A Single-tube, Low Input Protocol for Long Read Sequencing

Keerthana Krishnan, Luo Sun, Deyra N. Rodriguez, Brad Langhorst, Netsanet Gebremedhin, Yanxia Bei, Janine Borgaro, Shengxi Guan, Nicole Nichols, Eileen Dimalanta, and Theodore Davis

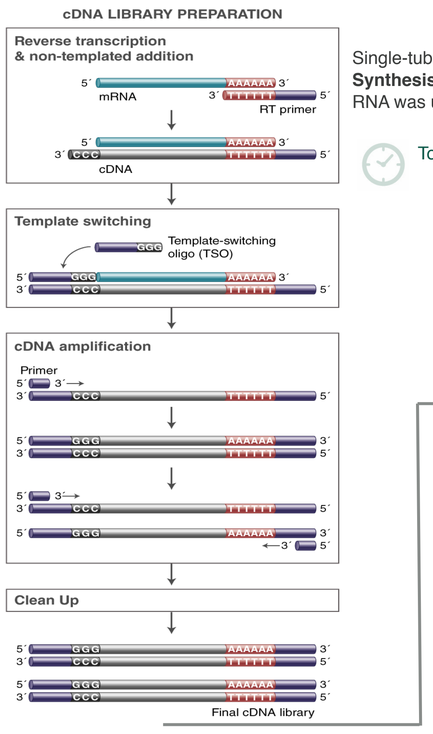
New England Biolabs, Ipswich, MA 01938, USA

INTRODUCTION

The human genome has ~20,000 genes, a majority of which produce multiple isoforms through pre-mRNA alternative splicing. These isoforms, or variants, are translated into proteins with different spatiotemporal dynamics, unique biological functions and, in some cases, disease-causing effects. Therefore, it is crucial to have a full complement of these spliced RNA isoforms to understand their heterogeneous functions. However, due to high sequence similarity among isoforms, confidently resolving their identities through de novo assembly using short-read sequencing data poses a challenge for bioinformatics.

Single molecule sequencing platforms such the PacBio® Sequel and Oxford Nanopore MinION™ have the potential to overcome some of the limitations of short-read sequencing, as they are capable of generating long-reads (>10kb). These technologies enable the sequencing of full-length cDNA, resulting in complete and unambiguous identification of each transcript isoform. Transcripts are sequenced with consistent coverage from 5' to 3' including full-length polyA tails. This is a significant improvement over short-read sequencing technologies, which require sequence assembly for transcriptome analysis. Prior to sequencing, preparation of a high quality cDNA library from full length mRNA is critical. Here, we demonstrate a novel workflow to generate cDNA from nanogram quantities of total RNA. The workflow beginning with total RNA to amplification of cDNA is carried out in a single tube with very little hands on time, producing consistent, full length cDNA libraries. Our workflow is a streamlined, efficient method to produce high quality cDNA libraries for both PacBio and Oxford Nanopore sequencing platforms, enabling robust detection of full-length isoforms and splicing variants from minimal RNA input.

METHODS



Single-tube cDNA Library Preparation using the **NEBNext® Single Cell/Low Input cDNA Synthesis & Amplification Module**. UHR (Universal Reference RNA from Agilent) total RNA was used as input RNA.

Total: 4-5hrs

PacBio SMRT DNA Library Prep

DNA End Repair → Adaptor ligation → Bead Cleanup → Exo Digestion → 2X Bead Cleanup

Amplified cDNA was end repaired using NEBNext End Repair Module, 25˚C for 5 min. The cDNA was ligated with a blunt end adaptor (PacBio) using the NEBNext Ultra II Ligation Master Mix for 10 min at room temperature followed by Ampure PB purification (1x reaction volume). Clean and eluted cDNA was digested with Exo III and VII (NEB) for 1 hr, followed by 2x Ampure PB purification at 1x and 0.6x reaction volume.

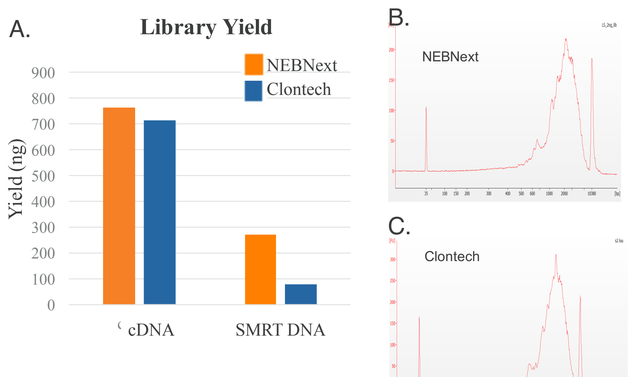
Oxford Nanopore (ONT) DNA Library Prep

DNA End Repair / dA-tailing → Adaptor ligation → Bead Cleanup

Amplified cDNA was end repaired and A-tailed using the NEBNext Ultra II End Repair/dA-Tailing Module, 25˚C for 5 min, 65˚C for 5 min. The cDNA was ligated with the ONT adapter using the NEBNext Ultra II Ligation Master Mix for 10 min at room temperature followed by NEBNext Sample Purification Beads (0.6x reaction volume). cDNA was cleaned using ABB Wash Buffer 2x and eluted in elution buffer (ONT). The eluted sample was then mixed with RBF running buffer and the LBB loading beads.

