenger*)
If no data or no difference: miR-142-5p (from the 5' arm)
miR-142-3p (from the 3' arm)

(from MiRBase database - http://www.mirbase.org/help/nomenclature.shtml

500 to 600 **verified** miRNAs in Humans. Low % of total predicted.

Some are conserved down to Fish. Conservation very important to predict functional miRNAs.

miRNA divided in Families
– same or similar target recognition.

Each miRNA (family) recognizes hundreds of target mRNA (regulon)

The target sets typically show specifically enriched GO terms

Functions of miRNA

The typical mode of target recognition by miRNA, which is limited to few base pairs, suggests that each miRNA may recognize several mRNAs.

This is exactly what experimentally was seen. Each miRNA down-regulates the expression of a set of mRNAs (regulon).

Strikingly, these targets are enriched by one to several GO terms that are coherent with the physiological action of that miRNA.

Micro RNA have been intensively studied during last ten years.

- ✓ Several miRNAs expressed in each cell
- ✓ Strong tissue- and cell-specificity
- ✓ Developmentally and signal transduction regulated
- ✓ Individual miRNA deletion or overexpression: severe effects on cell physiology
- ✓ miRNA expression profile severely altered in pathologies
- ✓ miRNAs quite stable in serum: possible biomarkers for diseases.

Note: in Humans, 96 miRNA genes (88 families) are conserved among placental mammals but not in Vertebrates. A third of these families are in two paternally imprinted clusters.

Functions of miRNA

miRNA are involved in the control of the expression of virtually all genes

involved in most biological phenomena and processes

important in regulatory circuits for feed-forward and feed-back actions.

Many simple circuits involving one miRNA and one mRNA in cell fate decision processes

Many studies attribute the phenotype of KD/KO to a single miRNA/mRNA interaction

Quite unlikely unless proven

Main proof: remove or mutate MRE from target mRNA and see whether this will phenocopy the miRNA.

---

miRNAs and mRNAs (and most likely also a nuber of lncRNAs) constitute a **network**

while it is possible that individual deletion phenotype can be attributed to one prevalen target (as also in the case of overexpression , e.g. cancer) their physiological role has to be seen in the context of a complex network

How is expression of micro-RNA regulated ?

miRNA and regulatory networks

Regulation of miRNA expression: Transcription

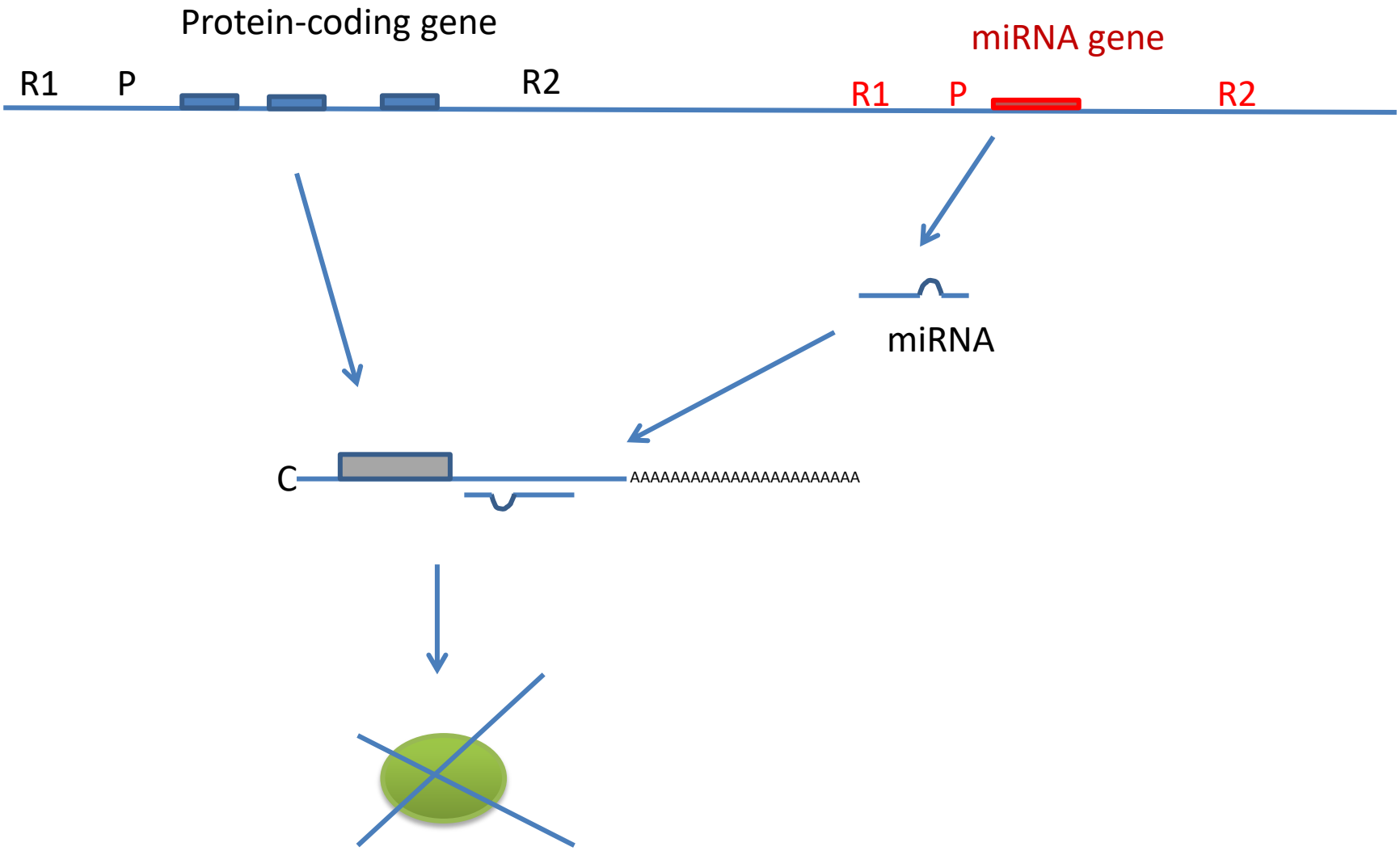
A recent large scale mapping of 175 human miRNA promoters through nucleosome positioning and chromatin immunoprecipitationon-genomic DNA microarray chip (or ChIP-onchip) analysis suggests that

the promoter structure of miRNA genes, including the relative frequencies of CpG islands, TATA box, TFIIB recognition, initiator elements, and histone modifications, is indistinguishable between the promoters of miRNA and mRNA.

Furthermore, DNA binding factors that regulate miRNA transcription largely overlap with those that control protein coding genes.

As a further proof, the expression of specific miRNA is also controlled by signal transduction pathways, in analogy to protein-coding genes.

A two-component regulatory circuit



Experimental:

Suppression of miRNA can be achieved by antisense technologies, especially using nonhydrolyzable oligos (anti-miR).

Over-expression: minigenes expressing pri-miRNA or pre-miRNA in constitutive or inducible vectors. Also synthetic miRNA mimics in transient transfection.

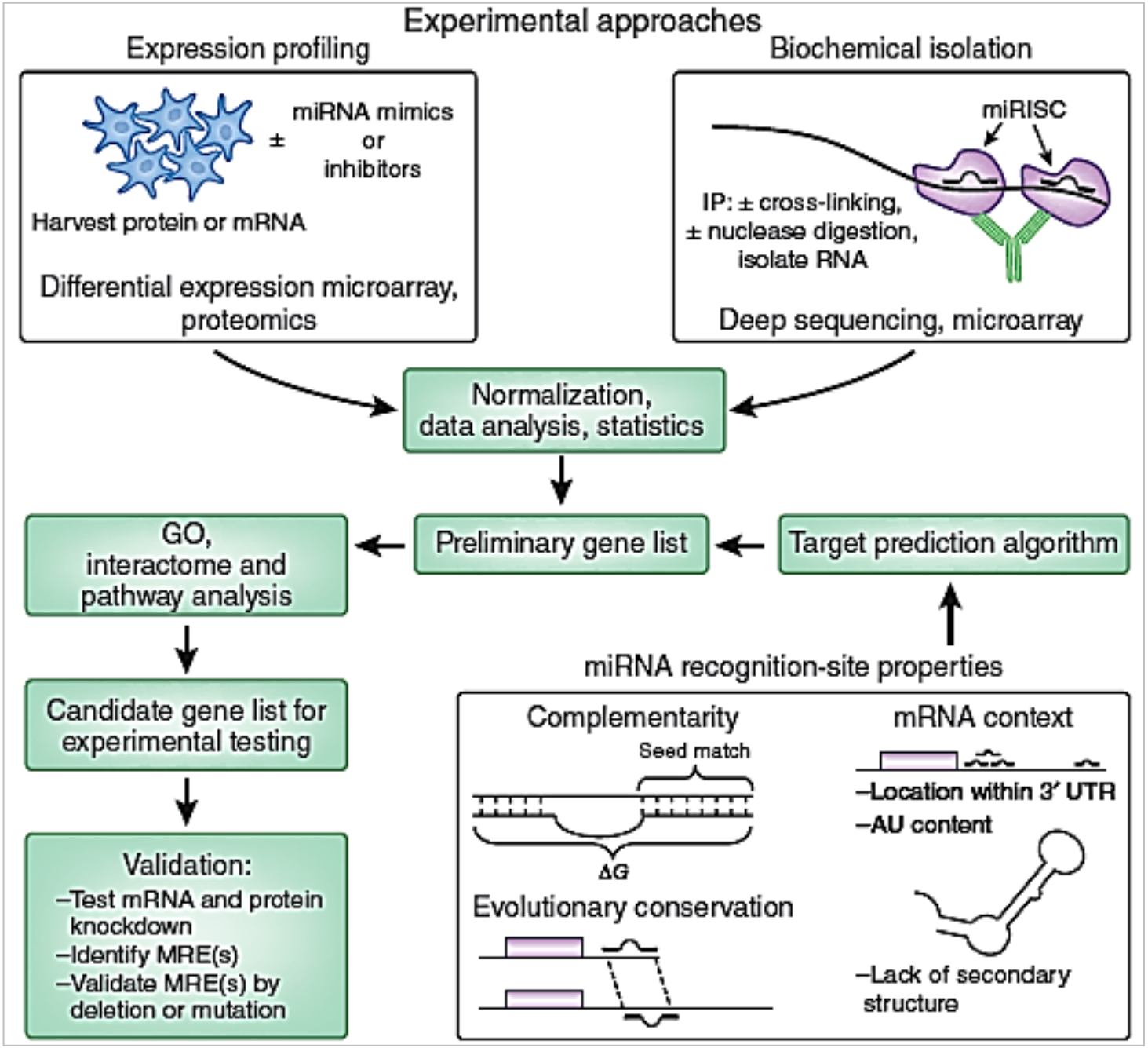
Today: CRISPR-mediated deletion

Expression analysis:

