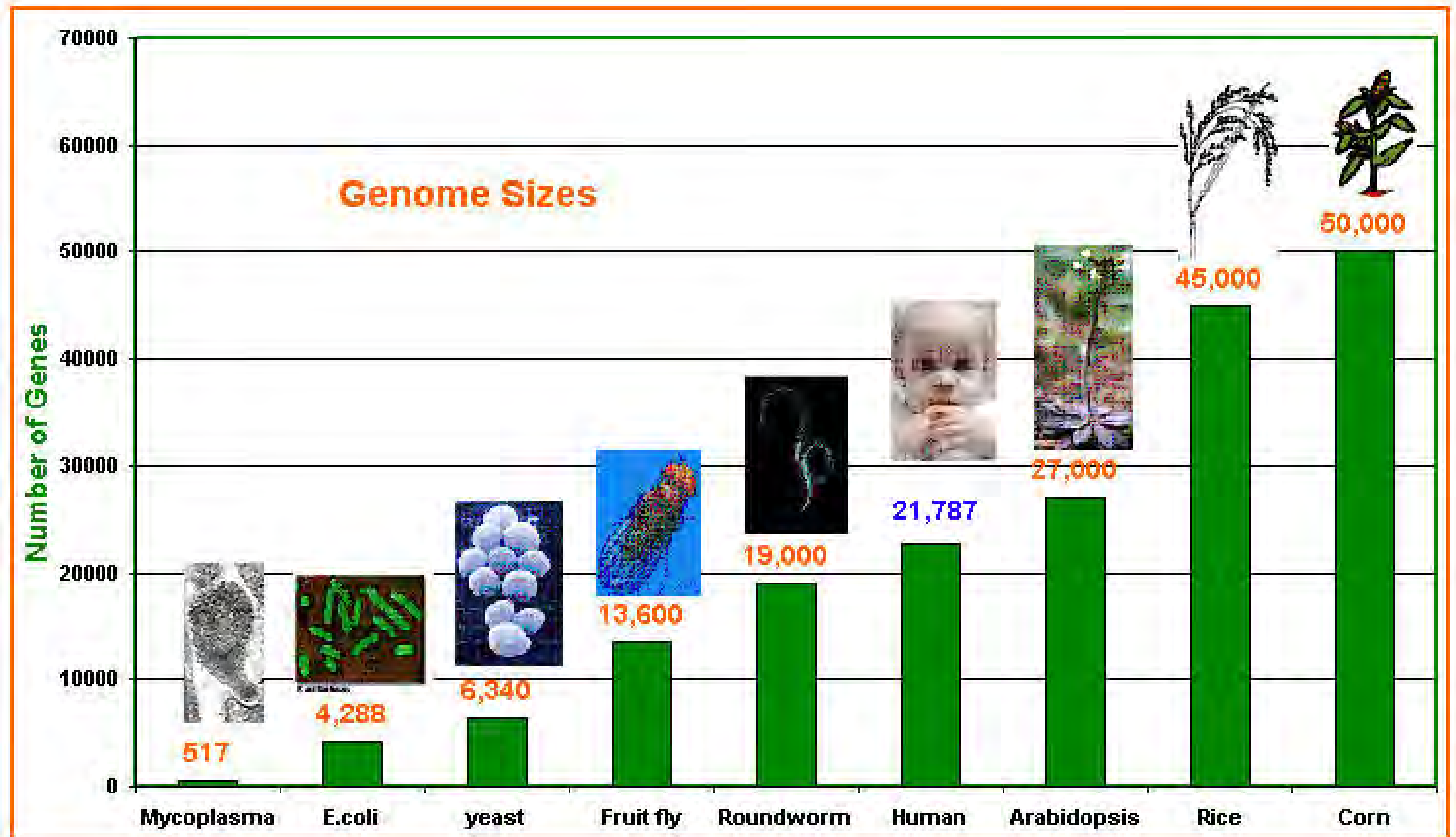
PROCOGEN 2nd training workshop

New technologies to sequence complex genomes

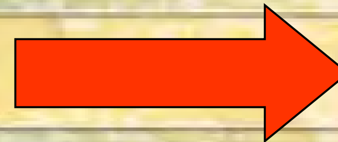
Prof. José Luis García
CIB-CSIC

20 February, 2014

number of genes



Pine Tree Genome



20,000 million
18,000 million
16,000 million
14,000 million
12,000 million
10,000 million
8,000 million
6,000 million
4,000 million
2,000 million

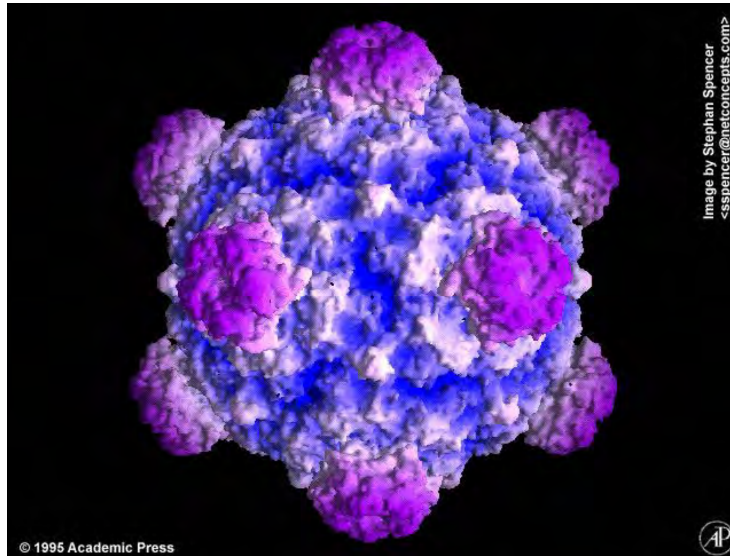
base pairs

E. coli	yeast	C. elegans	A. thaliana	fruit fly	pufferfish	chicken	zebrafish	dog	mouse	human	pine tree
4.7	12	100	125	130	335	1,054	1,560	2,400	2,615	3,100	18,000

