Disclaimer

This material is only for internal use and is given to the students to prepare the final evaluation for the course of Applied Genomics of the Master degree course of Bioinformatics.

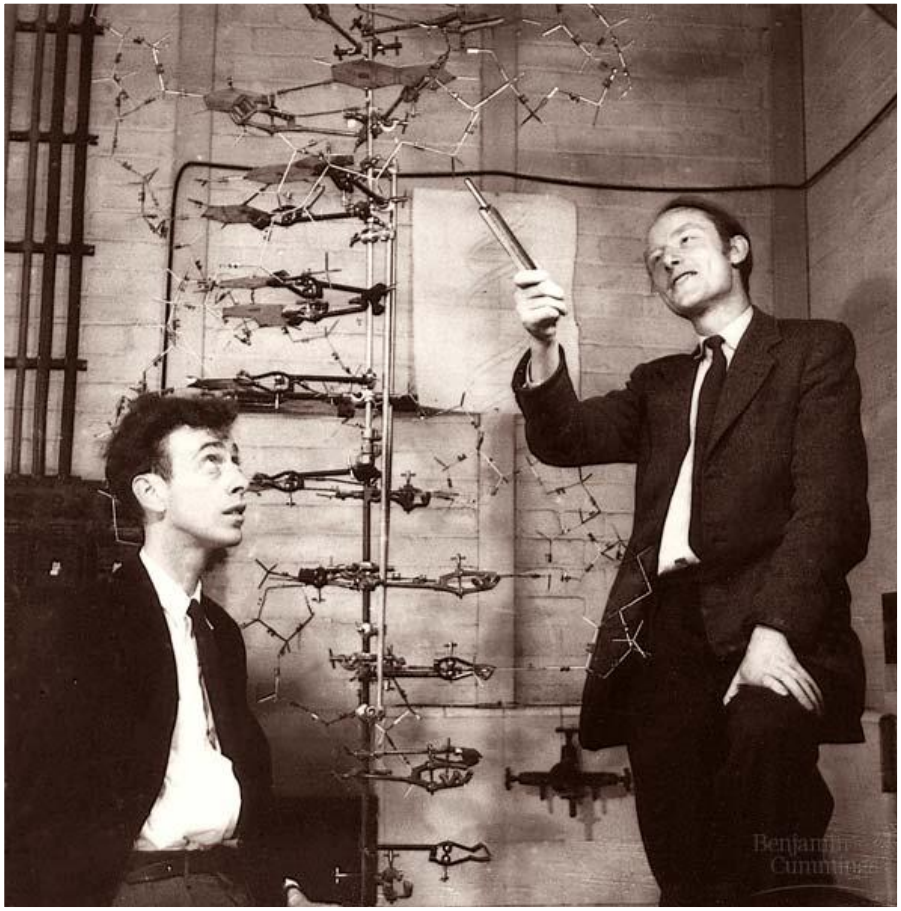
This file and its content are confidential and intended solely for the use of the individuals to whom they are given. If you have received this file it means that you are a student of the course of Applied Genomics of the master degree course in Bioinformatics, regularly enrolled for the current academic year. If you are not a student of this course you should not disseminate, distribute or copy this file. Please notify the professor immediately by e-mail if you have received this file.
In any case, also students of this course are notified that disclosing, copying, distributing or taking any action in reliance on the contents of this information is strictly prohibited.

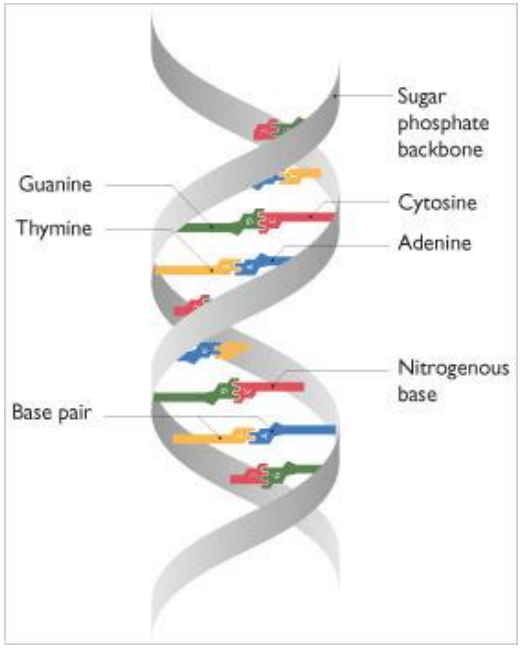
For the students: please note that the content of this file is not enough to pass the exam. The content of this file could contain a few errors as it has not been peer reviewed or edited after its preparation.

Applied Genomics

Program of the course:

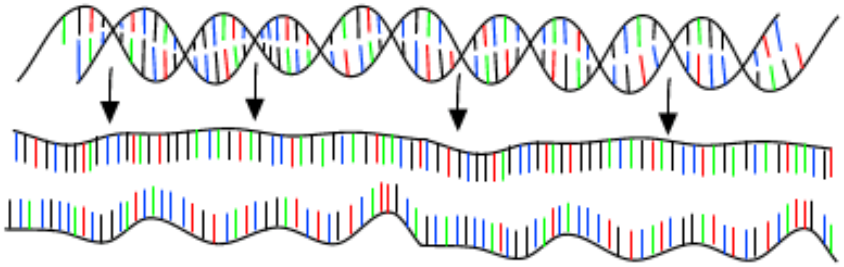
1) Foundational concepts in genetics (including population and quantitative genetics) and genomics.
2) Genome structure and variability in vertebrates
3) The transcriptional landscape of the mammalian genome
4) High throughput technologies for genotyping and next generation sequencing (NGS) platforms
5) Applications of NGS, array comparative genome hybridization
6) Linkage disequilibrium and linkage analysis, genetic mapping
7) QTL mapping, eQTL
8) Candidate gene analysis, genome wide association studies, selection signature
9) Relevant genomic projects: modENCODE, ENCODE, 1000 genome project, The Mammalian Genome Project, 10K Genome project.
10) Discussion of relevant scientific literature




Sugar phosphate backbone
Guanine
Cytosine
Thymine
Adenine
Nitrogenous base
Base pair

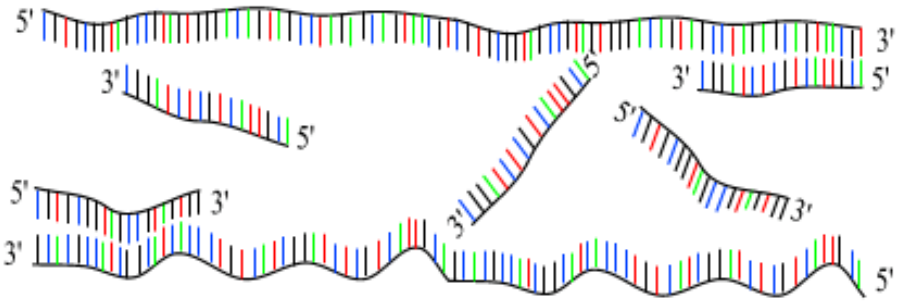
PCR : Polymerase Chain Reaction

30 - 40 cycles of 3 steps :



Step 1 : denaturation

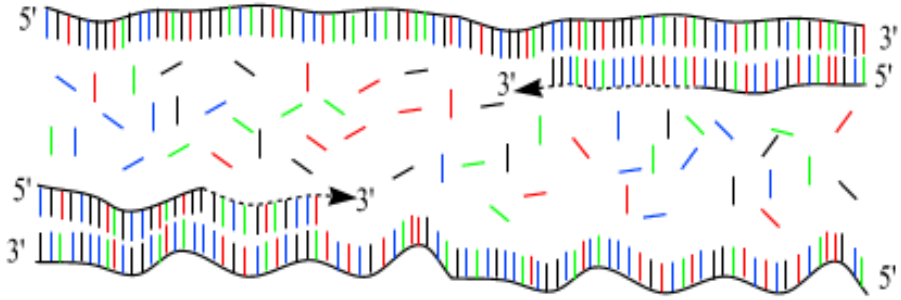
1 minut 94 °C



Step 2 : annealing

45 seconds 54 °C

forward and reverse primers !!!



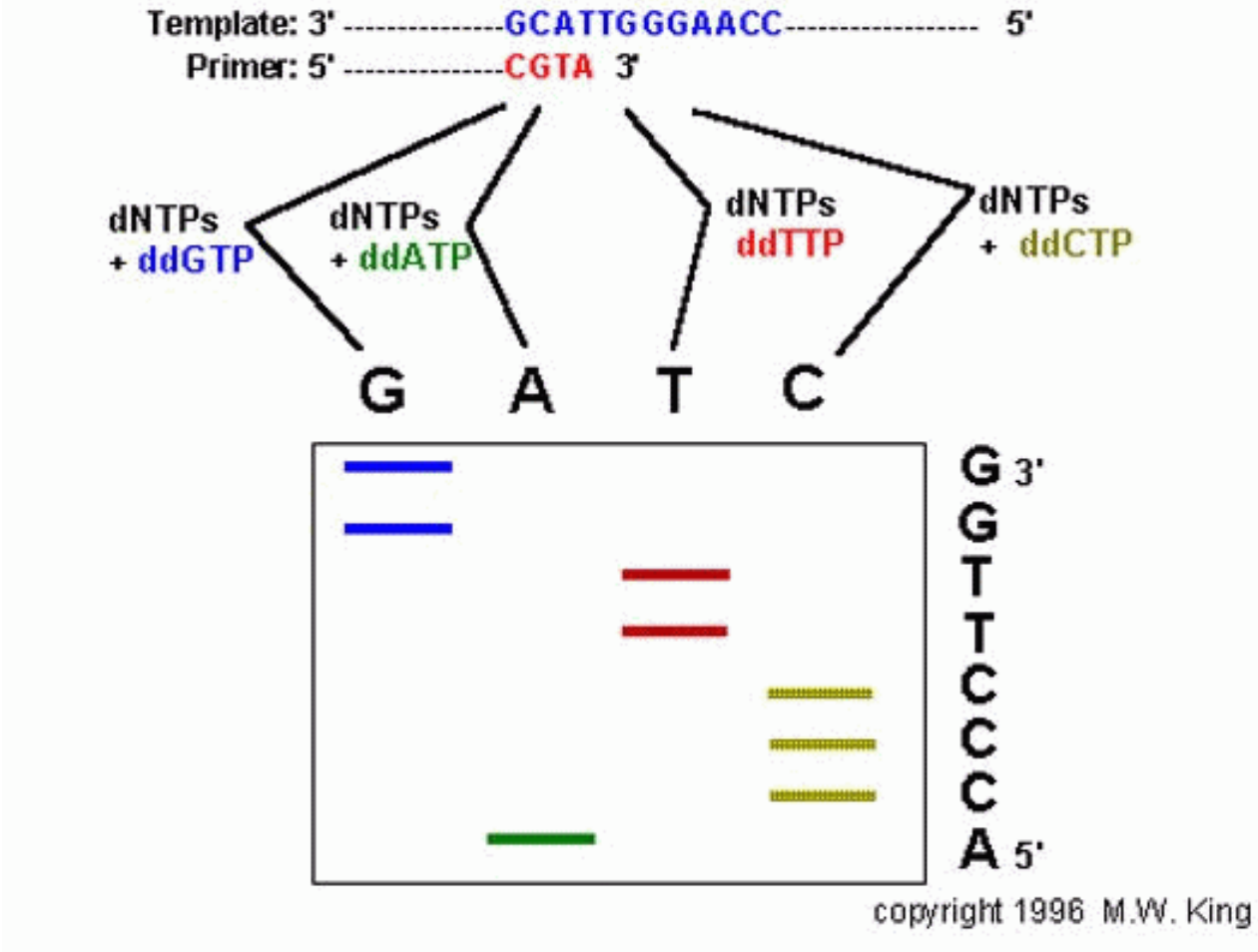
Step 3 : extension

2 minutes 72 °C

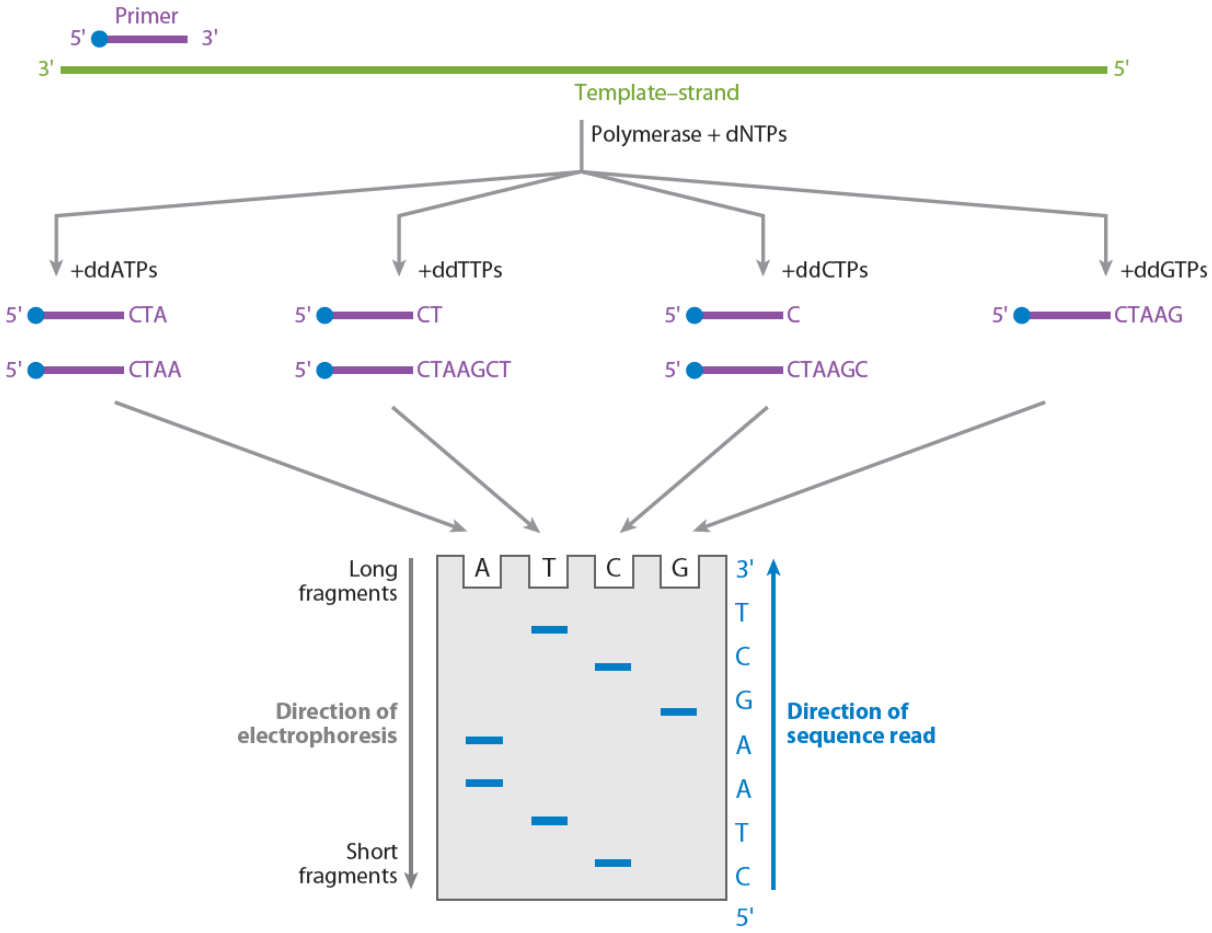
only dNTP's

(Andy Vierstraete 1999)

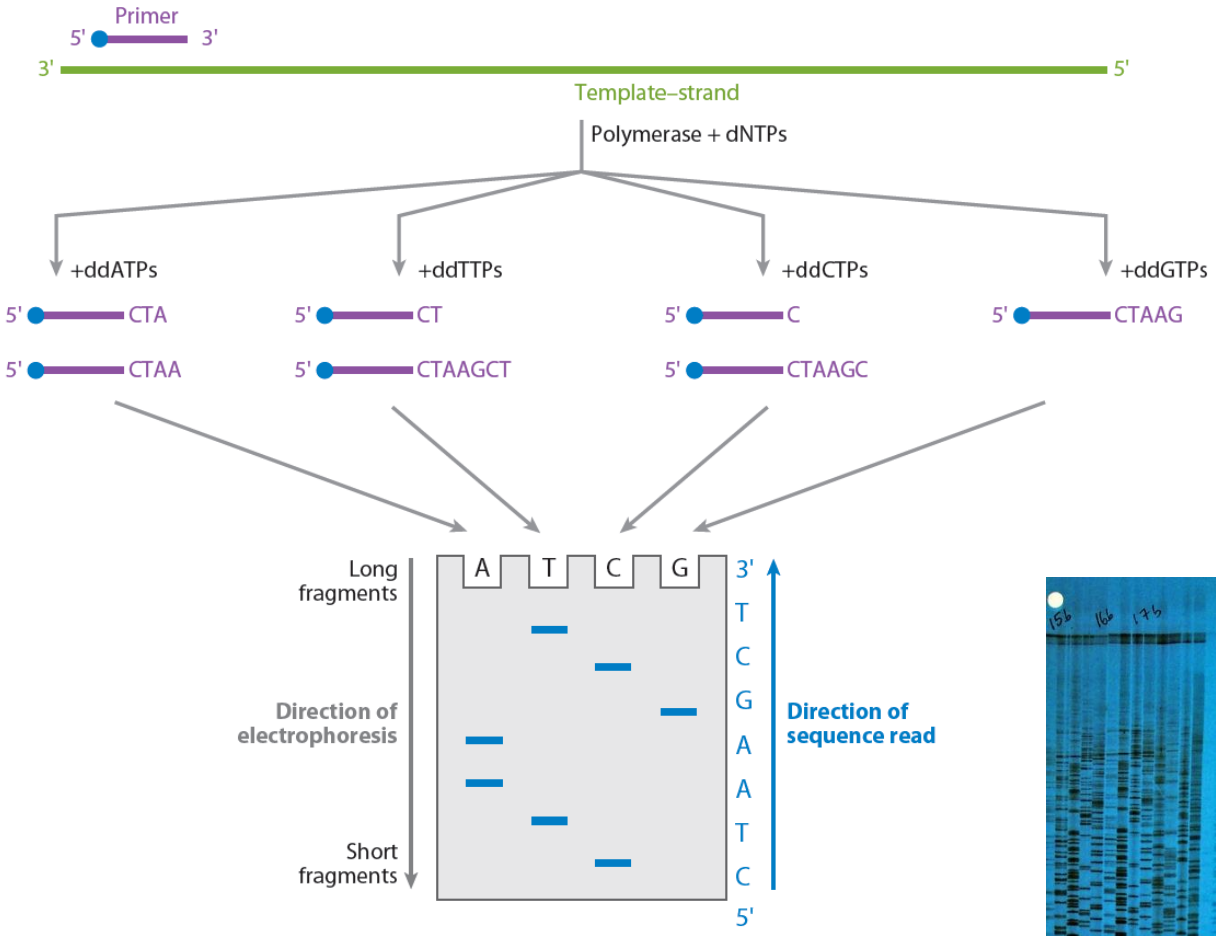
Sanger ddNTP Chain Termination Sequencing

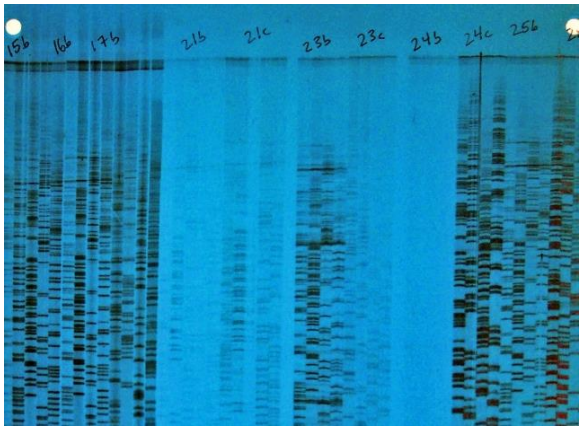


Sanger dideoxy sequencing = first generation sequencing



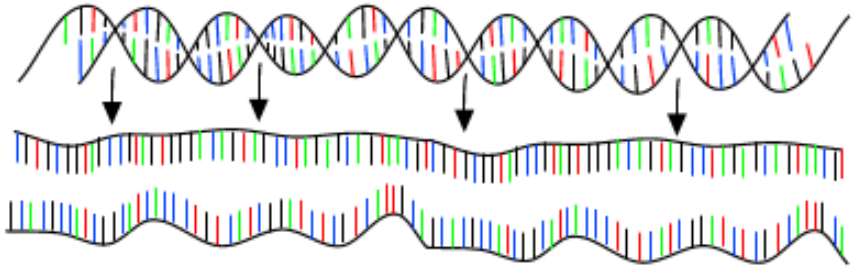
Sanger dideoxy sequencing = first generation sequencing





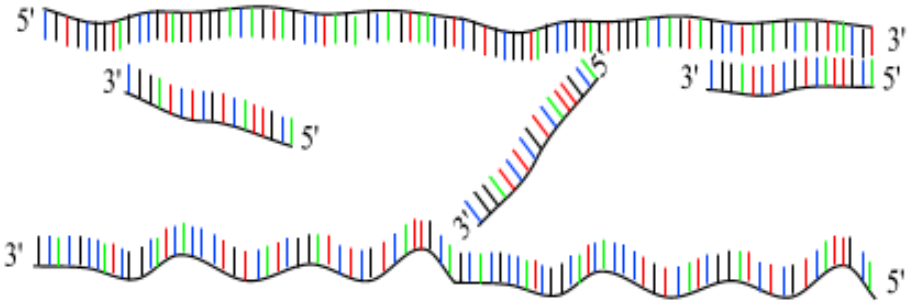
Sequencing

30 cycles of 3 steps :



Step 1 : denaturation

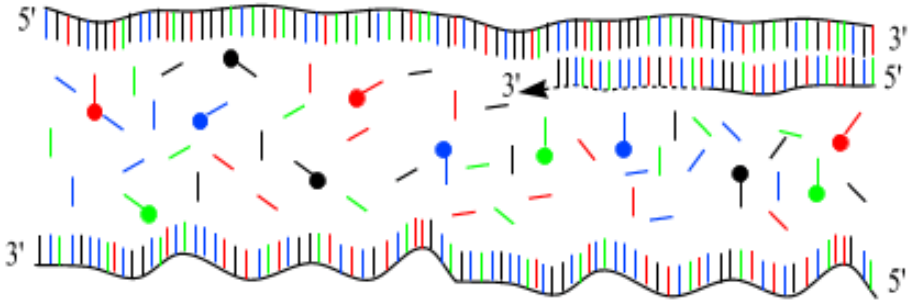
1 minut 94 °C



Step 2 : annealing

15 seconds 50 °C

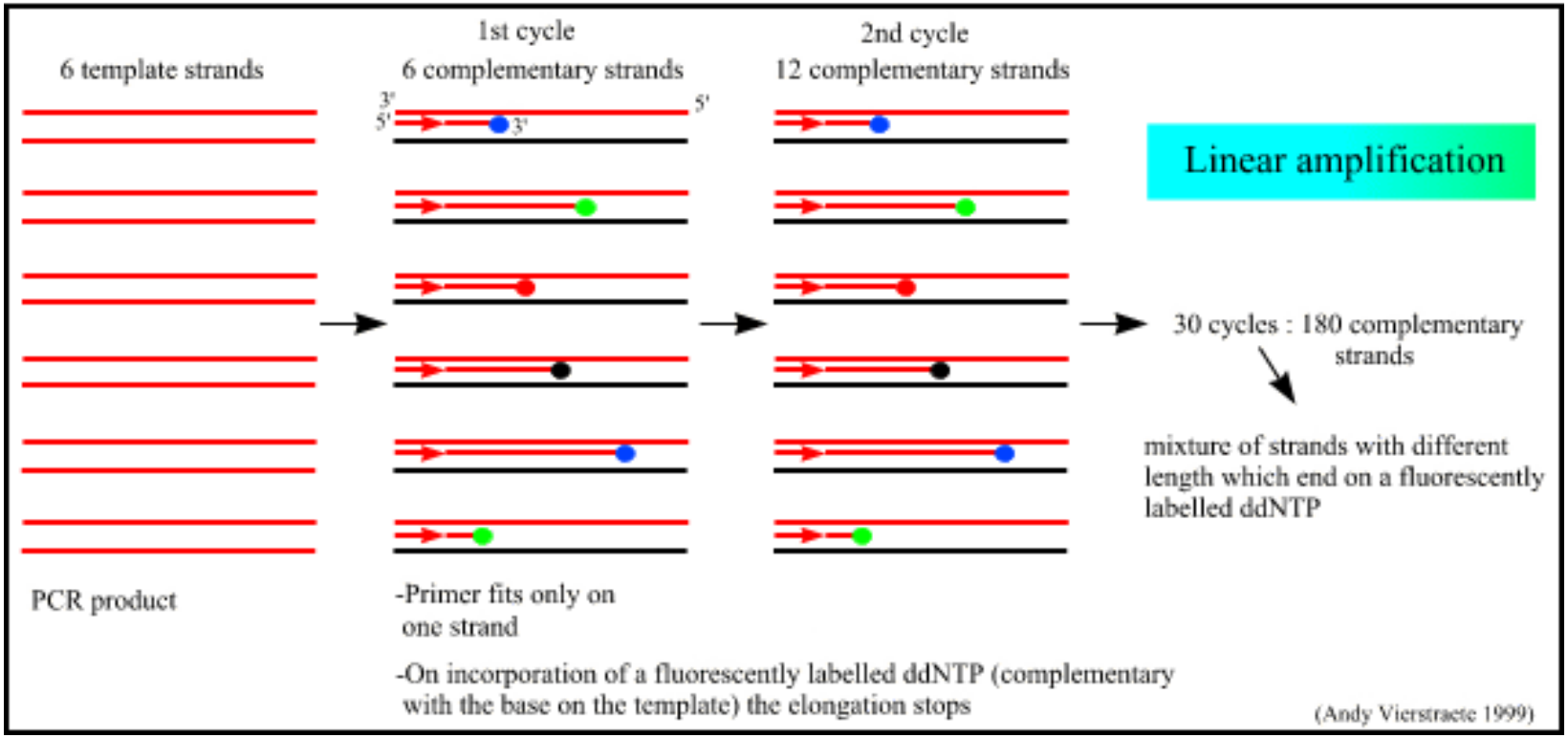
1 primer !!!!

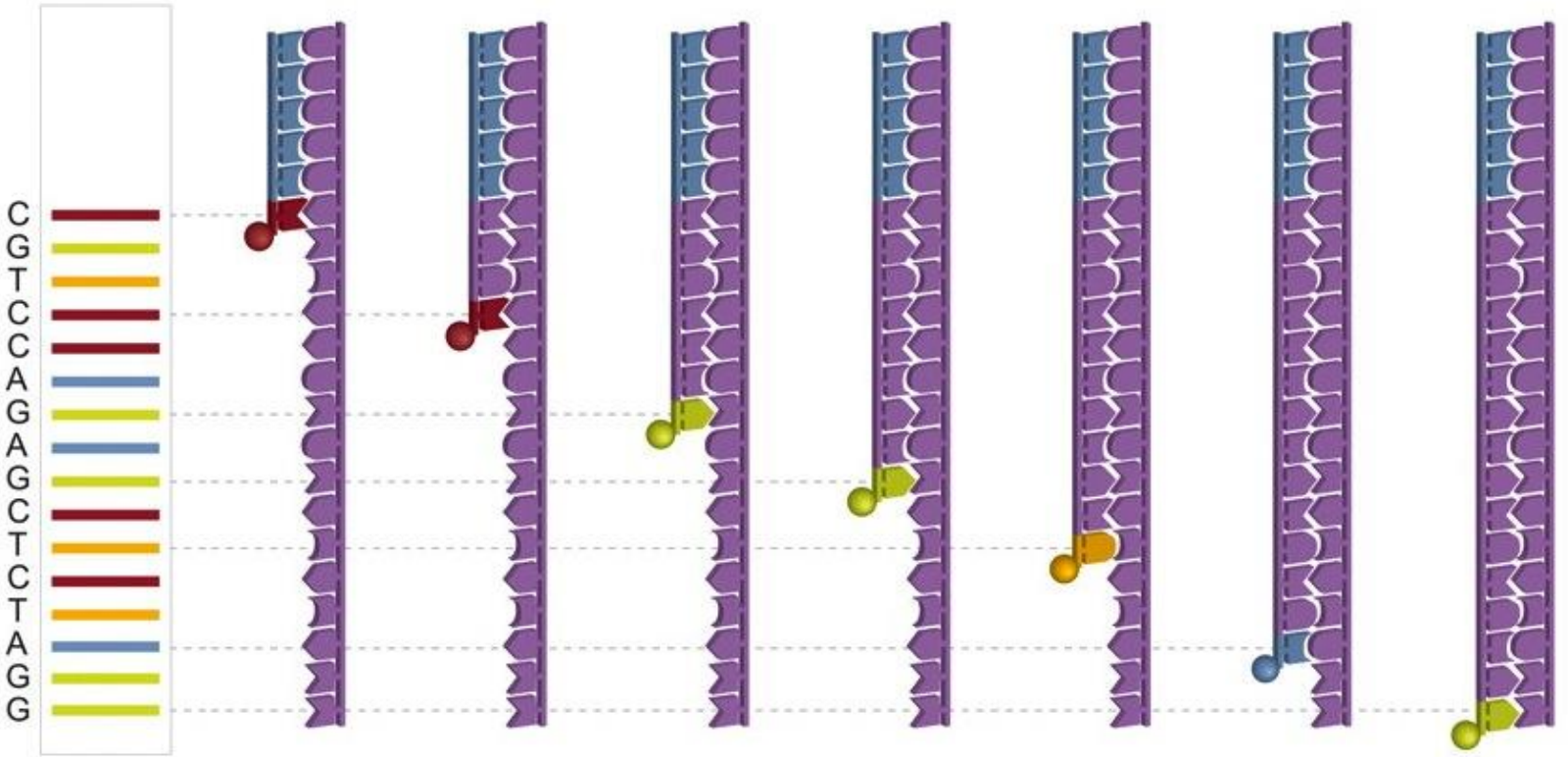


Step 3 : extension

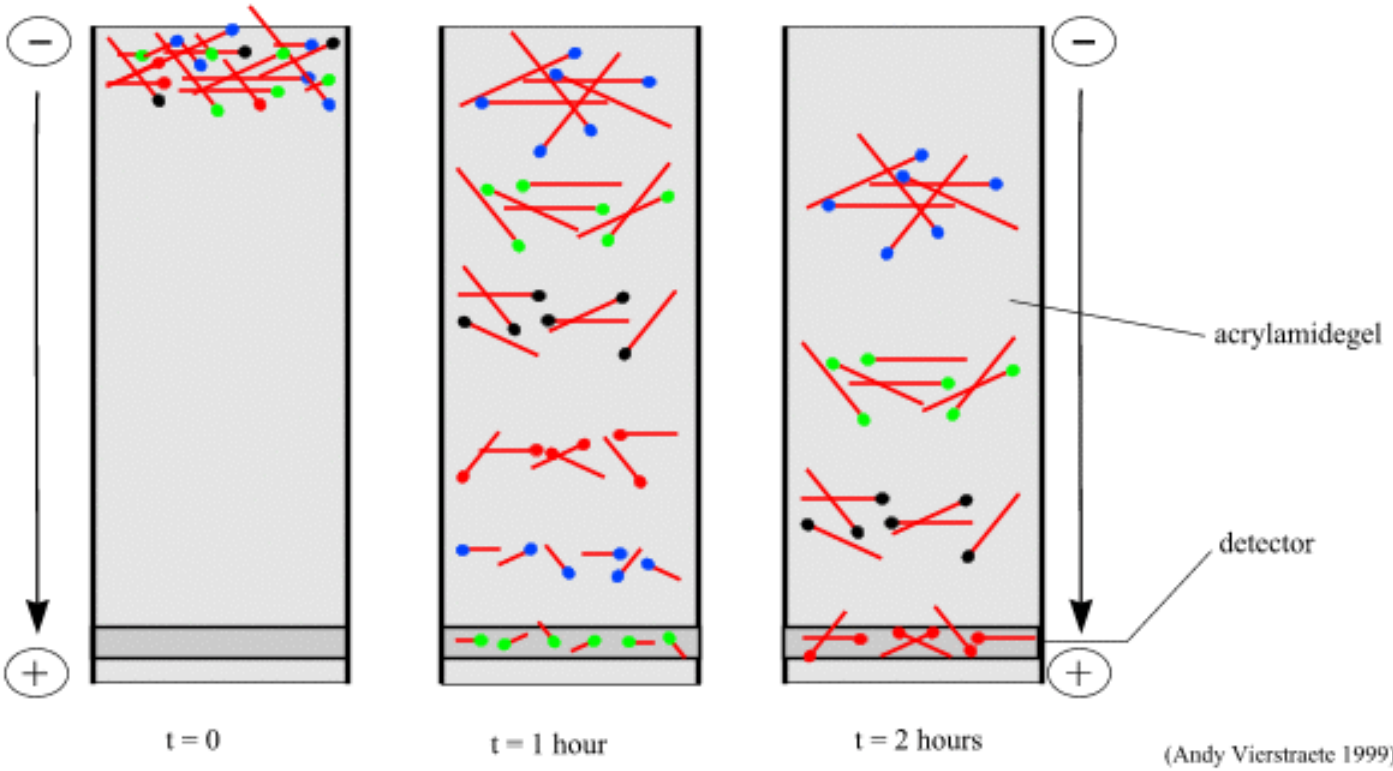
4 minutes 60 °C
mixture of dNTP's
and ddNTP's

