Transitioning to the HiSeq3000

Performance improvements over HiSeq 2000
Workflow differences – Cost benefits
Experimental design advantages

July 28th 2016
San Antonio, TX

Mehdi Keddache PhD
Sequencing Specialist
HiSeq X10
Population Scale Human Genome Sequencing

1.8T | 6B Reads | PE150 | <3 Days

$1,000 Genome* | $800 Consumables

18,000 Genomes | Year

*All estimates assume 85% capacity utilization and 4 year depreciation schedule
Patterned Flow Cell Technology

Nanowell substrate | Billions of ordered wells

- Optimal “fixed” cluster spacing
- Defined size
- Simplified imaging
- Increased cluster density
HiSeq 3000 and 4000 Systems
With patterned flow cell technology

HiSeq 3000
750 GB | 2.5B READS | PE150
3.5 DAYS

HiSeq 4000
1.5 TB | 5B READS | PE150
3.5 DAYS
Side by side performance comparison

Between HiSeq 2000 and HiSeq 3000 systems
### Maximum Output

<table>
<thead>
<tr>
<th></th>
<th>HiSeq 2000</th>
<th>HiSeq 3000</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Capacity</strong></td>
<td>3,000 M</td>
<td>1,500 M</td>
</tr>
<tr>
<td><strong>Read Length</strong></td>
<td>100</td>
<td>100</td>
</tr>
<tr>
<td><strong>Output</strong></td>
<td>600 Gb</td>
<td>300 Gb</td>
</tr>
<tr>
<td><strong>Run Time</strong></td>
<td>11 days</td>
<td>8 days</td>
</tr>
<tr>
<td><strong>Throughput</strong></td>
<td>55 GB/day</td>
<td>38 GB/day</td>
</tr>
</tbody>
</table>

*HiSeq 2000 High Output v3 Chemistry*
Maximum Output

<table>
<thead>
<tr>
<th></th>
<th>HiSeq 2000 High Output v3 Chemistry</th>
<th>HiSeq 3000</th>
</tr>
</thead>
<tbody>
<tr>
<td>Capacity</td>
<td>3,000 M</td>
<td>1,500 M</td>
</tr>
<tr>
<td>Read Length</td>
<td>100</td>
<td>100</td>
</tr>
<tr>
<td>Output</td>
<td>600 Gb</td>
<td>300 Gb</td>
</tr>
<tr>
<td>Run Time</td>
<td>11 days</td>
<td>8 days</td>
</tr>
<tr>
<td>Throughput</td>
<td>55 GB/day</td>
<td><strong>4 times faster</strong></td>
</tr>
</tbody>
</table>

$1,800 per lane 188 M clusters 2x100

$1,925 per lane 312 M clusters 2x150

4 times faster
### PE100 Run Setup

<table>
<thead>
<tr>
<th></th>
<th>HiSeq 2000 High Output v3 Chemistry</th>
<th>HiSeq 3000</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Capacity</strong></td>
<td>3,000 M</td>
<td>1,500 M</td>
</tr>
<tr>
<td><strong>Read Length</strong></td>
<td>100</td>
<td>100</td>
</tr>
<tr>
<td><strong>Output</strong></td>
<td>600 Gb</td>
<td>300 Gb</td>
</tr>
<tr>
<td><strong>Run Time</strong></td>
<td>11 days</td>
<td>8 days</td>
</tr>
<tr>
<td><strong>Throughput</strong></td>
<td>55 GB/day</td>
<td>38 GB/day</td>
</tr>
</tbody>
</table>
PE100 Run Setup

<table>
<thead>
<tr>
<th></th>
<th>HiSeq 2000 High Output v3 Chemistry</th>
<th>HiSeq 3000</th>
</tr>
</thead>
<tbody>
<tr>
<td>Capacity</td>
<td>3,000 M</td>
<td>1,500 M</td>
</tr>
<tr>
<td>Read Length</td>
<td>100</td>
<td>100</td>
</tr>
<tr>
<td>Output</td>
<td>600 Gb</td>
<td>300 Gb</td>
</tr>
<tr>
<td>Run Time</td>
<td>11 days</td>
<td>8 days</td>
</tr>
<tr>
<td>Throughput</td>
<td>55 GB/day</td>
<td>38 GB/day</td>
</tr>
</tbody>
</table>

$1,800 per lane
188 M clusters
2x100

$1,650 per lane
312 M clusters
2x100

45% off
Transitioning libraries from HiSeq 2000 to HiSeq 3000

Workflow adjustments
Kinetic Exclusion Amplification Clustering

• Seeding of library molecules in nanowells is much slower than their bridge amplification

• Only one library molecule per nanowell is detected

• Polyclonal nanowells are excluded (nPF)

• Occasionally larger fragments can “jump” to neighboring wells and create local duplicates
**Consequence of the new clustering method**
*Adapter dimers quickly create low diversity on the flowcell*

