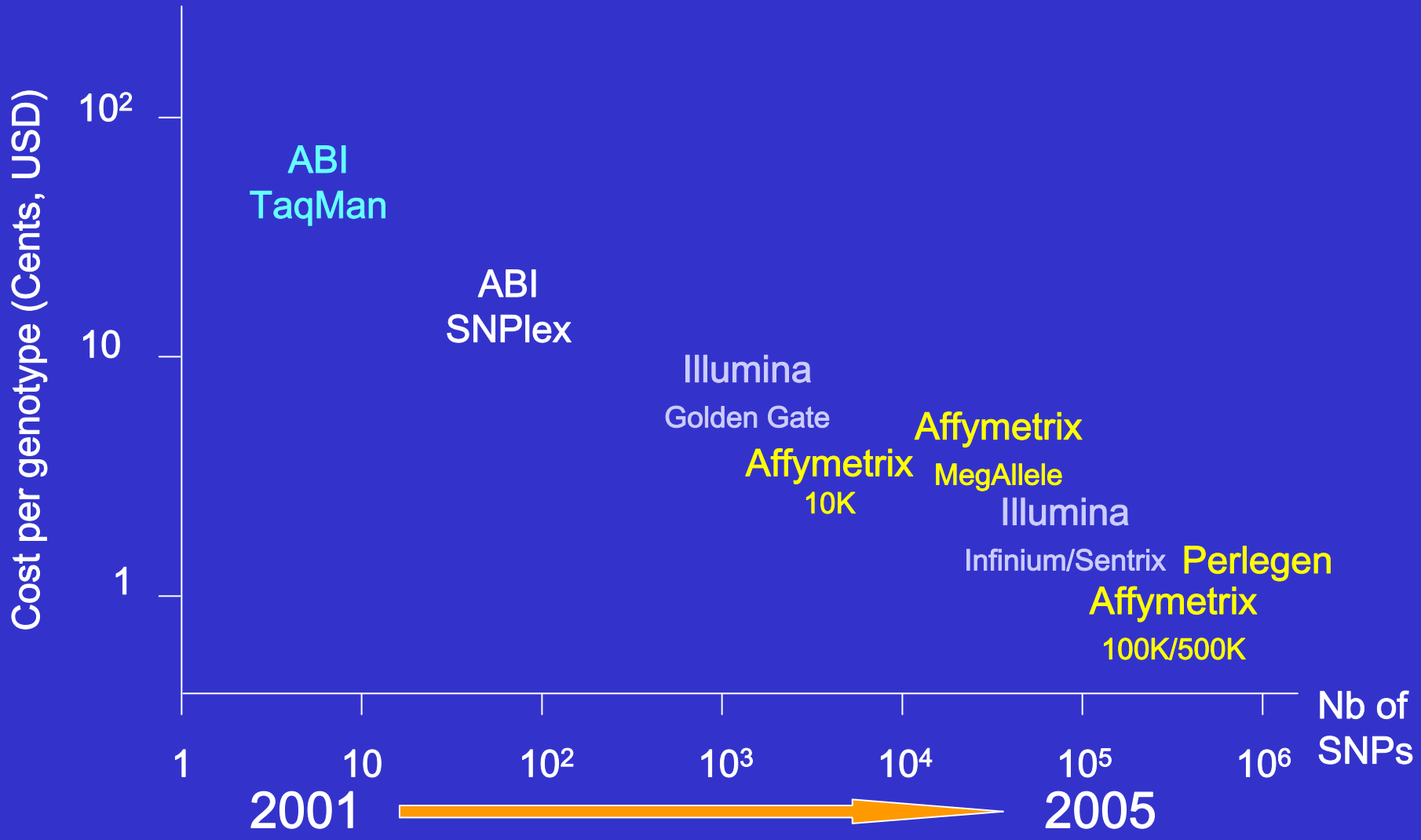
# Next Generations of Sequencing Technologies

**Jeffery A. Schloss, Ph.D.**
**Program Director, Technology Development Coordination**
**National Human Genome Research Institute**
**National Institutes of Health**
**U.S.A.**

**House of Lords Inquiry on Genomic Medicine**
**Visit to NHGRI**
**June 4, 2008**

# Progress in Genotyping Technology




Courtesy S. Chanock, NCI

# Genotyping vs. Sequencing Costs

…feasible to assay 375,000 or more single nucleotide polymorphisms (SNPs), capturing 80% or more of the HapMap-defined genomic variation, in roughly 2,000 subjects for roughly $1.7 million per study.

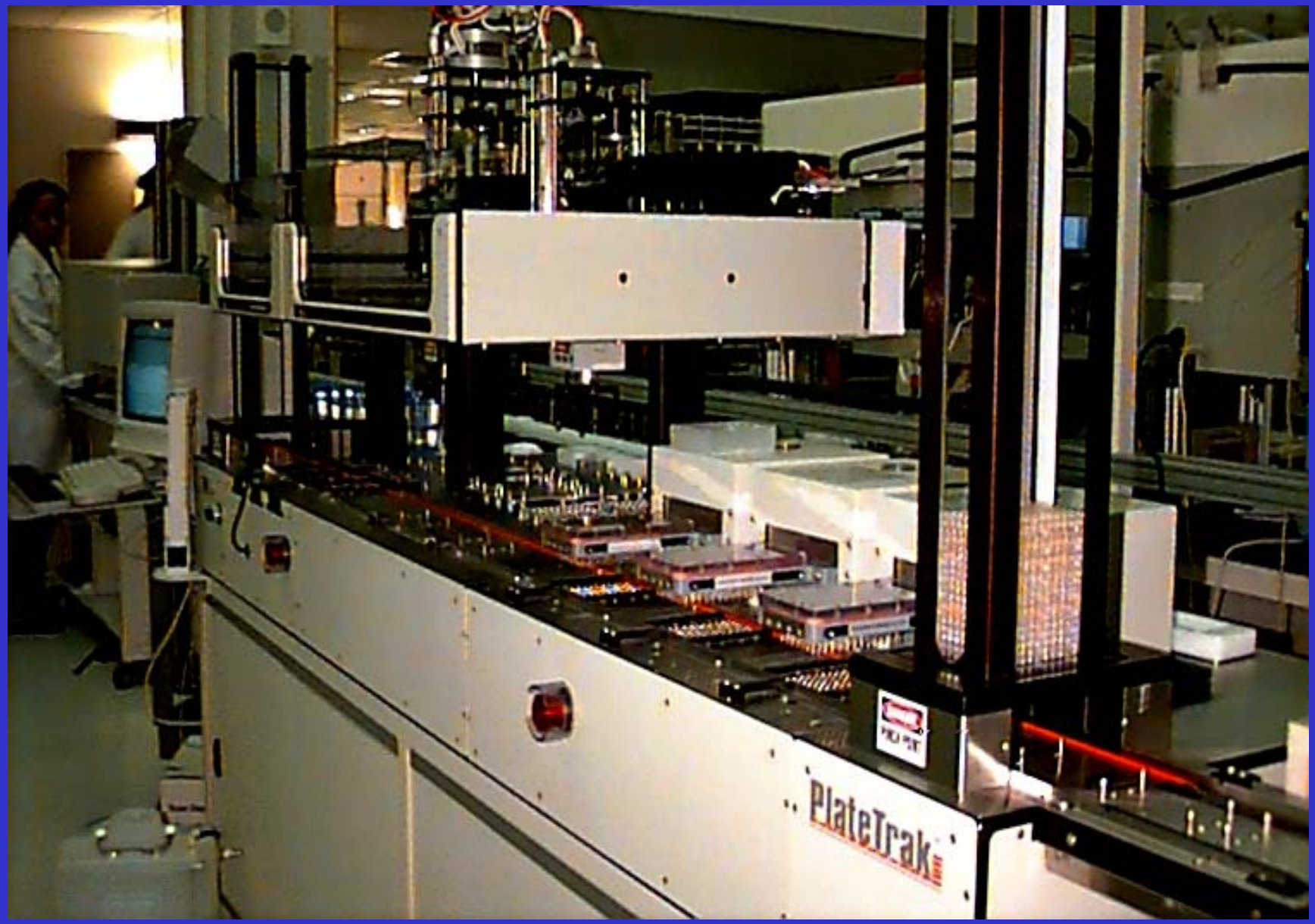
cost to genotype a human genome =$850

# The Applied Biosystems 3730*xl*™ DNA Analyzer

## ***Fully Integrated System***

- ✓ 96-capillaries
- ✓ Simultaneous injection and analysis of 96 samples
- ✓ Automated plate loading from a stacker that accommodates up to 16 plates (96 or 384 well)
- ✓ Internal barcode reader
- ✓ Bench top unit

