Streamlined hybridization capture workflow for targeted long-read sequencing

Elizabeth Tseng[1], Davy Lee[1], Camille Connor[1], Katelyn Larkin[2], Justin Jacques[2], Ashley Dvorak[2]

[1]PacBio, 1305 O'Brien Drive, Menlo Park, CA 94025, USA
[2]IDT, 1710 Commercial Park, Coralville, IA 52241, USA

Introduction

The PacBio Iso-Seq method can sequence full-length transcripts up to 10 kb without the need for computational assembly, offering exceptional insights into the complexity of alternative splicing. The recent release of Kinnex kits, which can be applied to cDNA generated using the Iso-Seq express 2.0 kit, further increases cDNA yield on PacBio's long-read sequencing platforms.

However, in certain applications where the genes of interests are known, targeted approaches can be an even more cost-effective solution. Hybrid capture can reduce sequencing costs, detect rare transcripts, and simplify analysis.

Here, we combine IDT hybridization capture with PacBio full-length RNA sequencing for deep isoform characterization. We show that the new IDT xGen Hyb and Wash Kit v3 is compatible with the PacBio Kinnex full-length RNA kit and can achieve an on-target rate of 85%, increasing detection of low abundance isoforms by several thousand-fold. Further, the transcript abundances are preserved through the capture, allowing for accurate quantification.

IDT-Kinnex targeted full-length RNA library workflow

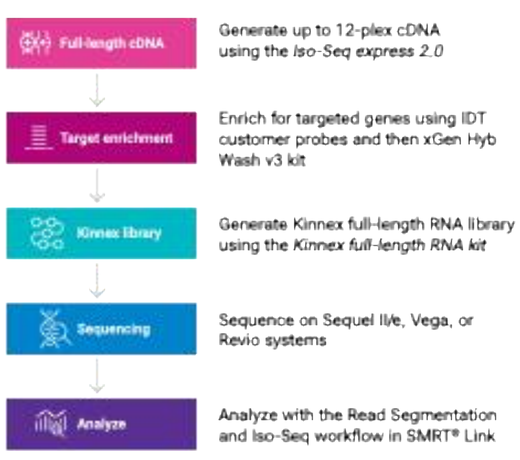

Figure 1. Library workflow using IDT capture with PacBio Kinnex library prep.

Full-length cDNA should be generated using the Iso-Seq express 2.0 kit which is compatible with IDT targeted enrichment and Kinnex full-length RNA kit.

The input to Iso-Seq express 2.0 kit is 300 ng of total RNA, ideally with RIN ≥ 7. Sequencing of the Kinnex libraries can be done on PacBio Sequel II/IIe, Revio, and Vega systems.

Targeted enrichment in human brain and UHRR samples

We applied the targeted IDT-Kinnex full-length RNA workflow to human brain and UHRR samples. Twenty-nine genes implicated in neurodegenerative, neuromuscular, and autoimmune disorders were selected for probe synthesis by IDT (Table 1).

ABCA7	*CD2AP*	*HTT*
APH1A	*CD33*	*INPP5D*
APOE	*CELF1*	*MAPT*
APP	*CLU*	*MEF2C-AS1*
BACE1	*CR1*	*MS4A6A*
BIN1	*EPHA1*	*NCSTN*
BSG	*FERMT2*	*NME8*
C9orf72	*FMR1*	*PICALM*
CASS4	*GRN*	*SPDEF*
	HLA-DRB1	
	HLA-DRB5	

Table 1. List of 29 target genes. Probes were synthesized by IDT.

Commercial human brain and UHRR samples were split into technical replicates, where half were subject to targeted enrichment. Sequencing was done on one Revio SMRT Cell and one Vega SMRT Cell each (Figure 2).

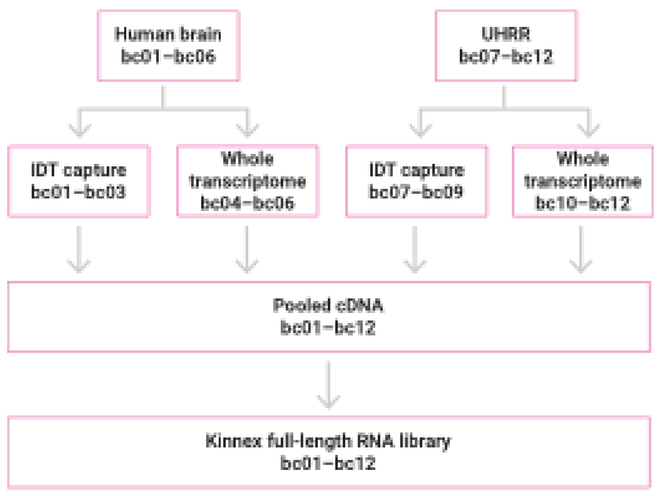

Figure 2. Experimental design for IDT-Kinnex targeted full-length RNA sequencing on human brain and UHRR samples.

