Introduction to Next-Generation Sequencing

DARRELL L. DINWIDDIE, PHD

DARYL B. DOMMAN, PHD

Instructors

- **Darrell L. Dinwiddie, PhD**
  - PhD, Mechanisms of Innate Immune Evasion by Respiratory Viruses, University of New Mexico School of Medicine
  - Post Doctoral Fellow, Human Genomics & Next-Generation Sequencing, National Center for Genome Resources
  - Director of Lab Operations, Center for Pediatric Genomic Medicine, Children's Mercy Hospital
  - Assistant Professor, Department of Pediatrics, University of New Mexico Health Sciences Center
  - Co-Director, SARS-CoV-2 Genomic Surveillance Rocky Mountain Consortium & State of New Mexico

Instructors

- Daryl B. Domman, PhD
  - PhD, Microbiology, University of Vienna
  - Postdoctoral Researcher, Wellcome Sanger Institute
  - Reines Distinguished Postdoctoral Fellow, Los Alamos National Laboratory
  - Assistant Professor, Department of Internal Medicine, Center for Public Health, University of New Mexico Health Sciences Center
  - Co-Director, SARS-CoV-2 Genomic Surveillance Rocky Mountain Consortium & State of New Mexico

Format

- Tuesday, Tutorials, 90 minutes
  - Lectures, Hands-on computer demonstrations
- Thursday, Office Hours, 90 minutes
  - Answer questions on previous tutorials
  - Both Dr. Dinwiddie & Dr. Domman will be present
- 14 Weeks
  - February 15- May 19, 2022

Agenda- weeks 1-7

- **Week 1- Introduction to Next-Generation Sequencing- Dr. Dinwiddie**
- **Week 2- Good Laboratory Practices for NGS Processes- Dr. Dinwiddie**
- **Week 3- Introduction to NGS Data and File types- Dr. Domman**
- **Week 4- Introduction to working on the command line and virtual machine- Dr. Domman**
- **Week 5-Overview of laboratory protocols for pathogen sequencing- Dr. Dinwiddie**
- **Week 6- Short read mapping and calling variants against reference genome (SARS-CoV-2 focus) - Dr. Domman**
- **Week 7- Illumina based SARS-CoV-2 genome sequencing protocols- Dr. Dinwiddie**

Agenda- weeks 8-14

- **Week 8- Genome assemblies (bacterial focus)- Dr. Domman**
- **Week 9- Nanopore based SARS-CoV-2 genome sequencing protocols - Dr. Dinwiddie**
- **Week 10- How to create and interpret Phylogenies- Dr. Domman**
- **Week 11- Comparison of Sequencing Technologies for Pathogen Sequencing- Dr. Dinwiddie**
- **Week 12- Genomic epidemiology and Nextstrain pipeline Dr. Domman**
- **Week 13- NGS troubleshooting, common mistakes & solutions - Dr. Dinwiddie**
- **Week 14- Data dissemination to public/private data repositories. - Dr. Domman**

Major Learning Outcomes

- Understand the differences in sequencing technologies
- Theory of lab practices for sequencing library protocols
- Understand different file formats related to sequencing data
- Map sequencing reads to a reference genome and call variants
- Construct phylogenetic trees for inferring evolutionary history
- Identify genomic differences between pathogens

Introduction to Next-Generation Sequencing

Contents

- Introduction to Next-Generation Sequencing
- Illumina
  - DNA library prep
  - RNA library prep
  - Cluster generation
  - Sequencing
- Oxford Nanopore Technologies
  - DNA library prep
  - RNA library prep
  - Sequencing

Next-Generation Sequencing

- Describes a variety of newer sequencing methods that are able to sequence samples in a massively parallel manner and much lower cost.
- Some error rates of NGS can be higher than Sanger sequencing, but increased depth of coverage of sequence improves consensus accuracy.
- May require large computer resources for analysis and interpretation.

NGS Systems

Current Systems

- Illumina
  - iSeq
  - MiniSeq
  - MiSeq
  - NextSeq 550
  - NextSeq 1000, 2000
  - NovaSeq
- Life Technologies/ ThermoFisher
  - Ion Torrent S5
  - Genexus
- Pacific Biosciences
  - RS II
  - Sequel
- Oxford Nanopore
  - MinION
  - GridION
  - PromethION

Obsolete Systems

- Polonator
- Helicos
- Life Technologies SOLiD 3, 4 & 5500
- Roche 454
- Complete Genomics
- Illumina
  - GA IIx
  - NextSeq 500/550
  - HiSeq 2500, 3000, 4000
  - X Five & X Ten
- Life Technologies/ ThermoFisher Ion Torrent
  - PGM
  - Proton

Types of Next Generation Sequencing

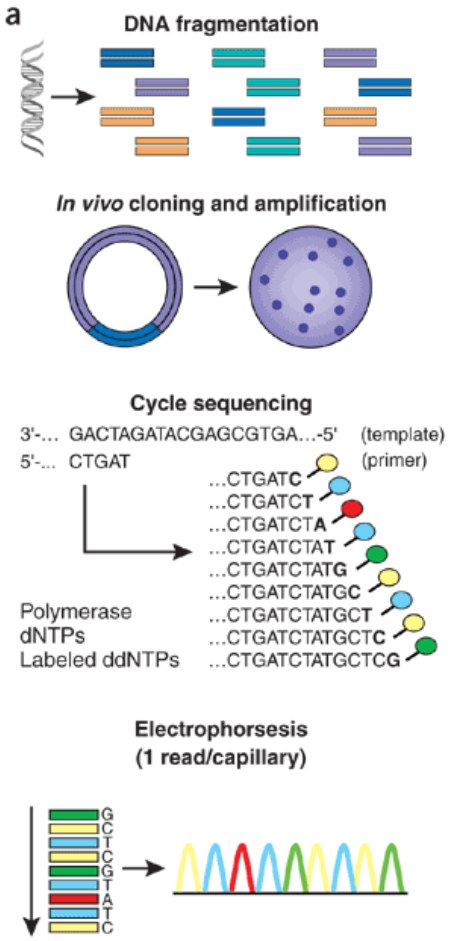
- DNA
  - Human, animal, microbes, etc.
  - Metagenomic
- RNA-seq (mRNA, ncRNA)
  - Poly-A selection
  - Ribosome depletion, includes non-coding RNA
  - Targeted enrichment
- Single Cell
- Amplicon
  - DNA or cDNA
- Directional RNA
  - Maintains strand information (sense vs antisense)
- ChIP-seq
  - Chromatin immunoprecipitation
- Methyl-seq
  - Bisulfite treatment
  - Direct detection of modifications (PacBio)
- Small RNA (miRNA)

Types of Next Generation Sequencing

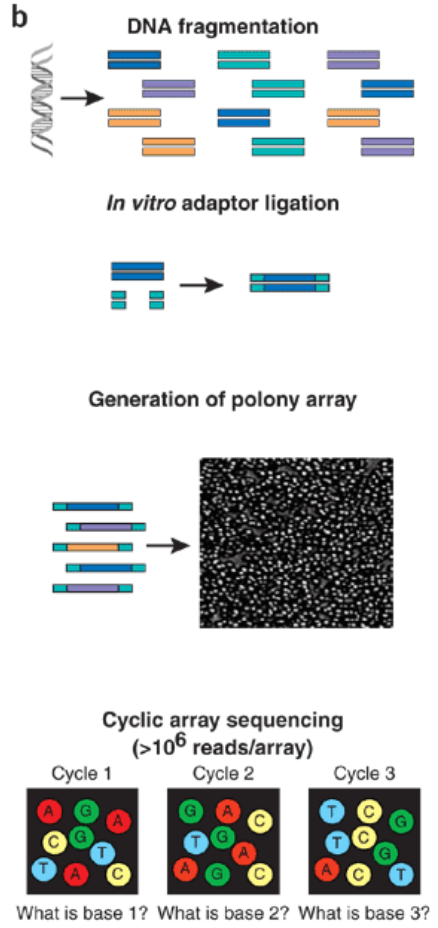
- Single end
  - Sequence from one end of DNA
  - Read lengths vary from 1x 36 bp to >1x 25,000 bp
- Paired-end
  - Sequences from both sides of the DNA insert, 2x 50 bp to 2x 300 bp
- Multiplexed
  - Can sequence multiple samples together (multiplex)
- Mate- Paired
  - Long-insert, paired-end (Illumina)

Comparison to Sanger Sequencing

Sanger Sequencing



Next-Generation Sequencing



Illumina Next-Generation Sequencing

Definitions

- Library
  - Completed sample, prepped and ready for sequencing. Includes two adapters and one or more indexes.
- Insert
  - Nucleic acid (DNA, cDNA) between Illumina adapters that can be sequenced.
  - The size of DNA library, can describe just the length of DNA or DNA + Adapters (~135 bp)
- Adapter
  - Specific DNA sequence added during library prep that enables DNA to attach to flowcell, includes sequence primer binding site, indexes. Illumina refers to them as P5 and P7.
- Index/barcode
  - Unique sequence (single or dual) that allows multiple samples to be sequenced together. Illumina refers to them as i5 & i7.
- Flowcell
  - Oligo-coated slide/cassette that library is hybridized to and in which sequencing reactions occur.