# Human Genome Project Sequencing Centers




Slide credit: Eric Green, NHGRI

# Human Genome Project Sequencing Centers

Slide credit: Eric Green, NHGRI

# Coordinating Committee Process


NIH NEWS RELEASE

National Institutes of Health — National Human Genome Research Institute

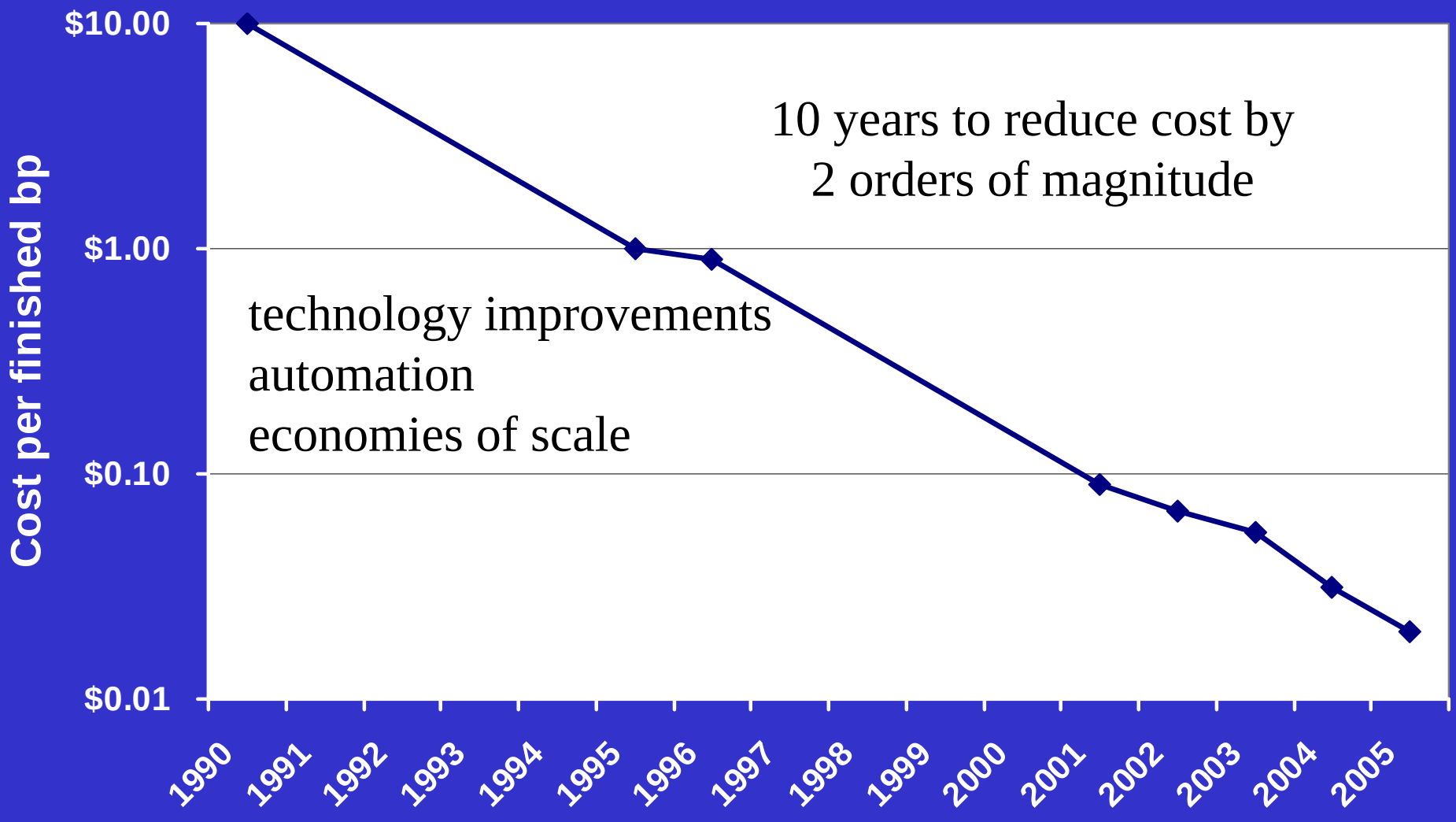
news release banner

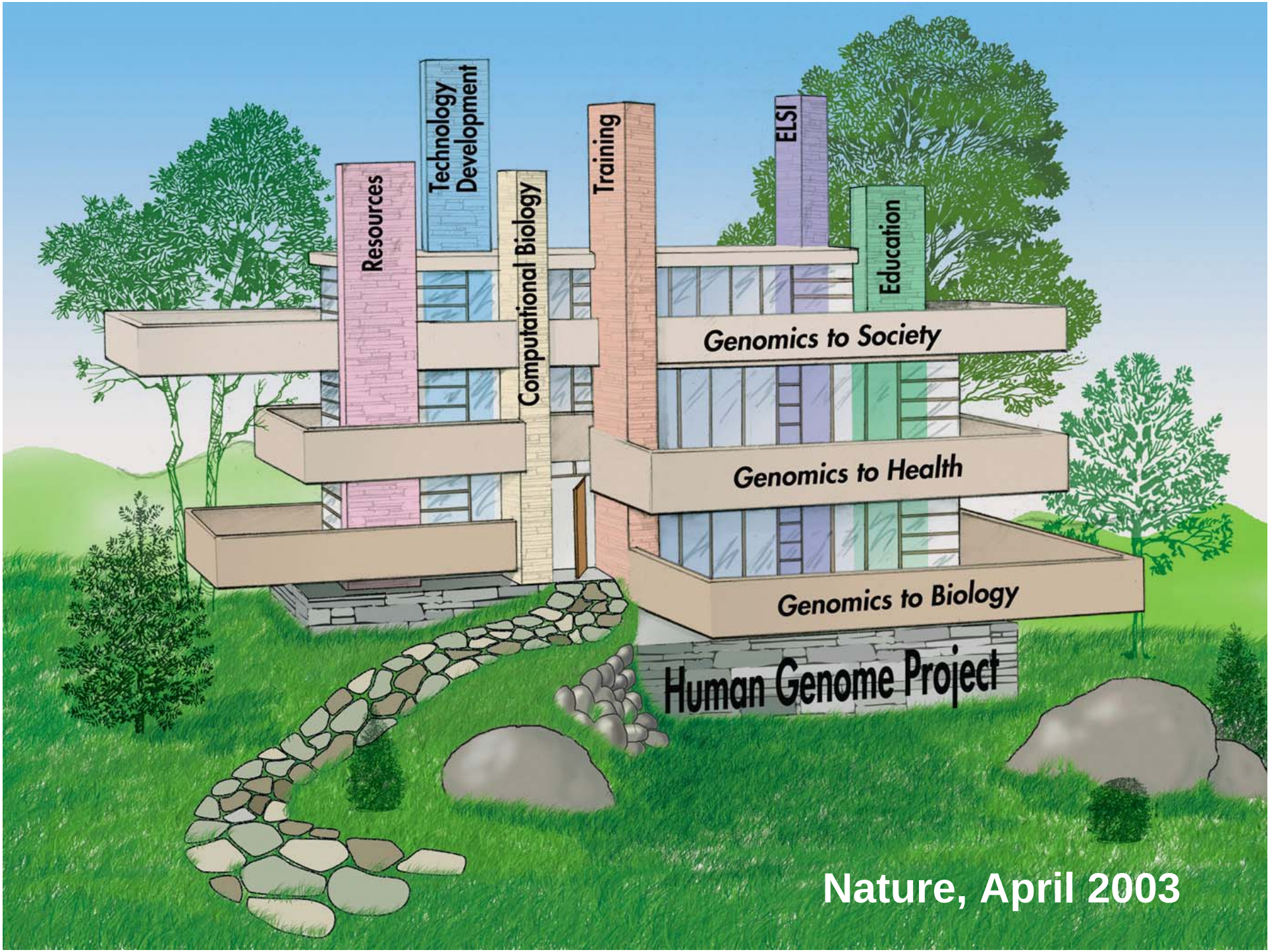
NHGRI Announces New Sequencing Targets

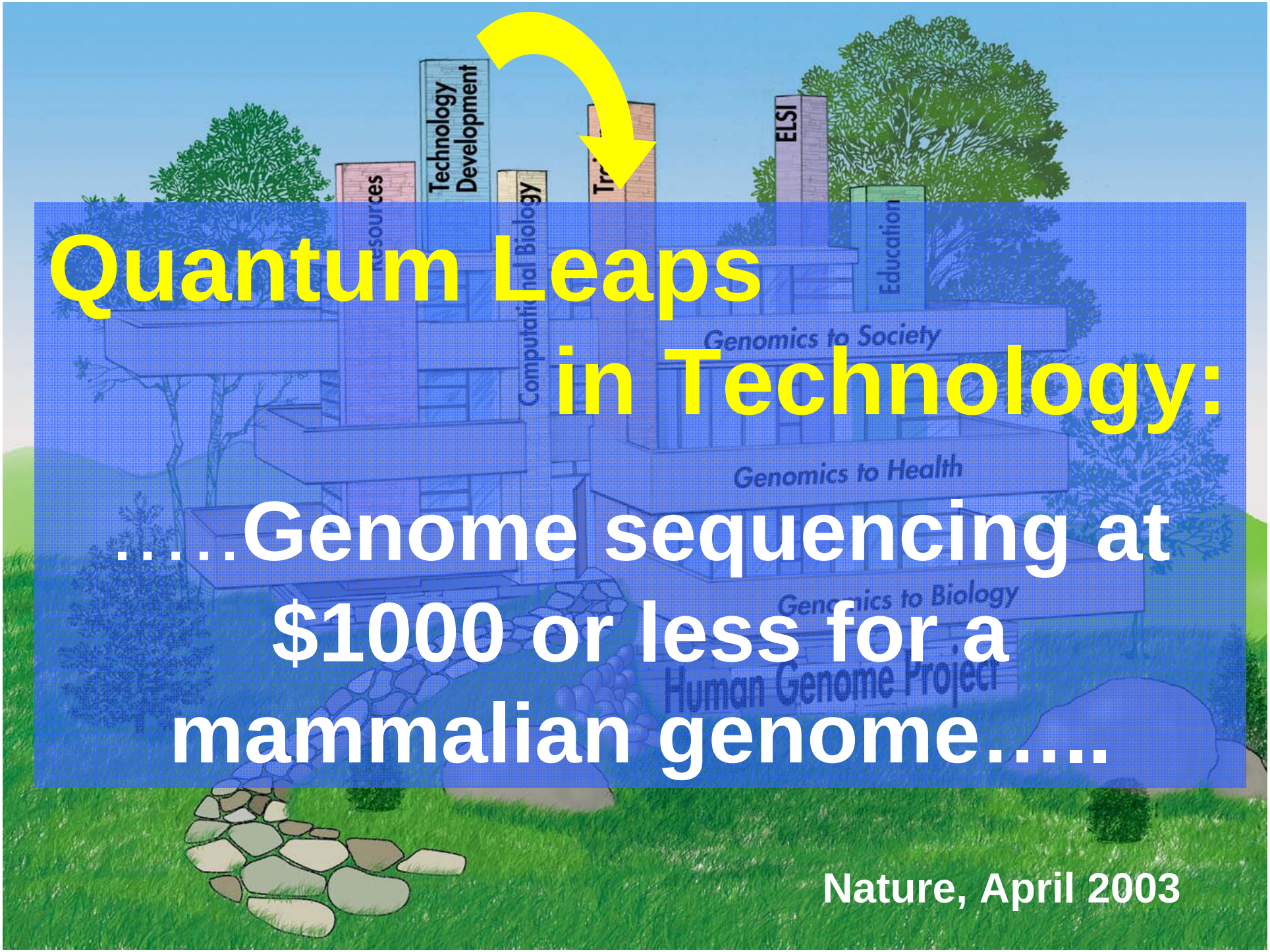
March 15, 2006

- structural variation in 48 HapMap samples (fosmid end sequencing). The genomes of any two humans are thought to differ by several hundred insertions, deletions and inversions.
- add DNA sequence to existing draft sequences of a number of primate species and add additional sequence information in regions of high biological interest for rhesus macacque, marmoset and orangutan
- low-density draft (2-fold coverage) for 8 mammals

# Decrease in the Cost of Finished DNA Sequencing




Resources
Technology Development
Computational Biology
Training
ELSI
Education
Genomics to Society
Genomics to Health
Genomics to Biology
Human Genome Project
Nature, April 2003




# Quantum Leaps in Technology:

## ….Genome sequencing at $1000 or less for a mammalian genome…..

Nature, April 2003

# NHGRI DNA Sequencing Technology Development Requests for Applications

- current technologies are able to produce the sequence of a mammalian-sized genome of the desired data quality (high-quality draft) for $10 to $50 million; the goal of this initiative is to reduce costs by at least two/four orders of magnitude.
- RFA goal is sequencing technology that produces assembled sequence at high accuracy ($10^{-4}$-$10^{-5}$ error rate), *de novo*. Ultimate goal is higher accuracy.
- applications that propose technology development for re-sequencing should explain how they will achieve the projected reduction in cost compared to technologies that can produce data of similar quality today

# NHGRI DNA Sequencing Technology Development Requests for Applications

- current technologies are able to produce the sequence of a mammalian-sized genome of the desired data quality (high-quality draft) for $10 to $50 million; the goal of this initiative is to reduce costs by at least two/four orders of magnitude.
- initial awards were made in 2004.
- goal: technologies for 100x reduction in cost by ~2009
- goal: technologies for 10,000x reduction in cost by ~2014

http://www.genome.gov/10000368#6 – “Advanced Sequencing Technology Awards”

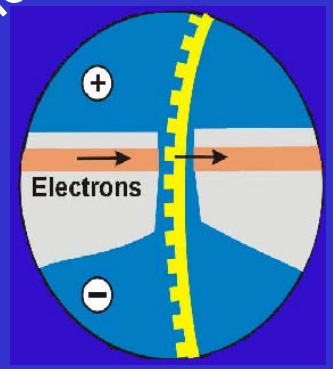
# NHGRI DNA Sequencing Technology Development
# $100,000 and $1,000 genome

Investment

| | |
|---|---|
| Round 1 | $39 M |
| Round 2 | $31 M |
| Round 3 | $13 M |
| Round 4 | $16 M |
| TOTAL | $99 M |

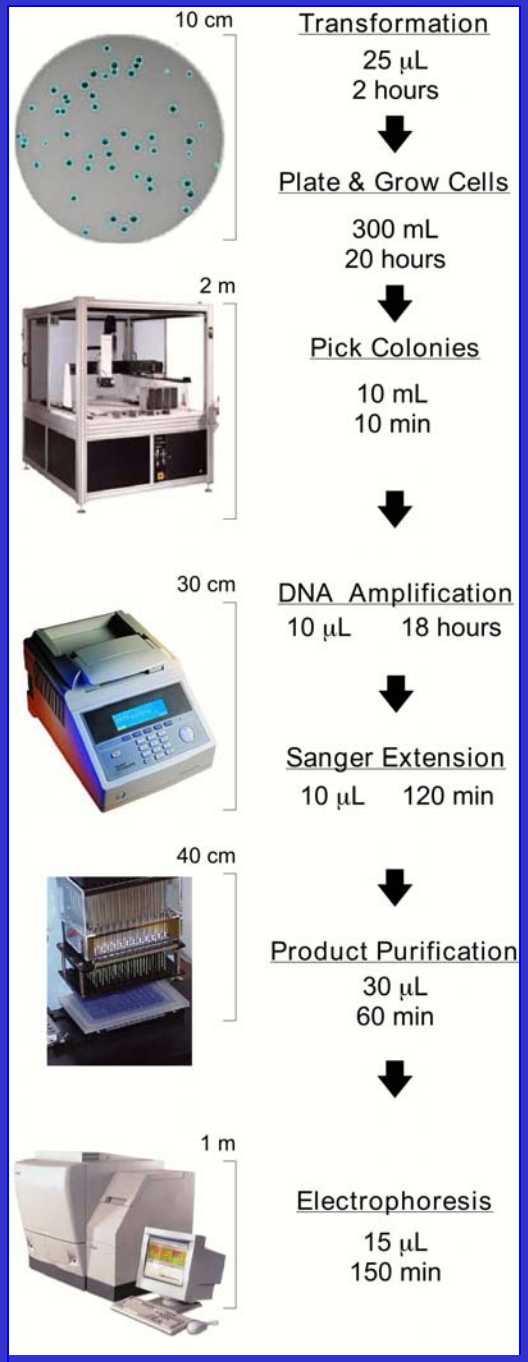
# NHGRI DNA Sequencing Technology Development
# $100,000 and $1,000 genome portfolio

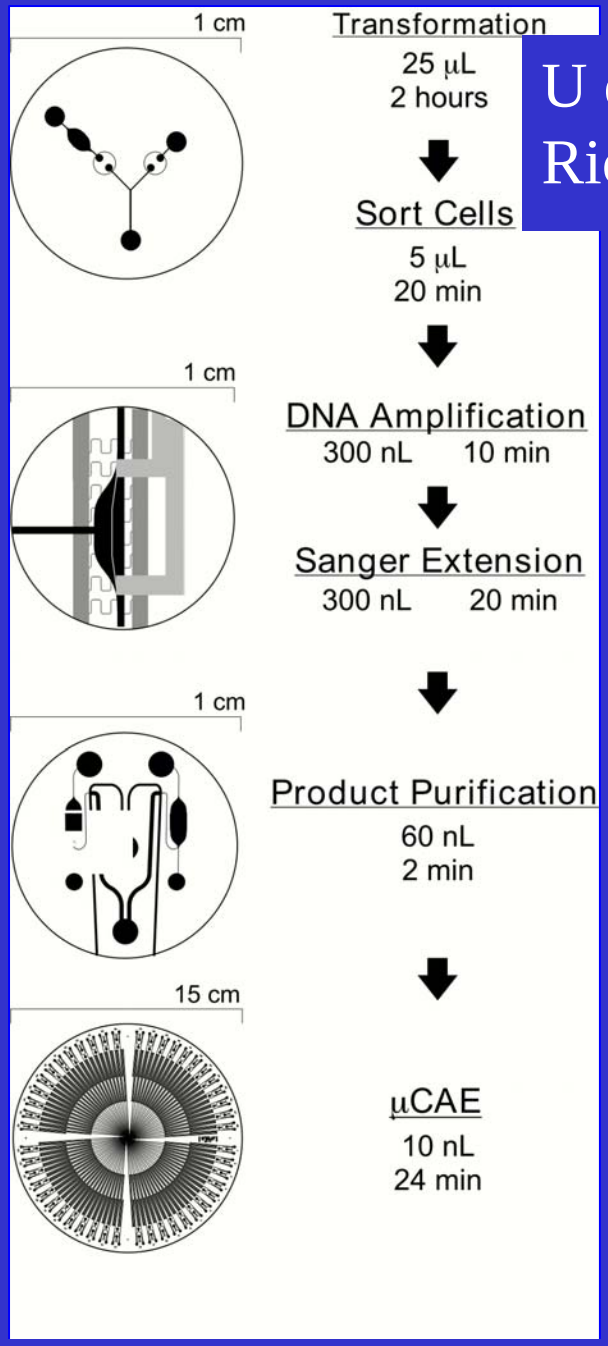
Sanger, microchannel integrated system (3)
Sanger, atomic force microscope separation (1)
Sequencing by cyclic synthesis (10)
Single molecule sequencing by synthesis (4)
Sequencing by ligation (4)
Sequencing-by-hybridization (4)
Other near-field sensing unmodified DNA (2)
Free-running polymerase/other enzyme (3)
Nanopore with modified analyte (4)
Biological nanopore (3)
Synthetic nanopore/gap (8)




Toward the $1,000 genome

# Sanger chemistry, miniaturized and integrated



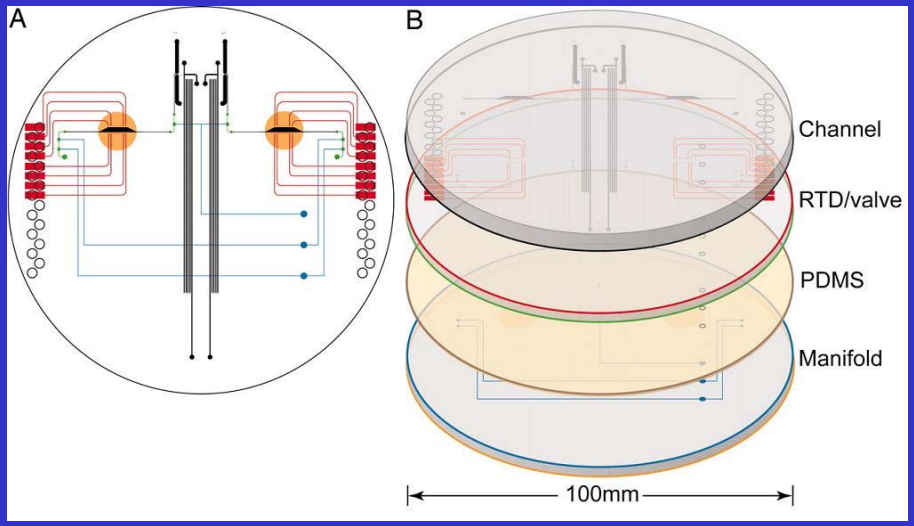




U of California, Berkeley
Richard A. Mathies, Ph.D.

"A microfabricated device for high-throughput DNA sequencing that couples clone isolation, template amplification, Sanger extension, purification, and electrophoretic analysis in a single microfluidic circuit is now attainable.".

B. Paegal, R. Blazej & R. Mathies, 2003, Curr. Opin. Biotechnol. *14*, 42

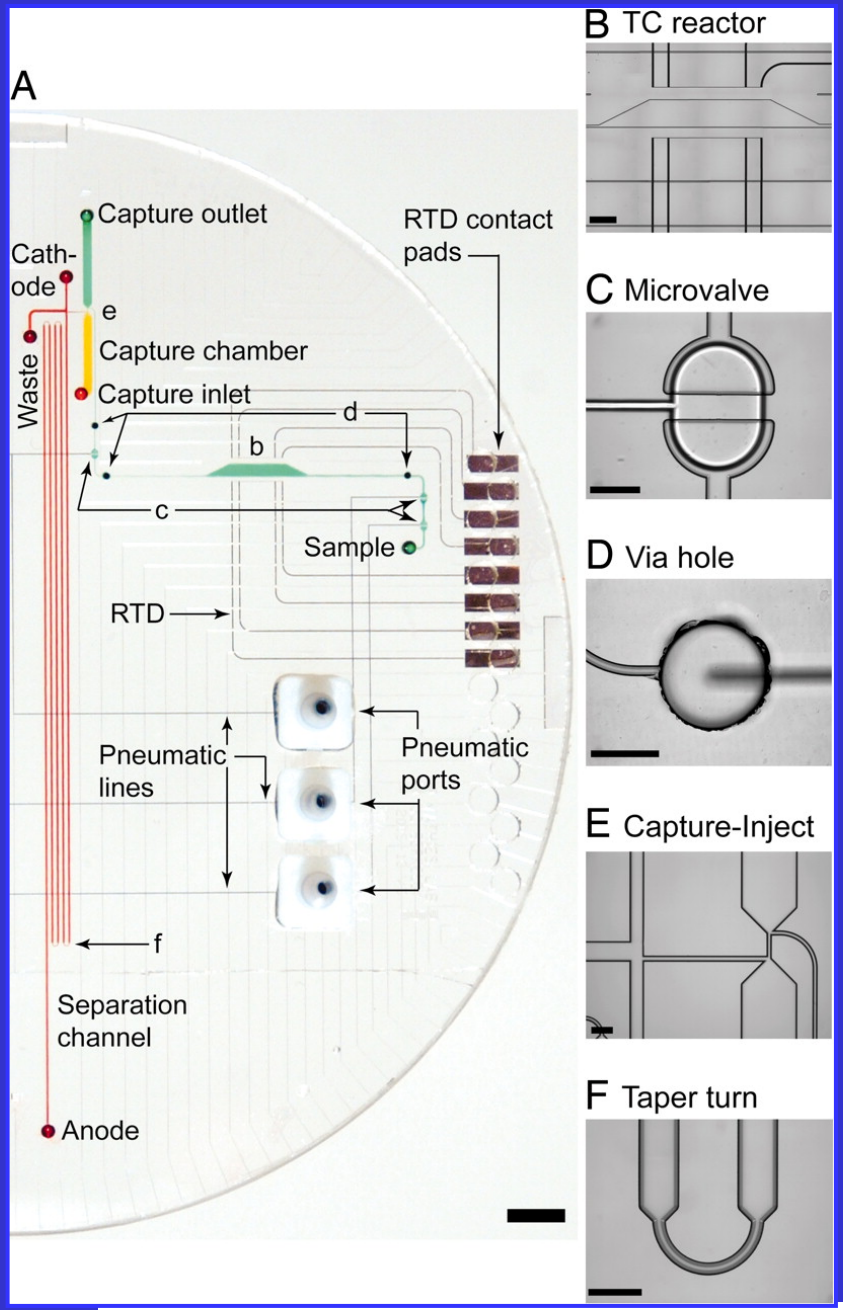
# Microfabricated bioprocessor for integrated nanoliter-scale Sanger DNA sequencing





Microchip Biotechnologies, Inc.
Stevan Jovanovich, Ph.D.

