

CHAPTER 3

Recombinant DNA Technology and DNA Cloning

PART 3 of 4

Studying Gene Expression

Northern Blotting

Techniques involve analyzing mRNA produced by a tissue

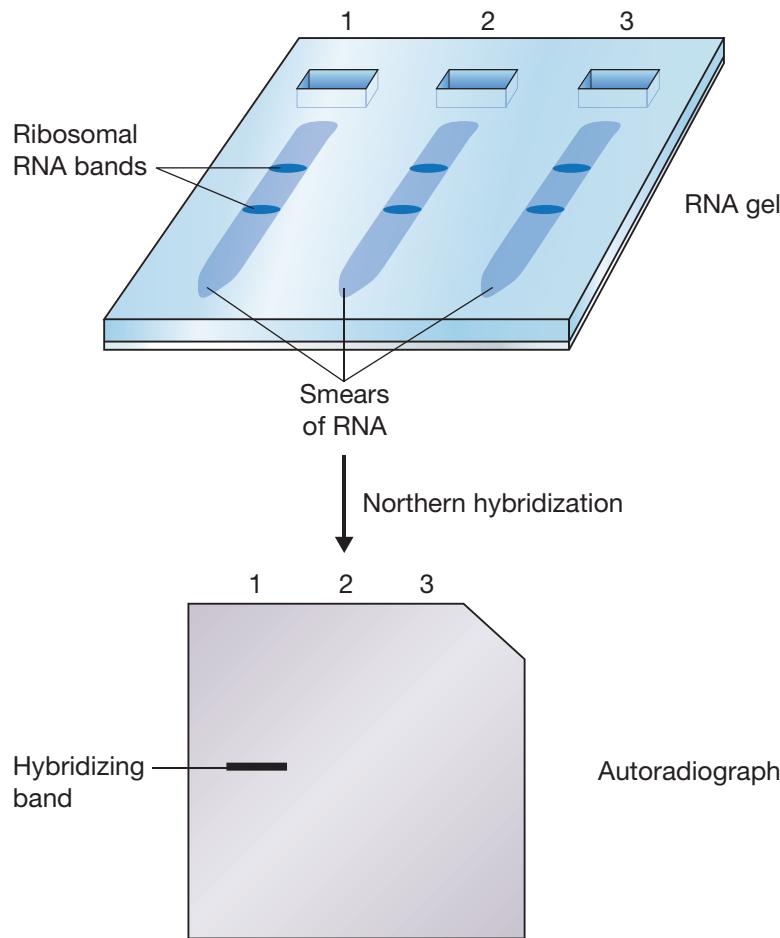
Basic method is similar to Southern blotting

RNA is isolated from a tissue of interest, separated by gel electrophoresis, blotted onto a membrane, and hybridized to a labeled DNA probe

Can compare and quantify amounts of mRNA present in different tissues

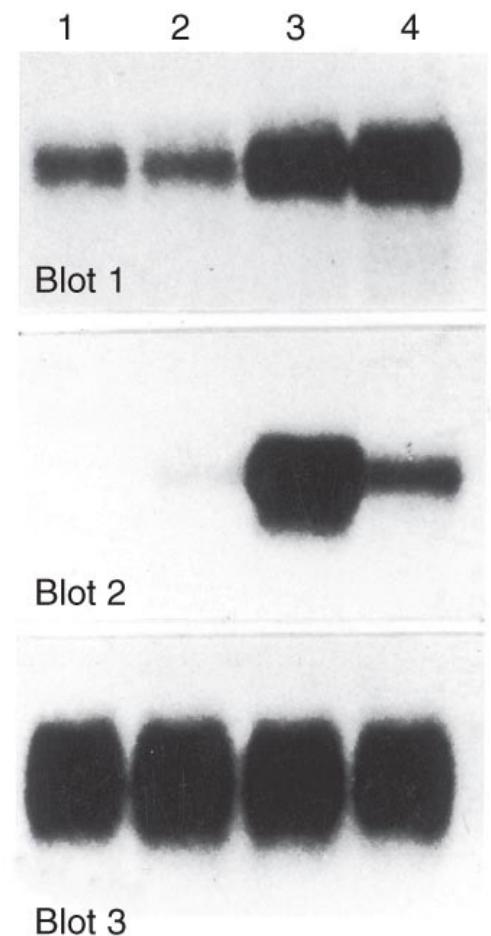
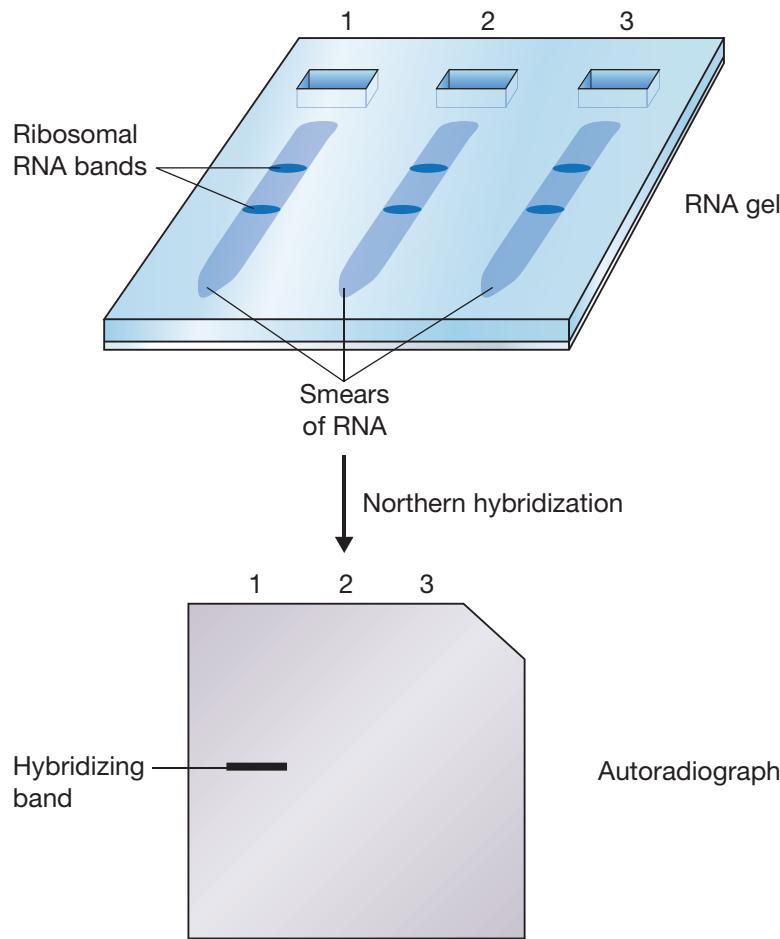
Studying Gene Expression

Northern Blotting



Studying Gene Expression

Northern Blotting



Studying Gene Expression

Reverse transcription PCR

Reverse transcription PCR- used to study mRNA levels when level of detection is below that of Northern

Reverse Transcriptase is used to make double stranded cDNA from mRNA

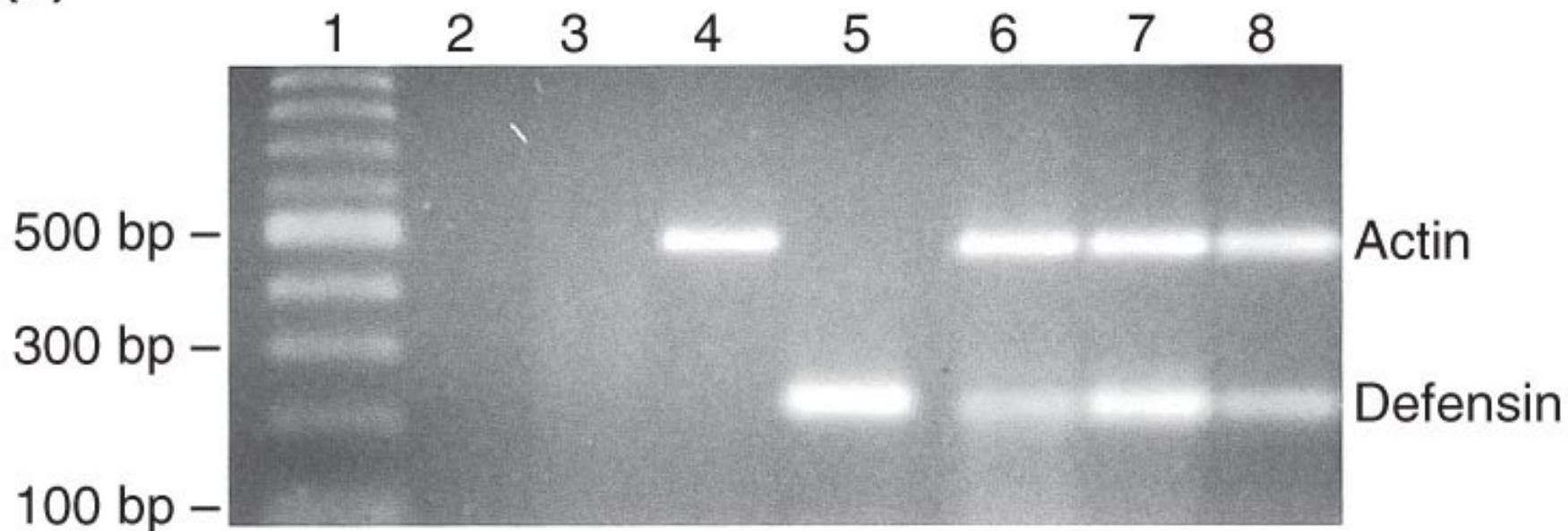
Use PCR to amplify region of cDNA with set of primers specific for gene of interest

Amount of cDNA produced in RT PCR reaction for gene of interest reflects amount of mRNA and level of gene expression

Studying Gene Expression

Reverse transcription PCR

(b)



© 2013 Pearson Education, Inc.

Studying Gene Expression

Real time or quantitative (qPCR)

Can quantify amplification reactions as they occur in real time

Need special thermal cyclers that use a laser to scan a beam of light through the top or bottom of each PCR reaction

Each reaction tube contains EITHER a dye containing probe or DNA binding dye that emits fluorescent light when illuminated by the laser

Studying Gene Expression

Real time or quantitative (qPCR)

Light emitted by the dyes correlates with amount of PCR product amplified

Light is captured by the detector which relays info. to the computer to provide readout on amount of fluorescence

Readout is plotted and analyzed to quantify the number of PCR products produced after each cycle

Real time or quantitative (qPCR)

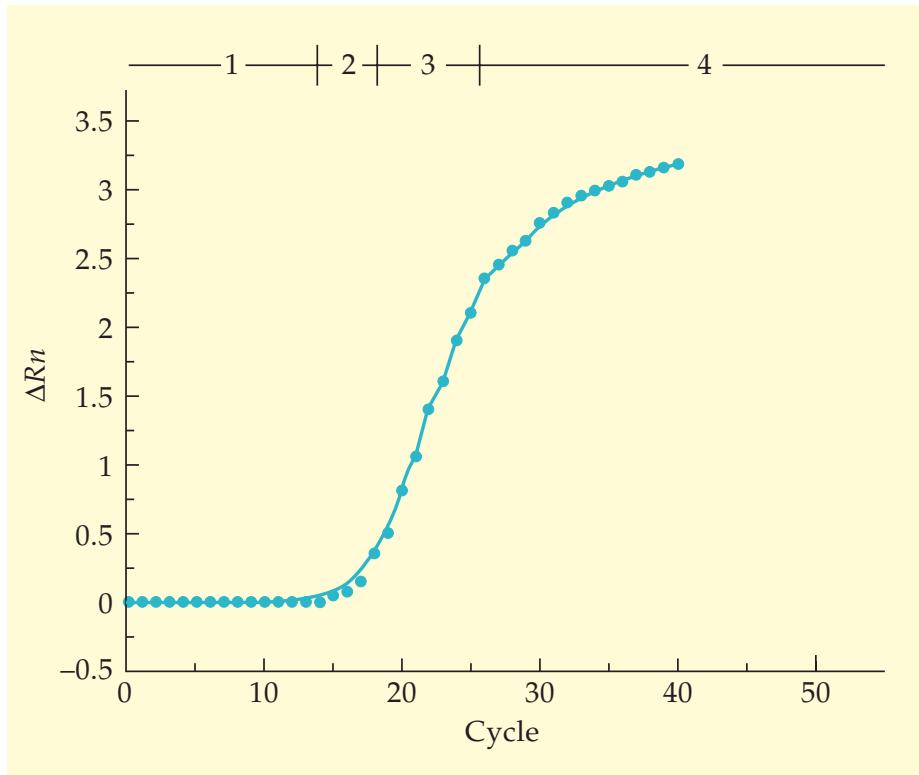
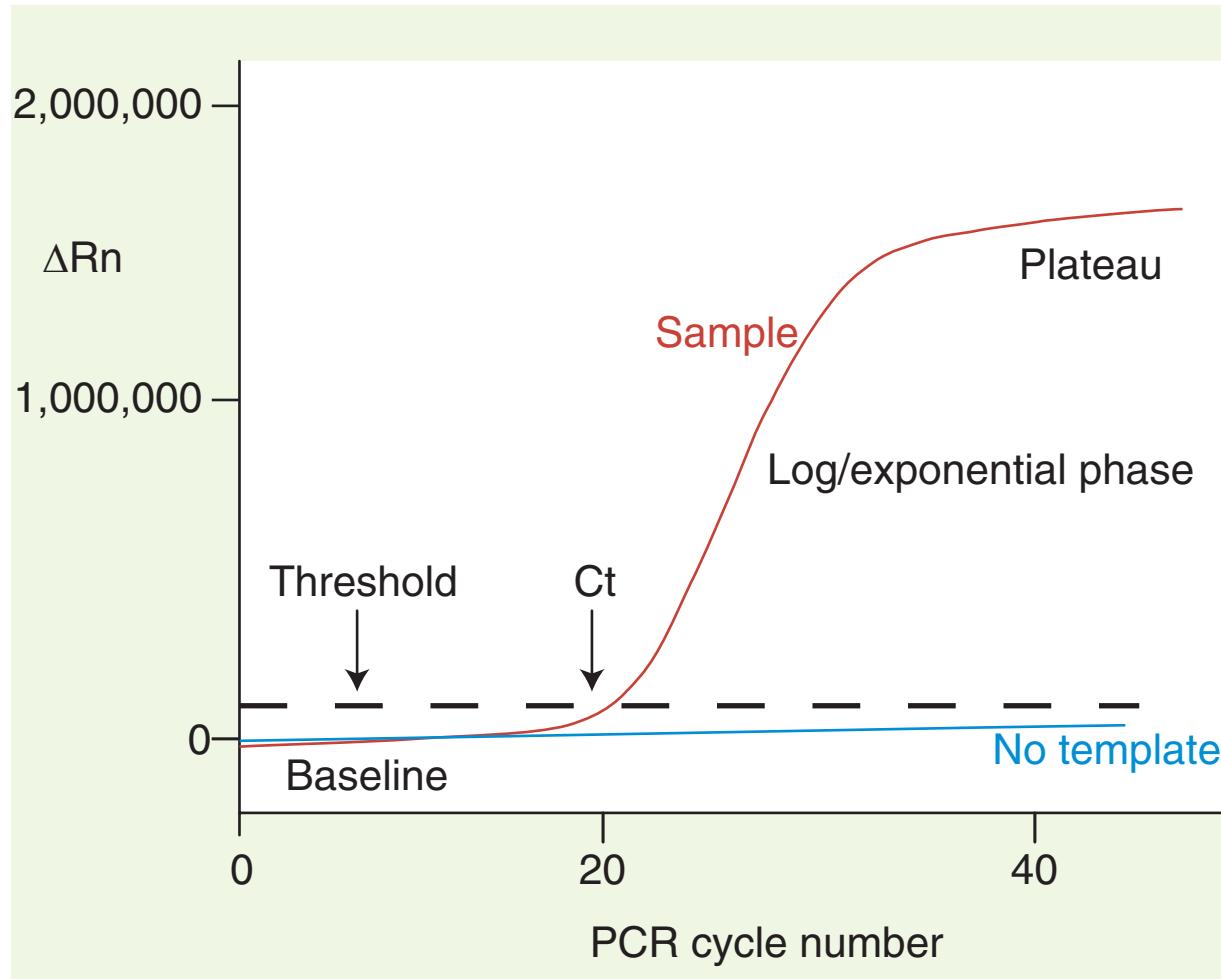
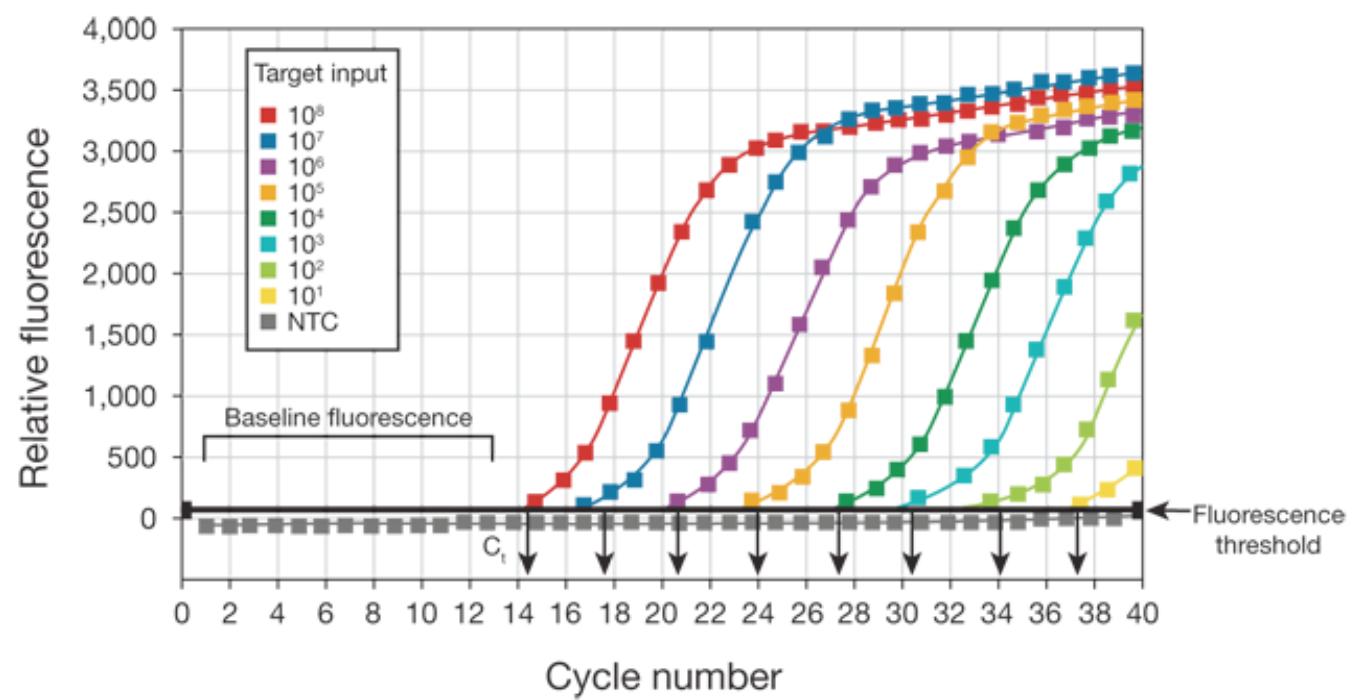
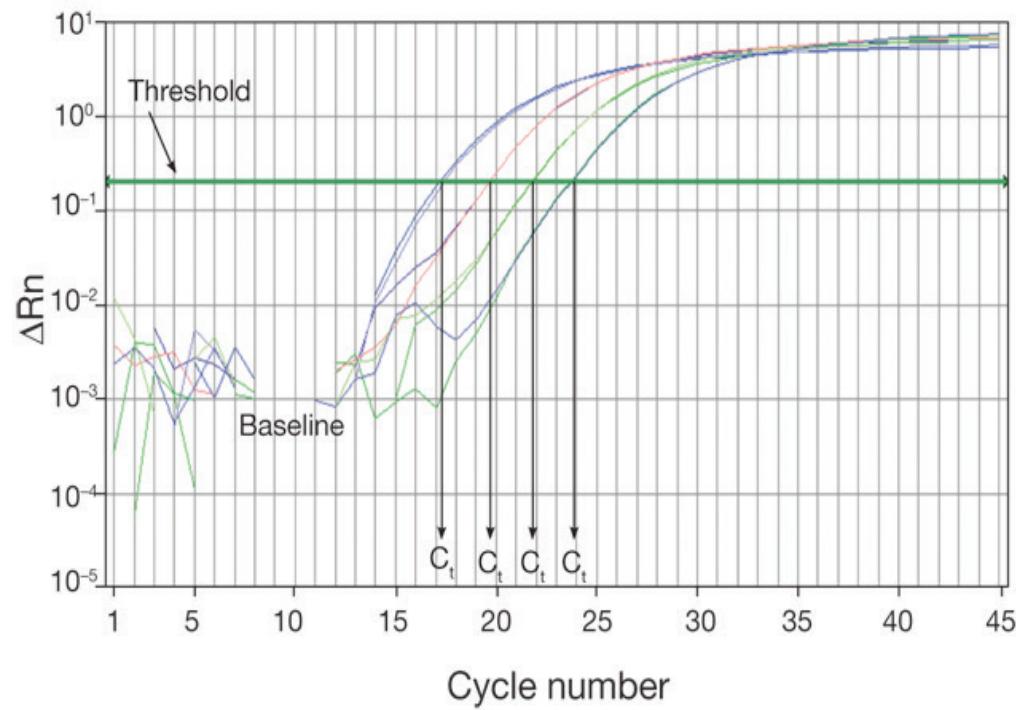


FIGURE 9.21 A plot of ΔRn (normalized fluorescence) versus cycle number in a real-time PCR experiment. Four phases of PCR are shown. (1) A linear phase, where fluorescence emission is not yet above background level. (2) An early exponential phase, where the fluorescence intensity becomes significantly higher than the background. The cycle at which this occurs is generally known as C_T . (3) An exponential phase, where the amount of product doubles in each cycle. (4) A plateau phase, where reaction components are limited and amplification slows down.

Real time or quantitative (qPCR)





Real time or quantitative (qPCR)

Baseline

The PCR cycles in which a reporter fluorescent signal is accumulating but is beneath the limits of detection of the instrument.

Normalized fluorescence (ΔR_n)

A computer software program calculates a ΔR_n using the equation $R_n = R_{nf} - R_{nb}$, where R_{nf} is the fluorescence emission of the product at each time point and R_{nb} is the fluorescence emission of the baseline.