FRED SANGER

WALTER GILBERT

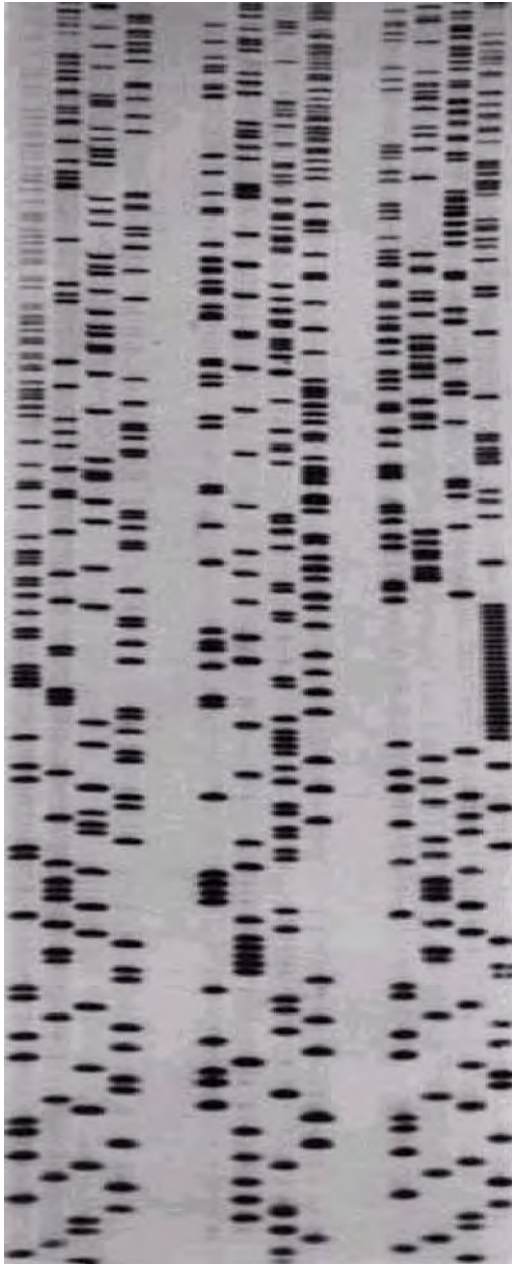
1977



Phage X174

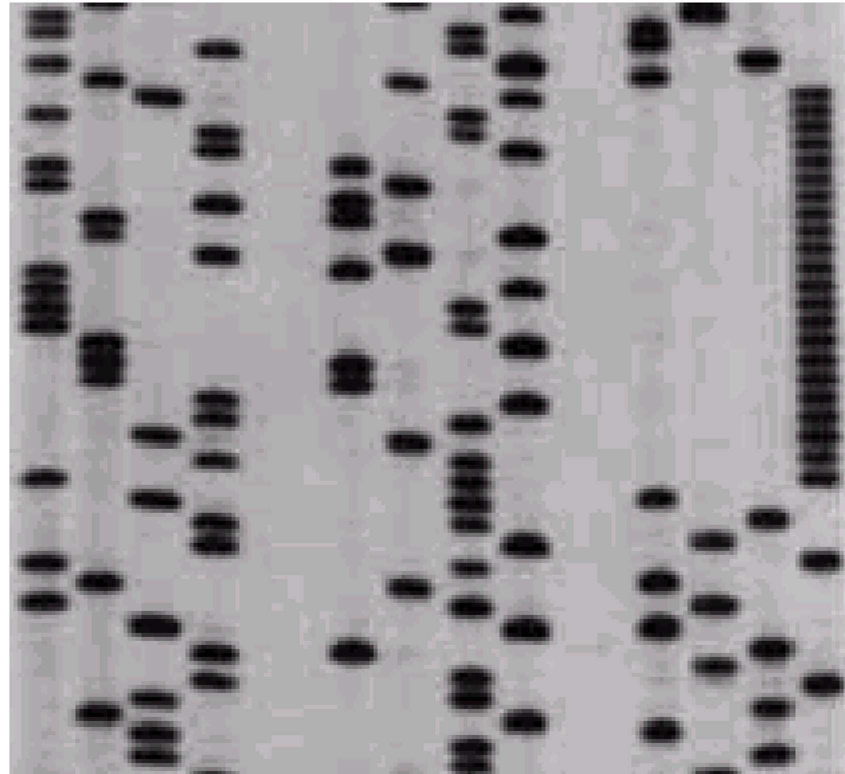
5368 bp

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

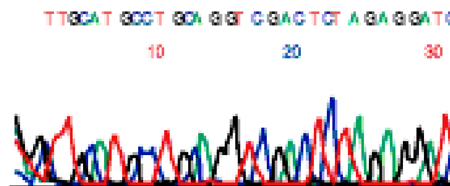


1980-1990

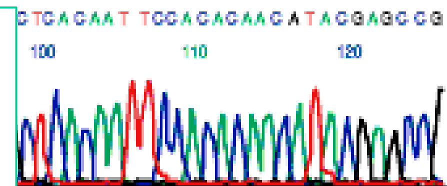
A G C T A G C T A G C T

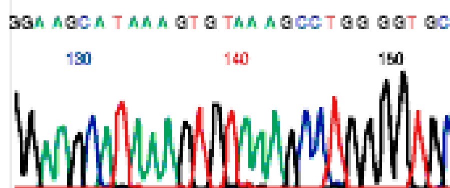


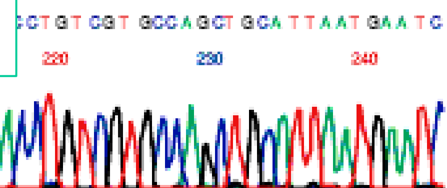
MANUAL

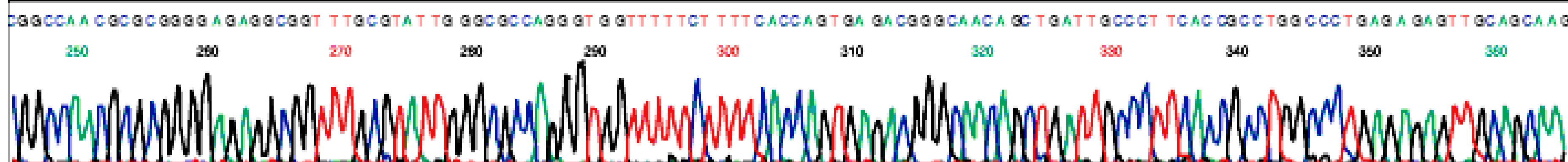


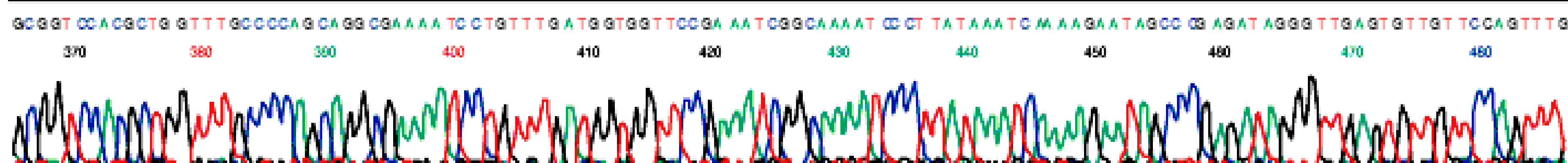
1990-2005

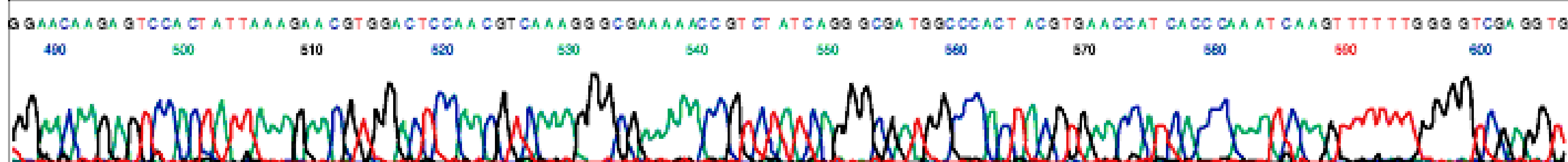


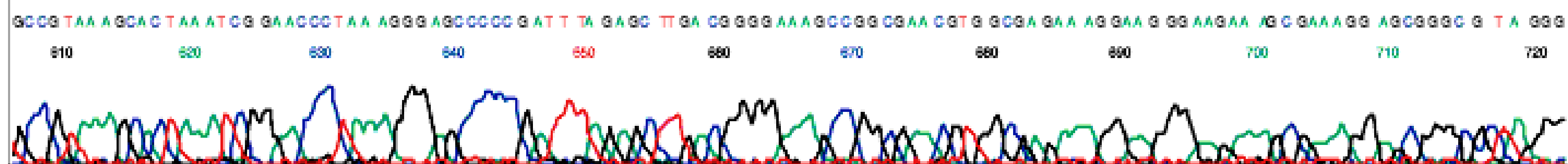






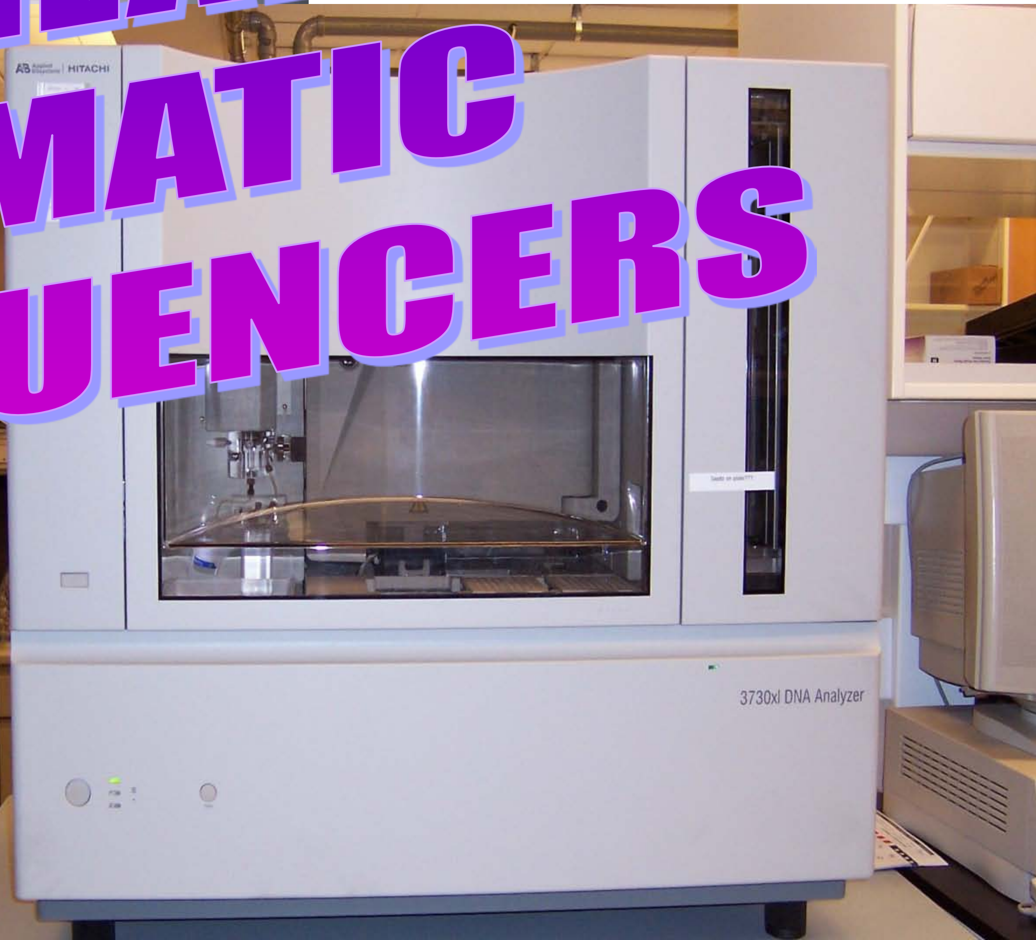






ABI 3700

CAPILAR AUTOMATIC DNA SEQUENCERS



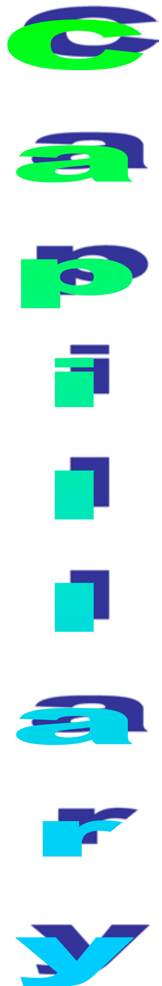
ABI 3730

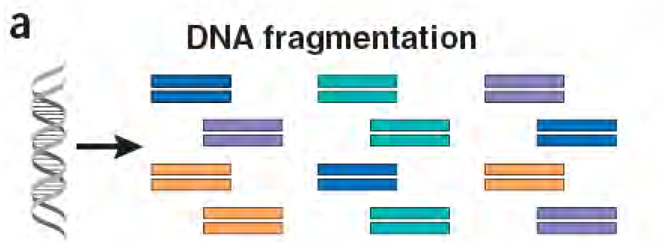
Next Generation Sequencing

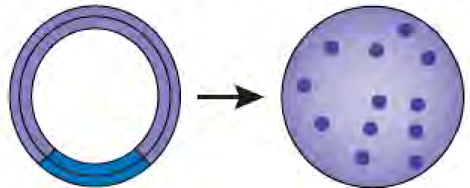
2005 - now

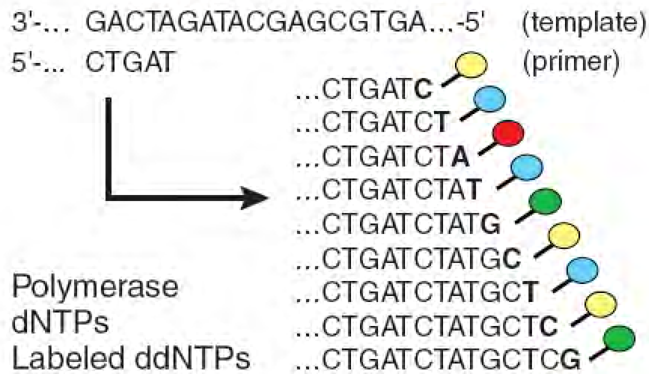
SEQUENCING METHODS

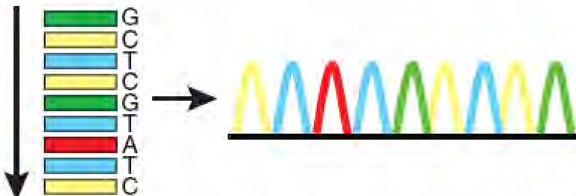
- Sequencing by Electrophoresis
  - Capillary Electrophoresis
  - Microelectrophoretic methods
- Cyclic array sequencing
  - Sequencing by ligation
  - Sequencing by synthesis
- Sequencing by hybridization
- Sequencing in real time

