

Informática Biomédica- FMRP/USP

**Disciplina
RIB0102- Genética
Molecular**

Aula

Princípios de Sequenciamento de DNA

Wilson Araújo da Silva Junior

Departamento de Genética da FMRP/USP

<http://rge.fmrp.usp.br>

wilsonjr@usp.br



1977

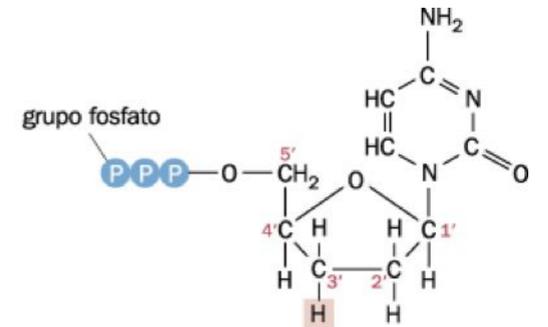
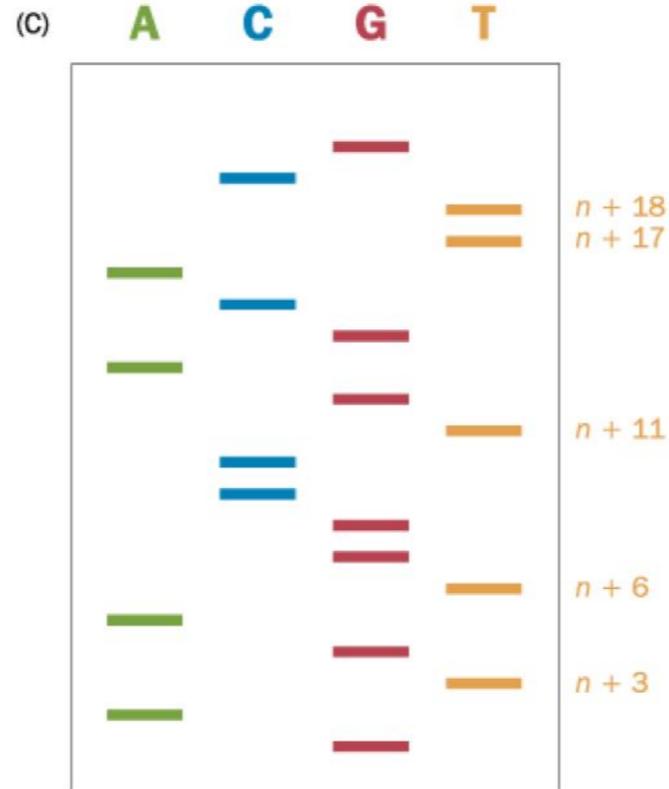
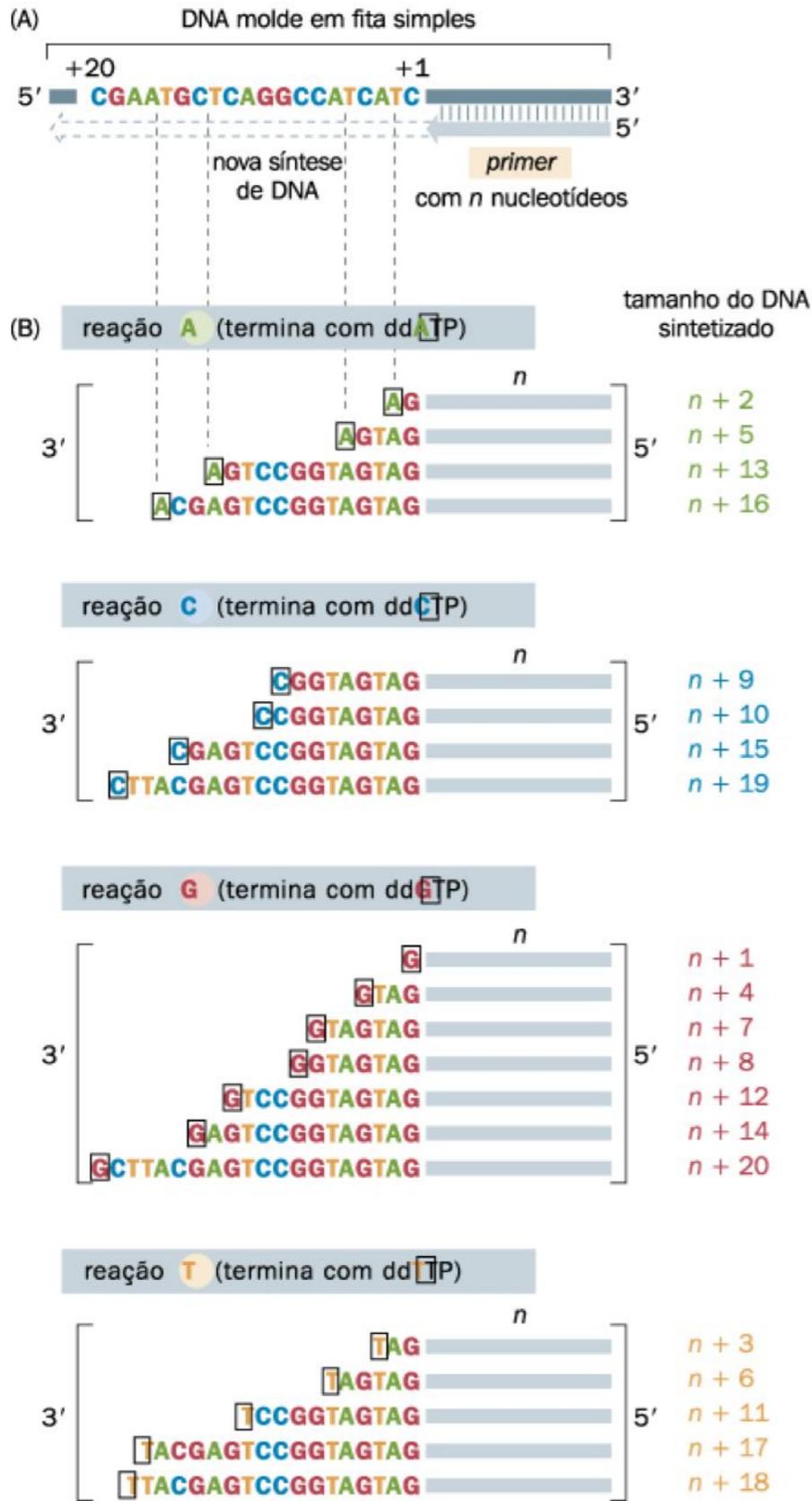
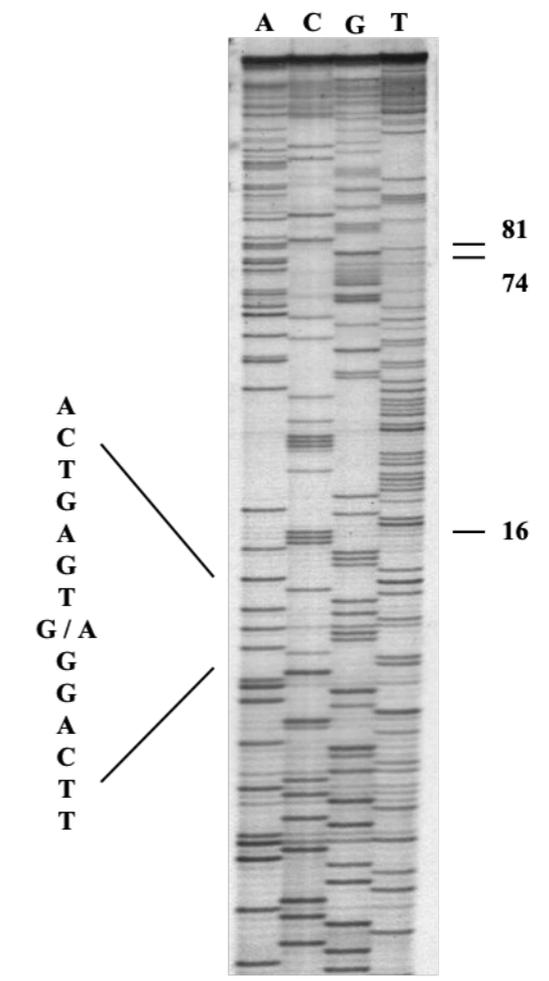
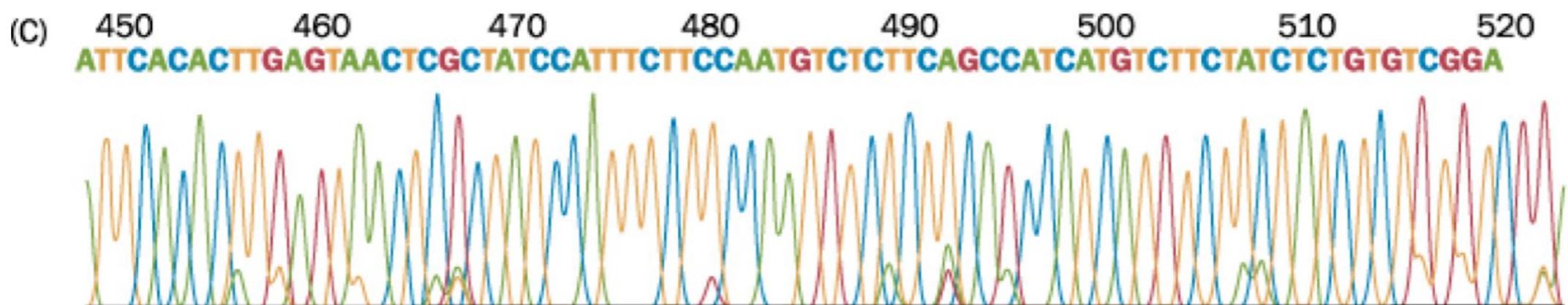
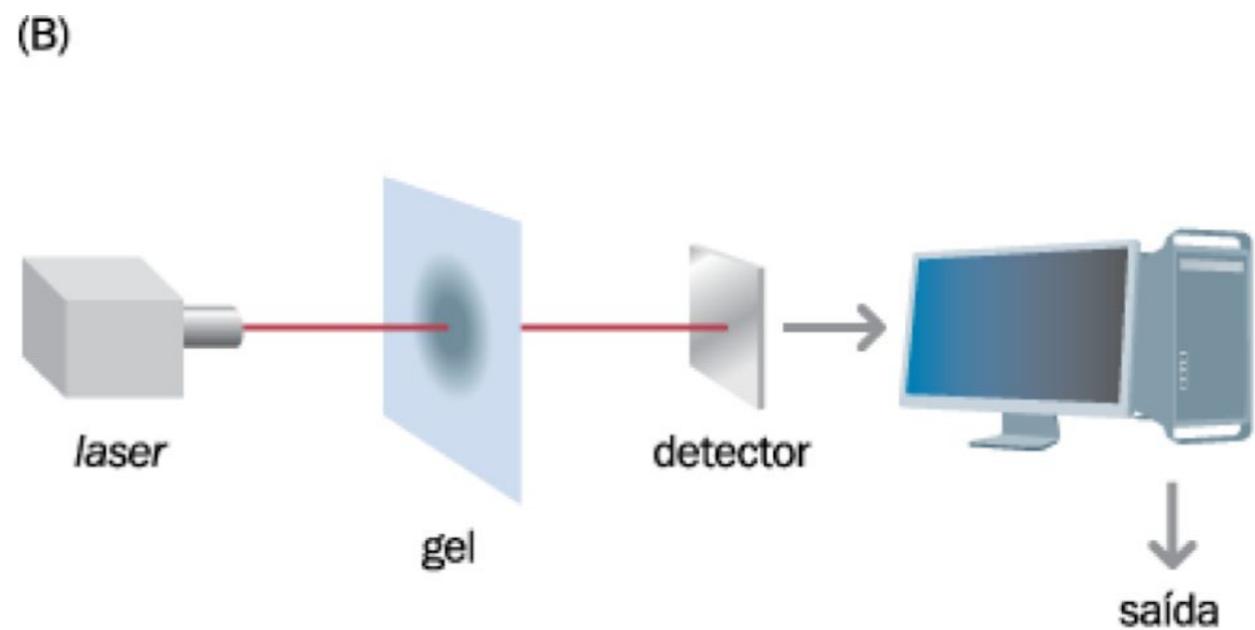
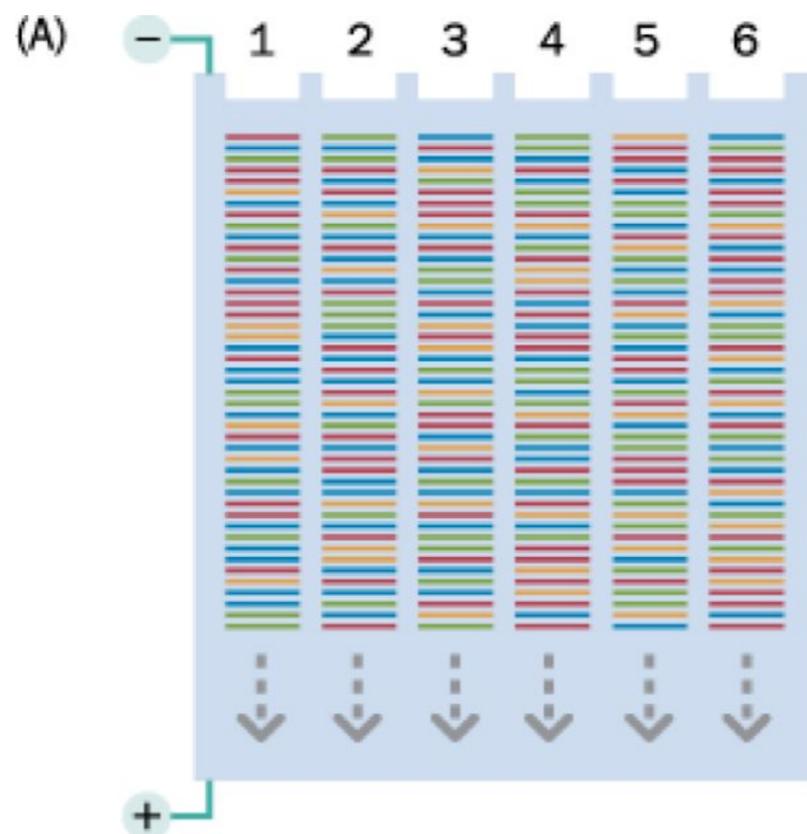


