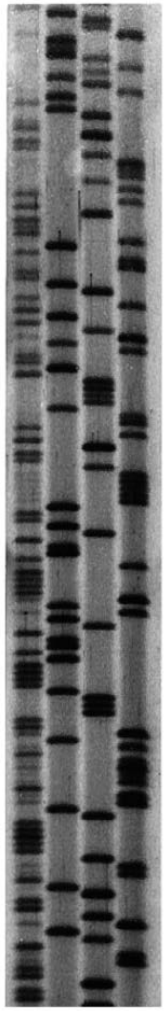


DNA-Sequencing

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

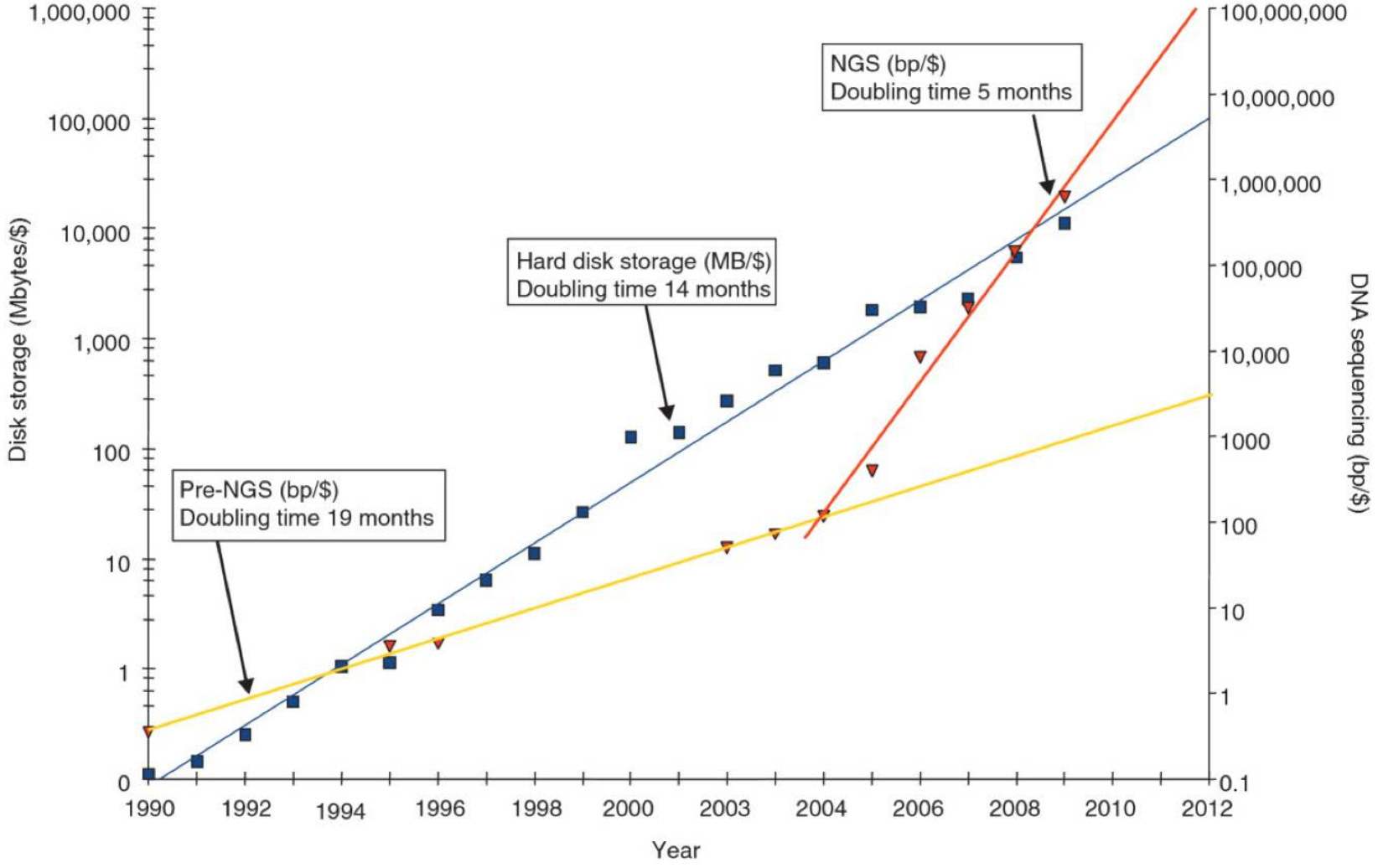
PacBio RS II

2/2015

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

Moore's law

Next Generation Sequencing - a game changer



http://ivory.idyll.org/permanent/lstein-ngs-capacity.png

Next Generation Sequencing - NGS

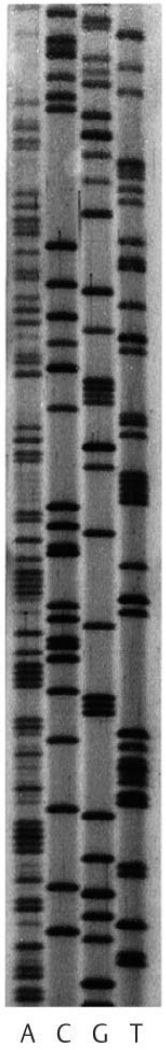
Terminologies

2nd generation

= massively parallel

3rd generation

= single-molecule

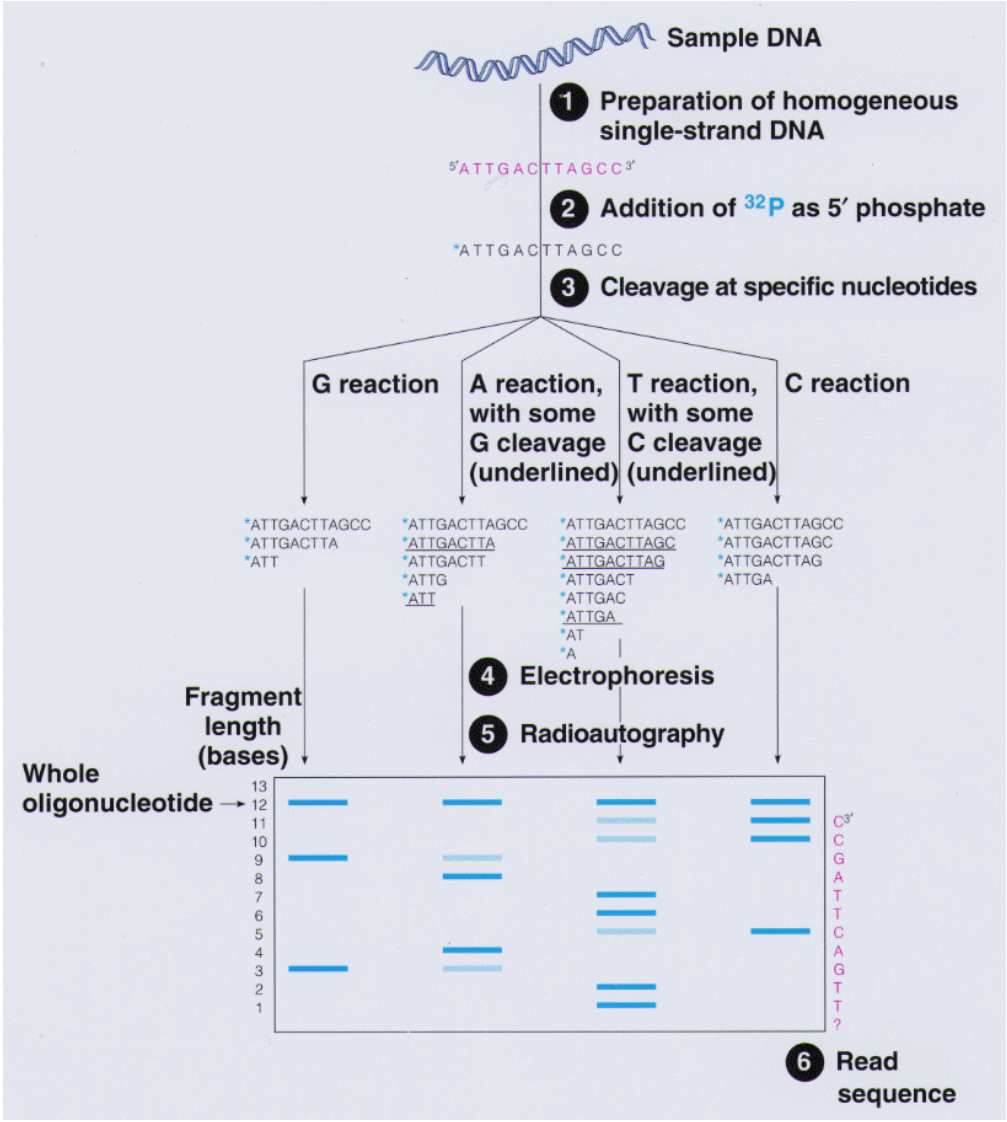


1st Generation

Maxam-Gilbert . Sanger . Pyro

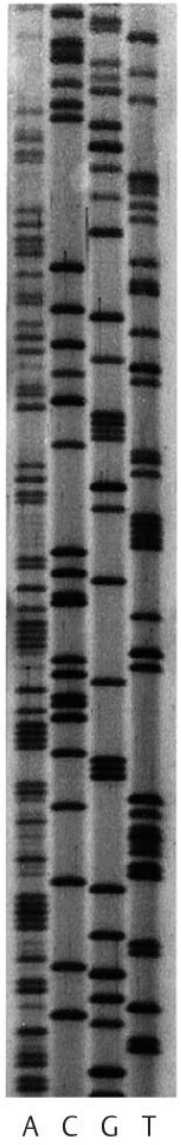
Maxam-Gilbert sequencing

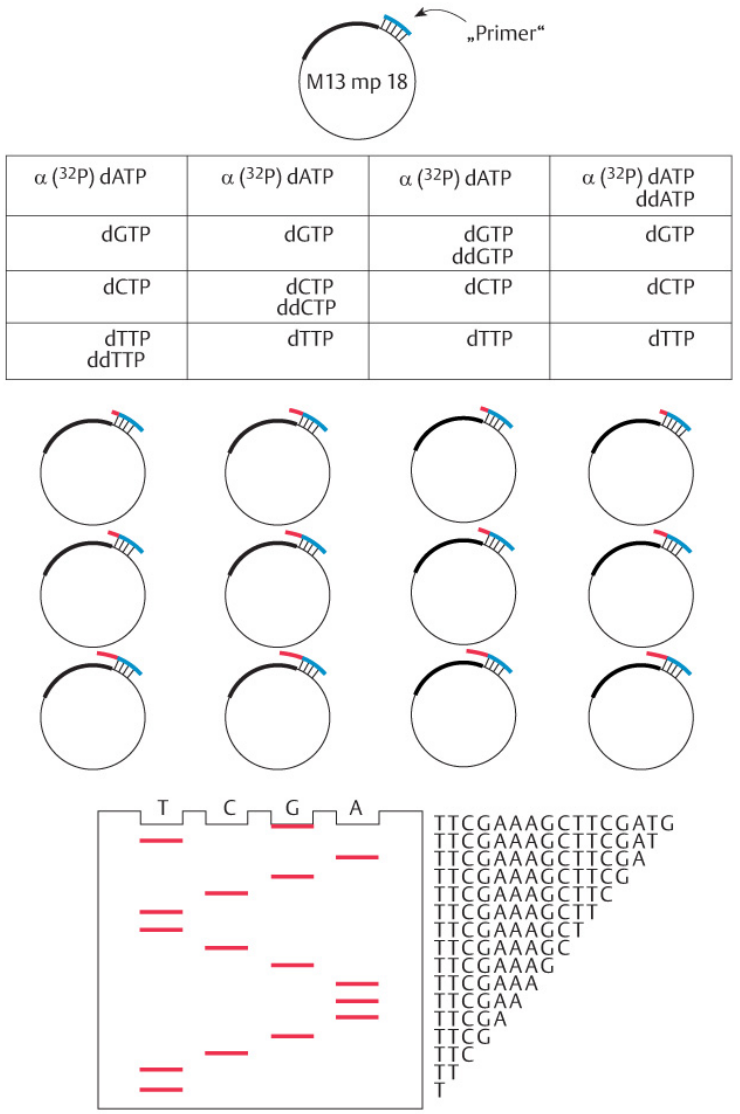
Chemical modifications, radioactive *1977



Sanger sequencing

Dideoxy chain termination, radioactive *1977



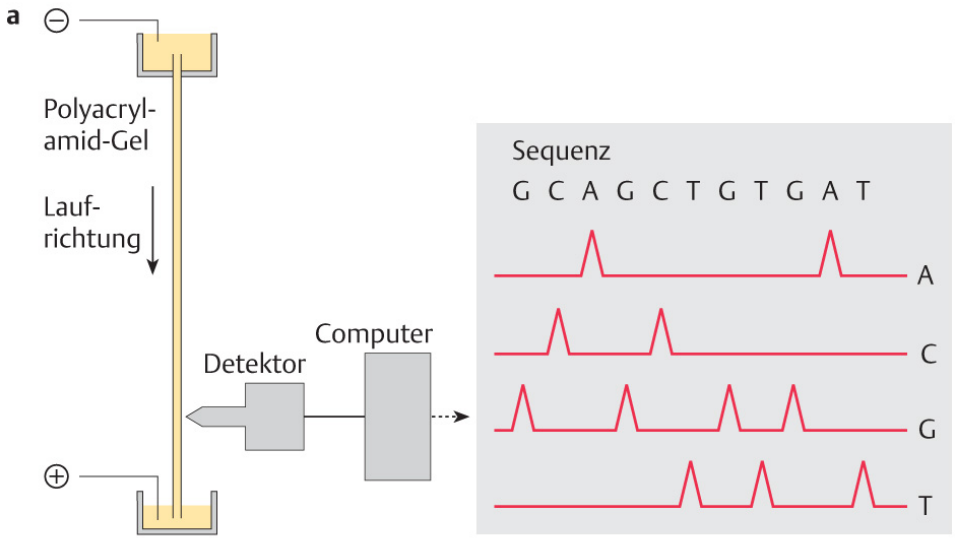


Fred Sanger

*Nobel prices 1958, **1980***

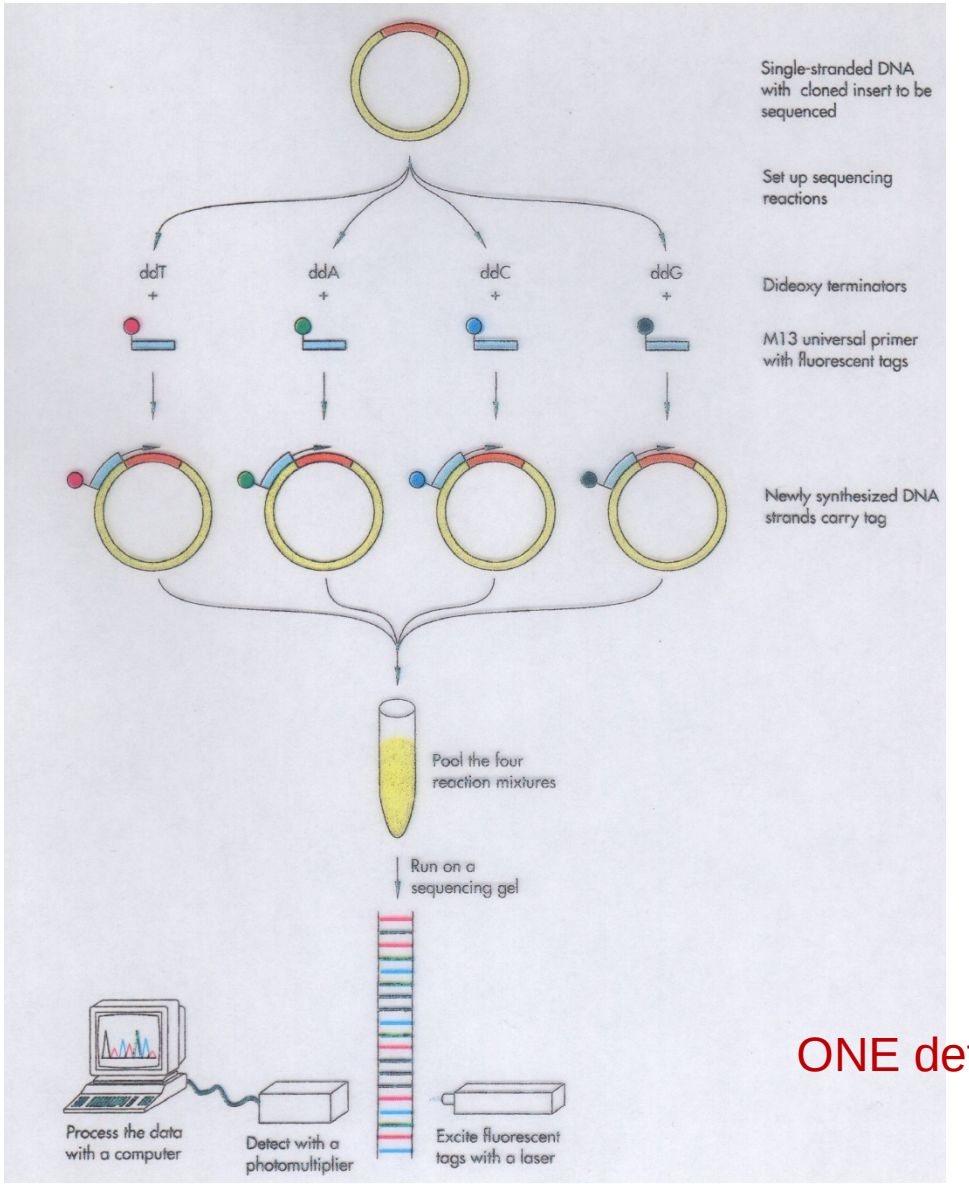
Sanger sequencing

Fluorescence labeling



Sanger sequencing

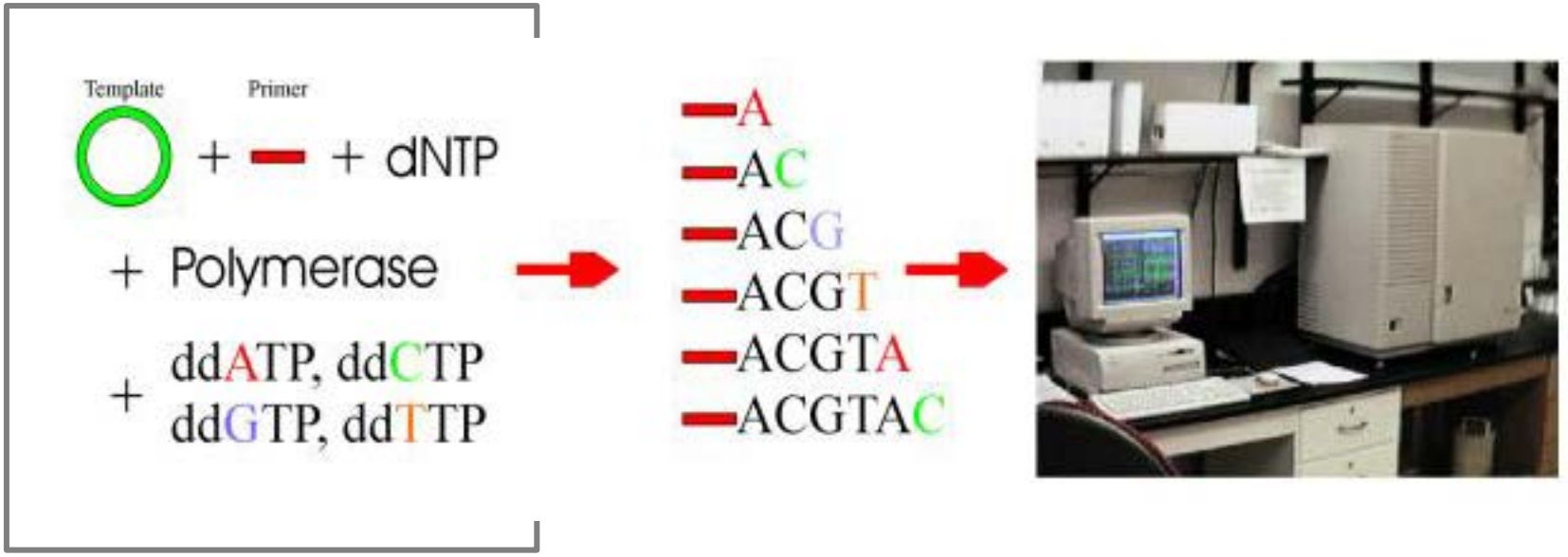
Four dyes > Dye primer technology



ONE detection

Sanger sequencing

Four dyes > Dye terminator technology

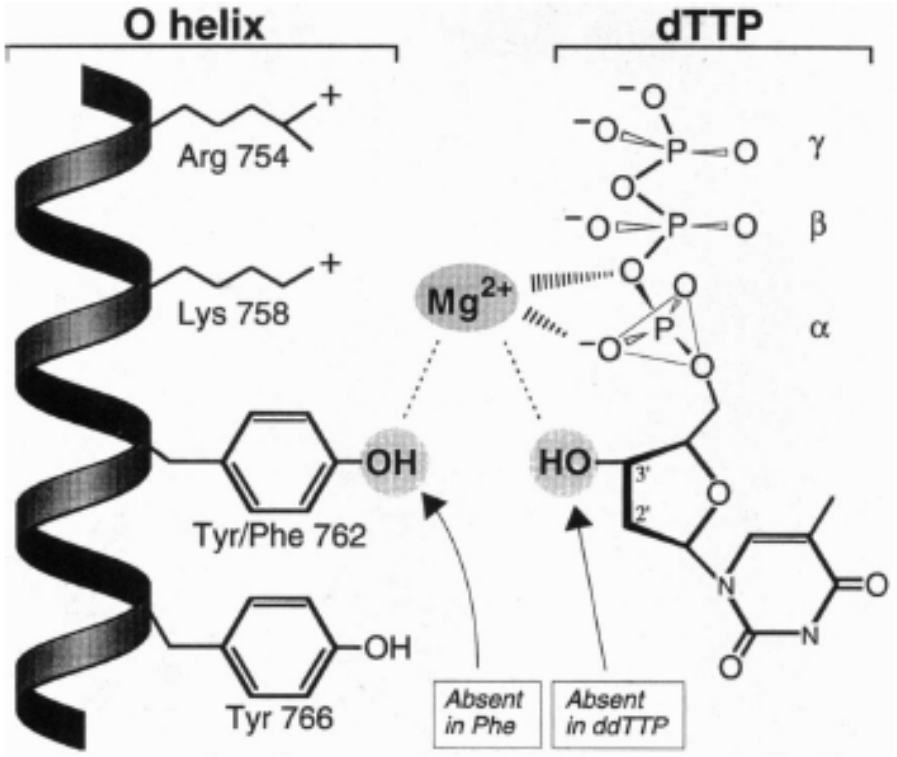


ONE reaction

but artificial substrates

Sanger sequencing

Engineered polymerase



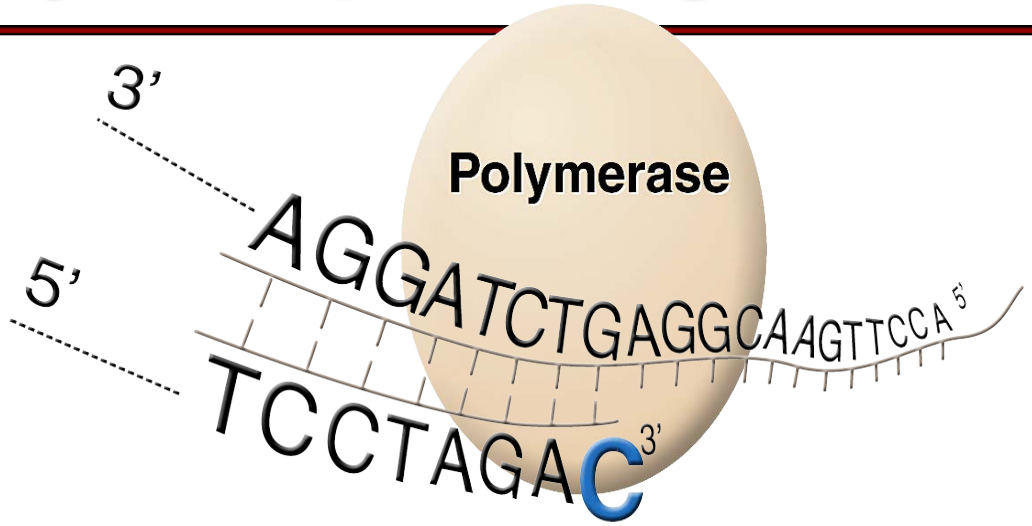
Thermo-Sequenase = Taq polymerase Phe>Thy mutated

