



NORWEGIAN SEQUENCING CENTRE

## SMRT sequencing

Ave Tooming-Klunderud, NCS/CEES/UoO,  
[ave.tooming-klunderud@ibv.uio.no](mailto:ave.tooming-klunderud@ibv.uio.no)



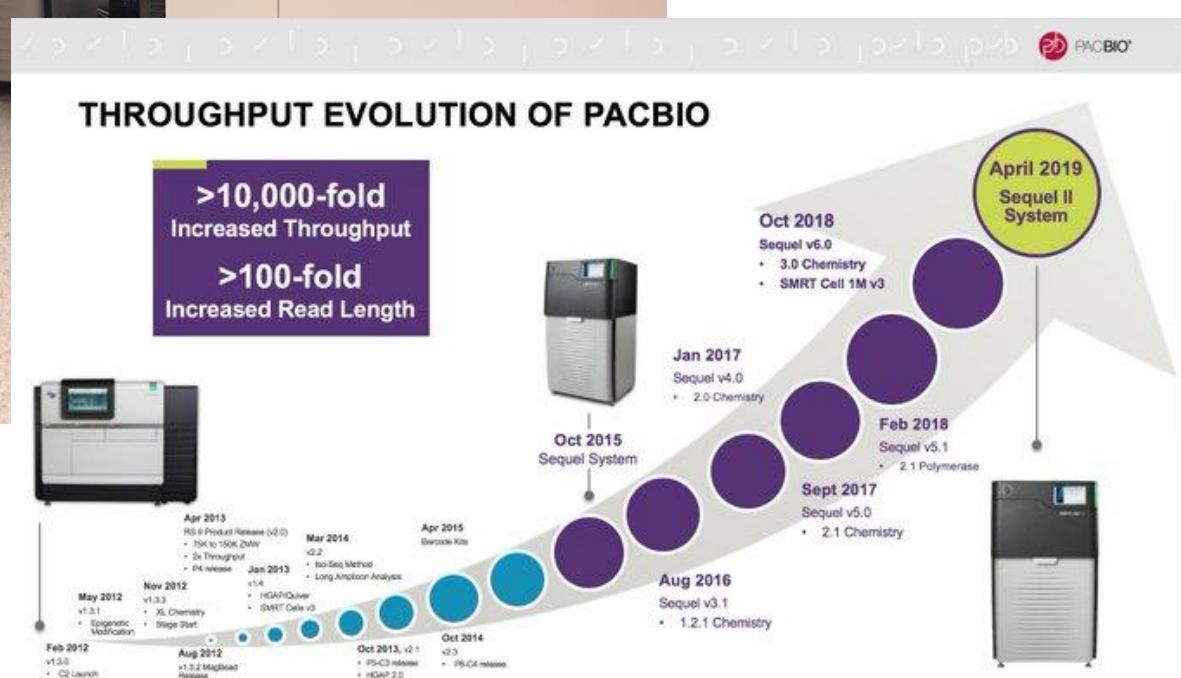
**CEES**

Centre for Ecological and  
Evolutionary Synthesis

HELSE • SØR-ØST



# PacBio sequencing since 2012



# Short read vs long read sequencing

## Short read sequencing

Amplification during sequencing

High read accuracy

## Long read sequencing

No amplification needed

Low single read accuracy/high consensus accuracy

### Requirements for input DNA:

Works with almost any DNA sample

Fragmented DNA

Low amount of DNA

High quality DNA needed

DNA fragments at least 40 kb long

High amount of DNA

**Low-input protocols available**

### Price:

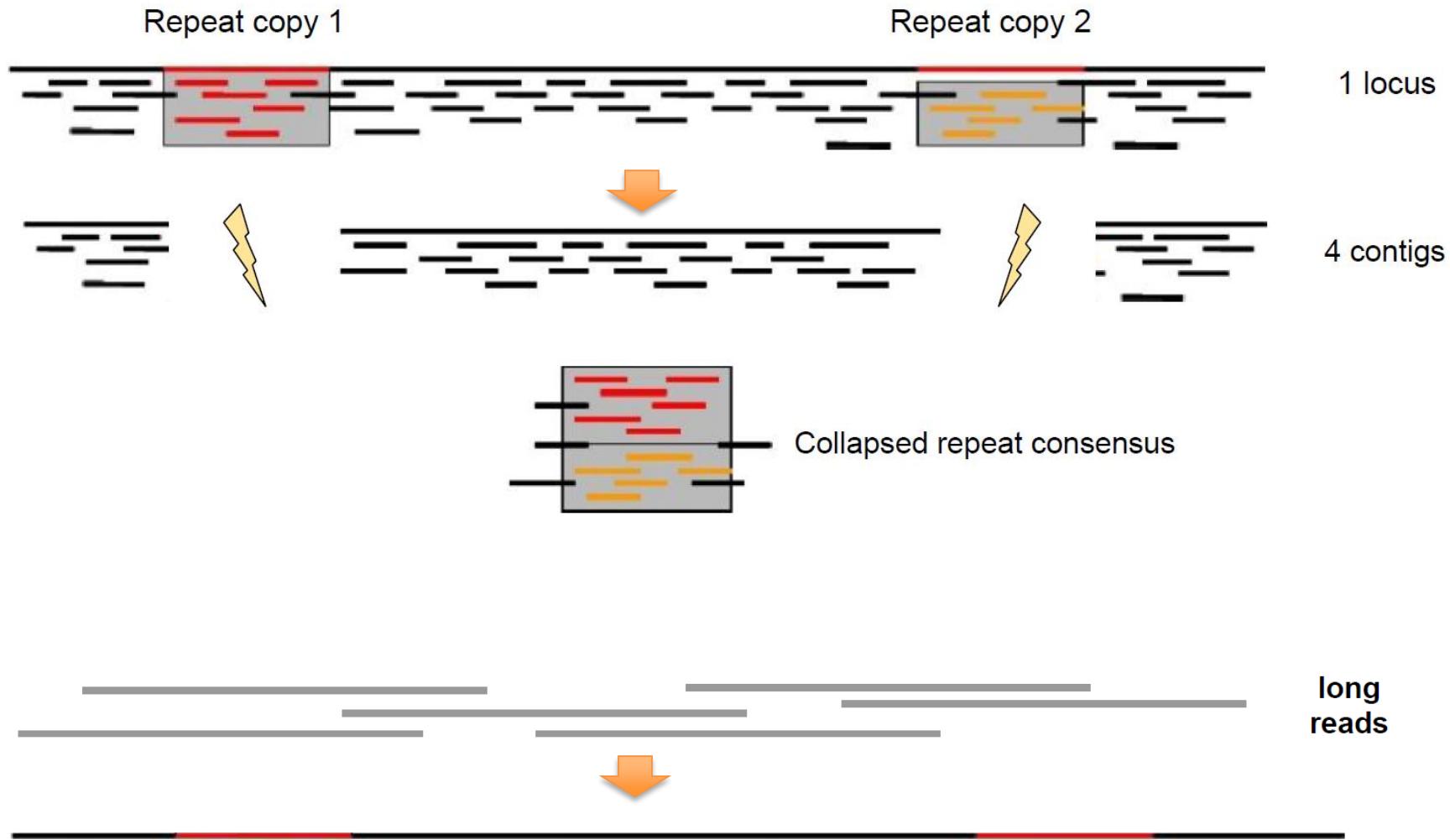
Low

Medium/High

**Low/medium**

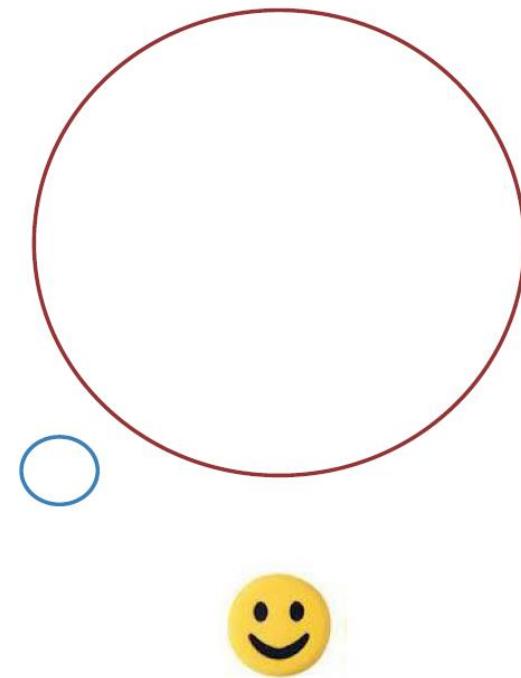
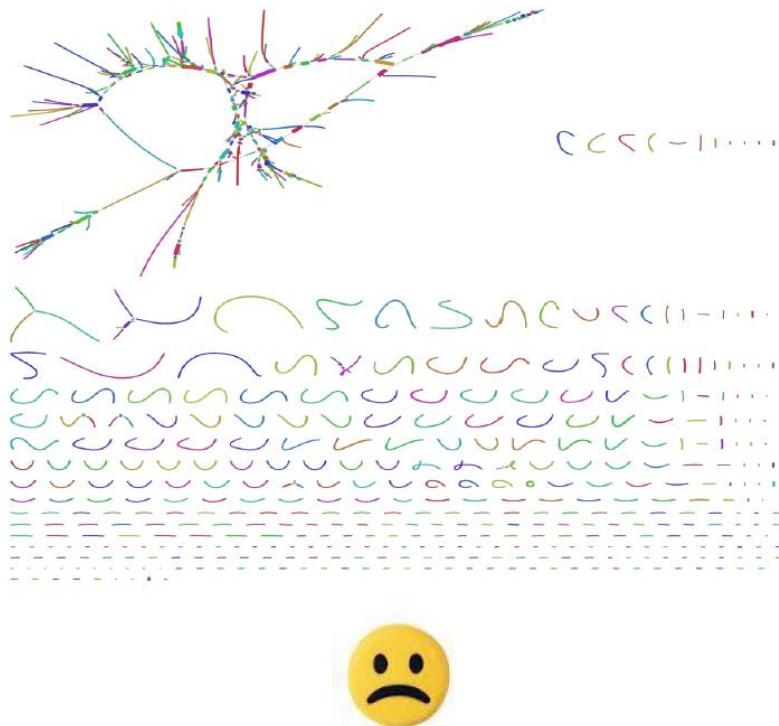
# Why long reads?

