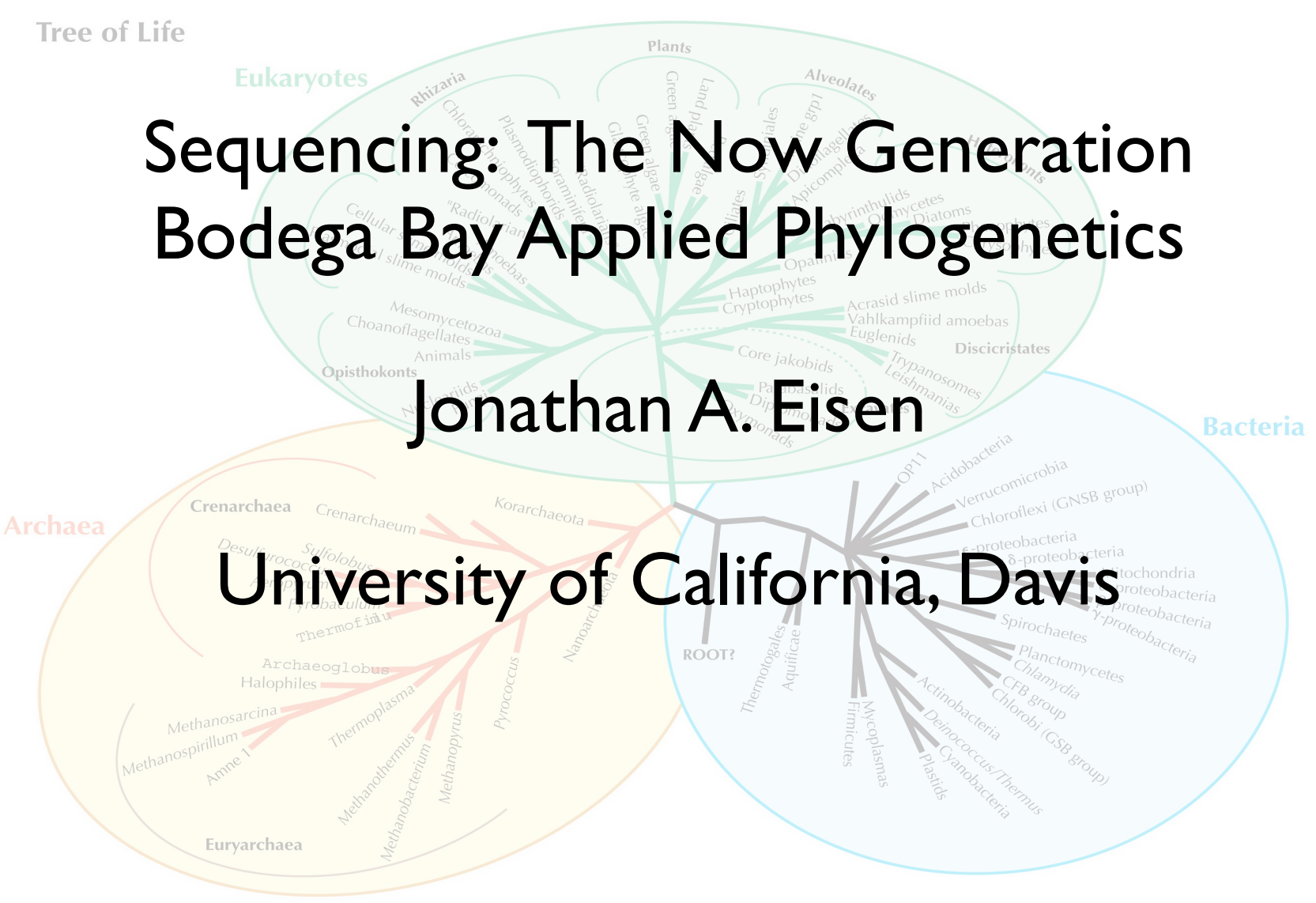
The Now Generation



Sequencing: The Now Generation
Bodega Bay Applied Phylogenetics

Jonathan A. Eisen

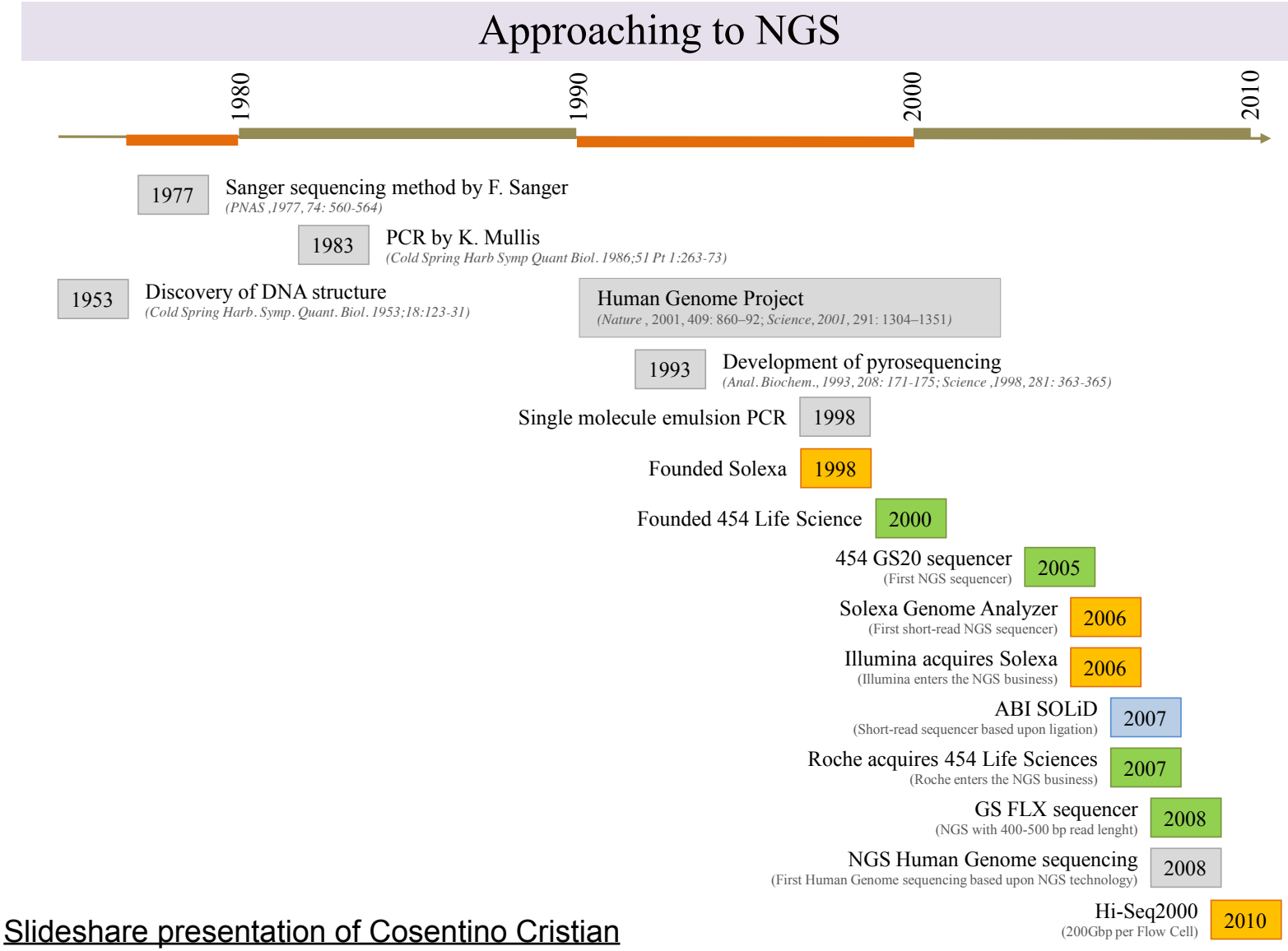
University of California, Davis

Sequencing Technology

Key Issues

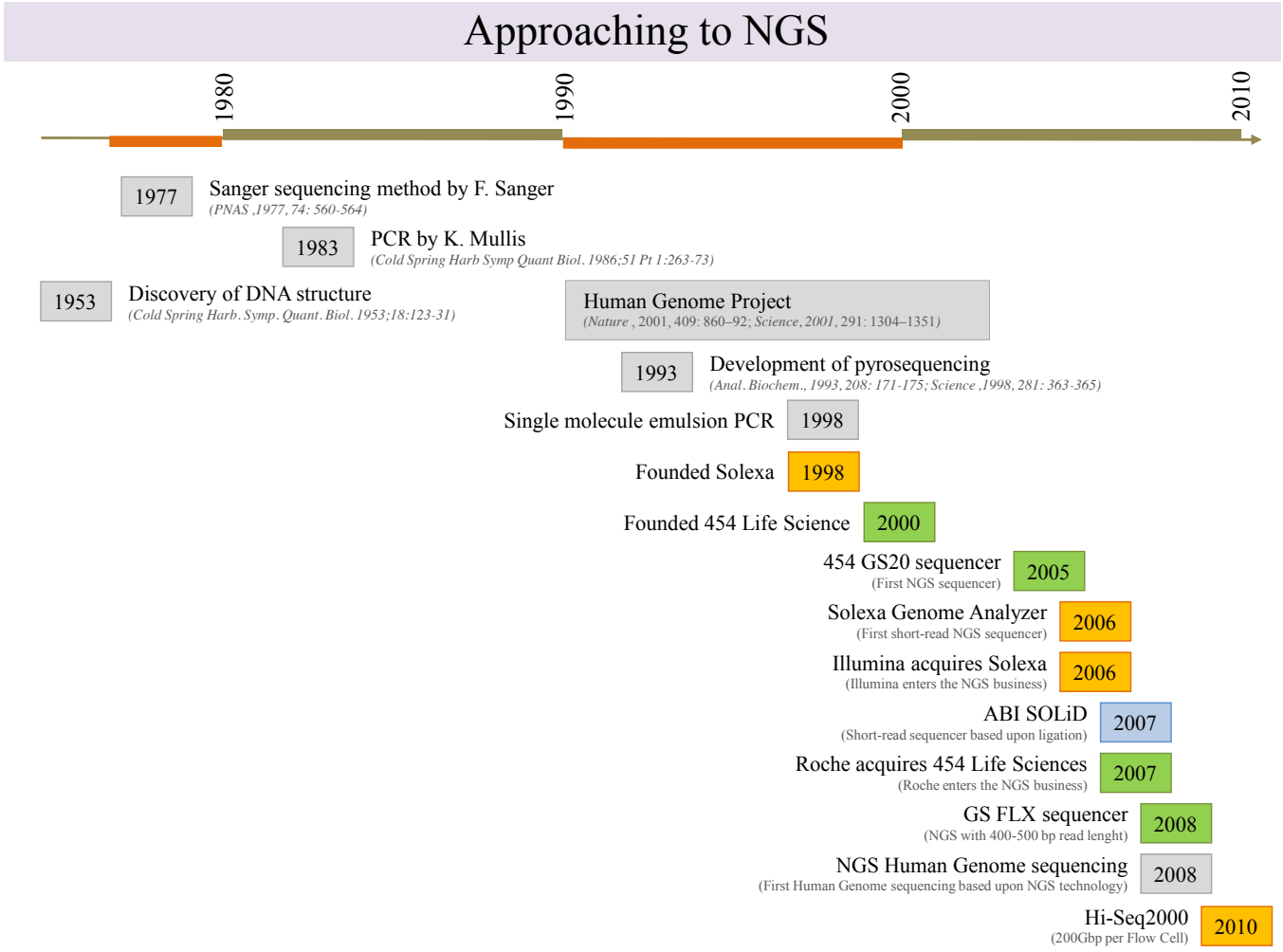
- Cost / bp
- Read length
- Paired end
- Ease of feeding
- Error profiles
- Barcoding potential

Timeline



From Slideshare presentation of Cosentino Cristian
http://www.slideshare.net/cosentia/high-throughput-equencing

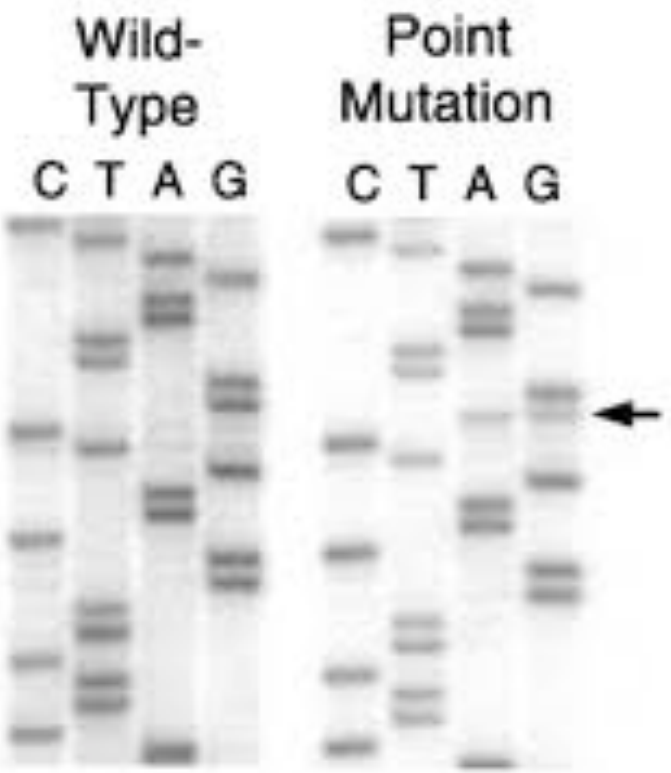
Timeline

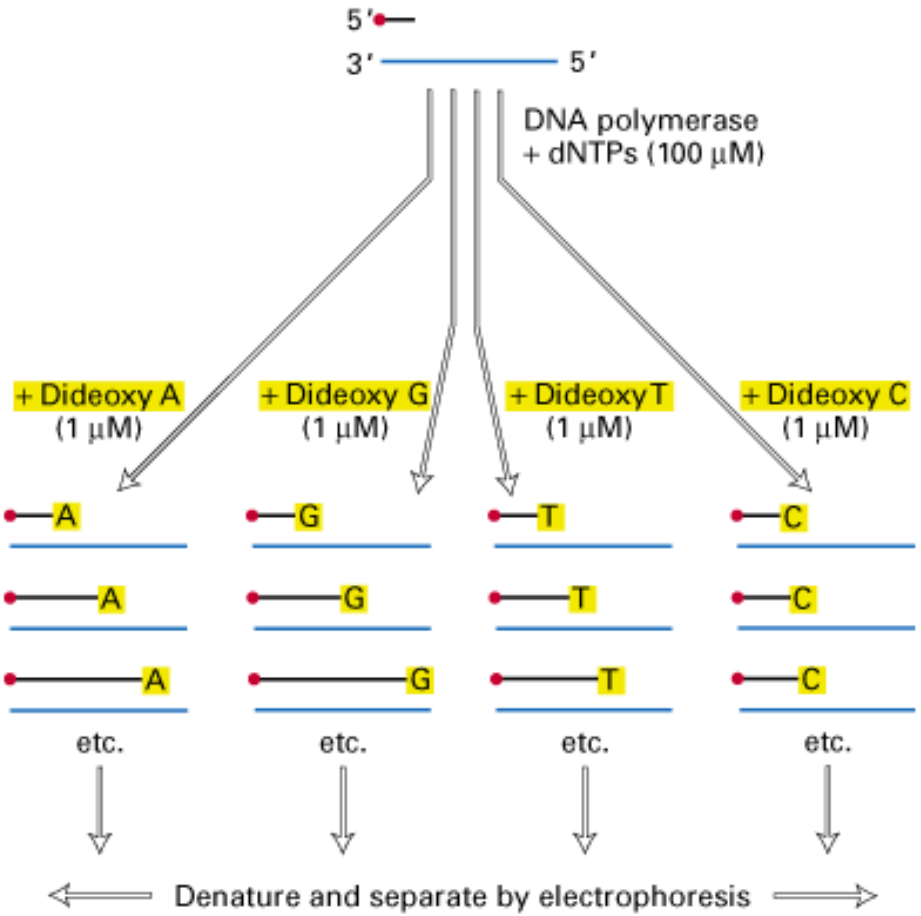


Miseq
Roche Jr
Ion Torrent
PacBio
Oxford

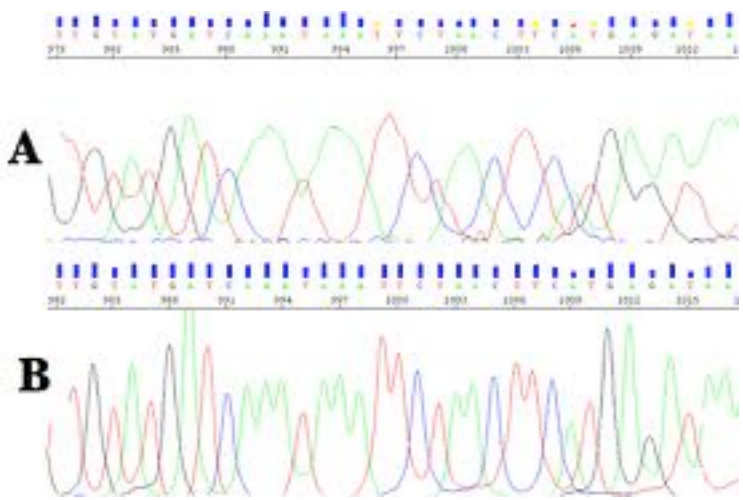
From Slideshare presentation of Cosentino Cristian
http://www.slideshare.net/cosentia/high-throughput-equencing