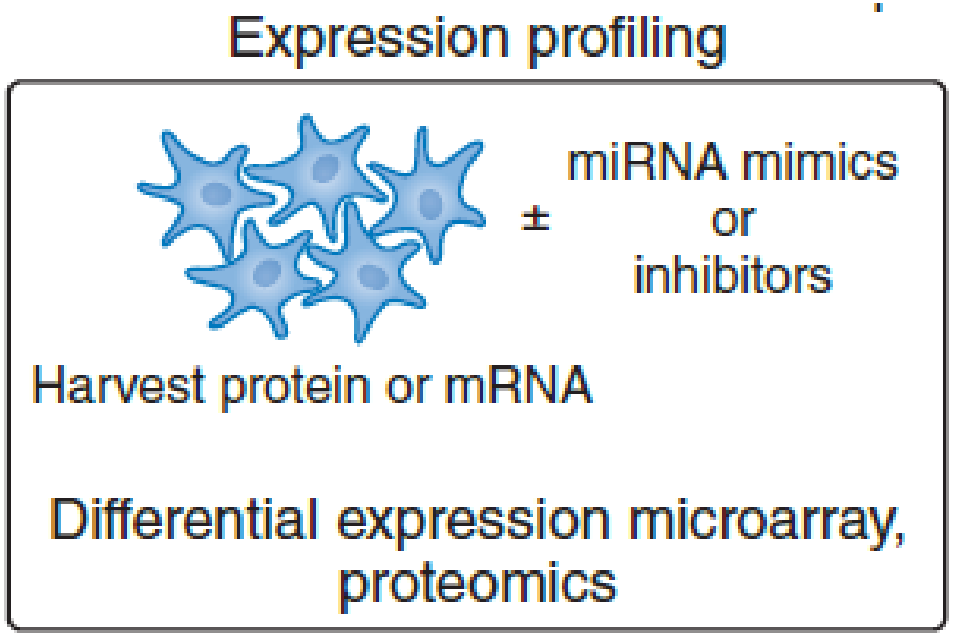
RT-PCR methods available (*problem: they are very short*!)

Microarrays carrying probes for all known miRNAs available, also made of LNA oligos instead of normal oligos to increase hybridization specificity

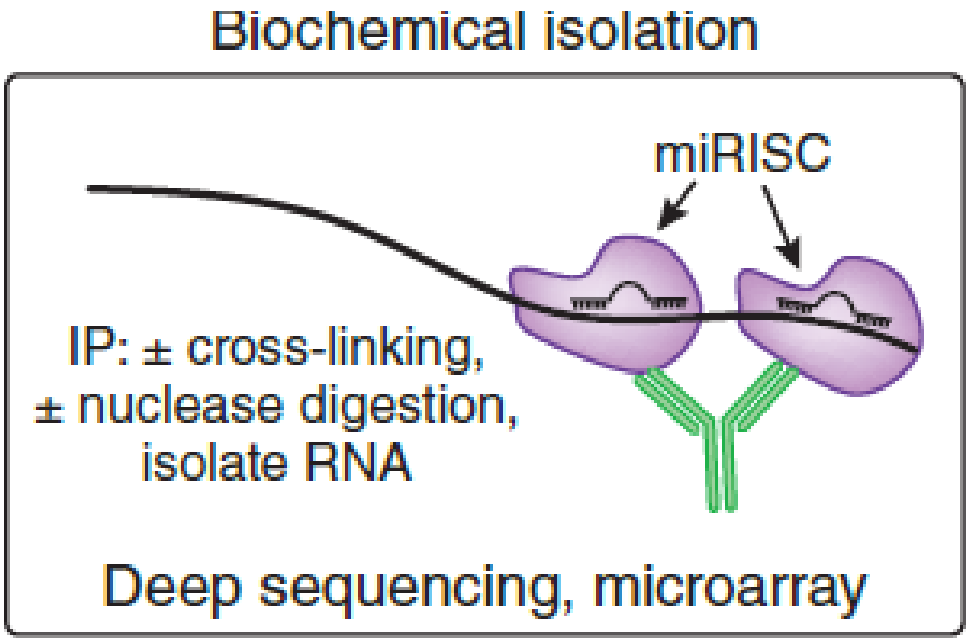
Specific protocol and application for **short RNA-seq NGS**.



MRE=miRNA Response Element

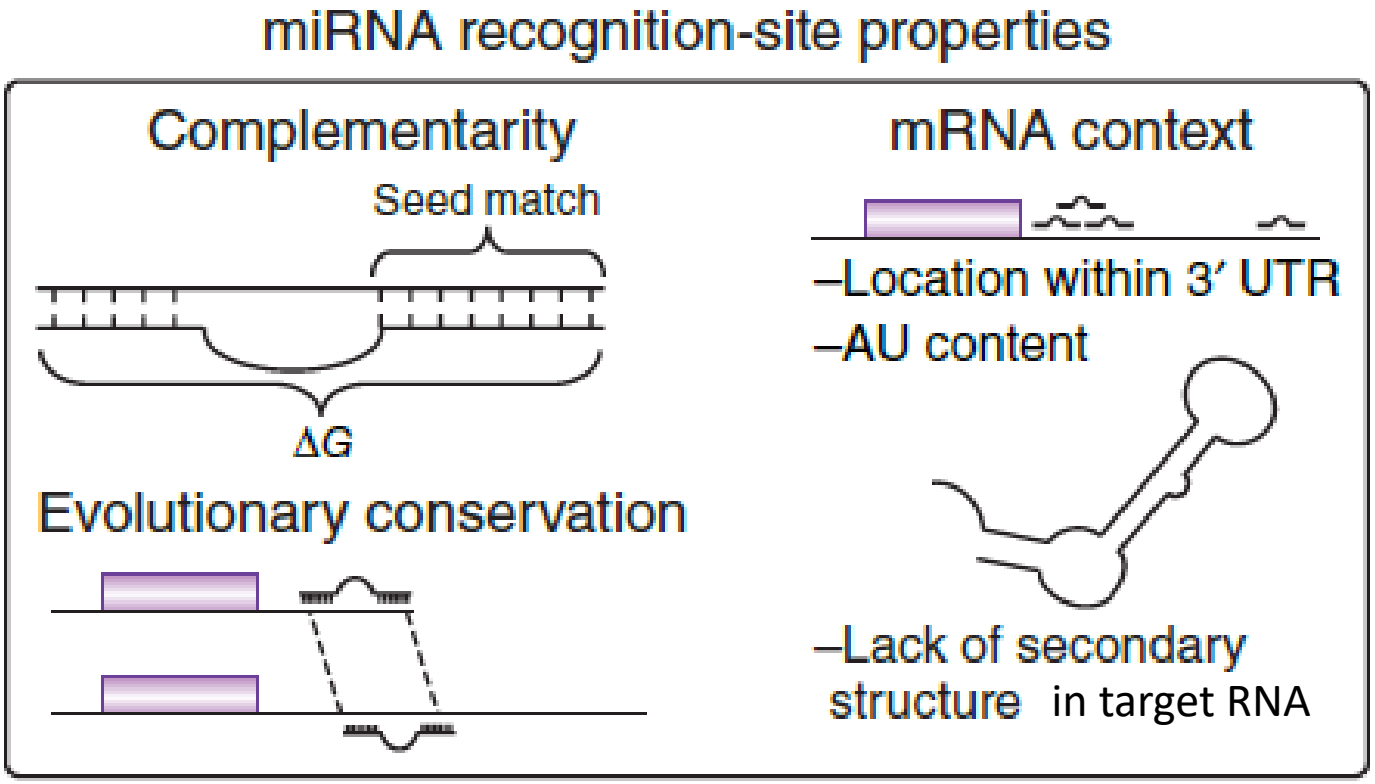


- «antagomir»
- Vectors to express miRNAs
- miRNA mimics
- CRISPR deletion



- RIP
- CLIP
- HITS-CLIP
- PAR-CLIP

Bioinformatics



Number of MREs very important

Most evolved models take into account all these aspects: performance close to that of Ligase-mediated HITS-CLIP **(Your Research Paper No. 4 !**)

Example of the RNA-Protein immunoprecipitation methods (RIP, CLIP, HITS-CLIP and further methods)

