

Introduction to HiFi Pacbio long reads in De novo assembly, Mapping, and Variants calling

Presenters: Phương Thảo, Mạnh Hùng

10 Nov 2024

De novo assembly and mapping with Pacbio Hifi reads

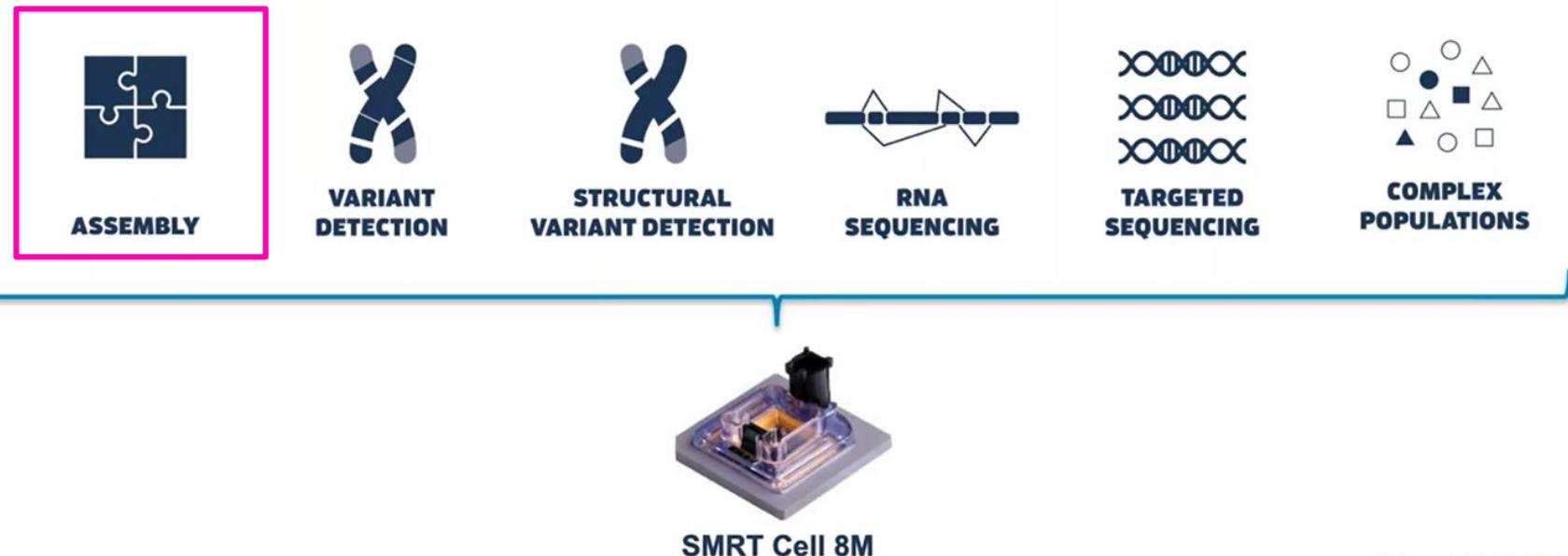
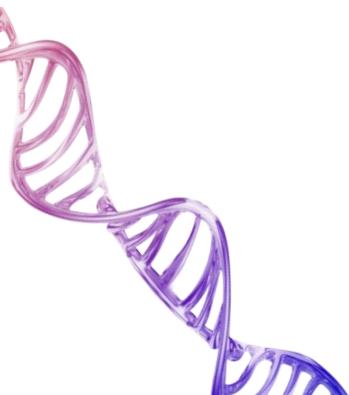


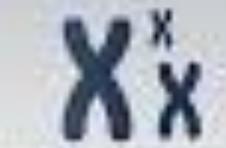
Table of contents

- 1. Pacbio HiFi (CCS) vs CLR long reads and Illumina short reads**
- 2. Advantages and limitations of Pacbio HiFi reads for genome assembly**
 - 2.1. Advantages**
 - High quality genome assembly/high performance metrics (correctness, completeness, contiguity)
 - Computational time and resources saving
 - Phase variants into haplotypes
 - 2.2. Limitation (typical errors of HiFi reads)**
- 1. Case study: Combination of long- and short-read sequencing fully resolves complex repeats of herpes simplex virus 2 strain MS complete genome**
 - 3.1. Quick overview of Herpes simplex virus 2 (HSV-2) genome**
 - 3.2. Code and results analysis**

Pacbio long reads

HiFi (CCS) vs Continuous Long Reads (CLR)





WHOLE GENOME
SEQUENCING



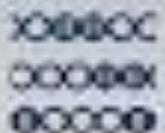
VARIANT
DETECTION



RNA
SEQUENCING



COMPLEX
POPULATIONS



TARGETED
SEQUENCING

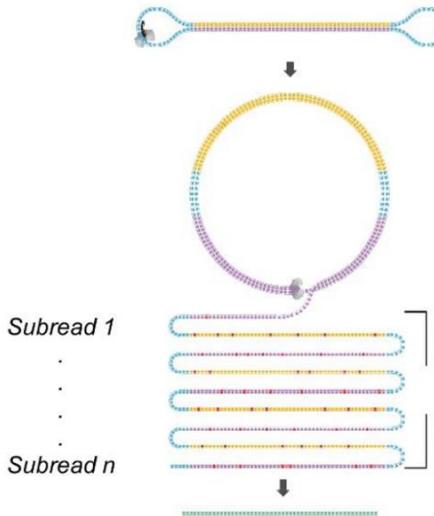


EPIGENETICS



Circular Consensus Sequencing (CCS) Mode

Inserts 10-20 kb



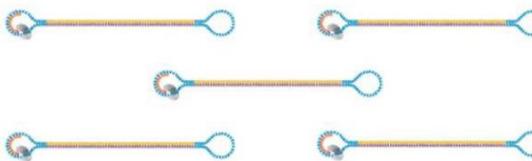
Single-molecule consensus sequence

99,98%
accuracy

Single molecule,
multiple reads

Continuous Long Read (CLR) Sequencing Mode

Inserts >25 kb, up to 175 kb



CLR 1

...

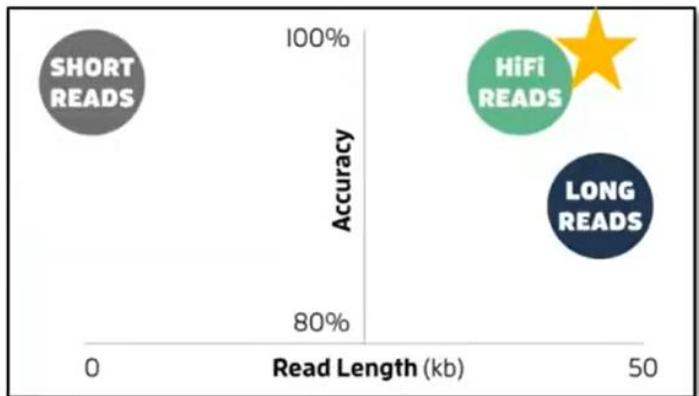
CLR n

Multi-molecule consensus sequence

89% accuracy

Multiple molecule,
single reads

HIFI READS ARE LONG AND ACCURATE



	Read length	Read accuracy	Genome characterization
NGS	300 bp	99.9%	single nucleotide variant, indel
PacBio CLR	>20 kb	89.0%	structural variant, assembly
PacBio HiFi	10-20 kb	99.8%	comprehensive

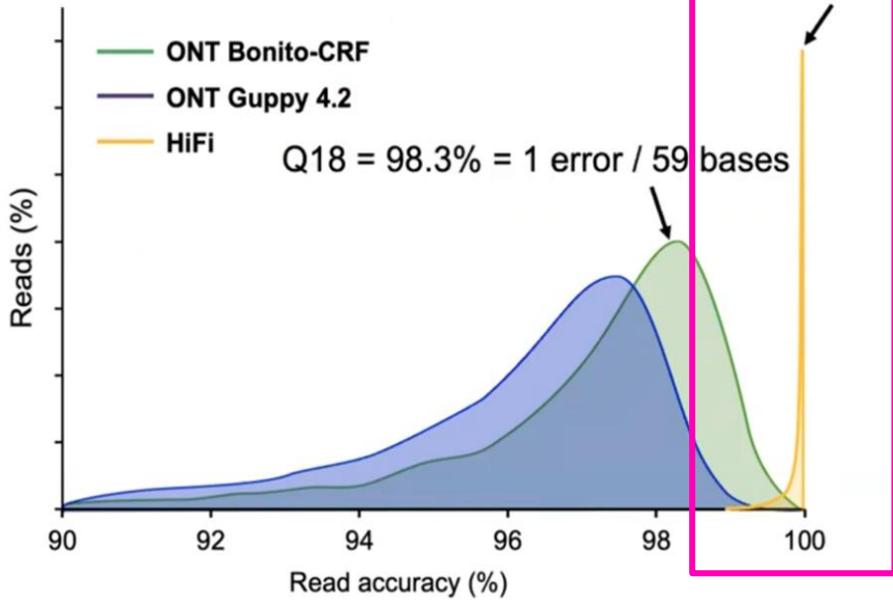
NGS



PacBio CLR

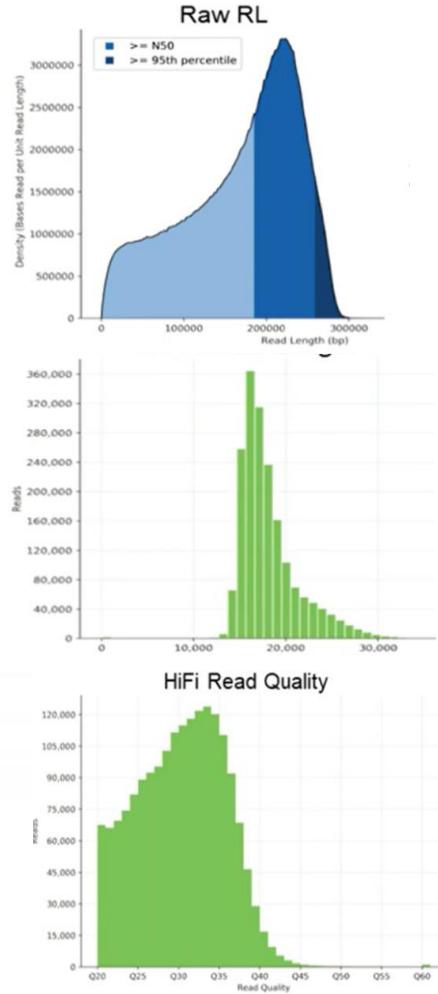
PacBio CCS





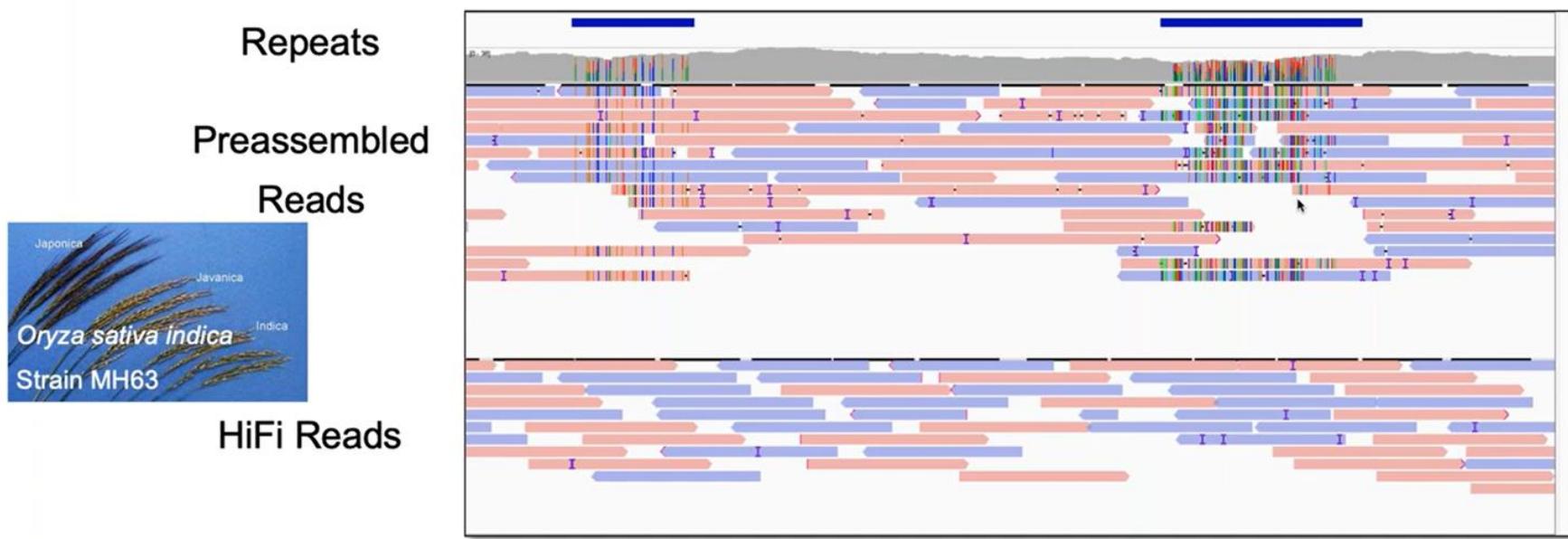
HiFi: HG003 18 kb library, Sequel II System Chemistry 2.0, [precisionFDA Truth Challenge V2](#)
 ONT: Bonito-CRF & Guppy 4.2 [NCM Nanopore Tech Update Dec. 2020](#)

