Technology Elements Driving Rapid Performance Growth of Highly Parallel DNA Sequencing Systems

29 June, 2011
Mark Pratt
Our Vision

Innovating for the Future of Genetic Analysis

From Genome Wide Discovery…

To Targeted Validation and Beyond…

To be the leading provider of integrated solutions that advance the understanding of genetics and health
Illumina Statistics

- Founded in 1998, San Diego headquarters
- Initial Public Offering on July 27, 2000 [NASDAQ: ILMN]
- Current market capitalization >$9 billion
- Over 500,000 sq. ft. of facilities in 6 countries
- More than 2,200 employees
- IP portfolio of >150 issued patents
- Ranked #1 on Forbes’ list of Fastest Growing Technology Companies in 2009 & 2007
- Added to the NASDAQ-100 listing in 2008
- Deloitte’s 2009 Technology Fast 500 Ranking
- MIT Technology Review’s 50 most innovative companies in the world list 2010
Industry Leading Growth

*Strong Performance While Reinvesting in New Opportunities*

**35% Revenue CAGR**

<table>
<thead>
<tr>
<th>Year</th>
<th>Non-GAAP Operating Profit</th>
<th>Revenue</th>
<th>Non-GAAP EPS</th>
</tr>
</thead>
<tbody>
<tr>
<td>2007</td>
<td>$94</td>
<td>$367</td>
<td>$0.42</td>
</tr>
<tr>
<td>2008</td>
<td>$169</td>
<td>$573</td>
<td>$0.68</td>
</tr>
<tr>
<td>2009</td>
<td>$210</td>
<td>$666</td>
<td>$0.80</td>
</tr>
<tr>
<td>2010</td>
<td>$286</td>
<td>$903</td>
<td>$1.06</td>
</tr>
</tbody>
</table>

Stock based compensation included in non-GAAP earnings per share but excluded in non-GAAP operating profit.
Revenue growth from new products* (x $1,000,000)
53% compound annual growth

In addition to addressing new markets, Illumina is a very successful at rapidly evolving its product line.

How does this process work?

* Revenue from products in first two years from introduction (unaudited)
Pre-Historic Sequencing… Circa 2005 (1 yr BGA)

- Large Sequencing Factories
- 100 to 150 Capillary Sequencers
- 5 to 10 Large Colony Picking Robots
- Dozens of PCR Machines
- Several Liquid Handling Robots
- Thousands of 384 well plates
- Dozens of lab personnel
- Multi-million dollar budgets

Broad, WashU, Sanger, Baylor, Venter, JGI, etc.
… replaced with a flow cell
The Impact of Scale in Sequencing

10^4 scale in throughput; 10^7 scale in parallelization in 5 years
An Illumina Sequencer for Every Lab
Making Next-Gen Sequencing Ubiquitous

CapEx Sensitivity vs. Price per Gb

- **MiSeq**
  - 1Gb
  - $100K

- **GA\text{IIx}**
  - 1Gb

- **HiScanSQ**
  - 100Gb

- **HiSeq 1000**
  - 300Gb

- **HiSeq 2000**
  - 600Gb

**Instrument Price**
- $100K
- $300K
- $500K
- $700K

**Price/Gb**
- $50
- $150
- $500

COMPANY CONFIDENTIAL – INTERNAL USE ONLY
Data Equivalence

60,000

HiSeq 2000
DNA SEQUENCING
What is DNA?

DNA is the repository of all genetic information. Information is encoded in a 4-letter code and stored in complimentary copies as strands in a double helix structure. It is also the physical instruction for building molecular machines (proteins) to support cell functions.

DNA is constructed by a special protein which uses a single strand as a template from which to make a complimentary copy.

RNA is directly transcribed from DNA using a related 4-letter code. Useful segments of DNA, called genes, are used as physical instructions to construct proteins from amino acids using three letter code groups.
Genetic analysis has something for everyone

- **Human health**
  - Understanding disease and aging
  - Cancer diagnosis & treatment
  - Preventative therapies
  - Drug development

- **Bioengineering**
  - Crops
  - Fuels

- **Conservation**
...and genetic self awareness

Your Genetic Data

<table>
<thead>
<tr>
<th>Who</th>
<th>Genotype</th>
<th>What It Means</th>
</tr>
</thead>
<tbody>
<tr>
<td>GG</td>
<td></td>
<td>Can taste certain bitter flavors.</td>
</tr>
<tr>
<td>Mark Pratt</td>
<td>CG</td>
<td>Special Powers</td>
</tr>
<tr>
<td>CC</td>
<td></td>
<td>Has about an 80% chance of not being able to taste certain bitter flavors.</td>
</tr>
</tbody>
</table>

Maternal Haplogroup: H5b

H5b is a subgroup of H5, which is described below.

Locations of haplogroup H5 circa 500 years ago, before the era of intercontinental travel.

