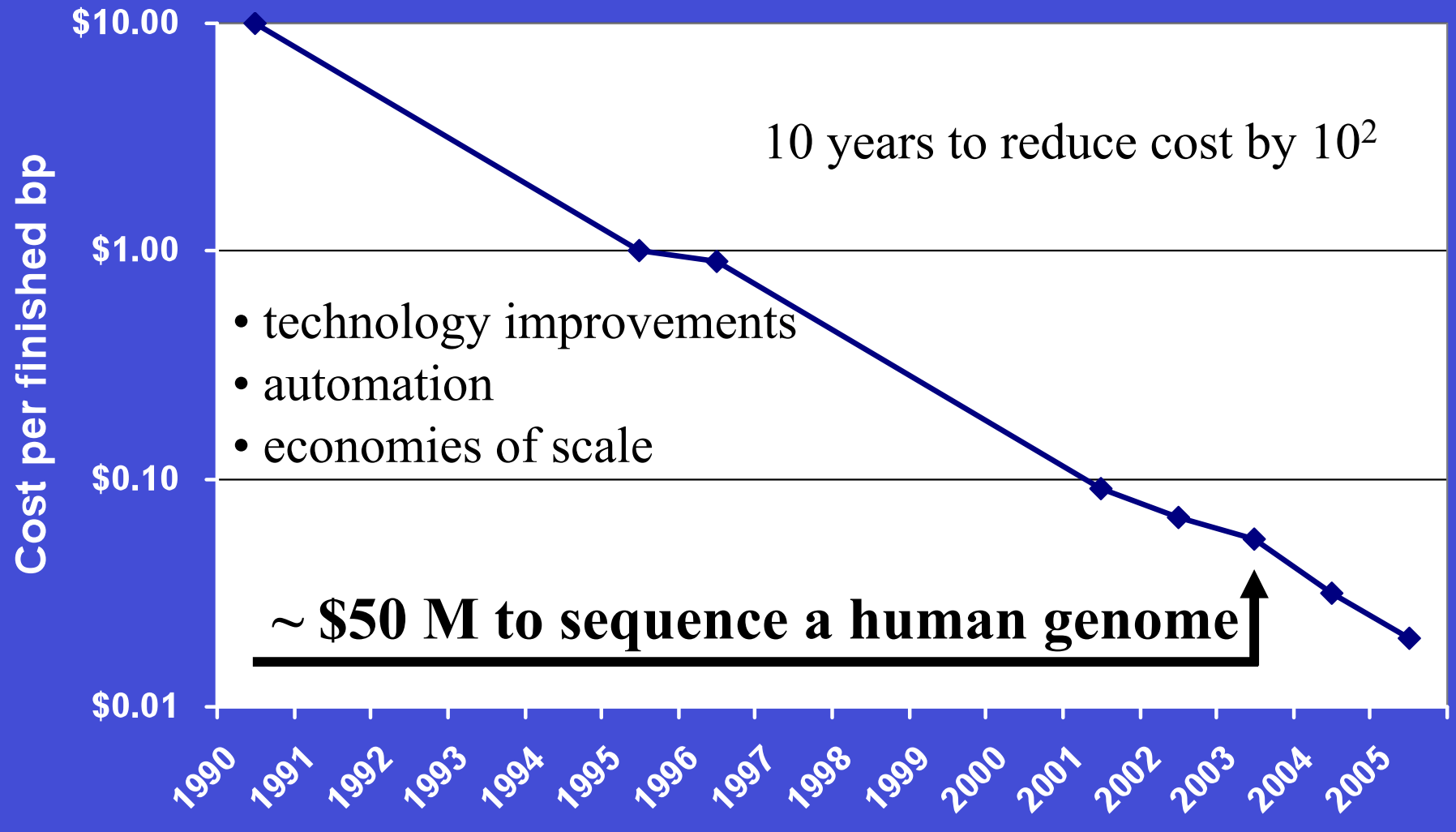
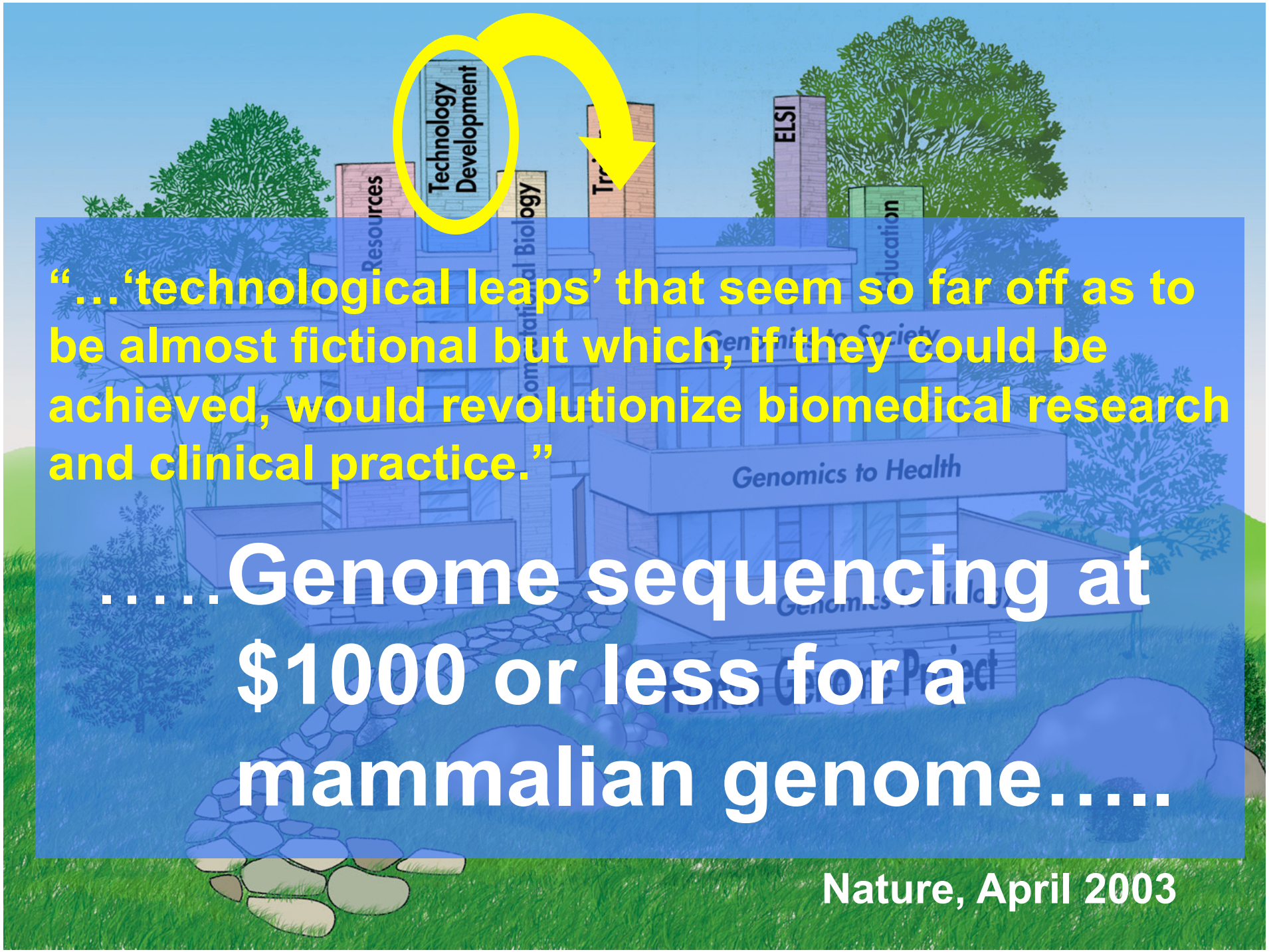
Decrease in the Cost of Finished DNA Sequencing





