Next Generation Sequencing Platforms

Speaker: Msc. Ngo Dai Phu
PI: PhD. Luu Phuc Loi

History and future of DNA sequencing

- 1953: Discovery of DNA structure by Watson and Crick
- 1967: First DNA sequence of 11 bp published (20 pages)

History and future of DNA sequencing

- 1953: Discovery of DNA structure by Watson and Crick
- 1967: First DNA sequence of 11 bp published (20 pages)

J. Mol. Biol. (1967) **30**, 507–527

Studies on the Bacteriophage MS2

IV†. The 3′-OH Terminal Undecanucleotide Sequence of the Viral RNA Chain

R. De Wachter and W. Fiers

Laboratory of Physiological Chemistry, State University of Ghent, Belgium

(*Received 1 May 1967, and in revised form 29 July 1967*)

The 3′-OH terminus of bacteriophage MS2 RNA was selectively labelled with ^{3}H. This was achieved by oxidation of the free 2′, 3′-diol group with sodium periodate to a dialdehyde, and reduction of the latter with tritiated sodium borohydride. Treatment of this RNA with alkali and separation of the hydrolysis products firmed each other. The results, together with the known specificity of the ribonuclease T_1, which had released the sequence, establish that MS2 RNA ends in ...GpUpUpApCpCpApCpCpCpA.

It is suggested that the termination signal for the translation into polypeptides

1. Introduction

Apart from several transfer RNA's, little is known about the primary structure of macromolecular RNA's. Particularly, one would like to gain information on the beginning and on the ending of a messenger RNA, as this might possibly be related to genetic signals for polypeptide chain initiation and termination. Viral RNA, although not a typical messenger *sensu stricto*, behaves nevertheless in many respects as a simple, polycistronic message. Sugiyama & Fraenkel-Conrat (1961) identified the 3′-OH terminal nucleoside of tobacco mosaic virus RNA as adenosine. Subsequently,

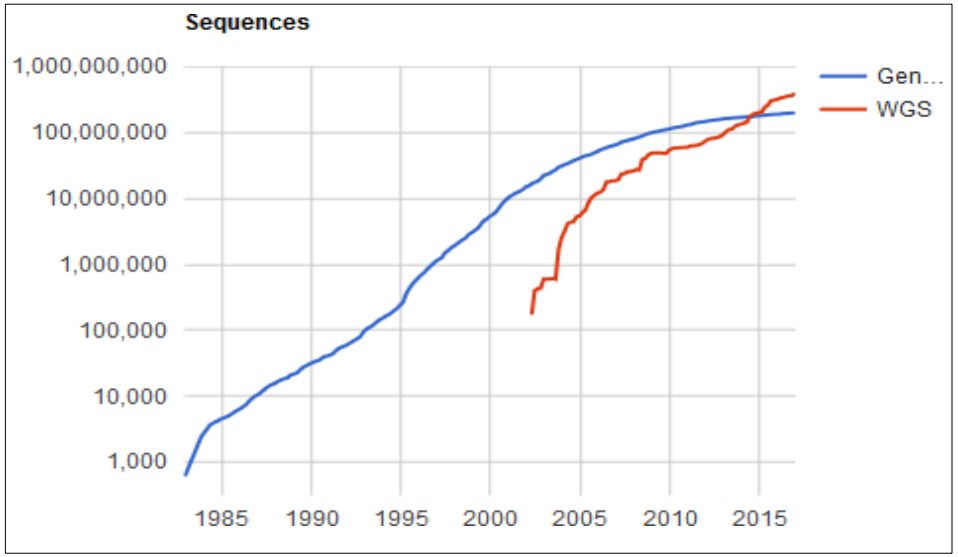
History and future of DNA sequencing

1953: Discovery of DNA structure by Watson and Crick

1967: First DNA sequence of 11 bp published (20 pages)
1976: First genome sequenced: Bacteriophage MS2 (3569 bp) by Walter Fiers
1977: Sanger sequencing method published
1980: Nobel Prize Wally Gilbert and Fred Sanger
1982: Genbank started
1983: development of PCR

1996: Capillary sequencer: ABI 310

2003: Human genome sequenced



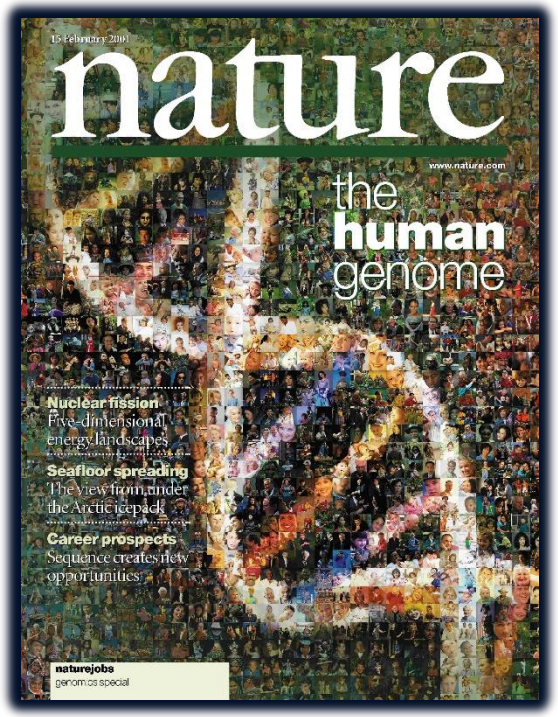
		GenBank		WGS	
Release	Date	Bases	Sequences	Bases	Sequences
3	Dec 1982	680338	606		
14	Nov 1983	2274029	2427		
20	May 1984	3002088	3665		
216	Oct 2016	220731315250	197390691	1676238489250	363213315
217	Dec 2016	224973060433	198565475	1817189565845	395301176

History and future of DNA sequencing

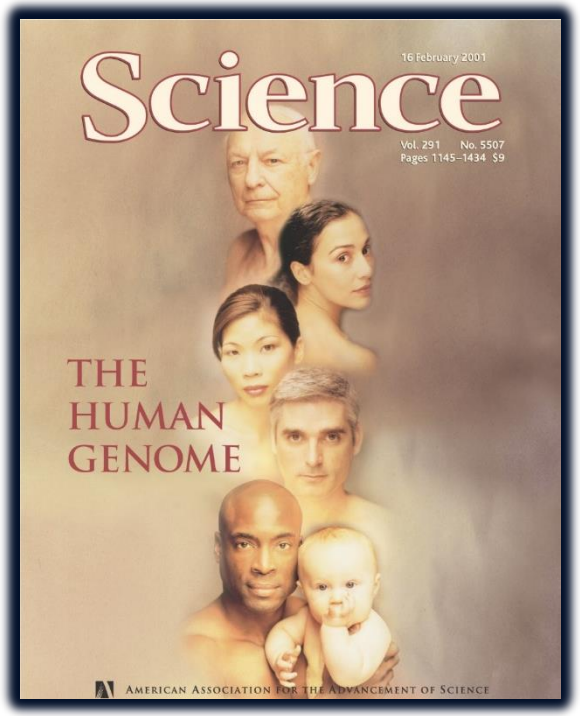
Dự án hệ gen người HGP (Oct 1990 - April 2003)

- Năm 2003, dự án hệ gen người HGP đã tạo ra một chuỗi trình tự gần 3 tỉ nucleotide chiếm hơn 90% bộ gen người.
- Đây là chuỗi gen hoàn chỉnh nhất có thể đạt được với công nghệ giải trình tự DNA thời điểm đầu những năm 2000.

=> Sự phát triển của công nghệ giải trình tự thế hệ mới (NGS).



HGP Paper



Venter/Celera Paper

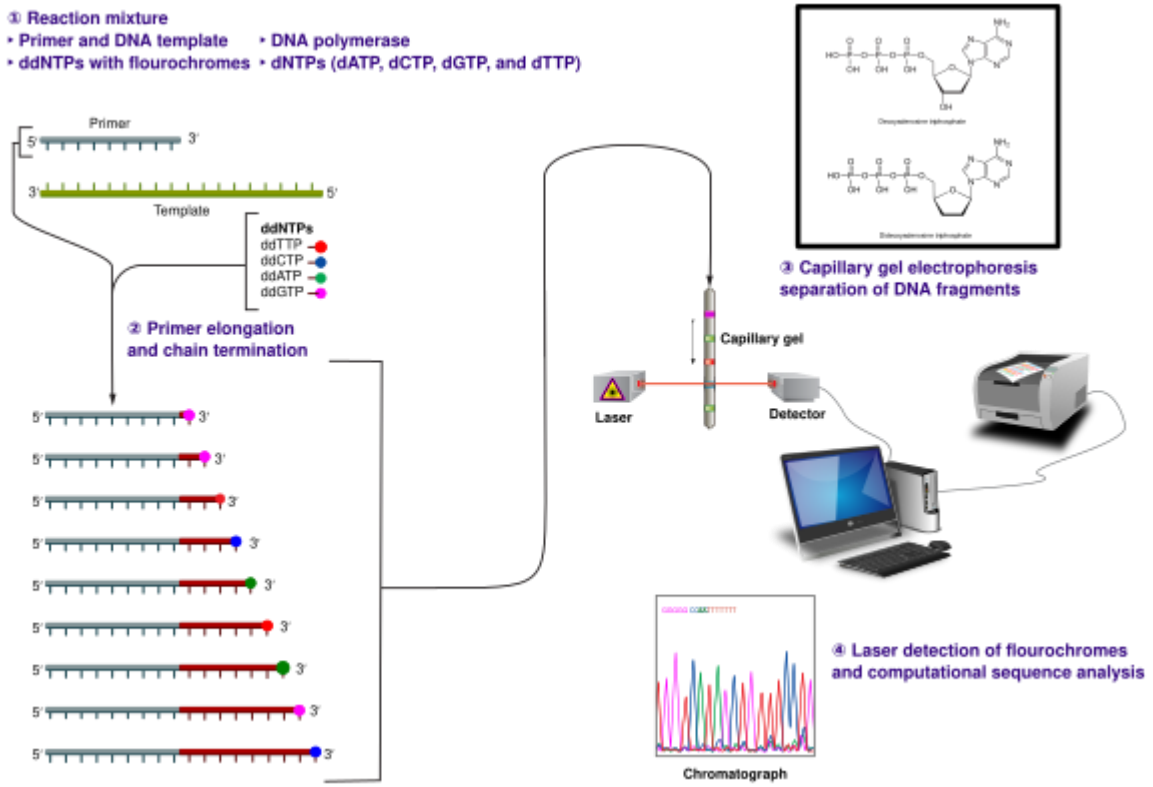
History and future of DNA sequencing

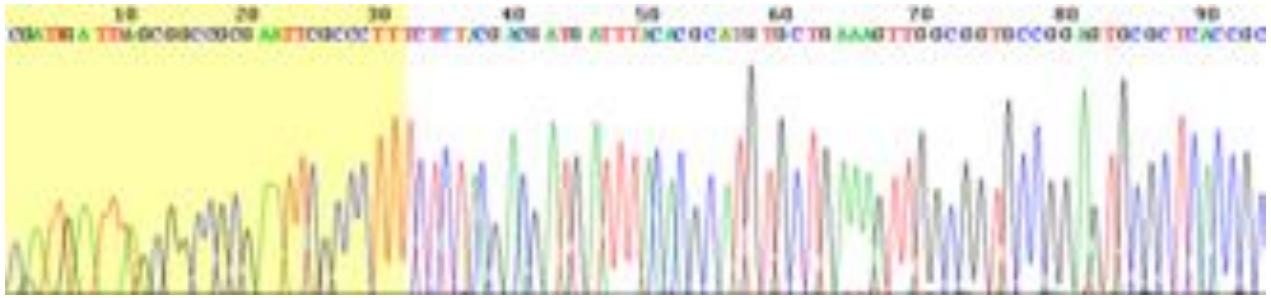
- 1953: Discovery of DNA structure by Watson and Crick
- 1967: First DNA sequence of 11 bp published (20 pages)
- 1976: First genome sequenced: Bacteriophage MS2 (3569 bp) by Walter Fiers
- 1977: Sanger sequencing method published
- 1980: Nobel Prize Wally Gilbert and Fred Sanger
- 1982: Genbank started
- 1983: development of PCR
- 1987: 1st automated sequencer: Applied Biosystems Prism 373
- 1996: Capillary sequencer: ABI 310
- 1998: Genome of Caenorhabditis elegans sequenced (100 million bp)
- 2003: Human genome sequenced (3,2 billion bp)
- 2005: 1st 454 Life Sciences Next Generation Sequencing system: GS 20 system († mid 2016)
- 2006: 1st Solexa Next Generation Sequencer: Genome Analyzer (Illumina)
- 2007: 1st Applied Biosystems Next Generation Sequencer: SOLiD († Dec 2017)
- 2009: 1st Helicos single molecule sequencer: Helicos Genetic Analyser System († Nov 2012)
- 2011: 1st Ion Torrent Next Generation Sequencer: PGM
  1st Pacific Biosciences single molecule sequencer: PacBio RS Systems
- 2012: Oxford Nanopore Technologies demonstrates ultra long single molecule reads
- 2014: Roche acquires Genia: development of NanoTag single molecule sequencing
- 2015: 1st BGI Next Generation Sequencer: BGISEQ-500 (sold in China only)
- 2016: 1st Oxford Nanopore Technologies sequencer: MinION
- 2017: SeqLL announces tSMS sequencer: single molecule (Helicos technology)

