Emerging NGS Platforms

Ultima Genomics

The UG100

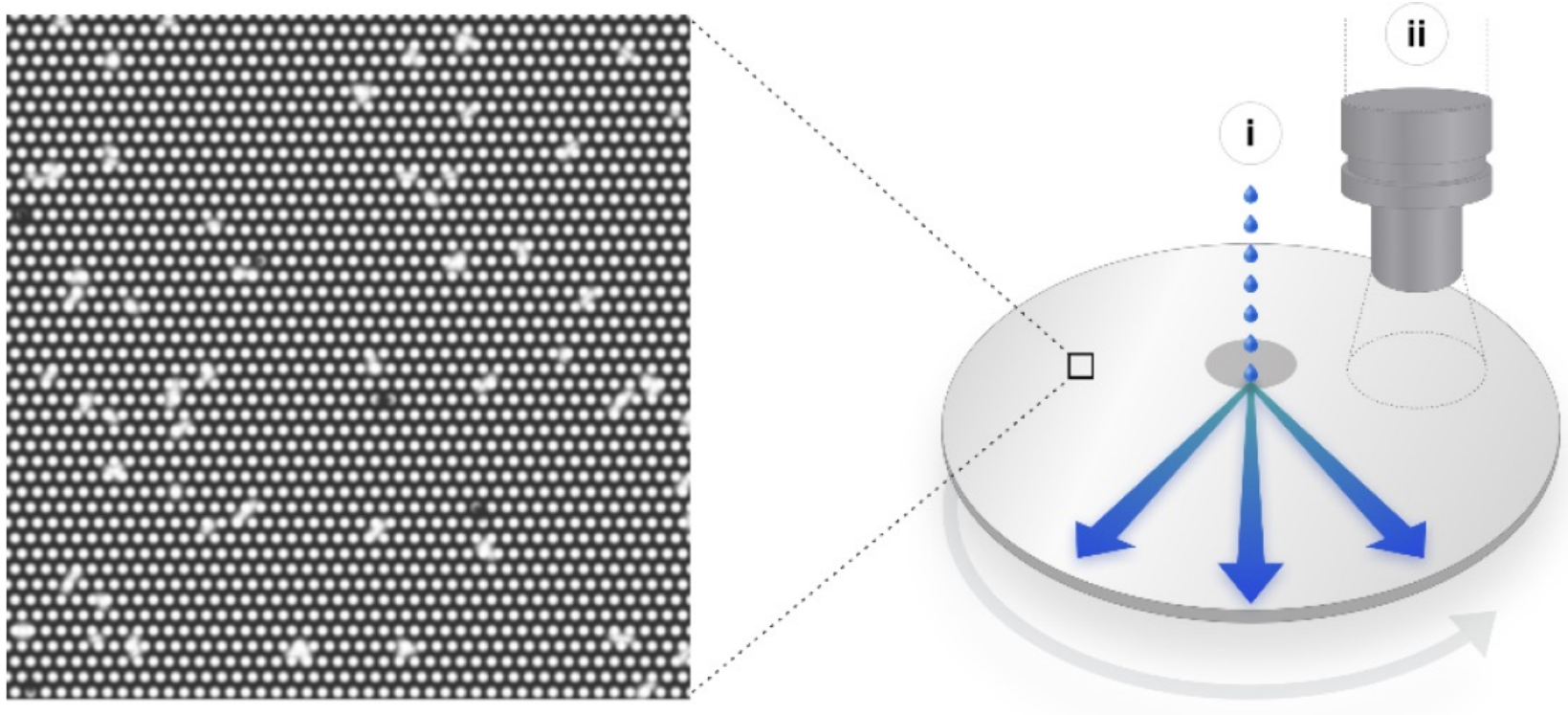
Footprint is the “Ultima”te of its kind

- Three instruments total
  - Clustering
  - Sequencing
  - Computational

Ultima Genomics

Three main innovative components: (1) *open fluidics and optics system*

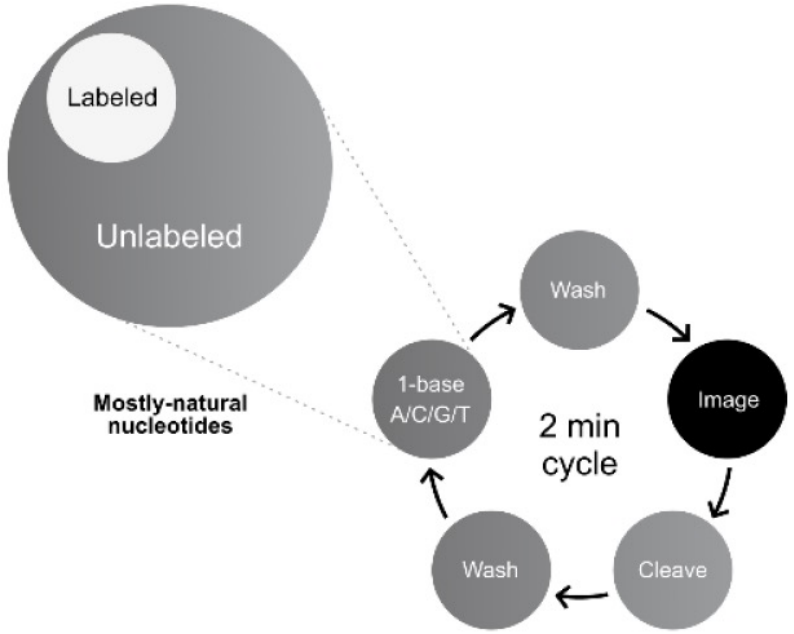
- Circular silicon wafer as an “open flow-cell”
- Patterned – dense array of electrostatic landing pads to bind sequencing beads
  - ~10 billion clonally amplified beads
- Spin-dispense system

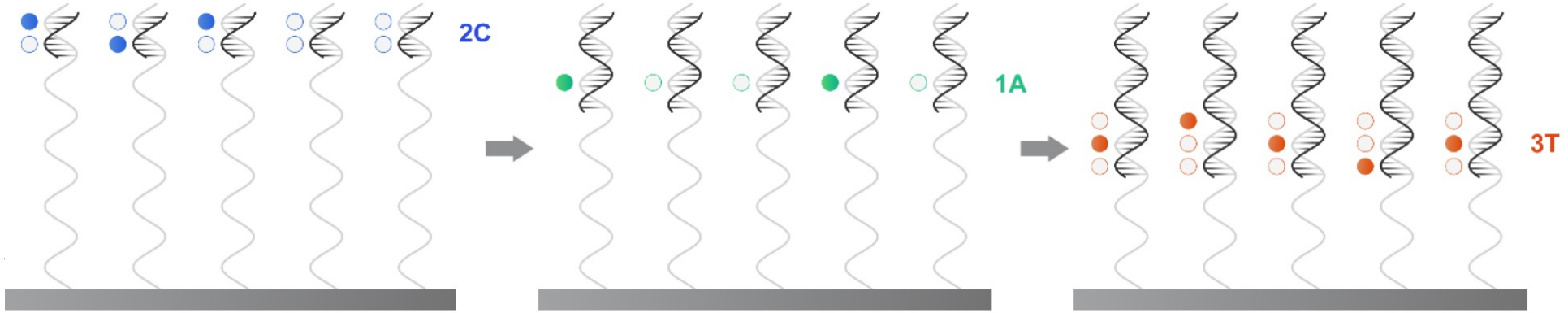


Ultima Genomics

Three main innovative components: (2) *mostly natural sequencing chemistry*

- Sequencing-by-synthesis (mnSBS) uses a mixture of native dNTPs and one-at-a-time fluorescently labeled dNTP
- Polymerase extends 0, 1 or several bases depending on respective homopolymer
- Detected signal proportional to length of homopolymer