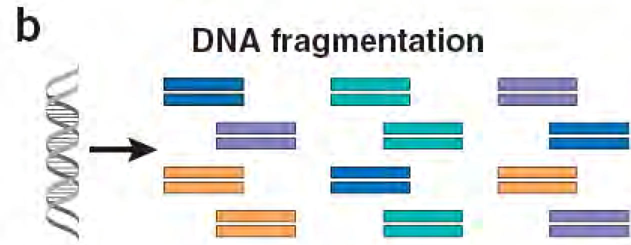


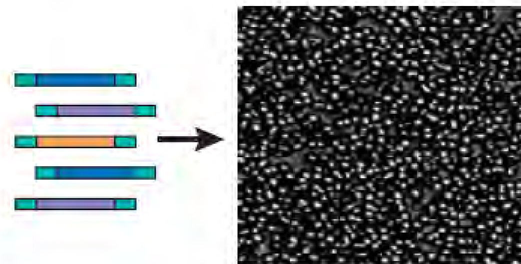


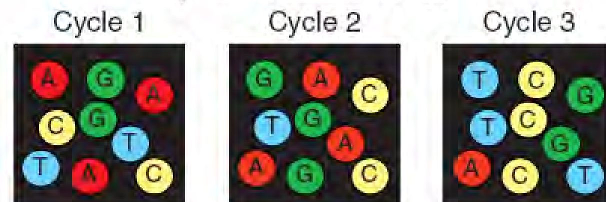






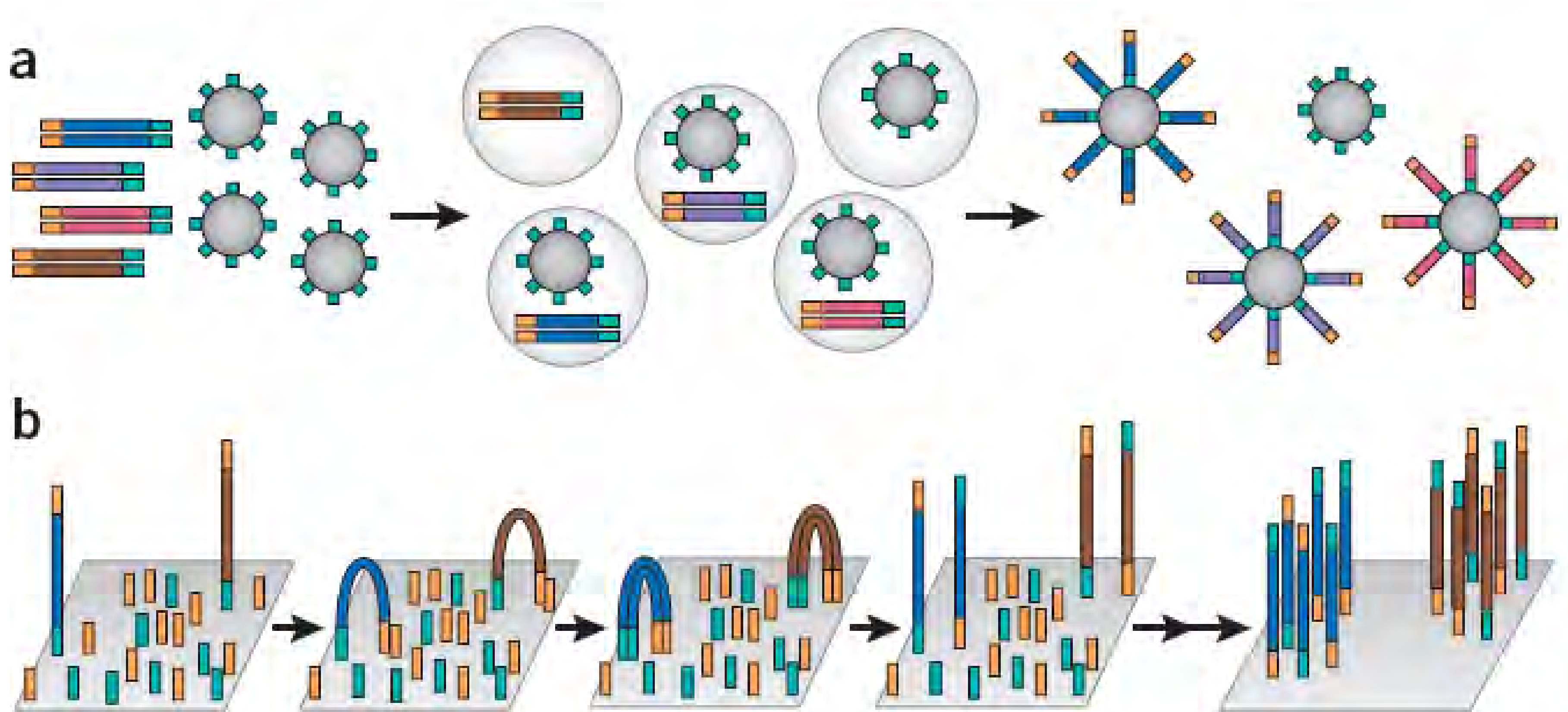




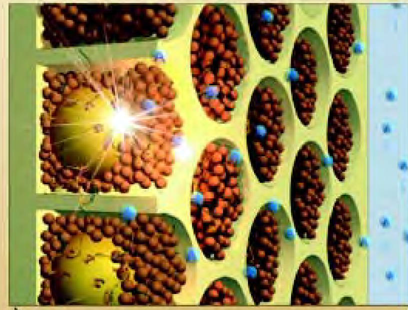


Nature Biotech

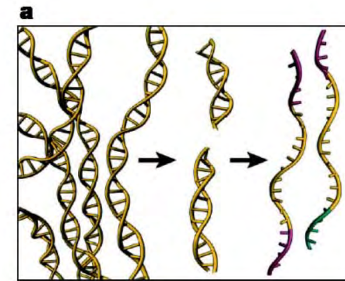
CLONAL AMPLIFICATION

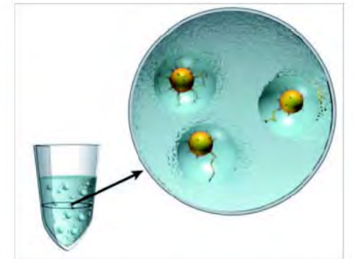


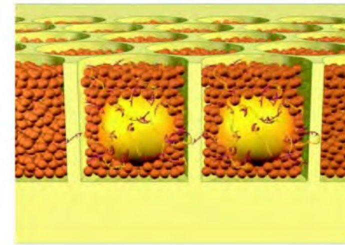
Nature Biotech

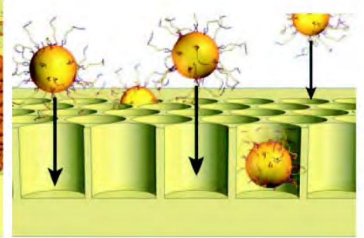


454







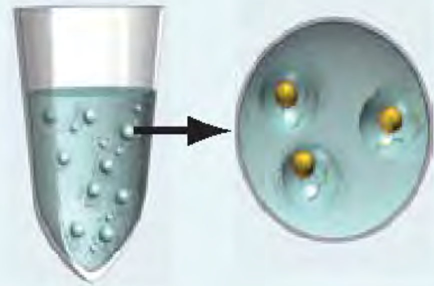


700 bp/read

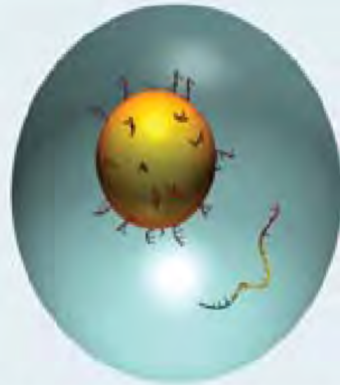
700 Mb/run

DNA Library Preparation and Titration | **emPCR** | **Sequencing**

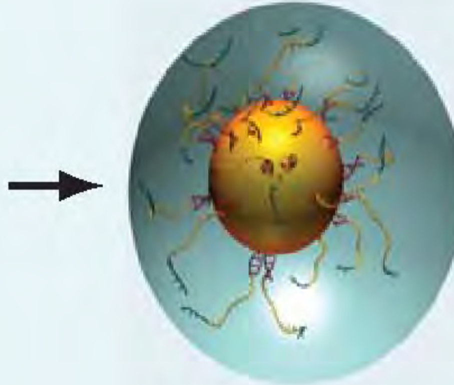
4.5 hours 10.5 hours | 8 hours | 4.5 hours



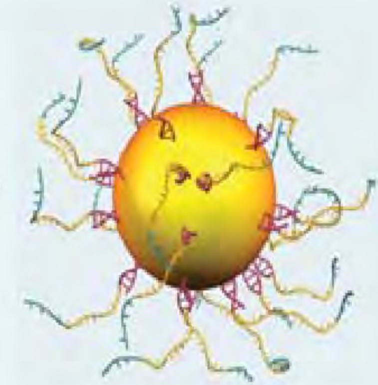
Anneal sstDNA to an excess of DNA Capture Beads



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



Clonal amplification occurs inside microreactors

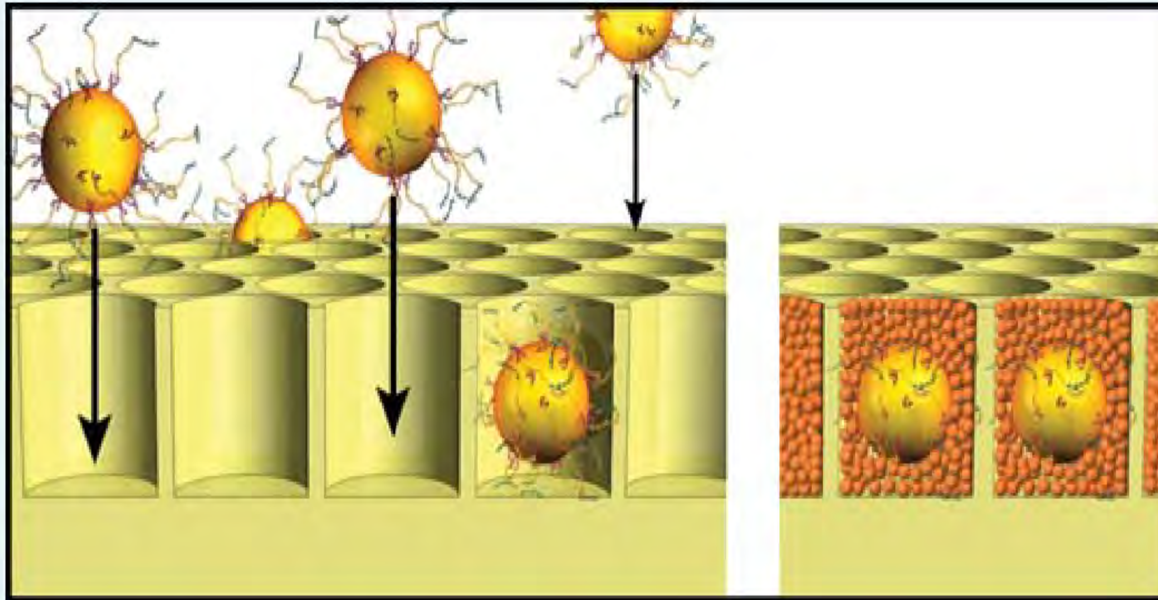


Break microreactors, enrich for DNA-positive beads

sstDNA library **Clonally-amplified sstDNA attached to bead**

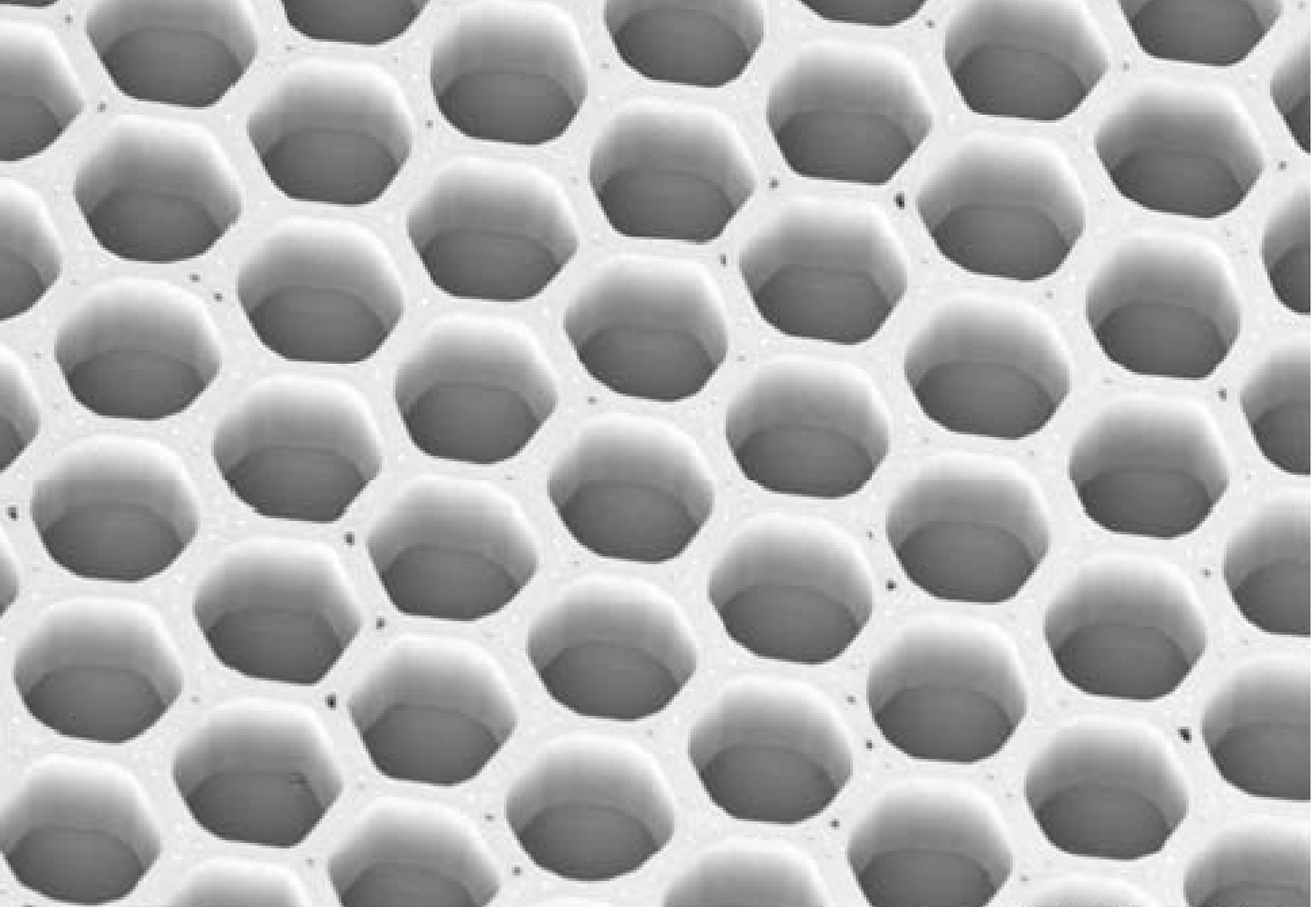
DNA Library Preparation and Titration → **emPCR** → **Sequencing**

