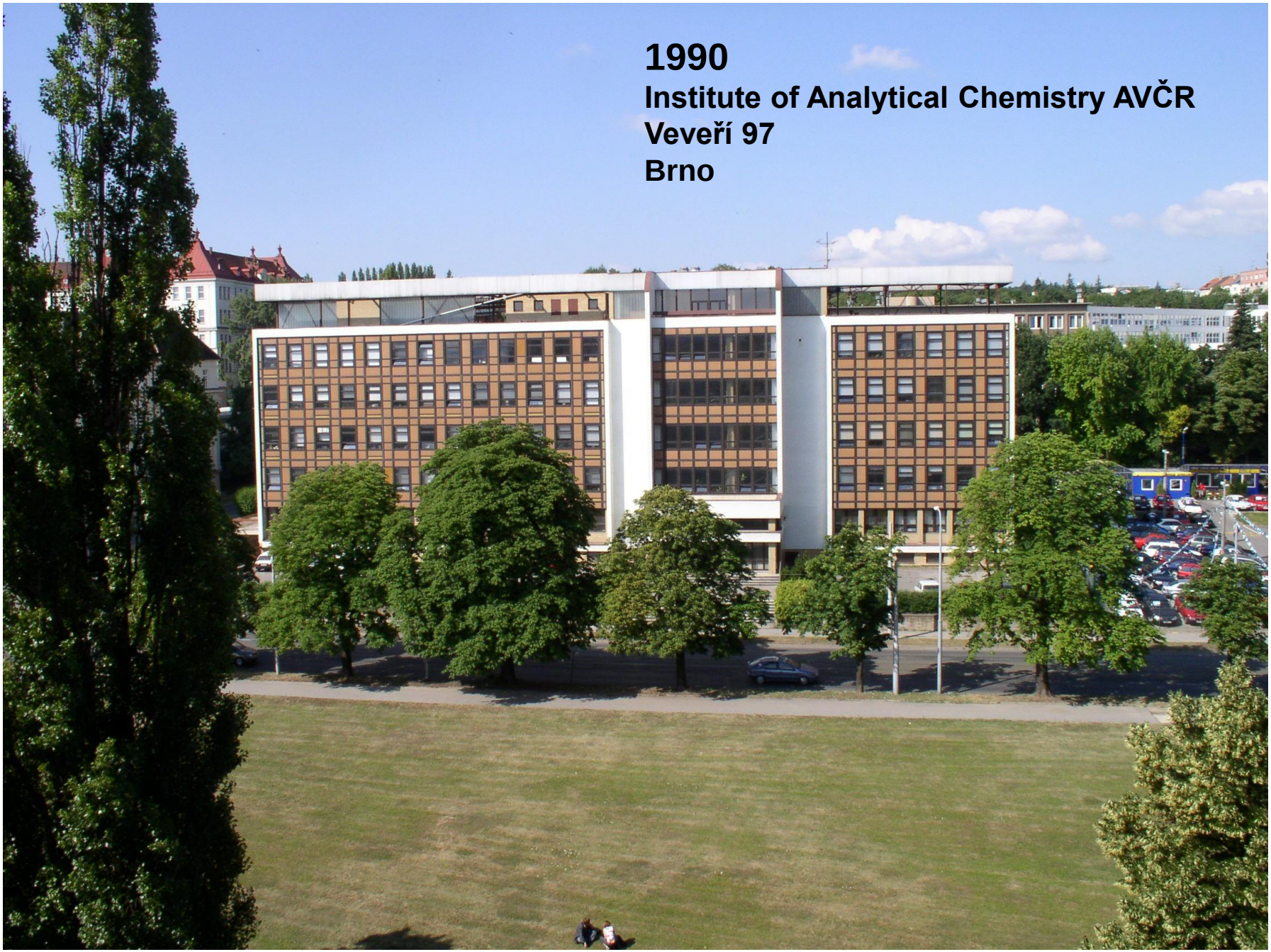
Modern analytical instrumentation for genetic research, medical diagnostics and molecular identification of organisms

Karel Klepárník
(klep@iach.cz)

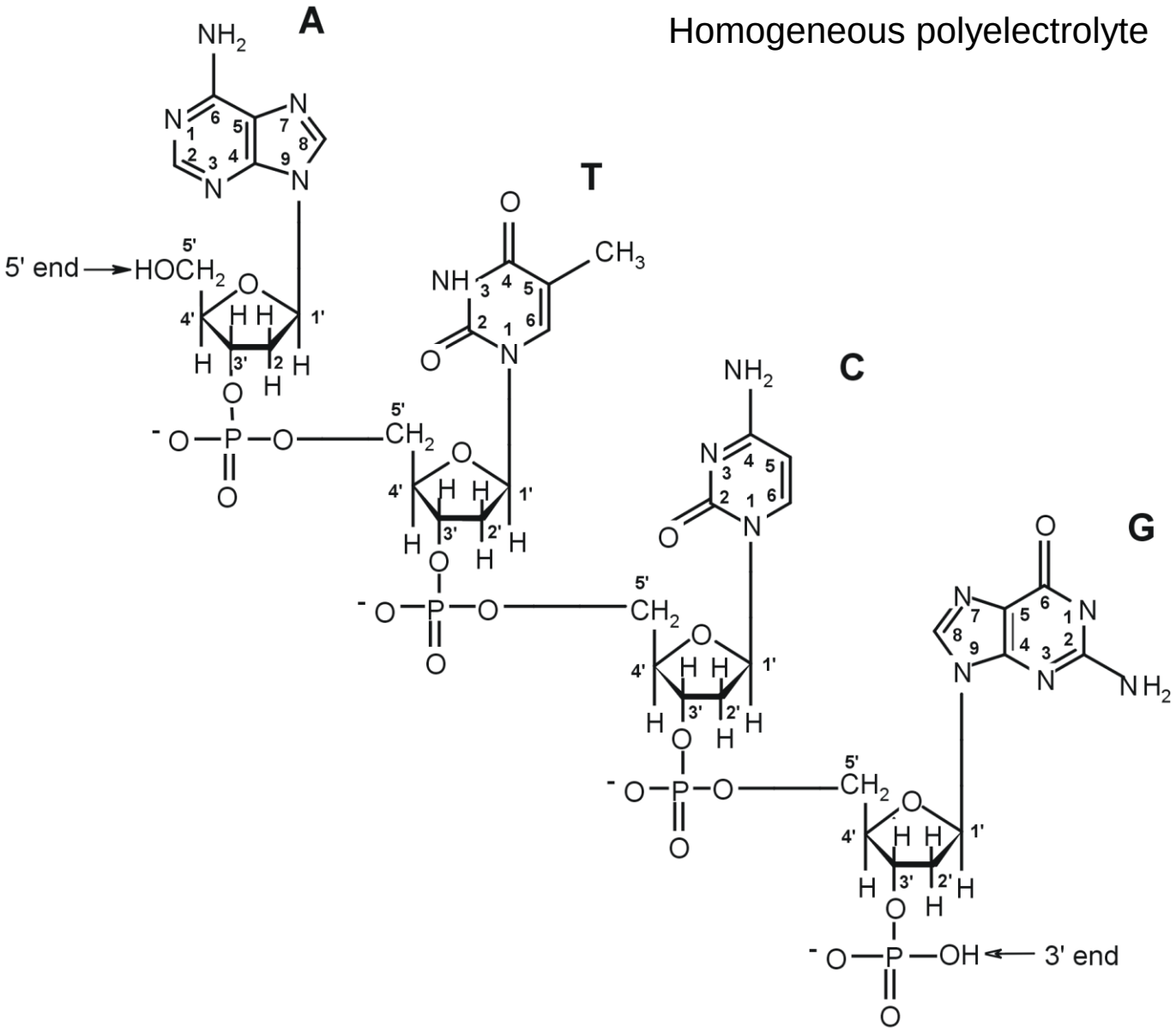
Department of Bioanalytical Instrumentation
Institute of Analytical chemistry
Czech Academy of Sciences
Brno
(www.iach.cz)

iac brno


1990
Institute of Analytical Chemistry AVČR
Veveří 97
Brno

DNA primary structure

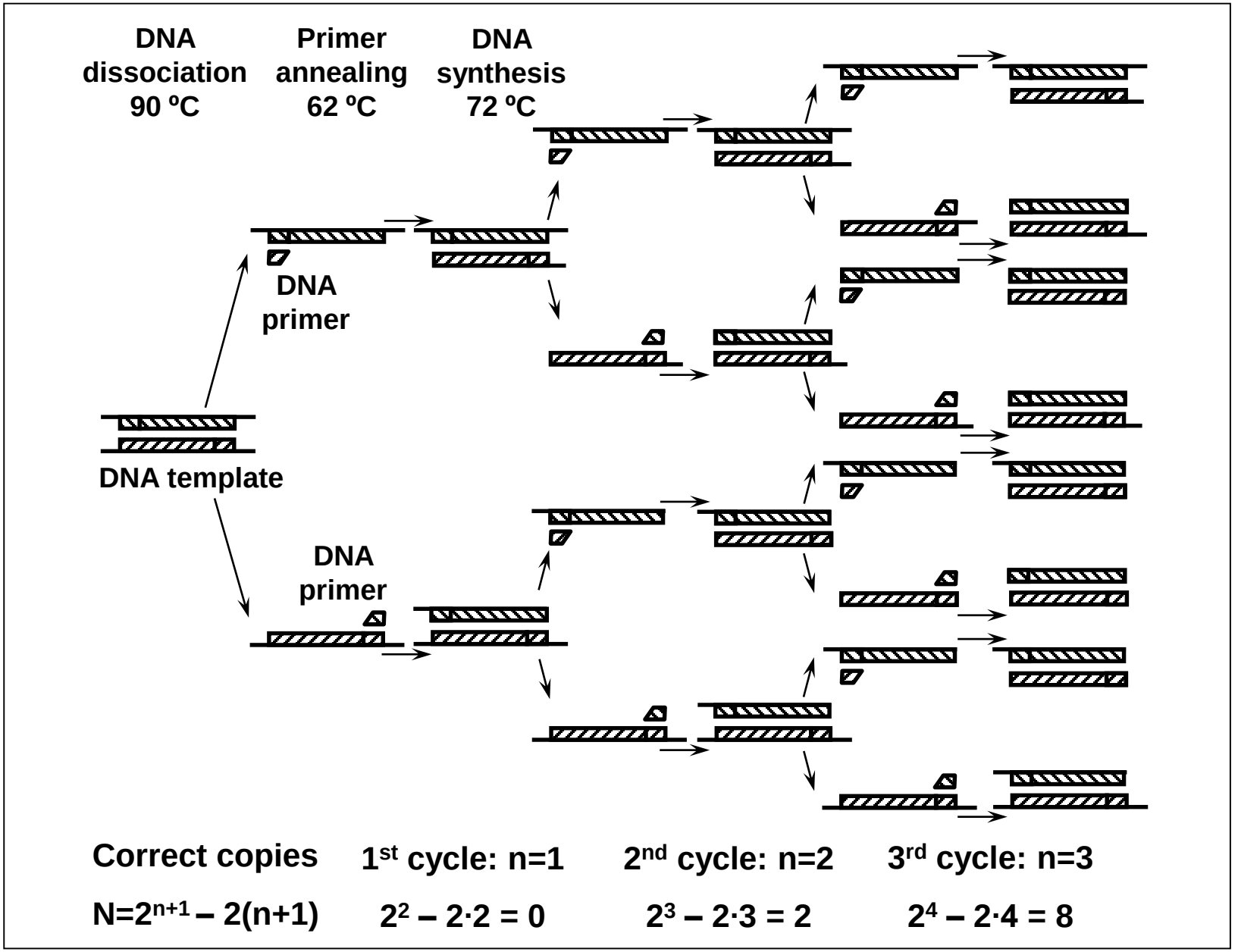
Homogeneous polyelectrolyte

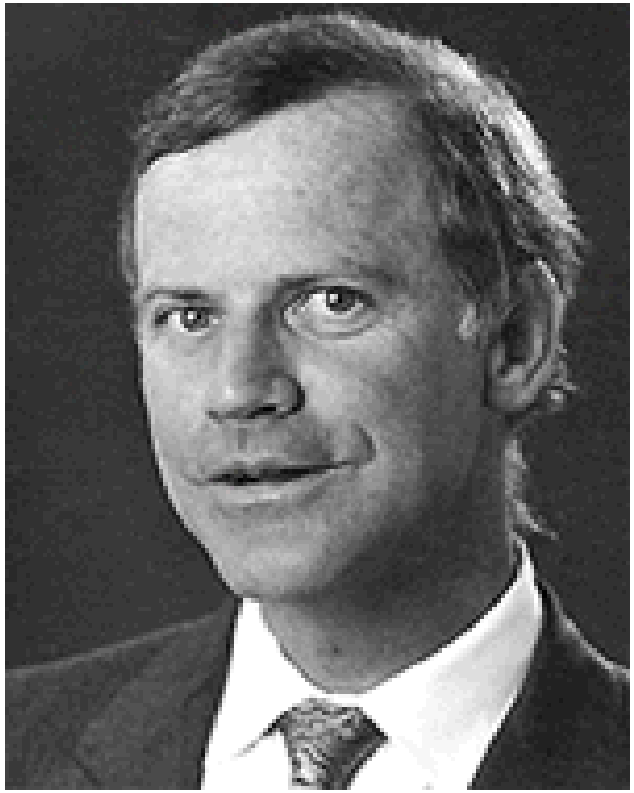


Polymerase chain reaction

PCR amplification

PCR amplification scheme





Kary B. Mullis

born 1944

La Jolla, CA, USA

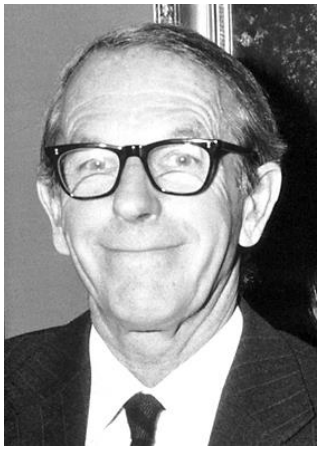
University of British Columbia

The Nobel Prize in Chemistry 1993

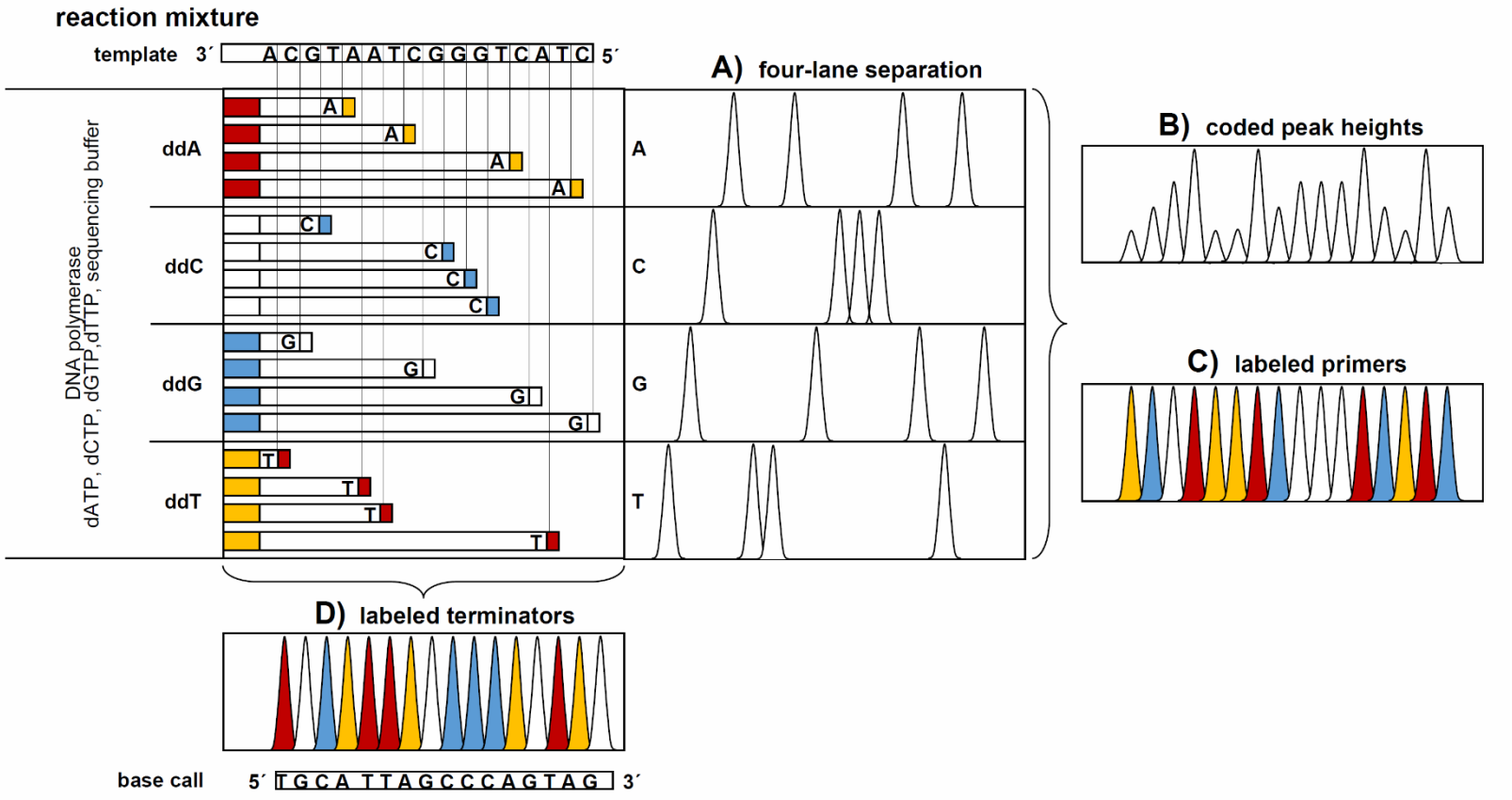
For his invention of the polymerase chain reaction (PCR) method

