PACIFIC BIOSCIENCES®

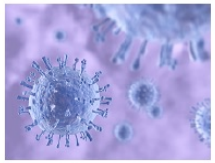
SMRT Sequencing – How it Works

Elizabeth Tseng, Principal Scientist, PacBio

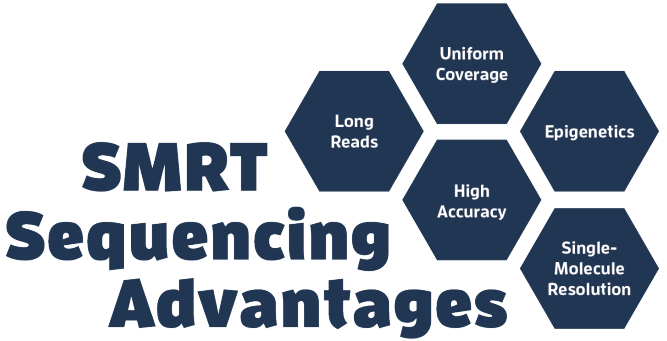
@Magdoll

SEQUENCE WITH CONFIDENCE

PacBio long-read sequencing provides access to the full spectrum of genetic variation, driving discovery across all fields of life science.

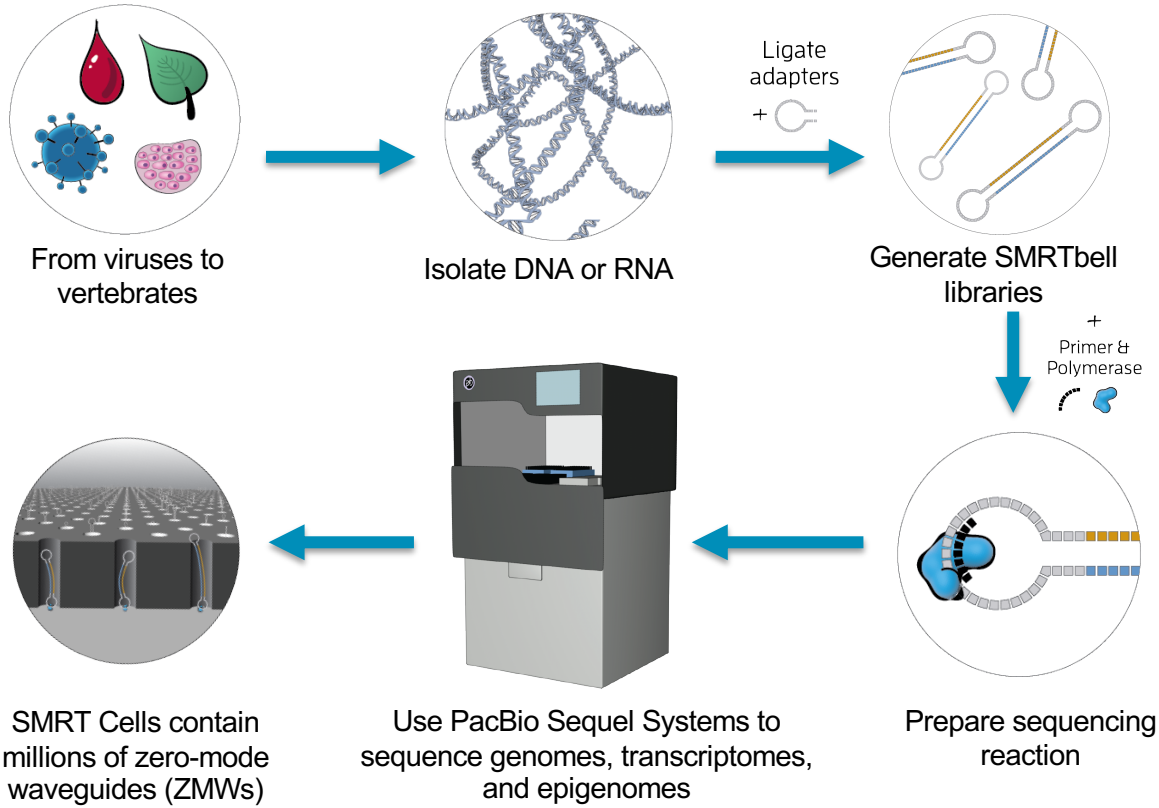



PacBio Systems are powered by **Single Molecule, Real-Time (SMRT) Sequencing**, a technology proven to produce exceptionally long reads with high accuracy.