## Long reads can span repeats



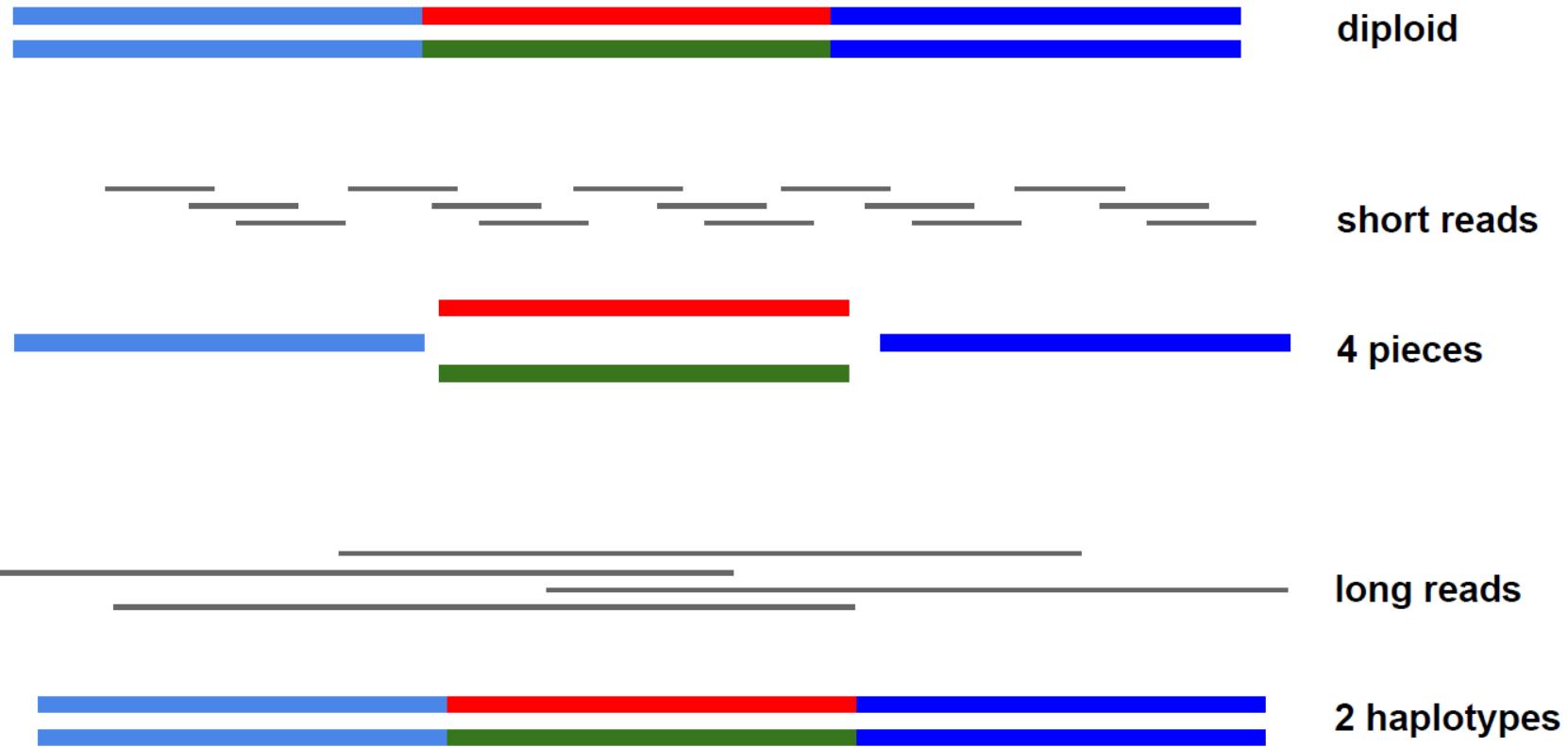
# Why long reads?

## Complete genomes



# Why long reads?

## Heterozygosity



# Why long reads?

## Phased haplotypes

Maternal ATGCTACGATCG**C**TCG

Paternal ATGG**T**ACGATCGATCG

Unphased: ATG <sup>C</sup>**G** TACGATCG <sup>C</sup>**A** TCG

A few phasing possibilities

Maternal ATGCTACGATCG**C**TCG

Paternal ATGG**T**ACGATCGATCG

Maternal ATGG**T**ACGATCGATCG

Paternal ATGCTACGATCG**C**TCG

Maternal ATGCTACGATCGATCG

Paternal ATGG**T**ACGATCG**C**TCG

Maternal ATGG**T**ACGATCG**C**TCG

Paternal ATGCTACGATCGATCG

.....ACTCAC.....GTAT**GGTGC**.....ACAG**TCTT**.....CTGAAGAT--AGCATT-----  
.....ACGCAC.....GTAT**CGTGC**.....ACACT**CTTT**.....CTGATGAT--AGCG**TTA**-----

↓ Sequencing

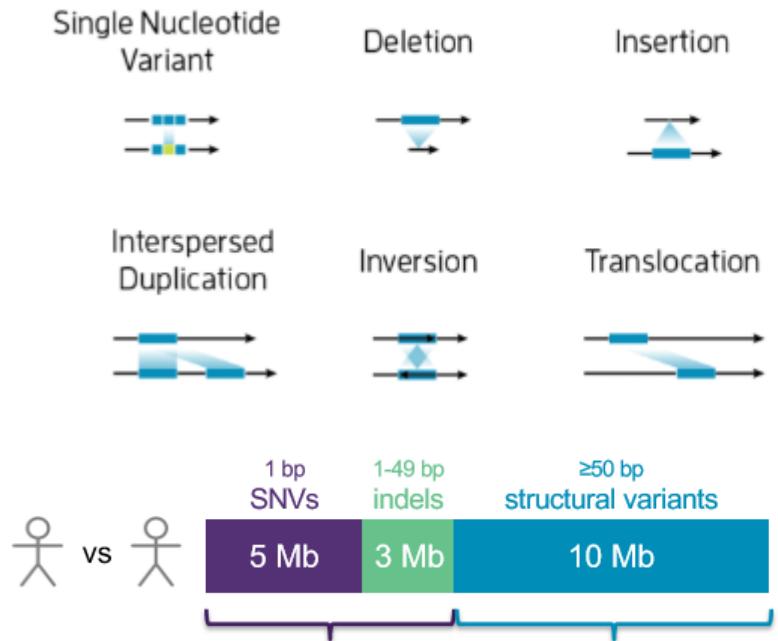
.....ACTCAC.....GTAT**GGTG**.....GTAT**CGTGC**.....TATCG**TGC**.....ACACT**TCT**.....ACAG**TCT**.....  
.....ACTCAC.....ACGCAC.....-----.....-----.....-----.....-----.....-----.....-----.....-----.....  
.....AGCG**TTA**.....GAAGAT--AGCATT

↓ Haplotype Assembly

.....T.....G.....G.....A.....A.....  
.....G.....C.....C.....T.....G.....

# Why long reads?

## Structural variation – the missing heritability, not just SNVs



"Small variants":

- Single Nucleotide Variants (SNVs)
- Indels <50 bp

Structural Variants (SVs):

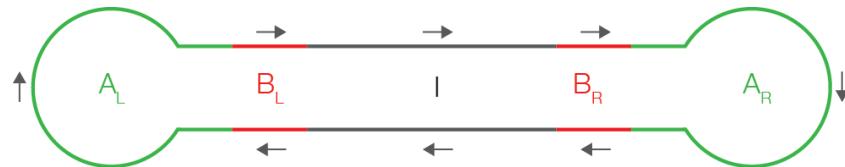
- Indels  $\geq 50$  bp
  - Duplications
  - Copy Number Variants (CNVs)
- Translocations
- Inversions



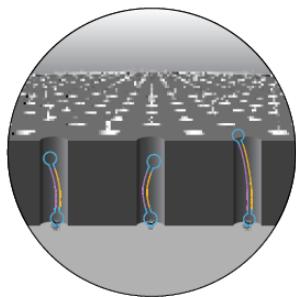
PacBio SMRT

1 bp SNVs	1-49 bp indels	$\geq 50$ bp structural variants
5 Mb	3 Mb	10 Mb

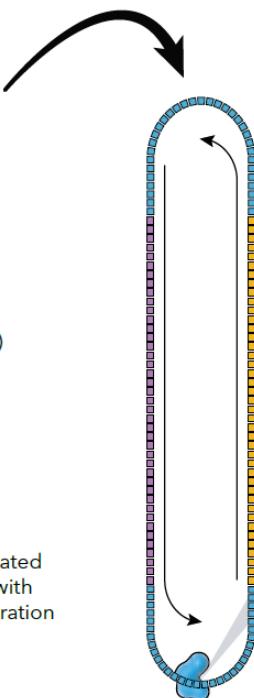
# The PacBio sequencing technology



## How SMRT Sequencing Works

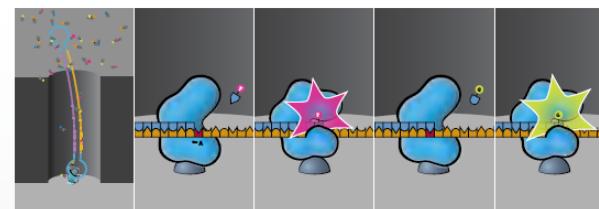


SMRT Cells contain millions of zero-mode waveguides (ZMWs)

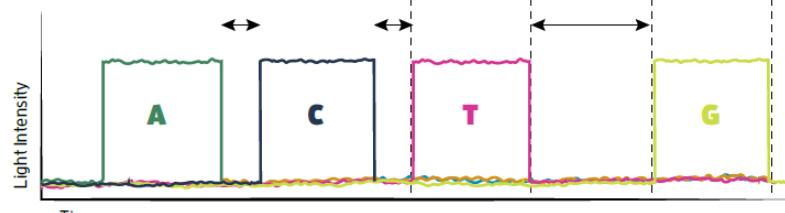


SMRTbell® templates enable repeated sequencing of circular template with real-time detection of base incorporation