DNA sequencing

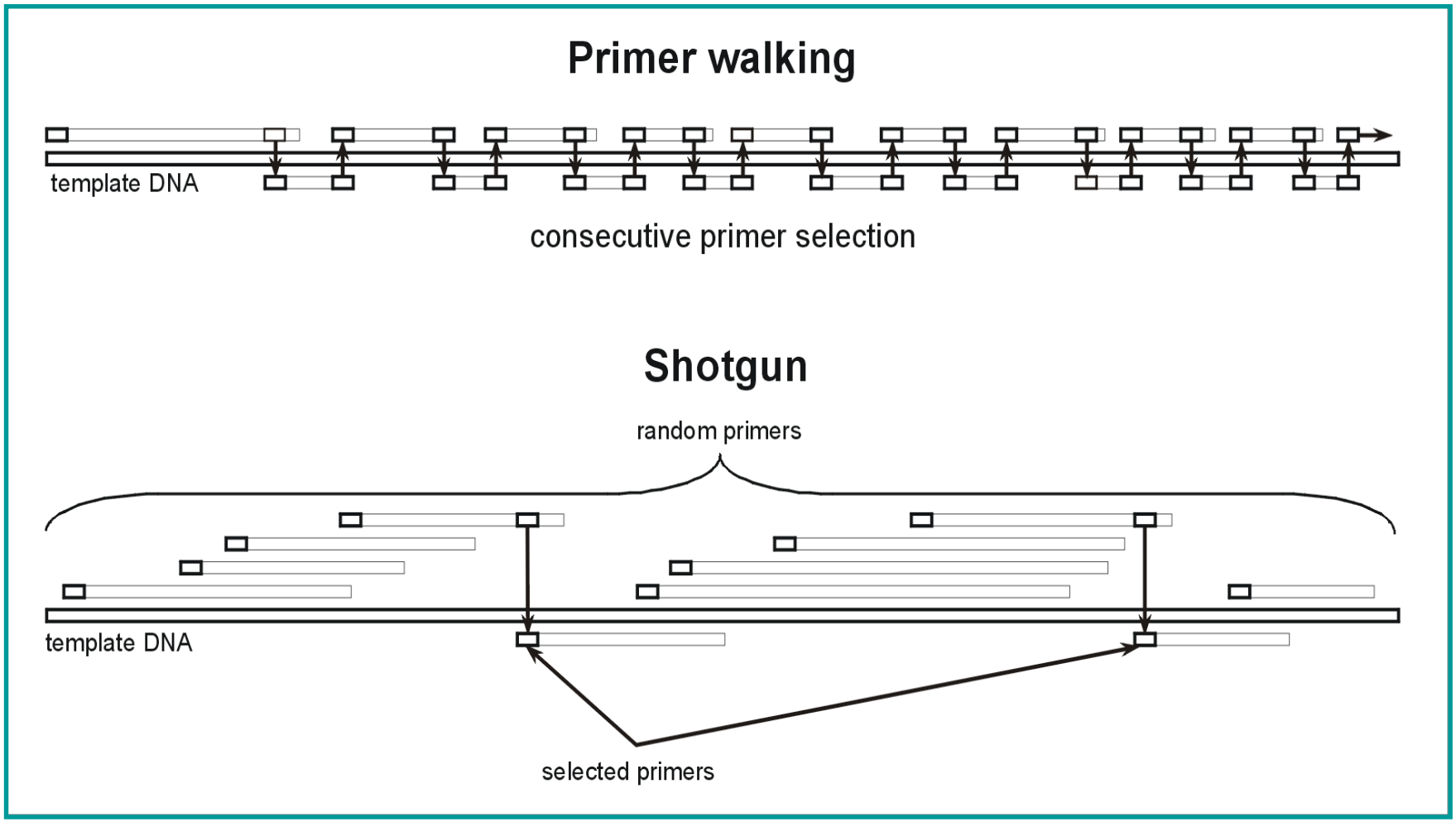
Synthesis of Sanger sequencing fragments



Frederick Sanger
MRC Laboratory of Mol. Biol.
Cambridge, UK
1918 – 2013
Nobel Price in Chemistry 1980



DNA sequencing strategy

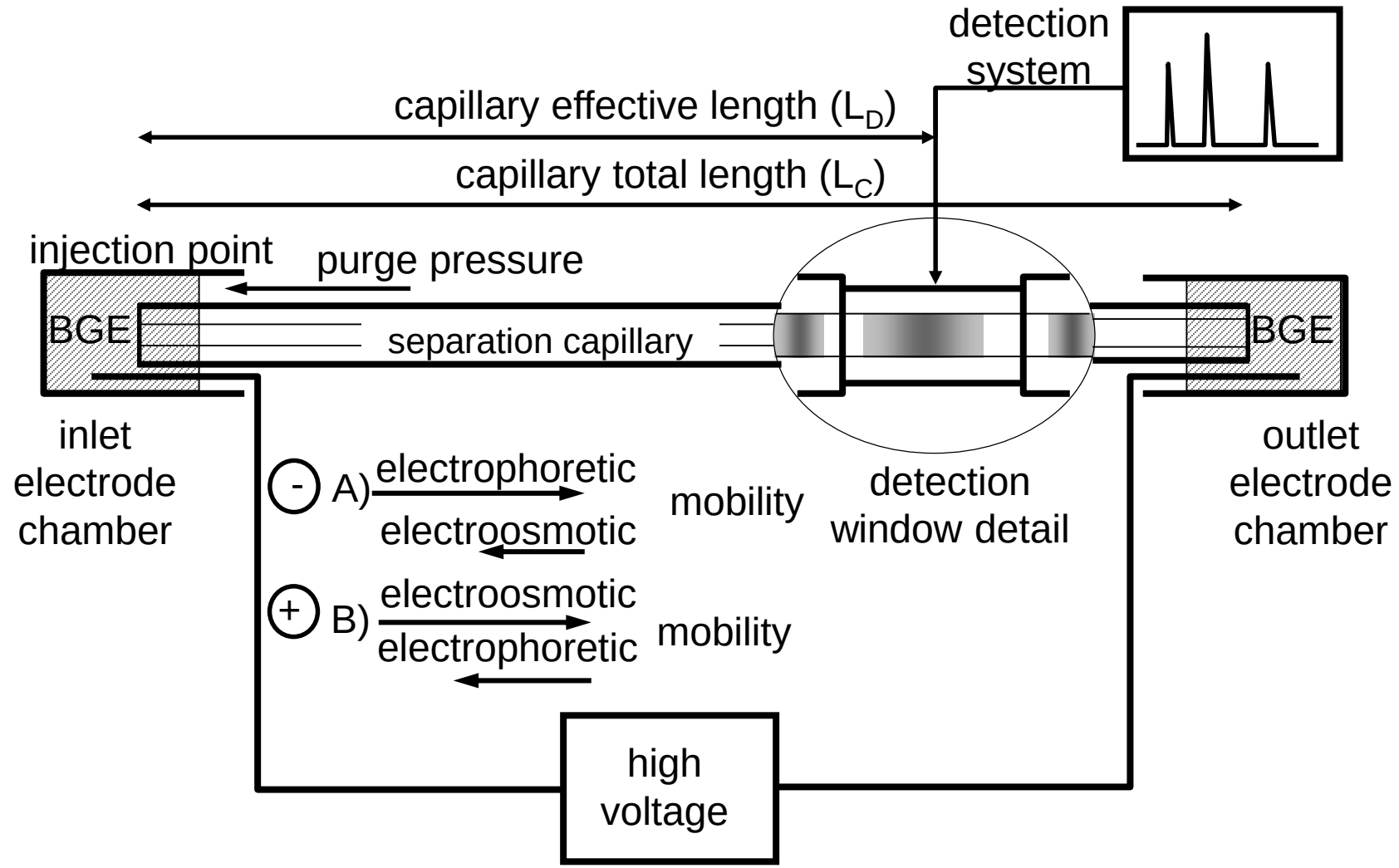


Separation methods

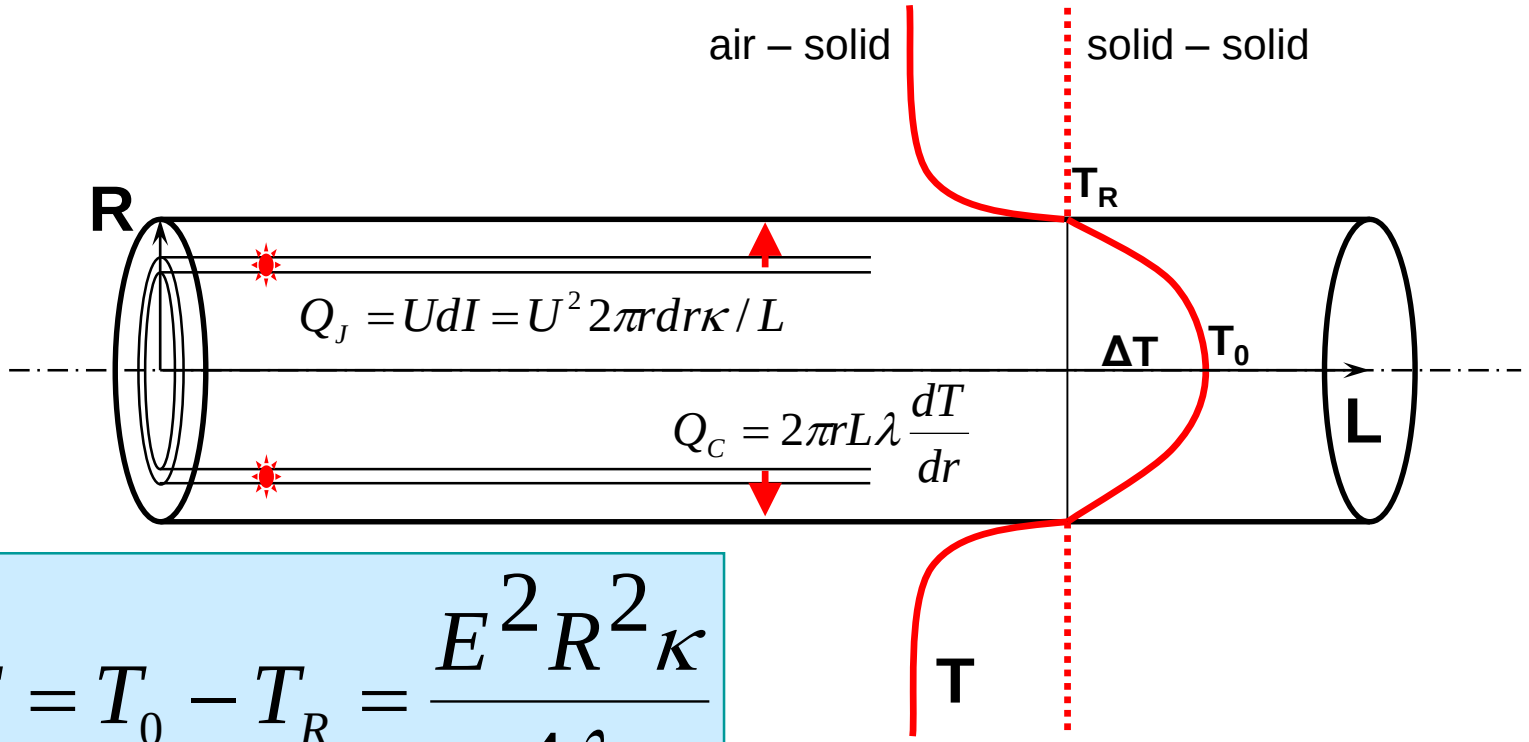
Capillary electrophoresis

CE

Capillary electrophoresis scheme



Why capillary electrophoresis?



$$\Delta T = T_0 - T_R = \frac{E^2 R^2 \kappa}{4\lambda}$$

Miniature capillary: low R => fast separation

1) high resistivity ➧ low current at high voltage ➧ low heat production

2) efficient heat transport ➧ low temperature difference inside the capillary

DNA electromigration

K. Klepárník, P. Boček, DNA diagnostics by Capillary Electrophoresis
Chemical Reviews 107, 5279 – 5317, 2007.

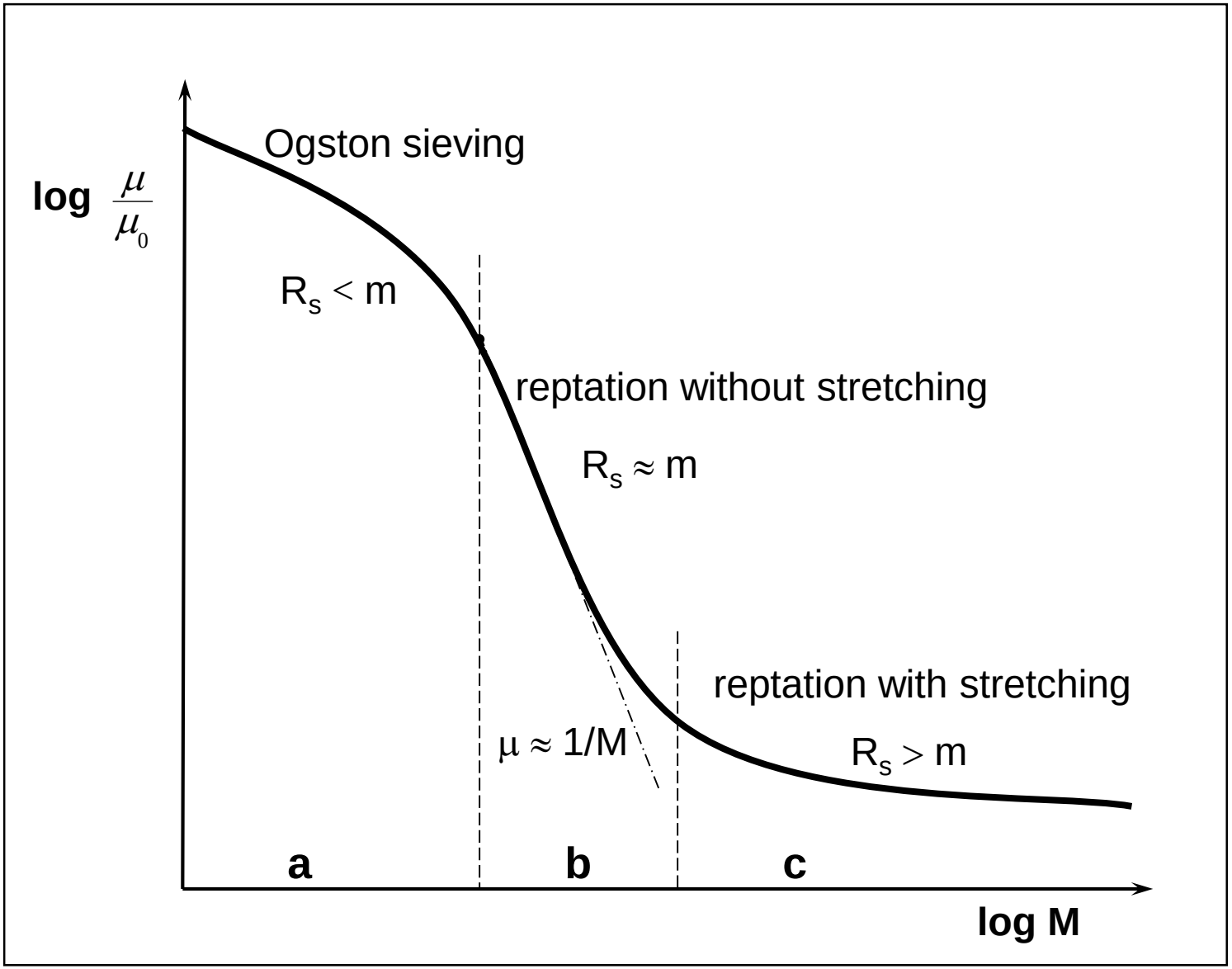
DNA electromigration regimes in sieving media

Size separations of homogeneous polyelectrolytes are impossible in free solutions

Short DNA fragments
Low concentration of media

Long DNA fragments
High concentration of media

Dependence of DNA electrophoretic mobility on molecular mass



Human Genome Project

J. CRAIG VENTER, Ph.D., PRESIDENT, CELERA GENOMICS
REMARKS AT THE HUMAN GENOME ANNOUNCEMENT
THE WHITE HOUSE
MONDAY, JUNE 26, 2000

Mr. President, Honorable members of the Cabinet, Honorable members of Congress, distinguished guests.
Today, June 26, 2000 marks an historic point in the 100,000-year record of humanity. We are announcing today that for the first time our species can read the chemical letters of its genetic code. At 12:30 p.m. today, in a joint press conference with the public genome effort, Celera Genomics will describe the first assembly of the human genetic code from the whole genome shotgun sequencing method. Starting only nine months ago on September 8, 1999, eighteen miles from the White House, a small team of scientists headed by myself, Hamilton O. Smith, Mark Adams, Gene Myers and Granger Sutton began sequencing the DNA of the human genome using a novel method pioneered by essentially the same team five years earlier at The Institute for Genomic Research in Rockville, Maryland. The method used by Celera has determined the genetic code of five individuals....

…There would be no announcement today, if it were not for the more than **$1 billion that PE Biosystems invested in Celera and in the development of the automated DNA sequencer** that both Celera and the public effort used to sequence the genome…

J. Craig Venter

The Institute for Genomic Research (**TIGR**)

The first president of **Celera Genomics**

The completed sequence of the human genome was published in February 2001 in Science.

Venter, C. J. et al. *Science* **2001**, *291*, 1304-1351.

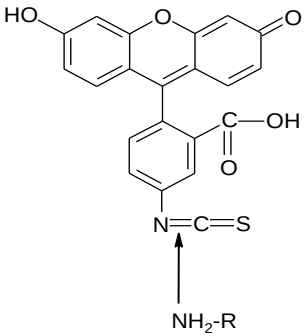
Fluorescence chemistry

Lloyd M. Smith

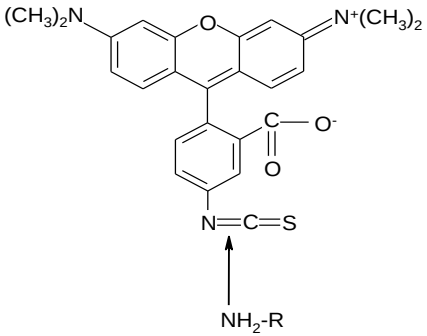
Born 1954
A.B. 1976, University of California - Berkeley
Ph.D. 1981, Stanford University
University of Wisconsin - Madison

Smith, L. M., Sanders, J. Z., Kaiser, R. J., Hughes, P., Dodd, C., Connell, C. R., Heiner, C., Kent, S. B. H. and Hood, L. E. Fluorescence detection in automated DNA sequence analysis *Nature, 321*, 674-679, **1986**.

