Long-read Sequencing (PacBio)

PacBio SMRT Technology

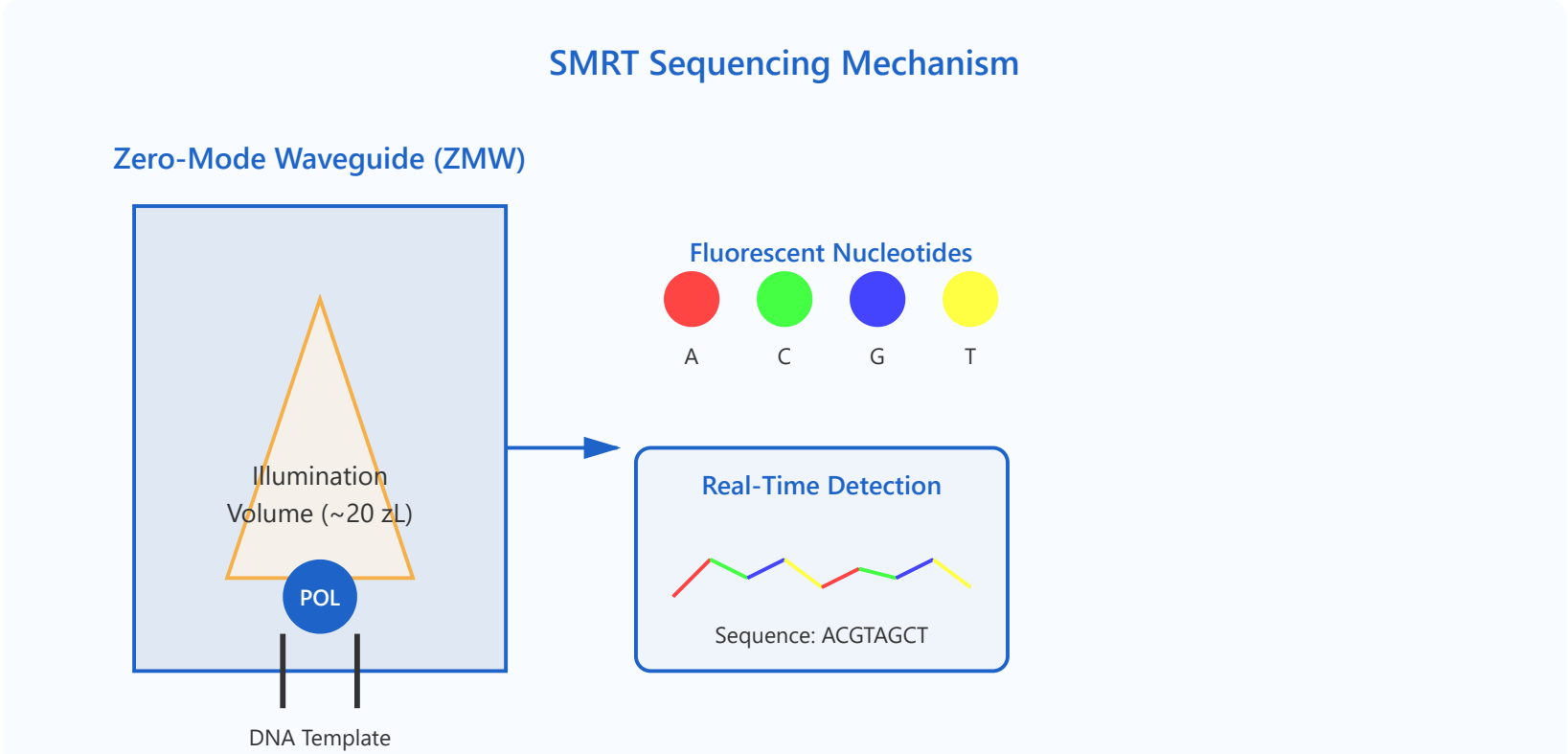
- Single Molecule Real-Time (SMRT) sequencing
- Watches DNA polymerase in real-time
- Zero-mode waveguides (ZMWs) for detection