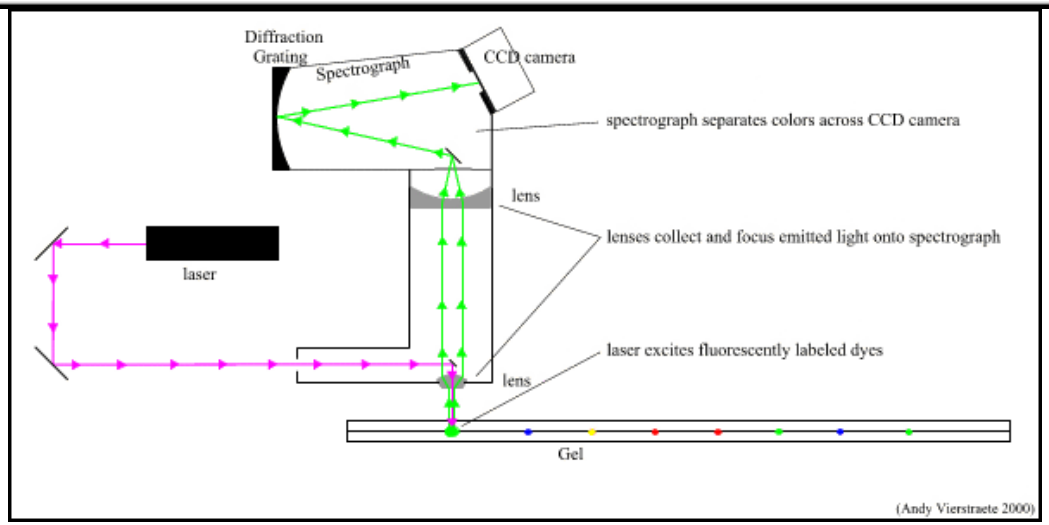
(Andy Vierstraete 1999)

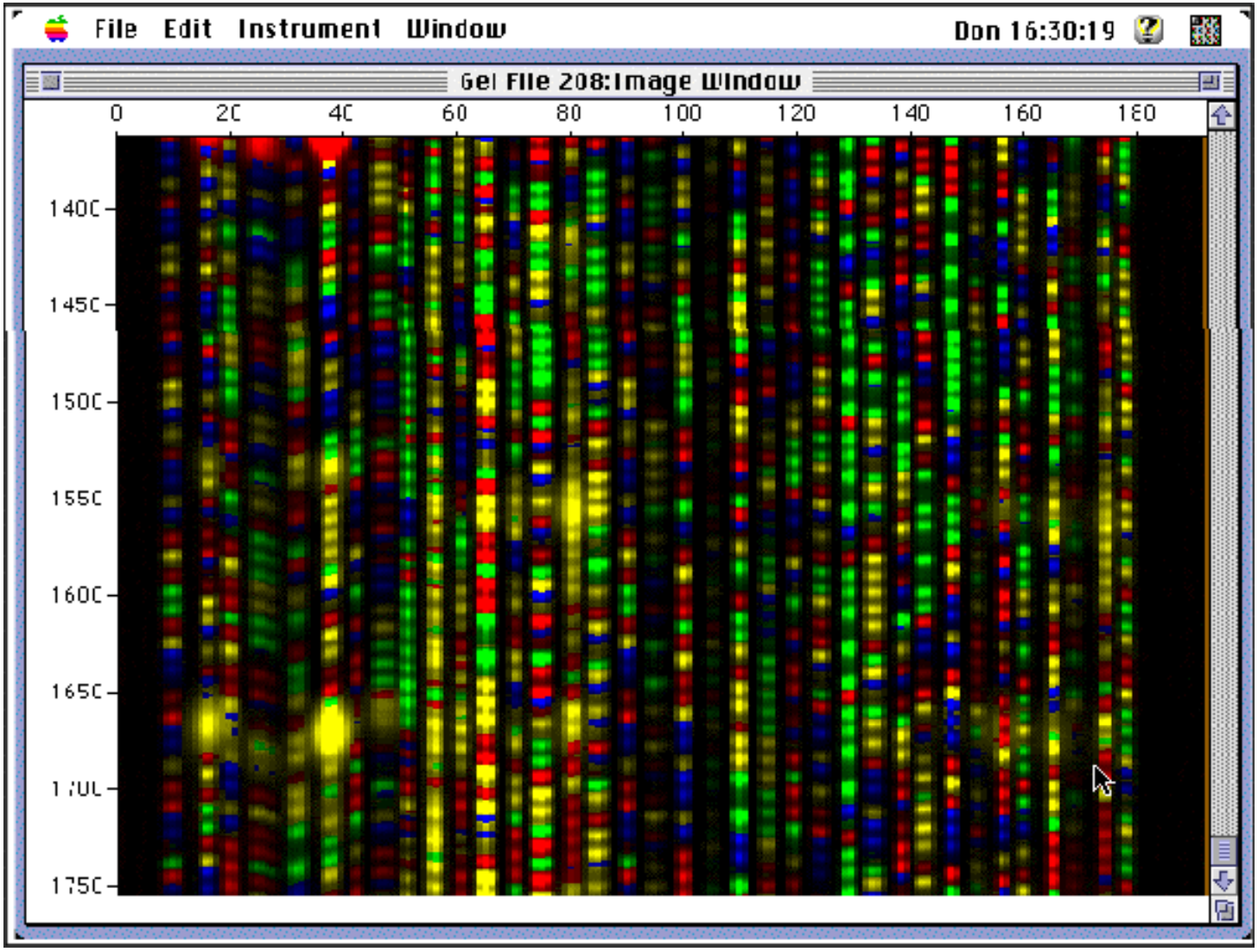

6 template strands
1st cycle
6 complementary strands
2nd cycle
12 complementary strands
3'
5'
5'
3'
Linear amplification
30 cycles : 180 complementary strands
mixture of strands with different length which end on a fluorescently labelled ddNTP
PCR product
-Primer fits only on one strand
-On incorporation of a fluorescently labelled ddNTP (complementary with the base on the template) the elongation stops
(Andy Vierstraete 1999)


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

Gel electrophoresis

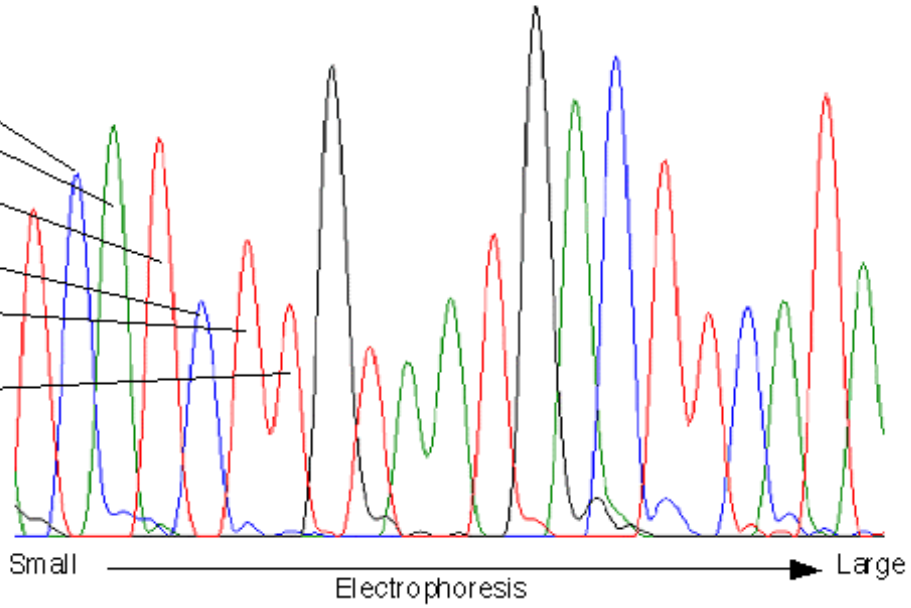


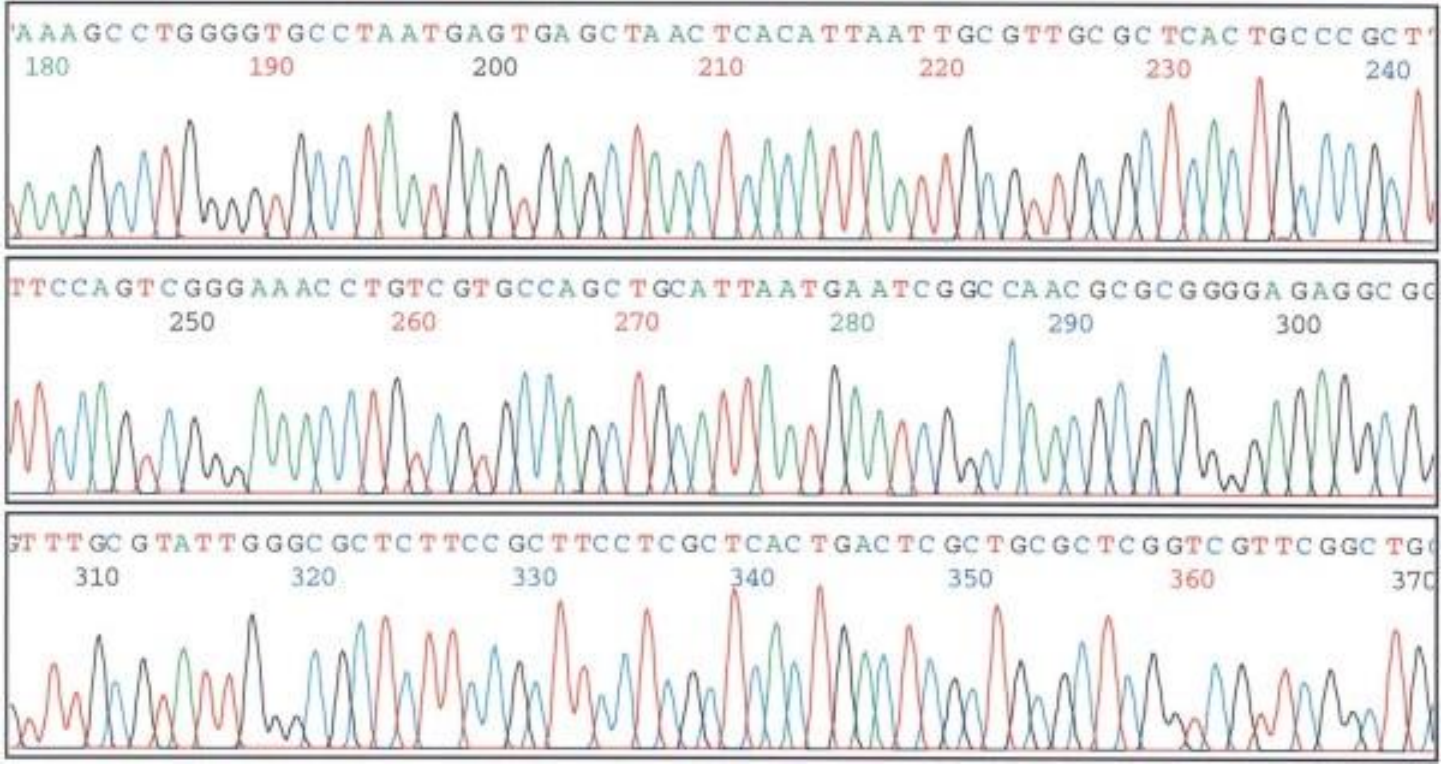



File Edit Instrument Window
Don 16:30:19
Gel File 208:Image Window
0
20
40
60
80
100
120
140
160
180
140
145
150
155
160
165
175

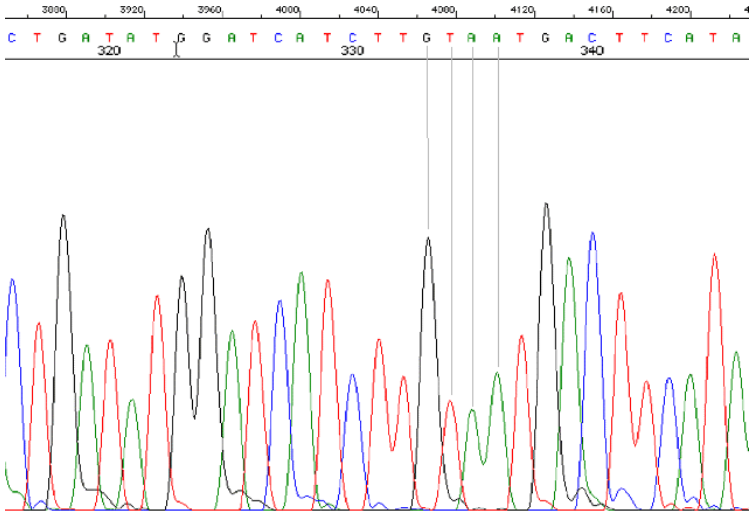
```
5'-Label-CTAGGCTC
        ::::::::
      3'-GATCCGAGTAGAACATTACTGAAG-5'

5'-Label-CTAGGCTCA
        :::::::::
      3'-GATCCGAGTAGAACATTACTGAAG-5'

5'-Label-CTAGGCTCAT
        ::::::::::
      3'-GATCCGAGTAGAACATTACTGAAG-5'

5'-Label-CTAGGCTCATC
        :::::::::::
      3'-GATCCGAGTAGAACATTACTGAAG-5'

5'-Label-CTAGGCTCATCT
        ::::::::::::
      3'-GATCCGAGTAGAACATTACTGAAG-5'

5'-Label-CTAGGCTCATCTT
        :::::::::::::
      3'-GATCCGAGTAGAACATTACTGAAG-5'
```

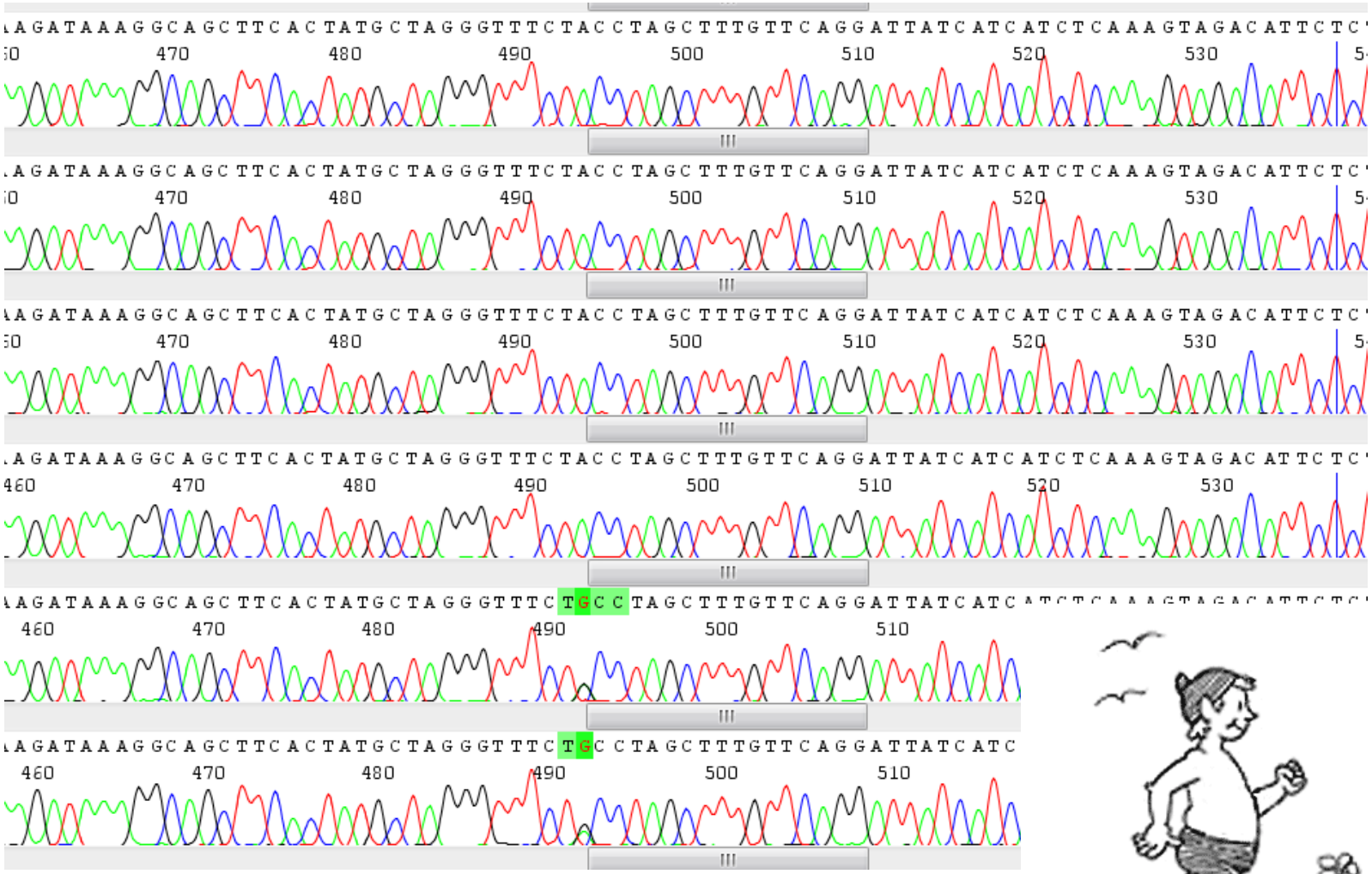
More typically now, sequencing reactions are denatured and the products are separated in a single gel lane or a single capillary tube. The products of the four reactions are labeled with a different fluorescent dye, and a single detector at the bottom of the apparatus detects the fluors as they emerge. The sequence can be read (automatically) from left to right.



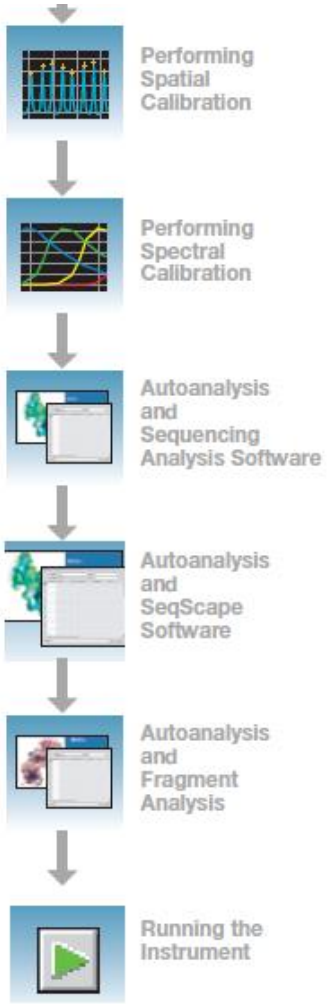


Output from automated sequencing

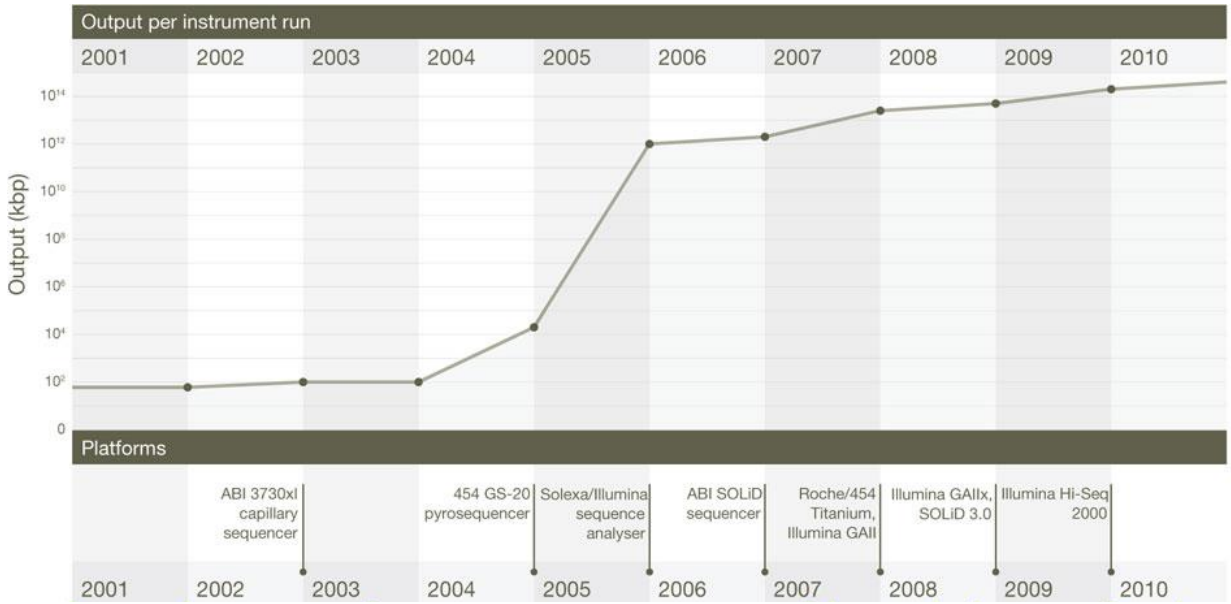



AGATAAAGGCAGCTTCACTATGCTAGGGTTTCTACCTAGCTTTGTTCAGGATTATCATCATCTCAAAGTAGACATTCTC
470
480
490
500
510
520
530
AGATAAAGGCAGCTTCACTATGCTAGGGTTTCTACCTAGCTTTGTTCAGGATTATCATCATCTCAAAGTAGACATTCTC
470
480
490
500
510
520
530
AGATAAAGGCAGCTTCACTATGCTAGGGTTTCTACCTAGCTTTGTTCAGGATTATCATCATCTCAAAGTAGACATTCTC
470
480
490
500
510
520
530
AGATAAAGGCAGCTTCACTATGCTAGGGTTTCTACCTAGCTTTGTTCAGGATTATCATCATCTCAAAGTAGACATTCTC
460
470
480
490
500
510
520
530
AAGATAAAGGCAGCTTCACTATGCTAGGGTTTCTGCCTAGCTTTGTTCAGGATTATCATC
460
470
480
490
500
510
AAGATAAAGGCAGCTTCACTATGCTAGGGTTTCTGCCTAGCTTTGTTCAGGATTATCATC
460
470
480
490
500
510

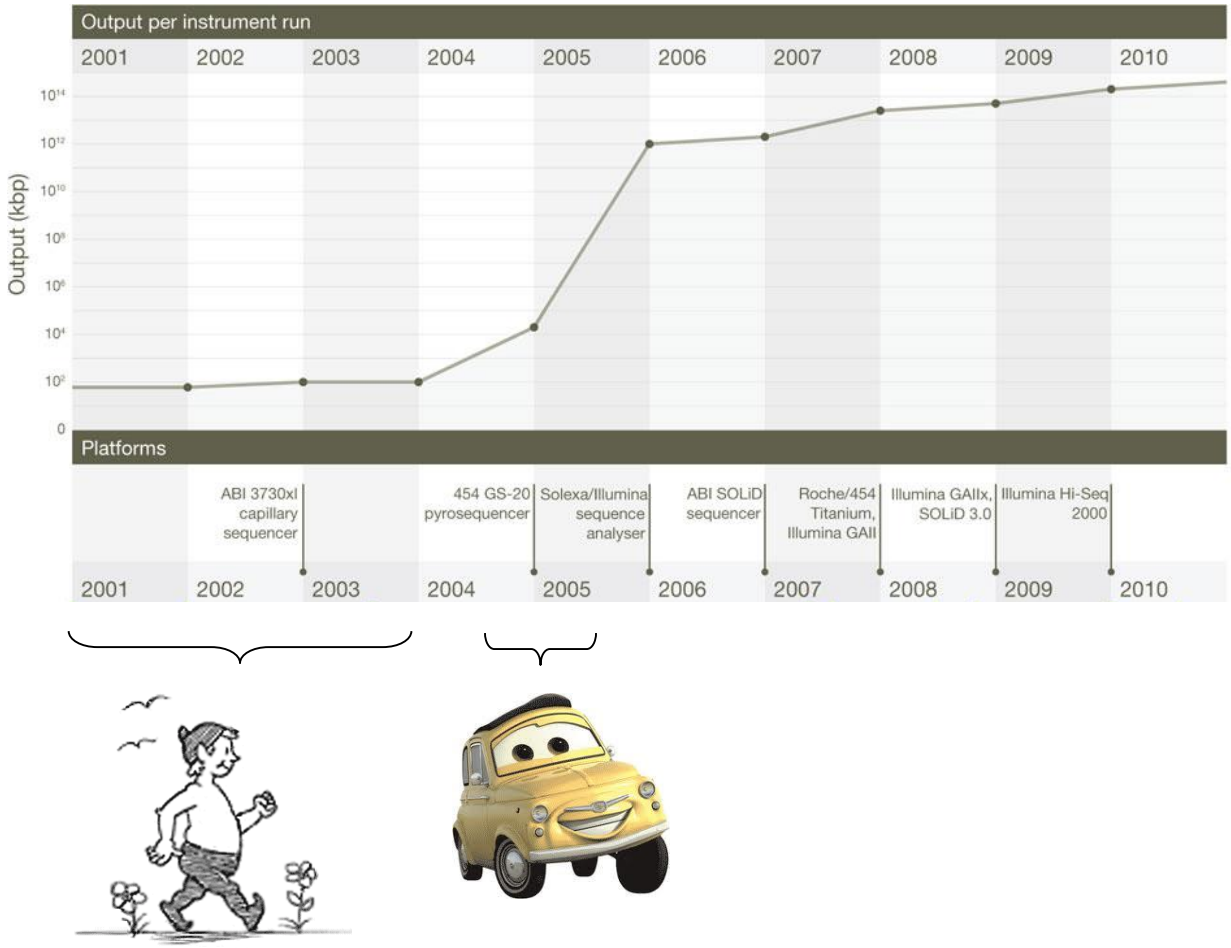
Sanger dideoxy sequencing = first generation sequencing