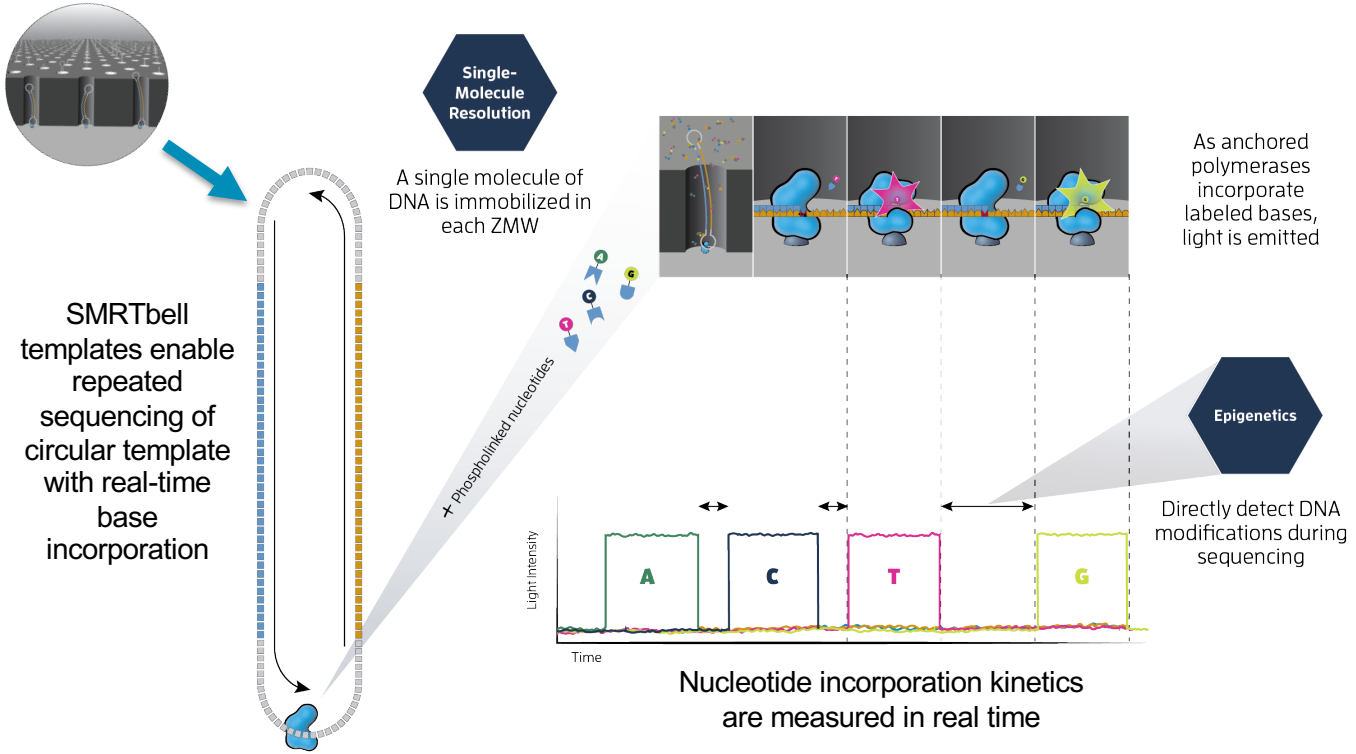


pacb.com/smrt

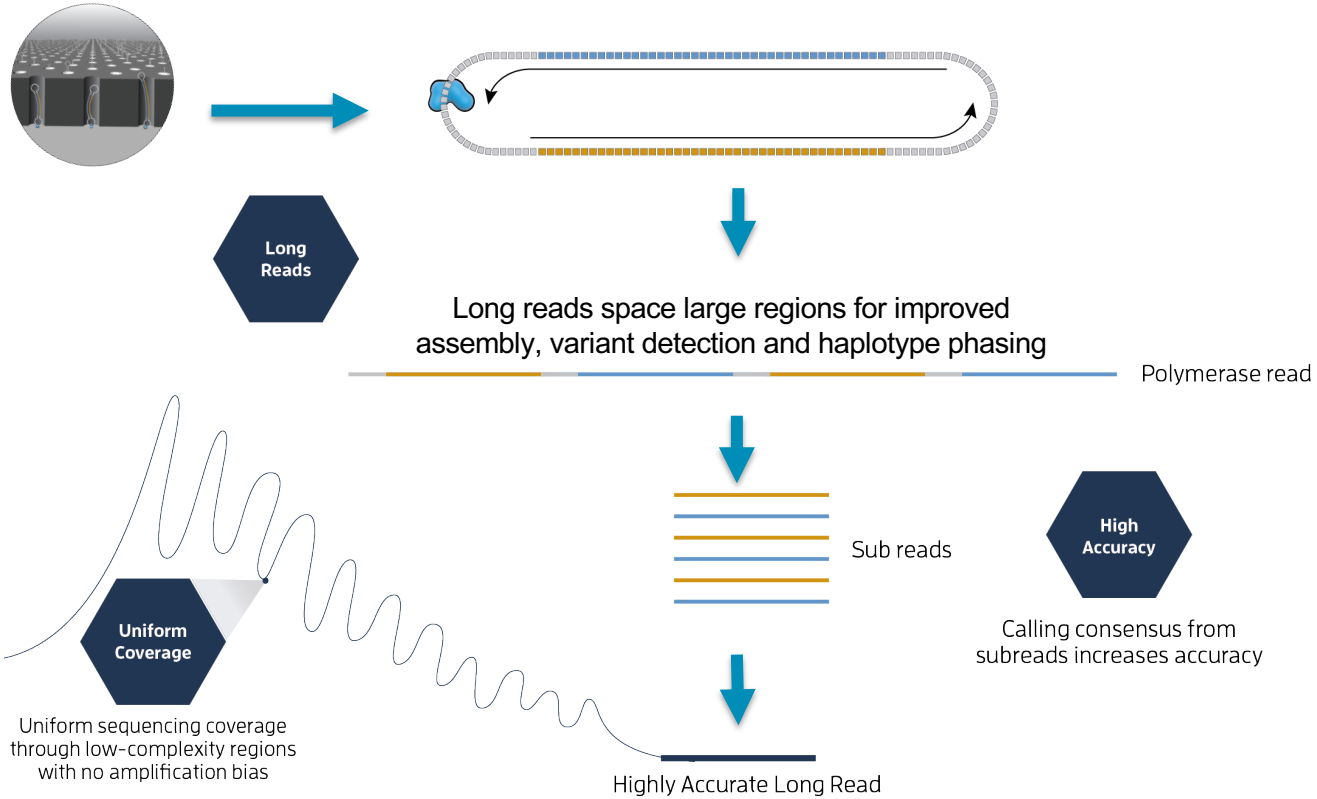
FROM SAMPLE TO SMRT SEQUENCING



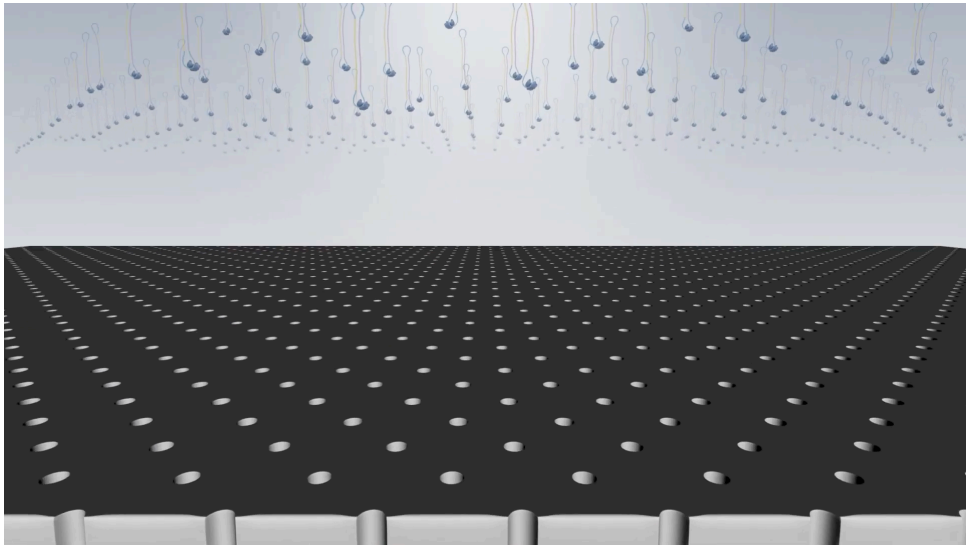
OUR CORE TECHNOLOGY



GENERATE HIGHLY ACCURATE LONG READS

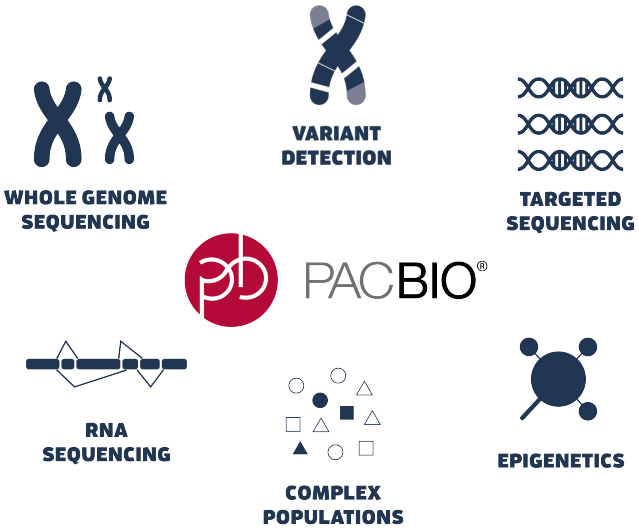


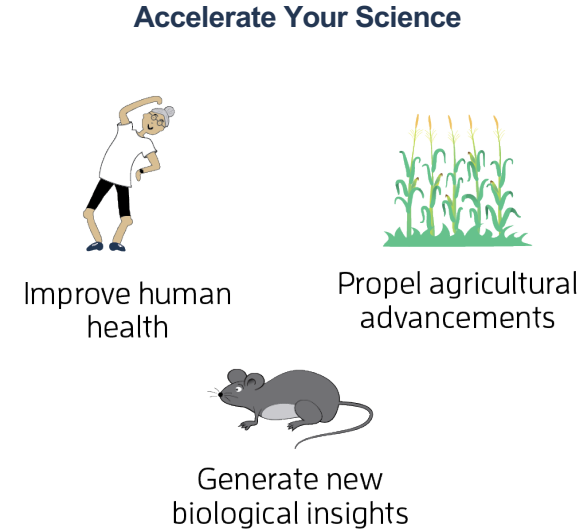
PACBIO SEQUENCING 101 CLIP