"…'technological leaps' that seem so far off as to be almost fictional but which, if they could be achieved, would revolutionize biomedical research and clinical practice."

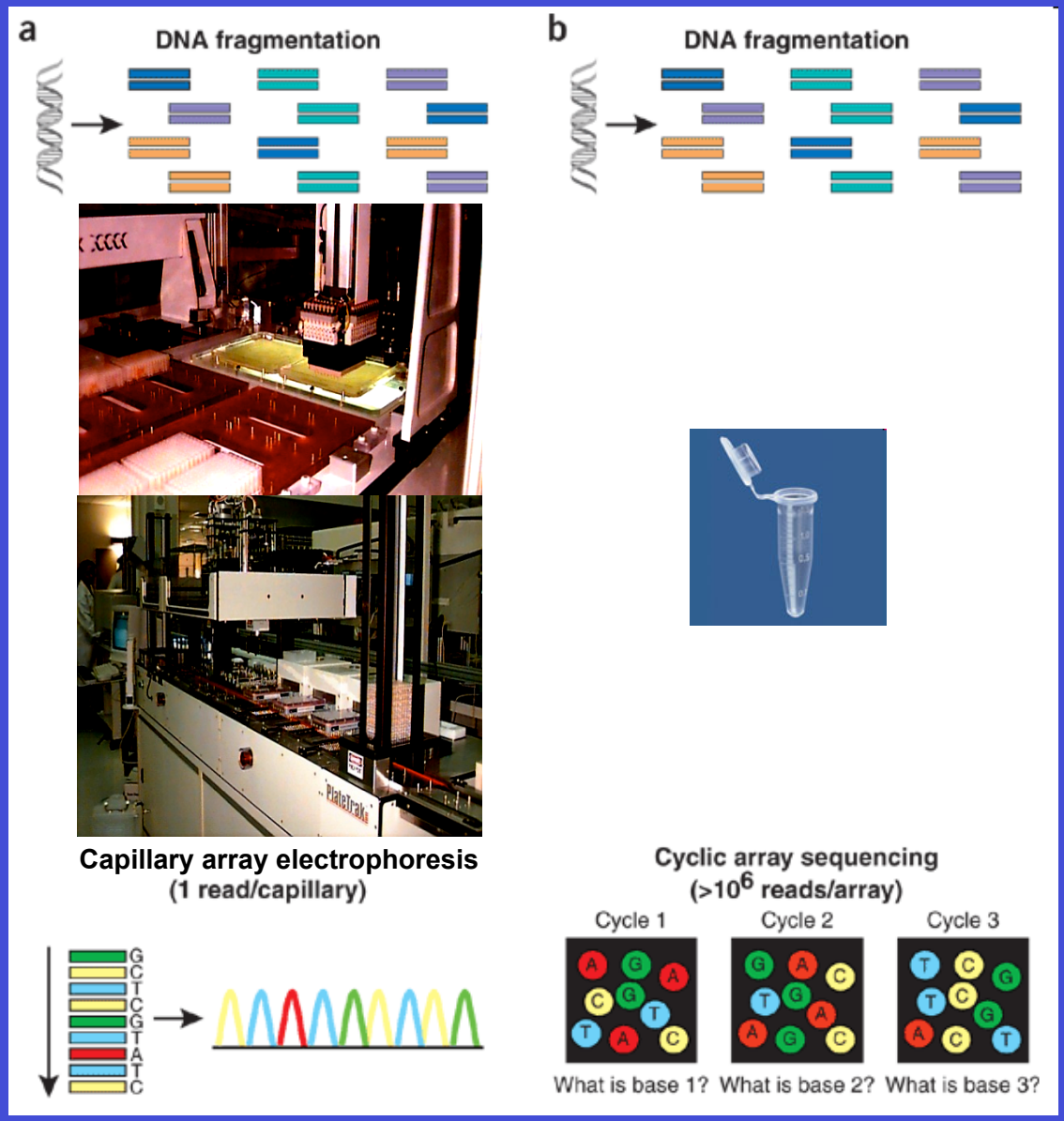
…..Genome sequencing at $1000 or less for a mammalian genome…..

