

A C G T

DNA-Sequencing

Technologies & Devices

Matthias Platzer

Genome Analysis
Leibniz Institute on Aging
- Fritz Lipmann Institute (FLI)

Genome analysis

DNA sequencing platforms

ABI 3730xl

4/2004 & 6/2006



1 Mb/day, 850 nt reads
2 Mb/day, 550 nt reads

Illumina MiSeq

1/2013



7,5 Gb/1d, 2x250 nt reads

HiSeq 2500

1/2013



2x300 Gb/10d, 2x100 nt reads
2x100 Gb/2d, 2x150 nt reads

PacBio RS II

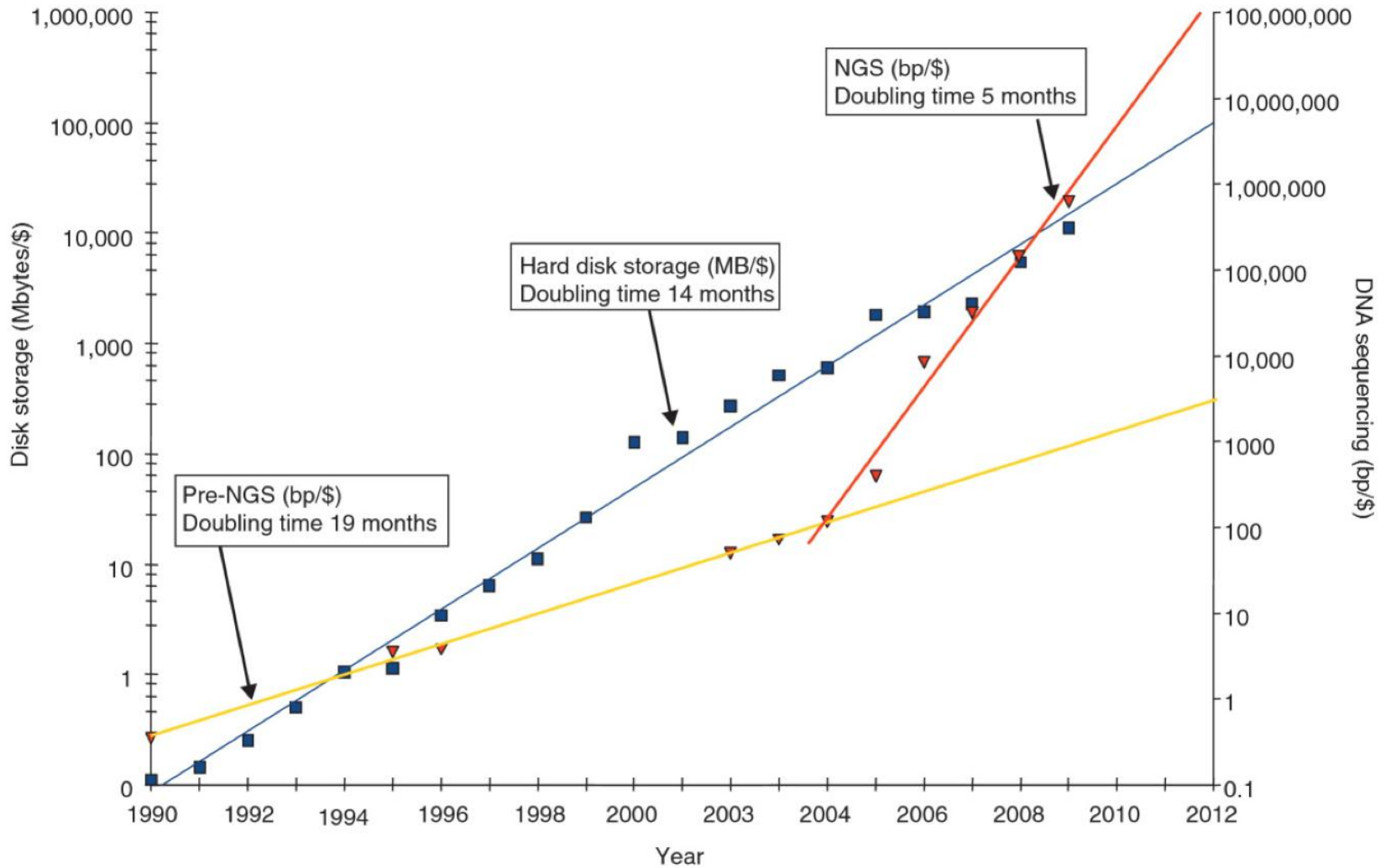
2/2015



0.5 Gb/6h, 10..40kb nt reads

Moore's law

Next Generation Sequencing - a game changer

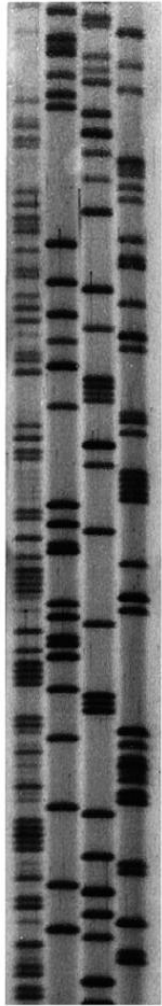


Next Generation Sequencing - NGS

Terminologies

2nd generation
= massively parallel

3rd generation
= single-molecule



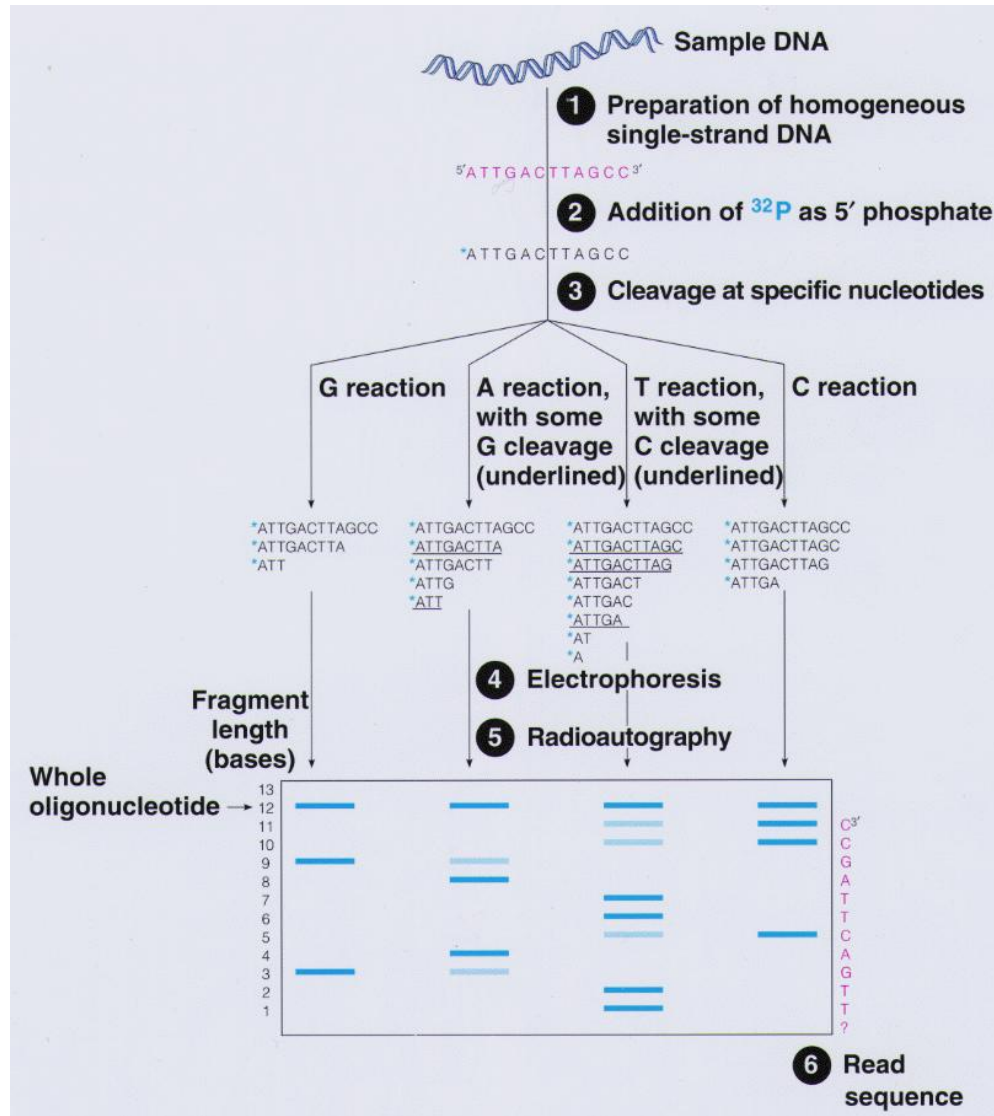
A C G T

1st Generation

Maxam-Gilbert . Sanger . Pyro

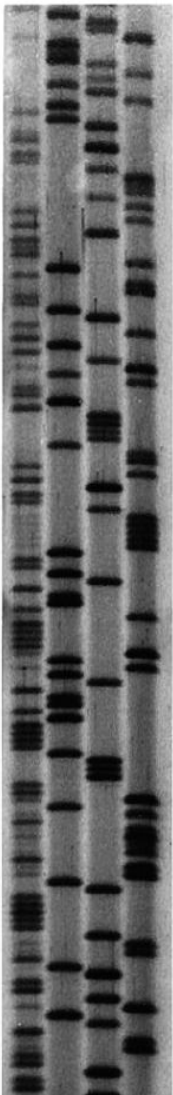
Maxam-Gilbert sequencing

Chemical modifications, radioactive *1977

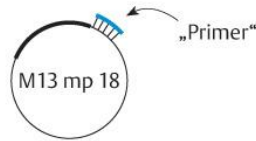


Sanger sequencing

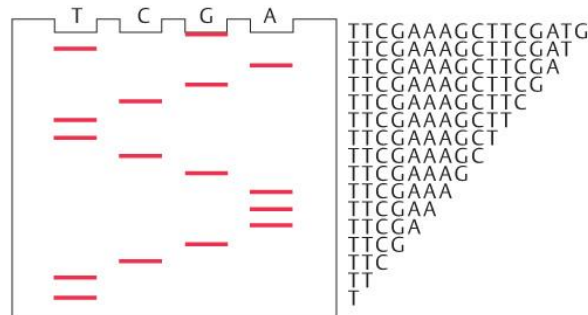
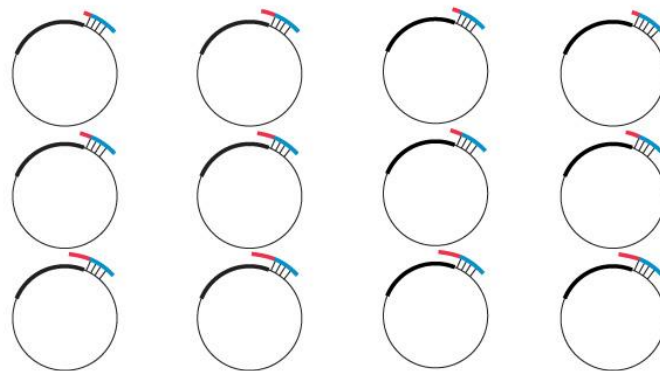
Dideoxy chain termination, radioactive *1977



A C G T



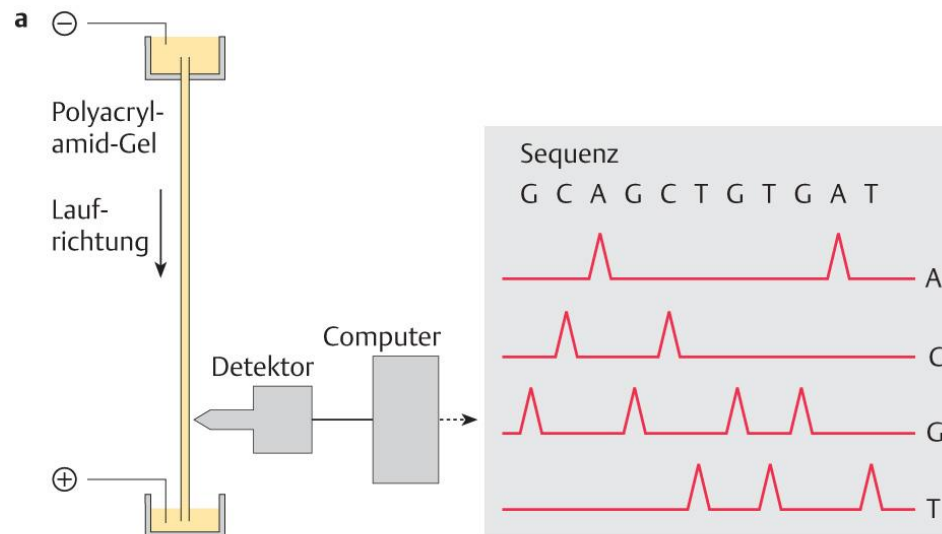
α (^{32}P) dATP	α (^{32}P) dATP	α (^{32}P) dATP	α (^{32}P) dATP ddATP
dGTP	dGTP	dGTP ddGTP	dGTP
dCTP	dCTP ddCTP	dCTP	dCTP
dTTP ddTTP	dTTP	dTTP	dTTP



Fred Sanger
Nobel prices 1958, 1980

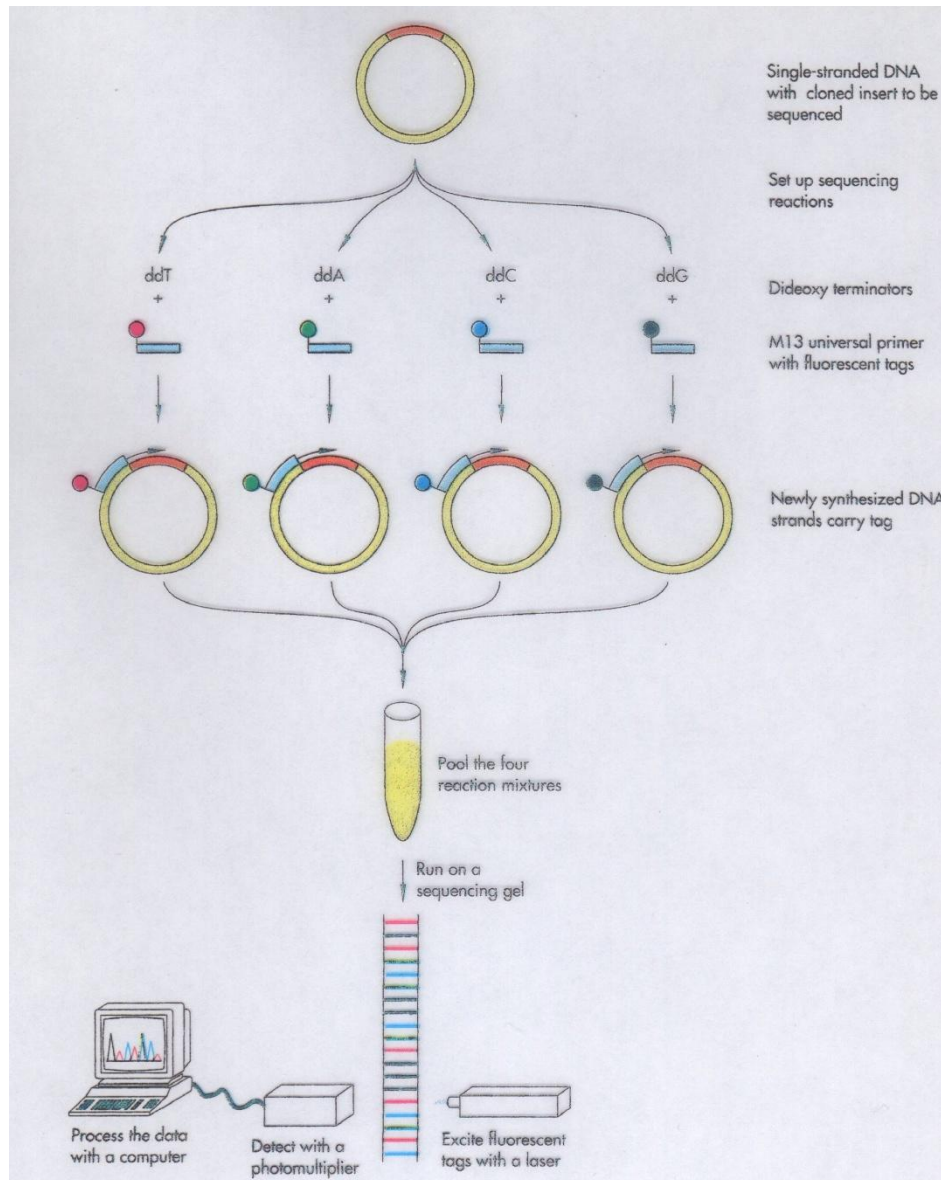
Sanger sequencing

Fluorescence labeling



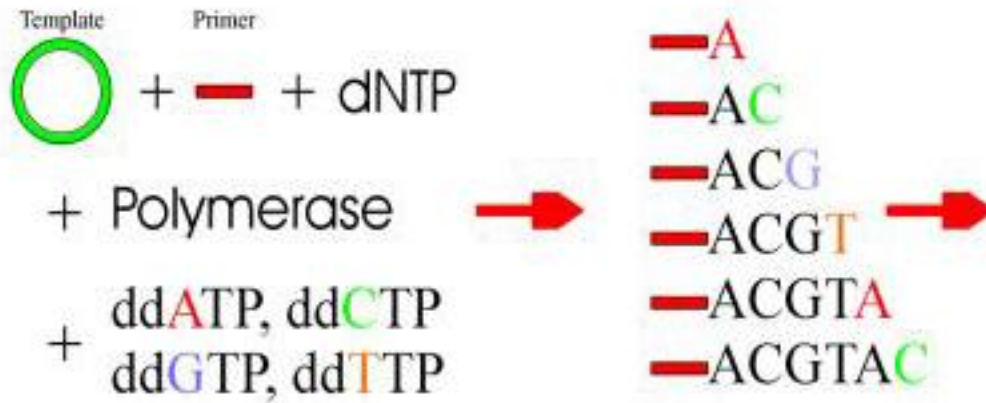
Sanger sequencing

Dye primer



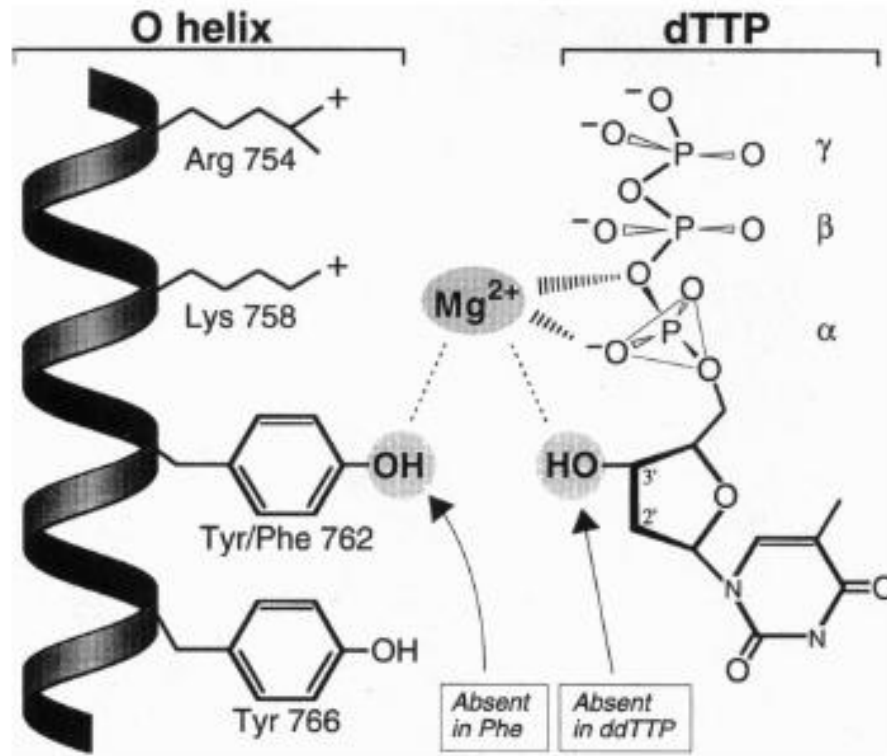
Sanger sequencing

Dye terminator



Sanger sequencing

Engineered polymerase



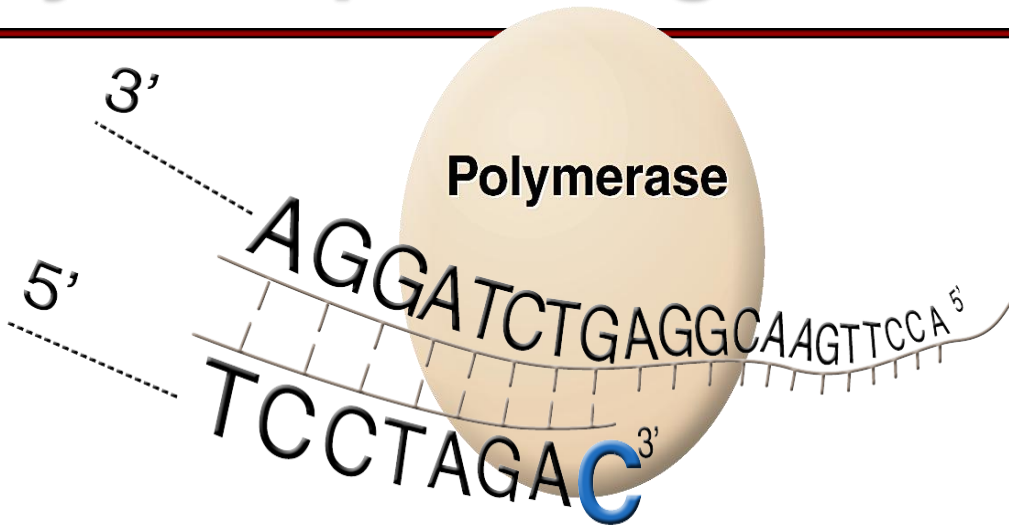
Thermo-Sequenase = Taq polymerase Phe>Thy mutated

