August 4, 2022

DNA Sequencing at Scale

Flash Memory Summit

DNA DATA STORAGE ALLIANCE
A SNIA Technology Affiliate

Boyan Boyanov

Sr. Principal Scientist
Advanced Platforms and Devices
Illumina Research & Development

illumina®

Illumina Overview

1998
Year founded

>9,100
Number of employees

$4.5 billion (2021)
Annual revenue

Low-throughput


MiSeq™


MiniSeq


iSeq 100

Mid-throughput


NextSeq 500


NextSeq 550


NextSeq 1000/2000

High-throughput


NovaSeq™ 6000


HiSeq 2000


HiSeq 2500


HiSeq X Ten

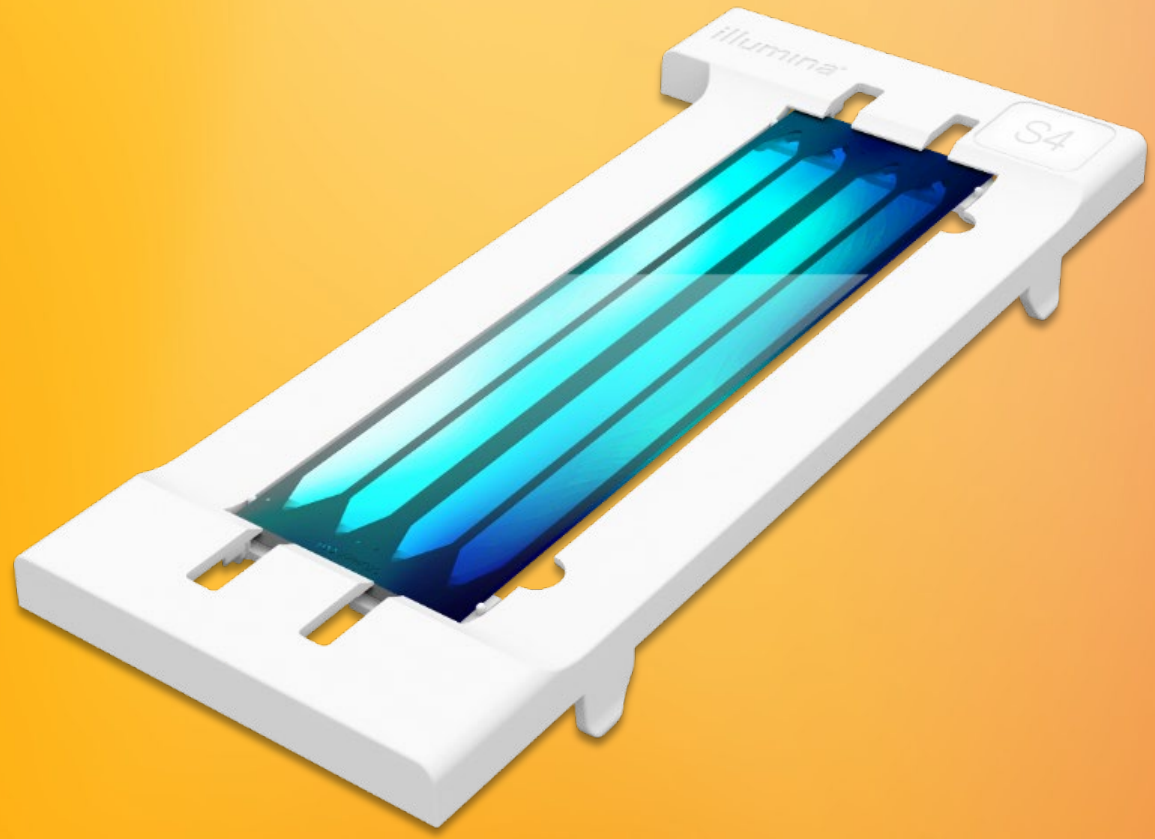

HiSeq 3000/4000

illumina®

Agenda

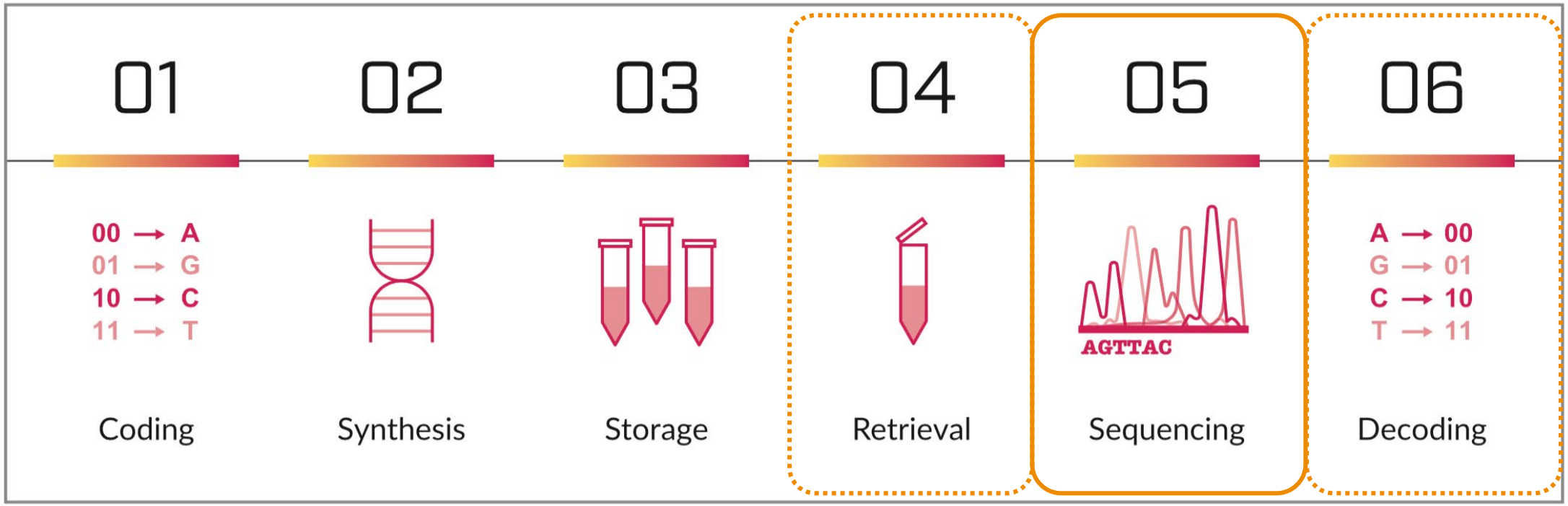
- Introduction to sequencing by synthesis (SBS)
- Sequencing at scale today
- Scaling output to meet the needs of DNA-based data storage

Sequencing by Synthesis (SBS)



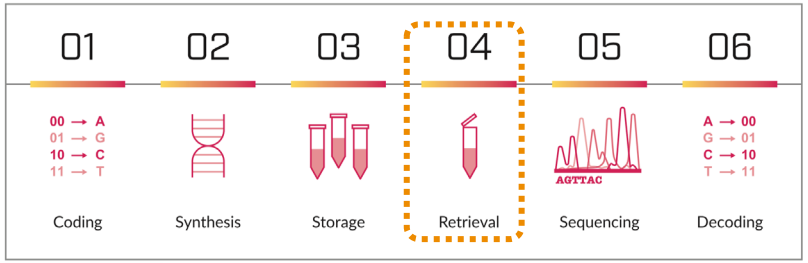
illumina®

Where does sequencing fit in the data storage pipeline?



