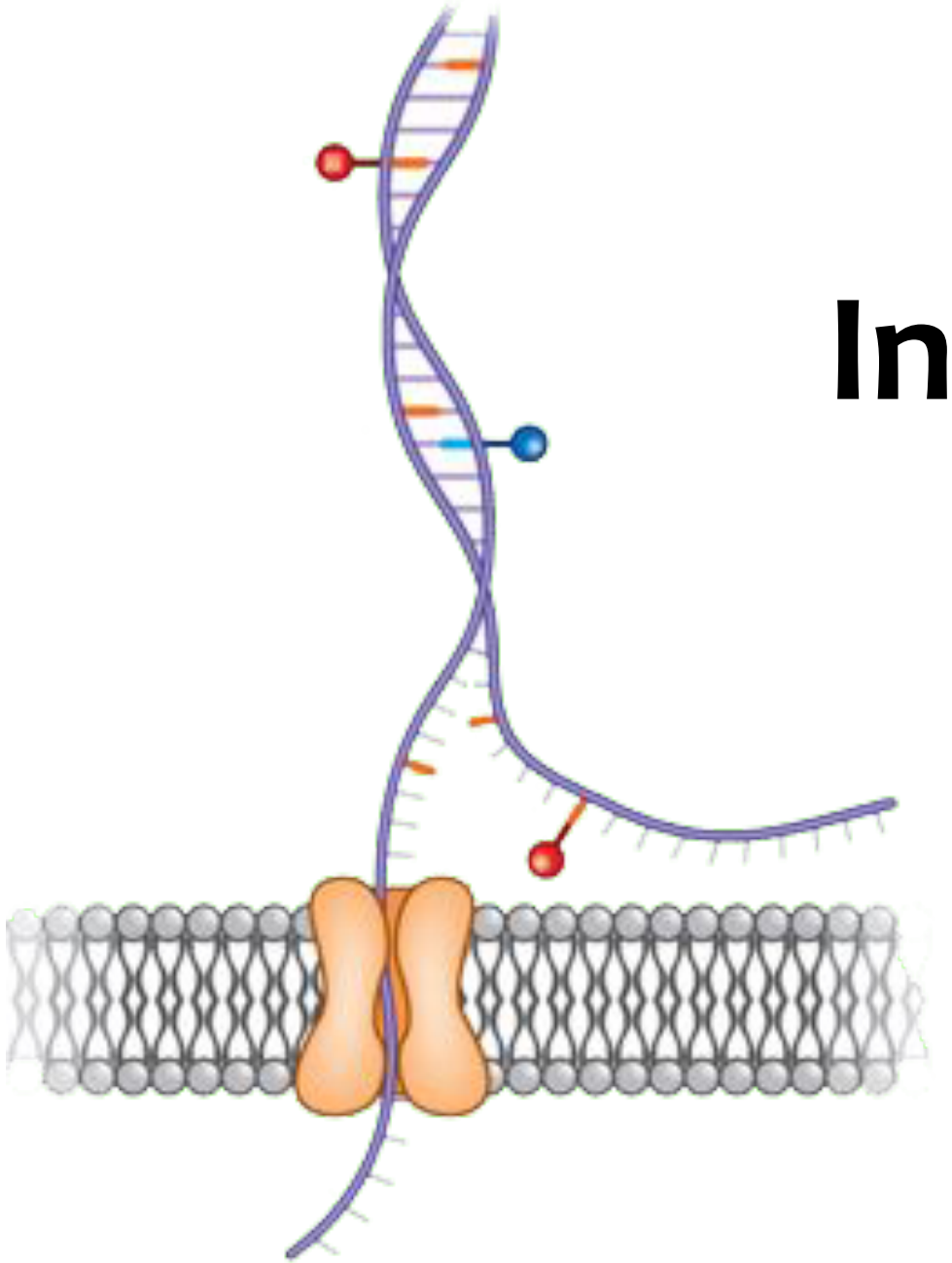
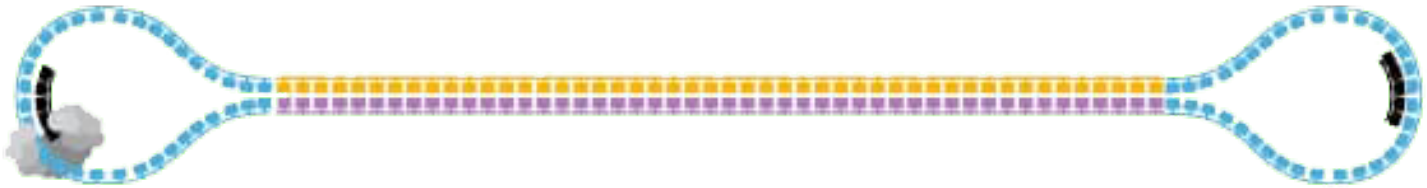


Intro to long-read sequencing



Intro to long-read sequencing

- How do long-read sequencing technologies work?
- When is long-read sequencing the right/wrong choice?
- Genomic and Transcriptomic applications

Intro to long-read sequencing

- **How do long-read sequencing technologies work?**

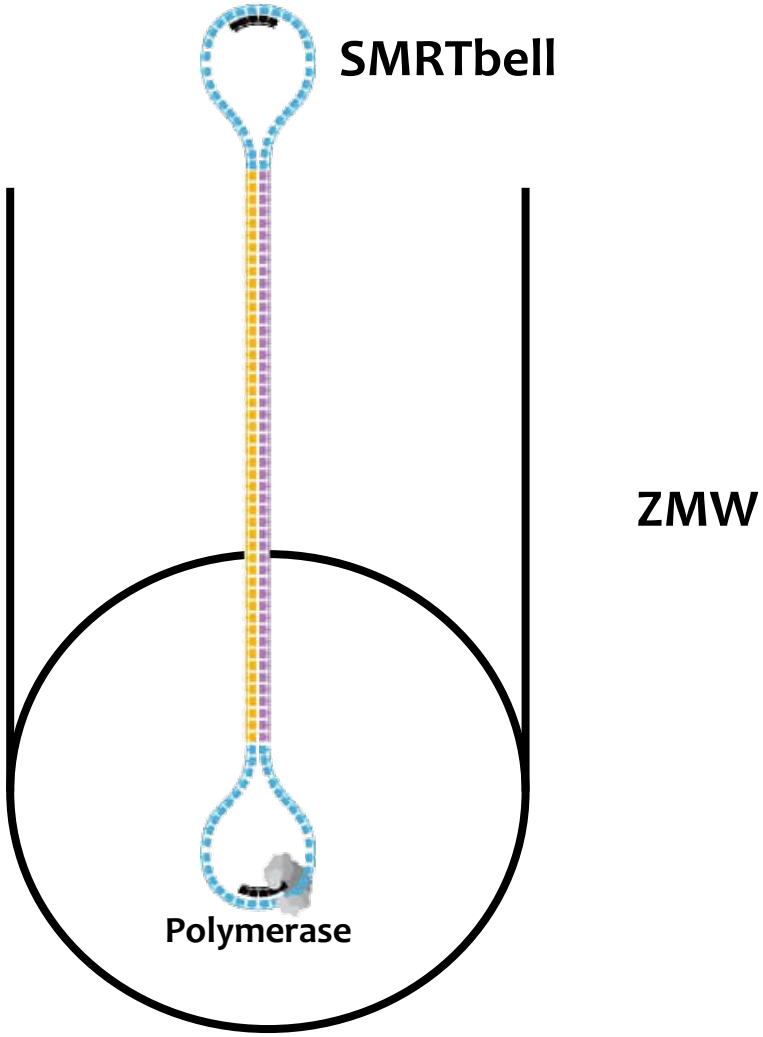
PacBio SMRTbells

Polymerase binds to SMRTbell (**S**ingle-**M**olecule **R**eal-**T**ime), performs **sequencing-by-synthesis** inside ZMWs

