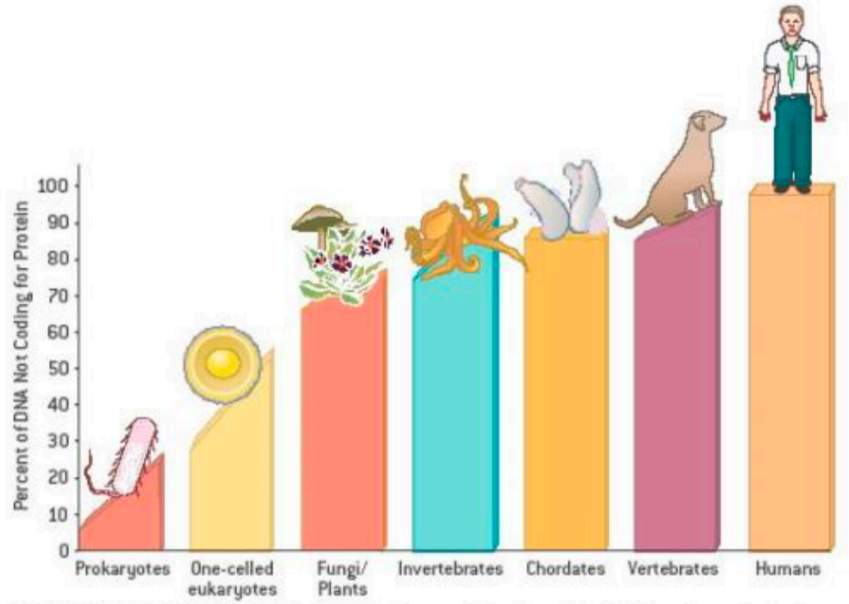
Loss of function approaches

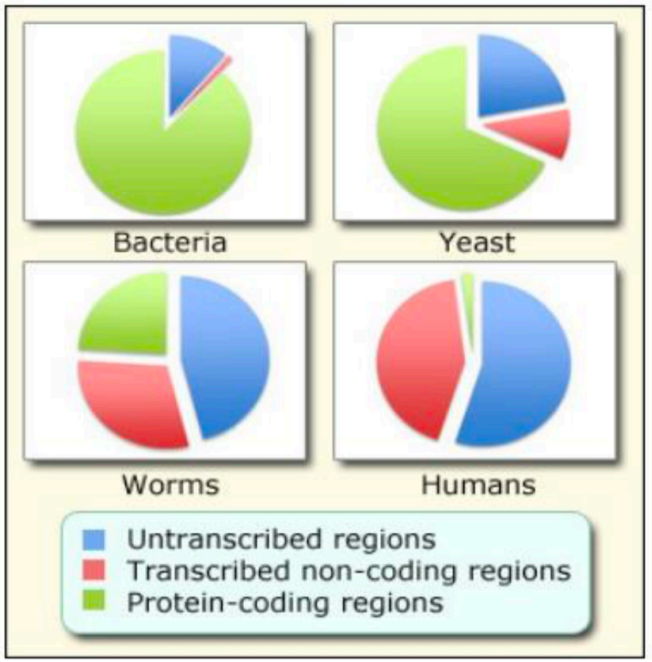
siRNA and
Genome editing by CRIPS/Cas9

Stefan Schoeftner

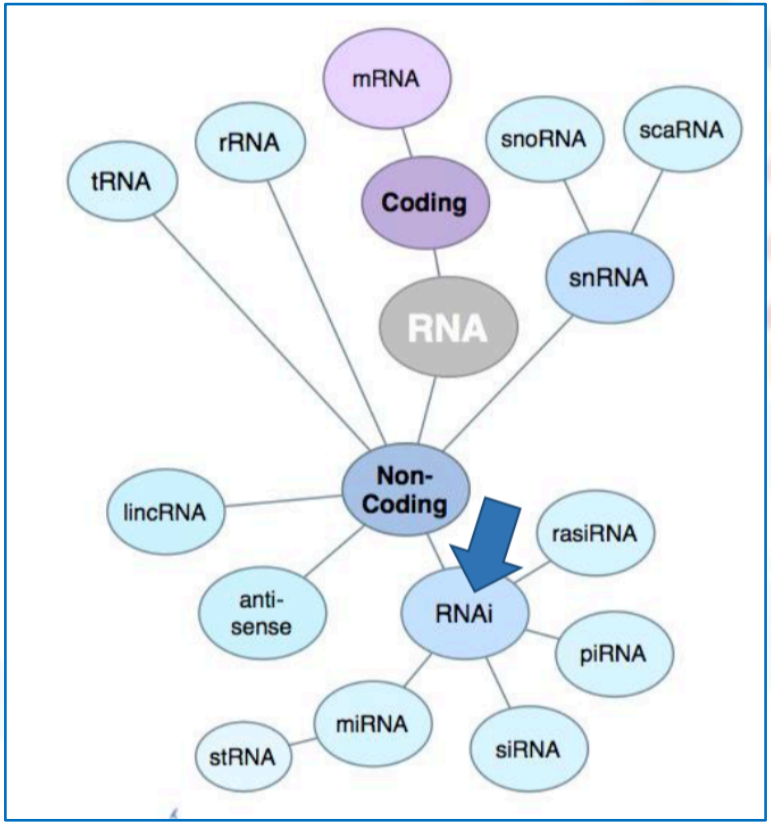
Introduction – Non coding sequences increase with complexity



NONPROTEIN-CODING SEQUENCES make up only a small fraction of the DNA of prokaryotes. Among eukaryotes, as their complexity increases, generally so, too, does the proportion of their DNA that does not code for protein. The noncoding sequences have been considered junk, but perhaps it actually helps to explain organisms' complexity.



Introduction – Non coding RNA (ncRNA) forms



Molecule	Function
mRNAs	messenger RNAs, code for proteins
rRNAs	ribosomal RNAs, form the basic structure of the ribosome and catalyze protein synthesis
tRNAs	transfer RNAs, central to protein synthesis as adaptors between mRNA and amino acids
snRNAs	small nuclear RNAs, function in a variety of nuclear processes, including the splicing of pre-mRNA
snoRNAs	small nucleolar RNAs, used to process and chemically modify rRNAs
scaRNAs	small cajal RNAs, used to modify snoRNAs and snRNAs
miRNAs	microRNAs, regulate gene expression typically by blocking translation of selective mRNAs
siRNAs	small interfering RNAs, turn off gene expression by directing degradation of selective mRNAs and the establishment of compact chromatin structures
Other non-coding RNAs	function in diverse cell processes, including telomere synthesis, X-chromosome inactivation, and the transport of proteins into the ER

Co-suppression of gene expression

The discovery of RNAi was preceded first by observations of transcriptional inhibition by antisense RNA expressed in transgenic plants.

Reports of unexpected outcomes in experiments performed by plant scientists in the United States and the Netherlands in the **early 1990s**.

Attempted to overexpress **chalone synthase** (anthrocyanin pigment gene) in petunia (trying to darken flower color) caused **the loss** of pigment .

Further investigation of the phenomenon in plants indicated that the downregulation was due to **post-transcriptional inhibition of gene expression** via an increased rate of mRNA degradation.

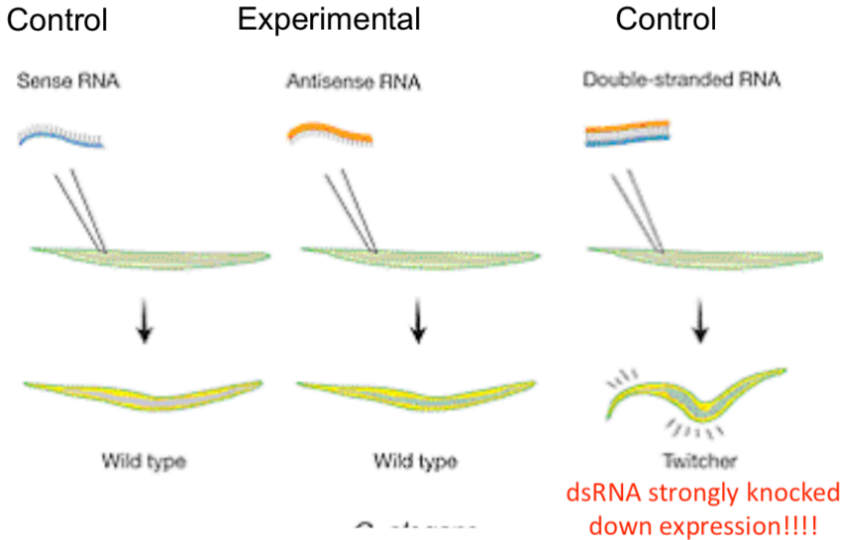
This phenomenon was called **co-suppression of gene expression**, **because suppressed expression of both endogenous gene and transgene** but the molecular mechanism remained unknown

RNA interference (RNAi)

Previously known as co-suppression or post transcriptional gene silencing (PTGS), now is known as **RNA interference (RNAi) a**s a process within living cells that moderates the activity of their genes.

Accidental Discovery of RNAi

- Goal: silence endogenous mRNAs with antisense RNA
- The *unc-22* gene encodes a myofilament protein.
- Decrease in *unc-22* activity is known to produce severe twitching movements.



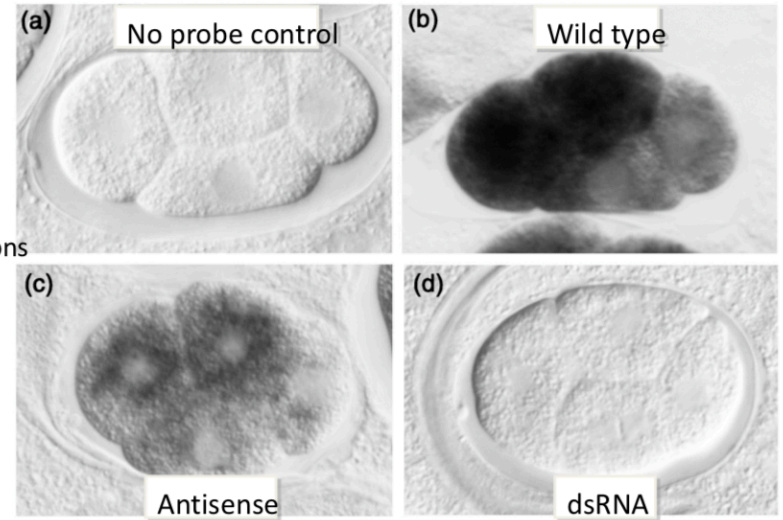
Fire *et al.* *Nature* 1997

Phenotypic effect after injection of ssRNA or dsRNA (*unc-22)* into the gonad of *C. elegans.*

RNA interference (RNAi)

Injection of dsRNA in *C. elegans* Shown To Cause Destruction of Specific mRNA

- Mello and colleagues, 1998
- Injection in gonads of dsRNA for *mex-3* (abundant RNA) gave much more efficient inhibition in embryos than antisense RNA
- dsRNA had to include exons; introns and promoter didn't work
- Effect was incredibly potent and even spread to other cells within the worm
- Termed 'RNA Interference'
- Incredibly useful as a tool for molecular biology



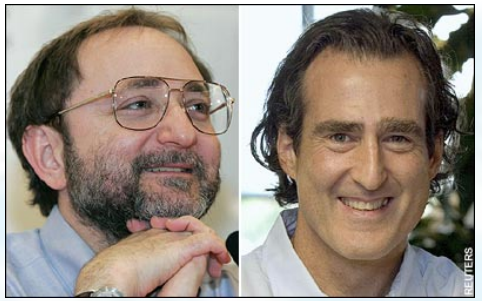
in situ hybridization four-cell stage embryo

Fire *et al.* *Nature* 1998

- dsRNA from mature mRNA elicits RNAi
- dsRNA from introns does not
- RNAi results in decreased mRNA levels
- RNAi is heritable (for a few generations)
- RNAi only requires a few molecules of dsRNA per cell
- RNAi is applicable to many different transcripts

RNA interference (RNAi)

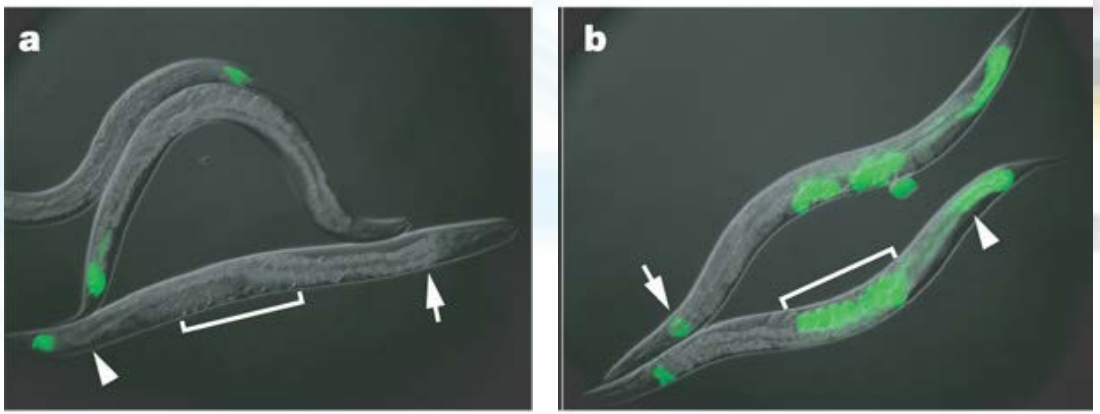
In **2006**, Andrew Fire and Craig C. Mello shared the Nobel Prize in Physiology or Medicine for their work on RNAi in the nematode worm C. elegans.



Andrew Fire Craig Mello

RNAi in *C.elegans*

– Silencing of a green fluorescent protein (**GFP**) reporter in *C. elegans* occurs when animals feed on bacteria expressing GFP dsRNA (**a**) but not in animals that are defective for RNAi (**b**).

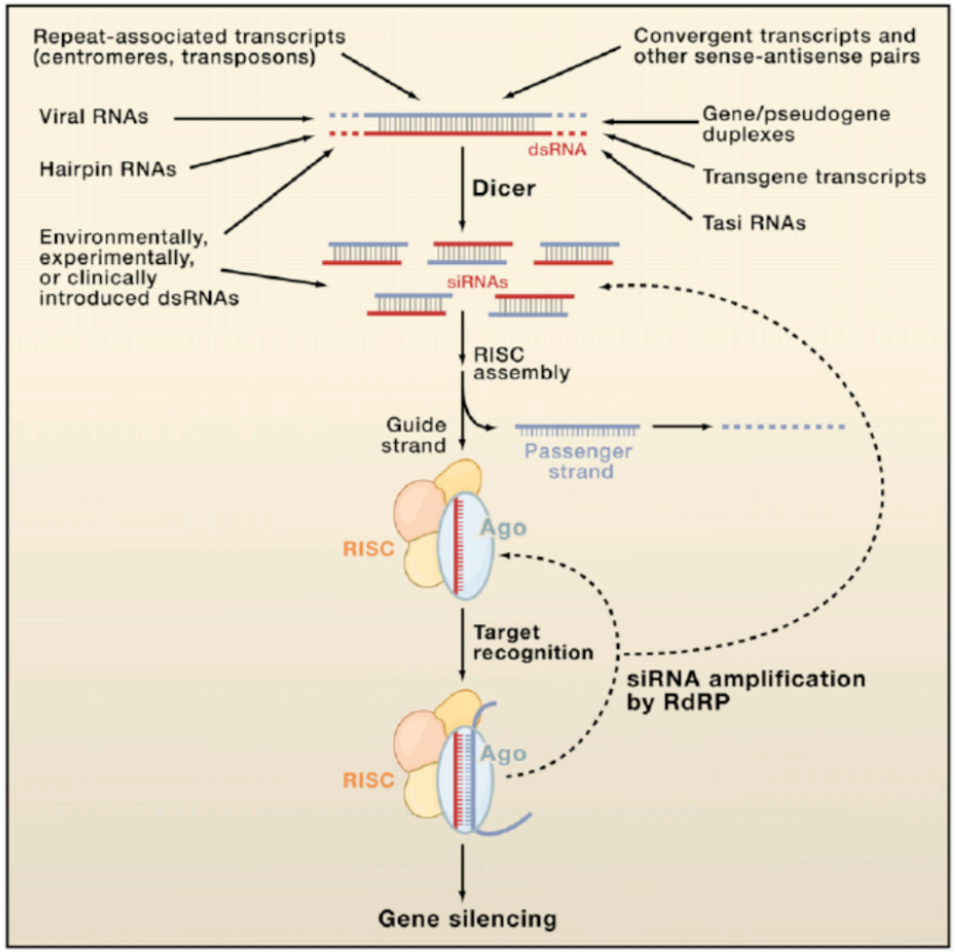


The lack of GFP-positive embryos in a (bracketed region) demonstrates the systemic spread

Endogenous sources of dsRNA for siRNA formation

Sources of dsRNA

- Some dsRNAs have viral origin, but not all
- Genomic repetitive sequences also are source of siRNA
- Some even regulate other genes (ta-siRNA for trans-acting in plants)
- exo siRNAs (viral etc)
- endo siRNAs –the precursor has a nuclear phase (hairpins, sense-antisense transcripts etc)



Carthew and Sontheimer, Cell (2009) 136, 642-655.

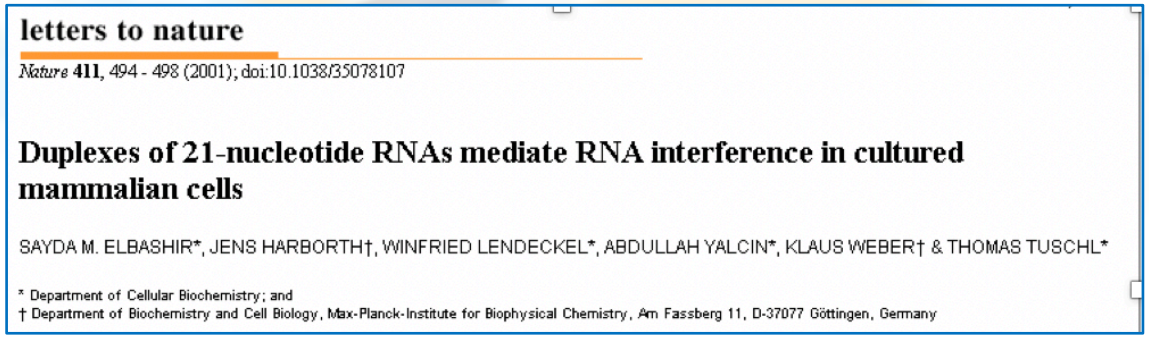
siRNA and miRNA

Two types of RNA molecules involved:
- **small interfering RNA (siRNA)**
- **microRNA (miRNA)**

They bind to other specific mRNAs and modulate their activity.

RNA interference has played an important role in defending cells against parasitic nucleotide sequences – viruses and transposons – but also in directing development as well as gene expression in general.

In 2001 first report of RNAi in MAMMALS



letters to nature

Nature **411**, 494 - 498 (2001); doi:10.1038/35078107

Duplexes of 21-nucleotide RNAs mediate RNA interference in cultured mammalian cells

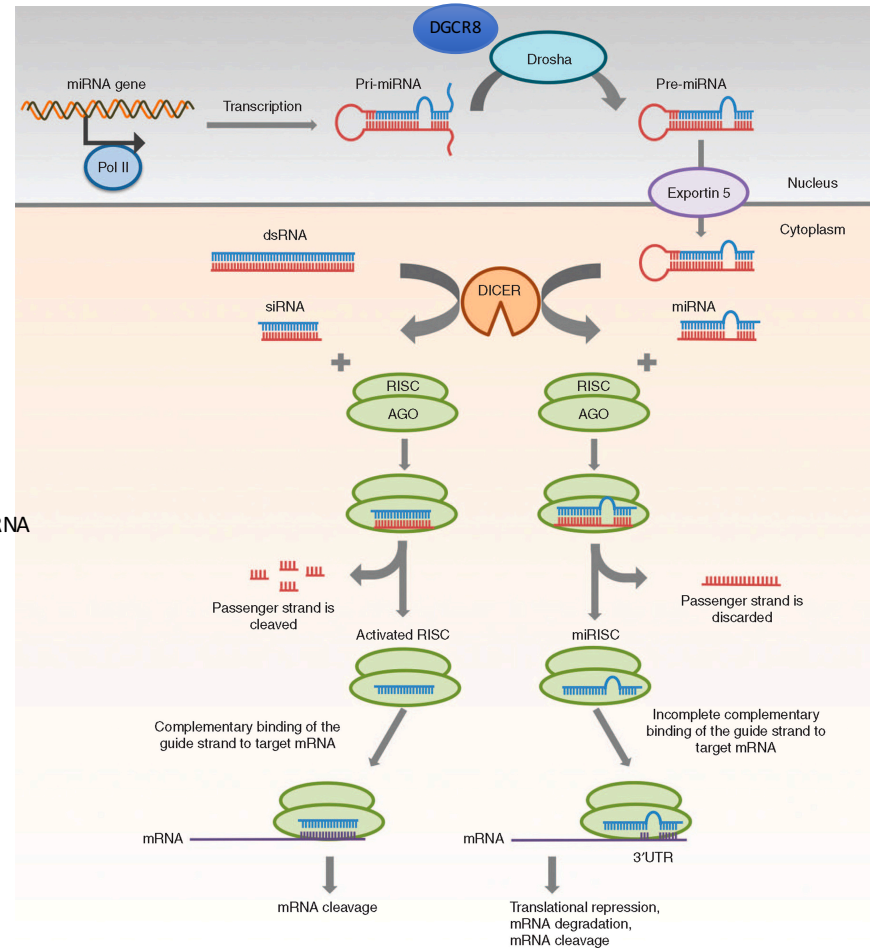
SAYDA M. ELBASHIR*, JENS HARBORTH†, WINFRIED LENDECKEL*, ABDULLAH YALCIN*, KLAUS WEBER† & THOMAS TUSCHL*

* Department of Cellular Biochemistry; and
† Department of Biochemistry and Cell Biology, Max-Planck-Institute for Biophysical Chemistry, Am Fassberg 11, D-37077 Göttingen, Germany

siRNA and miRNA biogenesis and gene regulation

siRNA biogenesis

1. dsRNA production and transfer to cytoplasma
2. RNaseII family enzyme Dicer processes pre-miRNA generating a 20-25 base dsRNA with overhang at the 3'end (2 bases)
3. Transfer of dsRNA to RISC complex (RNA induced silencing complex)
4. Selection of guide RNA → regulatory RNA passenger RNA → will be eliminated
7. RISC complex+guide RNA → regulatory function
   A. Perfect target RNA matching → RNA degradation = siRNA effect (cutting = "slicing")