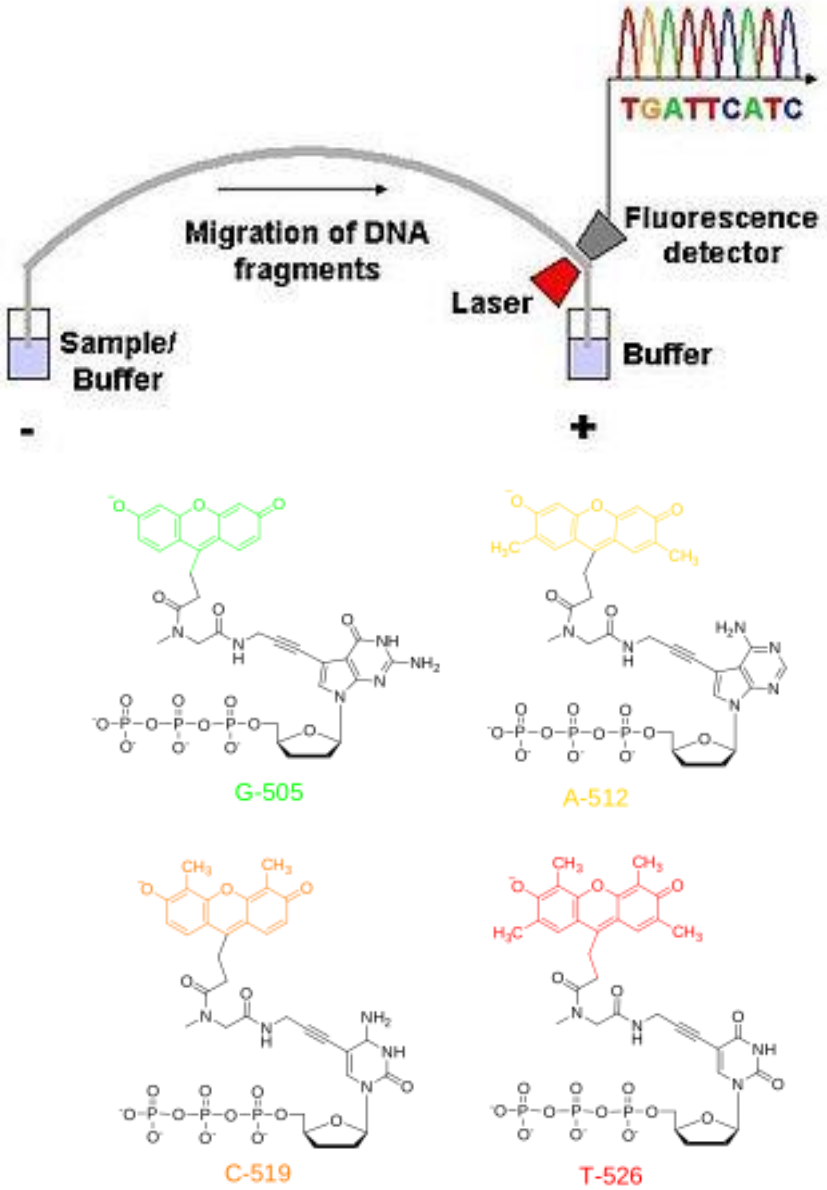
What is Next Generation Sequencing? For what does ‘Next’ stand?

History and future of DNA sequencing

What is Next Generation Sequencing?



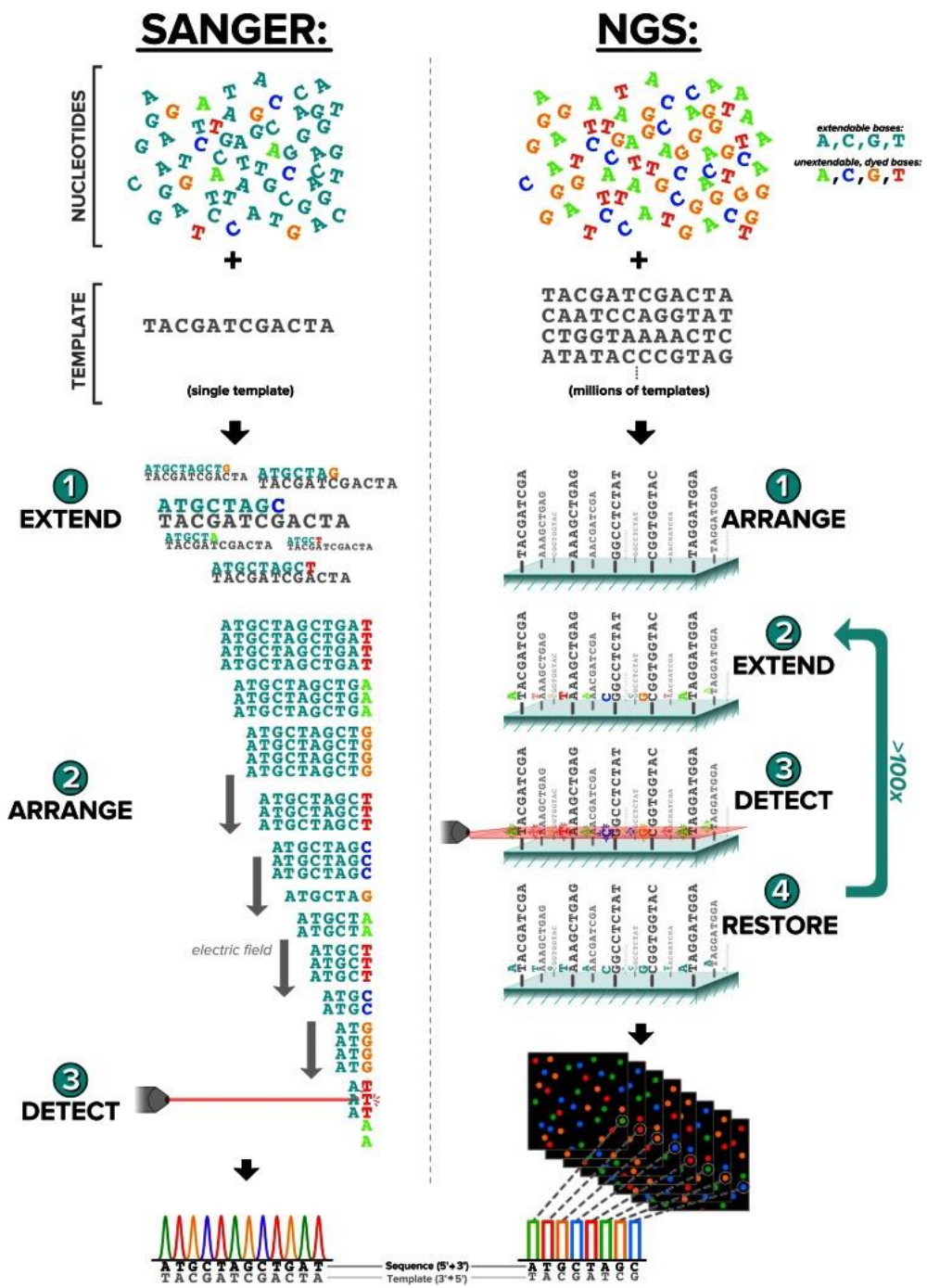




https://en.wikipedia.org/wiki/Sanger_sequencing

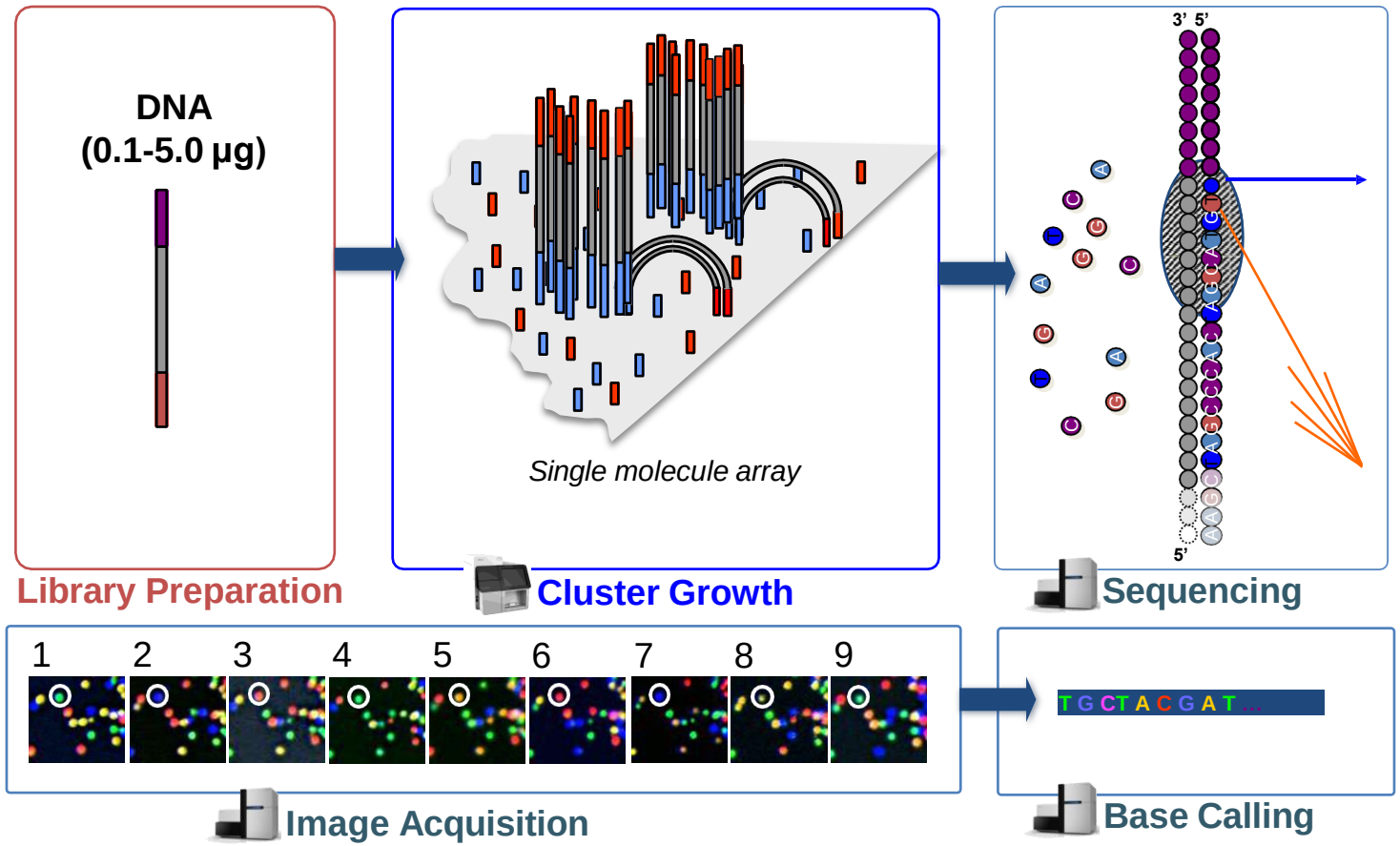
Giải trình tự gen thế hệ mới (NGS): giải trình tự song song

Read1:CTCGAATACG



Read1:CTCGAATACG
Read2:CTCGAATACG
Read3:CTCGAATACG
Read4:CTCGAATACG
Read5:CGCGAATACG
Read6:CGCGAATACG
Read7:CGCGACTACG
Read8:CGCGAATACG

10.1007/s40142-015-0076-8



Consider Cluster Size in Multiplexing

History and future of DNA sequencing

- 1953: Discovery of DNA structure by Watson and Crick
- 1967: First DNA sequence of 11 bp published (20 pages)
- 1976: First genome sequenced: Bacteriophage MS2 (3569 bp) by Walter Fiers
- 1977: Sanger sequencing method published
- 1980: Nobel Prize Wally Gilbert and Fred Sanger
- 1982: Genbank started
- 1983: development of PCR
- 1987: 1st automated sequencer: Applied Biosystems Prism 373
- 1996: Capillary sequencer: ABI 310
- 1998: Genome of Caenorhabditis elegans sequenced (100 million bp)
- 2003: Human genome sequenced (3,2 billion bp)
- 2005: 1st 454 Life Sciences Next Generation Sequencing system: GS 20 system († mid 2016)
- 2006: 1st Solexa Next Generation Sequencer: Genome Analyzer (Illumina)
- 2007: 1st Applied Biosystems Next Generation Sequencer: SOLiD († Dec 2017)
- 2009: 1st Helicos single molecule sequencer: Helicos Genetic Analyser System († Nov 2012)
- 2011: 1st Ion Torrent Next Generation Sequencer: PGM
  1st Pacific Biosciences single molecule sequencer: PacBio RS Systems
- 2012: Oxford Nanopore Technologies demonstrates ultra long single molecule reads
- 2014: Roche acquires Genia: development of NanoTag single molecule sequencing
- 2015: 1st BGI Next Generation Sequencer: BGISEQ-500 (sold in China only)
- 2016: 1st Oxford Nanopore Technologies sequencer: MinION
- 2017: SeqLL announces tSMS sequencer: single molecule (Helicos technology)

Different platforms

- Illumina (Solexa)
  - iSeq 100
  - MiniSeq
  - MiSeq
  - NextSeq 500 - 550
  - HiSeq 2500 - 3000 – 4000
  - HiSeq X Five – Ten
  - NovaSeq 5000 - 6000
- Thermo Fisher Scientific (Applied Biosystems -> Life Technologies)
  - Ion Torrent Personal Genome Machine (PGM)
  - Ion Torrent GeneStudio S5, S5 Plus, S5 Prime
  - Ion Torrent Proton

Next Generation Sequencing
Amplified Single Molecule Sequencing

- Pacific Biosciences
  - Sequel System
  - PacBio RS II
- Oxford Nanopore Technologies
  - SmidgION
  - MinION
  - GridION X5
  - PromethION
- SeqLL Elements Biosciences
  - tSMS sequencer

Third Generation Sequencing,
Next Next Generation Sequencing,
Single Molecule Sequencing

Different platforms

- Illumina (Solexa)
- GeneMind
- Thermo Fisher Scientific (Applied Biosystems -> Life Technologies)
- MGI