Figura 8.5 Estrutura de um dideoxinucleotídeo: 2', 3'-dideoxi CTP (ddCTP). O grupo hidroxila ancorado ao carbono 3' em nucleotídeos normais é trocado por um átomo de hidrogênio (sombreado).



Hb A/ β^0 -Talassemia (IVS-2 splice junction G-A)



Next Generation Sequencing



2001



2011

Biologia Computacional

Biologia de Sistemas Moleculares

Protocols:

DNA

- Whole Genome-Seq
- Exome-Seq
- Metagenomics
- 16S Ribosomal diversity

Structural and genomic alterations/mutation

Epigenomics

- Chip-Seq
- Methyl-Seq

Regulation of Gene Expression

RNA

- mRNA-Seq
- miRNA-Seq

Expression of Genes
(canonical and non-canonical)

Análise de dados em escala genômica

Next Generation Sequencing

1. Definição do problema biológico;
2. Discussão do desenho experimental;
 - Tipo de amostra, número amostral, abordagem de NGS, etc.
4. Discussão dos resultados;
5. Conclusão do estudo.

Análise de dados em escala genômica

Next Generation Sequencing



UNIVERSITY OF CALIFORNIA
SANTA CRUZ Genomics Institute



UCSC

Genome Browser Gateway

Browse/Select Species

POPULAR SPECIES








Human Mouse Rat Fruitfly Worm Yeast

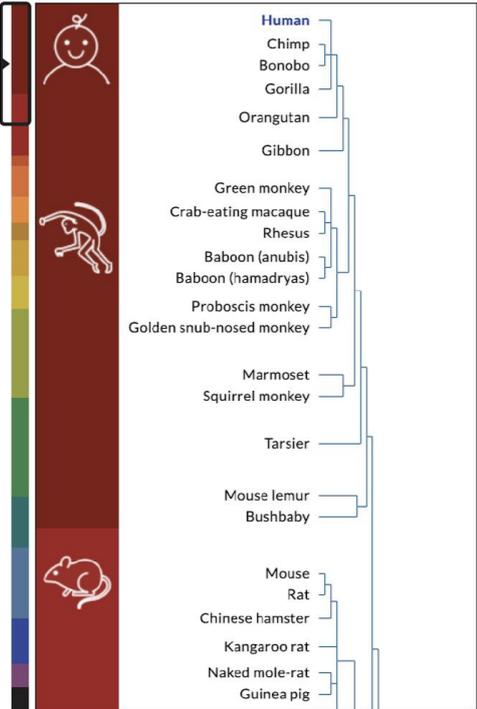
Find Position

Human Assembly
Feb. 2009 (GRCh37/hg19)

Position/Search Term
Enter position, gene symbol or search terms
Current position: chrX:153,295,836-153,296,474

GO

REPRESENTED SPECIES



Human Genome Browser - hg19 assembly [view sequences](#)

The February 2009 human reference sequence (GRCh37) was produced by the [Genome Reference Consortium](#). For more information about this assembly, see [GRCh37](#) in the NCBI Assembly database.

Sample position queries

A genome position can be specified by the accession number of a sequenced genomic clone, an mRNA or EST or STS marker, a chromosomal coordinate range, or keywords from the GenBank description of an mRNA. The following list shows examples of valid position queries for the human genome. See the [User's Guide](#) for more information.

Request:	Genome Browser Response:
chr7	Displays all of chromosome 7
chrUn_gl000212	Displays all of the unplaced contig gl000212
20p13	Displays region for band p13 on chr 20
chr3:1-1000000	Displays first million bases of chr 3, counting from p-arm telomere
chr3:1000000+2000	Displays a region of chr3 that spans 2000 bases, starting with position 1000000
RH18061;RH80175 15q11:15q13 rs1042522;rs1800370	Displays region between genome landmarks, such as the STS markers RH18061 and RH80175, or chromosome bands 15q11 to 15q13, or SNPs rs1042522 and rs1800370. This syntax may also be used for other range queries, such as between uniquely determined ESTs, mRNAs, refSeqs, etc.
D16S3046	Displays region around STS marker D16S3046 from the Genethon/Marshfield maps. Includes 100,000 bases on each side as well.
AA205474	Displays region of EST with GenBank accession AA205474 in BRCA1 cancer gene on chr 17
AC008101	Displays region of clone with GenBank accession AC008101
AF083811	Displays region of mRNA with GenBank accession number AF083811
PRNP	Displays region of genome with HUGO Gene Nomenclature Committee identifier PRNP
NM_017414	Displays the region of genome with RefSeq identifier NM_017414
NP_059110	Displays the region of genome with protein accession number NP_059110
pseudogene mRNA	Lists transcribed pseudogenes, but not cDNAs
homeobox caudal	Lists mRNAs for caudal homeobox genes
zinc finger	Lists many zinc finger mRNAs
kruppel zinc finger	Lists only kruppel-like zinc fingers
huntington	Lists candidate genes associated with Huntington's disease
zahler	Lists mRNAs deposited by scientist named Zahler
Evans,J.E.	Lists mRNAs deposited by co-author J.E. Evans

Use this last format for author queries. Although GenBank requires the search format *Evans JE*, internally it uses the format *Evans,J.E.*



Homo sapiens
(Graphic courtesy of CBSE)

NGS (Next Generation Sequencing)

Etapas:

1. Extração do DNA ou RNA
2. Preparo da biblioteca (fragmentação randômica)
3. Sequenciamento da biblioteca
4. Análise dos dados (Bioinformática)

NGS (Next Generation Sequencing)

1. Extração do DNA ou RNA

- Manual ou automatizada
- Qualidade (pureza e integridade)

