

# WHAT'S THE VALUE OF SEQUENCING FULL-LENGTH RNA TRANSCRIPTS?

## THE STUDY OF GENOMICS HAS REVOLUTIONIZED OUR UNDERSTANDING OF SCIENCE

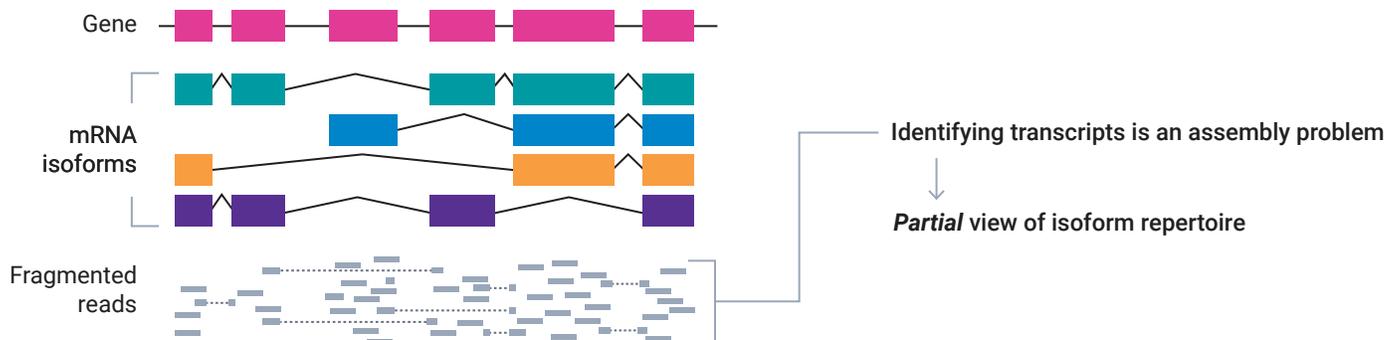
But the field of transcriptomics grew with the need to explore the functional impacts of genetic variation. While different tissues in an organism may share the same genomic DNA, they can differ greatly in what regions are transcribed into RNA and in their patterns of RNA processing. By reviewing the history of transcriptomics, we can see the advantages of RNA sequencing — using a full-length transcript approach — become clearer.

### REACHING FOR THE TRANSCRIPTOME

Even before genome sequencing became commonplace, scientists were able to measure gene expression activity using hybridization approaches like Northern blots, or later, microarrays. While these techniques provided a rigorous method for RNA measurement, they were limited in that they could only measure known transcripts. Additionally, many of these methods made the subtle differences between transcript isoforms difficult to decipher.

As next-generation sequencing (NGS) grew, researchers had the ability to interrogate the activity of all expressed genes including those that were previously unknown. This was a huge leap forward in the ability to characterize the transcriptome in an organism. The approach, known as RNA-seq, quickly replaced microarrays in many labs due to the enormous amount of information it could generate.

Understanding gene activity — not just the genes encoded by the genome, but the ones that are turned on at a specific time or in a specific cell or tissue — is critical for elucidating biological consequences.



The RNA-Seq approach yields reads that do not span multiple exon junctions

But even as RNA-seq expanded, scientists grew concerned about the accuracy of results when NGS short reads had to be pieced back together computationally to create whole transcripts. This is especially difficult in complex eukaryotes where one gene can generate many isoforms that differ in their transcription start site and alternative RNA processing.<sup>1</sup>

In a Mendelspod interview, Stanford genomics expert Mike Snyder described the RNA-seq process this way: “We take RNA, we blow it up into little fragments, and then we try and assemble them back together to see what the transcriptome looked like in the first place. ... You can’t always figure out which parts of the puzzle belong together.”

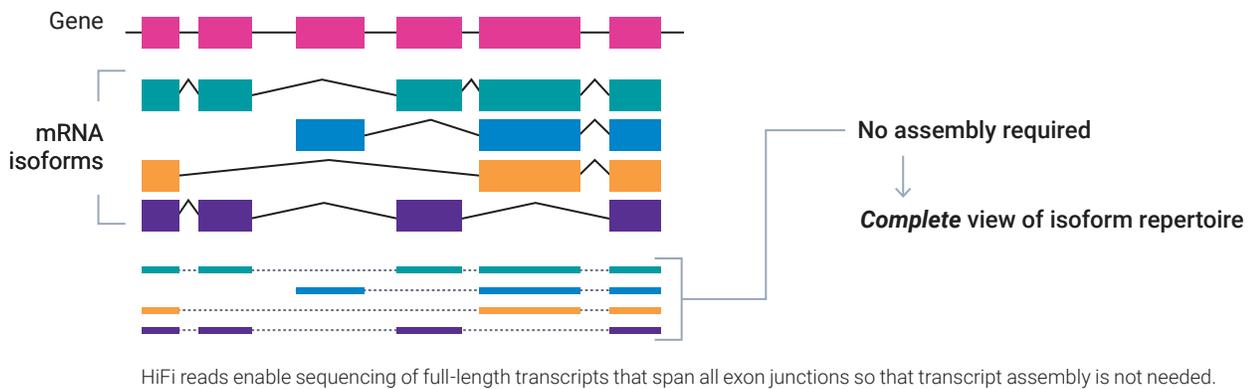
A similar concept was presented by Ian Korf in a *Nature Methods* article,<sup>2</sup> where he equated RNA-seq to reassembling magazine articles that had been put through a shredder. Suffice it to say, there have been concerns about the need for extensive computational inference when using an RNA-seq approach.