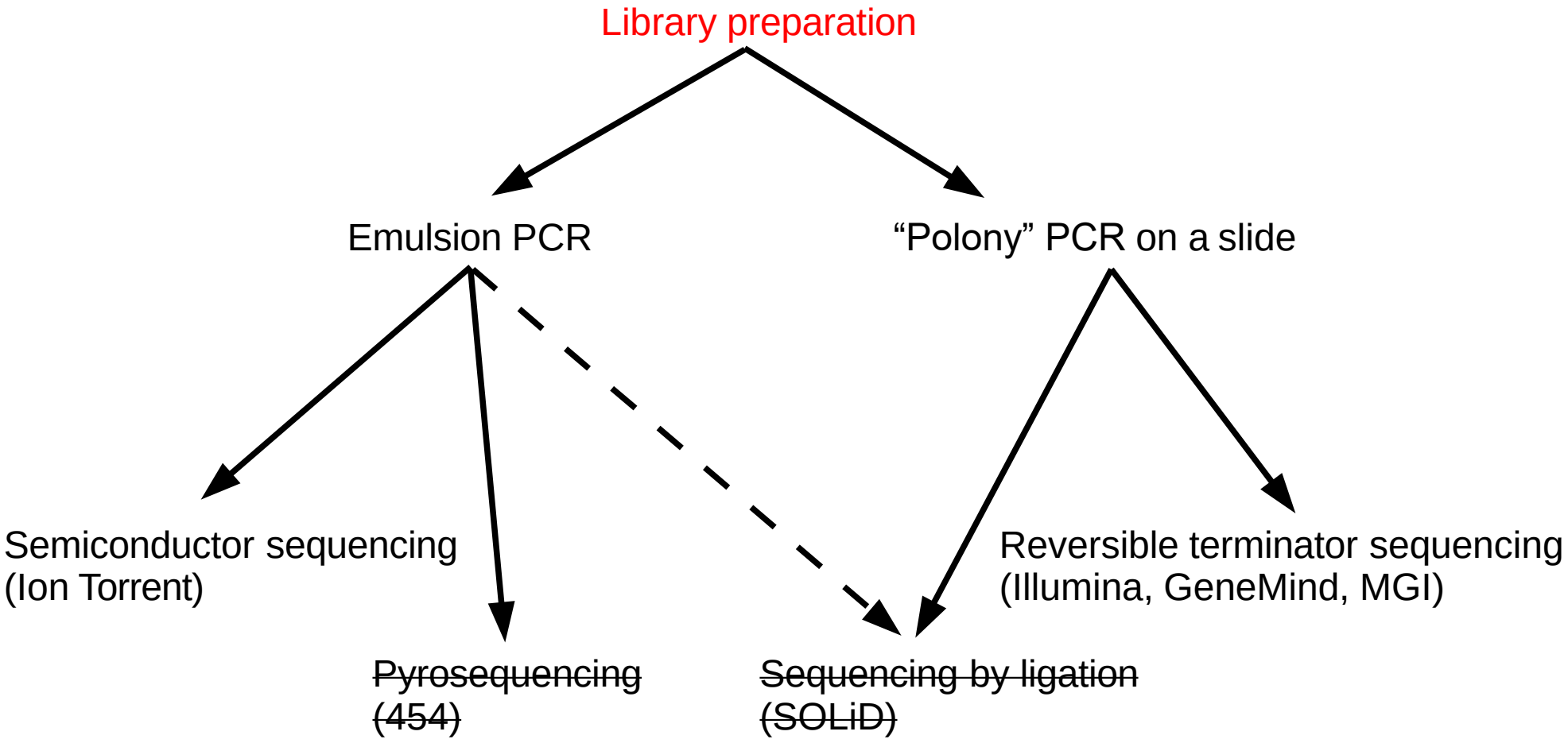
Next Generation Sequencing
Amplified Single Molecule Sequencing

- Pacific Biosciences
  - Sequel System
  - PacBio RS II
- Oxford Nanopore Technologies
  - SmidgION
  - MinION
  - GridION X5
  - PromethION
- SeqLL Elements Biosciences
  - tSMS sequencer

Third Generation Sequencing,
Next Next Generation Sequencing,
Single Molecule Sequencing

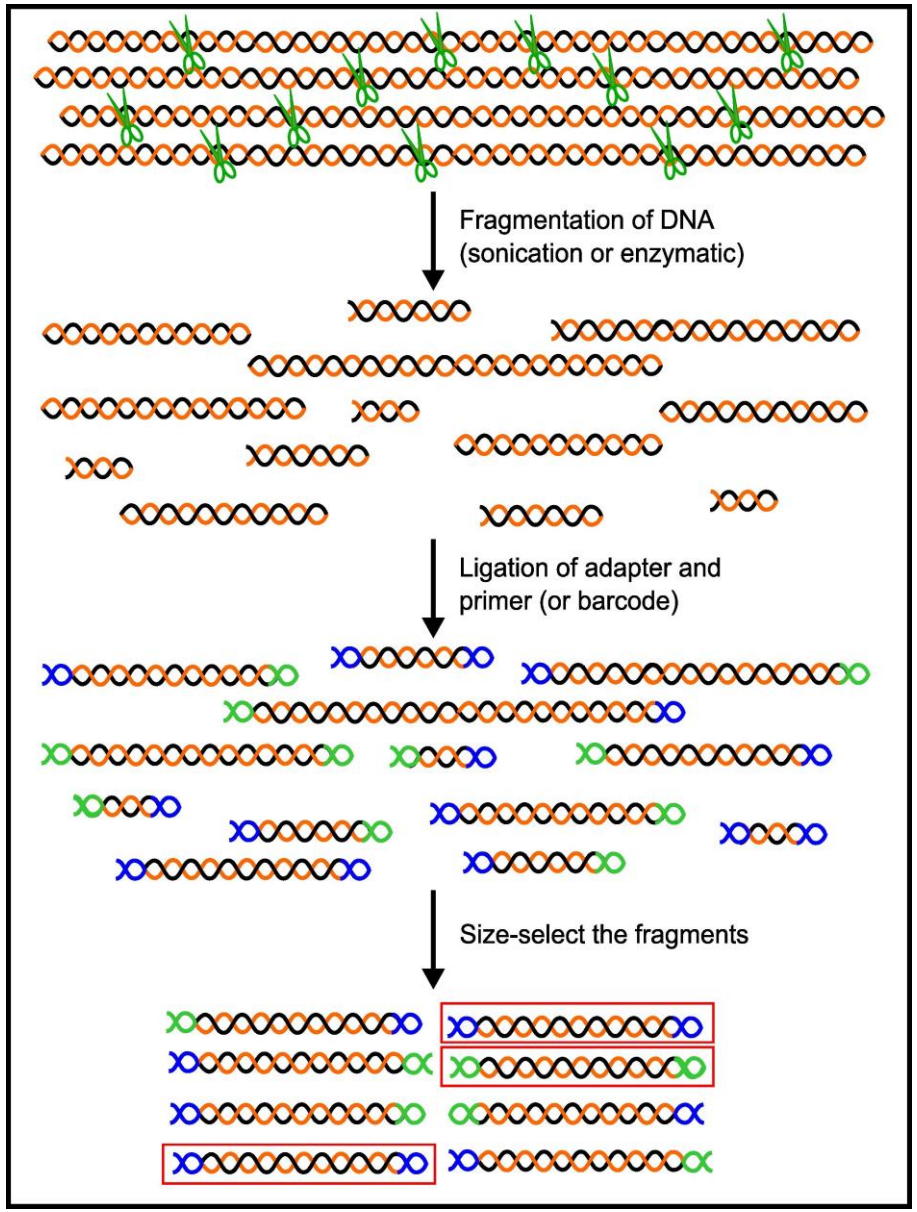
General Workflow

Next Generation Sequencing: Amplified Single Molecule Sequencing



General Workflow

Next Generation Sequencing: Amplified Single Molecule Sequencing

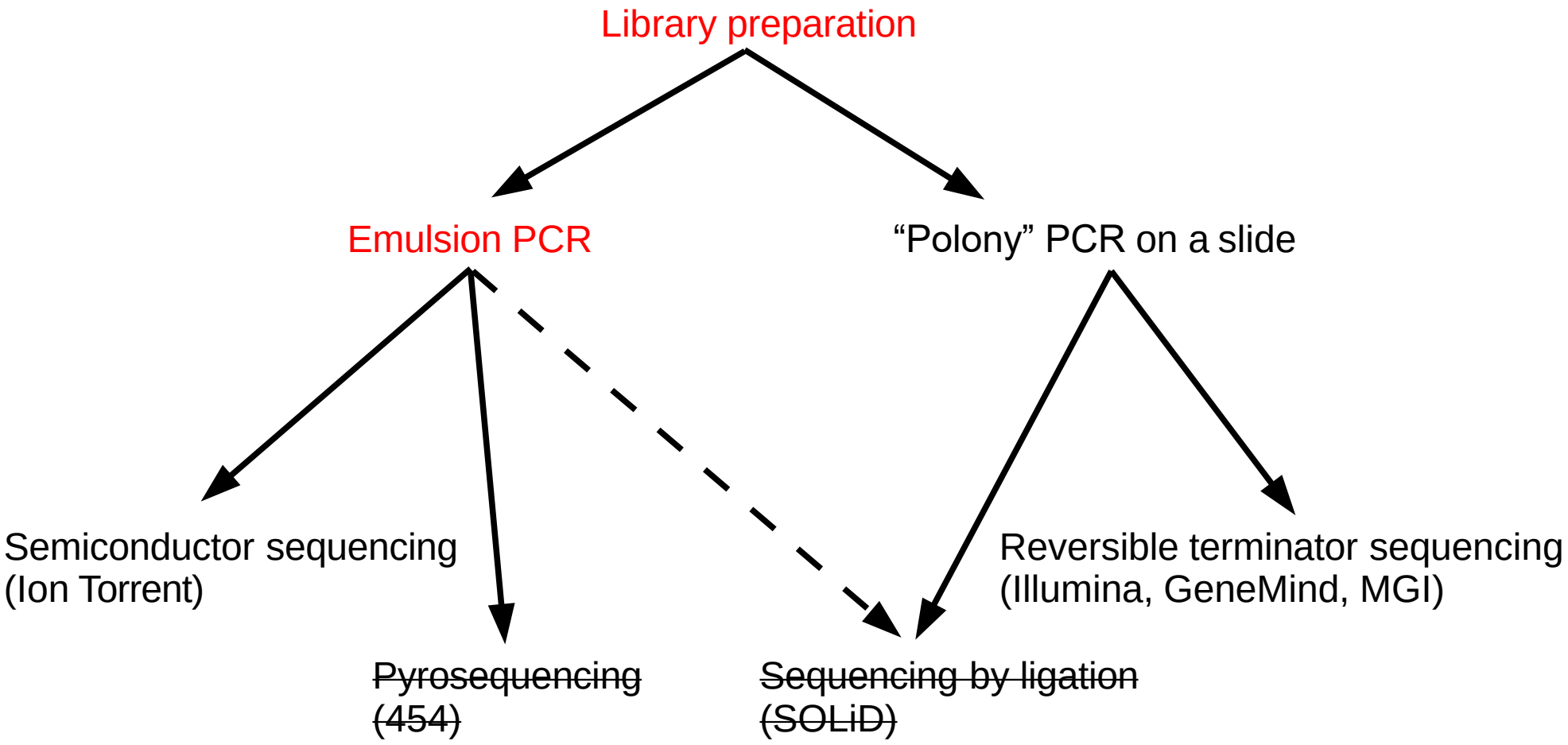


Library preparation

Good fragments:

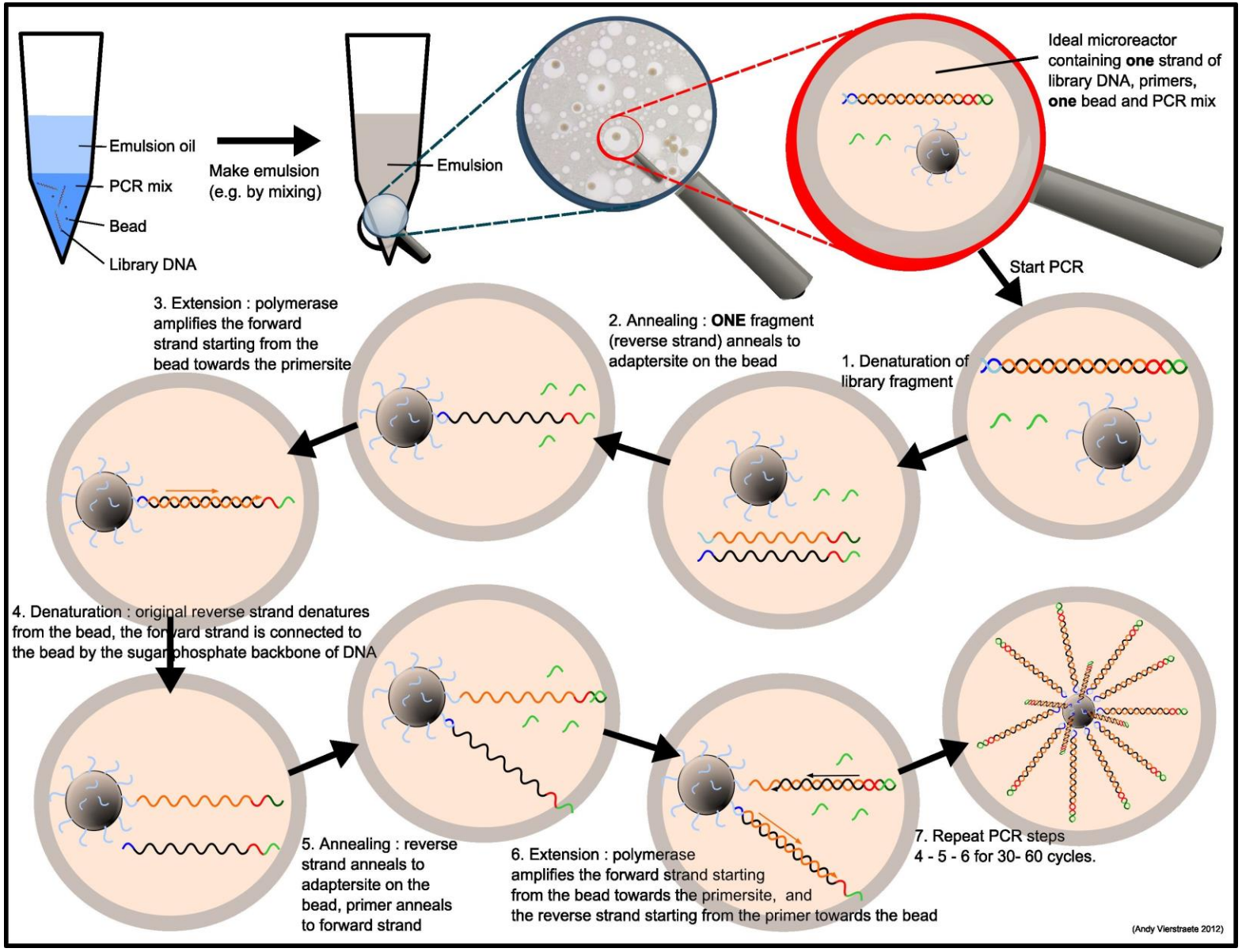
General Workflow

Next Generation Sequencing: Amplified Single Molecule Sequencing



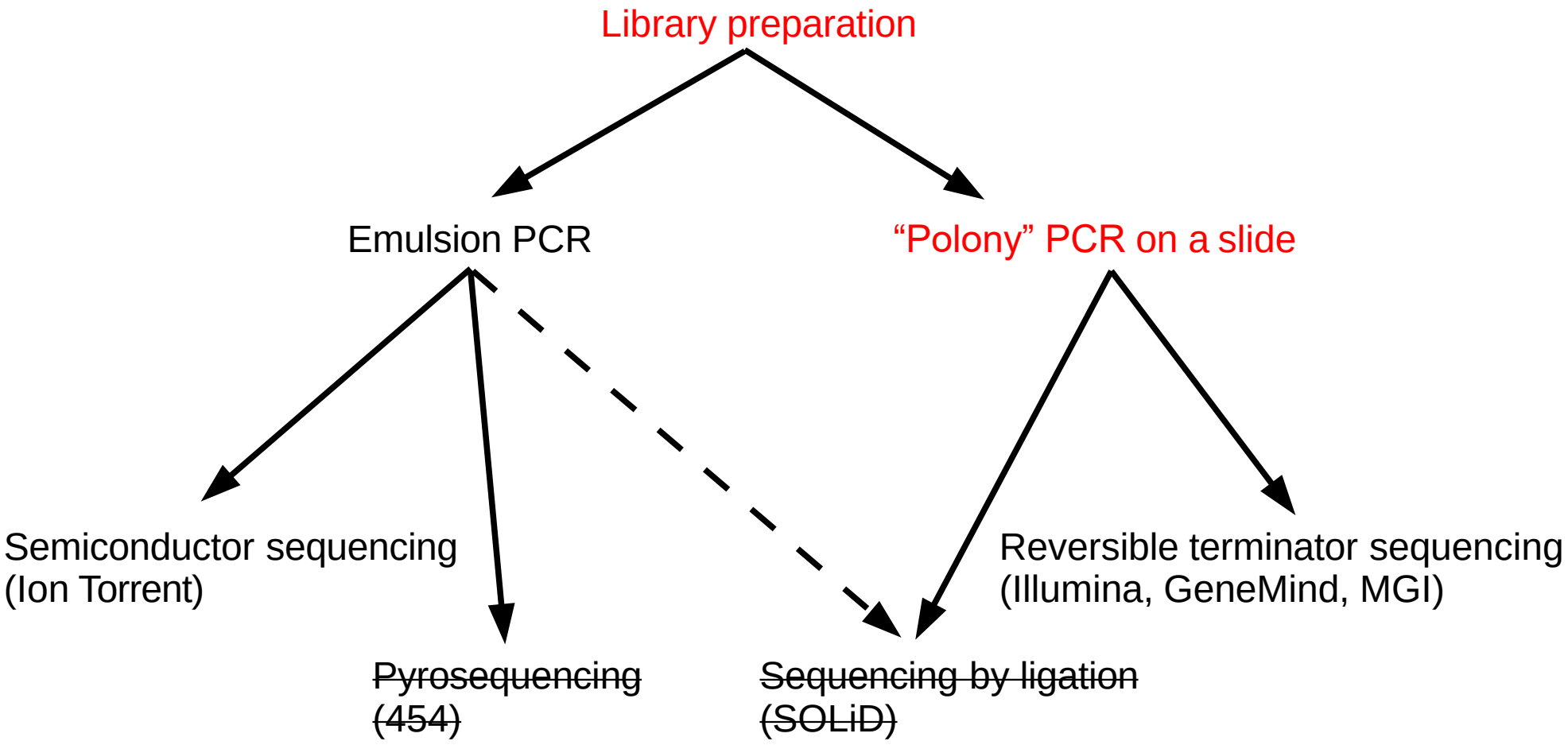
General Workflow

Next Generation Sequencing: Amplified Single Molecule Sequencing **Emulsion PCR**



General Workflow

Next Generation Sequencing: Amplified Single Molecule Sequencing

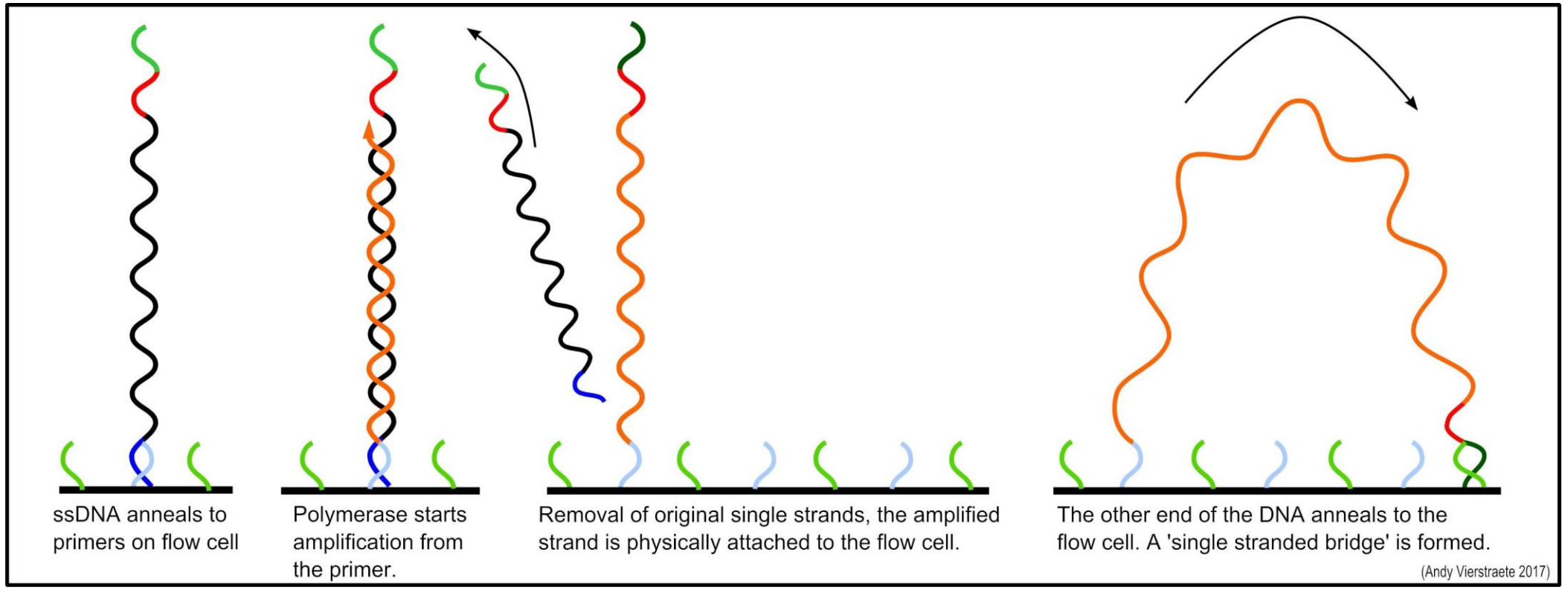


General Workflow

Illumina

Next Generation Sequencing: Amplified Single Molecule Sequencing **“Polony” PCR**

Bridge amplification: Illumina

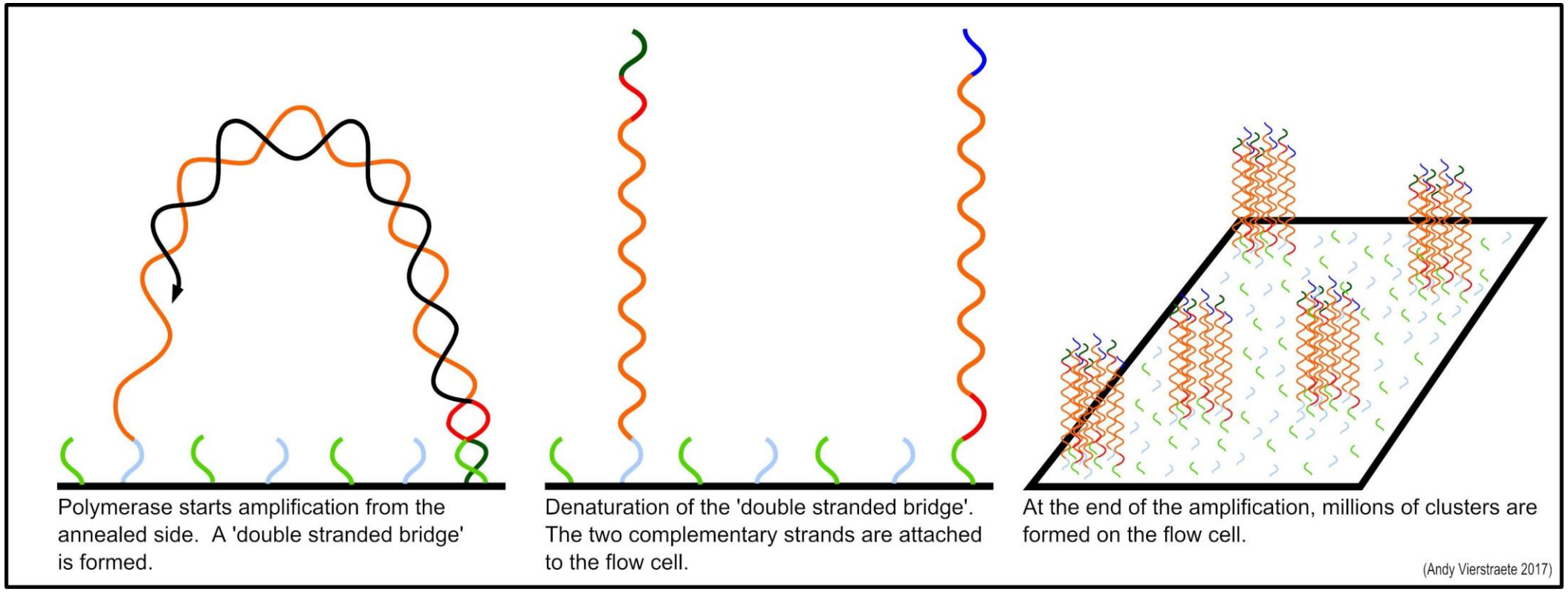


General Workflow

Illumina

Next Generation Sequencing: Amplified Single Molecule Sequencing **“Polony” PCR**

Bridge amplification: Illumina

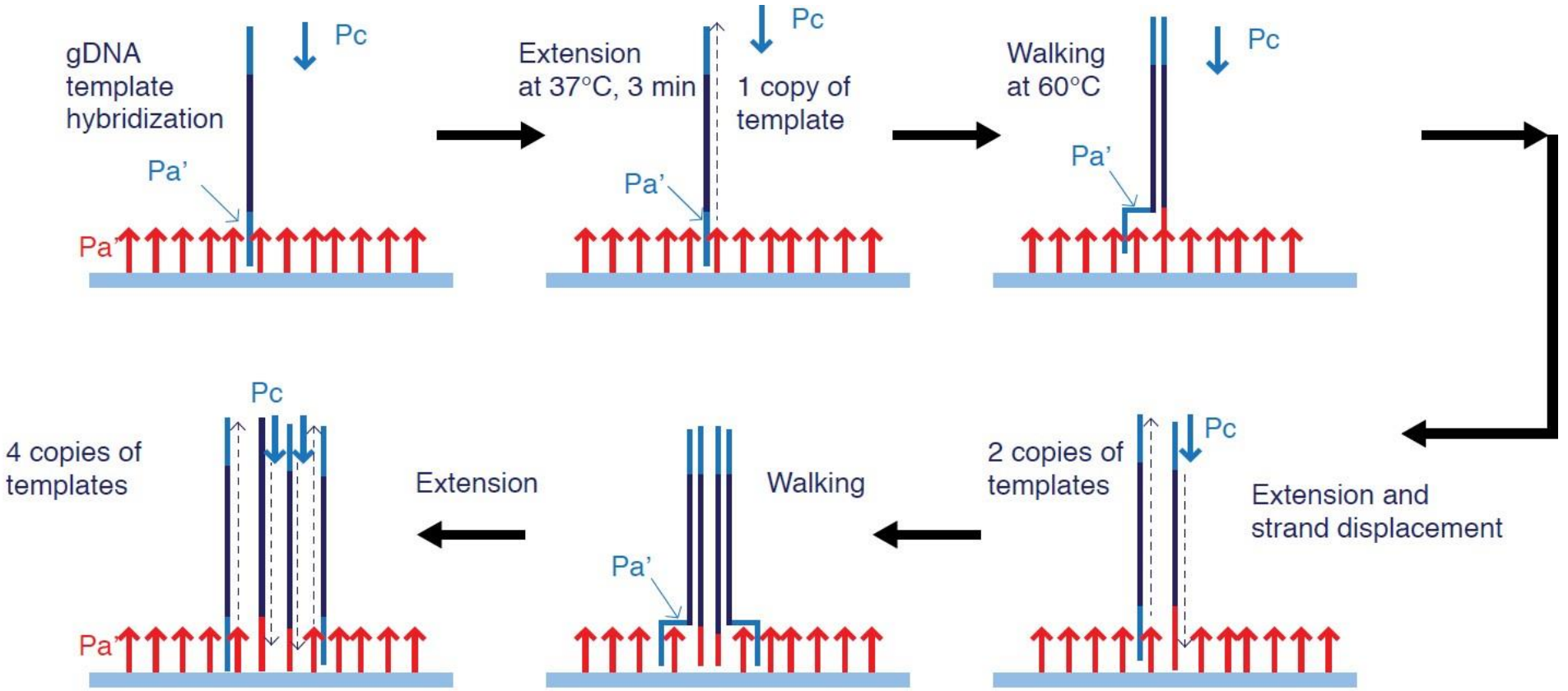


General Workflow

SOLiD

Next Generation Sequencing: Amplified Single Molecule Sequencing **“Polony” PCR**