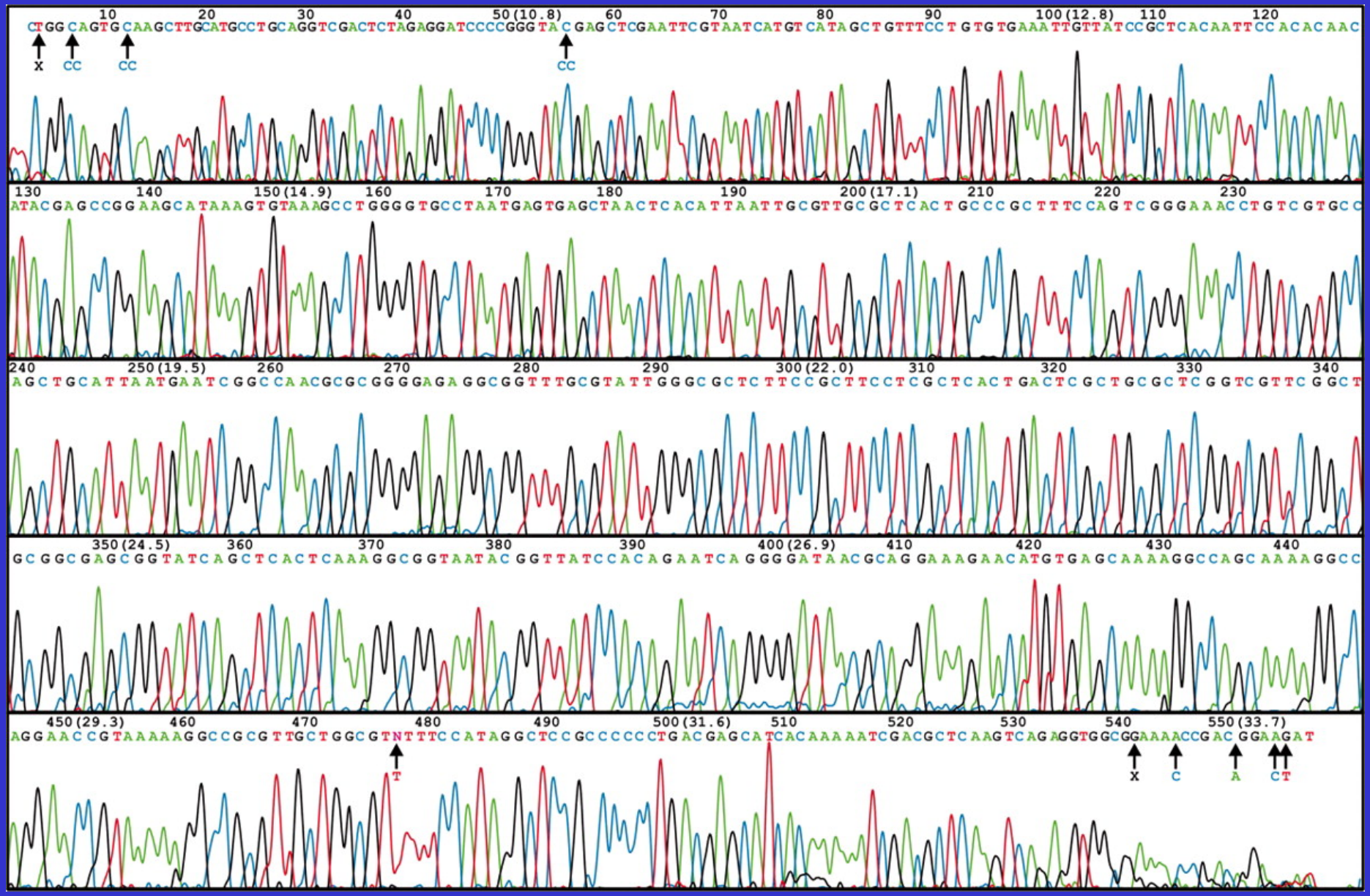
Richard A. Mathies, Ph.D.
U of Calif., Berkeley
Annelise Barron, Ph.D.
Northwestern University



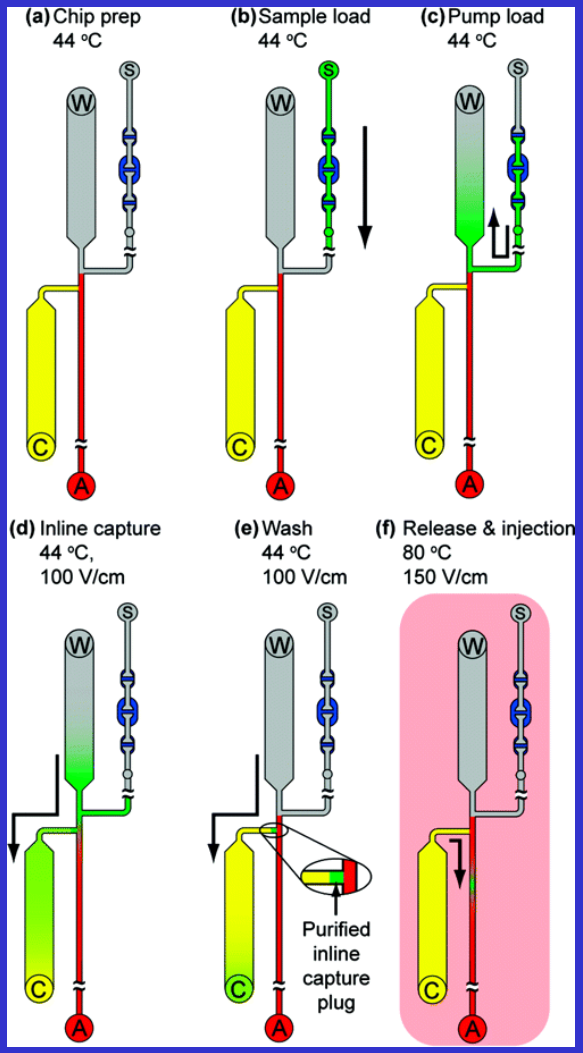
RG Blazej, P Kumaresan & RA Mathies, 2006, PNAS USA *103*:7240


NATIONAL HUMAN GENOME RESEARCH INSTITUTE

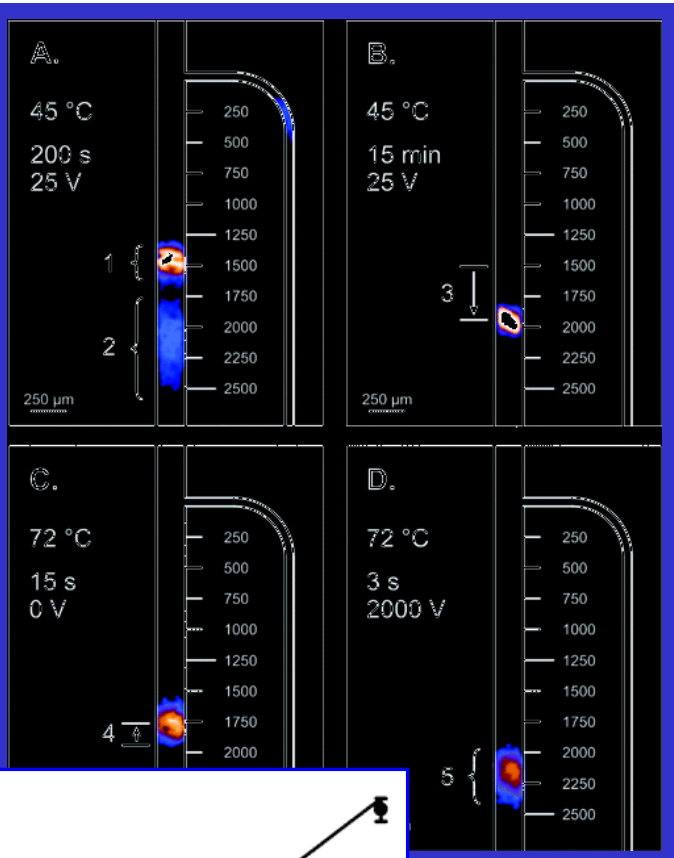


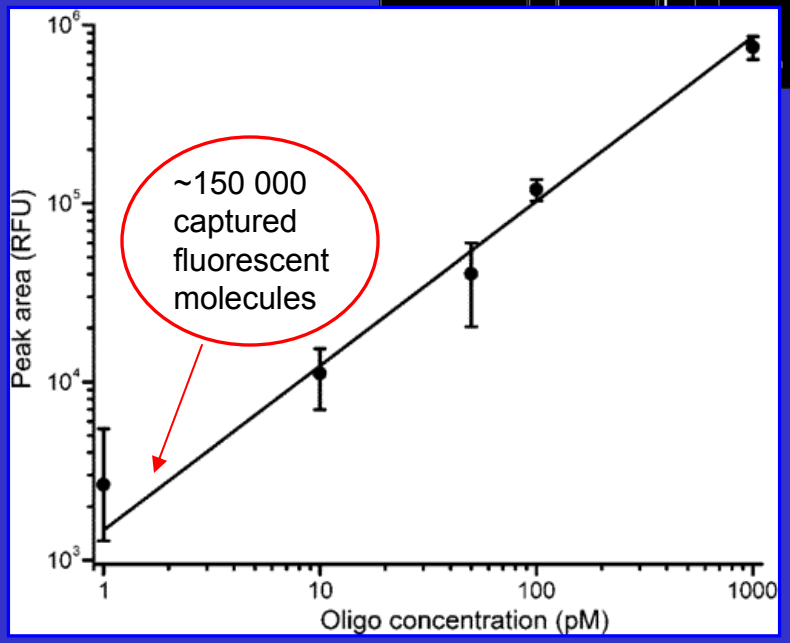
RG Blazej, P Kumaresan & RA Mathies, 2006, PNAS USA *103*:7240

# Integrated affinity capture, purification & inline injection capillary electrophoresis for attomole-scale Sanger DNA sequencing



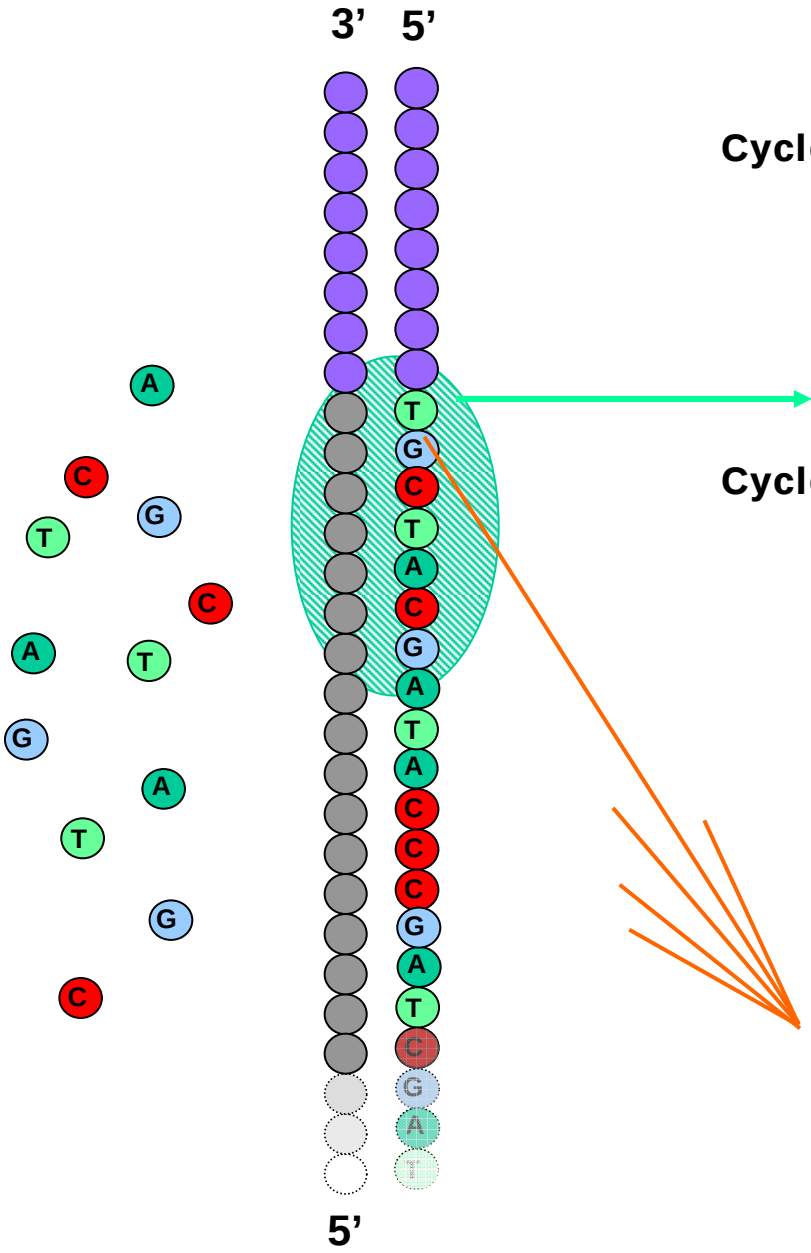

Mathies lab:
Blazej, et al., 2007, *Anal. Chem.*, 79 (12), 4499 -4506
Toriello,_ et al., 2007, *Anal. Chem.*, ASAP Article 10.1021/ac0712547 S0003-2700(07)01254-1



# Sequencing by cyclic synthesis

# Sequencing by cyclic synthesis (SBS)




**Cycle 1: Add sequencing reagents**

**First base incorporated**

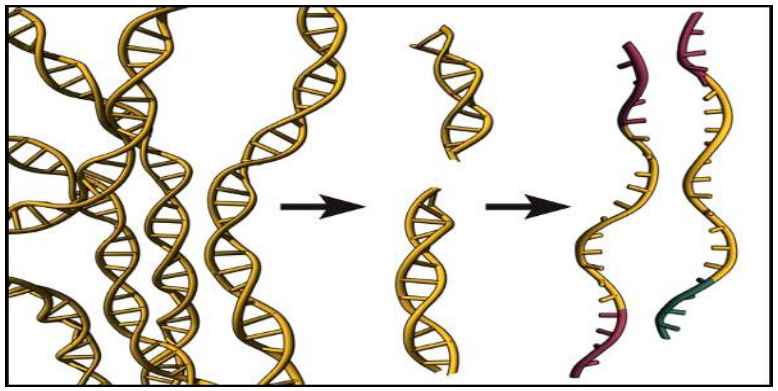
**Remove unincorporated bases**

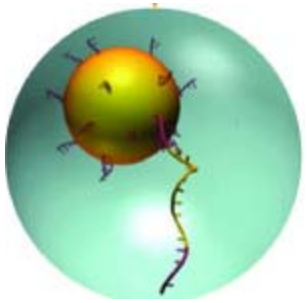
**Detect signal**

**Cycle 2-n: Add sequencing reagents and repeat**

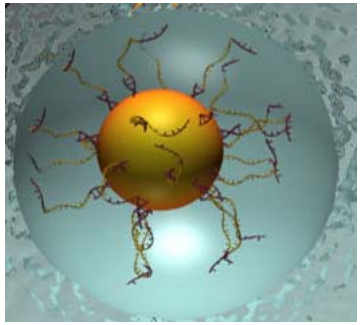
**Do this on a microarray with hundreds of thousands or millions of different molecules on a surface!**

# Process Overview

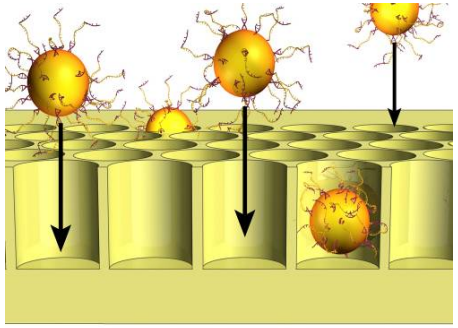




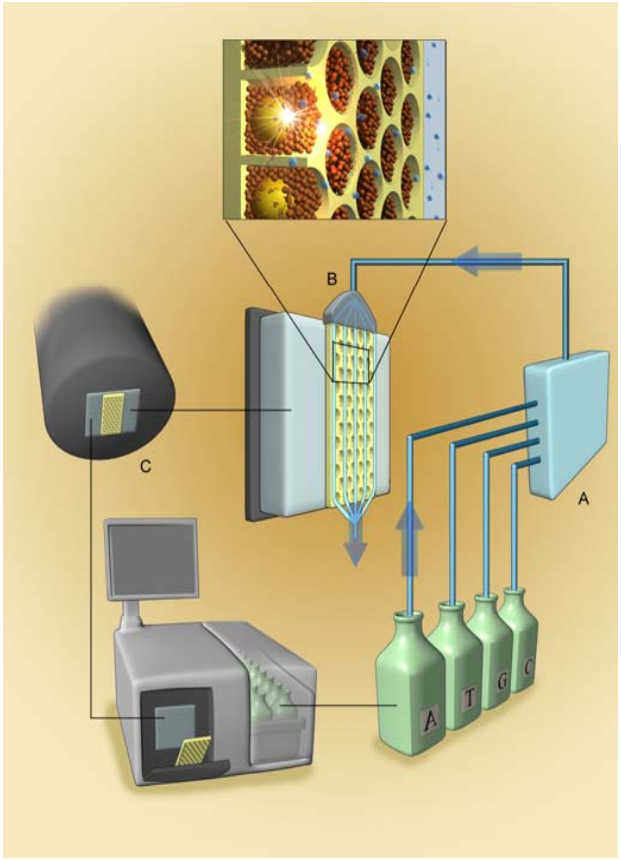
1) Prepare adapter ligated ssDNA library
2) Capture DNA fragments on excess of capture beads