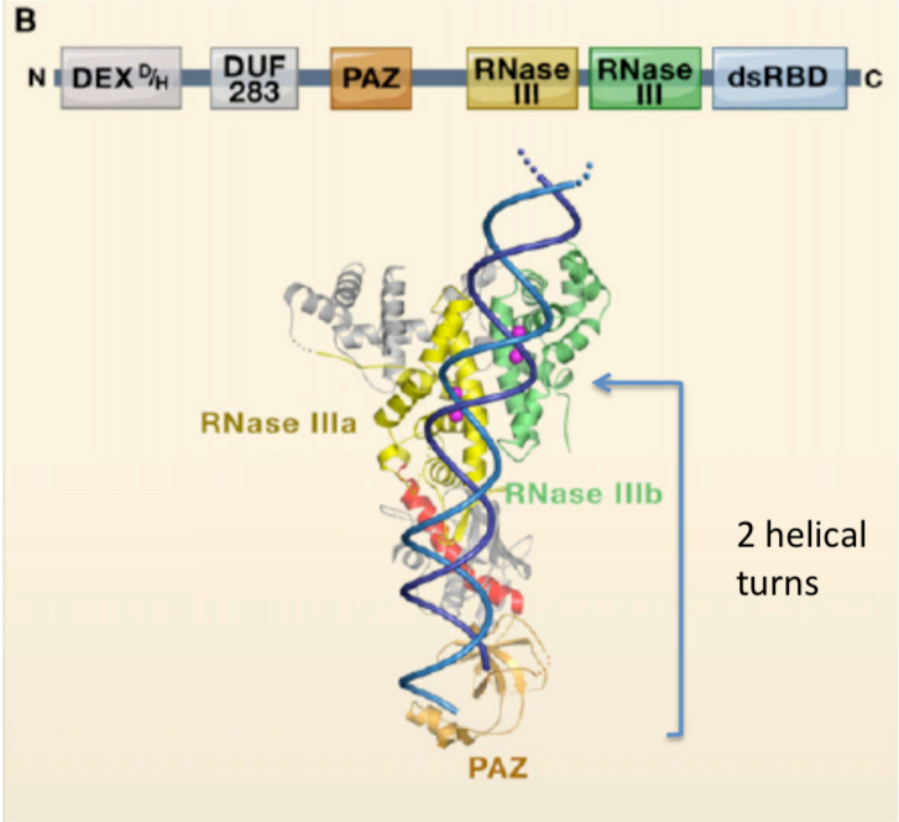


miRNA biogenesis

1. Long, unprocessed precursor dsRNA or stem loop RNA **(pri-miRNA)** produced in a independent and controlled manner from miRNA hosting gene
2. Processing in the nucleus by the RNaseIII family protein Drosha generates a stem-loop RNA with characteristic length of 65-70 nucleotides. Drosha is in complex with DGCR8 that is important for Drosha activity
3. Exportin 5-RanGTP transports pre-miRNA in ternary complex thought nuclear pore to cytoplasm. RanGAP stimulates GTP; pre-miRNA released from Exportin.
4. RNaseII family enzyme Dicer processes pre-miRNA generating a 20-25 base dsRNA with overhang at the 3'end (2 bases)
5. Transfer of dsRNA to RISC complex (RNA induced silencing complex)
6. Selection of guide RNA → regulatory RNA passenger RNA → will be eliminated
7. RISC complex+guide RNA → regulatory function
   A. RNA degradation = siRNA effect (cutting = "slicing"
   B. inhibition of mRNA translation =mRNA effect
   C. transfer to nucleus and chromatin regulation = siRNA mediated silencing

DICER: Producer of Small (21-23 bp) RNA fragments

- Structure solved by Doudna and colleagues (2006)
- PAZ domain binds RNA end, RNase III domains cut RNA to produce 2 nt 3'-overhang
- Roles of other domains (not present in structure) remain unclear

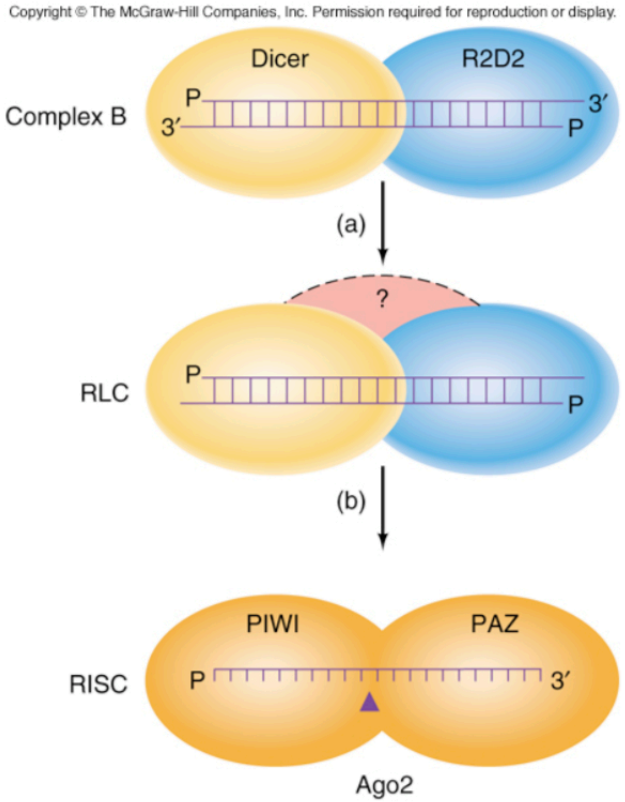


Carthew and Sontheimer, Cell (2009) 136, 642-655.

DICER partners and RISC

Assembly of the RNA-Induced Silencing Complex (RISC) Involves Additional Proteins

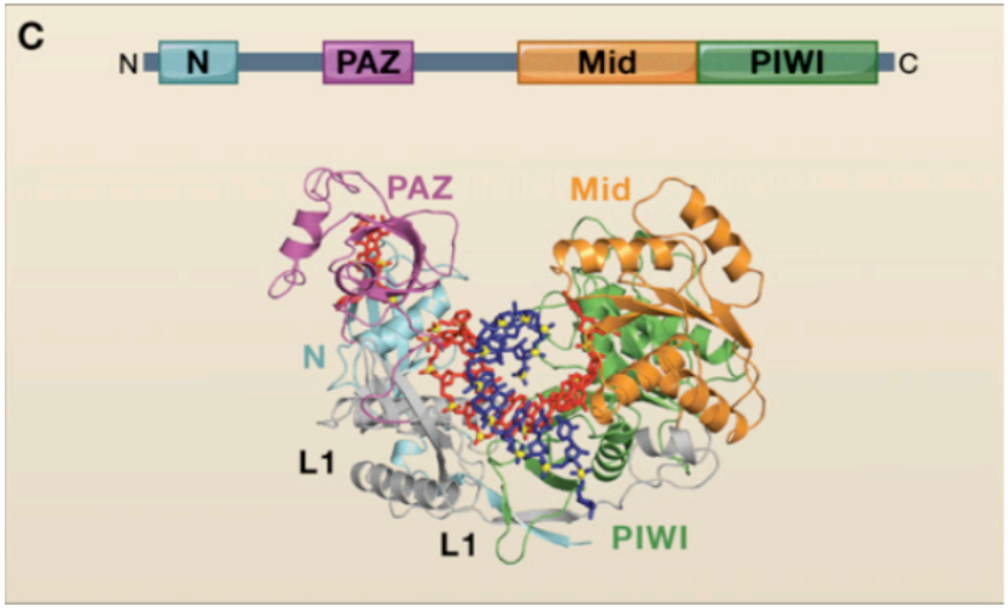
- Processing of dsRNAs into RISC requires accessory proteins: TRBP (R2D2 in *Drosophila*) forms complex with Dicer
- Other unknown proteins bind to form RISC Loading Complex
- Ago2 cleaves the passenger strand, leading to its ejection



ARGONAUTE: Central component of RISC

- One strand of the dsRNA produced by Dicer is retained in the RISC complex in association with Argonaute
- Structure first solved by Leemor-Tor and colleagues (2004), more recent structures by Patel and colleagues include RNAs mimicking guide ssRNA and target mRNA

- The PAZ domain has RNA 3' end binding activity
- In structure without mRNA, guide strand nucleotides 2-6 have bases exposed and available for base-pairing
- PIWI domain adopts RNase H fold and in some Ago proteins can cleave the 'passenger strand' : I.e. the mRNA



Carthew and Sontheimer, Cell (2009) 136, 642-655.

siRNA: Exogenous dsRNA molecules

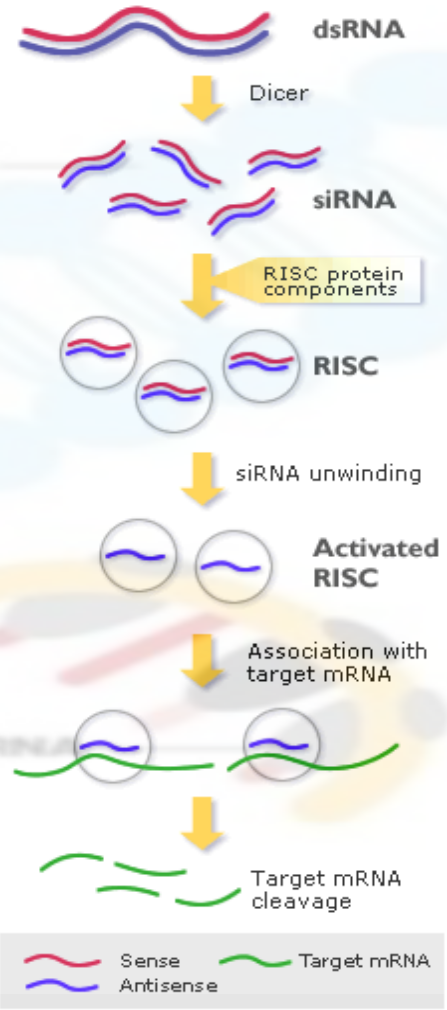
RNAi is controlled by **RISC** and is initiated by short dsRNA molecules in a cell's cytoplasm, where they interact with the catalytic RISC component argonaute.

dsRNAs is cleaved by the **Dicer enzyme** into short fragments of ~20 nucleotides that are called **siRNAs**.

Each siRNA is unwound into two single-stranded (ss) ssRNAs (**passenger** strand and the **guide** strand).

The passenger strand is degraded (red), and **the guide strand (blue) is incorporated into the RNA-induced silencing complex (RISC).**

The most well-studied outcome is post-transcriptional gene silencing, which occurs when the guide strand base pairs with a complementary sequence in a messenger RNA molecule (green) and induces **cleavage by Argonaute**, the catalytic component of the RISC complex.



miRNA: Endogenous RNA silencing

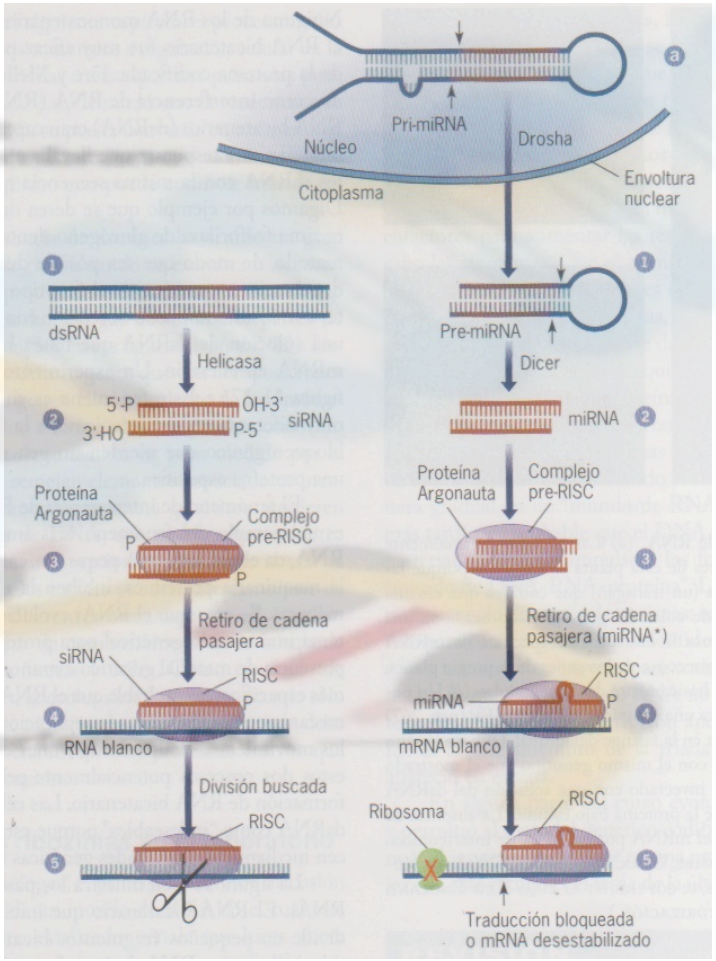
miRNAs are genomically encoded non-coding RNAs that regulate gene expression, particularly during development.

Mature miRNAs are structurally similar to siRNAs produced from exogenous dsRNA but **must undergo post-transcriptional modification**.

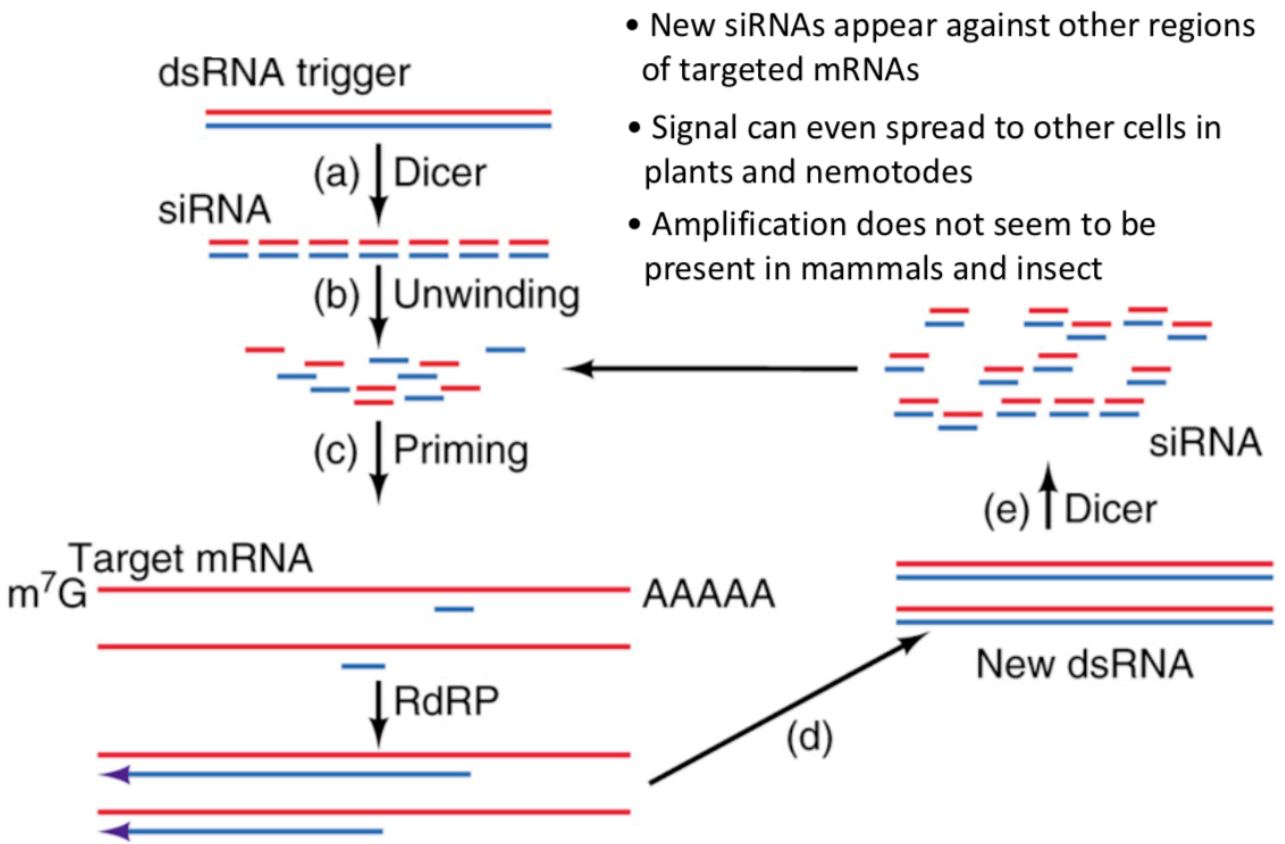
miRNA's are expressed from longer RNA-coding gene as a primary transcript (**pri-miRNA**) which is processed within the cell nucleus to a 70 bp stem-loop structure (**pre-miRNA**) by the microprocessor complex (RNase III **Drosha** and dsRNA binding protein DGCR8).

The dsRNA molecule is bound and cleaved by Dicer to produce the mature miRNA molecule that can be integrated into the RISC complex; thus, **miRNA and siRNA share the same cellular machinery downstream of their initial processing**.

miRNAs typically inhibit the translation of many different mRNAs with similar sequences. In contrast, siRNAs typically inhibit only a single, specific target.



In some organisms, siRNA signal is amplified and spread



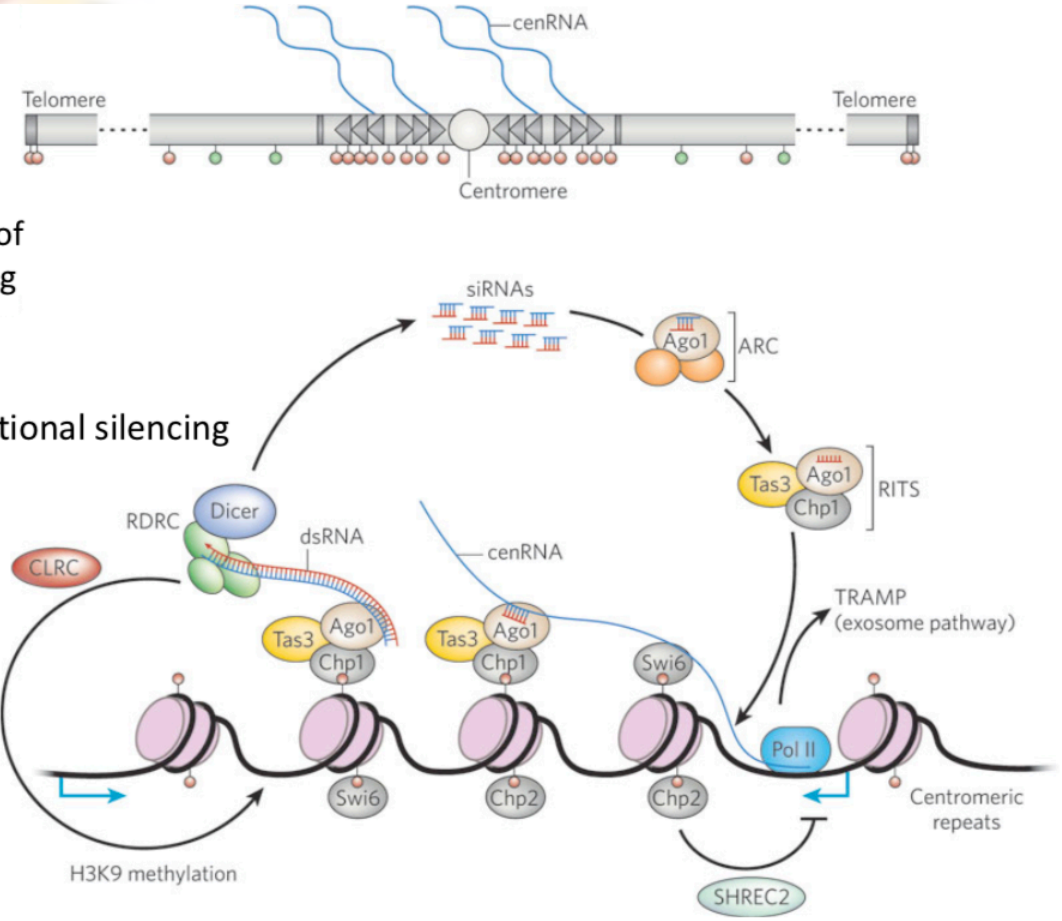
dsRNA processed to siRNAs in the nucleus: silencing by formation of heterochromatin

- Pathway best understood in *S. pombe*
- Silencing involves formation of heterochromatin and resulting transcriptional repression

the RNA-induced transcriptional silencing complex (RITS)

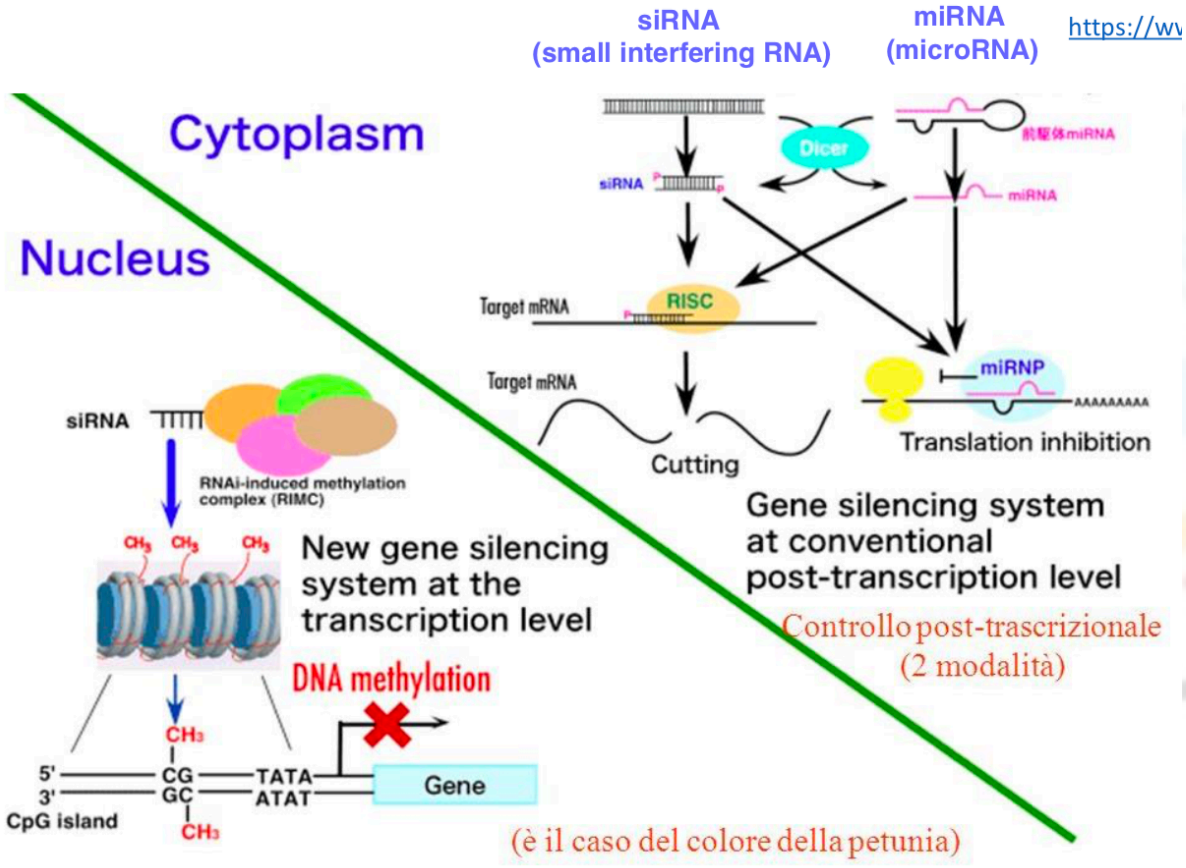
Clr4 methyltransferase complex (CLRC)

RNA-directed RNA polymerase complex (RDRC)



Moazed, *Nature* (2009) 457, 413-420

RNAi models


siRNA
(small interfering RNA)
miRNA
(microRNA)
https://wv
Cytoplasm
Nucleus
Dicer
siRNA
miRNA
前駆体miRNA
Target mRNA
RISC
Target mRNA
Cutting
miRNP
AAAAAAAAA
Translation inhibition
siRNA
RNAi-induced methylation complex (RIMC)
New gene silencing system at the transcription level
DNA methylation
CH3
5'
3'
CG
GC
TATA
ATAT
CpG island
Gene
Gene silencing system at conventional post-transcription level
Controllo post-trascrizionale
(2 modalità)
(è il caso del colore della petunia)

Medical application– Gene regulation

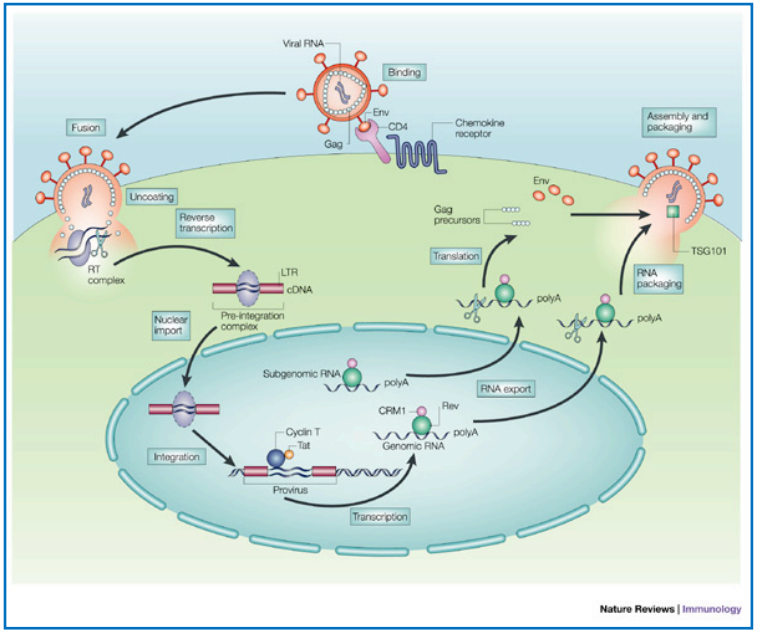
It is difficult to introduce long dsRNA strands into mammalian cells due to the interferon response, the use of siRNA mimics has been more successful.

