

INTRODUCTION AU SÉQUENÇAGE À HAUT DÉBIT POUR LA GÉNOMIQUE

Claude THERMES

INSTITUT DE BILOGIE INTÉGRATIVE DE LA CELLULE – I2BC

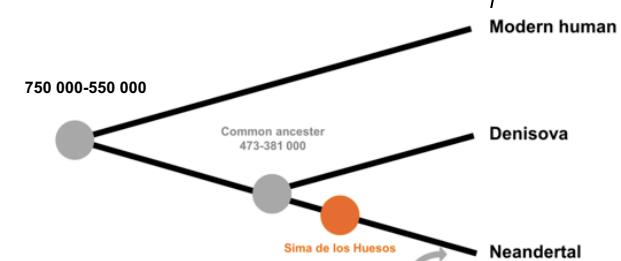
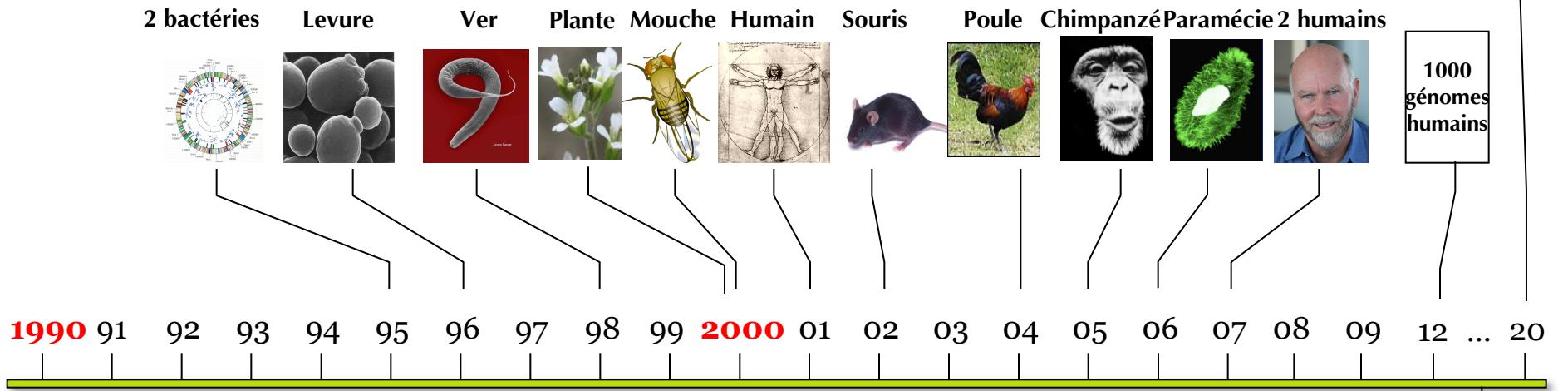
GIF-SUR-YVETTE

Diplôme Universitaire en Bioinformatique intégrative - DU-Bii - 09/03/2020



Premiers génomes entièrement séquencés

Médecine
personnalisée
>> milliers de génomes

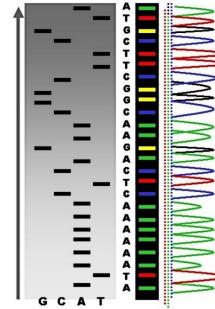


1st generation : Sanger sequencing

- Has been the major method up to 2005

Limitations

- Extremely high cost
- Long experimental set up times
- High DNA concentrations needed



