จีโนมิกส์ประเทศไทย: การแพทย์จีโนมิกส์เพื่อการยกระดับคุณภาพชีวิตคนไทย

Genomics Thailand: Enhancing our Quality of Life via Genomic Medicine

วันที่ 30 มีนาคม 2566 เวลา 09.30 – 16.00 น. อาคารสราญวิทย์ อุทยานวิทยาศาสตร์ประเทศไทย

NAC2023 18th NSTDA Annual Conference การประชุมวิชาการประจำปี สวทช. ครั้งที่ ๑๘

Update State of the Art Genomic Platforms

15.15 – 16.00 น.

นพ.วีรยุทธ ประพันธ์พจน์

บริษัท ศูนย์พันธุศาสตร์การแพทย์ จำกัด

อ.ดร.ธิดาทิพย์ วงศ์สุรวัฒน์

คณะแพทยศาสตร์ศิริราชพยาบาล มหาวิทยาลัยมหิดล

นายเฉลิมพล ศรีจอมทอง

ศูนย์ความเป็นเลิศทางการแพทย์ด้านเวชพันธุศาสตร์
โรงพยาบาลจุฬาลงกรณ์ สภากาชาดไทย

Q&A 10 นาที

Update State-of-the-Art Genomic Platforms: PacBio long-read sequencing

March 30, 2023

Chalurmpon Srichomthong, MSc

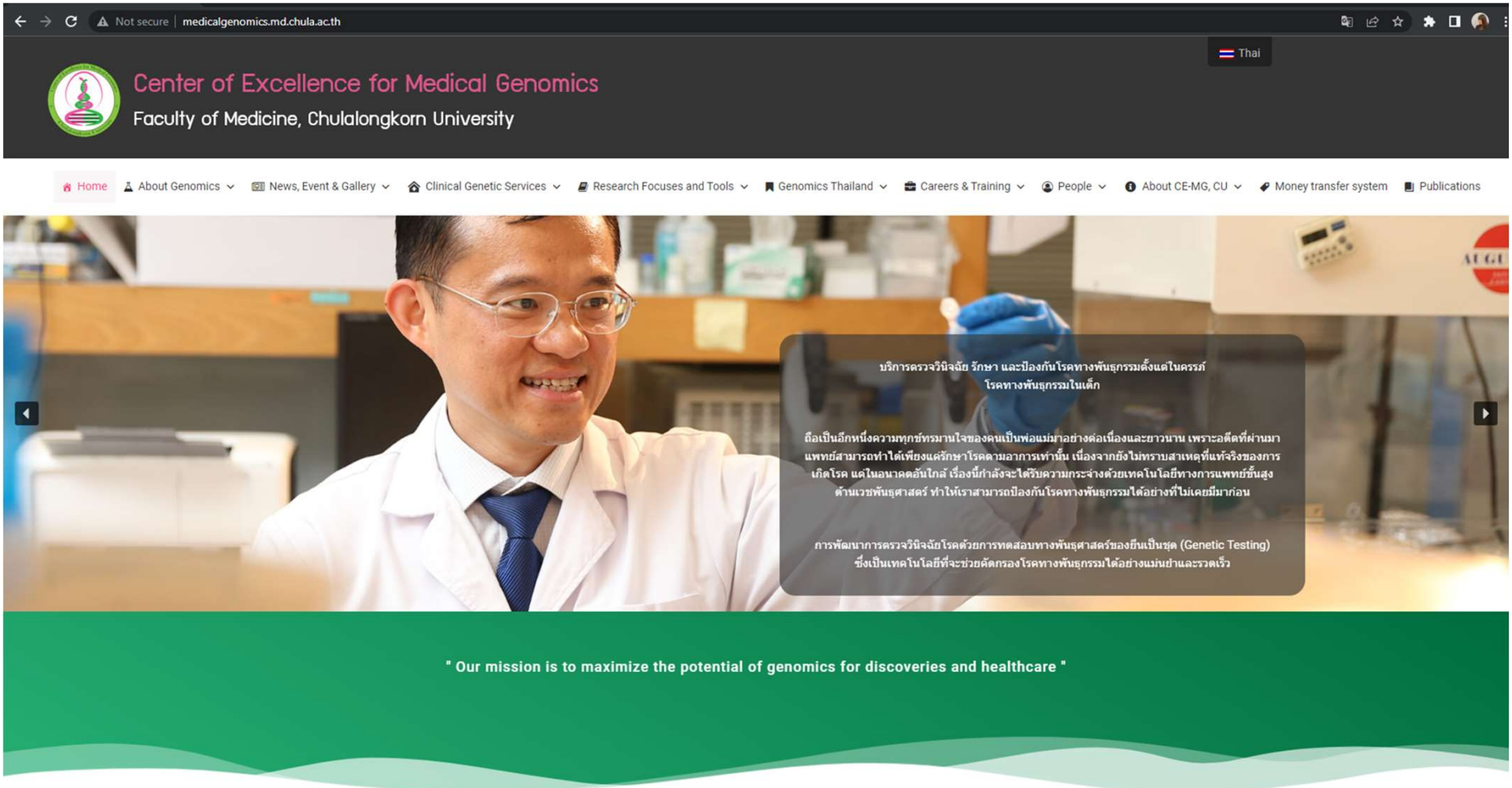
Excellence Center for Medical Genomics,
King Chulalongkorn Memorial Hospital, The Thai Red Cross Society

Center of Excellence for Medical Genomics, Department of Pediatrics
Faculty of Medicine, Chulalongkorn University