First applications to reach clinical trials were: the treatment of macular degeneration and respiratory syncytial virus,

RNAi has also been shown to be effective in the reversal of induced liver failure in mouse models.

Other proposed clinical uses center on antiviral therapies:

- HSV type 2
- knockdown of host HIV receptors
- silencing of HIV, HAV,HBV and flu genes
- inhibition of measles viral replication.



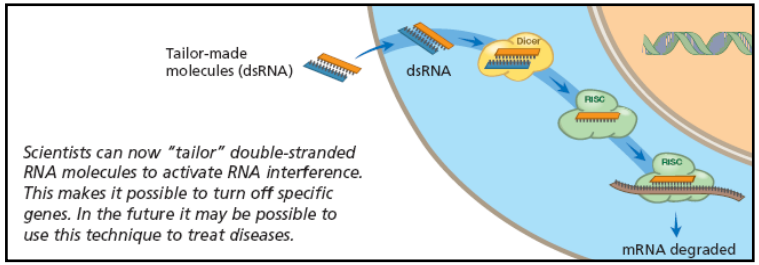
Viruses like HIV-1 are particularly difficult targets for RNAi-attack because they are escape-prone, which requires combinatorial RNAi strategies to prevent viral escape.

RNAi
dalla teoria alla pratica di laboratorio

convertire un meccanismo biologico in uno strumento per eliminare l'espressione di un gene di interesse in modo semplice --> siRNA sintetici

Come disegnare un siRNA in lab

Trasfezione transiente → liposomes



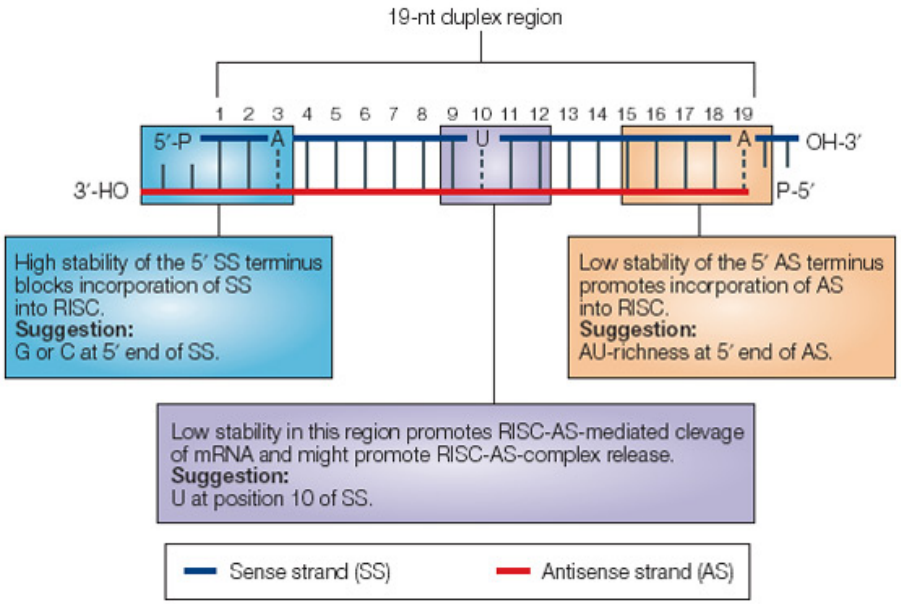
siRNA PROFILING TECHNOLOGIES

Disegno di siRNA

- il siRNA possiede un gruppo UU al terminale 3'
- 19 nucleotidi
- un contenuto in G/C < 50% è preferibile.

Algoritmo di nuova generazione

- Tm dell'siRNA
- Effetti della posizione nucleotidica
- Contenuto nucleotidico dei 3' overhangs
- Distribuzione nucleotidica
- Controllo della specificità



Come disegnare un siRNA in lab

Preventing Off-Target Effects

Overabundance of the siRNA activates the interferon pathway, as antiviral response

Low concentrations (~5-30nM) of single siRNA minimizes:
- **chances of off-target effect**
- **induction of interferon response**

It is currently preferable to use **ONE** highly potent siRNA than a **MIXTURE** of siRNAs that raise overall siRNA conc.

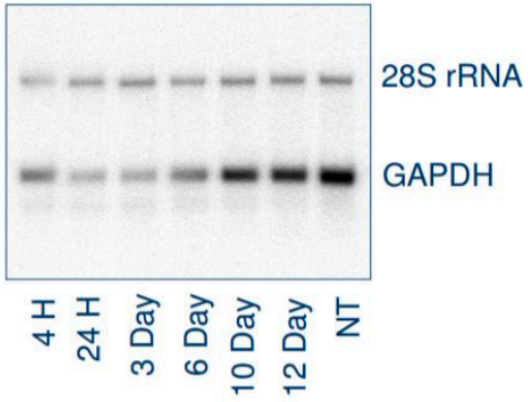
Verify specificity of RNAi effect by testing independent siRNAs to the same target

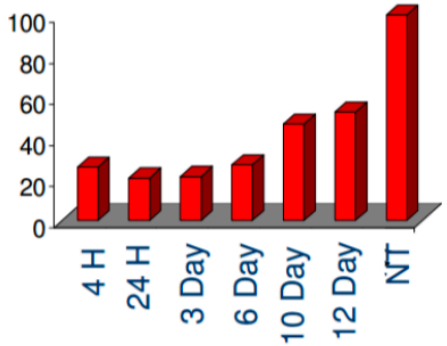
Durata del silenziamento transiente

Trasfezione transiente → liposomes

siGAPDH

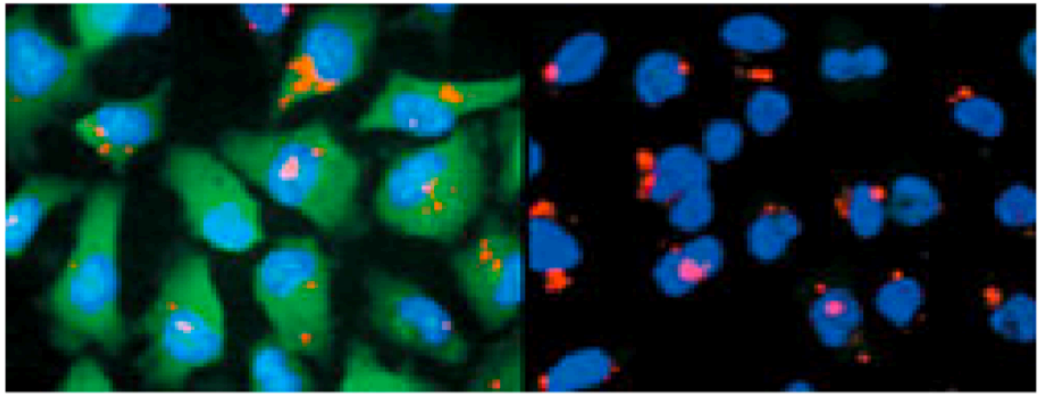
Northern blotting





CELLULE HeLa

In rosso: siRNA marcato
In blu: nuclei
In verde la proteina GAPDH



si RNA non specifico

si RNA contro GAPDH

Trattamento: 48h

Trasfezione con il siRNA: ottimizzazione delle condizioni

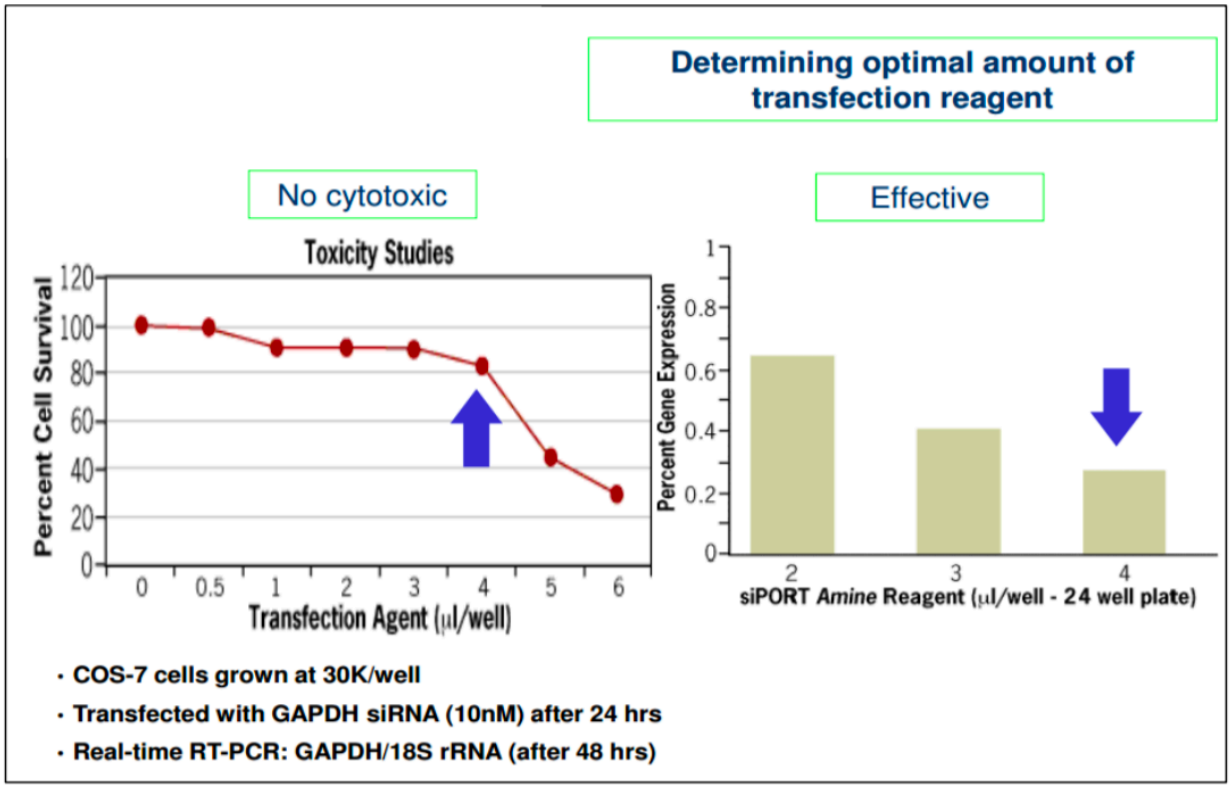
Prevenire effetti di spegnimento del target:

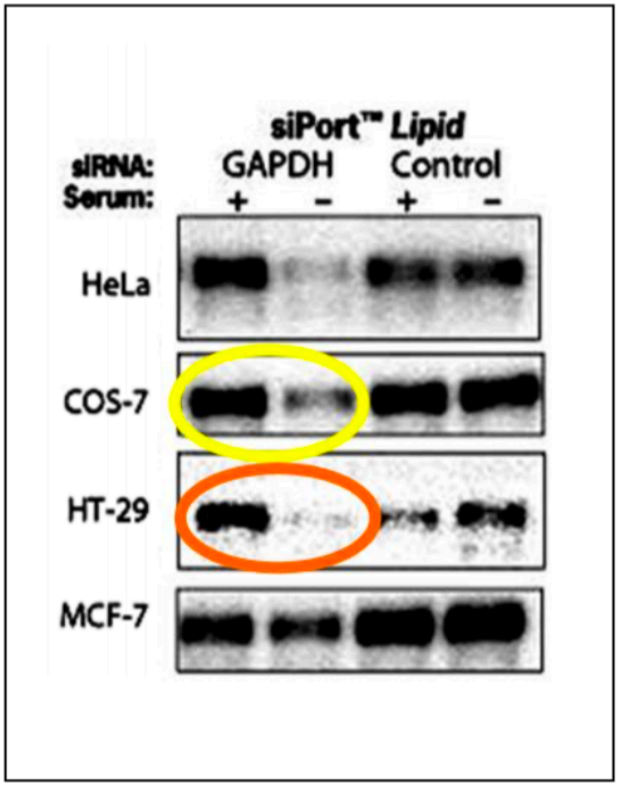
- **Basse concentrazioni** (~5-30 nM) di siRNA per minimizzare l'attivazione l'**interferon pathway** come risposta anti-virale
- E' preferibile usare un solo siRNA molto efficiente piuttosto che una miscela di siRNA meno potenti, la MIXTURE fa aumentare la concentrazione totale
- Usare RNAi specifici, dopo aver effettuato test di siRNA differenti sullo stesso mRNA bersaglio

Trasfezione con il siRNA: ottimizzazione delle condizioni

Scelta del reagente trasfettante

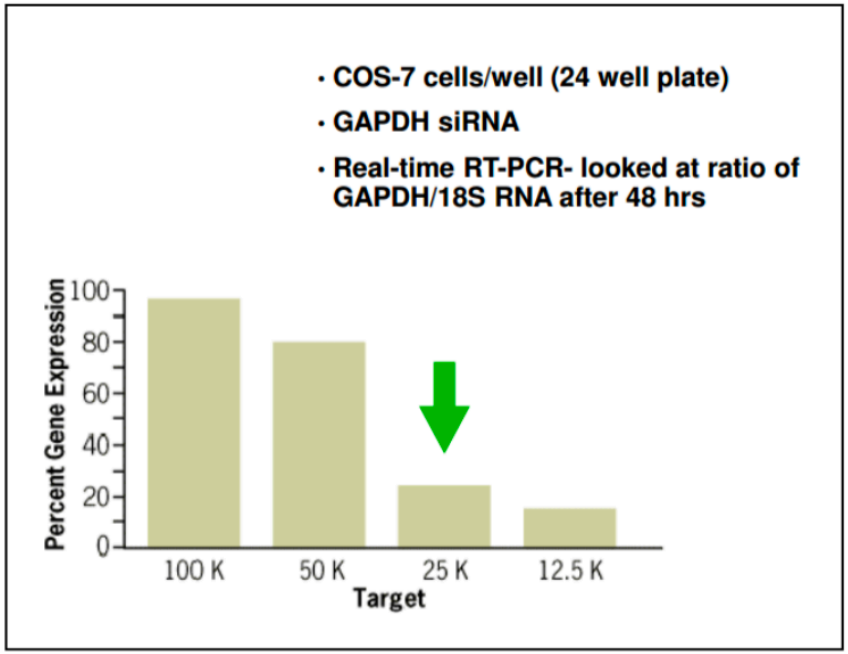
Presenza ed assenza di siero



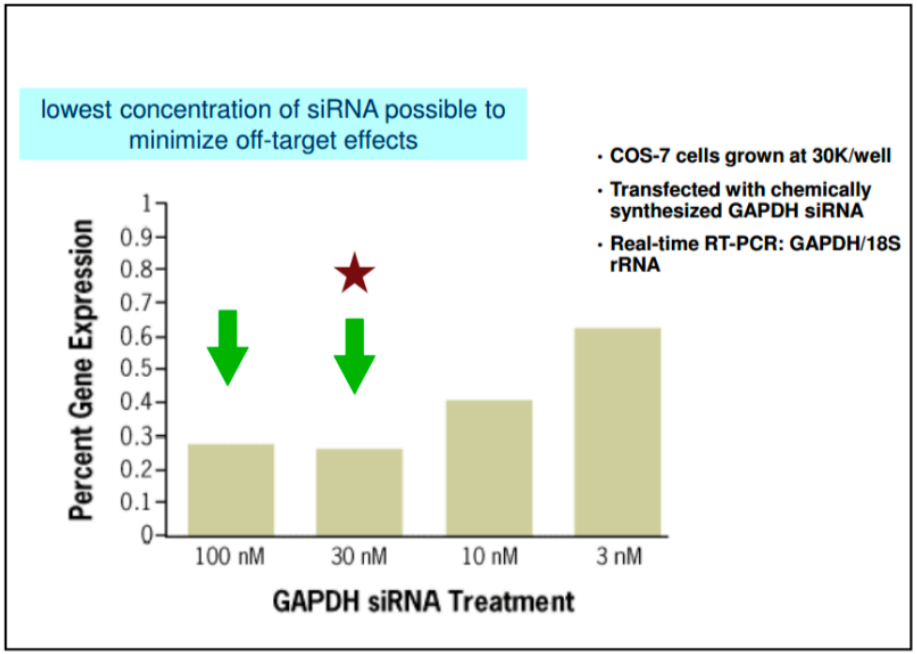


Trasfezione con il siRNA: ottimizzazione delle condizioni

Determinare la densità di semina delle cellule



Concentrazione di siRNA



Dal transiente alla trasfezione con vettori

PRO

- La trasfezione con siRNA è davvero molto efficiente in molti tipi di cellule
- Coi siRNA il silenziamento è immediato

CONTRO

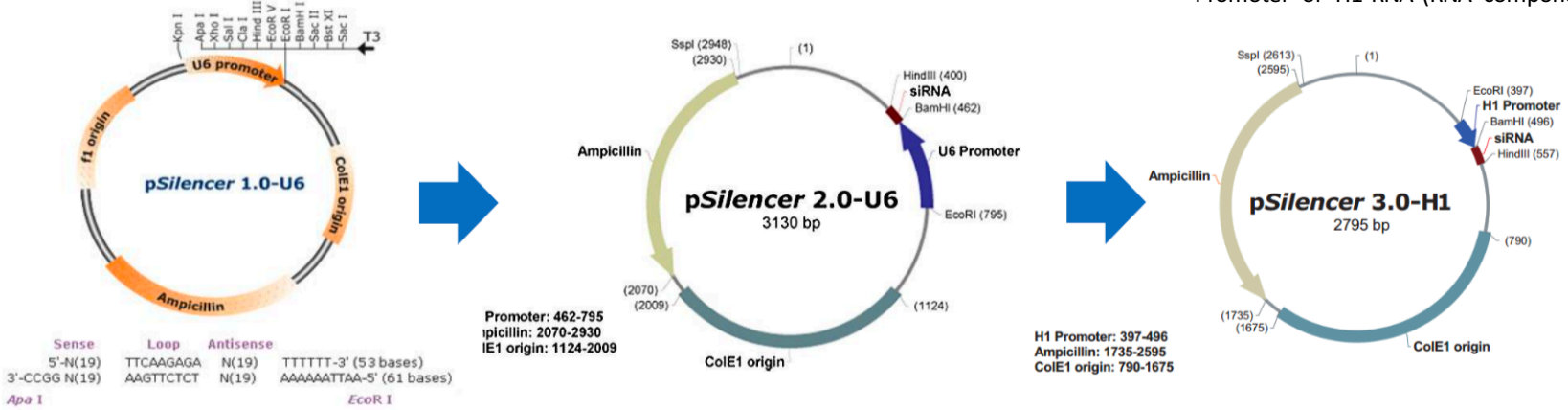
- Alcune cellule sono refrattarie alla trasfezione e la loro elettroporazione spesso causa morte cellulare
- I siRNA sono stabili, ma la trascrizione può risultare transiente se le cellule si duplicano molto in fretta diluendo il silenziamento e la vita media della proteina

Superamento del problema mediante……

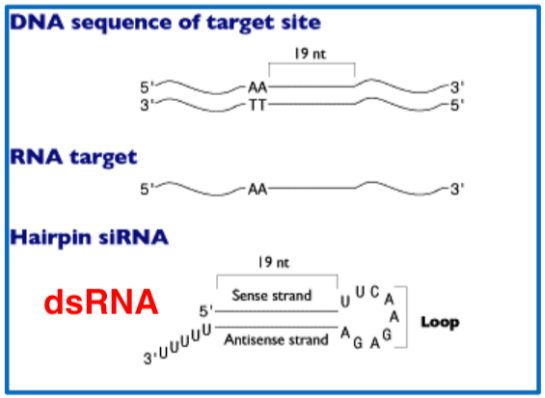
Dal transiente alla trasfezione con vettori

Silenziamento tramite vettori a DNA

Important:
Use RNA Pol III promoters (short RNA)
Promoter or U6 snRNA
Promoter of H1 RNA (RNA component of human RNase P)



Mimicking miRNA production



- Una sequenza stampo per un "hairpin siRNA" viene clonata in un opportuno vettore per trascrivere una molecola di RNA
- Produzione di siRNA *in vivo direttamente* nelle cellule trasfettate
- Trasfezione stabile nella linea cellulare di cui si vuol silenziare il gene target dell'RNAi
- **Silenziamento a lungo termine del gene target**

Dal transiente alla trasfezione con vettori

Sintesi di siRNA *in vivo*

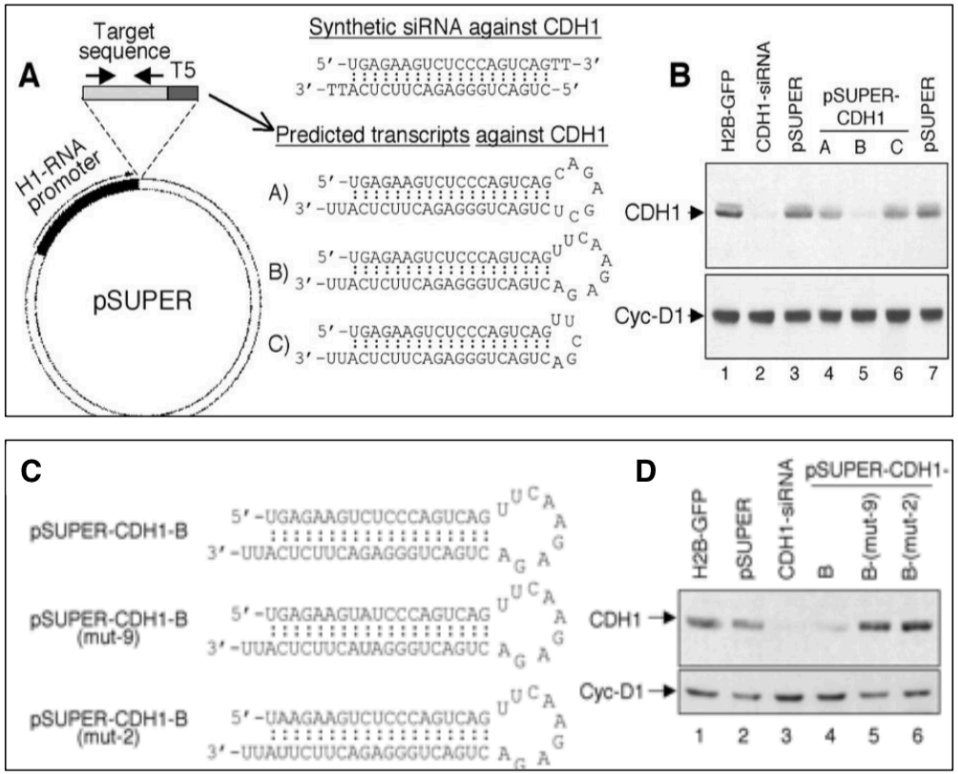
- Nessuna sequenza richiesta dopo start site per la trascrizione
- TTTT: sufficiente per terminazione

More efficiently processed by DICER!!