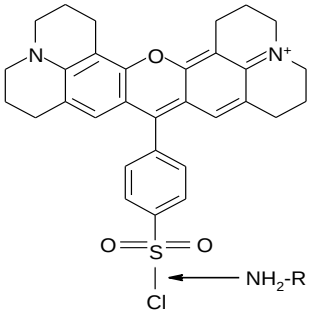
Fluorescent lebels



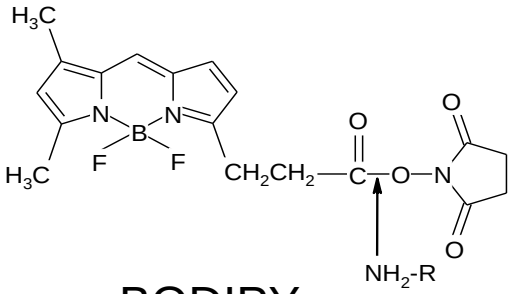
Fluorescein



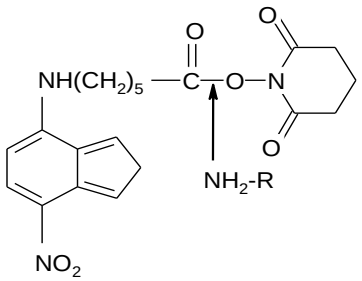
Rhodamine



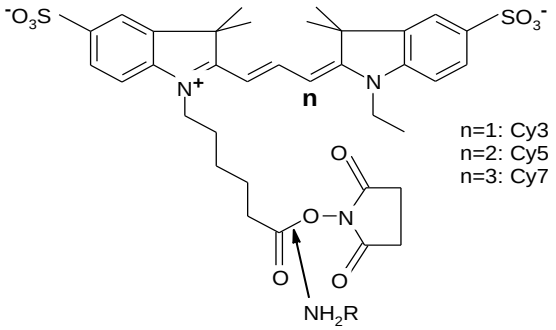
Texas Red



BODIPY



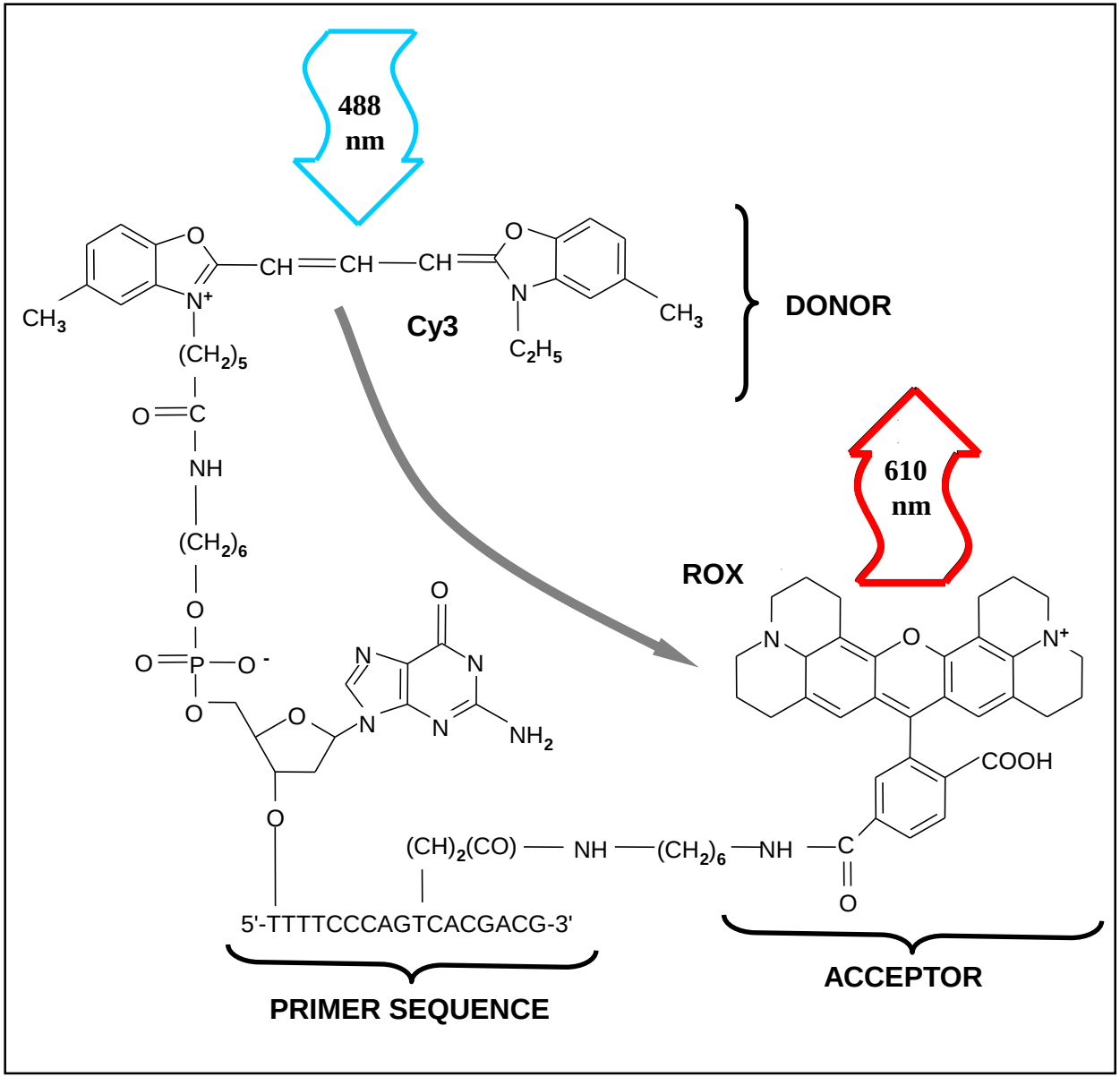
NBD



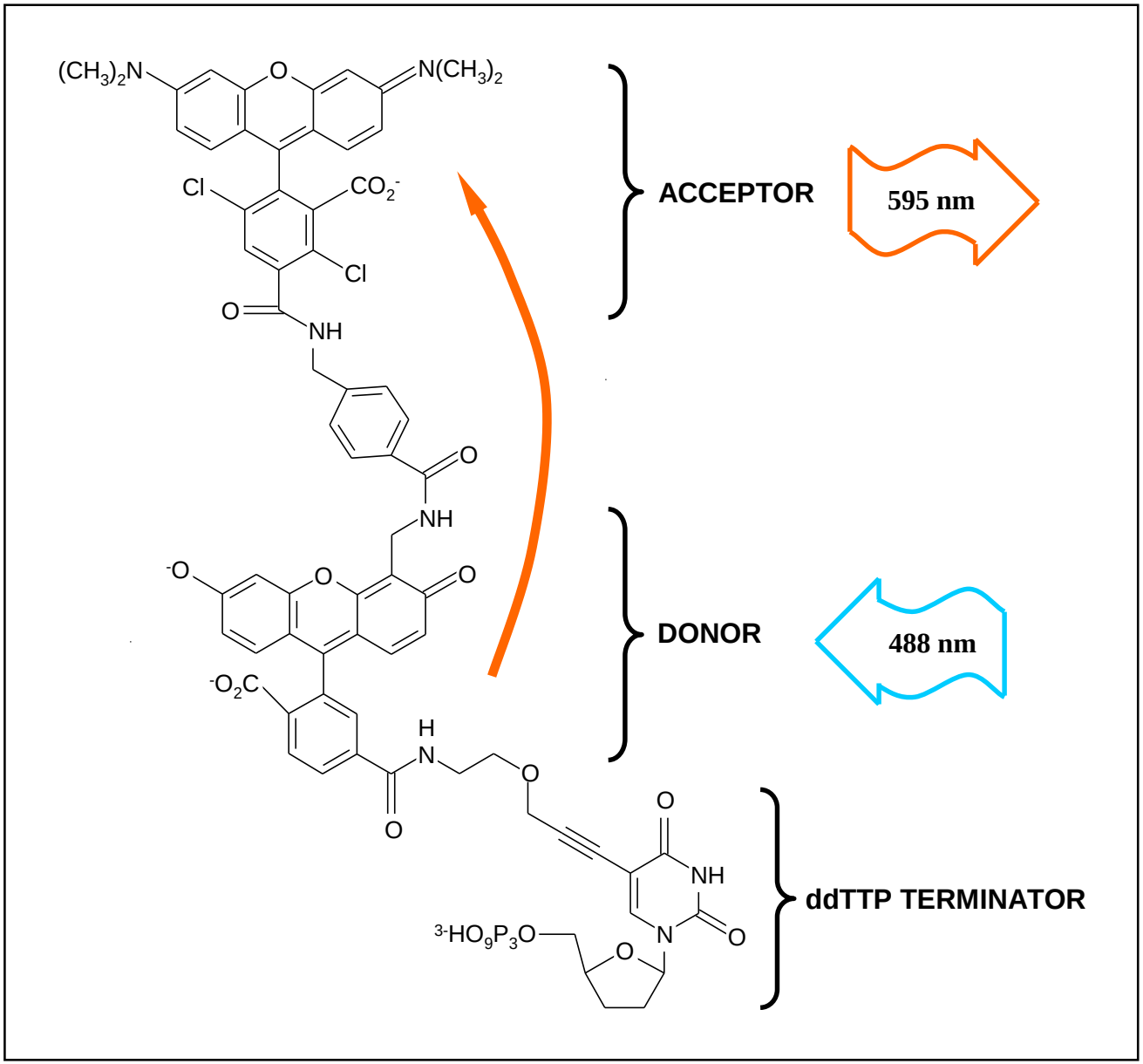
Cy3,5,7

Sequencing primer attached to Fluorescence Resonance Energy Transfer

Prof. Richard A. Mathies
University of California at Berkeley
Department of Chemistry
Berkeley, CA



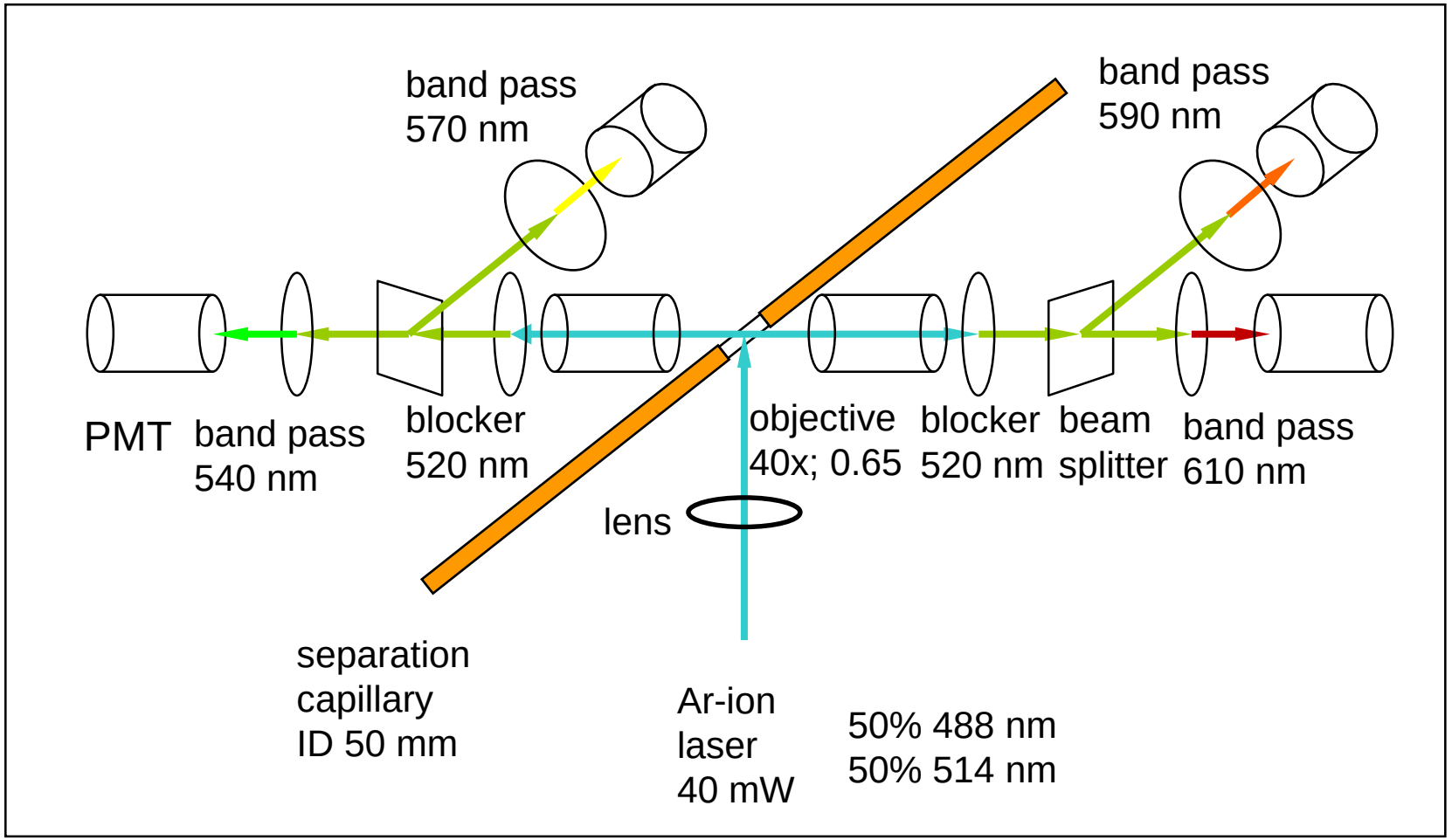
Dideoxy terminator attached to Fluorescence Resonance Energy Transfer



LIF detection

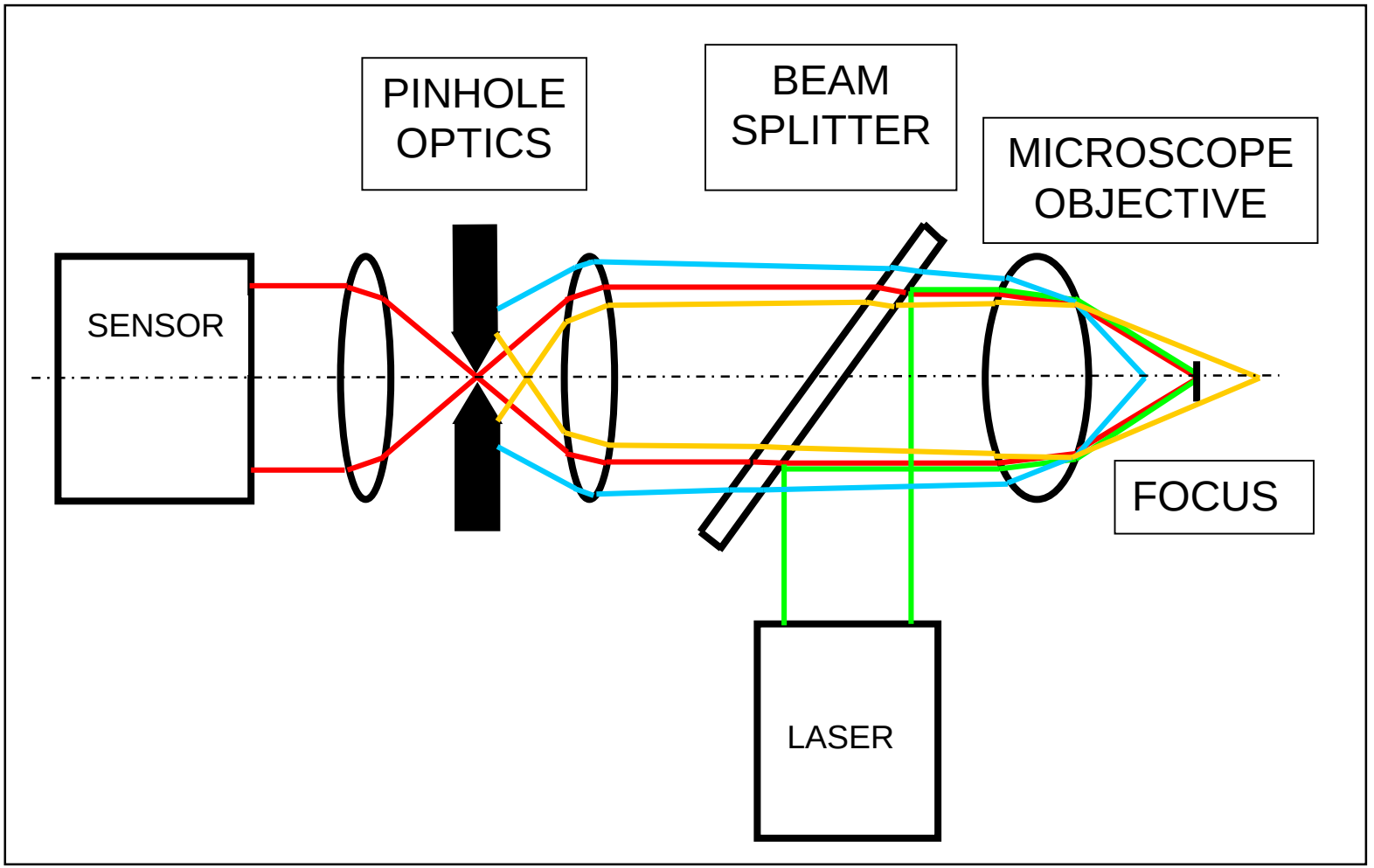
Spectral filtering

Four channel LIF detection arrangement



Space filtering

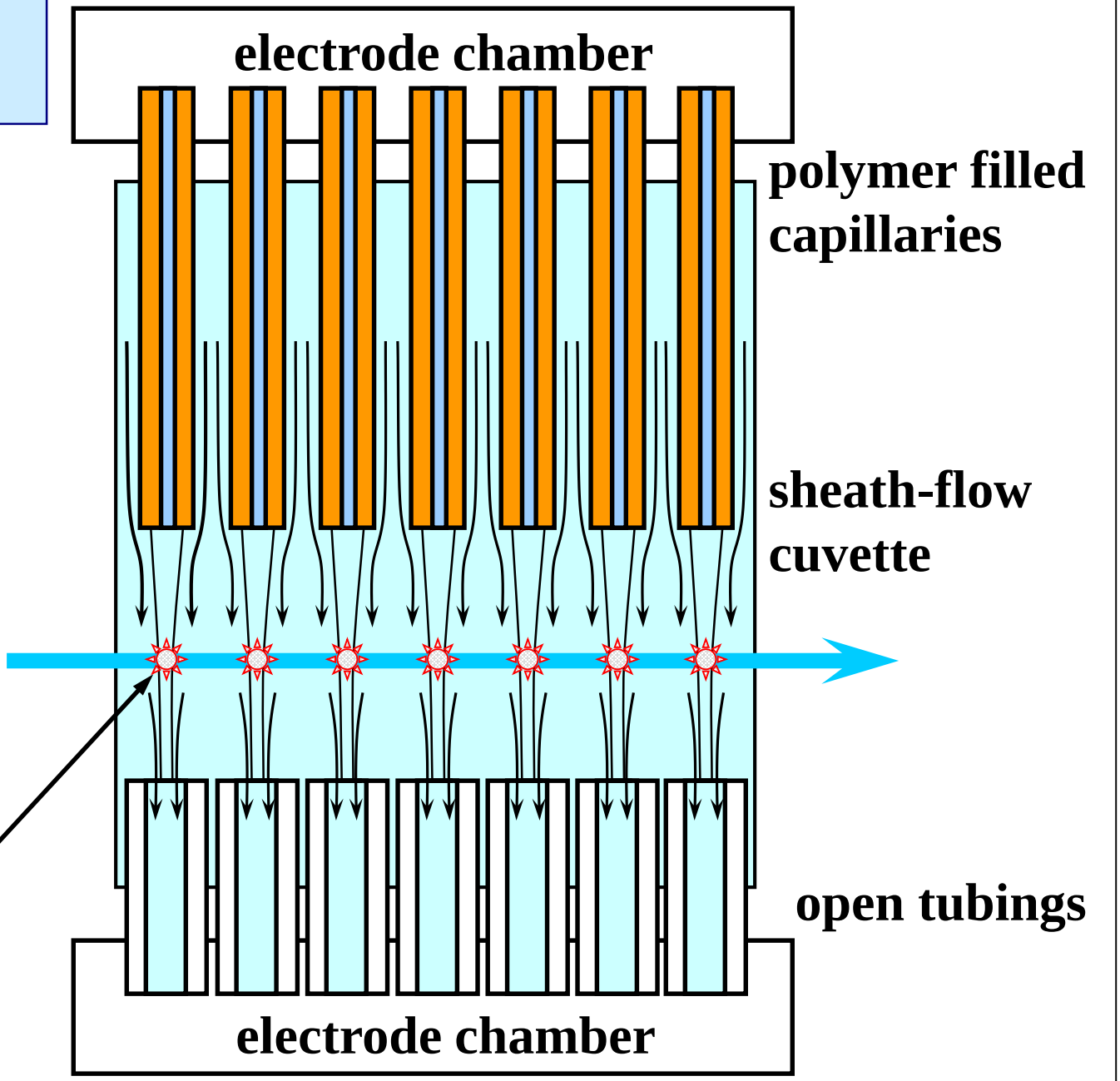
SCHEME OF CONFOCAL DETECTOR



Prof. Edward S. Yeung
Ames Laboratory
U.S. Department of Energy
Iowa State University.