Haplogroup H5 originated in the Caucasus region during the Ice Age, then spread out in several directions after the climate began to warm about 15,000 years ago. Today the haplogroup is common in Lebanon and many parts of Europe.
Unparalleled Publications Ramp

Fastest NGS publication rate, >1500 peer reviewed publications

Cumulative Illumina Peer Reviewed Sequencing Publications
Two basic architectures for DNA sequencing systems

Electrophoresis (fragment sorting)
Physical limitations on single channel rate and practical instrument limits on parallelism. Commercially peaked at 96-plex.

1-dimensional multiplexing (possible)

Sequencing by synthesis* (SBS)
Substantial physical limitations on individual reaction rates but extremely amenable to multiplexing. Commercially at several billion reactions proceeding in parallel.

2-dimensional multiplexing

* ligation, hybridization (biologically mediated)
Applications for DNA sequencing are enabled by an extremely fast growth in capability
Illumina Sequencing Technology
Robust Reversible Terminator Chemistry Foundation

DNA (0.1-1.0 ug)

Sample preparation

Cluster growth

Sequencing

Image acquisition

Base calling

COMPANY CONFIDENTIAL – INTERNAL USE ONLY
TruSeq Chemistry: Flow Cell

- **Simplified workflow**
  - Clusters in a contained environment (no need for clean rooms)
  - Sequencing performed in the flow cell on the clusters

Surface of flow cell coated with a lawn of oligo pairs

8 channels
Data Analysis

Detecting clusters
Measuring the color for each cluster
... for every cycle
SYSTEM ARCHITECTURE
At launch the Solexa sequencing system offered a 100-fold advantage in cost over other sequencing platforms. Since then Illumina has evolved the platform to reduce costs another 100-fold and improve throughput more than 1000-fold without changing the basic concept.
Start with a high fidelity SBS chemistry system comprising:
- Surface bound clonal clusters of single stranded DNA
- Step-wise synthesis using dye-labeled nucleotides

Add fluidic systems to host reaction, manage reagent deliveries and thermal environment.

Synchronously monitor system with automated fluorescent microscope system comprising:
- Optics and lasers
- Cameras and filters
- Motion control and autofocus systems

Tie it together with a control system, user interface & automated analysis pipeline.
Serial SBS System Cycle (Genome Analyzer)

Step: Image A → Image C → Image G → Image T

Image flowcell channel
Repeat for all channels
Repeat for all cycles

COMPANY CONFIDENTIAL – INTERNAL USE ONLY
Parallel SBS System Cycle (HiSeq 2000)

Dramatic improvements in
A. Peak acquisition rate due to multiplexing of 4 acquisition channels and motion, and
B. System utilization from two staggered processes using independent components.
Fundamentals of Fluorescence Detection

1. Excite dye molecule with a photon
2. Detect emitted photon of a lower energy

Because absorption cross sections are small, excitation photons outnumber emitted photons by many orders of magnitude. Scattered and reflected excitation photons must be attenuated by \( \sim 10^6 \) to obtain high SNR quantification of fluorescence.

Typical schemes (like ours) use narrow band excitation (lasers) coupled with optical thin film interference filters in the detection path to efficiently transmit fluorescence while rejecting excitation.
Illumina SBS Wavelength Management

The basic scheme relies on two excitation wavelengths to excite four dyes. As these dyes have similar Stokes shifts, this creates some systematic artifacts.

The general arrangement produces:
- an efficient, clean, narrow detection channel near the excitation wavelength
- a broader, less efficient and degenerate channel further from the excitation wavelength.

Together these pairs generate four independent (but non-orthogonal) detection channels with which to classify four species.

*While not perfect, the spectral cross-talk does have significant system benefits, particularly the ability to spatially map between spectral channels.*
System Elements Governing Data Rate

*(Start with a good sequencing chemistry)*

<table>
<thead>
<tr>
<th>Throughput</th>
<th>Data Rate</th>
<th>Duty Cycle</th>
</tr>
</thead>
<tbody>
<tr>
<td>Density</td>
<td>Speed</td>
<td>Architecture</td>
</tr>
<tr>
<td>Resolution</td>
<td>Bandwidth</td>
<td>Motion</td>
</tr>
<tr>
<td>Optics</td>
<td>SNR</td>
<td>Imaging</td>
</tr>
<tr>
<td>Pixel</td>
<td>Read Noise</td>
<td>Chemistry &amp; Fluidics</td>
</tr>
<tr>
<td>Algo.</td>
<td>Protocol</td>
<td>Serial / Parallel</td>
</tr>
<tr>
<td>Size</td>
<td>Vel.</td>
<td>Frame</td>
</tr>
<tr>
<td>SNR</td>
<td>Accel.</td>
<td></td>
</tr>
</tbody>
</table>

*COMPANY CONFIDENTIAL – INTERNAL USE ONLY*
Imaging Architecture Evolution

**Genome Analyzer (2006)**
- 1 Megapixel
- 0.12 mm²
- 10MHz readout

**Genome Analyzer IIx (2008)**
- 4 Megapixel
- 0.57 mm²
- 20MHz readout

- Larger format camera
- Improves efficiency of tiling
- (80% reduction in mechanical overhead)
- Improves readout rate

