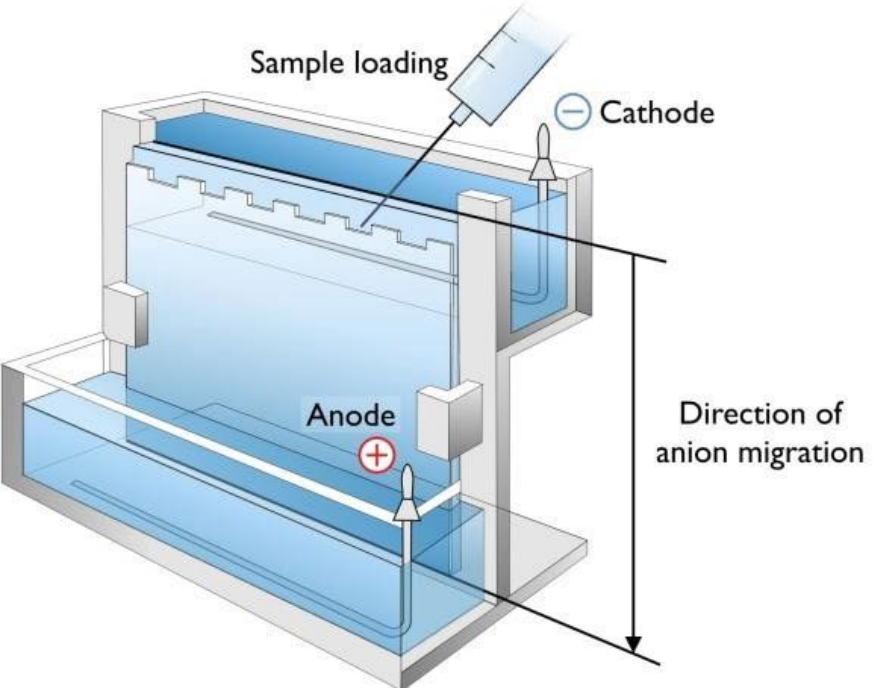
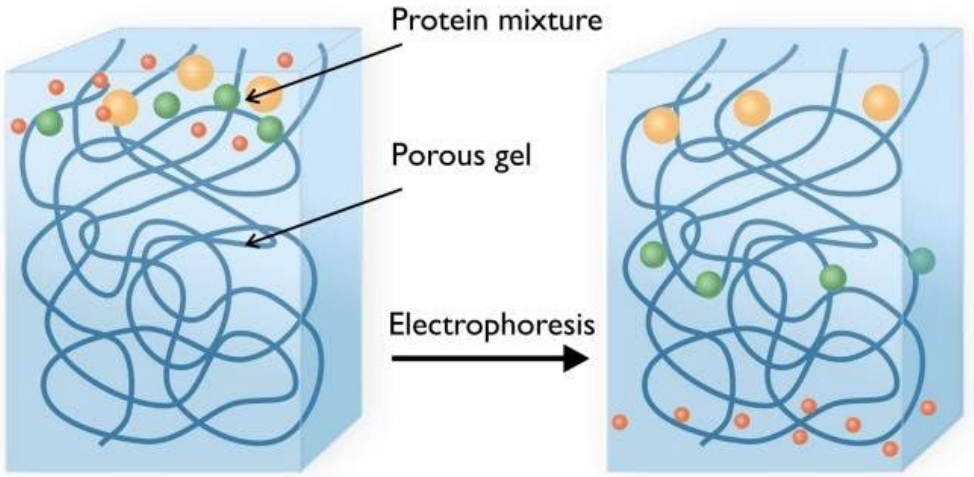
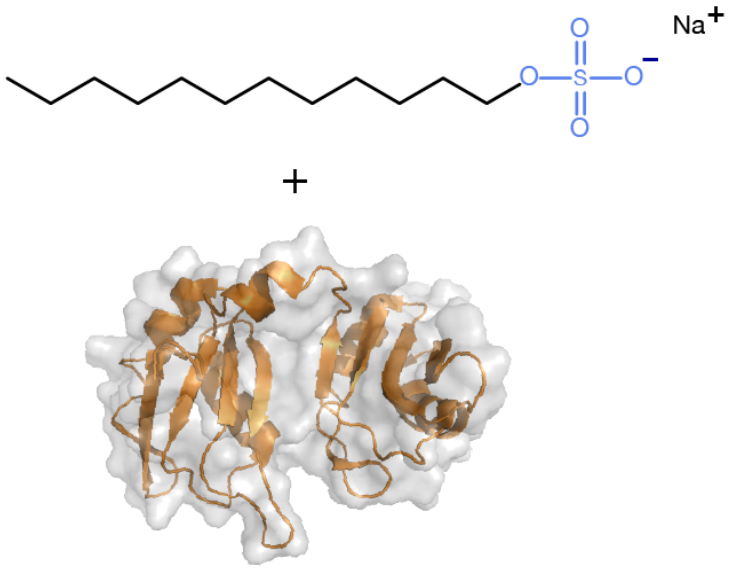
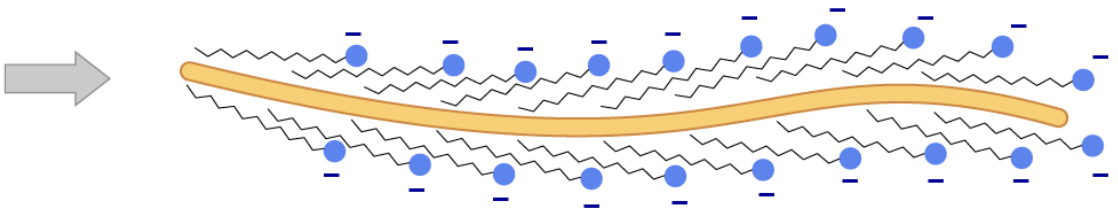
# Sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE)



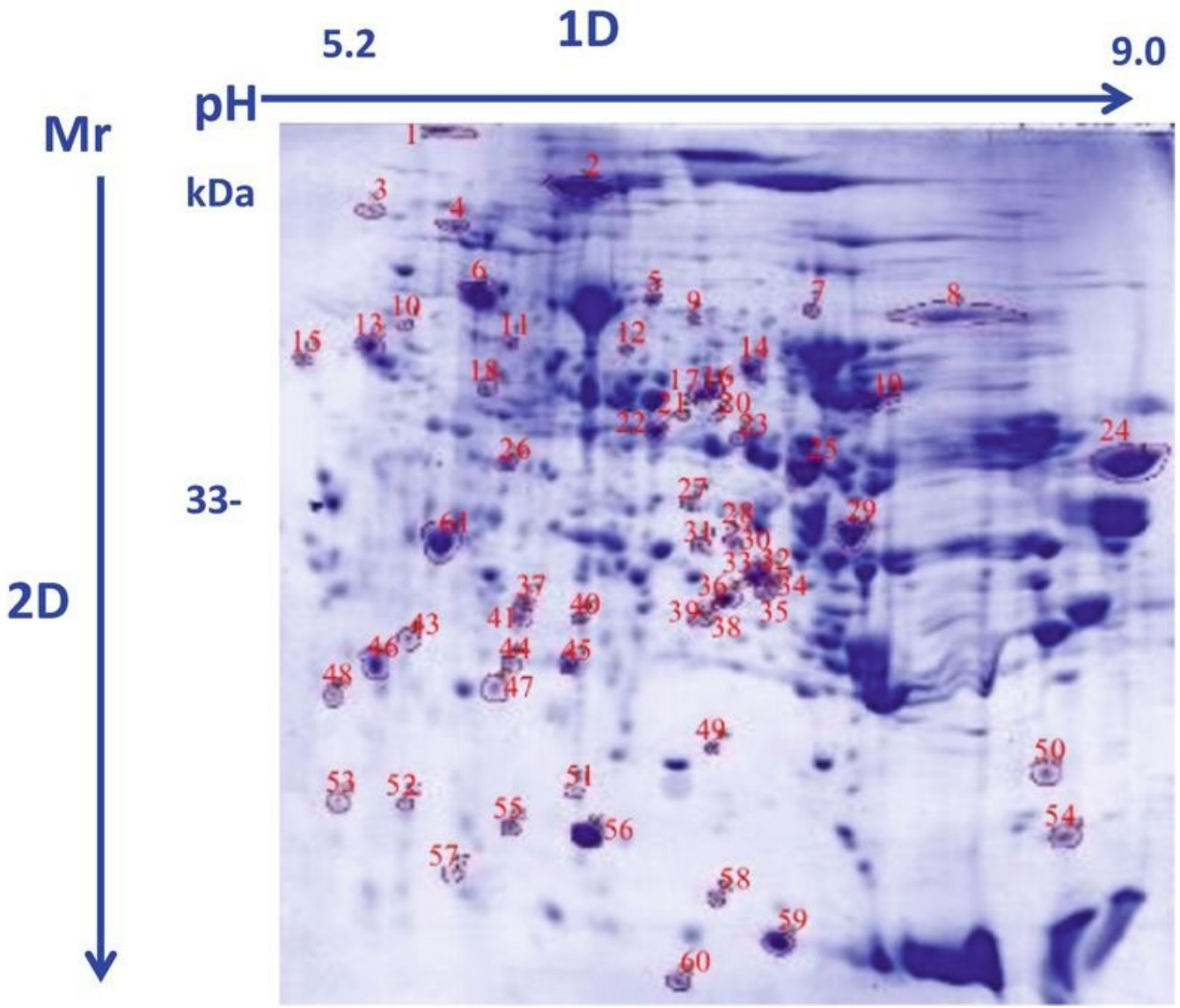



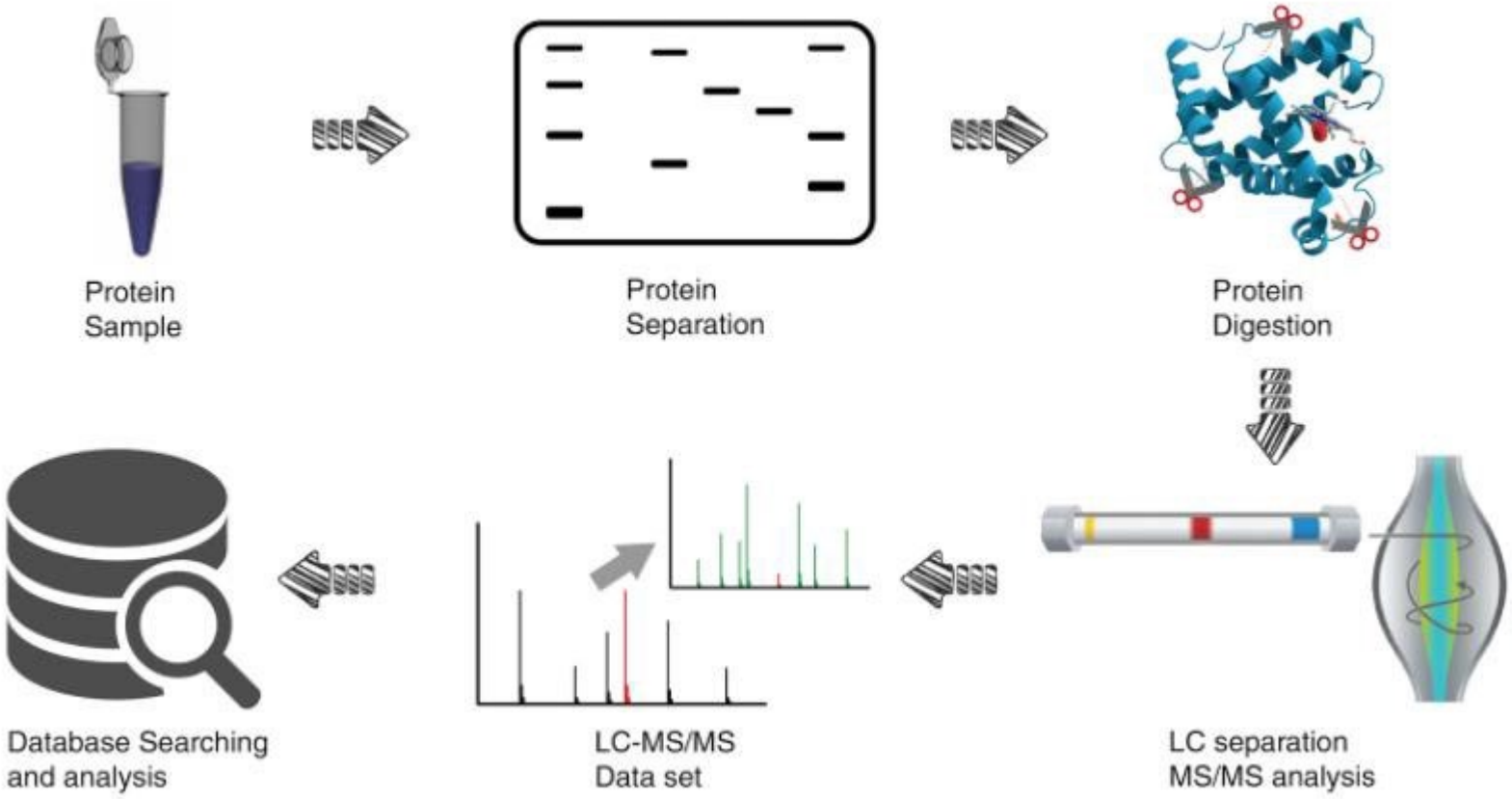
# Protein Denature



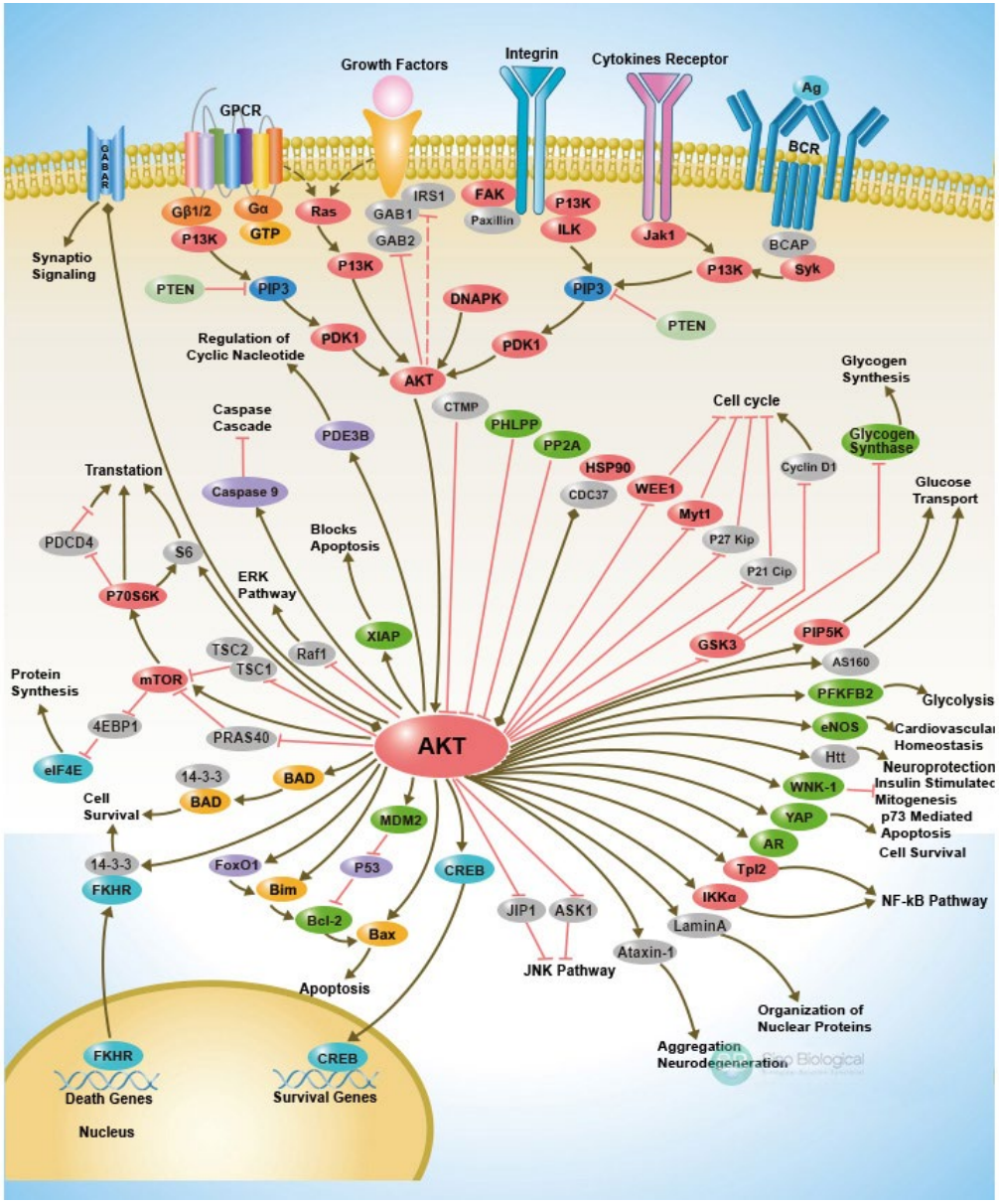



# 2D PAGE

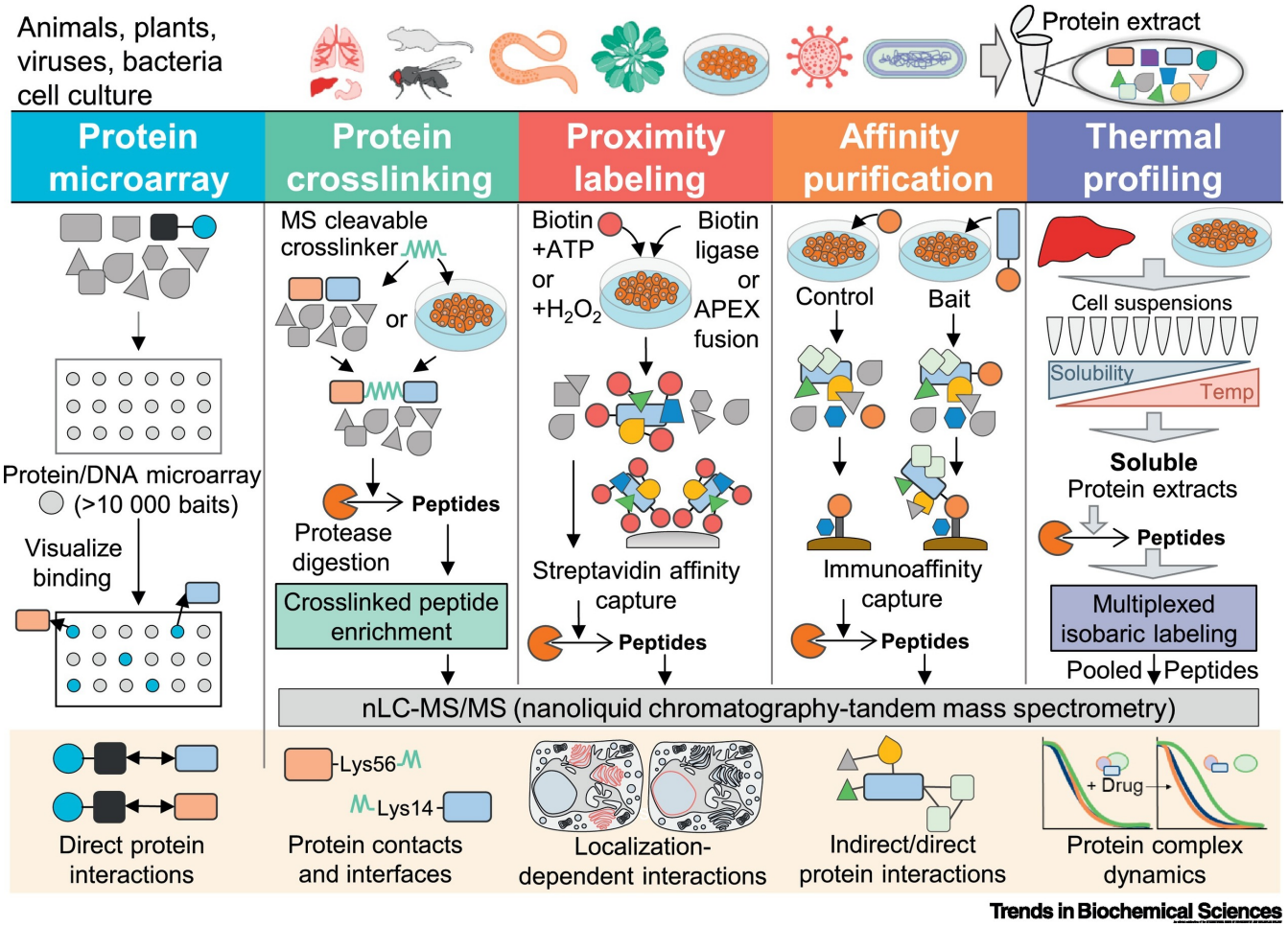



Protein
Sample
Protein
Separation
Protein
Digestion
Database Searching
and analysis
LC-MS/MS
Data set
LC separation
MS/MS analysis

# Signal Transduction

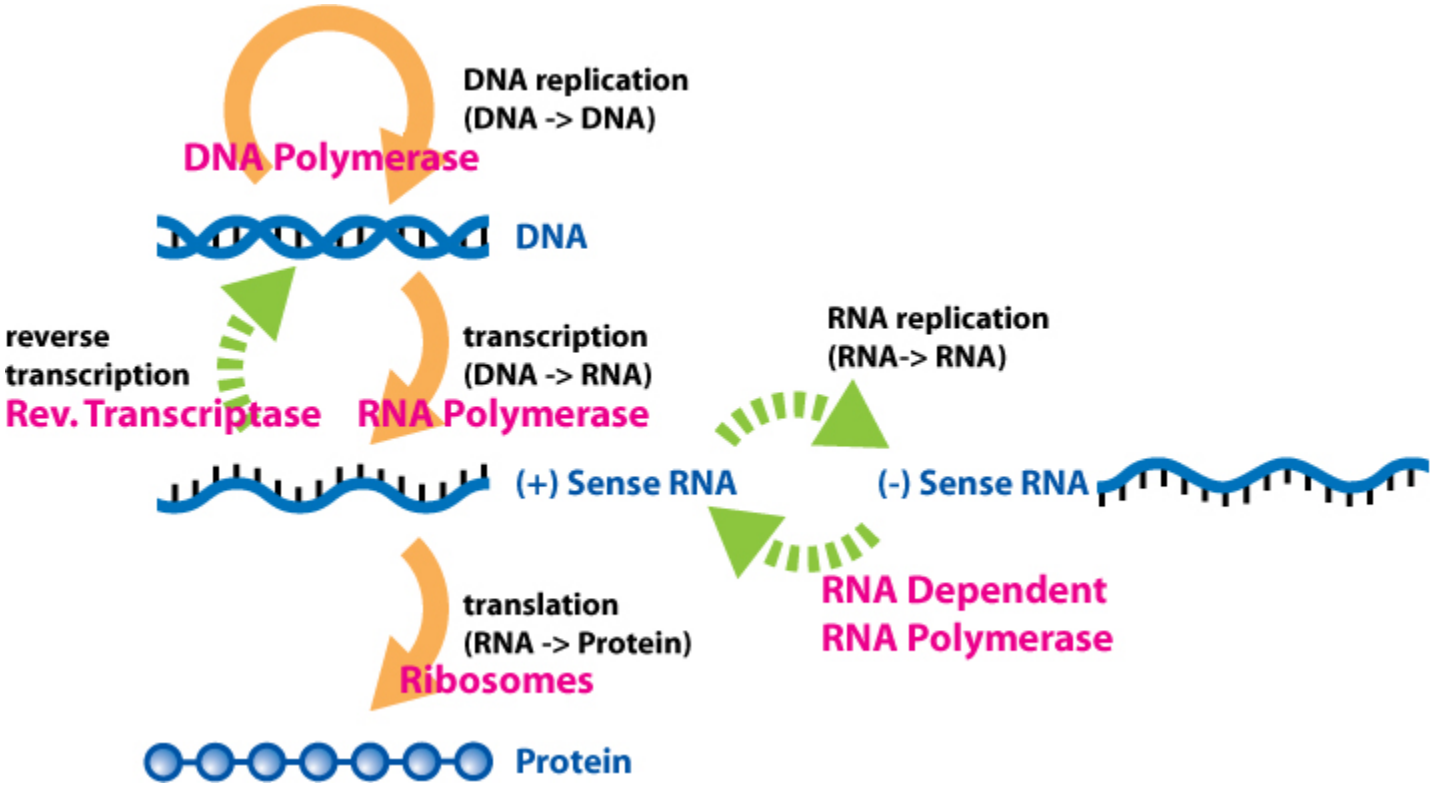


# Proteomic




*Trends in Biochemical Sciences* 2020 45454-455DOI: (10.1016/j.tibs.2020.01.003)

# Central Dogma

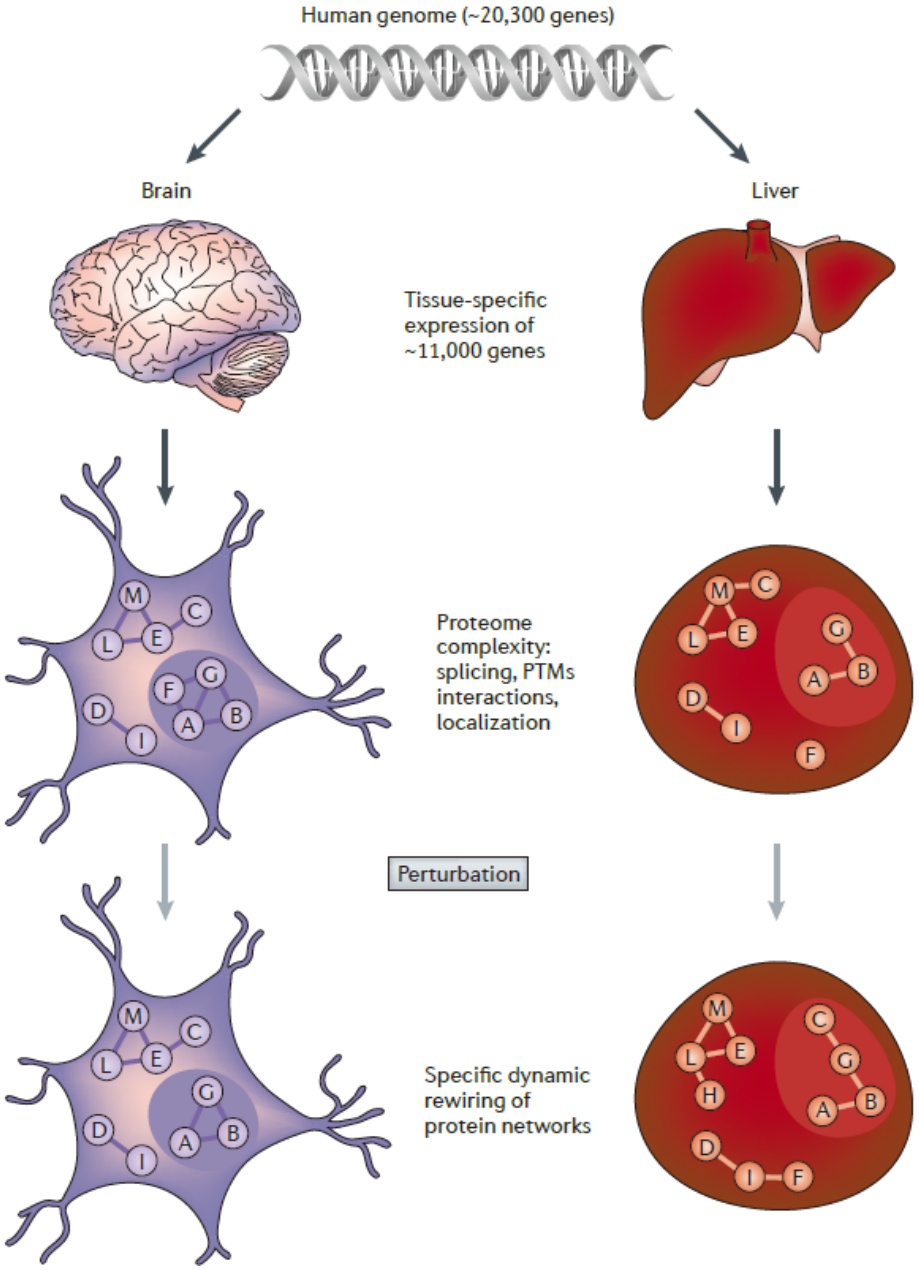


# From DNA to Protein

https://www.youtube.com/watch?v=gG7uCskUOrA

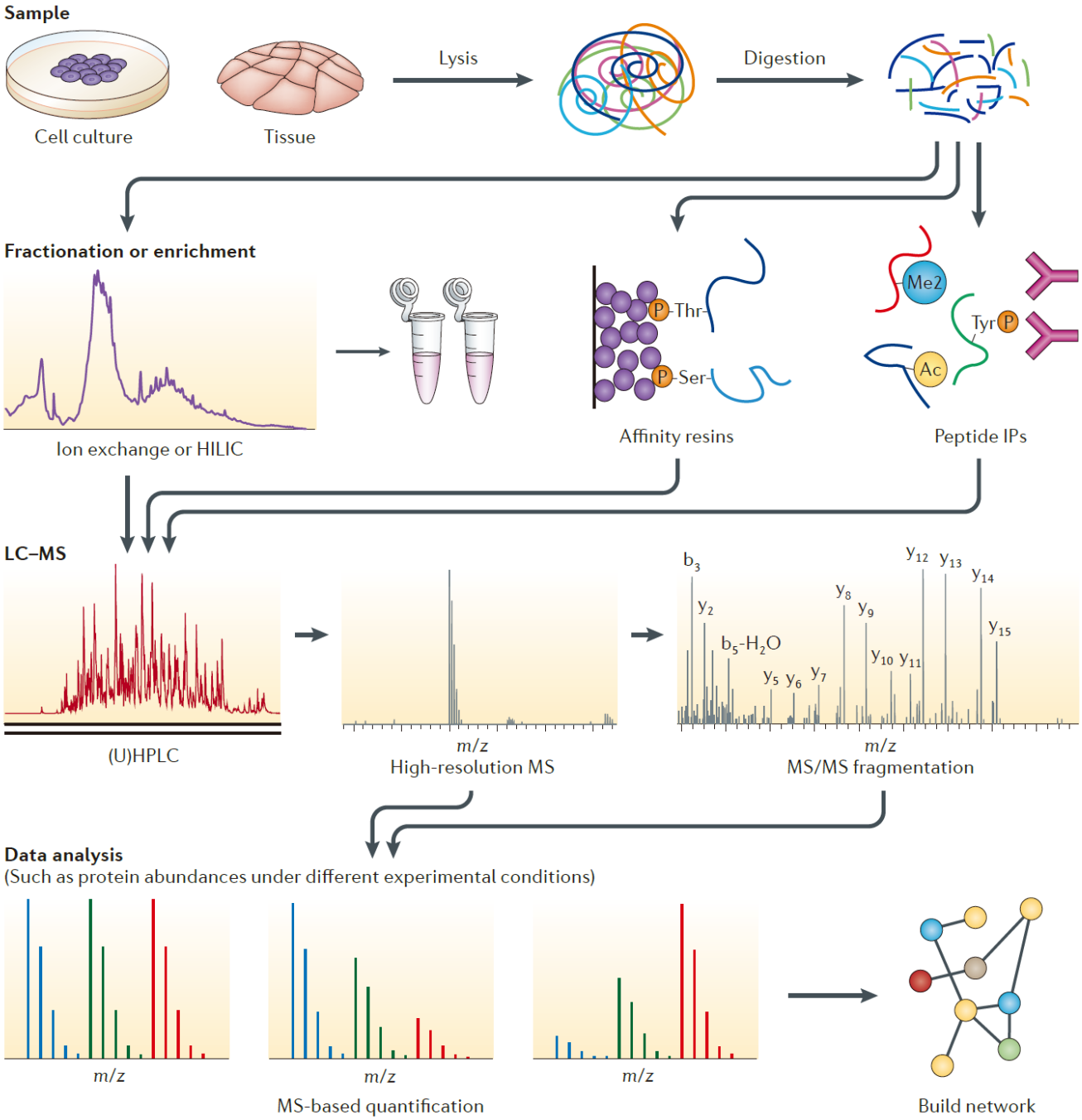
# Next-generation proteomics: towards an integrative view of proteome dynamics

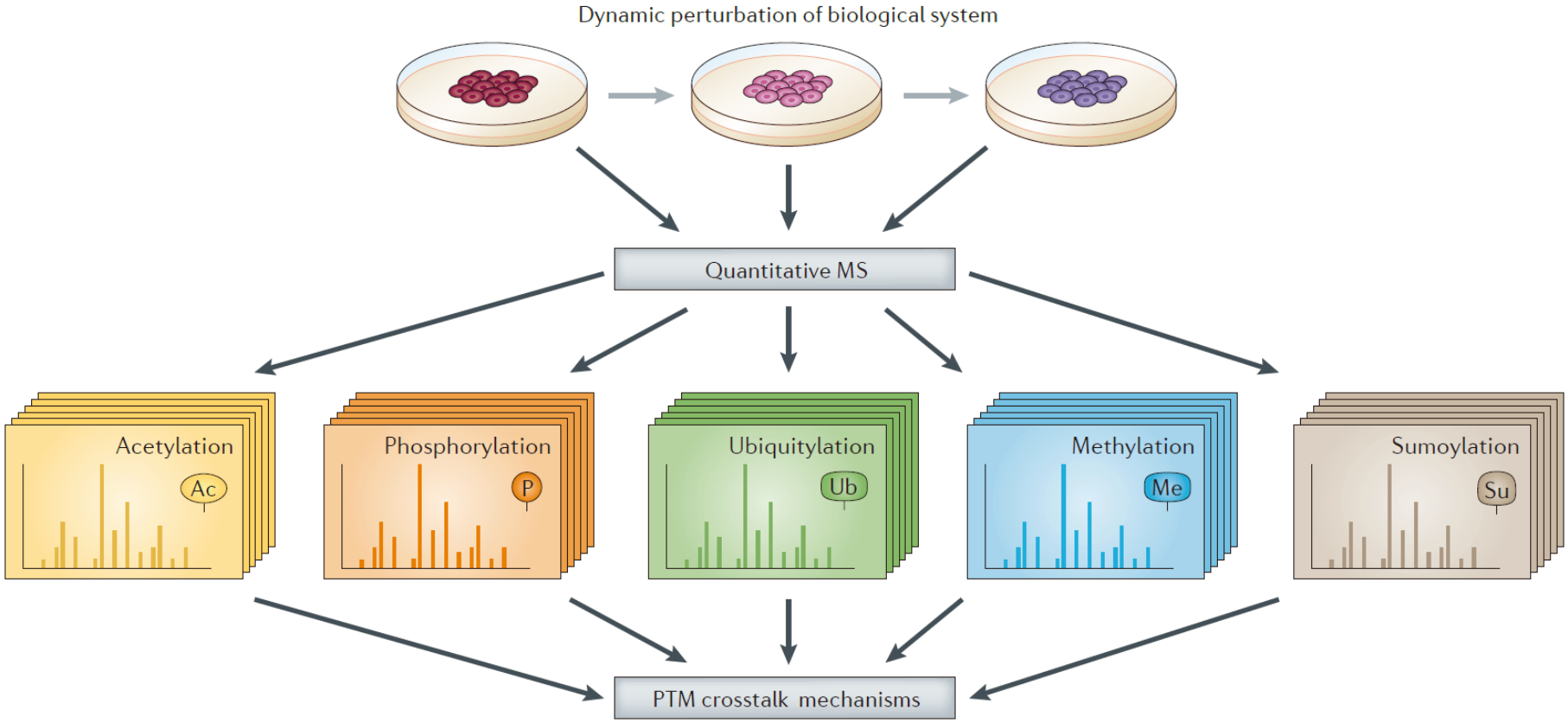
*A. F. Maarten Altelaar[1,2]*, Javier Munoz[1,2,3]* and Albert J. R. Heck[1,2]*



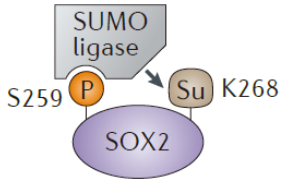
## Key Points

- Our understanding of cellular function depends on exquisite knowledge of all of the molecular components acting in a system. Mass spectrometry (MS)-based proteomics has matured immensely in the last decade, allowing quantitative system-wide analysis of the proteome, including post-translational modifications (PTMs), protein–protein interactions and cellular localization.
- Quantification of the entire set of proteins expressed in a complex biological system (for example, mammalian cells) is now possible with a high sensitivity and in a reasonable amount of time.
- With the availability of genomic information, the massive capacity for peptide identification by MS is being used to annotate gene sequences and to find new protein-coding genes and splicing variants.
- In combination with new approaches to isolate specific PTMs, MS-based studies are revealing a much higher order of proteome complexity in which most proteins are modified by several PTMs that crosstalk in intricate mechanisms to regulate protein function.
- Protein affinity strategies allow purification of candidate proteins and their interacting partners, which are subsequently identified by MS. These studies describe, with a high degree of detail, dynamic and context-specific protein–protein interaction networks and protein complexes.
- The improvements in sensitivity, robustness and high-throughput of MS-based proteomics now permits applications in the clinical field, including the possibility of discovering disease-related biomarkers and screening molecular targets of candidate drugs.