PacBio: Sequencing 101

SMRT SEQUENCING ENABLES THE FULL RANGE OF PACBIO APPLICATIONS



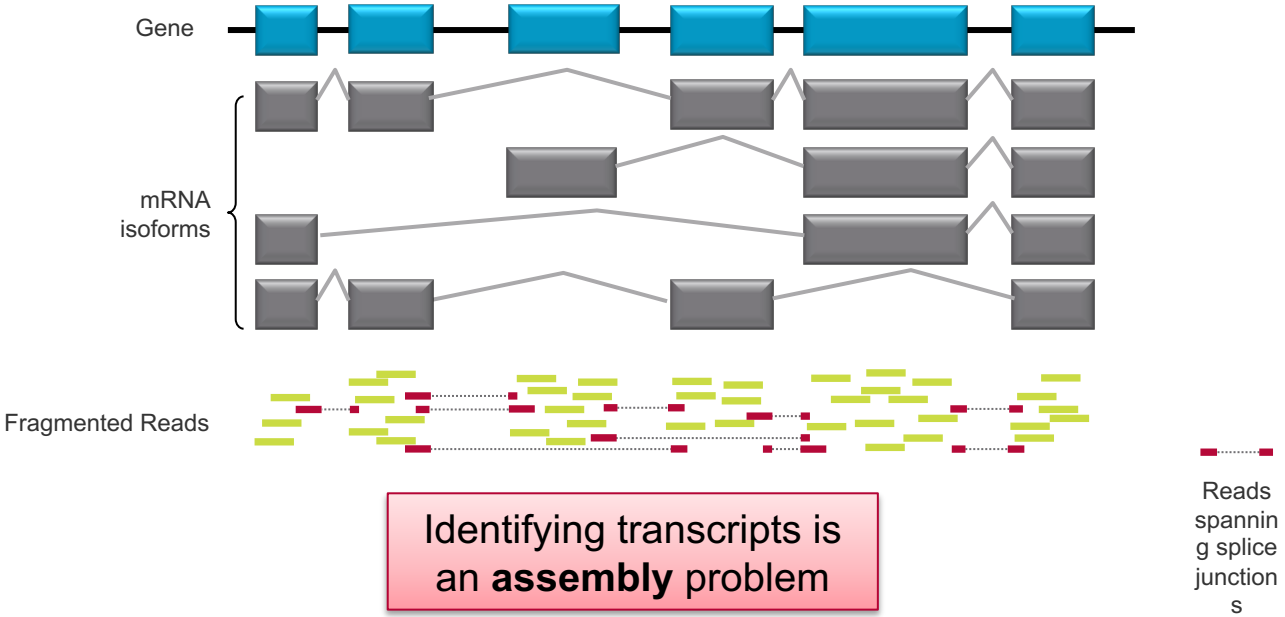


PACBIO®

The Iso-Seq Method

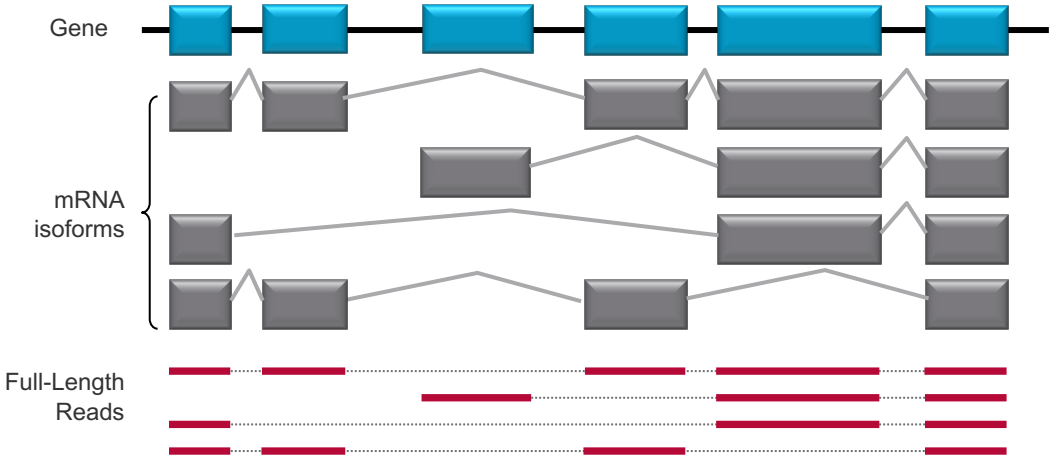
From Sample Prep to Analysis

TRADITIONAL RNA-SEQ

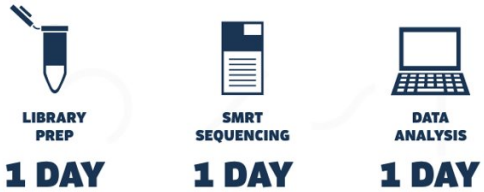


Identifying transcripts is an **assembly** problem

