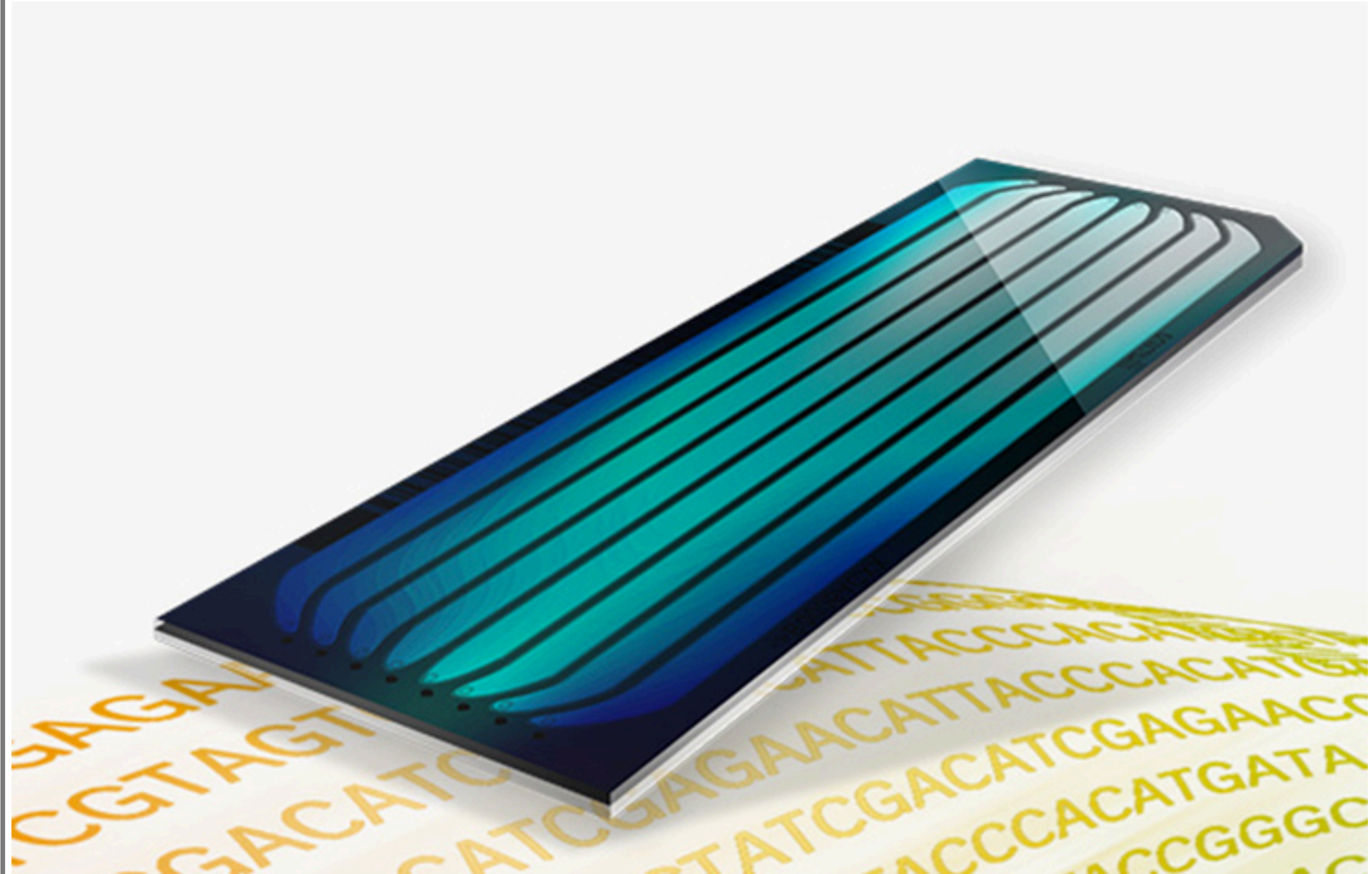
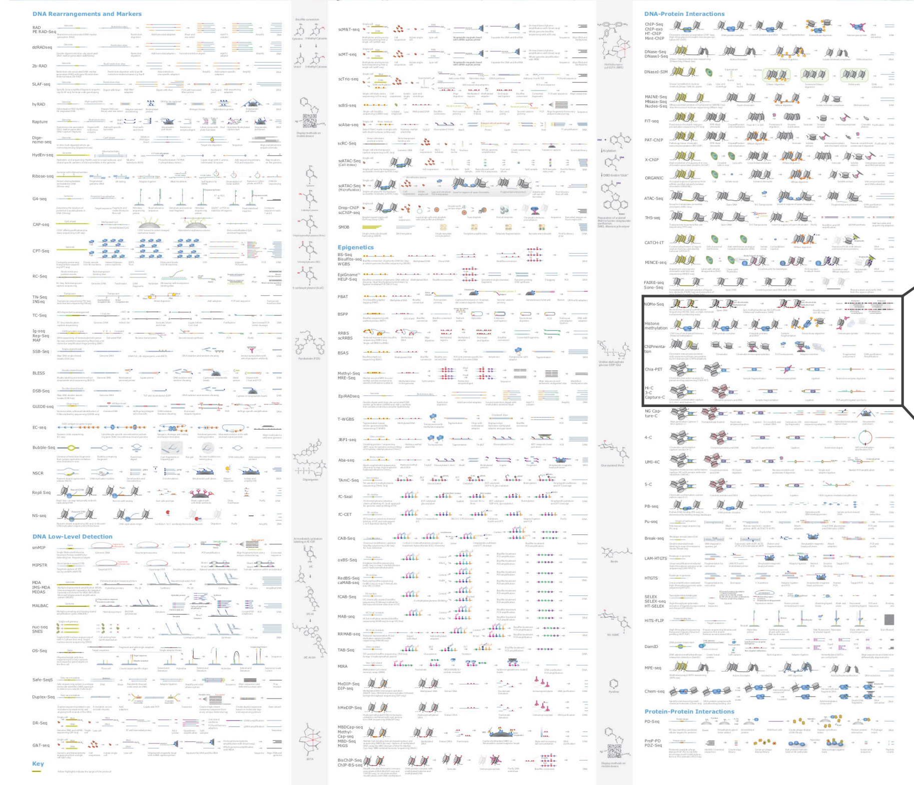
