Next Generation Sequencing Technologies

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

Different Platforms rely on Different Technology

HiSeq or GAIIx by Illumina

SOLiD by Applied Biosystems

454 Lifesciences by Roche

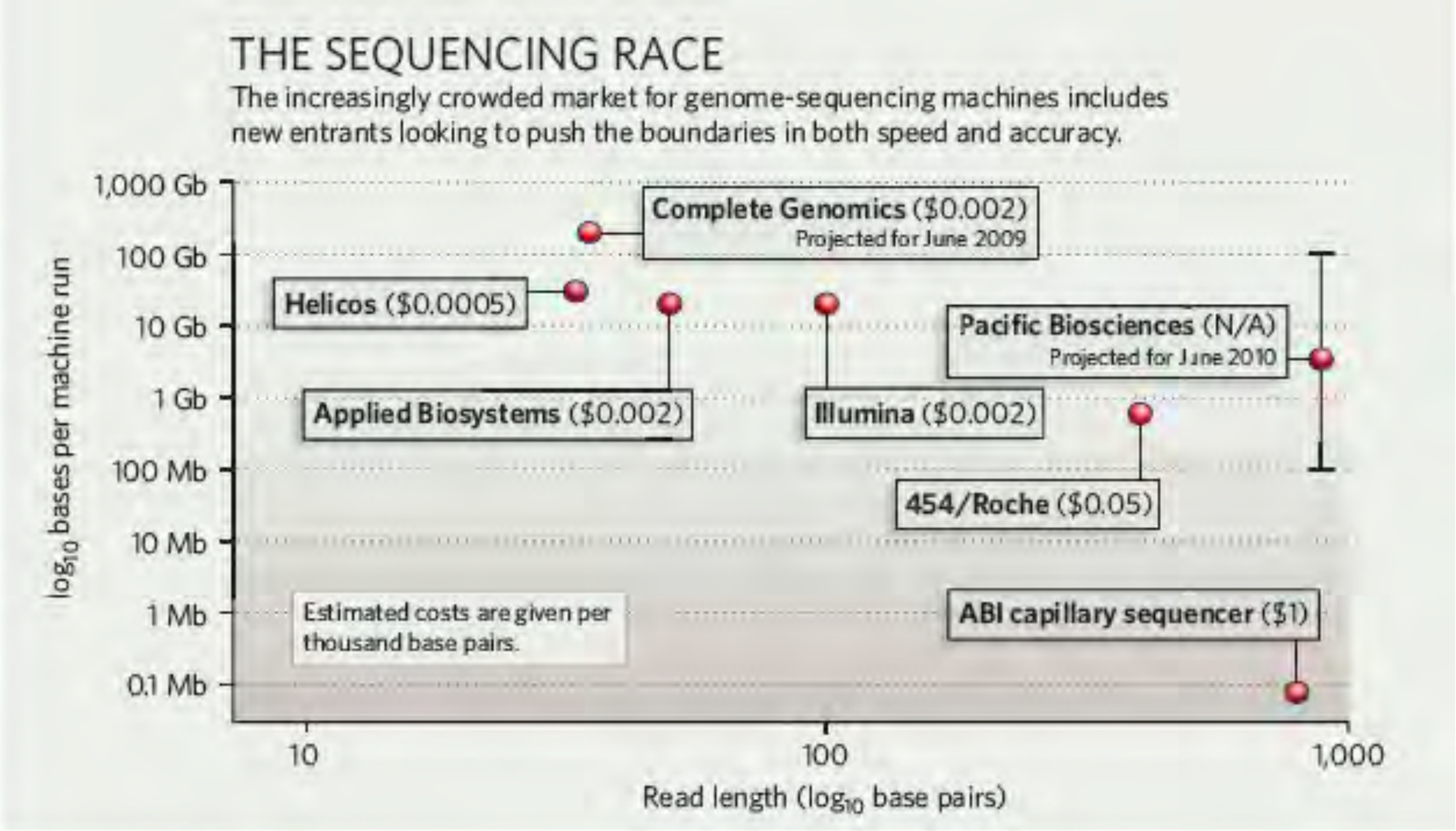
Ion Torrent by LifeTech

Ion Proton by LifeTech

PacBio

Complete Genomics

Basepair v. Cost



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 Platforms

- UAB Stem Cell Institute

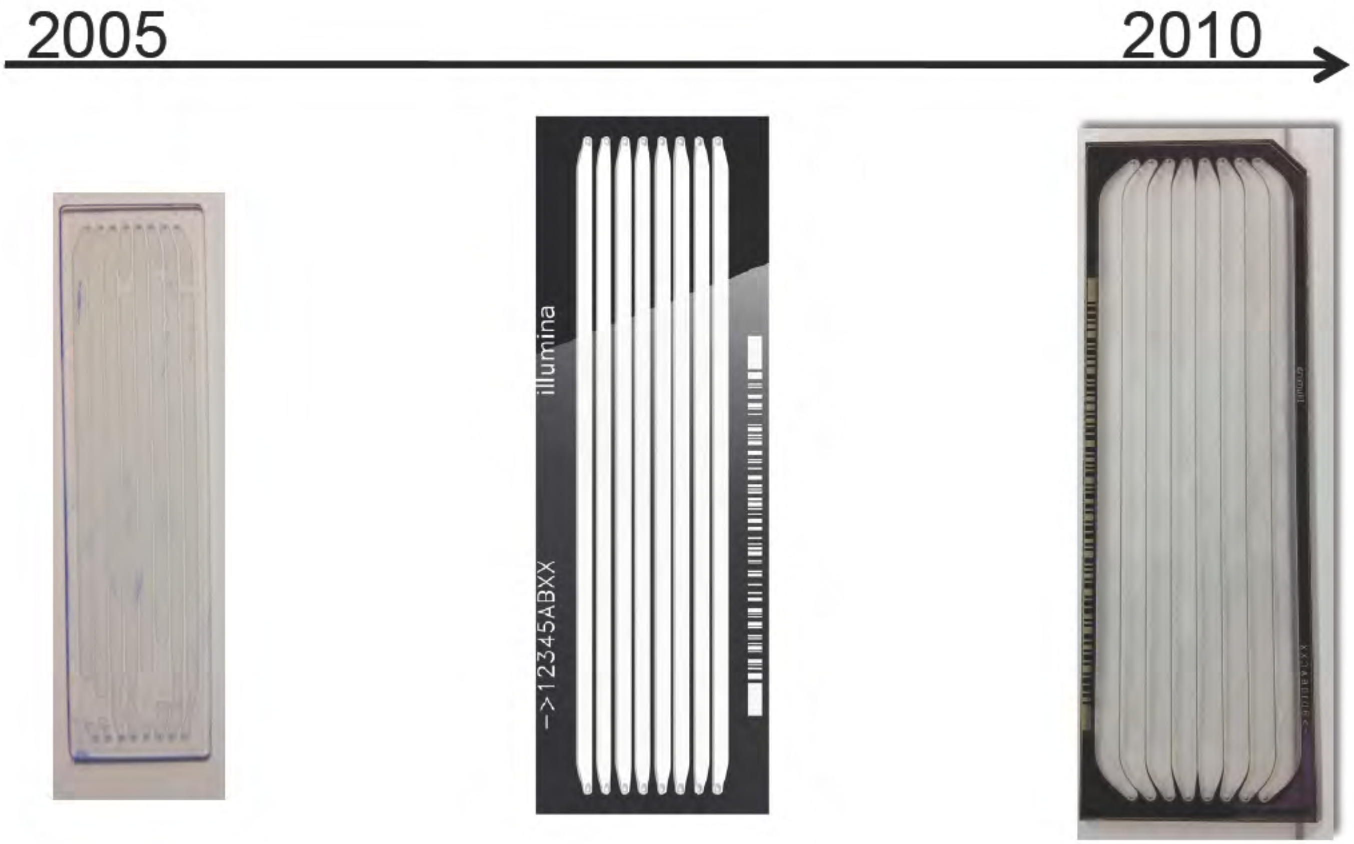
GAIIx

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

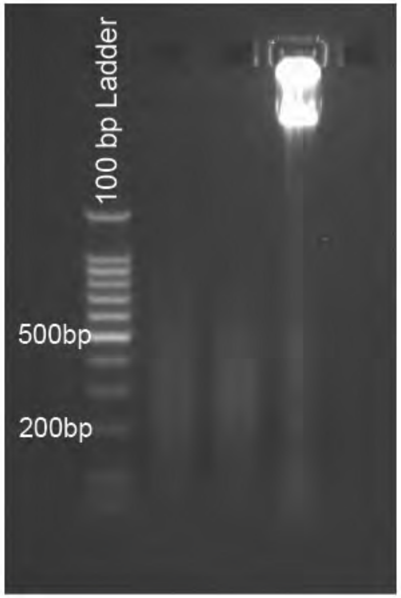
HiSeq2000

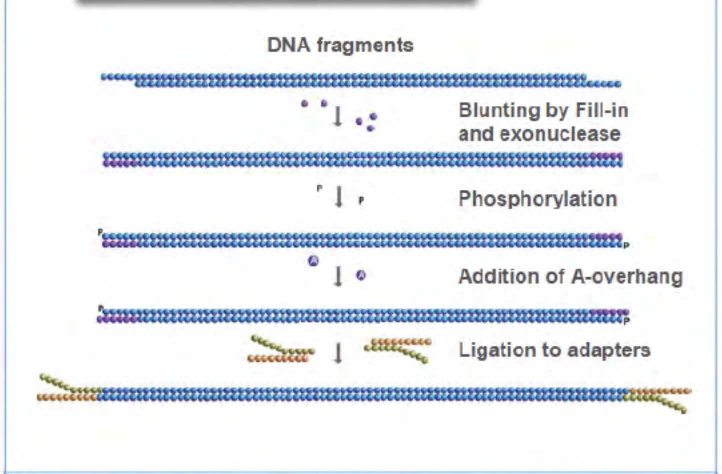
- Two flowcells
- ~600billion bases sequenced
- 50bp increments
- Lower cost per base sequenced
- Single reads and Paired end reads