Sanger sequencing

Cycle sequencing



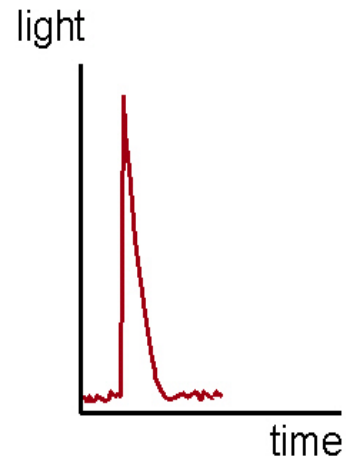
Pyrosequencing



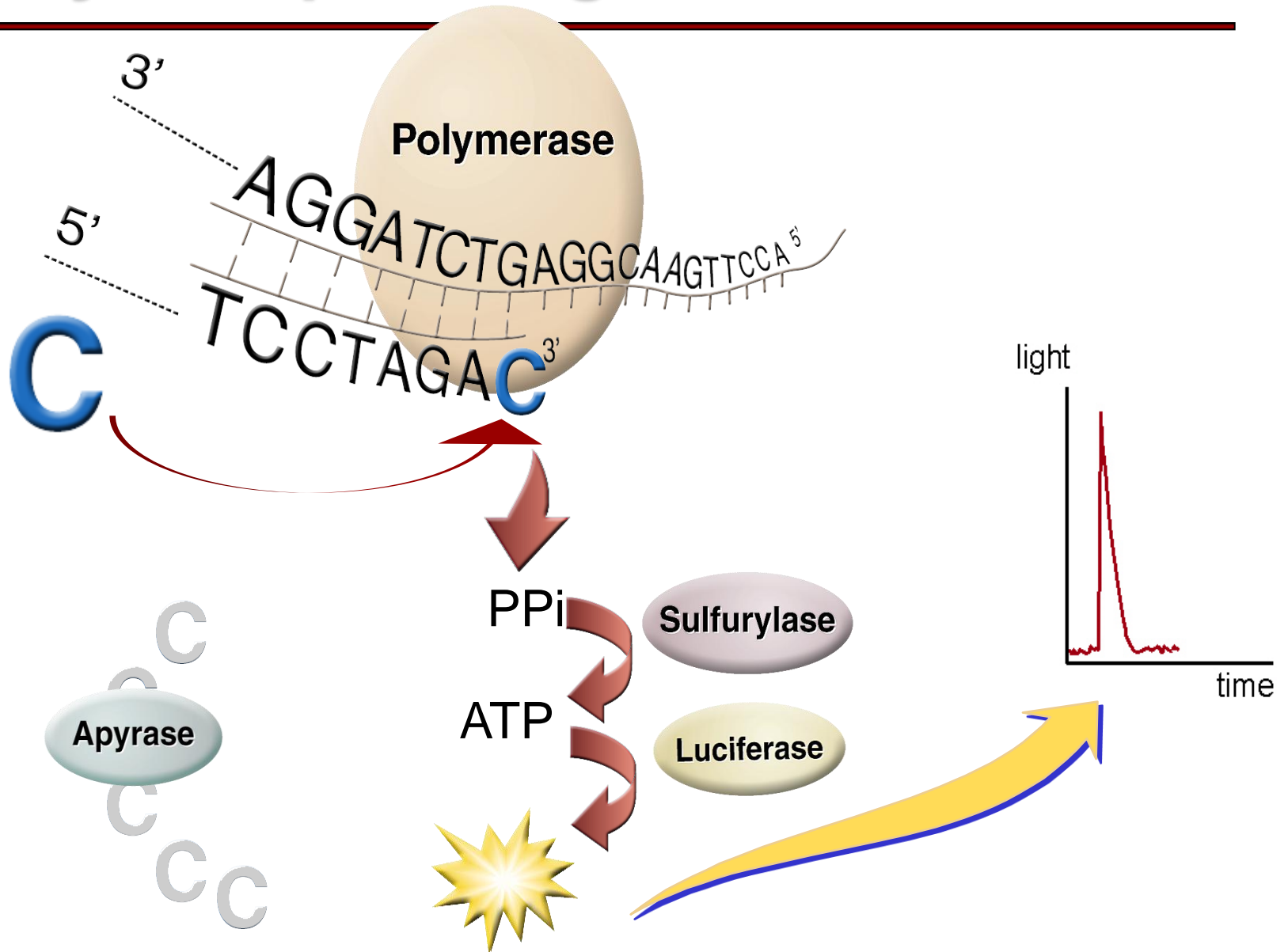
Apyrase

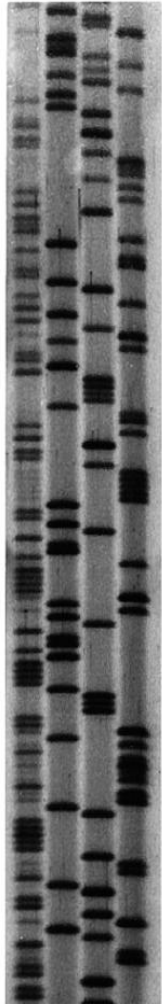
Sulfurylase

Luciferase



Pyrosequencing





A C G T

2nd Generation

Roche/454. Illumina/Solexa . ABI/Solid
ABI/IonTorrent . Complete Genomics

454

Genome Sequencer GS20/FLX/FLX+

**A Revolutionary Method
for Rapid Whole Genome
Sequencing and Assembly
And more.....**



ARTICLES

Genome sequencing in microfabricated high-density picolitre reactors

Marcel Margulies^{1*}, Michael Egholm^{1*}, William E. Altman¹, Said Attiya¹, Joel S. Bader¹, Lisa A. Bembgen¹, Jan Berka¹, Michael S. Braverman¹, Yi-Ju Chen¹, Zhoutao Chen¹, Scott B. Dewell¹, Lei Du¹, Joseph M. Fierro¹, Xavier V. Gomes¹, Brian C. Godwin¹, Wen He¹, Scott Helgesen¹, Chun He Ho¹, Gerard P. Irzyk¹, Szilveszter C. Jando¹, Maria L. I. Alenquer¹, Thomas P. Jarvie¹, Kshama B. Jirage¹, Jong-Bum Kim¹, James R. Knight¹, Janna R. Lanza¹, John H. Leamon¹, Steven M. Lefkowitz¹, Ming Lei¹, Jing Li¹, Kenton L. Lohman¹, Hong Lu¹, Vinod B. Makhijani¹, Keith E. McDade¹, Michael P. McKenna¹, Eugene W. Myers², Elizabeth Nickerson¹, John R. Nobile¹, Ramona Plant¹, Bernard P. Puc¹, Michael T. Ronan¹, George T. Roth¹, Gary J. Sarkis¹, Jan Fredrik Simons¹, John W. Simpson¹, Maithreyan Srinivasan¹, Karrie R. Tartaro¹, Alexander Tomasz³, Kari A. Vogt¹, Greg A. Volkmer¹, Shally H. Wang¹, Yong Wang¹, Michael P. Weiner⁴, Pengguang Yu¹, Richard F. Begley¹ & Jonathan M. Rothberg¹

The proliferation of large-scale DNA-sequencing projects in recent years has driven a search for alternative methods to reduce time and cost. Here we describe a scalable, highly parallel sequencing system with raw throughput significantly greater than that of state-of-the-art capillary electrophoresis instruments. The apparatus uses a novel fibre-optic slide of individual wells and is able to sequence 25 million bases, at 99% or better accuracy, in one four-hour run. To achieve an approximately 100-fold increase in throughput over current Sanger sequencing technology, we have developed an emulsion method for DNA amplification and an instrument for sequencing by synthesis using a pyrosequencing protocol optimized for solid support and picolitre-scale volumes. Here we show the utility, throughput, accuracy and robustness of this system by shotgun sequencing and *de novo* assembly of the *Mycoplasma genitalium* genome with 96% coverage at 99.96% accuracy in one run of the machine.

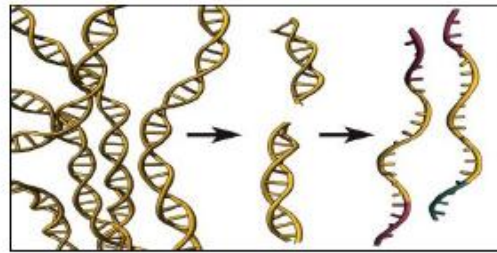
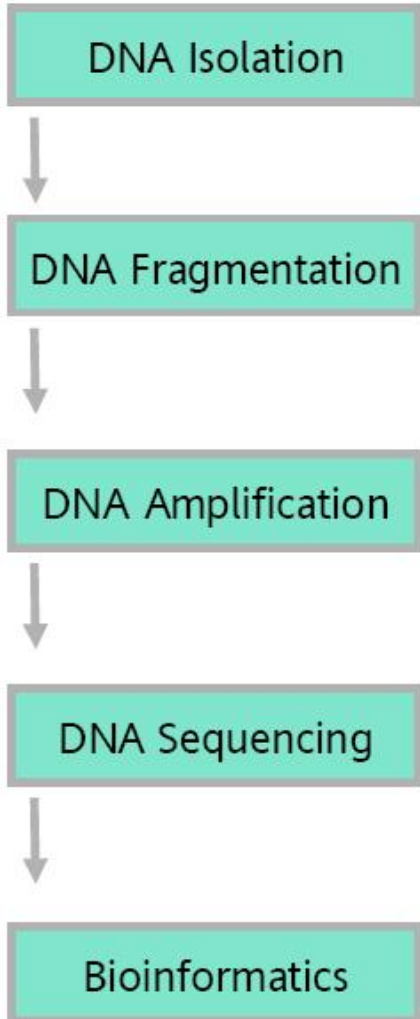
- 25 Mb
- 99% accuracy
- 4h

- *de novo* assembly
Mycoplasma genitalium (0.58 Mb): 96% coverage, 99,96% accuracy

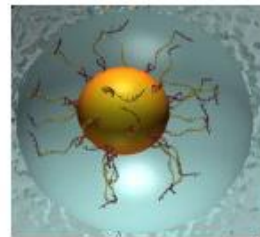
- 100x currant Sanger technology

Roche/454

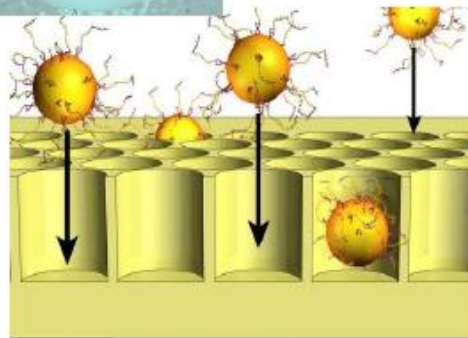
Workflow



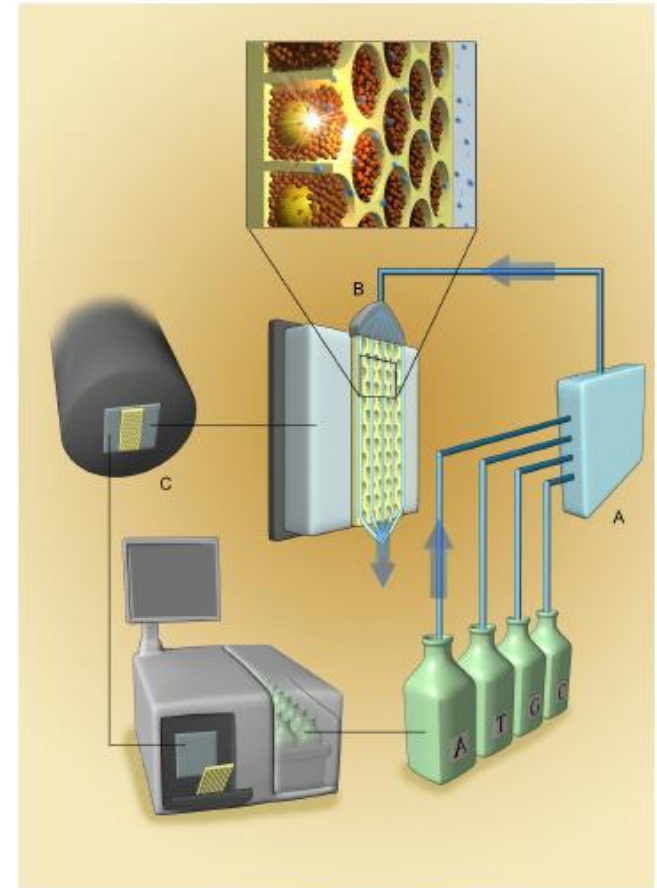
Adapter Ligated ssDNA Library



Clonal Amplification



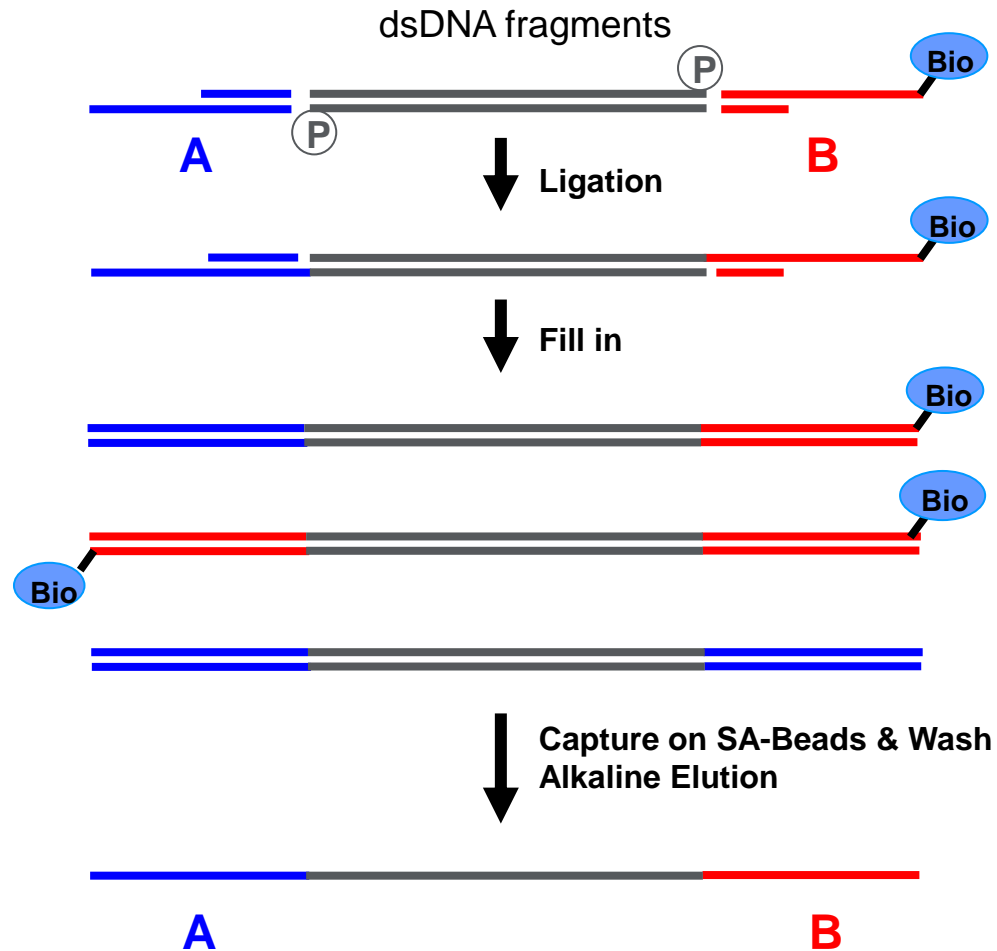
Beads enzymes in PicoTiter Plate™



- Contigs are constructed

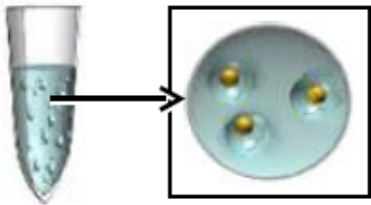
Roche/454

Preparation of A/B-fragments for sequencing

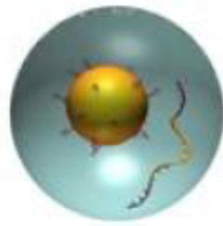


Roche/454 emPCR

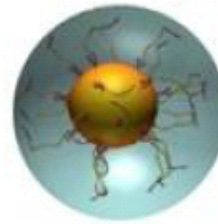
Emulsion-based clonal amplification



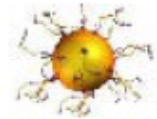
Anneal sstDNA
to an excess of
DNA Capture
Beads



Emulsify beads and
PCR reagents in
water-in-oil
microreactors



**Clonal amplification
occurs inside
microreactors**



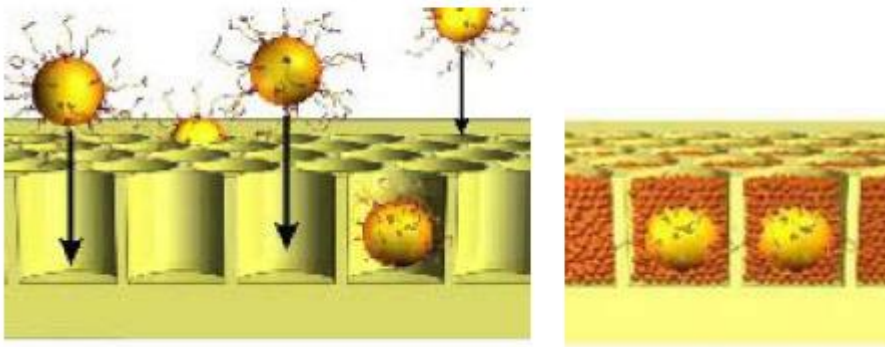
Break
microreactors,
enrich for DNA-
positive beads

sstDNA library → **Clonally-amplified sstDNA attached to bead**

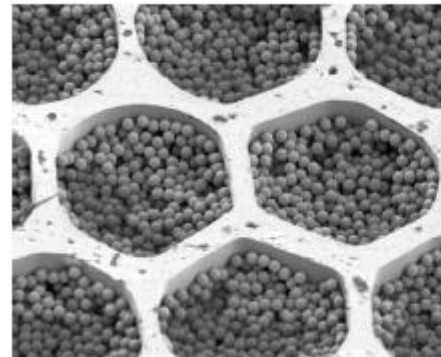
Roche/454

Sample loading

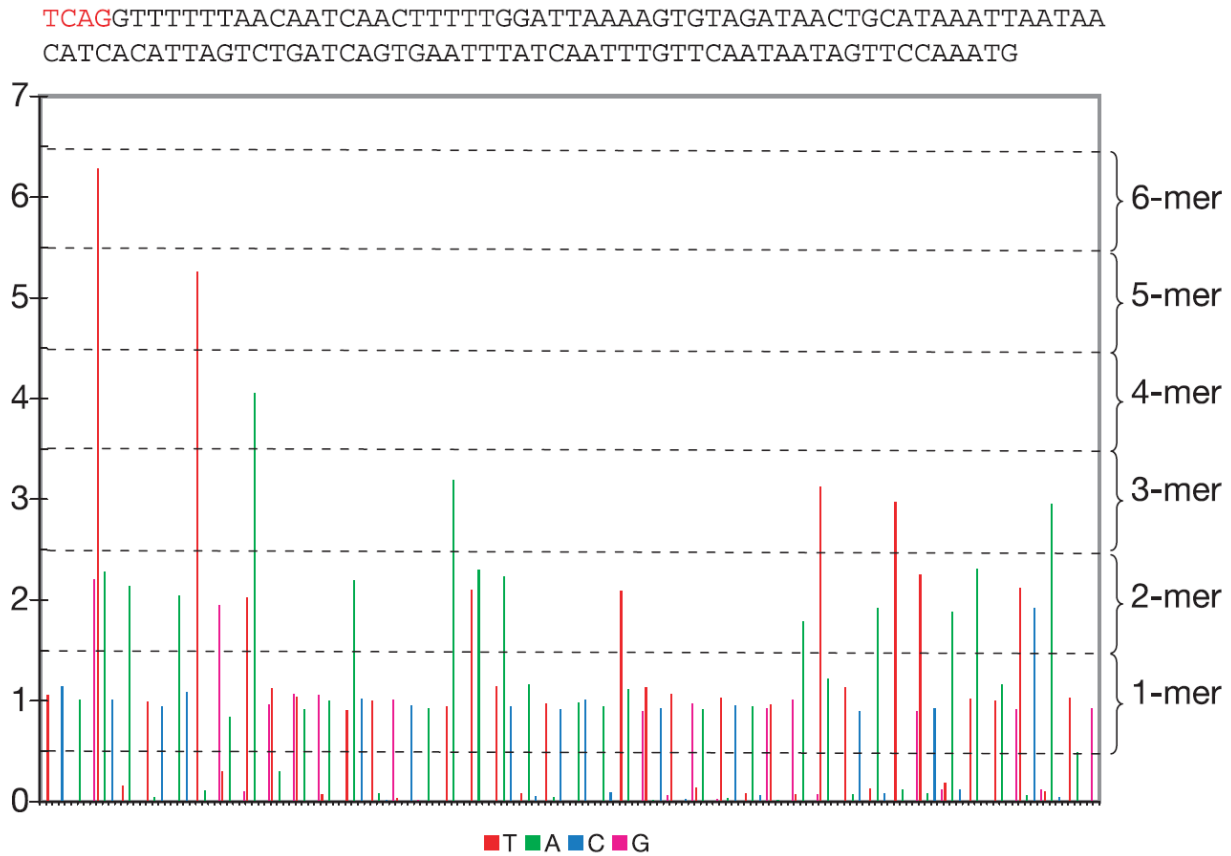
Depositing DNA beads into the PicoTiterPlate device