* Graphic courtesy of DNA Data Storage Alliance Whitepaper (https://dnastoragealliance.org/publications/)

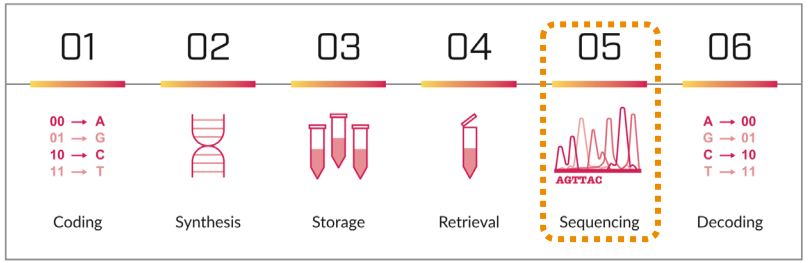
DNA retrieval (library preparation)

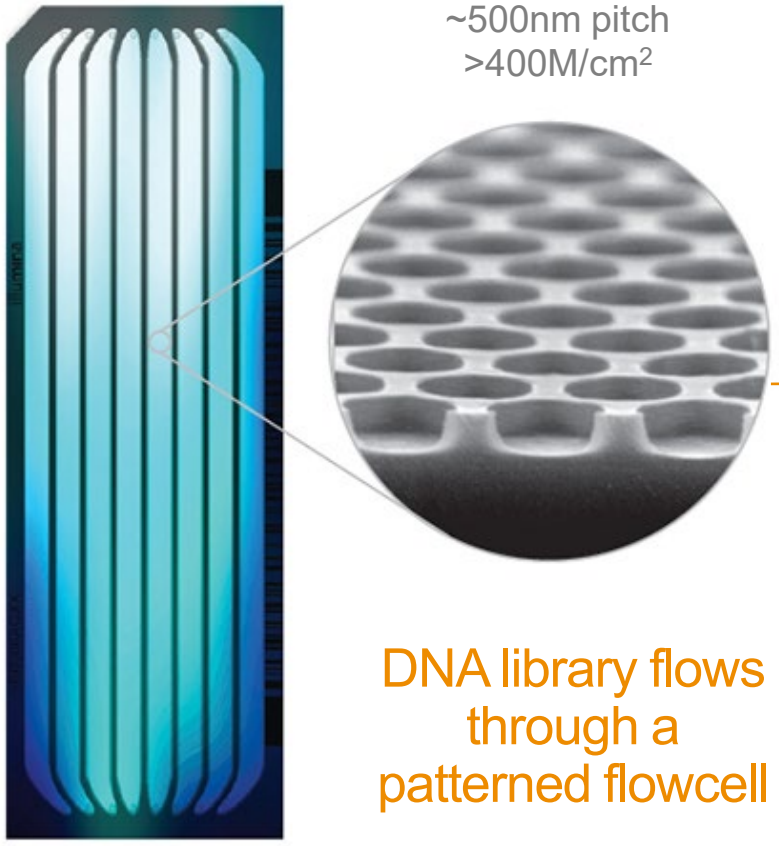


For Clustering	For Multiplexing	For Sequencing
Adapters (P5 and P7) bind the DNA to the surface of the flow cell	Unique barcode indices enable multiplexing and increased parallelism	Additional regions (Rd1 and Rd2) are required to initiate the sequencing

Reading out DNA requires special-purpose blocks to flank the encoding region

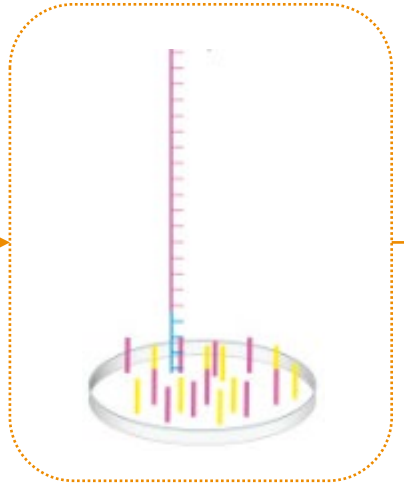
Prep for sequencing (cluster generation)



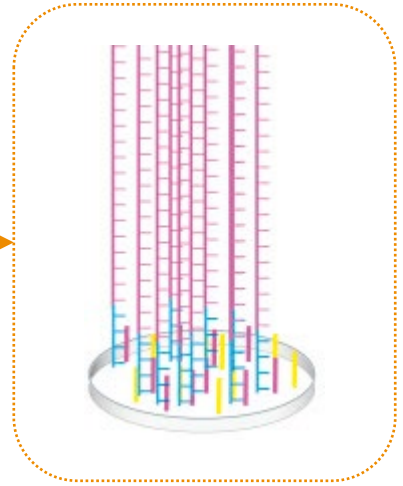


~500nm pitch
>400M/cm^2

DNA library flows through a patterned flowcell



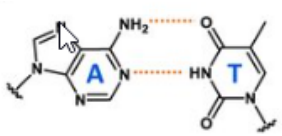
Fragments attach to nanowells on the flowcell surface

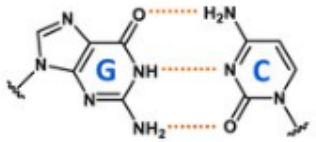


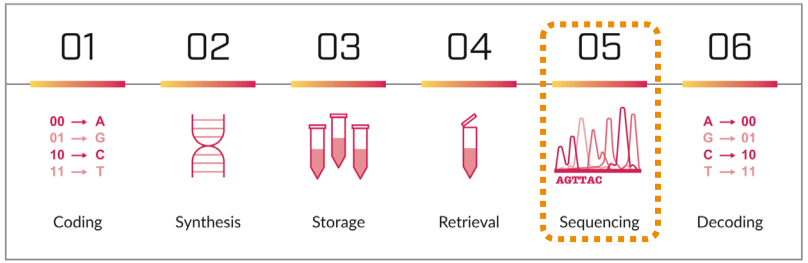
Each fragment is amplified ~ 1000x to create a monoclonal cluster