Flowcells through time



DNA Library Prep and Flow cell Production

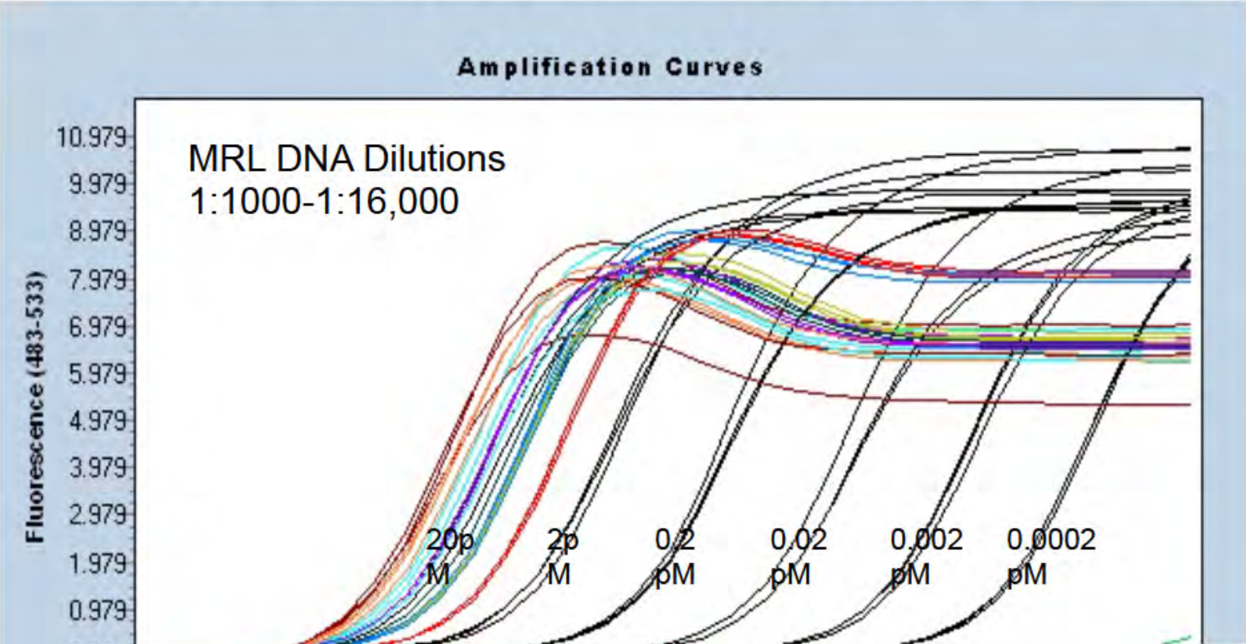


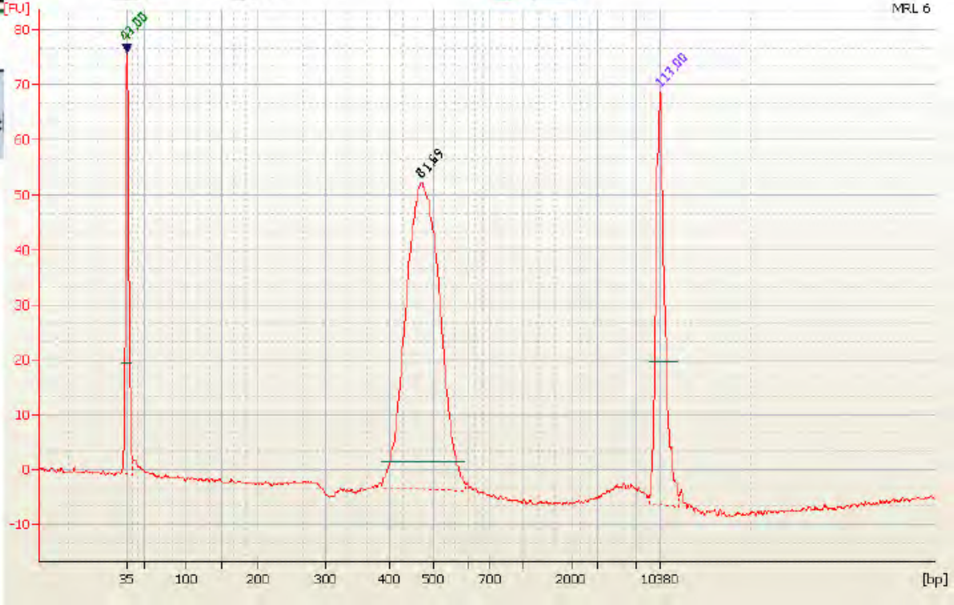


Illumina's Library Preparation Workflow

Version 3 HiSeq Flow Cell

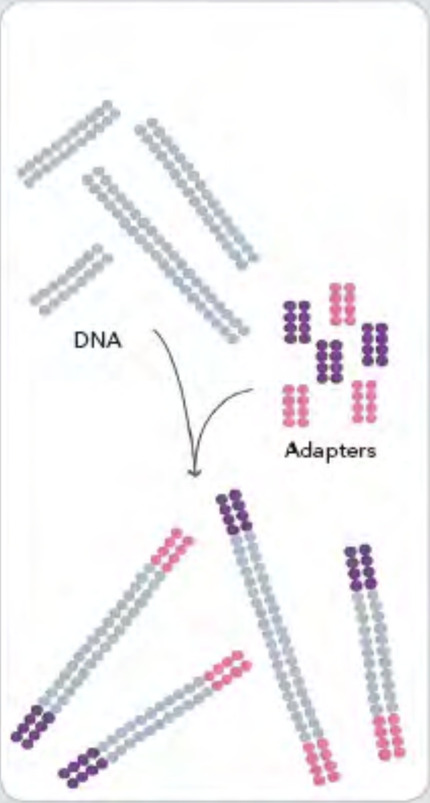
Library Assessment and Quantitation





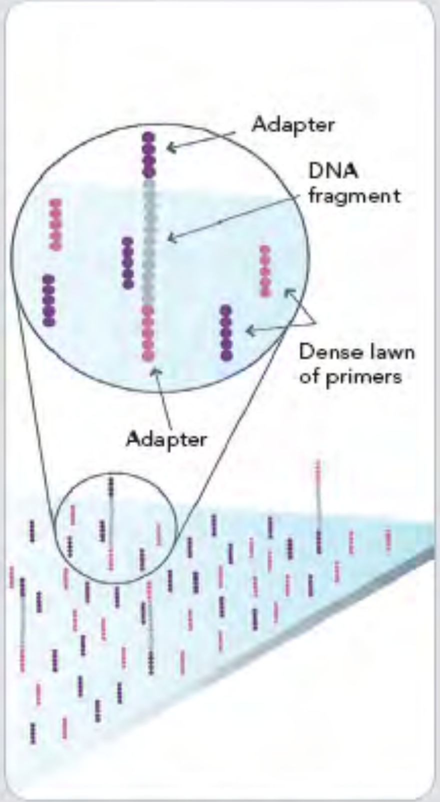
Illumina Cluster Generation

1. PREPARE GENOMIC DNA SAMPLE



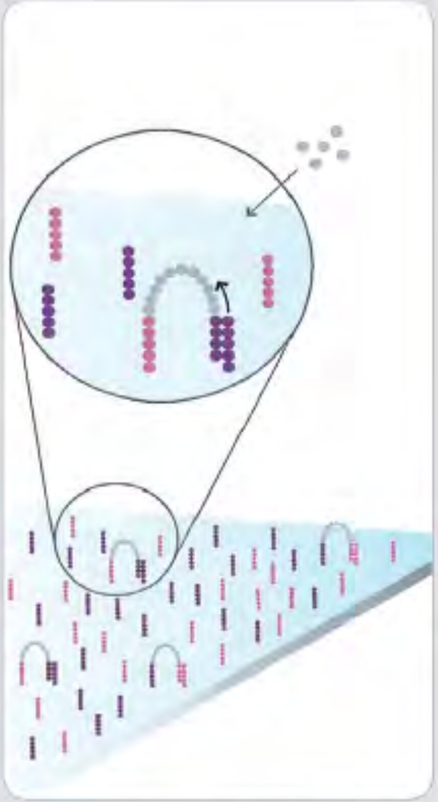
Randomly fragment genomic DNA and ligate adapters to both ends of the fragments.