Sample
Cell culture
Tissue
Lysis
Digestion
Fractionation or enrichment
Ion exchange or HILIC
P-Thr-
P-Ser-
Affinity resins
Me2
Tyr P
Ac
Peptide IPs
LC–MS
(U)HPLC
m/z
High-resolution MS
b3
y2
b5-H2O
y5 y6
y7
y8
y9
y10 y11
y12 y13
y14
y15
m/z
MS/MS fragmentation
Data analysis
(Such as protein abundances under different experimental conditions)
m/z
m/z
m/z
MS-based quantification
Build network

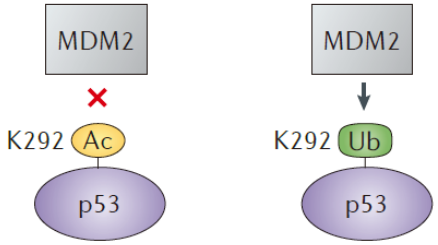



**Sequential**



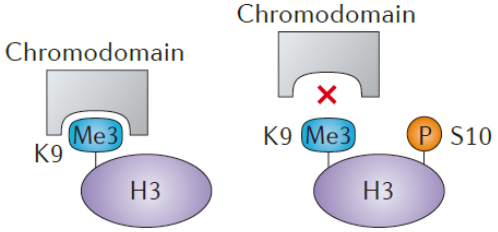

Phosphorylation-dependent SUMO modification

**Mutually exclusive**

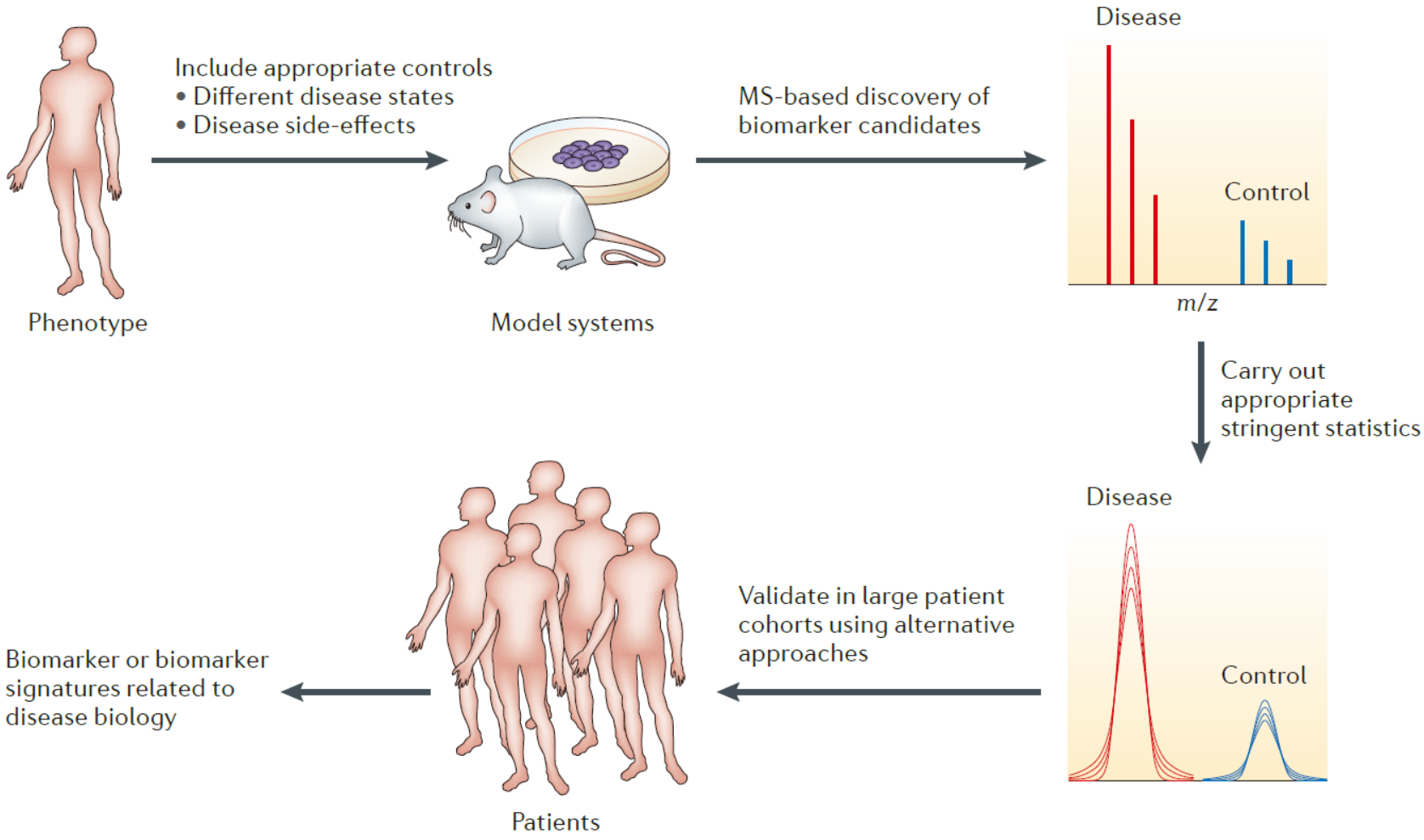



Acetylation prevents degradation of p53 by MDM2

**Antagonistic**




Phosphorylation disrupts H3 interaction with chromodomain


Phenotype
Include appropriate controls
• Different disease states
• Disease side-effects
Model systems
MS-based discovery of biomarker candidates
Disease
Control
m/z
Carry out appropriate stringent statistics
Disease
Control
Validate in large patient cohorts using alternative approaches
Patients
Biomarker or biomarker signatures related to disease biology

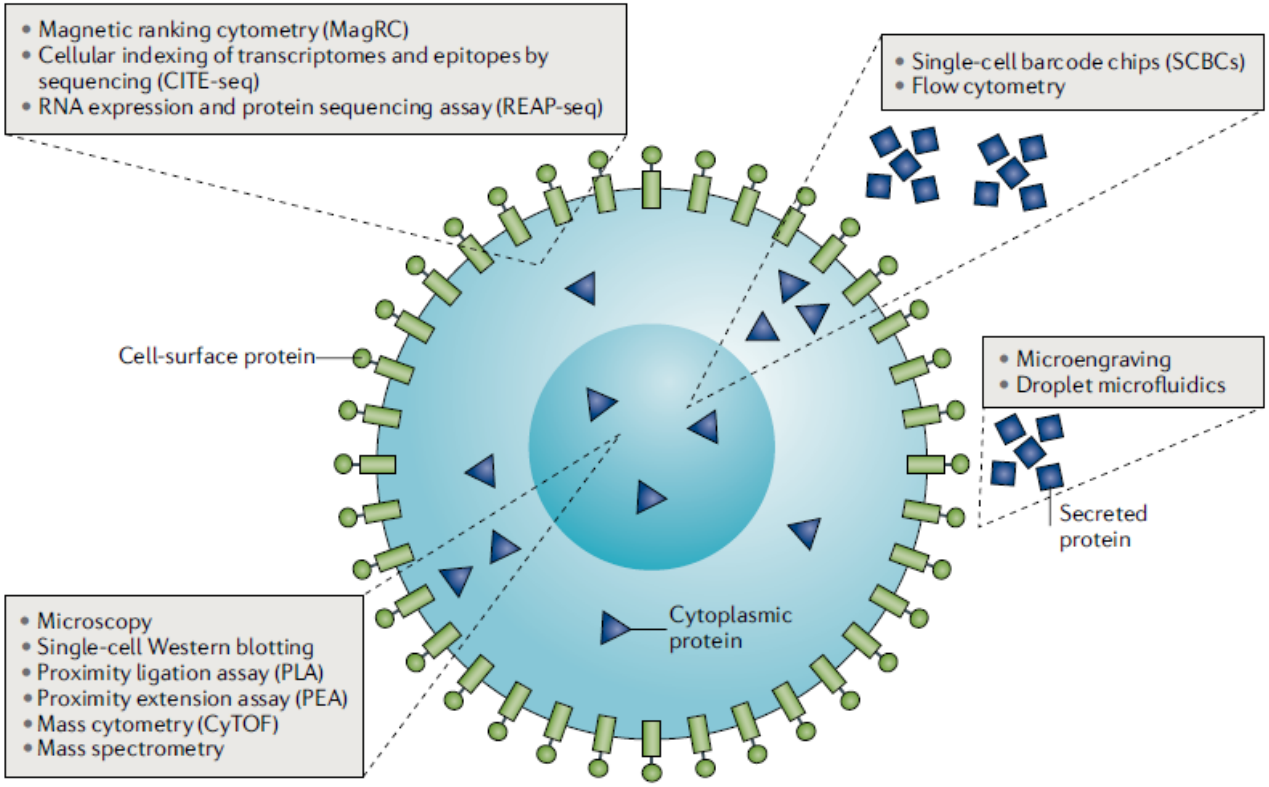
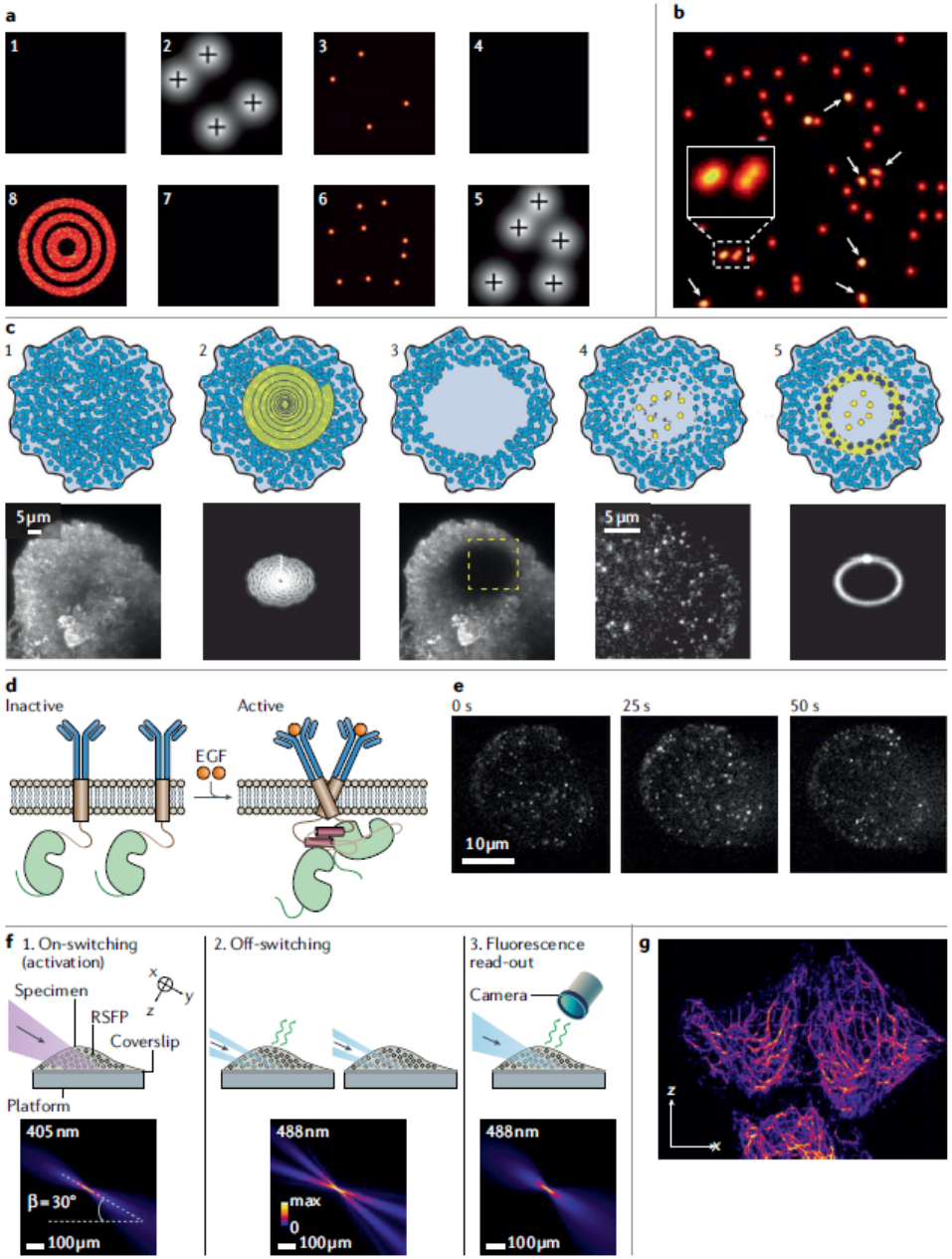
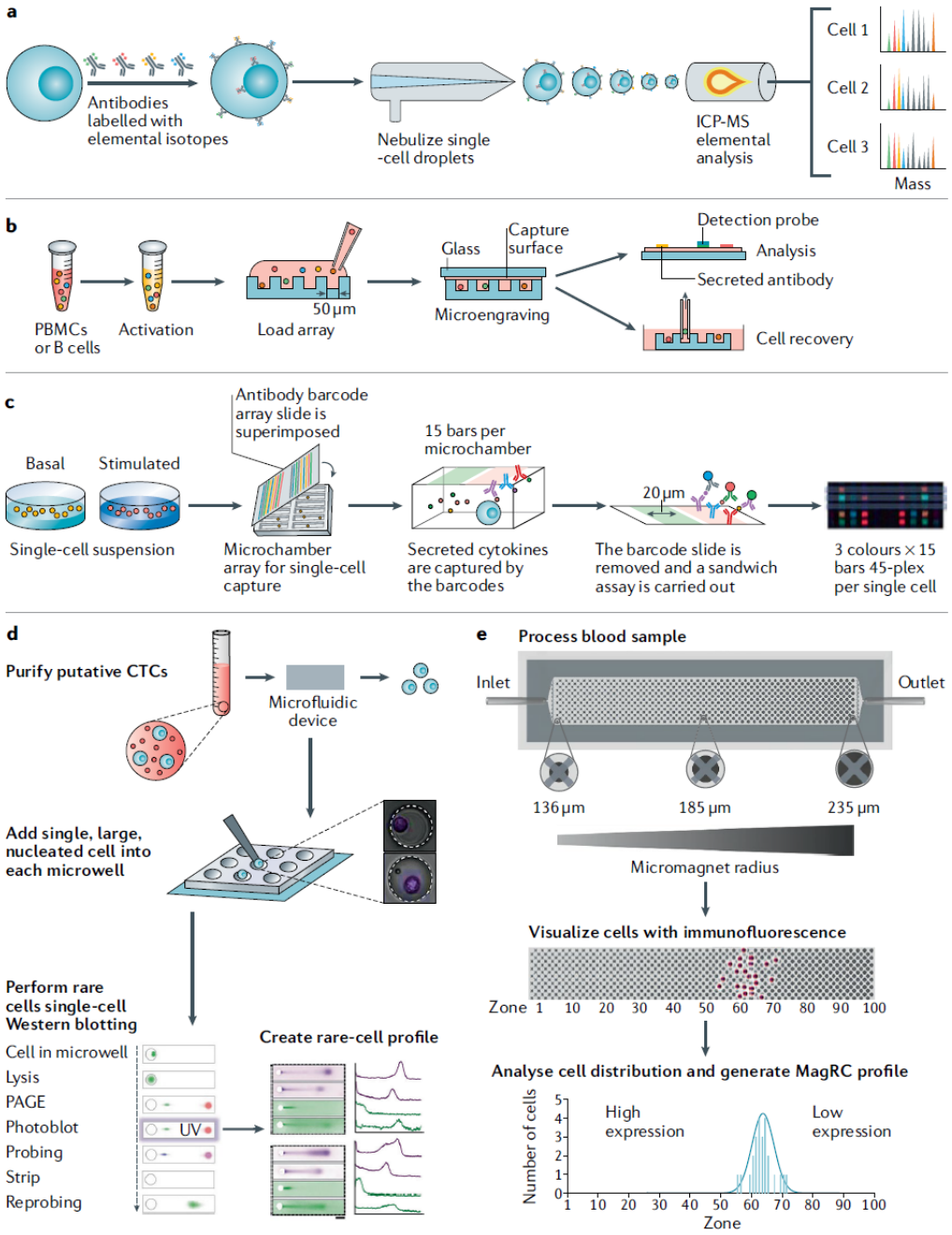
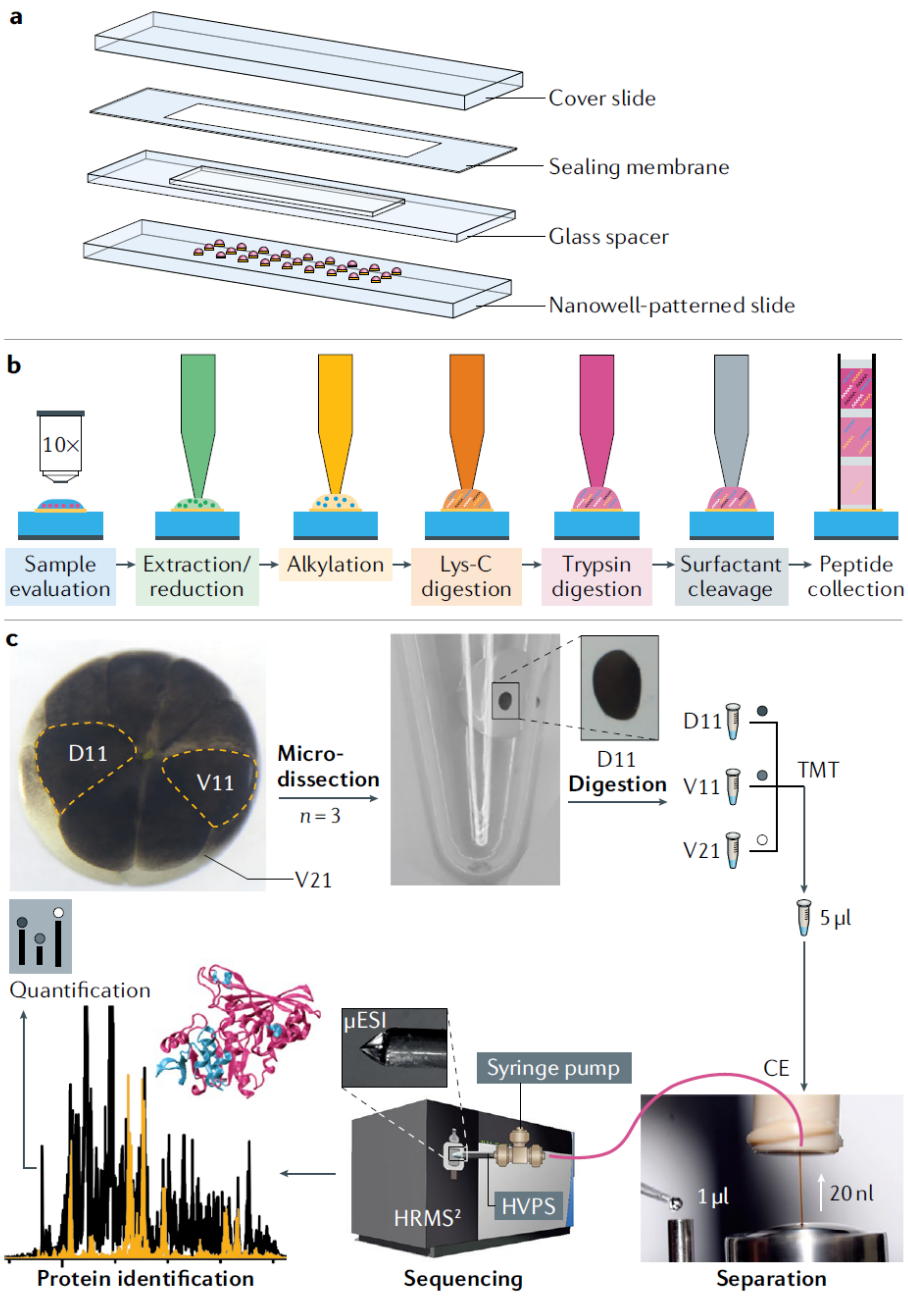



Fig. 1 | **Classification of single-cell protein analysis methods based on the location of target protein.** Cell-surface-protein analysis methods include magnetic ranking cytometry (MagRC), cellular indexing of transcriptomes and epitopes by sequencing (CITE-seq), and RNA expression and protein sequencing assay (REAP-seq). Methods that can be used for the analysis of cell-surface and cytoplasmic proteins include microscopy, single-cell Western blotting, proximity ligation assay (PLA), proximity extension assay (PEA), mass cytometry (cytometry by time of flight; CyTOF) and mass spectrometry. Methods utilized for secreted-protein analysis include droplet microfluidics and microengraving techniques. Methods used for comprehensive analysis of the three proteins include flow cytometry and single-cell barcode chips (SCBCs).


a
1
2
3
4
8
7
6
5
b
c
1
2
3
4
5
5μm
5 μm
d
Inactive
Active
EGF
e
0 s
25 s
50 s
10μm
f
1. On-switching (activation)
Specimen
RSFP
Coverslip
Platform
x
y
z
2. Off-switching
3. Fluorescence read-out
Camera
405 nm
β=30°
100μm
488nm
max
0
100μm
488nm
100μm
g
z
x


a
Antibodies labelled with elemental isotopes
Nebulize single -cell droplets
ICP-MS elemental analysis
Cell 1
Cell 2
Cell 3
Mass
b
PBMCs or B cells
Activation
50 µm
Load array
Glass
Capture surface
Microengraving
Detection probe
Analysis
Secreted antibody
Cell recovery
c
Basal
Stimulated
Single-cell suspension
Antibody barcode array slide is superimposed
Microchamber array for single-cell capture
15 bars per microchamber
Secreted cytokines are captured by the barcodes
20 µm
The barcode slide is removed and a sandwich assay is carried out
3 colours × 15 bars 45-plex per single cell
d
Purify putative CTCs
Microfluidic device
Add single, large, nucleated cell into each microwell
Perform rare cells single-cell Western blotting
Cell in microwell
Lysis
PAGE
Photoblot
UV
Probing
Strip
Reprobing
Create rare-cell profile
e
Process blood sample
Inlet
Outlet
136 µm
185 µm
235 µm
Micromagnet radius
Visualize cells with immunofluorescence
Zone 1 10 20 30 40 50 60 70 80 90 100
Analyse cell distribution and generate MagRC profile
Number of cells
High expression
Low expression
Zone

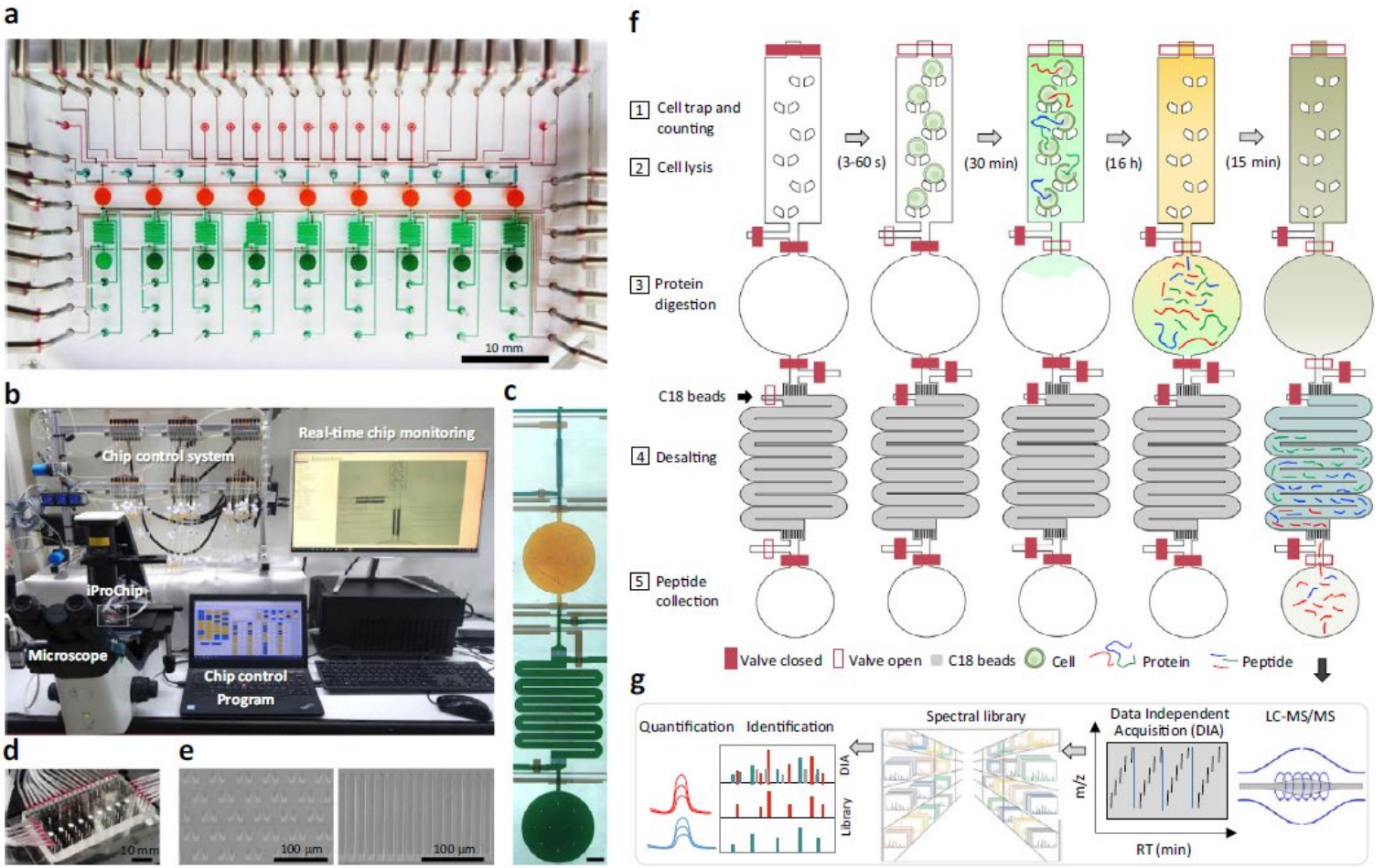

a
Cover slide
Sealing membrane
Glass spacer
Nanowell-patterned slide
b
10×
Sample evaluation
Extraction/ reduction
Alkylation
Lys-C digestion
Trypsin digestion
Surfactant cleavage
Peptide collection
c
D11
V11
V21
Micro-dissection
n = 3
D11 Digestion
D11
V11
V21
TMT
5 µl
Quantification
µESI
Syringe pump
CE
HRMS²
HVPS
1 µl
20 nl
Protein identification
Sequencing
Separation

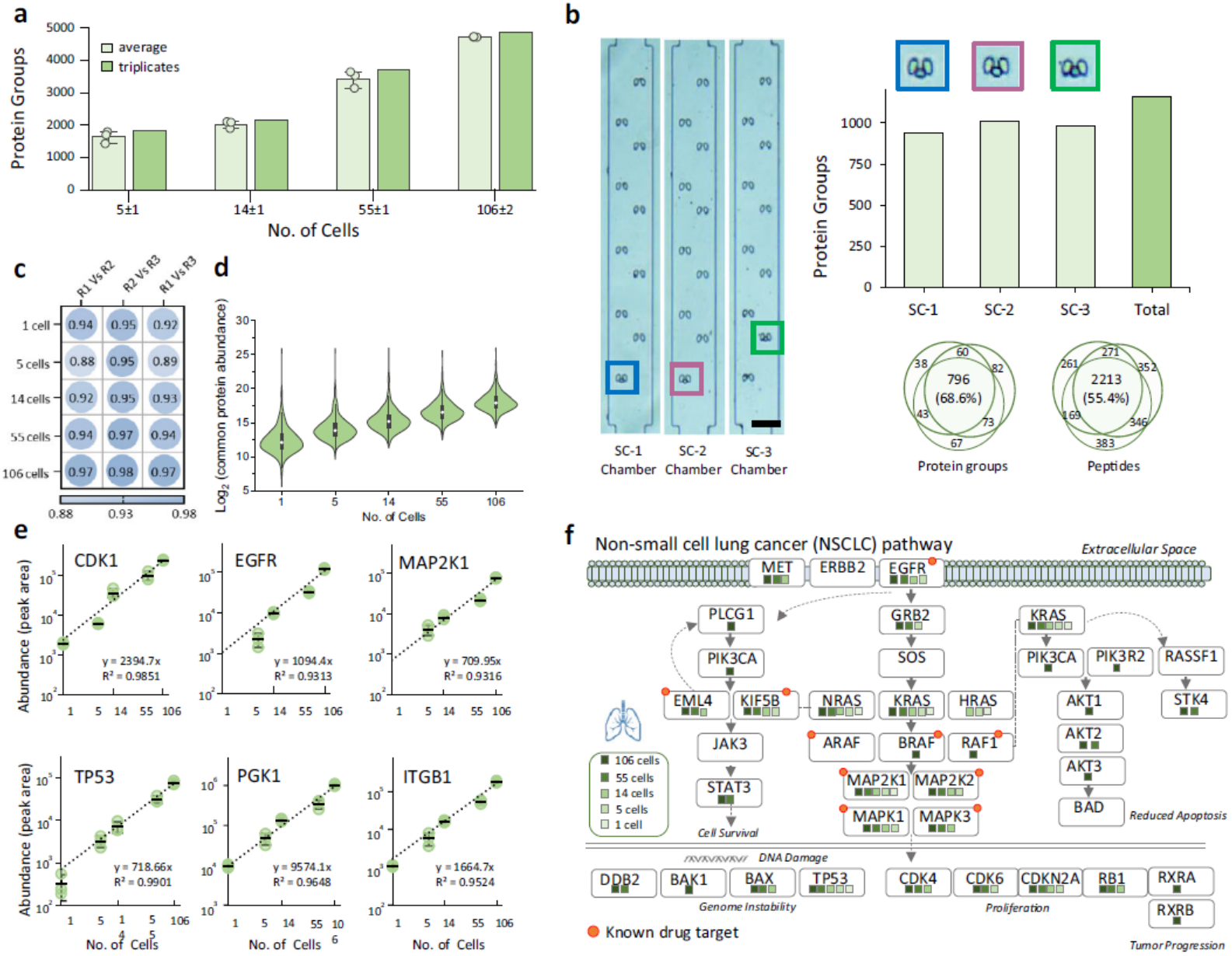
ARTICLE

https://doi.org/10.1038/s41467-021-27778-4 OPEN

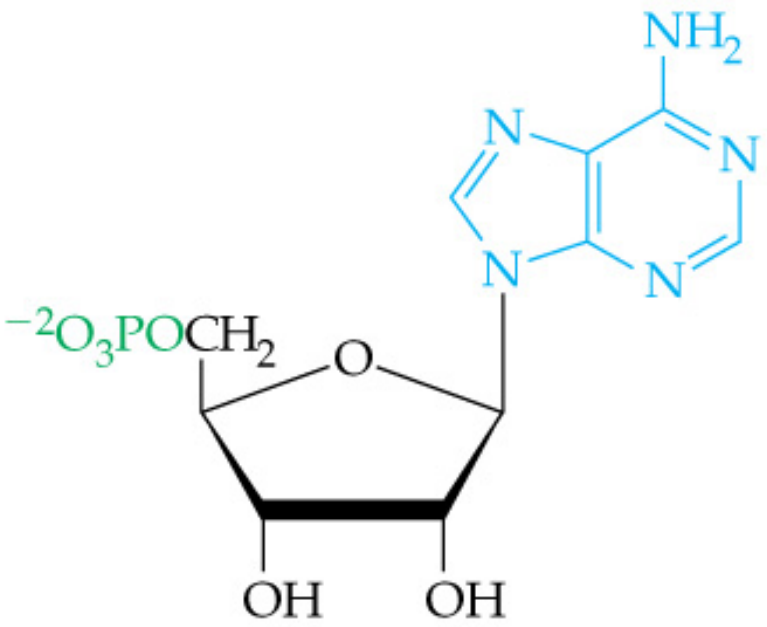
# Streamlined single-cell proteomics by an integrated microfluidic chip and data-independent acquisition mass spectrometry

Sofani Tafesse Gebreyesus[1,2,3,9], Asad Ali Siyal[1,4,5,9], Reta Birhanu Kitata[1], Eric Sheng-Wen Chen[1], Bayarmaa Enkhbayar[4,6], Takashi Angata[6], Kuo-I Lin[7], Yu-Ju Chen[1,3,4,8✉] & Hsiung-Lin Tu[1,2,4,8✉]

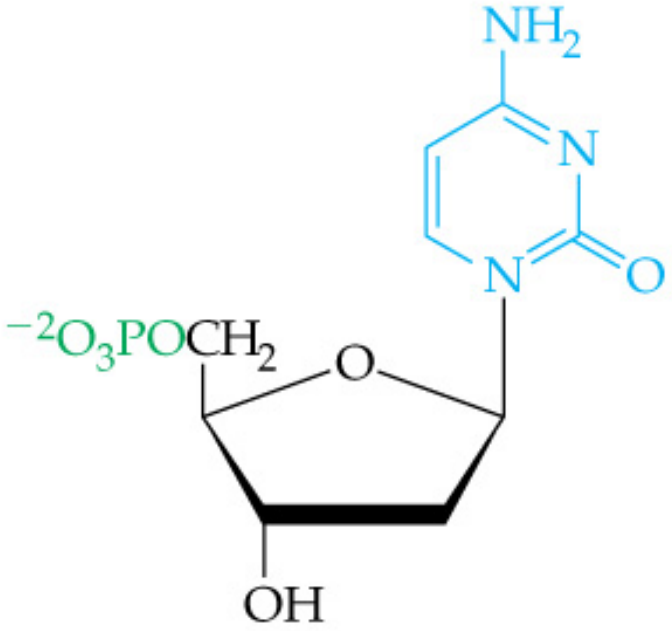



a
Protein Groups
average
triplicates
No. of Cells
5±1
14±1
55±1
106±2
b
SC-1 Chamber
SC-2 Chamber
SC-3 Chamber
Protein Groups
SC-1
SC-2
SC-3
Total
796 (68.6%)
2213 (55.4%)
Protein groups
Peptides
c
R1 Vs R2
R2 Vs R3
R1 Vs R3
1 cell
5 cells
14 cells
55 cells
106 cells
d
Log2 (common protein abundance)
No. of Cells
e
CDK1
EGFR
MAP2K1
TP53
PGK1
ITGB1
Abundance (peak area)
No. of Cells
f
Non-small cell lung cancer (NSCLC) pathway
Extracellular Space
MET
ERBB2
EGFR
PLCG1
GRB2
KRAS
PIK3CA
SOS
PIK3CA
PIK3R2
RASSF1
EML4
KIF5B
NRAS
KRAS
HRAS
AKT1
STK4
JAK3
ARAF
BRAF
RAF1
AKT2
AKT3
STAT3
MAP2K1
MAP2K2
BAD
MAPK1
MAPK3
Reduced Apoptosis
Cell Survival
DNA Damage
DDB2
BAK1
BAX
TP53
CDK4
CDK6
CDKN2A
RB1
RXRA
RXRB
Genome Instability
Proliferation
Tumor Progression
Known drug target

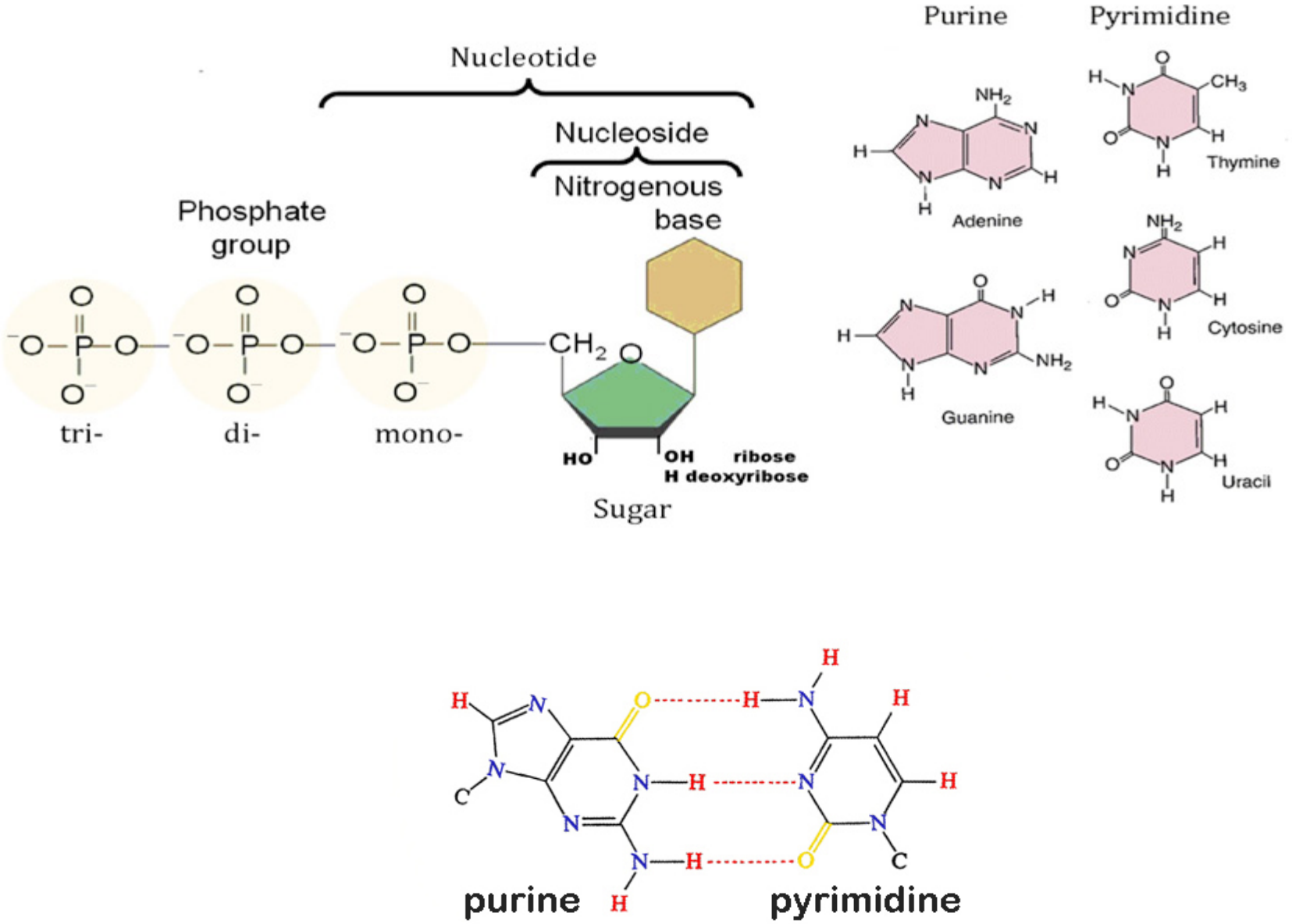
- In RNA, the sugar is ribose.
- In DNA, the sugar is deoxyribose.