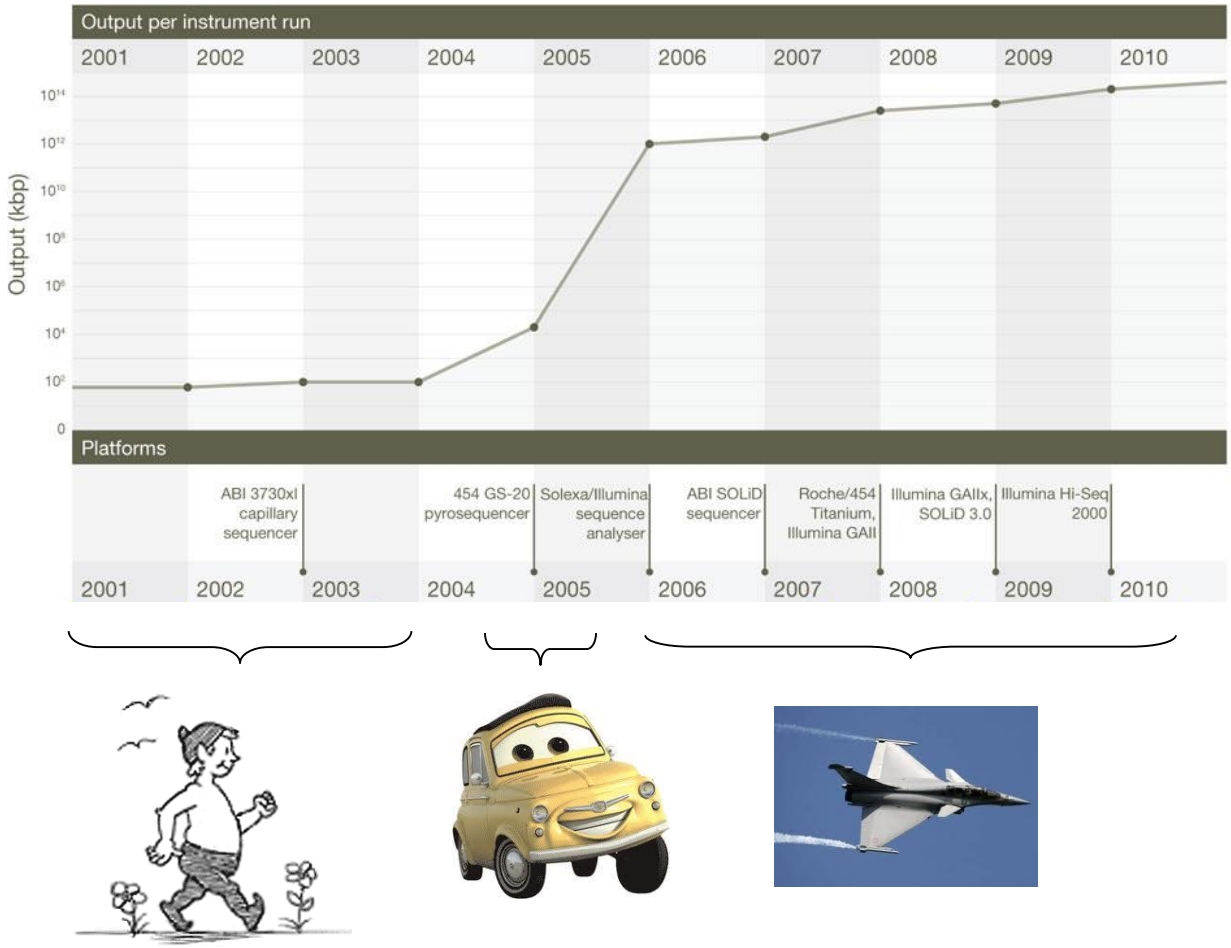
Mardis (2011) *Nature* **470**, 198-203



Mardis (2011) *Nature* **470**, 198-203

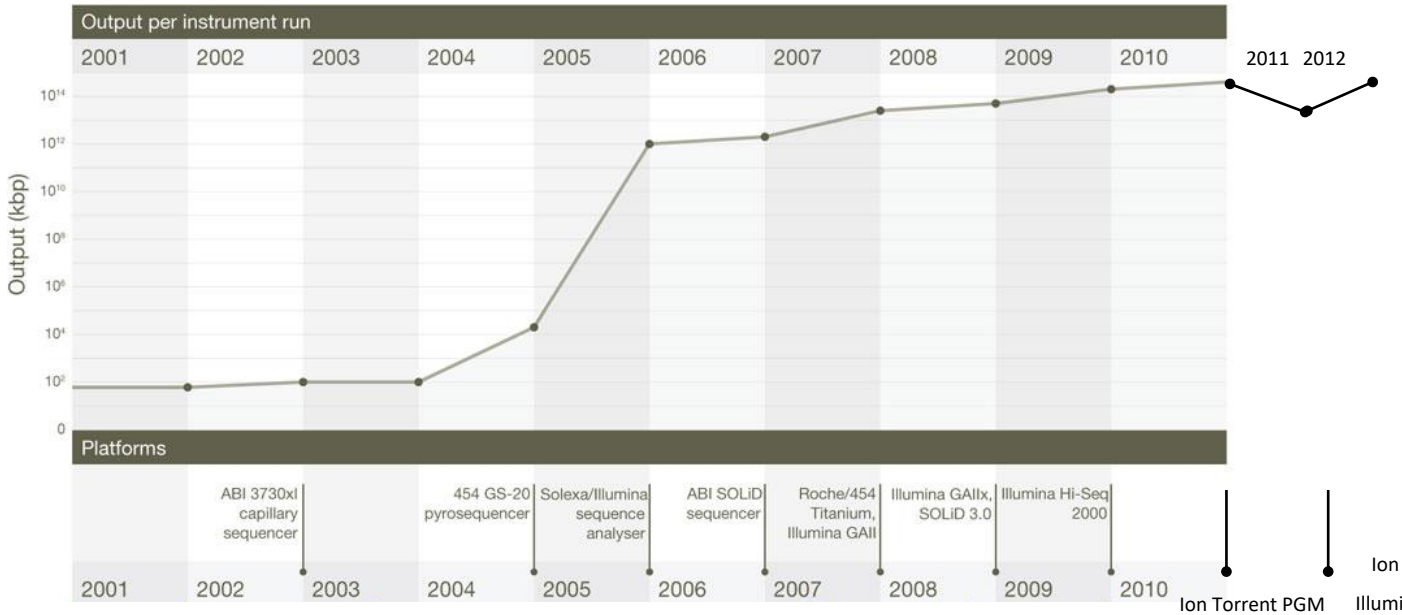


Mardis (2011) *Nature* **470**, 198-203



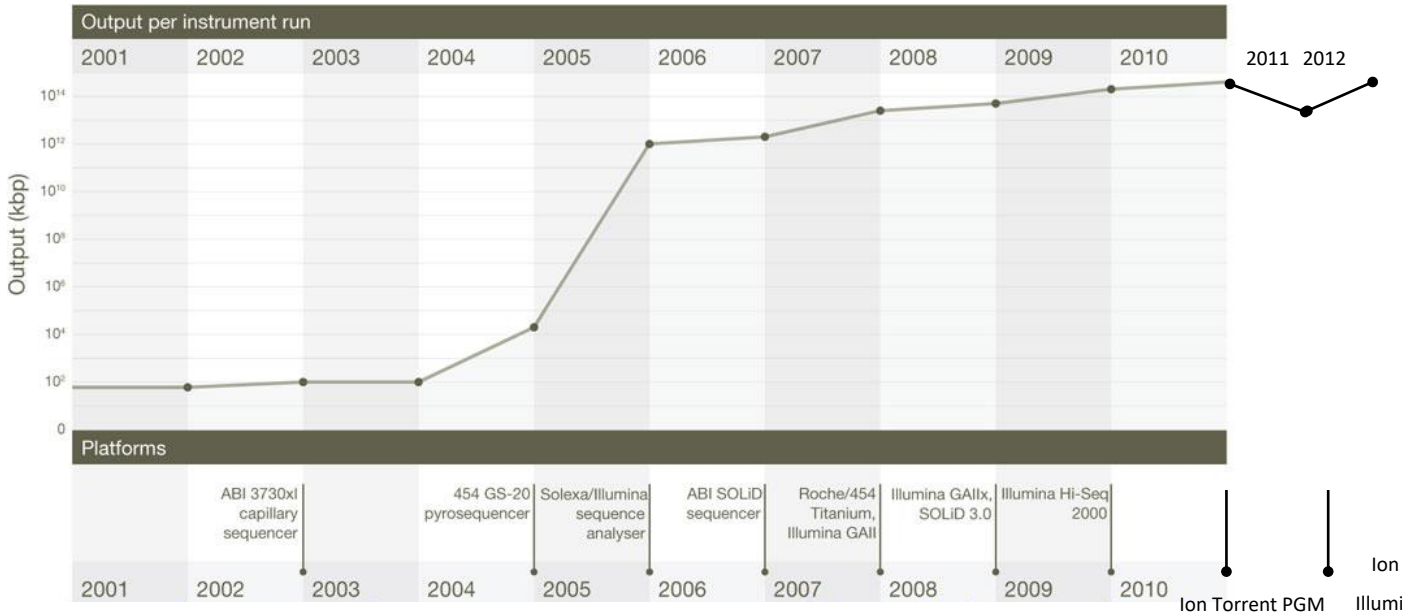
Mardis (2011) *Nature* **470**, 198-203

Changes in instrument capacity over the past decade

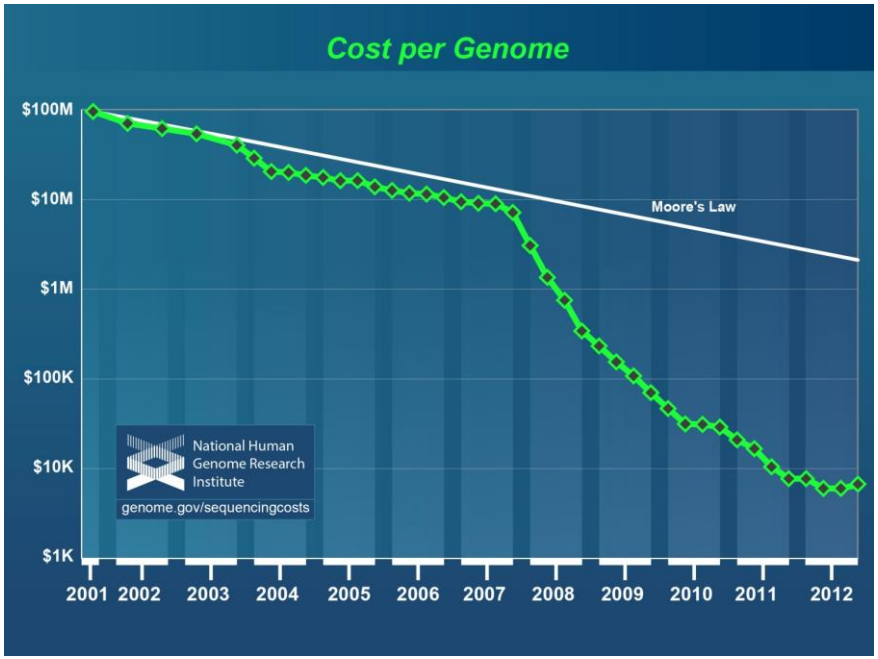


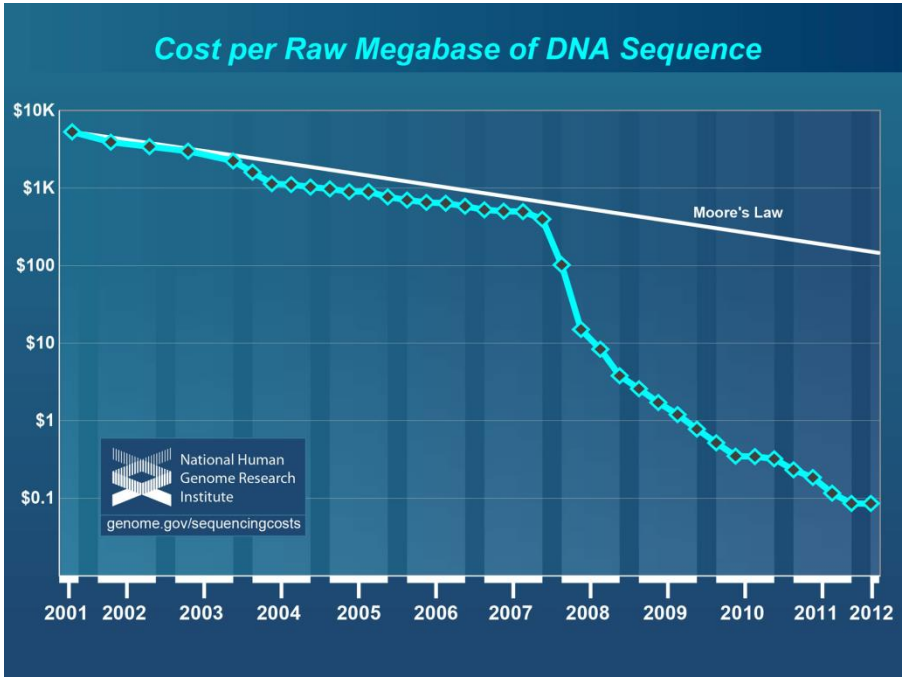
Mardis (2011) *Nature* **470**, 198-203

Changes in instrument capacity over the past decade



NGS


Cost per Genome
$100M
$10M
$1M
$100K
$10K
$1K
Moore's Law
National Human Genome Research Institute
genome.gov/sequencingcosts
2001 2002 2003 2004 2005 2006 2007 2008 2009 2010 2011 2012


Cost per Raw Megabase of DNA Sequence
$10K
$1K
$100
$10
$1
$0.1
Moore's Law
National Human Genome Research Institute
genome.gov/sequencingcosts
2001 2002 2003 2004 2005 2006 2007 2008 2009 2010 2011 2012

Experiments

Time x Effort = Cost

Effort

Planning

Data Production

Data Analysis

Time

Experiments

Time x Effort = Cost

ThermoFisher SCIENTIFIC

illumina®

Oxford NANOPORE Technologies

454 Roche

PACBIO®

ThermoFisher SCIENTIFIC

ARTICLE

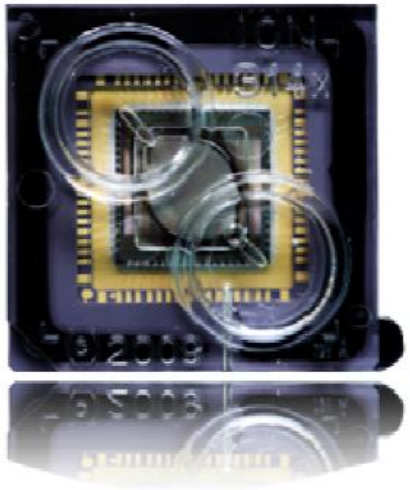
doi:10.1038/nature10242

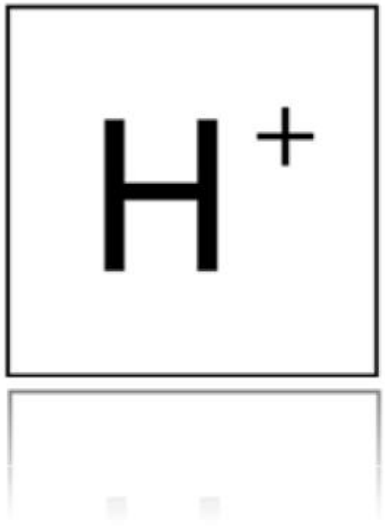
An integrated semiconductor device enabling non-optical genome sequencing

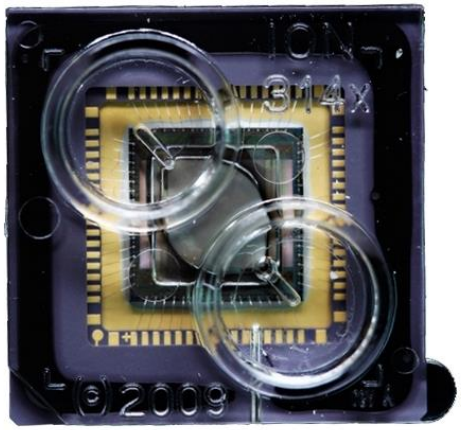
Jonathan M. Rothberg[1], Wolfgang Hinz[1], Todd M. Rearick[1], Jonathan Schultz[1], William Mileski[1], Mel Davey[1], John H. Leamon[1], Kim Johnson[1], Mark J. Milgrew[1], Matthew Edwards[1], Jeremy Hoon[1], Jan F. Simons[1], David Marran[1], Jason W. Myers[1], John F. Davidson[1], Annika Branting[1], 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]

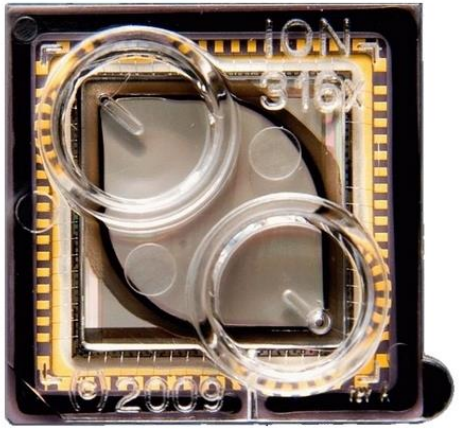
The seminal importance of DNA sequencing to the life sciences, biotechnology and medicine has driven the search for more scalable and lower-cost solutions. Here we describe a DNA sequencing technology in which scalable, low-cost semiconductor manufacturing techniques are used to make an integrated circuit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

ion torrent





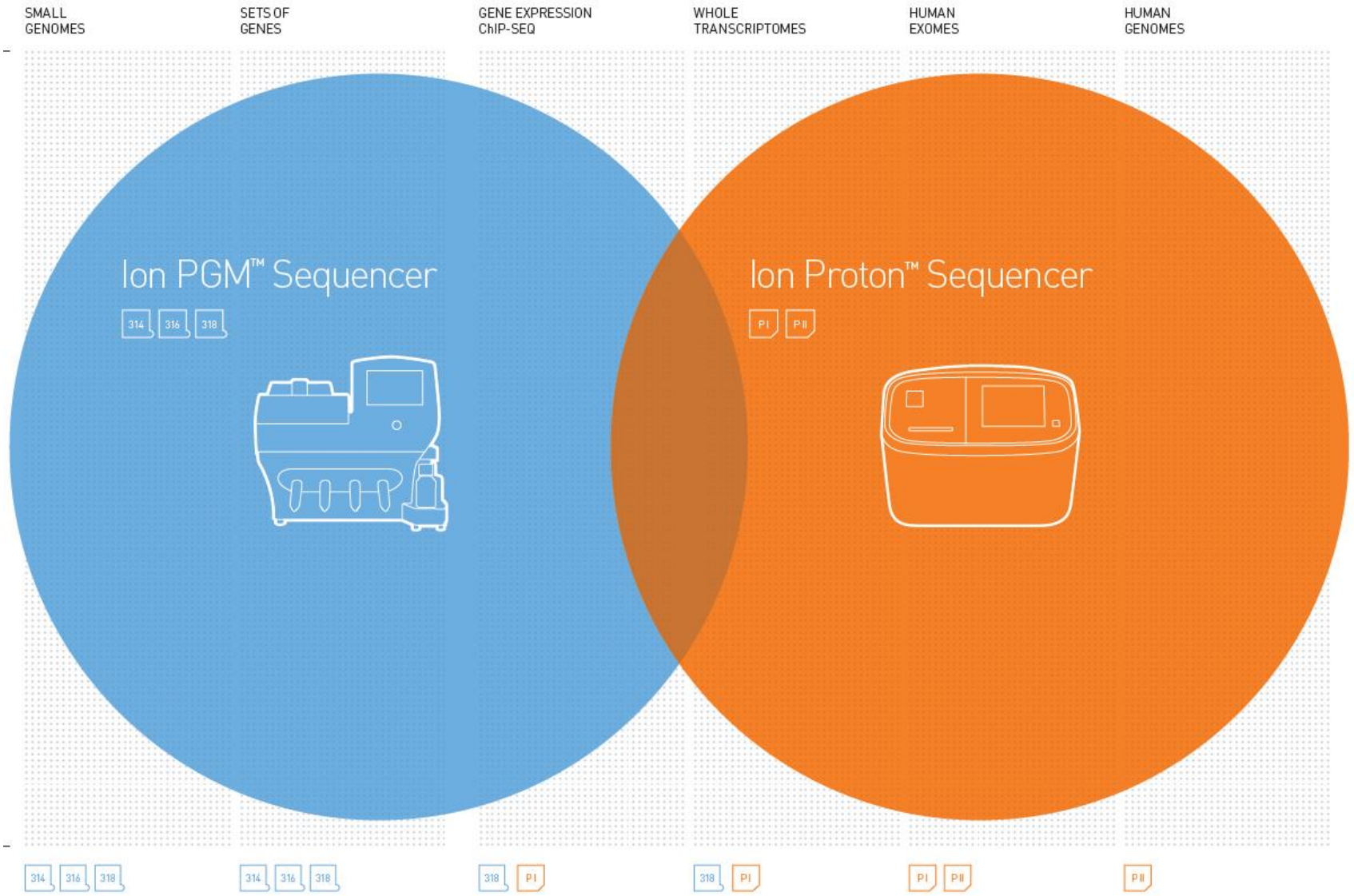




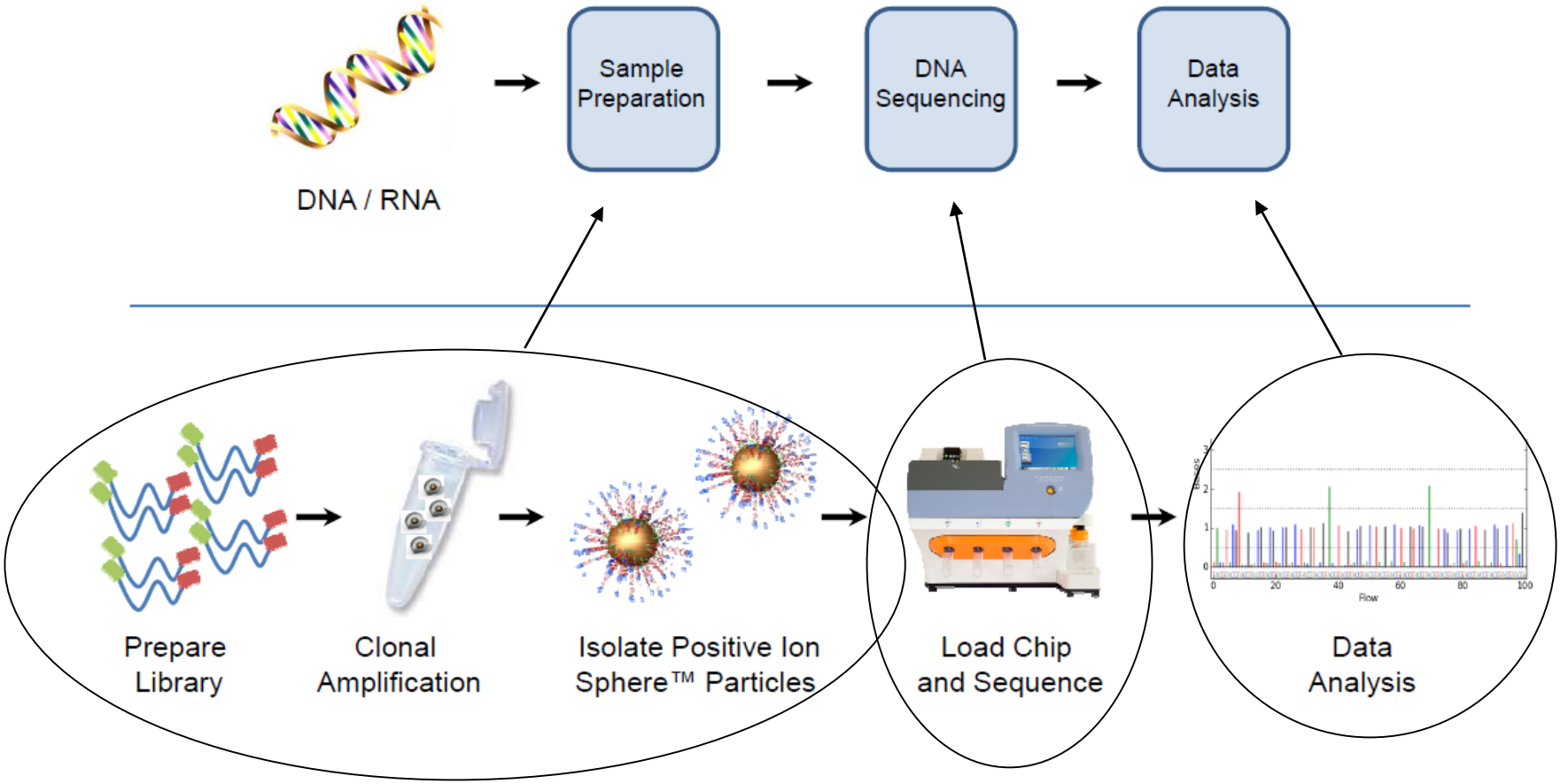
10 Mb

100 Mb

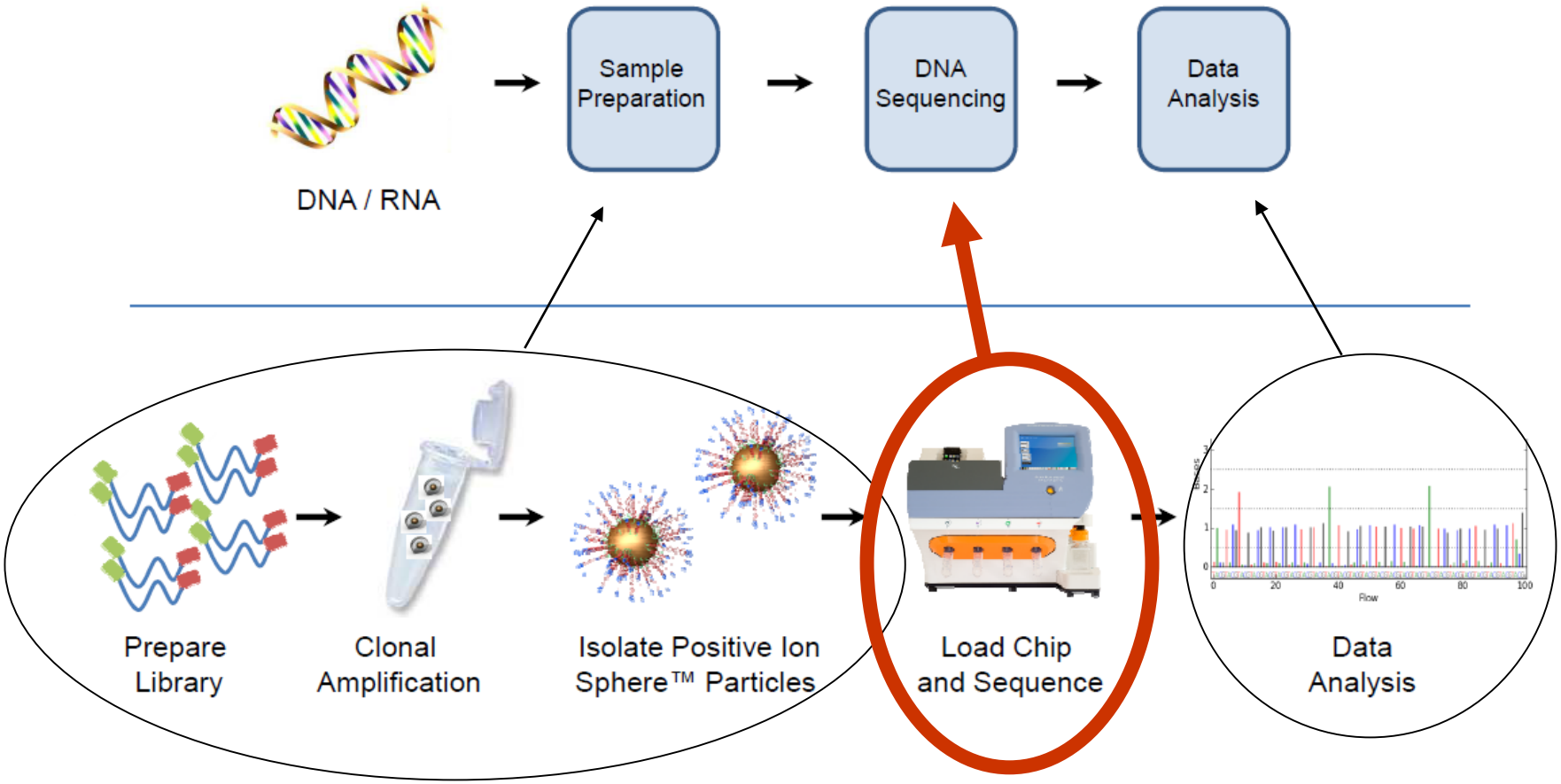
1 Gb

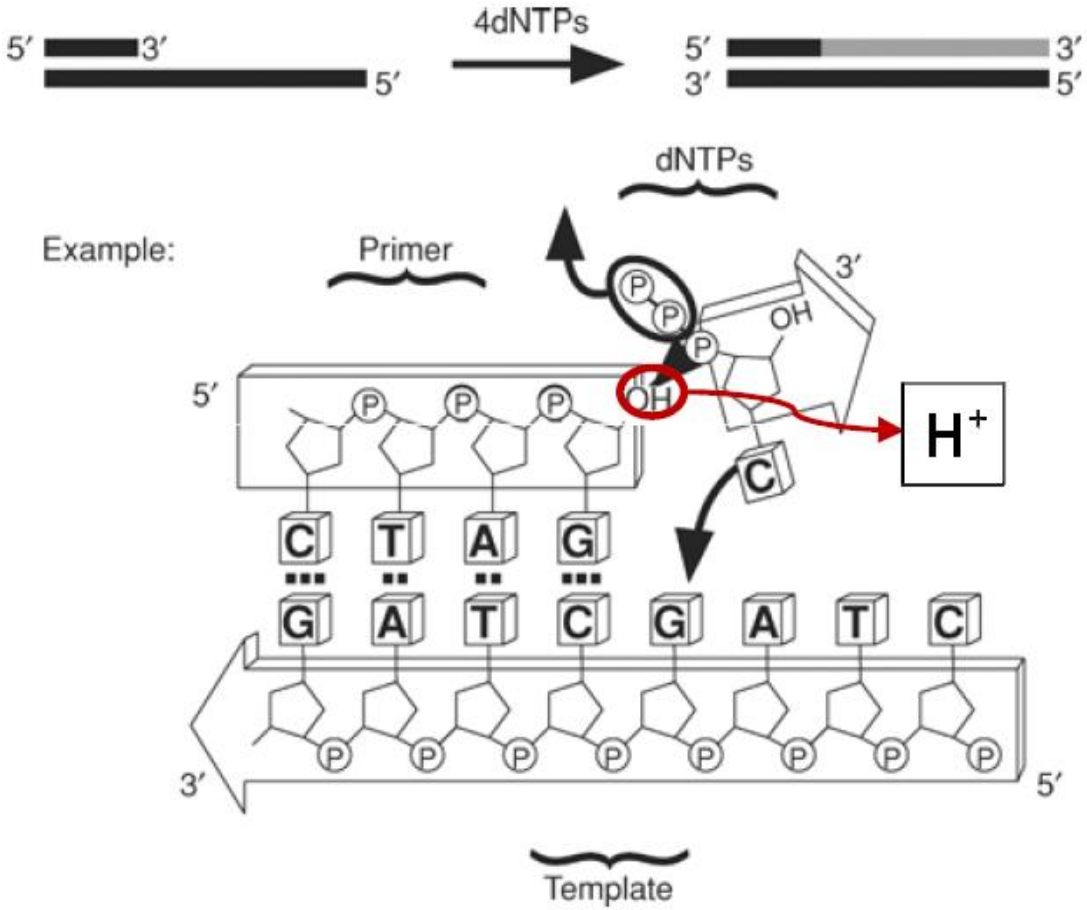

SMALL GENOMES
SETS OF GENES
GENE EXPRESSION ChIP-SEQ
WHOLE TRANSCRIPTOMES
HUMAN EXOMES
HUMAN GENOMES
Ion PGM™ Sequencer
314
316
318
Ion Proton™ Sequencer
PI
PII
314 316 318
314 316 318
318 PI
318 PI
PI PII
PII

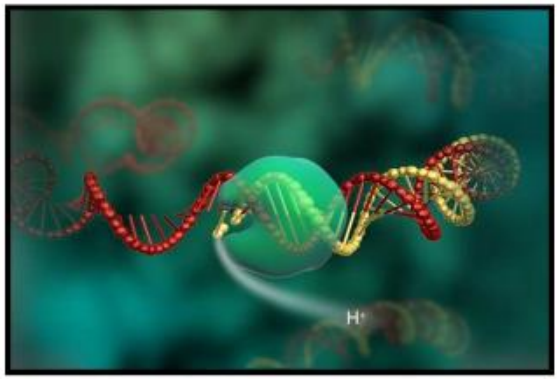
Ion Workflow Overview

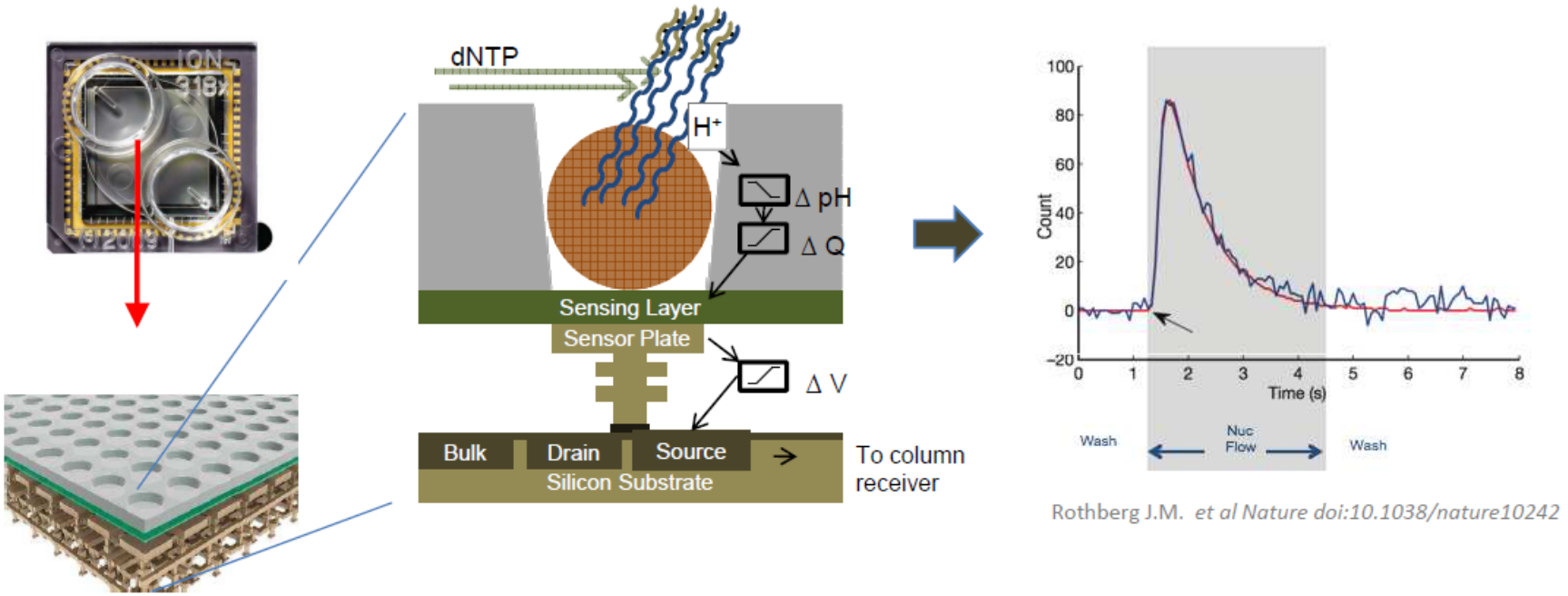


Ion Workflow Overview




5′
3′
5′
4dNTPs
5′
3′
3′
5′
dNTPs
Example:
Primer
3′
OH
5′
OH
H+
C
C T A G
G A T C G A T C
3′
5′
Template

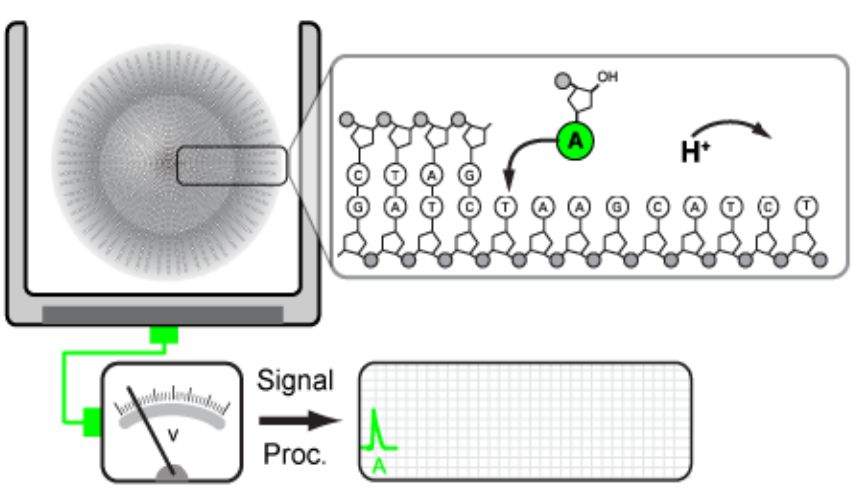


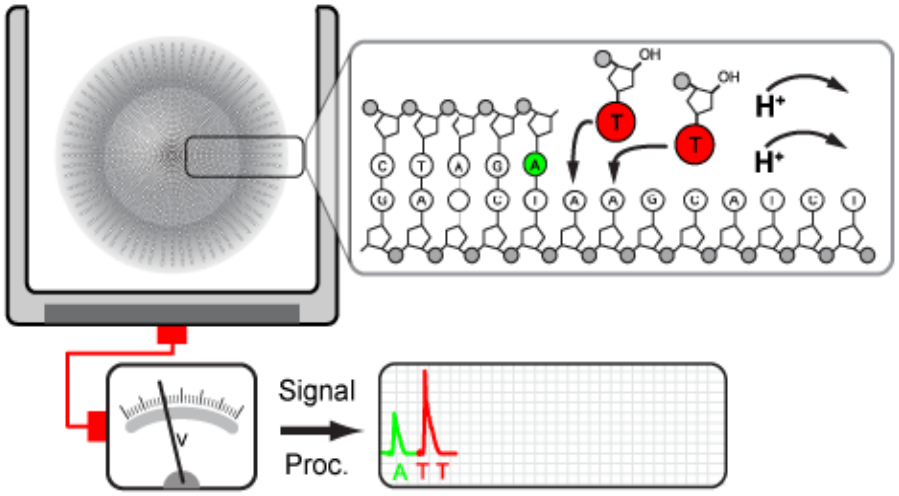


Rothberg J.M. *et al Nature doi:10.1038/nature10242*

DNA → Ions → Sequence

- Nucleotides flow sequentially over Ion semiconductor chip
- One sensor per well per sequencing reaction
- Direct detection of natural DNA extension
- Millions of sequencing reactions per chip
- Fast cycle time, real time detection