Ultima Genomics

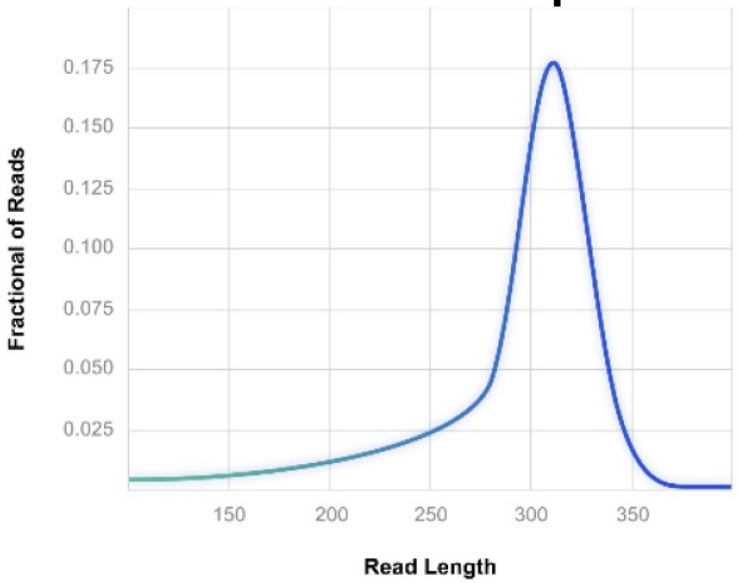
Three main innovative components: (3) *neural network-enabled base-calling*

- Machine learning and convolutional neural network (CNN) to convert raw signals to sequence reads; run-specific calibration process; CRAM file

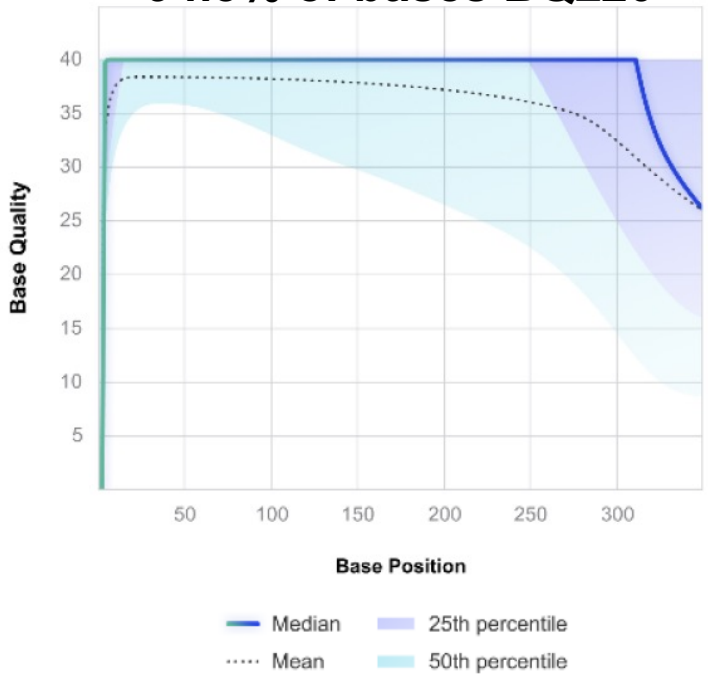
HG001-HG007 Generated Data

- 444 cycles and a 20 hr run time

Mode 310 bp

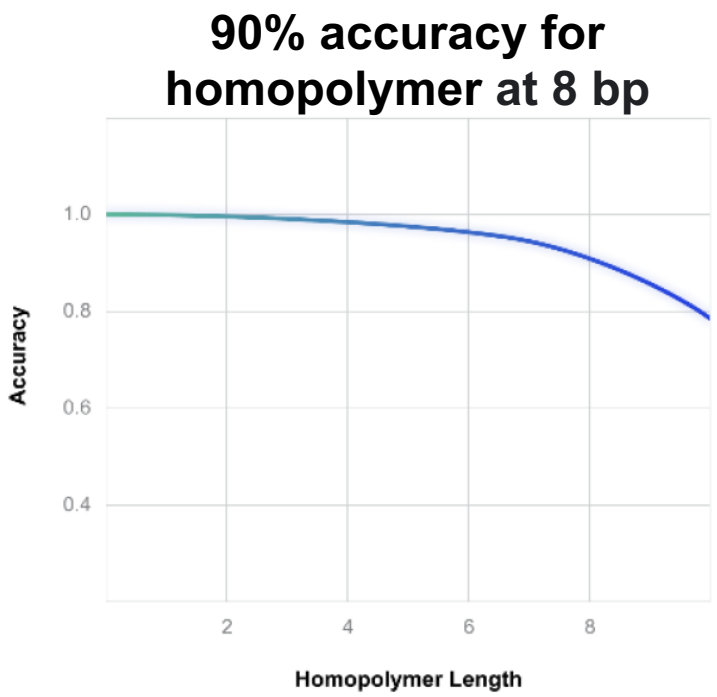


94.5% of bases BQ≥20



Ultima Genomics

Three main innovative components: (3) *neural network-enabled base-calling*



	HG001	HG002	HG003	HG004	HG005	HG006	HG007	GIAB Mean
Mean coverage	40.05	39.58	39.67	38.56	39.16	39.53	38.32	39.27
% >20X	95.9%	95.7%	95.9%	96.6%	96.8%	96.1%	96.7%	96.2%
% Duplication	4.1%	6.6%	6.2%	7.3%	7.6%	6.1%	8.2%	6.6%
F90*	1.483	1.466	1.469	1.377	1.398	1.464	1.368	1.432
F95*	1.821	1.885	1.803	1.677	1.632	1.797	1.666	1.754
PF reads (M)**	469	435	436	4338	430	443	433	440
% reads aligned	99.80%	99.96%	99.96%	99.97%	99.96%	99.95%	99.97%	99.94%
Mean read length	264.8	284.9	284.4	286.7	288.6	280.3	286.4	282.3
Median read length	291	302	302	303	304	301	302	300.7
Modal read length	309	311	311	311	311	310	311	310.6
% chimeras	2.3%	1.1%	1.2%	1.3%	1.4%	1.5%	1.3%	1.5%
Raw Indel error	0.27%	0.30%	0.29%	0.28%	0.29%	0.38%	0.28%	0.30%
HQ Mismatch error†	0.07%	0.08%	0.07%	0.07%	0.07%	0.10%	0.07%	0.07%
% BQ20 bases	95.5%	94.8%	95.0%	95.2%	95.0%	93.6%	95.2%	94.9%
% BQ30 bases	87.3%	86.4%	86.8%	87.6%	87.1%	84.6%	87.7%	86.8%
Ti/Tv ratio Exome	2.97	2.89	2.95	2.90	2.98	2.93	2.97	2.94
Ti/Tv ratio‡	2.09	2.09	2.09	2.09	2.09	2.09	2.09	2.09
SNP recall‡	99.7%	99.6%	99.6%	99.7%	99.7%	99.6%	99.7%	99.7%
SNP precision‡	99.6%	99.6%	99.6%	99.6%	99.6%	99.6%	99.6%	99.6%
SNP F1‡	99.7%	99.6%	99.6%	99.7%	99.7%	99.6%	99.7%	99.6%
Indel recall‡	96.7%	96.4%	96.6%	95.4%	96.0%	95.9%	96.0%	96.1%
Indel precision‡	97.0%	96.8%	97.1%	96.4%	97.0%	96.2%	96.7%	96.7%
Indel F1‡	96.9%	96.6%	96.8%	95.9%	96.5%	96.1%	96.3%	96.4%

