Next Generation Sequencing Technologies

Michael R. Crowley
Department of Genetics
Heflin Center for Genomic Sciences
Different Platforms rely on Different Technology

HiSeq, MiSeq or GAIIx by Illumina

SOLiD by Applied Biosystems

454 Lifesciences by Roche

Ion Torrent by LifeTech

Ion Proton by LifeTech

PacBio

Complete Genomics
Standard Sanger Sequencing

1. Reaction mixture
   - Primer and DNA template
   - DNA polymerase
   - dNTPs with fluorochromes
   - ddNTPs (dATP, dCTP, dGTP, dTTP)

2. Primer annealing
   - Primer and DNA template

3. Primer elongation and chain termination
   - Primer and DNA template
   - DNA polymerase
   - dNTPs with fluorochromes
   - ddNTPs (dATP, dCTP, dGTP, dTTP)

4. Primer elongation and chain termination
   - Primer and DNA template

5. Primer elongation and chain termination
   - Primer and DNA template

6. Primer elongation and chain termination
   - Primer and DNA template

7. Primer elongation and chain termination
   - Primer and DNA template

8. Primer elongation and chain termination
   - Primer and DNA template
Illumina Platforms

- UAB Stem Cell Institute

### GAIIx
- One flowcell
- ~95 billion bases sequenced
- 36bp increments
- Higher cost per base sequenced
- Single read and Paired end reads

### HiSeq2000
- Two flowcells
- ~600 billion bases sequenced
- 50bp-100bp increments
- Lower cost per base sequenced
- Single reads and Paired end reads
Flowcells through time

2005  

2010
DNA Library Prep and Flow cell Production

1kb Ladder

S-series
- Manual
- Single

DNA fragments
- Blunting by Fill-in and exonuclease
- Phosphorylation
- Addition of A-overhang
- Ligation to adapters

Illumina’s Library Preparation Workflow

Version 3 HiSeq Flow Cell
Library Assessment and Quantitation

MRL DNA Dilutions
1:1000-1:16,000
Useful Next-Gen Terms

- Cluster
  - Individual island of DNA molecules representing a single, unique template
- Clusters Passing filter
  - Number of clusters able to be distinguished by the software as individuals
- Fastq
  - DNA Sequence file that is able to be read by downstream analysis applications
- Q-Score
  - A quality score based on the Phred score from Sanger Sequencing which is the probability a base is incorrect at a given position. Example: Q30 means there is a 1:1000 chance the base is incorrect. Or stated another way it means the base call is 99.9% accurate
- Phasing/Prephasing
  - When the DNA sequencing reaction is either a base ahead or a base behind the majority of the other molecules
- Depth of Coverage
  - The average number of times a base is read within the genome
- Reads
  - Actual sequence
Illumina Cluster Generation

1. PREPARE GENOMIC DNA SAMPLE
   - Randomly fragment genomic DNA and ligate adapters to both ends of the fragments.

2. ATTACH DNA TO SURFACE
   - Bind single-stranded fragments randomly to the inside surface of the flow cell channels.

3. BRIDGE AMPLIFICATION
   - Add unlabeled nucleotides and enzyme to initiate solid-phase bridge amplification.
4. **FRAGMENTS BECOME DOUBLE-STRANDED**

The enzyme incorporates nucleotides to build double-stranded bridges on the solid-phase substrate.

5. **DENATURE THE DOUBLE-STRANDED MOLECULES**

Denaturation leaves single-stranded templates anchored to the substrate.

6. **COMPLETE AMPLIFICATION**

Several million dense clusters of double-stranded DNA are generated in each channel of the flow cell.
7. DETERMINE FIRST BASE

The first sequencing cycle begins by adding four labeled reversible terminators, primers, and DNA polymerase.

8. IMAGE FIRST BASE

After laser excitation, the emitted fluorescence from each cluster is captured and the first base is identified.

9. DETERMINE SECOND BASE

The next cycle repeats the incorporation of four labeled reversible terminators, primers, and DNA polymerase.
HiSeq can generate 600 Billion bases in one instrument run.

After laser excitation, the image is captured as before, and the identity of the second base is recorded.

The sequencing cycles are repeated to determine the sequence of bases in a fragment, one base at a time.