Fluorophore emits light at nucleotide incorporation

Movie for each ZMW is parsed to produce read calls

- 16hr, 20hr, 30hr movies

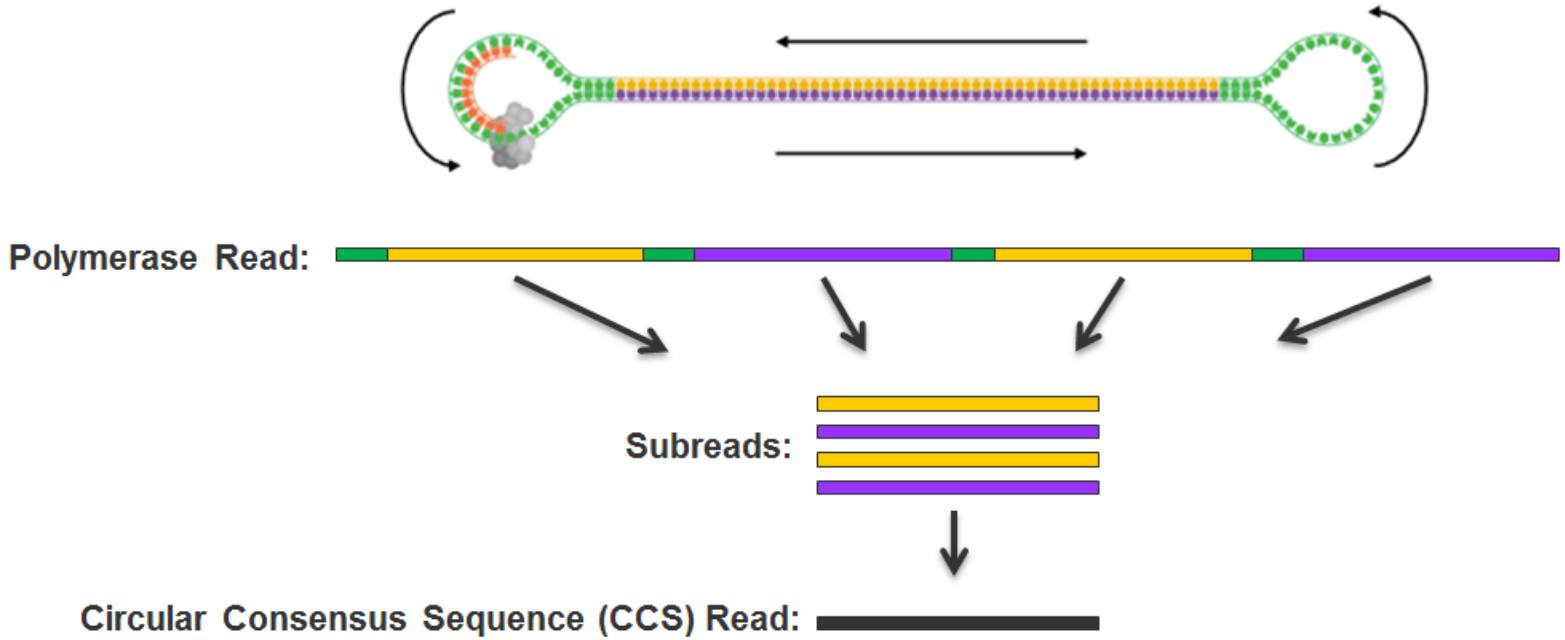


PacBio SMRTbells

Circular Consensus Sequencing Reads (**CCS Reads**) are produced when polymerase goes around SMRTbell ≥3 times

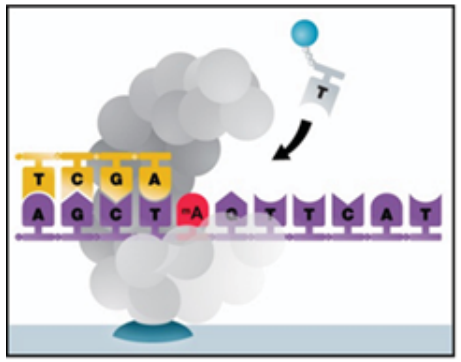
Can provide confidence for allele calling from single molecule, as a CCS read

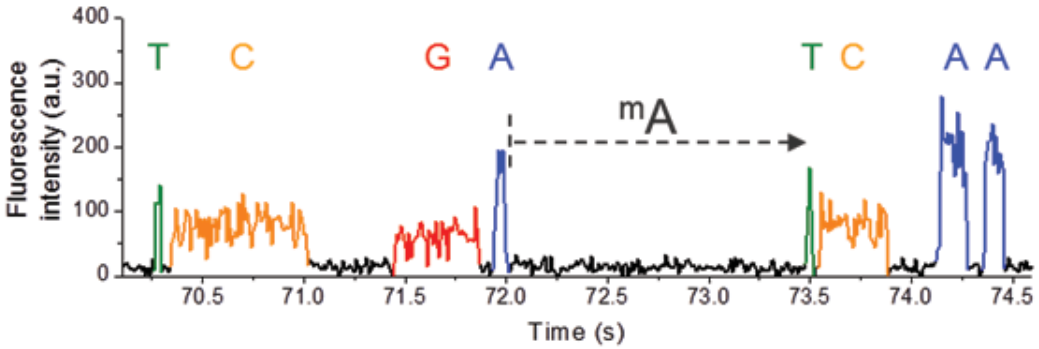
Large inserts (≥50 kbp) are unlikely to form CCS reads



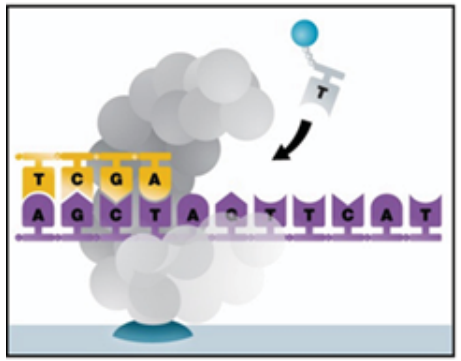
Detecting Base Modifications/Damage with PacBio SMRT bells

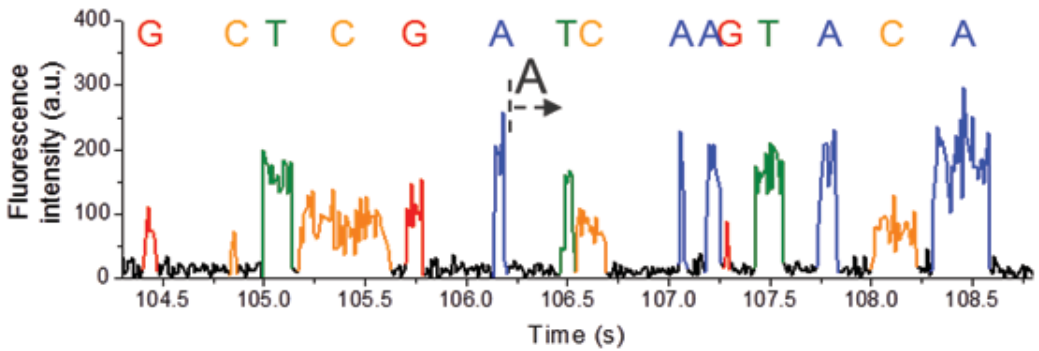
Base modifications impede polymerase processivity in a predictable manner