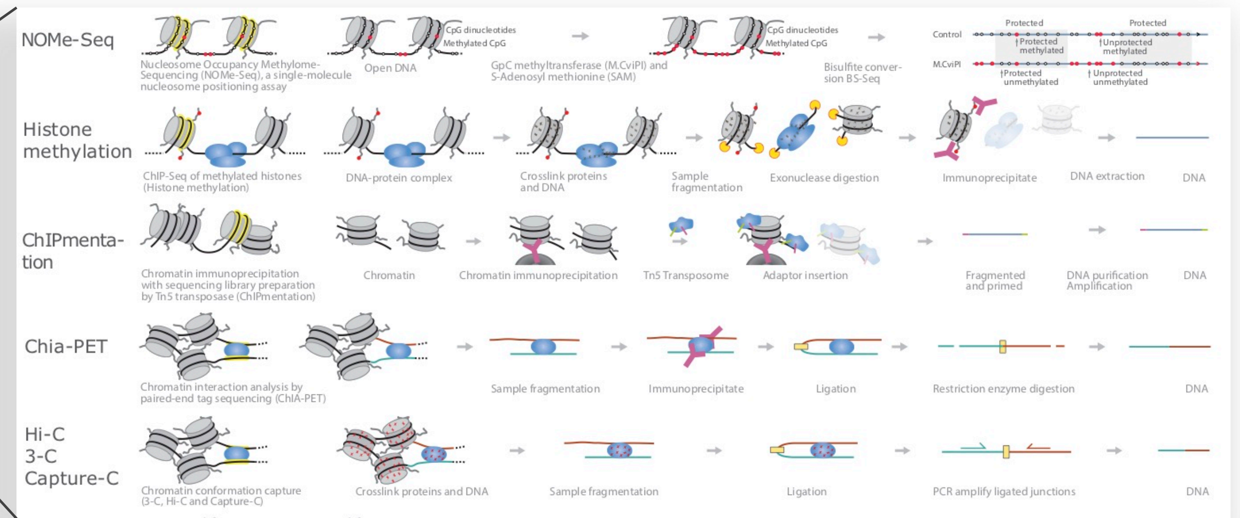
Introduction to Illumina Sequencing