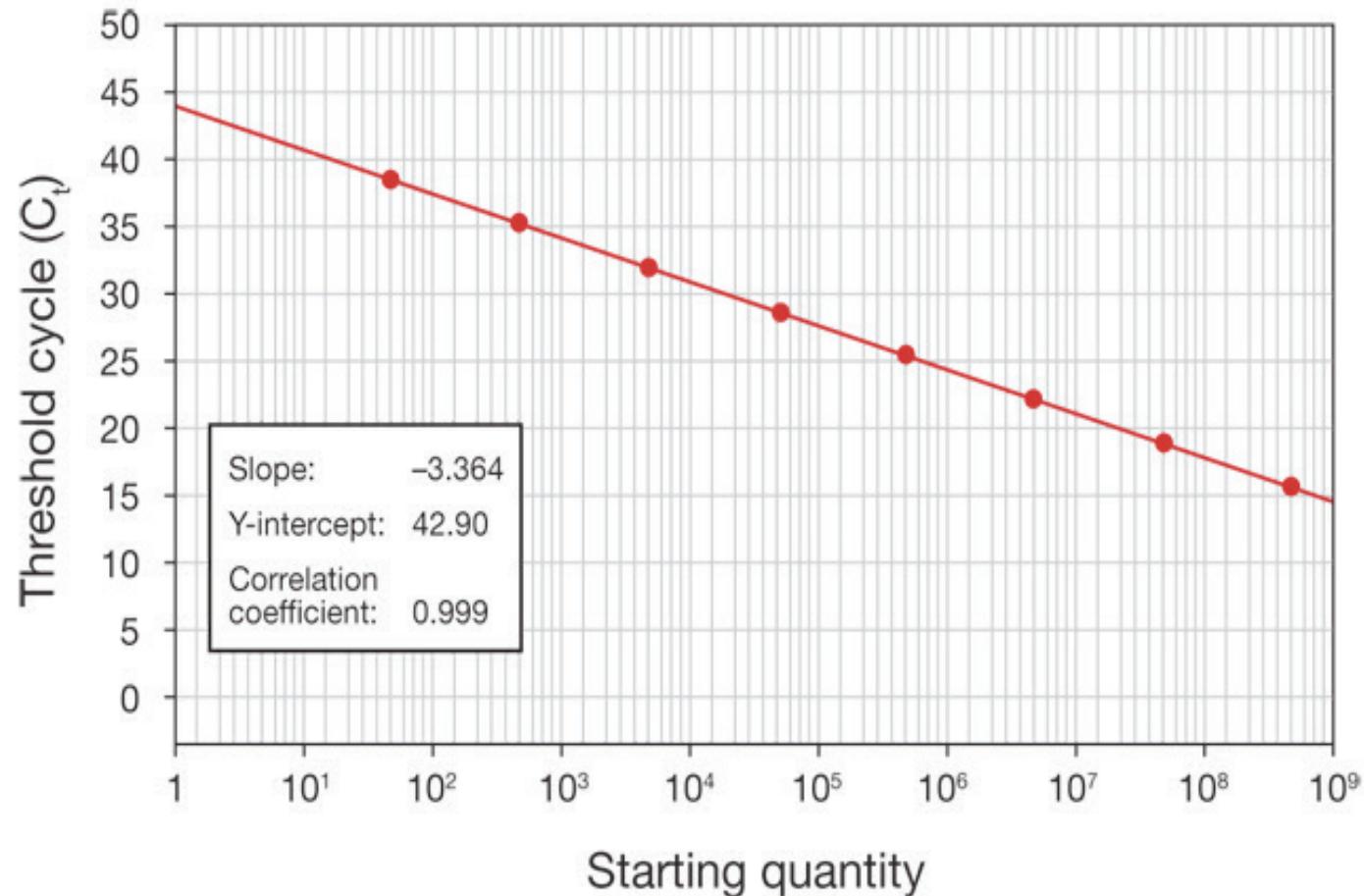
Threshold

An arbitrary threshold is chosen by the computers, based on the variability of the baseline. It is calculated as ten-times the standard deviation of the average signal of the baseline fluorescent signal between cycles three to 15. A fluorescent signal that is detected above the threshold is considered a real signal that can be used to define the threshold cycle (C_t) for a sample.

Threshold cycle (C_t)

The fractional PCR cycle number at which the reporter fluorescence is greater than the minimal detection level (i.e., the threshold). The C_t is a basic principle of real-time PCR and is an essential component in producing accurate and reproducible data.

Real time or quantitative (qPCR)



Studying Gene Expression

Real time or quantitative (qPCR)

Two major approaches:

Taqman probes are complimentary to specific regions of target DNA between forward and reverse primers for PCR

Taqman probes contain two dyes:

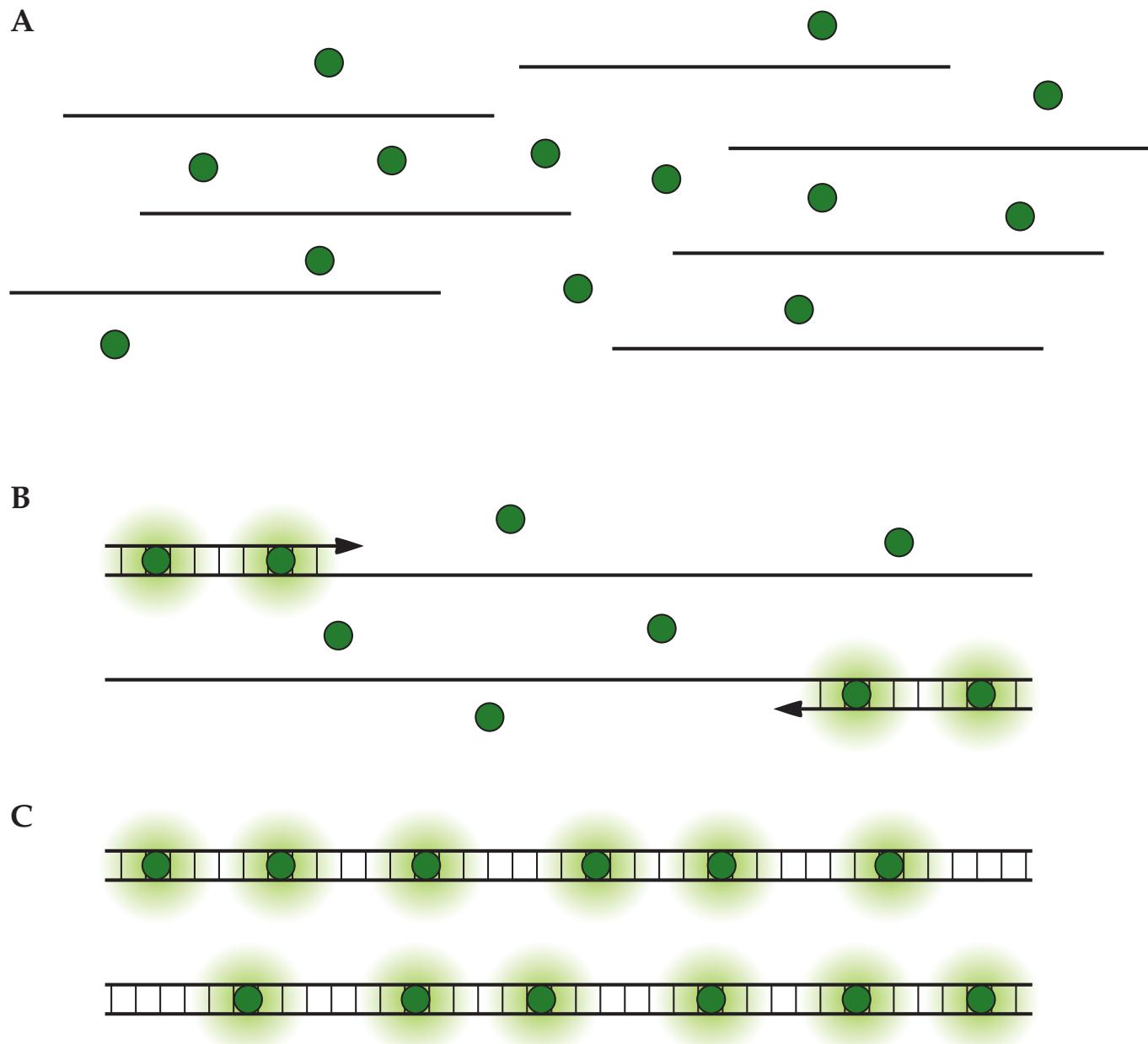
1. Reporter located at 5' end of probe and can release fluorescent light when excited by the laser
2. Quencher which is attached to 3' end of probe

Studying Gene Expression

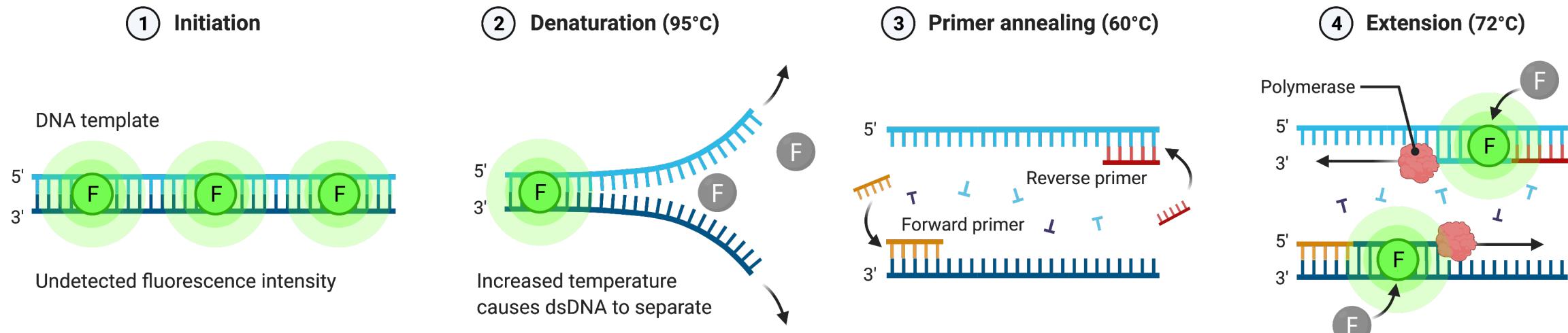
Real time or quantitative (qPCR)

SYBR green- binds double stranded DNA and as more double stranded DNA is copied with each round of qPCR there are more

DNA copies to bind SYBR Green which increases amount of fluorescent light emitted

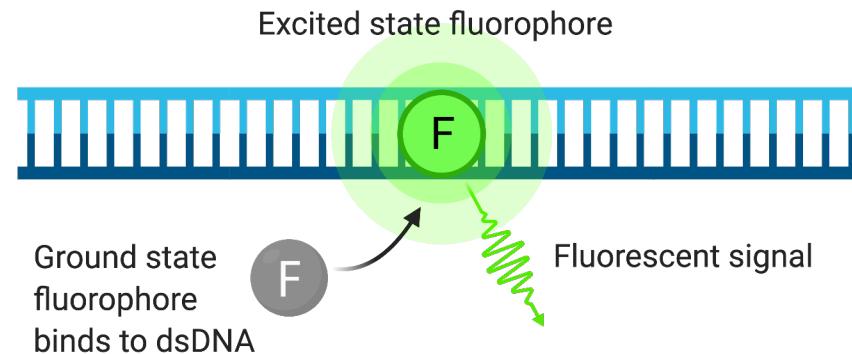


Studying Gene Expression: Dye-based detection

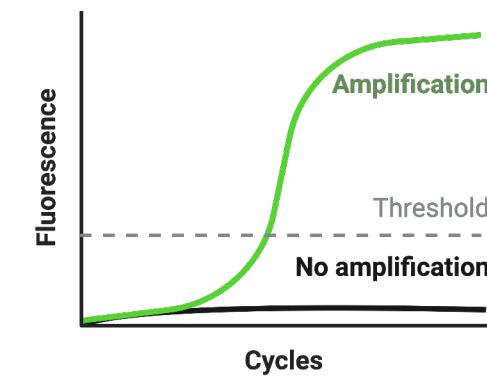


Key Concept

The higher the concentration of dsDNA the higher the fluorescence intensity signal.

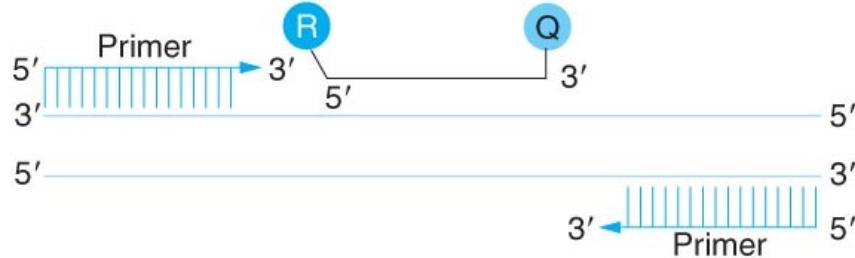


Results

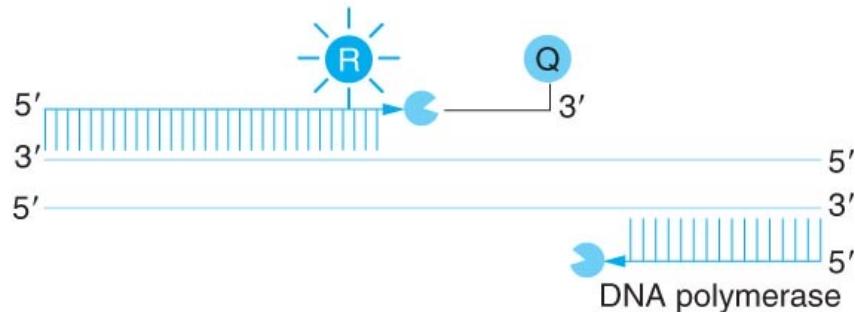


Studying Gene Expression

Real time PCR- Taqman

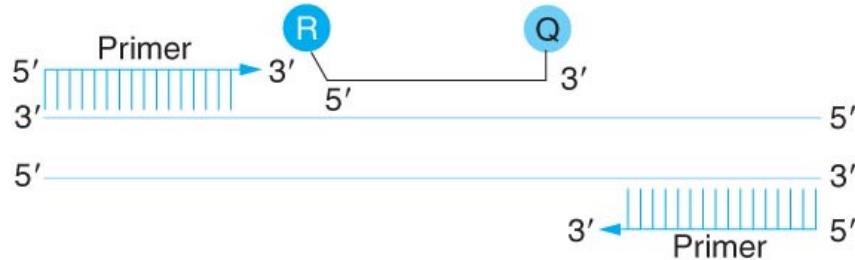


- 1. Hybridization.** Forward and reverse PCR primers bind to denatured target DNA. TaqMan probe with reporter (R) and quencher (Q) dye binds to target DNA between the primers. When probe is intact, emission by the reporter dye is quenched.
- 2. Extension.** As DNA polymerase extends the forward primer, it reaches the TaqMan probe and cleaves the reporter dye from the probe. Released from the quencher the reporter can now emit light when excited by a laser.



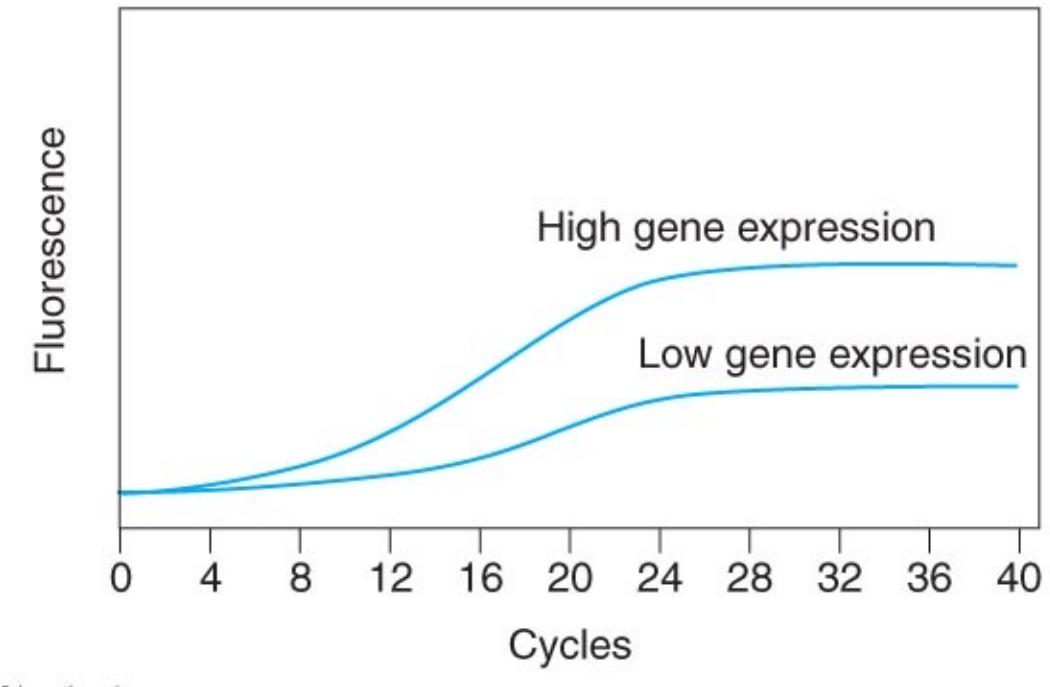
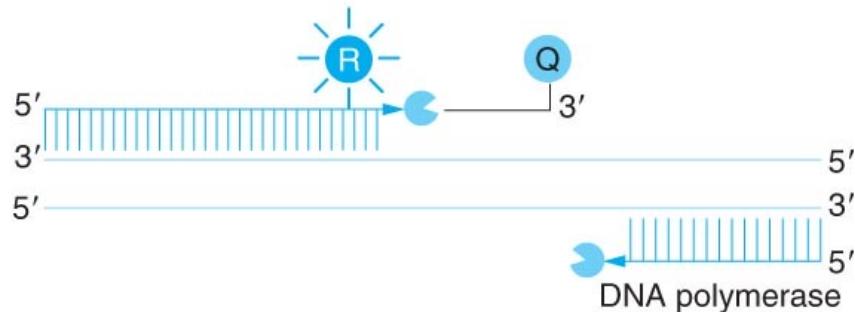
Studying Gene Expression

Real time PCR- Taqman

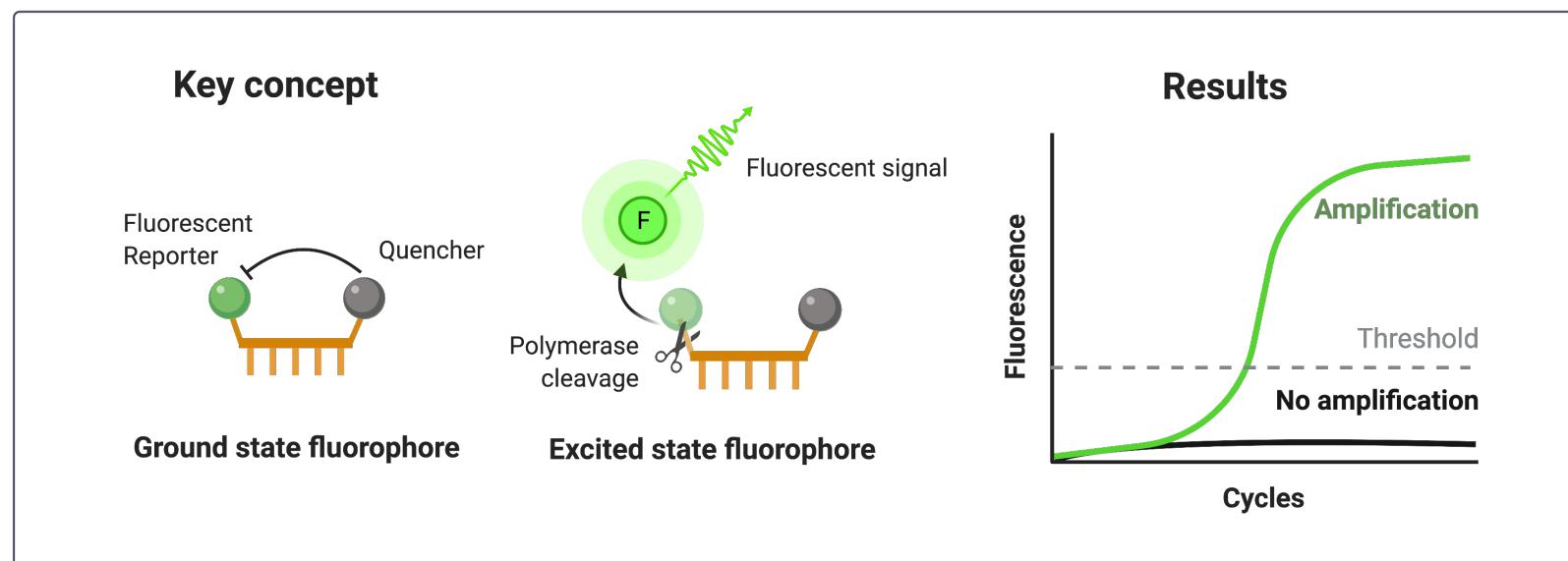
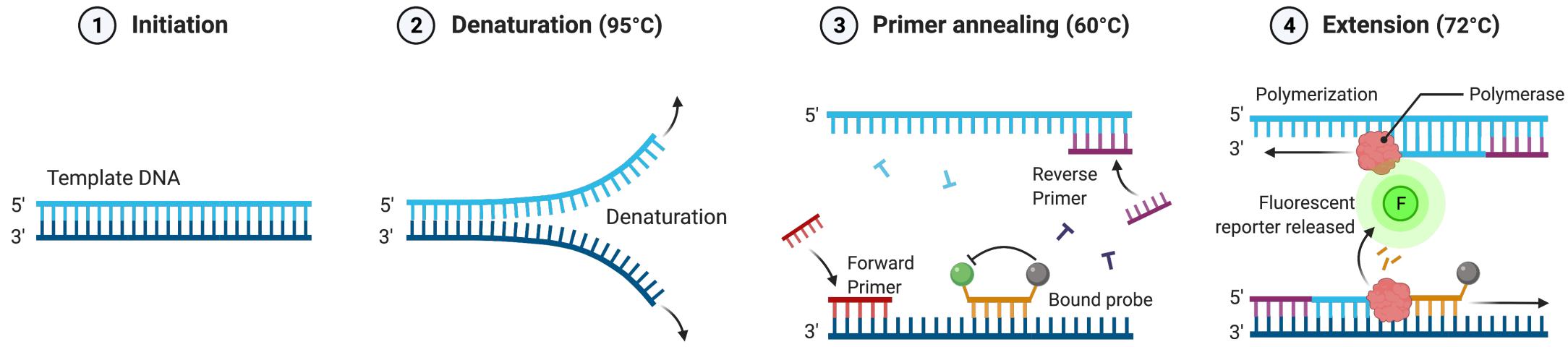


1. **Hybridization.** Forward and reverse PCR primers bind to denatured target DNA. TaqMan probe with reporter (R) and quencher (Q) dye binds to target DNA between the primers. When probe is intact, emission by the reporter dye is quenched.

2. **Extension.** As DNA polymerase extends the forward primer, it reaches the TaqMan probe and cleaves the reporter dye from the probe. Released from the quencher the reporter can now emit light when excited by a laser.



