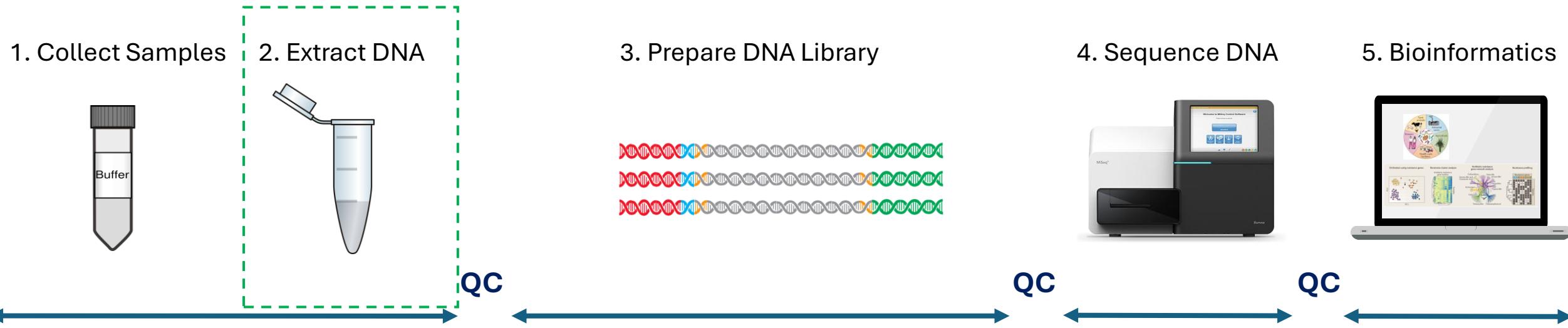


DNA Extractions

Speaker: Ngo Dai Phu, Msc.

PI: Luu Phuc Loi, PhD.

Research Workflow



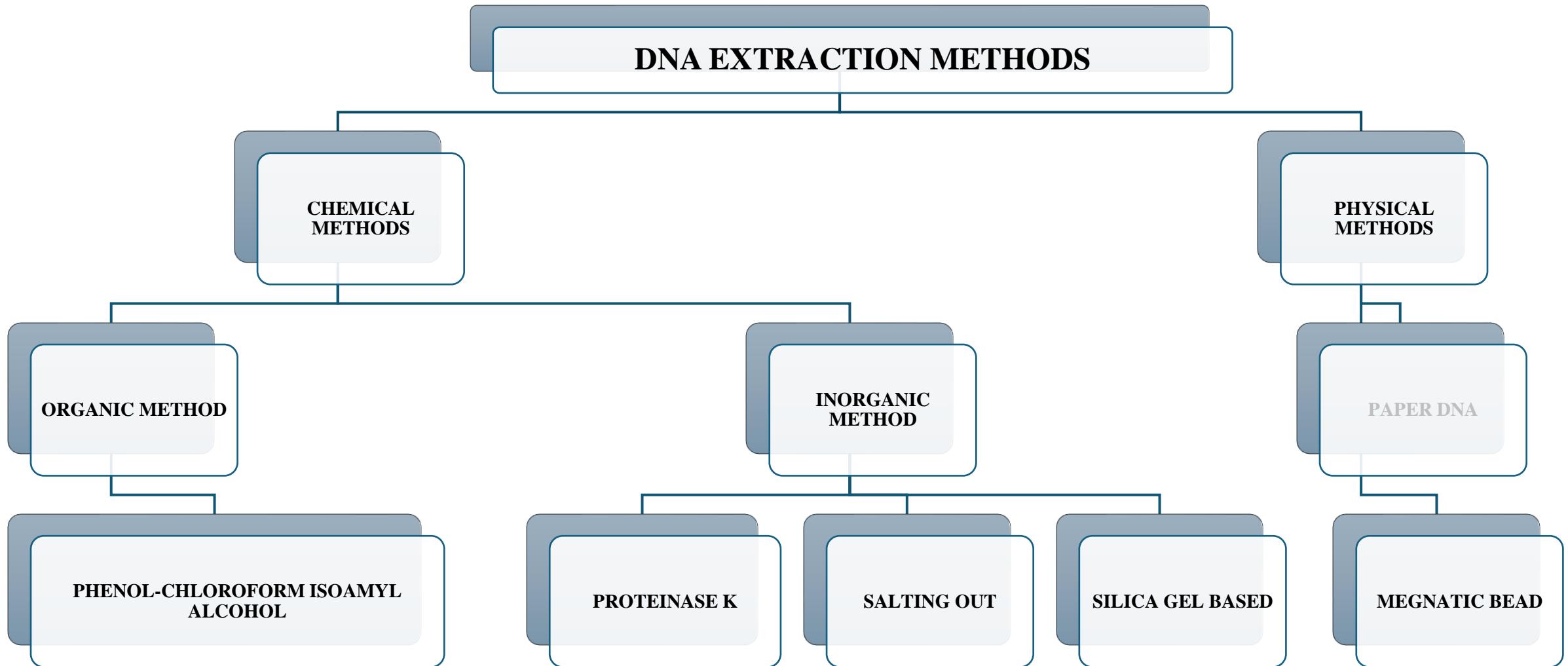
Introduction

- Several different types of DNA extraction methods are Phenol-chloroform isoamyl alcohol, Proteinase K, CTAB method, spin column-based methods and magnetic bead-based technique.
- Depending upon the type of sample, every DNA extraction method varies.
 - For example, the DNA extraction method for plant DNA is different from that of the blood.
 - Likewise, the bacterial DNA isolation method is different from other types.
 - Meaning, we need varieties of DNA isolation technique for different sample
- Main Steps: **cell lysis, precipitation, and elution**

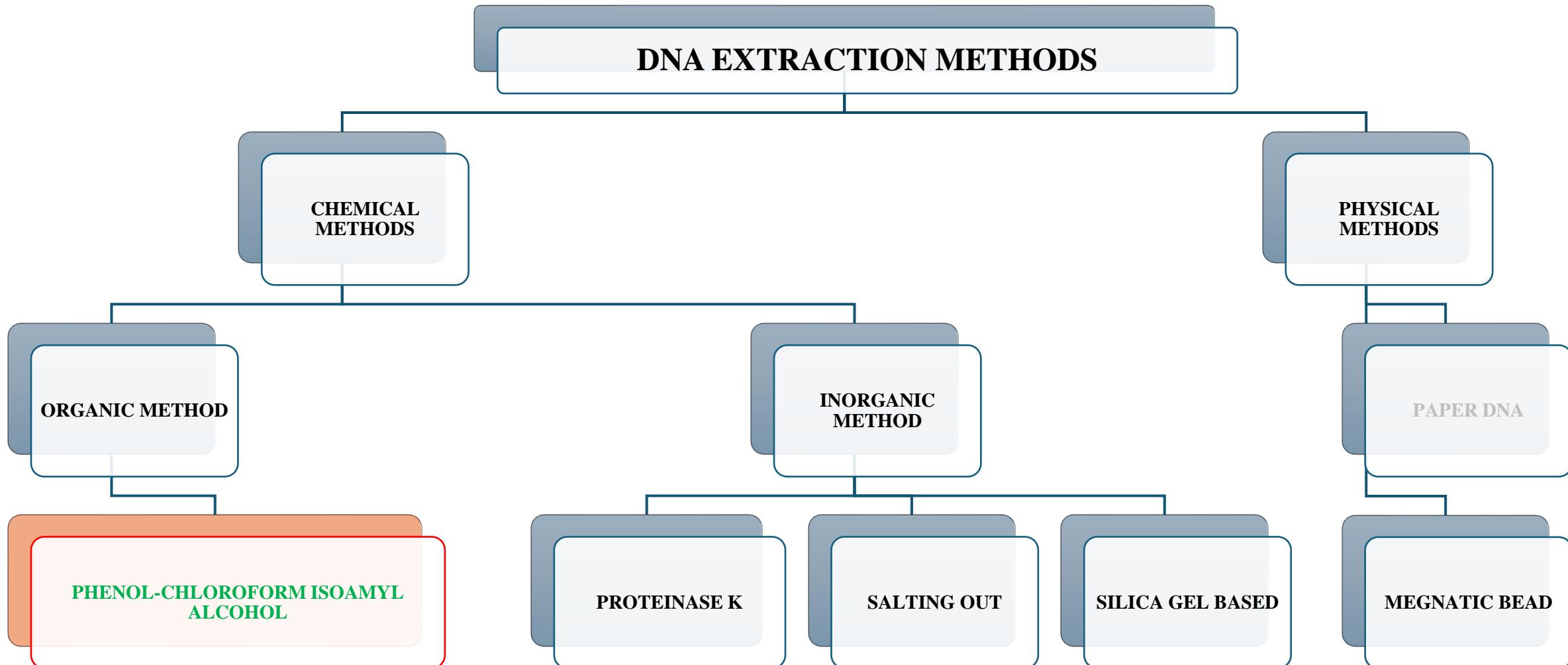
Different types of NA Extraction Methods

- DNA extraction methods are broadly categorized into two categories:
 - Chemical-based method
 - Physical-based method
 - Enzymatic-based Methods
 - Hybrid Methods (Physical + Chemical & Enzymatic + Chemical)
- The chemical DNA extraction methods are also known as solution-based methods

Different types of NA Extraction Methods

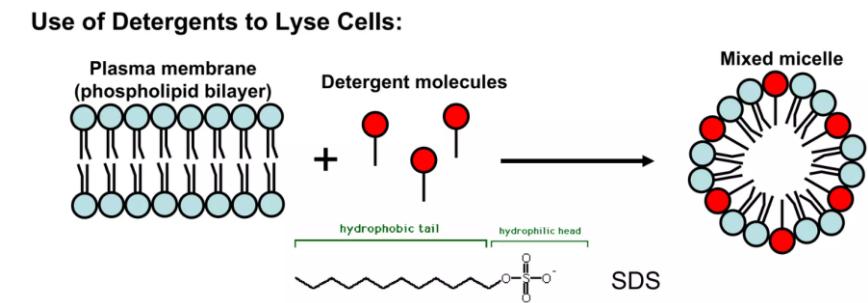


Different types of NA Extraction Methods



Chemical-based DNA extraction method

- Chemical or solution-based method uses various organic & inorganic solutions.
- Common chemicals used - SDS, CTAB, phenol, chloroform, isoamyl alcohol, Triton X 100, guanidium thiocyanate, Tris and EDTA
 - Dissolve cellular membranes
 - Inactivation of DNase and Rnase
 - Assist in the removal of contaminants
- The solution-based DNA extraction method is subdivided into
 - Organic solvent-based DNA extraction - phenol and chloroform
 - Inorganic solvent-based DNA extraction - Proteinase K
- **Phenol and chloroform** - harmful nature - less recommended- Yet **best** among all
- The proteinase K DNA extraction method - high DNA yield - But time-consuming.
- The lower stability of the enzyme is another major issue in this method.

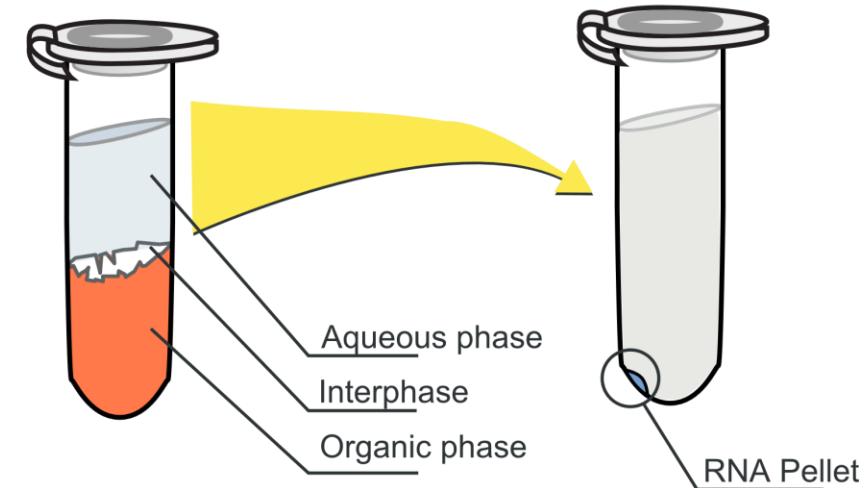


Phenol-chloroform method

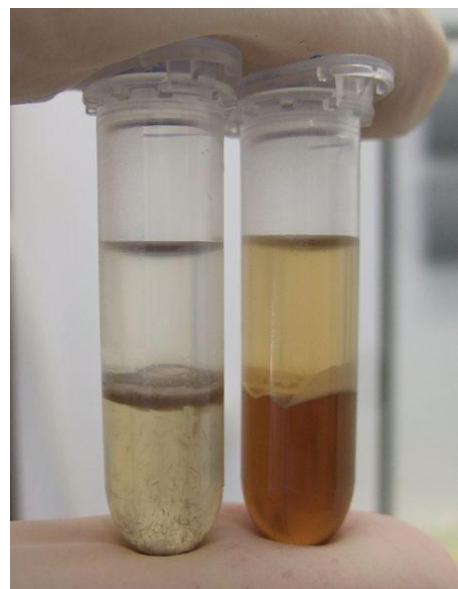
- One of the **best** methods of DNA extraction
- The yield and quality of DNA obtained by the PCI method is **very good** if we perform it well
- Also referred as a **phenol-chloroform and isoamyl alcohol or PCI method of DNA extraction.**
- Major chemicals - Lysis buffer, Phenol, chloroform, ethanol, isoamyl alcohol
- The **lysis buffer** contains Tris, EDTA, MgCl₂, NaCl, SDS, and other salts.
- Components of **lysis buffer** help in the **lysis of the cell membrane** as well as the **nuclear envelope.**
- The **organic component** of the technique - phenol and chloroform denature the protein portion of cells.

Phenol-chloroform method

Phase separation



Isopropanol precipitation



Samples (Blood, tissue, bacteria)

Soft cell wall

Add extraction buffer (Tris, EDTA, NaCl, MgCl₂, SDS, pH 8.0)

Mix and spin – 3000rpm

Pellet: add saturated phenol in Tris

Mix and spin – 12000rpm

Sup: add P:C:I = 25:24:1

Mix and spin – 12000rpm (optional)

Sup: add C:I = 24:1

(Precipitation)

Mix and spin – 12000rpm

Sup: add ethanol & sodium acetate = 24:1

(Washing 2 – 3 times)

Mix and spin – 12000rpm

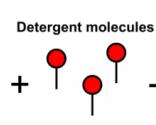
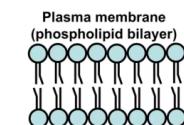
Pellet: add 70% ethanol & sodium acetate = 24:1

Dry pellet

Dissolve in TE buffer - storage

Help to get distinct organic & aqueous layers

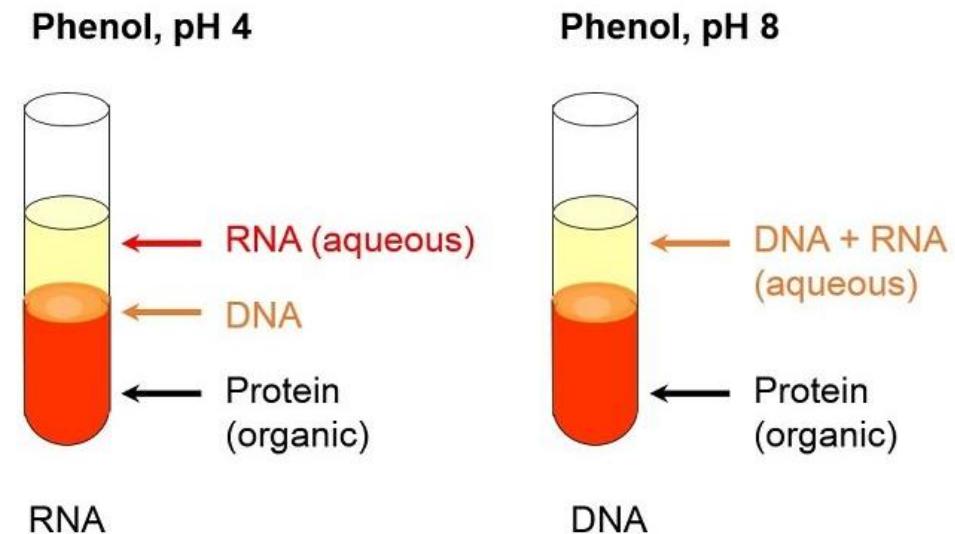
Use of Detergents to Lyse Cells:



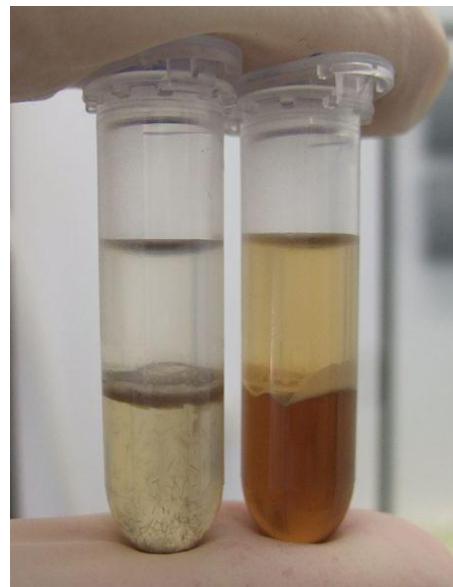
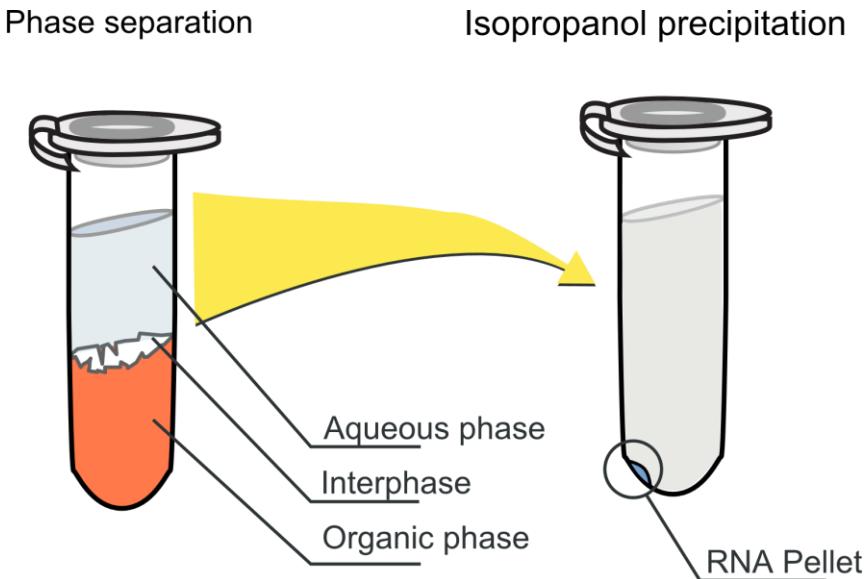
Chemical	Role in DNA extraction
Tris	DNA is pH sensitive. Maintains pH of solution (pH=8). interacts with the LPS & permeabilizes cell membrane for easy lysis
EDTA	Chelating agent & blocks the activity of DNase enzyme. Dnase- lyses the DNA with help of cofactor Ca ²⁺ . Chelator EDTA blocks the cofactor binding site. Works best in combination with Tris.
SDS	Anionic detergent. Denatures the membrane protein & break open the cell membranes and nuclear envelopes
NaCl	Prevents the denaturation of DNA Na ⁺ ion of NaCl creates the ionic bond with the negative charge of DNA and neutralizes it.
MgCl ₂	Protects DNA from mixing with other cell organelles
Phenol	Non polar solvent, unfolding & precipitation of protein impurities
Chloroform	Sharpens the fuzziness between organic & aqueous layers
Isoamyl alcohol	Prevent foaming between interphase. Anti-foaming agent
Sodium Acetate	Precipitate DNA (Na ⁺ interacts with PO ₃ ⁻)
Ethanol/isopropanol	Helps in DNA Precipitation (Interacts with water by hydrogen bonds)
TE buffer	It dissolves DNA (Tris & EDTA)

Phenol-chloroform method

- DNA:
 - Under neutral conditions (pH 7-8), both DNA and RNA partition into the aqueous phase.
- RNA:
 - Under acidic conditions (pH 4-6), DNA partitions into the organic phase while RNA remains in the aqueous phase.



Phenol-chloroform method



Samples (Blood, tissue, bacteria)

Soft cell wall

Add extraction buffer (Tris, EDTA, NaCl, MgCl₂, SDS, pH 8.0)

Mix and spin – 3000rpm

Pellet: add saturated phenol in Tris

Mix and spin – 12000rpm

Sup: add P:C:I = 25:24:1

Mix and spin – 12000rpm (optional)

Sup: add C:I = 24:1

(Precipitation)

Mix and spin – 12000rpm

Sup: add 100% ethanol & sodium acetate = 24:1

(Washing 2 – 3 times)

Mix and spin – 12000rpm, 0–8°C

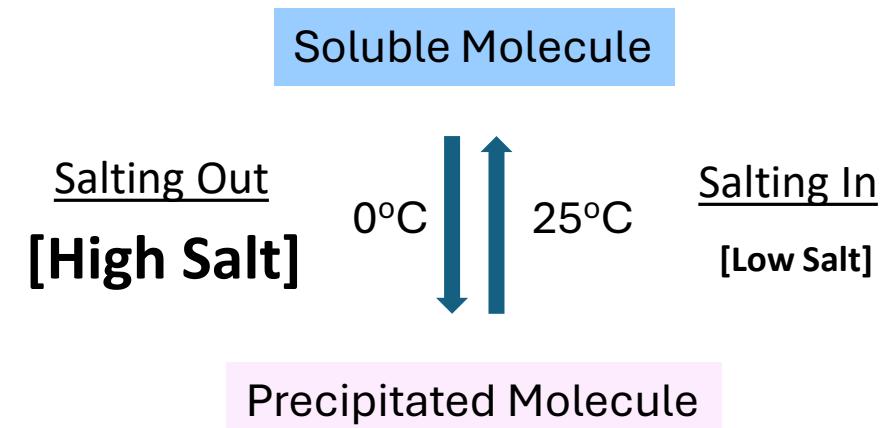
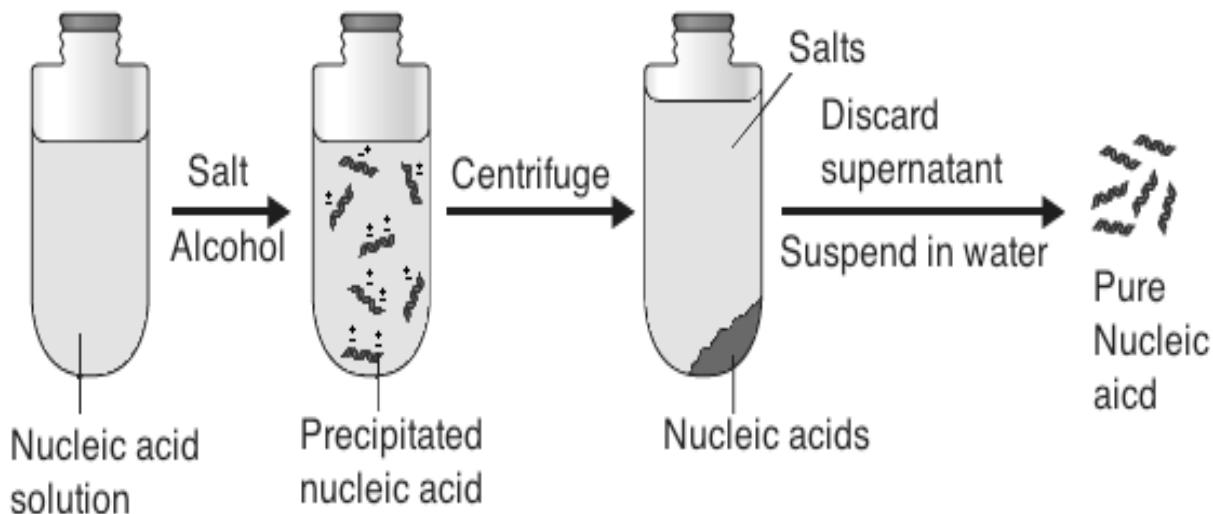
Pellet: add 70% ethanol & sodium acetate = 24:1

Mix and spin – 12000rpm, 0–8°C

Dry pellet

Dissolve in TE buffer - Storage

Ethanol Precipitation of DNA



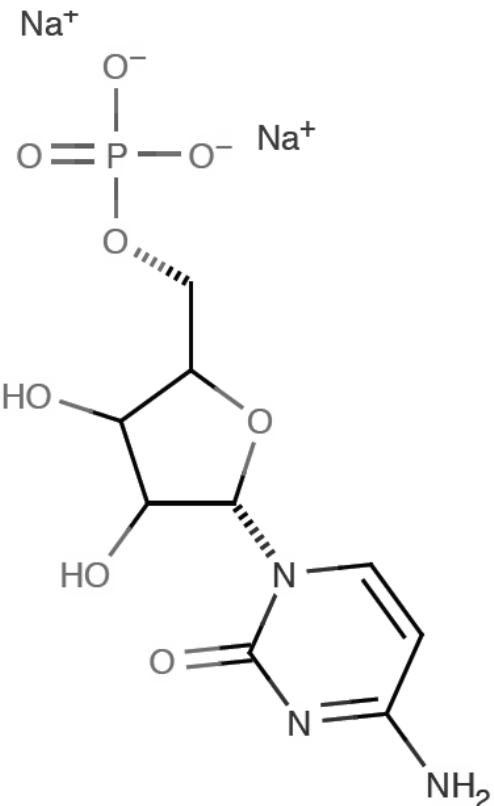
Ethanol Precipitation of DNA

Ethanol Precipitation is a method for purifying or concentrating DNA/RNA from aqueous solutions using Ethanol as anti-solvent. The precipitation is accompanied by the addition of salt and ethanol to a solution containing DNA or RNA. In presence of monovalent cations (eg: Na⁺) ethanol efficiently precipitates nucleic acids. The precipitate can be collected by centrifugation.