Pureza: Contaminação com RNA/proteínas

OD 260nm /280nm Razão = 1,8 – 2,0

Amostras com razão acima de 2,0 tem muito RNA

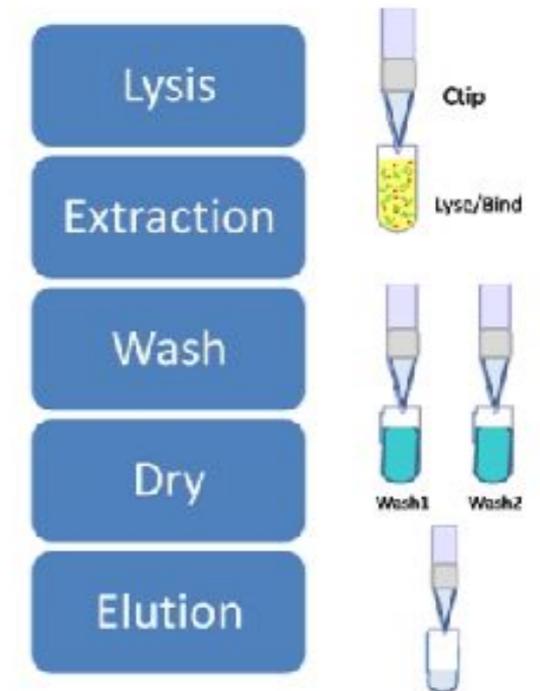
Amostras com razão abaixo de 1,8 tem contaminação com proteína

Contaminação com reagentes

OD 260nm /230nm Razão = 2,0 – 2,2

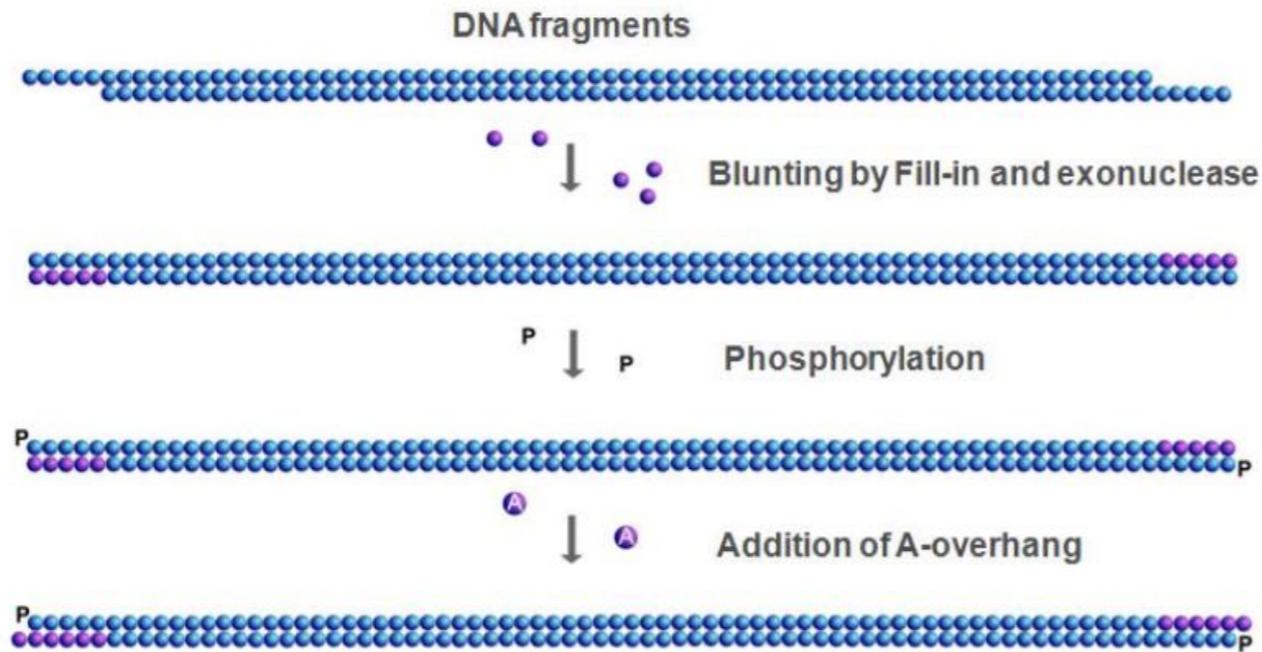
Amostras com razão abaixo do esperado tem contaminação reagentes

Integridade: Fragmentação do DNA ou RNA



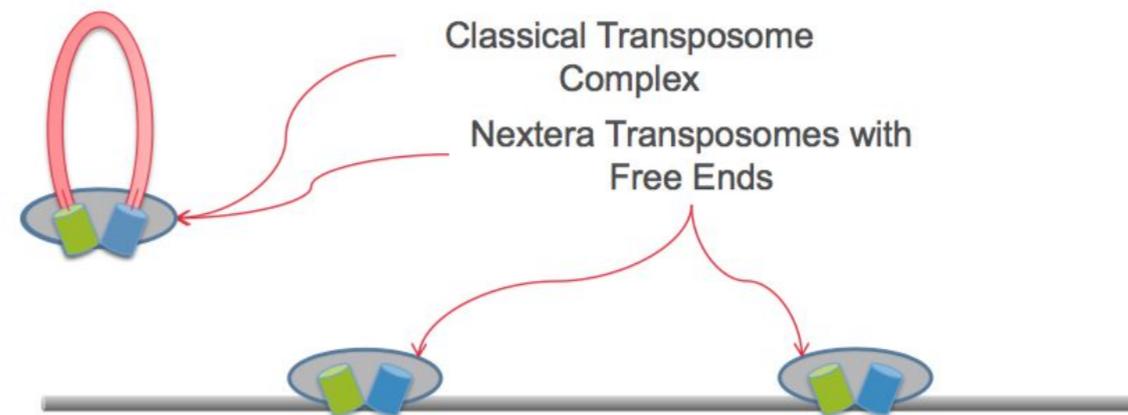
NGS (Next Generation Sequencing)

2. Preparo da biblioteca (fragmentação randômica)



Mecânica ou enzimática

Tagmentação (Transposase)

