Next Generation Sequencing Technologies

Michael R. Crowley
Department of Genetics
Heflin Center for Genomic Sciences

Different Platforms rely on Different Technology

HiSeq, MiSeq or GAIIx by Illumina

SOLiD by Applied Biosystems

454 Lifesciences by Roche

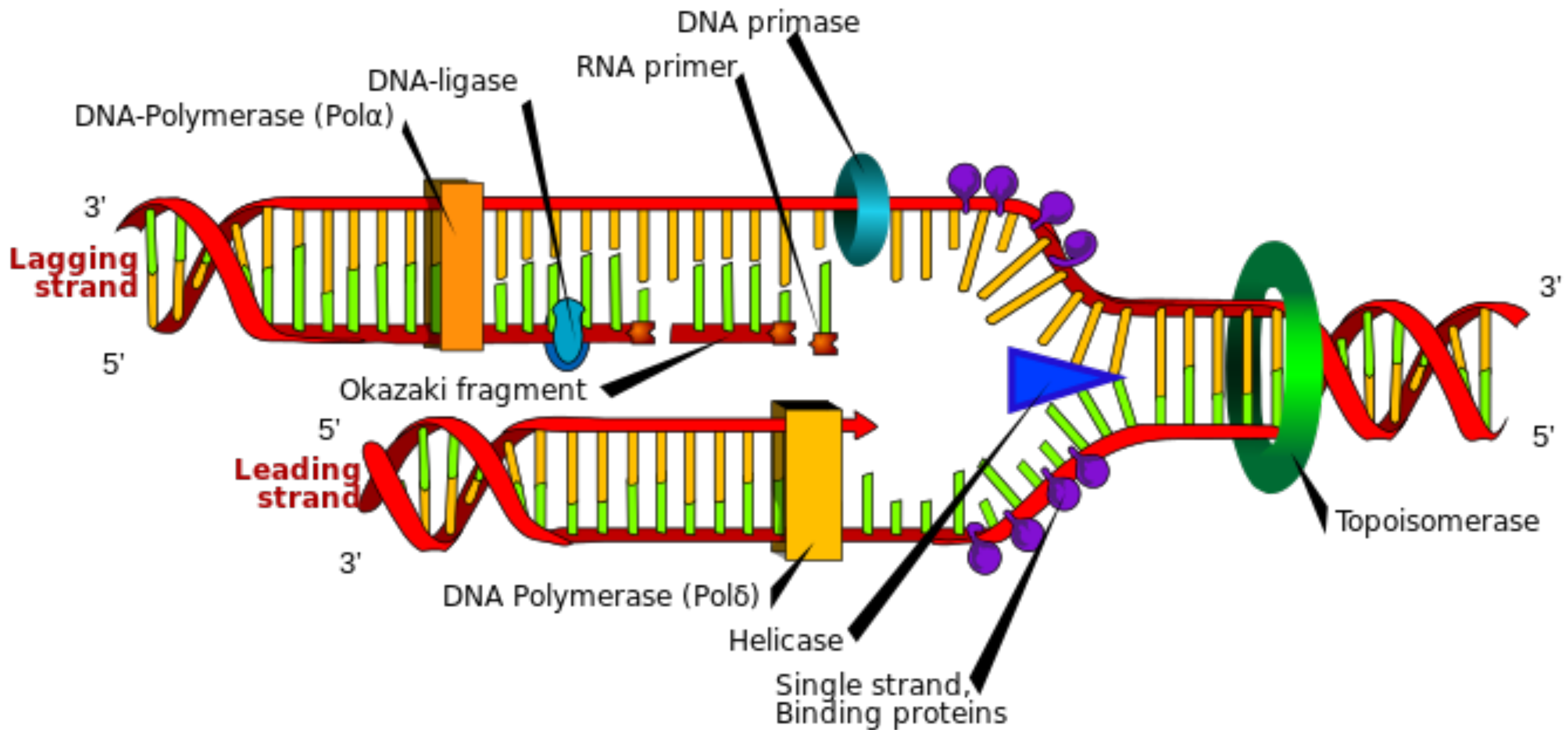
Ion Torrent by LifeTech

Ion Proton by LifeTech

PacBio

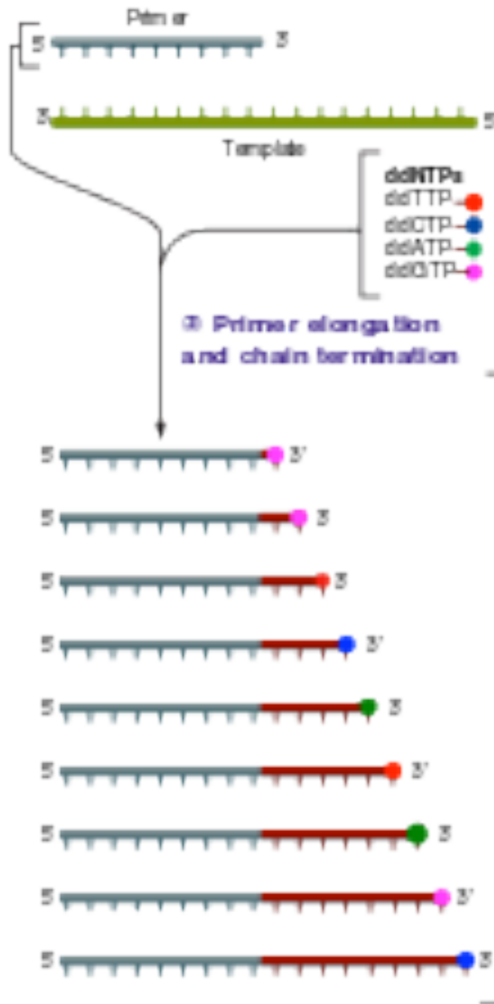
Complete Genomics

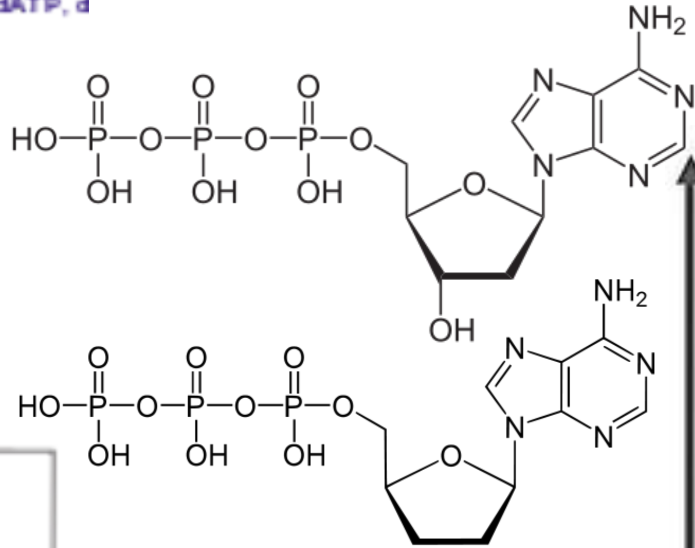
DNA Replication

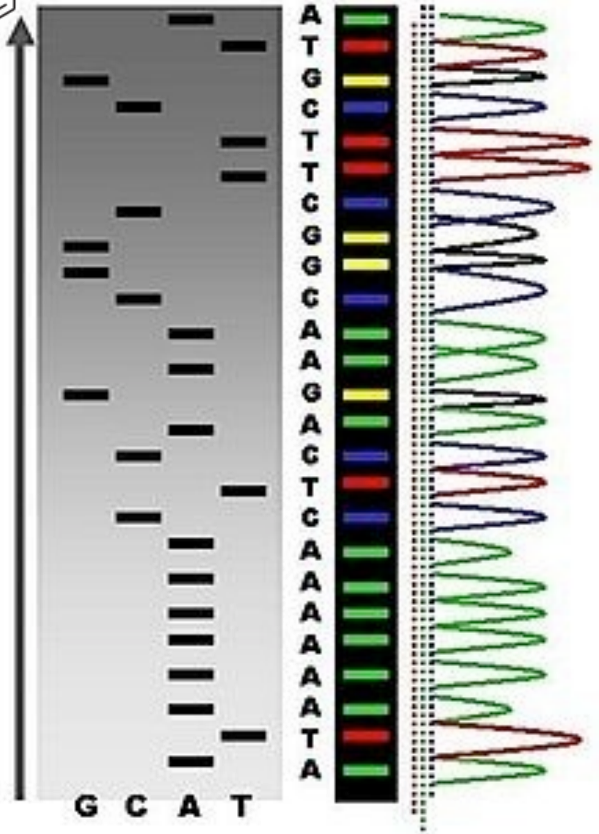


Standard Sanger Sequencing

① Reaction mixture
- Primer and DNA template
- DNA polymerase
- ddNTPs with flourochromes
- dNTPs (dATP, d