- Well diameter: average of 44 μm
- 200,000 reads obtained in parallel
- A single clonally amplified sstDNA bead is deposited per well



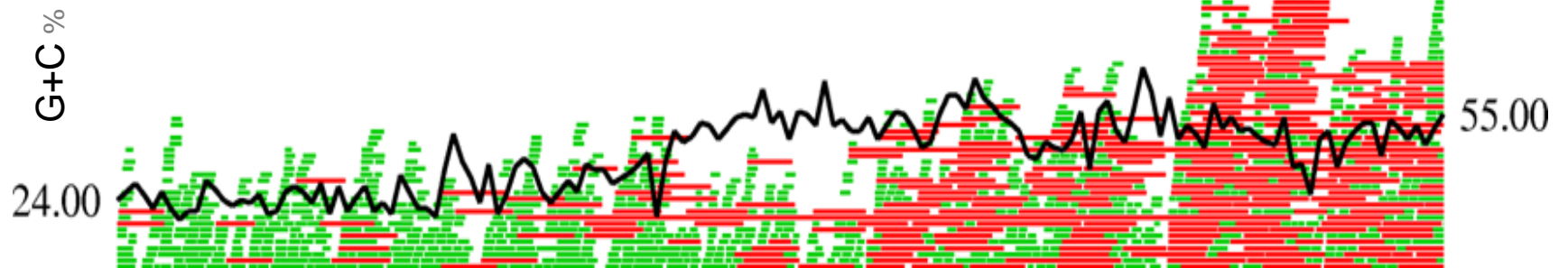
Roche/454 Flowgram



Roche/454

No cloning bias

green - 454 reads
red - Sanger reads
black - GC content

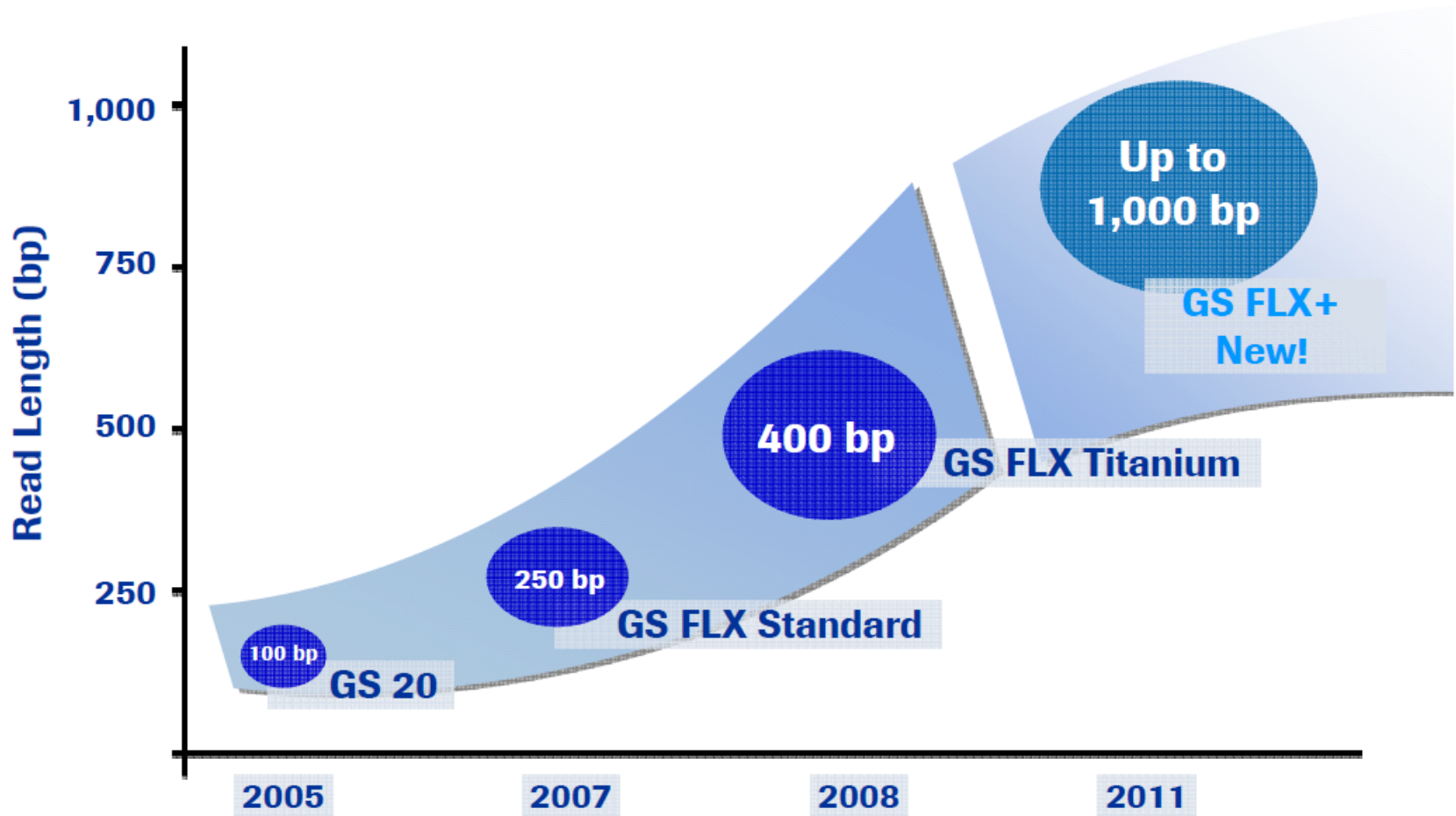


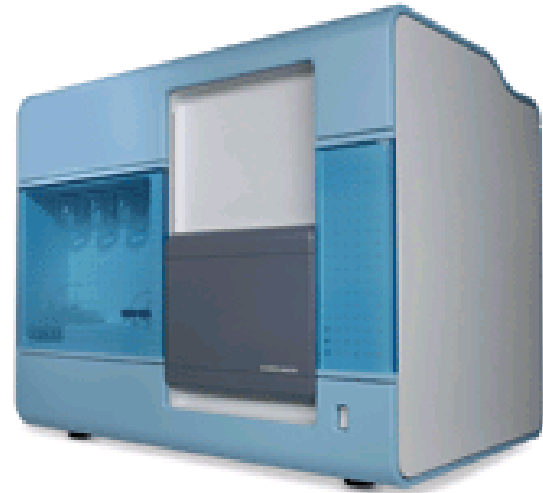
GC poor
low-cloned coverage

GC normal
high-cloned coverage

Roche/454

Read length





Advanced genetic analysis
one billion bases at a time

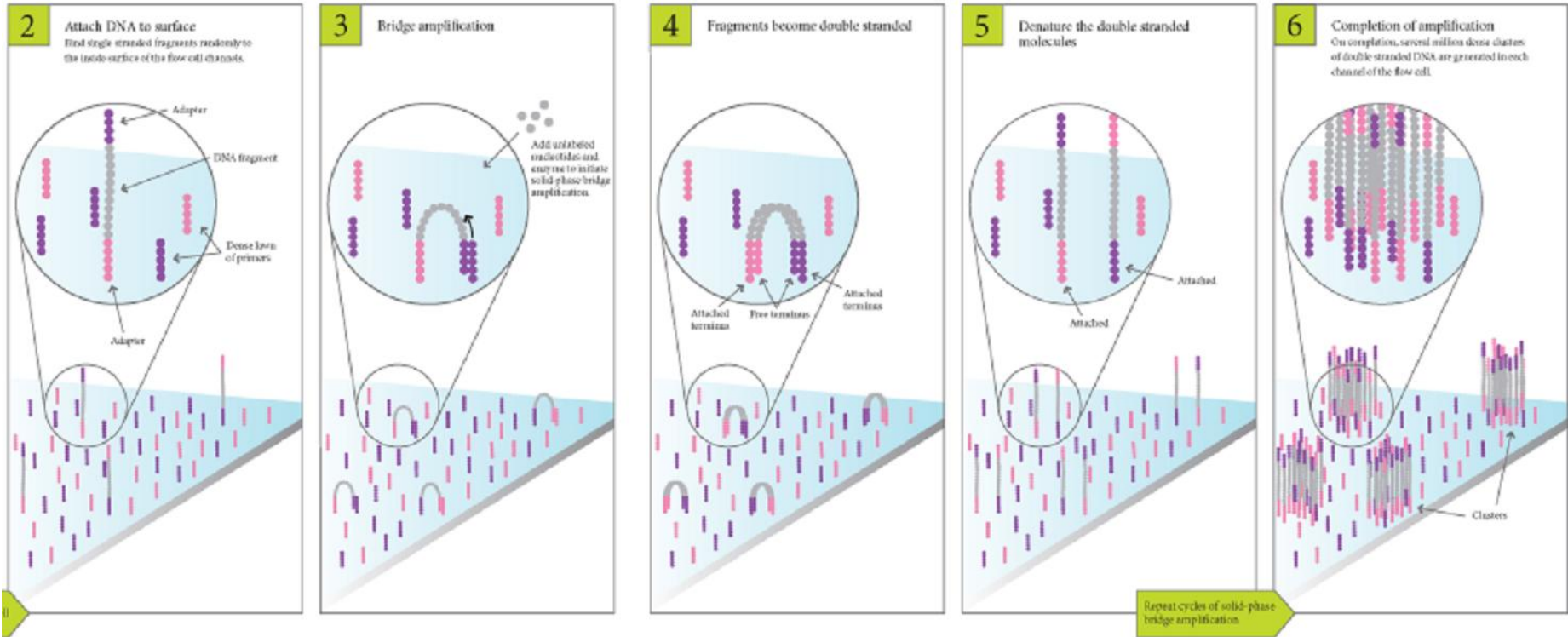
Solexa

Technology

- reversible fluorescent labels
- reversible 3'OH blocking

Illumina/Solexa

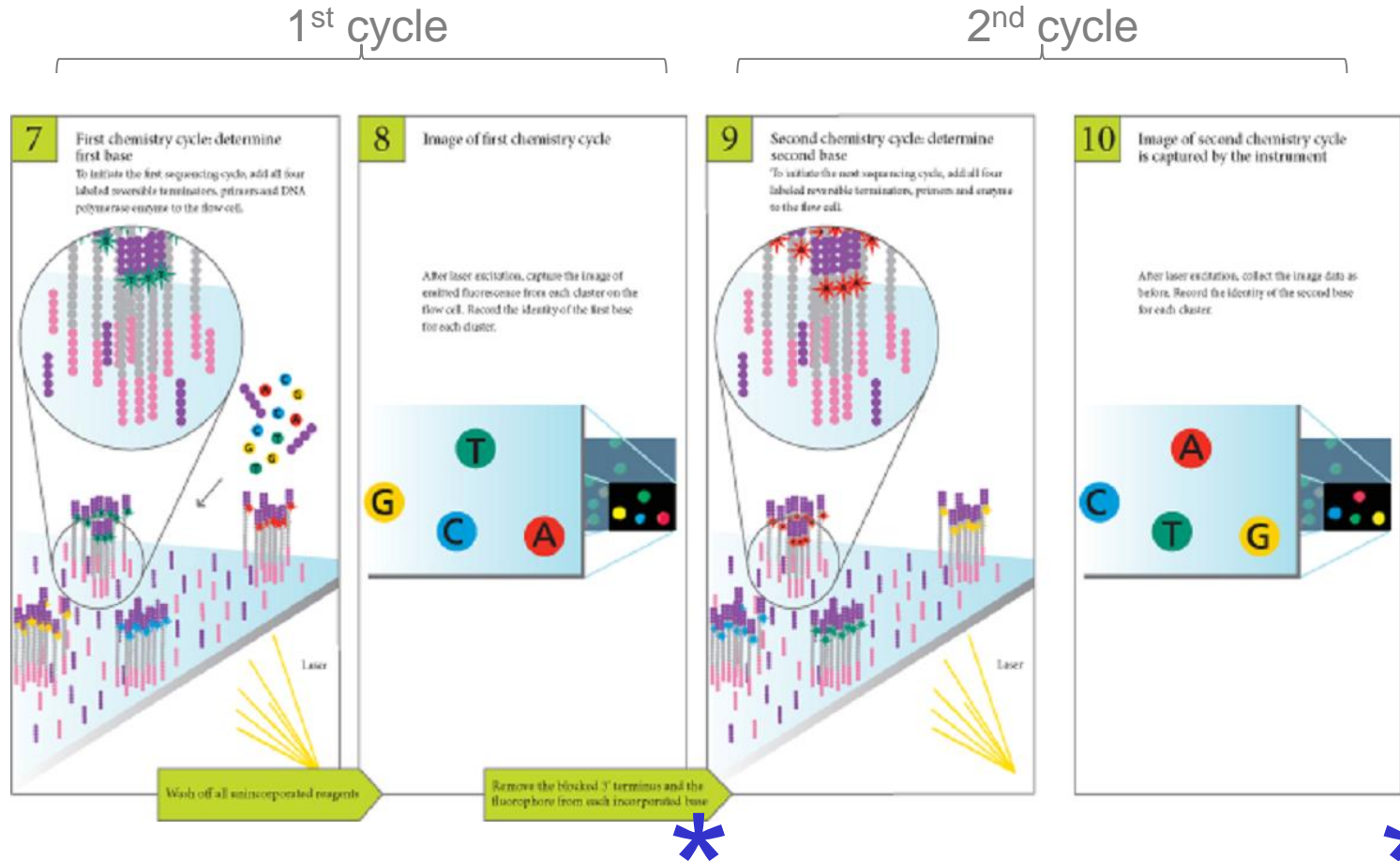
Solid-phase clonal single molecule PCR



“Bridge amplification”

Illumina/Solexa

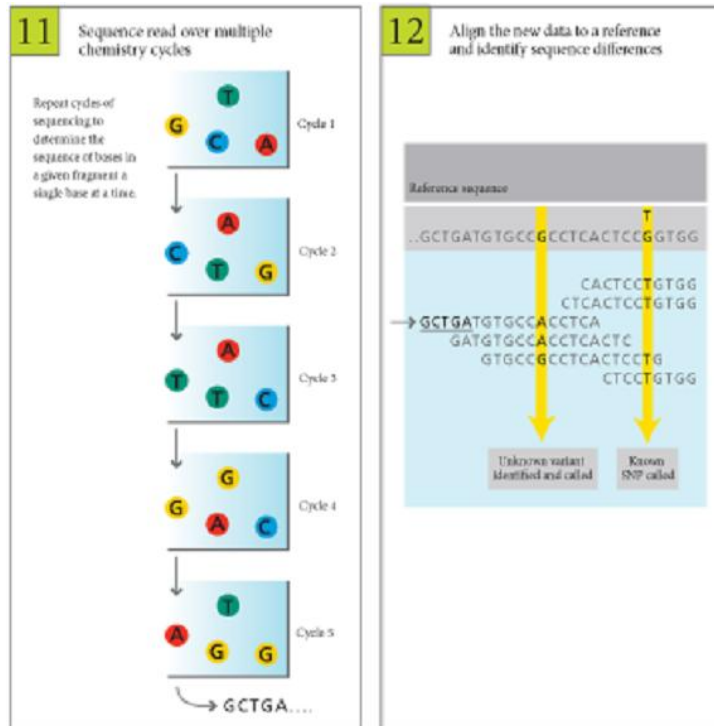
Sequencing by synthesis (SBS)



* removal of fluorescent labels & 3'OH blocking

Illumina/Solexa

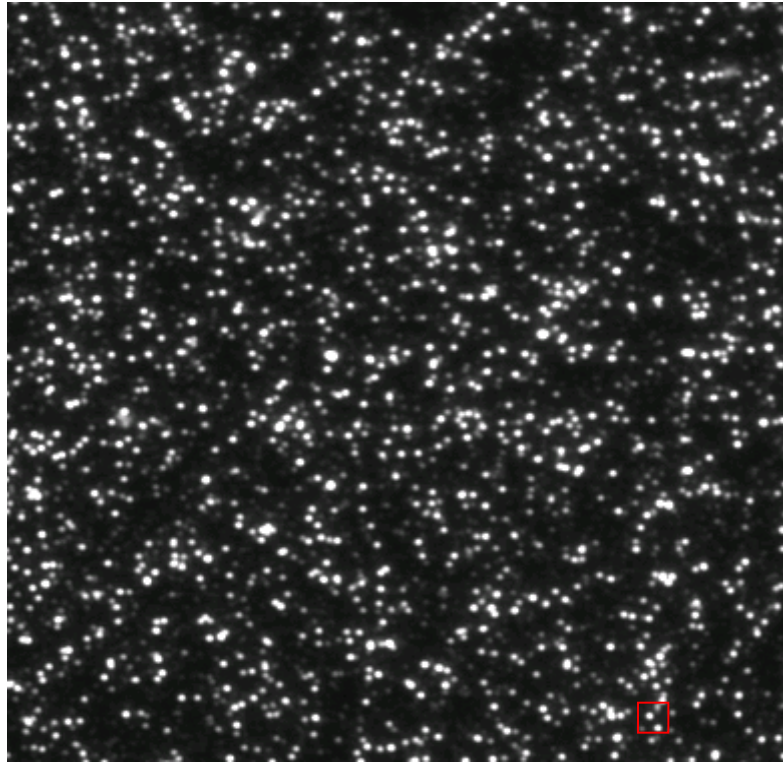
Assembly - Mapping



read length: 50 ... 250 nt

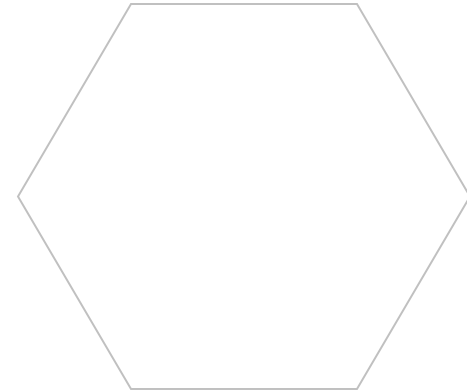
Illumina/Solexa

Solid-phase clonal single molecule PCR



100 μ m

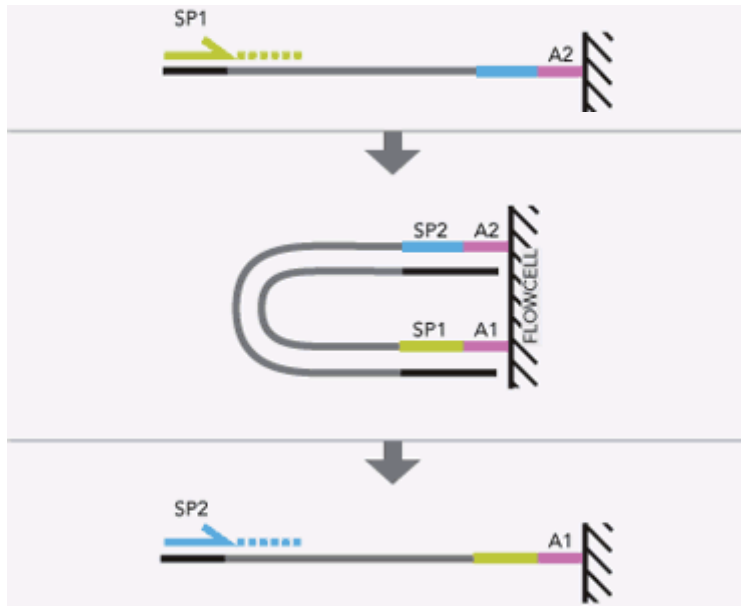
colony of \approx 1000 single-stranded DNA templates



*Single well of
454 Life Sciences
PicoTiterPlateTM
(to same scale)*

Illumina/Solexa

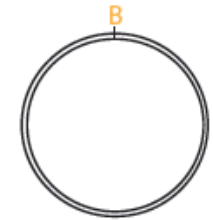
Paired ends & Mate pairs



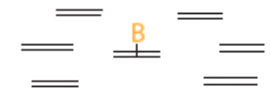
Fragment DNA
Perform End Repair
Biotin Label
Size-Select DNA