Vol 460|23 July 2009|doi:10.1038/nature08170 nature

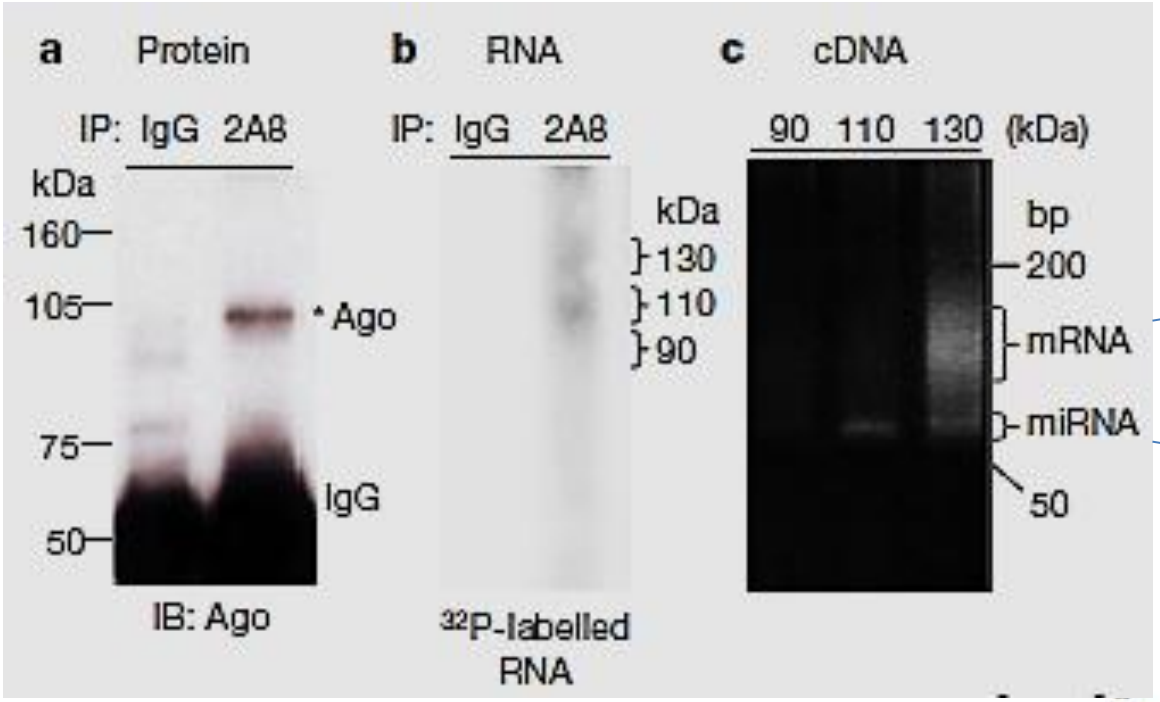
ARTICLES

Argonaute HITS-CLIP decodes microRNA–mRNA interaction maps

Sung Wook Chi[1], Julie B. Zang[1], Aldo Mele[1] & Robert B. Darnell[1]

MicroRNAs (miRNAs) have critical roles in the regulation of gene expression; however, as miRNA activity requires base pairing with only 6–8 nucleotides of messenger RNA, predicting target mRNAs is a major challenge. Recently, high-throughput sequencing of RNAs isolated by crosslinking immunoprecipitation (HITS-CLIP) has identified functional protein–RNA interaction sites. Here we use HITS-CLIP to covalently crosslink native argonaute (Ago, also called Eif2c) protein–RNA complexes in mouse brain. This produced two simultaneous data sets—Ago–miRNA and Ago–mRNA binding sites—that were combined with bioinformatic analysis to identify interaction sites between miRNA and target mRNA. We validated genome-wide interaction maps for miR-124, and generated additional maps for the 20 most abundant miRNAs present in P13 mouse brain. Ago HITS-CLIP provides a general platform for exploring the specificity and range of miRNA action *in vivo*, and identifies precise sequences for targeting clinically relevant miRNA–mRNA interactions.

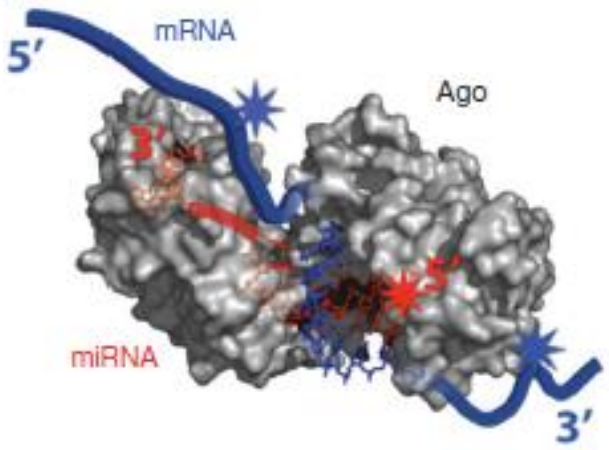
IP using the anti-AGO 2AB antibody reveals different complexes

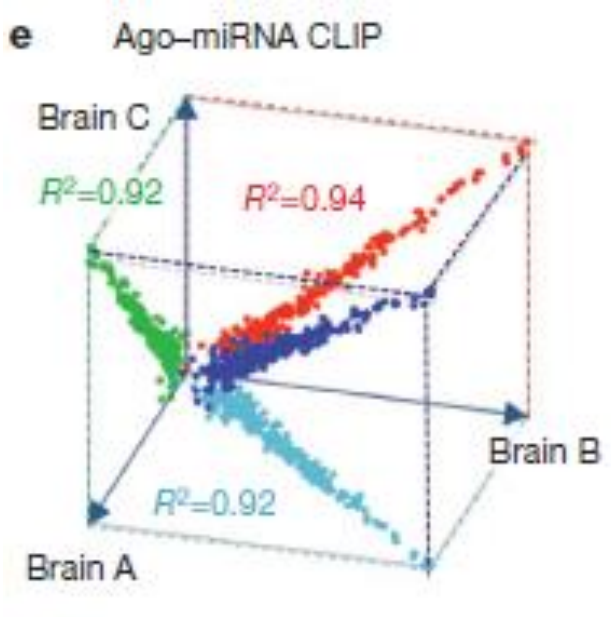


NGS

Short NGS

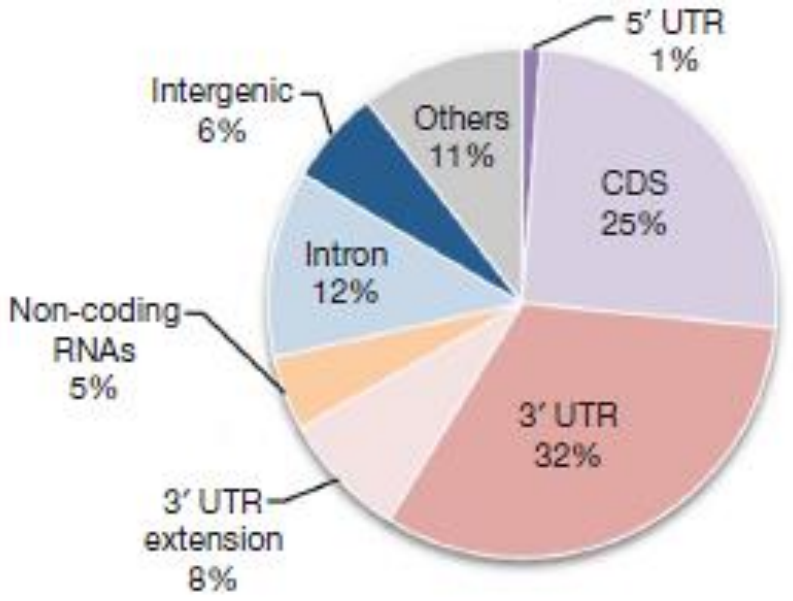
Extract bands, add 36-nt adapters, RT and PCR