2. ATTACH DNA TO SURFACE



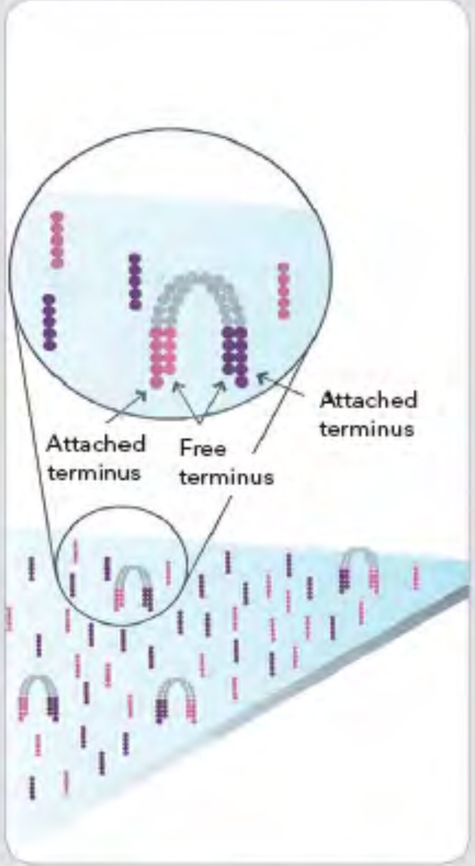
Bind single-stranded fragments randomly to the inside surface of the flow cell channels.

3. BRIDGE AMPLIFICATION



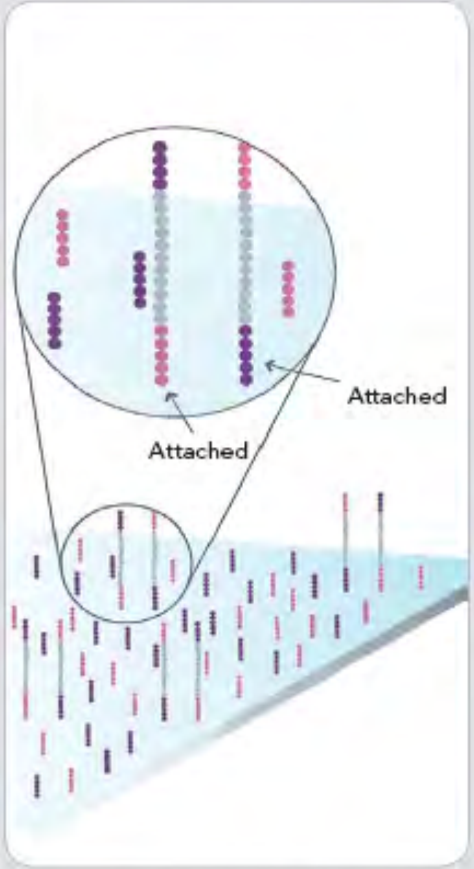
Add unlabeled nucleotides and enzyme to initiate solid-phase bridge amplification.

4. FRAGMENTS BECOME DOUBLE-STRANDED



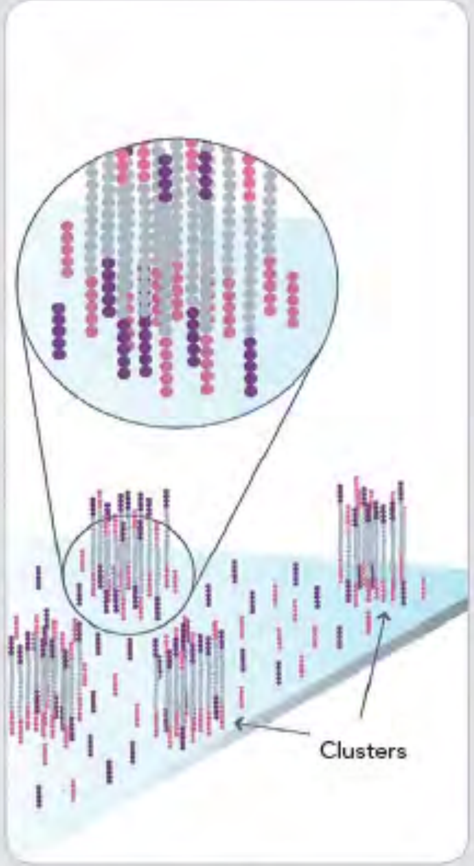
The enzyme incorporates nucleotides to build double-stranded bridges on the solid-phase substrate.