The Overwhelming Landscape of Next-Gen Sequencing Methods





Don't panic! The wide diversity of sequencing methods generally reflects minor variation on a few key themes.

https://www.illumina.com/content/dam/illumina-marketing/documents/applications/ngs-library-prep/ForAllYouSeqMethods.pdf

Short-Read vs Long-Read Next-Generation Sequencing Techniques

Short-Read Sequencing

Illumina

Long-Read (Single Molecule) Sequencing

Oxford Nanopore

PacBio

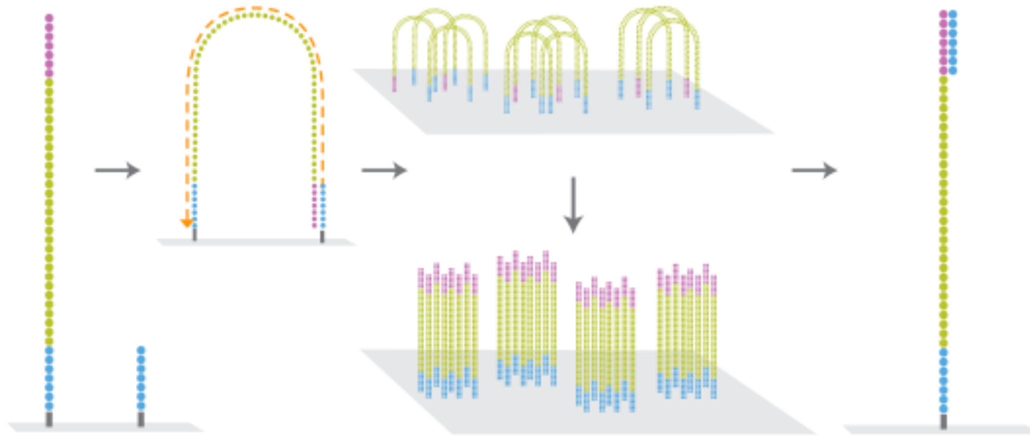
Illumina Sequencing Platforms

	iSeq	MiniSeq	MiSeq	NextSeq	HiSeq	NovaSeq
Max Yield	1.2 Gb	7.5 Gb	15 Gb	120 Gb	1800 Gb	6000 Gb
Max Length	150 bp	150 bp	300 bp	150 bp	150 bp	250 bp

- Illumina sequencers differ predominantly in amount of output (and cost) per run.
- The same sequencing library will generally work on all instruments.

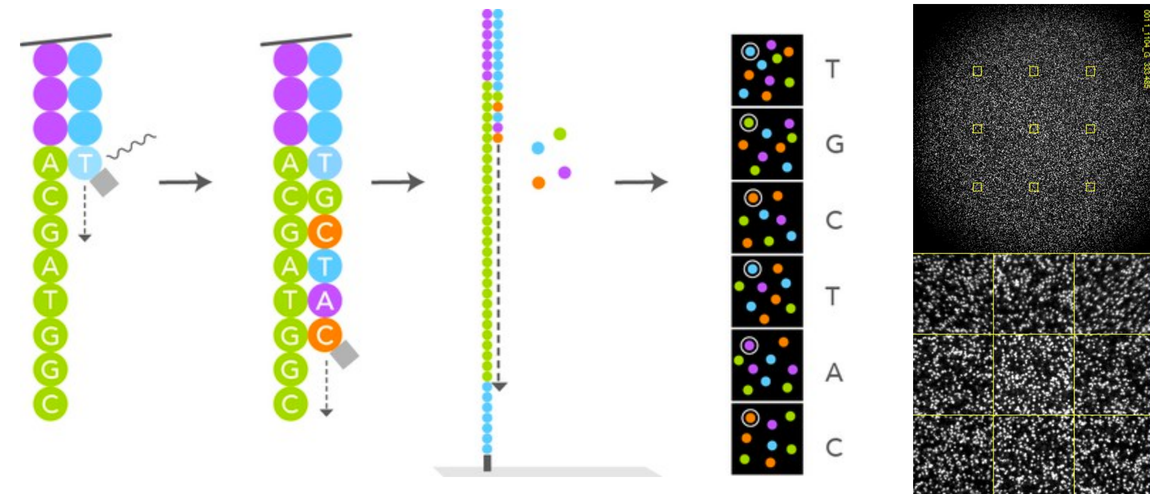
Illumina Sequencing by Synthesis

Cluster Generation



Sequencing libraries are loaded on a flow cell, where each library molecule seeds a cluster and is amplified into thousands of clonal copies by either bridge PCR or “exclusion amplification” (ExAmp).

