بِسْمِ اللَّهِ الرَّحْمَنِ الرَّحِيمِ

Topic

Next Generation Sequencing (NGS)

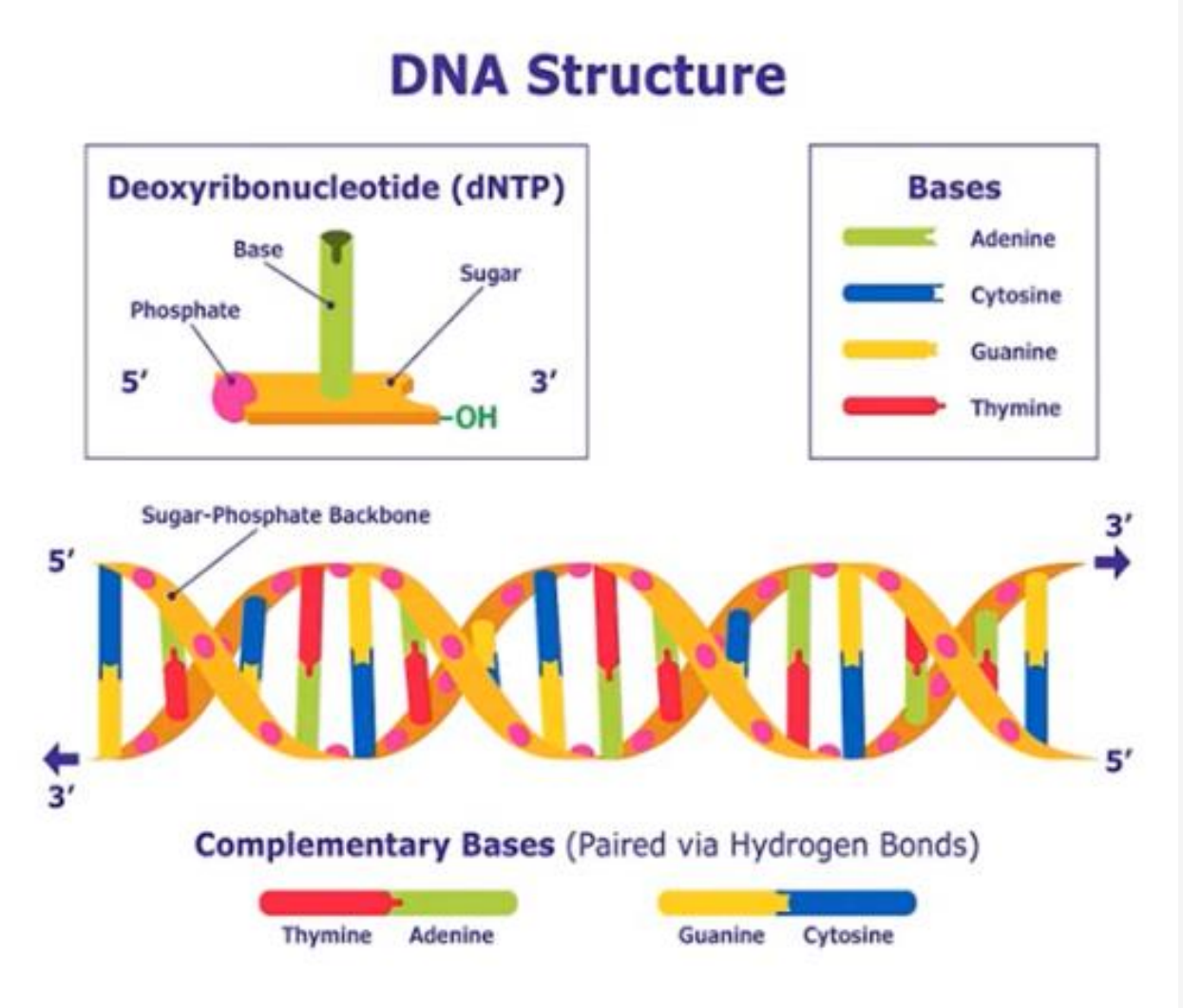
By:

Dr. Abbas Moridnia

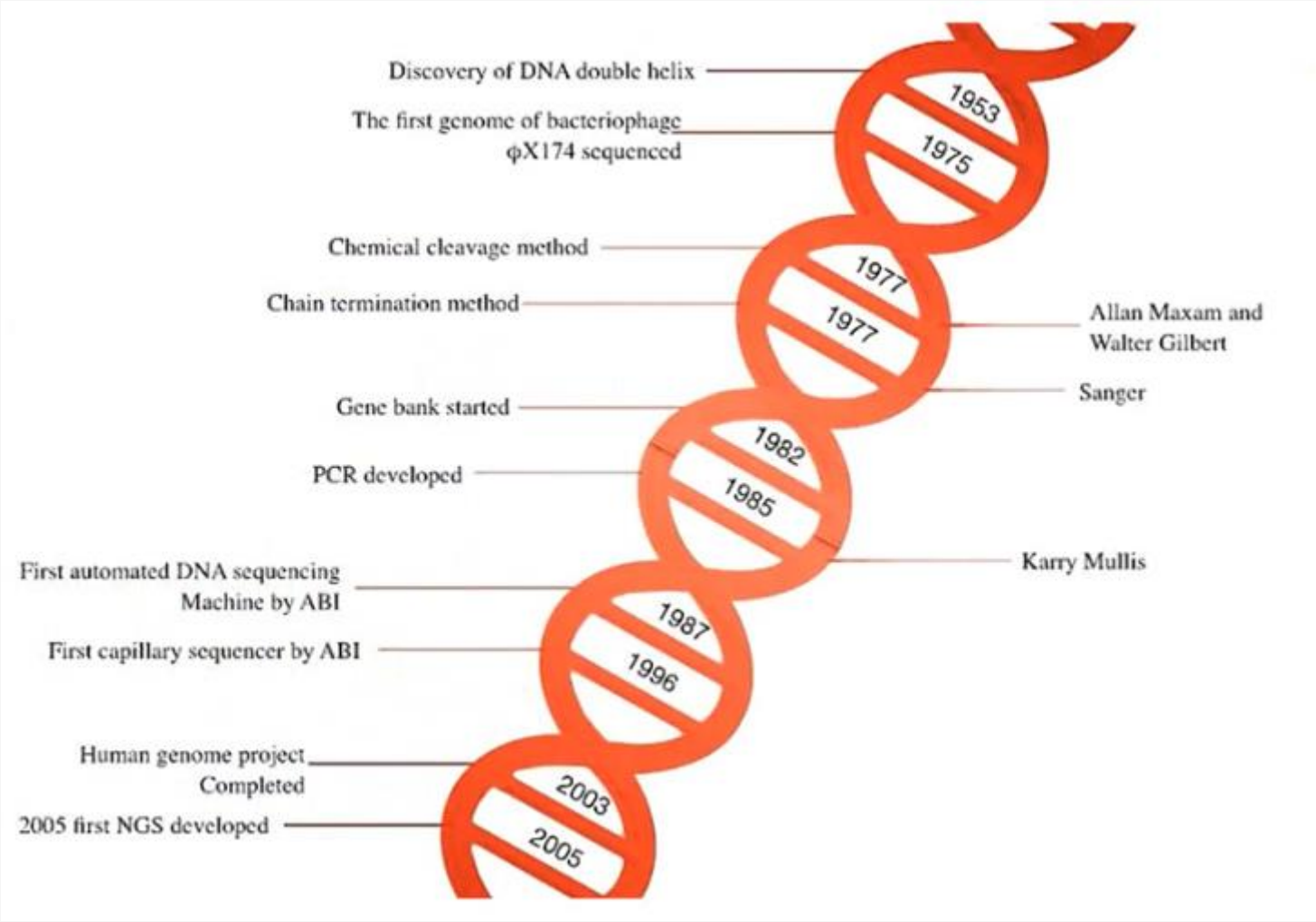
Ph.D of Molecular Medicine

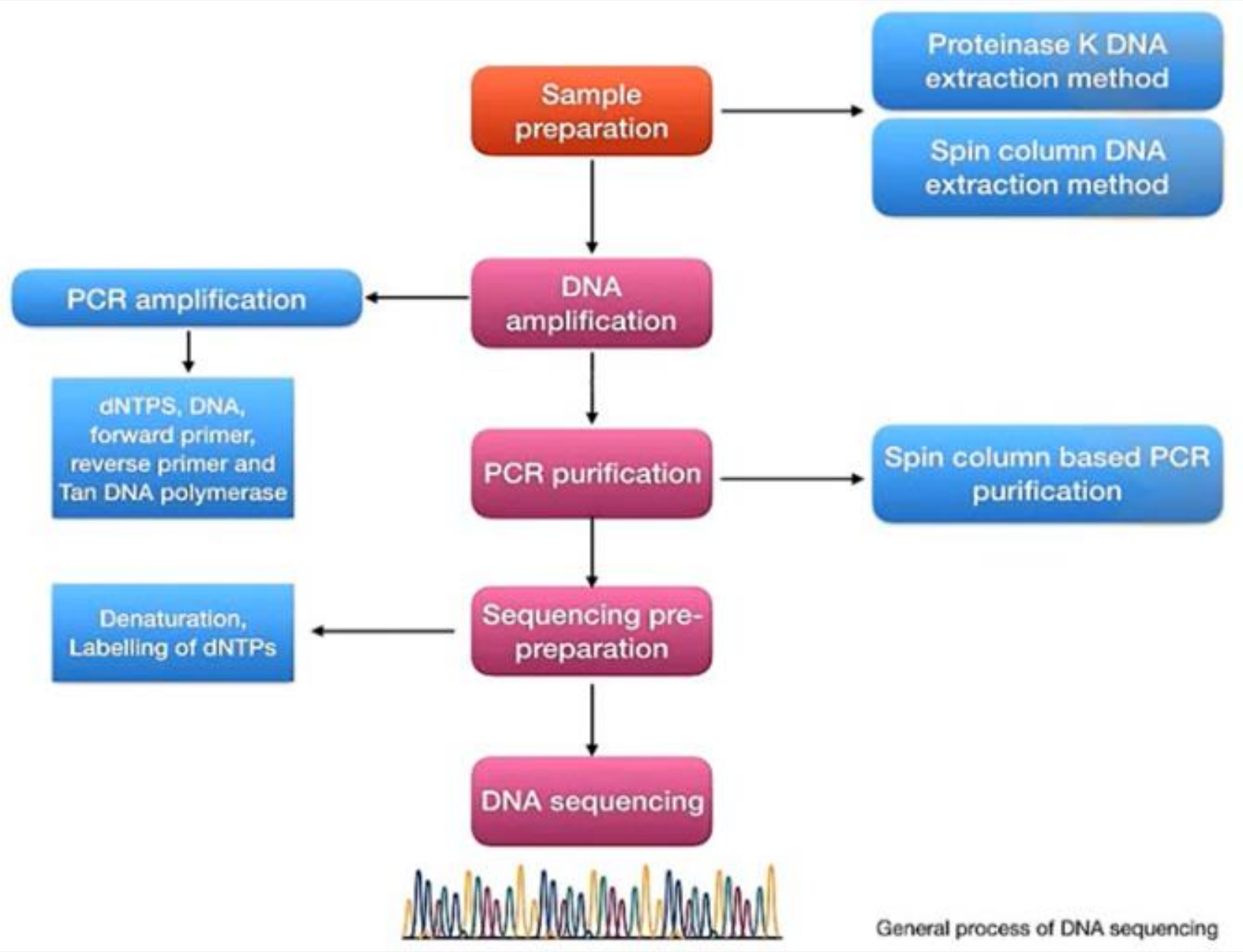
- مقدمه و روش های توالی یابی نسل اول (Sanger, Pyro, Bisulfite sequencing)
- مقدمه ای بر توالی یابی نسل جدید و سرویس های NGS
- روش های توالی یابی نسل دوم (Roch, Illumina, SOLiD)
- روش های توالی یابی نسل سوم (PacBio, Ion Torrent, Oxford Nanopore)
- Omics
- کاربردهای توالی یابی های نسل دوم و سوم

تعیین توالی DNA: تعیین توالی نوکلئیک اسیدها



تاریخچه توالی یابی

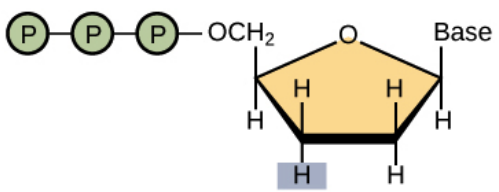




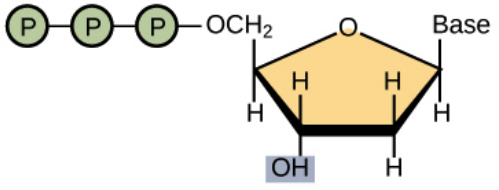
مراحل توالی یابی

- آماده سازی نمونه (استخراج DNA)
- تکثیر توالی هدف
- خالص سازی آمپلیکون ها
- آماده سازی قبل از تعیین توالی
- توالی یابی DNA هدف
- تحلیل داده ها

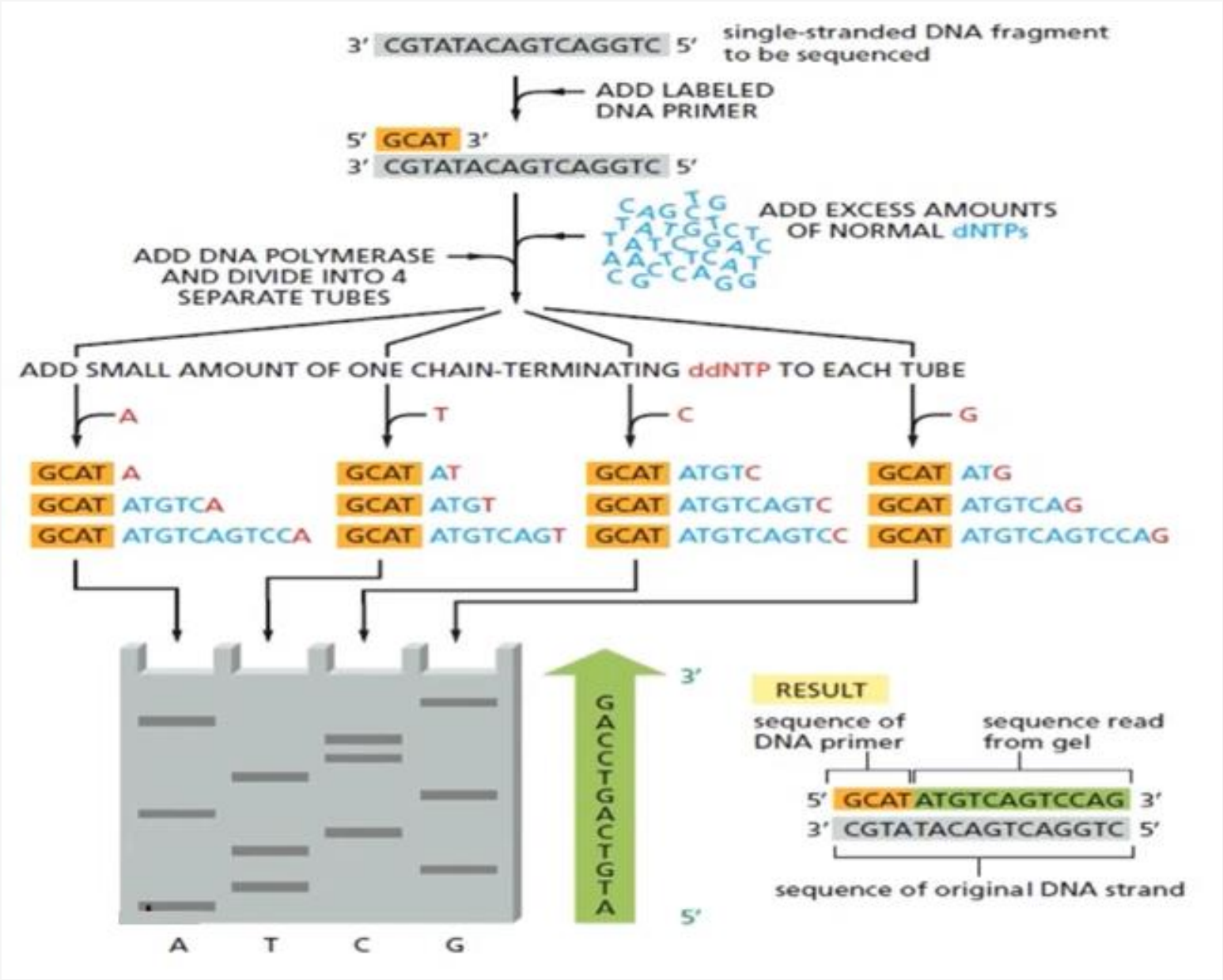
Sanger sequencing

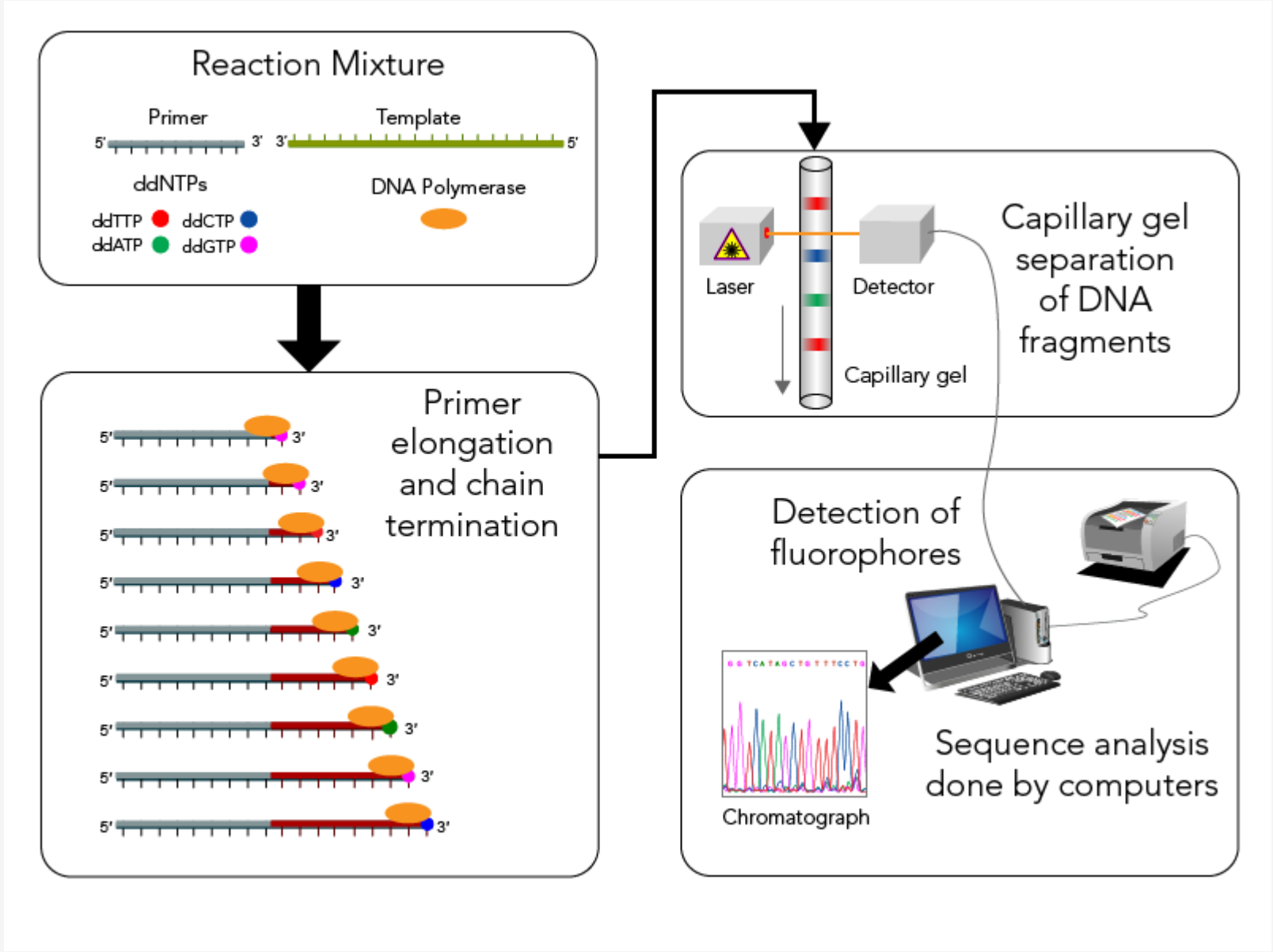


Dideoxynucleotide (ddNTP)



Deoxynucleotide (dNTP)



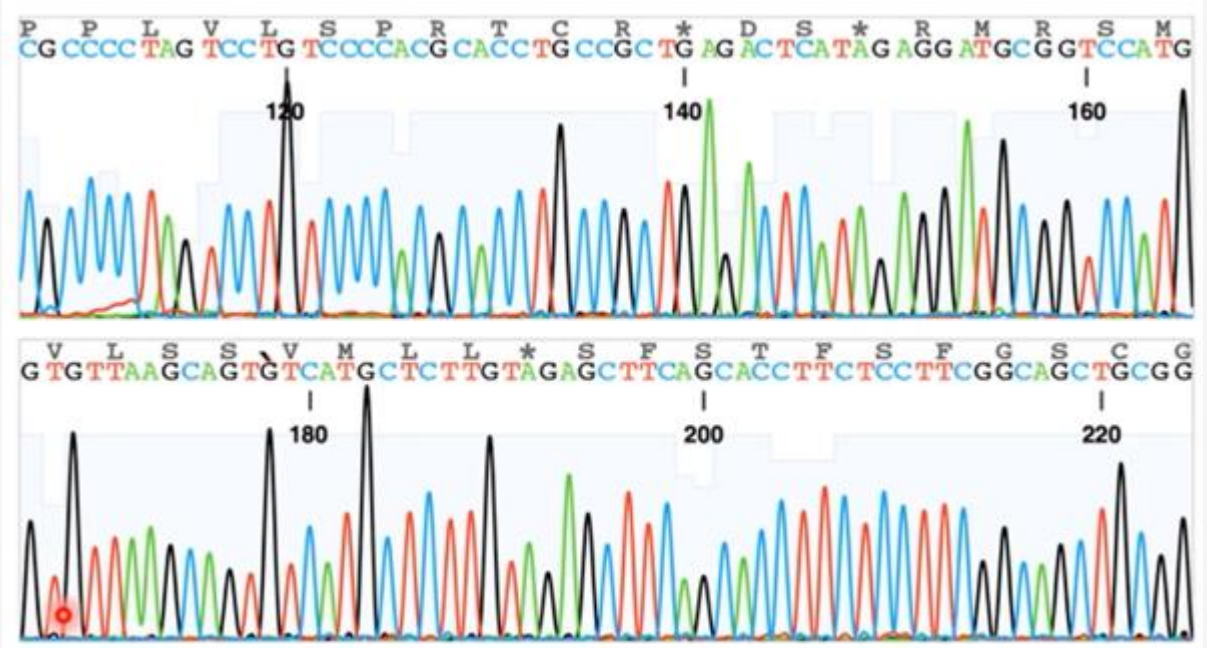

Reaction Mixture
Primer
Template
5'
3'
3'
5'
ddNTPs
DNA Polymerase
ddTTP
ddCTP
ddATP
ddGTP
Primer elongation and chain termination
Laser
Detector
Capillary gel
Capillary gel separation of DNA fragments
Detection of fluorophores
Chromatograph
Sequence analysis done by computers

Sequences are usually delivered as **ABI files**.

Usually thy have **.ab1** extension.

They can be open by:

- ✓ Chromas
- ✓ Sequence scanner (Applid Biosystemes)
- ✓ Finch TV
- ✓ Mega
- ✓ CLC work bench
- ✓ SnapGene
- ✓ ….