Illumina Library Structure

Rd1 SP- Read 1 Sequencing Primer
Rd2 SP- Read 2 Sequencing Primer

Illumina DNA Library Prep- Ligation

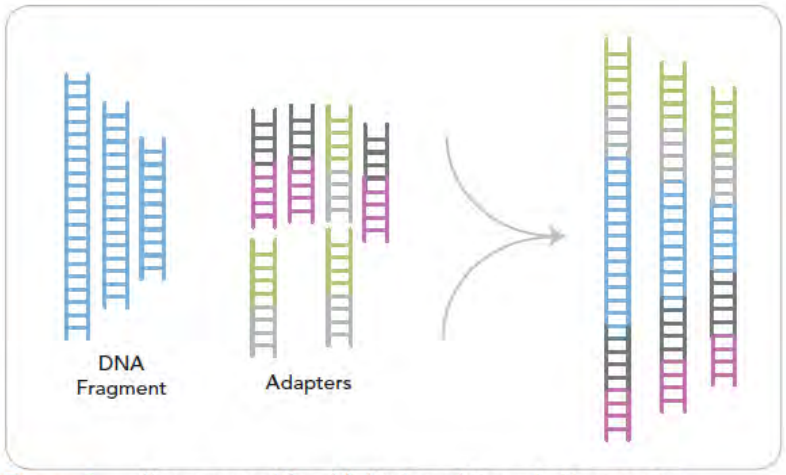


Figure 1 *Fragments After Multiplexed Sample Preparation*

Fragment DNA

↓

Repair ends / Add A overhang

↓

Ligate adapters

↓

PCR amplify DNA

Single Index/barcode

Adapter 2 (P7)

Sequencing Primer

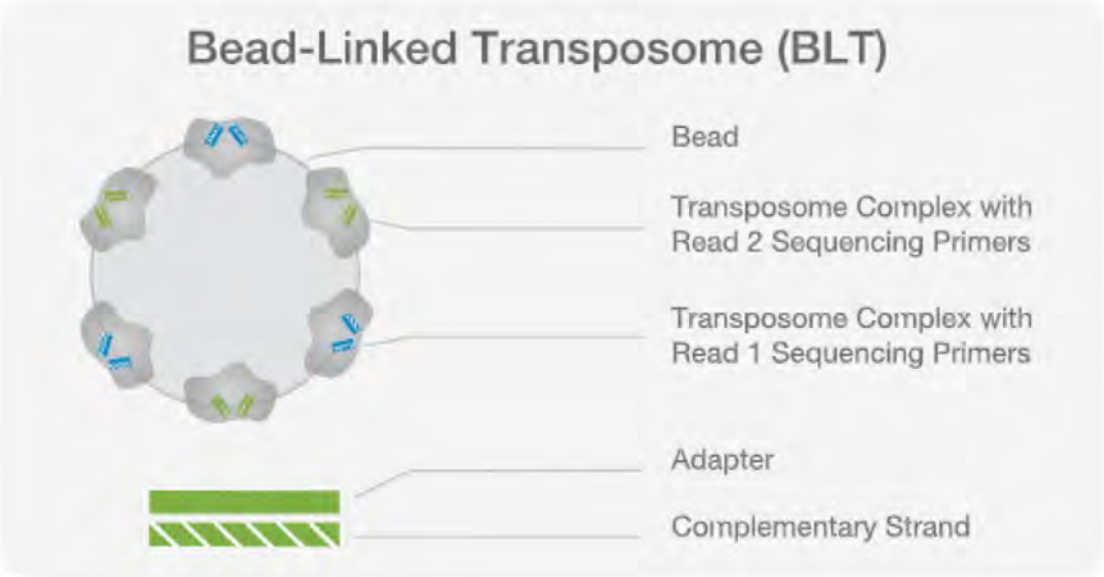
DNA/cDNA Insert

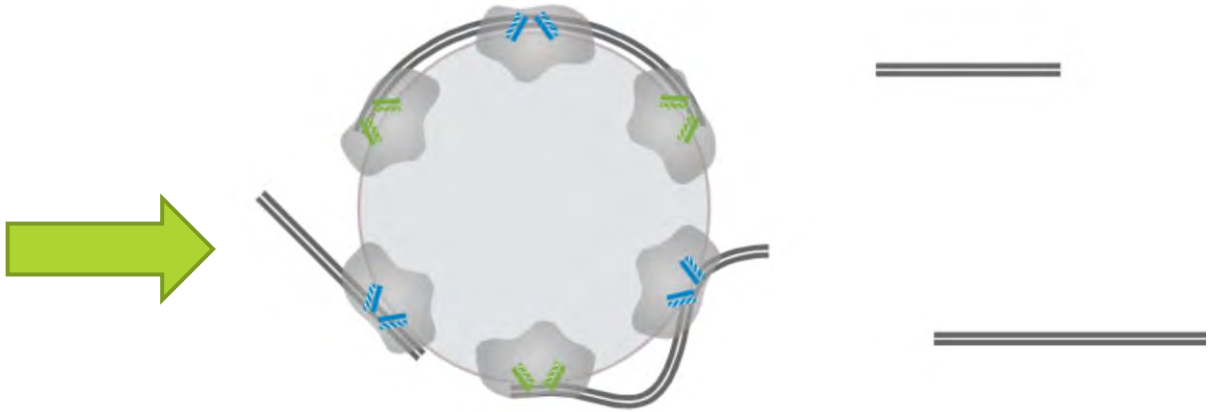
Index Primer

Index (i5)

Adapter 1 (P5)

Illumina DNA Library Prep- Tagmentation

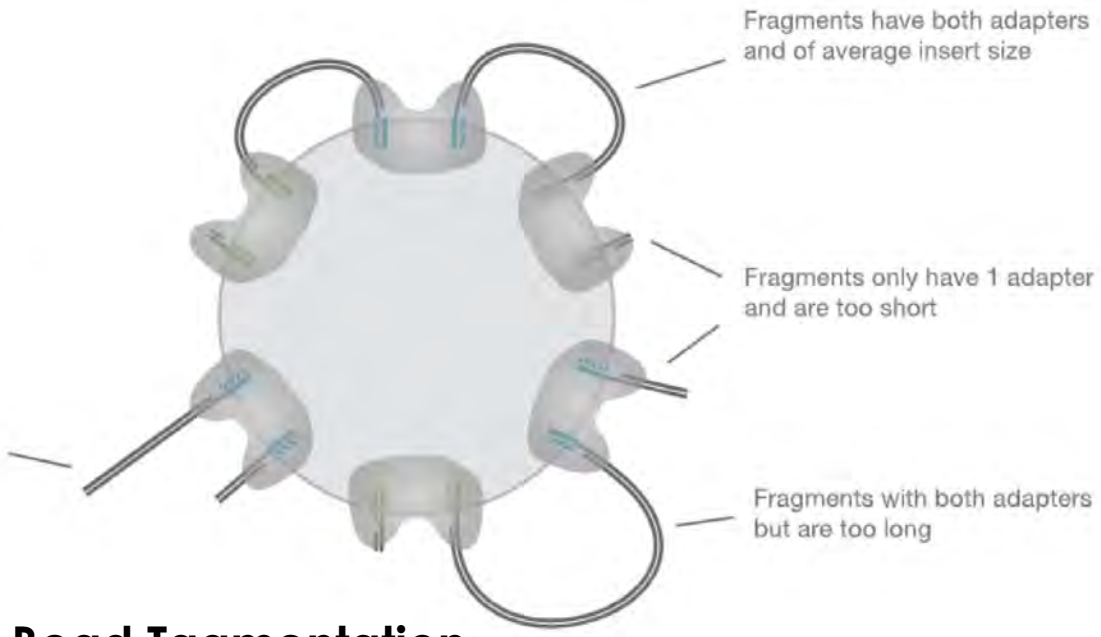




DNA binding & Normalization

Tagmentation is a method that uses a bead-linked transposome (BLT) to fragment dsDNA & insert part of the Illumina adapter sequence.

Illumina DNA Library Prep- Tagmentation



On Bead Tagmentation

Partially Completed Adapter/Library

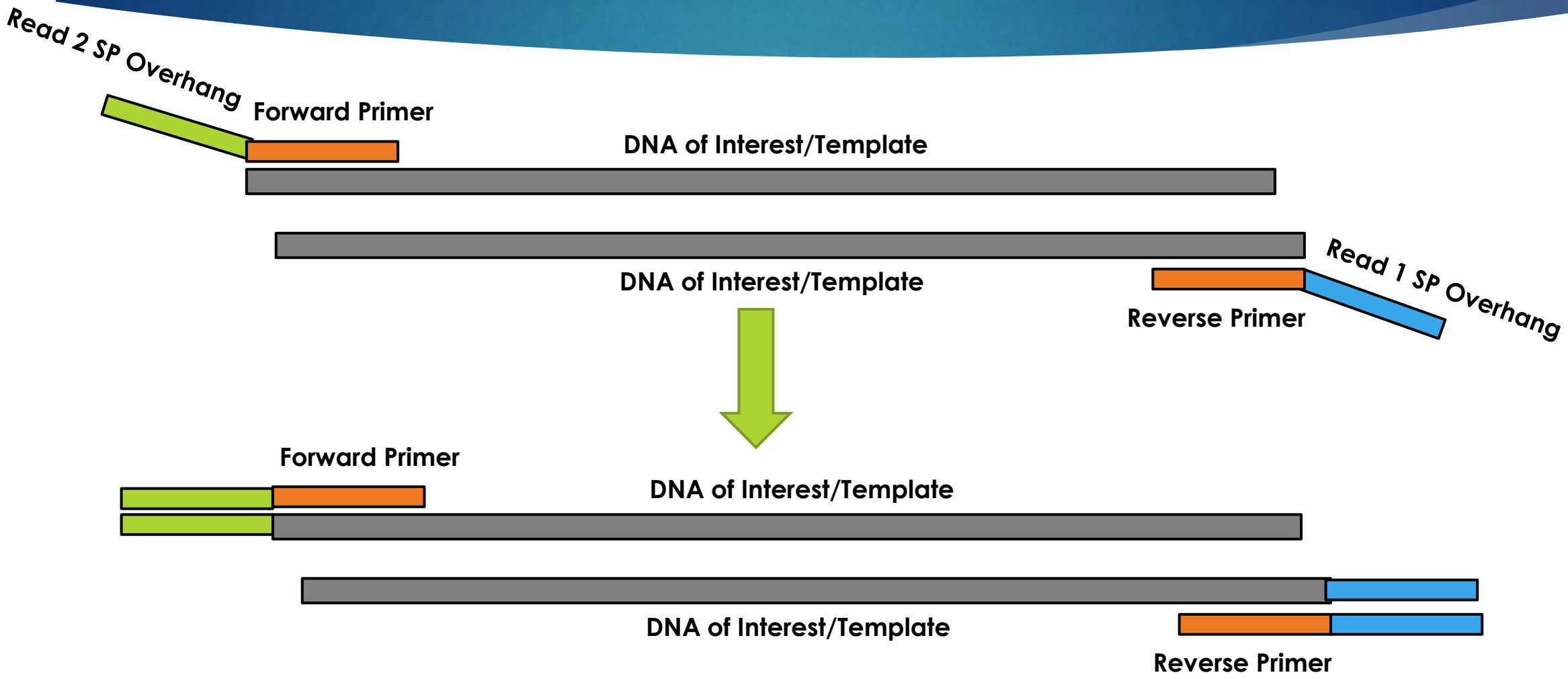
Illumina DNA Library Prep- Tagmentation

Nextera DNA Indexes (Primers)
IDT for Illumina UD Indexes (Primers)

