Poster #687

Full-length cDNA Sequencing of Prokaryotic Transcriptome and Metatranscriptome Samples

Matthew Boitano[1], Bo Yan[2], Ting Hon[1], Elizabeth Tseng[1], Laurence Ettwiller[2] and Tyson A. Clark[1]

[1]PacBio, 1305 O'Brien Drive, Menlo Park, CA 94025

[2]New England Biolabs, 240 County Road, Ipswich, MA 01938

Introduction

Next-generation sequencing has become a useful tool for studying transcriptomes. However, these methods typically rely on sequencing short fragments of cDNA, then attempting to assemble the pieces into full-length transcripts. Here, we describe a method that uses PacBio long reads to sequence full-length cDNAs from individual transcriptomes and metatranscriptome samples.

We have adapted the PacBio Iso-Seq protocol for use with prokaryotic samples by incorporating RNA polyadenylation and rRNA-depletion steps. In conjunction with SMRT Sequencing, which has average read lengths of 10-18 kb, we are able to sequence entire transcripts, including polycistronic RNAs, in a single read.

Here, we show full-length bacterial transcriptomes with the ability to visualize transcription of operons. We also highlight the ability to detect full-length transcription of operons with alternative start and stop sites. In the area of metatranscriptomics, long reads reveal unambiguous gene sequences without the need for post-sequencing transcript assembly.

Sample Preparation Methods

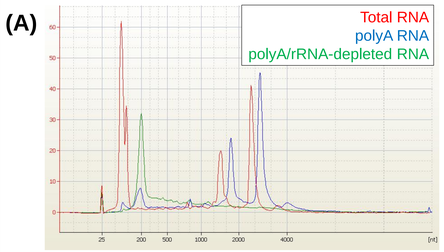
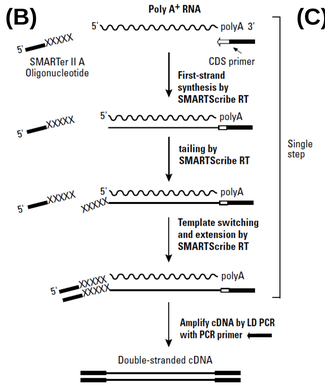
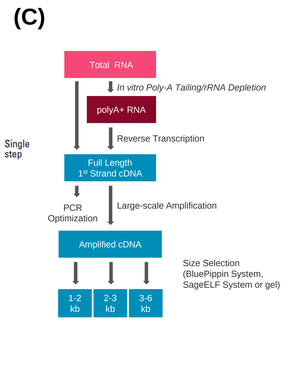
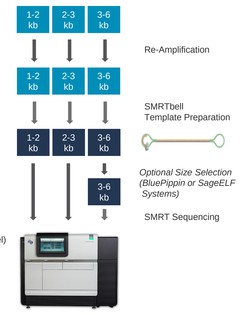


Figure 1. (A) Bioanalyzer traces of *E. coli* total RNA, polyadentylated RNA and polyadentylated/rRNA-depleted RNA. PolyA-tail reaction has been optimized in order to add ~200 nucleotides. Polyadentylated/rRNA-depleted RNA showed good reduction in rRNA peaks and was the input for the cDNA synthesis reaction.

(B) Clontech SMARTer PCR cDNA Synthesis Kit was used to generate double-stranded cDNA. (C) Double-stranded cDNA was size-fractionated using the Sage BluePippin System to sizes of 1-2, 2-3, 3-6 and 5-10 kb (if material is available at each size). This size-fractionated material was converted into SMRTbell libraries. Each library was sequenced on the PacBio RS II with P6C4 chemistry and 4 hour movies. Alternatively, non-size selected material could have been used to generate SMRTbell libraries.







Effects of rRNA Depletion

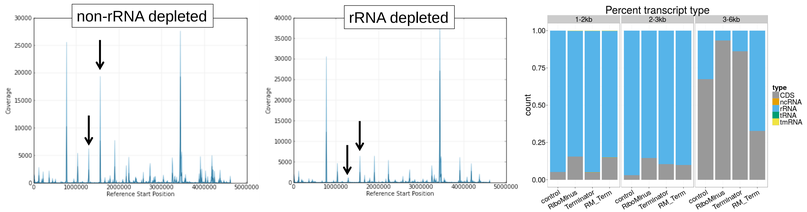


Figure 2. Sequence reads were mapped to the *E. coli* genome reference. Arrows show reduction in coverage of rRNAs after rRNA depletion. Shared peaks are most likely ribosomal associated genes. Out of all traditional rRNA depletion methods tested, RiboMinus had the highest rRNA depletion efficiency.