Sequencing Facility

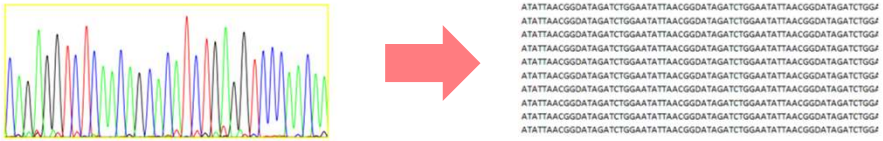
อาคารแพทยพัฒน์ ชั้น 8 ห้อง 808 คณะแพทยศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย

http://medicalgenomics.md.chula.ac.th/



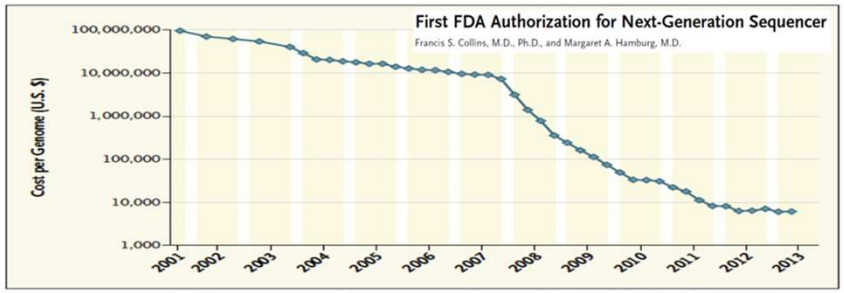
Mutation detection with NGS

DNA Sequencing Technologies

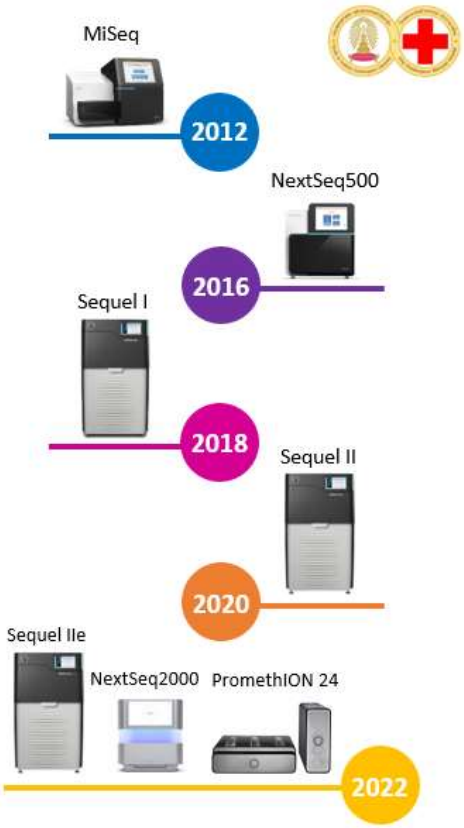


Million times faster

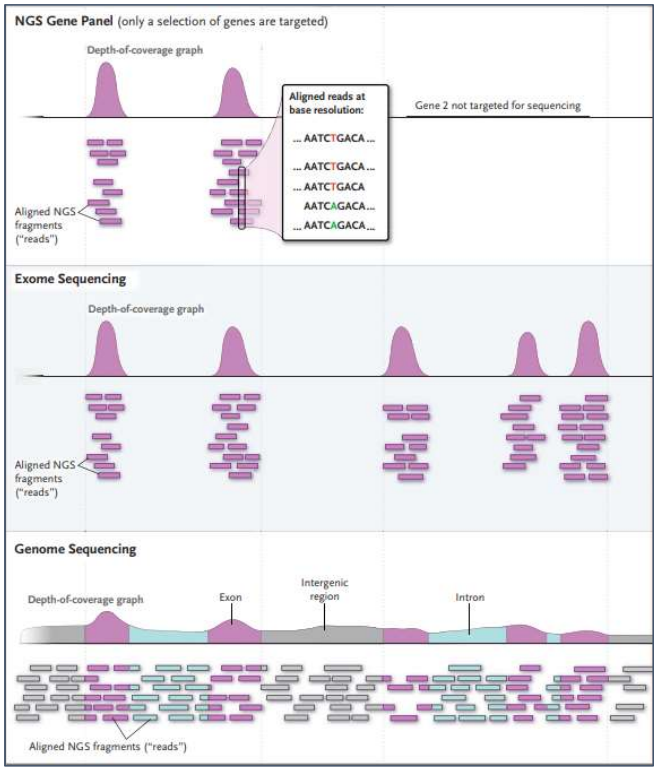
(1 million years → year)