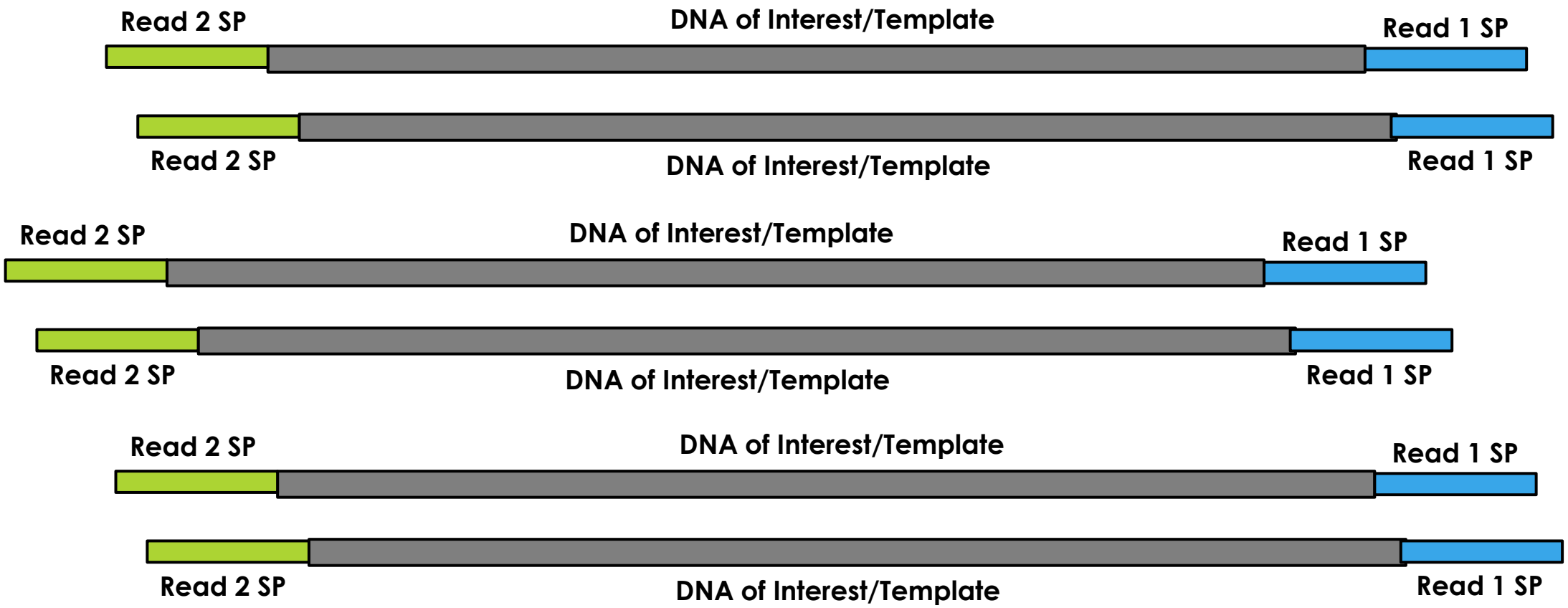
PCR

Completed Library

Illumina DNA Library Prep- PCR Overhang



Illumina DNA Library Prep- PCR Overhang



Illumina DNA Library Prep- PCR Overhang

Nextera DNA Indexes (Primers)
IDT for Illumina UD Indexes (Primers)

PCR

Completed Library

Illumina Library Prep- RNA

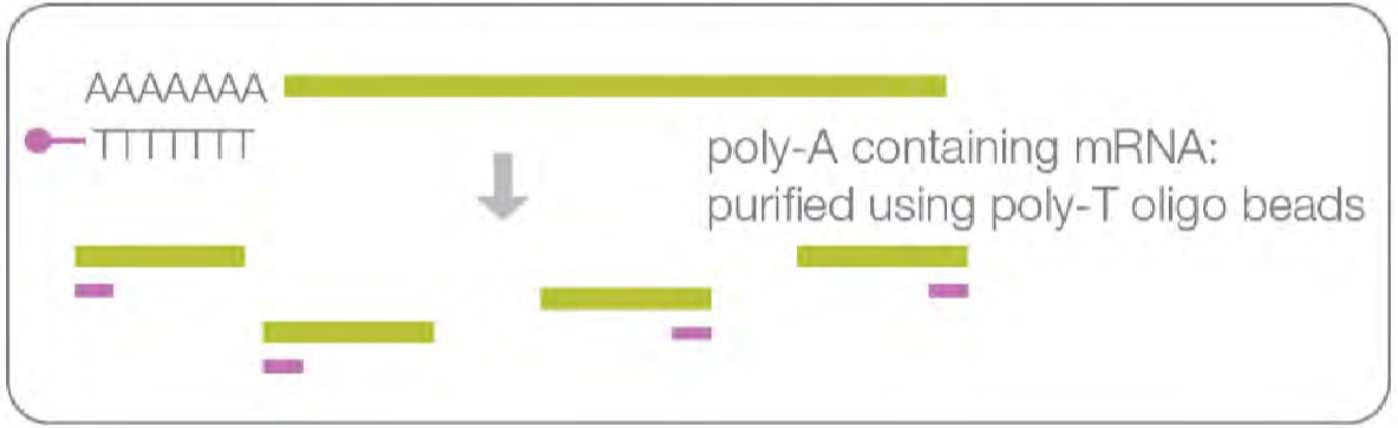
- Typically lower starting concentration as compared to DNA library prep
  - Ligation protocols use lower concentration of adapters to avoid adapter-dimers
- Need to remove rRNA
  - Poly-A selection
    - High quality RNA
    - mRNA (RNA must be poly-Adenylated)
  - rRNA depletion
    - Degraded RNA
    - Total RNA

Illumina Stranded RNA Library Prep- mRNA

Purify and Fragment mRNA

The Poly-A containing mRNA molecules are purified using poly-T oligo attached magnetic beads. Following purification, the mRNA is fragmented into small pieces using divalent cations under elevated temperature.

Figure 1 Purifying and Fragmenting mRNA



Illumina Stranded RNA Library Prep- mRNA

Synthesize First Strand cDNA

Cleaved RNA fragments are copied into first strand cDNA using reverse transcriptase and random primers. Adding Actinomycin D to FSA (First Stand Synthesis Act D mix) prevents spurious DNA-dependent synthesis, while allowing RNA-dependent synthesis, improving strand specificity.

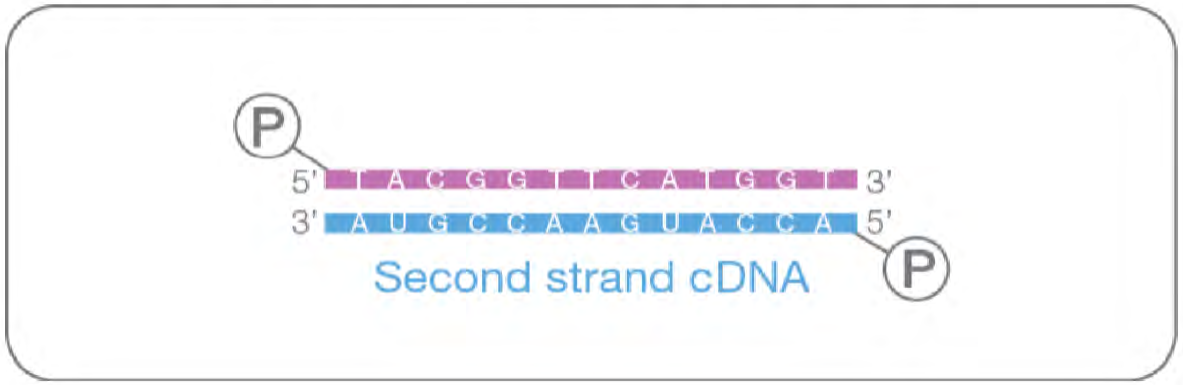
Figure 2 Synthesizing First Strand cDNA

Illumina Stranded RNA Library Prep- mRNA

Synthesize Second Strand cDNA

Strand specificity is achieved by replacing dTTP with dUTP in the SMM (Second Strand Marking Mix), followed by second strand cDNA synthesis using DNA Polymerase I and RNase H. The incorporation of dUTP in second strand synthesis quenches the second strand during amplification.