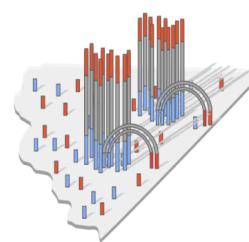
2^d generation

- Single DNA molecules replicated in clusters
- Very high throughput

Limitations

- Maximum read length \leq 300bp

Illumina



3rd generation

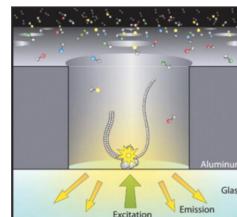
- Single molecules sequencing

- Very long reads

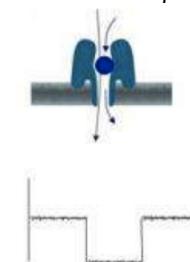
Limitations

- High error rates

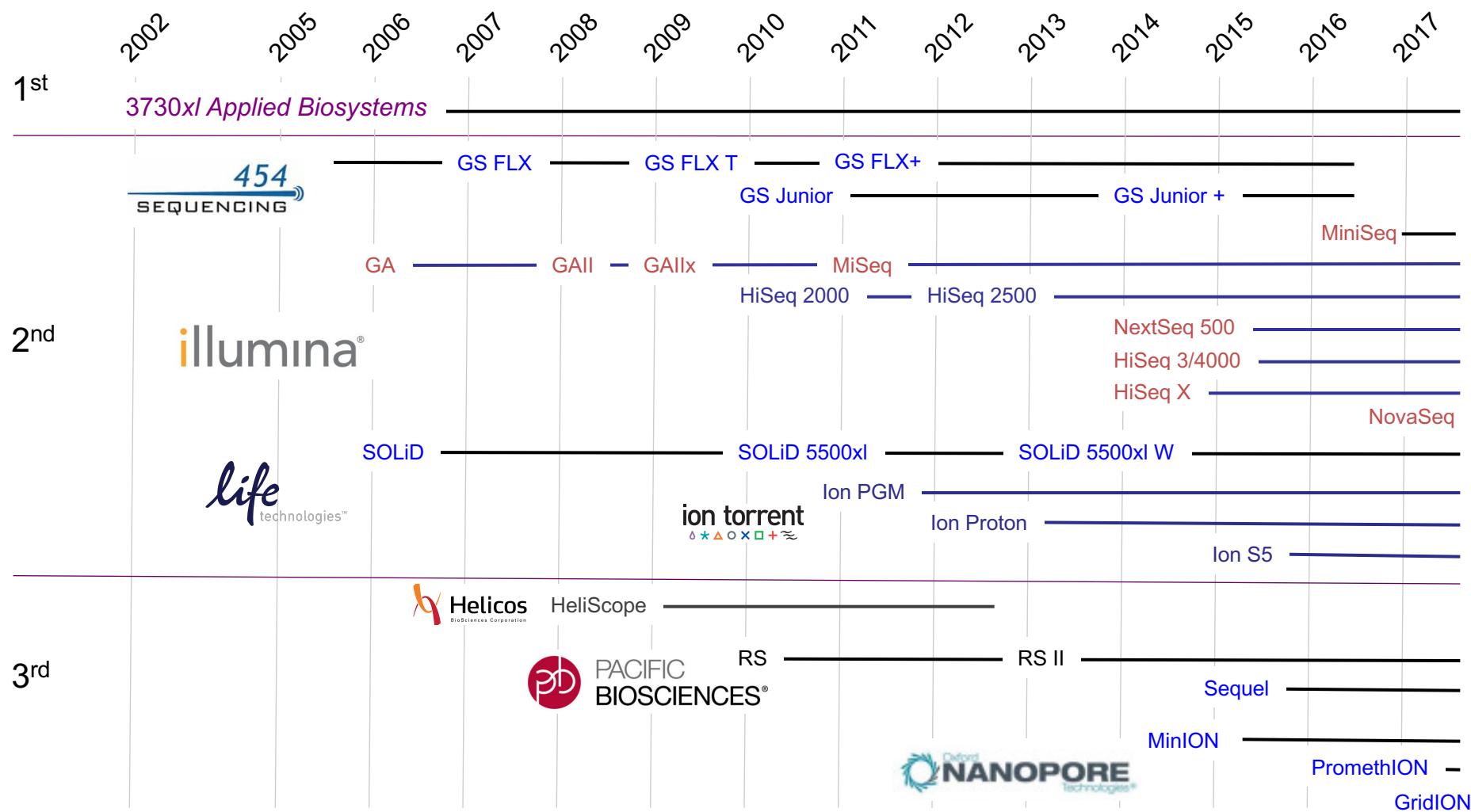
PacBio



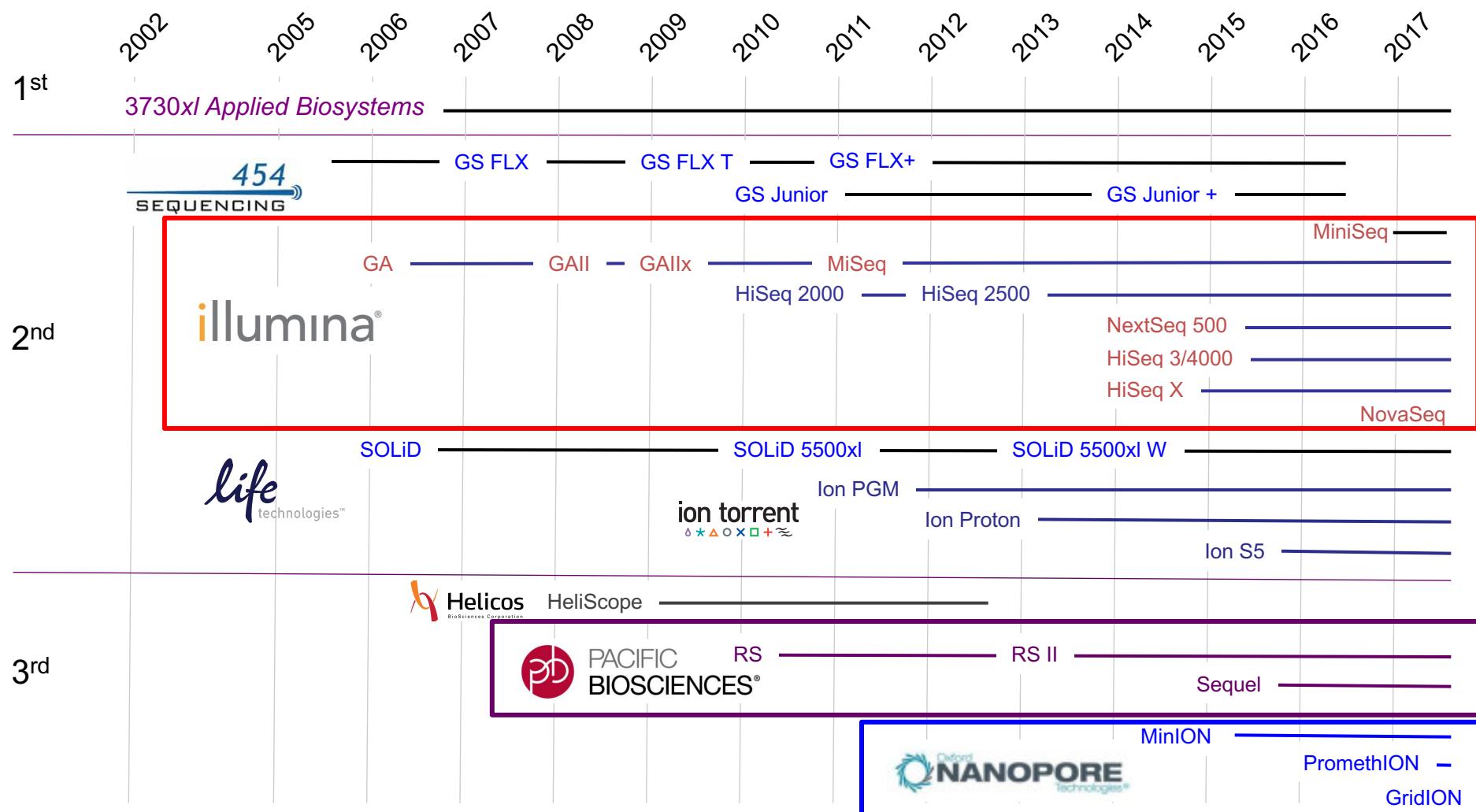
Oxford Nanopore



Sequencing technologies



Sequencing technologies



PART I

2^d GENERATION SEQUENCING

Illumina : the winning technology



MiniSeq
25 million reads



MiSeq
25 millions reads, 2 x 300 bp



NextSeq
400 million reads



HisSeq 4000