Ethanol Precipitation of DNA



DNA surrounded by water molecules
(Hydration Shell) in solution



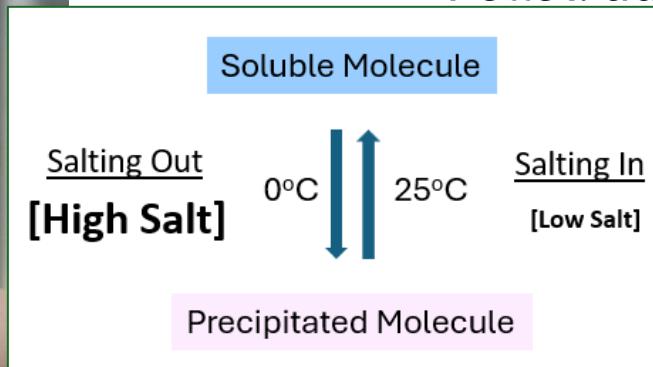
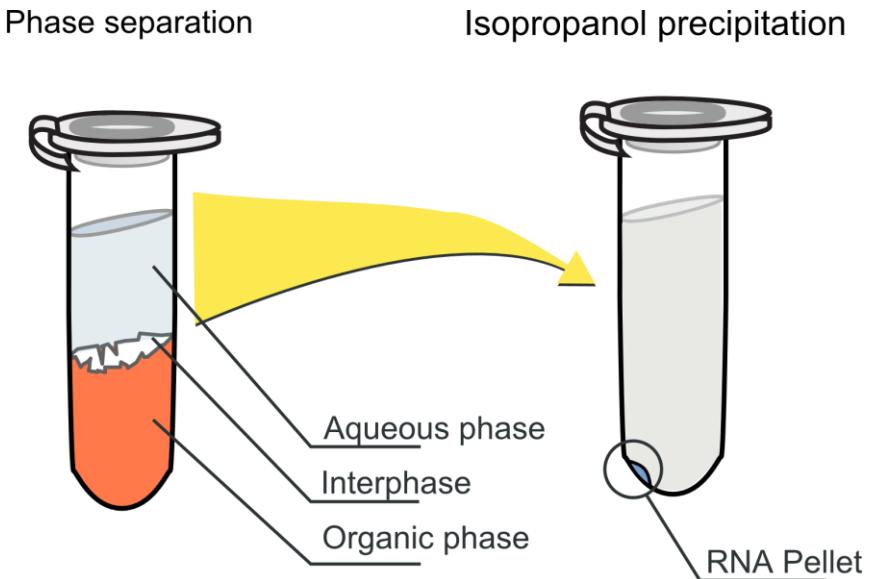
Sodium cations neutralizing a nucleotide.

Salt, Temperature, Alcohol

Sodium Acetate	Precipitate DNA (Na+ interacts with PO3-)
Ethanol/isopropanol	Helps in DNA Precipitation (Interacts with water by hydrogen bonds)
TE buffer	It dissolves DNA (Tris & EDTA)

When the cations and negatively charged nucleic acid backbone interact, nucleic acids are neutralized, therefore no longer dissolve in water and precipitate out of solution. Ethanol which has lower dielectric constant compared to water increases the interaction of the salt and the Coulomb force of attraction between the cations and the negatively charged nucleic acid backbone (that is, the resistance from the solvent's electric field sufficiently diminishes to permit efficient interaction; the solvation shells surrounding the solute's charges depletes).

Phenol-chloroform method



Samples (Blood, tissue, bacteria)



Pellet: add saturated phenol in Tris

Mix and spin – 12000rpm

Sup: add P:C:I = 25:24:1

Mix and spin – 12000rpm (optional)

Sup: add C:I = 24:1

(Precipitation) Mix and spin – 12000rpm

Sup: add 100% ethanol & sodium acetate = 24:1

(Washing 2 – 3 times) Mix and spin – 12000rpm, 0–8°C

Pellet: add 70% ethanol & sodium acetate = 24:1

Mix and spin – 12000rpm, 0–8°C

Dry pellet

Dissolve in TE buffer - Storage

Soft cell wall

DNA + debris

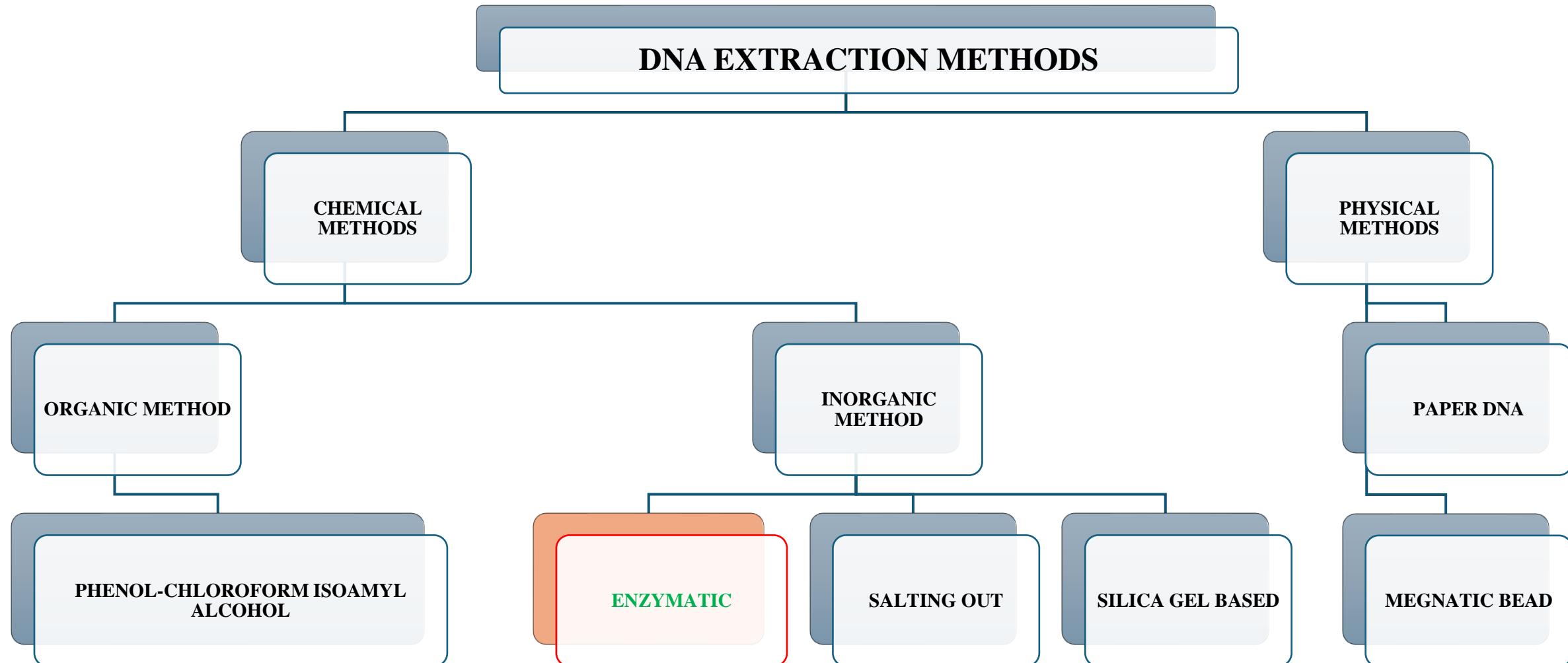
Help to get distinct organic & aqueous layers

Salt, Temperature, Alcohol, pH

Pros & Cons for Phenol-chloroform method

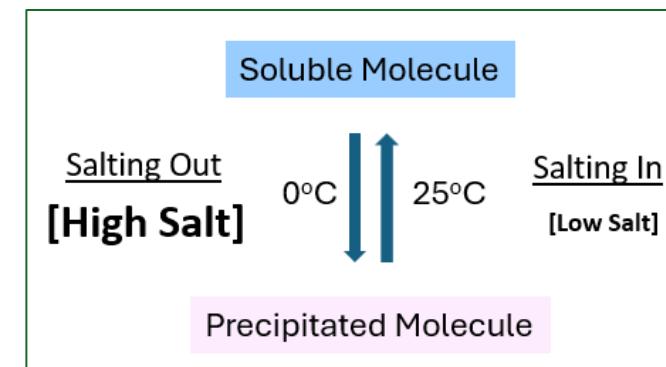
- Pros
 - Widely accepted.
 - Forensic departments trust the PCI method rather than the Kit method
 - Quantity of DNA obtained- very high (800 to 900 ng with great quality)
- Cons
 - If chemicals are prepared very well and performed it sincerely, result will be good by this method.
 - However, the amount of samples required for PCI DNA extraction is high.
 - It is difficult to isolate DNA from the samples such as hair and nail.
 - Also, the purity of DNA becomes a major issue if not performed well.

Different types of NA Extraction Methods



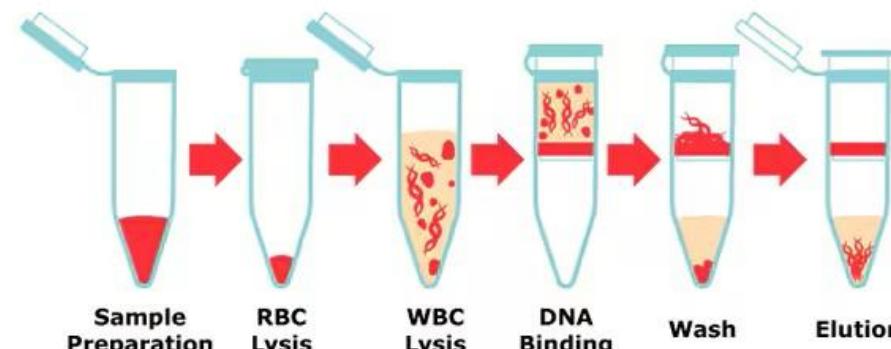
Enzymatic method of DNA extraction

- A combination of a salt method as well as an enzymatic method.
- Here the extraction buffer is used before going further on enzymatic digestion.
- Here phenol, chloroform or isoamyl alcohol is not used.
- Instead, the enzyme proteinase K is utilized for digesting the sample.
- The sample is incubated with proteinase K for 2 hours at 55-60°C (digest all the protein)
- Immediately after enzyme digestion, sample is centrifuged & precipitated by chilled alcohol.
- Finally, the DNA pellet is dissolved in **TE buffer**.
- This method of DNA extraction is **rapid and easy**.
- Even the yield is **very high**.
- However, the quality of DNA is a major concern for this method.

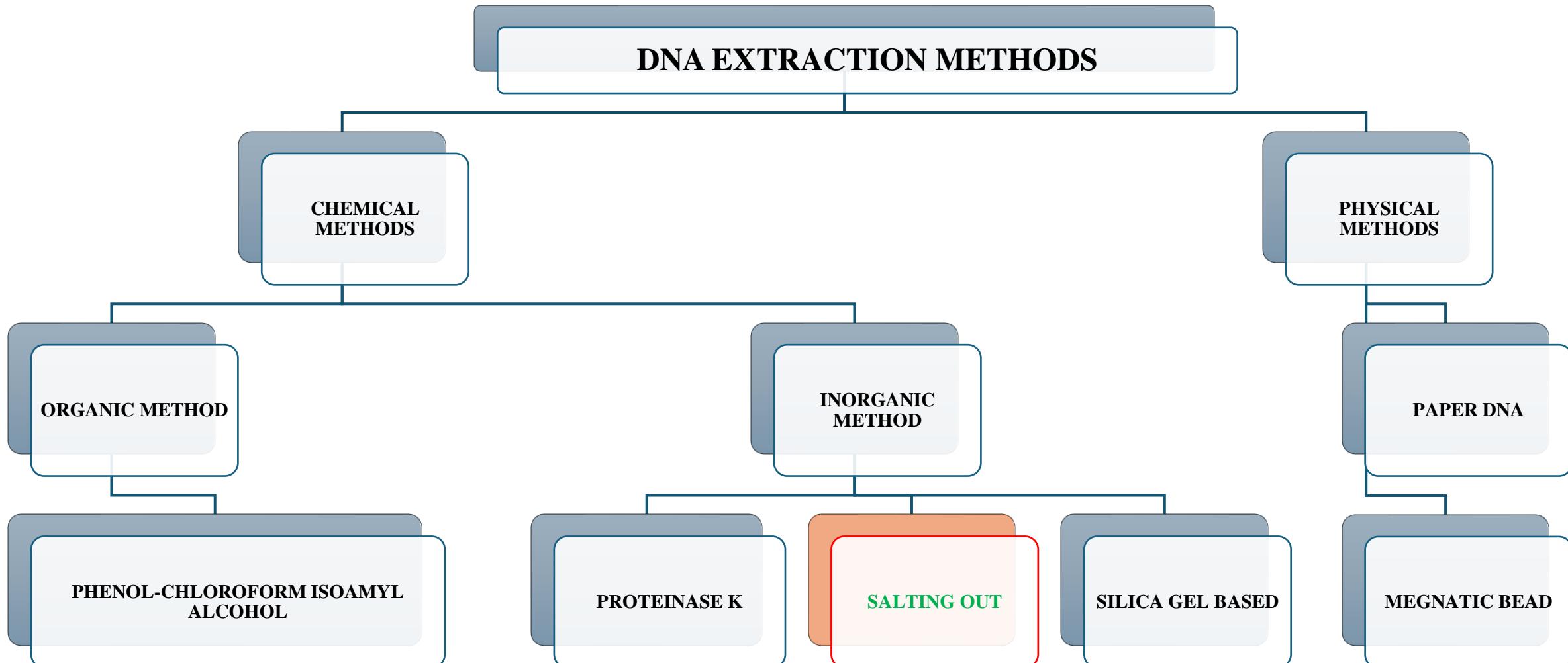


DNA extraction from blood protocol using proteinase K

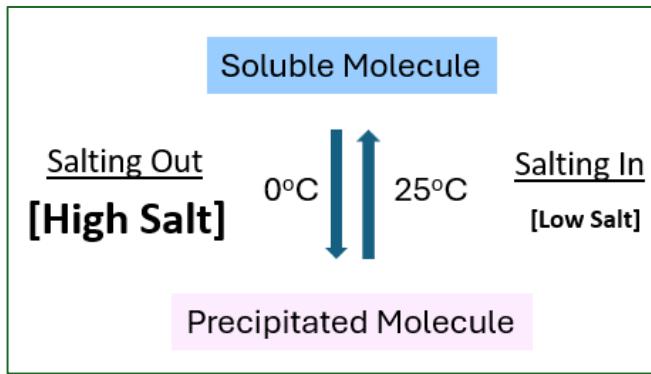
1. Take 2 mL of blood sample and add 10 to 20 µL of TE buffer to the sample, mix well.
2. Centrifuge the sample at 2,500 rpm for 20 minutes
3. Discard the supernatant and add 10 to 15 µL of TE buffer to the pellet and mix it gently.
4. Centrifuge the sample at 2,500 rpm for 15 minutes. Discard the supernatant and **add 20 µL of proteinase K solution** (commercially available ready to use) and **2 mL of DNA extraction buffer** to the Falcon tube.
5. **Incubate the sample at 56 °C to 60 °C for 1 to 2 hours** or until the pellet dissolve properly.
6. Now add 1 to 2 mL of cold isopropanol and a small amount of NaCl to the Falcon tube and invert the tube for some time to obtain the precipitate.
7. Centrifuge the tube at 8,000 to 10,000 rpm to settle the pellet at the bottom.
8. Discard the supernatant and add 1 mL ethanol and centrifuge it at 10,000 rpm for 1 to 2 minutes.
9. Discard the ethanol and dry the pellet in the dryer.
10. Add TE buffer (pH 8.0) to the pellet and put it in a water bath at 37 °C for 2 to 3 hours



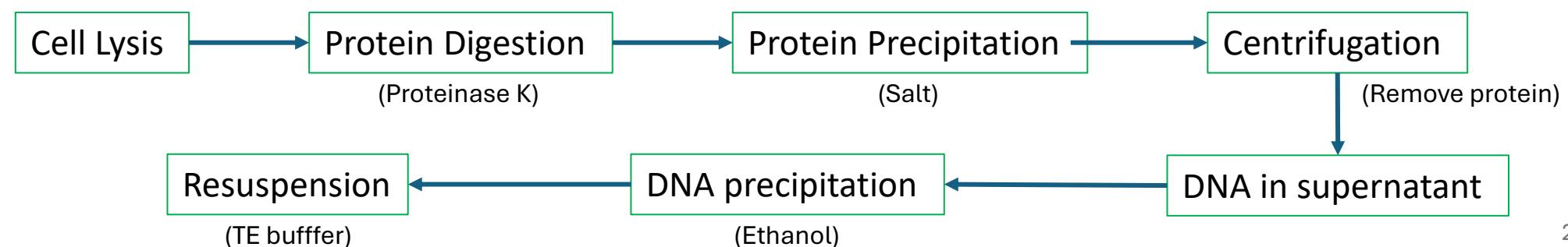
Different types of NA Extraction Methods



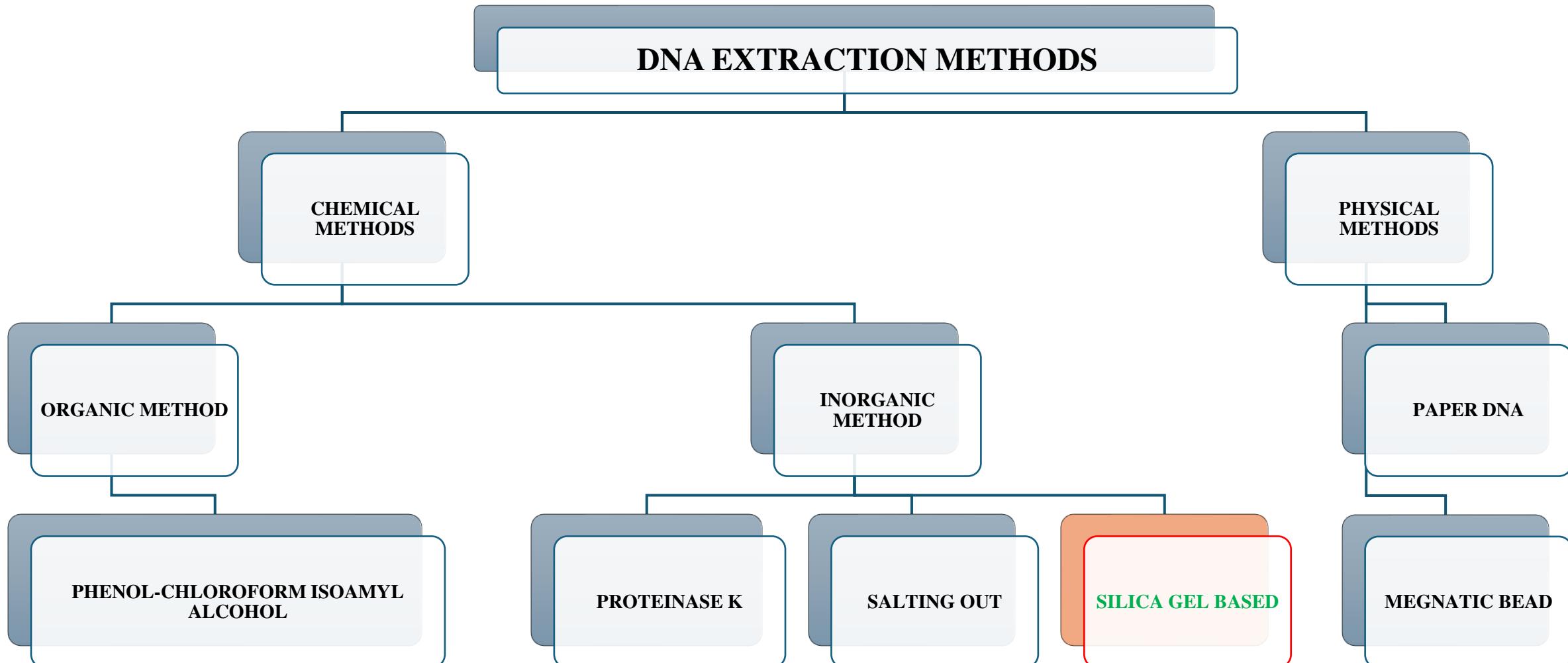
Salting out method



- At high salt concentration, proteins are dehydrated, lose solubility & precipitated (removed by centrifugation)
- The use of salts such as sodium chloride, potassium acetate & ammonium acetate helps in DNA extraction.
- DNA remains in supernatant
- However, the method gives excellent results in combination with proteinase K.
- Safer method than the PCl method.
- One of the major limitations of the salting-out method is the purity.
- Though enough yield can be obtained, the quality obtained might not be good.



Different types of NA Extraction Methods



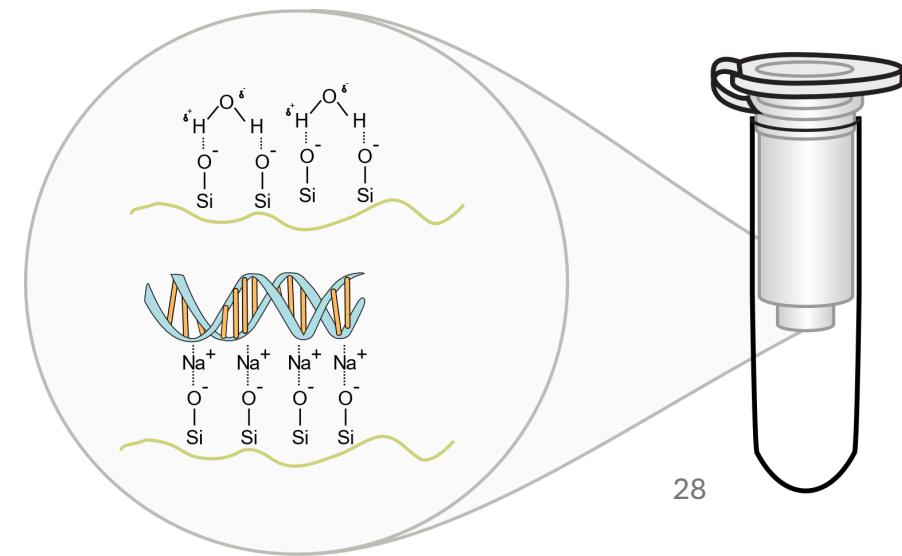
Solid-phase DNA extraction method

- Solid phase carrier adsorption method: spin column extraction method, glass bead adsorption method, silica matrix method, anion exchange method and nano magnetic bead extraction method
- Nowadays all the DNA extraction kits available are based on the unique chemistry of the solid/ liquid phase DNA extraction.
- Silica is a solid substance that binds with DNA during purification along with it
- The main advantage of silica gel-based DNA extraction is that it is rapid and gives "PCR ready DNA" for downstream applications
- No precipitation steps are required therefore this method is superior among all

Solid-phase DNA extraction method

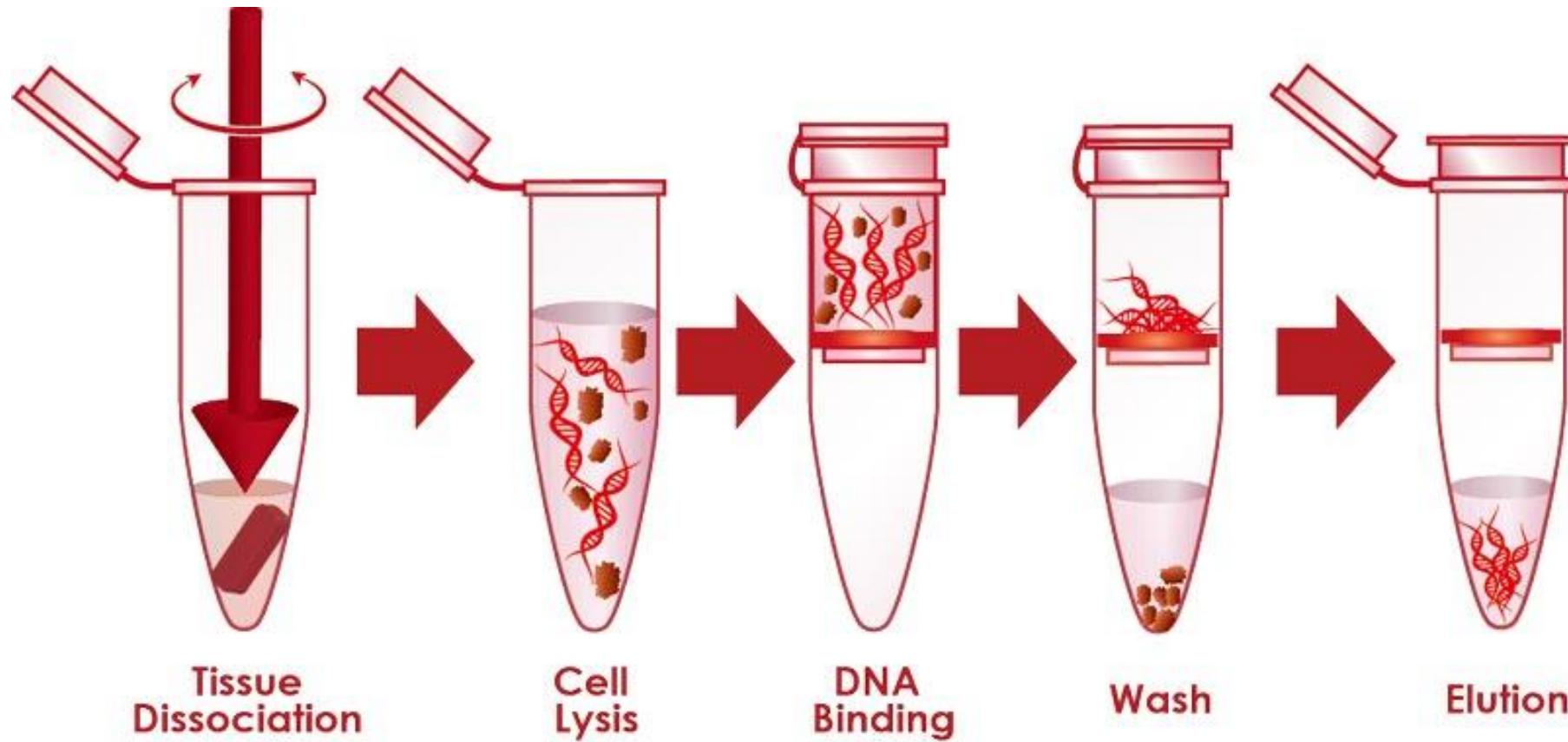
Silica column-based DNA extraction method

- Very unique and different from other DNA extraction methods.
- PCl or proteinase K method - Many times centrifugation, collection of aqueous phase or pellets depending upon the step of extraction.
- The silica-based DNA extraction method works on the unique chemistry of interaction between silica and DNA.
- A positively charged silica particles bind with the negatively charged DNA and hold it during centrifugation.
- The method was first described by McCormick in 1989.
- However, the idea was developed in 1979, when silica was used in DNA purification by Vogelstein.