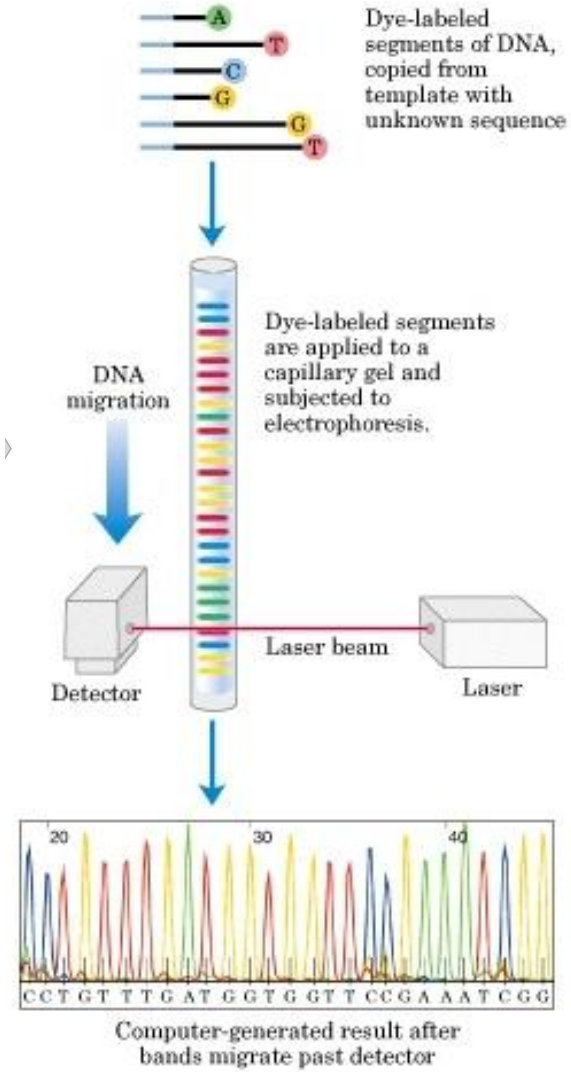
Generation I: Manual Sanger



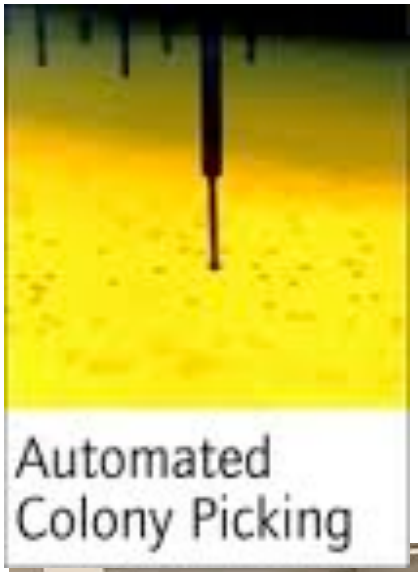


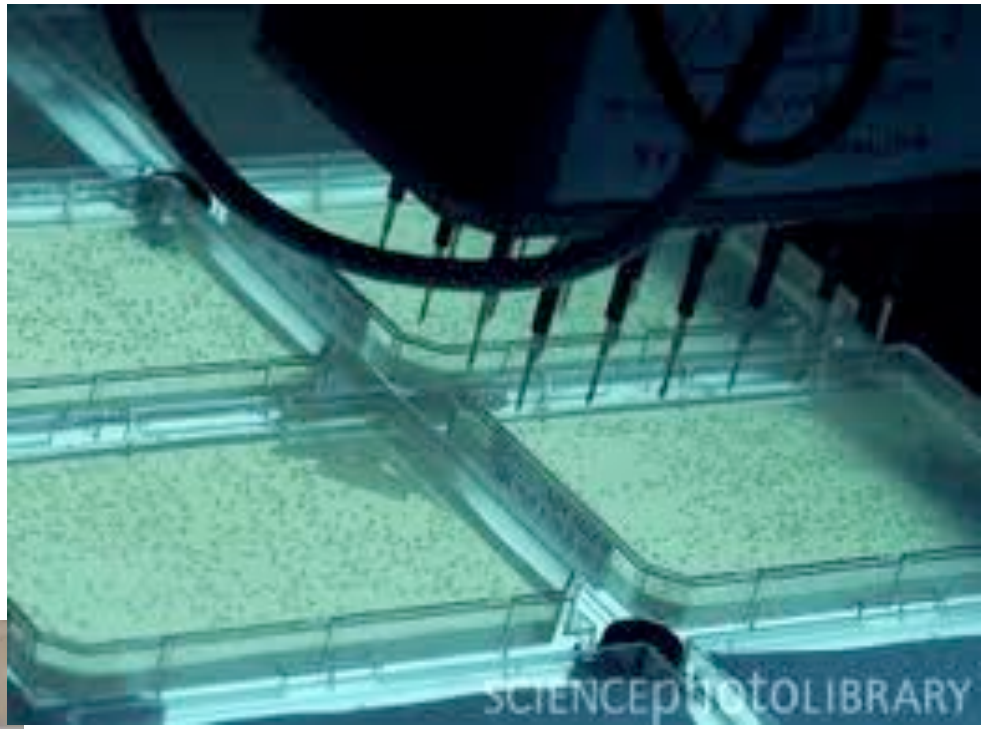
Generation II: Automation





Automation 2





Automated Sanger Highlights

- 1991: ESTs by Venter
- 1995: Haemophilus influenzae genome
- 1996: Yeast, archaea
- 1999: Drosophila genome
- 2000: Arabidopsis genome
- 2000: Human genome
- 2004: Shotgun metagenomics

Generation III: Clusters not clones

illumina

Roche 454 SEQUENCING

Applied Biosystems

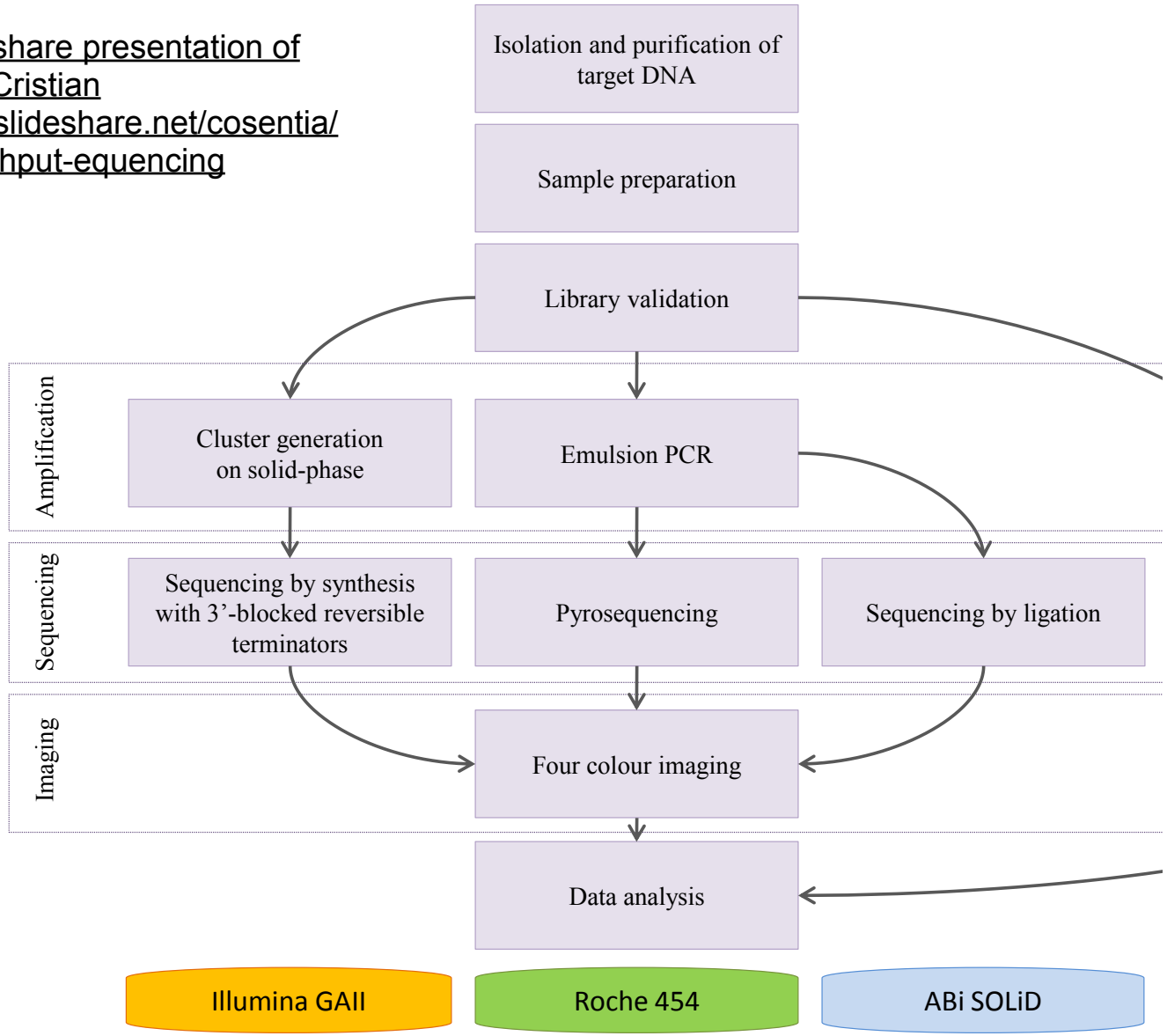
Next-Generation DNA Sequencing Methods

Elaine R. Mardis

Departments of Genetics and Molecular Microbiology and Genome Sequencing Center, Washington University School of Medicine, St. Louis MO 63108; email: emardis@wustl.edu

Annu. Rev. Genomics Hum. Genet. 2008. 9:387–402

From Slideshare presentation of Cosentino Cristian http://www.slideshare.net/cosentia/high-throughput-equencing



454

454

From Slideshare presentation of Cosentino Cristian
http://www.slideshare.net/cosentia/high-throughput-equencing

Roche 454

From Slideshare presentation of Cosentino Cristian
http://www.slideshare.net/cosentia/high-throughput-equencing

454 -> Roche

- 1st “Next-generation” sequencing system to become commercially available, in 2004
- Uses pyrosequencing
  - Polymerase incorporates nucleotide
  - Pyrophosphate released
  - Eventually light from luciferase released
- Three main steps in 454 method
  - Library prep
  - Emulsion PCR
  - Sequencing

Workflow

Sample Fragmentation

Library Preparation

emPCR Setup

emPCR Amplification

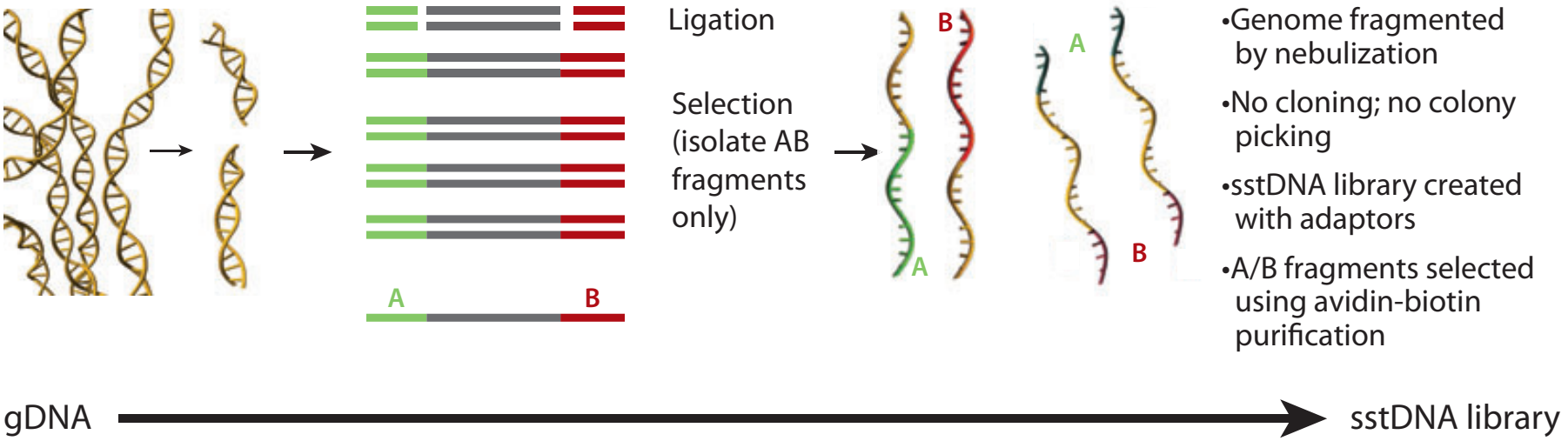
Pyrosequencing

Data Analysis

454 Step1: Library Prep

DNA library preparation

4.5 hours



gDNA fragmented by nebulization or sonication

Fragments are end-repaired and ligated to adaptors containing universal priming sites

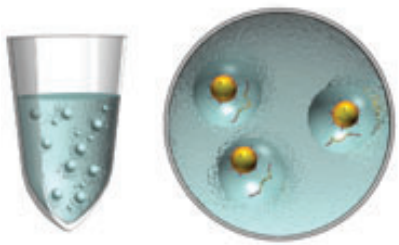
Fragments are denatured and AB ssDNA are selected by avidin/biotin purification (ssDNA library)

From Mardis 2008. Annual Rev. Genetics 9: 387.

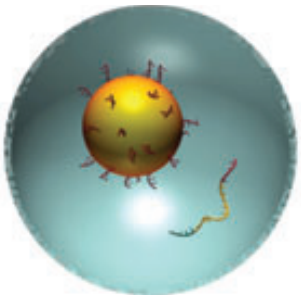
454 Step 2: Emulsion PCR

Emulsion PCR

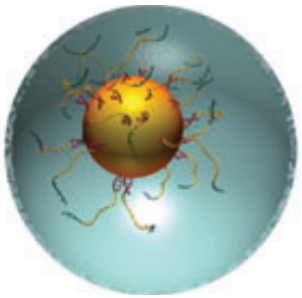
8 hours



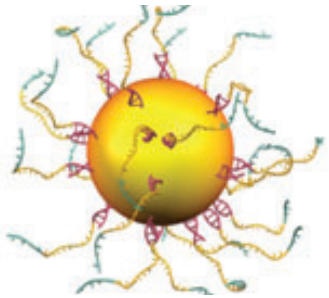
Anneal sstDNA to an excess of DNA capture beads



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



Clonal amplification occurs inside microreactors



Break microreactors and enrich for DNA-positive beads

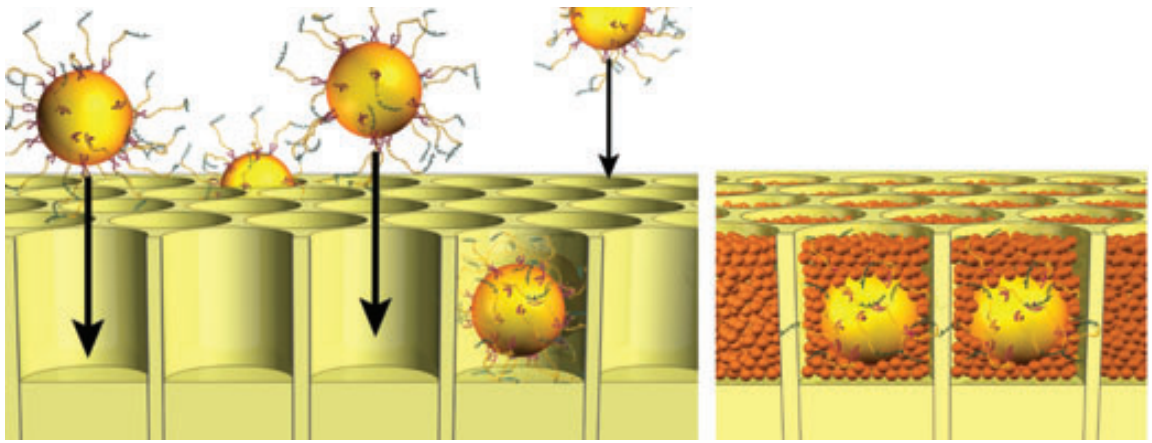
sstDNA library → Bead-amplified sstDNA library

From Mardis 2008. Annual Rev. Genetics 9: 387.

454 Step 3: Sequencing

Sequencing

7.5 hours

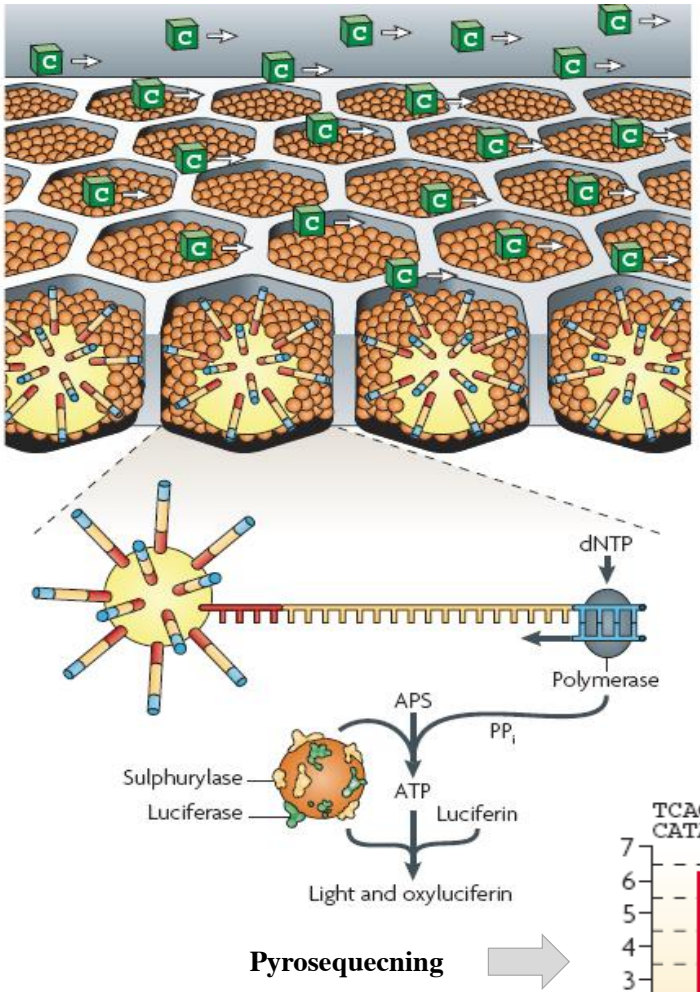


- Well diameter: average of 44 µm
- 400,000 reads obtained in parallel
- A single cloned amplified sstDNA bead is deposited per well

Amplified sstDNA library beads → Quality filtered bases

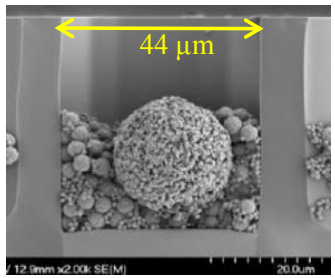
From Mardis 2008. Annual Rev. Genetics 9: 387.