5. DENATURE THE DOUBLE-STRANDED MOLECULES



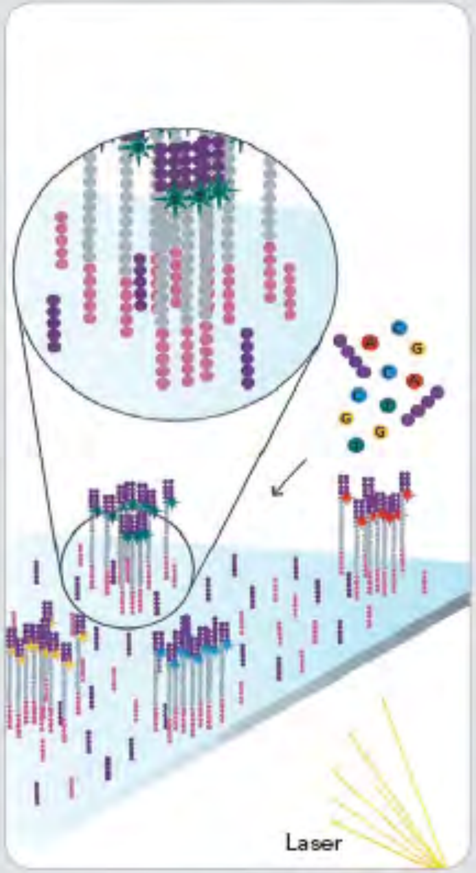
Denaturation leaves single-stranded templates anchored to the substrate.

6. COMPLETE AMPLIFICATION



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

7. DETERMINE FIRST BASE

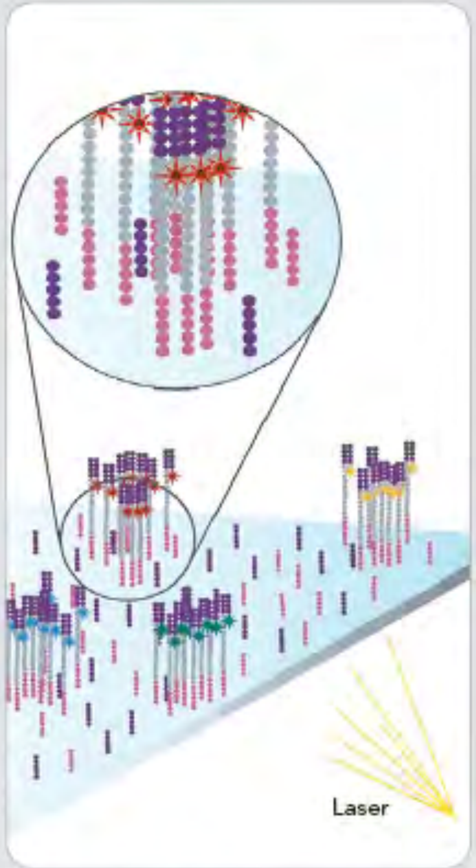


The first sequencing cycle begins by adding four labeled reversible terminators, primers, and DNA polymerase.

8. IMAGE FIRST BASE

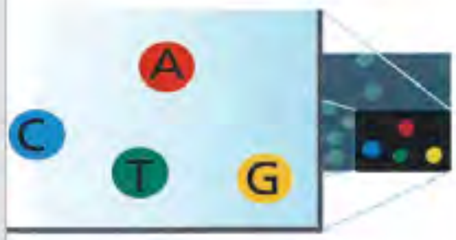
After laser excitation, the emitted fluorescence from each cluster is captured and the first base is identified.

9. DETERMINE SECOND BASE



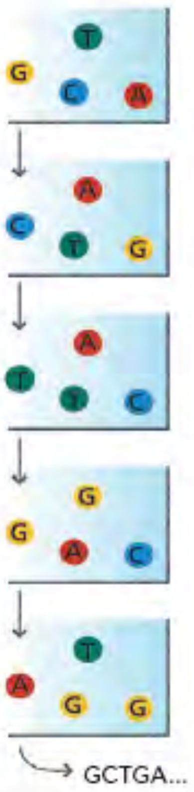
The next cycle repeats the incorporation of four labeled reversible terminators, primers, and DNA polymerase.

