The garden of misunderstood RNA: the molecular mechanisms of Lamarckian evolution

Andrew Kuznetsov

Freiburg, Germany

The Garden of Earthly Delights
[Hieronymus Bosch, 1490-1510]

The Garden of Earthly Delights
[Hieronymus Bosch, 1490-1510]

"We have fundamentally misunderstood the nature of the genetic programming of higher organisms because of the apparently reasonable but now evidently incorrect assumption that most genetic information is transacted by proteins. The vast majority of the human genome does not encode proteins, but is dynamically expressed as RNA, whose primary purpose appears to be to control the epigenetic processes ..."

Prof. John S. Mattick
Executive Director of the Garvan Institute of Medical Research
Sydney, Australia

Content

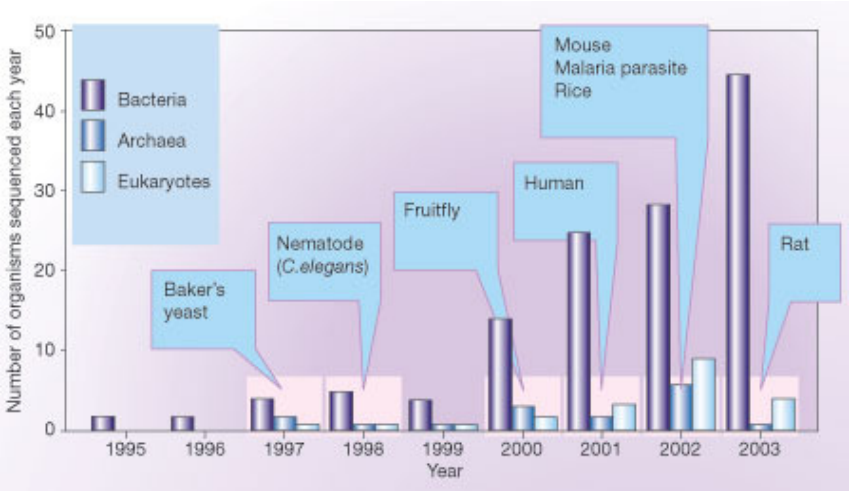
- Non-coding RNAs
  - increasing complexity, regulatory elements
  - information flow in the cell
- Classes of ncRNAs
  - siRNA (RNA interference, Dicer, RISC, Argonaute)
  - miRNA (Microprocessor complex, Drosha, Pasha)
  - piRNA (transposon silencing: ping-pong mechanism)
  - lncRNA (X-chromosome inactivation)
- Inheritance of acquired characteristics
  - chromatin remodeling (RITS)
  - CRISPR/Cas system in bacteria and archaea
  - epigenetic inheritance of viRNA in the nematoda *C.elegans*
  - RNAs as extracellular signaling and informational molecules, reverse transcription
- Crossroads between DNA and RNA, genetics and epigenetics
  - DNA-guided DNA interference
  - RNA-guided DNA modifications

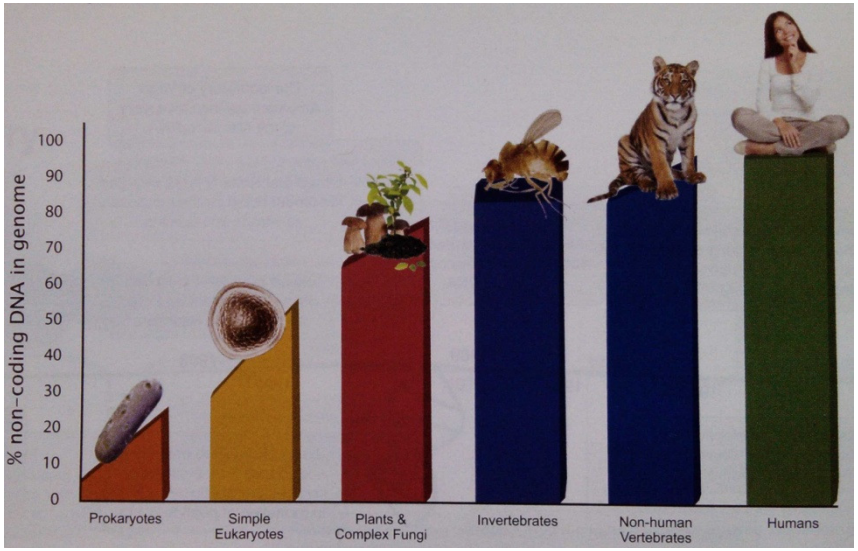
Abbreviations

- mRNA – messenger RNA
- ssRNA/dsRNA – single/double stranded RNA
- shRNA – small hairpin RNA
- ncRNA – non-coding RNA
- lncRNA – large non-coding RNA
- RNAi – RNA interference
- siRNA – small interfering RNA
- miRNA – micro-RNA
- miRNP – micro-RNA ribonucleoprotein complex
- pri-miRNA – primary-miRNA transcript
- pre-miRNA – precursor micro-RNA
- piRNA – piwi-interacting RNA
- RISC – RNA-Induced Silencing Complex
- RITS – RNA-Induced Transcriptional Silencing
- CRISPR – Clustered Regularly Interspaced Short Palindromic Repeats

Where is the information that programs human complexity?

The number of sequenced genomes has increased as technology decreases the cost of sequencing

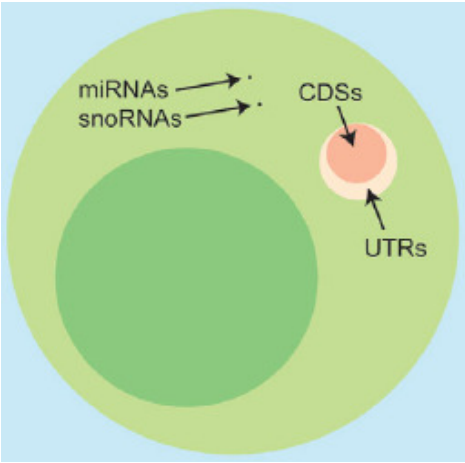




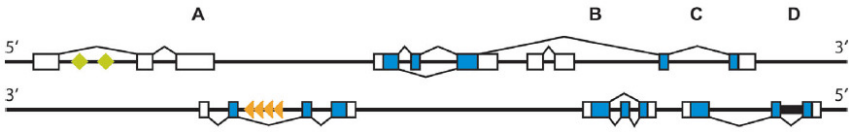
While the number of protein-coding genes expressed by organism does not scale with organism complexity (humans have approximately the same number as worm), the amount of non-protein-coding DNA does. Therefore, it is likely that RNA transcribed from these non-protein-coding regions allows for complex development and differentiation, as well as advanced cognitive potential

[Mattick, 2004]

The new RNA world



Transcription in mammals. The area of the box represents the genome. The large green circle is equivalent to the transcriptom, with the dark green area corresponding to transcripts from both strands. **CDSs** are protein-coding sequences, and **UTRs** are untranslated regions in mRNAs. The dots indicate (and in fact overstate) the proportion of the genome occupied by known miRNAs and snoRNAs



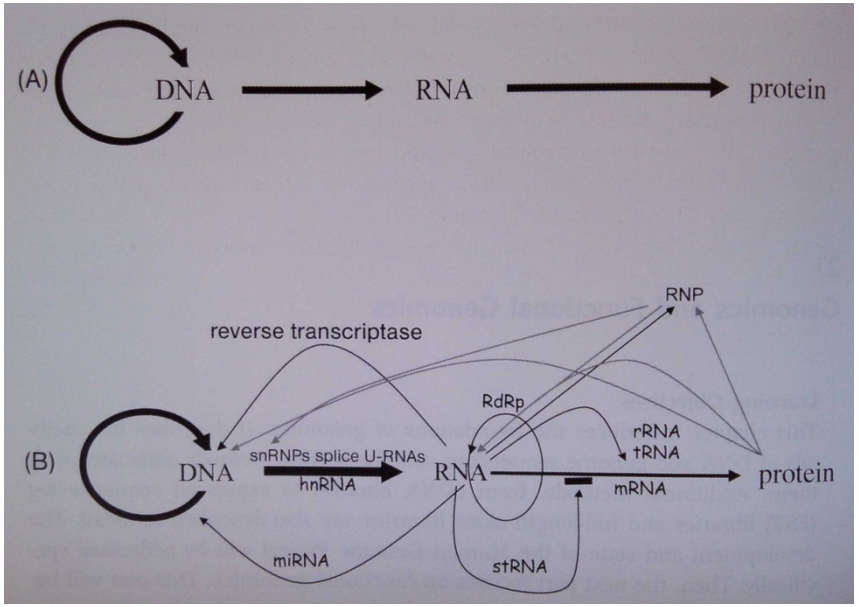
Complexity of the transcriptional landscape. White boxes represent non-coding exonic sequences and blue boxes protein-coding exonic sequences. Green diamonds represent snoRNAs and orange triangles represent miRNAs. Indicated are (**A**) antisense transcripts with overlapping exons, (**B**) nested transcripts on both strands, (**C**) antisense transcripts with interlacing exons and (**D**) retained introns

[Mattick, Makunin, 2006]

RNA throughout history

Year	Discovery
1958	The central dogma of molecular biology was proposed by Francis Crick
1960	mRNA discovered in the labs of Francois Jacob, Jacques Monod, Sydney Brenner, and Francis Crick
1966	The complete genetic code was cracked by labs of Nirenberg, Matthaei, Leder, and Korana
1968	Francis Crick, Carl Woese, and Leslie Orgel proposed that the primordial genetic molecule was RNA
1970	Reverse transcriptase discovered in the labs of Howard Temin and David Baltimore
1977	Introns and mRNA splicing discovered in labs of Phillip Sharp and Richard Roberts
1992	Harry Noller's lab presented evidence for the catalytic involvement of rRNA in formation of peptide bonds
1993	The lab of Victor Ambros published discovery of the first miRNA
1998	Andrew Fire and Craig Mello described RNAi in *C.elegans*
now	Many labs are investigating the role of ncRNAs in the gene regulation