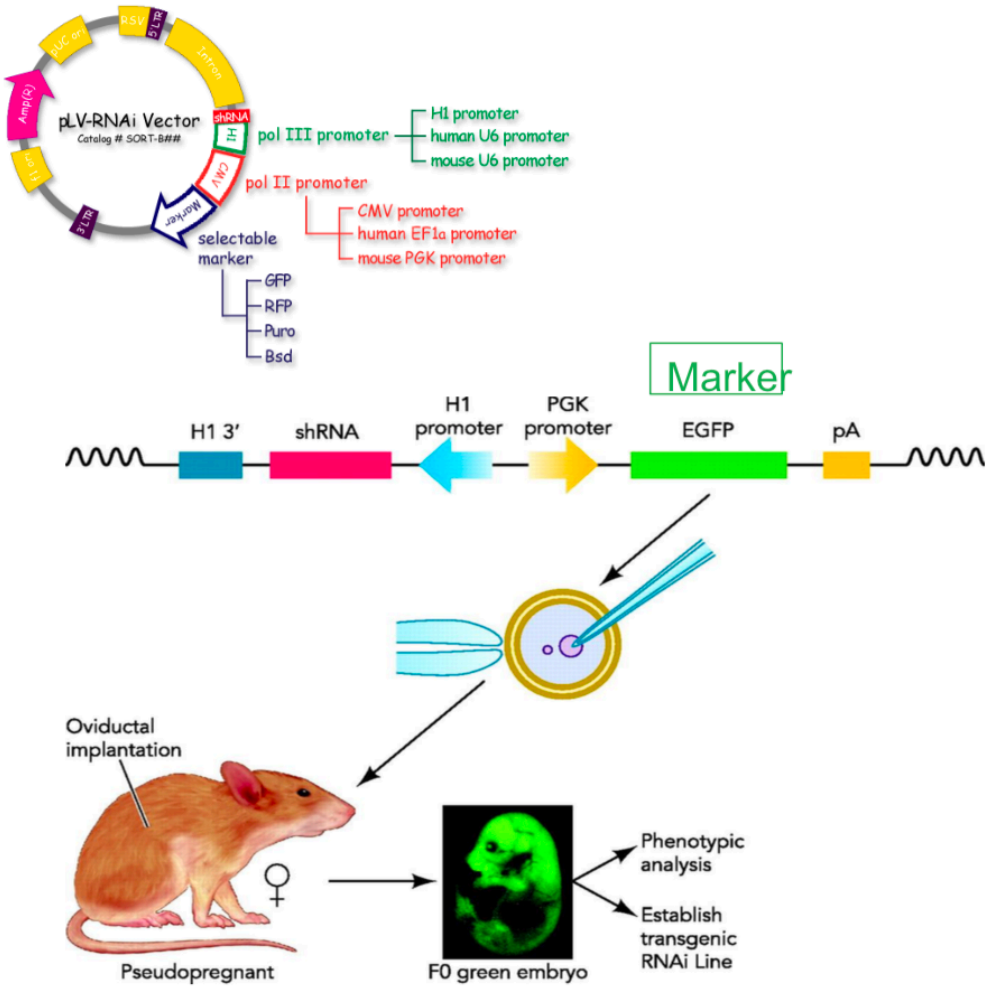
- **Clonati in vettori plasmidici** con promotori adatti per la produzione di RNA
- **Clonati in vettori virali**
  - **Oncoretrovirus**: MoMuLV o MSCV, le cellule devono duplicanti per poter essere infettate
  - **Lentivirus**: HIV-1, per infettare cellule quiescenti

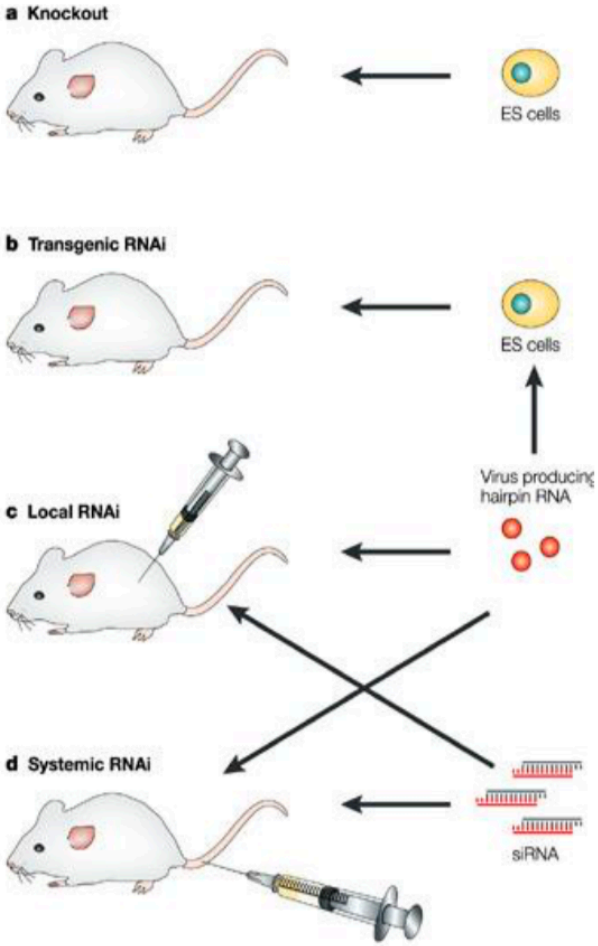
Espressione stabile di shRNA

Un sistema per l'espressione stabile di *short interfering RNA* in cellule di mammifero: vettore plasmidico



Espressione siRNA in vivo





Trasfezione con siRNA: le APPLICAZIONI

- Silenziamento genico specifico, efficiente e stabile nel tempo (economico e veloce)
- È un approccio di «genetica inversa»
- Screening delle funzioni genomiche (Genome-wide functional screenings)
- Terapia genica (es. antitumorale)
- Creazione di modelli per lo studio di agenti farmacologici (es. murini)
- Rivoluzione nello studio dei meccanismi di regolazione dell'espressione genica

siRNA library design

1. Grazie ai siRNA è possibile **silenziare uno alla volta** tutti i geni di un organismo.

2. Una tipica applicazione consiste **nell'identificare quali geni sono coinvolti in un certo processo**

3. Il punto di partenza è una **libreria di siRNA**, **specifica per un singolo gene del genoma**. Oggi esistono librerie in grado di coprire la maggior parte dei geni umani (≈ 20.000 siRNA).

siRNA "potenzialmente" funzionale:

- La regione target deve essere a valle del codone di inizio, ad una distanaza che varia da 50 a 100bp.
- Lunghezza compresa fra 19-22 bp.
- Contenuto in GC fra il 35-55%
- 2-nt 3´ overhangs di residui di uridina
- 5´-phosphate and 3´-hydroxyl group.

- **Stability**
- **Access to RISC**

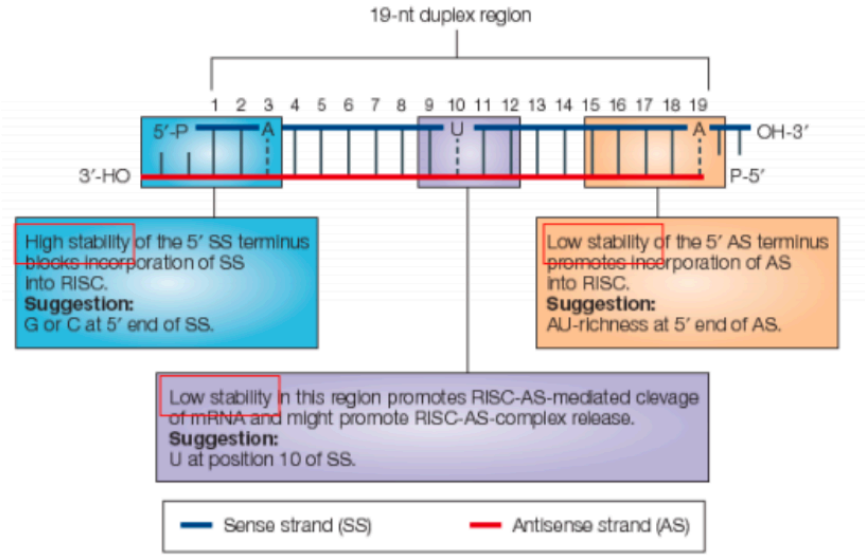
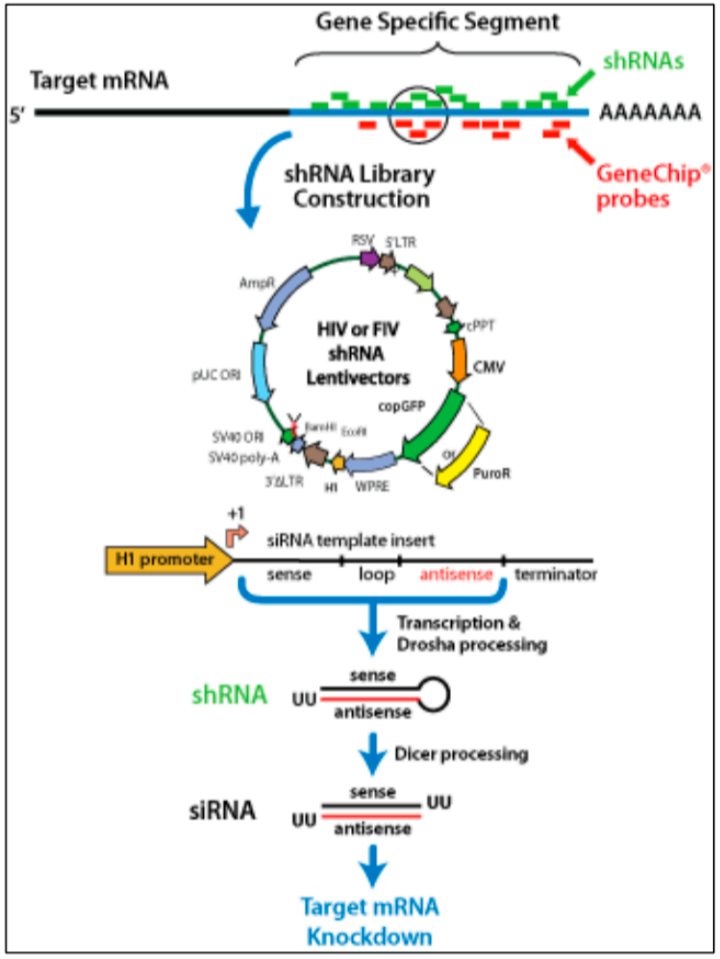


Figure 2 | **The generation of effective siRNA.** A small interfering RNA (siRNA) is a 21–23-nucleotide (nt) dsRNA that contains: a 19-nt duplexed region, symmetric 2–3-nt 3′ overhangs, and 5′-phosphate (P) and 3′-hydroxyl (OH) groups. The positions of each nucleotide in the 19-nt duplexed region of the sense strand are shown. On the basis of recently established design criteria, an effective siRNA has high stability at the 5′ terminus of the sense strand (blue box), lower stability at the 5′ antisense terminus (orange box) and at the cleavage site (purple box). In addition, the sequence-specific preferences at the following positions on the sense strand are important: the presence of an A at position 19, an A at position 3, a U at position 10 (BOX 2 lists other parameters). RISC, RNA-induced silencing complex.

Mittal, Nature Review Gentic, 2004

Rational siRNA design for RNA interference. Nature Biotechnology 22, 326 - 330 (2004)

Screening con siRNA

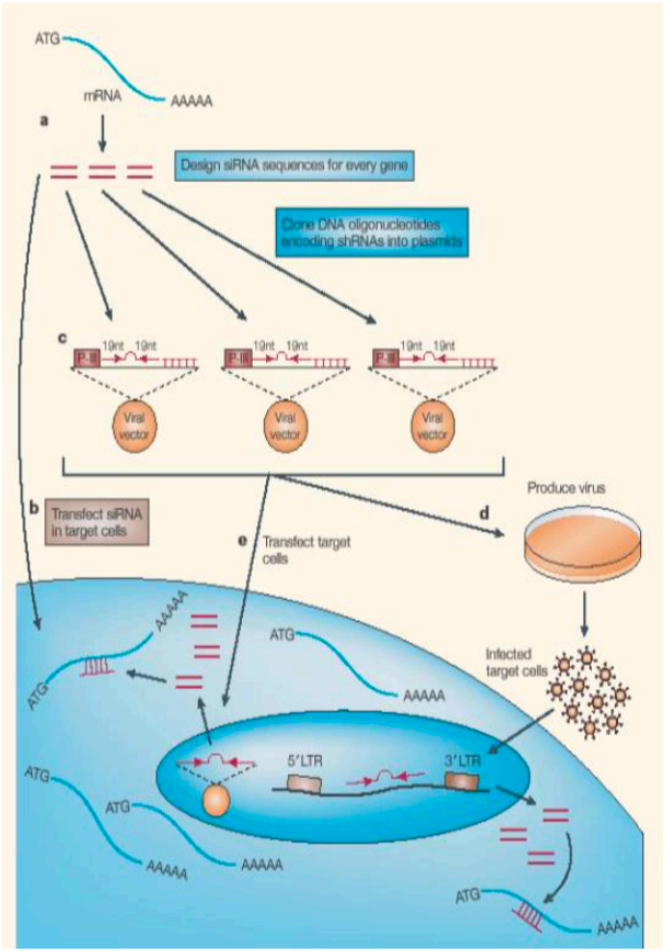


1. **Sintesi della libreria di siRNA, specifica per un singolo gene del genoma**. Oggi esistono librerie in grado di coprire la maggior parte dei geni umani (≈ 20.000 siRNA).

2. Trasfezione delle cellule con una libreria di siRNA diretti contro uno specifico gene target

3. Analisi espressione genica rispetto al controllo non trasfettato (Northern blotting; RT-PCR; gene-expression profiling) o ricerca della proteina analisi con saggi cellulari (FACS; ELISA)

4. **Identificazione del vettore con l'inserto in grado di inibire il gene target**



Limitazioni dei siRNA

Impossibile studiare geni essenziali
per la sopravvivenza cellulare (*housekeeping*) e sviluppo

Sviluppo di nuovi vettori
per l'espressione condizionale-inducibile dei shRNA

(tet OFF/ON H1 and U6 promoter system)

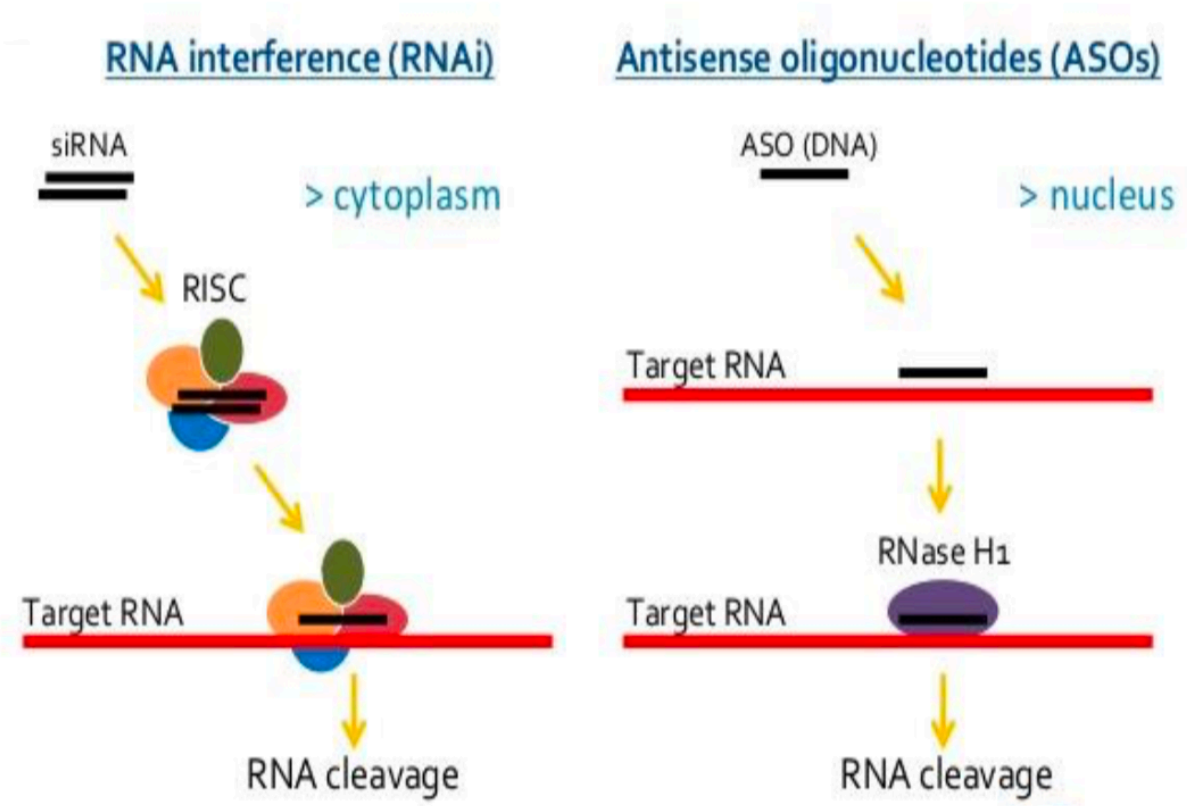
siRNA vs. oligonucleotidi antisense (a ssDNA)

Similarità

- Lunghezza
- Metodologia di *delivery* comune
- Induzione di silenziamento genico a livello post-trascrizionale
- Digestione di mRNA bersaglio da parte di endonucleasi
- Possibilità di stabilizzare con basi modificate
- Bio-distribuzione simile

Differenze

- Doppio filamento *vs.* singolo filamento
- Maggiore stabilità del siRNA
- Maggiore efficacia delle molecole in cellule in coltura
- Meccanismo d'azione mediato da RISC

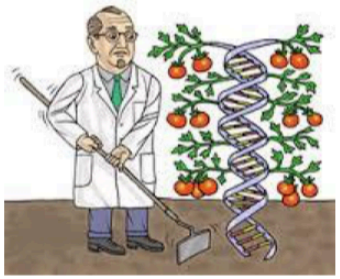


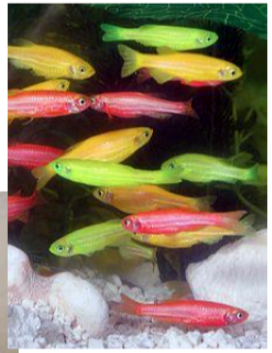
CRISPR/Cas technology

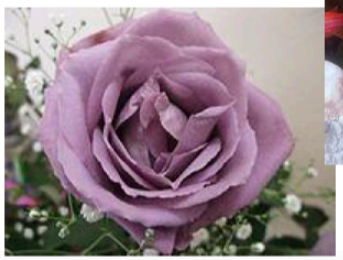
Introduction to CRISPR/Cas - Genetic Engineering

"The deliberate modification of the characteristics of an organism by manipulating its genetic material."

- Research
- Medicine (Protein/Enzyme production)
- Agriculture (Crops)
- Industrial Biotechnology (Biofuel production)
- Entertainment







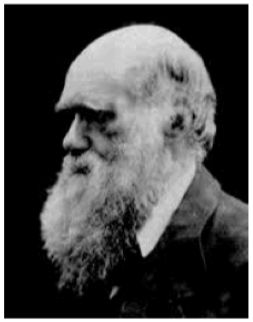
The way towars genetic Engineering

Rules

1859 Darwin
"*Origin of Species*"

1856-66 Mendel
"Mendelian inheritance"