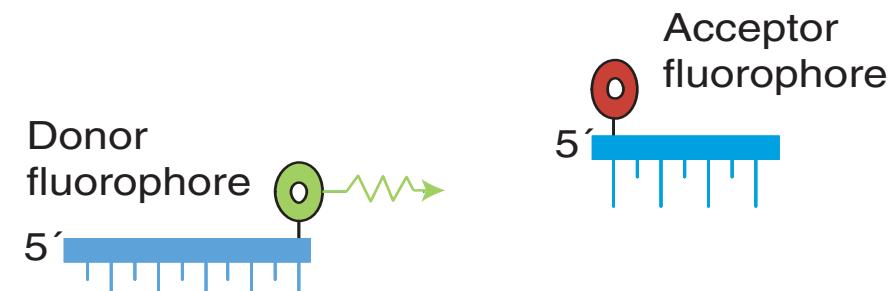
Studying Gene Expression: Probe-based detection



Studying Gene Expression

Real time PCR- Dual Hybridization Probes

One hybridization probe carries a donor fluorophore at its 3' end and the other is labeled with an acceptor fluorophore at its 5' end. During denaturation both hybridization probes remain separate in solution and any fluorescent emission from the donor fluorophore (e.g., green fluorescence, which occurs when excited by the LightCycler's light source) is disregarded by the detector

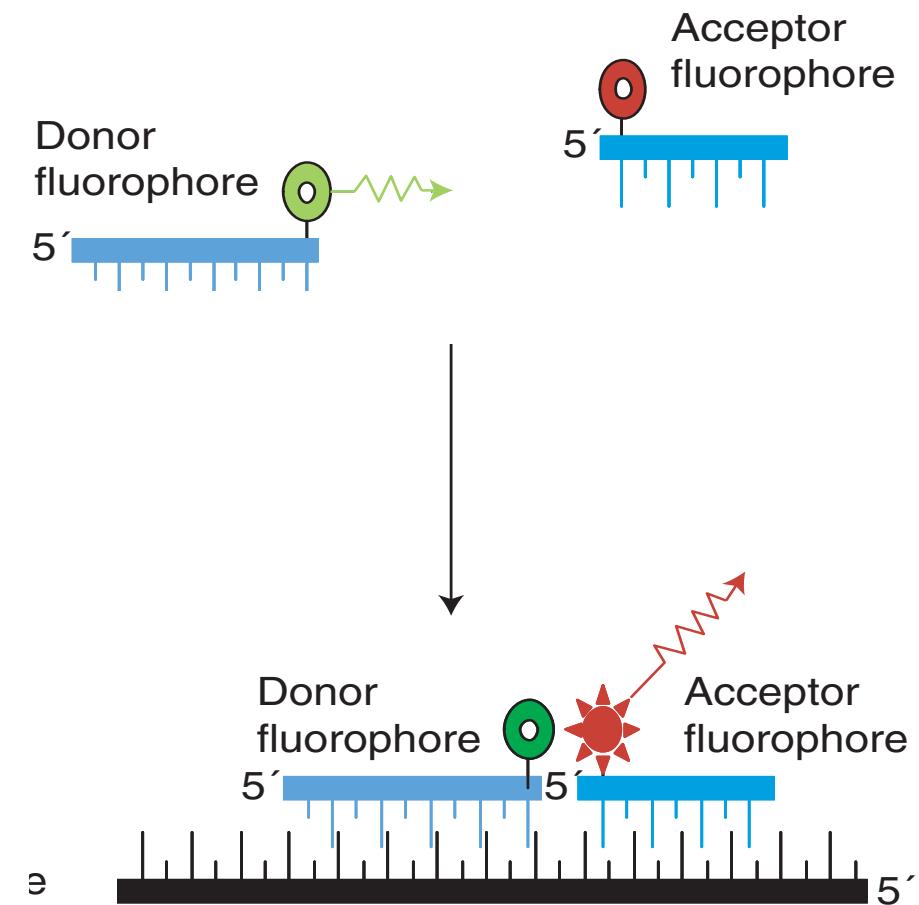


Studying Gene Expression

Real time PCR- Dual Hybridization Probes

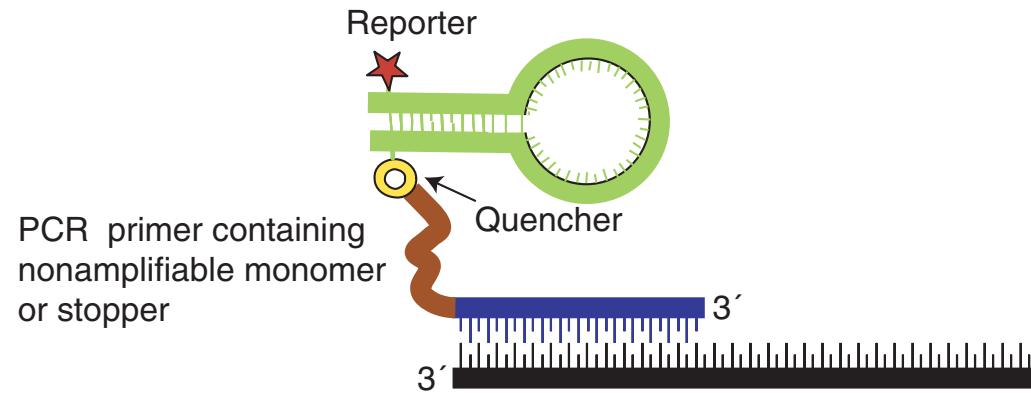
One hybridization probe carries a donor fluorophore at its 3' end and the other is labeled with an acceptor fluorophore at its 5' end. During denaturation both hybridization probes remain separate in solution and any fluorescent emission from the donor fluorophore (e.g., green fluorescence, which occurs when excited by the LightCycler's light source) is disregarded by the detector

During annealing the probes hybridize in a head-to-tail conformation, bringing the two dyes next to each other. Excitation of the donor leads to fluorescence resonance energy transfer to the acceptor resulting in a change of the fluorescent signal and emission of fluorescent light at a longer wavelength (red)



Studying Gene Expression

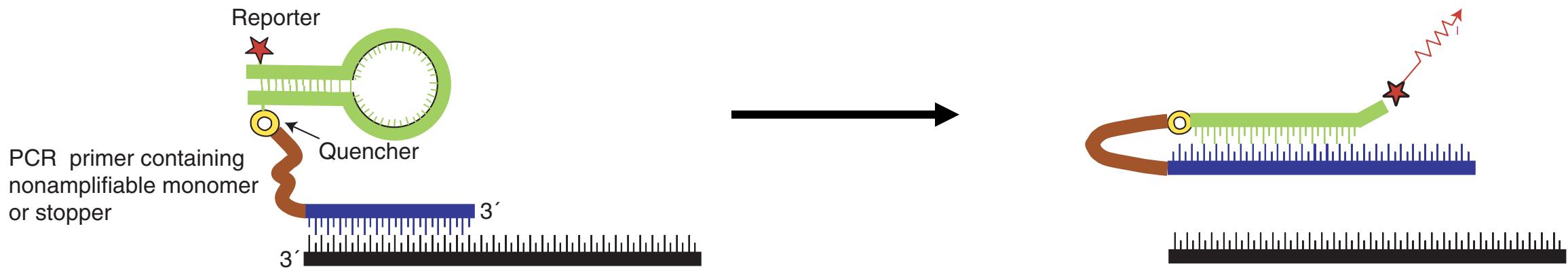
Real time PCR- Scorpion Probes



The Scorpion probe adopts a hairpin loop structure that is linked to the 5' end of a specific primer through a PCR stopper that prevents readthrough of the hairpin loop. During PCR, the Scorpion primers are extended to become part of the amplicon

Studying Gene Expression

Real time PCR- Scorpion Probes

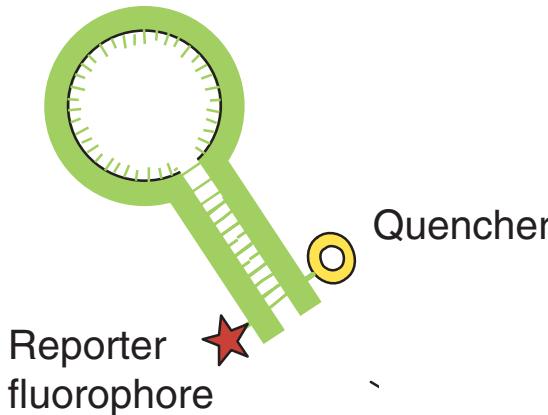


The Scorpion probe adopts a hairpin loop structure that is linked to the 5' end of a specific primer through a PCR stopper that prevents readthrough of the hairpin loop. During PCR, the Scorpion primers are extended to become part of the amplicon

During annealing, the probe sequence in the Scorpion hairpin loop hybridizes to the newly formed complementary target sequence in the PCR product, separating the fluorophore and quencher dyes and leading to emission of a fluorescent signal. As the tail of the Scorpion and the PCR product are now part of the same DNA strand, the interaction is intramolecular

Studying Gene Expression

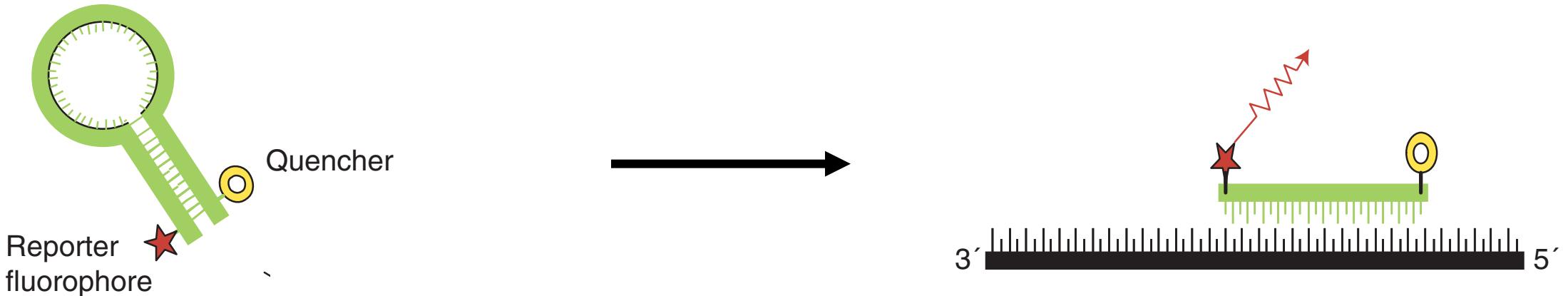
Real time PCR- Molecular Beacon



Molecular beacons adopt a hairpin structure whilst free in solution. The hairpin structure consists of a stem built of two complementary arms and a loop that is complementary to the target sequence. This configuration helps the reporter fluorescent dye and the quencher to remain extremely close and therefore no fluorescence is detected.

Studying Gene Expression

Real time PCR- Molecular Beacon



Molecular beacons adopt a hairpin structure whilst free in solution. The hairpin structure consists of a stem built of two complementary arms and a loop that is complementary to the target sequence. This configuration helps the reporter fluorescent dye and the quencher to remain extremely close and therefore no fluorescence is detected.

During annealing, the beacons hybridize to the target sequence, which changes their conformation & separates reporter & quencher dyes resulting in fluorescence being emitted

Digital PCR

Third generation, most advanced PCR technology designed for specific detection and absolute quantification of nucleic acids

Major features:

- Based on real-time PCR chemistry and workflow
- Reagents and amplification are the same as real-time PCR
- End-point detection
- Absolute quantification without a reference
- Binary signal: 1 – positive; 0 – negative
- Quantifies low abundance target molecules

Digital PCR

Third generation, most advanced PCR technology designed for specific detection and absolute quantification of nucleic acids

Major features:

- Based on real-time PCR chemistry and workflow
- Reagents and amplification are the same as real-time PCR
- End-point detection
- Absolute quantification without a reference
- Binary signal: 1 – positive; 0 – negative
- Quantifies low abundance target molecules

PCR techniques comparison

<i>First generation</i>	<i>Second generation</i>	<i>Third generation</i>
PCR	Quantitative PCR (qPCR)	Digital PCR (dPCR)

PCR techniques comparison

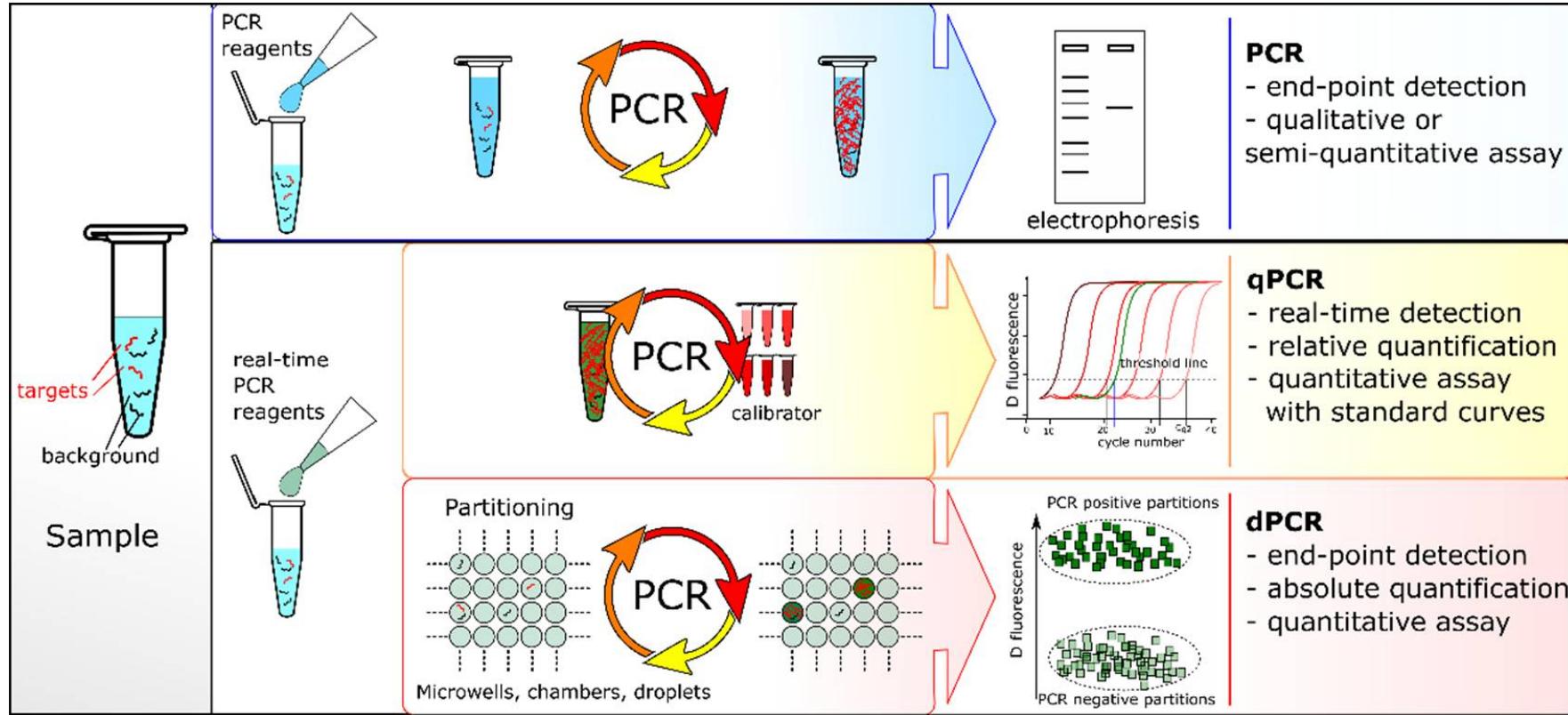
<i>First generation</i>	<i>Second generation</i>	<i>Third generation</i>
PCR	Quantitative PCR (qPCR)	Digital PCR (dPCR)
1983	1996	2006

PCR techniques comparison

<i>First generation</i>	<i>Second generation</i>	<i>Third generation</i>
PCR	Quantitative PCR (qPCR)	Digital PCR (dPCR)
1983	1996	2006
Qualitative	Relative quantification	Absolute quantification

PCR techniques comparison

<i>First generation</i>	<i>Second generation</i>	<i>Third generation</i>
PCR	Quantitative PCR (qPCR)	Digital PCR (dPCR)
1983	1996	2006
Qualitative	Relative quantification	Absolute quantification
Technically simple Multiplexing capabilities End-point detection Low cost	High accuracy High sensitivity & specificity Rapid cycling and throughput Non-specific amplification Real-time detection	No standard curves Higher precision & reproducibility Low sensitivity to inhibitors End-point detection



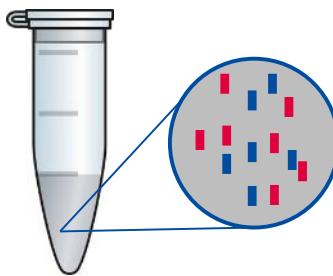
- In conventional PCR, the amplification products are analyzed at the end of the reaction (end-point PCR) by gel electrophoresis and detected after fluorescent staining.
- qPCR and dPCR use the same amplification reagents and fluorescent labeling systems.
- In qPCR, the amount of amplified DNA is measured at each cycle during the PCR reaction, i.e., in real-time.
- The ‘absolute’ quantity of target sequence is interpolated using a standard curve generated with a calibrator.
- In dPCR, the sample is first partitioned into many sub-volumes (in microwells, chambers or droplets) such that each partition contains either a few or no target sequences.
- After PCR, the proportion of amplification-positive partitions serves to calculate the concentration of concentration of the target sequence using Poisson’s statistics

Digital PCR

Step 1

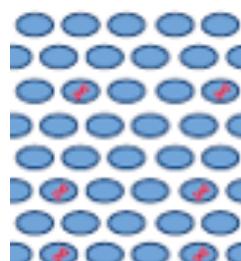
Sample dilution and PCR reaction mix setup

- Over concentrated sample – all positives partitions
- A number of negative partitions are required for accuracy
- Intercalating dye or hydrolysis probe-based reactions

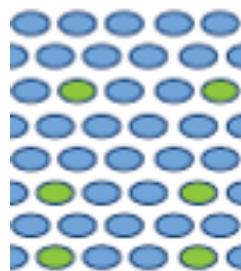


Target

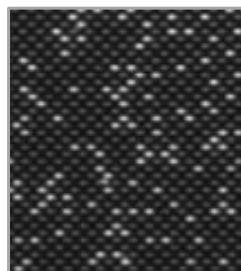
Background (gDNA, cDNA; primers/probes; master mix)



PCR reaction partitioning into thousands individual reactions



Positive reactions
Negative reactions



Absolute quantification

Step 2

Partitioning of PCR reaction

- Droplet generation
- Partition generation

Step 3

Amplification of partitioned PCR reaction

- End-point thermocycling

Step 4

Readout and quantification

- Positive and negative partitions are counted per reaction
- Data analysis

Digital assays vs. conventional analog biological assays

Simple mathematical computation of the Poisson law for a digital PCR experiment looks like this:

λ is the expectation value for these events and will tell what the most likely average value of copies/partitions is

$$\lambda = -\ln \left(\frac{\text{number of valid partitions} - \text{number of positive partitions}}{\text{number of valid partitions}} \right)$$

Computing the estimated concentration of target genes in copies/ μl

Given:
Valid partitions: 8000
Positive partitions: 4000

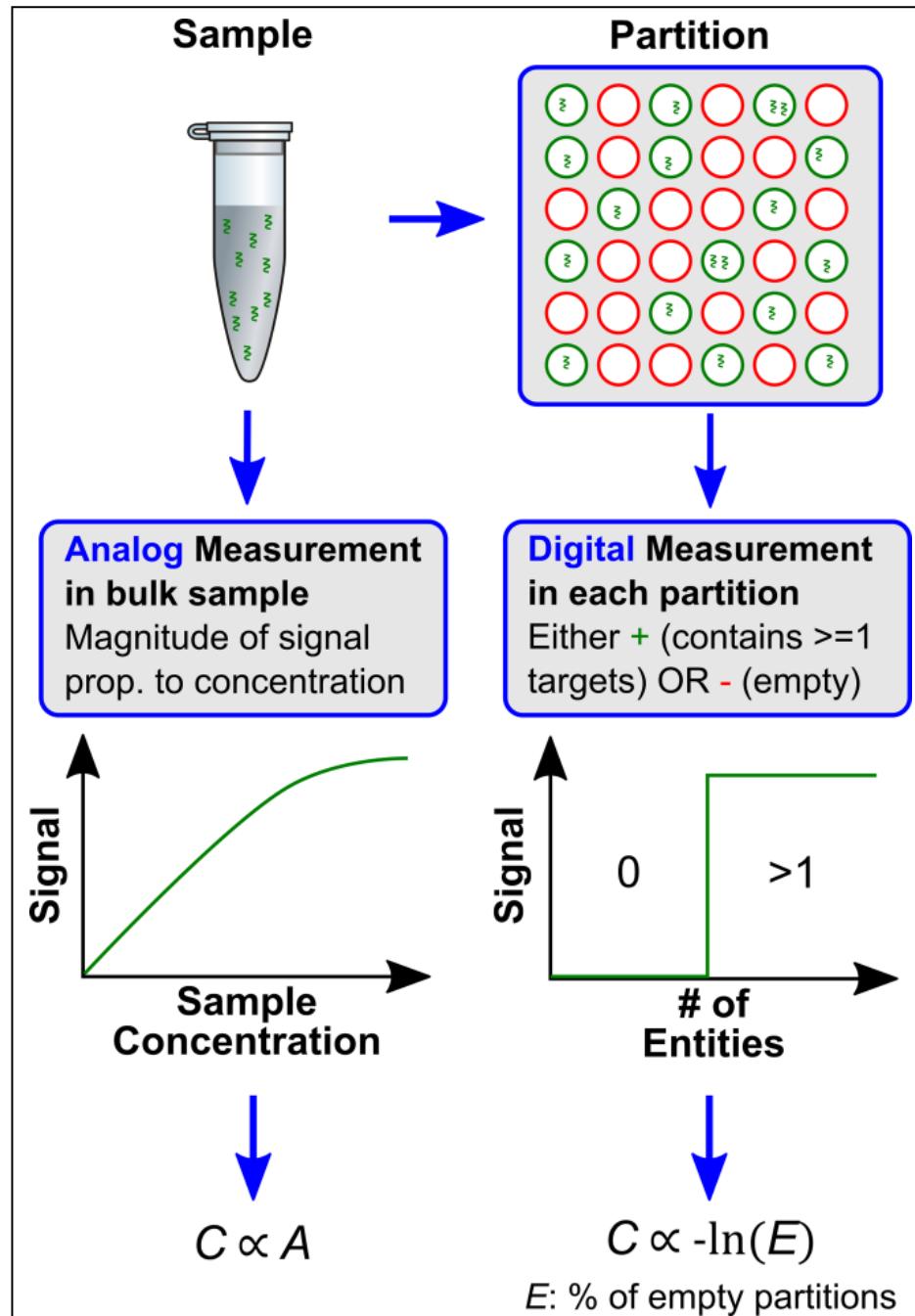
$$\lambda = -\ln \left(\frac{\text{number of valid partitions} - \text{number of positive partitions}}{\text{number of valid partitions}} \right)$$

$$\lambda = -\ln \left(\frac{8000 - 4000}{8000} \right) = 0.693 \text{ Calculated target copies per partition}$$

Given:
Partition volume is $V = 0.23 \text{ nL}$

$$\lambda_{\text{volume}} = \frac{\lambda}{V [\mu\text{l}]}$$

$$\lambda_{\text{volume}} = \frac{0.693}{0.23} \times 1000 = 3013 \text{ copies}/\mu\text{l} \text{ Calculated target copies per } \mu\text{l}$$



Studying Gene Expression

Gene microarrays

Used to study all of the genes expressed in a tissue very fast

Microarray (**gene chip**) is created with use of small glass microscope slide

Single stranded DNA molecules are spotted on a small glass slide using an arrayer (computer controlled robotic arm) which fixes DNA (multiple copies of cDNA) at different spots on the slide which is recorded by a computer

Studying Gene Expression

Gene microarrays

mRNA is extracted from a tissue of interest and cDNA is synthesized from mRNA and labeled with fluorescent dye

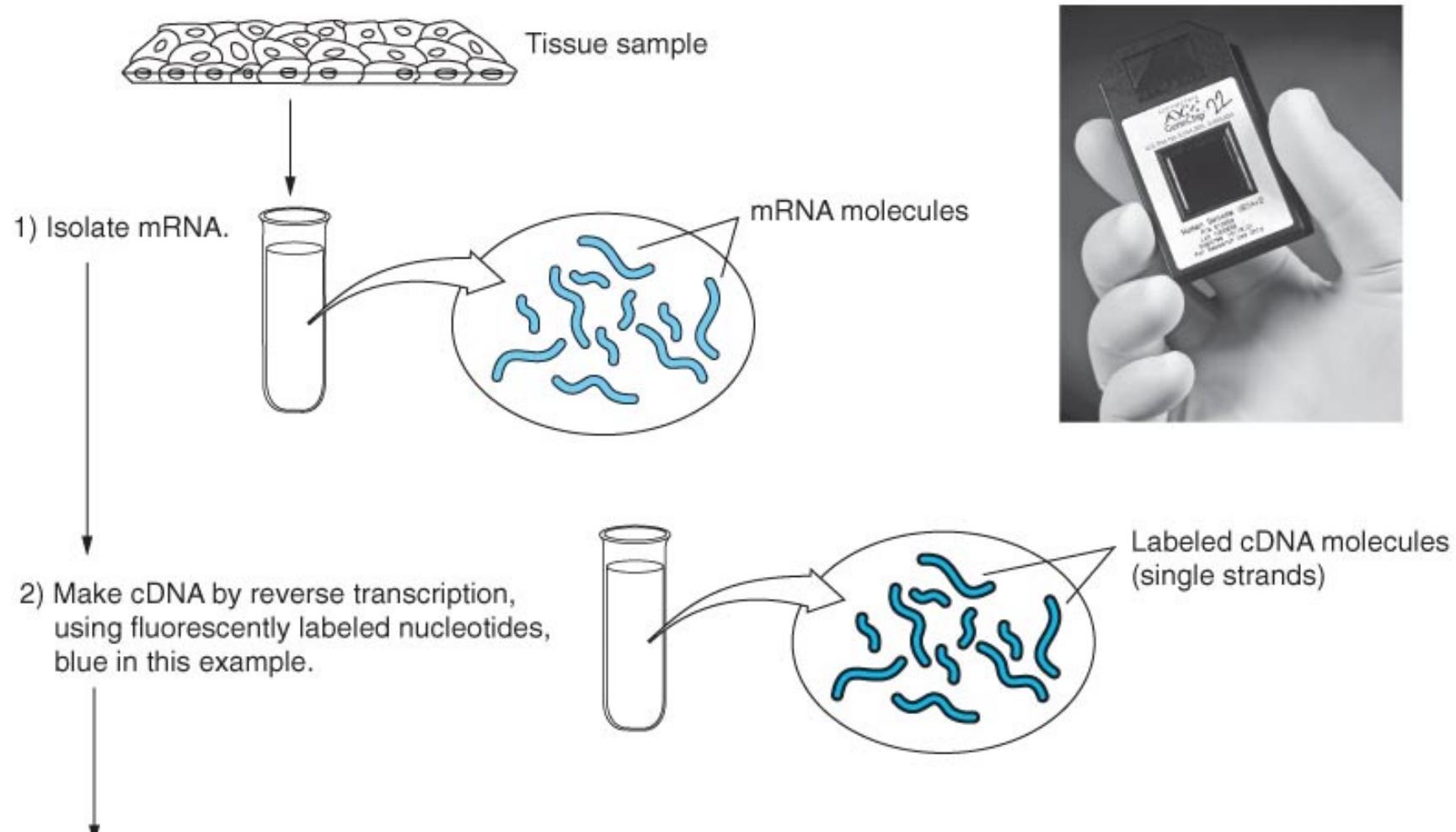
Labeled cDNA is incubated overnight with the array where it hybridizes with different spot on the array that contain complementary DNA sequences

Array is washed and scanned by a laser that causes cDNA hybridized to array to fluoresce

Fluorescent spots reveal which genes were regulated and Intensity of fluorescence indicates relative amount of gene expression

Studying Gene Expression

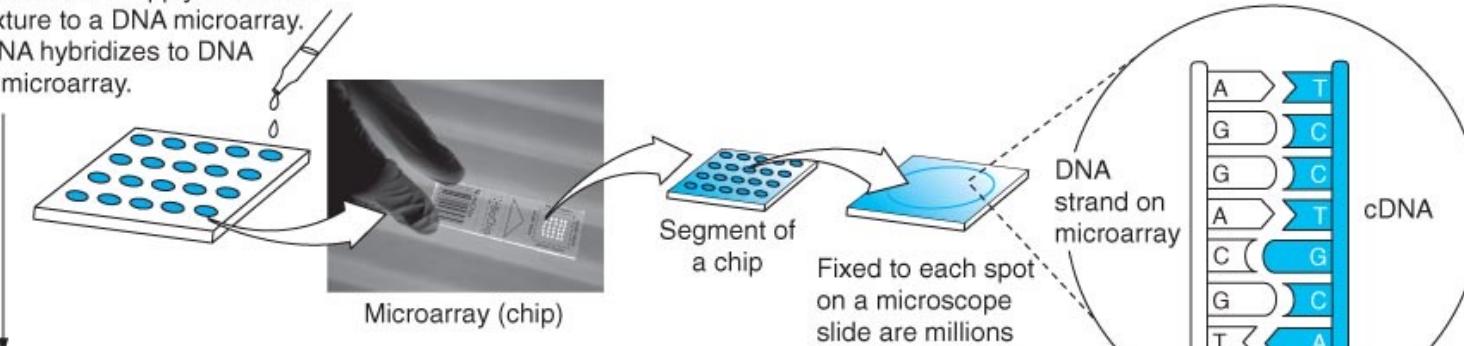
Gene microarrays



Studying Gene Expression

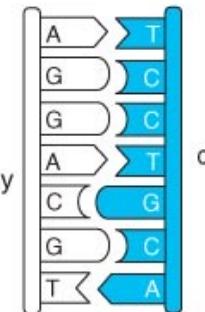
Gene microarrays

3) Hybridization: Apply the cDNA mixture to a DNA microarray. cDNA hybridizes to DNA on microarray.