Table 1: Performance metrics for Genome in a Bottle (GIAB) reference genomes HG001-7, and average performance metrics for 224 additional 1000 Genomes (1000G) reference genomes.

* F90/95: Ratio of coverage between the median and the 10th or 5th percentile lowest coverage, respectively.
** PF: Pass-filter reads. All other metrics were calculated over these reads.
† HQ Mismatch error rate was corrected for germline SNPs and alignment errors (see Methods section).
‡ Variant calling metrics were measured on GIAB HCR excluding long homopolymers and repetitive regions (UG-HCR, see Methods).

GIAB Mean
SNP: 99.7% recall; 99.6% precision
Indel: 96.1% recall; 96.7% precision

Ultima Genomics Summary

- Technology: mostly natural sequencing-by-synthesis (mnSBS)
- Output: 10 billion clonally amplified beads; 2 Wafers at a time
- Runtime: 20 hr
- Length of reads: 310 bp mode size (GIAB)
- Accuracy:
  - SNP: 99.7% recall; 99.6% precision
  - Indel: 96.1% recall; 96.7% precision
- Instrument cost: Unknown (comparable to Illumina NovaSeq 6000)
- Instrument size (relative): Several NovaSeqs worth
- Partnerships/acquisitions:
  - AI partnerships with Google DeepVariant, NVIDIA and Senteon
  - Exact Sciences, Regeneron Pharmaceuticals, NYGC and Broad Institute: beta testing—both NYGC and Broad have purchased

Element Biosciences – AVITI Benchtop Sequencer

Highlights

- **Two** independent, random-access flow cells (>1b paired-end reads output per flow cell)
- **FASTQ** file output (conversion performed by AVITI Operating Software Bases2Fastq workflow)
- Three library workflows (**Adept** – conversion of existing libraries ; **Elevate** – Aviti library preparation, **LoopSeq**—long read sequencing approach)
- **Tunable optical throughput** (full scan, ½ scan, ¼ scan etc..) to select desired TAT/read depth

Element Biosciences - AVITI

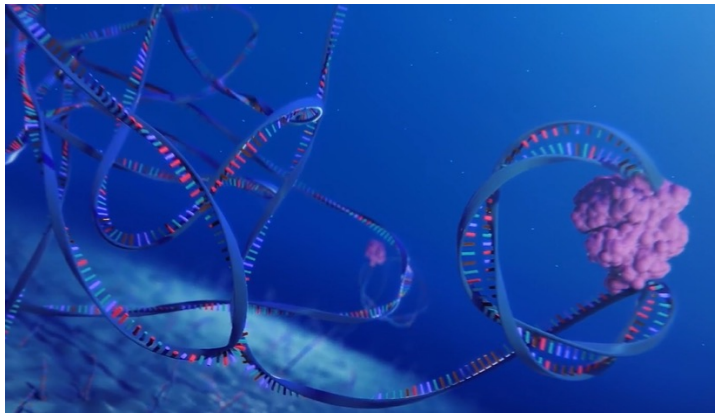
Circularization

Off-instrument library circularized prior to sequencing

"*Working to move onto the flow cell*"

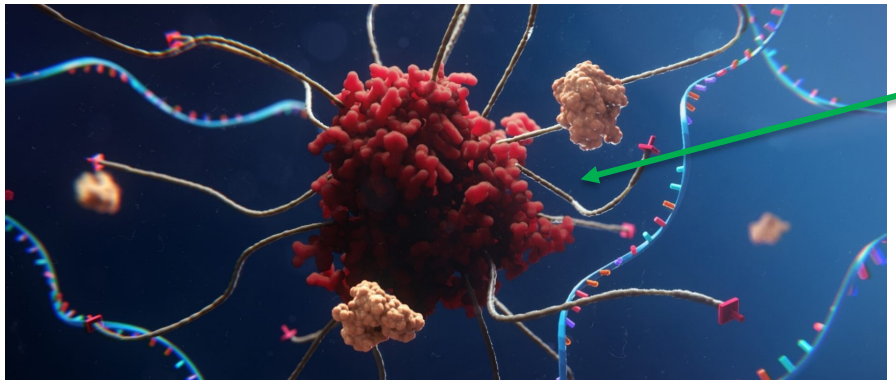
Pollination

Rolling circular amplification on flow cell surface creates clonal copies of the library molecule or *"polony"*



Sequencing

Instead of binding an individual labeled nucleotide at each location of the cluster, AVITI uses single fluor with many octopus-like tentacle arms (Avidite)