$Q33 = 99.95\% = 1 \text{ error} / 2,000 \text{ bases}$
(34x fewer errors)

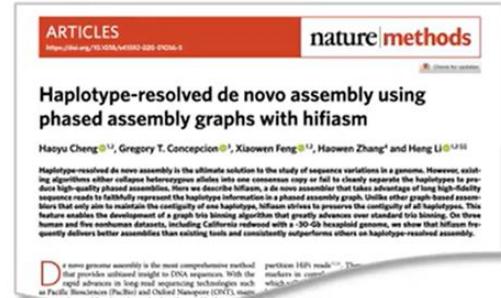
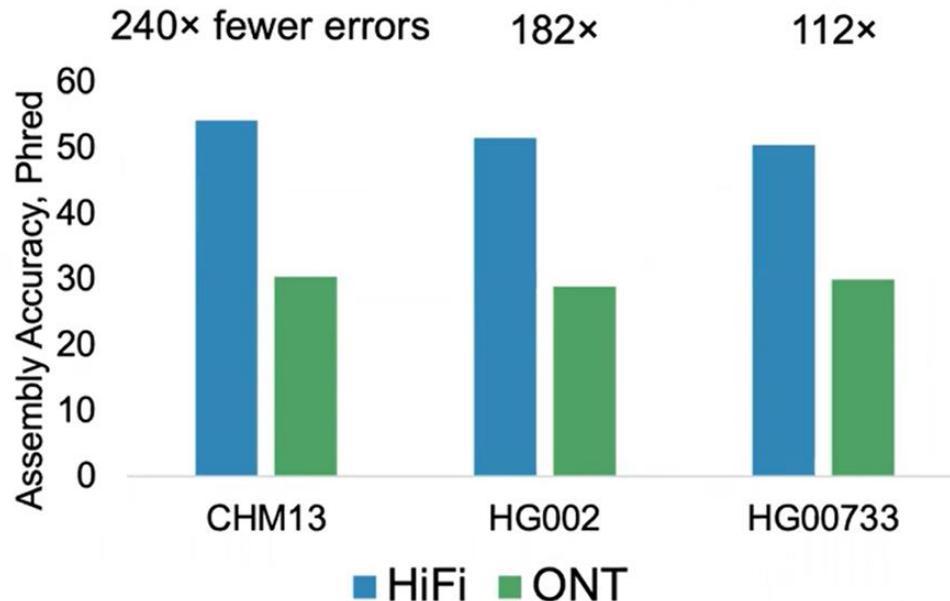


ERROR CORRECTION WITH CCS IS MORE ACCURATE

- HiFi reads use **single-molecule consensus**
- CLR error correction (preassembly) requires **multi-molecule consensus**



Hifi reads provide very high accurate denovo assembly



HiFi: hifiasm <https://doi.org/10.1038/s41592-020-01056-5>
ONT: Shasta <https://doi.org/10.1038/s41587-020-0503-6>

HIFI VERSUS CLR IN HUMAN

- HiFi is as contiguous as CLR
- HiFi is more accurate than CLR

	CHM13 (Vollger et al. 2019)	
Data Type	CLR	HiFi
Coverage	77-fold	24-fold
Contig N50	29.2	29.5
Median Base QV	40.7	45.0
Method	FALCON, Arrow	Canu, Racon

The bioRxiv preprint page for the CHM13 assembly. It shows the title "Improved assembly and variant detection of a haploid human genome using single-molecule, high-fidelity long reads", authors Mitchell R. Vollger et al., and a DOI link. The page includes a "Comment on this paper" section and a "Conformational Results" section.

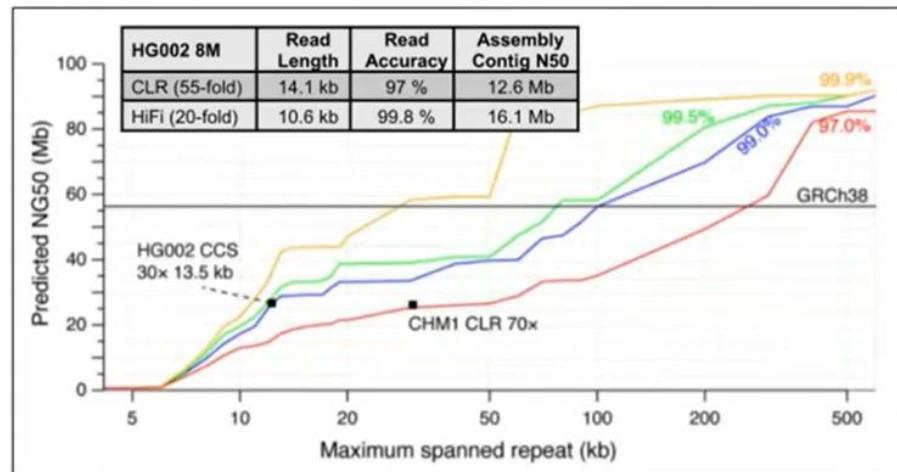
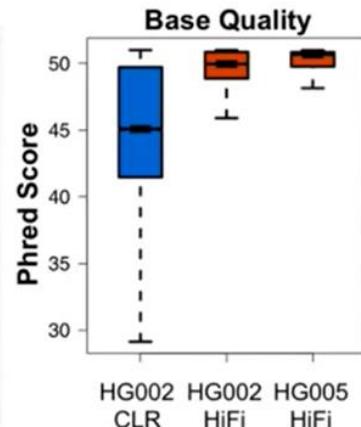
HG002 Data:

<https://bit.ly/2RW1b3I>

HG005 Data:

[PRJNA540706](https://www.ncbi.nlm.nih.gov/bioproject/PRJNA540706)

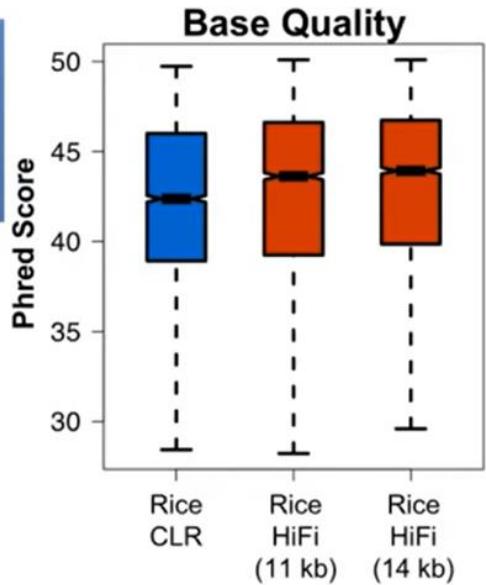
The bioRxiv preprint page for "Highly-accurate long-read sequencing improves variant detection and assembly of a human genome". It lists authors Aaron M. Wenger et al., and a DOI link. The page features a search bar and navigation links for HOME, ABOUT, SUBMIT, ALERTS / RSS, and CHANNELS.



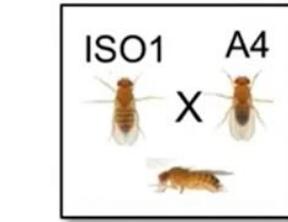
PLANT AND ANIMAL GENOMES: BASE QUALITY



CLR
30 kb, 60-fold
HiFi
11 kb, 30-fold
14 kb, 39-fold

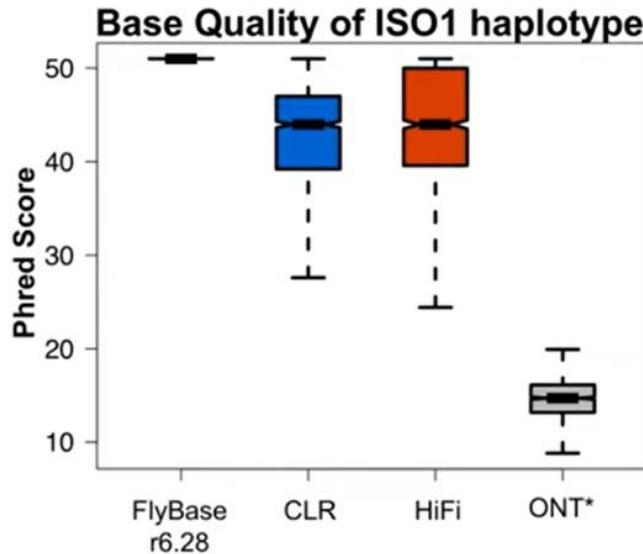


Drosophila melanogaster



Collaboration with J.J. Emerson
and Mahul Chakraborty

CLR
15 kb, 100-fold
HiFi
12 kb, 30-fold



Percent Embryophyta BUSCO Genes (N=1440)			
Complete	98.7	98.3	98.4
Single Copy	98.1	97.7	97.9
Duplicated	0.6	0.6	0.5
Fragmented	0.1	0.3	0.3
Missing	1.2	1.4	1.3

Percent Diptera BUSCO Genes (N=2799)

	Complete	98.6	98.8	98.6	97.1
Single Copy	98.1	98.0	98.0	96.7	
Duplicated	0.5	0.8	0.6	0.4	
Fragmented	0.8	0.6	0.8	1.7	
Missing	0.6	0.6	0.6	1.2	

* Solares et al. 2018 G3

PLANT AND ANIMAL GENOME: CONTIGUITY

<i>D. melanogaster</i>	CLR	HiFi
Coverage	71-fold	39-fold
Insert N50	27 kb	12 kb
Assembly Size	162 Mb	145 Mb
Contig N50	3.46 Mb	5.27 Mb
Number of Contigs	211	288



Collaboration with
J.J. Emerson and
Mahul Chakraborty

Pacific Bluefin Tuna	CLR	HiFi
Coverage	~200-fold	55-fold
Insert N50	23 kb	12.5 kb
Assembly Size	944 Mb	832 Mb
Contig N50	4.59 Mb	11.4 Mb
Number of Contigs	751	742



Collaboration with
Barbara Block,
Nate Truelove,
Luke Gardner

RICE GENOME: CONTIGUITY

HiFi Modifications

- Strict overlap filter: 99.7 %
- Ignore Indels

<i>O. sativa</i>	CLR	HiFi	HiFi	HiFi	HiFi
Coverage	60-fold	30-fold	39-fold	30-fold	39-fold
Insert Size	30 kb	11 kb	14 kb	11 kb	14 kb
FALCON settings				Strict ovlp filt Ignore indels	Strict ovlp filt Ignore indels
Contig N50	11.2 Mb	1.5 Mb	3.3 Mb	6.2 Mb	6.9 Mb

RICE REPEAT STRUCTURE

~14% of rice genome is gypsy-like LTRs retrotransposons
Length range: 10-13 kb
Mean length 11.7 kb