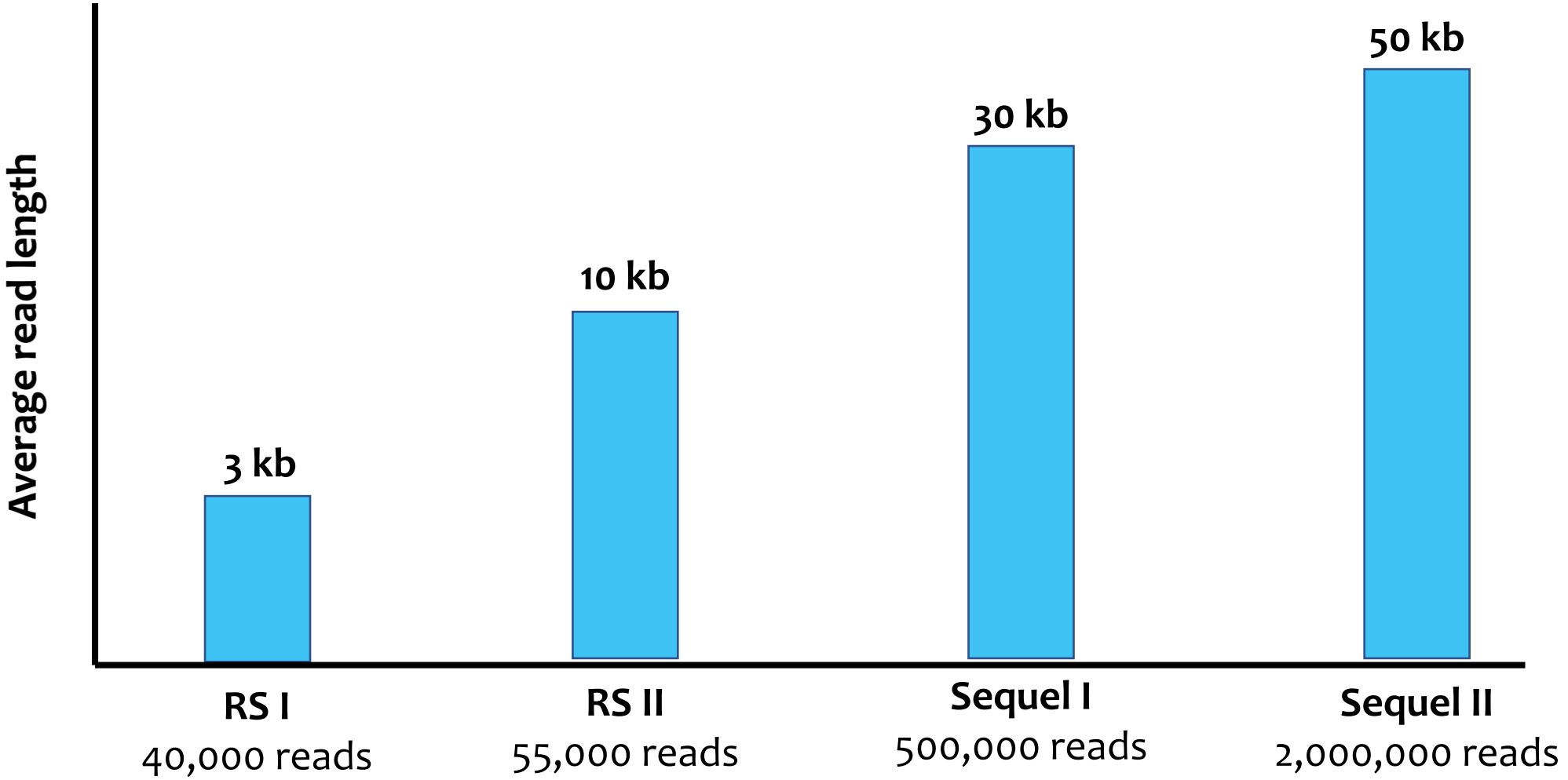


Can be measured with **Inter-pulse Distance (IPD)**





PacBio read length is increasing



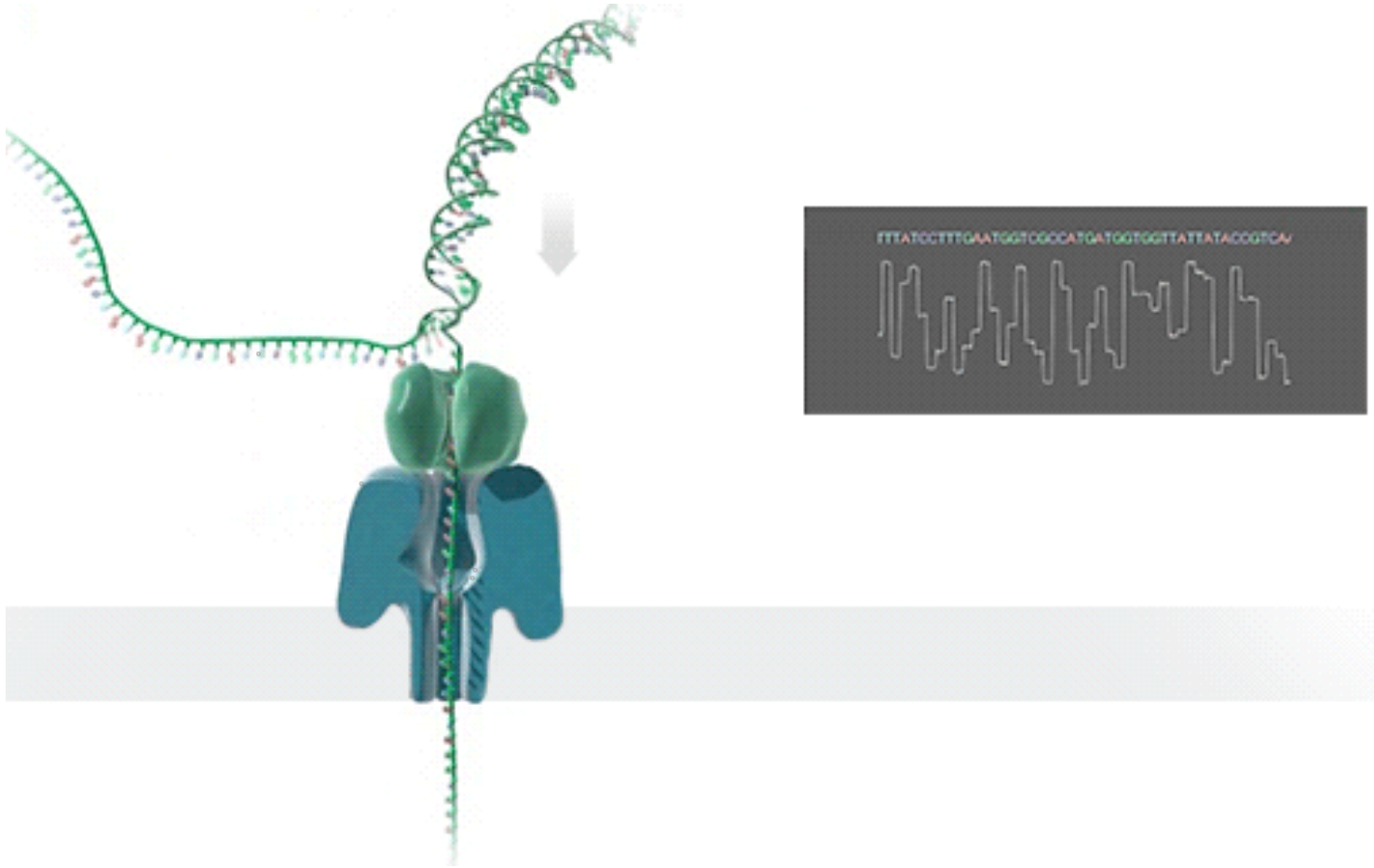
Oxford Nanopore

E. coli channel protein embedded in membrane nanopore

Double-stranded DNA is unwound and fed through a channel

Change in voltage across membrane measured by flow of ions through channel

The extent to which **ssDNA blocks the flow of ions** is the output signal



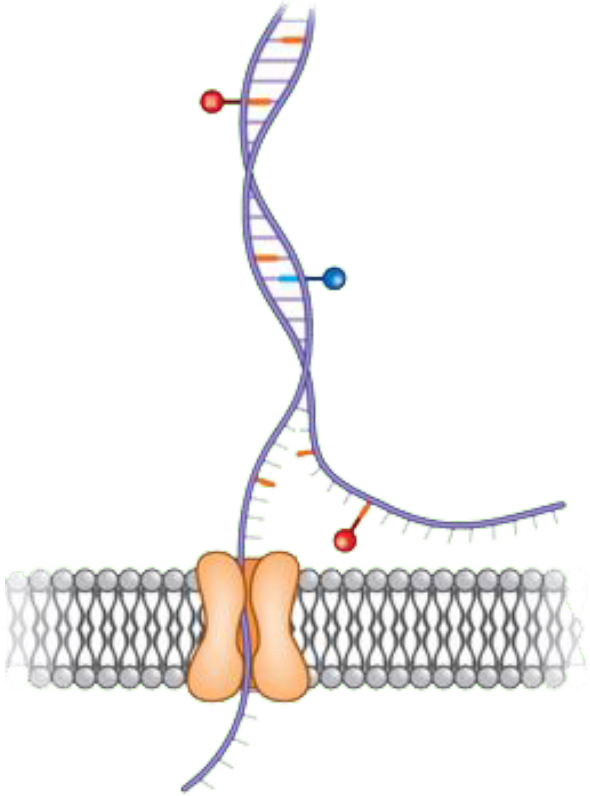
Oxford Nanopore

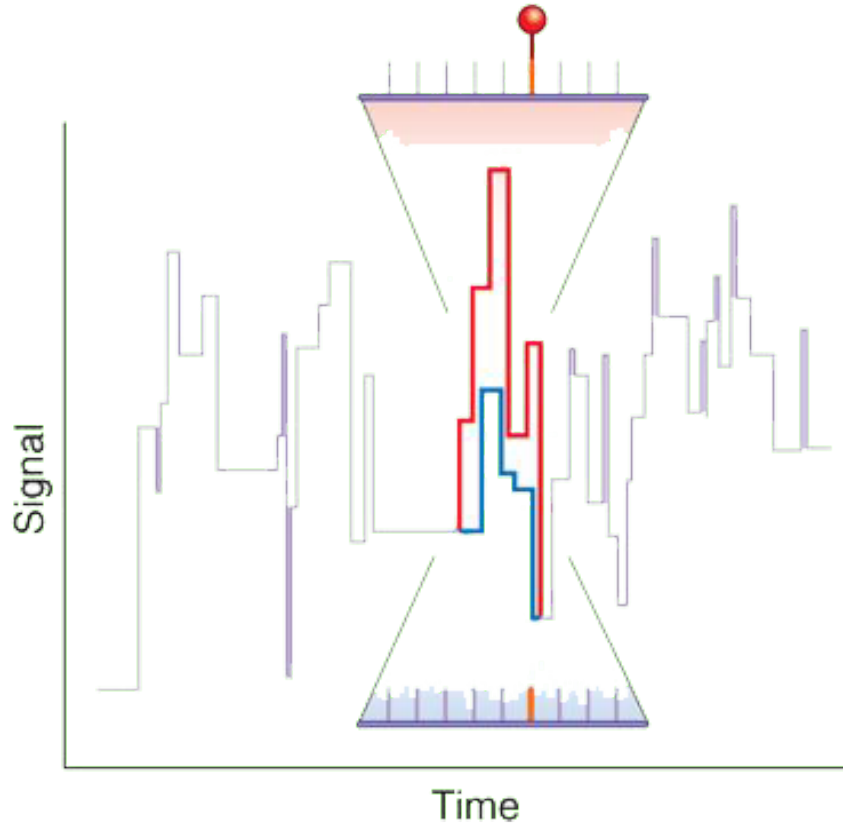
E. coli channel protein embedded in membrane nanopore

Double-stranded DNA is unwound and fed through a channel

Change in voltage across membrane measured by flow of ions through channel

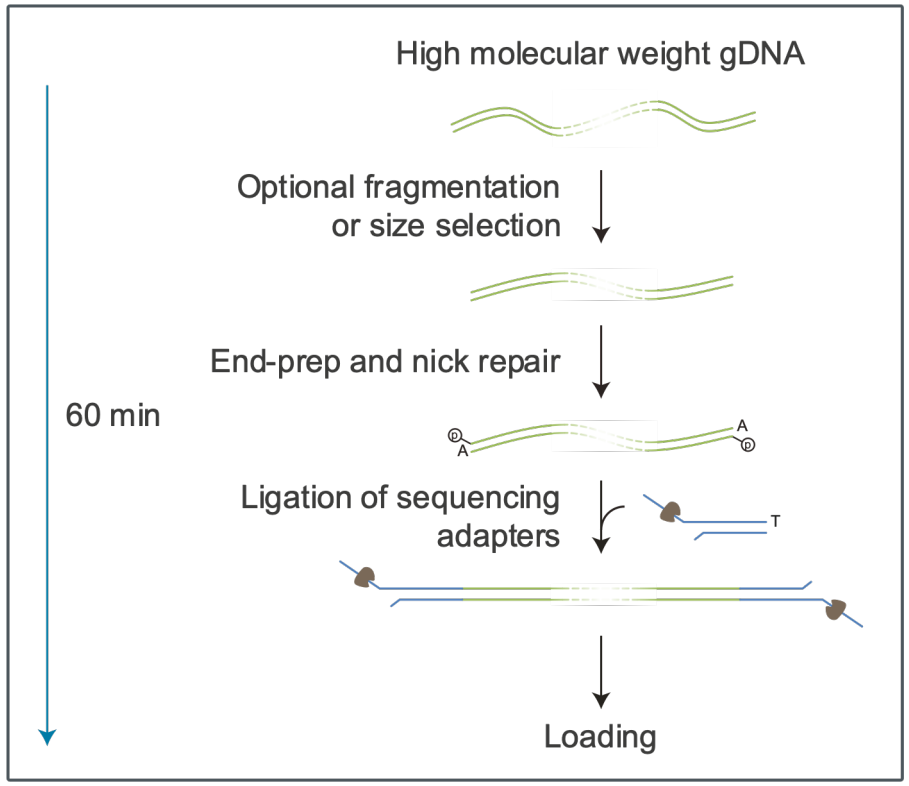
The extent to which **ssDNA blocks the flow of ions** is the output signal