5 billion reads

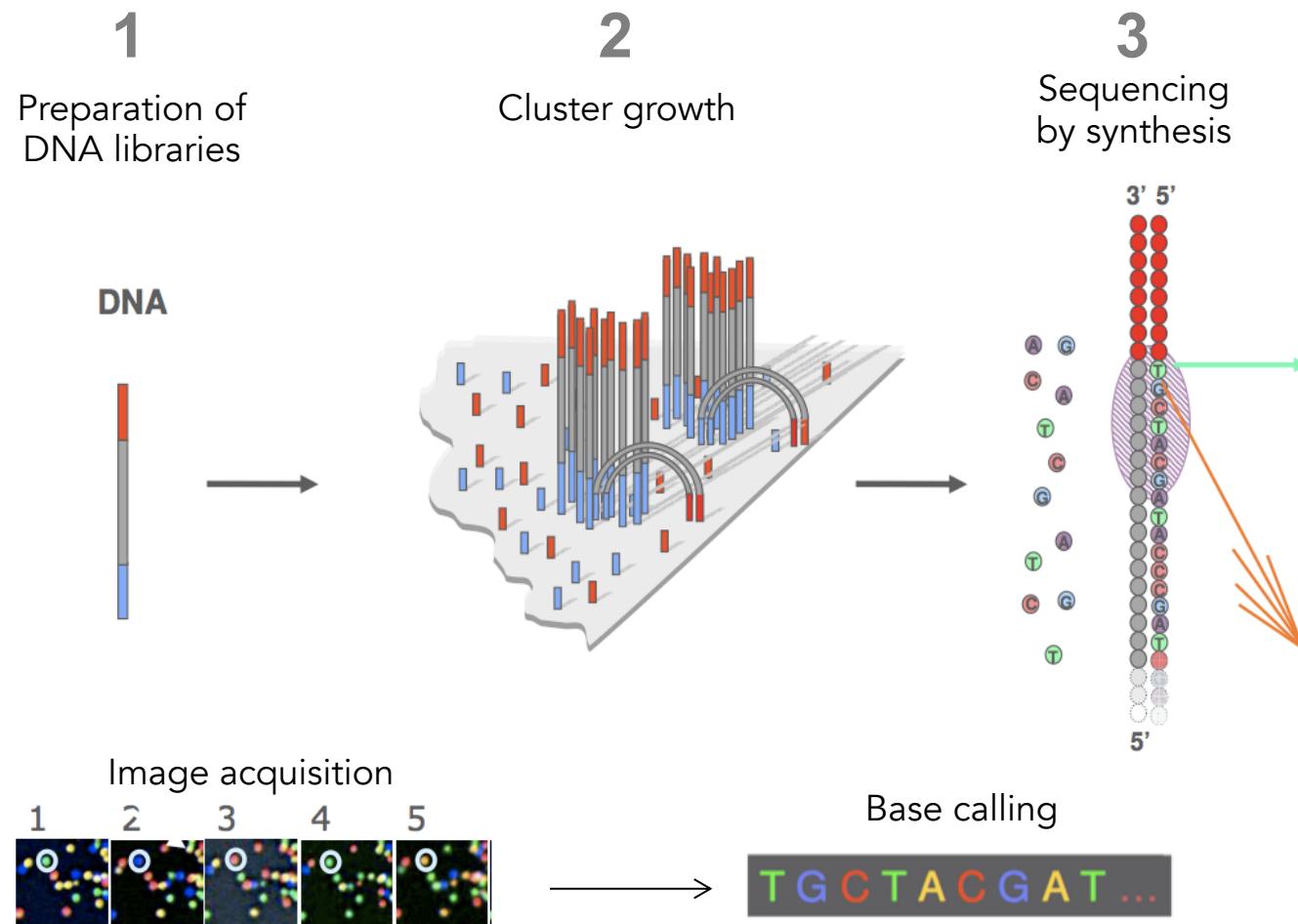


HisSeq X
6 billion reads



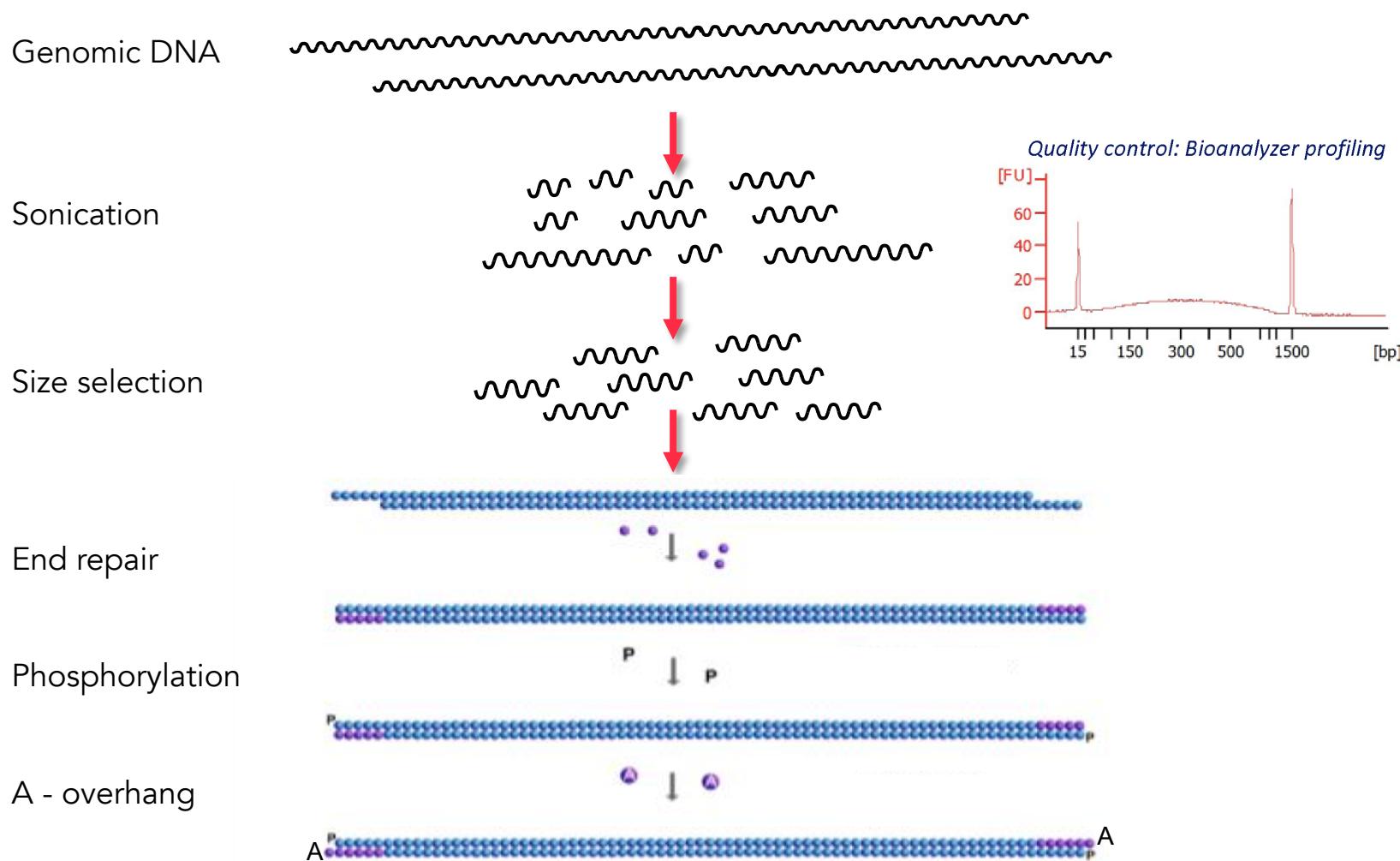
NovaSeq 6000
20 billion reads

General scheme of Illumina sequencing

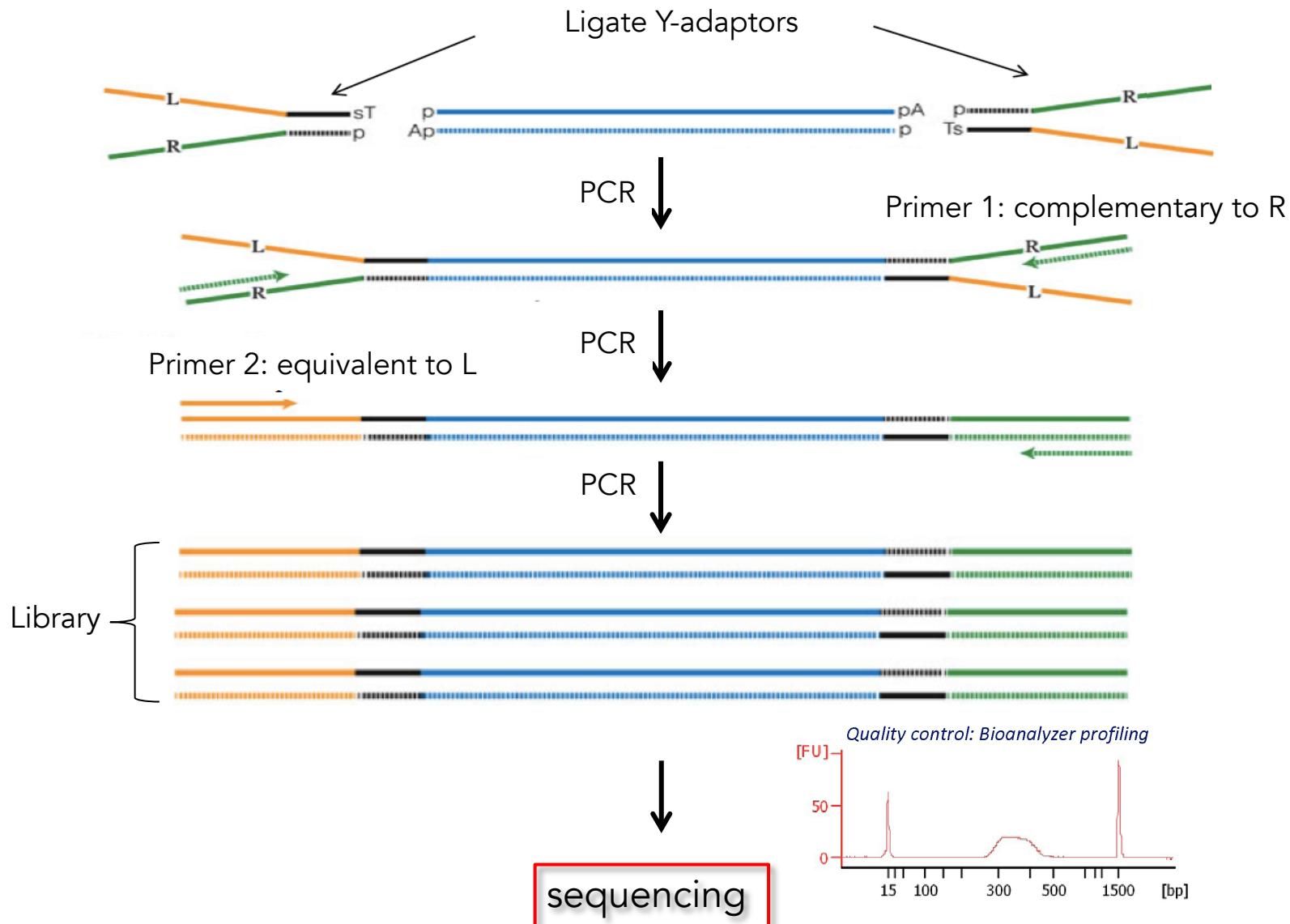


1 - Preparation of DNA-seq Libraries

Illumina TruSeq technology



1 - Preparation of DNA-seq Libraries



1 - Preparation of DNA-seq Libraries

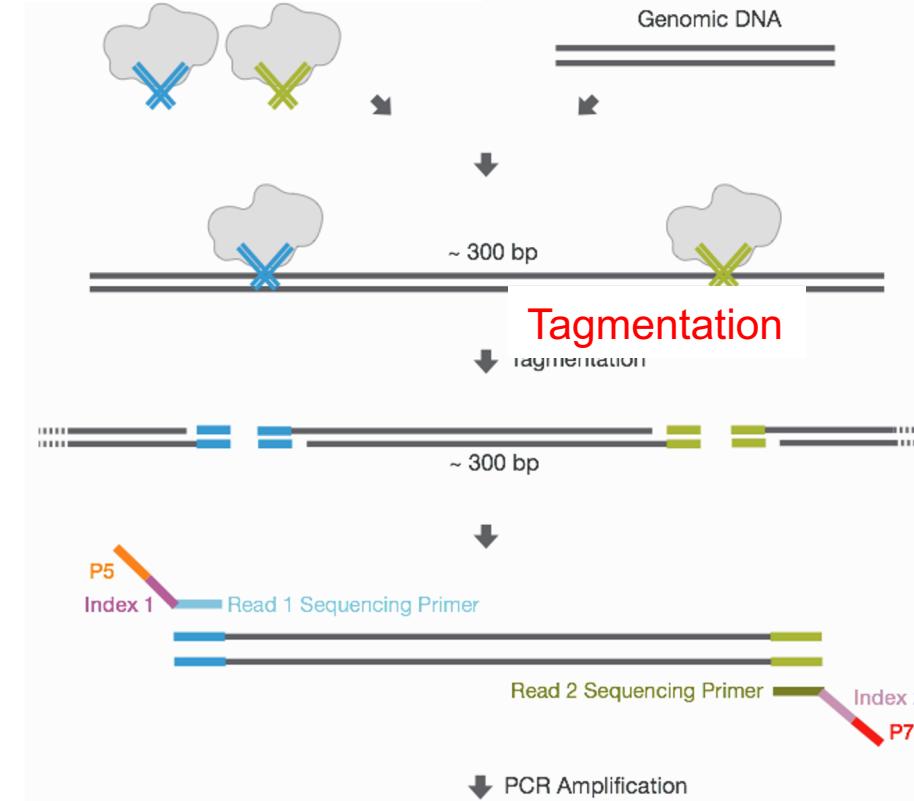
NexTera "tagmentation"

Tagment Enzyme fragments DNA and attaches junction adapters (blue and green) to both ends of the tagmented molecule

Dual barcode approach

up to 96 indexed samples

Transposomes / Tagment Enzyme



requires small quantities 1ng (bacteria) to 50 ng (human)