RESULTS: PacBio

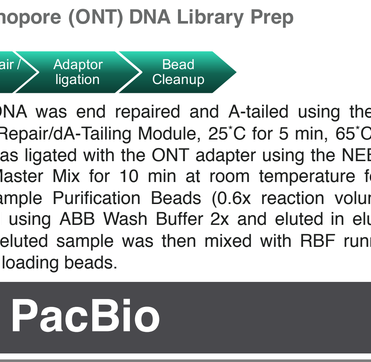
Library QC



Amplified cDNA libraries were made using 8ng of total UHR RNA, followed by 15 PCR cycles with either the NEBNext® Single Cell/Low Input cDNA Synthesis & Amplification Module or Clontech SMARTer PCR cDNA synthesis kit and PrimeSTAR GXL DNA Polymerase. The amplified cDNA was then used to make PacBio SMRT libraries using NEB reagents.
A) The cDNA and final SMRT DNA library yields were quantified using Qubit (ThermoFisher). Both the cDNA and final SMRT DNA library yields were higher for the NEBNext libraries compared to Clontech (B&C)The size distributions of final SMRT libraries (B. NEB, and C. Clontech) were analyzed on an Agilent 2100 Bioanalyzer. There was not a significant difference in the size distribution between NEB and Clontech libraries.

Sequencing Performance

CCS (Circular Consensus Sequence) reads from samples generated using NEB and Clontech reagents were analyzed using Iso-Seq. These results indicate that CCS reads have similar size and QV distribution for both NEB (A, B and C) and Clontech samples (D, E,and F). In addition, the read length of Full-length Non-Chimeric Reads are also similar for both NEB and Clontech samples.



The concordance of mapped subreads and the distribution of concordance vs read length are similar between both NEB and Clontech samples (Requencing, PacBio).

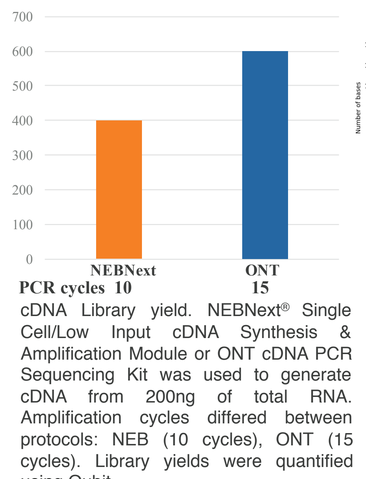
Iso-Seq Analysis

Transcript analysis of data using Iso-Seq through SMART link. Classification data indicates that the number of transcripts carrying 5' end, 3' end, poly-A tails, full-length reads and full length non-chimeric reads are all higher per GB data in the NEBNext versus Clontech samples.

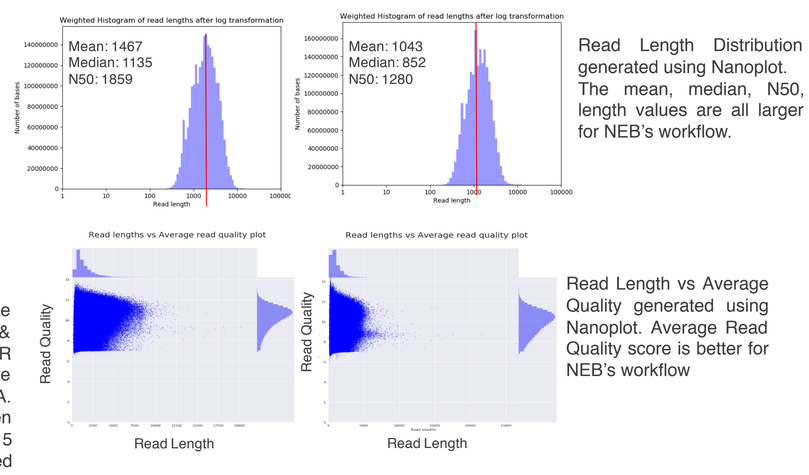
Transcript identification and coverage on gene ATF3 (Activating transcription factor 3, a known RAG with numerous promotors.). 3 isoforms were detected from the sample using NEBNext reagents, and 1 isoform was detected with Clontech reagents. In addition, two variants, G-T (chr 1, 212793348) and T-G (chr 1, 212782202), were detected only in the NEBNext workflow.

Oxford Nanopore Results

Library QC



cDNA Library yield. NEBNext® Single Cell/Low Input cDNA Synthesis & Amplification Module or ONT cDNA PCR Sequencing Kit was used to generate cDNA from 200ng of total RNA. Amplification cycles differed between protocols: NEB (10 cycles), ONT (15 cycles). Library yields were quantified using Qubit.