3) Clonal Amplification on 28 µ beads

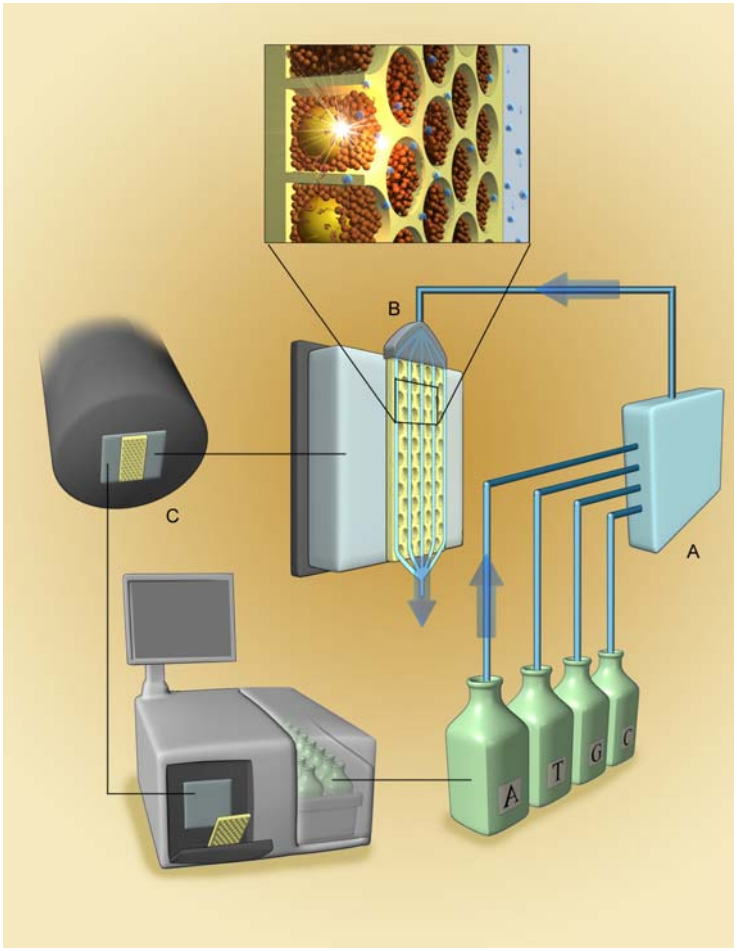


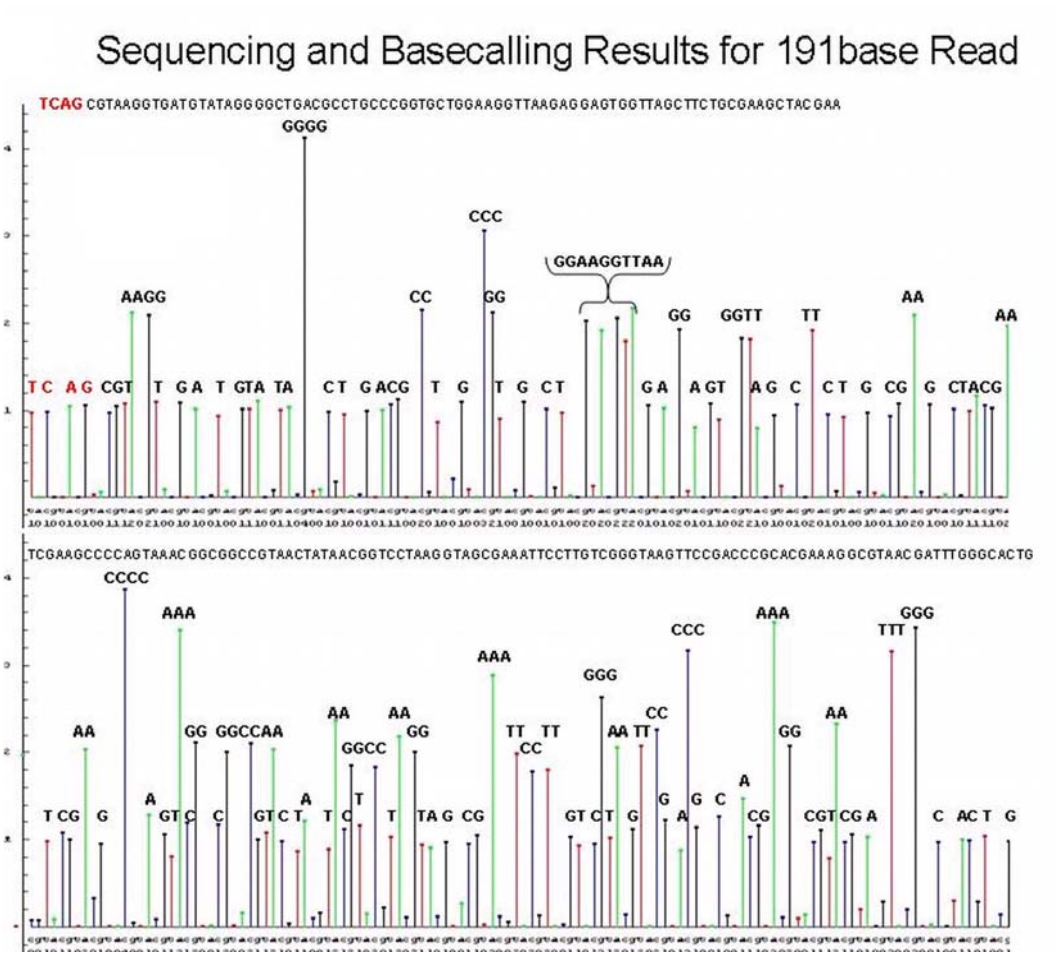
4) Load beads and enzymes in PicoTiter Plate™




5) Perform sequencing by synthesis on the 454 instrument

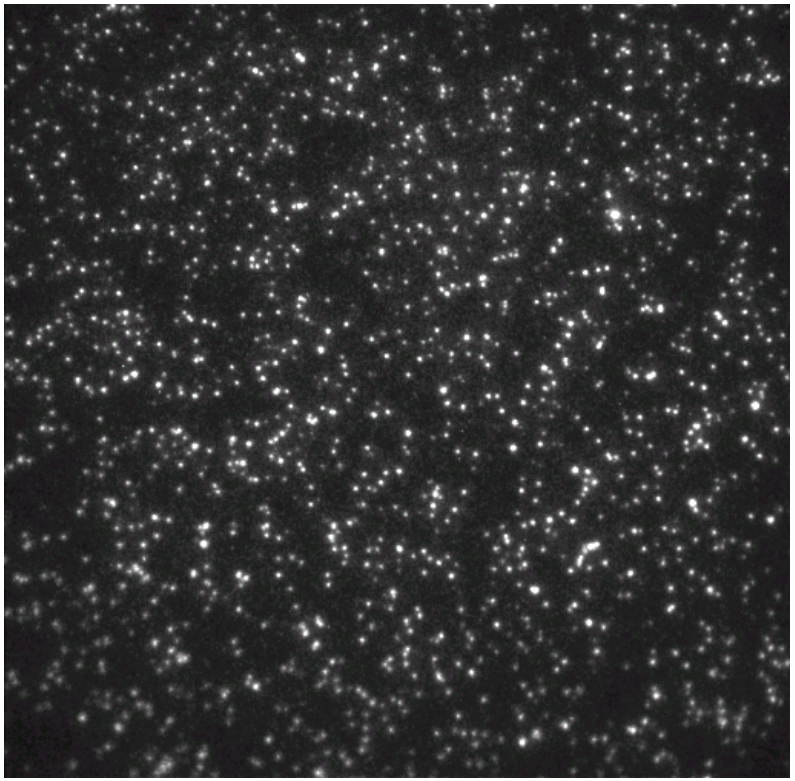
# 454 Technology - Sequencing Instrument




M Margulies, *et al.*, 2005, Nature advance online publication 31 July

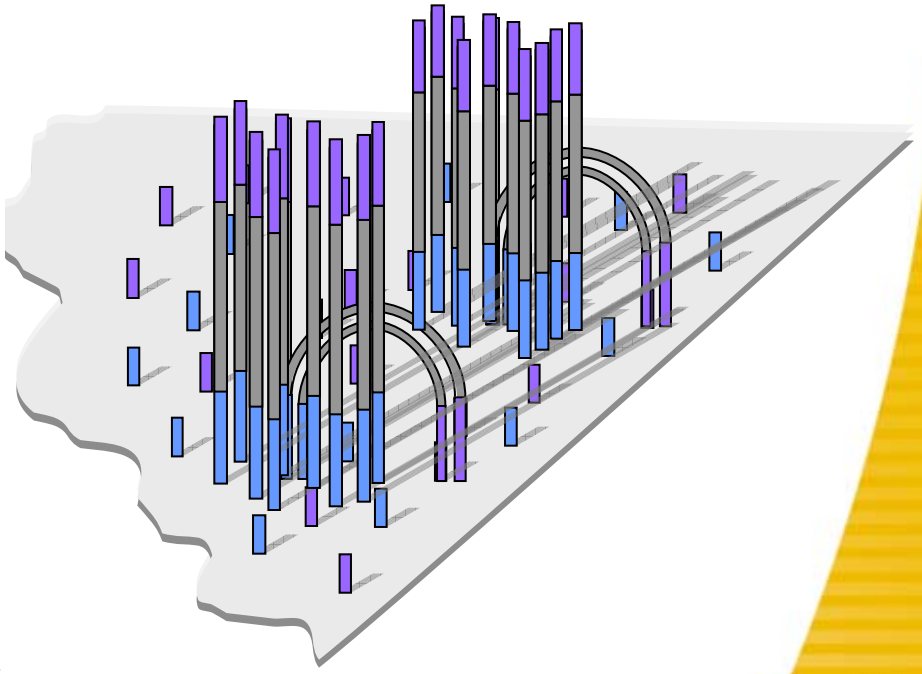
# Clonal Single Molecule Arrays™



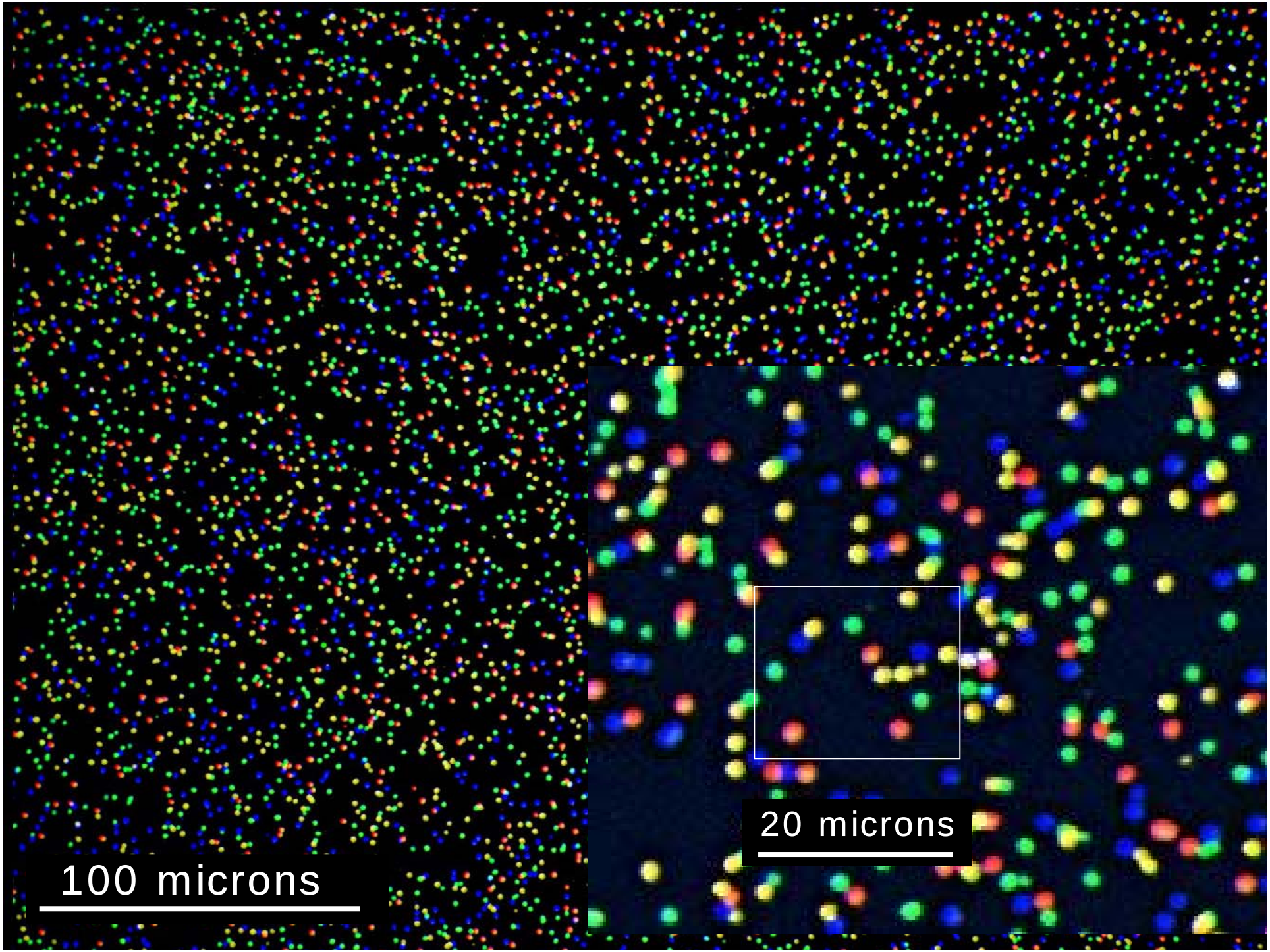
100um

Random array of clusters

**Attach single molecules to surface**
**Amplify to form clusters**

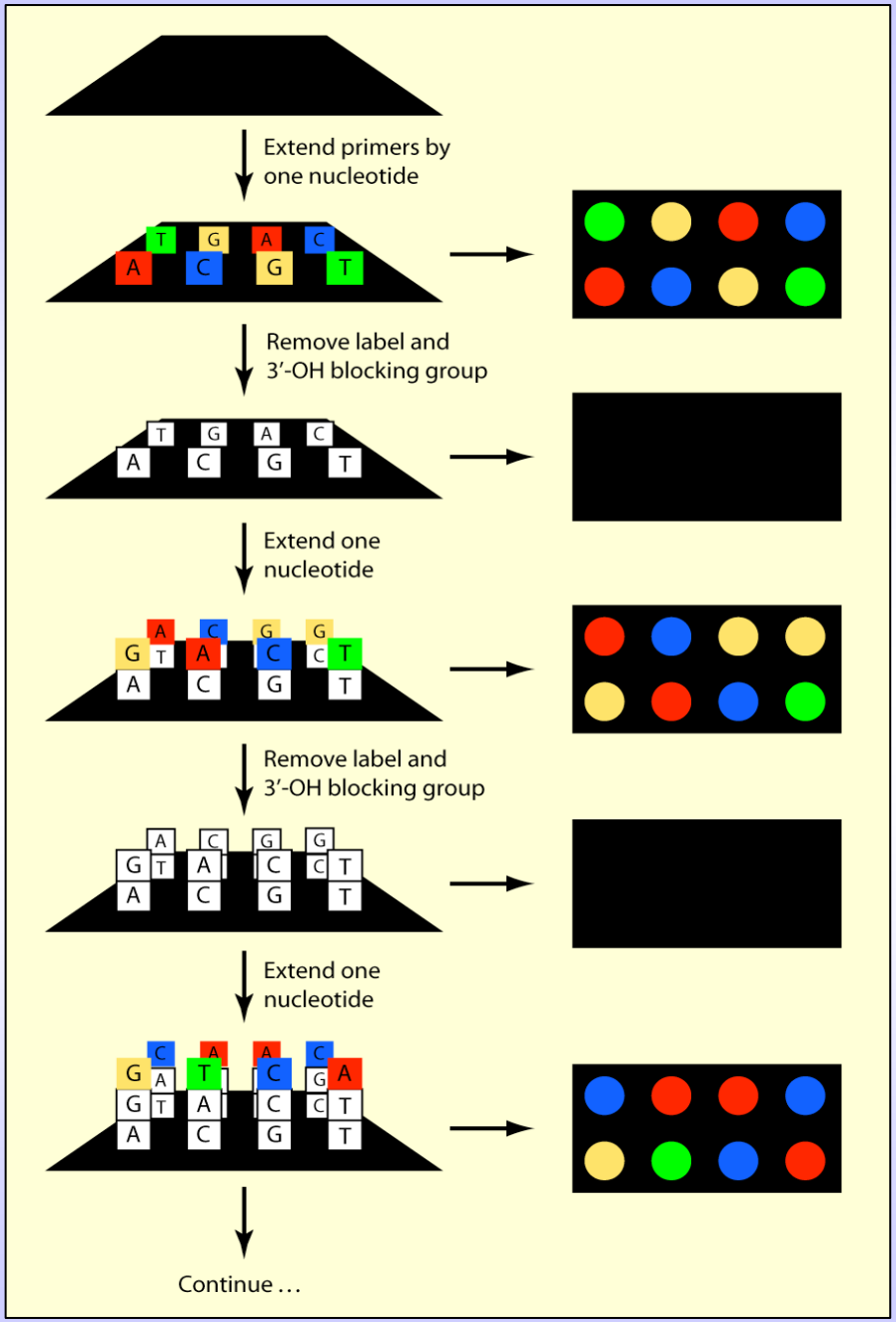


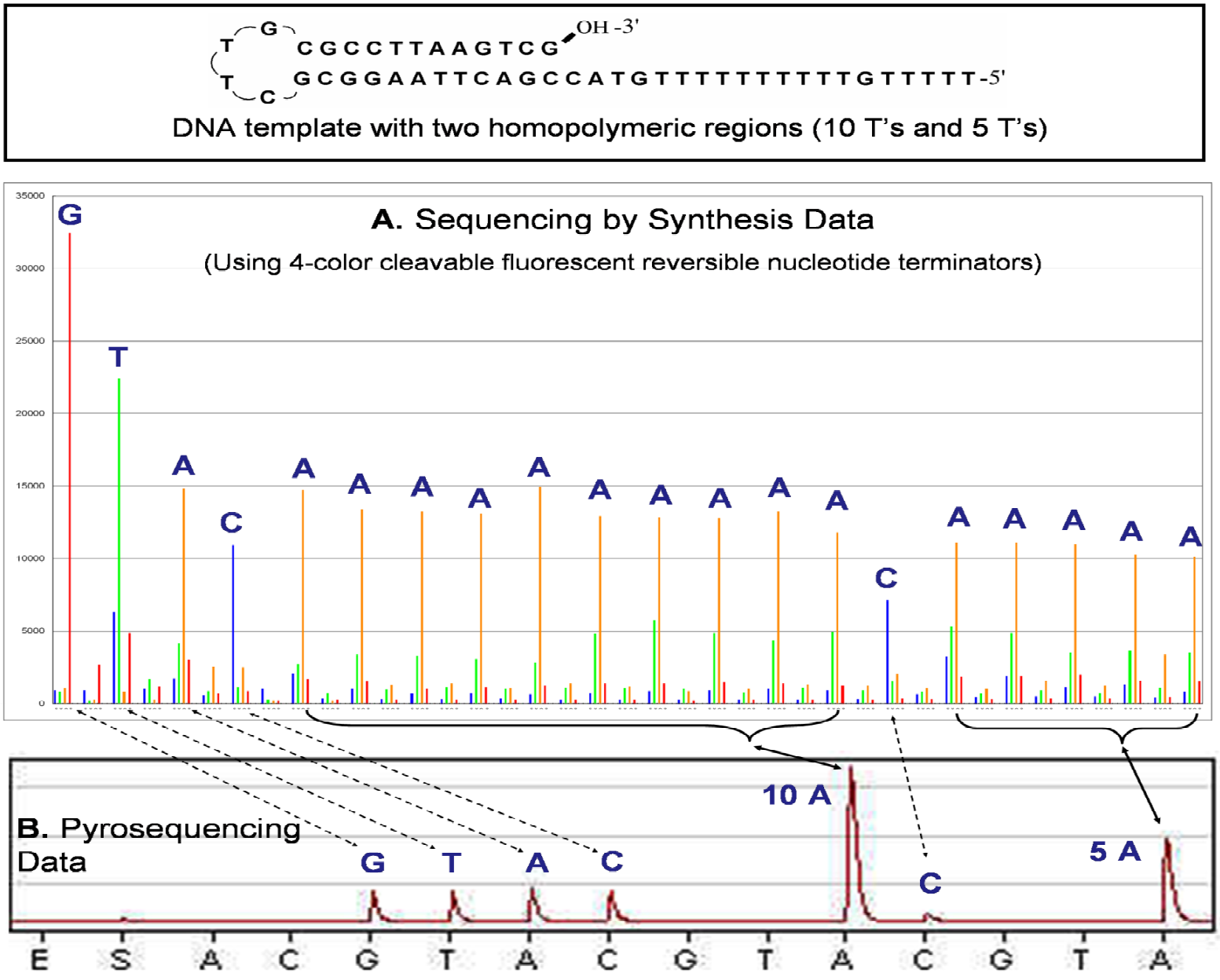
**1000 molecules per ~ 1 um cluster**
**1000 clusters per 100 um square**
**40 million clusters per experiment**


