Kinnex-IsoSeq-WDL: Workflows for High Throughput Analysis of HiFi Kinnex Full-Length RNA Data

8159T

Jocelyne Bruand[1], Tom Mokveld[1], Megan Schertzer[2], Elizabeth Tseng[1], Ales Varabyou[1], Xinchang Zheng[13], Zev Kronenberg[1], Khi Pin Chua[1], Jonathan R. Belyeu[1], Ryan Gossart[3], Margaret R. Starostik[4], Jonas Gustafson[5], Rebecca Martin[6], Joy Goffena[5,6], Katherine M. Munson[5], Kaitlyn Sun[5], Evan E. Eichler[5,7], Stephen B. Montgomery[9], Michael Schatz[4], Rob Patro[10], Michael Love[11], Winston Timp[4], Rajiv McCoy[4], Danny E. Miller[5,8,12], Michael A Eberle[1], Gloria Sheynkman[2], Fritz J. Sedlazeck[13,14], Yerbol Kurmangaliyev[3], Egor Dolzhenko[1]

[1] Pacific Biosciences, Menlo Park, CA, USA; [2] University of Virginia, Charlottesville, VA; [3] Brandeis University, Waltham, MA; [4] Johns Hopkins University, Baltimore, MD; [5] University of Washington, Seattle, WA; [6] Seattle Children's Research Institute, Seattle, WA; [7] Howard Hughes Medical Institute, Seattle, WA; [8] Seattle Children's Hospital, Seattle, WA; [9] Stanford University, Stanford, CA; [10] University of Maryland, College Park, MD; [11] University of North Carolina at Chapel Hill, Chapel Hill, NC; [12] Brotman Baty Institute for Precision medicine, Seattle, WA; [13] Baylor College of Medicine, Houston, TX; [14] Rice University, Houston, TX

Introduction

Long-read RNA sequencing enables direct observation of full-length RNA molecules, revealing transcript diversity not captured by short-read sequence data. We present Isocall, an efficient multi-sample isoform detection tool, together with workflows to enable population-scale full-length RNA analysis.

Long-read sequencing captures transcriptomic complexity

The vast array of cellular isoforms generated by alternative splicing introduces a layer of transcriptomic complexity that cannot be reliably resolved through short-read data. Long-read sequencing allows us to capture the true combination of alternative exons, transcription start sites, and polyadenylation signals.

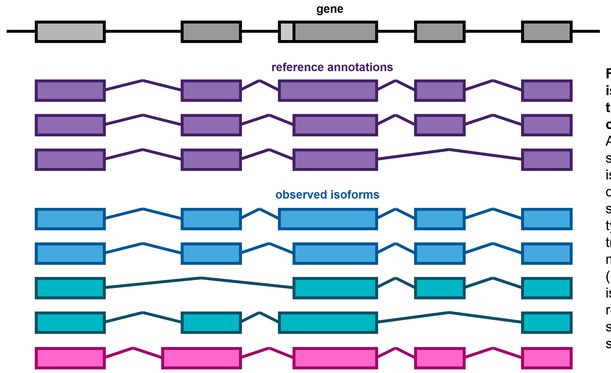


Figure 1. Full-length isoforms capture transcriptomic complexity. Alternative splicing drives structural diversity among isoforms. Reference databases catalog known structures (purple). Within a typical sample, observed transcripts include those matching known isoforms (blue) alongside novel isoforms, which either recombine known splice sites (teal) or feature novel splice sites (magenta).

Kinnex full-length RNA increases throughput for transcript sequencing

By concatenating transcripts into larger fragment libraries, Kinnex increases throughput for full-length RNA transcript sequencing, enabling large-scale studies at high-resolution. New tools and workflows improve and simplify data analysis.

