Next Generation Sequencing

Josef K Vogt
Slides by: Simon Rasmussen

2017
Second generation sequencing

454

Illumina

SOLiD

$10^6$-$10^9$

Ion Torrent (PGM)

90% market share
Library preparation

1. Create library molecules
2. Amplification (PCR)
3. Massive parallel sequencing

DNA from extract → Fragment & polish DNA → Adapters → Library molecule
Amplification and immobilization

Emulsion PCR (454, Solid, IonTorrent): Water, oil, beads, one DNA template/droplet

Bridge PCR (Illumina): One DNA template/cluster, primers on surface, grow by bridging primers

Bridge PCR

Metzker, NatGen Rev. 2010
Fluorescence detection

Illumina - Cyclic reversible termination

Add all dNTPs labelled w. diff dye
Create four-color image
Cleave dye and repeat next cycle

454 - Pyrosequencing

Load template beads into wells
Flow one dNTP across wells
Polymerase incorporates nucleotide
Release of PPi leads to light
Imaging next dNTP

Metzker, NatGen Rev. 2010
2G: Imaging handout

Illumina 1:__________
Illumina 2:__________

454: ________________
_____________________

Metzker, NatGen Rev. 2010
Quality of base call deteriorates after many cycles

Homopolymer runs are problematic, gives rise to indels

Metzker, NatGen Rev. 2010
Illumina: Quality deterioration

Can you think of why?

Efficiency of incorporation:
Polymerase incorporation of base
Enzyme that cleaves the dye
NextSeq/HiSeq3000/4000

- Chemistry is not based on 4 dyes (as before) but 2 dyes
  - T (red), C (green), A (both) and G (none = “dark”)
- Faster processing rate and cheaper reagents
- Slightly increases error rate
- Problem with G stretches because G is not dyed
Ion Torrent

Similar principle to 454
Library: Emulsion PCR
Based on semiconductors
Detection is based on H ions (pH) changes
a Life/AGP — Sequencing by ligation

Universal seq primer (n)

1,2-probes
x, y Interrogation bases
n Degenerate bases
z Universal bases

3' xynnnzzz 5'
3' xynnnzzz 5'
3' xynnnzzz 5'
3' xynnnzzz 5'

PI adapter

3' AT

Target sequence

3' POH

Ligase

Excite

Fluorescence, four-colour imaging

Cleavage agent

Universal seq primer (n – 1)

3' TA

PI adapter

1 base shift

Reset primer three more times

Repeat ligation cycles

Ligation cycle 1 2 3 4 5 6 7.. (n cycles)
Solid example

Double-base encoding
Colorscape
Low error rate
Errors propagates

@SRR349943.1 solid0420_20100825_FRAG/1
T302123023003302121122231210022323232112002222302010
+
!=:369A?:.<9=.-5=%3-:6%3&<2%(169%,0.3%&'(&.'%%%&&,
AAGT...
**Complete Genomics**

**ssDNA -> DNA nanoballs**

Place DNBs into each spot

**Use silicon chips with sticky spots**

Sequence using ligase and fluorescent labeled probes

You can't buy the machine - Acquired by BGI - delayed - Only humans!
3rd generation

No amplification (PCR introduces bias!)
Simple sample preparation

Helicos | Pacific Biosciences | Oxford Nanopore
Slowed down DNA polymerase, measure light emission, Long reads > 10kb, high error rate (but random)
Oxford Nanopore

Nano-scale pores, with current across
Drag DNA stretch through the pore
Measure change in current (pentamers)
Long reads (up to 200k), currently ~5k, some systematic errors
Synthetic Long Reads (2nd gen)

- Illumina Synthetic long-read sequencing (Moleculo)
- 10X Genomics
- Based on Illumina sequencing
- Using barcode system to create linked reads / read clouds

10X Genomics

Emulsion PCR
Arbitrarily long DNA is mixed with beads loaded with barcoded primers, enzyme and dNTPs

GEMs
Each micelle has 1 barcode out of 750,000

Amplification
Long fragments are amplified such that the product is a barcoded fragment ~350 bp

Pooling
The emulsion is broken and DNA is pooled, then it undergoes a standard library preparation

Linked reads
* All reads from the same GEM derive from the long fragment, thus they are linked
* Reads are dispersed across the long fragment and no GEM achieves full coverage of a fragment
* Stacking of linked reads from the same loci achieves continuous coverage
## Machine overview - I

<table>
<thead>
<tr>
<th>Company/technology</th>
<th>Current machine, key characteristics</th>
</tr>
</thead>
<tbody>
<tr>
<td>454</td>
<td>GS FLX+, 7-800bp length, 1M reads</td>
</tr>
<tr>
<td>Illumina</td>
<td>HiSeq400/HiSeqX Ten, 100-150bp length, up to 2-4B reads</td>
</tr>
<tr>
<td>Solid (Life)</td>
<td>5500XL, 50/75bp length, 1.5B reads</td>
</tr>
<tr>
<td>PGM (Life)</td>
<td>Ion Proton, 200/400bp length, 80M reads</td>
</tr>
<tr>
<td>Complete Genomics (BGISEQ)</td>
<td>BGISEQ-500, 50-100 bp, 200Gb in total</td>
</tr>
<tr>
<td>PacBio</td>
<td>Sequel, 8-12kb length, 350k reads</td>
</tr>
<tr>
<td>Oxford Nanopore</td>
<td>GridION, up to 200kb, &gt;100k reads</td>
</tr>
<tr>
<td>Illumina synthetic</td>
<td>up to 100kb synthetic length, 1000$ pr. Gb</td>
</tr>
<tr>
<td>10X Genomics</td>
<td>up to 100kb synthetic length, +500$ per sample</td>
</tr>
</tbody>
</table>

Excellent overview at Goodwin et al., Nature Reviews (2016)
# Machine overview - II

## Benchtop machines

<table>
<thead>
<tr>
<th>Company/technology</th>
<th>Current machine, key characteristics</th>
</tr>
</thead>
<tbody>
<tr>
<td>454</td>
<td>GS Junior, 4-500bp length, 100k reads</td>
</tr>
<tr>
<td>Illumina</td>
<td>MiSeq, 300bp length, 50M reads</td>
</tr>
<tr>
<td>PGM (Life)</td>
<td>Ion Torrent, 400bp length, 5M reads</td>
</tr>
<tr>
<td>Oxford Nanopore</td>
<td>MinION, up to 200kb, 100k reads</td>
</tr>
</tbody>
</table>

Excellent overview at Goodwin et al., Nature Reviews (2016)
More NGS material

- Elaine Mardis on NGS technology
- Youtube has many more …

- Excellent review on NGS technologies: Goodwin et al., Nature Reviews (2016)
- On Campusnet together with many other papers!