Circularize DNA
Digest Linear DNA



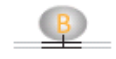
Fragment Circularized DNA



Purify Biotinylated DNA
Perform End Repair



A-tail
Adapter Ligation

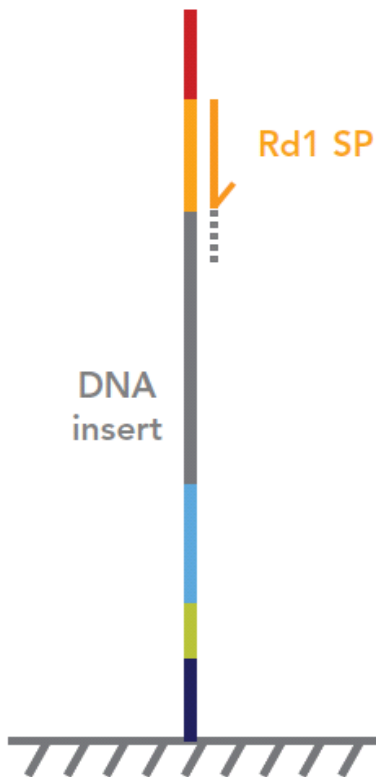


PCR Enrichment
Size-Select Library

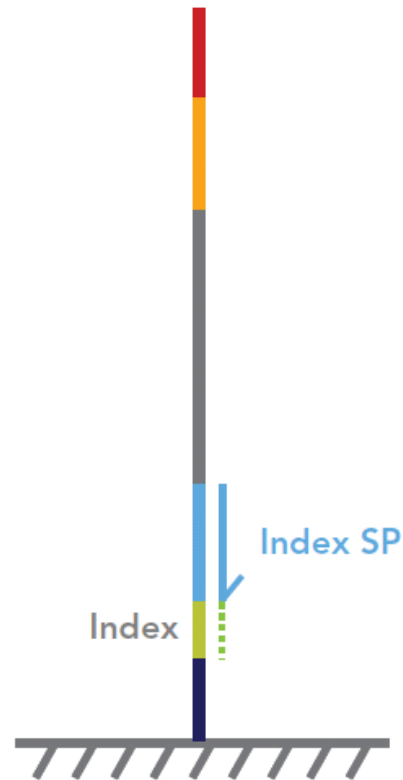


Illumina/Solexa Multiplexing

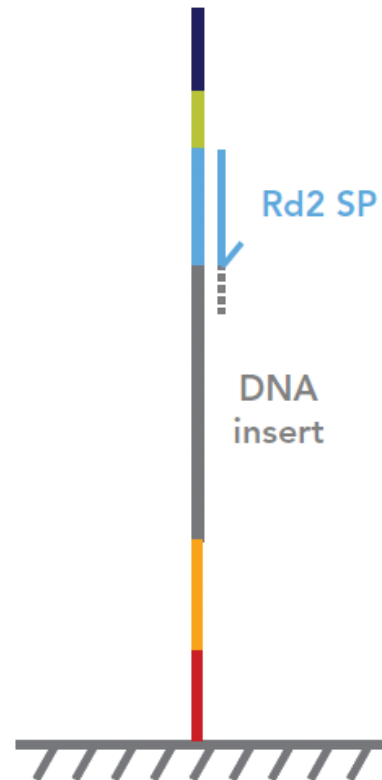
A. READ 1



B. INDEX READ



C. READ 2



Illumina/Solexa

Data output

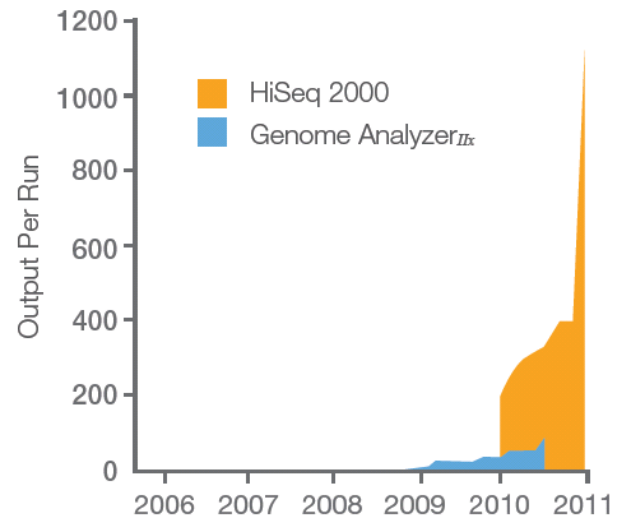
1,000-fold improvement in just four years.

Sequencing Run Parameters

Run form at: 2 x 150 bp	
Output per run	1.13 Tb
Output per day	81 Gb

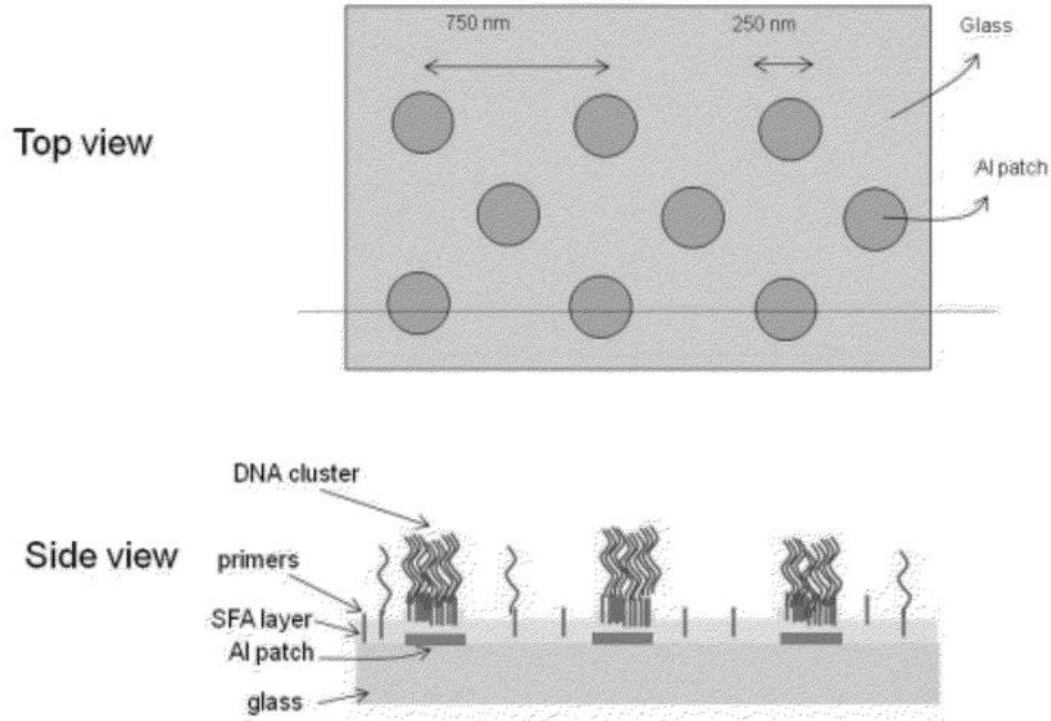
Over the last few years we've scaled from targeting 1 Gb on the Genome Analyzer, to generating in excess of 1 Tb* internally on the HiSeq 2000.

* R&D configuration not supported at launch



Illumina/Solexa

Patterned flow cells



Illumina/Solexa

Patterned flow cells

New platforms November 2014

NextSeq 500

30 Gb/day

1 human genome or 16 exomes



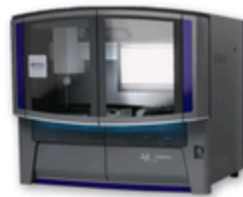
HiSeq3000/4000 HiSeqX Ten

150 Gb/day

5 human genome at **\$1000** each

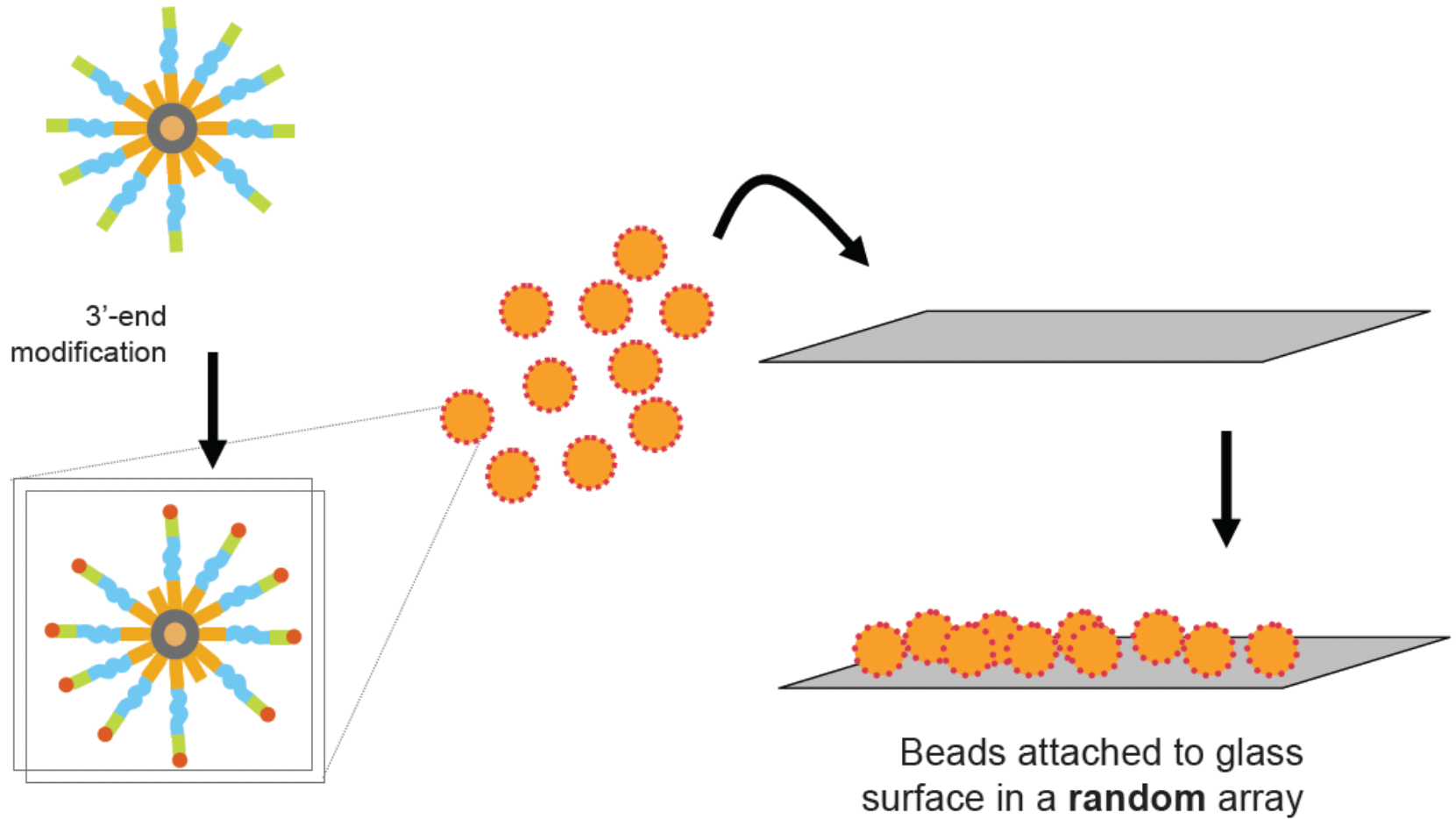


Agencourt



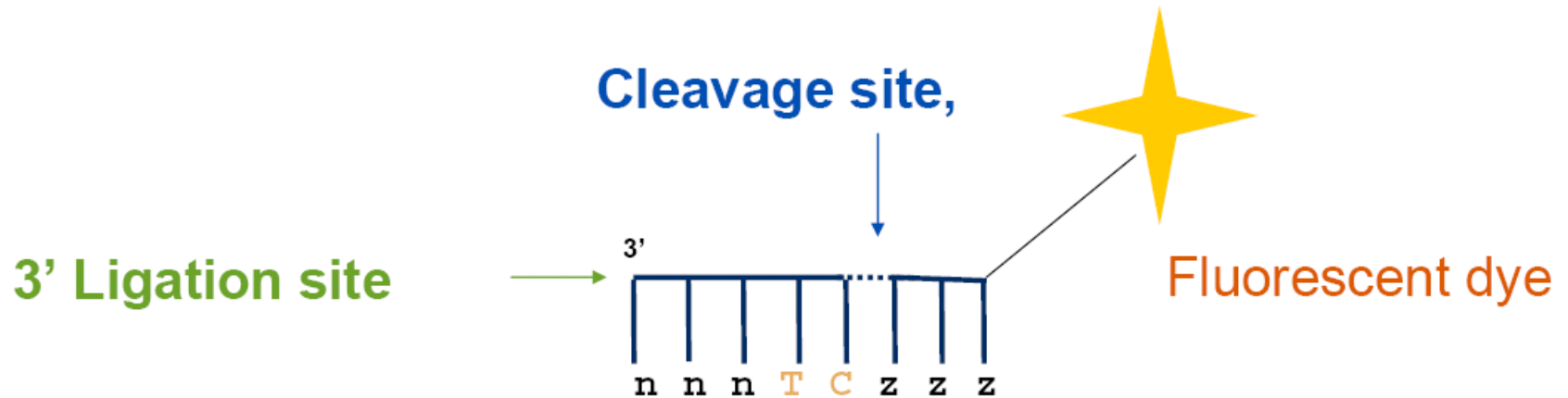
ABI/SOLID

Bead deposition



SOLID

Probes



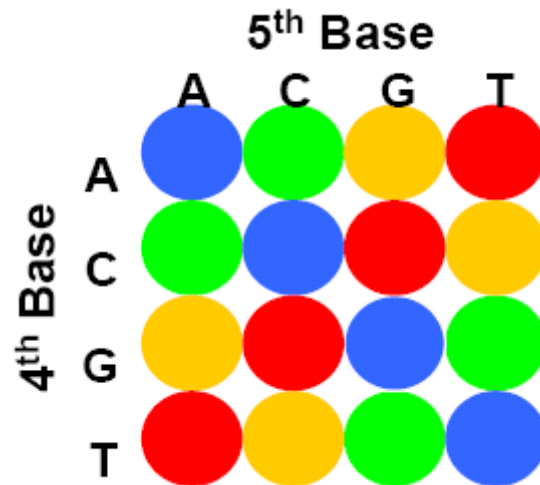
1,024 Octamer Probes (4^5)