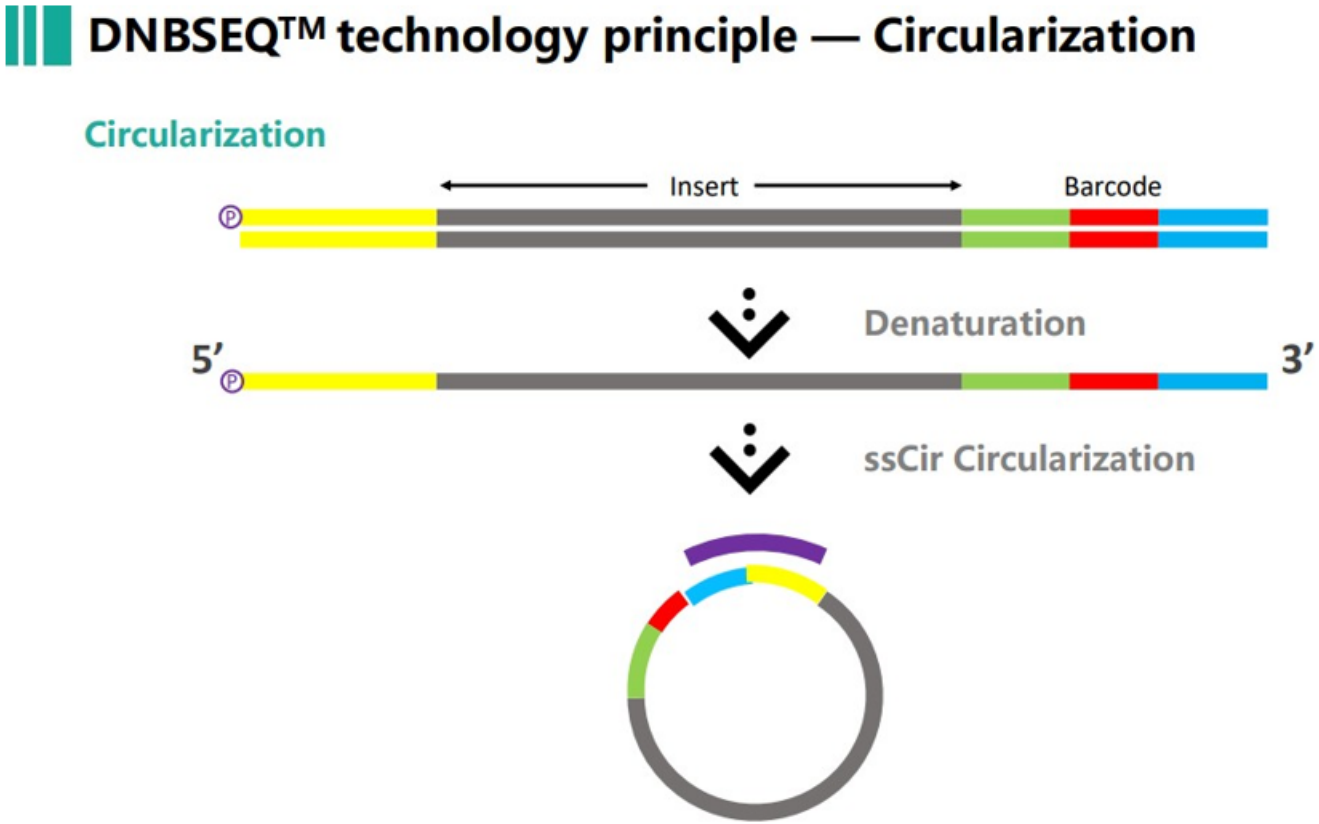
Wildfire amplification: SOLiD



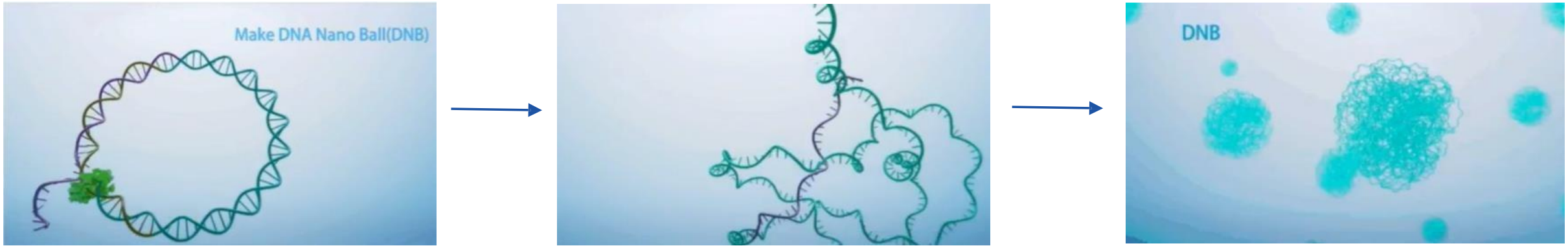
Wildfire chemistry schematic.

General Workflow

DNBSEQ: DNA NANOBALL SEQUENCING



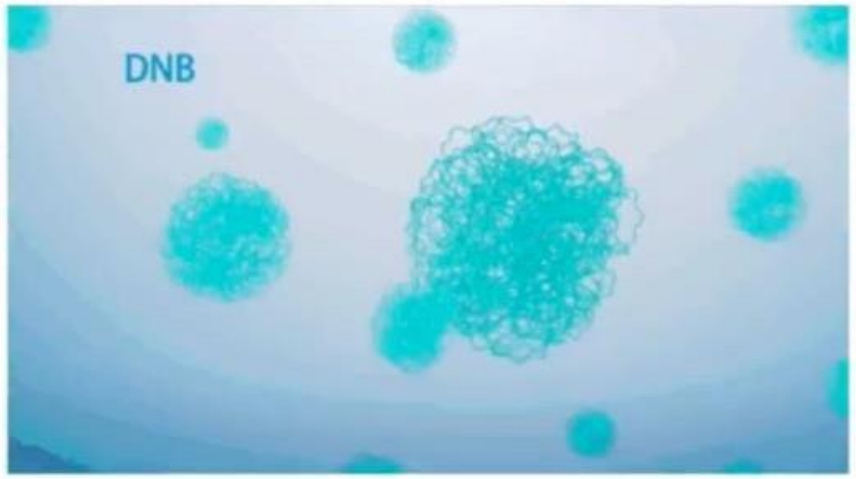
Rolling Circle Amplification

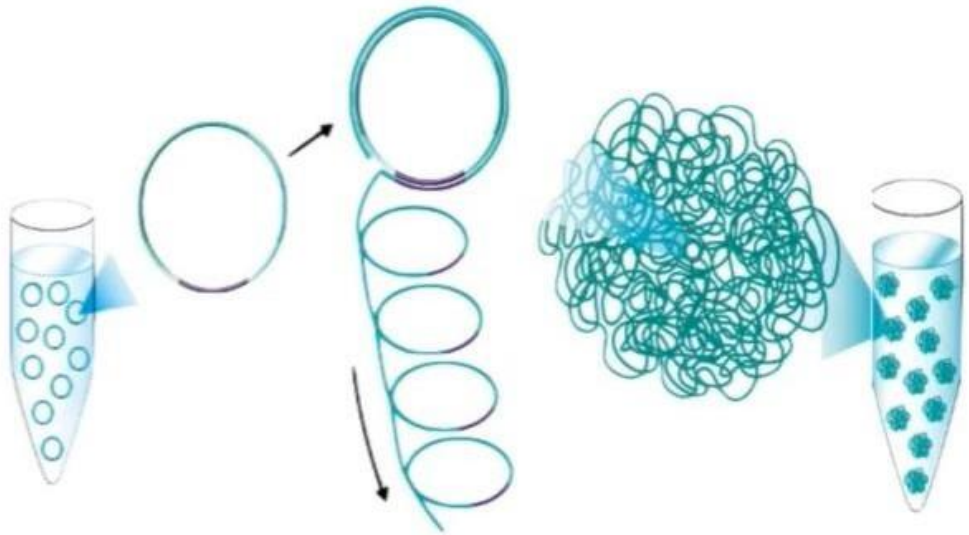


RCA: Rolling Circle Amplification

From **ssCirDNA** to **DNBs**

From 1 copy to 300-500 copies





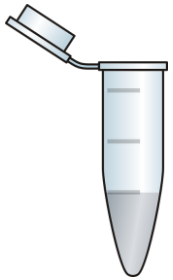
DNA Library for NGS

Speaker: Msc. Ngo Dai Phu
PI: PhD. Luu Phuc Loi

DNA Library

1. Collect Samples

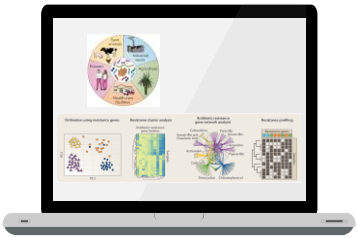
2. Extract DNA



3. Prepare DNA Library

4. Sequence DNA

5. Bioinformatics



QC QC QC

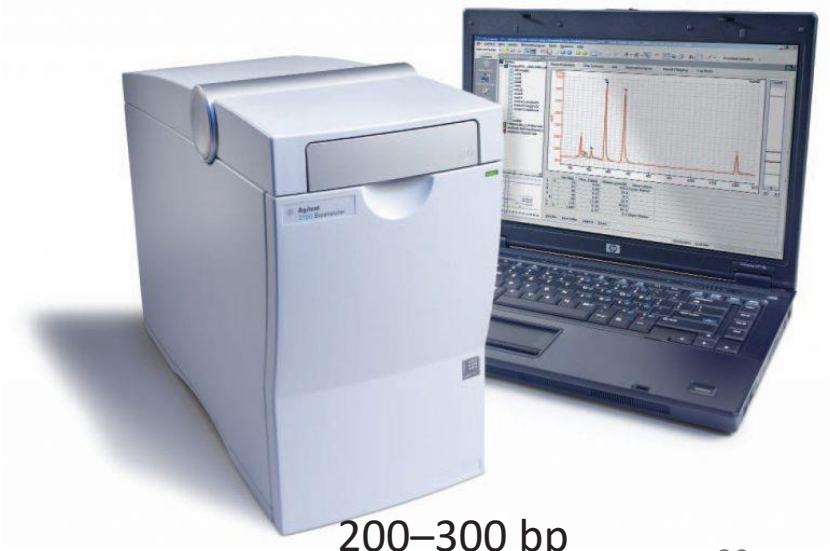
DNA Library Prep

QC of DNA Input

- The amount of input DNA: between 100 ng and 1 µg depending on the kit or protocol being used.
- Genomic DNA concentrations for Illumina libraries: Qubit
- Distribution of fragments: on average, ~200–300 bp and range ~100–1,000 bp

OD 260/280 Ratio: 1.7 -2
NanoDrop™ One - Thermo

100 ng and 1 µg
Qubit 4 Fluorometer (Thermo)



200–300 bp
Agilent's BioAnalyser

DNA Library Prep

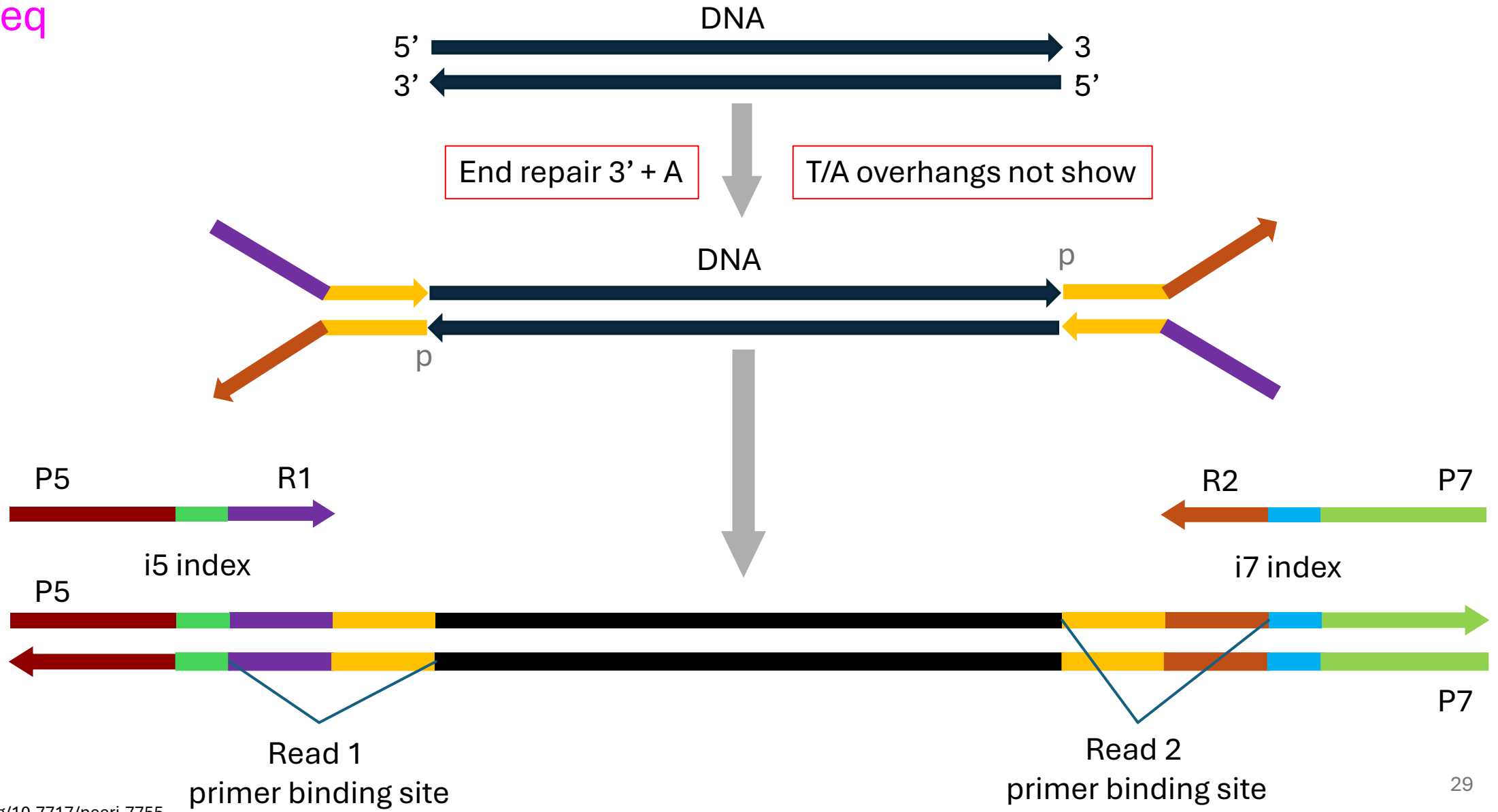
- Applications
  - Whole Genome Sequencing (WGS)
  - Whole exome sequencing (WES)
  - Targeted gene panels
- Methods
  - TruSeq/Ligated-based Methods
  - Nextera approach
  - Amplicon seq/PCR-based Methods

DNA Library Prep

- Applications
  - Whole Genome Sequencing (WGS) – TruSeq, Nextera approach
  - Whole exome sequencing (WES) – TruSeq, Amplicon seq, Nextera
  - Targeted gene panels – TruSeq, Amplicon seq
- Methods
  - TruSeq/Ligated-based Methods
  - Nextera approach
  - Amplicon seq/PCR-based Methods