Illumina Platforms

- UAB Stem Cell Institute

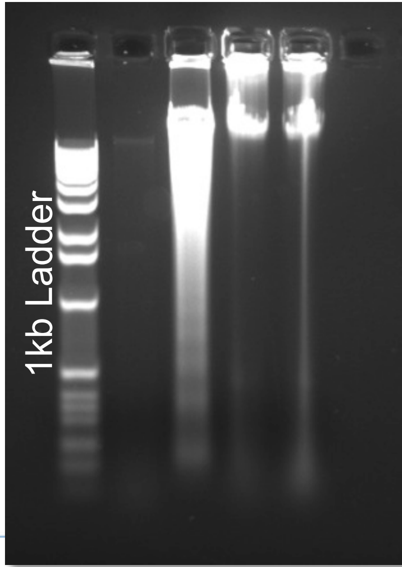
GAIIx
- One flowcell
- ~95billion bases sequenced
- 36bp increments
- Higher cost per base sequenced
- Single read and Paired end reads

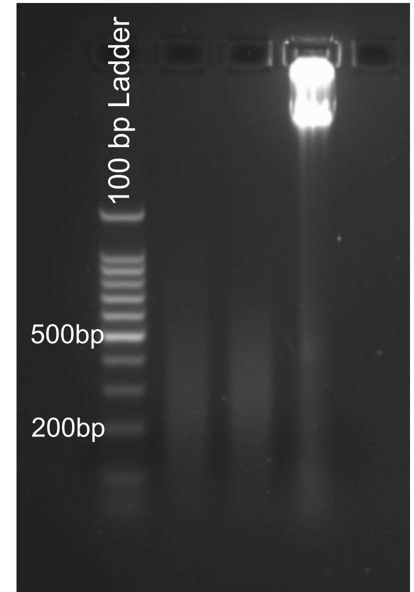
HiSeq2000
- Two flowcells
- ~600billion bases sequenced
- 50bp-100bp increments
- Lower cost per base sequenced
- Single reads and Paired end reads

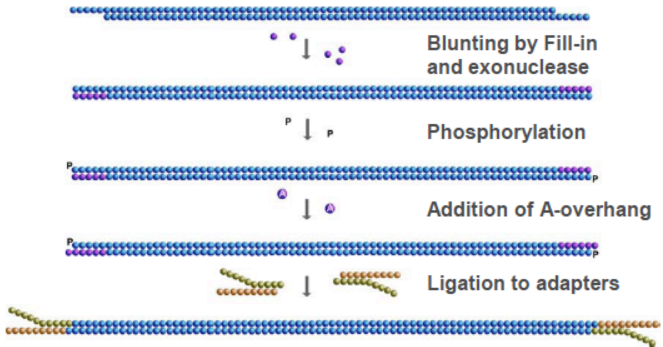
Flowcells through time

2005 → 2010

DNA Library Prep and Flow cell Production







Version 3 HiSeq Flow Cell

Library Assessment and Quantitation

Amplification Curves

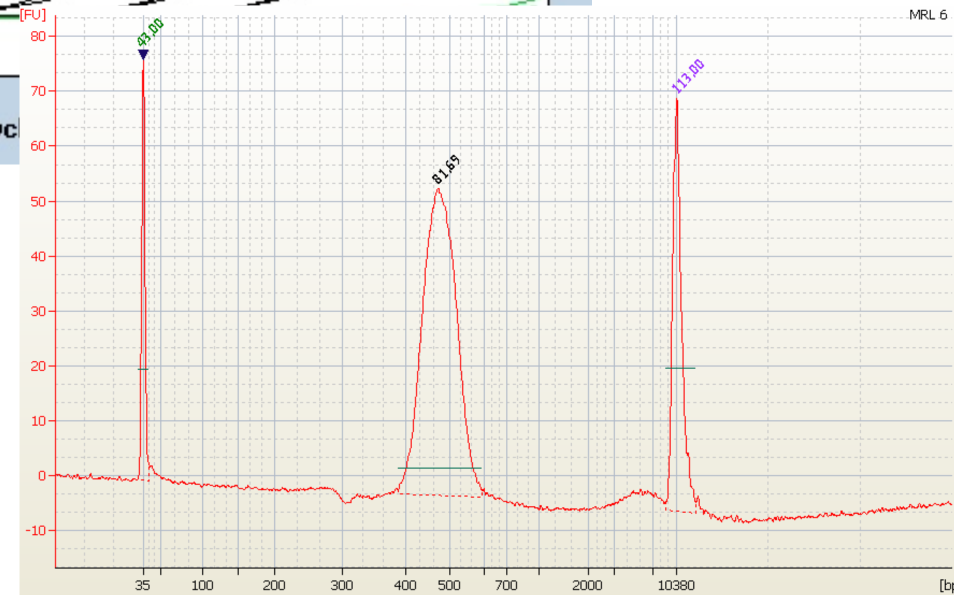
MRL DNA Dilutions
1:1000-1:16,000

Fluorescence (483-533)

10.979
9.979
8.979
7.979
6.979
5.979
4.979
3.979
2.979
1.979
0.979
-0.021

20pM
2pM
0.2 pM
0.02 pM
0.002 pM
0.0002 pM

5
10
15
Cyc

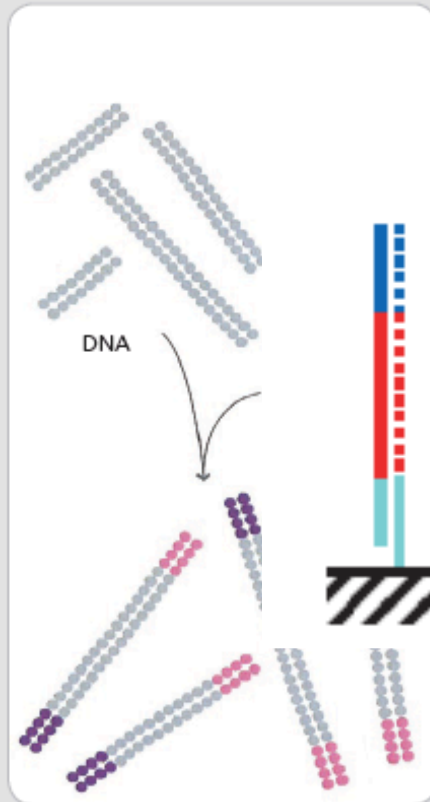


Useful Next-Gen Terms

- Cluster
  - Individual island of DNA molecules representing a single, unique template
- Clusters Passing filter
  - Number of clusters able to be distinguished by the software as individuals
- Fastq
  - DNA Sequence file that is able to be read by downstream analysis applications
- Q-Score
  - A quality score based on the Phred score from Sanger Sequenicng which is the probability a base is incorrect at a give position. Example: Q30 means there is a 1:1000 chance the base is incorrect. Or stated another way it means the base call is 99.9% accurate
- Phasing/Prephasing
  - When the DNA sequencing reaction is either a base ahead or a base behind the majority of the other molecules
- Depth of Coverage
  - The average number of times a base is read within the genome
- Reads
  - Actual sequence