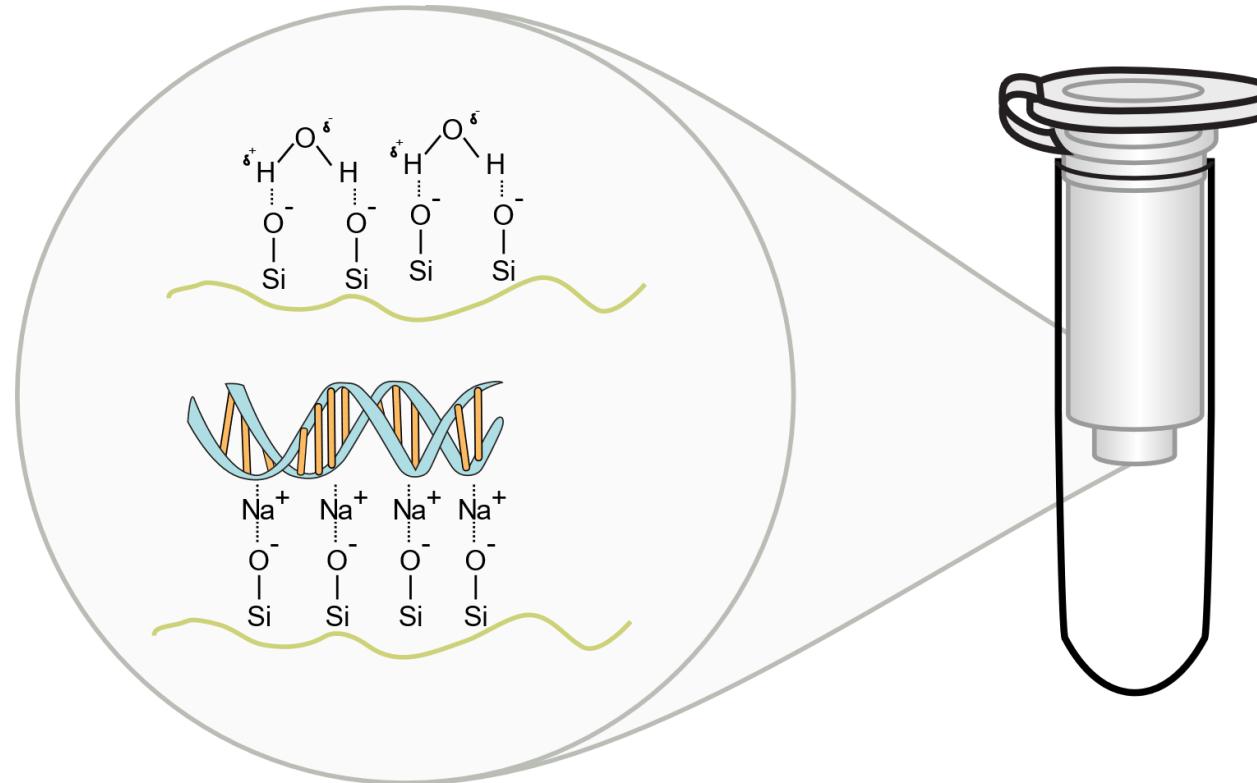
Solid-phase DNA extraction method

Silica column-based DNA extraction method



Solid-phase DNA extraction method

Silica column-based DNA extraction method

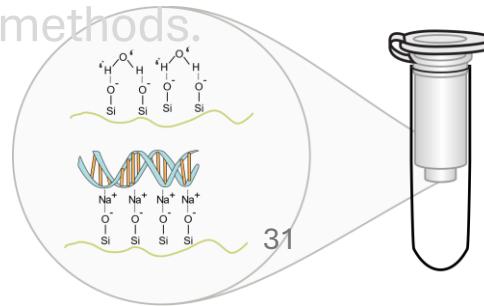


Silica in a spin column with water and with DNA sample in chaotropic buffer

Solid-phase DNA extraction method

Silica column-based DNA extraction method

- Commercially available method and it is most routinely used in diagnostic laboratories.
- Widely accepted - Good quality DNA yield & minimal simple operating system.
- The lysis buffer breaks the cell membrane and the nuclear envelope.
- The proteinase K digests all the protein.
- All the other impurities are removed by centrifugation.
- Here the DNA remains bounded with silica and other impurities pass through the silica column.
- Now the DNA can be washed twice for improving the purity.
- The impurities collected in collection tube are discarded.
- Finally, the DNA is dissolved into the TE buffer.
- The method is fast reliable, accurate and consumes less time as compared to other methods.



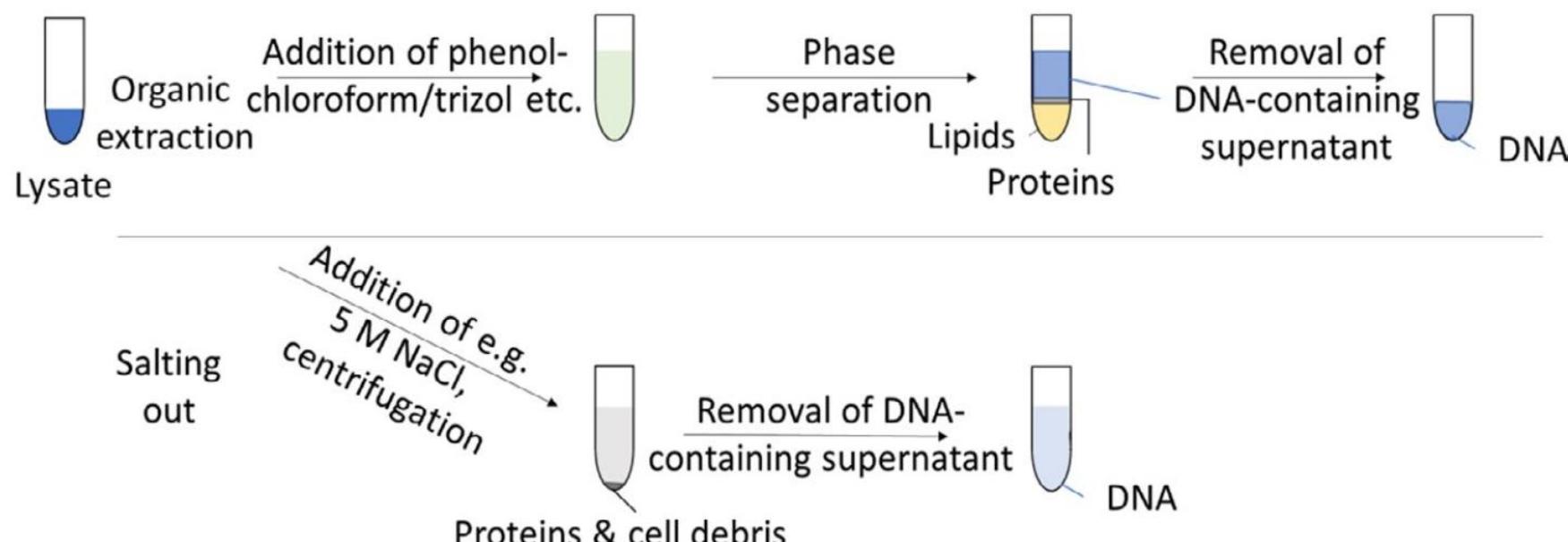
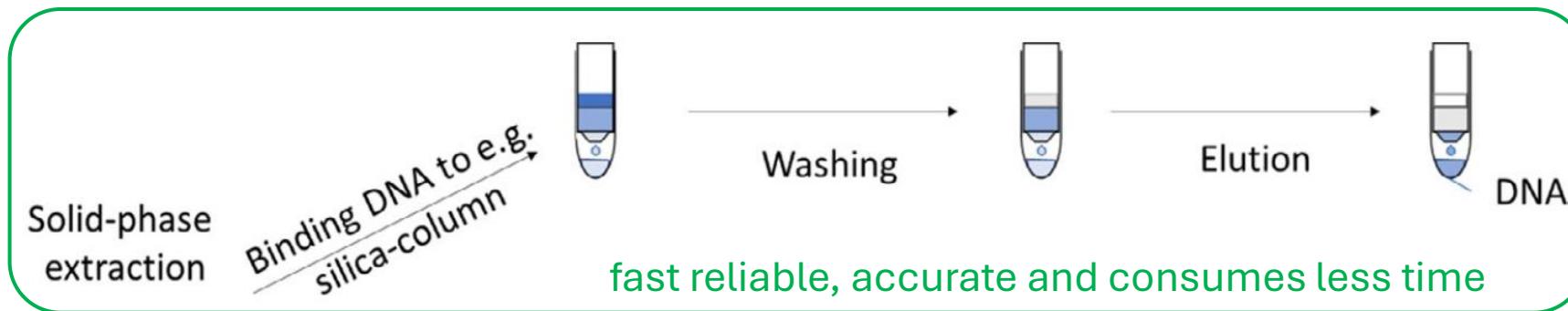
Solid-phase DNA extraction method

Silica column-based DNA extraction method: Lysis Buffer (LB)

- Conventional lysis solutions: detergents (such as SDS, Triton X-100, NP-40, Tween 20, etc.) and salts (such as Tris, EDTA, NaCl, etc.)
- Extraction buffer (20mM EDTA pH 8, 100mM Tris HCL pH 8 and 1.4M NaCl), PBS, Cetyltrimethylammonium bromide (CTAB).
- Chloroform, Isoamyl alcohol, Phenol
- Chaotropic Agents: urea (6–8 M), guanidine HCL (6 M) and lithium perchlorate (4.5 M), Guanidine thiocyanate.
 - **Guanidine thiocyanate** is a potent protein denaturant that is stronger than guanidine HCl and is often used during the **isolation of intact ribonucleic acid to eliminate RNase activity**
- *Lysozyme, Lysostaphin, Mutanolysin*
- *Sodium lauroyl sarcosinate, Polyvinylpolypyrrolidone, Sodium Metabisulfite and 2-Mercaptoethanol*
- Guanidinium thiocyanate-phenol/chloroform/2-mercaptoethanol, acidic guanidine thiocyanate-phenol mixture with chloroform (RNA).

Solid-phase DNA extraction method

Silica column-based DNA extraction method

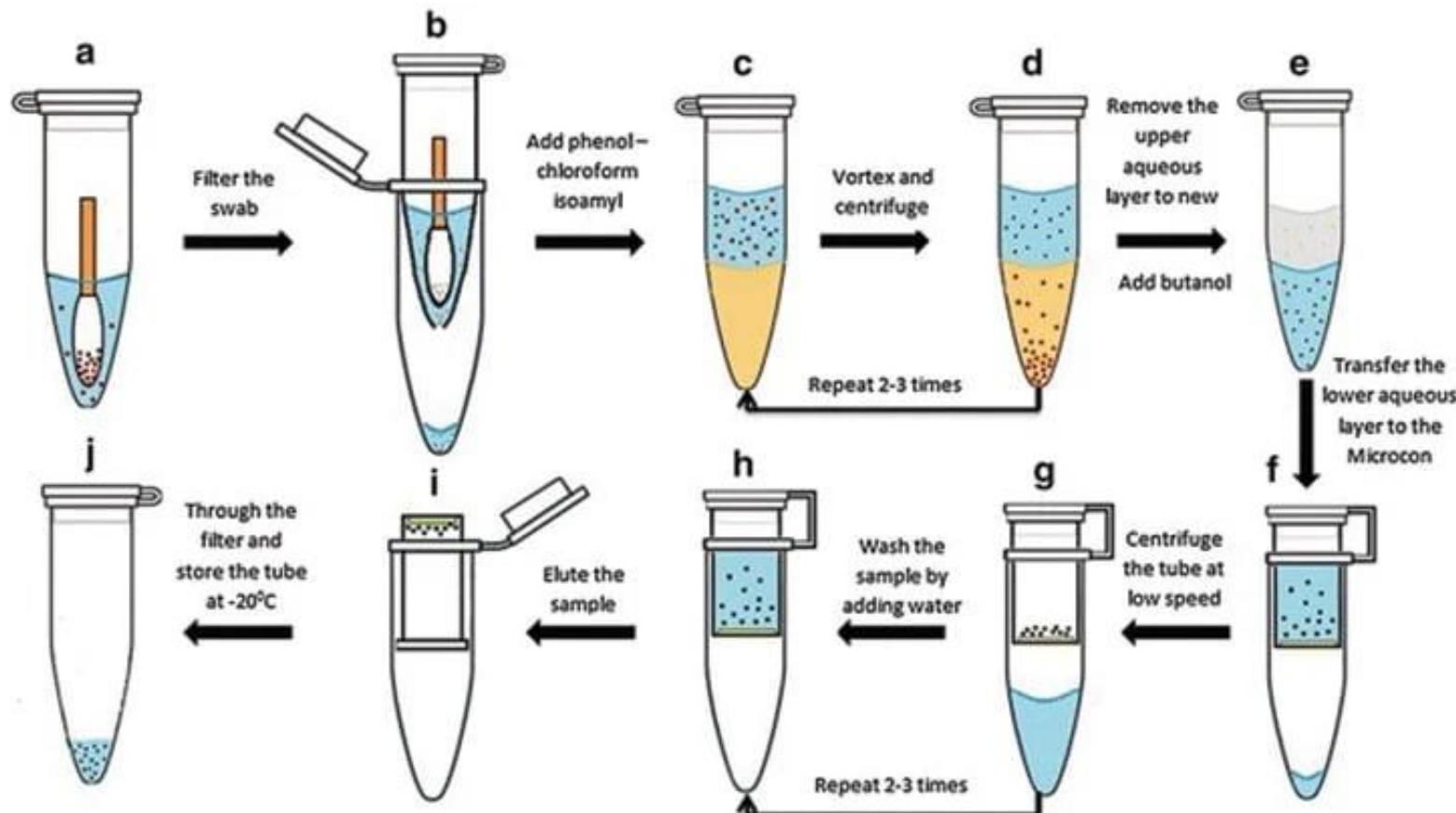


Overview of common DNA extraction protocols used on tissue lysate

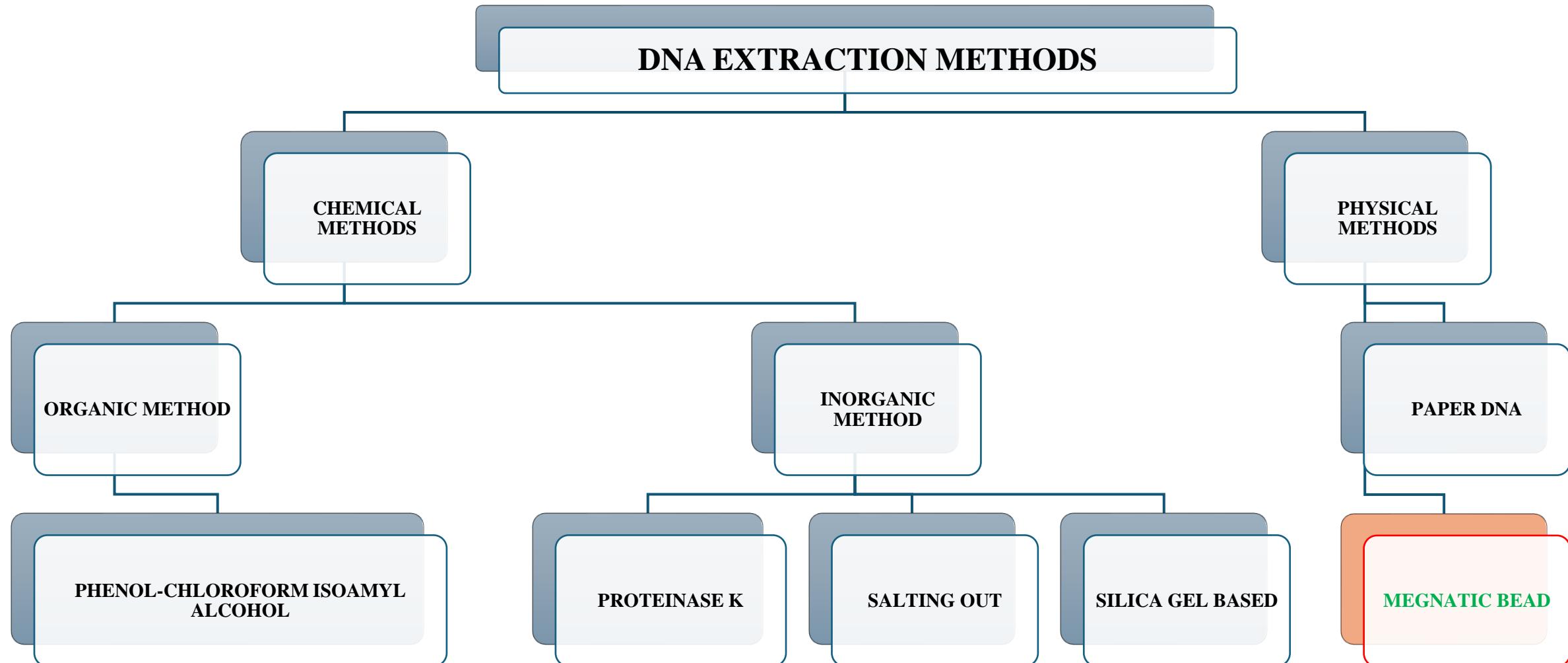
Solid-phase DNA extraction method

Silica column-based DNA extraction method

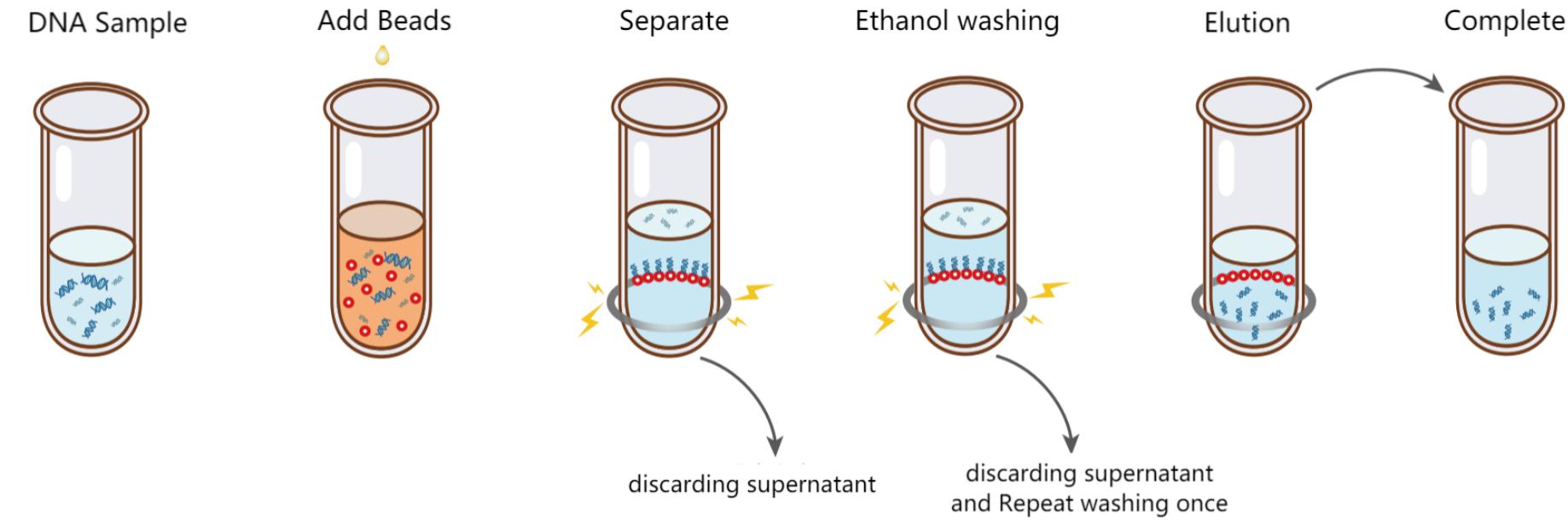
Phenol-chloroform extraction of prokaryotic DNA



Different types of NA Extraction Methods

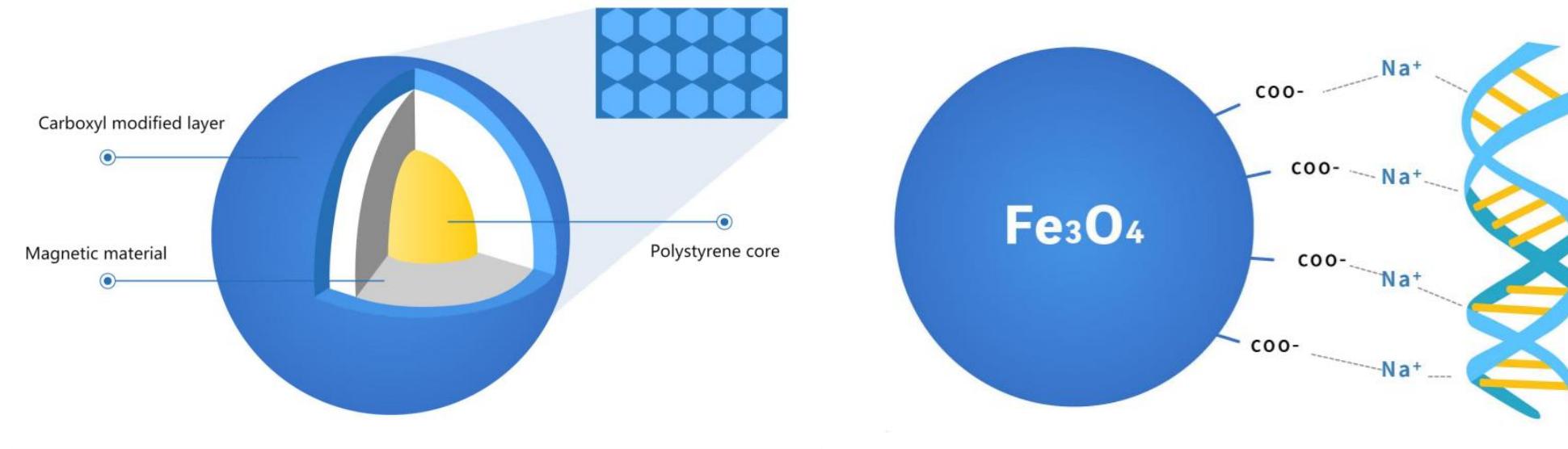


DNA extraction by magnetic beads

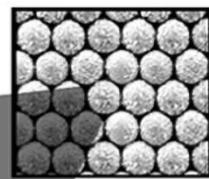
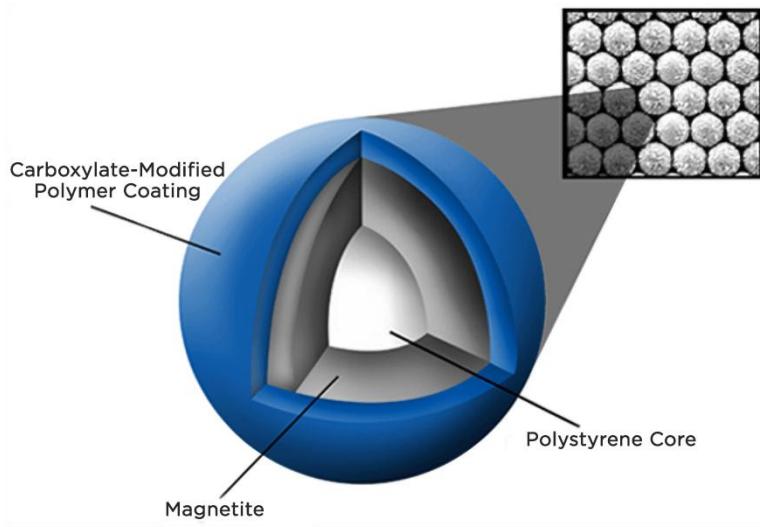


Solid-phase DNA extraction method

DNA extraction by magnetic beads



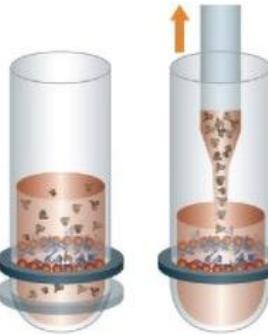
DNA extraction by magnetic beads



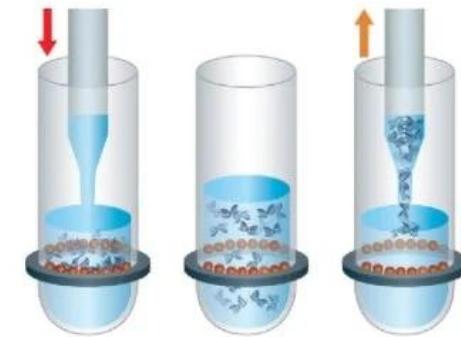
Each SPRI bead is $1 \mu\text{m} \pm 8\%$ in size, and consists of a buoyant polystyrene core surrounded by a thin layer of magnetite making it paramagnetic and easily manipulable with a magnetic field. On the surface, SPRI beads are coated by carboxyl molecules providing the charge groups for DNA binding.