Nature, April 2003

Workflow: Sequencing in 2003 vs. Sequencing today

100

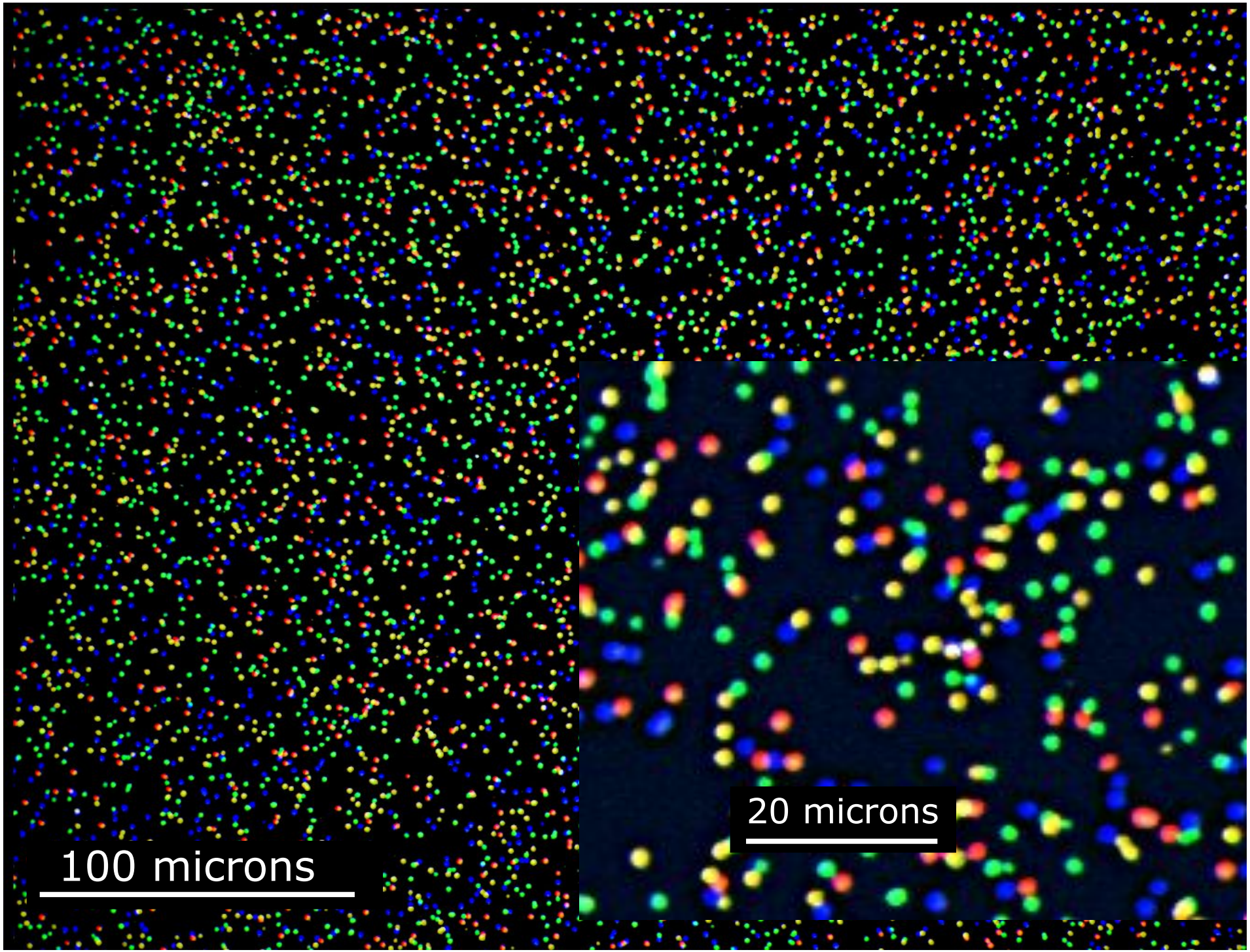
per run



>100

million

per run


100 microns
20 microns

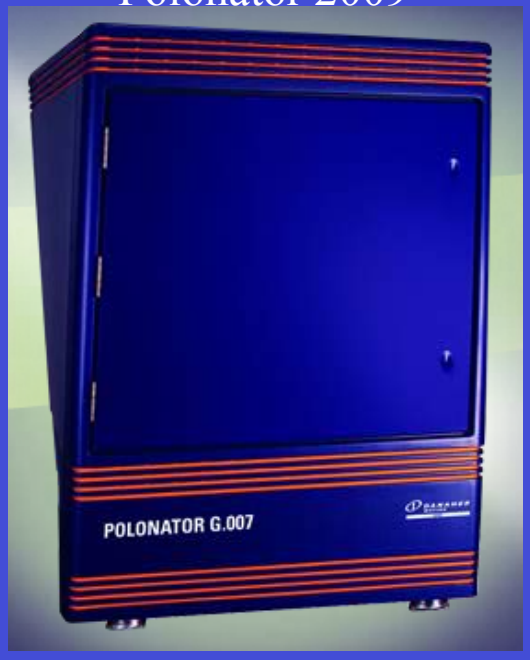
Roche/454 2005

Illumina / Solexa 2006

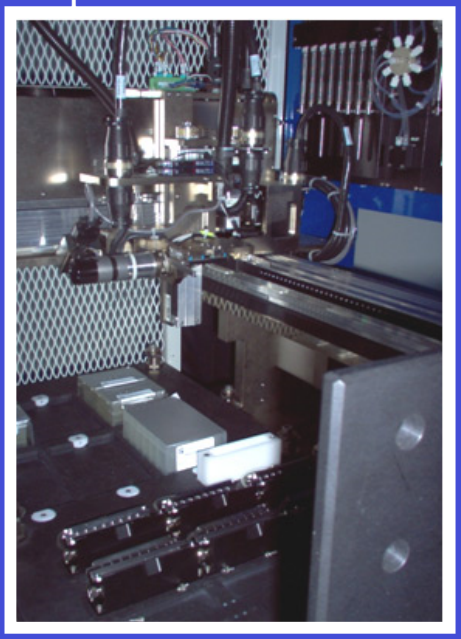
Applied Biosystems/Agencourt 2007

Helicos 2008

Polonator 2009



Complete Genomics 2010



Latest commercial systems 2010…

Roche / 454 FLX

Illumina HiSeq 2000

Applied Biosystems SOLiD 4

Helicos HeliScope

Pacific Biosciences RS

Ion Torrent PGM

…to scale

Throughput overview (estimated, snapshot)

Vendor (~cost/run)	method	read length	reads /run (B)	Gb /run	Gb /day	run time
Roche/454 ($ 7,000)	pyro-sequencing	400	0.001	0.5	1	10 hrs
Illumina/Solexa ($17,000)	reversible terminator	100	0.2	30	3	10 days
AB SOLiD ($30,000)	oligo ligation	50	1	50	3.5	14 days
Helicos	reversible terminator single molecule	35	1	35	4.5	8 days
Complete Genomics	probe-anchor ligation	10?	50?	2000?	?	30 days?

Applications

Unlike previous DNA sequencing technology, these new methods sample very large numbers of single molecules (or ensembles generated from single molecules), enabling new biological insights:

- genomes – re-sequencing, *de novo*
- sequence variation (SNP, indel, and larger)
  - rare variants, not just the common ones
- haplotypes (with difficulty)
- rearrangements (with difficulty)
- methylation status
- expression analysis – counting rather than ratios
- allele-specific expression analysis
- alternative splicing
- small RNAs
- ChIP-seq (proteins bound to DNA)
- rare samples (e.g., in mixtures) – high dynamic range

Resequence 1 Human Genome

Capillary Array Electrophoresis (2003)

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 (2007)

1 Gb/run, 2.5 days/run,

30x coverage of 6 Gb genome takes 1.5 year

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

~ 3-4 months with one machine

Resequence 1 Human Genome

Capillary Array Electrophoresis (2003)

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 (Sept 2009)

30 Gb/run, ~ 6 days/run

30x coverage of 6 Gb genome takes 3 runs

< 1 month with one machine

Resequence 1 Human Genome

Capillary Array Electrophoresis (2003)

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 (end of 2010, projected)

200-300 Gb/run, ~ 7-14 days/run

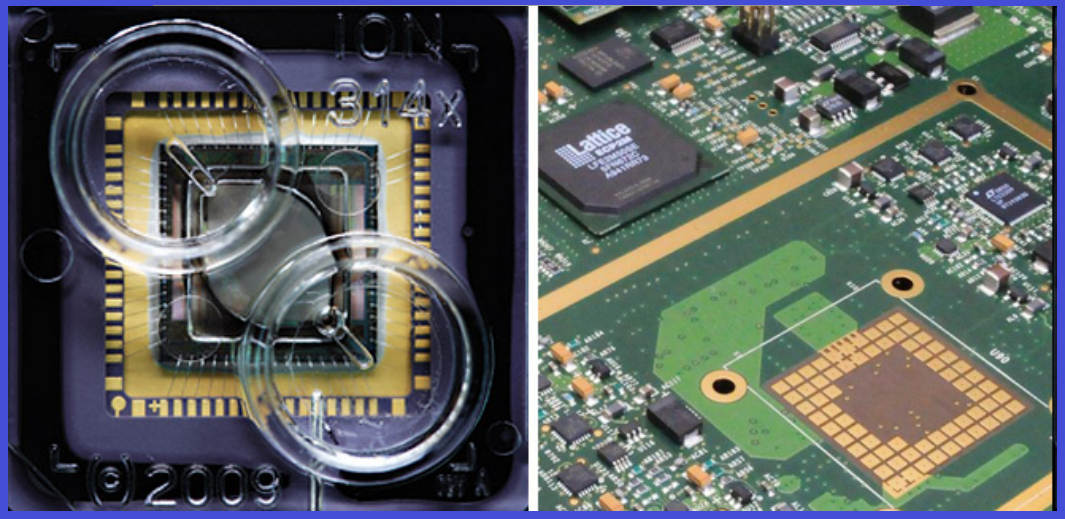
30x coverage of 6 Gb genome takes ~ 0.5 run