Figure 3 Synthesizing Second Strand cDNA



Illumina Stranded RNA Library Prep- mRNA

Figure 4 Adenylating 3' Ends

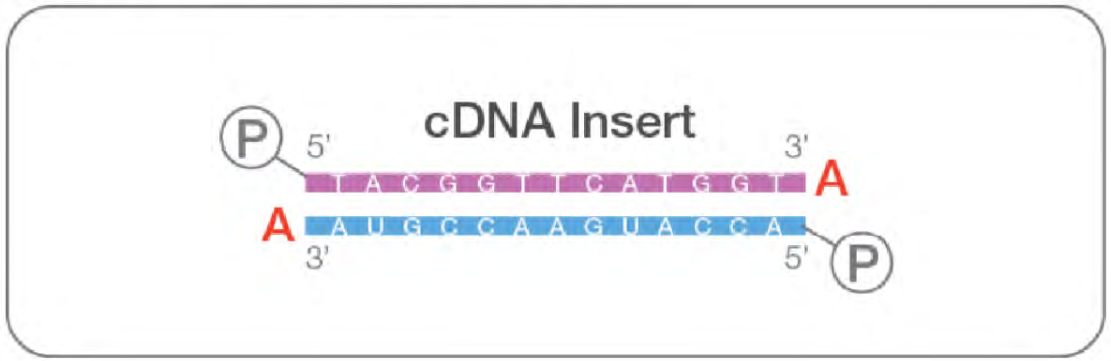
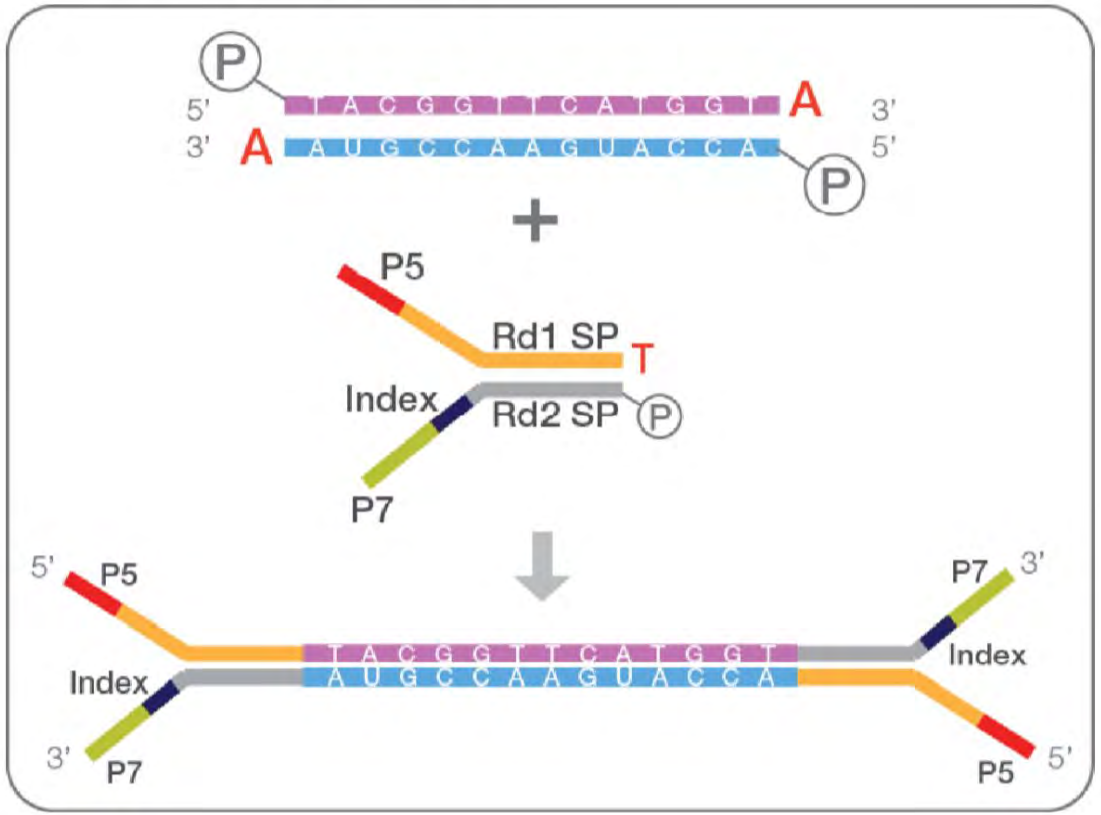
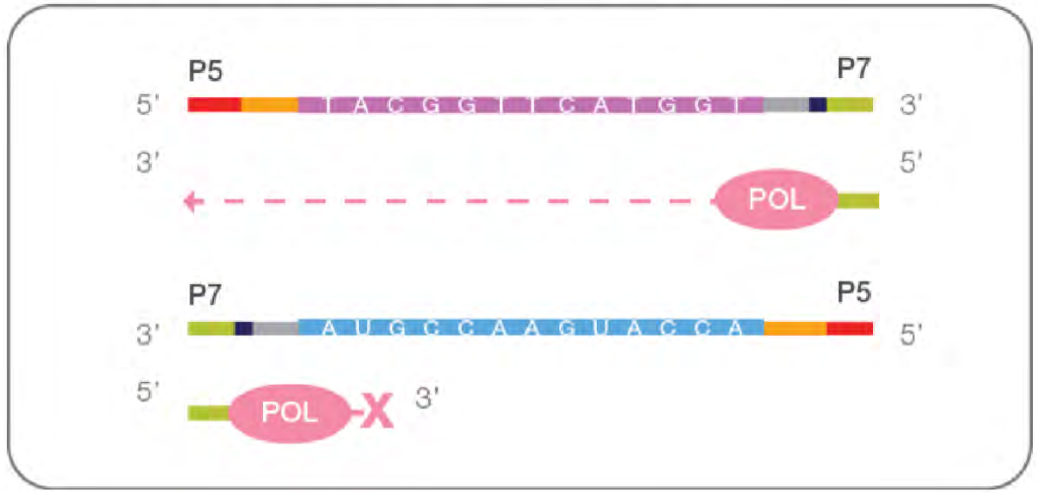


Figure 5 Ligating Adapters



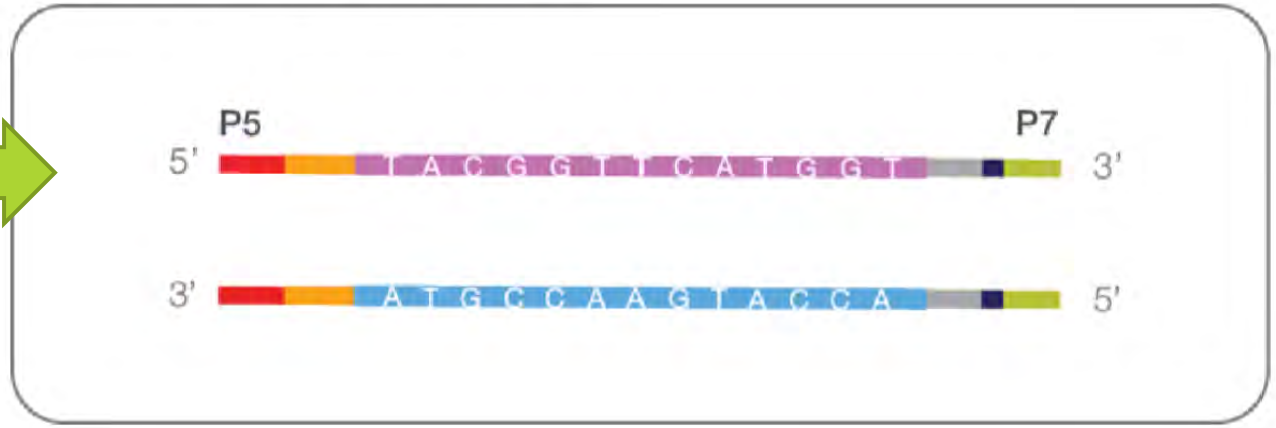
Illumina Stranded RNA Library Prep- mRNA

Figure 6 Enriching DNA Fragments



PCR

Figure 7 LS Final Library



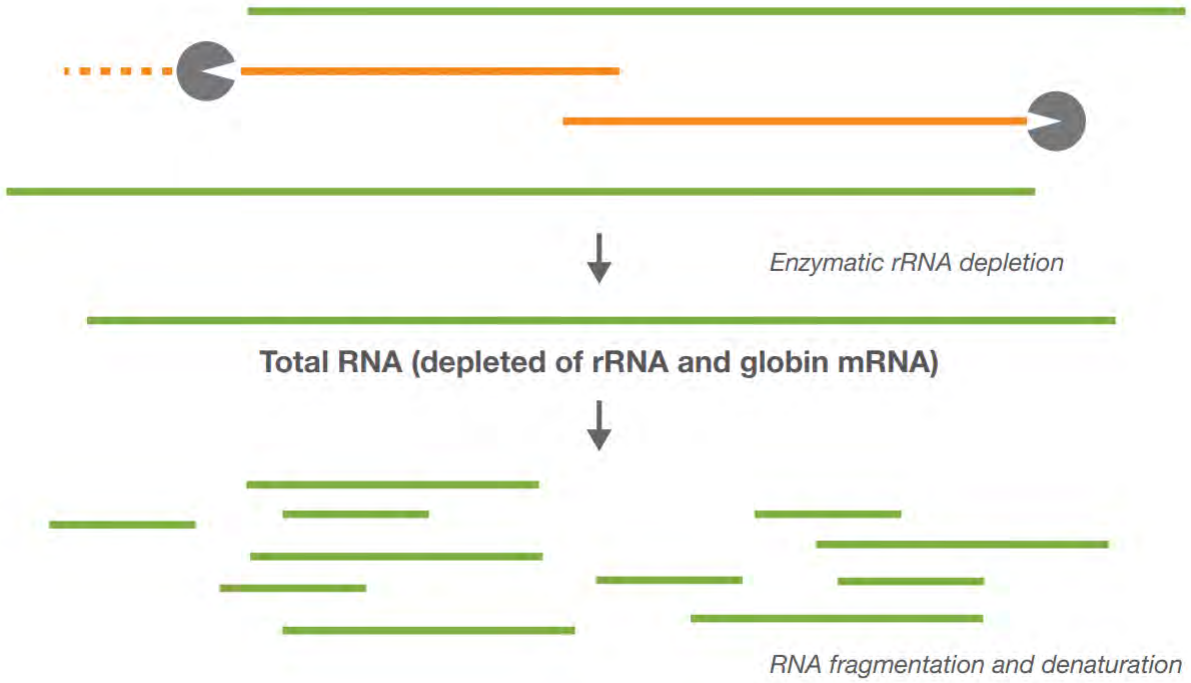
Number of PCR cycles is dependent on starting RNA concentration.
Lower concentration = More cycles