Information flow in the cell



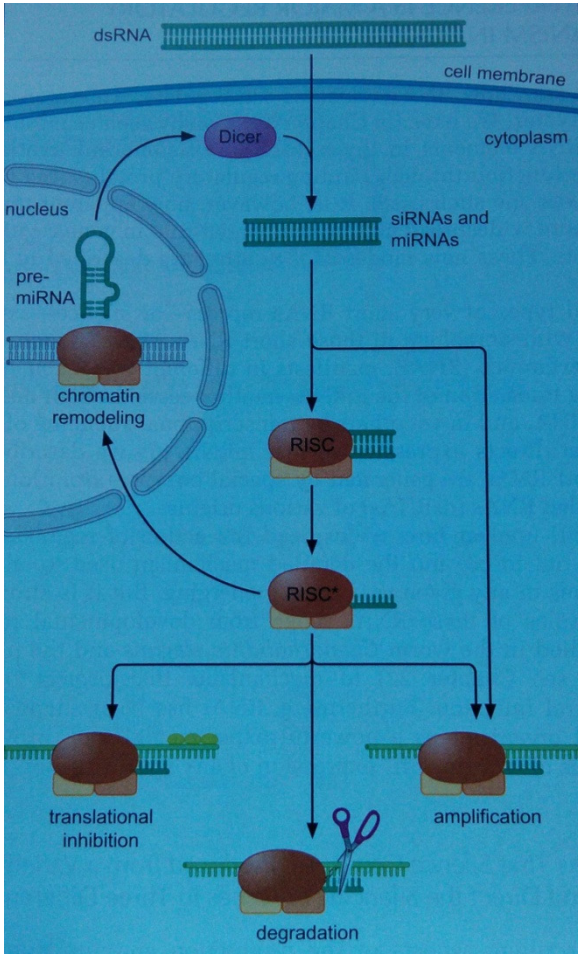
- The **central dogma** of molecular biology assumed an information flow from DNA to protein
- Connections between **DNA**, **RNA**, and **protein** are improved . The arrow drawn between DNA and RNA is thicker than the one between RNA and protein, since only a small amount of RNA is actually translated to protein. The gray arrows describe the role of proteins and RNA in cellular replication, transcription, and translation
- **hnRNA**, heterogeneous nuclear RNA (transcription product, still includes introns); **snRNP**, small nuclear ribonuclear protein (in splicosomes U-RNAs are small RNA, which are involved in splicing); **rRNA**, ribosomal RNA, **tRNA**, transfer RNA; **mRNA**, messenger RNA (fully processed messenger between gene and protein); **miRNA** and **stRNA** , micro RNA and small temporal RNA (~22 nt long functional RNAs, which play a part in cell cycle regulation, translation and degradation of RNA); **RdRp**, RNA-dependent RNA polymerase; **RNP**, ribonucleoprotein (complex consisting of proteins and RNAs, e.g., telomerase, splicosome)

[Wink, 2011]

Classes of small non-coding RNAs

	siRNA	miRNA	piRNA
Matching	perfect complementarity	partial matching	perfect complementarity
Binding	AGO	AGO	PIWI
Length	~21 nt	~22 nt	28-32 nt
Origin	exogenous (viral) dsRNAs, transposons	dsRNA precursors from genes	ssRNA precursors
Target	Anti(-)-strands, (-) transposons	(-) mRNAs	transposons, regulation of development
Abundance	all tissues	all tissues	mainly germ cells

Different RNA pathways



- dsRNA molecules switches off the gene expression when dsRNAs having homology to that gene are introduced, or made, in cell
- This effect involves processing of the dsRNA to make **siRNAs** and **miRNAs** by the enzyme **Dicer**. Another enzyme involved only in the case of miRNAs is **Drosha**
- The siRNAs and miRNAs direct the RNA-induced silencing complex (**RISC**) to repress genes in three ways:

1. it attacks and digests mRNA that has homology with the siRNA;
2. it interferes with translation of those mRNAs; or
3. it directs chromatin modifying enzymes to the promoters that direct expression of those mRNAs

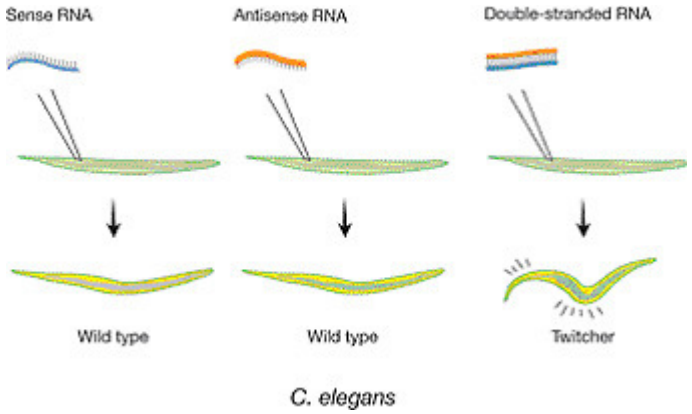
[Watson et al, 2008]

RNA interference

White regions in the petunia are the result of RNAi-mediated silencing of genes responsible for pigmentation

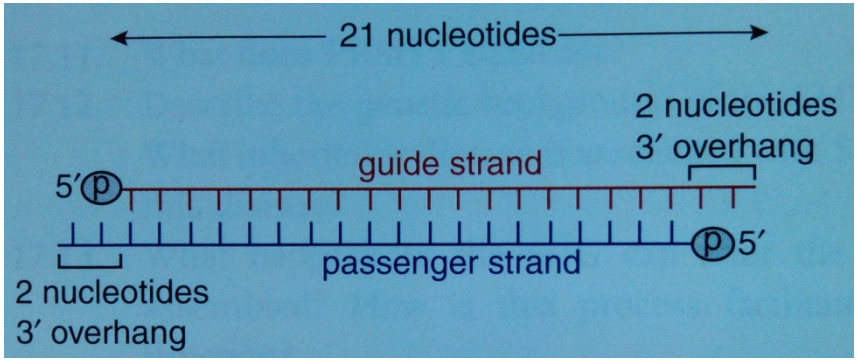
Crops have been engineered to express siRNA against viruses and insects. Other implementations of RNAi include caffeine-free beans and allergen-free fruit

The Nobel Prize in Physiology or Medicine 2006

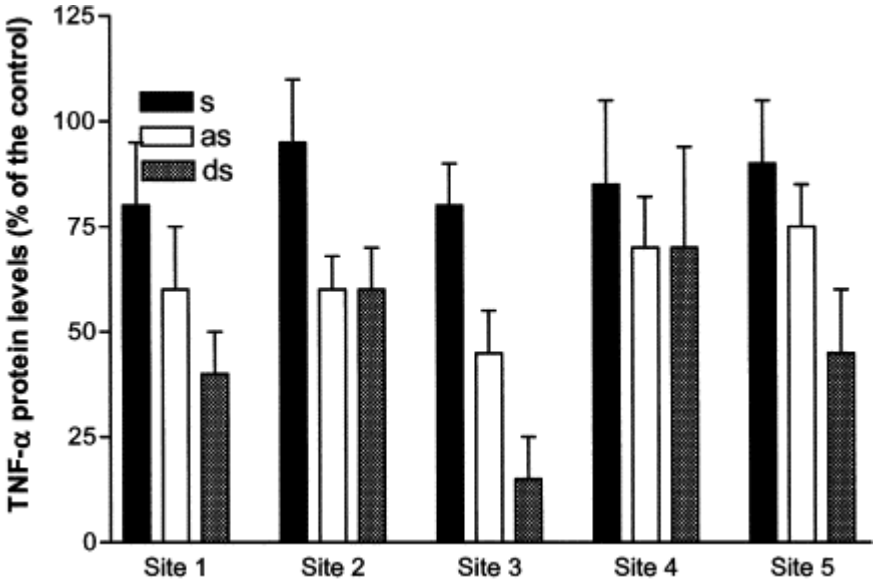


Their Majesties Queen Silvia and King Carl XVI Gustaf of Sweden (middle) posing with Nobel Laureate Andrew Z. Fire and his wife, Rachel Krantz (left), and Nobel Laureate Craig C. Mello and his wife, Edit Mello (right), at the Nobel Banquet, 10 December 2006

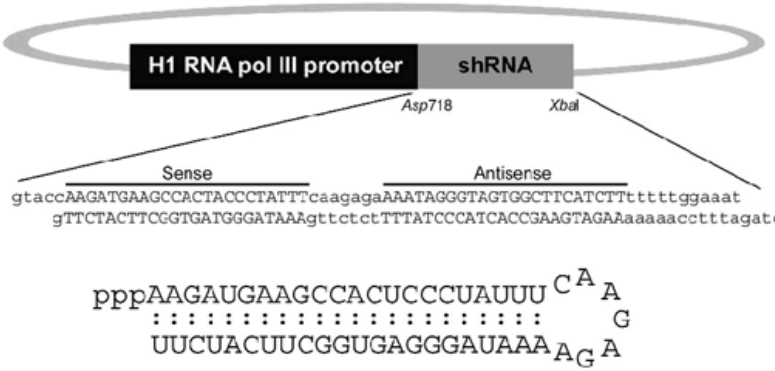
RNA-based gene knock-down



siRNA duplex. Two nucleotide 3‘ overhands and phosphate groups at the 5‘ ends characterize double stranded siRNAs. They are 21 nt long and contain a **guide strand**, which is incorporated into the **RISC** complex, and the **passenger strand**, which is removed from the cell after unwinding



in vitro study of **anti-TNF-α siRNAs** in peritoneal macrophages



RasGAP shRNA showing the sense and antisense region that target the ***Rasa1*** gene

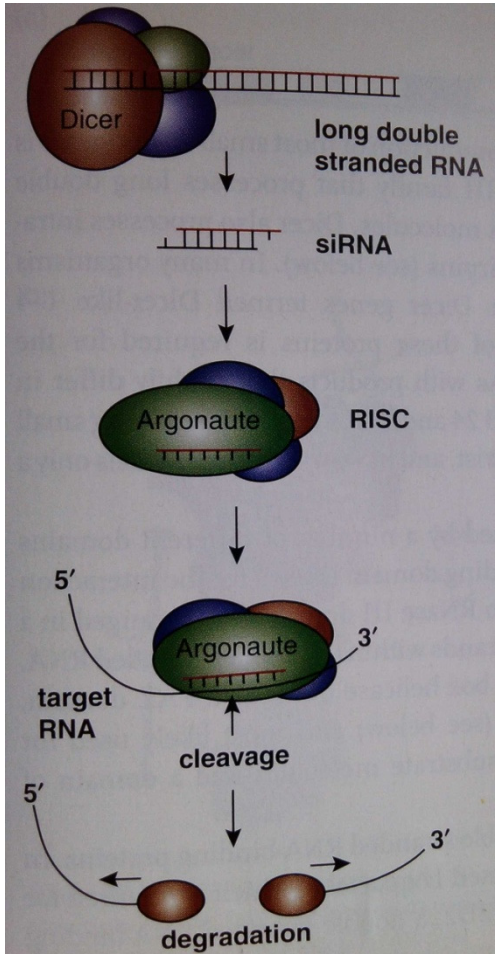
[Sorensen at al, 2003; Kunath et al, 2003]

Design of siRNA in Perl

```
# First read the sequence from a file named 'mrna.fa'
$seq = '';
open(IN, 'mrna.fa') or die "Could not open file mrna.fa\n";
while (<IN>) {
    unless (/>/) {
        chomp;
        $seq .= $_;
    }
}
close IN;

# Now analyze the sequence read from file
# Step through each position  of the sequence
for (  $i = 0 ; $i < length($seq) -22 ; $i++ ) {
     $testseq = substr( $seq, $i, 23 );

    # check if first two positions are AA and
    # last are TT
    if ( $testseq =~ /^AA.*TT$/ ) {

         # test GC content

         # count the number of G's and C's
         $gc_content = ( $testseq =~ tr/GC// ) / 23;

         # is the GC content within the range 30-50%?
         if ( ( $gc_content >= 0.3 ) && ( $gc_content <= 0.5 ) ) {

            # does the sequence contain stretches of As, Ts, Cs or Gs?
            unless ( ( $testseq =~ /A{4}/ )
                || ( $testseq =~ /T{4}/ )
                || ( $testseq =~ /G{4}/ )
                || ( $testseq =~ /C{4}/ )

            # avoid also regions of six positions with G or C
                ||  ( $testseq =~ /[GC]{6}/ ))
            {
                print "pos $i $testseq\n";
            }
        }
    }
}
```