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

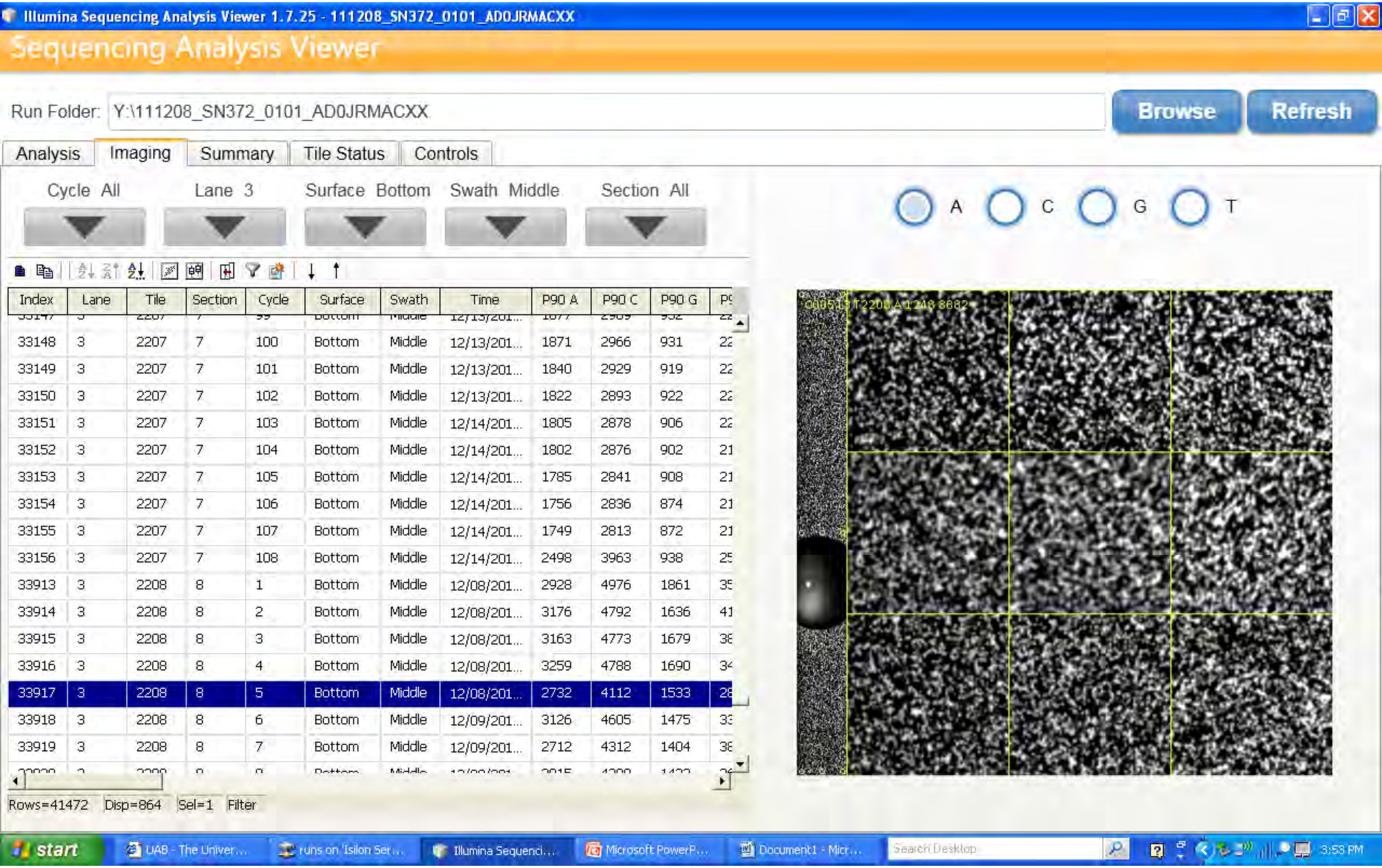


The sequencing cycles are repeated to determine the sequence of bases in a fragment, one base at a time.

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.