4.5 hours — 10.5 hours — 8 hours — 4.5 hours



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

Amplified sstDNA library beads **Quality filtered bases**


MAS 3.0kV 25.4mm x350 SE(U) 11/25/02
100um

illumina

~30x coverage of two human genomes in a single run for under $10,000 (per sample).

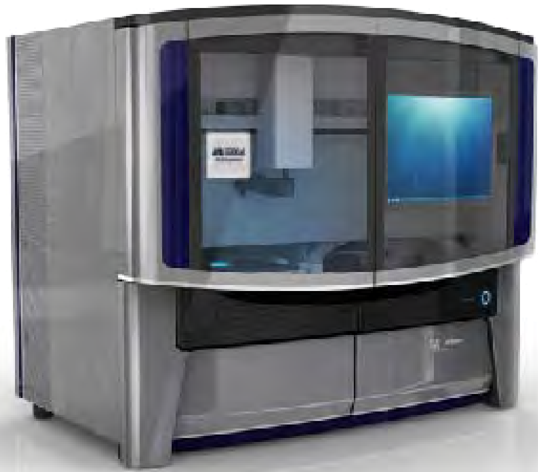
Two run modes. High output or rapid run.

Flexibility to batch process multiple samples with high output in a single run, or get rapid results with fewer samples for time-critical studies.

	HiSeq 2500		HiSeq 1500	
Run Mode	High Output	Rapid Run*	High Output	Rapid Run*
Output (2 × 100 bp)	600 Gb	120 Gb	300 Gb	60 Gb
Run Time (2 × 100 bp)	~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
Bases Above Q30***	> 85% (2 x 50 bp) > 80% (2 x 100 bp)			

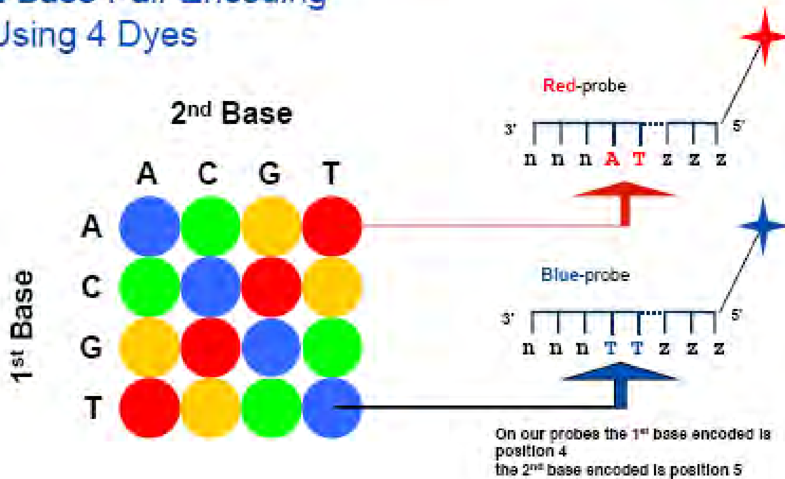
HiSeq 2500

VIDEO ILLUMINA



SOLiD 5500 XL

SECUENCIACIÓN POR LIGACIÓN



2 cells/run (6 x 2 lanes)

180 Gb/run microbeads

300 Gb/run nanobeads

75 bp fragment

75 x 35 bp paired-end

60 x 60 bp mate pair

1 day 35 bp 1 lane

7 days paired-end/mate pair 12 lanes

454 LIFE SCIENCES

35 Mb/run

100 000 reads

400 bp/read

GS Junior System

LOS OTROS PEQUEÑOS

MiSeq

SOLiD Pi

illumina

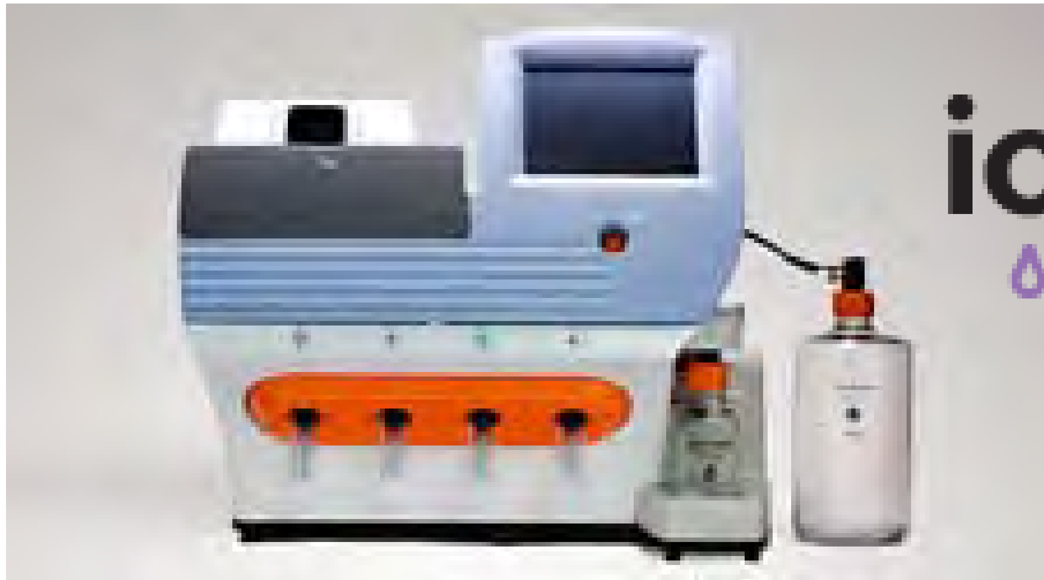
9 Exomas humanos

18 horas

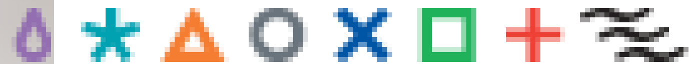
1 Genoma humano

29 horas

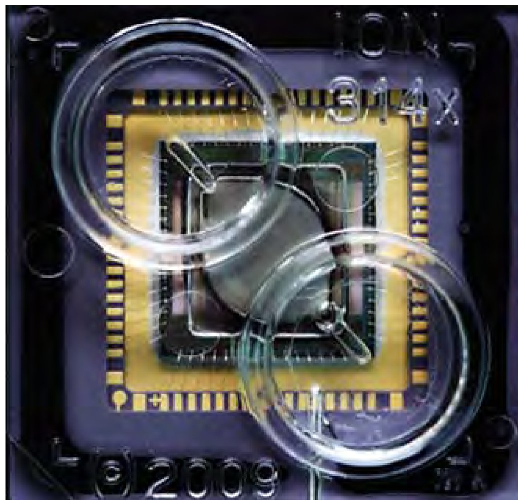
NextSeq 500 Desktop Sequencer

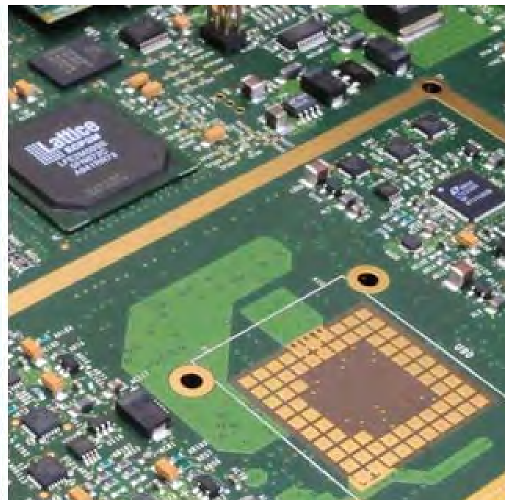


ion torrent



Ion Personal Genome Machine (PGM™)





50 000 $

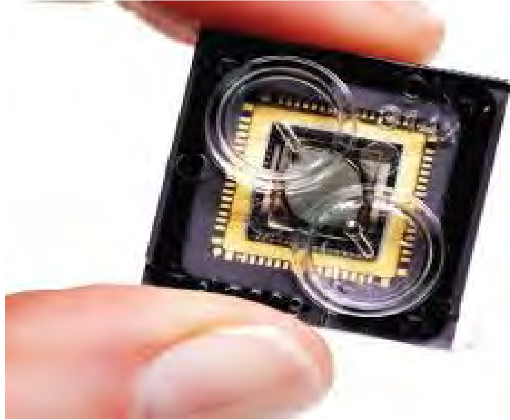
500 $ run

1.55 M wells

200-400 bp/read

1-2 h run

ION TORRENT CHIPS



ION 314 chip

1.3 M wells

10 Mb/run

ION 316 chip

6.3 M wells

100 Mb/run

ION 318 chip (new)

