RhinoSeq DNA Library Prep

A 12 minute, single tube, nanogram input library prep for multiple NGS applications

Joseph Dunham, Benjamin Akins, Christopher Angermayer, Christopher Mason

INTRODUCTION

Next Generation Sequencing (NGS) library construction is a workflow bottleneck. The processes of library construction for DNA and RNA sequencing are complex, error prone, costly, and in general are time consuming. Current commercial products offer multistep, multi-hour sample processing that can be cost and labor prohibitive. These issues have created a demand for a simple, rapid, and cost-effective library construction products that offer users application flexibility while minimizing construction complexity for low to ultra-high throughput sample processing. SeqOnce has developed a novel library technology that is rapid and minimizes sample processing complexity. The five tube kit is comprised of formatted master mixes for a simple and stable user workflow. The 12 minute library construction uses a single master-mix that when combined with fragmentation and PCR steps produces libraries in less than 50 minutes. A single size selection step occurs after PCR. The PCR free and pre-fragmented DNA protocols are less than 30 minutes.

The kit contains three master mixes and includes adapters, which maximally promotes product flexibility across multiple applications. With master-mixes that are stable for multiple days at ambient temperatures, automation on liquid handling platforms is effortless due to the simple workflow. The technology has been validated on a variety of human sample types and applications with input ranges of 10ng- 100ng. The sequencing data is either equivalent or superior to other competing library products, with high mapping rates and excellent performance across variable GC content. The technology is currently optimized for the Illumina platform and alternative NGS platforms are under evaluation for future development.

WORKFLOW & COMPONENTS

Single Tube Additive Reagent Library Prep

- Libraries in under 50 minutes
- Two step protocol, three steps with PCR
- Reagents stable for >24 hours at ambient temperature
- Minimal setup due to master mix format
  - Five reagent tube kit:

  - Three master mix tubes for:
    1. Fragmentation
    2. Library Construction
    3. PCR
  - Two remaining tubes:
    1. Stock Adapter
    2. Adapter Dilution Buffer

RHINOSEQ REAGENTS

Fragmentation Master Mix Characteristics:

- Single volume for 5 – 100ng inputs
- Adjustable for controlled fragment distribution
- No buffer/enzyme pipetting
- Single tube buffer/enzyme mix
- Identical conditions for microbial human gDNA fragmentation

FRAGMENTATION MASTER MIX REAGENT
SINGLE TUBE – SINGLE VOLUME

Fragmentation Distribution

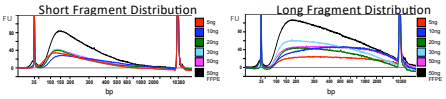


Figure 1: BioAnalyzer profile of RhinoSeq short or long fragmentation of Human DNA.
Various input amounts of Human gDNA were fragmented using the RhinoSeq kit to generate short (left) or long (right) fragment distributions. After fragmentation, samples were 2X SPRI® bead cleanup and analyzed using an Agilent High Sensitivity DNA Assay™.

Library Fragment Length Distribution

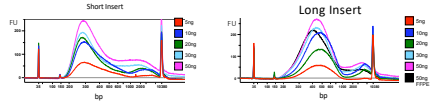


Figure 2: BioAnalyzer profile of RhinoSeq short or long insert Human gDNA Libraries.
Various input amounts of Human gDNA were fragmented to generate short (left) or long (right) fragment distributions and libraries constructed using the RhinoSeq kit. Short insert libraries were size selected for greater than 180bp and long insert libraries were dual size selected with SPRI® beads. Libraries were analyzed using an Agilent High Sensitivity DNA Assay™.

LIBRARY PREP MASTER MIX REAGENT
SINGLE TUBE – SINGLE VOLUME – 12 MIN PREP

Minimal Start Site/Coverage Bias

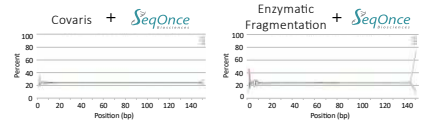


Figure 3: Read start-site bias for RhinoSeq *E. coli* libraries.
Nucleotide content over a 150bp window of sequencing read illustrates similar start-site complexity between Covaris (left) and random enzymatic fragmentation (right) using the RhinoSeq kit.