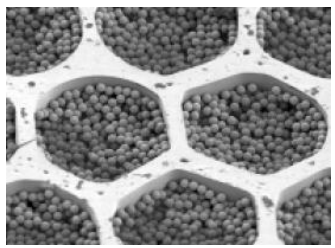
454 Step 3: Sequencing

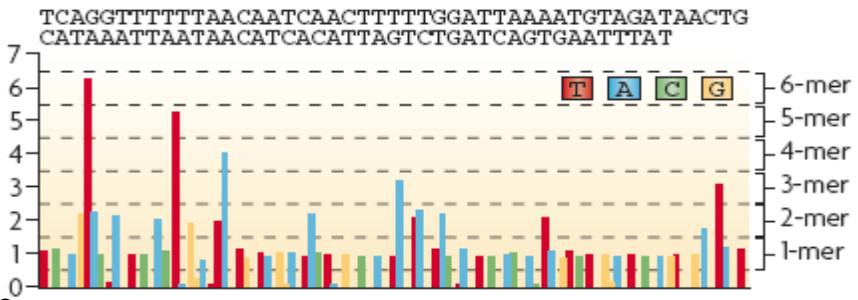


Pyrosequecning

Annu. Rev. Genomics Hum. Genet., 2008, 9: 387-402
Nature Reviews genetics, 2010, 11: 31-46







Reads are recorded as flowgrams

From Slideshare presentation of Cosentino Cristian
http://www.slideshare.net/cosentia/high-throughput-equencing

454 Key Issues

- Number of repeated nucleotides estimated by amount of light ... many errors
- Reasonable number of failures in EM-PCR and other steps
- Systems
  - GS20
  - FLX
  - FLX Titanium
  - FLX Titanium XL
  - Junior

454 Platform Updates

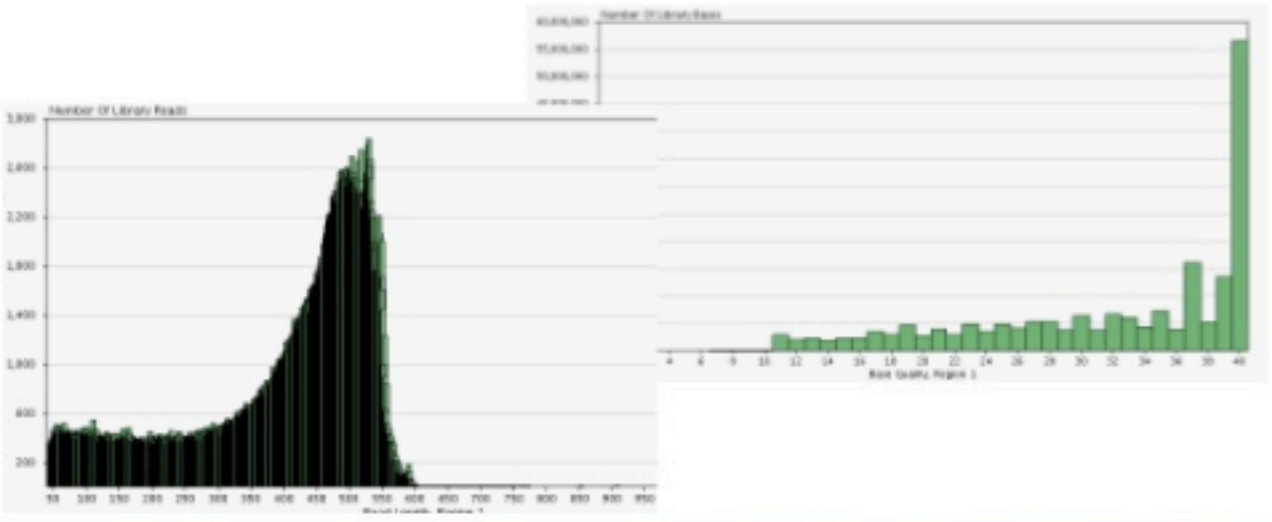
GS20	• 100bp reads, ~20Mbp / run
GS-FLX	• 250bp reads ~100 Mbp / run (7.5 hrs)
GS-FLX Titanium	• 400bp reads ~400 Mbp / run (10 hrs)
GS-FLX Titanium Plus	• 700 bp reads ~700 Mbp/run (18 hrs)
GS Junior	• 400 bp reads ~ 35Mbp/run (10 hrs)

454/Roche – Performance Specs

	GS FLX+	GS Jr.
Mb/run	700	35
Run time	23 hours	10 hours
Mb/day	700	35
Read length	Up to 1kb	~400b
# of single reads/run	1M	0.1M
Instrument cost	~$500k	$125k
Run cost	~$6k	~$1k

454 Sequencing Output

- *.sff *(standard flowgram format)*
- *.fna *(fasta)*
- *.qual *(Phred quality scores)*



Solexa

Solexa

From Slideshare presentation of Cosentino Cristian
http://www.slideshare.net/cosentia/high-throughput-equencing

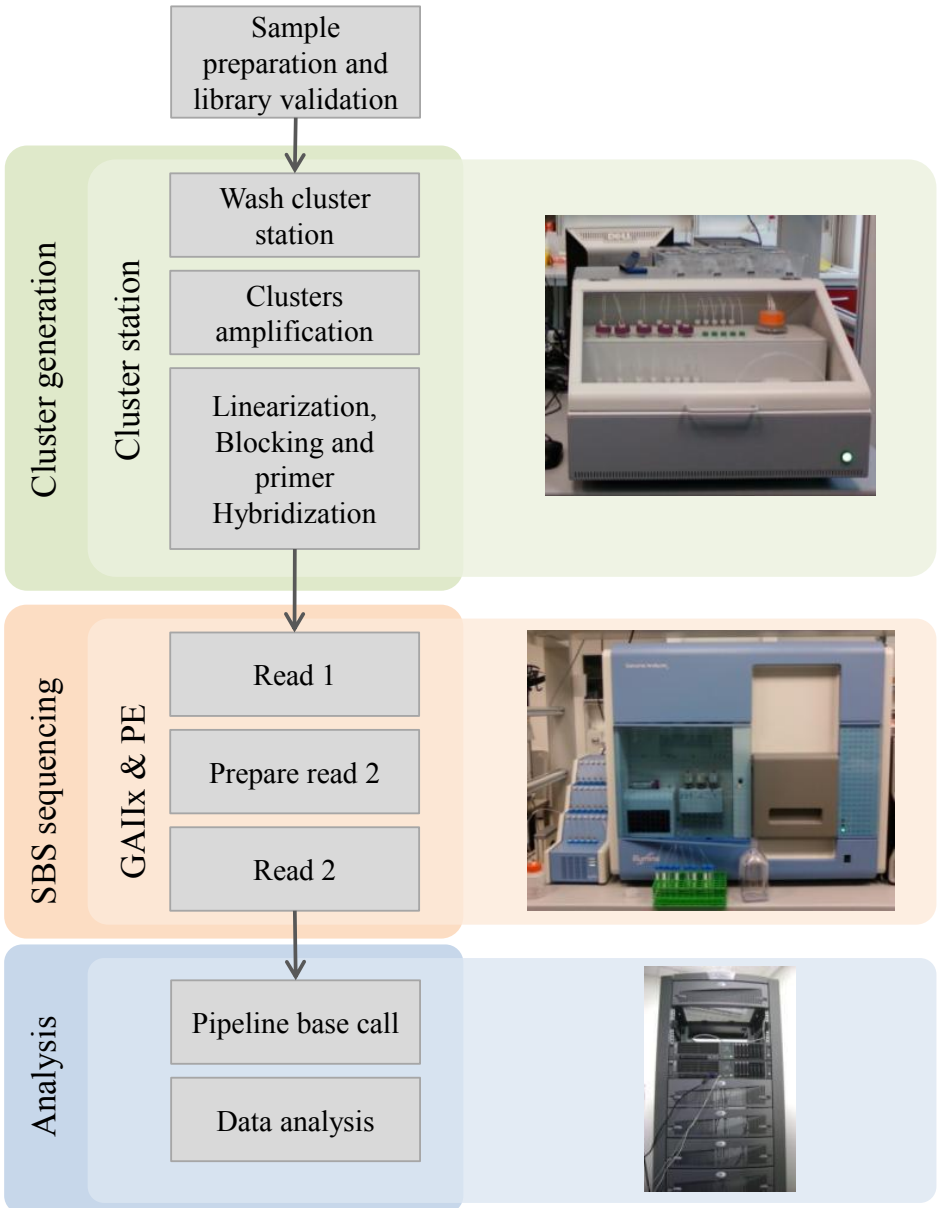
Illumina

From Slideshare presentation of Cosentino Cristian
http://www.slideshare.net/cosentia/high-throughput-equencing

Illumina

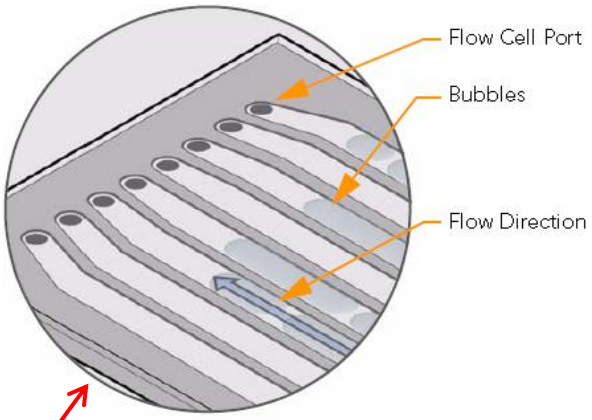
- Sequencing by synthesis
- 1st system released in 2006 by Solexa
- Acquired by Illumina in 2007
- Systems
  - GA
  - GAII
  - HiSeqs
  - HiScan
  - MiSeq

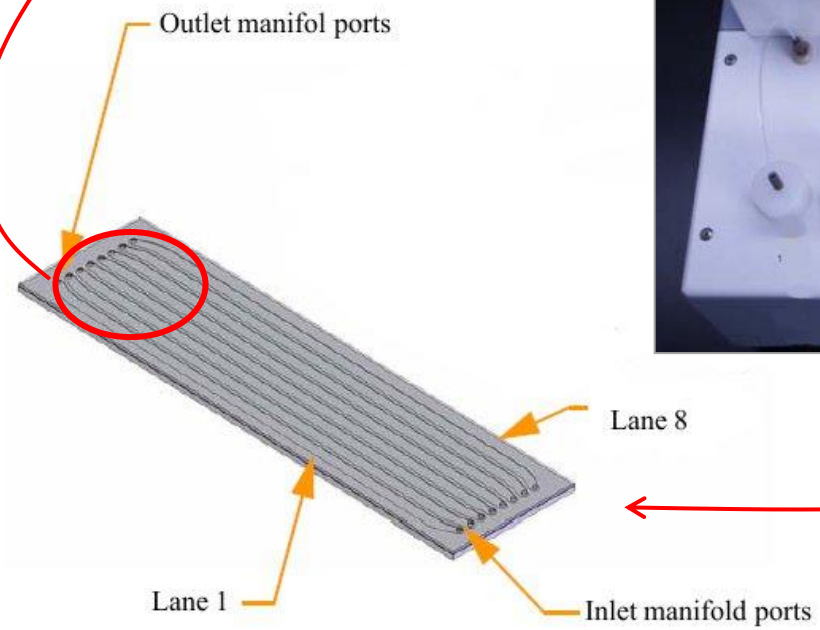
Illumina Outline



From Slideshare presentation of Cosentino Cristian http://www.slideshare.net/cosentia/high-throughput-equencing

Illumina Flow Cell



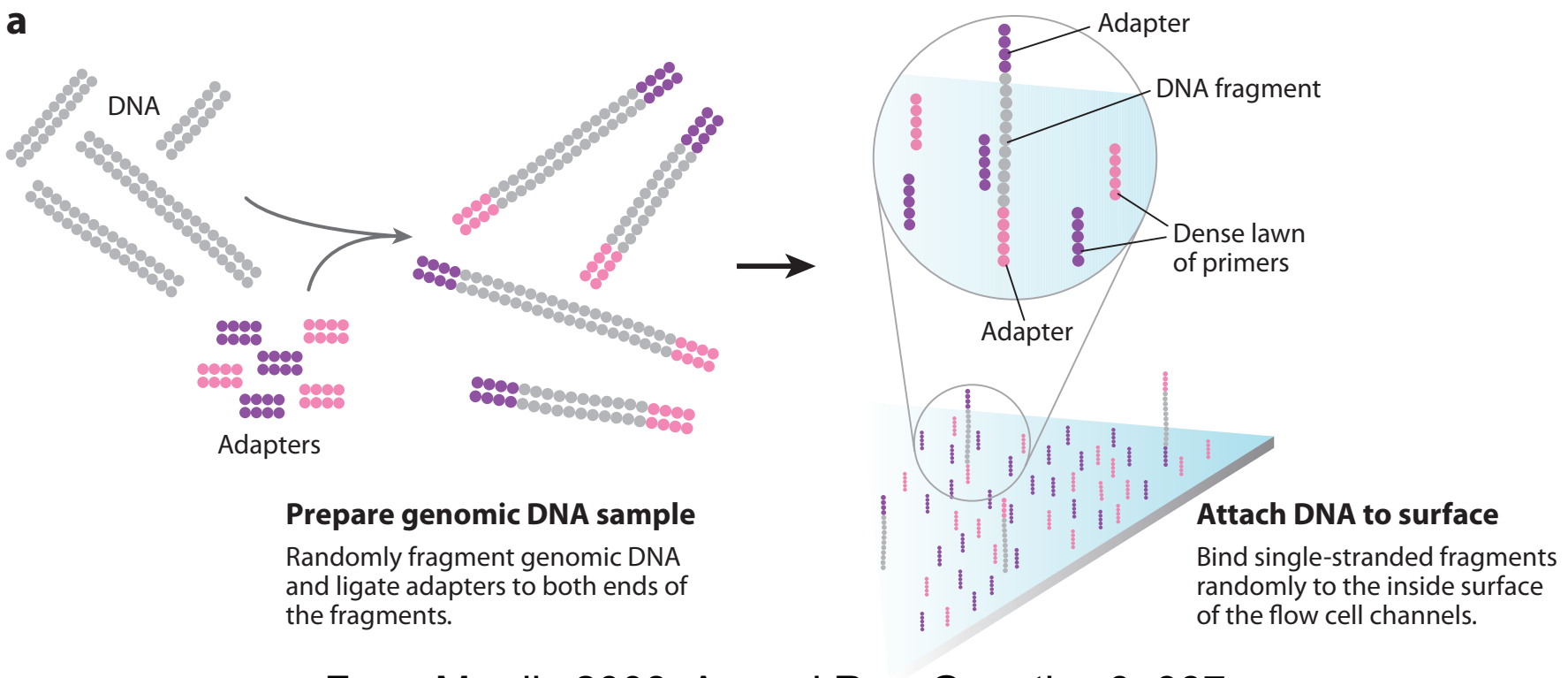


From Slideshare presentation of Cosentino Cristian http://www.slideshare.net/cosentia/high-throughput-equencing

Illumina Prep

- **Flow cell prep:**
  - **Coat flow cell with primers for bridge PCR**
- **DNA prep**
  - **Fragment DNA to size of interest**
  - **Add adapters to DNA of interest**
  - **Clean up fragments**

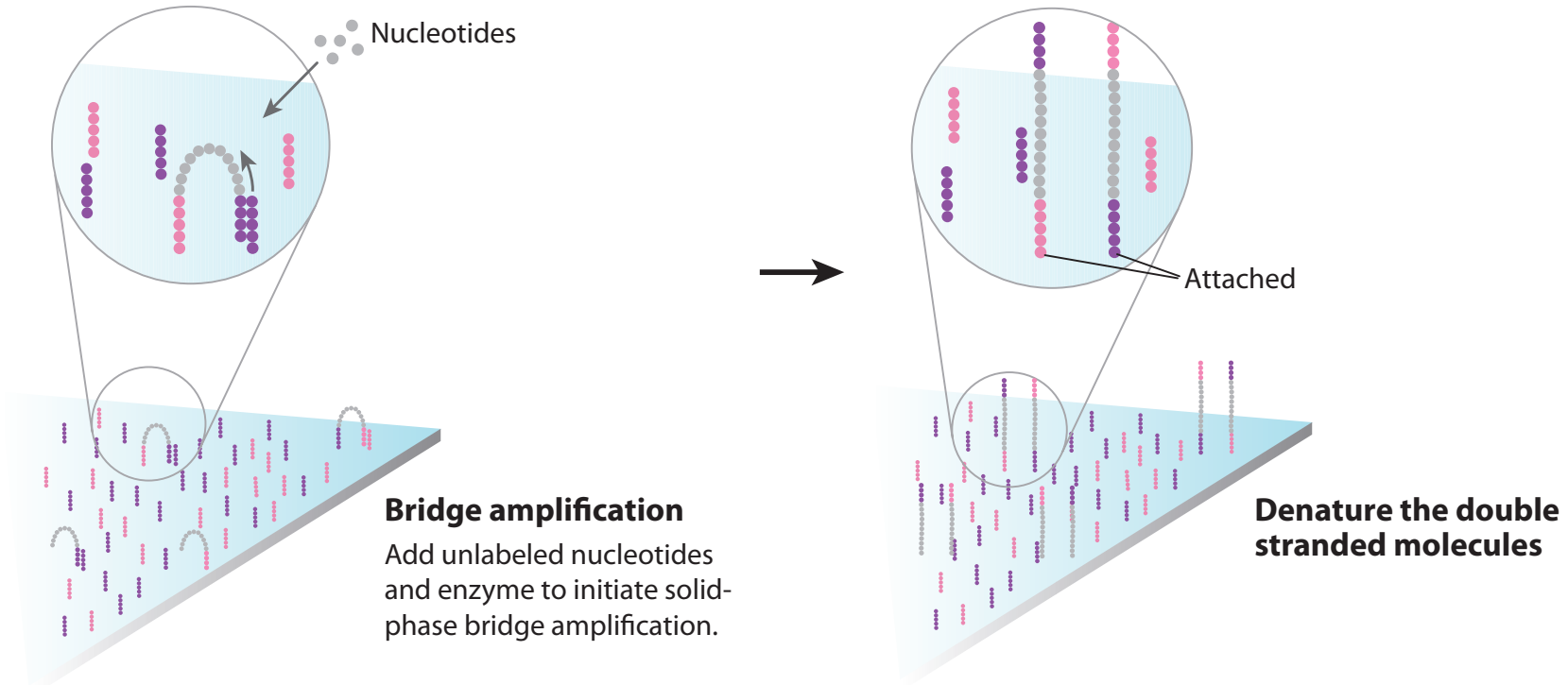
Illumina Step 1: Attach DNA



From Mardis 2008. Annual Rev. Genetics 9: 387.

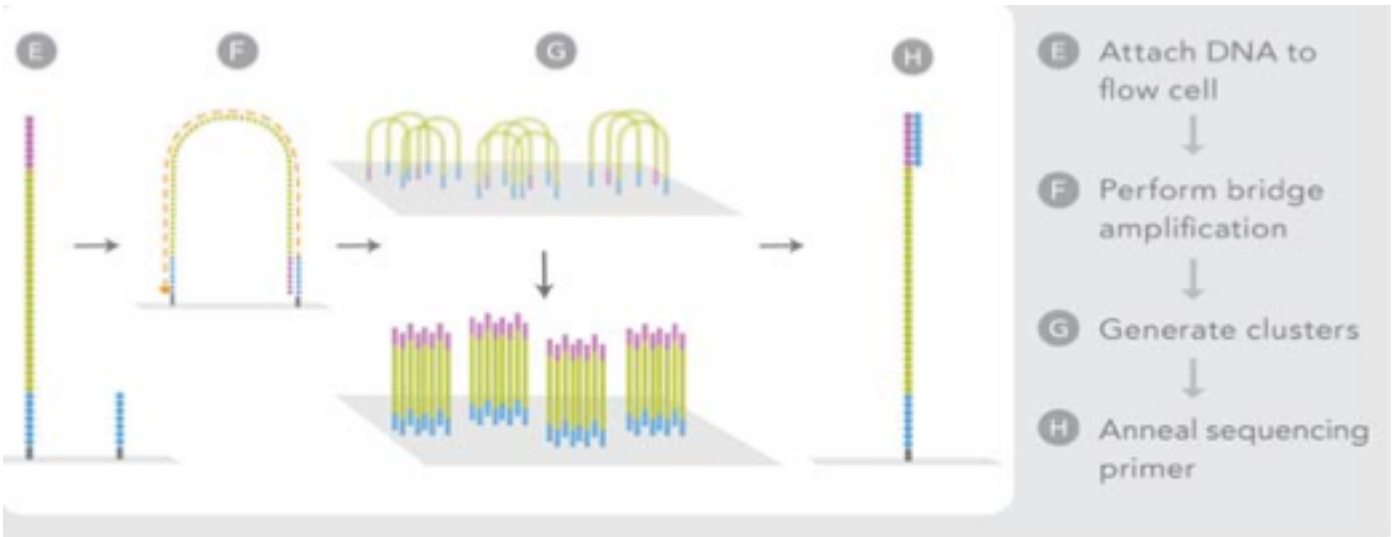
The Illumina sequencing-by-synthesis approach. Cluster strands created by bridge amplification are primed and all four fluorescently labeled, 3′-OH blocked nucleotides are added to the flow cell with DNA polymerase. The cluster strands are extended by one nucleotide. Following the incorporation step, the unused nucleotides and DNA polymerase molecules are washed away, a scan buffer is added to the flow cell, and the optics system scans each lane of the flow cell by imaging units called tiles. Once imaging is completed, chemicals that effect cleavage of the fluorescent labels and the 3′-OH blocking groups are added to the flow cell, which prepares the cluster strands for another round of fluorescent nucleotide incorporation

Illumina Step 2: Bridge PCR

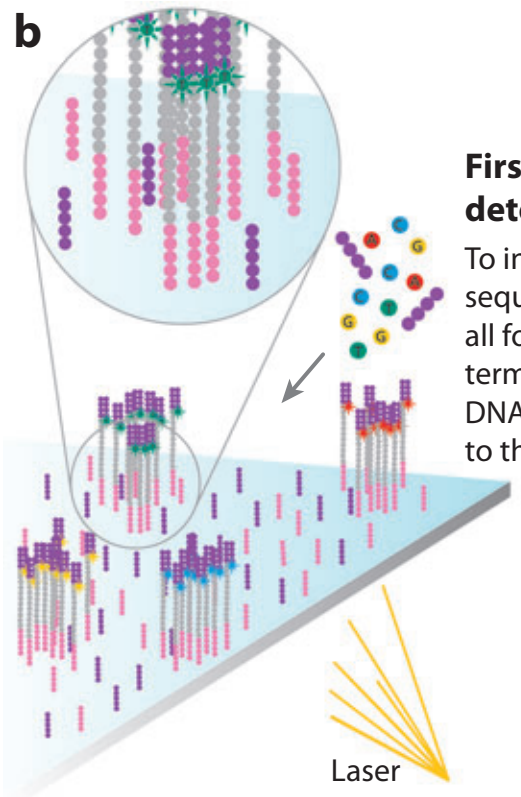


From Mardis 2008. Annual Rev. Genetics 9: 387.