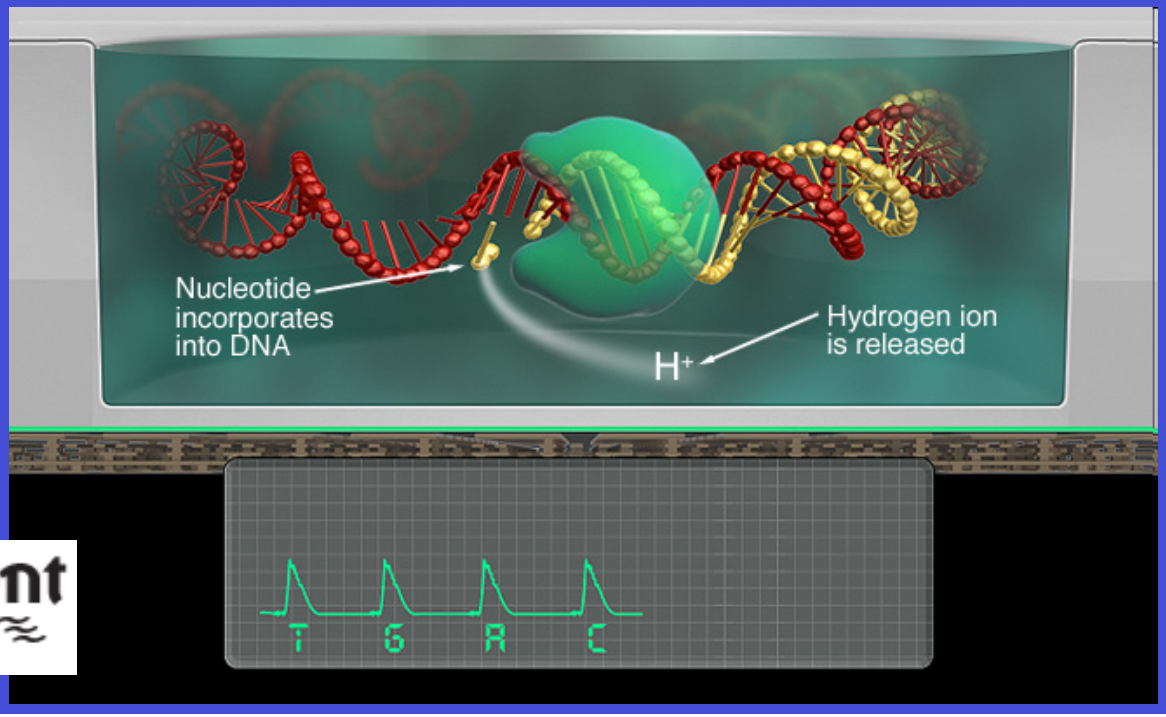
~ 1 week with one machine (~ 2 genomes)