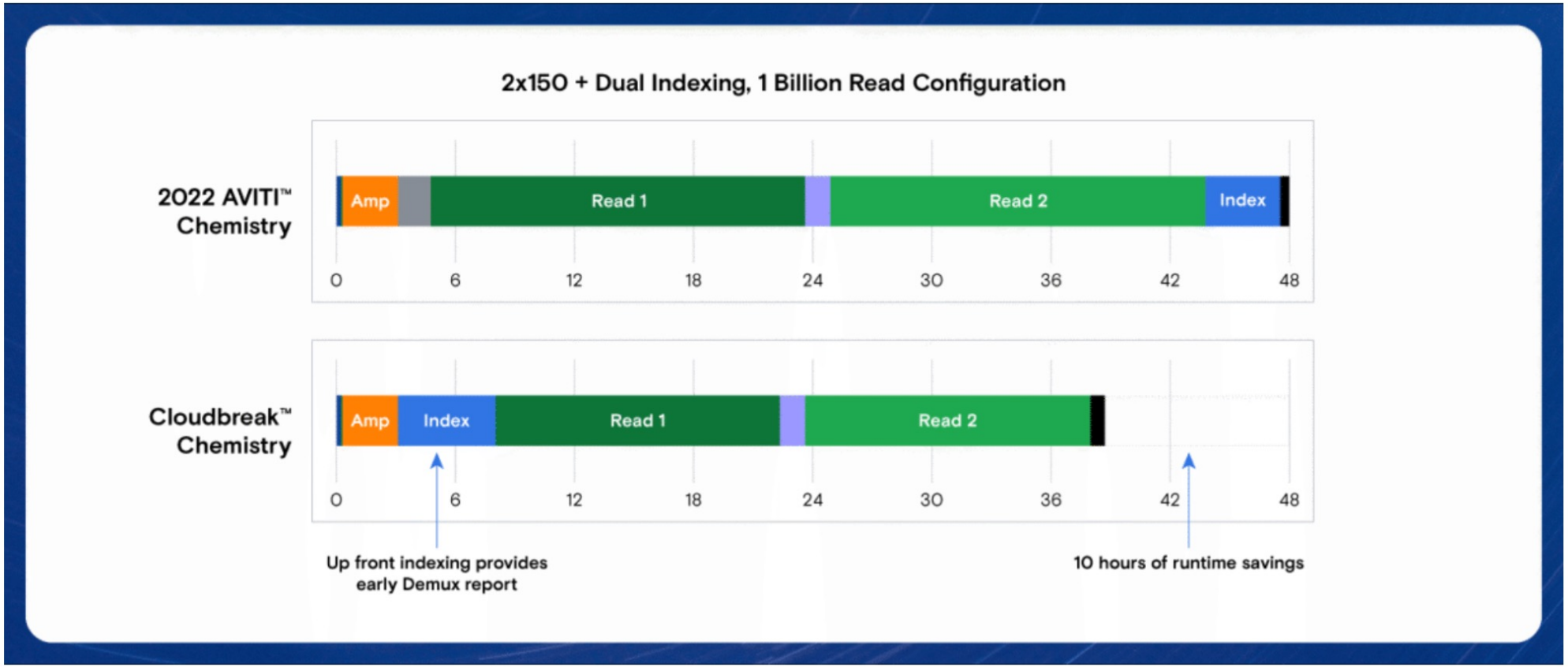
Avidite nucleotide substrate

Each arm, single nt type

One avidite = multiple binding sites within the polony = increased binding avidity

[Element Biosciences - Resources]

Aviti Cloudbreak Chemistry



- 20% faster run times
- Reads index 1 and 2 first for initial run/library quality control purposes
- Improved data quality at the ends of reads, and at homopolymers

Element Biosciences – AVITI Key Performance Metrics

Highly efficient kit choices cover a spectrum of applications and maintain cost savings across output requirements. In only 38 hours, two 2 x 150 runs with indexing produce up to 600 Gb of data and 2 billion reads.

READ LENGTH	HIGH OUTPUT FLOW CELL (GB/HOURS)*	MEDIUM OUTPUT FLOW CELL (GB/HOURS)	LOW OUTPUT FLOW CELL (GB/HOURS)	QUALITY SCORES
Read Count	1 billion**	500 million	250 million	Q30
2 x 75	150/24	75/20	Not applicable	> 90%
2 x 150	300/38	150/31	75/27 hours	> 90%
Read Count	300 million	100 million	Not applicable	Q30
2 x 300	180/60	60/51	Not applicable	> 85%

* Individually addressable lanes slightly extend run times.

** The numbers of reads are based on sequencing Element-prepared libraries. Actual read counts might differ based on lab-specific factors.

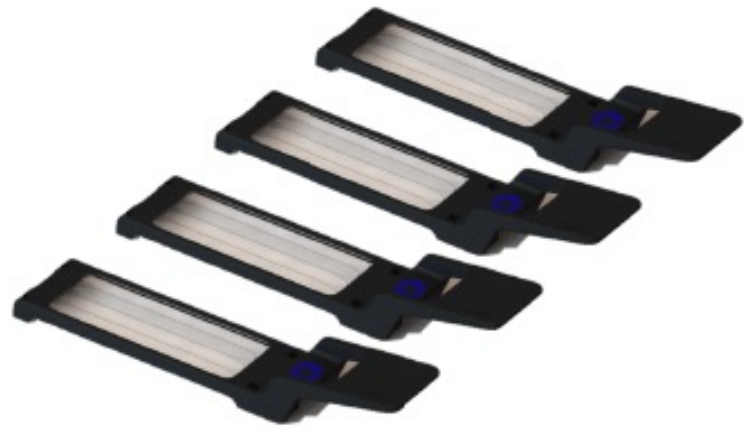
Singular Genomics G4 Platform

G4 Sequencing Platform
Bench Top

SYSTEM DESIGN

ENGINEERED FOR SPEED, THROUGHPUT, FLEXIBILITY

- Flexible run size: 1–4 flow cells
- 4 independent lanes/FC (up to 16 individual samples/run)
- Integrated cluster generation (amplification)
- Chemistry, optics, fluidics for rapid SBS: 2.5 min/cycle
- 400M Reads/FC or 1600M Reads/run



SINGULAR GENOMICS

Singular G4 Design

SBS FLOW CELL
WHERE IT ALL HAPPENS

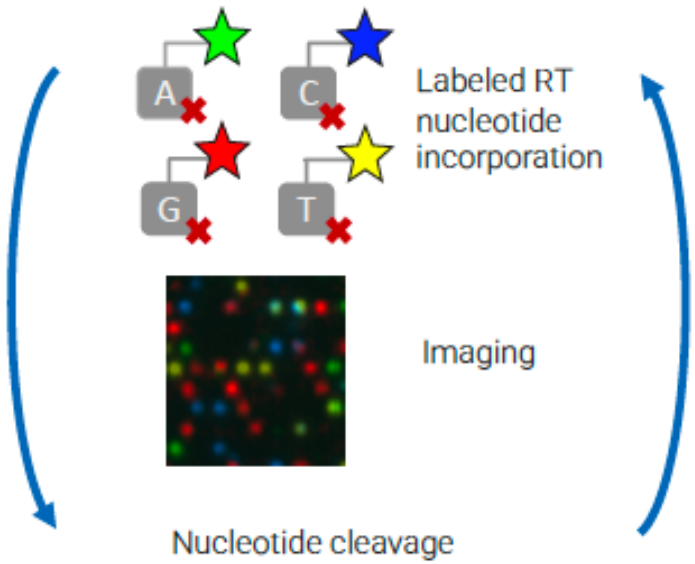
- Lithographically patterned flow cell
- Precise microfluidic channels
- Polymer “scaffold” for clustering & sequencing
- 4 independently-addressable lanes
- Ergonomic handle & frame

SINGULAR GENOMICS

HIGH-THROUGHPUT, 4-COLOR FLUORESCENCE IMAGING
DESIGNED FOR SPEED-READING DNA

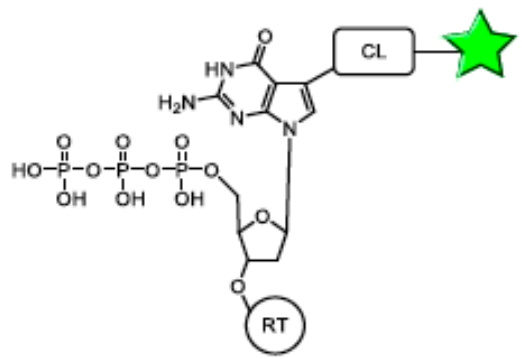
- Designed to image 4 dyes simultaneously
- Large aperture optics (high NA) for max resolution & light collection
- High-speed imaging at > 2B pixels/sec
- Real-time image processing