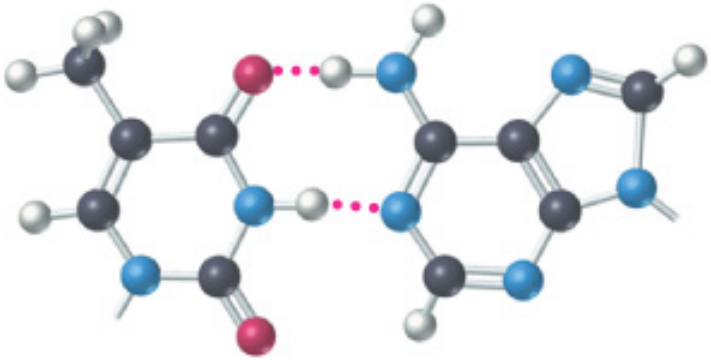

Adenosine 5′-monophosphate (AMP)
(a ribonucleotide)

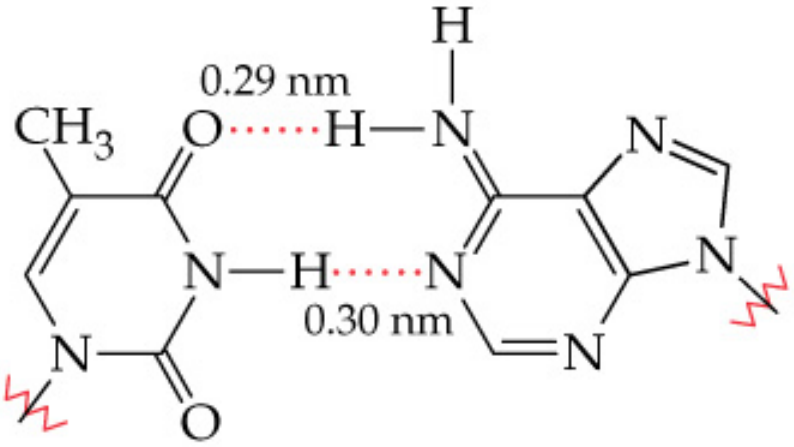



Deoxycytidine 5′-monophosphate (dCMP)
(a deoxyribonucleotide)

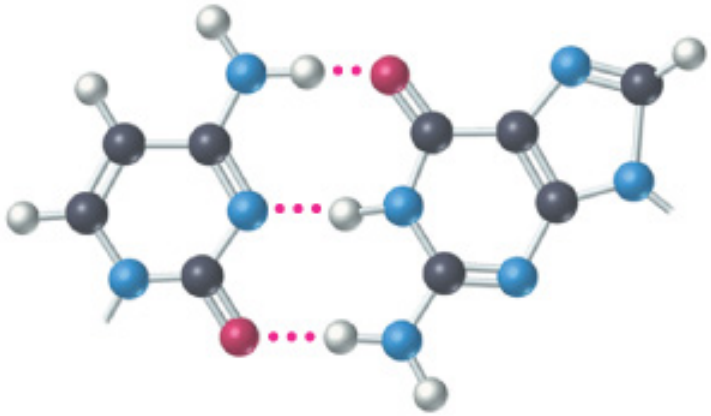

Purine
Pyrimidine
Nucleotide
Nucleoside
Nitrogenous base
Phosphate group
tri-
di-
mono-
HO
OH
H
ribose
deoxyribose
Sugar
Adenine
Thymine
Cytosine
Guanine
Uracil
purine
pyrimidine

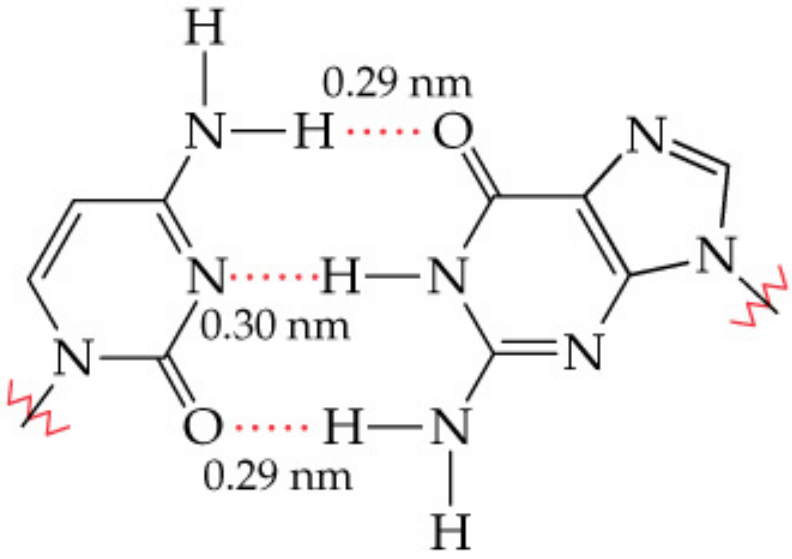
Thymine–Adenine

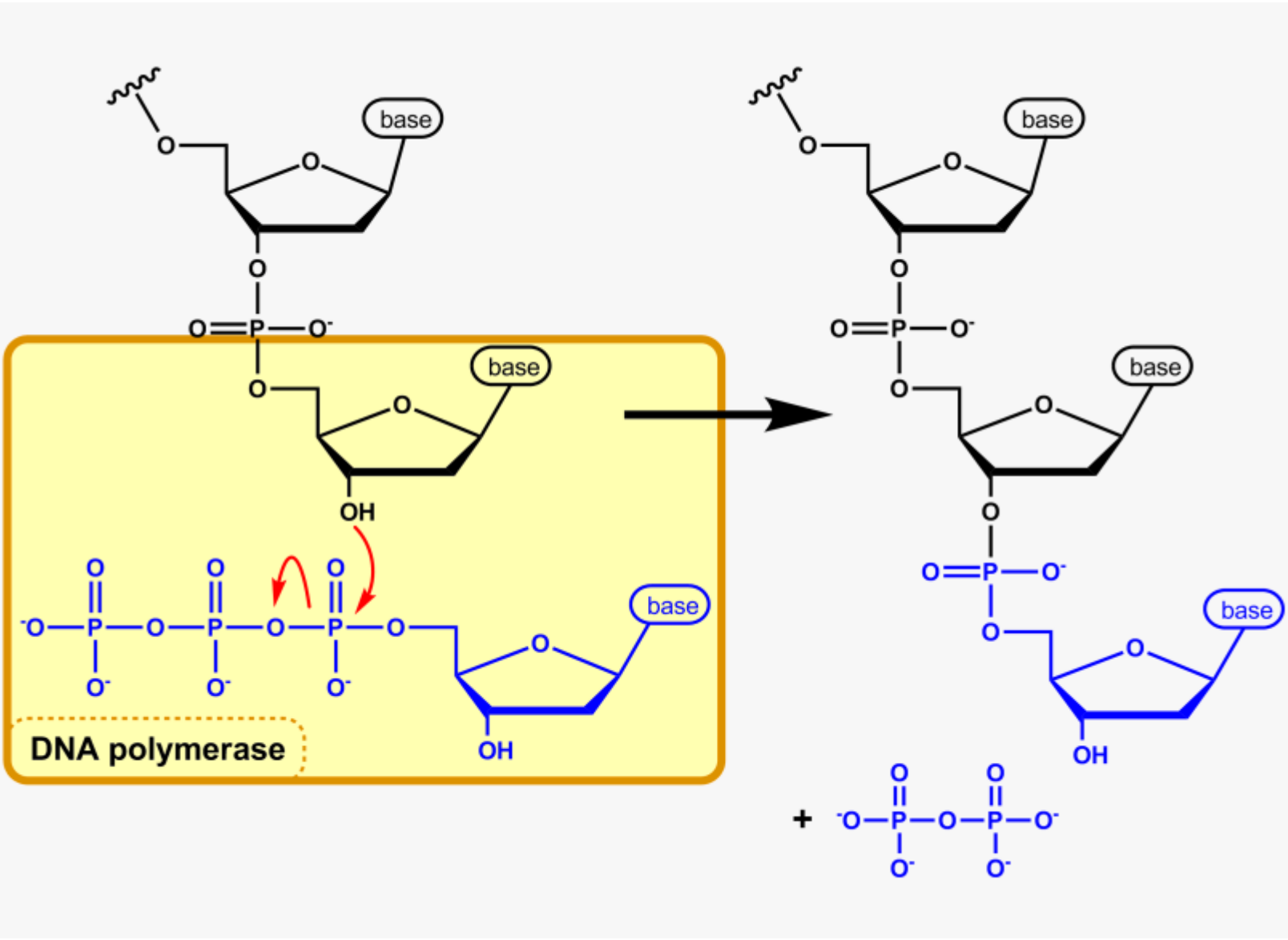


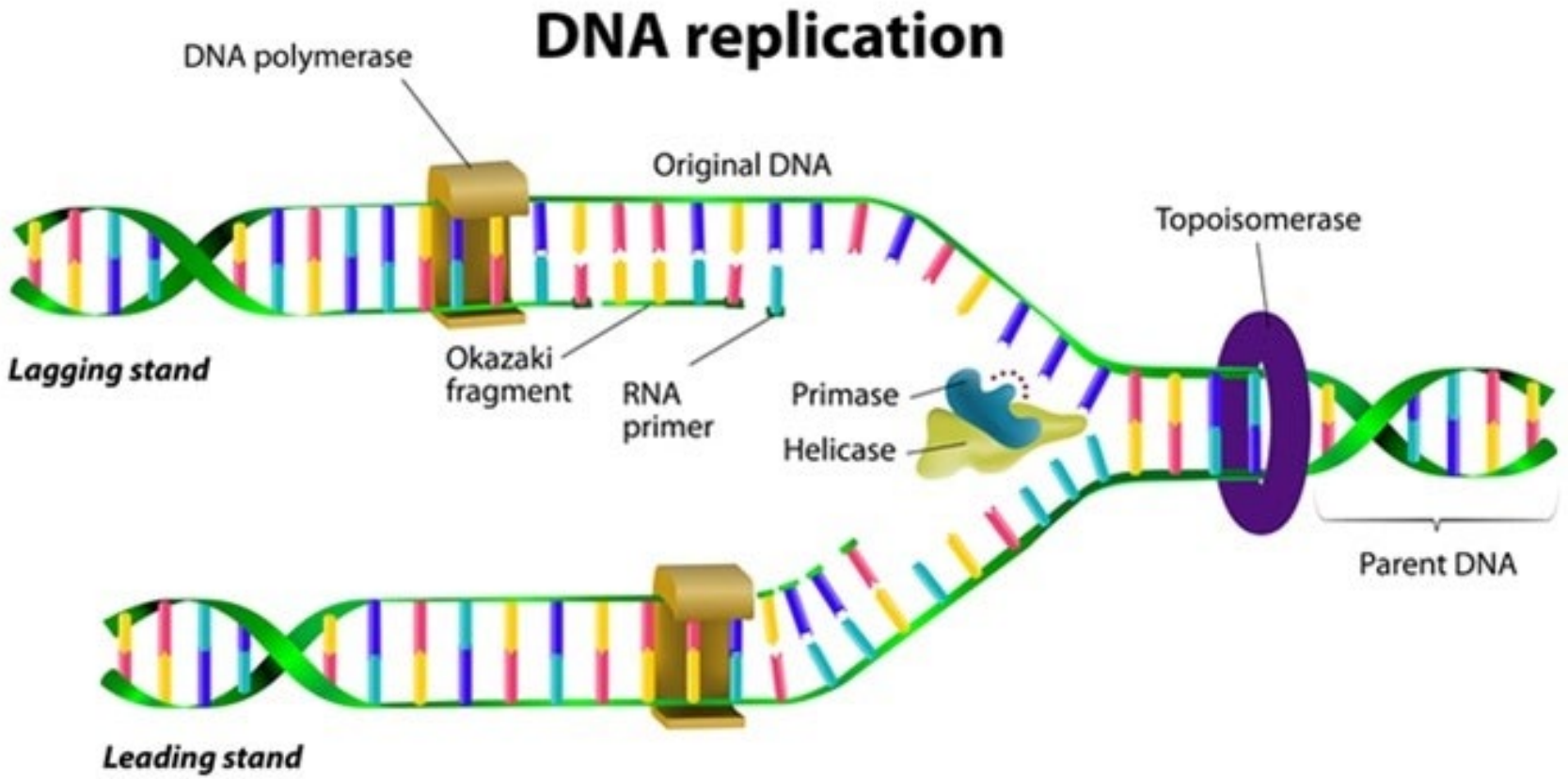



Cytosine–Guanine

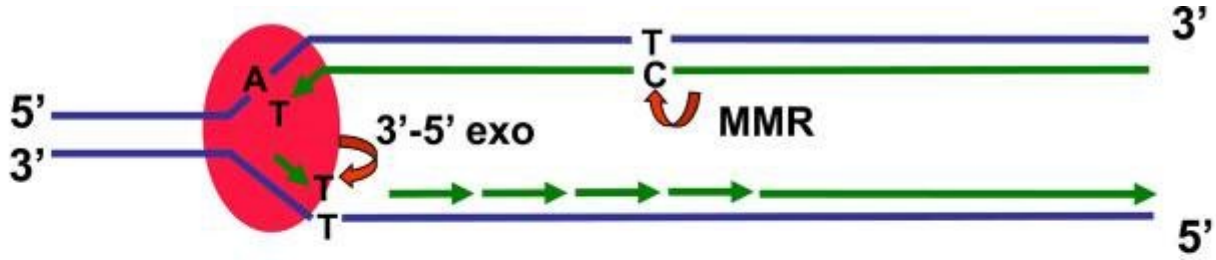





base
base
OH
base
DNA polymerase
OH
base
base
OH
+




https://youtu.be/TNKWgcFPHqw

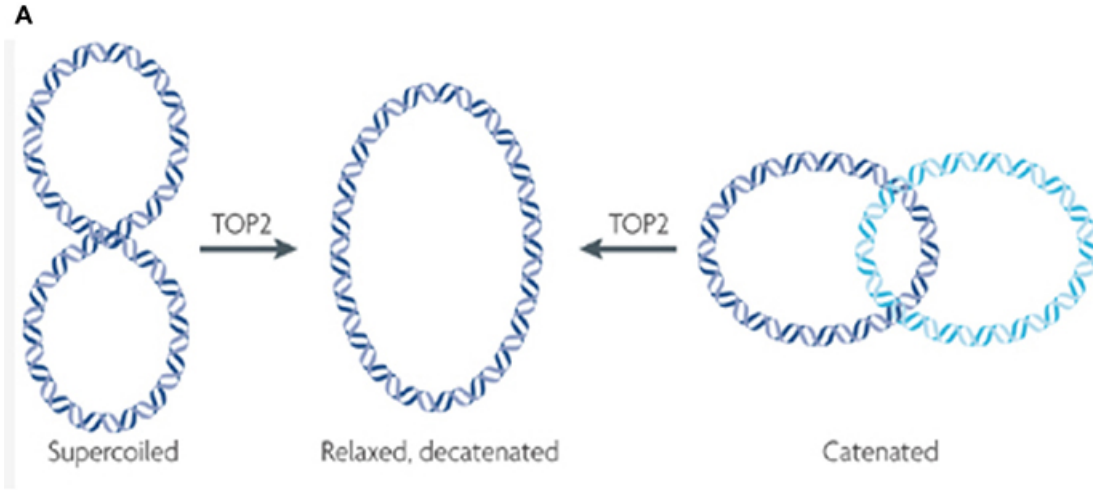


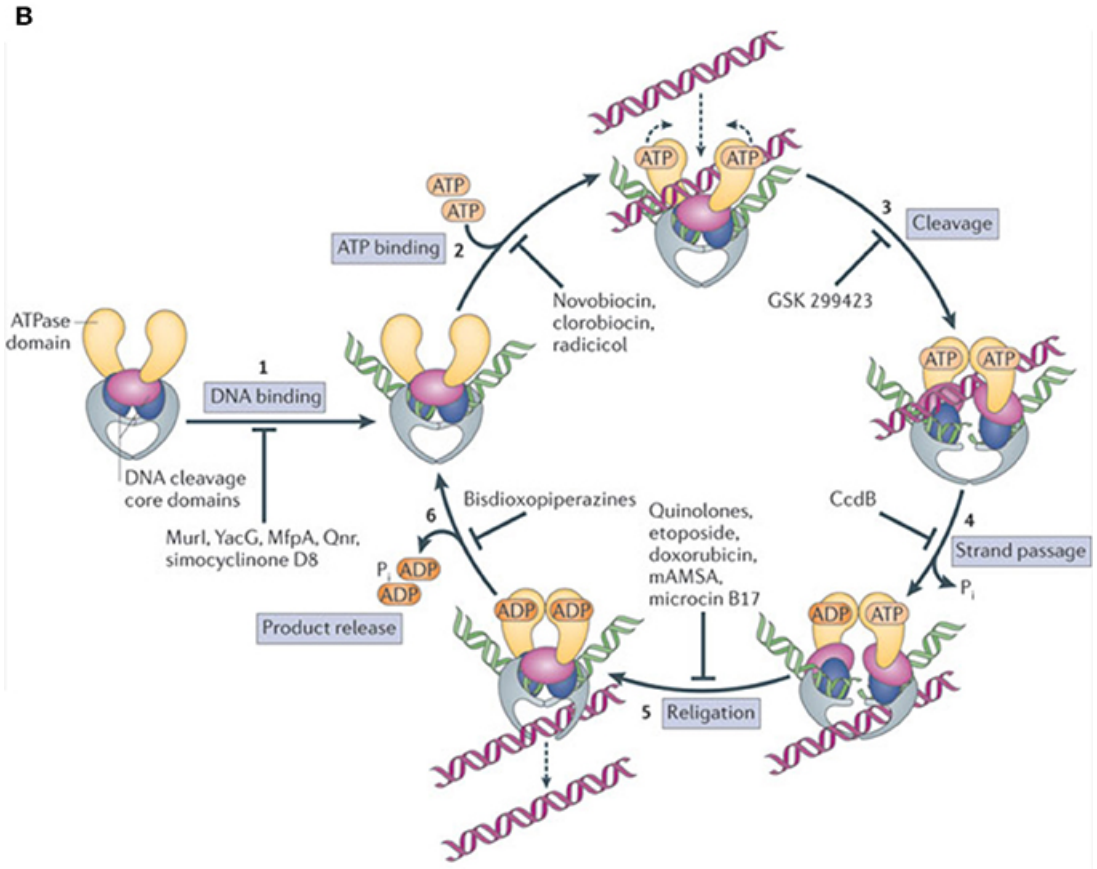

BASE SELECTION $10^{-4} – 10^{-5}$

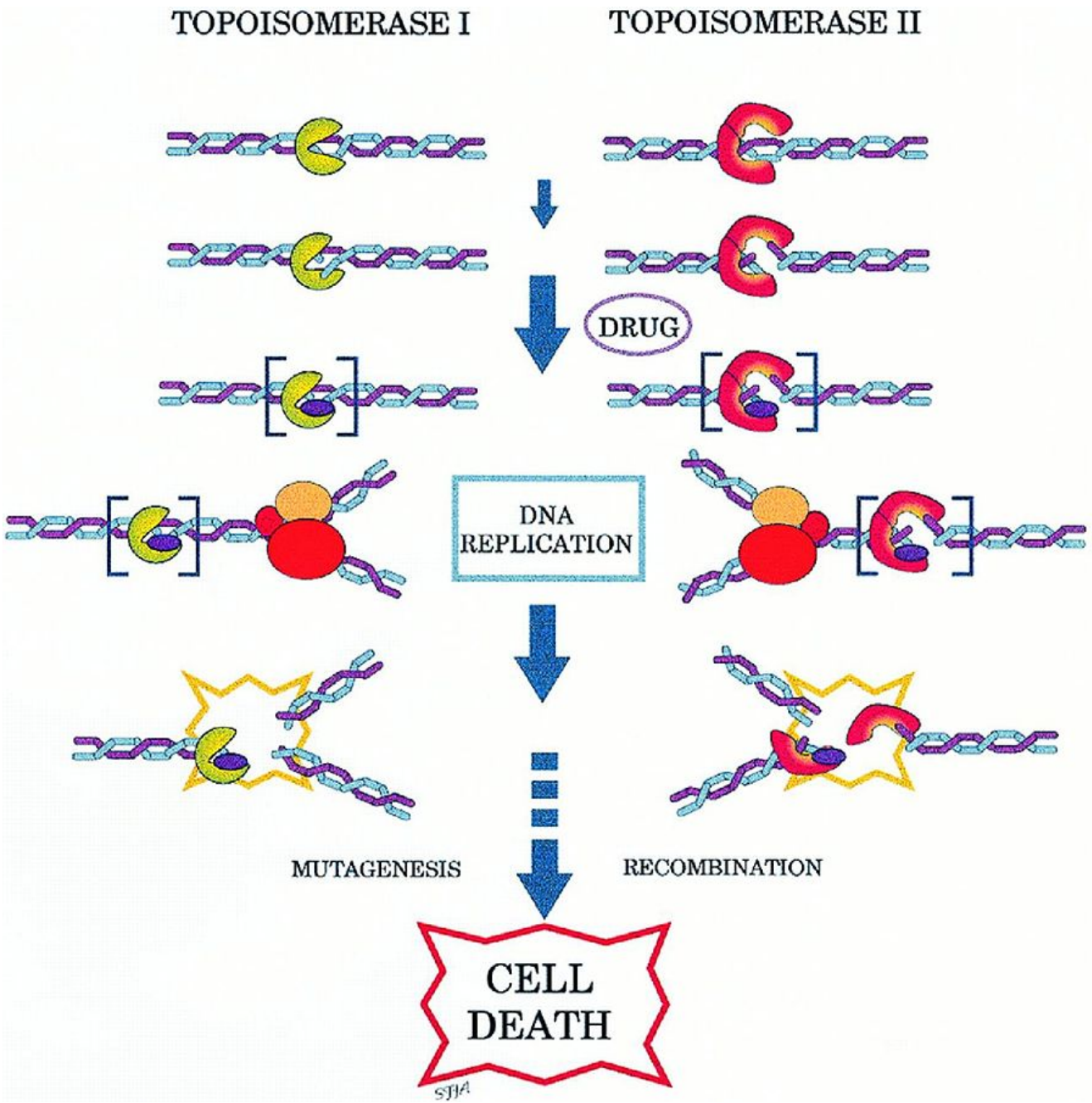
3'-5' EXONUCLEASE $10^{-2}$

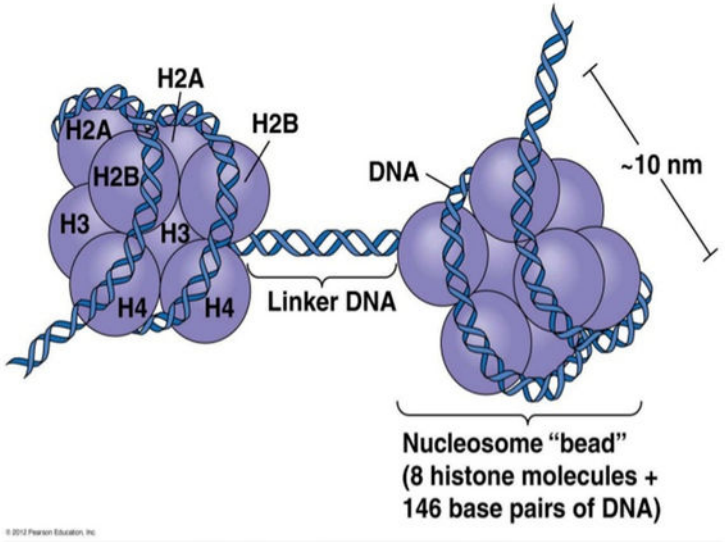
MISMATCH REPAIR $10^{-3}$

DNA REPLICATION FIDELITY $10^{-10}$

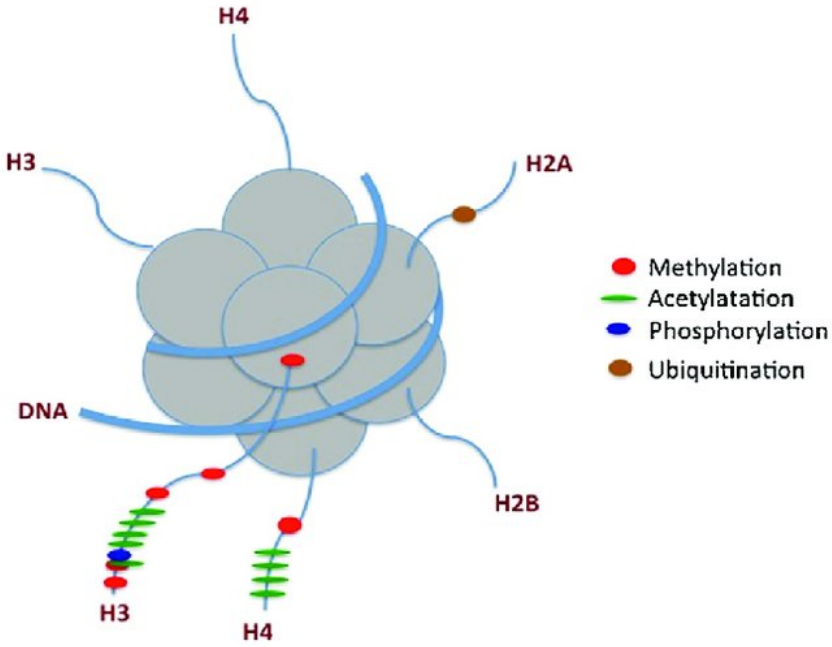

A
TOP2
TOP2
Supercoiled
Relaxed, decatenated
Catenated

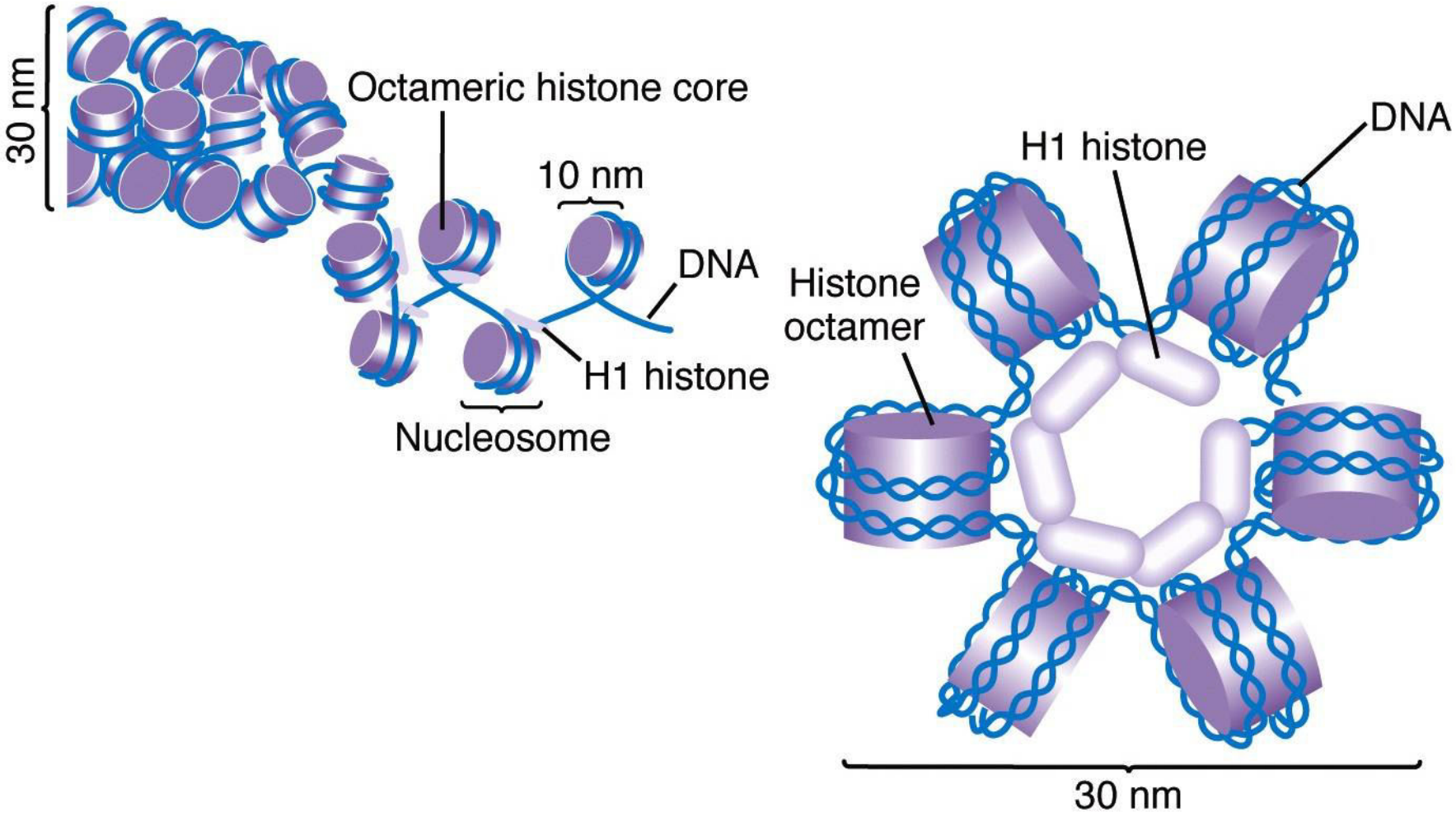

B
ATPase domain
DNA cleavage core domains
1
DNA binding
MurI, YacG, MfpA, Qnr, simocyclinone D8
ATP
ATP
ATP binding
2
Novobiocin, clorobiocin, radicicol
3
Cleavage
GSK 299423
4
Strand passage
CcdB
Pi
Quinolones, etoposide, doxorubicin, mAMSA, microcin B17
5
Religation
6
Bisdioxopiperazines
Pi
ADP
ADP
Product release

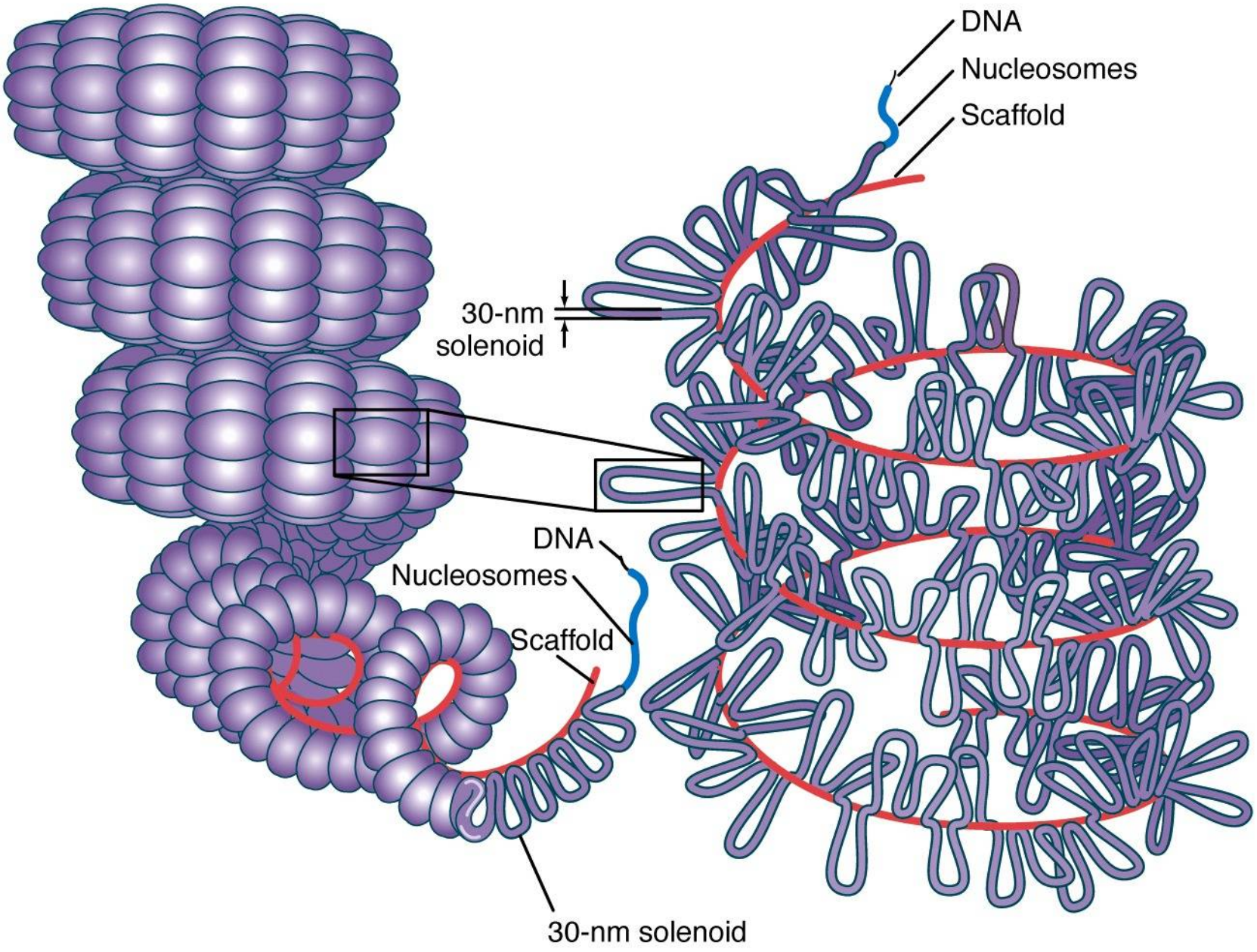

TOPOISOMERASE I
TOPOISOMERASE II
DRUG
DNA
REPLICATION
MUTAGENESIS
RECOMBINATION
CELL
DEATH
SJA

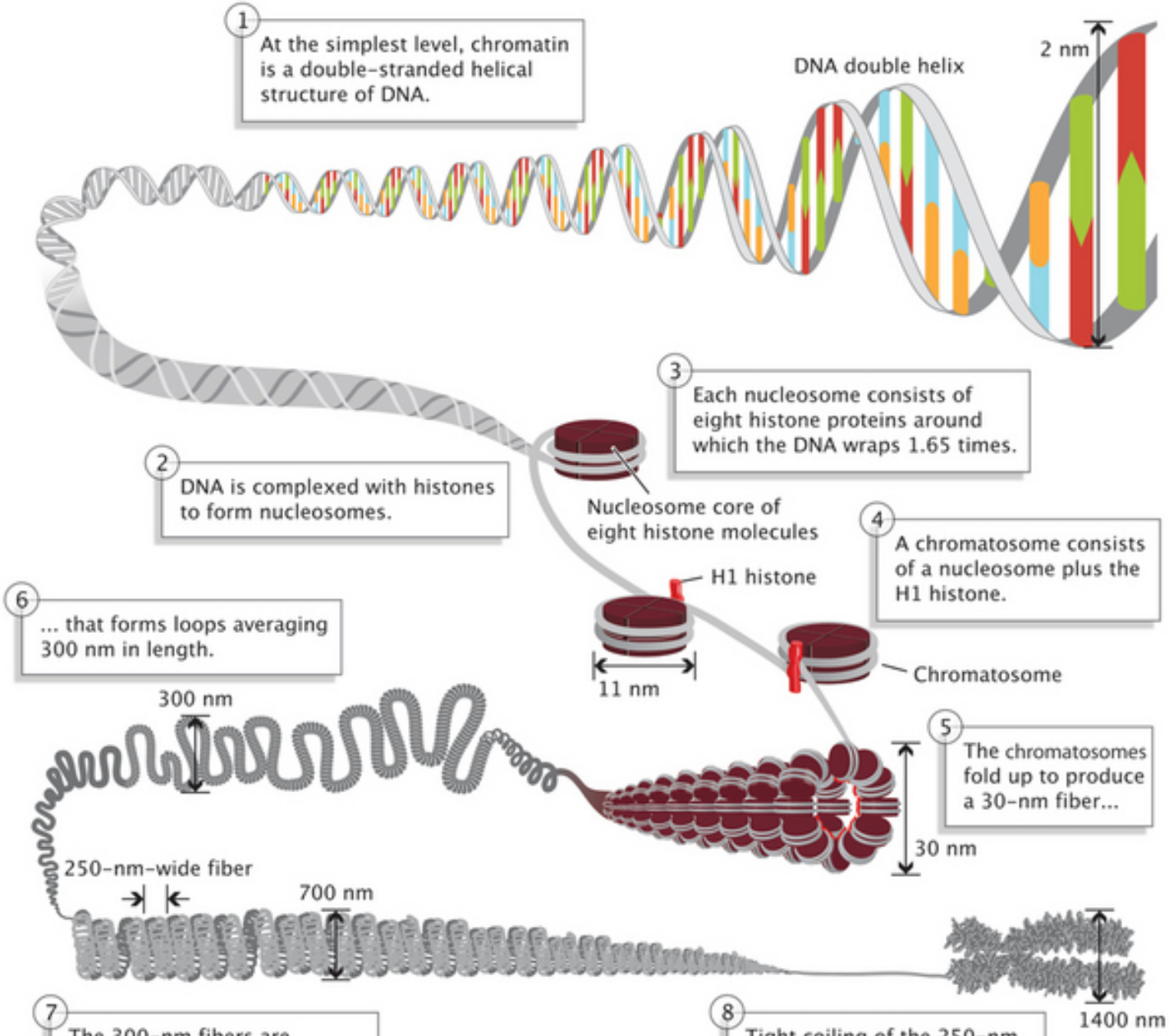


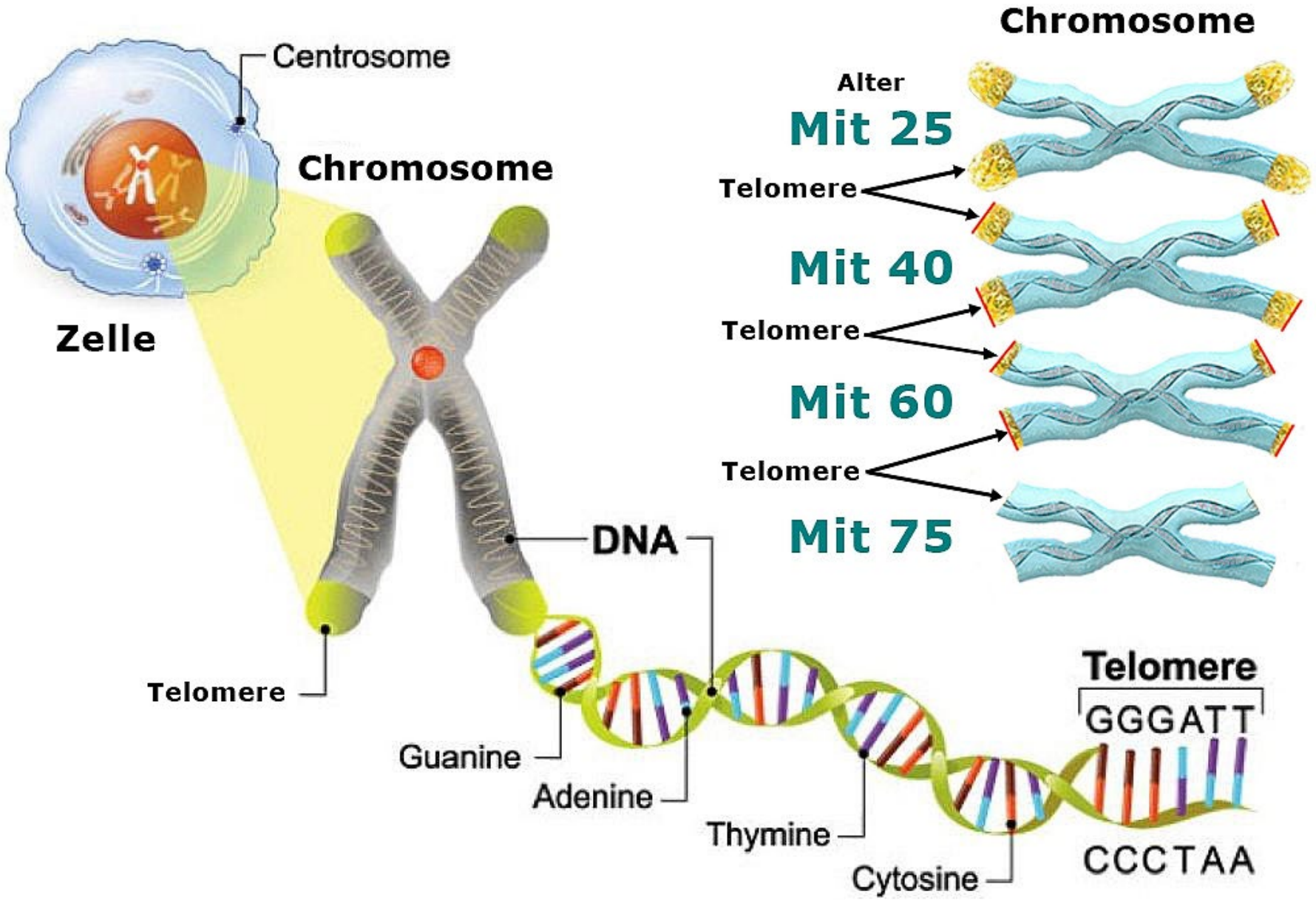

https://www.slideshare.net/jannatiftikhar/role-of-histone-in-dna-packaging

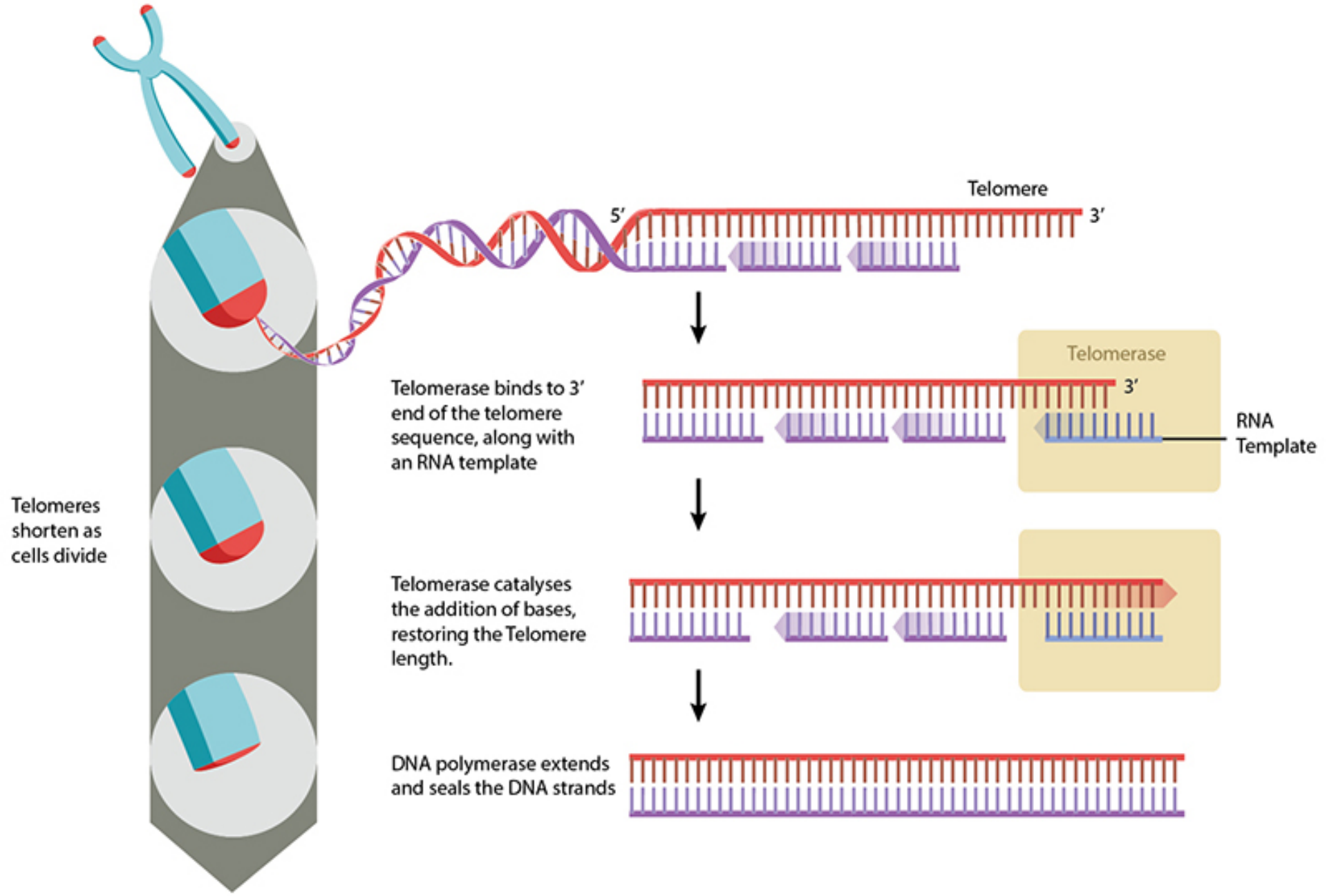



30 nm
Octameric histone core
10 nm
DNA
H1 histone
Nucleosome
DNA
H1 histone
Histone
octamer
30 nm


DNA
Nucleosomes
Scaffold
30-nm
solenoid
DNA
Nucleosomes
Scaffold
30-nm solenoid

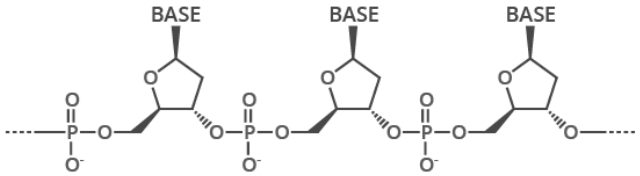

1
At the simplest level, chromatin is a double-stranded helical structure of DNA.
DNA double helix
2 nm
2
DNA is complexed with histones to form nucleosomes.
3
Each nucleosome consists of eight histone proteins around which the DNA wraps 1.65 times.
Nucleosome core of eight histone molecules
4
A chromatosome consists of a nucleosome plus the H1 histone.
H1 histone
11 nm
Chromatosome
6
... that forms loops averaging 300 nm in length.
300 nm
5
The chromatosomes fold up to produce a 30-nm fiber...
30 nm
250-nm-wide fiber
700 nm
1400 nm
7
The 300-nm fibers are compressed and folded to produce a 250-nm-wide fiber.
8
Tight coiling of the 250-nm fiber produces the chromatid of a chromosome.