Emerging sequencing technologies

Sequencing-by-synthesis, native DNA pol/dNTPs, pH detection

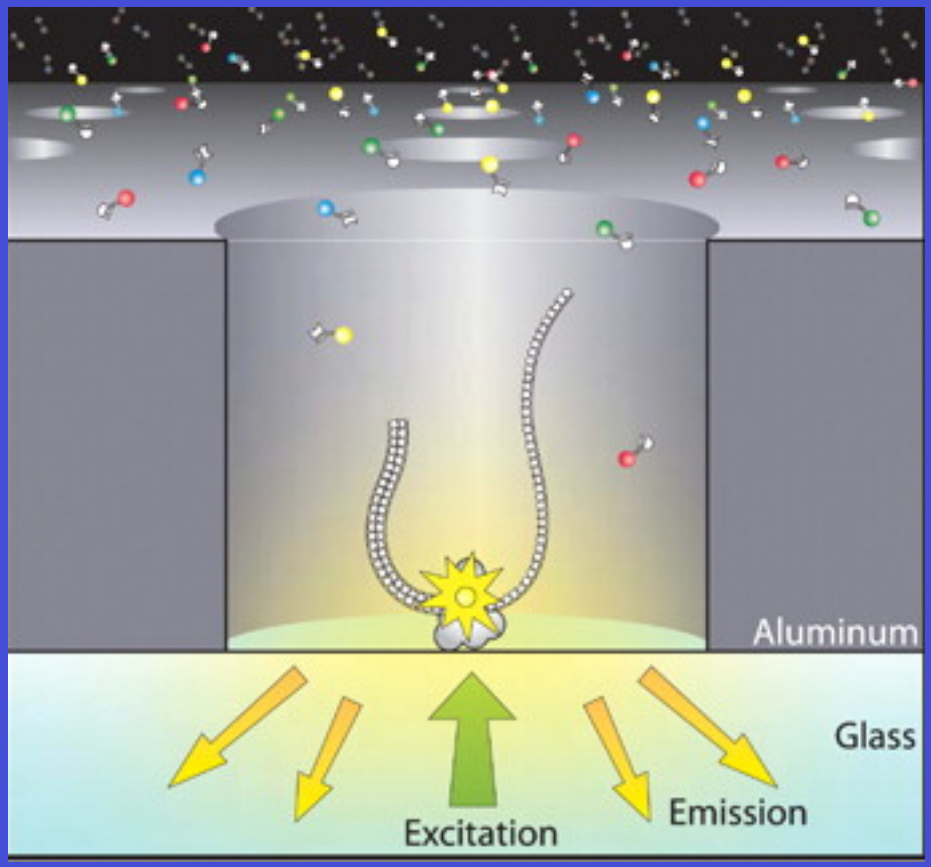
1.5 million pH meters





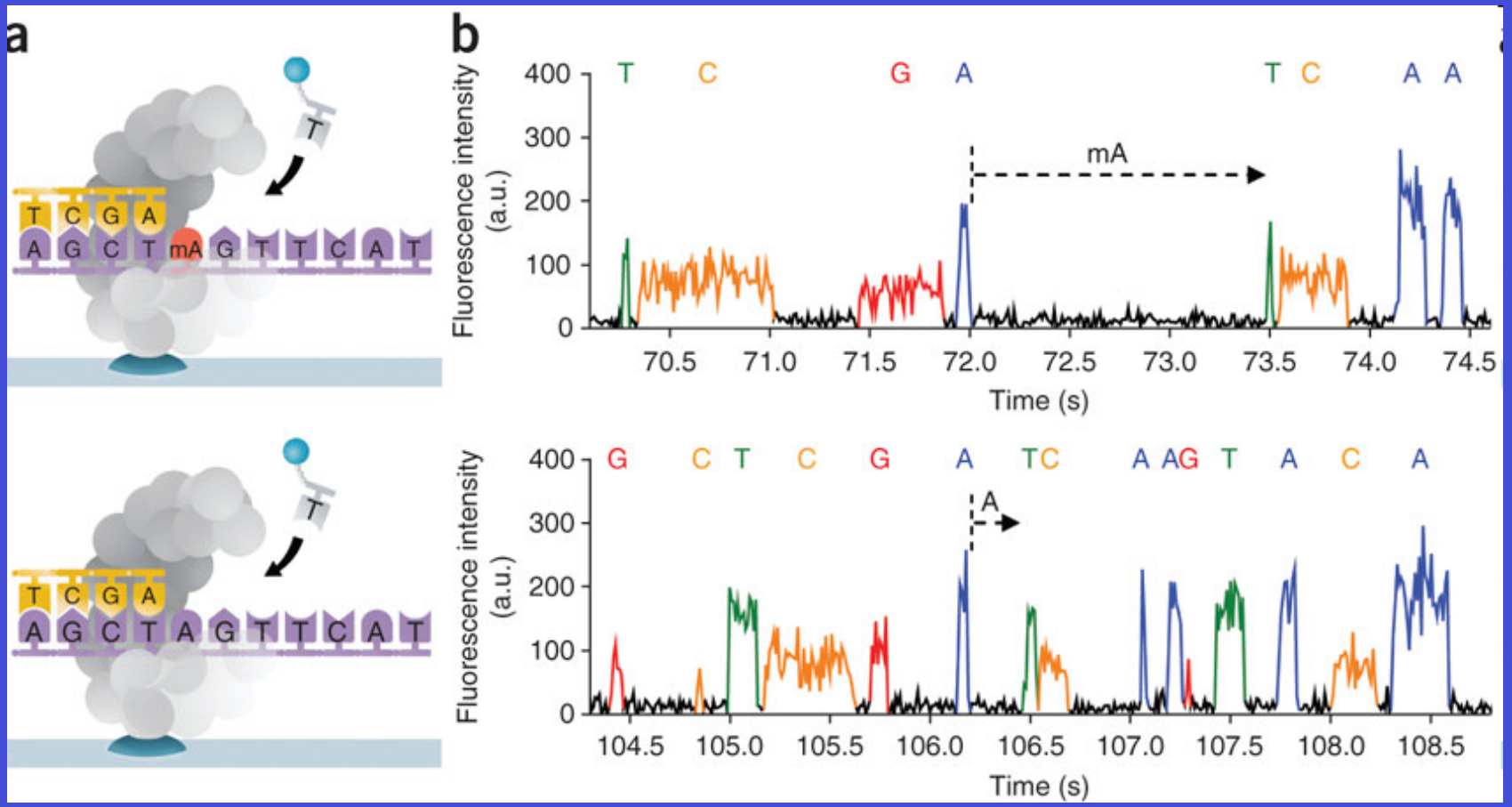
ion torrent

Free-running polymerase



PACIFIC BIOSCIENCES™

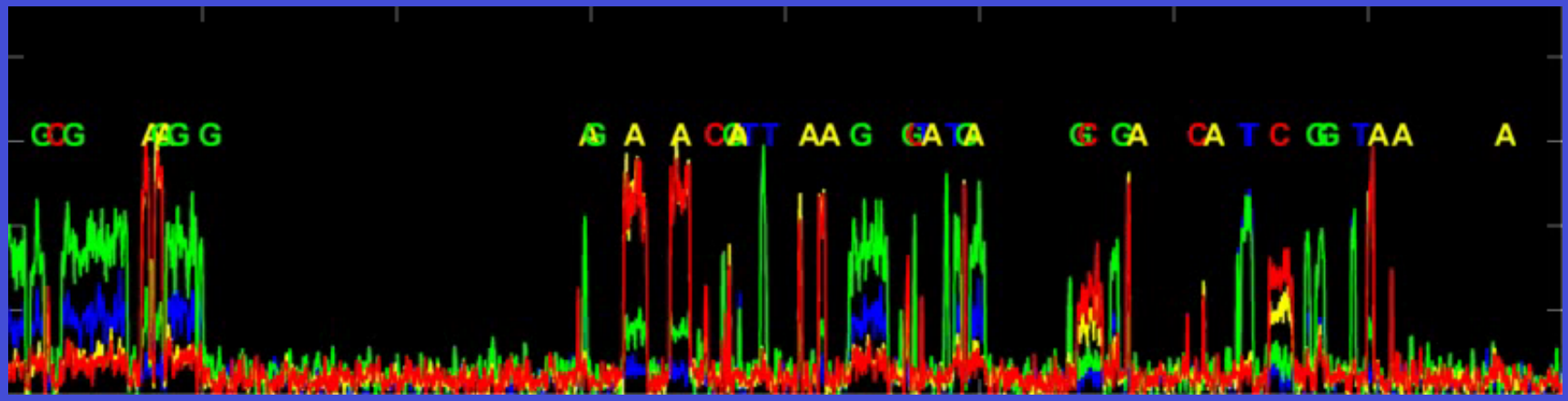
DNA polymerase as a sequence and methylation reader



Direct detection of DNA methylation during single-molecule, real-time sequencing.
B Flusberg, *et al.,* Nature Methods online 9 May 2010