A single molecule of DNA is immobilized in each ZMW



As anchored polymerases incorporate labeled bases, light is emitted

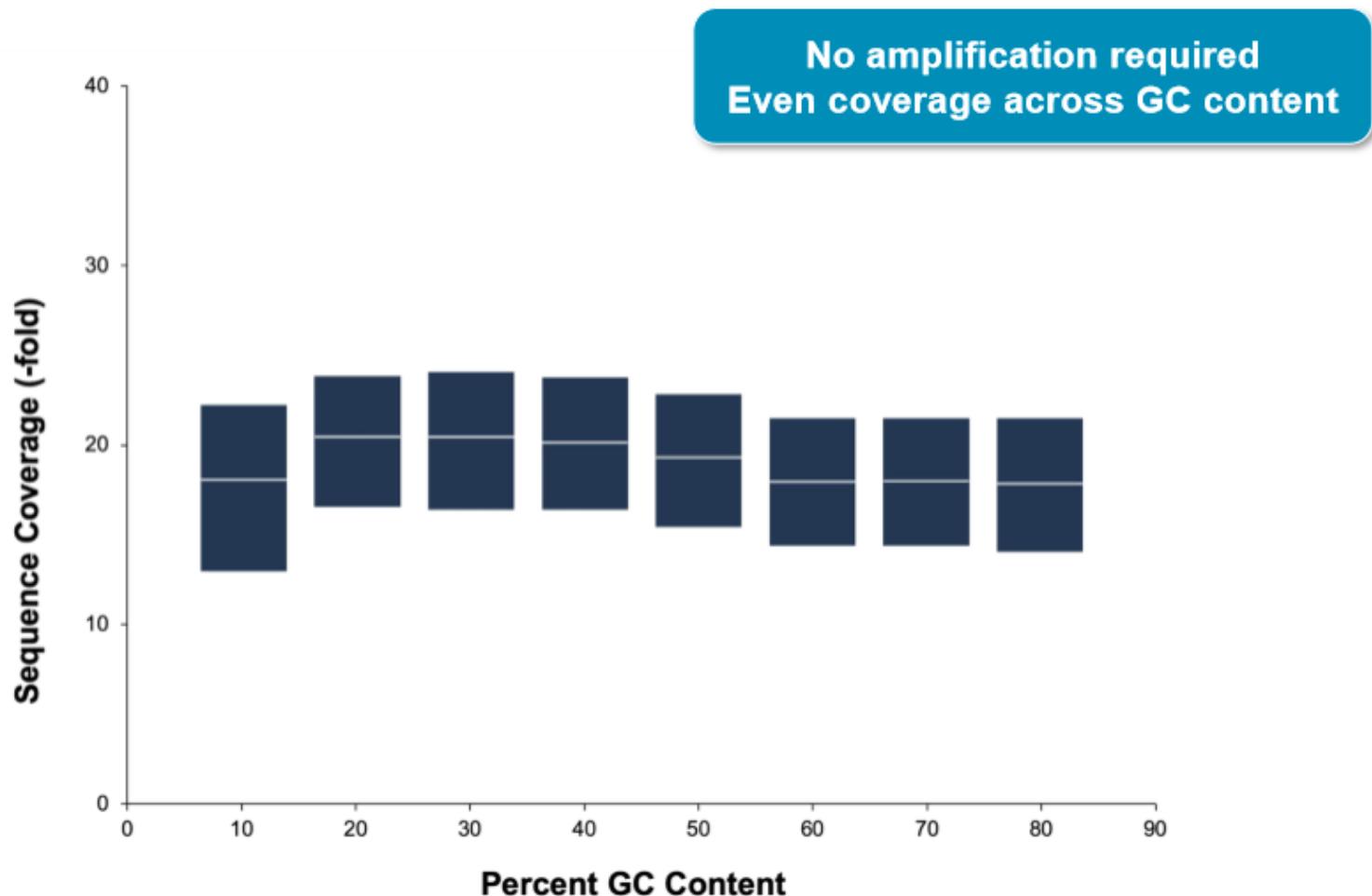


Nucleotide incorporation kinetics are measured in real time

Directly detect DNA modifications during sequencing

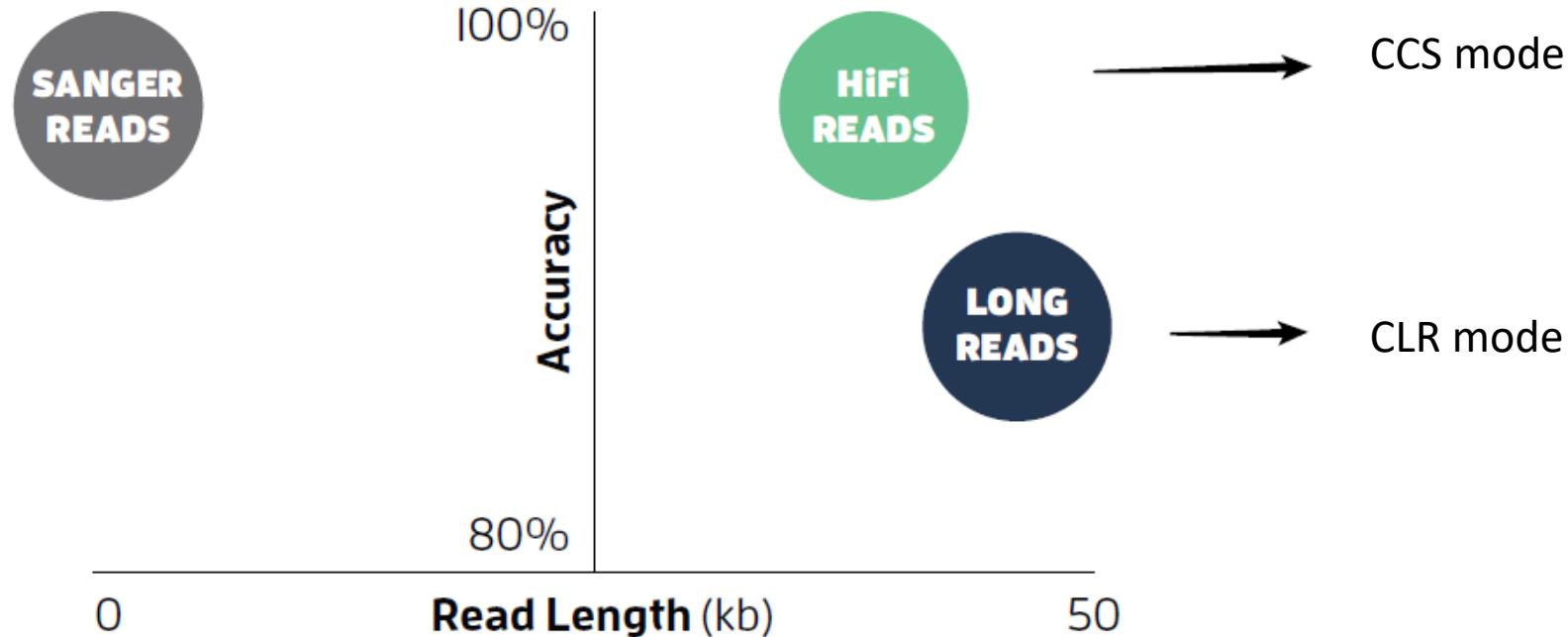
# Sequence performance: uniformity

## UNIFORM COVERAGE

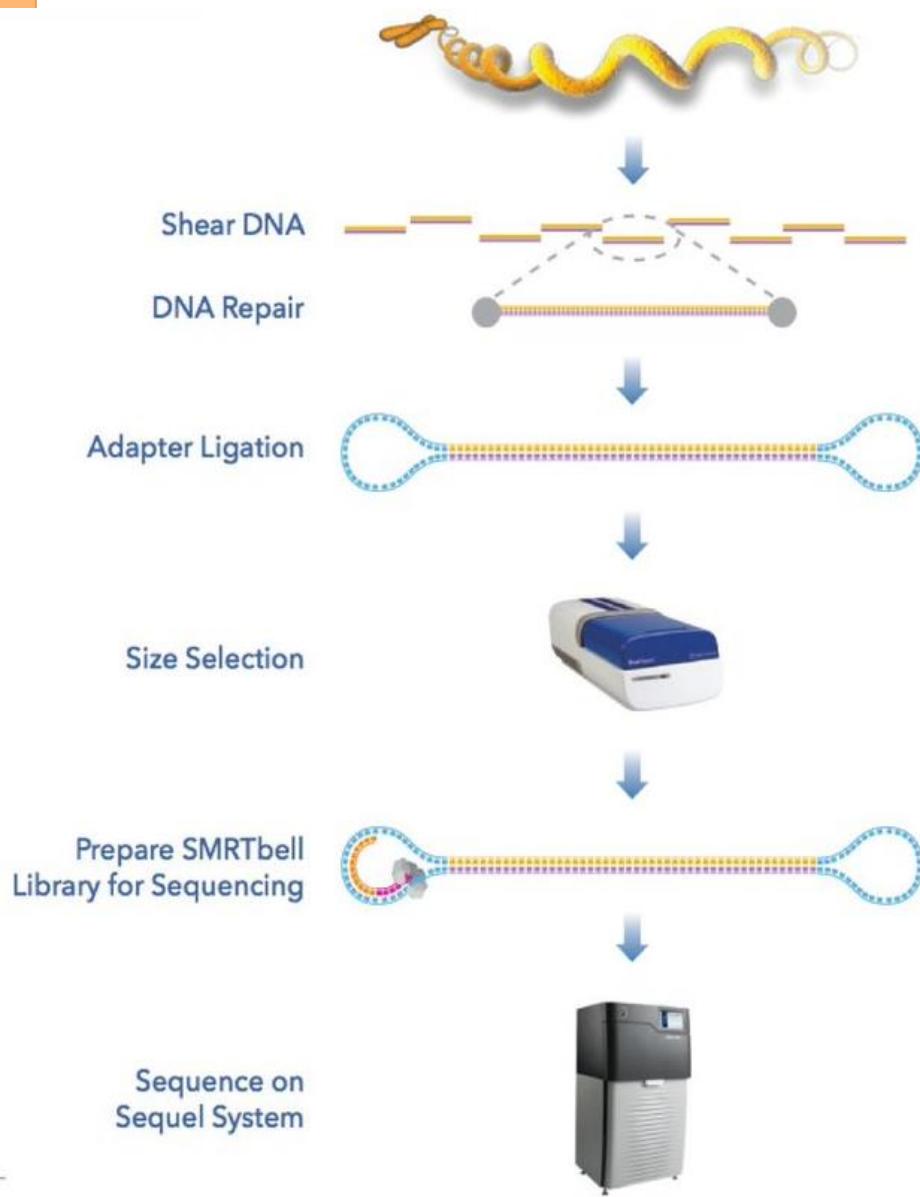


Mean coverage per GC window across a human sample. Data generated with a 15 kb human HiFi library on a Sequel II System using 2.0 Chemistry and Sequel II System Software v8.0