We obtained 59.7 million and 36.2 million segmented reads (S-reads), which are full-length cDNA reads, with a consistent on-target rate of 85-86% for IDT targeted enrichment vs 1.2-1.3% for whole transcriptome (Figure 3).

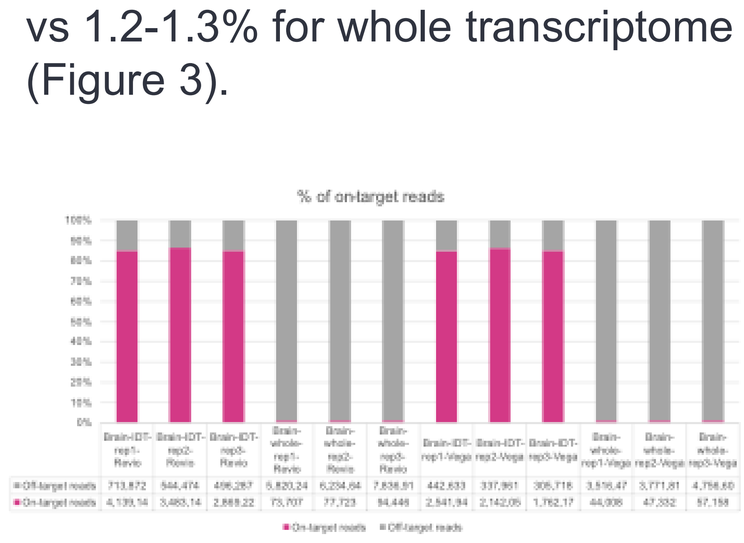

Figure 3. Percentage of on-target reads for IDT targeted enrichment vs whole transcriptome Kinnex full-length RNA data.

Targeted enrichment detects rare isoforms and retains abundances

Targeted enrichment enabled detection of more known and novel isoforms in targeted genes (Figure 4). Some genes (e.g., *CR1*, *NME8*) could only be detected with enrichment (Figure 4).

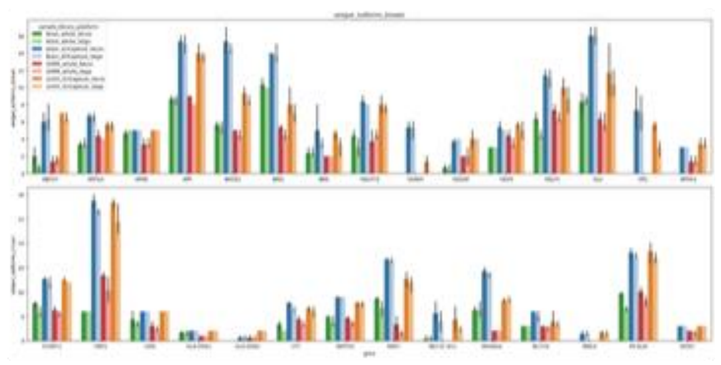

Figure 4. Number of unique isoforms detected for the 29 targeted genes. Missing bars means no transcripts from that gene was detected.

High replicability was observed for the same Kinnex libraries sequenced on Revio and Vega systems, with same-sample different-replicate correlations also high (Figure 5a). Further, transcript abundances correlated well between targeted and whole transcriptome data (Figure 5b).

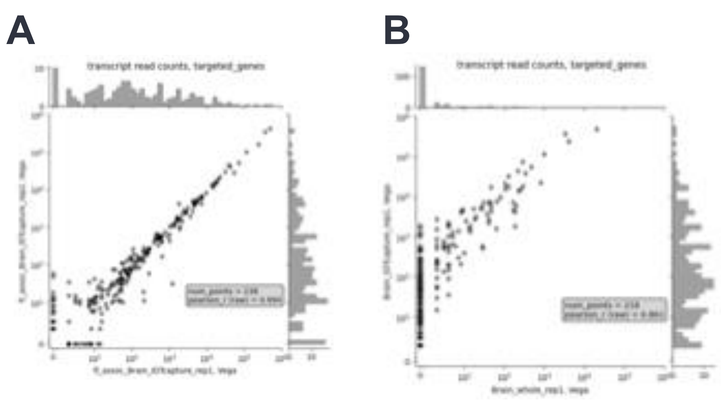

Figure 5. (a) High replicate reproducibility across technical replicates (b) Transcript abundances correlate between targeted enrichment vs whole transcriptome, while the capture data detects considerably more isoforms for the target genes.

Conclusion

Targeted enrichment is a cost-effective way to detect transcripts in genes of interest. We applied the IDT-Kinnex workflow to commercial human brain and UHRR samples and showed that:

- IDT targeted enrichment achieves ~85% on-target rate
- Targeted enrichment detects rare isoforms not seen in whole transcriptome data at equivalent depth
- Transcript abundances are retained through targeted enrichment and can be quantitative

References

To learn more about Kinnex kits, the IDT-Kinnex workflow, and the dataset, visit https://pacb.com/kinnex