ISO-SEQ METHOD = FULL-LENGTH TRANSCRIPT SEQUENCING

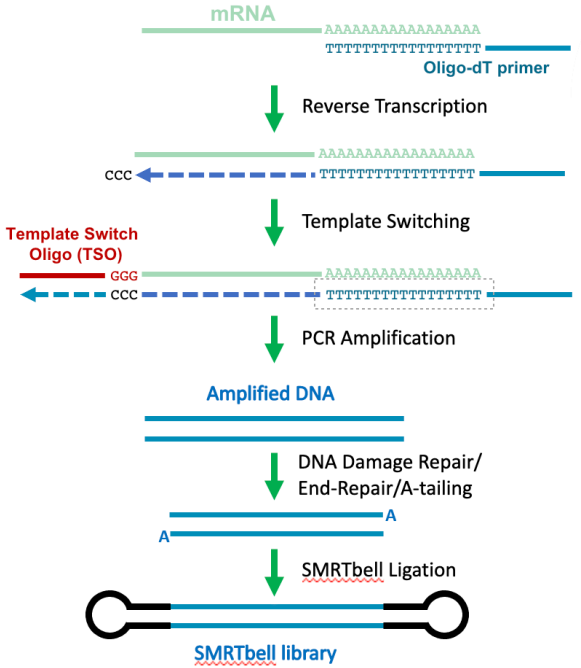


No assembly required


LIBRARY PREP
1 DAY
SMRT SEQUENCING
1 DAY
DATA ANALYSIS
1 DAY

Iso-Seq Express kit

- 60-300 ng total RNA
- Full-length cDNA
- Multiplexing support



	FL Reads	Unique Genes	Unique Transcripts
UHRR	4,734,362	16,328	183,689
Alzheimer Brain	4,277,293	17,670	162,290