[Samuelsson, 2011]

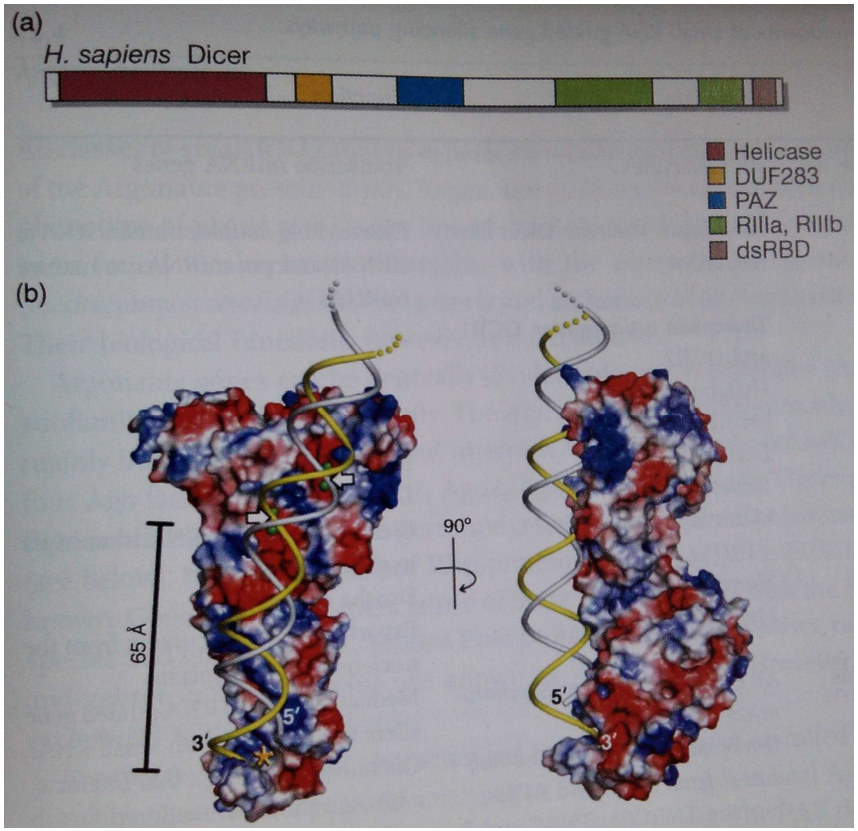
Mechanism of RNA-guided RNA interference



- Long dsRNA is cleaved by the RNase III enzyme **Dicer** to small double stranded siRNA. Dicer cleaves long RNA molecules preferentially from the ends
- The siRNA duplex is unwound and the **guide strand** is incorporated into the RNA-induced silencing complex (**RISC**). Within RISC, the guide strand interacts with an **Argonaute** protein that cleaves the target RNA
- The two cleaved products are removed from the cell

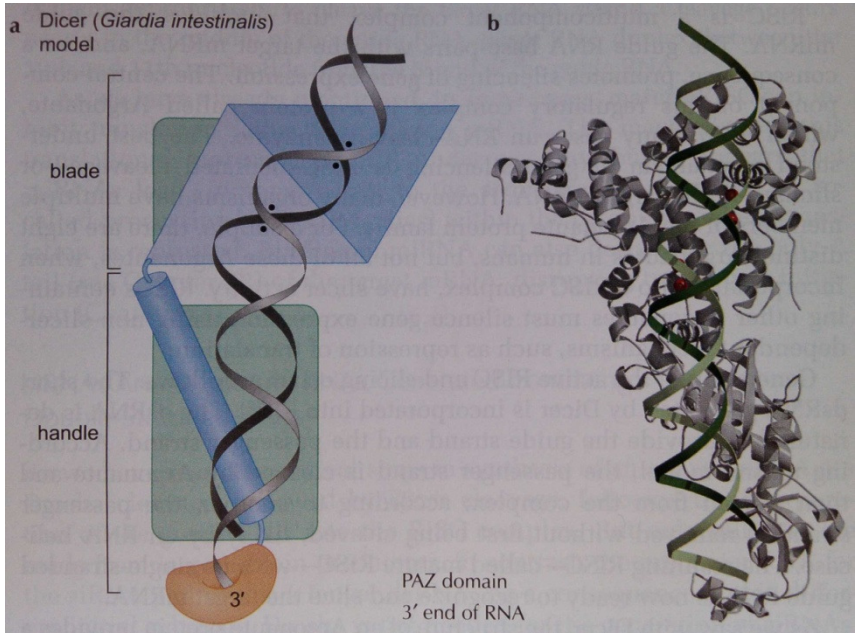
[Meister, 2011]

Structure of Dicer



(a) Dicer consists of **DEAD** box helicase domain, a domain of unknown function (**DUF283**), a **PAZ** domain, 2 **RNase III** domains and a **dsRNA-binding** domain

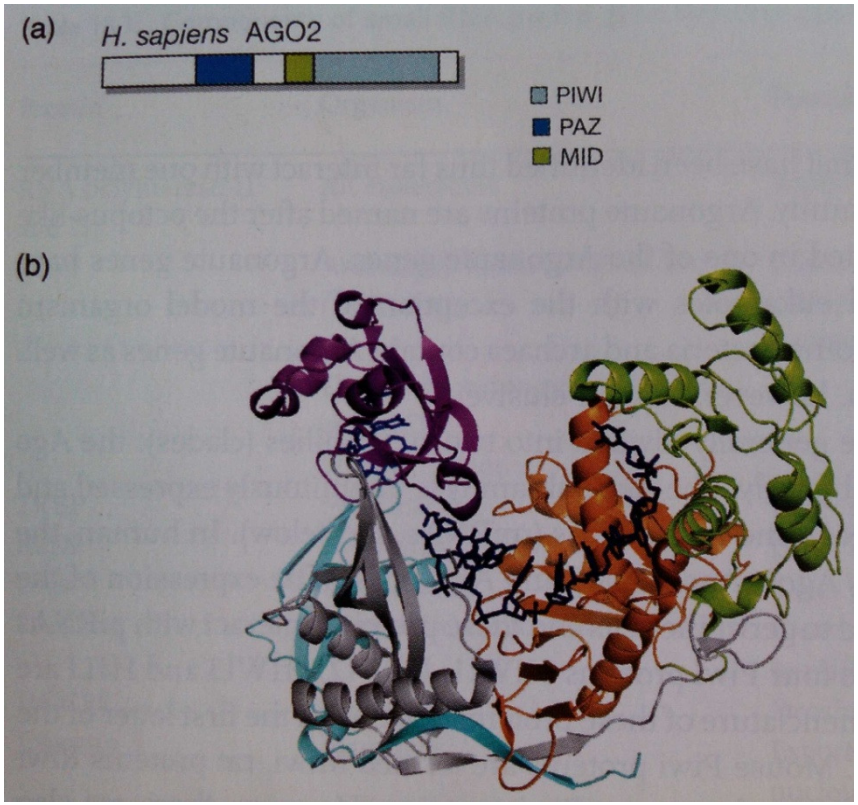
(b) Dicer binds the end of the long dsRNA (shown in yellow) and cleaves about 21 nt upstream resulting of a 21 nt dsRNA product



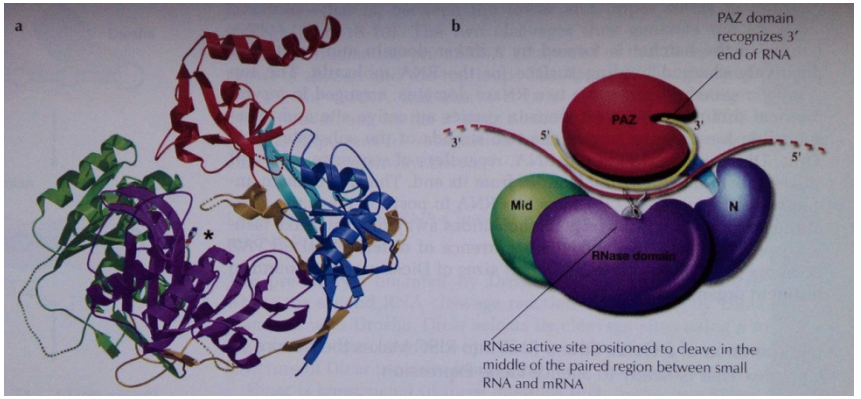
The structural model of Dicer with dsRNA. The protein is shown in gray, with nuclease active site indicated by the red spheres (and as black dots in part **a**). The RNA is in green. The structure shown contains only the **RNase III** and **PAZ** domains

[Macrae et al, 2006]

Structure of Argonaute protein



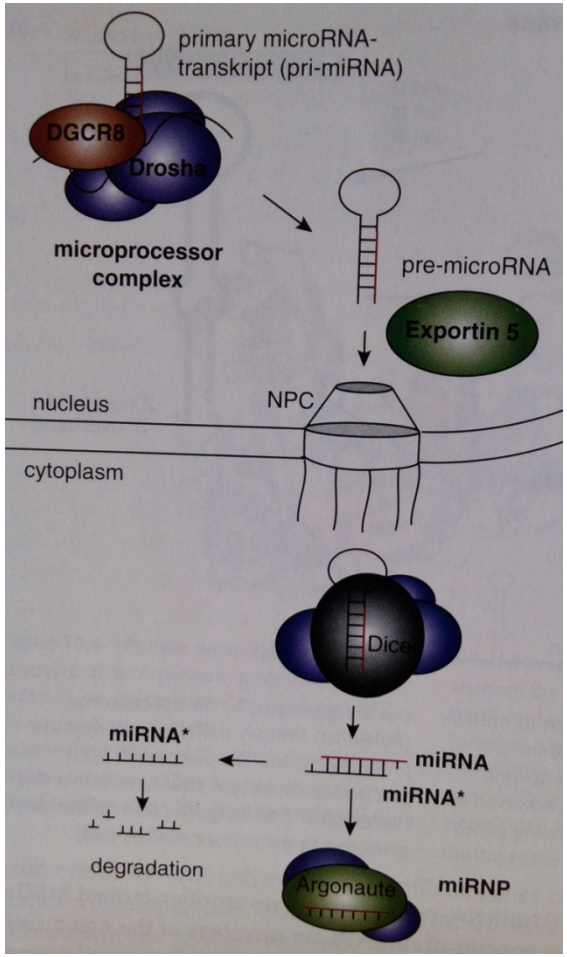
Ago proteins include PAZ, MID and PIWI domains. The **PAZ** domain (blue) binds 3’ end of the siRNA, while the 5’ end is anchored in the **MID** domain (green). The **PIWI** domain (orange) is structurally similar to RNase H and in some Ago proteins this domain can cleave target RNAs. Such Ago proteins are named **Slicers**