<table>
<thead>
<tr>
<th>Adapter in library</th>
<th>0.1%</th>
<th>0.5%</th>
<th>1%</th>
<th>5%</th>
<th>10%</th>
</tr>
</thead>
<tbody>
<tr>
<td>% Adapters in data</td>
<td>0.8</td>
<td>2.6</td>
<td>6.5</td>
<td>60.4</td>
<td>84.3</td>
</tr>
<tr>
<td>%PF</td>
<td>66</td>
<td>55</td>
<td>52</td>
<td>21</td>
<td>11</td>
</tr>
<tr>
<td>% Q30 R1</td>
<td>96.8</td>
<td>95.9</td>
<td>95.5</td>
<td>82.7</td>
<td>74.5</td>
</tr>
<tr>
<td>%Q30 R2</td>
<td>92.1</td>
<td>90.2</td>
<td>87.8</td>
<td>65.5</td>
<td>73.0</td>
</tr>
<tr>
<td>% Base</td>
<td><img src="image" alt="Graph" /></td>
<td><img src="image" alt="Graph" /></td>
<td><img src="image" alt="Graph" /></td>
<td><img src="image" alt="Graph" /></td>
<td><img src="image" alt="Graph" /></td>
</tr>
</tbody>
</table>

*2 x 36 cycle run*
Zymo Select-A-Size
Tunable size selection column

![Graph showing DNA sizing](image-url)
Robust performance of patterned flow cells over a broad range of input concentrations with quality libraries

- 75% PF yields 368 M reads per lane
- 50% PF yields 312 M reads per lane

Customers report very consistent lane yield
## Comparison Matrix – Paired End - Per GB

<table>
<thead>
<tr>
<th></th>
<th>PE100/125/150 Genomes</th>
<th>PE50/75 Exomes or RNA-seq</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>HD</strong></td>
<td>$ 21</td>
<td>$ 29</td>
</tr>
<tr>
<td><strong>2500 v4</strong></td>
<td>$ 32</td>
<td>$ 53</td>
</tr>
<tr>
<td><strong>2000 v3</strong></td>
<td>$ 48</td>
<td>$ 71</td>
</tr>
<tr>
<td><strong>2500 RM</strong></td>
<td>$ 55</td>
<td>$ 82</td>
</tr>
<tr>
<td><strong>NextSeq500</strong></td>
<td>$ 35</td>
<td>$ 44</td>
</tr>
</tbody>
</table>

- **2.3X**
- **2.5X**
## Comparison Matrix – Single Read - Per Million Read

<table>
<thead>
<tr>
<th></th>
<th>SR50/75</th>
<th>GEx or miR-seq</th>
</tr>
</thead>
<tbody>
<tr>
<td>HD</td>
<td>$2.4</td>
<td></td>
</tr>
<tr>
<td>2500 v4</td>
<td>$3.1</td>
<td></td>
</tr>
<tr>
<td>2000 v3</td>
<td>$4.2</td>
<td></td>
</tr>
<tr>
<td>2500 RM</td>
<td>$4.8</td>
<td></td>
</tr>
<tr>
<td>NextSeq500</td>
<td>$3.5</td>
<td></td>
</tr>
</tbody>
</table>

1.75X
Conversion to HiSeq 3000

- Offers the lowest price point for sequencing in the industry
  - Cost savings for current experiments
  - Competitiveness for grant applications and future projects

- Minimal workflow changes
  - No change in library preparation workflow
  - Optionally increases the sequencer run setup by ~1h (to remove adapters from pools)
  - No changes in bioinformatics pipeline *

- Cuts sequencing costs by about half, over v3 chemistry
Experimental designs enabled by the lower costs of sequencing

RNA applications
DNA applications
More Reads and Longer Reads for RNA-seq

- Increase limit of detection

- Quantitate alternatively spliced transcripts

Experiments with different read counts:
- 10M reads
- 20M reads
- 30M reads

Experiment size

mRNA
Short reads
Effect of deeper coverage and paired reads on mRNA-seq

Coefficient of Variation vs FPKM

- SE-50
- SE-100
- PE-50
- PE-100
- 6M
- 10M
- 20M
- 30M
More Reads and More Biological Replicates

Genome Sequencing on HiSeq 3000 – Per Lane 2x150

1 Human
3.2 Gb Genome
25X Coverage
$1,925

2 Humans
3.2 Gb Genome
12X Coverage
$962

3 Humans
3.2 Gb Genome
8X Coverage
$642

Sequencing costs only – Libraries start at $35/sample
Exome Sequencing on HiSeq 3000 – Per Lane 2x75

15 Routine Samples
50X Coverage
$91

7 Deep Samples
100X Coverage
$196

3 Somatic Samples
250X Coverage
$458

Sequencing costs only – Libraries start at $70/sample
Deeper and cheaper sequencing with HiSeq 3000

- Flexibility to increase the number of samples analyzed in RNA-seq experiments or to sequence deeper

- Increased detectability of isoforms because of longer read length

- Lowest cost possible for low pass genome sequencing, exome and ChIP/ATAC-seq
Thank you!