Illumina Stranded Total RNA Prep, Ligation with Ribo-Zero Plus

Illumina Stranded Total RNA Prep with Ribo-Zero Plus

7 hours

1. rRNA depletion
   Single-tube, enzymatic reaction
2. RNA fragmentation and denaturation
3. cDNA synthesis
   Integrated with library prep kit
4. A-tailing
5. Ligation
   Improved ligation-based chemistry
6. PCR amplification
7. Quantification and normalization
8. Ready for sequencing



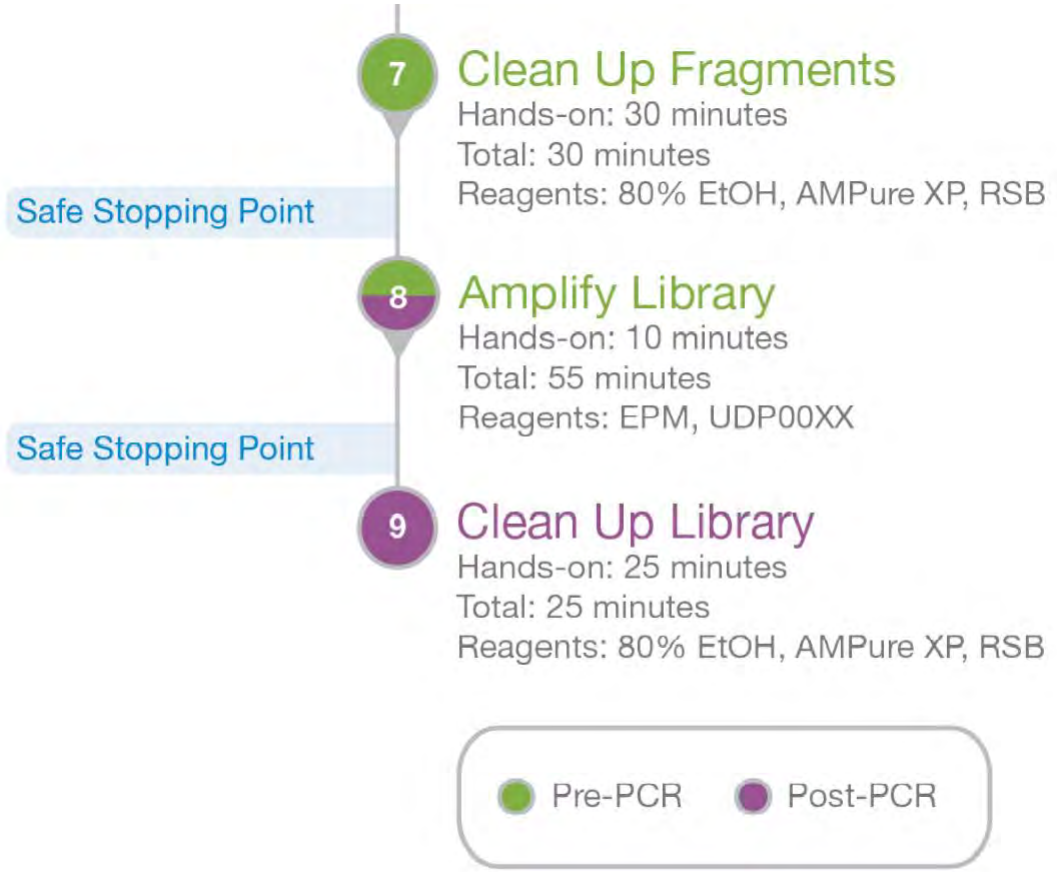
Illumina Stranded Total RNA Prep, Ligation with Ribo-Zero Plus

1. **Deplete Ribosomal RNA**
   Hands-on: 45 minutes
   Total: 1 hour 49 minutes
   Reagents: 80% EtOH, DB1, DP1, ELB, PRB, PRE, RDB, RDE, RNAClean XP
2. **Fragment and Denature RNA**
   Hands-on: 2 minutes
   Total: 7 minutes
   Reagents: EPH3
3. **Synthesize First Strand cDNA**
   Hands-on: 5 minutes
   Total: 50 minutes
   Reagents: FSA, RVT

Safe Stopping Point

4. **Synthesize Second Strand cDNA**
   Hands-on: 35 minutes
   Total: 1 hour 40 minutes
   Reagents: 80% EtOH, AMPure XP, RSB, SMM
5. **Adenylate 3´ Ends**
   Hands-on: 5 minutes
   Total: 45 minutes
   Reagents: ATL4
6. **Ligate Anchors**
   Hands-on: 10 minutes
   Total: 25 minutes
   Reagents: LIGX, RNA Index Anchors, RSB, STL

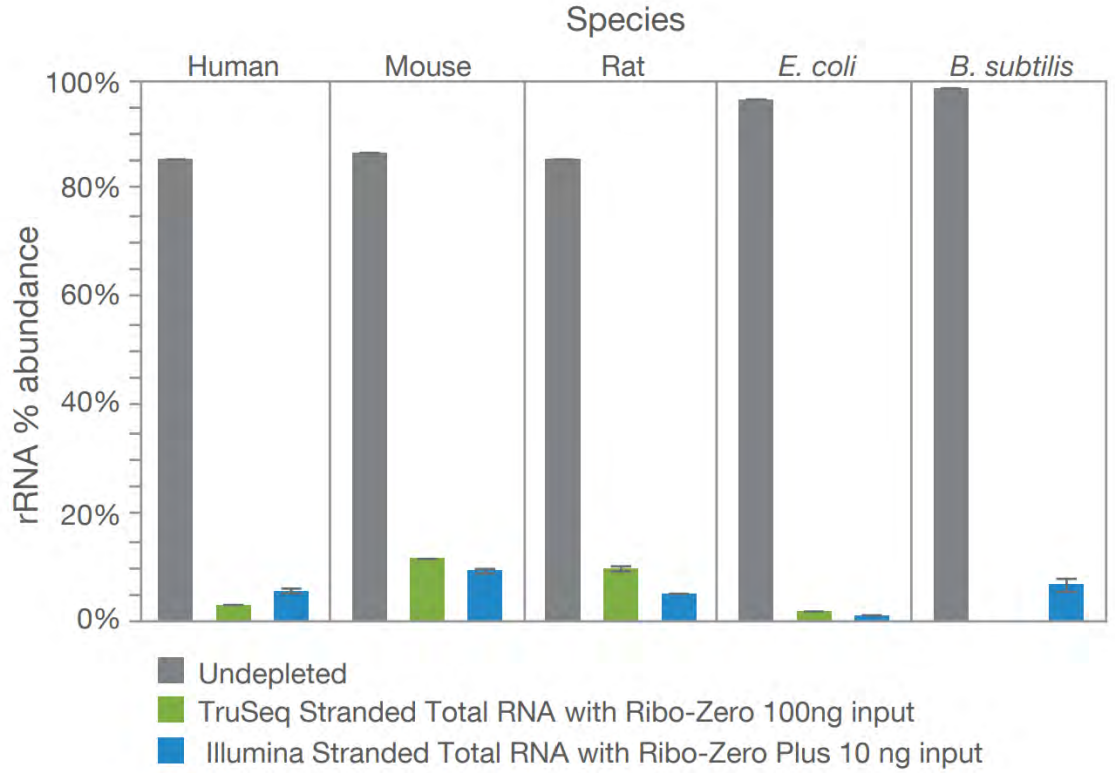
Illumina Stranded Total RNA Prep, Ligation with Ribo-Zero Plus



Illumina Stranded Total RNA Prep, Ligation with Ribo-Zero Plus

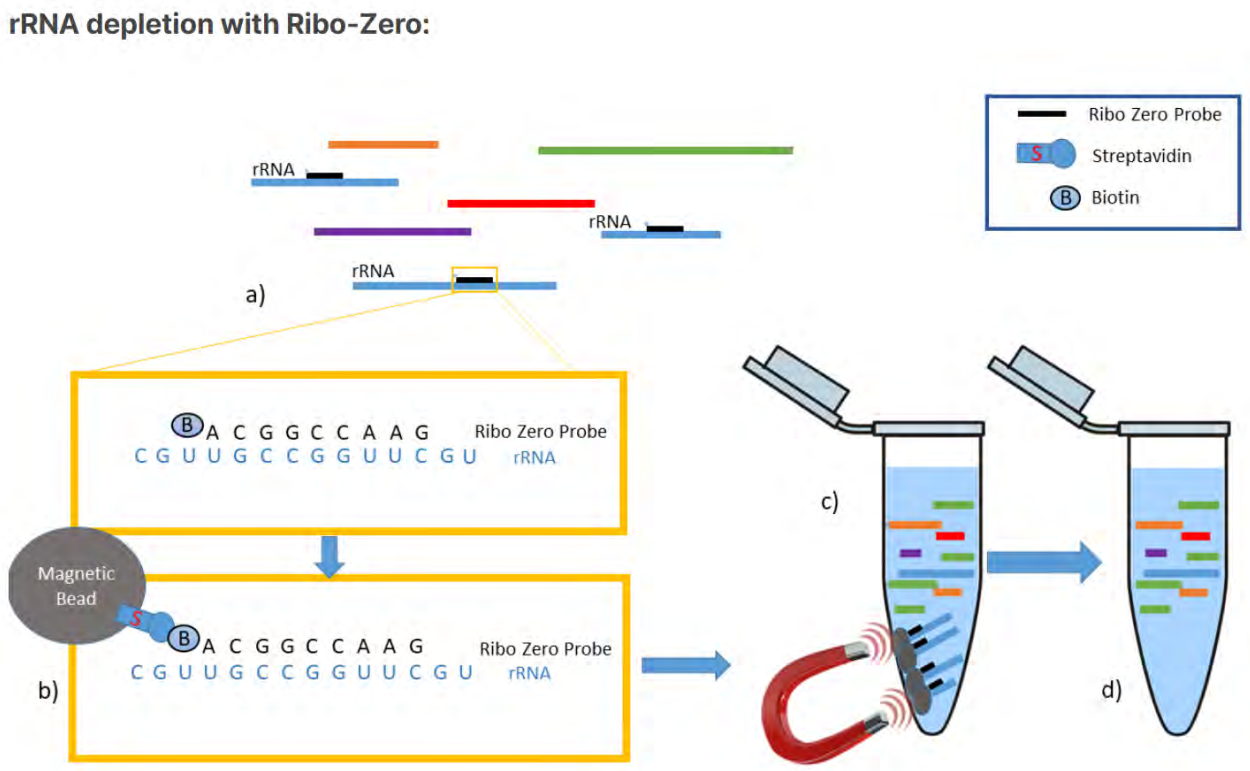
Table 2: RNA species targeted for reduction