Sequencing by Primer Extension

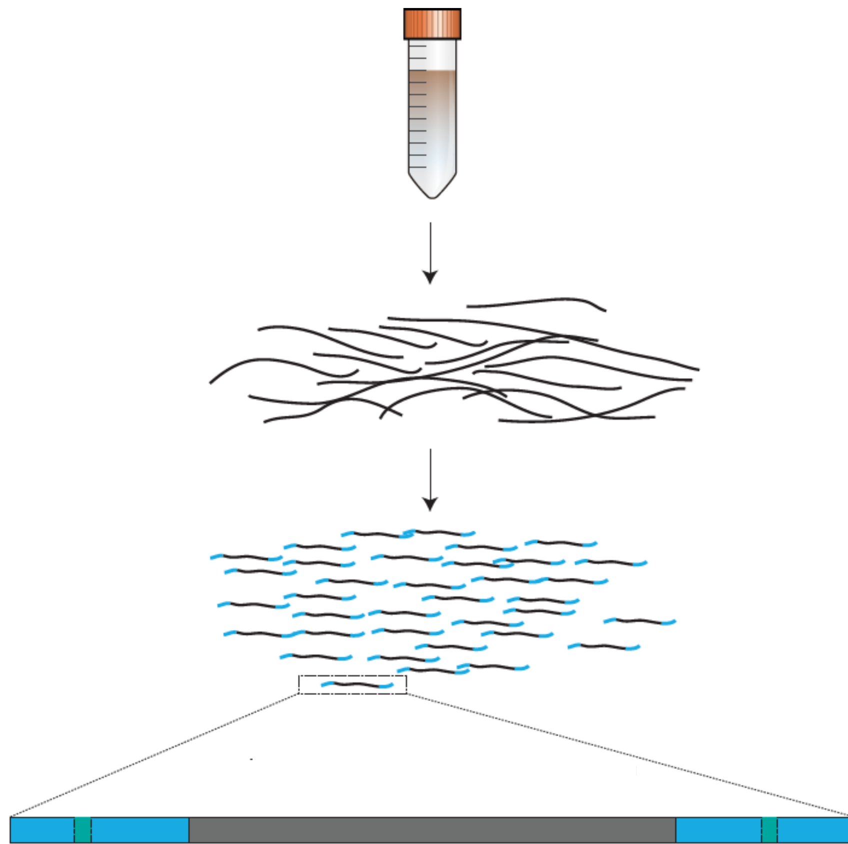


Sequencing proceeds by extending a primer one base at a time (a “cycle”) using reversible chain-terminating and fluorescently labeled nucleotides. The flow cell is imaged after each cycle before proceeding to the next base. Illumina instruments use either 4-color, 2-color, or 1-color chemistry.

https://www.youtube.com/watch?v=fCd6B5HRaZ8

Illumina Library Construction

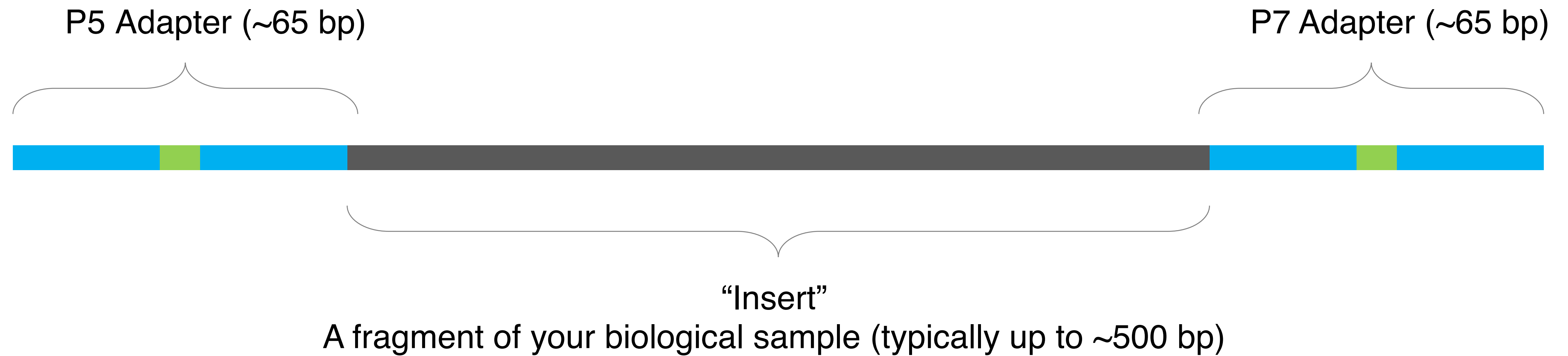
Library construction is the process of converting a nucleic acid sample into a form that is suitable for sequencing.