11 M wells

1 Gb/run

Ion Hi-Q™ Sequencing Chemistry

400 bp/read

ION PROTON

PROTON I chip

165 M wells

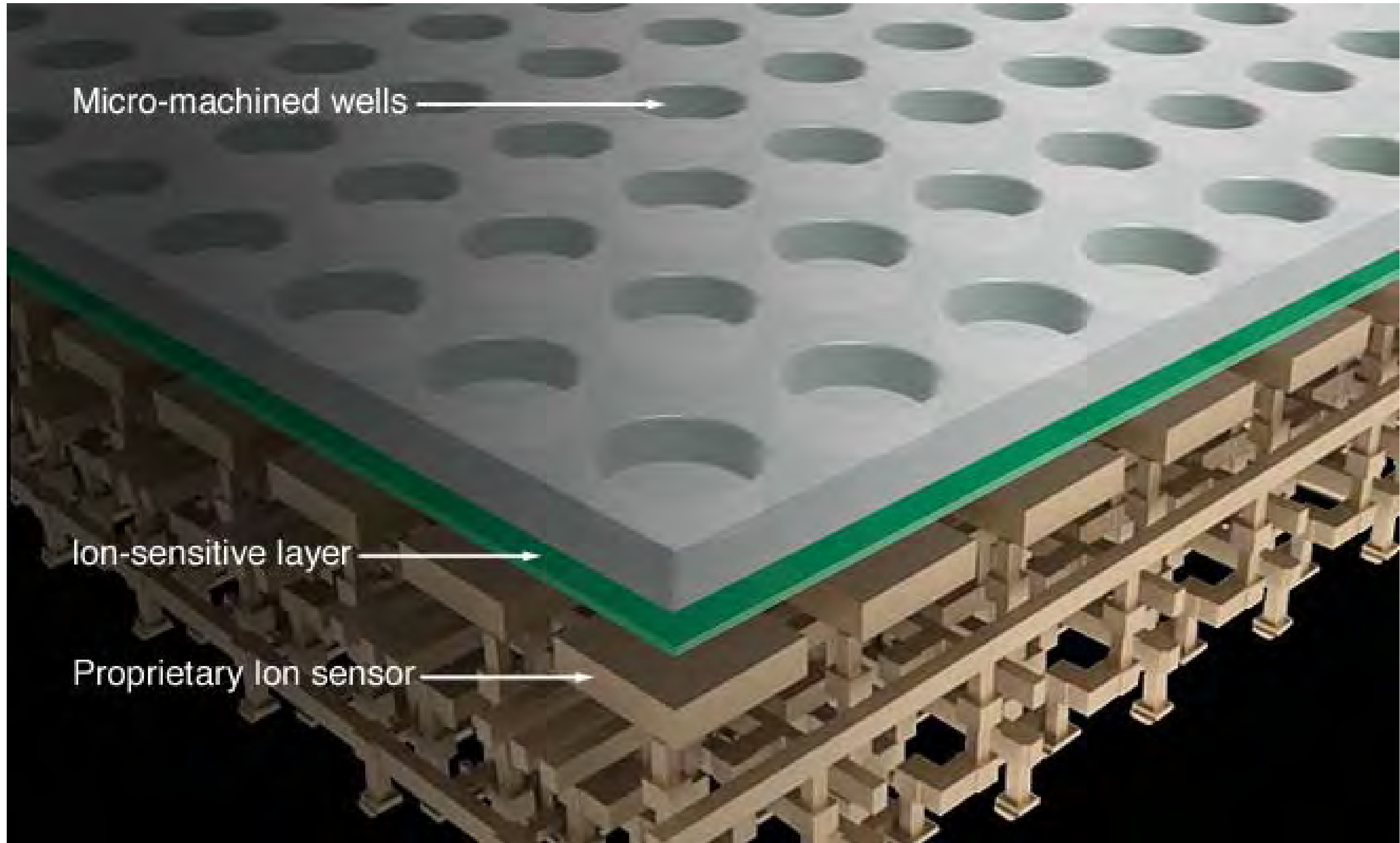
2 Human Exomes

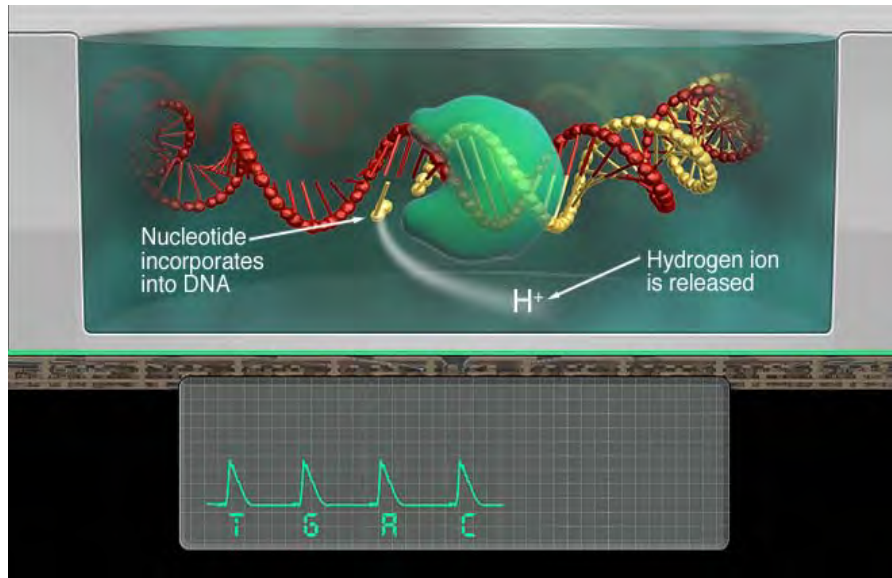
PROTON II chip

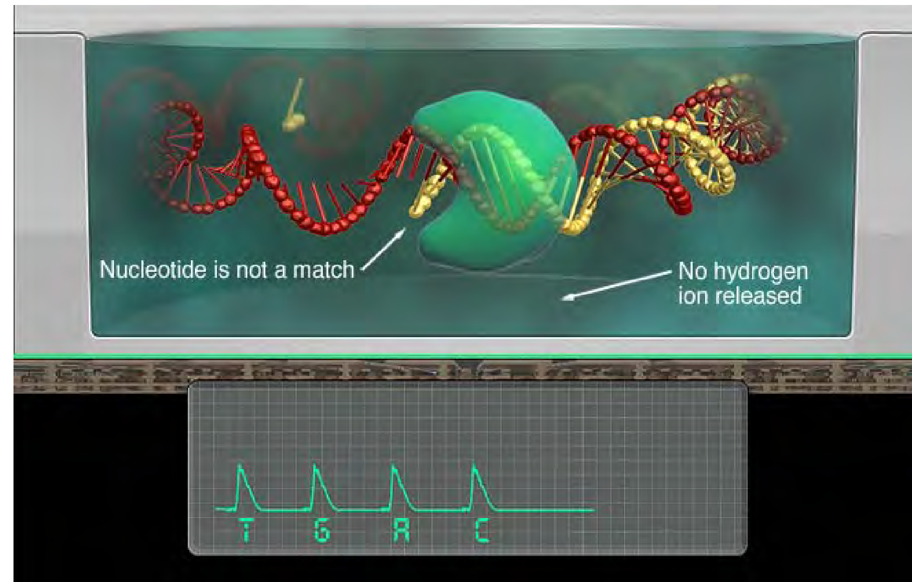
660 M wells

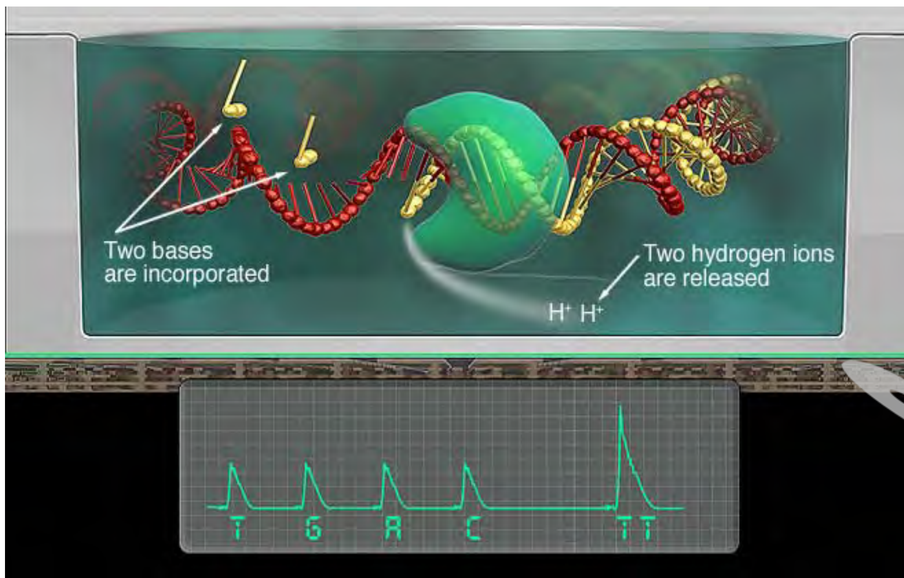
1 Human genome

Chip Technology









Sequencing

http://www.iontorrent.com/technology-ion-technology-video/

Reducing the complexity

moleculo

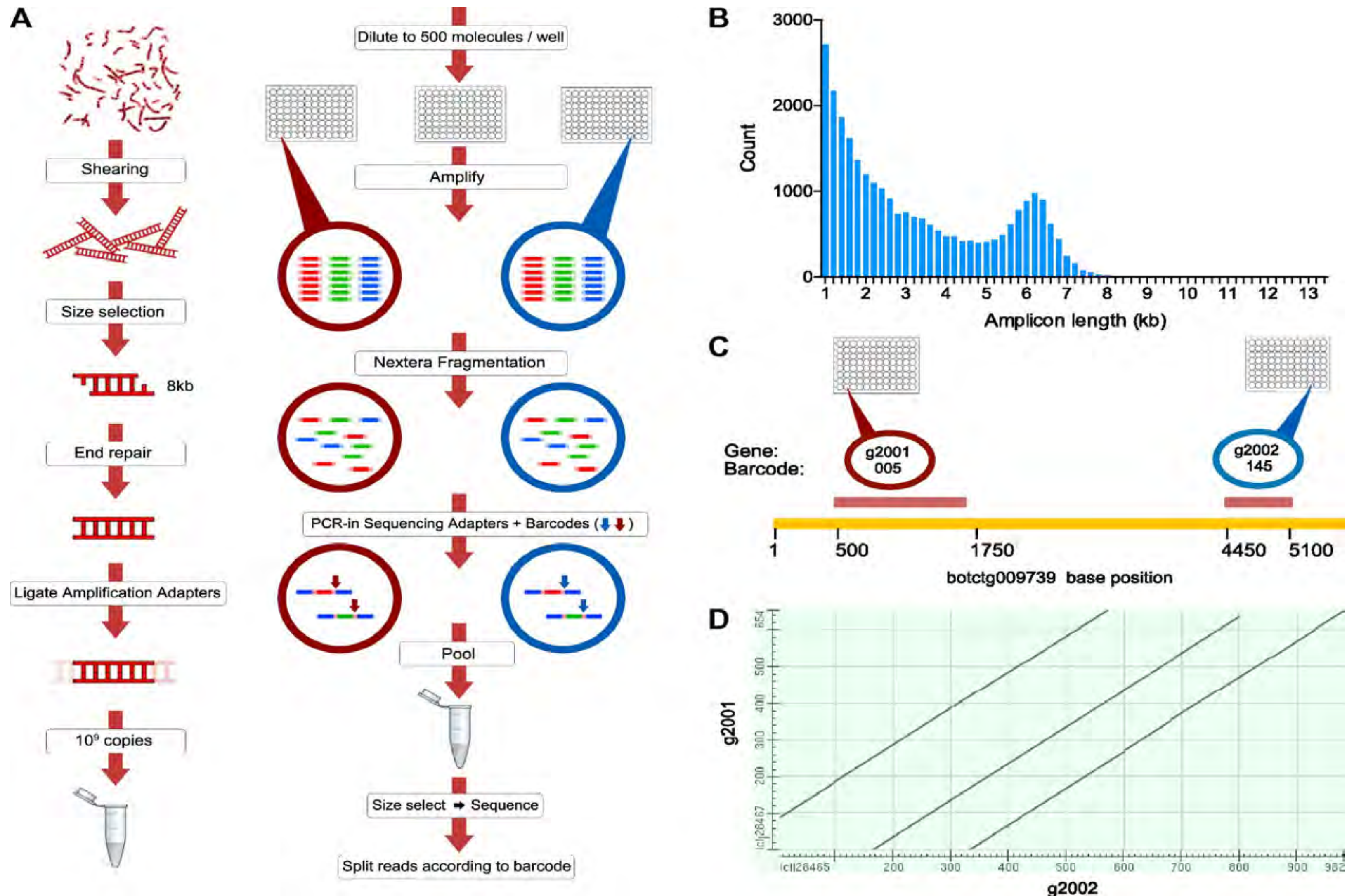
ILLUMINA

NEXT GENERATION SEQUENCING