Clusters are the primary information carriers that are used for sequencing

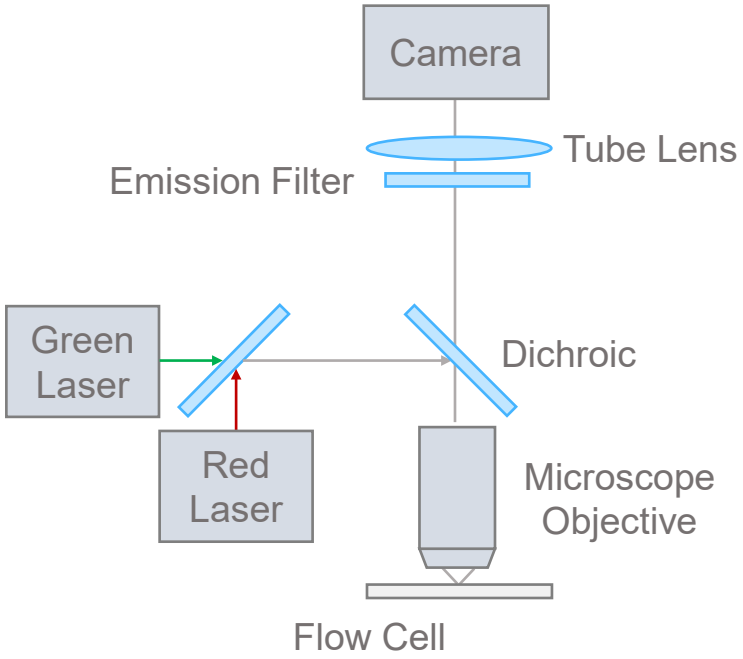
Sequencing

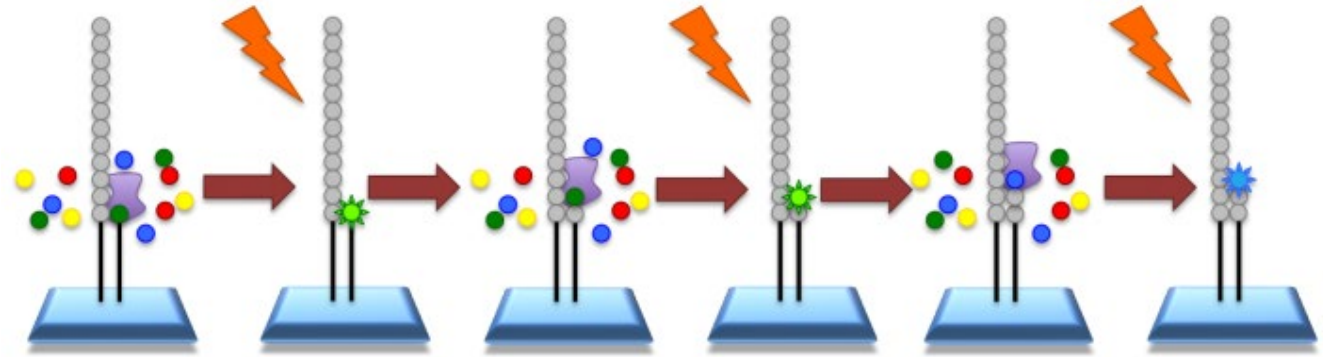






https://www.ebi.ac.uk/





Polymerase incorporates nucleotide

Fluorescent tag allows individual base to be ID'd

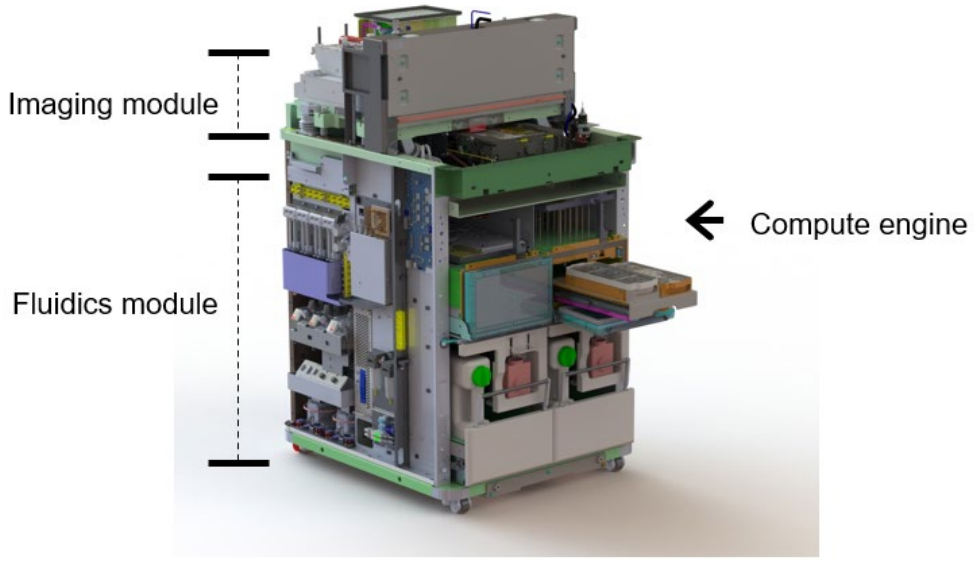
Chemical modifications prevents incorporation of more than one base

NovaSeq™ 6000

Highest throughput DNA Sequencer on the market

Highest capacity flowcell (S4) generates 3T bases in 44 hours
Instrument can run two S4 flow cells simultaneously