Centrosome
Chromosome
Zelle
Telomere
DNA
Guanine
Adenine
Thymine
Cytosine
Chromosome
Alter
Mit 25
Telomere
Mit 40
Telomere
Mit 60
Telomere
Mit 75
Telomere
GGGATT
CCCTAA


Telomeres shorten as cells divide
Telomere
5′
3′
Telomerase binds to 3′ end of the telomere sequence, along with an RNA template
Telomerase
3′
RNA Template
Telomerase catalyses the addition of bases, restoring the Telomere length.
DNA polymerase extends and seals the DNA strands

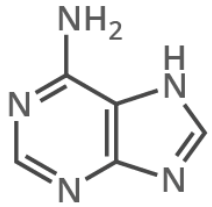
# THE CHEMICAL STRUCTURE OF DNA

## THE SUGAR PHOSPHATE 'BACKBONE'

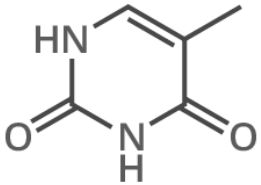



DNA is a polymer made up of units called nucleotides. The nucleotides are made of three different components: a sugar group, a phosphate group, and a base. There are four different bases: adenine, thymine, guanine and cytosine.

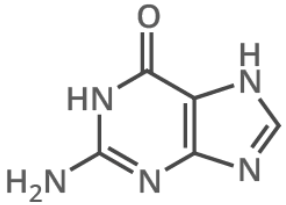
### A ADENINE



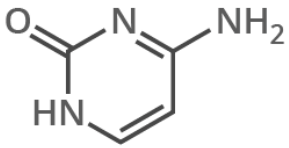
### T THYMINE



### G GUANINE

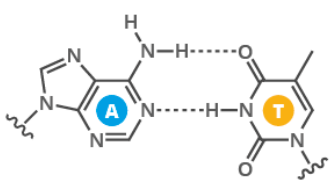


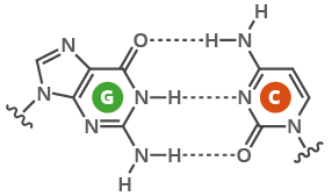
### C CYTOSINE



## WHAT HOLDS DNA STRANDS TOGETHER?

DNA strands are held together by hydrogen bonds between bases on adjacent strands. Adenine (A) always pairs with thymine (T), while guanine (G) always pairs with cytosine (C). Adenine pairs with uracil (U) in RNA.



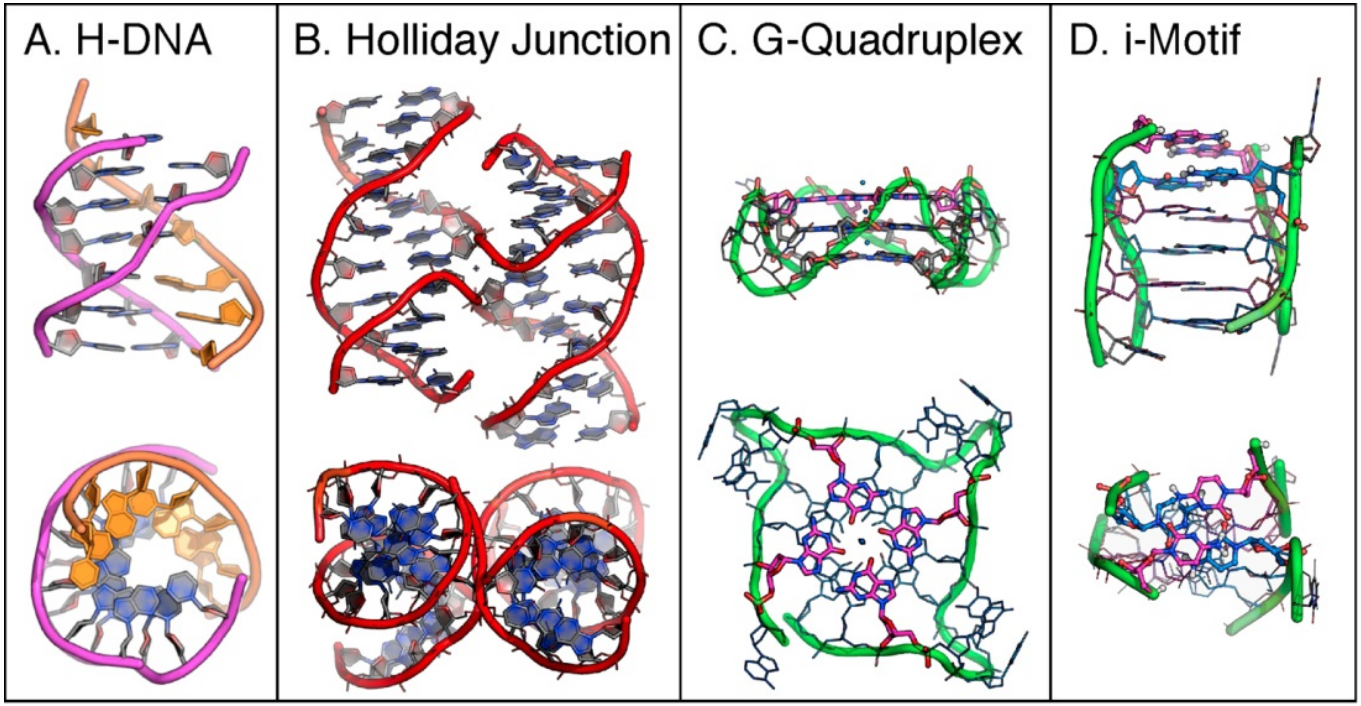

## FROM DNA TO PROTEINS

The bases on a single strand of DNA act as a code. The letters form three letter codons, which code for amino acids - the building blocks of proteins.

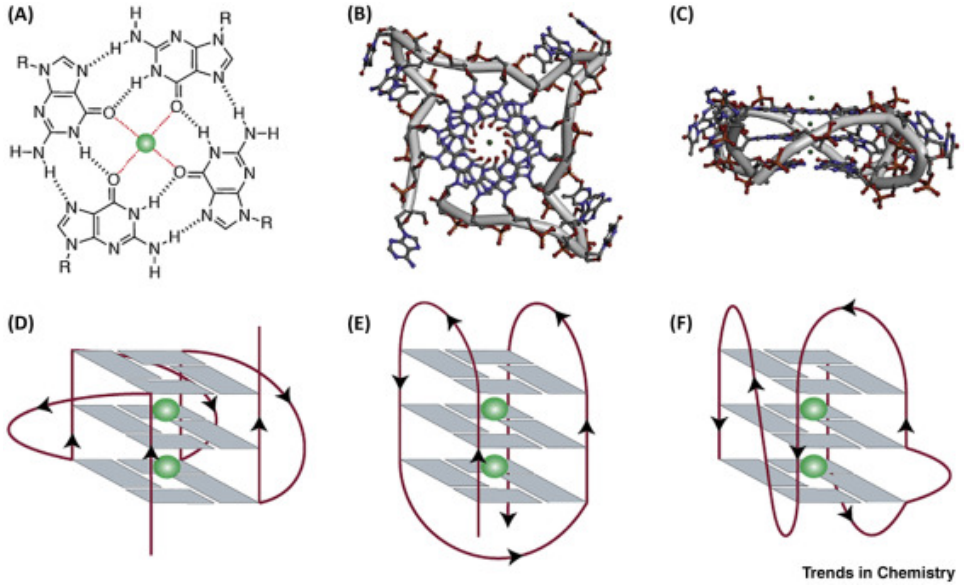

An enzyme, RNA polymerase, transcribes DNA into mRNA (messenger ribonucleic acid). It splits apart the two strands that form the double helix, then reads a strand and copies the sequence of nucleotides. The only difference between the RNA and the original DNA is that in the place of thymine (T), another base with a similar structure is used: uracil (U).

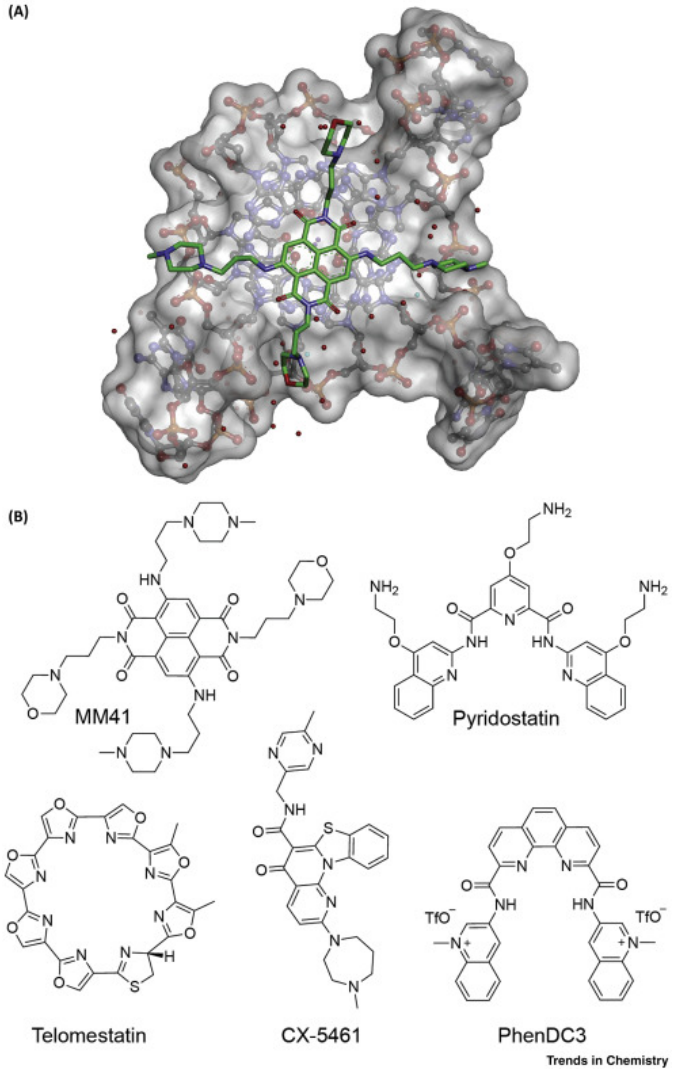

In multicellular organisms, the mRNA carries genetic code out of the cell nucleus, to the cytoplasm. Here, protein synthesis takes place. 'Translation' is the process of turning the mRNA's 'code' into proteins. Molecules called ribosomes carry out this process, building up proteins from the amino acids coded for.

# Triple and Quadruple Strained DNA

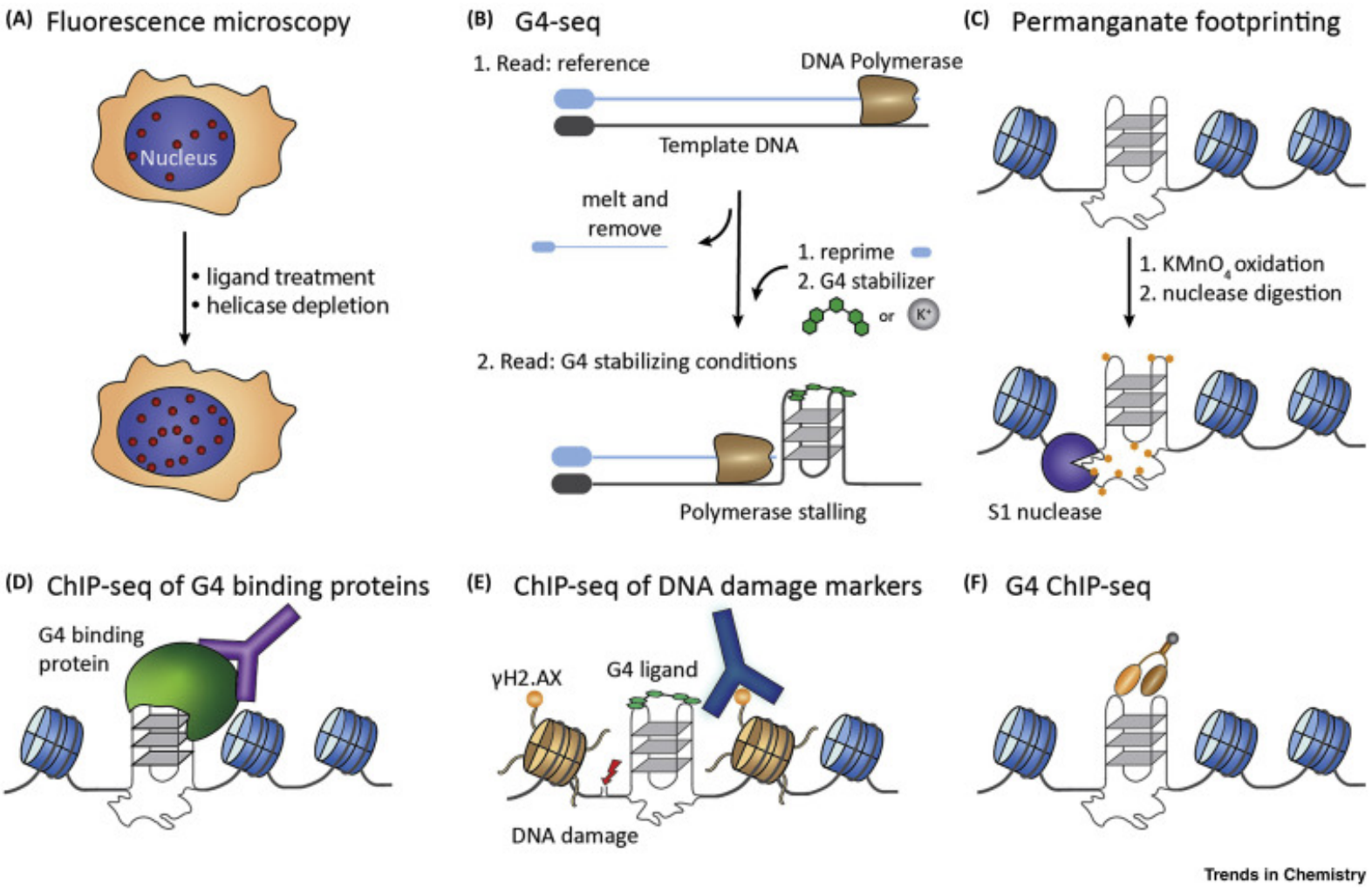


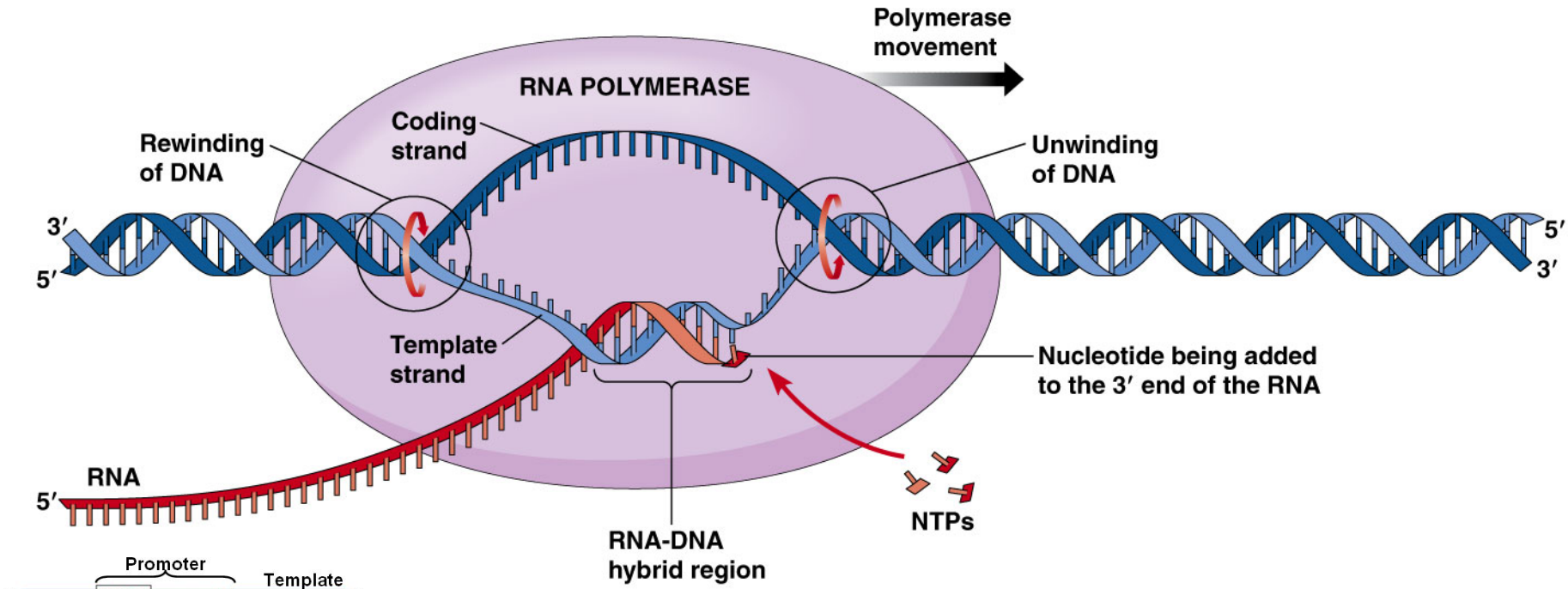
# DNA G-quadruplex (G4)

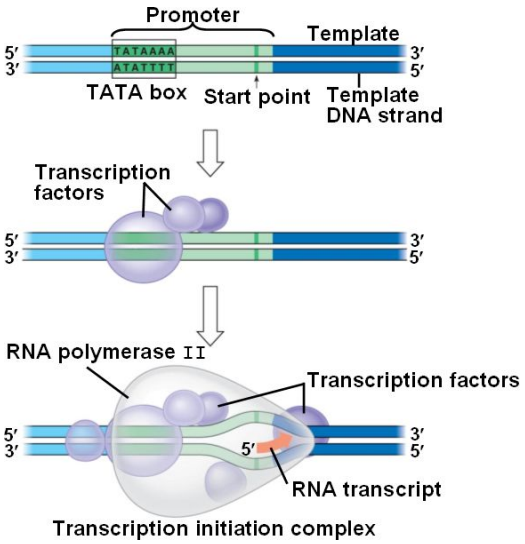


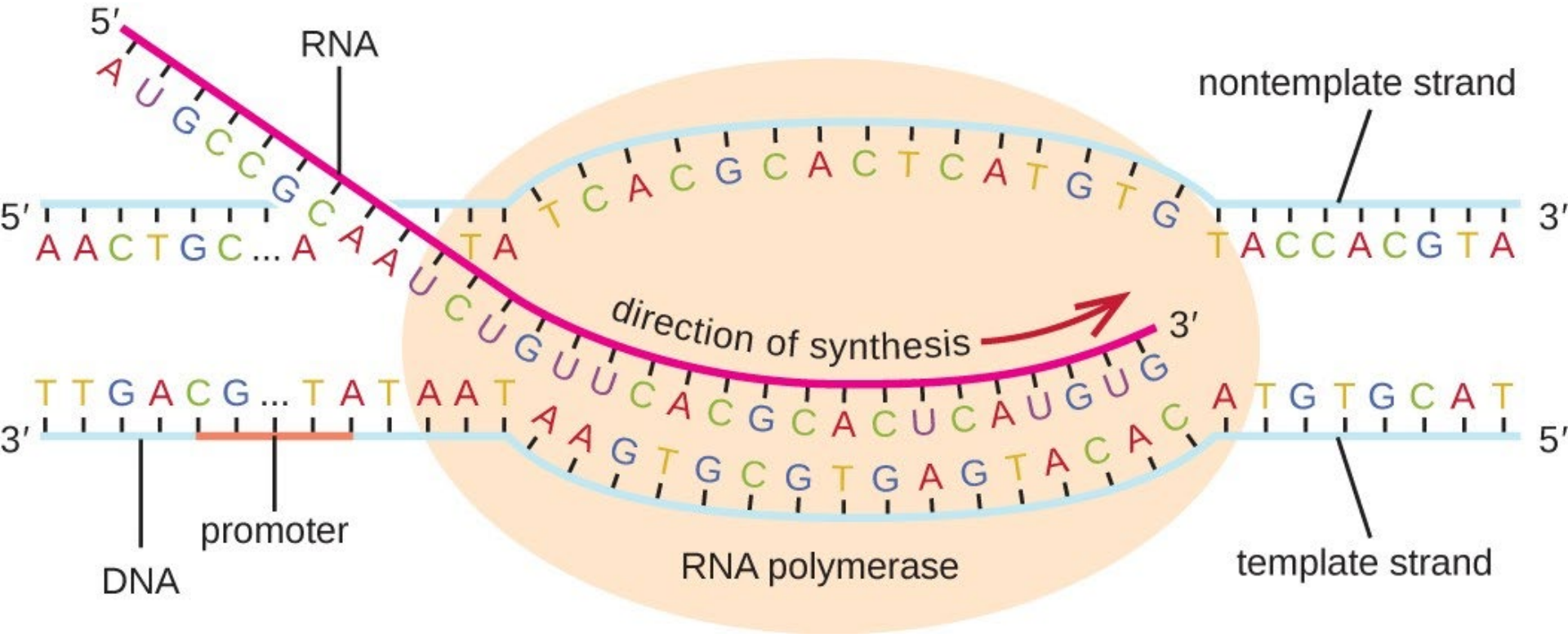



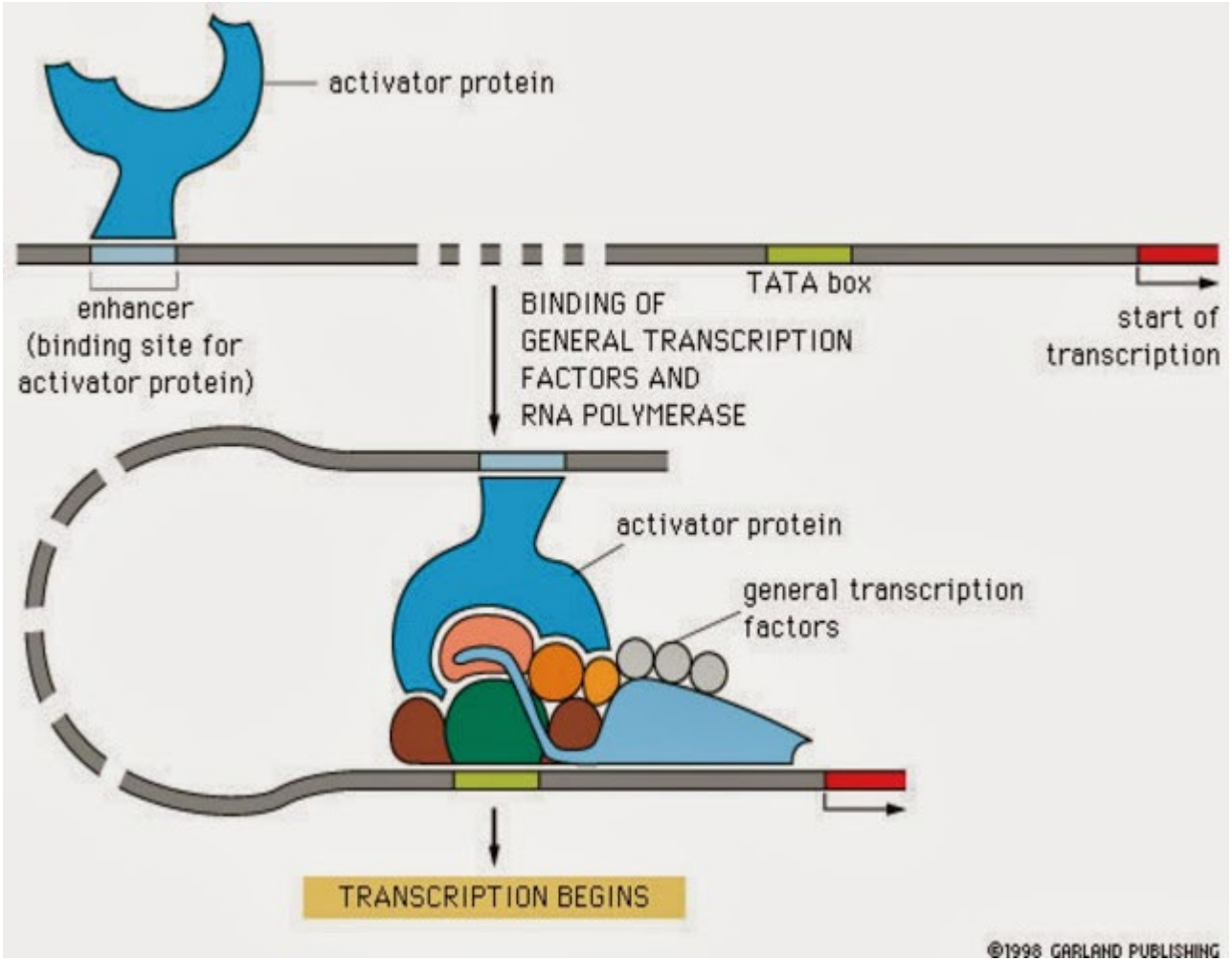

Trends in Chemistry, February 2020, Vol. 2, No. 2 P123

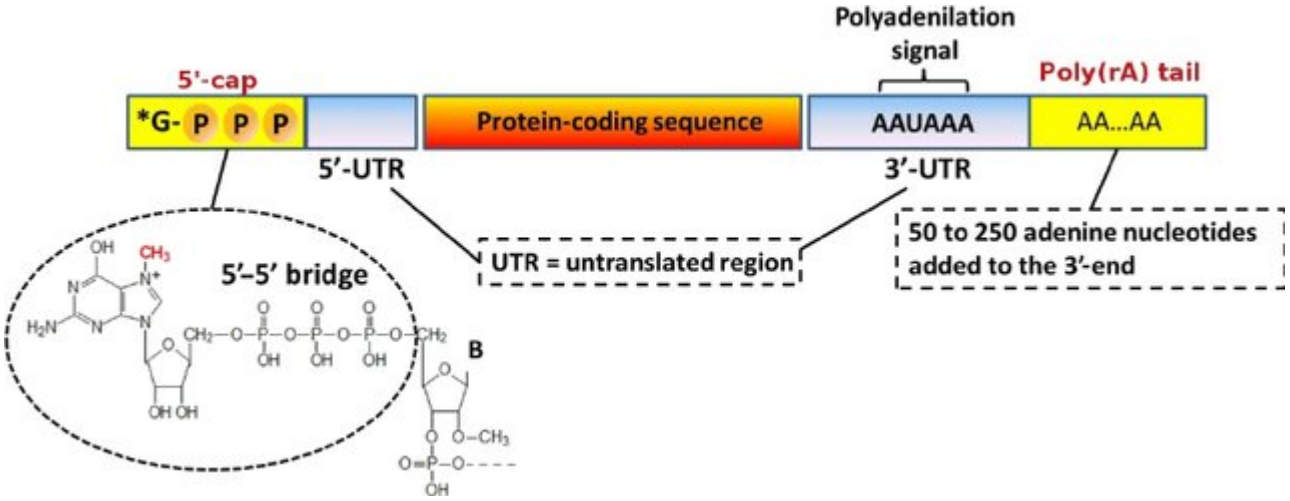

(A) Fluorescence microscopy
Nucleus
• ligand treatment
• helicase depletion
(B) G4-seq
1. Read: reference
DNA Polymerase
Template DNA
melt and remove
1. reprime
2. G4 stabilizer
or
K+
2. Read: G4 stabilizing conditions
Polymerase stalling
(C) Permanganate footprinting
1. KMnO4 oxidation
2. nuclease digestion
S1 nuclease
(D) ChIP-seq of G4 binding proteins
G4 binding protein
(E) ChIP-seq of DNA damage markers
γH2.AX
G4 ligand
DNA damage
(F) G4 ChIP-seq
Trends in Chemistry

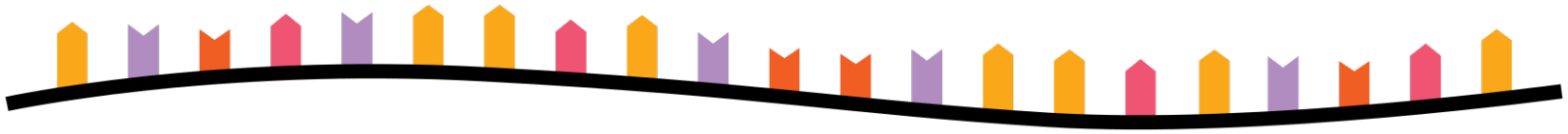

Polymerase movement
RNA POLYMERASE
Coding strand
Rewinding of DNA
Unwinding of DNA
3′
5′
5′
3′
Template strand
Nucleotide being added to the 3′ end of the RNA
RNA
5′
NTPs
RNA-DNA hybrid region


Promoter
Template
5′
3′
TATAAAA
ATATTTT
3′
5′
TATA box
Start point
Template DNA strand
Transcription factors
RNA polymerase II
Transcription factors
RNA transcript
Transcription initiation complex


5′
RNA
A U G C C G C A A U C U G U U U C A C G C A C U C A U G U G
nontemplate strand
5′
A A C T G C ... A T A
T C A C G C A C T C A T G T G
T A C C A C G T A
3′
direction of synthesis
3′
T T G A C G ... T A T A A T
A A G T G C G T G A G T A C A C
A T G T G C A T
3′
5′
promoter
DNA
RNA polymerase
template strand


activator protein
enhancer
(binding site for
activator protein)
TATA box
start of
transcription
BINDING OF
GENERAL TRANSCRIPTION
FACTORS AND
RNA POLYMERASE
activator protein
general transcription
factors
TRANSCRIPTION BEGINS
©1998 GARLAND PUBLISHING


Polyadenilation signal
5'-cap
Poly(rA) tail
*G- P P P
Protein-coding sequence
AAUAAA
AA...AA
5'-UTR
3'-UTR
5'–5' bridge
UTR = untranslated region
50 to 250 adenine nucleotides added to the 3'-end
B

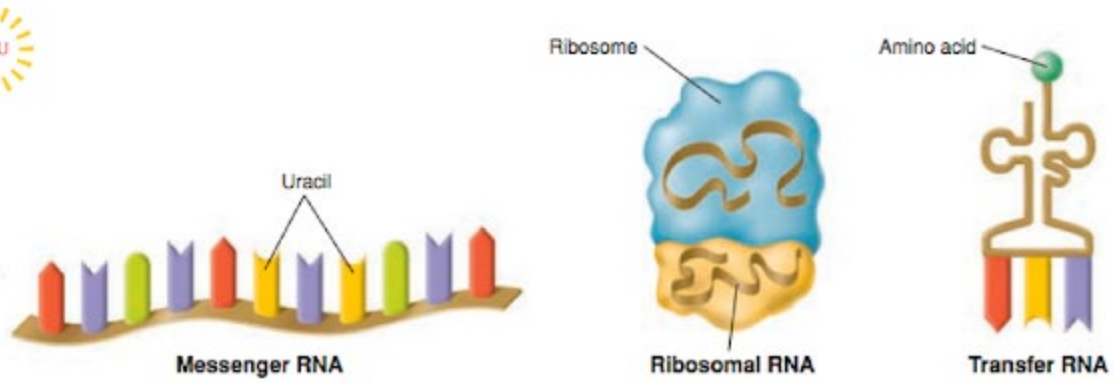


| | | | | | | | |
|---|---|---|---|---|---|---|---|
| **Base** | G C U | A C G | G A G | C U U | C G G | A G C | U A G |
| Codon | Codon 1 | Codon 2 | Codon 3 | Codon 4 | Codon 5 | Codon 6 | Codon 7 |
| **Aminoacid** | **Alanine** | **Threonine** | **Glutamate** | **Leucine** | **Arginine** | **Serine** | **Stop** |

**Second letter**

| First letter | U | C | A | G | Third letter |
|---|---|---|---|---|---|
| **U** | UUU, UUC: Phe<br>UUA, UUG: Leu | UCU, UCC, UCA, UCG: Ser | UAU, UAC: Tyr<br>**UAA Stop**<br>**UAG Stop** | UGU, UGC: Cys<br>**UGA Stop**<br>UGG Trp | U C A G |
| **C** | CUU, CUC, CUA, CUG: Leu | CCU, CCC, CCA, CCG: Pro | CAU, CAC: His<br>CAA, CAG: Gln | CGU, CGC, CGA, CGG: Arg | U C A G |
| **A** | AUU, AUC, AUA: Ile<br>**AUG Met** | ACU, ACC, ACA, ACG: Thr | AAU, AAC: Asn<br>AAA, AAG: Lys | AGU, AGC: Ser<br>AGA, AGG: Arg | U C A G |
| **G** | GUU, GUC, GUA, GUG: Val | GCU, GCC, GCA, GCG: Ala | GAU, GAC: Asp<br>GAA, GAG: Glu | GGU, GGC, GGA, GGG: Gly | U C A G |