Sequencer



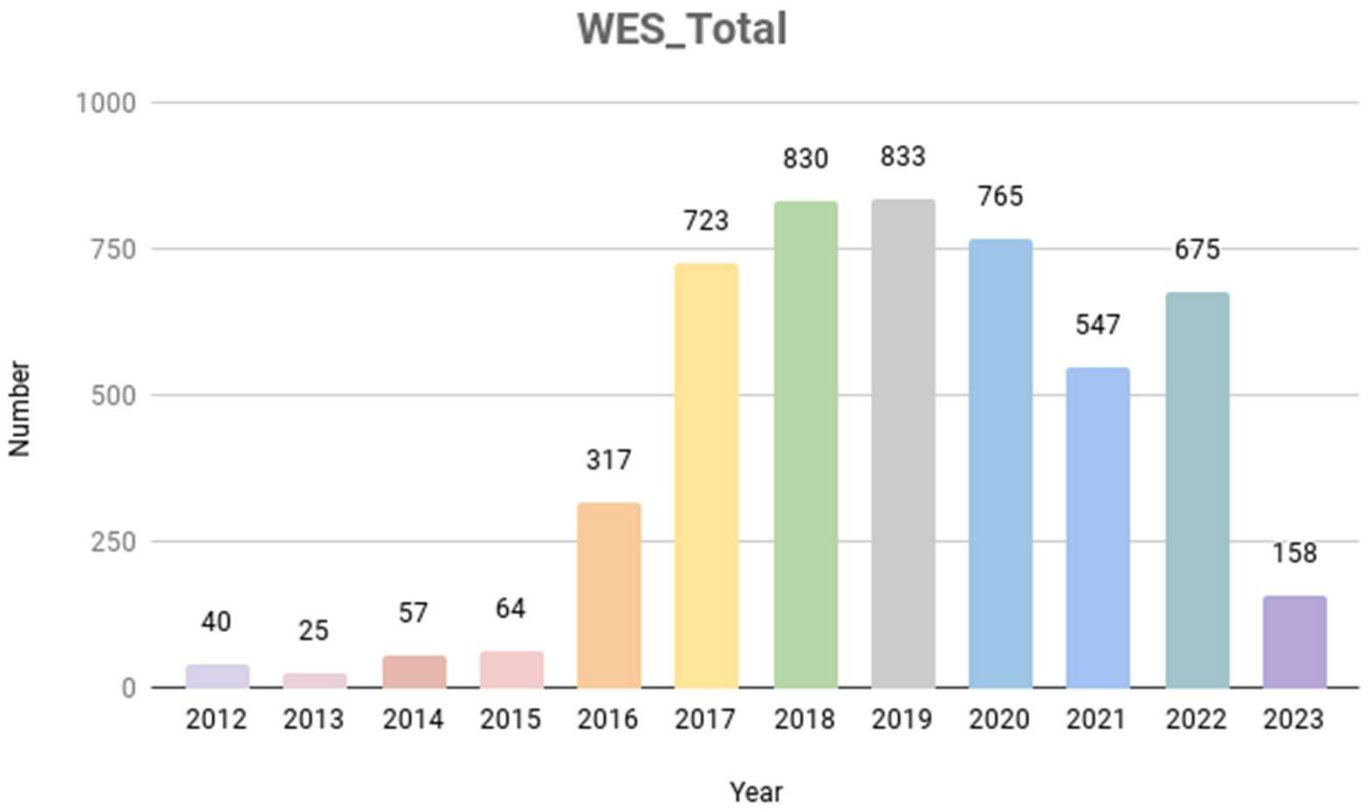
Application

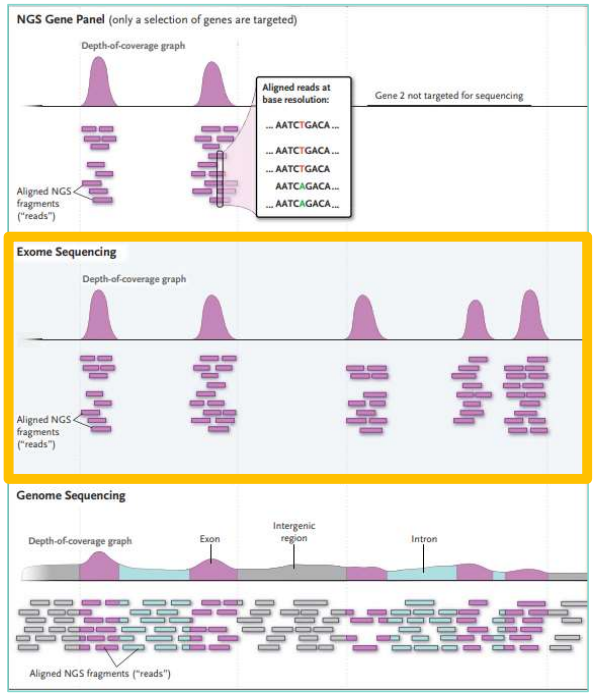


n engl j med 379;14 nejm.org October 4, 2018

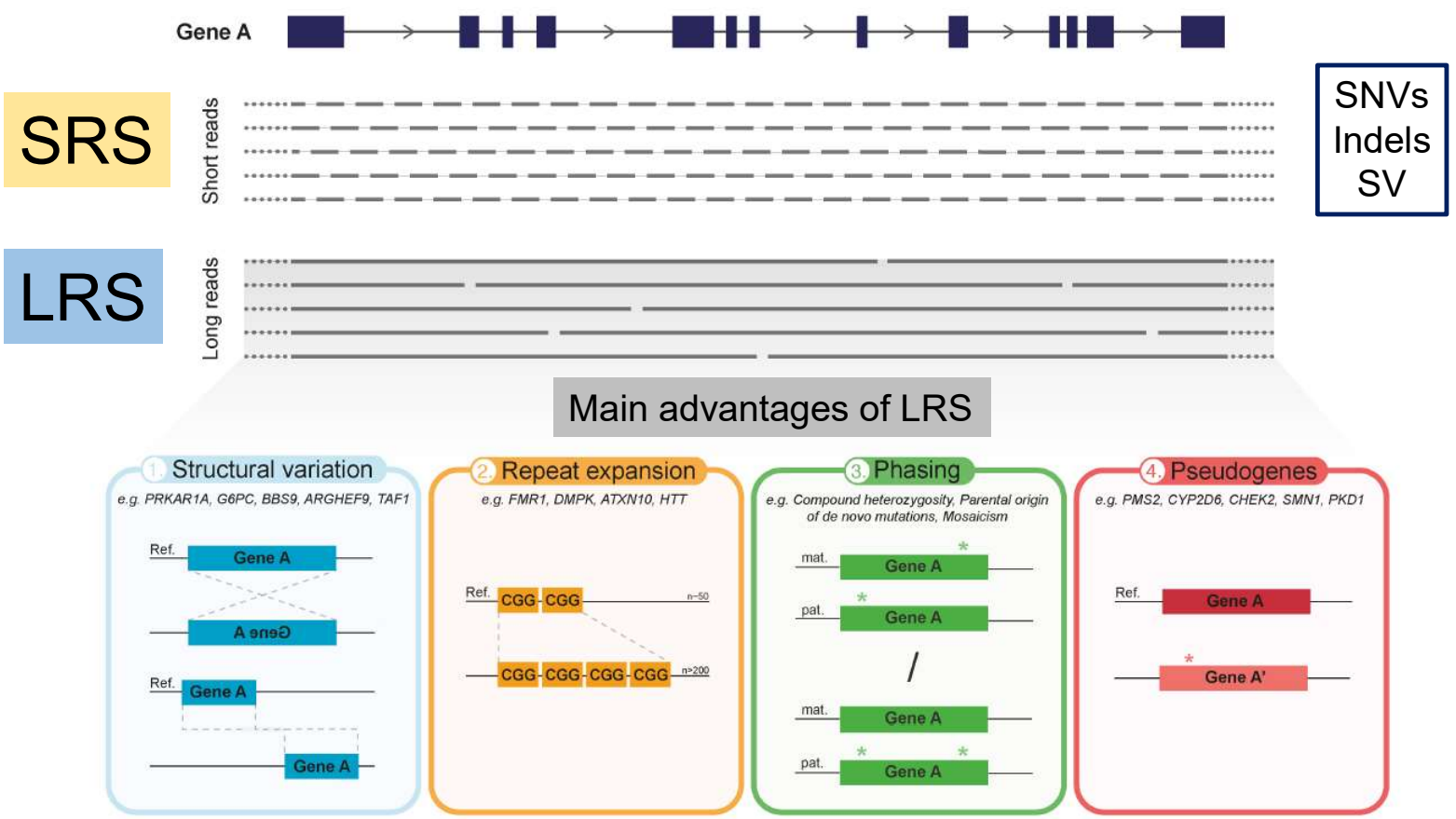
NGS short-read: Exome Sequencing

N = 5034 cases





Comprehensive Characterization of Human Genomes on Long-read sequencing



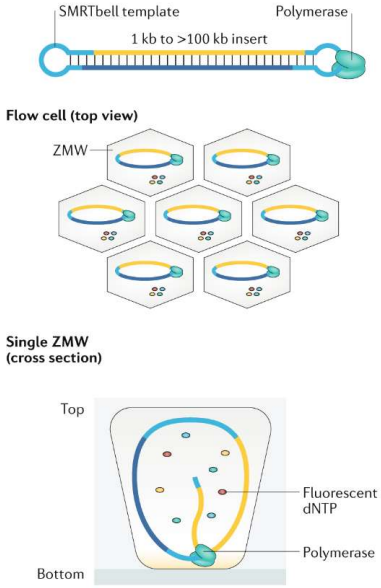
doi: 10.3389/fgene.2019.00426

NGS Long-read

Single molecule

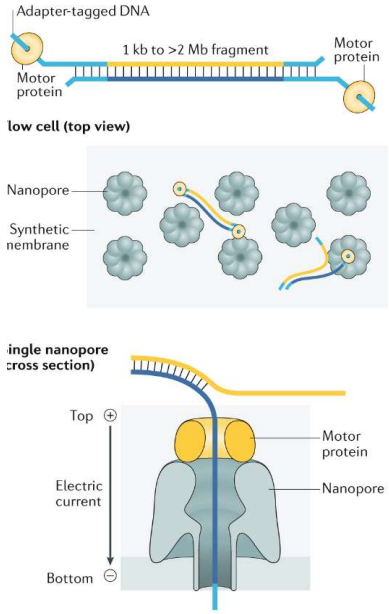
- **Pacific Biosciences**

Read length:15 – 20 kb, >100 kb
Accuracy: >99.9%, >99%



- **Oxford Nanopore**

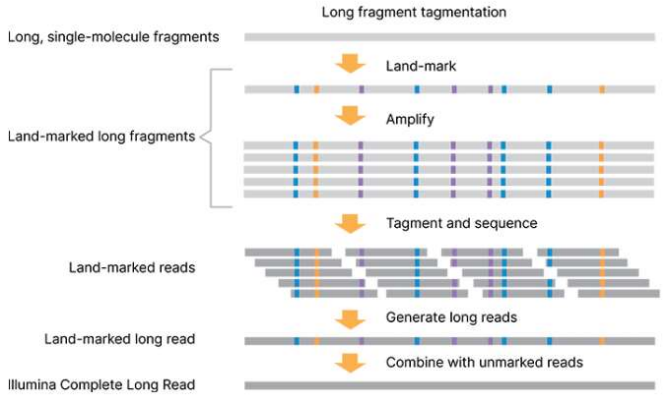
Read length: up to 4 mb
Accuracy: >99%, 99.9%



Synthetic long-read

- **illumina**

Read length: 5-7kb (N50)
Accuracy: 99.87%



https://sapac.illumina.com/