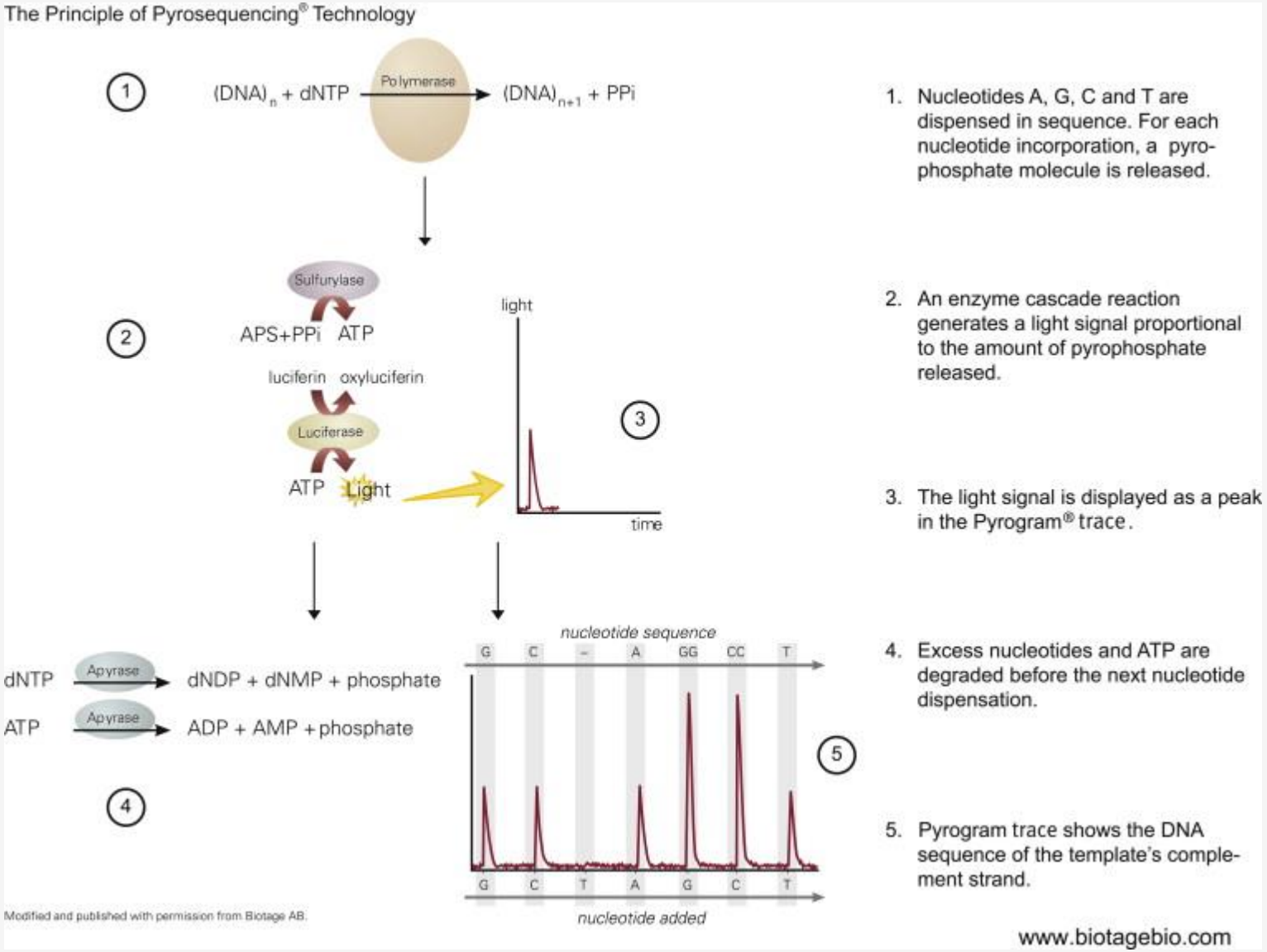


Advantages:

- ✓ Long reads (~900bp)
- ✓ Suitable for small projects

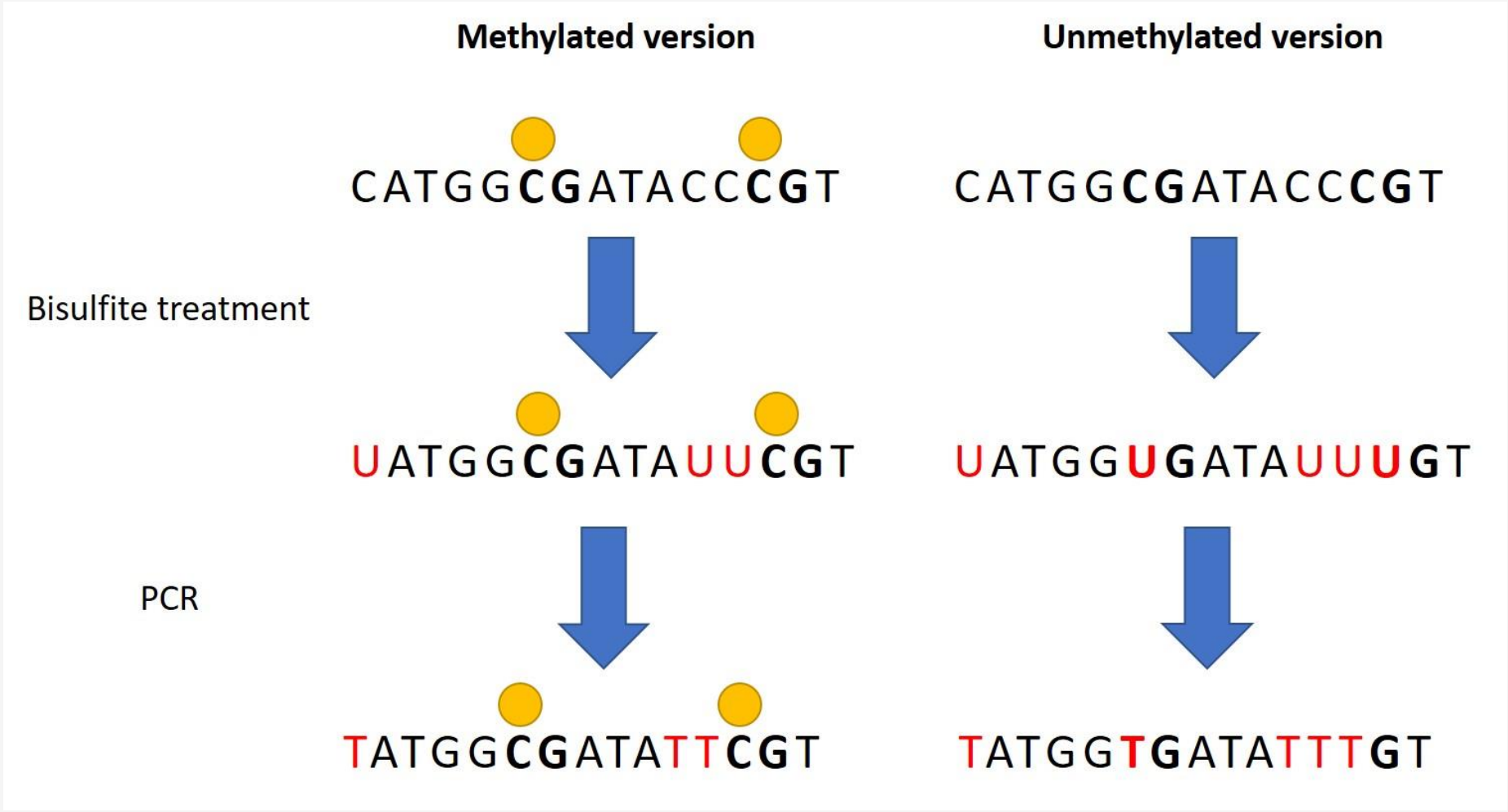
Disadvantages:

- ✓ Expensive
- ✓ Quality is worst at the begging and at the end
- ✓ Error rate 0.1% and error type is substitution.



- Read lengths are around 200-300 bases(short read length).
- Existence of more than 5 nucleotide run may lead to misunderstand in pyrogram analysis.

Bisulfite sequencing



Next generation sequencing (NGS) platforms perform massively parallel sequencing, during which millions of DNA fragments are sequenced in unison.

Advantage:

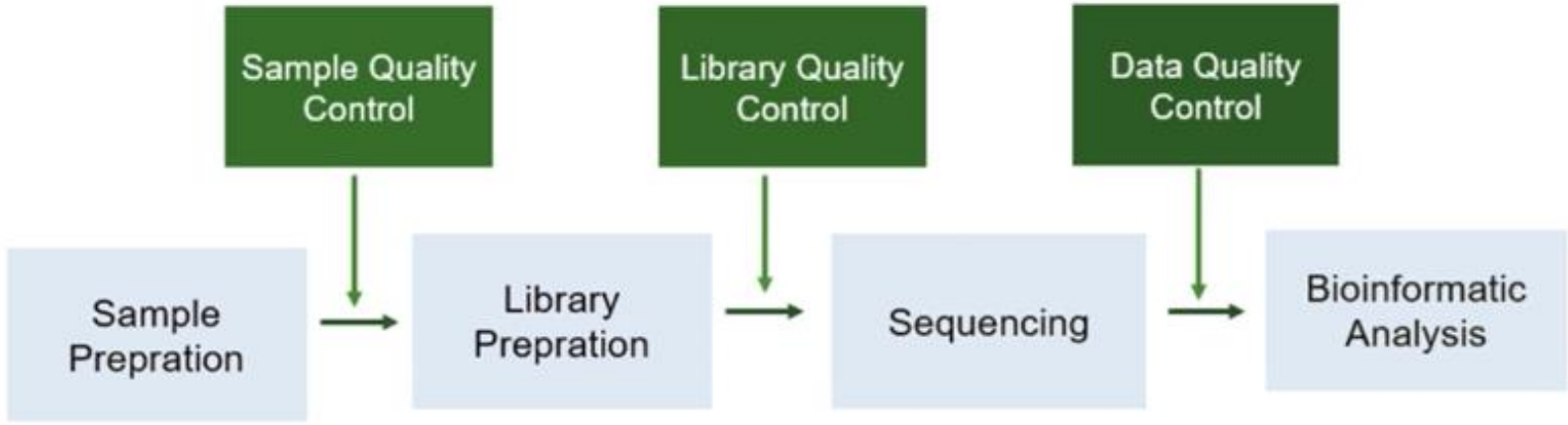
Rapid (sequence an entire genome in less than one day).

Low cost in comparison to traditional techniques (Sanger sequencing)

Sanger vs NGS sequencing

	Sanger	NGS
Num. sequences per reaction	1 clone	Millions of molecules
Max. parallelization	384	Several millions
Sequence quality	High	Low
Sequence length	600-800 bp	35-20000 (depends on the platform)
Throughtput	Low	High

Genome sequencing workflow



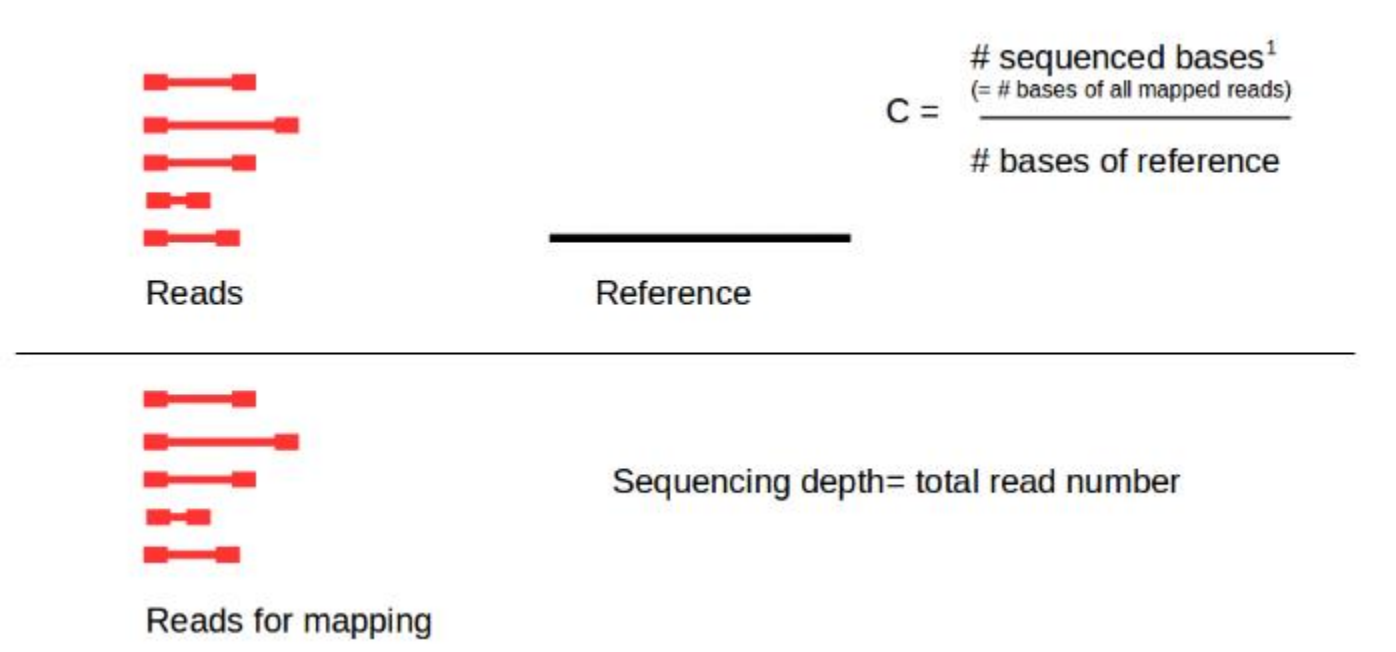
- **Quality** and **quantity** of the DNA and RNA that is extracted from the specimen is very important.

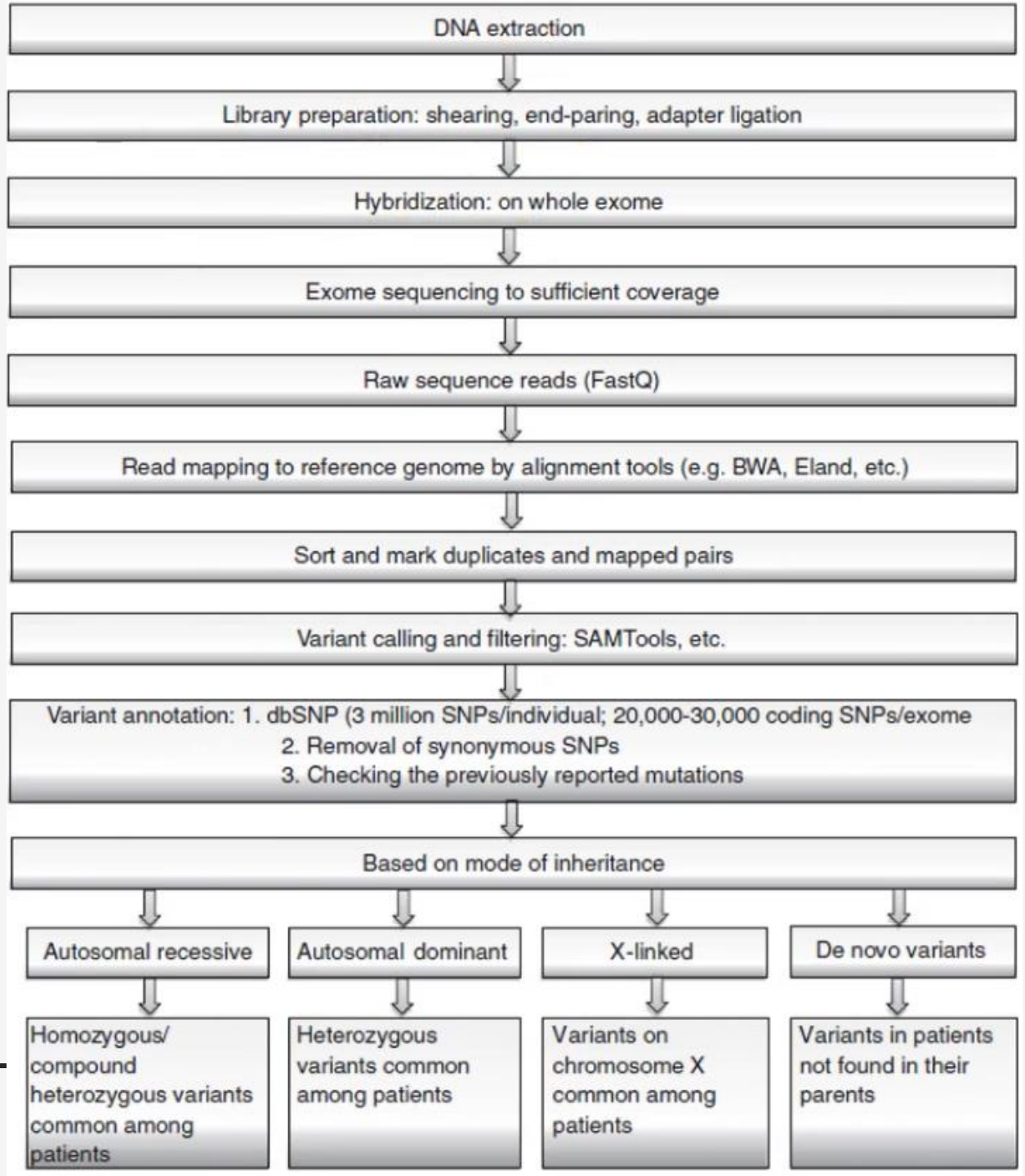
Too much DNA	low quality Data
Too little DNA	Reduced coverage/depth

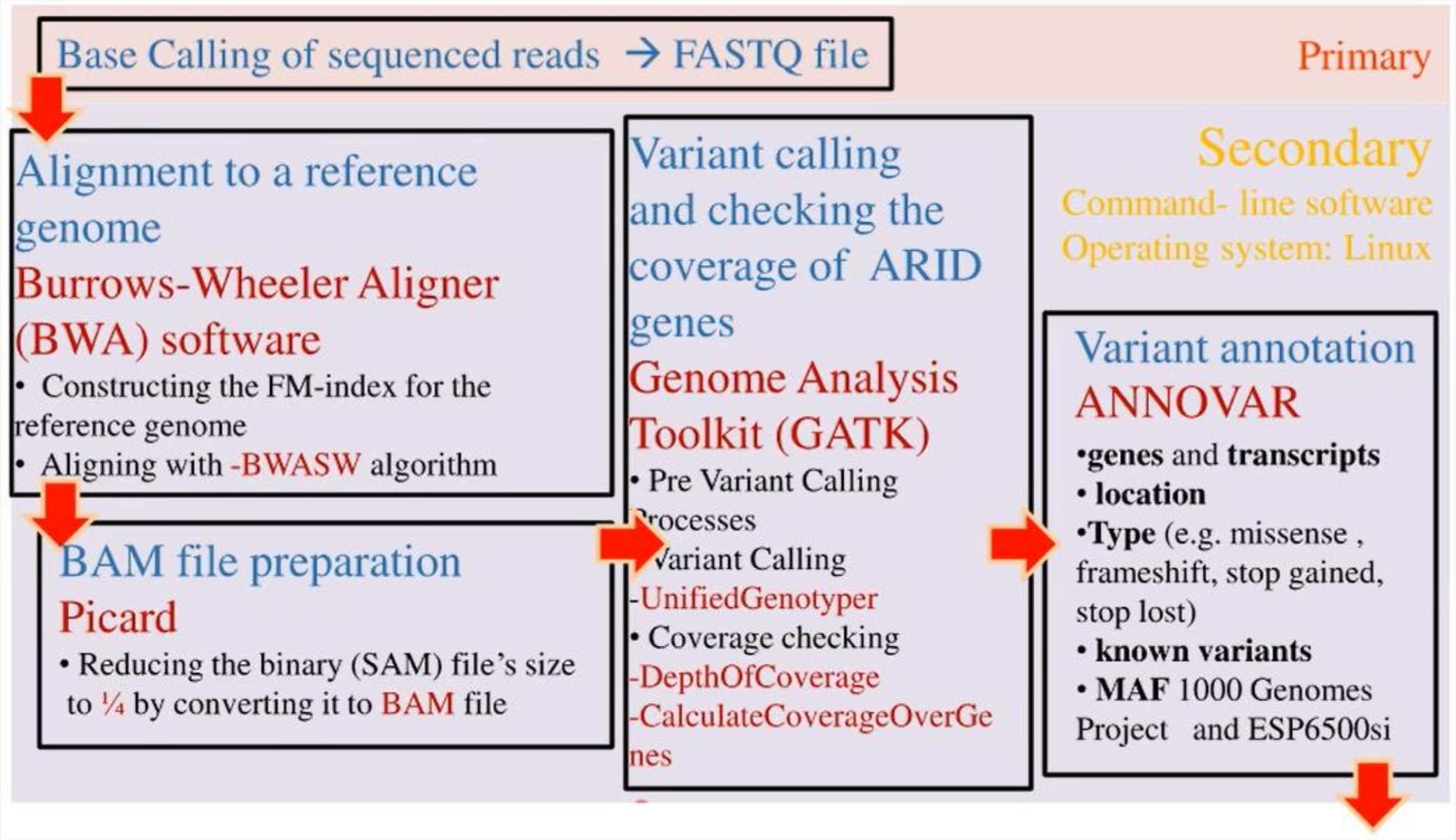
What does coverage mean in the context of NGS?

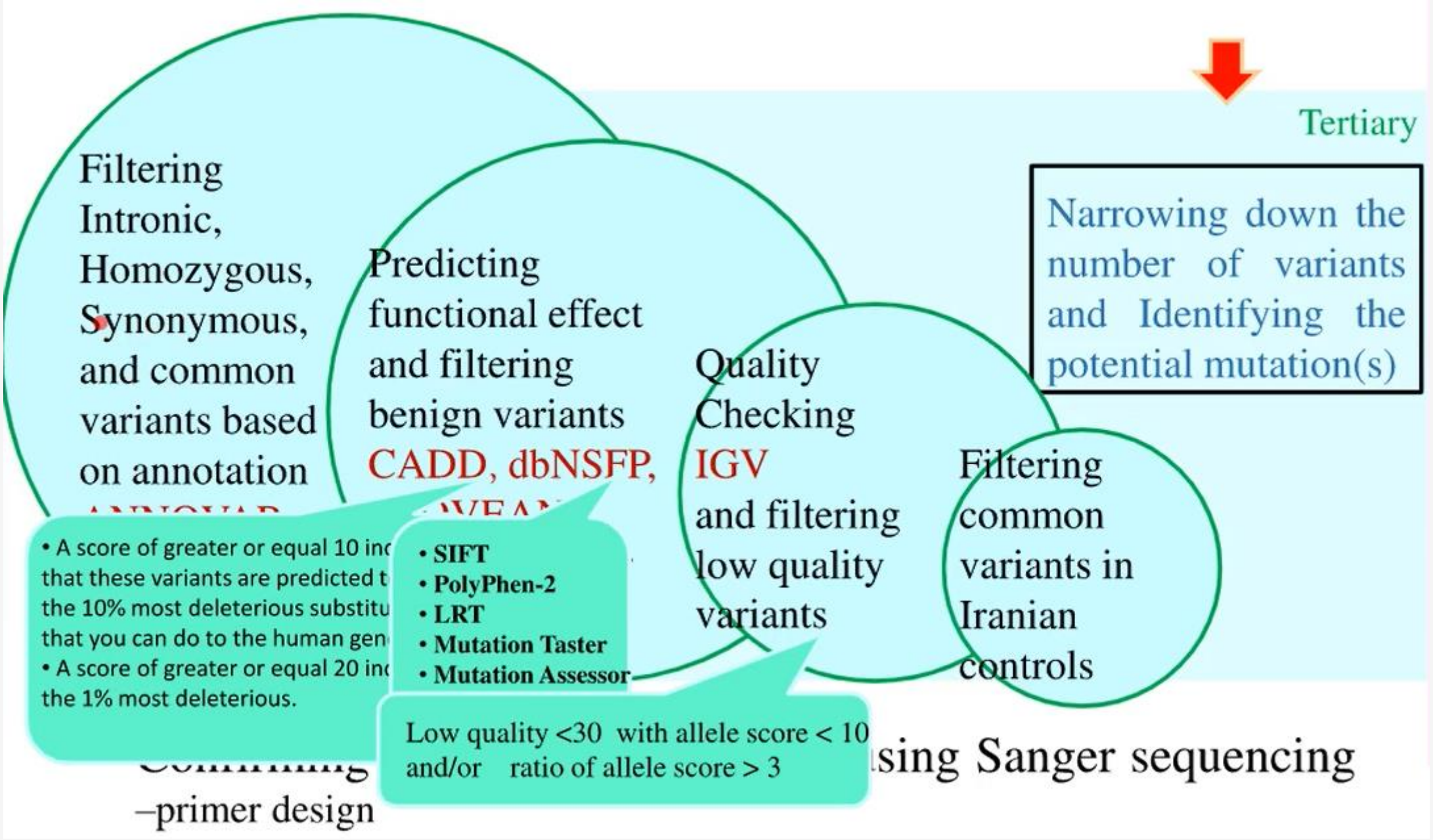
The term "**coverage**" in NGS always describes a relation between sequence reads and a reference (e.g. a whole genome or al locus), unlike sequencing depth which describes a total read number.

Sequencing depth: total number of usable reads from the sequencing machine (usually used in the unit "number of reads" (in millions). Especially used for RNA-seq.

