

GEP
GENOMICS EDUCATION
PARTNERSHIP

Long-Read Sequencing Technology

Katie M. Sandlin, M.S.

✉ kmsandlin1@gmail.com
 🐦 @KatieMSandlin
 👤 [linkedin.com/in/kmsandlin](https://www.linkedin.com/in/kmsandlin)

Last Update: December 29, 2025

1

Currently Available GEP Resources

thegep.org/sequencing/

2

Note to Instructors

- I have scripted the talk in the notes section of the PPT.
- Advanced versions of this presentation and lecture recordings were developed by Wilson Leung (Washington University in St. Louis).
 - Overview Third-generation Sequencing
 - Lecture Recording
 - Overview PacBio Iso-Seq
 - Lecture Recording

3

Agenda (1)

1. Review of RNA-Seq Short-read Sequencing
2. Overview and benefits of Long-read Sequencing
3. Oxford Nanopore Sequencing
4. Pacific Biosciences Sequencing

4

RNA Sequencing (RNA-Seq)

- RNA sequencing (or RNA-seq) lets us investigate and discover the transcriptome—the total cellular content of RNAs including mRNA, rRNA and tRNA—by using next-generation sequencing to examine the quantity and sequences of RNA in a sample.
 - RNA-seq analyzes the transcriptome, indicating which of the genes encoded in our DNA are turned on or off and to what extent.

The **transcriptome** is the complete set of transcripts in a cell, and their quantity, for a specific developmental stage or physiological condition.

Wang, Z., Gerstein, M., & Snyder, M. (2009). <https://doi.org/10.1016/j.cell.2009.12.011>

5

RNA-Seq: reads are typically 30–300 bp

① Isolate RNA from samples

Created with BioRender.com

6

Transcriptome reconstruction—akin to reassembling magazine articles after they have been through a paper shredder

Ward, J. 2013. Genomes: the code of life. In DNA-seq analysis. Nature Methods.

7

Short vs. Long Read RNA-Seq

Adapted from: Genetic Engineering & Biotechnology News.

8

Agenda (2)

1. Review of RNA-Seq Short-read Sequencing
2. Overview and benefits of Long-read Sequencing
3. Oxford Nanopore Sequencing
4. Pacific Biosciences Sequencing

Pacific Biosciences

9

Short vs. Long Read Sequencing

Sequencing instruments are categorized as long-read or short-read based on their underlying chemistry and the length of DNA fragments they analyze.

PacBio HiFi Sequencing - Unlock Your Next Great Discovery.

10

Benefits of assembling a genome with long-read sequencing

- A way to think about the difference between assembling a genome with short-reads versus long-read sequencing technology would be to imagine two different approaches to reconstructing a 368-page novel from randomized snippets of text.

Pacific Biosciences

11

Benefits of assembling a genome with long-read sequencing: an example using books (Short read sequencing = fragmented statements)

Those looking for serious land

the North Carolina coast

The shack sat back

hatchet and pack a buck

by a torn shoreline

Excerpts from: Where the Crawdads Sing

12

Benefits of assembling a genome with long-read sequencing: an example using books (Long read sequencing = entire paragraphs)

The marsh was quagmire by a twin shoal, labeled by early explorers as the "Greenway of the Atlantic" because rippling, furious winds, and shallow shoals wrecked ships like paper bats along what would become the North Carolina coast. One woman's journal read, "rang'd along the shore . . . but could discern no Entrance . . . A violent Storm overtook us . . . we were forced to get off to Sea, to secure Churches and Ship, and were driven by the Rapidity of a strong Current . . ."

Where the Crawdads Sing

The BRACK SAT BACK from the palmettos, which sprawled across sand flats to a necklace of green lagoons and, in the distance, all the marsh beyond. Miles of blade grass so tough it grew in salt water interrupted only by trees as bare they were the shape of the dead. Oak forests branched around the other sides of the shack and sheltered the closest lagoons. Its surface so rich in life it churned. Salt air and gull-sung drifted through the trees from the sea.