A
H+
V
Signal
Proc.
A


T
T
H+
H+
V
Signal
Proc.
A T T

Ion Workflow

A **DNA**

B **Compatible Library Prep**

C **Template Prep**

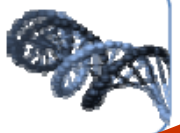
D **Sequencing**

E **Compatible FASTQ Data**

Ion Workflow

A **DNA**

B **Compatible Library Prep**



C **Template Prep**

D **Sequencing**

E **Compatible FASTQ Data**

Ion Workflow – Library Prep

gDNA

1. Fragmentation (Size dependent)
2. End repair (for physical shearing)
3. Adapter ligation & nick repair
4. Size selection
5. Amplification (if needed)
6. Qualify & Quantify

Total Time ~2-3 hours

Amplicon Library

1. Fragmentation (long amplicons)
2. End repair (small amplicons)
3. Adapter ligation & nick repair
4. Size selection
5. Amplification (if needed)
6. Qualify & Quantify

Total Time ~2-3 hours

1. PCR
2. Clean-Up
3. Quantification

Total Time ~3 hours

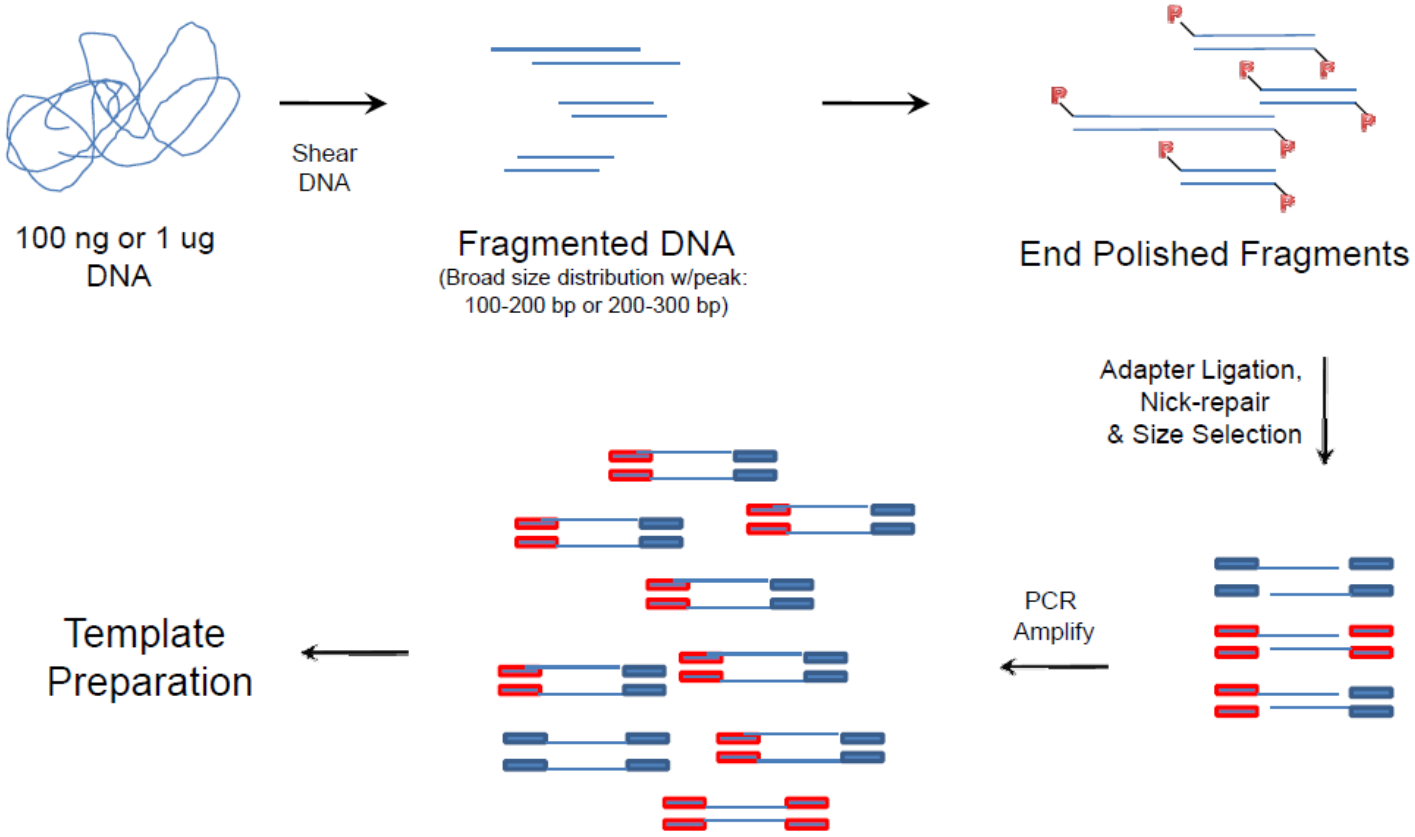
RNA Library

1. Prepare WT or miRNA
2. Hyb./ Adapter Ligation
3. Reverse Transcription
4. Size selection
5. Amplification
6. Qualify & Quantify

Total Time ~1.5 Days

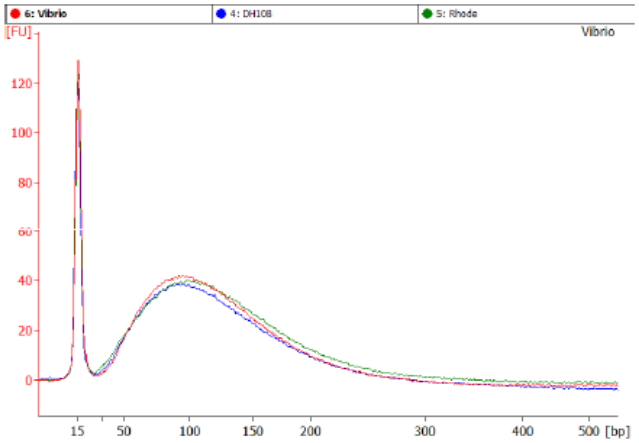
ion torrent
by Life technologies

Ion Fragment Library Workflow

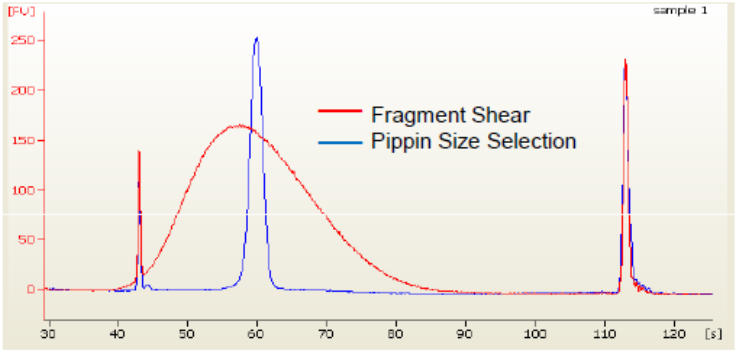


Overlay: Fragment Shear / Size Selection

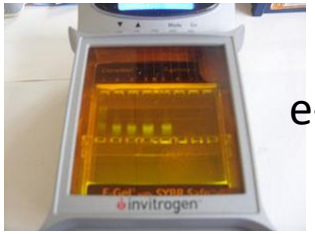
38 % GC ⟶ 65% GC



Consistent shearing across different genomes

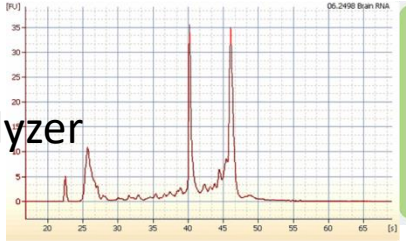


Size selection is necessary for maximization of sequencing yield



e-gel

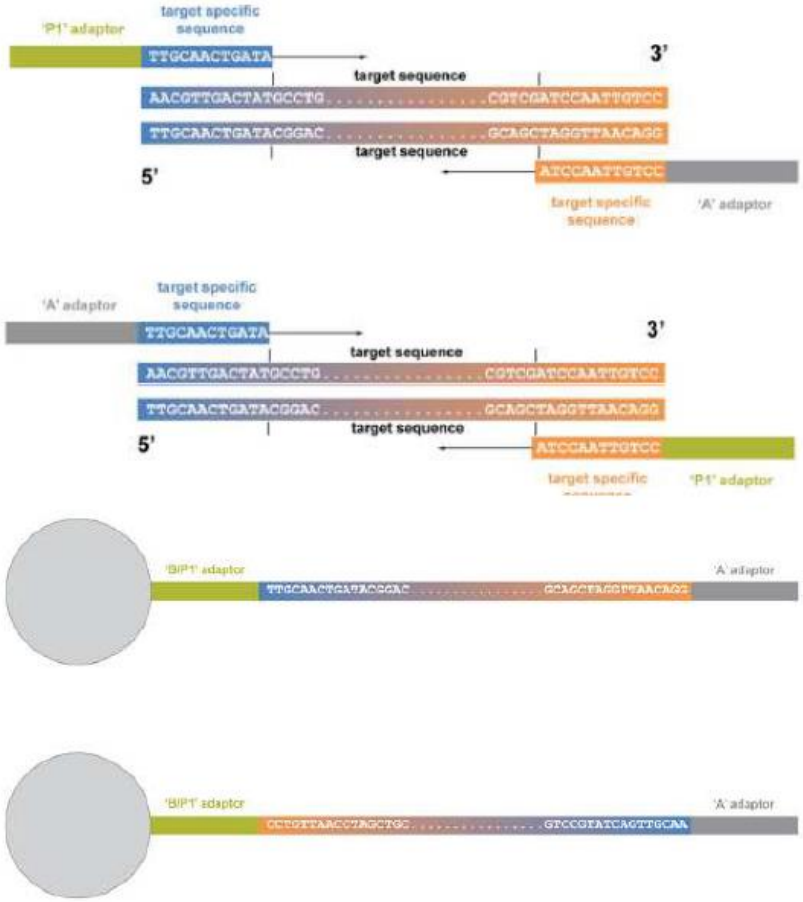
Bioruptor

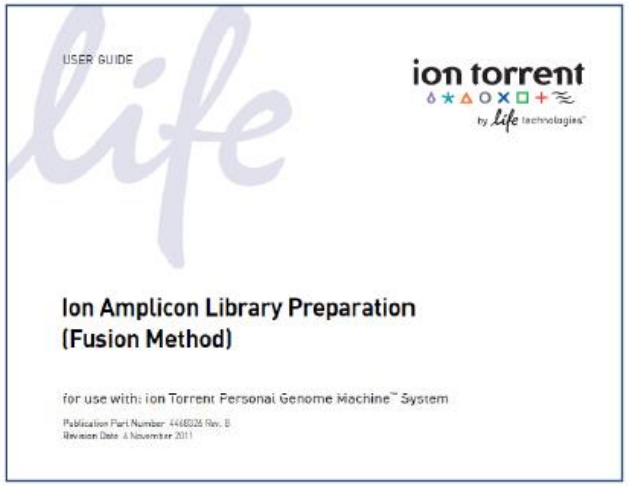


Bioanalyzer

Amplicon Library Construction Overview

Direct ligation of adapters OR fusion primer method





PCR amplify genomic DNA targets

▼

Purify the amplicon libraries

▼

Prepare an equimolar pool of the amplicon libraries

▼

Determine the dilution required for Template Preparation

ion torrent
by life technologies

Ion Workflow

A **DNA**

B **Compatible Library Prep**

C **Template Prep**

D **Sequencing**

E **Compatible FASTQ Data**

Ion Workflow – Template Prep

A DNA

B Compatible Library Prep

C **Template Prep**

D Sequencing

E Compatible FASTQ Data

1 **Assemble amplification reaction**
15 min. (15 min. hands-on)

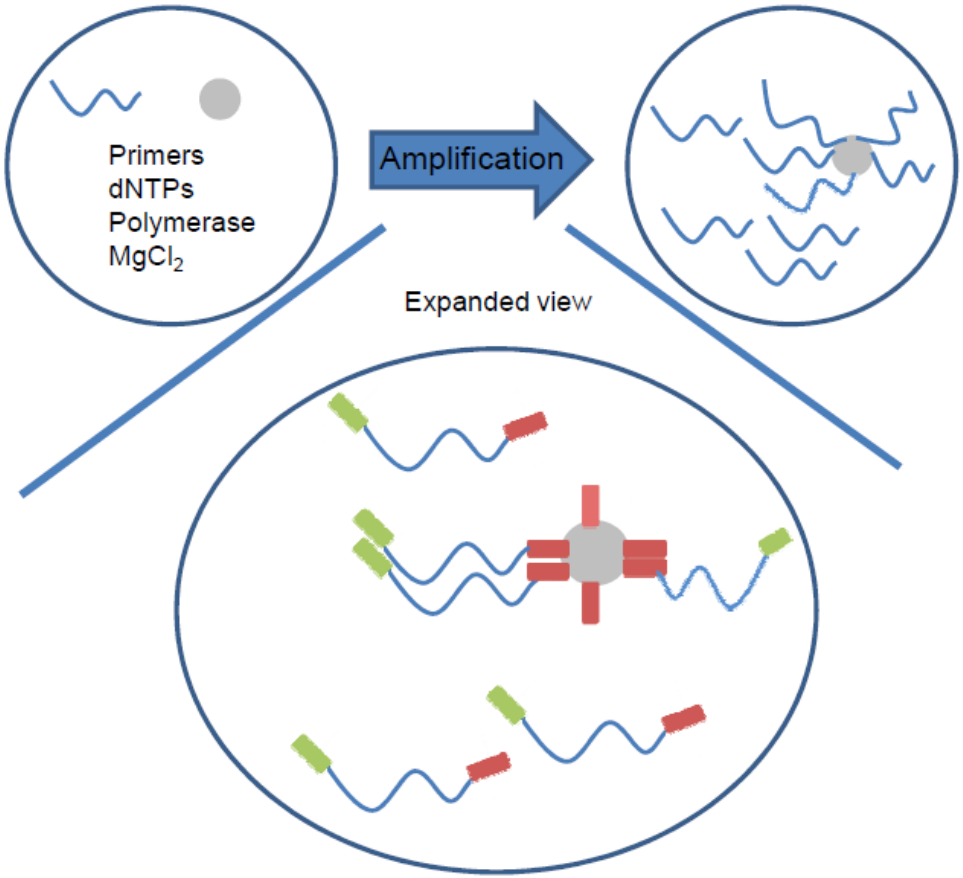
2 **Amplify, recover - Ion OneTouch™ Instrum.**
4 hours (5 min. hands-on)

3 **Enrich – Ion OneTouch™ ES**
35 min. (5 min. hands-on)

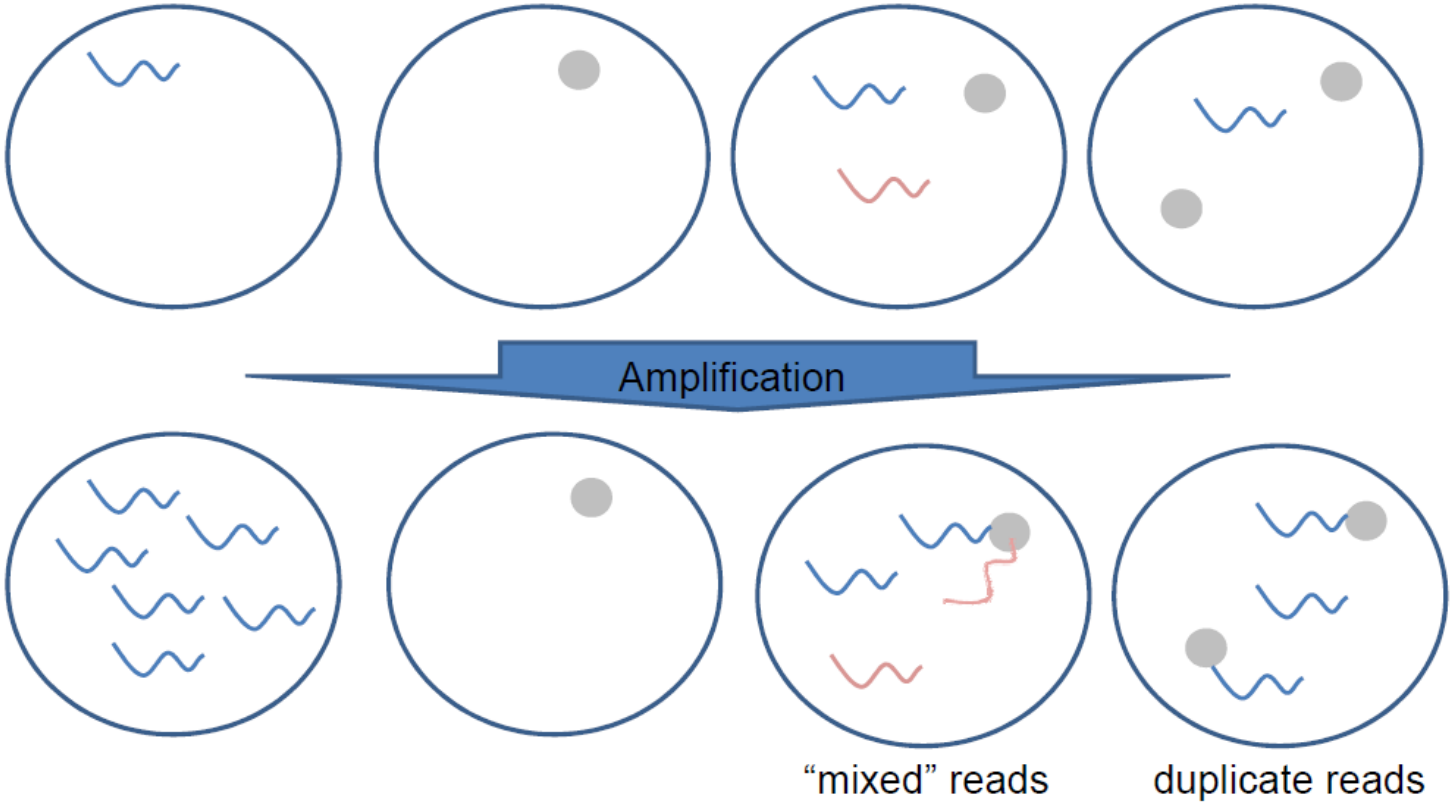
Qubit ® Fluorometer

Total Processing Time ~5 hours
Hands-On Time ~25 min.

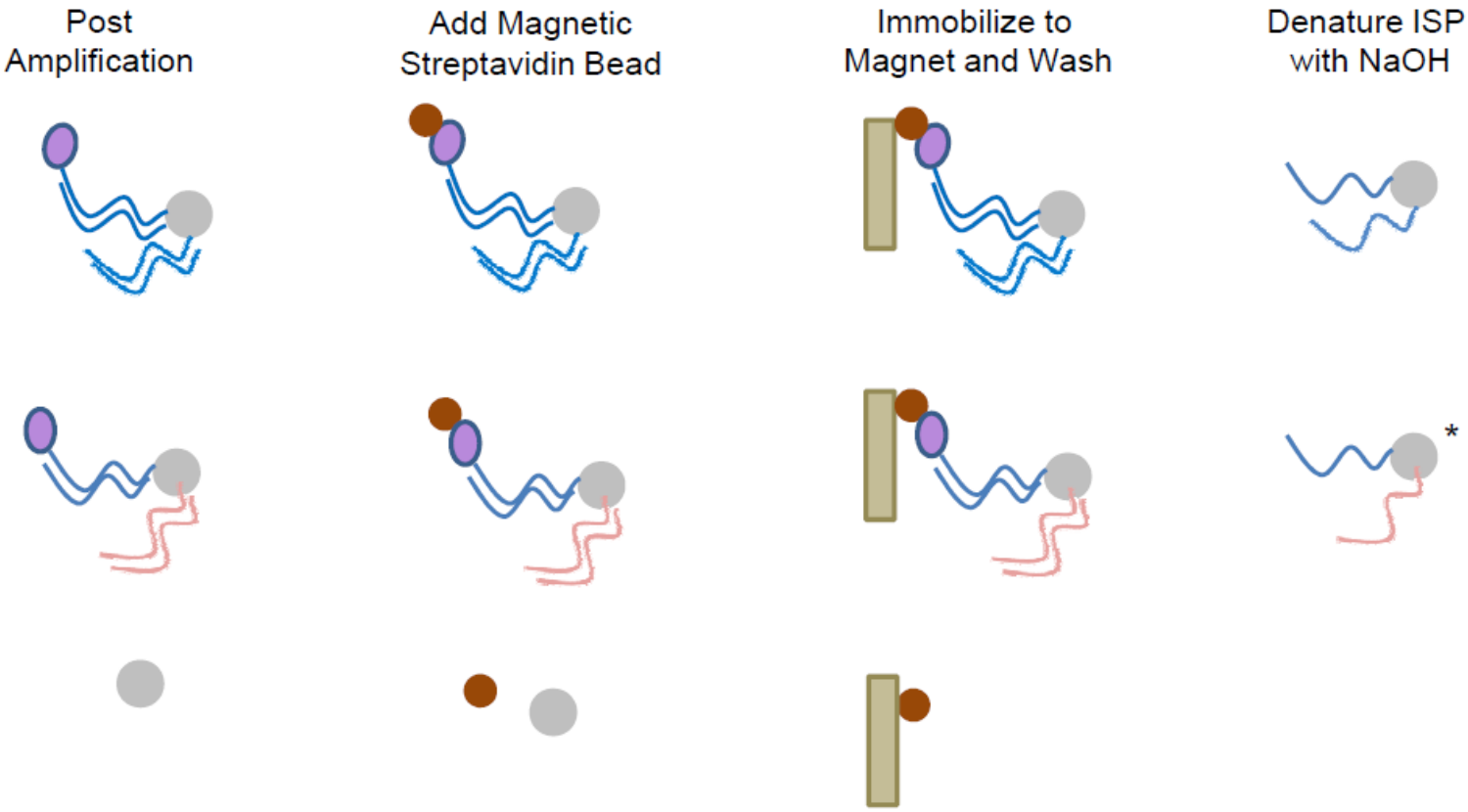
Ideal Clonal Amplification



Non-Ideal Cases

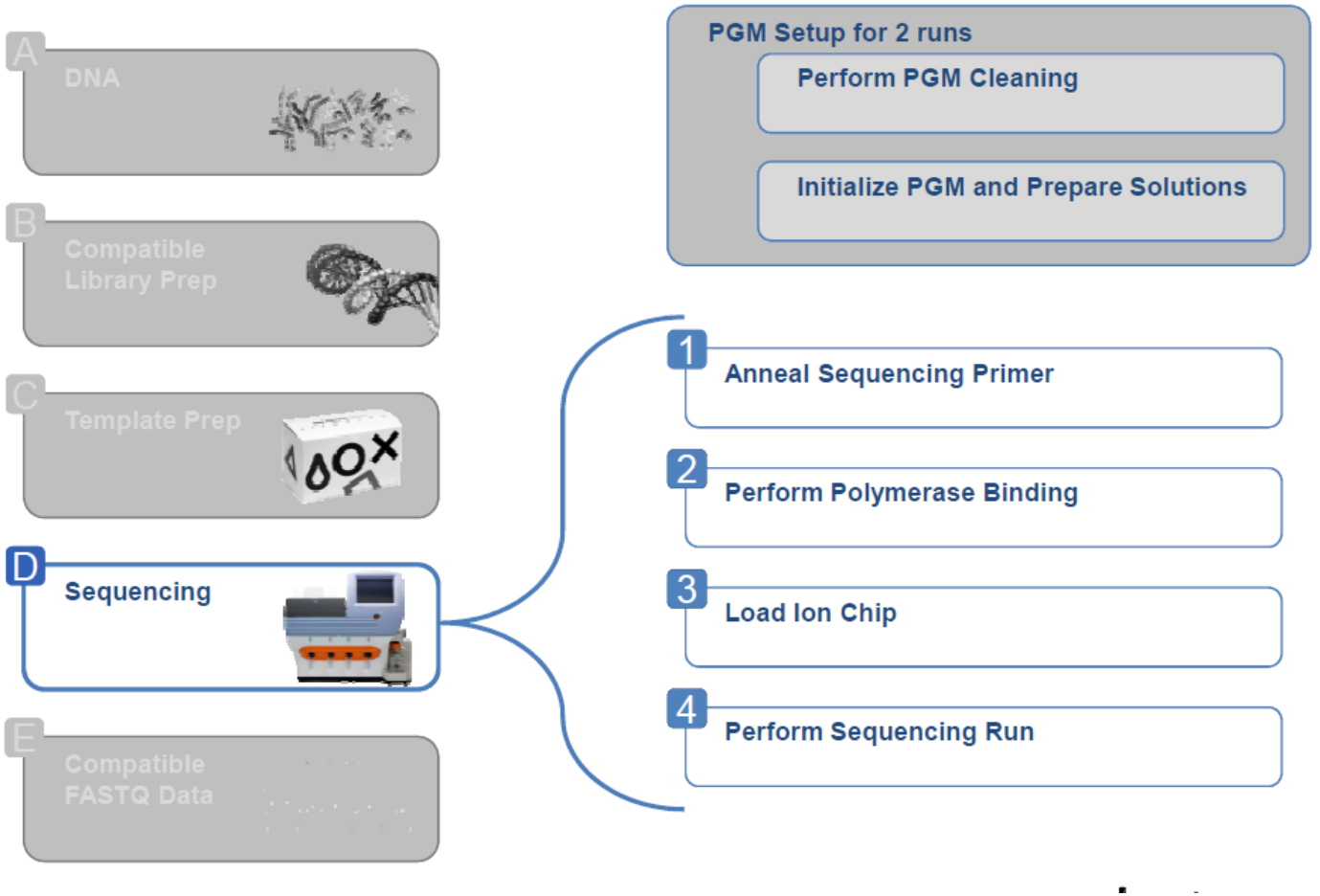


Ion Sphere™ Particle Enrichment



*This species can be minimized through proper dilution. Proper DNA to Ion Sphere™ Particle ratio is critical!

Ion Workflow – PGM™ System Sequencing Run



Sequencing: Flows

- A "flow" is the event of exposing the chip to one particular dNTP (T, A, C, or G), followed by a washing step
- The flow order repeats with pattern:
  - 'TACGTACGTCTGAGCATCGATCGATGTACAGC'

Flows 1-4 Flows 5-8 …9-12 … etc.

-----Primer------ **A G T C** A A G C G T C C C A T G …---Ion Sphere™ Particle

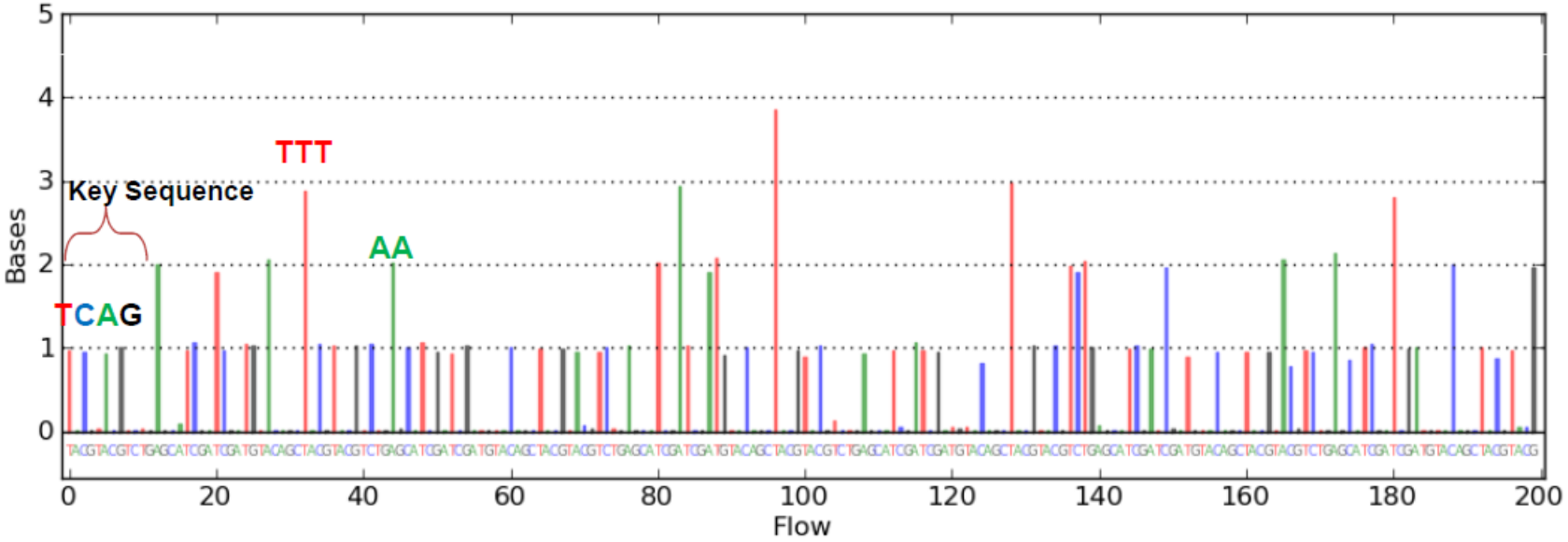
Key Sequence Sequence of Interest

Flows 1-4

A "cycle" is four consecutive dNTP flows: for instance, T-A-C-G = 1 cycle

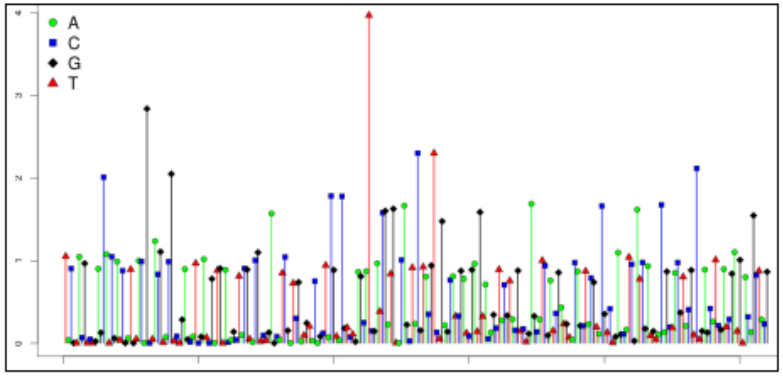
Sequencing: Ionograms

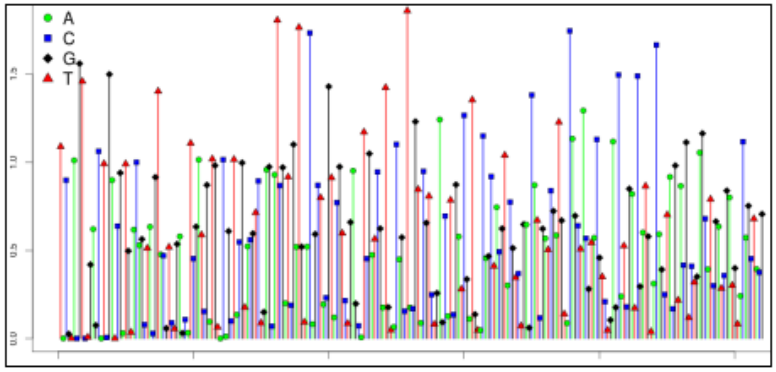
- An “ionogram” is the output of the signals in flow space
- Must be read “up-and-down” along with “left-to-right”
- Height of bar indicates how many nucleotides incorporated during flow
- “Negative” or “zero” flows indicate no nucleotide incorporation
  - These observations are omitted when converting to nucleotide space

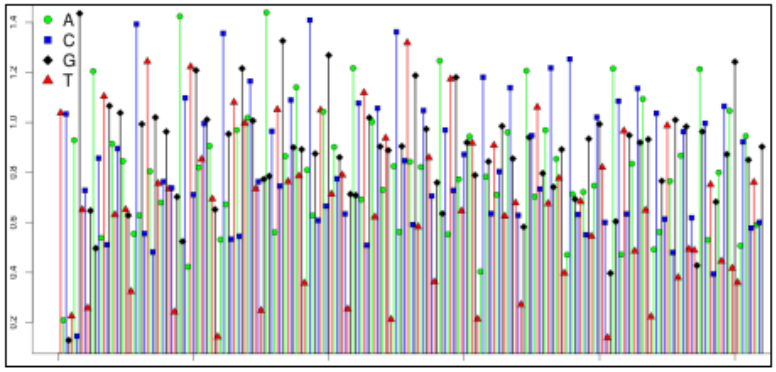


Sequence: AATCTTCTG…

ion torrent


A
C
G
T


A
C
G
T


A
C
G
T

Analysis Workflow

A **DNA**

B **Compatible Library Prep**

C **Template Prep**

D **Sequencing**

E **Compatible FASTQ Data**

```
@WRHH9:4:19
TTGTCATTCAGAGAATGTCTCTTGGACTTTTTCCAAACTGGACCATTTAGGGATTGCTCCACTGATTATGG
+
C@CCC@@<@ECCCC@DDCCCEE@B7A9>@@@->2:<3;@@<DD<DDE9CCC<AA>DD?>:>C@ACC9><:<
@WRHH9:5:6
AAAACGTTTGTCCAAAACCCCCAAAGATCTATTTTTGCTGGAGCTGAAATCCCCAGGGACCTCTGAAGCTG
+
CCC:BBBB=BDE@EEE8EEEE6<<6;@=??22222&2@AC=EDCCCEE;DDDD=B@@4..(.;=:<9?BAA
@WRHH9:5:7
CAATGGCCGCCGCTGCAGCCGGCGCACCGCGCTGATCCGATGGCAGGAGCCAGGATCCAGGTGCTTTC
+
CC@CE@E@CC@CDEEEECC@C?DCCBA=@CBBACCCC?CCCD?DCC?CCC>CD?DCC?CE=@@222*,
@WRHH9:5:13
CCTGCCTAGGAGACCCACGCTTCACCGAGGCAGACTTAGAGCCATGCTTCCCCACCAATGACGCCGACGCT
+
```

1 **Data Transfer from PGM Run**

2 **Convert Raw Signal to Base Calls**

3 **View Run Quality Data, Download Base Calls**

```
@WRHH9:4:19
TTGTCATTCAGAGAATGTCTCTTGGACTTTTTCCAAACTGGACCATTTAGGGATTGCTCCACTGATTATGG
+
C@CCC@@<@ECCCC@DDCCCEE@B7A9>@@@->2:<3;@@<DD<DDE9CCC<AA>DD?>:>C@ACC9><:<
@WRHH9:5:6
AAAACGTTTGTCCAAAACCCCCAAAGATCTATTTTTGCTGGAGCTGAAATCCCCAGGGACCTCTGAAGCTG
+
CCC:BBBB=BDE@EEE8EEEE6<<6;@=??22222&2@AC=EDCCCEE;DDDD=B@@4..(.;=:<9?BAA
@WRHH9:5:7
CAATGGCCGCCGCTGCAGCCGGCGCACCGCGCTGATCCGATGGCAGGAGCCAGGATCCAGGTGCTTTC
+
CC@CE@E@CC@CDEEEECC@C?DCCBA=@CBBACCCC?CCCD?DCC?CCC>CD?DCC?CE=@@222*,
@WRHH9:5:13
CCTGCCTAGGAGACCCACGCTTCACCGAGGCAGACTTAGAGCCATGCTTCCCCACCAATGACGCCGACGCT
+
```

FASTQ format stores sequences and Phred qualities in a single file.