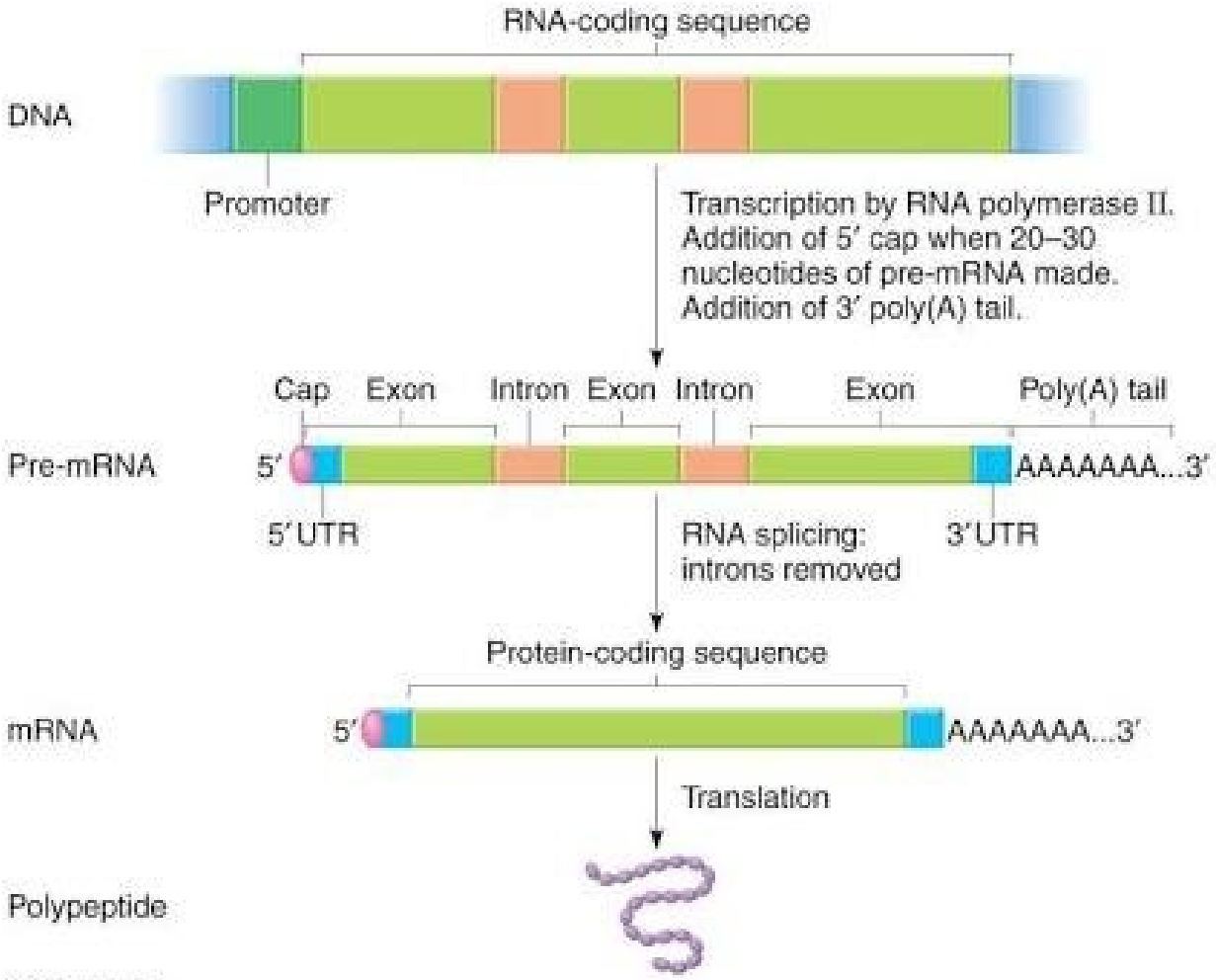
# 11 DIFFERENT TYPES OF RNA IN A CELL

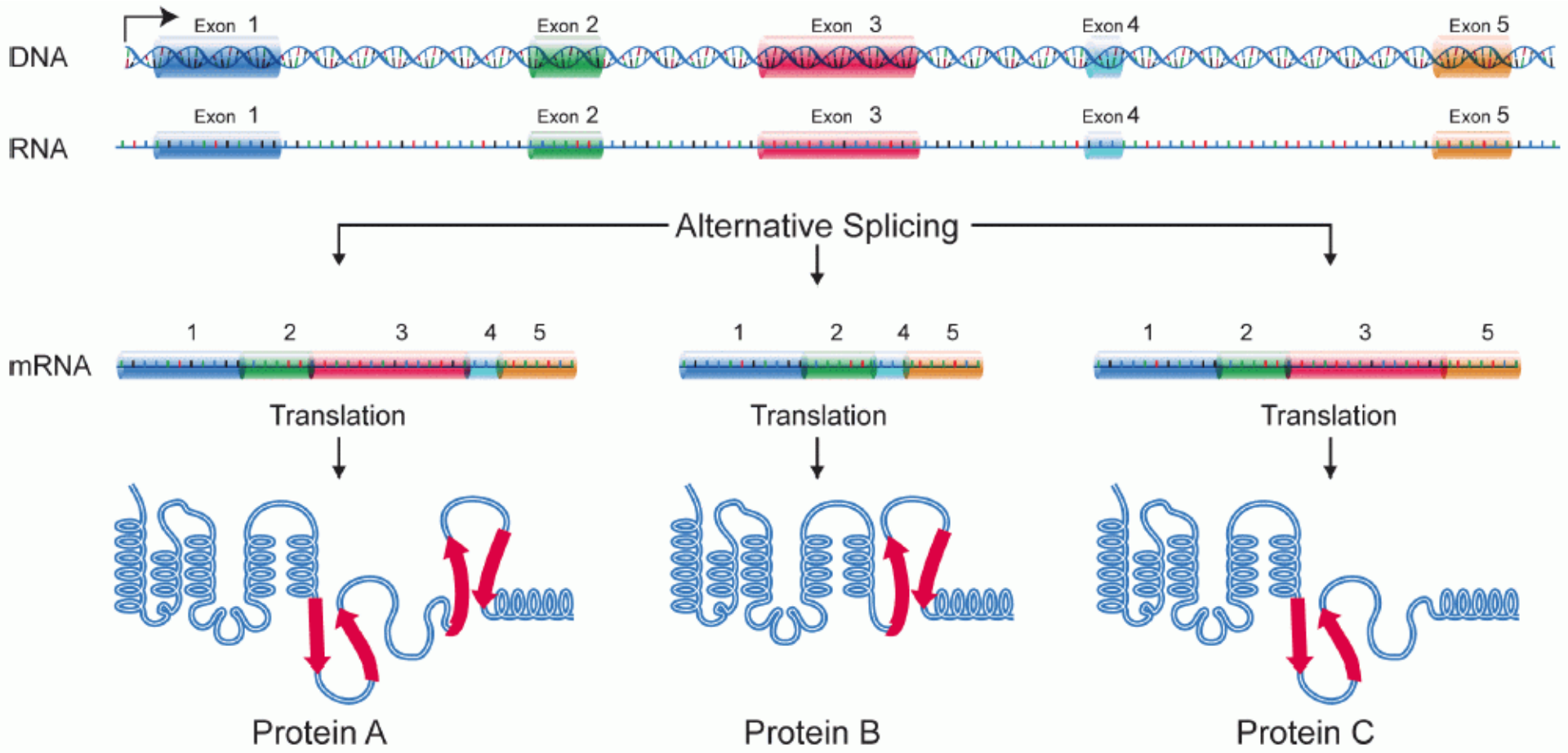



WWW.BIOLOGYEXAMS4U.COM

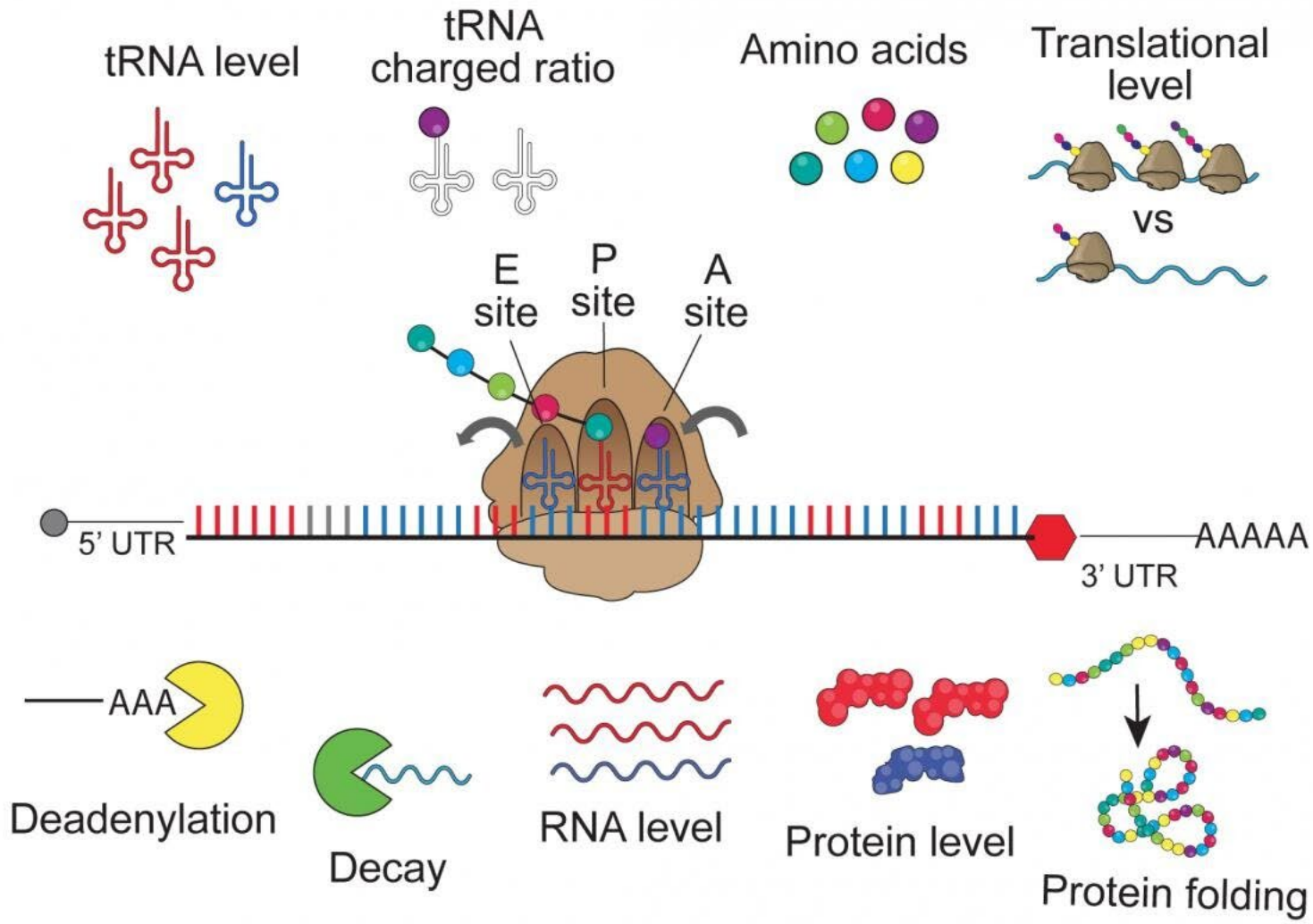
# Post Transcription Modification of RNA

1. RNA capping
2. PolyA tail
3. Splicing



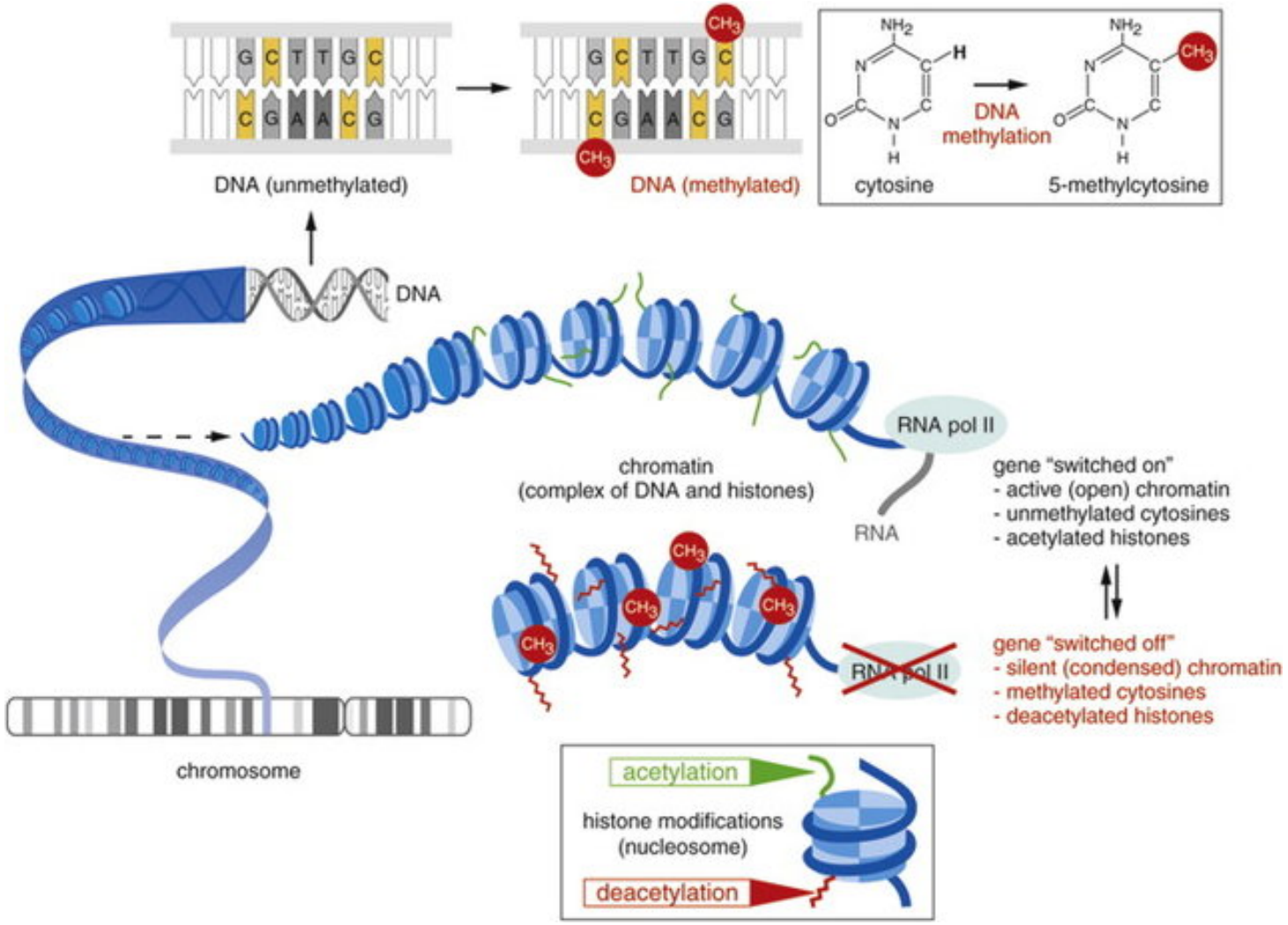

DNA
Exon 1
Exon 2
Exon 3
Exon 4
Exon 5
RNA
Exon 1
Exon 2
Exon 3
Exon 4
Exon 5
Alternative Splicing
mRNA
1 2 3 4 5
1 2 4 5
1 2 3 5
Translation
Translation
Translation
Protein A
Protein B
Protein C

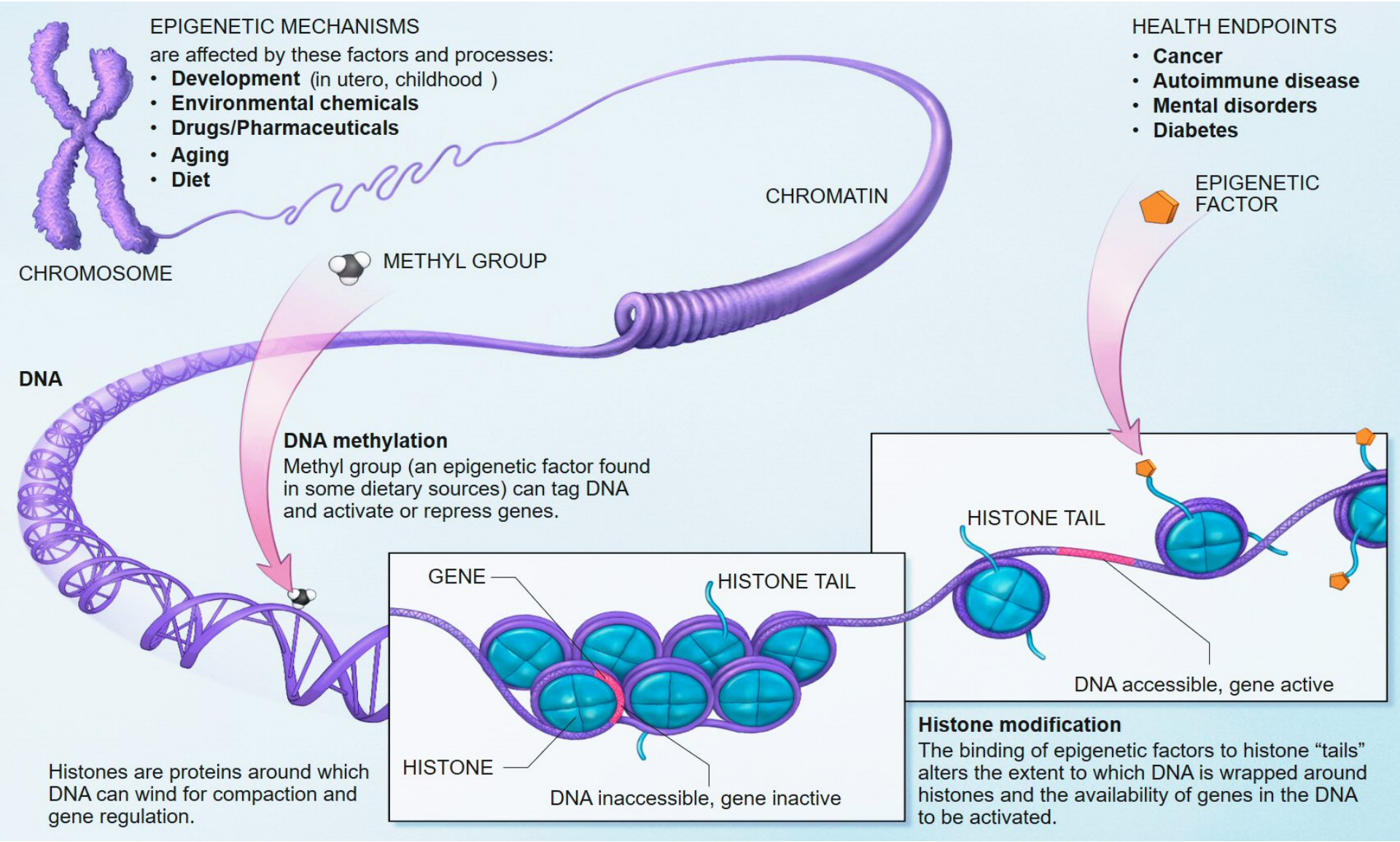
## Upstream regulator



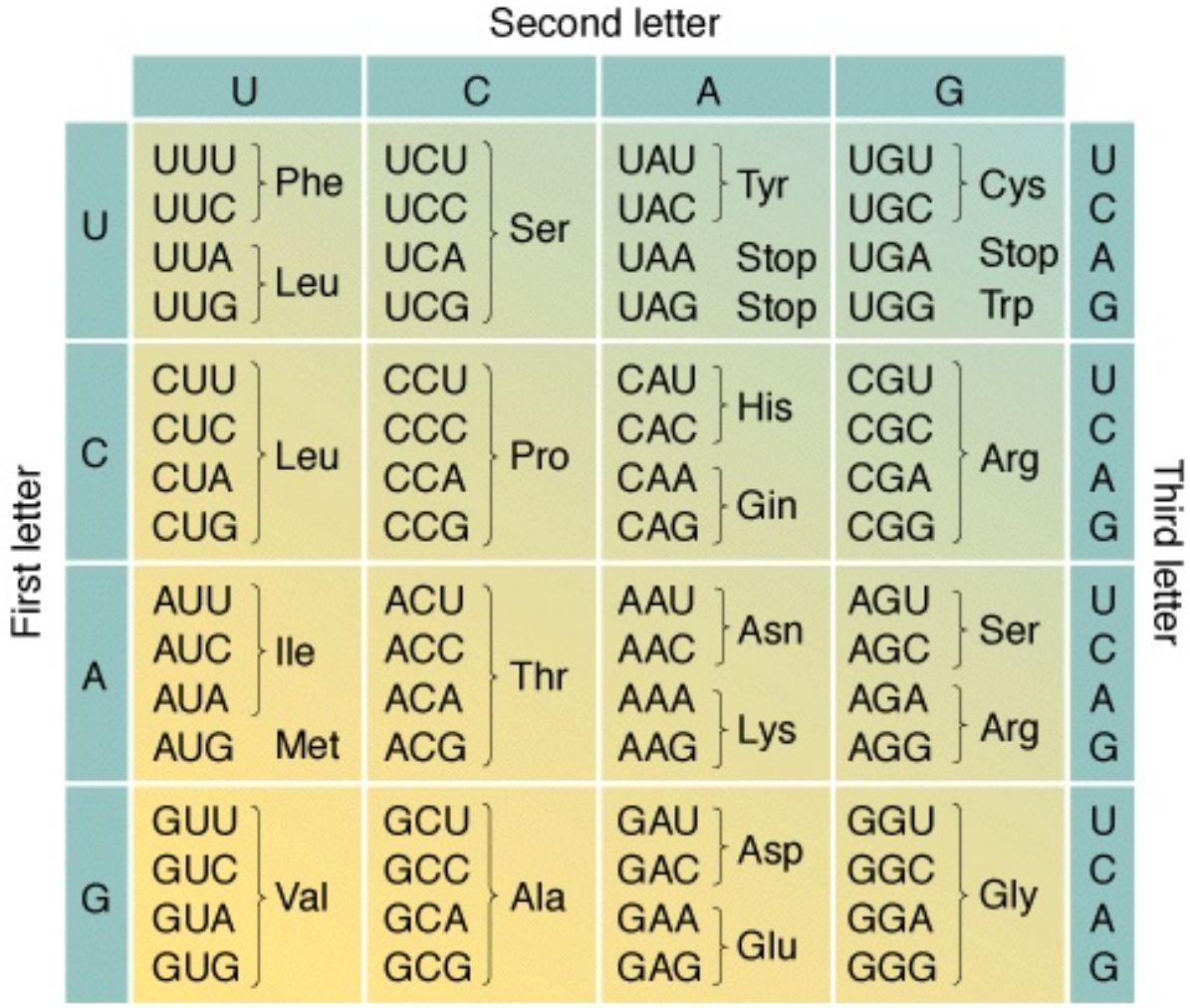

## Downstream effects

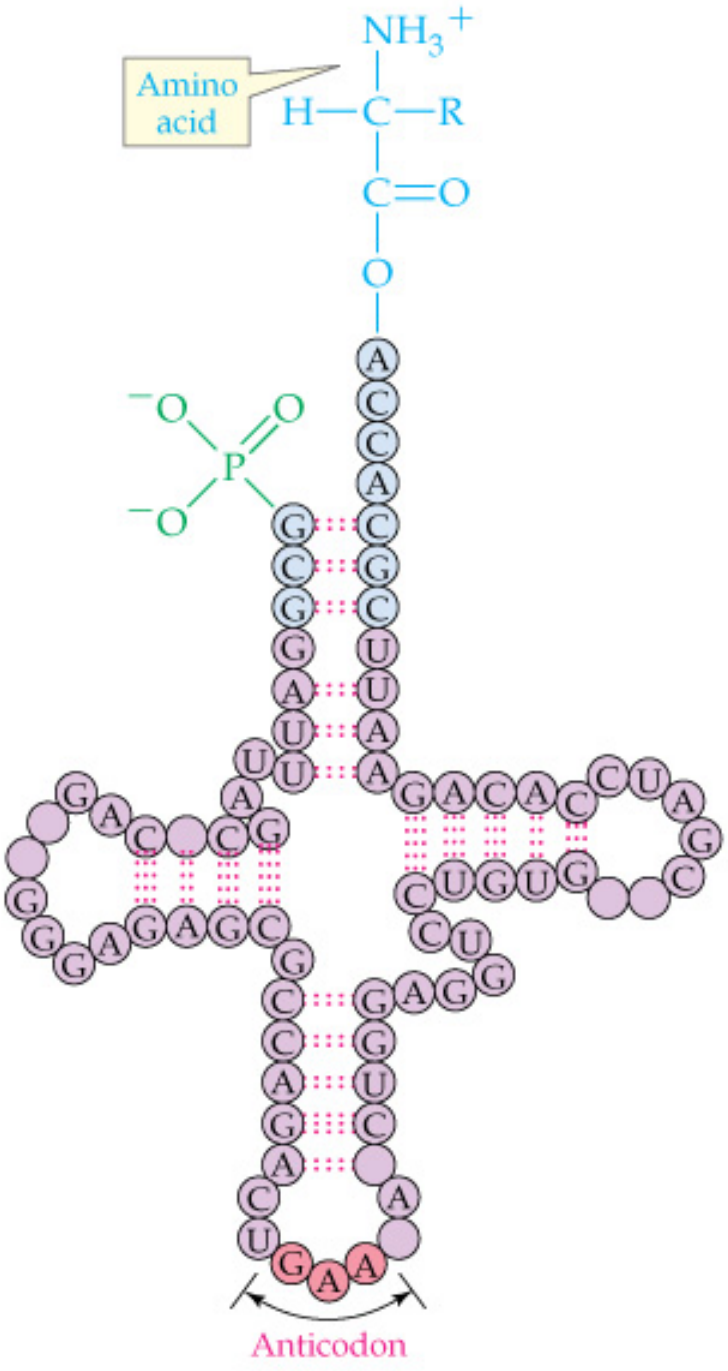
# DNA Methylation and Histone Acetylation



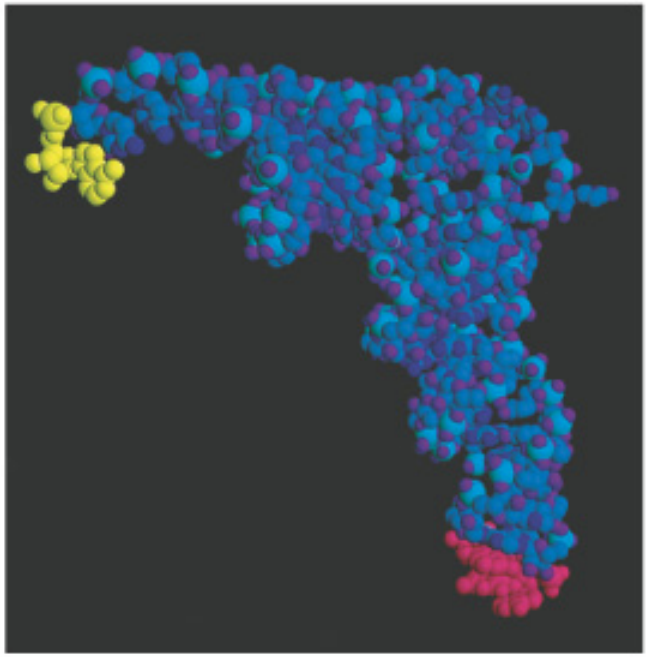

EPIGENETIC MECHANISMS
are affected by these factors and processes:
• Development (in utero, childhood )
• Environmental chemicals
• Drugs/Pharmaceuticals
• Aging
• Diet
HEALTH ENDPOINTS
• Cancer
• Autoimmune disease
• Mental disorders
• Diabetes
CHROMOSOME
CHROMATIN
METHYL GROUP
EPIGENETIC FACTOR
DNA
DNA methylation
Methyl group (an epigenetic factor found in some dietary sources) can tag DNA and activate or repress genes.
GENE
HISTONE TAIL
HISTONE
DNA inaccessible, gene inactive
HISTONE TAIL
DNA accessible, gene active
Histones are proteins around which DNA can wind for compaction and gene regulation.
Histone modification
The binding of epigenetic factors to histone "tails" alters the extent to which DNA is wrapped around histones and the availability of genes in the DNA to be activated.

# DNA Sequence

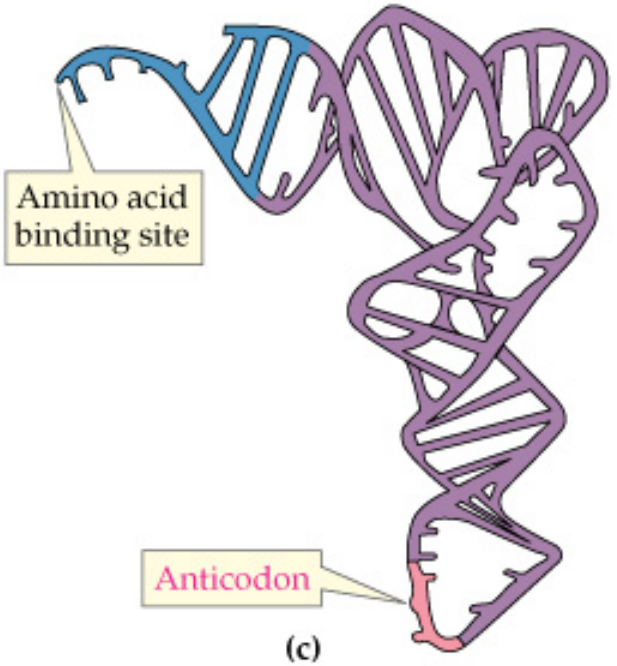



Amino acid
$NH_3^+$
H—C—R
C=O
O
$^-O$
O
P
$^-O$
Anticodon

(a)

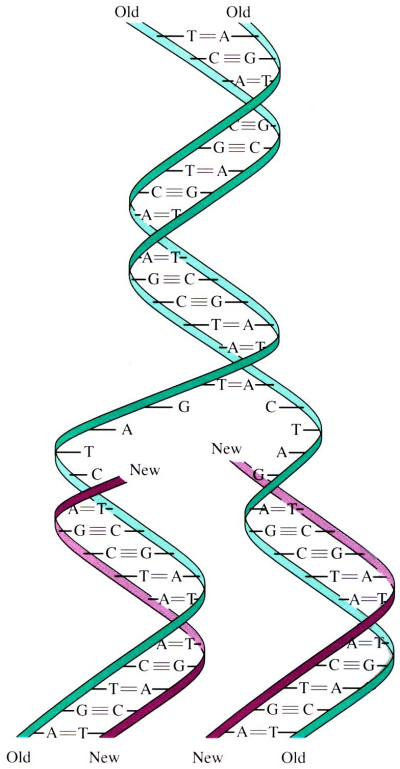


(b)

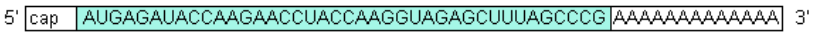

Amino acid binding site
Anticodon

(c)

# Self-Assembly Process in Nature

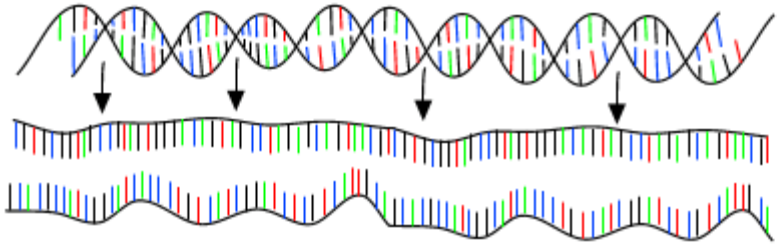






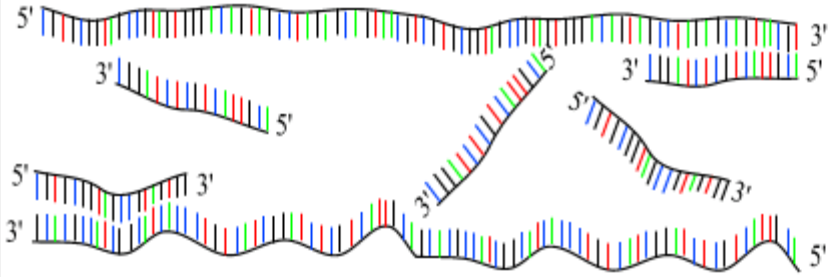
# PCR : Polymerase Chain Reaction

30 - 40 cycles of 3 steps :



**Step 1 : denaturation**

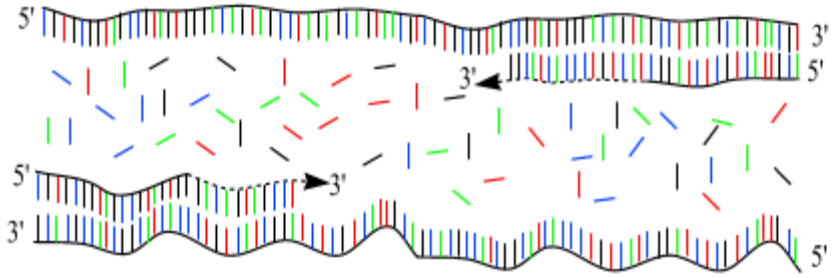
1 minut 94 °C




**Step 2 : annealing**

45 seconds 54 °C

forward and reverse primers !!!




**Step 3 : extension**

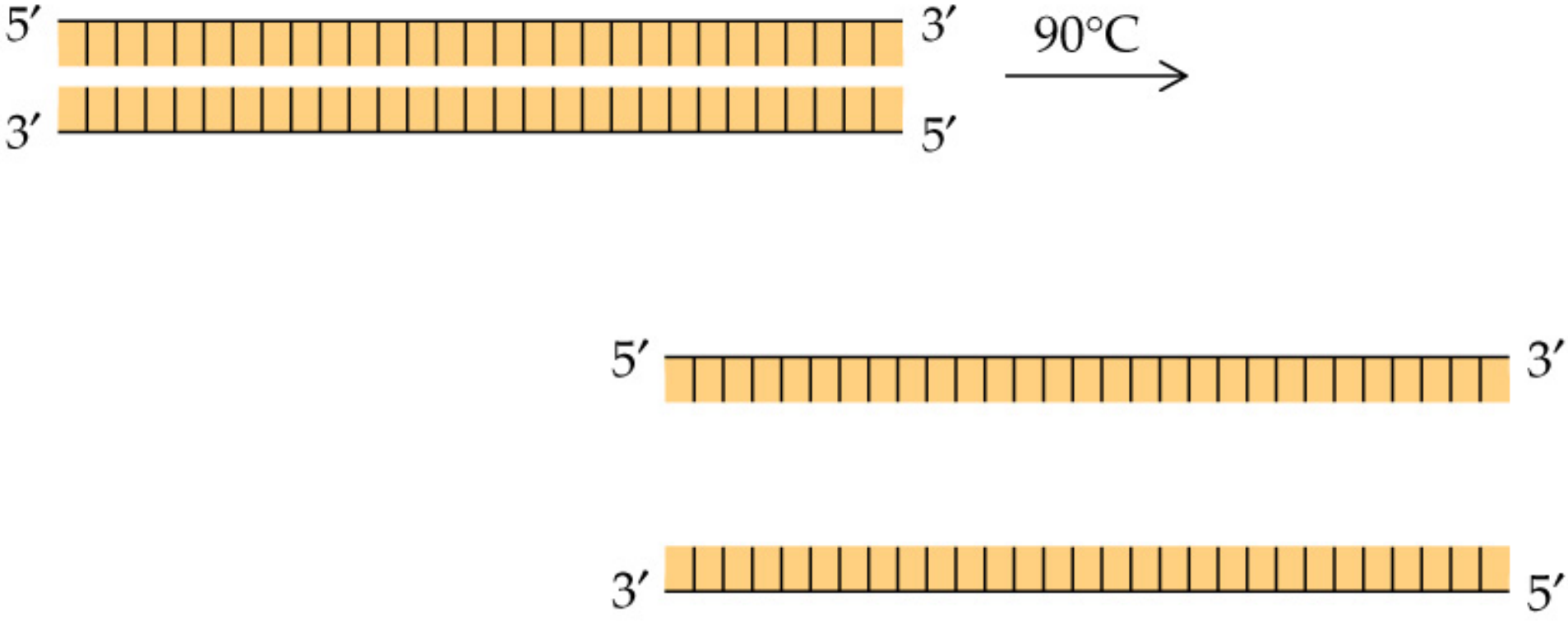
2 minutes 72 °C

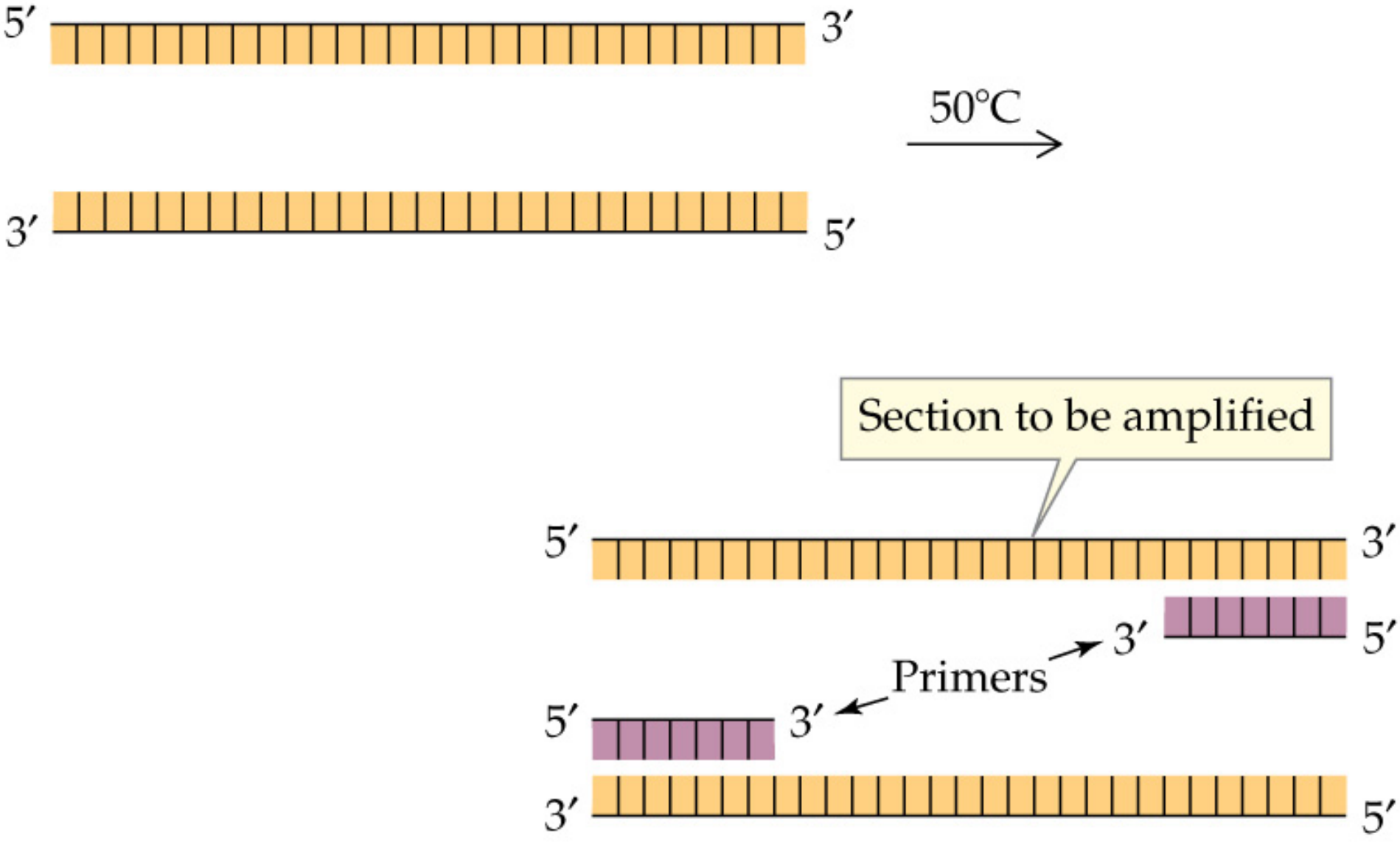
only dNTP's

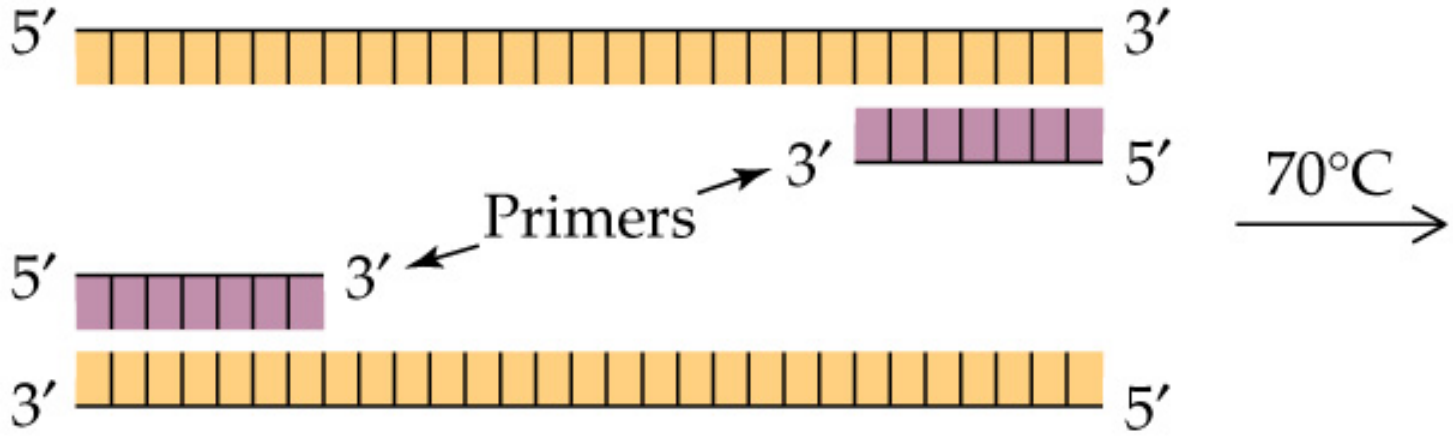
(Andy Vierstraete 1999)