"The Land . . . being marshy and Swamps, we return'd towards our Ship . . . Discouragement of all such as should hereafter come into those Parts to settle."

Claiming territory hadn't changed much since the 1500s. The scattered marsh holdings weren't legally described, just staked out natural—a creek boundary here, a dead oak there—by swagpods. A man doesn't set up a palmetto lawn in a bog unless he's on the run from somebody or at the end of his own road.

Those looking for arroyo land moved on, and this infamous marsh became a nest, scooping up a mishmash of mutinous sailors, contraband, debtors, and fugitives dodging wars, taxes, or laws that they didn't take to. The ones malaria didn't kill or the swamp didn't swallow bred into a woodsmen tribe of several races and multiple cultures, each of whom could fill a small town with a hatchet and pack a buck for miles. Like

Excerpts from *Where the Crawdads Sing* 13

13

Benefits of assembling a genome with long-read sequencing: an example using books

Where the Crawdads Sing

The BRACK SAT BACK from the palmettos, which sprawled across sand flats to a necklace of green lagoons and, in the distance, all the marsh beyond. Miles of blade grass so tough it grew in salt water interrupted only by trees as bare they were the shape of the dead. Oak forests branched around the other sides of the shack and sheltered the closest lagoons. Its surface so rich in life it churned. Salt air and gull-sung drifted through the trees from the sea.

Claiming territory hadn't changed much since the 1500s. The scattered marsh holdings weren't legally described, just staked out natural—a creek boundary here, a dead oak there—by swagpods. A man doesn't set up a palmetto lawn in a bog unless he's on the run from somebody or at the end of his own road.

The marsh was quagmire by a twin shoal, labeled by early explorers as the "Greenway of the Atlantic" because rippling, furious winds, and shallow shoals wrecked ships like paper bats along what would become the North Carolina coast. One woman's journal read, "rang'd along the shore . . . but could discern no Entrance . . . A violent Storm overtook us . . . we were forced to get off to Sea, to secure Churches and Ship . . . Discouragement of all such as should hereafter come into those Parts to settle."

Those looking for arroyo land moved on, and this infamous marsh became a nest, scooping up a mishmash of mutinous sailors, contraband, debtors, and fugitives dodging wars, taxes, or laws that they didn't take to. The ones malaria didn't kill or the swamp didn't swallow bred into a woodsmen tribe of several races and multiple cultures, each of whom could fill a small town with a hatchet and pack a buck for miles. Like

- More straightforward to construct a genome assembly using longer reads

Excerpts from *Where the Crawdads Sing* 14

14

3rd Generation (Long-Read) Sequencing Emerges

- To address the limitations of next-generation sequencing, scientists began developing real-time, single-molecule DNA sequencing platforms.
- In the mid 2000s, two single-molecule sequencing technologies emerged giving rise to 3rd generation sequencing:
 - single-molecule real-time (SMRT) sequencing by Pacific Biosciences
 - nanopore sequencing by Oxford Nanopore Technologies

15 Pacific Biosciences

15

Third-generation Sequencing Technologies

- Sequencing platforms from Pacific Biosciences (PacBio) and Oxford Nanopore Technologies (ONT)
- Key characteristics:
 - **Single-molecule, real-time** sequencing
 - Sequence samples directly without amplification
 - Can detect epigenetic modifications
 - DNA methylation
- Comparisons with Illumina:
 - Produce **longer reads**
 - Raw reads have **higher error rates** than Illumina

16

16

Agenda (3)

1. Review of RNA-Seq Short-read Sequencing
2. Overview and benefits of Long-read Sequencing
3. Oxford Nanopore Sequencing
4. Pacific Biosciences Sequencing

17

17

ONT Sequencing – Template Topology

- Yellow – forward DNA strand
- Dark Blue – reverse DNA strand
- Light Blue – sequencing adapters

18

18

ONT Sequencing – Flow Cell

- Yellow – forward DNA strand
- Dark Blue – reverse DNA strand
- Light Blue – sequencing adapters

Flow cell (top view)

Nanopore

Synthetic membrane

Logsdon, G.A., Volgger, M.R. & Eichler, E.E. [Long-read human genome sequencing and its applications](#). *Nat Rev Genet* 21, 597–614 (2020).