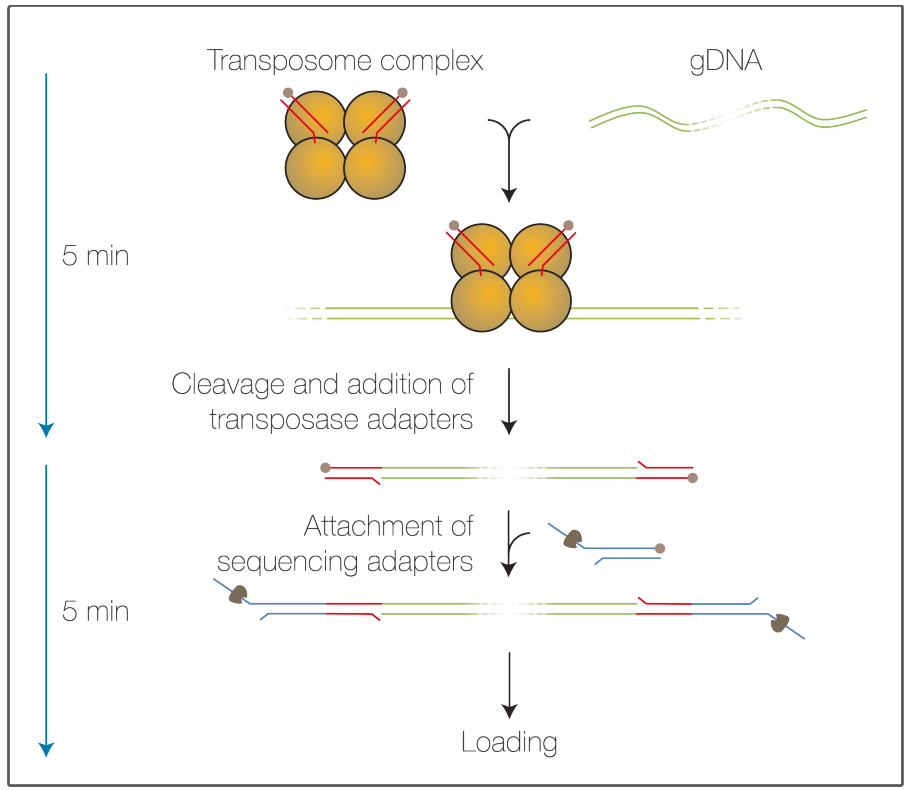


Oxford Nanopore Library Preps

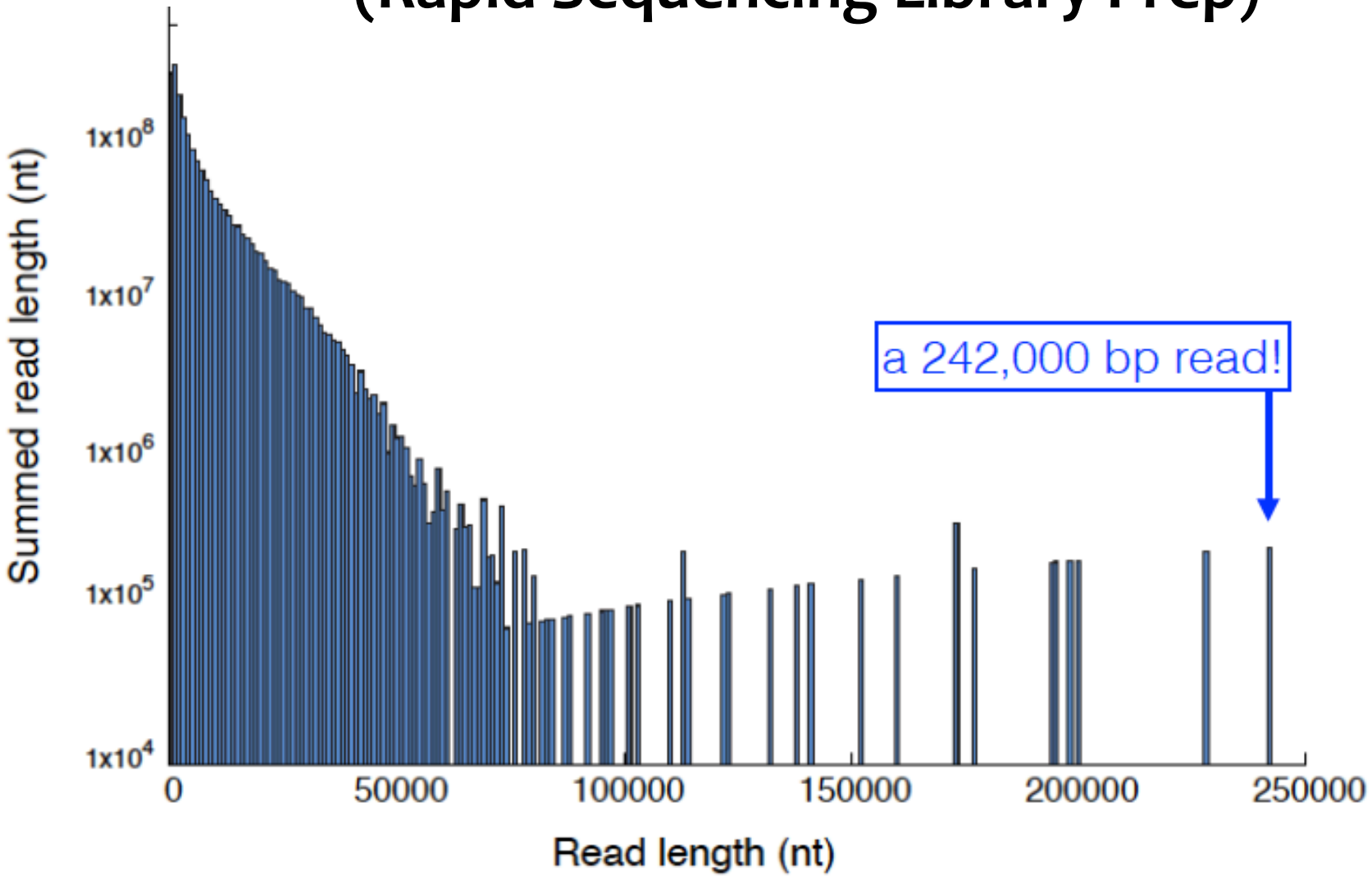
Ligation Prep
(longer reads, more prep time)



Rapid/Field Prep
shorter reads, less prep time



Nanopore read length distribution
(Rapid Sequencing Library Prep)



Oxford Nanopore Sequencing Platforms

SmidgION

-Will fill up your phone in seconds

Flongle

-Low-throughput (126 channels)
-Cheap
-Long queue

MinION

-Mid-throughput
30 Gb per flow cell
7-12 million reads
~$1000 starter kit
~$900 per flow cell after