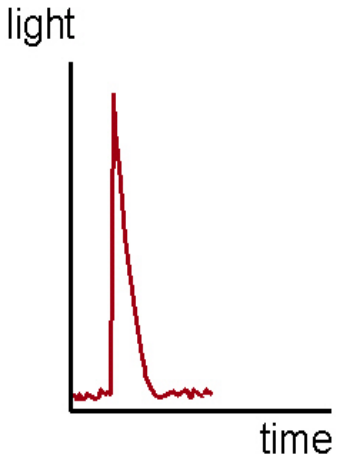
PNAS 92:6339 (1995)

Sanger sequencing

Cycle sequencing

Pyrosequencing



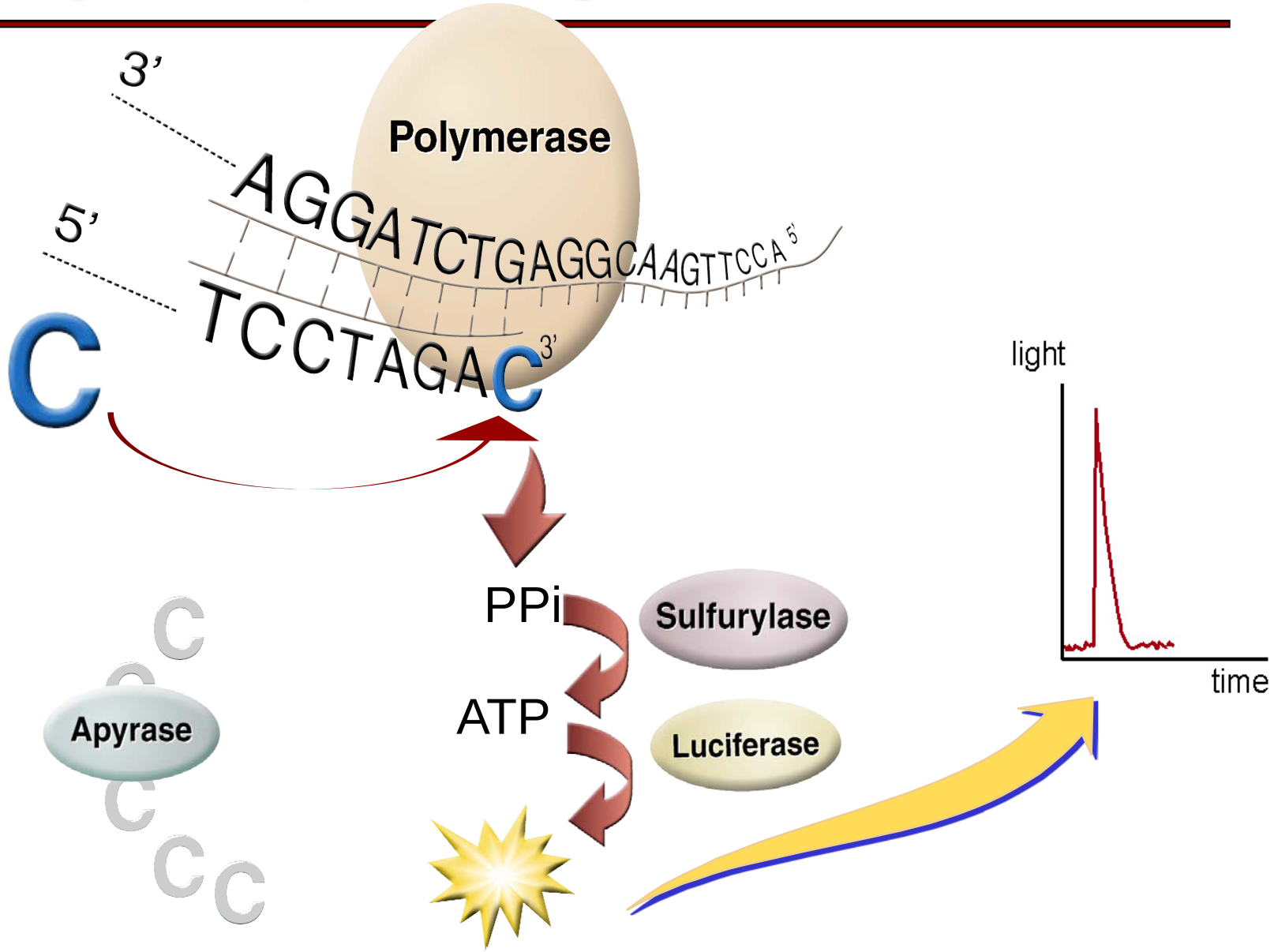


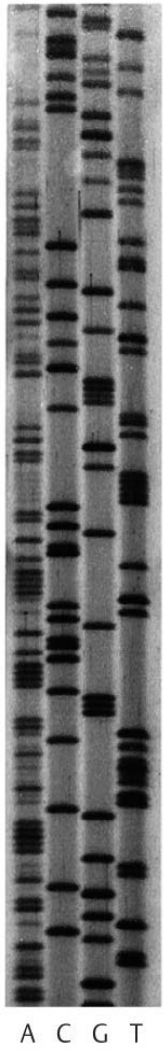
Sulfurylase

Apyrase

Luciferase

Pyrosequencing





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……..

Nature, July 31 2005

doi:10.1038/nature03959 nature

ARTICLES

Genome sequencing in microfabricated high-density picolitre reactors

Marcel Margulies[1]*, Michael Egholm[1]*, William E. Altman[1], Said Attiya[1], Joel S. Bader[1], Lisa A. Bemben[1], Jan Berka[1], Michael S. Braverman[1], Yi-Ju Chen[1], Zhoutao Chen[1], Scott B. Dewell[1], Lei Du[1], Joseph M. Fierro[1], Xavier V. Gomes[1], Brian C. Godwin[1], Wen He[1], Scott Helgesen[1], Chun He Ho[1], Gerard P. Irzyk[1], Szilveszter C. Jando[1], Maria L. I. Alenquer[1], Thomas P. Jarvie[1], Kshama B. Jirage[1], Jong-Bum Kim[1], James R. Knight[1], Janna R. Lanza[1], John H. Leamon[1], Steven M. Lefkowitz[1], Ming Lei[1], Jing Li[1], Kenton L. Lohman[1], Hong Lu[1], Vinod B. Makhijani[1], Keith E. McDade[1], Michael P. McKenna[1], Eugene W. Myers[2], Elizabeth Nickerson[1], John R. Nobile[1], Ramona Plant[1], Bernard P. Puc[1], Michael T. Ronan[1], George T. Roth[1], Gary J. Sarkis[1], Jan Fredrik Simons[1], John W. Simpson[1], Maithreyan Srinivasan[1], Karrie R. Tartaro[1], Alexander Tomasz[3], Kari A. Vogt[1], Greg A. Volkmer[1], Shally H. Wang[1], Yong Wang[1], Michael P. Weiner[4], Pengguang Yu[1], Richard F. Begley[1] & Jonathan M. Rothberg[1]

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.

454

Nature, July 31 2005

- 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

DNA Isolation

↓

DNA Fragmentation

↓

DNA Amplification

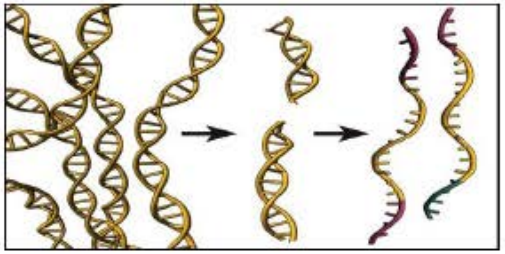
↓

DNA Sequencing

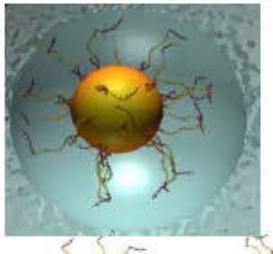
↓

Bioinformatics

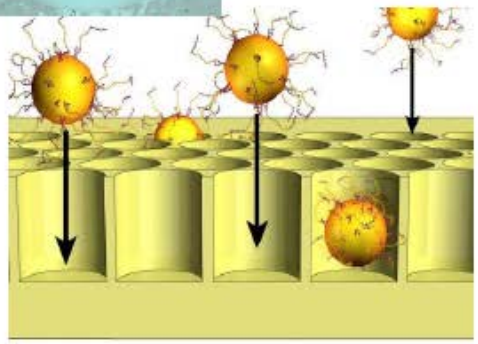
- Contigs are constructed



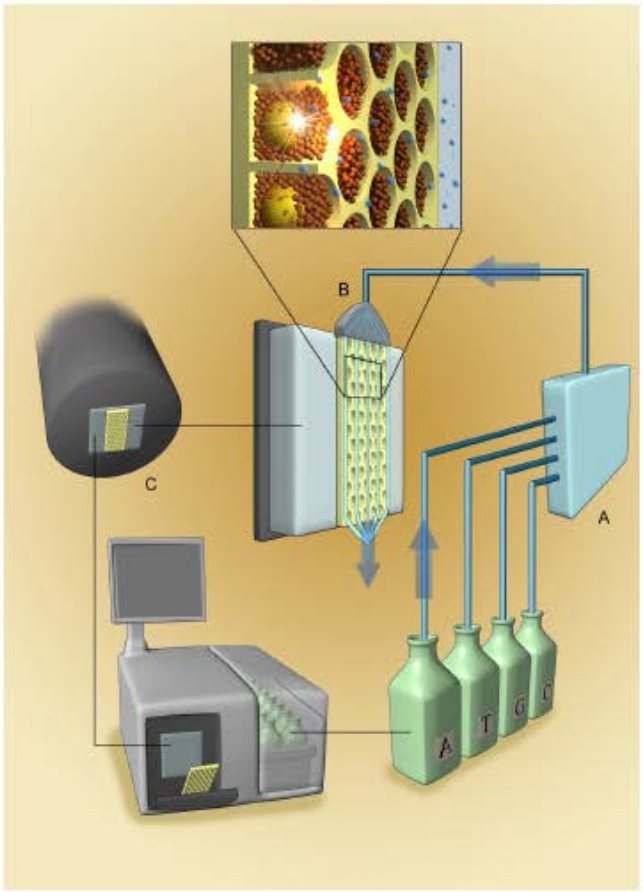
Adapter Ligated ssDNA Library



Clonal Amplification

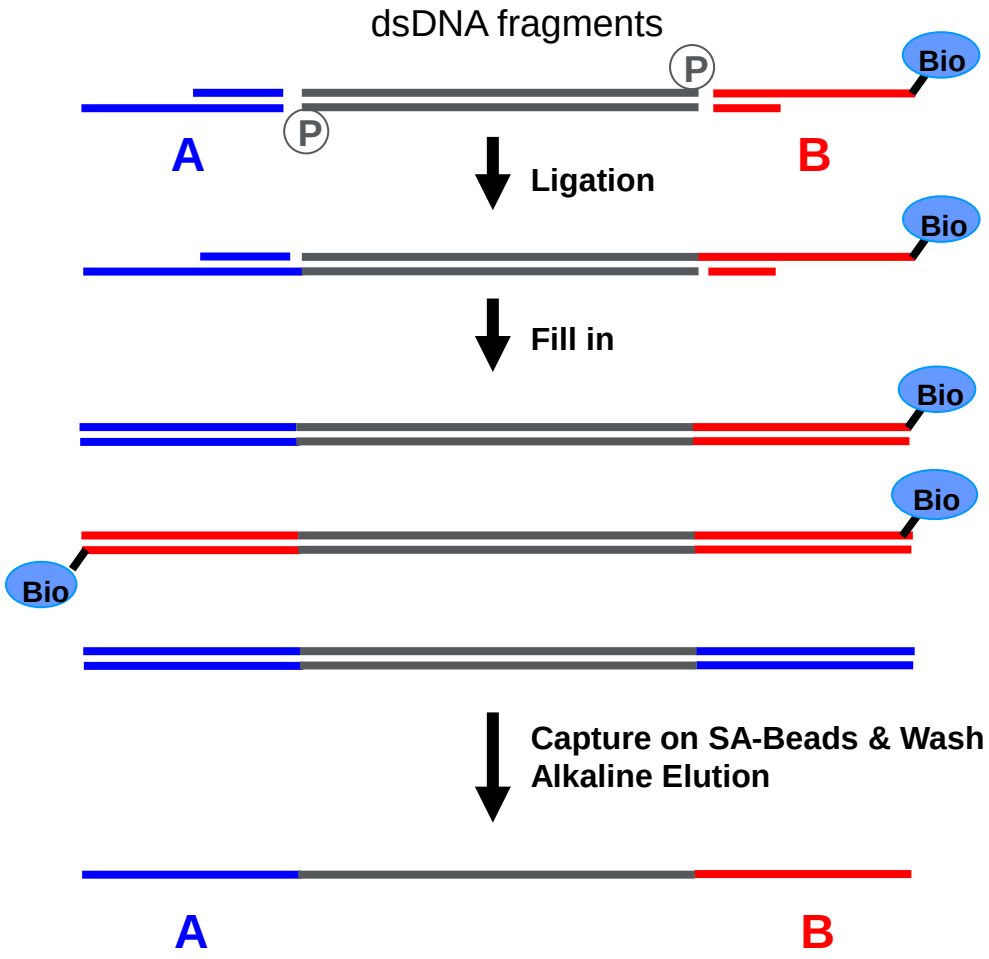


Beads enzymes in PicoTiter Plate™



Roche/454

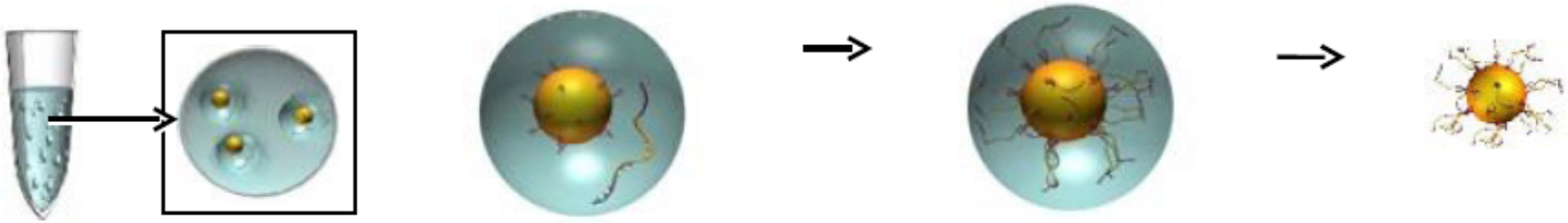
Preparation of A/B-fragments for sequencing



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

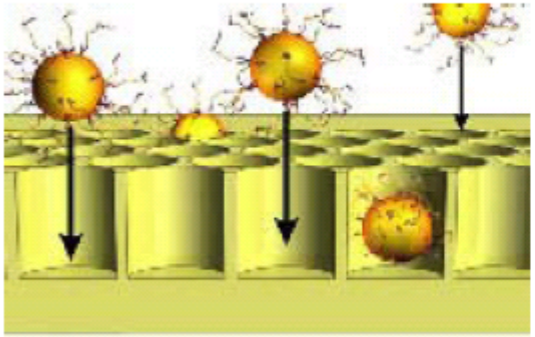
Break microreactors, enrich for DNA-positive beads

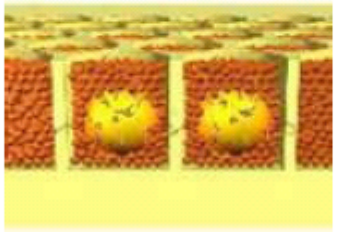
sstDNA library → Clonally-amplified sstDNA attached to bead

Roche/454

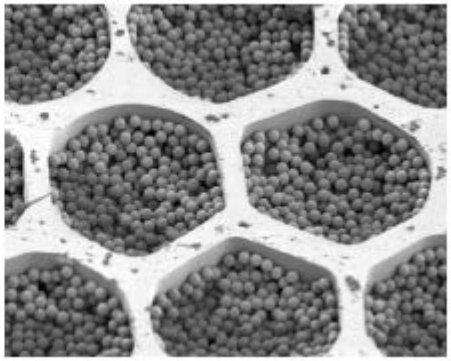
Sample loading

Depositing DNA beads into the PicoTiterPlate device



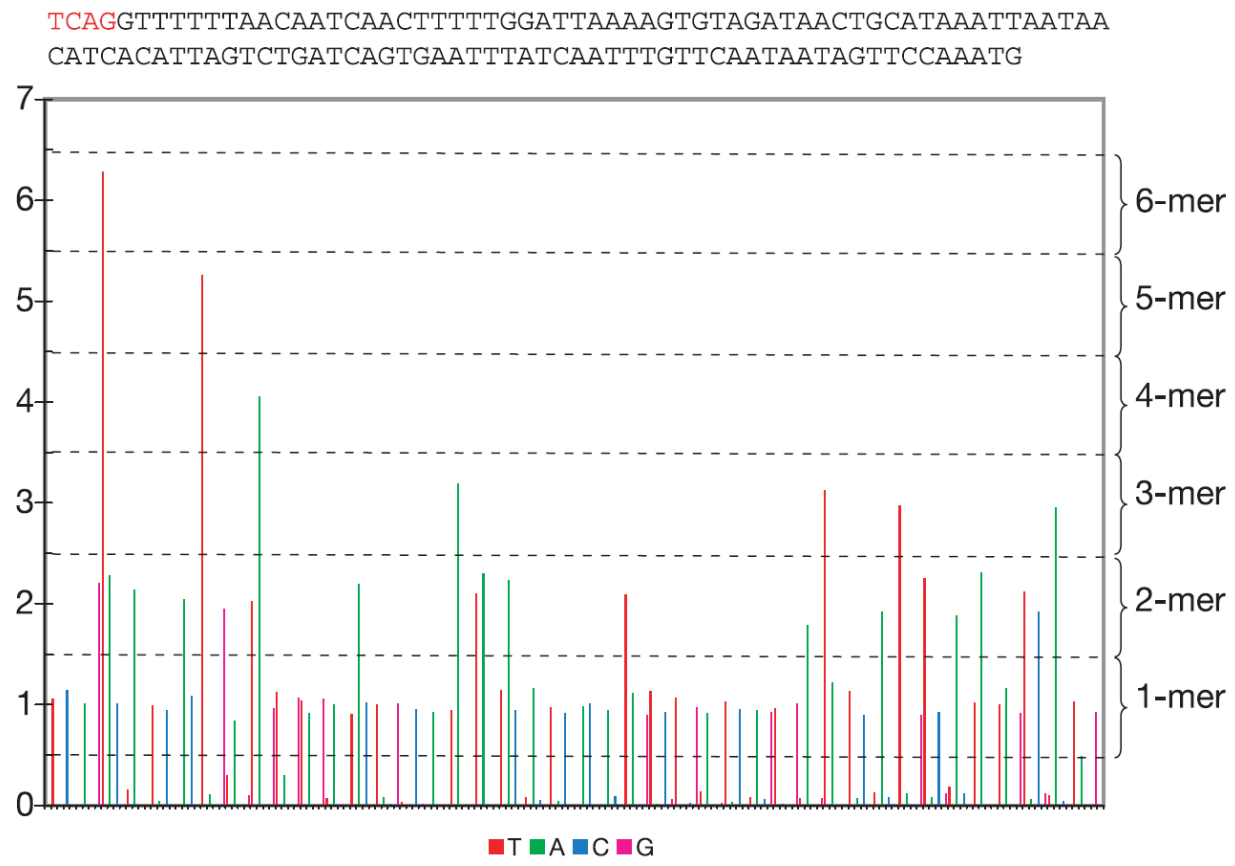


- Well diameter: average of 44 µm
- 200,000 reads obtained in parallel
- A single clonally amplified sstDNA bead is deposited per well