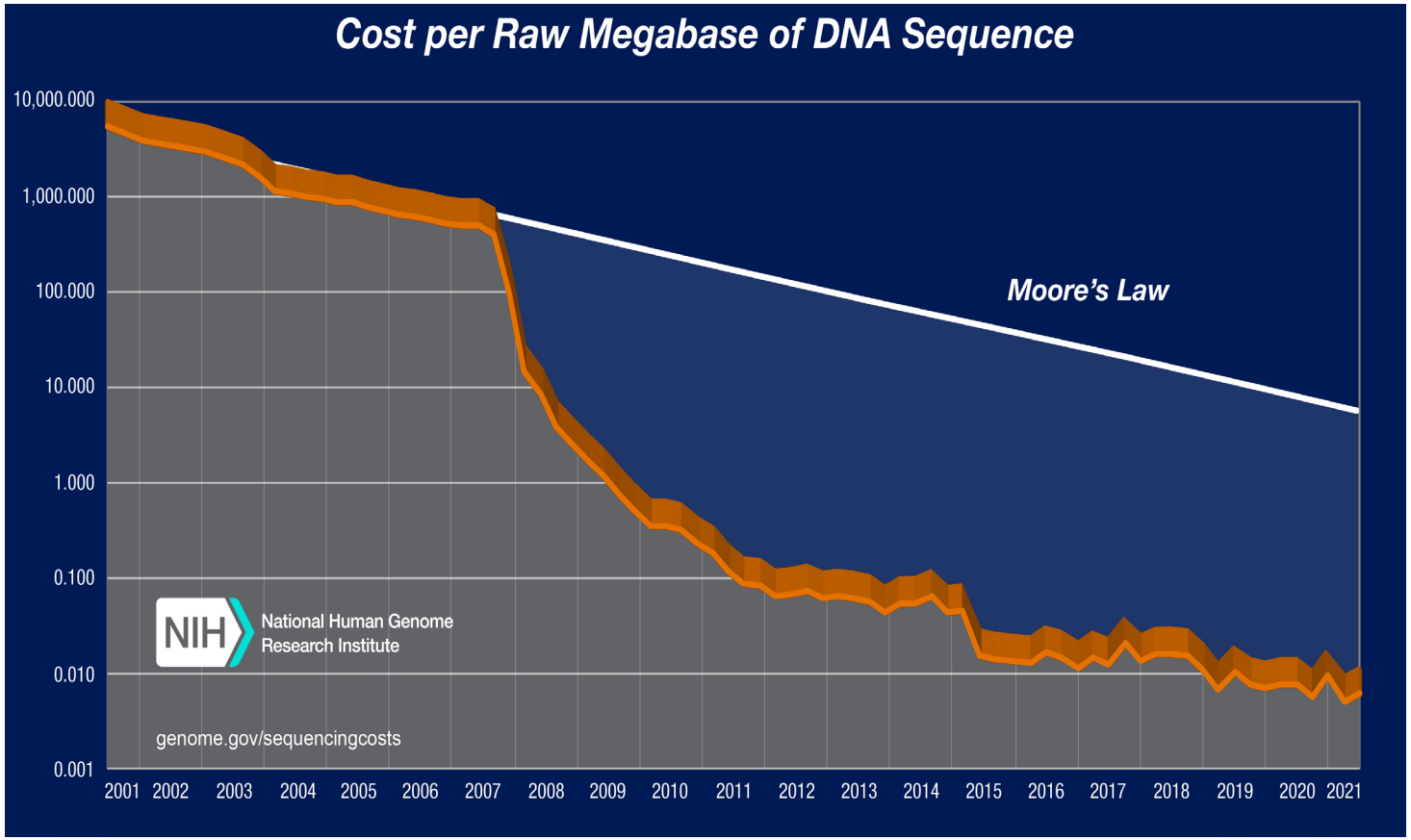
2 x 3T bases = 60 complete human genomes