GridION

-Mid/high-throughput
5 x Flow Cells

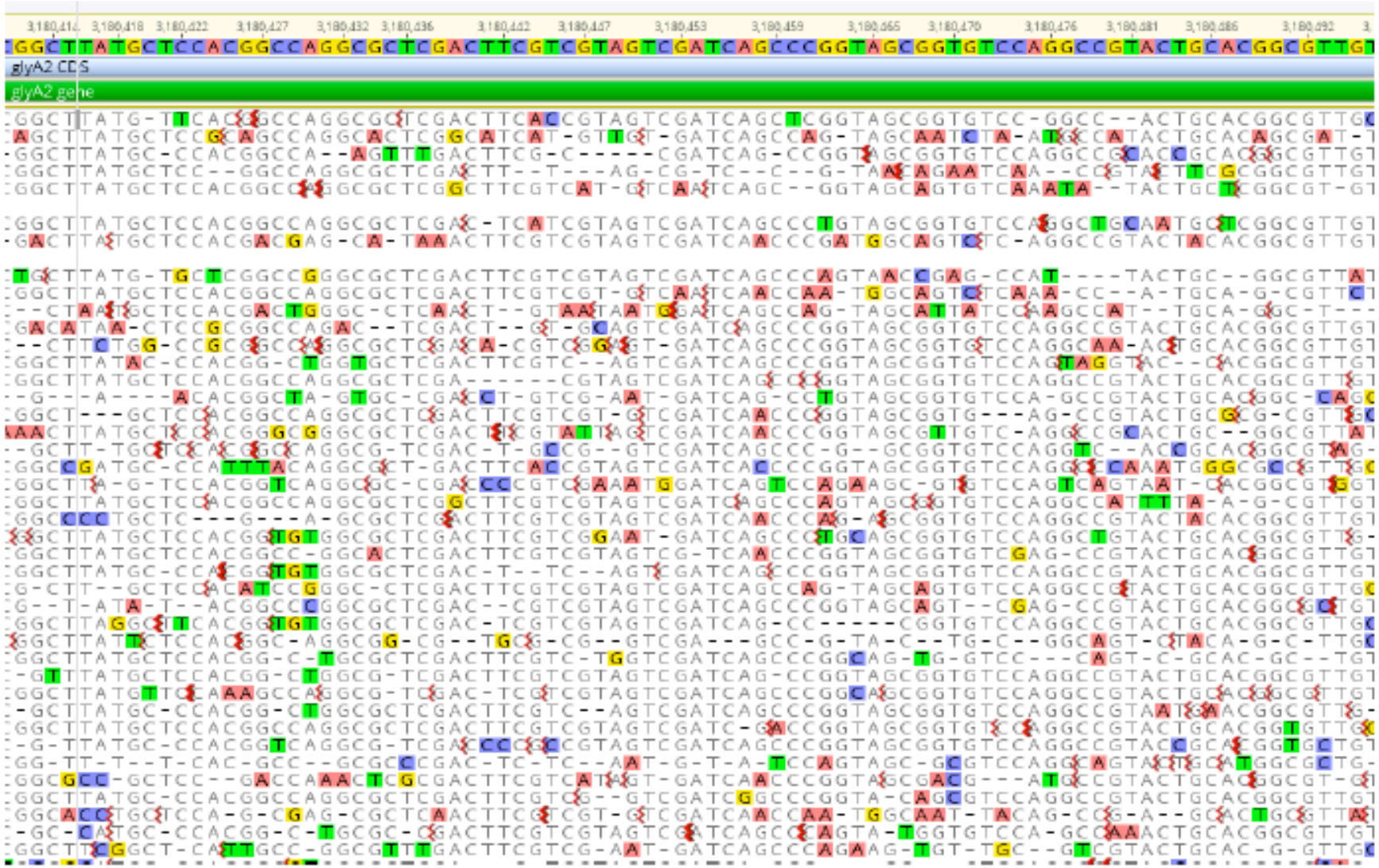
PromethION

-High-throughput
24/48 x Flow Cells

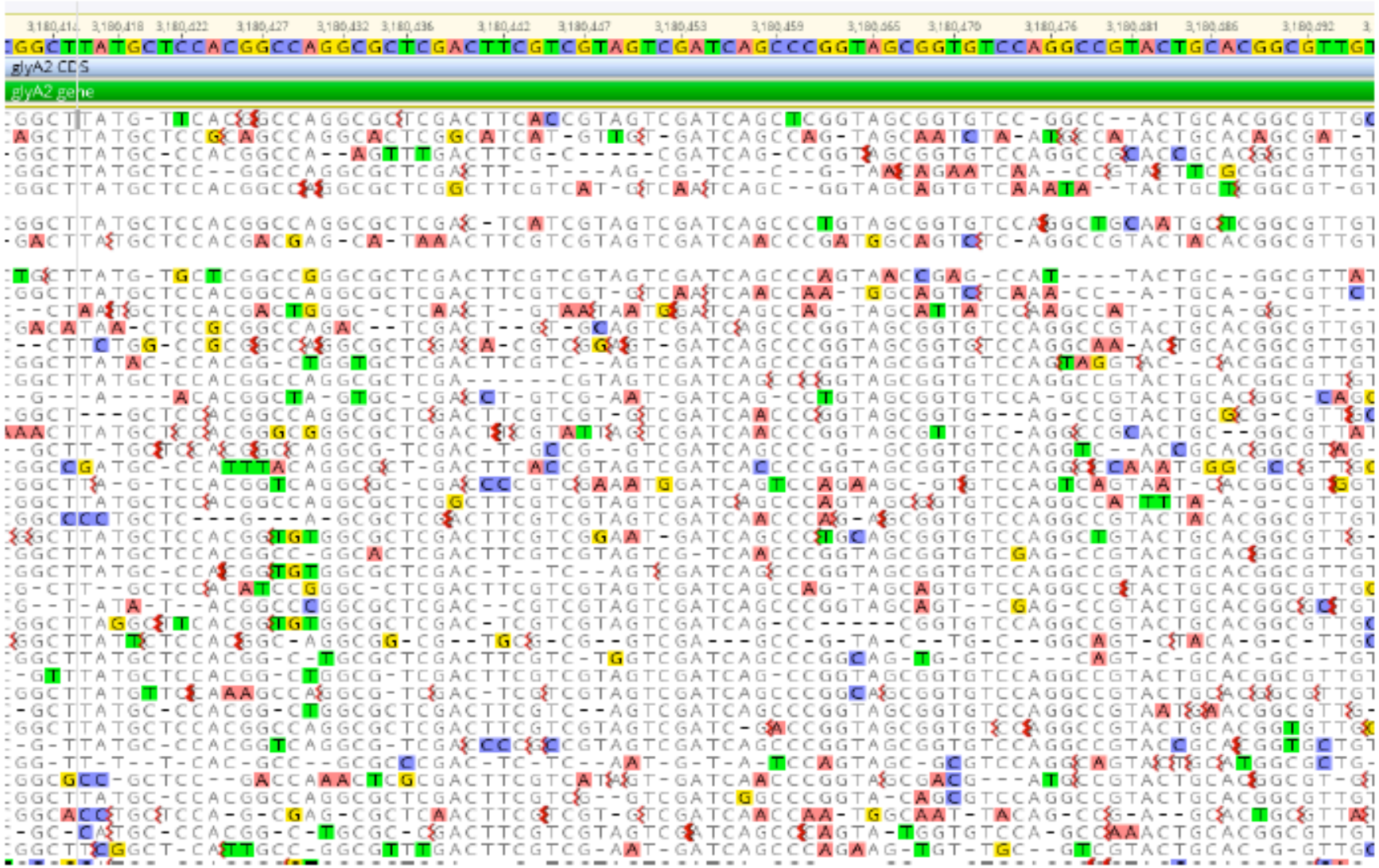
Intro to long-read sequencing

- **When is long-read sequencing the right/wrong choice?**

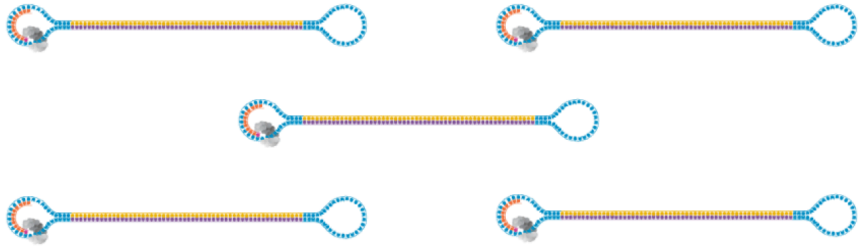
Long reads have high error rates

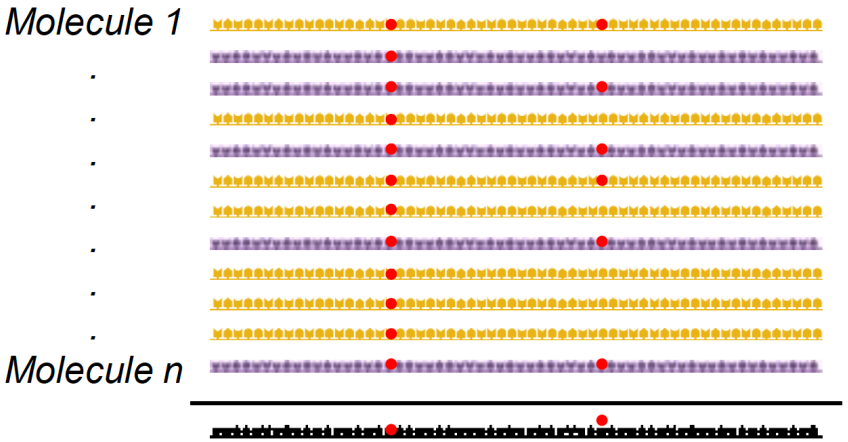


But you can use the consensus sequence for assembly

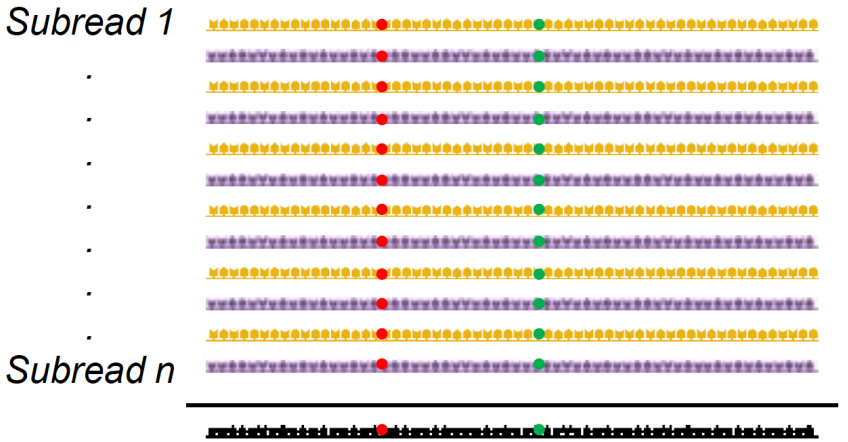


PacBio sequencing strategies





Long inserts, few CCS reads
De novo assembly



Short inserts, many CCS reads
Isoform Sequencing (**Iso-Seq**)

Which technology would you use?