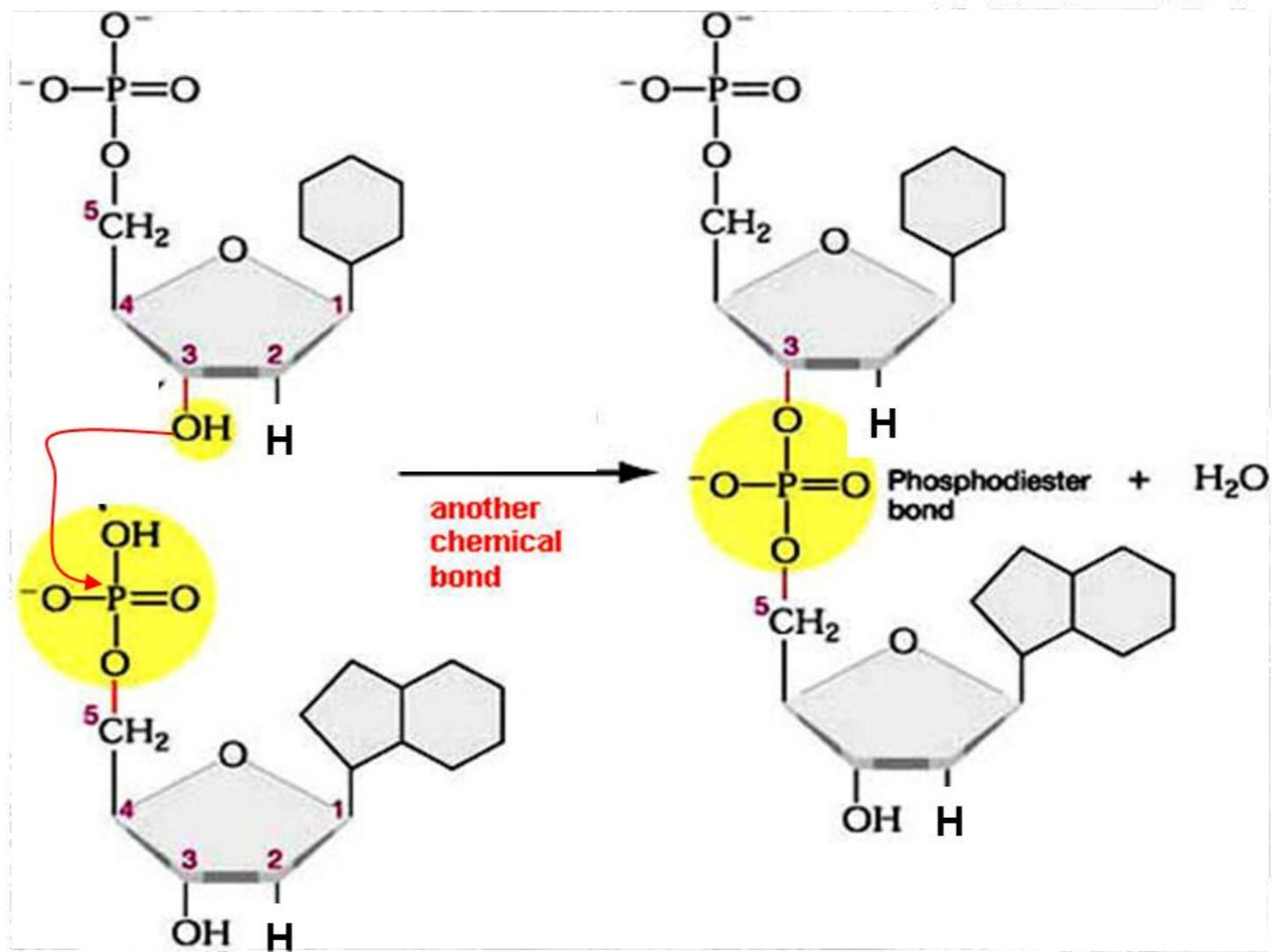
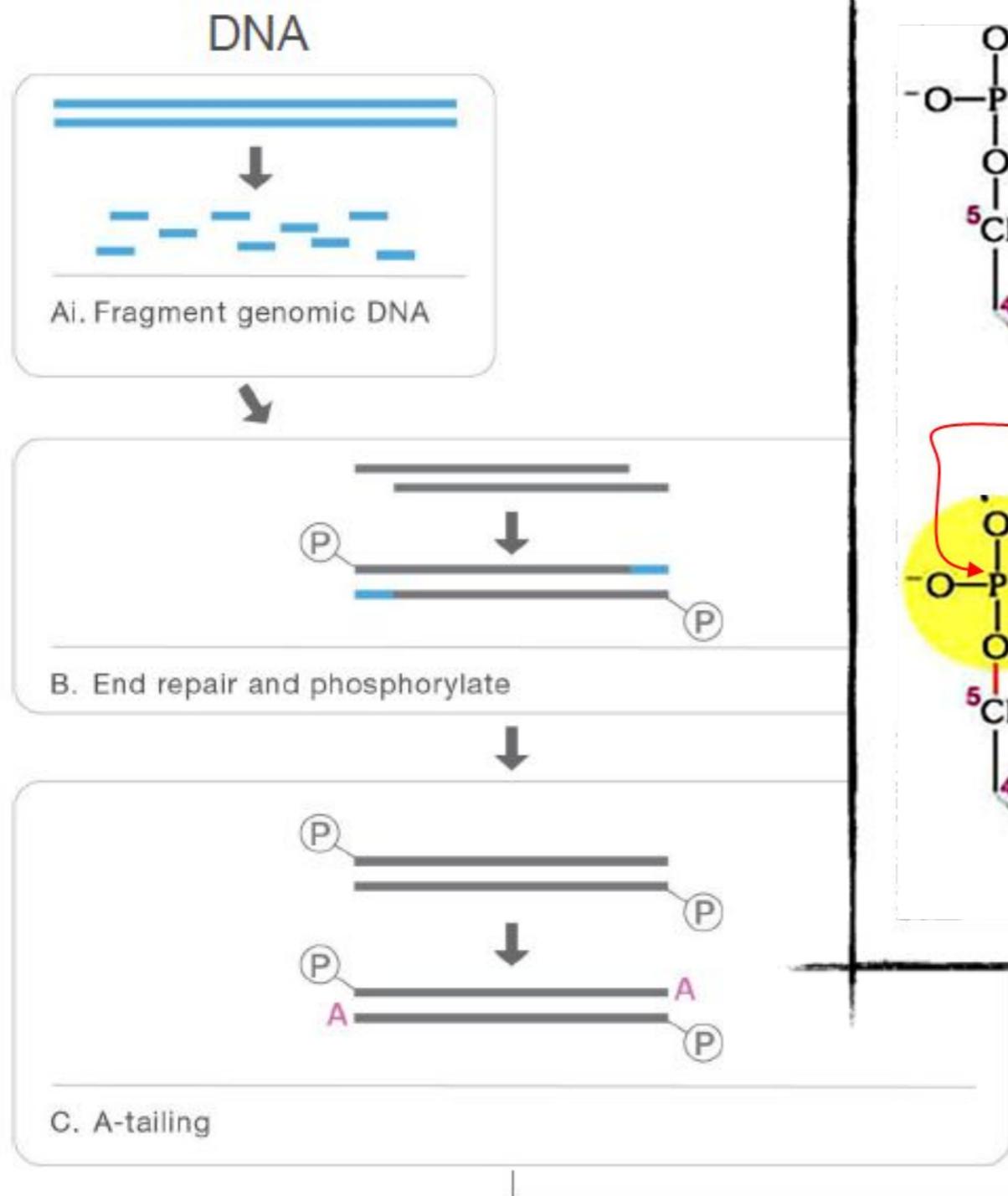