LONG READS

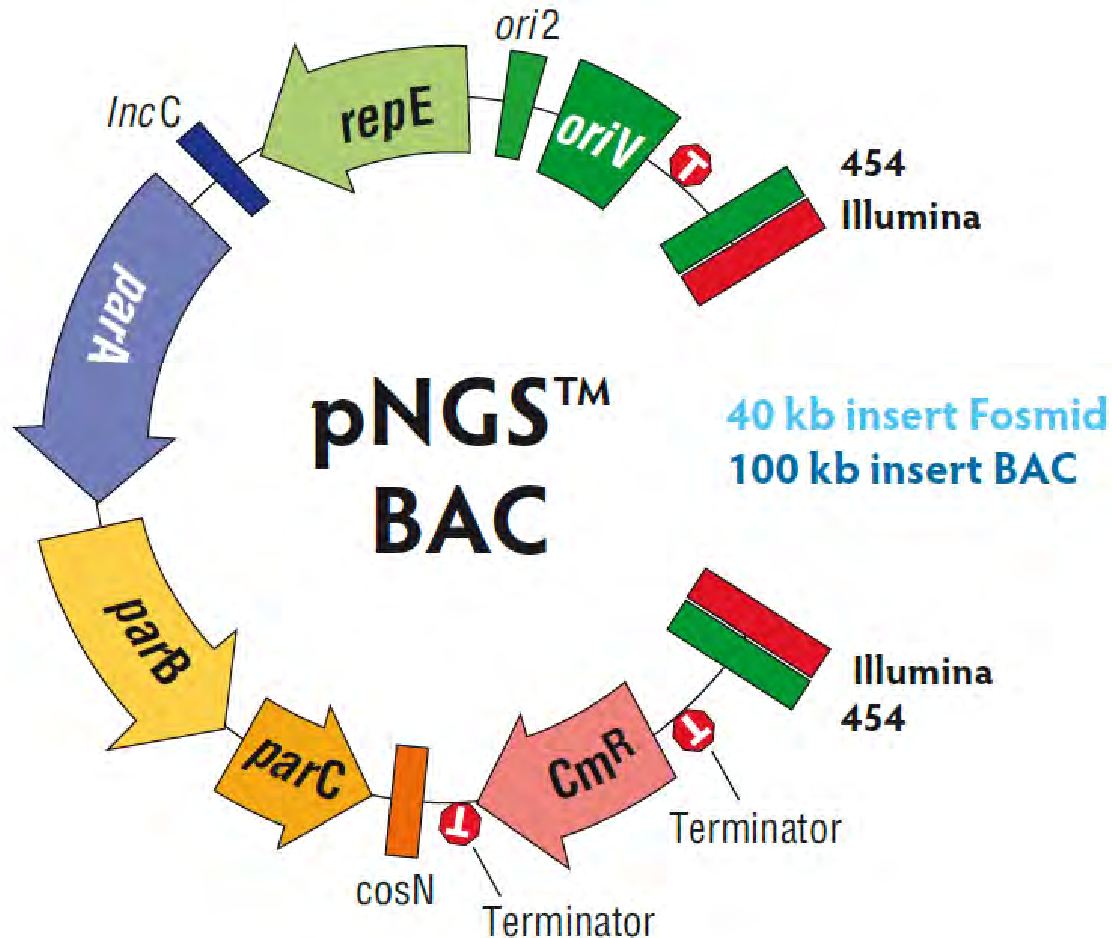
The genome sequence of the colonial chordate, *Botryllus schlosseri*

http://dx.doi.org/10.7554/eLife.00569

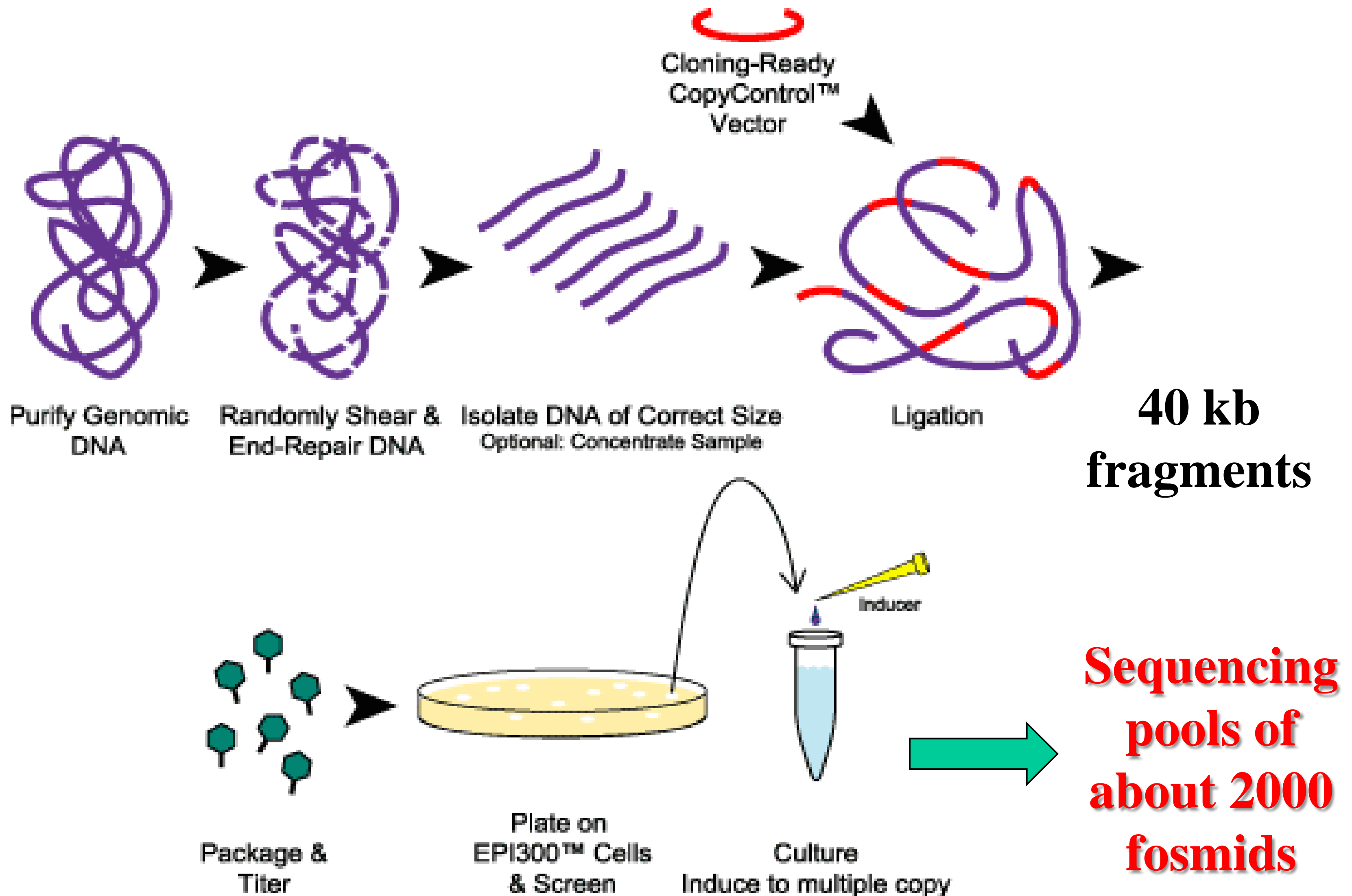


Lucigen® Corporation

SYNTHETIC FOS-BAC PLASMID



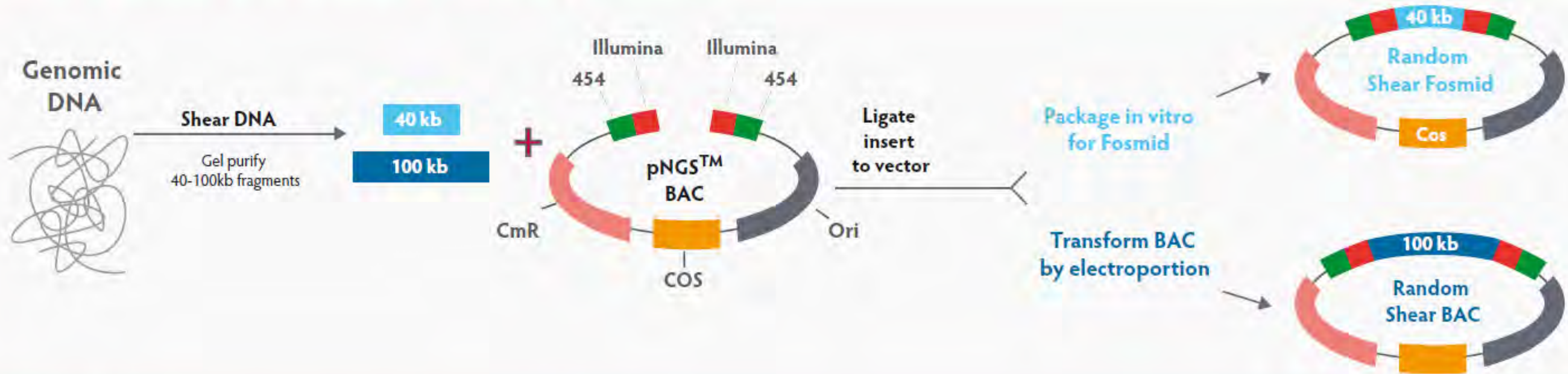
FOSMID Library Sequencing



40 kb fragments

Sequencing pools of about 2000 fosmids

STEP 1: Random Shear Fosmid or BAC Library Construction



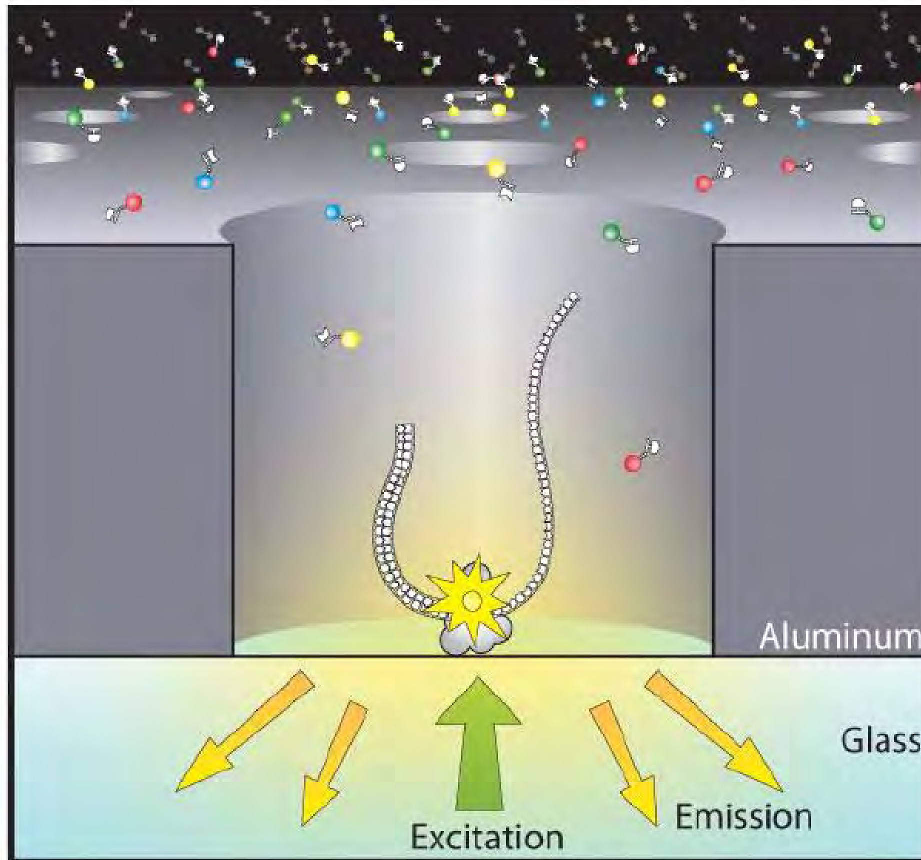
STEP 2: Generation of Terminal Sequence Tags

