Comparison of Library Prep Kits for Low Input Whole Exome Sequencing

B Marosy, B Craig, H Ling, KF Doheny

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™ 2S (Fig 1b) 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™ 1S 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 2S and Kapa Hyper Prep library kits and processed according to the manufacturer's protocol[1,2]. 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 CIDRs 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 1S and Kapa Hyper Prep library kits and processed according to the manufacturer's protocol[5]. 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 SureSelect™ XT 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 Run SBS chemistry, 2x100. Sequencing data for the Low Input experiments was downsampled (Picard; DownsampleSam) to the experiment with the lowest raw data yield and the intersection of the BED files between the capture products was used to ensure equal comparison between methods. Subsequent 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.

Figure 1: Library Prep Workflows for 1a) Kapa Hyper Prep Kit; 1b)Swift2S Library prep; 1c)Swift1S; 1d)CIDR's standard workflow using a modified Agilent XT workflow.

Figure 1a – Kapa Hyper Prep workflow

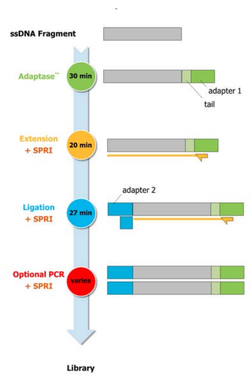

Figure 1c – Swift 1S workflow

Figure 1d – CIDR's standard workflow

Figure 1b – Swift2S workflow

Data

Low Input Experiments - Sequencing QC Metrics

Table 1 compares QC metrics generated during library prep, capture and sequencing. Overall – QC metrics are comparable between the three methods. Insert size for the Kapa Hyper prep method was lower than Swift 1S and CIDR's standard method. Percent duplication was similar between methods when comparing HapMap samples, however the experimental samples had a higher duplication rate. Percent selection metrics were based on data prior to downsampling.

Table 1. Library Prep, Capture and Sequencing Metrics

	Standard	LowInput		
Experiment	CIDR	KapaHyperPrep HapMap	Swift2S HapMap	Experimental KapaHyper
Library and Capture Metrics				
Sample Number	2	2	2	8
input dna [ng]	1000	50	50	50
# of PCR cycles - preHYB	8	8	8	8
Yield at PCR Amp [ng]	964	1513	635	1114
# of PCR cycles - postHYB	10	10	10	10
Capture	v5+UTR	v5+UTR	v5+UTR	v4
% Selection	**74.26**	**80.13**	**79.42**	**85.42**
Sequencing Data Quality				
Concordance	100	99.8388	99.8177	99.8148
Sensitivity (Het)	95.4471	94.7577	96.6945	96.7783
TiTv (ALL_Ratio)	**3.0988**	**3.1473**	**3.1012**	**3.0305**
Sequencing Data Output				
total reads (millions)	50.894	51.754	51.756	77.089
% on target 10x	**92.66**	**92.28**	**94.63**	**98.11**
% on target 20x	74.23	75.87	79.7	94.65
% zero bases	0.23	0.41	0.24	0.18
Raw GB	5.089	5.175	5.176	7.709
PF unique GB	4.742	4.459	4.848	7.038
Mean Insert Size	**299**	**233**	**270**	**227**
% Dups	**2.32**	**2.41**	**2.38**	**6.28**
Library Size (millions)	558	514	599	347
Mean Target Coverage	36	39	40	90
Read Length	100	100	100	100
plexed/sequencing	9/FC	6/lane	6/lane	5/lane

Low Input Experiments – GC/AT rich regions

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

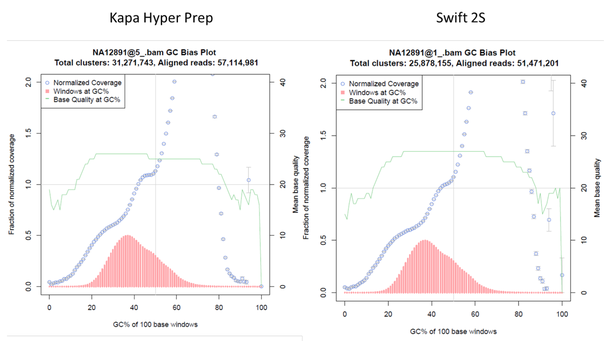

Figure 2: GC plots (Picard; CollectGcBiasMetrics.jar) of the HapMap samples tested were compared by library prep kit used (Left – Kapa; Right – Swift2S). The left y-axis depicts normalized coverage across % GC windows, and the right y-axis plots mean base quality across %GC windows.

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. Standard CIDR workflow covers first exons ~99% ≥8x (previous data not shown).

