PacBio

Whitepaper

Full-length and single-cell isoform sequencing for human disease research

Transcript isoforms are key drivers of human development and disease. Full-length isoform sequencing in bulk and single-cell transcriptomes enables characterization of complex alternative splicing, prediction of open reading frames (ORFs), and identification of cell type-specific, allele-specific isoform expressions. Short reads can only provide gene-level information and often present either an incomplete or mis-assembled representation of isoforms. The PacBio® Iso-Seq® method and Kinnex™ kits utilize highly accurate HiFi sequencing to capture full-length transcripts, with no assembly required. This enables a higher-resolution picture of transcriptomes at the isoform level, which is crucial for understanding functional cellular diversity and dynamic expression in human biology and disease.

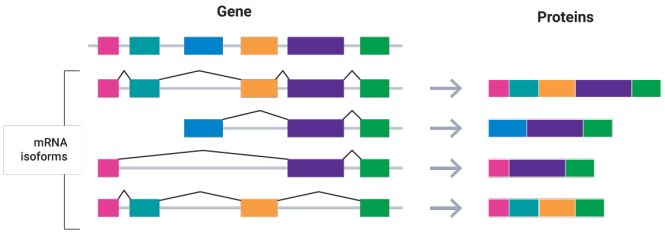


Figure 1. HiFi reads span the full length of individual transcripts without the need for assembly, which enables functional prediction of different isoforms.

In this whitepaper, we discuss the features of full-length and single-cell isoform sequencing with the Iso-Seq method and their applications in human genomics.

What can you do with the Iso-Seq method?

- Generate comprehensive whole transcriptomes at isoform-level resolution
- Detect isoforms as potential biomarkers for disease states or treatment response
- Discover key transcription factors and transcriptional signatures, alternative splice sites, and transcriptional fusion or readthroughs with high accuracy
- Predict ORFs for assessment of functional impact
- Identify allele-specific isoform expression
- Reveal cell type-specific isoform expression and new cell clusters in single-cell transcriptomes

102-326-576 REV02 NOV2023

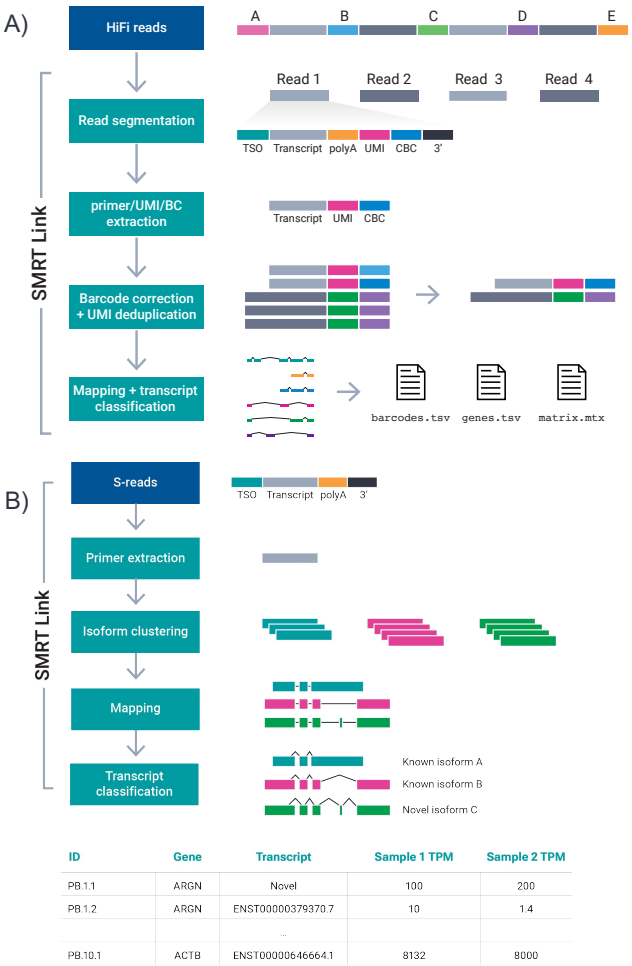
PacBio

Do more with Kinnex

With the Multiplexed Arrays Sequencing (MAS-Seq) (Al'Khafaji et al., 2023) method, full-length cDNAs are joined together into larger fragments, allowing for multiple transcripts to be sequenced on a single HiFi read. The result is higher throughput for cost-effective isoform sequencing. Compared to non-Kinnex methods, the *Kinnex single-cell RNA kit* increases throughput by 16-fold for single-cell RNA-Seq, while the *Kinnex full-length RNA kit* increases throughput by 8-fold for bulk RNA-Seq.