Read Length	Accuracy	Throughput
10-30 Kb	**99.9% (HiFi)**	**~30 Gb/run**

Advantages

- Sequence through repetitive regions
- Detect structural variants and complex rearrangements
- Better genome assembly - fewer gaps
- Native base modification detection (methylation)

1. Single Molecule Real-Time (SMRT) Sequencing



How SMRT Sequencing Works: The technology uses zero-mode waveguides (ZMWs), which are tiny wells with a diameter of approximately 70 nanometers at the bottom of the well. Light is directed through the bottom, creating an illumination volume of only about 20 zeptoliters (10^{-21} liters). This extremely small detection volume allows observation of single DNA polymerase molecules at work.

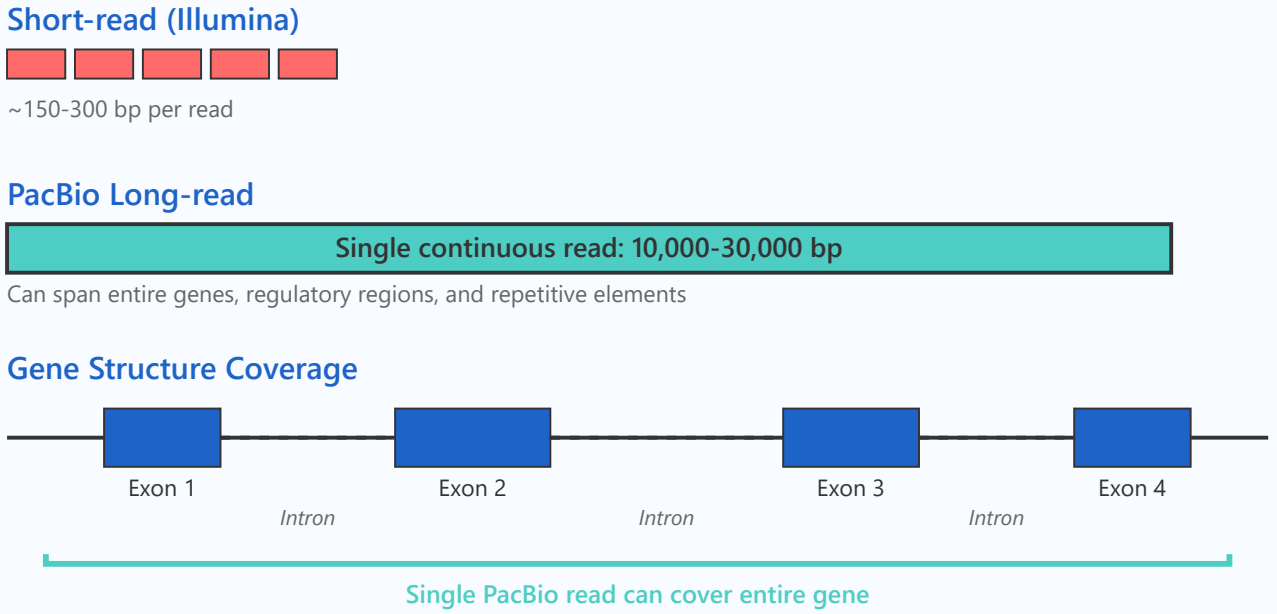
Each of the four DNA bases (A, C, G, T) is attached to a different fluorescent dye. When the polymerase incorporates a nucleotide into the growing DNA strand, the fluorescent label emits a light pulse that is detected in real-time. The dye is then cleaved off, allowing the next nucleotide to be incorporated without interference.

Example Application:

During sequencing of a bacterial genome, the SMRT system can continuously monitor a single DNA polymerase for several hours, generating reads of 10-30 kilobases without interruption. The real-time nature means that the sequencing speed is only limited by the natural rate of DNA polymerase (approximately 10 nucleotides per second).