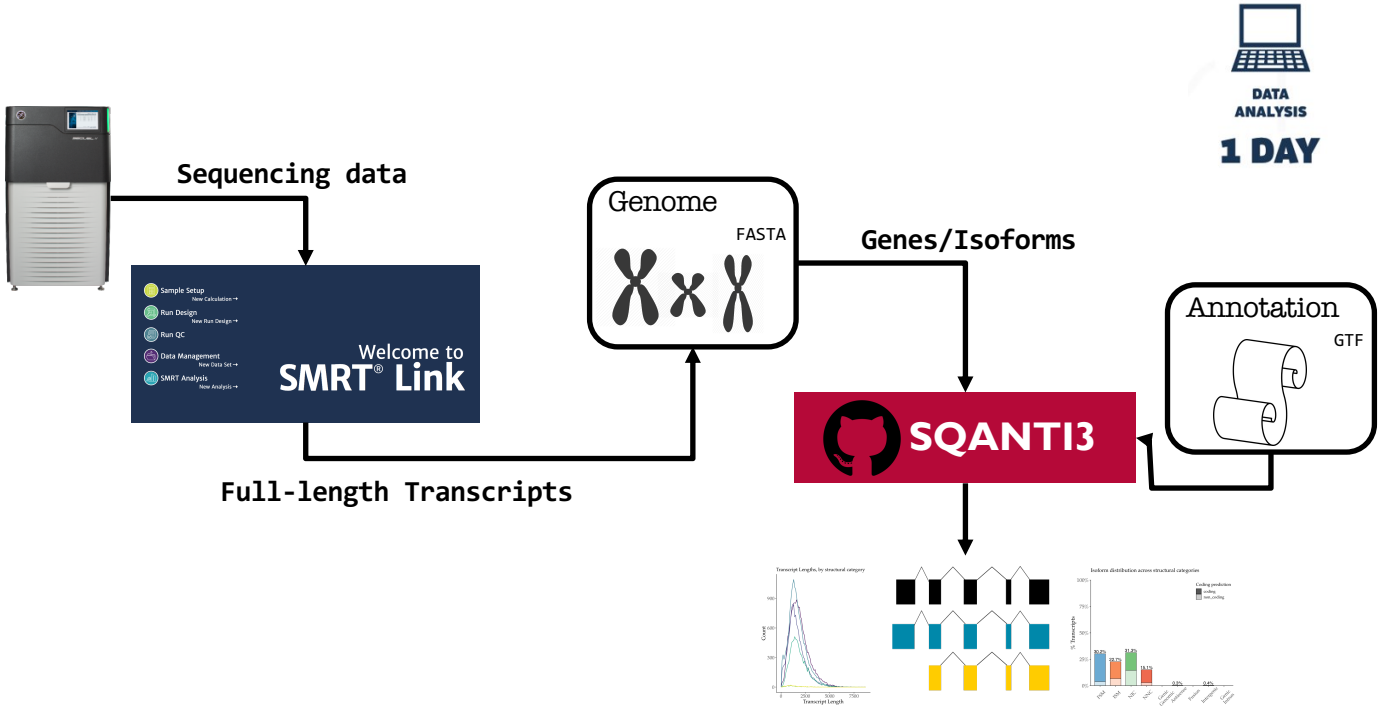
Dataset: UHRR

Dataset: Alzheimer Brain

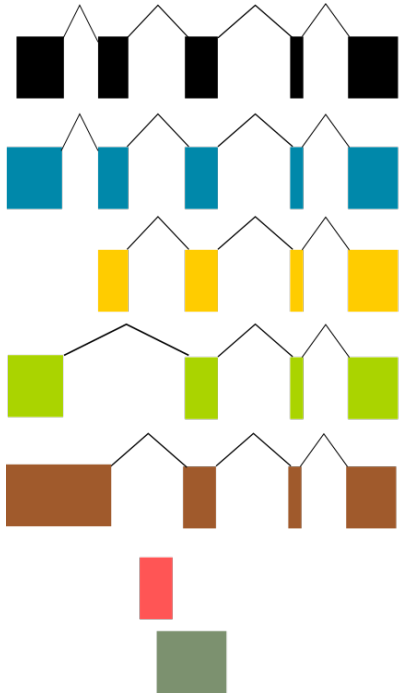
Sequel II System

- 1 SMRT Cell 8M for whole transcriptome
- Up to 4 million full-length reads

FULL BIOINFORMATICS SOLUTION



CLASSIFICATION OF TRANSCRIPTS



Reference

FSM Full Splice Match, matches reference perfectly.

ISM Incomplete Splice Matches, matches reference partially

NIC Novel In Catalog, novel isoform using known junctions

NNC Novel Not in Catalog, novel isoforms using novel junctions.

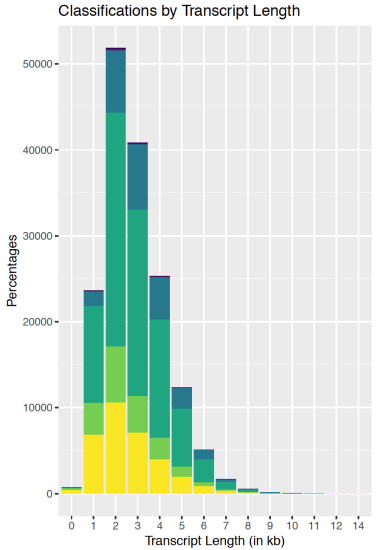
Genic Intron within intron

Genic Genomic Overlap with intron and exons

Tardaguila, M. *et al.* SQANTI: extensive characterization of long read transcript sequences for quality control in full-length transcriptome identification and quantification. 1–31 (2017). doi:10.1101/118083