1871 Mieska
Nucleic acids



Information

1944 Avery–MacLeod–McCarty
DNA as the genetic material

1953 Watson, Crick and Franklin
DNA structure

1961-1967 Genetic code

Basics

1970 Restriction Enzymes

1977 Sanger Sequencing

1983 PCR

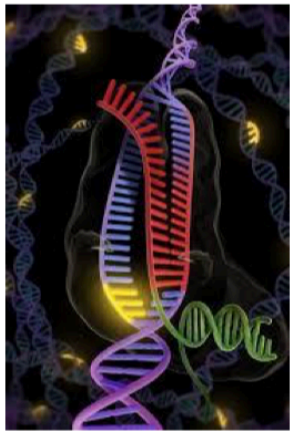
2003 Human Genome Project

Genome editing

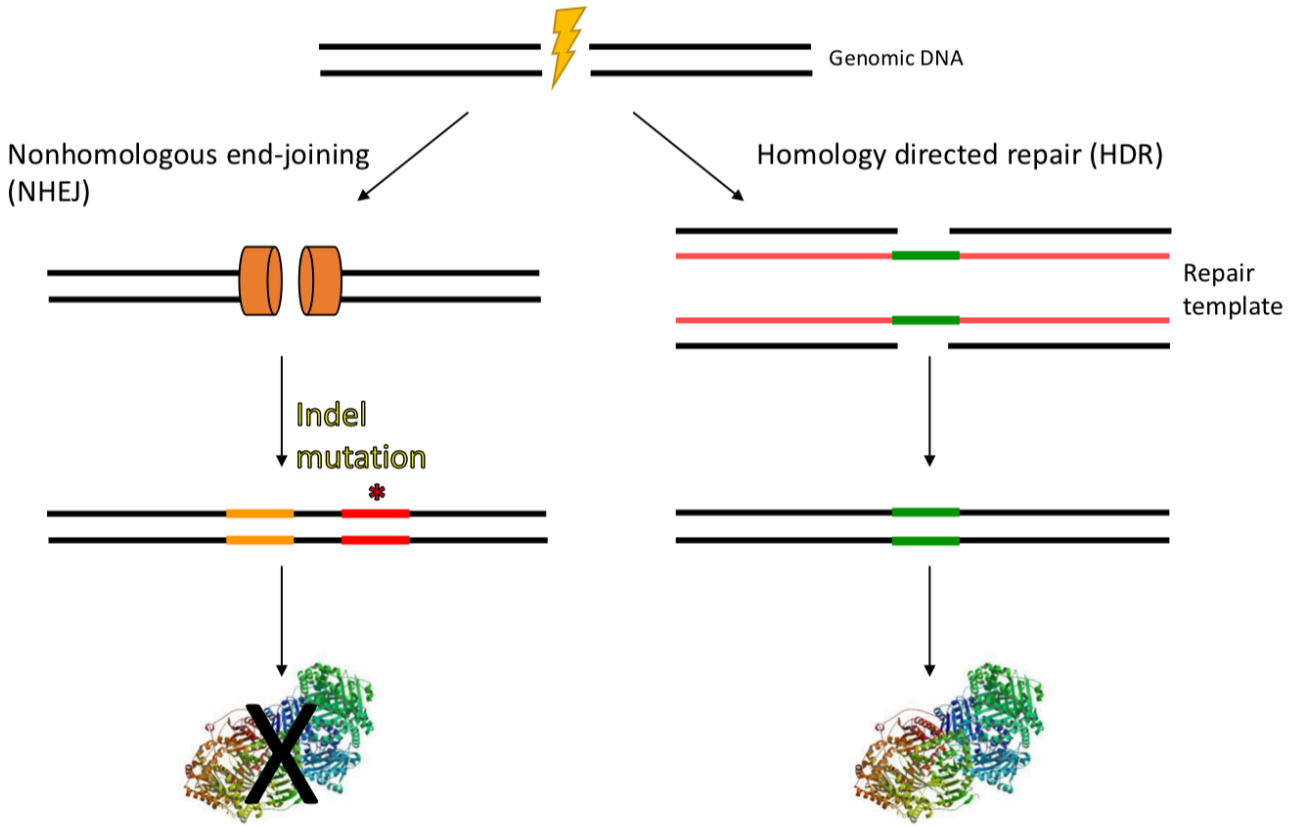
Zink Fingers

TALENs

CRISPRs



Genetic editing uses DNA repair pathways

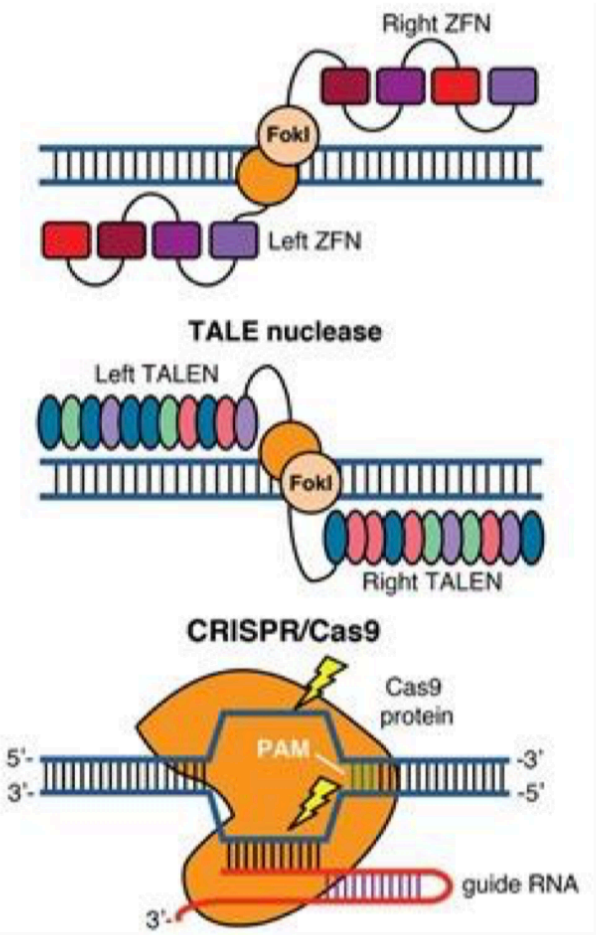


Genome Editing using Site Specific Nucleases

Zinc Fingers

TALEN

CRISPR

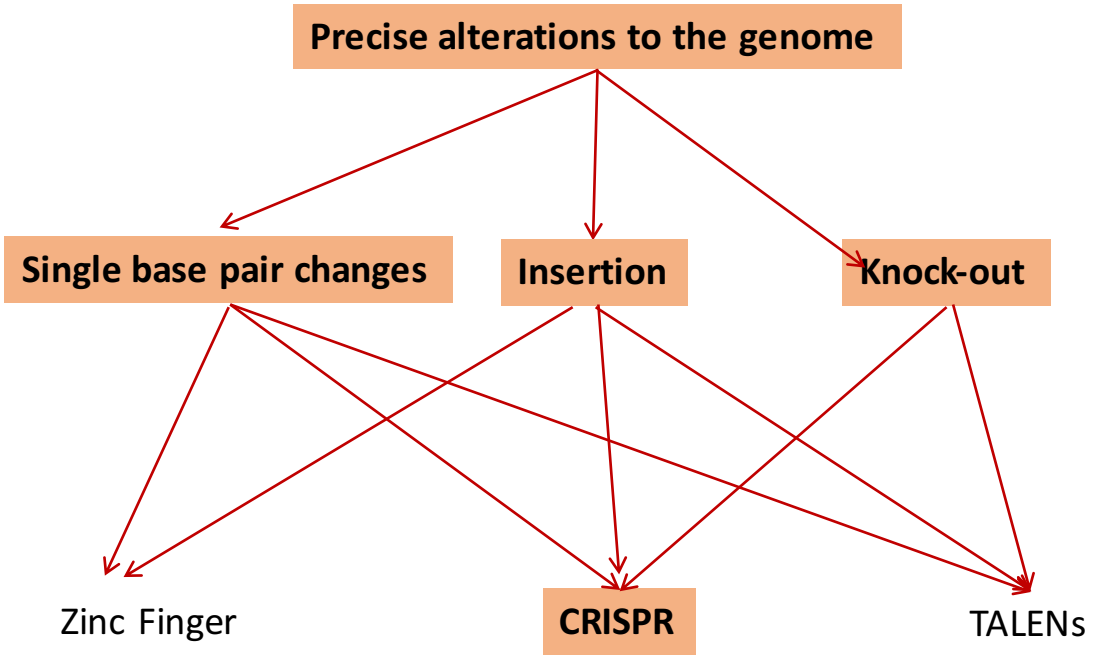


Protein based

RNA based

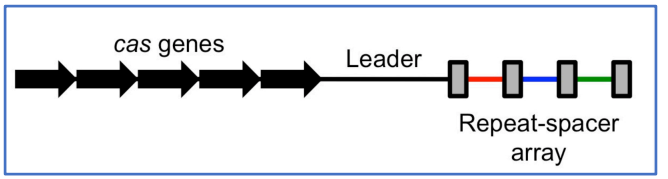
Adapted from Duke University

Genome Editing



CRISPR-Cas – Adaptive immune system in bacteria

A: CRISPR (**clustered regularly interspaced short palindromic repeats**) is a family of DNA sequences found within the genomes of prokaryotic organisms such as bacteria and archaea. These sequences are derived from DNA fragments from viruses that have previously infected the prokaryote and are used to detect and destroy DNA from similar viruses during subsequent infections. Hence these sequences play a key role in the antiviral defense system of prokaryotes.



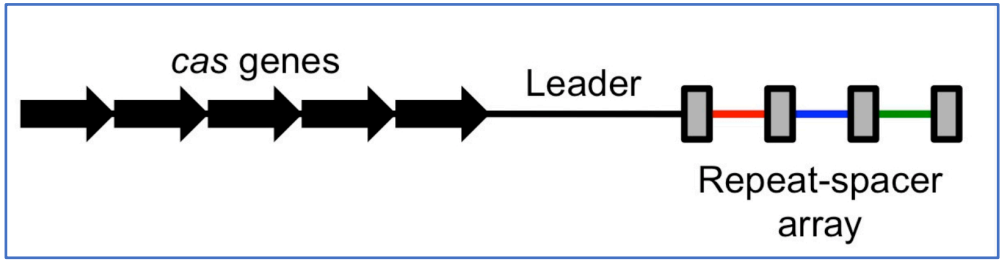
+

B. Cas9 (or "CRISPR-associated 9") is an enzyme that uses CRISPR sequences as a guide to recognize and cleave specific strands of DNA that are complementary to the CRISPR sequence.

Cas9 enzymes together with CRISPR sequences form the basis of a technology known as **CRISPR/Cas9** that can be used to edit genes within organisms

CRISPR-Cas – A. CRISPR locus

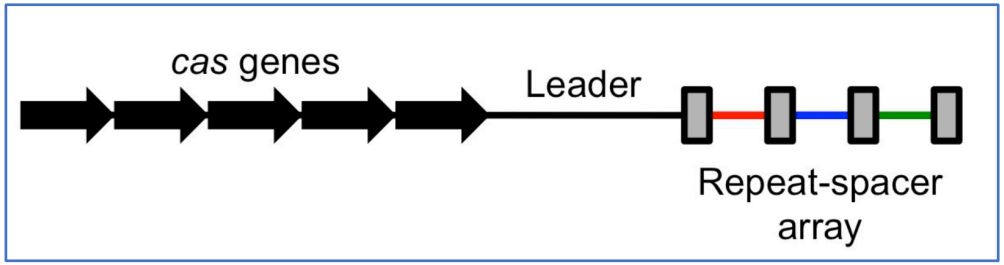
Simplified diagram of a CRISPR locus



The **three major components** of a CRISPR locus are shown: 1. ***cas* genes**, 2. **leader** and 3. **repeat-spacer array**. For the repeat-spacer array, repeats are shown as grey boxesc (typically range in size from 28 to 37 base pairs (bps), though there can be as few as 23 bp and as many as 55 bp), and spacers are colored bars

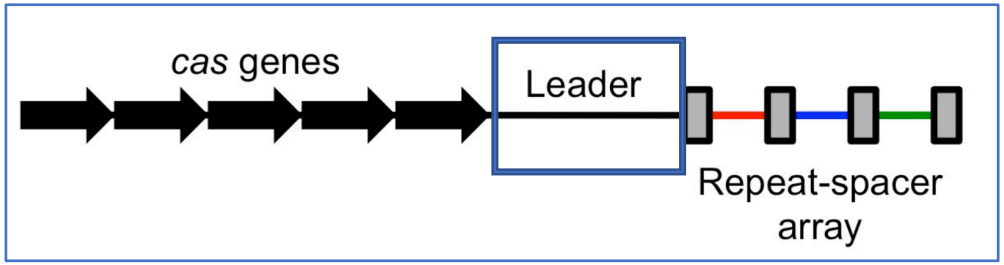
CRISPR-Cas – A. CRISPR locus

Simplified diagram of a CRISPR locus



CRISPR-Cas – A. CRISPR locus

Simplified diagram of a CRISPR locus

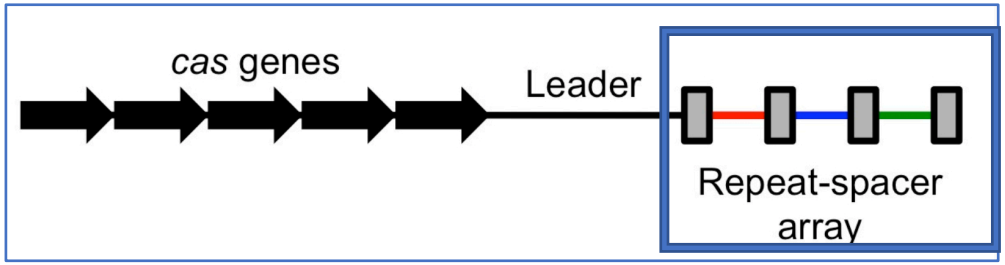


Leader Sequence

This sequence is an A-T reach sequence

CRISPR-Cas – A. CRISPR locus

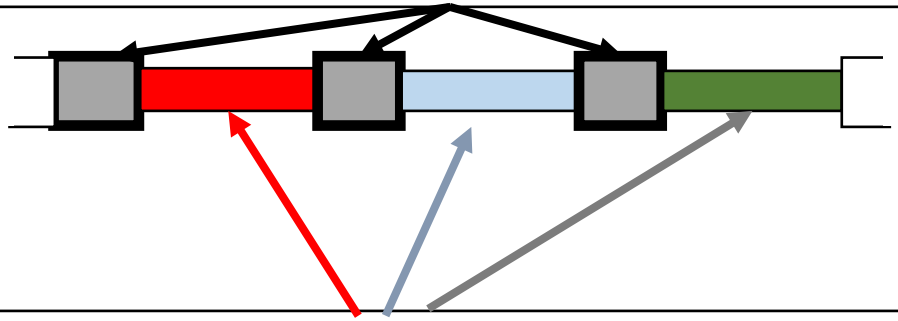
Simplified diagram of a CRISPR locus



REPEAT Spacer ARRAY

Shorts palindromic **repeats** (i.e. this is the same DNA sequence repeated in different places, gray boxes). These are part of the **bacterial genome** and typically range in size from 28 to 37 base pairs

Diagram of CRISPR locus in bacterial genome

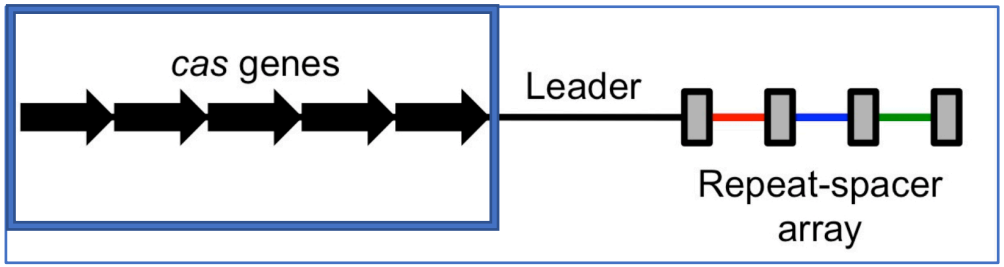


There are usually fewer than 50 units of the repeat-spacer sequence in a CRISPR array.

These bits/**spacers** are derived from **bacteriophage genome** (range 21 to 72 bp) and each one is different and these provide the guidance system for the adaptive immune system

CRISPR-Cas – A. CRISPR locus

Simplified diagram of a CRISPR locus



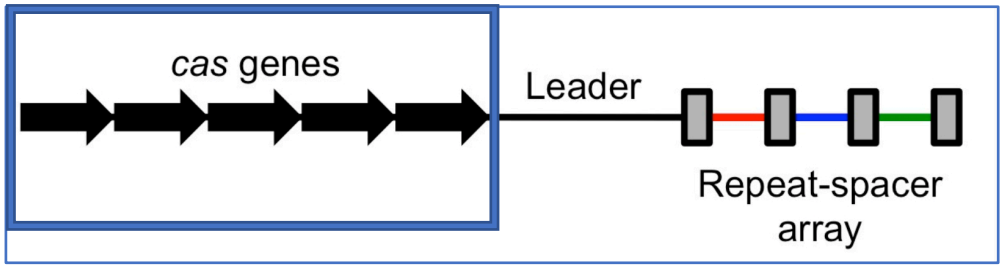
CAS genes

There are several other important regions of the bacterial DNA that are also always associated with the CRISPR locus and these provide the means for the palindromic repeat and the bacteriophage DNA sequences to actually destroy the bacteriophage.

These are called CRISPR Associated Sequences i.e. **Cas** genes

CRISPR-Cas – A. CRISPR locus

Simplified diagram of a CRISPR locus



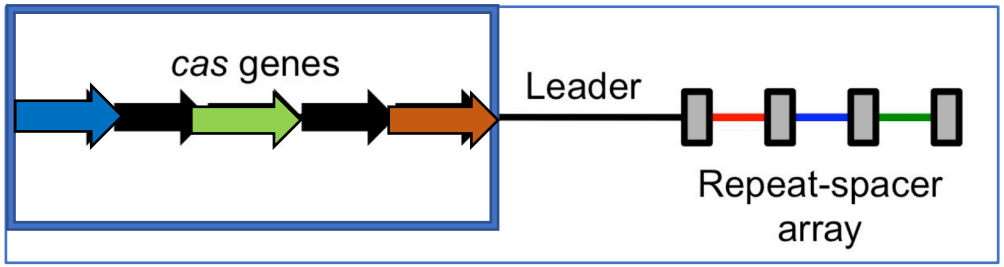
CAS genes

There are several other important regions of the bacterial DNA that are also always associated with the CRISPR locus and these provide the means for the palindromic repeat and the bacteriophage DNA sequences to actually destroy the bacteriophage.

These are called CRISPR Associated Sequences i.e. **Cas** genes

How does this genetic material in CRISPR locus then manage to kill bacteria ?

Simplified diagram of a CRISPR locus



CAS genes

The system can be slighty different in different types of bacteria but the best studies one is ***Streptococcus pyogenes*** so we will focus on that one

For the sake of simplicity lets focus on the 3 Cas genes (now colored arrows) most important for genetic engineering;

Codes for a **trans-activating CRISPR RNA (tracrRNA**) that will help in the process of ensuring the whole process only cuts bacteriophage DNA

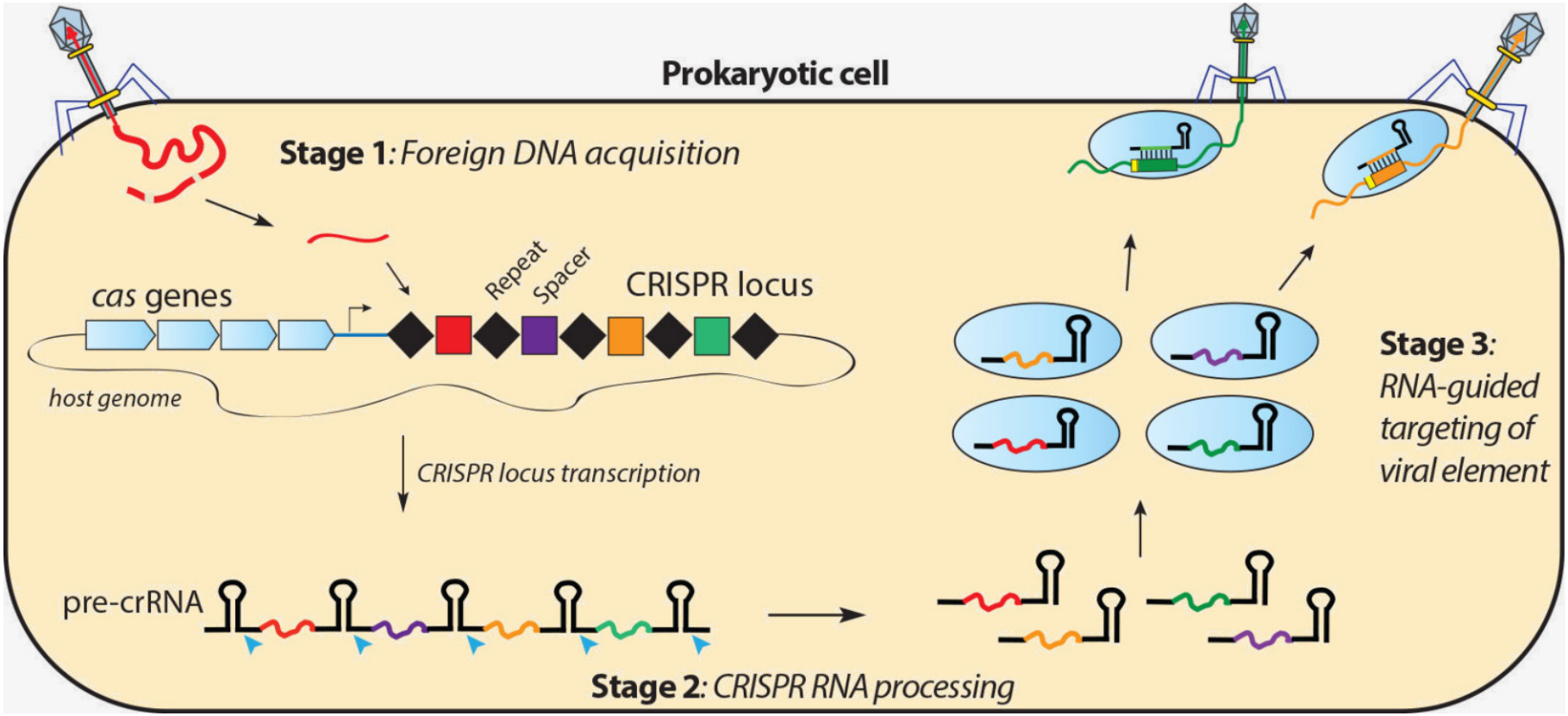
Codes **for a protein** that is a nuclease that cuts DNA but only if it is given a very specific set of signals to do so (otherwise it would potentially damage the bacteria's own DNA). The most common one used in genetic engineering approaches is called Cas9. ; additional Cas1 and Cas2 are responsible for spacer geration

Codes for a very specific piece of RNA (**crRNA or guide RNA**) that will help in the process of ensuring the whole process only cuts bacteriophage DNA

For now lets not worry about the other genes in the Cas locus

Acquisition of immunity

Adaptive immunity

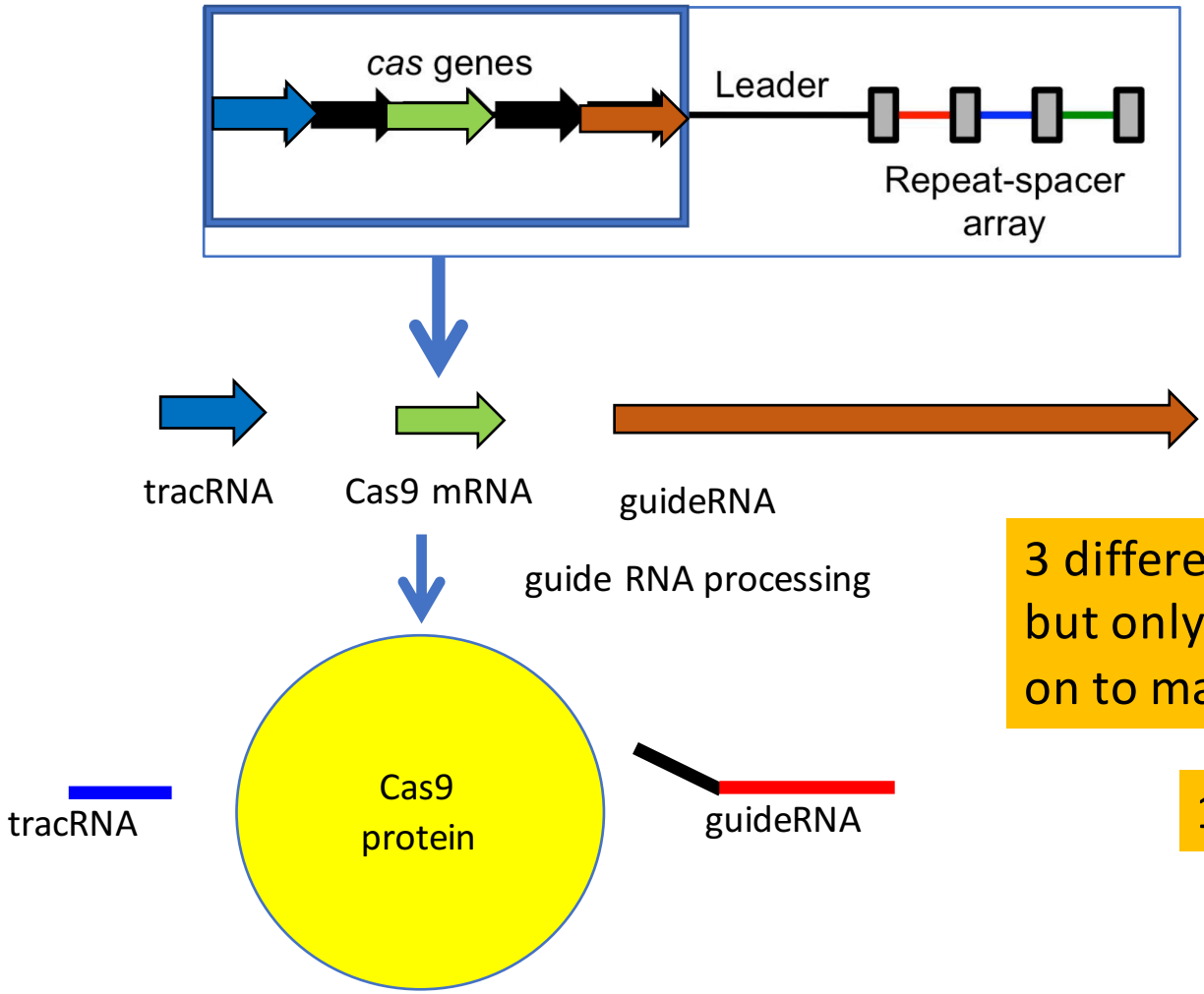


ACQUISITION OF IMMUNITY: Cas1 and Cas2 are responsible for spacer generation: Bioinformatic analysis of regions of phage genomes that were excised as spacers (termed protospacers) revealed that they were not randomly selected but instead were found adjacent to short (3 – 5 bp) DNA sequences termed protospacer adjacent motifs (PAM).