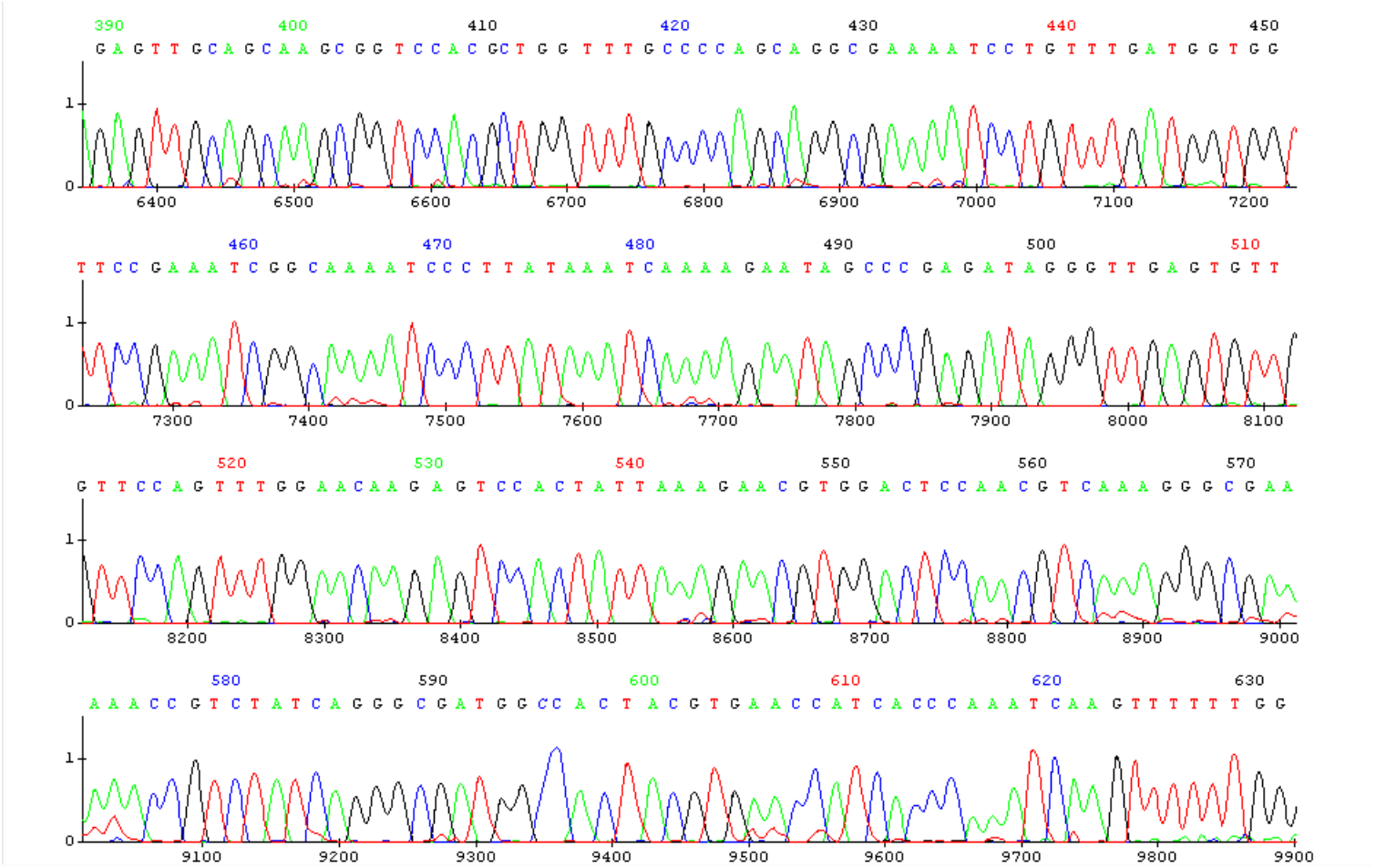
Sheath-flow cuvette

Prof. Norman Dovichi
University of Notre Dam
Indiana, USA



Prof. Hideki Kambara
Hitachi Central Research Laboratory
Tokyo, Japan

DNA sequencing record

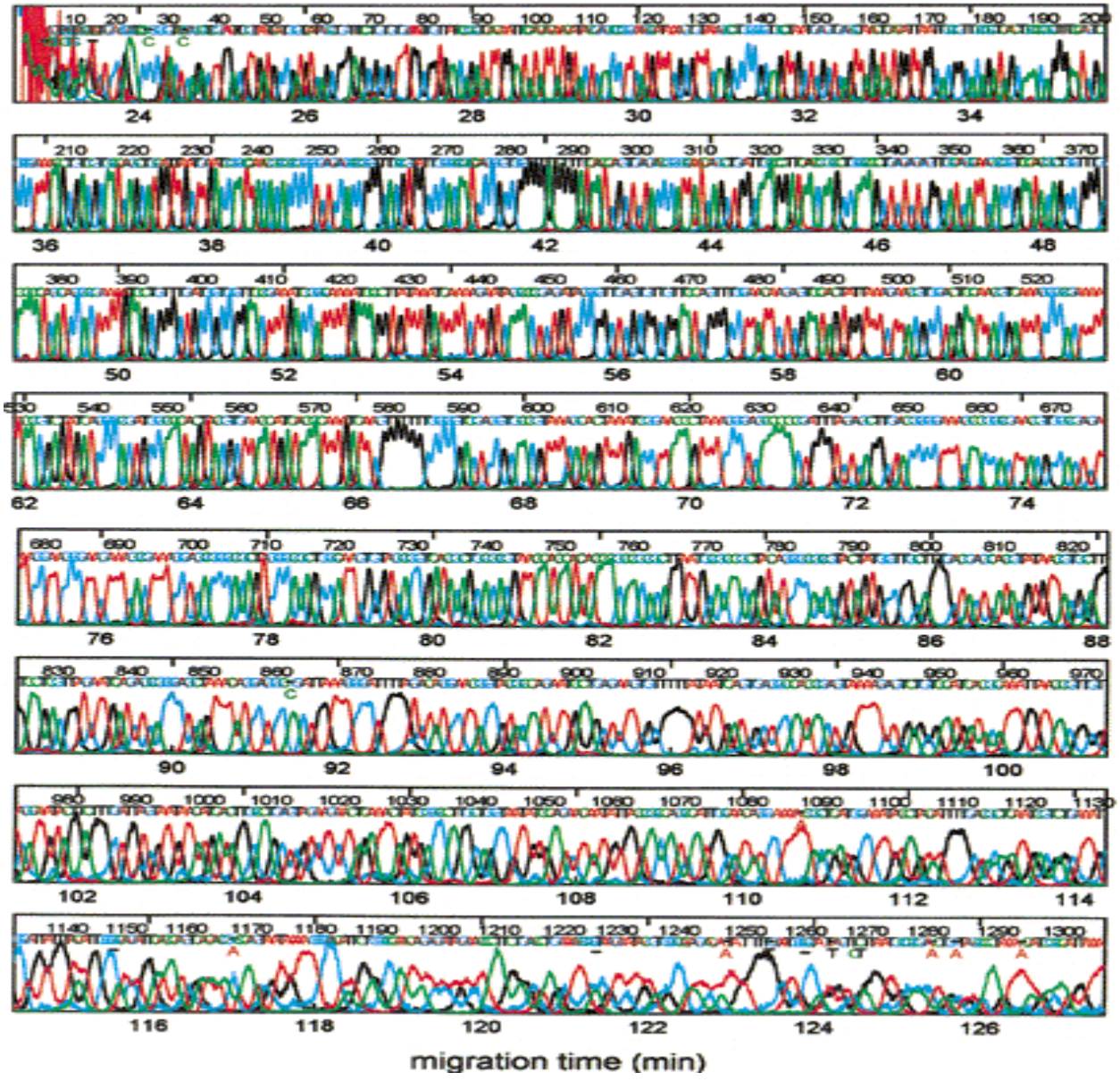


DNA sequencing up to 1300 bases in 2 hours

Separation matrix: LPA 2.0% (w/w) 17 MDa, 0.5% (w/w) 270 kDa

E: 125 V/cm, T: 70 °C

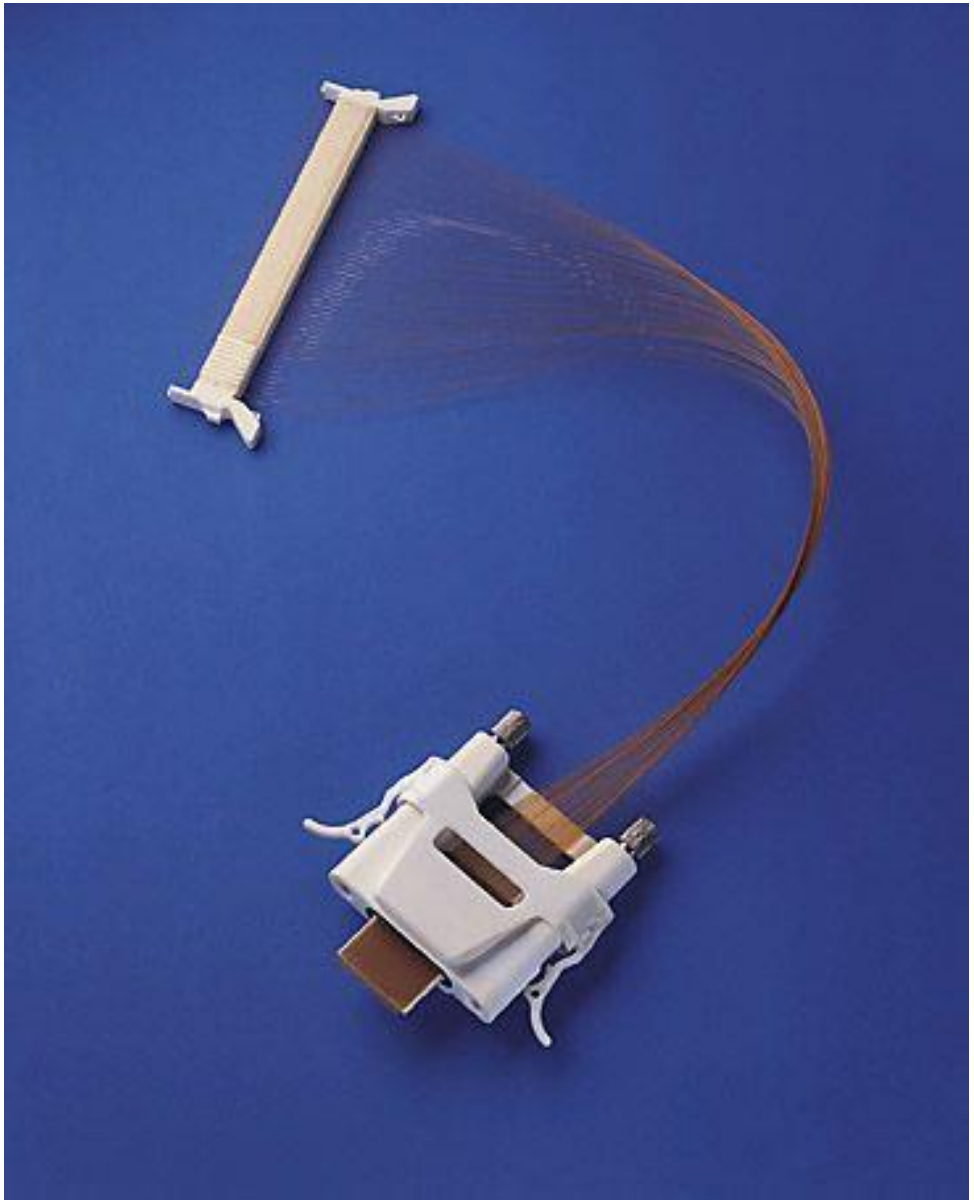
Barry L. Karger
The Barnett Institute
Northeastern University
Boston MA



ABI PRISM® 3700 DNA Analyzer

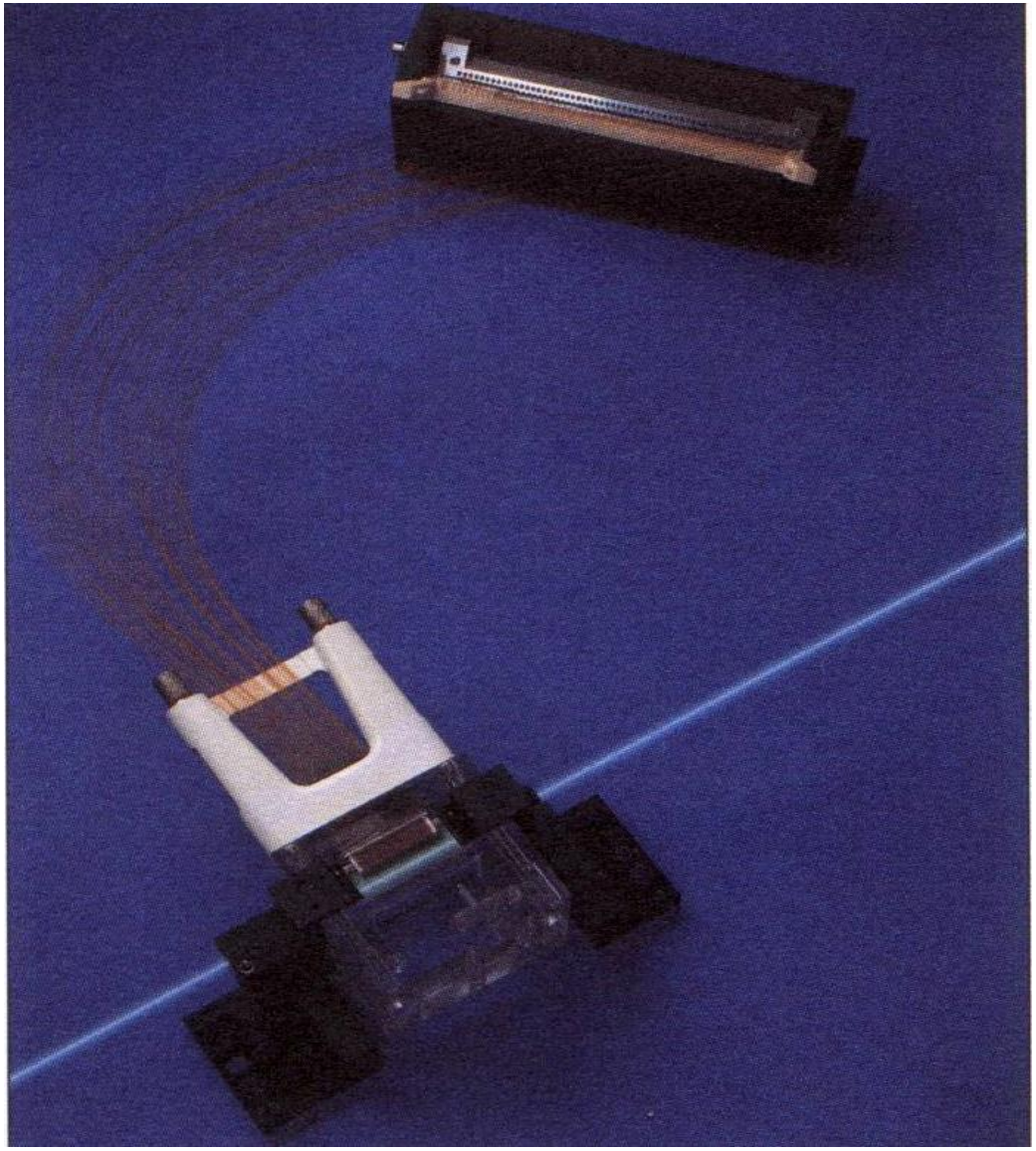
96 active

eight reserve capillaries

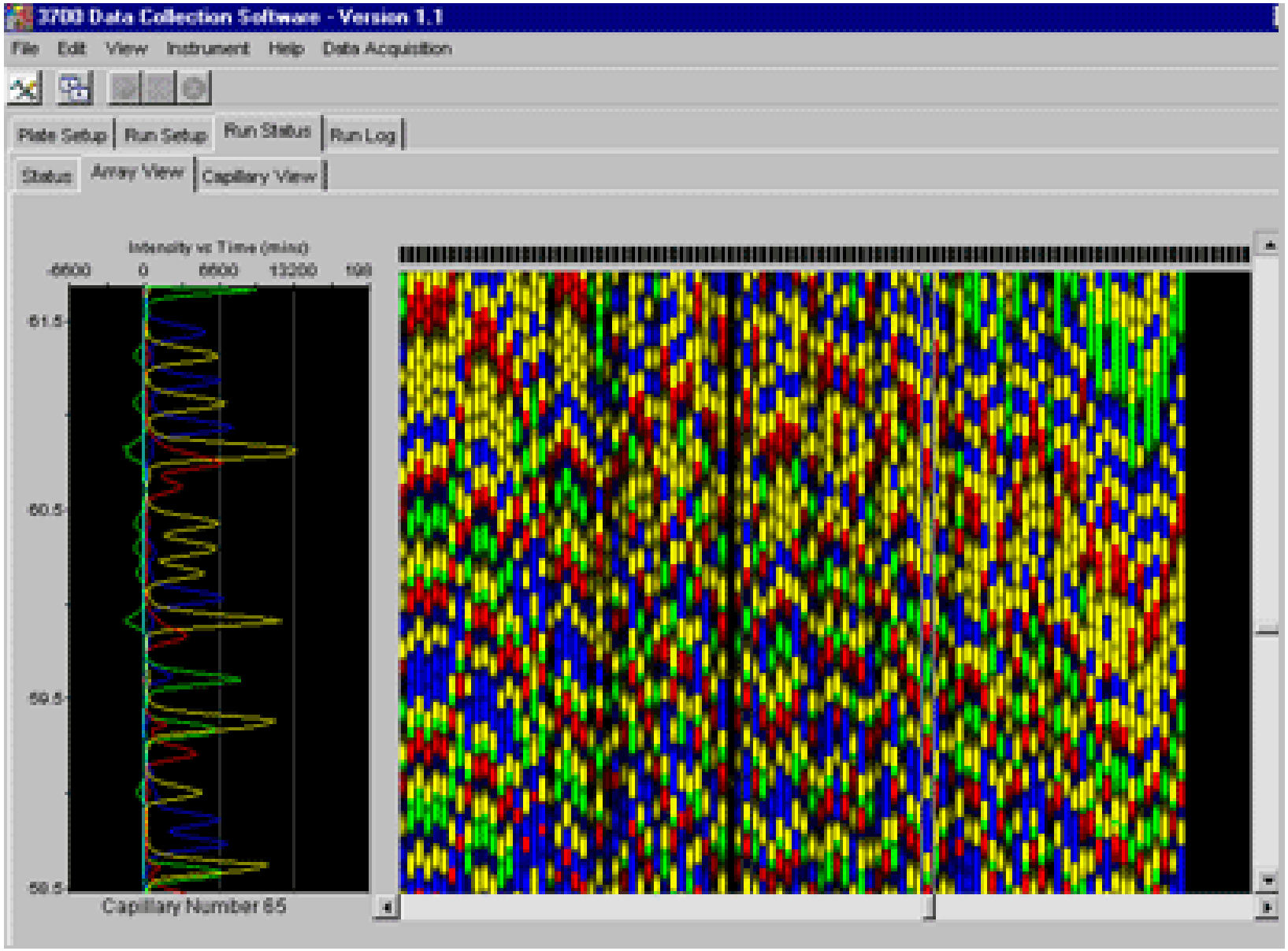


ABI PRISM® 3700 DNA Analyzer

Sheath flow cuvette



ABI PRISM® 3700 DNA Analyzer



PE Applied Biosystems

ABI PRISM 3700

accuracy > 98.5% to 550 base
96 samples per run in 3 hours
laser Ar-ion 488 and 514.5 nm
detection in sheath flow
concave spectrograph and cooled CCD