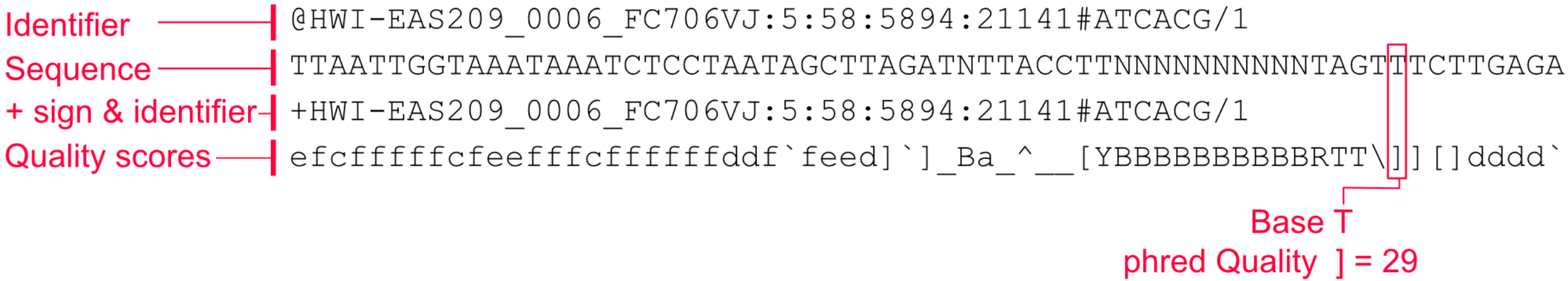


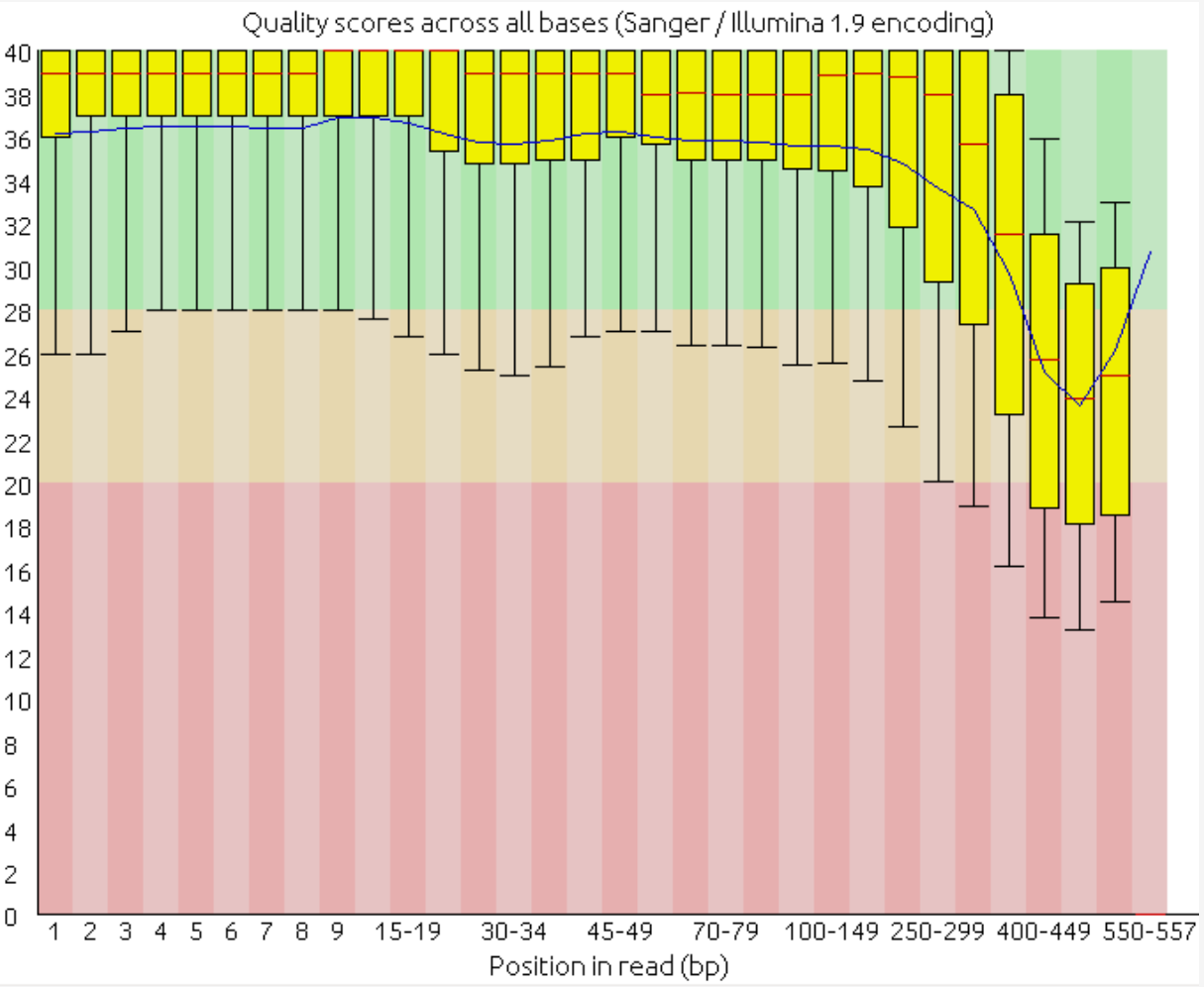

DNA extraction
Library preparation: shearing, end-paring, adapter ligation
Hybridization: on whole exome
Exome sequencing to sufficient coverage
Raw sequence reads (FastQ)
Read mapping to reference genome by alignment tools (e.g. BWA, Eland, etc.)
Sort and mark duplicates and mapped pairs
Variant calling and filtering: SAMTools, etc.
Variant annotation: 1. dbSNP (3 million SNPs/individual; 20,000-30,000 coding SNPs/exome
2. Removal of synonymous SNPs
3. Checking the previously reported mutations
Based on mode of inheritance
Autosomal recessive
Autosomal dominant
X-linked
De novo variants
Homozygous/ compound heterozygous variants common among patients
Heterozygous variants common among patients
Variants on chromosome X common among patients
Variants in patients not found in their parents

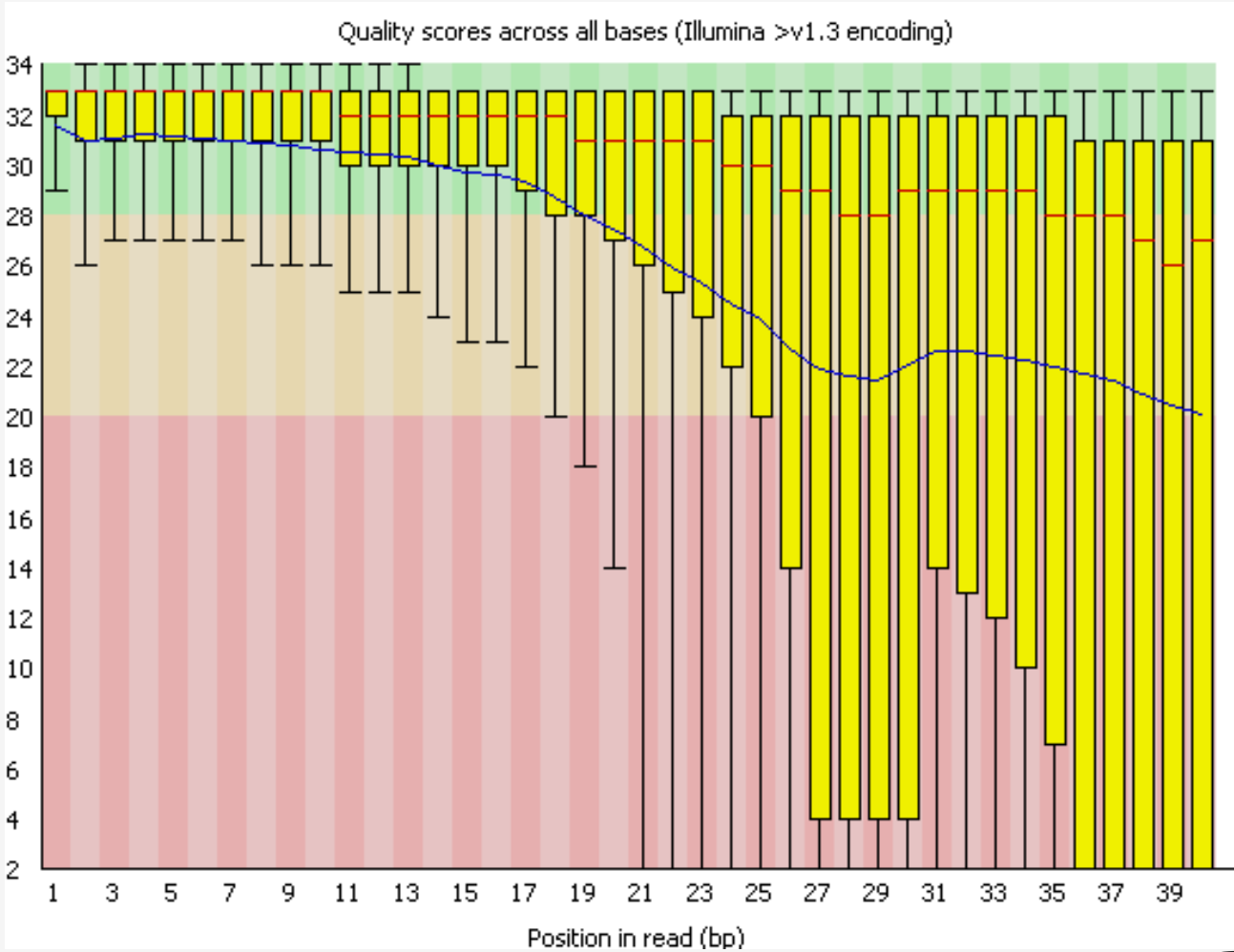

Base Calling of sequenced reads → FASTQ file
Primary
Secondary
Command- line software
Operating system: Linux
Alignment to a reference genome
Burrows-Wheeler Aligner (BWA) software
• Constructing the FM-index for the reference genome
• Aligning with -BWASW algorithm
BAM file preparation
Picard
• Reducing the binary (SAM) file's size to ¼ by converting it to BAM file
Variant calling and checking the coverage of ARID genes
Genome Analysis Toolkit (GATK)
• Pre Variant Calling Processes
• Variant Calling
-UnifiedGenotyper
• Coverage checking
-DepthOfCoverage
-CalculateCoverageOverGenes
Variant annotation
ANNOVAR
•genes and transcripts
• location
•Type (e.g. missense , frameshift, stop gained, stop lost)
• known variants
• MAF 1000 Genomes Project and ESP6500si


Tertiary
Narrowing down the number of variants and Identifying the potential mutation(s)
Filtering Intronic, Homozygous, Synonymous, and common variants based on annotation
Predicting functional effect and filtering benign variants
CADD, dbNSFP,
Quality Checking
IGV
and filtering low quality variants
Filtering common variants in Iranian controls
• A score of greater or equal 10 inc
that these variants are predicted t
the 10% most deleterious substitu
that you can do to the human gen
• A score of greater or equal 20 inc
the 1% most deleterious.
• SIFT
• PolyPhen-2
• LRT
• Mutation Taster
• Mutation Assessor
Low quality <30 with allele score < 10
and/or ratio of allele score > 3
sing Sanger sequencing
–primer design

FASTQ


Identifier
Sequence
+ sign & identifier
Quality scores
@HWI-EAS209_0006_FC706VJ:5:58:5894:21141#ATCACG/1
TTAATTGGTAAATAAATCTCCTAATAGCTTAGATNTTACCTTNNNNNNNNNNTAGTTTCTTGAGA
+HWI-EAS209_0006_FC706VJ:5:58:5894:21141#ATCACG/1
efcfffffcfeefffcffffffddf`feed]`]_Ba_^__[YBBBBBBBBBBRTT\]][]dddd`
Base T
phred Quality] = 29


Quality scores across all bases (Sanger / Illumina 1.9 encoding)
40
38
36
34
32
30
28
26
24
22
20
18
16
14
12
10
8
6
4
2
0
1 2 3 4 5 6 7 8 9 15-19 30-34 45-49 70-79 100-149 250-299 400-449 550-557
Position in read (bp)


Quality scores across all bases (Illumina >v1.3 encoding)
34
32
30
28
26
24
22
20
18
16
14
12
10
8
6
4
2
1 3 5 7 9 11 13 15 17 19 21 23 25 27 29 31 33 35 37 39
Position in read (bp)

Phred Quality Score	Probability of incorrect base call	Base call accuracy
10	1 in 10	90%
20	1 in 100	99%
30	1 in 1000	99.9%
40	1 in 10,000	99.99%
50	1 in 100,000	99.999%
60	1 in 1,000,000	99.9999%

NGS services

Whole Genome Sequencing (WGS)

Whole Exome Sequencing (WES)

De novo sequencing

Targeted sequencing (Panel)

Transcriptomics (RNA-Seq)

Epigenomics (Methyl-Seq)

Chip-Seq

What is Whole-Genome Sequencing?

Whole-genome sequencing (WGS) is a comprehensive method for analyzing entire genomes. Genomic information has been instrumental in identifying inherited disorders, characterizing the mutations that drive cancer progression, and tracking disease outbreaks. Rapidly dropping sequencing costs and the ability to produce large volumes of data with today's sequencers make whole-genome sequencing a powerful tool for genomics research.

Advantages of Whole-Genome Sequencing

- ✓ Provides a high-resolution, base-by-base view of the genome
- ✓ Captures both large and small variants that might be missed with targeted approaches
- ✓ Identifies potential causative variants for further follow-up studies of gene expression and regulation mechanisms
- ✓ Delivers large volumes of data in a short amount of time to support assembly of novel genomes

What is Exome Sequencing?

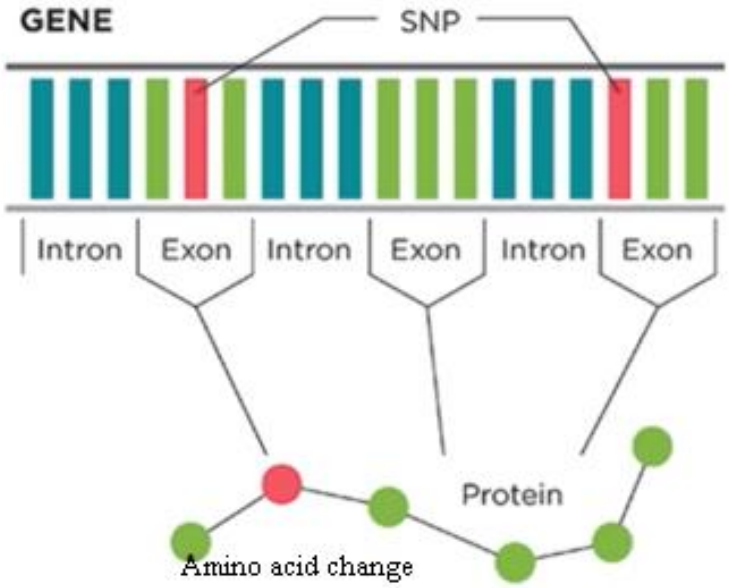
Whole-exome sequencing is a widely used next-generation sequencing (NGS) method that involves sequencing the protein-coding regions of the genome. The human exome represents less than 2% of the genome, but contains ~85% of known disease-related variants, making this method a cost-effective alternative to whole-genome sequencing.

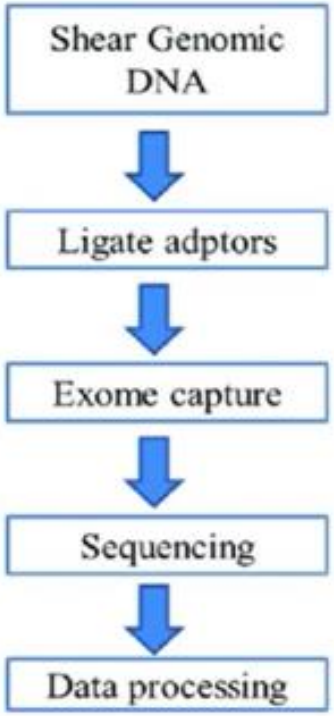
Exome sequencing using exome enrichment can efficiently identify coding variants across a broad range of applications, including population genetics, genetic disease, and cancer studies.

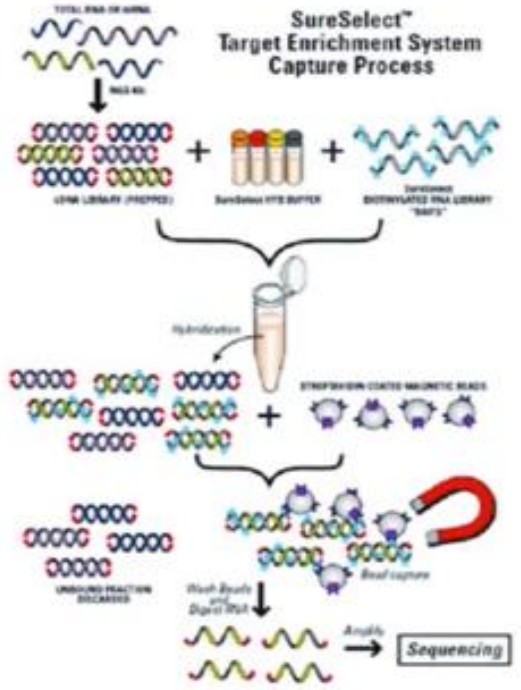
Advantages of Exome Sequencing

- ✓ Identifies variants across a wide range of applications
- ✓ Achieves comprehensive coverage of coding regions
- ✓ Provides a cost-effective alternative to whole-genome sequencing (4–5 Gb of sequencing per exome compared to ~90 Gb per whole human genome)
- ✓ Produces a smaller, more manageable data set for faster, easier data analysis compared to whole-genome approaches


DNA

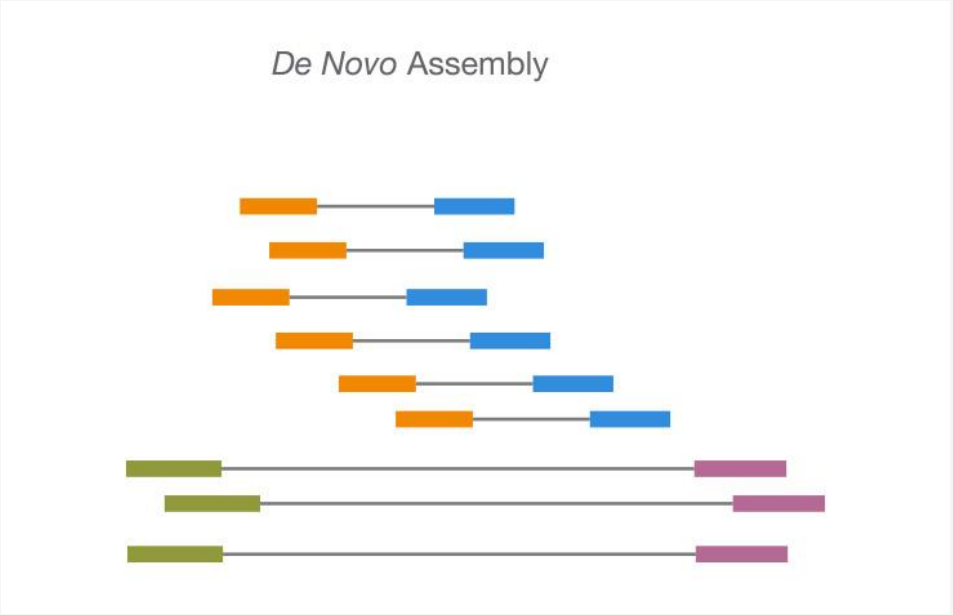

GENE
SNP
Intron
Exon
Intron
Exon
Intron
Exon
Protein
Amino acid change


Shear Genomic DNA
Ligate adptors
Exome capture
Sequencing
Data processing


SureSelect™
Target Enrichment System
Capture Process
Sequencing

What Is De Novo Sequencing?

De novo sequencing refers to sequencing a novel genome where there is no reference sequence available for alignment. Sequence reads are assembled as contigs, and the coverage quality of de novo sequence data depends on the size and continuity of the contigs (ie, the number of gaps in the data).



Advantages of De Novo Sequencing

- ✓ Generates accurate reference sequences, even for complex or polyploid genomes
- ✓ Provides useful information for mapping genomes of novel organisms or finishing genomes of known organisms
- ✓ Clarifies highly similar or repetitive regions for accurate de novo assembly
- ✓ Identifies structural variants and complex rearrangements, such as deletions, inversions, or translocations

Introduction to Targeted Gene Sequencing (Panel)

Targeted gene sequencing panels are useful tools for analyzing specific mutations in a given sample. Focused panels contain a select set of genes or gene regions that have known or suspected associations with the disease or phenotype under study. Gene panels can be purchased with preselected content or custom designed to include genomic regions of interest.

Targeted gene sequencing produces a smaller, more manageable data set compared to broader approaches such as whole-genome sequencing, making analysis easier.

Syndromic Hearing Loss Panel (89) Test code: EA0401	*ABHD12*	*CEP78*	*COL9A3*	*HOXB1*	*PAX3*	*SLC52A3*
	ACTG1	*CHD7*	*DCAF17*	*KCNE1*	*PCDH15*	*SLITRK6*
	ADGRV1	*CHSY1*	*DFNB31*	*KCNJ10*	*PDZD7*	*SMAD4*
	ALMS1	*CIB2*	*DLX5*	*KCNQ1*	*PEX1*	*SNAI2*
	ANKH	*CLPP*	*DNMT1*	*KIT*	*PEX26*	*SOX10*
	ATP6V1B1	*CLRN1*	*EDN3*	*LARS2*	*PEX6*	*TCOF1*
	ATP6V1B2	*COL11A1*	*EDNRB*	*LRP2*	*POLR1C*	*TFAP2A*
	BCS1L	*COL11A2*	*EYA1*	*MAN2B1*	*POLR1D*	*TIMM8A*
	BSND	*COL2A1*	*FDXR*	*MANBA*	*SALL1*	*TYR*
	BTD	*COL4A3*	*FGF3*	*MGP*	*SEMA3E*	*USH1C*
	C10ORF2	*COL4A4*	*FOXI1*	*MITF*	*SIX1*	*USH1G*
	CACNA1D	*COL4A5*	*GATA3*	*MYH9*	*SIX5*	*USH2A*
	CD151	*COL4A6*	*GJA1*	*MYO7A*	*SLC19A2*	*VCAN*
	CDH23	*COL9A1*	*HARS*	*NDP*	*SLC26A4*	*WFS1*
	CDKN1C	*COL9A2*	*HARS2*	*NLRP3*	*SLC52A2*	
Usher Syndrome Panel (15) Test code: OP1101	*ABHD12*	*CEP78*	*DFNB31*	*PCDH15*	*USH1C*	
	ADGRV1	*CIB2*	*HARS*	*PDZD7*	*USH1G*	
	CDH23	*CLRN1*	*MYO7A*	*PEX1*	*USH2A*	
Waardenburg Syndrome Panel (7) Test code: EA0101	*EDN3*	*KIT*	*PAX3*	*SOX10*		
	EDNRB	*MITF*	*SNAI2*			

Transcriptomics (RNA-Seq)

RNA-Seq (named as an abbreviation of RNA sequencing) is a sequencing technique which uses next-generation sequencing (NGS) to reveal the presence and quantity of RNA in a biological sample at a given moment, analyzing the continuously changing cellular transcriptome.

Specifically, RNA-Seq facilitates the ability to look at alternative gene spliced transcripts, post-transcriptional modifications, gene fusion, mutations/SNPs and changes in gene expression over time, or differences in gene expression in different groups or treatments.

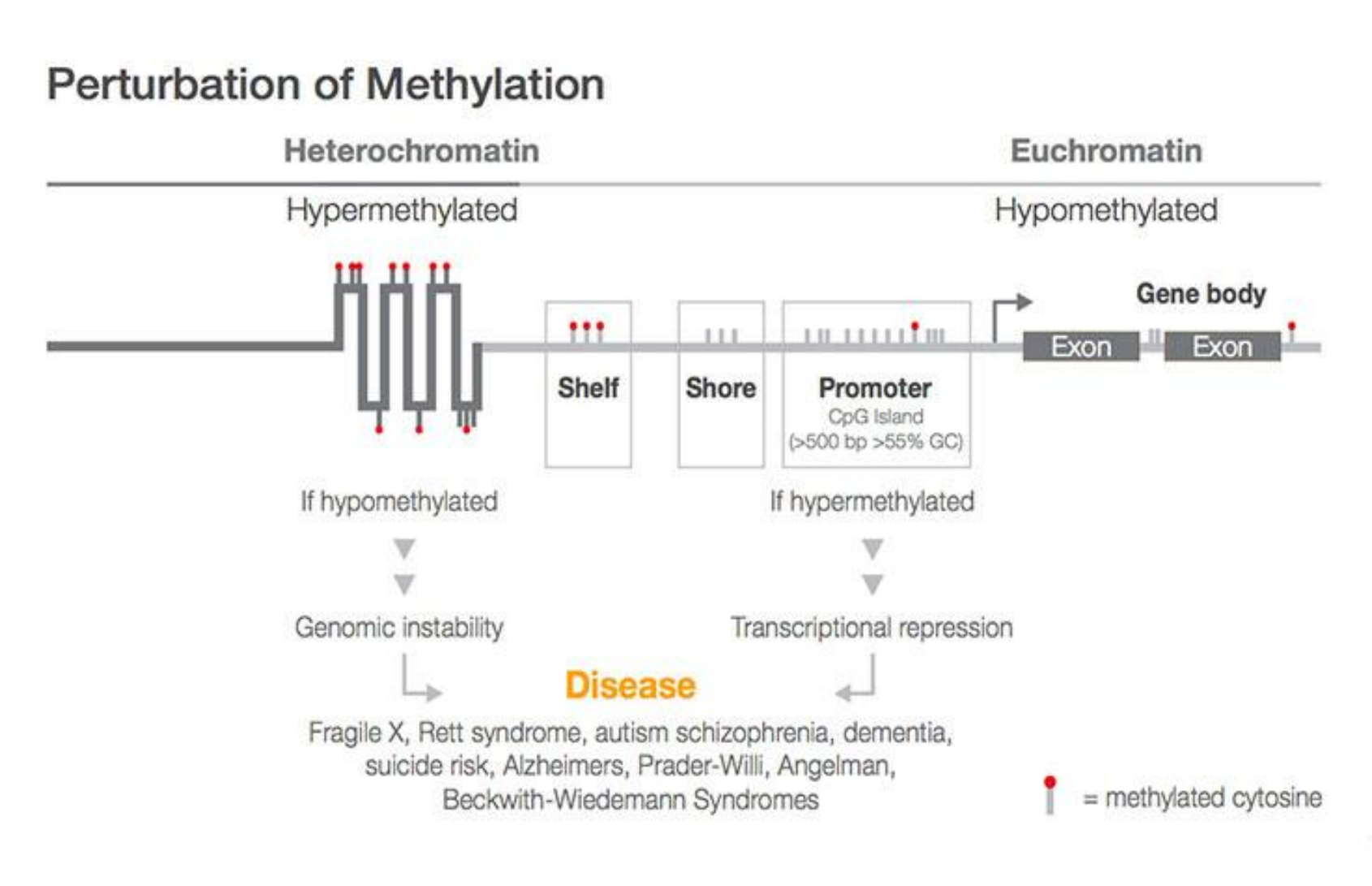
In addition to mRNA transcripts, RNA-Seq can look at different populations of RNA to include total RNA, small RNA, such as miRNA, tRNA, and ribosomal profiling. RNA-Seq can also be used to determine exon/intron boundaries and verify or amend previously annotated 5' and 3' gene boundaries. Recent advances in RNA-Seq include single cell sequencing, in situ sequencing of fixed tissue, and native RNA molecule sequencing with single-molecule real-time sequencing. Other examples of emerging RNA-Seq applications due to the advancement of bioinformatics algorithms are copy number alteration, microbial contamination, transposable elements, cell type (deconvolution) and the presence of neoantigens.

