Comparison of Library Prep Kits for Low Input Whole Exome Sequencing

B Marosy, B Craig, H Ling, KF Dohney
Center for Inherited Disease Research (CIDR), IGM, JHU-SOM, Baltimore, MD

Introduction
The Center for Inherited Disease Research (CIDR) provides high quality next-generation sequencing (NGS), genotyping and statistical genetics consultation to investigators working to discover genes that contribute to disease. CIDR is continually seeking new ways to improve its workflow and generate high quality data. Current challenges that were addressed to provide improved service for NGS include:

- Low Input - Historically, sample requirements for a variety of targeted selection and NGS platforms have been a limiting factor for investigators. Recently a number of commercially available kits offer reduced inputs to below 50ng. Here we compare the performance of the Kapa HyperPrep (Fig 1a) and Swift Accel-NGS™ 25 (Fig 1d) library prep kit.

- Small fragment size - Due to degradation, small fragment size for some samples presents a challenge for generating libraries using traditional workflows. The Swift Accel-NGS™ 15 DNA Library Kit uses single stranded ligation to prepare libraries (Fig 1c). The Kapa Hyper prep kit combines End Repair/A-Tailing into one step without a clean up prior to ligation (Fig 1a). Both kits provide for alternative workflows/chemistry that allow for smaller fragments to be successfully modified into libraries for whole exome capture and sequencing.

Methods
Library Prep – Low Input
50ng of HapMap DNA (NA12891 & NA12892) was used as input into library prep using the Swift 25 and Kapa Hyper Prep library kits and processed according to the manufacturer’s protocol1,2. An additional 8 experimental samples (50ng DNA input) were processed using the Kapa Hyper Prep kit. Additional data generated from these two HapMap samples was used from a previously processed project using CIDR’s current production protocol (Fig 1d)3,4.

Library Prep – Degraded/Small Fragment Size
2 HapMap samples (NA12891 & NA12892) underwent 40min of shearing using the Covaris® E210 to generate fragment sizes ~100bp in size. 500ng of DNA was used as input into library prep using the Swift 15 and Kapa Hyper Prep library kits and processed according to the manufacturer’s protocol5. An experimental sample – DNA extracted from a 38yr old tissue slide (~500ng DNA input) was processed using the Kapa Hyper prep kit, no shearing was performed on this sample since initial QC showed that the sample was highly fragmented ~100bp (data not shown).

Capture/Sequencing/Data Analysis
All samples were hybridized according to the Agilent SureSelectXTMXT protocol using either the Agilent Human All exon V4 or V5-UTR capture. Modifications to the hyb protocol included exchanging the universal blocker for base/base blocking to accommodate the indexed adapters used during ligation. Libraries were then clustered and sequenced on the Illumina® HiSeq™ 2500 platform using on-board clustering and Rapid SBS chemistry, 2x100. Sequencing data for the Low Input experiments was downsampled (Picard; DwnsampleSam) to the experiment with data analysis was performed using CIDRSeqSuite v3.0. Modifications to the pipeline for the degraded sample included trimming raw FASTQ files prior to alignment using CutAdapt (http://code.google.com/p/cutadapt/) to remove the default Illumina adapter sequences.

Low Input Experiments – GC/AT rich regions
To assess performance across GC/AT rich regions GC plots were compared (Fig 2). GC plots indicated similar profiles across both kits for normalized coverage and base quality at GC/AT rich regions.

Low Input Experiments - First Exon Analysis
First exons are typically higher in GC content. Per target interval metrics were filtered for first coding exons and binned according to mean coverage (8x vs ≥8x; Table 2). Counts of first exon intervals according to bin were compared between library kits.

Table 1. First Exon Coverage

<table>
<thead>
<tr>
<th>Sample</th>
<th>Kapa Hyper Prep</th>
<th>Swift 15</th>
</tr>
</thead>
<tbody>
<tr>
<td>NA12891</td>
<td>145 0.6 23745</td>
<td>99.4 437 1.8 22453 98.2</td>
</tr>
<tr>
<td>NA12892</td>
<td>91 0.4 22790</td>
<td>99.4 430 1.7 22457 98.2</td>
</tr>
<tr>
<td>NA12893</td>
<td>598 2.5 22292  97.5</td>
<td></td>
</tr>
</tbody>
</table>

Discussion/Conclusions
CIDR is committed to continued optimization and evaluation of new technology that will enable service to provide high quality data from a variety of platforms.

- Lowering DNA requirements is crucial for precious samples. High quality/low input DNA samples perform comparably to higher input samples. How this compares to FFPE samples/low quality DNA is a future consideration.

- Small fragment sized DNA can be used to generate libraries for NGS, however lower quality samples will impact data generation. Additional methods for repairing DNA need to be tested in conjunction with these low input methods to further evaluate the performance of these kits using low inputs of DNA.

Which library prep method to use may depend on a number of variables (cost differences, streamlined reagents, benefits from specific chemistry - single stranded adapter ligation). With any platform it is important to validate overall performance and quality within a laboratory and allow for optimization to account for site specific preferences and challenges that may arise when processing, in order to provide high quality sequencing data.

References:

www.cidr.jhmi.edu