Molecular Dynamics

MEGABACE 1000

accuracy > 98.5% to 550 base
96 samples per run in 2 hours
laser Ar-ion 488 nm
energy transfer dyes
confocal scanning with 4 filters and 2 PMTs

DNA mutation analysis

Restriction (amplification) fragment legth polymorphism

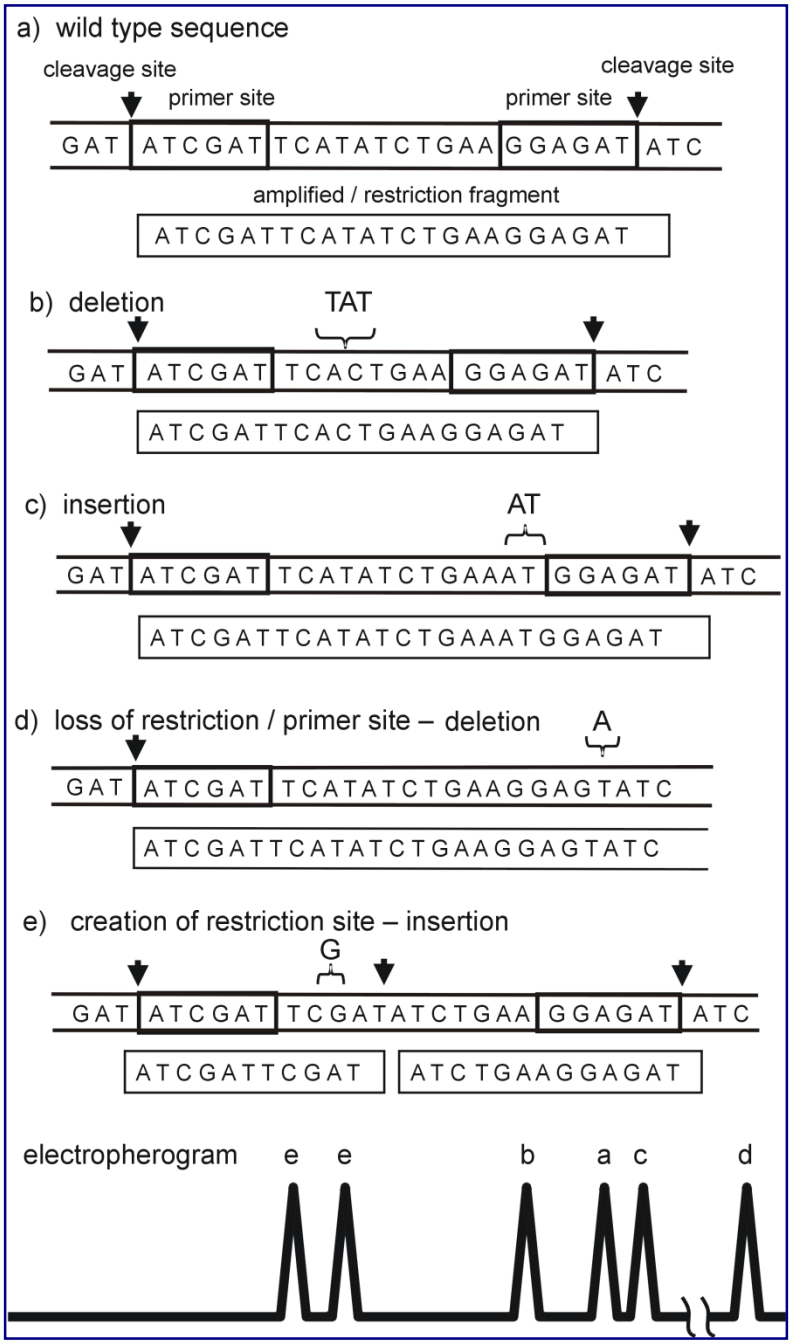
RFLP (AFLP)

Size based separation of ds or ss DNA fragments

Resolution:

ss > 1000

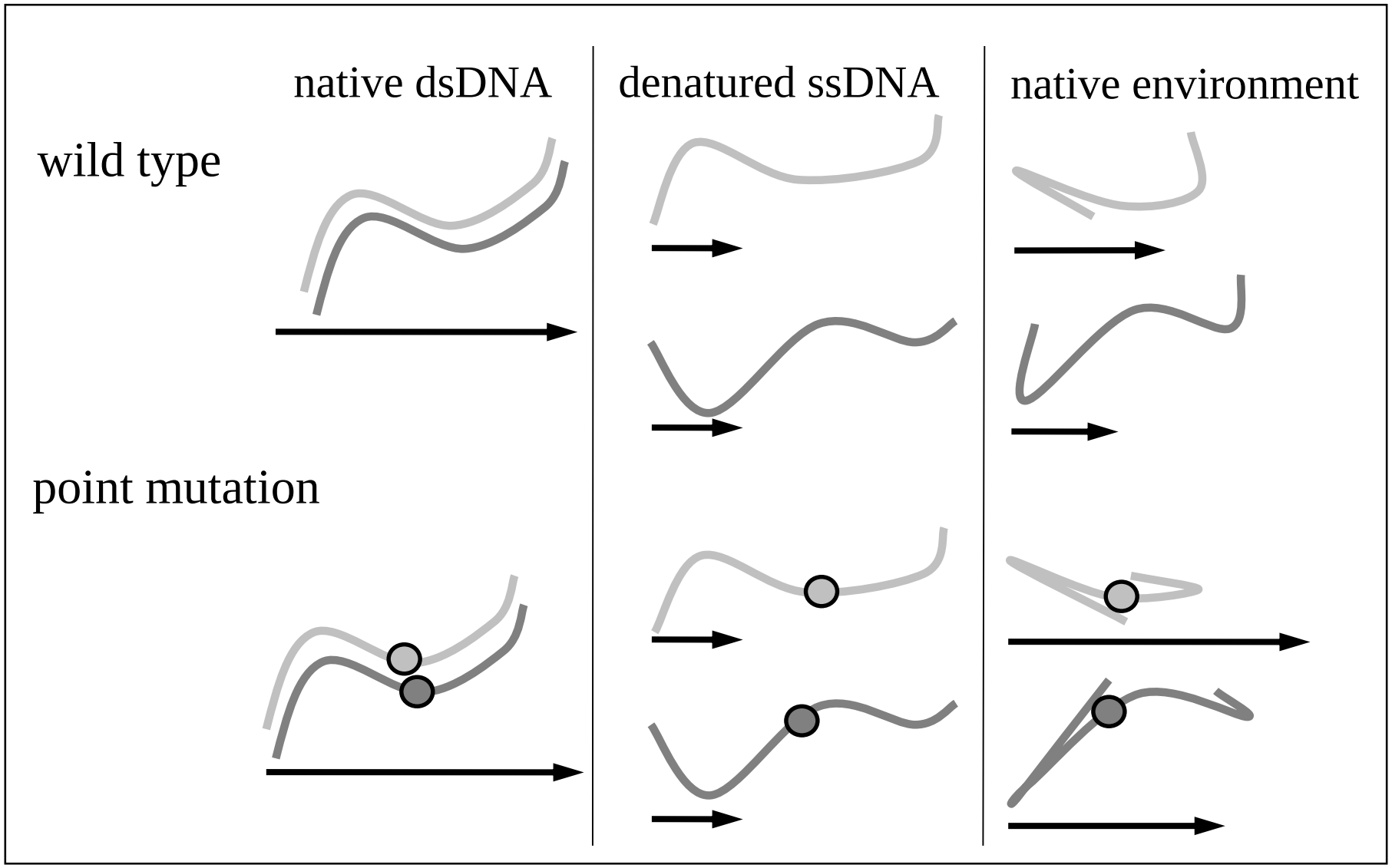
ds > 400



Single Strand Conformation Polymorphism

SSCP

Principle of SSCP technique

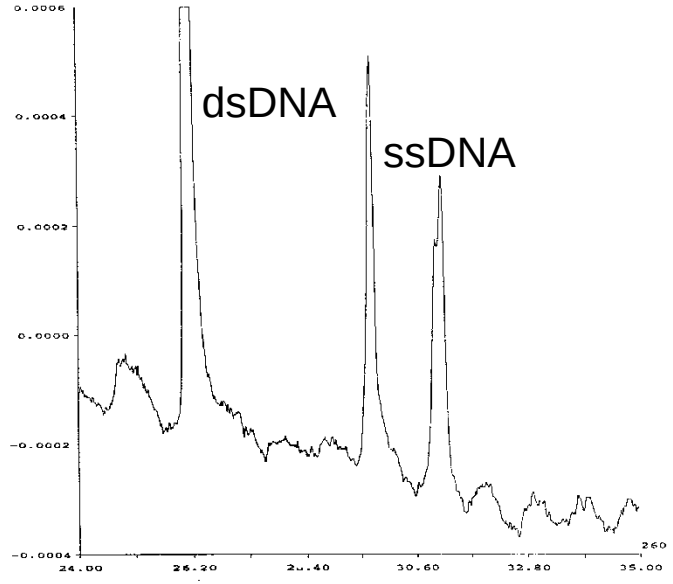


Phenylketonuria

SSCP analysis
Detection of point mutation C > T in phenylalanine hydroxylase gene on chromosome 12

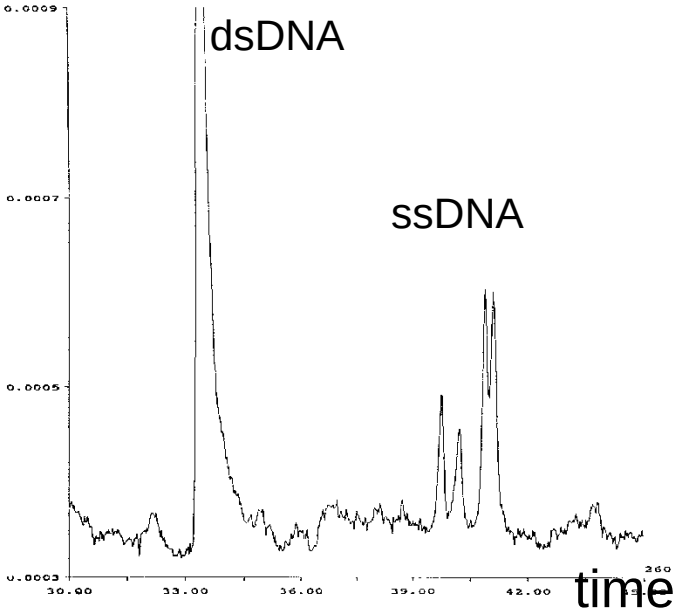
Separation conditions:
2% solution of agarose SeaPrep in 1xTBE with 10% formamide
T - 30 °C
LC - 55 cm
LD - 50 cm
E – a) 183 V/cm, b) 135 V/cm.

a) health homozygote



relative absorbance at 260 nm

b) heterozygote

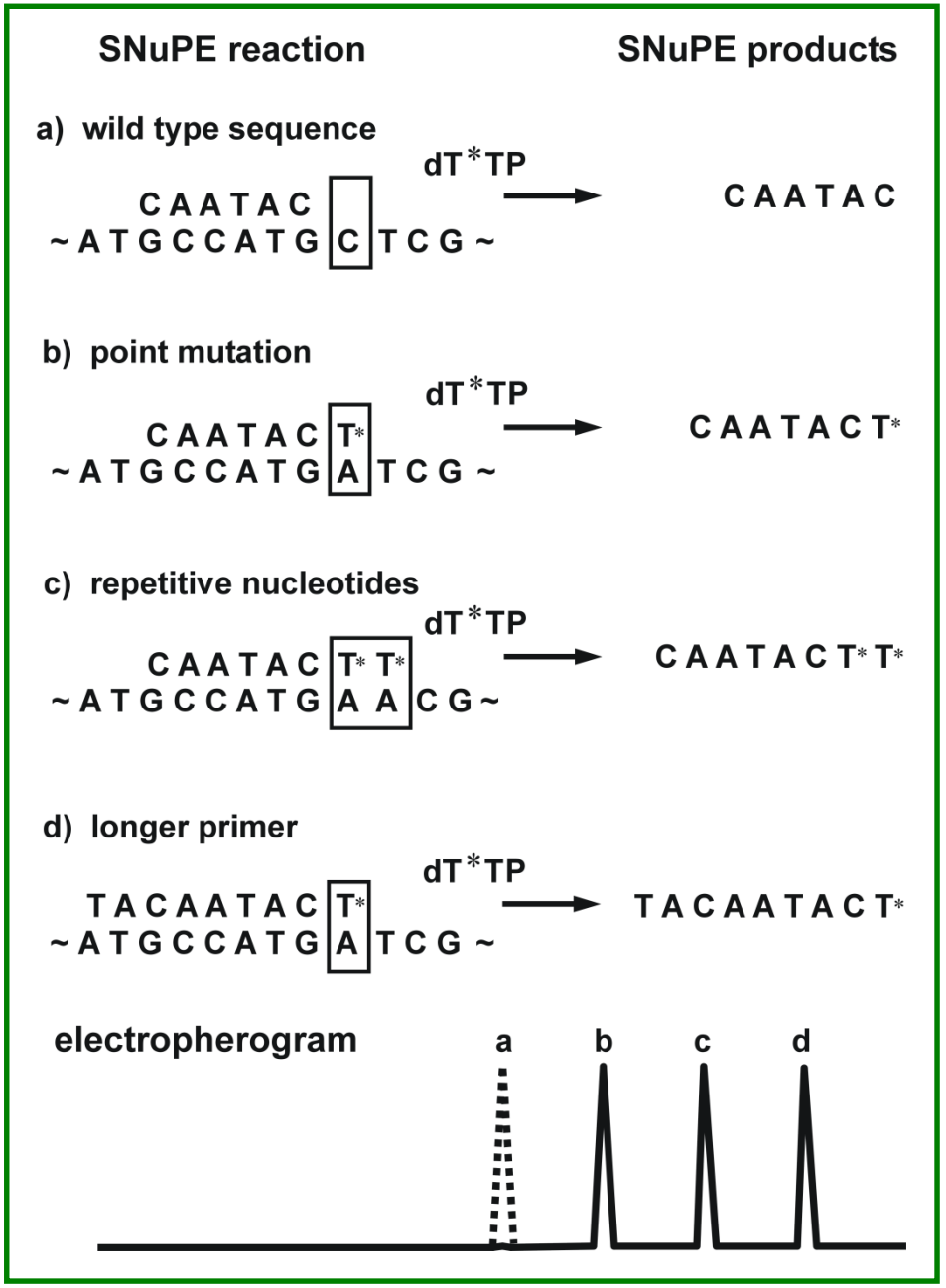


time

Single nucleotide primer extension

Minisequencing

SNuPE

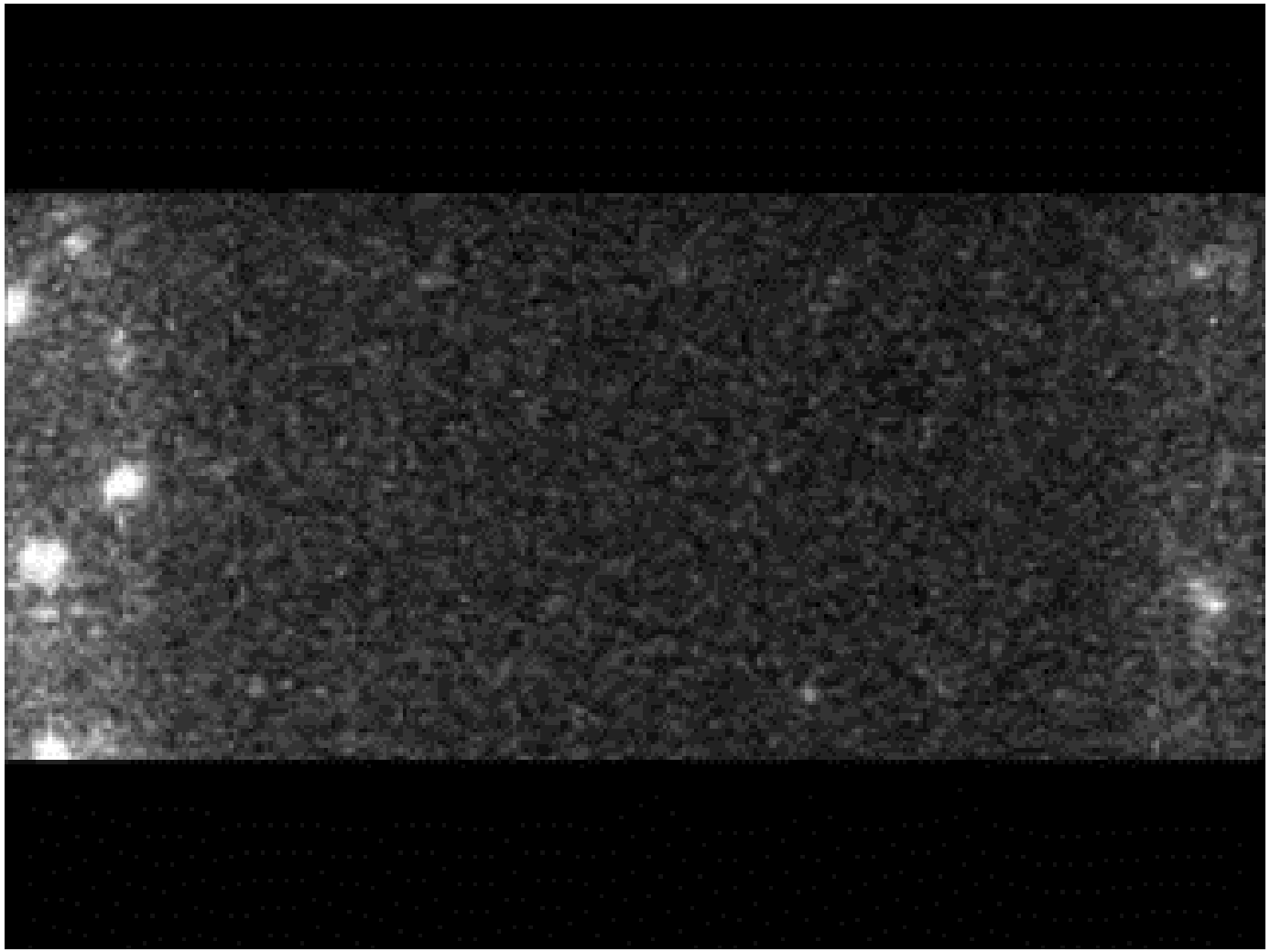


Next generation sequencing

Single molecule detection

Stretching of dsDNA in Nanochannels

- evaluation of size
- chromatography or electrophoresis
- detection of nucleotides consecutively cleaved by exonuclease