(a) **Crystal structure of Argonaute**. The domains are colored as in (b), with the blue domain being the amino-terminal part of the protein, and the green domain in the middle
(b) **The cartoon of the Argonaute domains**. The arrow shows the RNase active site positioned to cleave in the middle of the paired region

[Wang et al, 2009; Song et al, 2004]

miRNA pathway

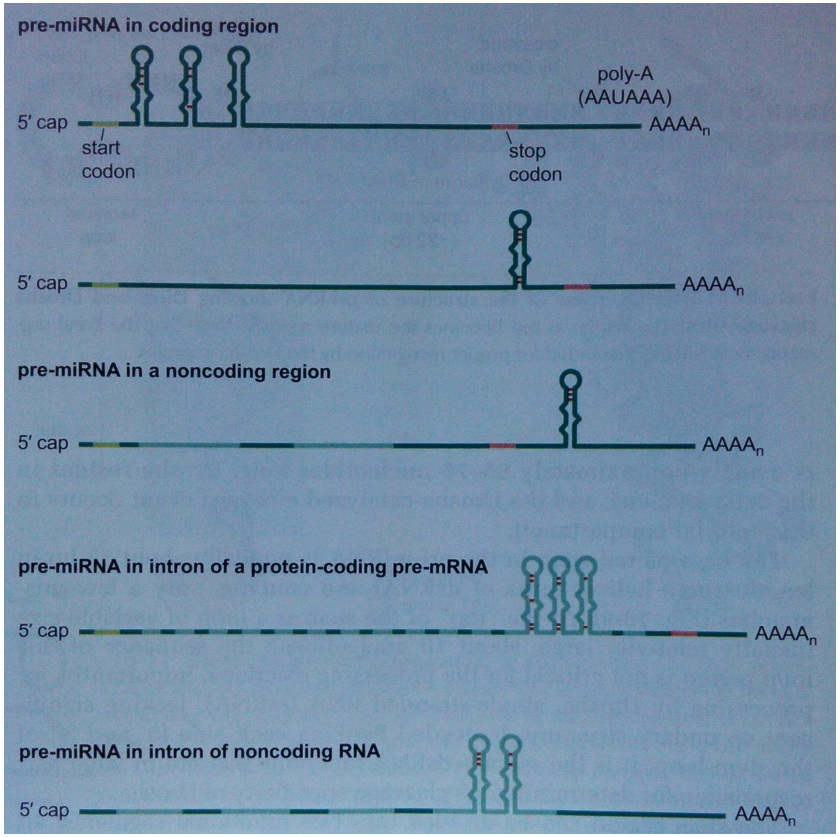


- The nuclear **Microprocessor complex** that contains RNase III enzyme **Drosha** and its interaction partner **Pasha** (**DGCR8**) process primary miRNA transcripts (**pri-miRNA**)
- The resulting miRNA precursors (**pre-miRNAs**) are transported into the cytoplasm by the **exportin-5**
- In the cytoplasm, **Dicer** further processes the pre-miRNA to a double stranded intermediate, which is further unwound
- The mature miRNA is incorporated into a miRNA-protein complex termed **miRNP**. The other strand (miRNA*) is destabilized and removed from the cell. Mature miRNAs directly interact with an **Argonaute** protein within the miRNP

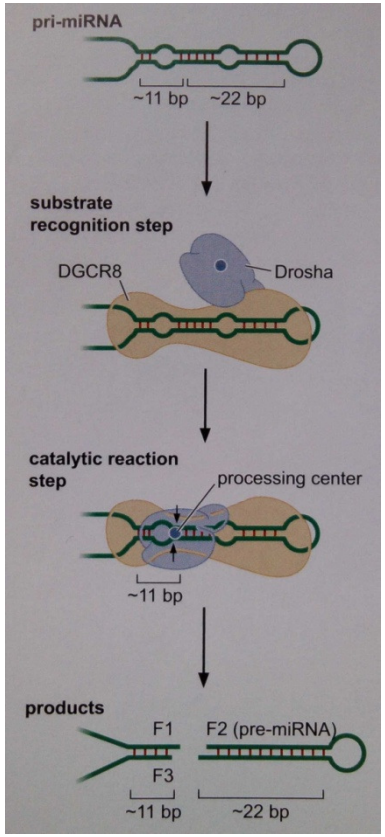
[Meister, 2011]

Cleavage of pri-miRNA by the Microprocessor complex

miRNAs are coded in both introns and exons within RNA

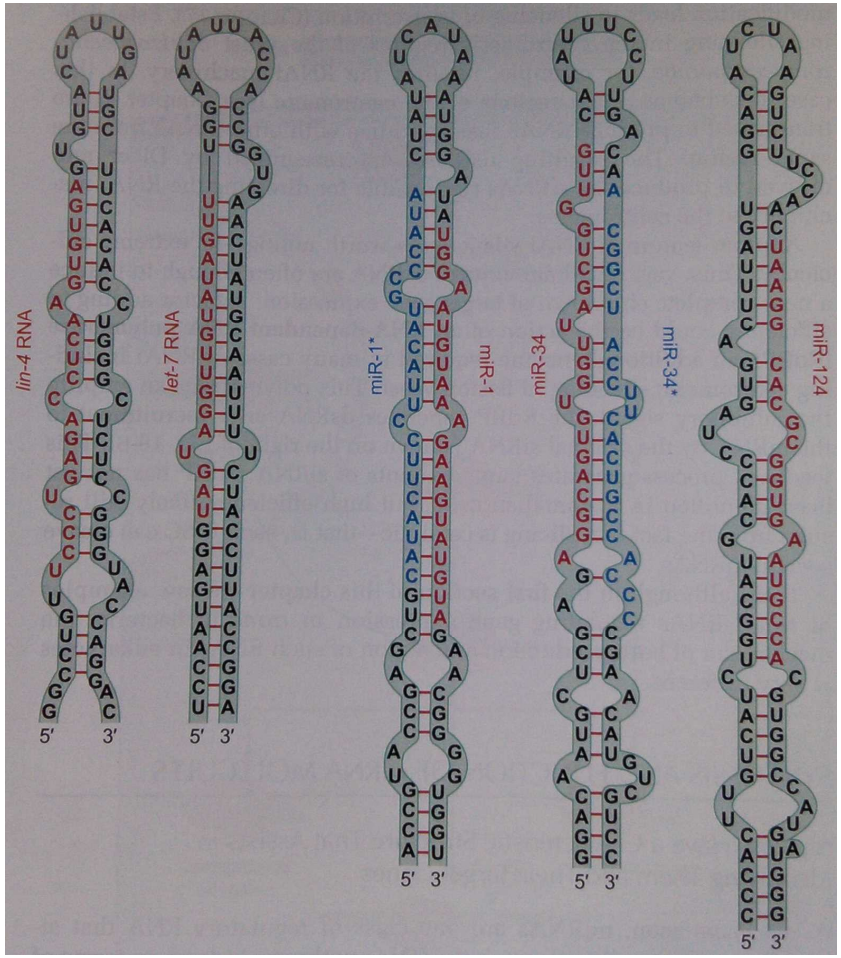


3 fragments are generated by cleavage, labeled F1, F3, and F2 (the pre-miRNA)



[Watson et al, 2008]

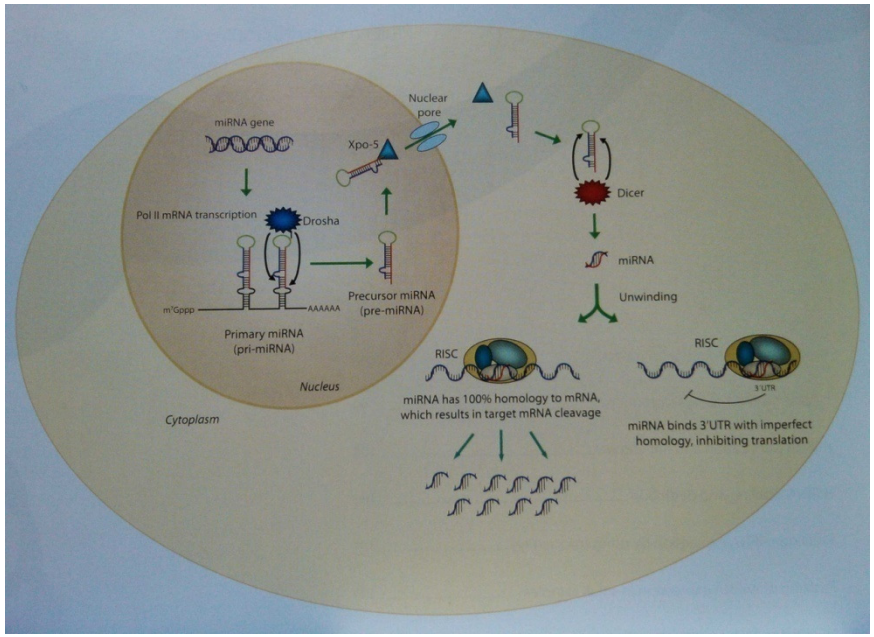
Structures of some pre-miRNAs from *C.elegans*



- The sequences in red are miRNAs. In some cases, both arms of a stem loop can generate a functional miRNA. In such cases, the second miRNA is shown in blue
- ***lin-4*** and ***let-7*** were identified genetically; those called **miR** are found by bioinformatics

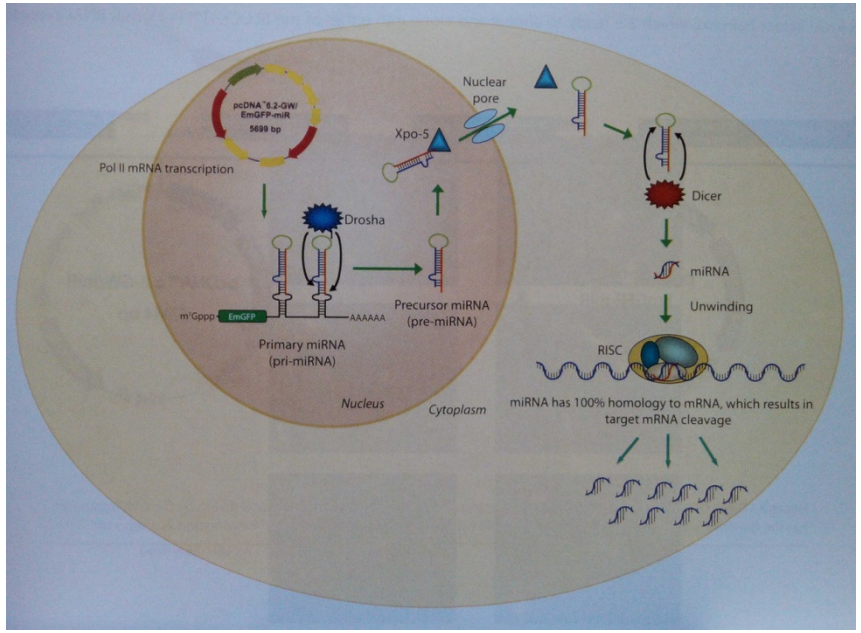
[Lim et al, 2003]