# Two sequencing modes:



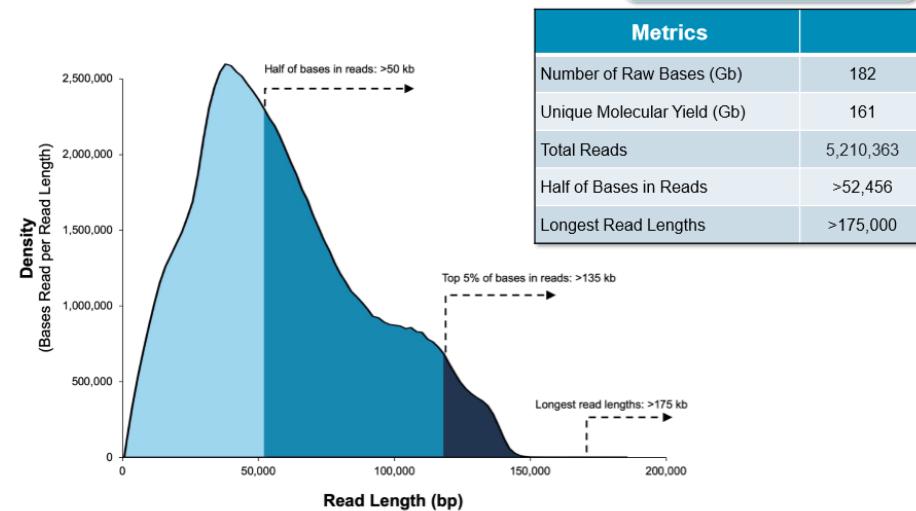
# Continuous long read sequencing mode



## Sequel II performance

### LONG READ SEQUENCING EXAMPLE: >35 KB SAMPLE – RAW DATA

Up to 160 Gb  
Average: 100 – 120 Gb



Data shown above from a 35 kb size-selected *E. coli* library using the SMRTbell Template Prep Kit on a Sequel II System (2.0 Chemistry, Sequel II System Software v8.0, 15-hour movie). Read lengths, reads/data per SMRT Cell 8M and other sequencing performance results vary based on sample quality/type and insert size.

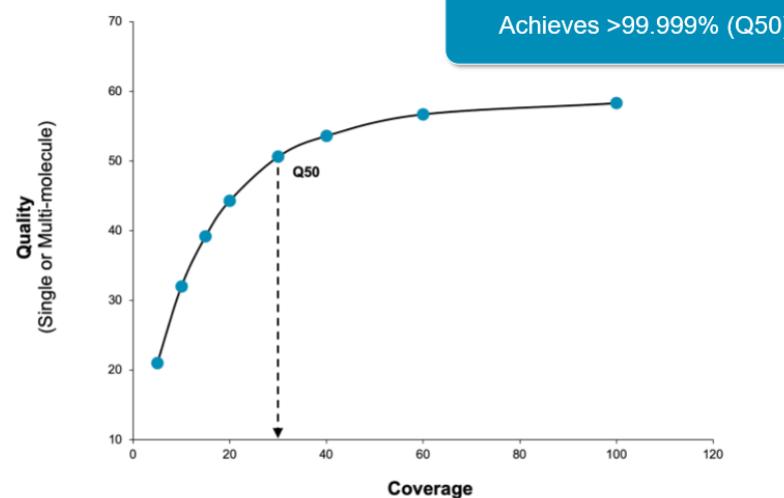
## Price:

100 Gb/SMRT cell: 230 NOK/Gb  
160 Gb/SMRT cell: 142 NOK/Gb

Typical use of data: *de novo* assembly

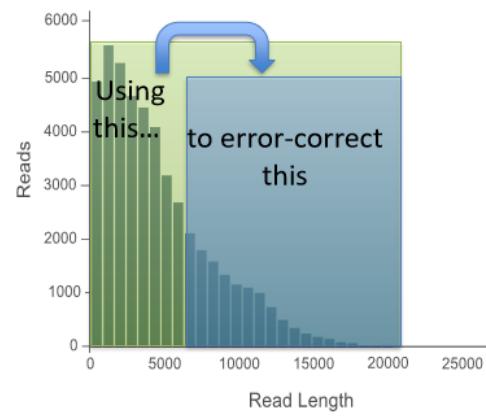
# CLR mode: accuracy

## HIGH CONSENSUS ACCURACY

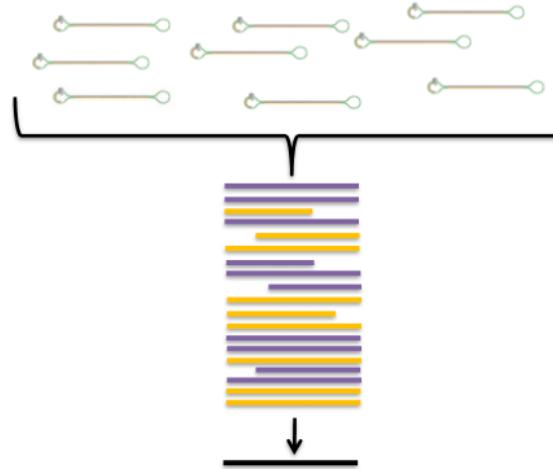


Consensus accuracy is a function of coverage and chemistry. The data above is based on a haploid bacterial genome run on the Sequel II System (2.0 Chemistry, Sequel II System Software v8.0). Single-molecule accuracy has similar coverage requirements.

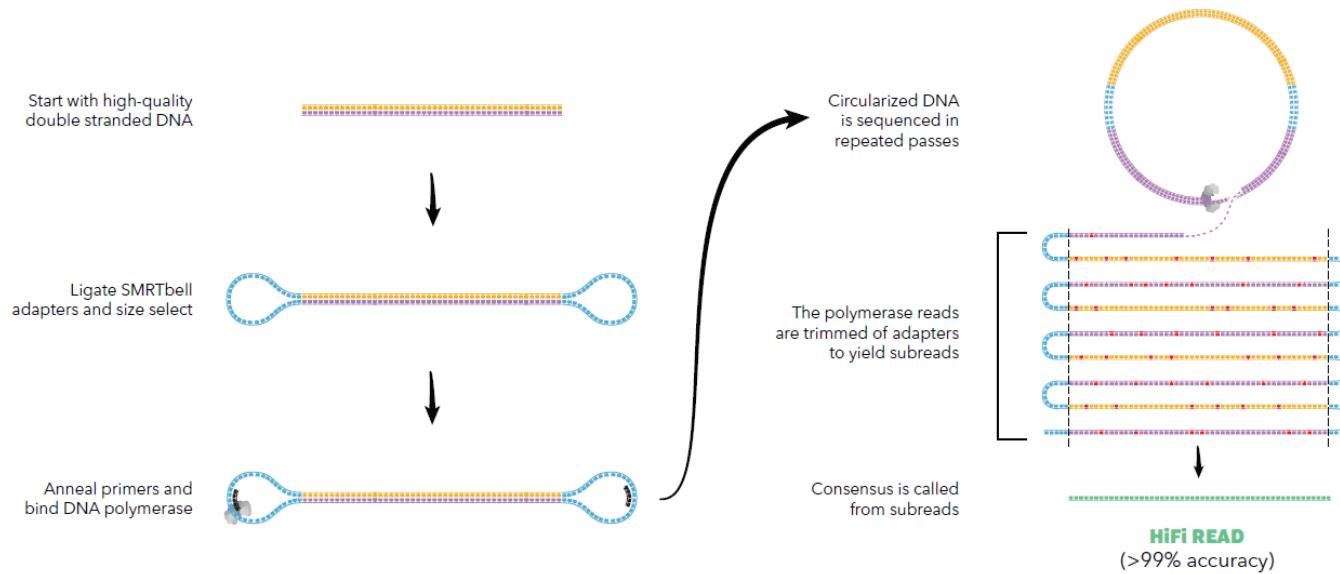
Genomic DNA: *de novo* assembly  
(60-100 x coverage needed)



Long amplicons: consensus is built using subreads from different fragments

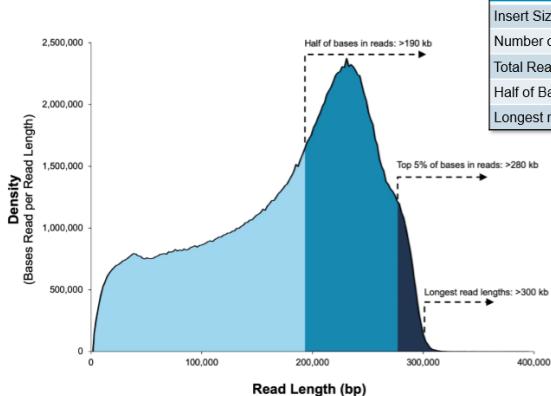


# Circular consensus sequencing mode



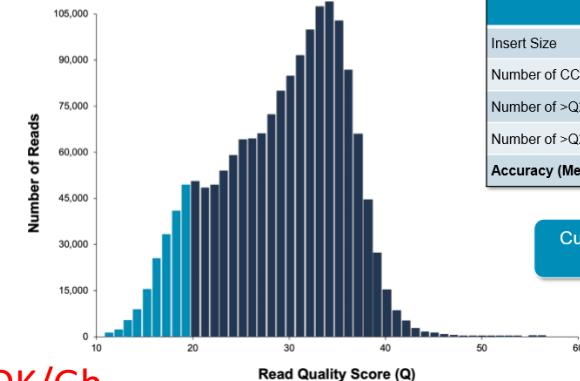
## Sequel II performance

### HIFI SEQUENCING (20 KB INSERT): RAW DATA



Price:  
CCS/HiFi data: 754 NOK/Gb

### HIFI SEQUENCING (20 KB INSERT): POST CCS PROCESSING



Data shown above from a 20 kb size-selected human library using the SMRTbell Template Prep Kit on a Sequel II System (2.0 Chemistry, Sequel II System Software v8.0, 30-hour movie). Read lengths, reads/data per SMRT Cell 8M and other sequencing performance results vary based on sample quality/type and insert size.

Data shown above from a 20 kb size-selected human library using the SMRTbell Template Prep Kit on a Sequel II System (2.0 Chemistry, Sequel II System Software v8.0, 30-hour movie). Read lengths, reads/data per SMRT Cell 8M and other sequencing performance results vary based on sample quality/type and insert size.