Single molecule reaction monitoring

Parallel single molecule sequencing by synthesis

Helicos

The HeliScope™ Sequencer

2 . 10^9 b/day

10^9 reads/run
25 – 55 bp read lengths

454 LIFE SCIENCES

Genome Sequencer FLX System

3 . 10^8 b/day

100 Mb/7.5 hour run
400 000 reads/7.5 hour
200 – 300 bp read lengths

Solexa

Illumina Genome Analyzer

6 . 10^8 b / day

3 . 10^9 b / 5 days run
50 . 10^6 oligo clusters
36 – 50 bp read lengths

Photocleavable dideoxy nucleotides

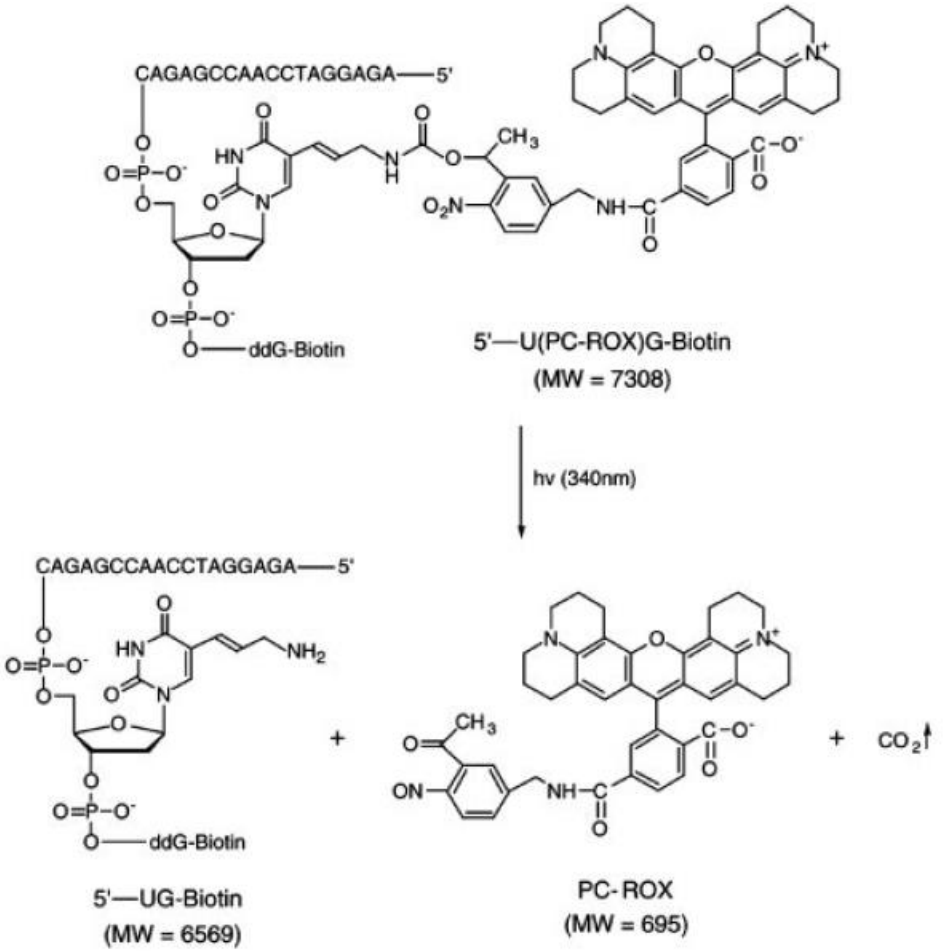


Fig. 2. DNA product 5′-U(PC-ROX)G-biotin, formed by incorporating a dUTP-PC-ROX into a primer in a polymerase reaction and its photocleavage, producing DNA fragment 5′-UG-biotin and PC-ROX. MW, molecular weight.

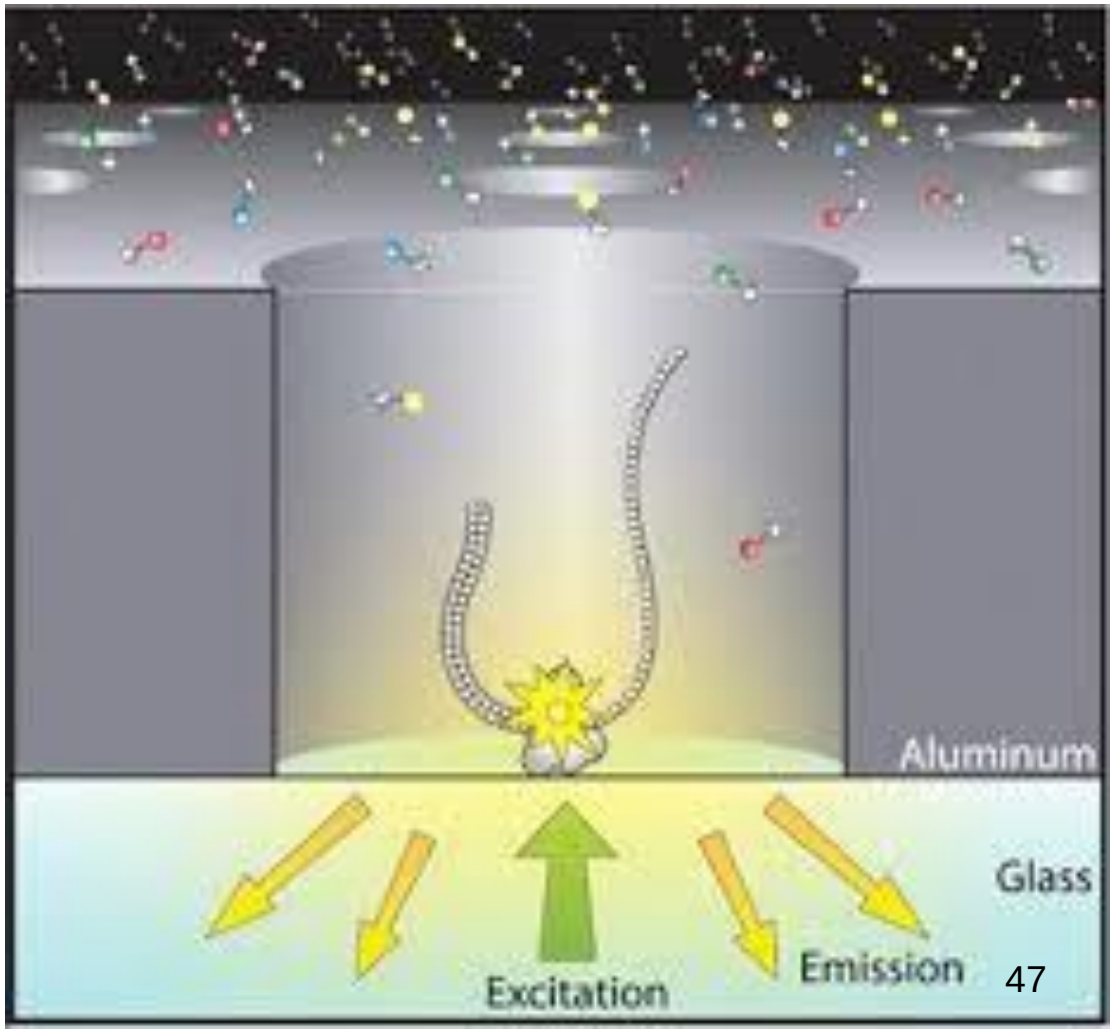
Next generation DNA sequencing

Single molecule real time sequencing (SMRT™)

Pacific Biosciences

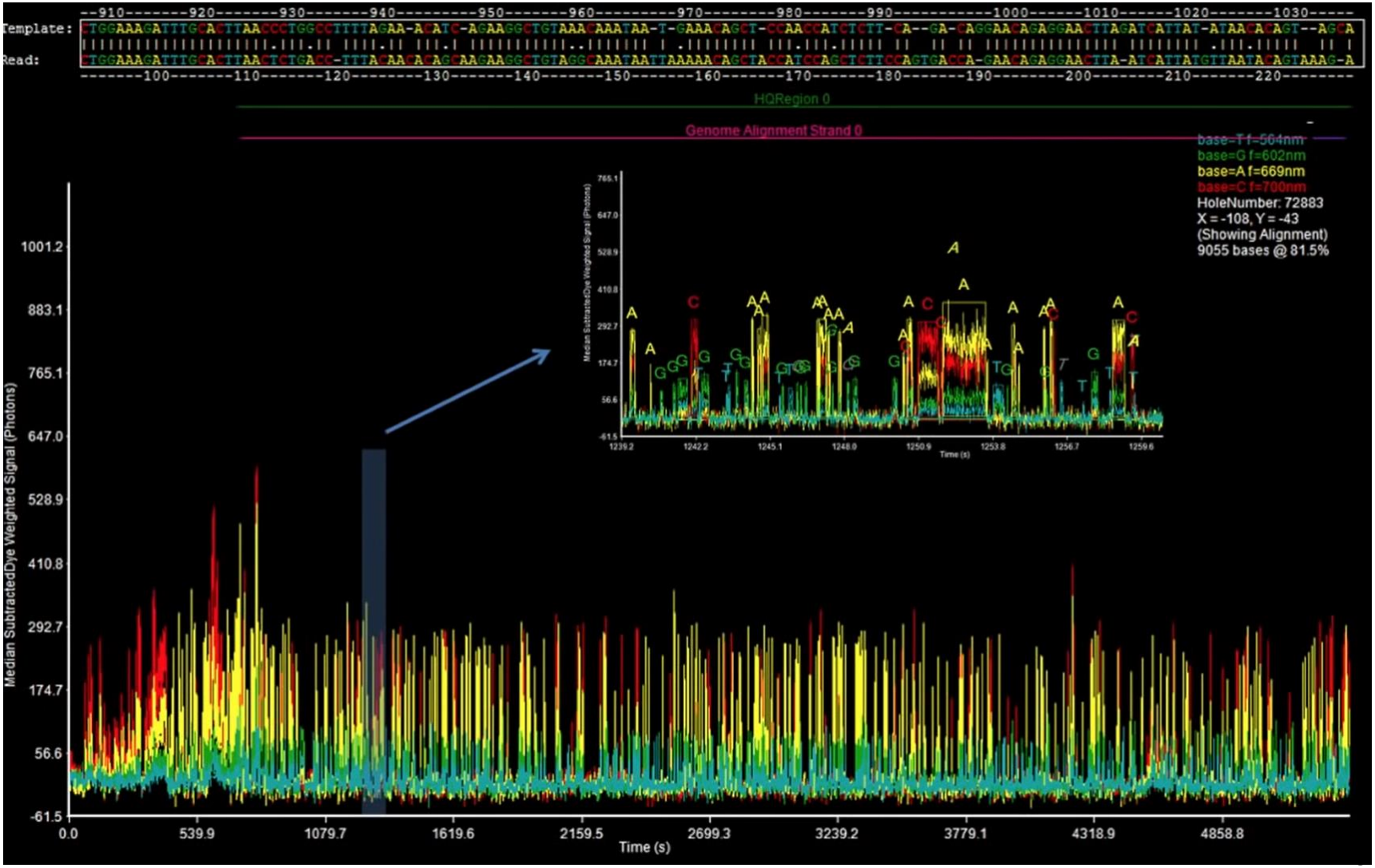
DNA sequencing – DNA polymerase
RNA sequencing – reverse transcriptase
Codone-resolved translation elongation by single ribosomes

Tens of nucleotide peaks in 1 sec
Read length 1 – 15 kb
80 000 detection points
15 min/genome: 50 n/s * 80 000 points * 15 min * 60 s = 3.6 Gb
DNA polymerase 529 processivity 20 kB – 400 b/s
Some enzymes are not processive
$ 100/genome



PacBio RS *instrument*

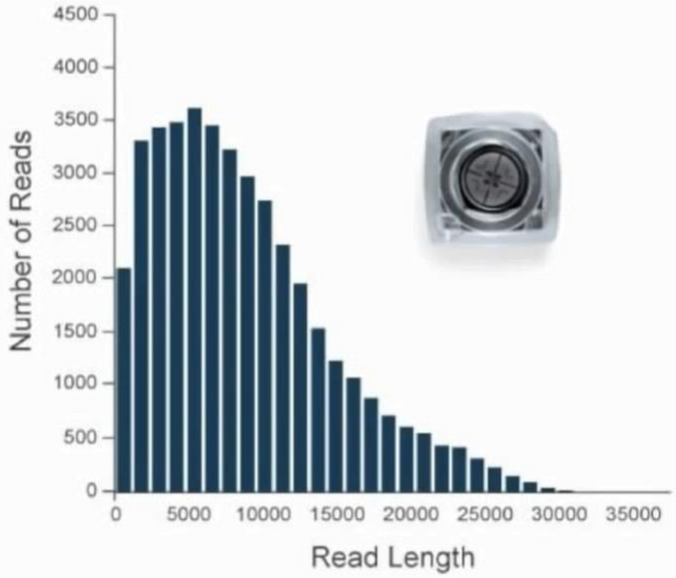
Single molecule real time sequencing



Pacific Biosciences Read Length

New P5-C3 Sequencing Chemistry

Read Length Distribution



Typical Results

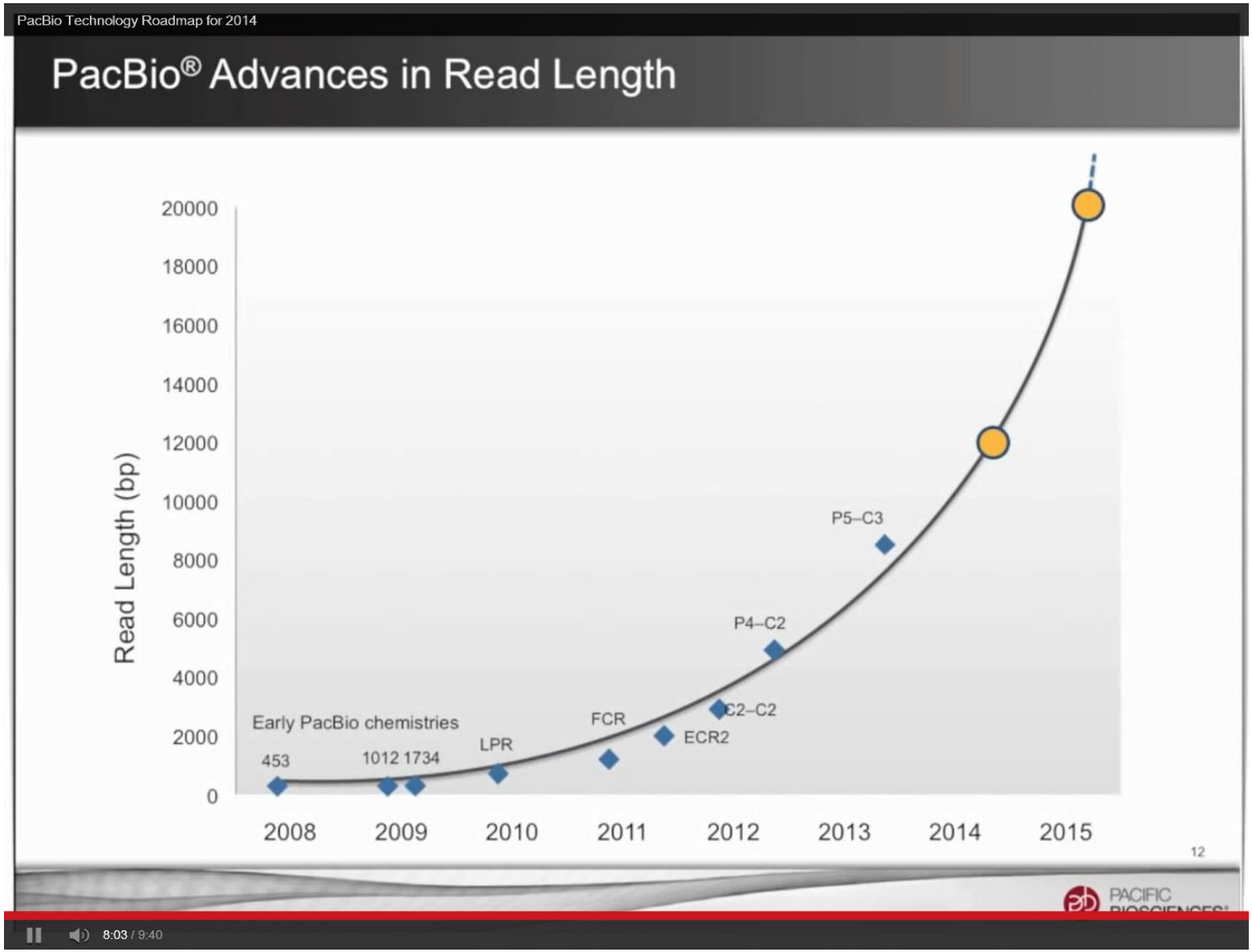
Read Length:

Average:	~ 8.5 kb
Maximum:	> 30 kb
Top 5% of reads:	> 18 kb
Half of data in reads:	> 10 kb
Data per SMRT® Cell:	~ 375 Mb

Based on data from a 20 kb size-selected E. coli library using a 180-minute movie.
Each SMRT Cell yields ~ 50,000 reads.