Biogenesis of miRNA & vector-mediated RNAi



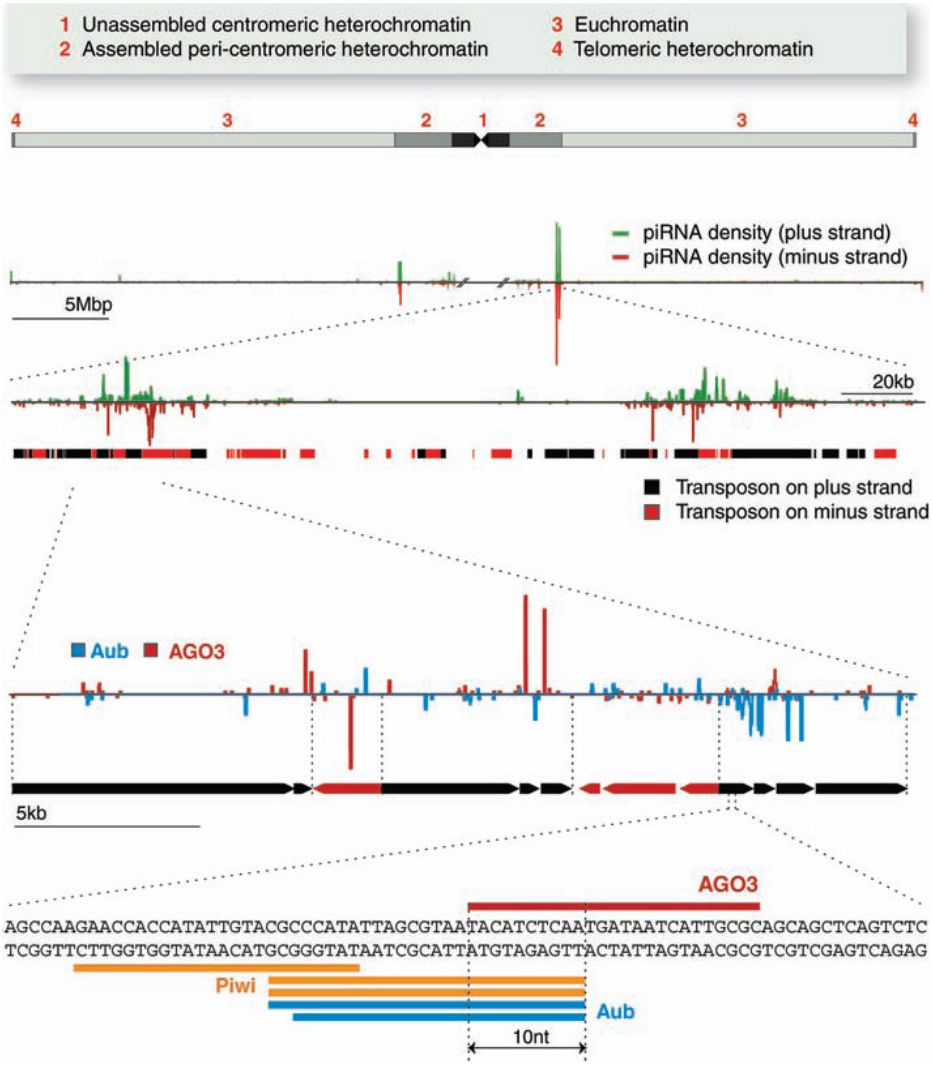
miRNA transcripts, generated by RNA polymerase II, are processed by RNase III enzyme **Drosha** (nuclear) and **Dicer** (cytoplasmic), yielding a 21-23 nt miRNA duplex. The less stable strand of the duplex is incorporated into the **RISC** complex, which regulates protein expression

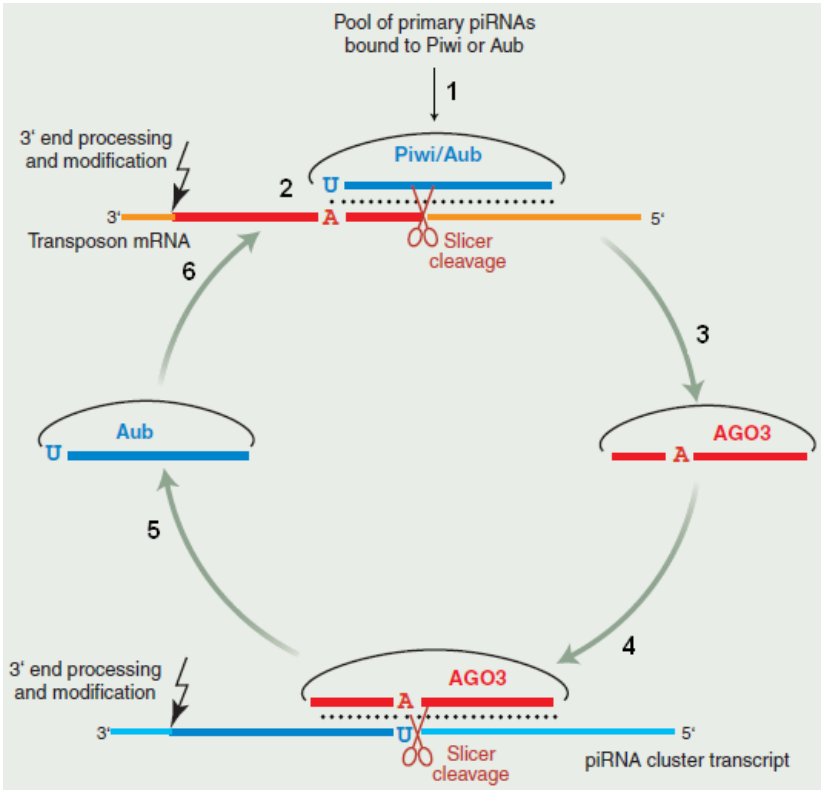
Expression of miR RNAi sequences using the BLOCK-IT Pol II miR RNAi Expression Vector, Invitrogen



[Invitrogen, RNAi]

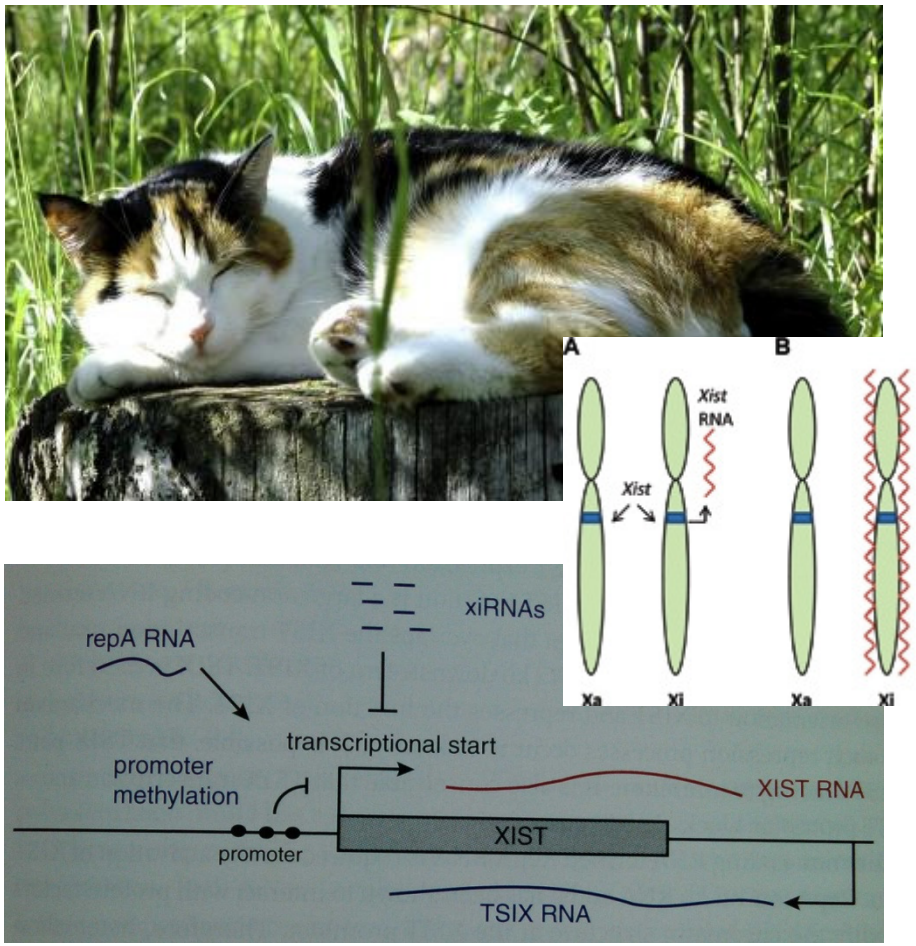
Relict innate immunity against foreign DNA/RNA, ping-pong mechanism





[Aravin et al, 2007]

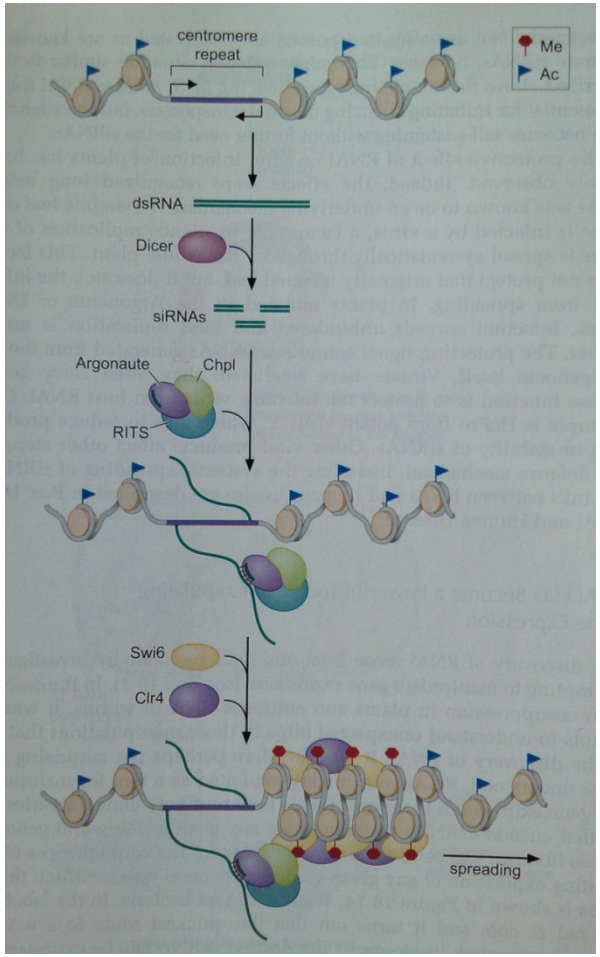
X-chromosome inactivation in mammals



- In calico cat, the patches of orange and black fur are a result of activity of different alleles on X chromosome
- During embryonic development one of the two X chromosomes is randomly chosen. From the X chromosome inactivation centre (**XIC**) the X-inactivation specific transcript (**XIST**) is expressed, which spreads over the X_i chromosome. XIST leads to the establishment of silenced heterochromatin on the inactivated chromosome X_i
- The major regulator of XIST is the lncRNA **TSIX** (reverse spelling of XIST), which is transcribed in antisense direction to XIST. TSIX may hybridize to XIST and inactivates it. **xiRNAs** are produced from the XIST/TSIX dsRNA and may influence XIST transcription negatively
- Another ncRNA that influences XIST expression is the **repA** ncRNA. repA has a positive effect on XIST transcription by chromatin modification in the XIST promoter region that lead to transcriptional activation
- Finally, **epigenetic modification** such as DNA methylation at the XIST promoter negatively regulate XIST expression