SEQUENCING BY SYNTHESIS (SBS) CHEMISTRY
NUCLEOTIDES WITH CLEAVABLE TERMINATORS AND DYES

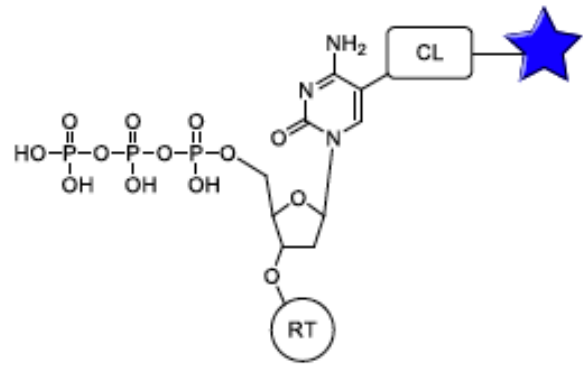


Desired properties in a new SBS chemistry:

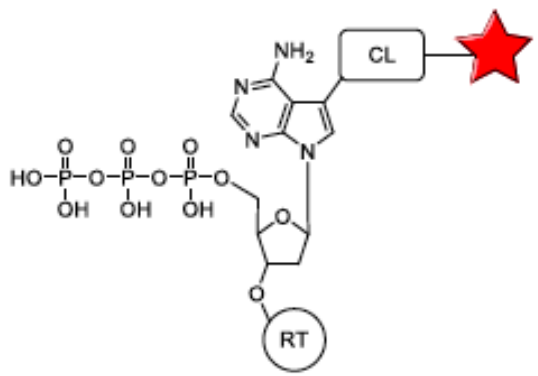
- Fast, accurate incorporation by the sequencing polymerase
- Rapidly cleavable terminator & dye linker
- Stable in storage and during sequencing



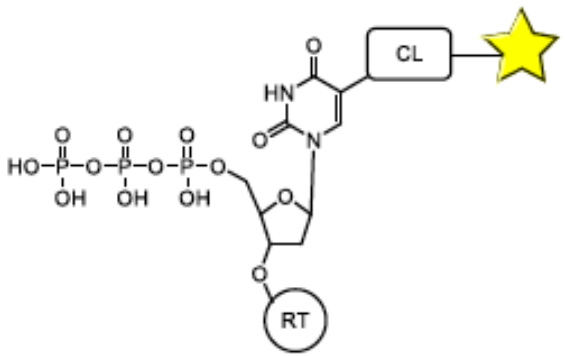
Adenine (A)



Cytosine (C)



Guanine (G)



Thymine (T)

G4 SEQUENCING REAGENTS

	Reagent Configuration[1]	Run Time[2]	Reads / Flow Cell[3]	Reads / Run[3]	Quality[4]
F2 Flow Cell	100 cycles	~11 hours	200M	1,000M	80–90% Bases ≥ Q30
	200 cycles	~15 hours			
	300 cycles	~19 hours			
F3 Flow Cell	50 cycles	8–11 hours	400M	1,600M	
	100 cycles	11–14 hours			
	200 cycles	15–19 hours			
	300 cycles	19–24 hours			

[1] Reagents include 50 additional cycles above what is represented to account for adapters and indices.
[2] Run time measured from run start through clustering, sequencing and instrument wash for non-indexed reads.
[3] Paired reads passing filter for F2 and F3 are dependent on application and read length.
[4] Performance metrics may be impacted by application, sample quality, library preparation, loading concentration, and other sequencing considerations. Metrics as generated on reference bacterial and human genomes.

FLOW CELL AND LANE COMPARTMENTALIZATION

ACCOMMODATING DIVERSE EXPERIMENTAL NEEDS

Lane Ownership

Multi-Lane Samples

Flow Cell Ownership

Flow Cell Ownership

<u>F2 Flow Cell</u>
~50M Reads per Lane

<u>F3 Flow Cell</u>
~100M Reads per Lane

$150–$250 per lane
$600–$1,000 per flow cell

SINGULAR GENOMICS

VERSATILITY ACROSS CORE APPLICATIONS

	F2 Flow Cell Samples / FC	F3 Flow Cell Samples / FC	Max Read Samples / FC	Samples / Run
Human Whole Genome (2x150 bp, 3 Gb at 30x coverage)	–	1	–	4
Exome (2x100 bp, 34 Mb at 100x coverage)	6	13	–	52
Target Enrichment (2x150 bp, 800 Kb at 4000x coverage)	8	16	–	64
Methyl-Seq (2x100 bp targeting 15x coverage)	1	2	–	8
Shotgun Metagenomics (2x150 bp, 10M reads)	20	40	–	160
RNA-Seq (2x100 bp, 50M reads)	4	8	–	32
Single Cell RNA-Seq (28x91 configuration, 10,000 cells/sample and 20,000 reads/cell)	1	2	4	16

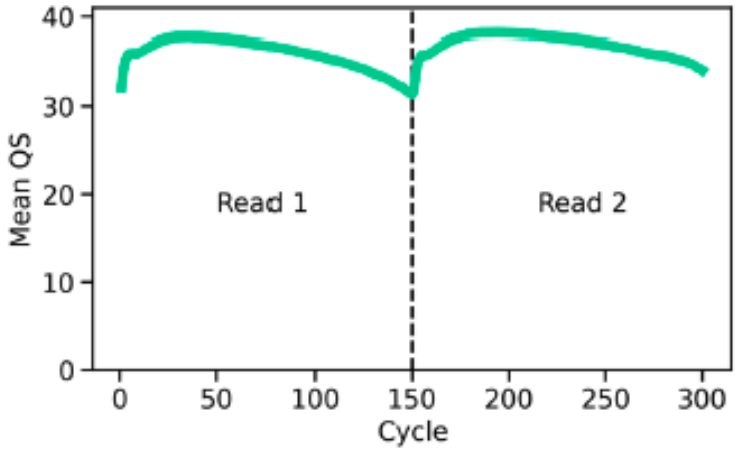
Sequencing multiplexing capabilities are approximations based on assumptions and configurations listed in table and expected, but not guaranteed, throughput on G4 flow cells. Results may vary based on experimental design and sample type. Please refer to kit specifications for more detail.

