PACIFIC BIOSCIENCES®

The Why, What, and How of the Iso-Seq Method:

Using Full-length RNA Sequencing to Annotate Genomes and Solve Diseases

Intro to the Iso-Seq Method

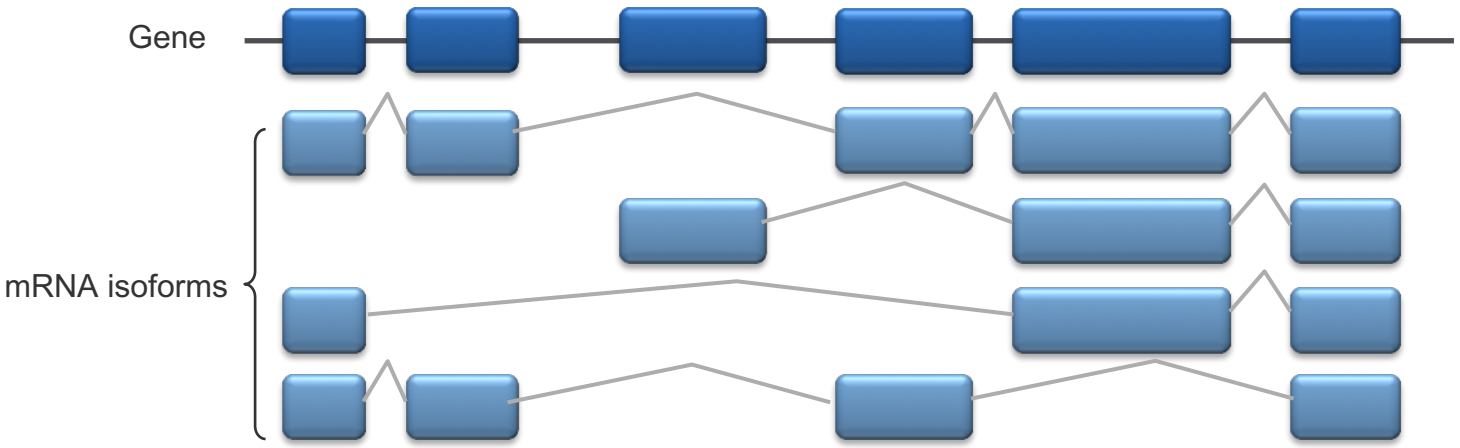
Jonas Korlach, Chief Scientific Officer

WHAT IS THE ISO-SEQ METHOD?

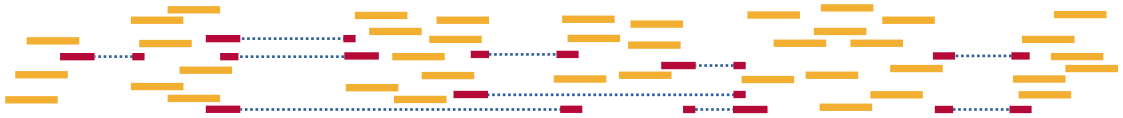
The PacBio Iso-Seq method is an end-to-end workflow for sequencing and analyzing full-length transcript isoforms.

1. Convert RNA → cDNA
2. cDNA → SMRTbell library
3. Sequence on the Sequel System
4. Generate circular consensus sequences (CCS)
5. Discover isoforms *de novo* with Iso-Seq analysis

WHY IS FULL-LENGTH RNA SEQUENCING USEFUL?



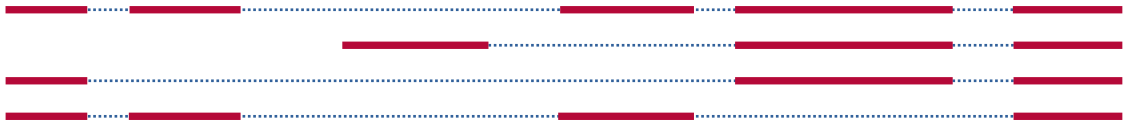
Short-read RNA-seq:



Reads spanning splice junctions

Insufficient Connectivity
Splice Isoform Uncertainty

Iso-Seq method:



Full-length cDNA Sequence Reads

Splice Isoform Certainty – No Assembly Required

KEY APPLICATIONS OF THE ISO-SEQ METHOD

Whole-genome Annotation

"I would like a reference catalog of all transcript isoforms detectable within a particular sample."

