Decrease in the Cost of Finished DNA Sequencing

- 10 years to reduce cost by $10^2$
- technology improvements
- automation
- economies of scale

~ $50 M to sequence a human genome
“…‘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

Capillary array electrophoresis
(1 read/capillary)

Cyclic array sequencing
(>10^6 reads/array)

100 per run

>100 million per run
Latest commercial systems 2010...

Roche / 454 FLX
Helicos HeliScope

Illumina HiSeq 2000

Applied Biosystems SOLiD 4

Pacific Biosciences RS

Ion Torrent PGM

...to scale
## Throughput overview (estimated, snapshot)

<table>
<thead>
<tr>
<th>Vendor</th>
<th>(~cost/run)</th>
<th>method</th>
<th>read length</th>
<th>reads /run (B)</th>
<th>Gb /run</th>
<th>Gb /day</th>
<th>run time</th>
</tr>
</thead>
<tbody>
<tr>
<td>Roche/454</td>
<td>($ 7,000)</td>
<td>pyro-sequencing</td>
<td>400</td>
<td>0.001</td>
<td>0.5</td>
<td>1</td>
<td>10 hrs</td>
</tr>
<tr>
<td>Illumina/Solexa</td>
<td>($17,000)</td>
<td>reversible terminator</td>
<td>100</td>
<td>0.2</td>
<td>30</td>
<td>3</td>
<td>10 days</td>
</tr>
<tr>
<td>AB SOLiD</td>
<td>($30,000)</td>
<td>oligo</td>
<td>50</td>
<td>1</td>
<td>50</td>
<td>3.5</td>
<td>14</td>
</tr>
<tr>
<td>Helicos days</td>
<td></td>
<td>ligation</td>
<td>35</td>
<td>1</td>
<td>35</td>
<td>4.5</td>
<td>8</td>
</tr>
</tbody>
</table>
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,
\( \sim 3 \) 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 \( \Rightarrow \) e.g., 4-6x improvement over next year
\( \sim 3-4 \) 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 \( \approx 1.8 \) 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
Free-running polymerase
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
Nanopore sequencing
Direct electronic readout of A, C, G, T and methyl-C

![Graphs showing event count vs. residual pore current for different nucleotides: dGMP, dTMP, Me-dCMP, dAMP, dCMP.](image_url)
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 \(\approx 1.8 \text{ Mb/day}\)
6x coverage of 3 Gb genome takes 26 years with 1 machine
\(\approx 3 \text{ months with 100 machines}\)

Sequencing by synthesis on array (end of 2010, projected)
200-300 Gb/run, \(\approx 7-14 \text{ days/run}\)
30x coverage of 6 Gb genome takes \(\approx 0.5 \text{ run}\)

\(\approx 1 \text{ week with one machine (\(\approx 2 \text{ genomes}\)}\)

Nanosensor (future)
1 msec per base
10x coverage of 6 Gb genome takes
\(\approx 2 \text{ years with single nanopore}\)
\(< 1 \text{ day with 1000 nanopore array – assembled?}\)