Sample	rRNAs targeted
Human cytoplasmic rRNAs	28S, 18S, 5.8S, 5S
Human mitochondrial rRNAs	12S, 16S
Human Beta Globin transcripts	HBA1, HBA2, HBB, HBG1, HBG2
Mouse and rat rRNA	16S, 28S
Gram (-) bacterial rRNAs	*E. coli* 5S, 16S, 23S
Gram (+) bacterial rRNAs	*B. subtilis* 5S, 16S, 23S



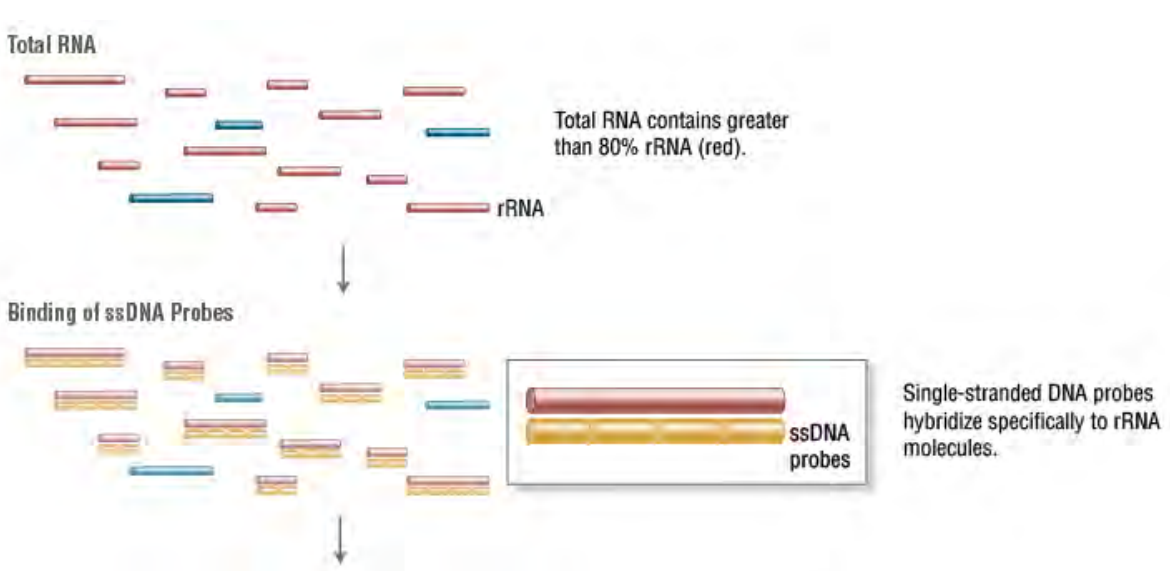
Illumina RNA Library Prep- Additional rRNA Depletion Methods

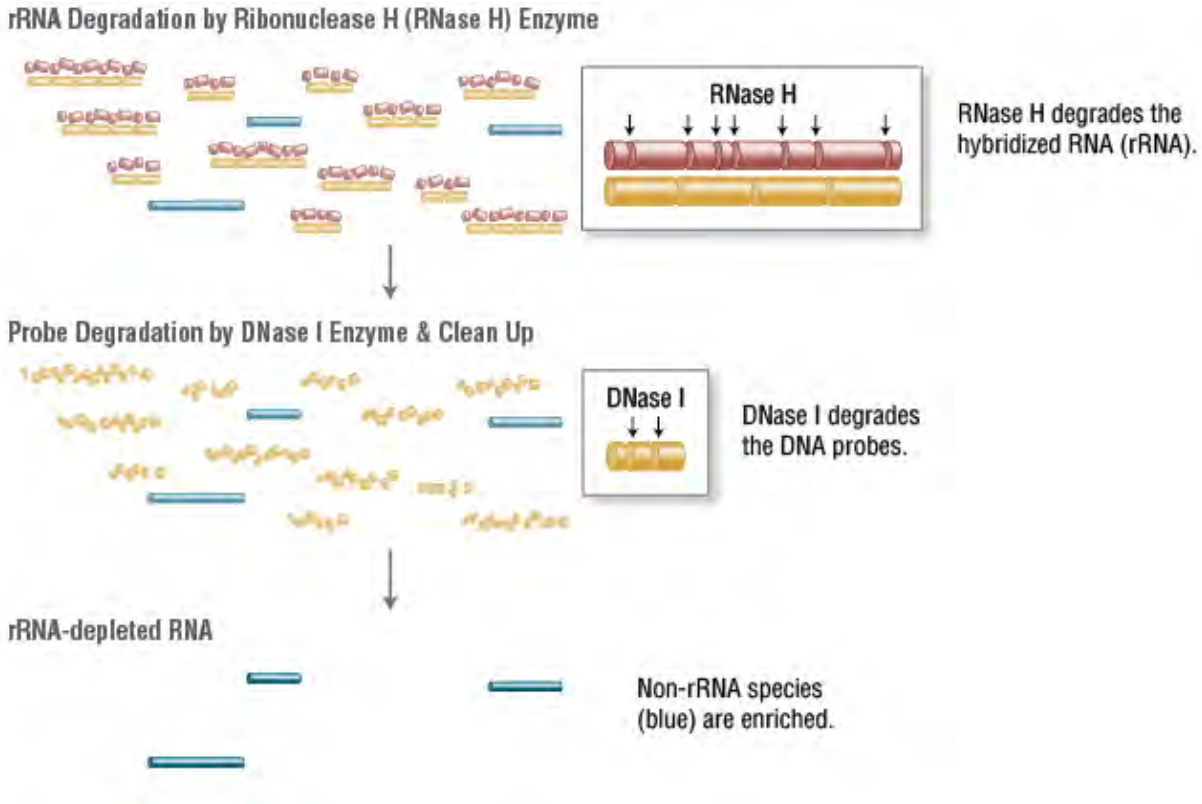
Illumina Original Ribo-Zero



Illumina RNA Library Prep- Additional rRNA Depletion Methods

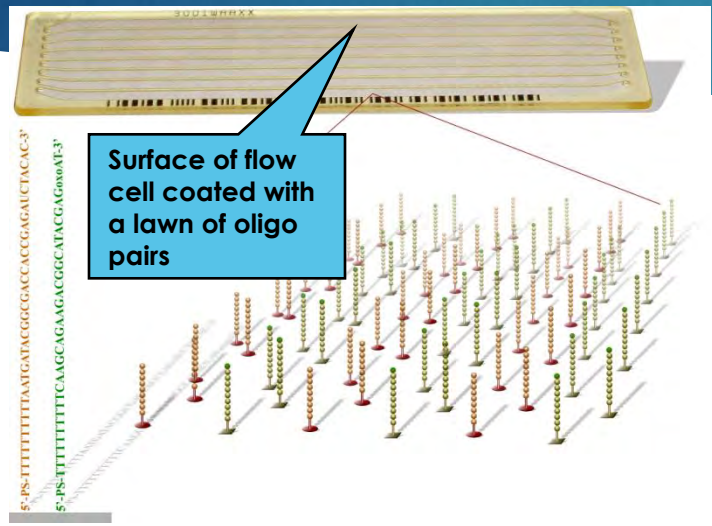
New England Biolabs NEBnext Ribosomal Depletion