100 microns
20 microns

# SBS system using Reversible Photocleavable Fluorescent Nucleotide Terminators

**Jingyue Ju**
**Columbia University**

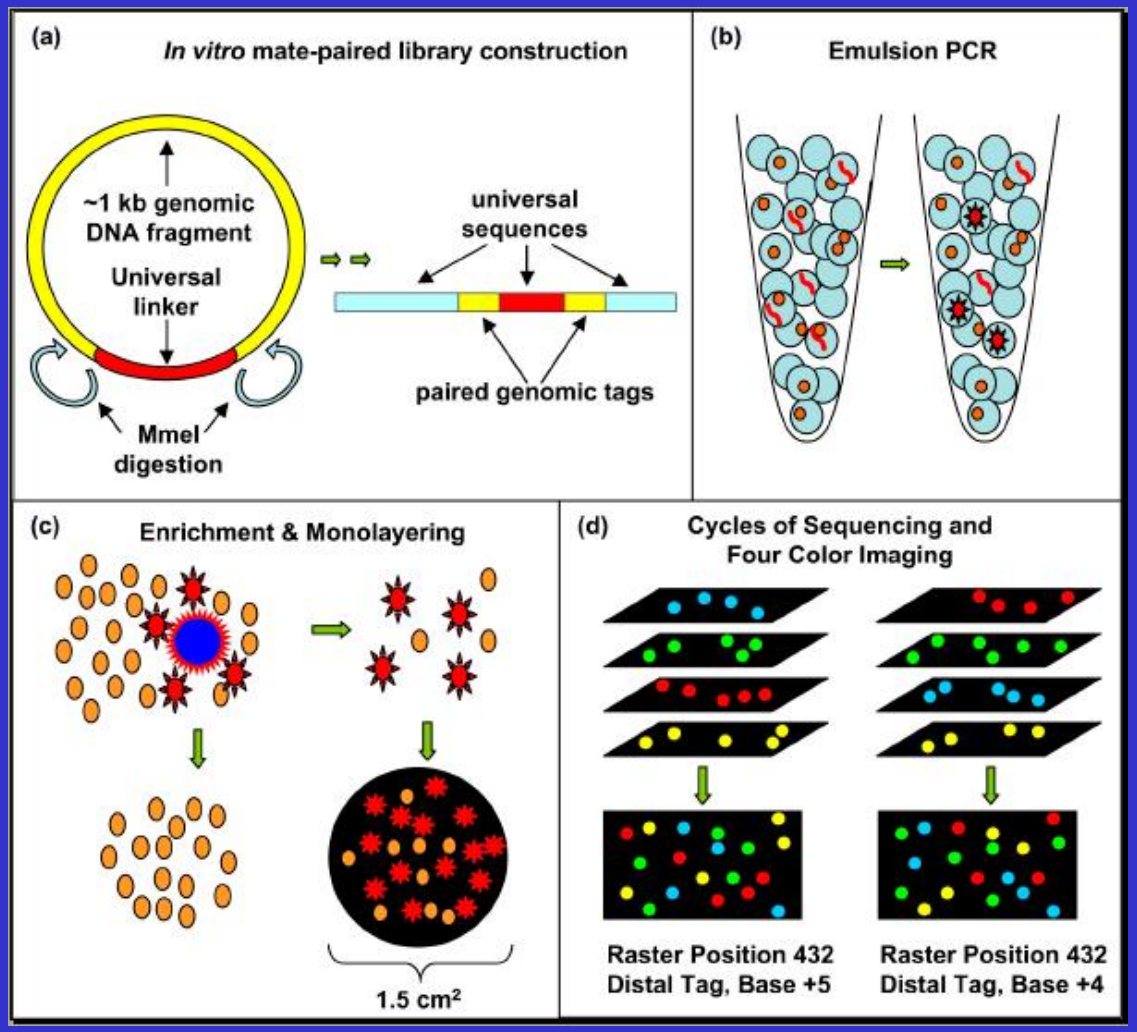





Jingyue Ju, Columbia University

Harvard Medical School
George M. Church

Single molecules generated from a cell-free, mate-paired library of *E. coli* genomic DNA, were amplified in parallel and attached to 1 micron beads, by emulsion polymerase chain reaction. Millions of beads were immobilized and subjected to automated cycles of sequencing by ligation and four-color imaging.




Off-the shelf reagents and an inexpensive epifluorescence microscope were used. Sequence accuracy is sufficient for re-sequencing applications.

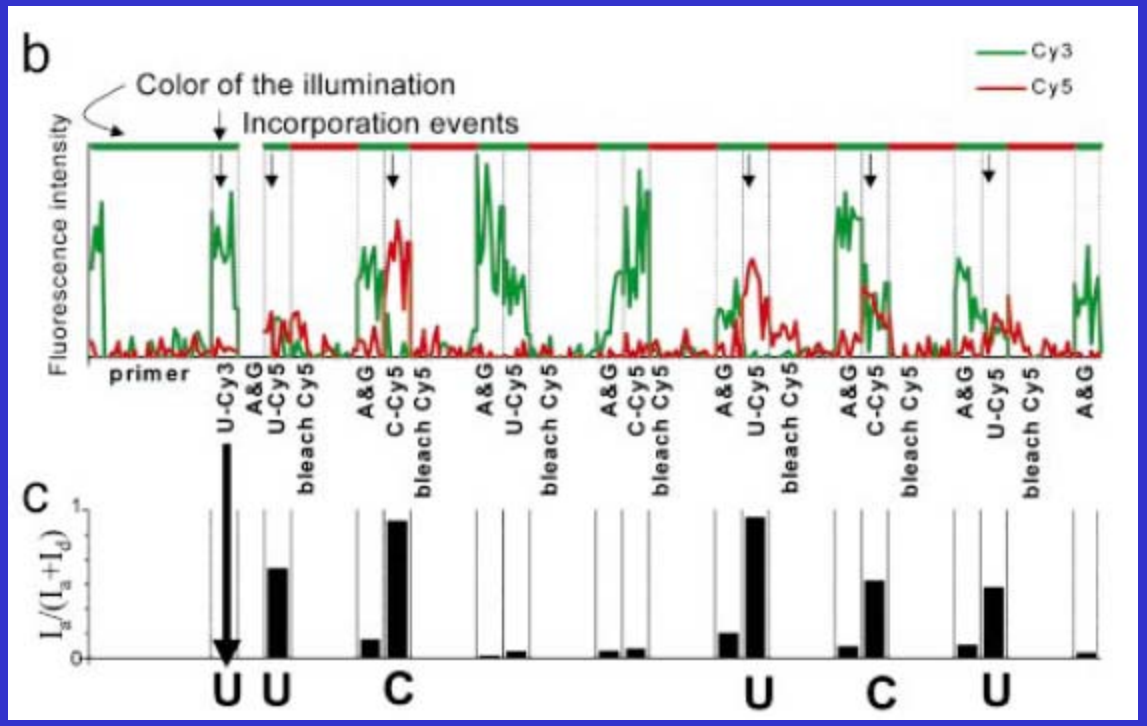
J Schendure, *et al.* Sciencexpress 4 August 2005
Science, 2005, *309*:1728-1732

NATIONAL HUMAN GENOME RESEARCH INSTITUTE

Stanford University
Steve Quake

Helicos Biosciences
Tim Harris

# Single molecules sequencing by synthesis.




Braslavsky, et al., 2003, PNAS *100*:3960-4

# Roche Applied Science -- Genome Sequencer FLX System (454)


First commercial sequencer available 2005; FLX released 2007

# Illumina Genome Analyzer (Solexa)


First commercial sequencer available 2006

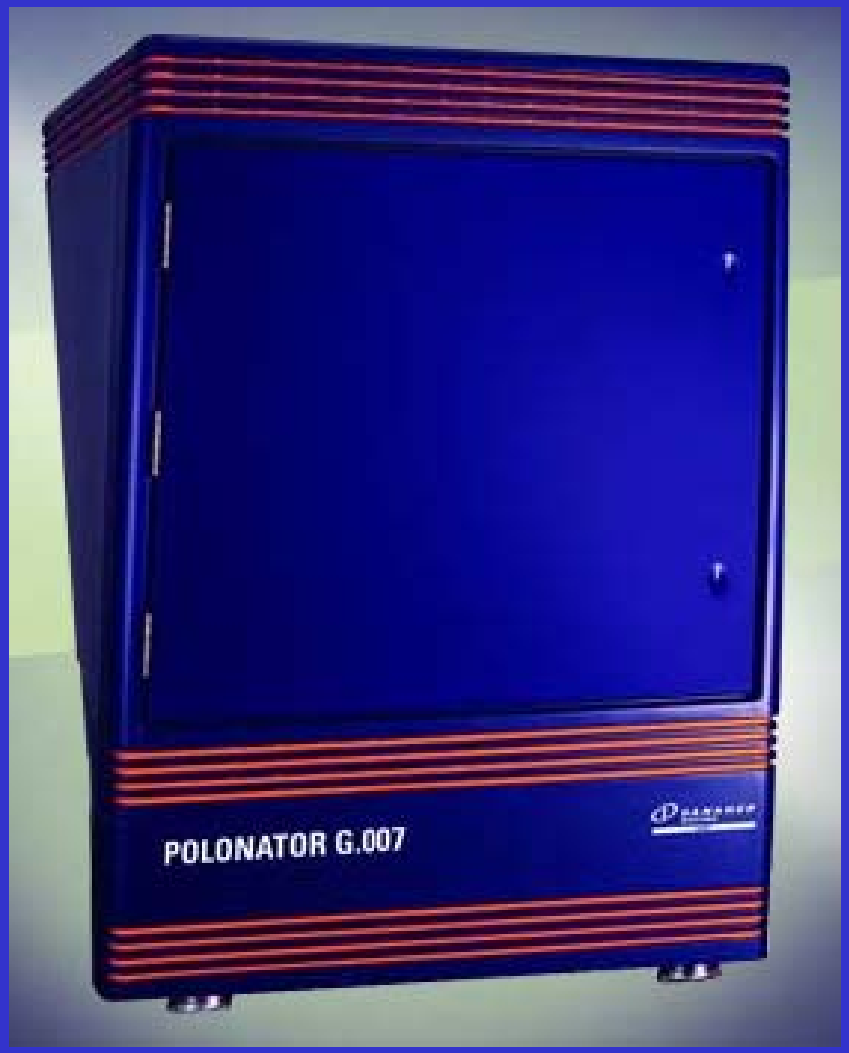
Applied Biosystems SOLiD™ System 2.0 (Agencourt)