The data are aligned and compared to a reference, and sequencing differences are identified.
c. Paired End Sequencing

don't know what to do with the image. It looks like a diagram of some kind.

d. Mate Paired End Sequencing

Adapters containing attachment sequences (A1 & A2) and sequencing primer sites (SP1 & SP2) are ligated onto DNA fragments (e.g., genomic DNA). The resulting library of single molecules is attached to a flow cell. Each end of every template is read sequentially.

Mate pair library preparation is designed to generate short fragments consisting of two segments that originally had a separation of several kilobases in the genome. Fragments of sample genomic DNA is end-biotinylated to tag the eventual mate pair segments. Self-circularization and refragmentation of these large fragments generates a population of small fragments, some of which contain both mate pair segments with no intervening sequence. These mate pair fragments are enriched using their biotin tag. Mate pairs are sequenced using a similar two-adaptor strategy as described for paired-end sequencing.
<table>
<thead>
<tr>
<th>Index</th>
<th>Lane</th>
<th>Tile</th>
<th>Section</th>
<th>Cycle</th>
<th>Surface</th>
<th>Swath</th>
<th>Time</th>
<th>P90 A</th>
<th>P50 C</th>
<th>P90 G</th>
<th>PK</th>
</tr>
</thead>
<tbody>
<tr>
<td>35148</td>
<td>3</td>
<td>2207</td>
<td>7</td>
<td>1.00</td>
<td>Bottom</td>
<td>Middle</td>
<td>12/13/2011</td>
<td>1871</td>
<td>936</td>
<td>936</td>
<td>22</td>
</tr>
<tr>
<td>35149</td>
<td>3</td>
<td>2207</td>
<td>7</td>
<td>1.01</td>
<td>Bottom</td>
<td>Middle</td>
<td>12/13/2011</td>
<td>1840</td>
<td>929</td>
<td>919</td>
<td>22</td>
</tr>
<tr>
<td>35150</td>
<td>3</td>
<td>2207</td>
<td>7</td>
<td>1.02</td>
<td>Bottom</td>
<td>Middle</td>
<td>12/13/2011</td>
<td>1822</td>
<td>899</td>
<td>922</td>
<td>22</td>
</tr>
<tr>
<td>35151</td>
<td>3</td>
<td>2207</td>
<td>7</td>
<td>1.03</td>
<td>Bottom</td>
<td>Middle</td>
<td>12/14/2011</td>
<td>1805</td>
<td>878</td>
<td>906</td>
<td>22</td>
</tr>
<tr>
<td>35152</td>
<td>3</td>
<td>2207</td>
<td>7</td>
<td>1.04</td>
<td>Bottom</td>
<td>Middle</td>
<td>12/14/2011</td>
<td>1802</td>
<td>876</td>
<td>902</td>
<td>22</td>
</tr>
<tr>
<td>35153</td>
<td>3</td>
<td>2207</td>
<td>7</td>
<td>1.05</td>
<td>Bottom</td>
<td>Middle</td>
<td>12/14/2011</td>
<td>1795</td>
<td>841</td>
<td>908</td>
<td>22</td>
</tr>
<tr>
<td>35154</td>
<td>3</td>
<td>2207</td>
<td>7</td>
<td>1.06</td>
<td>Bottom</td>
<td>Middle</td>
<td>12/14/2011</td>
<td>1785</td>
<td>846</td>
<td>874</td>
<td>22</td>
</tr>
<tr>
<td>35155</td>
<td>3</td>
<td>2207</td>
<td>7</td>
<td>1.07</td>
<td>Bottom</td>
<td>Middle</td>
<td>12/14/2011</td>
<td>1749</td>
<td>813</td>
<td>872</td>
<td>22</td>
</tr>
<tr>
<td>35156</td>
<td>3</td>
<td>2207</td>
<td>7</td>
<td>1.08</td>
<td>Bottom</td>
<td>Middle</td>
<td>12/14/2011</td>
<td>1746</td>
<td>806</td>
<td>878</td>
<td>22</td>
</tr>
<tr>
<td>35913</td>
<td>3</td>
<td>2208</td>
<td>8</td>
<td>1</td>
<td>Bottom</td>
<td>Middle</td>
<td>12/08/2011</td>
<td>2958</td>
<td>4976</td>
<td>1651</td>
<td>35</td>
</tr>
<tr>
<td>35914</td>
<td>3</td>
<td>2208</td>
<td>8</td>
<td>2</td>
<td>Bottom</td>
<td>Middle</td>
<td>12/08/2011</td>
<td>3176</td>
<td>4752</td>
<td>1630</td>
<td>41</td>
</tr>
<tr>
<td>35915</td>
<td>3</td>
<td>2208</td>
<td>8</td>
<td>3</td>
<td>Bottom</td>
<td>Middle</td>
<td>12/08/2011</td>
<td>3103</td>
<td>4773</td>
<td>1679</td>
<td>39</td>
</tr>
<tr>
<td>35916</td>
<td>3</td>
<td>2208</td>
<td>8</td>
<td>4</td>
<td>Bottom</td>
<td>Middle</td>
<td>12/08/2011</td>
<td>3259</td>
<td>4786</td>
<td>1640</td>
<td>31</td>
</tr>
<tr>
<td>35917</td>
<td>3</td>
<td>2208</td>
<td>8</td>
<td>5</td>
<td>Bottom</td>
<td>Middle</td>
<td>12/08/2011</td>
<td>2728</td>
<td>4112</td>
<td>1533</td>
<td>25</td>
</tr>
<tr>
<td>35918</td>
<td>3</td>
<td>2208</td>
<td>8</td>
<td>6</td>
<td>Bottom</td>
<td>Middle</td>
<td>12/09/2011</td>
<td>3156</td>
<td>4605</td>
<td>1475</td>
<td>30</td>
</tr>
<tr>
<td>35919</td>
<td>3</td>
<td>2208</td>
<td>8</td>
<td>7</td>
<td>Bottom</td>
<td>Middle</td>
<td>12/09/2011</td>
<td>2712</td>
<td>4312</td>
<td>1404</td>
<td>38</td>
</tr>
</tbody>
</table>