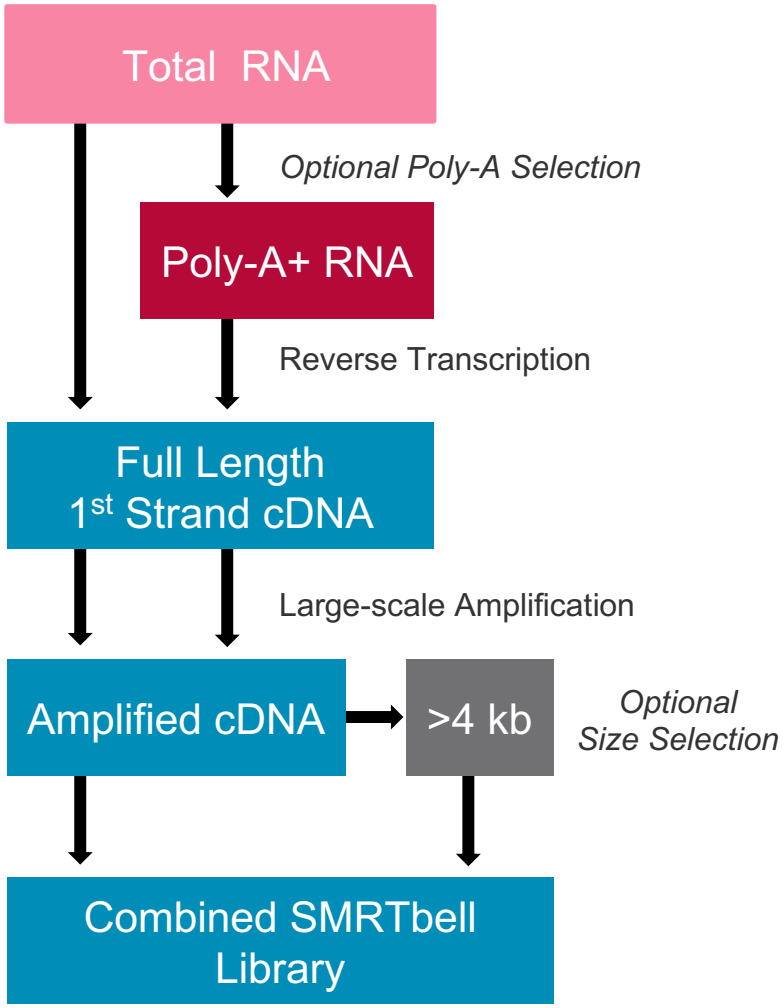
- Typically whole-transcriptome, non-quantitative
- Often included in *de novo* genome assembly projects
- Single tissue to several tissues
- Generates reference transcriptome for downstream RNA-seq studies

Gene-level Isoform Discovery

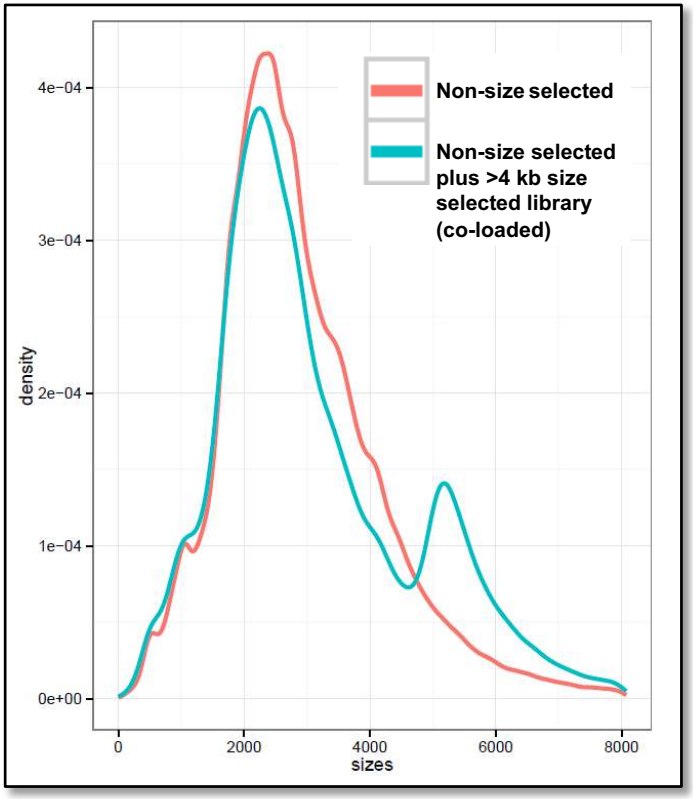
"Do alternative splicing or other transcription events play a role in a particular disease state?"