PACIFIC BIOSCIENCES®

Pacific Biosciences Read Length



Pacific Biosciences

Single molecule real time sequencing SMRT™

www.pacificbiosciences.com

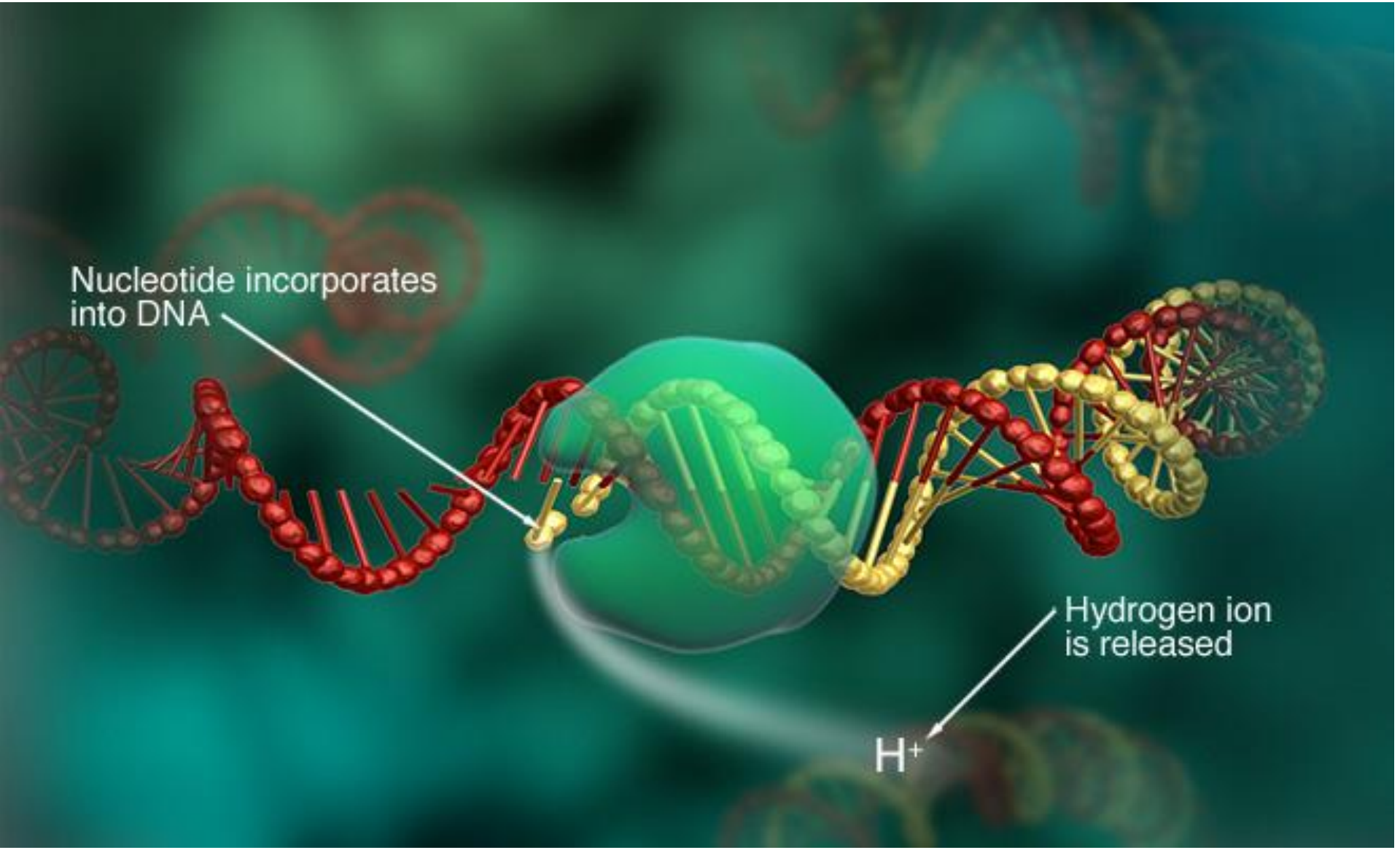
DNA sequencing – DNA polymerase
RNA sequencing – reverse transcriptase
Codone-resolved translation elongation by single ribosomes
Tens of nucleotide peaks in 1 sec
Read length 1 – 30 kb
80 000 detection points
15 min/genome: 50 n/s * 80 000 points * 15 min * 60 s = 3.6 Gb
DNA polymerase 529 processivity 20 kB – 400 b/s
Some enzymes are not processive
$ 100/genome

Ion Torrent
The Ion Personal Genome Machine (PGM™) sequencer

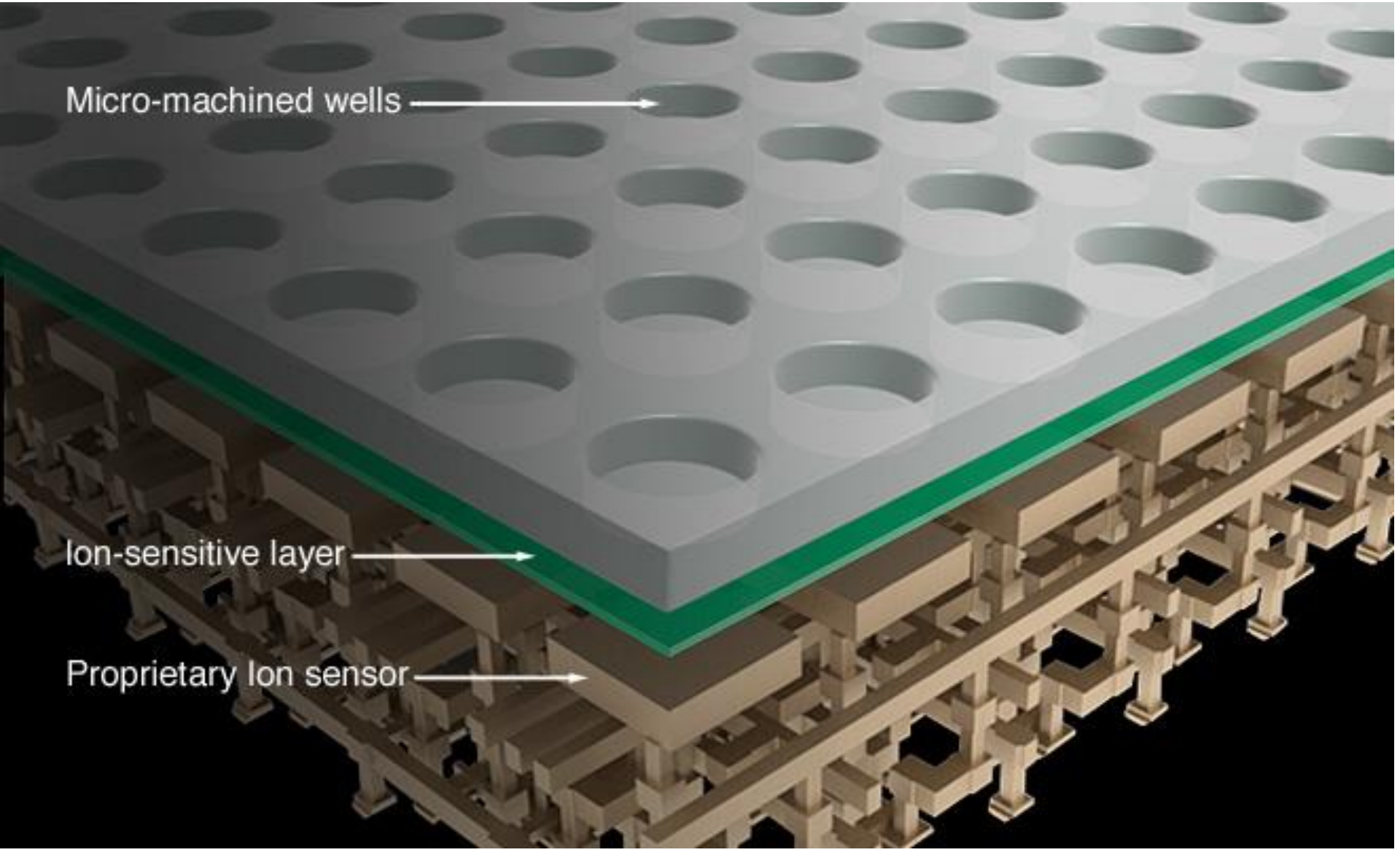
http://www.iontorrent.com/

- Different templates in microwells
- Washing steps by individual nucleotides G, C, T, A
- The world's smallest solid-state pH meter
- Digital output

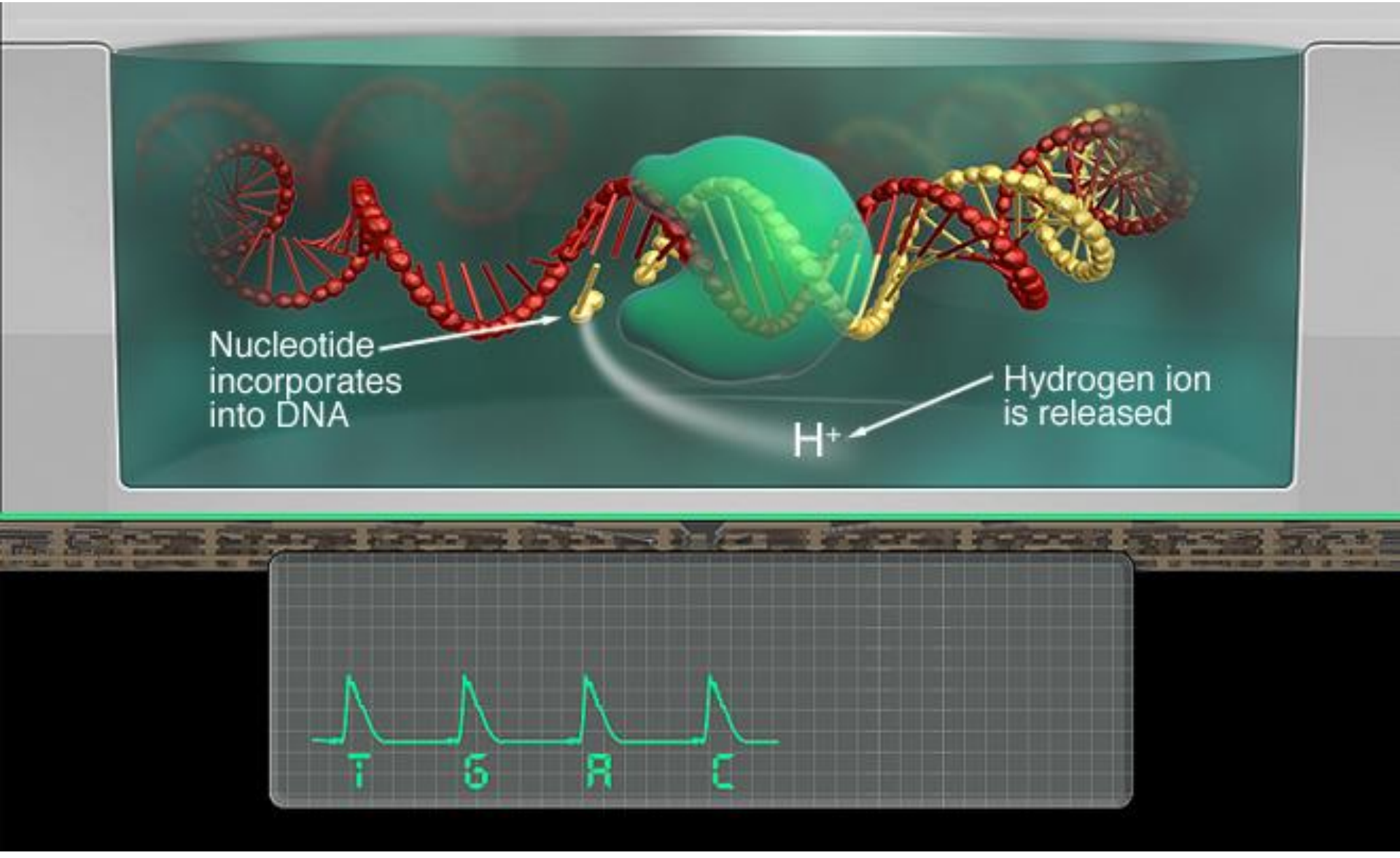
Hydrogen ion is released as a byproduct when a nucleotide is incorporated into a strand of DNA by a polymerase



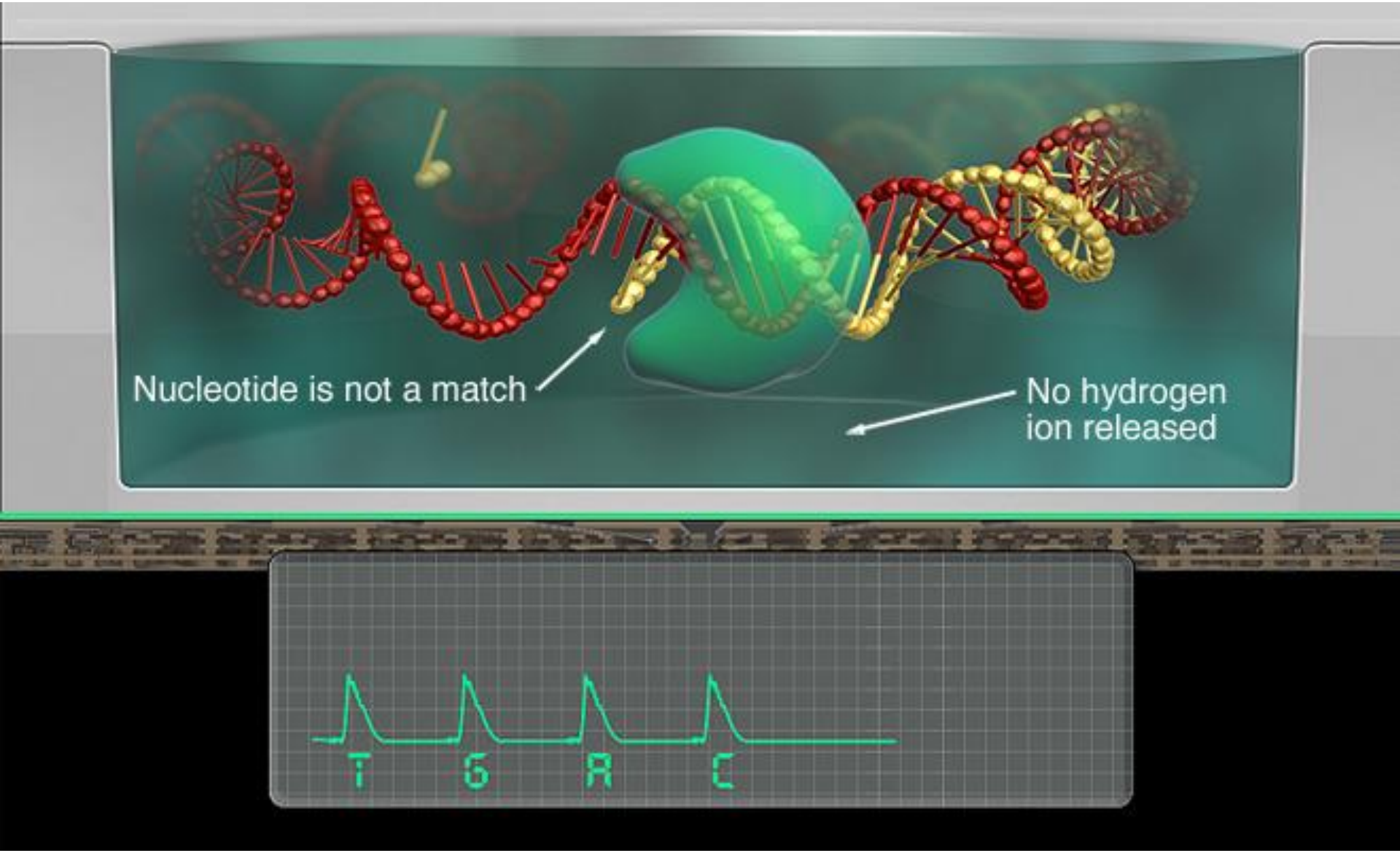
High-density array of micro-machined wells. Each well holds a different DNA template. Beneath the wells is an ion-sensitive layer and a proprietary ion sensor.



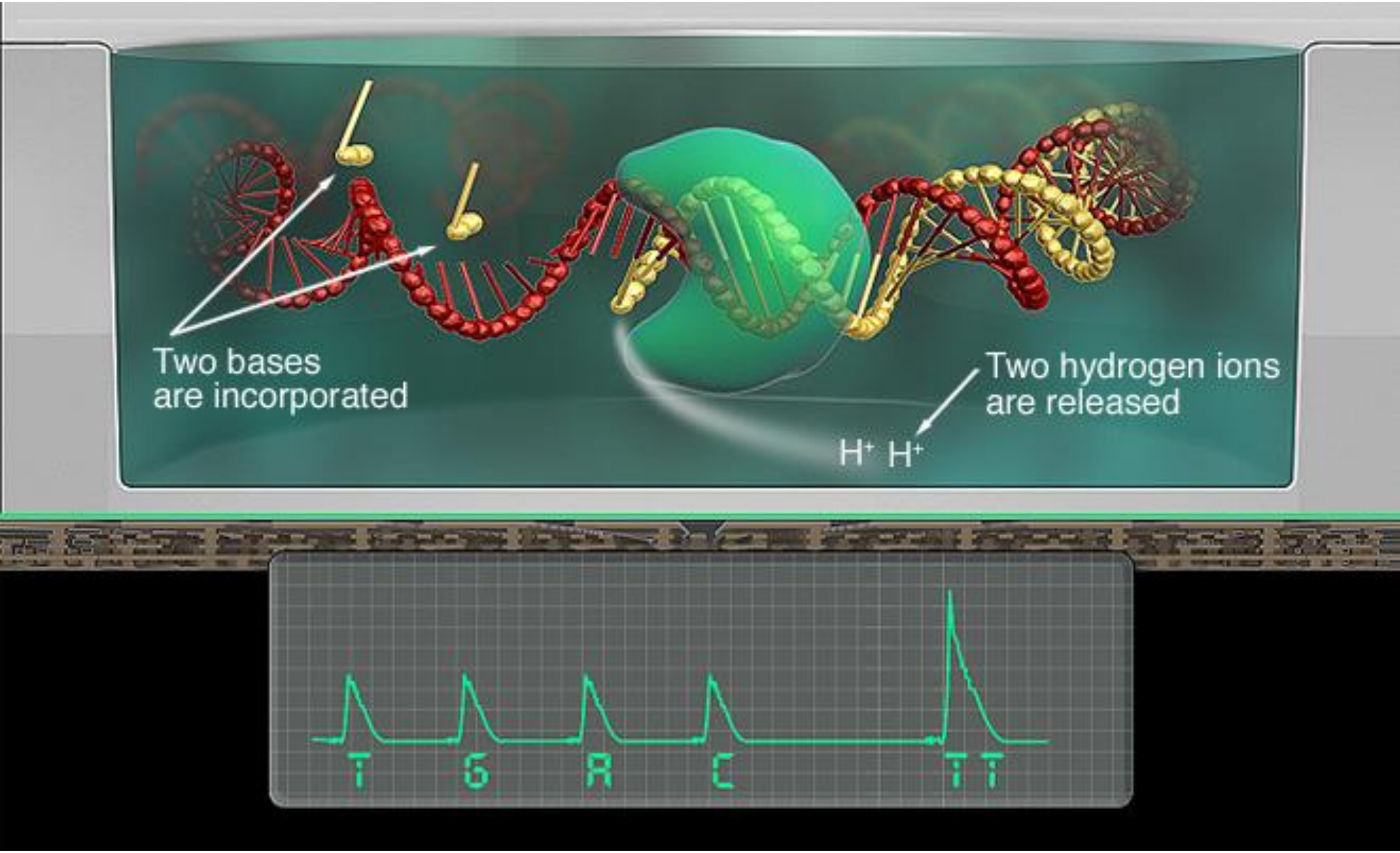
If a nucleotide is added to a DNA template and is then incorporated into a strand of DNA, a hydrogen ion is released. The charge from that ion will change the pH of the solution. The world's smallest solid-state pH meter—will call the base.