Logsdon, G.A., Vollger, M.R. & Eichler, E.E. Long-read human genome sequencing and its applications. *Nat Rev Genet* (2020).

NGS Long-read

❖ Single molecule real time sequencing

- **Single molecule real time (SMRT) from PacBio**

PacBio

Sequel system Revio system

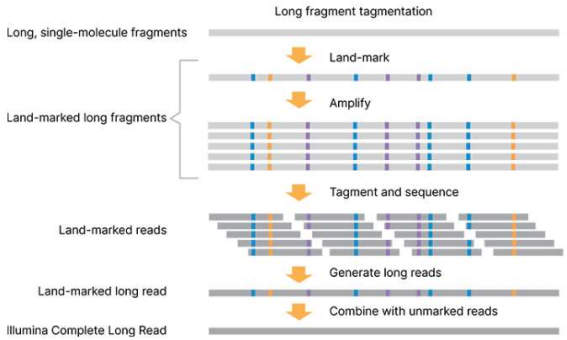
- **Oxford Nanopore Technologies (ONT)**

Oxford NANOPORE Technologies

❖ Synthetic long reads

illumina

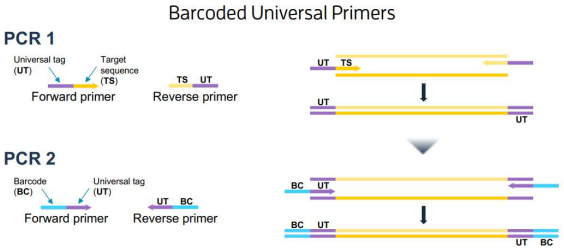
NovaSeq X series



PacBio long-read: Application for genomic

Targeted sequencing

- Amplicons sequencing



www.pacbio.com

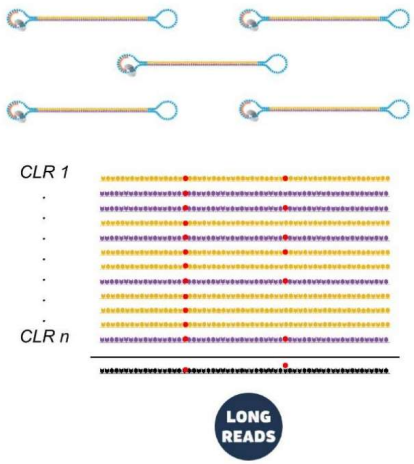
- No-Amp targeted sequencing

www.pacbio.com

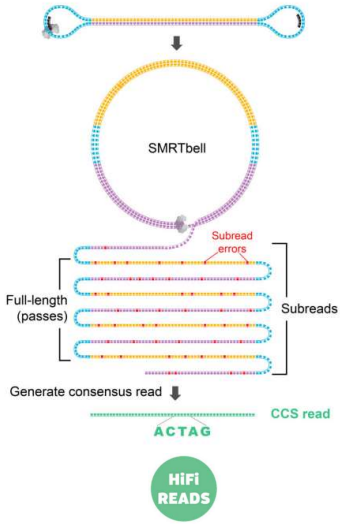
Whole genome sequencing

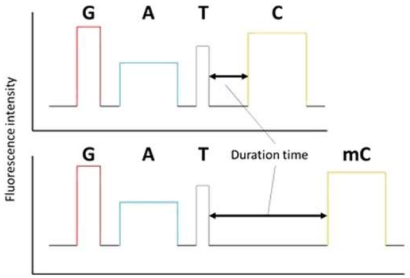
- Continuous Long Read (CLR)
- Circular Consensus Sequencing (CCS)/HiFi
- 5mC in CpG contexts