- Typically targeted, either cDNA amplicons or target capture
- Useful for detecting gene fusions, SNVs, allele-specific expression
- Cost-effectively multiplex many samples per single SMRT Cell
- Relative quantitation possible

SIMPLIFIED SEQUEL ISO-SEQ LIBRARY PREP



- Simplified library preparation
- Size selection optional



WHOLE-GENOME ANNOTATION: KEY PUBLICATIONS

Wang et al., **Unveiling the complexity of the maize transcriptome by single-molecule long-read sequencing**, *Nat Comm* (2016)

- First Iso-Seq application for whole genome annotation
- Multiplexed 6 different maize B73 tissues
- Obtained ~111 k high-quality transcripts
- Vastly improved existing annotation and incorporated to MaizeGDB v4

Wang et al., **A comparative transcriptional landscape of maize and sorghum obtained by single-molecule sequencing**, *Genome Research* (2018)

- Performed Iso-Seq method on maize and sorghum
- Comparative analysis of conserved and differentiated alternative splicing

WHOLE-GENOME ANNOTATION: KEY PUBLICATIONS

Kuo et al., **Normalized long read RNA sequencing in chicken reveals transcriptome complexity similar to human.** *BMC Genomics* (2017)

- Whole transcriptome sequencing of chicken
- Used 5’ cap normalized Iso-Seq libraries
- Obtained ~60 k high-quality transcripts (~29 k genes)
- Identified >20 k potential lncRNAs

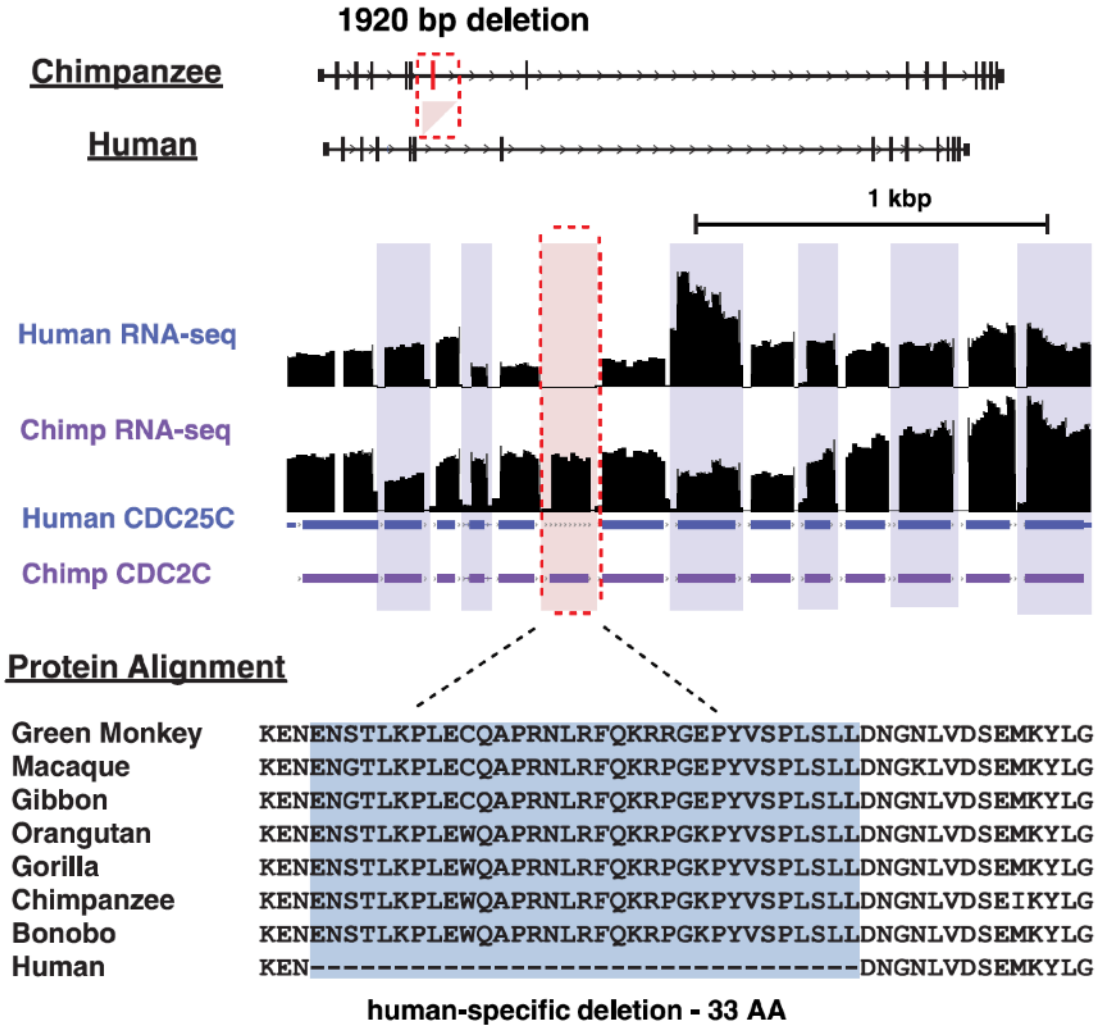
COMPARATIVE GENOME + TRANSCRIPTOME SEQUENCING

- Human, Chimp, and Orangutan
- *de novo* genome assembly using PacBio
- Iso-Seq + RNA-Seq for annotation