Target DNA is Simultaneously Fragmented and Tagged



NGS (Next Generation Sequencing)

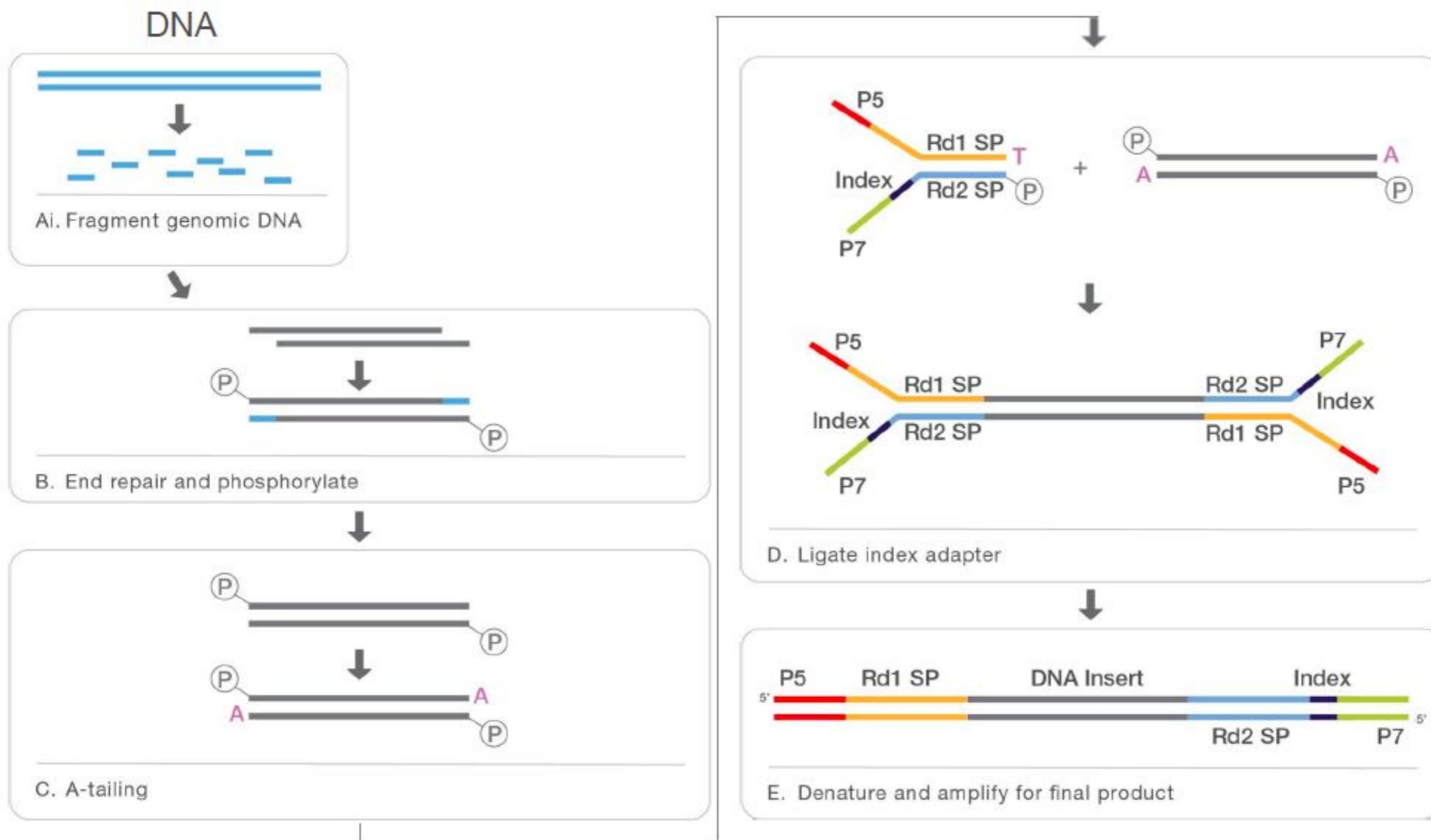
2. Preparo da biblioteca (fragmentação randômica)



E. Denature and amplify for final product

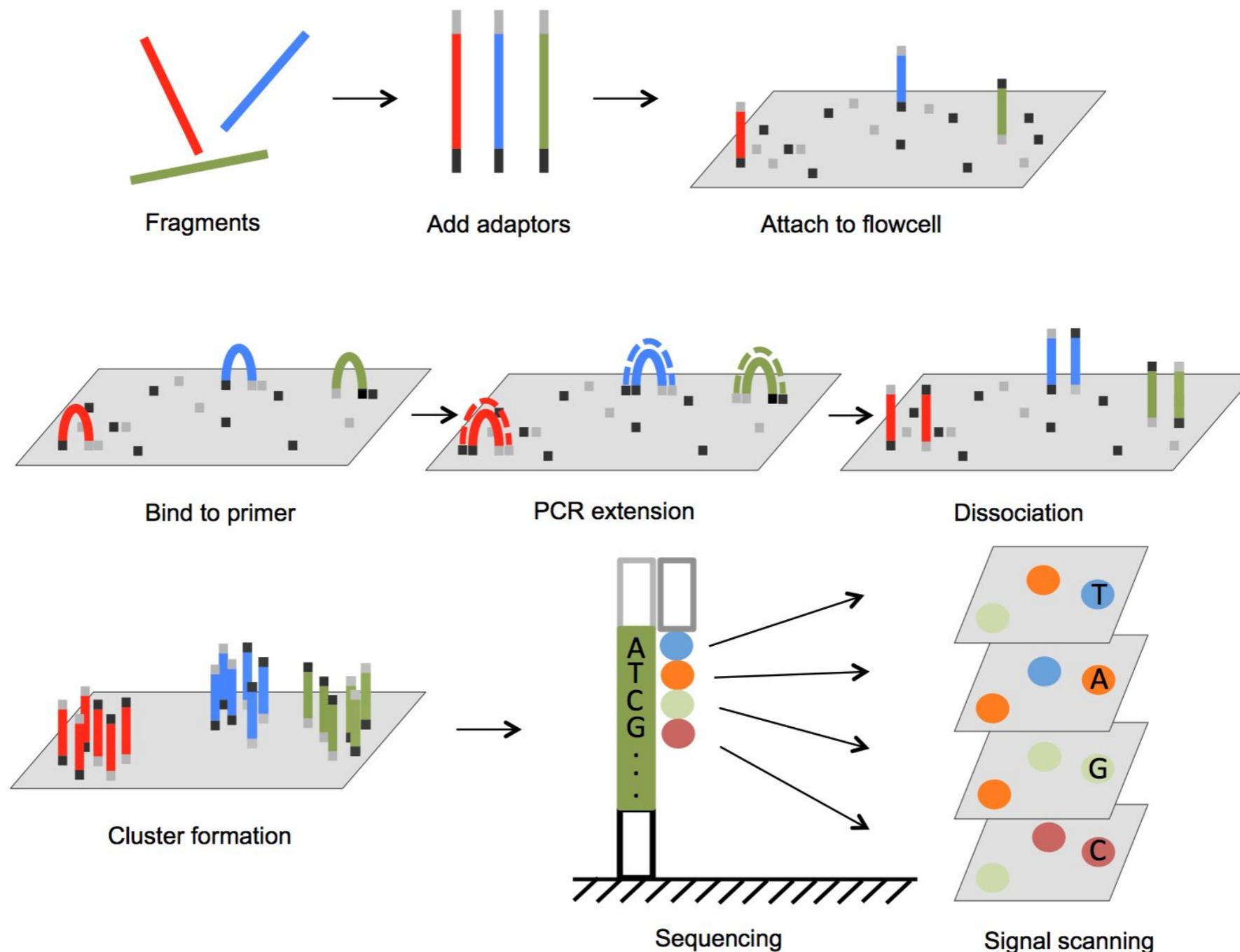
NGS (Next Generation Sequencing)