19

19

ONT Sequencing - Nanopore

- Yellow – forward DNA strand
- Dark Blue – reverse DNA strand
- Light Blue – sequencing adapters

Single nanopore (cross section)

Motor protein

Nanopore

Voltage-biased membrane

Logsdon, G.A., Volgger, M.R. & Eichler, E.E. [Long-read human genome sequencing and its applications](#). *Nat Rev Genet* 21, 597–614 (2020).

20

20

ONT Sequencing – Base Calling

- You can think of sequencing DNA or RNA like reading music — the bases being the notes, which when played in the correct order, produces a recognizable song.

Readout

Mean signal (pA)

Time (seconds)

Logsdon, G.A., Volgger, M.R. & Eichler, E.E. [Long-read human genome sequencing and its applications](#). *Nat Rev Genet* 21, 597–614 (2020).

21

21

Check For Understanding (A)

Raise your fingers (1-4) to share your answer.

1. Synthetic Membrane
2. Motor Protein
3. DNA or RNA
4. Nanopore

Cold Spring Harbor Laboratory – DNA Learning Center.

22

22

Check For Understanding (B)

Raise your fingers (1-4) to share your answer.

1. Synthetic Membrane
2. Motor Protein
3. DNA or RNA
4. Nanopore

Cold Spring Harbor Laboratory – DNA Learning Center.

23

23

Check For Understanding (C)

Raise your fingers (1-4) to share your answer.

1. Synthetic Membrane
2. Motor Protein
3. DNA or RNA
4. Nanopore

Cold Spring Harbor Laboratory – DNA Learning Center.

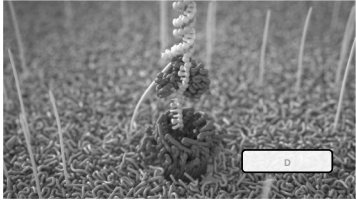
24

24

Check For Understanding (D)

Raise your fingers (1-4) to share your answer.

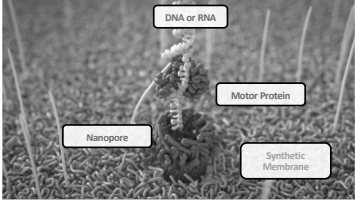
1. Synthetic Membrane
2. Motor Protein
3. DNA or RNA
4. Nanopore



Cold Spring Harbor Laboratory – DNA Learning Center 25

25

Check For Understanding



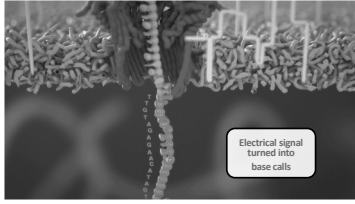
- Read length is up to 4 Mb
 - Long reads sequencing mode: 10–100kb
 - Ultra-long sequencing mode: 100–300kb

Cold Spring Harbor Laboratory – DNA Learning Center 26

26

Review – Base Calling

- This digital method of reading DNA or RNA sequence has multiple benefits:
 - Real-time analysis
 - PCR free, no amplification bias
 - **Modified base detection**
 - Read-length agnostic
 - **Direct sequencing of DNA or RNA**




Disclaimer: The real-time base caller has lower accuracy. For many applications, it might be beneficial to do the base calling afterwards with a more powerful GPU and a more compute intensive algorithm [i.e., duplex with super accuracy (SUP)].


Cold Spring Harbor Laboratory – DNA Learning Center 27

27

Flow Cell Where Nanopores are Housed



The **MinION** Flow Cell can generate up to 50 Gb of data for sequencing DNA, cDNA or native RNA in real-time.