Foreign DNA is inserted into CRISPR locus. IMPORTANT: immunity is passed on to the next generation of cells

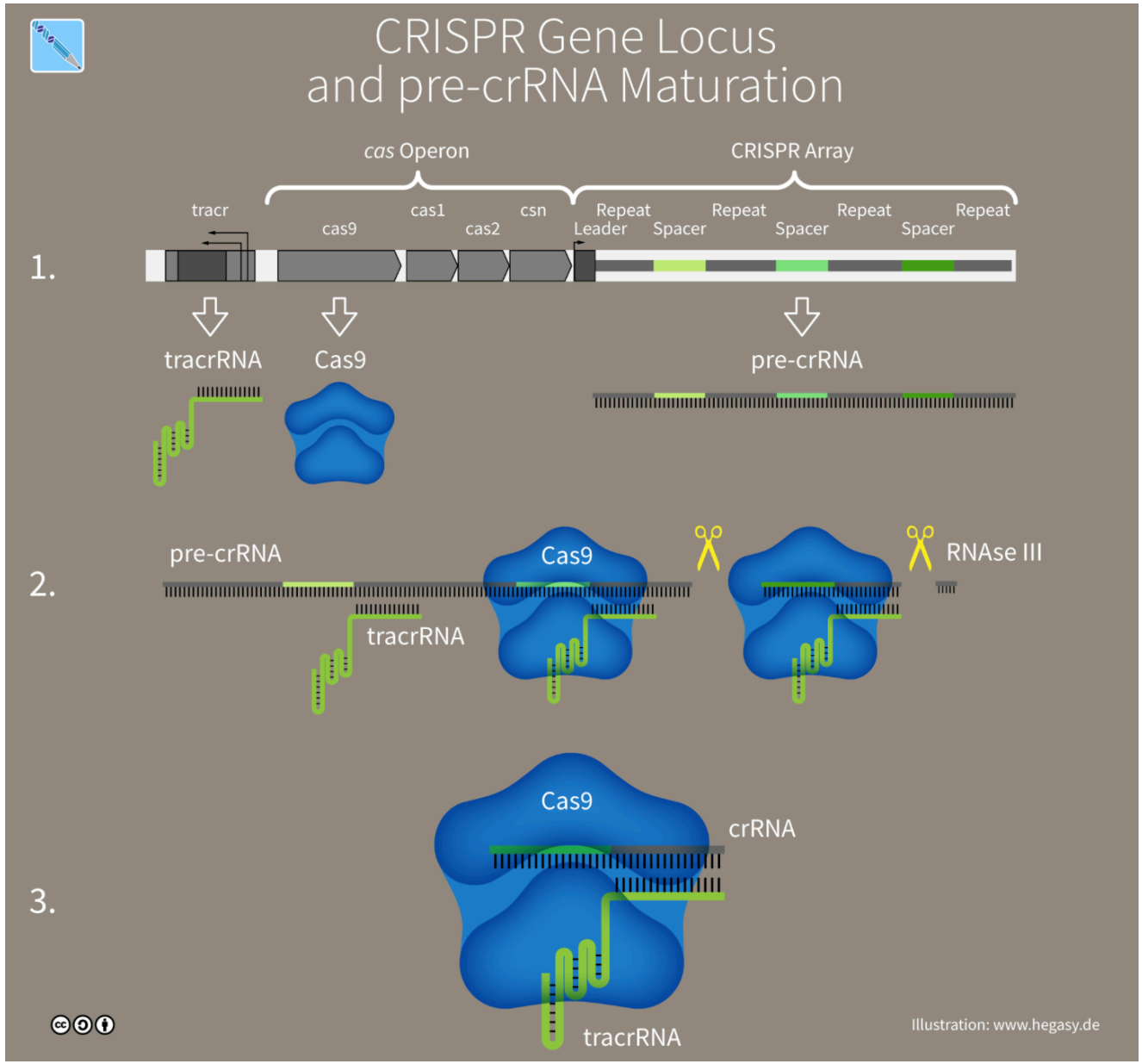
ADAPTIVE IMMUNITY: "daughter cells" are already immune to a bacteriophage that had infected the "mother/father" cell → inherited information

How does this genetic material in CRISPR locus then manage to kill bacteria ?



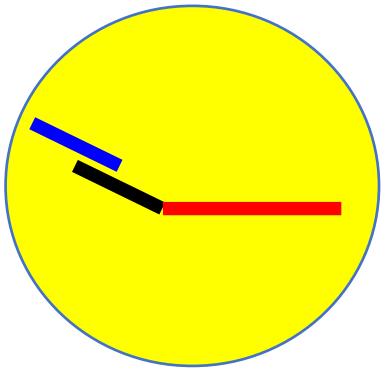
3 different RNAs generated but only one of these goes on to make a protein.

1 protein generated


CRISPR Gene Locus
and pre-crRNA Maturation
cas Operon
CRISPR Array
tracr
cas9
cas1
cas2
csn
Repeat
Leader
Spacer
Repeat
Spacer
Repeat
Spacer
Repeat
1.
tracrRNA
Cas9
pre-crRNA
2.
pre-crRNA
Cas9
tracrRNA
RNAse III
3.
Cas9
crRNA
tracrRNA
Illustration: www.hegasy.de

What is Cas9?

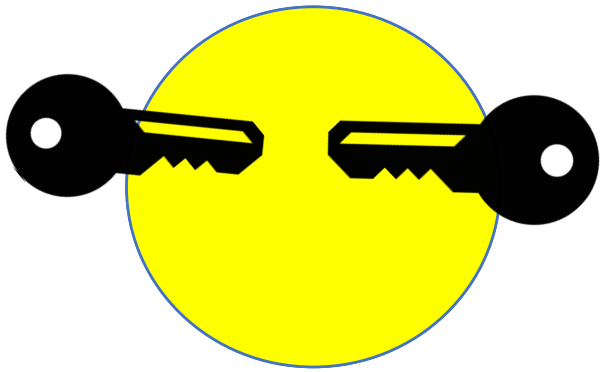
- Cas9 is an endonuclease that can cut double stranded DNA
- Cas 9 is only activated when the tracRNA and the guide RNA are associated with it (i.e it is a nucleoprotein). Imagine this a bit like the fail safe mechanism they use to prevent accidental launch of nuclear missiles where 2 people have to insert keys at exactly the same times
- In fact the tracRNA and the guide RNA have a short overlapping sequence that means they actually have to bind to each other in this complex for this to work properly



Active Cas9

What is Cas9?

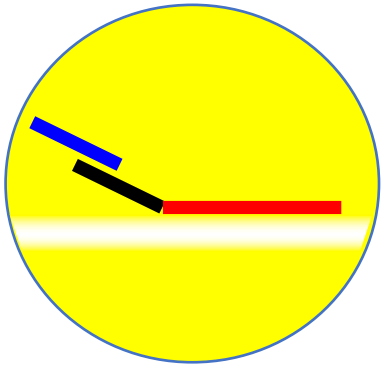
- Cas9 is an endonuclease that can cut double stranded DNA
- Cas 9 is only activated when the tracRNA and the guide RNA are associated with it (i.e it is a nucleoprotein). Imagine this a bit like the fail safe mechanism they use to prevent accidental launch of nuclear missiles where 2 people have to insert keys at exactly the same times
- In fact the tracRNA and the guide RNA have a short overlapping sequence that means they actually have to bind to each other in this complex for this to work properly



Active Cas9

How does Cas9 work?

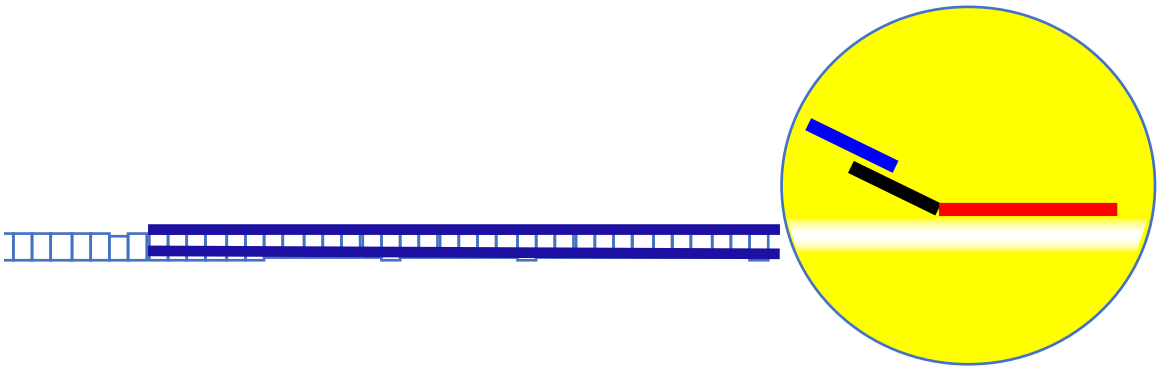
- Cas9 has a channel that DNA can fit into.
- It scans the DNA looking for sequence that match the guide sequence



Active Cas9

How Cas9 works?

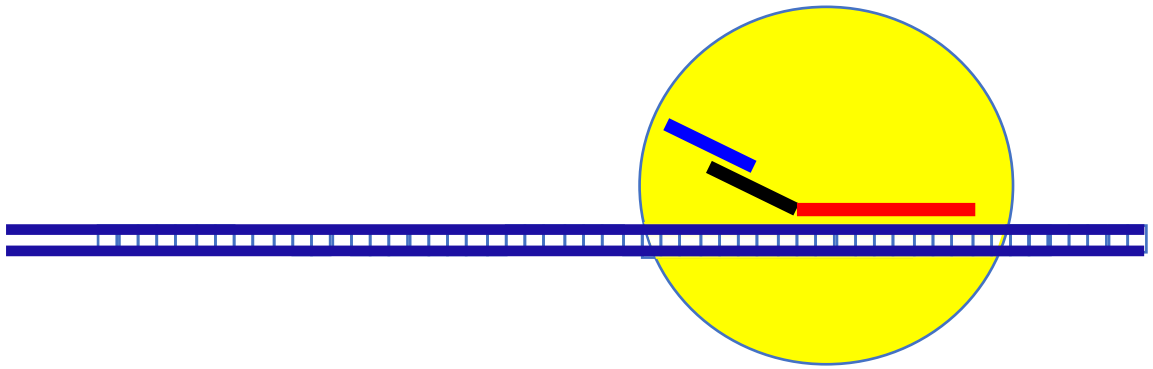
- Cas9 has a channel that DNA can fit into.
- It scans the DNA looking for sequence that match the guide sequence



Active Cas9

How does Cas9 work?

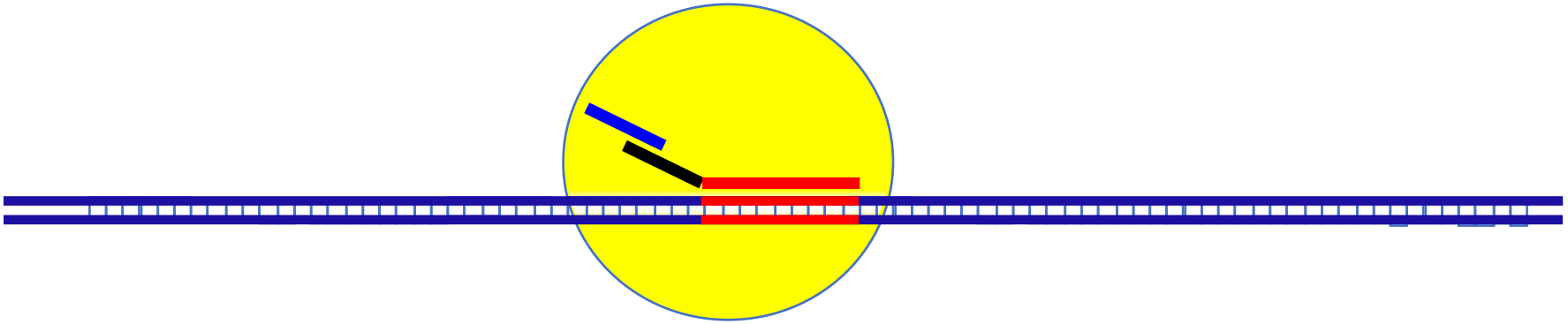
- Cas9 has a channel that DNA can fit into.
- It scans the DNA looking for sequence that match the guide sequence



Active Cas9

How does Cas9 work?

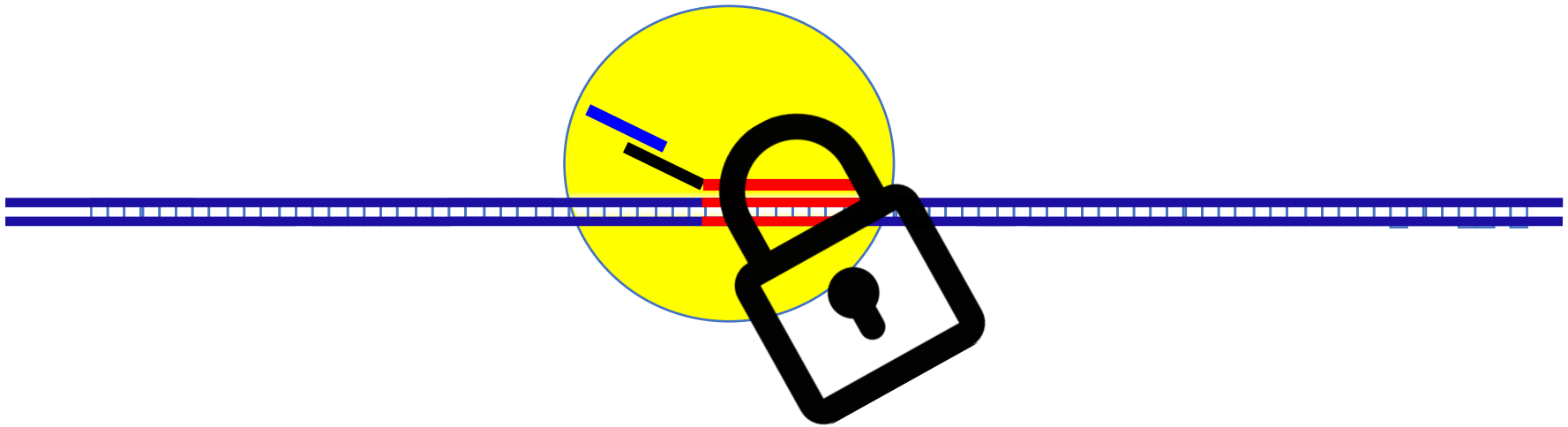
- When a DNA sequence complementary to the guide RNA is found the scanning stops



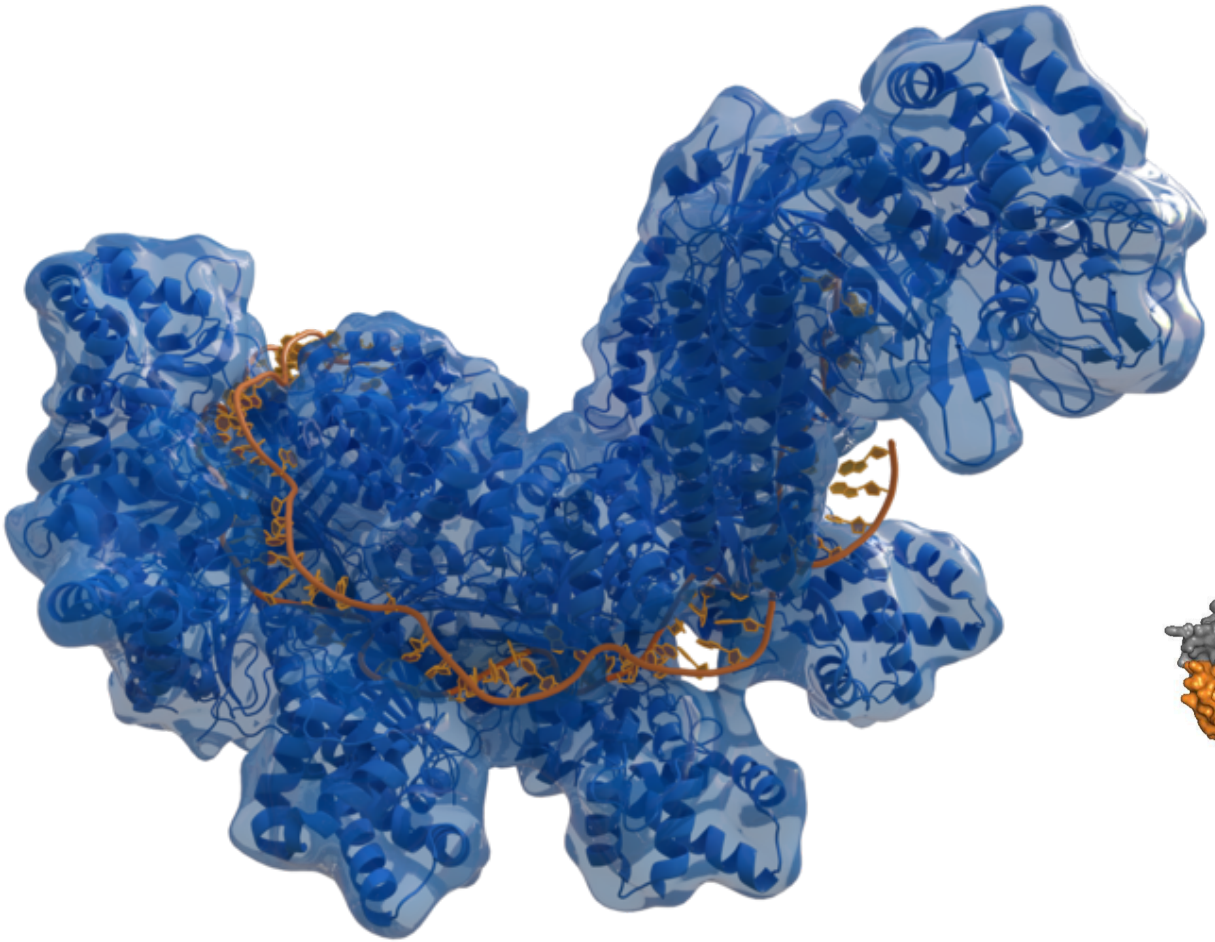
Active Cas9

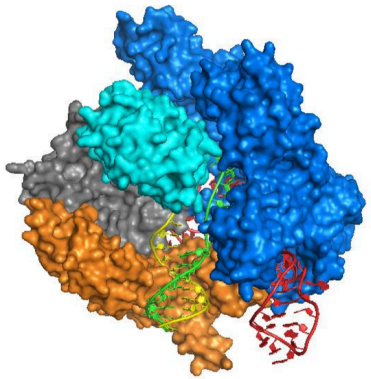
How does Cas9 work?

- When a DNA sequence complementary to the guide RNA is found the scanning stops



Structure of DNA bound to a Cas enzyme





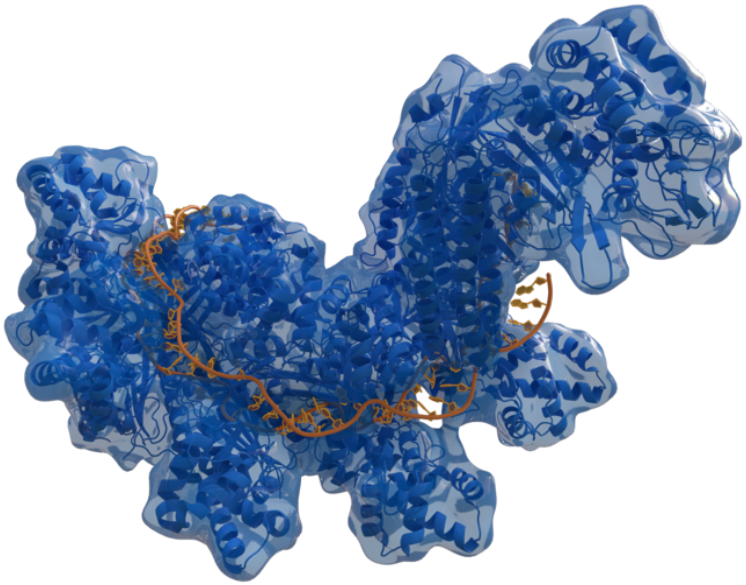
Cas9
(S. pyogenes)

- **Adaptation**
  Recognition of target site
- **Two nuclease domains**
  RuvC (gray) - cleaves non-target DNA strand
  HNH (cyan) - cleaves-target strand of DNA
- **PAM-interacting domain**
  (orange)

Jinek et al. *Science*, 2014

Structure of DNA bound to a Cas enzyme

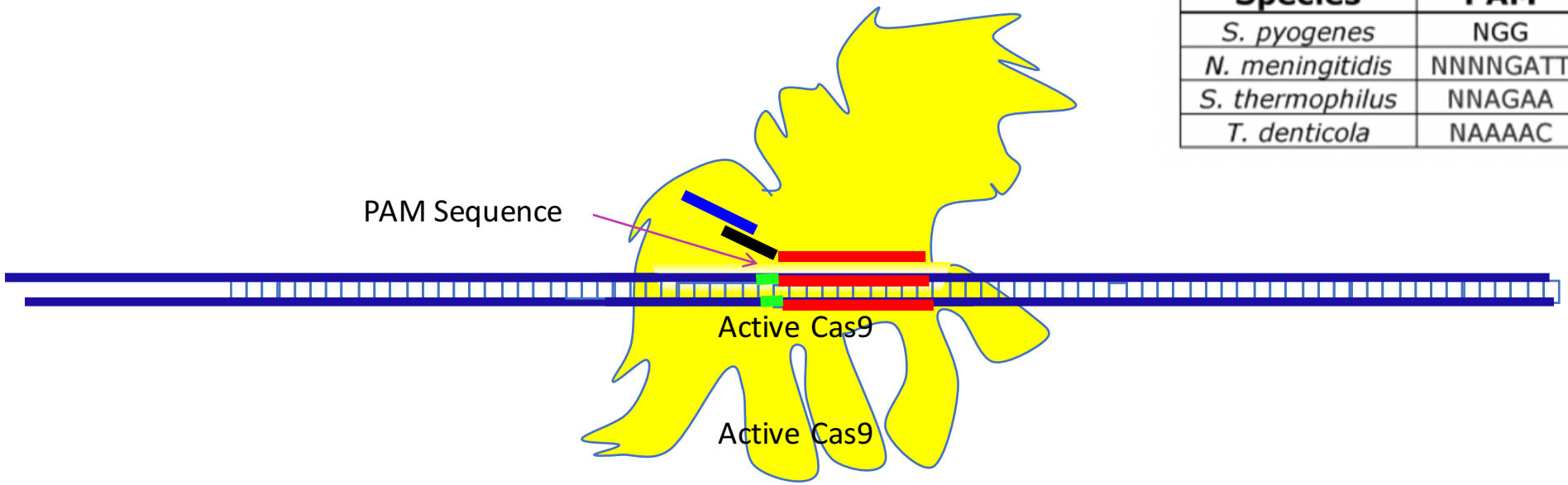
Completely irrelevant aside



How does Cas9 work?