Roche/454

Flowgram



Roche/454

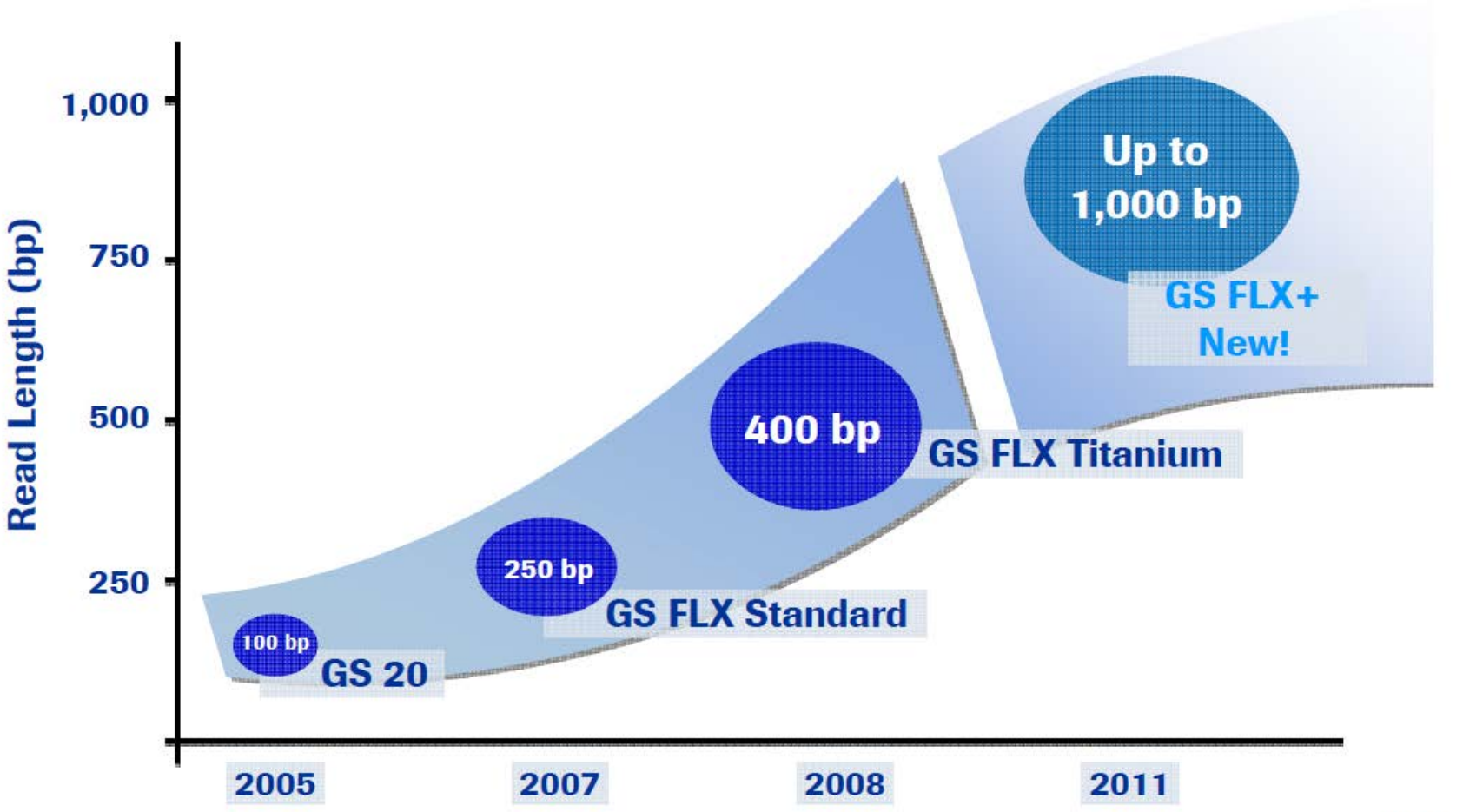
No cloning bias

GC poor
low-cloned coverage

GC normal
high-cloned coverage

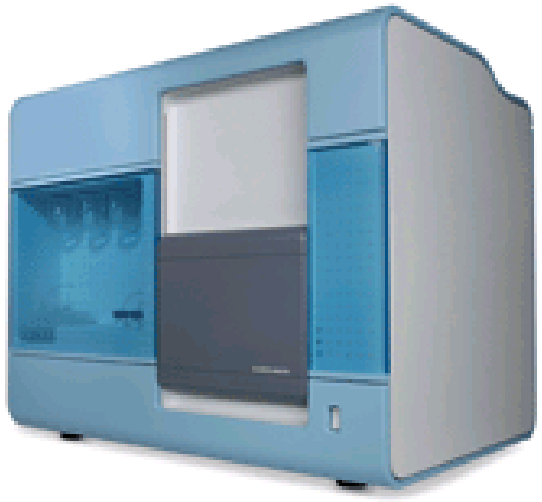
Roche/454

Read length



454 SEQUENCING

Solexa™



Advanced genetic analysis

one billion bases at a time

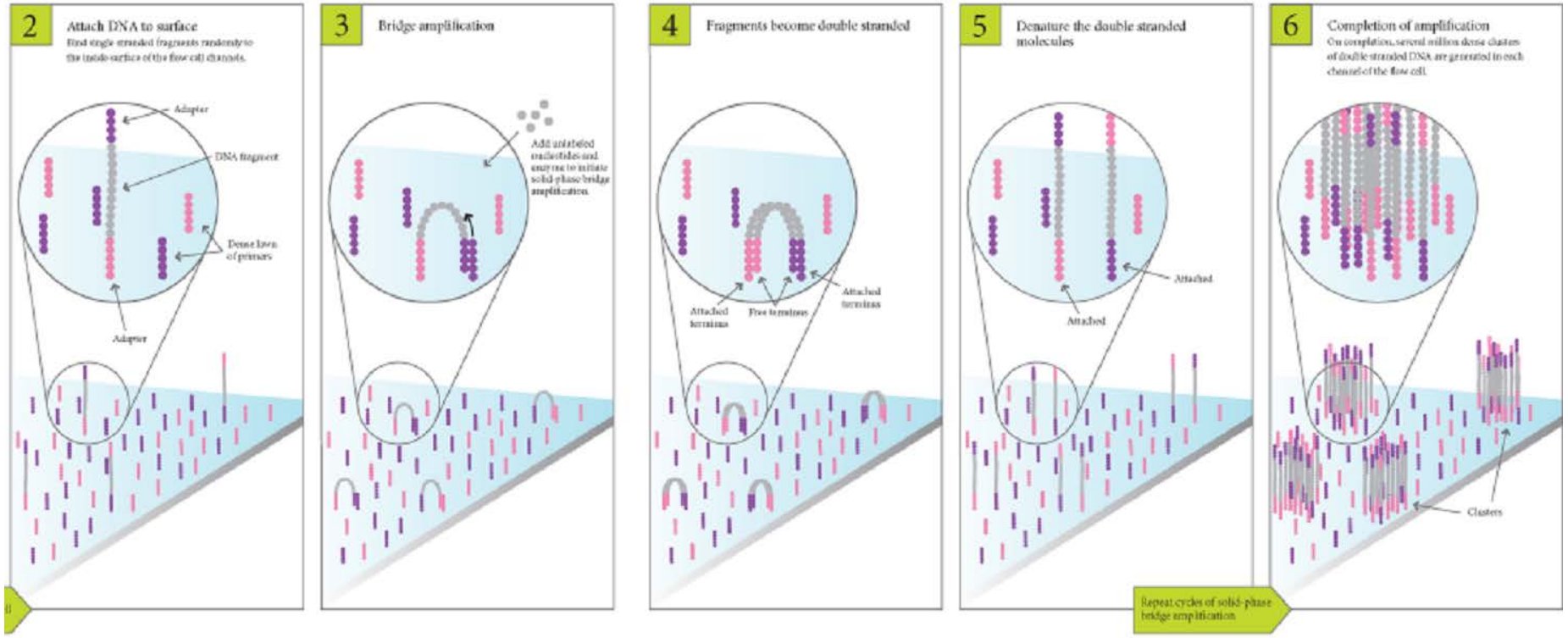
Solexa

Technology

- reversible fluorescent labels
- reversible 3’OH blocking

Illumina/Solexa

Solid-phase clonal single molecule PCR



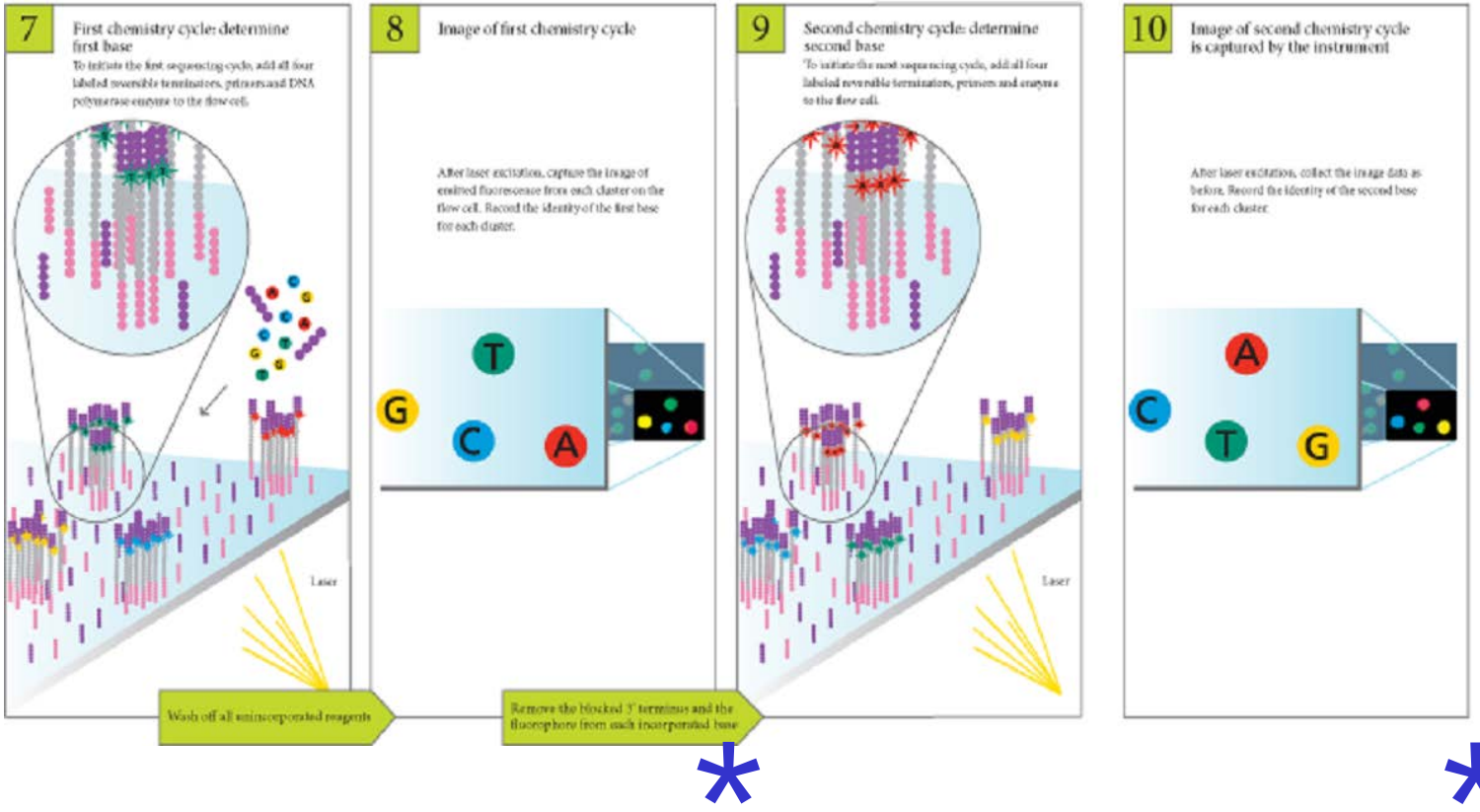
"Bridge amplification"

Illumina/Solexa

Sequencing by synthesis (SBS)

1st cycle

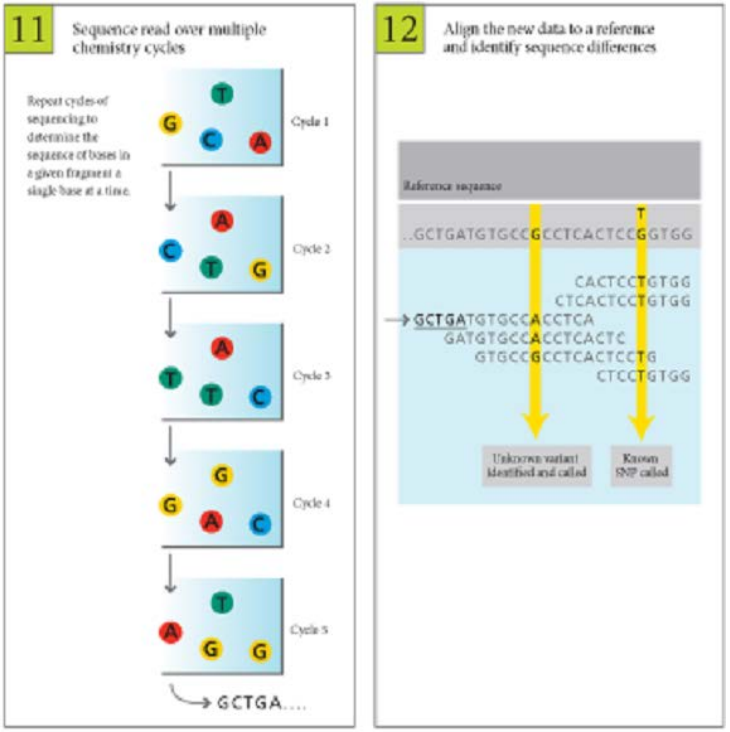
2nd cycle



*removal of fluorescent labels & 3'OH blocking

Illumina/Solexa

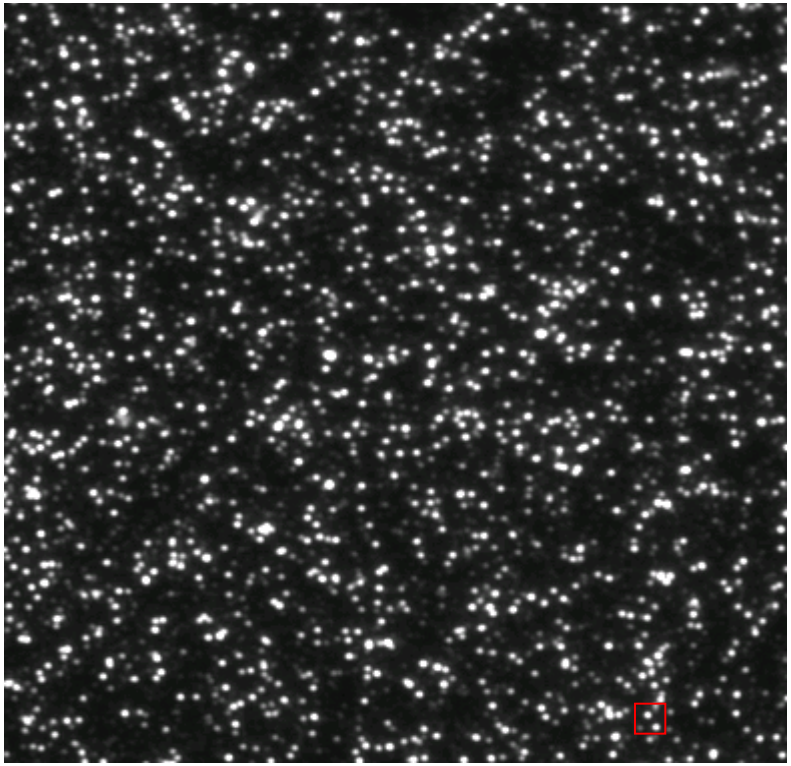
Assembly - Mapping



read length: 50 … 250 nt

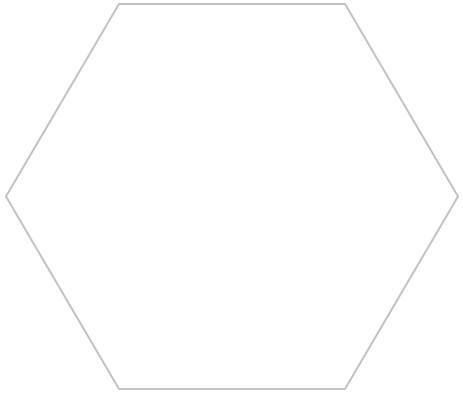
Illumina/Solexa

Solid-phase clonal single molecule PCR



100µm

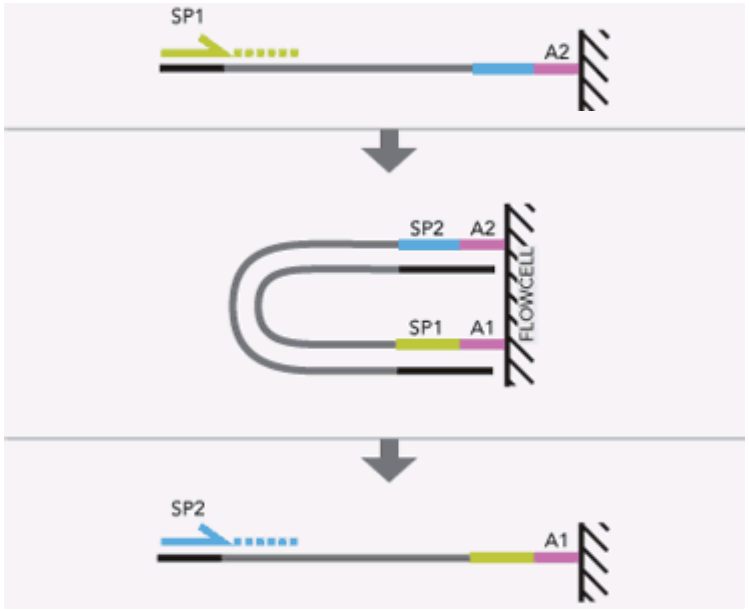
colony of ≈1000 single-stranded DNA templates

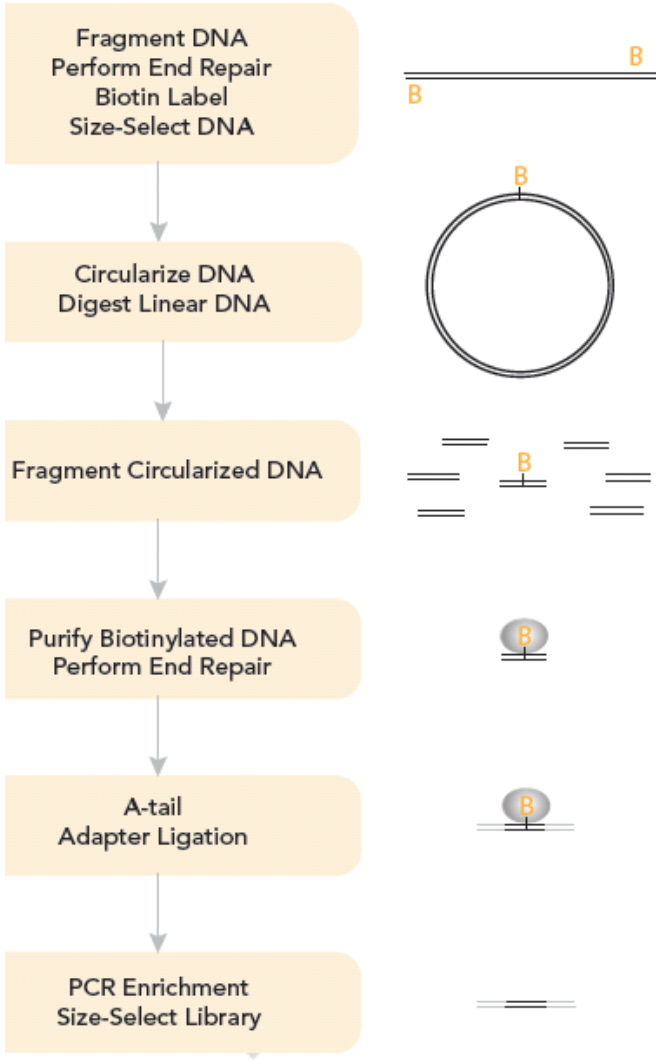


Single well of 454 Life Sciences PicoTiterPlate™ (to same scale)

Illumina/Solexa

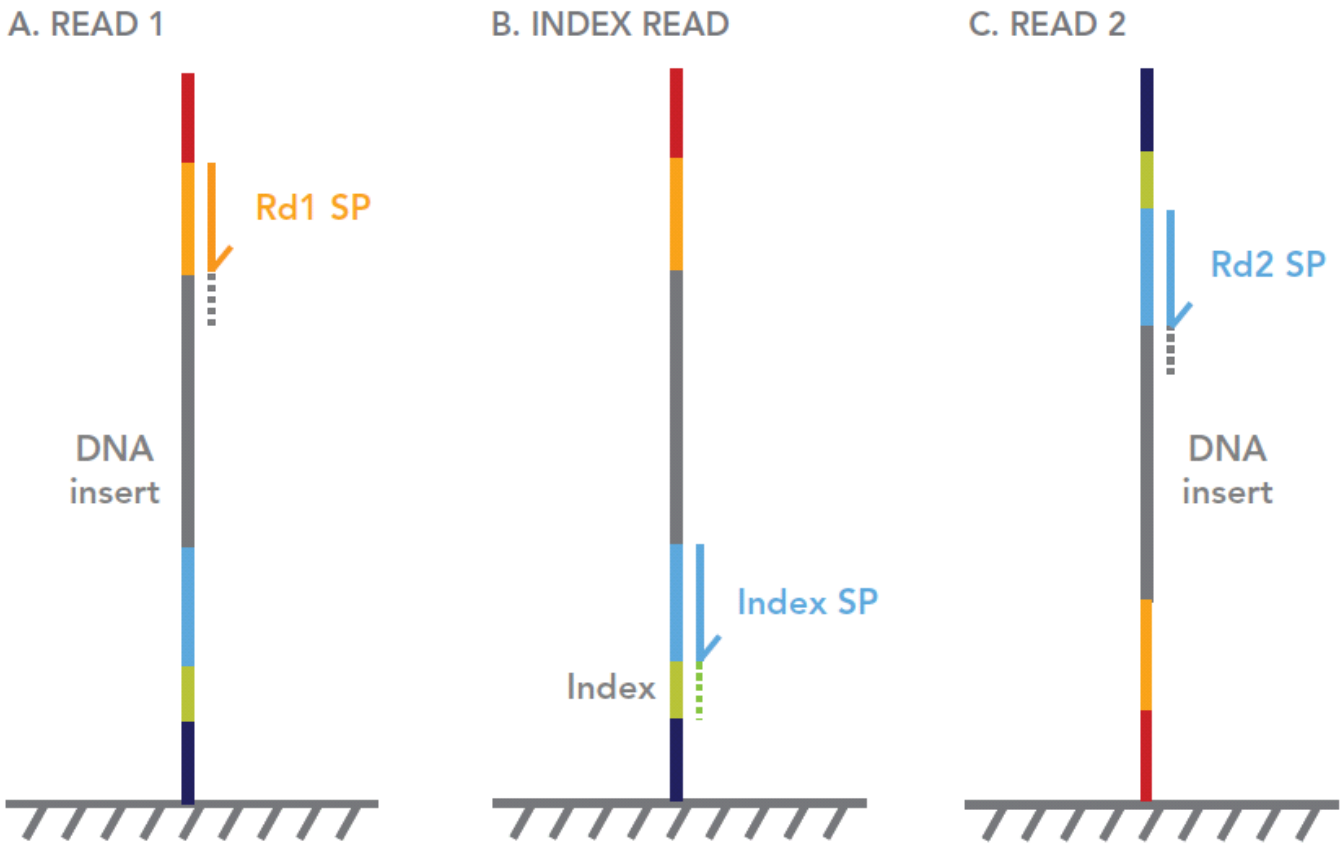
Paired ends & Mate pairs





Illumina/Solexa

Multiplexing



Illumina/Solexa

Data output

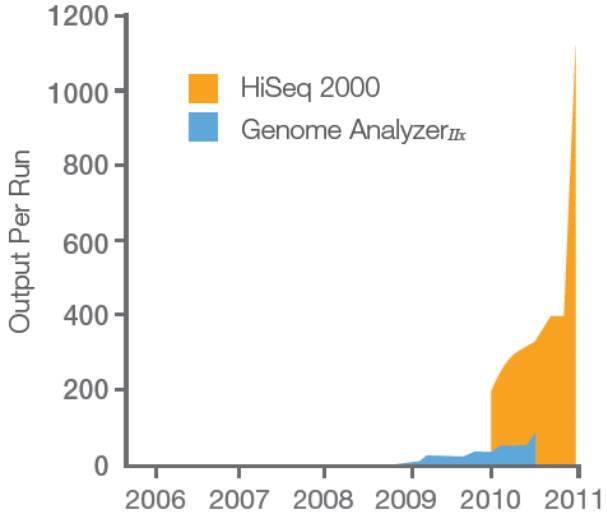
1,000-fold improvement in just four years.