Common Library Construction Steps
(not all steps required and not necessarily performed in this order)

- Isolation of nucleic acid (DNA or RNA)
- QC of nucleic acid isolates
- Enrichment (of nucleic acid subtypes you want) or subtraction (of those you do not want)
- Fragmentation of nucleic acid
- Addition of adapters to ends of library molecules
- Amplification of library
- Size selection
- Pooling of multiplexed samples
- QC and quantification of final libraries
- Loading on sequencer

Illumina Adapters and Library Molecule Structure



The four functions of Illumina adapters

- Library amplification (PCR primer binding)
- Flow cell binding
- Index sequences (barcodes) for multiplexing
- Sequencing primer binding

Illumina Adapters

1. Library Amplification

P5 Amplification Primer

P7 Amplification Primer

- Most library construction protocols include a PCR amplification step (in addition to the amplification that occurs on the flow cell during cluster generation).
- Adapters provide universal sequences such that all library molecules can be amplified with a common set of primers.

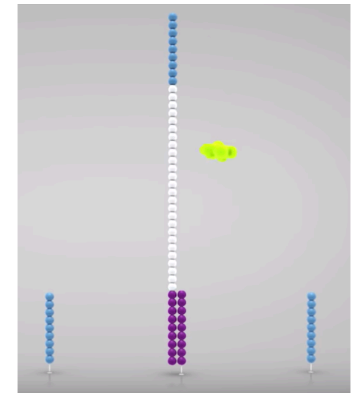
Illumina Adapters

2. Flow Cell Binding

P5 Flow Cell Binding Site

P7 Flow Cell Binding Site

- The same adapter regions used for initial library amplification are also complementary to the oligos that are anchored to the surface of Illumina flow cells.
- Flow cell binding enables subsequent cluster generation.



Illumina Adapters

3. Index Sequences (Barcodes) for Multiplexing

- "Multiplexing" involves pooling libraries from different biological samples to be sequenced together on the same flow cell.
- The i5 and i7 index sequences are barcodes that are shared by all molecules from the same library so that libraries can be distinguished from each other during data analysis.