[Meister, 2011]

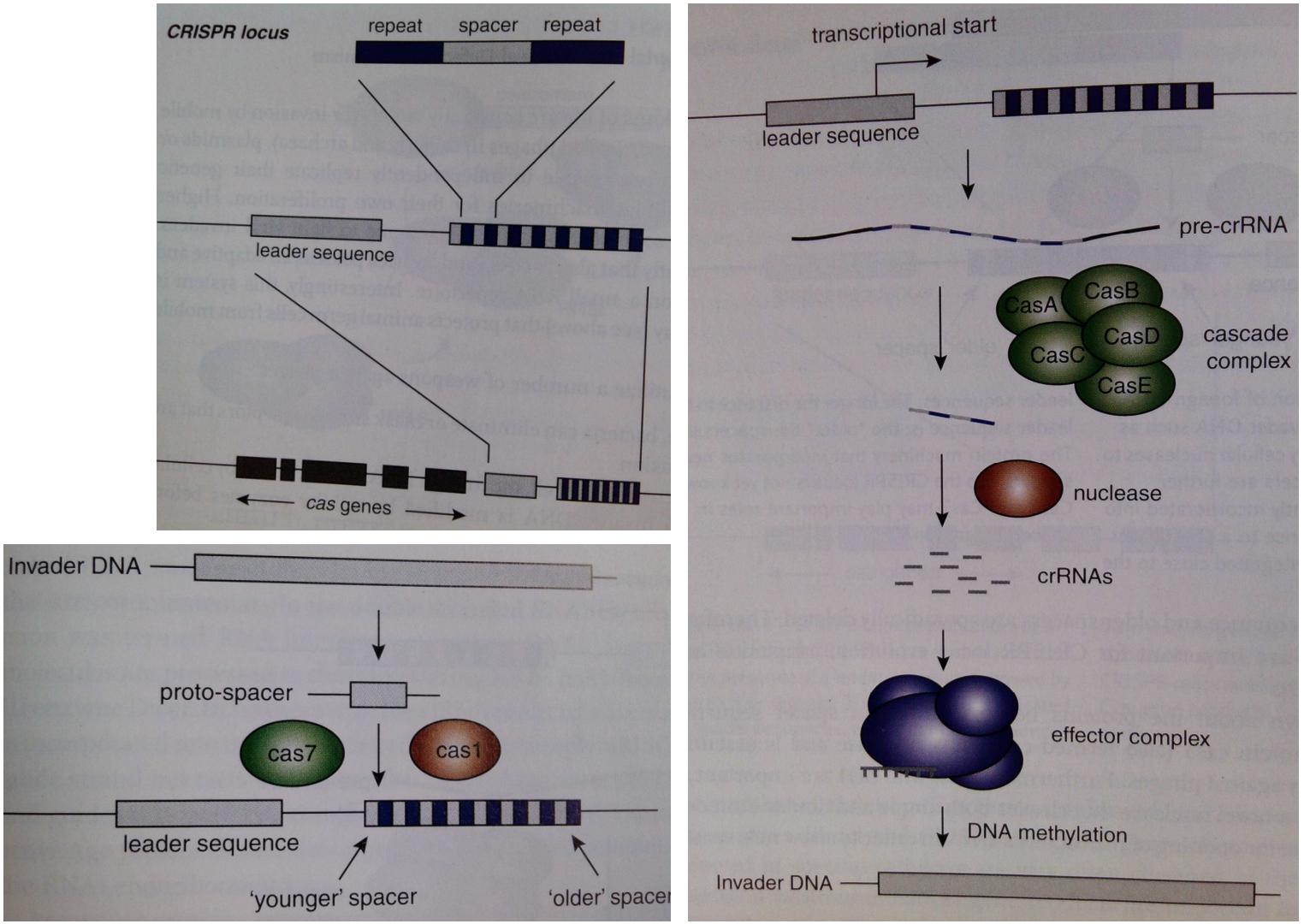
A model for RITS recruitment and the silencing of centromers in *S.pombe*



- The repeat sequences are transcribed from both strand by **RNA polymerase II**, generating dsRNA that is substrate for **Dicer**
- The produced siRNAs are loaded into the Argonaute-containing complex **RITS**
- The loaded RITS complex is then recruited back to the **PolII-tethered transcripts** through complementarity between the siRNA and the transcript
- This complex then recruits factors (**Clr4** and **Swi6**) that locally modify nucleosomes by adding the **H3K9** silencing markers
- Another subunit of RITS, **Chp1**, contains a **chromodomain**, which, by interacting with the methylated nucleosomes, stabilizes the binding of RITS
- "Slicing" of the transcripts by **Argonaute** (within RITS) generate substrate RNAs for the **RdRP**, which synthesized a complementary strand and thus generates further substrate for **Dicer**. This process is required for nucleosome modification to spread

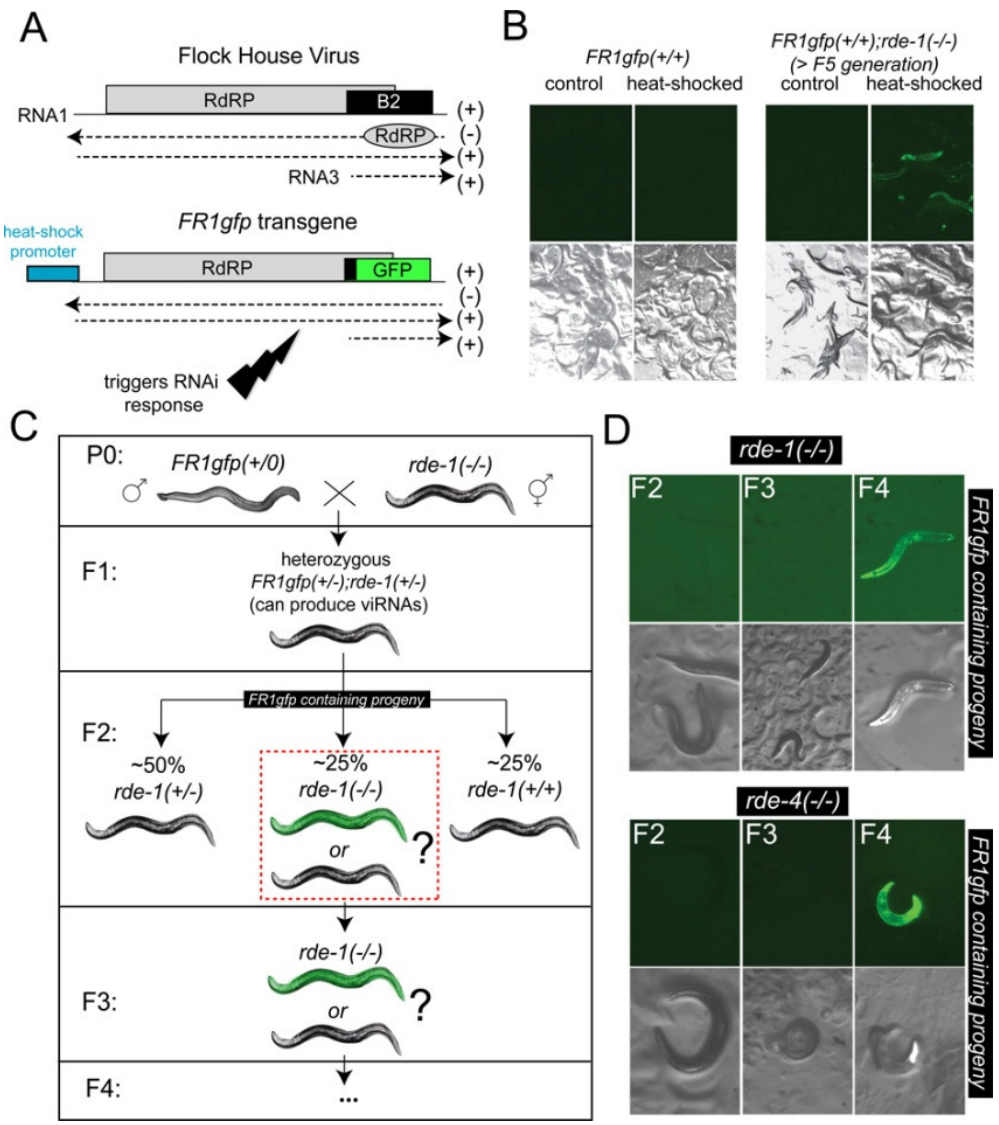
[Watson et al, 2008]

CRISPR defense mechanism in bacteria and archaea

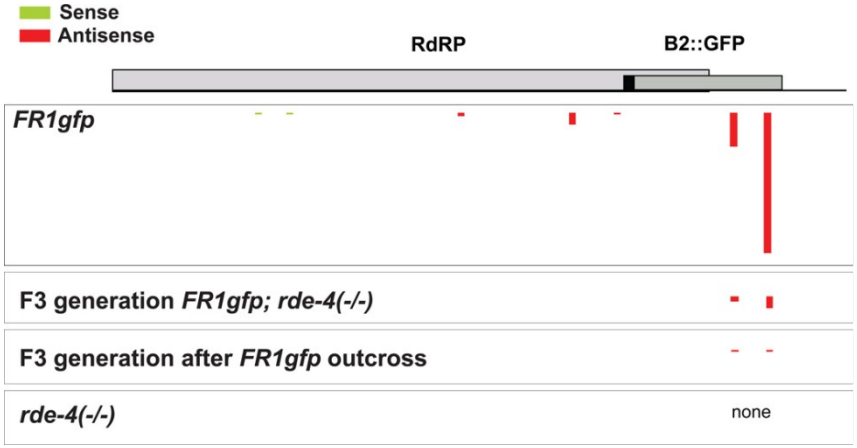


[Meister, 2011]

Inheritance of viRNA in *C.elegans*



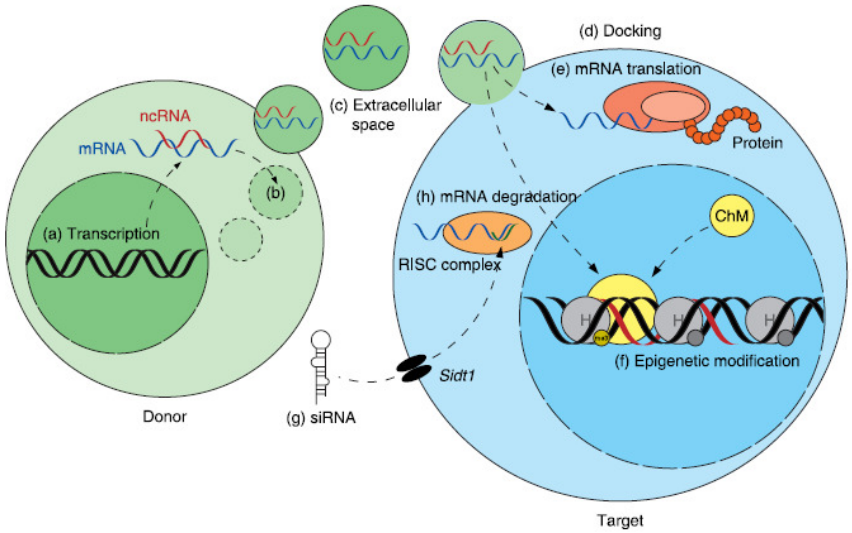
Expression of the recombinant **Flock House virus** in *C.elegans* results in the production of virus-derived, small interfering RNAs (viRNAs), which in turn silence the viral genome. The viRNA molecules are transmitted in non-Mendelian manner through 3 generations to silence viral genomes