2. Preparo da biblioteca (fragmentação randômica)



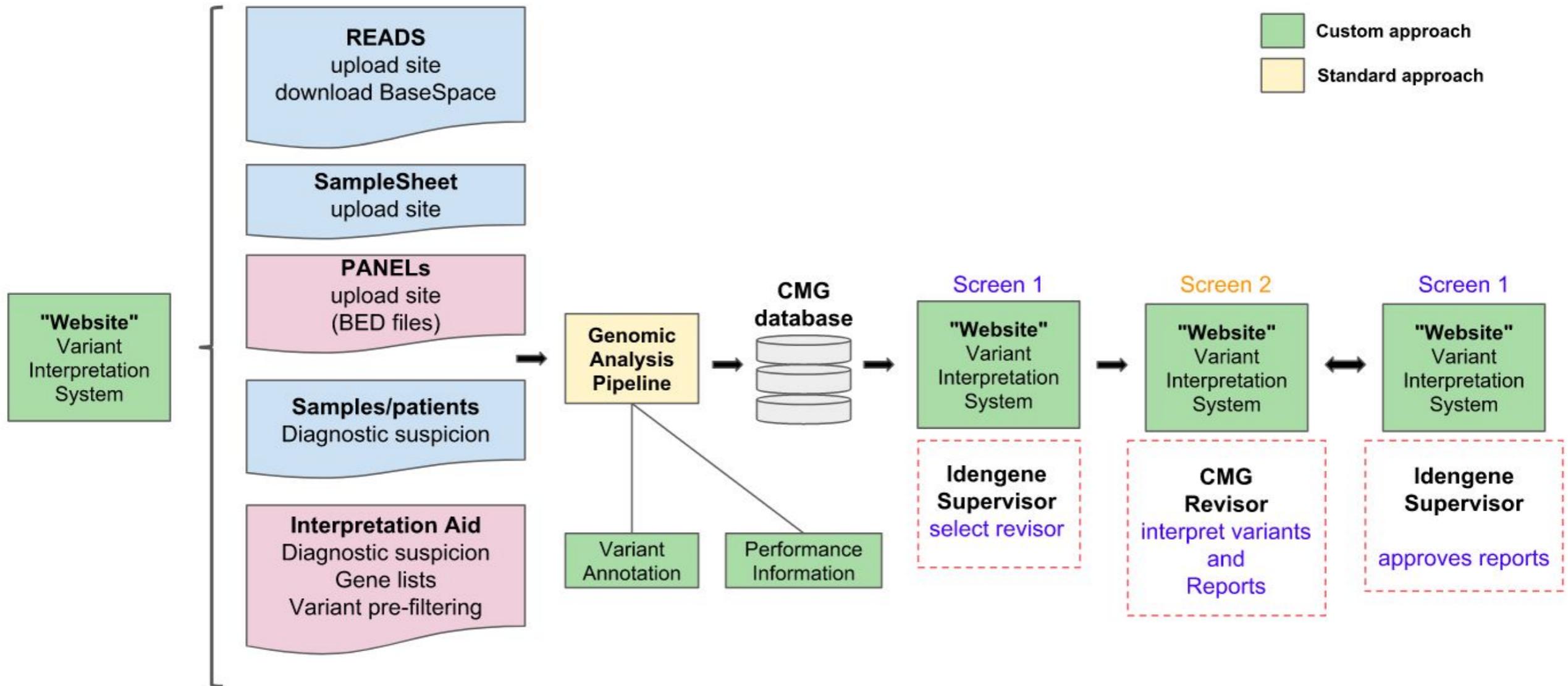
NGS (Next Generation Sequencing)

3. Sequenciamento da biblioteca



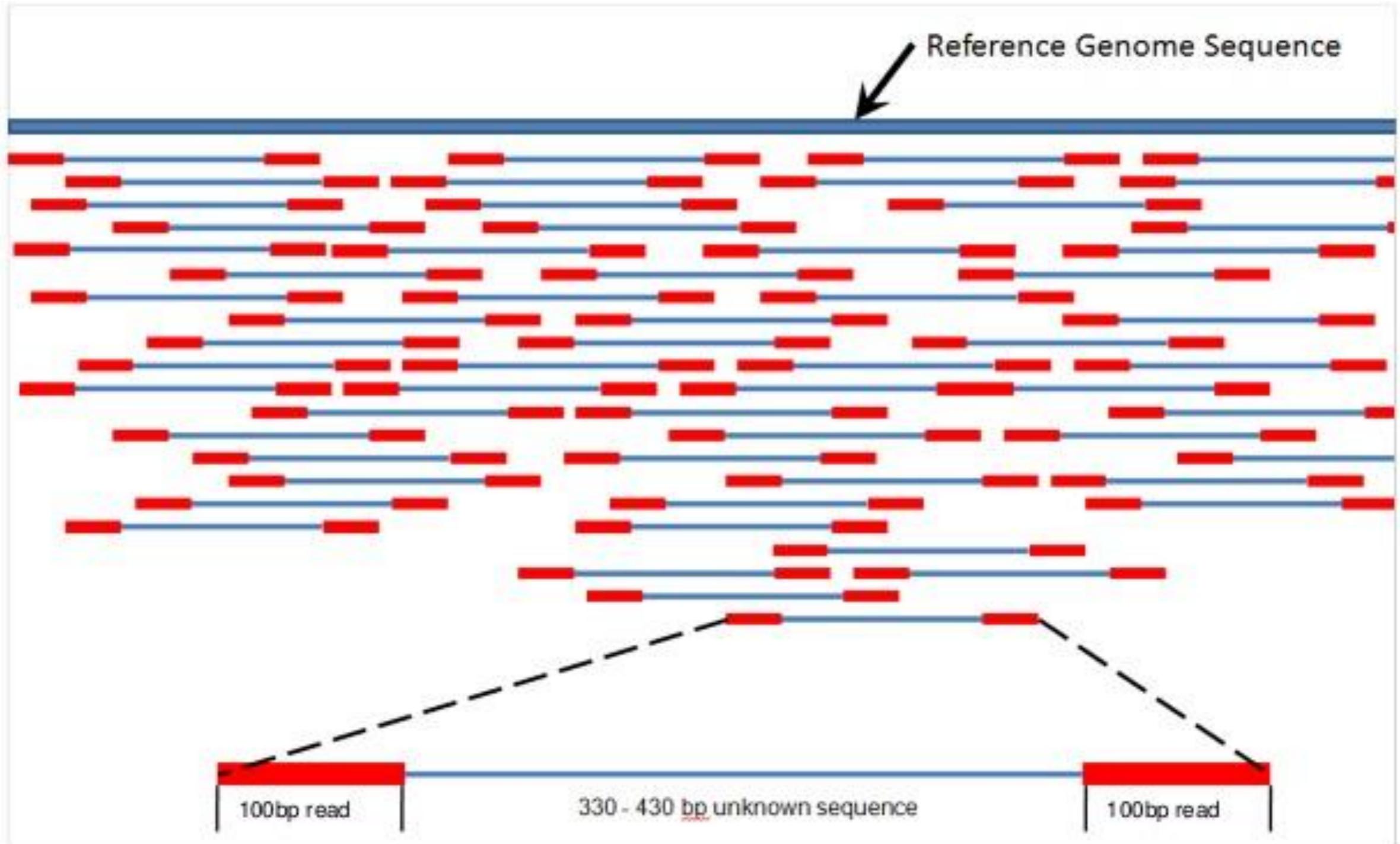
NGS (Next Generation Sequencing)