STEP 1: BINDING

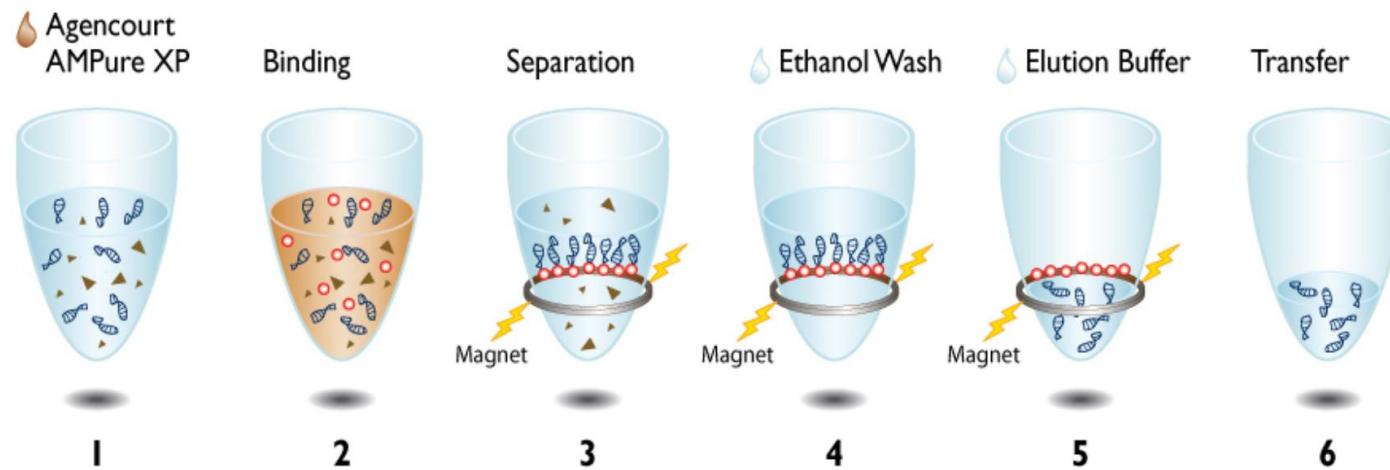


STEP 2: WASHING



STEP 3: ELUTION

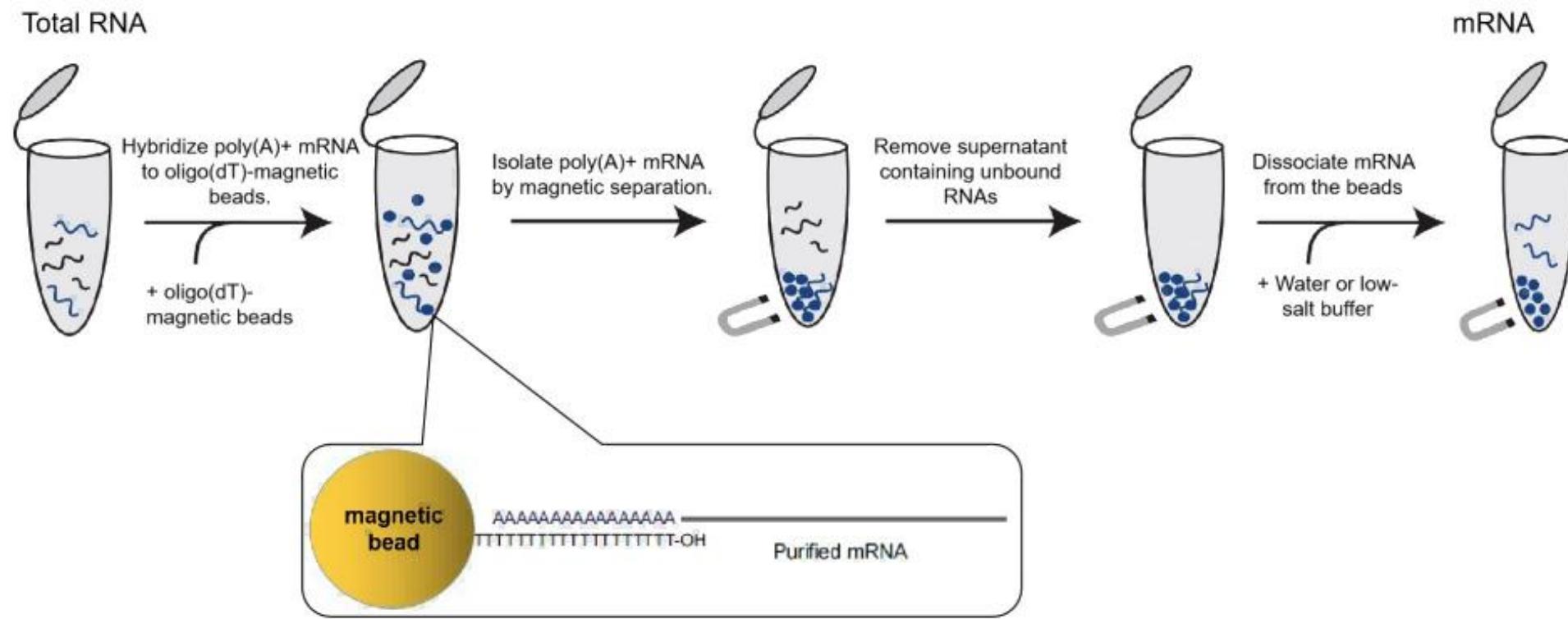
DNA extraction by magnetic beads



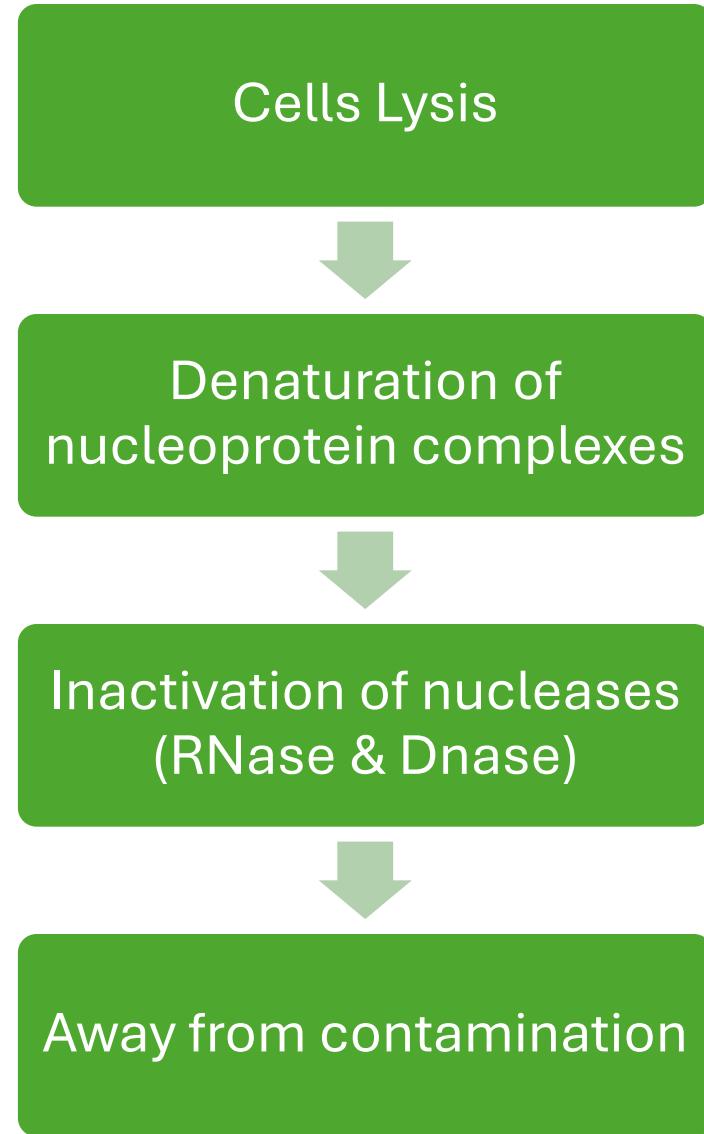
The workflow for the PCR purification process is as follows:

1. Add 1.8 µL AMPure XP per 1.0 µL of sample.
2. Bind DNA fragments to paramagnetic beads.
3. Separation of beads + DNA fragments from contaminants.
4. Wash beads + DNA fragments twice with 70% Ethanol to remove contaminants.
5. Elute purified DNA fragments from beads.
6. Transfer to new plate

DNA extraction by magnetic beads



DNA Extraction



Salt, Temperature, Alcohol, pH

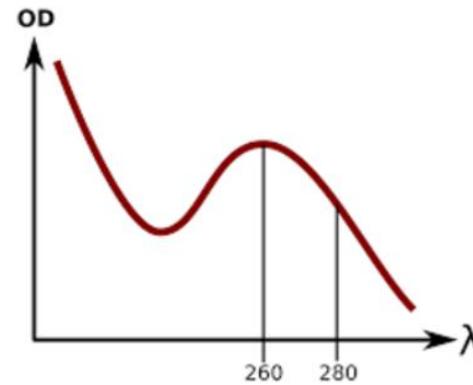
Nucleic Acid-extracted QC

Parameters of purified DNA

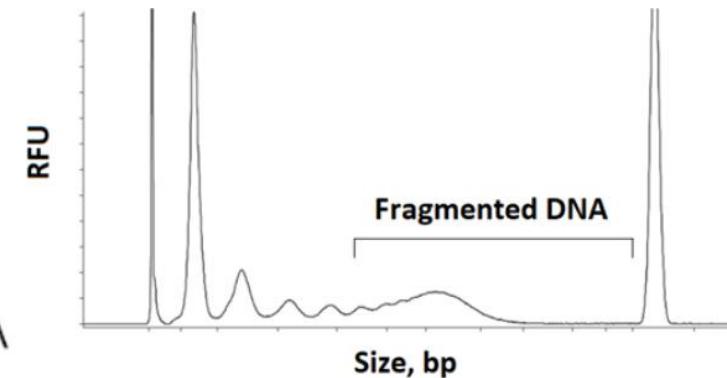
Concentraion
Qubit Fluorometry



Quality
Nanodrop Adsorption
Spectrum



Fragmentation
Agilent TapeStation
Capillary Electrophoresis



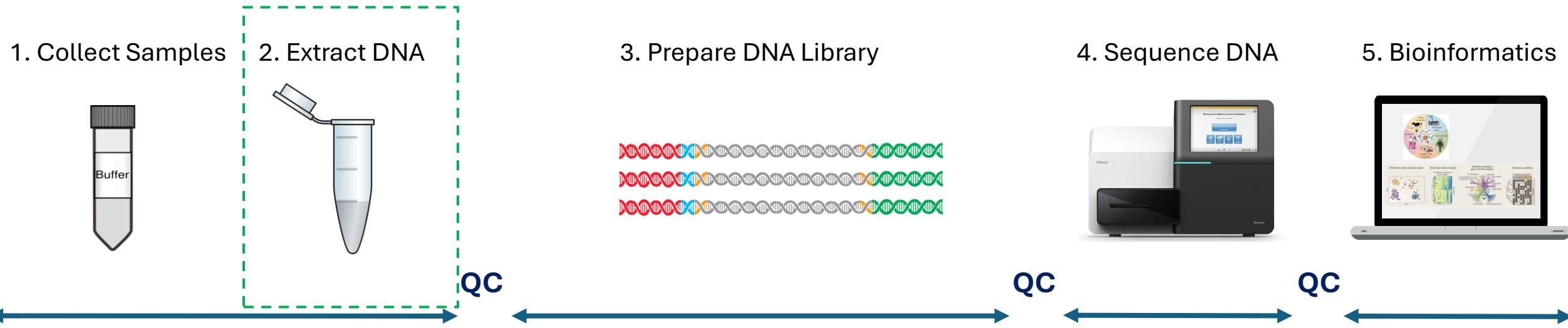
Additional purification steps were applied to some of the resulting DNA samples to follow the removal of contaminants and reduction of eukaryotic DNA load. All samples were evaluated for an indicated set of parameters to select the best DNA purification strategy

**THANK YOU FOR YOUR
ATTENTION!**

Special Topics for Nucleic Acid Extraction

- FFPE Tissue

Research Workflow

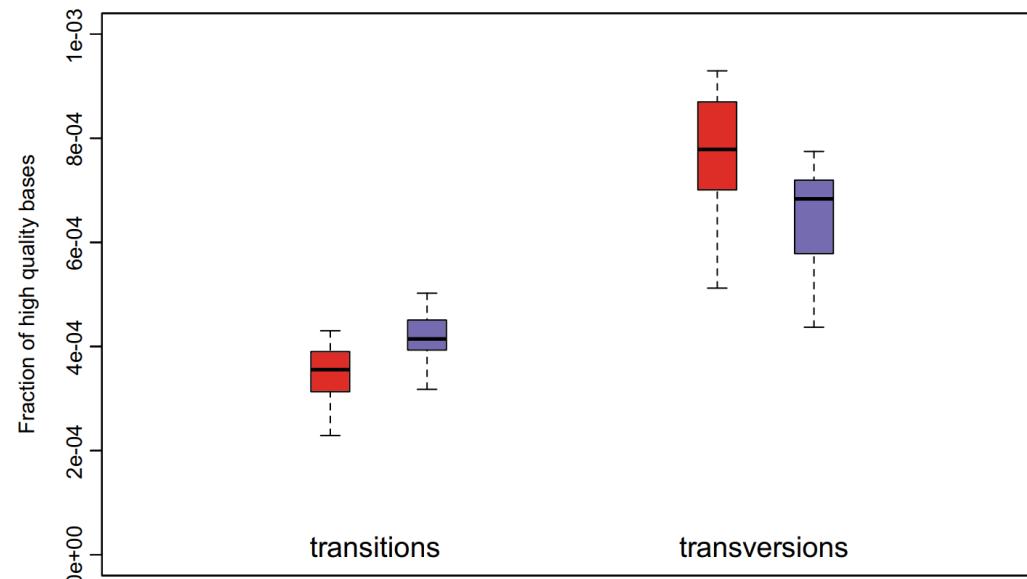


FFPE Tissue

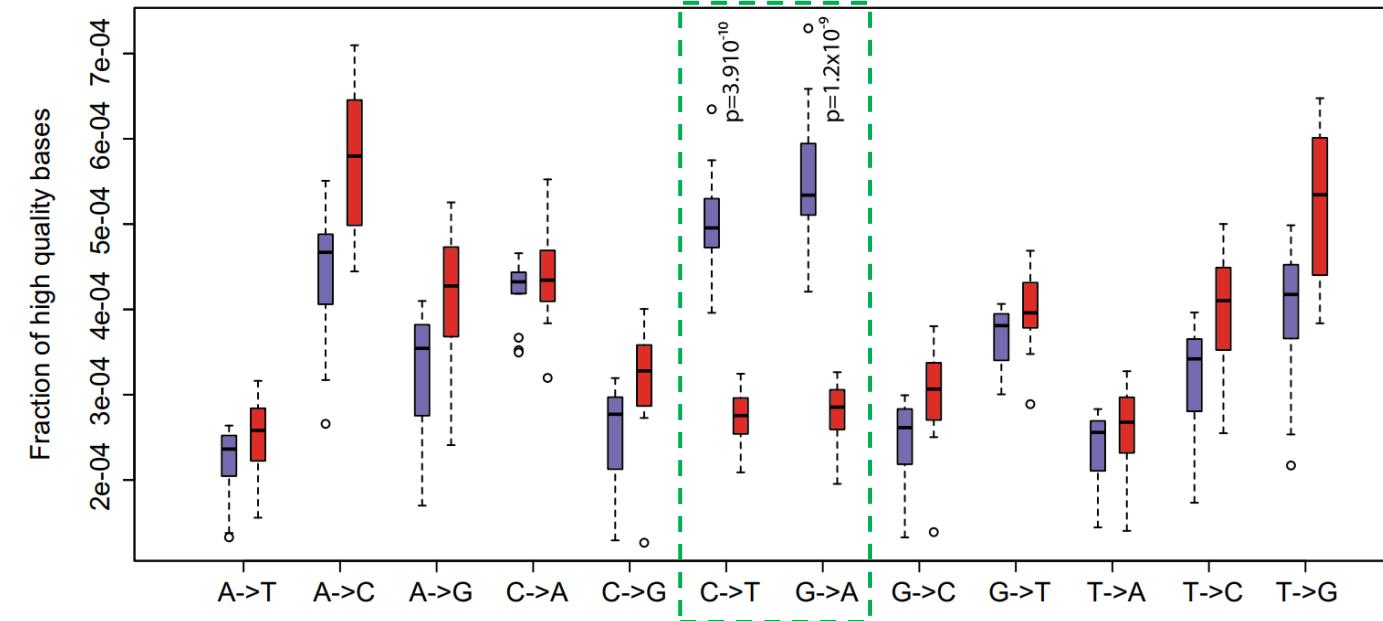
- DNA Degradation and Fragmentation
 - Extensive crosslinks between DNA and proteins, leading to DNA fragmentation and reduced amplifiable templates
 - Harsh paraffin removal and storage conditions over time can exacerbate DNA damage, resulting in shorter fragments.
- Over-Fixation and Variation in Fixation Times
 - Tissues fixed for prolonged periods (over 24–48 hours) can exhibit more extensive crosslinking, which makes extraction and downstream analyses more difficult.
 - Variability in clinical workflows means fixation times may not always be standardized.
- Low Yield and Purity
 - FFPE tissues often yield lower amounts of DNA due to sample age, type of tissue, or repeated block usage.
 - Contaminants such as residual paraffin, proteins, and chemicals can co-elute with DNA.
- Sample Heterogeneity and Tissue Quality
 - Different tissue types (e.g., fibrous vs. fatty tissues) may respond differently to standard protocols.
 - Necrotic or highly calcified areas can reduce DNA quality

FFPE Tissue

Transitions and Transversions



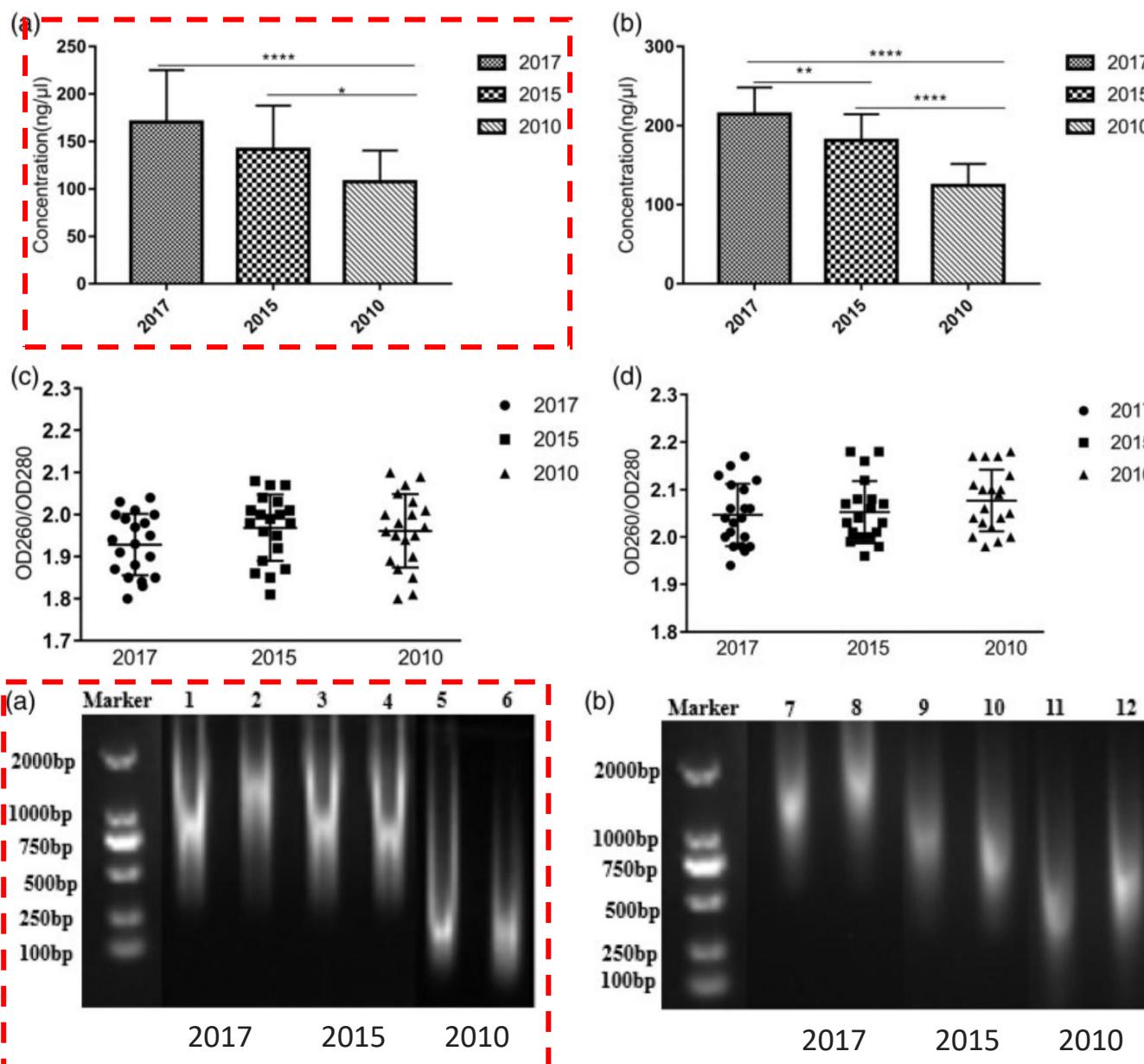
High Quality (Q>20) Non-Variant Base Changes



Observed spectrum of high-quality base changes in FFPE and frozen NGS data.

A: Distributions of the mean transition and transversion frequency for frozen (red) versus FFPE (purple) samples among high-quality discrepancies. The ratio of transitions to transversions was significantly larger in FFPE compared with fresh tissue ($P \approx 6.4 \times 10^{-8}$). B: Distributions of the mean frequencies for each possible base change for each sample type. Only C to T and G to A transitions were significantly different between fresh and frozen samples ($P \approx 3.9 \times 10^{-10}$ and 1.2×10^{-9} , respectively). In both panels, box plots display the median and interquartile range (IQR) of the per-sample mean frequency for each base change by sample type, with whiskers extending to the last data point within 1.5 times the IQR and outliers indicated by circles.

FFPE Tissue



Concentration and purity of DNA and RNA extracted from FFPE tissues stored for different lengths of time.

- (a) Concentration of DNA.
- (b) Concentration of RNA ($n \geq 20$).
- (c) Purity of DNA.
- (d) Purity of RNA ($n \geq 20$).

*P < 0.05, **P < 0.01, and ***P < 0.0001.

Degradation of DNA and RNA extracted from FFPE tissues stored for different lengths of time.

(a) Agarose gel electrophoresis of DNA. Lanes 1 and 2: FFPE tissues prepared in 2017; lanes 3 and 4: FFPE tissues prepared in 2015; lanes 5 and 6: FFPE tissues prepared in 2010.

(b) Agarose gel electrophoresis of RNA. Lanes 7 and 8: FFPE tissues prepared in 2017; lanes 9 and 10: FFPE tissues prepared in 2015; lanes 11 and 12: FFPE tissues prepared in 2010.

Troubleshooting for FFPE Tissue

- DNA Degradation and Fragmentation

- Extensive crosslinks between DNA and proteins, leading to DNA fragmentation and reduced amplifiable templates

Sample Type	Sample Management
FFPE	<ul style="list-style-type: none">• Importance of quick tissue removal and fixation• Use of 10% formalin solution (3.7% formaldehyde, 1–1.5% methanol)• Optimal use of neutral-buffered formalin• Ratio of formalin to tissue: at least 10:1• Avoid overfixation: no more than 24 hours• Use <u>fresh alcohol</u> and <u>xylene</u> for optimal results• Avoid additives in paraffin• Storage: RT for years• Transport: RT

FFPE
standardization
process

- Sample heterogeneity and tissue quality

- Different tissue types (e.g., fibrous vs. fatty tissues) may respond differently to standard protocols.
- Necrotic or highly calcified areas can reduce DNA quality

Troubleshooting for FFPE Tissue

Bioinformatics Step scientific reports



OPEN

DEEPOMICS FFPE, a deep neural network model, identifies DNA sequencing artifacts from formalin fixed paraffin embedded tissue with high accuracy

Dong-hyuk Heo, Inyoung Kim, Heejae Seo, Seong-Gwang Kim, Minji Kim, Jiin Park, Hongsil Park, Seungmo Kang, Juhee Kim, Soonmyung Paik & Seong-Eui Hong✉

Formalin-fixed, paraffin-embedded (FFPE) tissue specimens are routinely used in pathological diagnosis, but their large number of artifactual mutations complicate the evaluation of companion diagnostics and analysis of next-generation sequencing data. Identification of variants with low allele frequencies is challenging because existing FFPE filtering tools label all low-frequency variants as artifacts. To address this problem, we aimed to develop DEEPOMICS FFPE, an AI model that can classify a true variant from an artifact. Paired whole exome sequencing data from fresh frozen and FFPE samples from 24 tumors were obtained from public sources and used as training and validation sets at a ratio of 7:3. A deep neural network model with three hidden layers was trained with input features using outputs of the MuTect2 caller. Contributing features were identified using the SHapley Additive exPlanations algorithm and optimized based on training results. The performance of the final model (DEEPOMICS FFPE) was compared with those of existing models (MuTect filter, FFPolish, and SOBDetector) by using well-defined test datasets. We found 41 discriminating properties for FFPE artifacts. Optimization of property quantification improved the model performance. DEEPOMICS FFPE removed 99.6% of artifacts while maintaining 87.1% of true variants, with an F1-score of 88.3 in the entire dataset not used for training, which is significantly higher than those of existing tools. Its performance was maintained even for low-allele-fraction variants with a specificity of 0.995, suggesting that it can be used to identify subclonal variants. Different from existing methods, DEEPOMICS FFPE identified most of the sequencing artifacts in the FFPE samples while retaining more of true variants, including those of low allele frequencies. The newly developed tool DEEPOMICS FFPE may be useful in designing capture panels for personalized circulating tumor DNA assay and identifying candidate neoepitopes for personalized vaccine design. DEEPOMICS FFPE is freely available on the web (<http://deepomics.co.kr/ffpe>) for research.

**THANK YOU FOR YOUR
ATTENTION!**

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 ^3H . 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 T₁, which had released the sequence, establish that MS2 RNA ends in ...GpUpUpApCpCpApCpCpA.

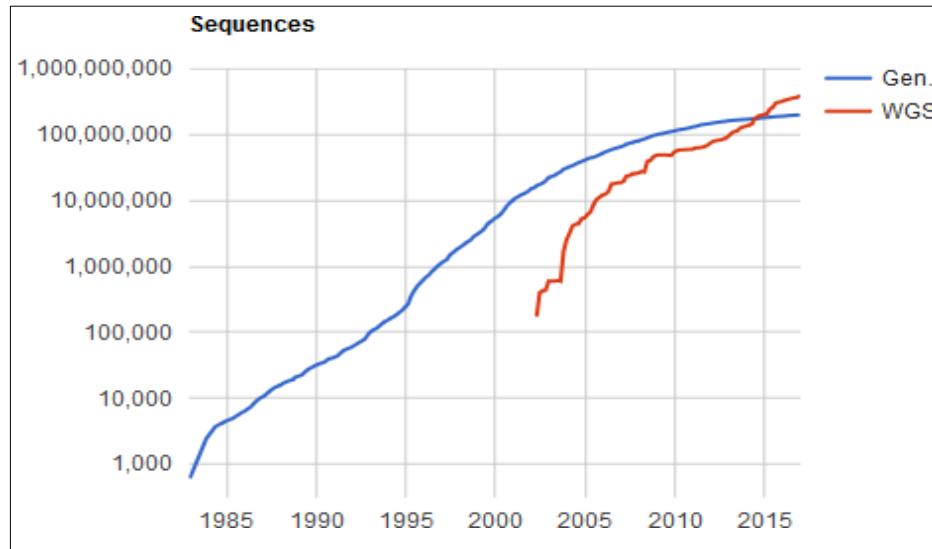
It is suggested that the termination signal for the translation into polypeptides

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**

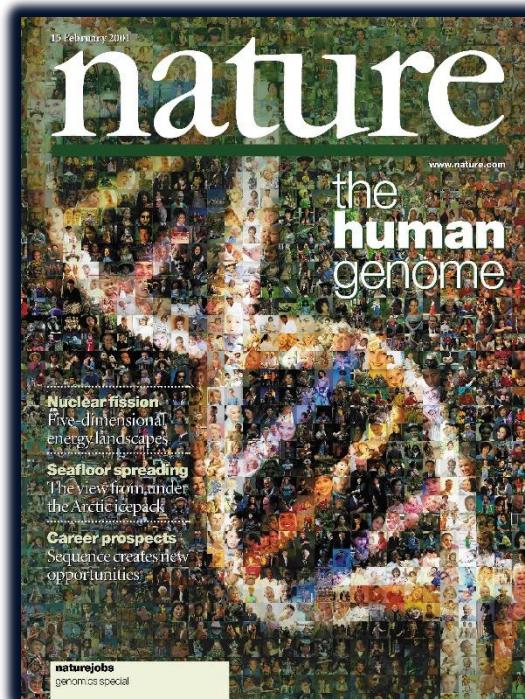


Release	Date	GenBank		WGS	
		Bases	Sequences	Bases	Sequences
3	Dec 1982	680338	606		
14	Nov 1983	2274029	2427		
20	May 1984	3002088	3665		
216	Oct 2016	220731315250	197390691	1676238489250	363213315
217	Dec 2016	224973060433	198565475	1817189565845	395301176