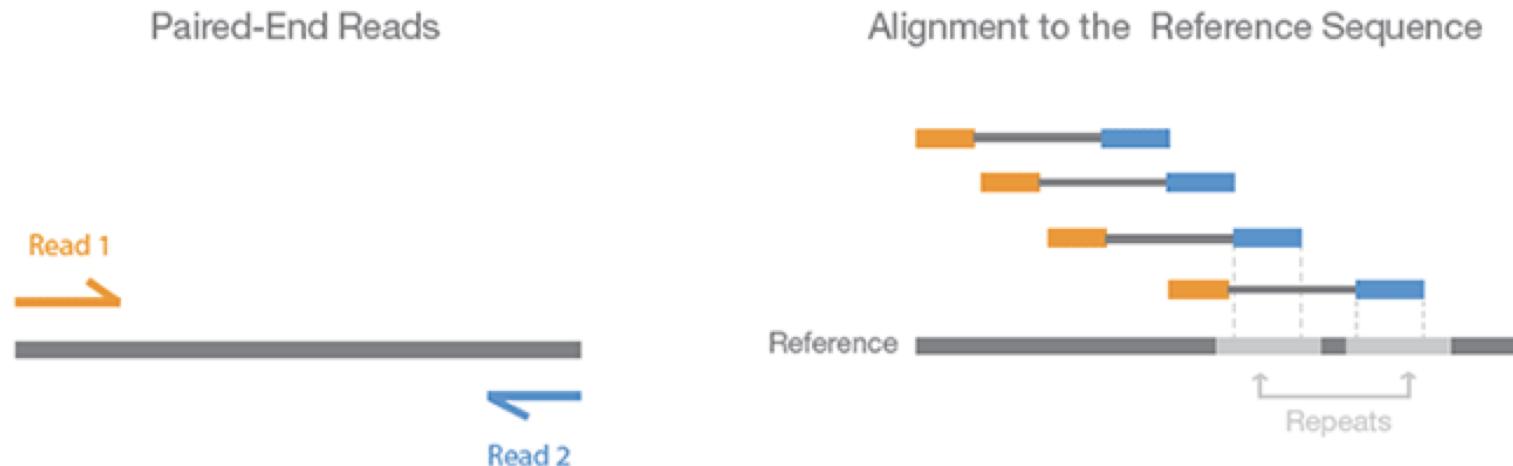
1 - Preparation of DNA-seq Libraries

SINGLE READ and PAIRED-END SEQUENCING

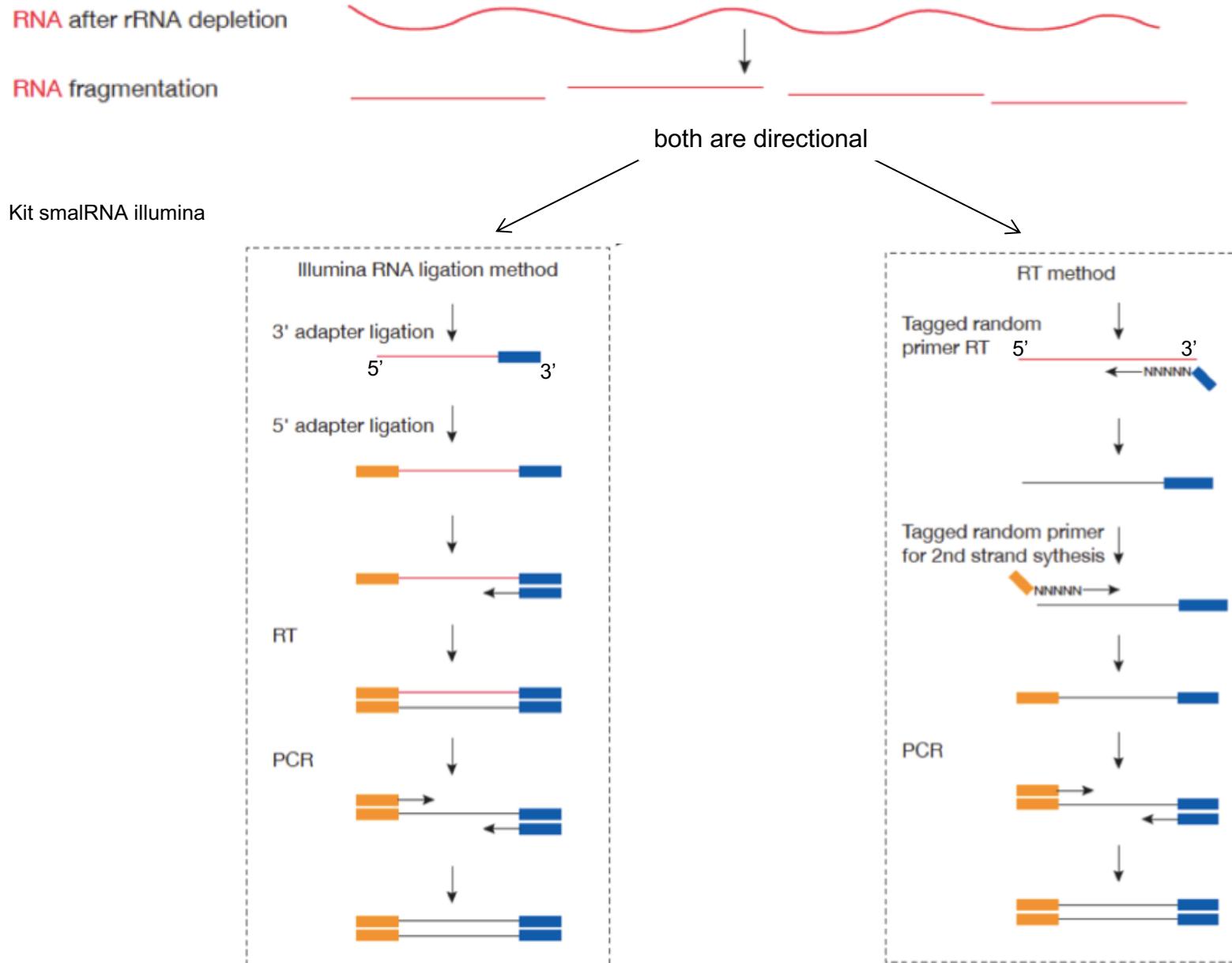
- **Single end:** Sequence one physical end of DNA fragment



- **Paired End:** Sequence both physical ends of DNA fragment
 - End distance: < 800nt



1 - Preparation of RNA-seq Libraries



Banques à partir d'ADN

Protocole

Illumina Truseq (nombreux kits disponibles dans le commerce, chez tous les fournisseurs de produits de bio mol)

PRINCIPE : Ligation d'adaptateurs sur fragments d'ADN

Matériel de départ

ADN génomique fragmenté (fragmentation enzymatique ou mécanique)

Fragments d'ADN double brin (par ex ChIP)

Toutes quantités de 1 à 5000 ng

Avantages

- Adaptable à tout type d'ADN double brin
- Bon contrôle de la taille finales (purif sur gel ou sur billes magnétiques)
- Fonctionne également sans PCR si la quantité de matériel suffisante (>100ng)
- Fonctionne bien même avec des petites quantités

Inconvénients

- Protocole long si fragmentation mécanique
- Possibilité de formation de dimères d'adaptateurs (perte de reads, perte de qualité) ; il existe des kits qui permettent de réduire la formation de dimères

Nextera (kit Nextera Illumina)

PRINCIPE : ajout d'adpatateurs et fragmentation par Tagmentation

ADN génomique, 50 ng (input fixe)

Très rapide (3h)

- Très sensible à qualité de l'ADN de départ (intégrité, pureté)
- Contrôle de la taille des fragments obtenus parfois difficile
- PCR obligatoire
- Extrémités des fragments de l'ADN de départ sont perdues (ex : génome de phage)

Banques à partir d'ARN

Protocole

TruSeq small RNA

(Kit Illumina ou autres fournisseurs (NEB))

TruSeq Stranded RNA

(Kit Illumina, autres fournisseurs possibles)

SMART-Seq V4 (Takara)

Matériel de départ

Petits ARNs pré-purifiés
Peut être utilisé sur de l'ARN fragmenté

100-2000 ng ARN total

ARN total (selection oligodT)
10 pg à 1ng
ARN NON DEGRADÉ

Principe

Ligation directe d'adaptateurs sur ARN
Suivi de Reverse transcription et PCR

- Sélection des messagers par polyA ou déplétion des ARNs ribosomiques par méthode RiboZero (au choix)
- RT par random priming pour obtention d'un cDNA
- Ajout des adaptateurs par ligation
- Enrichissement PCR
Directionnalité conservée

Directionnalité conservée

Synthèse du cDNA à partir de la queue polyA (oligo dT) puis amplification des messagers.
Construction des banques à partir du cDNA obtenu (Nextera ou Truseq)

Directionnalité non conservée

Avantages

- Petites quantités possibles
- Possibilité d'utiliser de l'ARN dégradé avec l'option RiboZero

- RNA-seq possible même si très petites quantités d'ARN total.
- cDNA longs

Inconvénients

- Long : 2-3 jours de manip
- Peu adapté aux messagers car faible rendement

- Sensible à contamination par ADN génomique

- ARN non dégradé seulement
- Coûteux

Remarques

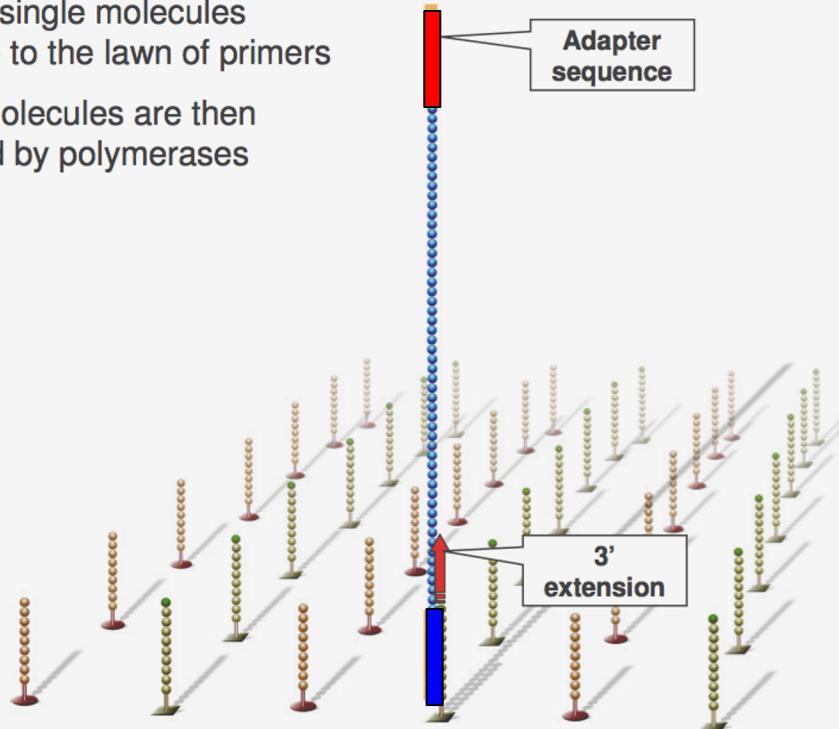
N'est plus utilisé à la PF pour le RNAseq classique

Peut être utilisé pour d'autres applications, comme la synthèse de cDNA longs pour séquençage nanopore

2 – Cluster growth

Cluster Generation: *Hybridize Fragment & Extend*

- ▶ > 100 M single molecules hybridize to the lawn of primers
- ▶ Bound molecules are then extended by polymerases



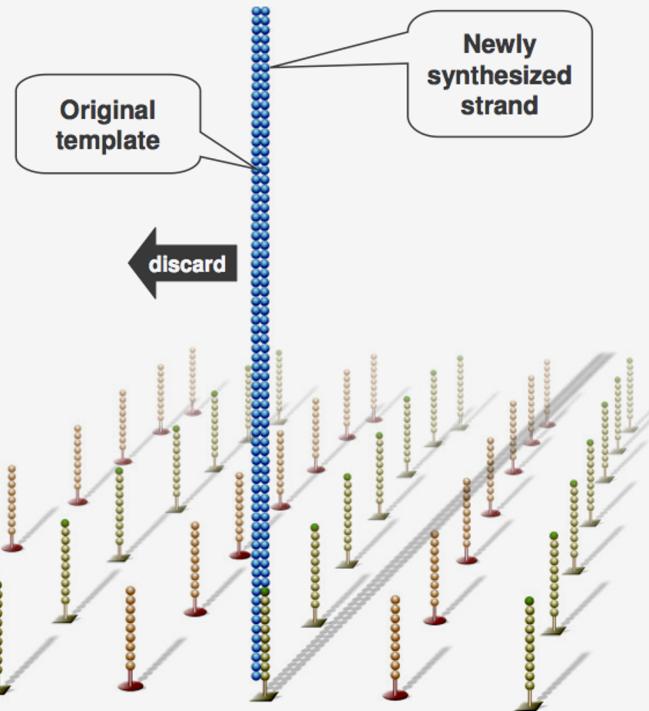
2 – Cluster growth

Cluster generation: *Denature double-stranded DNA*

Double-stranded
molecule is
denatured.