Sequencing Run Parameters

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

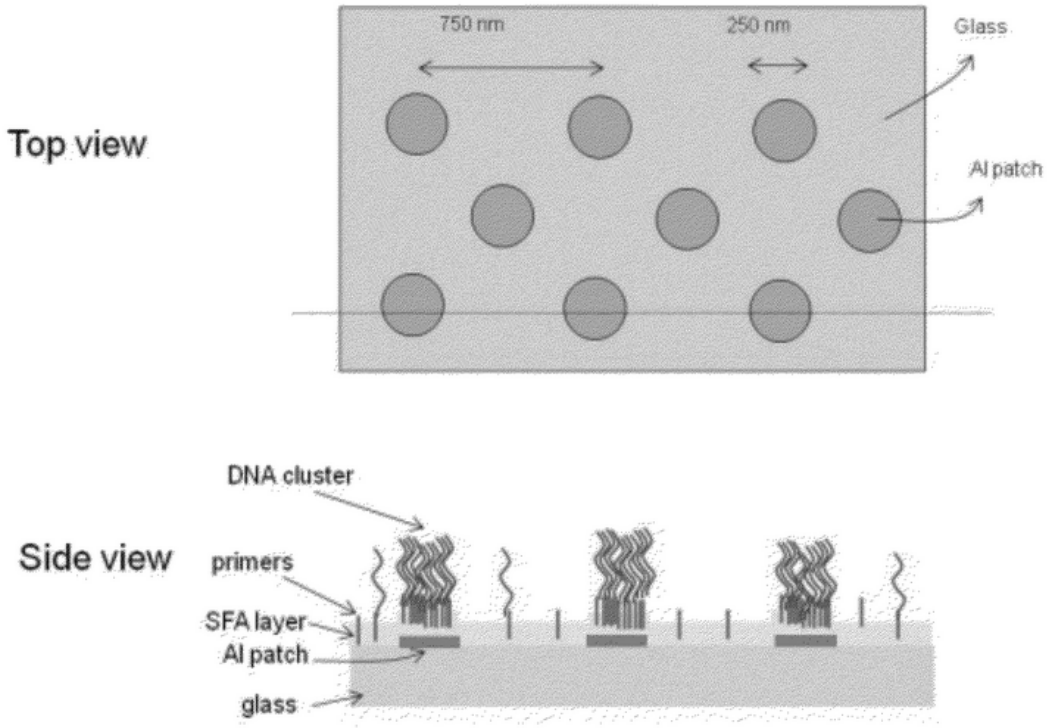
Over the last few years we've scaled from targeting 1 Gb on the Genome Analyzer, to generating in excess of 1 Tb* internally on the HiSeq 2000.

* R&D configuration not supported at launch



Illumina/Solexa

Patterned flow cells



Patent US 20120316086 A1 (2012)

Illumina/Solexa

Patterned flow cells

New platforms

2014

2017

NextSeq 500

30 Gb/day

1 human genome or 16 exomes

HiSeq3000/4000

HiSeqX Ten

150 Gb/day

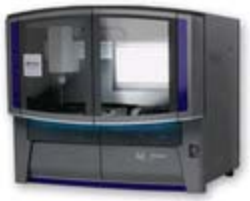
5 human genome at $1000 each

NovaSeq 6000

1.5 Tb/day

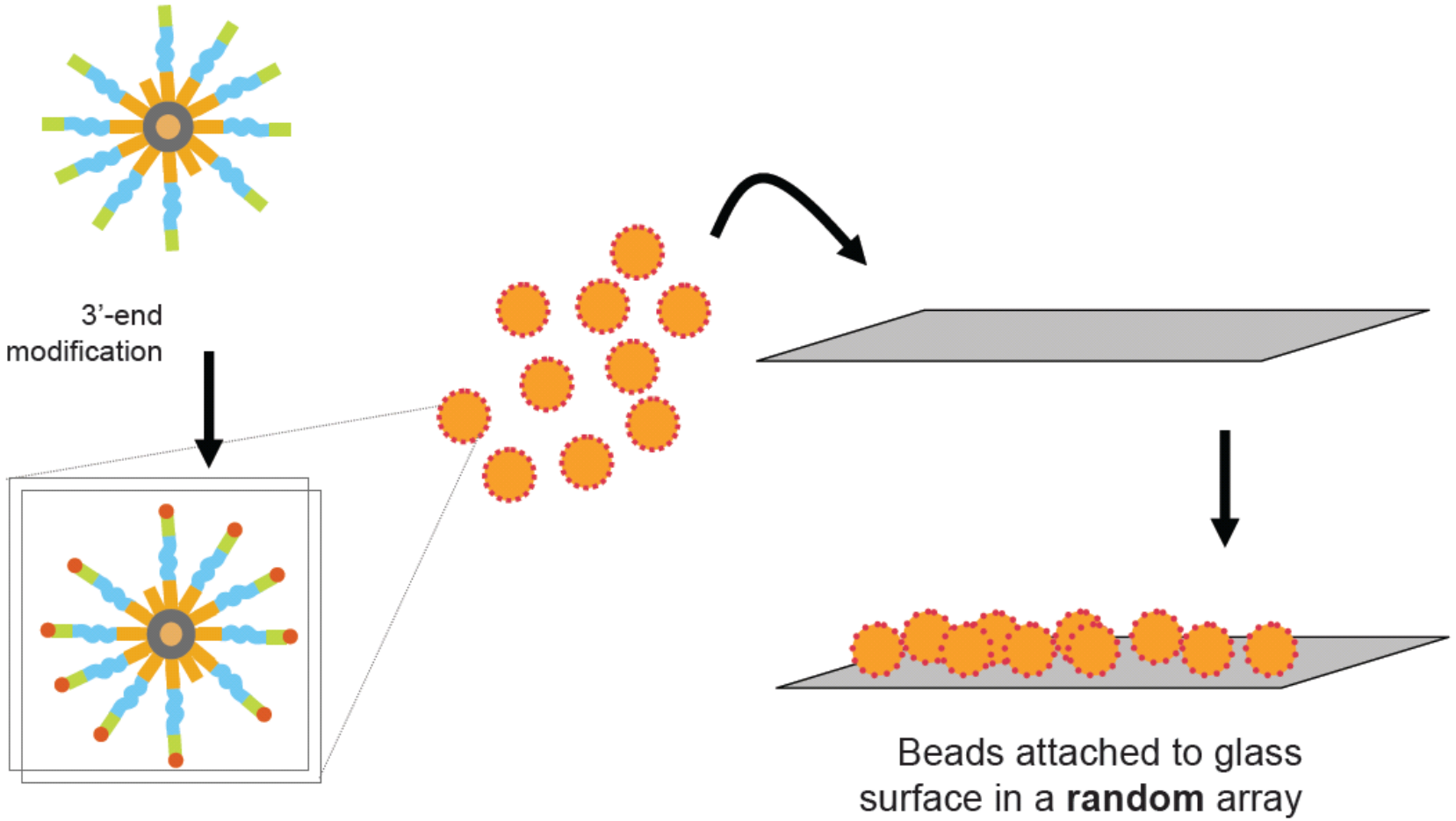
Agencourt


The Next Generation
is SOLiD™
Sequencing by Oligonucleotide Ligation and Detection



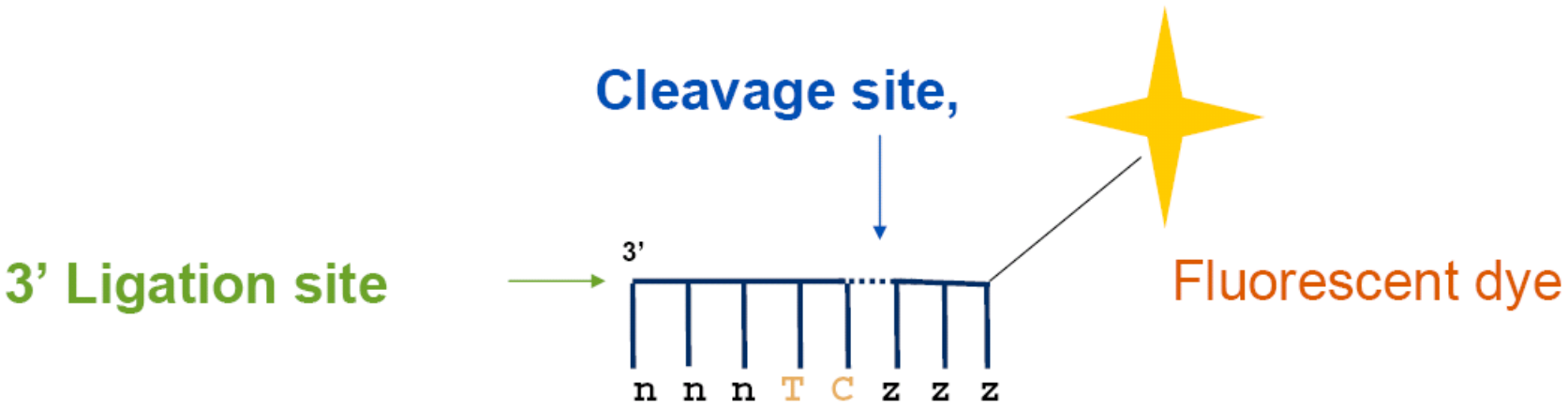
ABI/SOLID

Bead deposition



SOLID

Probes



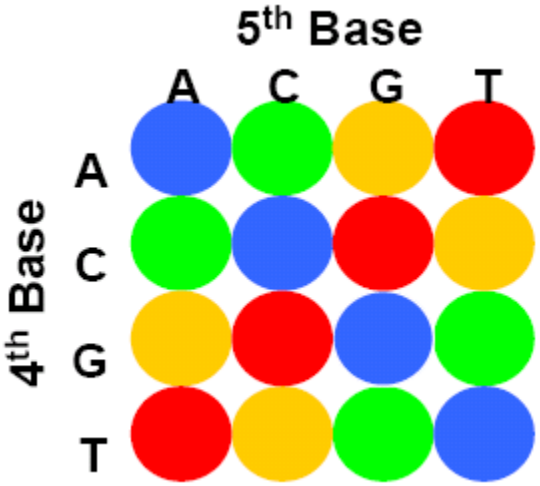
1,024 Octamer Probes (4^5)