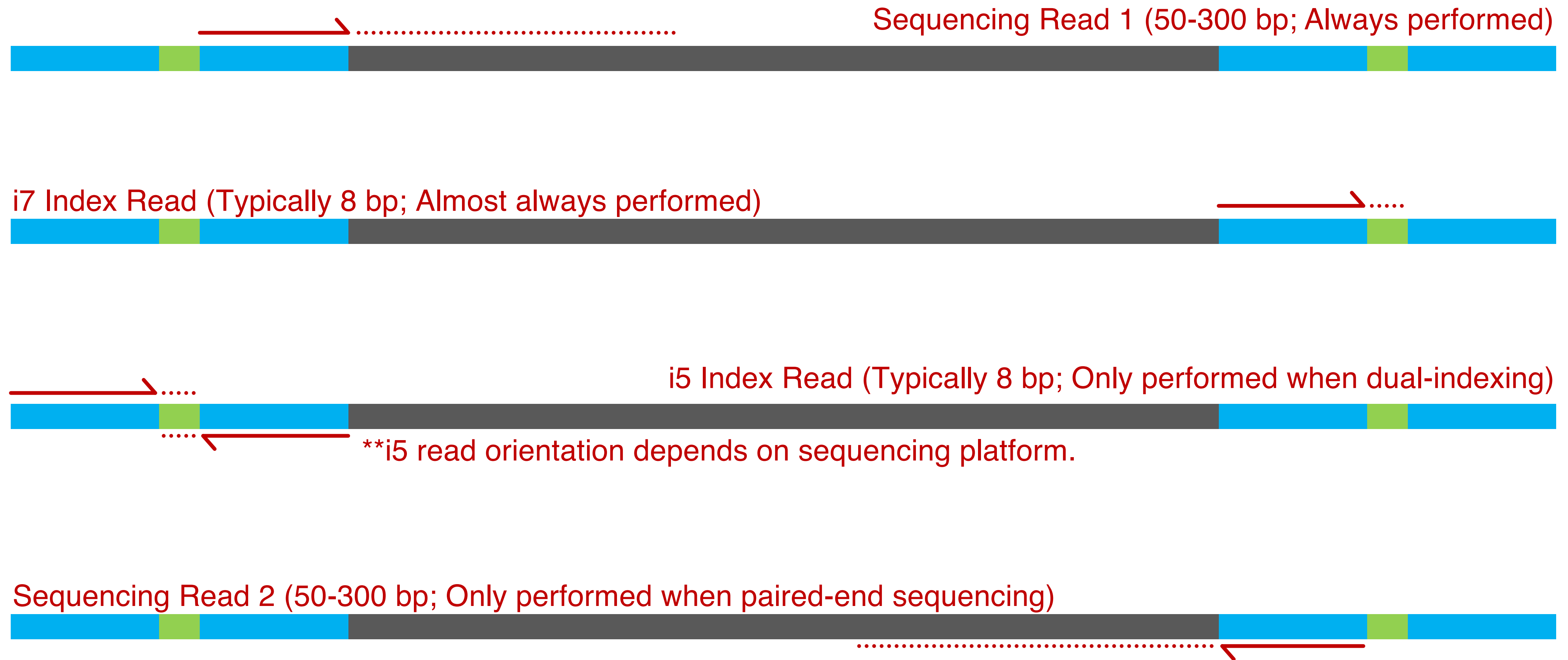
Illumina Adapters

4. Sequencing Primer Binding

- The Sequencing by Synthesis (SBS) process is initiated by primers that bind to specific regions of the Illumina adapters.
- A second sequencing read is initiated from the other sides of the insert when performing paired-end sequencing.

Illumina Reads

An Illumina run will actually produce up to four “reads” per molecule.



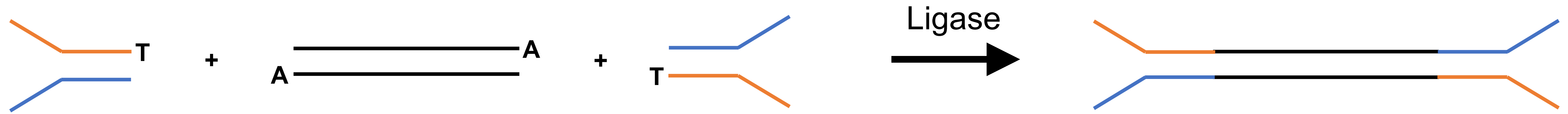
Library Construction Questions

- Imagine that you have fragmented a genomic DNA sample into ~200 bp pieces and then performed Illumina library construction. At what size would you expect to see a band if you run your library on a gel or similar electrophoresis device like a TapeStation?
- If you then sequence this Illumina library on a 2x150 bp run (paired-end sequencing with 150 cycles for each read), will your read pairs overlap? If so, by how much on average?
- If your sample is from an organism with a 2-Gb genome size and you want to obtain an average of 20x coverage to call variants relative to a reference, how many clusters (read pairs) will you have to sequence on your 2x150 run for this library?
- If your fragmentation produced some inserts with a smaller size of 50 bp, what will happen when these library molecules are sequenced on the 150-cycle run?

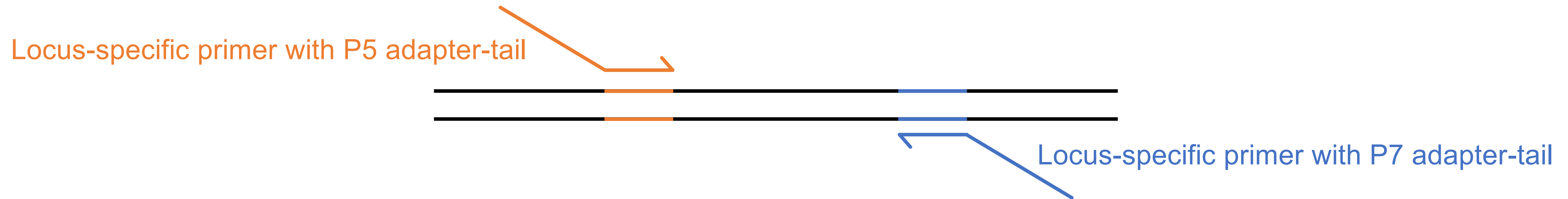
Methods to Attach Illumina Adapter

There are three primary ways to attach adapters to biological inserts.

1. Ligation



2. PCR Primers



3. Tagmentation: Simultaneous fragmentation and adapter incorporation by transposase