Original template
is washed away.

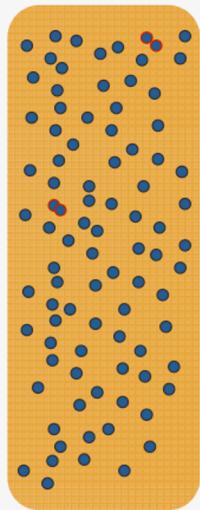
Newly synthesized
covalently
attached to the
flow cell surface.



2 – Cluster growth

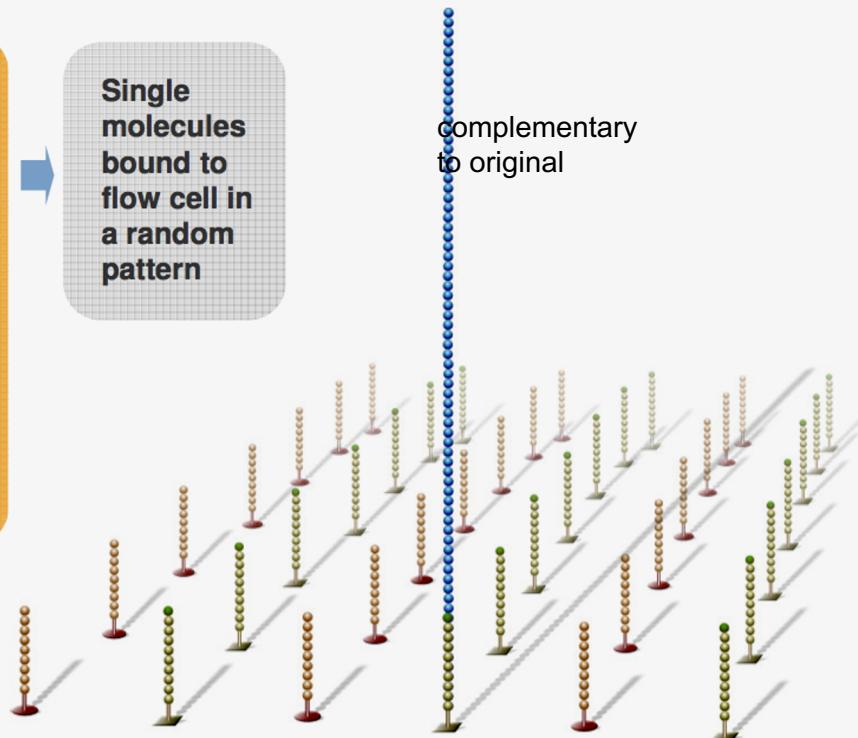
Cluster generation:

Covalently bound spatially separated single molecules



Single molecules bound to flow cell in a random pattern

complementary to original

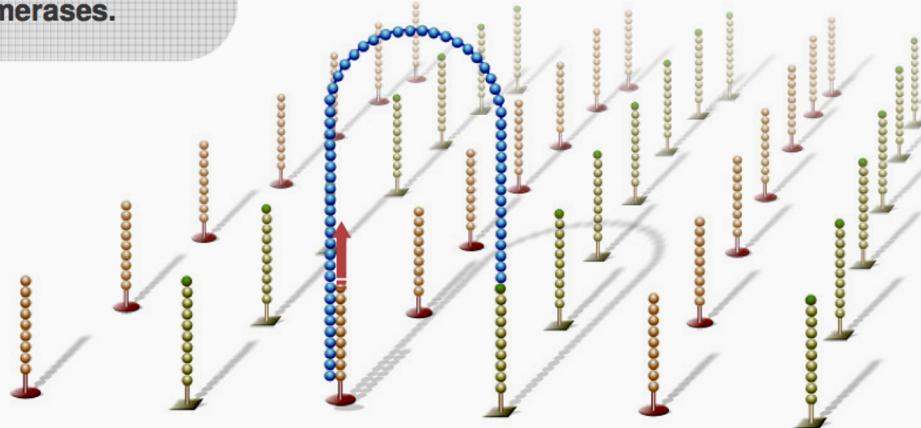


2 – Cluster growth

Cluster generation: *Bridge amplification*

Single-strand flips over to hybridize to adjacent primers to form a bridge.

Hybridized primer is extended by polymerases.

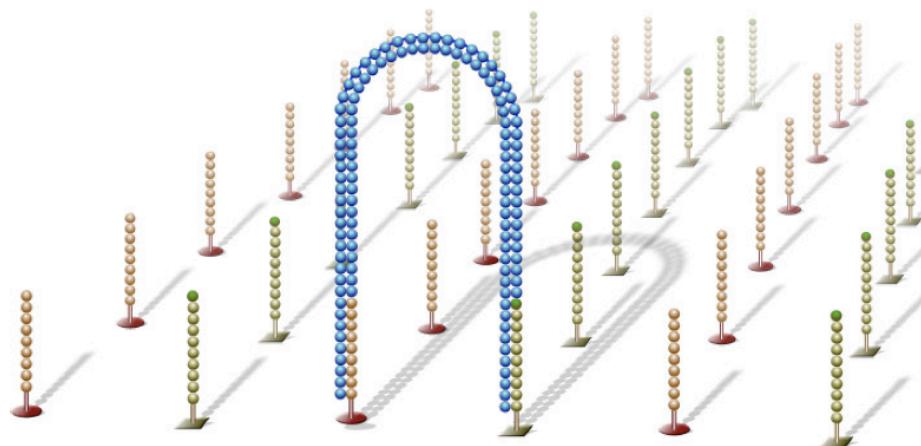


2 – Cluster growth

Cluster generation:

Bridge amplification

→ double-stranded
bridge is formed.



2 – Cluster growth

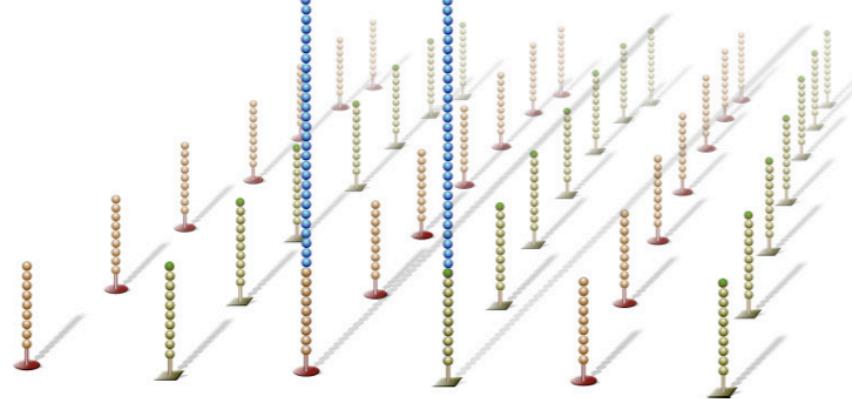
Cluster generation: *Bridge amplification*

Double-stranded bridge
is denatured.

Result: Two copies of
covalently bound single-
stranded templates.

identical
to original

complementary
to original



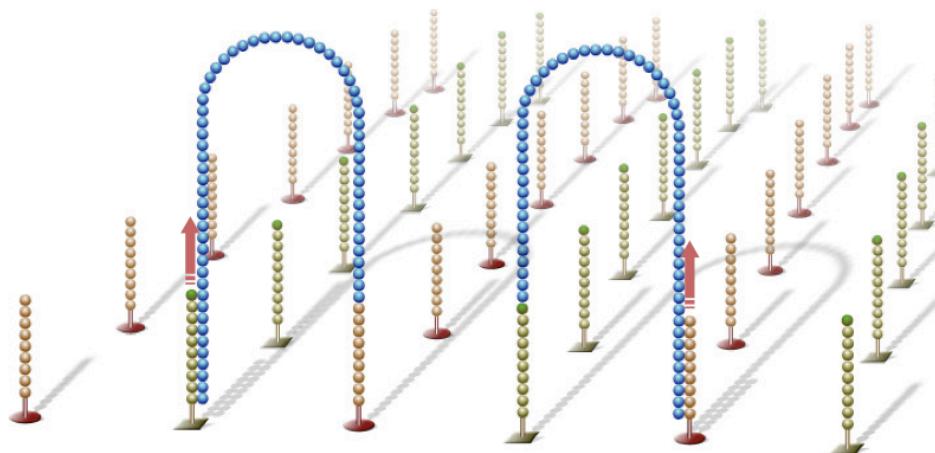
i

2 – Cluster growth

Cluster generation: *Bridge amplification*

Single-strands flip over
to hybridize to adjacent
primers to form bridges.

Hybridized primer is
extended by polymerase.