The Illumina sequencing-by-synthesis approach. Cluster strands created by bridge amplification are primed and all four fluorescently labeled, 3′-OH blocked nucleotides are added to the flow cell with DNA polymerase. The cluster strands are extended by one nucleotide. Following the incorporation step, the unused nucleotides and DNA polymerase molecules are washed away, a scan buffer is added to the flow cell, and the optics system scans each lane of the flow cell by imaging units called tiles. Once imaging is completed, chemicals that effect cleavage of the fluorescent labels and the 3′-OH blocking groups are added to the flow cell, which prepares the cluster strands for another round of fluorescent nucleotide incorporation


E
F
G
H
E Attach DNA to flow cell
F Perform bridge amplification
G Generate clusters
H Anneal sequencing primer

Illumina Step 3: Sequencing



First chemistry cycle: determine first base

To initiate the first sequencing cycle, add all four labeled reversible terminators, primers, and DNA polymerase enzyme to the flow cell.

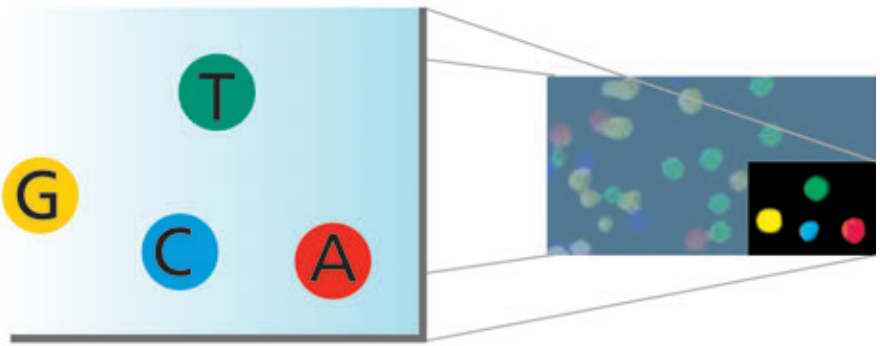


Image of first chemistry cycle

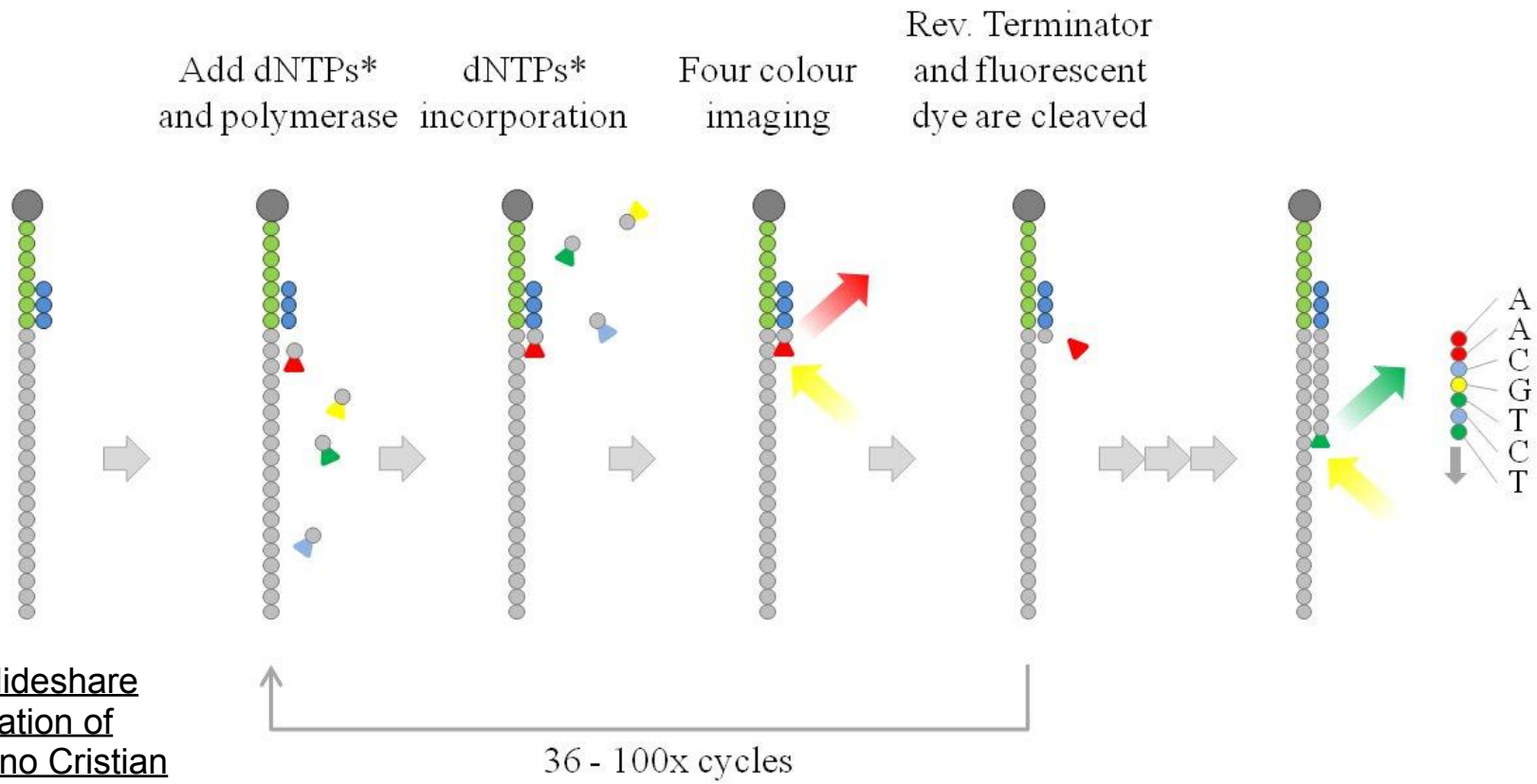
After laser excitation, capture the image of emitted fluorescence from each cluster on the flow cell. Record the identity of the first base for each cluster.

Before initiating the next chemistry cycle

The blocked 3' terminus and the fluorophore from each incorporated base are removed.

From Mardis 2008. Annual Rev. Genetics 9: 387.

Illumina SBS



From Slideshare presentation of Cosentino Cristian http://www.slideshare.net/cosentia/high-throughput-equencing

HiSeq 2500/1500

	HiSeq 2500		HiSeq 1500	
Run Mode	High Output	Rapid Output*	High Output	Rapid Output*
Output (2 × 100 bp)	600 Gb	120 Gb	300 Gb	60 Gb
Run Time	~11 days	~27 hours	~8.5 days	~27 hours
Cluster Generation	cBot	On board	cBot	On board
Paired-end Reads	6 Billion	1.2 Billion	3 Billion	600 Million
Single Reads	3 Billion	600 Million	1.5 Billion	300 Million
Maximum Read Length**	2 × 100 bp	2 × 150 bp	2 × 100 bp	2 × 150 bp
Quality Scores***	> 85% (2 x 50 bp) > 80% (2 x 100 bp)			

From Illumina Web Site

MiSeq

Approximate Run Duration and Output

Cluster Generation and Sequencing

READ LENGTH	TOTAL TIME FROM PREPPED LIBRARY THROUGH SEQUENCING*	OUTPUT**
1 × 36 bp	3.2-3.5 hours	440-550 Mb
2 × 25 bp	4.6-5.0 hours	640-800 Mb
2 × 100 bp	14.0-16.0 hours	2.5-3.1 Gb
2 × 150 bp	20.7-24.0 hours	3.7-4.6 Gb
2 × 250 bp	>35 hours	6.0-7.0 Gb

*Includes paired-end read, if applicable.
** Performance, output, amplification, and sequencing time for 2 x 250 bp read length depends on instrument upgrade, commerically available in the third quarter of 2012. Customers will be notified of upgrade availability. Upgrade dates subject to change.

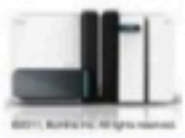
On-Instrument Analysis

READ LENGTH	APPROXIMATE DURATION
Paired-End 150 Cycles (2 × 150 bp)	Less than 2 hours for all supported analysis workflows, except 16s metagenomics. 16s metagenomics completes in less than 24 hours.

From Illumina Web Site

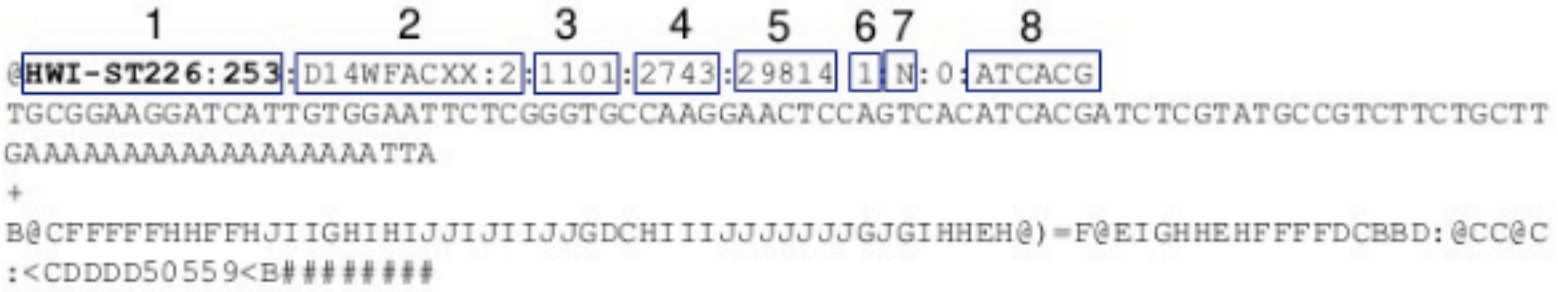
Illumina – Performance Specs




	HiSeq 2000/2500	HiSeq 1000/1500	GAIIx	HiScanSQ	MiSeq
Gb/run	600/120	300/60	95	150	7
Run time	11d/27hrs	8.5d/27hrs	14d	8.5d	35hrs
Gb/day	55/120	35/60	~7	18	~7
Read length	2x100/150	2x100/150	2x150	2x100	2x250
# of single reads/run	3B/600M	1.5B/300M	320M	750M	15M
Instrument cost	$690k/$740k	$590k/$640k	~$300k	~$400k	$125k
Run cost	~$23k	~$11k	~$17k	~$11k	~$1k

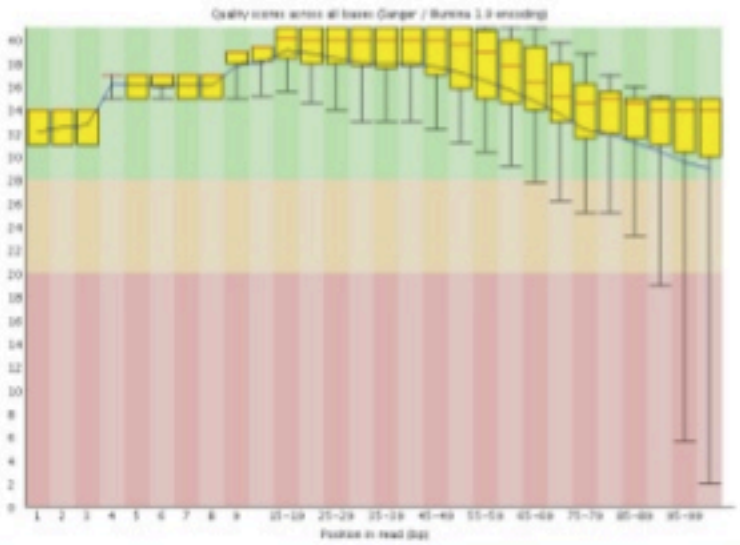
Illumina fastq



1. unique instrument ID and run ID
2. Flow cell ID and lane
3. tile number within the flow cell lane
4. 'x'-coordinate of the cluster within the tile
5. 'y'-coordinate of the cluster within the tile
6. the member of a pair, /1 or /2 *(paired-end or mate-pair reads only)*
7. N if the read passes filter, Y if read fails filter otherwise
8. Index sequence

Illumina Sequencing Output

- *.fastq *(sequence and corresponding quality score encoded with an ASCII character, phred-like quality score + 33)*



Platform Updates

Solexa 1G	• 18bp reads, ~1Gbp / run
Illumina GA	• 36bp reads ~3Gbp / run
Illumina GAII	• 75bp paired reads ~10Gbp / run (8 days)
Illumina GAIIx	• 75bp paired reads ~40Gbp / run (8 days)
Illumina HiSeq 2000	• 100 bp paired reads ~200 Gbp/ run (10 days)
Illumina HiSeq, v3 SBS	• 100bp paired reads ~600Gbp / run (12 days)
MiSeq	• 150 paired reads ~1.5 Gb/run (27 hrs)

Maximum yield / day 50,Gbp
~16x the human genome

Cluster Generation and Sequencing

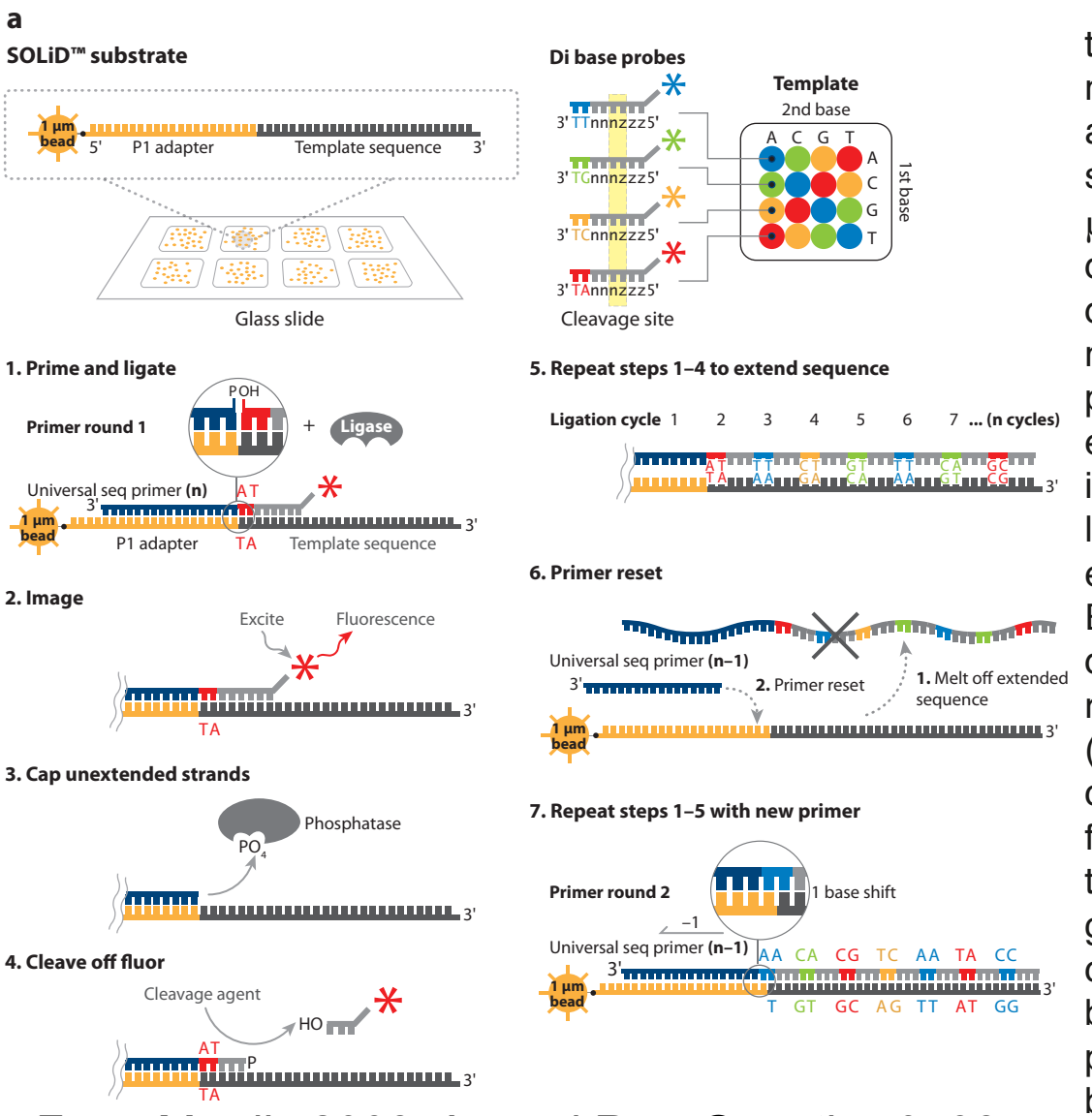
READ LENGTH (BP)	TOTAL TIME*	OUTPUT
1 × 36	~4 hrs	540-610 Mb
2 × 25	~5.5 hrs	750-850 Mb
2 × 100	~16.5 hrs	3.0-3.4 Gb
2 × 150	~24 hrs	4.5-5.1 Gb
2 × 250	~39 hrs	7.5-8.5 Gb
2 × 300**	> 48 hrs	~15 Gb

ABI Solid

ABI Solid

From Slideshare presentation of Cosentino Cristian http://www.slideshare.net/cosentia/high-throughput-equencing

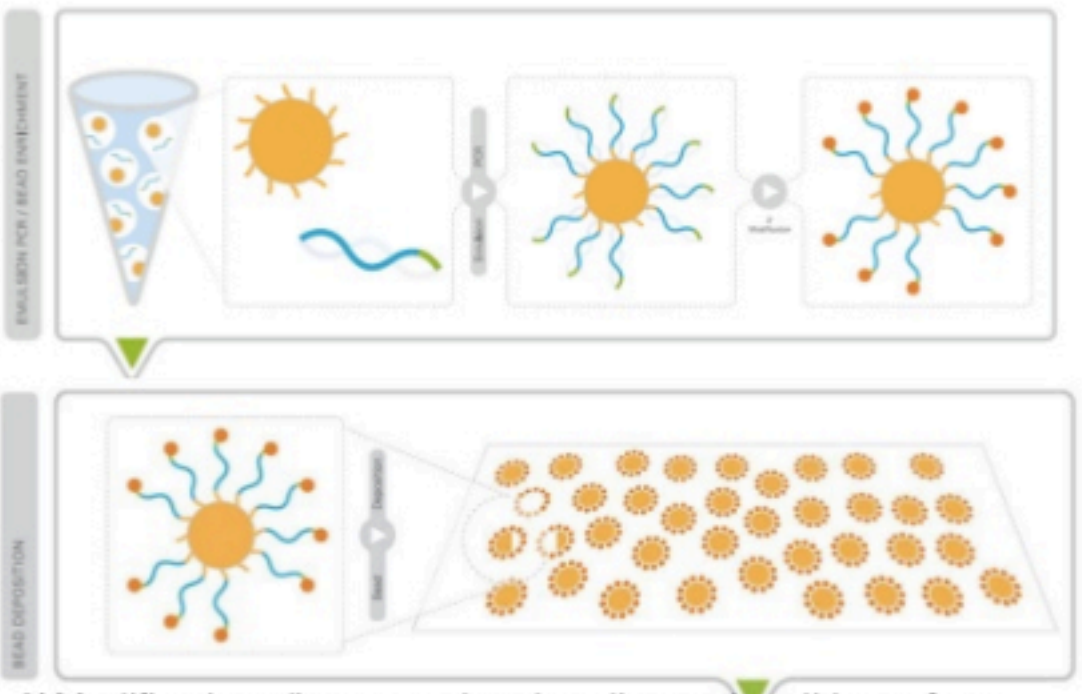
ABI Solid Details



From Mardis 2008. Annual Rev. Genetics 9: 387.