First commercial sequencer available 2007

# Helicos HeliScope™ Single Molecule Sequencer

First commercial sequencer available 2008

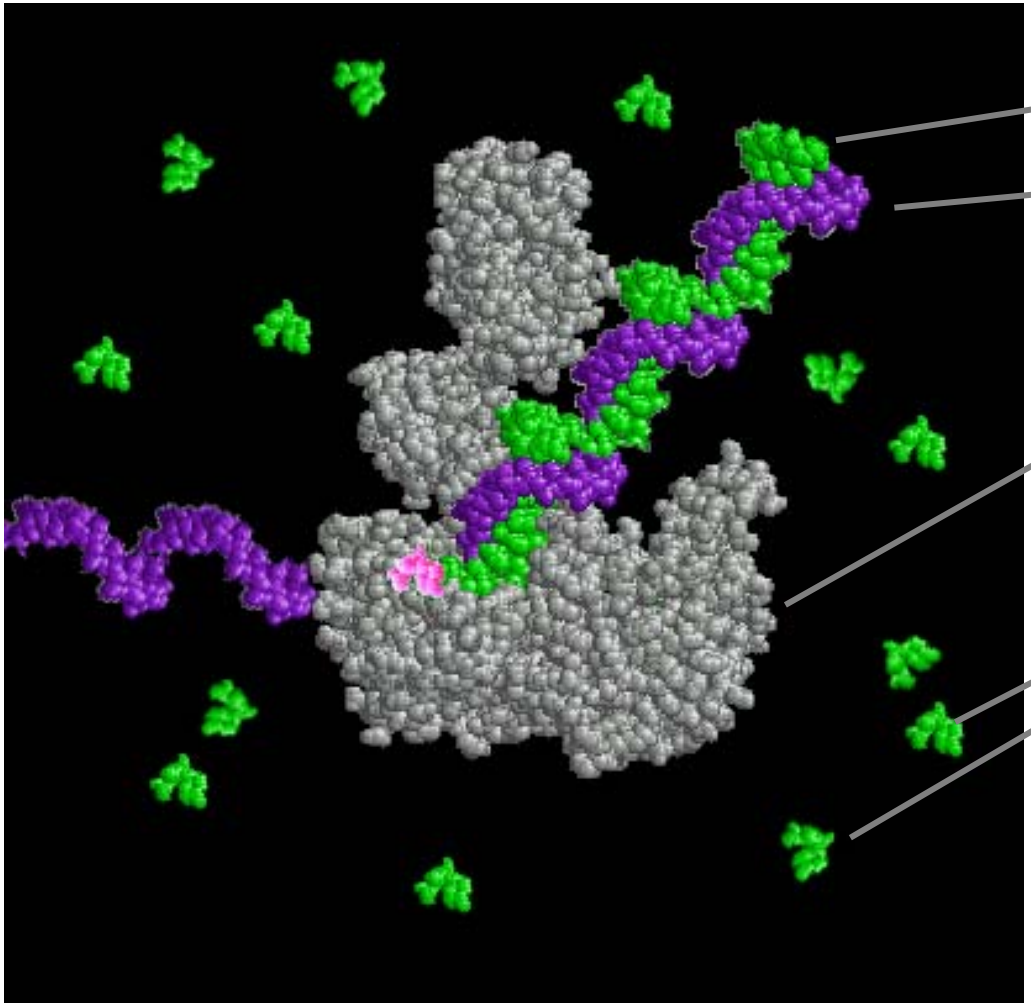
# Dover Systems The Polonator G.007




First commercial sequencer available 2008

# Free-running polymerase

# DNA Polymerase As a Sequence Reader



Primer

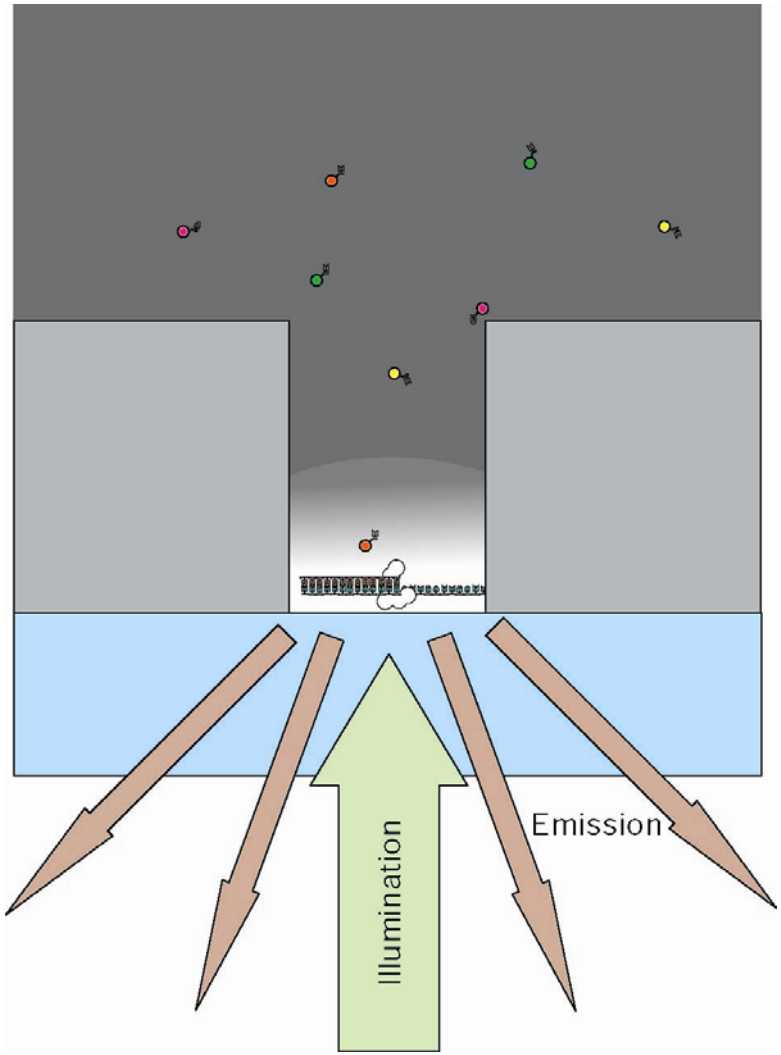
Template

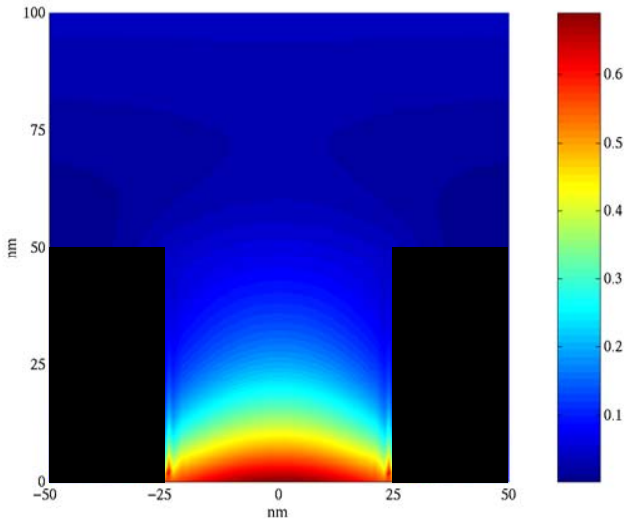
DNA polymerase

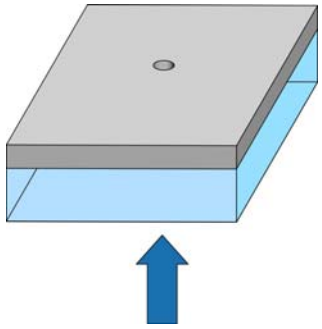
Nucleotides

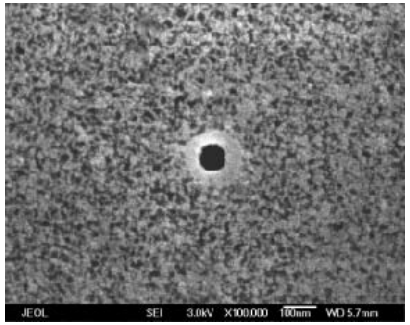
- Fast: 750 bps
- Frugal
- Faithful (1 in $10^5$ )
- Small

# Solution: Zero Mode Waveguide with Polymerase….



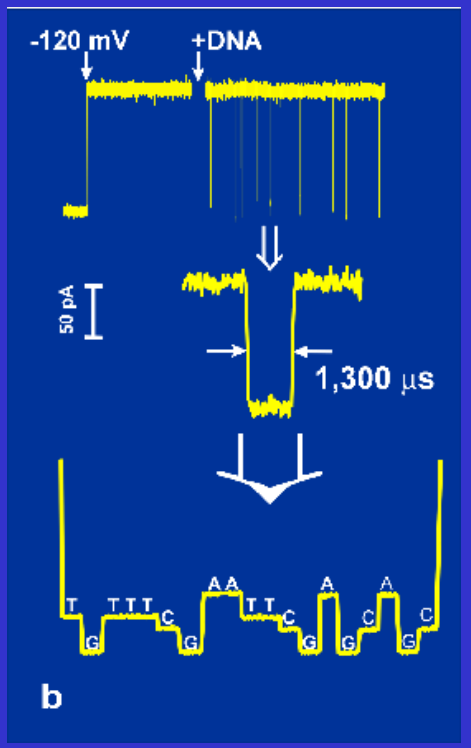





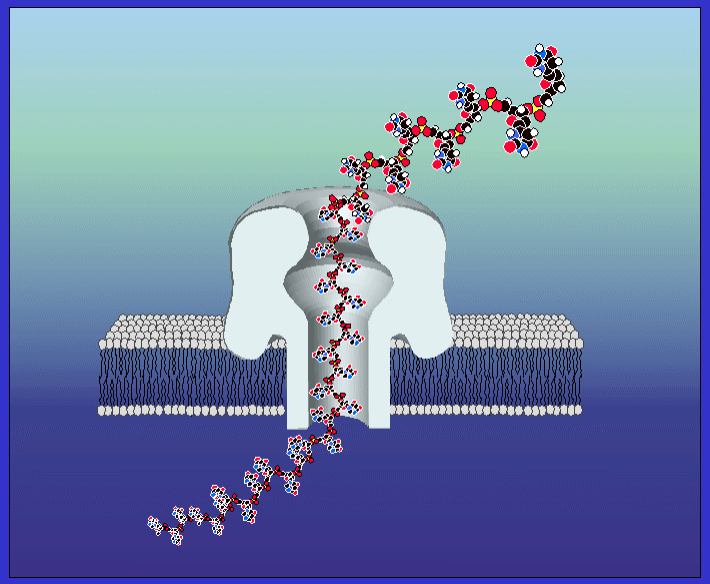
# Nanopore sequencing with electronic detection

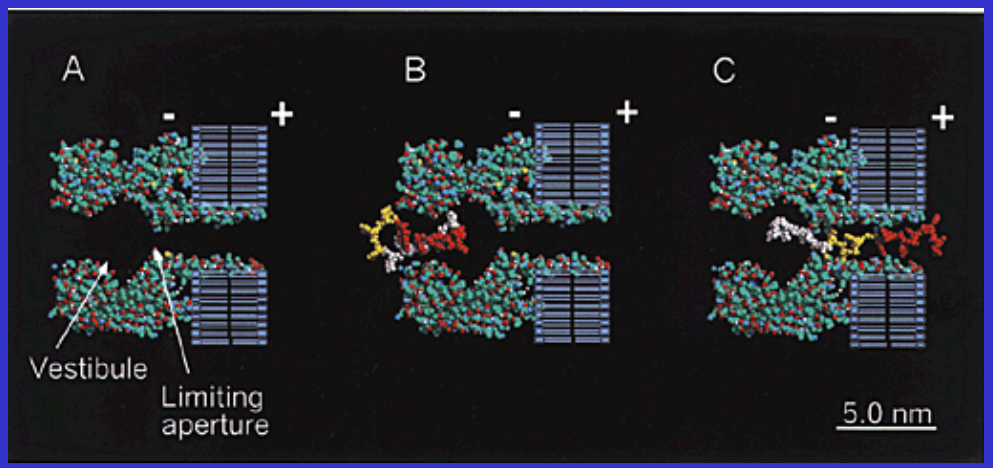
U of California, Santa Cruz
David W. Deamer, Ph.D
Mark Akeson, Ph.D.

Harvard University
Daniel Branton, Ph.D.
Jene Golovchenko, Ph.D




Single-stranded nucleic acid molecules passing through a nanometer-sized pore modulate the ionic conductance across the membrane. This observation may one day lead to a device for single molecule DNA sequencing.

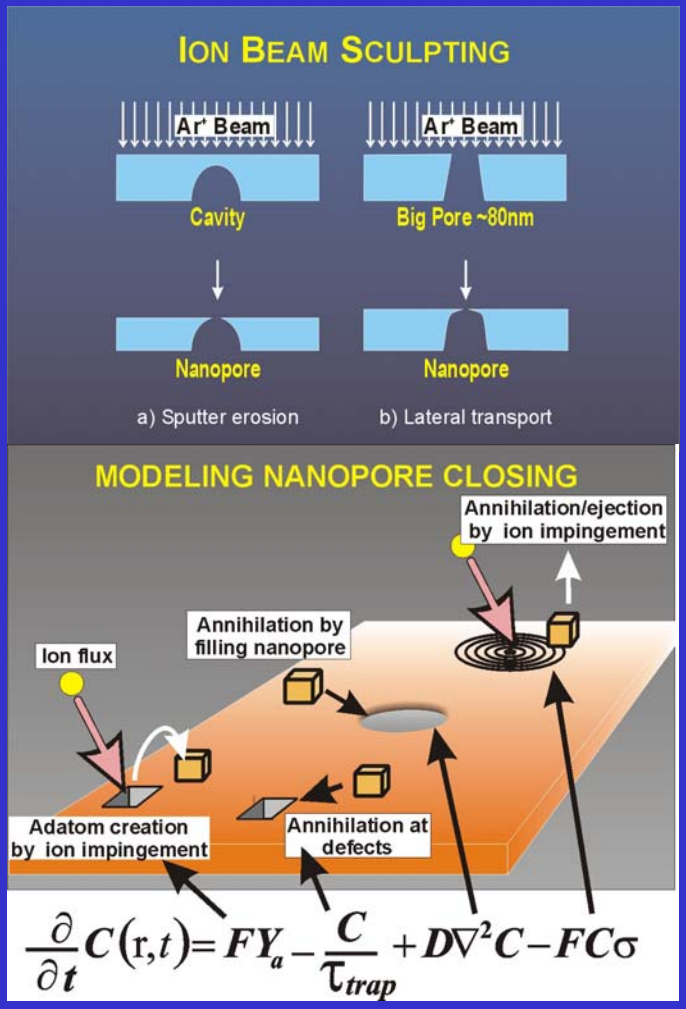





Nature Biotechnology 2001 *19*:248-252

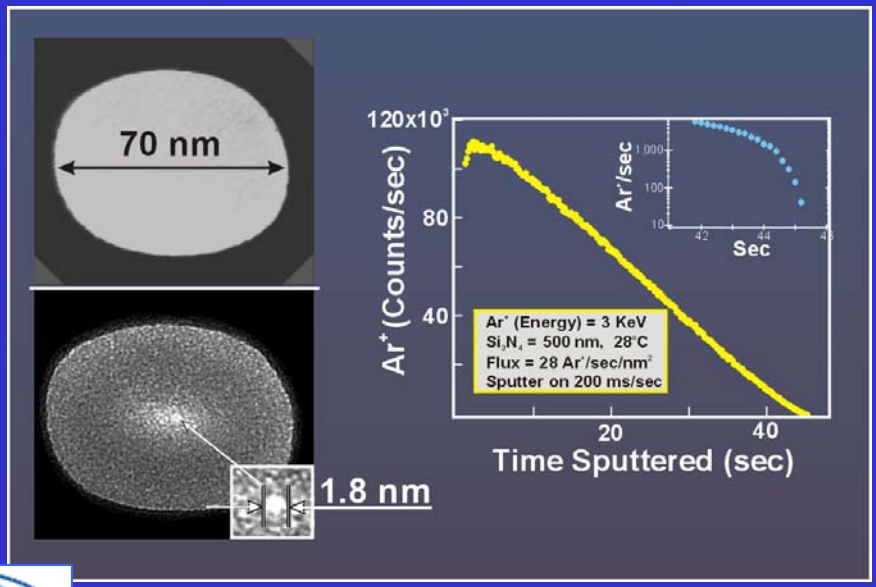
NATIONAL HUMAN GENOME RESEARCH INSTITUTE

Harvard University
Jene Golovchenko, Ph.D
Daniel Branton, Ph.D.




Nature 2001 *412:*166-169

Solid state fabrication methods were developed to create a pore small enough for single-stranded DNA analysis. A beam of massive argon ions closes a pre-made hole. Size control is achieved by monitoring ion flux through the pore. The result is a "robust electronic detector consisting of a single nanopore in a $Si_3N_4$ membrane, capable of registering single DNA molecules in aqueous solution."




DARPA

NATIONAL HUMAN GENOME RESEARCH INSTITUTE

Harvard University
Jene Golovchenko
Daniel Branton

U of California, Santa Cruz
David W. Deamer
Mark Akeson

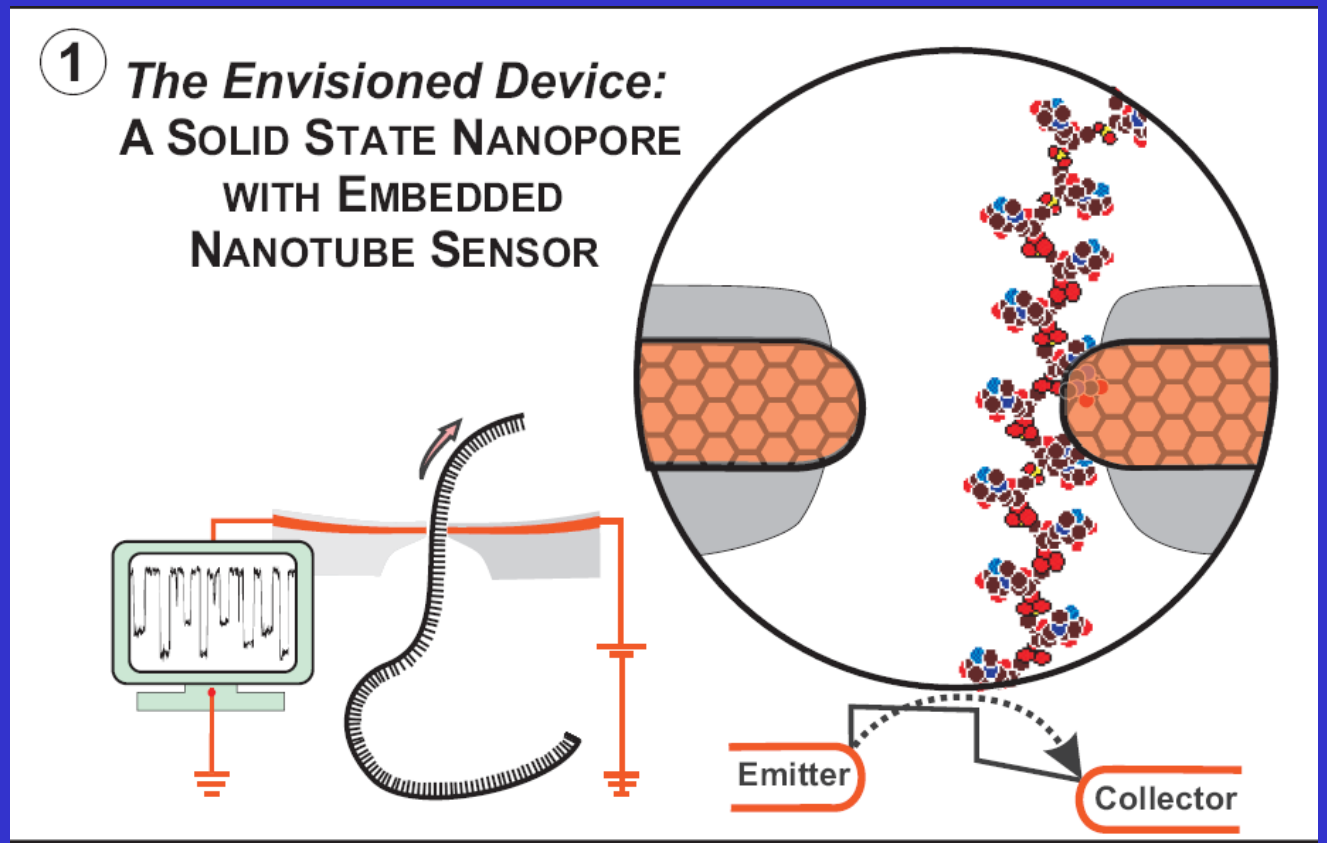



Figure 1. A biased nanopore translocates DNA molecules in sequential nucleotide order between probes that serve as emitter and collector of a tunneling "microscope"

http://www.mcb.harvard.edu/branton/

NATIONAL HUMAN GENOME RESEARCH INSTITUTE

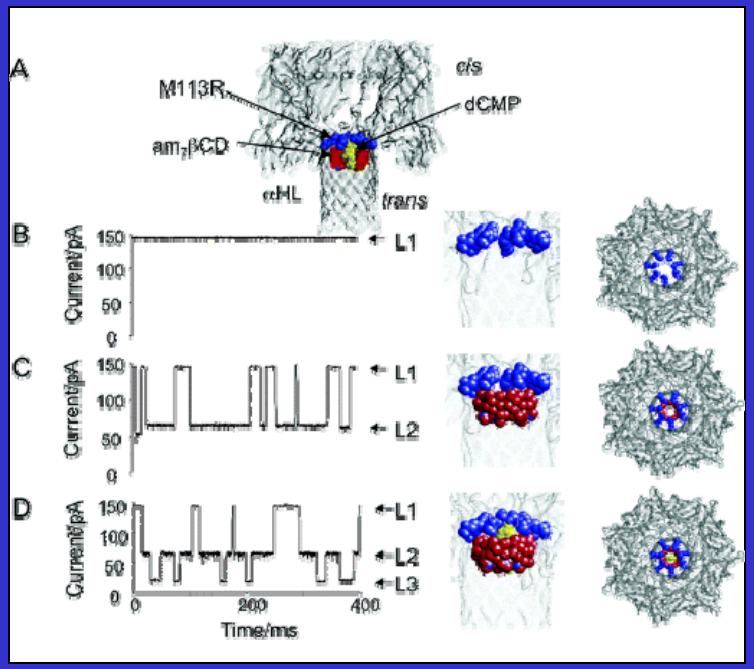
# DNA Sequencing Through Nanopore Sensors