**HiSeq 2000 (2010)**
- 4 x 2048 TDI CCDs
- 45.6 mm² swath
- 4x10MHz readout

- Multi-spectral TDI produces high data rate and acquisition duty cycle

**Glide 75% Duty Cycle (x 4)**

Step
~45% Duty Cycle

Step
40% Duty cycle
Data Density Evolution

**Genome Analyzer I**
50,000 / mm²

**Genome Analyzer II**
>200,000 / mm²

**Genome Analyzer IIx**
>700,000 / mm²

Reduced optical aberrations & improved algorithms

Improved surface chemistry & algorithms

20-fold evolutionary improvement over 5 years and three systems is the largest single driver in reduction of sequencing costs.

now

1,000,000 / mm²

Algorithms, chem.
Dose-Response, Noise Partition and SNR for Several Detection Schemes

<table>
<thead>
<tr>
<th></th>
<th>System A</th>
<th>System B</th>
<th>System C</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Filter</strong></td>
<td>A C G T</td>
<td>A C G T</td>
<td>A C G T</td>
</tr>
<tr>
<td><strong>Power</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(mW)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Dose</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(J/cm²)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Response</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(ke⁻² cm²/J)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>SNR</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(detection)</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Noise Contributors (Variance Partition):**

- **A**
- **C**
- **G**
- **T**

* at sample

Cycle 2 Data
Increased speed requires more efficient signal generation

- **Beam uniformity**
  
  Smoothing of laser modes to reduce dark spots and transients thereby improving data quality.

- **Beam shape**
  
  Custom design tailors power distribution to field of interest.

- **35% reduction of optical attenuation**

- **Increase input laser power**

Net: Approximately 8-fold improvement in effective power to sample to support increased data rate and quality.
Architecture of cost

Cost of Data

Consumables
- Reagents
- Flowcell
  - Density, Manufacturing, Quality

Instrument
- Amortization
  - Price, Speed, Robustness

Facilities
- Size, Efficiency, Reliability

Labor
- Prep
- Analysis
  - Automation, Ease of use, Accuracy

COMPANY CONFIDENTIAL – INTERNAL USE ONLY
A Decade of DNA Sequencing

Cost of Human Sequencing

Human Genome Project

“Moore’s Law”

Jim Watson

Yoruban Male

Everyone

$1,000,000,000

$100,000,000

$10,000,000

$1,000,000

$100,000

$10,000

$1,000

$1,000,000

2001 2002 2003 2004 2005 2006 2007 2008 2009

Untouched from 2009 and still ~ on track.
Architecture of Latency

Time to Answer

Cycle Time

Chemistry
- Incubation
- Thermal Control

Acquisition
- Fluidics
- Serial/Parallel

Read Length

Quality
- Chem. & Fluidics
- SNR

Enzyme
Additives
Rate
Dead Vol.
Sequencing by synthesis is a step-wise process, rate-limited by chemistry and chemistry support functions.

- 15 min. Better enzyme
- 28 min. Optimized incubation & temperature ramp times
- 45 min. Increased flow rate
- 55 min. Launch performance
- 5’ Reagents, process control fast fluidics and thermal cycling

10-fold evolutionary improvement in cycle times over 5 years provide for longer reads, faster time to answer (and smaller run sizes).
MiSeq – Optimizing something different
# Sequencers-R-Us

<table>
<thead>
<tr>
<th>MiSeq</th>
<th>Genome Analyzer IIx</th>
<th>HiScanSQ</th>
<th>HiSeq 1000</th>
<th>HiSeq 2000</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 FC x 5 M templates</td>
<td>1 FC x 320 M templates</td>
<td>1 FC x 750 M templates</td>
<td>1 FC x 1,500 M templates</td>
<td>2 FC x 1,500 templates</td>
</tr>
<tr>
<td>2 x 150 bp</td>
<td>2 x 150 bp</td>
<td>2 x 100 bp</td>
<td>2 x 100 bp</td>
<td>2 x 100 bp</td>
</tr>
<tr>
<td>1.5 Gb/ run</td>
<td>95 Gb/ run</td>
<td>150 Gb/ run</td>
<td>300 Gb/ run</td>
<td>600 Gb/ run</td>
</tr>
<tr>
<td>&lt; 14 Day/ run</td>
<td>&lt; 9 Day/ run</td>
<td>&lt; 9 Day/ run</td>
<td>&lt; 10 Day/ run</td>
<td>&lt; 10 Day/ run</td>
</tr>
<tr>
<td>$125,000/ system</td>
<td>$95,000-295,000</td>
<td>$405,000</td>
<td>$560,000</td>
<td>$690,000</td>
</tr>
<tr>
<td>&lt;$750/ run</td>
<td>&lt;$15,000/ run</td>
<td>&lt;$12,000/ run</td>
<td>&lt;$12,000/ run</td>
<td>&lt;$24,000/ run</td>
</tr>
</tbody>
</table>

- Cluster + Seq + analysis in one box
- Separate cBot, PE module, and computer
- Separate cBot and computer
- Separate cBot and computer

| Ships Q3 |