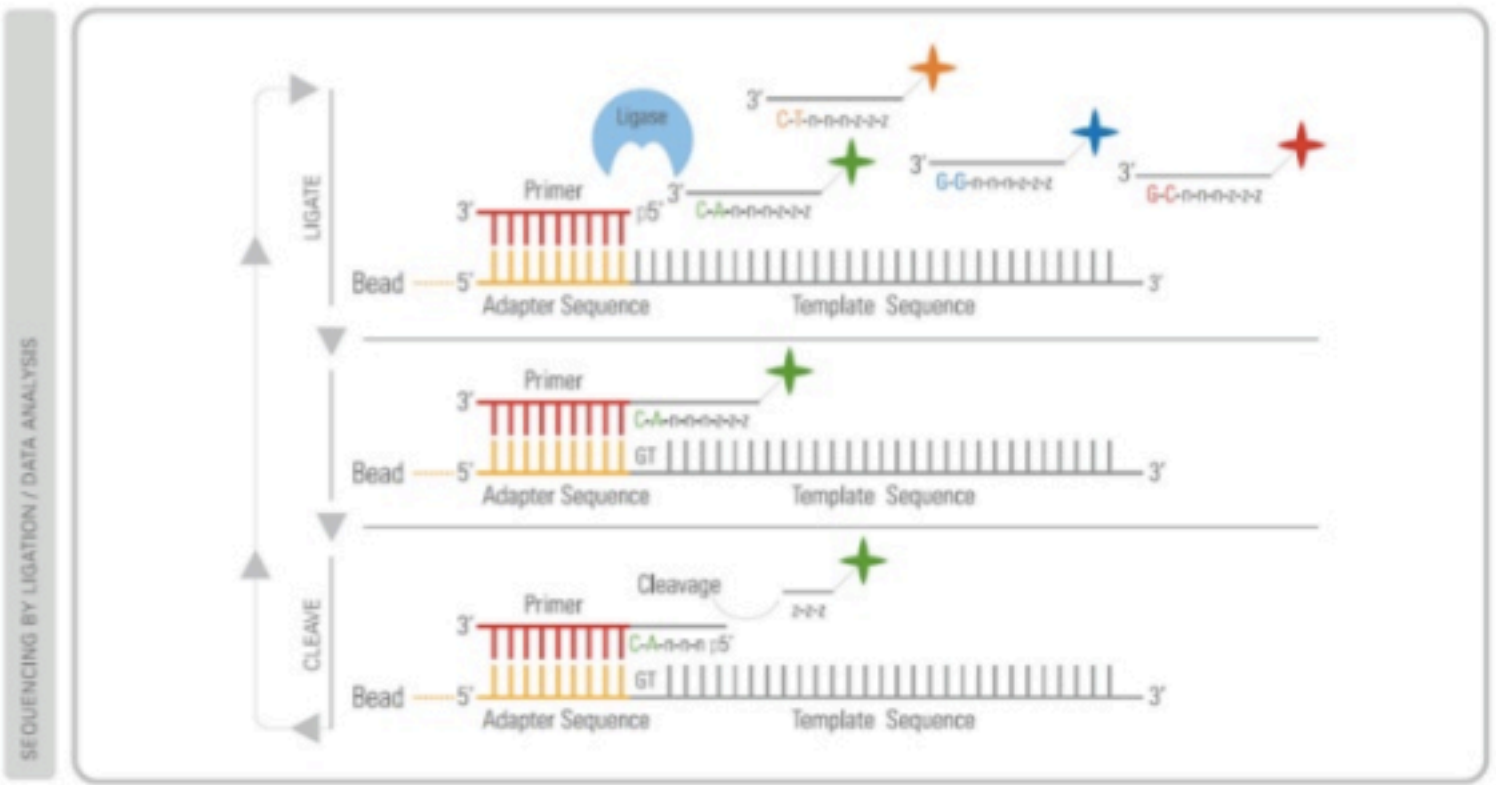
The ligase-mediated sequencing approach of the Applied Biosystems SOLiD sequencer. In a manner similar to Roche/454 emulsion PCR amplification, DNA fragments for SOLiD sequencing are amplified on the surfaces of 1-µm magnetic beads to provide sufficient signal during the sequencing reactions, and are then deposited onto a flow cell slide. Ligase-mediated sequencing begins by annealing a primer to the shared adapter sequences on each amplified fragment, and then DNA ligase is provided along with specific fluorescent-labeled 8mers, whose 4th and 5th bases are encoded by the attached fluorescent group. Each ligation step is followed by fluorescence detection, after which a regeneration step removes bases from the ligated 8mer (including the fluorescent group) and concomitantly prepares the extended primer for another round of ligation. (b) Principles of two-base encoding. Because each fluorescent group on a ligated 8mer identifies a two-base combination, the resulting sequence reads can be screened for base-calling errors versus true polymorphisms versus single base deletions by aligning the individual reads to a known high-quality reference sequence.

emPCR and Enrichment



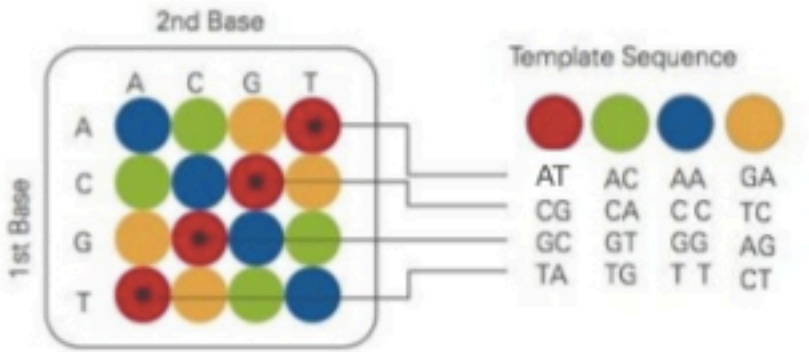
3' Modification allows covalent bonding to the slide surface

Sequencing by Ligation



2 Base encoding

Possible Dinucleotides Encoded By Each Color



Double Interrogation

With 2 base encoding each base is defined twice

Life: SOLiD – Performance Specs

	SOLiD 5500xl	SOLiD 5500xl W	SOLiD 5500	SOLiD 5500 W
Gb/run	95	240	48	120
Run time	6 days	10 days	6 days	10 days
Gb/day	~16	24	~8	12
Read length	2X60	2X50	2X60	2X50
# of single reads/run	~800M	2.4B	~400M	1.2B
Instrument cost	$595k	$70k upgrade	$349k	$70k upgrade
Run cost	~$10k	~$5k	~$5k	~$2.5k

Platform Updates

SOLiD 3	• 50bp Paired reads ~50Gbp / run (12 days)
SOLiD 4	• 50bp Paired reads ~100Gbp / run (12 days)
5500xl	• 75bp Paired reads ~300Gbp / run (14 days)

Maximum yield / day 21,000,000,000bp
7x the human genome
3.5 hours of sequencing for a 1 fold coverage.....

SOLiD Colour Space Reads

- *.csfasta *(colour space fasta)*
- *.qual *(Phred quality scores)*

```
>853_17_1660_F3
T32111011201320102312.....
```

AA	CC	GG	TT	0	Blue
AC	CA	GT	TG	1	Green
AG	CT	GA	TC	2	Yellow
AT	CG	GC	TA	3	Red

Comparison

From Slideshare presentation of Cosentino Cristian
http://www.slideshare.net/cosentia/high-throughput-equencing

Comparison in 2008

Sequencing	Amplif.	Chemistry	Read lenght (bp)	Run time (d)	Gbp/ day	DNA required (µg)	$/sequencer *(ref. 2008)*
Roche 454 GS FLX Titanium	emPCR	Pyrosequencing	250-400	0.35 *	1.3	3-5	500.000
ABi SOLiD	emPCR	Sequencing by ligation	25-50	7-14	3.6	0.1-20	595.000
Illumina GAII	Solid-phase	Reversible terminator	36-100	4-9	3.9	0.1-1	430.000

From Slideshare presentation of Cosentino Cristian
http://www.slideshare.net/cosentia/high-throughput-equencing

Comparison in 2008

Sequencing	Advantages	Disadvantages	$/Mbp *(in 2008)**
Roche 454	•Long reads even > 400 bp, improving *de novo* sequencing •Rare sustitution errors	•High indel in homopolymer stretches > 6 nucl. •High reagent cost •Longest reads only in single-read (2x150 bp)	60
ABi SOLiD	•Error correction with the two-base encoding system	•Long time run •Needs of cluster station to perform base calling and up to 1 week to align •Alignment must be performed against a reference db	2
Illumina GAII	•Most widely used platform (> 90 science/nature publication) •Sample preparation automatable •SBS , real-time analysis and base calling are performed simultaneously to the run •Automated cluster generation	•Low multiplexing capability •Substitution errors	2

**Nat. Biotech., 2008, 26: 1135-1145*

From Slideshare presentation of Cosentino Cristian
http://www.slideshare.net/cosentia/high-throughput-equencing

Comparison in 2008

	Roche (454)	Illumina	SOLiD
Chemistry	Pyrosequencing	Polymerase-based	Ligation-based
Amplification	Emulsion PCR	Bridge Amp	Emulsion PCR
Paired ends/sep	Yes/3kb	Yes/200 bp	Yes/3 kb
Mb/run	100 Mb	1300 Mb	3000 Mb
Time/run	7 h	4 days	5 days
Read length	250 bp	32-40 bp	35 bp
Cost per run (total)	$8439	$8950	$17447
Cost per Mb	$84.39	$5.97	$5.81

From “Introduction to Next Generation Sequencing” by Stefan Bekiranov prometheus.cshl.org/twiki/pub/Main/CdAtA08/CSHL_nextgen.ppt

Comparison in 2012

	Roche (454)	Illumina	SOLiD
Chemistry	Pyrosequencing	Polymerase-based	Ligation-based
Amplification	Emulsion PCR	Bridge Amp	Emulsion PCR
Paired ends/sep	Yes/3kb	Yes/200 bp	Yes/3 kb
Mb/run	100 Mb	1300 Mb	3000 Mb
Time/run	7 h	4 days	5 days
Read length	250 bp	32-40 bp	35 bp
Cost per run (total)	$8439	$8950	$17447
Cost per Mb	$84.39	$5.97	$5.81

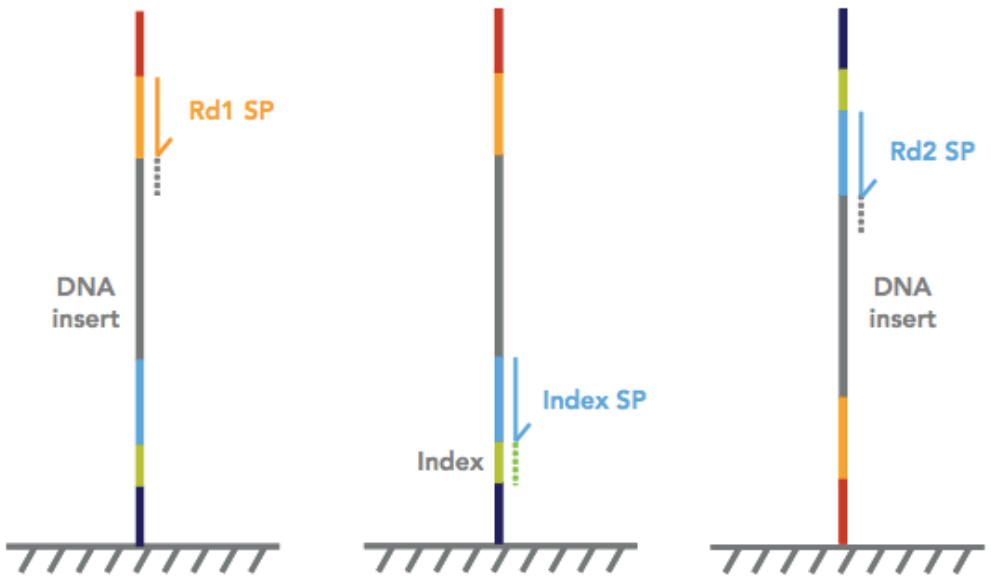
From “Introduction to Next Generation Sequencing” by Stefan Bekiranov prometheus.cshl.org/twiki/pub/Main/CdAtA08/CSHL_nextgen.ppt

Bells and Whistles

- Multiplexing
- Paired end
- Mate pair
- ????
-

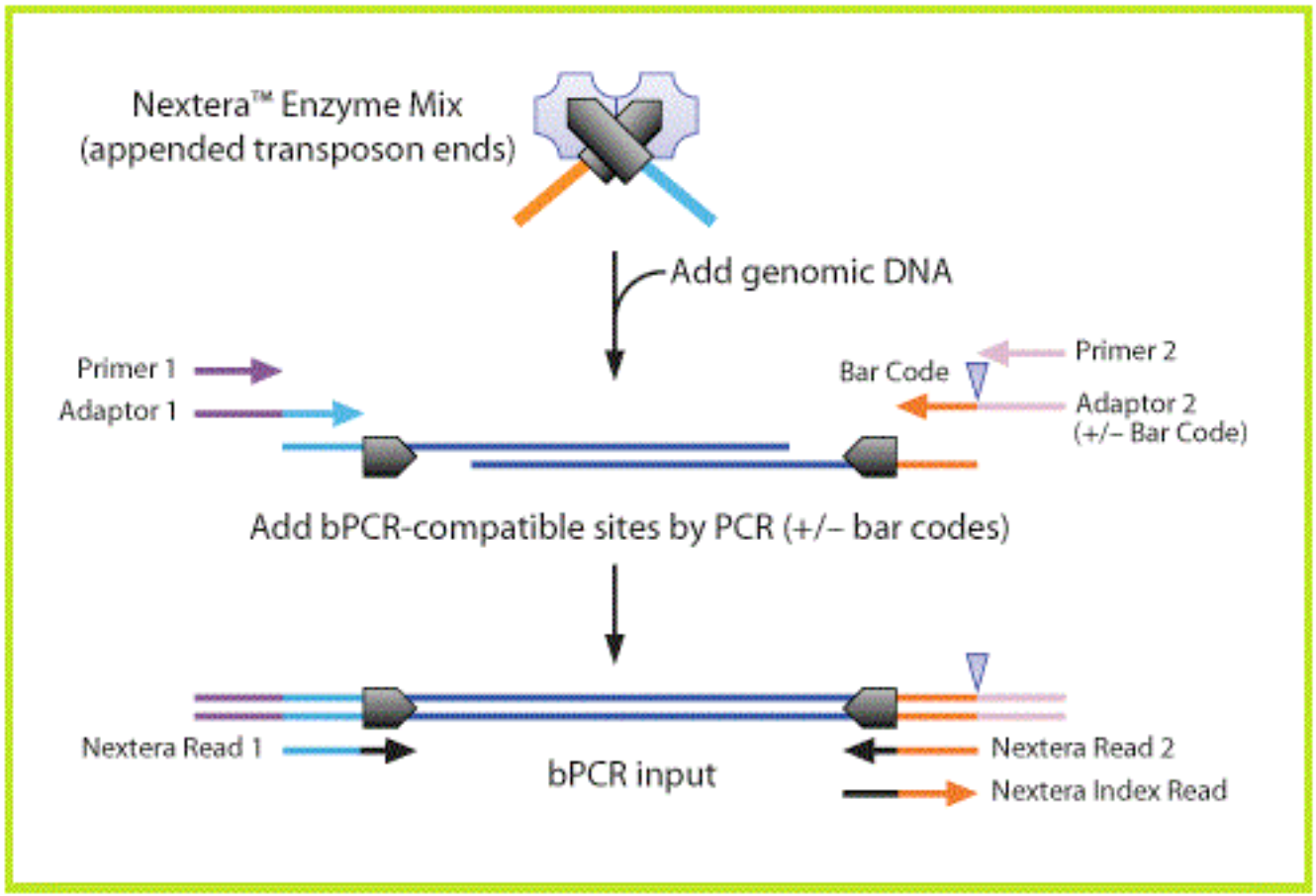
Multiplexing

FIGURE 1: MULTIPLEXED SEQUENCING PROCESS



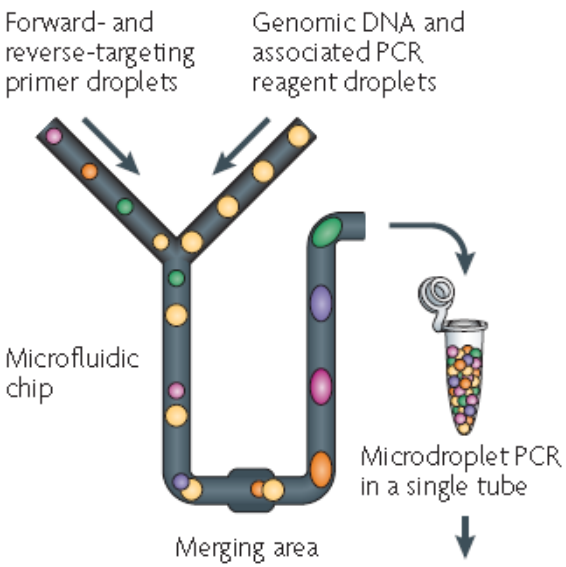
Sample multiplexing involves a total of three sequencing reads, including a separate index read, which is generated automatically on the Genome Analyzer equipped with the Paired-End Module. A: Application read 1 (dotted line) is generated using the Read 1 Sequencing Primer (Rd1 SP). B: The read 1 product is removed and the Index Sequencing Primer (Index SP) is annealed to the same strand to produce the 6-bp index read (dotted line). C: If a paired-end read is required, the original template strand is used to regenerate the complementary strand. Then, the original strand is removed and the complementary strand acts as a template for application read 2 (dotted line), primed by the Read 2 Sequencing Primer (Rd2 SP). Pipeline Analysis software identifies the index sequence from each cluster so that the application reads can be assigned to a single sample. Hatch marks represent the flow cell surface.