Detection of Poly-cistronic and Full-length Operon Transcripts

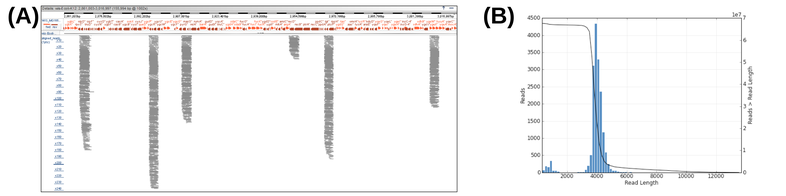


Figure 3. (A) Using long-read SMRT Sequencing, poly-cistronic and full-length operon reads are easily obtained without the need for assembly of short fragments. (B) Data shown are from 3-6 kb size bin, which have an average insert size of 3,917 bp.

Detection of Alternative Transcription Start/Stop Sites

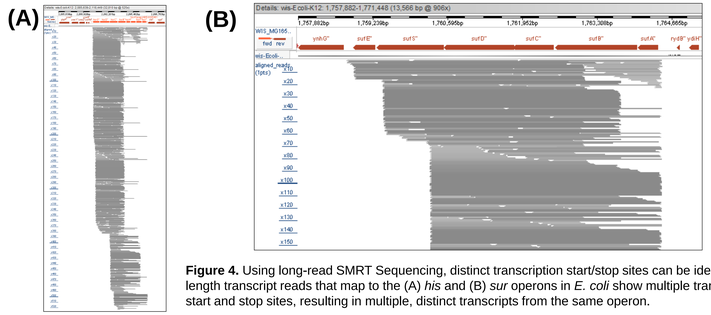


Figure 4. Using long-read SMRT Sequencing, distinct transcription start/stop sites can be identified. Full-length transcript reads that map to the (A) *his* and (B) *sur* operons in *E. coli* show multiple transcription start and stop sites, resulting in multiple, distinct transcripts from the same operon.

Enrich for Primary Transcripts with NEB Cappable-Seq

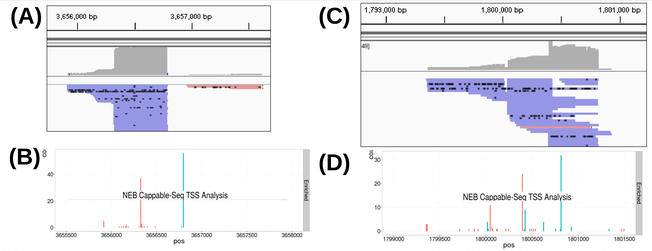


Figure 5. Primary transcripts were enriched utilizing a newly developed method from New England Biolabs called Cappable-Seq (1). Non-processed 5' ends of transcripts were capped with a selectable tag and then sequenced using the SMRT Sequencing Iso-Seq protocol. (A/B) Full-length transcripts, including their transcriptions start sites, were detected allowing for the detection of novel operons. (C/D) This method also provides phasing of transcription start sites and termination sites, even when dealing with overlapping transcripts with additional internal transcription start and termination sites.

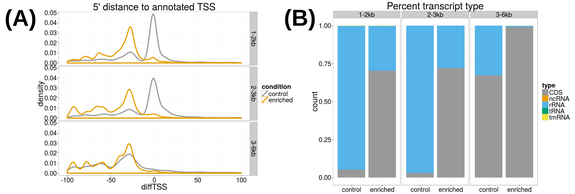


Figure 6. (A) Enrichment of primary transcripts allowed for the detection of transcription start sites by increasing the number of full-length transcript reads with an intact 5' end. (B) It also allowed for simultaneous rRNA depletion, with higher depletion rates compared to other methods.

Metatranscriptome Long-read Sequencing

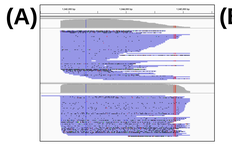


Figure 7. Cells for ZymoBIOMICS Microbial Community Standard were obtained. RNA was isolated using bead beading extraction. RNA was converted to full-length cDNA as described in the Sample Preparation section.

(A) Full-length, polycistronic transcripts were detected, as well as SNV's relative to a reference. These can be used for phasing homologous regions of different organisms.

(B) Percent reads mapped to each genome are dependent on the organism's expression level activity within the community.

Conclusion

- Bacterial SMRT Sequencing Iso-Seq protocol available on PacBio website (2)
- Newly developed method from New England Biolabs allows for enrichment of primary transcripts
- When combining SMRT Sequencing and Cappable-Seq users will be able to detect full-length, polycistronic transcripts from prokaryote transcriptome and metatranscriptome samples. Users will also be able to detect and phase transcription start and termination sites.

References

[1]Ettwiller L, *et al*. (2016) A novel enrichment strategy reveals unprecedented number of novel transcription start sites at single base resolution in a model prokaryote and the gut microbiome. *BMC Genomics*. 17, 199.

[2]http://www.pacb.com/wp-content/uploads/Unsupported-Protocol-Bacterial-Iso_Seq-Clontech-SMARTer-PCR-cDNA-Synthesis-Kit-BluePippin-Size-Selection.pdf