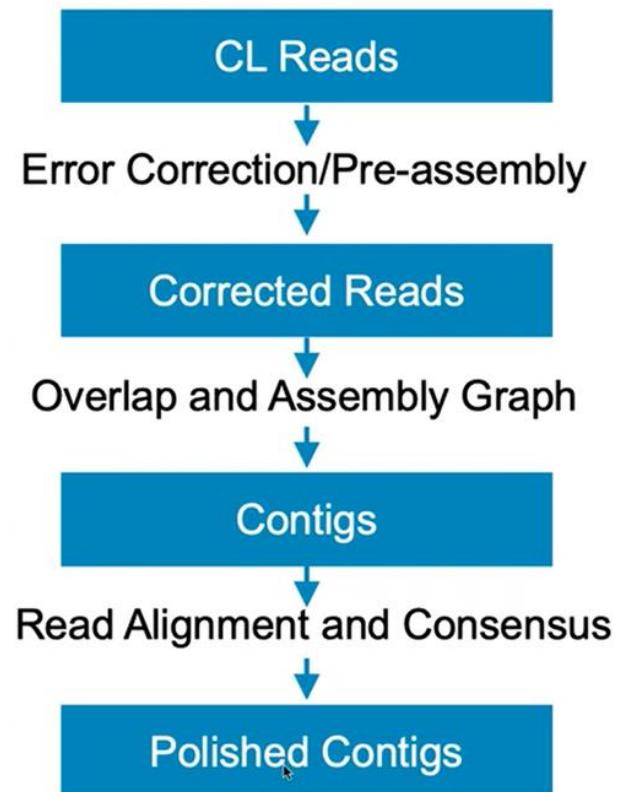
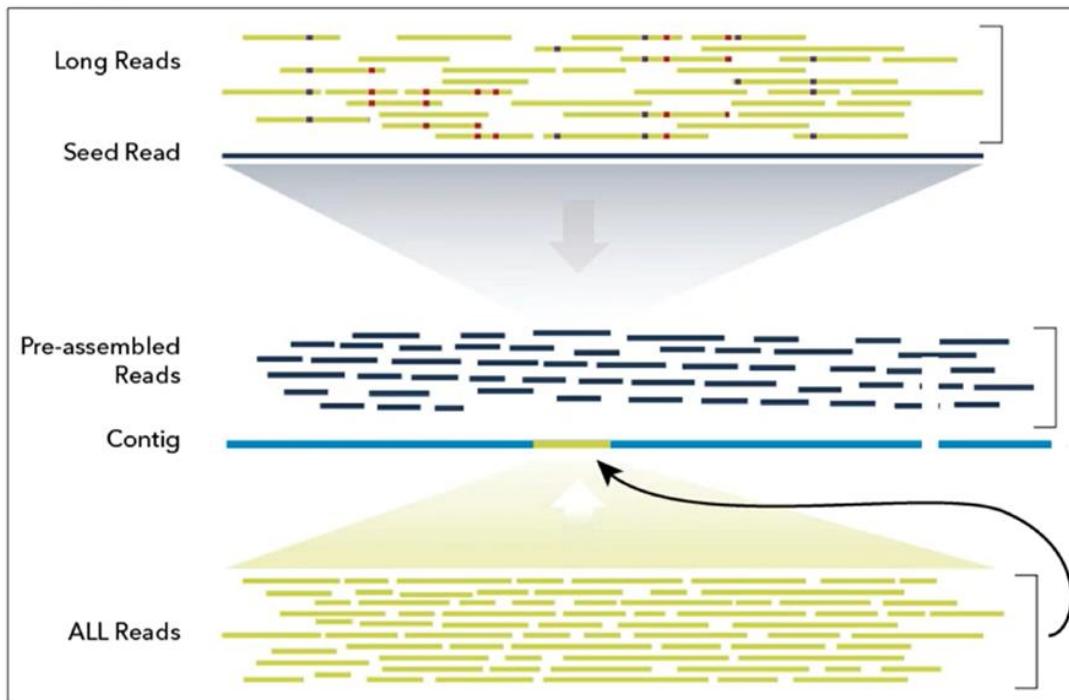
Repeats longer than 11kb

Family	Length	Copy Number
Osr28	18,005	<50
Osr41	15,655	300
Osr30	13,002	1500
Osr27	12,900	900
Osr34	12,797	450
Osr33	12,009	550
Osr40	11,421	600
Osr26	11,314	500

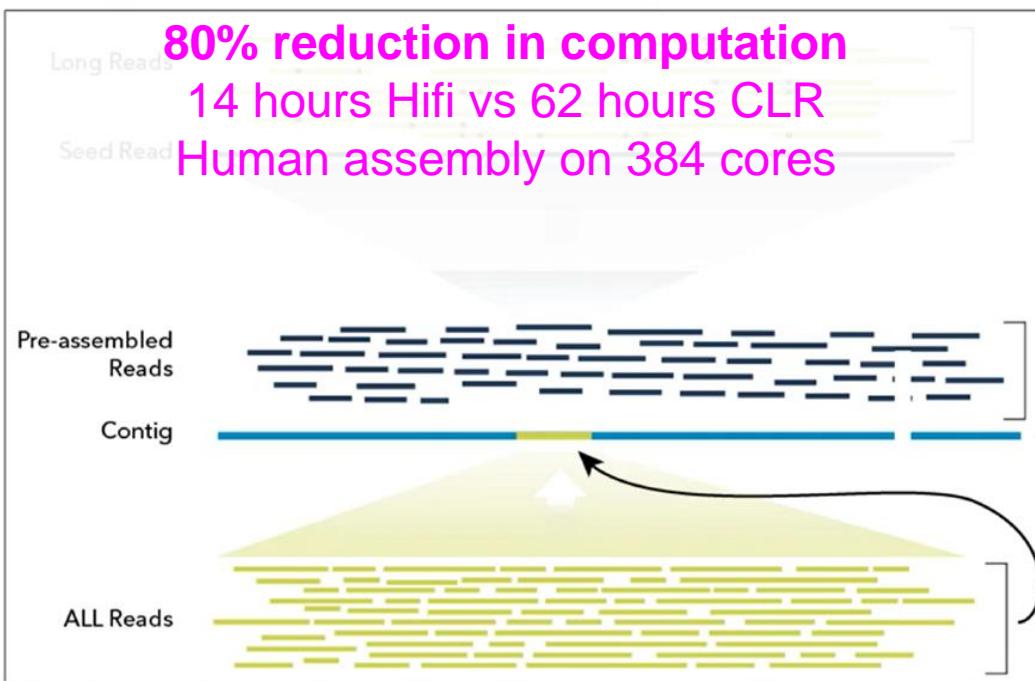


Library Size	17 kb	24 kb
1 SMRT Cell Yield	15 Gb	25 Gb
Coverage (400Mb)	38-fold	63-fold
FALCON Asm Length	403 Mb	405 Mb
# Contigs	209	211
Contig N50	14 Mb	20 Mb
# Chrom in 1 Contig	0	3

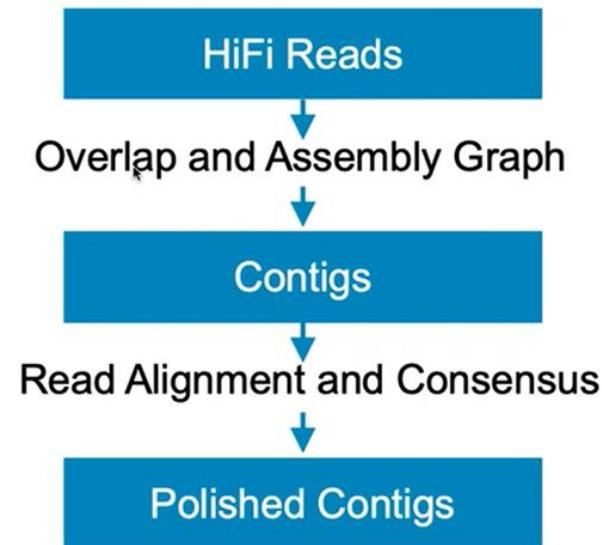
De novo assembly workflow with CLR (Falcon/Canu tools)



De novo assembly workflow with Hifi reads (Falcon/Canu)



(Saved time and computation)

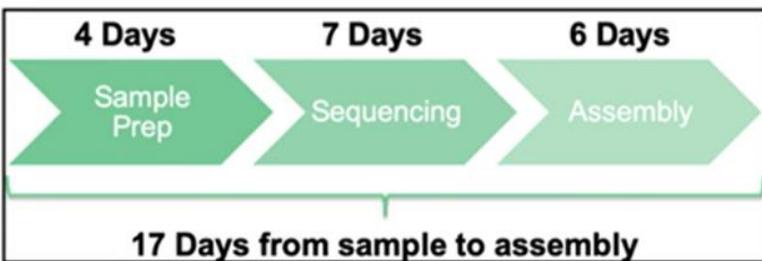


REDUCED ANALYSIS TIME WITH HIFI READS

Compute times for *de novo* assembly of a human genome:

Mode	HiFi	CLR	Ratio
Data Description	4 Cells (11kb ELF)	2 cells (20kb BP SS)	
Coverage	20-fold	50-fold	
File Type	fastq.bgz	subreads.bam	
Input File Size (GB)	44	323	14%
Error Correction Method	CCS Analysis	Preassembly	
CPU Hours: Error Correction	5,160	10,541	
CPU Hours: Assembly	1,164	2,630	
Total CPU hours	6,324	13,171	48%
Wall Hours: Error Correction	15.6	43.5	
Wall Hours: Assembly	13.7	18.9	
Total wall hours	29.3	62.4	47%
FALCON Dir Disk Usage	457 G	7.2 T	6%

EXAMPLE - CALIFORNIA REDWOOD GENOME PROJECT



Recent Conifer Genome Assembly Results			
Genome	California Redwood	California Redwood ¹	Douglas Fir ²
Methodology	PacBio HiFi	ONT + short reads	Short reads
Genome Coverage	22-fold	23-fold + 122-fold	61-fold
Assembly Size (Gb)	47.7	26.5	14.6
Contig N50 (Mb)	1.92	0.11	0.04
BUSCO Complete (embryophyta)	59%	56%	57%
Mapped transcripts with frameshift errors ³	0.12%	1.97%	32%
Assembly Time	6 days	5-6 months	-

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

2. The Douglas-Fir Genome Sequence Reveals Specialization of the Photosynthetic Apparatus in Pinaceae

3. Transcript set of *Abies alba* from Neale, D. et al. Varying number of transcripts aligned to each genome (4,958 mapped to PacBio HiFi redwood, 4,760 mapped to ONT redwood, 16,187 mapped to Douglas fir)

<https://medium.com/pacbio/a-genome-fit-for-a-giant-sequencing-the-california-redwood-ed722be9e49c>

<https://www.pacb.com/blog/tackling-a-giant-genome/>

HIFI ASSEMBLY USES LESS MEMORY: EXAMPLE HEXAPLOID WHEAT (12 GB GENOME)

—CLR Assembly

—Zimin et al. 2017

<https://doi.org/10.1093/gigascience/gix097>

—HiFi Assembly

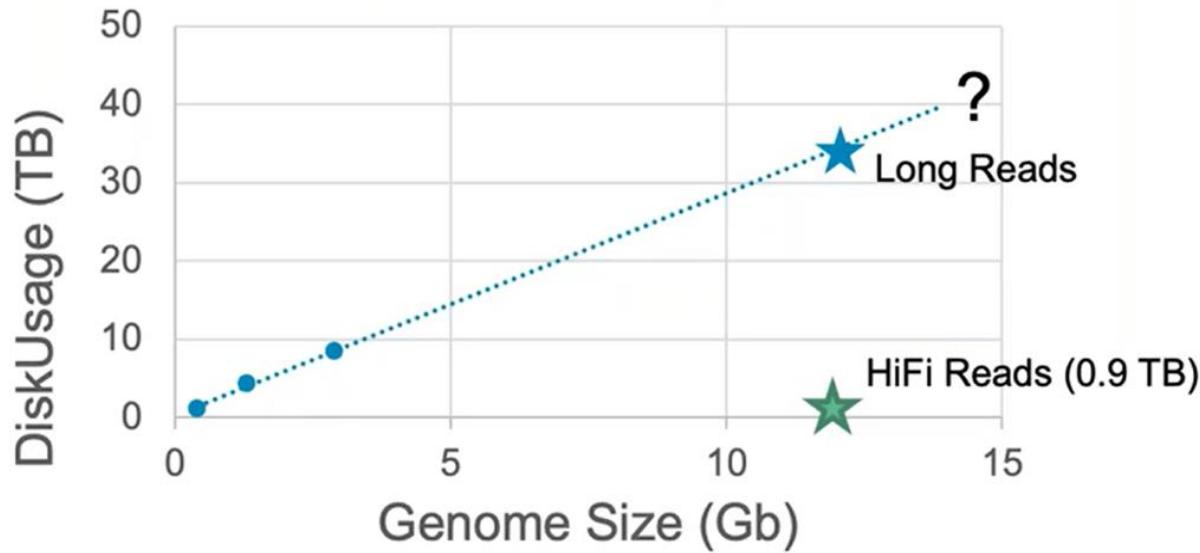
—Simulated HiFi reads using pbsim

—12-18 kb HiFi reads

—Error profile based on Sequel II v2.0
HiFi data

	Long Reads	HiFi Reads	
Coverage	32-fold	20-fold	
CPU Hours - Correction	116,000	21,000	
CPU Hours - Assembly	34,000	19,000	
Total CPU Hours	150,000	40,000	
Disk Usage	34 TB	0.9 TB	34x

DISK USAGE FOR LARGE GENOME ASSEMBLY



HUMAN DE NOVO ASSEMBLY COMPUTE AND STORAGE

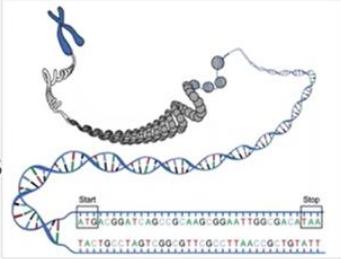
HG002 – 3 Gb Genome

Data Type	HiFi Reads	Long Reads	
Coverage	20-fold	50-fold	
Input File Type	CCS.FASTQ.GZ	SUBREADS.BAM	
Input File Size (GB)	48 (2 8M cells)	323 (3 8M cells)	
Read Correction Method	CCS Analysis	Pre-assembly	
Wall Hours	Read Correction	17.5	43.5
	Contig Assembly	13.7	18.9
	Total	31	62
CPU Hours	Read Correction	5,750	10,541
	Contig Assembly	1,164	2,630
	Total	6,914	13,171
Disk Usage (TB)	0.46	7.2	

2.5x
1.4x
2.0x
1.8x
2.3x
1.9x
15.7x!