# Hifi read example

>M64089\_191020\_002935/346/cce

19,820 bp HiFi read, predicted QV: 33

>M64089\_191020\_002935/346/cce

19,820 bp HiFi read, predicted QV: 33

**19,812 bp correct, 8 errors**

**99.96% accurate (QV34)**

# Sequel II applications

## SEQUEL II SYSTEM KEY APPLICATIONS



### Whole Genome Sequencing for *De novo* Assembly

- Single Molecule, Real-Time (SMRT) Sequencing on the Sequel II System enables easy and affordable generation of high-quality de novo assemblies. With megabase size contig N50s, accuracies >99.99%, and phased haplotypes, you can do more biology – capturing undetected SNVs, fully intact genes, and regulatory elements embedded in complex regions.



### Variant Detection Using Whole Genome Sequencing with HiFi Reads

- With highly accurate long reads (HiFi reads) from the Sequel II System, powered by SMRT Sequencing technology, you can comprehensively detect variants in a human genome. HiFi reads provide high precision and recall for single nucleotide variants (SNVs), indels, structural variants (SVs), and copy number variants (CNVs), including in difficult-to-map repetitive regions.



### Structural Variation Detection

- With the Sequel II System powered by SMRT Sequencing technology and SMRT Link v8.0, you can affordably and effectively detect structural variants (SVs), copy number variants, and large indels ranging in size from tens to thousands of base pairs with high precision and recall.

# Sequel II applications

## SEQUEL II SYSTEM KEY APPLICATIONS



### Long-Read RNA Sequencing (Iso-Seq Analysis)

- With SMRT Sequencing and the Sequel II System, you can easily and affordably sequence complete transcript isoforms in genes of interest or across the entire transcriptome. The Iso-Seq method allows users to generate full-length cDNA sequences up to 10 kb in length – with no assembly required – to confidently characterize full-length transcript isoforms.



### Metagenomic Sequencing of Complex Populations

- The ability to identify and understand the functions of the complex microbial populations living in, on, and around us requires comprehensive characterization of each community member. Highly accurate long reads – HiFi reads – with single-molecule resolution make SMRT Sequencing and the Sequel II System ideal for full-length 16S rRNA sequencing, long-read metagenomic profiling, and shotgun metagenomic assembly

# Sequel II applications – WGS

Application	Whole Genome Sequencing					
	De Novo Assembly - HiFi Reads	De Novo Assembly - Long Reads	De Novo Assembly - for Low DNA Input	Microbial De Novo Assembly	Variant Detection	Structural Variation Detection
<b>Experimental Design</b>						
<b>With 1 SMRT Cell 8M you can:</b>	Produce reference quality assemblies for genomes up to 2 Gb	Produce reference quality assemblies for genomes up to 3 Gb	Produce reference quality assemblies for genomes up to 1 Gb Multiplex up to 2 small genomes on the Sequel II System	Sequence up to 48 microbes	With 2 SMRT Cells 8M, Call SNVs, InDels, and SVs in a 3 Gb genome	Call SVs for up to 2 samples with ~3 Gb genomes
<b>Minimum Recommended Coverage</b>	>15-fold HiFi read coverage	≥30-fold Unique Molecular Coverage (UMC) per haplotype	≥30-fold UMC per haplotype	≥30-fold UMC coverage per microbial genome	>15-fold HiFi of a human genome	<a href="#">5- to 25-fold Unique Molecular Coverage (UMC) coverage depending on study goals</a>
<b>Library Insert Size</b>	15 - 20 kb	>30 kb	~20 kb	10 - 15 kb	15 - 20 kb	>15 kb
<b>Sample Preparation</b>						
<b>Minimum Input Amount</b>	15 µg	≥1 µg for 10 kb ≥3 µg for >15 kb ≥5 µg for >30 kb	150 ng per 300 Mb genome size	1 µg per microbe	15 µg	3 µg
<b>Multiplexing/SMRT Cell</b>	N/A	N/A	N/A	Up to 48 microbes / SMRT Cell 8M Up to 16 microbes / SMRT Cell 1M	N/A	Up to 2 human samples / SMRT Cell 8M N/A SMRT Cell 1M
<b>Sequencing Mode</b>	CCS	CLR	CLR / CCS	CLR	CCS	CLR

<https://www.pacb.com/wp-content/uploads/Overview-Sequel-Systems-Application-Options-and-Sequencing-Recommendations.pdf>

# HiFi vs CLR assemblies

DOI: 10.1111/ahg.12364

ORIGINAL ARTICLE

Annals of  
human genetics WILEY

## Improved assembly and variant detection of a haploid human genome using single-molecule, high-fidelity long reads

Mitchell R. Vollger<sup>1\*</sup>  | Glennis A. Logsdon<sup>1\*</sup>  | Peter A. Audano<sup>1</sup> 

Arvis Sulovari<sup>1</sup> | David Porubsky<sup>1</sup> | Paul Peluso<sup>2</sup> | Aaron M. Wenger<sup>2</sup> |

Gregory T. Concepcion<sup>2</sup> | Zev N. Kronenberg<sup>2</sup> | Katherine M. Munson<sup>1</sup> 

Carl Baker<sup>1</sup> | Ashley D. Sanders<sup>3</sup> | Diana C.J. Spierings<sup>4</sup> | Peter M. Lansdorp<sup>4,5,6</sup> |

Urvashi Surti<sup>7</sup> | Michael W. Hunkapiller<sup>2</sup> | Evan E. Eichler<sup>1,8</sup> 

TABLE 1 Statistics of the high-fidelity (HiFi) and continuous long-read (CLR) genome assemblies

	Polishing	Total size (Gbp)	N50 (Mbp)	No. of contigs	Median QV	No. of CPU hours for assembly
HiFi						
<i>CHM13 genome</i>						
<i>Canu assembly</i>	None	3.03	25.51	5,296	40.41	~2,800
	Arrow	3.03	25.51	5,296	43.29	~10,000
	Racon	3.03	25.51	5,296	44.95	~2,950
	2× Racon	3.03	25.51	5,296	45.25	~3,100
	2× Racon+	3.03	25.51	5,296	45.25	~4,200
CLR						
<i>CHM13 genome</i>						
<i>FALCON assembly</i>	None	2.88	29.26	1,916	27.49	>50,000
	Quiver	2.88	29.26	1,916	40.73	>55,000
	Quiver+	2.88	29.26	1,916	42.70	>55,000

# Assemblers for HiFi data available



HOME | ABOUT | ARCHIVE | SUBMIT | SUBSCRIBE | ADVERTISE | AUTHOR INFO | CONTACT | HELP

Institution: RIKSHOSPITALET HF Sign In via User Name/Password

## HiCanu: accurate assembly of segmental duplications, satellites, and allelic variants from high-fidelity long reads

Sergey Nurk<sup>1</sup>, Brian P Walenz<sup>1</sup>, Arang Rhie<sup>1</sup>, Mitchell R Vollger<sup>2</sup>,  
Glennis A Logsdon<sup>2</sup>, Robert Grothe<sup>3</sup>, Karen H Miga<sup>4</sup>, Evan E Eichler<sup>5</sup>,  
Adam M Phillippy<sup>1</sup> and Sergey Koren<sup>1,6</sup>

OPEN ACCESS ARTICLE

ACCEPTED MANUSCRIPT

### This Article

Published in Advance August 14, 2020, doi:  
10.1101/gr.263566.120

*Genome Res.* 2020.

Published by Cold Spring

Table 1. *D. melanogaster* ISO1×A4 assembly benchmarking results for PacBio CLR and HiFi.

Assembly	Size (Mbp)	NG50 (Mbp)	Quality (QV)	BUSCO complete	Phase block NG50 (Mbp)	Intra-block switch error	QUAST (diffs per Mbp)
Canu + Purge_dups	141.81	14.09	37.4	98.5%	0.42	3.86%	0.018
CLR	128.15	0.31	35.5	86.7%	0.25	2.97%	
Peregrine	141.59	12.68	32.9	98.2%	0.07	1.78%	0.062
HiFi	20.53	0.00	33.5	1.0%	0.00	3.71%	
Canu + Purge_dups	145.19	13.72	<b>51.9</b>	98.7%	2.04	<b>0.03%</b>	0.015
HiFi	130.23	1.28	<b>46.9</b>	93.7%	1.26	<b>0.03%</b>	
HiCanu + Purge_dups	146.27	<b>20.16</b>	51.0	<b>98.8%</b>	<b>7.62</b>	<b>0.03%</b>	0.025
HiFi	132.53	<b>4.54</b>	46.7	95.5%	<b>4.45</b>	<b>0.02%</b>	

# HiFi assembly of large genomes - redwood



PACIFIC BIOSCIENCES®

## Tackling a Giant Genome with Highly Accurate Long-read Sequencing

Emily Hatas, Senior Director, Business Development

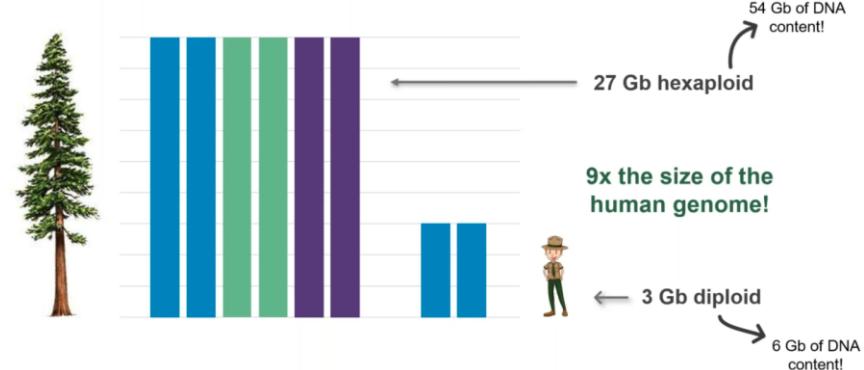
Recording available

### THE CALIFORNIA (COASTAL) REDWOOD GENOME



- One of the world's fastest-growing conifers
- Live for thousands of years
- Only 5% of the original old-growth coast redwood forest remains
- 27 Gb hexaploid genome
- Genome assemblies by ONT in 2019 and PacBio in 2020