Epigenomics

Epigenetics focuses on processes that regulate how and when certain genes are turned on and turned off, while epigenomics pertains to the analysis of epigenetic changes across many genes in a cell or entire organism.

Epigenetic processes control normal growth and development and this process is deregulated in diseases such as cancer. Diet and exposure to environmental chemicals throughout all stages of human development among other factors can cause epigenetic changes that may turn on or turn off certain genes. Changes in genes that would normally protect against disease, as a result, could make people more susceptible to developing that disease later in life.

Whole genome methylation coverage



What is ChIP-Seq?

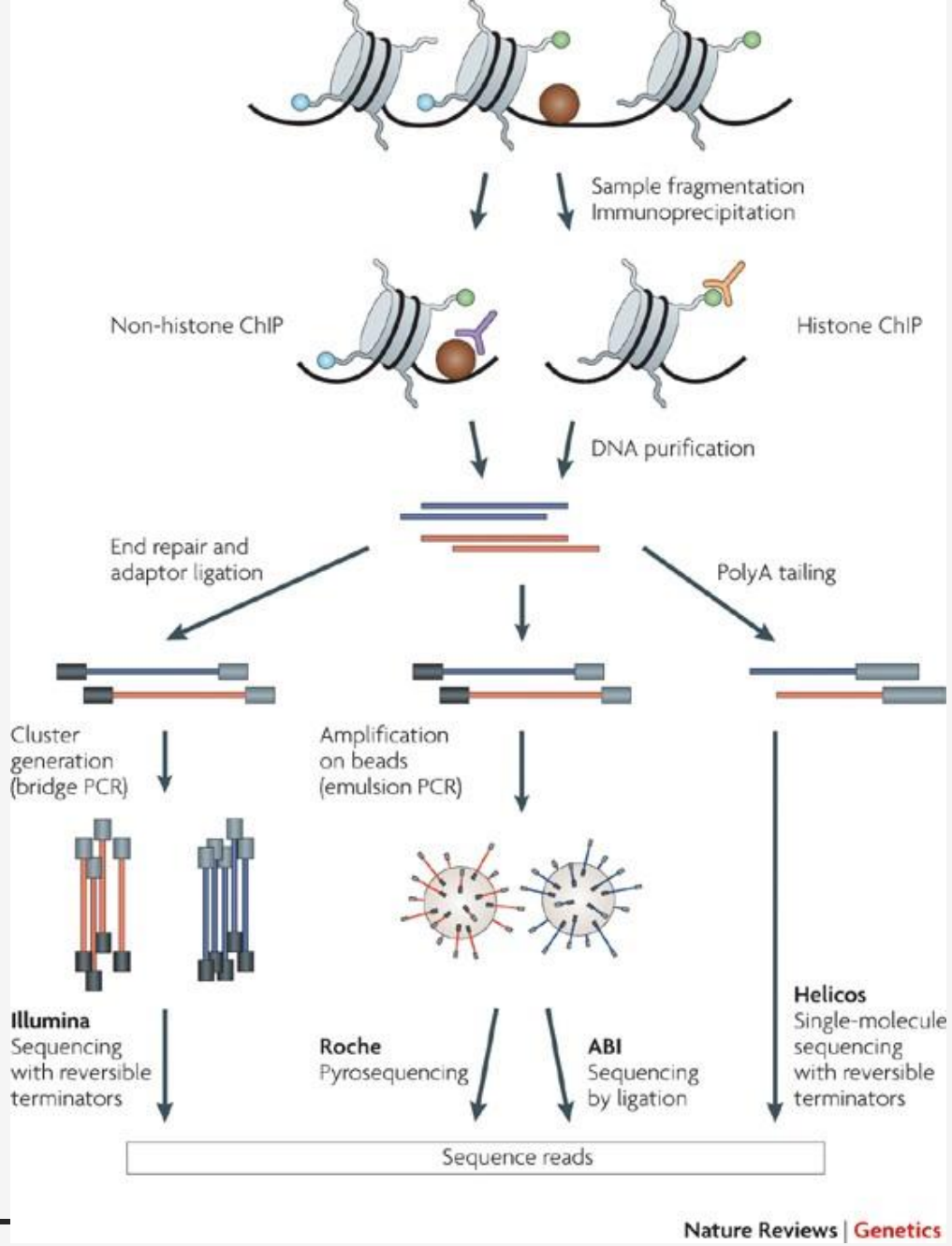
By combining chromatin immunoprecipitation (ChIP) assays with sequencing, ChIP sequencing (ChIP-Seq) is a powerful method for identifying genome-wide DNA binding sites for transcription factors and other proteins. Following ChIP protocols, DNA-bound protein is immunoprecipitated using a specific antibody. The bound DNA is then coprecipitated, purified, and sequenced.

The application of next-generation sequencing (NGS) to ChIP has revealed insights into gene regulation events that play a role in various diseases and biological pathways, such as development and cancer progression. ChIP-Seq enables thorough examination of the interactions between proteins and nucleic acids on a genome-wide scale.

Advantages of ChIP-Seq

Unlike arrays and other approaches used to investigate the epigenome, which are inherently biased because they require probes derived from known sequences, ChIP-Seq does not require prior knowledge. ChIP-Seq delivers genome-wide profiling with massively parallel sequencing, generating millions of counts across multiple samples for cost-effective, precise, unbiased investigation of epigenetic patterns. Additional advantages include:

- ✓ Captures DNA targets for transcription factors or histone modifications across the entire genome of any organism
- ✓ Defines transcription factor binding sites
- ✓ Reveals gene regulatory networks in combination with RNA sequencing and methylation analysis
- ✓ Offers compatibility with various input DNA samples


Sample fragmentation
Immunoprecipitation
Non-histone ChIP
Histone ChIP
DNA purification
End repair and
adaptor ligation
PolyA tailing
Cluster
generation
(bridge PCR)
Amplification
on beads
(emulsion PCR)
Illumina
Sequencing
with reversible
terminators
Roche
Pyrosequencing
ABI
Sequencing
by ligation
Helicos
Single-molecule
sequencing
with reversible
terminators
Sequence reads
Nature Reviews | Genetics

تکنولوژی های نسل دوم NGS

- 454 (Roche/ FLX) 2004
- Solexa (Illumina) 2006
- Applied Biosystem SOLiD 2007

454-life science technology

- اولین سری از تکنولوژی‌های نسل دوم
- سال ۲۰۰۴ ⟵ جاناتان روتبرگ

(FLX) 454 life science (Roche)

- Template preparation
- Imaging
- Sequencing

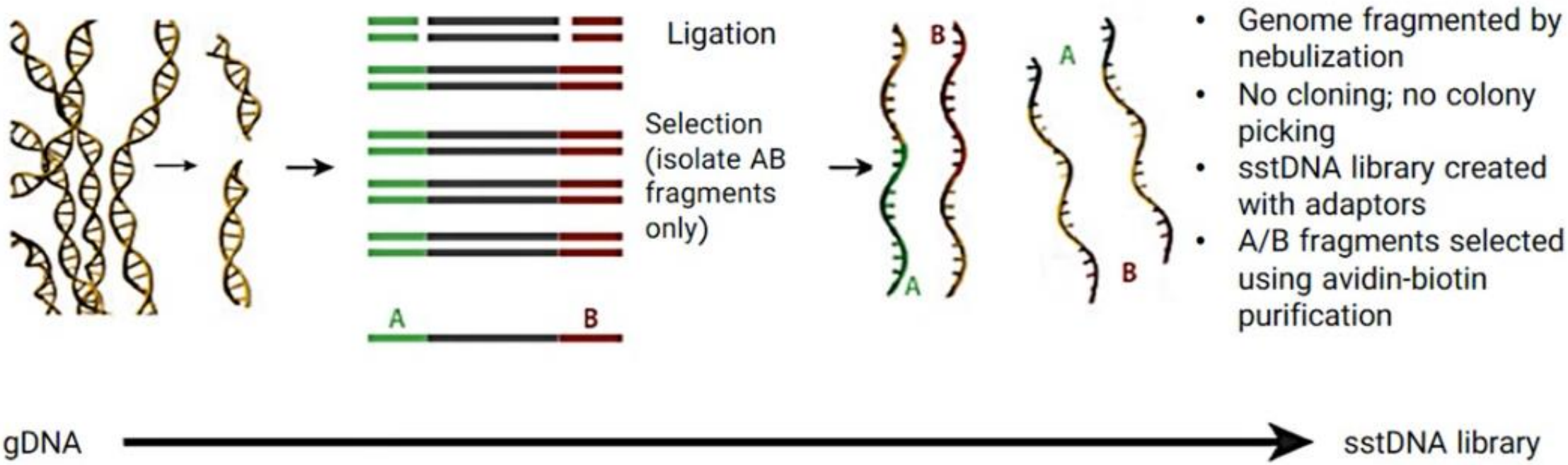
- Emulsion PCR (Em – PCR)
- Charge-coupled device (CCD camera)
- Single nucleotide Addition (SNA)

Template preparation

a)

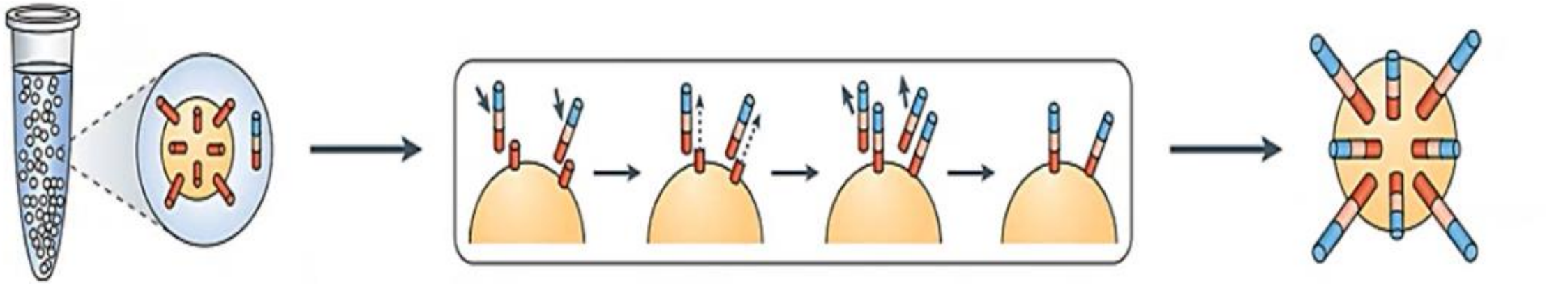
DNA library preparation

4.5 hours



- Genome fragmented by nebulization
- No cloning; no colony picking
- sstDNA library created with adaptors
- A/B fragments selected using avidin-biotin purification

Template preparation (Emulsion PCR)



Emulsion
Micelle droplets are loaded with primer, template, dNTPs and polymerase

On-bead amplification
Templates hybridize to bead-bound primers and are amplified; after amplification, the complement strand disassociates, leaving bead-bound ssDNA templates

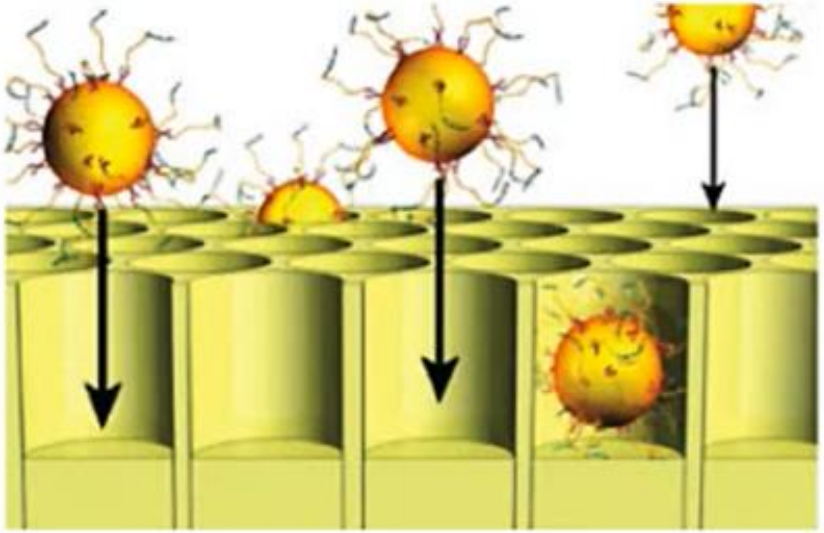
Final product
100–200 million beads with thousands of bound template

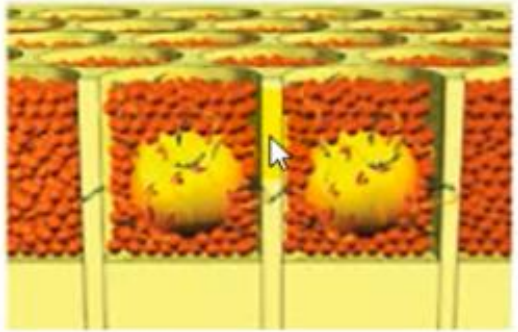
Sequencing

c)

Sequencing

7.5 hours

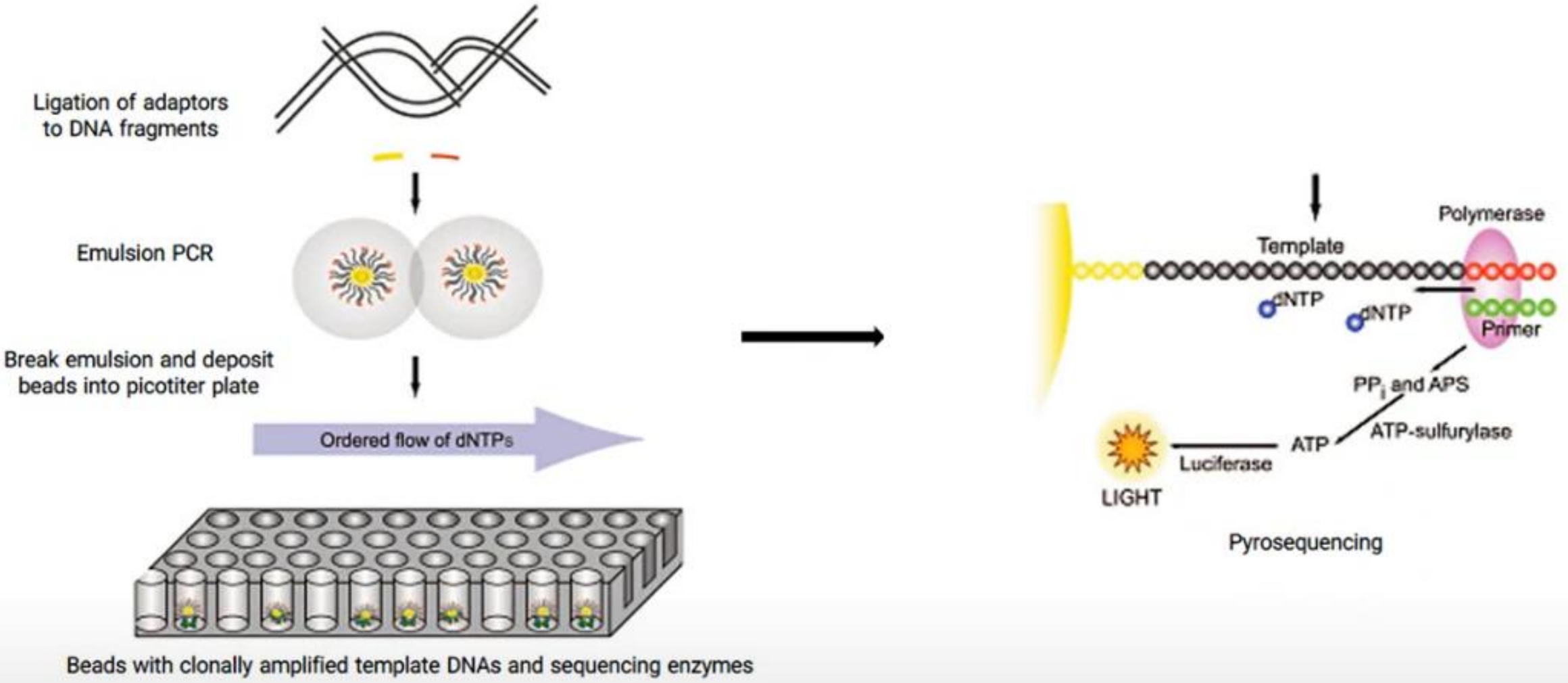




- Well diameter: average of 44 µm
- 400,000 reads obtained in parallel
- A single cloned amplified sstDNA bead is deposited per well

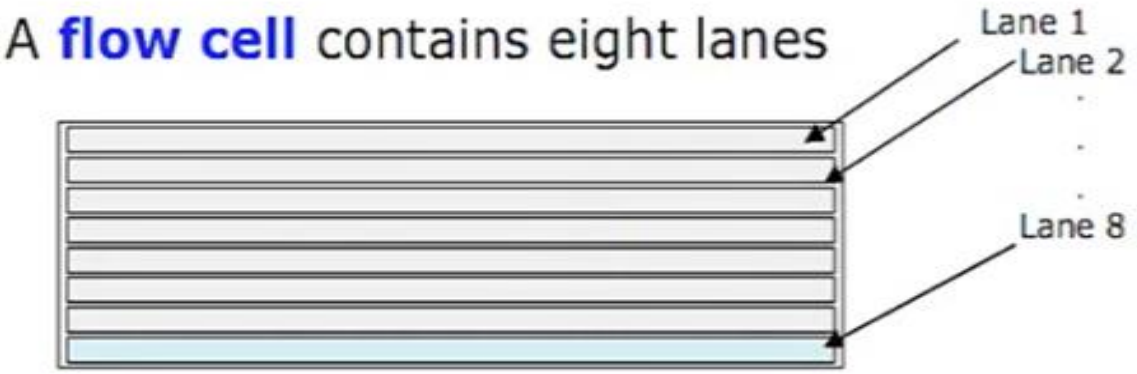
Amplified sstDNA library beads → Quality filtered bases

454 life science (FLX)



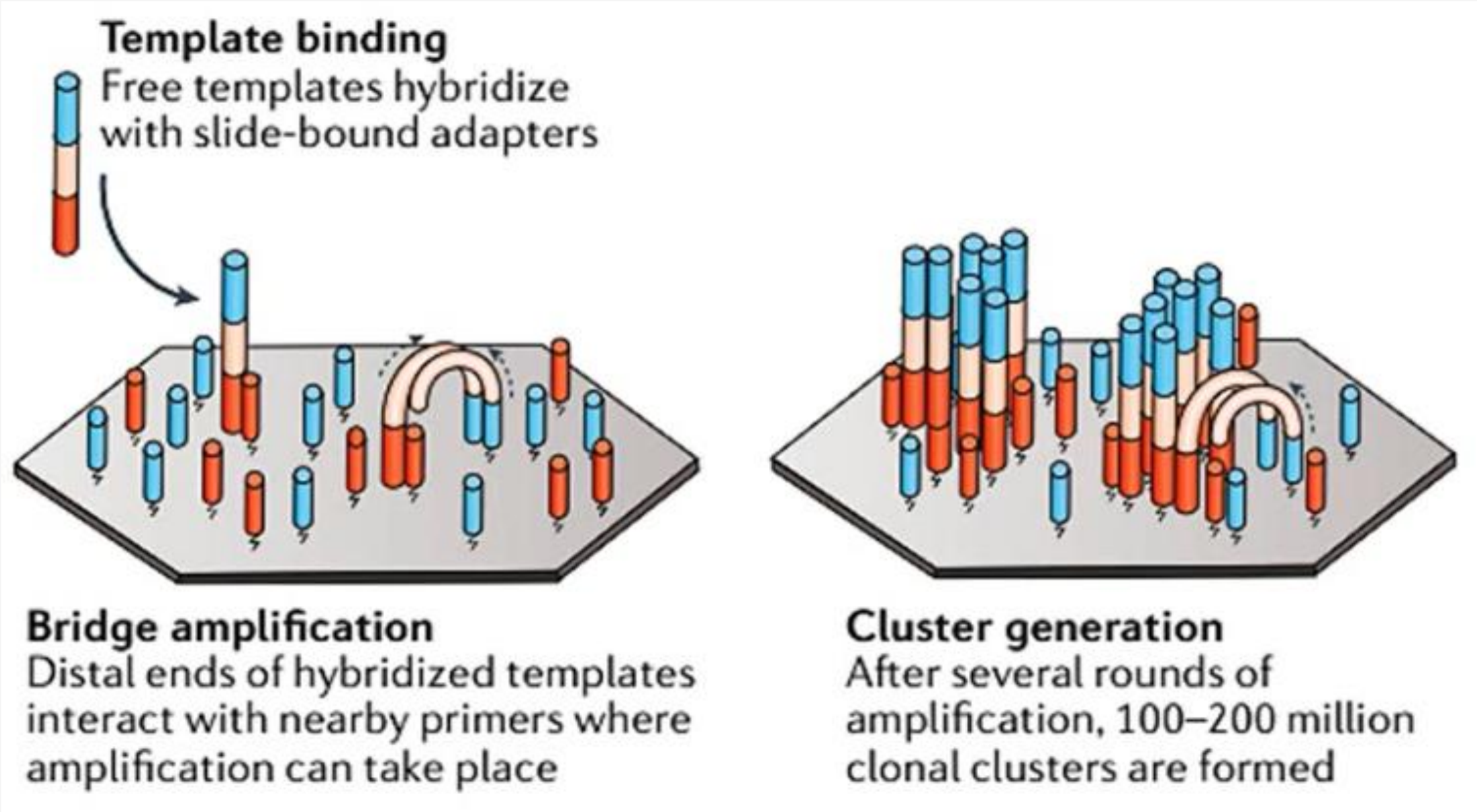
Solexa (Illumina)