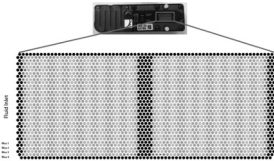


The **PromethION** Flow Cell can generate up to 290 Gb for sequencing DNA, cDNA or native RNA in real-time.

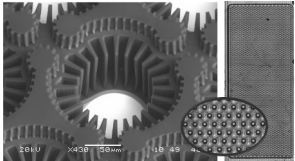
Oxford Nanopore Technologies 28

28

MinION Flow Cell



MinION Mk1B and Mk1C/GridION Flow Cell Channel Layout




SEM image of a single sensor well (left) and the hexagonal array of wells (right).

Oxford Nanopore Technologies 29


29

Nanopore Sequencing Devices




MinION
The only portable, real-time devices for DNA and RNA sequencing, giving complete control and creativity over when, where and how often you sequence, regardless of application.

[View MinION >](#)




GridION
Compact benchtop device designed to run and analyze up to five MinION Flow Cells

[View GridION >](#)



PromethION
High coverage nanopore sequencing in formats ranging from modular, fully-integrated devices, to high-throughput solutions. Each flow cell can deliver the lowest price per Gb for nanopore sequencing.

[View PromethION >](#)



MinION Mk1D:
Up to 15–35 Gb of data from a single flow cell

Oxford Nanopore Technologies 30

30

ONT Sequencing Workflow

A complete and streamlined workflow—real-time answers to your biological questions

Elysion:
Fully automated, sample-to-answer device

GEP

31

Nanopore for Educators eBook

- Community guide to integrate Nanopore DNA sequencing into the classroom
 - Quick Start Guide
 - Time and Cost Planning
 - Annotated Lab Protocols
 - Bioinformatics Support

GEP

32

Agenda (4)

1. Review of RNA-Seq Short-read Sequencing
2. Overview and benefits of Long-read Sequencing
3. Oxford Nanopore Sequencing
4. Pacific Biosciences Sequencing

GEP

33

PacBio SMRT Sequencing – SMRTbell Template

GEP

34

Key Components of SMRT Sequencing

8M SMRT Cell

Zero-Mode Waveguides (ZMWs)

Each SMRT Cell can contain up to 8 million ZMWs, all of which are liquid-filled chambers.

GEP

35

Key Components of SMRT Sequencing (Phospholinked Nucleotides)

GEP

36

Phospholinked Nucleotide Incorporation

1 DNA molecule and 1 polymerase in each well (ZMW)
4 colors flash in real-time as polymerase acts

Mantis 01: [http://nanoscale.computingplatform.com/Amir_Rev_Anal_Chem_2013_6237_303](#) 37

37

2 Modes of SMRT Sequencing (CCS)

Circular Consensus Sequencing (CCS) Mode
Inserts 10-20 kb

Single-molecule consensus sequence

Galaxy Training - Quality Control 38

38

2 Modes of SMRT Sequencing (CLR)

Circular Consensus Sequencing (CCS) Mode
Inserts 10-20 kb

Single-molecule consensus sequence

Continuous Long Read (CLR) Sequencing Mode
Inserts >25 kb, up to 175 kb

Multi-molecule consensus sequence

Galaxy Training - Quality Control 39

39

How are HiFi reads generated?

- SMRT sequencing can sequence the same DNA molecule **multiple times** generating highly accurate long reads, or **HiFi reads**.
- HiFi reads are produced using **circular consensus sequencing (CCS)** mode on PacBio long-read systems. HiFi reads provide base-level resolution with 99.9% single-molecule read accuracy.

CCS Docs 40

40

HiFi Reads vs. Long Reads

CCS Docs 41

41

HiFi Reads vs. Long Reads (Disclaimers)

15-20 kb (HiFi)

PacBio HiFi reads typically have lower quality than Illumina data. They also have systematic errors (particularly with homopolymers and specific types of repetitive sequences; see [McCartney et al., 2022](#)).