It is concise and compact. FASTQ is first widely used in the Sanger Institute and therefore we usually take the Sanger specification and the standard FASTQ format, or simply FASTQ format.

Although Solexa/Illumina read file looks pretty much like FASTQ, they are different in that the qualities are scaled differently. In the quality string, if you can see a character with its ASCII code higher than 90, probably your file is in the Solexa/Illumina format.

!"#$%&'()*+,-./0123456789:;<=>?@ABCDEFGHIJKLMNOPQRSTUVWXYZ[\]^_`abcdefghijklmnopqrstuvwxyz{|}~

Sanger format can encode a Phred quality score from 0 to 93 using ASCII 33 to 126 (although in raw read data the Phred quality score rarely exceeds 60, higher scores are possible in assemblies or read maps).
$Q = -10\log_{10}P$

A BAM file (.bam) is the binary version of a SAM file.

A SAM file (.sam) is a tab-delimited text file that
contains sequence alignment data.

These formats are described on the SAM Tools web site:
http://samtools.sourceforge.net.

Supported Applications

Microbial sequencing
- Accurate, fast bacteria and virus de-novo & resequencing

Mitochondrial sequencing
- Highly multiplexed mitochondrial sequencing for research, clinical, and forensic applications

Amplicon sequencing
- Multiplexed amplicon sequencing for rapid detection of germline and somatic mutations

Custom or fixed content amplicon panels for targeted resequencing by ultra-high multiplex PCR
- Revolutionary Ion AmpliSeq™ Target Selection technology simplifies targeted resequencing for research and clinical applications

Custom targeted resequencing by target enrichment
- Fast and simple workflows optimized for all major target enrichment providers

Validation of whole genome and whole exome mutation
- Orthogonal technology to validate SOLiD® System/Illumina whole genome/whole exome results

Library Assessment
- Rapid library complexity validation/QC prior to run on high throughput sequencing platforms

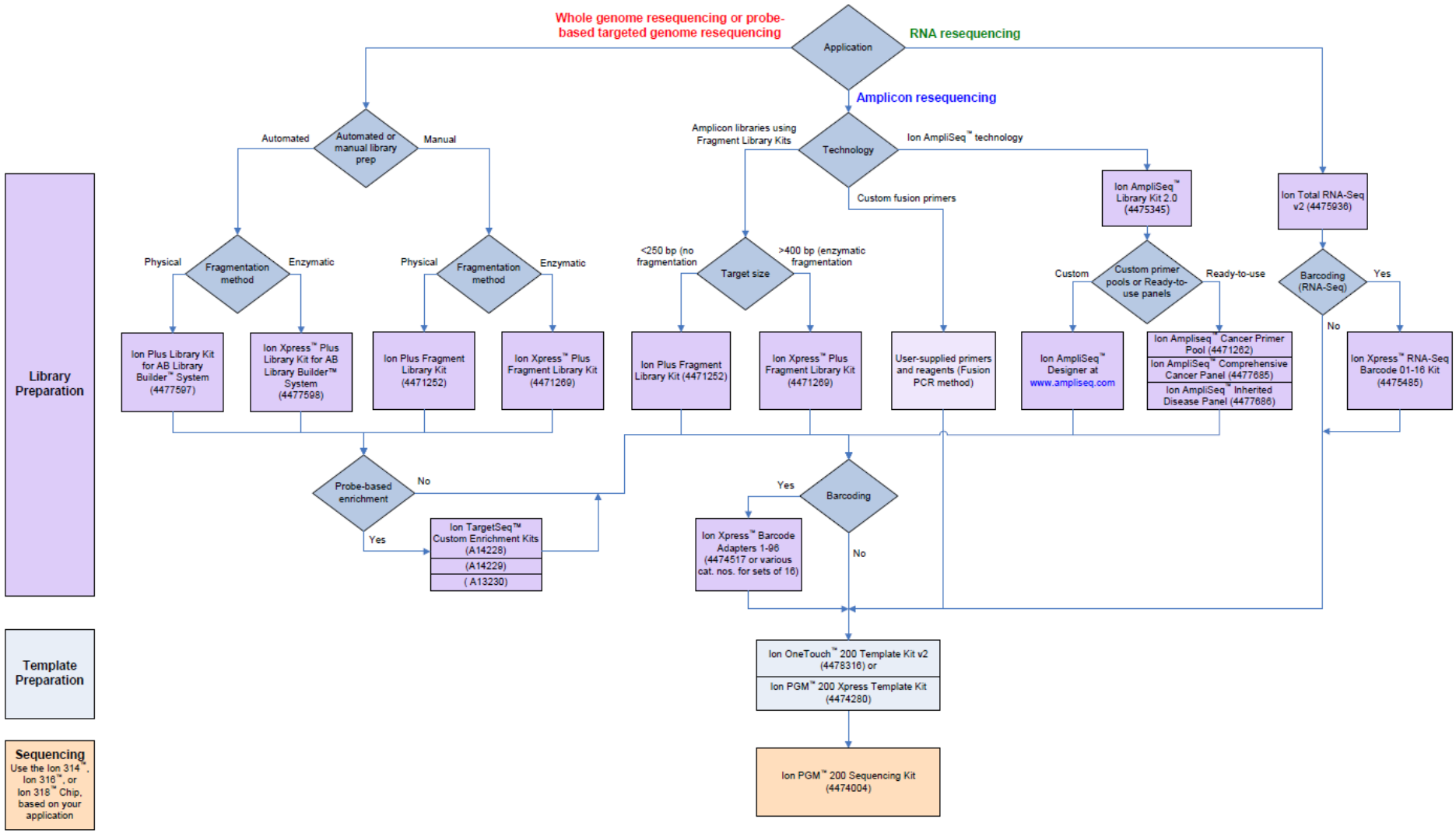
RNA-Seq
- Affordable, fast and simple RNA-Seq solution *(Initially focused on small RNAs & low complexity transcriptomes)*

Whole-transcriptome human RNA-Seq
- New RNA-Seq kits featuring faster workflow and lower RNA input for human whole transcriptome analysis
- Simplified and intuitive data analysis tools to make seamless transition from microarrays

Copy number detection
- Accurate Targeted copy-number detection for basic and clinical research application

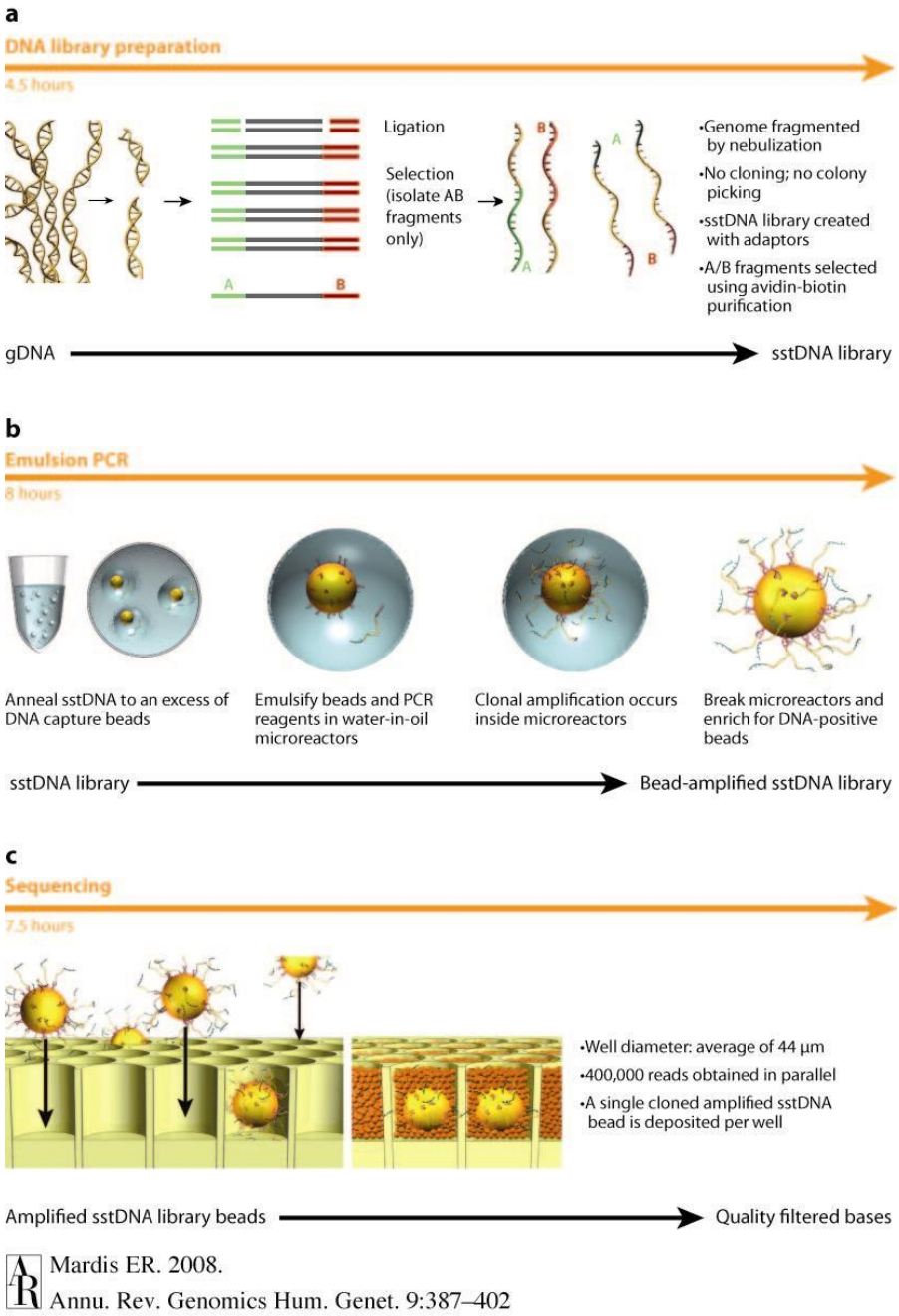
With validated protocols from Ion

Decision Tree for the Ion PGM™ System



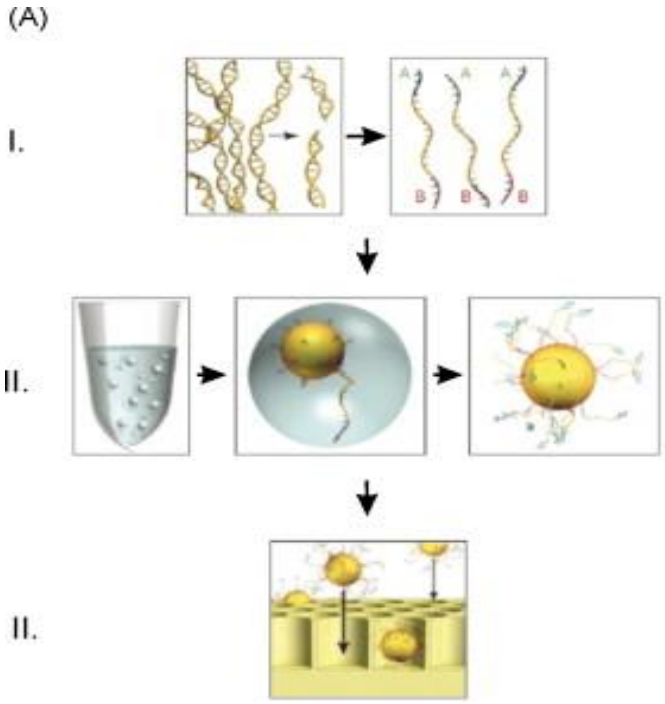
Roche 454

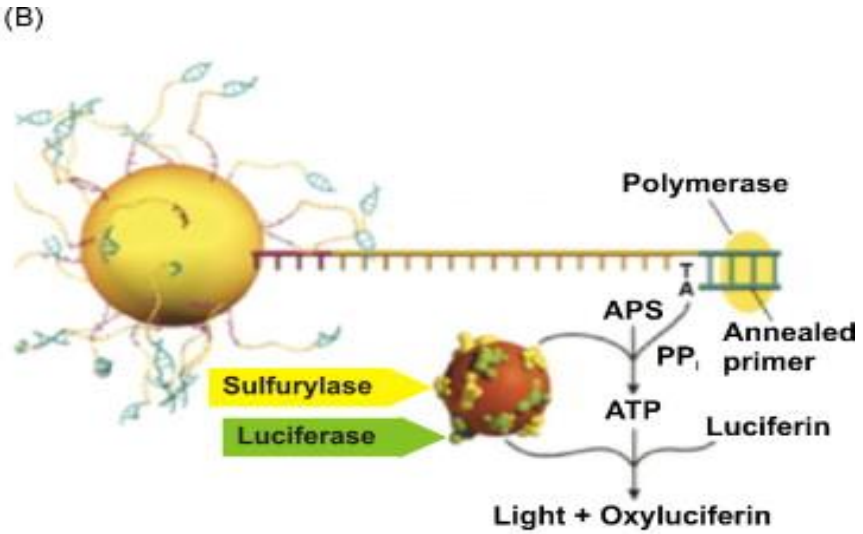
The method used by the Roche/454 sequencer to amplify single-stranded DNA copies from a fragment library on agarose beads. A mixture of DNA fragments with agarose beads containing complementary oligonucleotides to the adapters at the fragment ends are mixed in an approximately 1:1 ratio. The mixture is encapsulated by vigorous vortexing into aqueous micelles that contain PCR reactants surrounded by oil, and pipetted into a 96-well microtiter plate for PCR amplification. The resulting beads are decorated with approximately 1 million copies of the original single-stranded fragment, which provides sufficient signal strength during the pyrosequencing reaction that follows to detect and record nucleotide incorporation events. sstDNA, single-stranded template DNA.



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

Roche 454

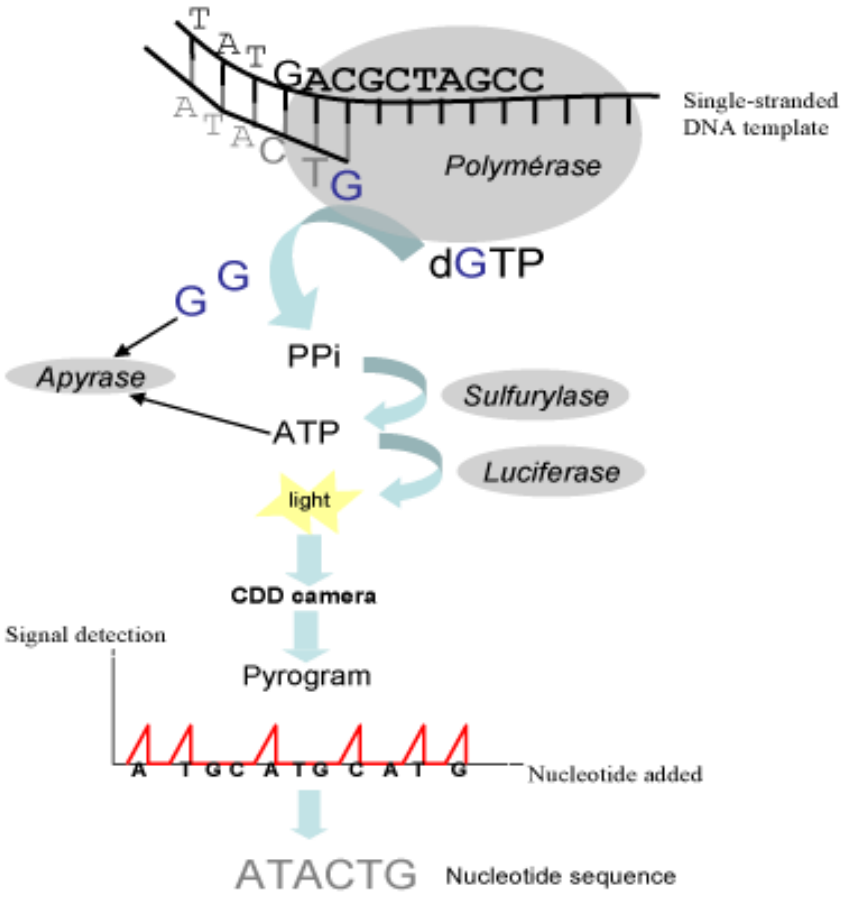




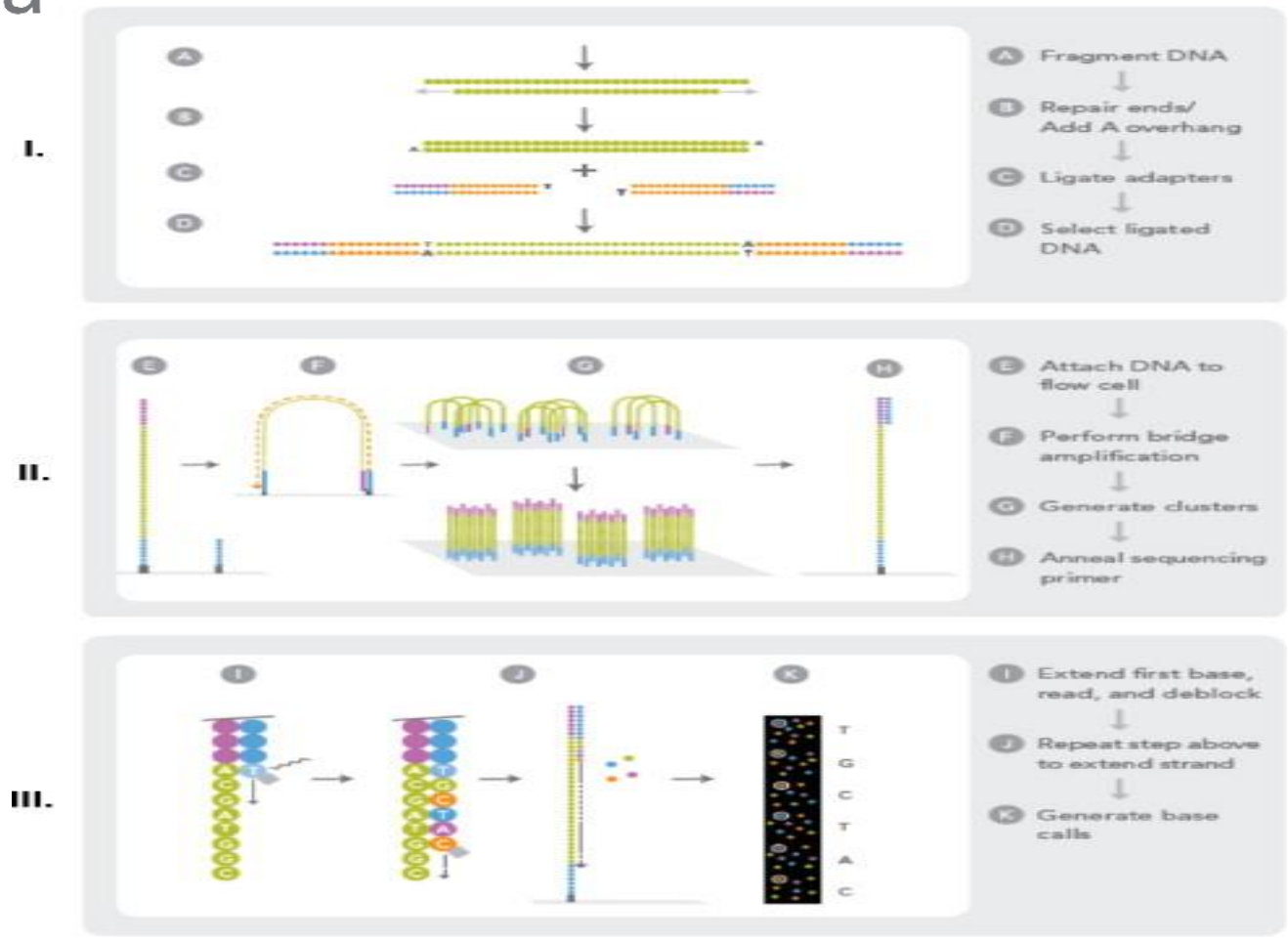
Wilhelm J. Ansorge

Next-generation DNA sequencing techniques

New Biotechnology Volume 25, Issue 4 2009 195 - 203



The primer for the sequencing step is hybridized to a single-stranded DNA template, and incubated with the enzymes, DNA polymerase, ATP sulfurylase, luciferase and apyrase, and the substrates. Deoxyribonucleotide triphosphate (dNTP) is added, one at a time, to the pyrosequencing reaction. The incorporation of a nucleotide is accompanied by release of pyrophosphate (PPi). The ATP sulfurylase quantitatively converts PPi to ATP. The signal light produced by the luciferase-catalyzed reaction in presence of ATP is detected by a charge coupled device (CCD) camera and integrated as a peak in a Pyrogram. The nucleotide degrading Apyrase enzyme continuously degrades ATP excess and unincorporated dNTPs. The process continues with addition of the next dNTP and the nucleotide sequence of the complementary DNA strand is inferred from the signal peaks of the pyrogram.



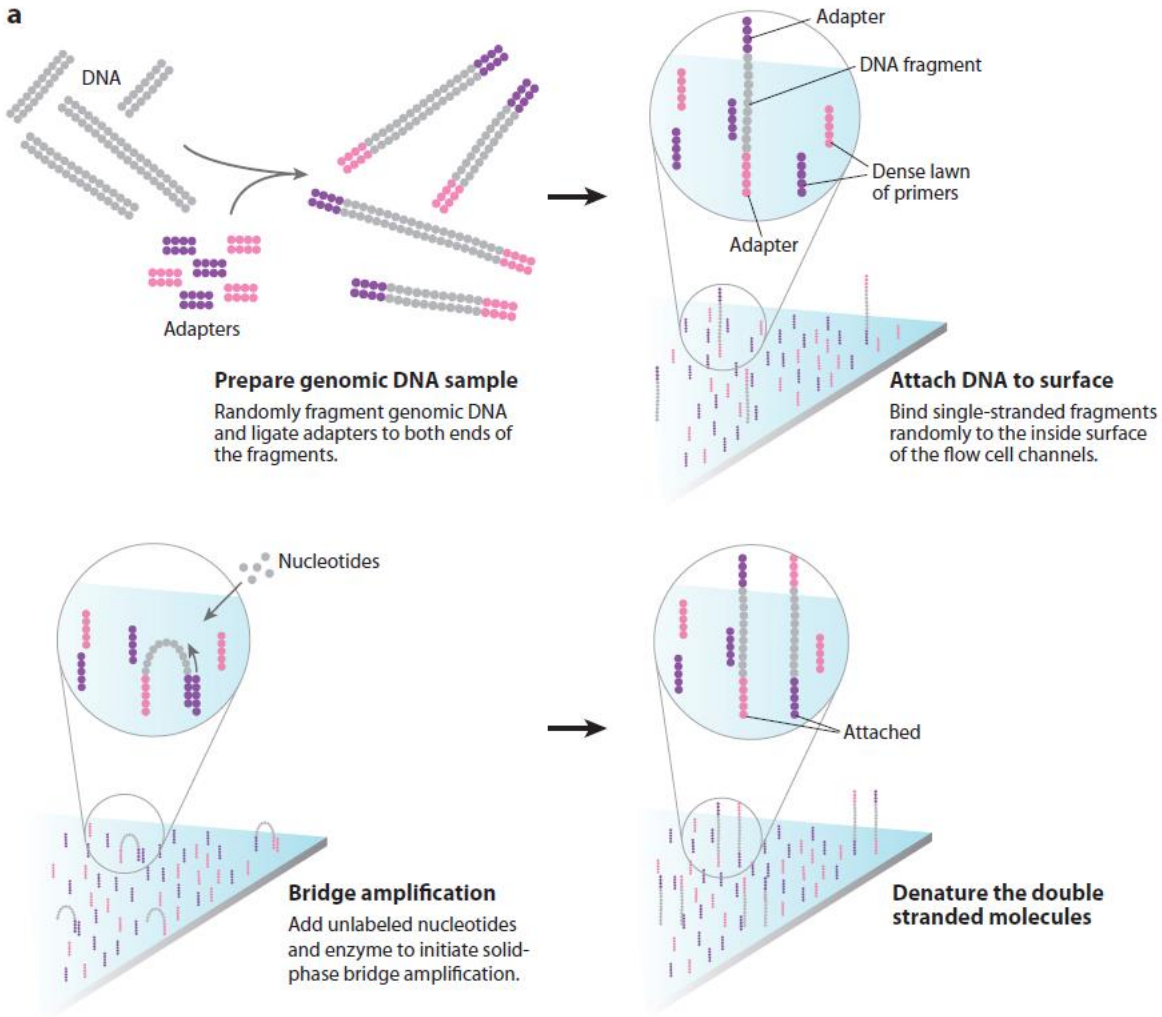
Wilhelm J. Ansorge

Next-generation DNA sequencing techniques

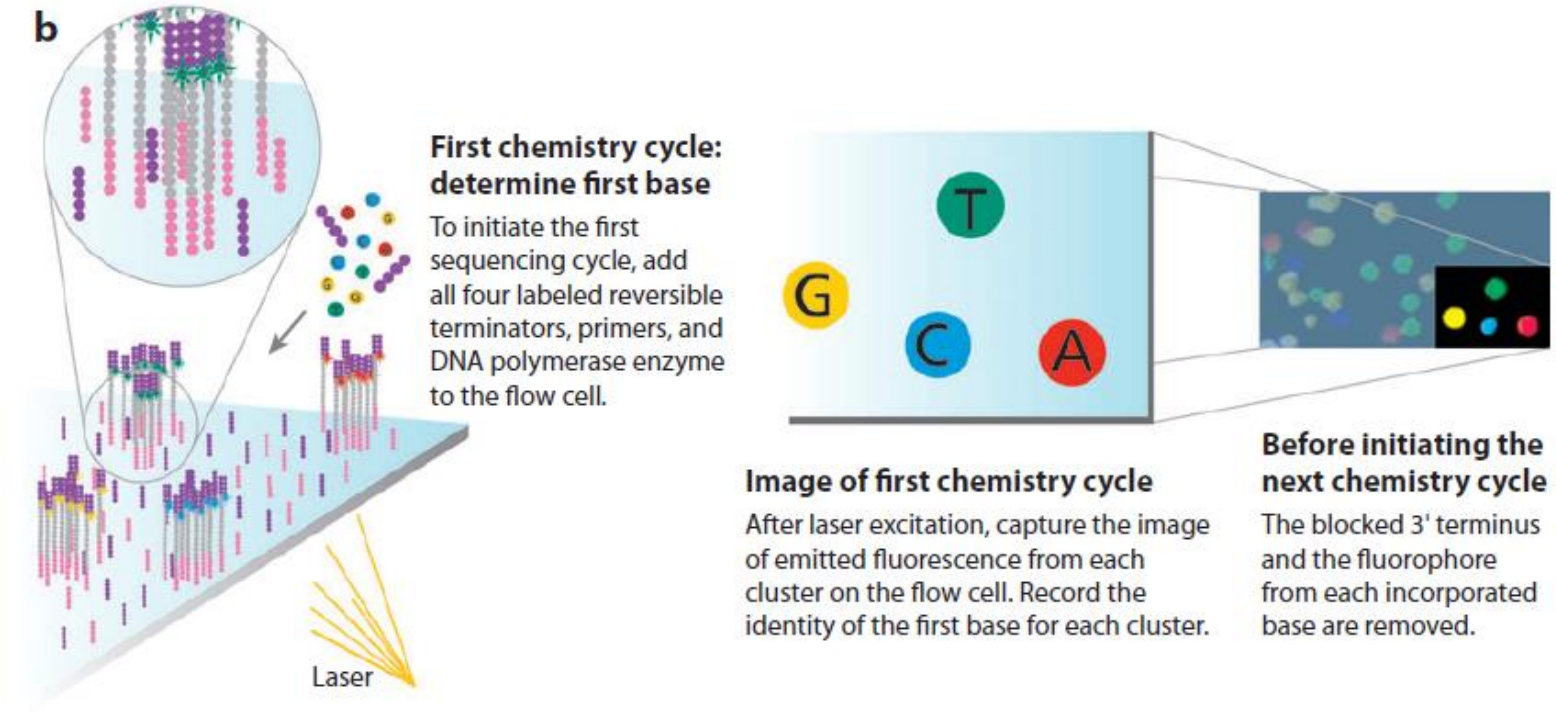
New Biotechnology Volume 25, Issue 4 2009 195 - 203

http://dx.doi.org/10.1016/j.nbt.2008.12.009

illumina®


a
DNA
Adapters
Prepare genomic DNA sample
Randomly fragment genomic DNA and ligate adapters to both ends of the fragments.
Adapter
DNA fragment
Dense lawn of primers
Adapter
Attach DNA to surface
Bind single-stranded fragments randomly to the inside surface of the flow cell channels.
Nucleotides
Bridge amplification
Add unlabeled nucleotides and enzyme to initiate solid-phase bridge amplification.
Attached
Denature the double stranded molecules

illumina®



First chemistry cycle: determine first base

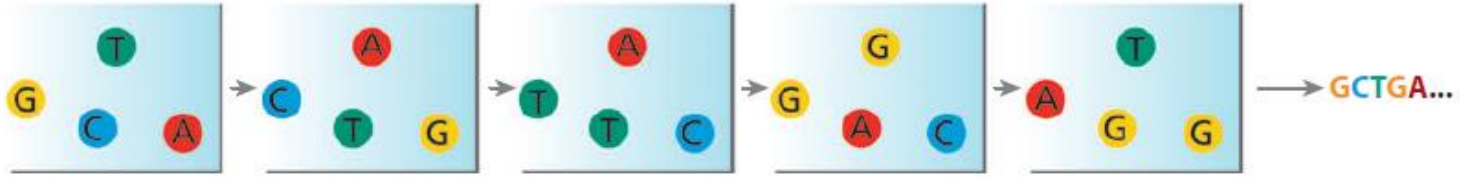
To initiate the first sequencing cycle, add all four labeled reversible terminators, primers, and DNA polymerase enzyme to the flow cell.