Illumina Sequencing Analysis Viewer 1.7.25 - 111208_SN372_0101_AD0JRMACXX

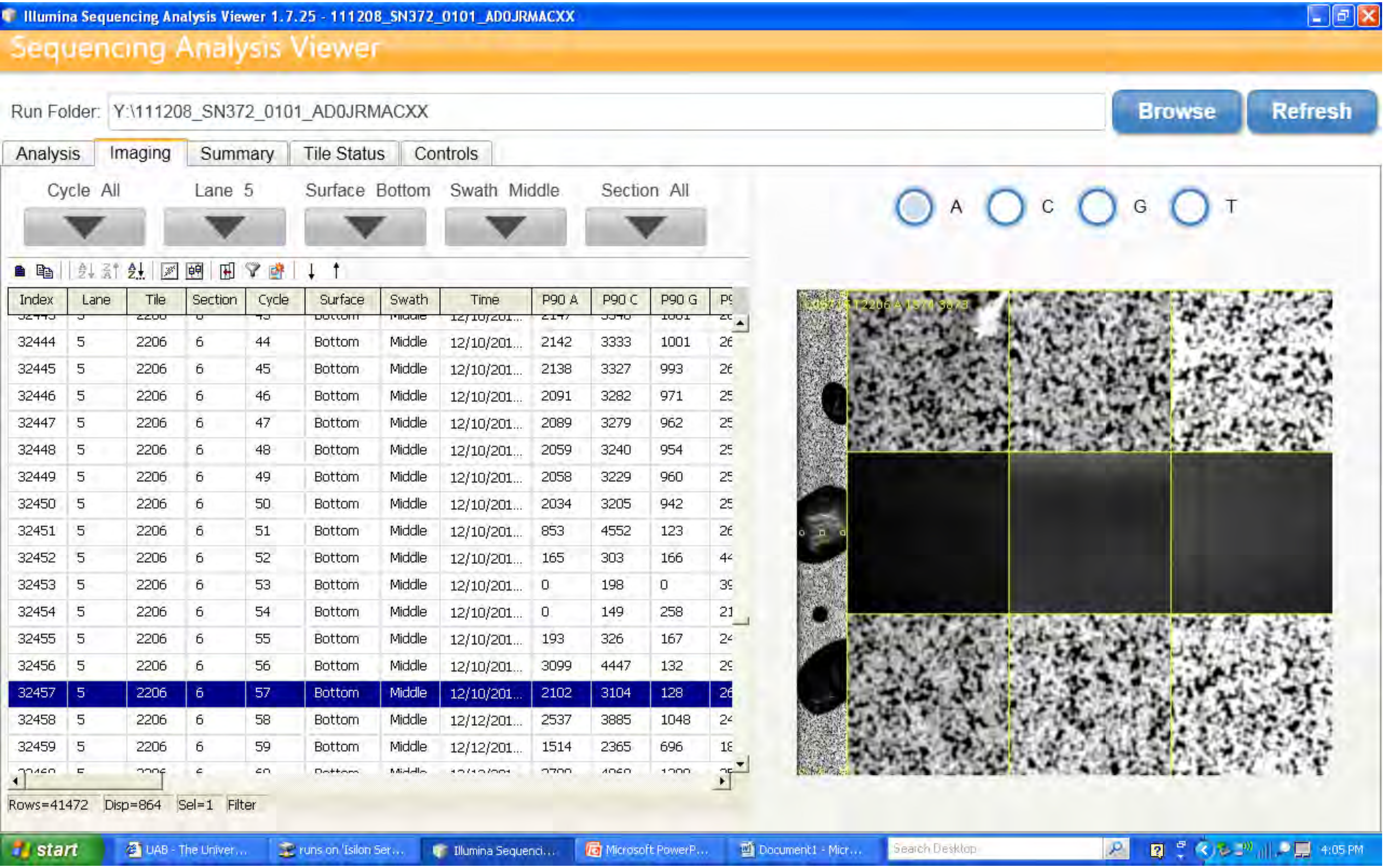
Sequencing Analysis Viewer

Run Folder: Y:\111208_SN372_0101_AD0JRMACXX

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

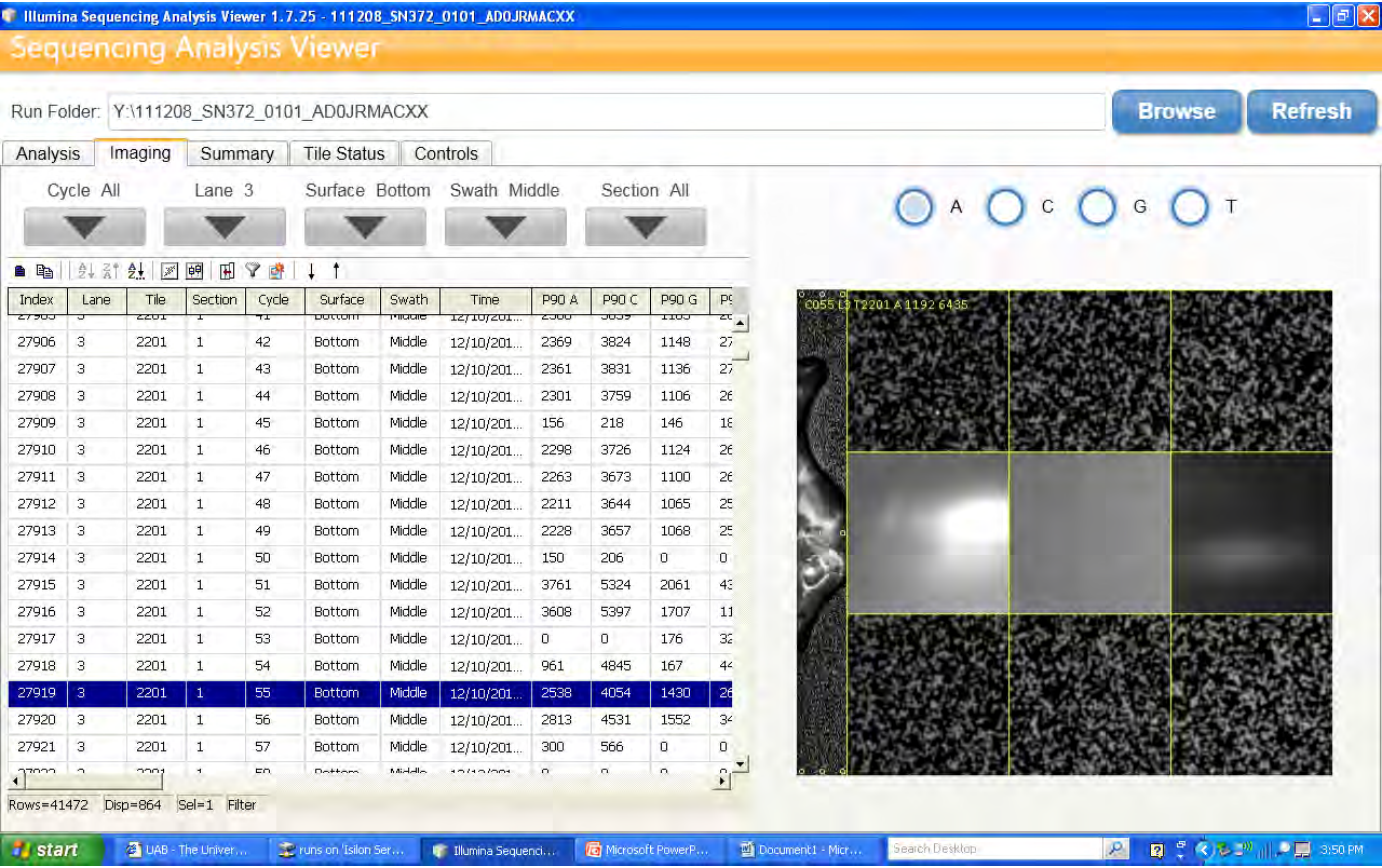


Run Folder: Y:\111208_SN372_0101_AD0JRMACXX

Cycle All
Lane 5
Surface Bottom
Swath Middle
Section All

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

Rows=41472 Disp=864 Sel=1 Filter



Illumina Sequencing Analysis Viewer 1.7.25 - 111208_SN372_0101_AD0JRMACXX

Sequencing Analysis Viewer

Run Folder: Y:\111208_SN372_0101_AD0JRMACXX

Cycle All

Lane 3

Surface Bottom

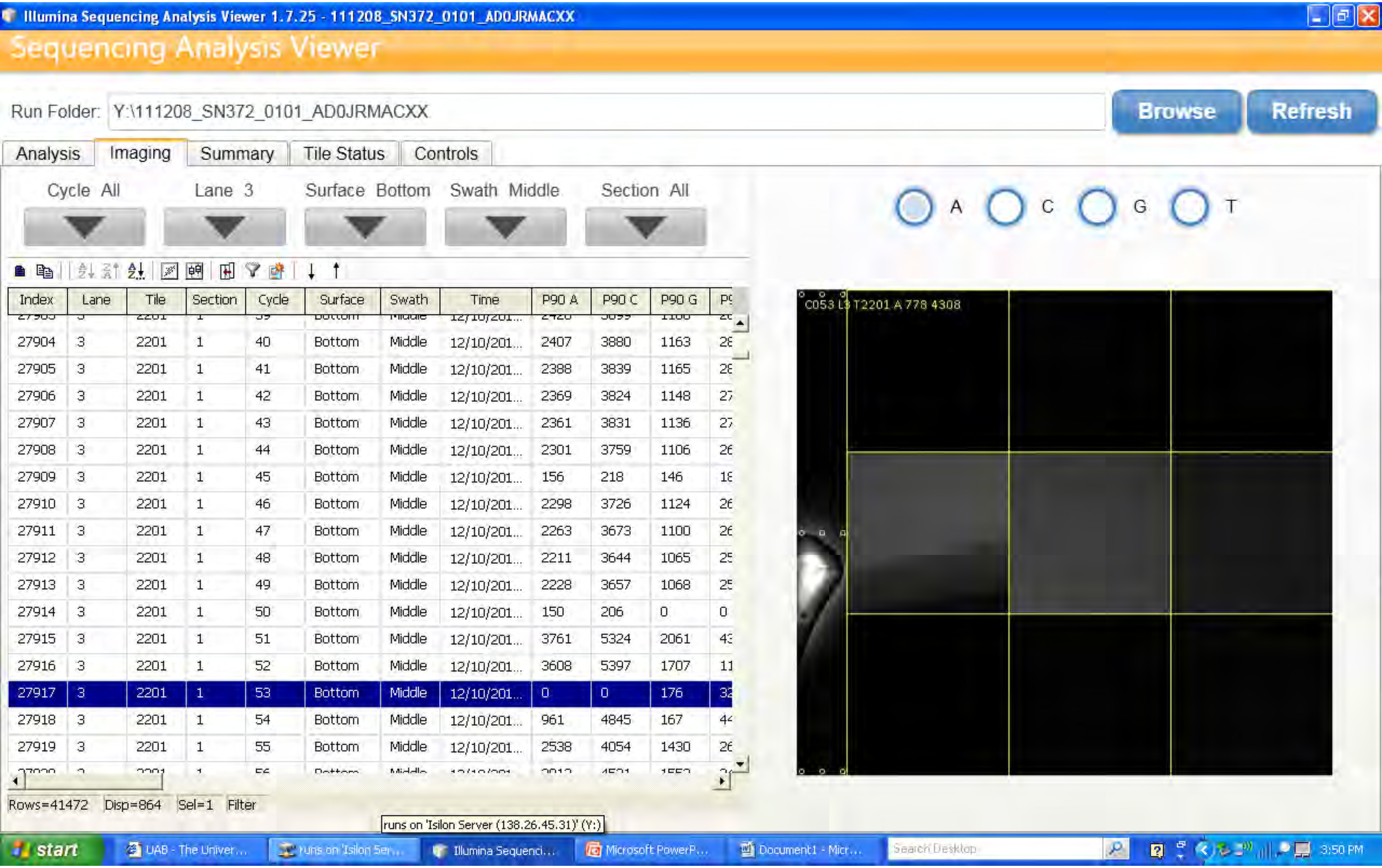
Swath Middle

Section All

○ A ○ C ○ G ○ T

Index	Lane	Tile	Section	Cycle	Surface	Swath	Time	P90 A	P90 C	P90 G	P9
27906	3	2201	1	42	Bottom	Middle	12/10/201...	2369	3824	1148	27
27907	3	2201	1	43	Bottom	Middle	12/10/201...	2361	3831	1136	27
27908	3	2201	1	44	Bottom	Middle	12/10/201...	2301	3759	1106	26
27909	3	2201	1	45	Bottom	Middle	12/10/201...	156	218	146	18