CLR read: >25 kb, up to 175 kb



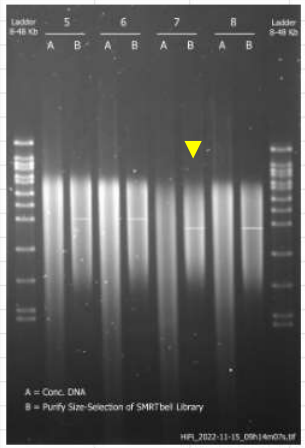
CCS/HiFi read: 15-20 kb

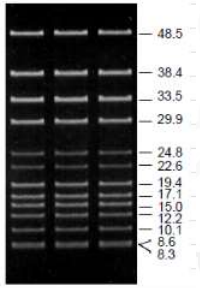


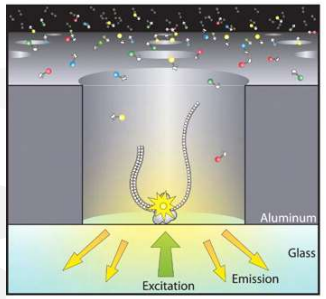


https://www.nature.com/articles/s10038-019-0679-0/figures/1

PacBio long-read: HiFi sequencing

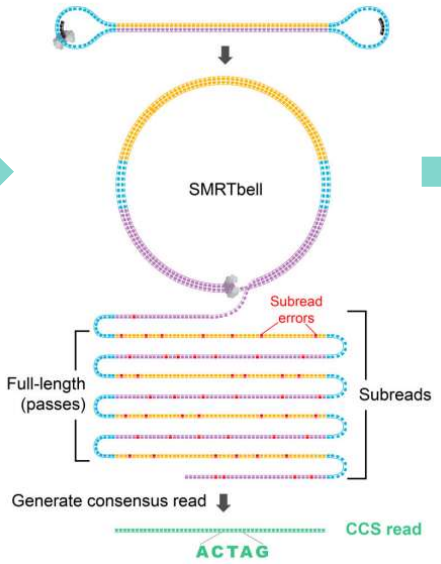


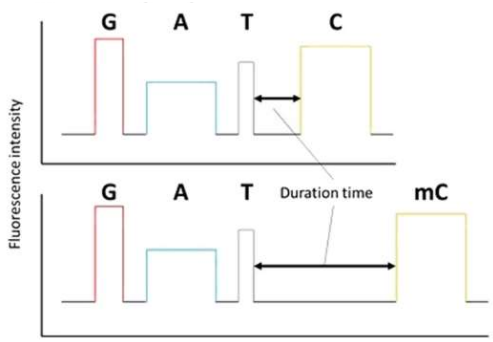




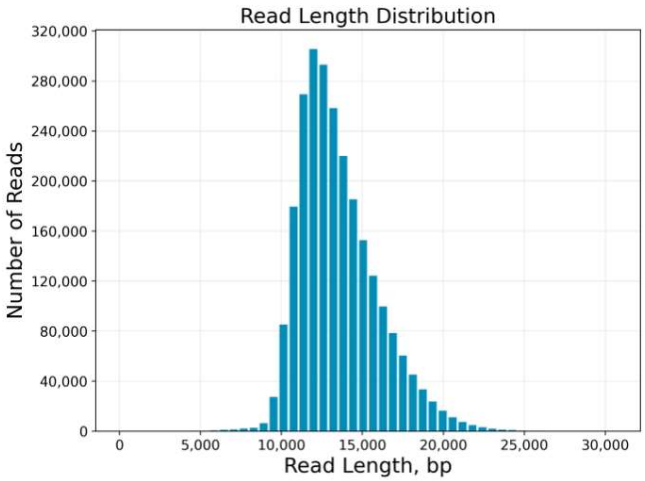
8M ZMW

https://www.science.org/doi/10.1126/science.1162986





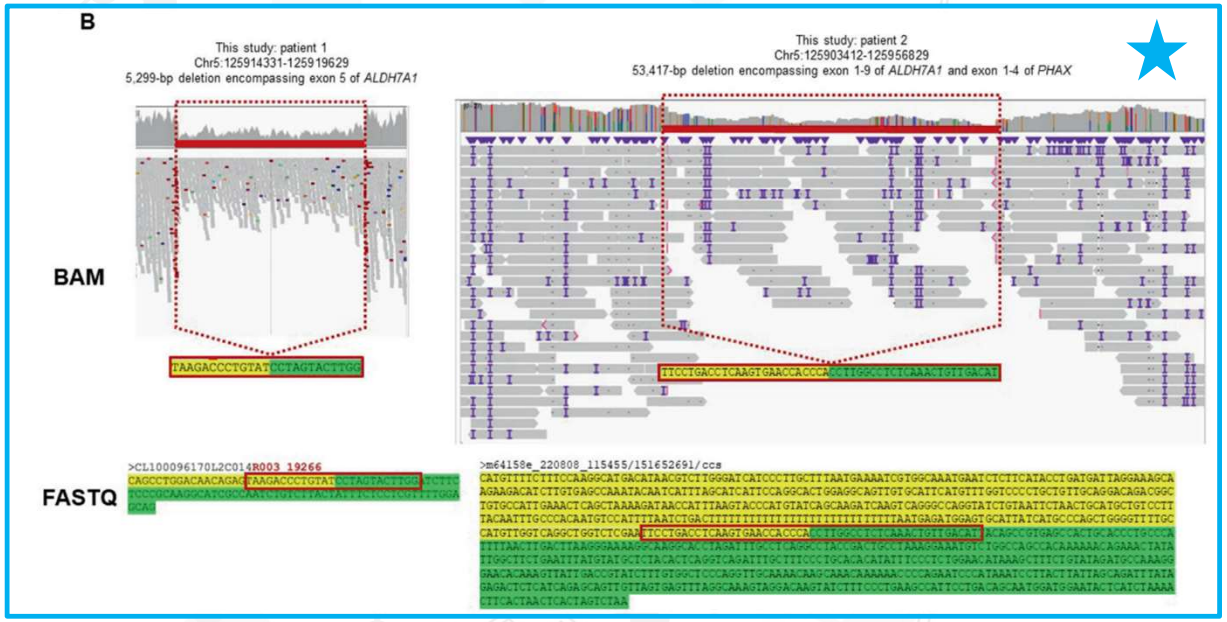
https://www.nature.com/articles/s10038-019-0679-0/figures/1