Illumina's High-Throughput Sequencing Lab in Hinxton, UK

Current capacity of ~5Pb bases of sequencing data per year*

Cost of DNA data generation



NovaSeq 6000 today offers < $6/Gbase of raw data

Direct line of sight exists to delivering $1/Gbase

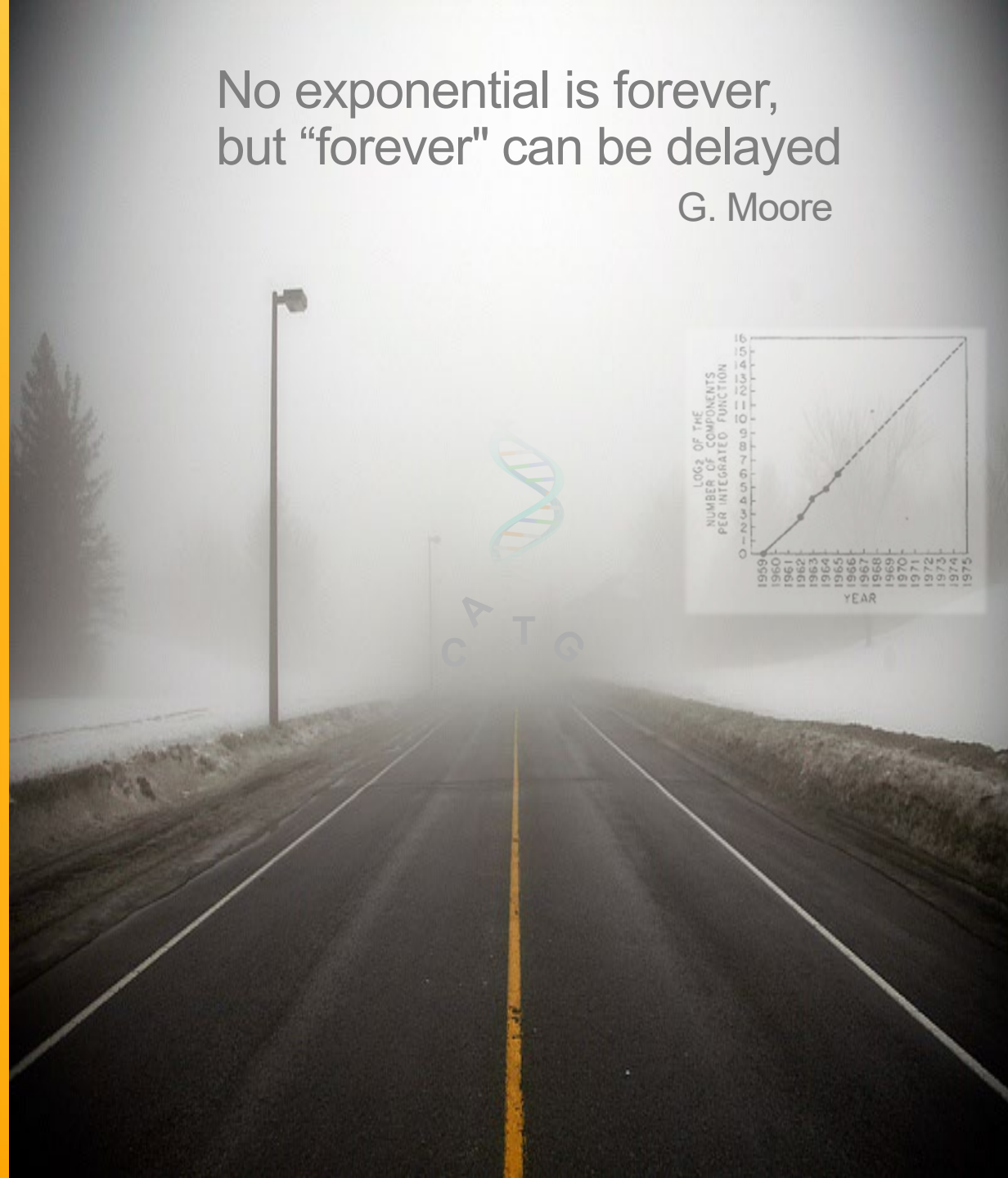
No conceptual hurdles to $0.1/Gbase

https://www.genome.gov/about-genomics/fact-sheets/Sequencing-Human-Genome-cost (downloaded May 2022)

Peering Through The Fog

The path to LTO-like cost

illumina®



The red brick wall is always 10 years away

*Table 46b MPU Interconnect Technology Requirements—Long Term Years***

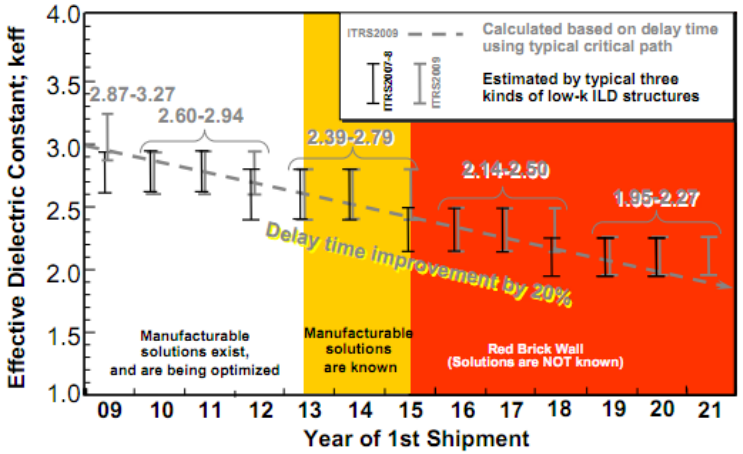
YEAR / TECHNOLOGY NODE	2008 70 nm	2011 50 nm	2014 35 nm
MPU ½ pitch	80	55	40
MPU gate length (nm)	45	33	22
Number of metal levels	9	9–10	10
Number of optional levels – ground planes/capacitors	3	4	4
Total interconnect length (m) – active wiring only (footnote for calculation)	51730	91532	148835
FITS/m X 10E-3 (fitting footnote)	0.10	0.05	0.03
Jmax (A/cm²)—wire (at 105°C)	2.1E6	3.7E6	4.6E6
Imax (mA)—via (at 105°C)	0.18	0.16	0.11
Local wiring pitch (nm)	185	130	95
Local A/R (for Cu)	1.9	2.1	2.3
Cu local dishing (nm), 5% × height	9	7	5
Intermediate wiring pitch (nm)	240	165	115
Intermediate wiring dual damascene A/R (Cu wire/via)	2.5/2.3	2.7/2.4	2.9/2.5
Cu intermediate wiring dishing (nm), 15 micron wide wire, 10% × height	30	22	17
Minimum global wiring pitch (nm)	390	275	190
Global wiring dual damascene A/R (Cu wire/via)	2.8/2.9	2.9/3.0	3.0/3.1
Cu global wiring dishing (nm), 15 micron wide wire, 10% × height	55	38	29
Conductor effective resistivity (μΩ-cm) Cu wiring	2.2	<1.8	<1.8
Barrier/cladding thickness (nm)	0	0	0
Barrier/cladding thickness (nm)	7	5	4
Interlevel metal insulator— effective dielectric constant (κ)	1.5	<1.5	<1.5
Interlevel metal insulator— effective dielectric constant (κ)	1.6	<1.6	<1.3
Interlevel metal insulator (minimum expected) — bulk dielectric constant (κ)	1.3	<1.3	1.1

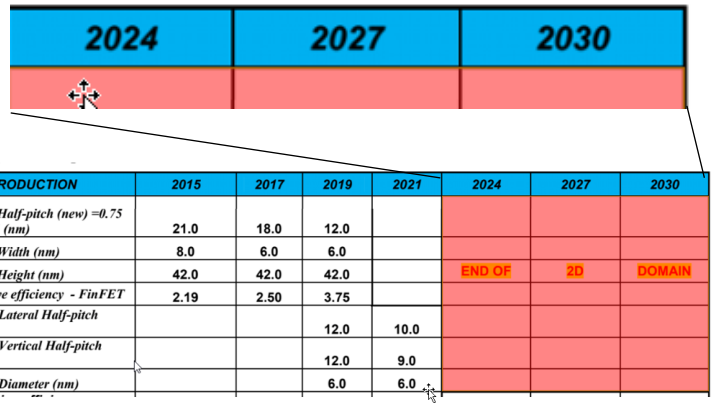
xist — Solutions Being Pursued — No Known Solutions

2008 70 nm	2011 50 nm	2014 35 nm
80	55	40

Depth of field will be reduced to about ± 0.2μ. Deep U.V. (λ = 200nm - 260nm) lenses will be difficult to build because of the lack of materials that are transparent at these wavelengths and yet have relatively high refractive indices. Elements made