2. Long Read Length (10-30 Kb)

Read Length Comparison



Advantages of Long Reads: The extended read length of PacBio sequencing (10-30 kb compared to 150-300 bp for Illumina) provides several critical advantages. Long reads can span entire genes including all exons and introns, making it possible to determine full-length transcript isoforms without computational assembly.

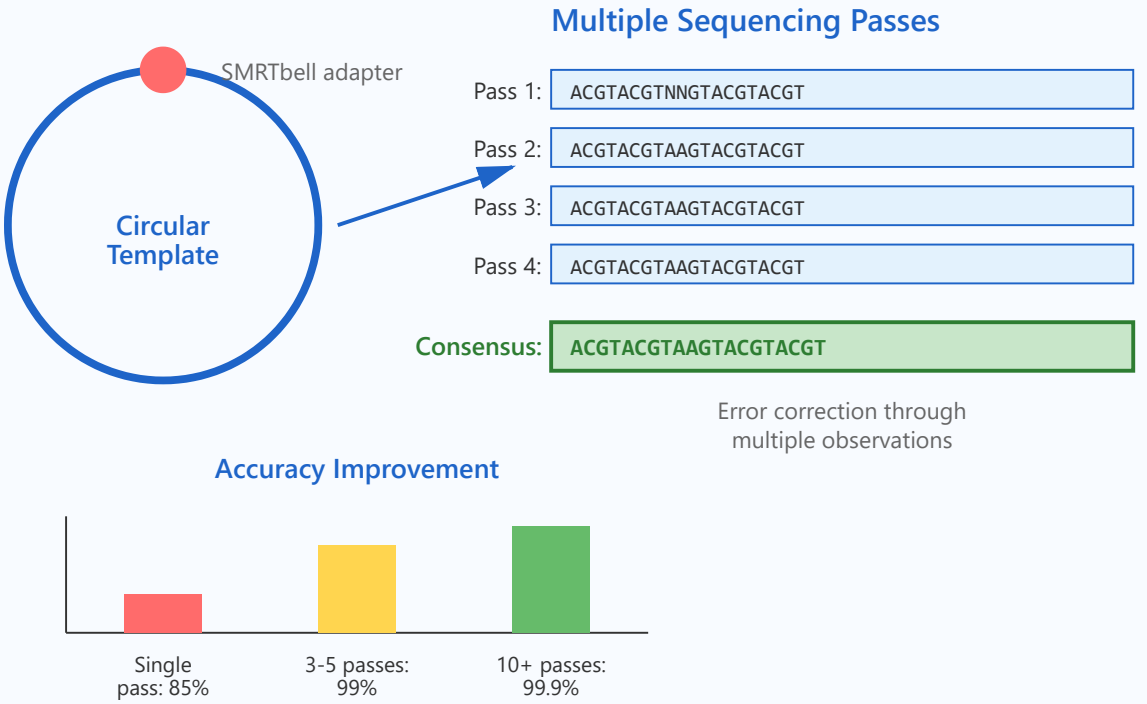
Long reads are particularly valuable for resolving complex genomic regions such as segmental duplications, tandem repeats, and other repetitive sequences that are typically fragmented or misassembled with short-read technologies. This capability significantly improves genome assembly contiguity and reduces the number of gaps in assembled genomes.

Example Application:

The human dystrophin gene (DMD) spans approximately 2.4 million base pairs with 79 exons. Short-read sequencing would require computational assembly of thousands of reads to reconstruct this gene, with potential for errors in repetitive regions. A single PacBio HiFi read can span multiple exons continuously, enabling direct detection of splicing patterns, structural variants, and disease-causing mutations without assembly artifacts.

3. High Accuracy (99.9% HiFi)

Circular Consensus Sequencing (CCS) for HiFi Reads



HiFi Technology: PacBio HiFi (High Fidelity) reads combine the advantages of long read lengths with high accuracy through Circular Consensus Sequencing (CCS). The DNA insert is ligated to hairpin adapters creating a circular template called a SMRTbell. The polymerase continuously sequences this circular template multiple times, generating multiple passes over the same DNA molecule.