CCS Analysis Report

Value	Analysis Metric
2,506,009	HiFi Reads
33,893,803,607	HiFi Yield (bp)
13,525	HiFi Read Length (mean, bp)
Q36	HiFi Read Quality (median)
14	HiFi Number of Passes (mean)

50 Kb deletion detected by using WGS HiFi



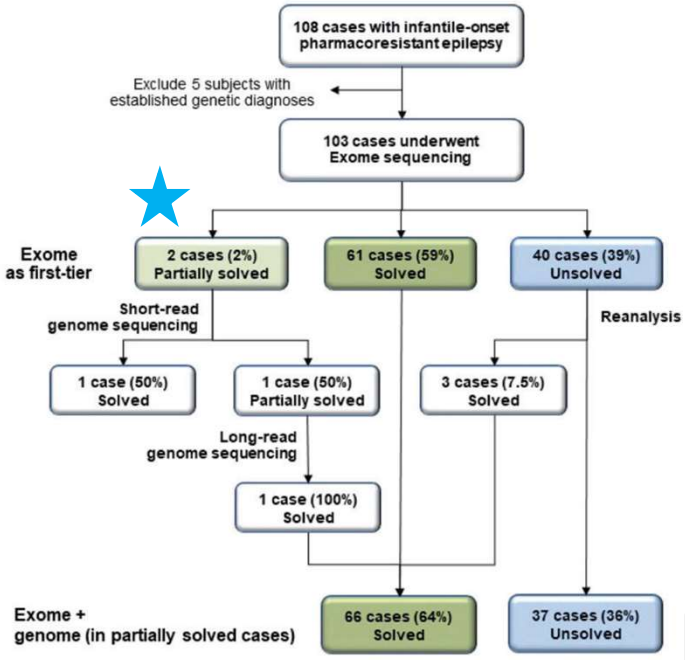
ESHG
www.nature.com/ejhg

ARTICLE

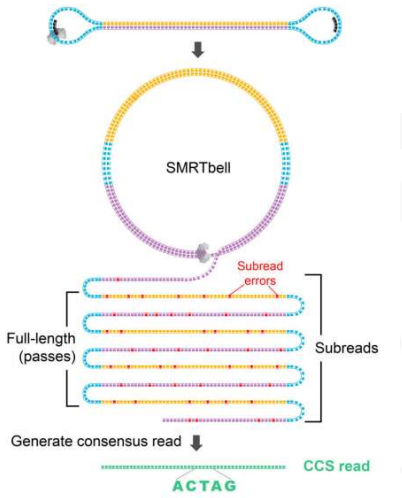
Check for updates

Exome sequencing as first-tier genetic testing in infantile-onset pharmacoresistant epilepsy: diagnostic yield and treatment impact

Ponghatai Boonsimma[1,2,4], Chupong Ittiwut[1,2,4], Wuttichart Kamolvisit[1,2], Rungnapa Ittiwut[1,2], Wanna Chetruengchai[1,2], Chureerat Phokaew[1,2], Chalurmpon Srichonthong[1,2], Sathida Poonmaksatit[3], Tayard Desudchit[3], Kanya Suphapeetiporn[1,2]✉ and Vorasuk Shotelersuk[1,2]



Repeat expansion detection by using WGS HiFi



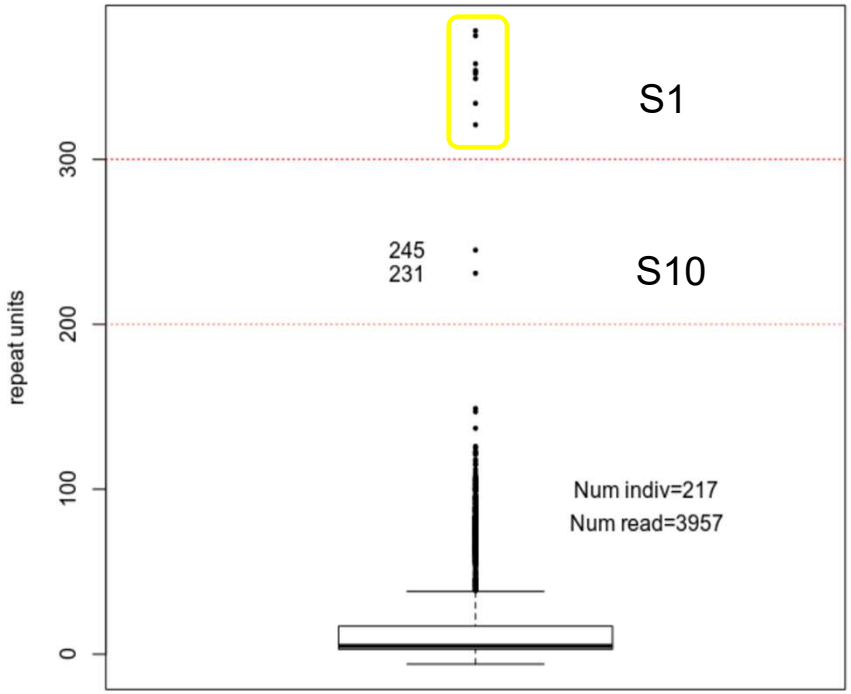
CCS Analysis Report

Value	Analysis Metric
2,506,009	HiFi Reads
33,893,803,607	HiFi Yield (bp)
13,525	HiFi Read Length (mean, bp)
Q36	HiFi Read Quality (median)
14	HiFi Number of Passes (mean)