## CAPTURING THE WHOLE TRANSCRIPTOME WITH THE ADVANTAGES OF LONG READS

What scientists really needed went beyond RNA-seq; for the most accurate view of biology possible, they had to have a way to capture isoforms in their entirety. The introduction of long-read sequencing answered this need.

With PacBio® Single Molecule, Real-Time (SMRT®) sequencing, researchers can generate highly accurate long reads – or HiFi reads – that are tens of kilobases long. That’s enough to capture most transcripts in a single sequencing read, with no downstream bioinformatic assembly required. This became known as the Iso-Seq® method because it enabled full-length isoform sequencing, which can be used to explore a transcriptome in its entirety or in a targeted fashion to examine individual genes.

As scientists began using SMRT sequencing for transcriptome projects, a common theme emerged: long-read studies, even of the most well characterized organisms, routinely showed far more transcript diversity than previously observed. It quickly became clear that short-read NGS platforms greatly underestimated the diversity of expressed isoforms.



### The Iso-Seq method can be used to:

- Identify alternative splicing events
- Characterize alternative start/end sites
- Find fusion transcripts
- Study allele-specific isoforms
- Discover non-coding RNAs

The Iso-Seq method provides end-to-end workflow, requiring one day each for library preparation, SMRT sequencing, and bioinformatics analysis.

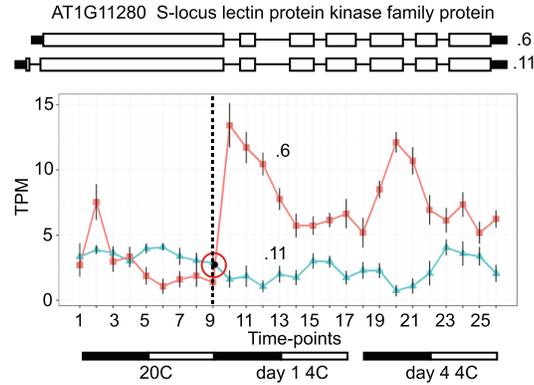


# 1

## ALTERNATIVE START SITES

Identify transcription start site (TSS) changes associated with important traits.

### SHORTER TRANSCRIPT USING ALTERNATIVE TRANSCRIPTION START SITE ASSOCIATED WITH COLD RESPONSE IN ARABIDOPSIS



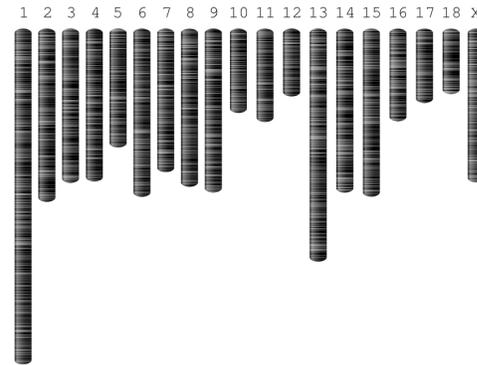
Zhang et al. (2022) A high-resolution single-molecule sequencing-based *Arabidopsis* transcriptome using novel methods of Iso-seq analysis. *Genome Biology*.

# 2

## GENOME ANNOTATION

Gain a complete view of the porcine transcriptome including the identification of genes missing in the assembly.

### THE ISO-SEQ METHOD VERIFIED KNOWN GENES AND RESCUED MISSING GENES DURING ANNOTATION OF THE NEW PIG GENOME ASSEMBLY



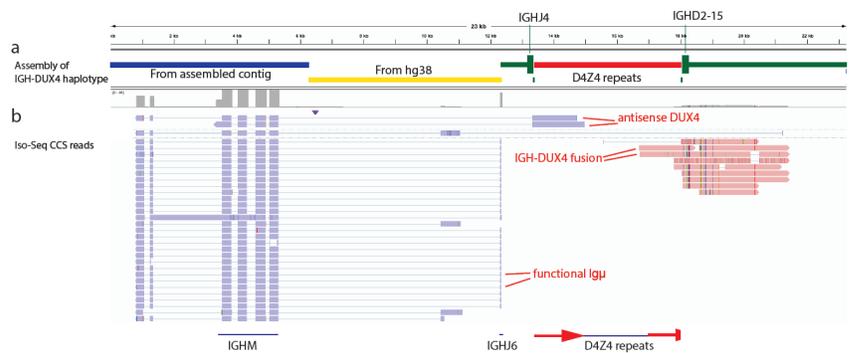
Warr, A., et al. (2020) An improved pig reference genome sequence to enable pig genetics and genomics research. *GigaScience*, 9, 1–14.