Analyses run with PacBio recommended compute infrastructure
384 cores, 5 Gb RAM per slot

Key Advantages of HiFi Read over CL Read in de novo assembly

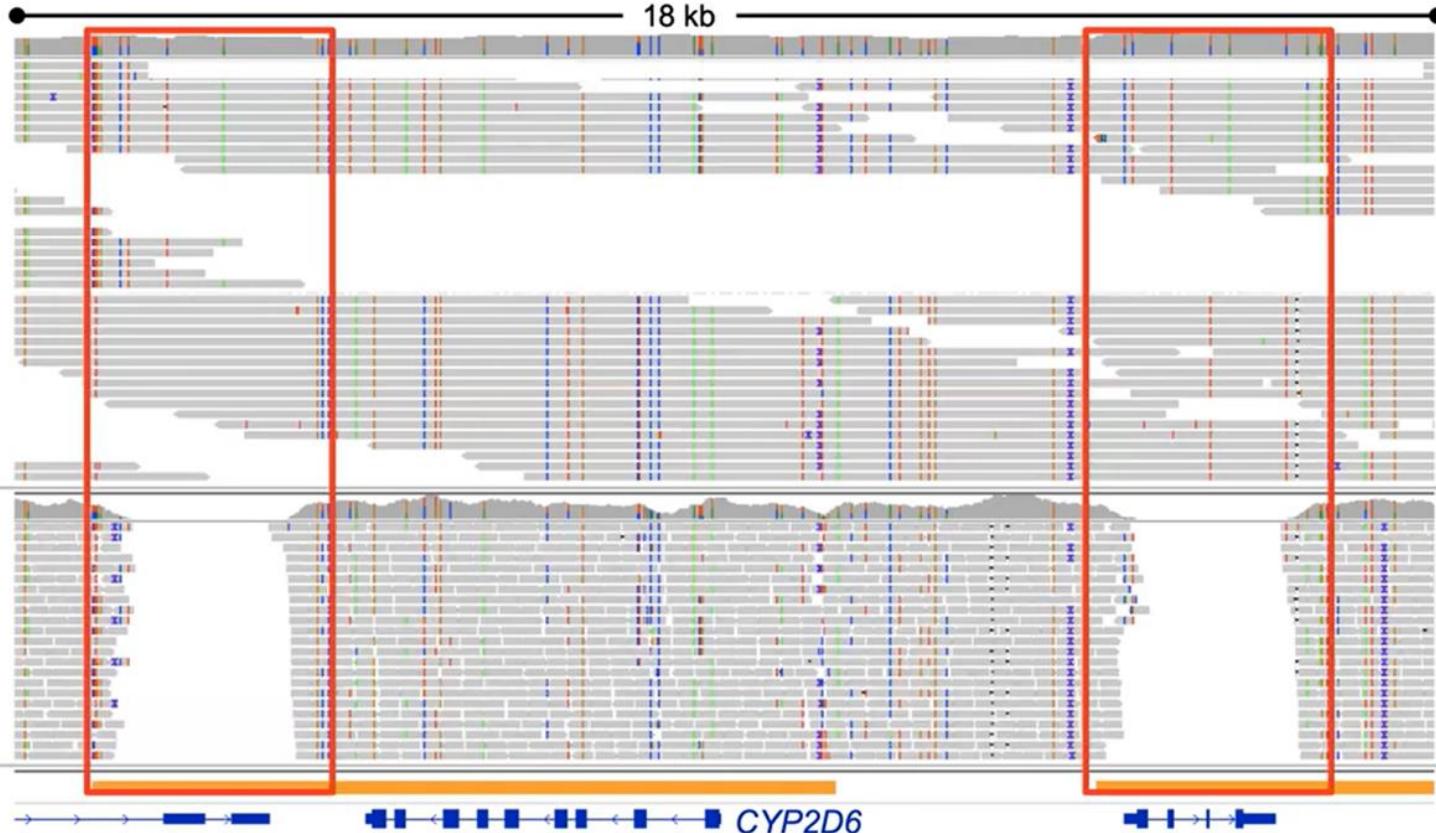
Contiguity <ul style="list-style-type: none">- Resolve Repetitive Regions- High Contig N50		Correctness <ul style="list-style-type: none">- Base QV- Phasing accuracy 
Completeness <ul style="list-style-type: none">- Gene Space- Repetitive Regions		Compute <ul style="list-style-type: none">- CPU / Wall Time- RAM- Disk Storage 

Pacbio HiFi (CCS) vs NGS (Illumina)



Human, HG002

18 kb

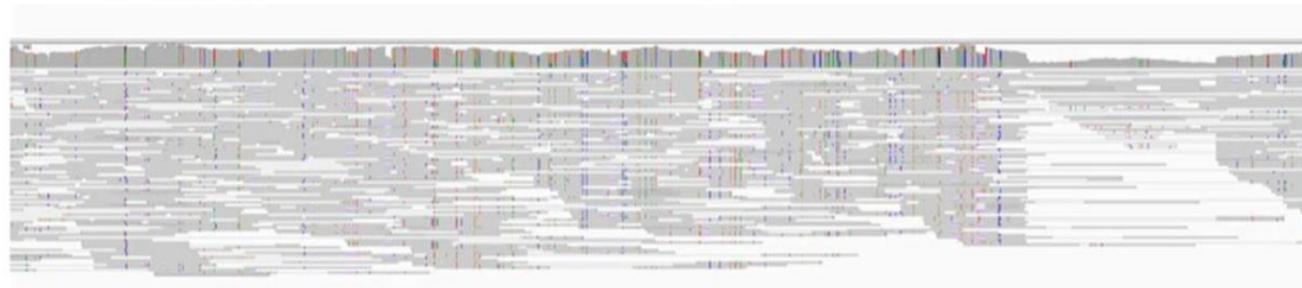


HiFi and short reads from Genome in a Bottle: <https://jimb.stanford.edu/giab>

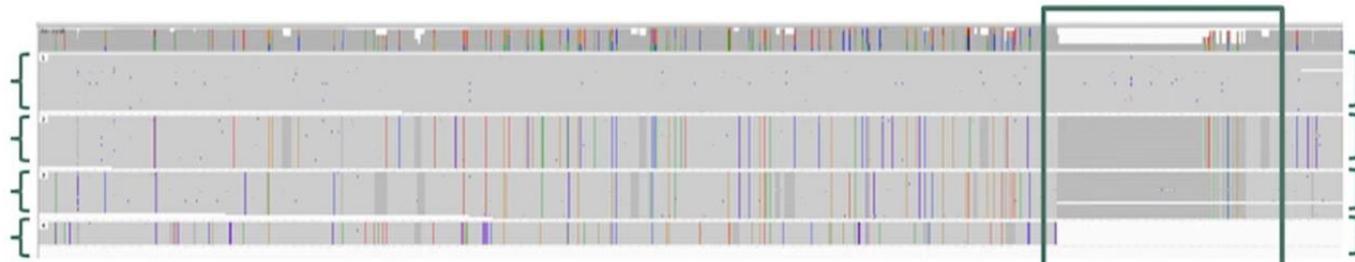
EXAMPLE - TETRAPLOID ROSE GENOME

HiFi long reads improves haplotyping

DH 'Old Blush' ref



Illumina



PacBio HiFi

ASSEMBLY RESULTS

	Primary Assembly	Haplotype
Assembly size	2.95 Gb	2.56 Gb (87%)
# of contigs	1,638	15,451
Contig N50	28.97 Mb	0.360 Mb*
Accuracy†	QV 47.3	QV 48.5
Ensembl transcripts mapped against GRCh38 (n=36,694)‡	C:98.3% [S:98.0%,D:0.3%] F:0.3%,M:1.4%	C:80.9% [S:80.8%,D:0.1%] F:6.5%,M:12.6%
BUSCO (mammalia, n=4,104)§	C:95.0% [S:94.1%,D:0.9%] F:2.3%,M:2.7%	

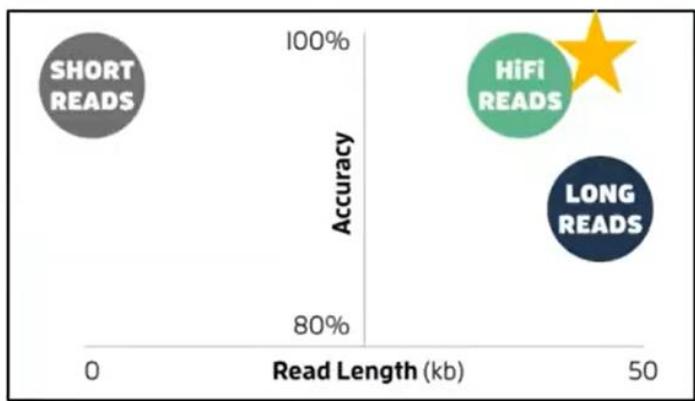
*Haplotype N50 ≈ Phase block N50
(since haplotypes are fully phased)

† Using k-mer analyzer yak (<https://github.com/lh3/yak>) against Seq2500 reads from HGSCV (ftp://ftp.1000genomes.ebi.ac.uk/vol1/ftp/data_collections/hgsv_sv_discovery/illumina_wgs.sequence.index)

‡ Mapping Ensembl r94 cDNA file (ftp://ftp.ensembl.org/pub/release-94/fasta/homo_sapiens/cdna/) to GRCh38 and compare against assembly (Heng Li, personal communication)

§ BUSCO against GRCh38 for reference: C:94.6% [S:91.3%,D:3.3%] F:2.5%,M:2.9%

HIFI READS ARE LONG AND ACCURATE



	Read length	Read accuracy	Genome characterization
NGS	300 bp	99.9%	single nucleotide variant, indel
PacBio CLR	>20 kb	89.0%	structural variant, assembly
PacBio HiFi	10-20 kb	99.8%	comprehensive

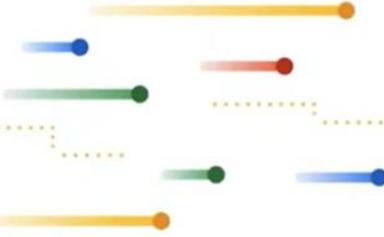
NGS



PacBio CLR

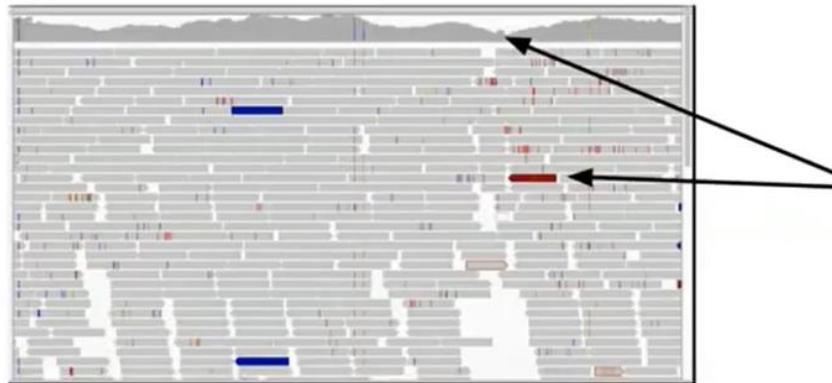
PacBio CCS





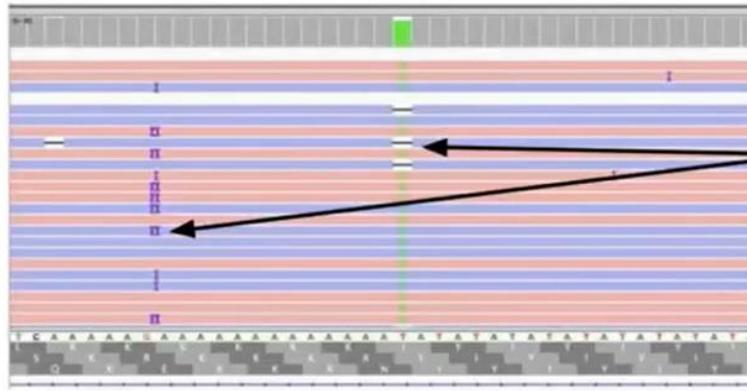
HiFi Errors are Different vs Short Reads

illumina®
(short reads)



Mapping complexity and coverage variability play larger roles in Illumina data.