27910	3	2201	1	46	Bottom	Middle	12/10/201...	2298	3726	1124	26
27911	3	2201	1	47	Bottom	Middle	12/10/201...	2263	3673	1100	26
27912	3	2201	1	48	Bottom	Middle	12/10/201...	2211	3644	1065	25
27913	3	2201	1	49	Bottom	Middle	12/10/201...	2228	3657	1068	25
27914	3	2201	1	50	Bottom	Middle	12/10/201...	150	206	0	0
27915	3	2201	1	51	Bottom	Middle	12/10/201...	3761	5324	2061	43
27916	3	2201	1	52	Bottom	Middle	12/10/201...	3608	5397	1707	11
27917	3	2201	1	53	Bottom	Middle	12/10/201...	0	0	176	32
27918	3	2201	1	54	Bottom	Middle	12/10/201...	961	4845	167	44
27919	3	2201	1	55	Bottom	Middle	12/10/201...	2538	4054	1430	26
27920	3	2201	1	56	Bottom	Middle	12/10/201...	2813	4531	1552	34
27921	3	2201	1	57	Bottom	Middle	12/10/201...	300	566	0	0

Rows=41472 Disp=864 Sel=1 Filter



Illumina Sequencing Analysis Viewer 1.7.25 - 111208_SN372_0101_AD0JRMACXX

Sequencing Analysis Viewer

Run Folder: Y:\111208_SN372_0101_AD0JRMACXX

Cycle All

Lane 3

Surface Bottom

Swath Middle

Section All

● A ○ C ○ G ○ T

Index	Lane	Tile	Section	Cycle	Surface	Swath	Time	P90 A	P90 C	P90 G