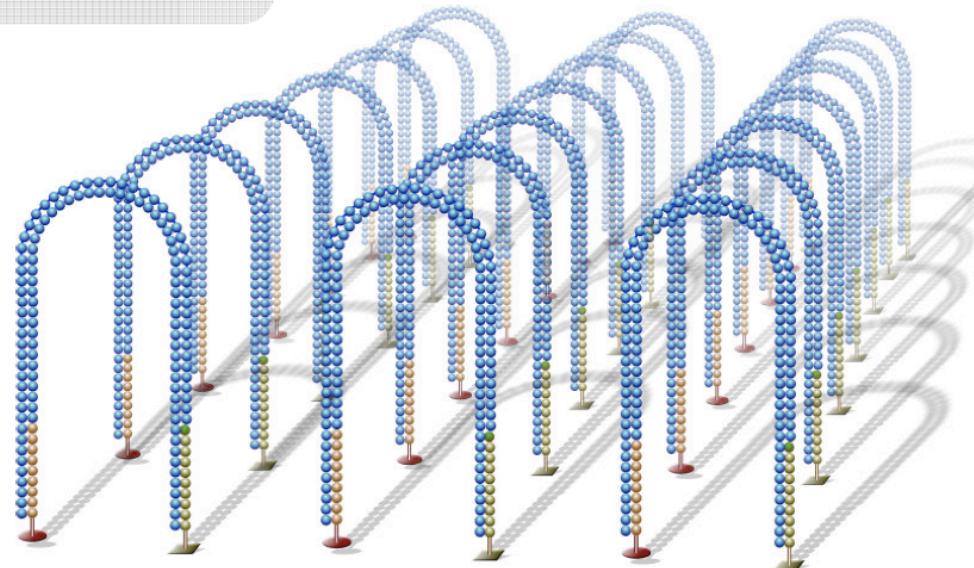


2 – Cluster growth

Cluster generation:

Bridge amplification

Bridge amplification
cycle repeated till
multiple bridges
are formed

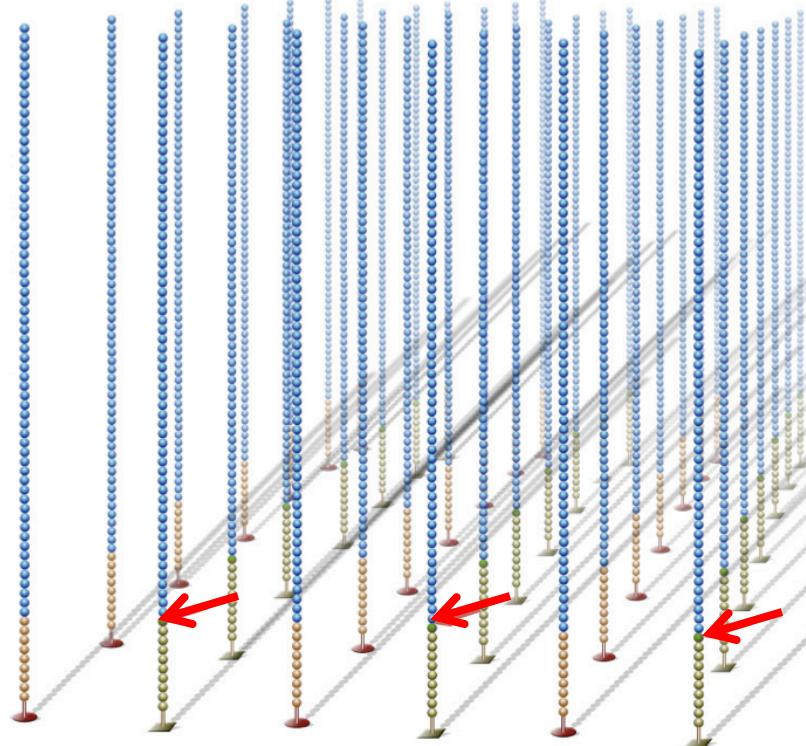


2 – Cluster growth

Cluster generation

dsDNA
bridges
denatured.

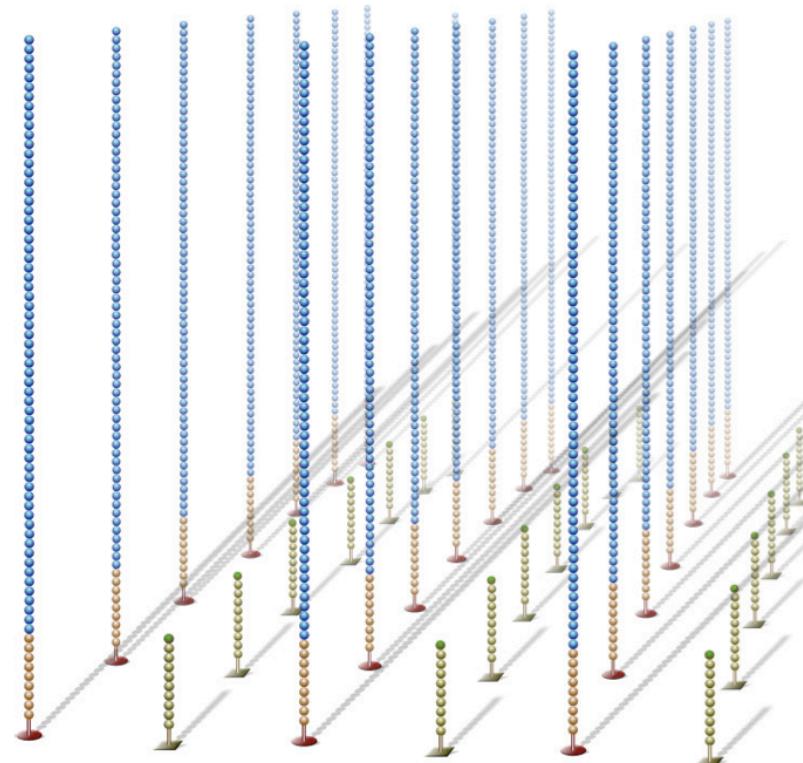
Reverse
strands
cleaved
and
washed
away.



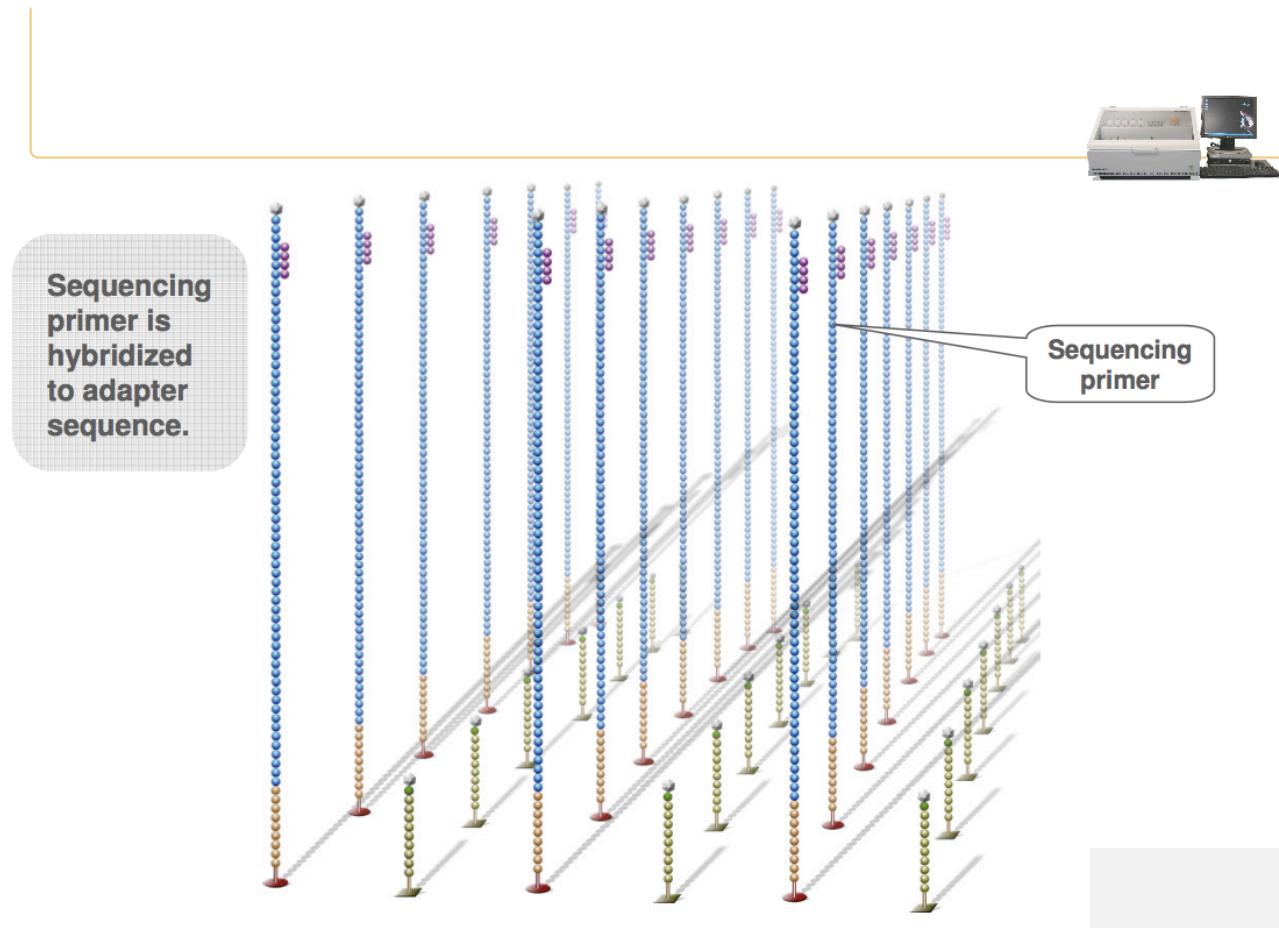
2 – Cluster growth

Cluster generation

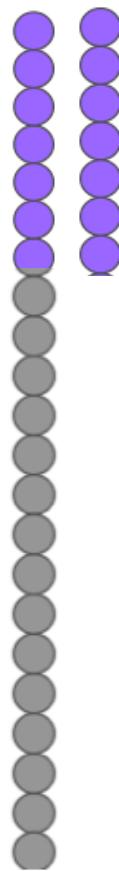
... leaving
a cluster
with forward
strands only.