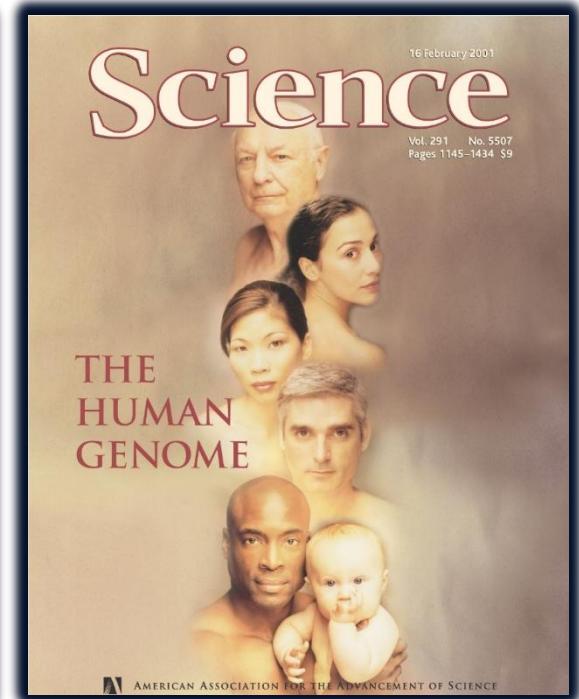
History and future of DNA sequencing

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

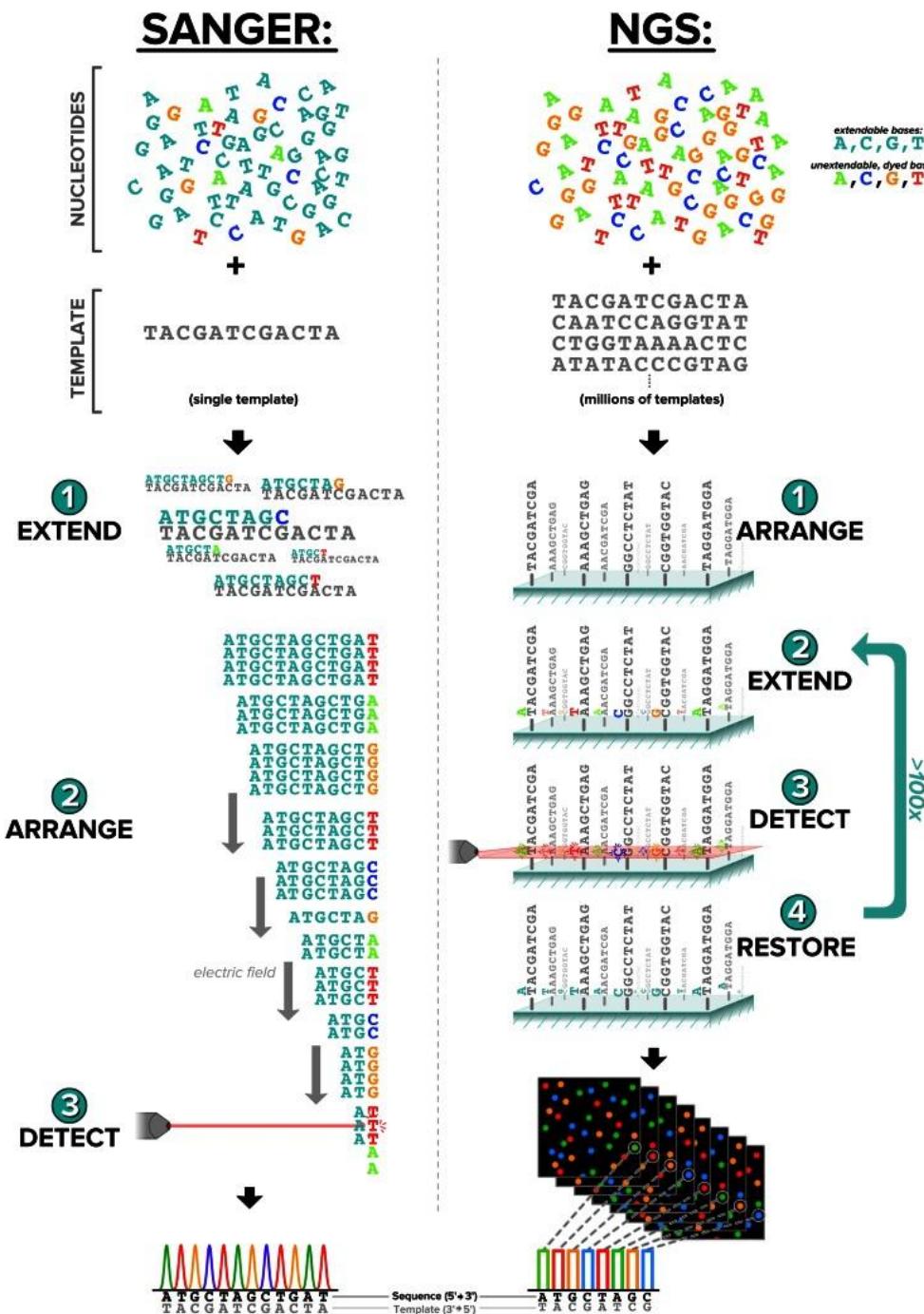
History and future of DNA sequencing

- 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?**

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

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

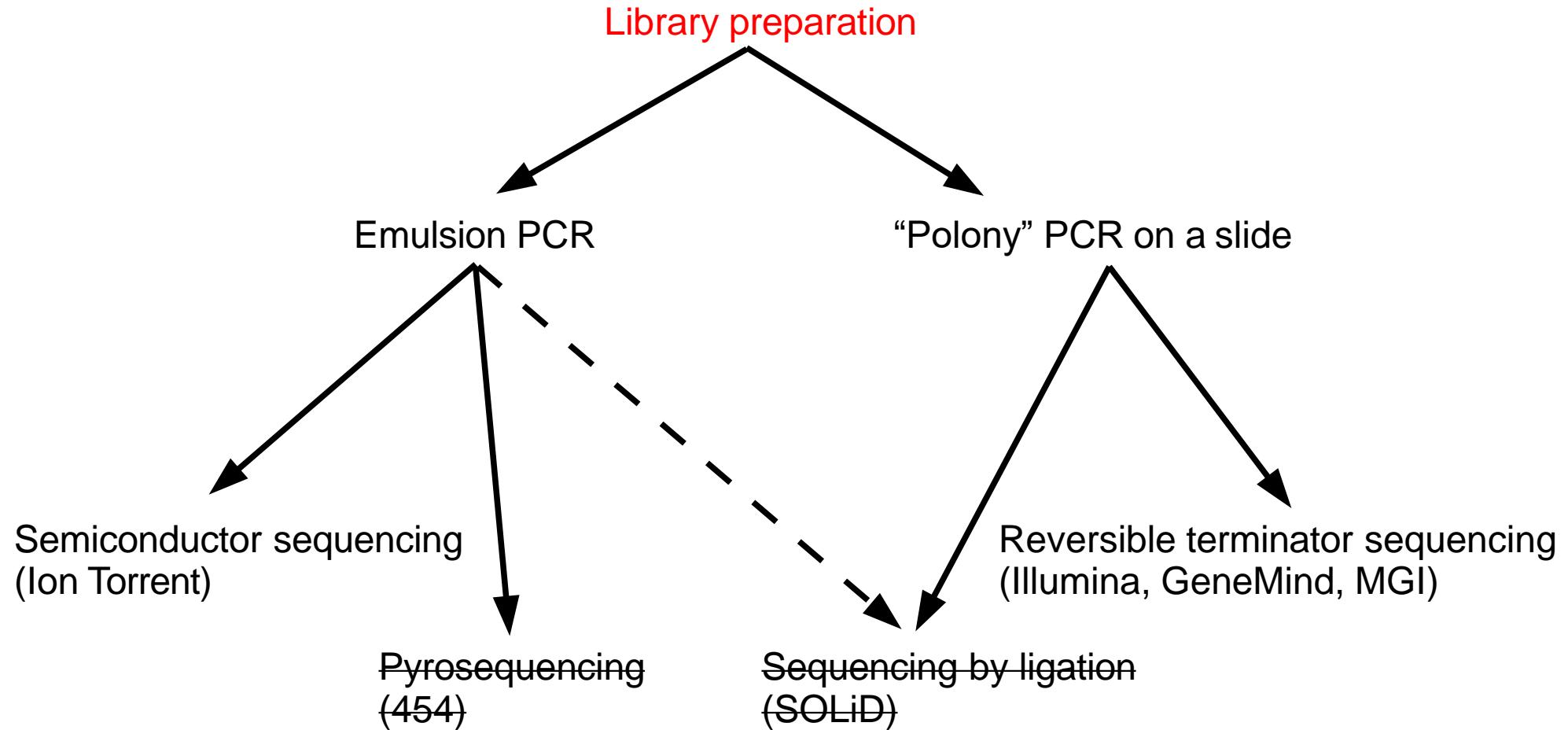
Next Generation Sequencing
Amplified Single Molecule Sequencing

- 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

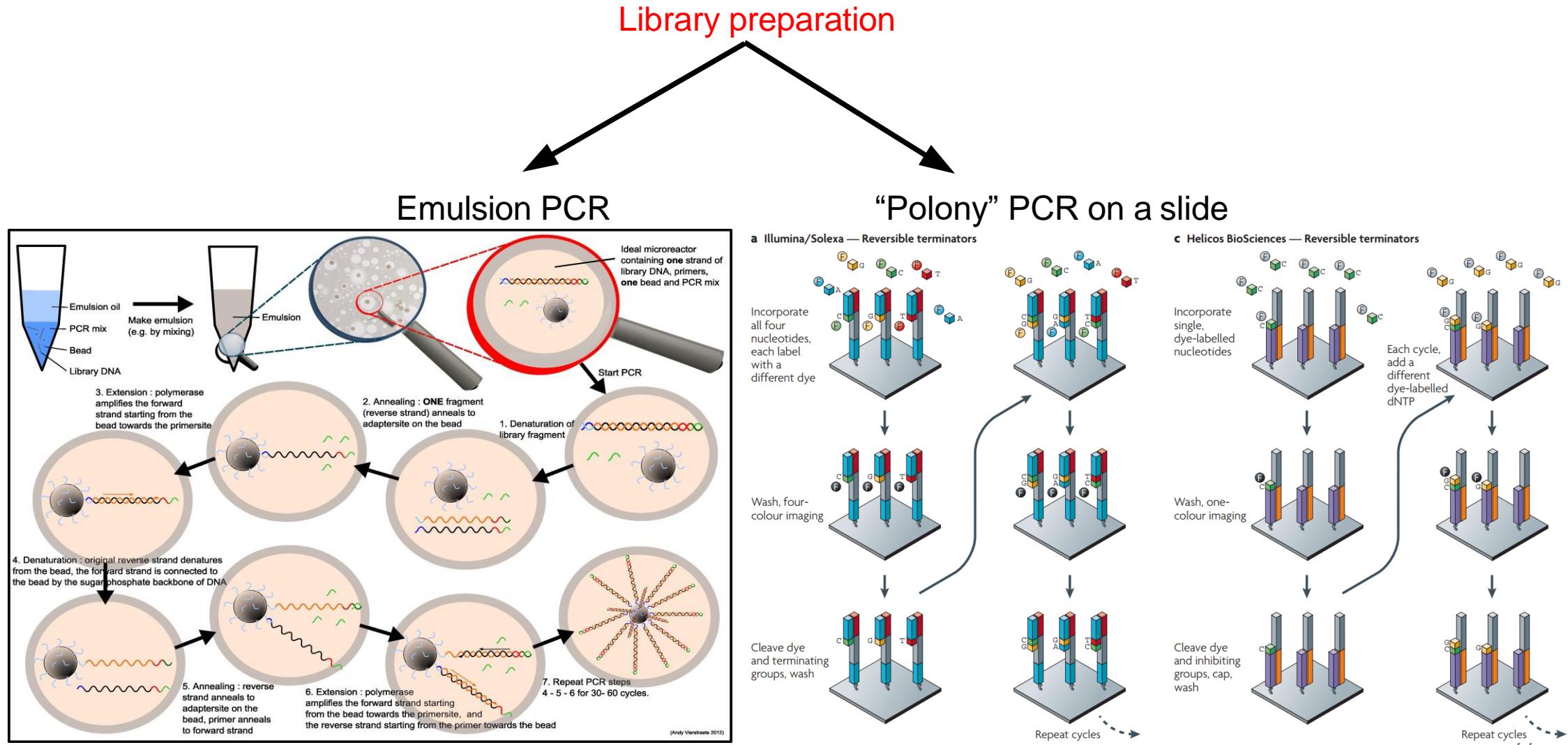
General Workflow

Next Generation Sequencing: Amplified Single Molecule Sequencing



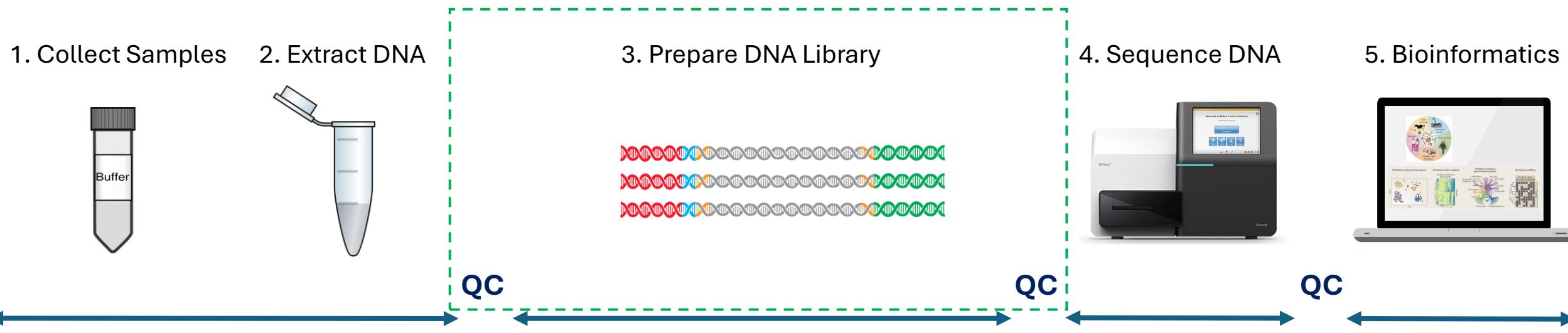
General Workflow

Next Generation Sequencing: Amplified Single Molecule Sequencing



**THANK YOU FOR YOUR
ATTENTION!**

Library Preparation Approaches



DNA Library Prep

QC of DNA Input

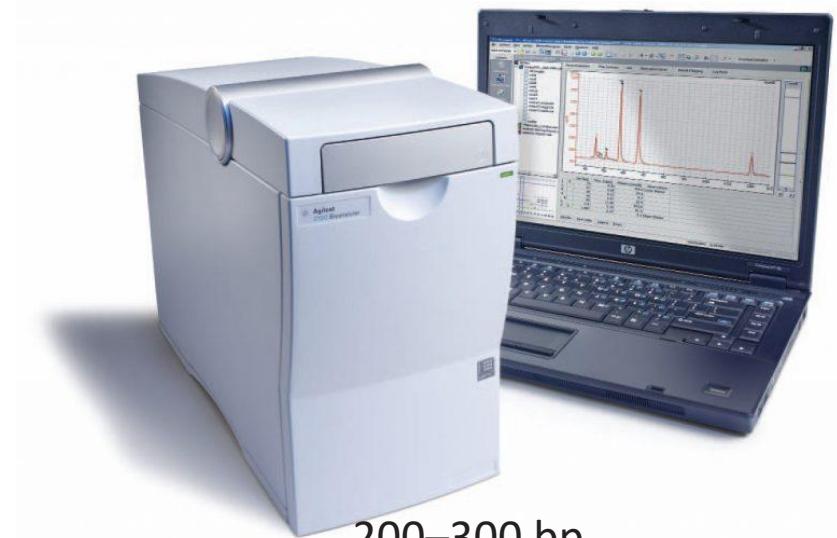
- The amount of input DNA: between 100 ng and 1 µg depending on the kit or protocol being used.
- Genomic DNA concentrations for Illumina libraries: Qubit
- Distribution of fragments: on average, ~200–300 bp and range ~100–1,000 bp



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

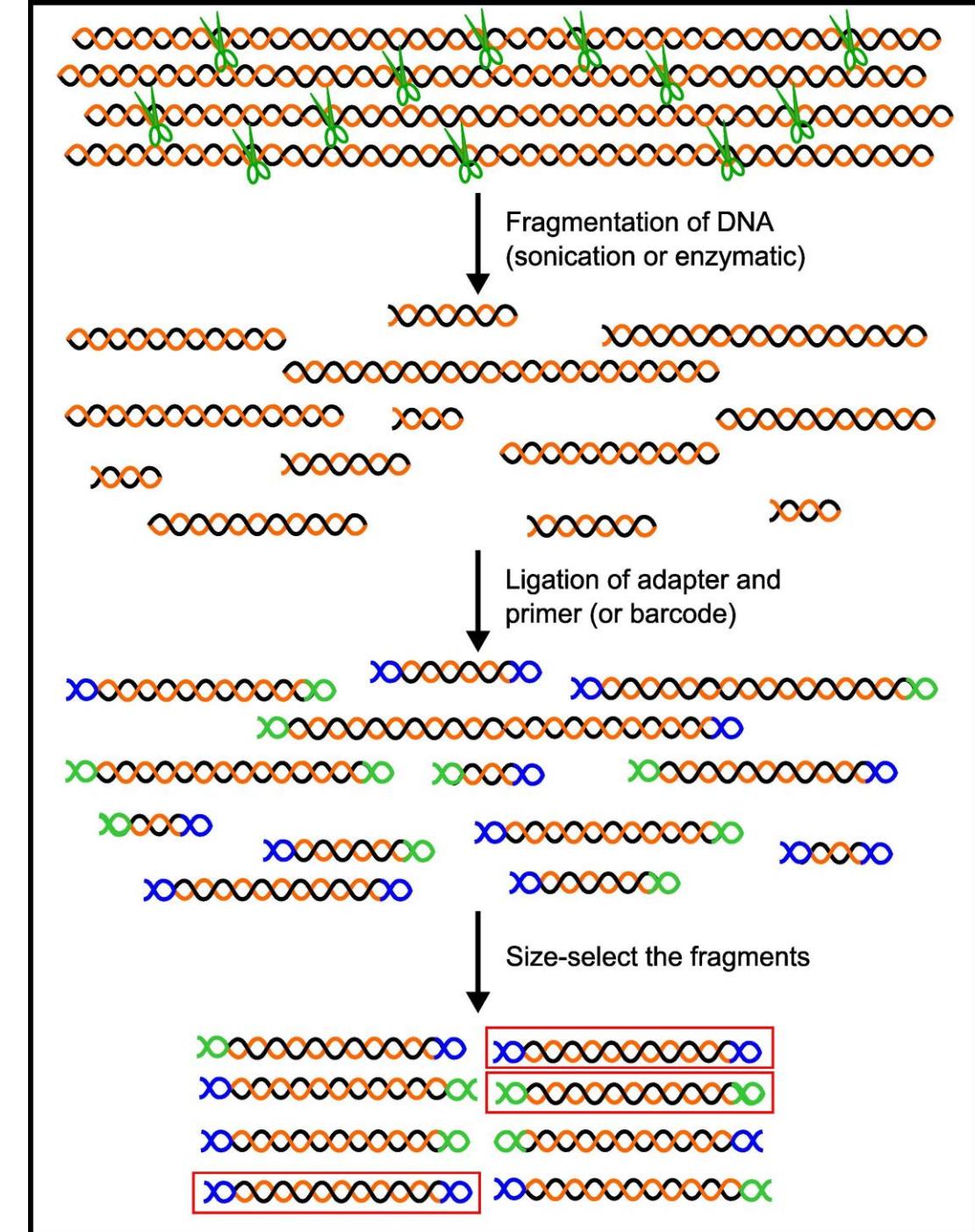
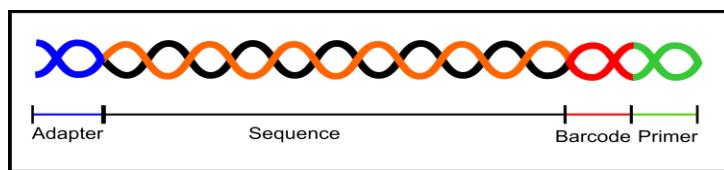


100 ng and 1 µg
Qubit 4 Fluorometer (Thermo)



200–300 bp
Agilent's BioAnalyser

Library preparation



DNA Library Prep

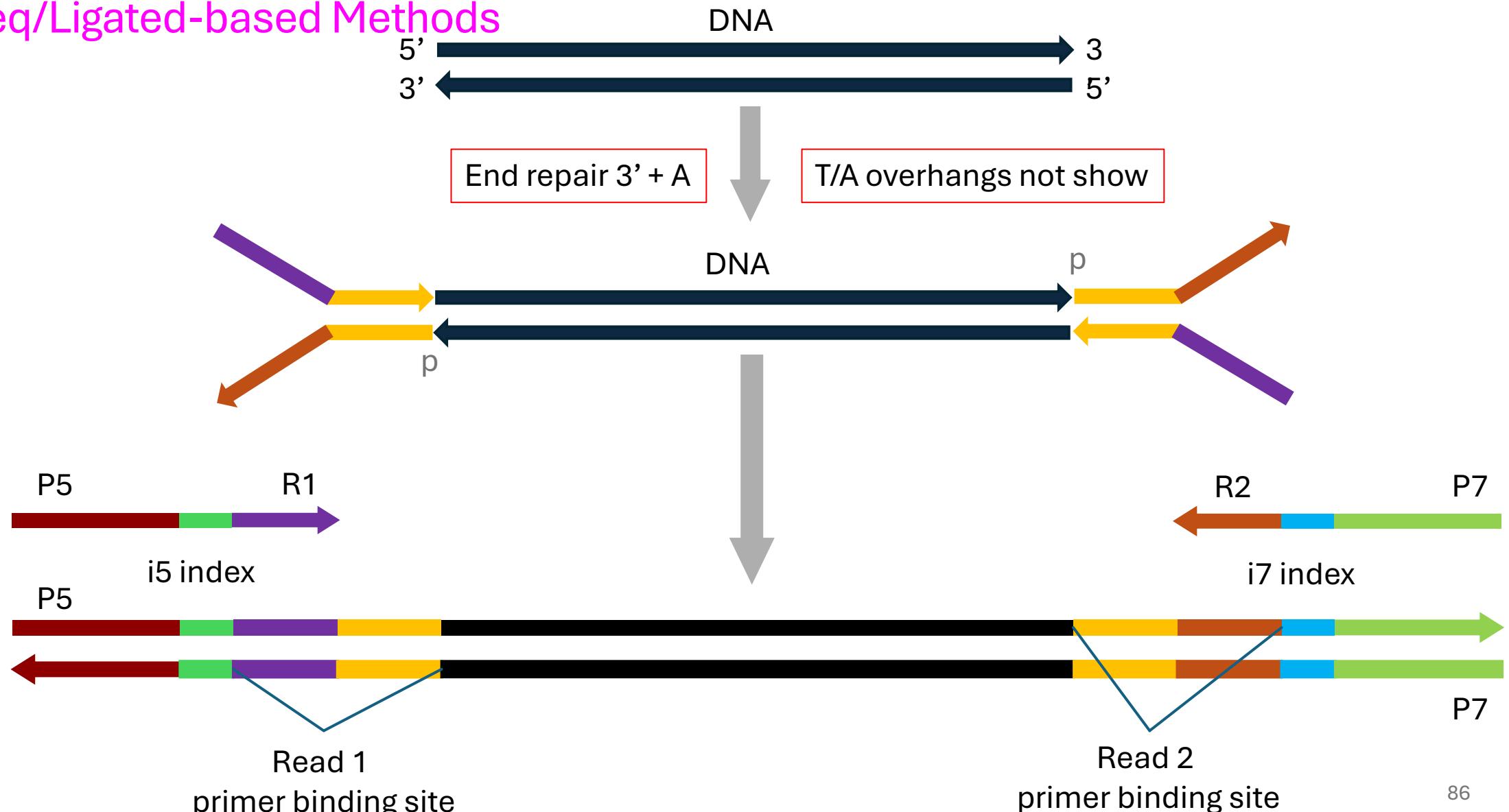
- Methods
 - TruSeq/Ligated-based Methods
 - Nextera approach
 - Meyer and Kircher approach (Mey Kir Approach)
- Applications
 - Whole Genome Sequencing (WGS)
 - Targeted sequencing
 - Whole exome sequencing (WES)
 - Gene-targeted panels

DNA Library Prep

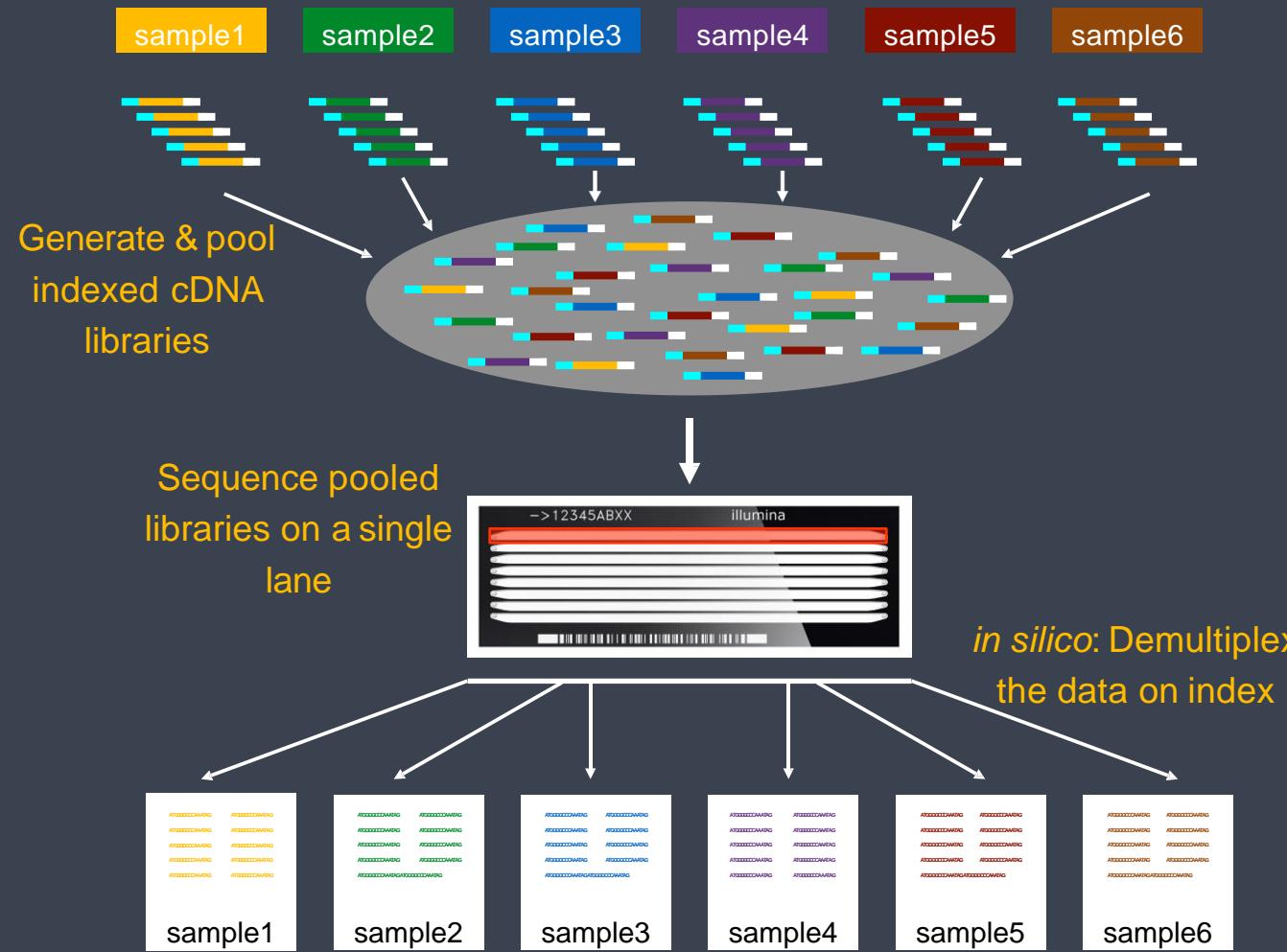
- Methods
 - TruSeq/Ligated-based Methods
 - Nextera approach
 - Rohland and Reich approach (Roh Rei Approach)
- Applications
 - Whole Genome Sequencing (WGS) - TruSeq, Nextera, Roh Rei
 - Targeted sequencing - TruSeq, Nextera, Roh Rei
 - Whole exome sequencing (WES)
 - Gene-targeted panels