Analysis tools

The analysis of HiFi reads to generate accurate, full-length transcript isoforms is streamlined and fully supported with Iso-Seq and Single-cell Iso-Seq analyses in SMRT® Link that are compatible with popular tertiary analysis tools.



ID	Gene	Transcript	Sample 1 TPM	Sample 2 TPM
PB.1.1	ARGN	Novel	100	200
PB.1.2	ARGN	ENST00000379370.7	10	1.4
...				
PB.10.1	ACTB	ENST00000646664.1	8132	8000

Figure 2. Iso-Seq analysis workflow through SMRT Link for a) single-cell and b) bulk isoform sequencing.

Powering human disease research

Neurobiology

Full-length RNA sequencing can give a snapshot of average isoform expression in specific brain regions. The Iso-Seq method can be used to help better understand neurological conditions, such as Alzheimer's disease and amyotrophic lateral sclerosis (ALS), as well as psychiatric diseases and brain injuries. Long-read sequencing technology provides comprehensive coverage to fully characterize transcriptomes of neurons and glial cells, providing insight into brain development, plasticity, and circuit function. Single-cell long-read sequencing can help uncover cellular heterogeneity and transcription signatures in specific cell types in the brain, as well as support research on developmental trajectories and cell lineages. Comprehensively, the Iso-Seq method can be leveraged to deepen understanding of brain function, development, and disease.

Iso-Seq in action

The full-length and single-cell Iso-Seq methods were used in a study by Patowary, et al. (2023) to specify cell types and characterize isoform diversity in the developing human brain. Iso-Seq data enabled detection of over 200,000 isoforms, of which 72.6% were determined to be novel and many predicted to be functionally impactful. The authors also identified 27 Mb of the genome that is transcriptionally active during brain development but not currently annotated in GENCODE. Single-cell Iso-Seq data enabled identification of sub-clusters and transitory states not discernible using short-read gene-based single-cell RNA sequencing.

Rare disease

While whole exome or whole genome sequencing is often used to solve rare disease cases, a lack of functional evidence hinders variant interpretation and the understanding of disease etiology. In cases where DNA sequencing is uninformative, a growing body of research suggests that RNA sequencing may have diagnostic utility in rare disease. With the Iso-Seq method, researchers can discover aberrant expression or splicing, prioritize candidate variants, and identify

allele-specific expression patterns. The addition of the Iso-Seq method to the sequencing toolbox furthers the understanding of the role of isoforms in disease pathogenesis.

Iso-Seq in action

Stergachis, et al. (2023) utilized the Iso-Seq method to investigate the function of an intronic SNP identified in an autosomal recessive case of axonal type 2a Charcot-Marie-Tooth disease. The candidate variant disrupted a branch point and led to aberrant splicing in intron 6. With the Iso-Seq method, the authors characterized five distinct altered splicing transcripts. The authors also found that despite the presence of the canonical *MFN2* transcript, the variant led to deficient MFN2 protein, a causal mechanism for the disease.

Association studies linking isoforms to function

Genome-wide association studies (GWAS) have revealed many loci associated with alternative splicing, and short-read RNA sequencing has resulted in splicing QTLS (sQTLs) shown to colocalize with GWAS loci.

However, GWAS and sQTL alone cannot provide isoform-level information on these potentially disease-causal genes. Using Kinnex sequencing, which can predict full-length proteins, the effects of sQTL can be better interpreted by providing a functional context.

Iso-Seq in action

Abood et al. (2023) used the Iso-Seq method to characterize isoform expression changes associated with bone mineralization in a hFOB (human fetal osteoblast) system. Using GWAS, sQTL, mass spectrometry data, and the Iso-Seq method, the authors were able to infer the effects of sQTLs on the protein isoform products they encode. For example, a sQTL colocalization implicated *TPM2* in association with bone mineral density (BMD). Iso-Seq data showed four primary *TPM2* isoforms expressing a pair of mutually exclusive exons (exons 6 and 7) and alternative last exons (exons 10 and 11). siRNA knockdown confirmed KD of isoforms containing exon 6 resulted in significant increase in mineralization, while KD of isoforms containing exon 7 resulted in a decrease in mineralization. The authors conclude that "TPM2 *isoforms have distinct isoforms with respect to osteoblast activity and are likely regulars of BMD*".