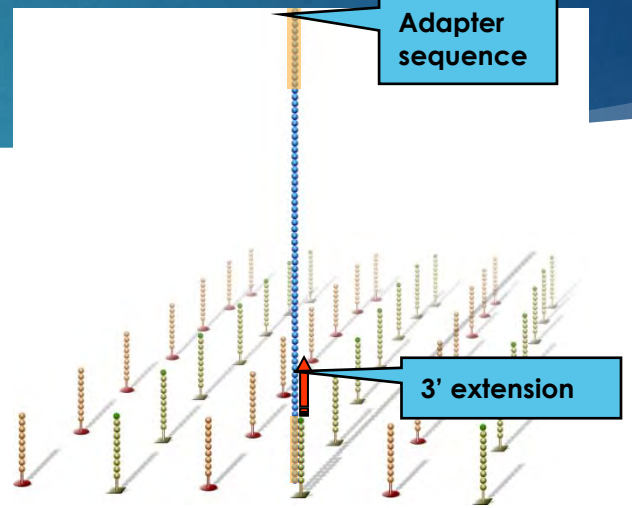


Illumina Sequencing

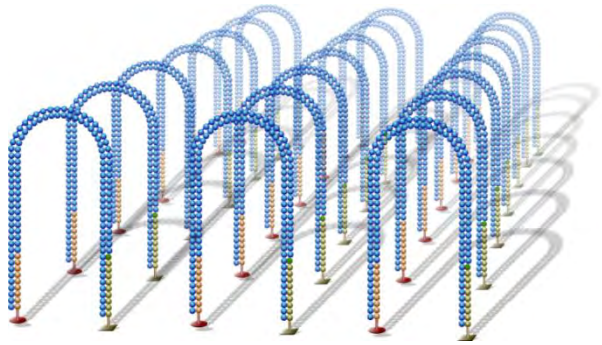
Illumina Cluster Generation



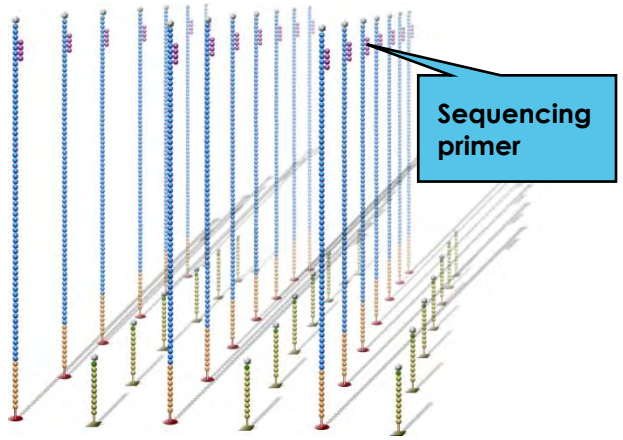
Hybridize library & extend



Bridge formation & amplification

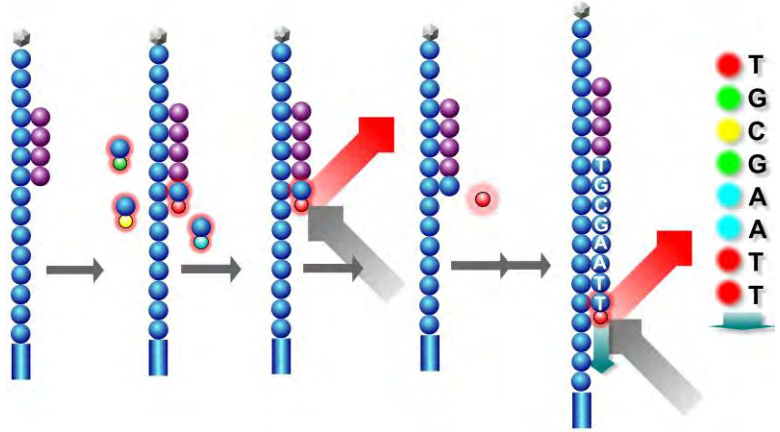


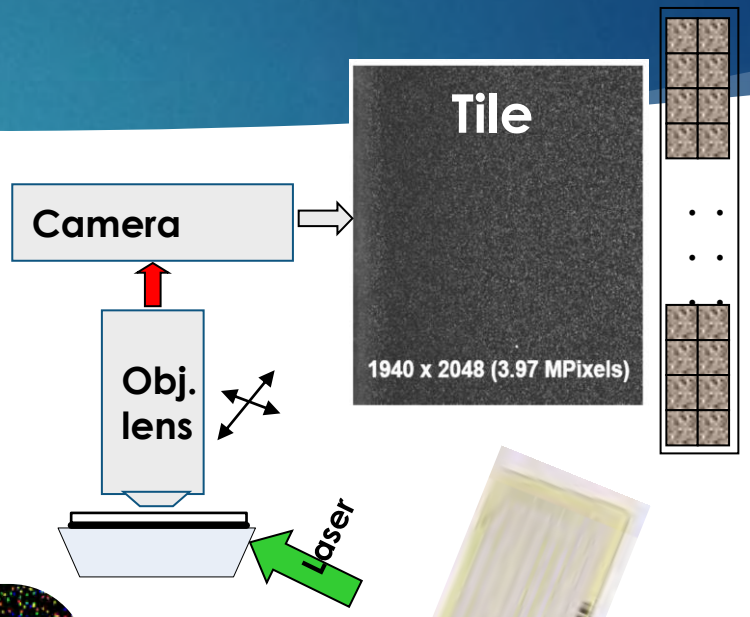
Denature bridges, cleave reverse strands, block 3' ends, & add sequencing primer