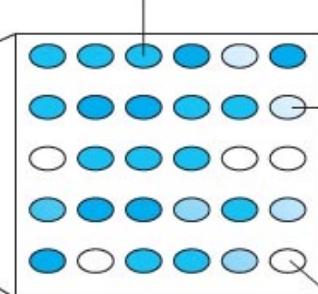
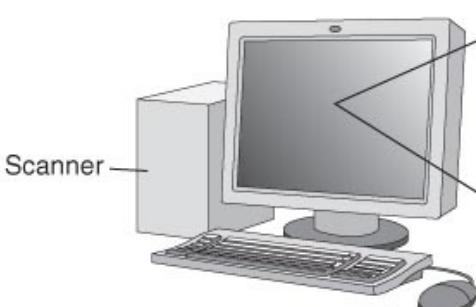
Segment of a chip

Fixed to each spot on a microscope slide are millions of copies of short single-stranded DNA molecules, a different gene or probe in each spot



4) Rinse off excess cDNA, put the microarray in a scanner to measure fluorescence of each spot. Fluorescence intensity indicates the amount of gene expressed in the tissue sample.

Readout



In this image blue spots indicate bright fluorescence and white spots indicate no fluorescence.

Bright fluorescence: highly expressed gene in tissue sample

Moderate fluorescence: low gene expression

Light: no fluorescence, gene not expressed in tissue sample

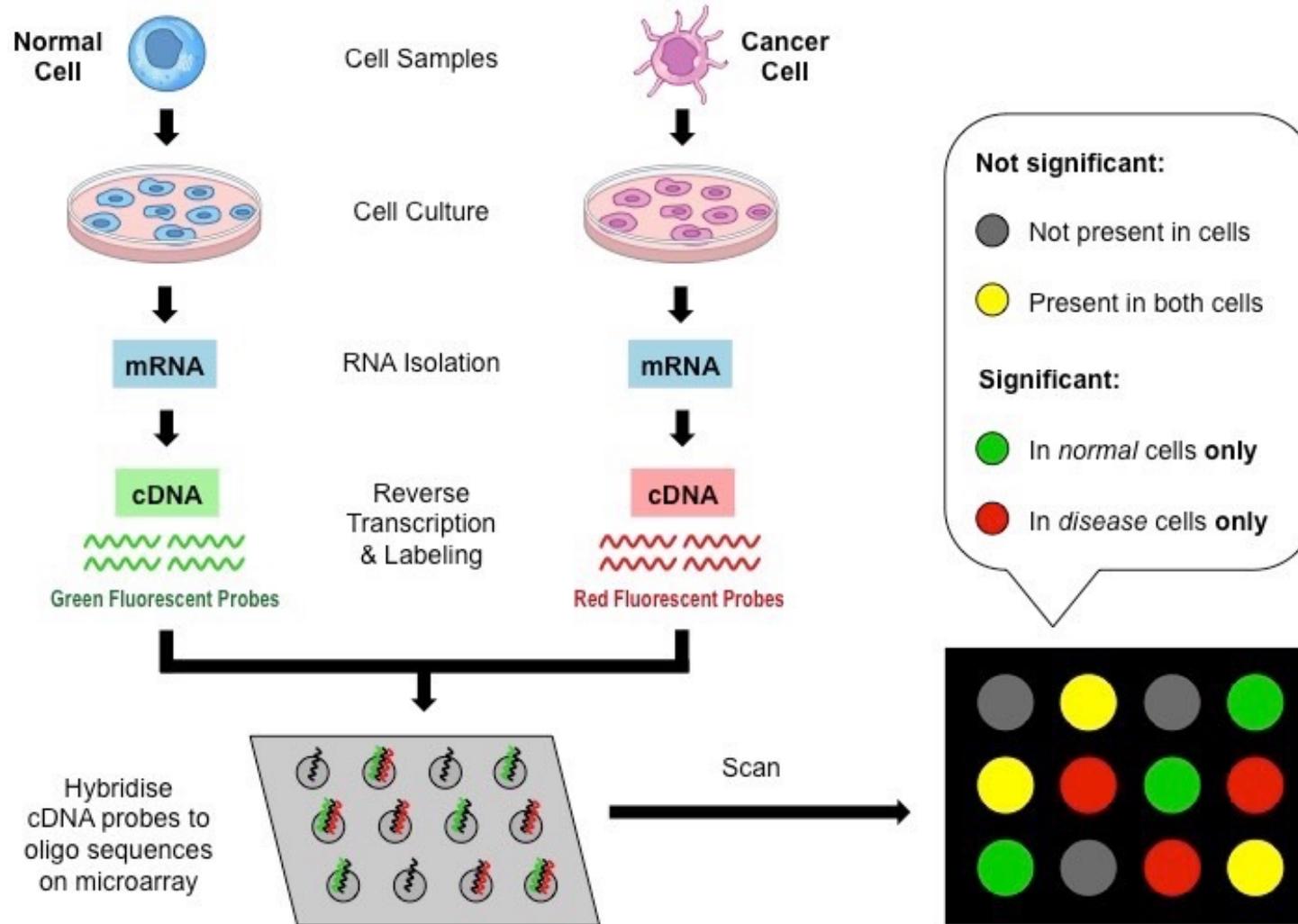
Studying Gene Expression

Gene microarrays

Microarrays are used to compare two different conditions with 2 different colored dyes (one for treatment and other for control condition)

Laser is scanned at different wavelengths for each probe and then the images are overlaid to make direct comparisons between the treatment and control

DNA microarray



RNA interference

RNA interference (RNAi) is naturally occurring mechanism for inhibiting gene expression Procedure:

Double-stranded RNA can be bound by **Dicer** enzyme that cuts dsRNA into 21–25 nucleotide snippets called **small interfering RNA (siRNA)**

siRNA then bound to protein: RNA complex called **RNA induced silencing complex (RISC)**

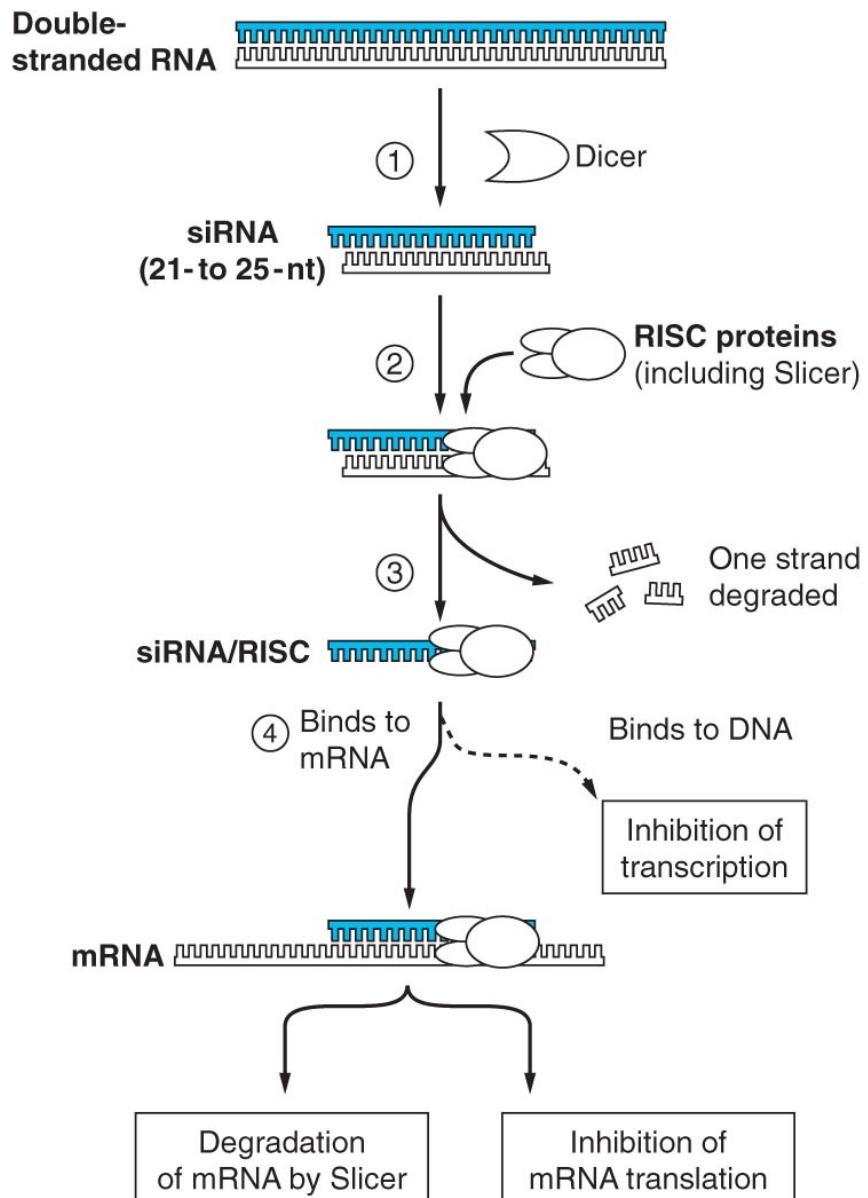
RNA interference

RISC unwinds dsRNA releasing single stranded RNA that bind complementary mRNA

Binding of siRNA to mRNA leads to degradation of mRNA by Slicer enzyme or blocks translation by interfering with ribosome binding

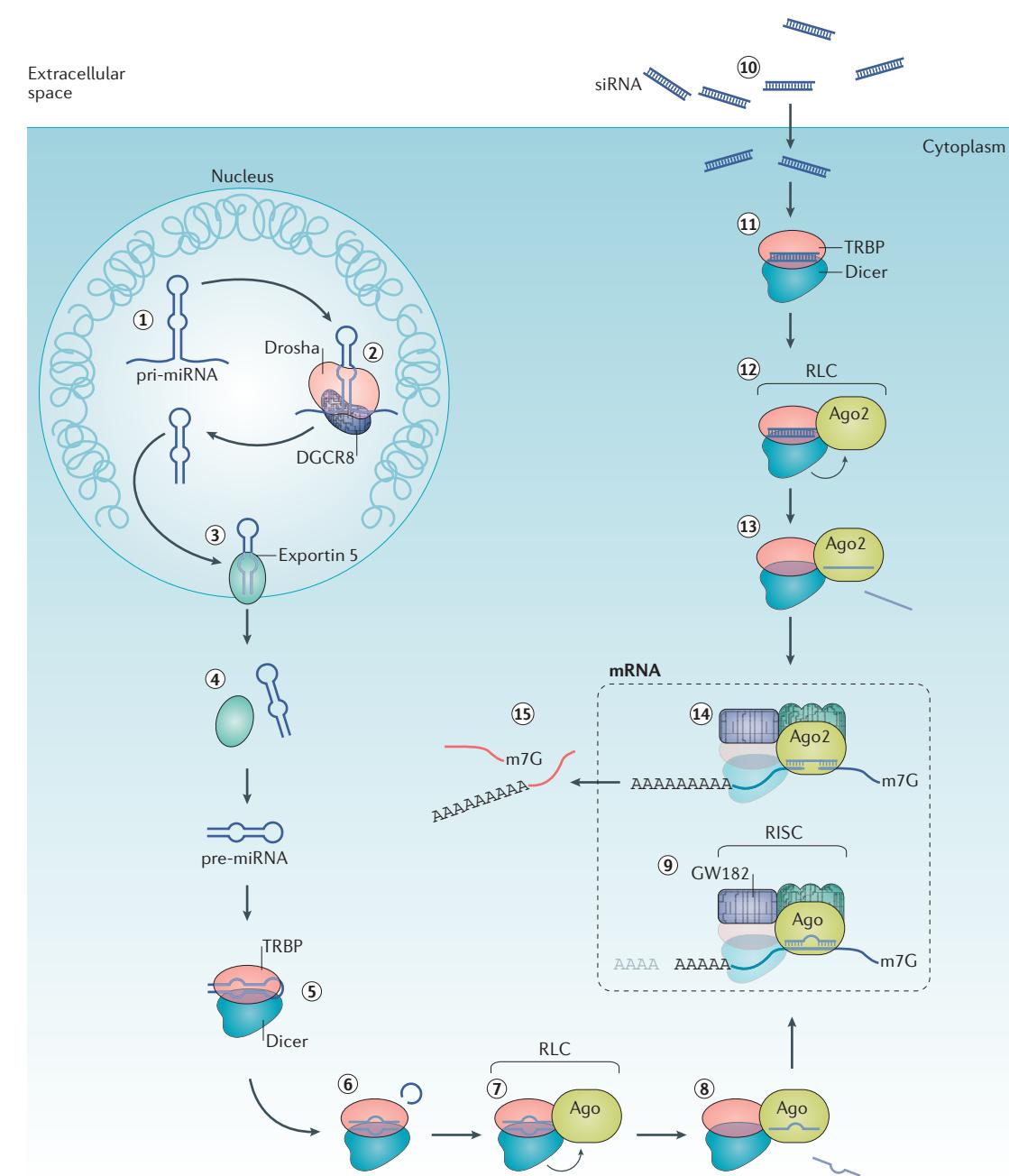
Currently developing techniques to use RNAi to silence gene expression

RNA interference



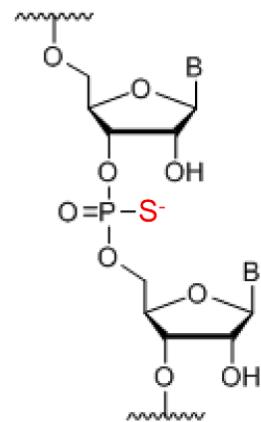
Pathways for mammalian miRNA biogenesis, synthetic RNAi trigger processing and RNAi silencing.

1. Mammalian primary microRNA (miRNA) transcripts (pri-miRNA) are transcribed in the nucleus
2. The pri-miRNA are cleaved by the Microprocessor complex (Drosha–DGCR8) to produce (~30 bp) short hairpin RNAs (shRNAs) called pre-miRNA.
3. Exportin 5 binds and transports the pre-miRNA to the cytoplasm
4. pre-miRNA disengages from exportin 5
5. pre-miRNA binds with Dicer and TAR RNA-binding protein (TRBP)
6. Dicer cleaves the terminal loop of pre-miRNA (7) and induces formation of an RNA-induced silencing complex (RISC)-loading complex (RLC) with an Argonaute (Ago1–Ago4) protein.
7. A guide strand (antisense) is selected²⁹⁵ and loaded into Ago1–Ago4 and the passenger (sense) strand is discarded.
8. The mature RISC can regulate gene expression by inhibiting mRNA translation, inducing mRNA sequestration in cytoplasmic P-bodies and/or GW-bodies, promoting mRNA degradation and directing transcriptional gene silencing of the target gene loci.
9. Argonaute, GW182 and the guide strand are essential for the mRNA-silencing activities of RISC. TRBP and DICER can dissociate from mature RISC after guide strand loading. mRNAs with as few as 7 bases of complementarity to the seed region (bases 2–8 from the 5' end) of guide strands can be affected by RNAi.
10. Synthetic small interfering RNAs (siRNAs) enter the cytosol via endocytosis followed by rare endosomal escape events.
11. siRNAs then interact directly with the cytosolic RNA interference (RNAi) enzymes (Dicer and TRBP)
12. The RLC is formed via Dicer-mediated or non-Dicer-mediated pathways
13. The RLC undergoes strand selection to produce mature RISC.
14. siRNA guide strands usually have full complementarity to a single target mRNA to induce potent and narrowly targeted gene silencing.
15. Ago2 is particularly important for RNAi therapeutics as it has intrinsic slicer activity to efficiently cleave mRNA targets. m7G, 7-methylguanosine.

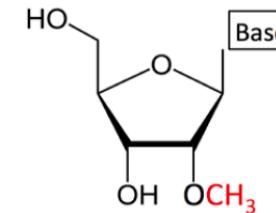


Chemical modifications in backbone and sugar and delivery platforms of the siRNA

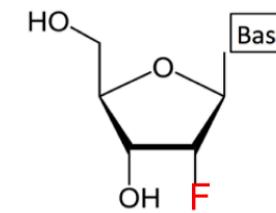
Chemical Modifications



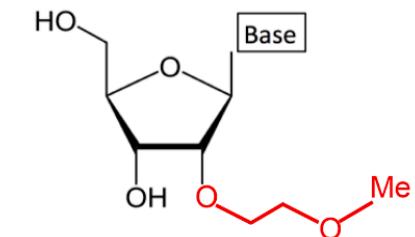
Phosphorothioate (PS)



2-O-methyl (2'-OMe)

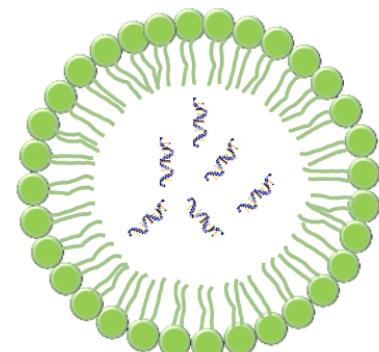


2'-fluoro (2'-F)

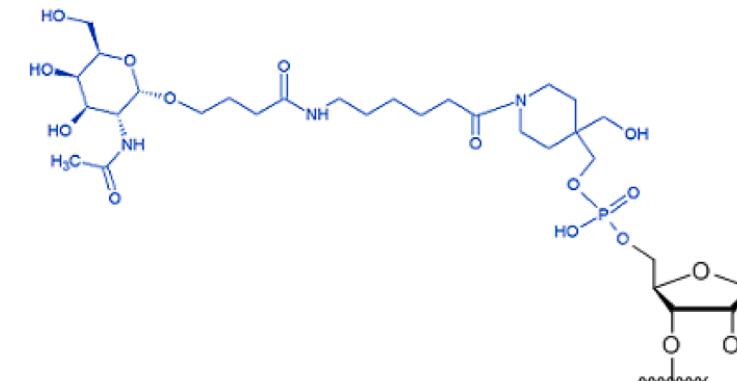


2'-O-methoxyethyl (2'-MOE)

Delivery methods

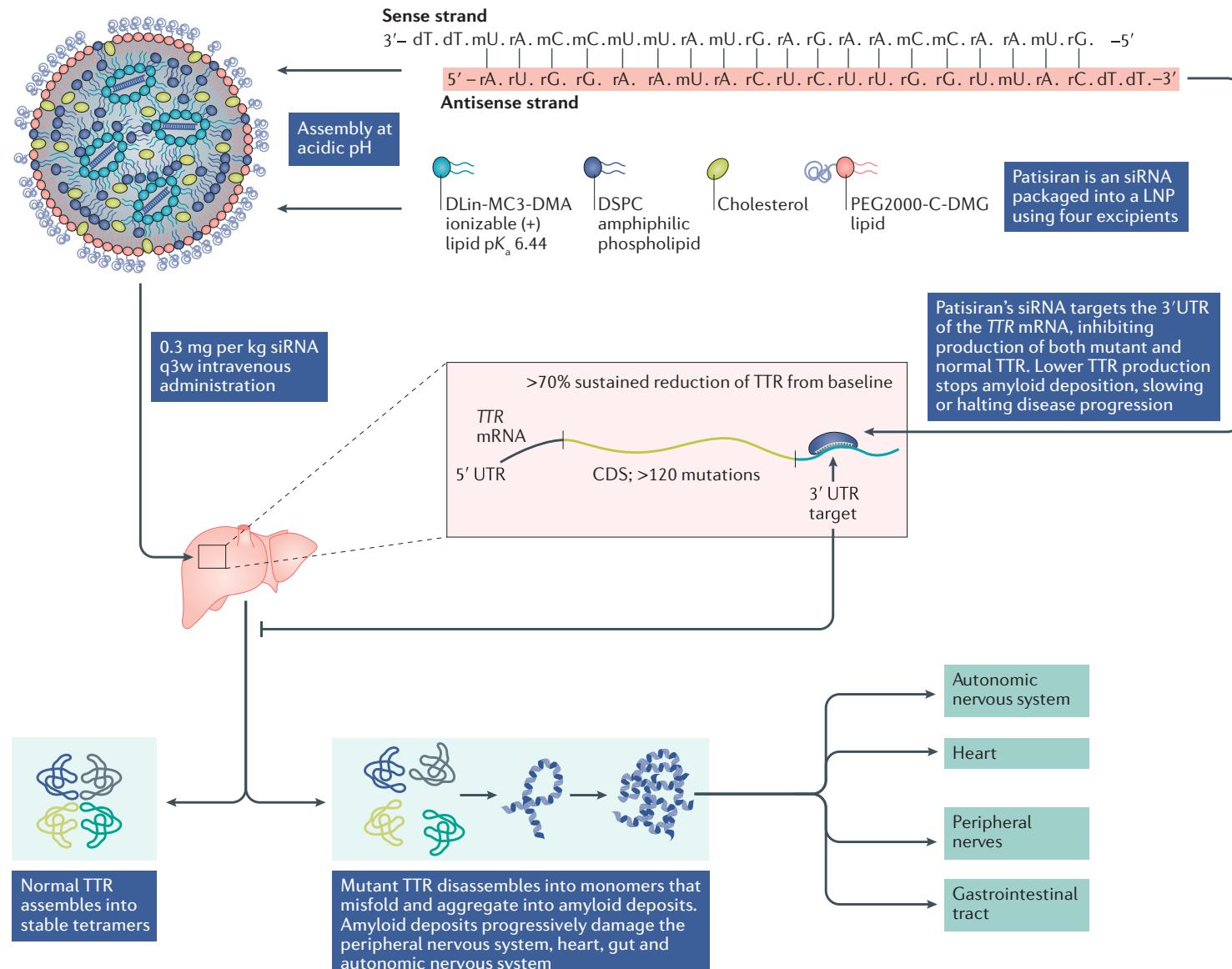


Lipid nanoparticle (LNP)



N-acetylgalactosamine (GalNAc)

ONPATTRO: The first approved RNA interference drug



Gene editing

- Engineered nucleases can be designed to target specific sites in the genome, creating double-strand breaks (DSBs) at desired locations.
- The natural repair mechanisms of the cell repair the break by either homologous recombination (HR) or non-homologous end joining (NHEJ).
- HR is more precise, since it requires a template, allowing the introduction of foreign DNA into the target gene. Homologous DNA “donor sequences” can be used with homology-directed repair (HDR) to introduce a defined new DNA sequence.
- DSB repair by NHEJ is likely to introduce errors such as insertions or deletions (indels), leading to a nonfunctional gene.