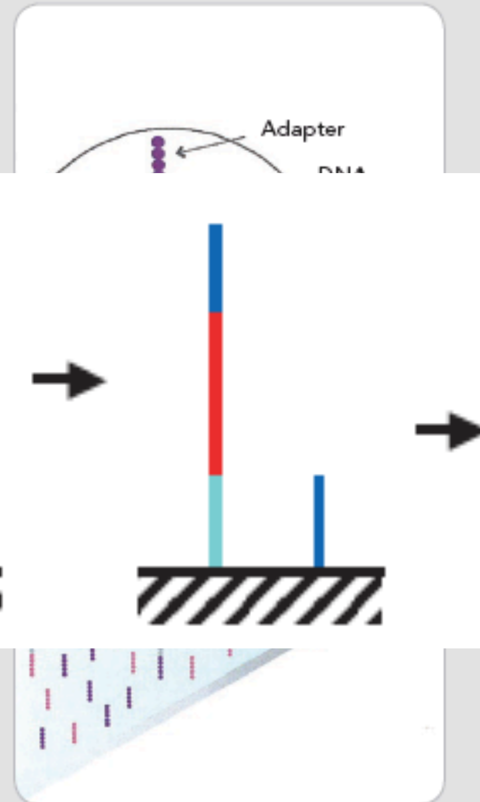
Illumina Cluster Generation

1. PREPARE GENOMIC DNA SAMPLE



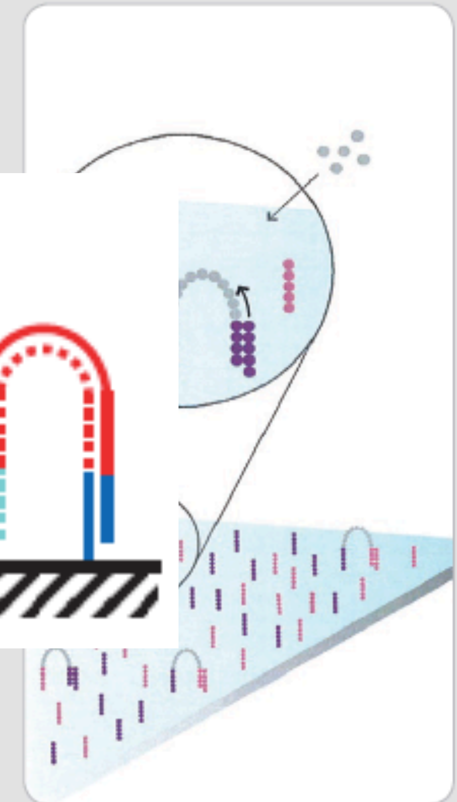
Randomly fragment genomic DNA and ligate adapters to both ends of the fragments.

2. ATTACH DNA TO SURFACE



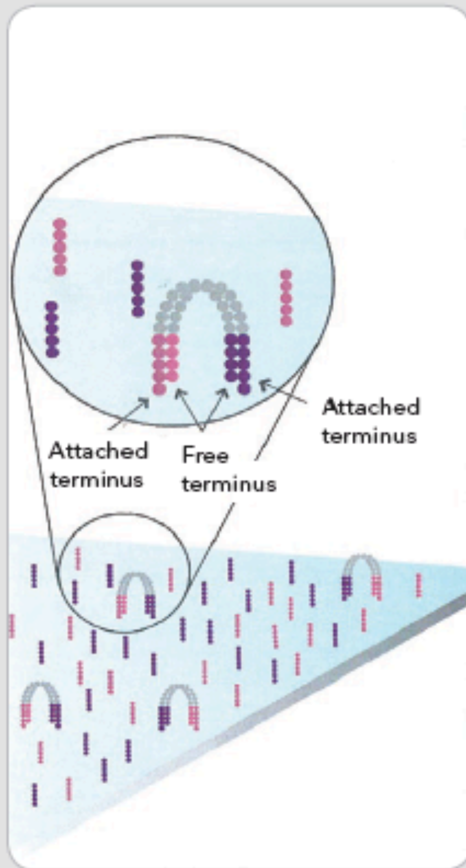
Bind single-stranded fragments randomly to the inside surface of the flow cell channels.

3. BRIDGE AMPLIFICATION



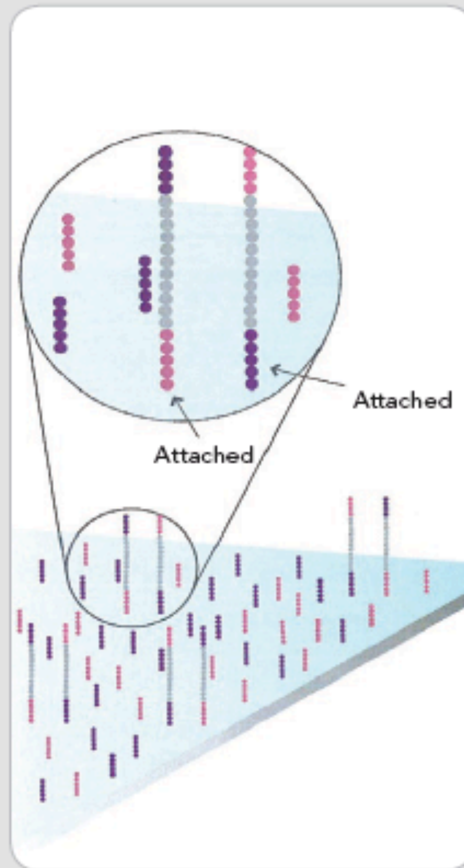
Add unlabeled nucleotides and enzyme to initiate solid-phase bridge amplification.

4. FRAGMENTS BECOME DOUBLE-STRANDED



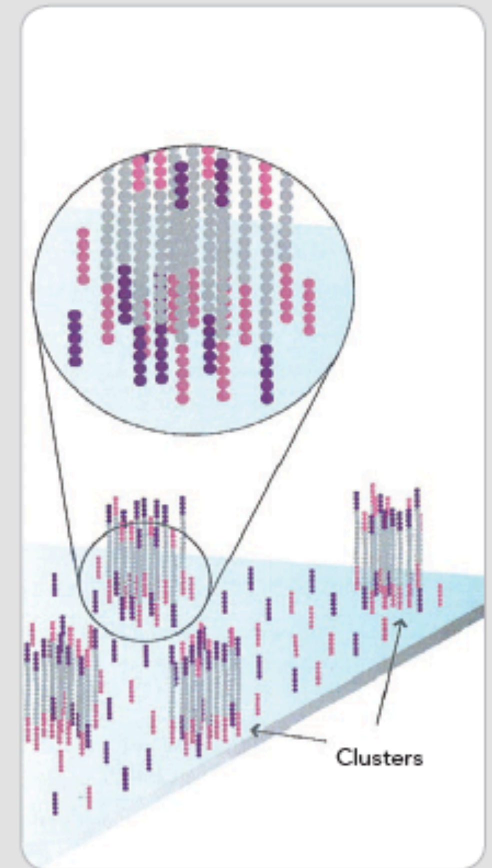
The enzyme incorporates nucleotides to build double-stranded bridges on the solid-phase substrate.

5. DENATURE THE DOUBLE-STRANDED MOLECULES



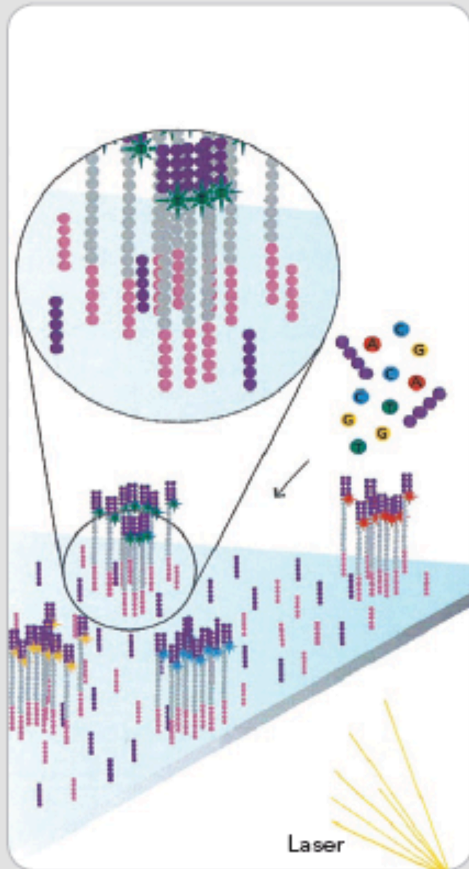
Denaturation leaves single-stranded templates anchored to the substrate.