tandem-genotypes

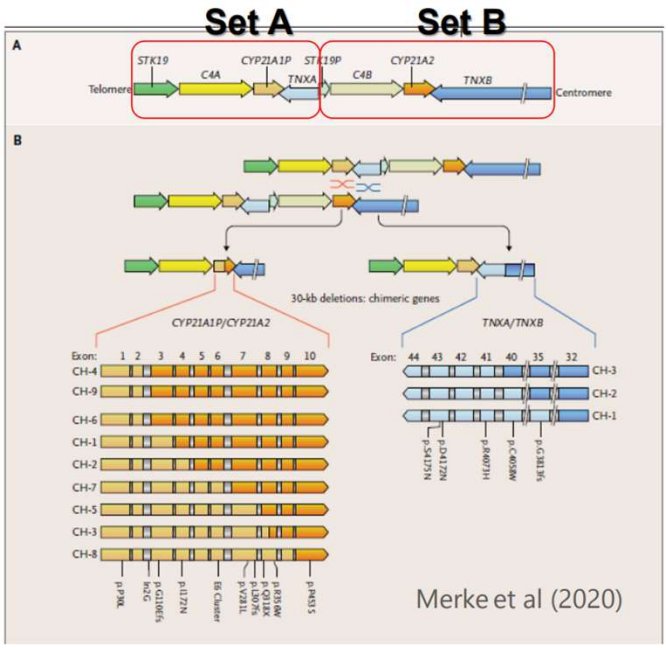
chr	[redacted]
from	[redacted]
to	[redacted]
pattern	[redacted]
gene	[redacted]
location	intron

S1	_1_fwd	5,5,5,321
S1	_1_rev	5,6,6,352,354
S1	_2_fwd	334,349,358,375
S1	_2_rev	6,6,6,6,353,378
S2	_1_fwd	1,3
S2	_1_rev	3,3,3
S2	_2_fwd	2,2,3,3
S2	_2_rev	3
S2	_3_fwd	2,2,3,3
S2	_3_rev	2,2,3,3
S3	_1_fwd	3
S3	_1_rev	4
S3	_2_fwd	3,4,4
S3	_2_rev	3,4
S3	_3_fwd	3,3,3,4
S3	_3_rev	3,3,4,4,4,4



PacBio long-read: Amplicon sequencing

21-Hydroxylase Deficiency: molecular challenges



Merke et al (2020)

1.tandem: pseudogene
2.Huge: 100s Kb
3.SNV + SV

The *CYP21A2* gene is

HOMOLOGOUS

at 98% in exons

and at 96% in introns

to the non-functional

CYP21A1P pseudogene.

(Balsamo A. *et al.*, 2010)

The Journal of Clinical Endocrinology & Metabolism, 2022, **107**, 1939–1947
https://doi.org/10.1210/clinem/dgac187
Advance access publication 1 April 2022
Clinical Research Article

Long-read Amplicon Sequencing of the *CYP21A2* in 48 Thai Patients With Steroid 21-Hydroxylase Deficiency

Nithiphut Tantirukdham,[1,2,*] Taninee Sahakitrungruang,[3,*] Ratikorn Chaisiwamongkol,[3] Monnat Pongpanich,[4,5] Chalurmpon Srichomthong,[6,7] Adjima Assawapitaksakul,[6,7] Aayalida Buasong,[6,7] Siraprapa Tongkobpetch,[6,7] Patra Yeetong,[8] and Vorasuk Shotelersuk[6,7]

Long length PCR (one primer pair) 8.5 kb

↓

CCS sequencing

PacBio long-read: MDCU

Sequencer	Application	Sample	SMRT Cell
Sequel I	WGS (CLR)	12	115
2018 - 2021	Amplicon Sequencing: CYP2D6 CYP21A2, BAFME	156	4
	RNA sequencing (Iso-Seq)	5	10
	Microbial Assembly (Multiplex)	18	3
	Amplicon Sequencing: HLA	672	8
Sequel II/IIe	WGS HiFi	223	494
2020 - 2023	No-Amp targeted sequencing	1	1
	Amplicon Sequencing: HLA	96	1
	Total:	1183	636

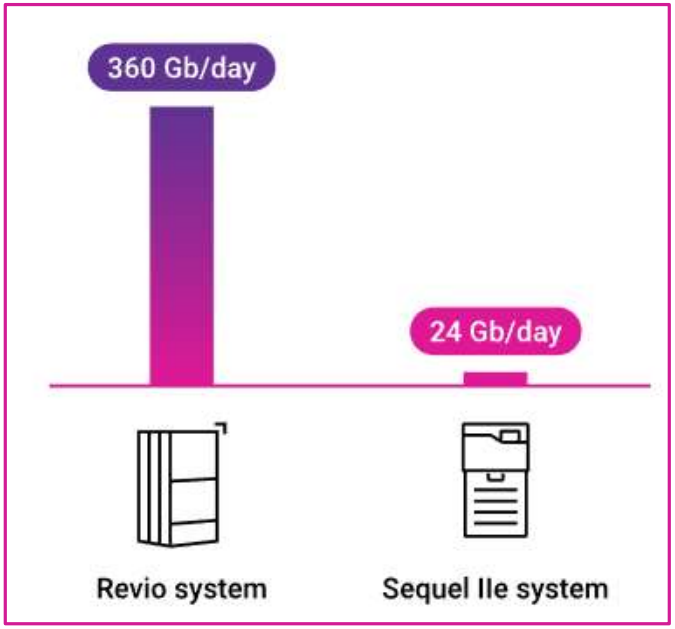
PacBio long-read: Application for genomic

Utility of long-read sequencing for All of Us

M. Mahmoud[1,2], Y. Huang[3], K. Garimella[3], P. A. Audano[4], W. Wan [3], N. Prasad[5], R. E. Handsaker[6], S. Hall[5], A. Pionzio[5], M. C. Schatz[7], M. E. Talkowski[8,9], E. E. Eichler[10,11], S. E. Levy[12], F. J. Sedlazeck[1,2,13]

