#PB3396

Scalable, cost-effective isoform sequencing with Kinnex™ full-length RNA kit using long-read sequencing

Elizabeth Tseng[1], Heather Ferrao[1], Jocelyne Bruand[1], Armin Toepfer[1], Gloria Sheynkman[2], Madison Mehlferber[2], Vasilli Pavelko[2]
1 PacBio, 1305 O'Brien Drive, Menlo Park, CA 94025
2 Dept of Molecular Physiology and Biological Physics, University of Virginia, Charlottesville VA 22093

Introduction

Isoforms – not genes – are the drivers of biology and disease [1]. Traditional RNA-Seq fragments cDNA for short-read sequencing (100–200 bp), which are often followed by computational assembly or mapping to infer the original transcript isoforms. However, given the complexity of alternative splicing, many isoforms share highly similar structures, and the inferred transcripts are often inaccurate. PacBio HiFi reads sequence full-length RNA isoforms without the need for cDNA fragmentation and transcript assembly (Figure 1), allowing for unambiguous full-length isoform detection.

The *Kinnex full-length RNA kit* takes total RNA as input and outputs a sequencing-ready library that results in an 8-fold throughput increase compared to typical Iso-Seq libraries. Combined with the Iso-Seq analysis in SMRT Link software, PacBio offers cost-effective isoform sequencing that does not require orthogonal sequencing methods.

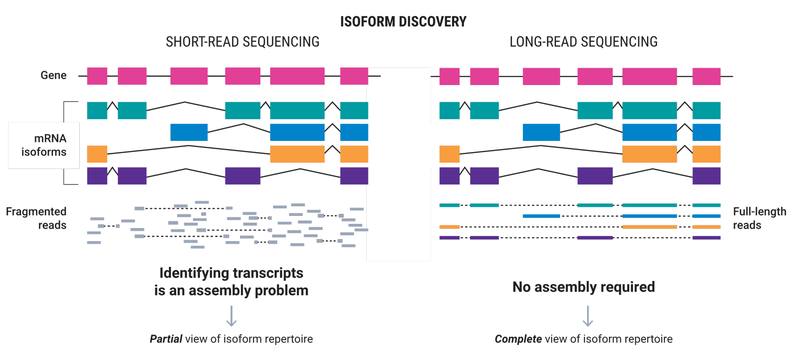

Figure 1. Long-read RNA sequencing eliminates the need for transcript assembly, which cannot accurately resolve the isoform structure. Long-read RNA-Seq using PacBio (the Iso-Seq method) sequences the entire full-length cDNA to provide an unambiguous view of the transcriptome.

Kinnex full-length RNA workflow

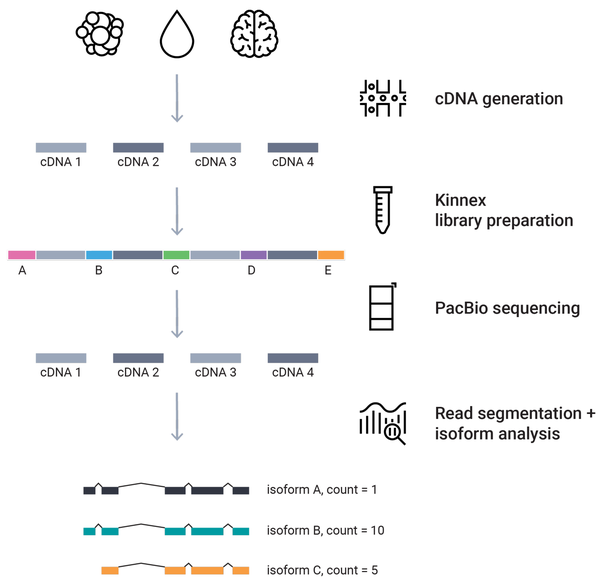

Figure 2. Example workflow for cells, blood or brain tissue from cDNA generation through isoform analysis

- Total RNA as input (300ng, RIN >= 7)
- Generates barcoded cDNA (up to 12-plex) using Iso-Express 2.0 kit
- Create Kinnex libraries by concatenating 8 cDNA into an array
- Sequence on PacBio Sequel II, IIe, or Revio systems
- SMRT Link outputs isoform read count information
- Achieve 15 million reads on Sequel II and IIe systems and 40 million reads on Revio system

Kinnex full-length RNA kit increases throughput for full-length RNA-Seq without bias

Using a UHRR (Universal Human Reference RNA) sample, we observed that Kinnex libraries can capture full-length transcripts up to 10 kb without skewing transcript lengths and abundances (Figure 3, 4, 5).

Yield comparison for regular vs Kinnex libraries

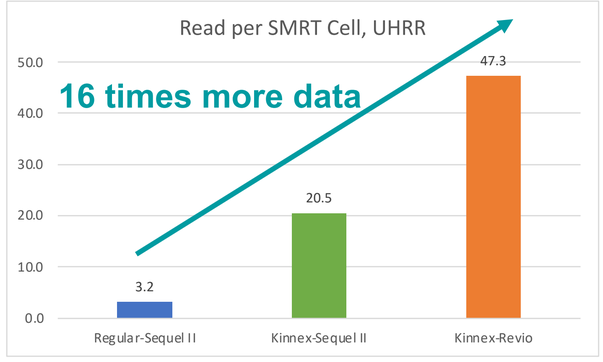

Figure 3. UHRR sample run in three sequencing run modes: (A) regular Iso-Seq library on Sequel II system, (B) Kinnex library on Sequel II system, and (C) Kinnex library on Revio system yielding 16-times more data in mode C vs mode A.