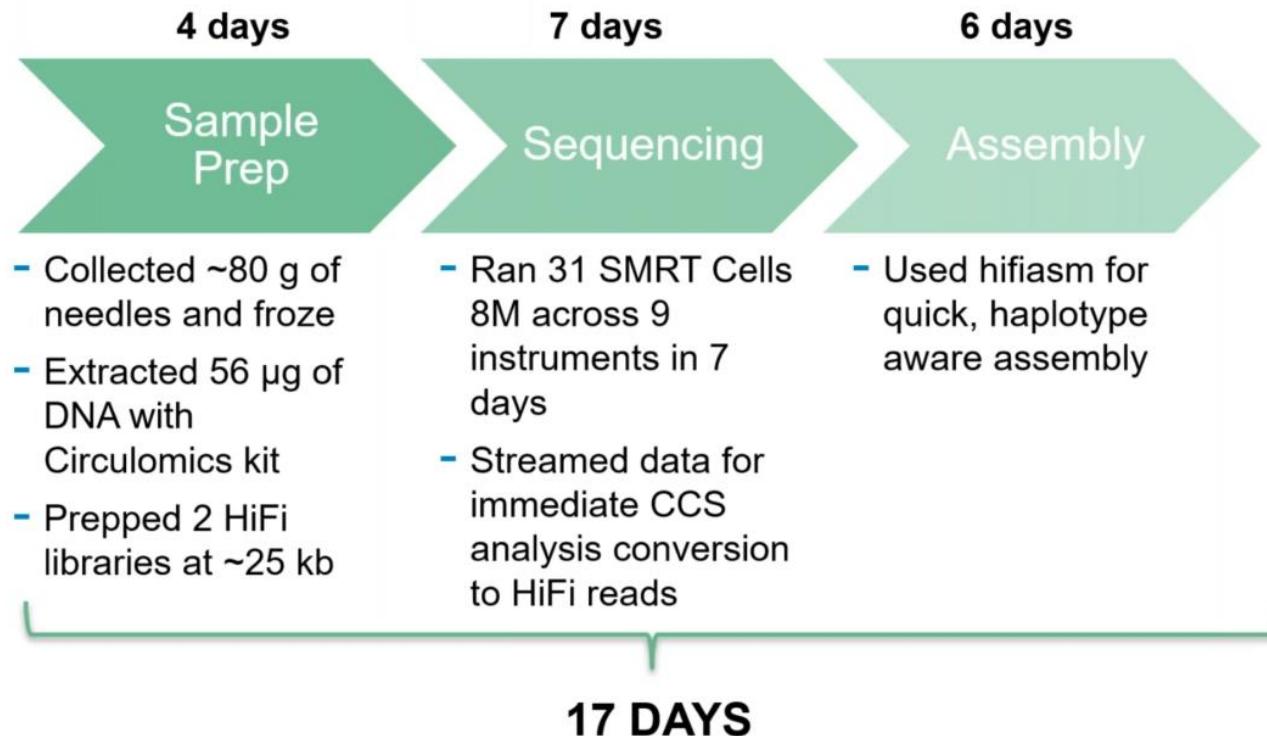
### THE REDWOOD GENOME IS LARGE AND COMPLEX



Courtesy of Pacific Biosciences of California, Inc.

# California Redwood project

## THE PROJECT WORKFLOW



# California Redwood project

## RESULTING GENOME ASSEMBLY

- Standard running parameters – no iteration
- Run on 64 cores with 512 Gb of RAM – no specialized or particularly large compute cluster

California Redwood Genome Assembly Results		
Methodology	PacBio HiFi reads	ONT + short reads <sup>1</sup>
Genome Coverage	22-fold	23-fold + 122-fold
Assembly Size (Gb)	47.7	26.5
Contig N50 (Mb)	1.92	0.11
BUSCO Complete	59%	56%
Mapped transcripts with frameshift errors <sup>2</sup>	0.12%	1.97%

BUSCO does not work well in conifers due to very long introns

### PacBio HiFi reads<sup>1</sup>

- 64 cores with 512 Gb of RAM
- ~46,000 CPU hours for HiFi generation (“error correction”)
- 6 days wall time, ~7,200 CPU hours for assembly

**6 days** vs **5-6 months** of wall time for just genome assembly

### ONT + short reads<sup>2</sup>



1. Sequencing and assembling mega-genomes of mega-trees: the giant sequoia and coast redwood genomes.

2. Using transcript set of *Abies alba* from [Neale, D. et al.](#) consisting of 22,561 transcript sequences

# What about hybrid approaches?

✗ **HiFi reads + short reads:** no benefit for contig building or polishing

✗ **HiFi reads + long reads:** may have marginal benefit to contiguity, but no readily available tools

✓ **HiFi + scaffolding:** technologies like optical maps and HiC help assign your high-quality HiFi genome assemblies into chromosomes

## HIFI READS FOR GENOME ASSEMBLY OF OTHER PLANTS & ANIMALS

With HiFi reads you can assemble reference-quality genomes with one technology

- Reach high contiguity, completeness, and correctness ensuring downstream utility
- Phase haplotypes for allele-specific genomic information
- Generate complete genomes in half the assembly time of traditional long reads



### Small-bodied Species

- 150 Mb genome
- 14.4 Mb contig N50
- 99.999% accuracy (Q50)



### Newly Sequenced Species

- 800 Mb genome
- 26.5 Mb contig N50
- 98.1% of genome phased



### Large, Complex Species

- 11 Gb genome
- 20 Mb contig N50
- Assembly in 12 hours

# HiFi sequencing data available

## NEW HIFI DATASETS – “TRY BEFORE YOU BUY”

 **bioRxiv**  
THE PREPRINT SERVER FOR BIOLOGY

New Results

**Highly accurate long-read HiFi sequencing data for five complex genomes**

Ting Hon, Kristin Mars, Greg Young,  Yu-Chih Tsai, Joseph W. Karalius, Jane M. Landolin,  Nicholas Maurer,  David Kudrna, Michael A. Hardigan,  Cynthia C. Steiner,  Steven J. Knapp,  Doreen Ware,  Beth Shapiro,  Paul Peluso,  David R Rank

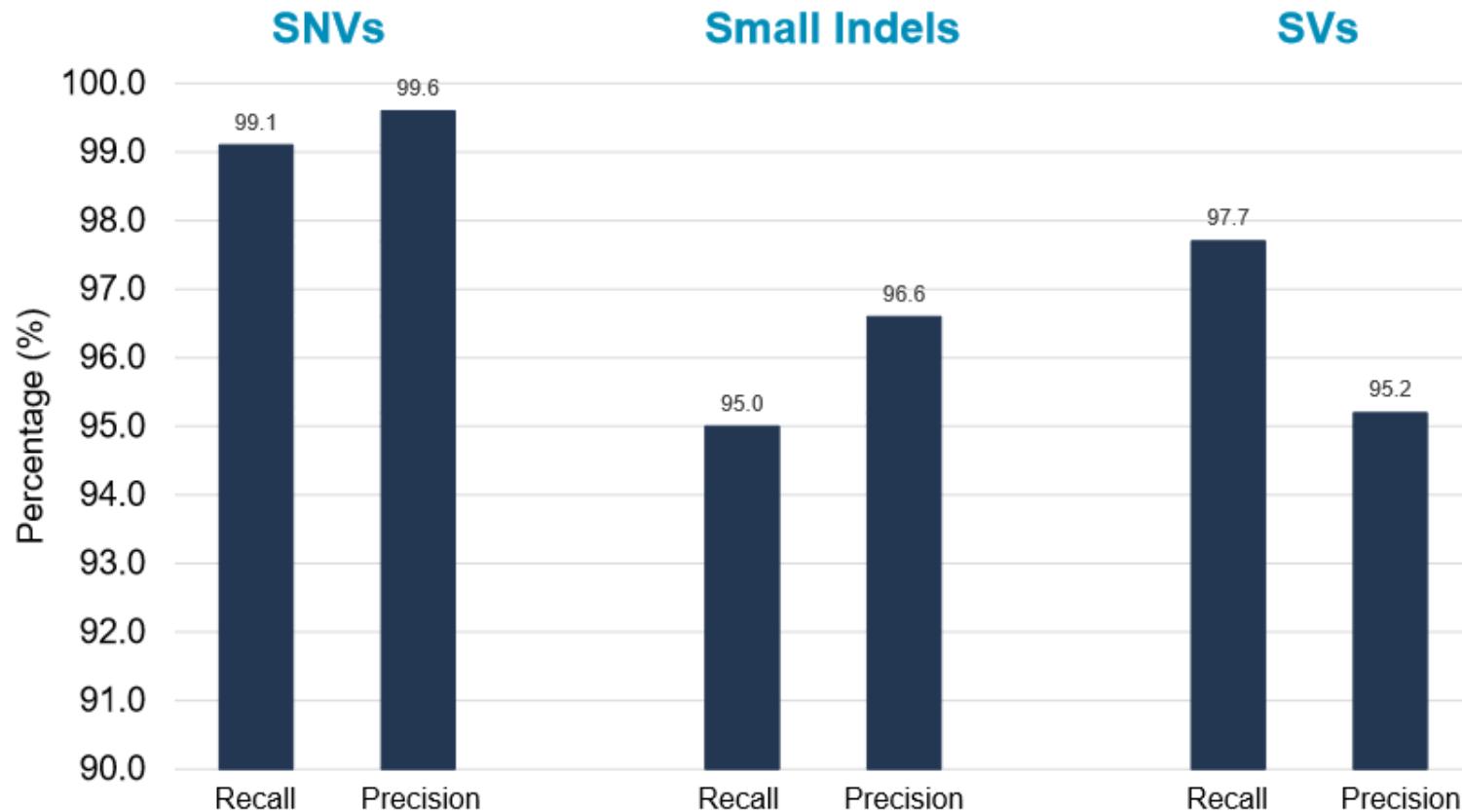
doi: <https://doi.org/10.1101/2020.05.04.077180>



# Variant calling



## EXAMPLE: VARIANT CALLING WITH HIFI READS

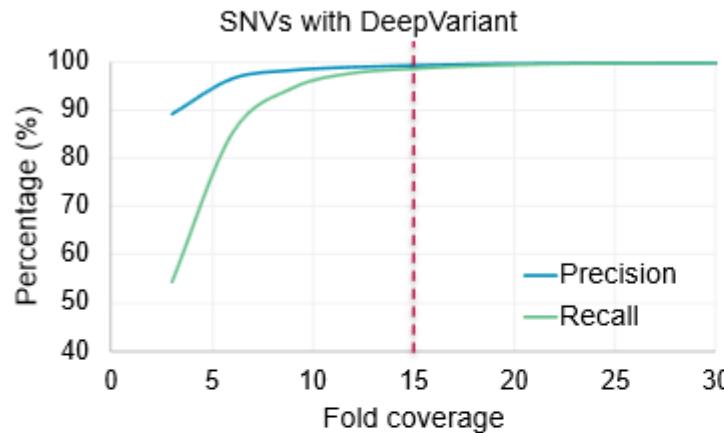


Variant calls from ~15-fold HiFi read coverage of a human genome (HG002) were measured against the Genome in a Bottle small variant benchmark (v3.3.2) for SNVs and indels using Deep Variant and SMRT Link 8.0 for SVs. Libraries were generated using a 15 kb insert and sequenced using Chemistry 2.0.