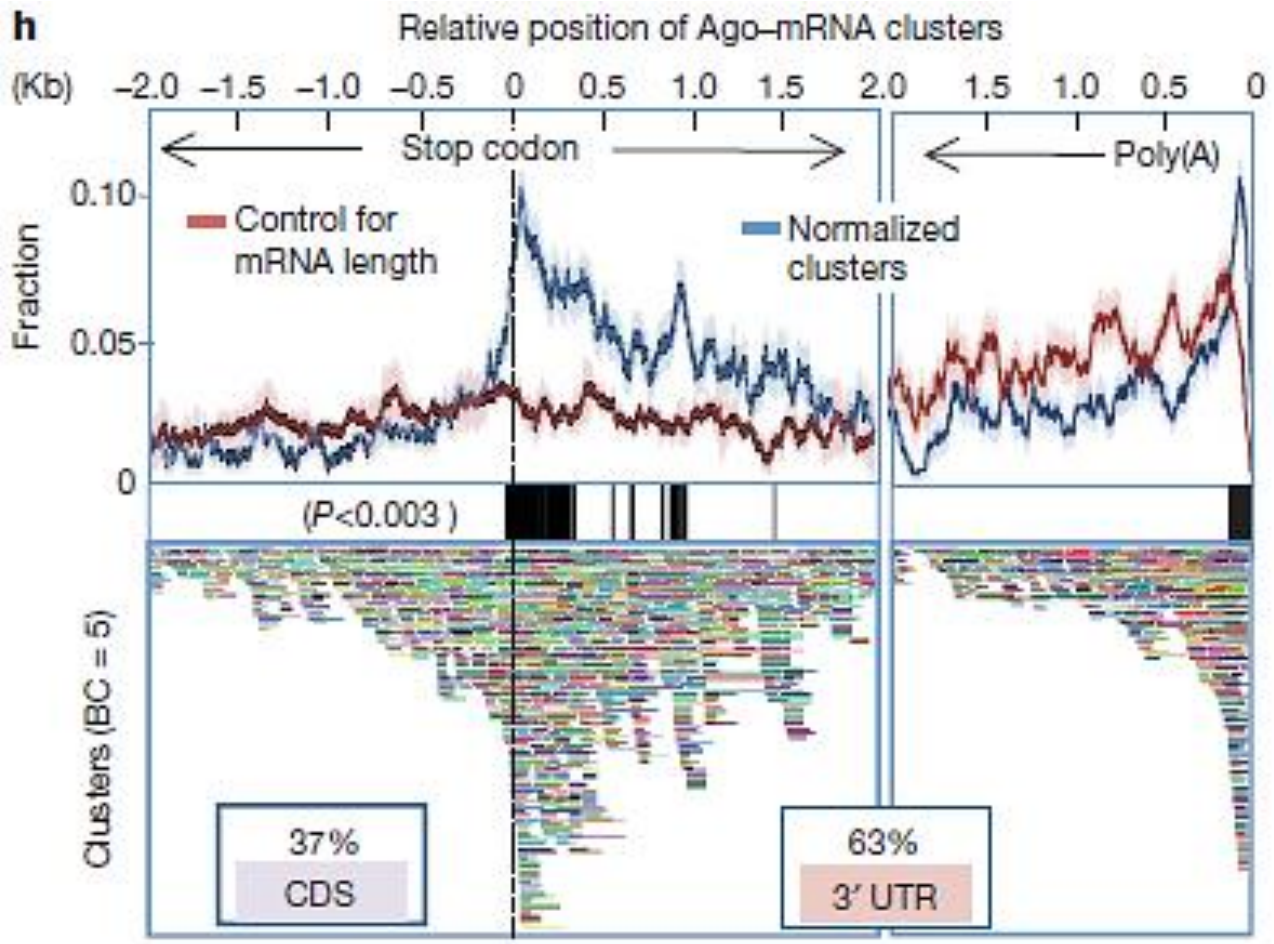




Three replicates from brain

Mapping of targets




h
Relative position of Ago–mRNA clusters
(Kb)
–2.0 –1.5 –1.0 –0.5 0 0.5 1.0 1.5 2.0
1.5 1.0 0.5 0
Stop codon
Poly(A)
Fraction
0.10
0.05
0
Control for mRNA length
Normalized clusters
(P<0.003)
Clusters (BC = 5)
37%
CDS
63%
3′ UTR

Ch 4 – L4.3

targets & circuits

Main problems associated with CLIP and derivatives:

1. cross-linking requires intense, short-wavelength irradiation and this yields quite high non-specific results, due to AGO occasional «touch» with flanking mRNA sequences or completely unrelated RNAs.
2. two separate libraries must be generated, and the lists matched bioinformatically (guess level).
3. target (m)RNA sequence protected by AGO footprint is quite long and gives uncertainty to the possible target sequence.

Trying to bypass this problem, researchers developed PAR-CLIP (next slide)

Transcriptome-wide Identification of RNA-Binding Protein and MicroRNA Target Sites by PAR-CLIP

Markus Hafner,[1,5] Markus Landthaler,[1,4,5] Lukas Burger,[2] Mohsen Khorshid,[2] Jean Hausser,[2] Philipp Berninger,[2] Andrea Rothballer,[1] Manuel Ascano, Jr.,[1] Anna-Carina Jungkamp,[1,4] Mathias Munschauer,[1] Alexander Ulrich,[1] Greg S. Wardle,[1] Scott Dewell,[3] Mihaela Zavolan,[2,*] and Thomas Tuschl[1,*]

[1]Howard Hughes Medical Institute, Laboratory of RNA Molecular Biology, The Rockefeller University, 1230 York Avenue, Box 186, New York, NY 10065, USA

[2]Biozentrum der Universität Basel and Swiss Institute of Bioinformatics (SIB), Klingelbergstr. 50-70, CH-4056 Basel, Switzerland

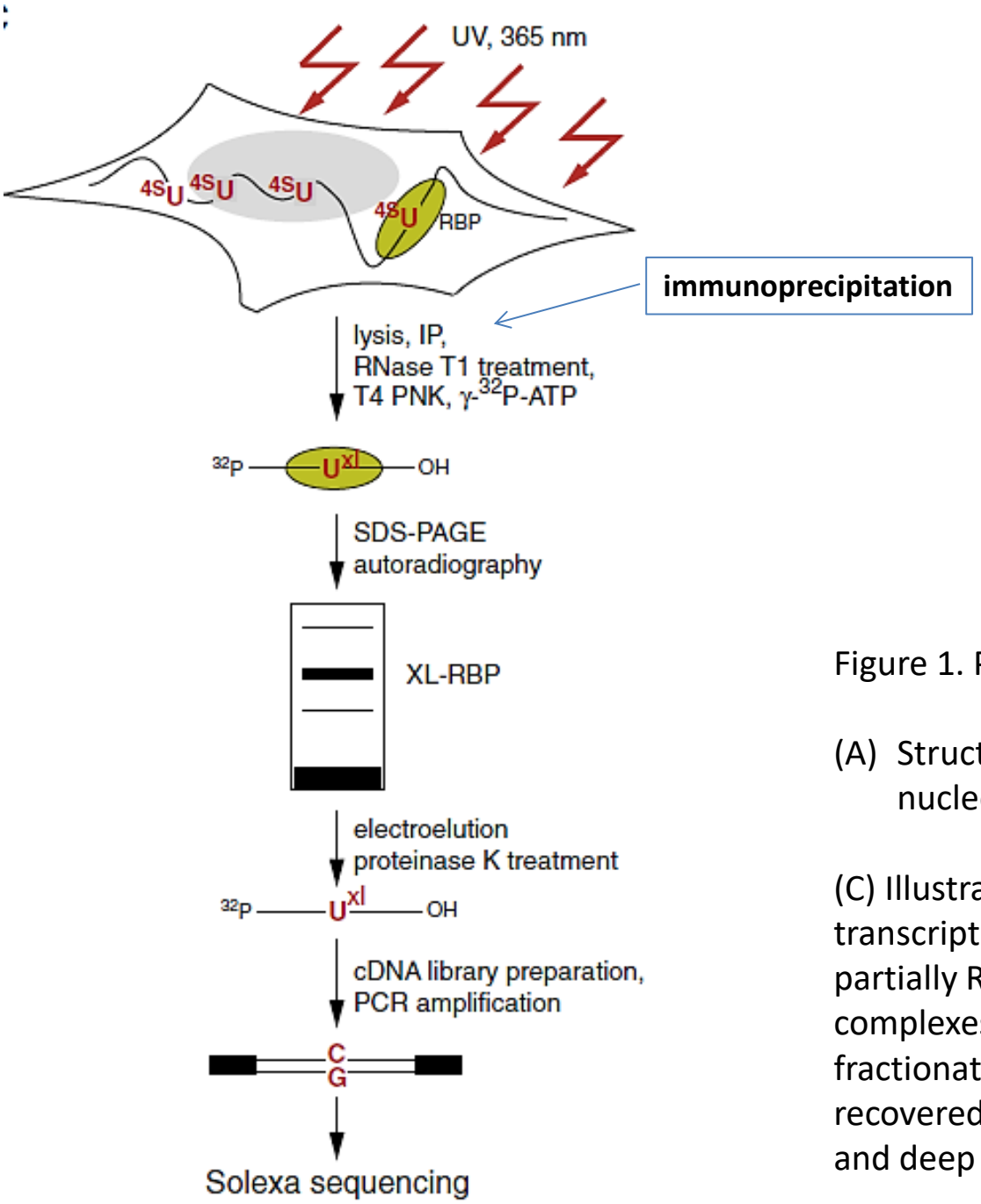
[3]Genomics Resource Center, The Rockefeller University, 1230 York Avenue, Box 241, New York, NY 10065, USA

[4]Present address: Berlin Institute for Medical Systems Biology, Max-Delbrück-Center for Molecular Medicine, 13125 Berlin, Germany

[5]These authors contributed equally to this work

*Correspondence: mihaela.zavolan@unibas.ch (M.Z.), ttuschl@rockefeller.edu (T.T.)

RNA transcripts are subject to posttranscriptional gene regulation involving hundreds of RNA-binding proteins (RBPs) and microRNA-containing ribonucleoprotein complexes (miRNPs) expressed in a cell-type dependent fashion. We developed a cell-based crosslinking approach to determine at high resolution and transcriptome-wide the binding sites of cellular RBPs and miRNPs. The crosslinked sites are revealed by thymidine to cytidine transitions in the cDNAs prepared from immunopurified RNPs of 4-thiouridine-treated cells. We determined the binding sites and regulatory consequences for several intensely studied RBPs and miRNPs, including PUM2, QKI, IGF2BP1-3, AGO/EIF2C1-4 and TNRC6A-C. Our study revealed that these factors bind thousands of sites containing defined sequence motifs and have distinct preferences for exonic versus intronic or coding versus untranslated transcript regions. The precise mapping of binding sites across the transcriptome will be critical to the interpretation of the rapidly emerging data on genetic variation between individuals and how these variations contribute to complex genetic diseases.