DNA Library Prep

TruSeq/Ligated-based Methods



Multiplexing (barcodes and indices)



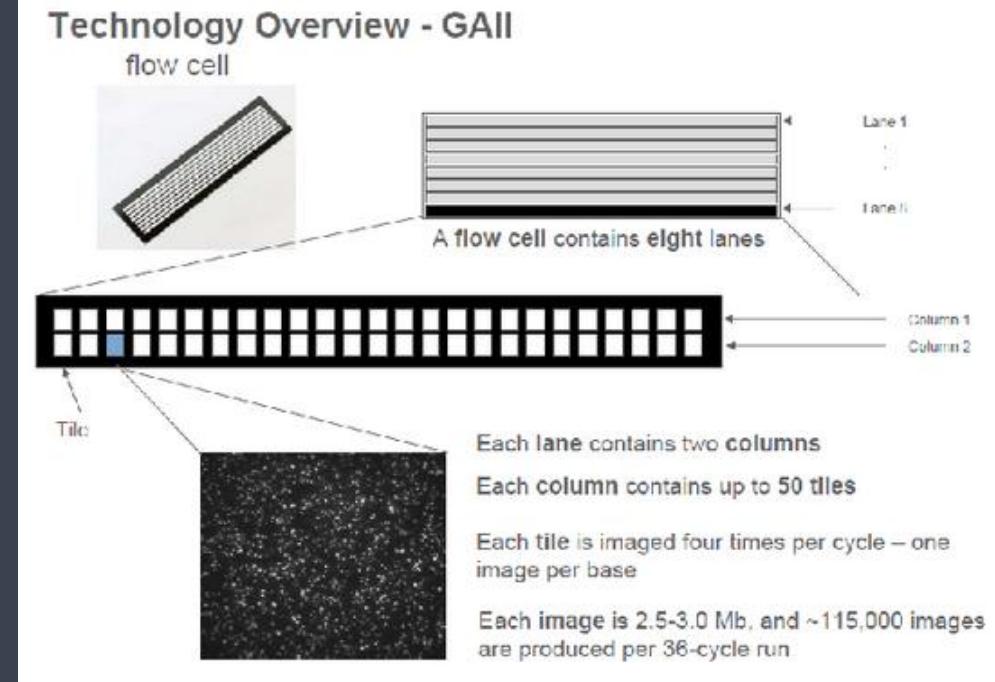
Fulllength Adapter P5

5' - AATGATAACGGCGACCACCGAGATCTACACACCGACAAACACTTTCCCTACACGACGCTCTCCGATCTXXXXX//XXXXXAGATCGGAAGAGCACACGTCTGAACCTCCAGTCACAGGTCACTATCTCGTATGCCGTCTCTGCTTG-3'
3' - TTACTATGCGCGCTGGTGGCTCTAGATGTGTGGCTGTTGTGAGAAAAGGGATGTGCTGCAGAGAAGGCTAGAXXXX//XXXXXTCTAGCCTCTCGTGTGCAAGACTTGAGGTCACTGTCAGTGTAGAGCATAACGGCAGAAGACGAAC-5'

i5 index

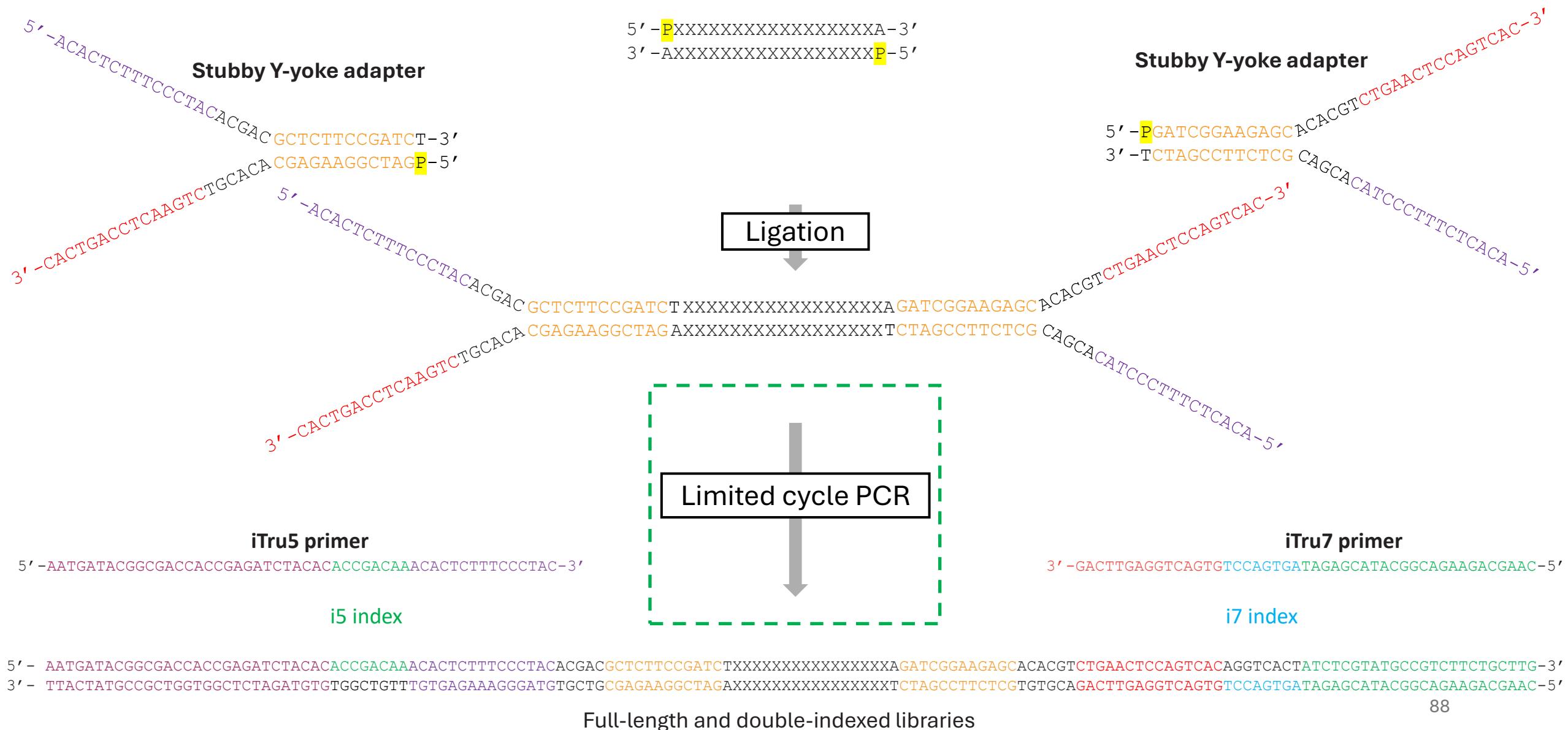
Full-length and double-indexed libraries

i7 index



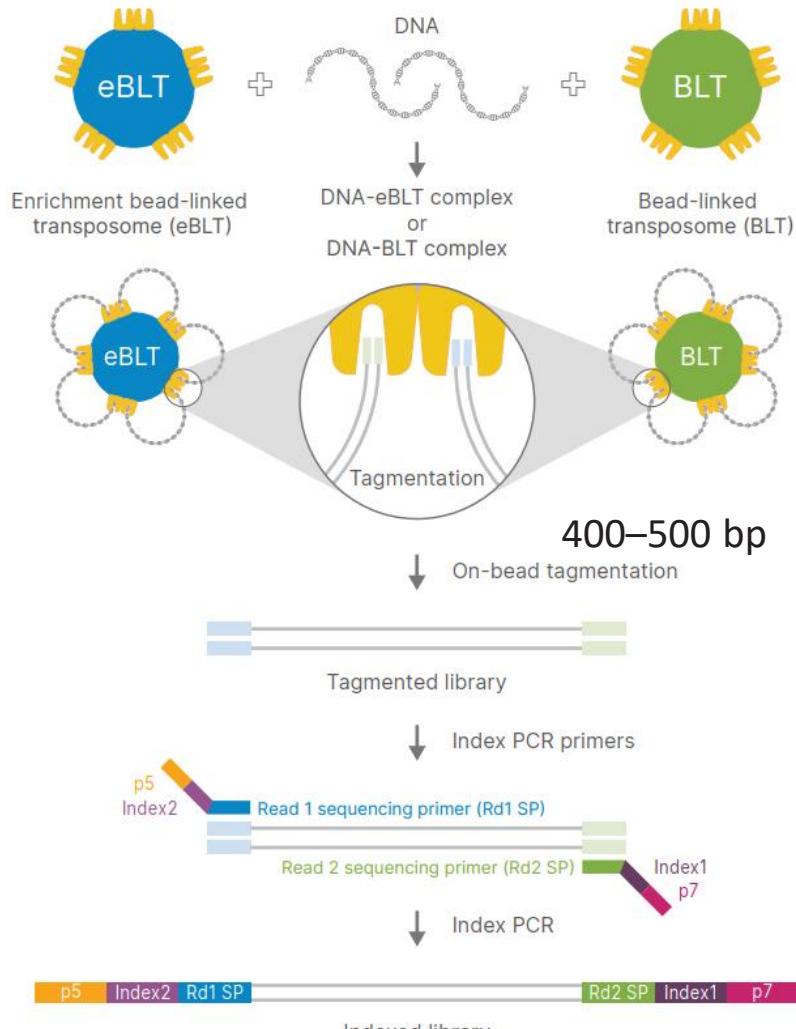
Traditional approach (TruSeq Kits)

Step 5: Limited cycle PCR

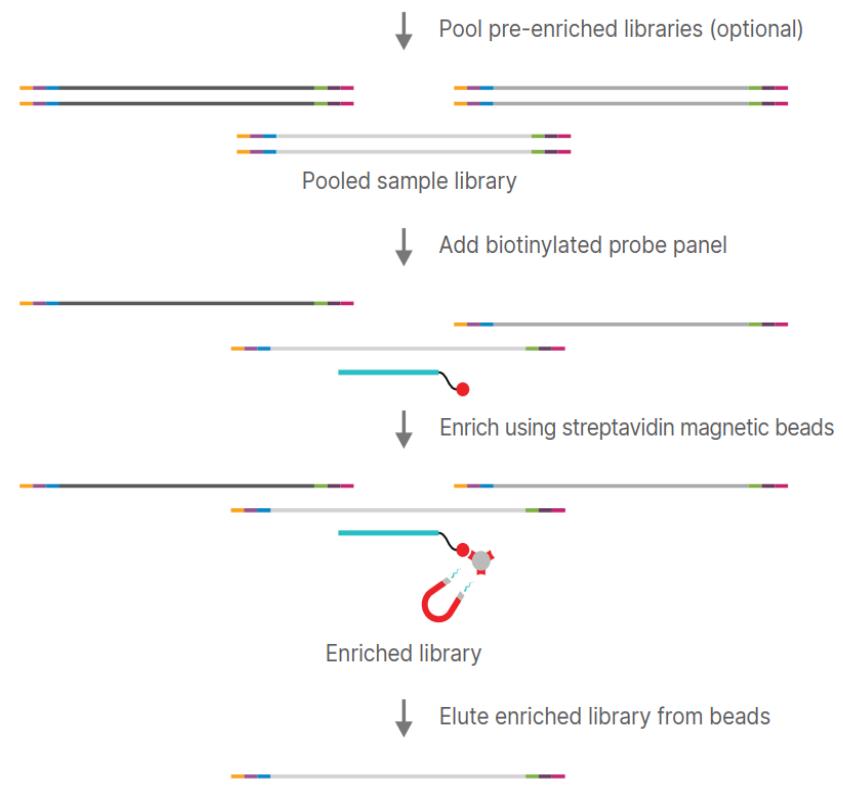


DNA Library Prep

Nextera approach



Illumina DNA Prep library ready for sequencing



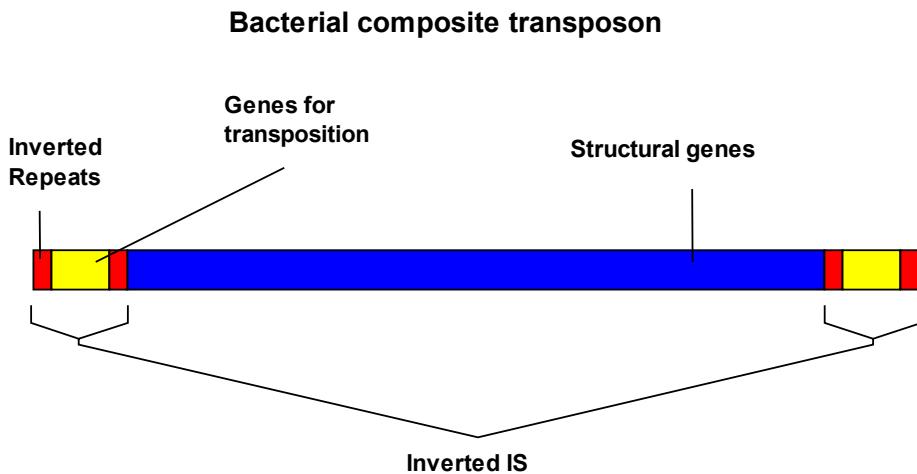
Illumina DNA Prep with Enrichment library ready for sequencing

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.

DNA Library Prep

NexTera approach: Transposable element

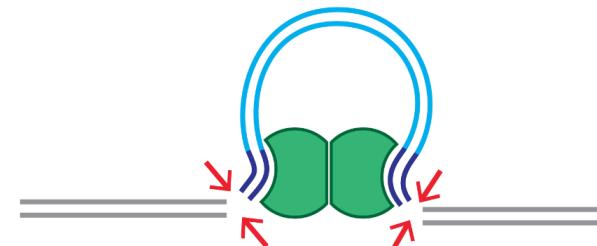


- (1) Two separate transposase subunits work together to excise a Tn5 transposon from the host DNA;
- (2) The two subunits form a transposase homodimer that is bound to the transposon; and
- (3) This newly formed “synaptic complex” then binds to a new target location in the host DNA and inserts the transposon

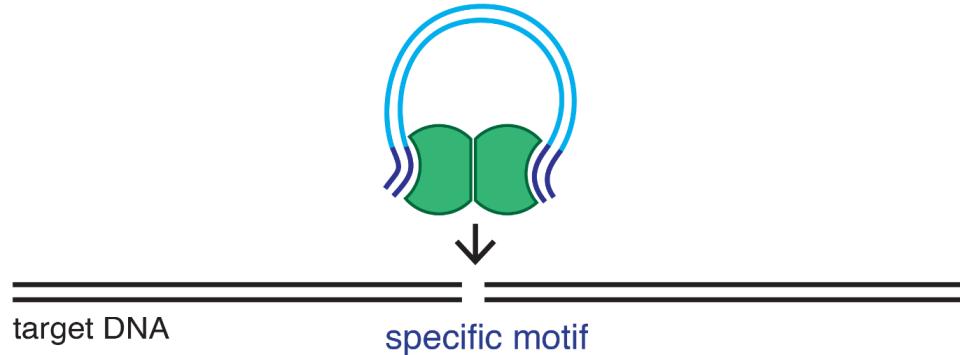
Transposase binding



Cleavage



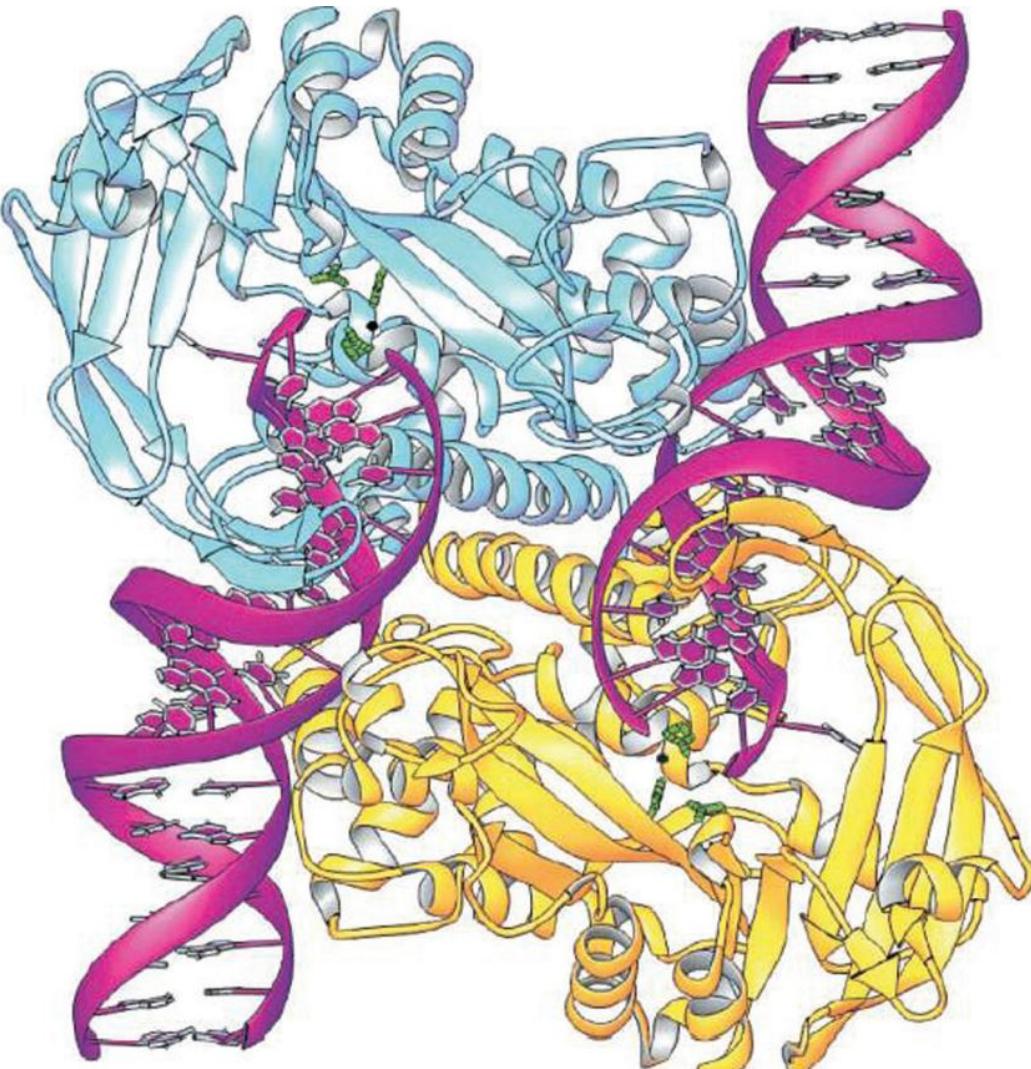
Target capture and strand transfert



- The wildtype form of Tn5 is not useful for *in vitro* transposition owing to its exceedingly low activity
- A “hyperactive” (mutated) version of the Tn5 transposase to conduct *in vitro* transposition experiments

DNA Library Prep

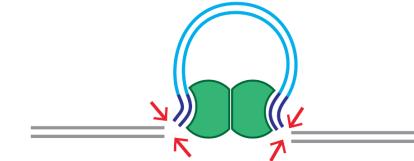
Nextera approach: Tn5 transposase



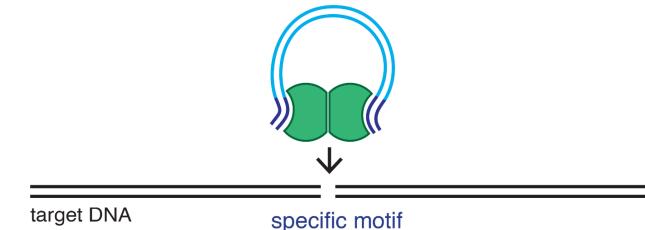
Transposase binding



Cleavage



Target capture and strand transfert



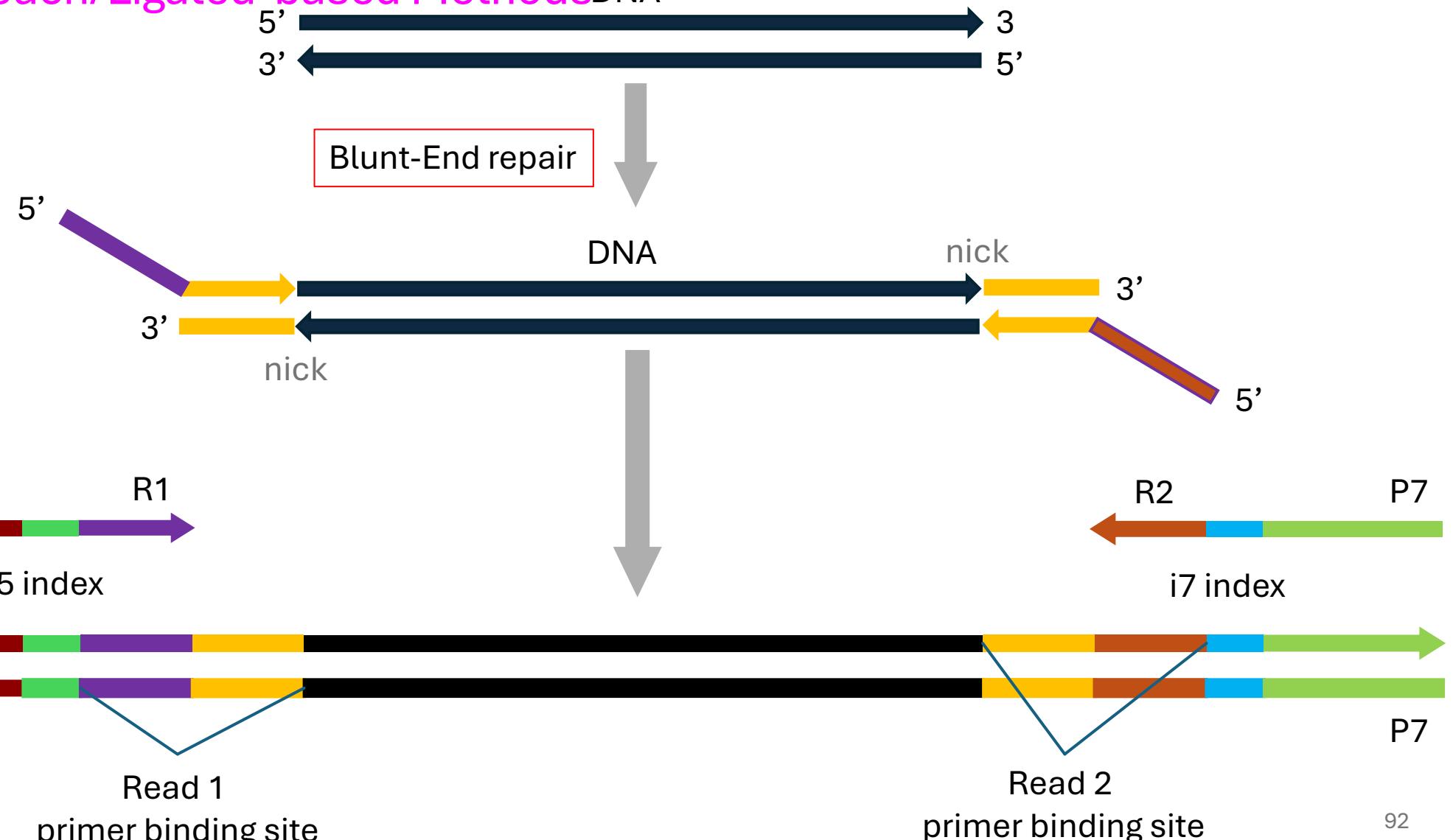
19 1
5' AGATGTGTATAAGAGACAG 3' <= 19 bp ME sequence (transferred strand)
 |||||
3' TCTACACATATTCTCTGTC 5' <= Complement (nontransferred strand)

The 19-bp “mosaic end” ME sequence for Tn5 transposase. (Bhasin, A. et al. 2000. *J Mol Biol* 302:49–63.)

Crystal structure of Tn5 transposase/DNA complex. The transposase synaptic complex is comprised of a homodimer between two transposase subunits (yellow and blue ribbons) with two DNA transposon ends (purple) protruding from the active sites of the subunits. Catalytic residues in the active site are shown as green ball-and-stick structures and the associated Mn²⁺ ion is black. (Reprinted from Davies, D. R. et al. 2000. *Science* 289:77–85.)

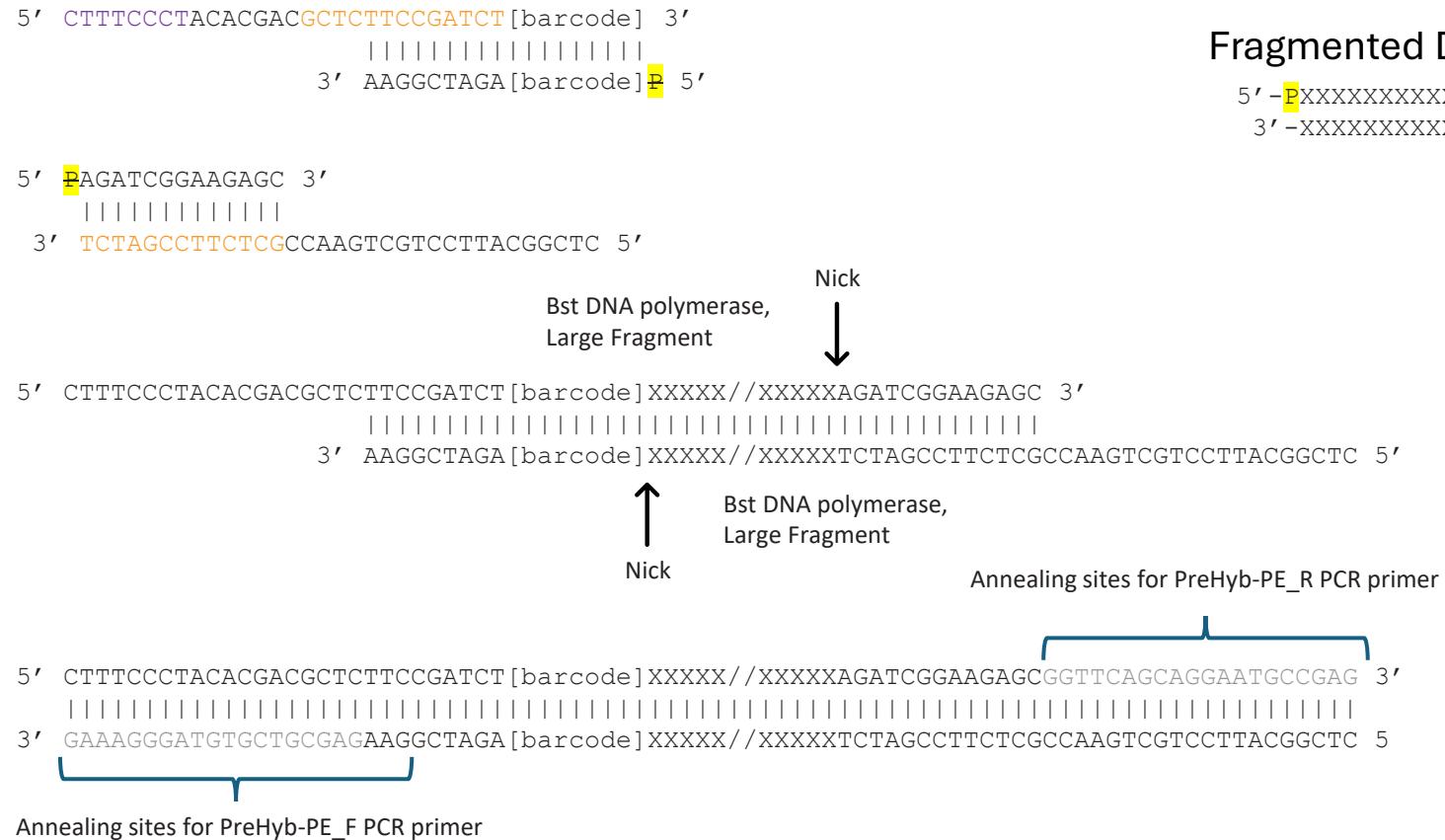
DNA Library Prep

Roh Rei Approach/Ligated-based Methods



DNA Library Prep

Roh Rei Approach/Ligated-based Methods



Structure of the Rohland and Reich Illumina adapters and initial adapter-fragment constructs.