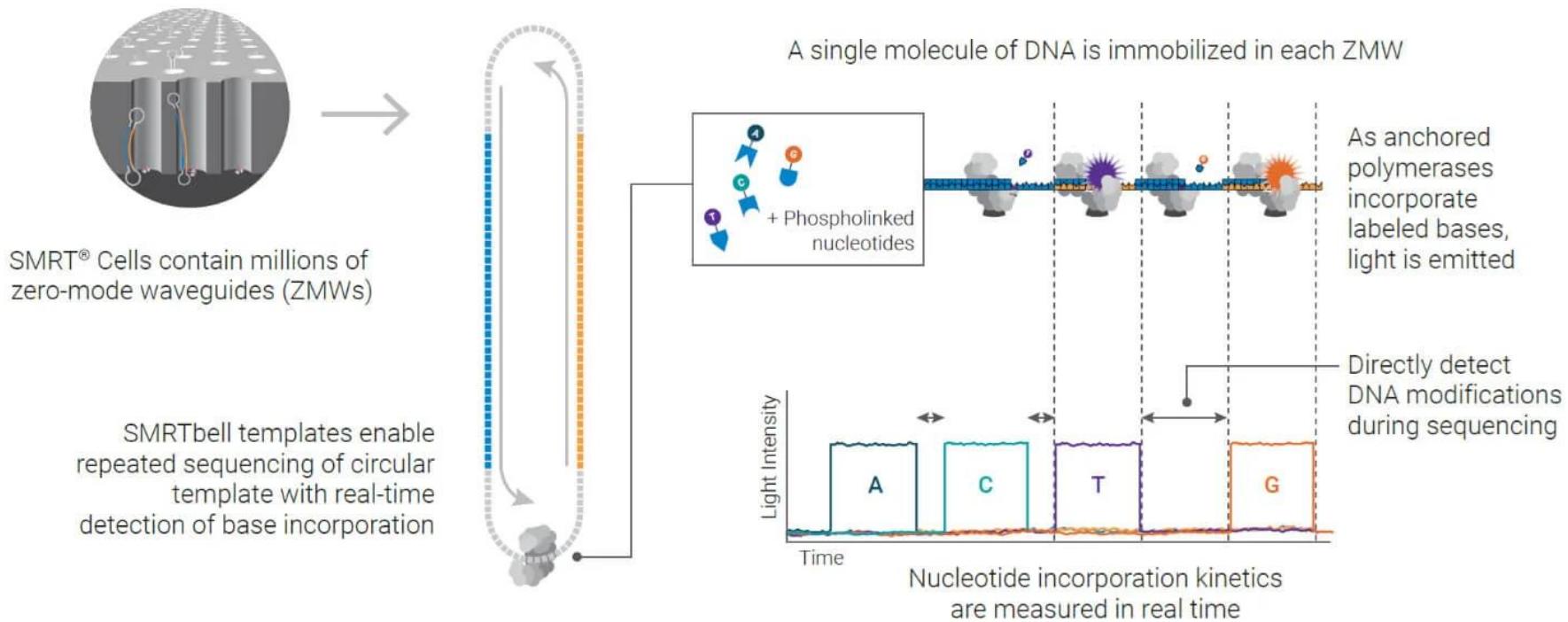
PacBio
(HiFi Reads)



Homopolymers have noisier indel lengths in HiFi data

(Ref is 13A[TA]x. Evidence in reads for 14A[TA]x, 15A[TA]x. DeepVariant correctly calls HET ALT 14A[TA]x)

Single Molecule Real-Time (SMRT) Sequencing Principle



Case study: Combination of long- and short-read sequencing fully resolves complex repeats of herpes simplex virus 2 strain MS complete genome

ebi.ac.uk/ena/browser/view/PRJEB40042

Tube Maps Gmail

ENA European Nucleotide Archive

Project: PRJEB40042

HSV-2 strain MS genome sequencing using both Illumina and PacBio technologies

Secondary Study Accession: ERP123638

Study Title: HSV-2 strain MS genome sequencing using both Illumina and PacBio technologies

Center Name: Universidad Autonoma de Madrid;CBMSO

Study Name: HSV-2 strain MS genome sequencing using both Illumina and PacBio technologies

ENA-FIRST-PUBLIC: 2020-09-23

ENA-LAST-UPDATE: 2020-08-26

General

- View: XML
- XML (STUDY)
- Download: XML
- XML (STUDY)
- Cross References: Show
- Publications: Show
- ORCID Data Claims: Show
- Related ENA Records: Show

Data

Read Files

Show Column Selection

Download report: JSON TSV

Get download script Download selected files

Sample Accession	Run Accession	Scientific Name	Generated FASTQ files: FTP	SRA files: FTP	Bam files: FTP
SAMEA5565540	ERR3278849	Human alphaherpesvirus 2	<input type="checkbox"/> ERR3278849_1.fastq.gz <input type="checkbox"/> ERR3278849_2.fastq.gz	<input type="checkbox"/> ERR3278849	N/A
SAMEA5565544	ERR3278853	Human alphaherpesvirus 2	<input type="checkbox"/> ERR3278853_subreads.fastq.gz	<input type="checkbox"/> ERR3278853	N/A

Illumina reads:

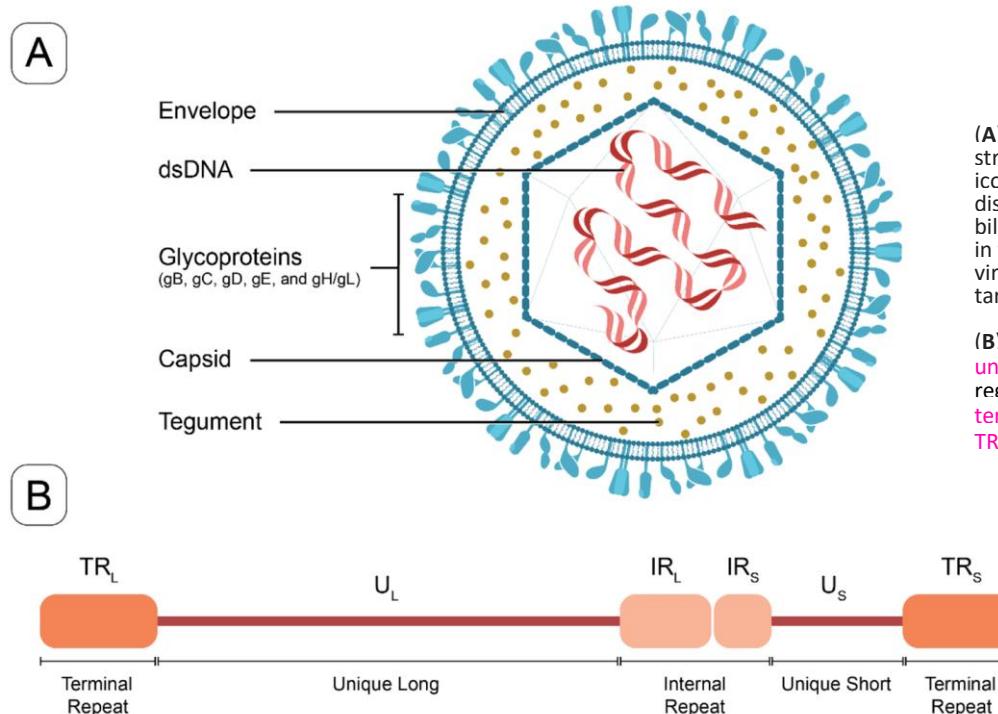
ERR 3278849_1.fastq.gz

and

ERR 3278849_2.fastq.gz

Pacbio reads:
ERR 3278853

The number of complete genomic sequences for HSV-2 isolates and strains is limited, making it insufficient for accurate evolutionary comparisons and investigations into the genomic determinants of virulence.

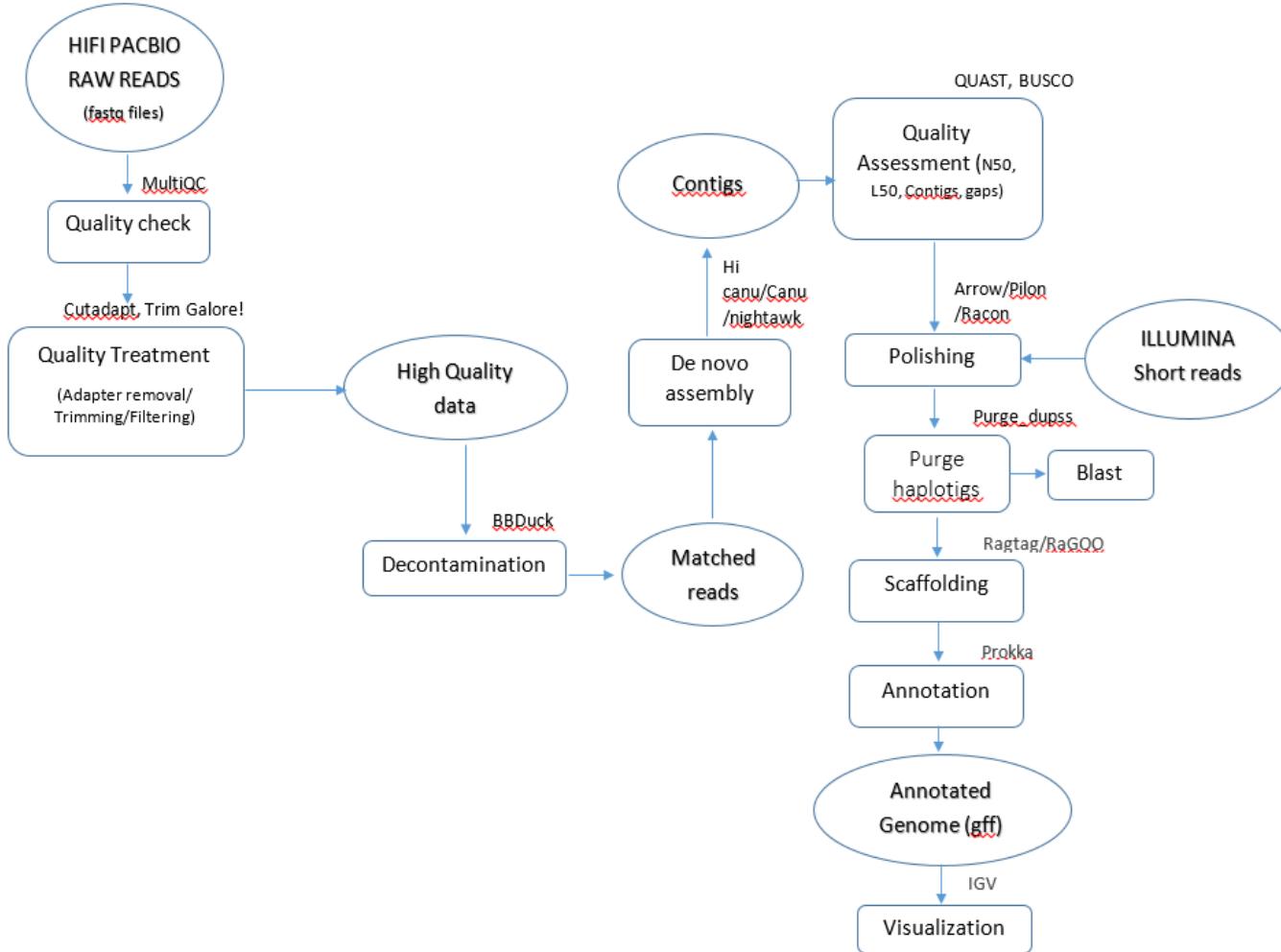


(A) The HSV virion consists of a linear double-stranded DNA genome enclosed within an icosahedral capsid. Tegument proteins are distributed between the capsid and the lipid bilayer envelope. Glycoproteins are embedded in the outer surface of the envelope, facilitating viral entry (key glycoproteins in vaccine targeting include gB, gC, gD, and gE).