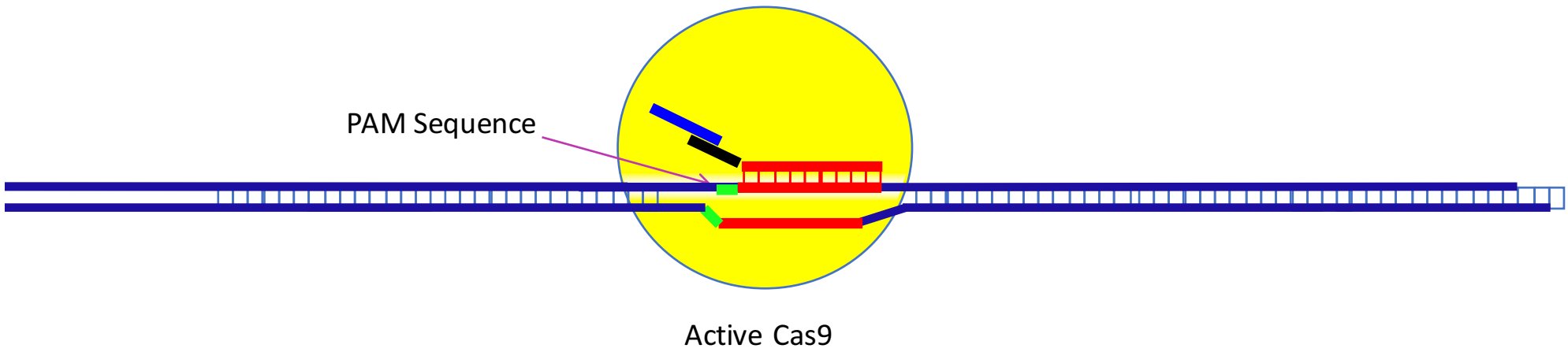
- There is one additional check
- In this control step the target site in the bacteriophage DNA needs to have the PAM sequence (**P**rotospacer **A**djacent **M**otif)
- PAM seqeunces DO NOT APPEAR in the bacterial genome
- PAM seqeunces are requried for Cas9 endonucease activity
- PAM sequences are specific for bacterial strains and protect the Cas locus from beeing cut by Cas9

Species	PAM
S. pyogenes	NGG
N. meningitidis	NNNNGATT
S. thermophilus	NNAGAA
T. denticola	NAAAAC



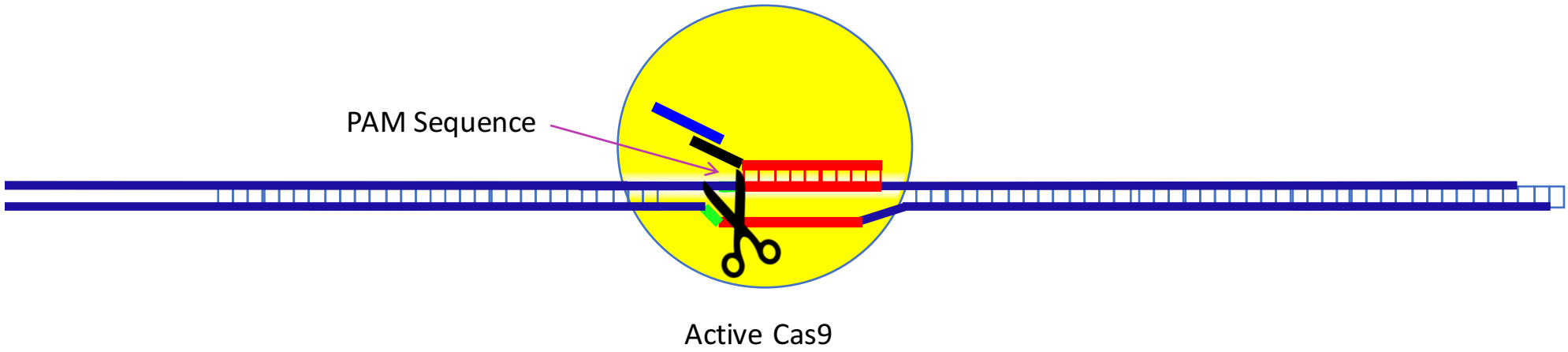
How does Cas9 work?

- Now the RNA binds to the complementary strand of the DNA and opens up the DNA helix



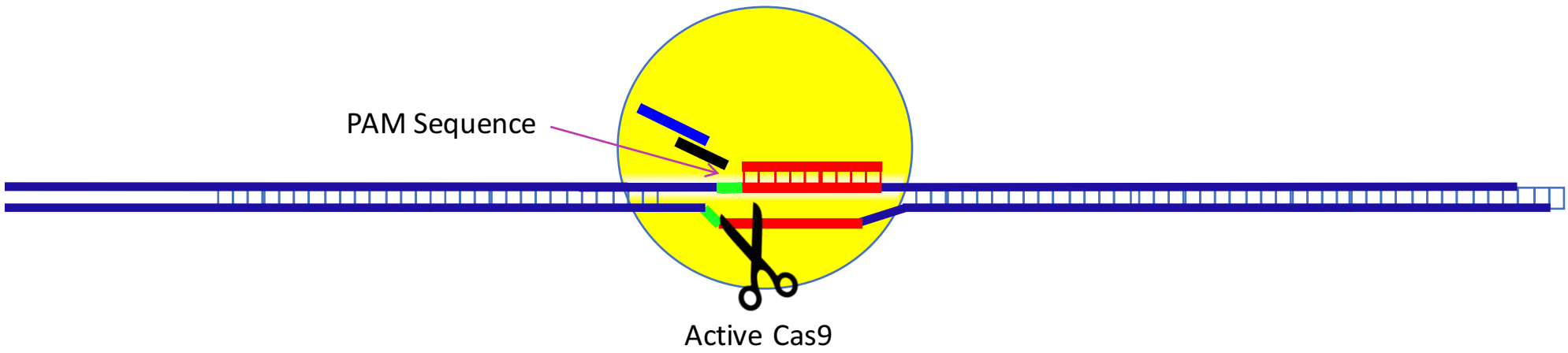
How does Cas9 work?

- Now the bacteriophages DNA gets cut very close to the PAM site



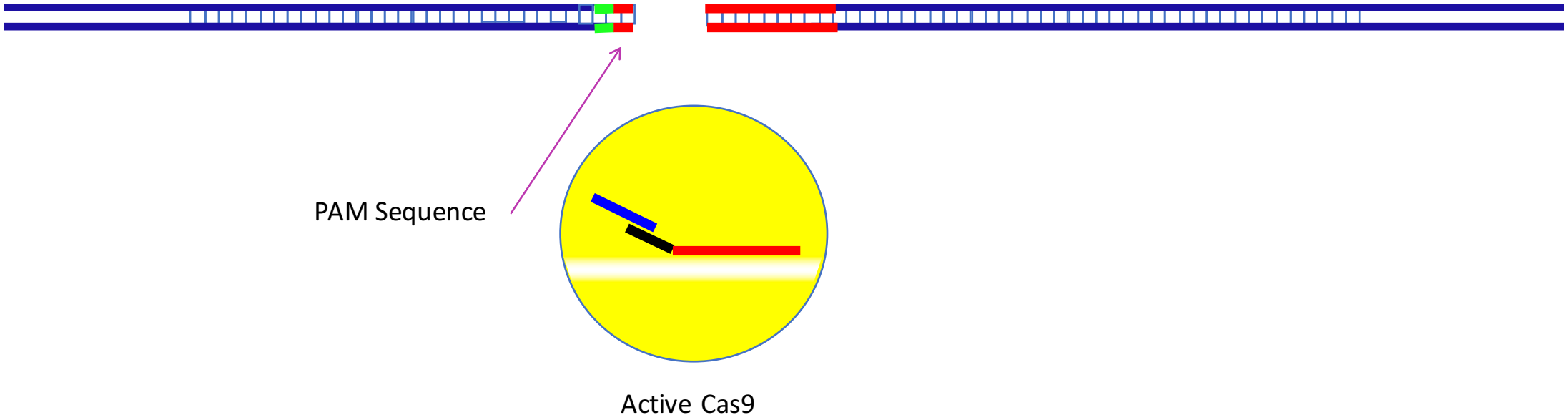
How does Cas9 work?

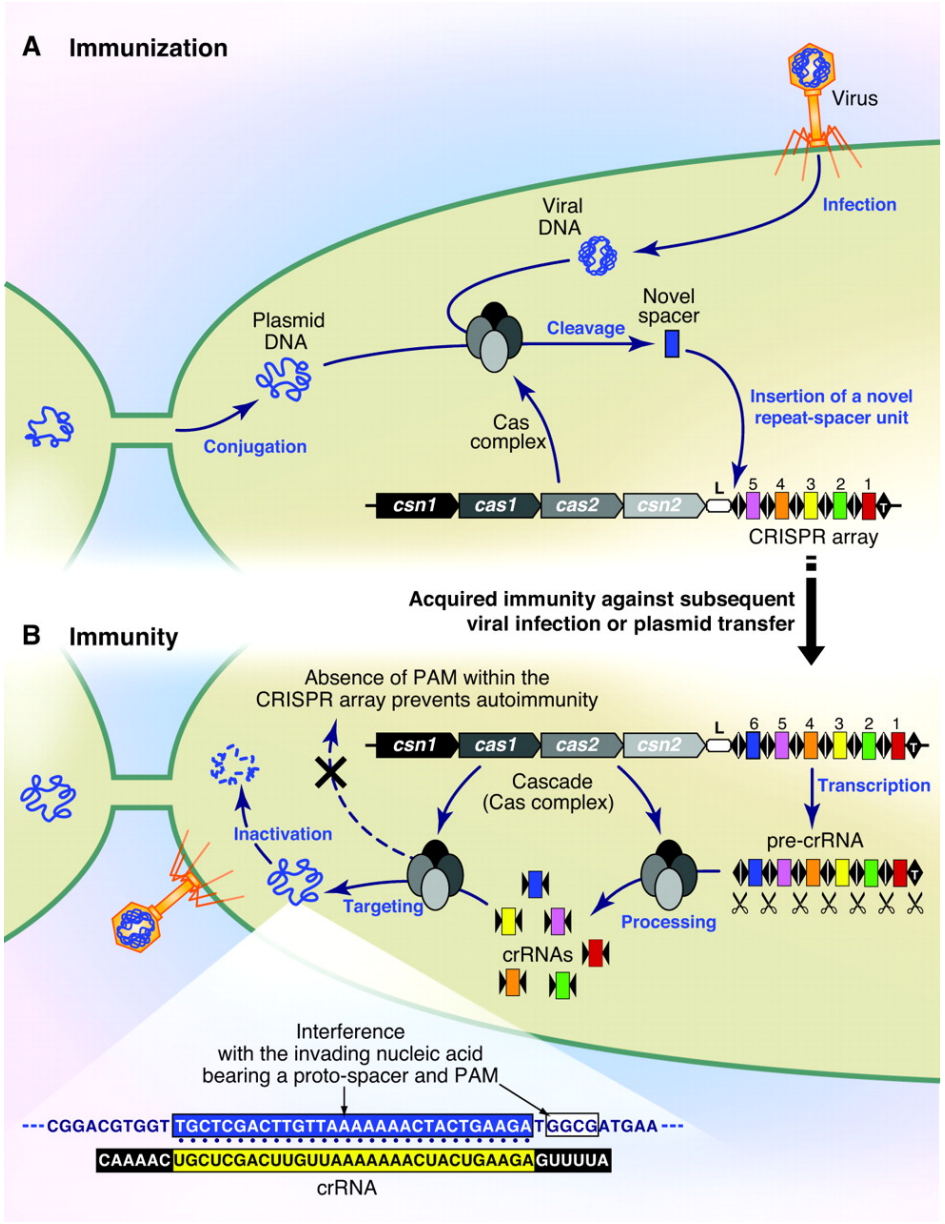
- Now the bacteriophages DNA gets cut very close to the PAM site



How does Cas9 work?

- Now the bacteriophages DNA gets cut very close to the PAM site, it looks like this and the bacteriophage is essentially inactivated





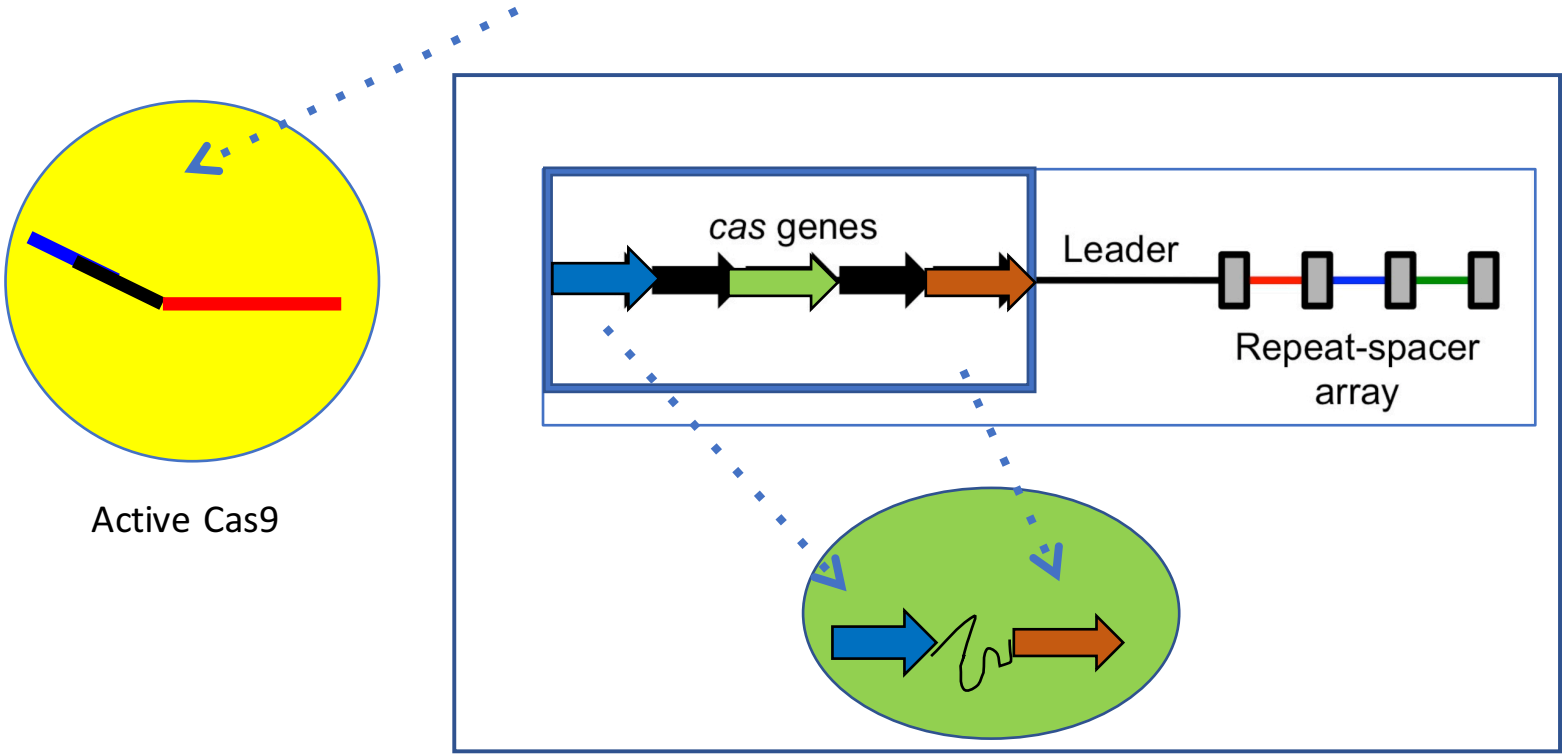
Protospacer adjacent motif (PAM) is a 2-6 base pair DNA sequence immediately following the DNA sequence targeted by the Cas9 nuclease in the CRISPR bacterial adaptive immune system.

PAM is a component of the invading virus or plasmid, but is not a component of the bacterial CRISPR locus. Cas9 will not successfully bind to or cleave the target DNA sequence if it is not followed by the PAM sequence.

PAM is an essential targeting component (not found in bacterial genome) which distinguishes bacterial self from non-self DNA, thereby preventing the CRISPR locus from being targeted and destroyed by nuclease.

How can we use CRISPR/Cas9 for genetic engineering?

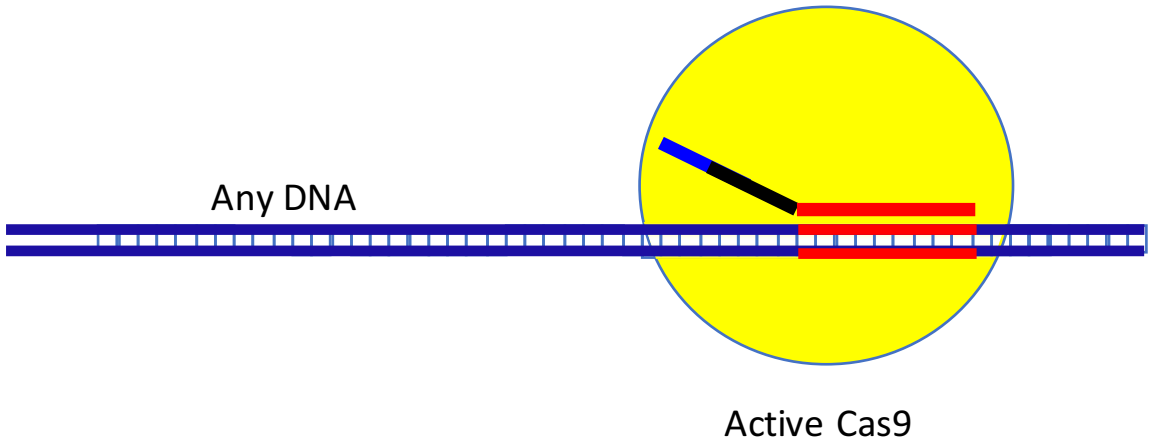
- Jennifer Doudna and Emmanuelle Charpentier re-engineered the Cas9 endonuclease into a more manageable two-component system by fusing the two RNA molecules tracRMA and guide RNA (or crRNA) into a "**SINGLE-GUIDE RNA**" (**sgRNA)** that, when combined with Cas9, could find and cut the DNA target specified by the guide RNA



Active Cas9

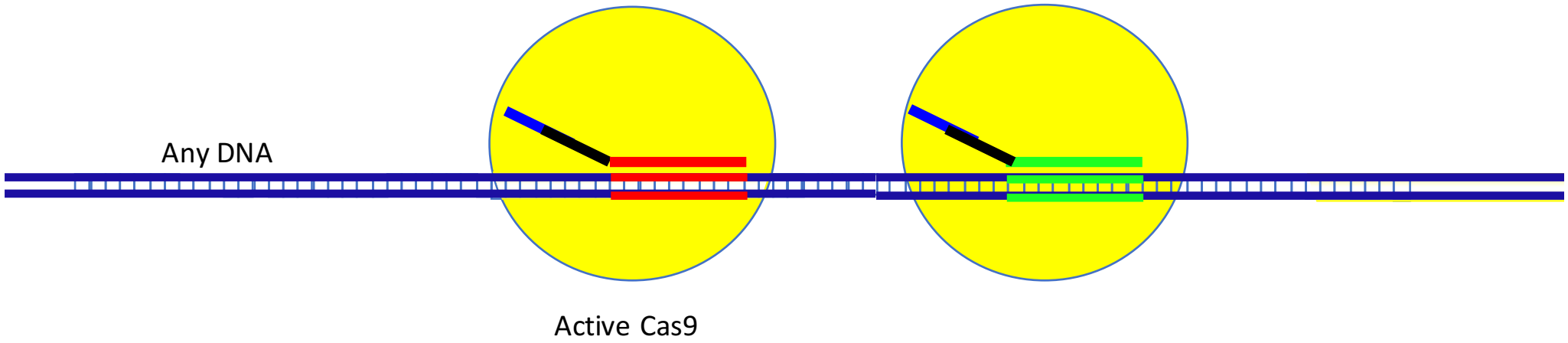
How can we use CRISPR/Cas9 for genetic engineering?

- This means we can artificially make a sgRNA that can be designed to target **any part of the genome** (as long as it has an appropriate PAM sequence nearby)
- **All we have to do is artificially express the Cas9 and the sgRNA together**



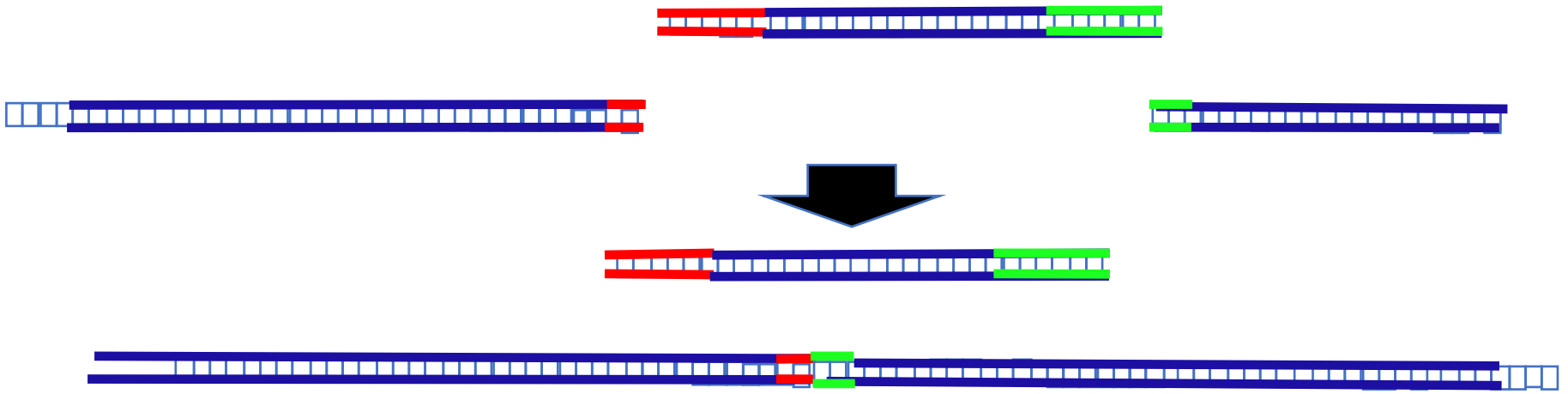
How can we use CRISPR/Cas9 for genetic engineering?

- We can put two different sgRNA into the same protein and cut at 2 places in the genome . → we can cut out large regions of DNA.



How can we use CRISPR/Cas9 for genetic engineering?

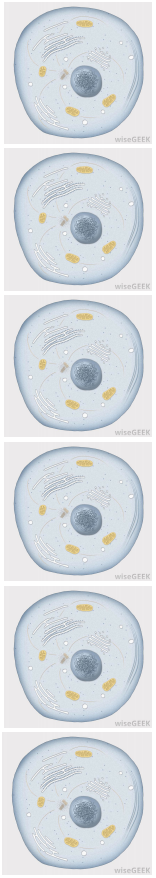
This allows us to selectively “knock out” regions of the genome



Just an example: Knockout of VEGF-a gene

1

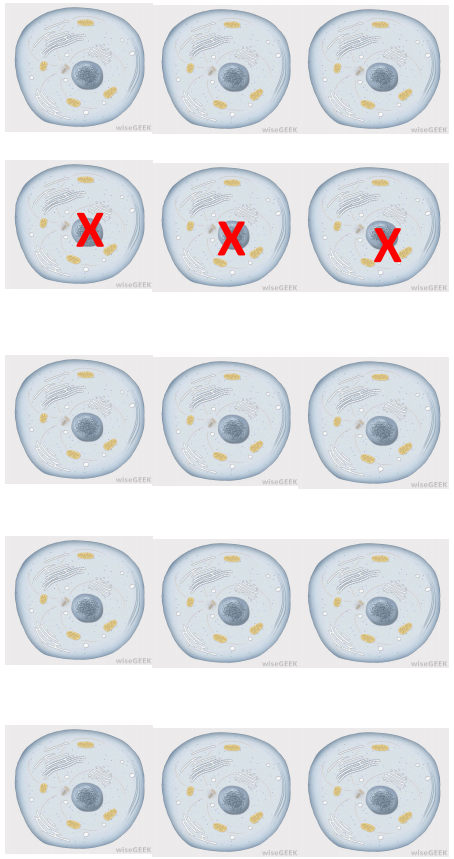
Take lots of cells and add the Cas9 protein plus 2 sgRNA that specifically bind to VEGFA gene



2

Isolate single cells (i.e select clones)

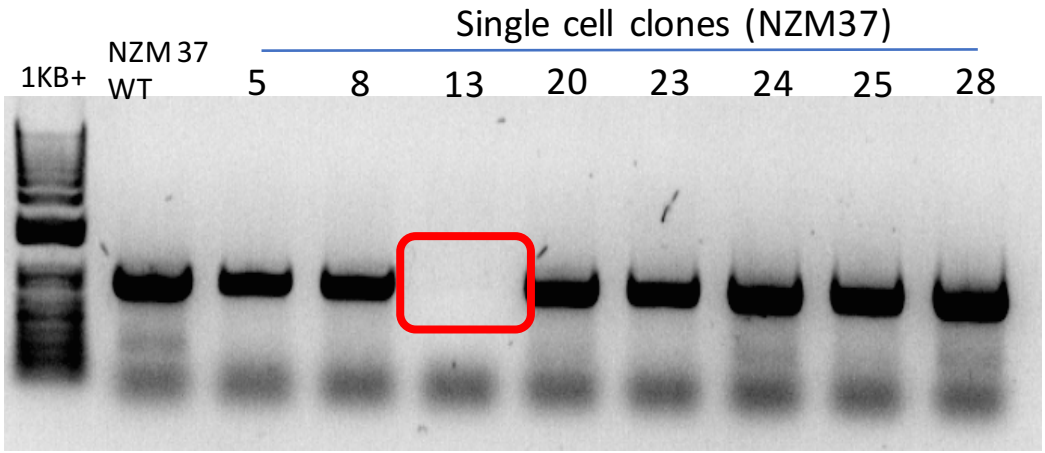
3 Grow cells



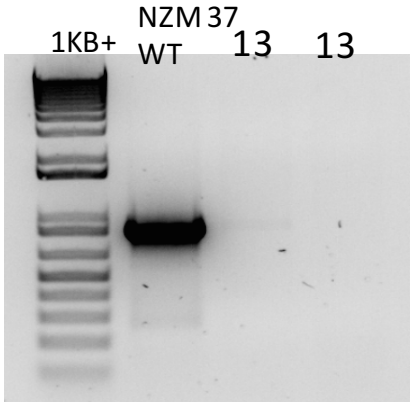
4

Isolate DNA from cells and find cells that have the gene knocked out

Just an example: Knockout of VEGF-a gene



Repeat PCR

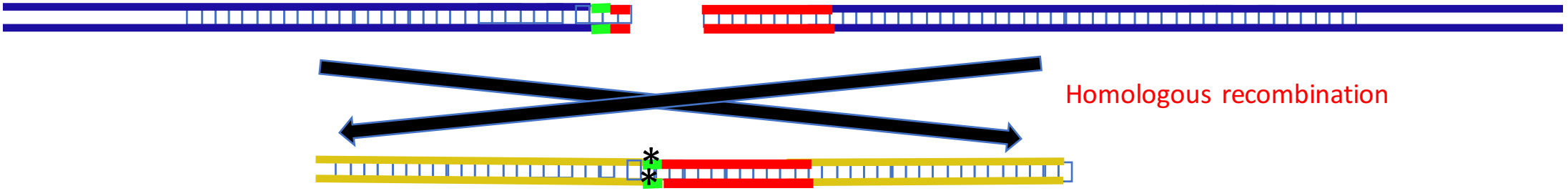


Here is an example of PCR of the VEGFA gene of melanoma cells where we have tried to use CRISPR to "knockout the VEGFA gene (achieved in clone 13)

Using CRISPR/Cas9 to "knockin" bits of DNA

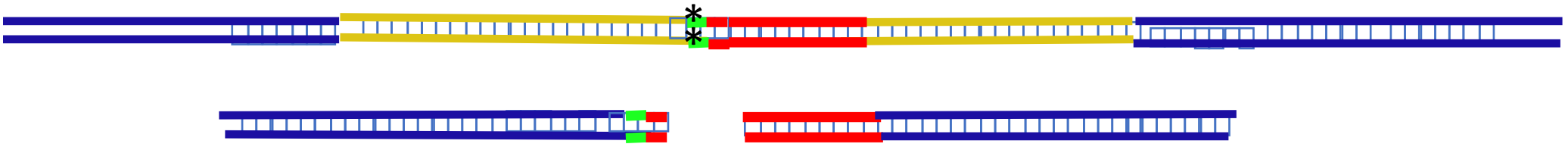
- We can use an artificial piece of DNA that is identical to the cleaved region of DNA (with "corrected" sequence -∗-) ------
- when the cell tries to repair its own chromosomal DNA it will sometimes accidentally incorporate this into its own DNA by homologous recombination!