4 Dyes, 16 dinucleotides, 256 probes per dye

N= degenerate bases Z= Universal bases

SOLID

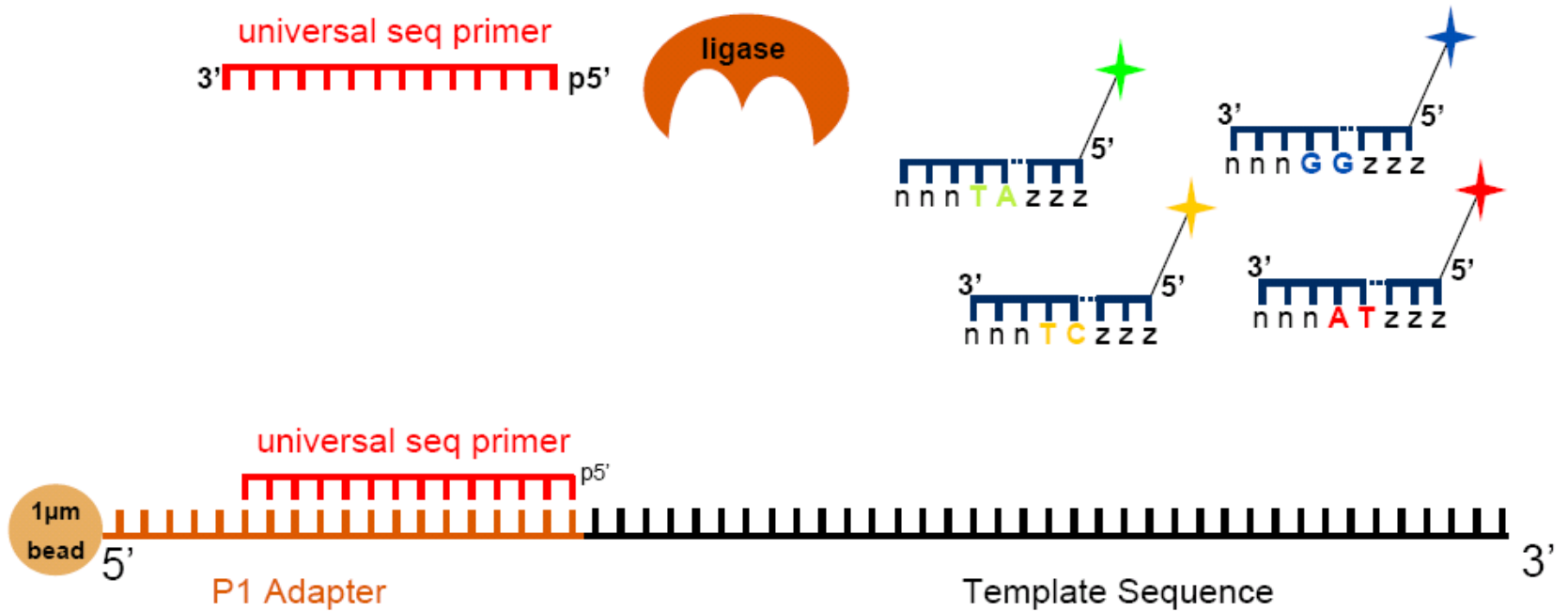
2-base encoding



16 dinucleotides / 4 dyes = 4 dinucleotides/dye

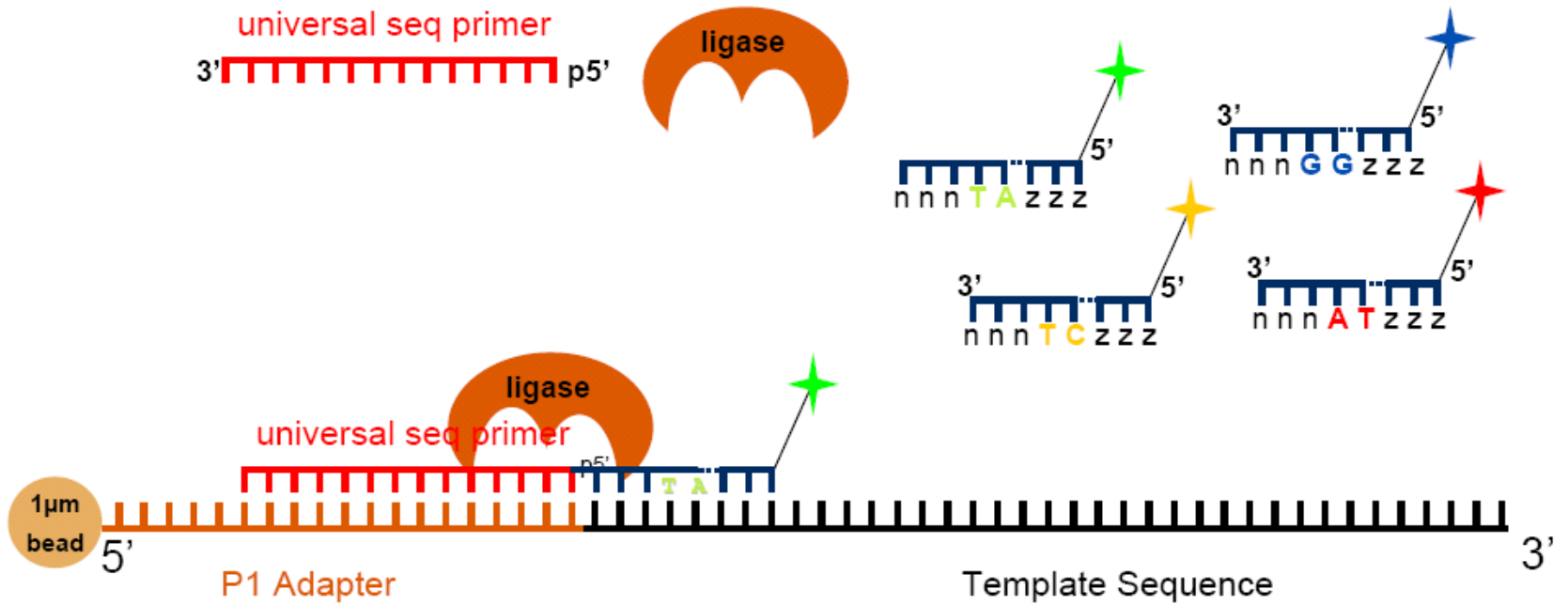
SOLID

4-color ligation



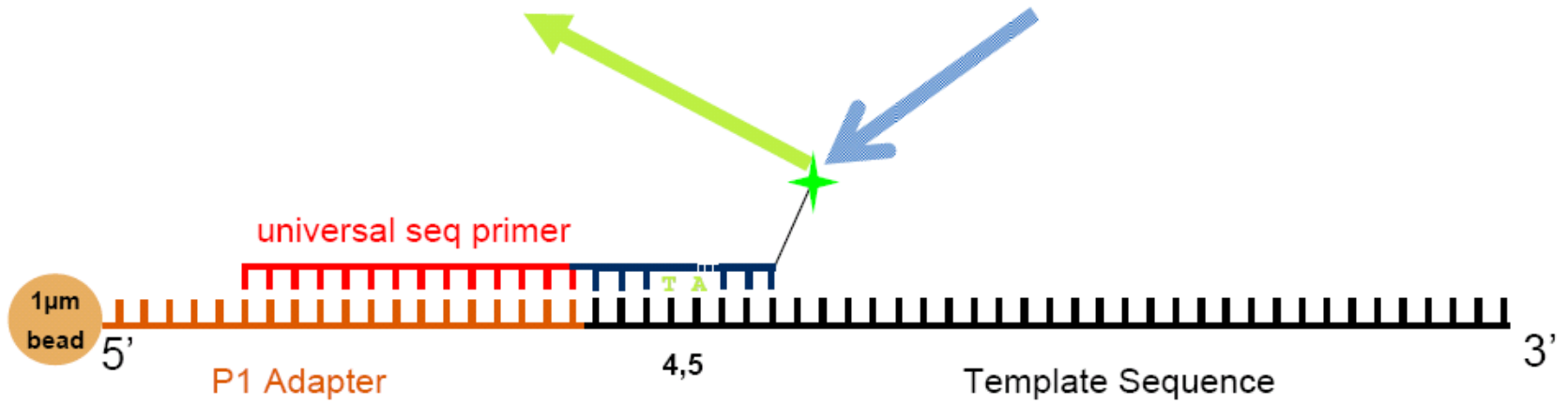
SOLID

4-color ligation



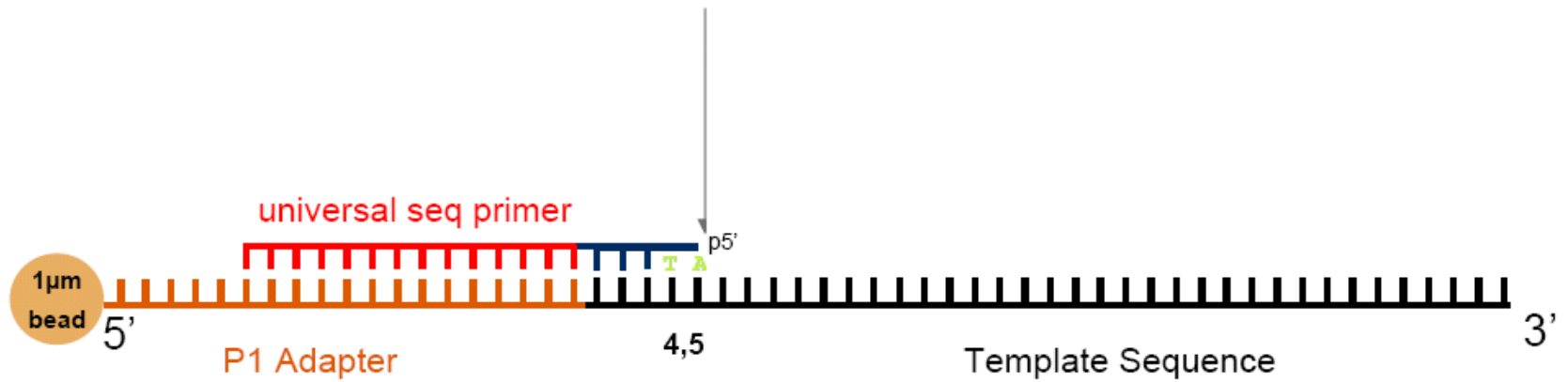
SOLID

Detection



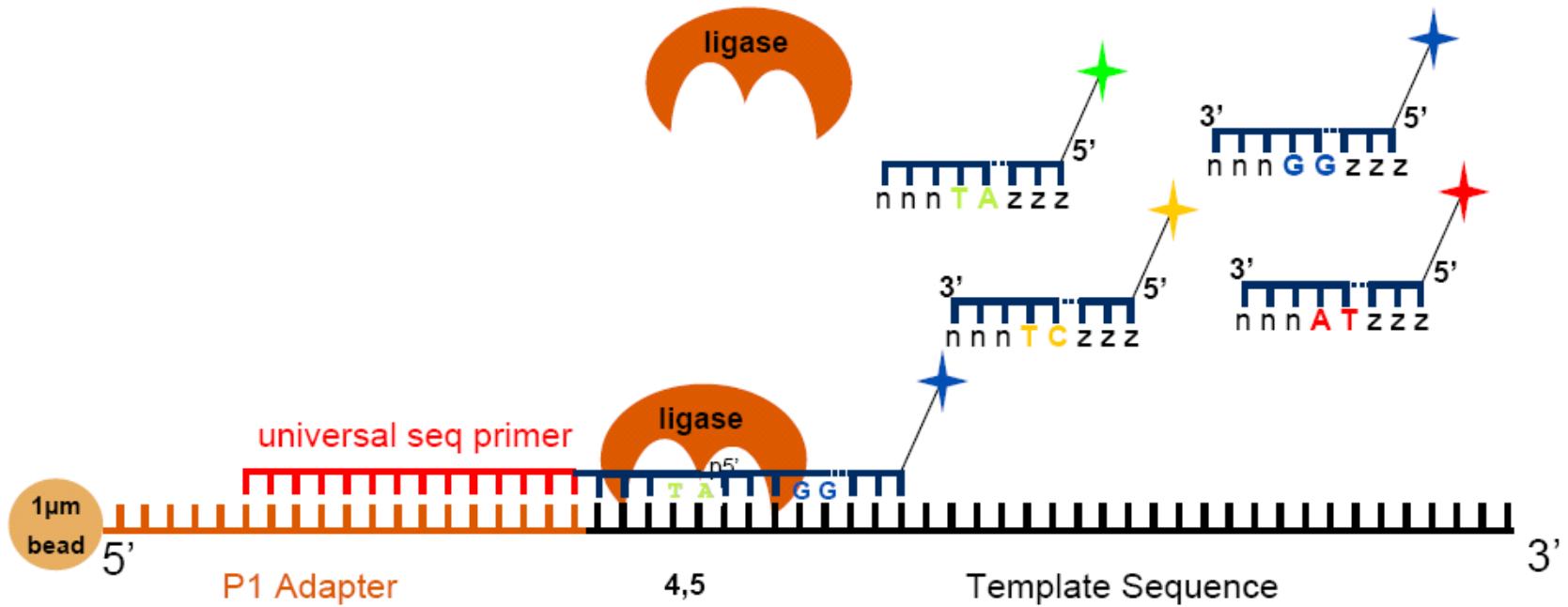
SOLID

Cleavage



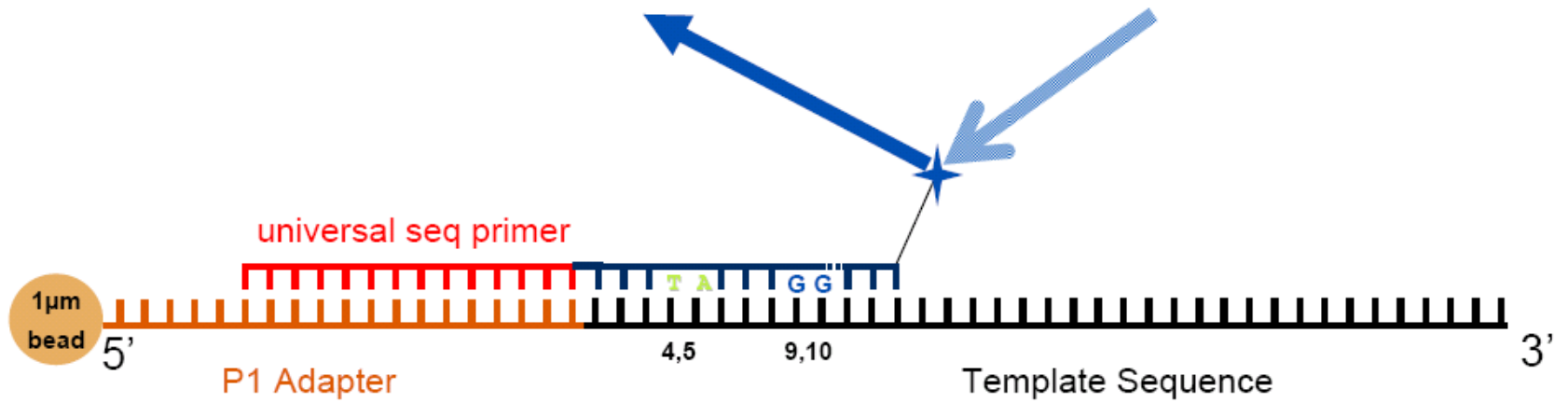
SOLID

2nd cycle ligation



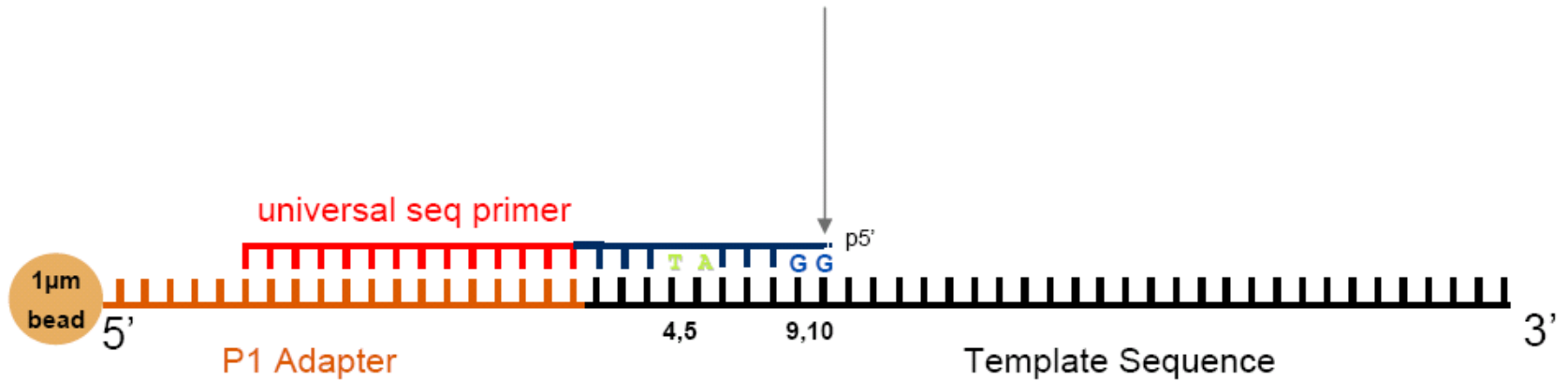
SOLID

2nd cycle detection



SOLID

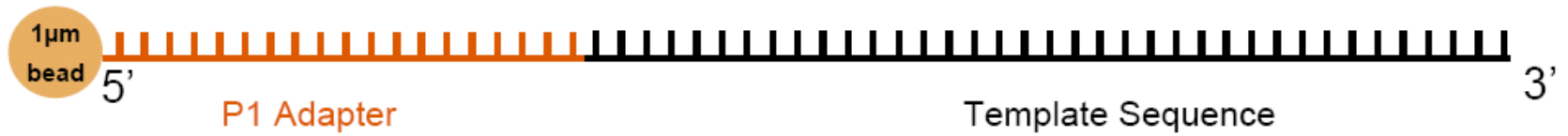
2nd cycle cleavage



+ additional 3 ligation cycles

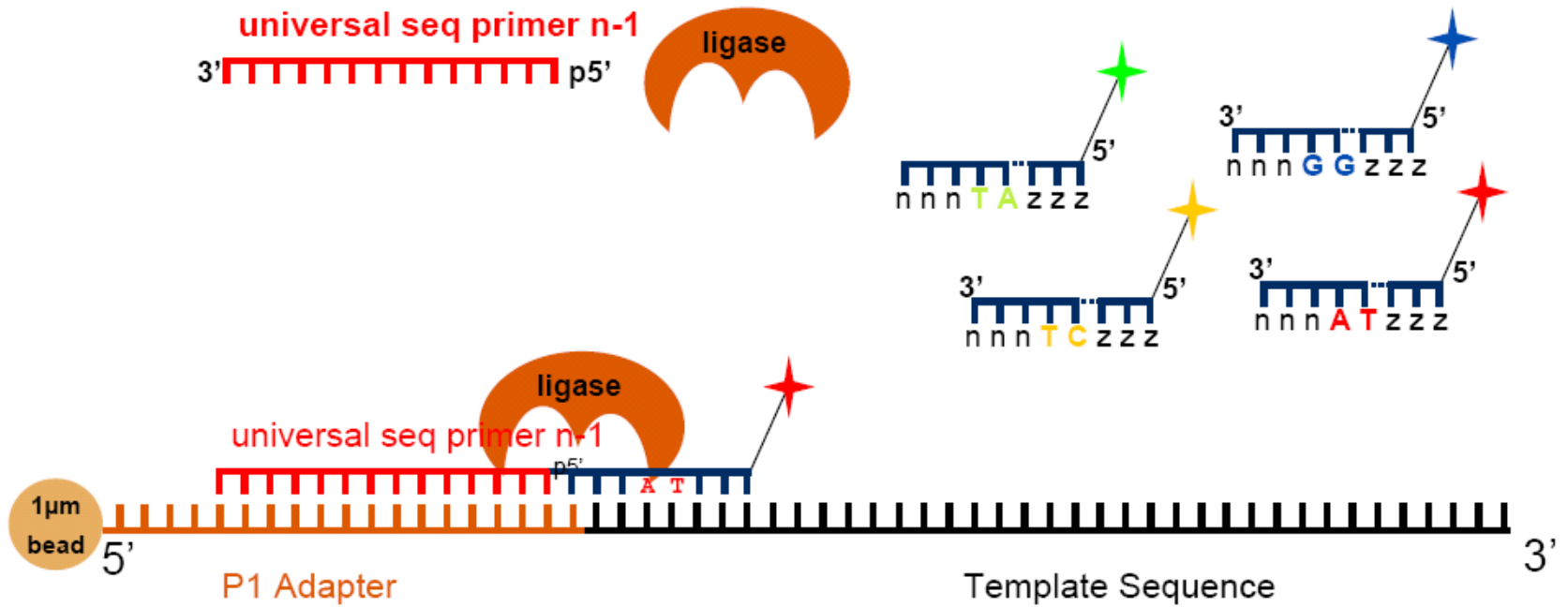
SOLID

Reset after 5th cycle



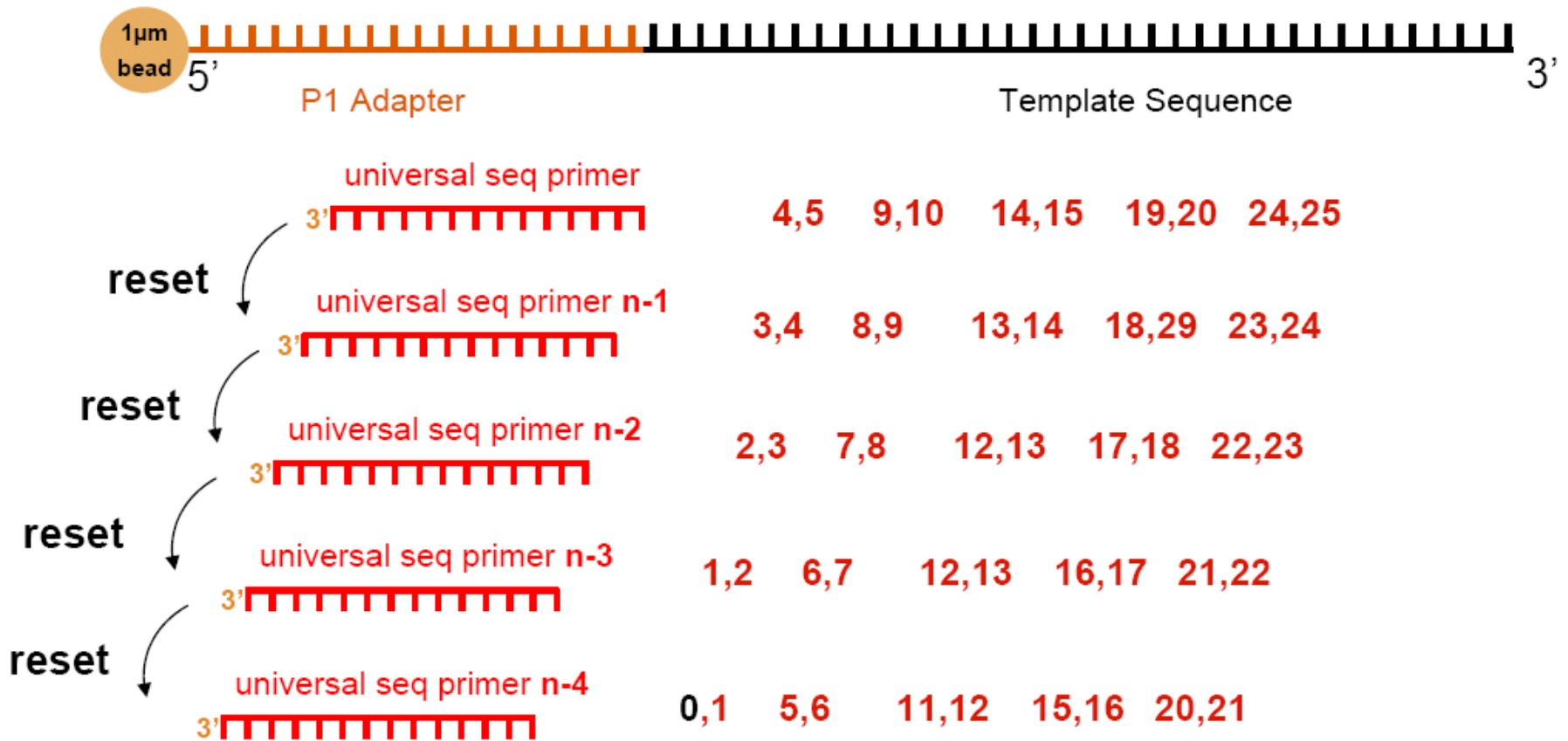
SOLID

6th cycle ligation



SOLID

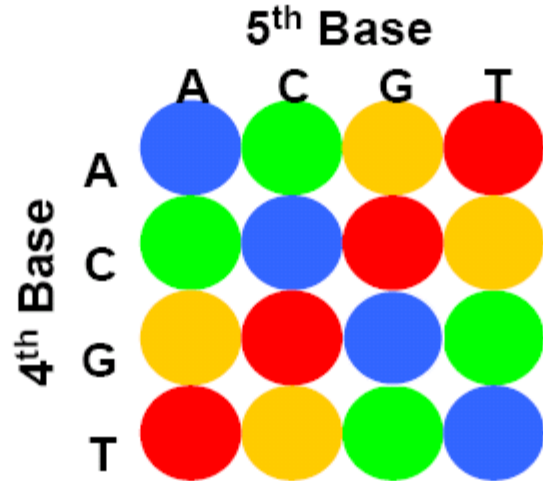
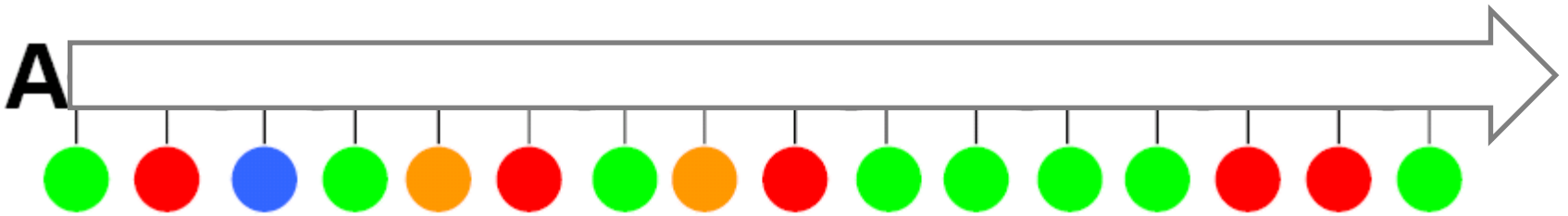
Cycling scheme



SOLID

2-base encoding

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





Complete Genomics Wins
**2009 Emerging Technology
Award—Biotech**

Silicon Valley/San Jose Business Journal



Accurate, Affordable,
Complete Human Genome Studies

Are Within Your Reach.

[View Data](#)



Commercial-scale Genome Center

Complete Genomics' innovative **sequencing technology** will be uniquely offered as a service through its own commercial-scale genome center. Customers will have access to large-scale, complete human genomic data analysis without making a major in-house investment in instruments or high-performance computing resources.



About Complete Genomics

Complete Genomics' high quality, affordable DNA sequencing enables commercial-scale research of the genetic mechanisms underlying drug responses and complex diseases.

Complete Genomics combines innovative technology with a disruptive market approach that will revolutionize DNA sequencing.