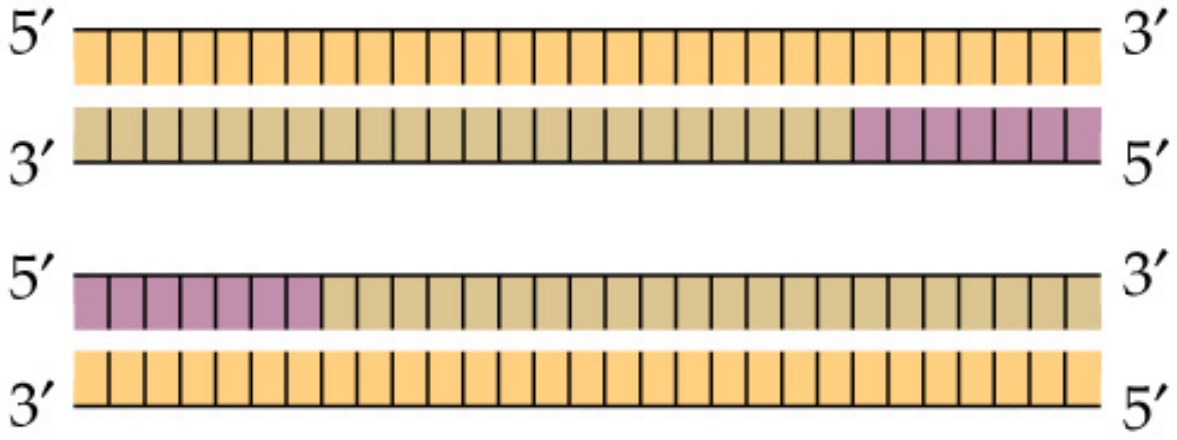
# PCR

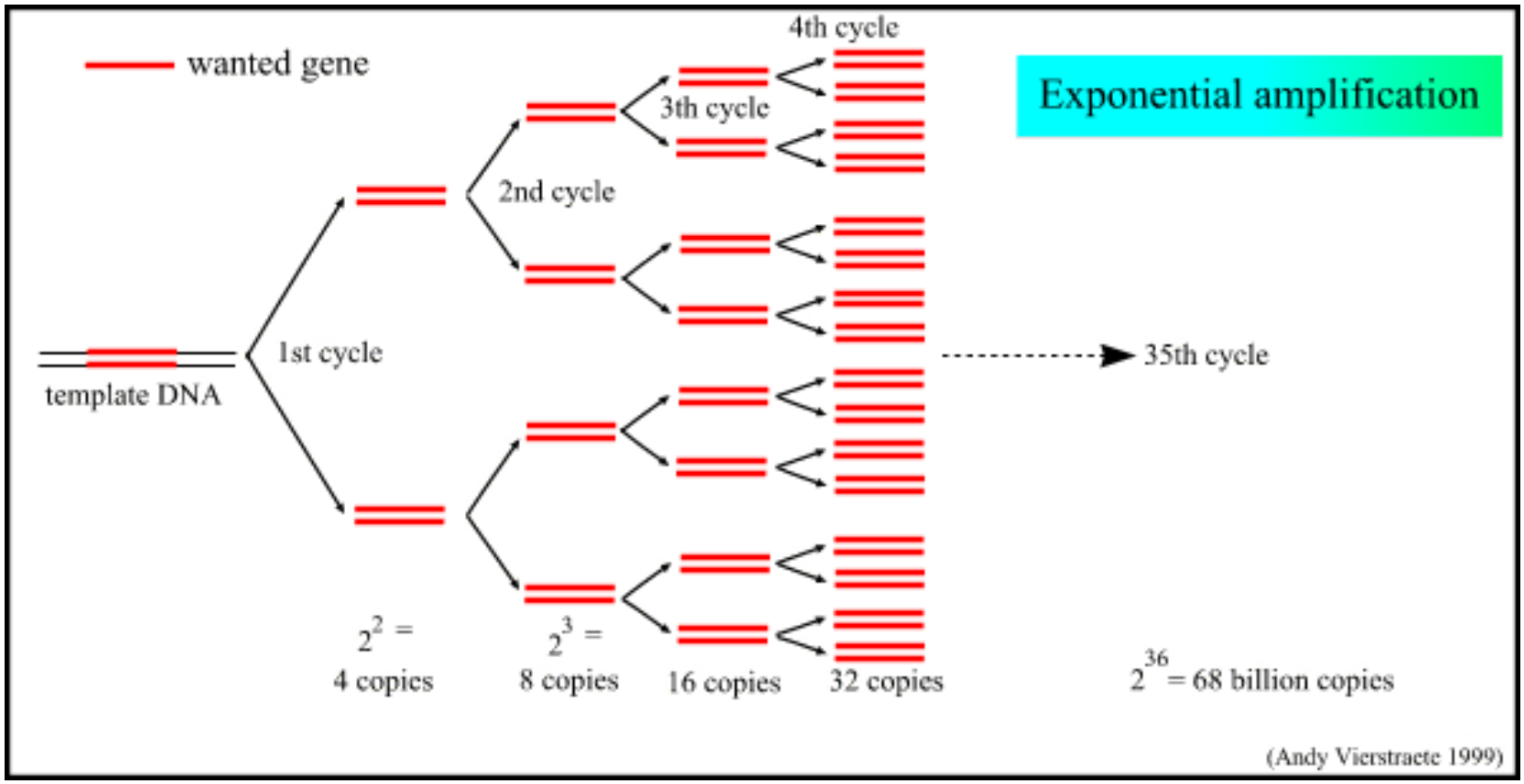
https://www.youtube.com/watch?v=MyLrs_h1OIE


5′
3′
90°C
3′
5′
5′
3′
3′
5′


5′
3′
50°C
3′
5′
Section to be amplified
5′
3′
3′
5′
Primers
5′
3′
3′
5′

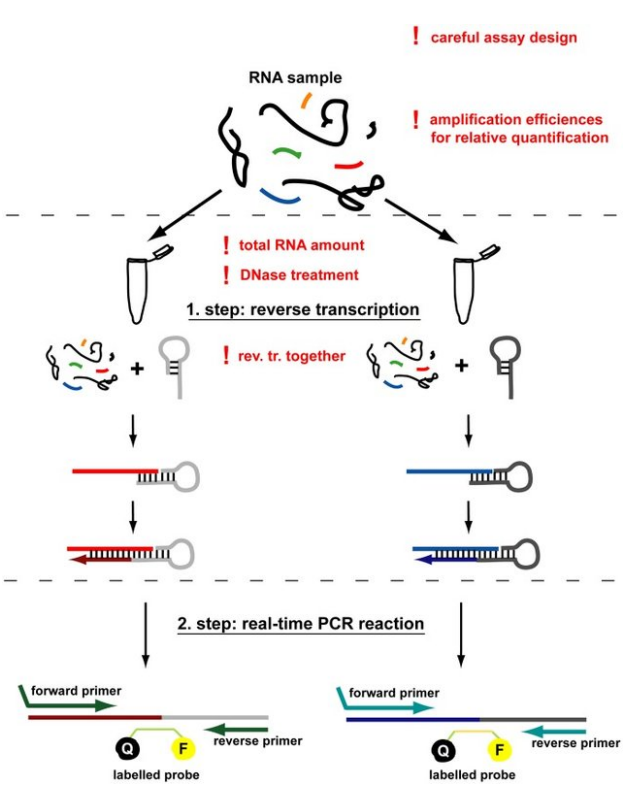

5′
3′
3′
5′
Primers
5′
3′
3′
5′
70°C

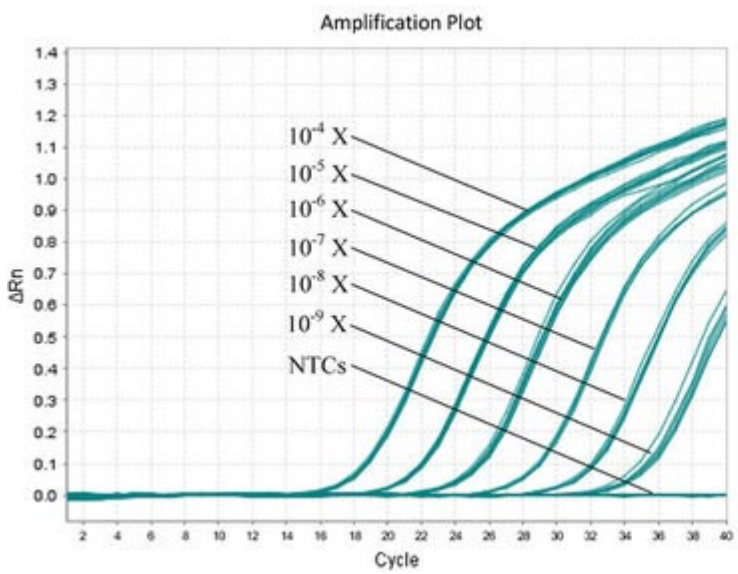

5′
3′
3′
5′
5′
3′
3′
5′

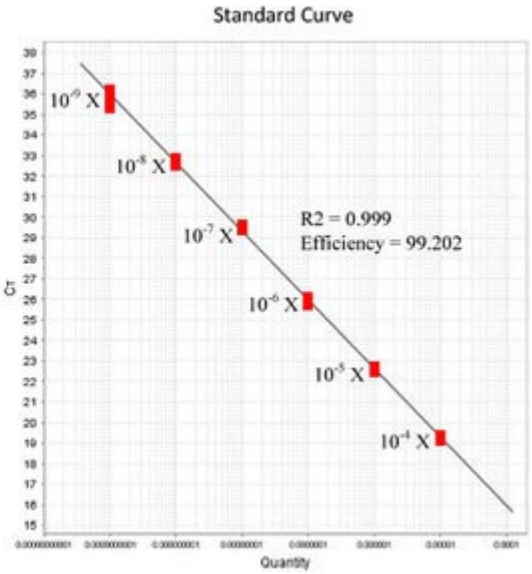

wanted gene
Exponential amplification
4th cycle
3th cycle
2nd cycle
1st cycle
template DNA
35th cycle
$2^2$ =
4 copies
$2^3$ =
8 copies
16 copies
32 copies
$2^{36}$ = 68 billion copies
(Andy Vierstraete 1999)

# Real-time PCR

https://www.youtube.com/watch?v=1kvy17ugl4w


! careful assay design
RNA sample
! amplification efficiences for relative quantification
! total RNA amount
! DNase treatment
1. step: reverse transcription
! rev. tr. together
2. step: real-time PCR reaction
forward primer
reverse primer
Q
F
labelled probe
forward primer
reverse primer
Q
F
labelled probe


Amplification Plot
ΔRn
Cycle
10-4 X
10-5 X
10-6 X
10-7 X
10-8 X
10-9 X
NTCs


Standard Curve
Ct
Quantity
10-9 X
10-8 X
10-7 X
10-6 X
10-5 X
10-4 X
R2 = 0.999
Efficiency = 99.202

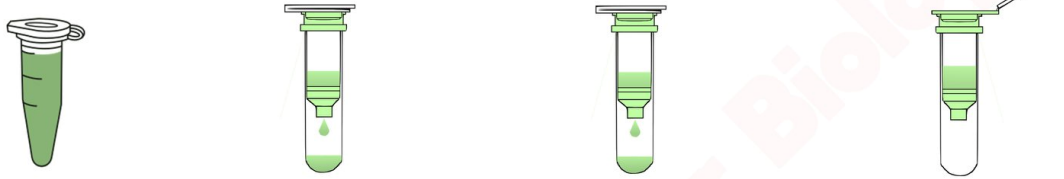
# PROTOCOL OF SARS-COV-2 DETECTION USING REAL-TIME RT-PCR

**Target gene →** RdRp gene (Corman *et al.* 2020)

**PCR amplification regions →** nCoV_IP2/12621-12727 and nCoV_IP4/14010-14116 (Institut Pasteur, Paris)

**Primer sets and probes** → designed based on the first sequences of SARS-CoV-2 available on the GISAID database

## RNA extraction → NucleoSpin® RNA Virus or viral RNA mini kit (QIAGEN)



**Sample lysis**

5 min incubation of sample in Lysis Buffer containing Proteinase K

**Binding of viral RNA**

Ethanol addition and transfer of lysate to Column

**Washing**

1st Wash Buffer (high salt concentration)
2nd Wash Buffer (low salt concentration)

**Elution of viral RNA**

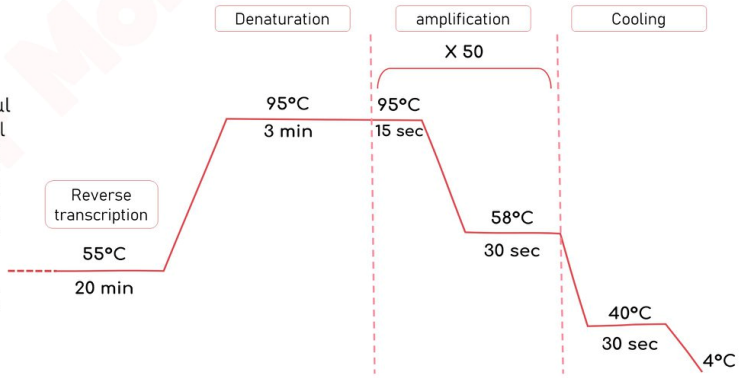
Elution in 20-50 µl RNase-free water or Elution Buffer

## Real-time Multiplex RT-PCR (Institut Pasteur, Paris)

### Multiplex Mix (nCoV_IP2&IP4)

| | |
|---|---|
| Sample RNA | 5 µl |
| H20 | 1.3 µl |
| Reaction mix 2X | 12.50 µl |
| MgSO4 (50mM) | 0.40 µl |
| Forward Primer1 (10µM) | 1.00 µl |
| Reverse Primer1 (10µM) | 1.00 µl |
| Forward Primer2 (10µM) | 1.00 µl |
| Reverse Primer2 (10µM) | 1.00 µl |
| Probe 1 (10µM) | 0.4 µl |
| Probe 2 (10µM) | 0.4 µl |
| SuperscriptIII RT/Platinum Taq Mix | 1.00 µl |

### Amplification Cycles (Lightcycler System)



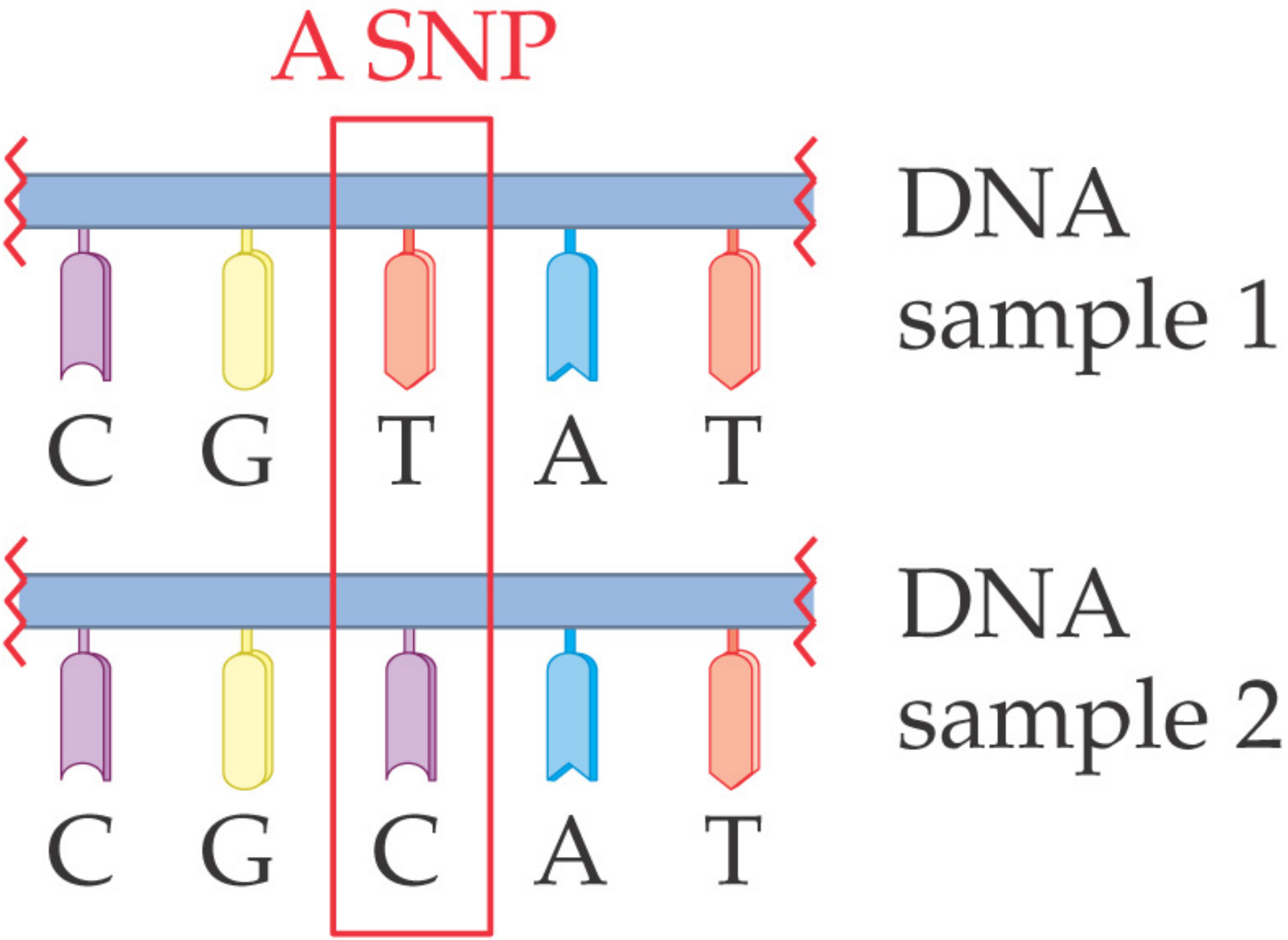

### POSITIVE CONTROL

Positive control for real-time RT-PCR is the in vitro transcribed RNA derived from strain BetaCoV_Wuhan_WIV04_2019. The transcript contains the **amplification regions** of the RdRp and E gene as positive strand.

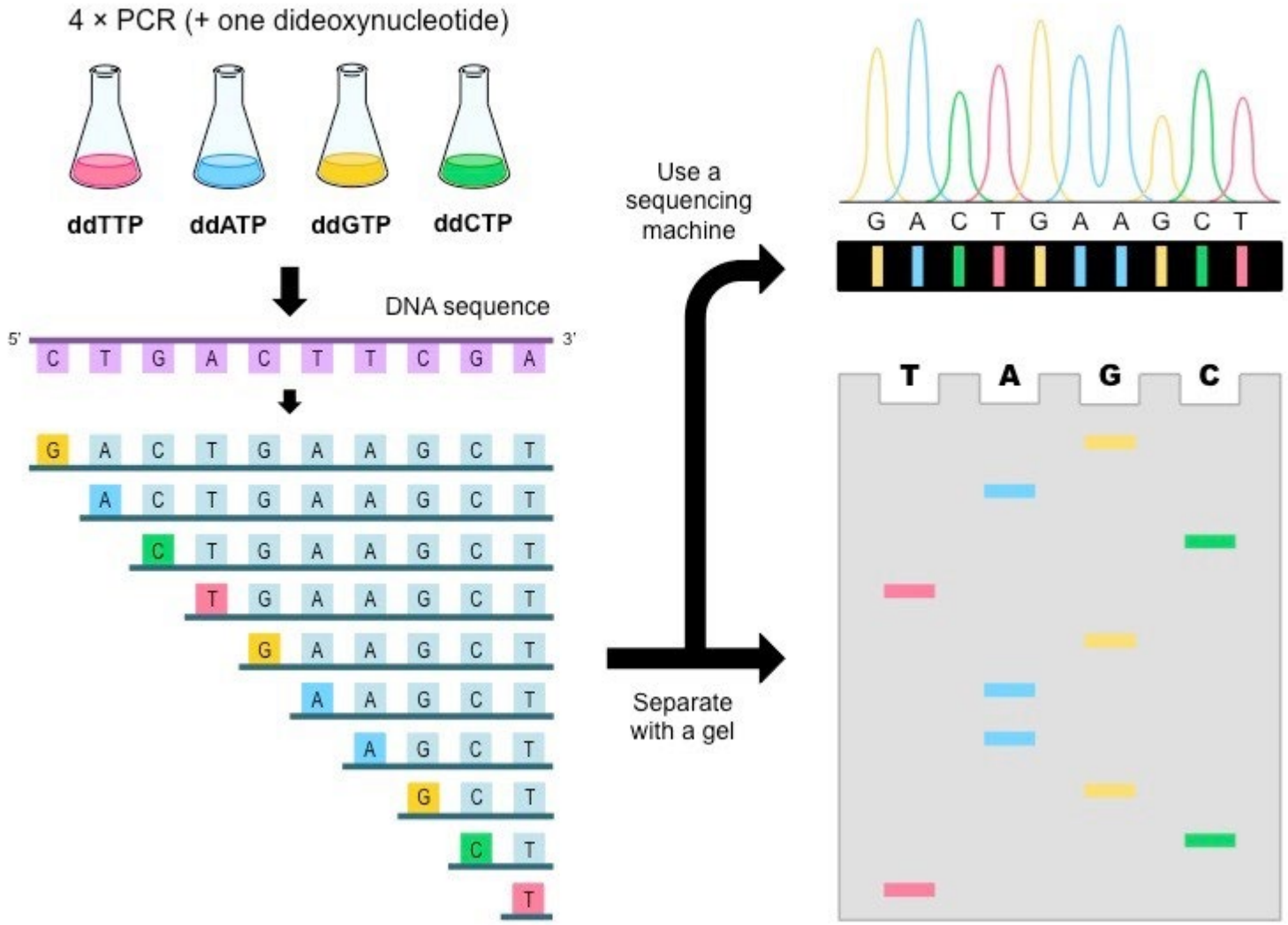
M. MERZOUG

## References


1. Institut Pasteur, Paris, « Protocol: Real-time RT-PCR assays for the detection of SARS-CoV-2 », OMS, 2 mars 2020.
2. Corman VM, Landt O, Kaiser M, et al. Detection of 2019 novel coronavirus (2019-nCoV) by real-time RT-PCR. Euro Surveill 2020;25.

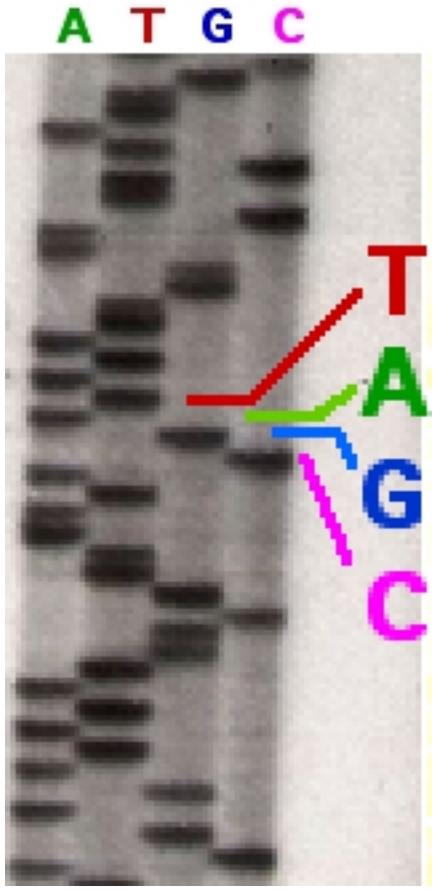

A SNP
C G T A T
DNA sample 1
C G C A T
DNA sample 2

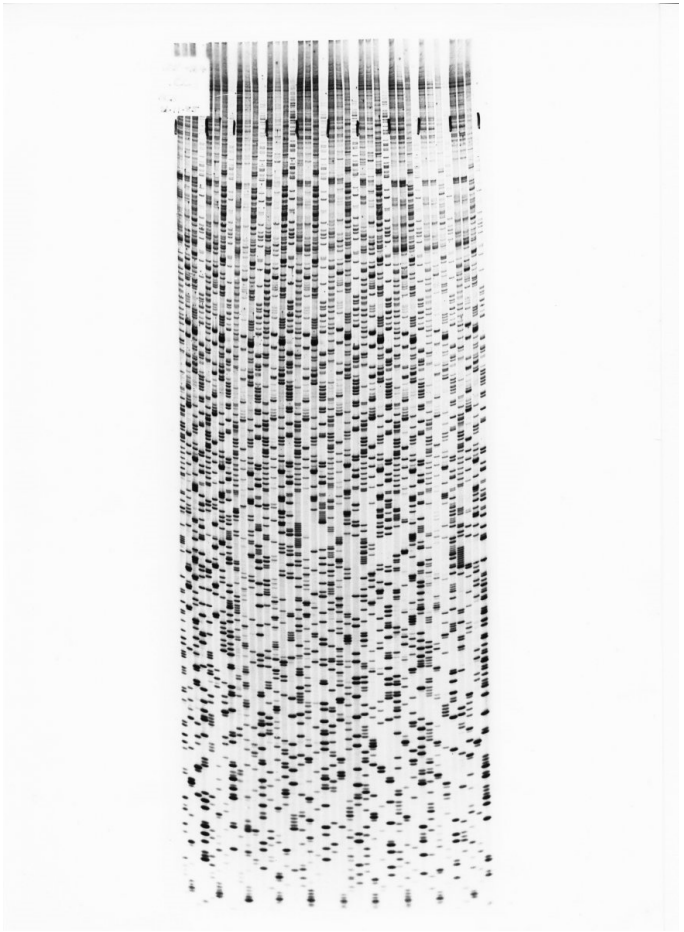
# DNA Sequencing



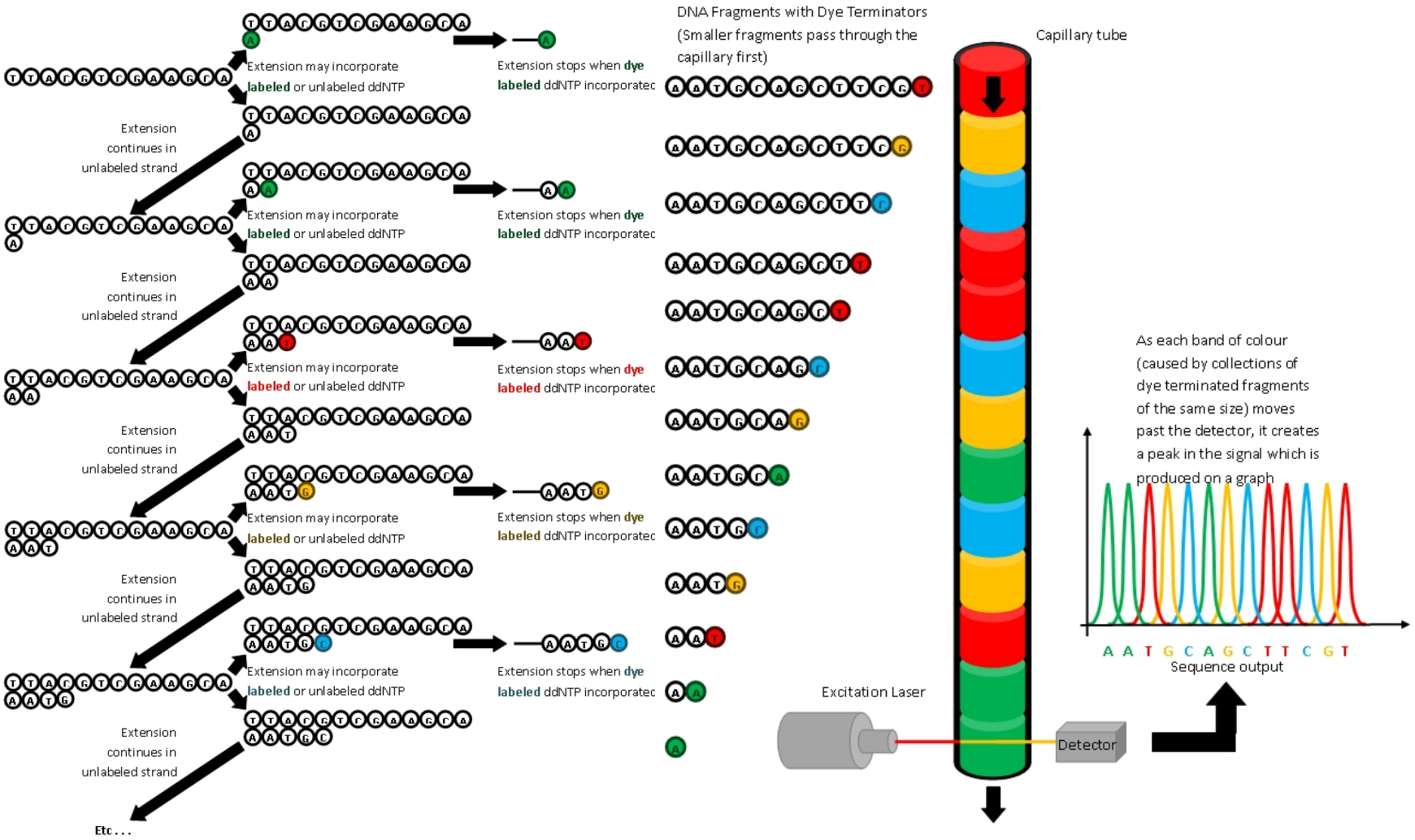
# DNA Sequencing

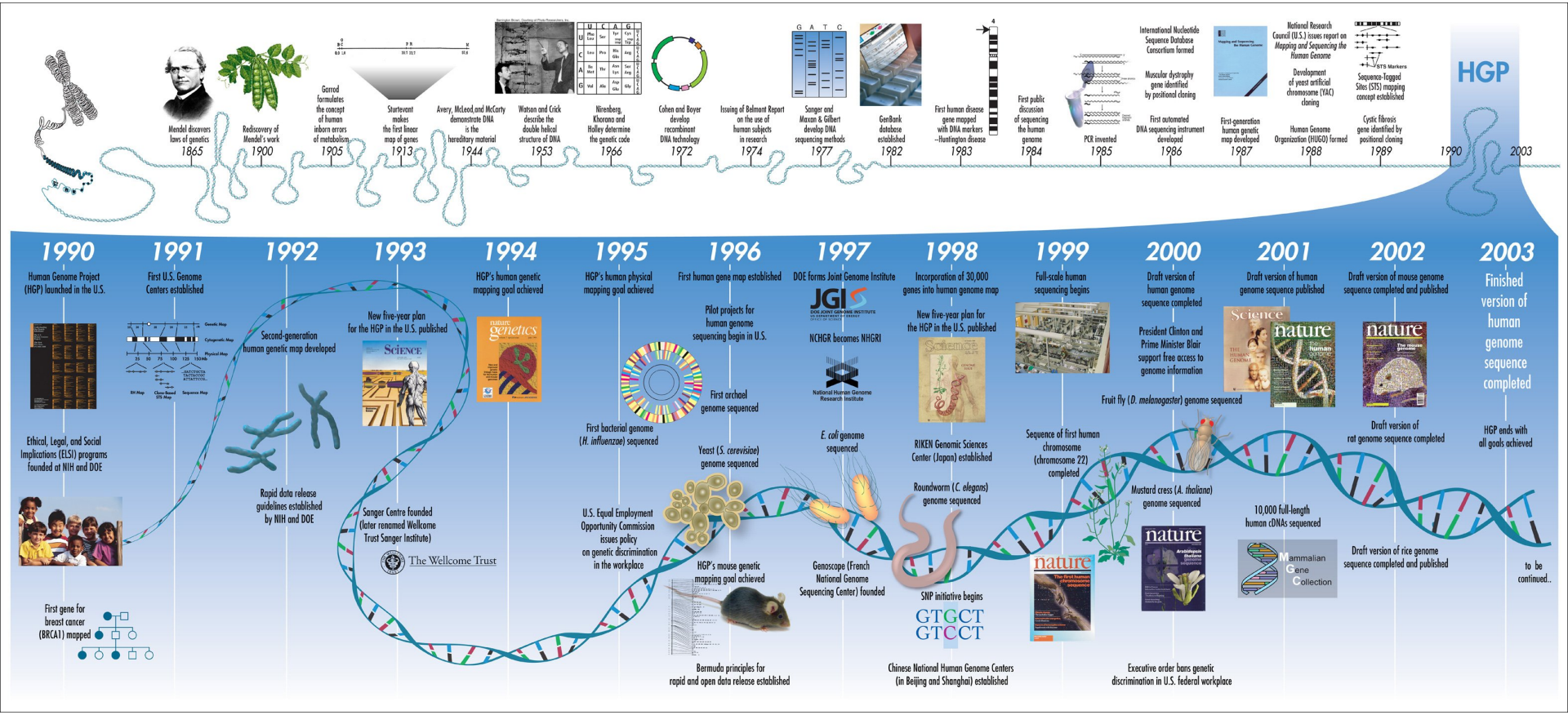
https://www.youtube.com/watch?v=vK-HlMaitnE



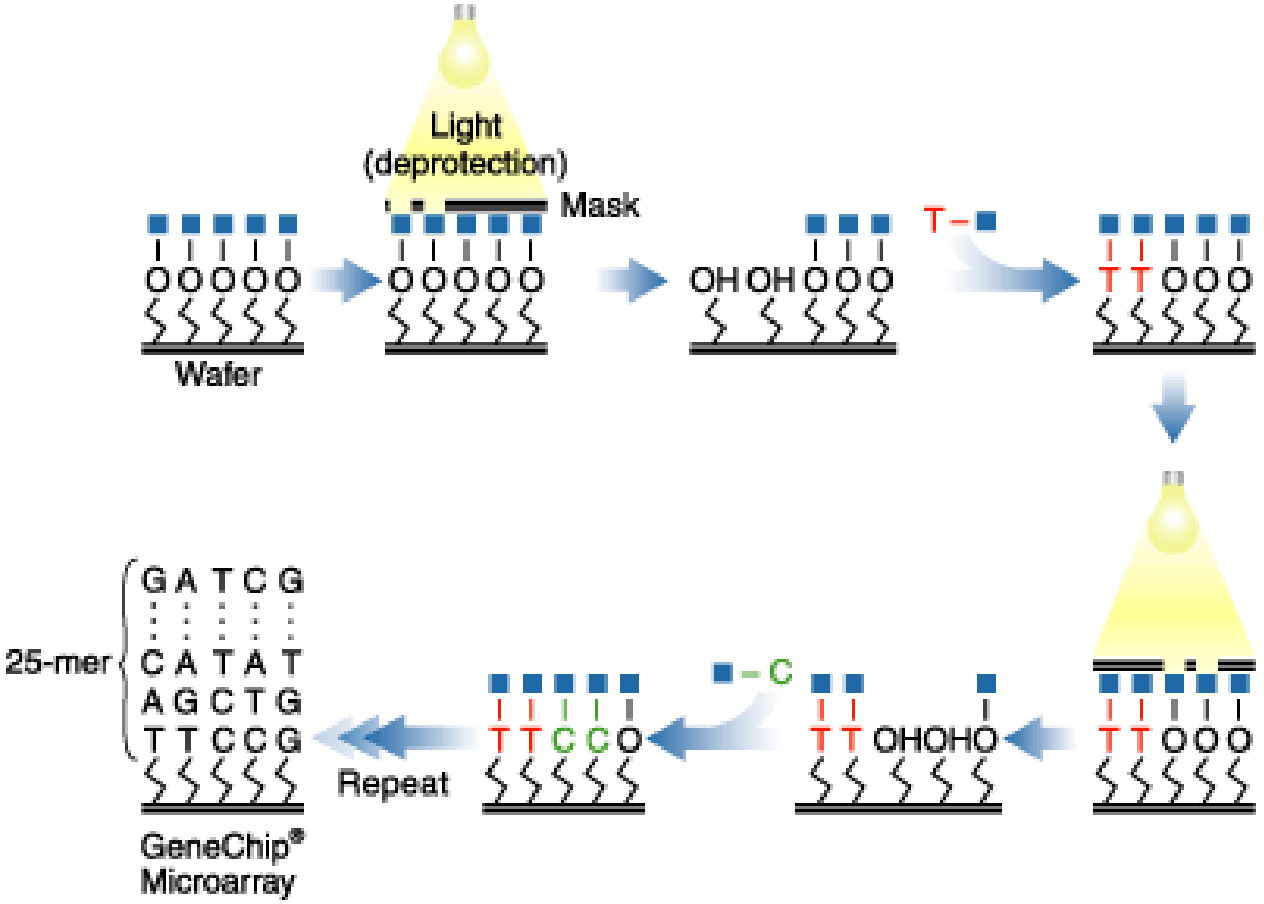



# Dye Terminations

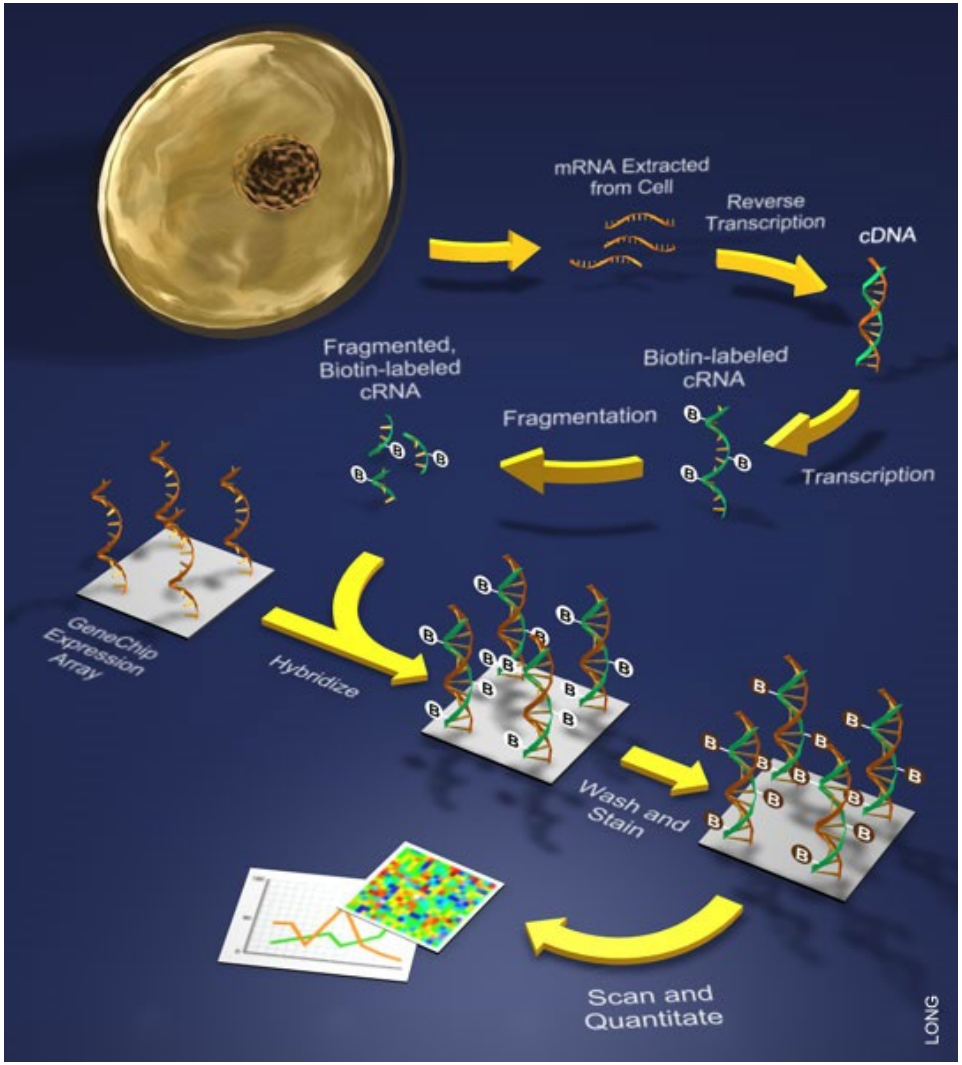



HGP
1865 Mendel discovers laws of genetics
1900 Rediscovery of Mendel's work
1905 Garrod formulates the concept of human inborn errors of metabolism
1913 Sturtevant makes the first linear map of genes
1944 Avery, McLeod, and McCarty demonstrate DNA is the hereditary material
1953 Watson and Crick describe the double helical structure of DNA
1966 Nirenberg, Khorana and Holley determine the genetic code
1972 Cohen and Boyer develop recombinant DNA technology
1974 Issuing of Belmont Report on the use of human subjects in research
1977 Sanger and Maxam & Gilbert develop DNA sequencing methods
1982 GenBank database established
1983 First human disease gene mapped with DNA markers –Huntington disease
1984 First public discussion of sequencing the human genome
1985 PCR invented
1986 International Nucleotide Sequence Database Consortium formed
Muscular dystrophy gene identified by positional cloning
First automated DNA sequencing instrument developed
1987 First-generation human genetic map developed
1988 National Research Council (U.S.) issues report on Mapping and Sequencing the Human Genome
Development of yeast artificial chromosome (YAC) cloning
Human Genome Organization (HUGO) formed
1989 STS Markers
Sequence-Tagged Sites (STS) mapping concept established
Cystic fibrosis gene identified by positional cloning
1990
2003
1990
Human Genome Project (HGP) launched in the U.S.
Ethical, Legal, and Social Implications (ELSI) programs founded at NIH and DOE
First gene for breast cancer (BRCA1) mapped
1991
First U.S. Genome Centers established
1992
Second-generation human genetic map developed
Rapid data release guidelines established by NIH and DOE
1993
New five-year plan for the HGP in the U.S. published
Sanger Centre founded (later renamed Wellcome Trust Sanger Institute)
The Wellcome Trust
1994
HGP's human genetic mapping goal achieved
1995
HGP's human physical mapping goal achieved
First bacterial genome (H. influenzae) sequenced
U.S. Equal Employment Opportunity Commission issues policy on genetic discrimination in the workplace
1996
First human gene map established
Pilot projects for human genome sequencing begin in U.S.
First archael genome sequenced
Yeast (S. cerevisiae) genome sequenced
HGP's mouse genetic mapping goal achieved
Bermuda principles for rapid and open data release established
1997
DOE forms Joint Genome Institute
JGI
NCHGR becomes NHGRI
National Human Genome Research Institute
E. coli genome sequenced
Genoscope (French National Genome Sequencing Center) founded
1998
Incorporation of 30,000 genes into human genome map
New five-year plan for the HGP in the U.S. published
RIKEN Genomic Sciences Center (Japan) established
Roundworm (C. elegans) genome sequenced
SNP initiative begins
GTGCT
GTCCT
Chinese National Human Genome Centers (in Beijing and Shanghai) established
1999
Full-scale human sequencing begins
Sequence of first human chromosome (chromosome 22) completed
2000
Draft version of human genome sequence completed
President Clinton and Prime Minister Blair support free access to genome information
Fruit fly (D. melanogaster) genome sequenced
Mustard cress (A. thaliana) genome sequenced
Executive order bans genetic discrimination in U.S. federal workplace
2001
Draft version of human genome sequence published
10,000 full-length human cDNAs sequenced
Mammalian Gene Collection
2002
Draft version of mouse genome sequence completed and published
Draft version of rat genome sequence completed
Draft version of rice genome sequence completed and published
2003
Finished version of human genome sequence completed
HGP ends with all goals achieved
to be continued..