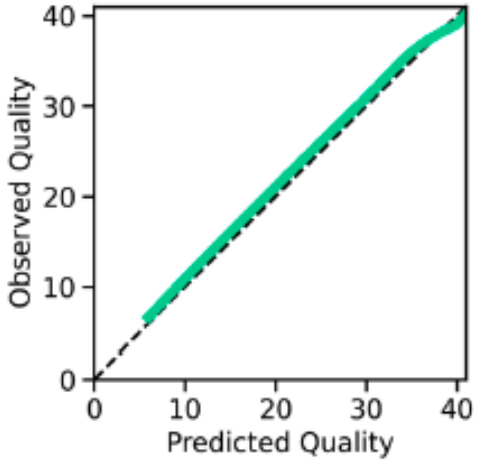
RAPID HUMAN GENOME ON F3 FLOW CELL

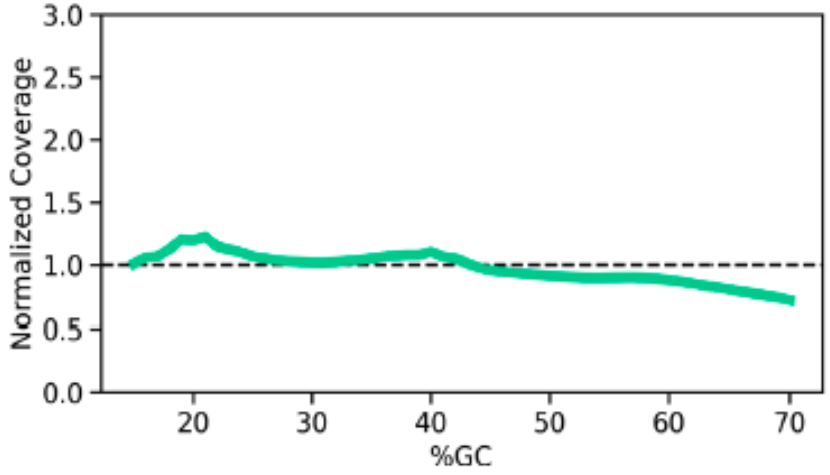
- 1 FC = 1 genome
- 1 - 4 genomes/run; no batching needed; < 24 hr run time
- F3 FC output: 400M Read pairs, 2 x 150 bp
- Genome coverage (HG002): > 30X
- Single-pass accuracy: 99.9%
- Q-scores: 80-90% >Q30

Metric	Value
%PF Reads Aligned	99.9
Median Insert Size (bp)	328
Mean Coverage (X)	33.6
SNP Precision	99.86
SNP Recall	99.18
SNP F1-Score	99.52
Indel (<50bp) Precision	98.33
Indel (<50bp) Recall	97.43
Indel F1-Score	97.88
Het:Hom Ratio	1.51
Ti:Tv Ratio	2.00

Details available in WGS Technical Report 2.0







SINGULAR GENOMICS

SINGLE CELL RNA-SEQ IN MAX READ FORMAT

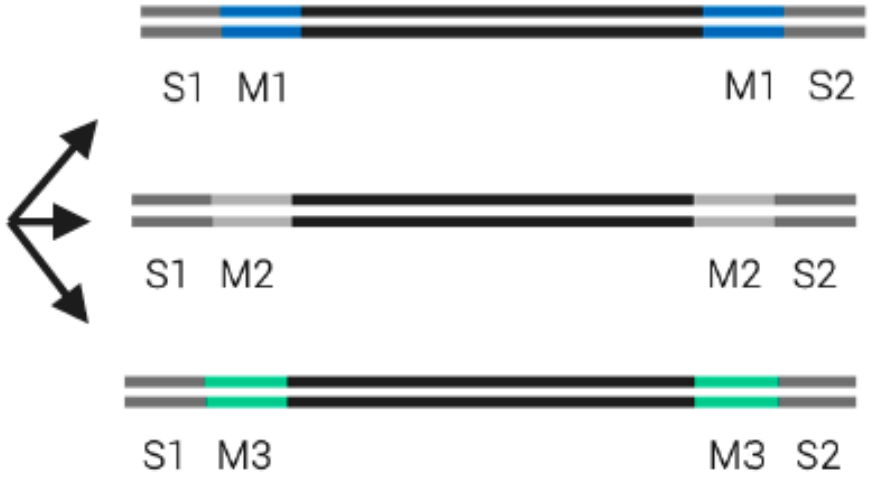
GETTING MORE READS WITH MULTIPLE PRIMERS

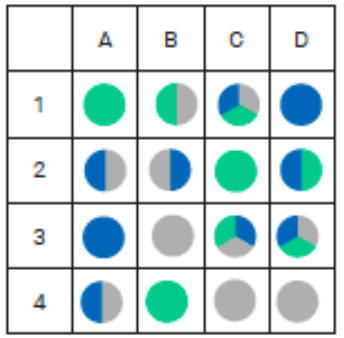
scRNA-Seq library

Split and add Max Read sequencing primers

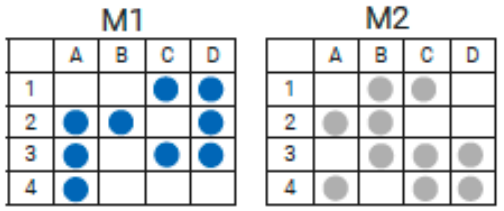
Pool and Sequence on the G4

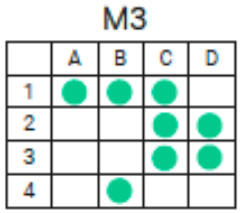
Analyze with CellRanger™ or SpaceRanger™



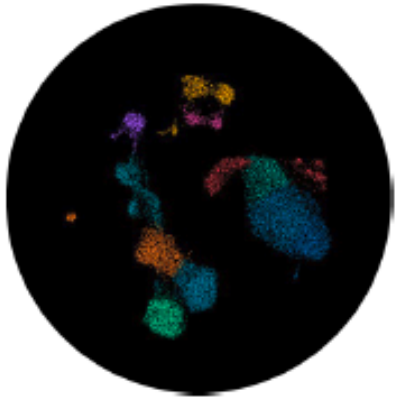


Multi-templated nanowells





Selective sequential priming



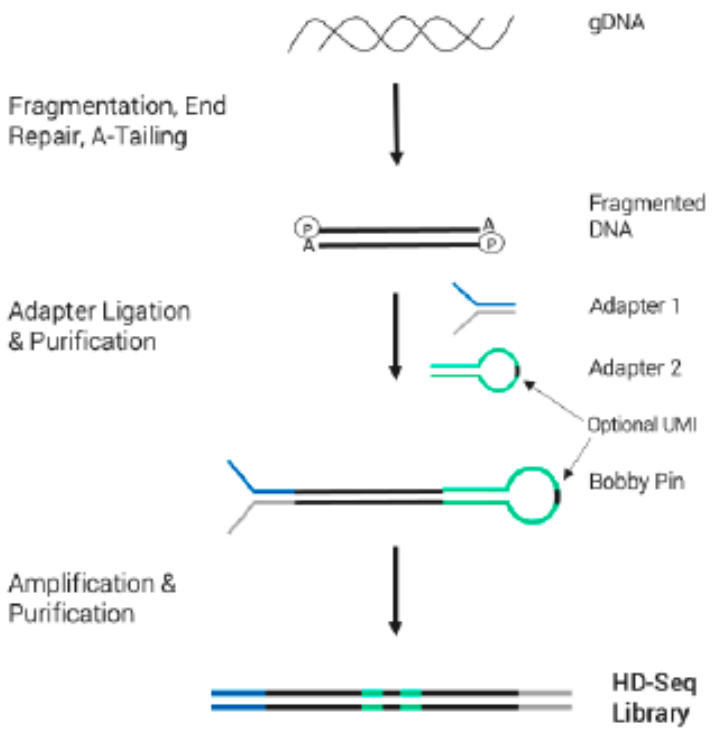
Lane	Flow Cell	Run
1 sample	4 samples	16 samples
200M Reads	800M Reads	3200M Reads