→ Bring in CRISP-Cas9 components to make a specific cut
→ Bring in a DNA fragement that contains the desired genetic alteration (wt→ mutant; mutant→wt)
→ Strand invasion by cut sequence and HR

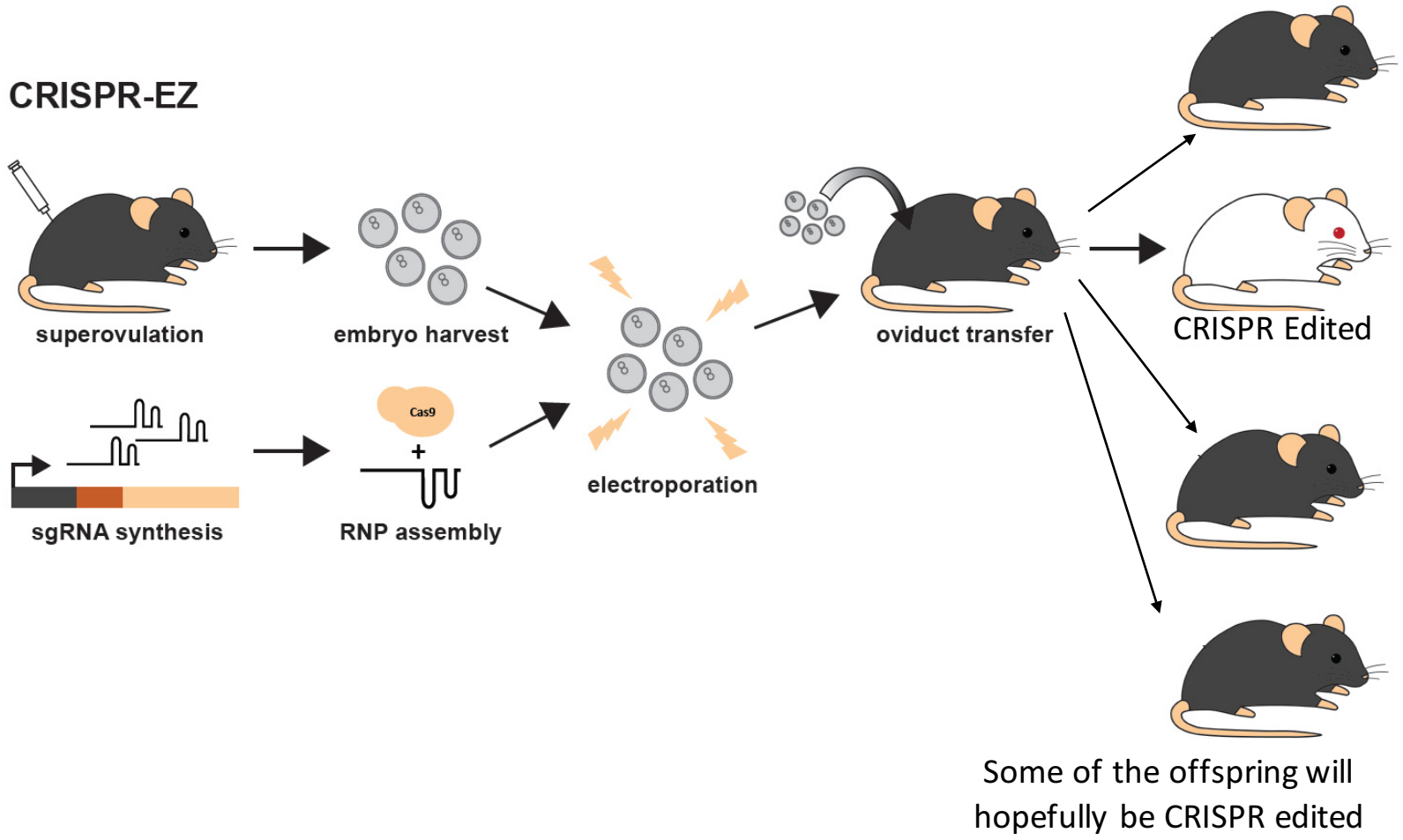


Using CRISPR/Cas9 to "knockin" bits of DNA

- Now the artificially produced piece of DNA is "knocked in" to the genome

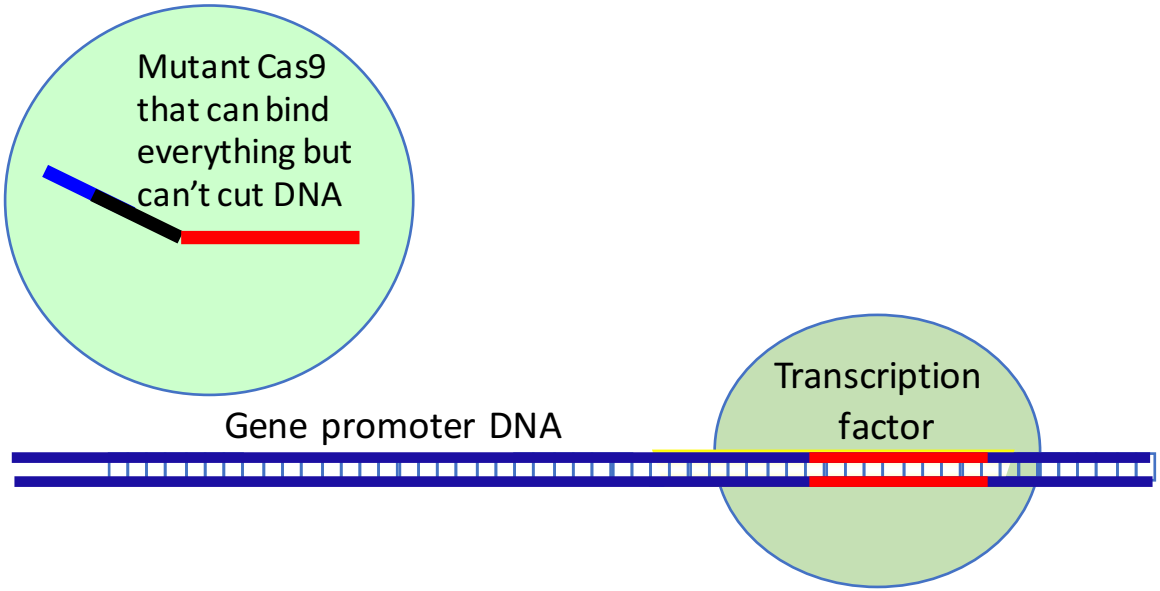


Making mice where genes are knocked out is easier and cheaper



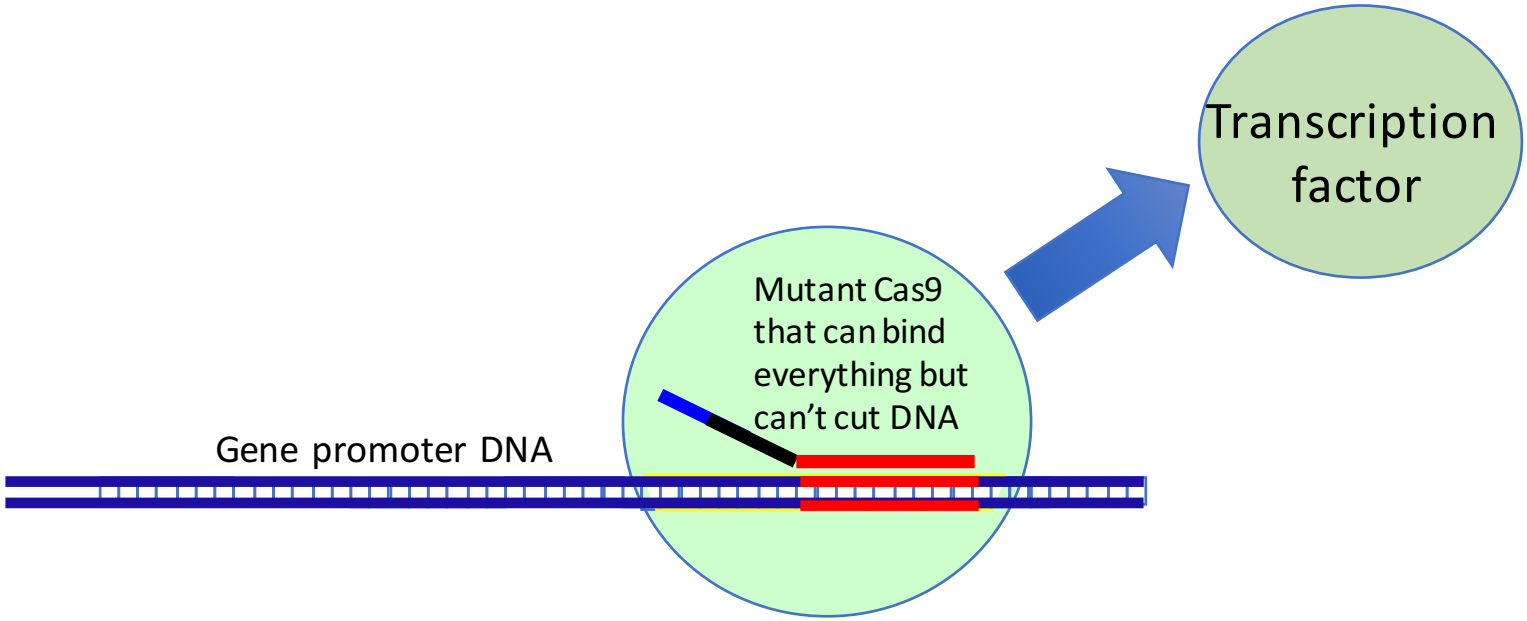
CRISPR/Cas9 to SWITCH ON or OFF GENES

- Uses a mutant **Cas9 that can bind everything but can't cut DNA**
- This means it locks on tightly to the DNA that matches the guide sequence and stops on that region of DNA
- An example of how this can be used is by having a big Cas9 protein sitting at specific transcription factor binding site: we can block the transcription factor from coming into the gene promoter and therefore we switch off the expression of that specific gene controlled by TF in a highly targeted way.
- Negative aspect: keeps DNA in a short RNA:DNA hybrid state



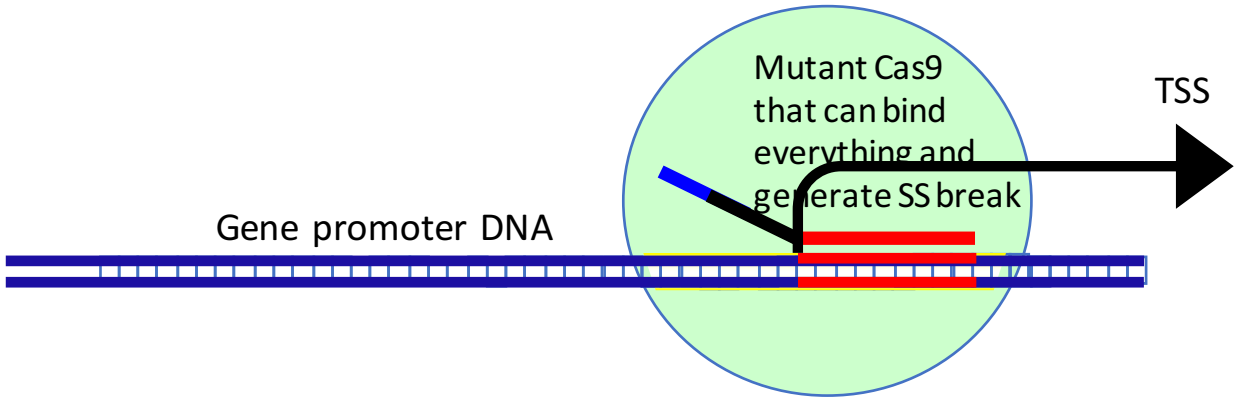
CRISPR/Cas9 to SWITCH ON or OFF GENES

- Uses a mutant **Cas9 that can bind everything but can't cut DNA**
- This means it locks on tightly to the DNA that matches the guide sequence and stops on that region of DNA
- An example of how this can be used is by having a big Cas9 protein sitting at specific transcription factor binding site: we can block the transcription factor from coming into the gene promoter and therefore we switch off the expression of that specific gene controlled by TF in a highly targeted way.



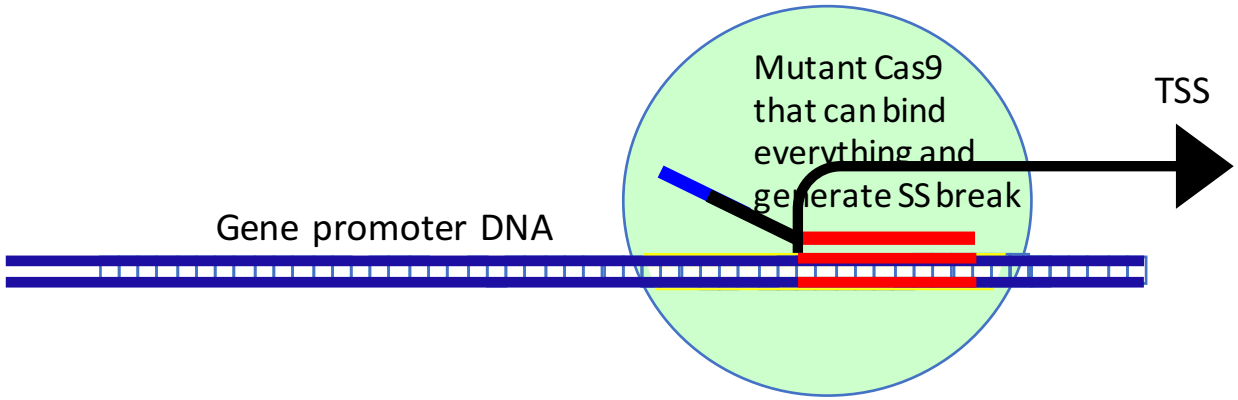
CRISPR/Cas9 to SWITCH ON or OFF GENES

- Alterantively we can use a mutant **Cas9 that can bind everything but generates a single strand cut DNA**
- This means it locks on tightly to the DNA that matches the guide sequence, stops and generate a single strand break that turn-off gene expression



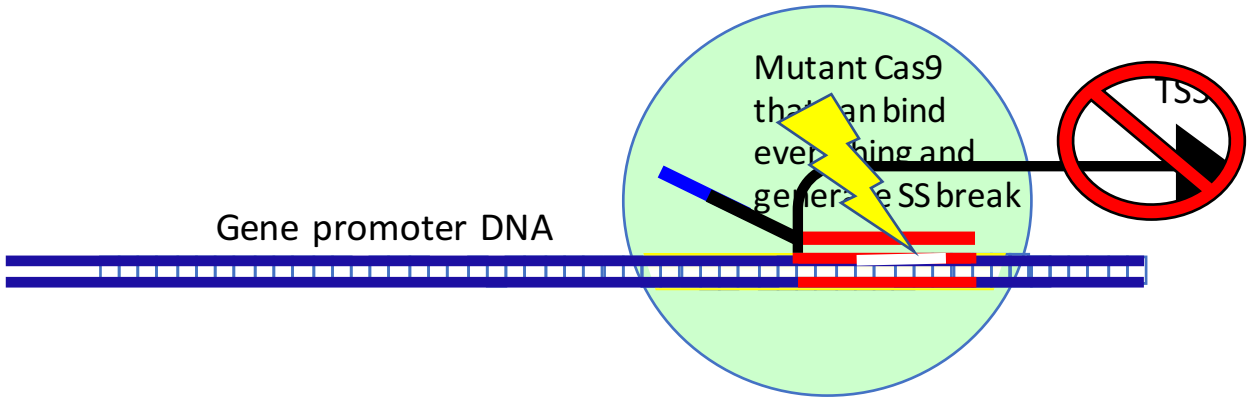
CRISPR/Cas9 to SWITCH ON or OFF GENES

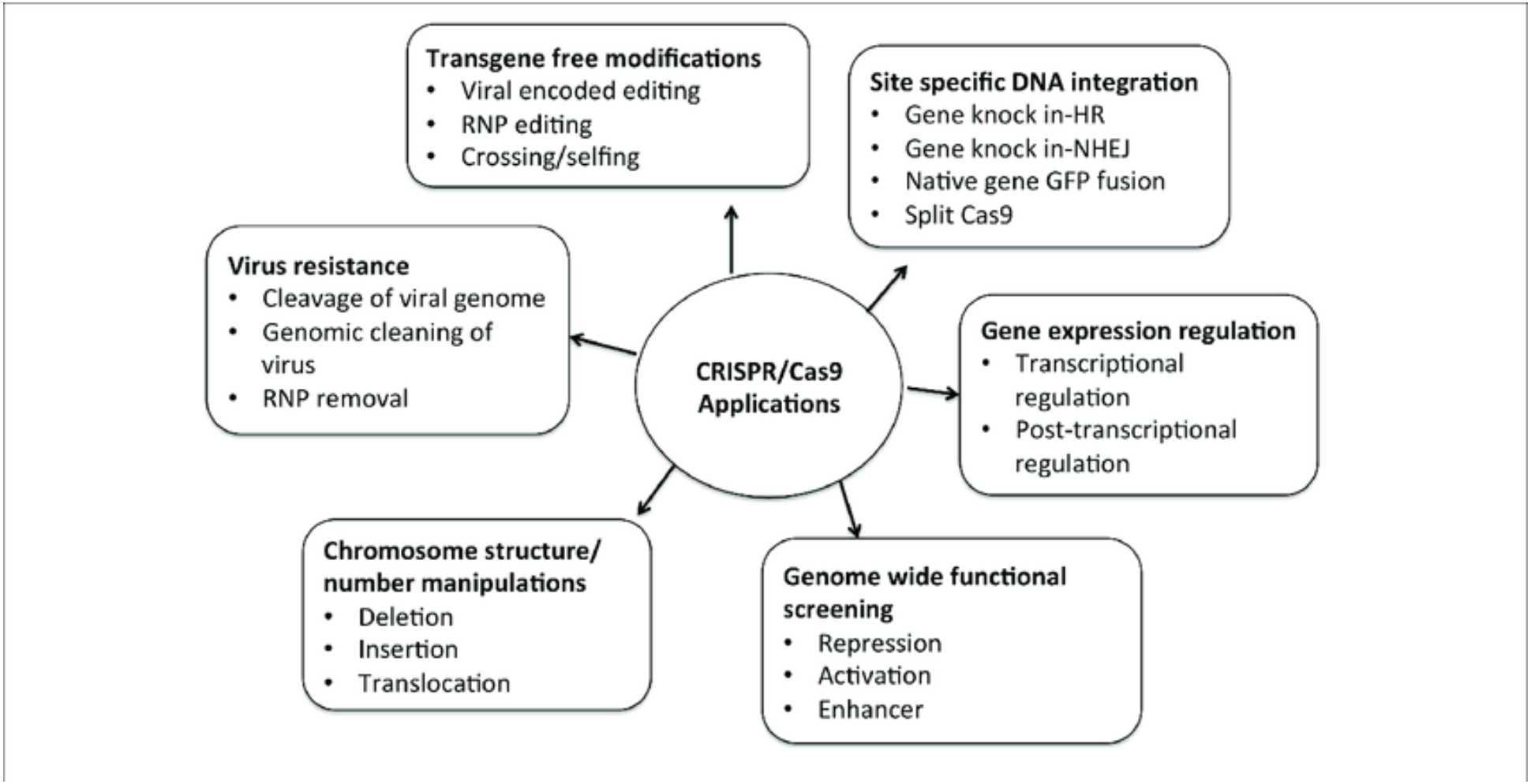
- Alterantively we can use a mutant **Cas9 that can bind everything but generates a single strand cut DNA**
- This means it locks on tightly to the DNA that matches the guide sequence, stops and generate a single strand break that turn-off gene expression



CRISPR/Cas9 to SWITCH ON or OFF GENES

- Alterantively we can use a mutant **Cas9 that can bind everything but generates a single strand cut DNA**
- This means it locks on tightly to the DNA that matches the guide sequence, stops and generate a single strand break that turn-off gene expression




Transgene free modifications
• Viral encoded editing
• RNP editing
• Crossing/selfing
Site specific DNA integration
• Gene knock in-HR
• Gene knock in-NHEJ
• Native gene GFP fusion
• Split Cas9
Virus resistance
• Cleavage of viral genome
• Genomic cleaning of virus
• RNP removal
CRISPR/Cas9 Applications
Gene expression regulation
• Transcriptional regulation
• Post-transcriptional regulation
Chromosome structure/ number manipulations
• Deletion
• Insertion
• Translocation
Genome wide functional screening
• Repression
• Activation
• Enhancer

Using CRISPR a weapon to wipe out mosquitos

A gene drive is a genetic engineering technology that can propagate a particular suite of genes throughout a population. Gene drives can arise through a variety of mechanisms. They have been proposed to provide an effective means of genetically modifying specific populations and entire species.