4 Dyes,16dinucleotides, 256 probes per dye

N= degenerate bases Z= Universal bases

SOLID

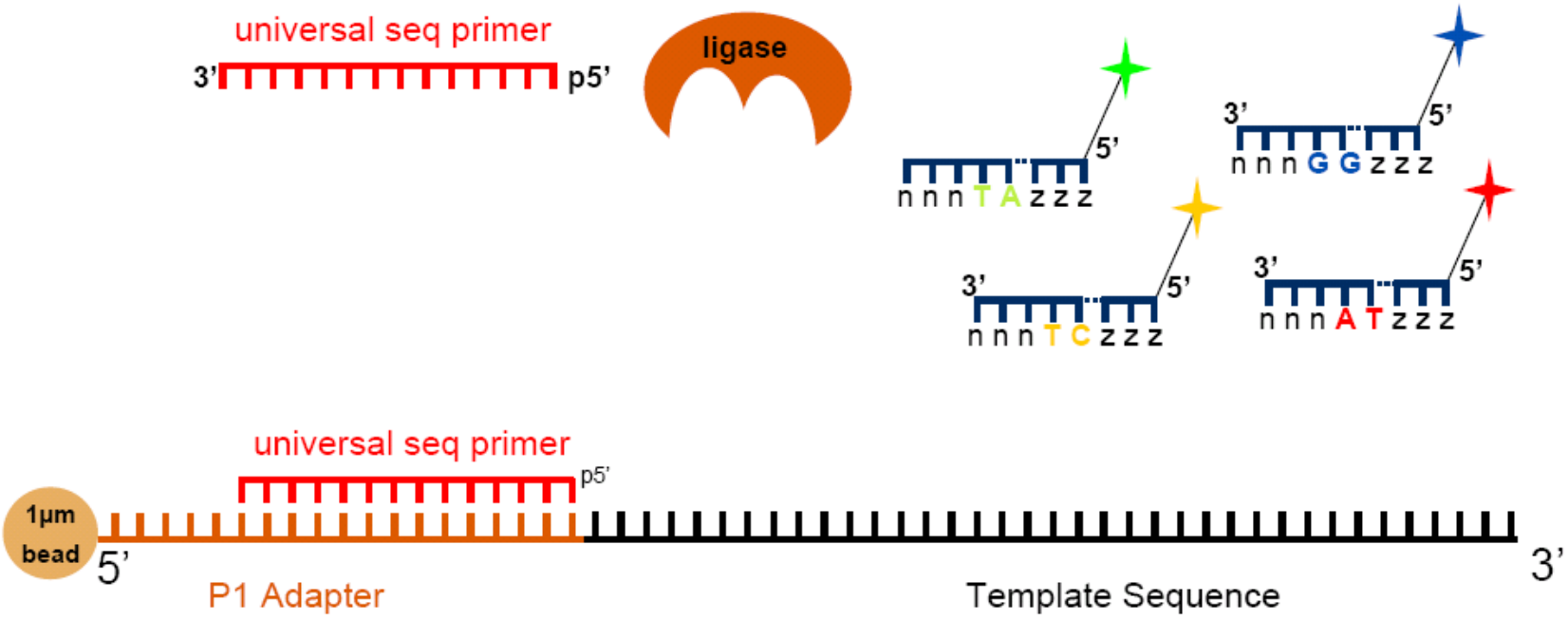
2-base encoding



16 dinucleotides / 4 dyes = 4 dinucleotides/dye

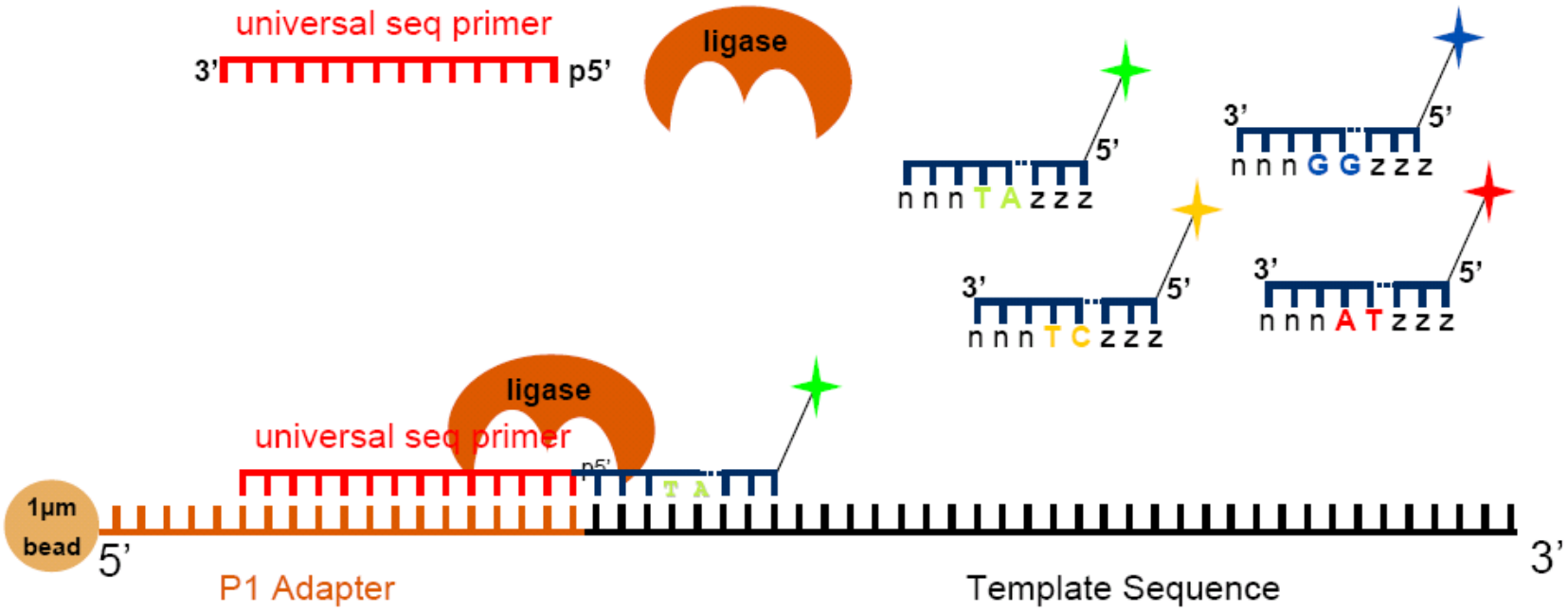
SOLID

4-color ligation



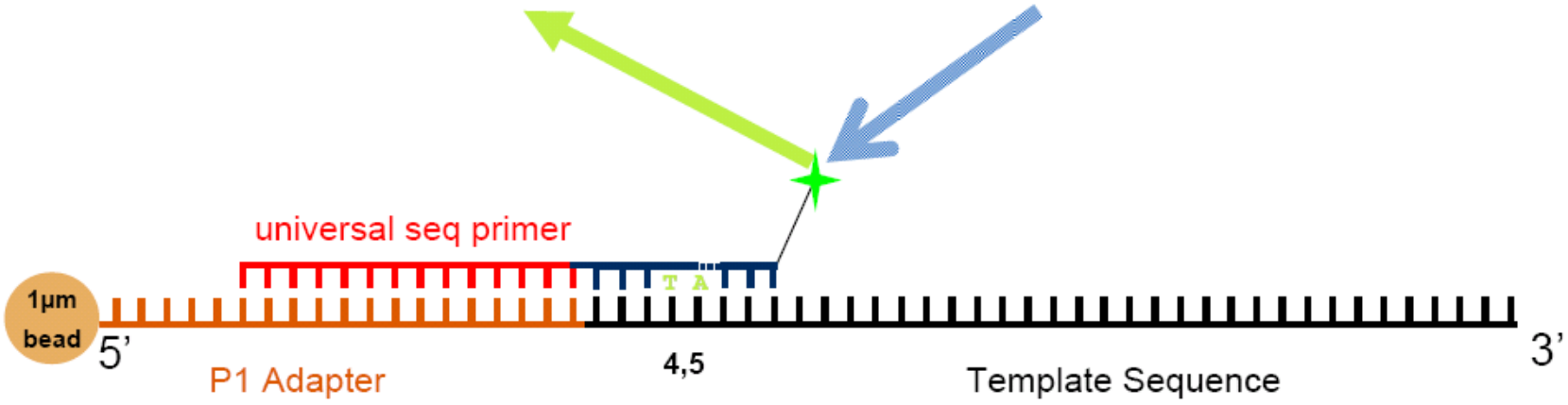
SOLID

4-color ligation



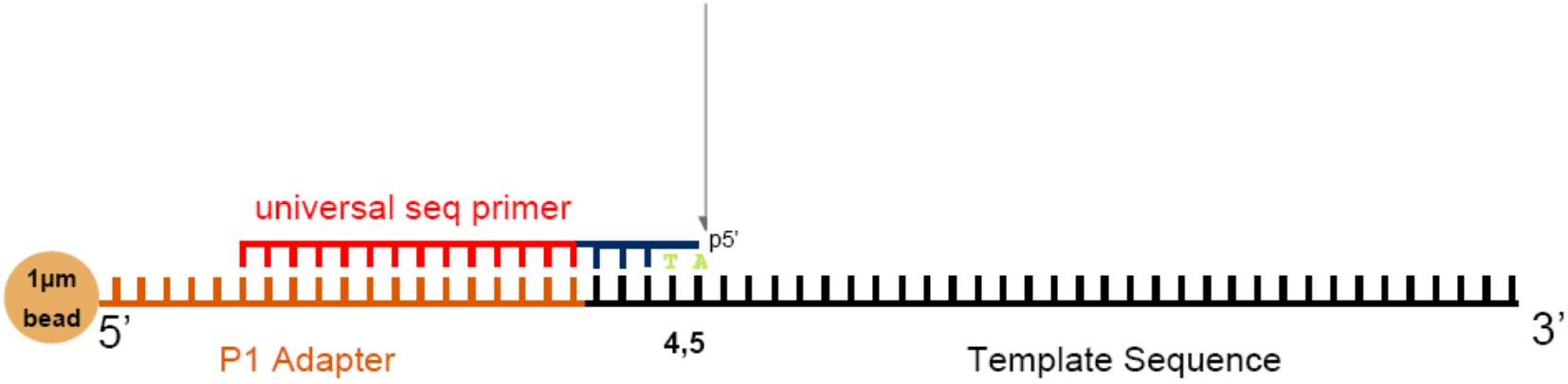
SOLID

Detection



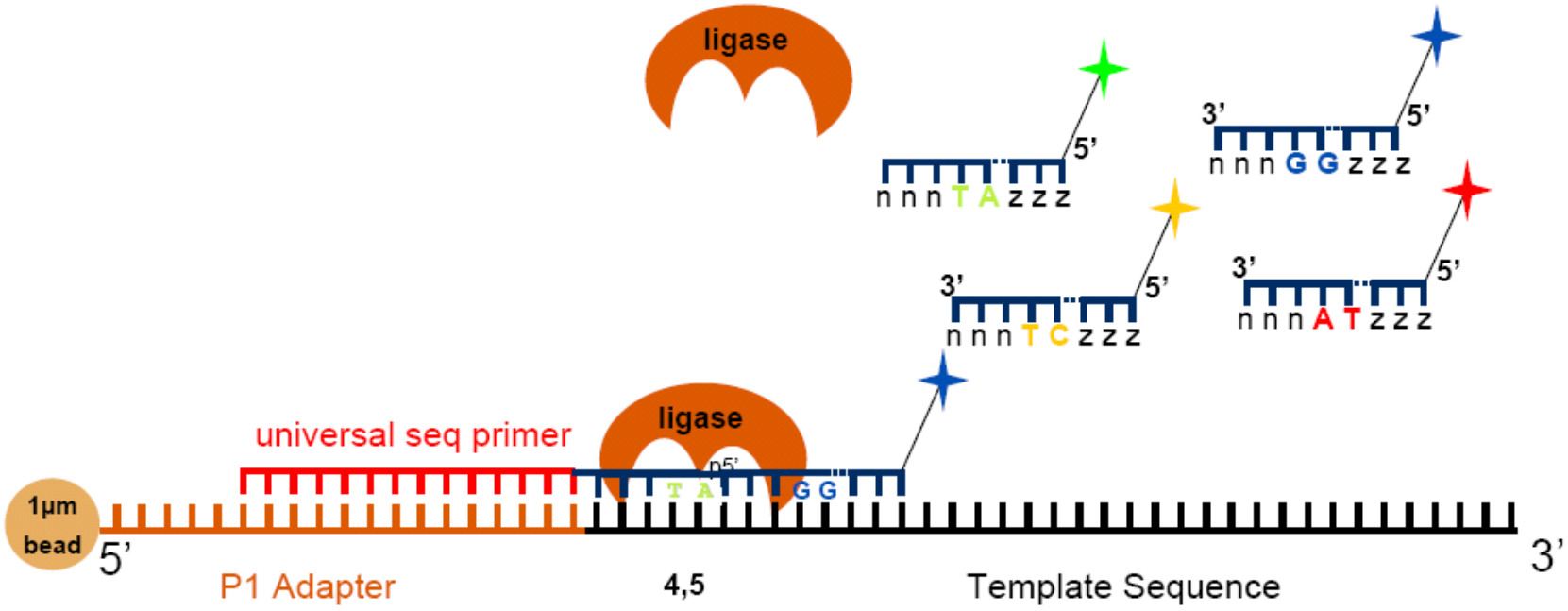
SOLID

Cleavage



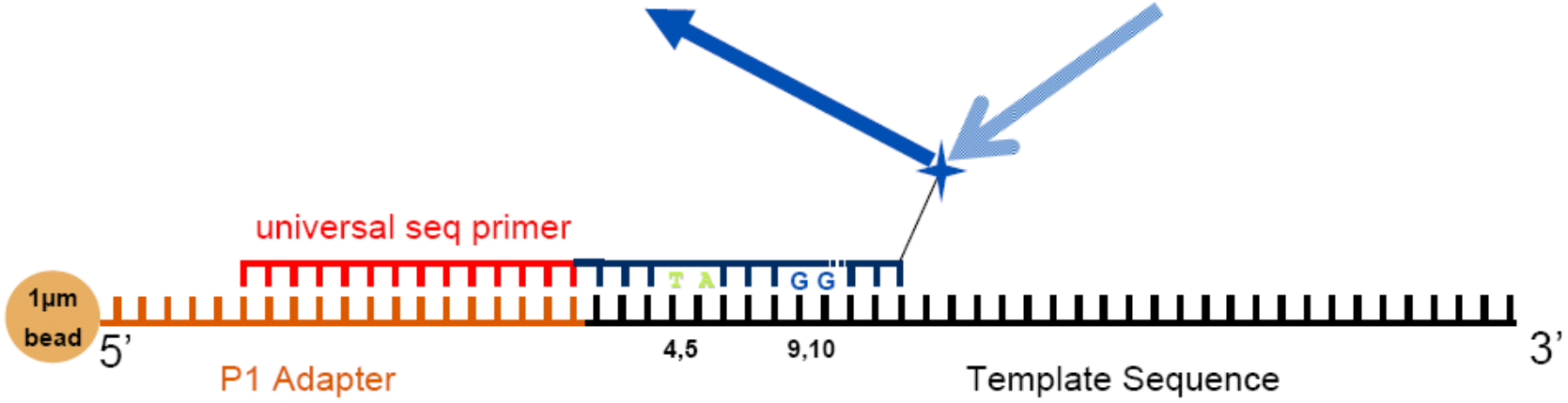
SOLID

2nd cycle ligation



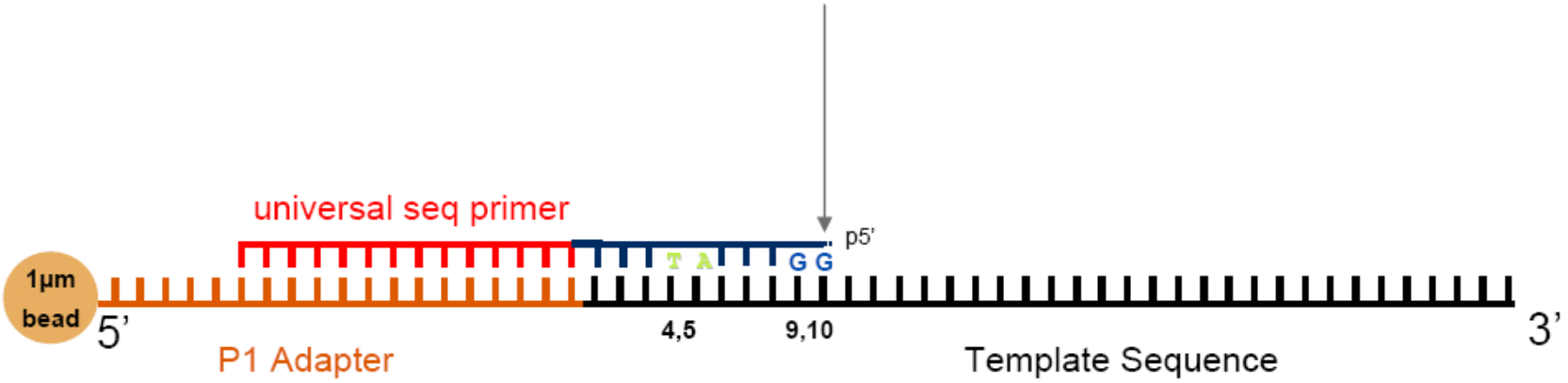
SOLID

2nd cycle detection



SOLID

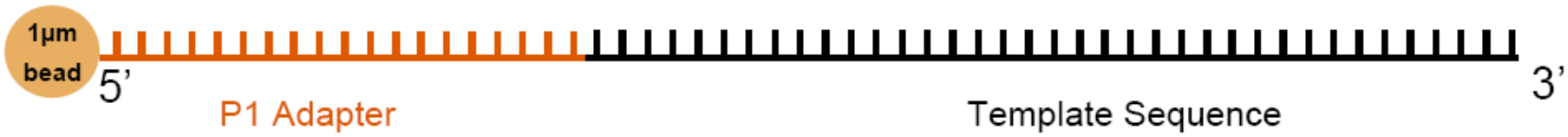
2nd cycle cleavage



+ additional 3 ligation cyles

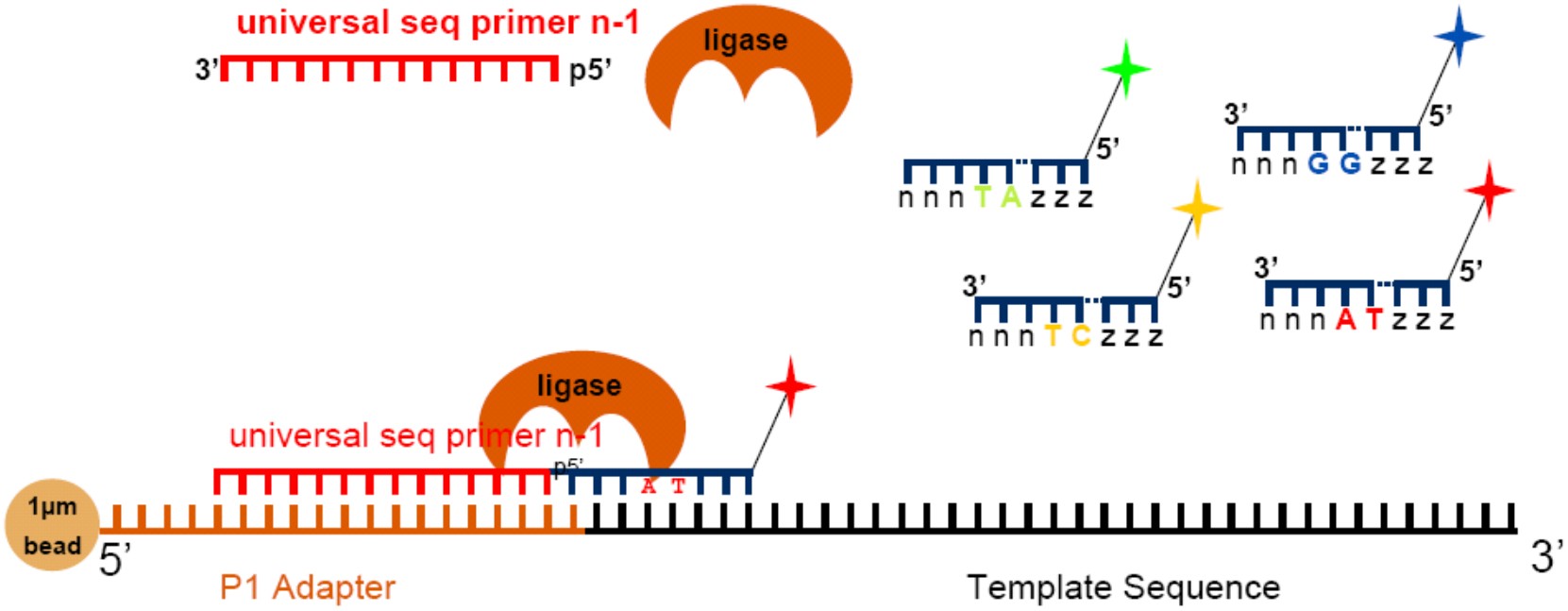
SOLID

Reset after 5th cycle



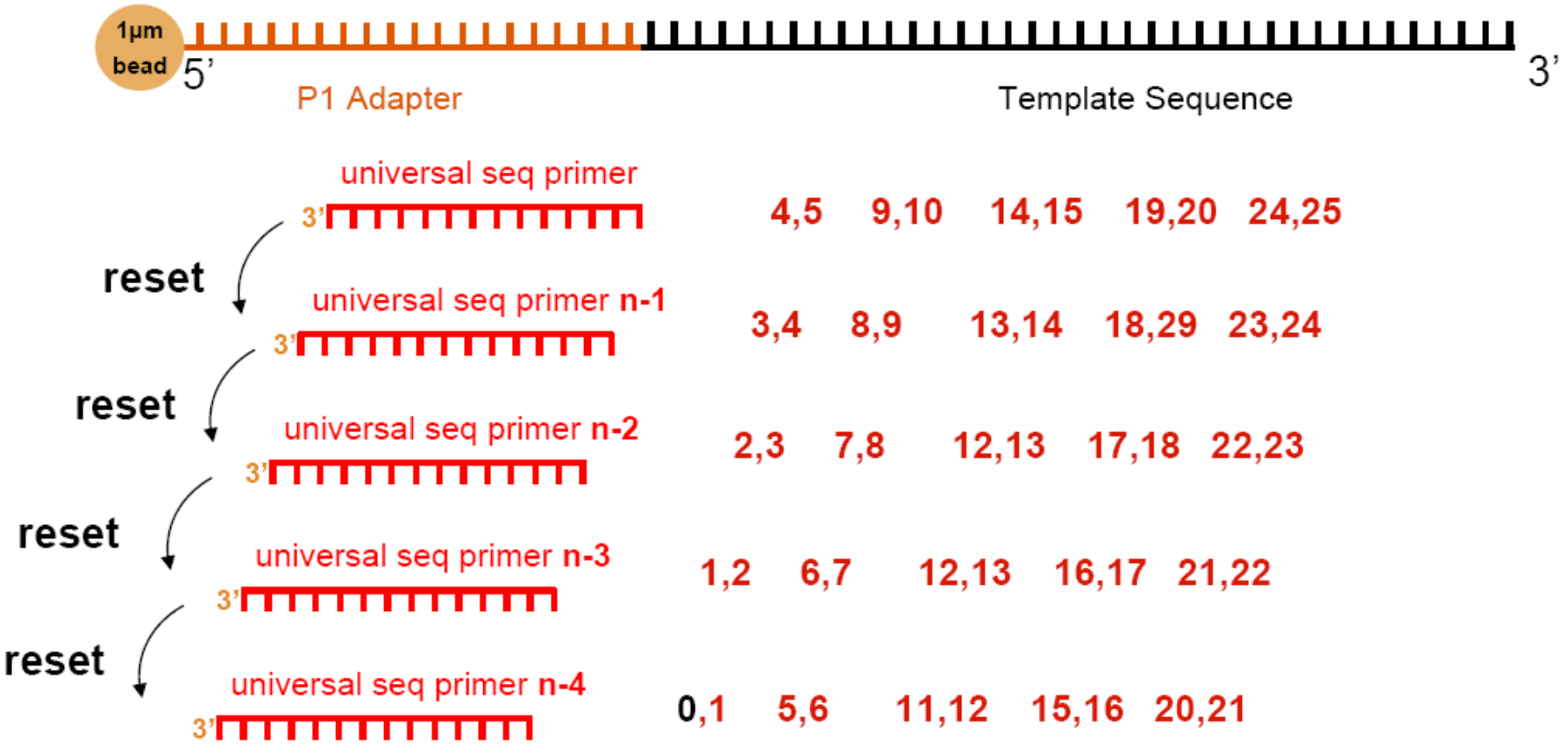
SOLID

6th cycle ligation



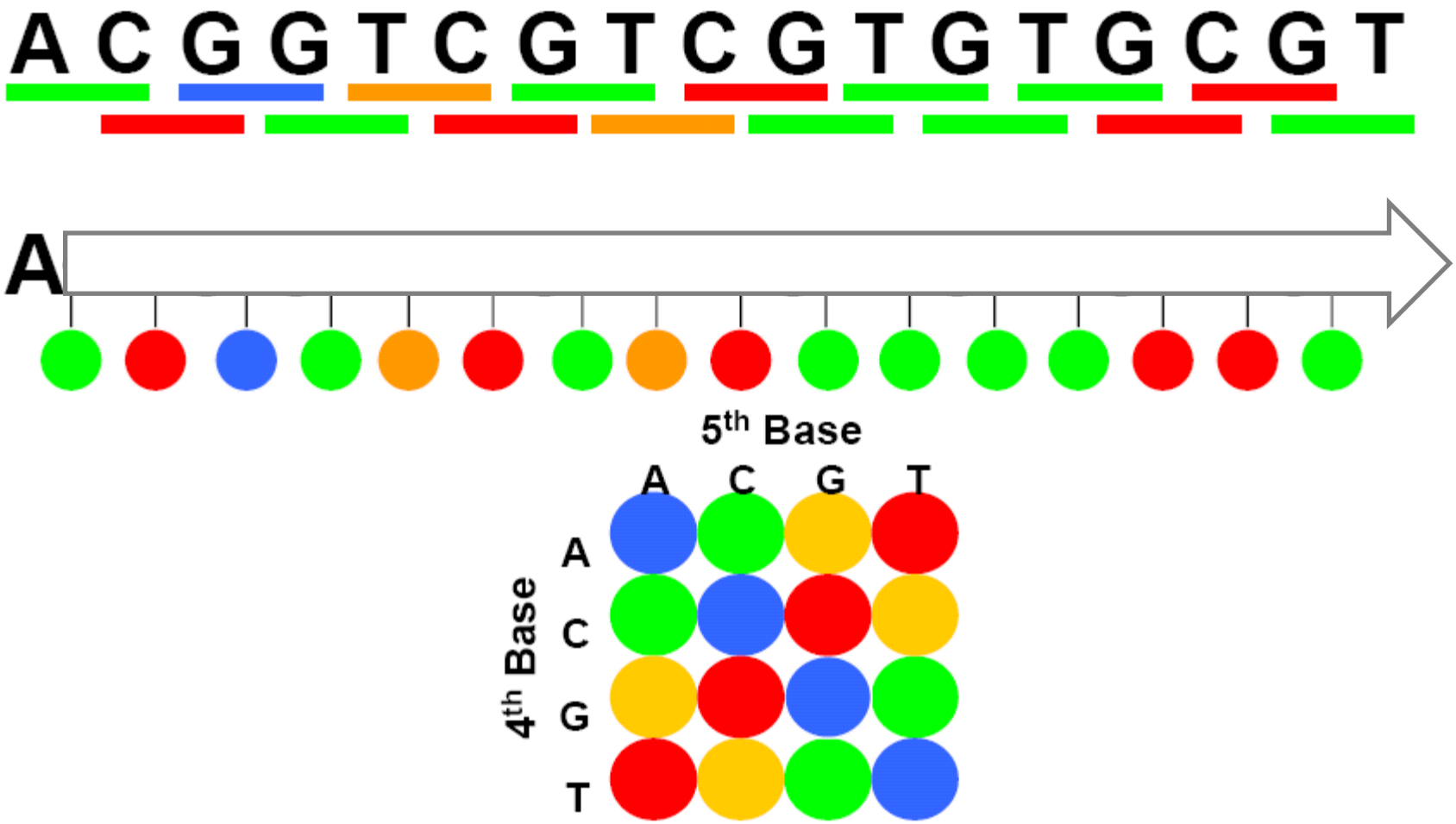
SOLID

Cycling scheme



SOLID

2-base encoding



Complete Genomics

Powering large-scale human genome studies

Corporate | Technology | Services | Data Release | Future Applications | Resources | Contact Us

Complete Genomics Wins 2009 Emerging Technology Award–Biotech

Silicon Valley/San Jose Business Journal

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.