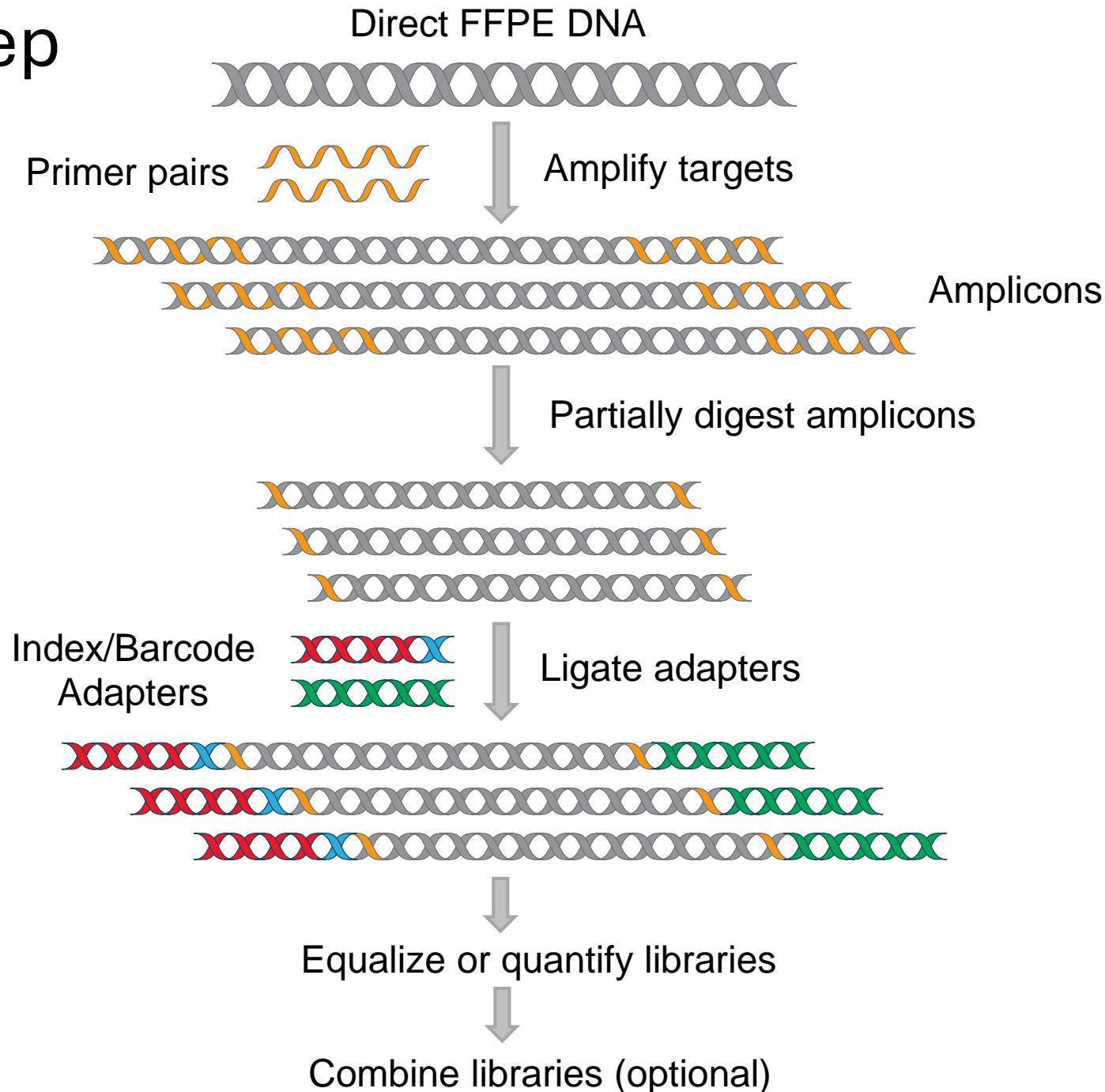
- One adapter contains a 6-bp “barcode” sequence that is used to mark individual libraries prior to making hybridization pools (vertical dashes indicate hydrogen bonding between complementary bases).
- The second adapter does not contain barcode or index sequences and resembles one of the Meyer Kircher adapters (see Figure 7.15b).
- The initial ligation product contains two nicks in the phosphate backbone and two long 5' overhangs. Before this construct can be used in the first limited cycle PCR, the nicks must be repaired and the overhangs filled in.
- Following the fill-in reaction, the adapter-fragment construct is double-stranded end-to-end and contains annealing sites for the PreHyb-PE_F and PreHyb-PE_R PCR Primers⁹³ which are used in the first limited cycle PCR. The middle section of the adapter-fragment construct contains the target DNA (a string of Xs).

DNA Library Prep

- Methods
 - TruSeq/Ligated-based Methods
 - Nextera approach
 - Rohland and Reich approach (Roh Rei Approach)
- Applications
 - Whole Genome Sequencing (WGS)
 - Targeted sequencing - TruSeq, Nextera, Roh Rei
 - Whole exome sequencing (WES)
 - PCR-based Method / Amplicon – **1000 primers**
 - Hybridized-based Method
 - Gene-targeted panels
 - PCR-based Method / Amplicon
 - Hybridized-based Method

DNA Library Prep

PCR-based Method



A Specific Example

PCR-based BRCA Assay

- Low Input: DNA of FFPE samples
- Allele frequencies: > 5%
- Expectation: somatic and germline mutations
- Faster Turnaround

BRCA1/2 Assay

- AmpliSeq for Illumina BRCA Panel (Illumina)



AmpliSeq for Illumina BRCA Panel
Targeted research panel investigating somatic and germline variants in *BRCA1* and *BRCA2*.

[AmpliSeq for Illumina BRCA Panel Data Sheet](#)
[Data sheet](#) | [HTML externalFile](#)

5 hours (...) Assay time

<1.5 hr Hands-on time

1–100 ng ... Input quantity

[See full details in the specifications table](#)

- Oncomine BRCA Assay (Ion Torrent)



[ion torrent](#)

Evaluation of the Oncomine BRCA Research Assay for variant detection by next-generation sequencing

[Download](#)

Thermo Fisher SCIENTIFIC

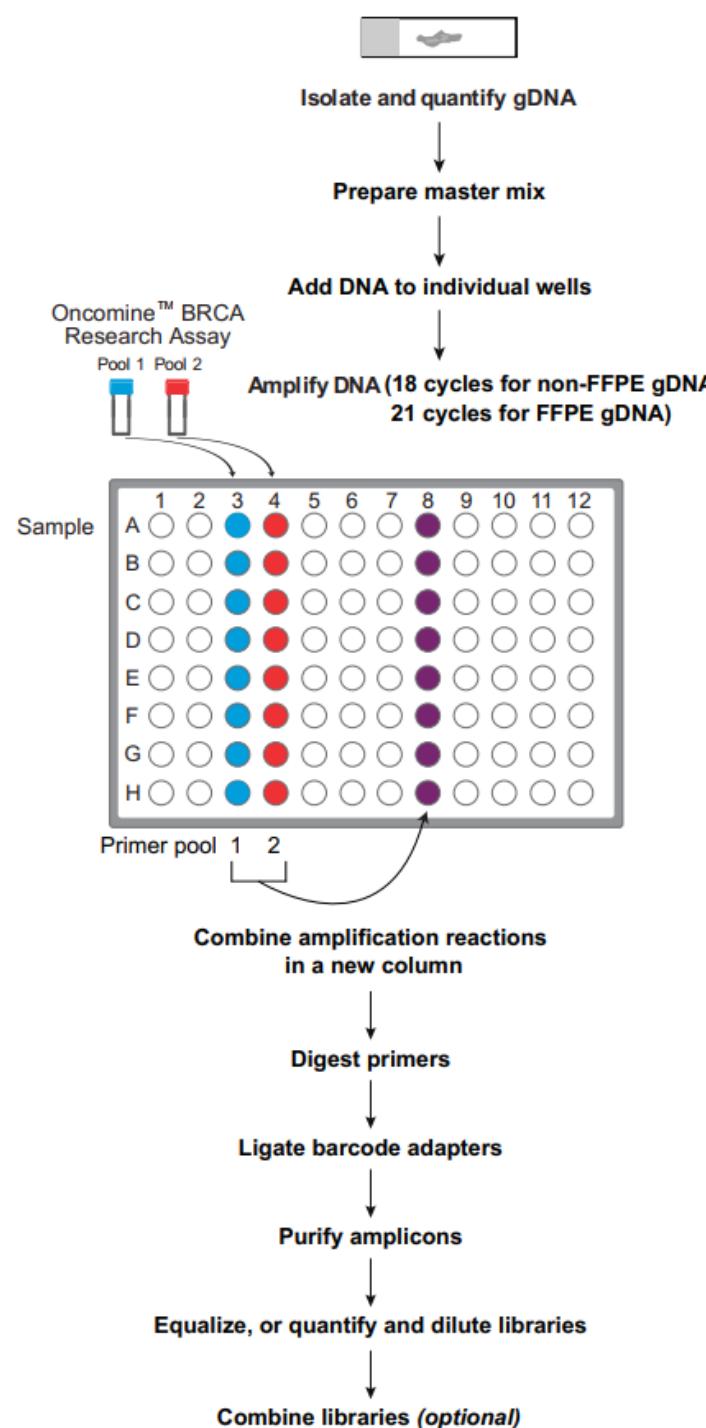
White Paper: Evaluation of the Oncomine BRCA Research Assay for variant detection by next-generation sequencing

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

DNA Library Prep

PCR-based Method

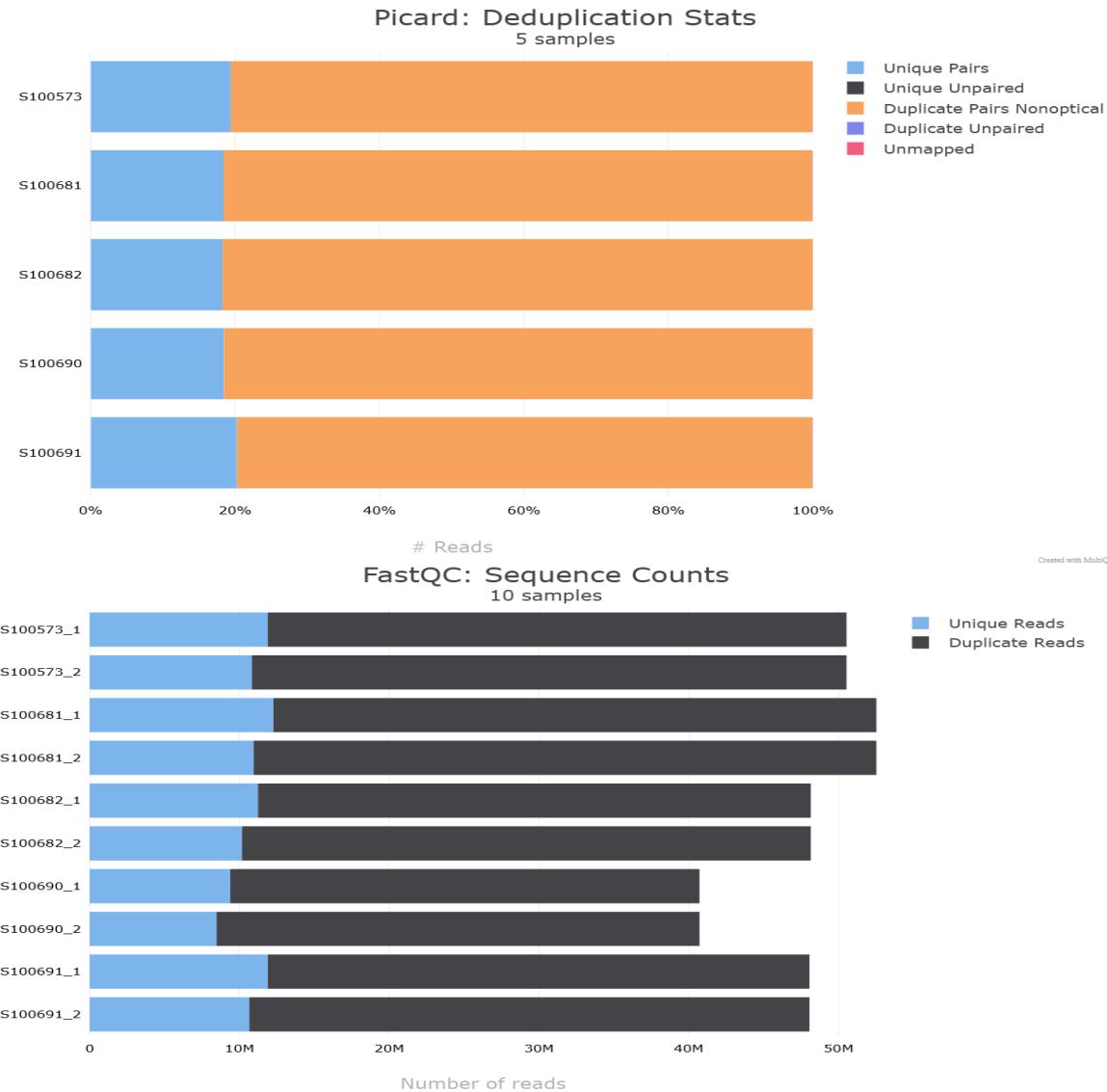


DNA Library Prep

PCR-based Method

DNA Input: 50 ng

Stage	Step	Temperature	Time
Hold	Activate the enzyme	99°C	2 minutes
Cycle (9 cycles for non-FFPE DNA, 21 cycles for FFPE DNA)	Denature	99°C	15 seconds
	Anneal and Extend	60°C	4 minutes
Hold	—	10°C	Hold (16 hours maximum)



DNA Library Prep

PCR-based Method

DNA Input: 5 ng

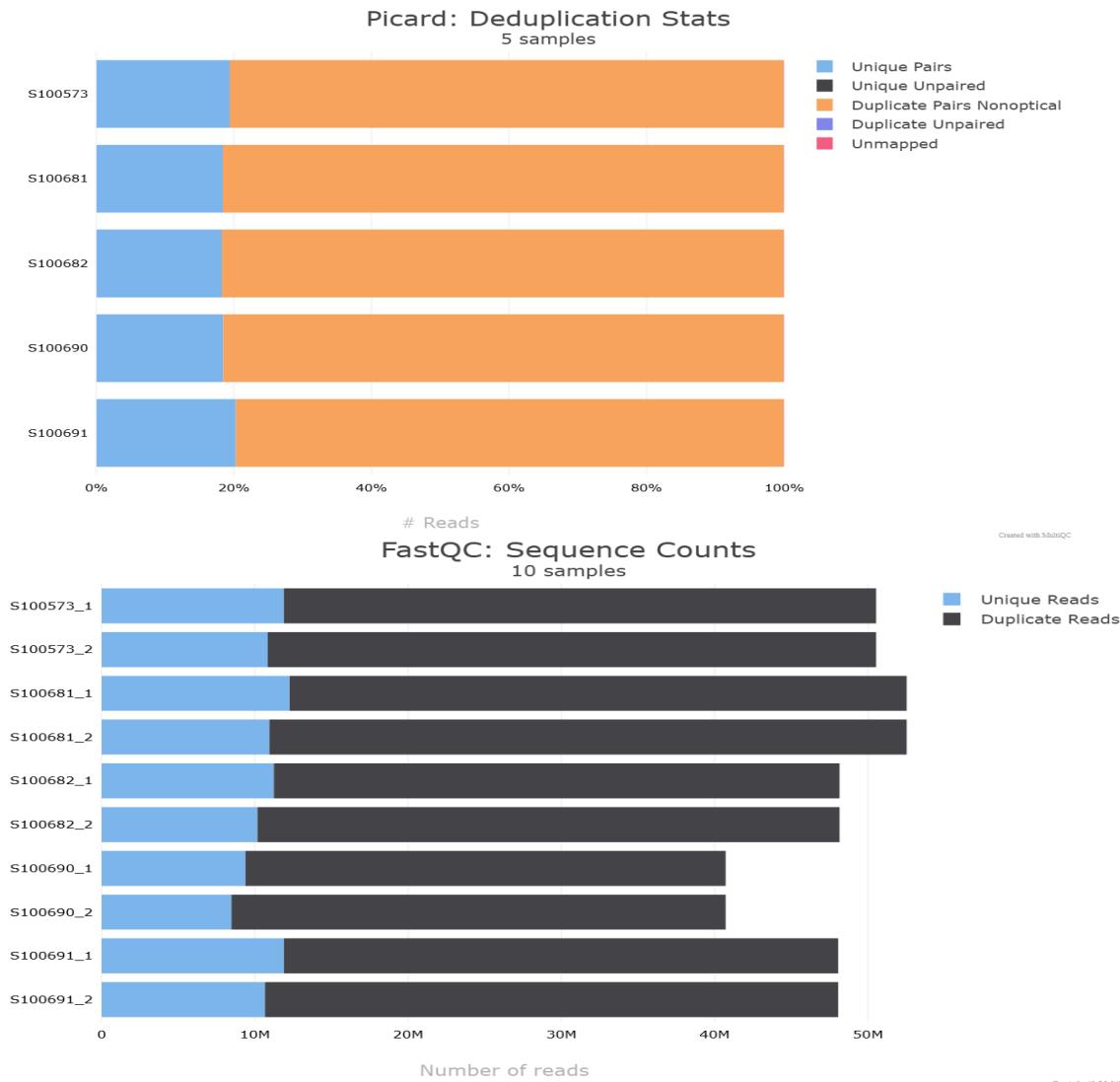
Stage	Step	Temperature	Time
Hold	Activate the enzyme	99°C	2 minutes
Cycle (18 cycles for non-FFPE DNA, 21 cycles for FFPE DNA)	Denature	99°C	15 seconds
	Anneal and Extend	60°C	4 minutes
Hold	—	10°C	Hold (16 hours maximum)

Pros

- High Sensitivity for Low VAF:** 1–5% allele frequency.
- Lower Input DNA Requirement:** Typically requires less DNA Input.
- Faster Turnaround:** suitable for smaller, routine panels.

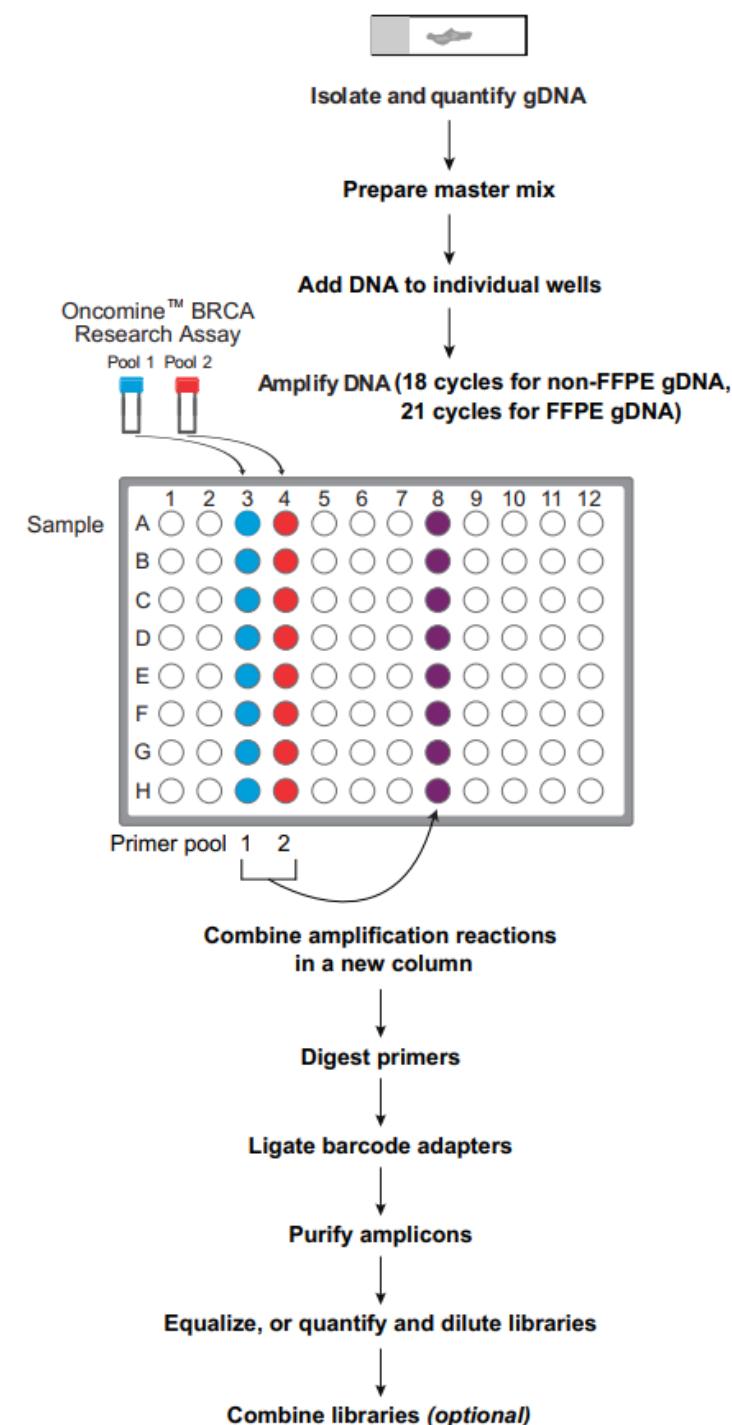
Cons

- Amplification Bias:** Certain regions.
- Limited to Known Targets.**
- Primer-Dimer and Off-Target Issues:** Non-specific amplification.



DNA Library Prep

PCR-based Method



Pros

- High Sensitivity for Low VAF:** 1–5% allele frequency.
- Lower Input DNA Requirement:** Typically requires less DNA input.
- Faster Turnaround:** suitable for smaller, routine panels.

Cons

- Amplification Bias:** Certain regions.
- Limited to Known Targets.**
- Primer-Dimer and Off-Target Issues:** Non-specific amplification.