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.

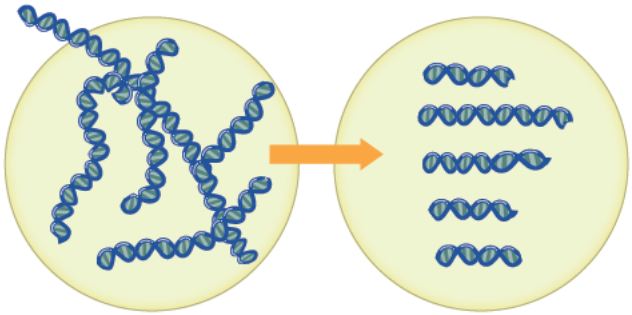


Sequence read over multiple chemistry cycles

Repeat cycles of sequencing to determine the sequence of bases in a given fragment a single base at a time.

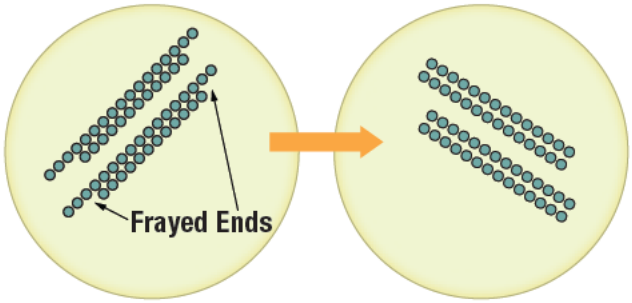
illumina®

1 SONICATION



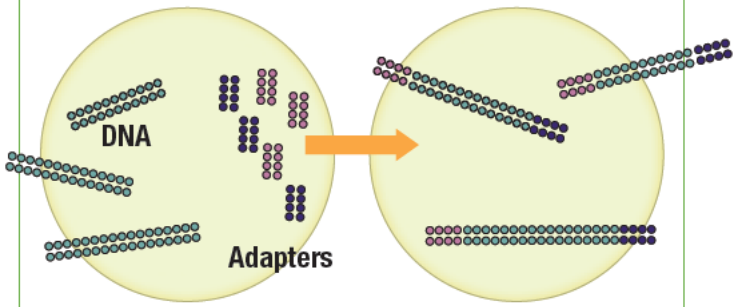
Genomic DNA is fragmented into 100-500 base pair fragments by sonication to create a library.

2 FRAGMENT END REPAIR



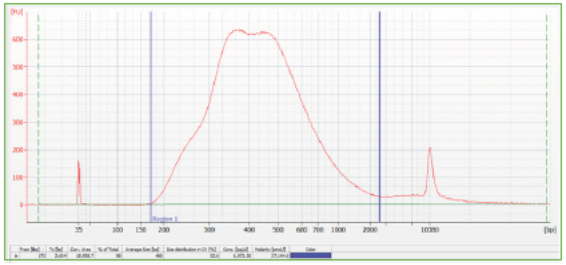
Sonication creates frayed DNA ends which must be blunted or repaired.

3 A-TAILING AND ADAPTER LIGATION



Adapters are ligated to each end of the A-tailed DNA fragment.

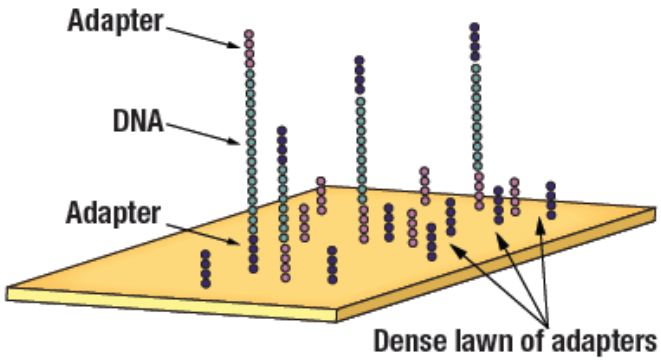
4 QC CHECK



The electropherogram shows the size and concentration of the final library. This library size also confirms the ligation of adapters.

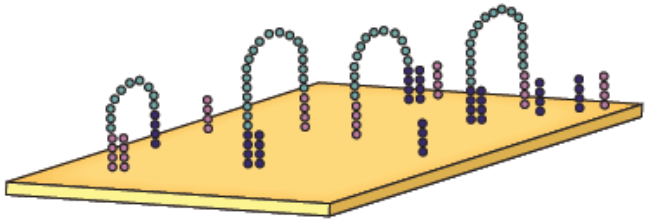
illumina®

5 cBOT CLUSTER GENERATION SYSTEM



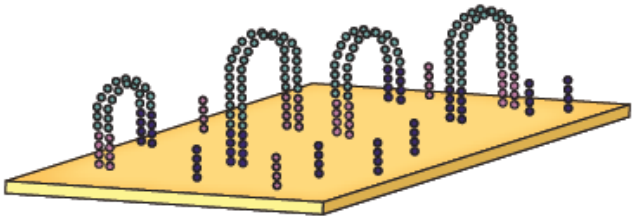
Sodium hydroxide creates single-stranded DNA. Randomly bind these single-stranded DNA to the top and bottom of each channel in the flow cell.

6 BRIDGE FORMATION



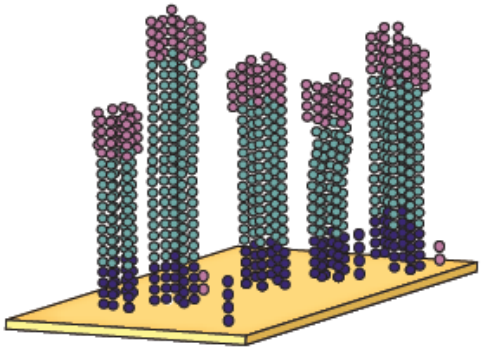
Free DNA end binds to complimentary primer to form a bridge.

7 BRIDGE AMPLIFICATION



Add unlabeled nucleotides and enzyme to initiate solid-phase bridge amplification. Fragments become double-stranded DNA bridges.Thirty-five (35) cycles of amplification create clusters of identical DNA fragments.

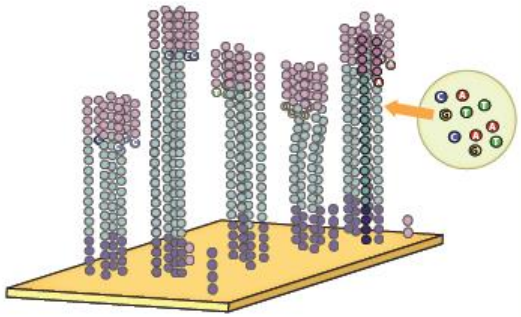
8 FINISHED FLOWCELL



By completion of amplification, several million dense clusters of single-stranded DNA have been generated in each channel of the flow cell with a sequencing primer attached.

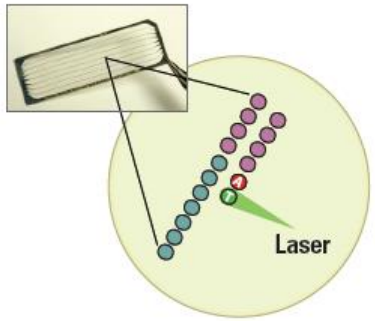
illumina®

9 DNA SEQUENCING



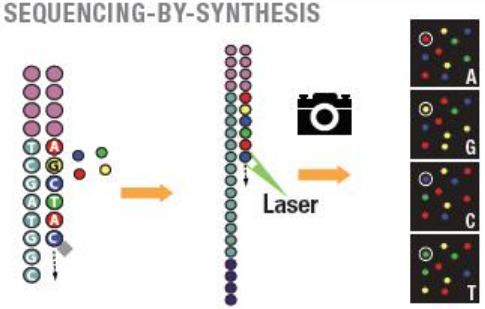
To initiate the first sequencing cycle and determine the first base, all four labeled reversible terminators and DNA polymerase enzyme are first added. Only one base can incorporate at a time.

10 BASE CALLING



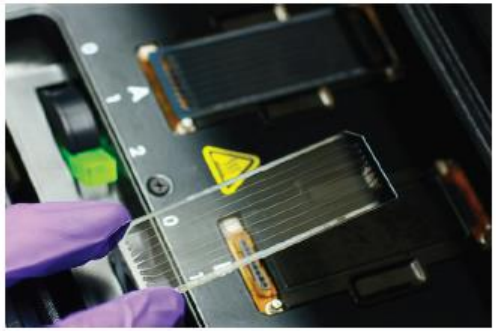
Lasers excite the fluorescent tags and the images are captured via CCD camera. The identity of the first base in each cluster is recorded, and then the fluorescent tag is removed.

11 SEQUENCING-BY-SYNTHESIS

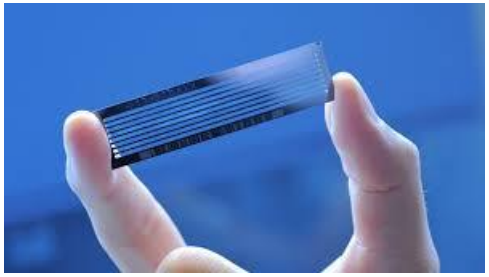


In the first cycle, the first base is incorporated. Its identity is determined by the signal given off and then recorded. In subsequent cycles, the process of adding sequencing reagents, removing unincorporated bases and capturing the signal of the next base to identify is repeated.

12 DUAL FLOW CELLS



Once the top surface of the flow cell channel has been scanned, the imaging step is repeated on the bottom surface.



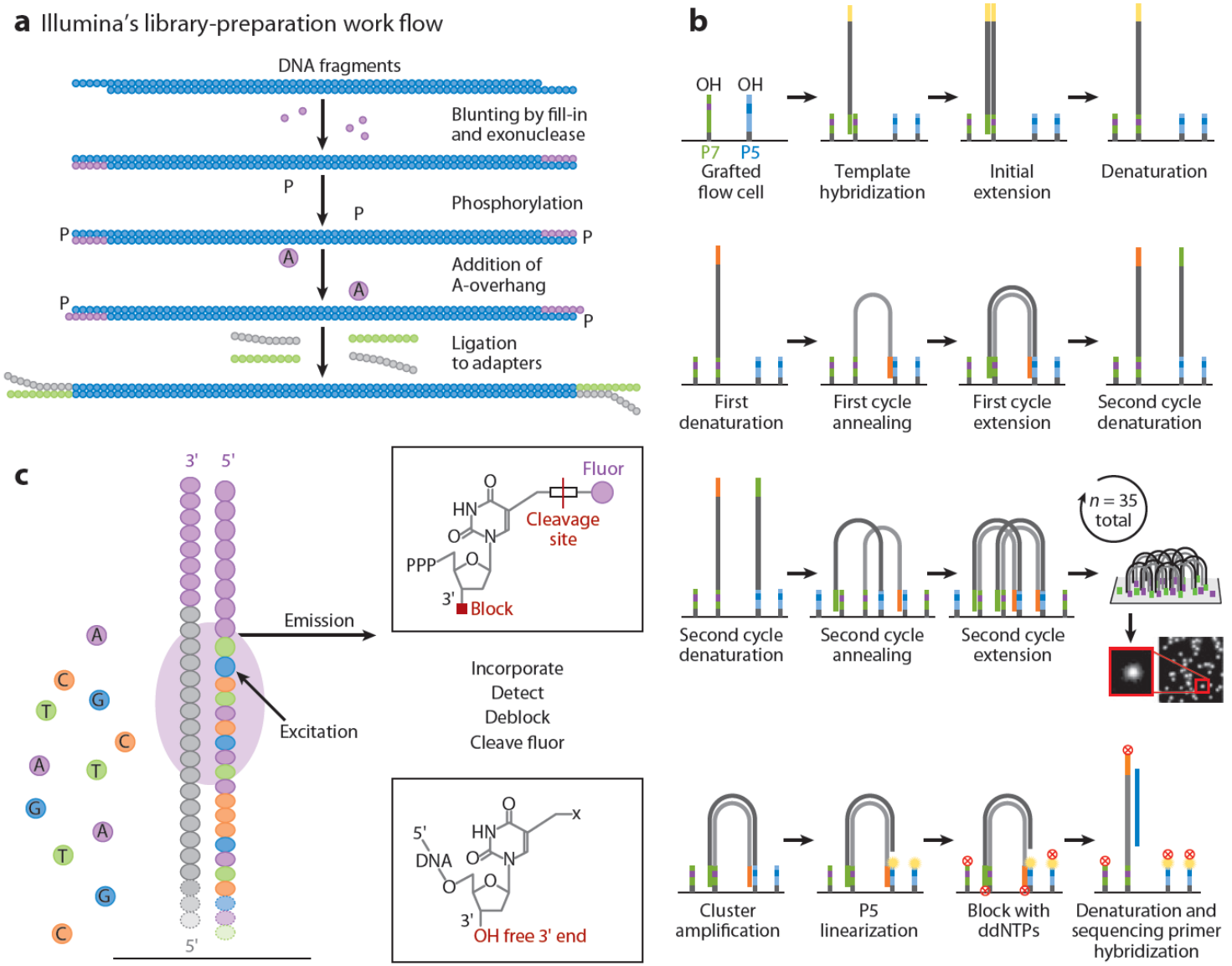
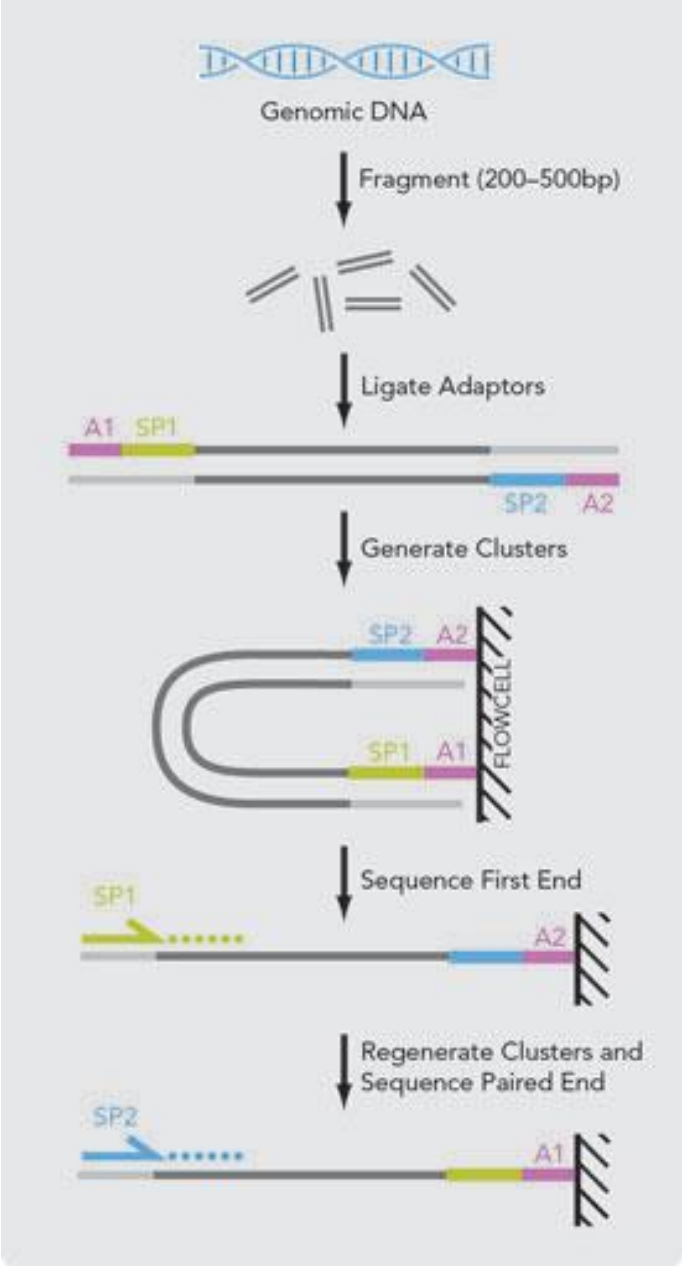


Figure 3

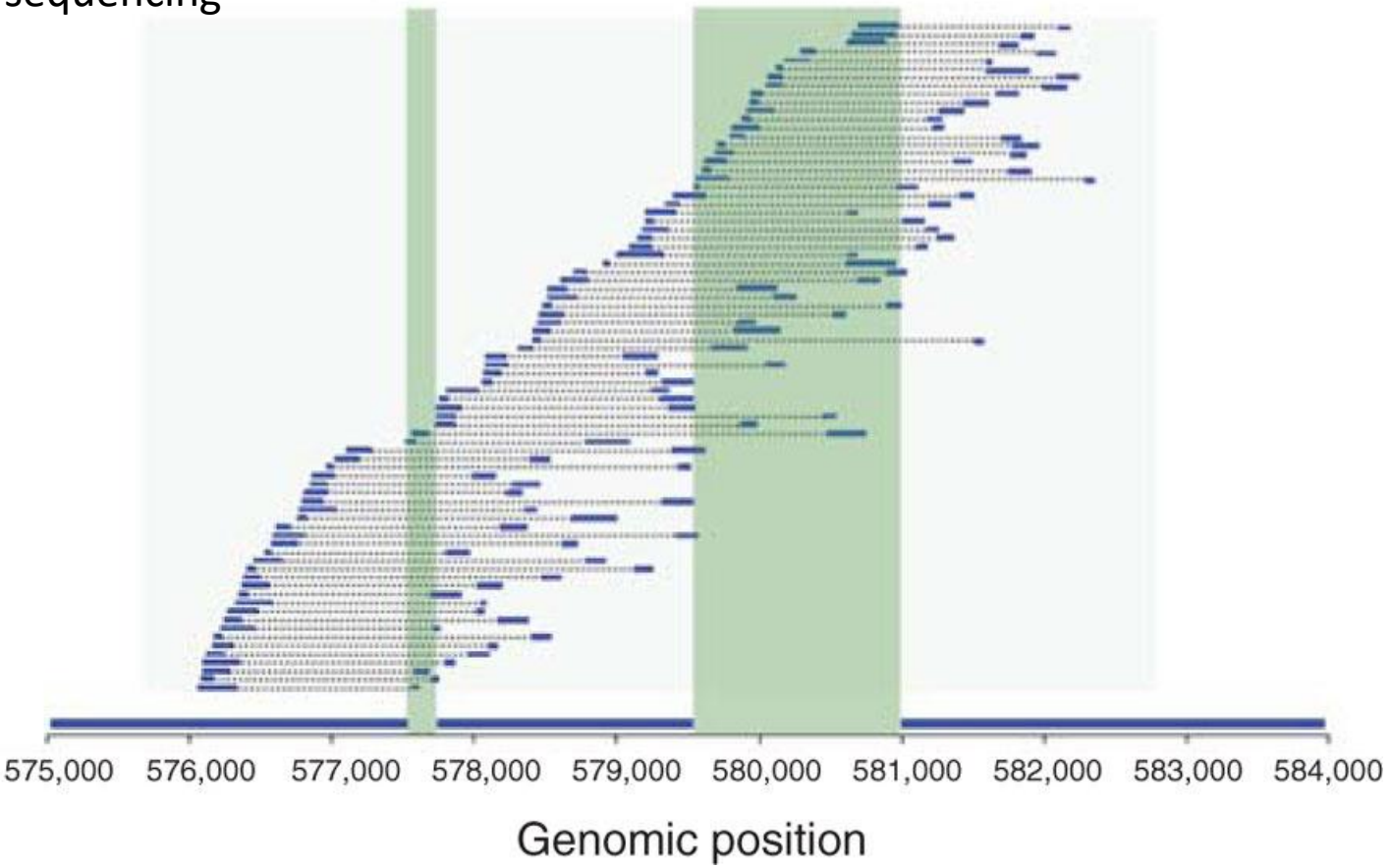
(*a*) Illumina® library-construction process. (*b*) Illumina cluster generation by bridge amplification. (*c*) Sequencing by synthesis with reversible dye terminators.

illumina®

Paired end sequencing

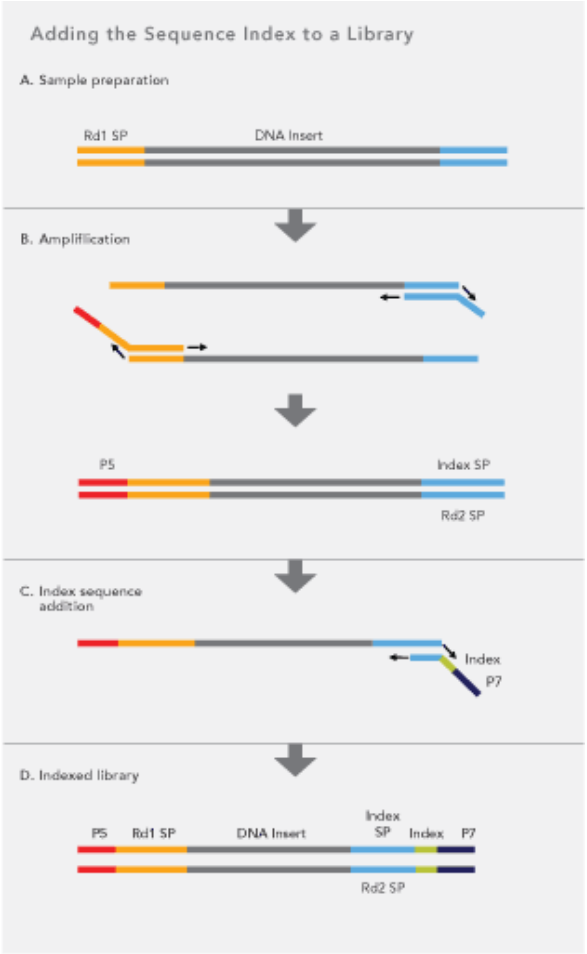


Paired end sequencing



illumina®

Adding the Sequence Index to a Library



(A) During sample preparation, adapters are ligated to the DNA fragments. One adapter contains the sequencing primer site for application read 1 (Rd1 SP). (B) Prepared samples are amplified via PCR using two universal primers. One primer contains an attachment site (P5) for the flow cell, while the other contains the sequencing primer sites for the index read (Index SP) and for application read 2 (Rd2 SP). (C) A third primer in the PCR adds the Index as well as a second flow cell attachment site (P7) to the PCR product shown in step 2. (D) The indexed library is ready for sequencing using the Genome Analyzer system.

	 HiSeq 2500/2000	 HiSeq 1500/1000	 HiSeq 2500	 HiSeq 1500	 HiScan SQ	 MiSeq
Mode	High Output	High Output	Rapid Run	Rapid Run		
Output (maximum)	600Gb	300Gb	180Gb	90Gb	150Gb	8.5Gb
Run Time	2 - 11 Days	2 - 8.5 Days	7 - 40 Hr.	7 - 40 Hr.	1.5 - 8.5 Days	4 - 39 Hr.
Paired-end Reads (maximum)	6 Billion	3 Billion	1.2 Billion	600 Million	1.5 Billion	15 Million
Single-end Read (maximum)	3 Billion	1.5 Billion	600 Million	300 Million	750 Million	7.5 Million
Max Read Length	2x100 bp	2x100 bp	2x150 bp	2x150 bp	2x100 bp	2x250 bp
Bases above Q30 (2x100bp)	>80% (2x100 bp)	>80% (2x100 bp)	>80% (2x100 bp)	>80% (2x100 bp)	>80% (2x100 bp)	>85% (2x100 bp)
Required Input	50 ng with Nextera 100 ng – 1 µg with TruSeq	50 ng with Nextera 100 ng – 1 µg with TruSeq	50 ng with Nextera 100 ng – 1 µg with TruSeq	50 ng with Nextera 100 ng – 1 µg with TruSeq	50 ng with Nextera 100 ng – 1 µg with TruSeq	1ng with NexteraXT 50 ng with Nextera 100 ng – 1 µg with TruSeq

What type of sequencing should I choose for the Illumina sequencing project?

HiSeq 2000/2500 – 100-160mln single end sequencing reads per lane.

- **ChIPseq** – Single End 50 cycles (2-3 human samples per lane)
- **RNAseq** – Single End 50 cycles (2-3 human samples per lane)

If you are interested in splicing variants and fusion genes both Single End 100cycles and Paired End 2x50cycles will be better option for you.

-**Whole Genome Sequencing** – Paired End 2x100cycles (2-3 lanes per genome)
-**Exome Capture** - Paired End 2x100cycles (4 samples per lane)

MiSeq – 3-7 mln single end sequencing reads per lane. Custom projects , fast turnaround.
Metagenomics - 16S profile – Paired End 2x150cycles up to 24 samples per lane.
-Whole Microbial Genome Sequencing - Paired End 2x150cycles

SHORT READ PLATFORMS at UNC

HiSeq 2000

Initially capable of up to 600Gb per run in 13 days.

Cost of resequencing one human genome:
Now UNC PI - (30x coverage) about $6,000
Now for outside of UNC - (30x coverage) about $9,000

HiSeq 2500

Initially capable of up 100Gb per run in 27hours.
Cost per genome - ???

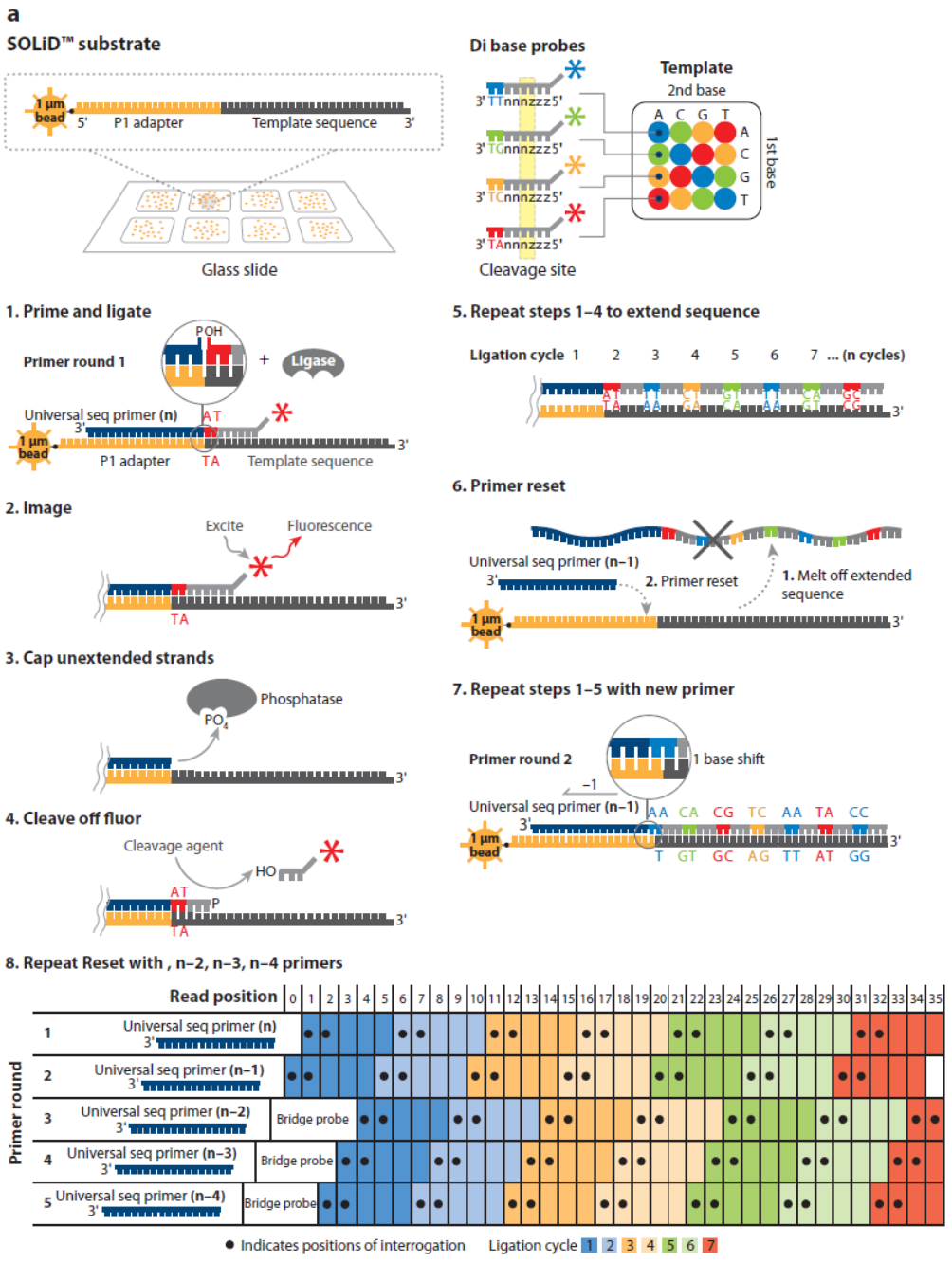
MiSeq

- Small capacity system. PE 2x150cycles in 27hours.