27904	3	2201	1	40	Bottom	Middle	12/10/201...	2407	3880	1163
27905	3	2201	1	41	Bottom	Middle	12/10/201...	2388	3839	1165
27906	3	2201	1	42	Bottom	Middle	12/10/201...	2369	3824	1148
27907	3	2201	1	43	Bottom	Middle	12/10/201...	2361	3831	1136
27908	3	2201	1	44	Bottom	Middle	12/10/201...	2301	3759	1106
27909	3	2201	1	45	Bottom	Middle	12/10/201...	156	218	146
27910	3	2201	1	46	Bottom	Middle	12/10/201...	2298	3726	1124
27911	3	2201	1	47	Bottom	Middle	12/10/201...	2263	3673	1100
27912	3	2201	1	48	Bottom	Middle	12/10/201...	2211	3644	1065
27913	3	2201	1	49	Bottom	Middle	12/10/201...	2228	3657	1068
27914	3	2201	1	50	Bottom	Middle	12/10/201...	150	206	0
27915	3	2201	1	51	Bottom	Middle	12/10/201...	3761	5324	2061
27916	3	2201	1	52	Bottom	Middle	12/10/201...	3608	5397	1707
27917	3	2201	1	53	Bottom	Middle	12/10/201...	0	0	176
27918	3	2201	1	54	Bottom	Middle	12/10/201...	961	4845	167
27919	3	2201	1	55	Bottom	Middle	12/10/201...	2538	4054	1430

Rows=41472 Disp=864 Sel=1 Filter

C053 L3 T2201 A 778 4308

runs on 'Isilon Server (138.26.45.31)' (Y:)