Small amounts of DNA



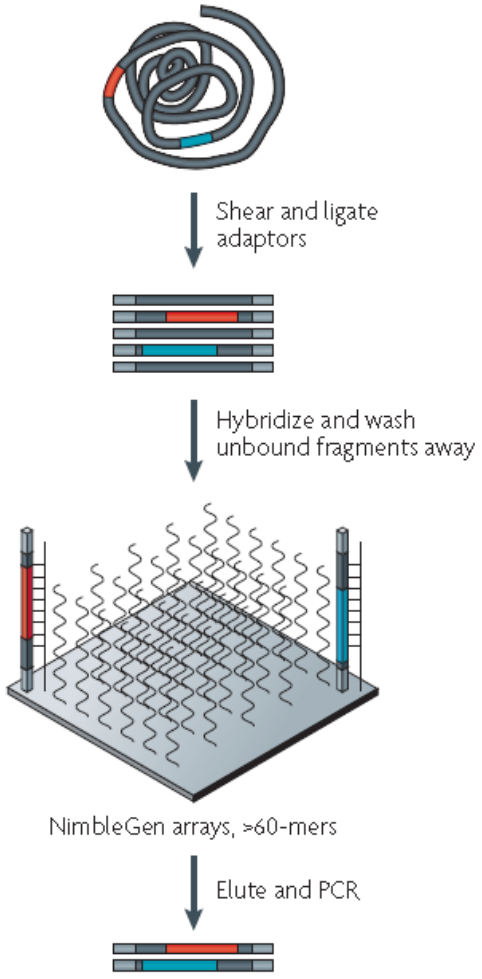
Capture Methods

RainDance
Microdroplet PCR



Reported 84% of capture efficiency

Roche Nimblegen
Salid-phase capture with custom-designed oligonucleotide microarray

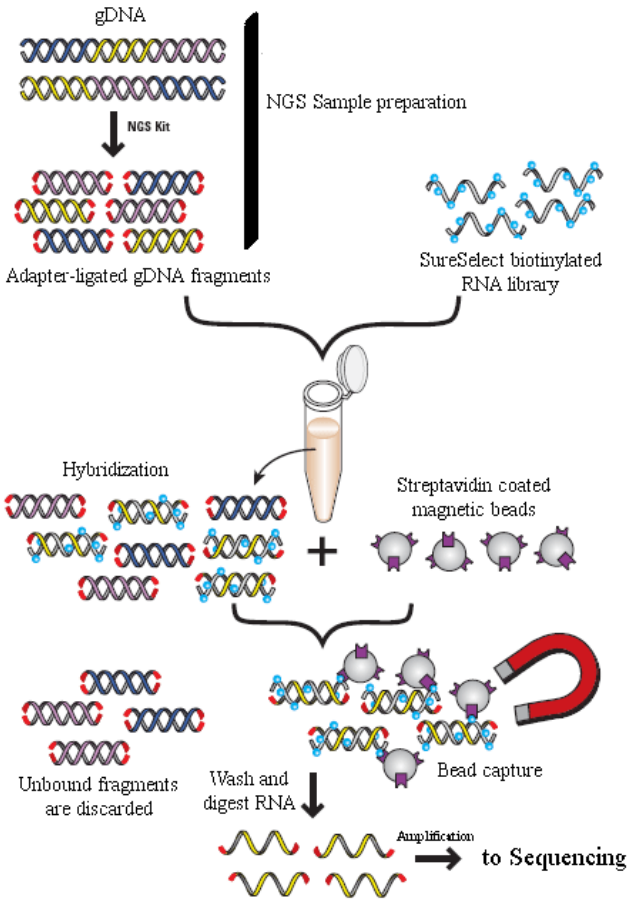


Reported 65-90% of capture efficiency

From Slideshare presentation of Cosentino Cristian
http://www.slideshare.net/cosentia/high-throughput-equencing

Agilent SureSelect

Solution-phase capture with
streptavidin-coated magnetic beads



Reported 60-80% of capture efficiency

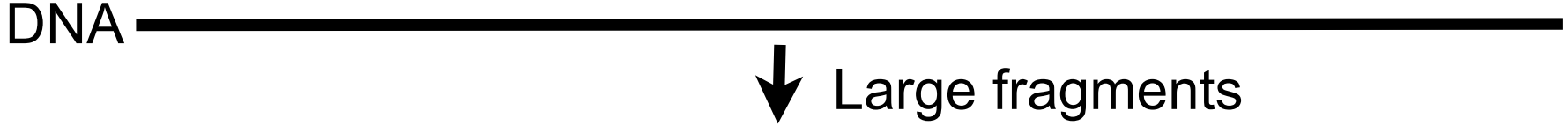
From Slideshare presentation of Cosentino Cristian http://www.slideshare.net/cosentia/high-throughput-equencing

Moleculo

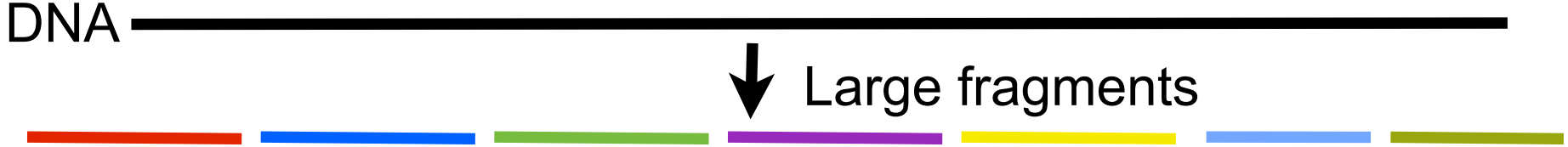
Moleculo

DNA

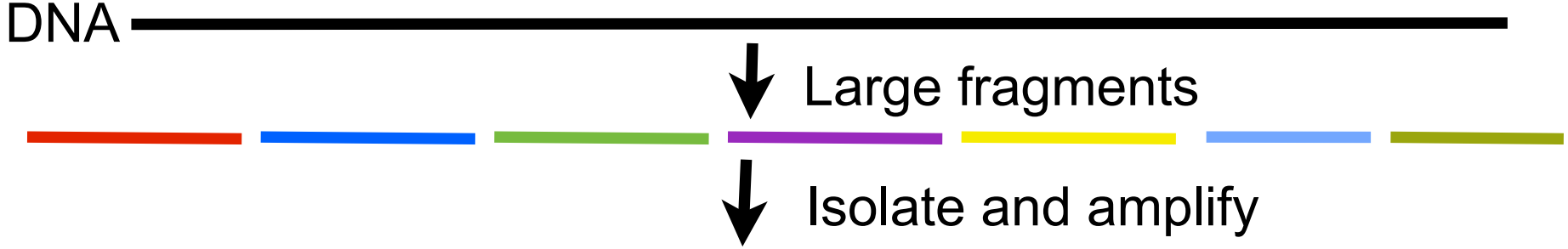
Moleculo



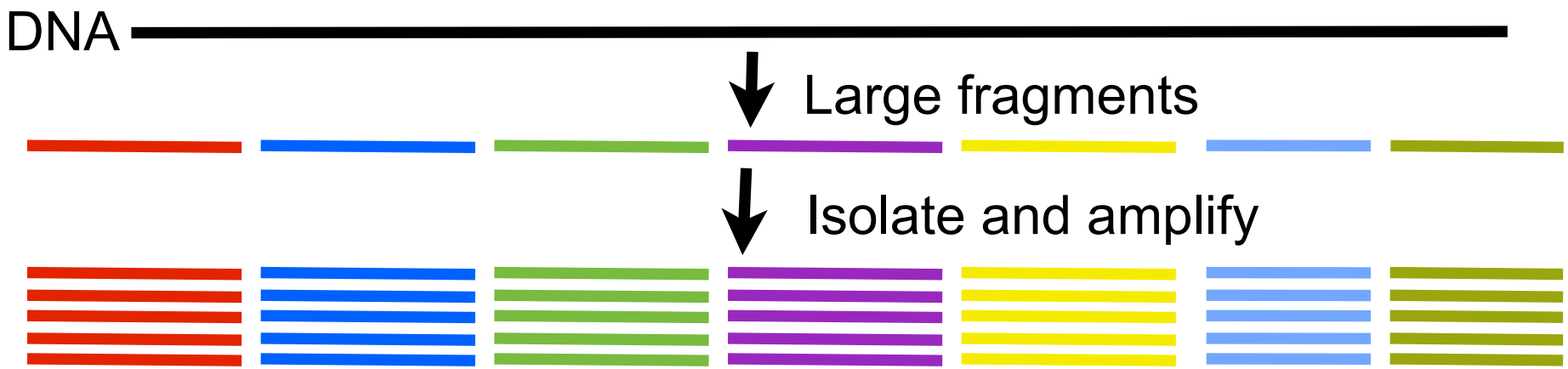
Moleculo



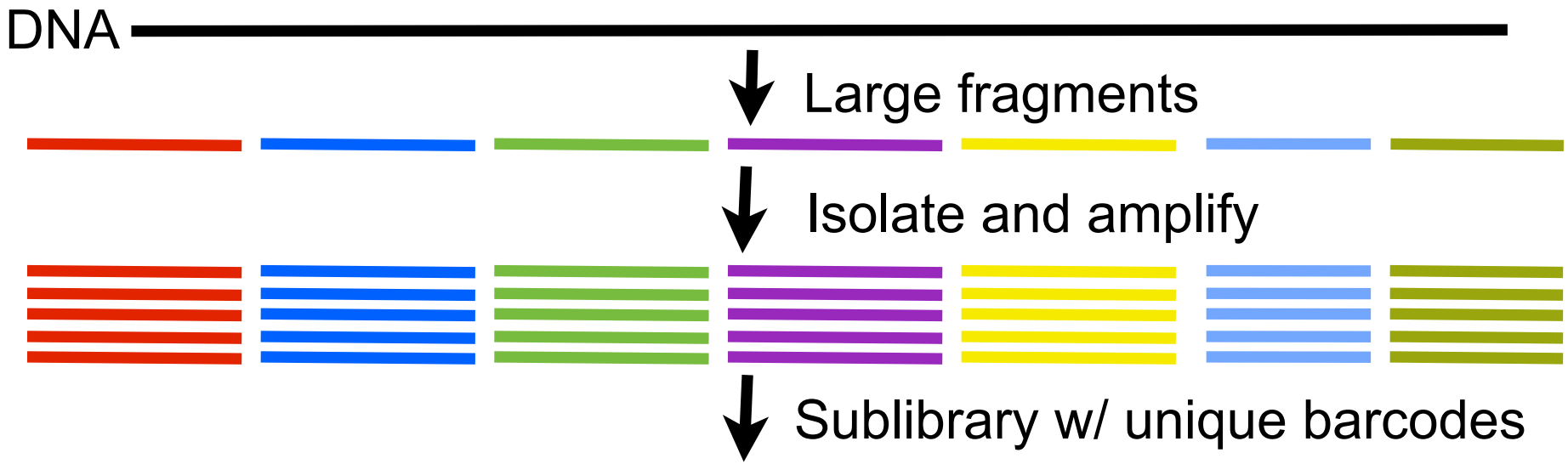
Moleculo



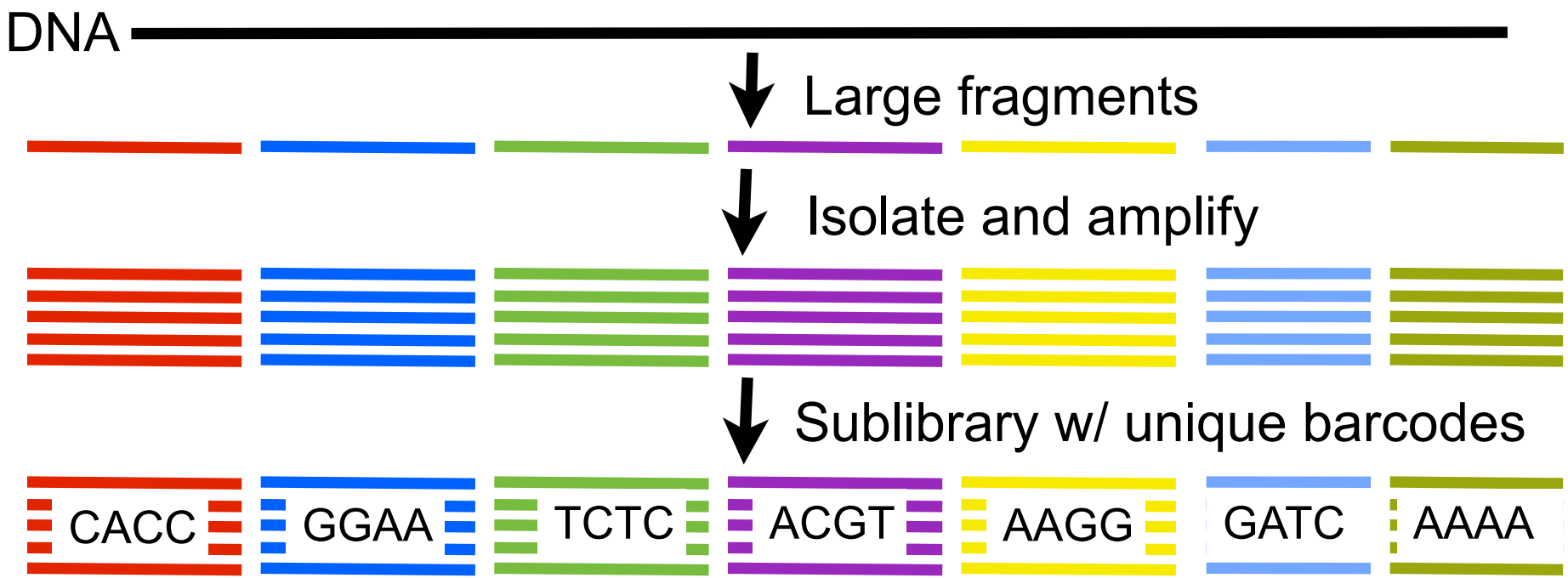
Moleculo



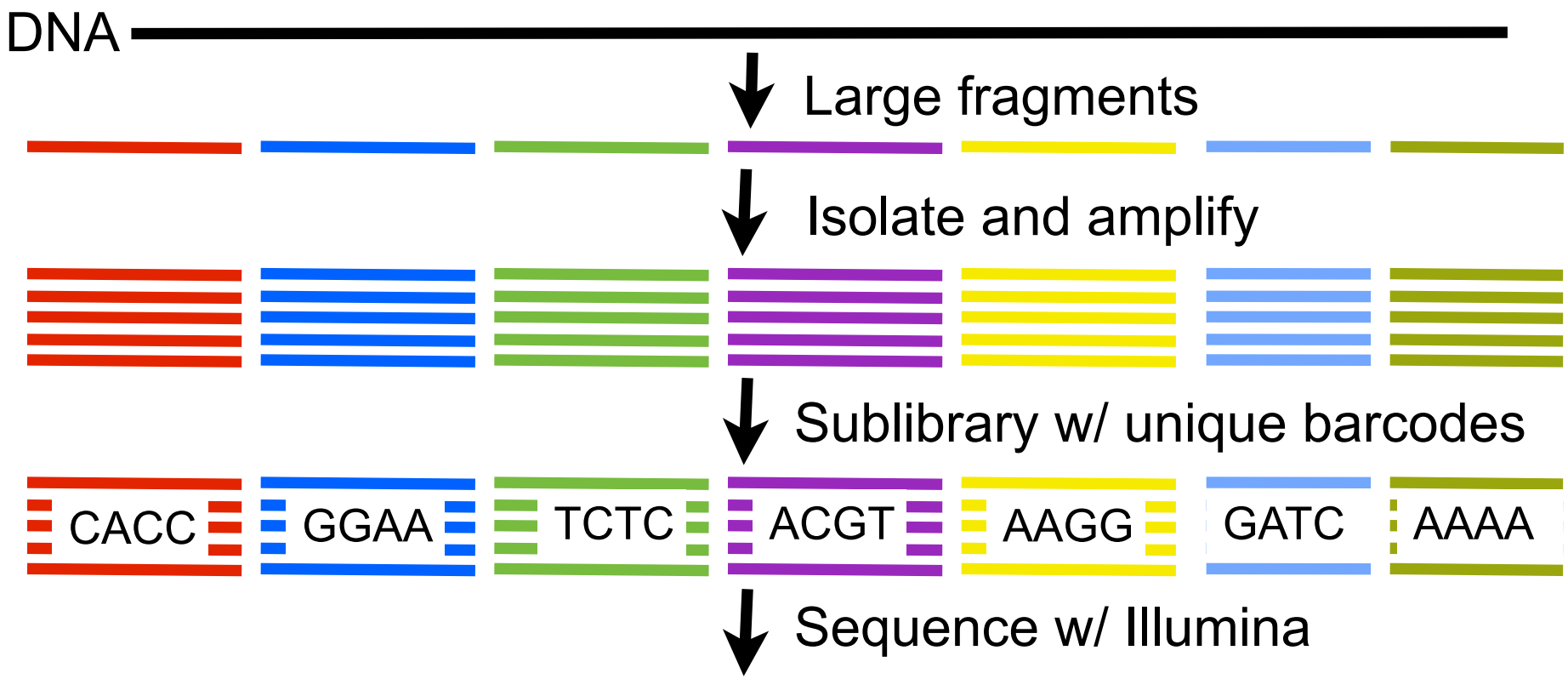
Moleculo


DNA
Large fragments
Isolate and amplify
Sublibrary w/ unique barcodes

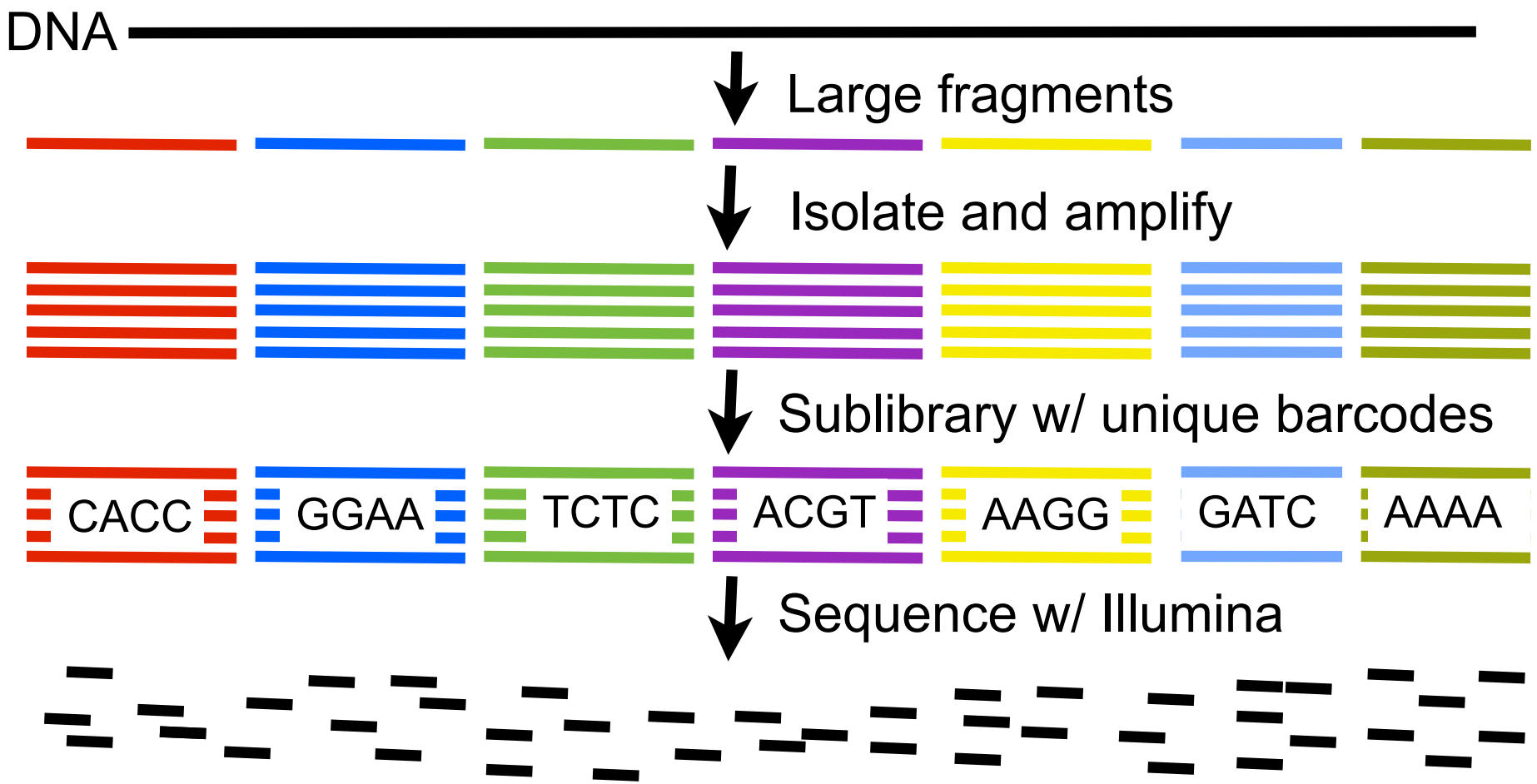
Moleculo


DNA
Large fragments
Isolate and amplify
Sublibrary w/ unique barcodes
CACC
GGAA
TCTC
ACGT
AAGG
GATC
AAAA