ISO-SEQ METHOD ON THE SEQUEL II SYSTEM

- **Comprehensive**
- Full-Length
- Highly accurate



162,290 transcripts

80 – 14,288 bp
(mean: 3.3 kb)

Dataset: Alzheimer brain on 1 SMRT Cell 8M

ISO-SEQ METHOD ON THE SEQUEL II SYSTEM

- **Comprehensive**
- Full-Length
- Highly accurate

	Known	Novel	Total
Genes	17,051	619	**17,670**
Isoforms	51,660	110,630	**162,290**

Dataset: Alzheimer brain on 1 SMRT Cell 8M

ISO-SEQ METHOD ON THE SEQUEL II SYSTEM

- Comprehensive
- **Full-Length**
- Highly accurate

Category	Count	Description
FSM	32,649	Perfect match
ISM	19,011	Incomplete match
NIC	84,610	Novel isoform using known junctions
NNC	25,323	Novel isoform using at least novel junction
Antisense	321	Anti-sense to known gene
Intergenic	376	Intergenic

Dataset: Alzheimer brain on 1 SMRT Cell 8M

ISO-SEQ METHOD ON THE SEQUEL II SYSTEM

- Comprehensive
- **Full-Length**
- Highly accurate

Category	Count	CAGE peak within 50 bp
FSM	32,649	70%
ISM	19,011	37%
NIC	84,610	36%
NNC	25,323	57%
Antisense	321	24%
Intergenic	376	24%

Dataset: Alzheimer brain on 1 SMRT Cell 8M

ISO-SEQ METHOD ON THE SEQUEL II SYSTEM

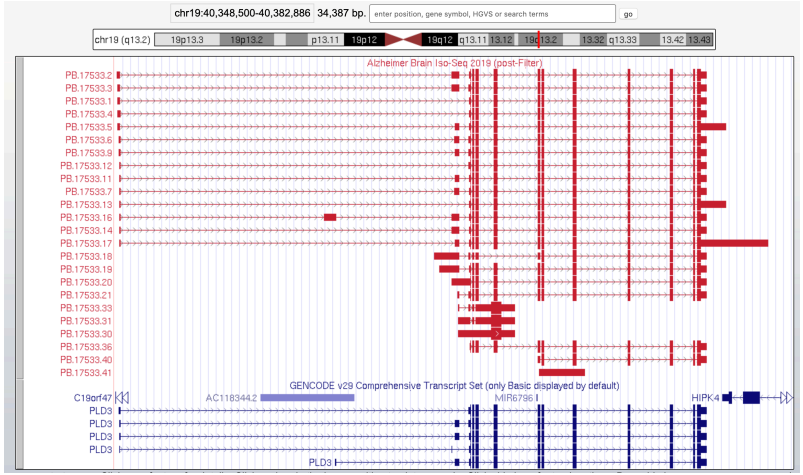
- Comprehensive
- **Full-Length**
- Highly accurate

Category	Count	CAGE peak within 50 bp	polyA Motif Detected
FSM	32,649	70%	72%
ISM	19,011	37%	62%
NIC	84,610	36%	55%
NNC	25,323	57%	72%
Antisense	321	24%	43%
Intergenic	376	24%	38%

Dataset: Alzheimer brain on 1 SMRT Cell 8M

ISO-SEQ METHOD ON THE SEQUEL II SYSTEM

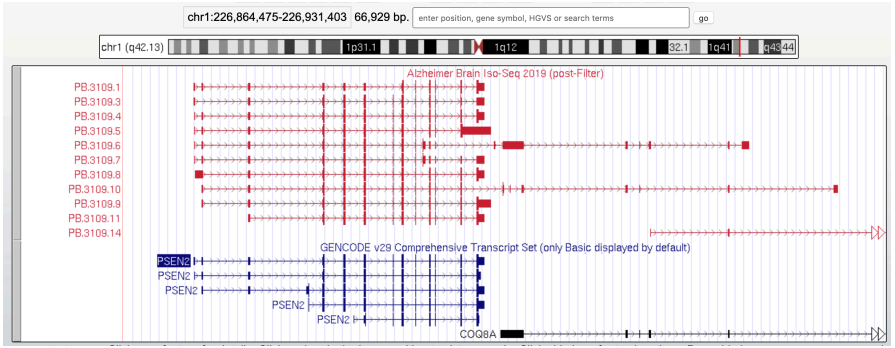
- Comprehensive
- **Full-Length**
- Highly accurate