- Quantifying gene expression among different isoforms in a non-model species
- Linkage analysis between SNPs that are on average 10kb apart
- Assemble a plant mitochondrial genome

Use long reads

- When linkage is more important than nucleotide identity
- Identify structural variants
- Resolve complex DNA structures
- Sequence though repeats
- Identify distinct splice variants
- Assembling a reference genome

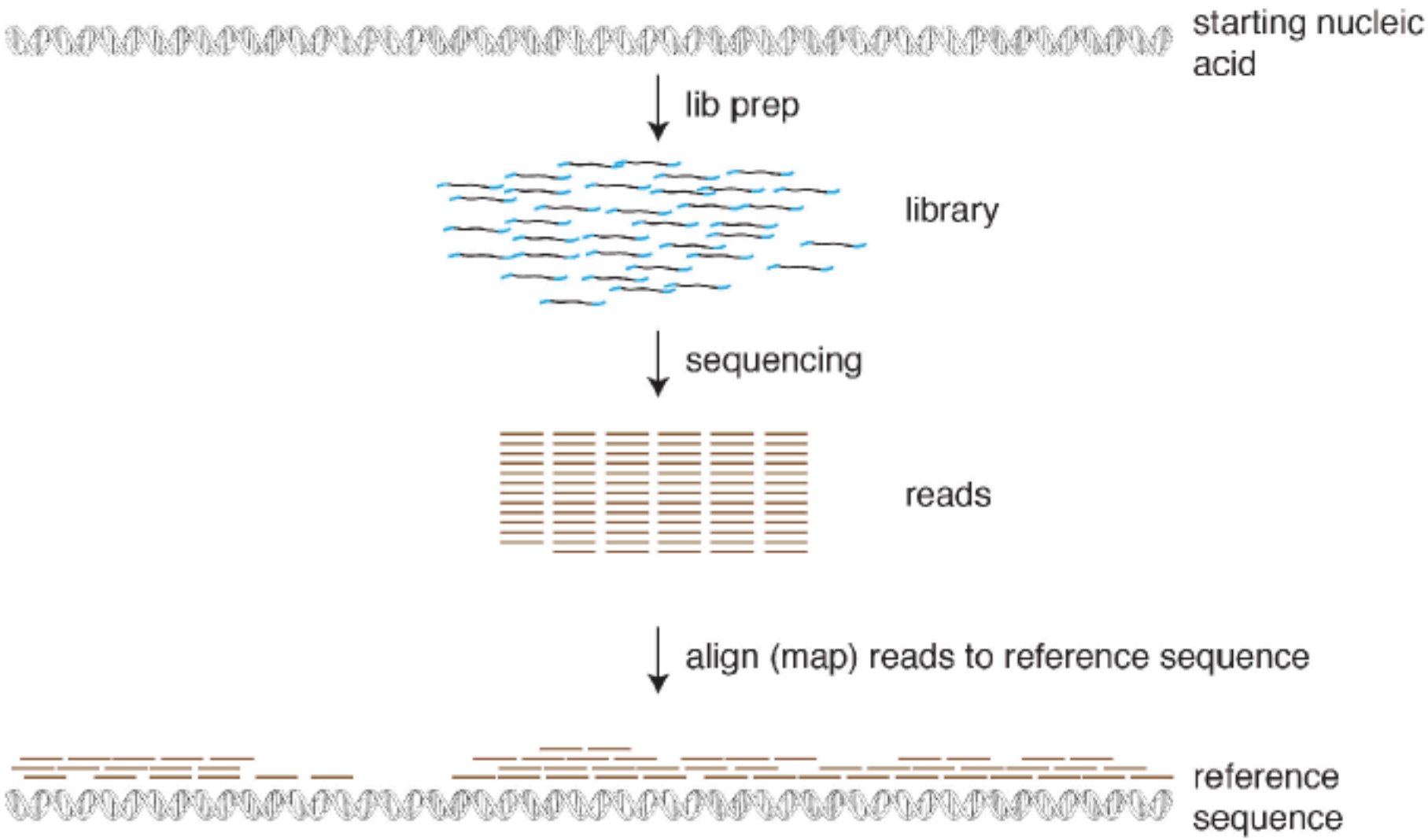
Don’t use long reads

- When nucleotide identity is more important than linkage
- Identify low-frequency SNVs
- Quantify gene expression
- Re-sequencing in populations (for now)

Intro to long-read sequencing

- **Genomic and transcriptomic applications**

Long reads can **map reads uniquely** in a reference



Using long-read overlaps to perform **de novo assembly**

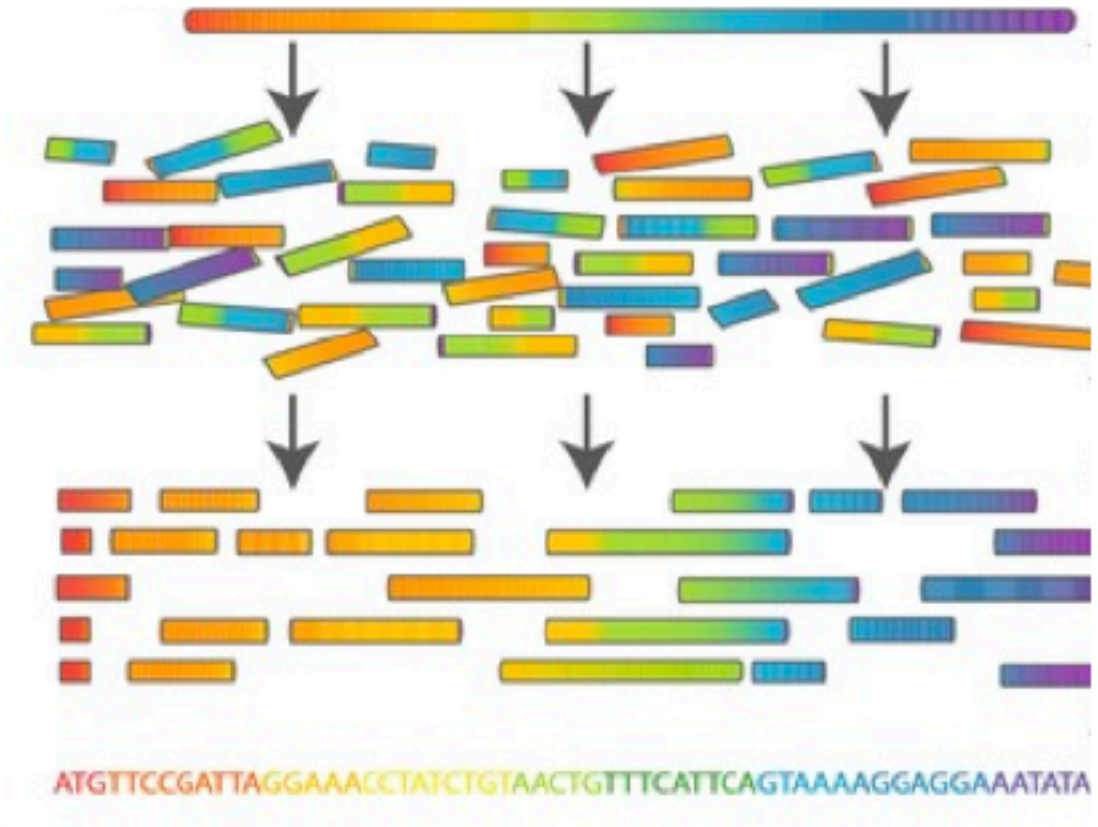
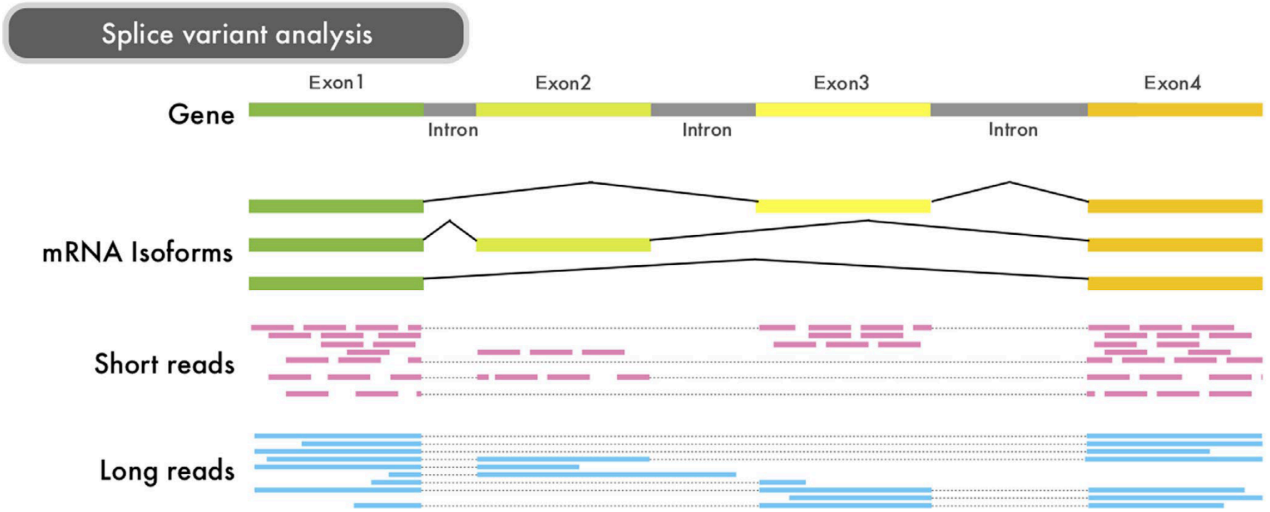
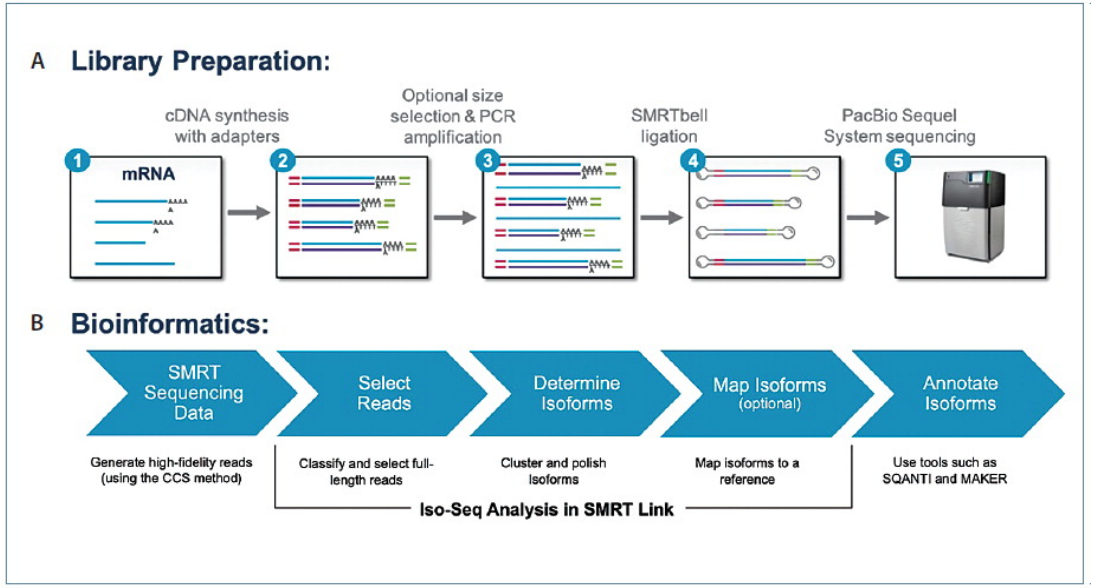