MICROBIAL SEQUENCING

WHOLE GENOME

Third Party- Independent Tester

Genome Coverage & GC Bias

Species	GC %	Library Input	Coverage Depth	Genome Coverage (unique reads)
Staphlococcus	32.7	30	10 - 20	+93%
Escherichia coli	51	10, 20, 30, 40	19 - 24	+96%
Rhodobacter	70	30	9.5 - 32	+96%

Table 1: Bacterial sequencing and coverage statistics.
Genome structure and sequencing statistics of *Staphlococcus*, *E. coli*, and *Rhodobacter*.

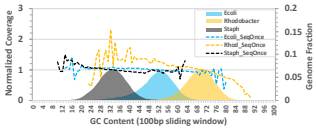


Figure 4: GC Bias Comparison for Microbial Samples.
Rhodobacter (grey), *E. coli* (blue), and *Staphlococcus* (yellow) GC bias was assessed by using Picard CollectGCBiasMetrics. A normalized coverage value of 1 is representative of an absence of sequencing bias. The RhinoSeq kit results in minimal GC bias in AT or GC rich genomes.

MICROBIAL COMMUNITY

Community Characteristics

Species	GC %	Gram Strain	gDNA Abun. (%)
Pseudomonas aeruginosa	66.2	-	12
Escherichia coli	56.8	-	12
Salmonella enterica	52.2	-	12
Lactobacillus fermentum	52.8	+	12
Enterococcus faecalis	37.5	+	12
Staphylococcus aureus	32.7	+	12
Listeria monocytogenes	38.0	+	12
Bacillus subtilis	43.8	+	12
Saccharomyces cerevisiae	38.4	Yeast	2
Cryptococcus neoformans	48.2	Yeast	2

Table 2: Community Standard Characteristics
The Zymo Research Microbial Community DNA Standard was used to assay library construction efficiency using the SeqOnce RhinoSeq kit. The standard contains a defined and characterized composition of 5 Gram-Positive and 3 Gram-Negative bacteria plus 2 yeast species. The GC content of the standard range is 15%-85%.

Community Sequencing

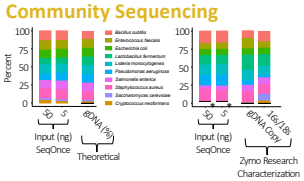


Figure 5: Taxonomic Identification
The Zymo Research Microbial Community DNA Standard was used to make short insert libraries with 50ng and 5ng input using the RhinoSeq kit. Profiling was with 2x75bp shotgun sequencing (15.2 million reads, Zymo: 178 million reads). Left panel, data analysis was conducted by reference read mapping and counting. Right panel, microbial composition was determined by using mOTU (http://www.bork.embl.de/software/mOTU). Relative-abundance deviation in average: <15% (indicated by vendor).
**S. cerevisiae* and *C. neoformans* were not included in the analysis.

HUMAN WGS

DEEP COVERAGE

Joseph Boland – Independent Tester
NCI/NIH, Leidos Biomedical Research Inc.

Quality

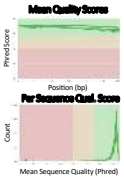


Figure 6: Sequence Quality
Base phred quality (top) and mean read quality scores (bottom) show high quality data from 50ng and 100ng NA12878 and NA12892 DNA libraries

Coverage

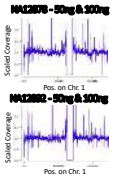


Figure 7: Human Chr. 1 Coverage. Coverage across Chr. 1 for NA12878 and NA12892 50ng (blue) and 100ng (pink) input libraries.

ROC Plots

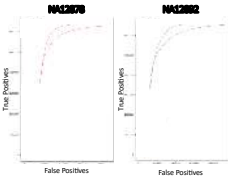


Figure 8: ROC Plots. True positive calls (y-axis) versus false positive calls (x-axis) for 50ng and 100ng NA12878 and NA12892 DNA libraries are plotted. Fifty nanogram libraries are in blue and 100ng libraries are in orange.

SHALLOW COVERAGE

Mapping

Avg. Insert Size (bp)	Input (ng)	Depth (bp)	Genome Coverage	Mapped Reads
102	50	1-10 (avg: 3.6)	+80%	91%

Table 3: NA12878 Library Mapping Statistics
The RhinoSeq kit was used to generate Human NA12878 50ng input short insert gDNA libraries and sequenced (2x150bp reads) to 3.6X average coverage.