6. COMPLETE AMPLIFICATION



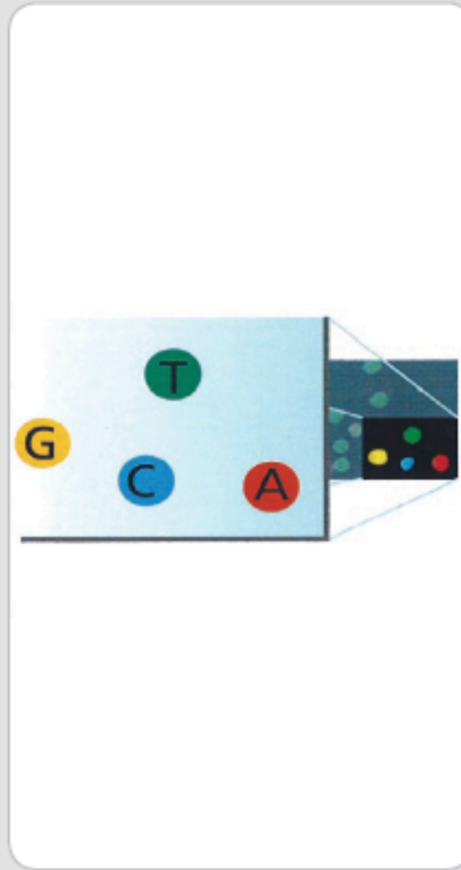
Several million dense clusters of double-stranded DNA are generated in each channel of the flow cell.

7. DETERMINE FIRST BASE



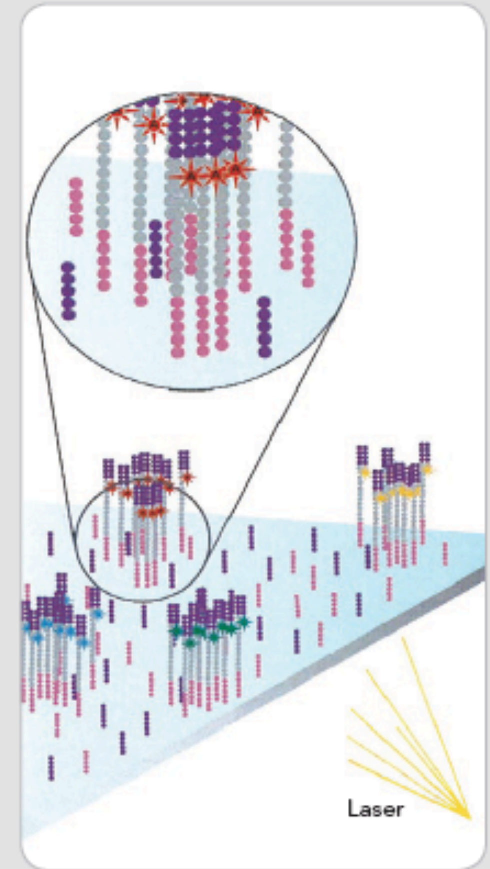
The first sequencing cycle begins by adding four labeled reversible terminators, primers, and DNA polymerase.

8. IMAGE FIRST BASE



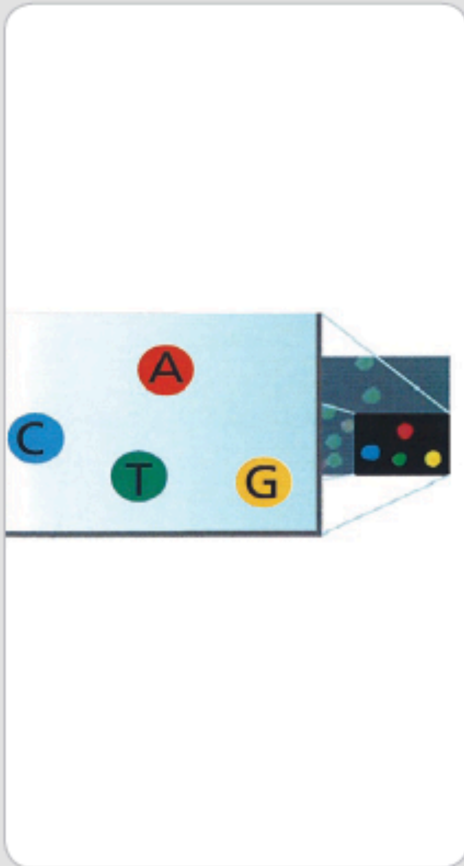
After laser excitation, the emitted fluorescence from each cluster is captured and the first base is identified.

9. DETERMINE SECOND BASE



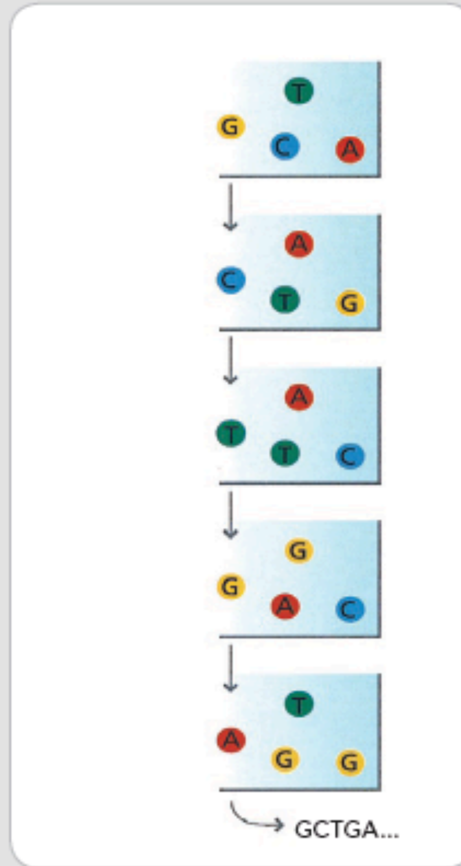
The next cycle repeats the incorporation of four labeled reversible terminators, primers, and DNA polymerase.

10. IMAGE SECOND CHEMISTRY CYCLE



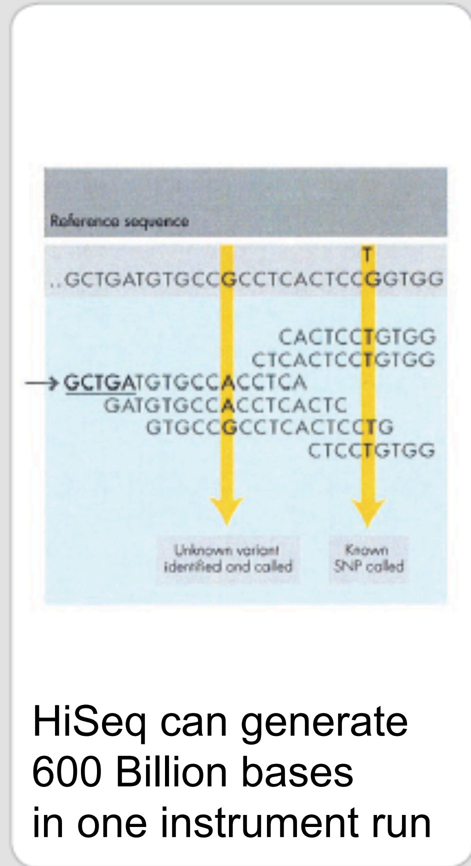
After laser excitation, the image is captured as before, and the identity of the second base is recorded.

11. SEQUENCING OVER MULTIPLE CHEMISTRY CYCLES



The sequencing cycles are repeated to determine the sequence of bases in a fragment, one base at a time.