A Specific Example

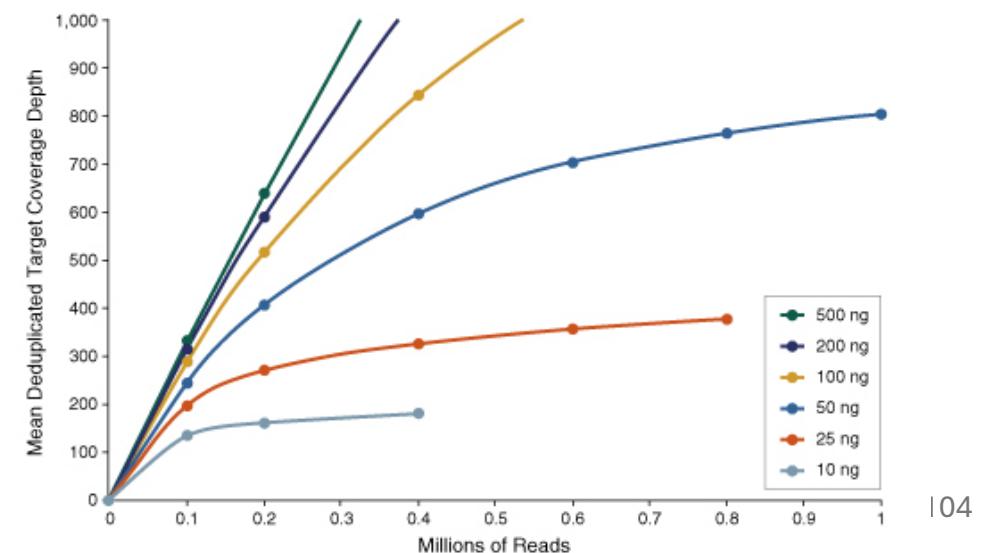
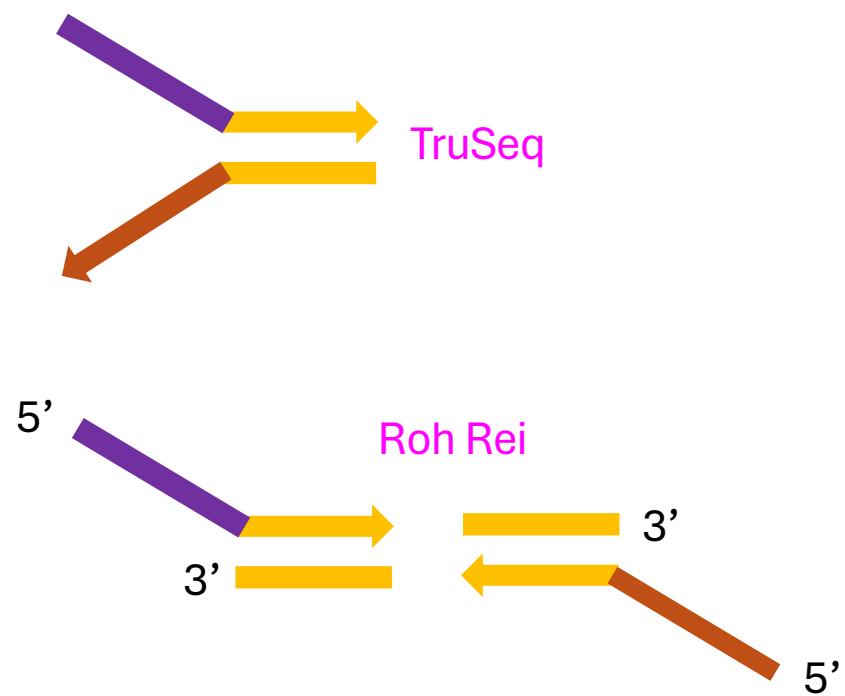
Hybridized-based BRCA Assay

- High Input: DNA of blood samples
- Allele frequencies: 1-5%
- Expectation: somatic and germline mutation
- Reduced Primer-Related Bias
- Highly scalable

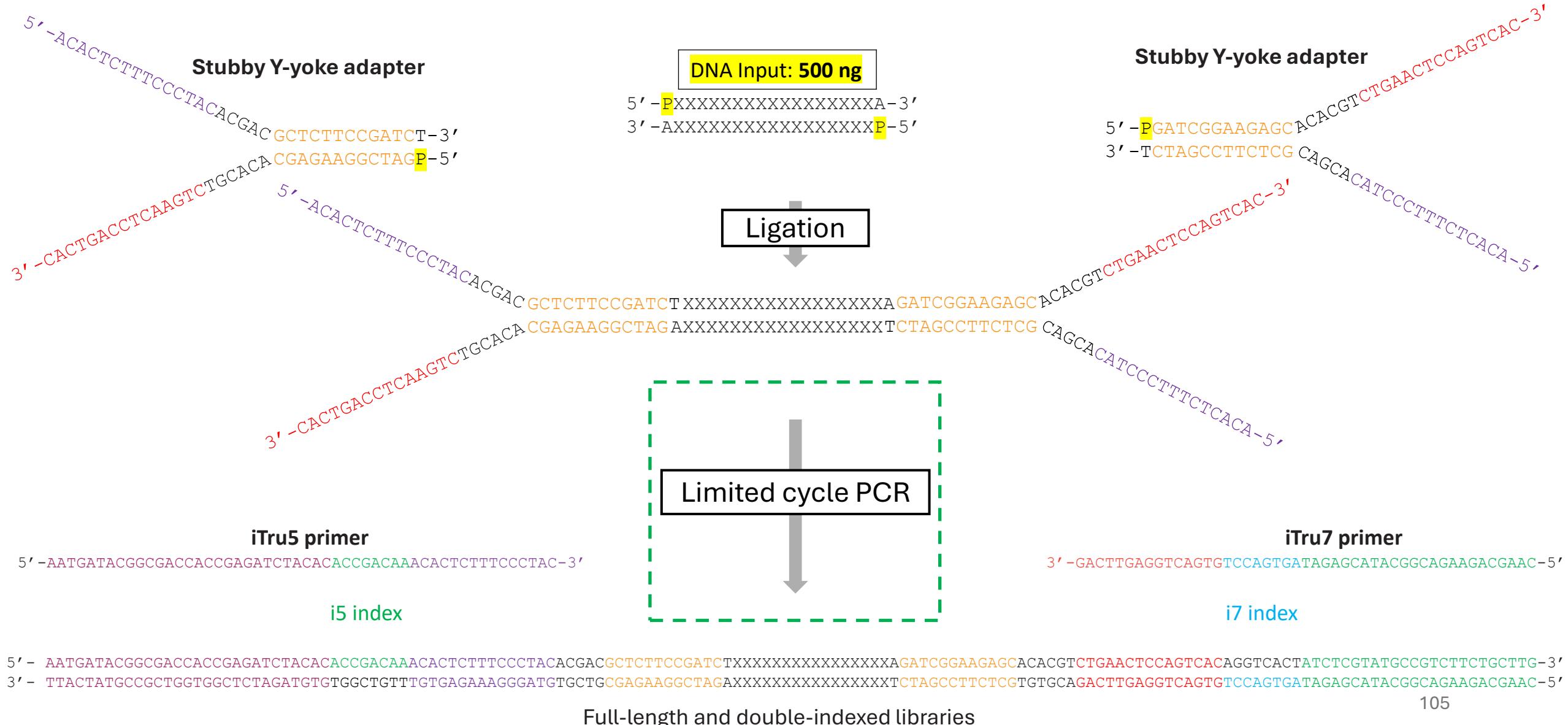
BRCA1/2 Assay



- NEB NextDirect® BRCA1/BRCA2 Panel



Traditional approach (TruSeq Kits)

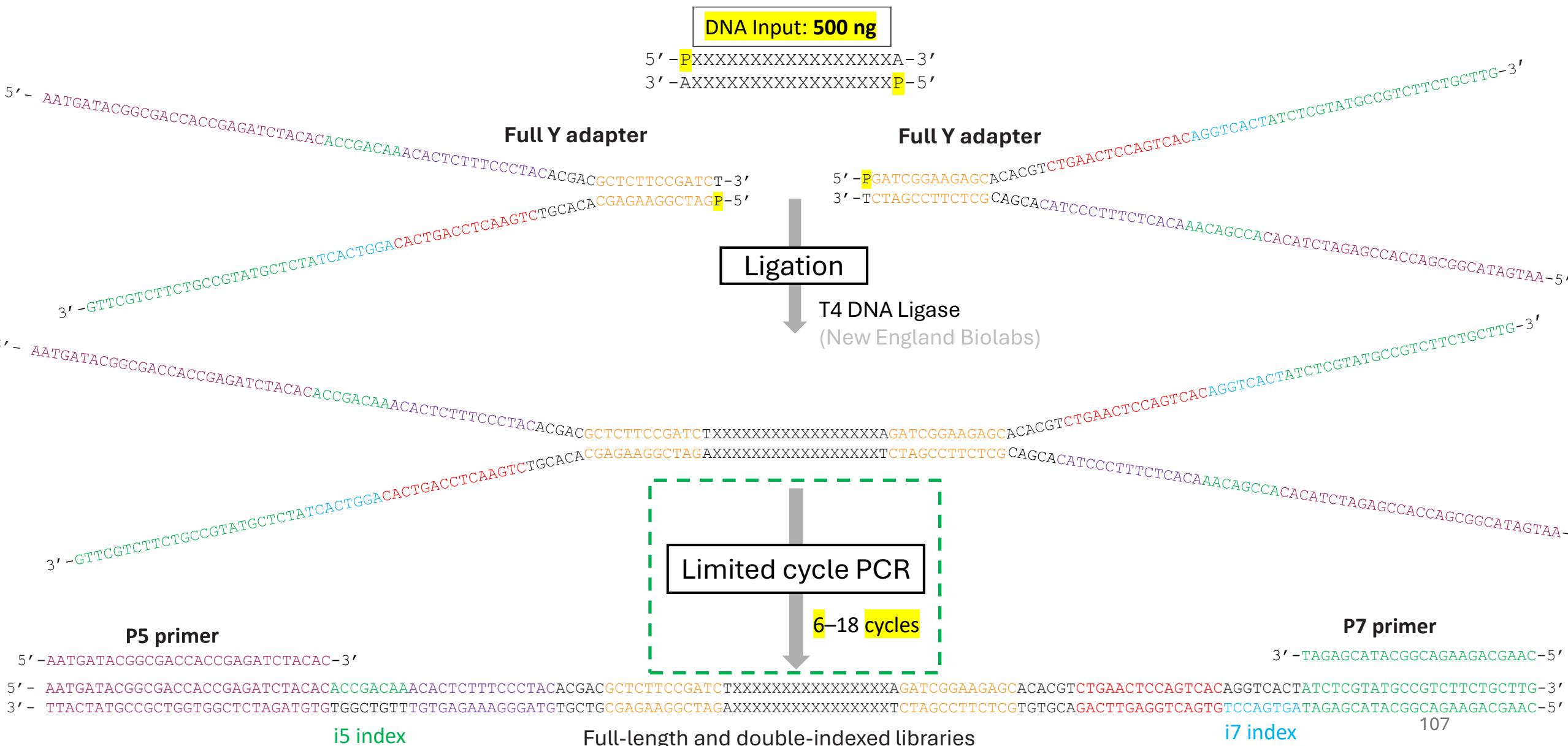


A Specific Example

Hybridized-based BRCA Assay

- High Input: DNA of blood samples
- Allele frequencies: very low VAF detection, ~~Allele frequencies: < 5%~~
- Expectation: somatic and germline mutation
- Reduced Primer-Related Bias
- Highly scalable

Traditional approach (TruSeq Kits)

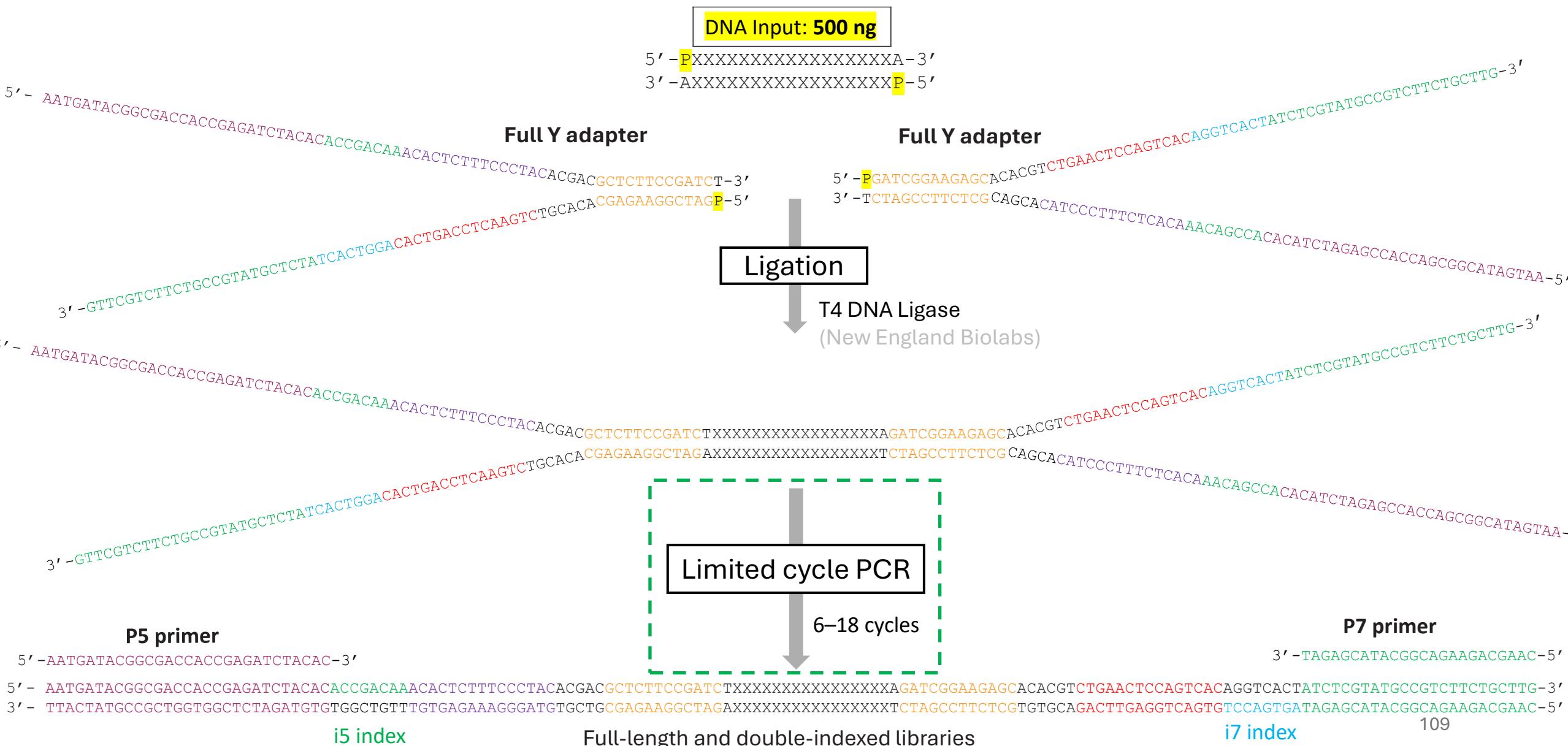


A Specific Example

Hybridized-based BRCA Assay

- High Input: DNA of blood samples
- Allele frequencies: very low VAF detection, ~~Allele frequencies: < 5%~~
- Expectation: somatic and germline mutation
- Reduced Primer-Related Bias and Artifacts
- Highly scalable

Traditional approach (TruSeq Kits)

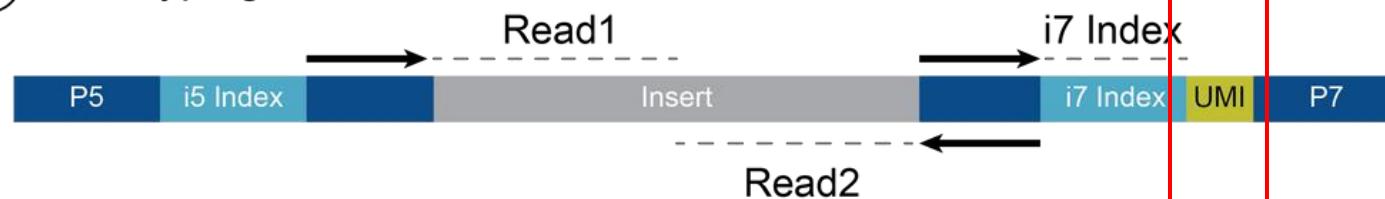


xGen Dual Index UMI Adapters—Tech Access

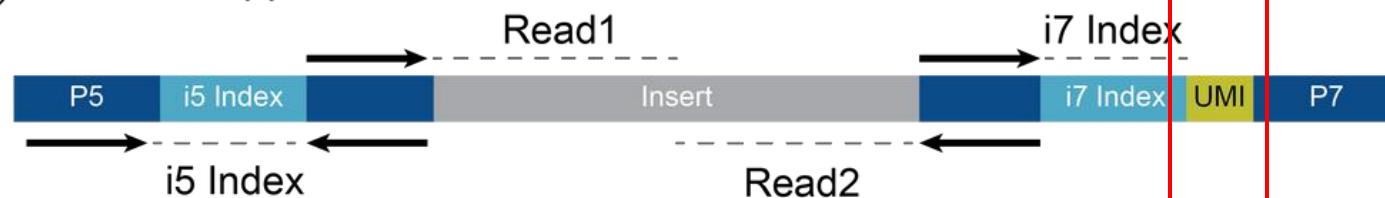
3-in-1 design



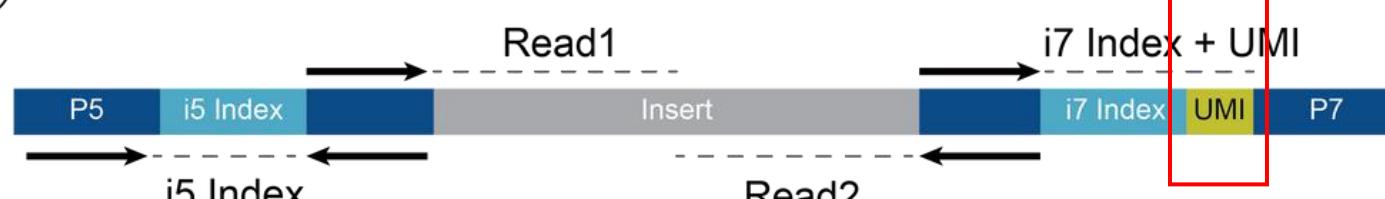
① Genotyping



② Sensitive applications

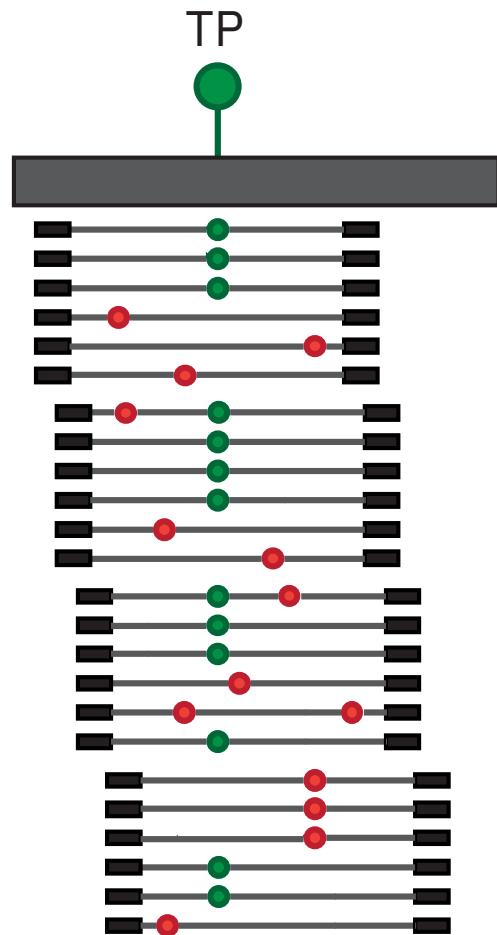


③ Rare variant detection



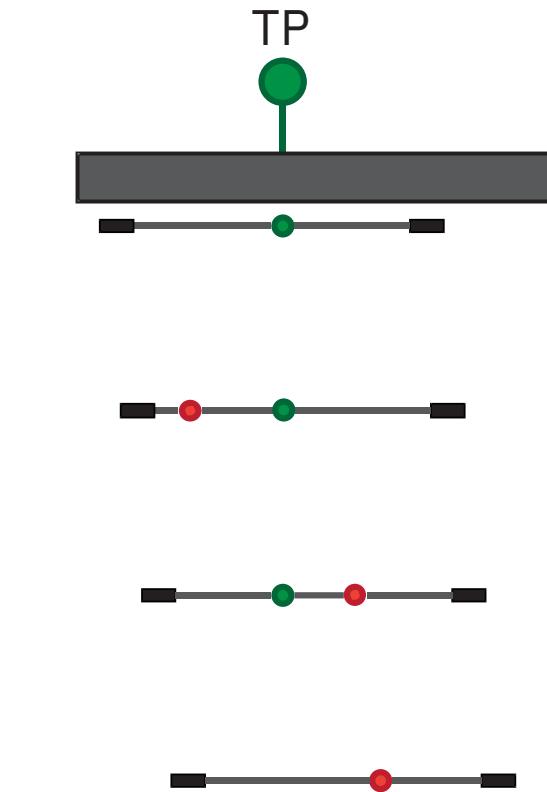
Unique Molecular Identifiers (UMIs)

Consensus calling reduces artifacts in sequencing data

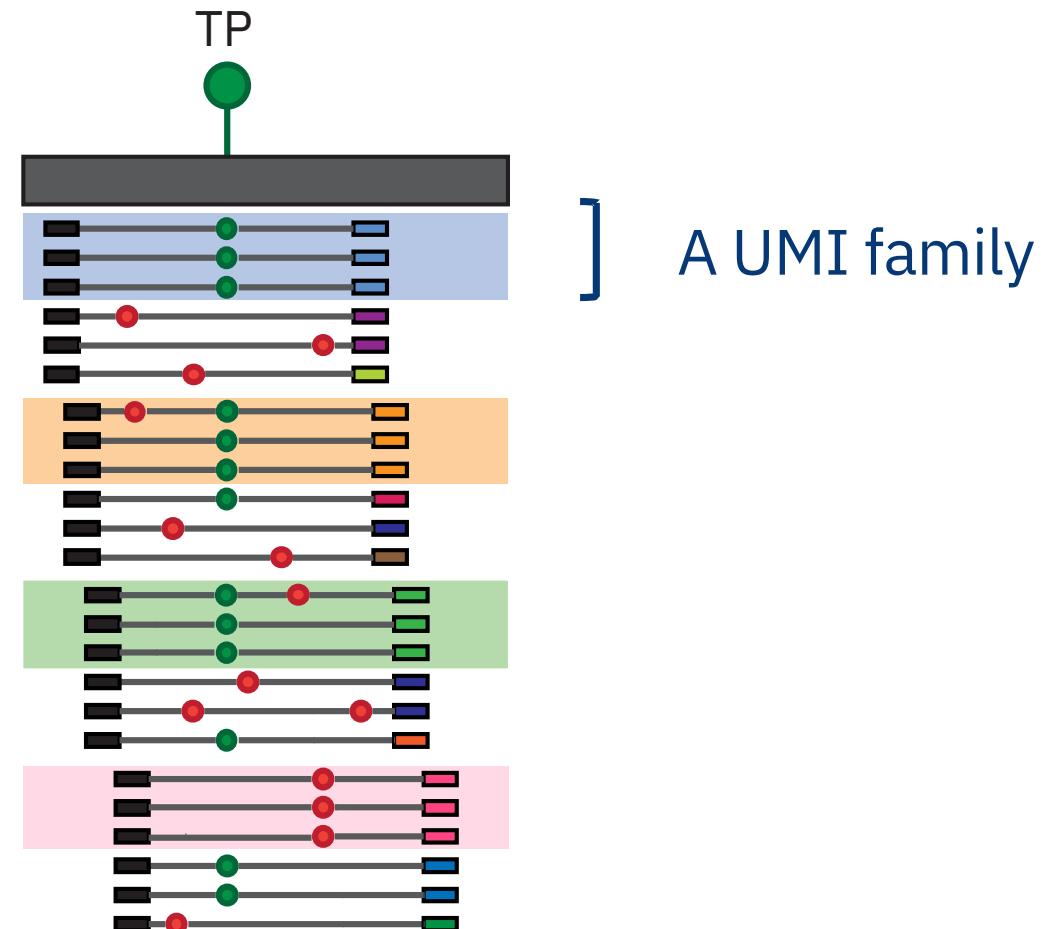


DedupbyT osttarl tr/estaodps
positions

Consensus calling reduces artifacts in sequencing data

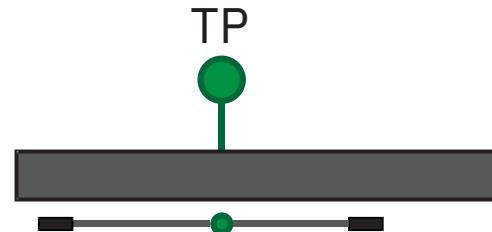


DedupbTyo tsatla
rret/asdtosp positions

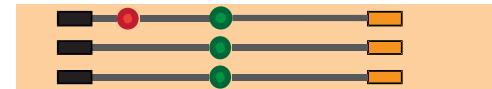
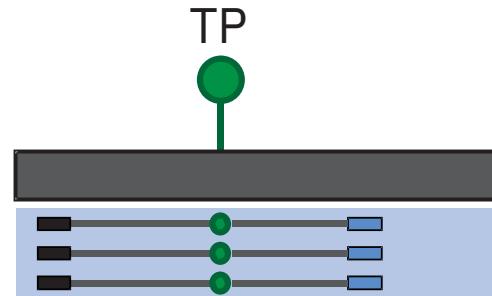


Consensus reads
(Min3)

Consensus calling reduces artifacts in sequencing data



Deduppy start/stop positions



Consensus reads
(Min3)

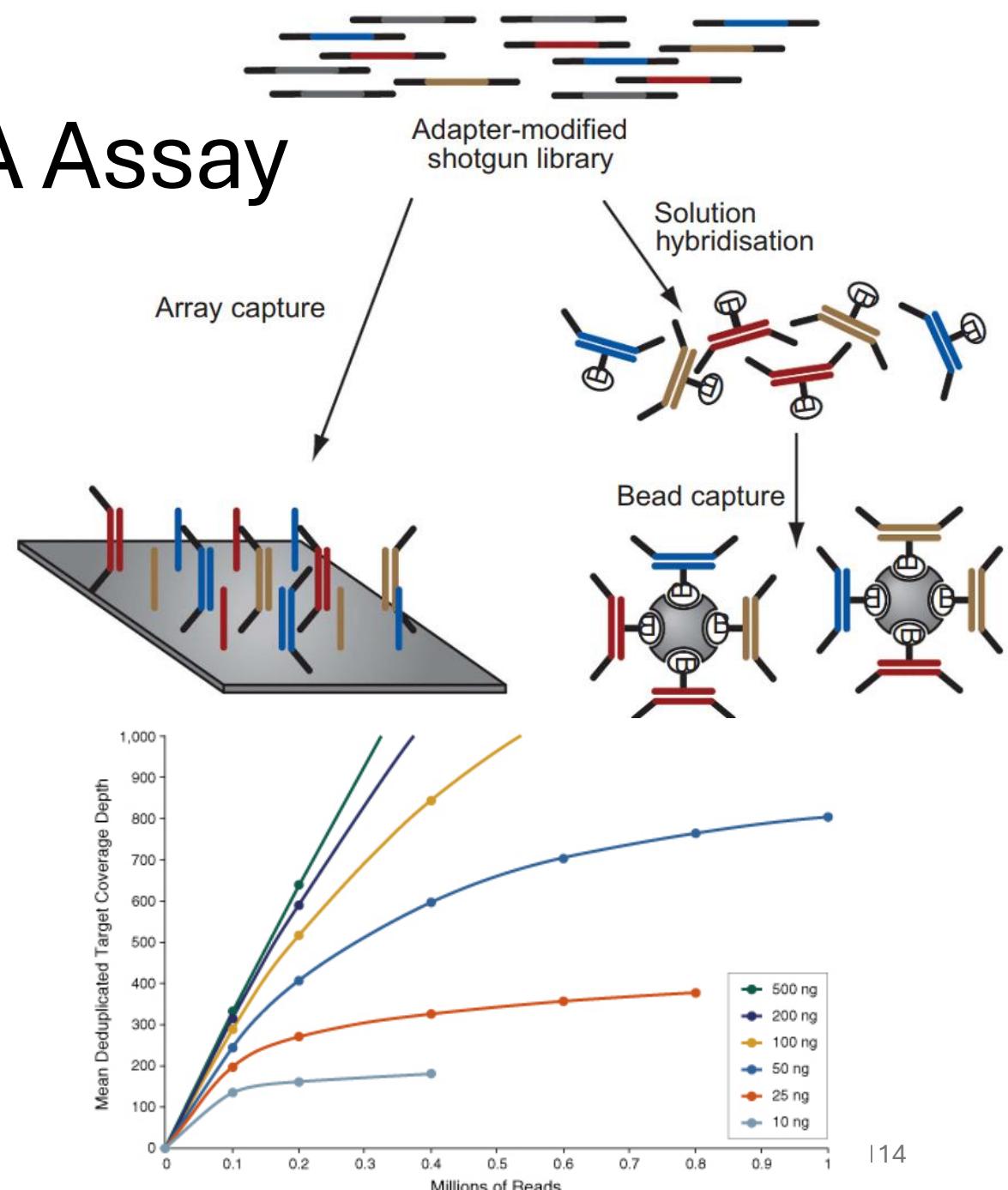
Hybridized-based BRCA Assay

Pros

- Broad Range of Variants:** point mutations, small indels, and sometimes larger rearrangements.
- Scalability for Large Panels:** Facilitates multi-gene
- Reduced Primer-Related Bias:** Fewer issues with region-specific amplification failure
- Low-Frequency Variants:** <5% VAF with deep sequencing.

Cons

- Higher Input DNA Requirements:** Usually needs more starting material (e.g., $\geq 50\text{--}200$ ng of DNA)
- More Complex Workflow**



**THANK YOU FOR
YOUR ATTENTION!**