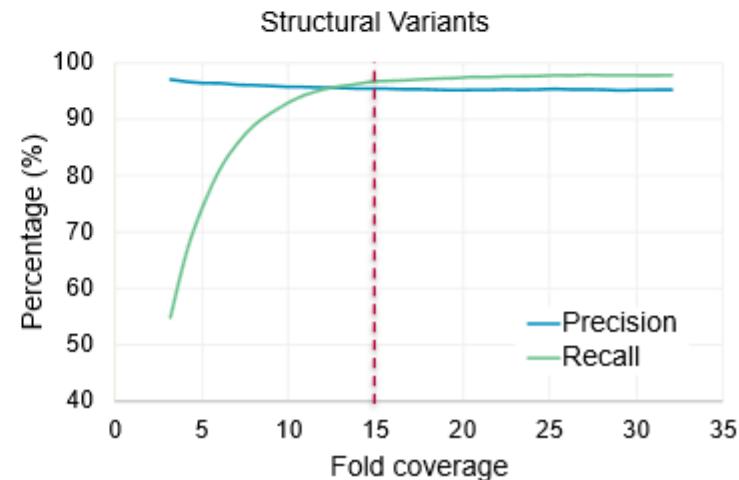
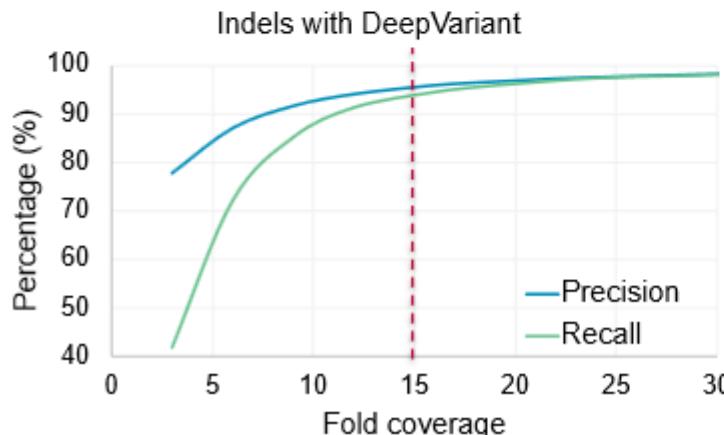
# Variant calling:



## VARIANT DETECTION COVERAGE RECOMMENDATION



**15-fold HiFi ( $\geq Q20$ ) Coverage**  
[2 SMRT Cells 8M] provides a  
good trade-off between cost  
and results

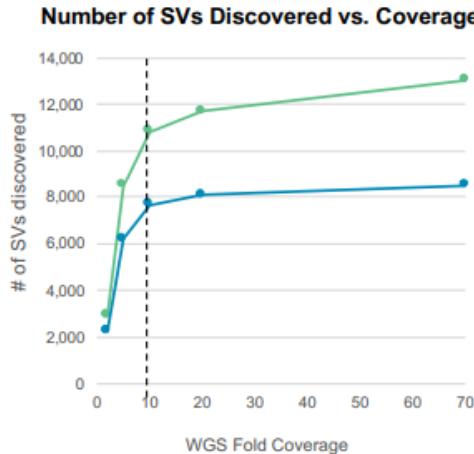


# Structural Variant Detection

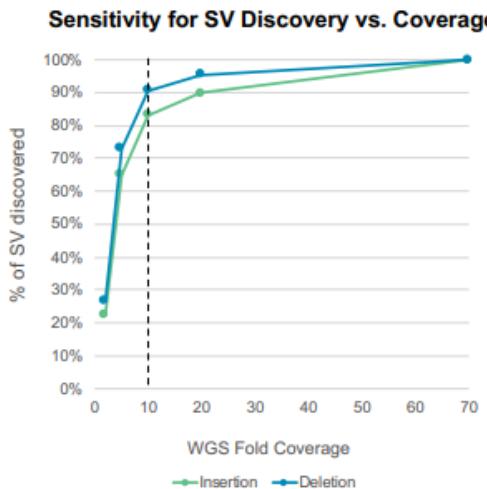
Sequence to desired coverage based on study needs:

- 5 to 10-fold: population genetics studies –  
sensitivity limited per individual, but high for variants shared in the population using joint calling
- 10-fold: rare undiagnosed disease studies –  
sensitivity high per individual allowing discovery of pathogenic SVs
- 10 to >20-fold: genetic disease studies –  
identify a variant or gene that causes disease in a cohort of individuals with a shared phenotype; higher coverage required for de novo SV detection in trios

## SV DISCOVERY POWER AT VARIOUS COVERAGE LEVELS

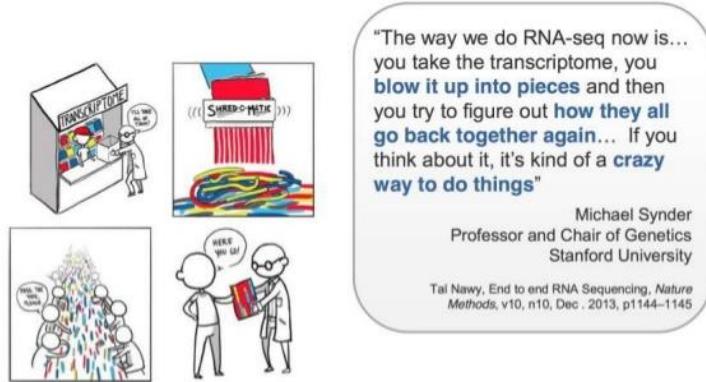


A diploid human (HG00733) was sequenced to 70-fold coverage on the Sequel System. The reads were randomly sampled to various coverage levels, and the SV calls at each coverage were evaluated against the calls at full 70-fold coverage.



Sensitivity increases sharply with coverage until about 10-fold, where it begins to level off. At 10-fold coverage, 10,854 insertions and 7,692 deletions are called (83% and 90.5% sensitivity, respectively)<sup>4</sup>.

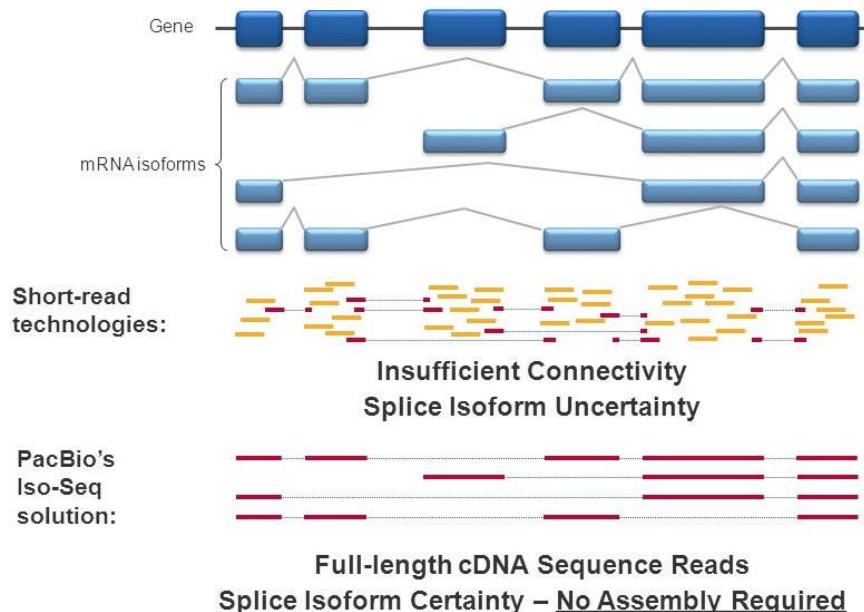
# Sequel II applications – RNA sequencing



Iso-Seq method generates high-quality, full-length transcripts – no assembly required.

Consider Iso-Seq if you need to transcriptome data for:

- Genome annotation
- Isoform discovery
- Fusion gene detection
- Creating *de novo* reference transcripts for RNA-seq quantification



# Sequel II applications – RNA sequencing

## DATA ANALYSIS SOLUTIONS WITH THE PACBIO ANALYTICAL PORTFOLIO

- Generate highly accurate long reads (HiFi reads), with single-molecule resolution using circular consensus sequencing (CCS) mode
- Use the Iso-Seq analysis in SMRT Link to output high-quality, full-length transcript FASTA sequences, with no assembly required, to characterize transcripts and splice variants<sup>3,4</sup>
- Run Iso-Seq analysis with or without a reference genome, and annotate the genome using community tools such as SQANTI<sup>5</sup>, TAMA<sup>6</sup>, and LoReAn<sup>7</sup>



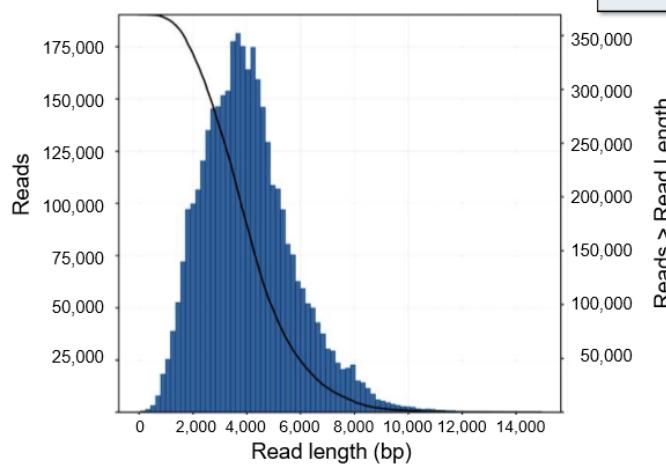
# Sequel II applications – RNA sequencing

RNA Sequencing	
Iso-Seq Method	Single-Cell Iso-Seq
Characterize alternative splicing/annotate a genome with full length transcripts	Characterize alternative splicing with full length transcripts up to 3M full length reads with cell barcode and UMI information
One human transcriptome per SMRT Cell 8M	1,000 unique reads/ single cell for 3000 cells 10,000 unique reads/ single cell for 300 cells
<2 kb to >3 kb	<2 kb to >3 kb
300 ng total RNA for 1st Strand cDNA Synthesis	>160 ng cDNA AFTER reamplification
The protocol supports up to 12 barcodes available.	Detects cell barcodes and UMIs
CCS	CCS



## ISO-SEQ PERFORMANCE

- Iso-Seq: Universal Human Reference RNA



Data shown above from a Universal Human Reference RNA (human) and Lexogen SIRV spike-in controls. The library was constructed using the Iso-Seq Express workflow including the SMRTbell Express Template Prep Kit 2.0 on a Sequel II System (Sequel II Sequencing kit 2.0, Sequel II Binding Kit 2.0, Sequel II System Software v8.0, 24-hour movie). Read lengths, reads/data per SMRT Cell 8M and other sequencing performance results vary based on sample quality/type and insert size.