DNA Library Prep

TruSeq



http://dx.doi.org/10.7717/peerj.7755

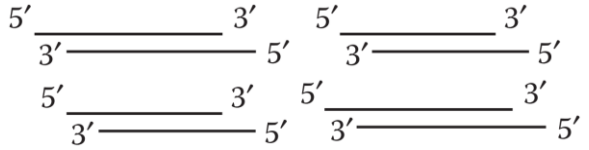
DNA Library Prep

Nextera approach

- The kit has been optimized for input of 50 ng of DNA. If a larger amount of DNA is used then the resulting library fragment distribution will consist of fragments that are, on average, larger than desired, whereas if <50 ng is used, then the size distribution will shift toward smaller-sized fragments.
- The previous library-making methods require the starting DNA material to be fragmented (e.g., via sonication), end-repaired, A-tailed (traditional approach only), and ligated to adapters. In contrast, the Nextera method only involves a one-step 5-minute enzymatic reaction—a process called *tagmentation*, which results in adapter-ligated fragments ready for the first limited cycle PCR step.
- Performing the tagmentation reaction step is simple as it consists of first adding the input DNA, enzyme buffer, and transposase enzyme/adapter complexes to a reaction tube followed by a 5-minute incubation period at 55°C in a thermocycler

Step 1: Tagmentation of input DNA followed by SPRI

Input DNA
(e.g., genomic DNA)

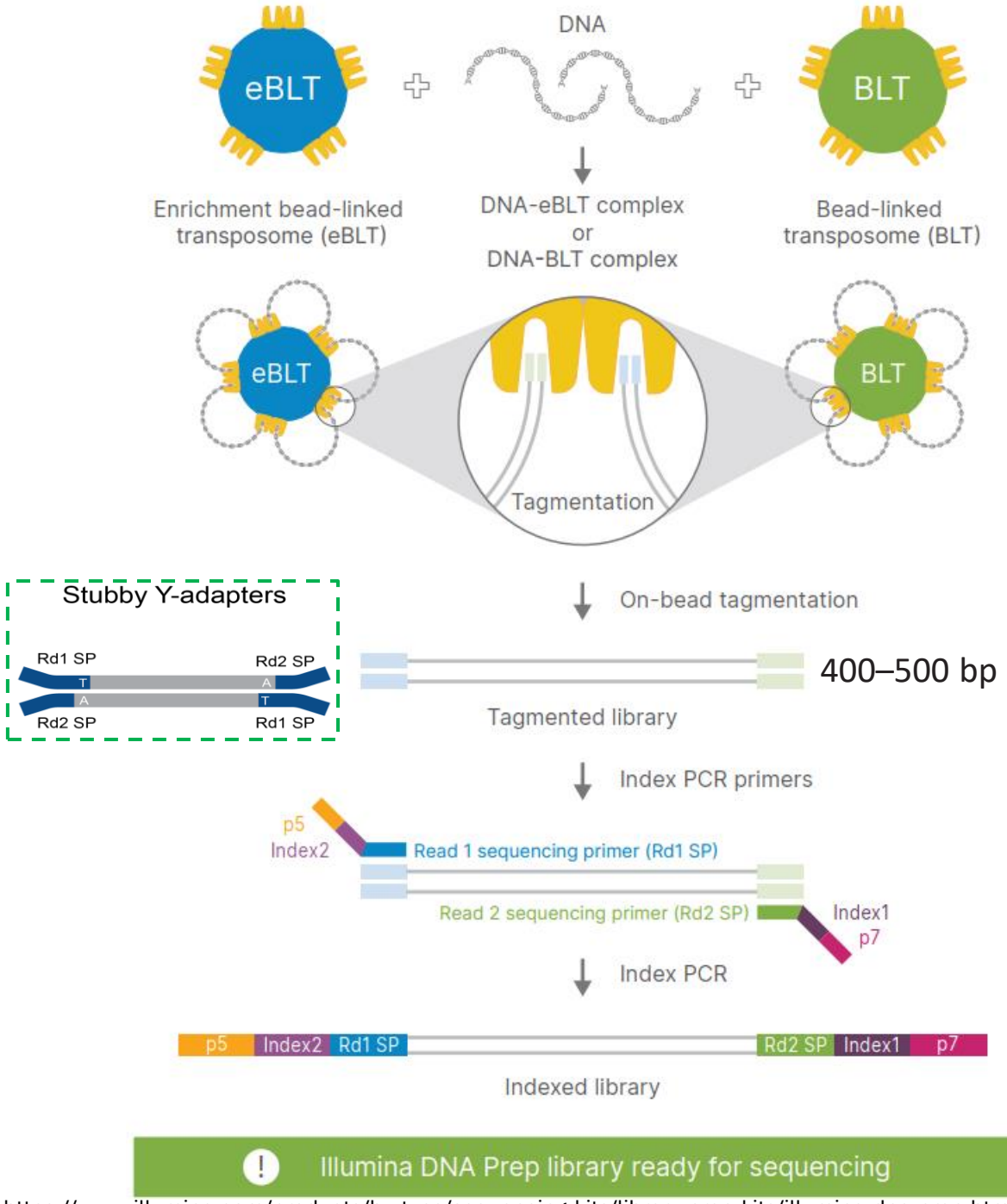


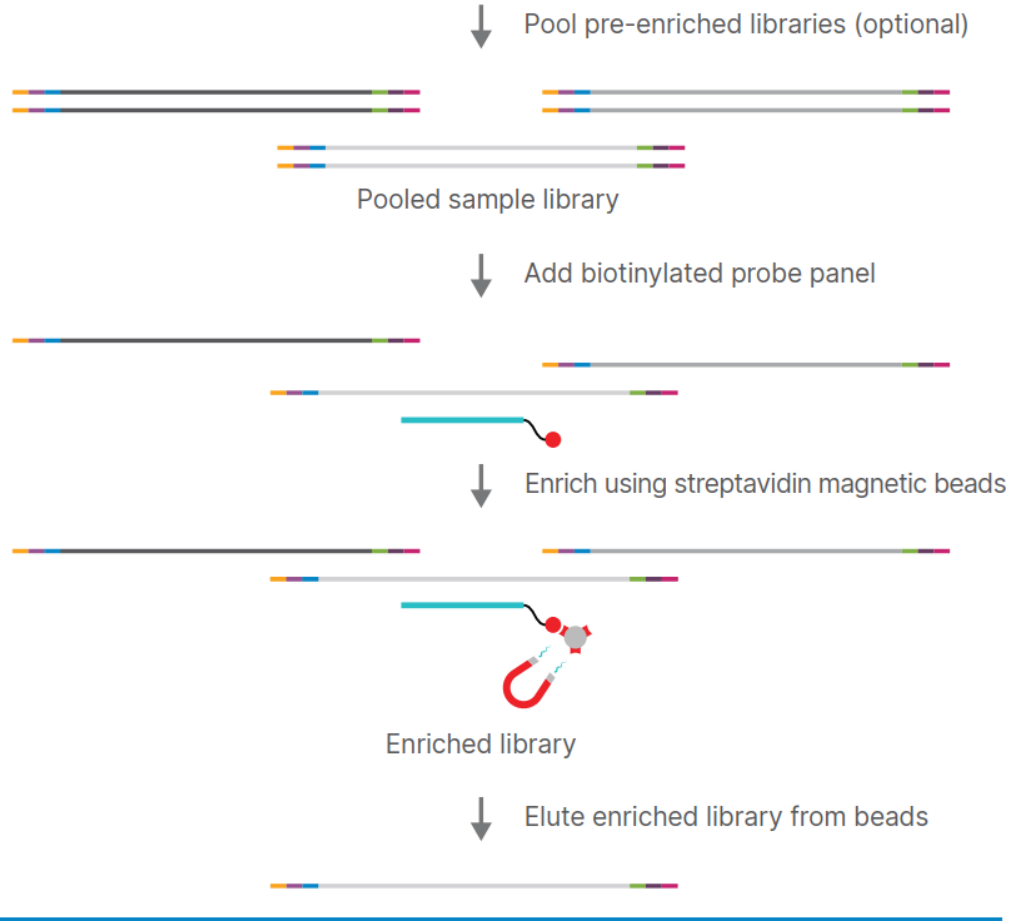
Step 2: First limited cycle PCR followed by SPRI

Step 3: Quantify libraries and verify fragment distributions

Step 4: Normalize and pool* libraries (e.g., 4–10 libraries per pool)

In-solution hybrid selection





Illumina DNA Prep with Enrichment library ready for sequencing

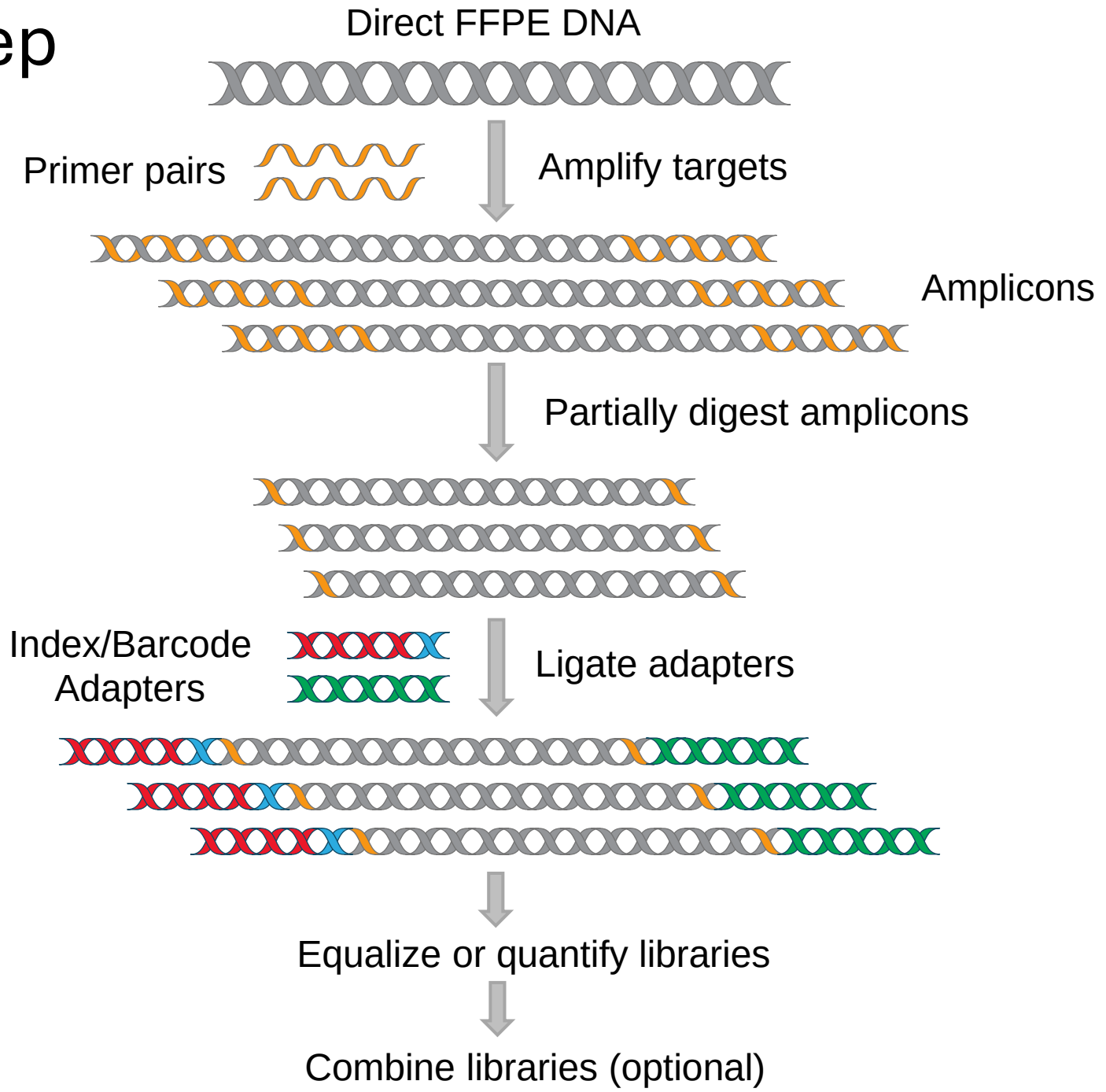
Illumina DNA Prep bead-linked transposome Chemistry

On-bead tagmentation mediates the simultaneous fragmentation of gDNA and the addition of Illumina sequencing primers. Reduced-cycle PCR amplifies DNA fragments and adds indexes and adapters. Sequencing-ready Illumina DNA Prep libraries are pooled. Illumina DNA Prep with Enrichment libraries are pooled and undergo a single hybridization reaction to produce an enriched library ready for sequencing.

https://www.illumina.com/products/by-type/sequencing-kits/library-prep-kits/illumina-dna-prep.html#tabs-0e810c0b06-item-7eade5e2d5-documentation