Repair of DBS

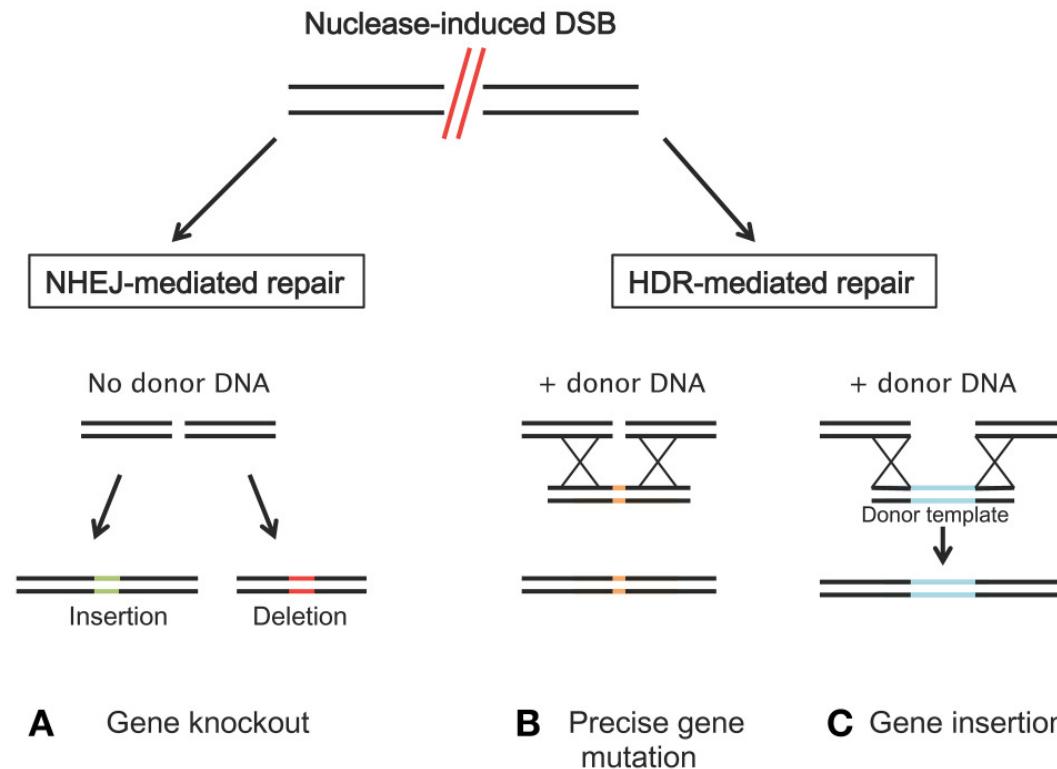
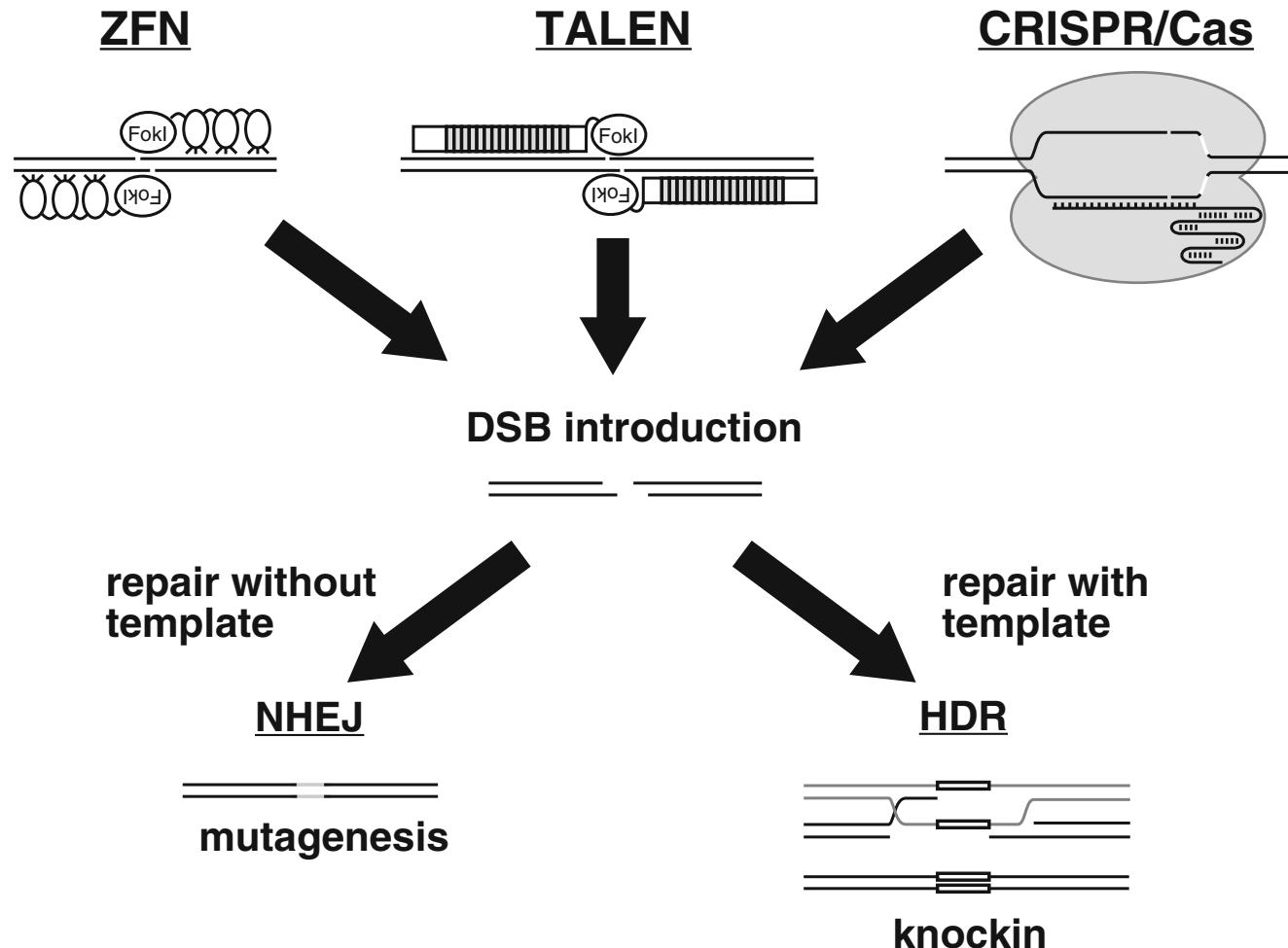


FIGURE 3 | Induced double strand breaks (DSBs) of the target DNA by nucleases can be repaired by either non-homologous end-joining (NHEJ) or homology directed repair (HDR). **(A)** NHEJ usually leads to gene knockout by the insertion (green) or deletion (red) of random base pairs. **(B)** If a donor template, that shares regions of homology to the sequence next to the DSB is available, HDR can introduce precise gene modification or **(C)** specific nucleotide/gene insertion.

Gene editing tools

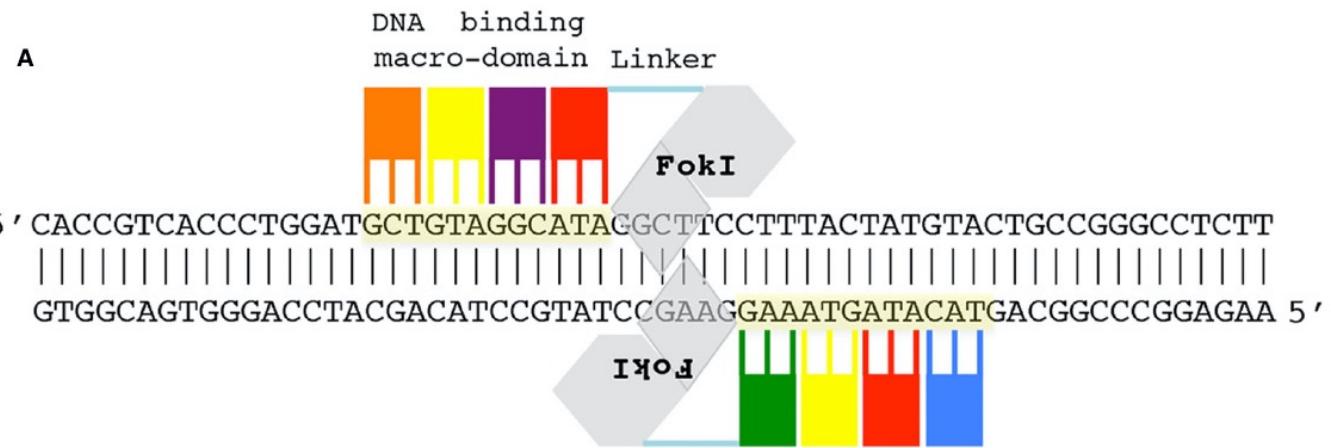
1. Zinc-finger nucleases (ZFNs)
2. Transcription activator-like effector nucleases (TALENs)
3. Clustered regularly interspaced short palindromic repeat (CRISPR)/ CRISPR-associated (Cas)

Gene editing tools

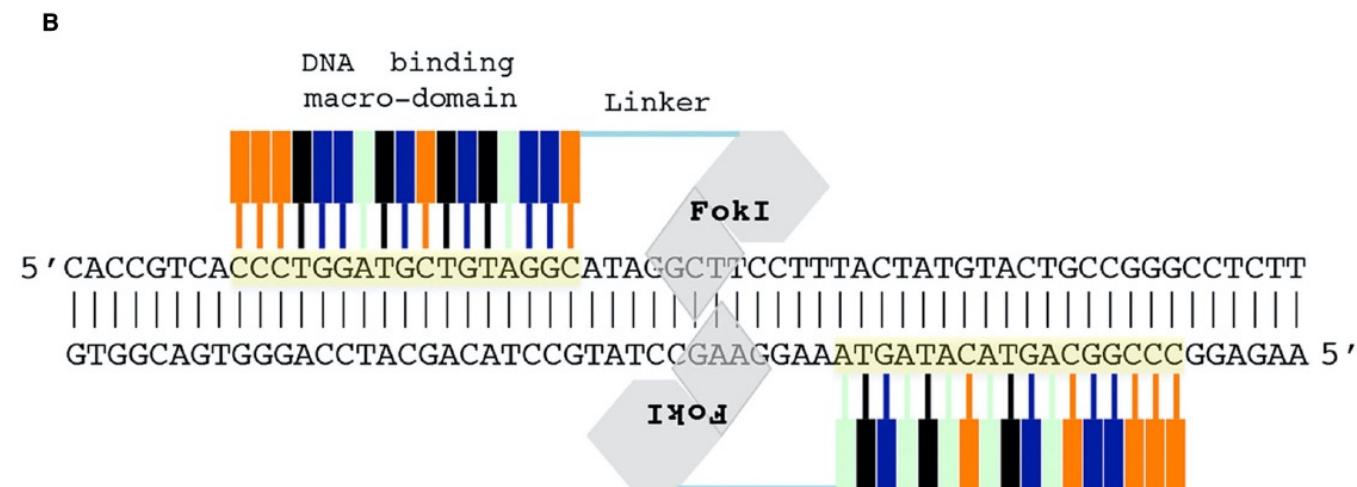


Gene editing tools

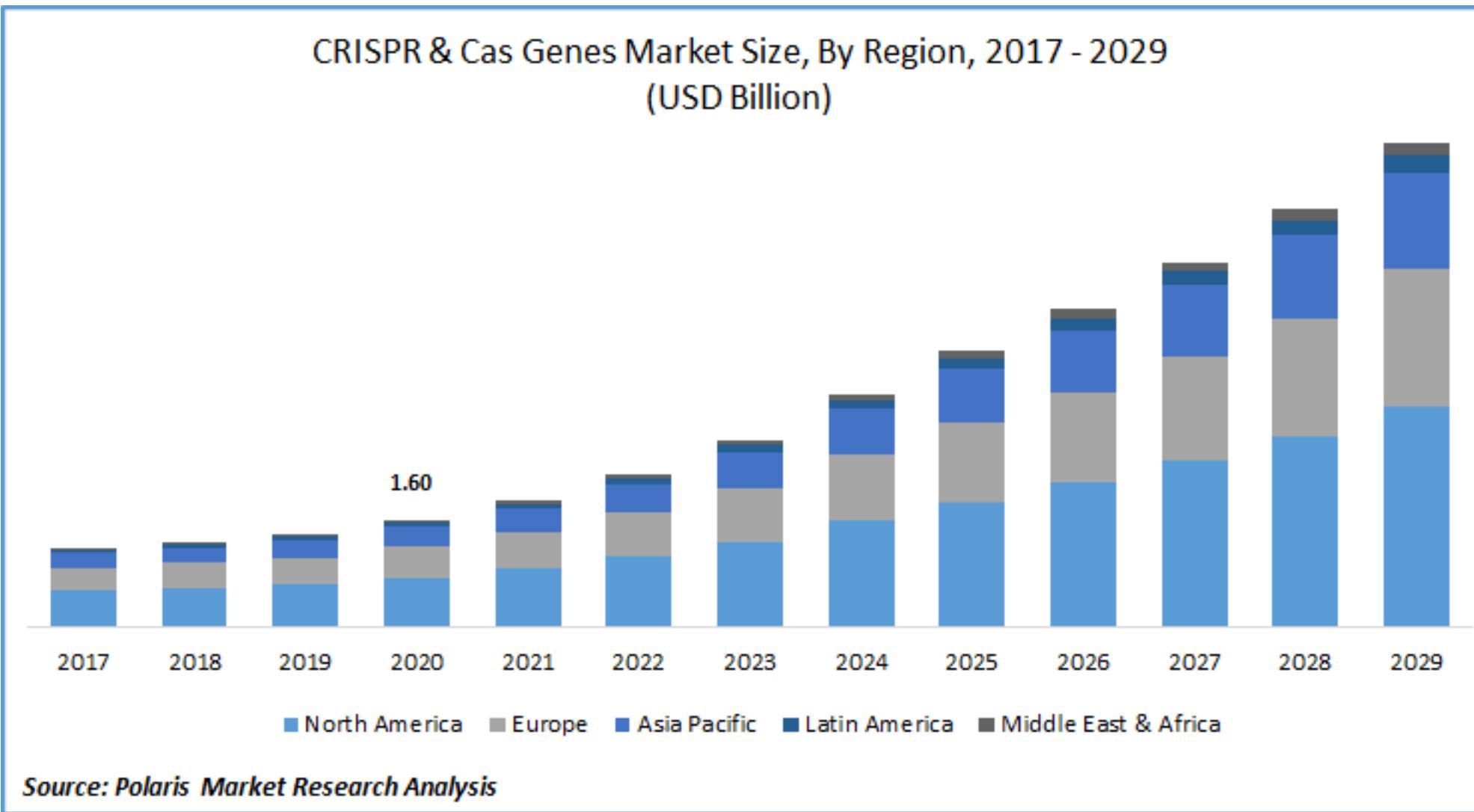
Zinc-finger nucleases (ZFNs)



Transcription activator-like effector nucleases (TALENs)



The rise of CRISPR



CRISPR/Cas system

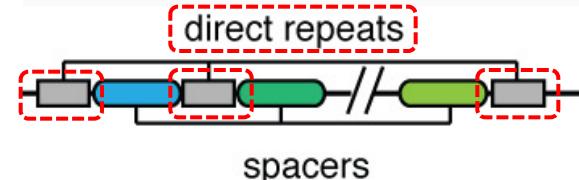
- The guide RNA (gRNA) is fused with the DNA sequence targeting the host gene of interest.
- The gRNA recognizes specific regions on the host RNA and complexes with Cas9, which recognizes the protospacer adjacent motif (PAM) on the target and exerts its endonuclease function to cause double stranded breaks (DSBs).
- This triggers two mechanisms for repair: one is non-homologous end-joining (NHEJ), which introduces mutations in the DSB site.
- The other mechanism is homologous recombination (HR) which enables the donor DNA information to be inserted at the break site.

CRISPR discovery

Identification of genes that are associated with DNA repeats in prokaryotes

Ruud. Jansen,^{1*} Jan. D. A. van Embden,²
Wim. Gaastra¹ and Leo. M. Schouls²

Using *in silico* analysis we studied a novel family of repetitive DNA sequences that is present among both domains of the prokaryotes (Archaea and Bacteria), but absent from eukaryotes or viruses. This family is characterized by direct repeats, varying in size from 21 to 37 bp, interspaced by similarly sized non-repetitive sequences. To appreciate their characteristic structure, we will refer to this family as the clustered regularly interspaced short palindromic repeats (CRISPR). In most species with two or more

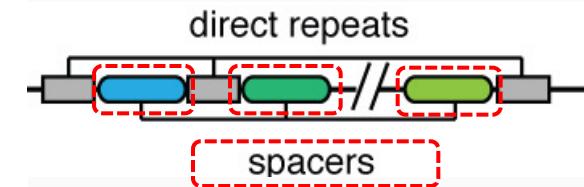


CRISPR discovery

Identification of genes that are associated with DNA repeats in prokaryotes

Ruud. Jansen,^{1*} Jan. D. A. van Embden,²
Wim. Gaastra¹ and Leo. M. Schouls²

Using *in silico* analysis we studied a novel family of repetitive DNA sequences that is present among both domains of the prokaryotes (Archaea and Bacteria), but absent from eukaryotes or viruses. This family is characterized by direct repeats, varying in size from 21 to 37 bp, interspaced by similarly sized non-repetitive sequences. To appreciate their characteristic structure, we will refer to this family as the clustered regularly interspaced short palindromic repeats (CRISPR). In most species with two or more

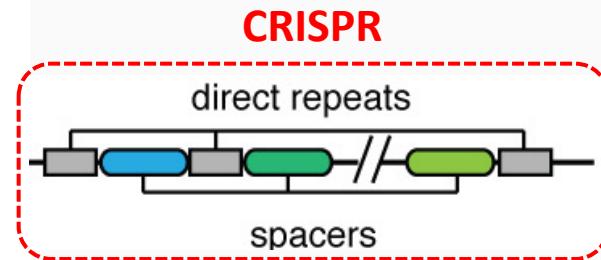


CRISPR discovery

Identification of genes that are associated with DNA repeats in prokaryotes

Ruud. Jansen,^{1*} Jan. D. A. van Embden,²
Wim. Gaastra¹ and Leo. M. Schouls²

Using *in silico* analysis we studied a novel family of repetitive DNA sequences that is present among both domains of the prokaryotes (Archaea and Bacteria), but absent from eukaryotes or viruses. This family is characterized by direct repeats, varying in size from 21 to 37 bp, interspaced by similarly sized non-repetitive sequences. To appreciate their characteristic structure, we will refer to this family as the clustered regularly interspaced short palindromic repeats (CRISPR). In most species with two or more



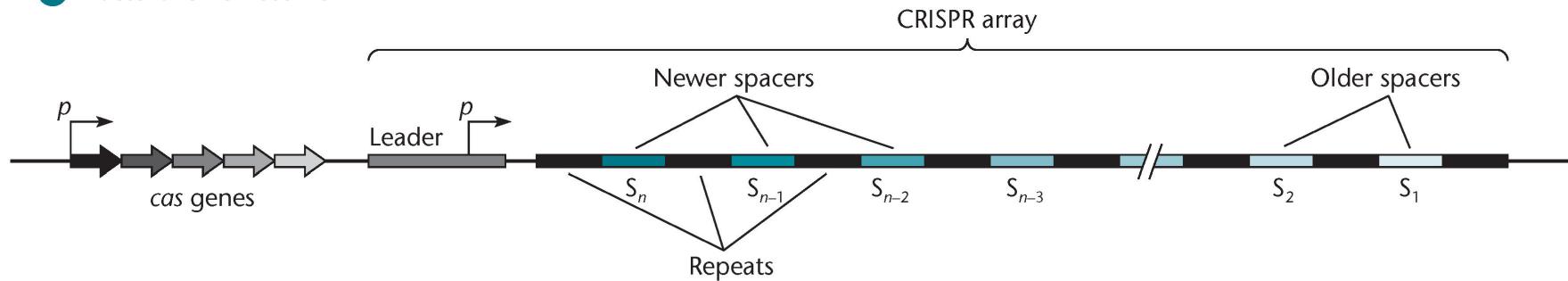
CRISPR/Cas9 glossary

CRISPR	Clustered Regularly Interspaced Short Palindromic Repeats - CRISPR refers to prokaryotic DNA elements involved in adaptive immunity which are characterized by clusters of identical repeats interspaced with non-identical segments called spacers. CRISPR has evolved to refer more generally to the use of Cas9 for genome editing.
DSB	Double Strand Breaks result from cleavage of both strands of DNA. This can be achieved through the use of wild type Cas9 or by employing two Cas9 nickases targeting opposite DNA strands.
gRNA	Guide RNAs bind to Cas9 and direct the complex to a specific genomic location. Naturally occurring guide RNAs consist of two parts: a CRISPR RNA (crRNA) and a trans-activating crRNA (tracrRNA). Alternatively, the crRNA and tracrRNA can be combined into a single chimeric oligonucleotide called a single guide RNA (sgRNA).
HDR	Homology Directed Repair is a mechanism of DNA repair that uses a homologous DNA template to rebuild sites of genomic damage. HDR can be leveraged in genome editing experiments to make precise genomic alterations by supplying the desired sequence for insertion flanked by segments of DNA that are homologous to the sequence surrounding the Cas9-induced DSB.
indel	(Insertion/Deletion) Following creation of a DSB by Cas9, the cell initiates DNA damage repair. Repair by the error-prone NHEJ pathway can result in small insertions and/or deletions at the site of cleavage. These indels can cause frameshift mutations or premature stop codons resulting in a genetic knock-out.
NHEJ	Non-Homologous End-Joining is the predominant DNA DSB repair mechanism in mammalian cells. Unlike HDR, NHEJ directly ligates the ends of the DSB and does not require a homologous repair template. Researchers capitalize on the error-prone nature of NHEJ to create indels following targeted cleavage with Cas9.
PAM	Protospacer-Adjacent Motif - In the naturally occurring prokaryotic CRISPR/Cas system, the DNA sequences recognized by gRNA are called protospacers. The PAM is a short sequence next to the target site that is required for Cas9 targeting both in prokaryotic adaptive immunity and in mammalian genome editing experiments.
sgRNA	Single Guide RNA , a chimeric RNA composed of crRNA and tracrRNA, connected by a short RNA linker
tracrRNA	Trans-Activating crRNA , one of two RNAs required to form a functional gRNA. The tracrRNA forms base pairs with the crRNA and is required for Cas9-mediated target cleavage.
Target sequence	A 20 nucleotide genomic DNA sequence which base-pairs with gRNA and is cleaved by Cas9.
Cas9	CRISPR Associated Protein 9 - Cas9 is an RNA-guided DNA endonuclease from the type Cas9 II CRISPR system of <i>Streptococcus pyogenes</i> that has been adapted for use in genome editing applications.
crRNA	CRISPR RNA - One of two RNAs required to form a functional gRNA. The crRNA contains crRNA the sequence complementary to the DNA target and a segment of RNA that base pairs with tracrRNA

CRISPR: The Adaptive Phase

A CRISPR-associated elements

1 Bacterial chromosome



CRISPR (clustered regularly interspaced short palindromic repeats)

CRISPR array: one to hundreds of repeats

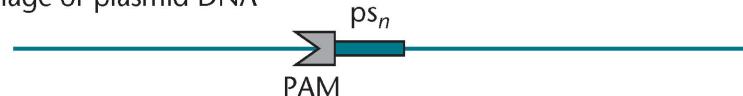
Repeat: 21–47 bp (CAGTTCCCCGCCAGCGGGATAAACCG)

Spacer sample: ~30 bp (CTTCGCAGACGCCGCGGATACGCTACGCA)_n

Leader: Several hundred noncoding base pairs

cas genes (cascade): cas complex for CRISPR-associated antivirus defense; acquisition of new spacers; immunity, e.g., processing pre-crRNAs