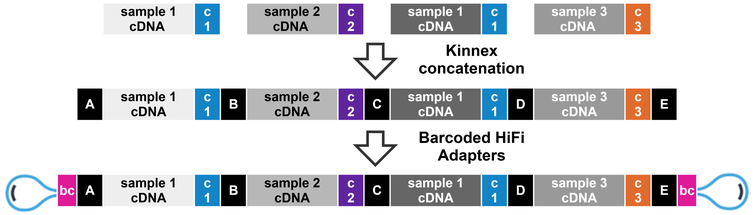


Figure 2. Overview of Kinnex Full-Length RNA. After cDNA generation, Kinnex adapters are attached to the transcripts via PCR. These are then concatenated into a long molecule during array formation. The resulting library can then be cleaned up and sequenced. The library is barcoded at two levels: the original barcode incorporated on the transcript and the HiFi adapter barcode incorporated after concatenation.

Isocall enables population-scale joint isoform calling

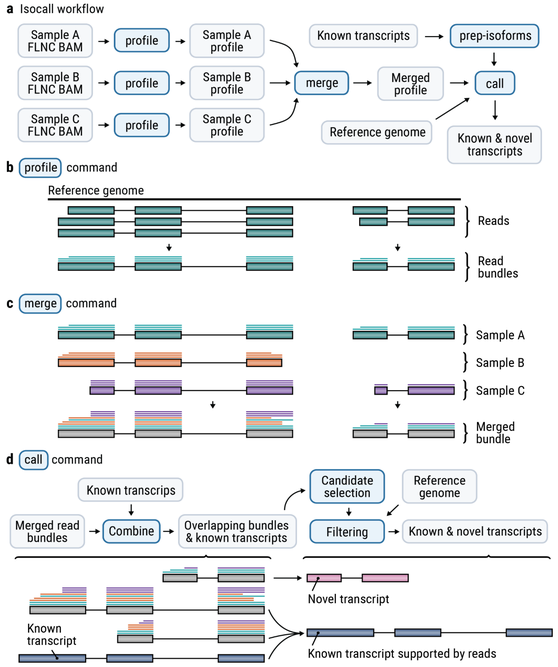


Figure 3. Overview of the Isocall workflow.
(a) Full-length non-concatemer (FLNC) reads aligned to the reference genome are converted into per-sample transcript profiles, merged across samples, and jointly analyzed with the known transcript annotations to call known and novel transcripts present in the dataset.
(b) `Isocall profile` extracts exon chains from aligned reads and groups reads with matching intron chains into compact read bundles.
(c) `Isocall merge` combines matching bundles across samples while retaining each read's sample of origin. Colors depict different samples.
(d) `Isocall call` integrates merged read bundles with known transcripts. Bundles matching an annotated intron chain are reported as known transcripts and the reads they explain are not used for novel transcript discovery; annotated transcripts not supported by reads are not reported. Novel transcripts are then selected from the remaining read bundles by iteratively choosing the candidate with the most compatible reads and are filtered for internal priming artifacts and low relative abundance.

Kinnex-IsoSeq-WDL streamlines the end-to-end analysis

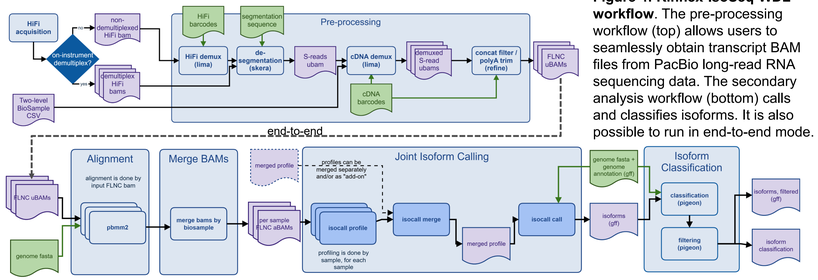


Figure 4: Kinnex-IsoSeq-WDL workflow. The pre-processing workflow (top) allows users to seamlessly obtain transcript BAM files from PacBio long-read RNA sequencing data. The secondary analysis workflow (bottom) calls and classifies isoforms. It is also possible to run in end-to-end mode.