DNA Library Prep

Amplicon



A Specific Example

BRCA Assay

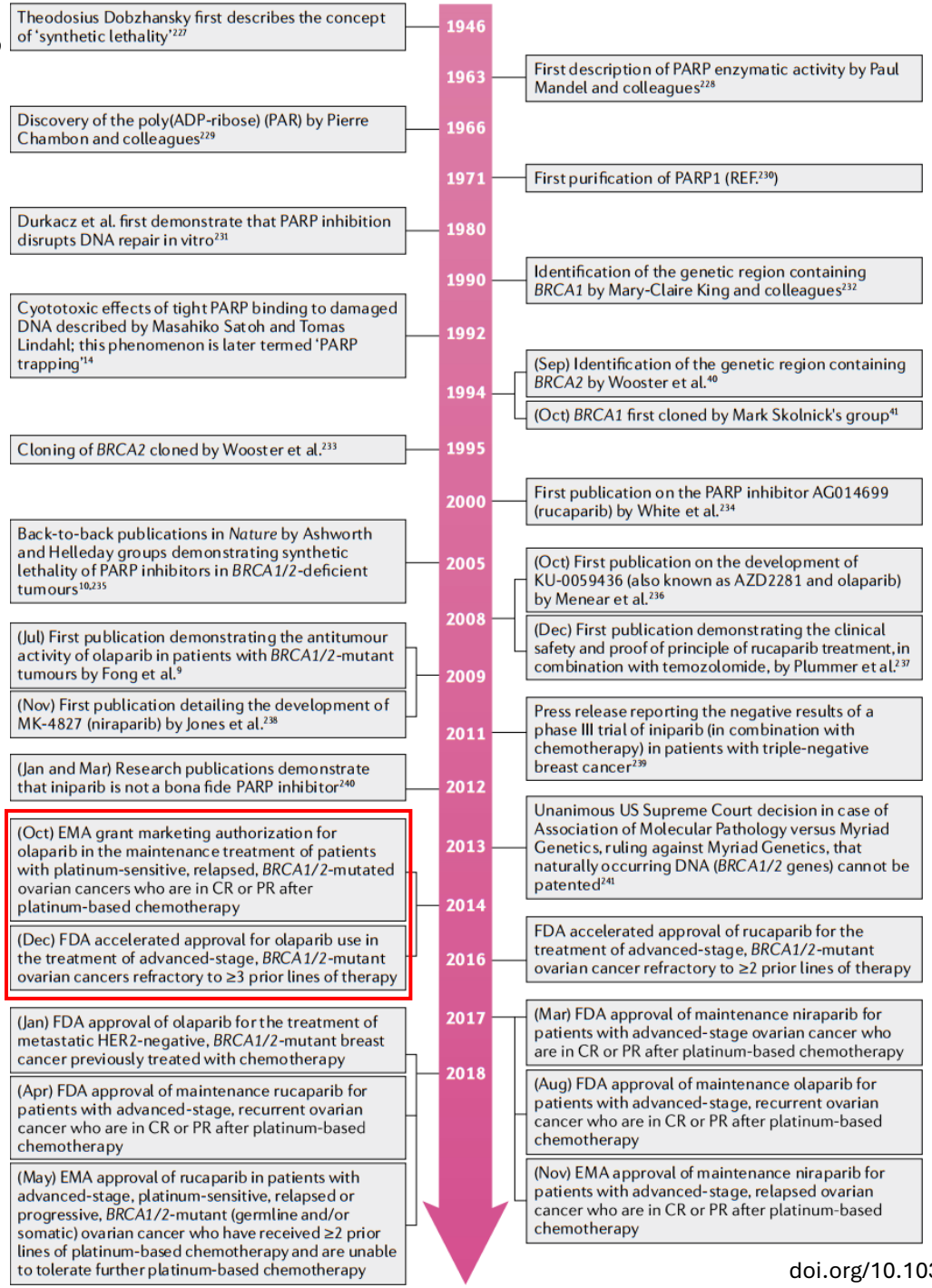
Speaker: Msc. Ngo Dai Phu
PI: PhD. Luu Phuc Loi

DNA Library Prep

- Applications
  - Whole Genome Sequencing (WGS) – TruSeq, Nextera approach
  - Whole exome sequencing (WES) – TruSeq, Amplicon seq
  - Targeted gene panels – Amplicon seq
- Methods
  - TruSeq – ligated based methods
  - Nextera approach
  - Amplicon seq/PCR-based Methods

Landmark discoveries and advances in the development of poly(ADP-ribose) polymerase (PARP) inhibitors

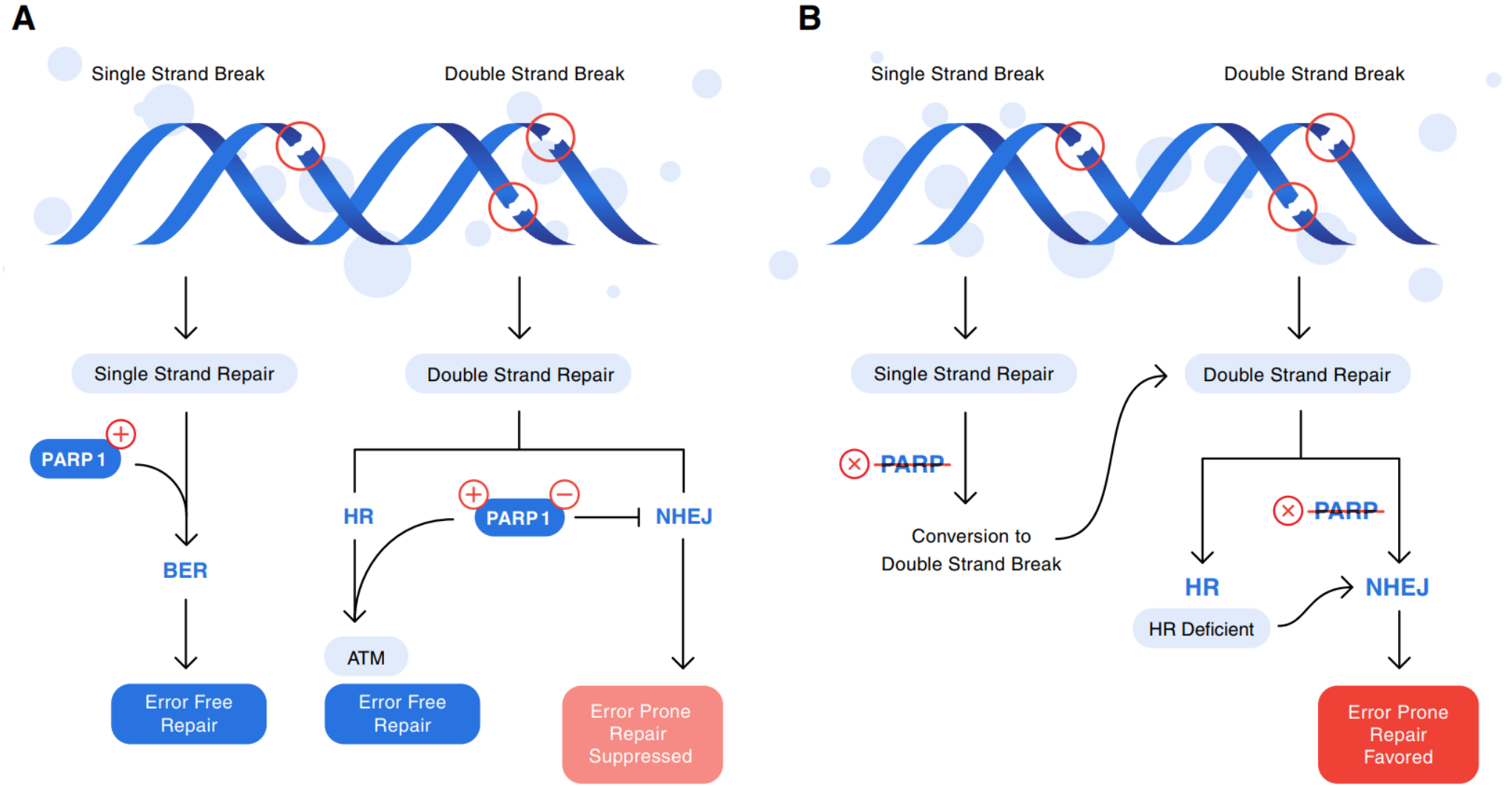
2014: Olaparib, BRCA1/2-mutant ovarian cancers
2016: Rucaparib, BRCA1/2-mutant ovarian cancer
2018: Olaparib for the treatment of metastatic HER2-negative, BRCA1/2-mutant breast cancer
2018: Rucaparib in patients with advanced-stage, platinum-sensitive, relapsed or progressive, BRCA1/2-mutant (germline and/or somatic) ovarian cancer



Timeline of key events leading to FDA approvals of PARP inhibitors in cancer medicine.
Landmark discoveries and advances in the development of poly(ADP-ribose) polymerase (PARP) inhibitors are indicated, together with the current approved indications for these agents in the USA and the EU. CR , complete remission; PR , partial remission .

doi.org/10.1038/s41571-018-0114-z

Rationale for PARP Inhibitor Use in EOC

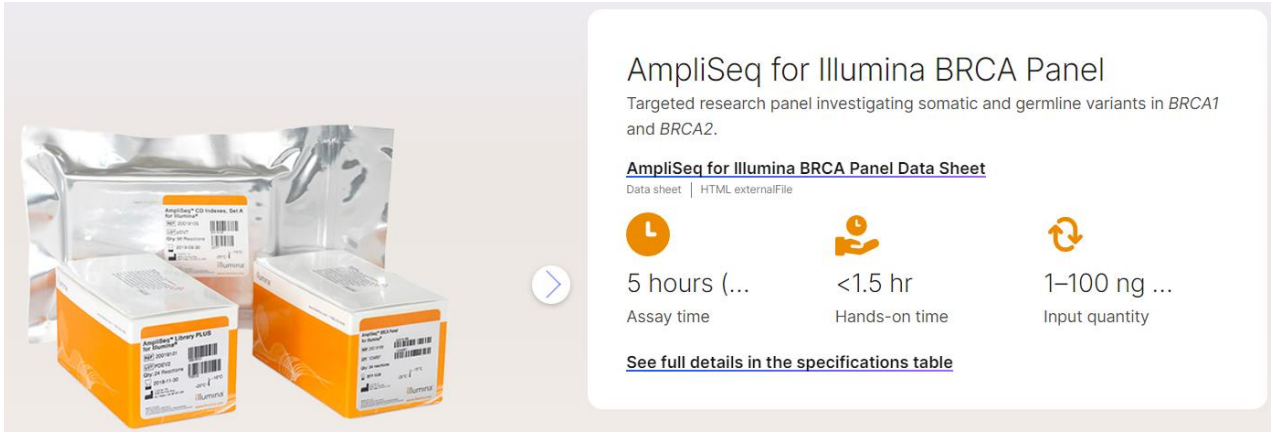


Poly-(ADP-ribose) polymerase (PARP) maintains genomic integrity by supporting single strand (SS) and double strand (DS) DNA break repair.
A When PARP function is intact, SS DNA breaks are repaired by base excision repair (BER), an error-free mechanism. DS DNA breaks are repaired by homologous recombination (HR), an error-free process, supported by PARP activation of the DNA damage kinase ataxia-telangiectasia mutated (ATM) protein. Simultaneously, PARP suppresses error-prone mechanisms of DS DNA break repair to maintain genomic integrity.
B When PARP function is impaired, SS DNA breaks are converted to DS DNA breaks and the suppression of error-prone mechanisms of DS DNA repair is removed. If HR is also impaired (e.g. BRCA1/2 mutation, or other HR defciencies), then DNA repair occurs via error-prone mechanisms, leading to genomic instability and cell death. NHEJ non-homologous endjoining

DOI: 10.1007/s40259-019-00347-4

BRCA Assay

- AmpliSeq for Illumina BRCA Panel (Illumina)



- Oncomine BRCA Assay (Ion Torrent)

White Paper: Evaluation of the Oncomine *BRCA* Research Assay for variant detection by next-generation sequencing

Fill out the short form below to access the white paper and learn more about:

- Important inherited and somatic mutations in *BRCA1/2*, their associations with several cancer types, and their importance as biomarkers for approved cancer therapies
- The technology behind the Oncomine *BRCA* Research Assay, which enables automated, rapid sequencing with low DNA input from blood, cell lines, or FFPE samples
- The sensitivity and specificity of the assay in the detection of SNVs, MNVs, indels, and whole-exon, multiple-exon, or entire-gene aberrations
- Ion Reporter Software and Ion Torrent Oncomine Reporter informatics packages

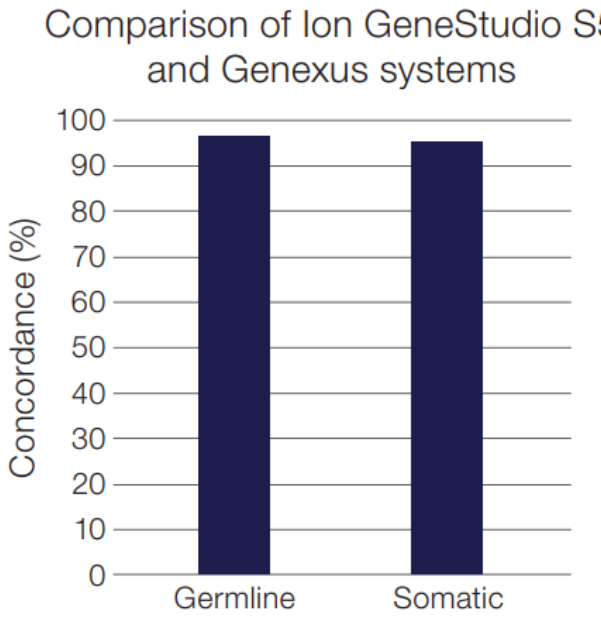
https://www.illumina.com/products/by-type/sequencing-kits/library-prep-kits/ampliseq-brca-panel.html
https://www.thermofisher.com/vn/en/home/clinical/preclinical-companion-diagnostic-development/oncomine-oncology/oncomine-brca-research-assay.html

Specifications of Oncomine *BRCA* Research Assay

A

Control sample variants	Platform	Library	SNV		Indel	
			Sensitivity (%)	PPV (%)	Sensitivity (%)	PPV (%)
5% allele frequency	Ion GeneStudio S5 System	Manual Chef	100 100	98 92	98 99	92 99
	Genexus System	Genexus	100	100	98	100
50%, 100% allele frequency	Ion GeneStudio S5 System	Manual Chef	100 100	100 100	100 100	100 100
	Genexus System	Genexus	100	100	94	100

B



Superior accuracy in detecting somatic and germline variants, and high consistency, independent of the workflow.

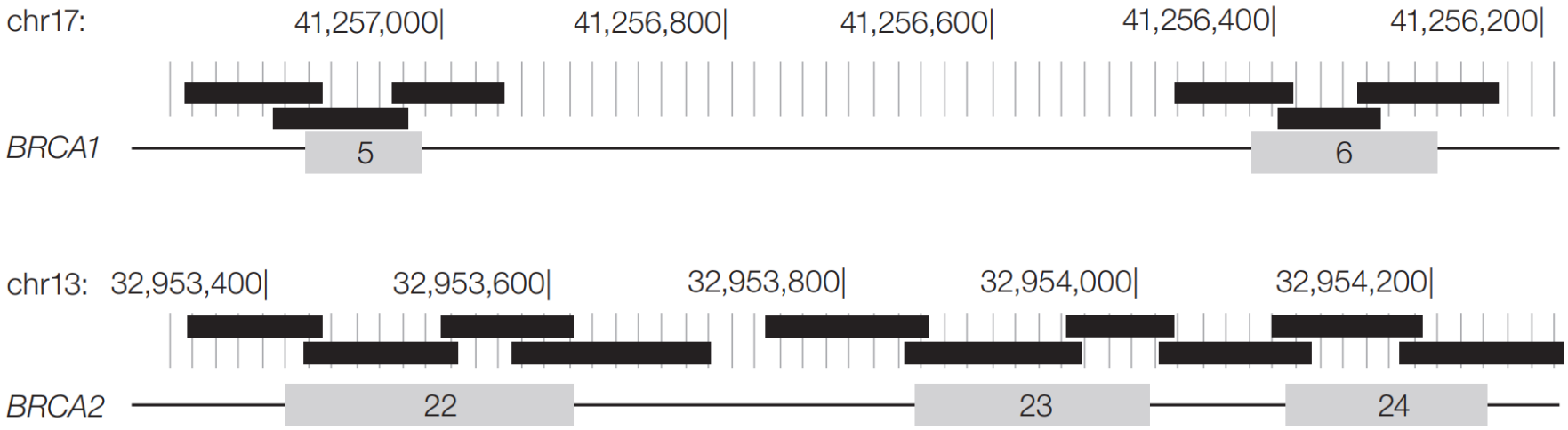
(A) Sensitivity and positive predictive value (PPV) for detecting single-nucleotide variants (SNV) and indels. Positive predictive value = true positives/total number of positives. Sensitivity = true positives/(true positives + false positives).