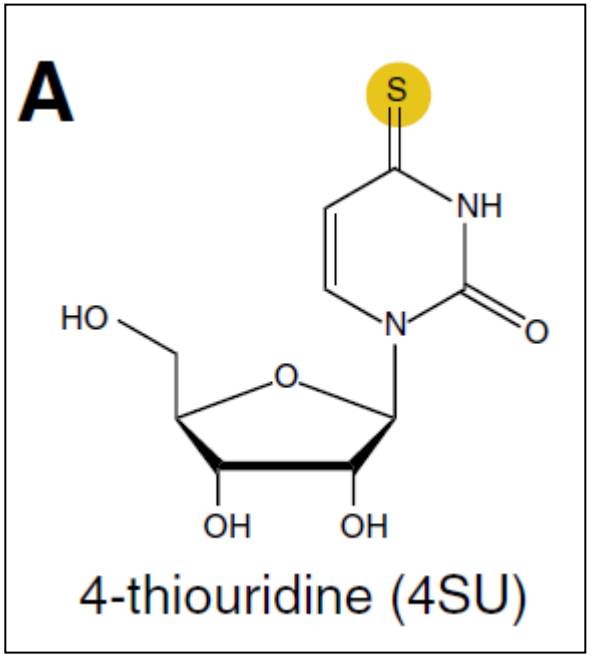
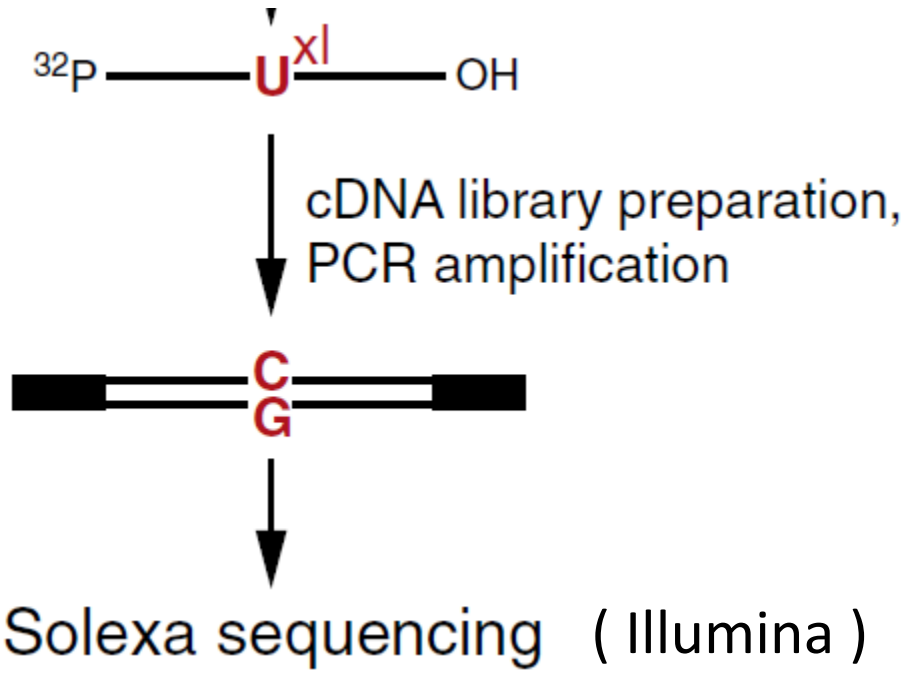


Figure 1. PAR-CLIP Methodology.

(A) Structure of photoactivatable nucleosides.

(C) Illustration of PAR-CLIP. 4SU-labeled transcripts were crosslinked to RBPs and partially RNase-digested RNA-protein complexes were immunopurified and size-fractionated. RNA molecules were recovered and converted into a cDNA library and deep sequenced.



The «U» that was directly cross-linked to the RBP is identified since it is converted to «C» and consequently the targets are univocally identified.

Major problems for these approaches is that they are indirect, i.e. they are based on the generation of **separate libraries**, one for miRNAs and the other for targets (mRNA, lncRNAs, etc). Matching is always based on complementarity searches.

For this reason, Darnell's group developed a strategy to ligate miRNAs and targets and sequence them together, called CLEAR-CLIP

nature COMMUNICATIONS

Research Paper 4

ARTICLE

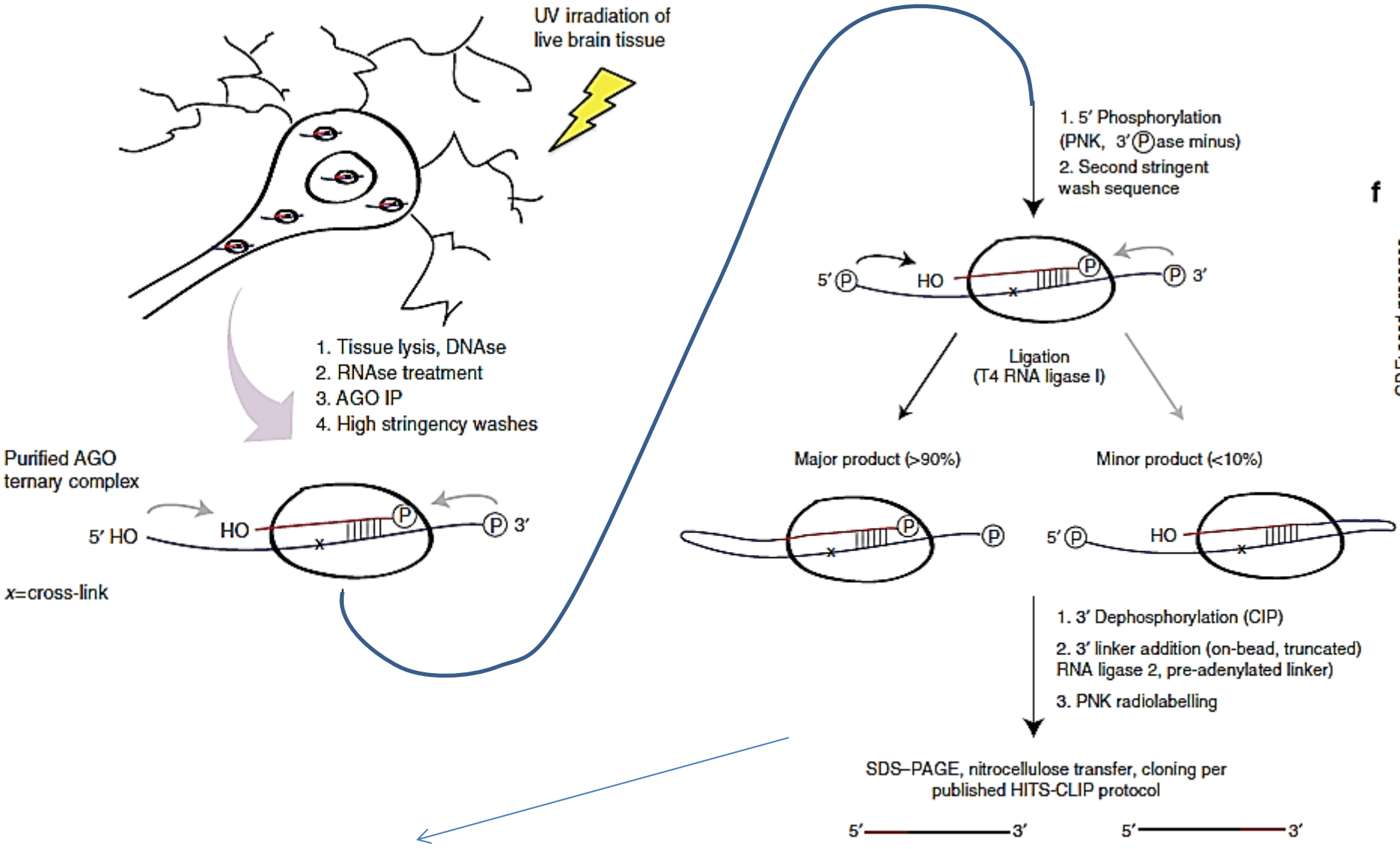
Received 4 May 2015 | Accepted 12 Oct 2015 | Published 25 Nov 2015

DOI: 10.1038/ncomms9864 OPEN

miRNA–target chimeras reveal miRNA 3′-end pairing as a major determinant of Argonaute target specificity

Michael J. Moore[1], Troels K.H. Scheel[2,3,4], Joseph M. Luna[1,2], Christopher Y. Park[1,5], John J. Fak[1], Eiko Nishiuchi[2], Charles M. Rice[2] & Robert B. Darnell[1,5]