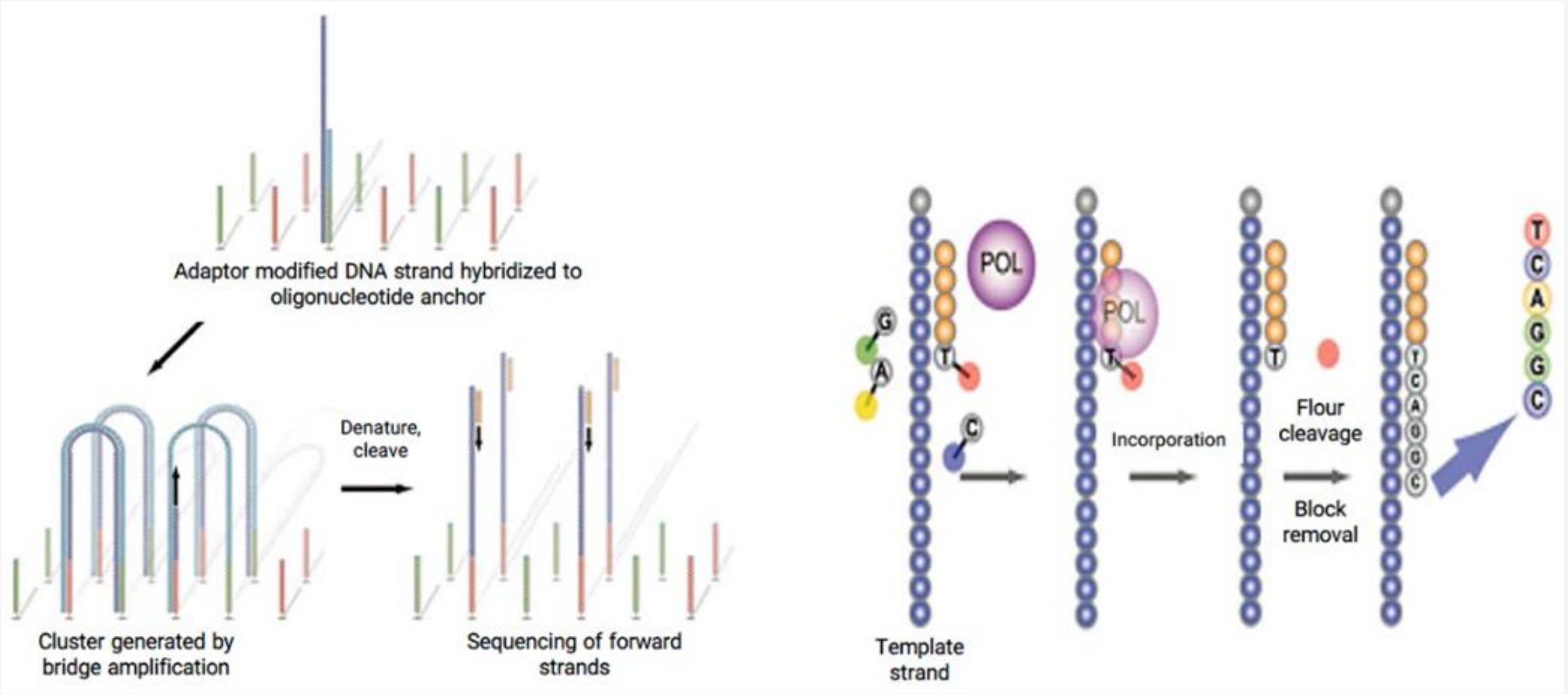
2006 •

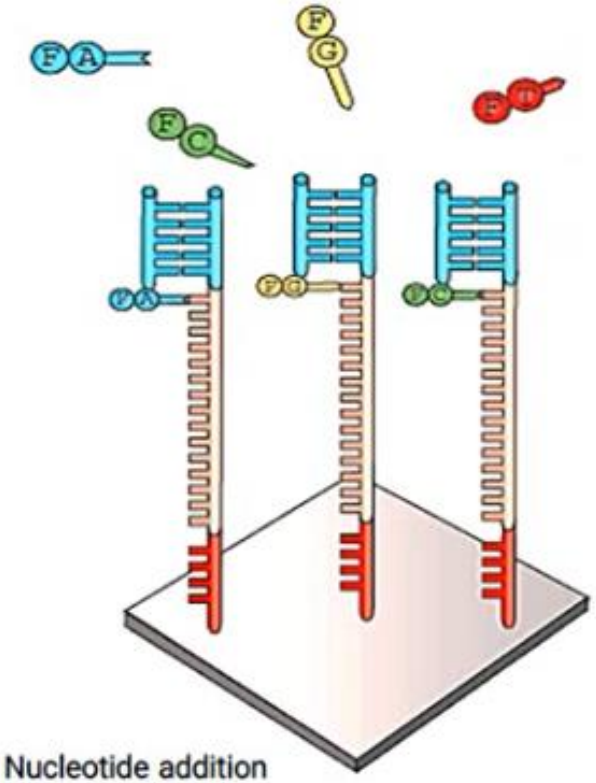


- Template preparation
- Imaging
- Sequencing

- Solid Phase amplification
- Total internal reflection fluorescence (TIRF)
- Cycle reversible termination (CRT)

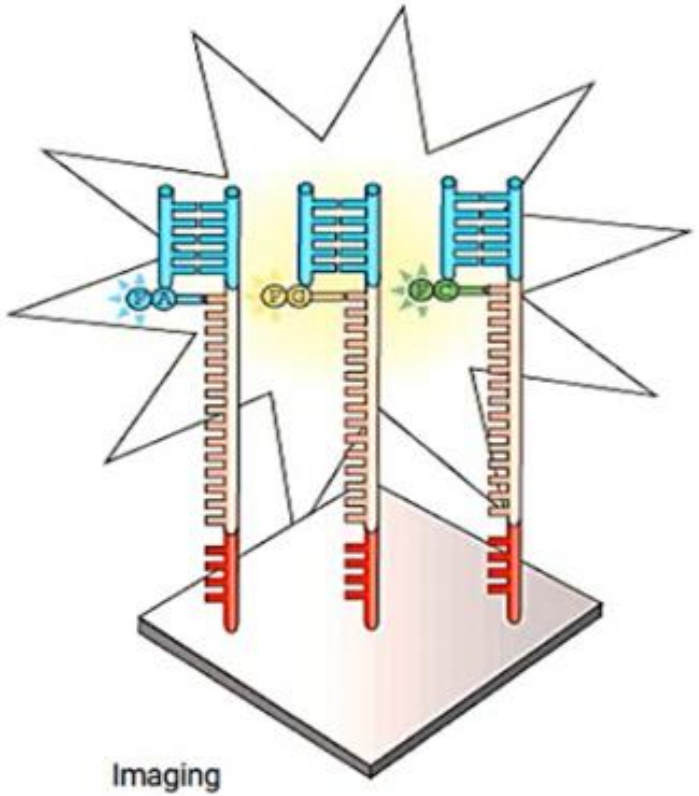

Template binding
Free templates hybridize
with slide-bound adapters
Bridge amplification
Distal ends of hybridized templates
interact with nearby primers where
amplification can take place
Cluster generation
After several rounds of
amplification, 100–200 million
clonal clusters are formed


Adaptor modified DNA strand hybridized to oligonucleotide anchor
Denature, cleave
Cluster generated by bridge amplification
Sequencing of forward strands
POL
G
A
T
C
Template strand
Incorporation
Flour cleavage
Block removal
T
C
A
G
G
C



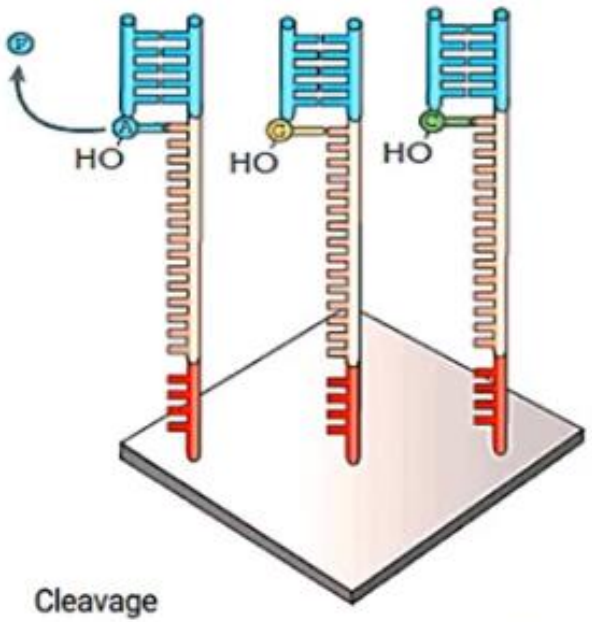
Nucleotide addition

Fluorophore-labelled, terminally blocked nucleotides hybridize to complementary base. Each cluster on a slide can incorporate a different base.



Imaging

Slides are imaged with either two or four laser channels. Each cluster emits a colour corresponding to the base incorporated during this cycle.



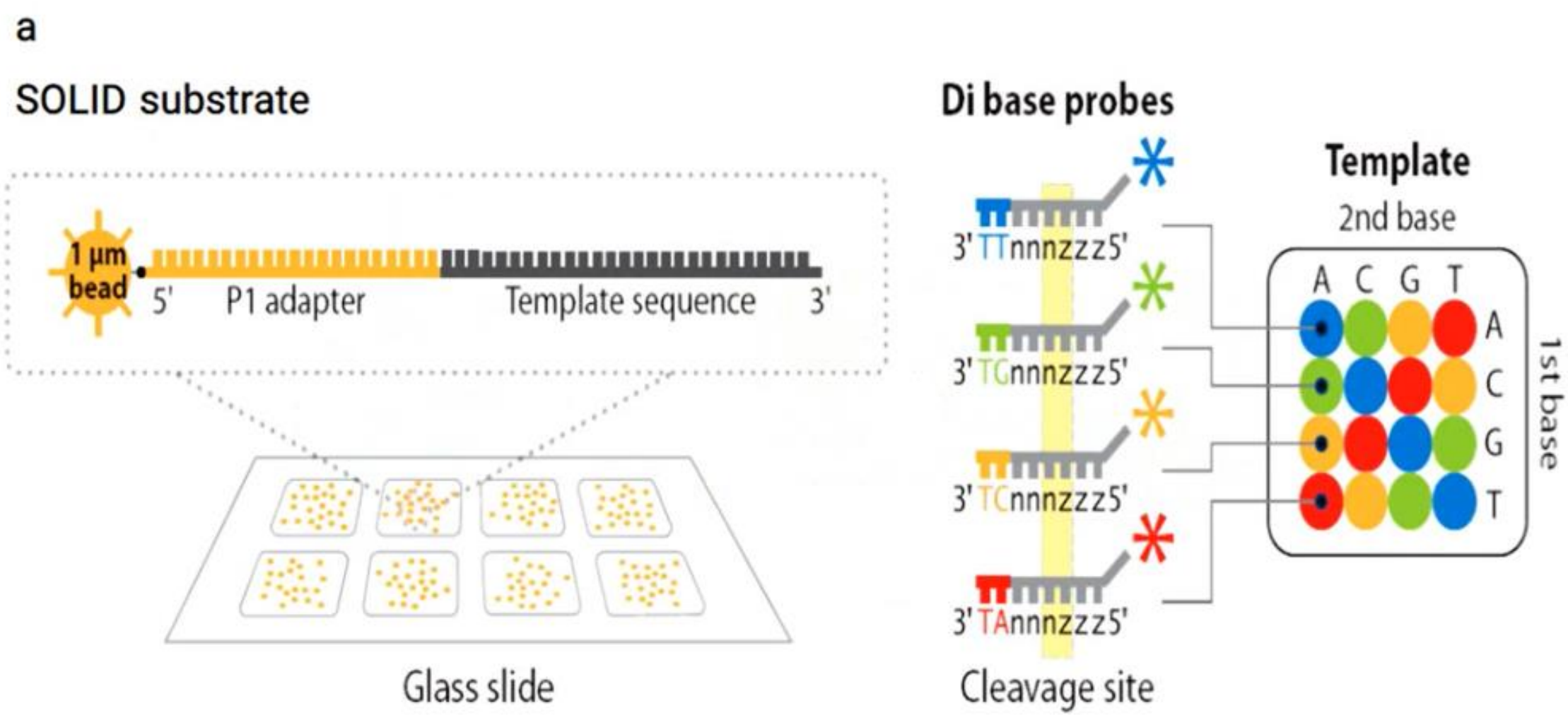
Cleavage

Fluorophores are cleaved and washed from flow cells and the 3′-OH group is regenerated. A new cycle begins with the addition of new nucleotides.

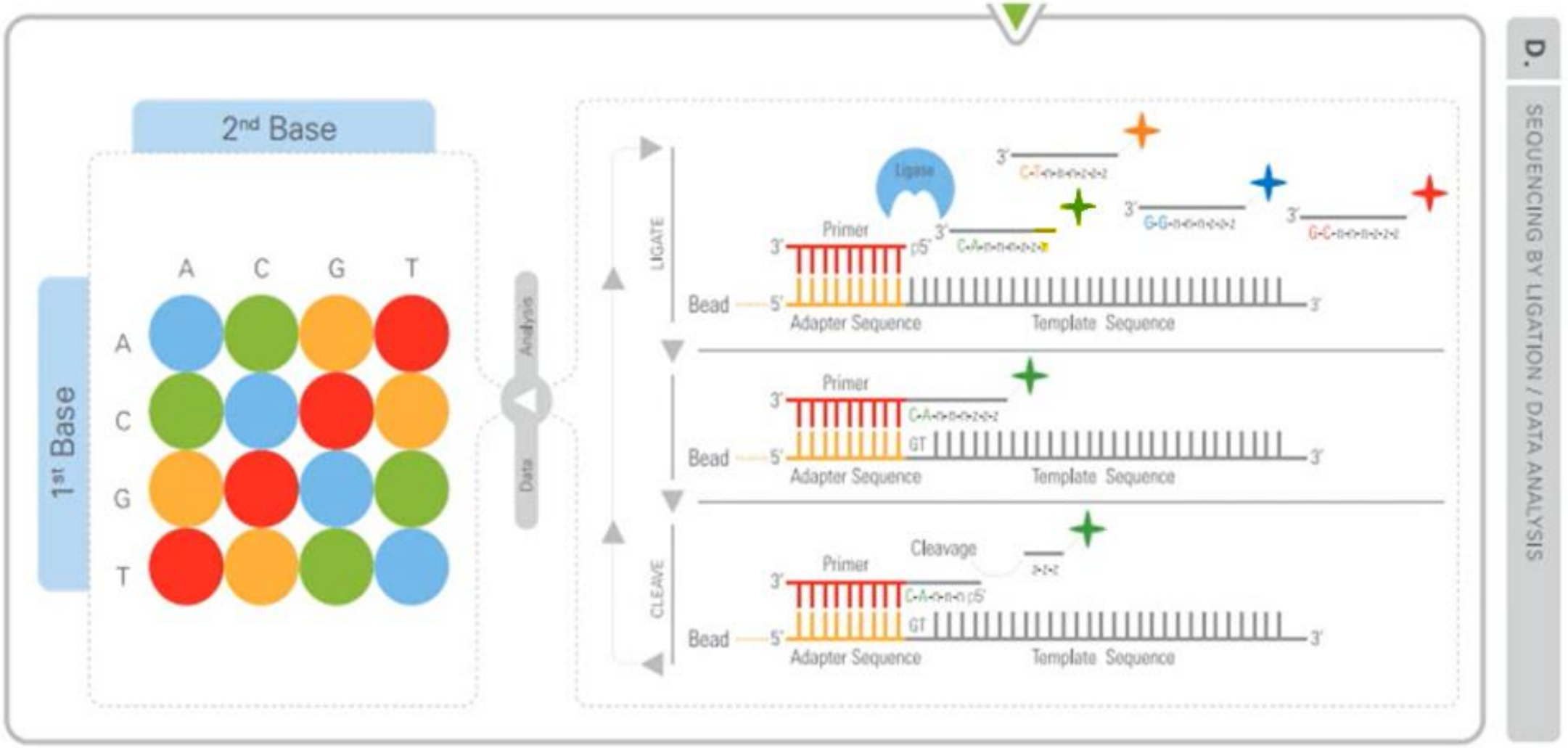
Solexa (Illumina)

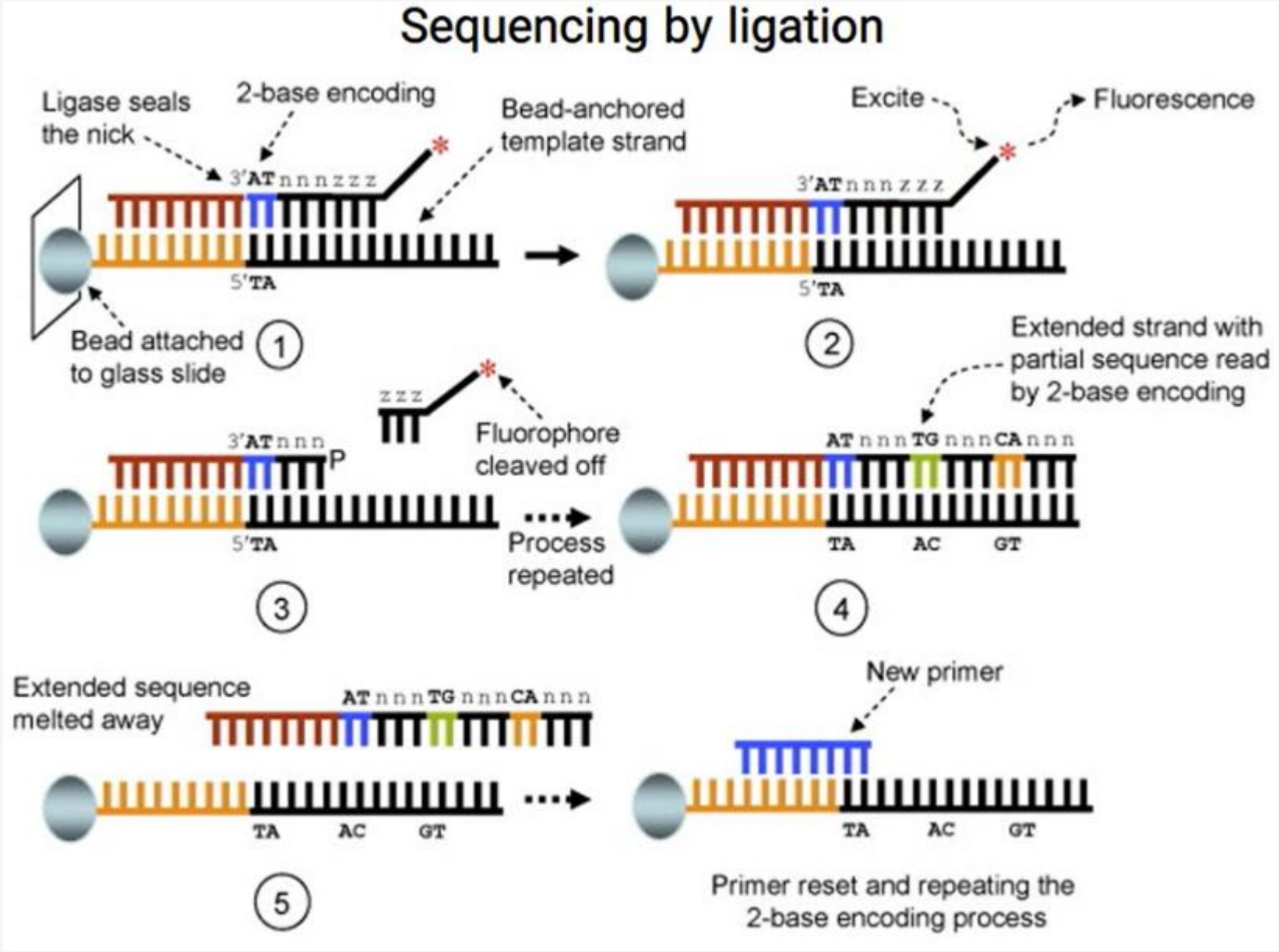
	MiSeq	NextSeq	HiSeq 2500	HiSeq X Ten
Run Time	56 hours	29 hours	3.6 days	3 days
Read length	2× 300 bp	2× 150 bp	2× 125 bp	2× 150 *bp*
Read number	25 Million	400 Million	4 Billion	6 Billion
output	15 GB	120 GB	1000 GB	1800 GB
cost	$99K	$250K	$740K	$10M

ABI Solid Sequencing



Sequencing by ligation




Sequencing by ligation
Ligase seals the nick
2-base encoding
Bead-anchored template strand
3'ATnnnzzz
5'TA
Bead attached to glass slide
1
Excite
Fluorescence
2
Extended strand with partial sequence read by 2-base encoding
zzz
Fluorophore cleaved off
3'ATnnn
P
ATnnnTGnnnCAnnn
TA AC GT
Process repeated
3
4
Extended sequence melted away
New primer
5
Primer reset and repeating the 2-base encoding process

ABI Solid Sequencing

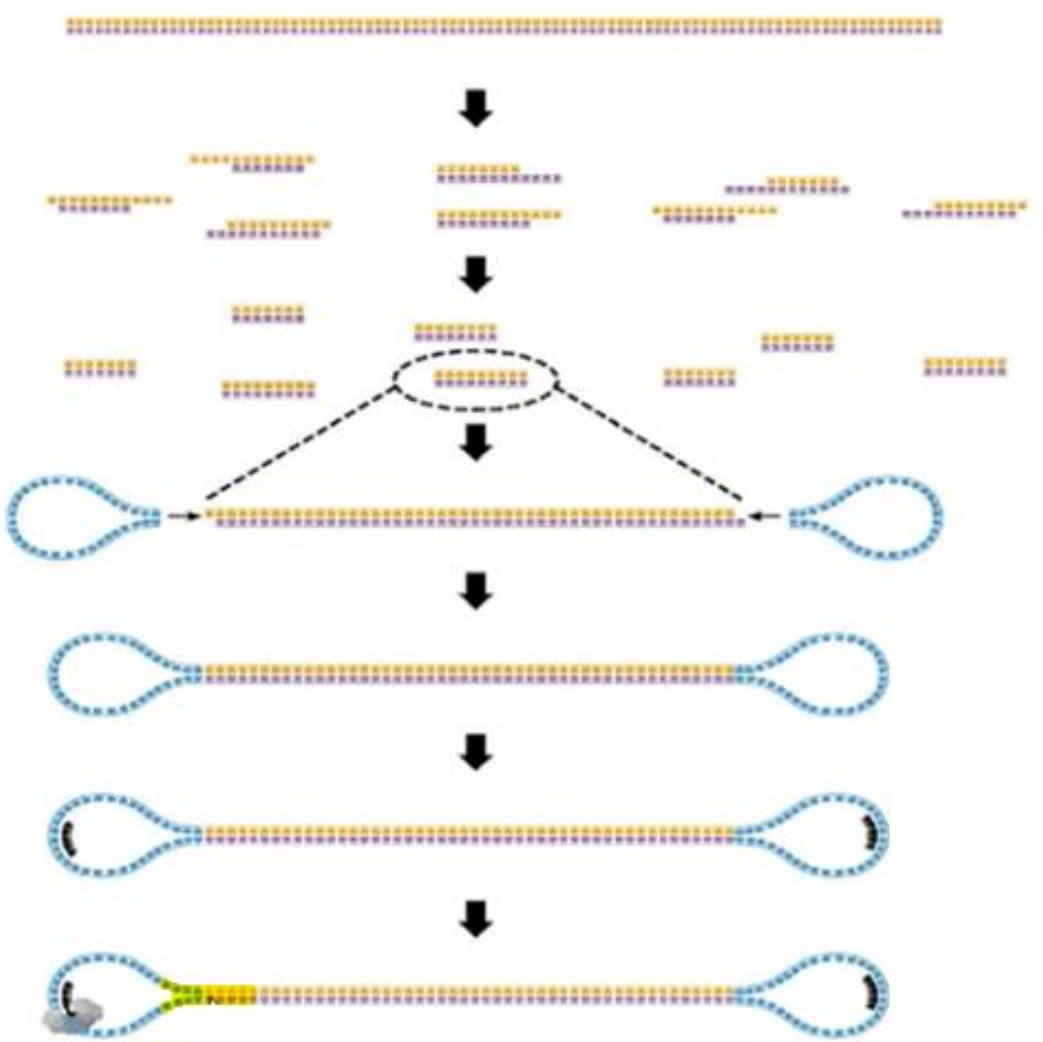
تکنیک های توالی یابی نسل سوم (PacBio, Ion Torrent, Oxford Nanopore)

Pacific bioscience

• Template Preparation	• Single Molecule Template (SMT)
• Imaging	• CCD camera
• Sequencing	• Single Molecule Real Time (SMRT)

Template Preparation

- DNA sample
- Fragment DNA
- Repair ends
- Ligate Adaptors
- Purify DNA
- Primer annealing
- Polymerase binding



Principal of Pacific Bioscience

- A single DNA polymerase is immobilized on the bottom of a reaction cell.
- Reaction cell called a ZMW (Zero-mode waveguide).
- Φ29 DNA polymerase is used.

PACIFIC BIOSCIENCES®

Chemistry of Pacific Bioscience

- استفاده از یک dNTP Phospolinked.
- هر dNTP شامل fluoroporeهای متفاوت.
- طی فرایند سکانس: ساطع‌شدن سیگنال فلوروسانت در زمان بسیار کوتاهی از ZMW.
- ورود dNTP جدید و خروج dNTP قبلی.