Rows=41472  Disp=564  Sel=1  Filter
Percent Q30 Scores per cycle for all lanes and both surfaces
454 LifeSciences

With Titanium Chemistry can generate up to 1000bp/template
700 Million bases/run
**Sequencing**

7.5 hours

- Well diameter: average of 44 μm
- 400,000 reads obtained in parallel
- A single cloned amplified ssDNA bead is deposited per well

Amplified ssDNA library beads → Quality filtered bases
Pacific Biosciences Technology

Complete Genomics Technology

R Drmanac et al. Science 2010;327:78-81
Next Next-Gen Sequencing (3G Seq)

Figure 1. Third-generation sequencing platforms. (a) Pacific Biosciences SMRT (single-molecule real-time) DNA sequencing method. The platform uses a DNA polymerase anchored to the bottom surface of a ZMW (pictured in cross section). Differentially labeled nucleotides enter the ZMW via diffusion and occupy the ‘detection volume’ (white translucent halo area) or microseconds. During an incorporation event, the labeled nucleotide is ‘held’ within the detection volume by the polymerase for tens of milliseconds. As each nucleotide is incorporated, the label, located on the terminal phosphate, is cleaved off and diffuses out of the ZMW. (b) Life Technologies FRET sequencing platform uses base fluorescent labeling technology, a DNA polymerase modified with a quantum dot and DNA template molecules immobilized onto a solid surface. During an incorporation event, energy is transferred from the quantum dot to an acceptor fluorescent moiety on each labeled base. Light emission can only emanate from labeled nucleotides as they are being incorporated. (c) The Oxford nanopore sequencing platform uses an exonuclease coupled to a modified α-hemolysin nanopore (purple, pictured in cross section) positioned within a lipid bilayer. As sequentially cleaved bases are directed through the nanopore, they are transiently bound by a cyclodextrin moiety (blue), disturbing current through the nanopore in a manner characteristic for each base. (d) The Ion Torrent sequencing platform uses a semiconductor-based high-density array of microwell reaction chambers positioned above an ion-sensitive layer and an ion sensor. Single nucleotides are added sequentially, and incorporation is recorded by measuring hydrogen ions released as a by-product of nucleotide chain elongation.