Image: Commins, J., Toft, C., Fares, M. A (2009)

Isoform profiling with long reads removes the assembly step



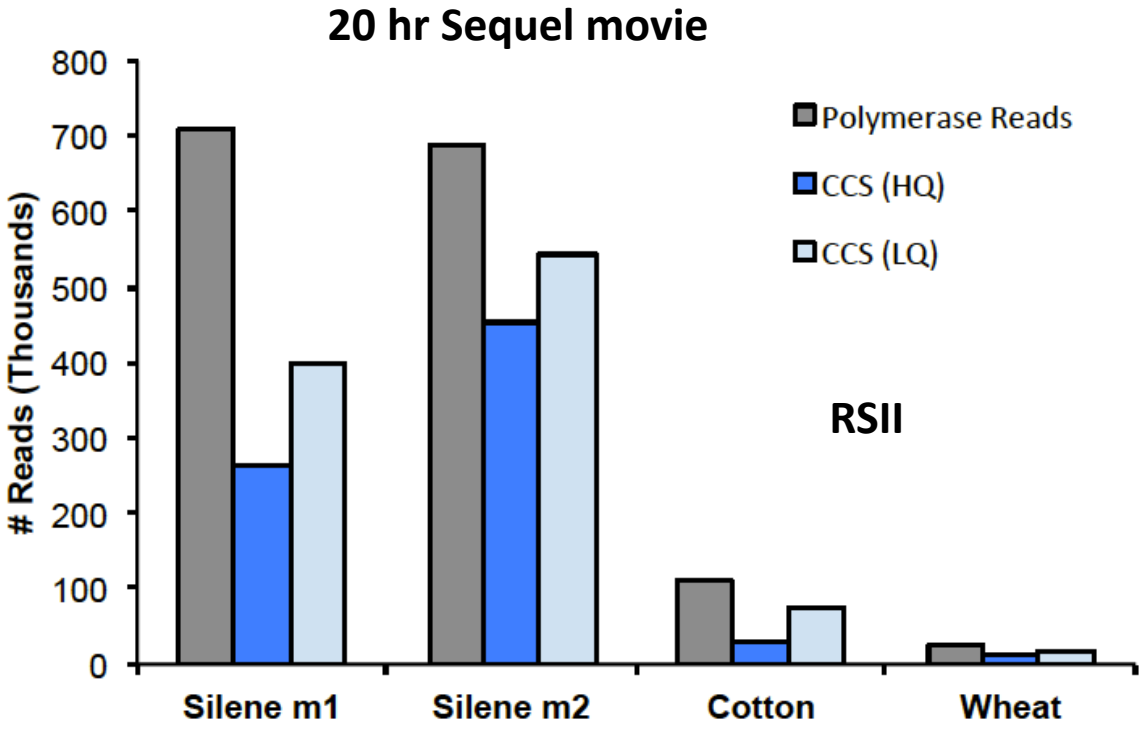
Nanopore RNA sequencing



PacBio IsoSeq

PacBio Iso-Seq Transcriptomics

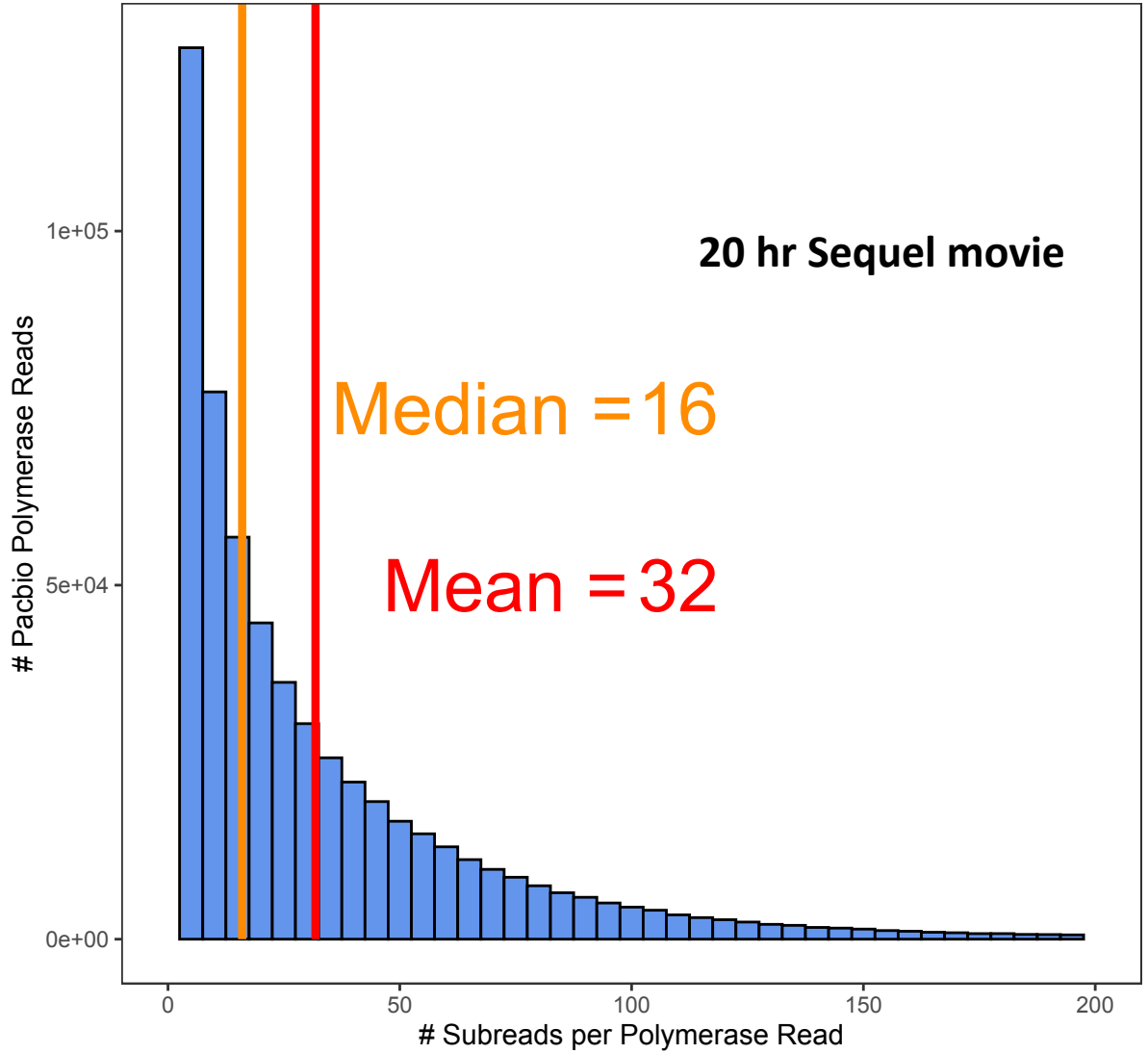
Polymerase reading the (subread + adapter) 3x = 1 CCS read