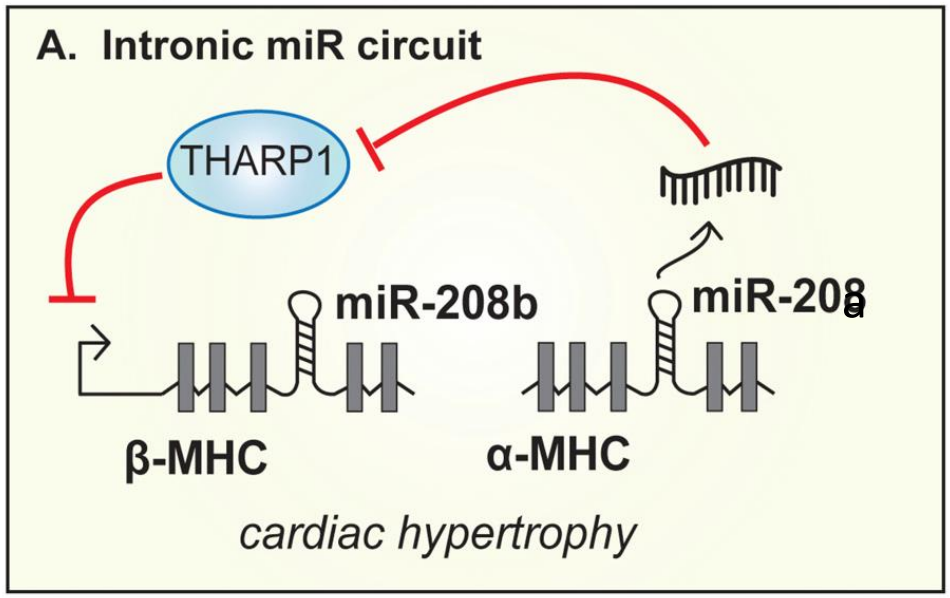
Here we report a modified AGO HITS-CLIP strategy termed CLEAR (covalent ligation of endogenous Argonaute-bound RNAs)-CLIP, which enriches miRNAs ligated to their endogenous mRNA targets. CLEAR-CLIP mapped ca. 130,000 endogenous miRNA–target interactions in mouse brain and B40,000 in human hepatoma cells. Motif and structural analysis define expanded pairing rules for over 200 mammalian miRNAs. Most interactions combine seed-based pairing with distinct, miRNA-specific patterns of auxiliary pairing. At some regulatory sites, this specificity confers distinct silencing functions to miRNA family members with shared seed sequences but divergent 30-ends. This work provides a means for explicit biochemical identification of miRNA sites in vivo, leading to the discovery that miRNA 3’-end pairing is a general determinant of AGO binding specificity.



One library for NGS with joint target-miRNA

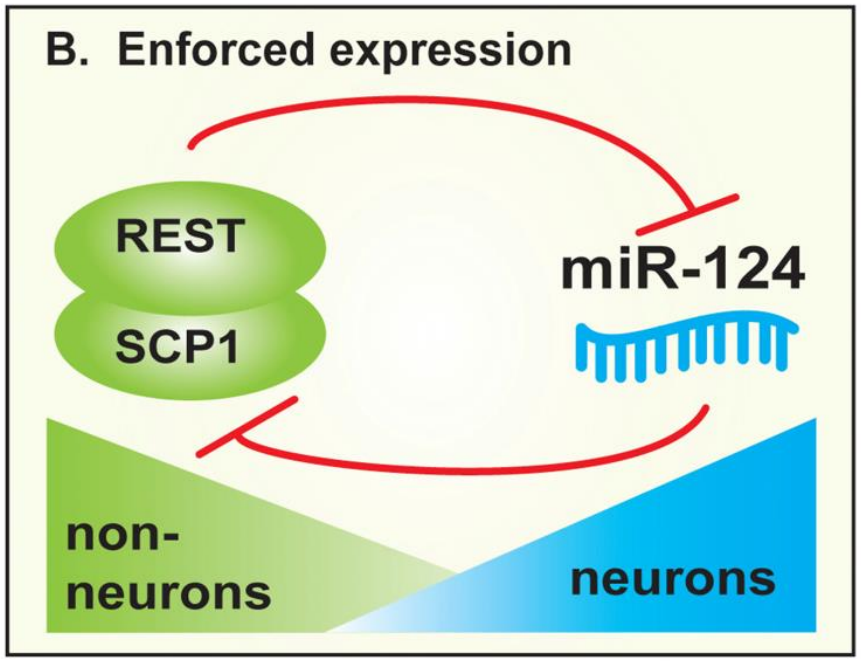
circuits involving miRNAs

miRNA regulatory circuits



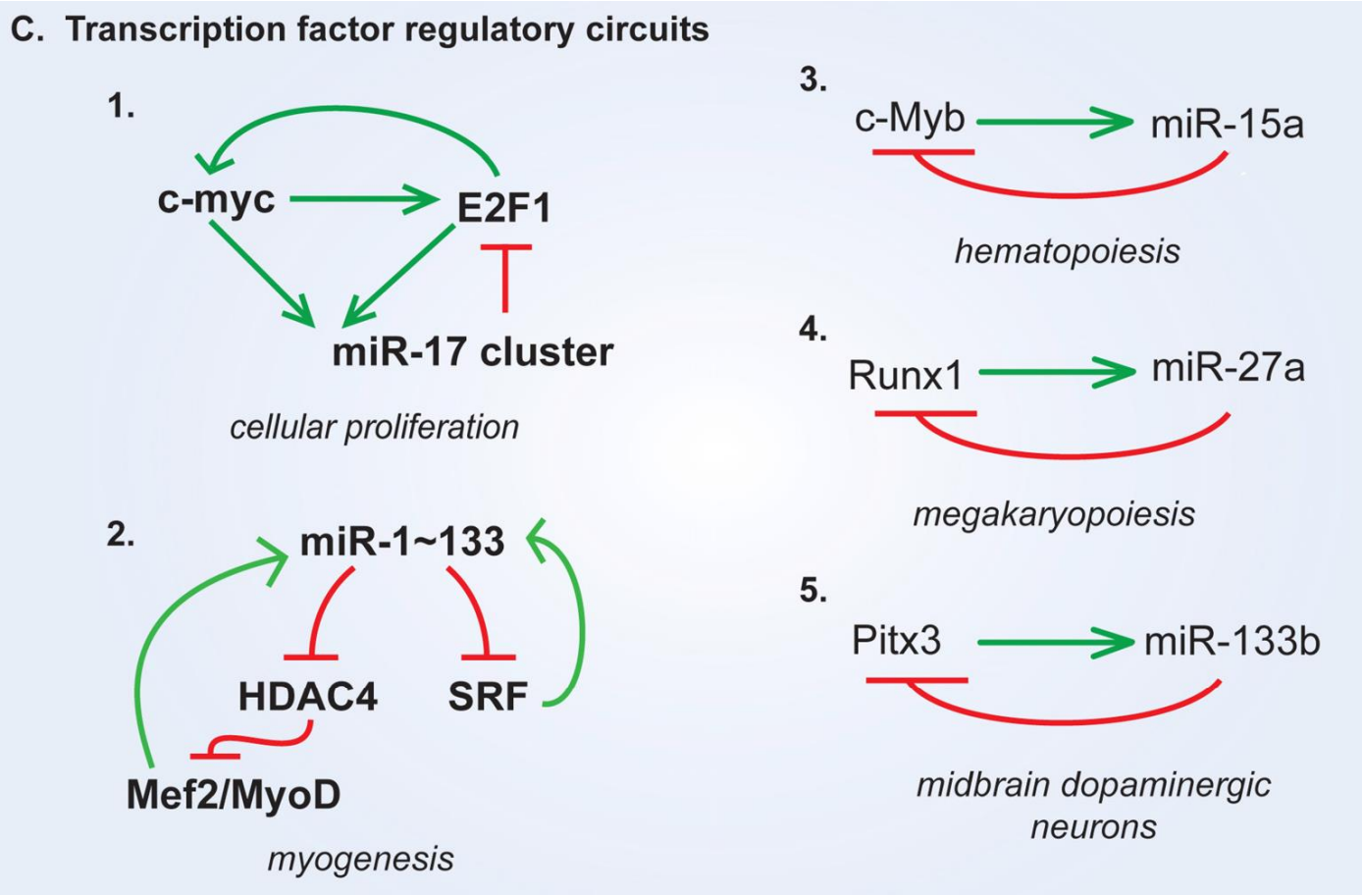
The cardiac specific miR-208 family is encoded within the introns of myosin heavy chain (MHC) genes. miR-208a targets THARP1, and will reduce its level, thus increasing expression of β-MHC.

This is an auto-boosting circuit that is found altered in cardiac hypertrophy.



The expression of miR-124 is negatively regulated by the binding of the RE1 silencing transcription (REST) factor to the promoter in non-neuronal cells

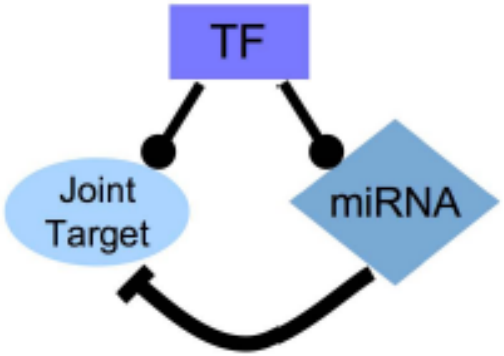
In neurons, miR-124 represses translation of SCP1 that is one component of the REST complex.

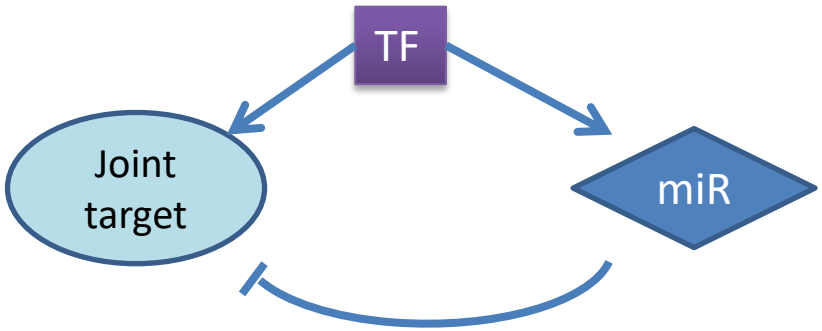


Examples of feed-back regulation of microRNA transcription through the repression of transcription factors.