2 Phage or plasmid DNA

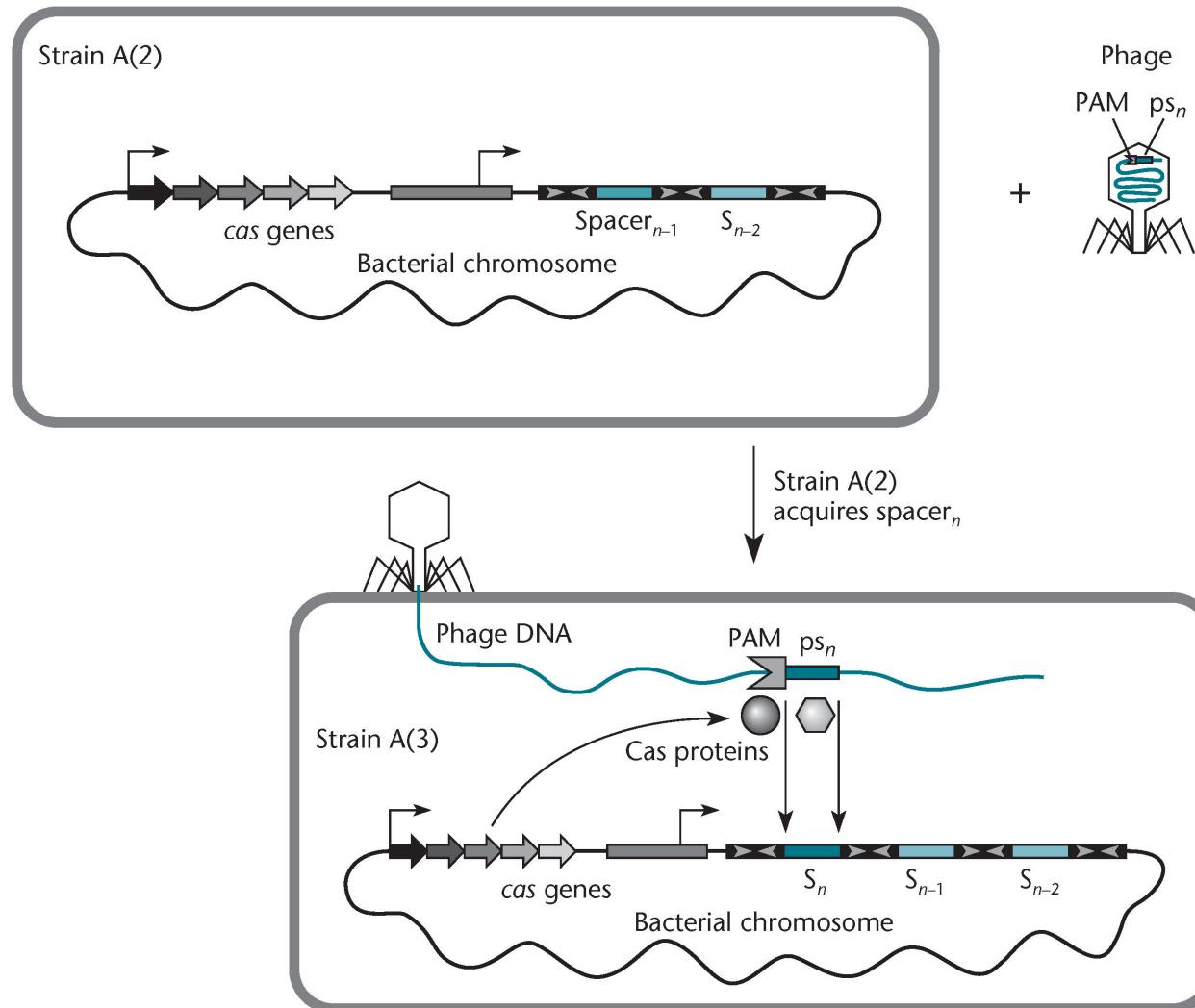


PAM: protospacer-adjacent motif (example) (CT/AT)

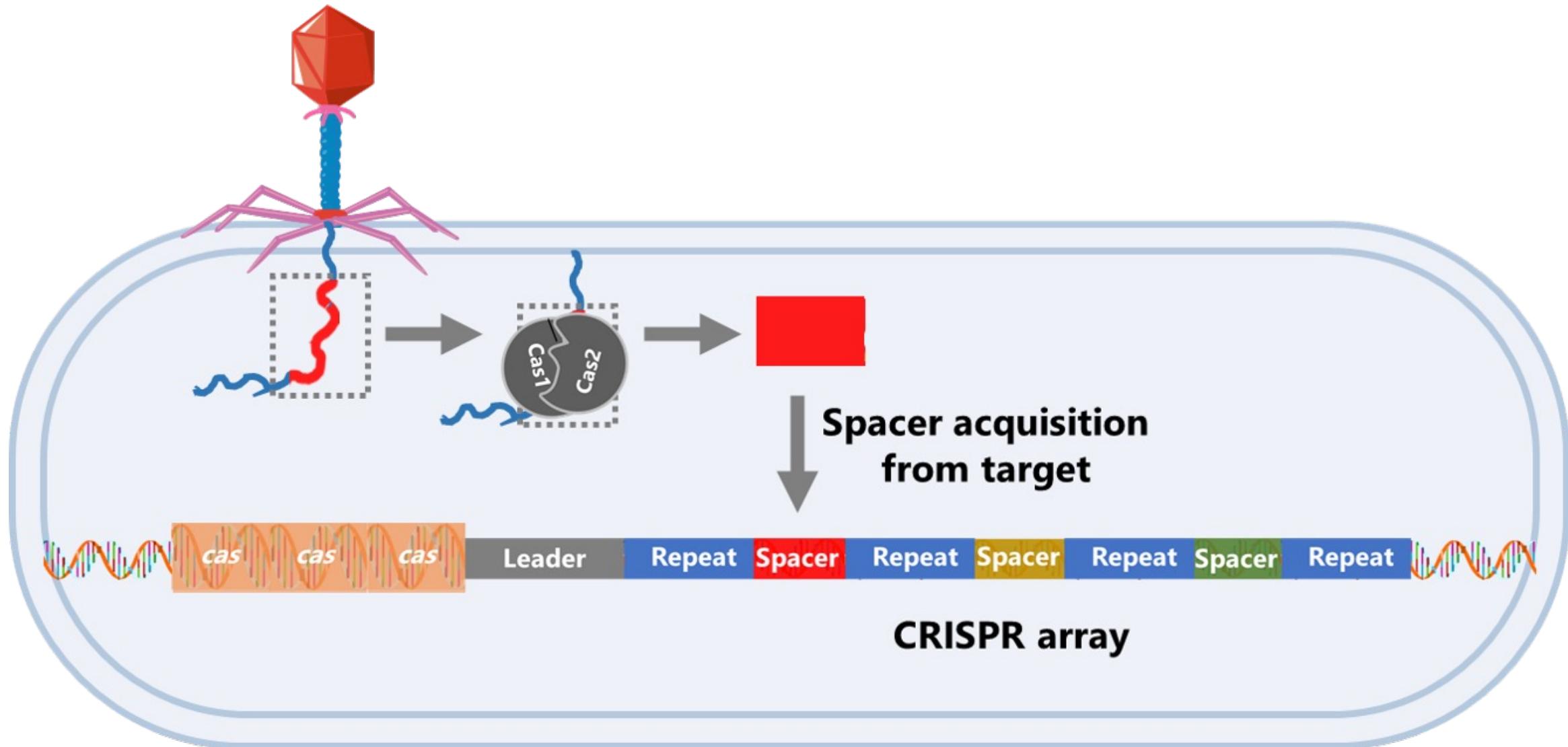
ps_n: proto spacer (CTTCGCAGACGCCGCGGATACGCTACGCA)_n

CRISPR: The Adaptive Phase

B Construction of CRISPR array: adaptive phase



CRISPR: The Adaptive Phase

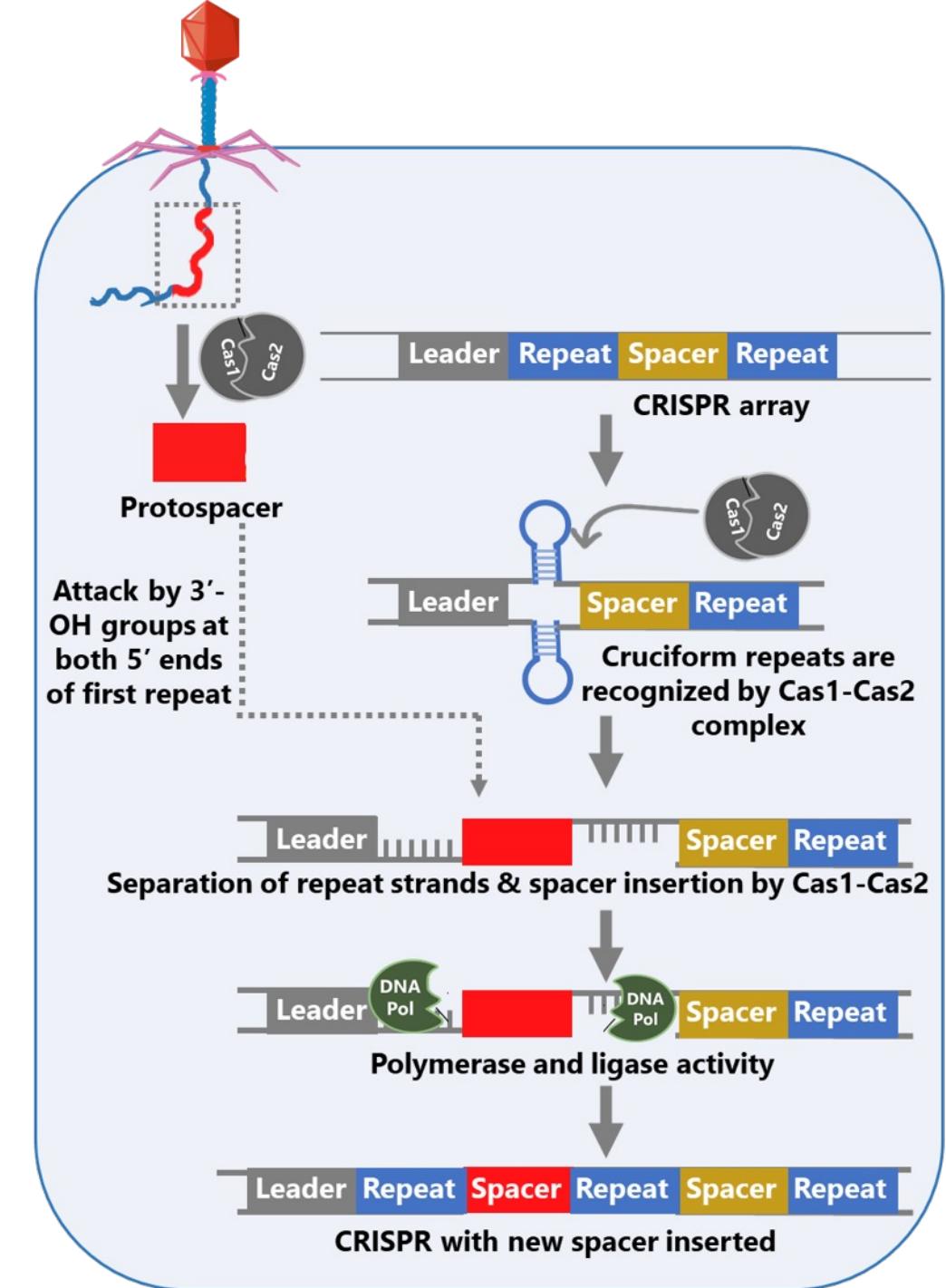


CRISPR: The Adaptive Phase

- The first step for CRISPR-Cas mediated defense is known as adaptation or acquisition.
- This step is required to provide genetic memory to the bacterial cell.
- In this step, the invading phage's genetic material is incorporated into the CRISPR locus, thus providing the bacteria with a way of recognizing and adjusting its response to further invasions by the same phage.
- This process begins when a protein complex called **Cas1** and **Cas2** identifies the invading viral DNA and excises a segment of a specific length from the viral DNA. This segment of DNA is known as the **protospacer** or simply **spacer**.
- Then the protospacer is added into the front of the CRISPR array between the two repeat sequences
- By this mechanism, bacteria generate the embedded memory of the invading virus.

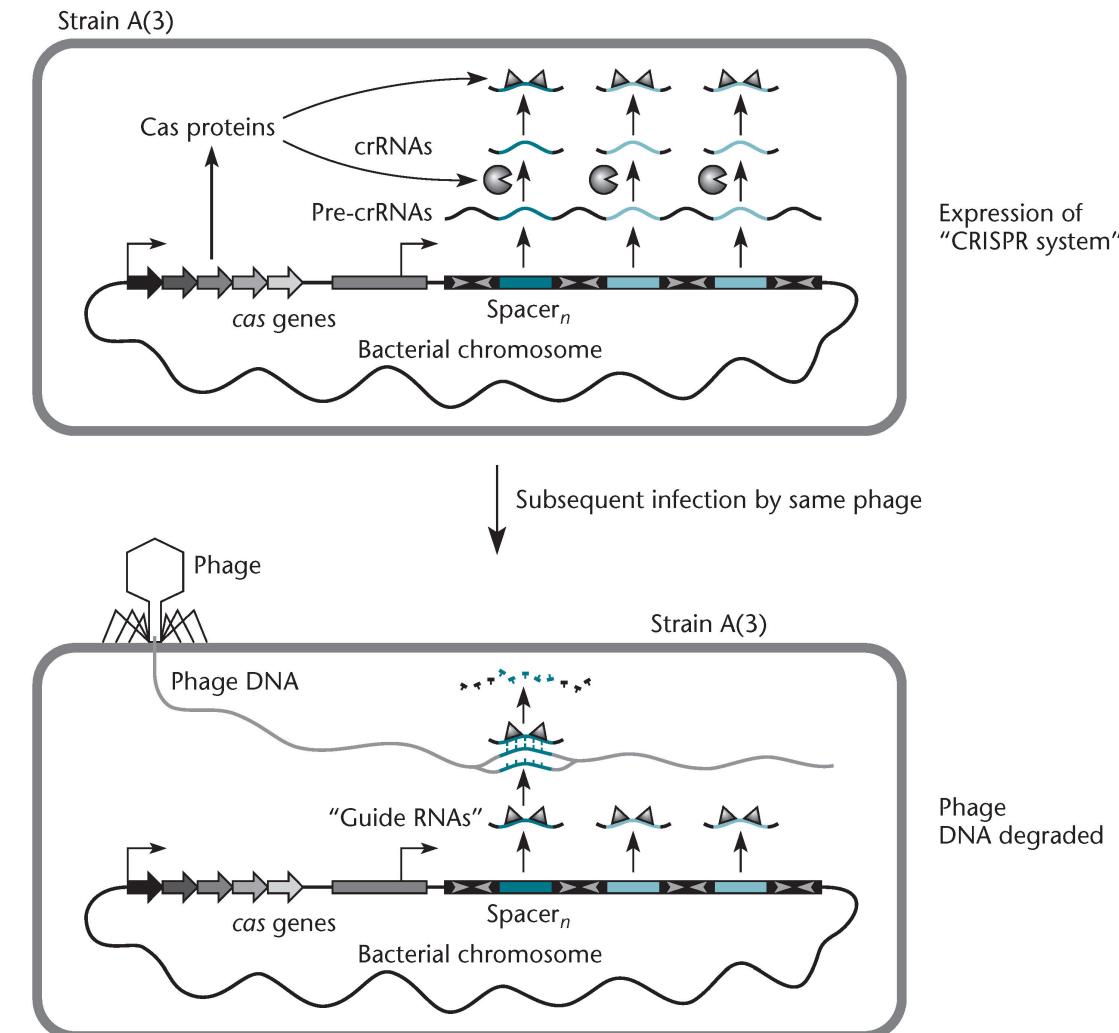
CRISPR

The Adaptive Phase

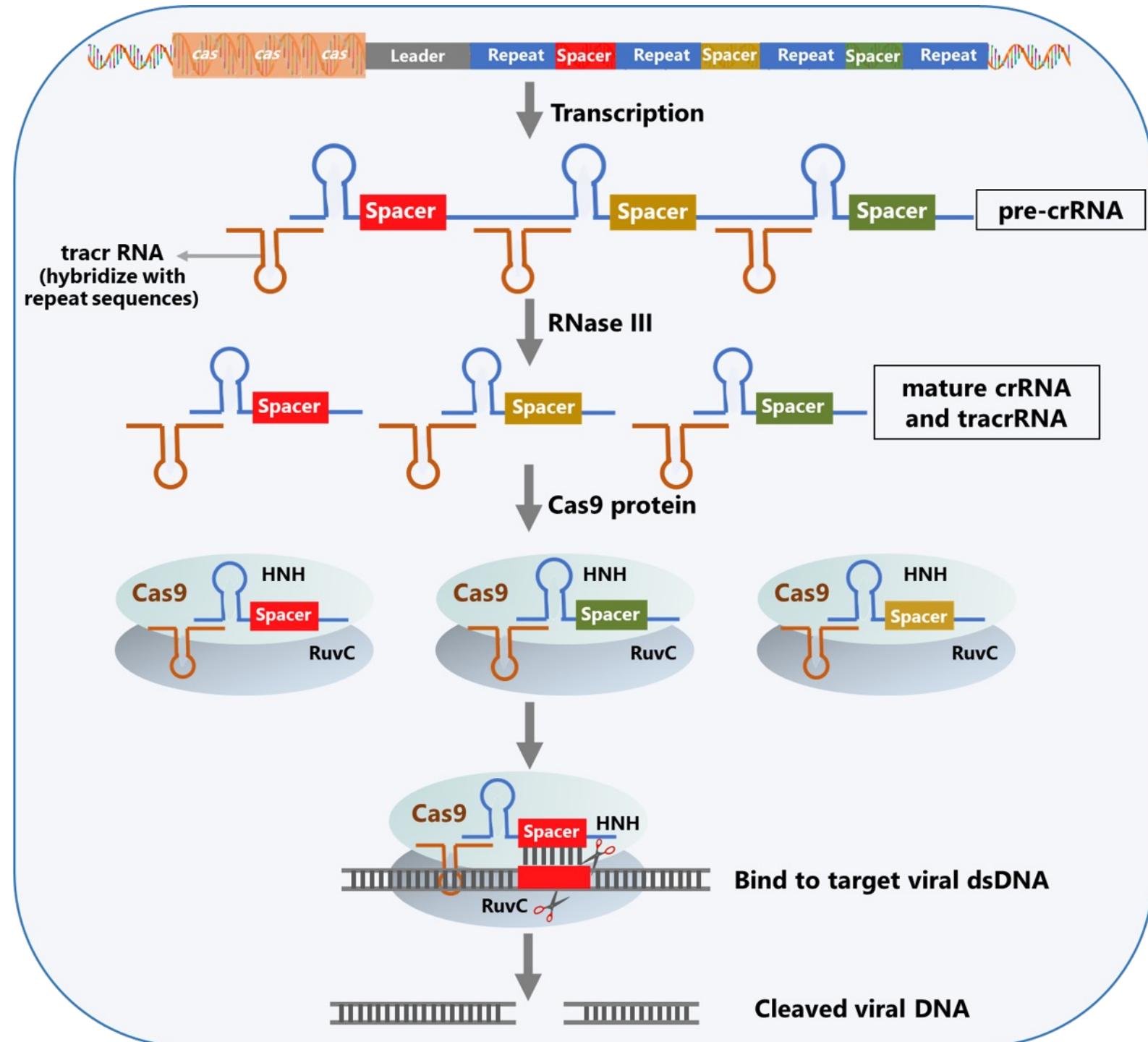


CRISPR: The Immunity Phase

C Roles of CRISPR functions in defense: immunity phase



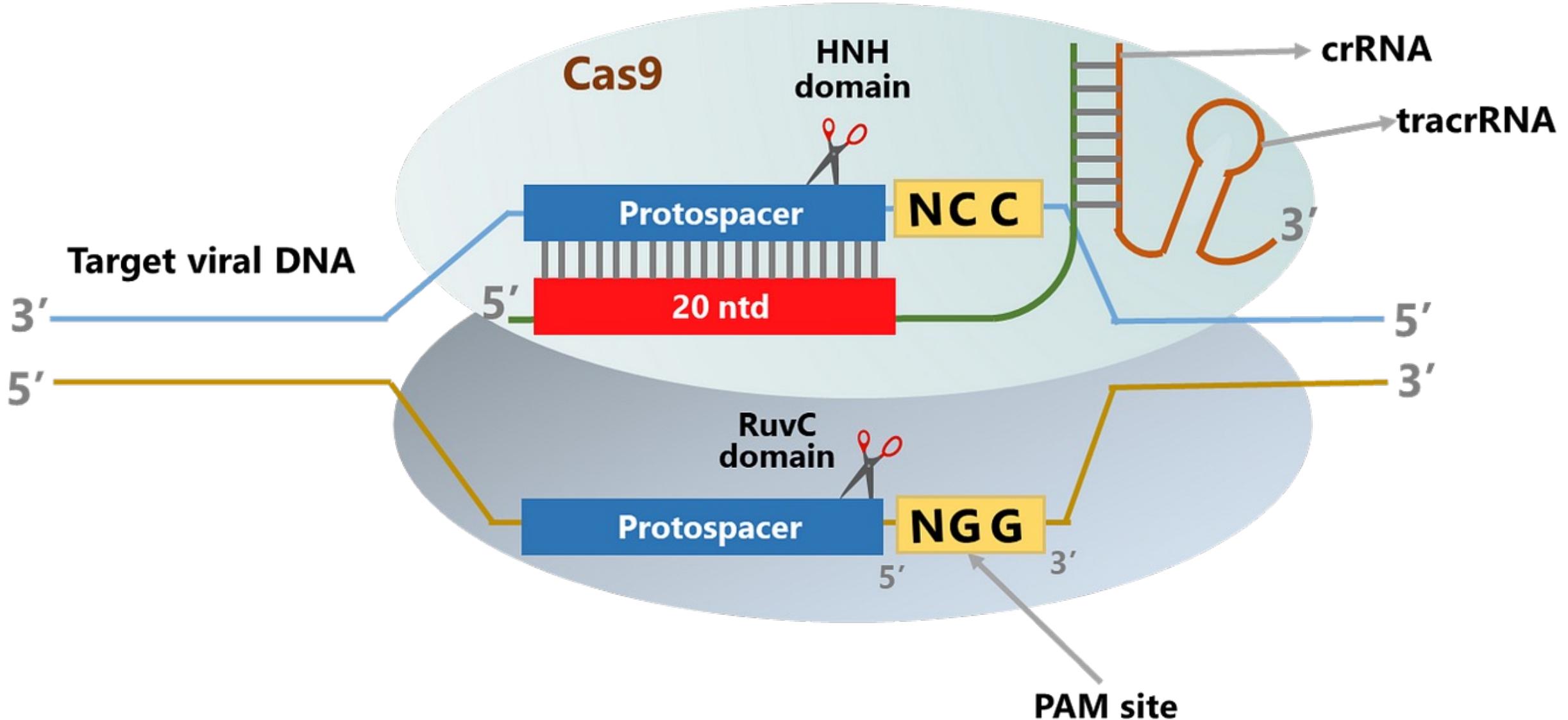
CRISPR The Immunity Phase



CRISPR: The Adaptive Phase

- To integrate protospacers (spacers) into the CRISPR array, leader sequence and leader proximal repeat are required.
- The spacer acquisition is usually directional, with new spacers added to the leader-proximal end of the array.
- One new repeat is generated because the protospacer is added into the front of the CRISPR array between two repeats. For this, the leader-proximal repeat serves as a template for the synthesis of the new repeat.
- The palindromic repeats form a **cruciform** structure that recruits **Cas1** and **Cas2** proteins that mediate protospacer DNA integration.
- Cas1 and Cas2 proteins have both **nuclease** and **integrase** activity. The nuclease activity mediates the excision of protospacer DNA from the viral DNA. The integrase activity of the Cas1-Cas2 complex mediates the integration of protospacer DNA into the CRISPR array.
- The 3'-OH groups of the protospacer attacks at both 5' ends of the first repeat of the CRISPR array. The result is an expanded CRISPR array with a new spacer in between two incomplete ssDNA repeats.
- The polymerase enzyme then generates the new complementary strands to generate dsDNA repeats, and the ligase enzyme seals the nicks. Thus, the protospacer is inserted between two repeats, out of which one is a newly synthesized repeat.