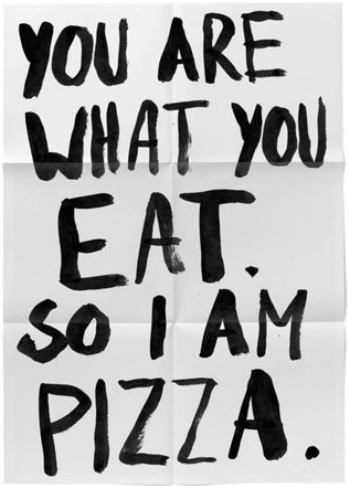


Small **viRNAs** were extracted, sequenced and mapped to the viral genome. 20-30 nt viRNAs match to the main two epitopes of the FHV genome. These results demonstrated the inheritance of an acquired trait, induced by the exposure of animals to viral infection

[Rechavi et al, 2011]

RNAs as extracellular signaling molecules





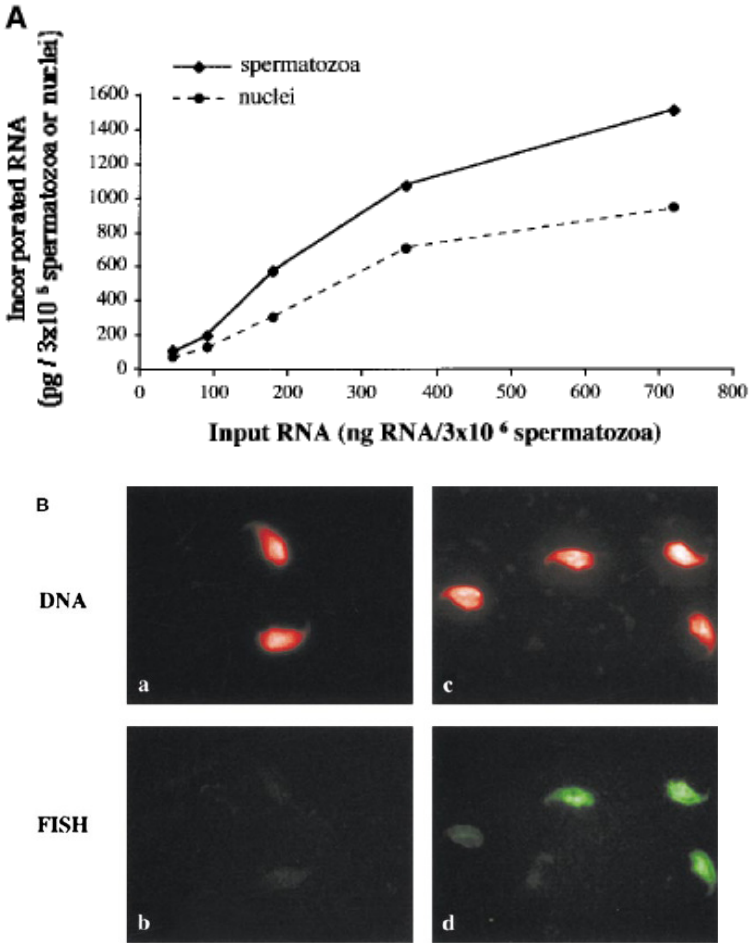
- Signaling **mRNA** (blue) as well as **ncRNA** (red) are (**a**) transcribed in the donor cell. These RNAs are then (**b**) trafficked and packaged into **vesicles**, which are emitted into (**c**) the extracellular environment. (**d**) The vesicles then dock and fuse with the target cell, releasing their RNA content. The mRNA may then be (**e**) translated in the target cell and the ncRNA may guide the chromatin modifying proteins (**ChM**) to establish (**f**) the new epigenetic state

- In addition (**g**), extracellular RNA molecules, such as **siRNAs** or **miRNAs**, may be transferred across the plasma membrane by specific receptors and channels, such as **Sid-1**. (**h**) These RNAs may regulate the inhibition of translation or mRNA degradation in the cell

- **MIR168a** -> **LDLRAP1**

[Dinger et al, 2008]

Evidence of RNA-uptake by mouse spermatozoa

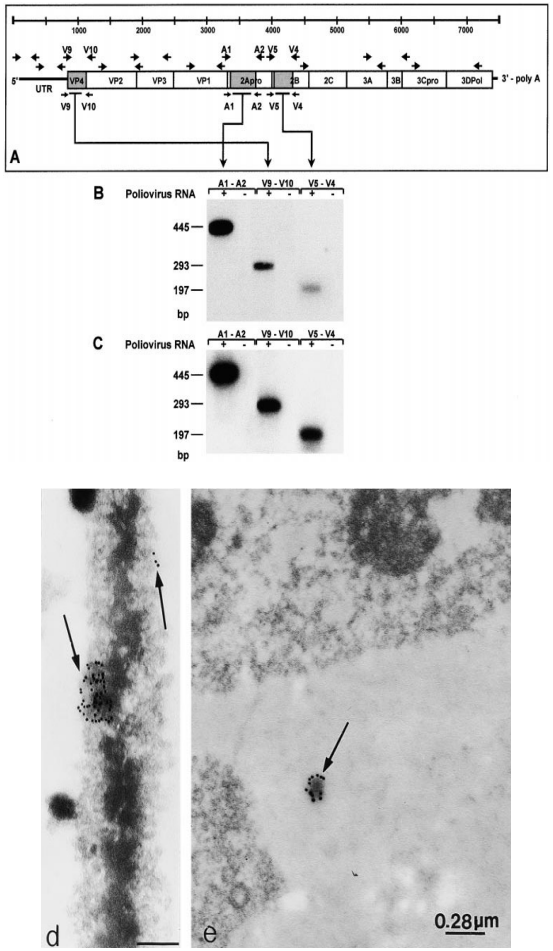


- **Association of the radioactive end-labeled poliovirus RNA with mouse epididymal sperm cells**. Spermatozoa were incubated with **poliovirus RNA**. After 30 min, spermatozoa from the mixture were washed and divided in 2 aliquots. The first one was dissolved in scintillation cocktail and counted to measure the RNA uptake; the second aliquot was used for nuclei purification, then treated and counted like whole cells to measure nuclear internalization

- **FISH of poliovirus RNA in isolated nuclei from RNA-treated spermatozoa**. (**a** and **c**) Sperm nuclei stained with DAPI and pseudocolored in red; (**b**) FISH with nuclei from spermatozoa incubated with buffer only; (**d**) FISH with nuclei from spermatozoa incubated with poliovirus RNA. The signal of the biotinylated probe is pseudocolored in green

[Giordano et al, 2000]

Reverse transcriptase activity was found in mouse sperm cells

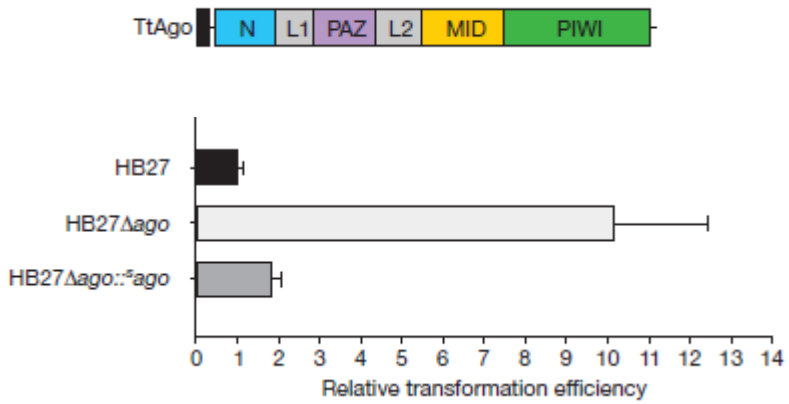


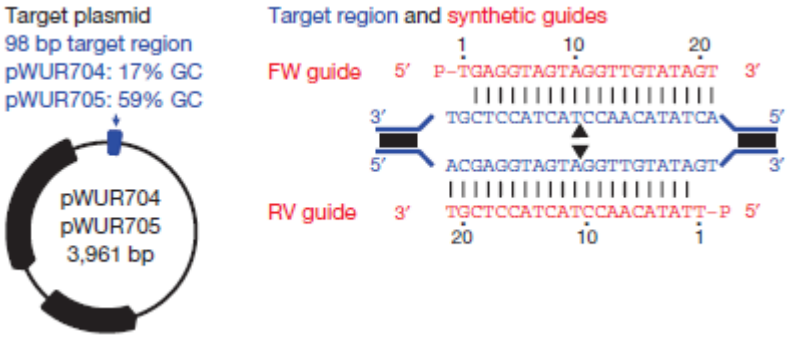
- **Human poliovirus RNA** replicates through a RNA(–) strand with no DNA intermediate. The poliovirus RNA was taken up by sperm cells, then reversetranscribed to cDNA copies, which were transferred to oocytes during IVF and further transmitted to 2-cell embryos

- **PCR of cDNA copies in sperm cells and 2-cell embryos after incubation with poliovirus RNA**. (**A**) Map of poliovirus RNA chromosome, (**B**) spermatozoa, and (**C**) 2-cell embryos. Amplified cDNAs were visualized by hybridization with internal oligonucleotide probes increasing specificity
- **Immunoelectronmicroscopy with anti-RT antibody showed that RT molecules were associated with the sperm nuclear scaffold**. (**d**) Sperm nuclear scaffolds, (**e**) HIV-infected T-lymphocyte

[Giordano et al, 2000]

