Progress in Whole Genome Sequencing

Dr. Paul Billings MD PhD
Chief Medical Officer
Overview of Life Technologies

1,500+ SCIENTISTS
600,000+ CITATIONS
10,400 EMPLOYEES

4,000+ PATENTS
75,000+ CUSTOMERS
$3.7B SALES
Life Technologies Genetic Analysis Portfolio - Rapid Discovery, Confirmation and Screening

- **Ion Proton™ System**
  - Single Day, Benchtop Sequencing of Exomes and Genomes

- **Ion PGM™ System**
  - Rapid, Benchtop gene sequencing & gene expression

- **Capillary Electrophoresis**
  - for variant confirmation

- **Real Time PCR Systems**
  - for variant ID & rapid screening of large numbers of samples

- **TaqMan® Assays**
  - for largest catalog of assays and streamlined custom pipeline

**Diagram:**
- Discovery
- Confirmation
- Screening
Summary of first, second and third generation sequencing

1. ‘Sanger’
   - Chain termination

2. ‘Next’ generation
   - Reversible termination
     - Illumina/ Solexa
     - HiSeq
   - Pyrosequencing
     - Roche/454 Lifesciences
     - Genome Analyser
   - Ligation
     - Life Technologies
     - SOLID

3. ‘Next-next generation’
   - Fluorescence
     - Labelled polymerase
       - Life Technologies
     - Tethered polymerase
       - Pacific Biosciences
     - Reversible termination
       - Helicos
   - Electronic
     - pH
       - Ion Torrent/Life Technologies
   - Biological nanopores
     - Oxford Nanopore
   - Solid-state nanopores
     - IMB/Roche, NABSys
   - Microscopy
     - Halcyon Molecular, ZS Genetics

Key: blue = sequencing by synthesis; green = sequencing by ligation; yellow = nanopore sequencing; orange = sequencing by electron microscope
Evolution of Sequencing Technologies

- **Gel-based Sanger sequencing**
- **High-throughput Sanger sequencing**
- **Massively parallel ‘next generation’ sequencing**

**Year**
- 1975
- 1980
- 1985
- 1990
- 1995
- 2000
- 2005
- 2010
- 2015

- Sanger method published
- PCR invented
- DNA microarrays developed
- Capillary gel electrophoresis developed
- Single molecule sequencing?
- Short read sequencers launched

**Kilobases of DNA per day per machine**
- 10
- 100
- 100,000
- 1,000,000
- 10,000,000
- 100,000,000
- 1,000,000,000
- 10,000,000,000
- 100,000,000,000

**Generations**
- 1st generation
- 2nd generation
- 3rd generation
Sequencing Costs Plummeting

The cost of sequencing fell dramatically a few years after high-throughput sequencing machines were introduced.
Next Generation Sequencing goes global
Bioinformatics pipeline for analyzing whole genome sequencing data

1° analysis
BASE-CALLING

Signal analysis
Base-calling
Quality scoring (raw)

Reference sequence

Reference sequence

Mapping reads to reference
Quality scoring (consensus)

Variant calling
- Read depth analysis
- Paired-end analysis
- Haplotype inference

Visualisation

Databases of variation

2° analysis
ALIGNMENT

Variant analysis
- Genetic filtering
- Functional filtering

Validation of candidates

Clinical interpretation and decision-making
- Diagnostic
- Therapeutic

3° analysis
INTERPRETATION
Clinical Implementation Challenges for WGS

**Economics**
- Expensive developmental costs
- Economic studies to show cost savings
- Need for solid business model for private sector to obtain good ROI. Reimbursement.

**Users**
- Large allelic variation and “dark” regions
- Need for greater population data
- Participation from patient community still too low in research.

**Storage**
- Whole Genome Sequencing will generate GBs of data
- Storage of this data will present many challenges including cost and security

**Standards**
- Regulatory requirements and uncertainty
- Need for standards to compare data from different platforms and controls
The need for more population data

- **HOW DO WE ENCOURAGE MORE PARTICIPATION?**

<table>
<thead>
<tr>
<th>Disease Relevant Alleles</th>
<th>Odds Ratio (Effect size)</th>
<th>Number of individuals needed</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mendelian Diseases</td>
<td>Common Diseases e.g. Cancer</td>
<td></td>
</tr>
<tr>
<td>1 - several</td>
<td>100,000 +</td>
<td></td>
</tr>
<tr>
<td>10 – 100+</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Genomic Data Storage Challenge

Digital information storage (in petabytes)

- **28 PB**: The digital storage that would be needed to hold seven billion genomes at four megabytes each.
- **100 PB**: The amount of digital information Facebook stores in the form of photos and videos.
- **85,000 PB**: The storage that would be required if all 850 million Facebook users had 100 gigabytes of recorded genomic data.

- **721,181 PB**: The world's total installed digital storage capacity in 2010.
The Need for Standards

- Clinical quality standards for each step of sample to answer.
- Controls.
- Clinical utility.
- Adoption of competing technologies by systems and payers.
- Consents.
- Trial design.
- Health economics
The Economic Challenge

- Equity of access demands a low cost, universal solution in health care.
- R&D are costly and require an ROI.
- Proving clinical utility is costly and standards not available.
- Distribution, supportive services and informatic infrastructure add expense.
- Macro- and health care related economics have not recently been favorable for innovation and new technologies in the developed world.
- Coding and reimbursement changes as well as significant new payments for medical devices and therapeutics make WGS problematic to provide.
Conclusions

- Life Technologies provides important solutions for research and will deliver clinical applications of WGS.
- WGS will become universally available if current trends continue.
- Many challenges remain to insure that:
  - remarkable innovation continues
  - revealing data becomes available
  - new tests are used properly
  - unmet needs are addressed
  - innovation is properly valued.
Thank you!