CRISPR & gene editing

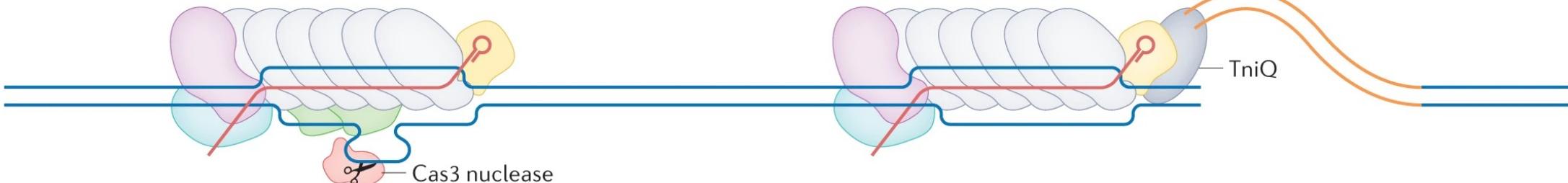


Classification of the CRISPR system

- The CRISPR-Cas system is functionally divided into two classes
- **Class I**
 - Characterized by multisubunits of effector nuclease complexes
 - Routine applications of the class as a genome editing tool have been limited because of not only limited knowledge but also restrictions in the cloning of the system in a functional vector or in the use of a ribonucleoprotein protein (RNP) complex.
- **Class 2**
 - Consists of a single effector protein
 - Subclassified into types II, V, and VI according to the factors necessary for pre-crRNA processing and the diversity of the domains constituting the effector protein.
 - Each type shows different specificity for nucleotide substrates and PAM, cleavage pattern,
 - Suitable tool for tailored and fine-tuned genome engineering

Natural Cas systems

Class I



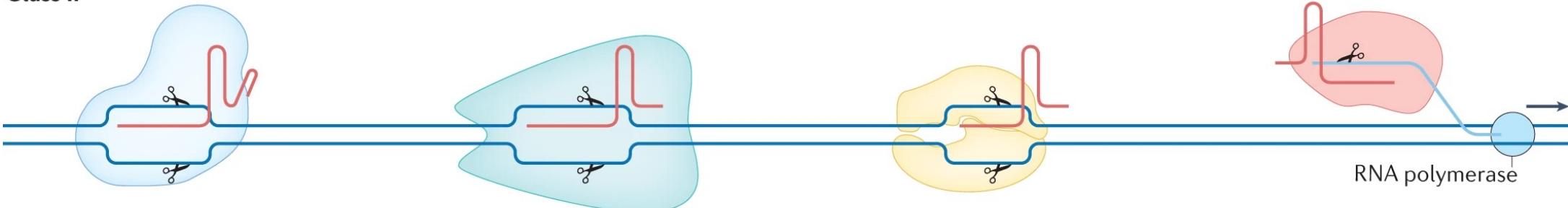
a Type I-E (Cascade)

- Cuts dsDNA
- Multimer of Cas3, Cas5, Cas6, Cas7, Cas8 and Cas11
- Longer crRNA = higher specificity

b Type I-F

- Inserts dsDNA
- Multimer of Cas6, Cas7, Cas5-8 and TniQ
- Requires other transposon machinery

Class II



c Type II-A (Cas9)

- Cuts dsDNA
- Can utilize sgRNA
- High GC PAM
- First in human

d Type V-A (Cas12a)

- Cuts dsDNA
- Can process its own crRNA
- High AT PAM

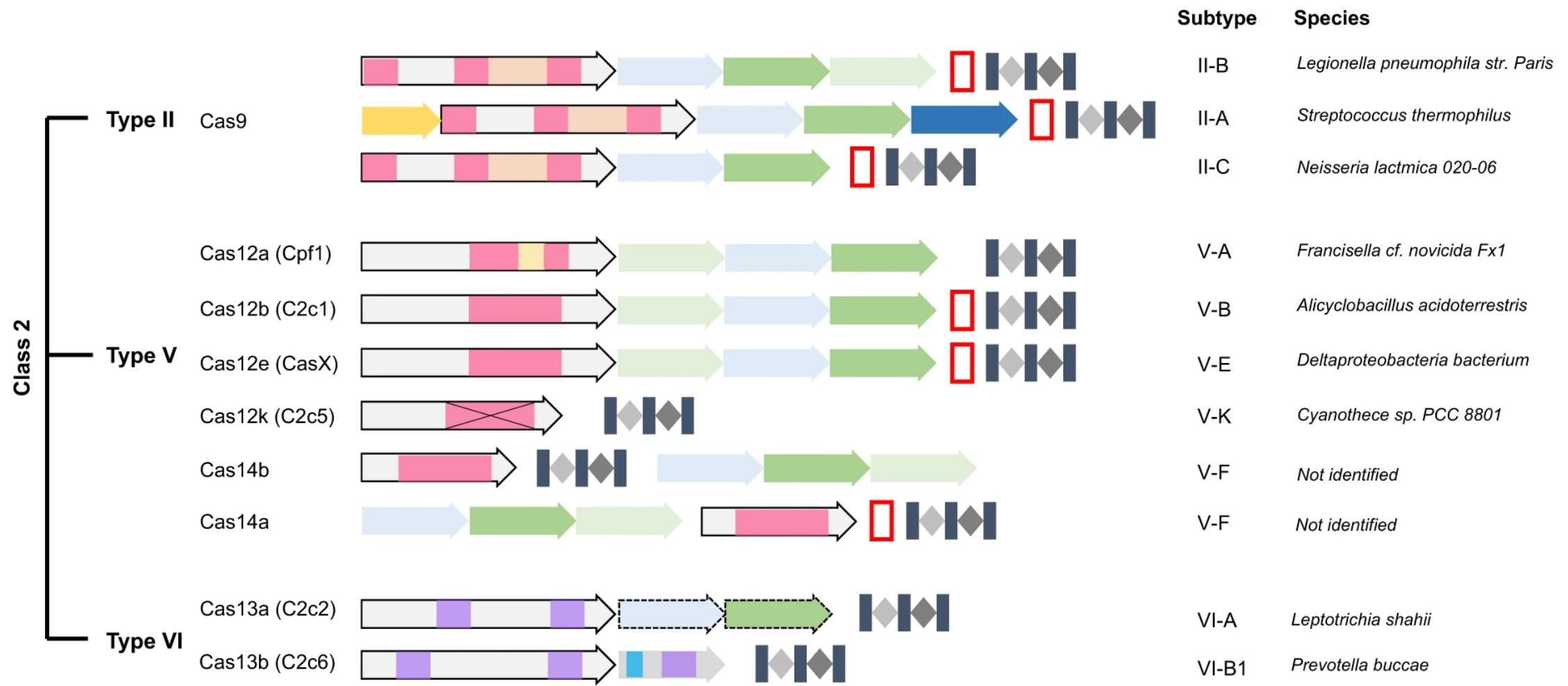
e Type V-F (Cas12f)

- Cuts dsDNA and ssDNA
- Hyper compact
- High AT PAM

f Type VI-D (Cas13d)

- Cuts ssRNA
- No PAM requirement
- Edits the transcriptome

Type II CRISPR systems



Gene: cas1 cas2 cas4 csn2 csx28 effector RNase III

Domain : RuvC HNH Nuc HEPN TM

Guide RNA : tracrRNA CRISPR array

Functional modules and core genes

- All cas genes can be subdivided into four distinct, although partially overlapping, functional modules:

The adaptation module

- Includes the gene encoding the key enzyme involved in spacer insertion (the Cas1 integrase) and the structural subunit of the adaptation complex Cas2, as well as the Cas4 nuclease in several CRISPR–Cas subtypes, the Csn2 protein in sub-type II-A.

The expression processing module

- Responsible for pre-crRNA processing. In most class 1 systems, Cas6 is the enzyme that is directly responsible for processing. In type II systems, processing is catalyzed by the bacterial RNase III (a non-Cas protein)

Functional modules and core genes

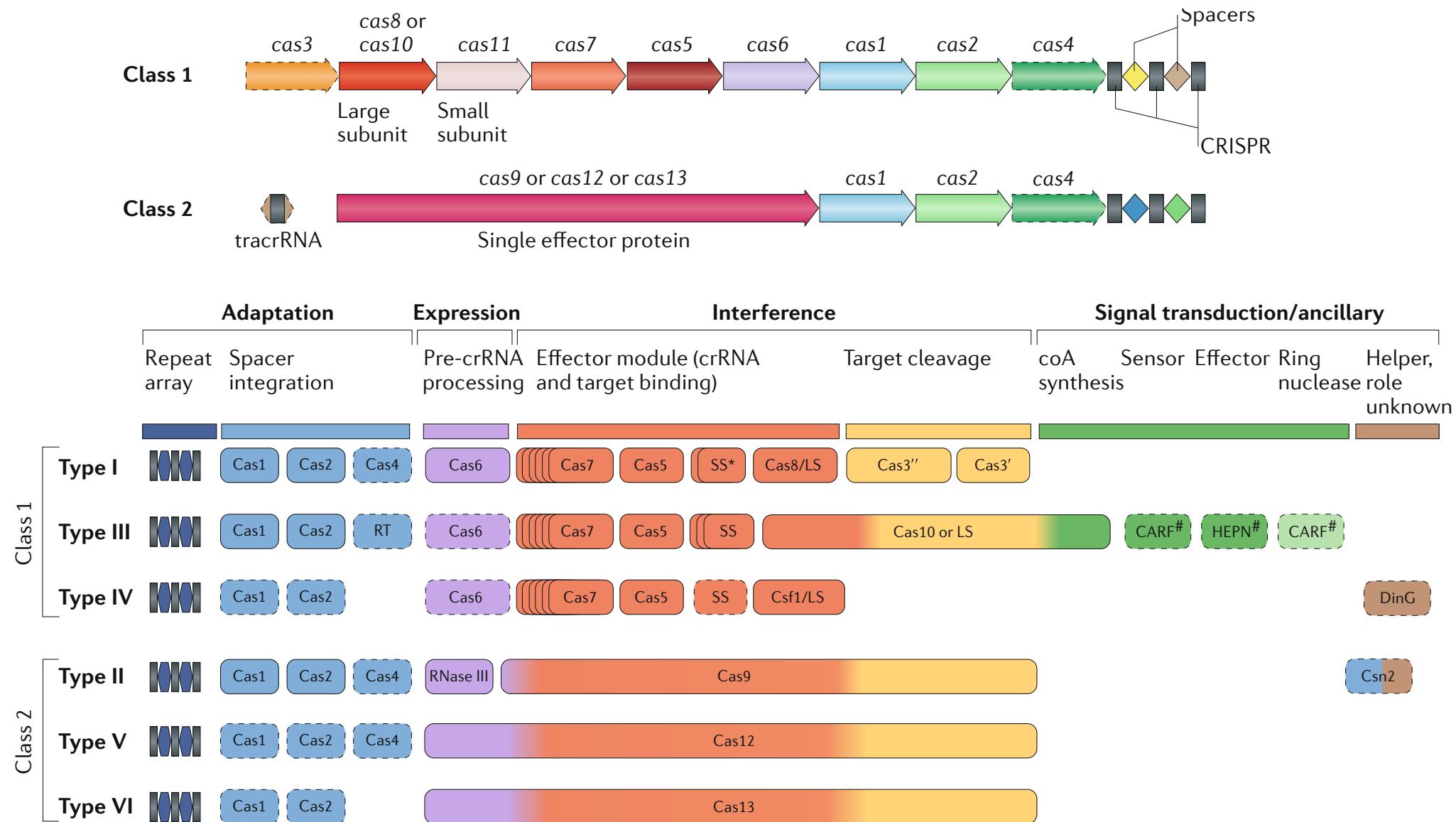
The interference or effector module

- Involved in target recognition and nucleic acid cleavage. In class 1 CRISPR–Cas systems, the effector module consists of multiple Cas proteins — namely, Cas3 (sometimes fused to Cas2), Cas5–Cas8, Cas10 and Cas11
- In class 2 systems, the effector module is represented by a single, large protein — Cas9, Cas12 or Cas13.

The signal transduction or ancillary module

- A collection of CRISPR-linked genes, most of which have roles in CRISPR–Cas systems that are, at best, tentatively predicted.

Functional modules and core genes



Type II Class 2 CRISPR systems

- The era of CRISPR technology has emerged from applications of the type II-A
- In complex with mature gRNA, Cas9 recognizes a G-rich PAM sequence and is directed to a target DNA that is complementary to the spacer sequence of crRNA.
- PAM recognition induces a structural alteration in Cas, resulting in unwinding of target DNAs to generate an R-loop formation between gRNA and target DNA.
- The HNH and RuvC domains are involved in the cleavage of target and nontarget strands, respectively.
- Previously, it was believed that SpCas9 creates blunt-end DSBs, but a recent publication revealed the formation of a staggered end with 5'-overhangs due to the postcleavage trimming activity of the RuvC domain¹⁸.

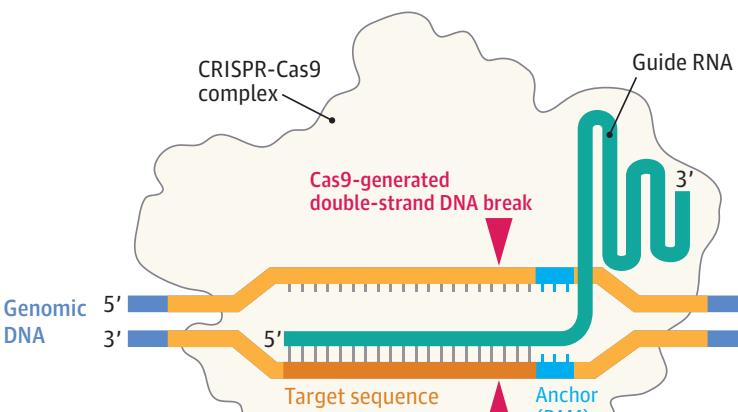
Type V Class 2 CRISPR systems

- The type V, Class 2 system, represented by Cas12a, is subdivided into 10 known subtypes from A-I to U according to the similarity of the domain organization
- Unlike type II, they possess only the RuvC domain
- The composition of gRNA is different among subtypes. Whereas Cas12a requires only crRNA, an additional tracrRNA is necessary for Cas12b, Cas12e, and Cas14a
- Cas12a possesses RNA editing activity and thus trims pre-crRNA to mature crRNA^{[31](#)}.
- The type V system usually shows specificity toward T-rich PAM sequences located 3'-upstream of a protospacer
- Cas14 exhibits collateral ssDNA degradation activity upon recognition of a target sequence

Type VI Class 2 CRISPR systems

- The type VI CRISPR-Cas system possesses unique RNase activity
- Cas13a (C2c2 / VI-A) grabs only crRNA that carries a 20-nt target-binding sequence.
- A pair of helical 1 domain or higher eukaryotes and prokaryotes nucleobinding (HEPN) domains replace the RuvC domain in other Cas proteins and are involved in RNA maturation and target RNA cleavage
- Type VI targets ssRNA and requires a protospacer flanking sequence (PFS) instead of the PAM required for dsDNA unwinding.
- The HEPN domain activated by target ssRNA cleavage also exhibits collateral activity toward nontarget ssRNAs

CRISPR-Cas9-Targeted DNA Breaks



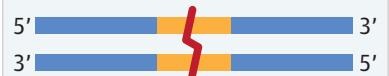
Single guide RNA



Two guide RNAs



NHEJ (error prone)



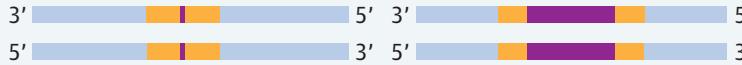
**Disruption of DNA sequence
(insertion or deletion)**

Potential clinical applications:
Dyslipidemia (PCSK9)
Familial amyloid cardiomyopathy (TTR)

HDR

Repair template

Repair template



**Correction or introduction
of a single nucleotide mutation**

Potential clinical application:
Cardiomyopathies (gene correction)

NHEJ



**Deletion of gene
(or other part of chromosome)**

Potential clinical application:
Duchenne muscular dystrophy
(DMD exon deletion)

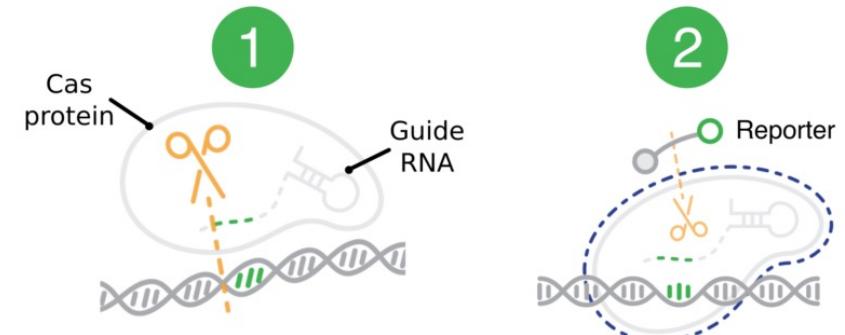
Cas12

The Cas12 proteins can cut either single or double stranded DNA.

Once a Cas12 protein cuts its DNA target, it begins to shred single stranded DNA non-specifically.

Thus Cas12-based diagnostics can only directly detect DNA. They must be combined with proteins that convert RNA into DNA to detect RNA.

Cas 12



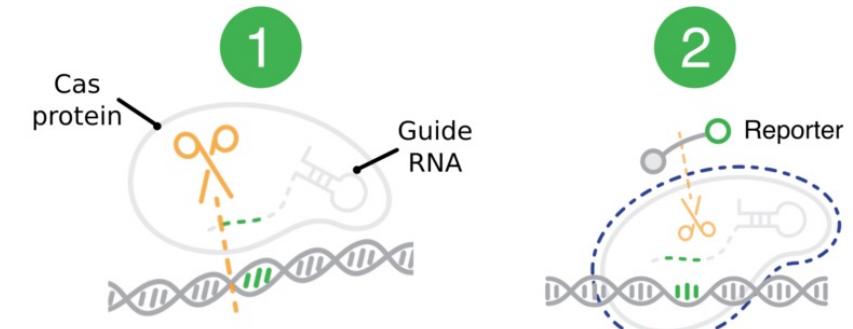
Cas13

The Cas13 proteins can cut can directly detect RNA, but not DNA..

Once a Cas13 protein cuts its DNA target, it begins to shred single stranded DNA non-specifically.

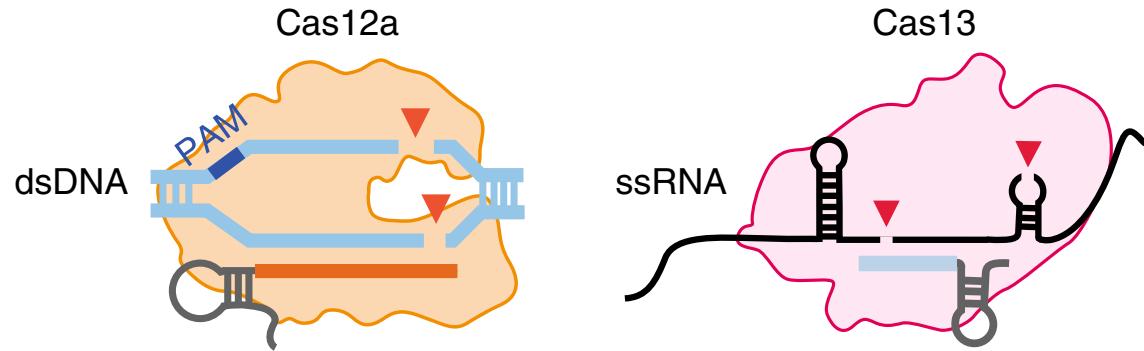
To detect DNA, Cas13-based diagnostics must be combined with proteins that convert DNA into RNA

Cas 13



Cas12 and Cas13 for diagnostics

To date, both the Cas13 and Cas12 protein families of CRISPR systems have been shown to have collateral activity, making them useful for nucleic acid detection applications^{6,8,9}. The schematic diagrams and table below show the key differences between the Cas13 and Cas12a enzymes.



PAM required	Yes	No
PAM identity	TTTV	Not applicable
Cleavage	Single staggered cut	Many cleavage sites
Target type	ssDNA, dsDNA	ssRNA only
Collateral	Yes	Yes

Cas12 and Cas13 for diagnostics

Cas12a is targeted to a specific DNA sequence, such as the Human papilloma virus (HPV) genome, via a crRNA.

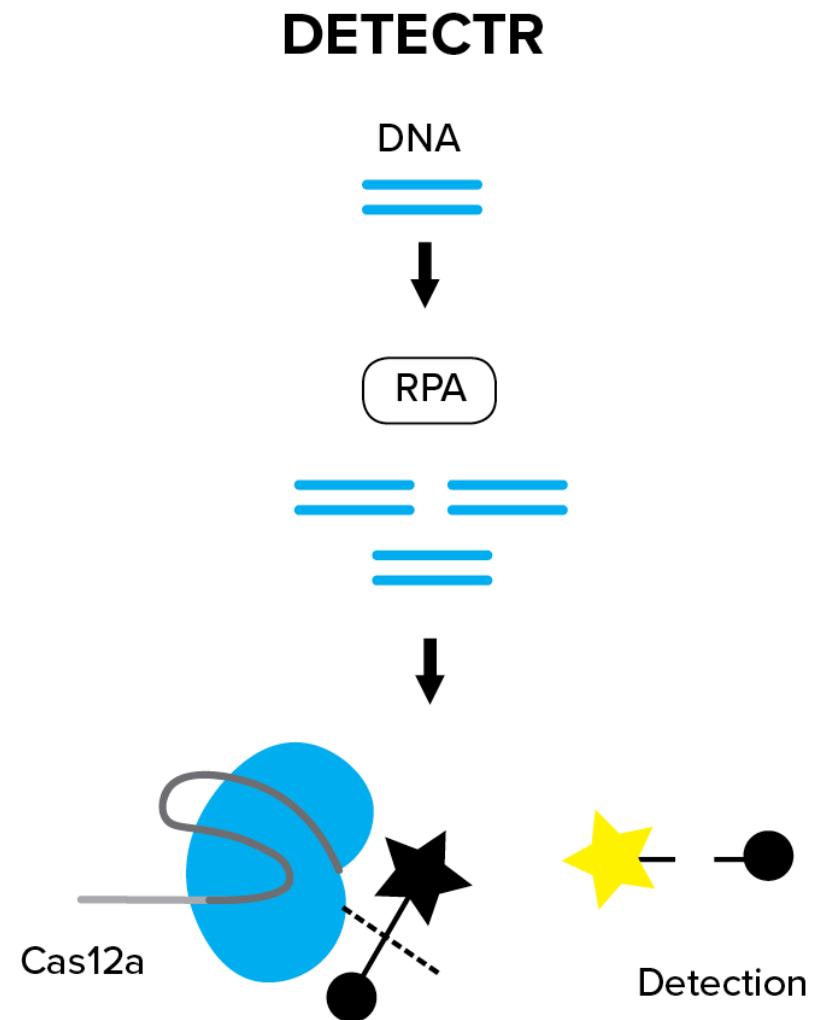
An ssDNA-fluorescently quenched (FQ) reporter, which will produce a signal when the ssDNA is degraded, is added to the reaction.

To enhance sensitivity, the DNA is first amplified through isothermal amplification by RPA.

When Cas12a-cRNA base pairs with the dsDNA of interest, the DNase activity of Cas12a is initiated.

Surrounding trans-ssDNA, including the ssDNA-FQ reporter are subsequently degraded at a rate of ~1,250 cuts per second.

A quantifiable fluorescent signal designates the presence of your DNA of interest, in this case HPV.