DNA-guided DNA interference



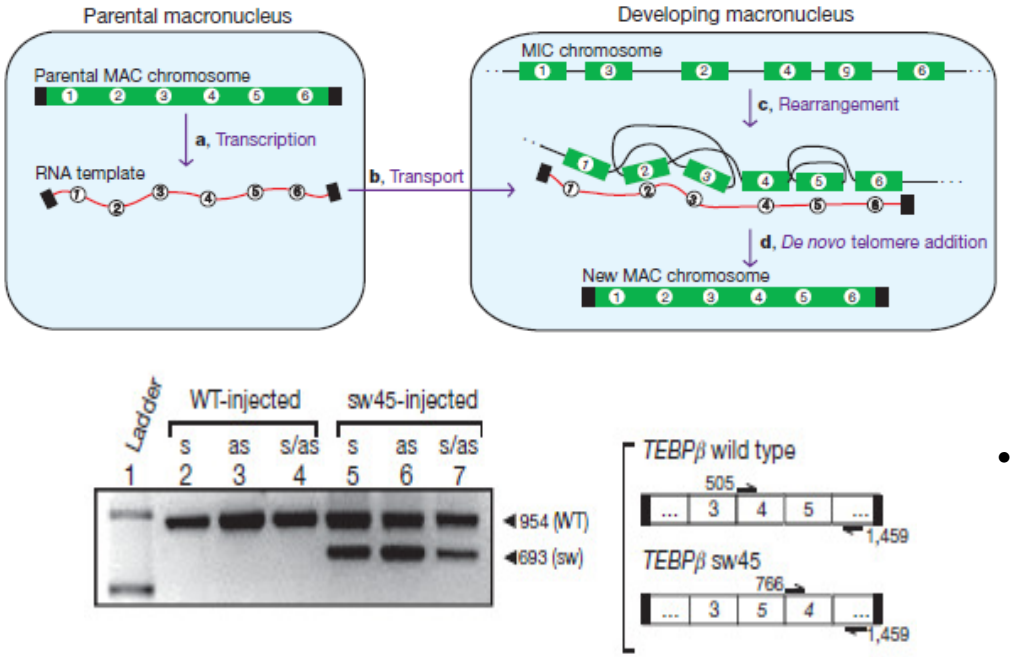


- The structure of **TtAgo** protein from the bacterium *Thermus thermophilus* is similar to the eukaryotic analog (**N**, **PAZ**, **MID**, and **PIWI** are structural domains, L1 and L2 are linkers)
- Transformation efficiency of wild type *T.thermophilus* HB27 significantly lower than HB27*Δago* mutant strain
- The guide targets invader plasmid independent of GC content
- **TtAgo-siDNA** complex cleaves a phosphate ester bond between the target nucleotides corresponding to guide nucleotides 10 and 11

[Swarts et al, 2014]

RNA-guided DNA modifications

Oxytricha trifallax is a unicellular eukaryote with 2 nuclei: **germline micronucleus** and **somatic macronucleus**. Micronucleus is transcriptionally inert, but transmits the germline genome to next generation. Macronucleus provides gene expression, but degrades during fertilization process. Deletion of transposons in ciliates during the formation of new macronucleus leads to genome fragmentation into 2 Kb '**nanochromosomes**' with just 1 gene, accomplishing 95% genome reduction and compressing 1 Gb germline (30,000 genes) into 50 Mb. It is an accurate mechanism of DNA rearrangements by maternal RNA cache

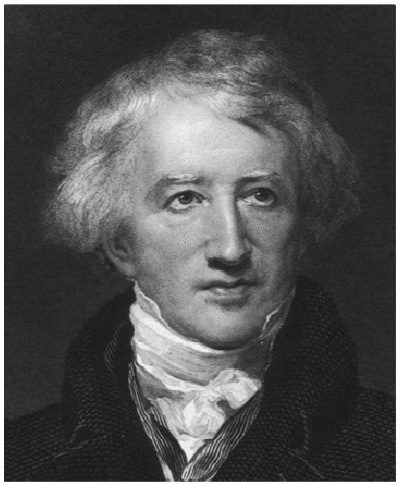


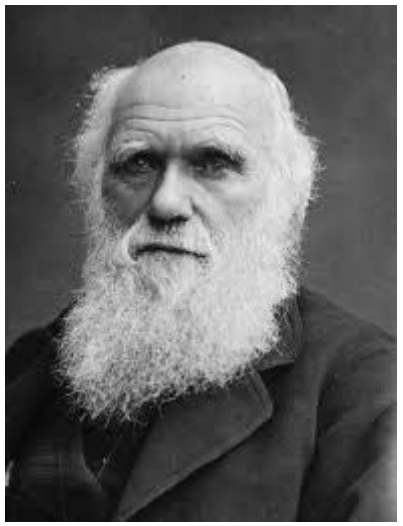
- **RNA-guided DNA rearrangements during macronuclear development**. **a**, RNA transcription of all DNA in the old macronucleus before its degradation. **b**, Transport of these RNA to the newly developing macronucleus, where they may act as scaffolds to guide DNA rearrangements (**c**). **d**, Telomere addition (black rectangles) and amplification of new macronuclear nanochromosomes

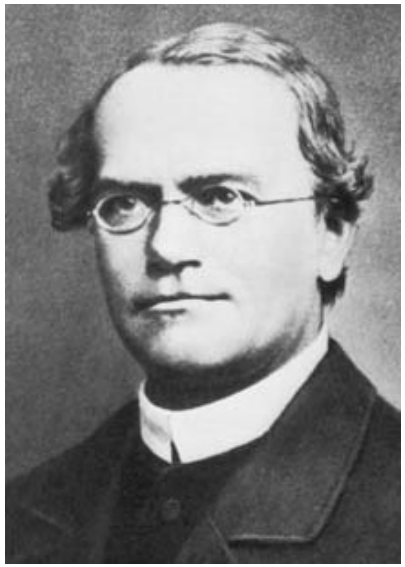
- **RNA template microinjection of Telomer-End-Binding Protein subunit β**. Sense (s), antisense (as), and combined (s/as) RNA were microinjected in both wild-type and switched orientations. Lanes 5–7 display the segments 4 and 5 have been switched (lower band)

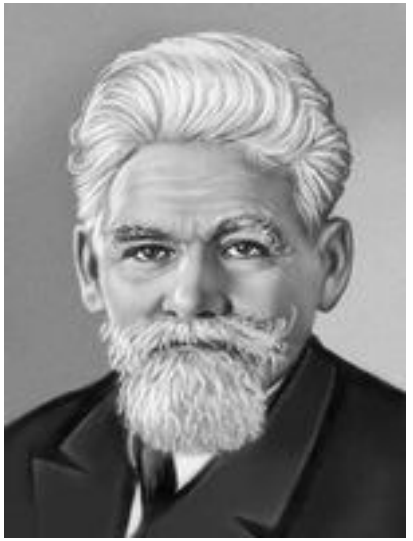
[Nowacki et al, 2007]

Crossroads in biology





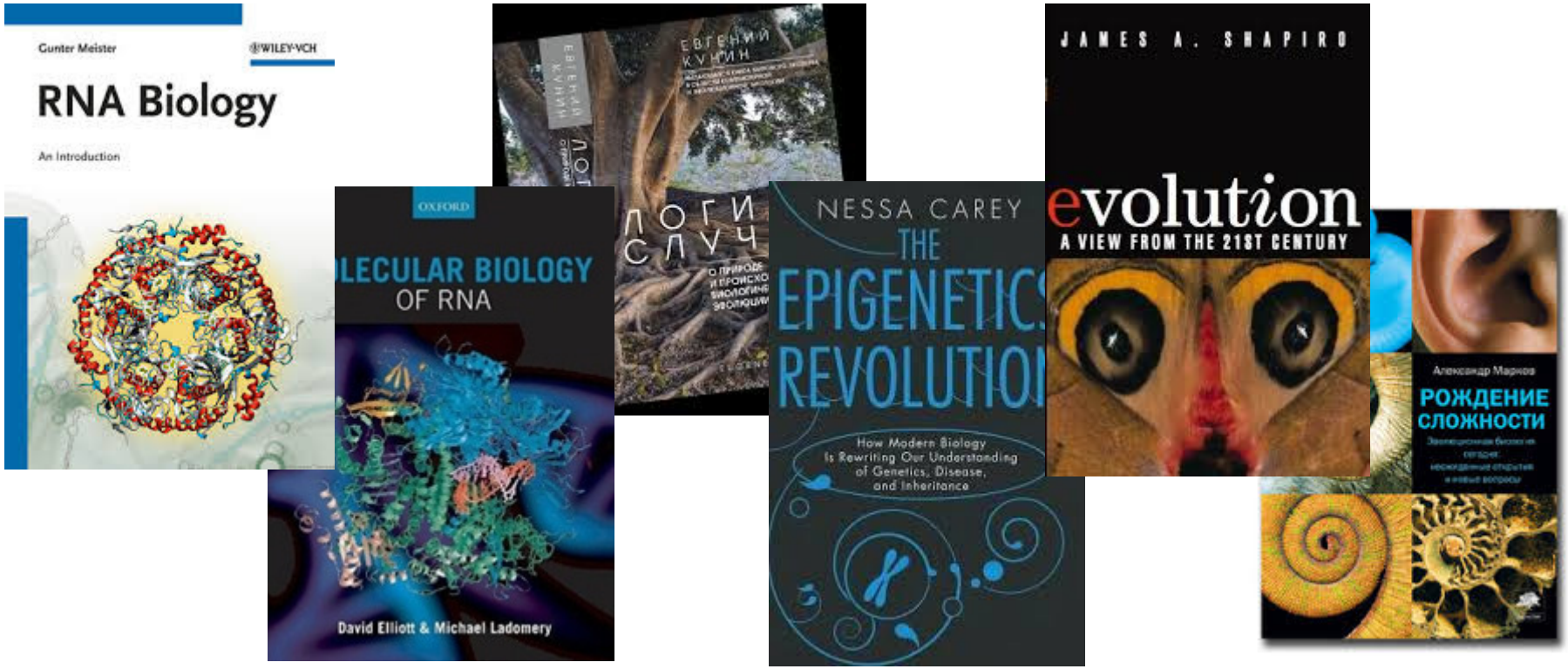




Conclusion

- The beauty of science to make things true. Small RNAs have gone from being “junk RNA”, transcriptional noise, and degradation intermediates to being the “**most important molecules**”
- **Drosha** and **Dicer** recognize dsRNA and generate from that short RNAs (21-22-nt) that are used for gene silencing. Both enzymes have RNase III domains and cut the substrate RNA on the bases of size and structure, rather than specific sequence
- Once produced, **siRNA** and **miRNA** act in essentially the same way. They are incorporated into **RISC** where guide RNA strand directs the molecular machine to complementary target RNA
- The guide RNA can also direct RISC with associated histone-modifying proteins to promoter regions where **it silences genes transcriptionally**
- Many of ncRNAs (**piRNAs**) are derived from transposons and silence transposon activity. An ancient RNAi machinery might have protected organisms from transposons and viruses that is similar to immunity
- There is still much more to learn

Additional reading



- Aravin *et al*, The Piwi-piRNA pathway provides an adaptive defense in the transposon arms race // Science. 2007 Nov 2;318(5851):761-4
- Giordano *et al*, Reverse transcriptase activity in mature spermatozoa of mouse // J Cell Biol. 2000 Mar 20;148(6):1107-13
- Rechavi *et al*, Transgenerational inheritance of an acquired small RNA-based antiviral response in *C.elegans* // Cell. 2011 Dec 9;147(6):1248-56.7
- Swarts *et al*, DNA-guided DNA interference by a prokaryotic Argonaute // Nature. 2014 Mar 13;507(7491):258-61
- Nowacki *et al*, RNA-mediated epigenetic programming of a genome-rearrangement pathway // Nature. 2008 Jan 10;451(7175):153-8