Ultrasequencing: Methods and Applications of the New Generation Sequencing Platforms

Laura Moya Andé rico
Master in Advanced Genetics
Genomics Class
December 16th, 2015
Brief Overview

* First-generation sequencing— Sanger
  • Sample preparation: DNA fragments are labeled with dyes
  • Capillary electrophoresis: arranges fragments by length
  • Average read length ~800 bp
  • Expensive, low throughput, high turnaround time

* Second-generation sequencing— Illumina
  • Higher throughput
  • “Wash-and-scan” process
  • Higher turnaround time
  • Requires PCR amplification prior to sequencing
  • Dephasing yields many problems
Transition to TGS

* Ion Torrent semiconductor sequencer
  - Simplifies sequencing process
  - Pros and cons

* Helicos Genetic Analysis Platform
  - First available instrument with SMS technology
  - High turnaround time but no PCR is required
  - Process is similar to “wash-and-scan” in SGS

* Single molecule sequencing—improved sequencing by synthesis technology (SBS) used in SGS
## Table Comparing First, Second, and Third Generation Sequencing

<table>
<thead>
<tr>
<th></th>
<th>First generation</th>
<th>Second generation</th>
<th>Third generation</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Fundamental technology</strong></td>
<td>Size-separation of specifically end-labeled DNA fragments, produced by SBS or degradation</td>
<td>Wash-and-scan SBS</td>
<td>SBS, by degradation, or direct physical inspection of the DNA molecule</td>
</tr>
<tr>
<td><strong>Resolution</strong></td>
<td>Averaged across many copies of the DNA molecule being sequenced</td>
<td>Averaged across many copies of the DNA molecule being sequenced</td>
<td>Single-molecule resolution</td>
</tr>
<tr>
<td><strong>Current raw read accuracy</strong></td>
<td>High</td>
<td>High</td>
<td>Moderate</td>
</tr>
<tr>
<td><strong>Current read length</strong></td>
<td>Moderate (800–1000 bp)</td>
<td>Short, generally much shorter than Sanger sequencing</td>
<td>Long, 1000 bp and longer in commercial systems</td>
</tr>
<tr>
<td><strong>Current throughput</strong></td>
<td>Low</td>
<td>High</td>
<td>Moderate</td>
</tr>
<tr>
<td><strong>Current cost</strong></td>
<td>Low cost per base</td>
<td>Low cost per base</td>
<td>Low-to-moderate cost per base</td>
</tr>
<tr>
<td><strong>High cost per run</strong></td>
<td>High cost per run</td>
<td>Low cost per run</td>
<td>Low cost per run</td>
</tr>
<tr>
<td><strong>RNA-sequencing method</strong></td>
<td>cDNA sequencing</td>
<td>cDNA sequencing</td>
<td>Direct RNA sequencing and cDNA sequencing</td>
</tr>
<tr>
<td><strong>Time from start of sequencing reaction to result</strong></td>
<td>Hours</td>
<td>Days</td>
<td>Hours</td>
</tr>
<tr>
<td><strong>Sample preparation</strong></td>
<td>Moderately complex, PCR amplification not required</td>
<td>Complex, PCR amplification required</td>
<td>Ranges from complex to very simple depending on technology</td>
</tr>
<tr>
<td><strong>Data analysis</strong></td>
<td>Routine</td>
<td>Complex because of large data volumes and because short reads complicate assembly and alignment algorithms</td>
<td>Complex because of large data volumes and because technologies yield new types of information and new signal processing challenges</td>
</tr>
<tr>
<td><strong>Primary results</strong></td>
<td>Base calls with quality values</td>
<td>Base calls with quality values</td>
<td>Base calls with quality values, potentially other base information such as kinetics</td>
</tr>
</tbody>
</table>

Third-generation Sequencing

* Single-molecule real-time sequencing — PacBio

- Observes single molecule of DNA polymerase synthesize a strand of DNA using zero-mode waveguide technology
- Pros: fast turn-around time, requires minimal amounts of reagents/sample preparation, no PCR amplification needed, >1000bp average read lengths
- Cons: >5% read error rates, low throughput
Third-generation Sequencing (cont.)

* Single-molecule real-time sequencing — PacBio

Figure 2. Srinivasan S, Batra J (2014). *Next Generat Sequenc & Applic* 1: 107.
Third-generation Sequencing (cont.)