Metrics	
Number of Raw Bases (Gb)	502
Total Reads	5,433,706
Full Length Non-chimeric Reads	3,693,801 (86%)
CCS Passes (Mean)	8

# Sequel II applications – metagenomics

Application	Metagenomics	
	Full-length 16S rRNA Sequencing	Shotgun Metagenomic Profiling or Assembly
With 1 SMRT Cell 8M you can:	Multiplex up to 96 samples to provide strain level resolution	Generate near-complete assemblies of high-complexity sample(s) (e.g. gut microbiome)
Minimum Recommended Coverage	8,000 reads/sample	See Best practices guide
Library Insert Size	1 - 2 kb	10 kb
Minimum Input Amount	500 ng - 1 µg	1.5 µg
Multiplexing/SMRT Cell	Up to 96 samples/ SMRT Cell 8M Up to 12 samples/ SMRT Cell 1M	Profile up to 4 communities/ SMRT Cell 8M Profile one community/ SMRT Cell 1M
Sequencing Mode	CCS	CCS

Depending on coverage needs, up to 96 libraries can be pooled for one MiSeq run. For metagenomics **samples**, >100,000 **reads per sample** is sufficient to fully survey the bacterial composition. This **number of reads** allows for **sample** pooling to the maximum level of 96 libraries, given the MiSeq output of > 20 million **reads**.

[support.illumina.com](http://support.illumina.com) > 16s-metagenomic-library-prep-guide-15044223-b PDF

16S Sample Preparation Guide - Support Illumina

Price (16S):  
Sequel II: 22 635 NOK  
Illumina MiSeq (300 bp PE): 24 810 NOK

# Metagenomics: 16S rRNA

The ISME Journal (2016) 10, 2020–2032  
© 2016 International Society for Microbial Ecology All rights reserved 1751-7362/16  
[www.nature.com/ismej](http://www.nature.com/ismej)

OPEN

## ORIGINAL ARTICLE

### High-resolution phylogenetic microbial community profiling

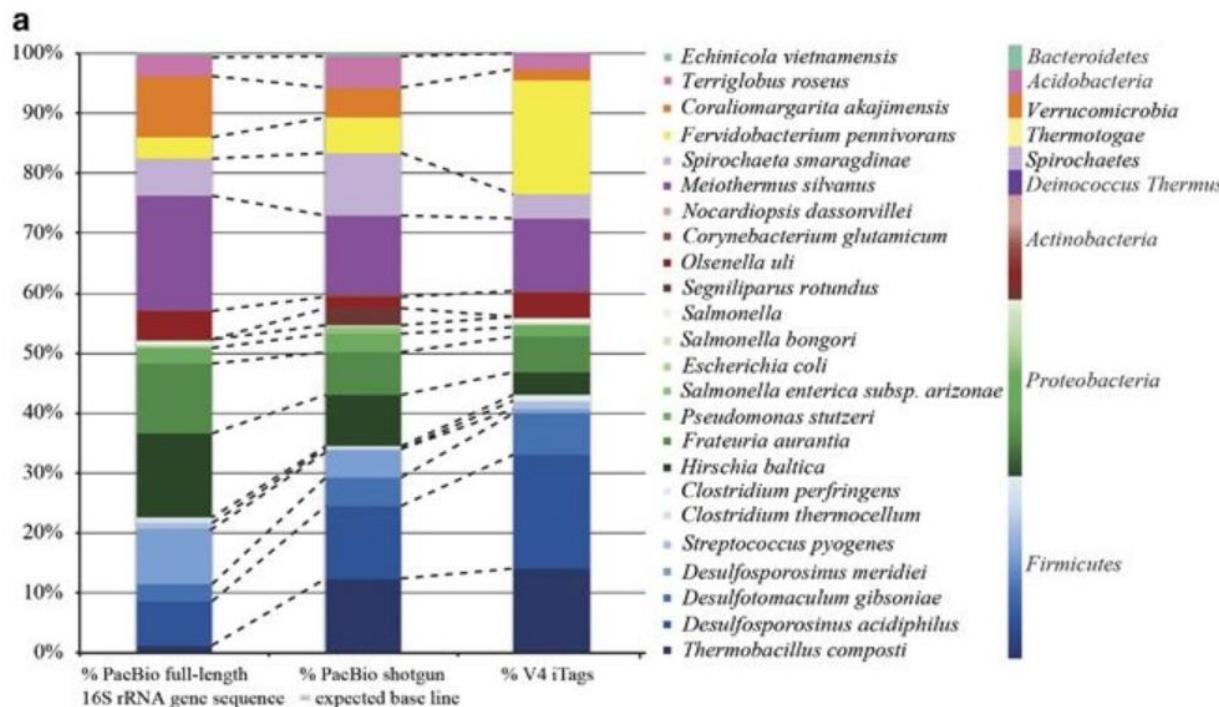
Esther Singer<sup>1</sup>, Brian Bushnell<sup>1</sup>, Devin Coleman-Derr<sup>1,2</sup>, Brett Bowman<sup>3</sup>, Robert M Bowers<sup>1</sup>, Asaf Levy<sup>3</sup>, Esther A Gies<sup>4</sup>, Jan-Fang Cheng<sup>1</sup>, Alex Copeland<sup>1</sup>, Hans-Peter Klenk<sup>5</sup>, Steven J Hallam<sup>4</sup>, Philip Hugenholtz<sup>6</sup>, Susannah G Tringe<sup>1</sup> and Tanja Woyke<sup>1</sup>

<sup>1</sup>US Department of Energy, Joint Genome Institute, Walnut Creek, CA, USA; <sup>2</sup>USDA-ARS, Albany, CA, USA;

<sup>3</sup>Pacific Biosciences, Menlo Park, CA, USA; <sup>4</sup>University of British Columbia, Vancouver, BC, Canada;

<sup>5</sup>Newcastle University, School of Biology, Newcastle upon Tyne, UK and <sup>6</sup>Australian Centre for Ecogenomics, School of Chemistry and Molecular Biosciences and Institute for Molecular Bioscience, The University of Queensland, St Lucia, QLD, Australia

- Study of **mock** community made up of 23 bacterial and 3 archaeal species and microbial community in Sakinaw Lake.
- **Conclusion:** Comparison with V4 iTag, using PacBio sequencing enables more accurate phylogenetic resolution of microbial communities and predictions on their metabolic potential.

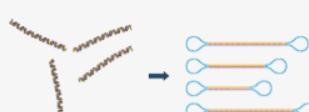


# Sequel II applications – targeted sequencing

Application	Targeted Sequencing	
	Amplicon Sequencing	No-Amp Targeted Sequencing
With 1 SMRT Cell 8M you can:	Sequence 384 barcoded amplicons	Sequence 5 targeted regions in a multiplex of 10 samples
Minimum Recommended Coverage	30-fold $\geq$ Q20 CCS read coverage for variant detection 6,000-fold $\geq$ Q20 CCS read coverage for minor variant detection (1% sensitivity)	$\geq$ 100-fold $\geq$ Q20 CCS read coverage per target locus
Library Insert Size	500 bp - 15 kb	4-6 kb or larger
Minimum Input Amount	250-500 ng for 250-1000 bp 500-1000 ng for 1-3 kb bp 1000-2000 ng for 3-10 kb 3000 ng for 15 kb	5 to 10 $\mu$ g (represented by either a single sample or the total of multiple samples that will be multiplexed)
Multiplexing/SMRT Cell	Up to 1,000+ samples/ SMRT Cell 8M or SMRT Cell 1M	Up to 10 samples/SMRT Cell
Sequencing Mode	CCS	CCS

When targeting >50 kb genomic regions – use probe-based capture using DNA oligo hybridization.  
 Protocols available for:

- IDT xGen Lockdown probes
- Nimbelgen SeqCap EZ



PCR 50 KB

IDT 300 KB

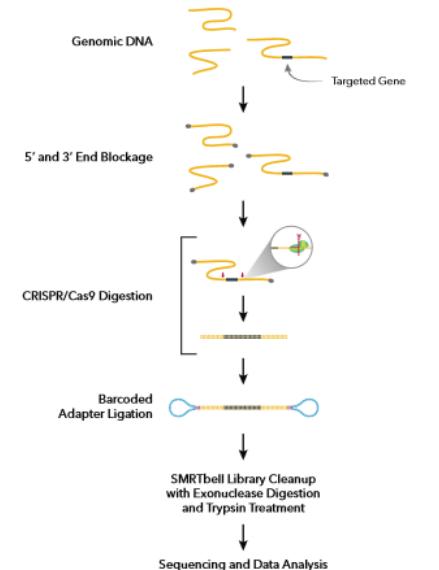
NimbleGen

50 MB

No-amplification targeted sequencing using CRISPR/Cas9 system:

- Challenging regions for PCR amplification (repeat expansions, low complexity regions)
- No PCR bias
- Preserves epigenetic modification signals

FROM gDNA TO COMPLETE REPEAT EXPANSION SEQUENCE



Sequence on Sequel® Systems



Analysis in SMRT Link using CCS



Visualization of Results

# PacBio sequencing is available at CEES/IBV/Uo



Meet the Award-Winning  
Sequel II System

[cees-drift@sequencing.uio.no](mailto:cees-drift@sequencing.uio.no)

NSC team at CEES: Ave Tooming-Klunderud

Morten Skage

Spyros Kollias