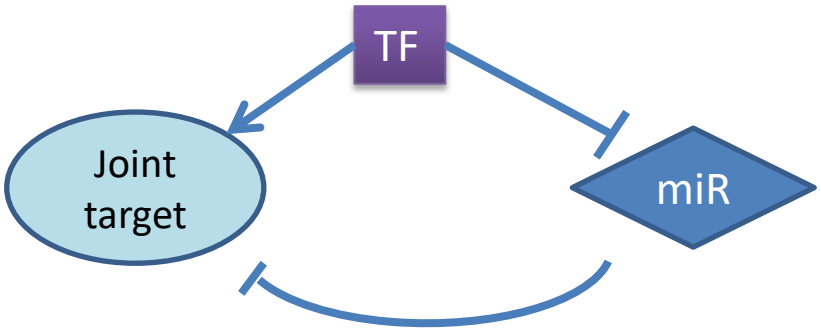
mRNAs that encode for Transcription Factors are very often controlled by miRNA.

This suggests a model where feed-forward and feed-back control loops exist.





uncoherent



coherent

Friard et al. BMC Bioinformatics 2010, 11:435
http://www.biomedcentral.com/1471-2105/11/435

BMC Bioinformatics

DATABASE Open Access

CircuitsDB: a database of mixed microRNA/transcription factor feed-forward regulatory circuits in human and mouse

Olivier Friard[1], Angela Re[2], Daniela Taverna[1,3,4], Michele De Bortoli[1,3], Davide Corá[1,5*]

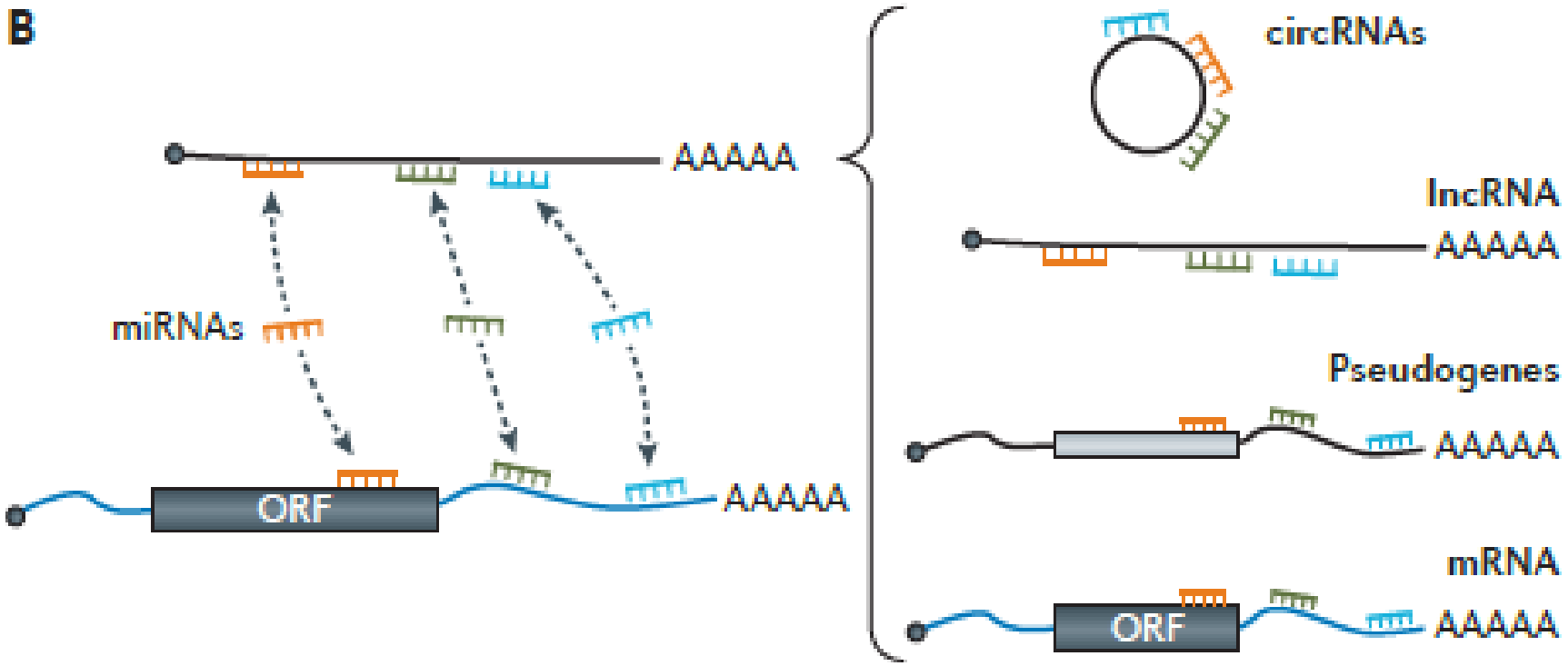
Abstract

Background: Transcription Factors (TFs) and microRNAs (miRNAs) are key players for gene expression regulation in higher eukaryotes. In the last years, a large amount of bioinformatic studies were devoted to the elucidation of transcriptional and post-transcriptional (mostly miRNA-mediated) regulatory interactions, but little is known about the interplay between them.

Description: Here we describe a dynamic web-accessible database, `CircuitsDB`, supporting a genome-wide transcriptional and post-transcriptional regulatory network integration, for the human and mouse genomes, based on a bioinformatic sequence-analysis approach. In particular, `CircuitsDB` is currently focused on the study of mixed miRNA/TF Feed-Forward regulatory Loops (FFLs), i.e. elementary circuits in which a master TF regulates an miRNA and together with it a set of Joint Target protein-coding genes. The database was constructed using an ab-initio oligo analysis procedure for the identification of the transcriptional and post-transcriptional interactions. Several external sources of information were then pooled together to obtain the functional annotation of the proposed interactions. Results for human and mouse genomes are presented in an integrated web tool, that allows users to explore the circuits, investigate their sequence and functional properties and thus suggest possible biological experiments.

Conclusions: We present `CircuitsDB`, a web-server devoted to the study of human and mouse mixed miRNA/TF Feed-Forward regulatory circuits, freely available at: http://biocluster.di.unito.it/circuits/

ceRNAs (competing endogenous RNA)



Base pairing is the mode of action of competing endogenous RNAs. In this case, however, the complementarity is between microRNAs (miRNAs) and different targets (including circular RNAs (circRNAs), lncRNAs, pseudogene transcripts and mRNAs).

Fatica & Bozzoni 2014

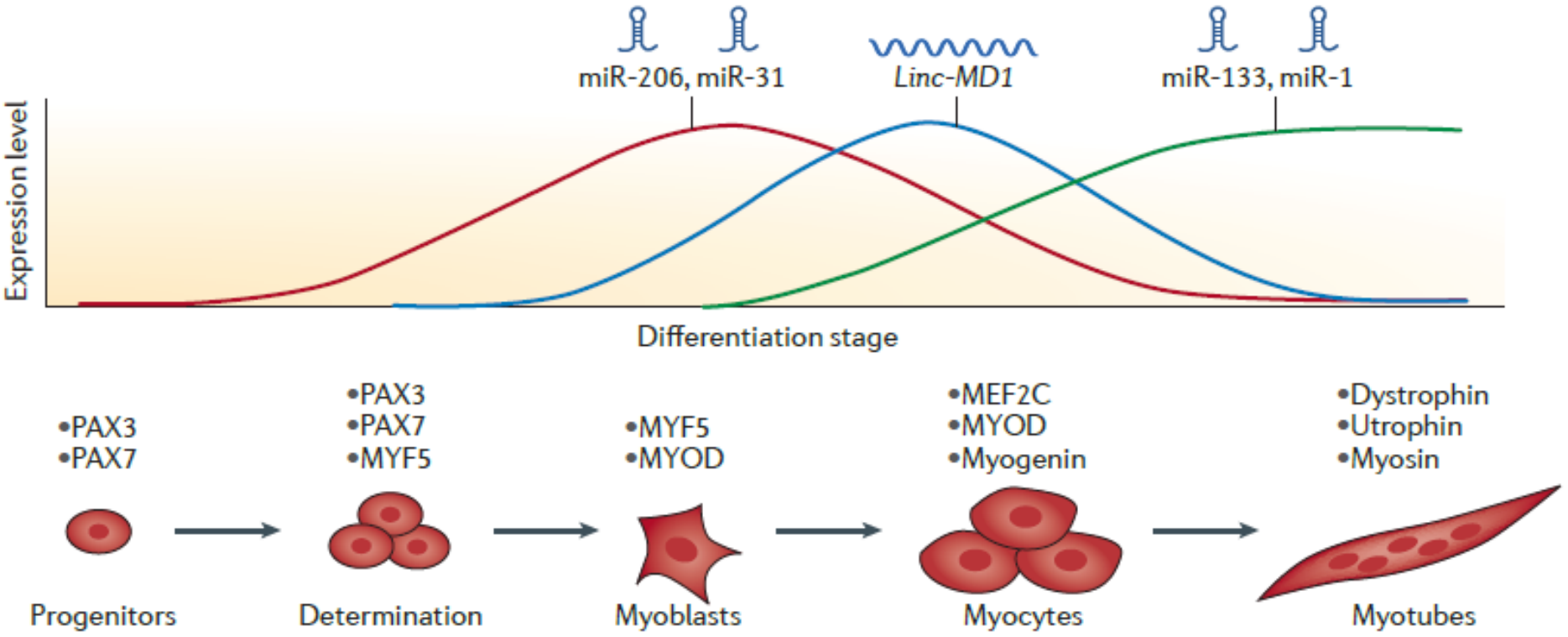
Another layer of regulation.....