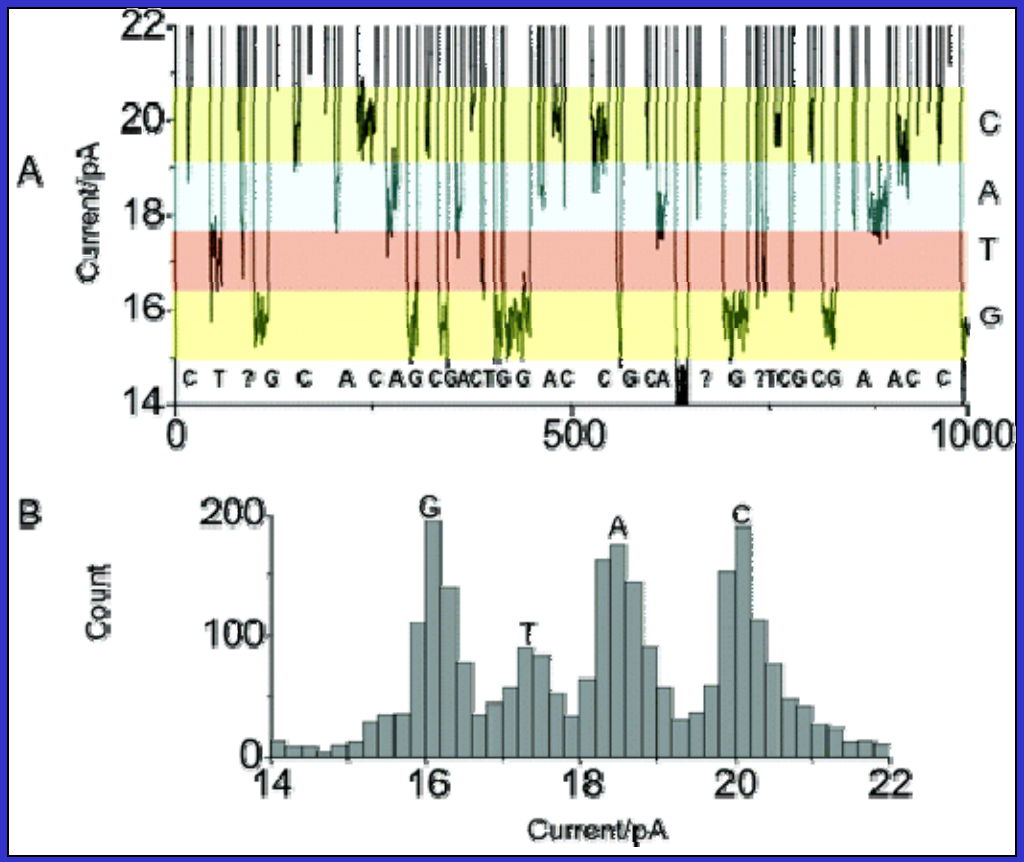
*Recent progress:*

- further demonstrations of single-base discrimination, though not yet sequencing
- fabricating pores/sensors
- understanding the physics of DNA transport through pores of the same diameter as the molecule
- analyzing the potential to distinguish between the four bases as the molecule passes the sensor

University of Oxford
J Hagan P Bayley

# Identification of Deoxyribonucleoside 5'-Monophosphates by Using an Engineered Protein Nanopore Equipped with a Molecular Adapter

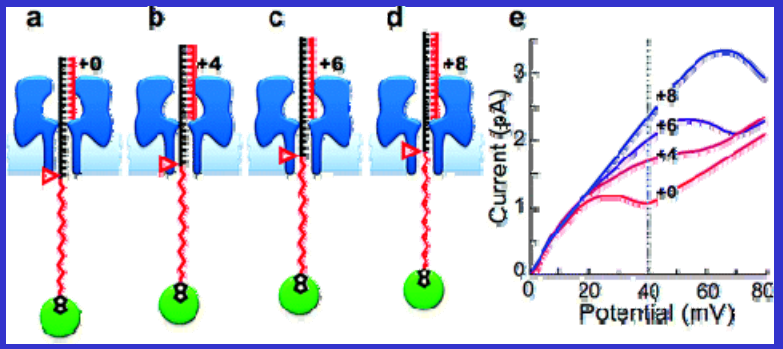


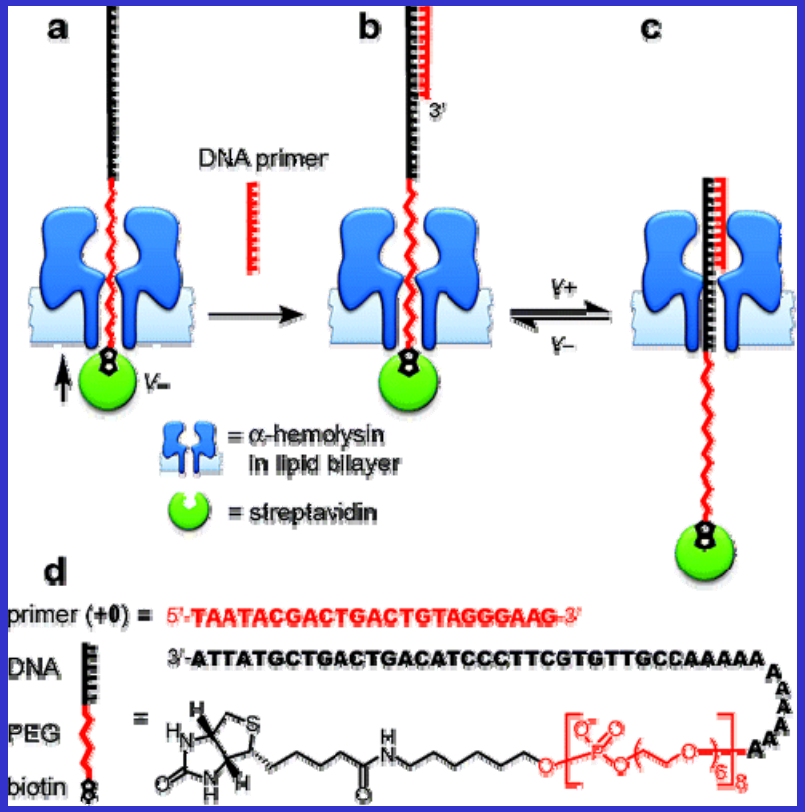

Yann Astier, Orit Braha, and Hagan Bayley
2006,*J. Am. Chem. Soc.*,128 (5), 1705 -1710

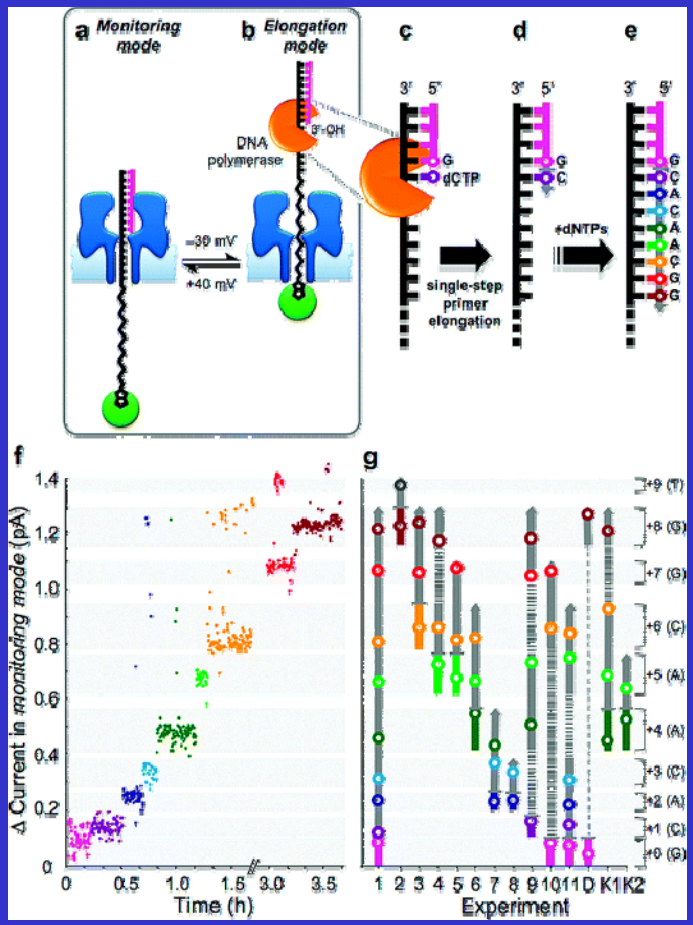
Scripps Research Institute
M Reza Ghadiri

# A Single-Molecule Nanopore Device Detects DNA Polymerase Activity with Single-Nucleotide Resolution



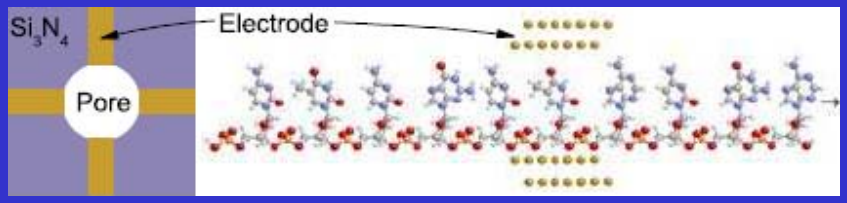


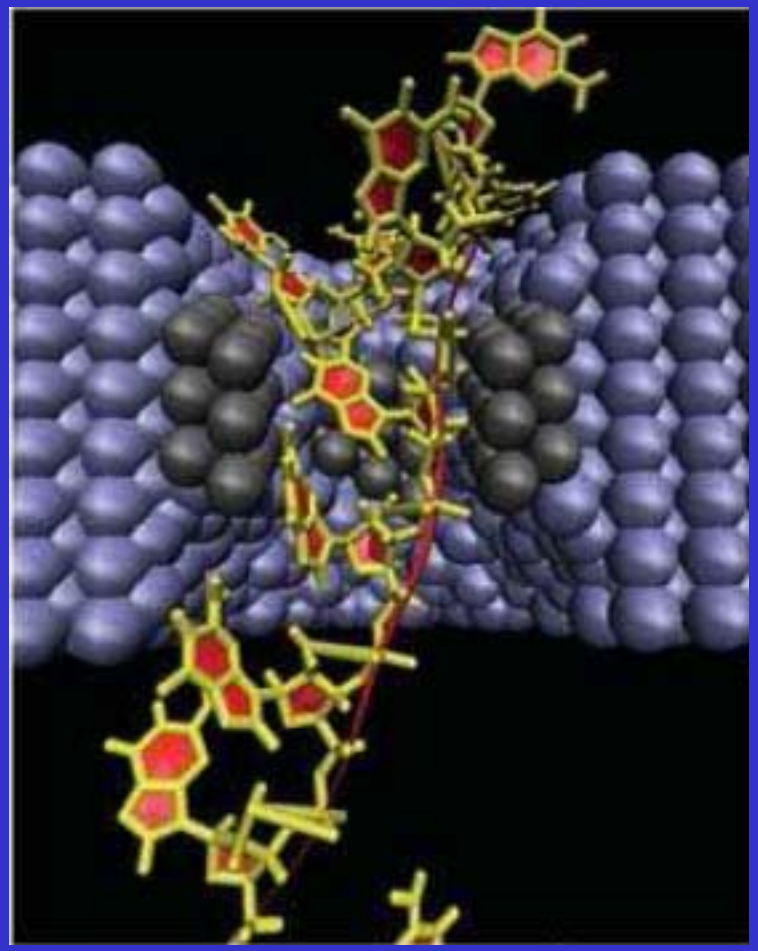




Cockroft, et al. 2008 J. Am. Chem. Soc. *130*:818-820

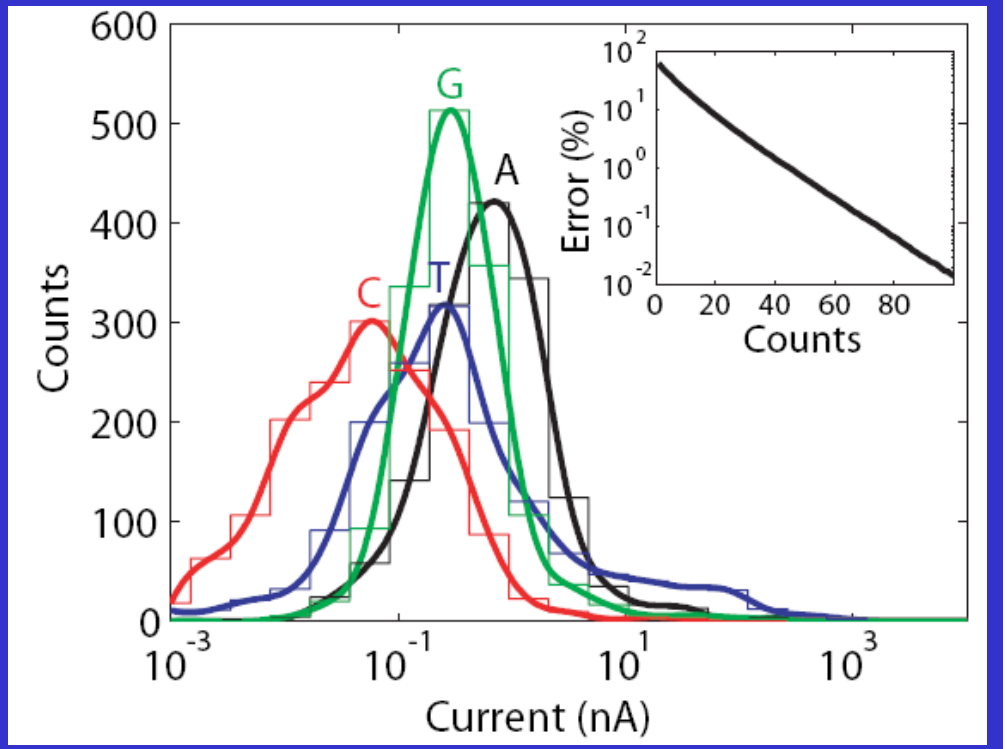
# University of North Carolina Chapel Hill J. Michael Ramsey



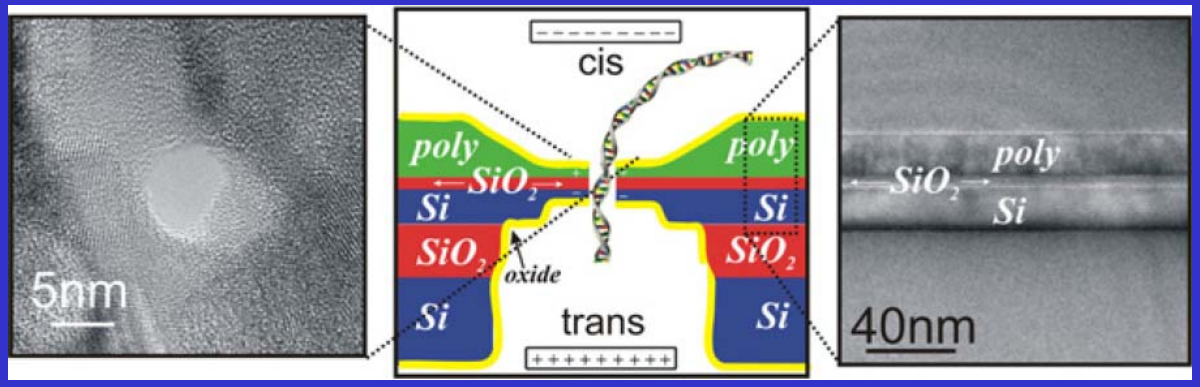

For the sensor configuration shown at left, with electrodes in the walls of a nanopore, modeling shows that the distributions of current values for each nucleotide will be sufficiently different to allow for rapid sequencing.



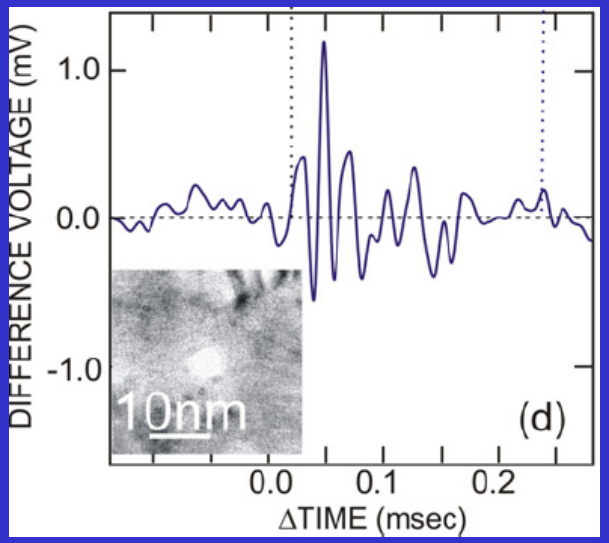
M. Di Ventra, 2006, Nano Lett. *6*:779-782

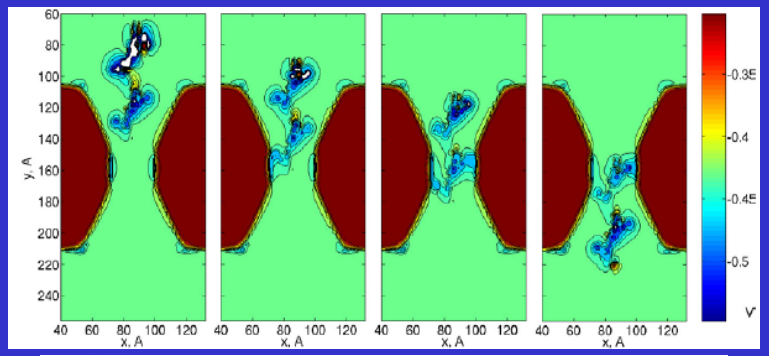


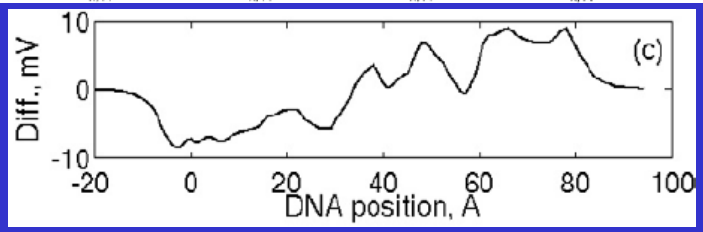
University of Illinois UC Gregory Timp




Measurement (lower left) of electrical signal difference between "poly" and "Si" electrodes (shown above) as DNA molecule passes through a 7 nm pore through the sensor. Computer simulation (lower right) of helical poly$C_{20}$ through the device produces "movie" frames in which DNA charge centers are seen moving through the sensor. Peaks in the simulated trace (lower right) show that electrical signals should be strong enough to detect individual bases. Narrower pores are needed to reduce noise.