PacBio HiFi reads are defined as reads with an accuracy of Q20 or above (error rate of 1/100; 99% accuracy). For the Revio system, the mode for the read quality is ~Q30 (99.9% accuracy; see the [Revio system brochure](#)).

For the Illumina XLEAP-SBS chemistry, the read quality is Q40 (99.99% accuracy; see the [XLEAP-SBS chemistry enables Q40 and above data quality on NovaSeq X and NextSeq 1000/2000](#) blog post).

CCS Docs 42

42

PacBio Sequencing – How it Works

PacBio Sequencing – How it Works

GEP

43

Overview of PacBio Sequencers

Images provided by Pacific Biosciences

GEP

44

PacBio Revio Sequencing System

\$779,000

Wall, Cornell Medicine

GEP

45

PacBio Vega Benchtop System

\$169,000
\$1,100 per run

PacBio Vega Benchtop System

GEP

46

PacBio SMRT Sequencing Workflow

Although each sequencing project is unique, there are three main steps to go from DNA to discovery with SMRT sequencing.

Library prep

[Learn more about kits for fast and easy library preparation](#)

SMRT sequencing

[Explore sequencing interactive](#)

Data analysis

[SMRT Link user interface and SMRT Analysis](#)

Pacific Biosciences – Steps of SMRT Sequencing

GEP

47

Isoform Sequencing (Iso-Seq)

- Using SMRT sequencing to do RNA-Sequencing = marketing term is Iso-Seq
- Iso-Seq data is obtained via Circular Consensus Sequencing (CCS)

Iso-Seq
Scalable De Novo Isoform Discovery from PacBio HiFi Reads

Genentech

GEP

48

Copy Google Doc Worksheet

	Oxford Nanopore Technologies (ONT)	Pacific Biosciences (PacBio)
Sequencing chemistry/ technology name		
Read length		
What is detected in the sequencing process?		
How is that detected?		
Enzyme used in the sequencing process		

thegep.org/lrs

Adapted from: Bowling, et al. [Using Next-Generation Sequencing into the Classroom through a Comparison of Molecular Biology Techniques](#). *The American Biology Teacher*, 74 (6), 396-401.

55

Answer Key

	Oxford Nanopore Technologies (ONT)	Pacific Biosciences (PacBio)
Sequencing chemistry/ technology name	Nanopore Sequencing	Single Molecule, Real-Time (SMRT)
Read length	Up to 4 Mb; Long reads sequencing mode: 10-100kb; Ultra-long sequencing mode: 100-300kb	15-20 kb (HiFi)
What is detected in the sequencing process?	Changes in current as nucleotides pass through the pore	Fluorophore associated with the phospholinked nucleotides emit different color light
How is that detected?	Different collection of nucleotides that occupy a nanopore will cause different changes in the electric current. Machine learning algorithms (neural networks) are used to analyze the continuous changes to the electric current (squiggle) to infer the sequences that pass through the nanopore.	Each phospholinked nucleotide (A, C, G, T) is labeled with a different fluorescent dye. When a nucleotide is incorporated by the polymerase, it emits different color light that corresponds to the fluorescent dye for the nucleotide.
Enzyme used in the sequencing process	Motor protein (helicase)	DNA polymerase

Adapted from: Bowling, et al. [Using Next-Generation Sequencing into the Classroom through a Comparison of Molecular Biology Techniques](#). *The American Biology Teacher*, 74 (6), 396-401.

56

Acknowledgements

- Special thanks to Wilson Leung (Washington University in St. Louis) for his suggestions and feedback.
- Thank you to Jason Williams and Anna Feitzinger at the CSHL DNA Learning Center for leading a Nanopore Sequencing Workshop.
- This presentation is based upon work supported by the National Science Foundation under Grant No. 1915544 and National Institute of General Medical Sciences of the National Institutes of Health under award number R25GM130517 to LKR.

57

Questions?

58