In the News

Silicon Valley/San Jose Business Journal Honors Complete Genomics With Its 2009 Biotechnology Emerging Technology Award

Complete Genomics Publishes Three Genomes in *Science*

ISB and Complete Genomics to Conduct 100-Genome Study

[more>>](#)



[Careers](#)



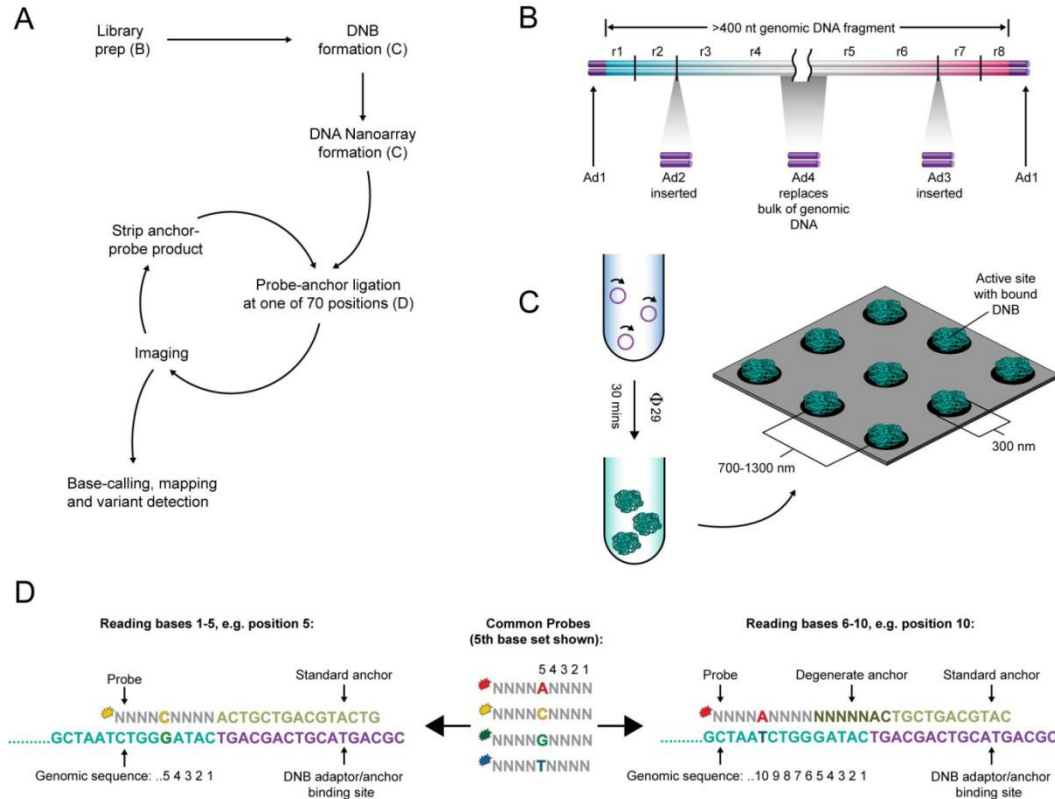
[News and Events](#)



[Our Mission](#)

Complete Genomics

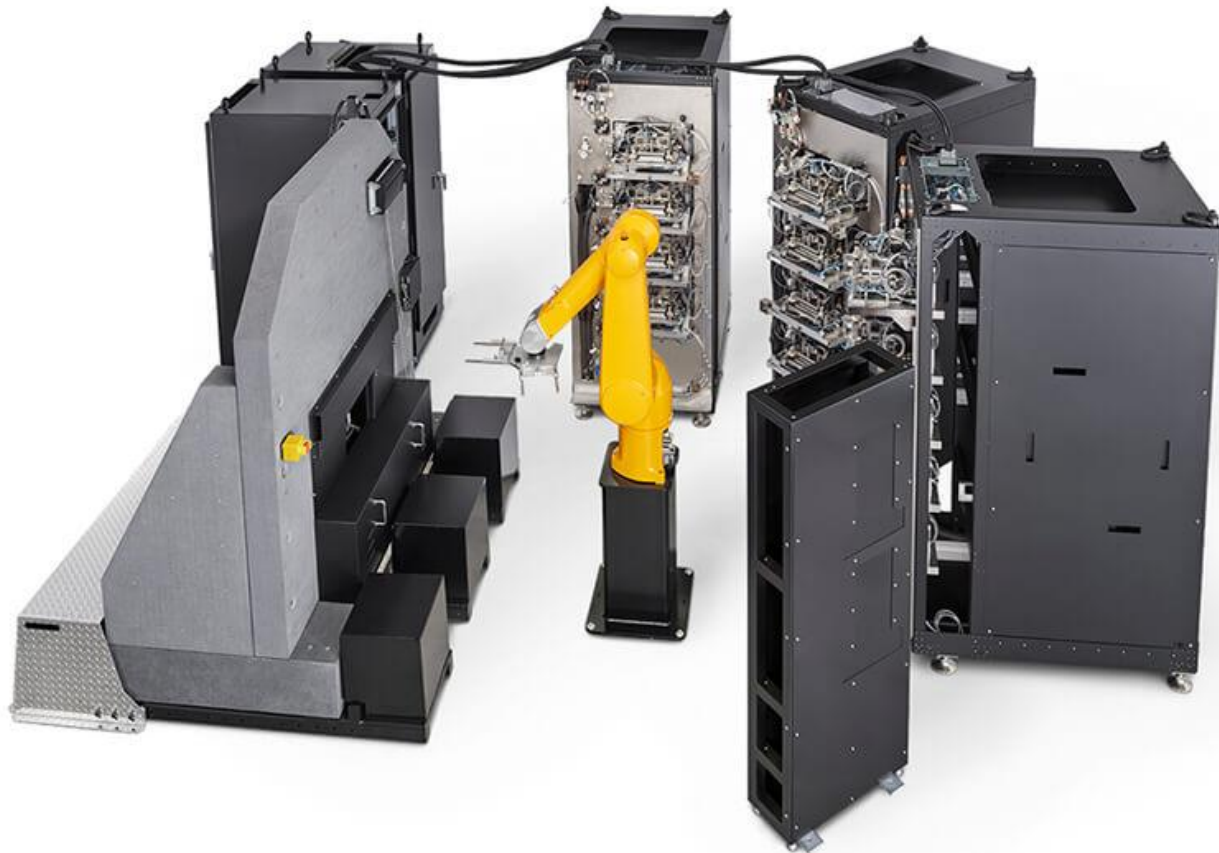
In-house technology



- self-assembling DNA nanoballs (DNB)
- patterned nanoarrays
- combinatorial probe anchor ligation (cPAL)

Complete Genomics - BGI

Revolocity



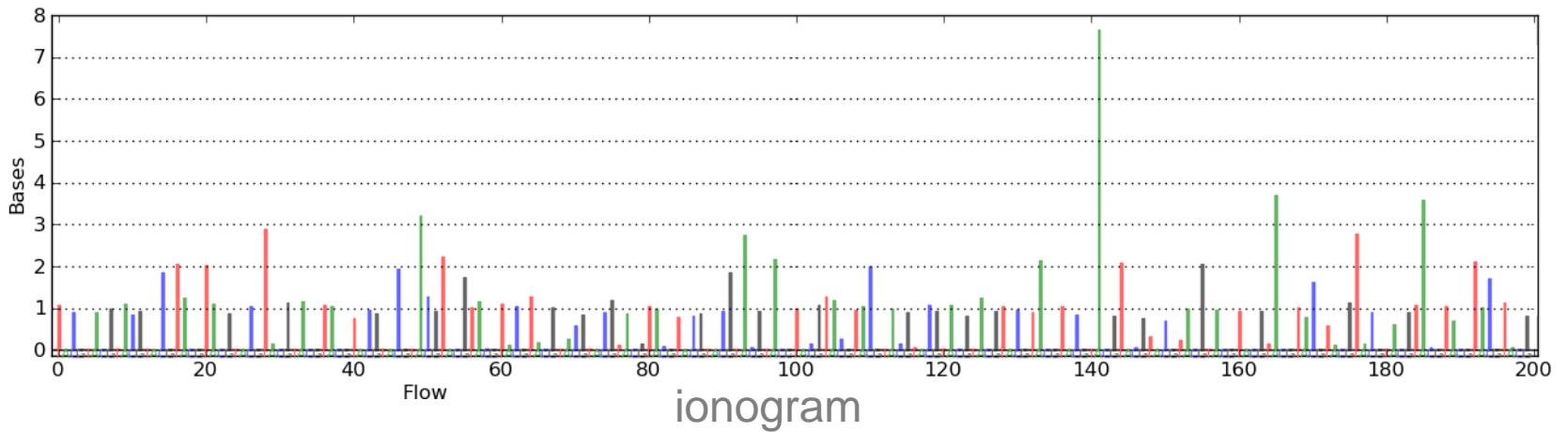
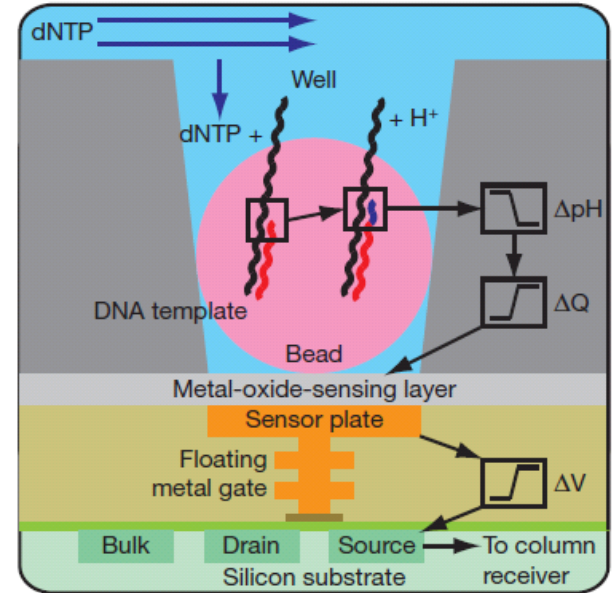
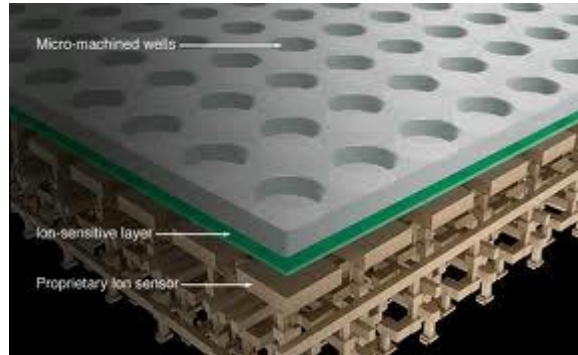
10,000 ... 30.000 human genomes / year



PGM - Personal Genome Machine

Ion Torrent

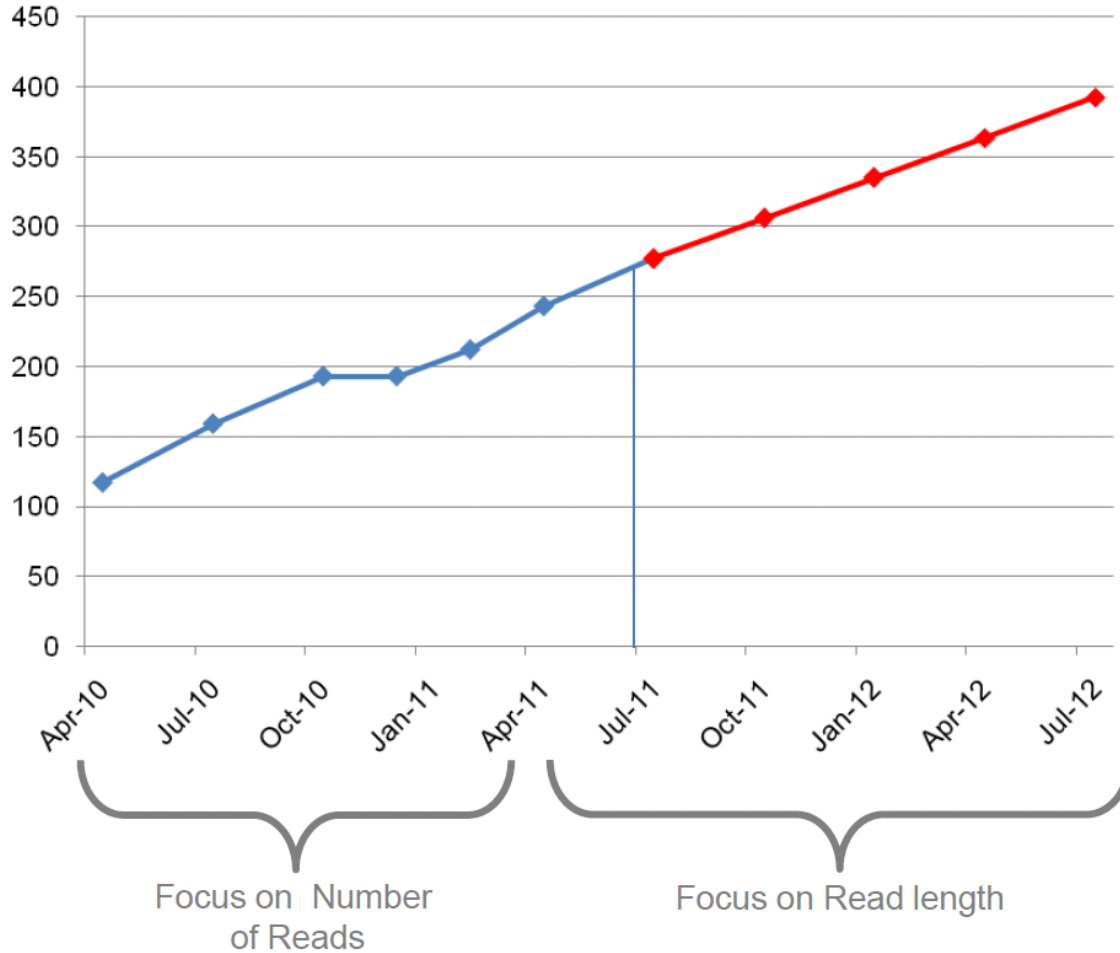
The chip is the machine



Ion Torrent

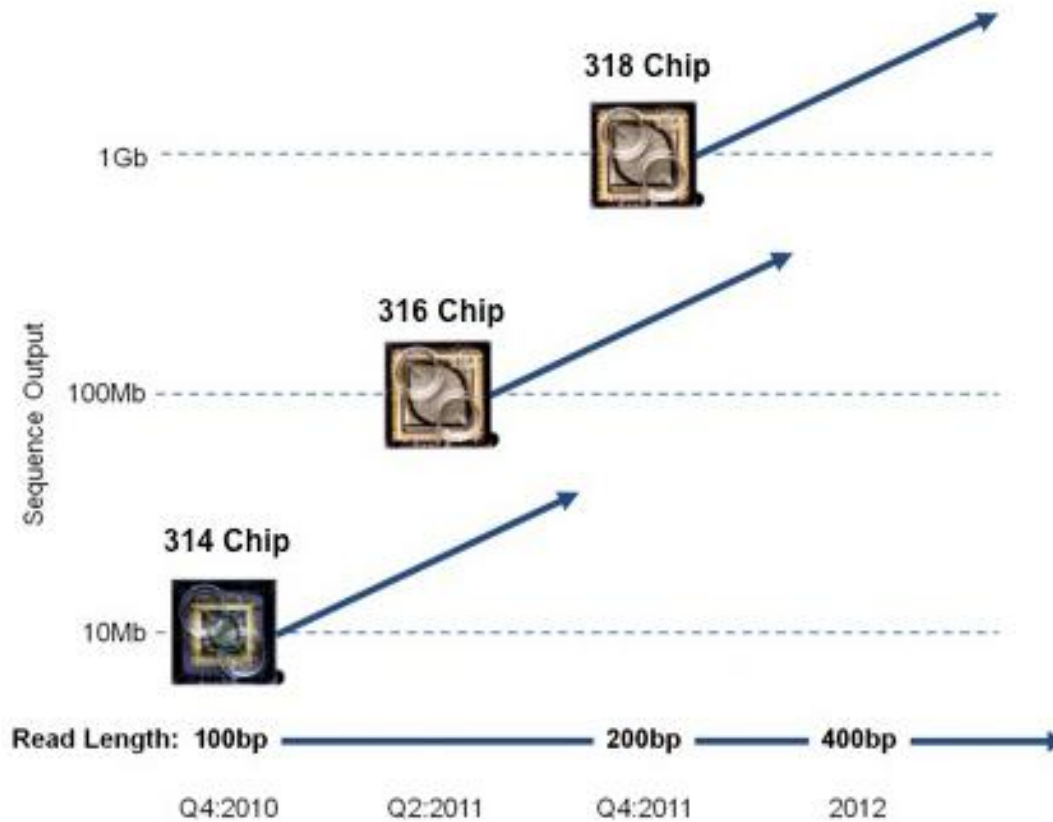
Read length

Longest Perfect Reads



ABI/Ion Torrent

Data output



Chip Type	314	316	318
Wells (in millions)	1.2	6.2	11.1
Bases (in Mb)	>10	>100	>1000

Ion Torrent Data output

SMALL
GENOMES

SETS OF
GENES

GENE EXPRESSION
ChIP-SEQ

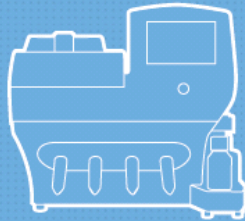
WHOLE
TRANSCRIPTOMES

HUMAN
EXOMES

HUMAN
GENOMES

Ion PGM™ Sequencer

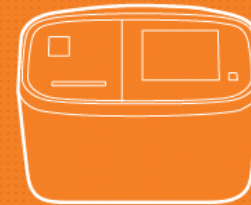
314 316 318



30 Mb – 1 Gb

Ion Proton™ Sequencer

P1 P2



10 Gb

314 316 318

314 316 318

318 P1

318 P1

P1 P2

P2

ABI/Ion Torrent

'Post-light' sequencing of Gordon Moore

An integrated semiconductor device enabling non-optical genome sequencing

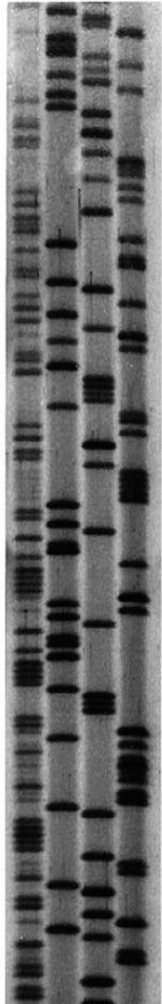
Jonathan M. Rothberg¹, Wolfgang Hinz¹, Todd M. Rearick¹, Jonathan Schultz¹, William Mileski¹, Mel Davey¹, John H. Leamon¹, Kim Johnson¹, Mark J. Milgrew¹, Matthew Edwards¹, Jeremy Hoon¹, Jan F. Simons¹, David Marran¹, Jason W. Myers¹, John F. Davidson¹, Annika Branting¹, John R. Nobile¹, Bernard P. Puc¹, David Light¹, Travis A. Clark¹, Martin Huber¹, Jeffrey T. Branciforte¹, Isaac B. Stoner¹, Simon E. Cawley¹, Michael Lyons¹, Yutao Fu¹, Nils Homer¹, Marina Sedova¹, Xin Miao¹, Brian Reed¹, Jeffrey Sabina¹, Erika Feierstein¹, Michelle Schorn¹, Mohammad Alanjary¹, Eileen Dimalanta¹, Devin Dressman¹, Rachel Kasinskas¹, Tanya Sokolsky¹, Jacqueline A. Fidanza¹, Eugeni Namsaraev¹, Kevin J. McKernan¹, Alan Williams¹, G. Thomas Roth¹ & James Bustillo¹

The seminal importance of DNA sequencing to the life sciences, biotechnology and medicine has driven the search for more scalable and lower-cost solutions. Here we describe a DNA sequencing technology in which scalable, low-cost semiconductor manufacturing techniques are used to make an integrated circuit able to directly perform non-optical DNA sequencing of genomes. Sequence data are obtained by directly sensing the ions produced by template-directed DNA polymerase synthesis using all-natural nucleotides on this massively parallel semiconductor-sensing device or ion chip. The ion chip contains ion-sensitive, field-effect transistor-based sensors in perfect register with 1.2 million wells, which provide confinement and allow parallel, simultaneous detection of independent sequencing reactions. Use of the most widely used technology for constructing integrated circuits, the complementary metal-oxide semiconductor (CMOS) process, allows for low-cost, large-scale production and scaling of the device to higher densities and larger array sizes. We show the performance of the system by sequencing three bacterial genomes, its robustness and scalability by producing ion chips with up to 10 times as many sensors and sequencing a human genome.

2nd Generation sequencing

Applications

- Re-sequencing
- RNA-seq
- miRNA-seq
- ChIP-seq
- Ribosome-footprinting
- ...