(B) (B) The percent of concordance for germline analysis and somatic cell analysis with the Ion GeneStudio S5 and Genexus systems.

https://www.thermofisher.com/vn/en/home/clinical/preclinical-companion-diagnostic-development/oncomine-oncology/oncomine-brca-research-assay.html

Specifications of Oncomine *BRCA* Research Assay

The Oncomine *BRCA* Research Assay covers 100% of the coding sequences of *BRCA1/2*, including all splice and acceptor sites, with an average extension of 63 bp into adjoining introns.
The recommended input is 20 ng of DNA extracted from either FFPE or blood samples.



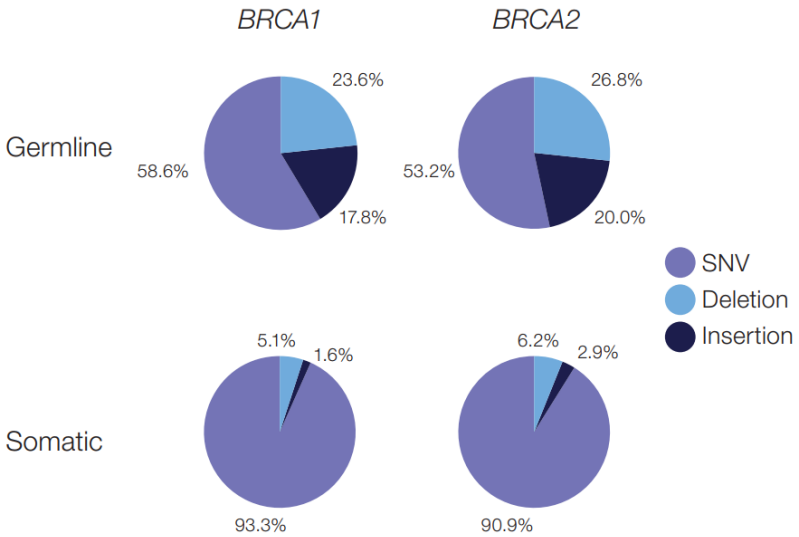
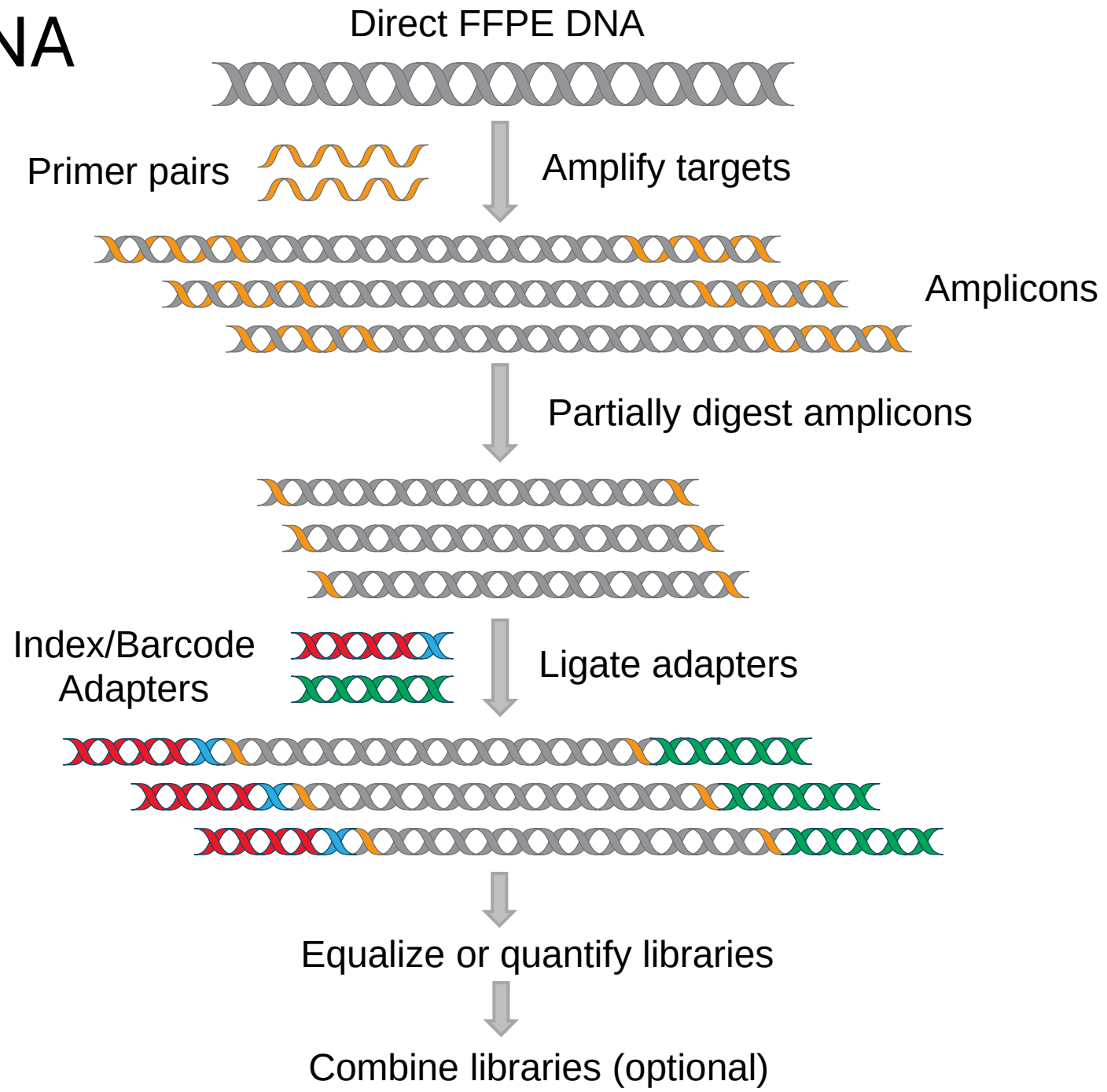


Table 8. Distribution of variant types, including indels associated with HP sequence ≥5, in TCGA somatic mutation data.

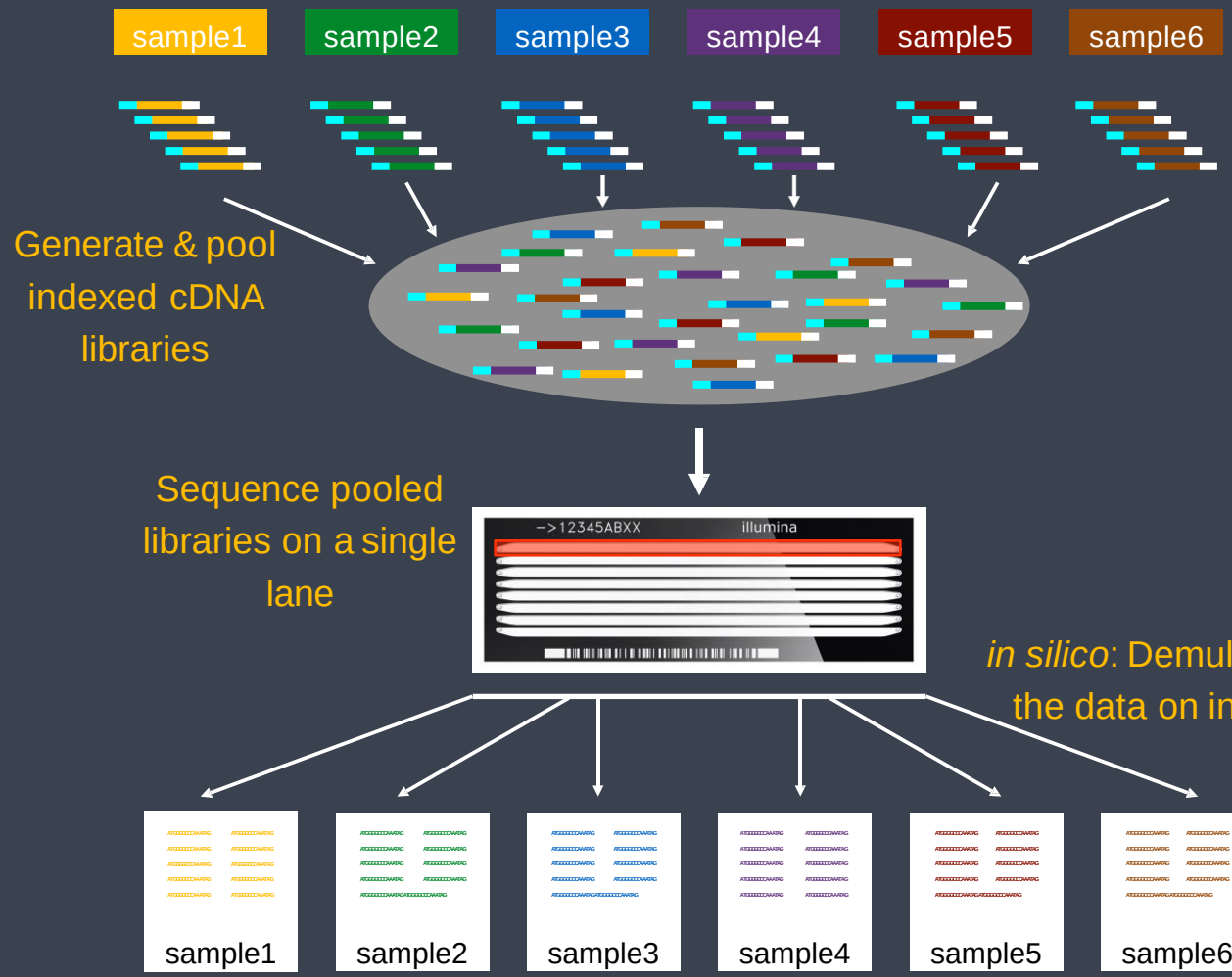
	BRCA1	*BRCA2*	*BRCA1/2* (combined)
Positive samples	263	431	694
Positive variants	314	689	1,003
SNV	293 (93.3%)	626 (90.9%)	919 (91.6%)
Indel	21 (6.7%)	63 (9.1%)	84 (8.3%)
Indel with HP ≥5	2 (0.64%)	22 (3.2%)	24 (2.4%)

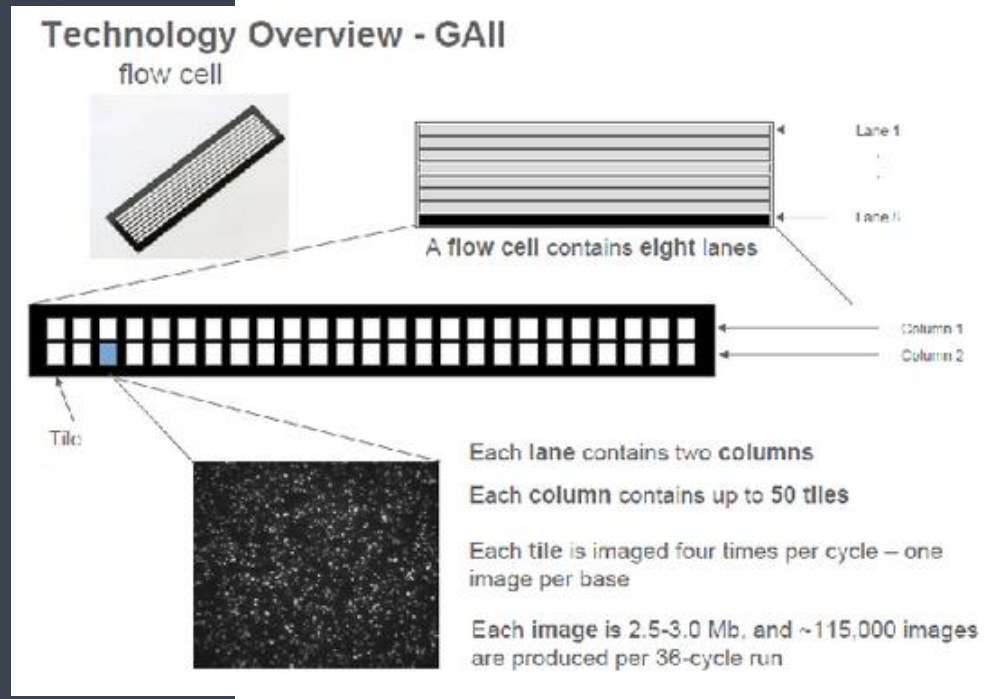
https://www.thermofisher.com/vn/en/home/clinical/preclinical-companion-diagnostic-development/oncomine-oncology/oncomine-brca-research-assay.html

Workflow of DNA Library Prep



Multiplexing (barcodes and indices)





Fulllength Adapter P5

Fulllength Adapter P7

```
5'- AATGATACGGCGACCACCGAGATCTACACACCGACAAACACTCTTTCCCTACACGACGCTCTTCCGATCTXXXXX//XXXXXAGATCGGAAGAGCACACGTCTGAACTCCAGTCACAGGTCACTATCTCGTATGCCGTCTTCTGCTTG-3'
3'- TTACTATGCCGCTGGTGGCTCTAGATGTGTGGCTGTTTGTGAGAAAGGGATGTGCTGCGAGAAGGCTAGAXXXXX//XXXXXTCTAGCCTTCTCGTGTGCAGACTTGAGGTCAGTGTCCAGTGATAGAGCATACGGCAGAAGACGAAC-5'
```

i5 index

Full-length and double-indexed libraries

i7 index

Thank You!