Mean Coverage / Chr.

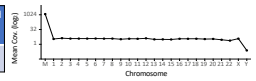


Figure 9: Mean Coverage Per Chromosome. The mean coverage per chromosome was consistent, ranging between 3-4x coverage.

GC Bias

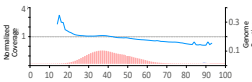


Figure 10: NA12878 RhinoSeq GC Bias
The RhinoSeq kit resulted in minimal GC bias despite sequencing to shallow depth. GC bias was assessed by using Picard CollectGCBiasMetrics.

HUMAN FFPE

Mild FFPE

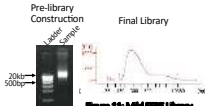


Figure 11: Mild FFPE Library
Pre- and post- library prep was assayed. A 50ng library was constructed with the standard RhinoSeq kit.

Severe FFPE

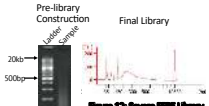


Figure 12: Severe FFPE Library
Pre- and post- library construction was assayed. The pre-fragmented DNA version of the RhinoSeq kit was used with 50ng DNA input and produced libraries in 30m.

Mapping

	GENOME TERRITORY	MEAN COV	SD COV	GENOME COV
WHOLE	3066216270	16.594398	9.511458	90%
NON-ZERO	2764750736	18.403834	8.187294	

	GENOME TER.	MEAN COV	SD COV	GENOME COV
Severe - 1	3066216270	0.703822	1.508231	30 - 40%
	1176392479	1.834482	1.963387	
Severe - 2	3066216270	0.5536	1.315136	30 - 40%
	1024199224	1.65735	1.829937	

HLA TYPING

Independent Tester – HLA Typing Company

HLA Library BioA

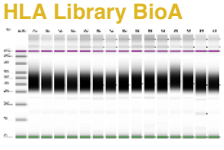
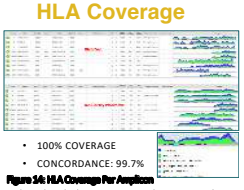


Figure 13: Bioanalyzer HLA Amplicon Libraries
Plot represents a subset of Bioanalyzer traces of 48 libraries generated with the RhinoSeq kit using 22-53ng HLA multiplexed amplicons (4-7kb).

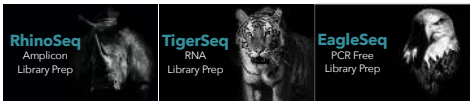
HLA Coverage



- 100% COVERAGE
- CONCORDANCE: 99.7%

Figure 14: HLA Coverage Per Amplicon
Each amplicon had 100% coverage and 99.7% concordance.

FUTURE PRODUCTS



CONCLUSION

The RhinoSeq kit allows for DNA library prep that is both straightforward and rapid. Total library processing time (less than 50 minutes) is faster than other commercial kits (three to four times faster). The simplicity of the five tube two-step protocol, three-step with PCR, enables experienced and inexperienced users to easily prepare libraries for sequencing. The flexibility of this system also allows users to readily scale the number of samples processed, easy to automate, and can be applied to a variety of sequencing applications.

- Bioinformatics was performed by independent testers where indicated or outsourced to NextBioInformatics.
- Microbial abundance, activity and population genomic profiling with mOTUs2. Milanese, A et al. Nature Communications 10, Article number: 1014 (2019).
- The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. McKenna A et al. 2010 *GENOME RESEARCH 20:1297-303*
- A framework for variation discovery and genotyping using next-generation DNA sequencing data DePristo M et al. 2011 *NATURE GENETICS 43:491-498*
- From FastQ Data to High-Confidence Variant Calls: The Genome Analysis Toolkit Best Practices Pipeline Van der Auwera GA et al. 2013 *CURRENT PROTOCOLS IN BIOINFORMATICS 43:11.10.1-11.10.33*
- MultiQC: Summarize analysis results for multiple tools and samples in a single report. *Philip Ewels, Måns Magnusson, Sverker Lundin and Max Käller. Bioinformatics (2016).*