Dataset: Alzheimer brain on 1 SMRT Cell 8M

ISO-SEQ METHOD ON THE SEQUEL II SYSTEM

- Comprehensive
- **Full-Length**
- Highly accurate

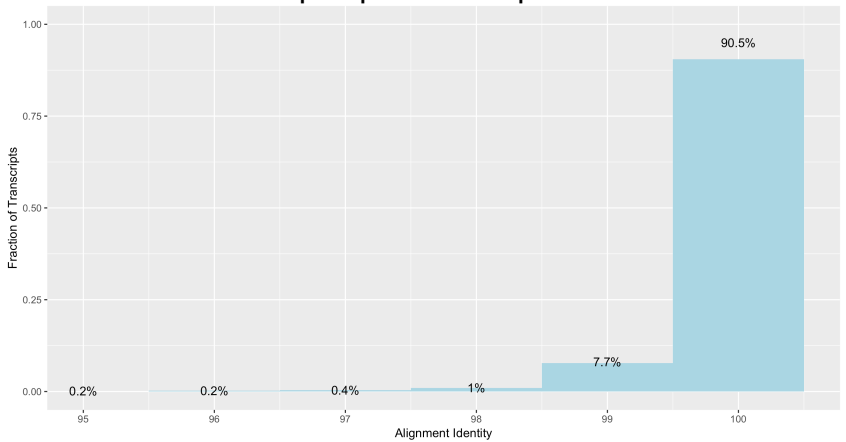


Dataset: Alzheimer brain on 1 SMRT Cell 8M

ISO-SEQ METHOD ON THE SEQUEL II SYSTEM

- Comprehensive
- Full-Length
- **Highly accurate**

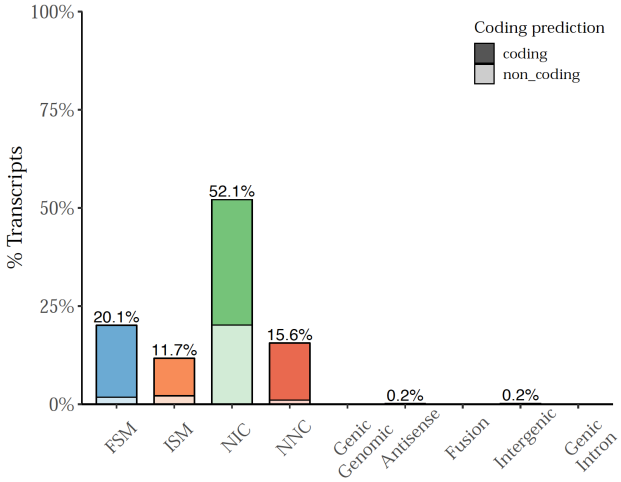
> 99% of Iso-Seq output transcript is >99% accurate



Dataset: Alzheimer brain on 1 SMRT Cell 8M

ISO-SEQ METHOD ON THE SEQUEL II SYSTEM

- Comprehensive
- Full-Length
- **Highly accurate**



Dataset: Alzheimer brain on 1 SMRT Cell 8M

Iso-Seq Express kit

- 60-300 ng total RNA
- Full-length cDNA
- Multiplexing support

Bioinformatics

- Iso-Seq3 in SMRT Analysis
- Reads to ORF in 1 day
- Downstream community tools

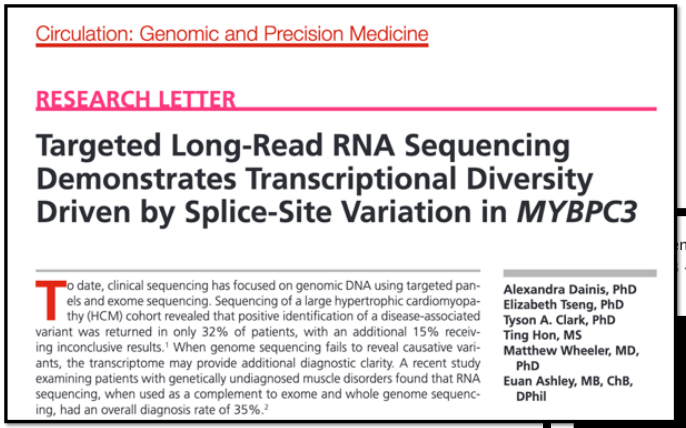
Sequel II System

- 1 SMRT Cell 8M for whole transcriptome
- 4 million full-length reads

PACBIO®

Research Highlights

GENETIC DIAGNOSIS FOR RARE DISEASES



Circulation: Genomic and Precision Medicine

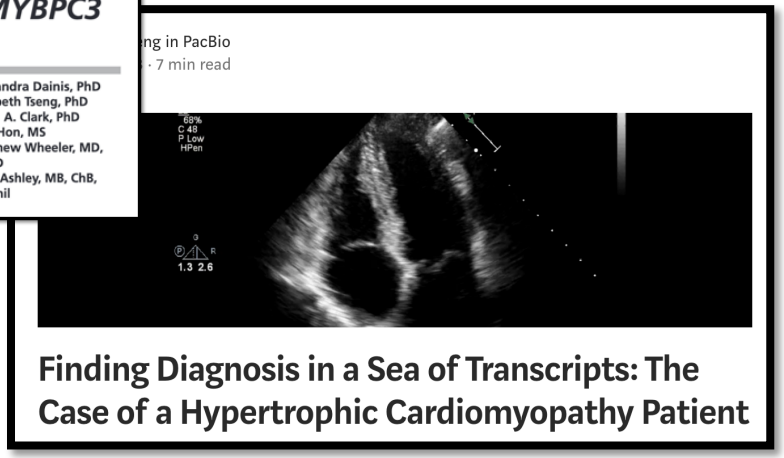
RESEARCH LETTER

Targeted Long-Read RNA Sequencing Demonstrates Transcriptional Diversity Driven by Splice-Site Variation in *MYBPC3*

To date, clinical sequencing has focused on genomic DNA using targeted panels and exome sequencing. Sequencing of a large hypertrophic cardiomyopathy (HCM) cohort revealed that positive identification of a disease-associated variant was returned in only 32% of patients, with an additional 15% receiving inconclusive results.[1] When genome sequencing fails to reveal causative variants, the transcriptome may provide additional diagnostic clarity. A recent study examining patients with genetically undiagnosed muscle disorders found that RNA sequencing, when used as a complement to exome and whole genome sequencing, had an overall diagnosis rate of 35%.[2]