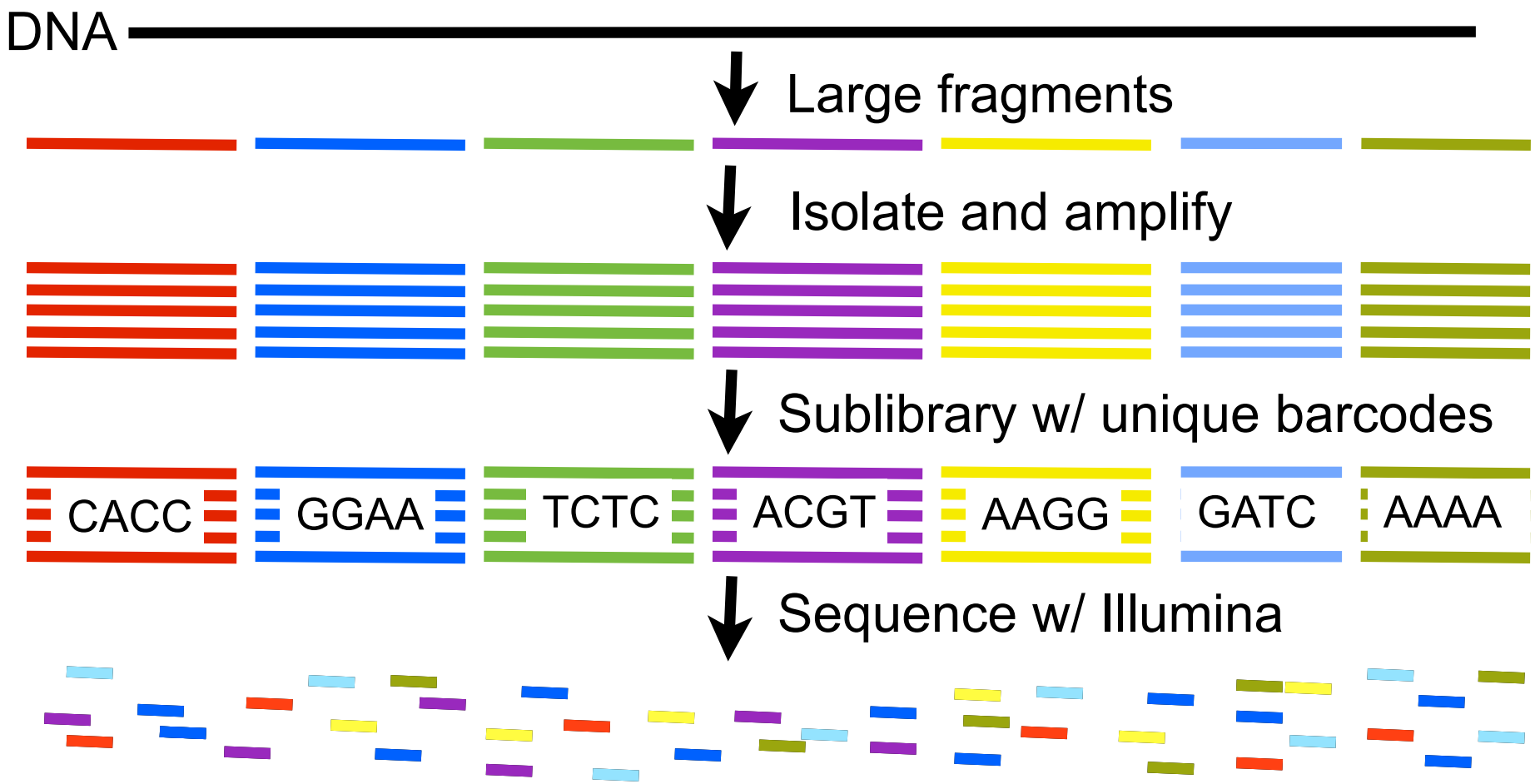
Moleculo


DNA
Large fragments
Isolate and amplify
Sublibrary w/ unique barcodes
CACC
GGAA
TCTC
ACGT
AAGG
GATC
AAAA
Sequence w/ Illumina

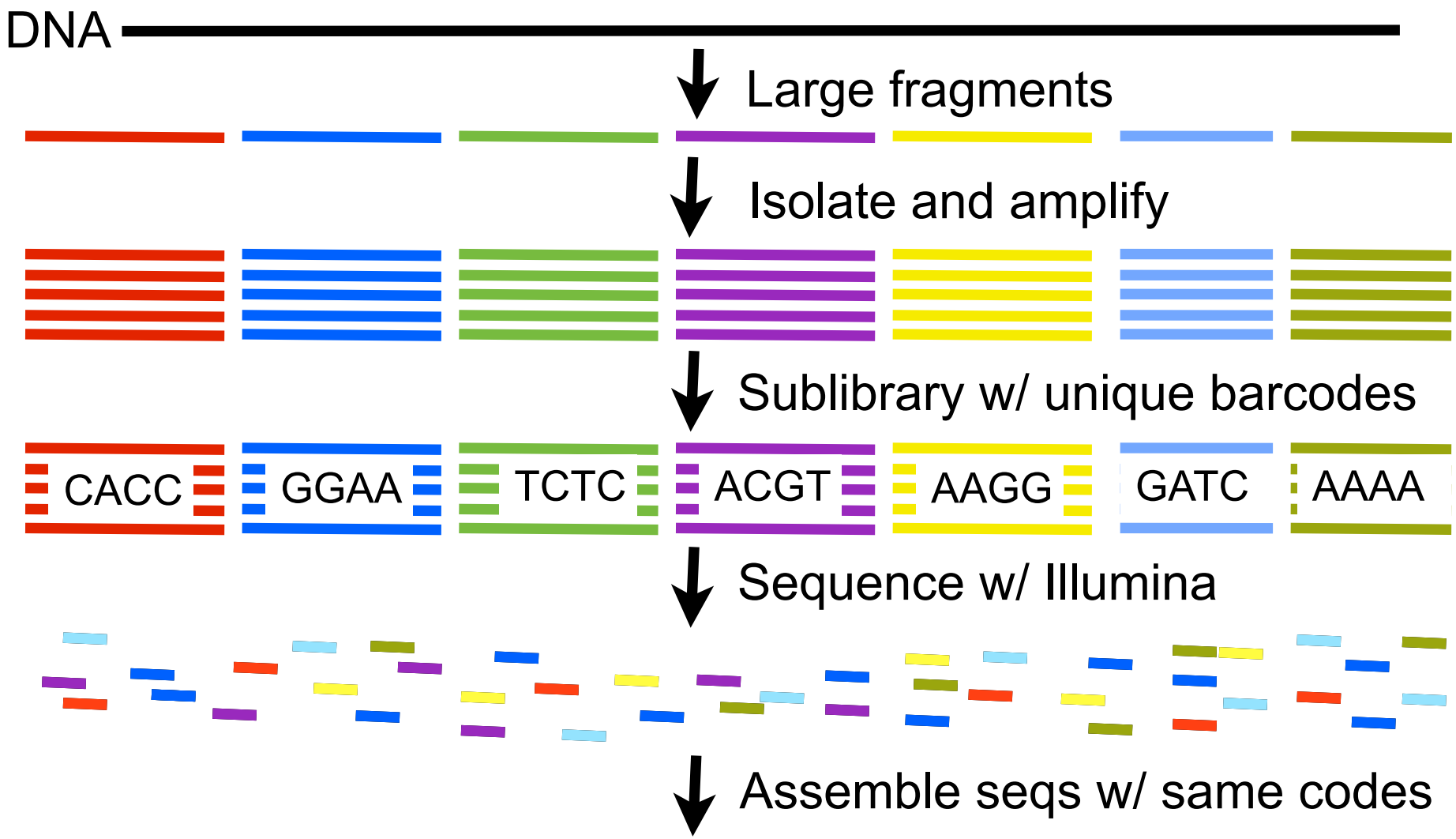
Moleculo


DNA
Large fragments
Isolate and amplify
Sublibrary w/ unique barcodes
CACC
GGAA
TCTC
ACGT
AAGG
GATC
AAAA
Sequence w/ Illumina

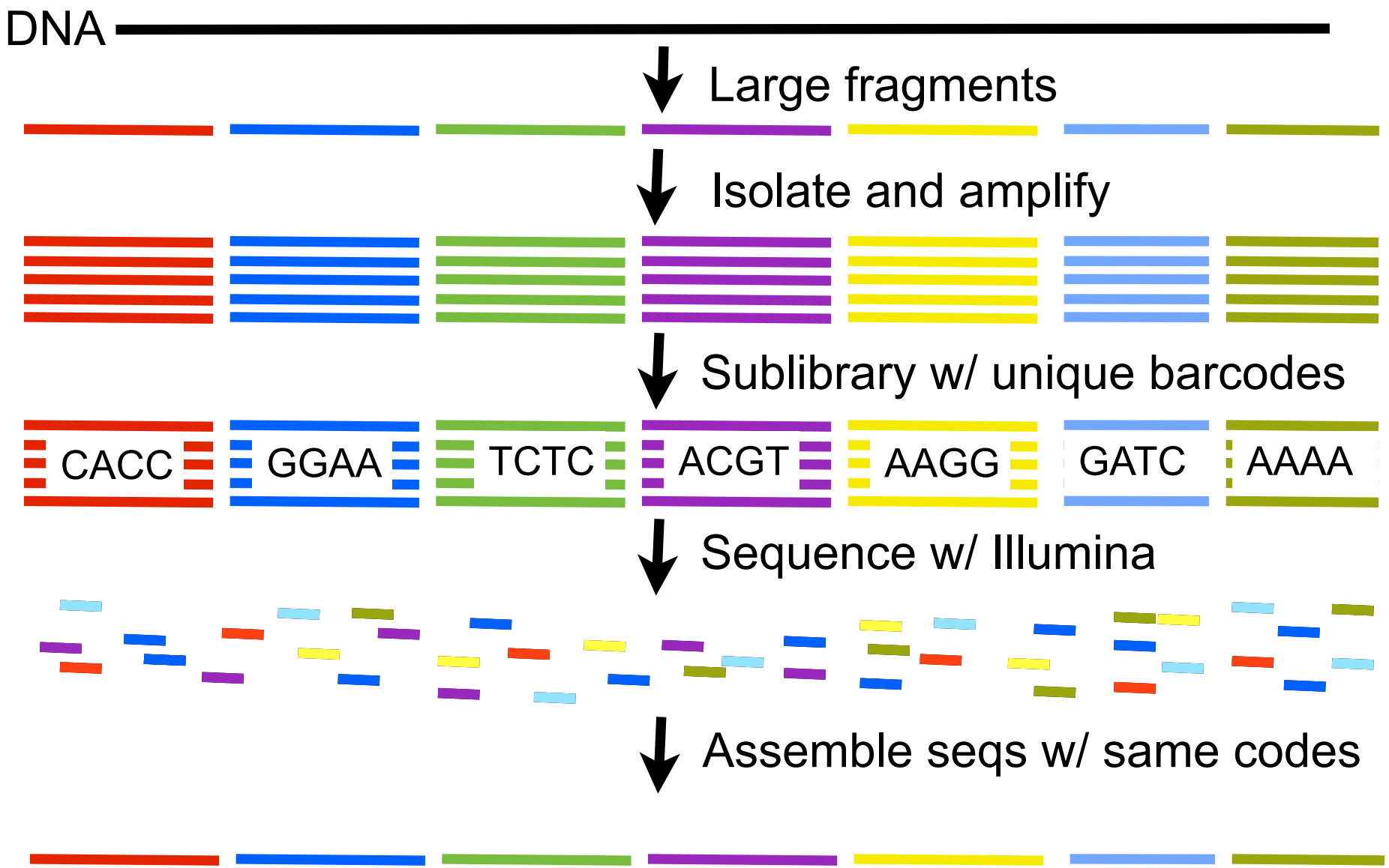
Moleculo


DNA
Large fragments
Isolate and amplify
Sublibrary w/ unique barcodes
CACC
GGAA
TCTC
ACGT
AAGG
GATC
AAAA
Sequence w/ Illumina

Moleculo


DNA
Large fragments
Isolate and amplify
Sublibrary w/ unique barcodes
CACC
GGAA
TCTC
ACGT
AAGG
GATC
AAAA
Sequence w/ Illumina
Assemble seqs w/ same codes

Moleculo


DNA
Large fragments
Isolate and amplify
Sublibrary w/ unique barcodes
CACC
GGAA
TCTC
ACGT
AAGG
GATC
AAAA
Sequence w/ Illumina
Assemble seqs w/ same codes

Generation 3.5

- **Even faster**

Ion Torrent PGM

Ion Torrent PGM

Applied Biosystems Ion Torrent PGM

Applied Biosystems Ion Torrent PGM

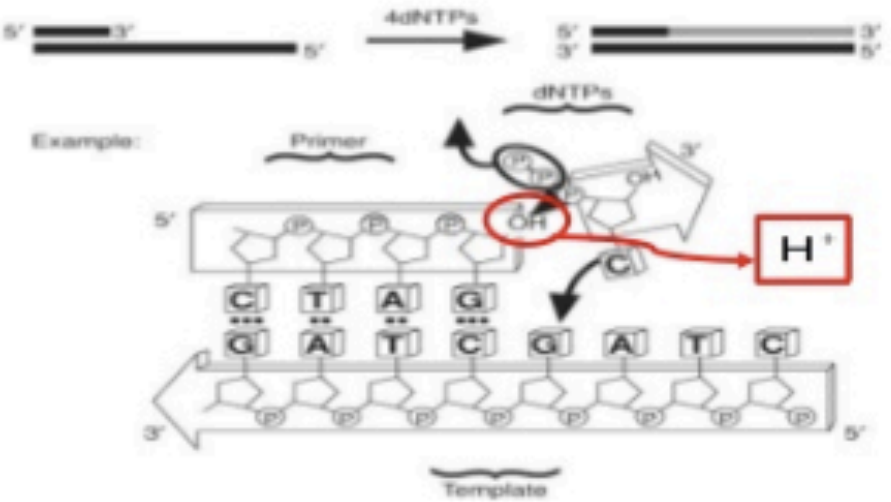
Workflow similar to that for Roche/454 systems.

Not surprising, since invented by people from 454.

Applied Biosystems Ion Torrent PGM

Ion Torrent

- Ion Semiconductor Sequencing
- Detection of hydrogen ions during the polymerization DNA
- Sequencing occurs in microwells with ion sensors
- No modified nucleotides
- No optics



Ion Torrent

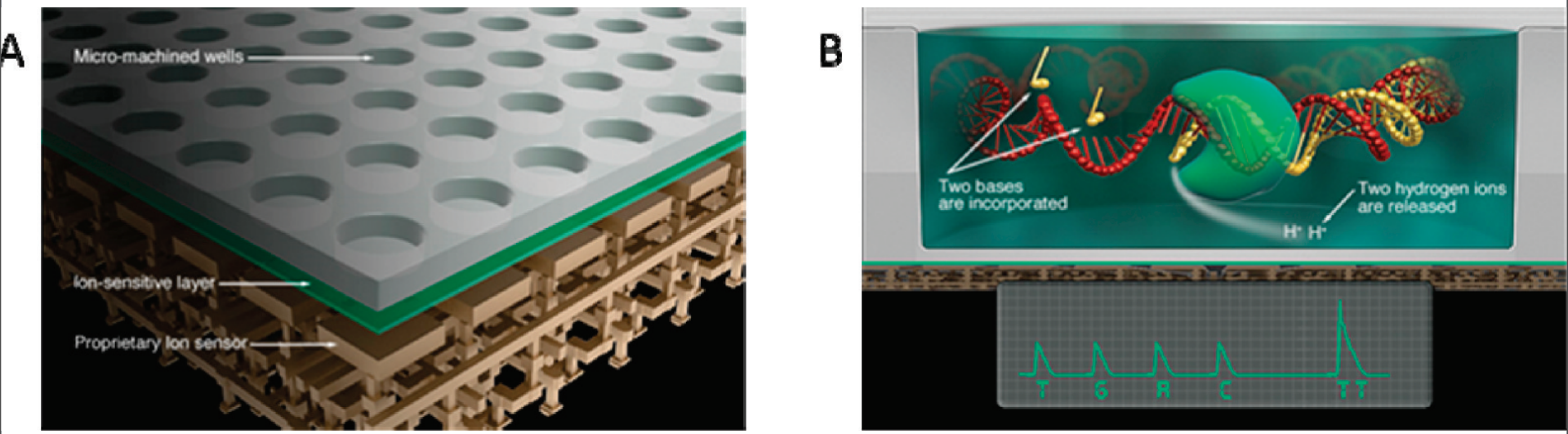


Figure4. Layout of IonTorrent's semiconductor sequencing chip technology.(A)A layer-by-layer view of the chip revealing the structural design. The top layer contains the individual DNA polymerization reaction wells, and the bottom two layers comprise the FET ion sensor. Each well has a corresponding FET detector that identifies a change in pH. (B) A side view of an individual reaction well depicting DNA polymerase incorporation of a repeat of two TTP nucleotides on a sequencing fragment. The hydrogen ions released during this process are detected by the FET below. Reprinted with permission from Ion Torrent (Wes Conrad).

From Niedringhaus et al. Analytical Chemistry 83: 4327. 2011.

Life: Ion Torrent – Performance Specs

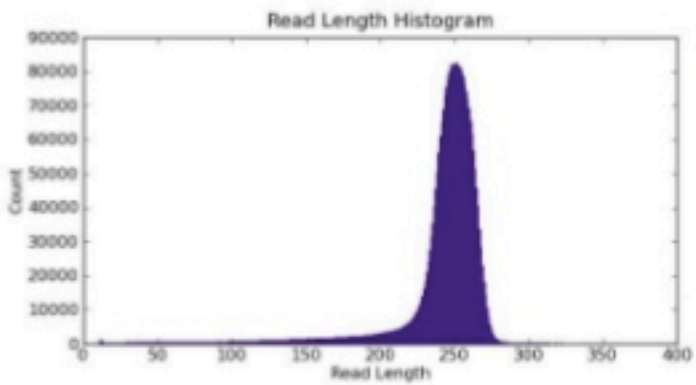
	PGM 314	PGM 316	PGM 318	Proton 1	Proton 2
Gb/run	10-40 Mb	100-400 Mb	1 Gb	~10 Gb	~100 Gb
Run time	2 hours	2 hours	2 hours	~4 hours	~4 hours
Gb/day	~120 Mb	~1.2 Gb	~3 Gb	~30 Gb	~ 300 Gb
Read length	200b	200b	200b	200b	>200b
# of single reads/run	~0.6M	~3M	~5.5M	~82M	~330M
Instrument cost	$50k	$50k	$50k	$149k	$149k
Run cost	$349	$549	$749	~$1k	~$1k

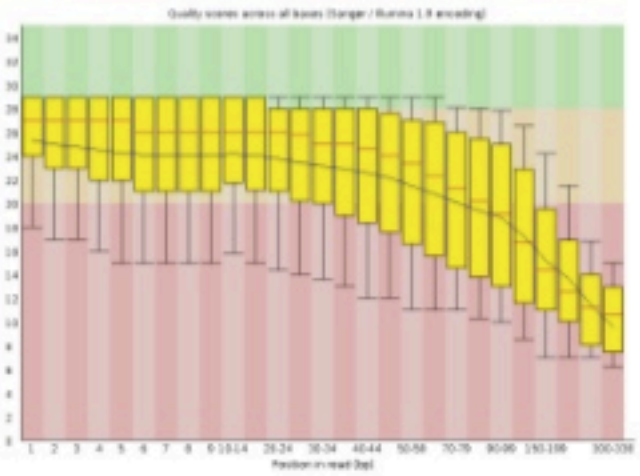
Ion Torrent: System Updates

314 Chip	• 100bp reads ~10 Mb/run (1.5 hrs)
316 Chip	• 100 bp reads ~100 Mbp / run (2 hrs) • 200 bp reads ~200 Mbp/run (3 hrs)
318 Chip	• 200 bp reads ~1 Gbp / run (4.5 hrs)

Ion Torrent Reads

- *.sff *(standard flowgram format)*
- *.fastq (*sequence and corresponding quality score encoded with an ASCII character, phred-like quality score + 33)*





Generation IV??