Lithography Scaling Limitations
From Broers [1] IEDM Plenary Session 1980





2024	2027	2030

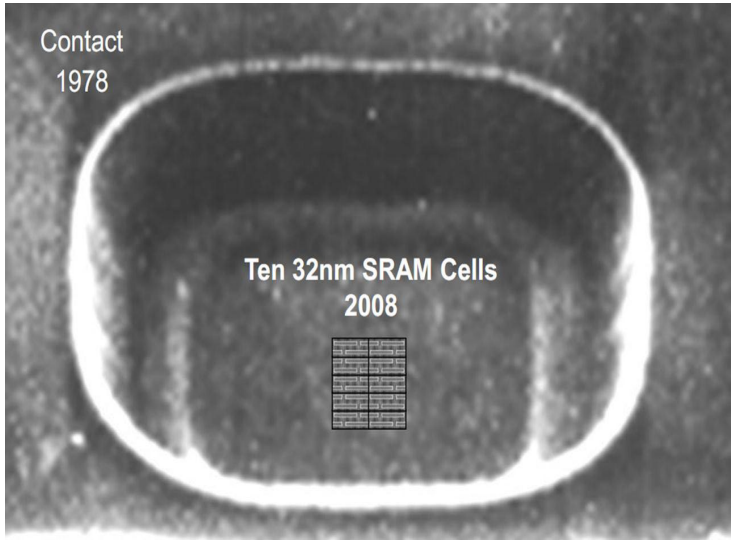
YEAR OF PRODUCTION	2015	2017	2019	2021	2024	2027	2030
FinFET Fin Half-pitch (new) =0.75 or 1.0 M0/M1 (nm)	21.0	18.0	12.0				
FinFET Fin Width (nm)	8.0	6.0	6.0				
FinFET Fin Height (nm)	42.0	42.0	42.0		END OF	2D	DOMAIN
Footprint drive efficiency - FinFET	2.19	2.50	3.75				
Lateral GAA Lateral Half-pitch (nm)			12.0	10.0			
Lateral GAA Vertical Half-pitch (nm)			12.0	9.0			
Lateral GAA Diameter (nm)			6.0	6.0			

ITRS 2000
(interconnect)

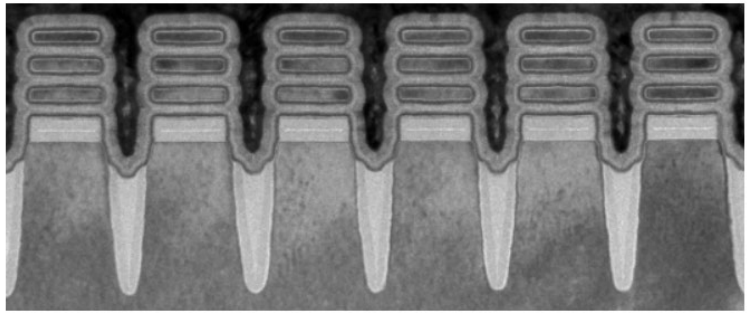
ITRS 2009
(Dielectrics)

ITRS 2015
(FinFET)

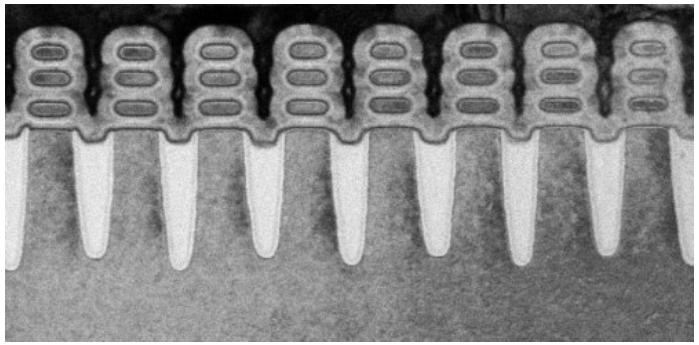
Yet the semiconductor industry keeps marching on



M. Bohr, S. Sivakumar, Intel



Cross section of IBM 2nm silicon manufacturing process. Source: IBM.

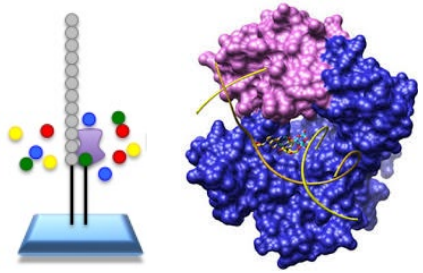


Samsung

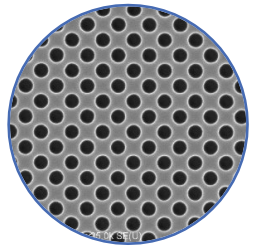
Our track record of predicting technology cliffs is not great

Drivers of sequencing cost for data storage

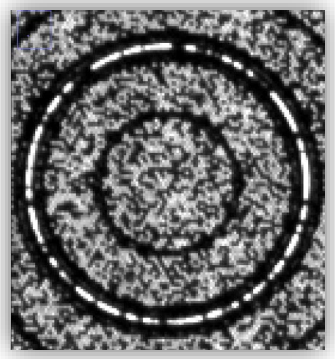
Library prep



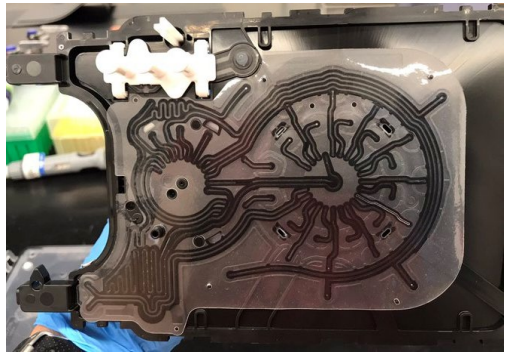
Enzymatics



Cluster density



Imaging time and resolution



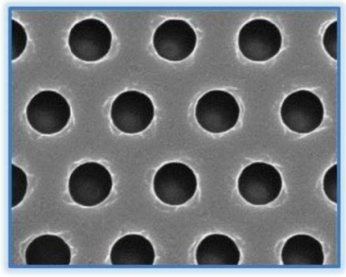
Speed and efficiency of fluidic exchanges

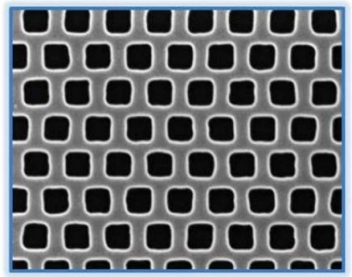
Compute

Increasing cluster density

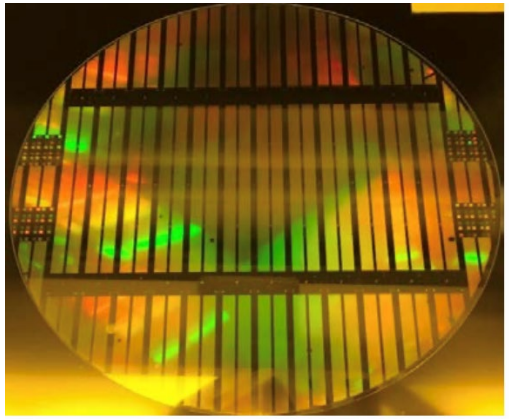
300mm substrates, NIL, and 5x well density increases are expected to reduce flow cell COGS reductions by up to 90%