* Single molecule sequencing — Transmission electron microscope (TEM) by Halcyon Molecular

  - Directly images and chemically detects atoms that identify the nucleotides comprising a DNA template
  - DNA bases are detected with scanning tunneling microscopy
  - Challenges: requires preparation of stretched ssDNA on a surface, expensive microscopes are needed, and requires more specific equipment
TEM-based DNA sequencing using electron microscopy — ZS Genetics

- Directly visualizes DNA by labeling atoms within the nucleotides using annular dark-field scanning transmission electron microscopy (ADF-STEM)
- Promises >10kb base reads at a rate of 1.7 billion bases a day
- Challenges: reduce label losses, increase detectable differences between DNA base types, and increase the fraction of bases labeled within a molecule
Third-generation Sequencing (cont.)

* TEM-based DNA sequencing using electron microscopy — ZS Genetics

Figure 1a, 3b. Bell David, et al. 2012. Microscopy and Microanalysis 18(5):1049-1053.
Third-generation Sequencing (cont.)

* Nanopore Sequencing — Oxford Nanopore
  - Single DNA molecules pass through a nanopore chamber
  - Uses microwells within an array chip for sample preparation, detection, and analysis
  - Predictions: Billon DNA bases in ~6 hours for ~800 €
  - Nanopore sequencing variations: exonuclease-assisted, NanoTag-SBS, Optipore, sequencing by electronic tunneling, among others
Nanopore Sequencing — Oxford Nanopore

FIGURE 2 | Schematic illustration of a nanopore sequencing device. (A) A U-tube supports the lipid bilayer membrane bathed in electrolyte solution in which a 120 mV bias is applied. During DNA translocation through the nanopore, ionic current is recorded by a PCA connected to the cis (negative) and trans (positive) chambers. (B) When an ssDNA molecule traverses through α-HL from cis to trans chamber, the open pore current drops to four different levels ($I_b$: current blockage) each at a certain time ($\tau$: dwell time). (C) Current signals could reveal sequence information (Deamer and Branton, 2002; Bayley, 2006). Reproduced by copyright permissions of American Chemical Society and Elsevier.

that MspA pore had great potential to sequence DNA (Derrington et al., 2010). The feature of single recognition site of MspA pore seems to be more advantageous than α-HL (Manrao et al., 2011). This group replaced negatively-charged amino acids with neutral asparagine residues in pore’s constriction site, and with positively charged basic residues in pore’s entrance through genetic engineering (Butler et al., 2008), which enabled easy DNA capture and deceleration of DNA translocation through the pore. They later demonstrated that the engineered MspA pore exhibited better base resolution than α-HL pore by generating larger signal difference between bases (Derrington et al., 2010). However, development of new methods is needed to avoid signal overlapping between different bases, particularly deoxynucleotides adenine and guanine (Derrington et al., 2010; Manrao et al., 2011), to increase accuracy. For precise SBR, the length of the recognition region of a nanopore shouldn’t be larger than $\sim 0.5$-nm, equivalent to phosphorus-phosphorus distance of a nucleotide (or base spacing) in an ssDNA strand (Wilson et al., 2009; Cherf et al., 2012). The constriction region of MspA is about 0.6 nm long (Manrao et al., 2012), which means signal interference from adjacent bases (Manrao et al., 2011, 2012; Laszlo et al., 2013, 2014). Their previous work showed that about four bases together around the constriction region contribute to the overall current blockage (Manrao et al., 2011, 2012). Recently, they have resorted to tetramer maps, which are the standard electric signal curves collected by measuring current blockage signals when each of the combination of 256 possible 4-mers is translocating through the pore, and algorithms to circumvent this problem (Laszlo et al., 2014).

Single base resolution by graphene and other solid-state nanopores. Besides α-HL, researchers have long been investigating solid-state nanopores, with an initial hope to tackle problems in the protein nanopores, such as instability and dimension-tuning difficulties. In 2001, Golovchenko, Branton and colleagues demonstrated that nanopores as small as 1.8 nm in diameter...
Third-generation Sequencing (cont.)

* TGS Applications
  * Gene expression analysis
  * ChIP-seq
  * De novo genome sequencing
  * Metagenomics
  * Non-invasive prenatal testing
  * Disease gene identification
  * Pharmacogenomics

* All applications require a minimal amount of reads of a pre-defined length to be able to formulate proper conclusions
Conclusions

* Sequencing techniques have progressed greatly within a short time span
  - Higher throughput
  - Longer read lengths
  - More economical — closer to the $100 genome

* DNA sequencing is not the only utility with TGS technologies

* TGS has many applications — will they be added to the clinical diagnostic setting?

* Must have collaboration between scientists and informatics to keep up with the large amount of data generated
References


Thank you for listening!

Questions?