SINGULAR GENOMICS

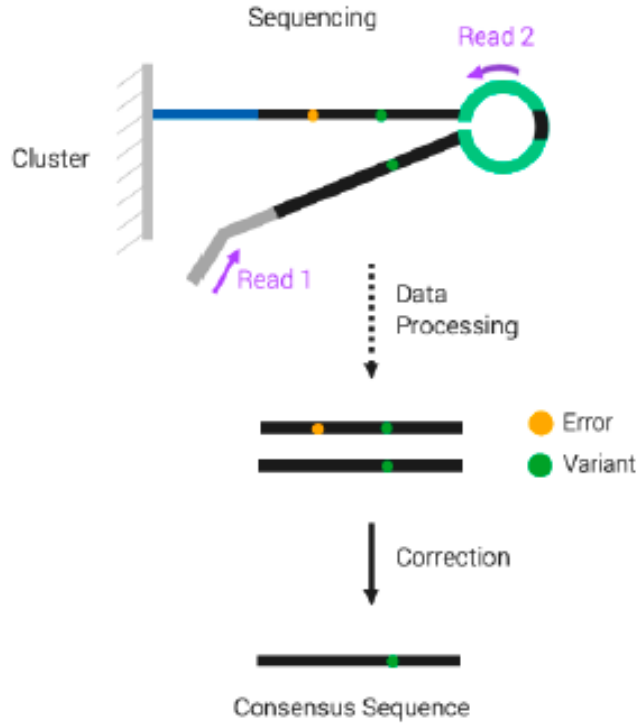
HIGH-DEFINITION SEQUENCING (HD-SEQ)

DOUBLE-STRANDED SEQUENCING FOR RARE VARIANT DETECTION

Methods – Library Preparation



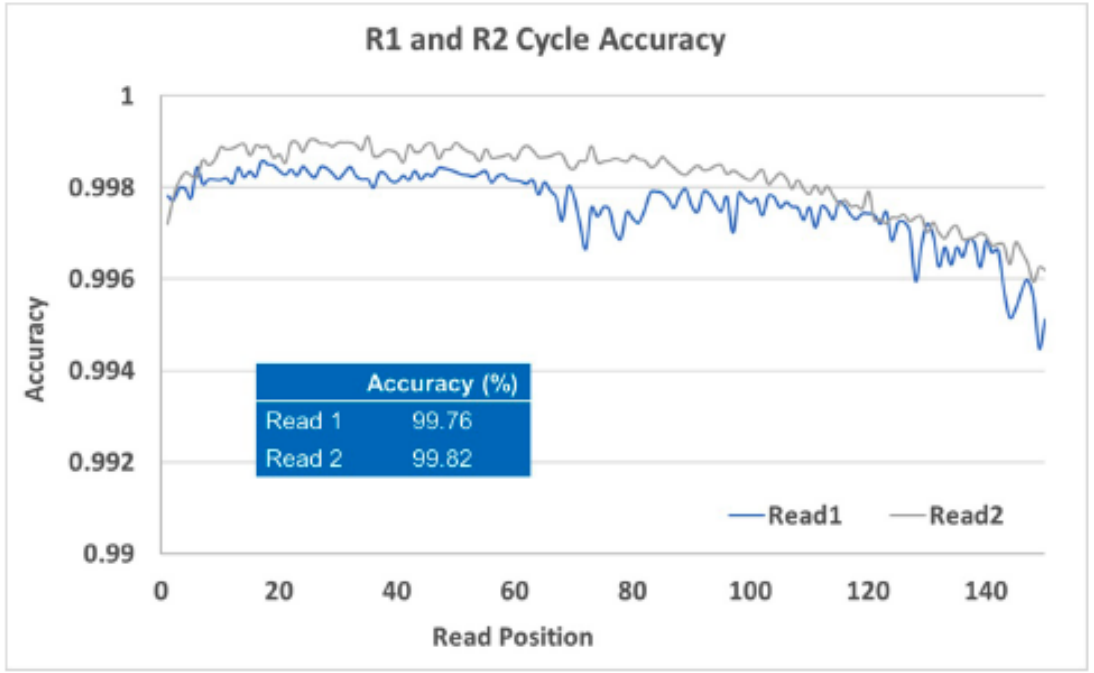
Methods – Sequencing & Bioinformatics

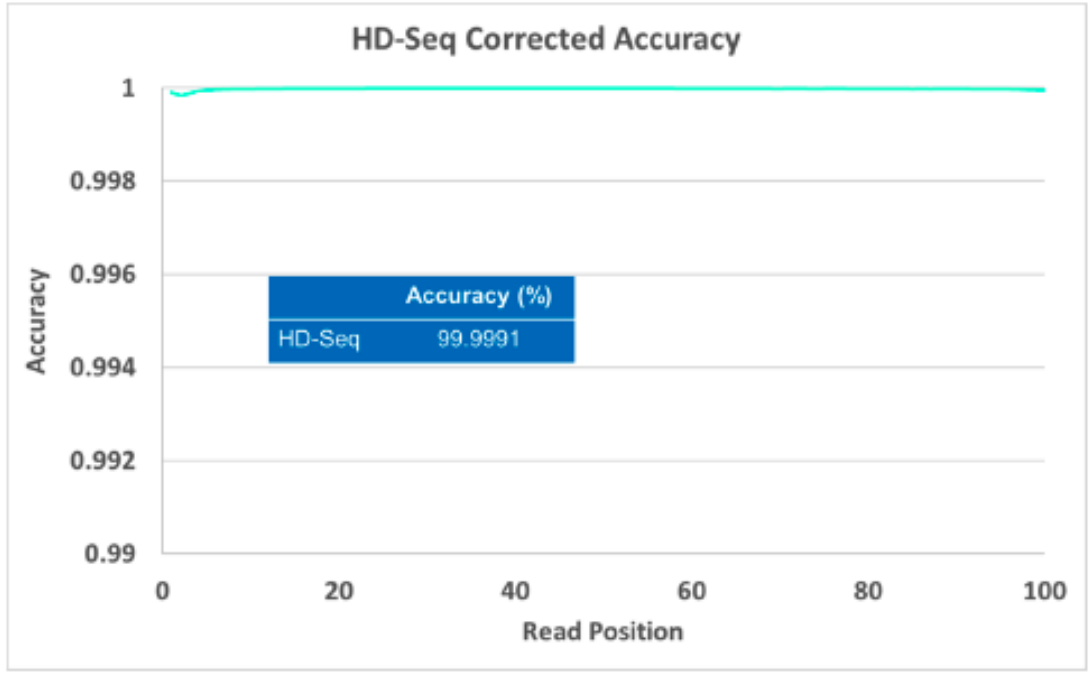


- HD-Seq enables much greater accuracy by physically linking double stranded DNA from library prep through clustering & sequencing
- Ideally suited for rare variant detection in cell-free DNA

HIGH-DEFINITION SEQUENCING (HD-SEQ)