Table 2. First Exon Coverage

	Kapa Hyper Prep				Swift 2S			
Sample	Count	%<8x	Count	%≥8x	Count	%<8x	Count	%≥8x
NA12891	807	3.38	23083	**96.62**	322	1.35	23568	**98.65**
NA12892	1018	4.26	22872	**95.74**	319	1.34	23571	**98.66**
Experimental (n=8)	29	0.12	23302	**99.88**				

Low Input Experiments – Concordance

We compared the variant calls between each method (Table 3) using HapMap NA12892. More calls were missing by Kapa when compared to other methods. However when comparing to the Experimental-HapMap, which used Kapa processing, the Exp-HapMap had fewer missing calls.

Table 3. Concordance

	CIDR (A) v Swift (B)		CIDR (A) v Kapa (B)		CIDR (A) v Exp-HapMap (B)		Swift (A) v Exp-HapMap (B)		Kapa (A) v Exp-HapMap (B)	
Comp	Count	%	Count	%	Count	%	Count	%	Count	%
Diff	60	0.12	61	0.14	80	0.15	90	0.17	81	0.16
MissByA	2568	5.03	2095	4.14	4225	8.02	3555	**6.77**	5605	**10.74**
MissByB	2064	**4.04**	3971	**7.85**	1939	**3.68**	1773	3.37	1443	2.76
Same	46354	**90.81**	44446	**87.88**	46459	**88.15**	47119	**89.69**	45078	**86.34**

Degraded/Small Fragment Size Sample Experiments – Sequencing QC Metrics

Table 4 compares QC metrics generated during library prep, capture and sequencing. Comparing methods between lab generated degraded samples, Kapa had a higher sensitivity than Swift1S. Duplication rate was lower with the Swift 1S libraries. The experimental sample had a much lower TiTv ratio and higher duplication rate. Further analysis of this sample revealed a large amount of degradation (data not shown) which impacted the quality of the data. This was expected due to the age and condition of the sample.

Table 4. Library Prep, Capture and Sequencing Metrics

	Degraded		
Experiment	Degraded Kapa	Degraded Swift1S	Experimental Kapa
Library and Capture Metrics			
Sample Number	2	3	1
input dna [ng]	500	500	500
# of PCR cycles - preHYB	8	8	8
Yield at PCR Amp [ng]	524	711	607
# of PCR cycles - postHYB	10	10	10
Capture	v5+UTR	v5+UTR	v4
% Selection	**81.83**	**66.63**	**48.65**
Sequencing Data Quality			
Concordance	99.7489	99.5694	na
Sensitivity (Het)	86.3392	78.9622	na
TiTv (ALL_Ratio)	3.1265	3.1312	2.0063
Sequencing Data Output			
total reads (millions)	66.632	55.385	57.100
% on target 10x	**87.5**	**80.7**	**73.88**
% on target 20x	74.15	53.49	45.6
% zero bases	1	1.16	0.71
Raw GB	5.612	5.231	4.998
PF unique GB	4.959	3.647	3.116
Mean Insert Size	**97**	**145**	**100**
% Dups	**7.68**	**4.97**	**23.68**
Library Size (millions)	203	322	501
Mean Target Coverage	48	27	23
Read Length	100	100	100
plexed/sequencing	5/lane	5/lane	9/FC

Degraded/Small Fragment Sample Experiments – First Exon Analysis

First exons are typically higher in GC content. Per target intervals were filtered for first coding exons and binned according to mean coverage (8x vs ≥8x; Table 5). Counts of first exon intervals according to bin were compared between library kits.

Table 5. First Exon Coverage

	Kapa Hyper Prep				Swift 1S			
Sample	Count	%<8x	Count	%≥8x	Count	%<8x	Count	%≥8x
NA12891	145	0.6	23745	**99.4**	437	1.8	23453	**98.2**
NA12892	91	0.4	23799	**99.6**	413	1.7	23477	**98.3**
NA12878					598	2.5	23292	**97.5**

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 generated. 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:

1-Kapa Hyper Prep Kit Illumina platforms Technical Data Sheet – KR0961-v1.14

2-Swift Accel-NGS 2S DNA Library Kit for the Illumina platform instruction manual – cat. no. DL-ILM2S-12/48; ver. 04291444

3-SureSelectXT Target Enrichment System for Illumina Paired-End Sequencing Library; SureSelectXT Target Enrichment for Illumina Multiplexed Sequencing Protocol v1.1.1, January 2011 p/n G7530-90000

4-Fisher *et al*; A scalable, fully automated process for construction of sequence-ready human exome targeted capture libraries. *Genome Biology* 2011, 12:R1

5-Swift Accel-NGS 1S DNA Library Kit for the Illumina platform Instruction manual – cat. no. DL-ILM1S-12/48; ver. 12011444

To obtain an electronic copy of this poster, please scan the barcode or visit http://www.cidr.jhmi.edu/nih/publications.html

www.cidr.jhmi.edu