Isoform abundance concordance for regular vs Kinnex libraries

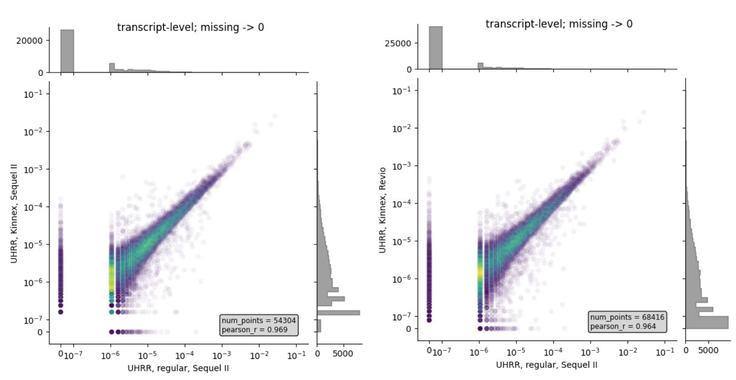

Figure 4. UHRR isoform abundances are highly concordant using (left) regular Iso-Seq vs Kinnex library on the Sequel II system and (right) regular Iso-Seq vs Kinnex library on the Revio system. Isoform abundances were based on normalized read counts.

Transcript lengths captured with Kinnex libraries

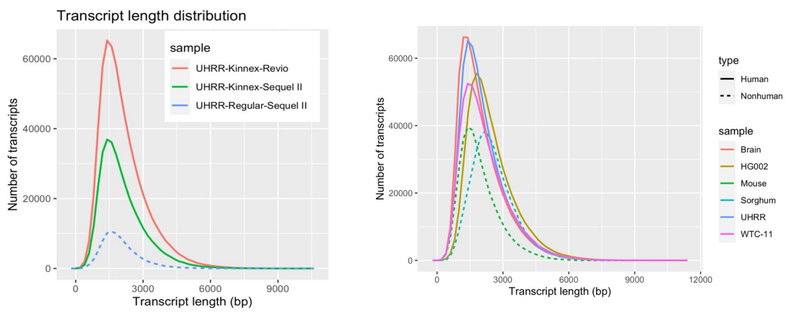

Figure 5. (left) UHRR transcript length capture and data consistent across three run modes outlined in Figure 3 and (right) transcript lengths captured with Kinnex for various species and tissues.

Kinnex dataset release

We release a dataset consisting of the following:

Type	# of samples	Total # of reads	# genes and isoforms
UHRR	Sequel II and Revio, 2 total	66,096,124	~20k genes, ~200k isoforms
HG002	One sample on Revio	38,740,671	~20k genes, ~300k isoforms

Additional datasets from collaborator samples including WTC-11, mouse, sorghum etc were used as part of this analysis (to be published)

Dataset publicly available at www.pacb.com/connect/datasets/

Transcript capture sensitivity analysis

Kinnex RNA data shows high technical reproducibility

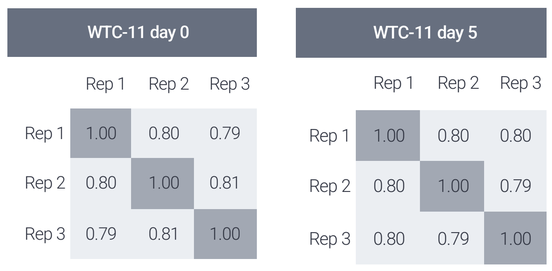

Figure 6. Good technical reproducibility in Kinnex libraries. Three technical replicates each from collaborator WTC-11 sample show high isoform abundance correlation for both day 0 and day 5, similar to observed correlation values for matching Illumina technical replicates (0.78–0.82, data not shown).

Saturation of known isoforms at 10 million reads per sample

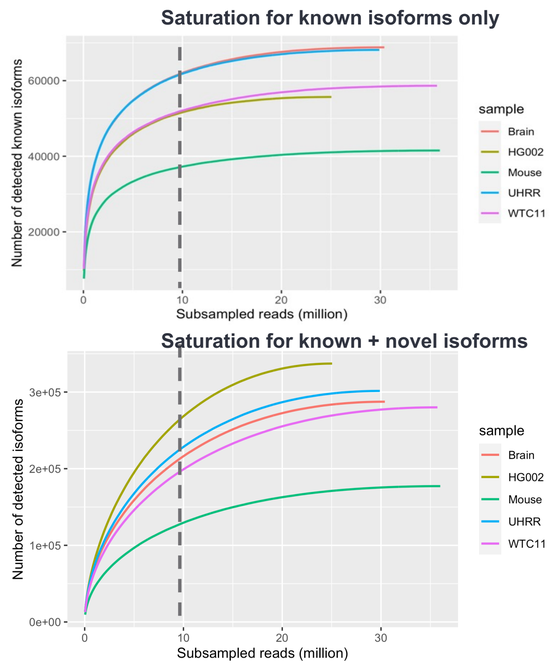

Figure 7. Saturation analysis of (top) known isoforms and (bottom) all isoforms for different Kinnex samples. At ~10 million reads, 80% of known isoforms have been detected, while capture of increasing rare novel isoforms continue with deeper coverage.

Conclusion

- Long-read RNA-Seq using Kinnex libraries with PacBio systems sequence the entire full-length cDNA to provide an unambiguous view of the transcriptome.
- Use of Kinnex libraries does not alter transcript sizes and abundances compared to regular Iso-Seq libraries
- There is little technical variability resulting from library preparation or the PacBio sequencer used.
- 10 million reads per sample is sufficient to detect most known isoforms, while deeper sequencing detects more rare, novel isoforms

References

1. Castaldi P, et al. (2022) Bridging the splicing gap in human genetics with long-read RNA sequencing: finding the protein isoform drivers of disease, *Human Mol Genet*

Acknowledgements

The authors would like to thank everyone who helped generate data for the poster.

For more information, go to pacb.com/RNA-Seq