(B) The HSV-1 genome schematic includes unique long (UL) and unique short (US) regions, which are flanked by internal or terminal repeats, (IRL and TRL) and (IRS and TRS)

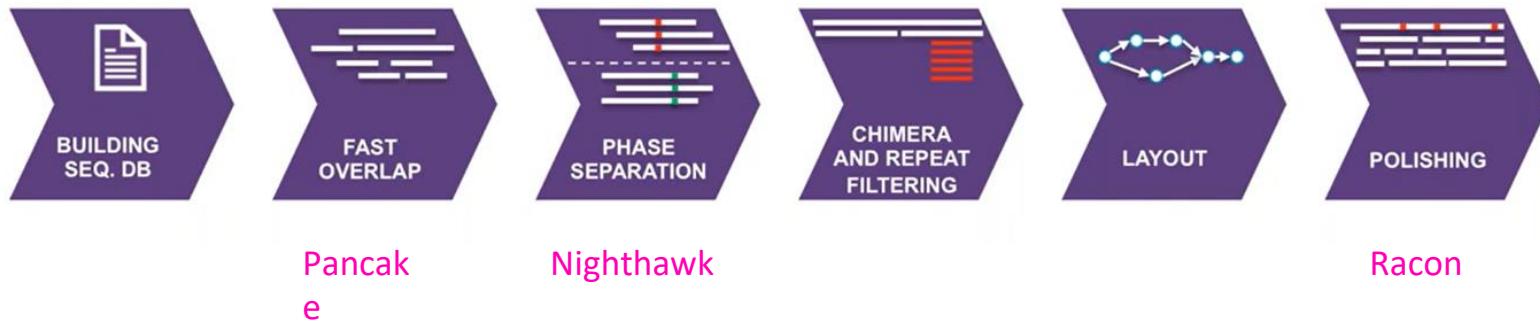
Variability in HSV genome size is mainly caused by the presence of variable nucleotide repeats, including microsatellites and tandem repetitions. The characterization of the flanking inverted repeats at both unique regions has been difficult because previous sequencing technologies could not resolve their length and sequence composition with accurate resolution.

De novo assembly and annotation HiFi reads pipeline



Improved Phased Assembly using HiFi data (IPA)

<https://github.com/PacificBiosciences/pipa>





Thanks for your attention!
(to be cont.)

Variants Calling with Pacbio Hifi reads

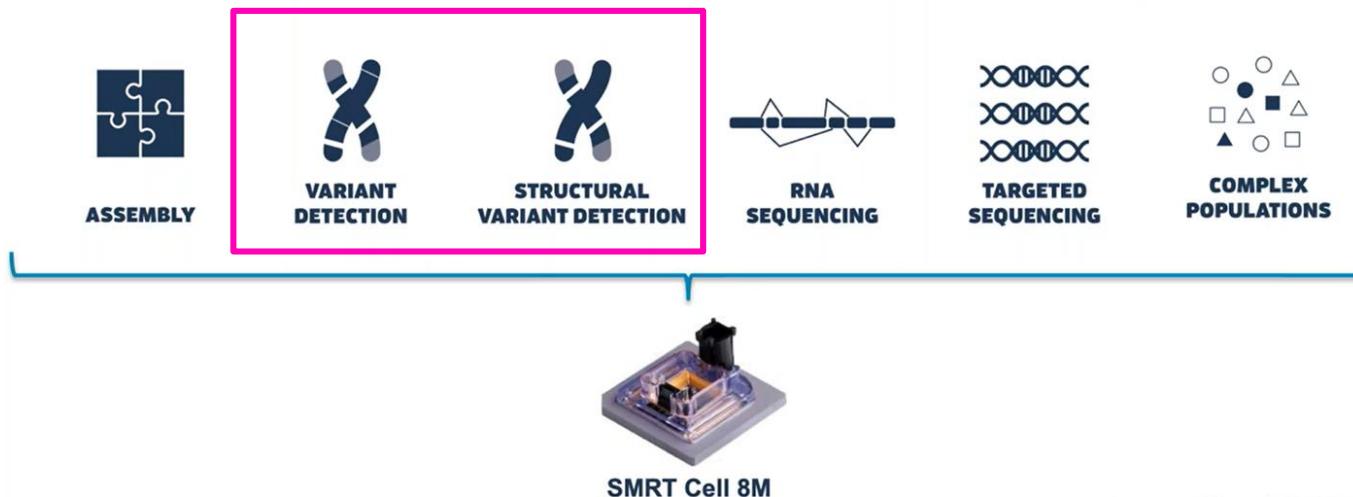


Table of contents

1. Advantages and limitations of Pacbio HiFi reads for variants calling
2. Case study: Combination of long- and short-read sequencing fully resolves complex repeats of herpes simplex virus 2 strain MS complete genome
3. Colab code and results analysing