Steps of Pacific Bioscience

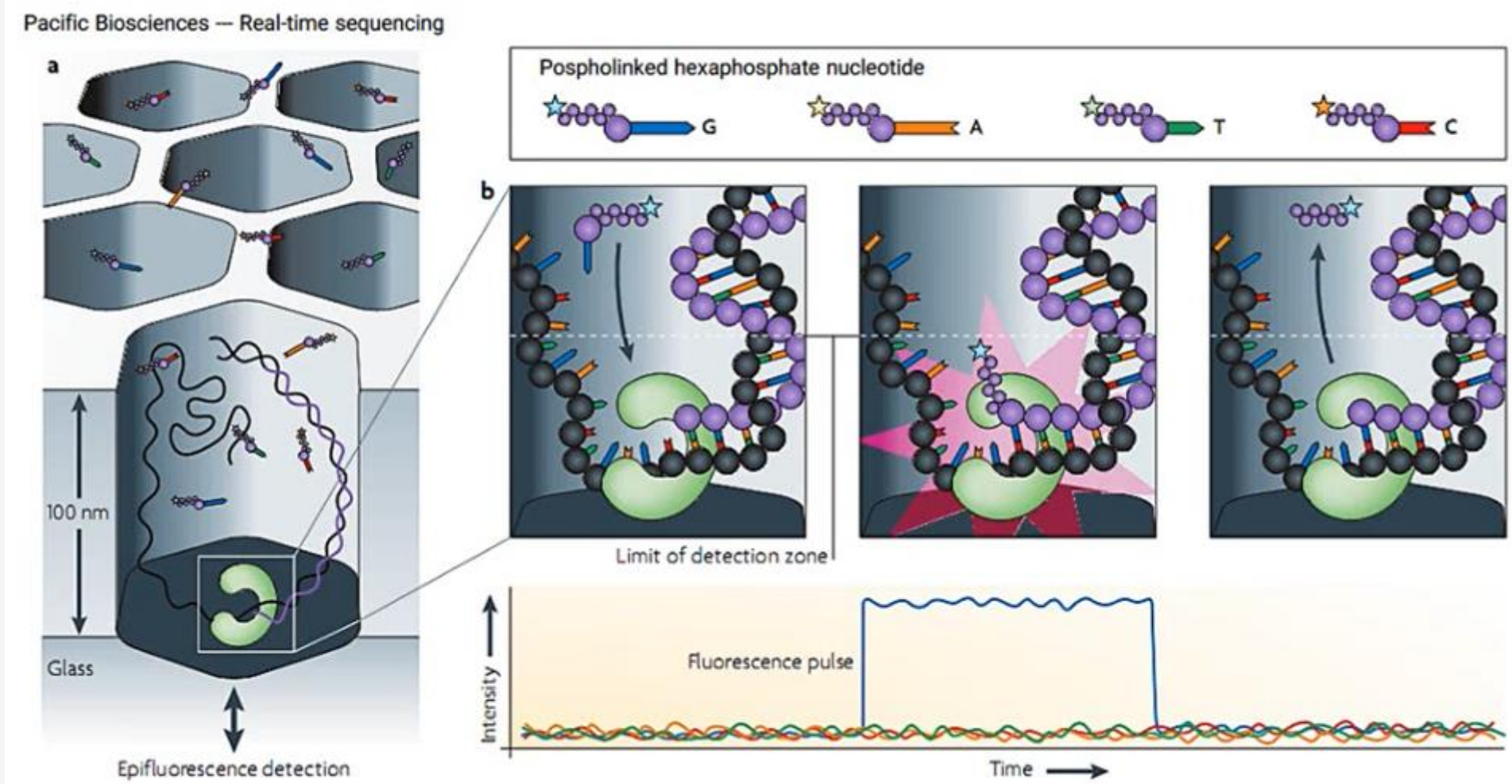
۱. اتصال نوکلئوتیدهای نشان‌دارشده به ZMW.

۲. تولید ده هزارم ثانیه نوری توسط بازی که در واکنش شرکت می‌کند.

۳. شکافته‌شدن زنجیره فسفات و رهاشدن مولکول رنگ‌شده.

۴. تکرار این پروسه.

ZMW (Zero Mode Waveguide)

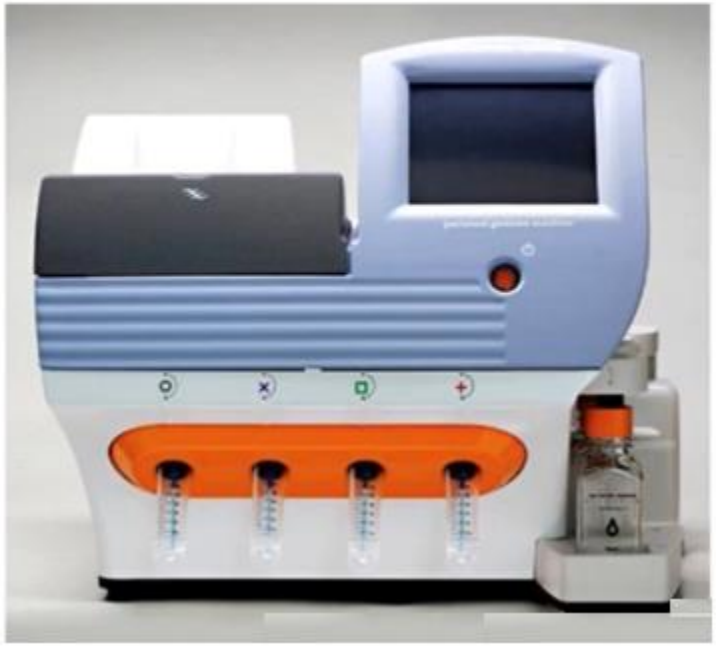


Pacific Bioscience

Ion Torrent

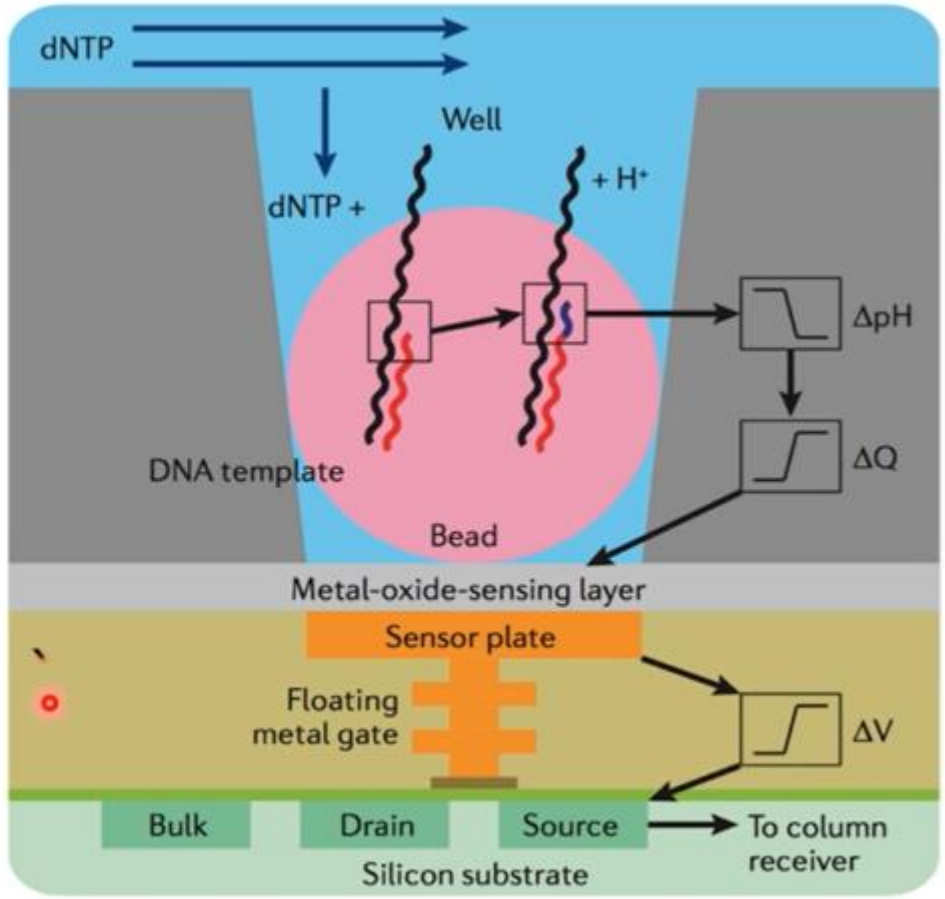
- Around 60-80 M reads.
- 200 bp length.
- Sequences based on H+ production
- Error rates lower than other 2nd generation
- Error pattern similar to 454, with homopolymer problem.
- Very cheap per run.
- Belongs to Life technologies (Applied Biosystems)

Method:
Proton detection



Ion Torrent Sequencing platform

- Similar to pyrosequencing but using semiconducting chip to detect dNTP incorporation.
- The chip measures differences in PH.
- shown to have coverage bias with GC-rich regions.
- Ion proton promises higher output and longer reads.



A simplified schematic diagram of an ion chip sensor well. Image is reproduced from Rothberg *et al.*

Advantages:
Low startup costs
Medium/low cost per base
Low error rate
Fast runs

Disadvantages:
Costs higher than e.g. Illumina
Read length between Illumina and PacBio
Higher error rate than Illumina

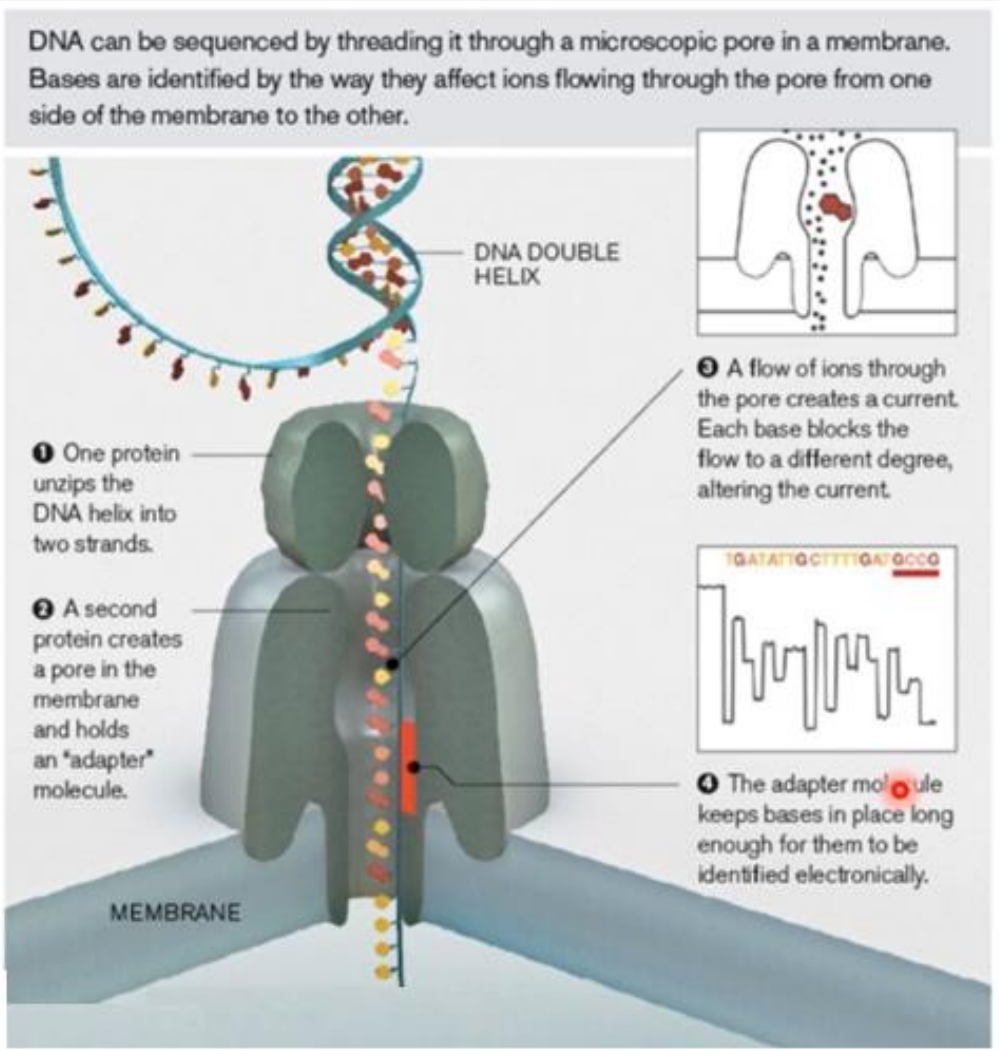
Oxford Nanopore

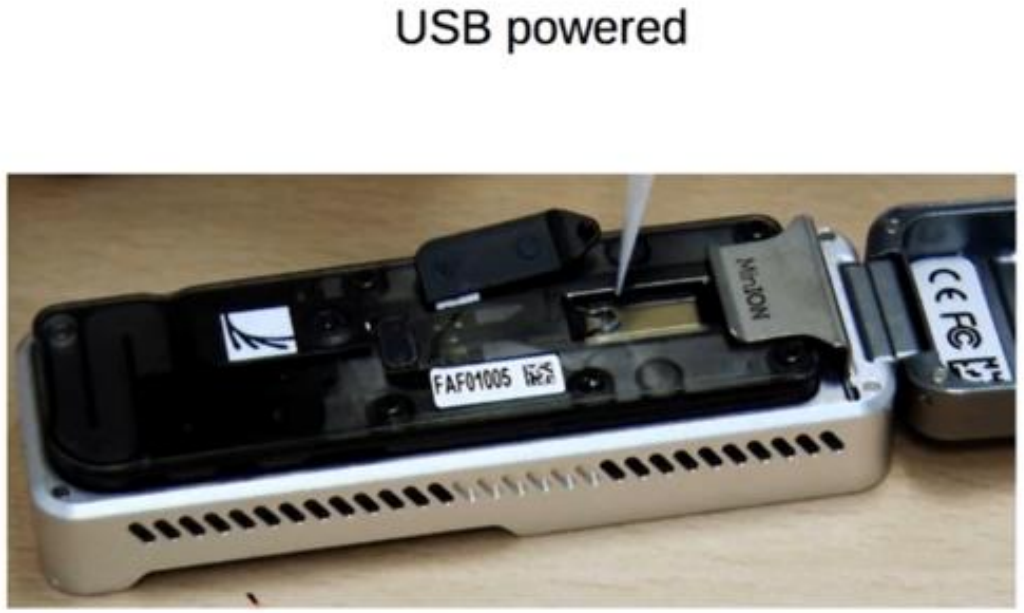
Nanopore sequencing is a third generation approach used in the sequencing of biopolymers — specifically, polynucleotides in the form of DNA or RNA.

Using nanopore sequencing, a single molecule of DNA or RNA can be sequenced without the need for PCR amplification or chemical labeling of the sample. Nanopore sequencing has the potential to offer relatively low-cost genotyping, high mobility for testing, and rapid processing of samples with the ability to display results in real-time.

Publications on the method outline its use in rapid identification of viral pathogens, monitoring ebola, environmental monitoring, food safety monitoring, human genome sequencing, plant genome sequencing, monitoring of antibiotic resistance, haplotyping and other applications.

Oxford Nanopore





Advantages:

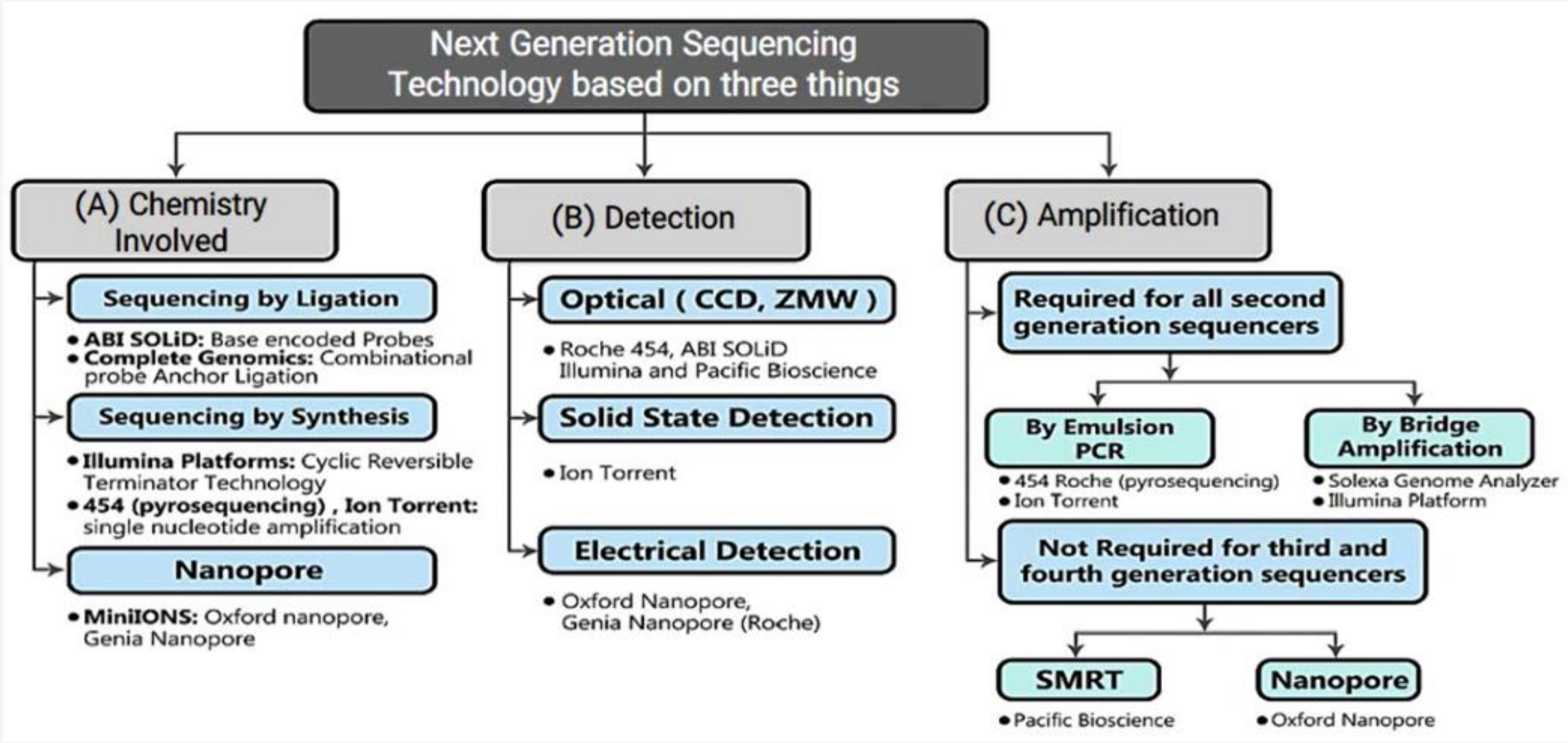
Minion is a USB devise

Extremely low cost

Extremely long read feasible and short sequencing time

Disadvantages:

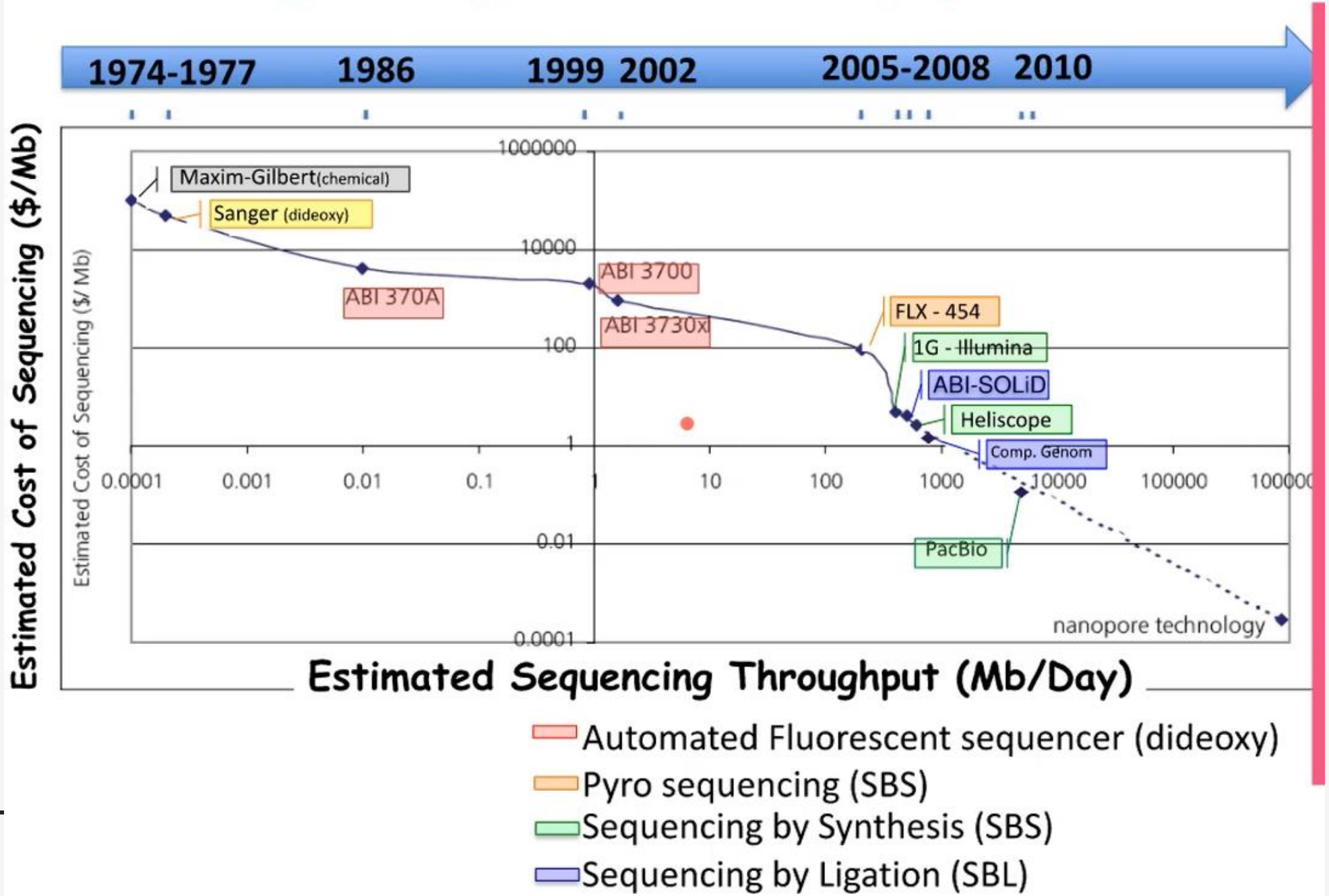
Unknown error rate


Next Generation Sequencing Technology based on three things
(A) Chemistry Involved
Sequencing by Ligation
• ABI SOLiD: Base encoded Probes
• Complete Genomics: Combinational probe Anchor Ligation
Sequencing by Synthesis
• Illumina Platforms: Cyclic Reversible Terminator Technology
• 454 (pyrosequencing) , Ion Torrent: single nucleotide amplification
Nanopore
• MiniIONS: Oxford nanopore, Genia Nanopore
(B) Detection
Optical (CCD, ZMW)
• Roche 454, ABI SOLiD Illumina and Pacific Bioscience
Solid State Detection
• Ion Torrent
Electrical Detection
• Oxford Nanopore, Genia Nanopore (Roche)
(C) Amplification
Required for all second generation sequencers
By Emulsion PCR
• 454 Roche (pyrosequencing)
• Ion Torrent
By Bridge Amplification
• Solexa Genome Analyzer
• Illumina Platform
Not Required for third and fourth generation sequencers
SMRT
• Pacific Bioscience
Nanopore
• Oxford Nanopore

Comparison of different NGS platforms

	Throughput	Length	Quality	Costs
Sanger	6 Mb/day	800nt	10^{-4} - 10^{-5}	500$/Mb
454	750Mb/day	400nt	10^{-3} - 10^{-4}	~20$/Mb
Ion Torrent	1600Mb/day	200nt	10^{-2} - 10^{-3}	~10$/Mb
Illumina	100000Mb/day	125nt	10^{-2} - 10^{-3}	~0.40$/Mb
SOLiD 4	100000Mb/day	125nt	10^{-2} - 10^{-3}	~0.40$/Mb
Helicos	5000Mb/day	32nt	10^{-2}	~0.40$/Mb

DNA Sequencing Cost vs Throughput Timeline



Omics

واژه Omics اشاره به یک رشته مطالعاتی در علوم بیولوژیک دارد که با omics– پایان می یابد.