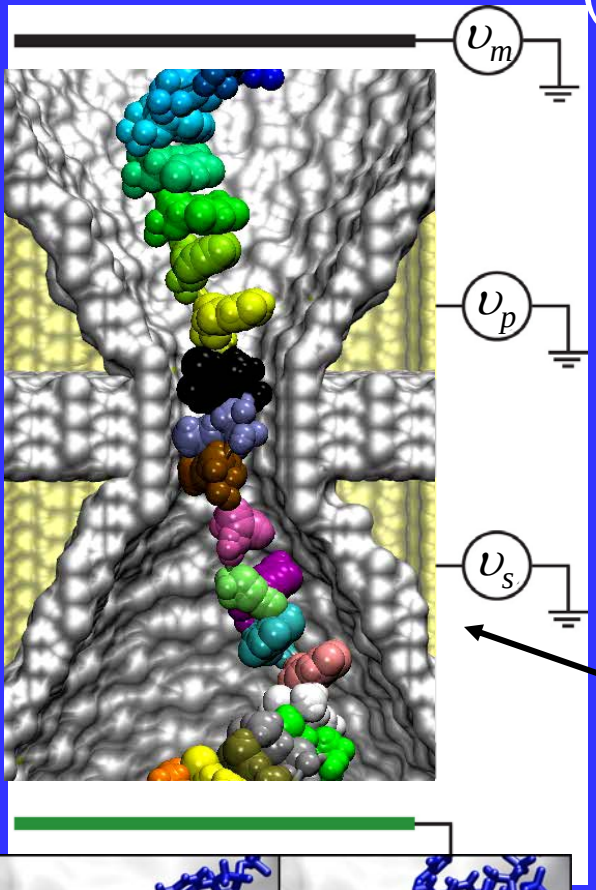



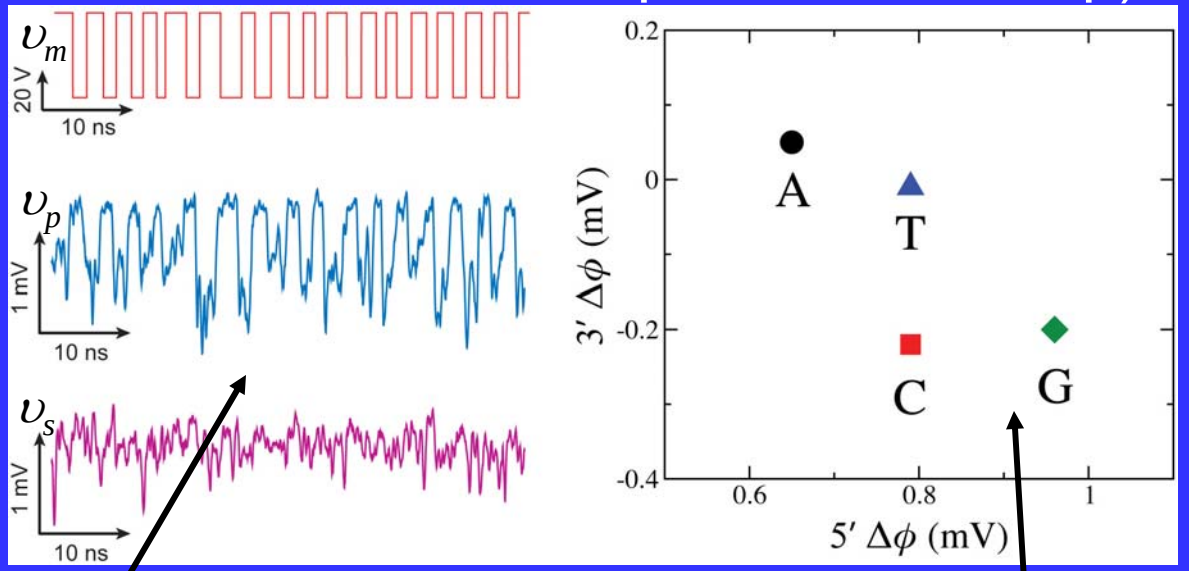




J-P Leburton, 2006, Nanotechnology *17*:622-633

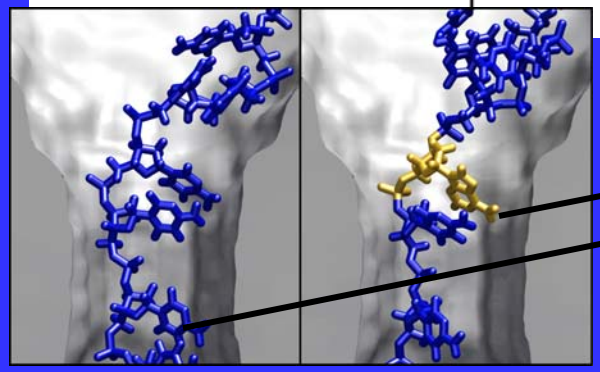
# Sequencing Using an AC-field in a Pore

(MD simulations of 1nm pore in MOS cap)

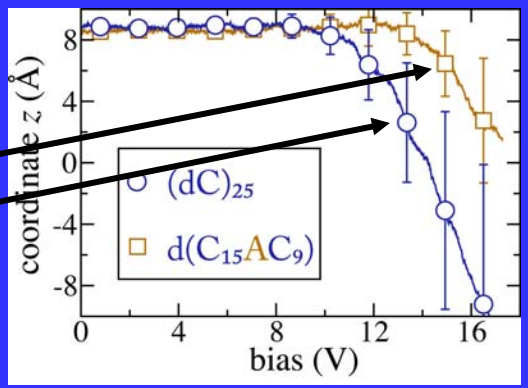



- bases tilt with AC $\mathscr{E}$-field in pore induced potential changes due to switching polarity of the applied bias
- **predict:** that nucleotides can be discriminated



dC$_{25}$ d(C$_{15}$AC$_9$)



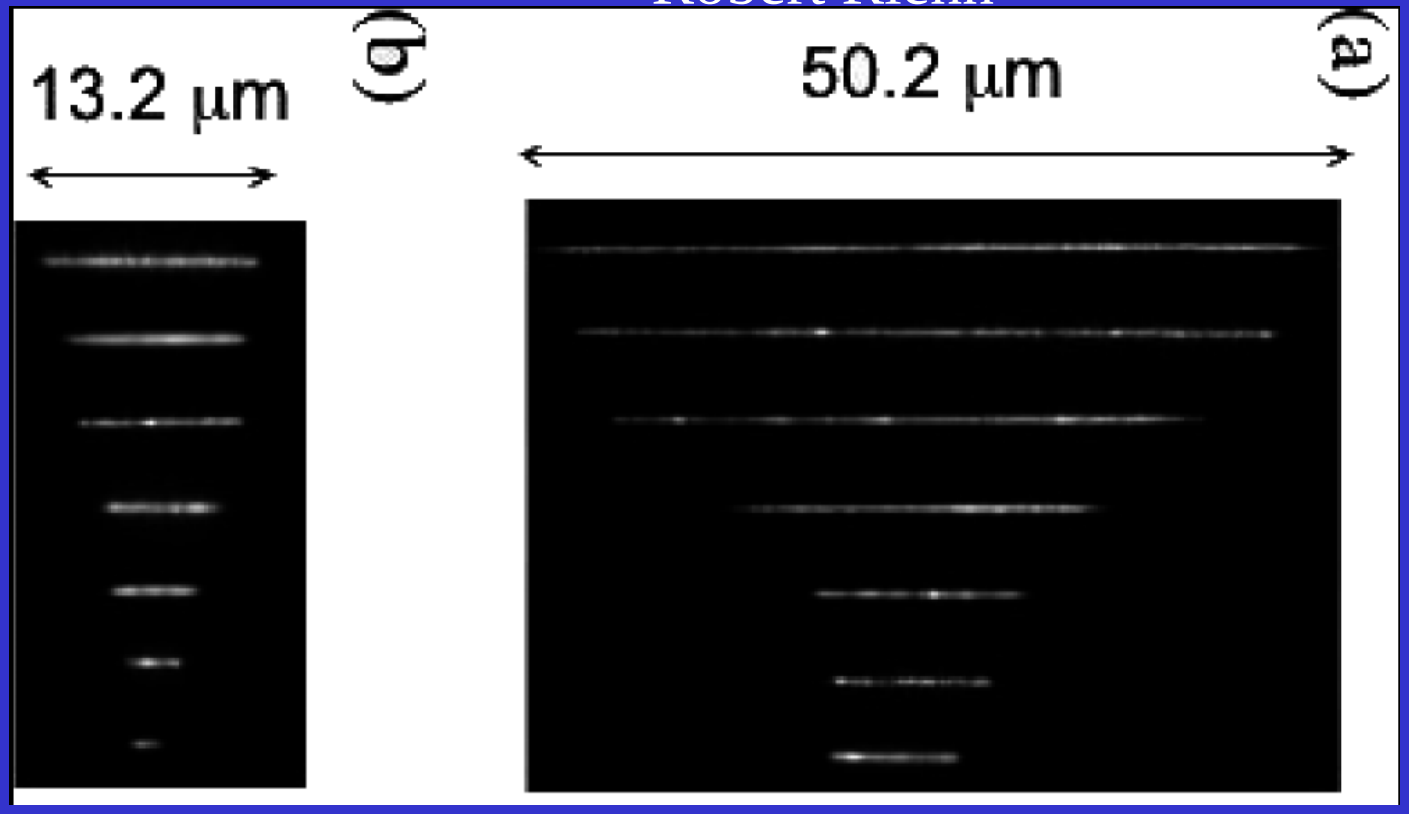

- **predict:** sequence can be determined with single nucleotide precision by gradually increasing bias

G Timp & A Aksimentiev

Nano Letters 2008 *8*,56

Princeton University
Edward C. Cox
Robert Austin

Cornell University
Harold G. Craighead
North Carolina State University
Robert Riehn




W. Reisner, et al., 2005, Phys. Rev. Lett. 94, 196101

"…nanochannels would confine the DNA over its entire length rather than at a single point, and…one would expect to see a straight DNA molecule moving by without any kinks."

RH Austin (2003) Nature Materials 2, 567-568

# Throughput of DNA Sequencing Technologies

Baseline:

Human Genome “Reference” Sequence

$300 million
~8 years

# Throughput of DNA Sequencing Technologies

Capillary Array Electrophoresis

96 channels x 24 runs/day x 800 bp per run ≈ 1.8 Mb/day

6x coverage of 3 Gb genome takes 26 years with 1 machine,

~ 3 months with 100 machines

# Throughput of DNA Sequencing Technologies

Capillary Array Electrophoresis

96 channels x 24 runs/day x 800 bp per run ≈ 1.8 Mb/day

6x coverage of 3 Gb genome takes 26 years with 1 machine,

~ 3 months with 100 machines

Sequencing by synthesis on array

1 Gb/run, 2.5 days/run,

20x coverage of 6 Gb genome takes 1 year

these are still early days for this collection of emerging technologies → e.g., 4-6x improvement over next year

~2 months with one machine

# Throughput of DNA Sequencing Technologies

Capillary Array Electrophoresis

96 channels x 24 runs/day x 800 bp per run ≈ 1.8 Mb/day

6x coverage of 3 Gb genome takes 26 years with 1 machine,

~ 3 months with 100 machines

Sequencing by synthesis on array

1 Gb/run, 2.5 days/run,

20x coverage of 6 Gb genome takes 1 year

these are still early days for this collection of emerging technologies → e.g., 4-6x improvement over next year

~2 months with one machine

Nanosensor

1 msec per base

10x coverage of 6 Gb genome takes

~ 2 years with single nanopore;

< 1 day with 1000 nanopore array

# Applications

- sequence variation (SNP, indel, and larger)
  - rare variants, not just the common ones
- haplotypes
- rearrangements
- expression analysis -- COUNTING
- allele-specific expression analysis
- alternative splicing
- microRNAs
- rare samples (e.g., in mixtures) – dynamic range
- genomes – re-sequencing, *de novo?*
- targeted regions (some)
- methylation status

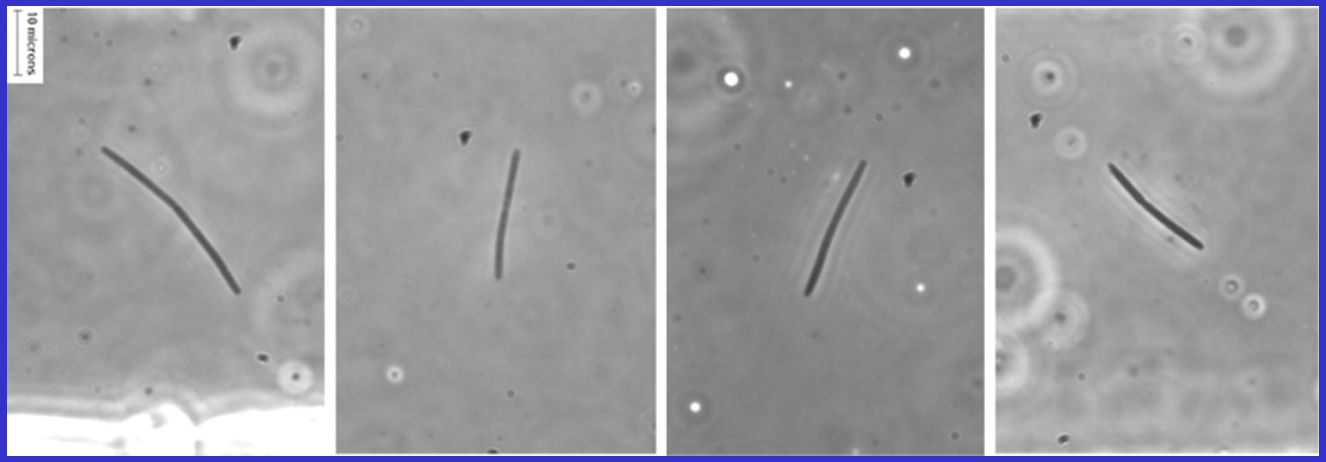
# Next-next generation genome sequencing?

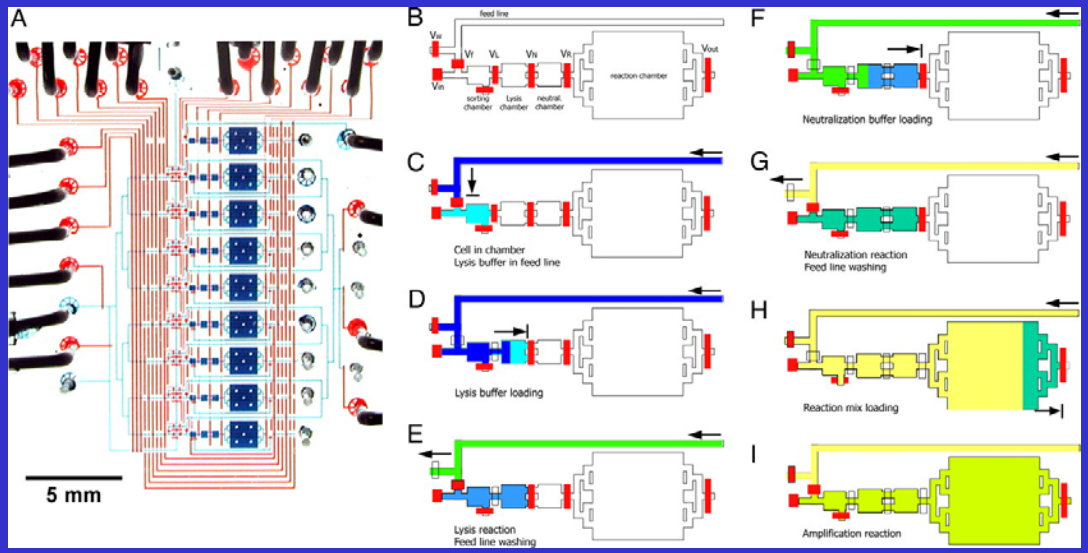
- Cost
- MUCH longer reads
  - Assembly, haplotypes, microbiomes…
- Read all 5 bases
- Read RNA (directly?), protein
- Re-read same template – impact on data quality
- …
- Medical care for individual patients

# Single-Cell Analysis

to get at the ‘unculturables’

Requires ability to read a LOT of genomes!




Marcy et al., 2007, PNAS 104:11889