Careers

News and Events

Our Mission

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*

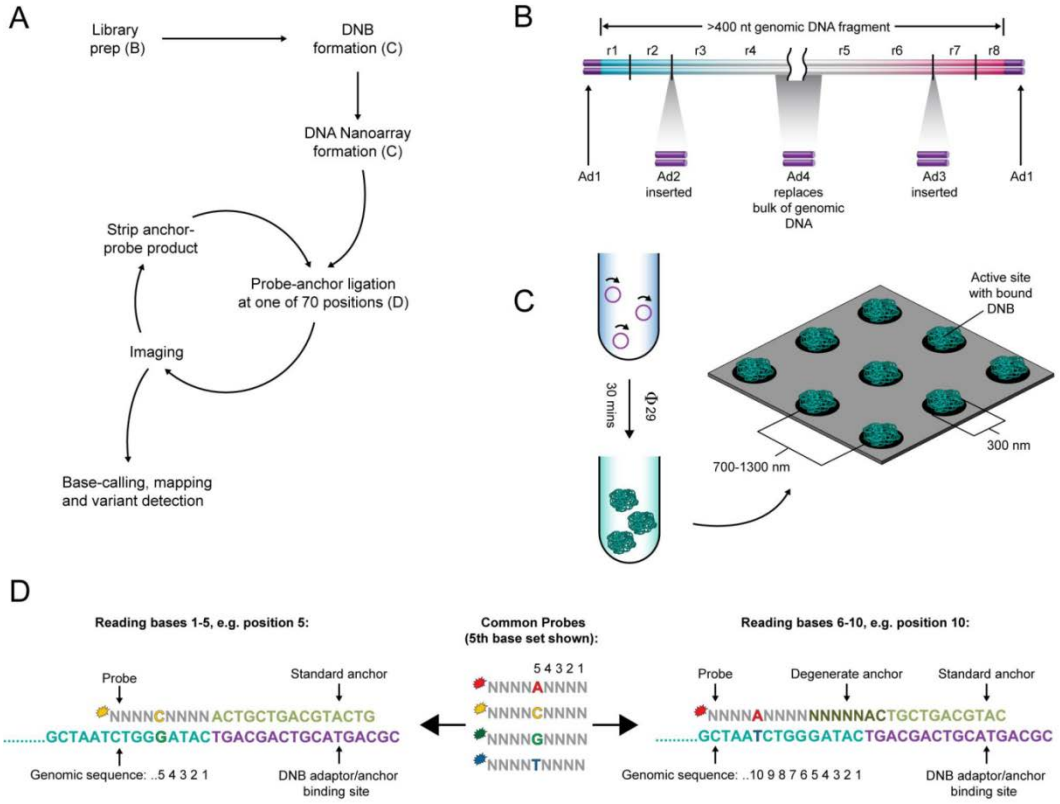
ISB and Complete Genomics to Conduct 100-Genome Study

more>>

www.completegenomics.com

Complete Genomics

In-house technology



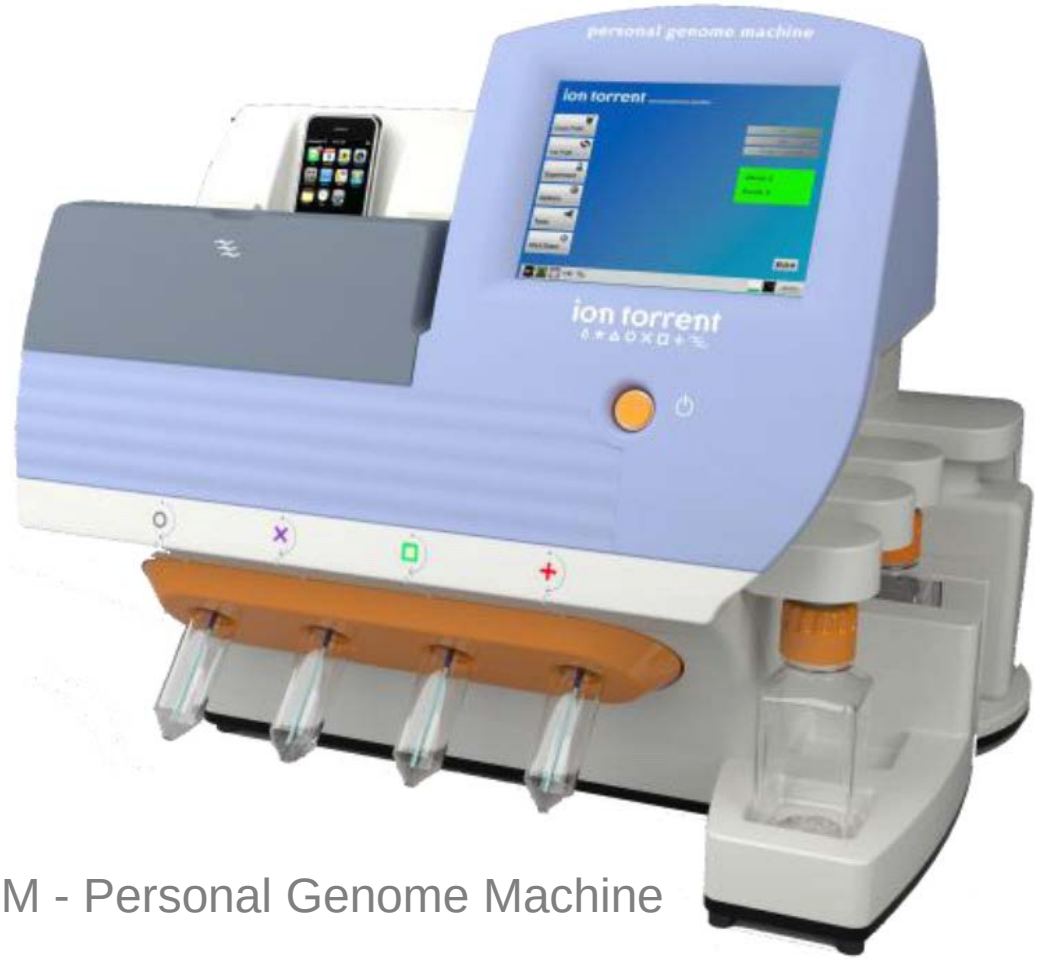
- self-assembling DNA nanoballs (DNB)
- patterned nanoarrays
- combinatorial probe anchor ligation (cPAL)

Drmanac et al. Science 327:78 (2009)

Complete Genomics - BGI

Revolocity

10,000 … 30.000 human genomes / year



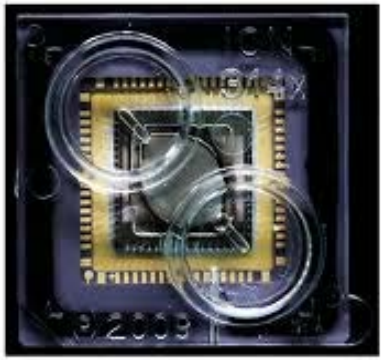
PGM - Personal Genome Machine

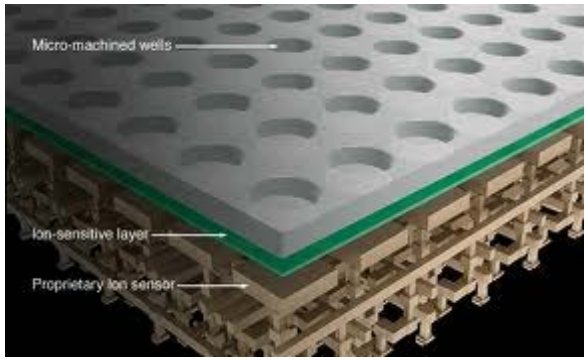
ion torrent
by life technologies

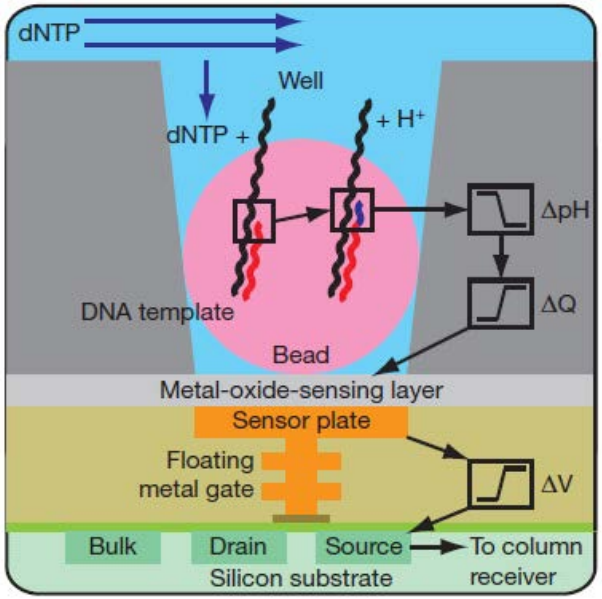
www.iontorrent.com

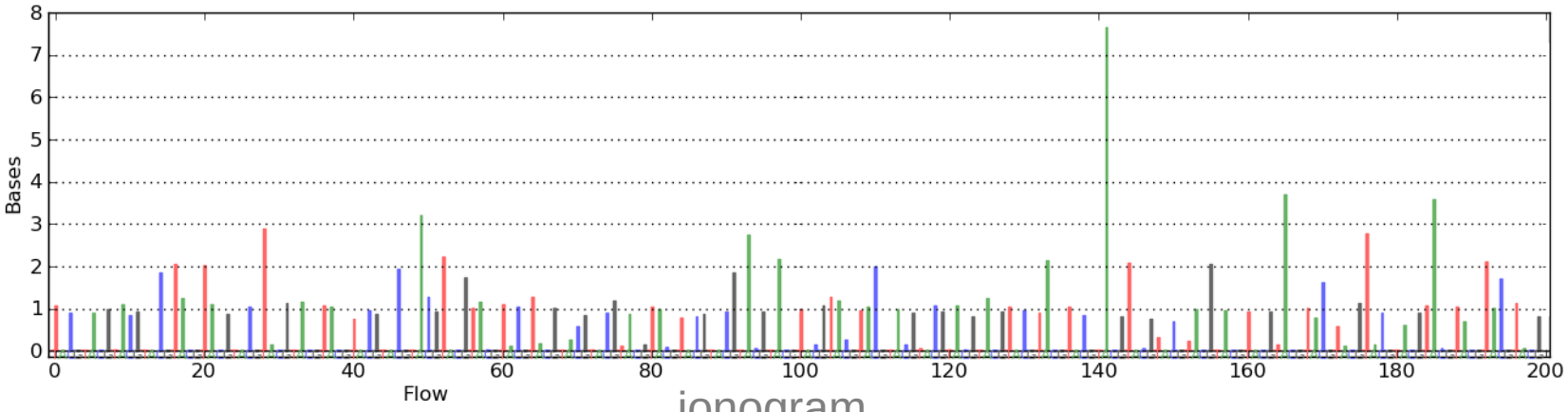
Ion Torrent

The chip is the machine







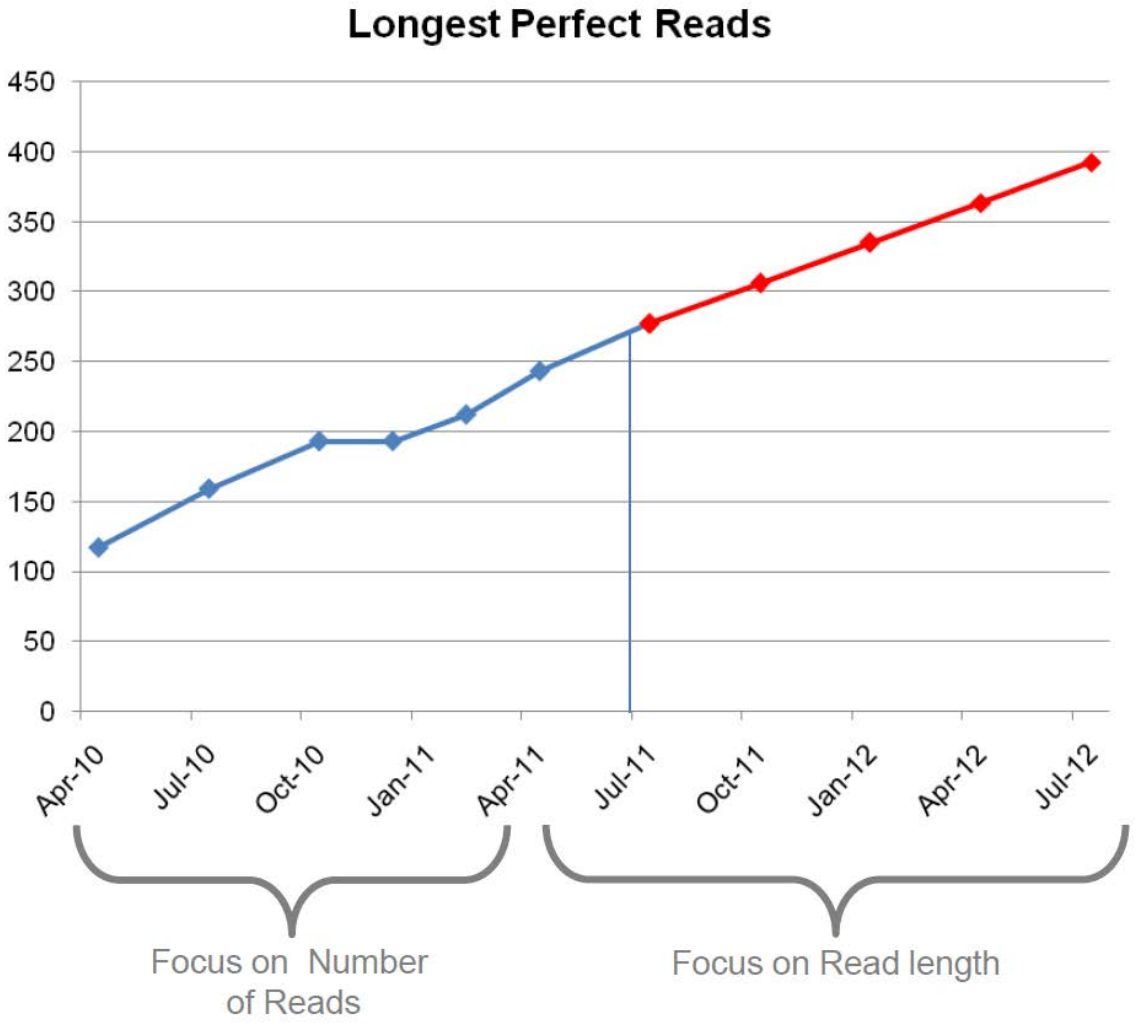


ionogram

www.iontorrent.com

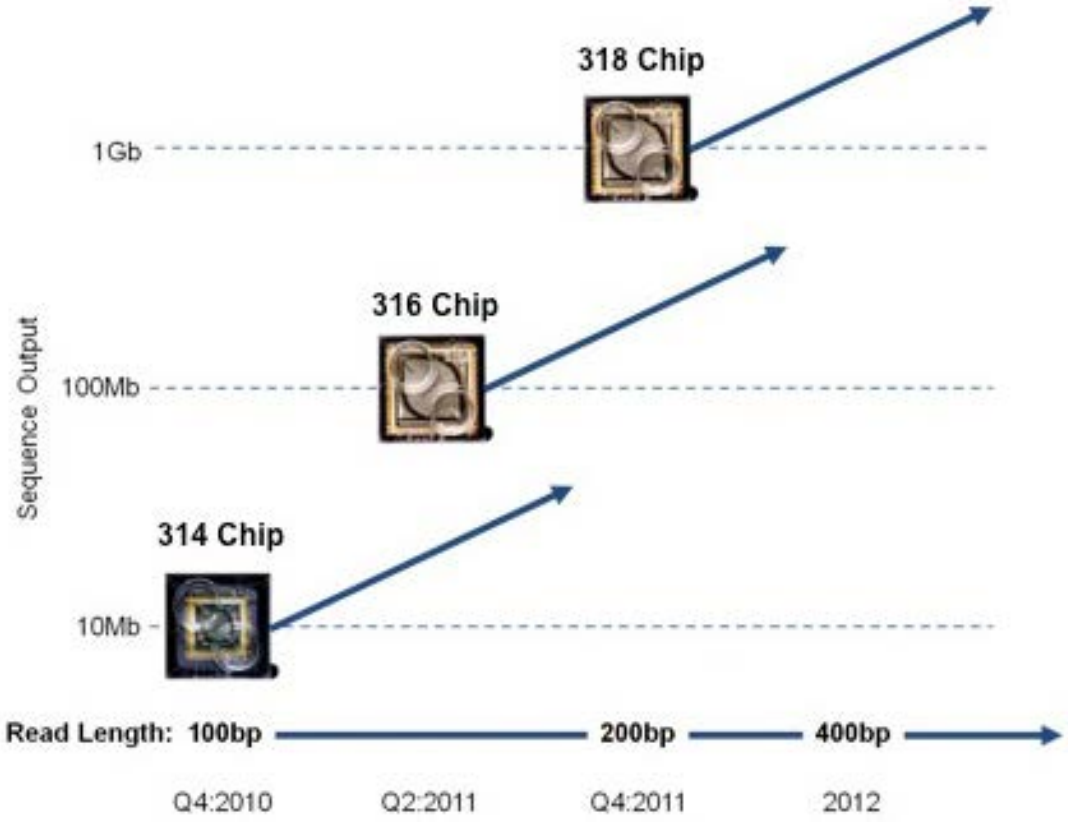
Ion Torrent

Read length



ABI/Ion Torrent

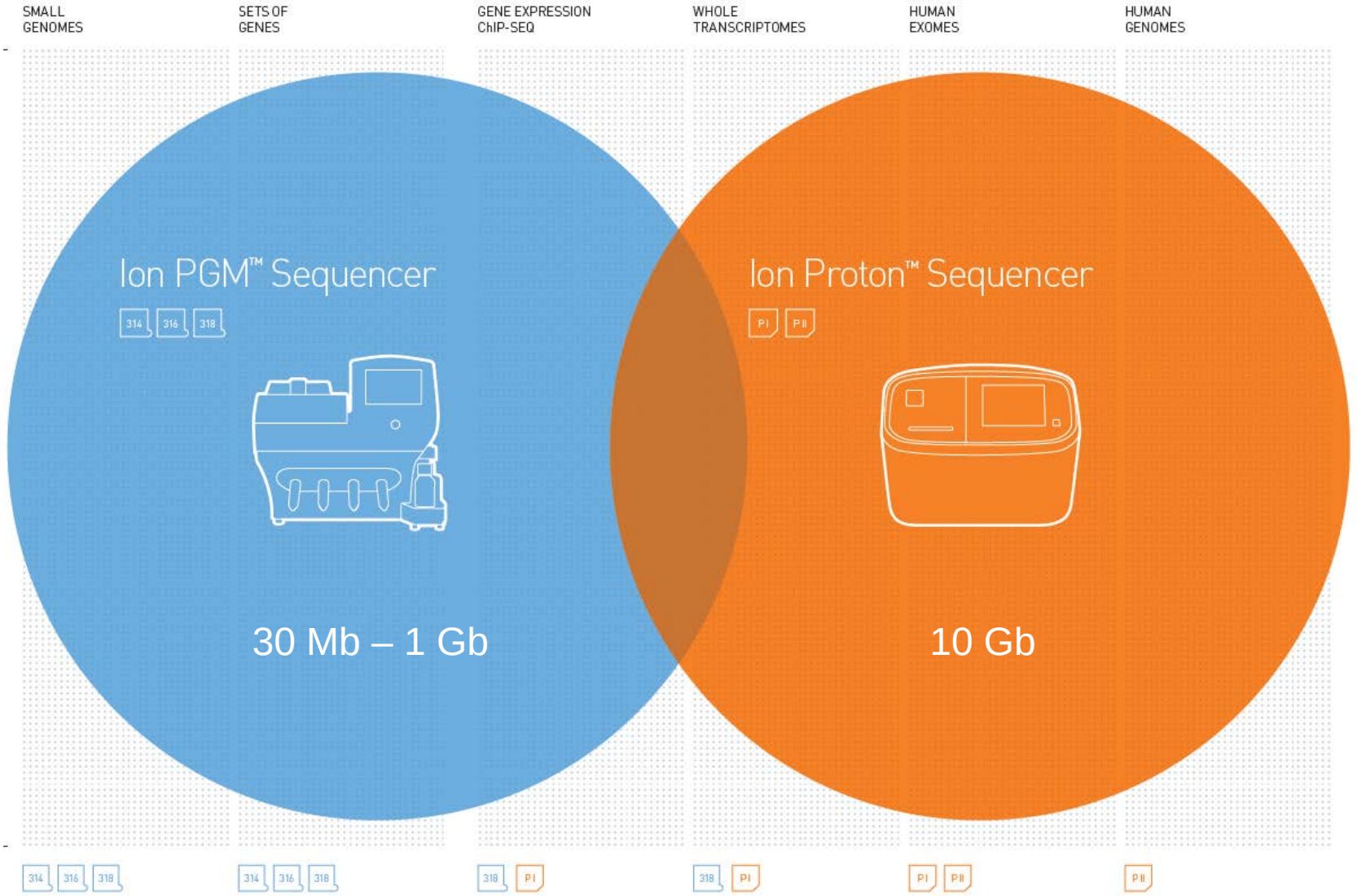
Data output



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

Ion Torrent

Data output



ABI/Ion Torrent

‘Post-light’ sequencing of Gordon Moore

An integrated semiconductor device enabling non-optical genome sequencing

Jonathan M. Rothberg[1], Wolfgang Hinz[1], Todd M. Rearick[1], Jonathan Schultz[1], William Mileski[1], Mel Davey[1], John H. Leamon[1], Kim Johnson[1], Mark J. Milgrew[1], Matthew Edwards[1], Jeremy Hoon[1], Jan F. Simons[1], David Marran[1], Jason W. Myers[1], John F. Davidson[1], Annika Branting[1], John R. Nobile[1], Bernard P. Puc[1], David Light[1], Travis A. Clark[1], Martin Huber[1], Jeffrey T. Branciforte[1], Isaac B. Stoner[1], Simon E. Cawley[1], Michael Lyons[1], Yutao Fu[1], Nils Homer[1], Marina Sedova[1], Xin Miao[1], Brian Reed[1], Jeffrey Sabina[1], Erika Feierstein[1], Michelle Schorn[1], Mohammad Alanjary[1], Eileen Dimalanta[1], Devin Dressman[1], Rachel Kasinskas[1], Tanya Sokolsky[1], Jacqueline A. Fidanza[1], Eugeni Namsaraev[1], Kevin J. McKernan[1], Alan Williams[1], G. Thomas Roth[1] & James Bustillo[1]

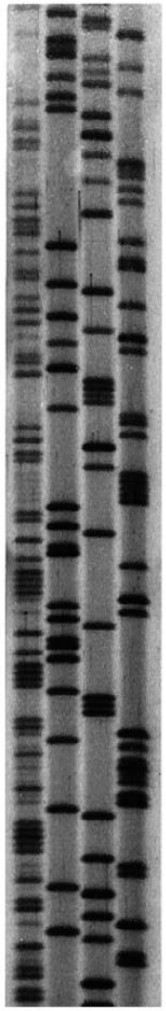
The seminal importance of DNA sequencing to the life sciences, biotechnology and medicine has driven the search for more scalable and lower-cost solutions. Here we describe a DNA sequencing technology in which scalable, low-cost semiconductor manufacturing techniques are used to make an integrated circuit 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.