12. ALIGN DATA

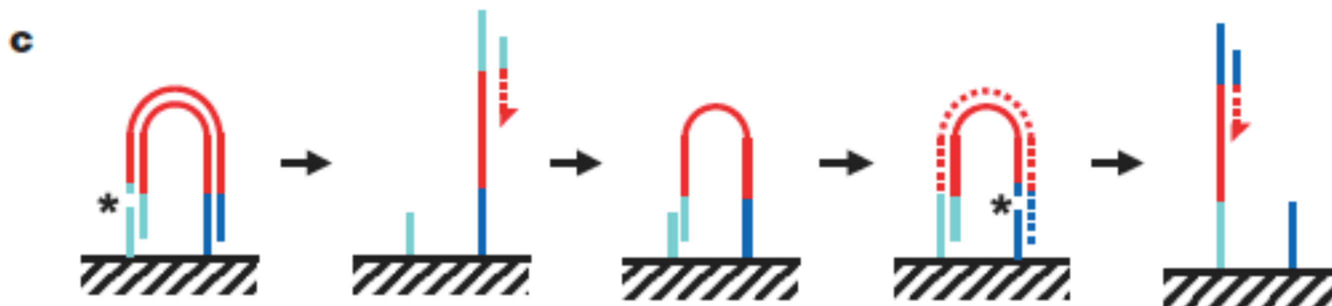


HiSeq can generate 600 Billion bases in one instrument run

The data are aligned and compared to a reference, and sequencing differences are identified.

c. Paired End Sequencing

d. Mate Paired End Sequenicng



c.

Adapters containing attachment sequences (A1 & A2) and sequencing primer sites (SP1 & SP2) are ligated onto DNA fragments (e.g., genomic DNA). The resulting library of single molecules is attached to a flow cell. Each end of every template is read sequentially.

d.

Mate pair library preparation is designed to generate short fragments consisting of two segments that originally had a separation of several kilobases in the genome. Fragments of sample genomic DNA is end-biotinylated to tag the eventual mate pair segments. Self-circularization and refragmentation of these large fragments generates a population of small fragments, some of which contain both mate pair segments with no intervening sequence. These mate pair fragments are enriched using their biotin tag. Mate pairs are sequenced using a similar two-adaptor strategy as described for paired-end sequencing.

Illumina Sequencing Analysis Viewer 1.7.25 - 111208_SN372_0101_AD0JRMACXX

Sequencing Analysis Viewer

Run Folder: Y:\111208_SN372_0101_AD0JRMACXX

Browse | Refresh

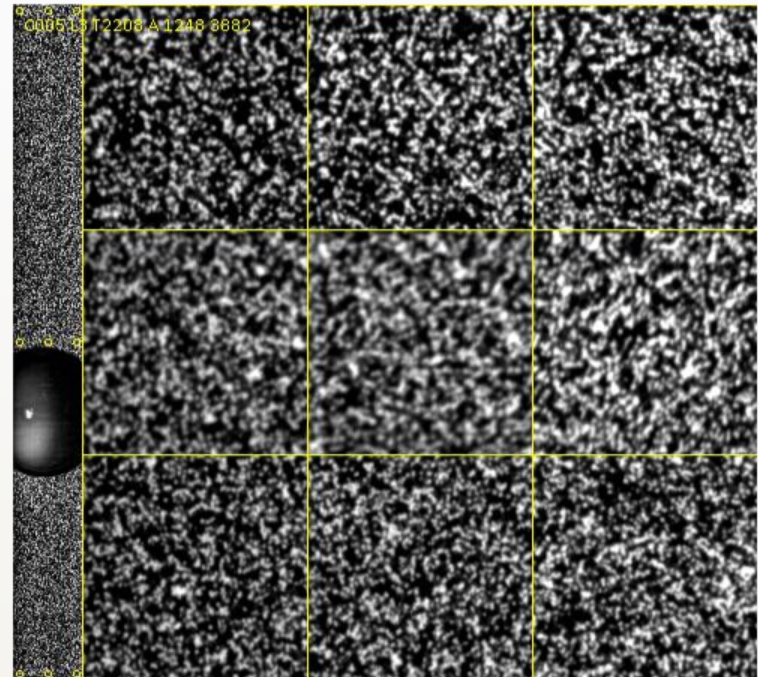
Analysis | Imaging | Summary | Tile Status | Controls

Cycle All | Lane 3 | Surface Bottom | Swath Middle | Section All

Index	Lane	Tile	Section	Cycle	Surface	Swath	Time	P90 A	P90 C	P90 G	P9
33148	3	2207	7	100	Bottom	Middle	12/13/201...	1871	2966	931	22
33149	3	2207	7	101	Bottom	Middle	12/13/201...	1840	2929	919	22
33150	3	2207	7	102	Bottom	Middle	12/13/201...	1822	2893	922	22
33151	3	2207	7	103	Bottom	Middle	12/14/201...	1805	2878	906	22
33152	3	2207	7	104	Bottom	Middle	12/14/201...	1802	2876	902	21
33153	3	2207	7	105	Bottom	Middle	12/14/201...	1785	2841	908	21
33154	3	2207	7	106	Bottom	Middle	12/14/201...	1756	2836	874	21
33155	3	2207	7	107	Bottom	Middle	12/14/201...	1749	2813	872	21
33156	3	2207	7	108	Bottom	Middle	12/14/201...	2498	3963	938	25
33913	3	2208	8	1	Bottom	Middle	12/08/201...	2928	4976	1861	35
33914	3	2208	8	2	Bottom	Middle	12/08/201...	3176	4792	1636	41
33915	3	2208	8	3	Bottom	Middle	12/08/201...	3163	4773	1679	38
33916	3	2208	8	4	Bottom	Middle	12/08/201...	3259	4788	1690	34
33917	3	2208	8	5	Bottom	Middle	12/08/201...	2732	4112	1533	28
33918	3	2208	8	6	Bottom	Middle	12/09/201...	3126	4605	1475	33
33919	3	2208	8	7	Bottom	Middle	12/09/201...	2712	4312	1404	38

Rows=41472 | Disp=864 | Sel=1 | Filter

A C G T



start | UAB - The Univer... | runs on 'Isilon Ser... | Illumina Sequenci... | Microsoft PowerP... | Document1 - Micr... | Search Desktop | 3:53 PM

Illumina Sequencing Analysis Viewer 1.7.25 - 111208_SN372_0101_AD0JRMACXX

Sequencing Analysis Viewer

Run Folder: Y:\111208_SN372_0101_AD0JRMACXX

Browse | Refresh

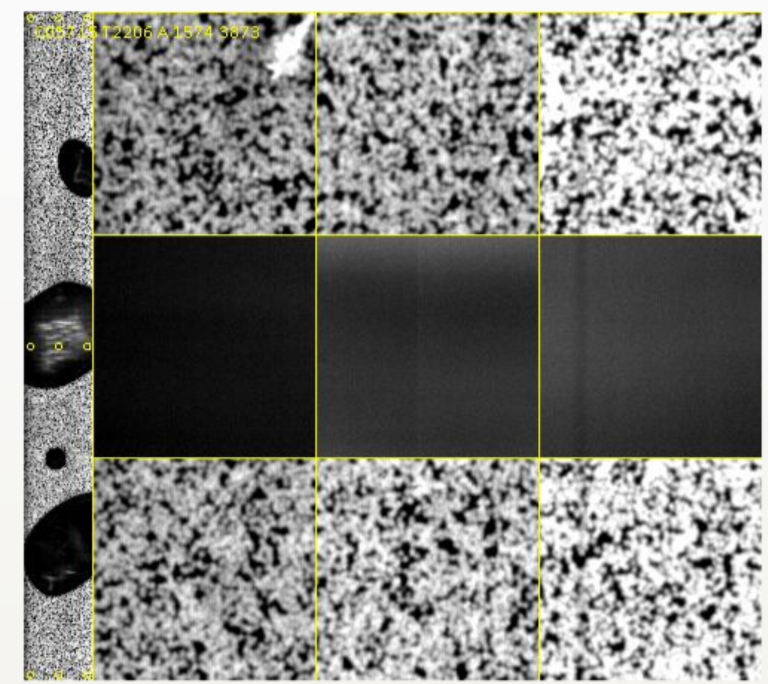
Analysis | Imaging | Summary | Tile Status | Controls