Isocall provides tunable precision

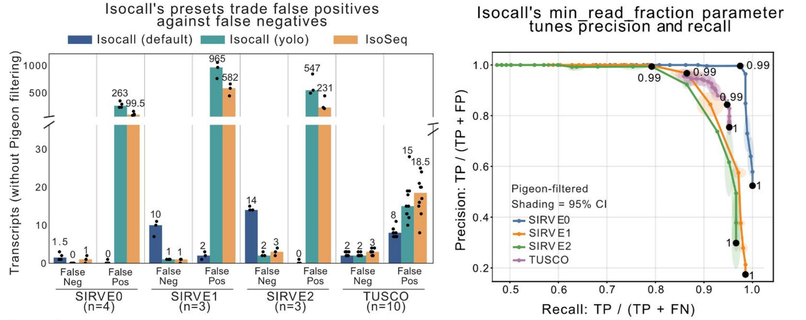


Figure 5. Isocall provides tunable precision.
(left) False negatives and false positives, without Pigeon filtering, called by Isocall (default and yolo presets) and IsoSeq across SIRV spike-in mixes E0, E1, and E2 and the TUSCO endogenous gene set in samples from Wissel et. al[1]
(right) Pigeon-filtered precision and recall across min_read_fraction values (0.85 to 1.00) for SIRV mixes and the TUSCO gene set.

Isocall provides efficient cohort-scale analysis

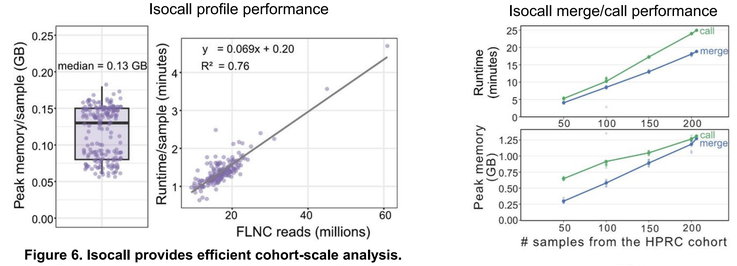


Figure 6. Isocall provides efficient cohort-scale analysis.
(left) Peak memory and runtime for the Isocall profile step, which runs independently on each HPRC sample.
(right) Runtime and peak memory for the Isocall merge and call steps, which run jointly across all samples.

Conclusions and References

- Isocall enables population-scale joint isoform calling by analyzing hundreds of samples.
- RNA workflows streamline analysis, providing a sample-to-answer solution.

1. Wissel, David, Madison M. Mehlferber, Khue M. Nguyen, et al. 2026. "A Systematic Benchmark of High-Accuracy PacBio Long-Read RNA Sequencing for Transcript-Level Quantification." Genome Biology 27 (1): 110.
2. Liao, Wen-Wei, Mobin Asri, Jana Ebler, et al. 2023. "A Draft Human Pangenome Reference." Nature 617 (7960): 312–324.

PacBio Kinnex

Isocall GitHub

Isocall pre-print
10.64898/2026.09.08.749180v1

Kinnex-IsoSeq-WDL
(GitHub)

Research use only. Not for use in diagnostic procedures. ©2026 Pacific Biosciences of California, Inc. ("PacBio"). All rights reserved. Information in this document is subject to change without notice. PacBio assumes no responsibility for any errors or omissions in this document. Certain notices, terms, conditions and/or use restrictions may pertain to your use of PacBio products and/or third-party products. Refer to the applicable PacBio terms and conditions of sale and to the applicable license terms at pacb.com/license. Pacific Biosciences, the PacBio logo, PacBio, Circulomics, Omniome, SMRT, SMRTbell, Iso-Seq, Sequel, Nanobind, Revio, Revio SPRQ, Kinnex, PureTarget, SPRQ, Vega and Vega SPRQ are trademarks of PacBio.