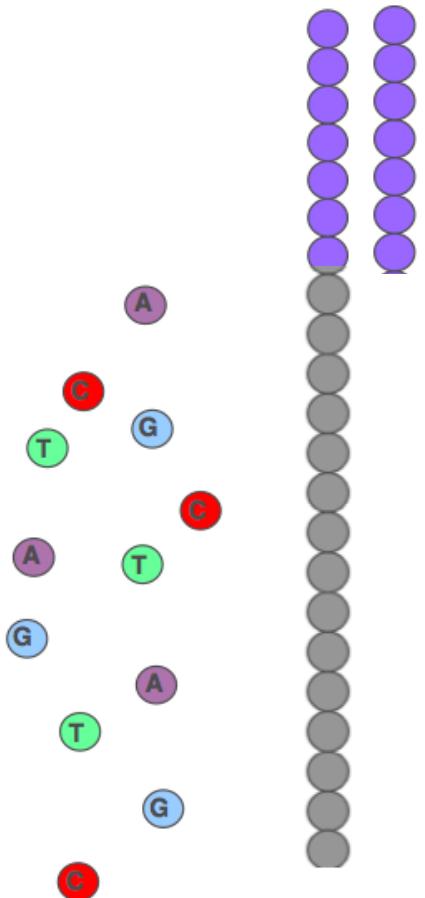
3 – Sequencing



3 - Sequencing By Synthesis (SBS)

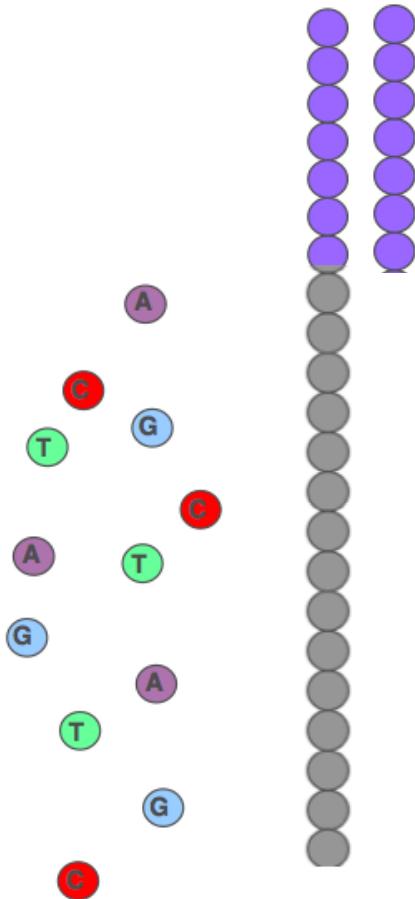


3 - Sequencing By Synthesis (SBS)



fluorescent-labeled terminator bound to each dNTP

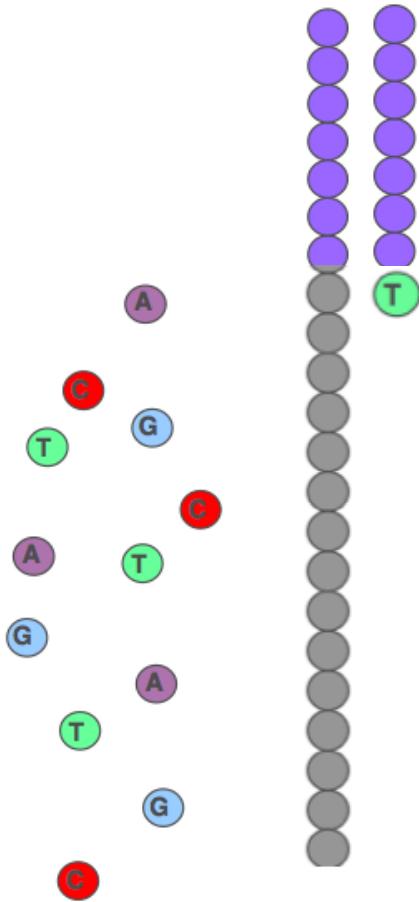
3 - Sequencing By Synthesis (SBS)



fluorescent-labeled terminator bound to each dNTP

Cycle 1 : add sequencing reagents

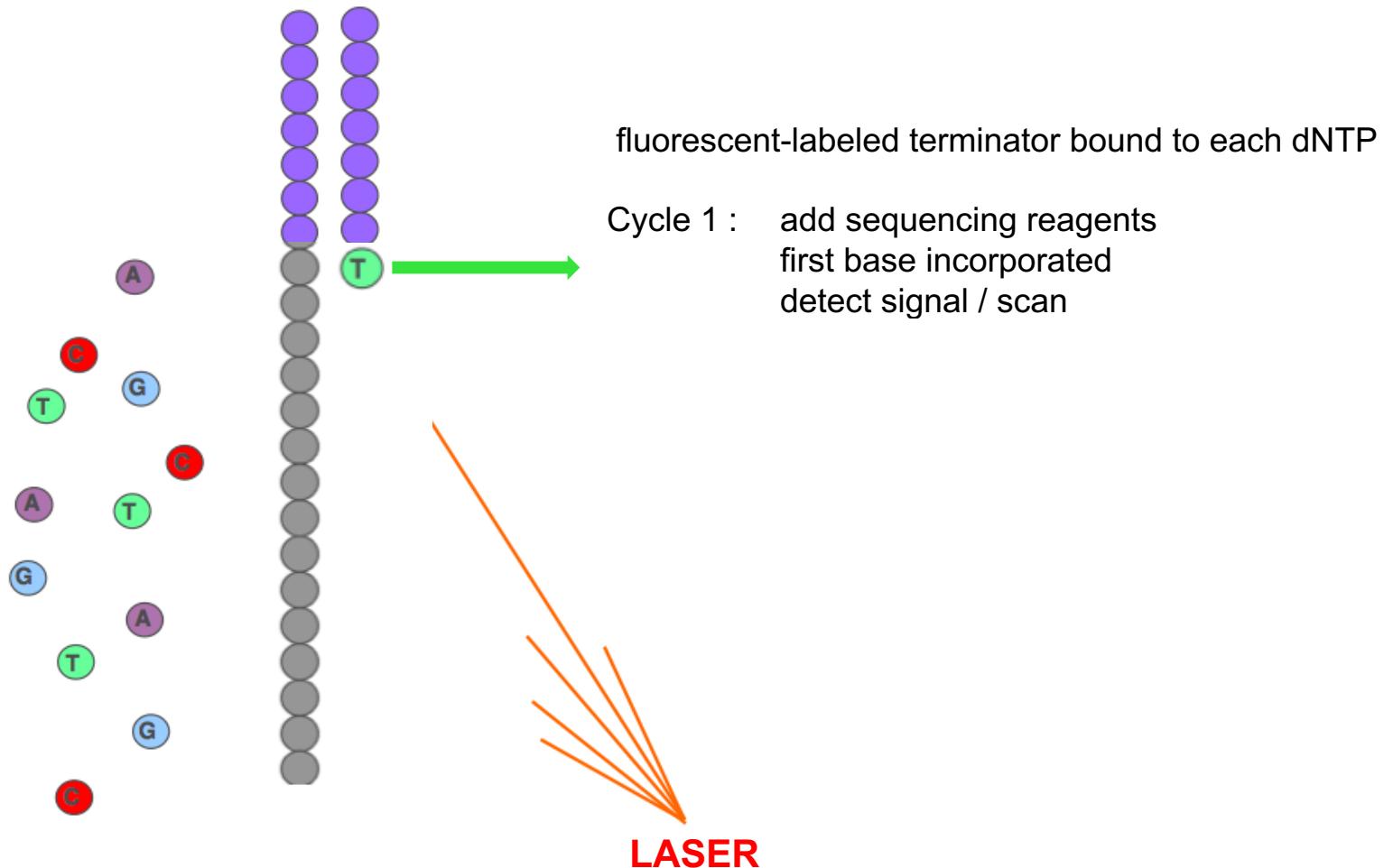
3 - Sequencing By Synthesis (SBS)



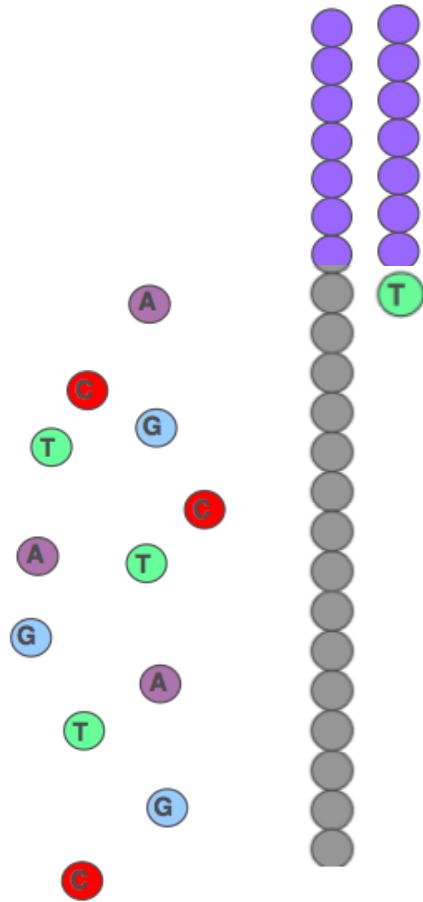
fluorescent-labeled terminator bound to each dNTP

Cycle 1 : add sequencing reagents
first base incorporated

3 - Sequencing By Synthesis (SBS)