Cas12 and Cas13 for diagnostics

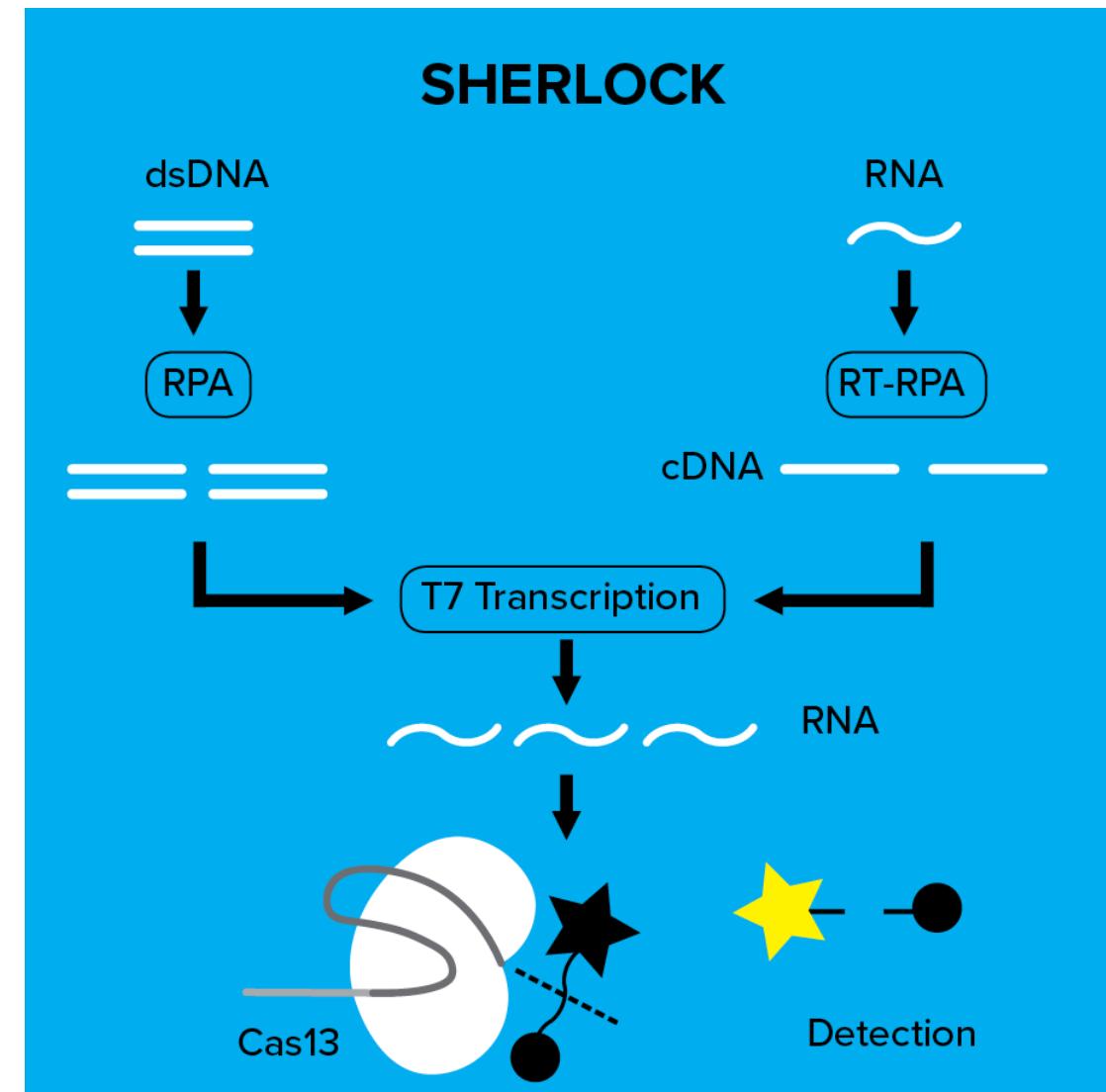
Cas13 can be programmed with crRNA to target an ssRNA of interest.

Once Cas13 recognizes and binds to the programmed sequence, it can promiscuously cleave surrounding ssRNA molecules.

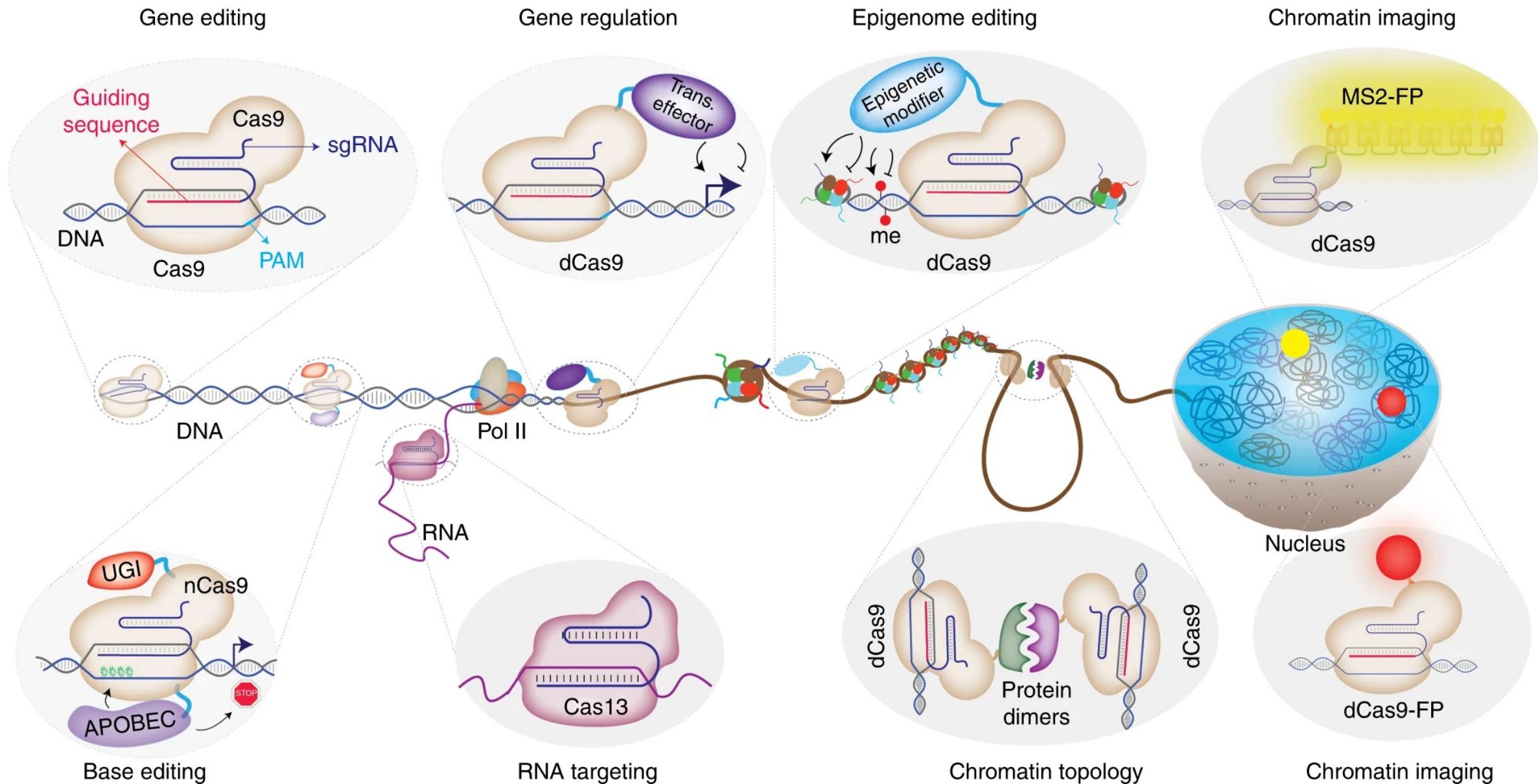
Cleavage of the quenched fluorescent RNA by the “activated” Cas13 produces a quantifiable signal that indicates the presence of your targeted nucleic acid.

Targeted DNA or RNA from a sample is first amplified using RPA (recombinase polymerase amplification) or reverse transcriptase (RT)-RPA, respectively.

RPA is coupled with T7 transcription to convert amplified DNA to RNA for subsequent detection by Cas13.



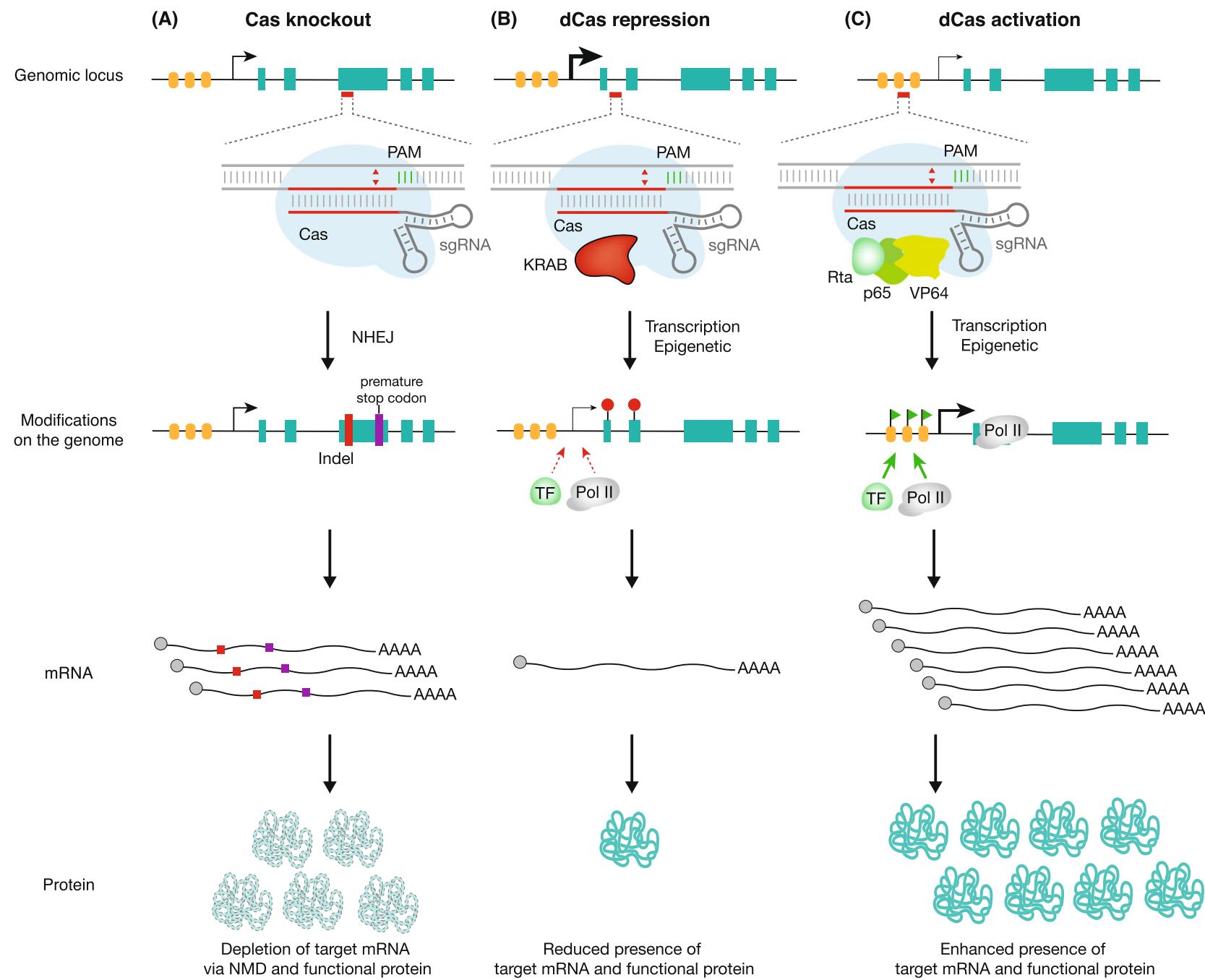
CRISPR technology: Beyond genome editing



CRISPR KO (genome editing)

CRISPRi (transcriptional repression)

CRISPRa (transcriptional activation)



3.5 Genomics and Bioinformatics: Hot Disciplines of Biotechnology

Genomics – cloning, sequencing, and analyzing entire genomes

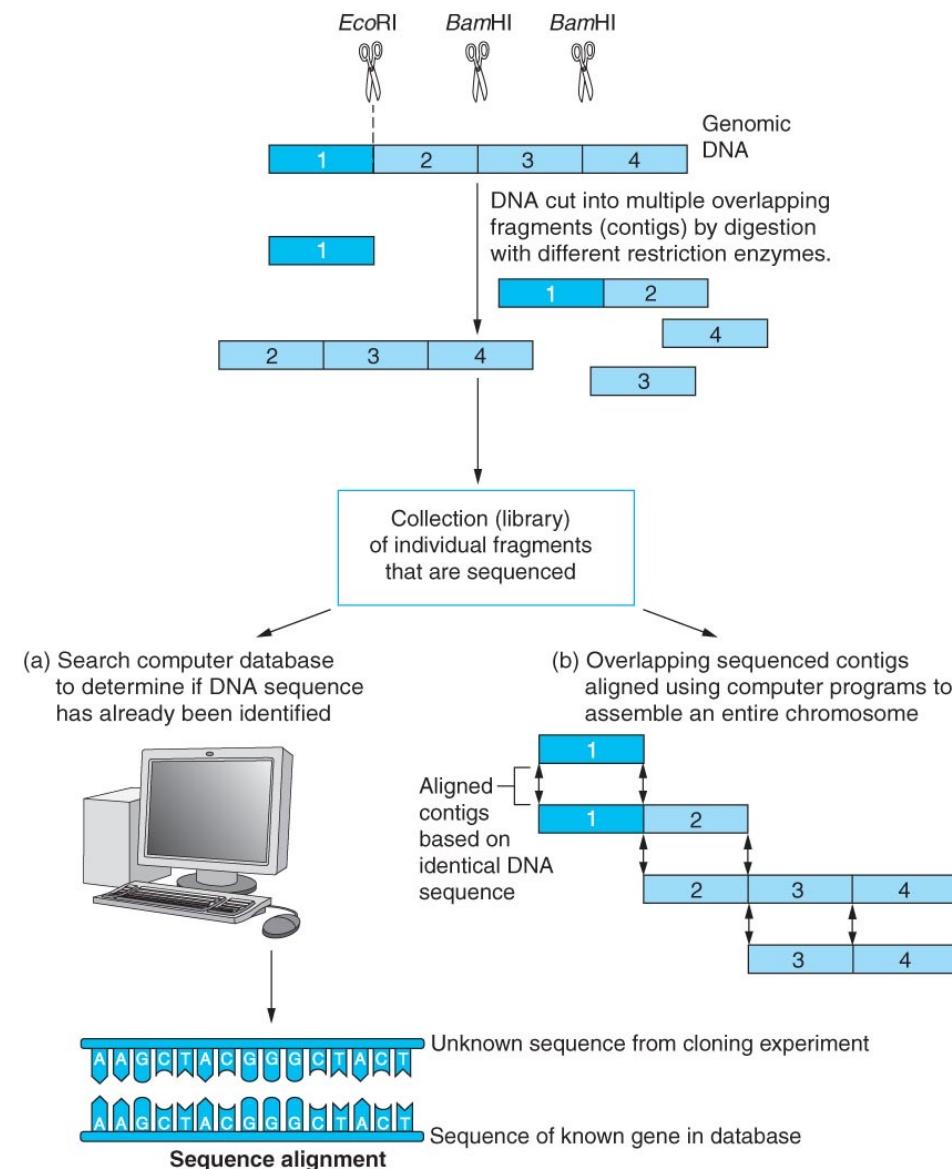
Whole genome **shotgun sequencing or shotgun cloning**

Use restriction enzymes to digest pieces of entire chromosomes

Produces thousands of overlapping fragments called **contigs** which are each sequenced.

Computer programs are used to align the sequenced fragments based on overlapping sequence pieces

Shotgun sequencing



Bioinformatics: Merging molecular biology with computer technology

An interdisciplinary field that applies computer science and information technology to promote an understanding of biological processes

Application of Bioinformatics:

Databases to store, share, and obtain the maximum amount of information related to gene structure, gene sequence and expression, and protein structure and function

The Human Genome Project

Started in 1990 by the U.S. Department of Energy

International collaborative effort to identify all human genes and to sequence all the base pairs of the 24 human chromosomes

Mostly done by 20 centers in 18 countries: China, France, Germany, Great Britain, Japan, and the United States

Competitor was a private company, Celera Genomics, directed by Dr. J. Craig Venter

The Human Genome Project

The Human Genome Project designed to accomplish the following:

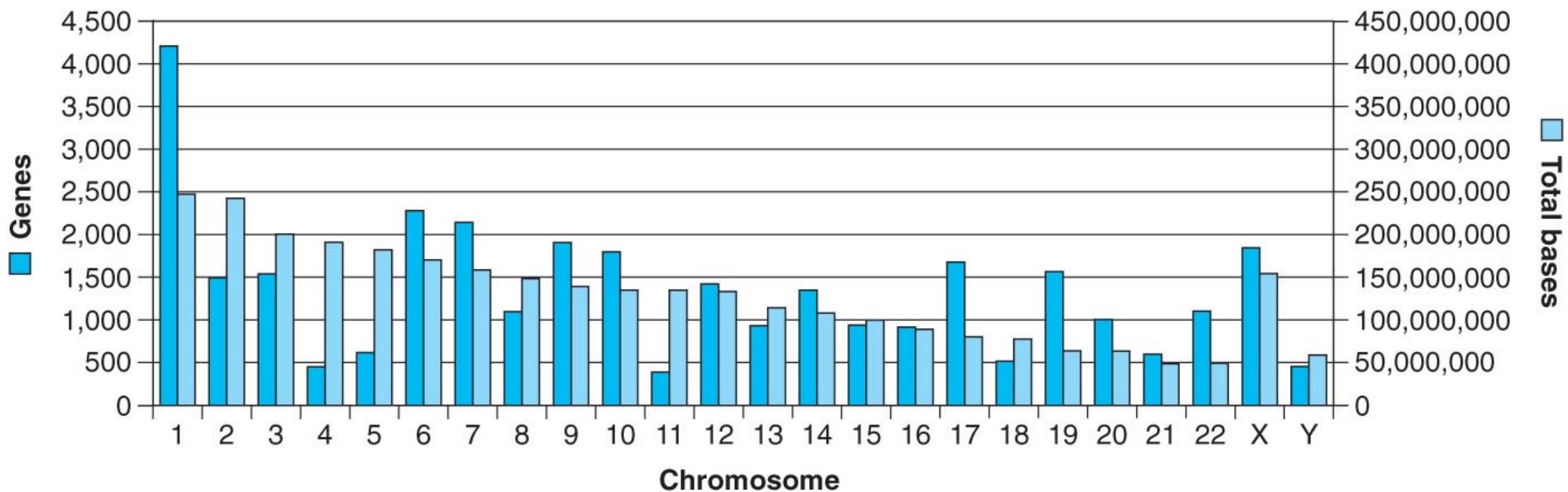
1. Analyze genetic variations among humans. This included the identification of single-nucleotide polymorphisms
2. Map and sequence the genomes of **model organisms**, including bacteria, yeast, roundworms, fruit flies, mice, and others
3. Develop new laboratory technologies such as high-powered automated sequencers and computing technologies, as well as widely available databases of genome information, which can be used to advance our analysis and understanding of gene structure and function
4. Disseminate genome information among scientists and the general public
5. Consider the ethical, legal, and social issues that accompany the HGP and genetic research

The Human Genome Project

April 14, 2003: map of the human genome was completed

Consists of 20,000 protein-coding genes

Map was complete with virtually all bases identified and placed in their order and potential genes assigned to chromosomes



What we have learned from HGP

1. The human genome consists of approximately 3.1 billion base pairs
2. The genome is approximately 99.9% the same between individuals of all nationalities
3. Single-nucleotide polymorphisms (SNPs) and **copy number variations (CNVs)**—such as long deletions, insertions and duplications in the genome—account for much of the genome diversity identified between humans
4. Less than 2% of the genome codes for genes

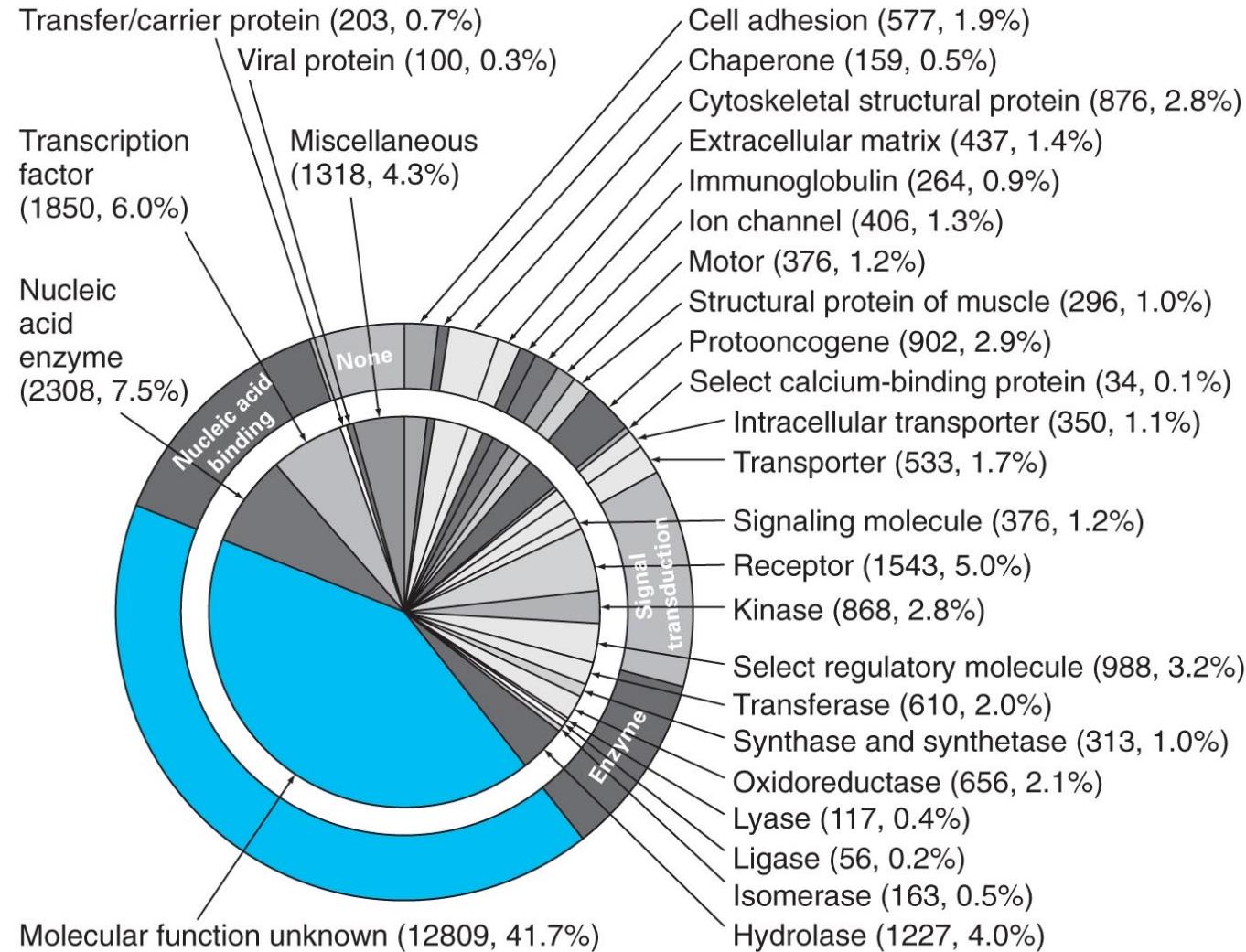
What we have learned from HGP

5. The vast majority of our DNA is non-protein coding, and repetitive DNA sequences account for at least 50% of the noncoding DNA
6. The genome contains approximately 20,000 protein-coding genes
7. Many human genes are capable of making more than one protein, allowing human cells to make at least 100,000 proteins from only about 20,000 genes.

What we have learned from HGP

8. Chromosome 1 contains the highest number of genes. The Y chromosome contains the fewest genes.
9. Many of the genes in the human genome show a high degree of sequence similarity to genes in other organisms.
10. Thousands of human disease genes have been identified and mapped to their chromosomal locations

What we have learned from HGP



The Human Genome Project started an "omics" revolution

Proteomics—studying all proteins in a cell

Metabolomics—studying proteins and enzymatic pathways involved in cell metabolism

Glycomics—studying carbohydrates of a cell

Transcriptomics—studying all genes transcribed in a cell

The Human Genome Project started an "omics" revolution

Metagenomics—analysis of genomes of organisms in an environment

Pharmacogenomics—customized medicine based on person's genetic profile for a particular condition

Nutrigenomics—interaction between genes and diet

Comparative Genomics

Mapping and sequencing genomes from a number of model organisms

Allows researchers to study gene structure and function in these organisms in ways designed to understand gene structure and function in other species including humans

Stone Age Genomics (paleogenomics)

Analyzing "ancient" DNA

Analysis of DNA from a 2,400-year-old Egyptian mummy, mammoths, platypuses, Pleistocene-age cave bears, and Neanderthals are some of the most prominent examples

In 2009, a rough draft of the Neanderthal (*Homo neanderthalensis*) genome was completed