Cycle All | Lane 5 | Surface Bottom | Swath Middle | Section All

A C G T

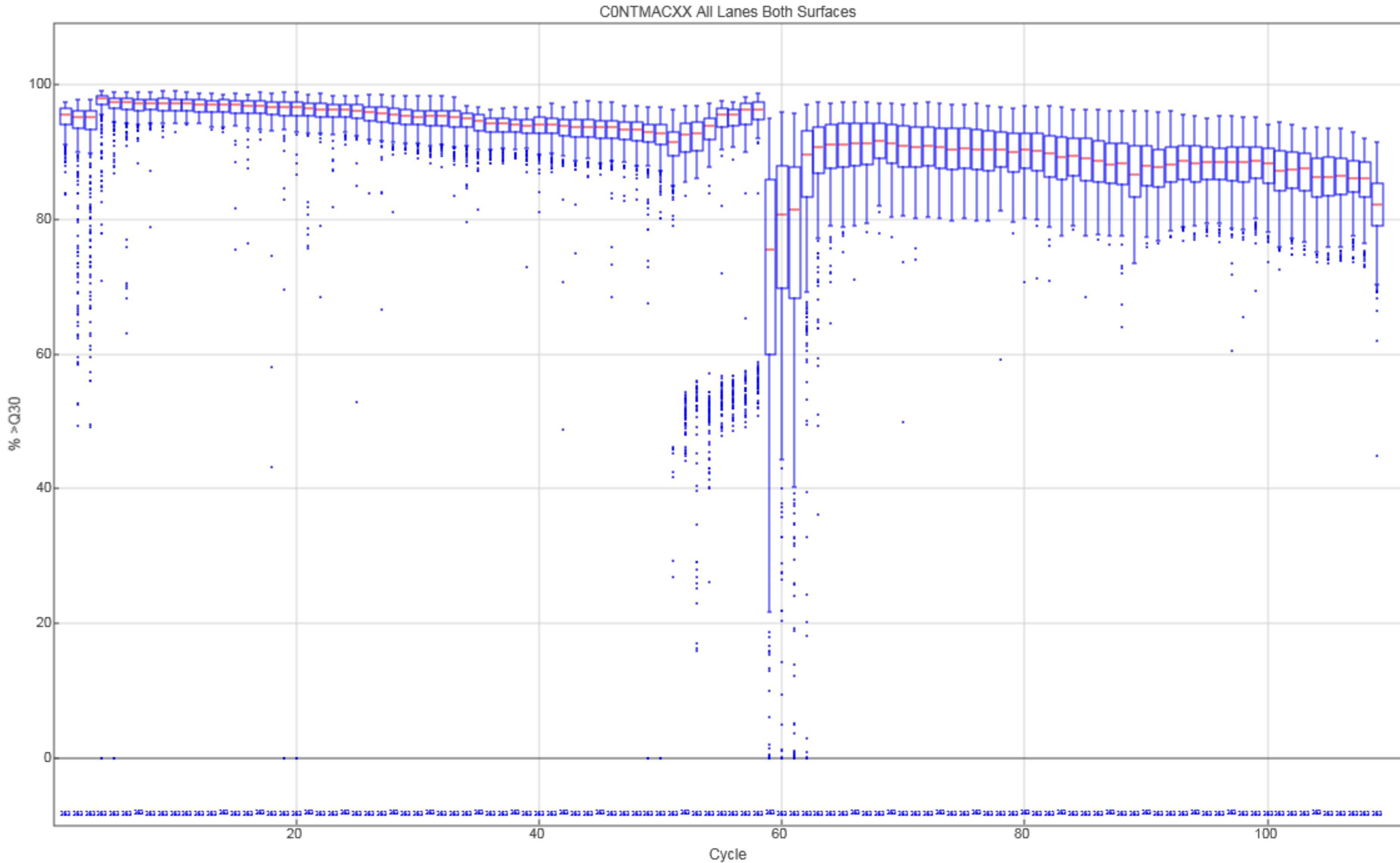
Index	Lane	Tile	Section	Cycle	Surface	Swath	Time	P90 A	P90 C	P90 G	P9
32444	5	2206	6	44	Bottom	Middle	12/10/201...	2142	3333	1001	26
32445	5	2206	6	45	Bottom	Middle	12/10/201...	2138	3327	993	26
32446	5	2206	6	46	Bottom	Middle	12/10/201...	2091	3282	971	25
32447	5	2206	6	47	Bottom	Middle	12/10/201...	2089	3279	962	25
32448	5	2206	6	48	Bottom	Middle	12/10/201...	2059	3240	954	25
32449	5	2206	6	49	Bottom	Middle	12/10/201...	2058	3229	960	25
32450	5	2206	6	50	Bottom	Middle	12/10/201...	2034	3205	942	25
32451	5	2206	6	51	Bottom	Middle	12/10/201...	853	4552	123	26
32452	5	2206	6	52	Bottom	Middle	12/10/201...	165	303	166	44
32453	5	2206	6	53	Bottom	Middle	12/10/201...	0	198	0	39
32454	5	2206	6	54	Bottom	Middle	12/10/201...	0	149	258	21
32455	5	2206	6	55	Bottom	Middle	12/10/201...	193	326	167	24
32456	5	2206	6	56	Bottom	Middle	12/10/201...	3099	4447	132	29
32457	5	2206	6	57	Bottom	Middle	12/10/201...	2102	3104	128	26
32458	5	2206	6	58	Bottom	Middle	12/12/201...	2537	3885	1048	24
32459	5	2206	6	59	Bottom	Middle	12/12/201...	1514	2365	696	18