The sequencer sequentially floods the chip with one nucleotide after another. If the next nucleotide that floods the chip is not a match, no voltage change will be recorded.



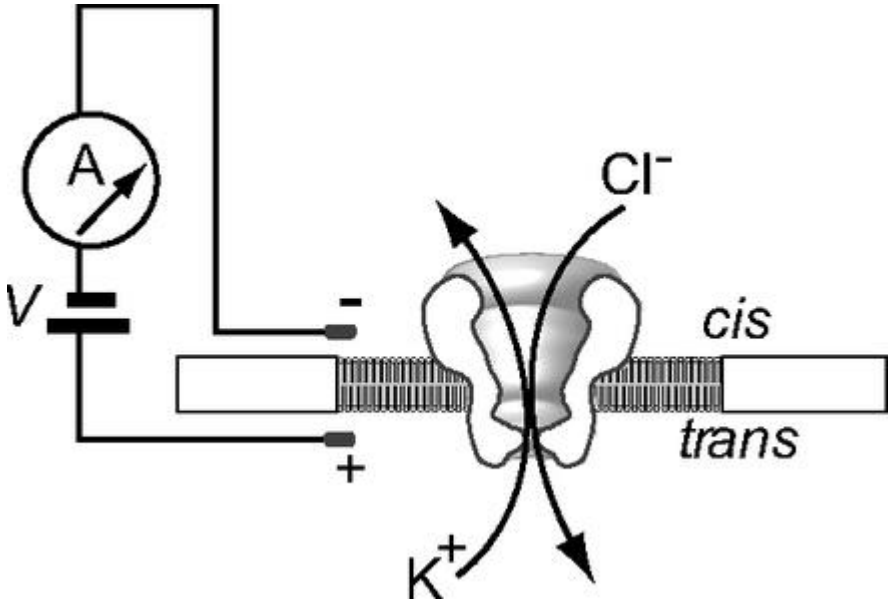
If there are two identical bases on the DNA strand, the voltage is double, and the chip records two identical bases.



Single molecule passage through a pore

Oxford Nanopore Technologies

Schematic of the nanopore device.



DNA sequencing development

2001: Genome draft of 5 individuals in 9 months
– more than billion $

2015: Complete human genome in an hour – ~100 $

Sample preparation
for next gen. DNA/RNA sequencing

single cell profiling

Single Cell RNA-Seq

- transcriptomes of thousands of single cells varying in type and state

Traditional Techniques:

- analysis of a few genes in thousands of individual cells (e.g., in situ hybridization)
- expression profile of thousands of genes only on a tissue homogenate.

Examples of *Single Cell RNA-Seq* applications:

- Understanding tumor heterogeneity and clonal evolution – lineage analysis, cancer stem cells, and drug resistant and metastatic clones.

- Understanding complex tissues (e.g. neural tissues - the first look at the entire transcriptional profile in individual neurons activated by external stimuli - a critical step in ultimately discovering how a memory is captured and stored).

- High resolution identification of cells types and markers, and understanding differentiation pathways in developmental and systems biology.

Experimental conditions for single-cell sequencing

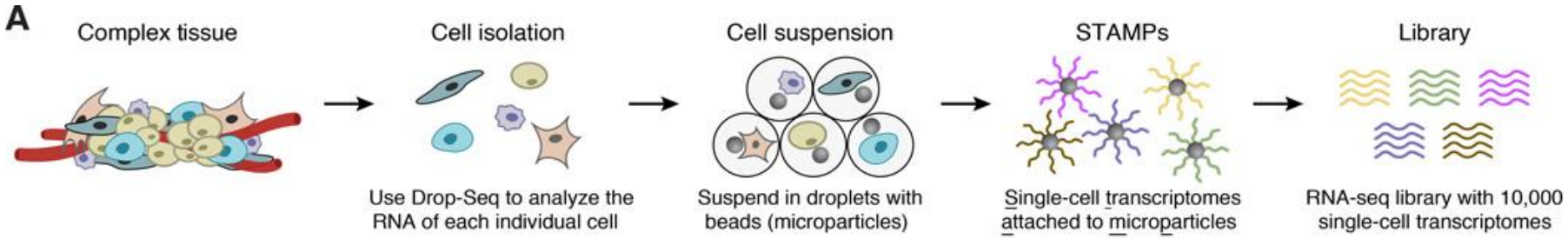
Thousands of cells from a tissue	– capturing containers (10^5 droplets/min)
Gene coding regions	– RNA
Complete transcriptome	– excess of capturing oligo primers
Cell identification	– cell barcode for each RNA fragment
Sequence identification	- one sequence could be analyzed many times
RNA constructs amenable to	- reverse transcription
	- PCR
	- high throughput next gen. sequencing

Drop-RNA seq

enables highly parallel analysis of thousands of individual cells by RNA-seq

(Macosko et al., Cell, 2015, 161,1202-14)

- Analysis of RNA or transcriptome variation in identified cells



separation of thousands of cells in suspension

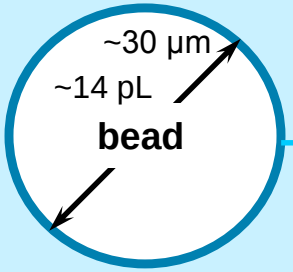
cell – RNA assignment – barcoding

analyses of cellular transcriptomes

Molecular barcoded cellular transcriptomes

in 0.5 nL droplets

10^8 reads on a single bead



PCR handle	cell barcode	mol. identifier	poly dT_{30}	cellular mRNA hybridized
identical for all beads	12 nts (4^{12} = 1.7*10^7) identical for all primers on a bead, i.e. for the cell in the drop	8 nts (4^8 = 65536) different on each primer (reveals PCR duplicates)	captures polyA on mRNA and primes reverse transcription	

1000 beads in µL

outside droplets

reverse transcription - cDNA

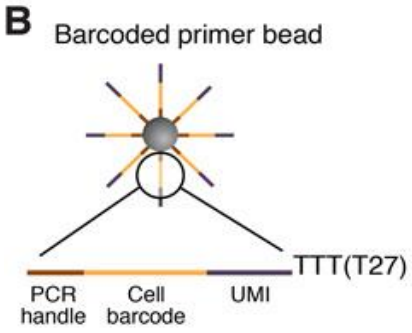
PCR amplified cDNA

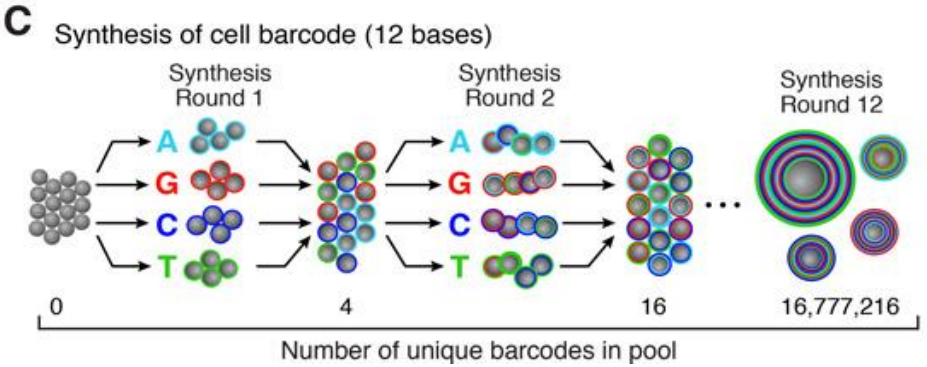
high throughput sequencing

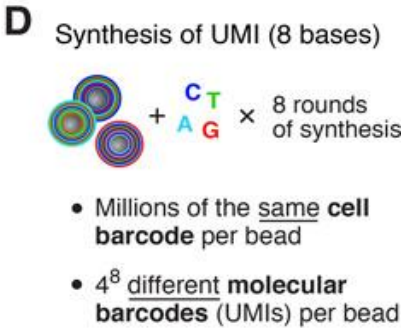
Synthesis of cellular barcodes and molecular identifiers on microparticles

Millions of primers on a microparticle

“split-and-pool“ strategy	- the same sequence of all primers on a single bead
„bar codes“	- 4^{12} (16,777,216) possible barcodes after 12 rounds
	- different microparticles have different sequences
degenerative synthesis	- 8 synthesis rounds with 4 DNA bases
„univ. mol. identifier“ (UMI)	- 4^{8} (65,536) possible sequences on each particle
	- specific sequences for each primer
30 dT sequence	- complementary for polyA RNA

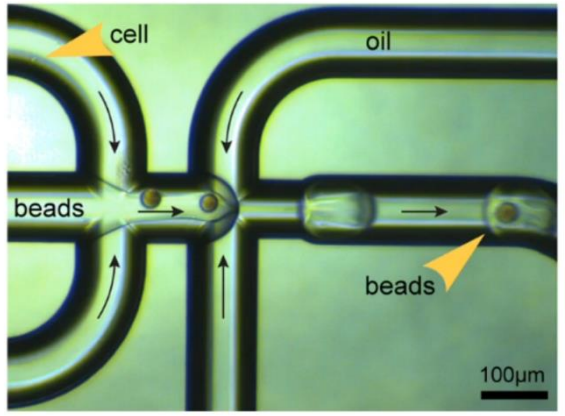


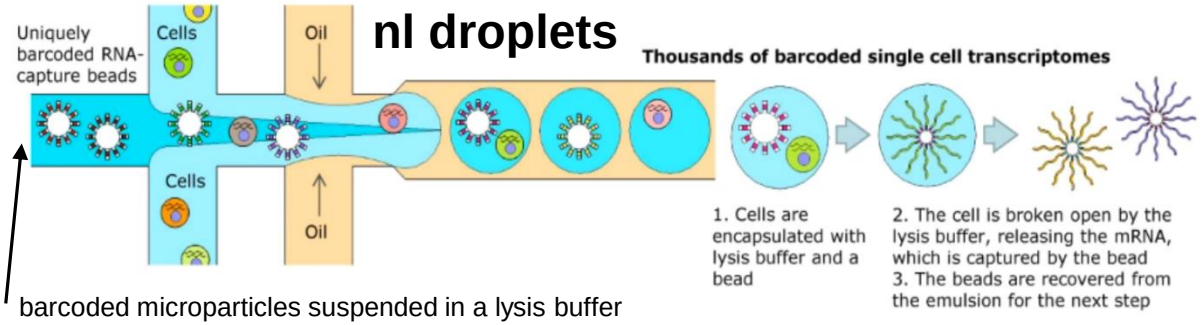




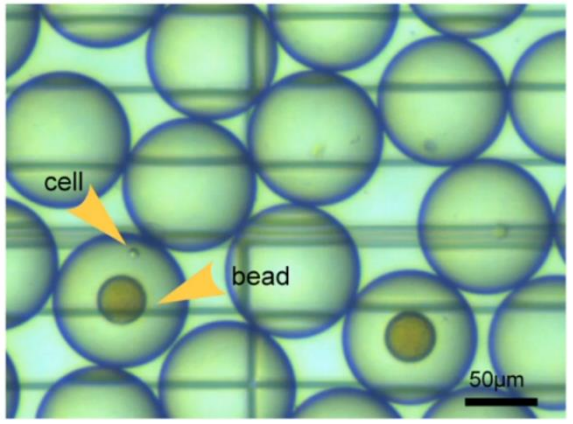
Single cell RNA-Seq

- Individual cells are captured in droplets with mRNA-capture beads
  - Each bead has a unique DNA sequence 'barcode'
  - There are 16 million unique barcodes in the bead library
  - Beads capture ~11% of mRNAs/cell (i.e., ~20,000 transcripts)

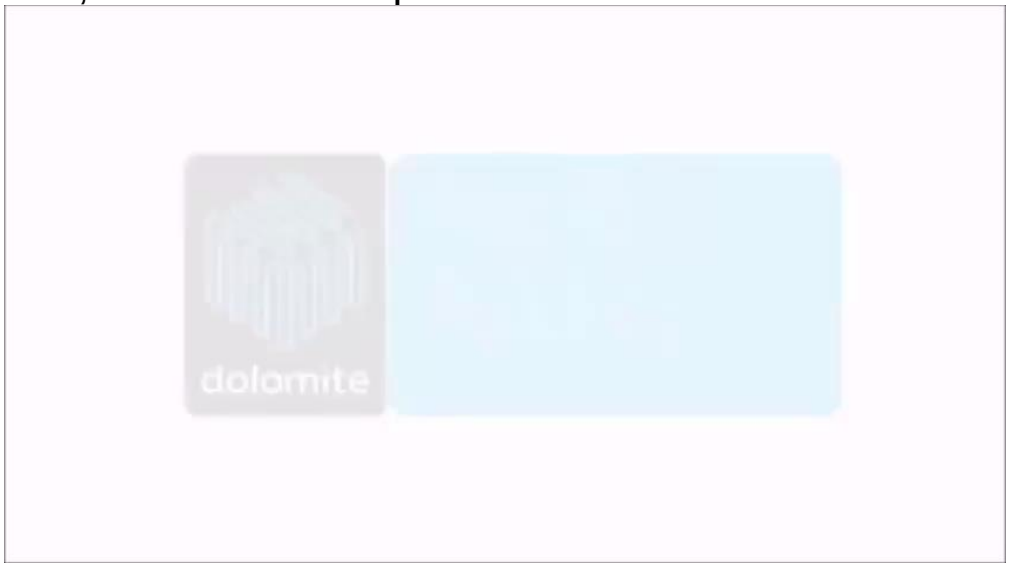




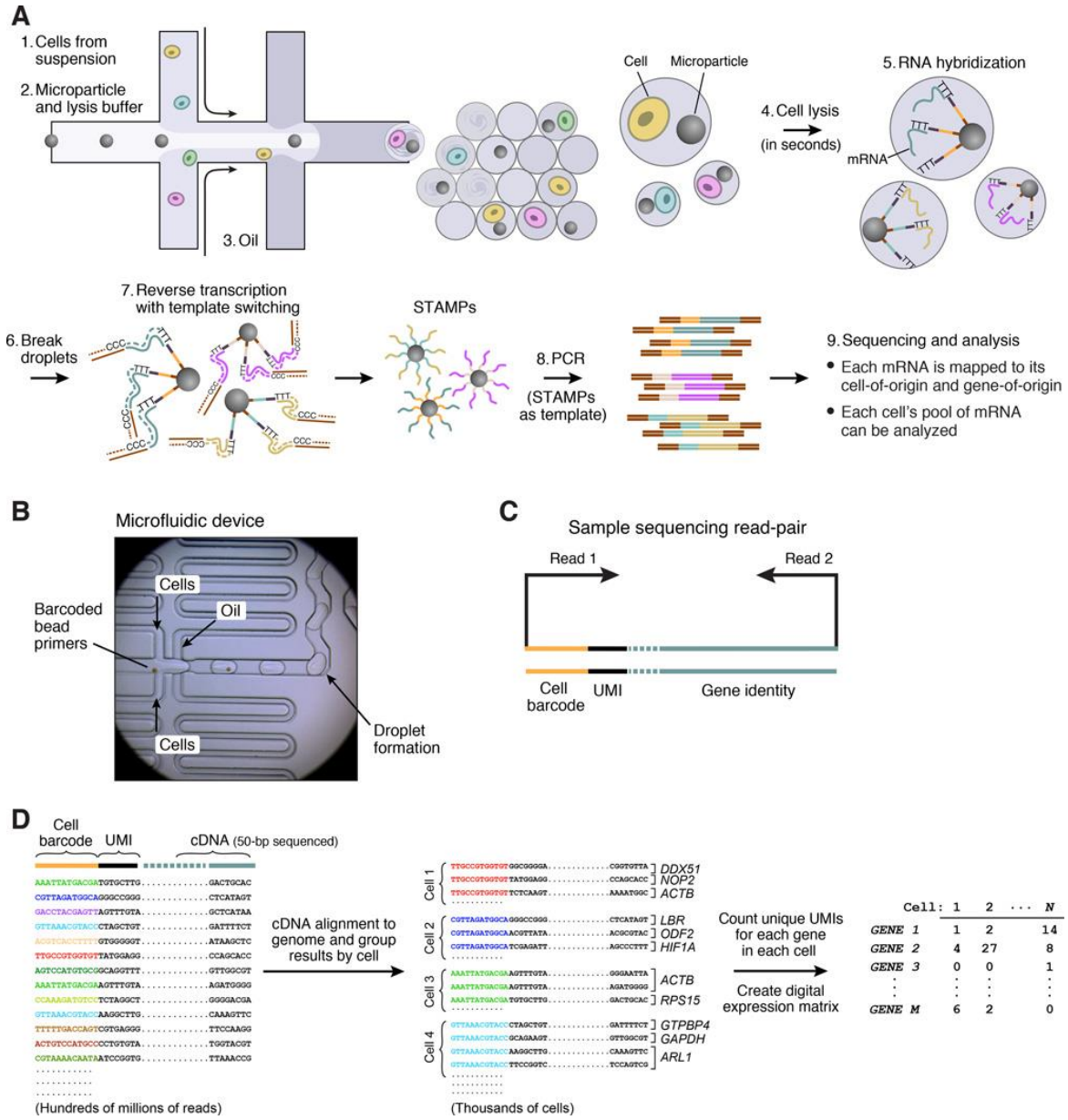
barcoded microparticles suspended in a lysis buffer



100,000 nl-sized droplets/min



Single Cell RNA-Seq



Complex neural mouse retina tissue

- transcriptomes from 44,808 mouse retinal cells analyzed
- 39 transcriptionally distinct cell populations identified