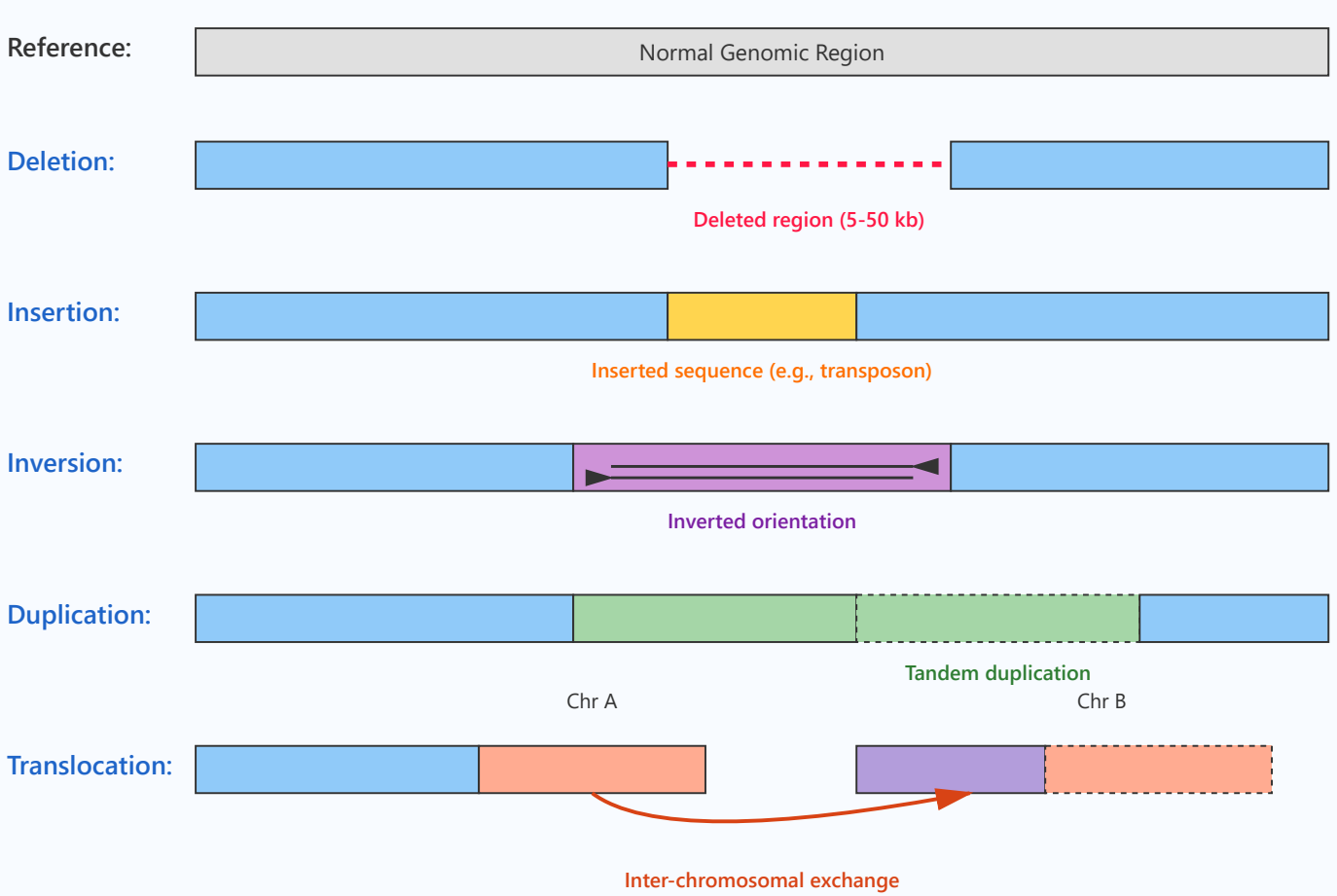
Each pass may contain random errors, but by sequencing the same molecule 10-20 times, these errors can be identified and corrected through consensus calling. The final HiFi read achieves greater than 99.9% (Q30) accuracy while maintaining read lengths of 10-25 kb, matching or exceeding the accuracy of short-read platforms while preserving long-range information.