- Improved genome contiguity by 30- to 500-fold
- 83% of ape genome now in multi-species alignment
- Systematic SV discovery (~600 k in ape)
- Rare human-specific exonic deletion detected

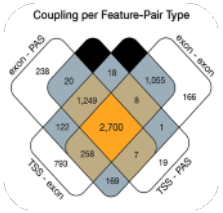
HUMAN SPECIFIC DELETIONS DETECTED BY CROSS-SPECIES ISO-SEQ COMPARISON

[Blog: Finding Human by sequencing our Ape relatives]



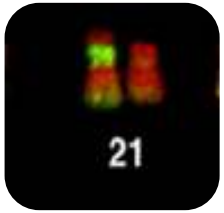
ISO-SEQ PUBLICATIONS: HUMAN GENES AND DISEASES

Treutlein et al., **Cartography of neurexin alternative splicing mapped by single-molecule long-read mRNA sequencing.** *Proc Natl Acad Sci* (2014)



Anvar et al., **Full-length mRNA sequencing uncovers a widespread coupling between transcription initiation and mRNA processing.** *Genome Biol.* (2018)

Kohli et al., **Androgen Receptor Variant AR-V9 Is Coexpressed with AR-V7 in Prostate Cancer Metastases and Predicts Abiraterone Resistance**, *Clinical Cancer Research* (2017)



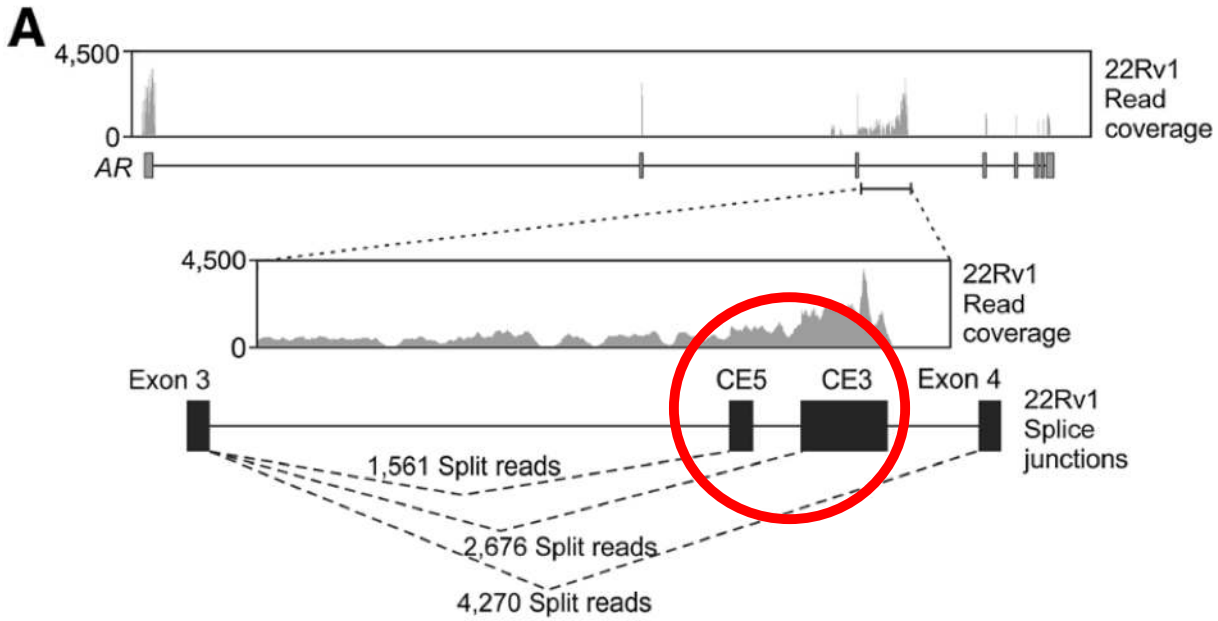
Deveson et al., **Universal Alternative Splicing of Noncoding Exons.** *Cell Systems* (2018)