TYPES OF GENOMIC VARIATION

Single Nucleotide Variant



Deletion



Insertion



Tandem Duplication



Interspersed Duplication



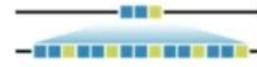
Inversion



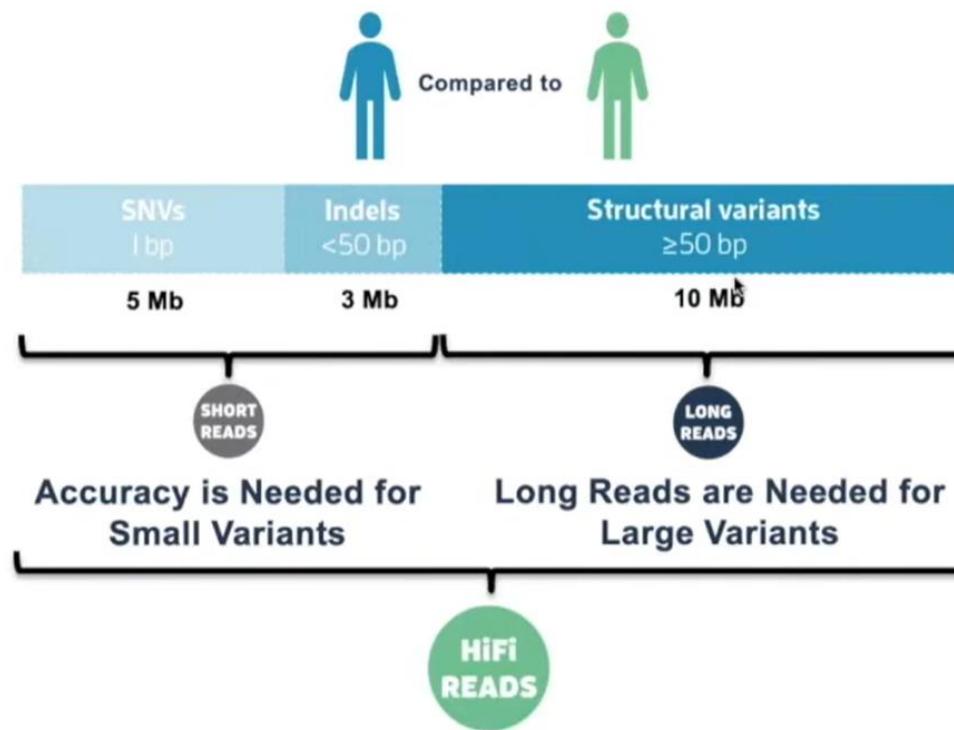
Translocation



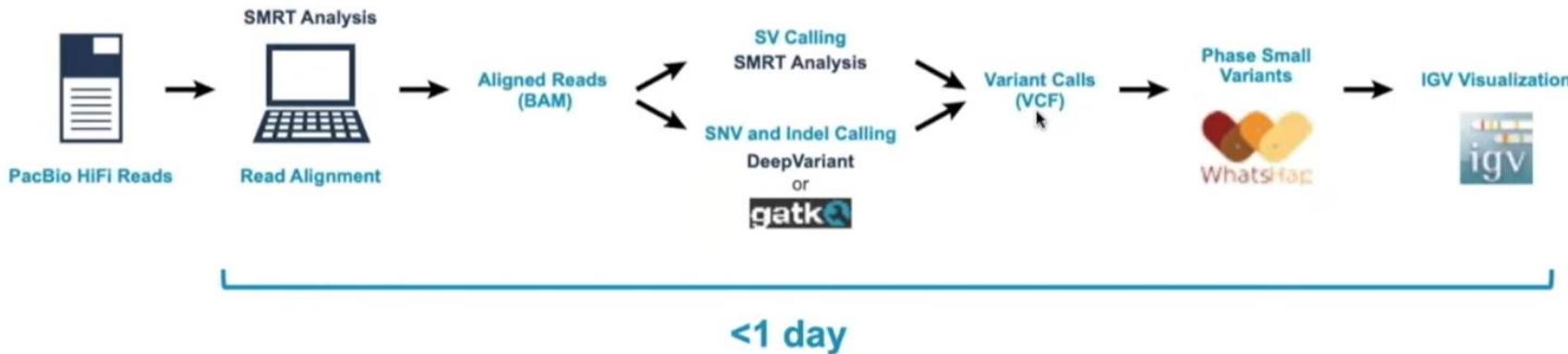
Copy Number Variant



GENETIC VARIATION OCCURS AT SMALL AND LARGE SCALES

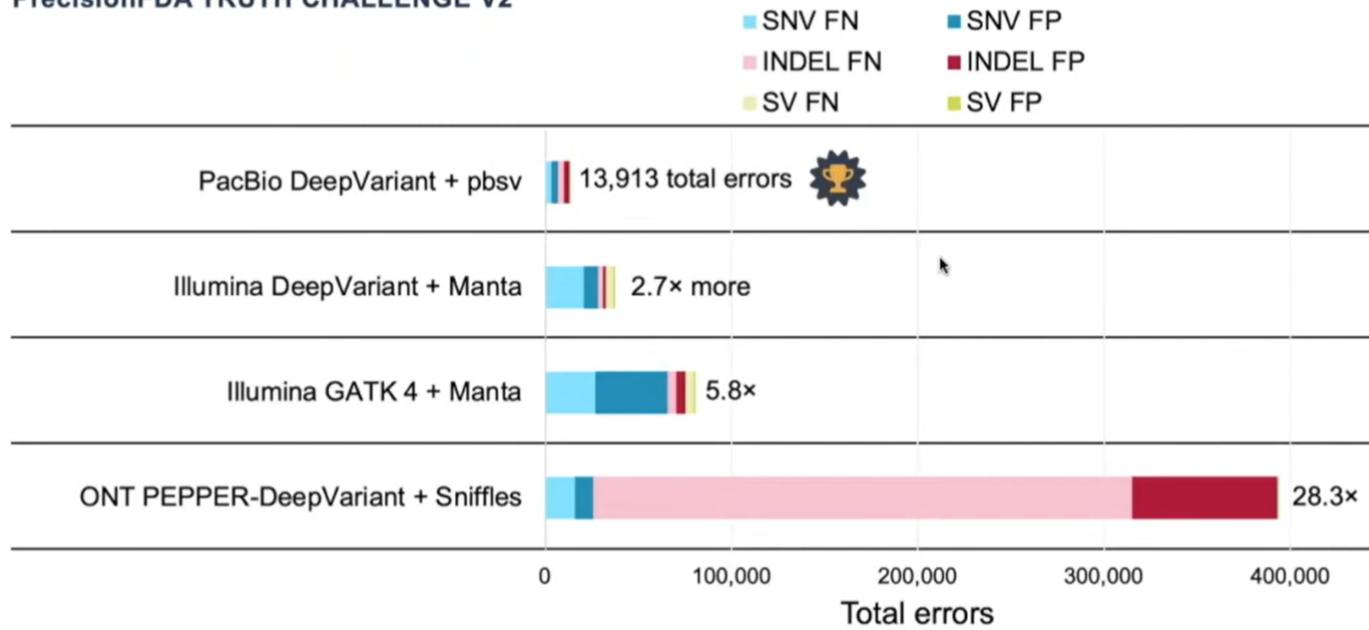


DATA ANALYSIS WORKFLOW



VARIANT CALLING COMPARISON, HG003

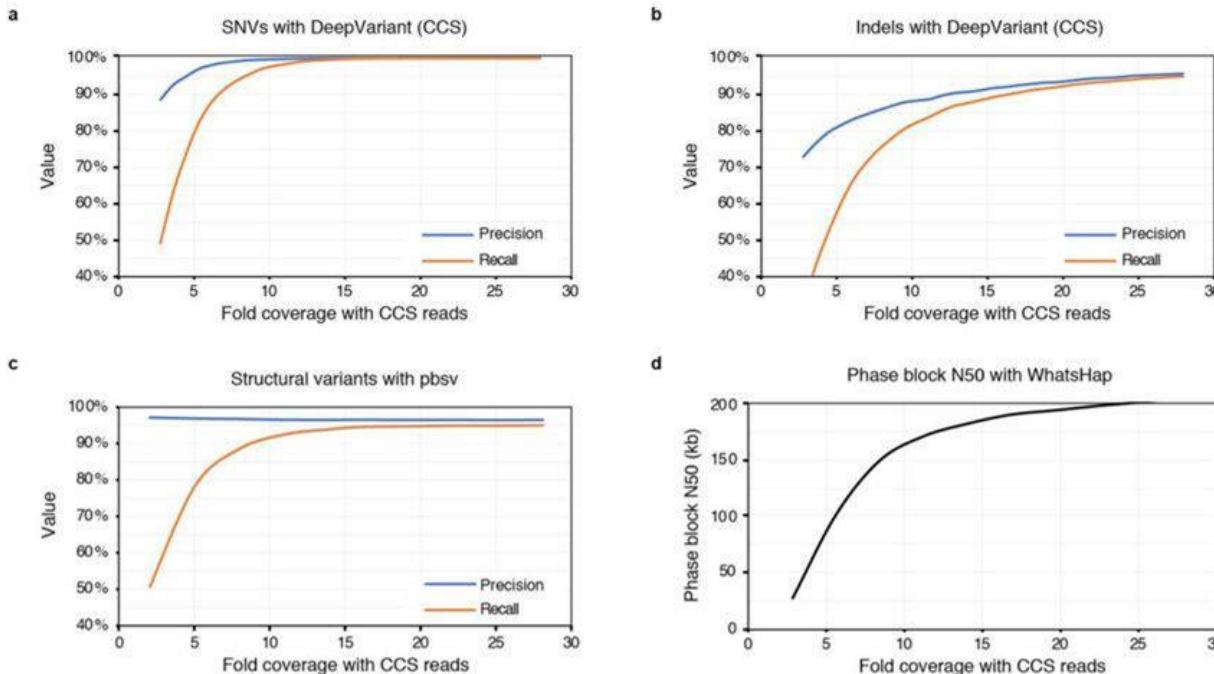
PrecisionFDA TRUTH CHALLENGE V2



PacBio: 35× HiFi reads (Sequel II, Chemistry 2.0); Illumina: 35× NovaSeq ; ONT: 60× PromethION R9.4.1

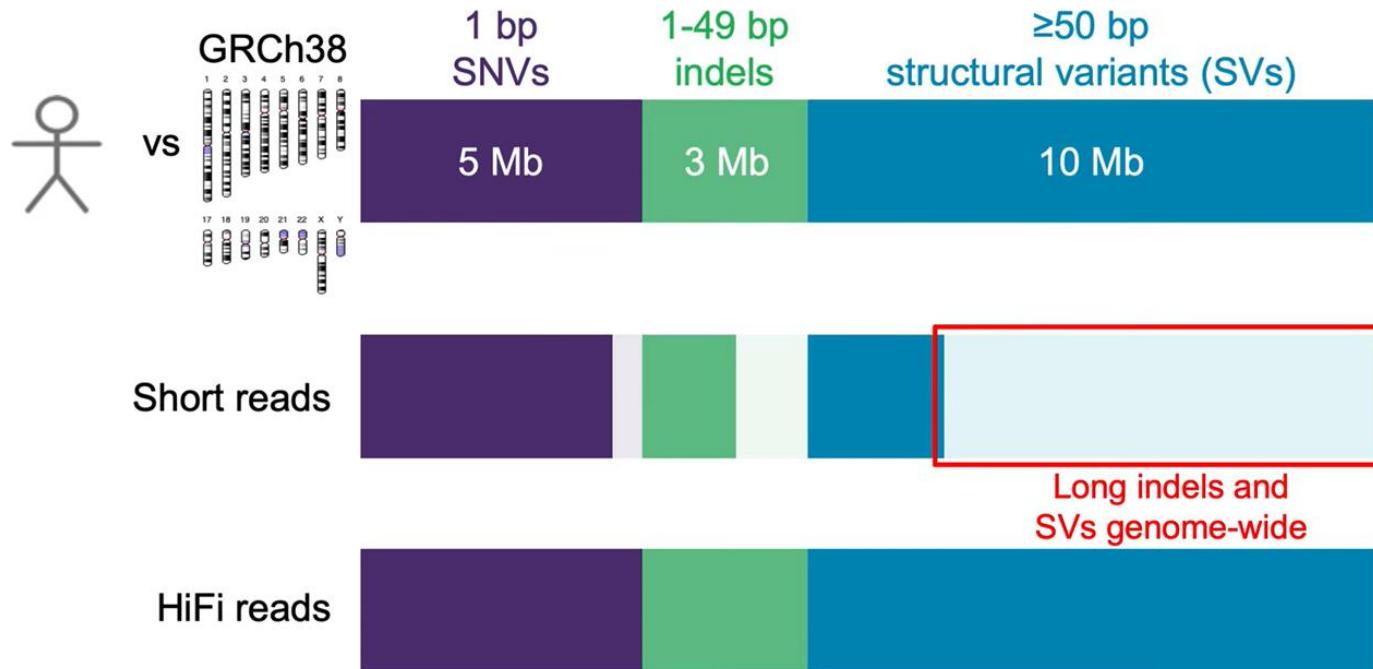
Small variant calling: <https://precision.fda.gov/challenges/10>; Structural variant calling: HG002 against Genome in a Bottle v0.6 benchmark

15-FOLD HIFI READ COVERAGE RECOMMENDATION FOR COMPREHENSIVE VARIANT DETECTION APPLICATIONS

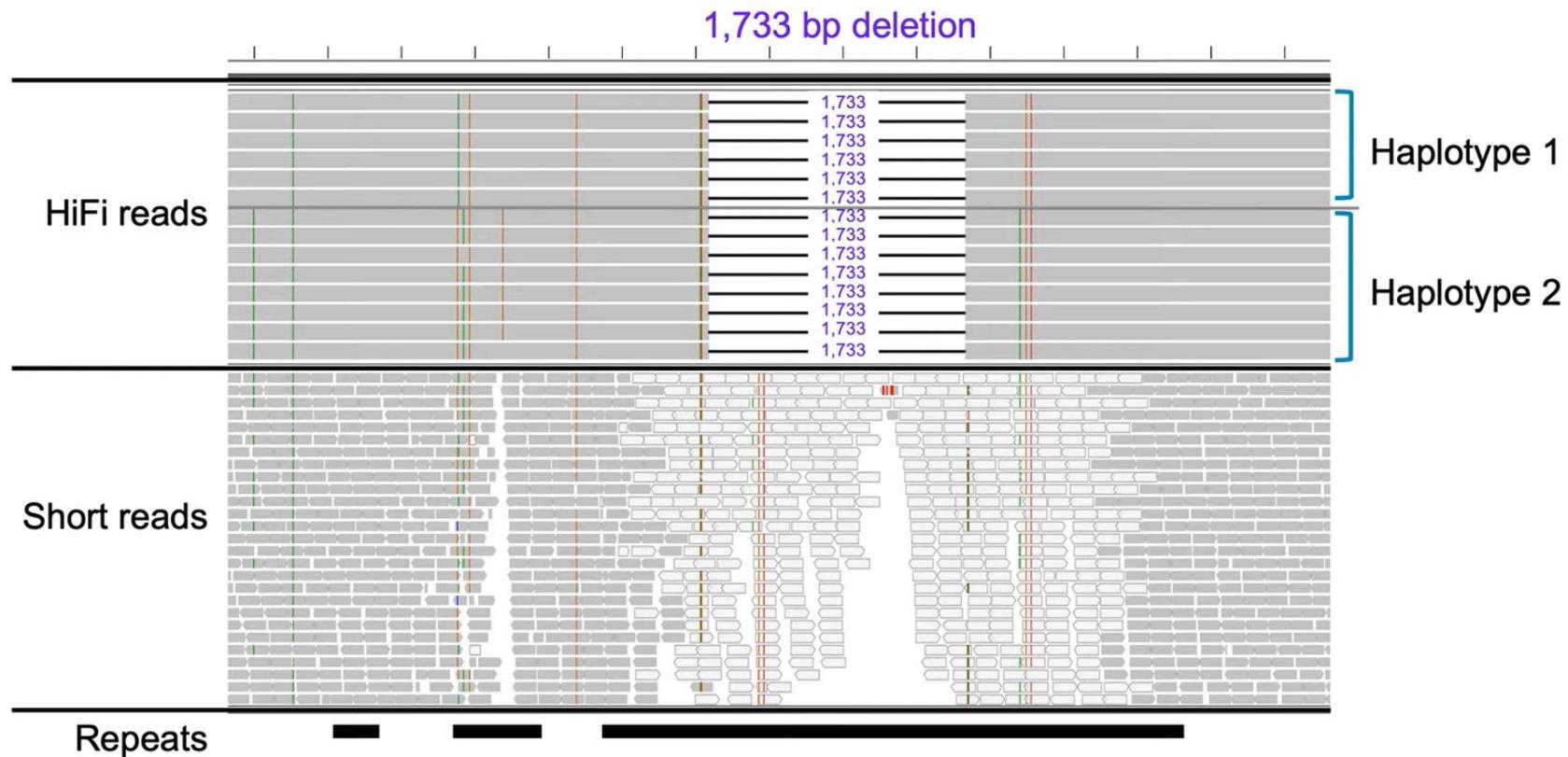


15-fold HiFi (>Q20) Coverage
[2 SMRT cells 8M for 3Gb genome]