STEP 3: Illumina or 454 Sequencing

SEQUENCING FOSMID ENDs IN POOLS

Single molecule sequencing

Single Molecule Real Time Sequencing



SMRT

http://www.pacificbiosciences.com/video_lg.html

VIDEO PACBIO

PACBIO RS

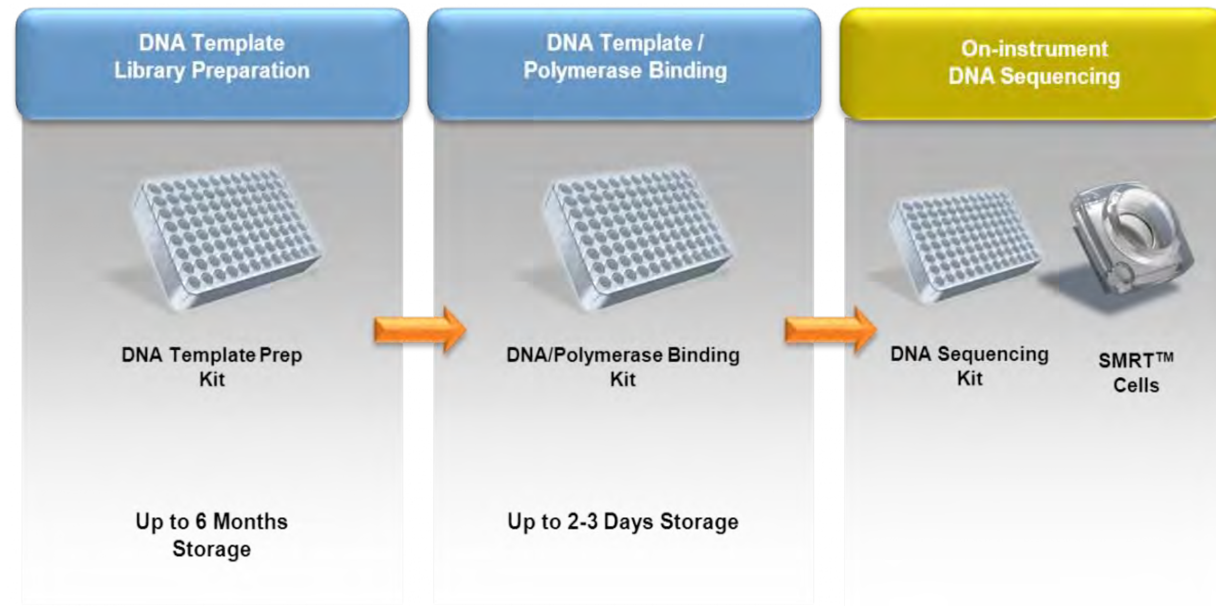
2 x 75.000 cell

35.000 reads/cell

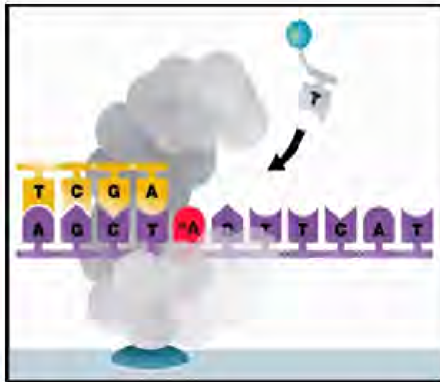
Up to 10.000 bp/read

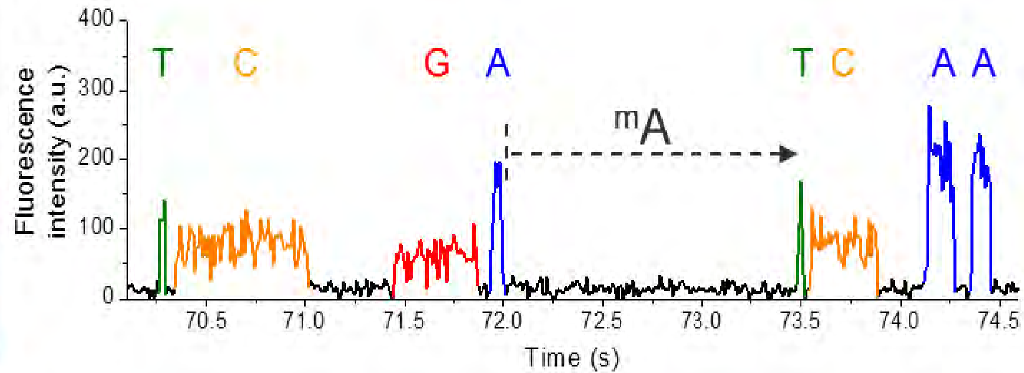
Minutes/run

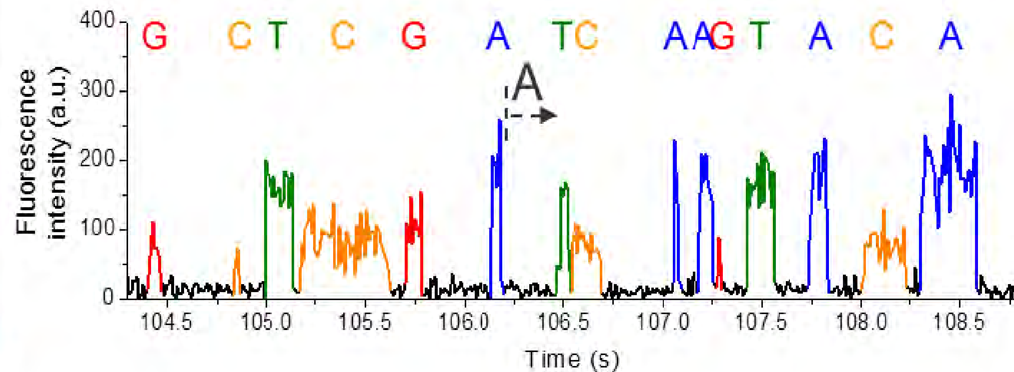
50 Mb/Cell



Detection of base modifications





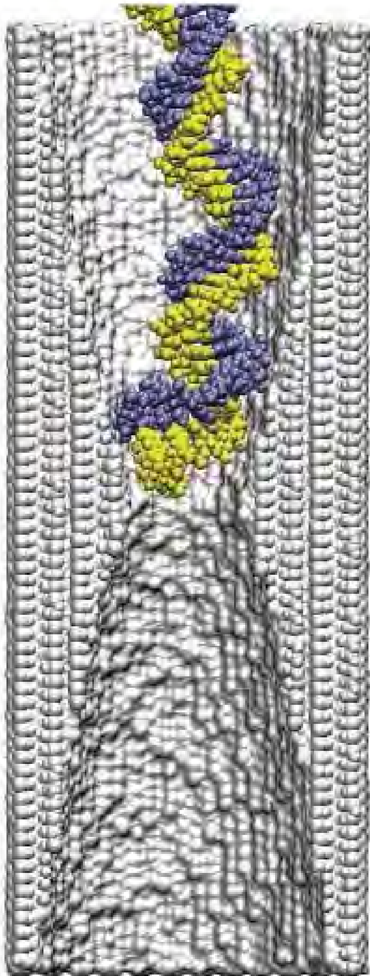
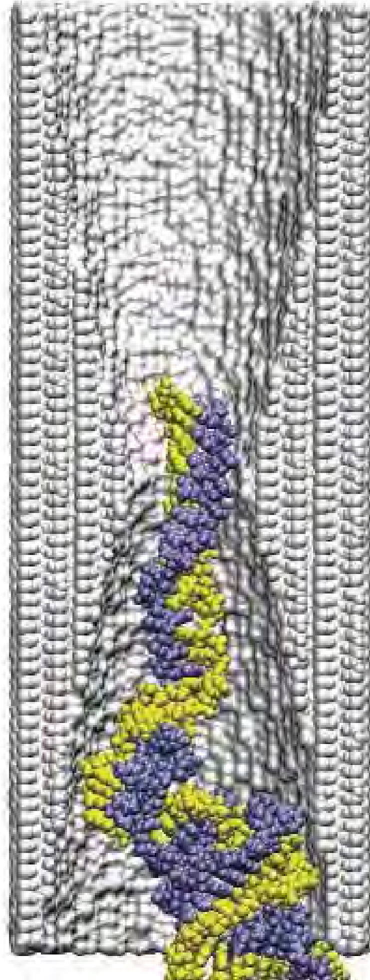
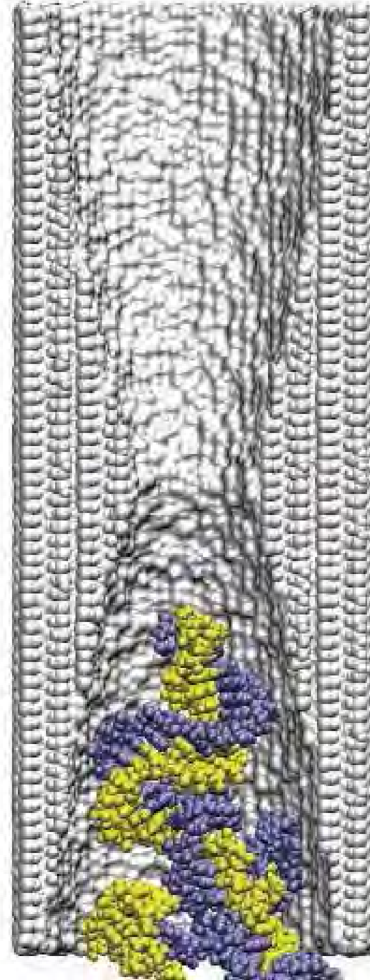