Nature 475:348 (2011)

2nd Generation sequencing

Applications

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



3rd Generation

Pacific Biosciences
Oxford Nanopore

PACIFIC BIOSCIENCES℠

www.pacificbiosciences.com

Pacific Biosciences

Fluorescent lable

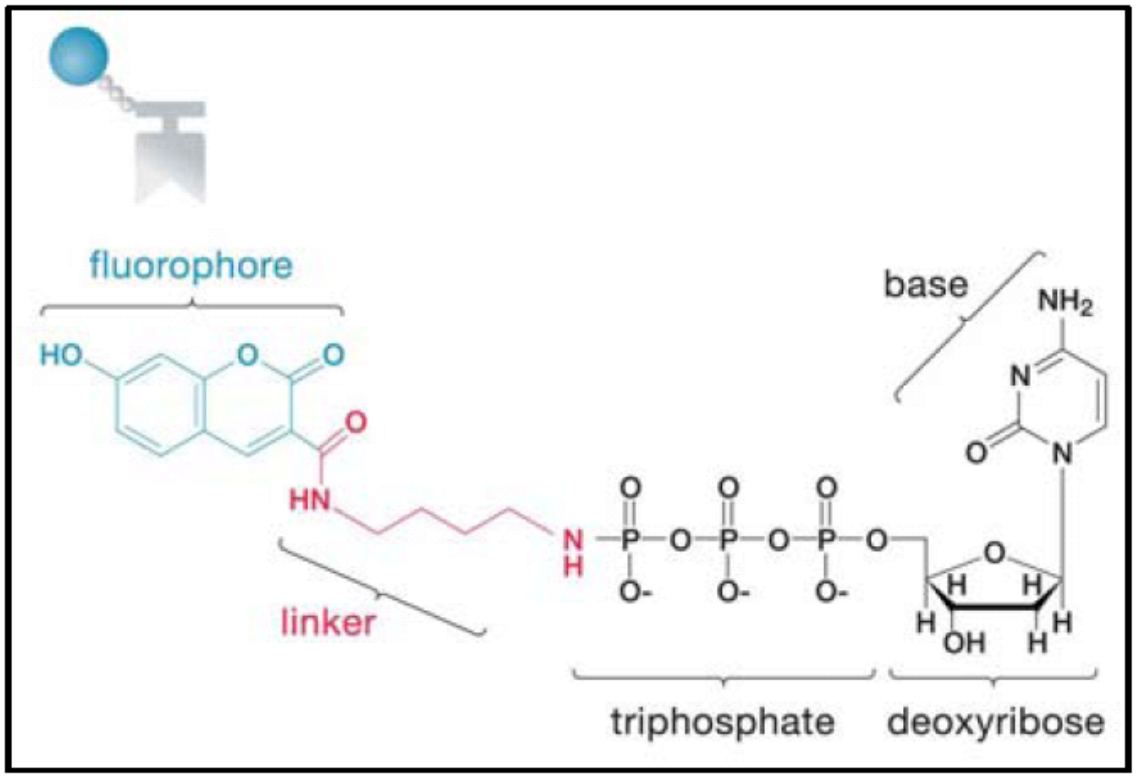
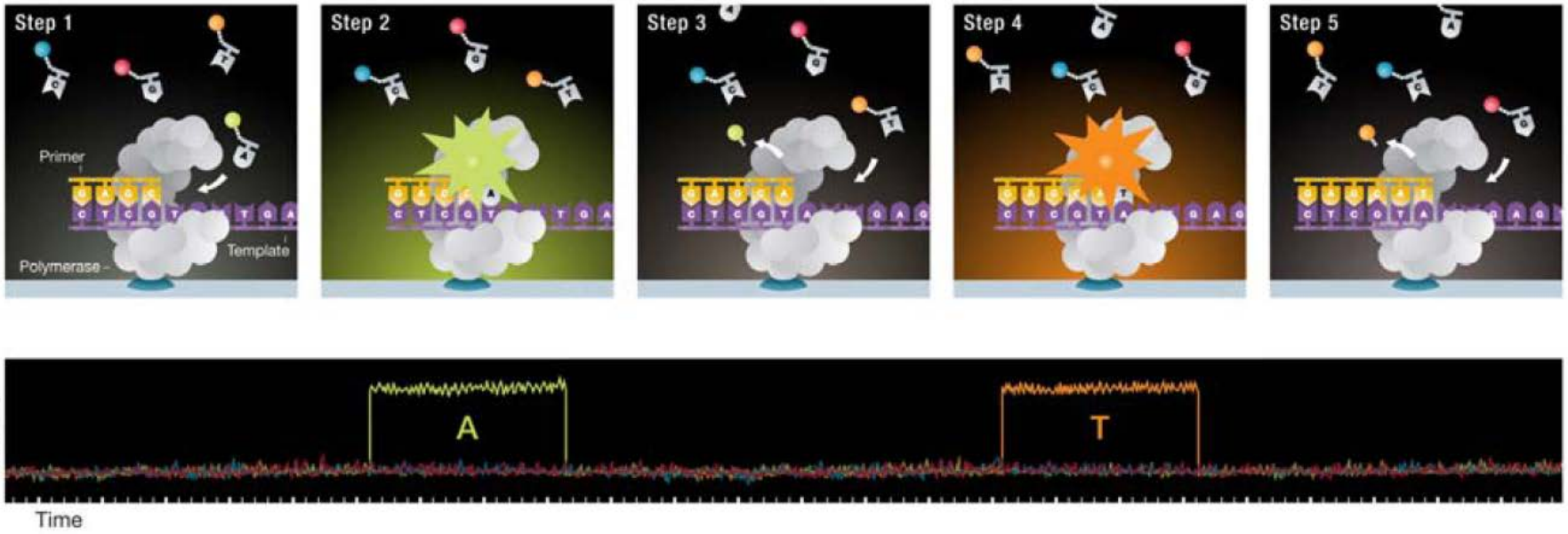


Figure 9. Phospholinked nucleotides

Pacific Biosciences

Immobilised polymerase

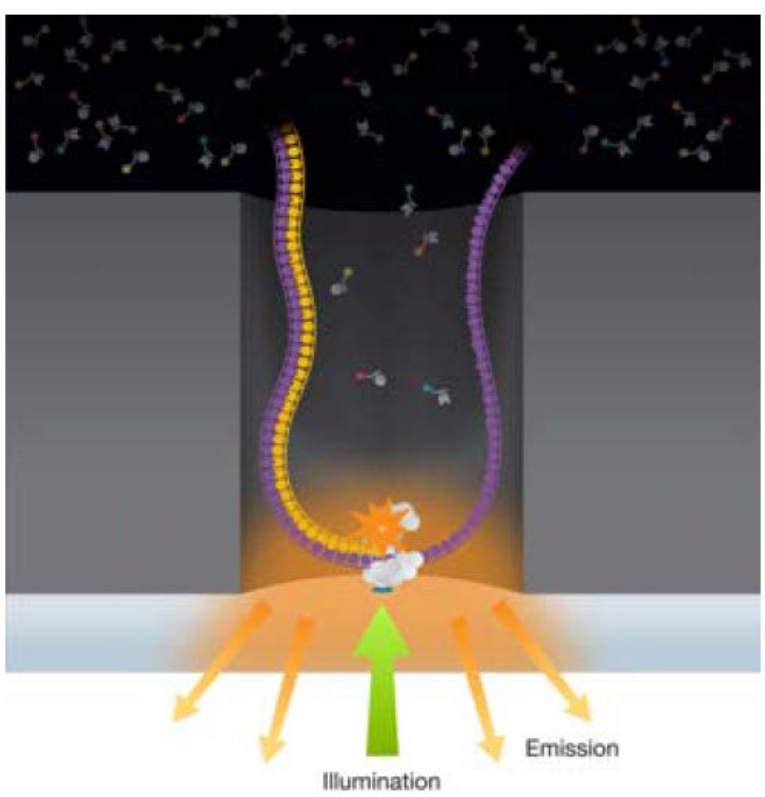


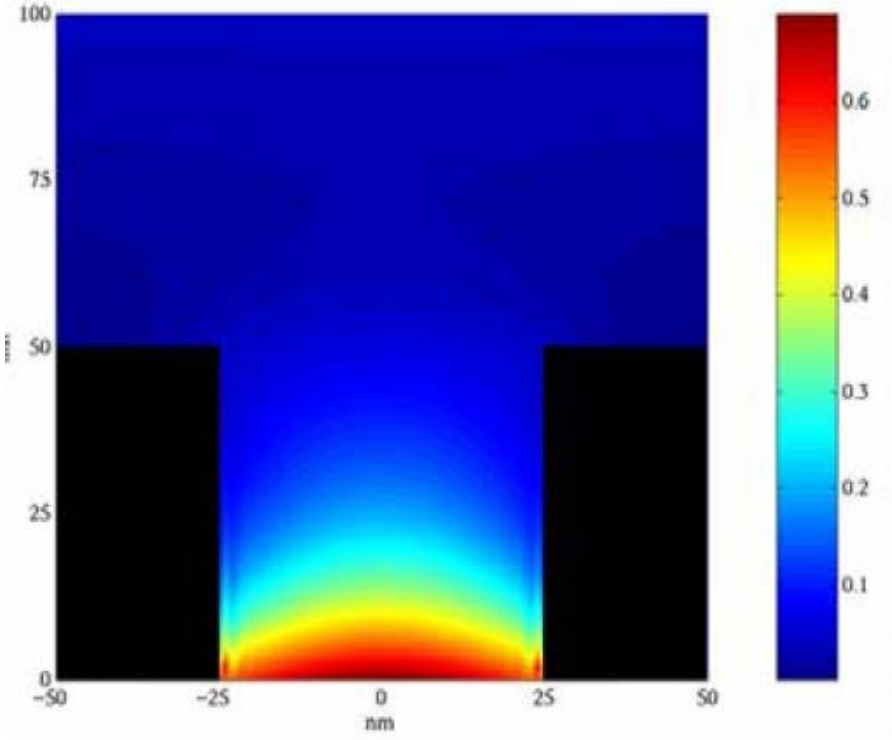
long sequences

www.pacificbiosciences.com

Pacific Biosciences

Zero-mode waveguide chip



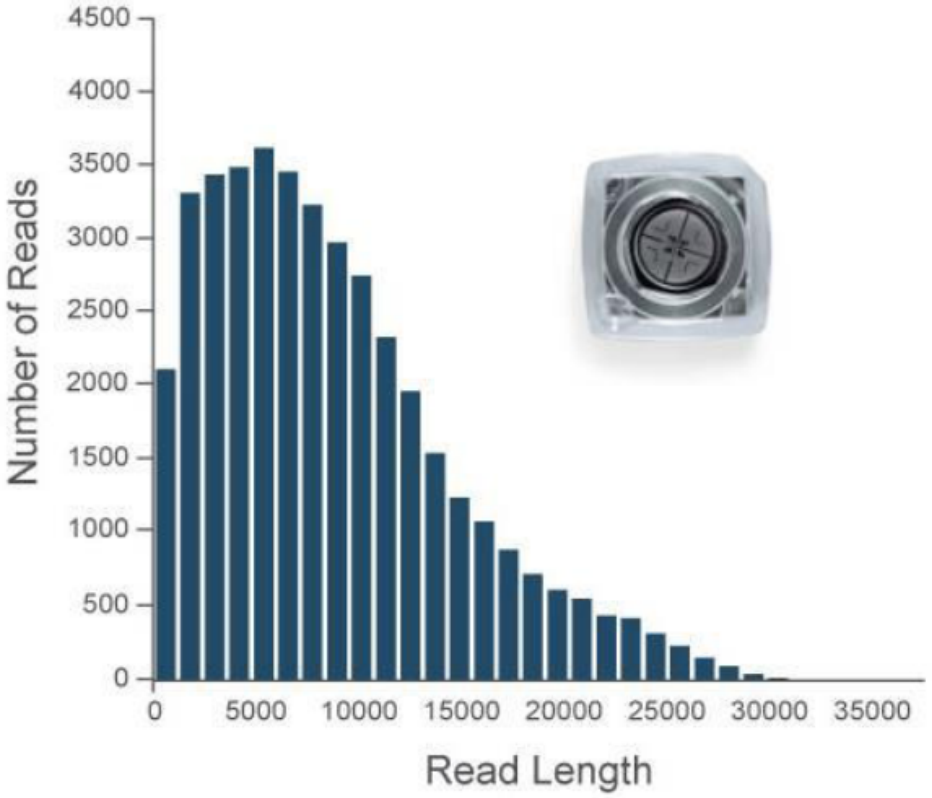


www.pacificbiosciences.com

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

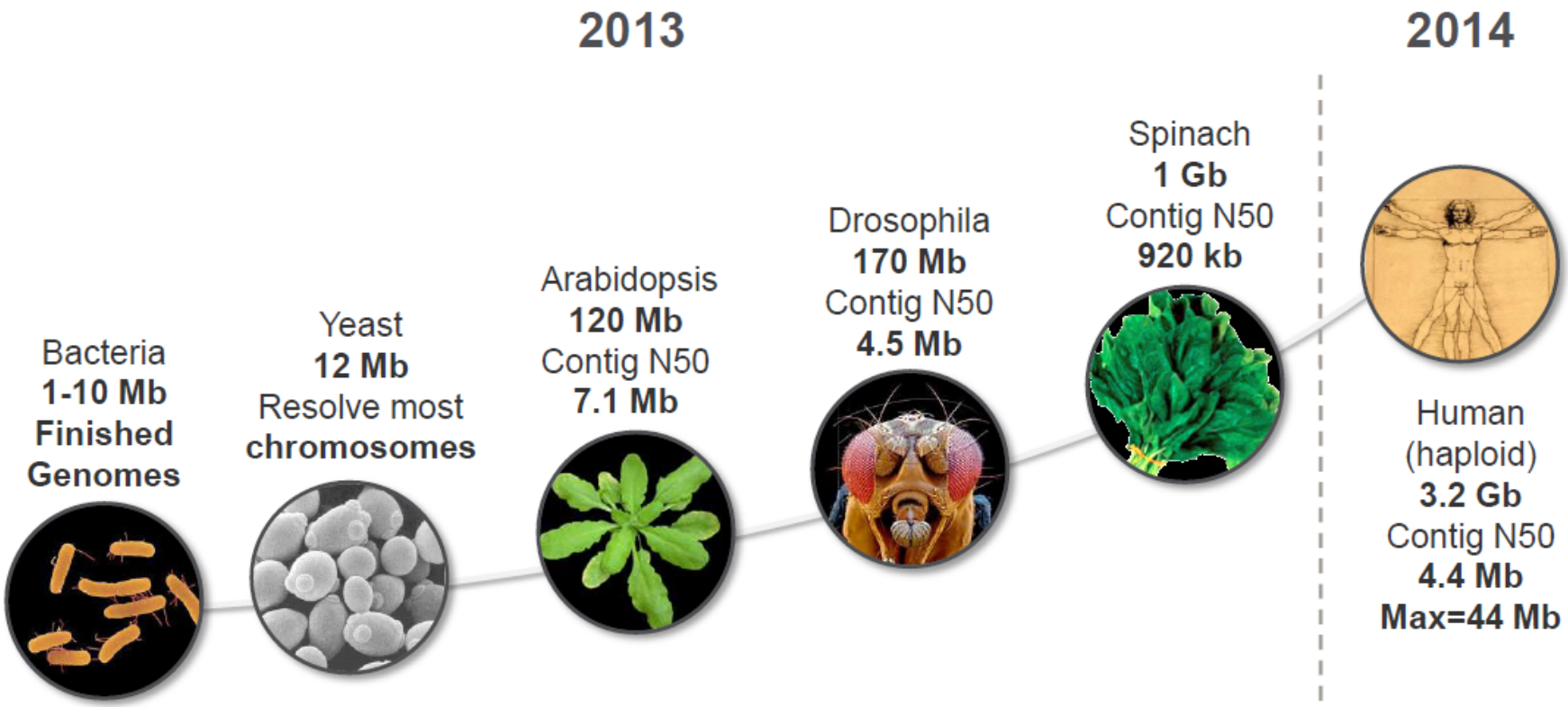
Based on data from a 20 kb size-selected E. coli library using a 180-minute movie.

Each SMRT Cell yields ~ 50,000 reads.

PACIFIC BIOSCIENCES®

Pacific Biosciences

Genome *de novo* assemblies



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

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

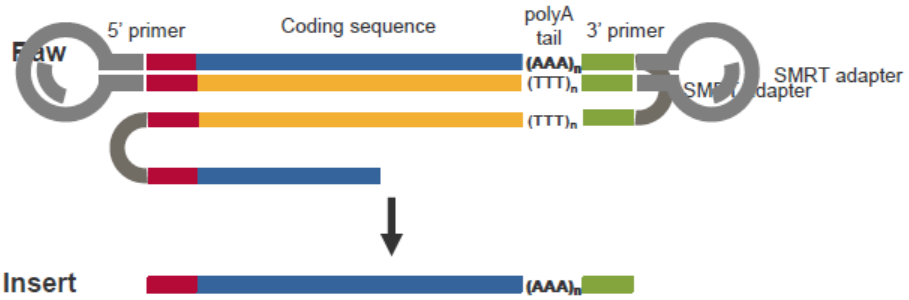
PACIFIC BIOSCIENCES

Pacific Biosciences

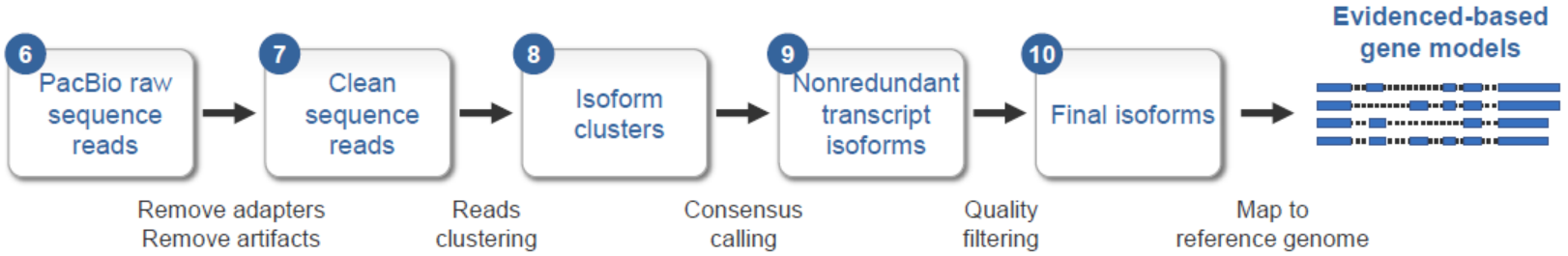
Full length transcripts

Experimental Pipeline

SampleNet: Iso-Seq Method with Clonetech cDNA Synthesis Kit



Informatics Pipeline

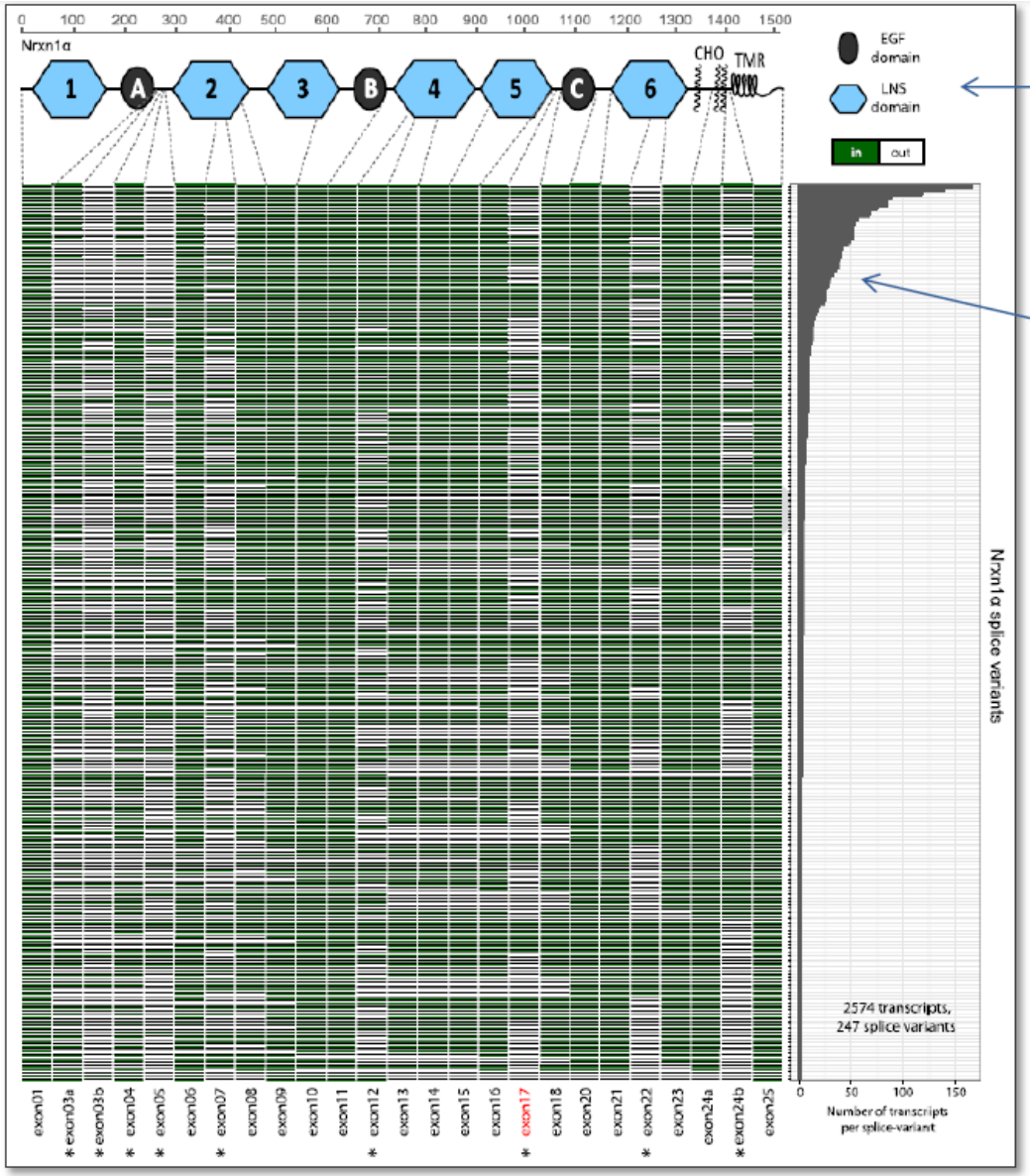


DevNet: Iso-Seq wiki page

PACIFIC BIOSCIENCES®

Pacific Biosciences

Alternative neurexin 1α transcripts



Nrxn1α domain structure

Splice isoform abundance

- 2,574 full-length Nrxn1α mRNAs sequence reads
- Rows: 247 unique alternatively spliced isoforms
- Columns: Exons
  - green – present
  - white – absent

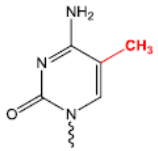
Treutlein *et al.* (2014) *PNAS.* doi:10.1073/pnas.1403244111

PACIFIC BIOSCIENCES

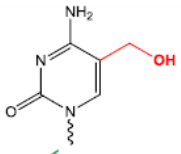
Pacific Biosciences

Epigenetics

Epigenetic markers:

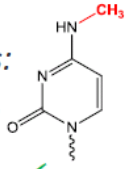


✓ 5-mC

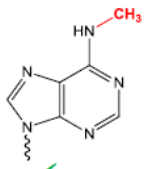


✓ 5-hmC

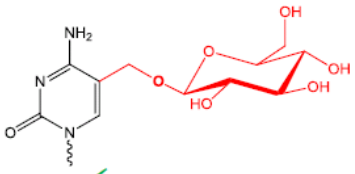
Identity markers: (host-pathogen interactions)



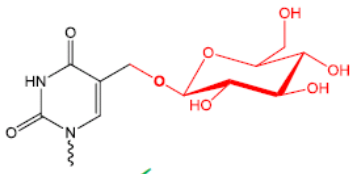
✓ 4-mC



✓ 6-mA

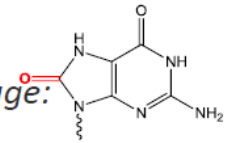


✓ Glucosyl-5-hmC

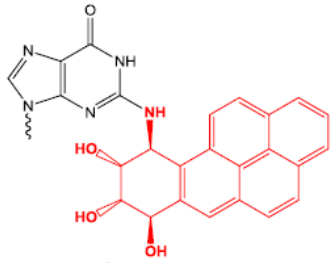


✓ Base J

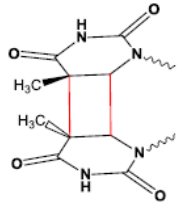
DNA damage:



✓ 8-oxoG

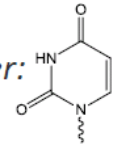


✓ N^2-BPDE-G



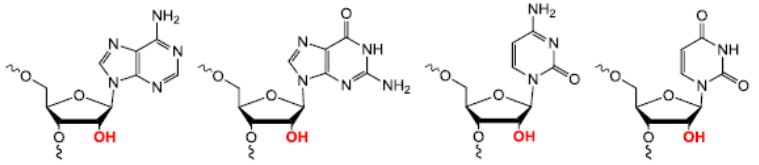
✓ Thymidine dimer

Other:



✓ dU

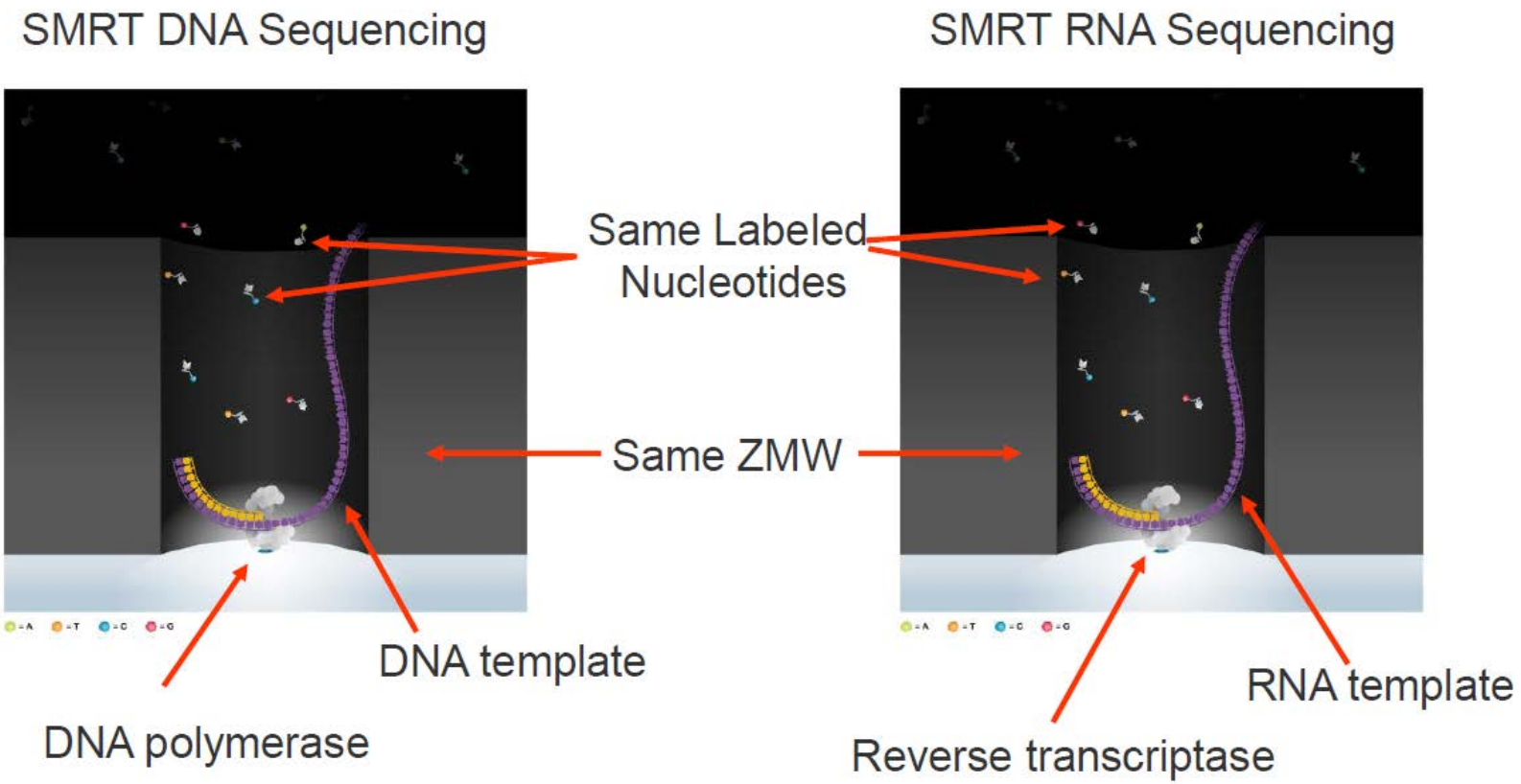
✓ 5Br-dU



✓ Ribonucleosides

Pacific Biosciences

Direct RNA sequencing



www.pacificbiosciences.com

Pacific Biosciences

Translation

nature Vol 464|15 April 2010|doi:10.1038/nature08925

ARTICLES

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

Sotaro Uemura[1,2], Colin Echeverría Aitken[1,3]†, Jonas Korlach[4], Benjamin A. Flusberg[4], Stephen W. Turner[4] & Joseph D. Puglisi[1,3]

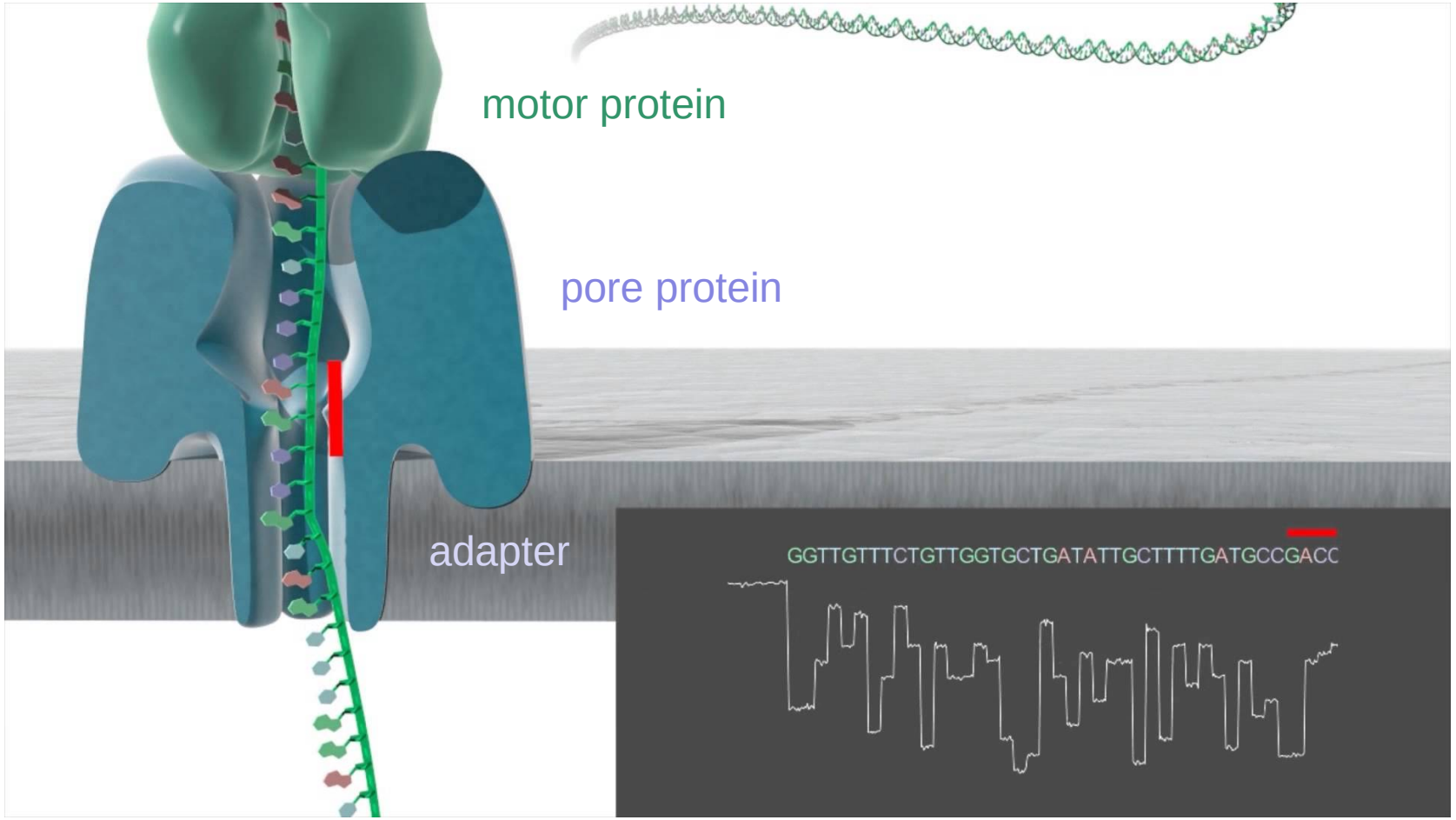
Translation by the ribosome occurs by a complex mechanism involving the coordinated interaction of multiple nucleic acid and protein ligands. Here we use zero-mode waveguides (ZMWs) and sophisticated detection instrumentation to allow real-time observation of translation at physiologically relevant micromolar 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 Technologies™

nanoporetech.com

Oxford Nanopore

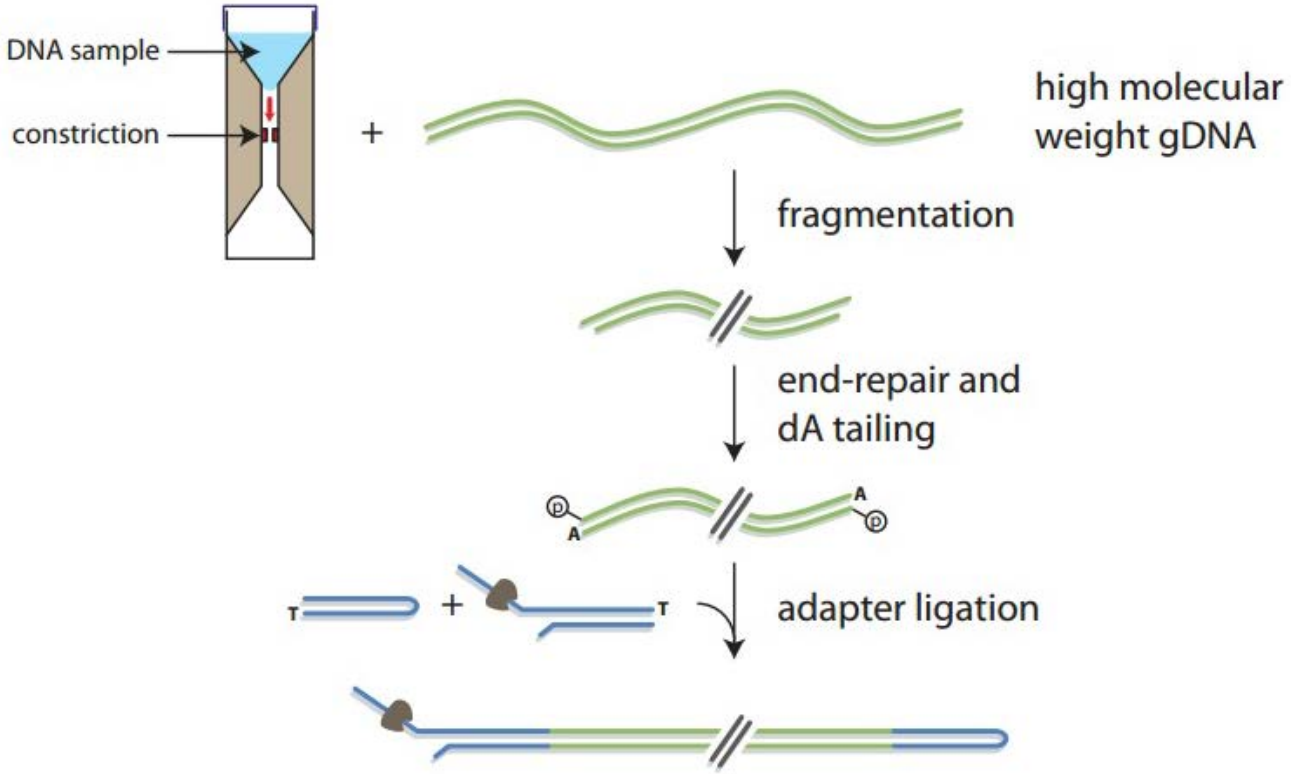
Principle



nanoporetech.com

Oxford Nanopore

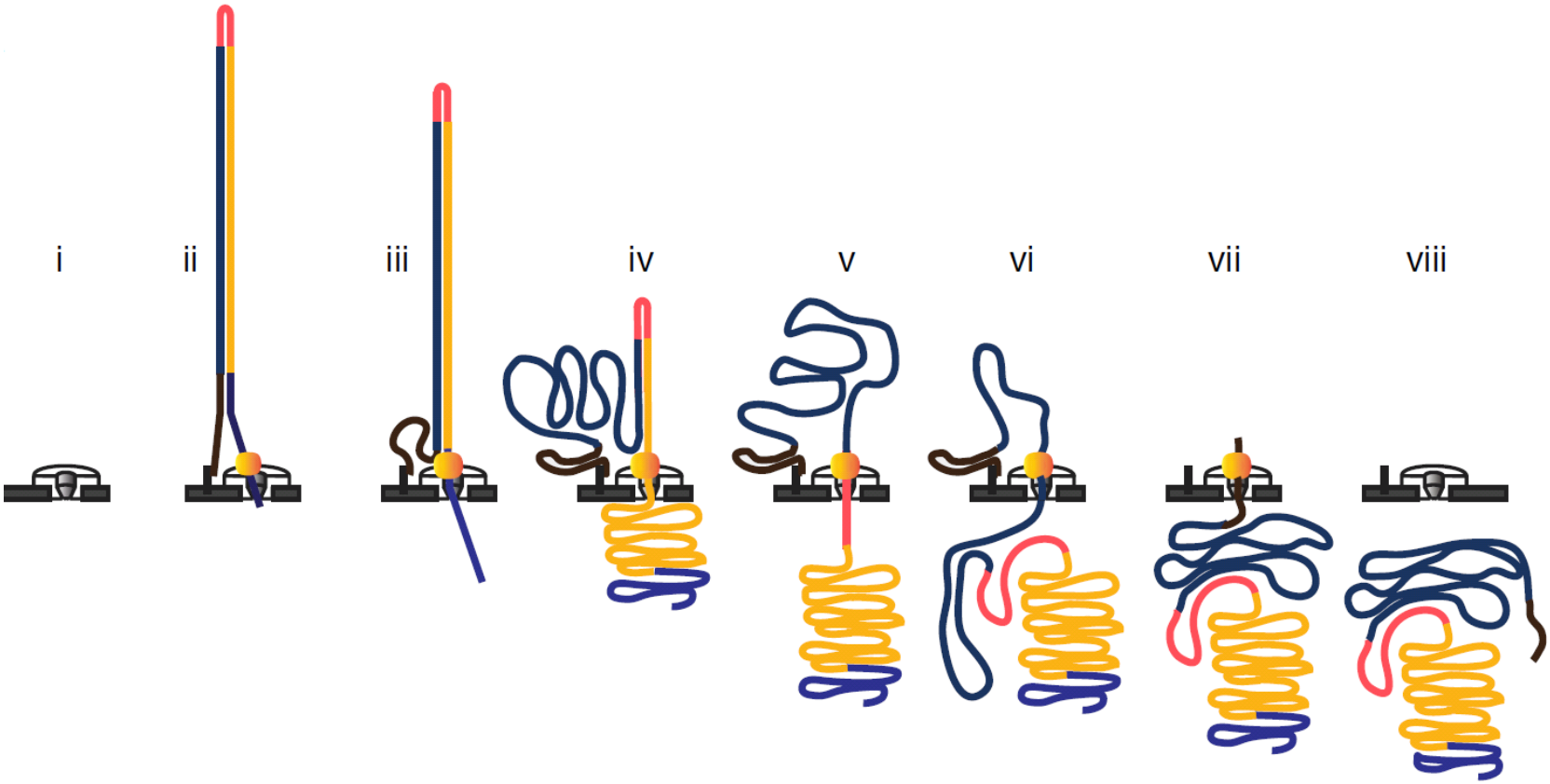
Library



nanoporetech.com

Oxford Nanopore

Sequencing

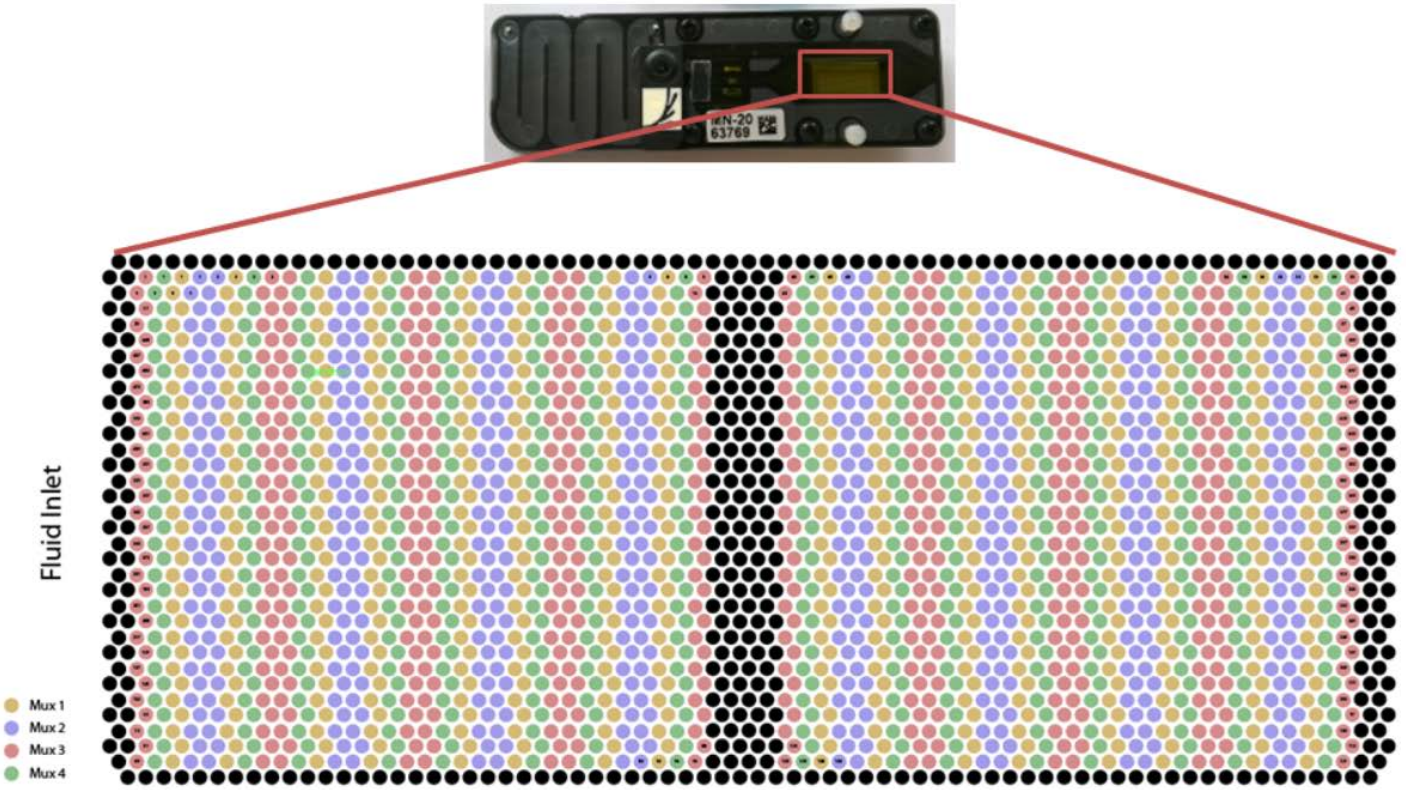


error rate: 10-30%

Jain et al. Nat Methods 12:351 (2015)

Oxford Nanopore

Data output



512 sequencing channels → each channel sequences multiple molecules

~ 400 Mb

nanoporetech.com

Oxford Nanopore

Data output

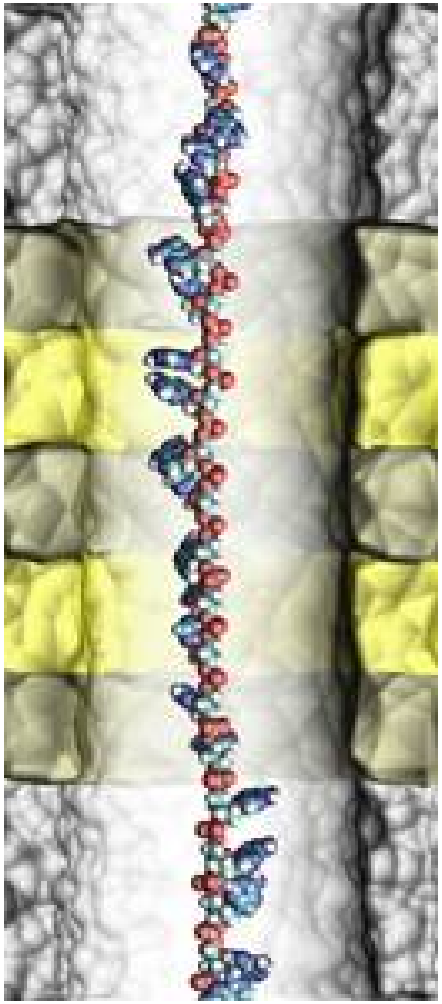
PROMETHION

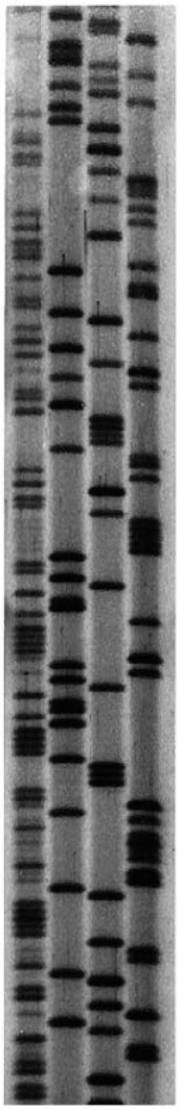
- 300X greater throughput than MinION
- 2016

nanoporetech.com

IBM

Waver pore





genome.leibniz-fli.de
Lectures