Example of a very long read

10,351 Bases…

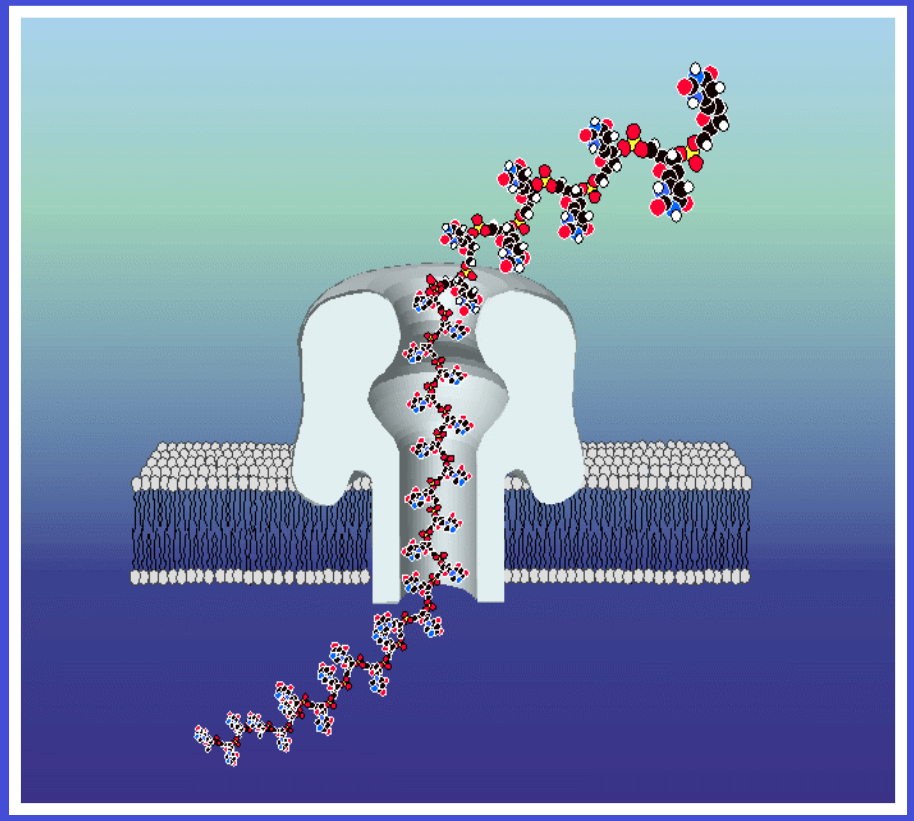


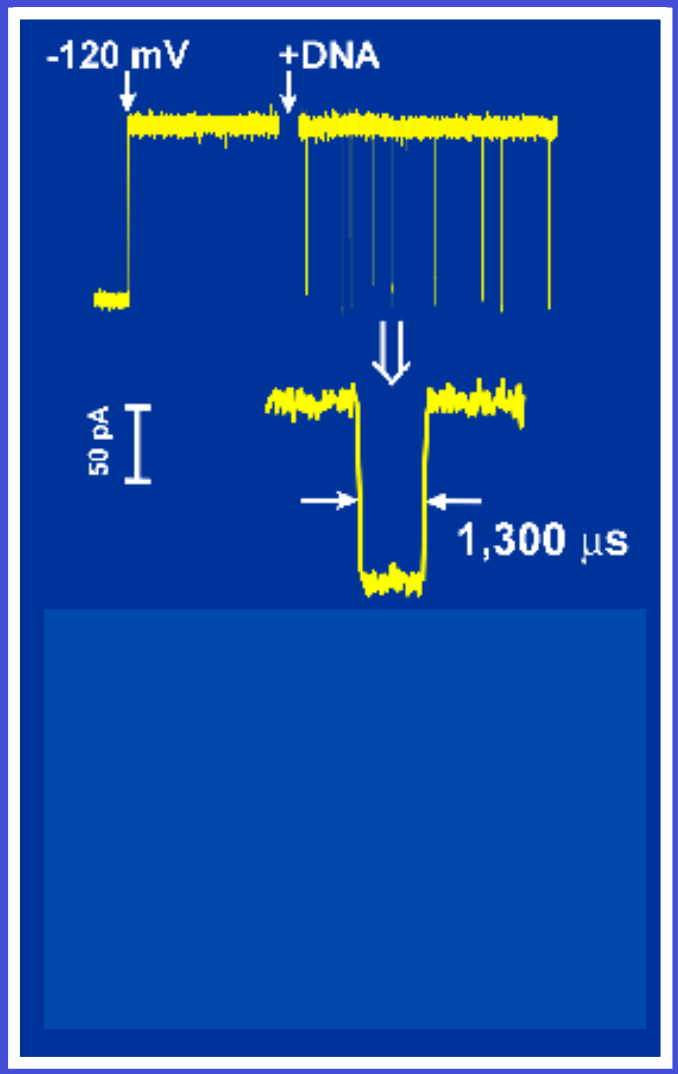
From the *E.coli* genome

PACIFIC BIOSCIENCES™

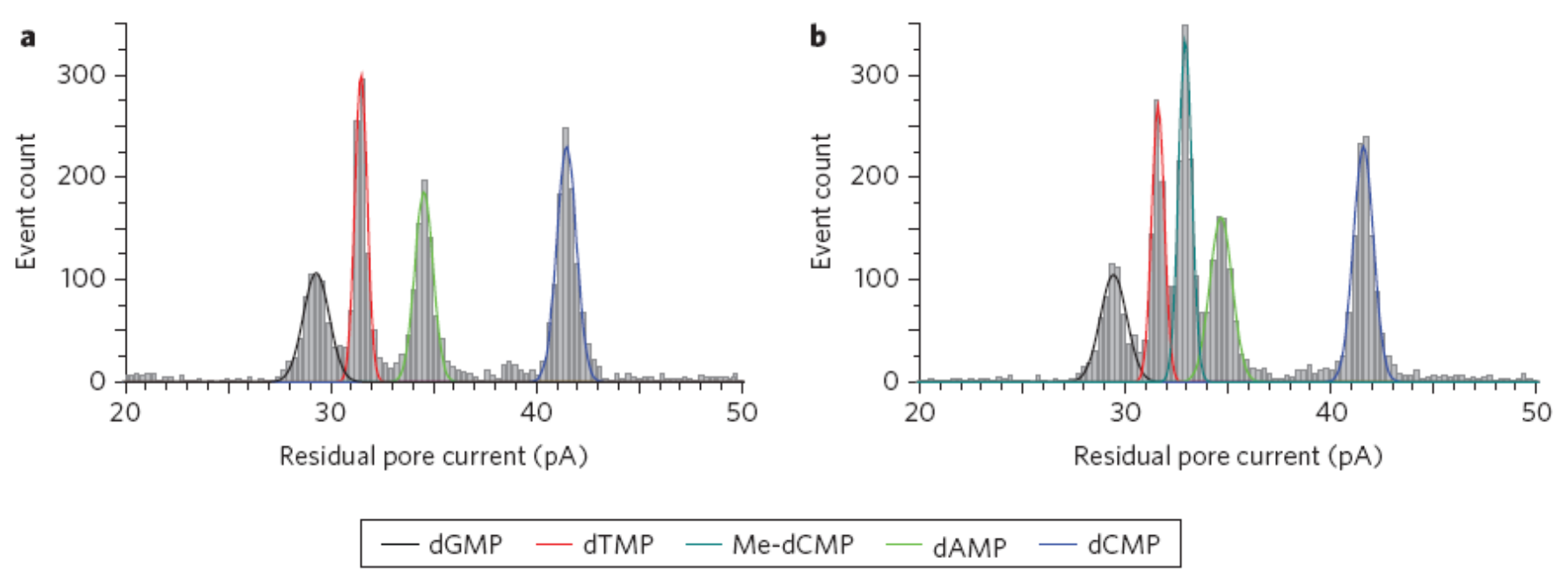
Nanopore sequencing

Nanopore sequencing concept





Direct electronic readout of A, C, G, T and methyl-C



Nanopore sequencing advantages

- Sequence genomic DNA directly – no conversion or amplification; no reagent costs except extraction
- Very long reads – assembly, haplotype information, structural variants. De novo sequencing
- Microbiome sequencing would be immensely simplified
- Non-destructive of the DNA sample
- A, C, G, T and modified bases
- RNA, too? Gene Expression, allele-specific G.E., splice variants, small RNAs, …
- Digital (gene expression, copy number variants)
- Very fast
- Fully electronic; takes advantage of integrated chip technology infrastructure
- Portable, hand-held devices

Resequence 1 Human Genome

Capillary Array Electrophoresis (2003)

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 (end of 2010, projected)

200-300 Gb/run, ~ 7-14 days/run

30x coverage of 6 Gb genome takes ~ 0.5 run

~ 1 week with one machine (~ 2 genomes)

Nanosensor (future)

1 msec per base

10x coverage of 6 Gb genome takes

~ 2 years with single nanopore

< 1 day with 1000 nanopore array – assembled?