[1]Human Genome Sequencing Center, Baylor College of Medicine, Houston, TX, USA,
[2]Department of Molecular and Human Genetics, Baylor College of Medicine, Houston, TX, USA,
[3]Data Sciences Platform, Broad Institute of MIT and Harvard, Cambridge, MA 02141
[4]The Jackson Laboratory for Genomic Medicine, Farmington, CT 06032 USA
[5]Discovery Life Sciences, Huntsville, AL 35806, USA
[6]Department of Genetics, Harvard Medical School, Boston, Massachusetts, USA
[7]Department of Computer Science, Johns Hopkins University, Baltimore, Maryland, USA
[8]Program in Medical and Population Genetics, Broad Institute of MIT and Harvard, Cambridge, MA 02141, USA
[9]Center for Genomic Medicine, Massachusetts General Hospital, Boston, MA, USA,
[10]Genome Sci, University of Washington, Seattle, WA, USA,
[11]Howard Hughes Medical Institute, University of Washington, Seattle, WA, USA,
[12]HudsonAlpha Institute for Biotechnology, Huntsville, AL 35806,
[13]Department of Computer Science, Rice University, Houston, Texas, USA

"HiFi reads produced the most accurate results for both small and large variants."



1,300 30X HiFi human /year	88 30X HiFi human /year

https://doi.org/10.1038/s41592-022-01759-x

Method of the Year 2022: long-read sequencing

Long-read sequencing powers a more complete reading of genomic information.

This June, we published a special issue highlighting the success of the Telomere-to-Telomere (T2T) Consortium in presenting the first complete human genome. This achievement was made possible by a wide range of experimental and computational efforts. Among them was long-read sequencing, the main sequencing technology responsible for generating the T2T data, which arguably laid the foundation of this feat. Yet the work from the T2T Consortium is only one example of the vast number of discoveries long-read sequencing is enabling in reading genomes, transcriptomes and epigenomes in humans and other species. For its momentous methodological advancement and broad application, we have chosen long-read sequencing as our Method of the Year 2022.

Since the advent of next-generation sequencing nearly two decades ago, the pace of technological innovation has never slowed. Although powerful algorithms strive to connect short reads relying on overlapping sequences, the sheer length and complexity of many genomes pose severe hurdles in generating complete sequences, often resulting in many missing parts and errors. This motivated the development of various strategies for long-read sequencing. The two most widely used commercial technologies are Pacific Biosciences' Single Molecule Real-Time (SMRT) sequencing (average read length ~20 kb with >99.9% accuracy for HiFi reads) and Oxford Nanopore Technologies' nanopore sequencing (average read length ~100 kb for ultra-long reads, ~99% accuracy for R10.4). Their distinct sequencing principles and approaches to data generation yield sequencing reads with varied lengths, error rates and throughputs. Researchers may find one long-read sequencing technology better to meet their research goals and resource requirements, depending on the application, and both techniques are continually evolving. In a News Feature in this issue, Vivien Marx highlights voices from several researchers developing and applying long-read sequencing in various areas, including interesting stories from its early days and perspectives on the future.

As in many other fields where new technologies are emerging, computational methods are vital role to translating the rich information embedded in long-read sequences to biological discoveries. A Comment from Michael Schatz and colleagues highlights such developments. Active method development is ongoing for many long-read data analysis tasks, ranging from identifying different bases and chemical modifications in DNA and RNA to genome assembly and genome variation detection. One promising direction is to apply advanced statistical and machine learning methods, which have shown remarkable performance in other fields for many computationally challenging tasks. They are increasingly becoming the core elements of the toolbox for long-read data analysis, and we expect the trend to continue into the future.

Enabled by the multitude of method developments, long-read sequencing has found applications in almost all the major areas of genomics. Karen Miga, who co-leads the T2T Consortium, and colleagues present a Comment on applying long-read sequencing in discovering and analyzing genetic variation. As demonstrated by their T2T work, long-read data shines light on many previously dark regions of the genome, such as telomeres and other highly repetitive regions and complex structural variations. With the launch of other large-scale endeavors such as the Vertebrate Genomes Project, more high-quality genomes from human and other species are on the horizon.

Besides genomes, the study of transcriptomes, which are dynamic and tissue- and cell-type-specific in nature, also benefits considerably from long-read sequencing. As explained in a Comment from Hagen Tilgner and colleagues, long-read sequencing holds the potential to unveil the hidden complexity of transcriptomes, such as isoform structure and expression, down to the level of a single cell. Given the paramount role of gene regulation and intra- and intermolecular interaction in isoform diversity, this knowledge will lead to a more quantitative and complete understanding of transcriptomic dynamics and its underlying mechanisms.

Another exciting dimension of genomics where long-read sequencing is seeing substantial traction is epigenomics and epitranscriptomics. A Comment from Eva Maria Novoa and colleagues provides an overview of this fast moving area, which is boosted by long-read sequencing's ability to detect chemical modifications in DNA and RNA. Unlike standard chemical- or antibody-based detection methods, direct analysis of nanopore sequencing signal, as an example, has been shown to enable reading of different types of modifications. Considering the vast number of different DNA and RNA modifications, with many being underdetected and understudied, long-read sequencing opens a door to exciting discoveries about their distribution and functional significance.

The final Comment comes from Mads Albertsen, who highlights the surging area of applying long-read sequencing to microbial genomics and metagenomics. One common challenge when studying microbial genomes is that samples are often composed of a community of microbes, with individual species hard to separate or culture. With the help of long-read sequencing, high-quality metagenome-assembled genomes are now more than ever within reach. Such efforts will greatly accelerate our exploration of genomic information spanning the whole tree of life.

Despite its power, long-read sequencing technology does not reach perfection. Besides the unceasing race to generate longer reads with higher accuracy, optimizing cost effectiveness is another crucial consideration for improving its accessibility to more research communities. It also does not exist in isolation. Combined with other genomic methods, long-read sequencing has nurtured new frontiers for method development and biological research. We hope you share our excitement when reading this special issue, in which we also cover a number of Methods to Watch. We wish you a very happy 2023!

Published online: 12 January 2023

Long-read sequencing powers a more complete reading of genomic information.

Center of Excellence for Medical Genomics, MDCU
Excellence Center for Genomics and Precision Medicine, KCMH

NAC2023
18th NSTDA Annual Conference
การประชุมวิชาการประจำปี สวทช. ครั้งที่ ๑๘

Update State of the Art Genomic Platforms

ช่วงถามตอบ

GENOMICS THAILAND