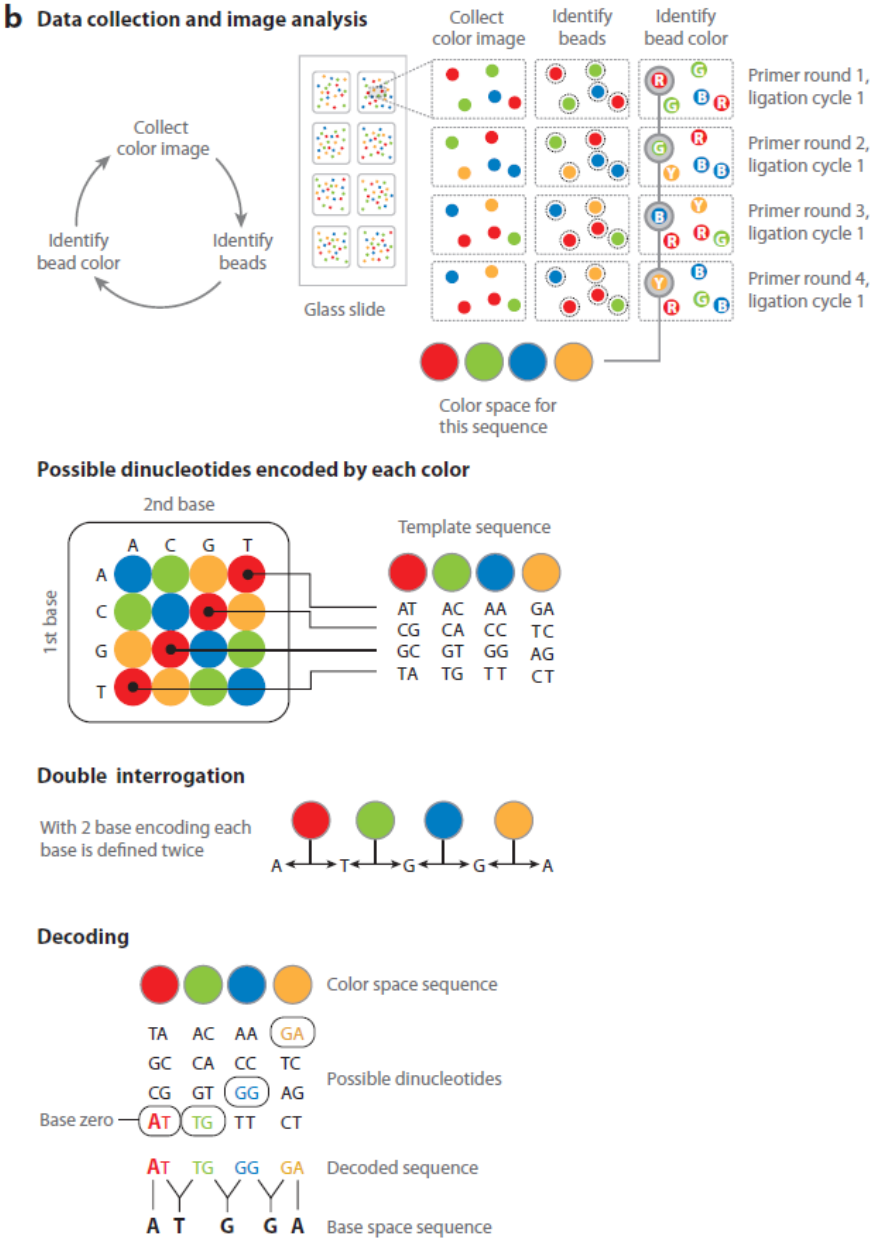
- PE 2 x 250bp coming soon – error rate for read 1 – less than 1%; read 2 about 1.2%.
- In preparation – PE 2 x 400bp – error rate for read1 about 2%; read 2 about 4%.
- In preparation – Longer insert size possible 1.5kb

AB SOLiD

Sequencing by Oligo Ligation Detection



AB SOLiD


b Data collection and image analysis
Collect color image
Identify beads
Identify bead color
Collect color image
Identify bead color
Identify beads
Glass slide
Primer round 1, ligation cycle 1
Primer round 2, ligation cycle 1
Primer round 3, ligation cycle 1
Primer round 4, ligation cycle 1
Color space for this sequence
Possible dinucleotides encoded by each color
2nd base
1st base
A C G T
Template sequence
AT AC AA GA
CG CA CC TC
GC GT GG AG
TA TG TT CT
Double interrogation
With 2 base encoding each base is defined twice
A T G G A
Decoding
Color space sequence
TA AC AA GA
GC CA CC TC
CG GT GG AG
Base zero
AT TG TT CT
Possible dinucleotides
AT TG GG GA
Decoded sequence
A T G G A
Base space sequence

AB SOLiD

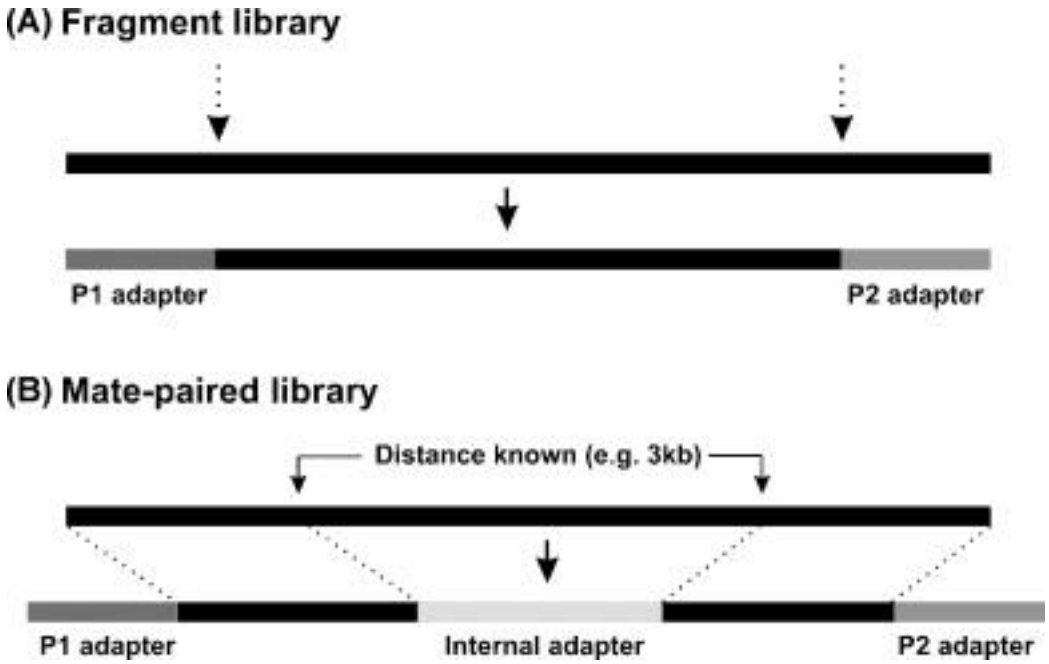


Figure 3 Library preparation for DNA sequencing using the SOLiD DNA sequencing platform. (A) <ce:italic> Fragment library</ce:italic> : After whole genome DNA is randomly fragmented (indicated by the dashed arrows), two different 25<ce:hsp sp="...

Wilhelm J. Ansorge

Next-generation DNA sequencing techniques

New Biotechnology Volume 25, Issue 4 2009 195 - 203

http://dx.doi.org/10.1016/j.nbt.2008.12.009

AB SOLiD

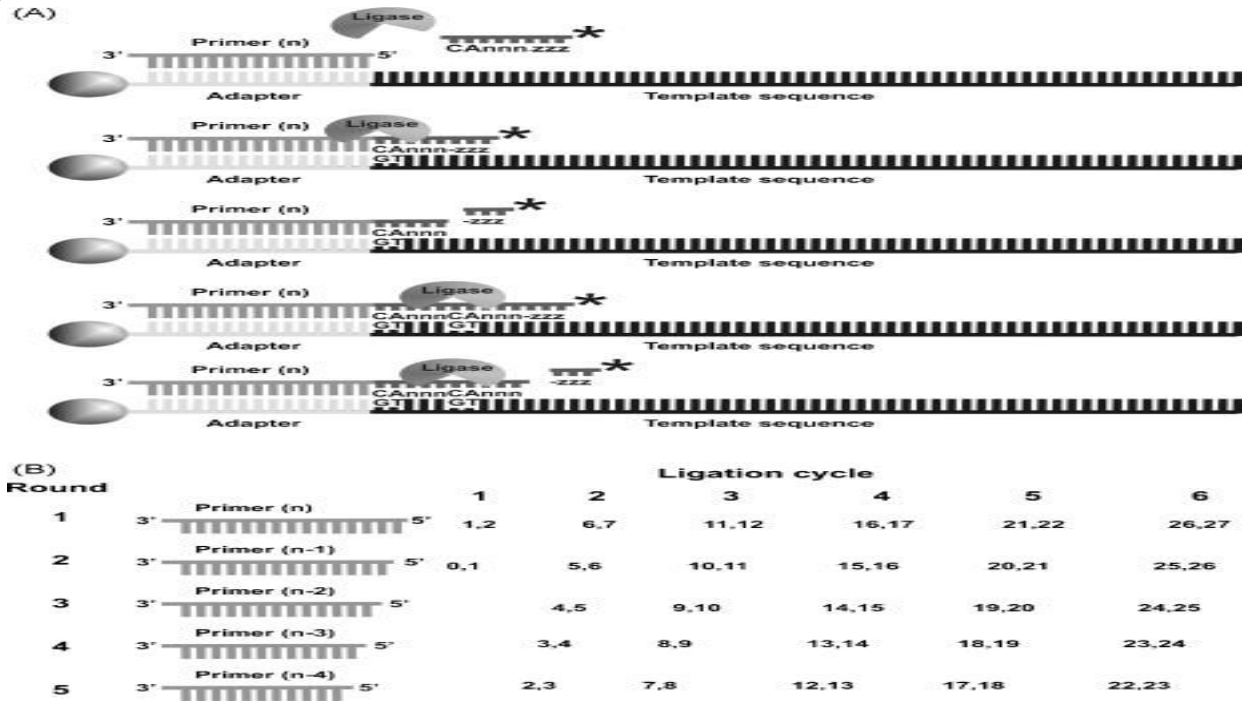


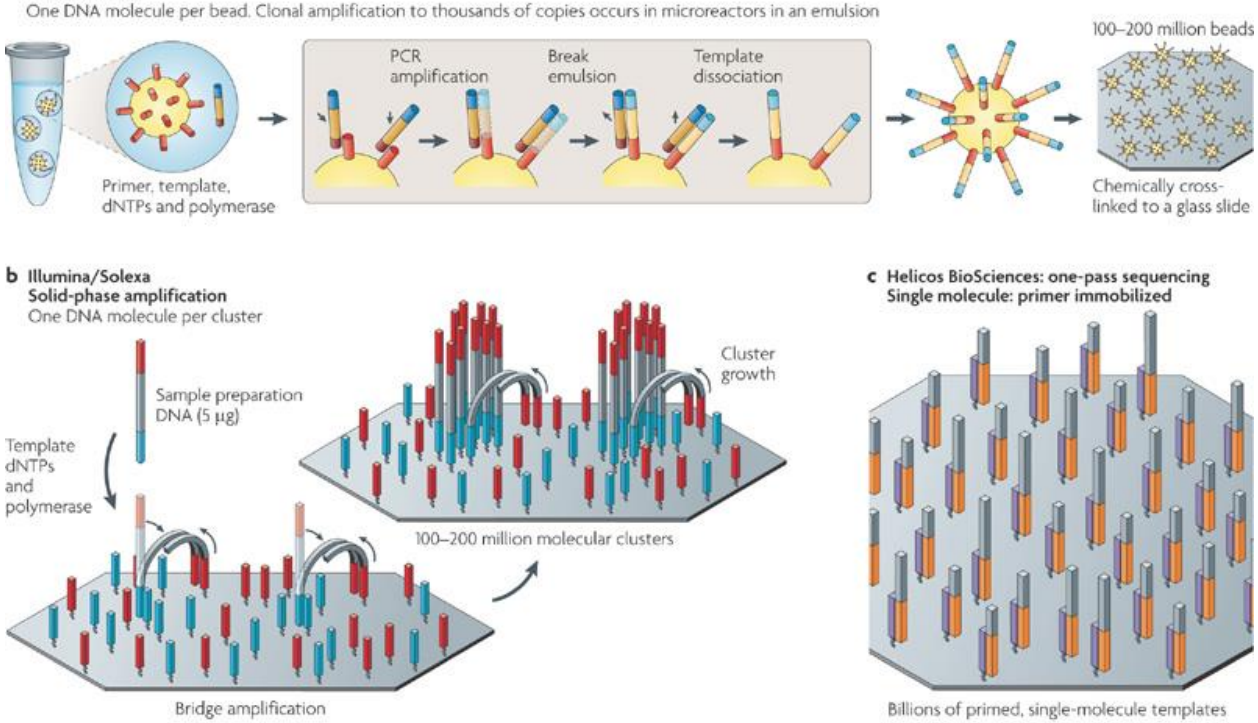
Figure 4 Sequencing-by-ligation, using the SOLiD DNA sequencing platform. (A) Primers hybridise to the P1 adapter within the library template. A set of four fluorescence-labelled di-base probes competes for ligation to the sequencing primer. Th...

Wilhelm J. Ansorge

Next-generation DNA sequencing techniques

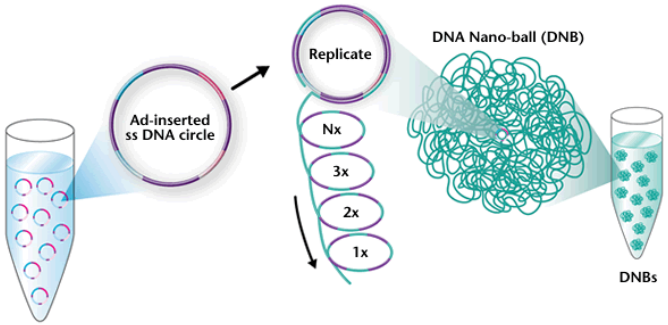
New Biotechnology Volume 25, Issue 4 2009 195 - 203

http://dx.doi.org/10.1016/j.nbt.2008.12.009

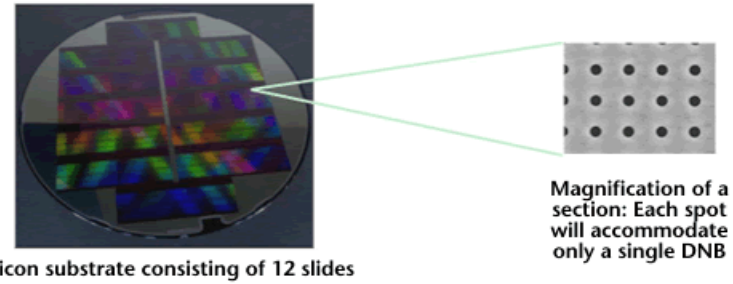

a Roche/454, Life/APG, Polonator
Emulsion PCR
One DNA molecule per bead. Clonal amplification to thousands of copies occurs in microreactors in an emulsion
Primer, template, dNTPs and polymerase
PCR amplification
Break emulsion
Template dissociation
100–200 million beads
Chemically cross-linked to a glass slide
b Illumina/Solexa
Solid-phase amplification
One DNA molecule per cluster
Sample preparation DNA (5 µg)
Template dNTPs and polymerase
Cluster growth
100–200 million molecular clusters
Bridge amplification
c Helicos BioSciences: one-pass sequencing
Single molecule: primer immobilized
Billions of primed, single-molecule templates
d Helicos BioSciences: two-pass sequencing
Single molecule: template immobilized
Billions of primed, single-molecule templates
e Pacific Biosciences, Life/Visigen, LI-COR Biosciences
Single molecule: polymerase immobilized
Thousands of primed, single-molecule templates

Complete genomics
A BGI Company

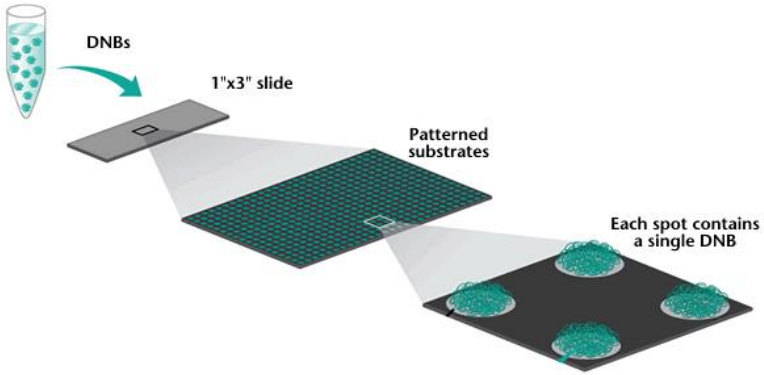
1



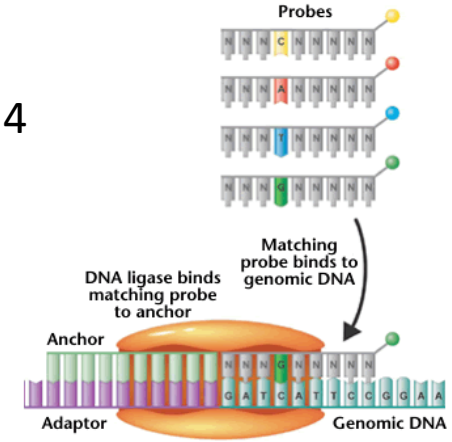
2

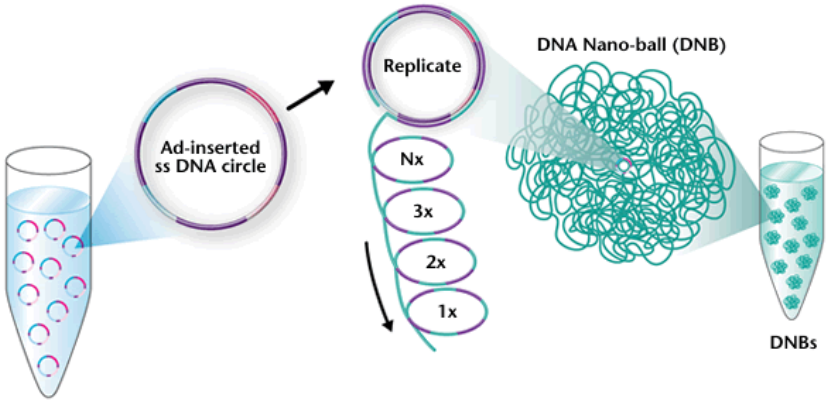


3

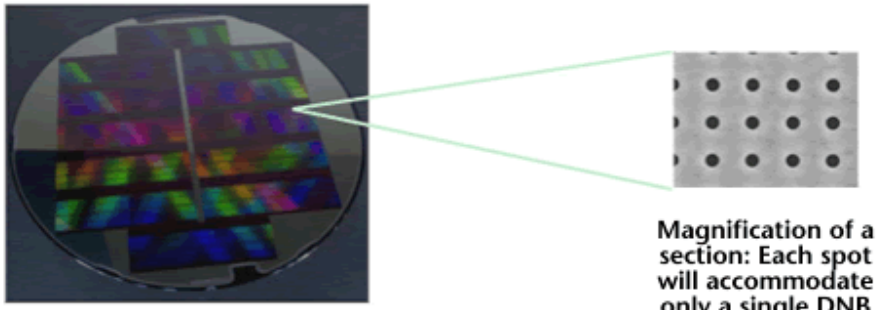


4





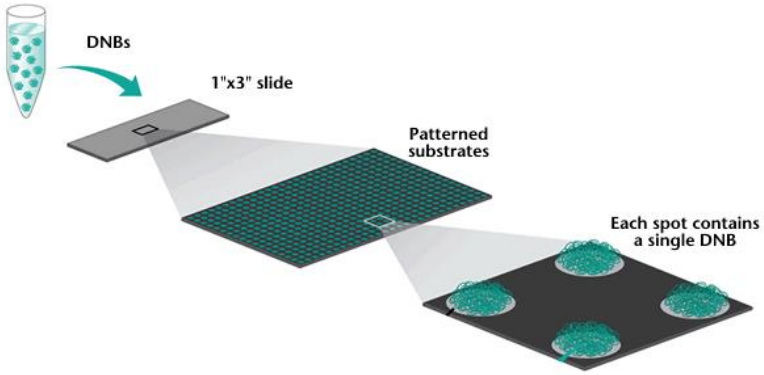
DNB Nanoball™ Arrays: a novel approach to preparing fragmented DNA which can be packed onto a silicon chip very efficiently. During library preparation, genomic DNA is fragmented and each fragment is then copied in a manner that results in a long single molecule containing hundreds of copies of the same fragment. Proprietary techniques developed by Complete Genomics cause each long single molecule to consolidate, or ball up, into a small particle of DNA referred to as a DNA nanoball, or DNB™. DNBs are approximately 200 nanometers in diameter and a library contains millions of DNBs that together represent the complete genome.



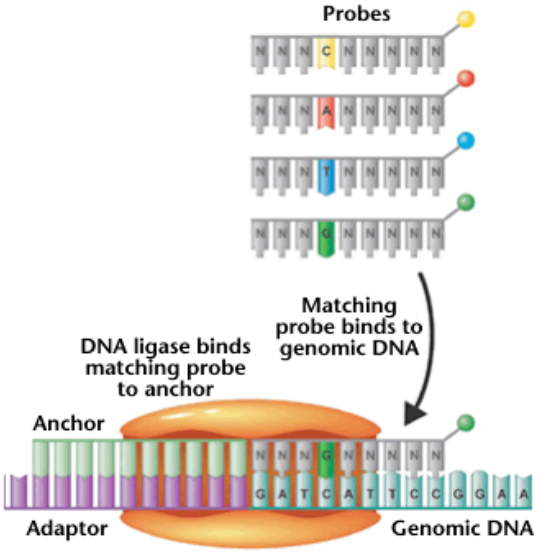
Silicon substrate consisting of 12 slides

Magnification of a section: Each spot will accommodate only a single DNB

The small size and biochemical characteristics of the DNBs enable the system to pack them together very tightly on a silicon chip. The system uses established photolithography processes developed in the semiconductor industry to create a silicon chip that has a grid pattern of small spots, which it is referred to as “sticky spots”.



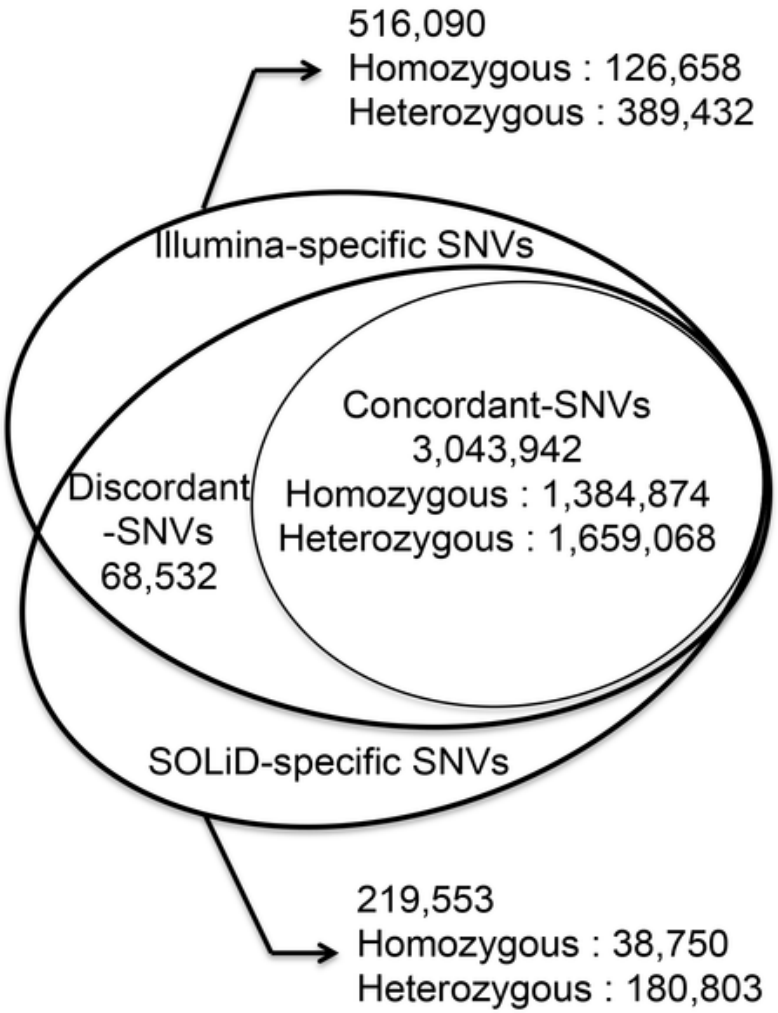
Complete Genomics has a proprietary process that fills over 90% of the sticky spots with exactly one DNB, without adherence of the DNA to the areas between the spots. A DNA nanoball array is a silicon chip filled with DNBs. Each finished DNA nanoball array contains up to 180 billion bases of genomic DNA prepared for imaging.



Combinatorial Probe-Anchor Ligation

Complete Genomics has developed a unique and highly accurate cPAL™ technology that allows the sequence of each DNB to be read very efficiently on the sequencing platform. The DNB contains both genomic DNA sequence and adapter sequence. During sequencing, an anchor probe binds to the adapter sequence. A ligase enzyme then attaches one of four possible fluorescent-labeled probes to the anchor, depending on the sequence being read in the fragment. By imaging the fluorescence during each ligation step, it is possible to subsequently determine the sequence of nucleotides in each DNB.

Figure 1. Concordance of SNVs identified by the two different sequencing platforms.



Kim D, Kim W-Y, Lee S-Y, Lee S-Y, et al. (2013) Revising a Personal Genome by Comparing and Combining Data from Two Different Sequencing Platforms. PLoS ONE 8(4): e60585. doi:10.1371/journal.pone.0060585
http://www.plosone.org/article/info:doi/10.1371/journal.pone.0060585

Table 1. Classification of sequenced SNVs and their chip-concordance.

			Illumina			SOLiD		
			# of SNVs	Chip-concordance	Median depth	# of SNVs	Chip-concordance	Median depth
		Total	390,494	98.92%	20	390,494	98.92%	35
	Chip-concordant	HOM	192,914	48.87%	20	192,914	48.87%	32
Concordant SNVs between platforms		HET	197,580	50.05%	20	197,580	50.05%	38
		Total	4,244	–	21	4,244	–	37
	Chip-discordant	HOM	564	–	19	564	–	29
		HET	3,680	–	21	3,680	–	38
		Total	2,489	90.08%	18	271	9.81%	29
	Chip-concordant	HOM	2,440	88.31%	17	127	4.60%	31
Discordant SNVs between platforms		HET	49	1.77%	18	144	5.21%	27
		Total	274	–	13	2,492	–	21
	Chip-discordant	HOM	144	–	12	50	–	10.5
		HET	130	–	13	2,442	–	21
Illumina-specific SNVs	Chip-concordant	Total	5,879	97.82%	18	–	–	–
	Chip-discordant	Total	131	–	20	–	–	–
SOLiD-specific SNVs	Chip-concordant	Total	-	–	–	3,565	97.11%	33
	Chip-discordant	Total	-	–	–	106	–	27

HOM, homozygous calls; HET, heterozygous calls; Median depth, median sequencing depth

doi:10.1371/journal.pone.0060585.t001

Kim D, Kim W-Y, Lee S-Y, Lee S-Y, et al. (2013) Revising a Personal Genome by Comparing and Combining Data from Two Different Sequencing Platforms. PLoS ONE 8(4): e60585. doi:10.1371/journal.pone.0060585
http://www.plosone.org/article/info:doi/10.1371/journal.pone.0060585

PacBio RS

Single molecule resolution in real time

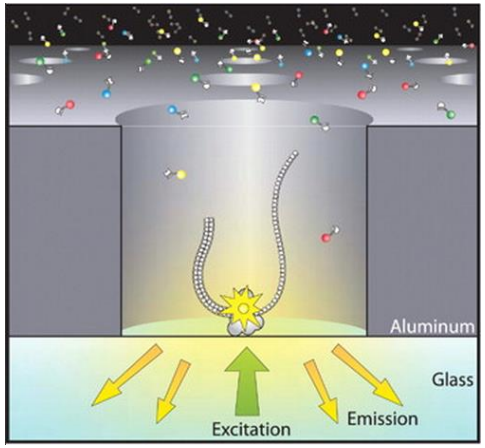
- Short waiting time for result and simple workflow
  - Generate basecalls in <1 day
  - Polymerase speed ≥1 base per second
- No amplification required
  - Bias not introduced
  - More uniform coverage
- Direct observation
  - Distinguish heterogeneous samples
  - Simultaneous kinetic measurements
- Long reads
  - Identify repeats and structural variants
  - Less coverage required
- Information content
  - One assay, multiple applications
    - Genetic variation (SVs to SNPs)
    - Methylation
    - Enzymology

C2 chemistry – installed March 2012
-Long reads 6-10kb
-Median size of molecules 3kb
-Still 15% error rate
-No strobe sequencing

Software focus on:
De novo assembly
Hi quality CCS consensus reads

In preparation

-Load long molecules by magnetic beads
-Modified nucleotides detection



PacBio RS – two sequencing modes

LS – long sequencing reads

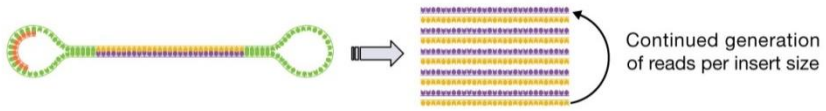
Sample Preparation

Standard

- Large insert sizes (2kb-10kb)
- Generates one pass on each molecule sequenced

CCS (Circular Consensus Sequencing) – high quality sequencing reads

Circular Consensus



- Small insert sizes 500bp
- Generates multiple passes on each molecule sequenced

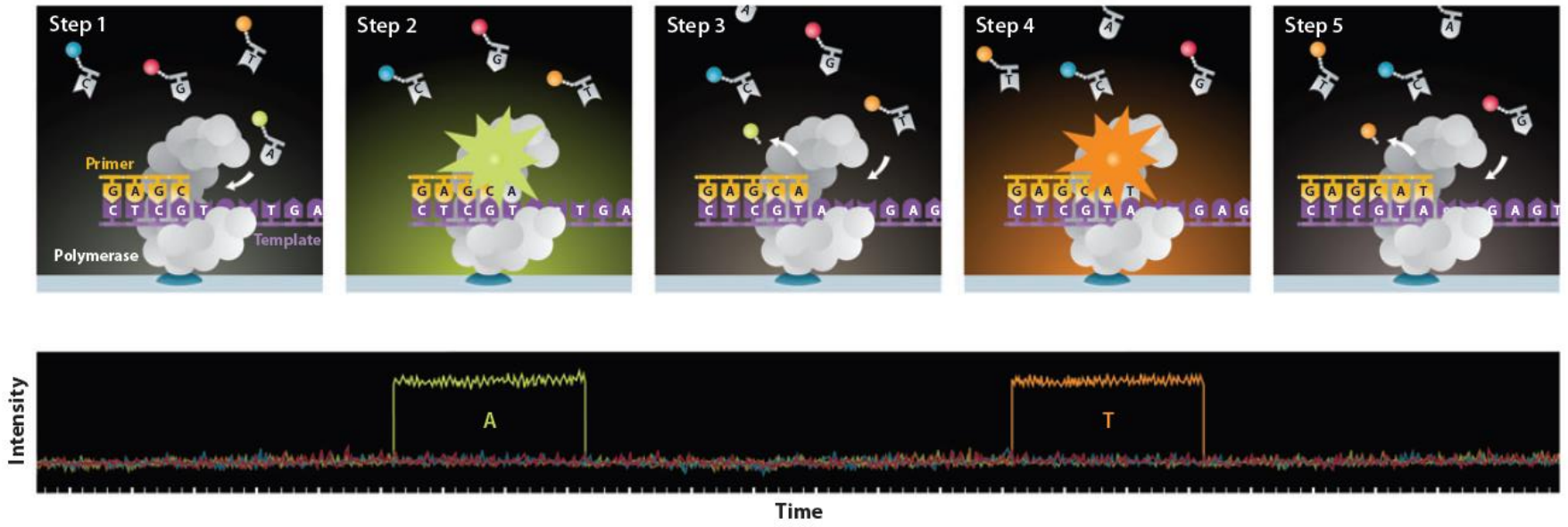
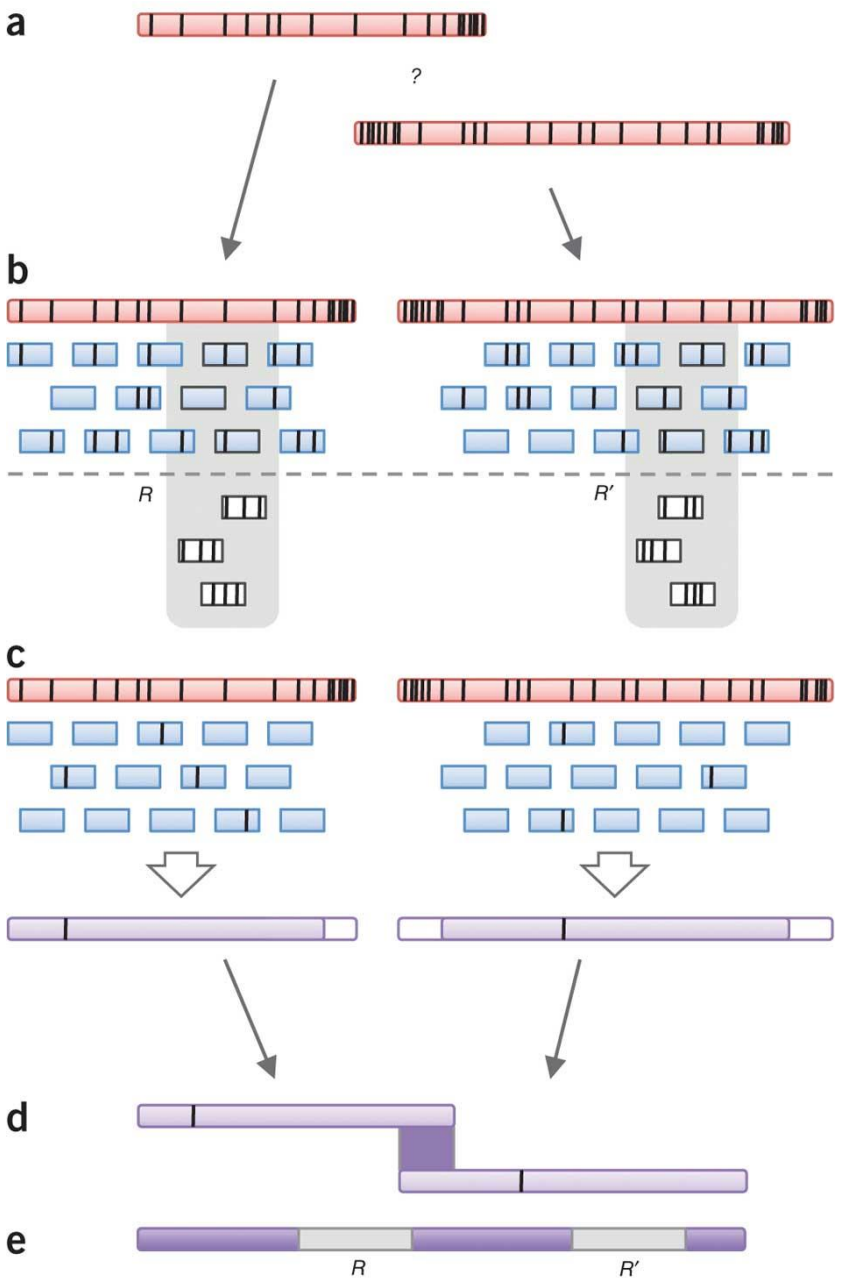


Figure 5

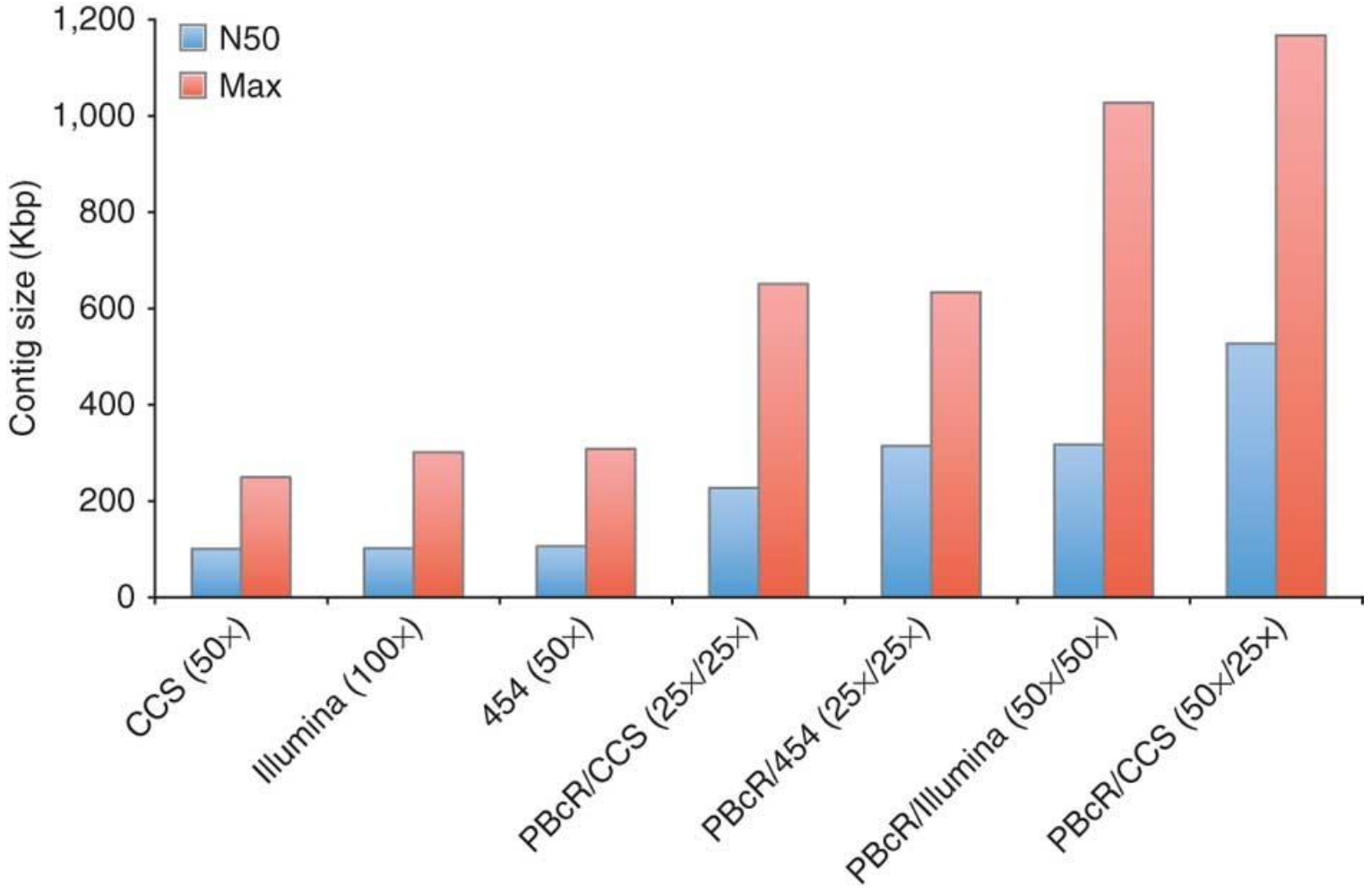
Single-molecule sequencing using Pacific Biosciences' zero-mode waveguides.