A Long Noncoding RNA Controls Muscle Differentiation by Functioning as a Competing Endogenous RNA

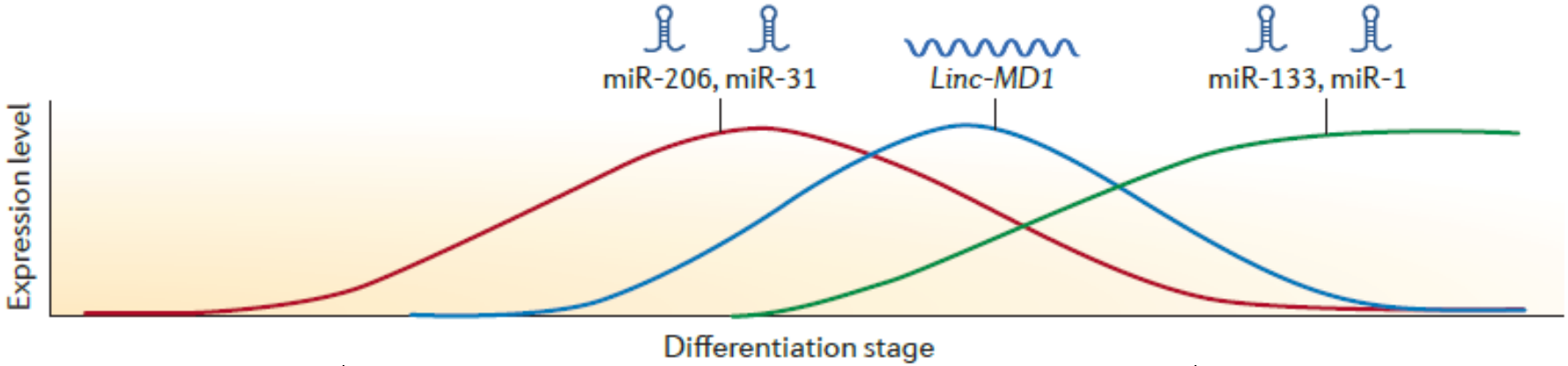
Marcella Cesana,[1,6] Davide Cacchiarelli,[1,6] Ivano Legnini,[1] Tiziana Santini,[1] Olga Sthandier,[1] Mauro Chinappi,[2] Anna Tramontano,[2,3,4] and Irene Bozzoni[1,3,4,5,*]
[1]Department of Biology and Biotechnology "Charles Darwin"
[2]Department of Physics
[3]Institut Pasteur Fondazione Cenci-Bolognetti
[4]Center for Life Nano Science @Sapienza, Istituto Italiano di Tecnologi
[5]IBPM of Consiglio Nazionale delle Ricerche (CNR)
Sapienza University of Rome, P.le A. Moro 5, 00185 Rome, Italy
[6]These authors contributed equally to the work
*Correspondence: irene.bozzoni@uniroma1.it

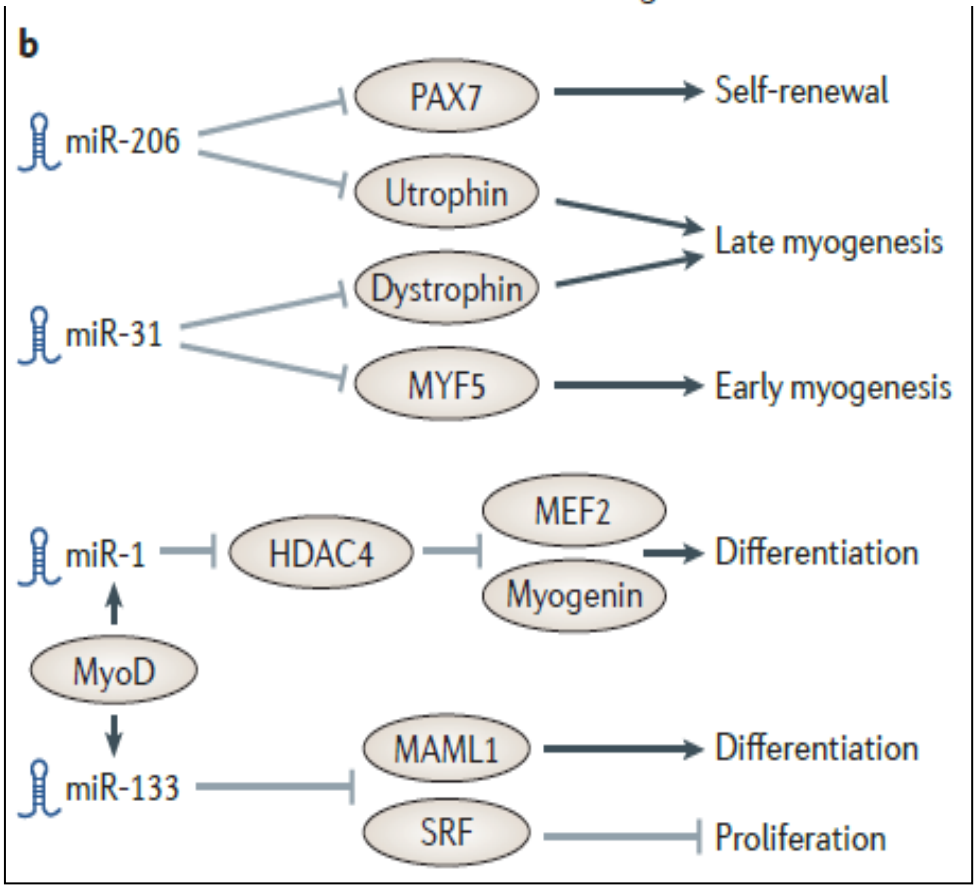
Recently, a new regulatory circuitry has been identified in which RNAs can crosstalk with each other by competing for shared microRNAs. Such competing endogenous RNAs (ceRNAs) regulate the distribution of miRNA molecules on their targets and thereby impose an additional level of post-transcriptional regulation. Here we identify a muscle-specific long noncoding RNA, linc-MD1, which governs the time of muscle differentiation by acting as a ceRNA in mouse and human myoblasts. Downregulation or overexpression of linc-MD1 correlate with retardation or anticipation of the muscle differentiation program, respectively. We show that linc-MD1 "sponges" miR-133 and miR-135 to regulate the expression of MAML1 and MEF2C, transcription factors that activate muscle-specific gene expression. Finally, we demonstrate that linc-MD1 exerts the same control over differentiation timing in human myoblasts, and that its levels are strongly reduced in Duchenne muscle cells. We conclude that the ceRNA network plays an important role in muscle differentiation.

Cell *147*, 358–369, October 14, 2011



Fatica & Bozzoni 2014


miR-206, miR-31
Linc-MD1
miR-133, miR-1
Expression level
Differentiation stage


b
PAX7
Self-renewal
miR-206
Utrophin
Late myogenesis
Dystrophin
miR-31
MYF5
Early myogenesis
MEF2
miR-1
HDAC4
Differentiation
Myogenin
MyoD
MAML1
Differentiation
miR-133
SRF
Proliferation

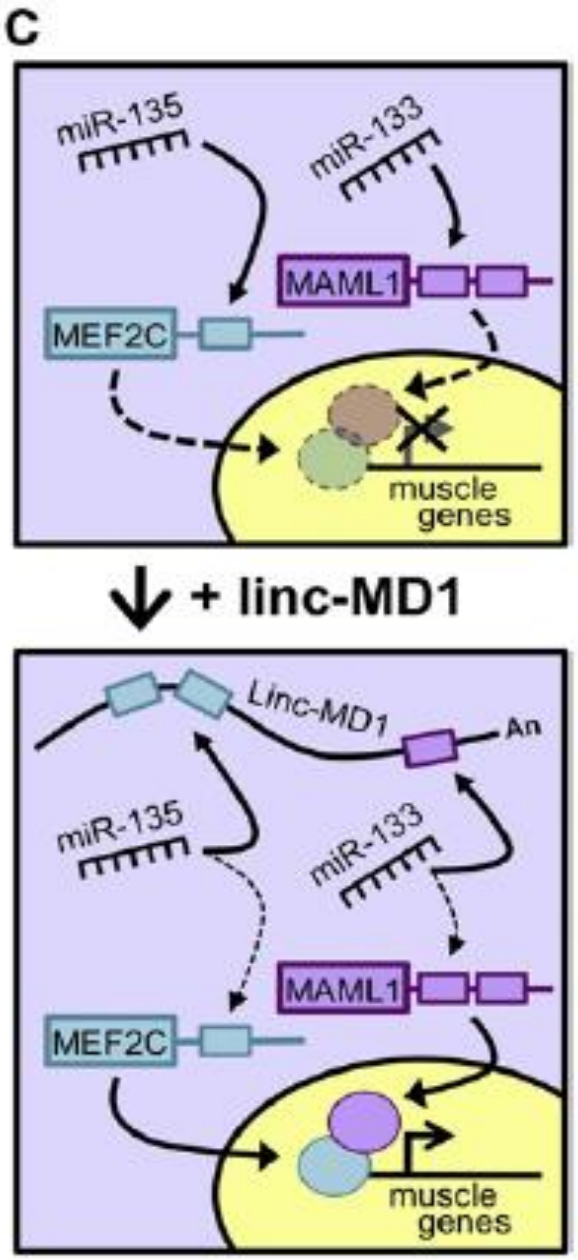


Figure 7. linc-MD1 Is Conserved in Humans, and It Improves Differentiation of Duchenne Myoblasts

(C) Schematic representation of the circuitry linking linc-MD1, miR-135, miR-133, and muscle differentiation. (Cesana et al., 2011)