Genomics and epigenomics

The study of the DNA sequence and associated heritable biochemical modifications

Transcriptomics

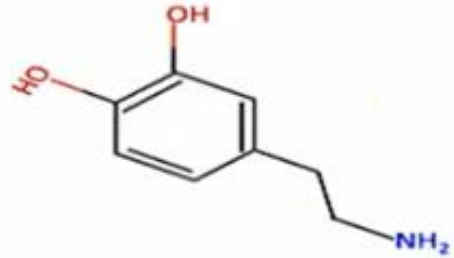
The study of the RNA molecules present in a sample

Phenotype

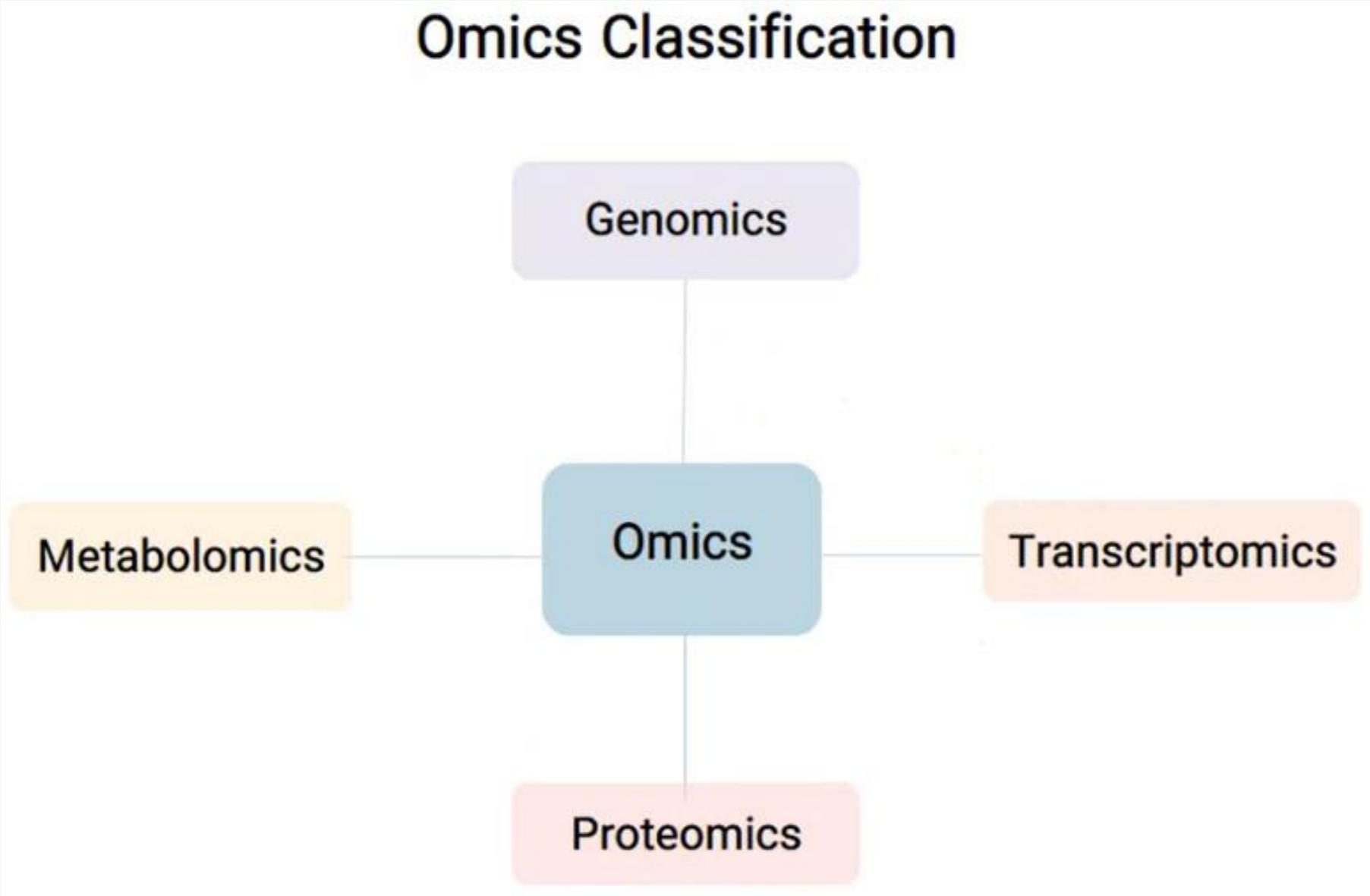
Proteomics

The study of the proteins present in a sample.

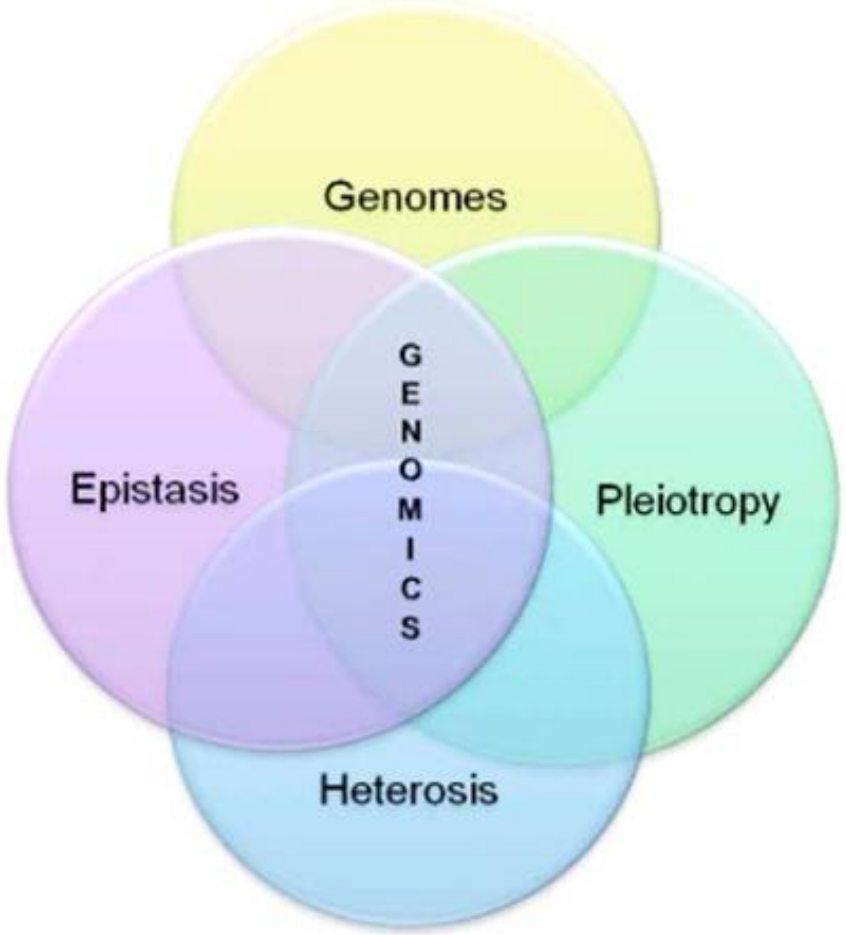
Metabolomics



The study of the metabolites present in a sample.


Omics Classification
Genomics
Metabolomics
Omics
Transcriptomics
Proteomics

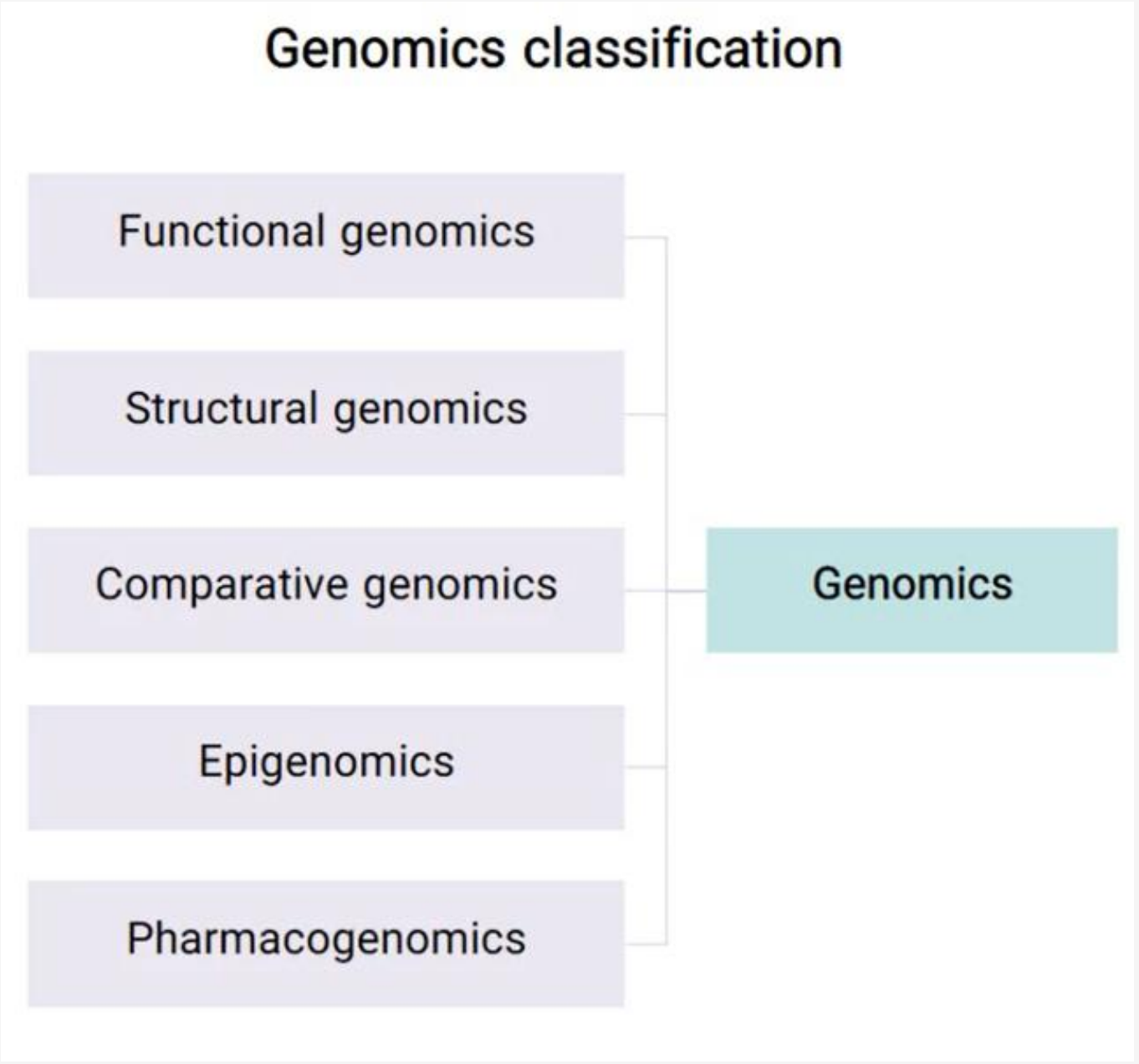
Genomics



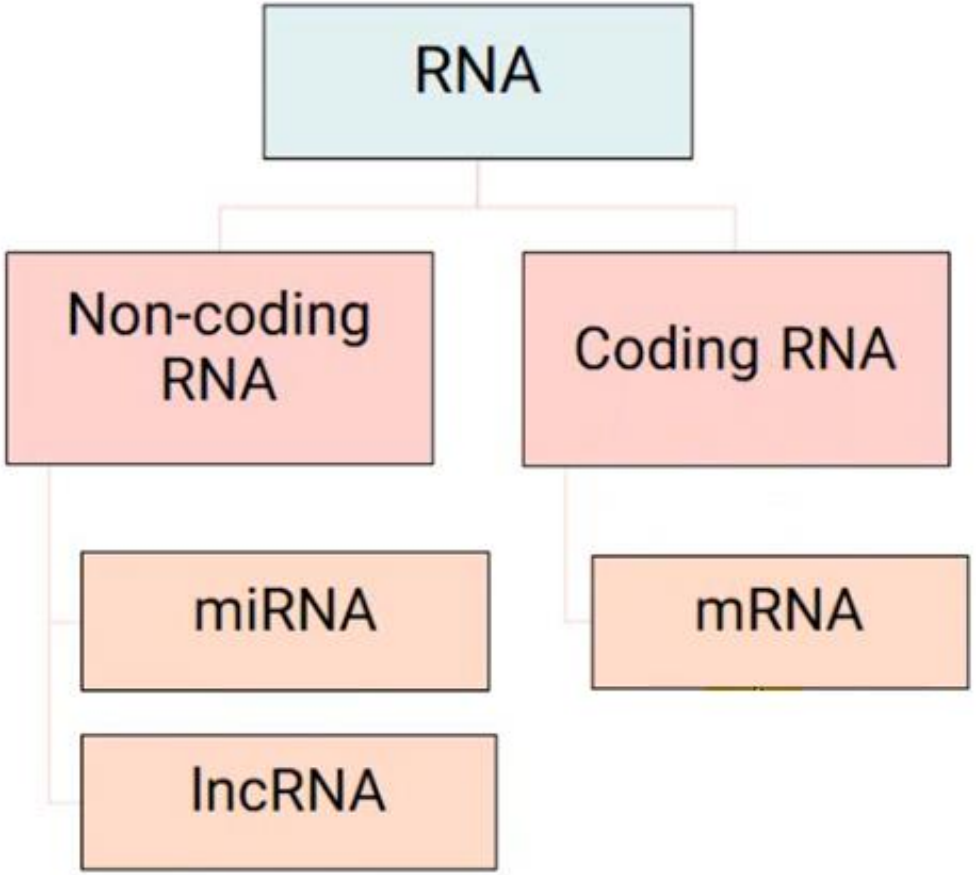
- مطالعه کل ژنوم موجودات
- تلفیقی از بیوانفورماتیک و تکنیک‌های توالی‌یابی نسل دوم DNA

آنالیز ساختار و عملکرد DNA

- متمرکز بر تعاملات بین جایگاه‌ها و آلل‌های ژنوم و سایر فعل‌وانفعالات مانند اپیستازی، پلیوتروپی و هتروزیس


Genomics classification
Functional genomics
Structural genomics
Comparative genomics
Epigenomics
Pharmacogenomics
Genomics

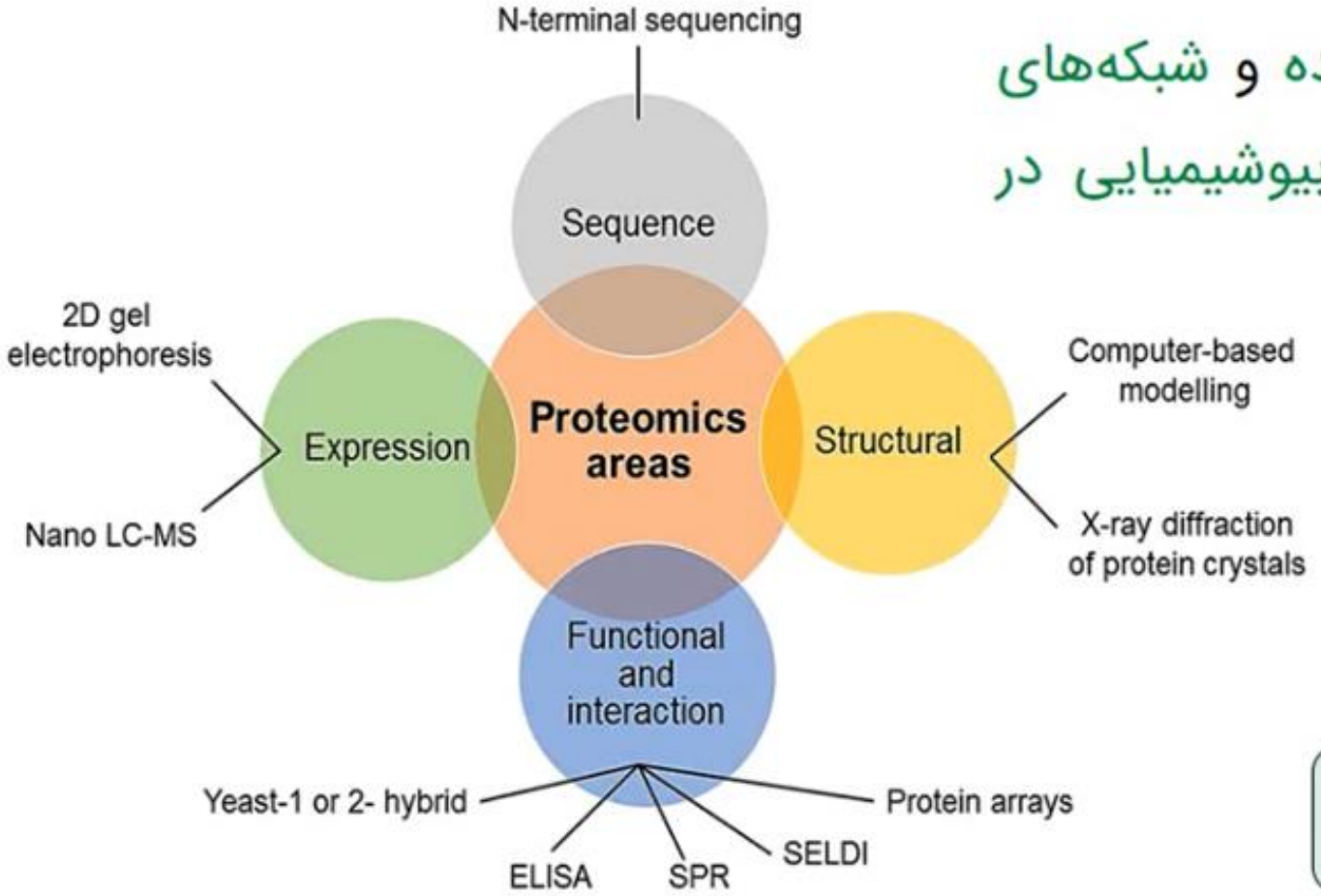
Transcriptomics



- مطالعه ترنسکریپتوم
- مقایسه رونویسی‌ها ⟸ شناسایی ژن‌هایی که به طور متفاوت در جمعیت سلول‌های متمایز یا در پاسخ به درمان‌های مختلف بیان می‌شوند.
- توسط آنالیزهای دقیق میکرواری
- ۲۵۰۰ رونوشت
- توالی RNA نسل بعدی فناوری‌ها ⟸ فراهم‌کردن درک عمیق از تغییرات و بیان ژن‌ها

Proteomics

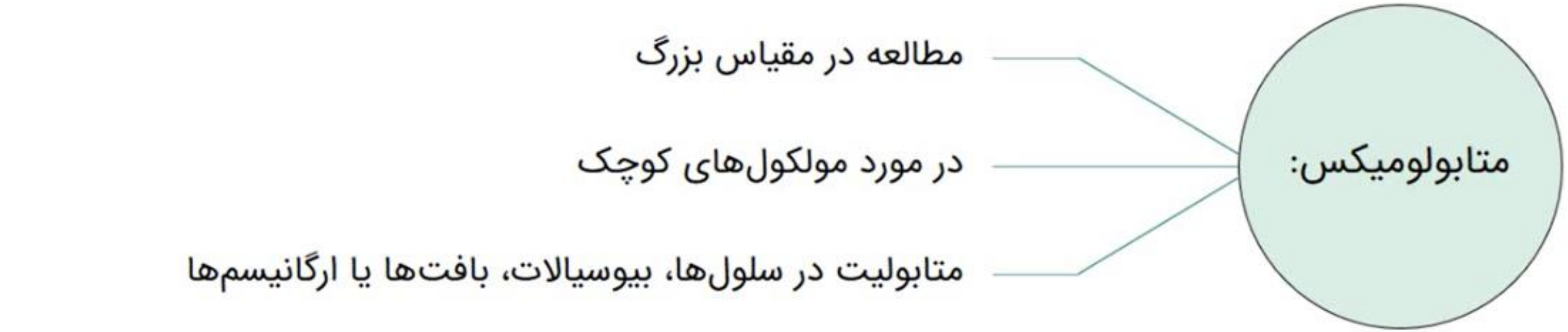
- مطالعه پروتئوم
- مجموعه‌ای از تمام پروتئین‌های بیان‌شده و شبکه‌های خانواده پروتئین متقابل و مسیرهای بیوشیمیایی در سلول، بافت یا ارگانیسم



هنوز تعداد دقیق پروتئین‌ها / پپتیدها مشخص نیست

تخمین: حدود چند صد هزار

Metabolomics

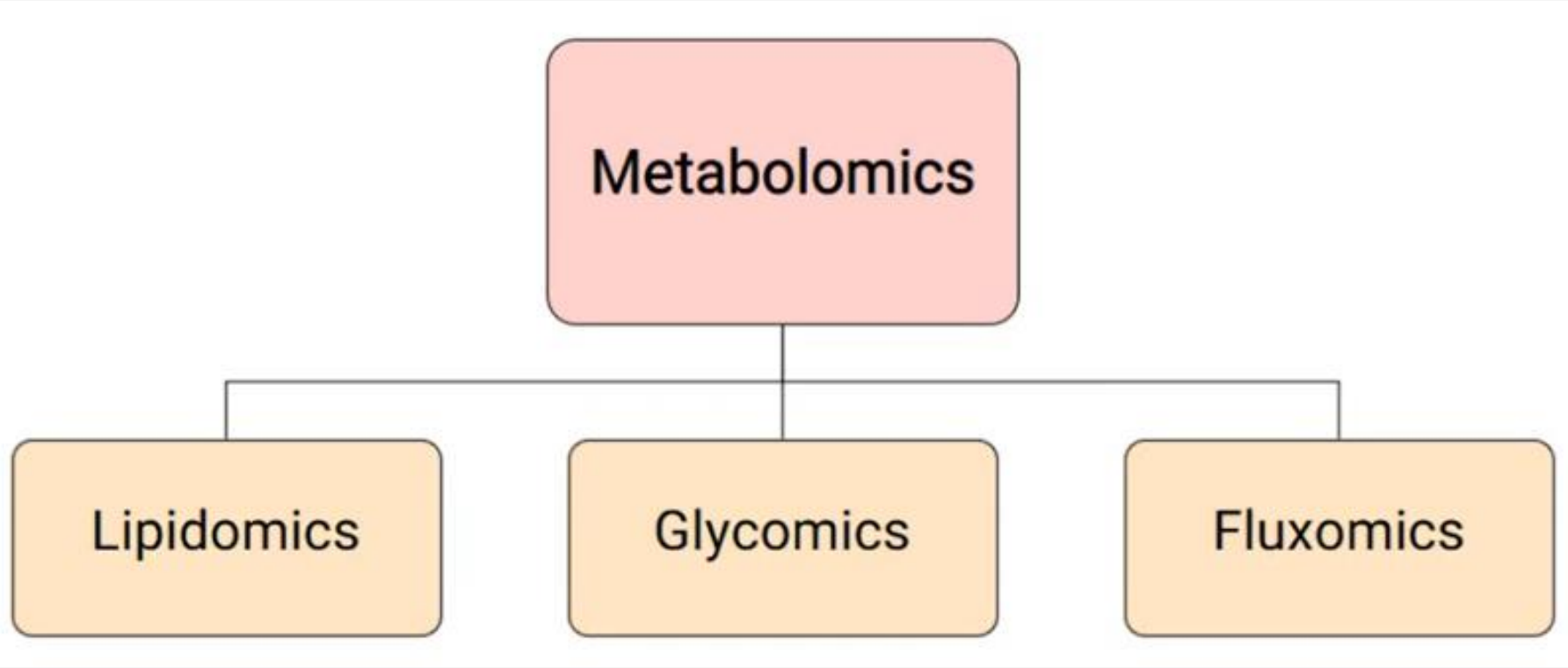


مطالعه هم‌زمان متابولیت‌ها

دستگاه‌های مدرن و جدید طیف‌سنجی جرمی یا طیف‌سنجی تشدید مغناطیس هسته

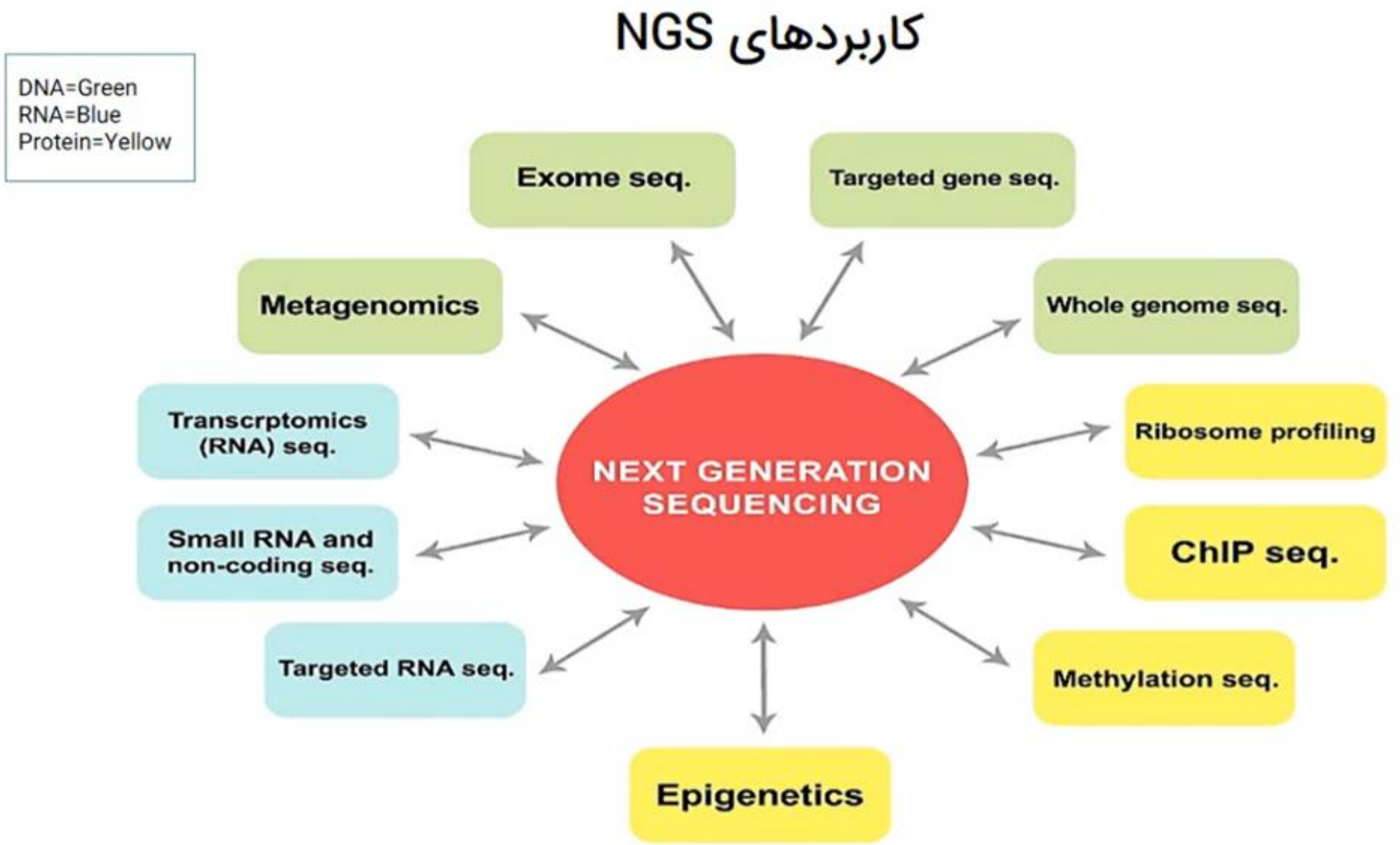
دستگاه‌های مدرن آنالیز LC/Ms/Ms و NMR با قدرت تفکیک‌پذیری بالا (۵۰۰–۹۰۰ مگاهرتز)