Flow cell micrographs of same unit area

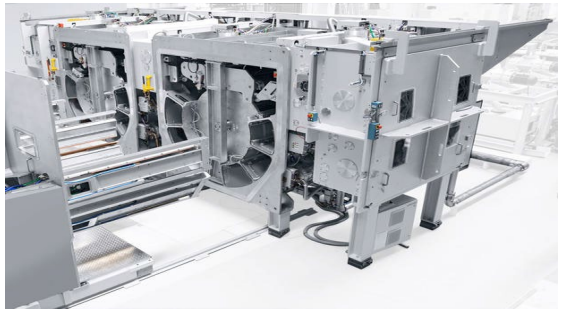




Density improvements



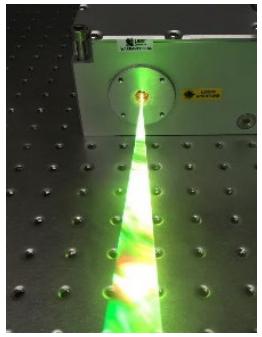
300mm wafers and low-cost patterning



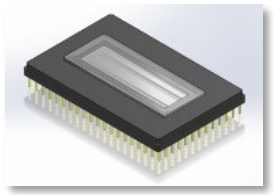
https://www.appliedmaterials.com/products

Further reduction in the COGS is possible by leveraging progress in large format substrate technologies

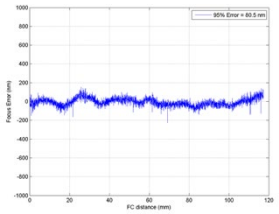
Pushing past the diffraction and SNR limits