4. Análise dos dados (Bioinformática)



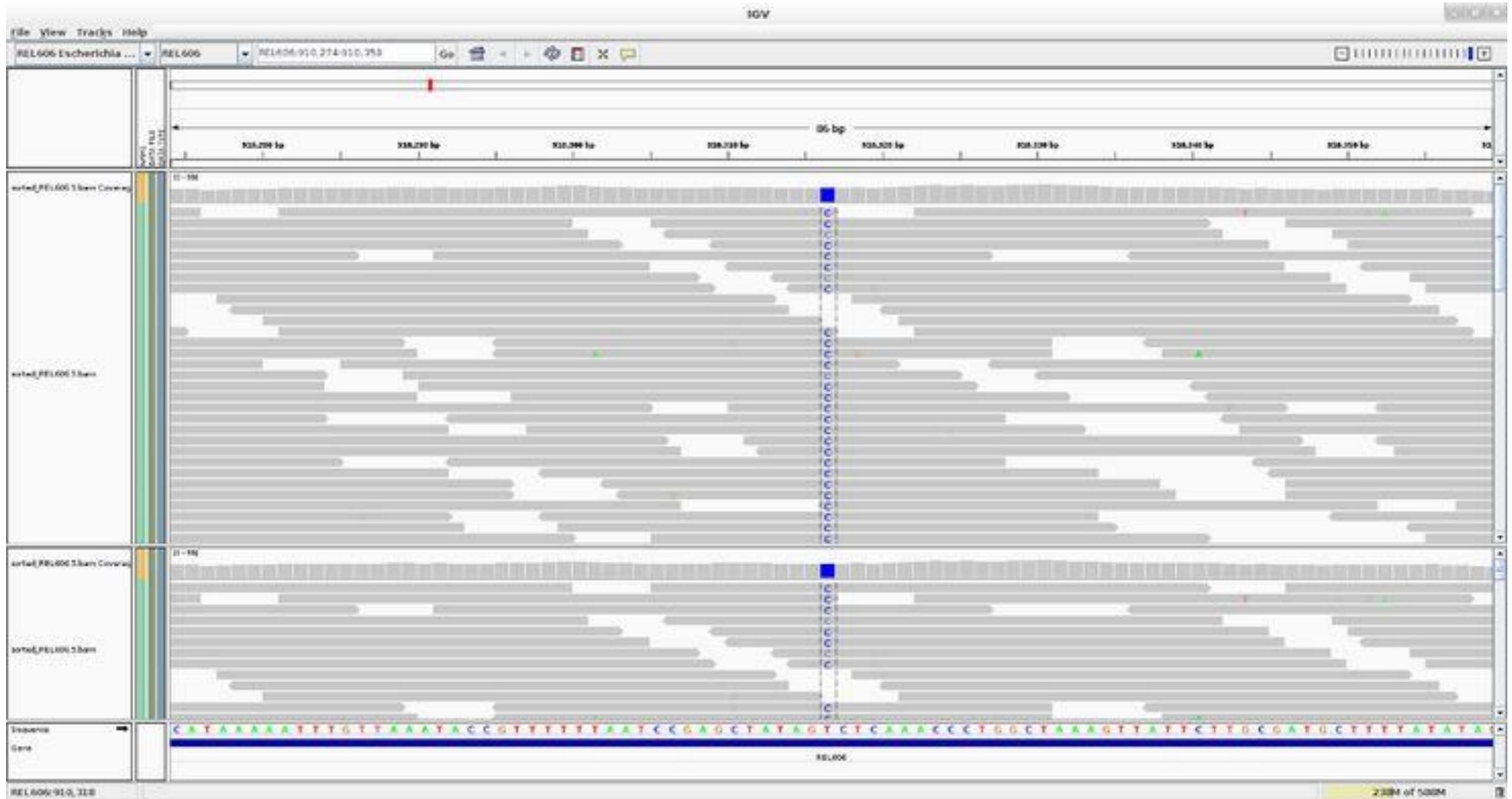
Parâmetros importantes

Cobertura!!!



Parâmetros importantes

Cobertura!!!



Parâmetros importantes

The Lander/Waterman equation

Equação Geral: $C = LN/G$

C: cobertura

L: tamanho da read (sequência)

N: número de reads

G: tamanho do genoma haploide

$$C = (100 \text{ bp}) * (189 \times 10^6) / (3 \times 10^9 \text{ bp}) = 6.3$$



iSeq 100 System



MiniSeq System



MiSeq Series



NextSeq Series

	iSeq 100 System	MiniSeq System	MiSeq Series	NextSeq Series
Run Time	9–17.5 hours	4–24 hours	4–55 hours	12–30 hours
Maximum Output	1.2 Gb	7.5 Gb	15 Gb	120 Gb
Maximum Reads Per Run	4 million	25 million	25 million [†]	400 million
Maximum Read Length	2 x 150 bp	2 x 150 bp	2 x 300 bp	2 x 150 bp