Metabolomics classification



کاربردهای توالی یابی های نسل دوم و سوم در مطالعات مختلف Omics

کاربردهای NGS



توالی‌یابی کل ژنوم (Whole Genome Sequencing)

DNA هسته

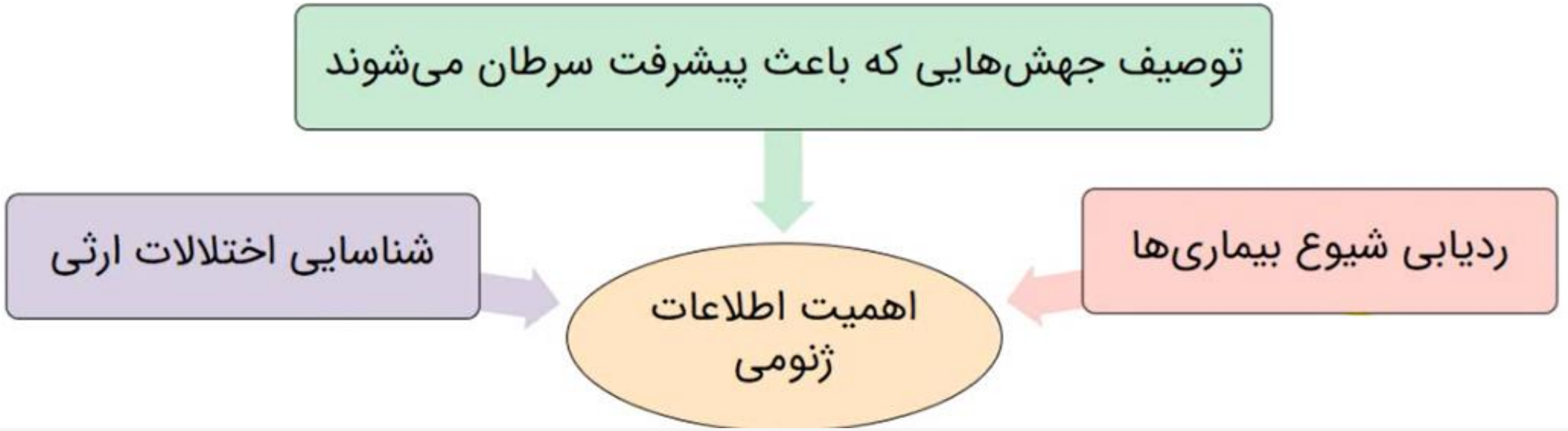
میتوکندری

- مشخص‌کردن توالی کامل DNA انسان (نواحی کدکننده و غیر کدکننده)

- امکان‌پذیری شناسایی کلیه جهش‌های نقطه‌ای، حذف و مضاعف‌شدگی‌ها و بازآرایی‌های کروموزومی

توالی‌یابی کل ژنوم (Whole Genome Sequencing)

تعیین توالی ژنوم:
روش جامع برای تجزیه‌وتحلیل کل ژنوم‌ها



روش توالی‌یابی هدف‌دار Targeted Gene Sequencing

- استفاده برای تشخیص وجود جهش‌ها یا غربالگری حامل.
- تعیین افزایش یا کاهش عملکرد یک ژن منجر به بیماری‌های مختلف؛ مانند:
  - آلفا تالاسمی
  - فیبروز کیستیک
- غربال ژنوم برای SNPها.

توالی‌یابی اگزوم Exome Sequencing

توالی‌یابی نواحی کدکننده ژنوم انسان

شناسایی جهش‌های نقطه‌ای و حذف و مضاعف‌شدگی‌های کوچک

استفاده گسترده

شامل تعیین توالی مناطق کدکننده پروتئین ژنوم

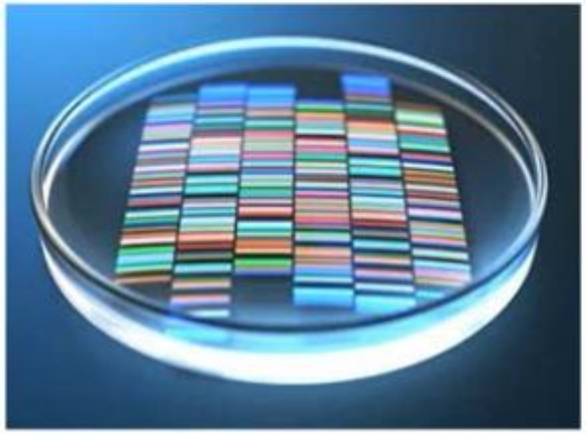
نمایش اگزوم انسان کمتر از ۲٪ ژنوم

شامل ۸۵٪ از انواع شناخته‌شده مربوط به بیماری

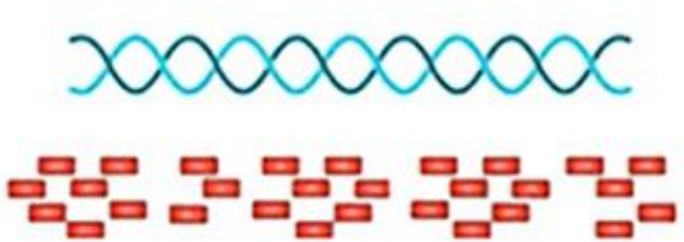
مقرون‌به‌صرفه

مزایای تعیین توالی Exome

- فراهم‌کردن یک جایگزین مقرون‌به‌صرفه برای توالی‌یابی کل ژنوم.
- برای تجزیه‌وتحلیل داده‌ها ⟵ سریع‌تر و آسان‌تر
- در مقایسه با رویکردهای کل ژنوم ⟵ تولید یک مجموعه داده کوچکتر و قابل کنترل‌تر

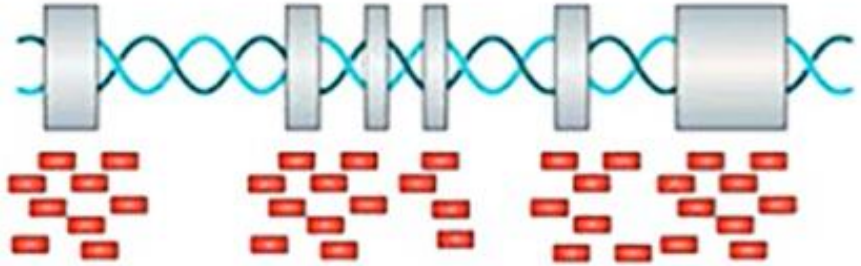


Whole genome sequencing



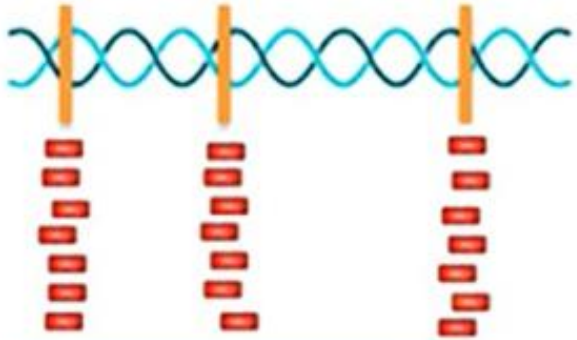
- Sequencing region : whole genome
- Sequencing Depth: >30X
- Covers everything – can identify all kinds of variants including SNPs, INDELs and SV.

Whole exome sequencing



- Sequencing region: whole exome
- Sequencing Depth : >50X ~ 100X
- Identify all kinds of variants including SNPs, INDELs and SV in coding region.
- Cost effective

Targeted sequencing



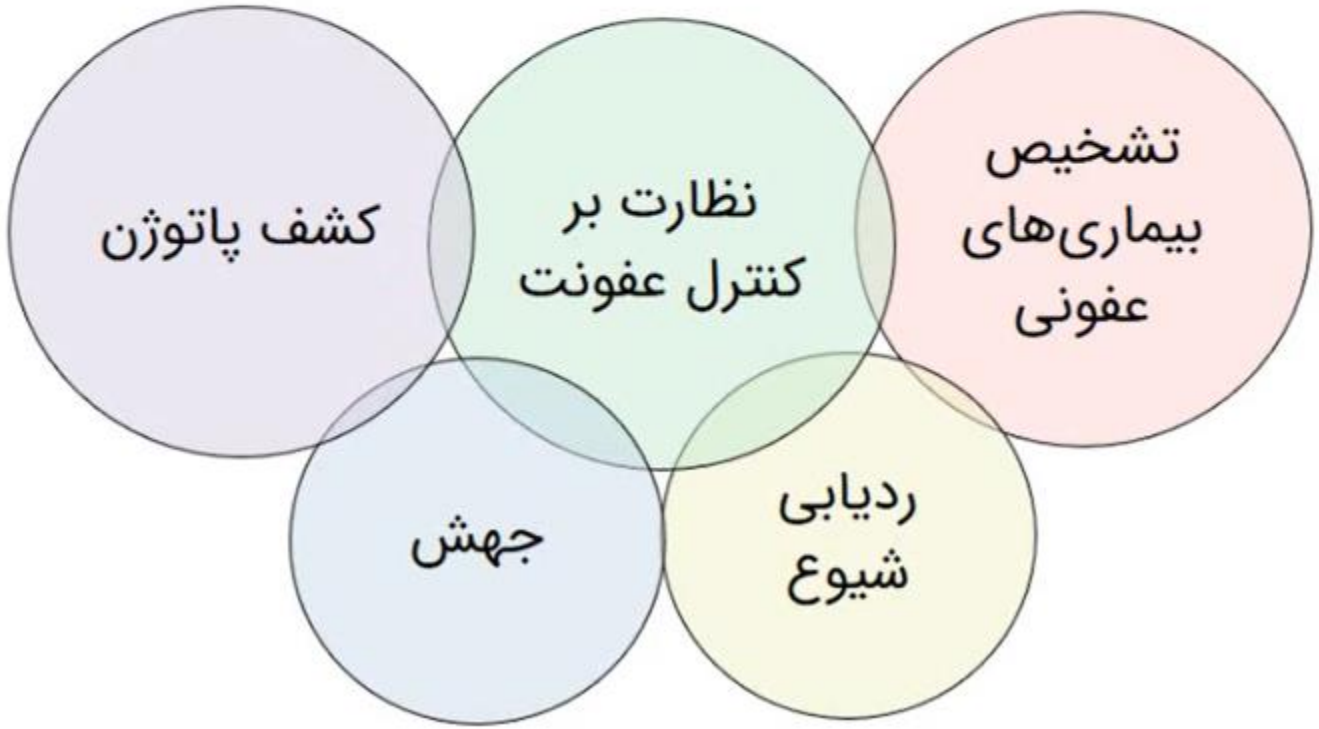
- Sequencing region: specific regions (could be customized)
- Sequencing Depth : >500X
- Identify all kinds of variants including SNPs, INDELs in specific regions
- Most Cost effective

Metagenomics

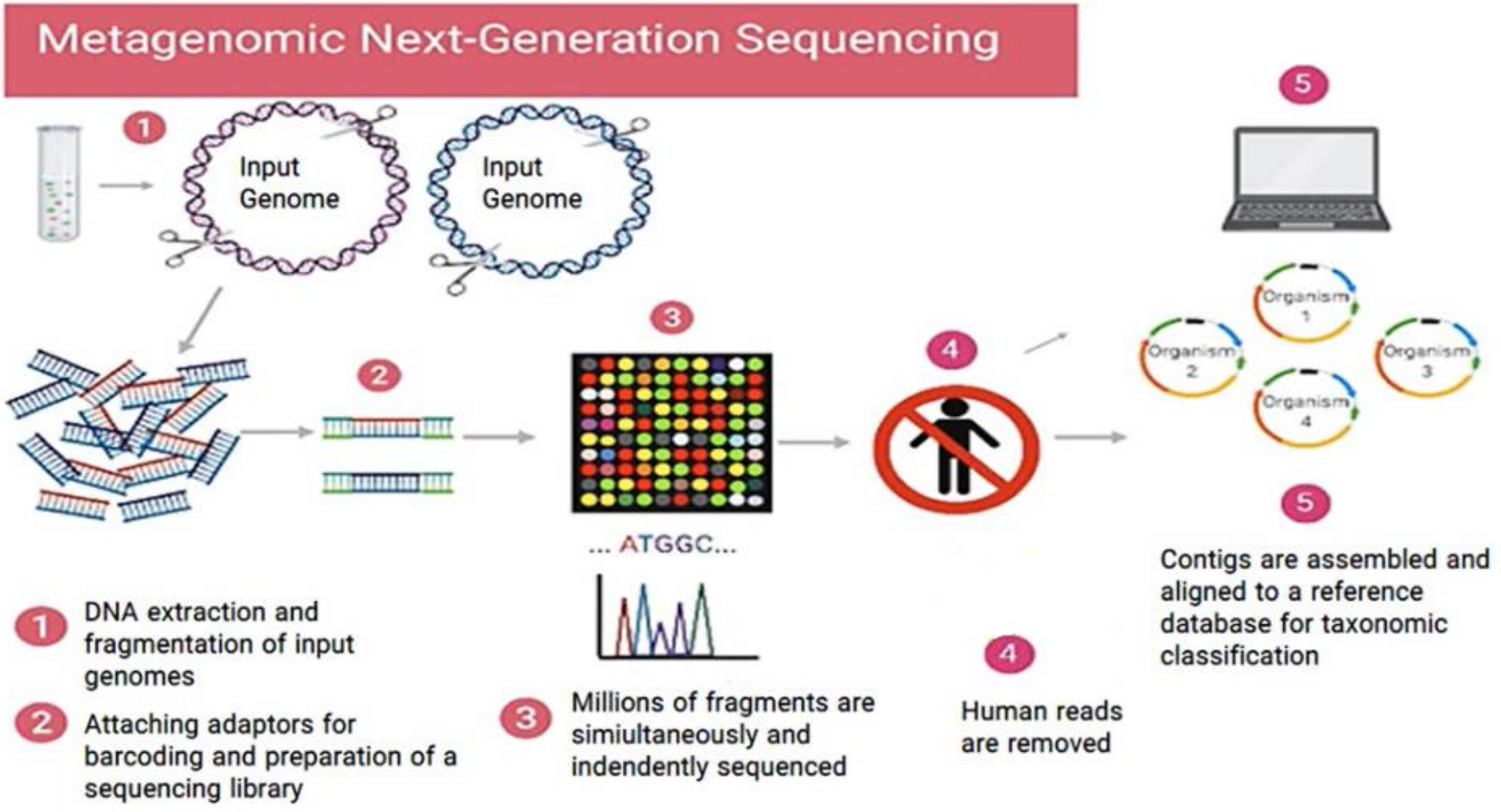
NGS متاژنومی به سادگی تمام اسیدهای نوکلئیک موجود در یک نمونه را که ممکن است شامل جمعیت مخلوطی از میکروارگانیسم‌ها باشد، اجرا می‌کند و این موارد را به ژنوم‌های مرجع آن‌ها اختصاص می‌دهد تا بفهمد:

Metagenomics

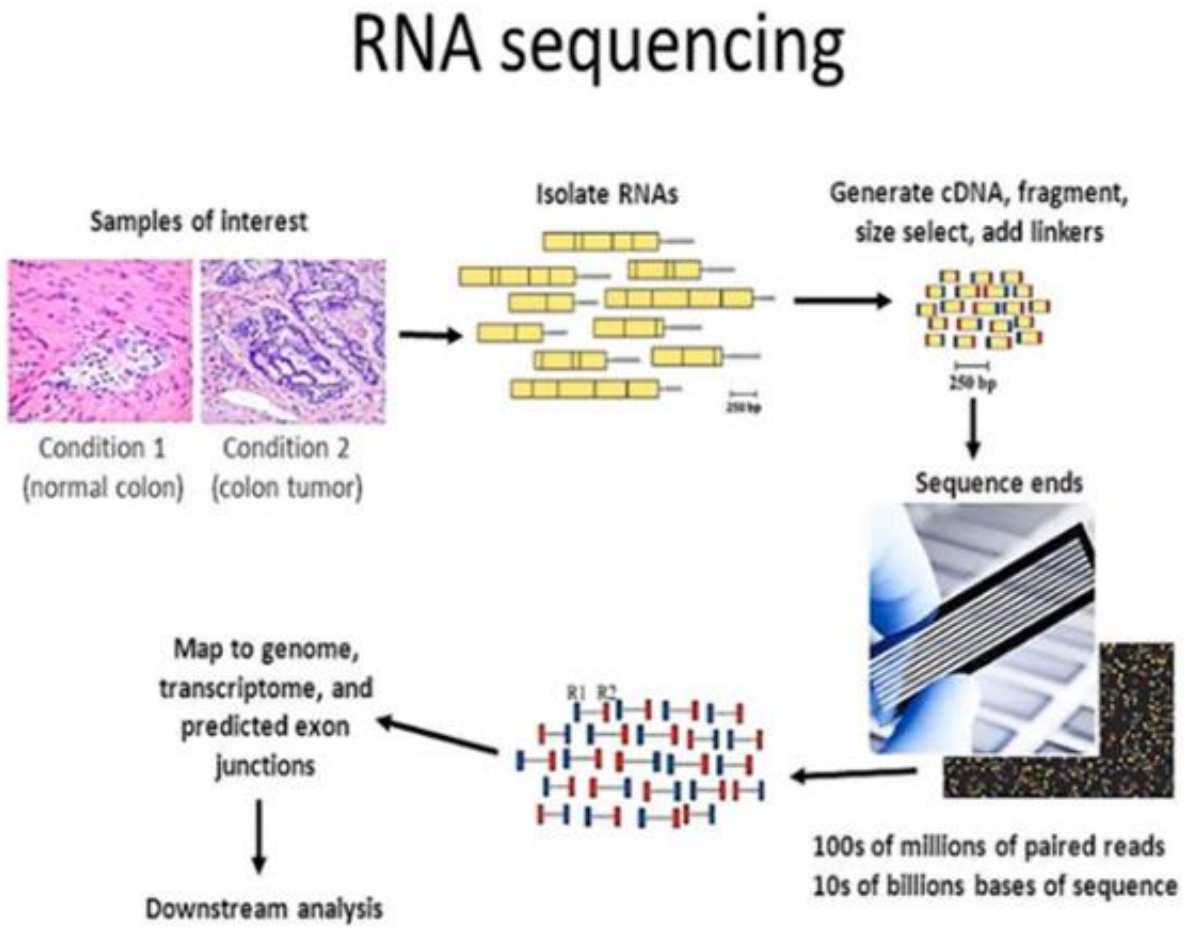
- تعیین توالی و شناسایی اسیدهای نوکلئیک از چندین گونه مختلف
- برای تجزیه‌وتحلیل متاژنومی



Metagenomics



Transcriptomics (RNA) Seq



:RNA-seq

بررسی کمیت و توالی RNA در یک نمونه با استفاده از توالی نسل بعدی.

این نسخه متن از الگوهای بیان ژن رمزگذاری‌شده در RNA ما را تجزیه‌وتحلیل می‌کند.

Targeted RNA seq

تعیین توالی RNA هدفمند

روشی کاملا دقیق برای انتخاب و تعیین توالی‌های خاص

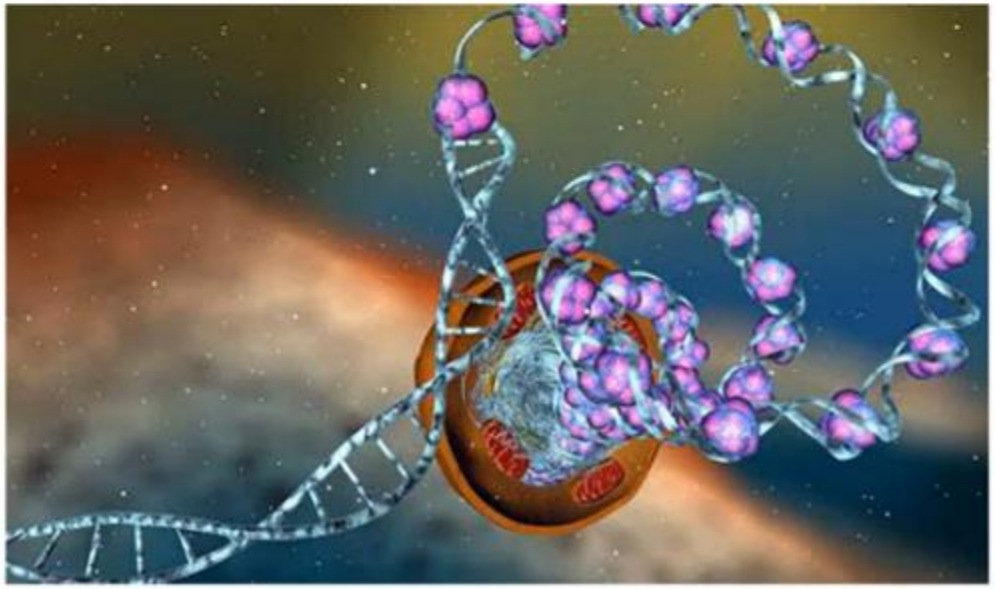
ارائه اطلاعات کمی و هم کیفی

هدف RNA-Seq را می‌توان از طریق غنی‌سازی یا رویکرد مبتنی بر آمپلیکن به دست آورد که هر دو تجزیه‌وتحلیل بیان ژن را در یک مجموعه متمرکز از ژن‌های مورد نظر امکان‌پذیر می‌کند.

Epigenetics

اپی‌ژنتیک: مطالعه تنظیم ژن از طریق عواملی غیر از تغییر در توالی نوکلئوتید.

به عنوان مثال تغییرات هیستون پس از ترجمه مانند متیلاسیون DNA که نشان داده شده است برای تنظیم قابلیت دسترسی ژن برای رونویسی است.



Methylation Sequencing

متیلاسیون DNA

- یک تغییر مهم اپی‌ژنتیکی
- تاثیرگذاری بر چندین فرایند رشد مهم

انحراف در این فرایند، مانند هیپو یا هیپرمتیلاسیون سیتوزین-گوانین (CpG) دی نوکلئوتیدها

↓

بی‌ثباتی ژنومی یا خاموش‌شدن رونویسی

↓

توسعه بیماری‌های مختلف: سرطان، دیابت، بیماری‌های قلبی عروقی و التهابی یا اختلالات روانی

Methylation Sequencing

تعیین دقیق وضعیت متیلاسیون DNA از این رو حیاتی است.

توالی متیلاسیون (توالی متیل seq یا بی‌سولفیت):

- ابزاری قدرتمند برای درک متیلاسیون در کل ژنوم با رزولوشن تک‌نوکلئوتیدی
- یک استاندارد طلایی

ChIP Sequencing

- Chromatin Immunoprecipitation Sequencing

- روشی برای تجزیه‌وتحلیل فعل‌وانفعالات پروتئین با DNA
- رسوب‌گذاری ایمنی کروماتین (CHIP) NGS همراه با توالی مقیاس بزرگ NGS

تهیه نقشه‌های کل ژنوم از محل فاکتورهای رونویسی یا هیستون‌های اصلاح‌شده

- مطالعات ChIP-seq در مورد اصلاحات هیستون ⟸ افزایش دانش پیرامون تنظیم ژن در تحقیقات سرطان

Ribosome Profiling (Foot Printing)

پروفایل ریبوزوم:

- یک کاربرد جدید جذاب NGS در تجزیه‌وتحلیل پروتئین.
- شناسایی موقعیت ریبوزوم‌های فعال که به یک mRNA هدف متصل می‌شوند.
- ضبط فرایند NGS، mRNA محافظت‌شده از ریبوزوم
- مشخص‌کردن مکان‌های اتصال ریبوزوم توالی mRNA
- ترسیم توالی خوانده‌شده به ژنوم انسان
- شناسایی پروتئین‌های ترجمه‌شده


Thanks for attachment