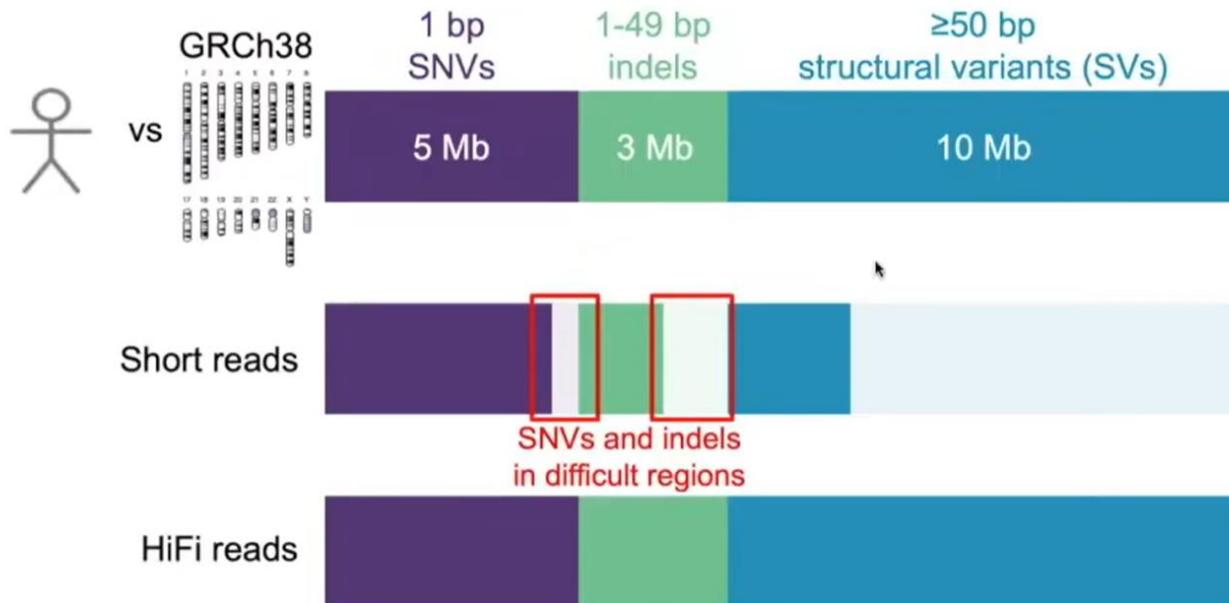
HIFI READS PROVIDE A COMPREHENSIVE VIEW OF VARIATION IN THE HUMAN GENOME



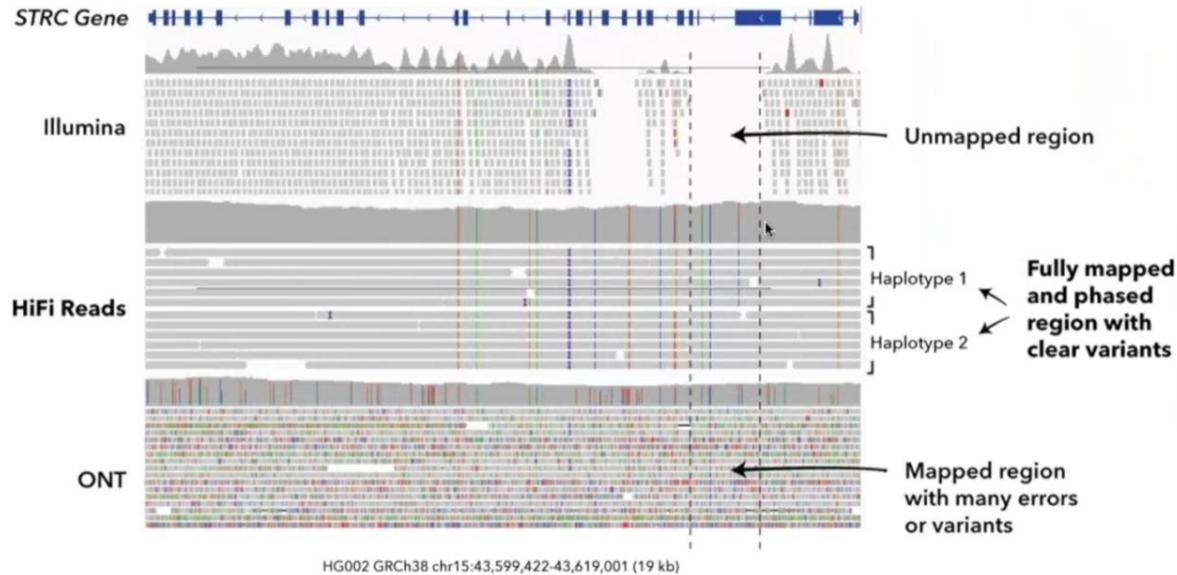
HIFI READS SPAN STRUCTURAL VARIANTS



HIFI READS PROVIDE A COMPREHENSIVE VIEW OF VARIATION IN THE HUMAN GENOME



HIFI READS MAP AND REVEAL VARIANTS IN DIFFICULT REGIONS



STRC gene alignments from [Genome in a Bottle \(GIAB\)](#), HG002 NA24385 son. (IGV settings)

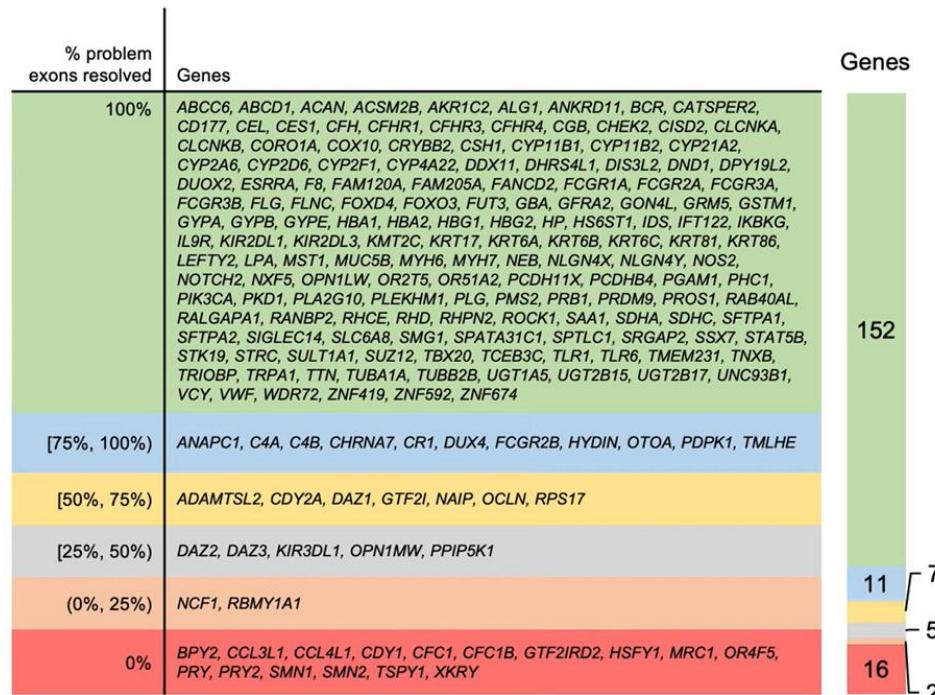
LONG-READ SEQUENCING HAS MANY BENEFITS FOR VARIANT DETECTION

“A comparison of the same individuals sequenced with the Illumina short-read and PacBio long- read platforms, for example, showed that 47% of the deletions and nearly 78% of insertions were missed by Illumina whole-genome sequencing even after application of 11 different variant callers designed to detect insertions, deletions, inversions and duplications in genomes.”

MORE VARIANTS IN MEDICALLY-RELEVANT GENES

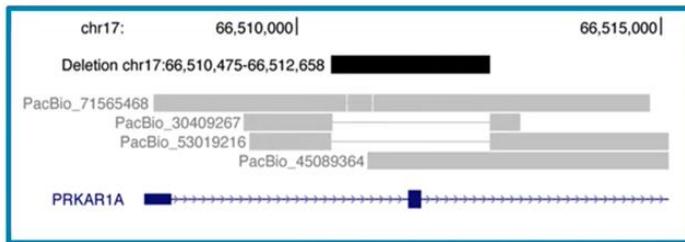
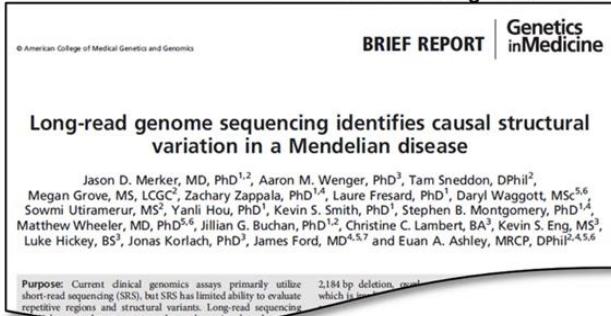


NGS missed ~5% of medical exome
Hifi reads detect medically-relevant
genes missed by short reads.

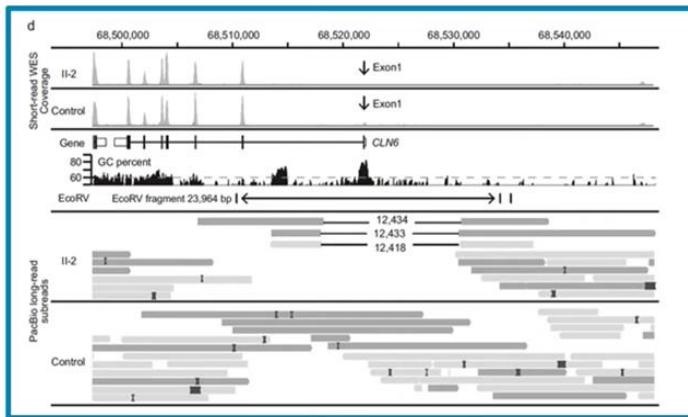
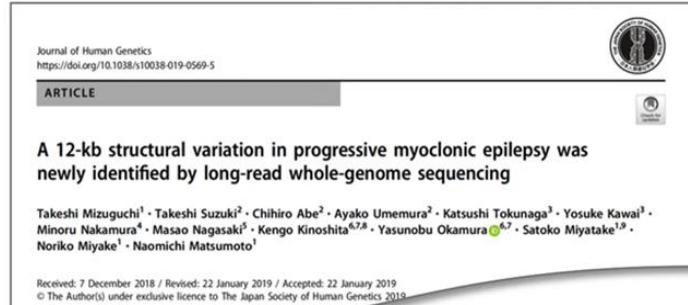


PATHOGENIC VARIANTS DETECTED WITH PACBIO

doi:10.1038/gim.2017.86



doi:10.1038/s10038-019-0569-5



SOLVING THE PREVIOUSLY UNSOLVABLE CASES

 CSH
Cold Spring Harbor Laboratory

bioRxiv
THE PREPRINT SERVER FOR BIOLOGY

New Results

Long-read genome sequencing for the diagnosis of neurodevelopmental disorders

Susan M. Hiatt, James M.J. Lawlor, Lori H. Handley, Ryne C. Lori Beth Boston, Melissa Williams, Jerry Jenkins, David E. Jane Grimwood, Jeremy Schmutz, Gregory M. Cooper
[doi: https://doi.org/10.1101/2020.07.02.185447](https://doi.org/10.1101/2020.07.02.185447)

Journal of Human Genetics

Article | Published: 13 February 2019

A 12-kb structural variation in progressive myoclonic epilepsy was newly identified by long-read whole-genome sequencing

Takeshi Mizuguchi, Takeshi Suzuki, Chihiro Abe, Ayako Umemura, Katsushi Tokunaga, Yosuke Kawai, Minoru Nakamura, Masao Nagasaki, Kengo Kinoshita, Yasunobu Okamura, Satoko Miyatake, Noriko Miyake & Naomichi Matsumoto 

[Journal of Human Genetics 64, 359–368\(2019\) | Cite this article](https://doi.org/10.1007/s00438-019-01080-w)

BRIEF REPORT | Genetics in Medicine

© American College of Medical Genetics and Genomics

Long-read genome sequencing identifies causal structural variation in a Mendelian disease

Jason D. Merker, MD, PhD^{1,2}, Aaron M. Wenger, PhD³, Tam Sneddon, DPhil², Megan Grove, MS, LCGC², Zachary Zappala, PhD^{1,4}, Laure Fresard, PhD¹, Daryl Wagrott, MS^{5,6}, Sowmi Utiramerur, MS², Yanli Hou, PhD¹, Kevin S. Smith, PhD¹, Stephen B. Montgomery, PhD^{1,4}, Matthew Wheeler, MD, PhD^{5,6}, Jillian G. Buchan, PhD^{1,2}, Christine C. Lambert, BA³, Kevin S. Eng, MS³, Luke Hickey, BS⁵, Jonas Korlach, PhD⁴, James Ford, MD^{4,5,7} and Euan A. Ashley, MRCP, DPhil^{1,4,5,6}

 Journal of Human Genetics

Review Article | Published: 27 September 2019

Long-read sequencing for rare human genetic diseases

Satomi Mitsuhashi  & Naomichi Matsumoto

[Journal of Human Genetics 65, 11–19\(2020\) | Cite this article](https://doi.org/10.1007/s00438-019-01080-w)

Pathologist

SMRT Long-Read Sequencing Solves Genetic Mysteries

Solving rare disease with single molecule, real-time sequencing

Luke Hickey | 05/08/2020 | Longer Read

Many thanks for your attention!



[View in Icarus contig browser](#)

All statistics are based on contigs of size ≥ 500 bp, unless otherwise noted (e.g., "# contigs (≥ 0 bp)" and "Total length (≥ 0 bp)" include all contigs).

Statistics without reference  **hsv.contigs**

# contigs	3
# contigs (≥ 0 bp)	3
# contigs (≥ 1000 bp)	3
# contigs (≥ 5000 bp)	3
# contigs (≥ 10000 bp)	2
# contigs (≥ 25000 bp)	1
# contigs (≥ 50000 bp)	1
Largest contig	133 009
Total length	160 798
Total length (≥ 0 bp)	160 798
Total length (≥ 1000 bp)	160 798
Total length (≥ 5000 bp)	160 798
Total length (≥ 10000 bp)	154 064
Total length (≥ 25000 bp)	133 009
Total length (≥ 50000 bp)	133 009
N50	133 009
N90	21 055
auN	113 061
L50	1
L90	2
GC (%)	69.09

Mismatches

# N's per 100 kbp	0
# N's	0

3 Assembly	pilon_polished
4 # contigs (>= 0 bp)	1
5 # contigs (>= 1000 bp)	1
6 # contigs (>= 5000 bp)	1
7 # contigs (>= 10000 bp)	1
8 # contigs (>= 25000 bp)	1
9 # contigs (>= 50000 bp)	1
10 Total length (>= 0 bp)	151124
11 Total length (>= 1000 bp)	151124
12 Total length (>= 5000 bp)	151124
13 Total length (>= 10000 bp)	151124
14 Total length (>= 25000 bp)	151124
15 Total length (>= 50000 bp)	151124
16 # contigs	1
17 Largest contig	151124
18 Total length	151124
19 GC (%)	70.09
20 N50	151124
21 N90	151124
22 auN	151124.0
23 L50	1
24 L90	1
25 # N's per 100 kbp	6.62