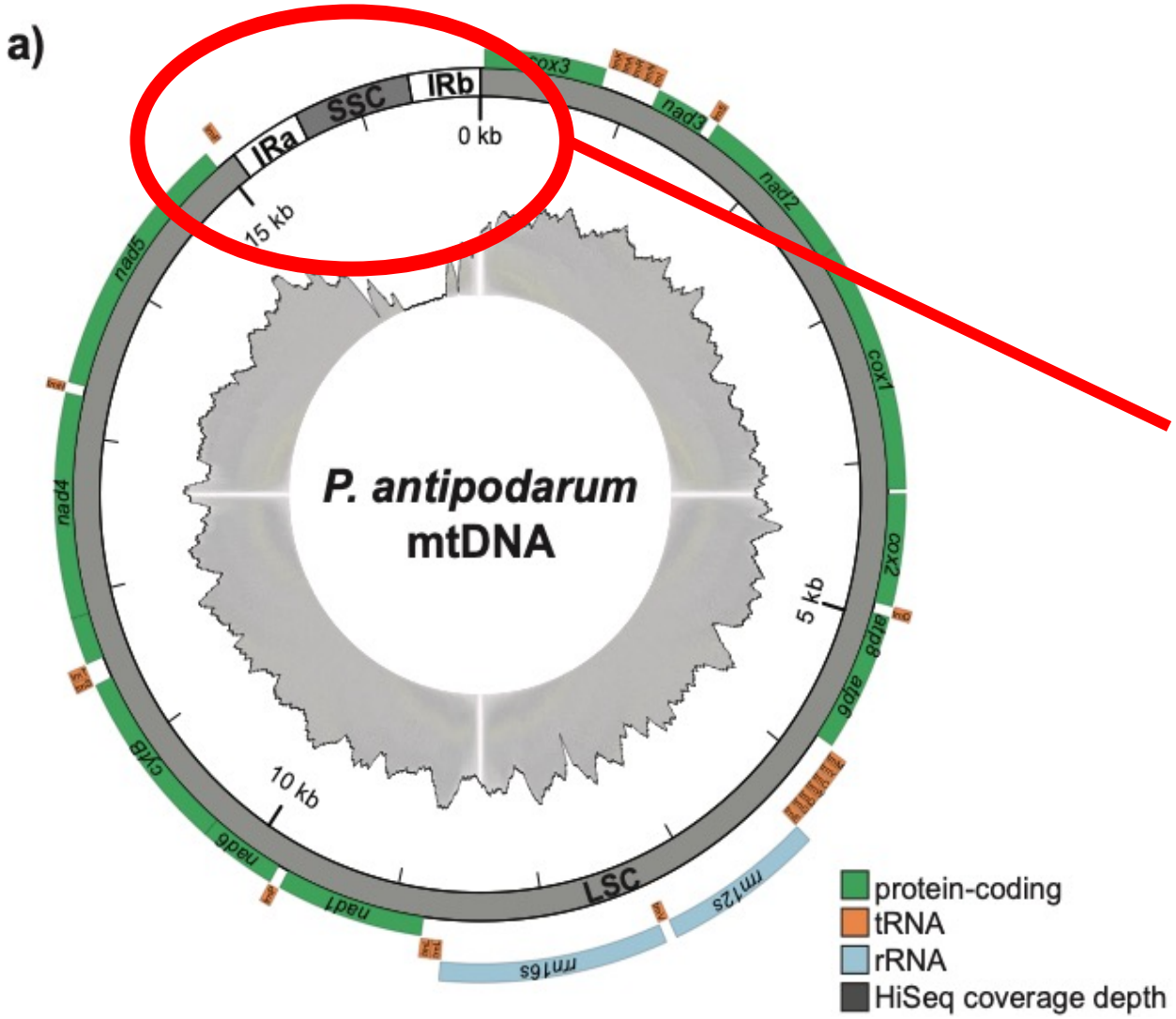
PacBio Iso-Seq Transcriptomics

Base modifications impede polymerase processivity in a predictable manner

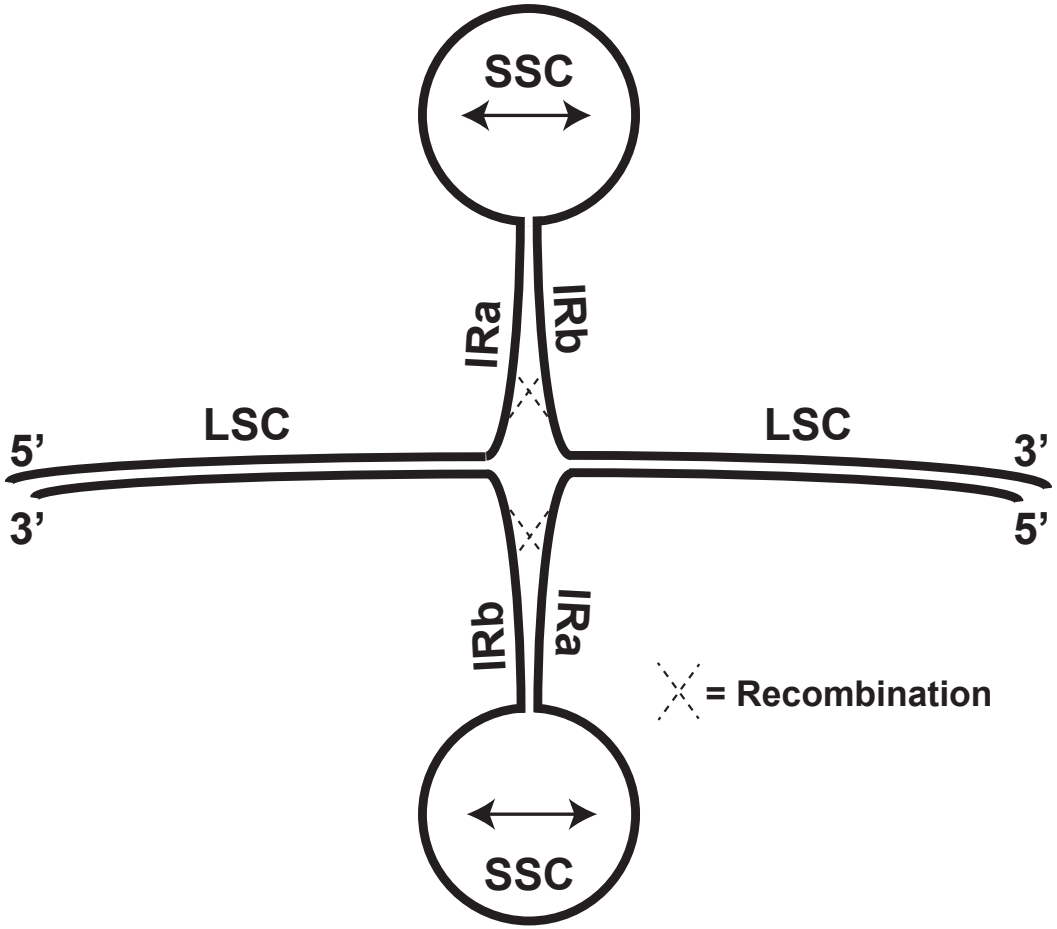
Can be measured with **Inter-pulse Distance (IPD)**



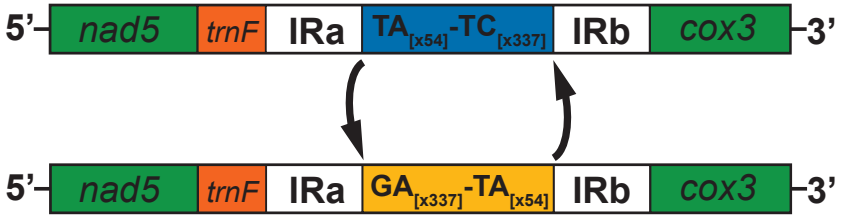
Resolving complex genomic features



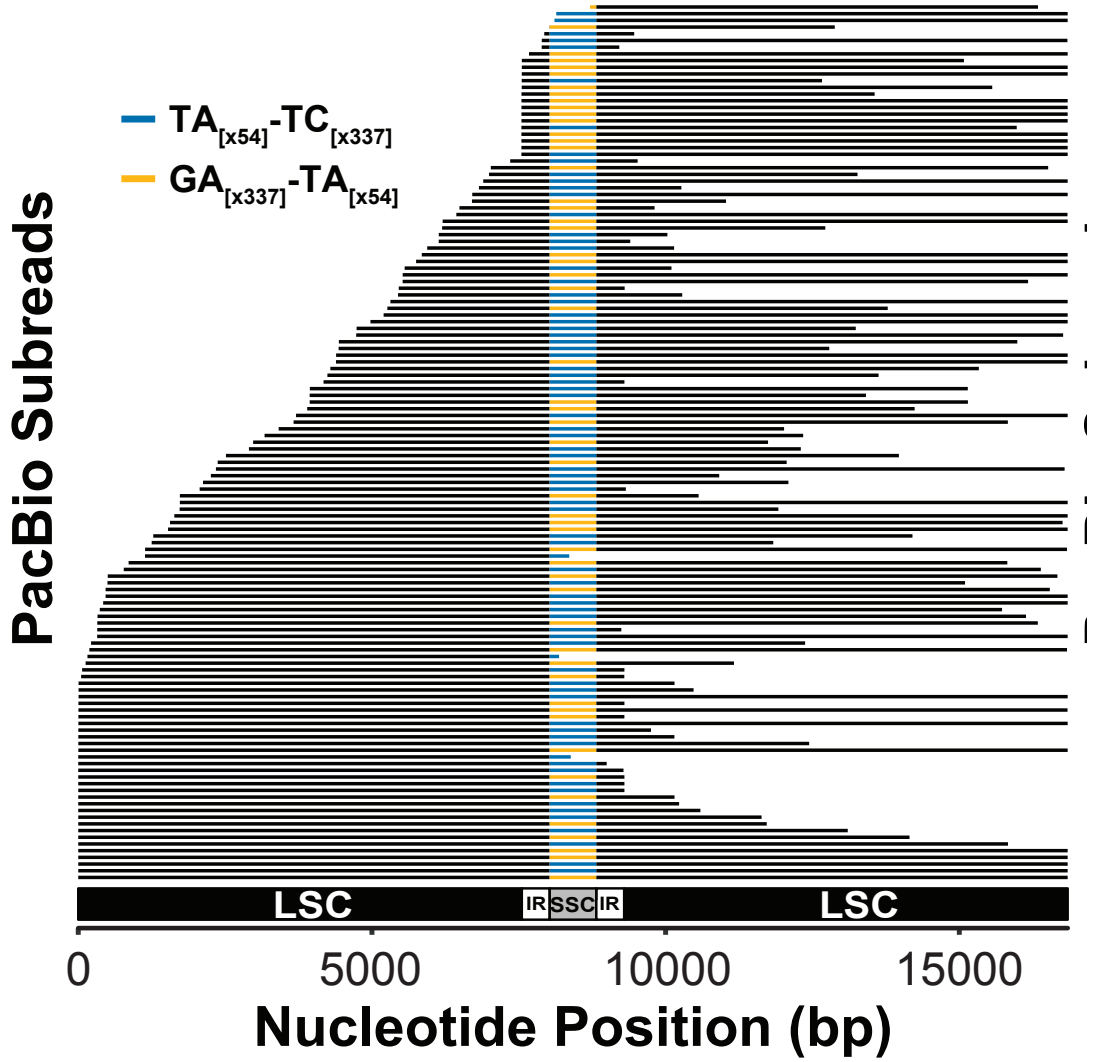
Flip-flop recombination



Resolving complex genomic features



Long reads can identify structural variants



Intro to long-read sequencing

- How do long-read sequencing technologies work?
- When is long-read sequencing the right/wrong choice?
- Genomic and transcriptomic applications