Full-length vs. single-cell resolution

	Full-length	Single-cell
PacBio advantages	• Achieve full-length isoform resolution (1–10 kb) without the need for assembly • Identify rare or novel transcripts and alternative splicing events, and gene fusions • Predict ORFs • Discover allele-specific isoform expression • Reveal differential isoform expression and isoform usage	• Provide full-length isoform information (1–5 kb*) at the individual cell level • Reveal cellular heterogeneity and signals • Identify cell-specific, allele-specific isoform expression
Example research topics	• Explore cell atlas projects to characterize different cell types and interactions • Identify novel splicing events to investigate mechanisms of rare disease • Compare isoform expression between diseased vs. healthy or treated vs. untreated samples to support discovery of biomarkers or therapeutic targets	• Reveal the complexity and diversity of cell types in the heart to better understand mechanisms of cardiac function and disease • Map immune cell activities, their microenvironment interactions, and response to treatment or infectious diseases • Understand human organ development • Follow cell proliferation and migration
Supporting products & platforms	• Iso-Seq express 2.0 kit and *Kinnex full-length RNA kit* • Sequel® II/IIe and Revio™ systems	• *Kinnex single-cell RNA kit* • Sequel II/IIe and Revio systems
Example datasets	Kinnex Universal Human Reference RNA dataset	Kinnex PBMC dataset

*Transcript length distribution limited by single-cell platforms used to generate single-cell cDNA

102-326-576 REV02 NOV2023

Resources and references

Resources

1. Procedure & checklist – Preparing Kinnex libraries using the Kinnex full-length RNA kit
2. Procedure & checklist – Preparing Kinnex libraries using Kinnex single-cell RNA kit
3. Application note – *Kinnex full-length RNA kit* for isoform sequencing
4. Application note – *Kinnex single-cell RNA kit* for single-cell isoform sequencing
5. Application note – Bioinformatics tools for full-length isoform sequencing
6. Iso-Seq Bioinformatics resources

References

Abood, A., et al. (2023). Long read proteogenomics to connect disease-associated sQTLs to the protein isoform effectors of disease. *bioRxiv*, 10.1101/2023.03.17.531557.

Al'Khafaji et al. (2023). High-throughput RNA isoform sequencing using programmed cDNA concatenation. *Nature Biotechnology*, 1-5. https://doi.org/10.1038/s41587-023-01815-7

Koboldt, D. C., et al. (2021). PTEN somatic mutations contribute to spectrum of cerebral overgrowth. *Brain*, *144*(10), 2971-2978. 10.1093/brain/awab173.

Leung, S. K., et al. (2021). Full-length transcript sequencing of human and mouse cerebral cortex identifies widespread isoform diversity and alternative splicing. *Cell Reports*, *37*(7), 110022. 10.1016/j.celrep.2021.110022.

Mehlferber, M. M., et al. (2022). Characterization of protein isoform diversity in human umbilical vein endothelial cells via long-read proteogenomics. *RNA Biology*, *19*(1), 1228-1243. 10.1080/15476286.2022.2141938.

Palmer, C. R., et al. (2021). Altered cell and RNA isoform diversity in aging Down syndrome brains. *Proceedings of the National Academy of Sciences*, *118*(47), e2114326118. 10.1073/pnas.211432611.

Patowary, A., et al. (2023). Cell-type-specificity of isoform diversity in the developing human neocortex informs mechanisms of neurodevelopmental disorders. bioRxiv, 10.1101/2023.03.25.534016.

Salz, R., et al. (2022). SUsPECT: A pipeline for variant effect prediction based on custom long-read transcriptomes for improved clinical variant annotation. *bioRxiv*, 10.1101/2022.10.23.513417.

Stergachis, A. B., et al. (2023). Full-length isoform sequencing for resolving the molecular basis of Charcot-Marie-Tooth 2A. *bioRxiv*, 10.1101/2023.02.07.526487.

Tseng, E., et al. (2019). The landscape of SNCA transcripts across synucleinopathies: new insights from long reads sequencing analysis. *Frontiers in Genetics*, *10*, 584. 10.3389/fgene.2019.00584.

Learn more about RNA sequencing: https://www.pacb.com/human-rna