Aneichyk et al., **Dissecting the Causal Mechanism of X-Linked Dystonia-Parkinsonism by Integrating Genome and Transcriptome Assembly**. *Cell* (2018)

PROSTATE CANCER AWARENESS

Kohli et al., **Androgen Receptor Variant AR-V9 Is Coexpressed with AR-V7 in Prostate Cancer Metastases and Predicts Abiraterone Resistance**, *Clinical Cancer Research* (2017)

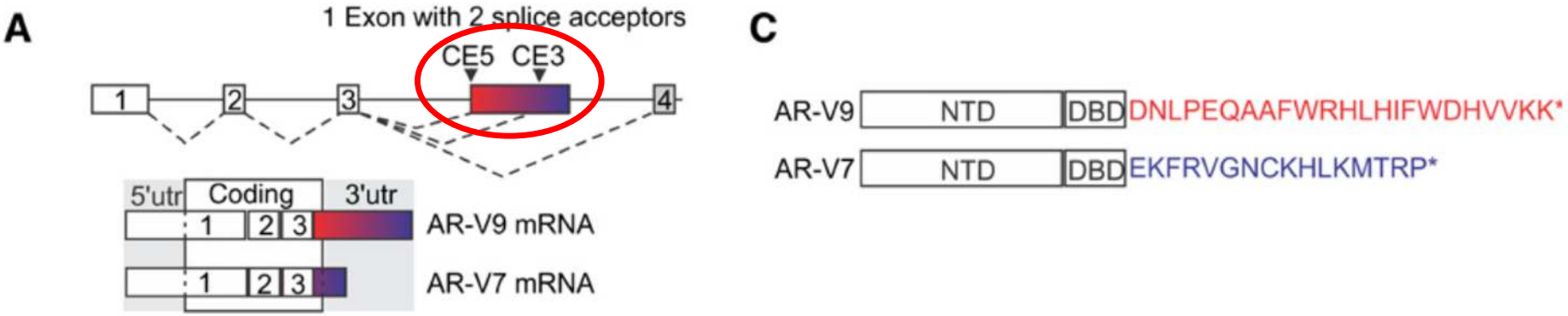
- Sequenced only Andogren Receptor gene (AR) in prostate cancer
- AR-V7 is a known variant that prohibits successful therapy in castration-resistant prostate cancer



PROSTATE CANCER AWARENESS

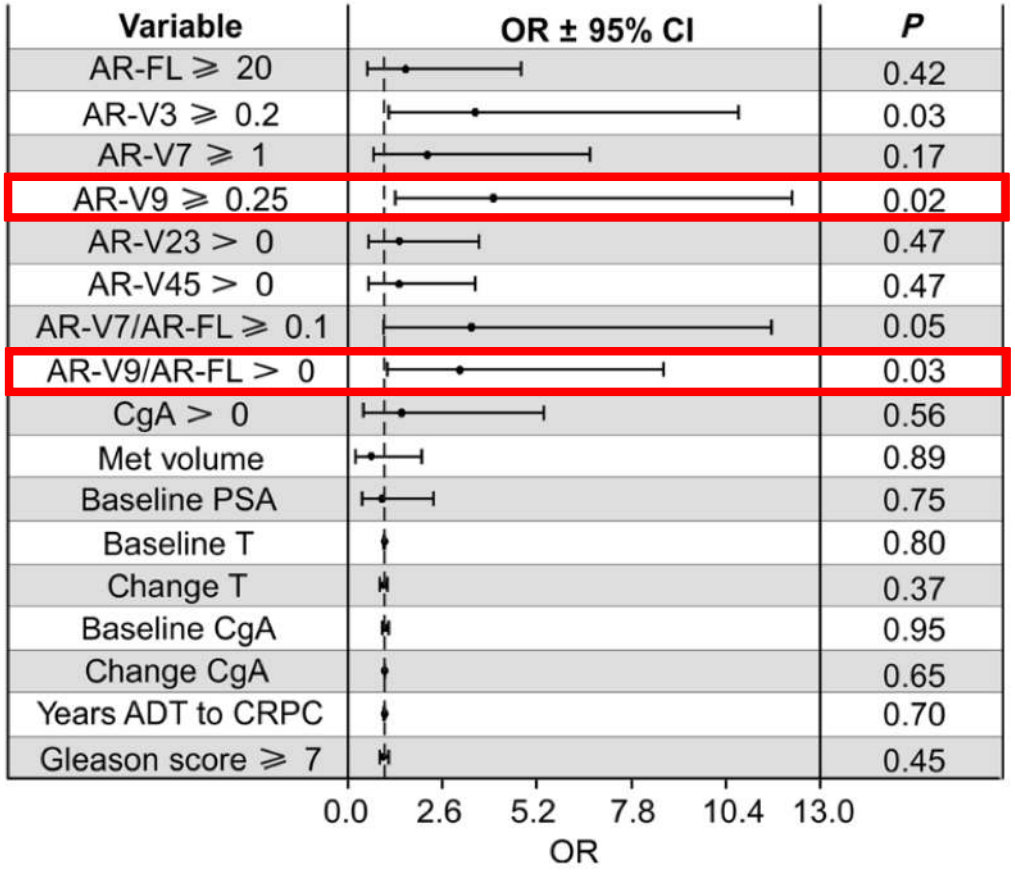
Kohli et al., **Androgen Receptor Variant AR-V9 Is Coexpressed with AR-V7 in Prostate Cancer Metastases and Predicts Abiraterone Resistance**, *Clinical Cancer Research* (2017)