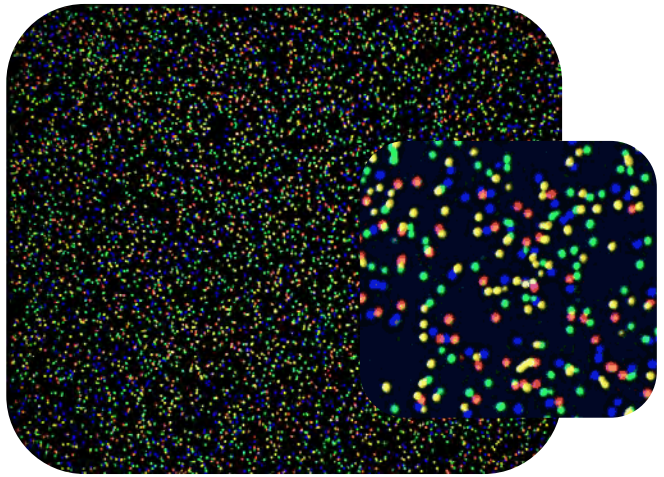


1 cluster has ~1,000 copies of 1 strand of DNA library

Illumina Sequencing by Synthesis

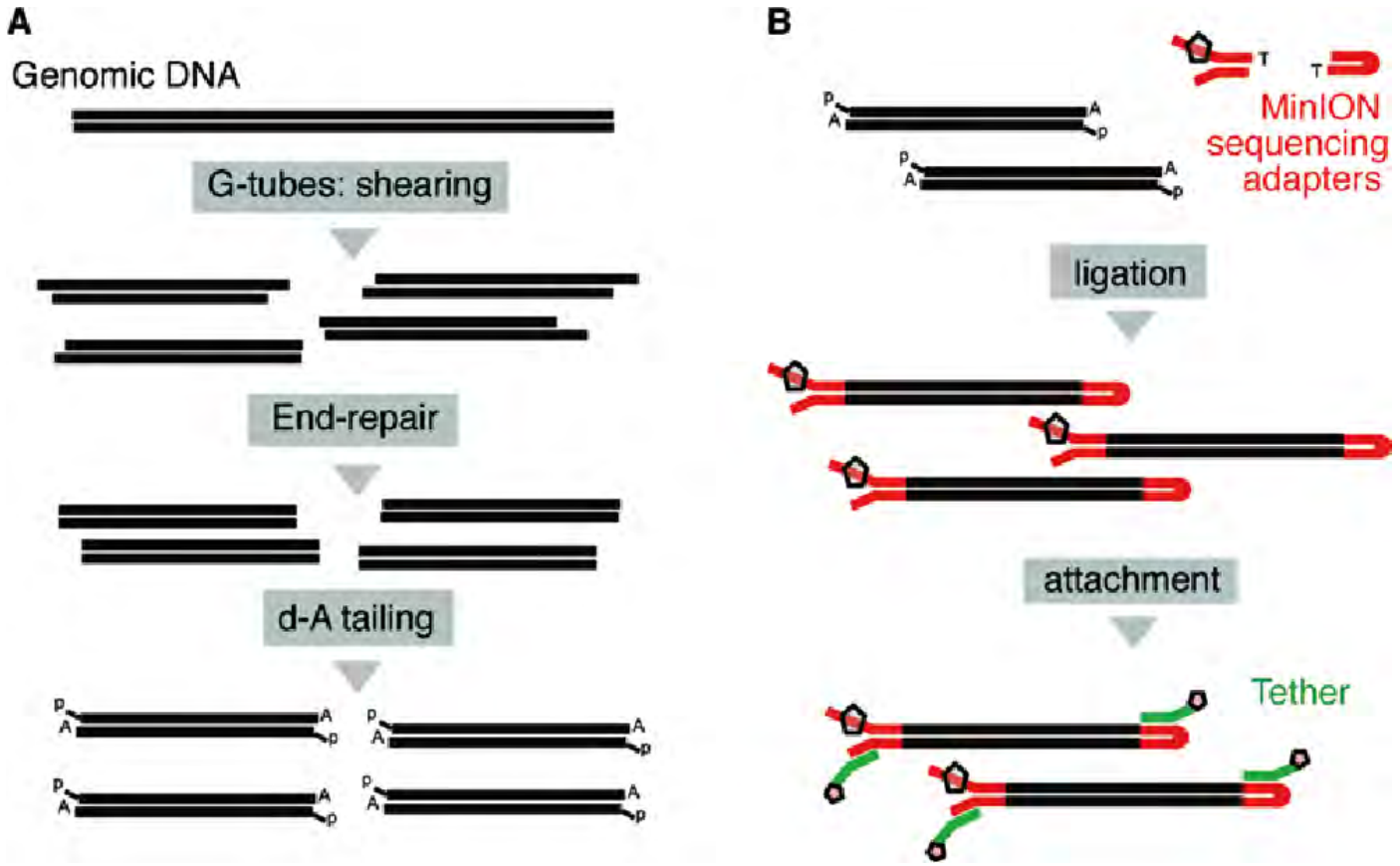




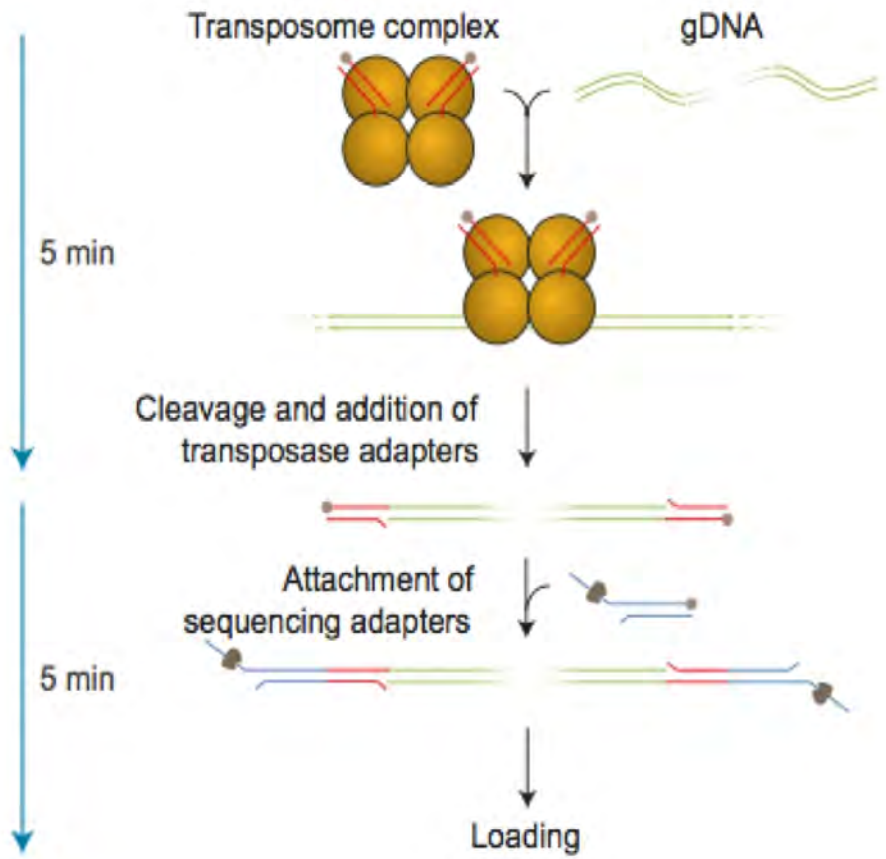


Oxford Nanopore Technologies Next-Generation Sequencing

ONT DNA library prep- Ligation



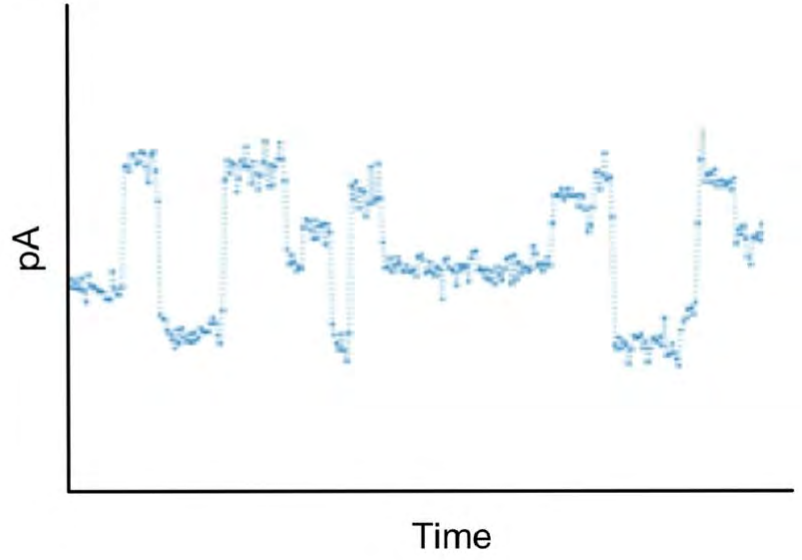
ONT DNA library prep- Rapid barcoding



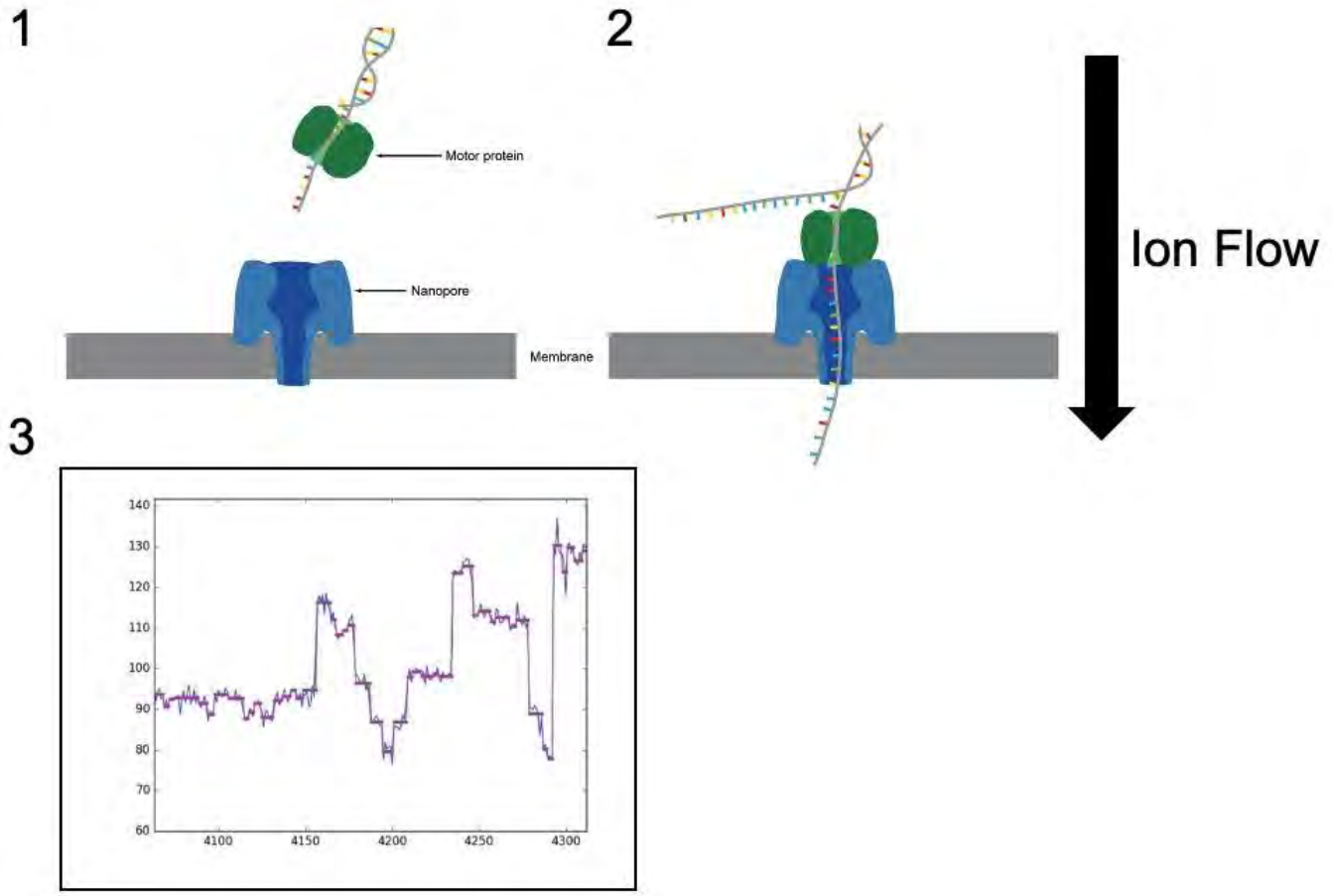
ONT Sequencing

C

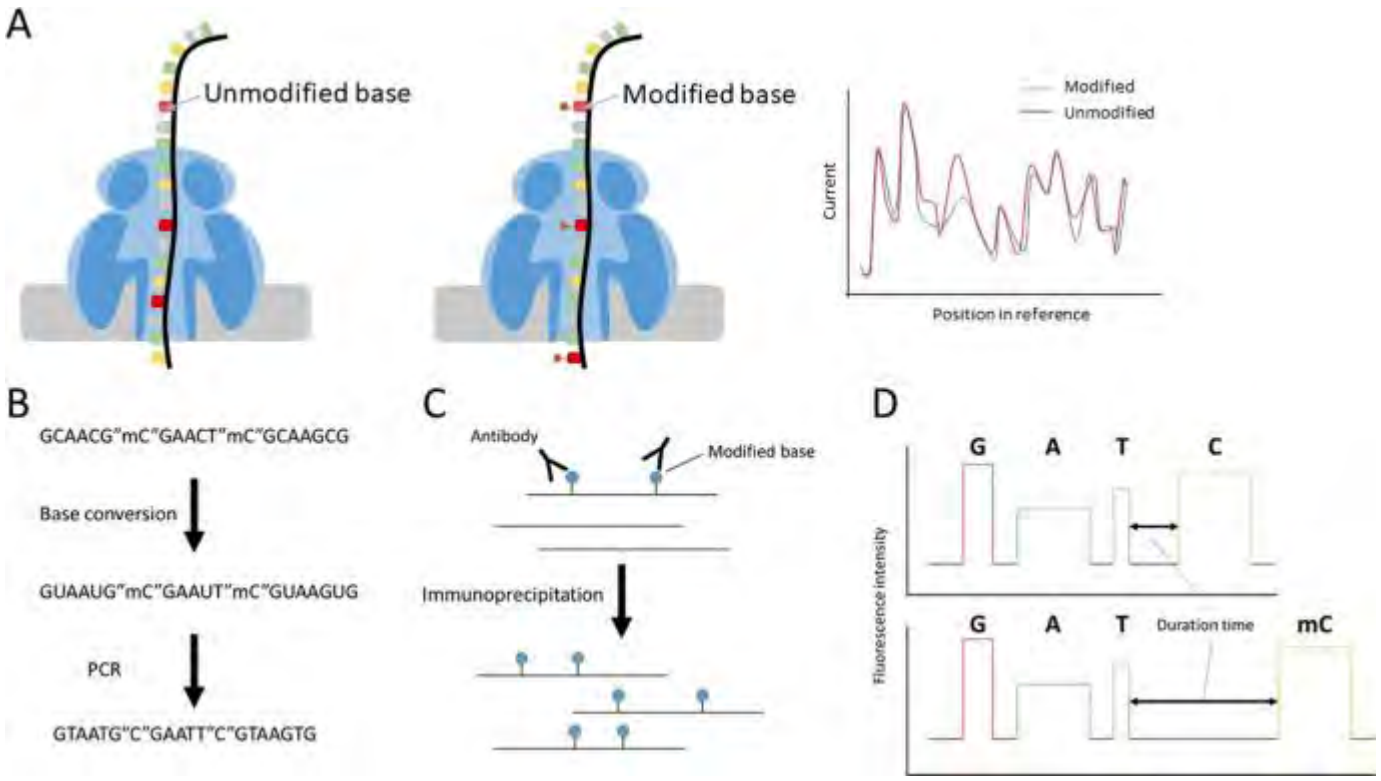
D



ONT Sequencing



ONT Sequencing



Summary

- Next-Generation Sequencing (NGS) describes a variety of technologies that are able to sequencing in a massively parallel manner.
- Library preps are similar across NGS technologies.
  - Ligation-based: fragment, end repair, A-Tail, Ligation of Adapters, PCR
  - Transposase-based: Enzymatic fragmentation coupled with addition of part or all of adapter
- Choice of library prep depends on sample type, sequencing method, and desired data output.