- **Single molecule sequencing**

Pacific Biosciences

Pacific Biosciences

Pacific Biosciences

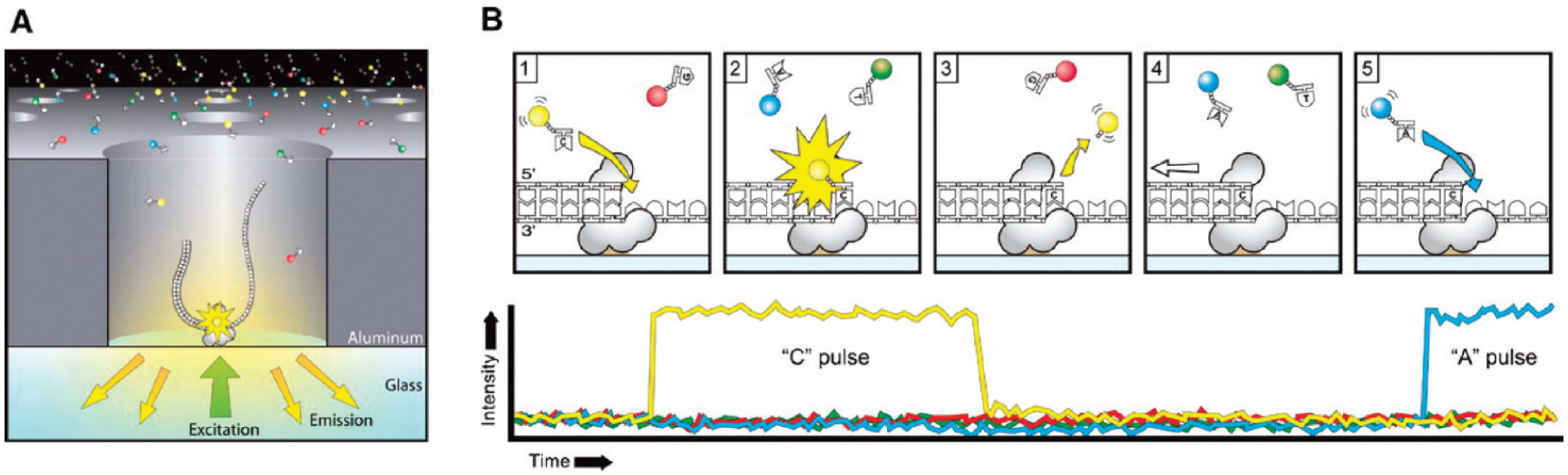


Figure 2. Schematic of PacBio's real-time single molecule sequencing. (A) The side view of a single ZMW nanostructure containing a single DNA polymerase (Φ29) bound to the bottom glass surface. The ZMW and the confocal imaging system allow fluorescence detection only at the bottom surface of each ZMW. (B) Representation of fluorescently labeled nucleotide substrate incorporation on to a sequencing template. The corresponding temporal fluorescence detection with respect to each of the five incorporation steps is shown below.

From Niedringhaus et al. Analytical Chemistry 83: 4327. 2011.

Pacific Biosciences – Performance Specs

	PacBio RS 'C2'
Mb/run	120
Run time	40 min
Gb/day	~1 Gb
Read length	3kb (avg)
# of single reads/run	~50k
Instrument cost	~$700k
Run cost	$100

Complete Genomics

Complete Genomics

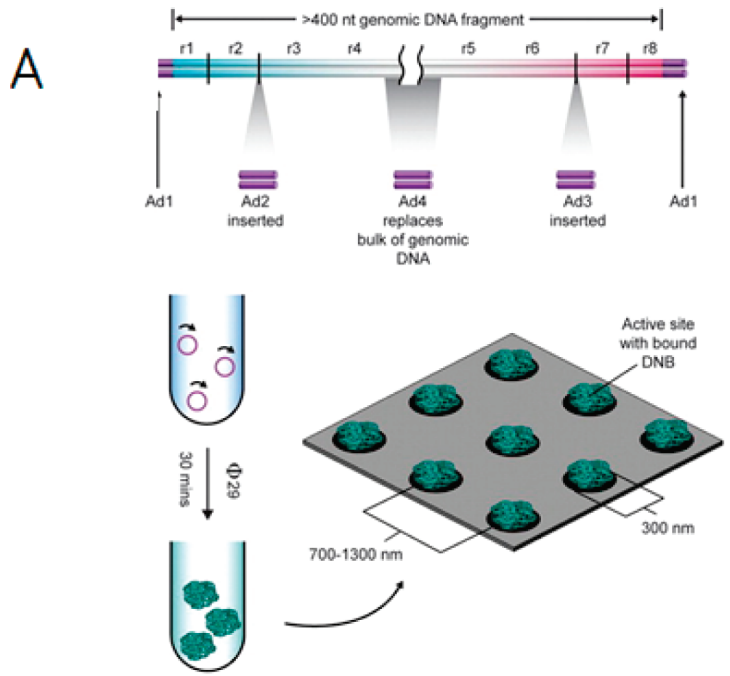
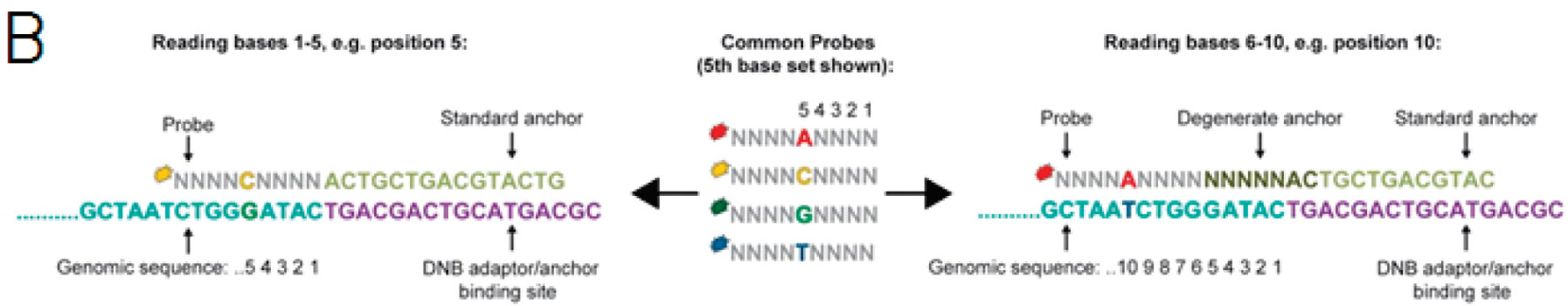


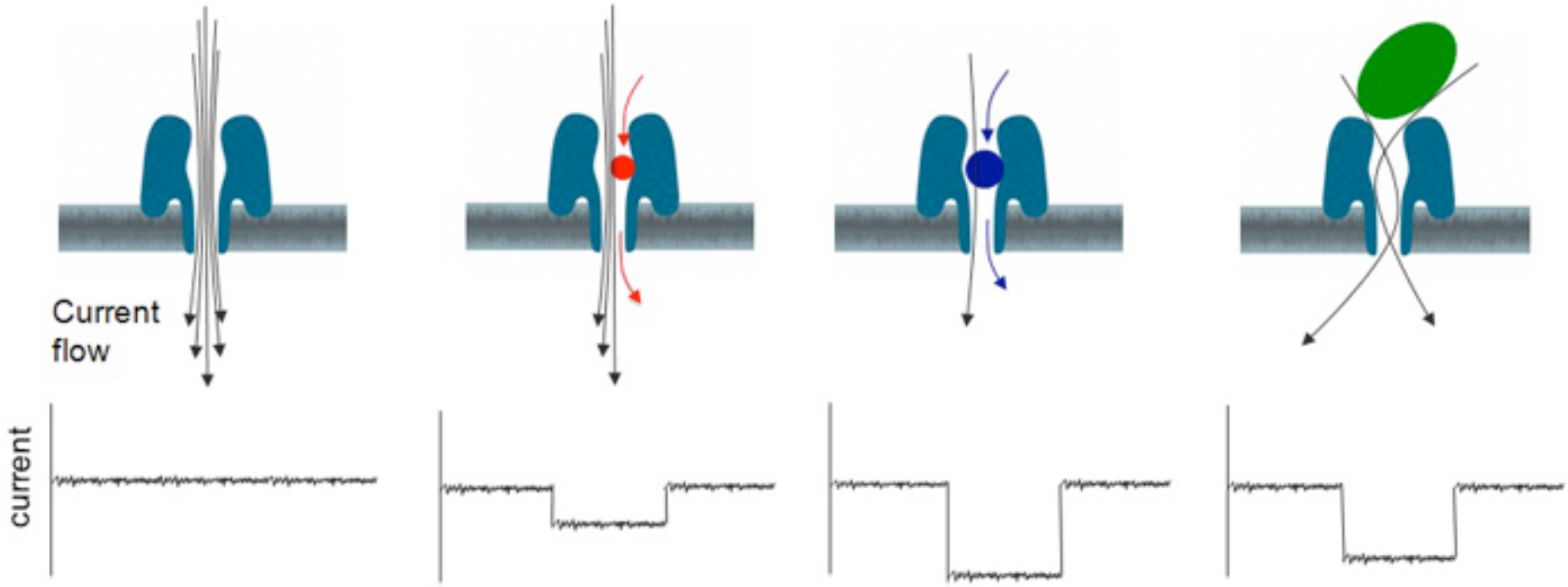
Figure 3. Schematic of Complete Genomics' DNB array generation and cPAL technology. (A) Design of sequencing fragments, subsequent DNB synthesis, and dimensions of the patterned nanoarray used to localize DNBs illustrate the DNB array formation. (B) Illustration of sequencing with a set of common probes corresponding to 5 bases from the distinct adapter site. Both standard and extended anchor schemes are shown.



From Niedringhaus et al. Analytical Chemistry 83: 4327. 2011.

Oxford Nanopore

Oxford Nanopore



This diagram shows a protein nanopore set in an electrically resistant membrane bilayer. An ionic current is passed through the nanopore by setting a voltage across this membrane.

If an analyte passes through the pore or near its aperture, this event creates a characteristic disruption in current. By measuring that current it is possible to identify the molecule in question. For example, this system can be used to distinguish the four standard DNA bases and G, A, T and C, and also modified bases. It can be used to identify target proteins, small molecules, or to gain rich molecular information for example to distinguish the enantiomers of ibuprofen or molecular binding dynamics.

From Oxford Nanopores Web Site

Oxford Nanopore

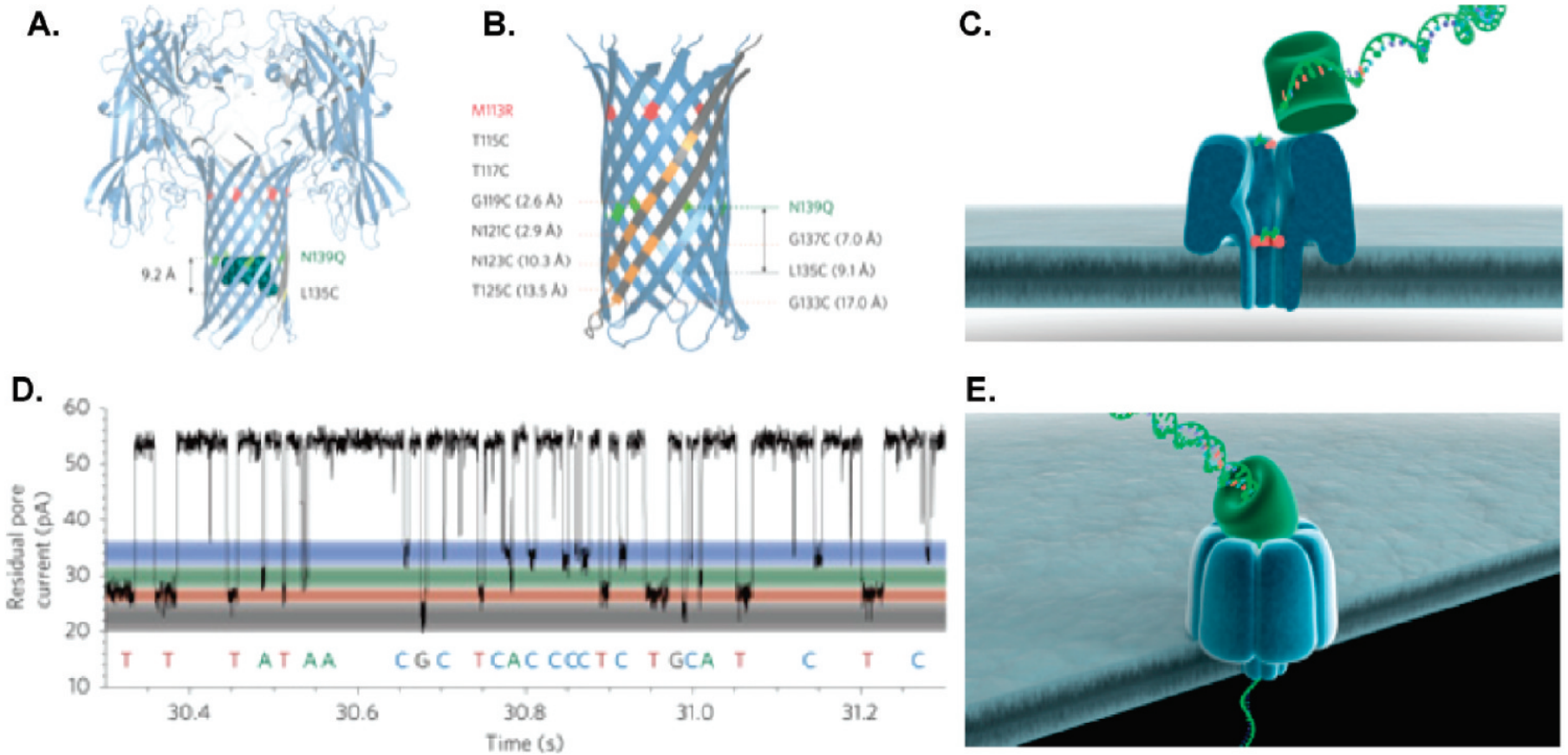
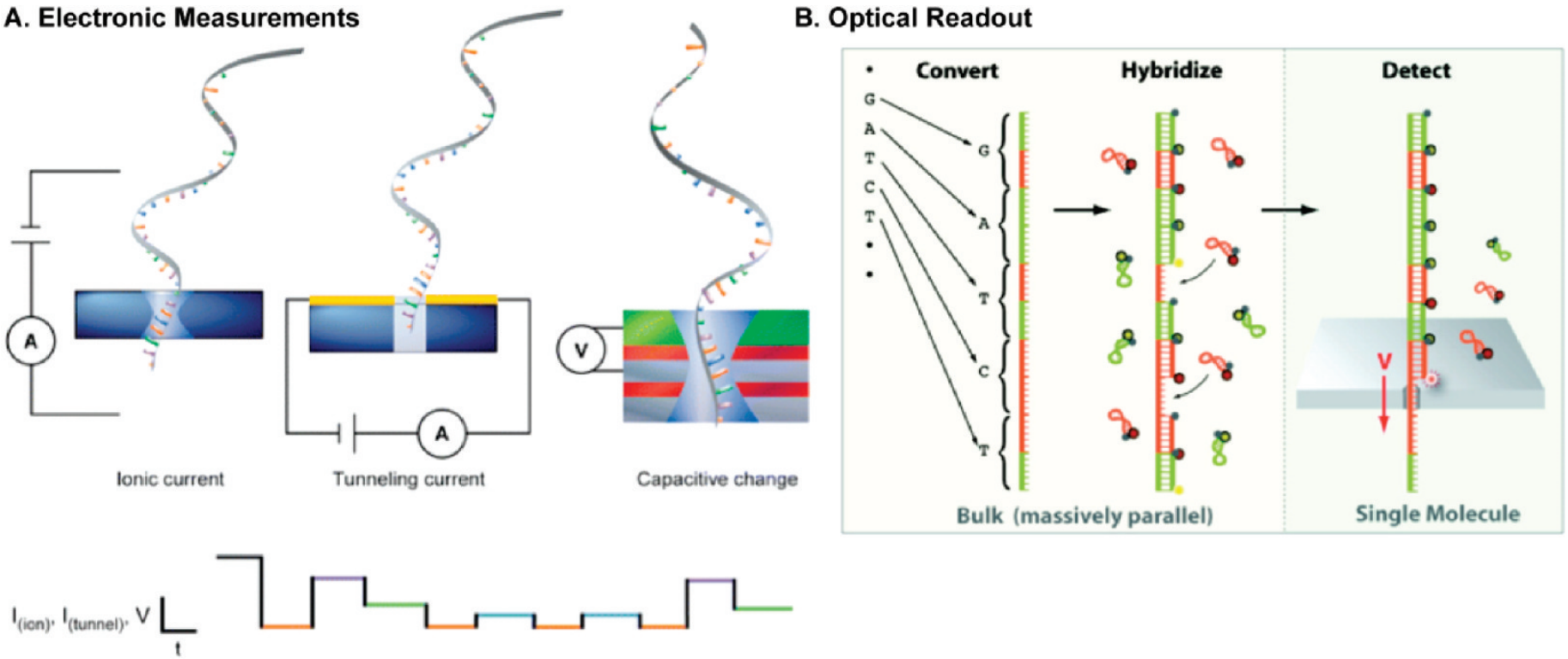


Figure6. BiologicalnanoporeschemeemployedbyOxfordNanopore.(A)SchematicofRHLproteinnanoporemutantdepictingthepositionsofthe cyclodextrin (at residue 135) and glutamines (at residue 139). (B) A detailed view of the β barrel of the mutant nanopore shows the locations of the arginines (at residue 113) and the cysteines. (C) Exonuclease sequencing: A processive enzyme is attached to the top of the nanopore to cleave single nucleotides from the target DNA strand and pass them through the nanopore. (D) A residual current-vs-time signal trace from an RHL protein nanopore that shows a clear discrimination between single bases (dGMP, dTMP, dAMP, and dCMP). (E) Strand sequencing: ssDNA is threaded through a protein nanopore and individual bases a identified, as the strand remains intact. Panels A, B, and D reprinted with permission from ref 91. Copyright 2009 Nature Publishing Group. Panels C and E reprinted with permission from Oxfor Nanopore Technologies (Zoe McDougall).

From Niedringhaus et al. Analytical Chemistry 83: 4327. 2011.

Nanopores



Nanopore DNA sequencing using electronic measurements and optical readout as detection methods.(A)In electronic nanopo schemes, signal is obtained through ionic current,[73] tunneling current, and voltage difference measurements. Each method must produce a characteristic signal to differentiate the four DNA bases. (B) In the optical readout nanopore design, each nucleotide is converted to a preset oligonucleotide sequence and hybridized with labeled markers that are detected during translocation of the DNA fragment through the nanopore.

From Niedringhaus et al. Analytical Chemistry 83: 4327. 2011.

Oxford Nanopore

From Oxford Nanopores Web Site

BlueSEQ

Other Platforms...