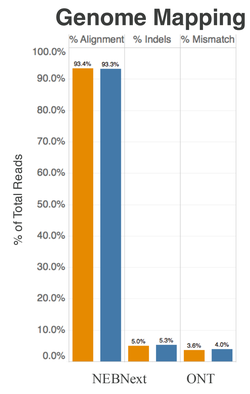
Read Length Distribution generated using Nanoplot. The mean, median, N50, length values are all larger for NEB's workflow.

Read Length vs Average Quality generated using Nanoplot. Average Read Quality score is better for NEB's workflow

Reference Mapping

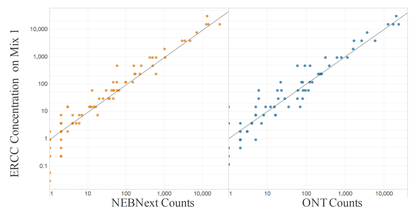
For mapping analysis reads were adaptor trimmed (Porechop 0.2.3 [1]), aligned to GRCh38 or Gencodev28 +ERCC transcripts using Minimap2 (v2.5 [2]), and downsampled to 1.5M reads.

Genome Mapping Stats



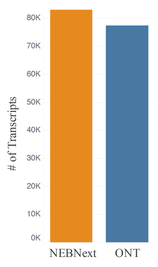
Alignment, indel and mismatch rates were assessed using Picard's alignment summary metrics tool [3]. Similar mapping efficiency is observed between NEBNext and ONT samples.

ERCC Correlation



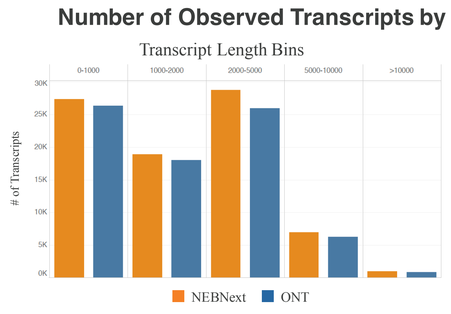
Counts of ERCC transcripts were correlated with their expected concentrations. NEBNext has a tighter distribution against the expected ERCC counts.

Number of Observed Transcripts



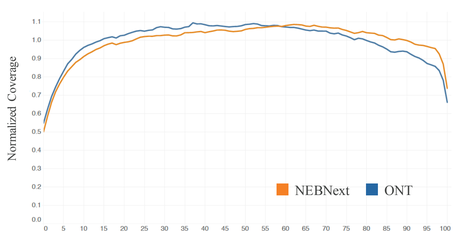
Transcripts with 2 or more reads were considered "observed". More transcripts are observed for NEBNext.

Number of Observed Transcripts by Length



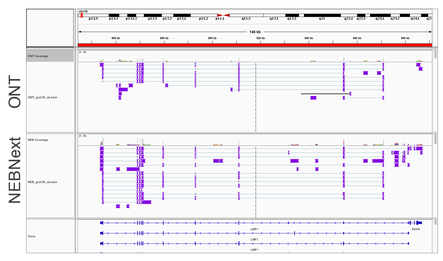
Transcripts with 2 or more reads were considered "observed". To assess differences related to transcript length, these were divided into length bins. More or similar number of transcripts are observed for NEBNext.

5-3' Coverage



Aggregate 5'-3' coverage was assessed using Picard's RNA-seq metrics tool [3]. 5'-3' coverage is flatter across the transcript length for NEBNext.

Example of Isoform Coverage on LMF1



Identification and coverage on gene LMF1 (Lipase Maturation Factor 1). LMF1 gene length is 113620 base pairs. Higher and better isoform coverage was observed in the NEBNext sample.

CONCLUSIONS

Here we describe the **NEBNext® Single Cell/Low Input cDNA Synthesis & Amplification Module** for preparing cDNA libraries in a single tube reaction. This protocol requires minimal hands on time with less pipetting steps for cDNA synthesis and amplification. This module is compatible with both PacBio and ONT DNA Library Prep Workflows for long read sequencing.

Pac Bio

Comparison against Clontech SMARTer PCR cDNA synthesis kit shows that the size distribution of final SMRT DNA libraries is similar in both NEBNext and Clontech samples. Final library yield is higher for NEBNext compared to Clontech. The sequencing performance metrics (CCS read length, QC, and subread mapping concordance) are comparable between NEBNext and Clontech. A higher number of transcripts is detected for NEBNext vs Clontech.

ONT

Comparison against ONT's cDNA PCR Sequencing Kit shows that the mean and N50 read lengths are longer for NEBNext libraries. A lower number of PCR cycles (5 cycles less) was required for the NEBNext sample. The number of transcripts detected is higher for NEBNext, and the 5'-3' coverage across transcript length is more uniform with NEBNext. We observe higher and better isoform coverage on NEBNext samples vs ONT.

The NEBNext single cell/low input cDNA synthesis Module enables a robust detection of full-length isoforms and variants from low RNA input.

REFERENCES

[1] Ryan Wick https://github.com/rrwick/Porechop
[2] Heng Li (2017). Minimap2: fast pairwise alignment for long nucleotide sequences
[3] http://broadinstitute.github.io/picard/