- Iso-Seq data identified AR-V9 often co-expressed with AR-V7
- Iso-Seq data re-annotated the cryptic exons CE3 and CE5 as a single 3' exon with different splice sites

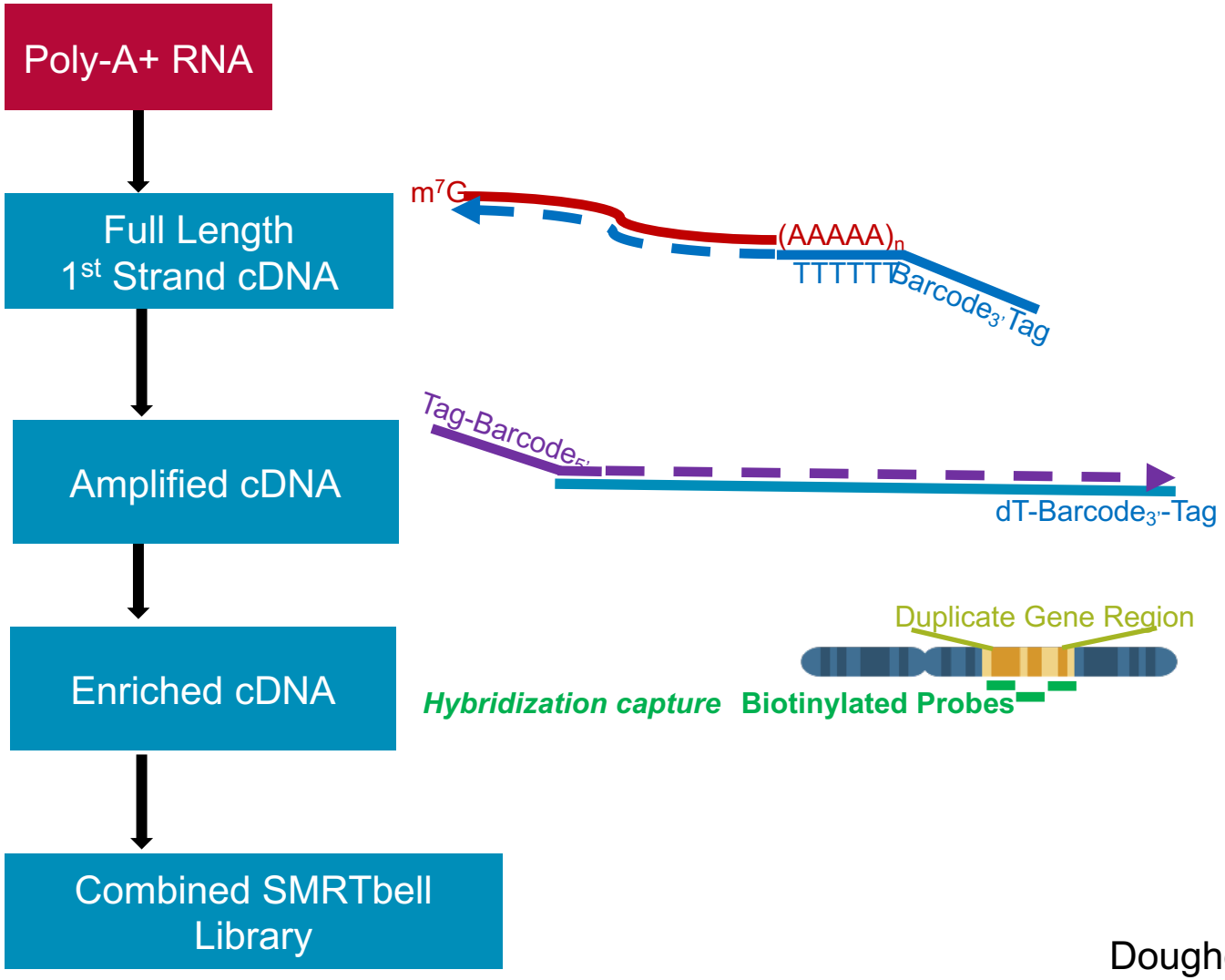


Kohli et al., **Androgen Receptor Variant AR-V9 Is Coexpressed with AR-V7 in Prostate Cancer Metastases and Predicts Abiraterone Resistance**, *Clinical Cancer Research* (2017)