Rows=41472 | Disp=864 | Sel=1 | Filter

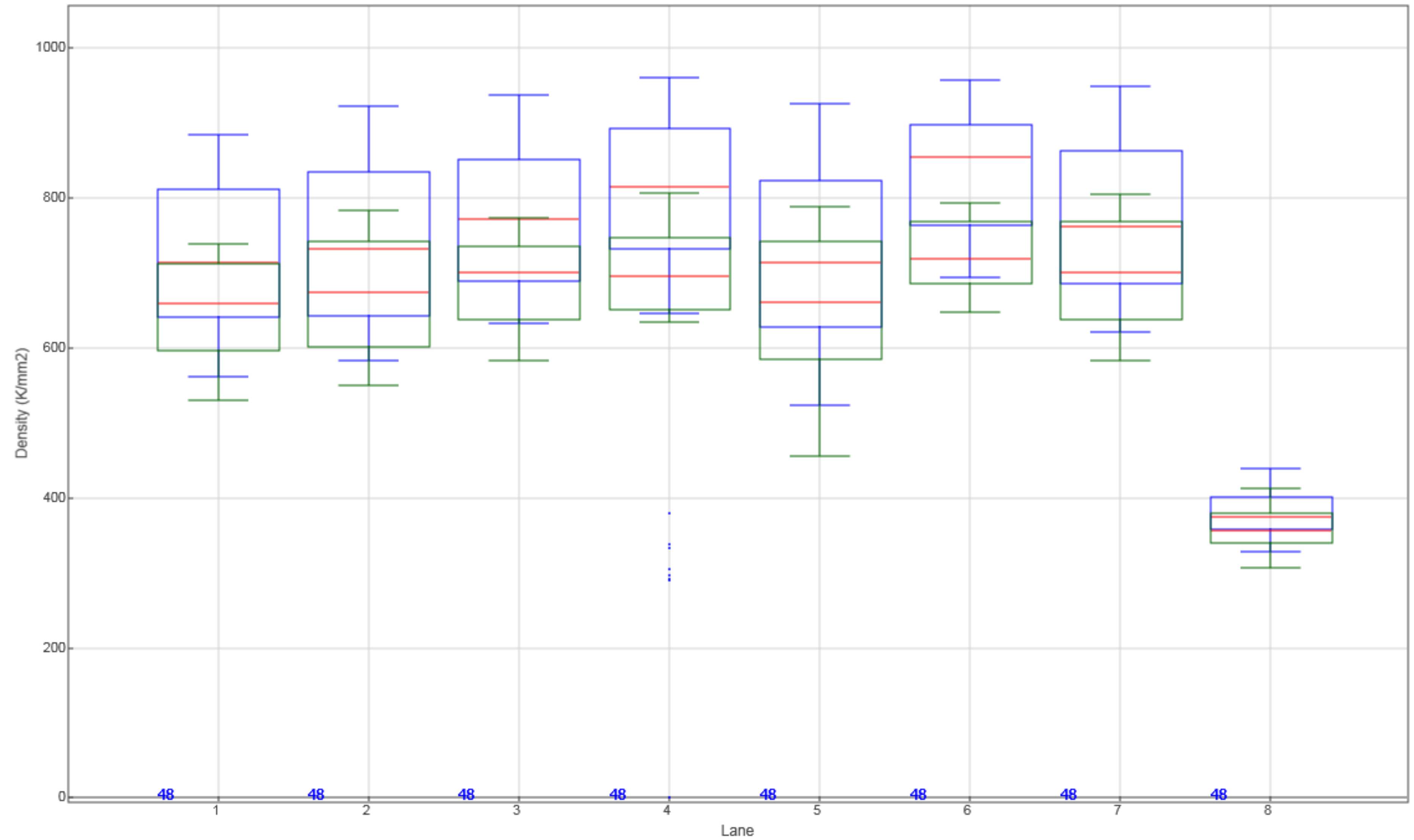


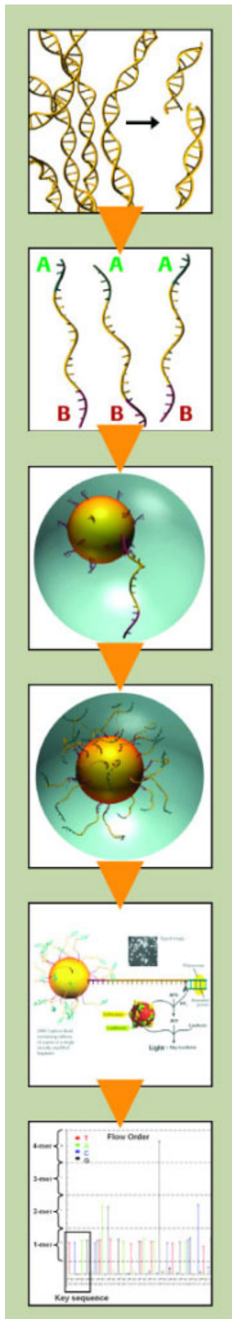
start | UAB - The Univer... | runs on 'Isilon Ser... | Illumina Sequenci... | Microsoft PowerP... | Document1 - Micr... | Search Desktop | 4:05 PM

Percent Q30 Scores per cycle for all lanes and both surfaces

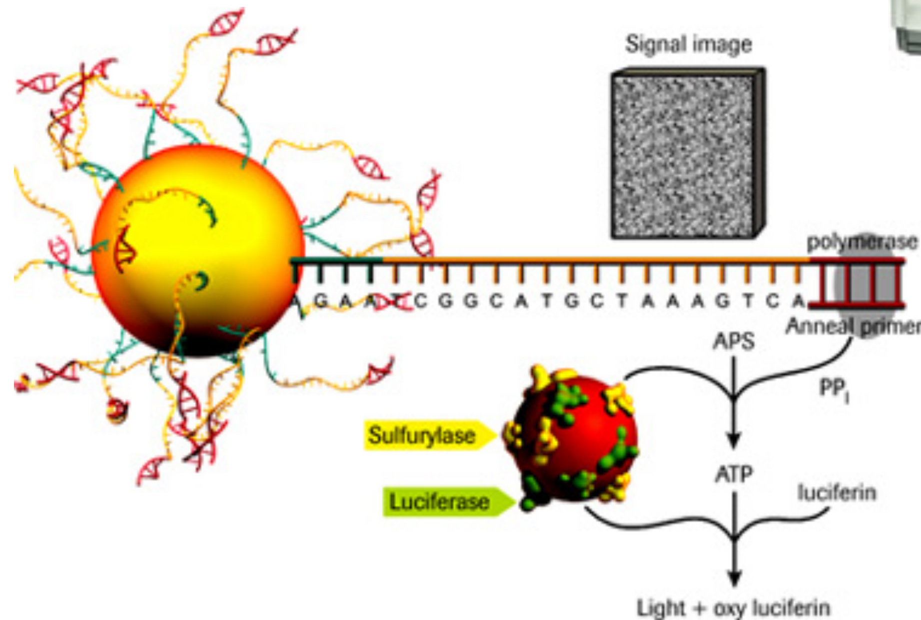


Cluster Density





454 LifeSciences



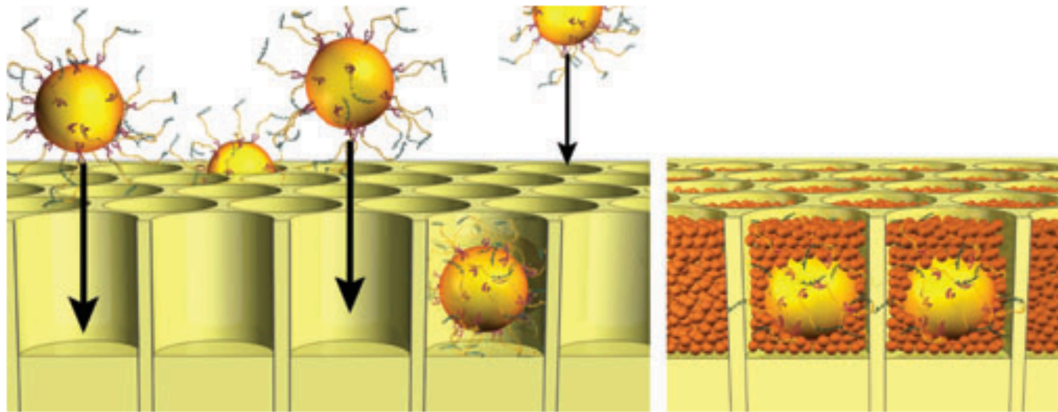
With Titanium Chemistry can generate up to 1000bp/template
700 Million bases/run

454

c

Sequencing

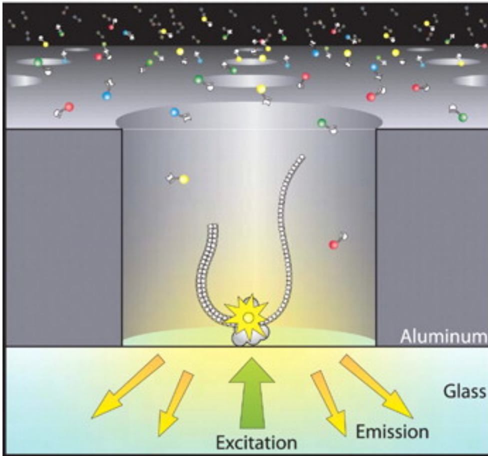
7.5 hours

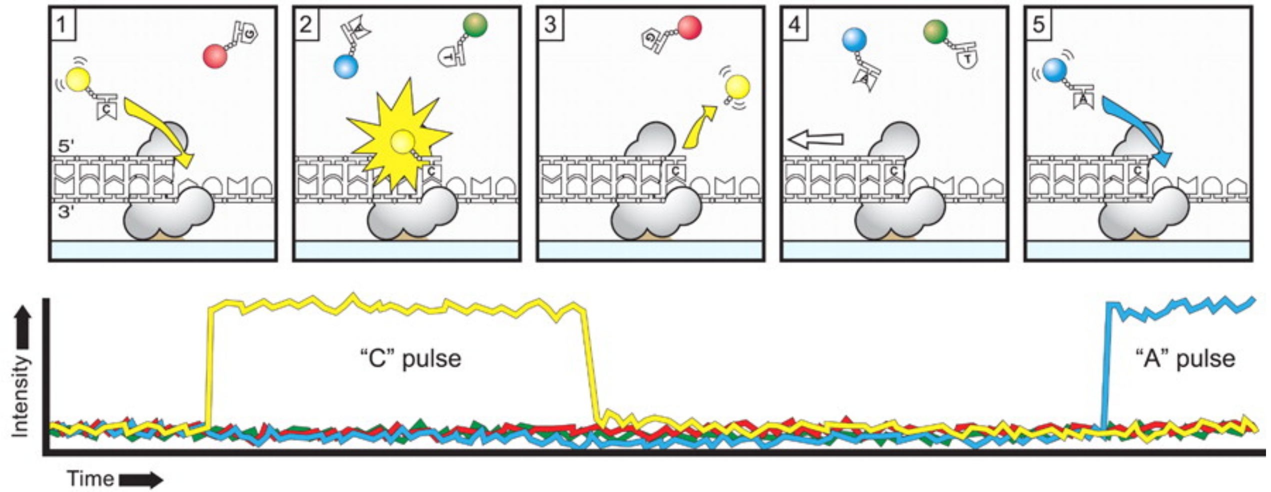


- Well diameter: average of 44 µm
- 400,000 reads obtained in parallel
- A single cloned amplified sstDNA bead is deposited per well