DOUBLE-STRANDED SEQUENCING FOR RARE VARIANT DETECTION



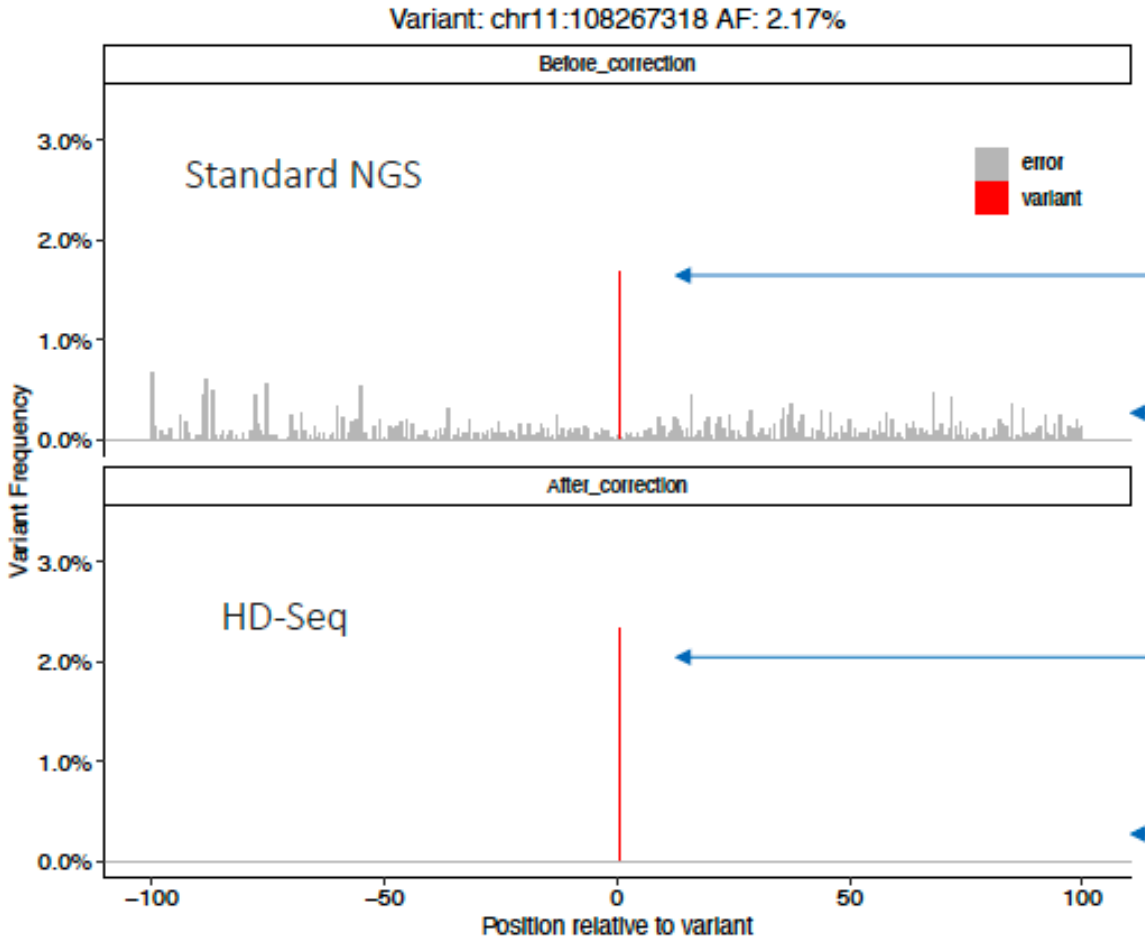


99.8% Single-Read Accuracy → 99.999% HD-Seq Accuracy (Q50)

- HD-Seq eliminates many sources of errors: sequencing errors, PCR amplification errors, and errors due to DNA damage

HIGH-DEFINITION SEQUENCING (HD-SEQ)

DOUBLE-STRANDED SEQUENCING FOR RARE VARIANT DETECTION



True variant, present at ~ 2% AF

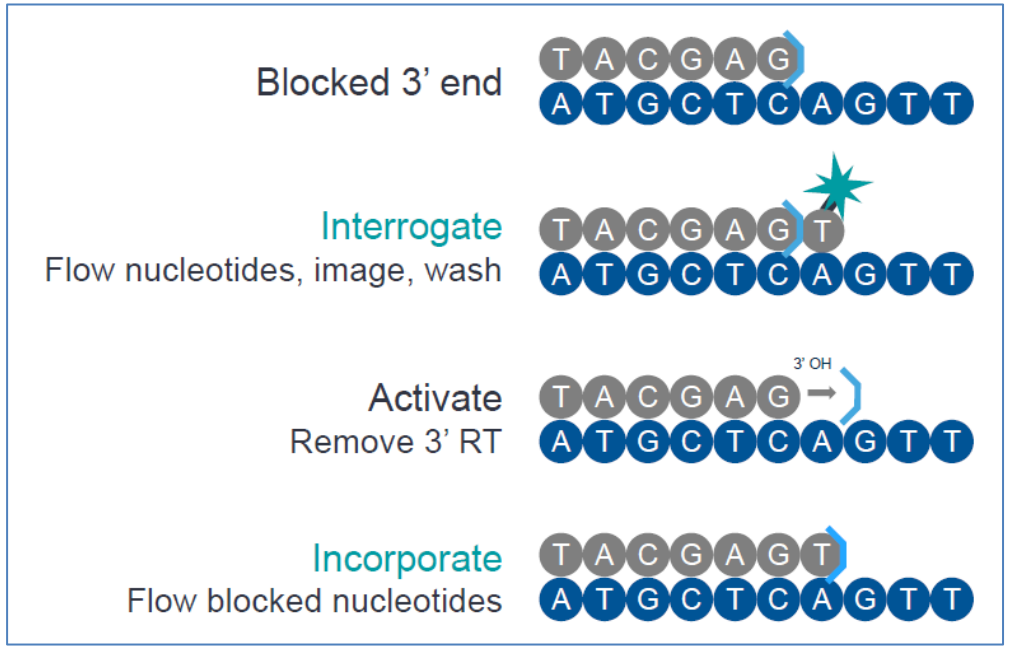
Background errors from sequencing, PCR amplification, and DNA damage

True variant, present at ~ 2% AF

Greatly reduced error rate

Pacific Biosciences: Onso

- Sequencing By Binding (SBB) Technology: interrogation followed by incorporation
  - Accuracy: 90% of bases >Q40
  - Improved homopolymer accuracy
- Output: Up to 500M paired-end reads
  - Demultiplexed FASTQ
- Runtime: <24 hours
- Length of reads: 2x150, or 1x200
- 50-65Gb output
- Accuracy:
  - Google DeepVariant calling
  - 99.25% Indels
  - 99.7% SNVs



(1) Each cycle initiates with a 3' reversible blocked nucleotide.
(2) Fluorescently labelled nucleotides are then flowed over the flow cell allowing for the appropriate base to bind. Unbound nucleotides are washed away so that the base can be interrogated with reduced background signal.
(3) The 3' end of the nucleotide is activated via removal of the reversible terminator.
(4) Native, unlabeled, reversible blocked nucleotides incorporate the cognate base into the growing strand.