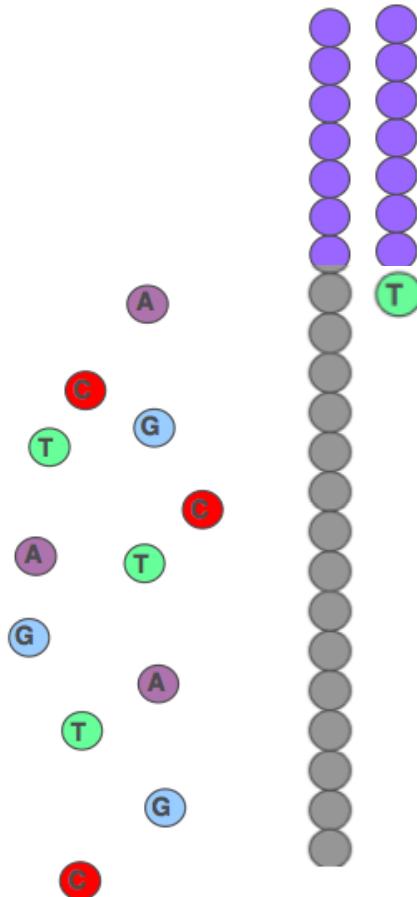
3 - Sequencing By Synthesis (SBS)



fluorescent-labeled terminator bound to each dNTP

Cycle 1 : add sequencing reagents
first base incorporated
detect signal / scan
cleave terminator and dye

3 - Sequencing By Synthesis (SBS)

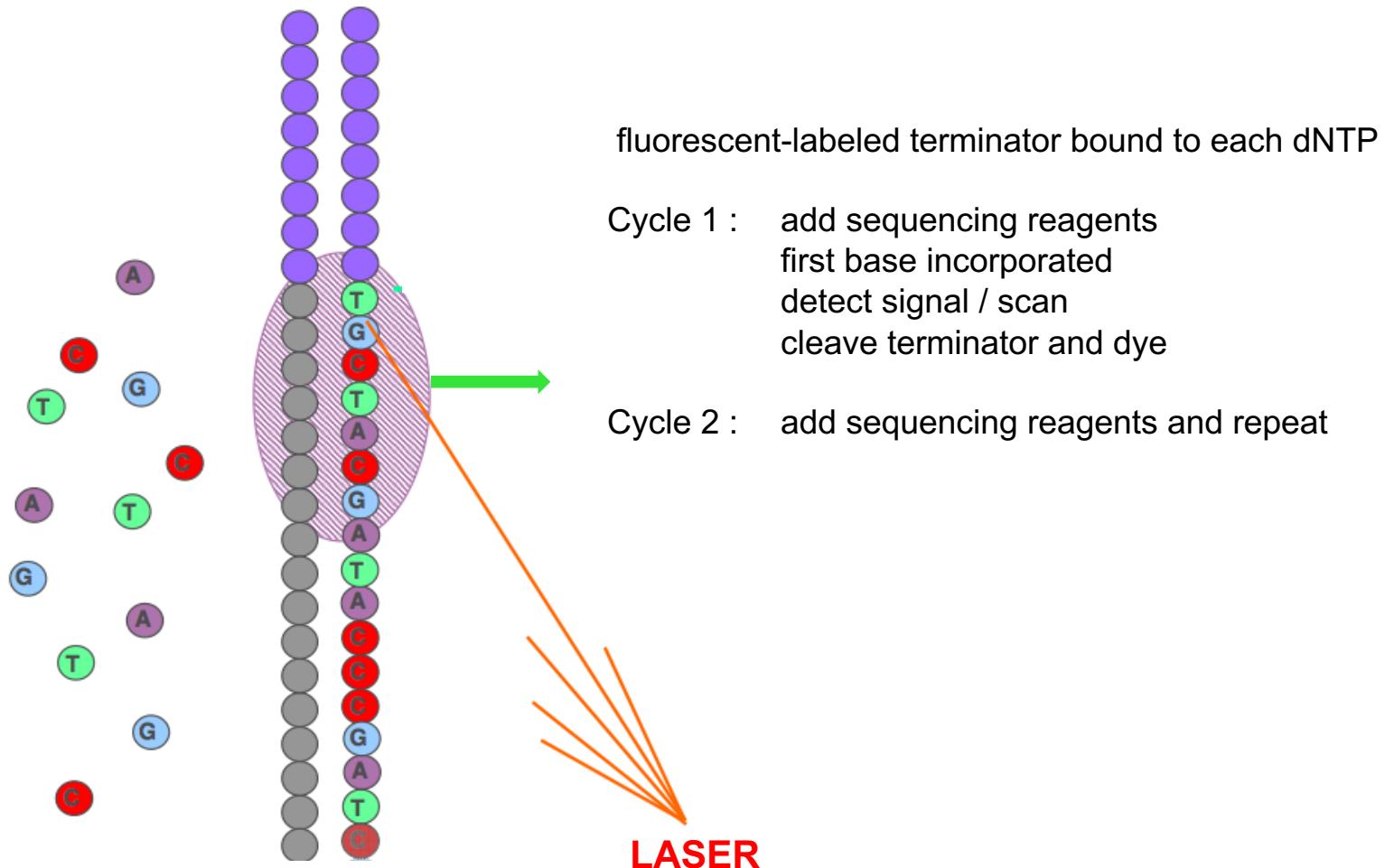


fluorescent-labeled terminator bound to each dNTP

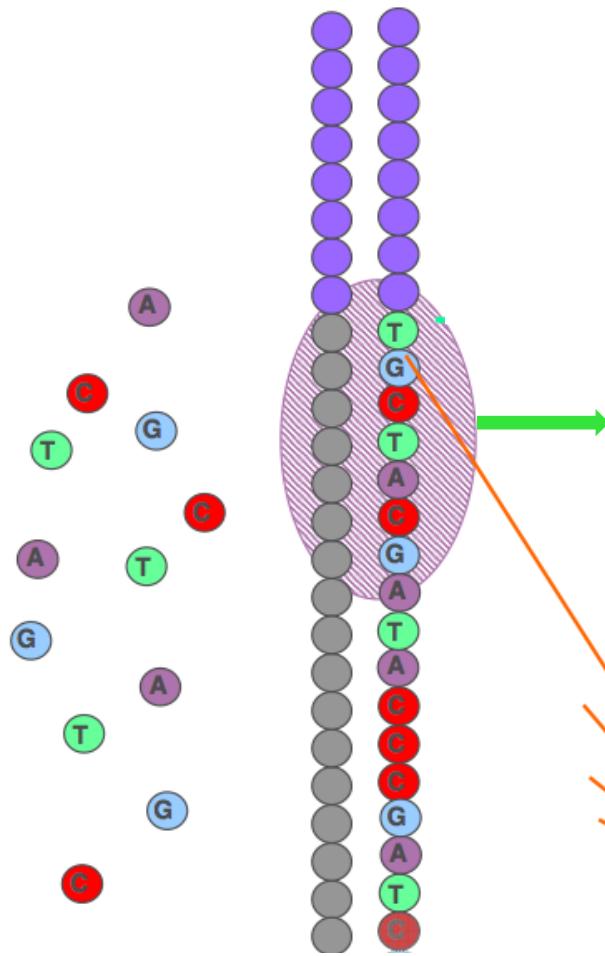
Cycle 1 : add sequencing reagents
first base incorporated
detect signal / scan
cleave terminator and dye

Cycle 2 : add sequencing reagents and repeat

3 - Sequencing By Synthesis (SBS)



3 - Sequencing By Synthesis (SBS)



fluorescent-labeled terminator bound to each dNTP

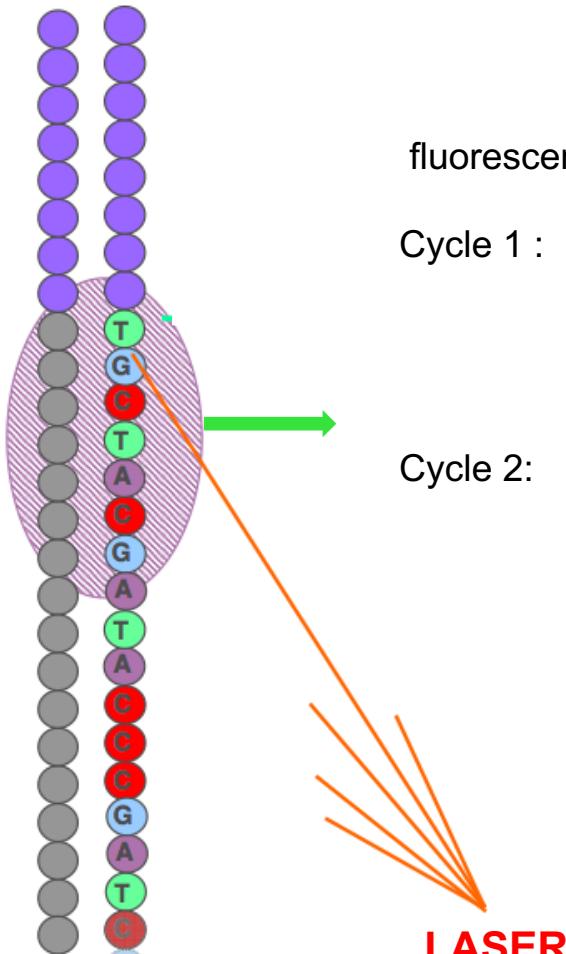
Cycle 1 : add sequencing reagents
first base incorporated
detect signal / scan
cleave terminator and dye

Cycle 2 : add sequencing reagents and repeat

4 terminator-bound dNTPs present during each cycle

natural competition minimizes incorporation bias

3 - Sequencing By Synthesis (SBS)



fluorescent-labeled terminator bound to each dNTP

Cycle 1 : add sequencing reagents
first base incorporated
detect signal / scan
cleave terminator and dye

Cycle 2: add sequencing reagents and repeat

$$(C_{\text{eff}})^{\text{RL}} = 0.5$$

Cycle Efficiency (%)	Read-length (bases)
90.0	7
95.0	14
98.0	35
99.0	69
99.5	149
99.9	693