Example Application:

In clinical diagnostics, HiFi sequencing enables accurate detection of single nucleotide variants (SNVs) and small insertions/deletions (indels) across complex genomic regions. For example, HiFi reads can accurately sequence through the highly polymorphic HLA genes (human leukocyte antigen) which are critical for transplant matching, resolving both allelic variations and structural differences with clinical-grade accuracy.

4. Structural Variant Detection

Types of Structural Variants Detected by Long Reads



Structural Variant Detection Capabilities: Long reads excel at detecting structural variants (SVs) that are difficult or impossible to identify with short reads. SVs include deletions, insertions, inversions, duplications, and translocations that affect segments typically larger than 50 base pairs. These variants play crucial roles in genome evolution, genetic disease, and cancer development.

PacBio reads can span entire structural variants, including their breakpoints, providing direct evidence of the variant structure. This is particularly valuable in repetitive regions where short reads cannot uniquely map.

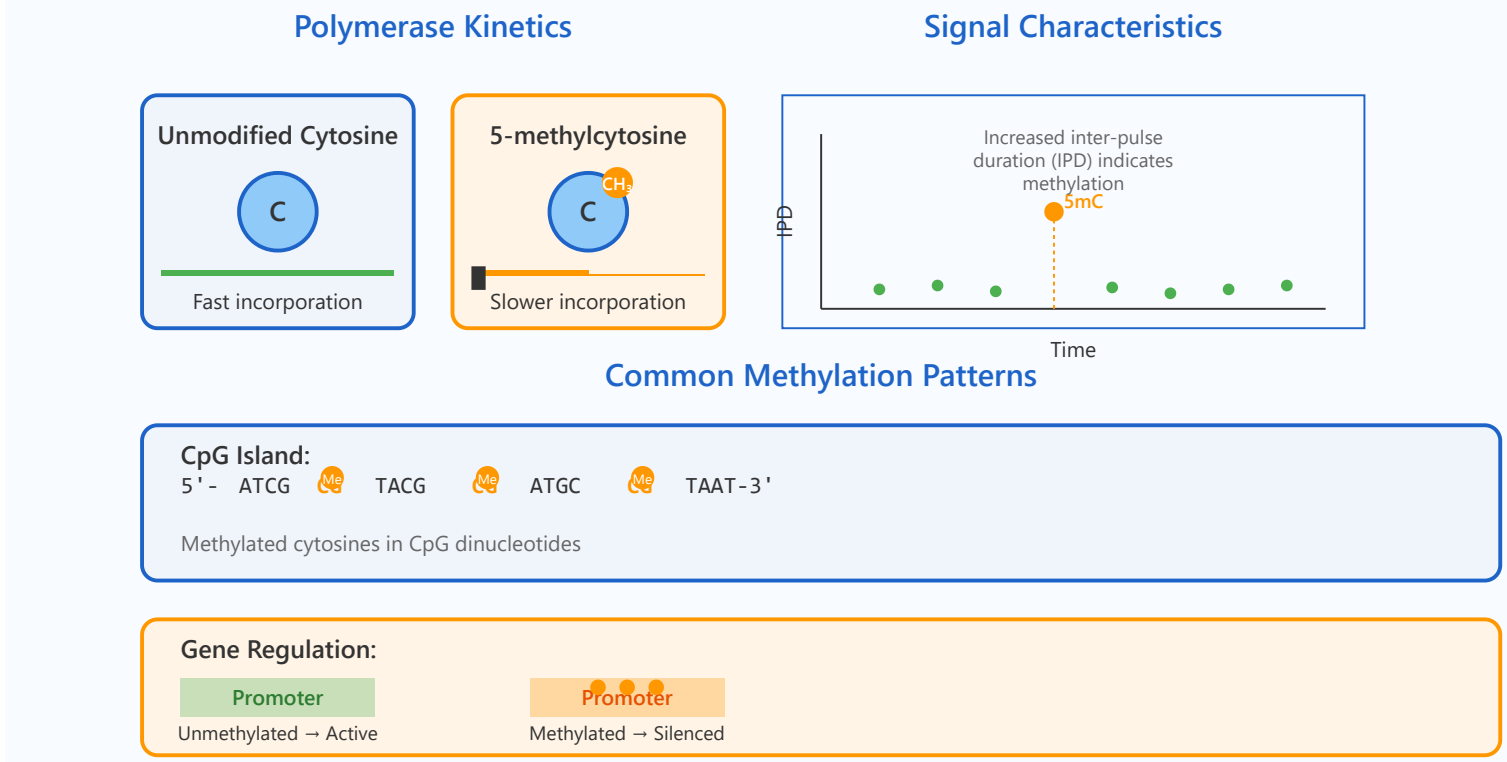
Long reads can also resolve complex rearrangements involving multiple events and determine their phase on individual chromosomes.