# 3

## FUSION TRANSCRIPTS

Discover novel fusion transcripts in clinically relevant genes to understand disease mechanisms.

### THE ISO-SEQ METHOD REVEALS A FUNCTIONAL IGH, A GENE FUSION EVENT BETWEEN IGH AND DUX4, AS WELL AS AN ANTISENSE DUX4 GENE COPY



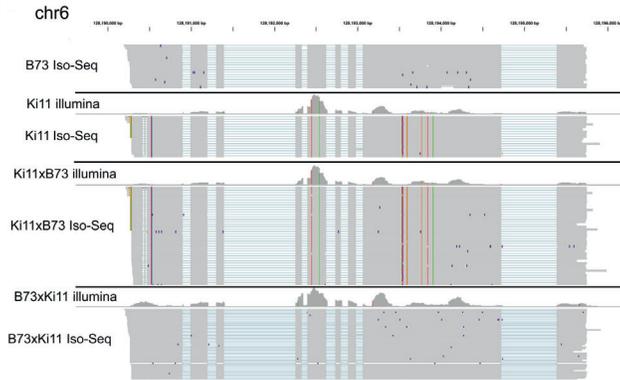
Tian, L., et al. (2019). Long-read sequencing unveils IGH-DUX4 translocation into the silenced IGH allele in B-cell acute lymphoblastic leukemia. *Nature Communications*, 10(1), 2789.

# 4

## ISOFORM VARIANT PHASING

Detect and phase mutations in disease genes for clinically relevant insights.

### ALLELE-SPECIFIC ISOFORM EXPRESSION IDENTIFIED USING THE ISO-SEQ METHOD IN MAIZE



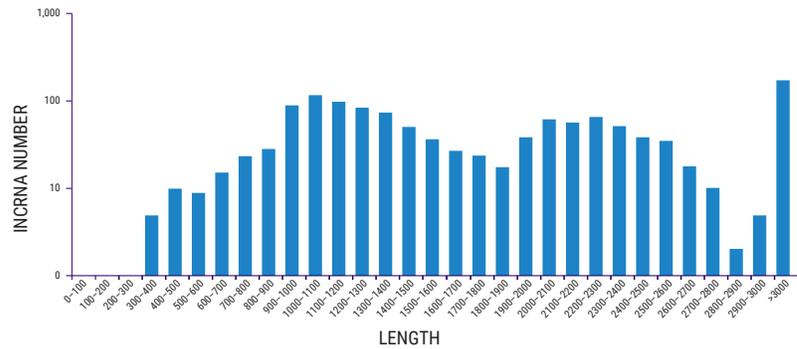
Wang et al. (2020) Variant phasing and haplotypic expression from long-read sequencing in maize. *Comm Biol*.

# 5

## LONG, NON-CODING RNAS

Identify long noncoding RNA for a more comprehensive view of transcriptional and translation regulation.

### THE LENGTH DISTRIBUTION OF lncRNA



Teng, K., et al., (2019) PacBio single-molecule long-read sequencing shed new light on the complexity of the *Carex breviculmis* transcriptome. *BMC Genomics*, 20(1), 789.

## READY TO GET STARTED WITH YOUR NEXT SEQUENCING PROJECT?

Full-length transcript sequencing offers many benefits to scientists looking to interrogate transcriptomes. Whether exploring the human transcriptome to better understand health and disease or studying plants and animals to advance agricultural or conservation efforts.



Learn more about the Iso-Seq method [pacb.com/isoseq](https://pacb.com/isoseq)



Connect with a PacBio scientist to get started: [pacb.com/scientist](https://pacb.com/scientist)

### KEY REFERENCES

1. Martin, J. A., & Wang, Z. (2011) Next-generation transcriptome assembly. *Nature reviews. Genetics*, 12(10), 671–682.
2. Korf, I. (2013) Genomics: the state of the art in RNA-seq analysis. *Nat Methods* 10, 1165–1166.

Research use only. Not for use in diagnostic procedures. © 2022 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](https://pacb.com/license). Pacific Biosciences, the PacBio logo, PacBio, Circulomics, Omniome, SMRT, SMRTbell, Iso-Seq, Sequel, Nanobind, and SBB are trademarks of PacBio. All other trademarks are the sole property of their respective owners.