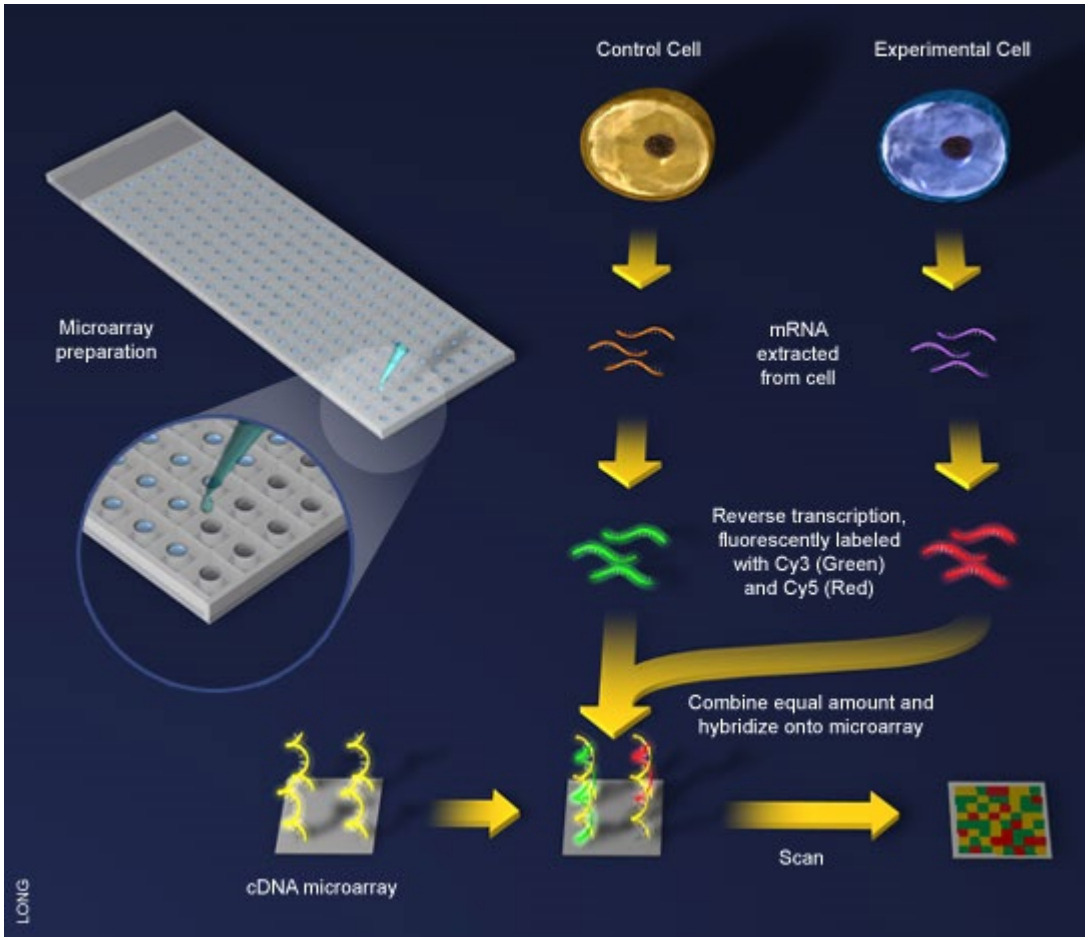
# GeneChip


Light
(deprotection)
Mask
Wafer
T
C
Repeat
25-mer
G A T C G
C A T A T
A G C T G
T T C C G
GeneChip®
Microarray

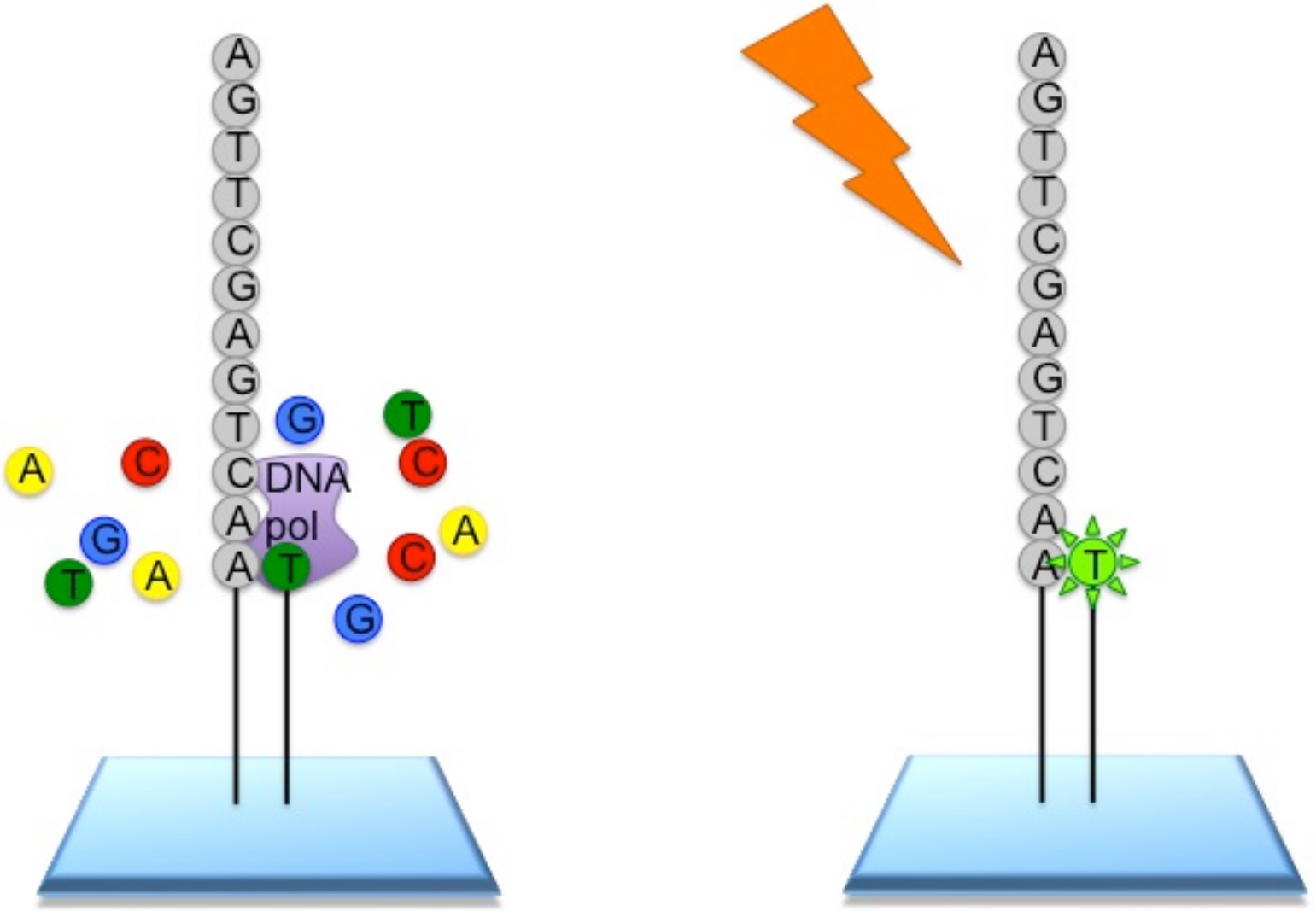
# Scheme


mRNA Extracted from Cell
Reverse Transcription
cDNA
Fragmented, Biotin-labeled cRNA
Biotin-labeled cRNA
Fragmentation
Transcription
GeneChip Expression Array
Hybridize
Wash and Stain
Scan and Quantitate
LONG

# cDNA Microarray

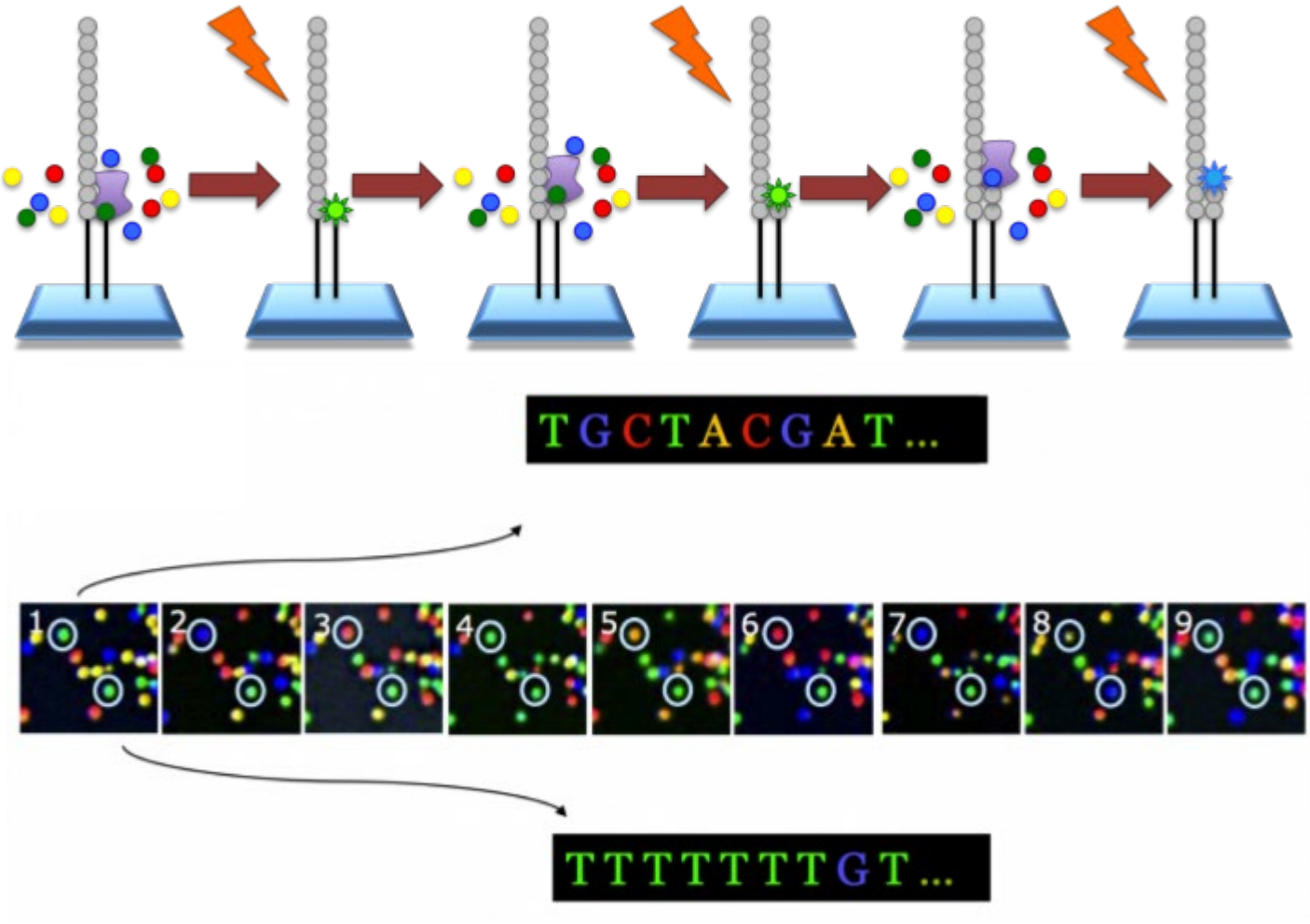


# NGS Illumina

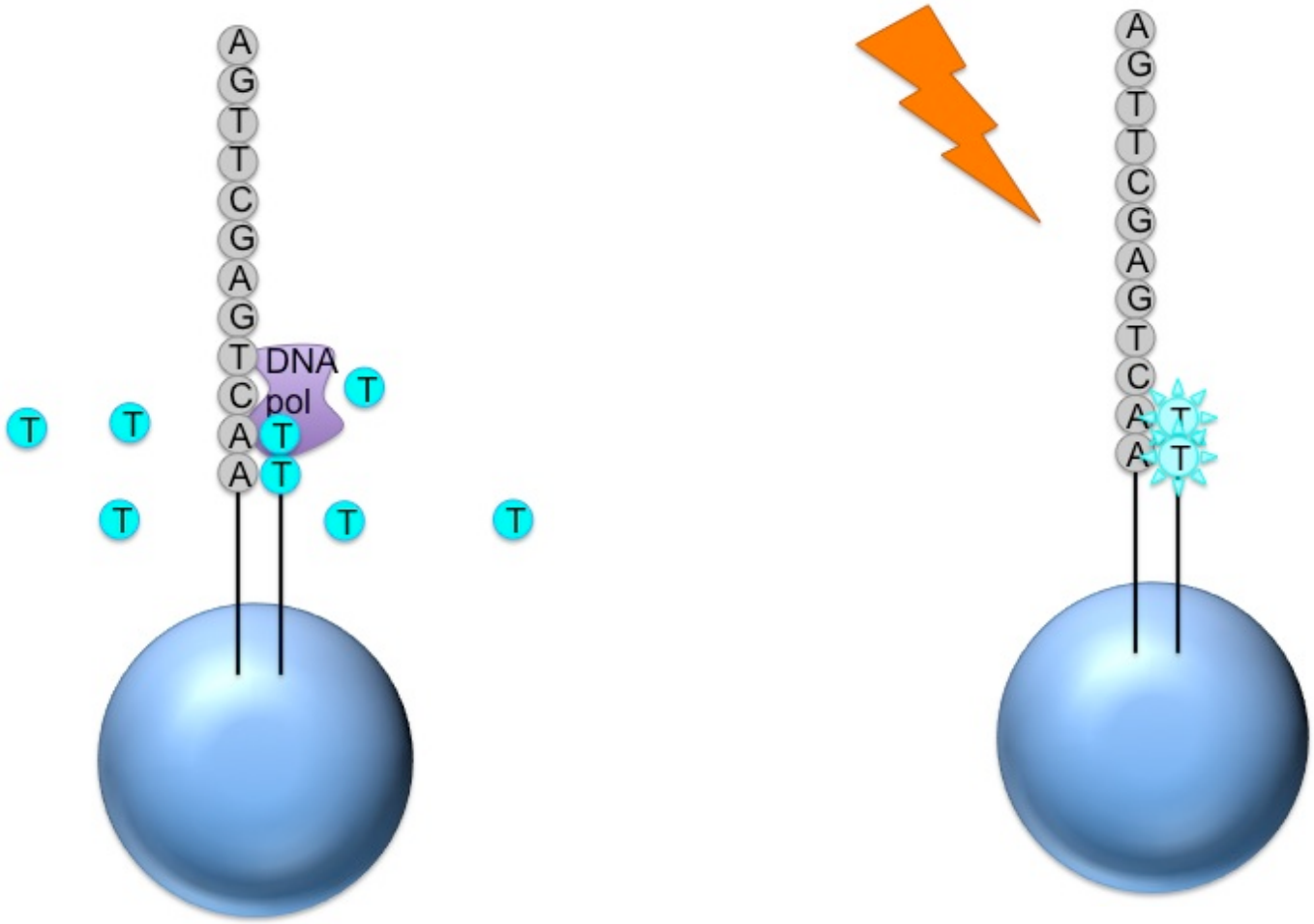


# NGS Illumina

https://www.youtube.com/watch?v=fCd6B5HRaZ8

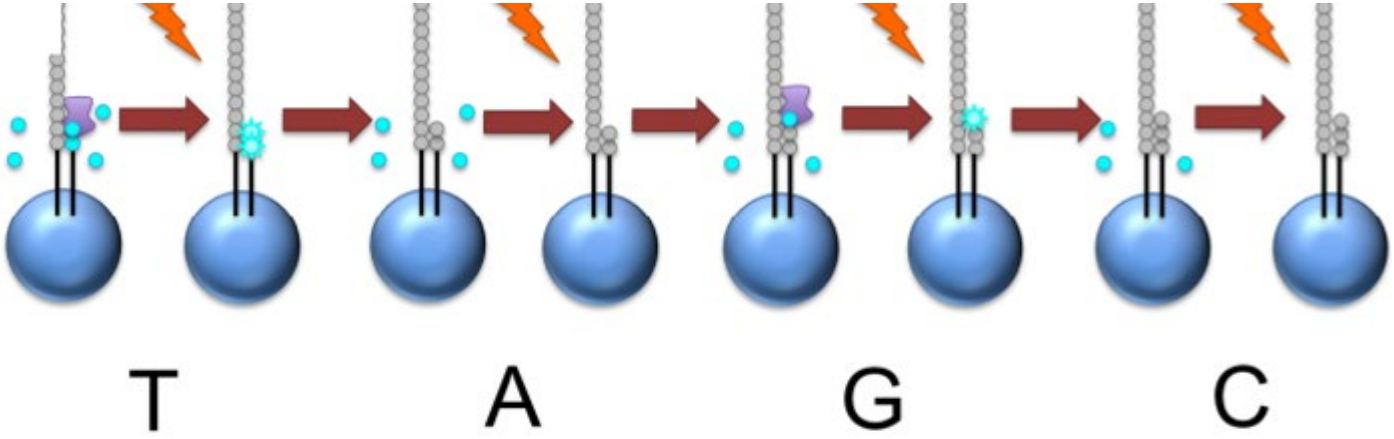


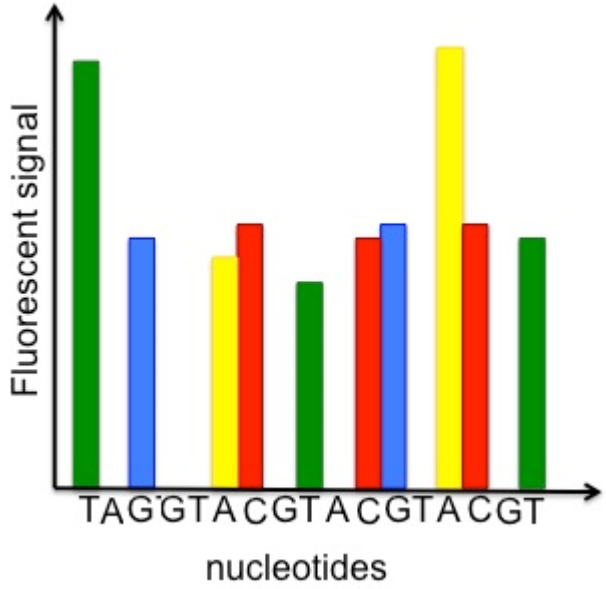
# Roche 454 sequencing




1000 bp

# Roche 454 sequencing



The four main advantages of NGS over classical Sanger sequencing are:

**speed**
**cost**
**sample size**
**accuracy**

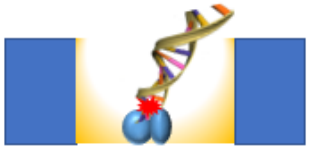
NGS is significantly cheaper, quicker, needs significantly less DNA and is more accurate and reliable than Sanger sequencing.

NGS is quicker than Sanger sequencing in two ways. Firstly, the chemical reaction may be combined with the signal detection in some versions of NGS, whereas in Sanger sequencing these are two separate processes. Secondly and more significantly, only one read (maximum ~1kb) can be taken at a time in Sanger sequencing, whereas NGS is massively parallel, allowing 300Gb of DNA to be read on a single run on a single chip.
The first human genome sequence cost in the region of £300M. Using modern Sanger sequencing methods, aided by data from the known sequence, a full human genome would still cost £6M. Sequencing a human genome with Illumina today would cost only £6,000.

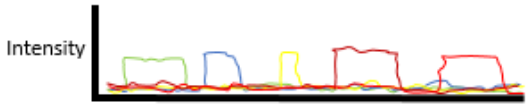
# Third Generation Sequencing

PacBio SMRT seq

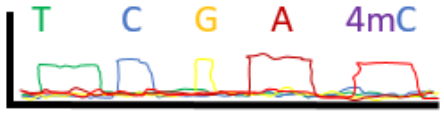
DNA passes thru polymerase in an illuminated volume



Raw output is fluorescent signal of the nucleotide incorporation, specific to each nucleotide

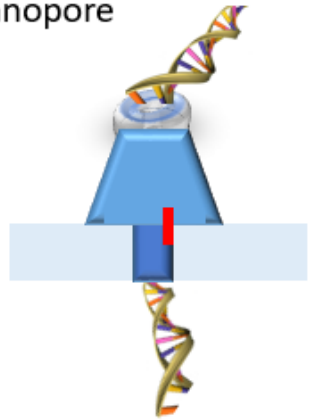



A,C,T,G have known pulse durations, which are used to infer methylated nucleotides

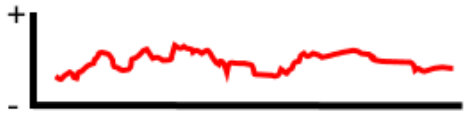



Oxford Nanopore

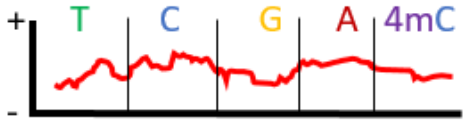
DNA passes thru nanopore



Raw output is electrical signal caused by nucleotide blocking ion flow in nanopore

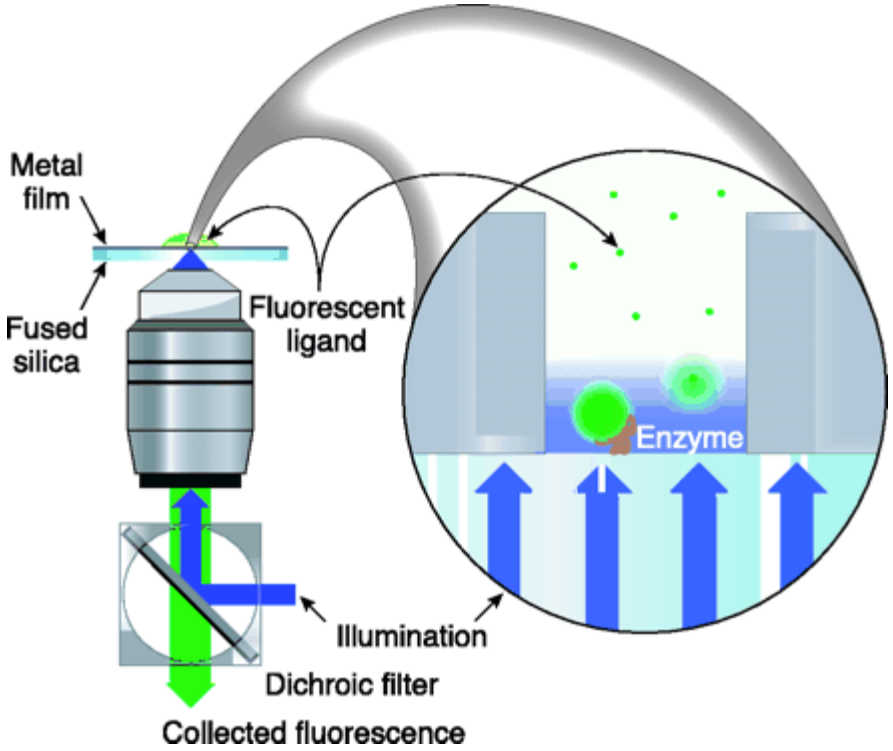



Each nucleotide has a specific electric “signature”



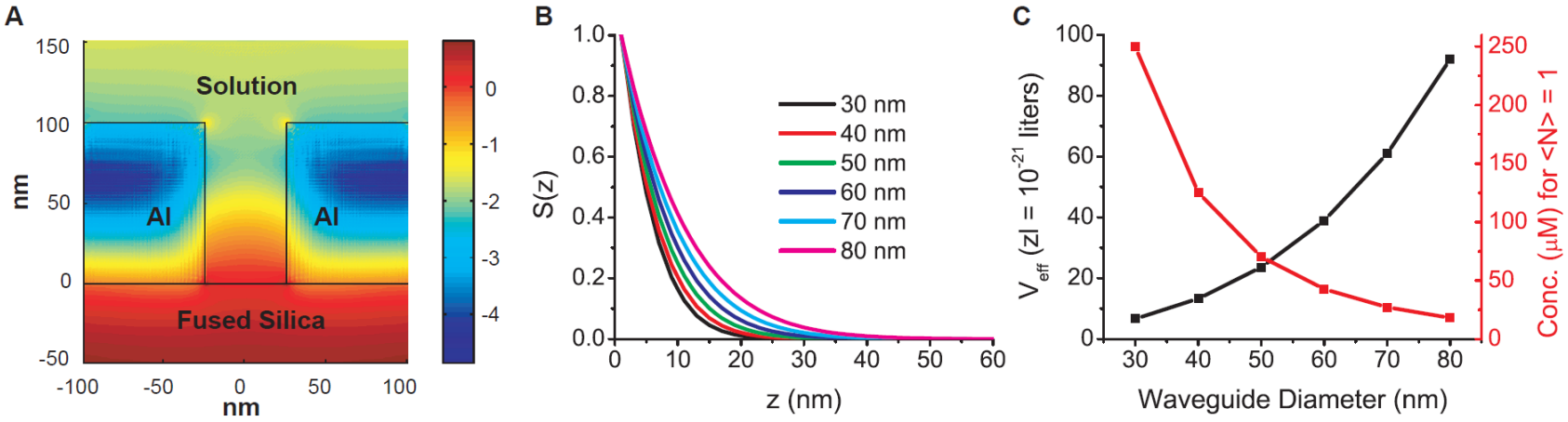
# Zero Mode Waveguide

https://www.youtube.com/watch?v=NHCJ8PtYCFc




SCIENCE VOL 299 31 JANUARY 2003

# Zero Mode Waveguide




50 x 50 x 10 $nm^3$ = 2.5 x $10^4$ x $10^{-21}$ cc = 2.5 x $10^{-20}$ L= 25 zeptoliter

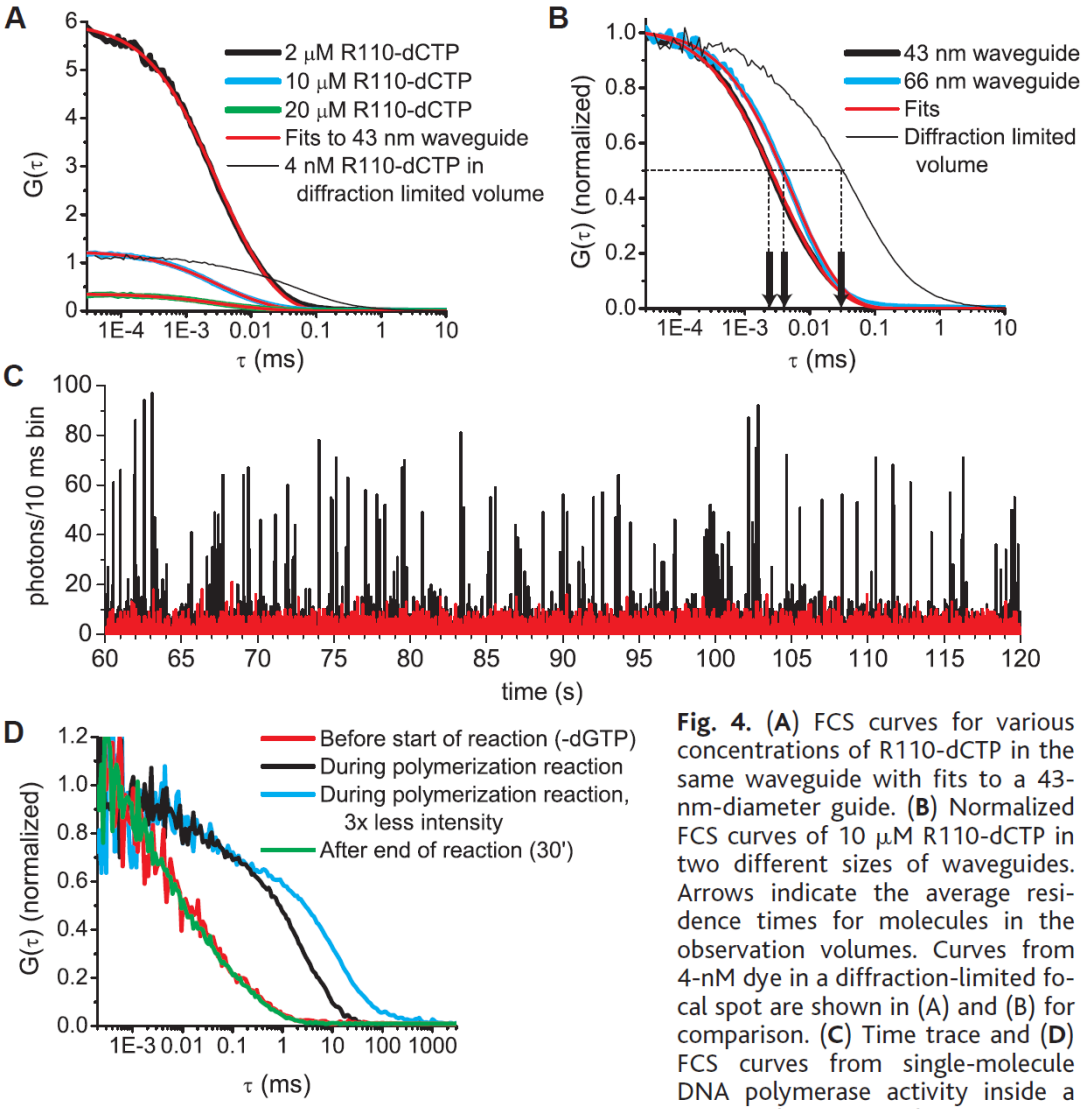
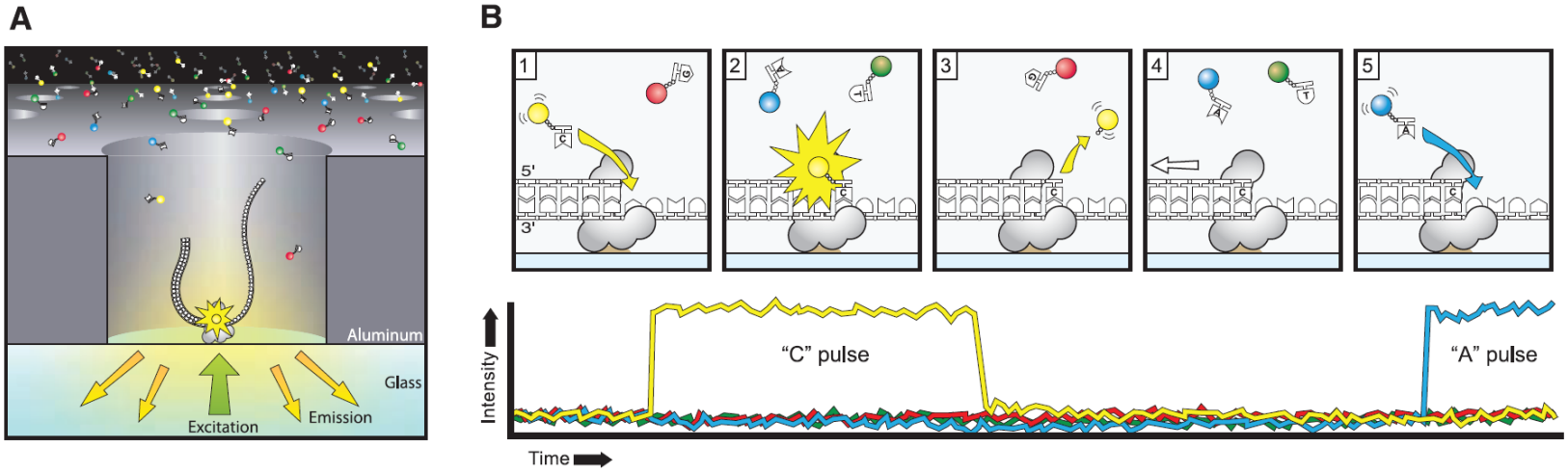



**Fig. 4.** (**A**) FCS curves for various concentrations of R110-dCTP in the same waveguide with fits to a 43-nm-diameter guide. (**B**) Normalized FCS curves of 10 μM R110-dCTP in two different sizes of waveguides. Arrows indicate the average residence times for molecules in the observation volumes. Curves from 4-nM dye in a diffraction-limited focal spot are shown in (A) and (B) for comparison. (**C**) Time trace and (**D**) FCS curves from single-molecule DNA polymerase activity inside a zero-mode waveguide. Incorporation events and subsequent photobleaching of coumarin-dCTP appear as distinct fluorescence bursts in the black time trace (10-ms time bins). This results in a long-time shoulder in the corresponding FCS curves during polymerization (black and blue curves) in (D). Decreasing the intensity results in slower photobleaching as seen by the longer residence time in the blue curve. The red curves in (C) and (D) are the corresponding negative controls (absence of one native nucleotide) in the same waveguide before initiation of the reaction. The green curve in (D) is the control after the reaction has stopped.