A C G T

3rd Generation

Pacific Biosciences
Oxford Nanopore



Pacific Biosciences

Fluorescent label

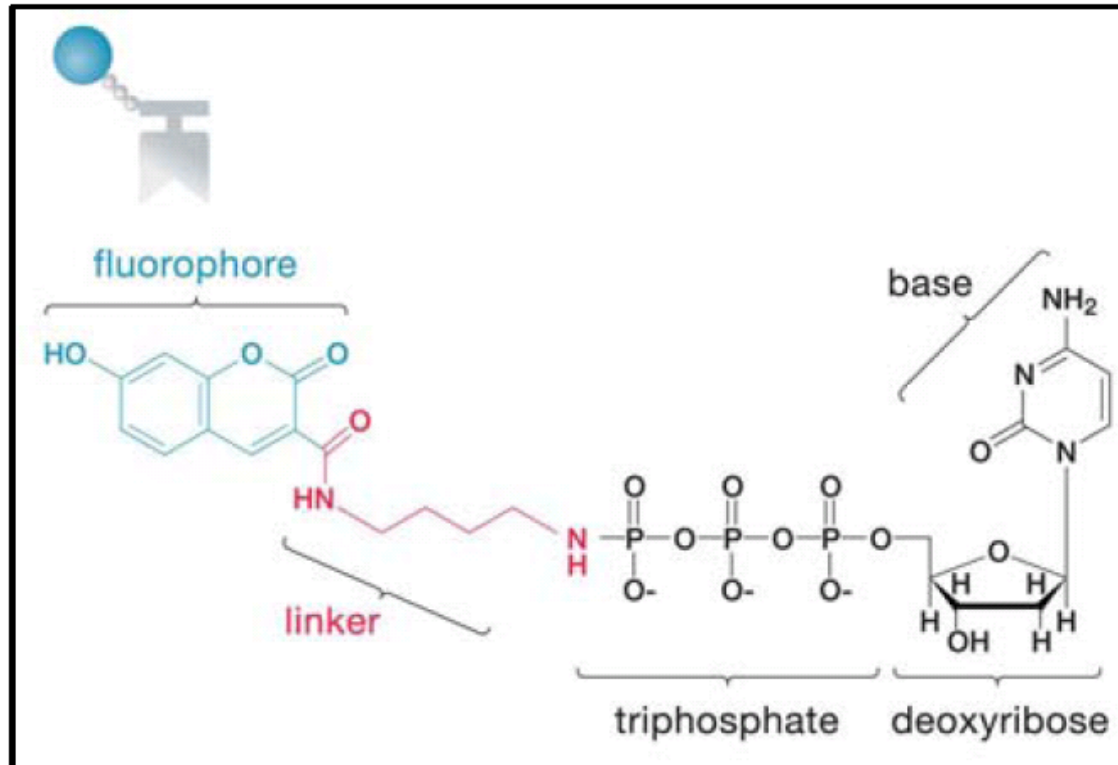
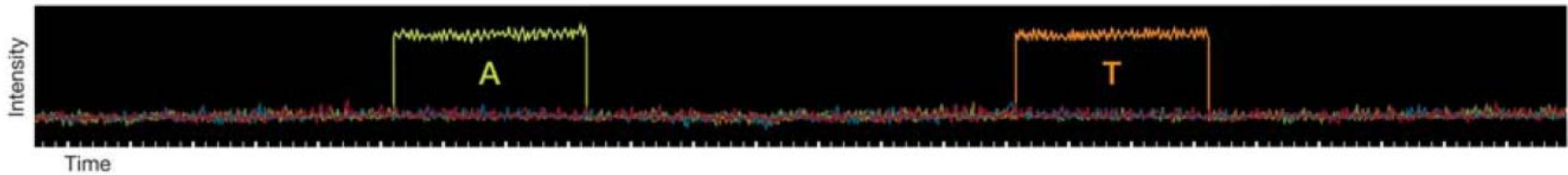


Figure 9. Phospholinked nucleotides

Pacific Biosciences

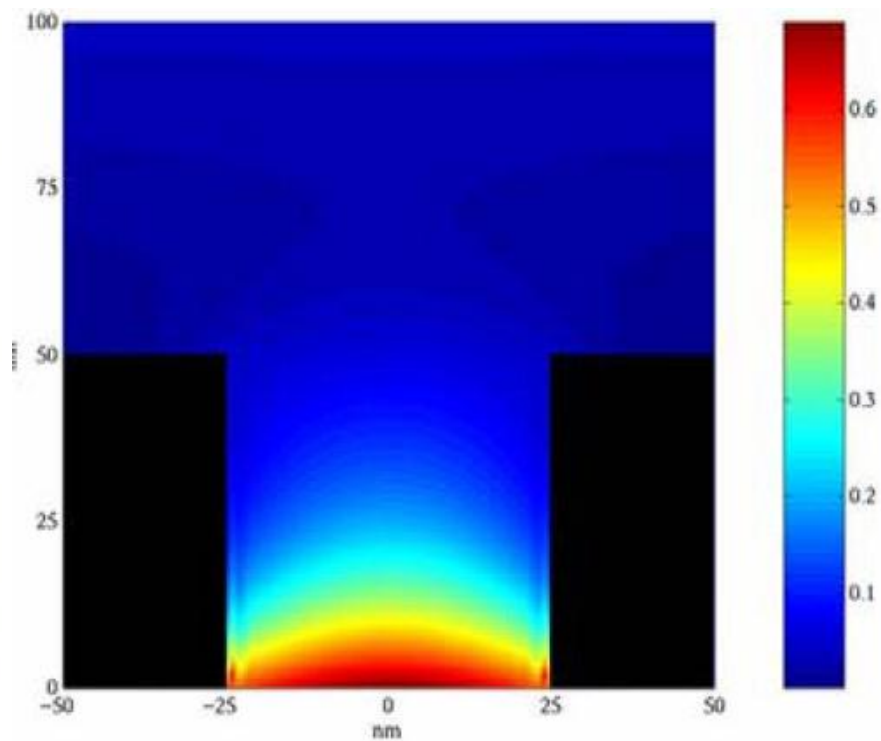
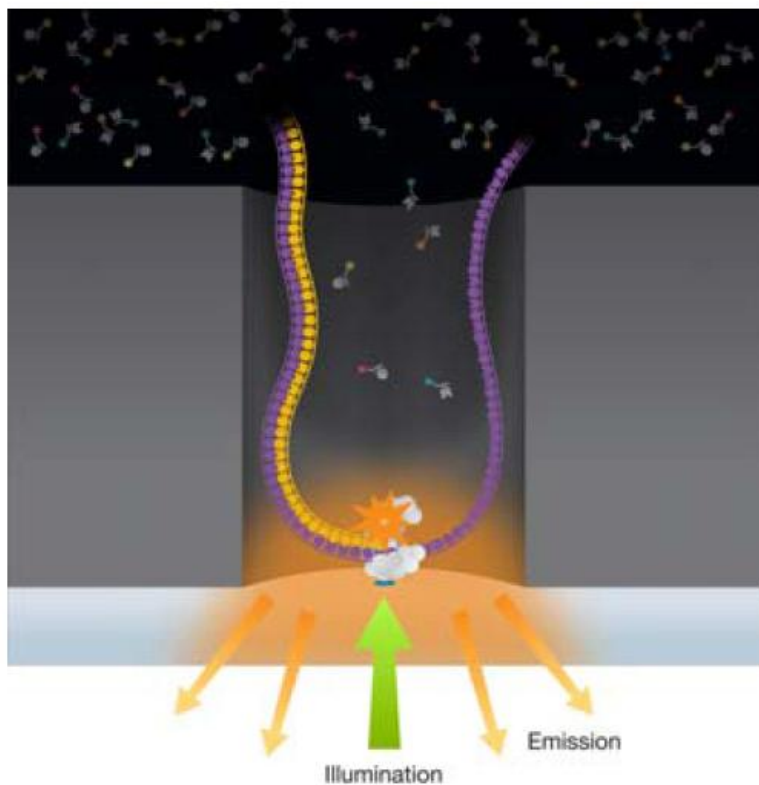
Immobilised polymerase



long sequences

Pacific Biosciences

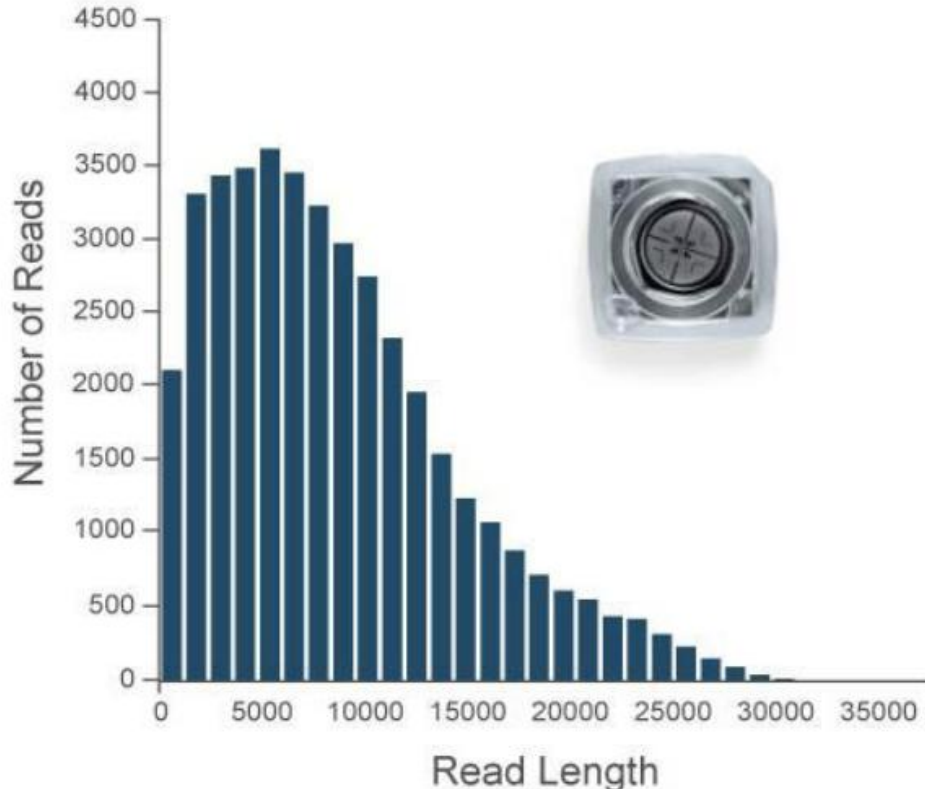
Zero-mode waveguide chip



Pacific Biosciences

Read length P5-C3

Read Length Distribution



Typical Results

Read Length:

Average: ~ 8.5 kb

Maximum: > 30 kb

Top 5% of reads: > 18 kb

Half of data in reads: > 10 kb

Data per SMRT® Cell: ~ 375 Mb

Based on data from a 20 kb size-selected E. coli library using a 180-minute movie.

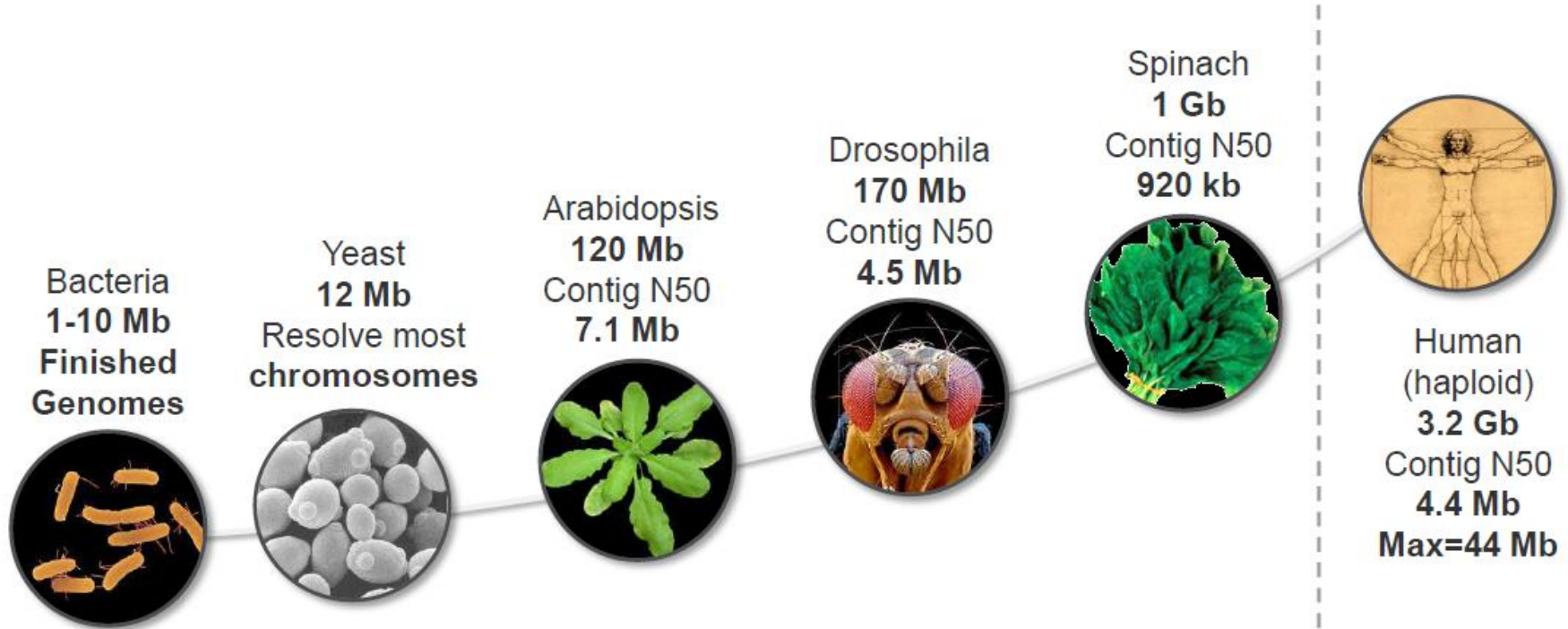
Each SMRT Cell yields ~ 50,000 reads.

Pacific Biosciences

Genome *de novo* assemblies

2013

2014



Addressing the missing 15% of genomes important for function and structural integrity

Genes • Regulatory Elements • Transposable Elements • Retrotransposable Elements • Centromere sequences • Telomeres • Pericentromeric heterochromatin • Structural Variation

Pacific Biosciences

Read coverage



PacBio data at <http://datasets.pacb.com/2014/Human54x/fast.html>

HiSeq data at <http://www.ncbi.nlm.nih.gov/sra/SRR642636>

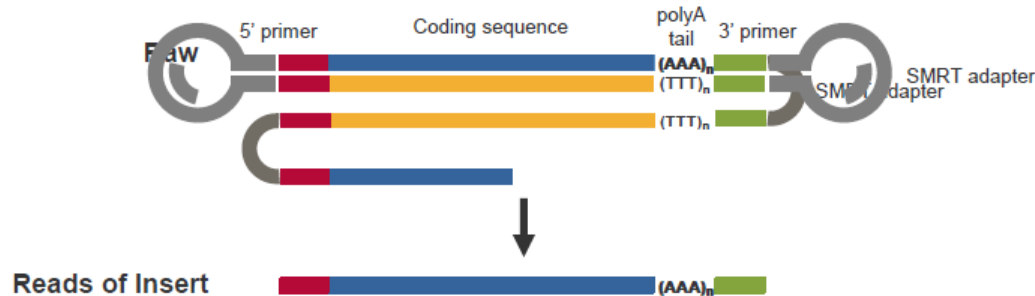
Pacific Biosciences

Full length transcripts

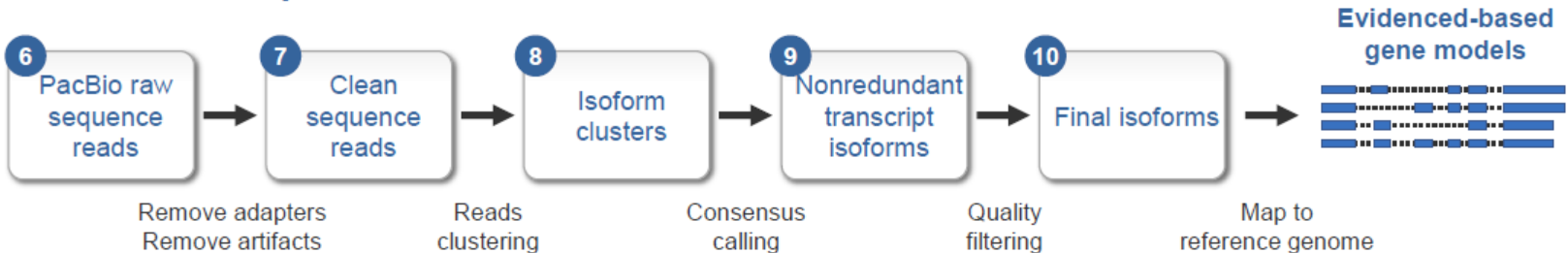
Experimental Pipeline



[SampleNet: Iso-Seq Method with Clontech cDNA Synthesis Kit](#)

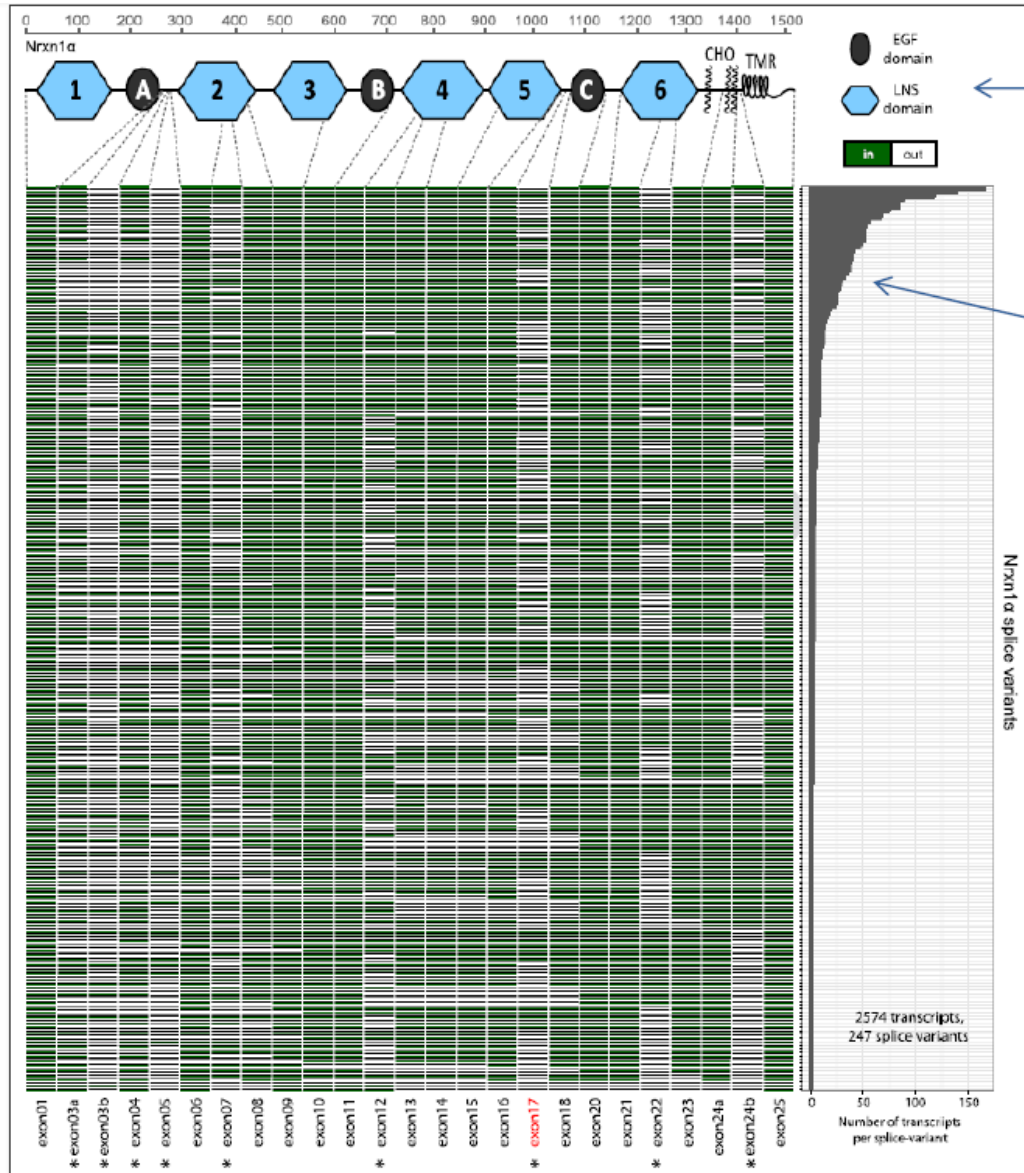


Informatics Pipeline



Pacific Biosciences

Alternative neurexin 1 α transcripts



← Nrnx1 α domain structure

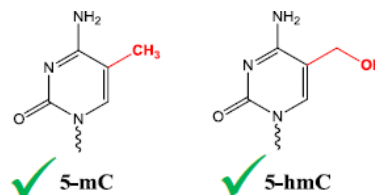
← Splice isoform abundance

- 2,574 full-length Nrnx1 α mRNAs sequence reads
- Rows: 247 unique alternatively spliced isoforms
- Columns: Exons
green – present
white – absent

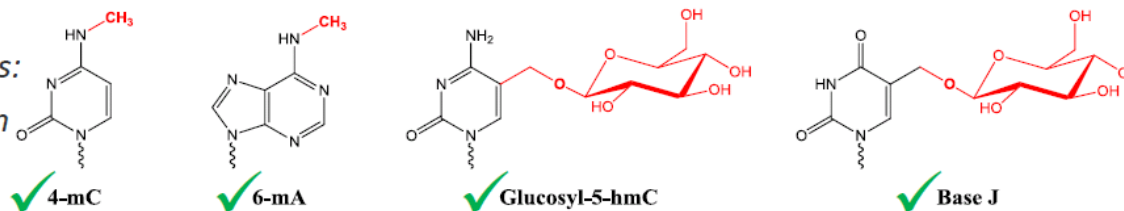
Pacific Biosciences

Epigenetics

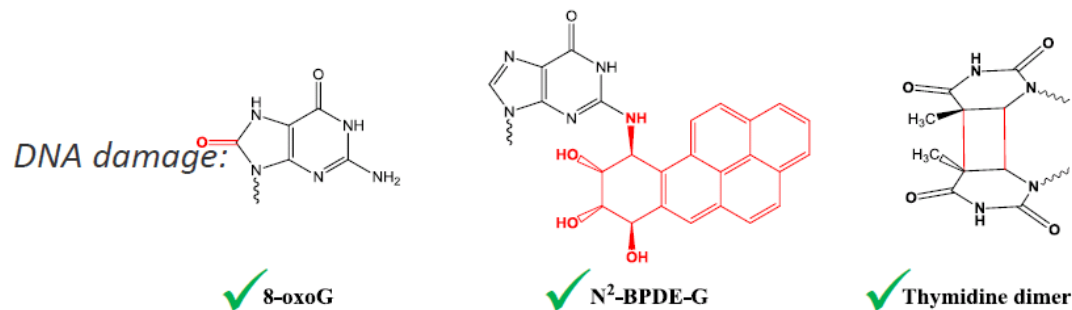
Epigenetic markers:



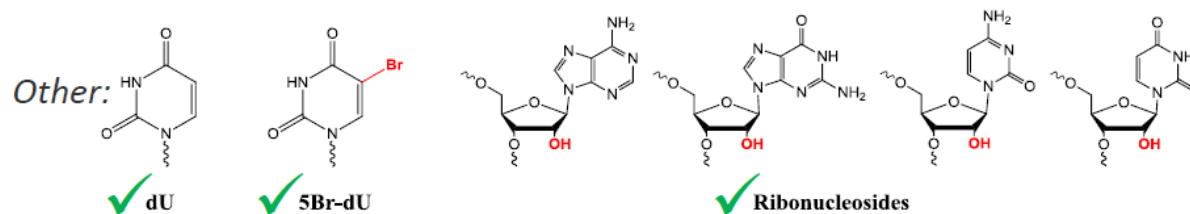
Identity markers:
(host-pathogen interactions)



DNA damage:



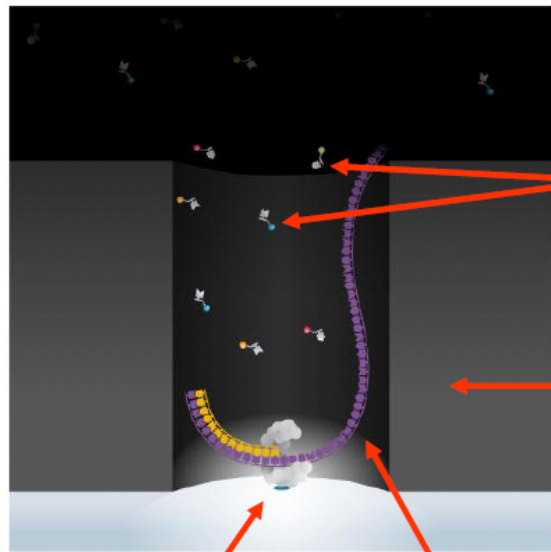
Other:



Pacific Biosciences

Direct RNA sequencing

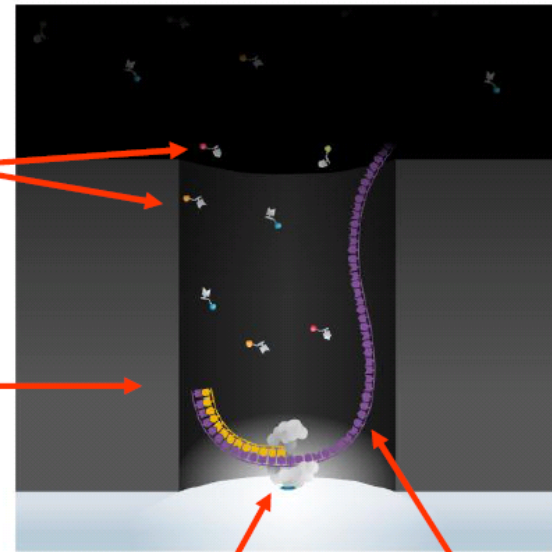
SMRT DNA Sequencing



DNA polymerase

DNA template

SMRT RNA Sequencing



Reverse transcriptase

RNA template

Same Labeled Nucleotides

Same ZMW

ARTICLES

Real-time tRNA transit on single translating ribosomes at codon resolution

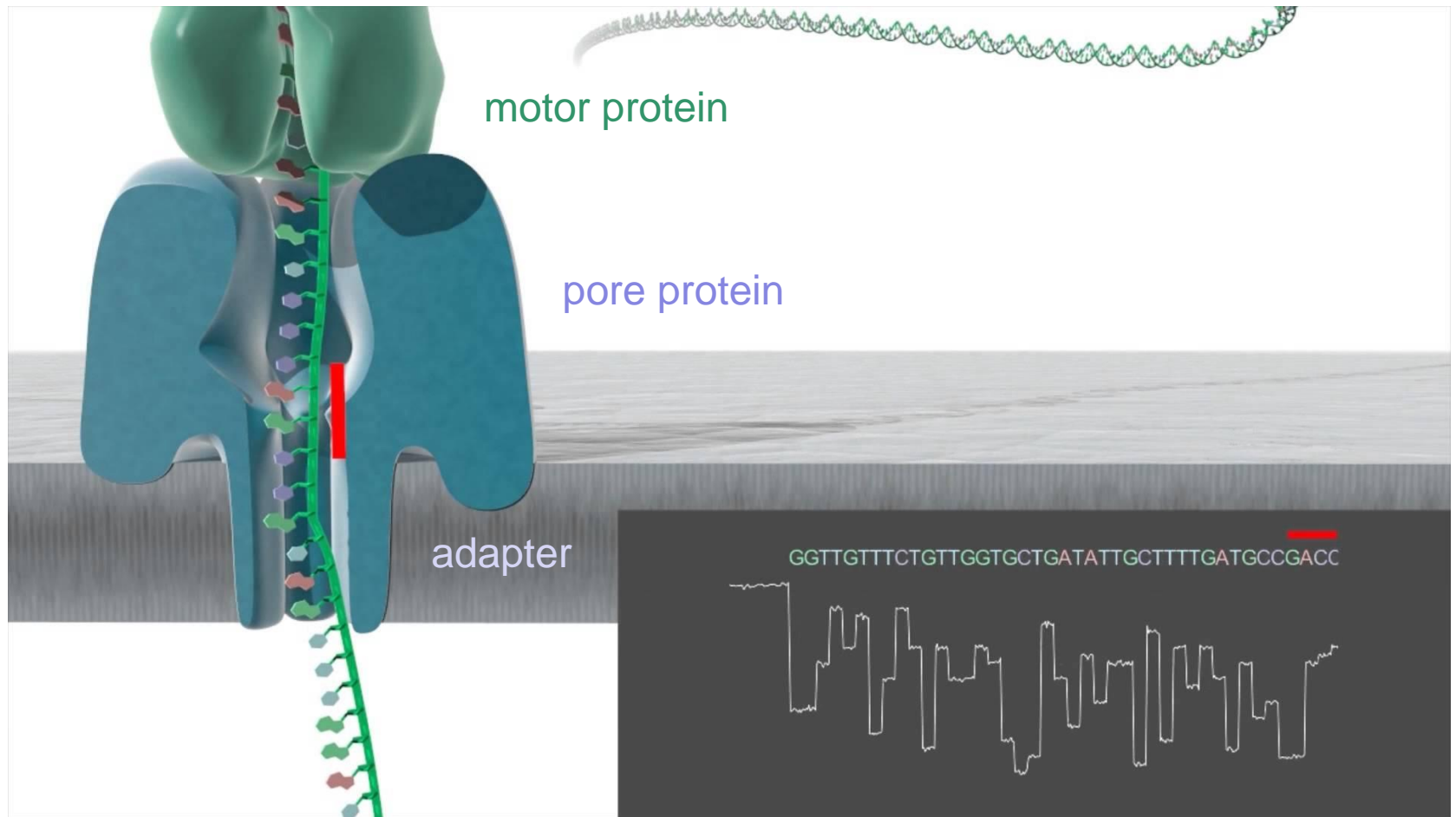
Sotaro Uemura^{1,2}, Colin Echeverría Aitken^{1,3†}, Jonas Korlach⁴, Benjamin A. Flusberg⁴, Stephen W. Turner⁴
& Joseph D. Puglisi^{1,3}

Translation by the ribosome occurs by a complex mechanism involving the coordinated interaction of multiple nucleic acid and protein ligands. Here we use zero-mode waveguides (ZMWs) and sophisticated detection instrumentation to allow real-time observation of translation at physiologically relevant micromolar ligand concentrations. Translation at each codon is monitored by stable binding of transfer RNAs (tRNAs)—labelled with distinct fluorophores—to translating ribosomes, which allows direct detection of the identity of tRNA molecules bound to the ribosome and therefore the underlying messenger RNA (mRNA) sequence. We observe the transit of tRNAs on single translating ribosomes and determine the number of tRNA molecules simultaneously bound to the ribosome, at each codon of an mRNA molecule. Our results show that ribosomes are only briefly occupied by two tRNA molecules and that release of deacylated tRNA from the exit (E) site is uncoupled from binding of aminoacyl-tRNA site (A-site) tRNA and occurs rapidly after translocation. The methods outlined here have broad application to the study of mRNA sequences, and the mechanism and regulation of translation.

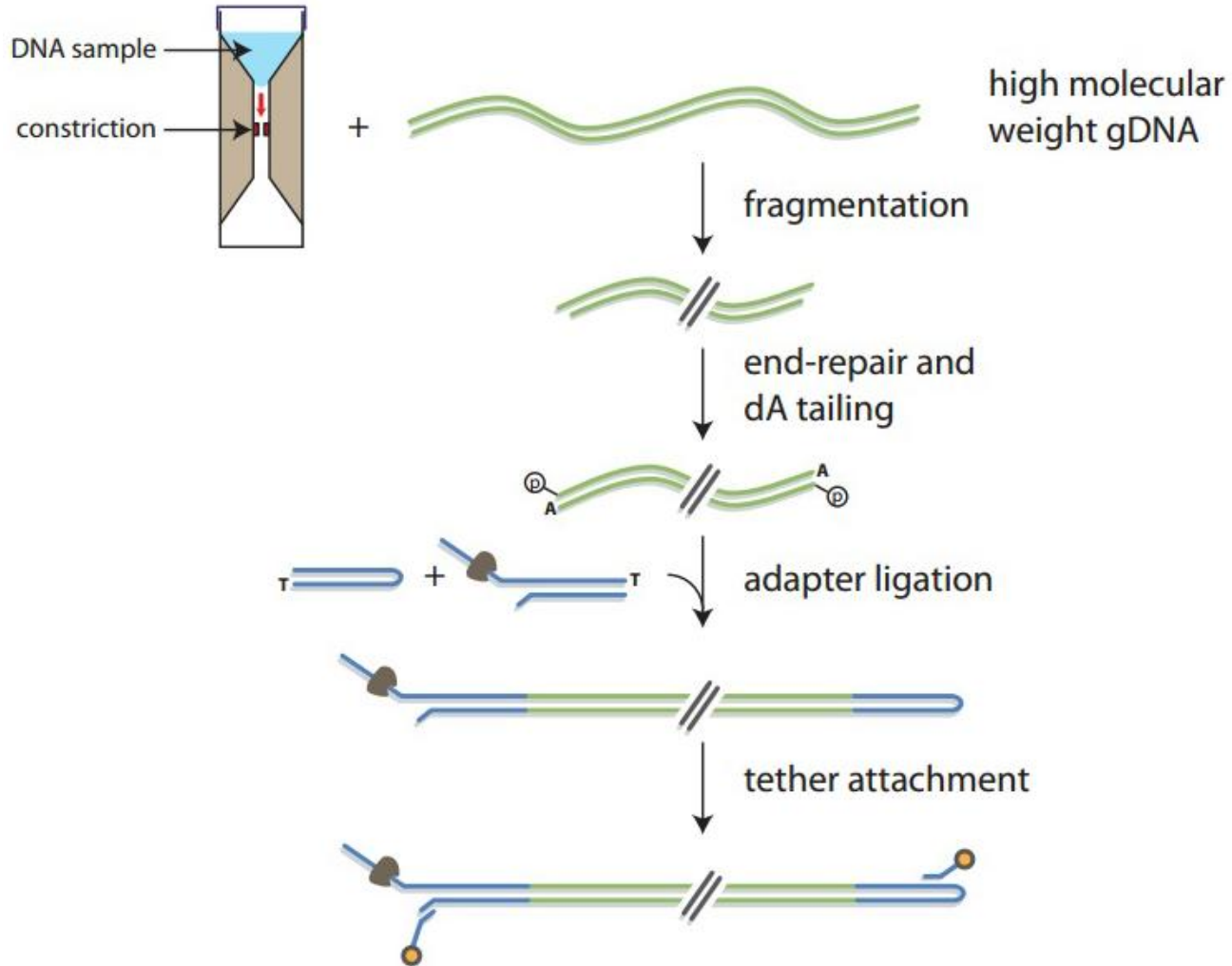


Oxford Nanopore

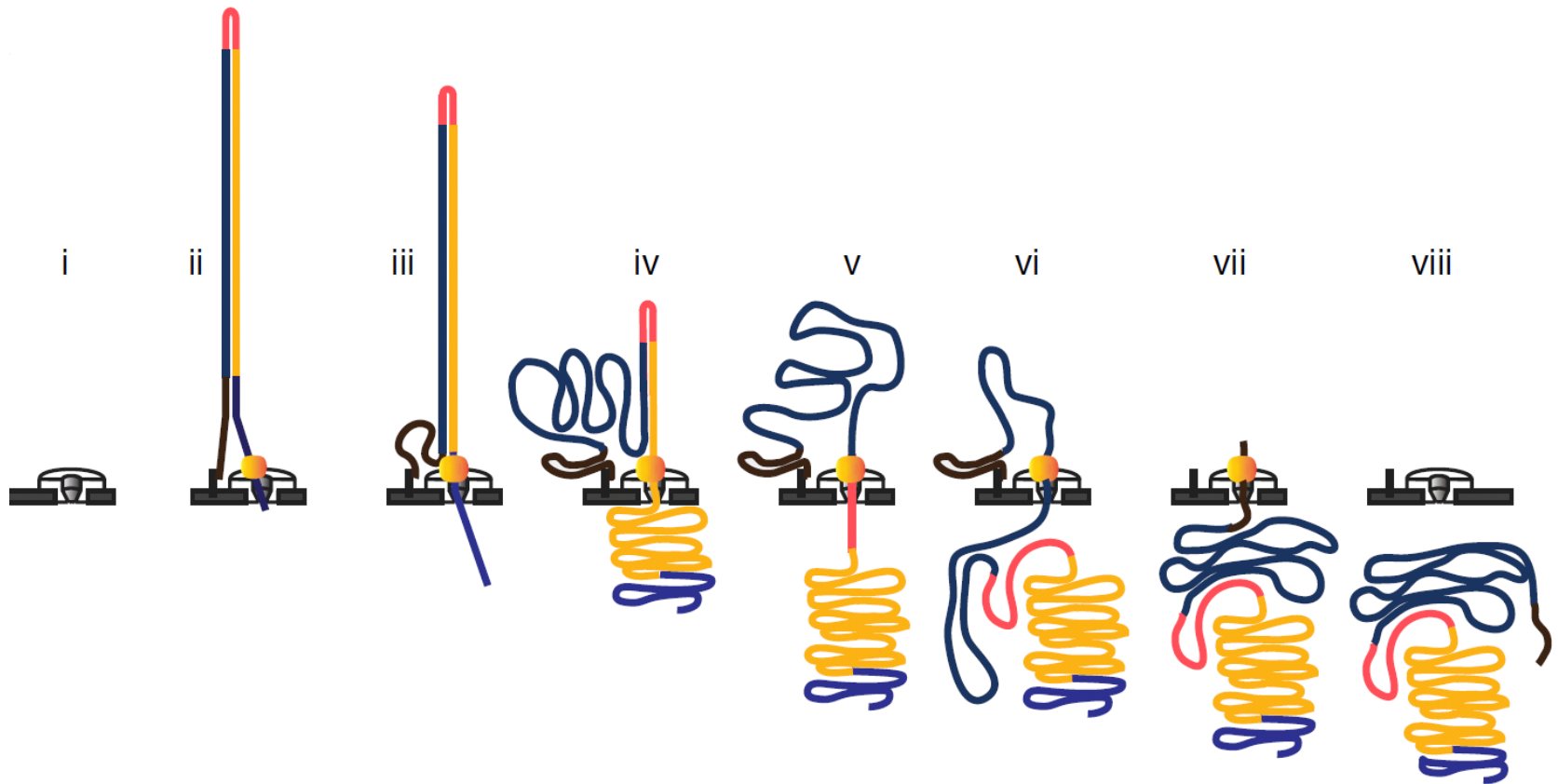
Principle



Oxford Nanopore Library



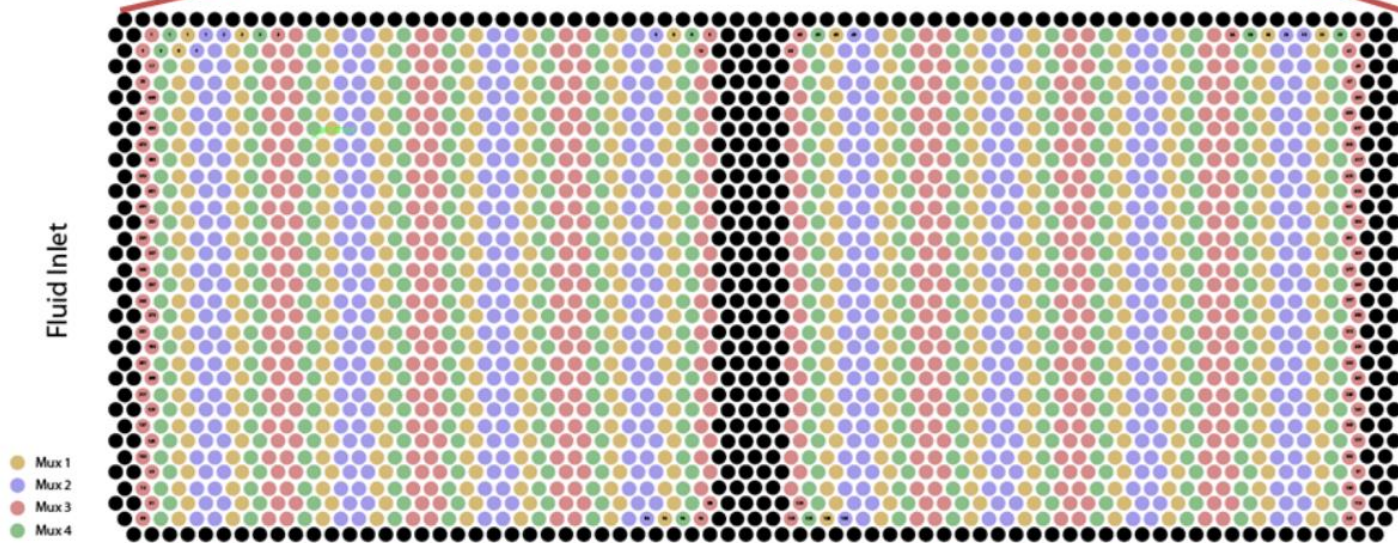
Oxford Nanopore Sequencing



error rate: 10-30%

Oxford Nanopore

Data output



512 sequencing channels → each channel sequences multiple molecules

~ 400 Mb

Oxford Nanopore

Data output

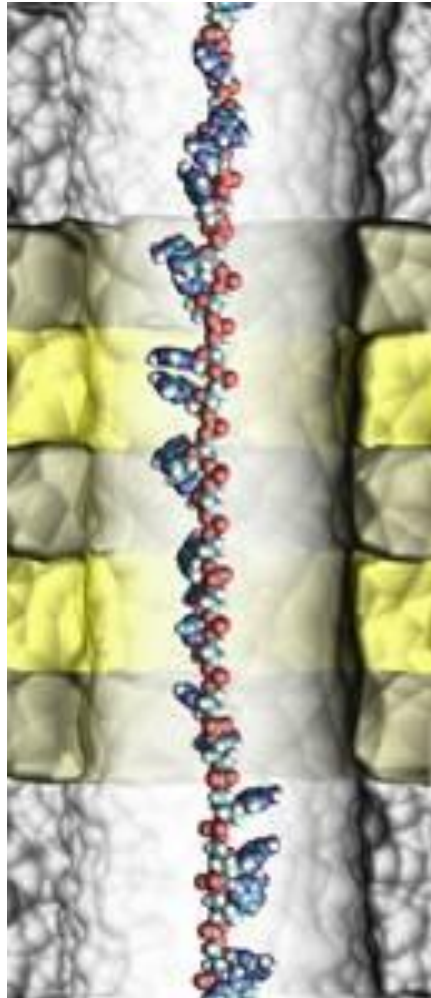


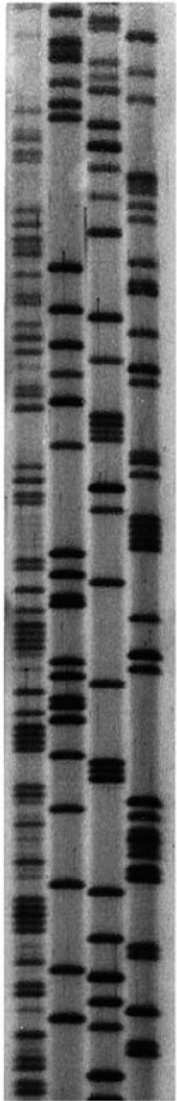
PROMETHION

- 300X greater throughput than MinION
- 2016

IBM

Waver pore





A C G T

genome.leibniz-fli.de Lectures