Amplified sstDNA library beads Quality filtered bases

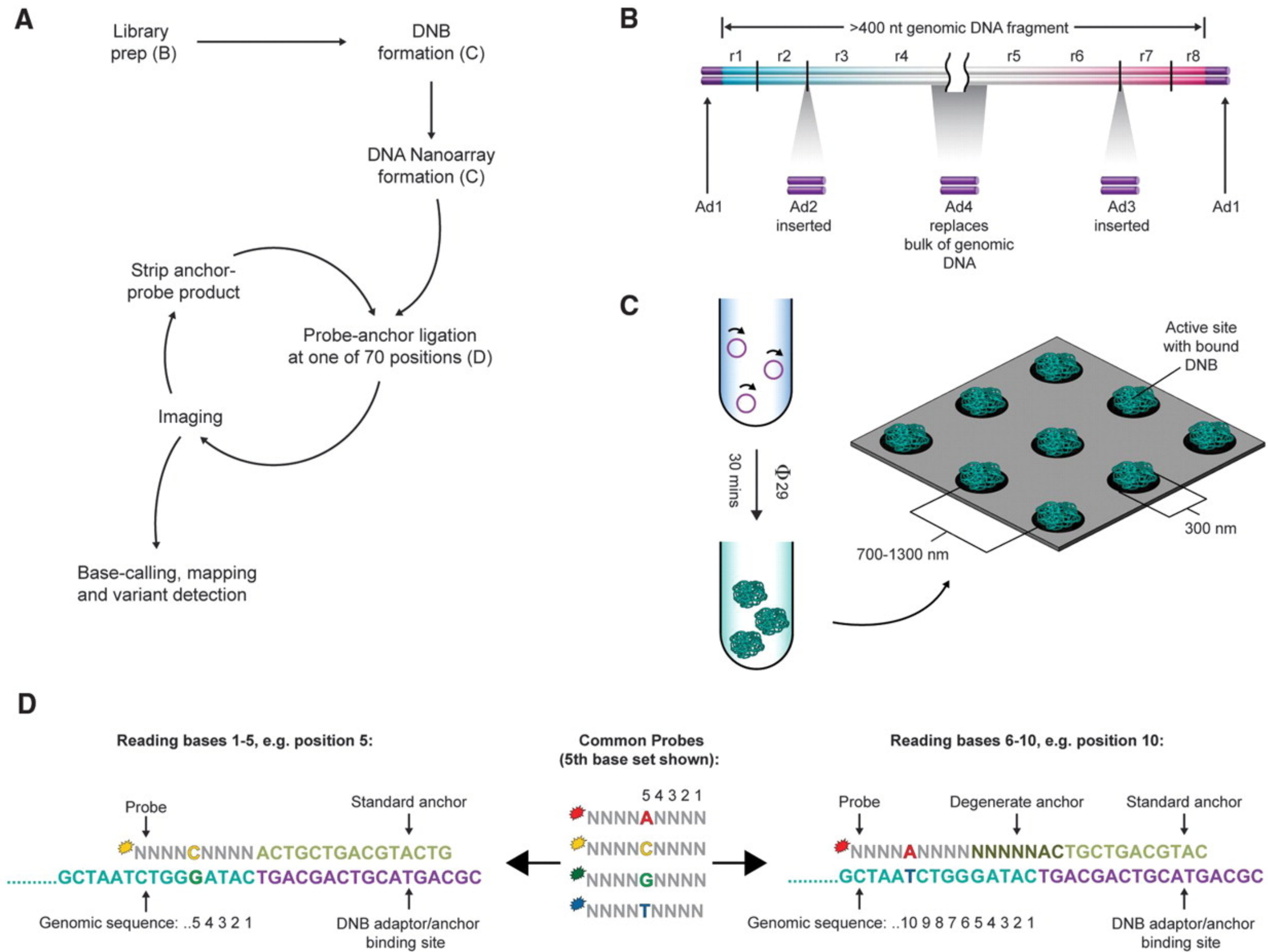
Pacific Biosciences Technology





J Eid et al. Science 2009;323:133-138

Complete Genomics Technology



R Drmanac et al. Science 2010;327:78-81

Next Next-Gen Sequencing (3G Seq)

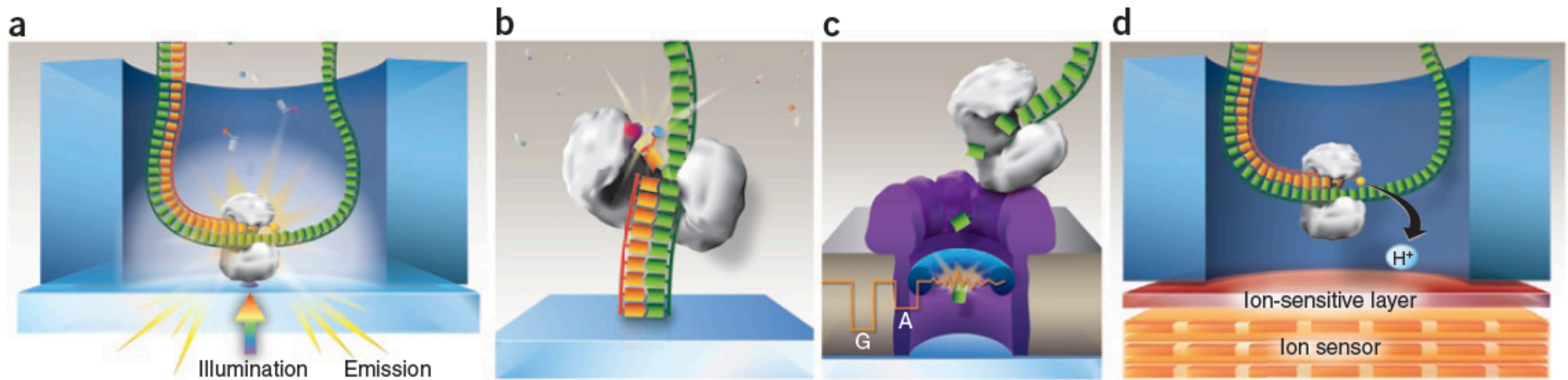


Figure 1 Third-generation sequencing platforms. (**a**) Pacific Biosciences SMRT (single-molecule real-time) DNA sequencing method. The platform uses a DNA polymerase anchored to the bottom surface of a ZMW (pictured in cross section). Differentially labeled nucleotides enter the ZMW via diffusion and occupy the 'detection volume' (white translucent halo area) or microseconds. During an incorporation event, the labeled nucleotide is 'held' within the detection volume by the polymerase for tens of milliseconds. As each nucleotide is incorporated, the label, located on the terminal phosphate, is cleaved off and diffuses out of the ZMW. (**b**) Life Technologies FRET sequencing platform uses base fluorescent labeling technology, a DNA polymerase modified with a quantum dot and DNA template molecules immobilized onto a solid surface. During an incorporation event, energy is transferred from the quantum dot to an acceptor fluorescent moiety on each labeled base. Light emission can only emanate from labeled nucleotides as they are being incorporated. (**c**) The Oxford nanopore sequencing platform uses an exonuclease coupled to a modified α-hemolysin nanopore (purple, pictured in cross section) positioned within a lipid bilayer. As sequentially cleaved bases are directed through the nanopore, they are transiently bound by a cyclodextrin moiety (blue), disturbing current through the nanopore in a manner characteristic for each base. (**d**) The Ion Torrent sequencing platform uses a semiconductor-based high-density array of microwell reaction chambers positioned above an ion-sensitive layer and an ion sensor. Single nucleotides are added sequentially, and incorporation is recorded by measuring hydrogen ions released as a by-product of nucleotide chain elongation.