High Power, multi-watt Lasers

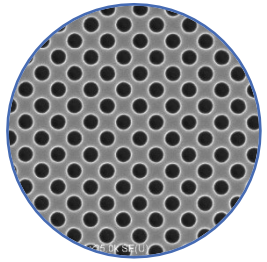


Wide-format TDI CCD camera

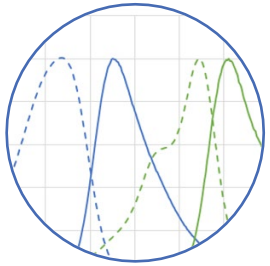


Nanometer-scale laser-based focus tracking

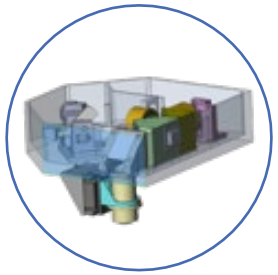
Patterned Flow Cell



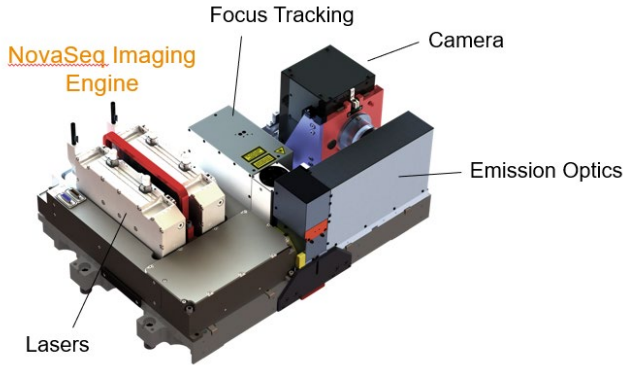
Blue-green imaging

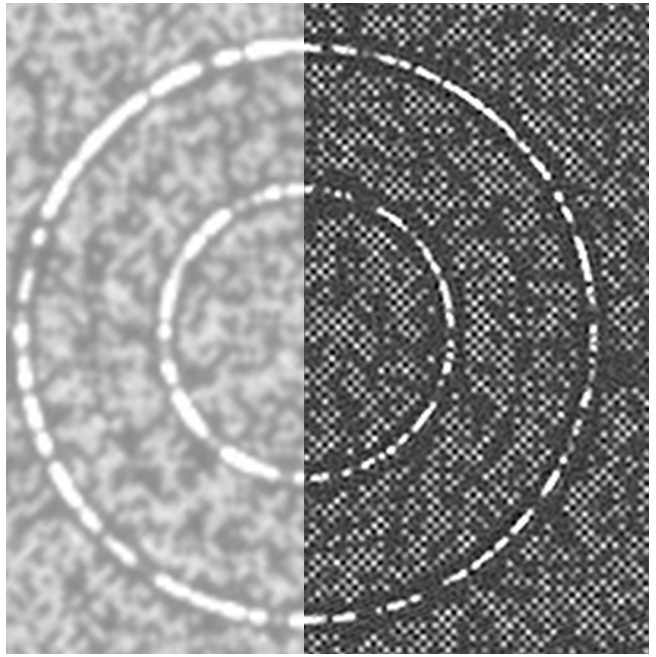


Super-Resolution Optics



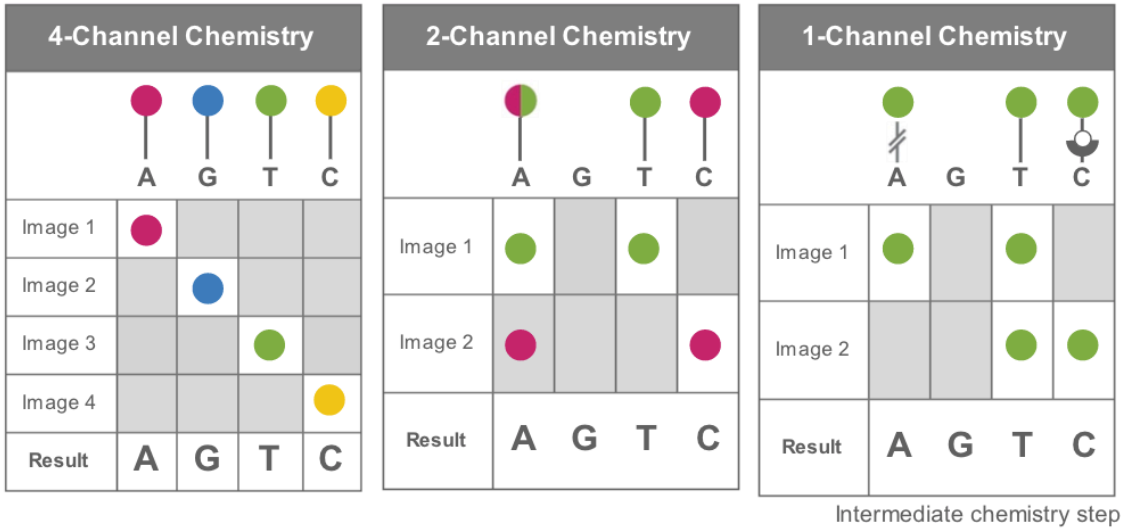
NextSeq™ 2000





Reducing imaging time

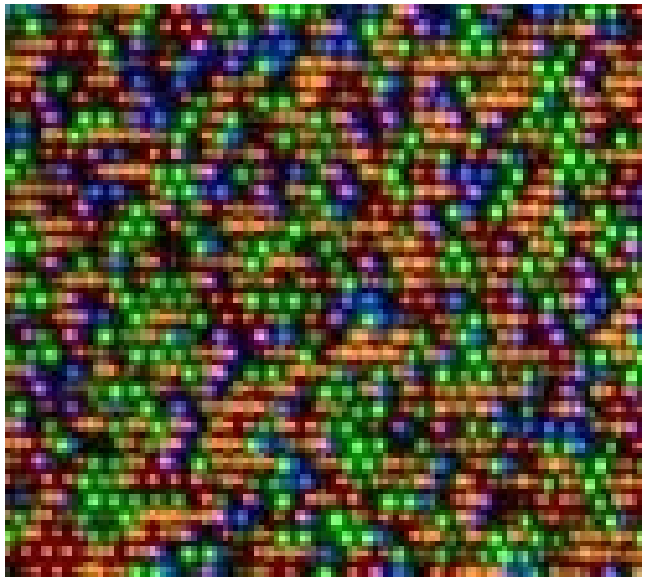
50% reduction in imaging time with biochemistry



Uses four fluorescent dyes (one for each base), and four images per sequencing cycle.

Uses two fluorescent dyes and two images per cycle to determine all four base calls.

Uses CMOS technology to determine base calls using two images per cycle.



4-color image of a patterned flowcell

Continuing improvements and cost reduction in high-speed area imagers and motion stages provides further benefits

Liquid distribution and waste management

~10x reduction in reagent volumes reduces cost and waste

~30% reduction in number of reagents further reduces size and simplifies design

single storage condition with integrated cartridge

Cartridge as delivered

Cosmetic cover removed

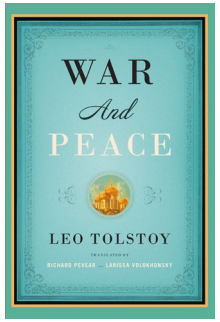
Piercing plate removed

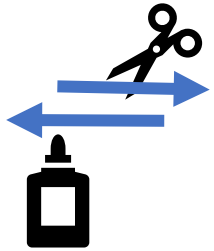
Underside reagent distribution layer

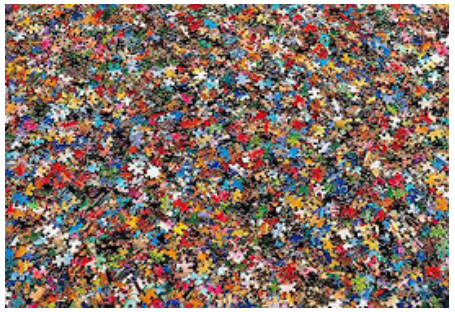
Reduced reagent consumption and improved reagent reuse are a major driver for continued cost reduction

Hardware acceleration of genomic analysis

The DRAGEN Platform uses reconfigurable field-programmable gate array technology (FPGA) to provide hardware-accelerated implementations of genomic analysis algorithms







Hardware Acceleration: Provides ultra-efficient workflow; can fully process 4Tbases in ~36 minutes

Lossless Compression: Reduction of FASTQ file sizes by up to 5x decreases operational and storage costs

Continuous Innovation: Graph reference genome and machine learning driving unprecedented accuracy

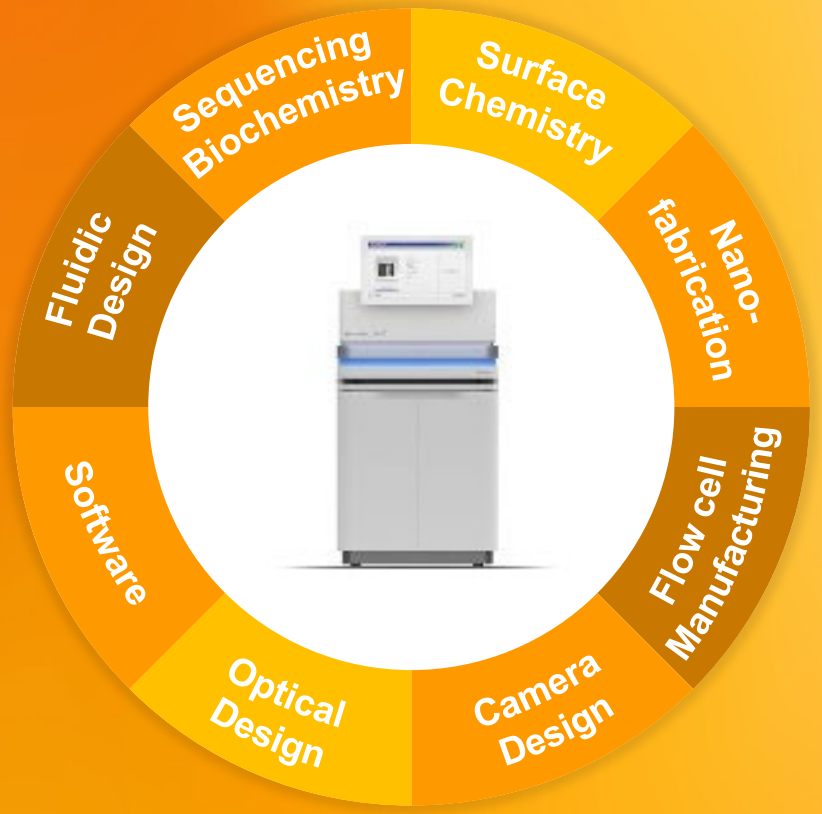
Summary

- Sequencing by Synthesis (SBS) based platforms offer the highest throughput and lowest cost sequencing in the market today

- Roadmaps exist for all primary drivers of sequencing cost without perceived red brick walls down to at least $0.1/Gbase

- Roadmaps for cost drivers indicate sufficient headroom for further cost decrease by leveraging ongoing developments in semiconductors, light sources, imagers, fluidic technology and compute infrastructure

Orthogonal disciplines working together



The multi-disciplinary nature of the technology suggests that the super-exponential reduction in sequencing costs is likely to persist well into the future

Thank you!

Boyan Boyanov

Sr. Principal Scientist
Advanced Platforms and Devices
Illumina Research & Development

illumina®