- AR-V9 expression predictive of therapy resistance



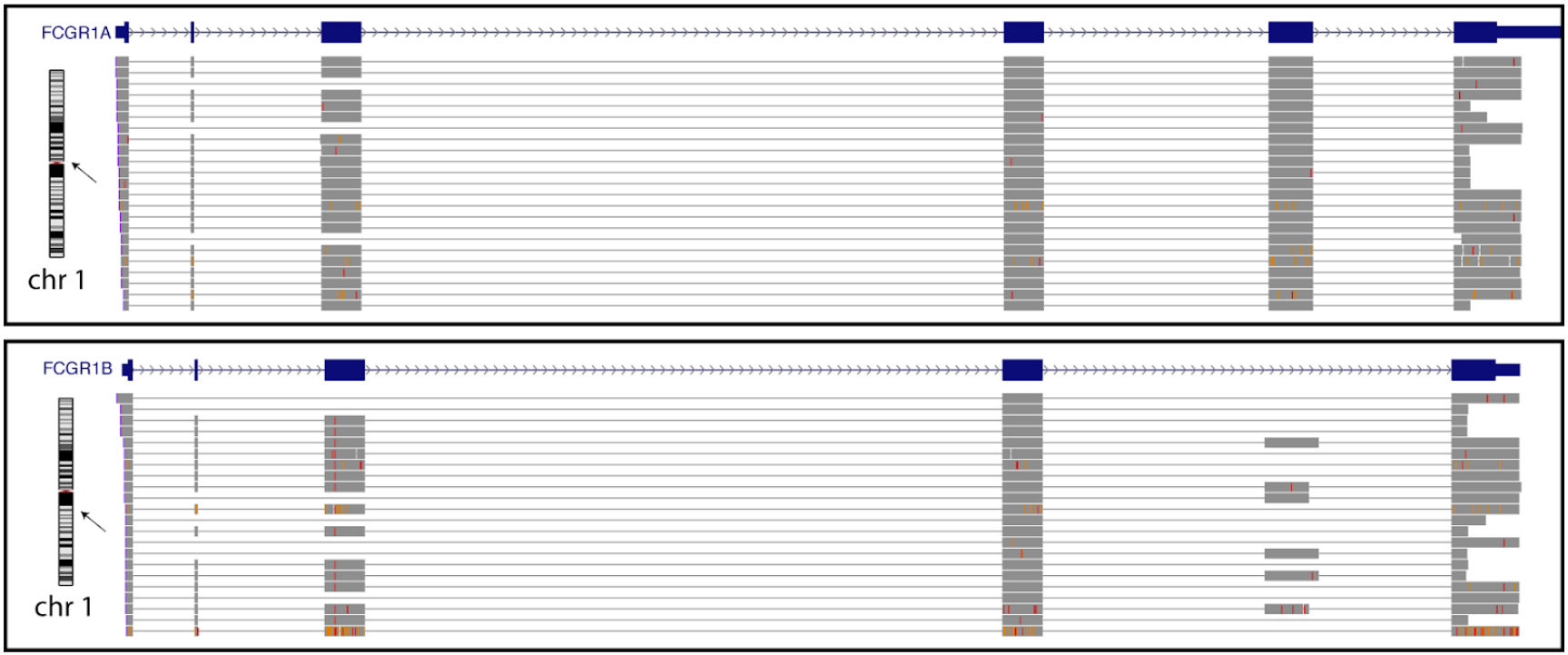
TARGETED ENRICHMENT OF SEGMENTAL DUPLICATED GENES



Dougherty et al. (accepted)

ISO-SEQ ANALYSIS CAPTURES SEGMENTAL DUPLICATED GENES

- FCGR1A and FCGR1B are > 99% similar



Dougherty et al. (accepted)

