Illumina NGS Technology

Pernille Albertus
Sr Marketing Technical Specialist
palbertus@illumina.com
Illumina, US company, head quarters in California

- >4,100 employees
- 90% sequencing data
- 1,000$ human genome
- 85,000 publications
Platform evolution
- *the original* Genome Analyzer (2007)
Platform evolution
Platform evolution
- the MiSeq (2011)
The markets we serve

- Genetics
- Cancer
- Consumer Genomics
- Reproductive Health
- Agriculture
- Research
- Forensic Genomics
- Microbiology
- BioPharma
The Illumina NGS workflow
The Illumina NGS Workflow

1. Isolate DNA
2. Prepare library
3. Run sequence
4. Analyze data

Sanger Sequencing
Library

isolate DNA (or RNA)

fragment (<1000 bp)

flank DNA with adapter sequence

Index 1 = 6-base barcode unique to this sample

Index 1 sequencing primer annealing site

Index 2 = 6-base barcode unique to this sample

Index 2 sequencing primer annealing site

Read 1 sequencing primer annealing site

Read 2 sequencing primer annealing site

PCR primer annealing site

P7 (complementary to oligo on flow cell surface)

P5 (complementary to oligo on flow cell surface)
TruSeq library prep workflow

A. Fragment genomic DNA

Aii. Double-stranded cDNA (from figure 2B)

B. End repair and phosphorylate

C. A-tailing

D. Ligate index adapter

E. Denature and amplify for final product
TruSeq Nano workflow (manual)

1. Quantify DNA
2. Shear DNA
3. Cleanup
4. End repair
5. Cleanup & size
6. Adenylate
7. Ligate adapters
8. Cleanup
9. PCR
10. Cleanup
11. Quantify
12. QC library
TruSeq Nano workflow (automated)

1. Quantify DNA
2. Shear DNA
3. NeoPrep
Nextera library prep workflow
NexteraXT hardware workflow

1. Quantify DNA
2. Tagment
3. Index & PCR
4. Clean & Size
5. Normalize
6. Pool & load
A deep dive into SBS chemistry
Massively parallel sequencing

1. DNA (0.05-1.0 µg)
2. Library preparation
3. Cluster growth
4. Sequencing
5. Image acquisition
6. Base calling

DNA: TGCTACGAT...
Hybridize fragment and extend

- Single DNA libraries are hybridized to the primer lawn
- Bound libraries are then extended by polymerases
Denature double-stranded DNA

- The double-stranded molecule is denatured
- The original template is washed away
Anchor the template to the surface

- The newly synthesized strand is covalently attached to the flow cell surface
Bridge amplification

- The single-stranded molecule flips over and forms a bridge by hybridizing to an adjacent, complementary primer.
- The hybridized primer is extended by polymerases.
Bridge amplification

- A double-stranded bridge is formed
Denature the double-stranded bridge

- The double-stranded bridge is denatured
- Result: Two copies of covalently bound single-stranded templates
Bridge amplification

- Single-stranded molecules flip over to hybridize to adjacent primers
- Hybridized primer is extended by polymerase
Bridge amplification

- Bridge amplification cycle repeated until multiple bridges are formed
Linearization

- ds-DNA bridges are denatured
- Reverse strands are cleaved
Reverse strand cleavage

- .. and washed away, leaving a cluster with forward strands only
Blocking

- Free 3’ ends are blocked to prevent unwanted DNA priming
A sequencing primer is introduced to the flow cell and hybridized to the adapter sequence annealing site.
Sequencing by synthesis (SBS) close-up

1. Fluorescently labeled ddNTPs are added together with a polymerase.

2. A base is incorporated; unincorporated nucleotides are washed away.

3. Light excites fluorophore.

4. Camera captures emission colour.

5. Fluorophore is cleaved off and terminator is removed.

6. The next cycle begins...

CYCLE 1 CYCLE 2 CYCLE 3 CYCLE 4 CYCLE 5 CYCLE 6 CYCLE 7 CYCLE 8 CYCLE 9

TGCTACGAT

BASE CALLING
Paired-end sequencing

- After completion of sequencing of the forward strand, the sequenced product is stripped off
- 3’-ends of template strands and lawn primer are unblocked
Paired-end sequencing

- Single-stranded template loops over to form a bridge by hybridizing with a lawn primer
- 3’-ends of lawn primer is extended (double-stranded stretch allows the polymerase to build the complementary strand)
Paired-end sequencing

Double stranded DNA
Paired-end sequencing

- Bridges are linearized and the original forward template is cleaved off
Paired-end sequencing

- Free 3’-ends of the reverse template and lawn primers are blocked to prevent unwanted DNA priming.

- The Read 2 sequencing primer is introduced to the flow cell and hybridized to the adapter sequence annealing site.
Dual index sequencing

Region complementary to P5 grafting primer

P5 index

P5 primer region

Insert

SBS sequencing primer

Sequence

P7 primer region

P7 index primer

P7 index

Sequence 8 bp

P7 grafting primer

Flow cell surface

36
Dual index sequencing

Flow cell surface

Continue into Read 2 resynthesis
8 cycles sequencing P5 index
7 dark cycles (chemistry with no imaging)
P5 grafting primer
Number of reads and depth

Raw data:
8 clusters produce a total of 16 sequences = reads (paired-end sequencing)

Filtered, high quality data:
6 clusters, corresponding to 12 reads

The reads are mapped to a reference genome. Depth is the number of reads that align to a specific base positions in the reference genome. A few examples of depth are indicated (i.e. 2x, 5x, 1x, and 3x)
How many samples can you pool together in a single run/lane?

- Pooling samples depends on genome size, coverage, and read length

<table>
<thead>
<tr>
<th>organism</th>
<th>genome size</th>
<th>coverage</th>
<th>read length</th>
<th>no of samples</th>
</tr>
</thead>
<tbody>
<tr>
<td>Staphylococcus aureus (MRSA)</td>
<td>2.8 Mb</td>
<td>100x</td>
<td>2x300</td>
<td>45</td>
</tr>
<tr>
<td>Mycobacterium tuberculosis (TB)</td>
<td>4.4 Mb</td>
<td>100x</td>
<td>2x300</td>
<td>29</td>
</tr>
<tr>
<td>E. coli</td>
<td>4.6 Mb</td>
<td>100x</td>
<td>2x300</td>
<td>27</td>
</tr>
<tr>
<td>Plasmodium falciparum (malaria)</td>
<td>22.9 Mb</td>
<td>100x</td>
<td>2x300</td>
<td>5</td>
</tr>
</tbody>
</table>

*1 lane HO
Running costs – examples from publications

- **Plasma HIV RNA**
  - HIV-RT amplicons
  - MiSeq 2x250
  - Variant frequency detected >0.5

- **RNA virus**
  - without culture
  - MiSeq 2x150
  - 36-hours from sample to answer

- **MRSA outbreak**
  - MiSeq 2x150
  - 48-hours from sample to answer

- **Tuberculosis detection and susceptibility**
  - MiSeq 2x150
  - 5-day TAT
Running costs – examples from publications

- **31 US$**
  - Plasma HIV RNA
  - HIV-RT amplicons MiSeq 2x250
  - variant frequency detected >0.5

- **134 US$**
  - RNA virus without culture MiSeq 2x150
  - 36-hours from sample to answer

- **97 £**
  - MRSA outbreak
  - MiSeq 2x150
  - 48-hours from sample to answer

- **150 US$**
  - Tuberculosis detection and susceptibility
  - MiSeq 2x150
  - 5-day TAT
Illumina – driving the future of genomics