The PBcR single-molecule read correction and assembly method



(**a**) Errors, indicated by black vertical bars, in single-pass PacBio RS reads (pink rectangles) make it difficult to determine whether reads overlap. (**b**) Aligning high-fidelity short reads to error-prone long reads. Accurate alignments can be computed because the error between a short, high-accuracy sequence (~99% identical to the truth) and a PacBio RS sequence is half the error between two PacBio RS sequences. In this example, black bars in the short-reads indicate 'mapping errors' that are a combination of the sequencing error in both the long and short reads. In addition, a two-copy inexact repeat is present (outlined in gray) leading to pile-ups of reads at each copy. To avoid mapping reads to the wrong repeat copy, the algorithm selects a cutoff, *C*, and only the top *C* hits for each short read are used. The spurious mappings (in white) are discarded. (**c**) The remaining alignments are used to generate a new consensus sequence (purple), trimming and splitting long reads whenever there is a gap in the short-read tiling. Sequencing errors, indicated in black, may propagate to the PBcR read in rare cases where sequencing error co-occurs. (**d**) After correction, overlaps between long PBcR sequences can be easily detected. (**e**) The resulting assembly is able to span repeats that are unresolvable using only the short reads.

Contig sizes for various combinations of sequencing technologies



Assemblies are for *E. coli* C227-11 (assemblies including Illumina and PacBio CCS) and *E. coli* JM221 (assemblies including 454). Both genomes have similar repeat content, PacBio read length and coverage. Assemblies of only second-generation data are comparable and average N50 ≈ 100 Kbp. By comparison, adding 25× or 50× of PBcR to these data sets increases N50 as much as fivefold and results in a maximum contig size of greater than 1 Mbp (for the PBcR and CCS combination).

Library Preparation from Low Quantities of DNA or RNA

Microfluidics stationary and portable systems

Mondrian SP System – NuGEN Technologies

- Human libraries from 5ng of total DNA. Only 10-15% of duplicate reads.
- Ultralow DNA library systems

Soon:

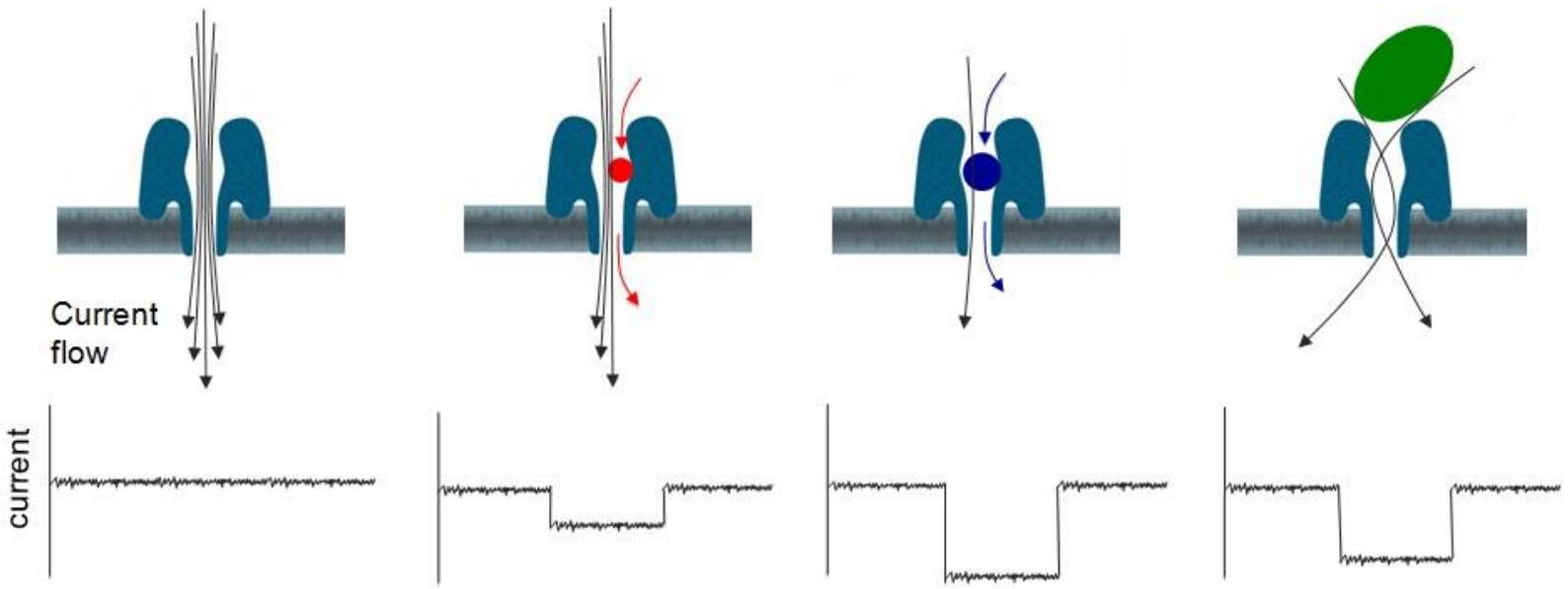
- Ultralow RNA library systems
- Libraries from total RNA with rRNA depletion.

Advanced Liquid Logic from RTP

Table 1
The platforms and the detailed information for the NGS technologies.

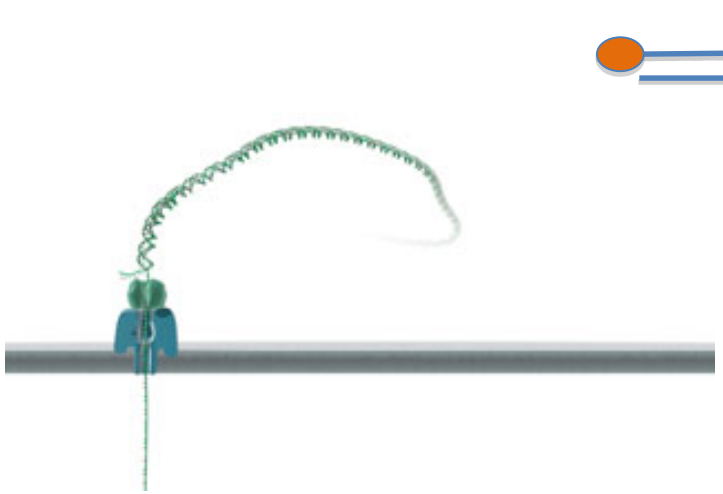
Technology	Amplification	Read length	Throughput	Sequence by synthesis
Currently available				
Roche/GS-FLX Titanium	Emulsion PCR	400−600 bp	500 Mbp/run	Pyrosequencing
Illumina/HiSeq 2000, HiScan	Bridge PCR (Cluster PCR)	2 × 100 bp	200 Gbp/run	Reversible terminators
ABI/SOLiD 5500xl	Emulsion PCR	50−100 bp	>100 Gbp/run	Sequencing-by-ligation (octamers)
Polonator/G.007	Emulsion PCR	26 bp	8−10 Gbp/run	Sequencing-by-ligation (monomers)
Helicos/Heliscope	No	35 (25−55) bp	21−37 Gbp/run	True single-molecule sequencing (tSMS)
In development				
Pacific BioSciences/RS	No	1000 bp	N/A	Single-molecule real time (SMRT)
Visigen Biotechnologies	No	>100 Kbp	N/A	Base-specific FRET
U.S. Genomics	No	N/A	N/A	Single-molecule mapping
Genovoxx	No	N/A	N/A	Single-molecule sequencing by synthesis
Oxford Nanopore Technologies	No	35 bp	N/A	Nanopores/exonuclease-coupled
NABsys	No	N/A	N/A	Nanopores
Electronic BioSciences	No	N/A	N/A	Nanopores
BioNanomatrix/nanoAnalyzer	No	400 Kbp	N/A	Nanochannel arrays
GE Global Research	No	N/A	N/A	Closed complex/nanoparticle
IBM	No	N/A	N/A	Nanopores
LingVitae	No	N/A	N/A	Nanopores
Complete Genomics	No	70 bp	N/A	DNA nanoball arrays
base4innovation	No	N/A	N/A	Nanostructure arrays
CrackerBio	No	N/A	N/A	Nanowells
Reveo	No	N/A	N/A	Nano-knife edge
Intelligent BioSystems	No	N/A	N/A	Electronics
LightSpeed Genomics	No	N/A	N/A	Direct-read sequencing by EM
Halcyon Molecular	No	N/A	N/A	Direct-read sequencing by EM
ZS Genetics	No	N/A	N/A	Direct-read sequencing by TEM
Ion Torrent/PostLight	No	N/A	N/A	Semiconductor-based pH sequencing
Genizon BioSciences/CGA	No	N/A	N/A	Sequencing-by-hybridization

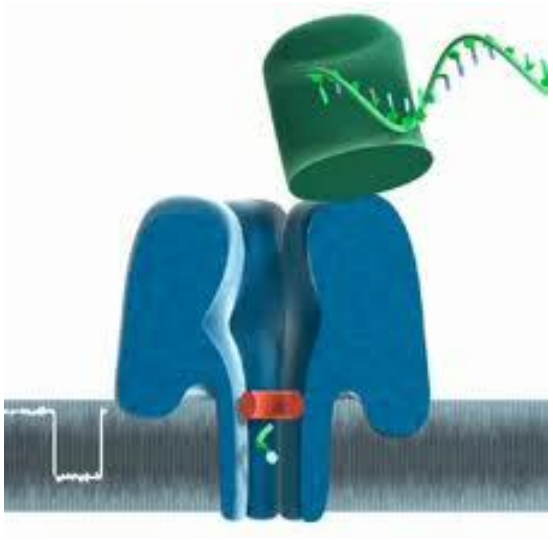
Oxford Nanopore

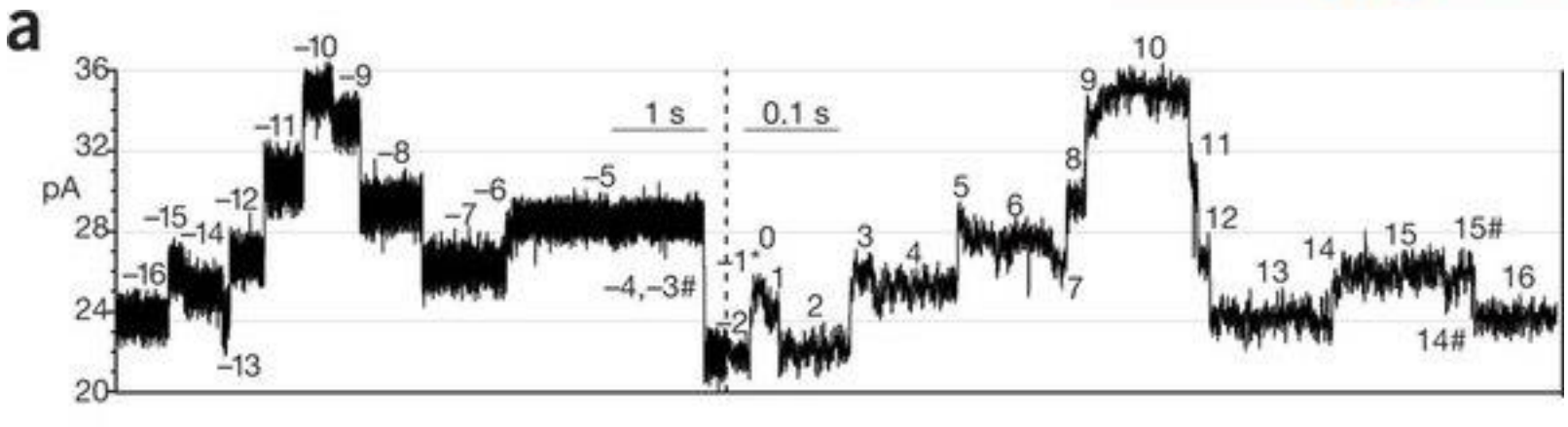


Hemolysin – pore - inner diameter of 1nm, about 100,000 times smaller than that of a human hair.

Oxford Nanopore



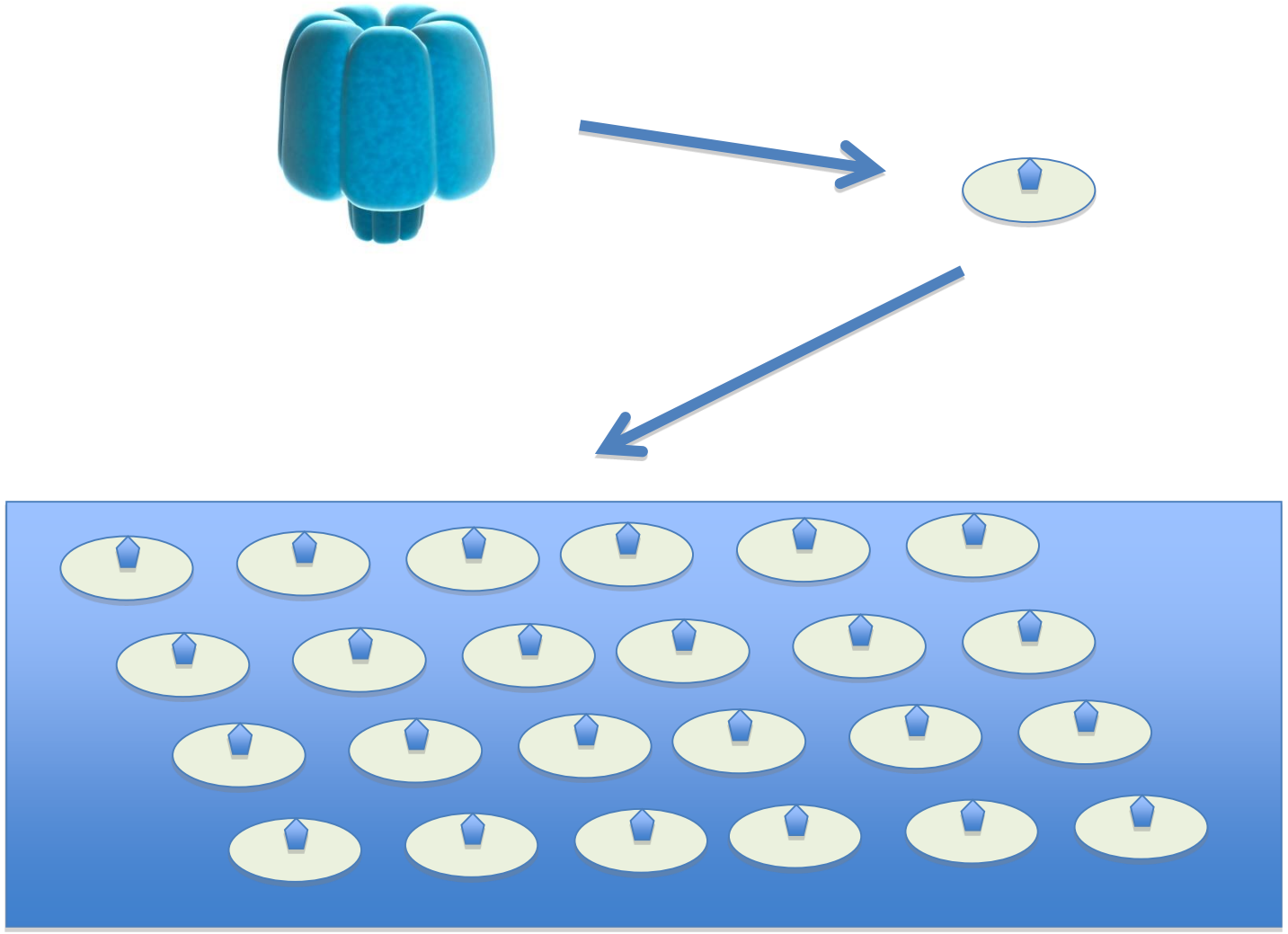




DNA sequencing

Error rate 4%, prediction for end of the year 0.1 – 2%.

Nanopore array



Oxford Nanopore – new concepts

MinION

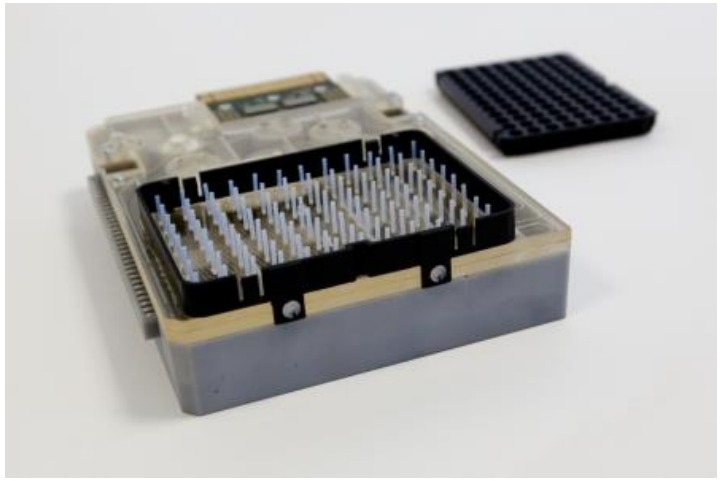
- 150Mb per run
- Tested 48kb read length
-$900 per instrument
-500 pores per device

GridION

- XXXMb per run
- Tested 48kb read length
-$XXX per instrument
-2000 pores per device, soon 8000 pores
-Cost per human genome $1500.

Oxford Nanopore – applications

- DNA sequencing
- Protein detection
- Protein DNA interaction
- Small molecule detection



- 96 well plates for 96 samples
- Controlled time of sequencing

Intelligent BioSystems Mini20 System (manufactured by Azco Biotech)

- Amplification by rolony method
- Sequencing by Synthesis with announced 100 base reads, but expect to compete with Sanger down the road
- Designed for clinical labs
- 20 independent flow cells, no queue for loading, run asynchronously
- 20M reads/flow cell, 4 GB/ flow cell
- Potential problems with repeats
- System cost $120K, $150 flow cell (disposable), full costs per sample not clear yet.

Genia Technologies

- **Very** early stage announcement – Backed by Life Technologies
  (at least 1 year away)
- Describe system as a cross between Ion Torrent and Oxford Nanopore
- Electronic "Active Control" technology enables highly efficient nanopore-membrane assembly and control of DNA movement through the channel
- Initially used α-Hemolysin and claimed 98% raw accuracy with that but now are using an undisclosed pore for further development.
- Claim sensitivity 1-2 orders of magnitude greater than Oxford Nanopore.
- Ramping up pore density to 100K pores/chip.
- Plan to market a mobile reader for <$1K and per sample costs <$100

Quail et al. BMC Genomics 2012, 13:341
http://www.biomedcentral.com/1471-2164/13/341

BMC Genomics

RESEARCH ARTICLE Open Access

A tale of three next generation sequencing platforms: comparison of Ion Torrent, Pacific Biosciences and Illumina MiSeq sequencers

Michael A Quail*, Miriam Smith, Paul Coupland, Thomas D Otto, Simon R Harris, Thomas R Connor, Anna Bertoni, Harold P Swerdlow and Yong Gu

Table 1 Technical specifications of Next Generation Sequencing platforms utilised in this study

Platform	Illumina MiSeq	Ion Torrent PGM	PacBio RS	Illumina GAIIx	Illumina HiSeq 2000
Instrument Cost*	$128 K	$80 K**	$695 K	$256 K	$654 K
Sequence yield per run	1.5-2Gb	20-50 Mb on 314 chip, 100-200 Mb on 316 chip, 1Gb on 318 chip	100 Mb	30Gb	600Gb
Sequencing cost per Gb*	$502	$1000 (318 chip)	$2000	$148	$41
Run Time	27 hours***	2 hours	2 hours	10 days	11 days
Reported Accuracy	Mostly > Q30	Mostly Q20	<Q10	Mostly > Q30	Mostly > Q30
Observed Raw Error Rate	0.80 %	1.71 %	12.86 %	0.76 %	0.26 %
Read length	up to 150 bases	~200 bases	Average 1500 bases**** (C1 chemistry)	up to 150 bases	up to 150 bases
Paired reads	Yes	Yes	No	Yes	Yes
Insert size	up to 700 bases	up to 250 bases	up to 10 kb	up to 700 bases	up to 700 bases
Typical DNA requirements	50-1000 ng	100-1000 ng	~1 μg	50-1000 ng	50-1000 ng

* All cost calculations are based on list price quotations obtained from the manufacturer and assume expected sequence yield stated.
** System price including PGM, server, OneTouch and OneTouch ES.
*** Includes two hours of cluster generation.
**** Mean mapped read length includes adapter and reverse strand sequences. Subread lengths, i.e. the individual stretches of sequence originating from the sequenced fragment, are significantly shorter.

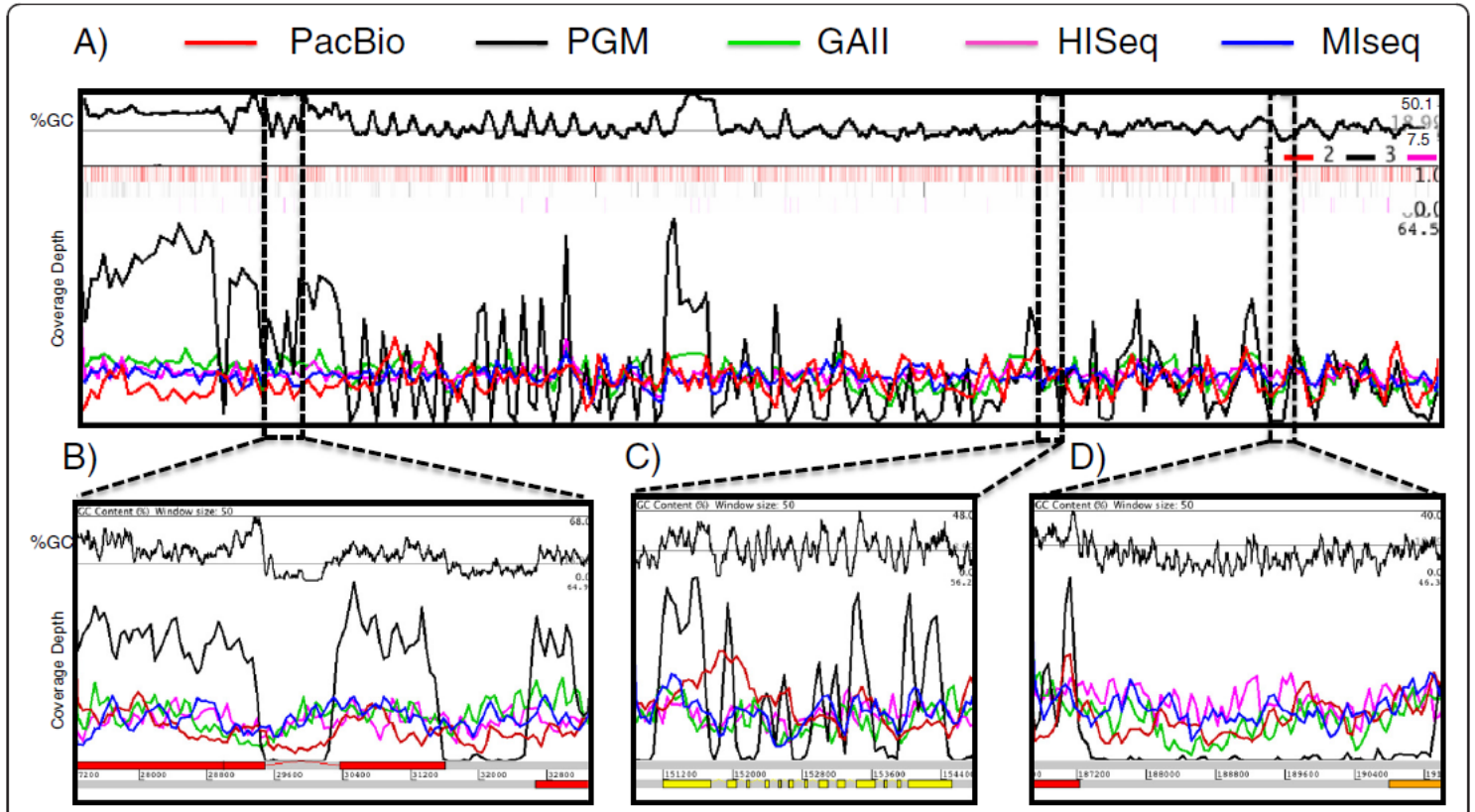


Figure 2 Artemis genome browser [8] **screenshots illustrating the variation in sequence coverage of a selected region of *P. falciparum* chromosome 11, with 15x depth of randomly normalized sequence from the platforms tested.** In each window, the top graph shows the percentage GC content at each position, with the numbers on the right denoting the minimum, average and maximum values. The middle graph in each window is a coverage plot for the dataset from each instrument; the colour code is shown above graph a). Each of the middle graphs shows the depth of reads mapped at each position, and below that in B-D are the coordinates of the selected region in the genome with gene models on the (+) strand above and (–) strand below. **A**) View of the first 200 kb of chromosome 11. Graphs are smoothed with window size of 1000. A heatmap of the errors, normalized by the amount of mapping reads is included just below the GC content graph (PacBio top line, PGM middle and MiSeq bottom). **B**) Coverage over region of extreme GC content, ranging from 70% to 0%. **C**) Coverage over the gene PF3D7_1103500. **D**) Example of intergenic region between genes PF3D7_1104200 and PF3D7_1104300. The window size of B, C and D is 50 bp.

Hindawi Publishing Corporation
Journal of Biomedicine and Biotechnology
Volume 2012, Article ID 251364, 11 pages
doi:10.1155/2012/251364

Review Article

Comparison of Next-Generation Sequencing Systems

Lin Liu, Yinhu Li, Siliang Li, Ni Hu, Yimin He, Ray Pong, Danni Lin, Lihua Lu, and Maggie Law

NGS Sequencing Department, Beijing Genomics Institute (BGI), 4th Floor, Building 11, Beishan Industrial Zone, Yantian District, Guangdong, Shenzhen 518083, China

Table 1: (a) Advantage and mechanism of sequencers. (b) Components and cost of sequencers. (c) Application of sequencers.

(a)

Sequencer	454 GS FLX	HiSeq 2000	SOLiDv4	Sanger 3730xl
Sequencing mechanism	Pyrosequencing	Sequencing by synthesis	Ligation and two-base coding	Dideoxy chain termination
Read length	700 bp	50SE, 50PE, 101PE	50 + 35 bp or 50 + 50 bp	400~900 bp
Accuracy	99.9%*	98%, (100PE)	99.94% *raw data	99.999%
Reads	1 M	3 G	1200~1400 M	—
Output data/run	0.7 Gb	600 Gb	120 Gb	1.9~84 Kb
Time/run	24 Hours	3~10 Days	7 Days for SE 14 Days for PE	20 Mins~3 Hours
Advantage	Read length, fast	High throughput	Accuracy	High quality, long read length
Disadvantage	Error rate with polybase more than 6, high cost, low throughput	Short read assembly	Short read assembly	High cost low throughput

(b)

Sequencers	454 GS FLX	HiSeq 2000	SOLiDv4	3730xl
Instrument price	Instrument $500,000, $7000 per run	Instrument $690,000, $6000/(30x) human genome	Instrument $495,000, $15,000/100 Gb	Instrument $95,000, about $4 per 800 bp reaction
CPU	2* Intel Xeon X5675	2* Intel Xeon X5560	8* processor 2.0 GHz	Pentium IV 3.0 GHz
Memory	48 GB	48 GB	16 GB	1 GB
Hard disk	1.1 TB	3 TB	10 TB	280 GB
Automation in library preparation	Yes	Yes	Yes	No
Other required device	REM e system	cBot system	EZ beads system	No
Cost/million bases	$10	$0.07	$0.13	$2400

(c)

Sequencers	454 GS FLX	HiSeq 2000	SOLiDv4	3730xl
Resequencing		Yes	Yes	
De novo	Yes	Yes		Yes
Cancer	Yes	Yes	Yes	
Array	Yes	Yes	Yes	Yes
High GC sample	Yes	Yes	Yes	
Bacterial	Yes	Yes	Yes	
Large genome	Yes	Yes		
Mutation detection	Yes	Yes	Yes	Yes

(1) All the data is taken from daily average performance runs in BGI. The average daily sequence data output is about 8 Tb in BGI when about 80% sequencers (mainly HiSeq 2000) are running.
(2) The reagent cost of 454 GS FLX Titanium is calculated based on the sequencing of 400 bp; the reagent cost of HiSeq 2000 is calculated based on the sequencing of 200 bp; the reagent cost of SOLiDv4 is calculated based on the sequencing of 85 bp.
(3) HiSeq 2000 is more flexible in sequencing types like 50SE, 50PE, or 101PE.
(4) SOLiD has high accuracy especially when coverage is more than 30x, so it is widely used in detecting variations in resequencing, targeted resequencing, and transcriptome sequencing. Lanes can be independently run to reduce cost.