Example Application:

In cancer genomics, structural variants such as gene fusions drive oncogenesis. For example, the BCR-ABL1 fusion in chronic myeloid leukemia results from a translocation between chromosomes 9 and 22. PacBio long reads can span the fusion breakpoint, precisely determining the exact junction sequence and any additional complexity (e.g., insertions or deletions at the breakpoint) that may affect treatment response to tyrosine kinase inhibitors.

5. Native Base Modification Detection (Methylation)

Direct Detection of DNA Methylation



Direct Methylation Detection: Unlike other sequencing platforms that require chemical conversion (bisulfite sequencing) to detect methylation, PacBio directly detects DNA modifications during sequencing. Modified bases, such as 5-methylcytosine (5mC) and N6-methyladenine (6mA), affect the kinetics of DNA polymerase incorporation.

The SMRT system measures the inter-pulse duration (IPD), which is the time between successive nucleotide incorporations. When the polymerase encounters a modified base, it pauses slightly longer, resulting in an increased IPD. By analyzing these kinetic signatures across the genome, PacBio can simultaneously determine DNA sequence and map methylation patterns without additional sample preparation or loss of sequence information.

Example Application:

In cancer epigenetics research, aberrant DNA methylation patterns are hallmarks of tumorigenesis. PacBio sequencing can identify hypermethylation of tumor suppressor gene promoters (such as BRCA1 or MLH1) that leads to gene silencing without requiring separate methylation assays. This integrated approach enables researchers to correlate structural variants, sequence mutations, and epigenetic modifications in a single experiment, providing a comprehensive view of cancer genome architecture.

Feature	Bisulfite Sequencing	PacBio Native Detection
Sample preparation	Chemical conversion required	No conversion needed
DNA degradation	Significant (~90% loss)	No degradation
Read length	Reduced due to treatment	Full long-read length maintained
Modification types	5mC only	5mC, 6mA, and other modifications
Sequence context	C/T ambiguity	Original sequence preserved
Phasing information	Lost in short reads	Maintained across long reads

Summary: PacBio SMRT Sequencing Advantages

PacBio's Single Molecule Real-Time (SMRT) sequencing technology represents a paradigm shift in genomics by providing long, accurate reads with native modification detection. The key advantages include:

Long-Range Information

Spanning 10-30 kb enables resolution of complex genomic regions, complete gene structures, and haplotype phasing

High Accuracy

HiFi reads achieve >99.9% accuracy through multiple sequencing passes, suitable for clinical applications

Structural Variant Detection

Direct observation of large insertions, deletions, inversions, and complex rearrangements

Epigenetic Profiling

Simultaneous detection of DNA sequence and modifications without additional library preparation

PacBio technology is particularly valuable for de novo genome assembly, characterization of complex genetic diseases, cancer genomics, and microbiome research where comprehensive genomic information is essential.