SINGLE-CELL APPLICATION

G&T-seq: parallel sequencing of single-cell genomes and transcriptomes

Iain C Macaulay[1], Wilfried Haerty[2,10], Parveen Kumar[3,10], Yang I Li[2,9], Tim Xiaoming Hu[2], Mabel J Teng[4], Mubeen Goolam[5], Nathalie Saurat[6], Paul Coupland[7], Lesley M Shirley[7], Miriam Smith[7], Niels Van der Aa[3], Ruby Banerjee[8], Peter D Ellis[7], Michael A Quail[7], Harold P Swerdlow[7,9], Magdalena Zernicka-Goetz[5], Frederick J Livesey[6], Chris P Ponting[1,2,11] & Thierry Voet[1,3,11]

The simultaneous sequencing of a single cell's genome and transcriptome offers a powerful means to dissect genetic variation and its effect on gene expression. Here we describe G&T-seq, a method for separating and sequencing genomic DNA and full-length mRNA from single cells. By applying G&T-seq to over 220 single cells from mice and humans, we discovered cellular properties that could not be inferred from DNA or RNA sequencing alone.

Karlsson and Linnarsson *BMC Genomics* (2017) 18:126
DOI 10.1186/s12864-017-3528-6

BMC Genomics

RESEARCH ARTICLE Open Access

Single-cell mRNA isoform diversity in the mouse brain

Kasper Karlsson[1] and Sten Linnarsson[2*]

Abstract

Background: Alternative mRNA isoform usage is an important source of protein diversity in mammalian cells. This phenomenon has been extensively studied in bulk tissues, however, it remains unclear how this diversity is reflected in single cells.

Results: Here we use long-read sequencing technology combined with unique molecular identifiers (UMIs) to reveal patterns of alternative full-length isoform expression in single cells from the mouse brain. We found a surprising amount of isoform diversity, even after applying a conservative definition of what constitutes an isoform. Genes tend to have one or a few isoforms highly expressed and a larger number of isoforms expressed at a low level. However, for many genes, nearly every sequenced mRNA molecule was unique, and many events affected coding regions suggesting previously unknown protein diversity in single cells. Exon junctions in coding regions were less prone to splicing errors than those in non-coding regions, indicating purifying selection on splice donor and acceptor efficiency.

Conclusions: Our findings indicate that mRNA isoform diversity is an important source of biological variability also in single cells.

Cold Spring Harbor Laboratory

bioRχiv

THE PREPRINT SERVER FOR BIOLOGY

HOME | ABO

Search

New Results

Single-cell isoform RNA sequencing (ScISOr-Seq) across thousands of cells reveals isoforms of cerebellar cell types.

Ishaan Gupta, Paul G Collier, Bettina Haase, Ahmed Mahfouz, Anoushka Joglekar, Taylor Floyd, Frank Koopmans, Ben Barres, August B Smit, Steven Sloan, Wenjie Luo, Olivier Fedrigo, M Elizabeth Ross, Hagen U Tilgner

doi: https://doi.org/10.1101/364950

This article is a preprint and has not been peer-reviewed [what does this mean?].

Abstract Info/History Metrics Preview PDF

ADDITIONAL REFERENCES

Study Target	Approach	Publications
Single gene	cDNA amplicons Targeted enrichment	Tseng et al. (FMR1) Kohli et al. (AR) Aneichyk et al. (XDP)
10-200 genes	Targeted enrichment	Goldfeder (AGBT2018) Deveson et al. (chr21)
lncRNA	Normalization Targeted enrichment	Kuo et al. (chicken) Lagarde (GENCODE)
Differential expression	Combine with RNA-seq	Chen et al. (garlic)
Whole Transcriptome	Standard cDNA library	Anvar et al. (MCF-7)
Single Cell Sequencing	Combine with UMIs	Macaulay (G&T-Seq) Karlsson (mouse brain) Tilgner (ScISOr-Seq)

SEQUEL SYSTEM ISO-SEQ EXPERIMENT SIZE

SMRT Cells (per sample)	Experimental Goals
<1	Targeted, gene-specific isoform characterization
1	General survey of full-length isoforms in a transcriptome (moderate to high expression levels) with or without size selection
1-2	A comprehensive survey of full-length isoforms in the transcriptome (per sample)
2+	Deep sequencing for comprehensive isoform discovery and identification of low abundance transcripts (per sample)

Sequel Performance (5.1):

- Up to 20 Gb per SMRT Cell
- 20-hour movie time
- 250 kb - 350 kb full-length non-chimeric (FLNC) reads

Analysis:

- IsoSeq2 or Iso-Seq3 (beta) for whole-genome annotation and targeted experiments

Planned Improvements:

- IsoSeq3 in SMRT Link 6.0
- Up to 40 Gb per SMRT Cell (6.0)
- More high-quality long transcripts

www.pacb.com