# Real-Time DNA Sequencing from Single Polymerase Molecules

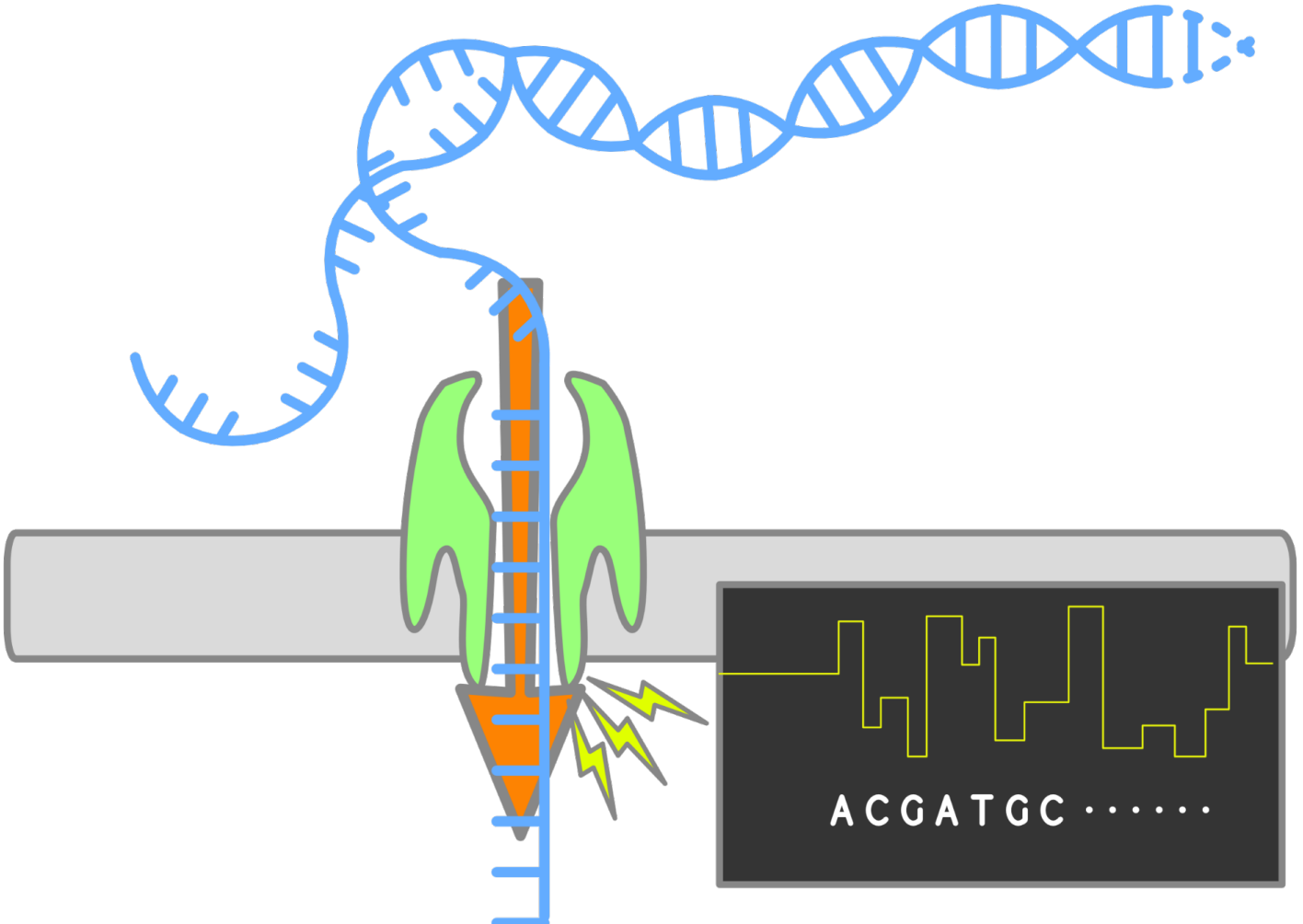


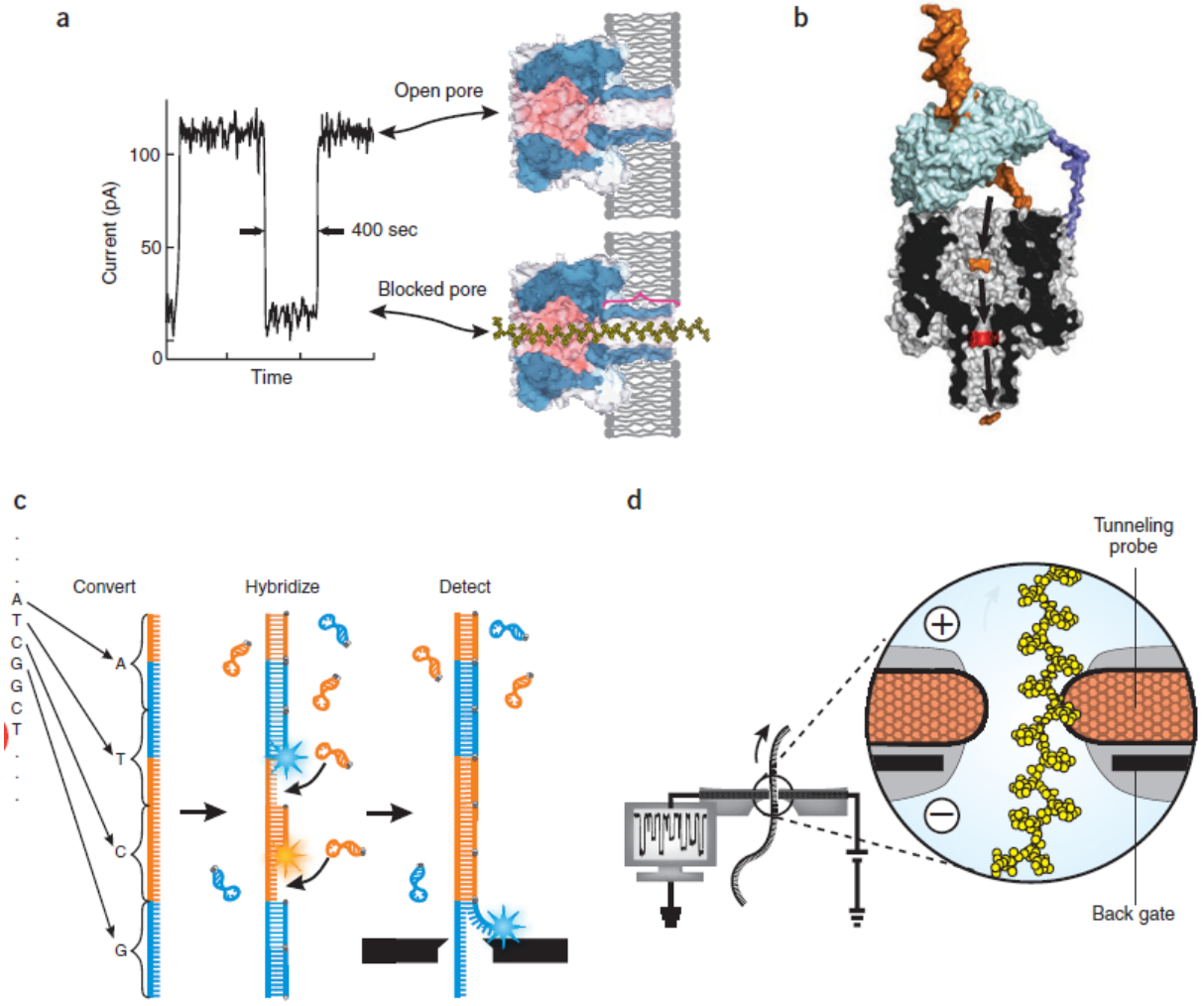

SCIENCE VOL 323 2 JANUARY 2009

# Nanopore Sequencing

https://www.youtube.com/watch?v=RcP85JHLmnI

https://www.youtube.com/watch?v=qzusVw4Dp8w




a
Open pore
100
50
0
Current (pA)
400 sec
Blocked pore
Time
b
c
Convert
Hybridize
Detect
A
T
C
G
G
C
T
A
T
C
G
d
Tunneling probe
Back gate

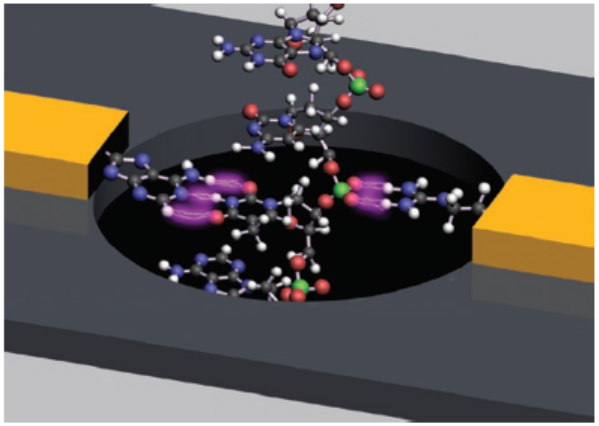
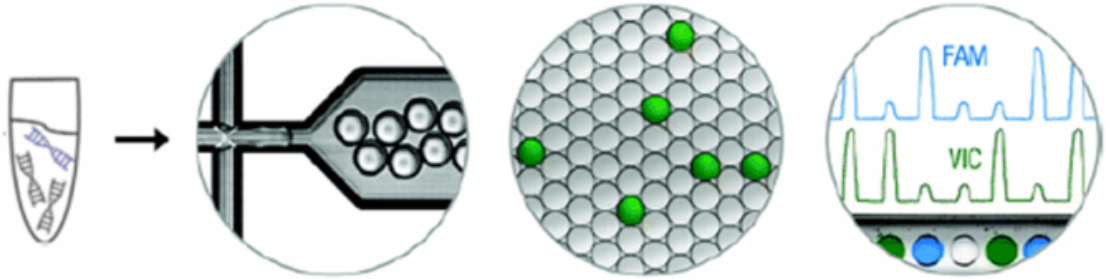


**Figure 2** A nanopore reader with chemically functionalized probes. As a strand of DNA emerges from a nanopore, a 'phosphate grabber' on one functionalized electrode and a 'base reader' on the other electrode form hydrogen bonds (light blue ovals) to complete a transverse electrical circuit through each nucleotide as it is translocated through the nanopore.

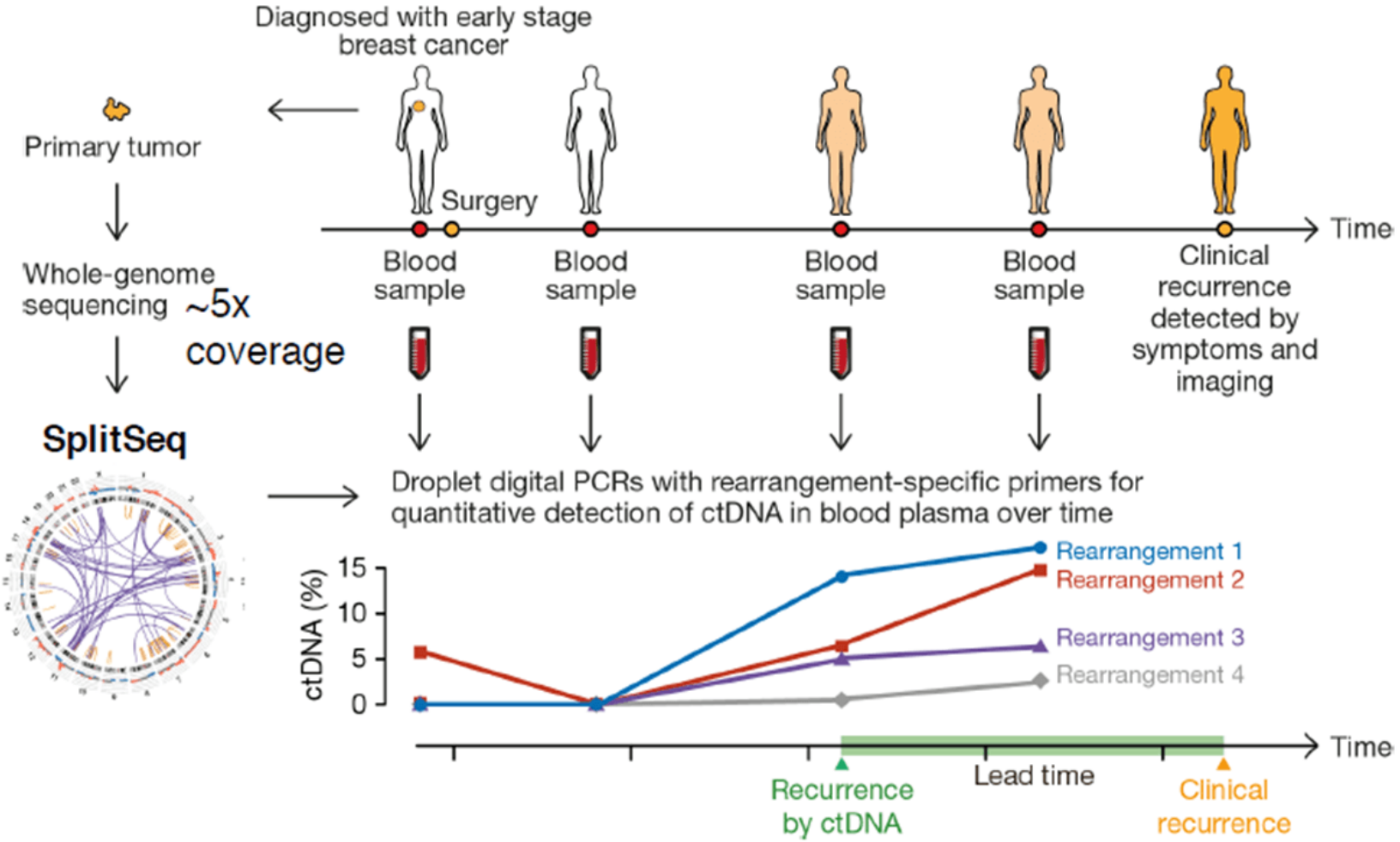
# Digital PCR


## Droplet digital PCR

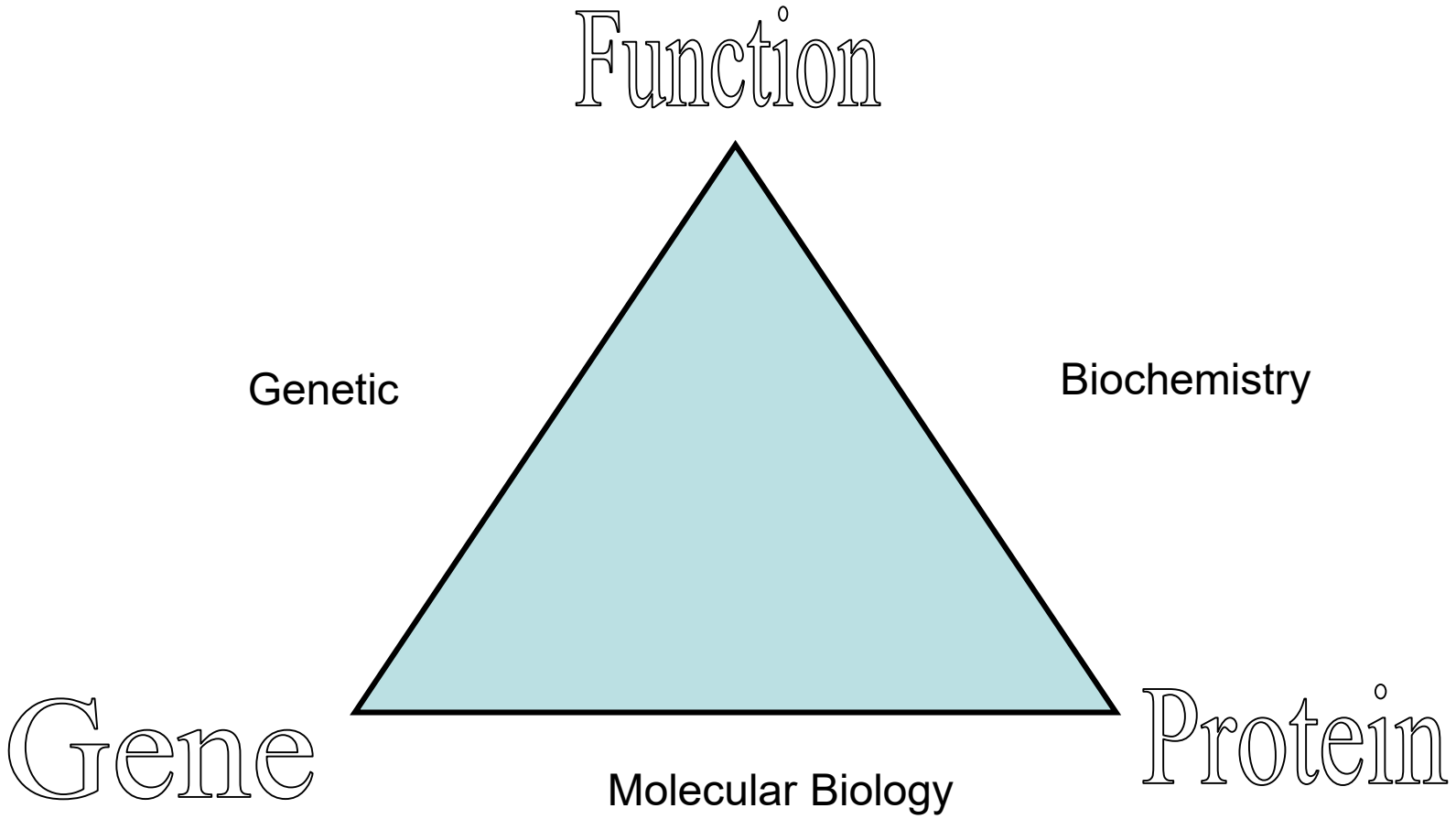


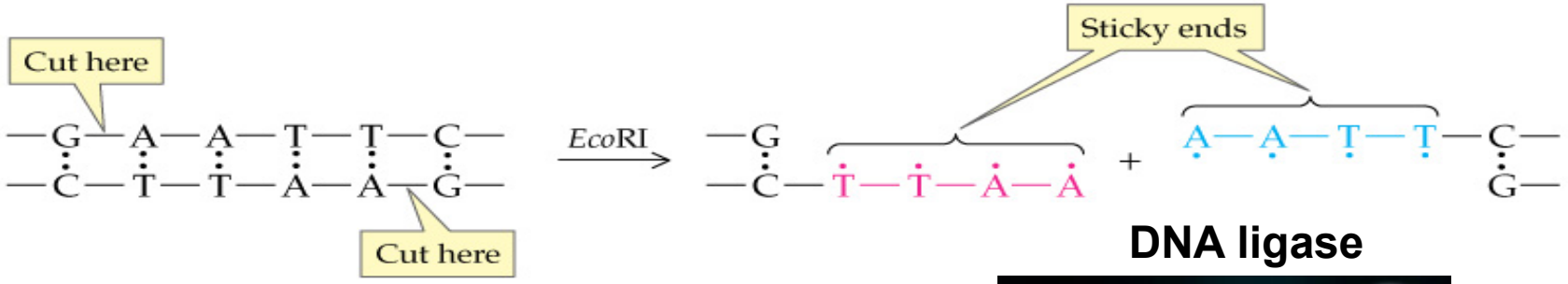

Bio-Rad QX100

| 1. MAKE | 2. CYCLE | 3. READ | |
|---|---|---|---|
| Sample is partitioned into 20,000 droplets | Run PCR cycles in all droplets simultaneously | Measure fluorescence intensity in each droplet | Calculate concentration from number of positive droplets |

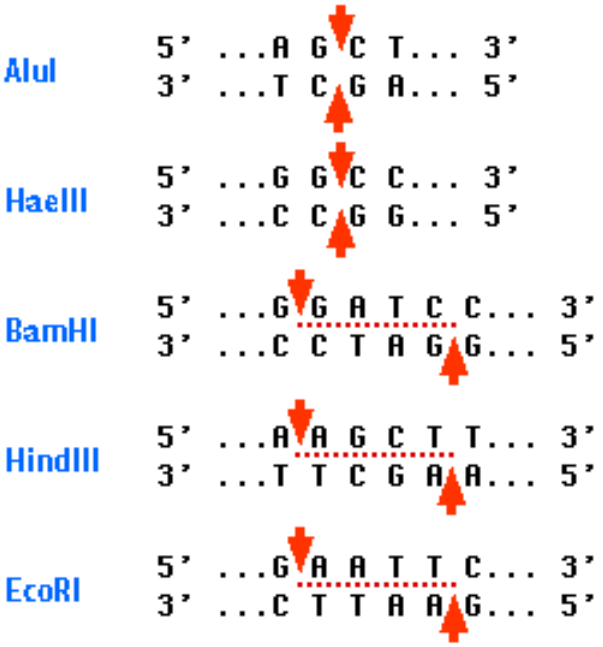

Diagnosed with early stage breast cancer
Primary tumor
Surgery
Time
Whole-genome sequencing
~5x coverage
Blood sample
Blood sample
Blood sample
Blood sample
Clinical recurrence detected by symptoms and imaging
SplitSeq
Droplet digital PCRs with rearrangement-specific primers for quantitative detection of ctDNA in blood plasma over time
ctDNA (%)
15
10
5
0
Rearrangement 1
Rearrangement 2
Rearrangement 3
Rearrangement 4
Time
Lead time
Recurrence by ctDNA
Clinical recurrence

# Recombinant DNA





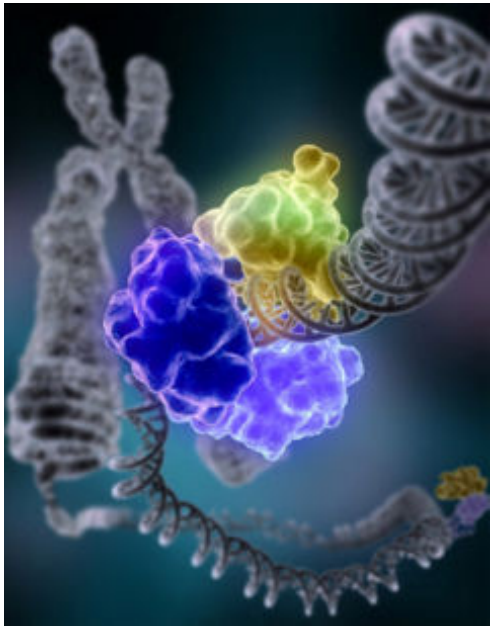

## Restriction Enzyme

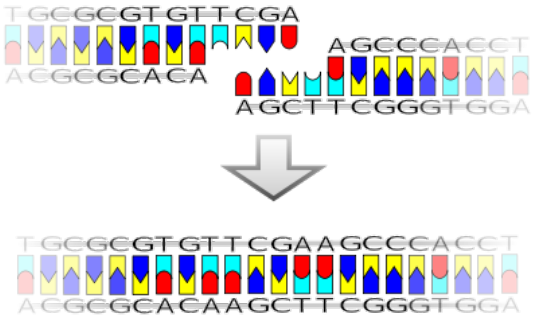


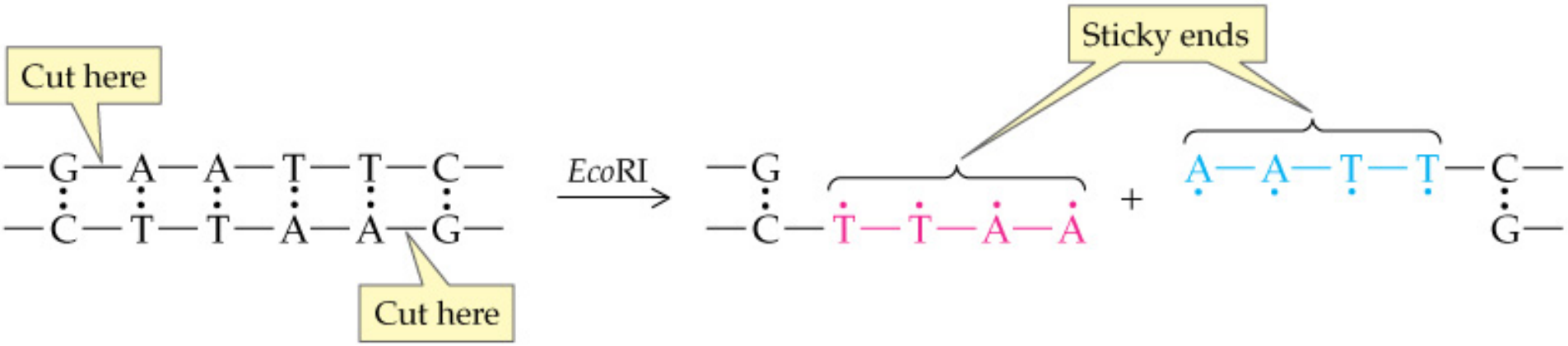

**AluI** and **HaeIII** produce blunt ends

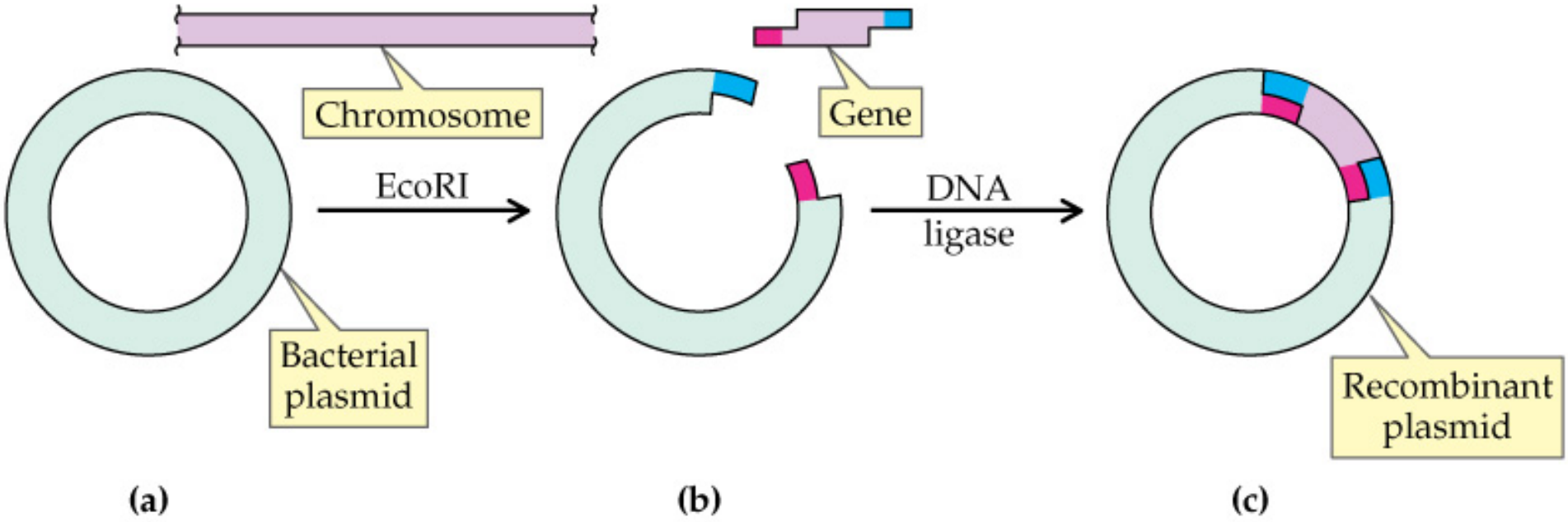
**BamHI** **HindIII** and **EcoRI** produce "sticky" ends

## DNA ligase



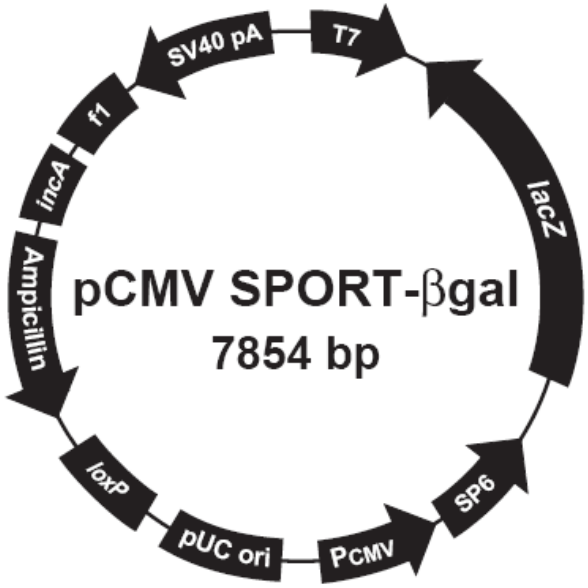


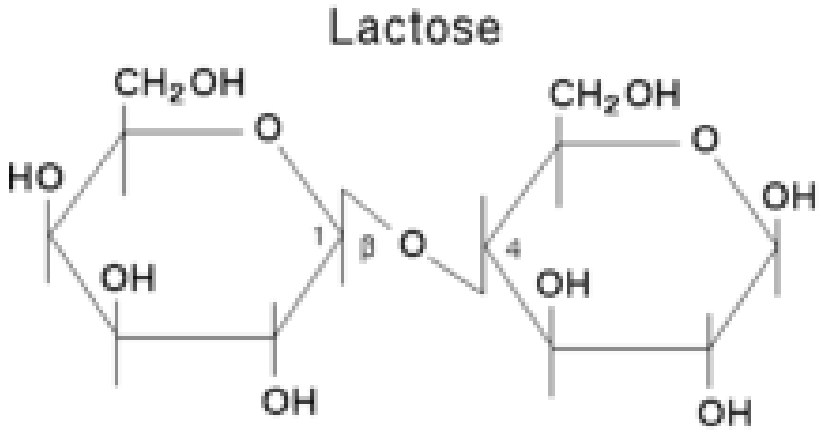

Cut here
G A A T T C
C T T A A G
Cut here
EcoRI
Sticky ends
G
C T T A A
+
A A T T C
G


Chromosome
Gene
EcoRI
DNA
ligase
Bacterial
plasmid
Recombinant
plasmid
(a)
(b)
(c)

# β-Galactosidase

The enzyme that splits lactose into glucose and galactose. Coded by a gene (lacZ) in the lac operon of Escherichia coli.

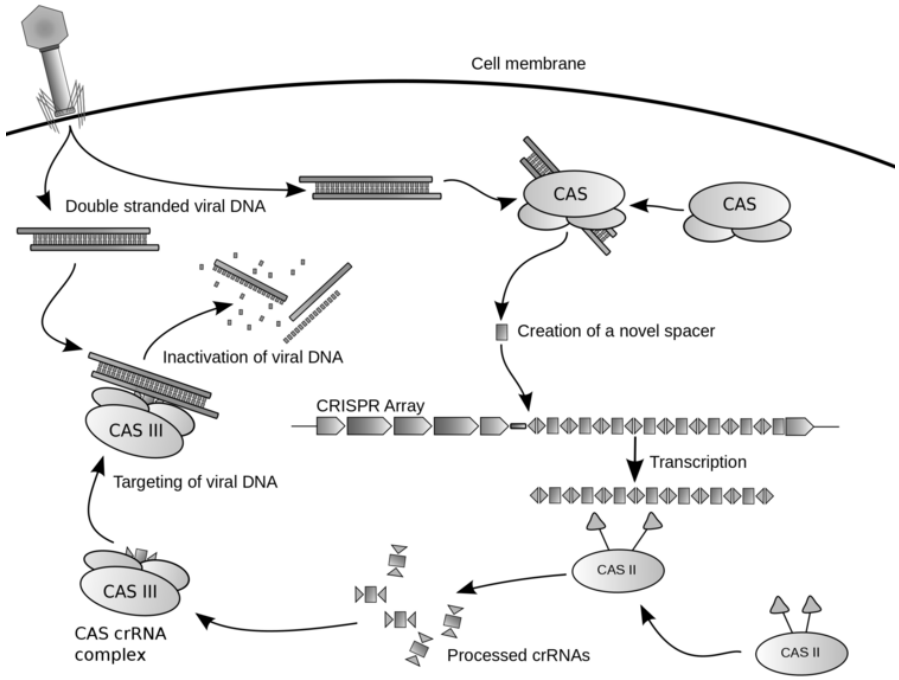



PUC is a family of plasmids that have an ampicillin resistance gene and more importantly a *lacZ* gene. A functional lacZ gene will produce the protein β - galactosidase. Bacterial colonies in which β - galactosidase is produced, will form blue colonies in the presence of the substrate 5 - bromo - 4 - chloro - 3 - indolyl - b - D - galactoside or as it is more commonly referred to, X-gal.

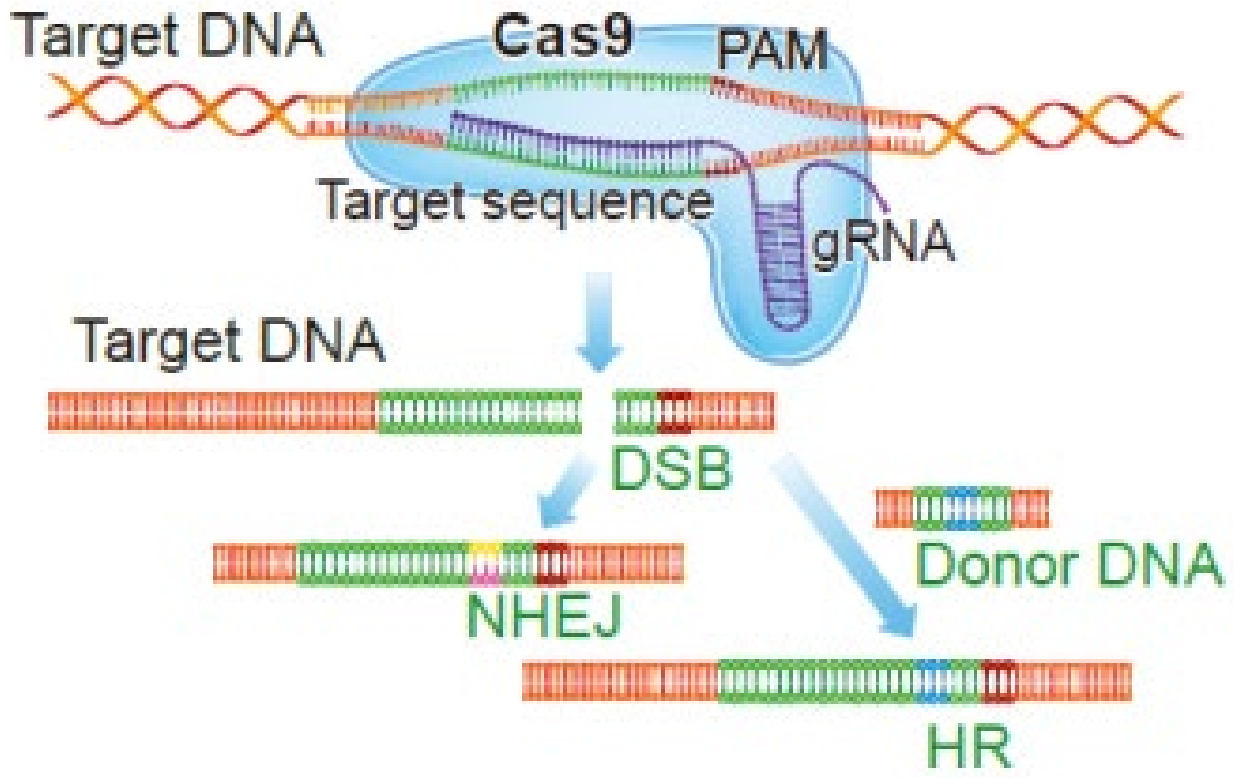
# CRISPR CAS9

**CRISPRs** (**clustered regularly interspaced short palindromic repeats**) are segments of prokaryotic DNA containin g short repetitions of base sequences. Each repetition is followed by short segments of "spacer DNA" from previous exposures to a bacterial virus or



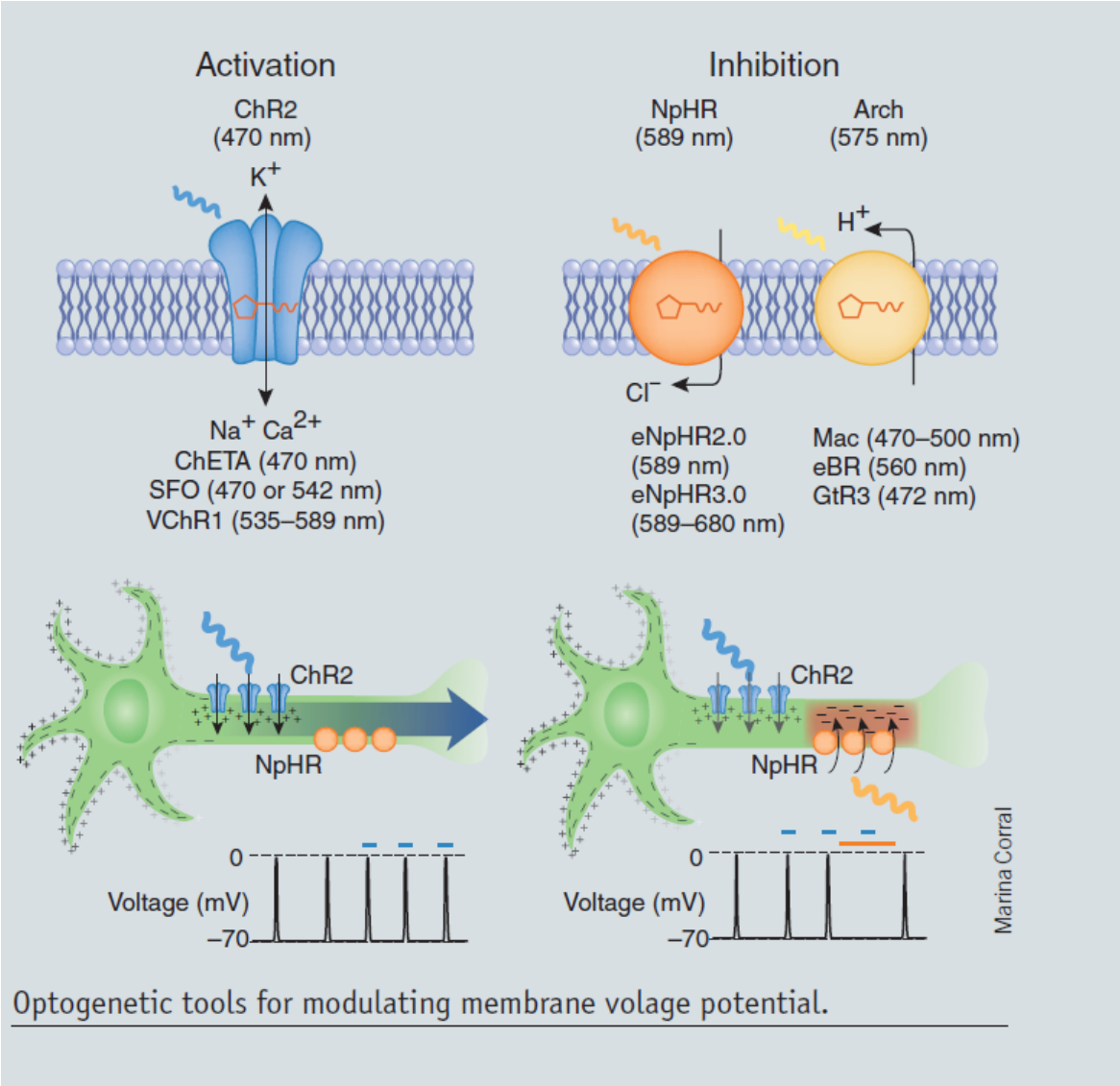
# Movie

- https://youtu.be/2pp17E4E-O8


Target DNA
Cas9
PAM
Target sequence
gRNA
Target DNA
DSB
NHEJ
Donor DNA
HR

# Optogenetics

https://www.youtube.com/watch?v=I64X7vHSHOE




Optogenetic tools for modulating membrane volage potential.