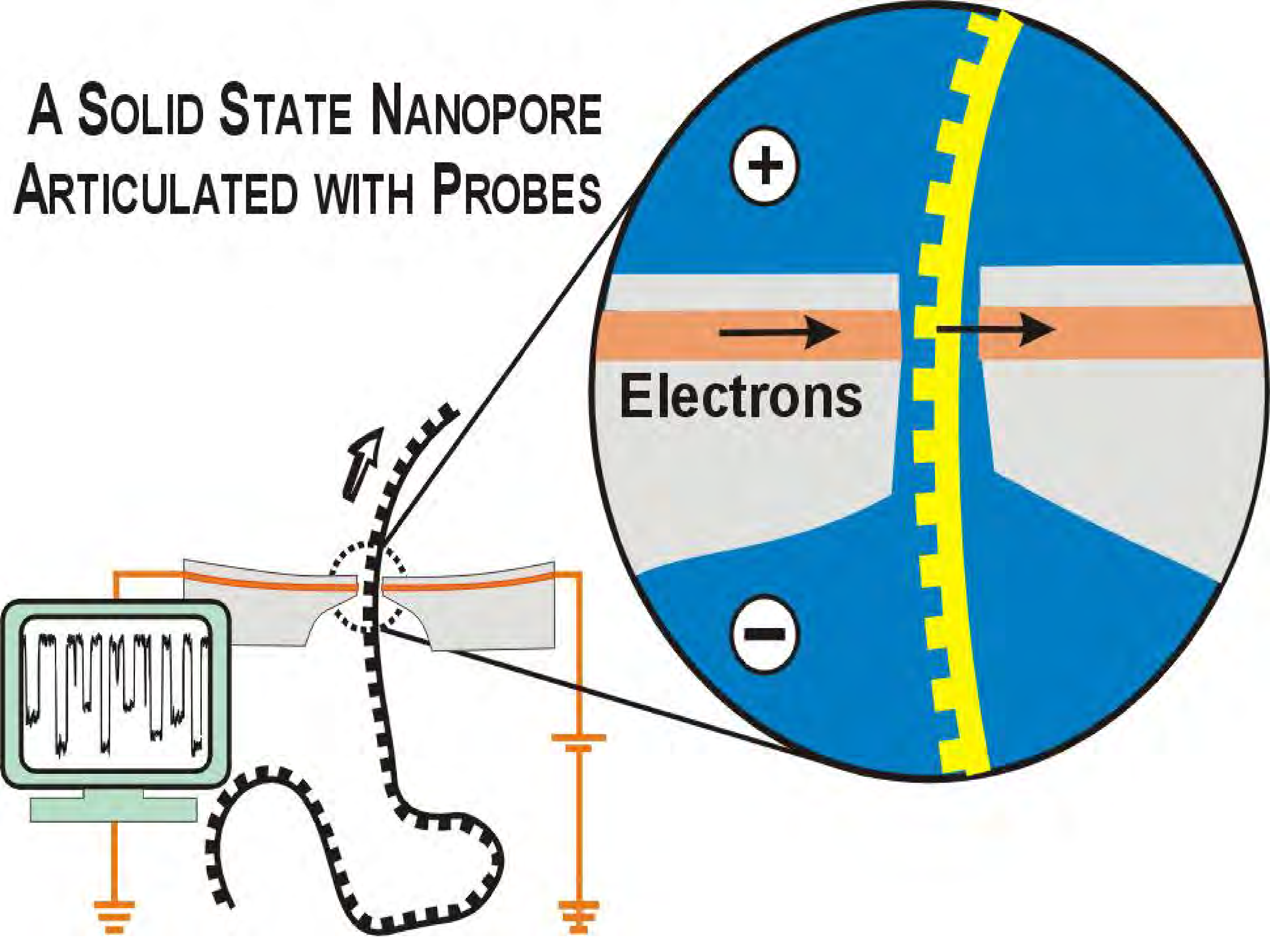
Epigenetics on real time

NANOPORES

Oxford NANOPORE Technologies

SEQUENOM®


A Solid State Nanopore
Articulated with Probes
+
Electrons
−

GridION

single molecule analysis system

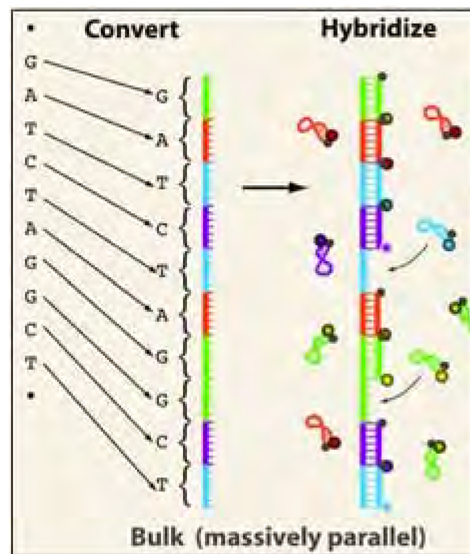
VIDEO SECUENCIA DNA NANOPORE

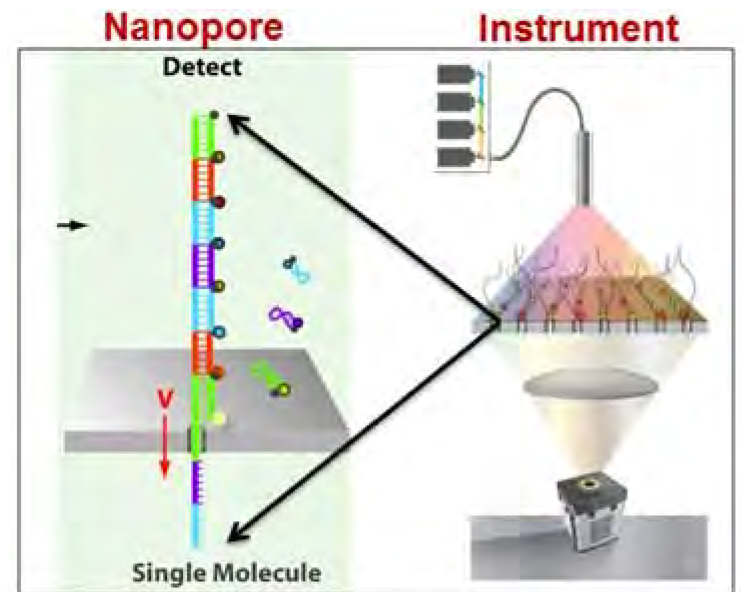
MINION

The future

Noblegen
BIOSCIENCES

OPTIPORE SEQUENCING





Hybridization → Ligation → Cleavage / melting / washing

GNNNNNN
TNNNNNN
ANNNNNN
CNNNNNN

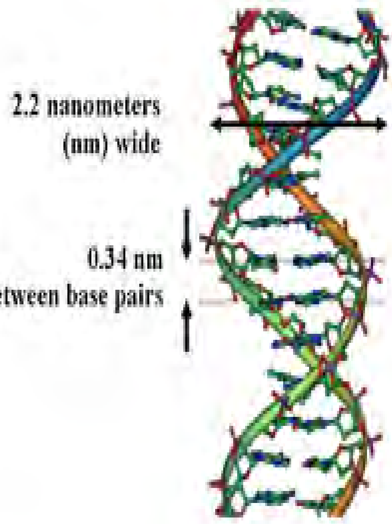
Probe library

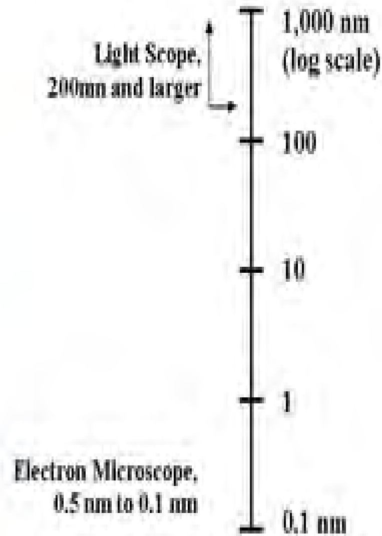
\+

T G A C

R T R G R A R C

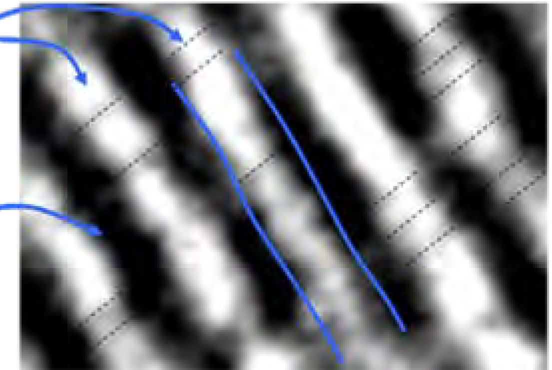
ZS Genetics Inc



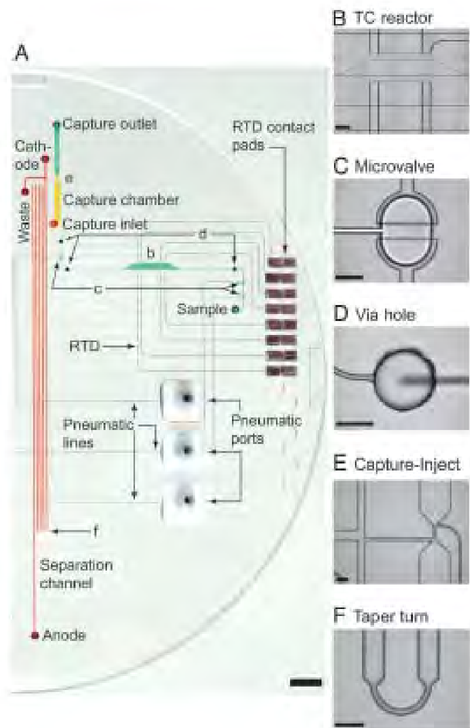


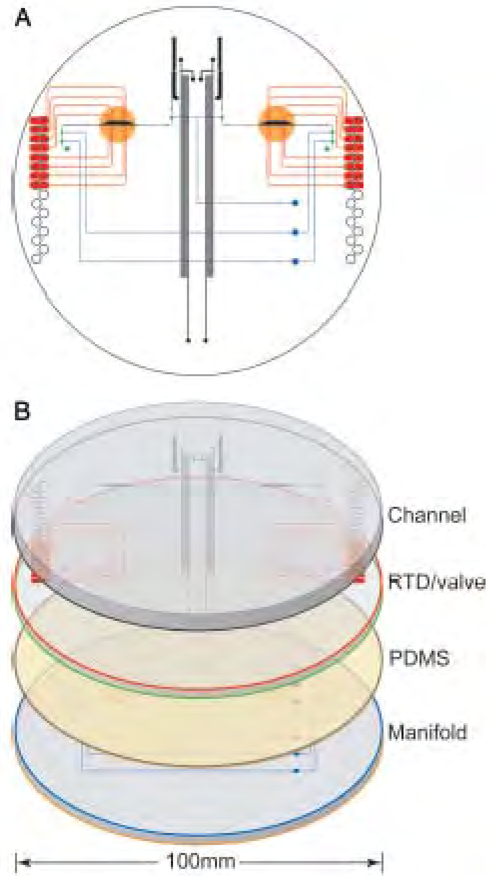
Fact

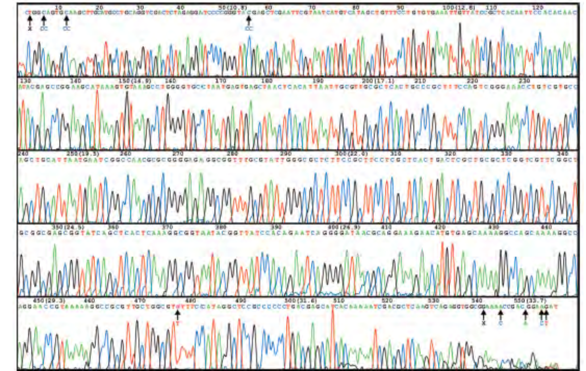
Labeling atoms cast dark "shadows".
Ladders form a monolayer of parallel molecules.
Some unlabeled nitrogenous bases can be seen.



Single Atom Labels Enable Electron Microscopes To Directly See DNA







lab-on-a-chip for DNA sequencing

Richard A. Mathies PNAS (2006)103, 7240

Many Thanks