Alexandra Dainis, PhD
Elizabeth Tseng, PhD
Tyson A. Clark, PhD
Ting Hon, MS
Matthew Wheeler, MD, PhD
Euan Ashley, MB, ChB, DPhil

https://medium.com/@magdoll



ng in PacBio

· 7 min read

Finding Diagnosis in a Sea of Transcripts: The Case of a Hypertrophic Cardiomyopathy Patient

Dainis et al., "Targeted Long-Read RNA Sequencing Demonstrates Transcriptional Diversity Driven by Splice-Site Variation in MYBPC3." *Circulation. Genomic and Precision Medicine* (2019)

GENETIC DIAGNOSIS FOR RARE DISEASES

- 21y female patient with severe hypertrophic cardiomyopathy (HCM)
- HCM panel identified single base (c.1898-1G>A) mutation in MYBPC3
- Mutation expected to affect splicing between exon 19-20

Circulation: Genomic and Precision Medicine

RESEARCH LETTER

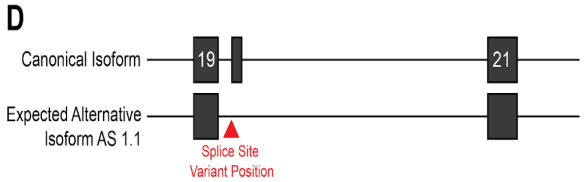
Targeted Long-Read RNA Sequencing Demonstrates Transcriptional Diversity Driven by Splice-Site Variation in *MYBPC3*

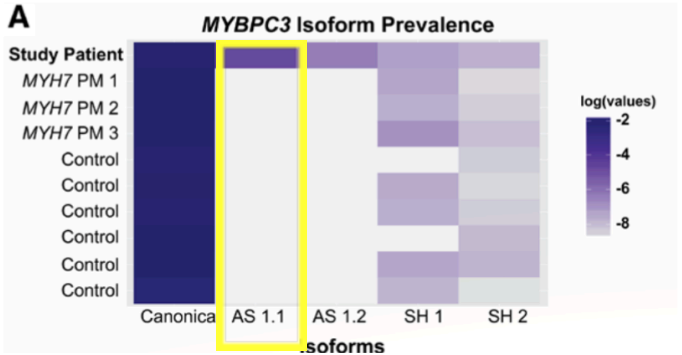
To date, clinical sequencing has focused on genomic DNA using targeted panels and exome sequencing. Sequencing of a large hypertrophic cardiomyopathy (HCM) cohort revealed that positive identification of a disease-associated variant was returned in only 32% of patients, with an additional 15% receiving inconclusive results.[1] When genome sequencing fails to reveal causative variants, the transcriptome may provide additional diagnostic clarity. A recent study examining patients with genetically undiagnosed muscle disorders found that RNA sequencing, when used as a complement to exome and whole genome sequencing, had an overall diagnosis rate of 35%.[2]

Alexandra Dainis, PhD
Elizabeth Tseng, PhD
Tyson A. Clark, PhD
Ting Hon, MS
Matthew Wheeler, MD, PhD
Euan Ashley, MB, ChB, DPhil

Dainis et al., "Targeted Long-Read RNA Sequencing Demonstrates Transcriptional Diversity Driven by Splice-Site Variation in MYBPC3." *Circulation. Genomic and Precision Medicine* (2019)

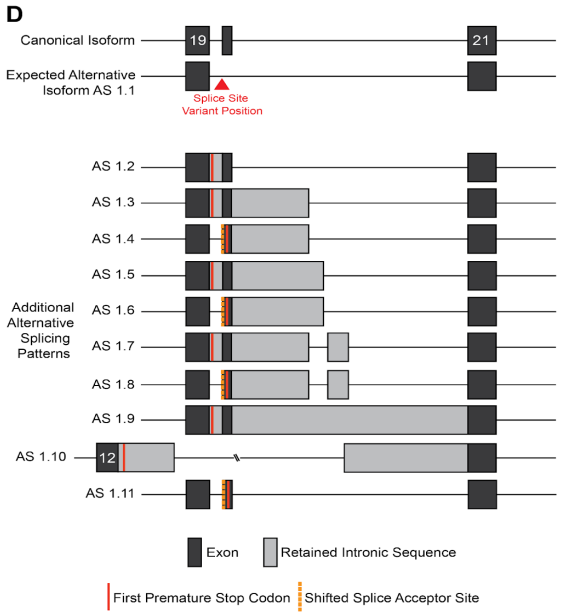
GENETIC DIAGNOSIS FOR RARE DISEASES

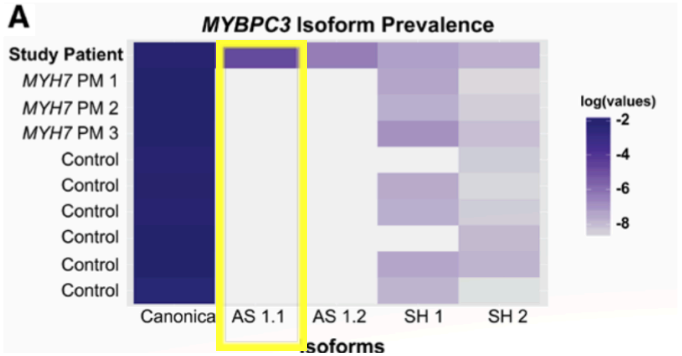




Dainis et al., "Targeted Long-Read RNA Sequencing Demonstrates Transcriptional Diversity Driven by Splice-Site Variation in MYBPC3." *Circulation. Genomic and Precision Medicine* (2019)

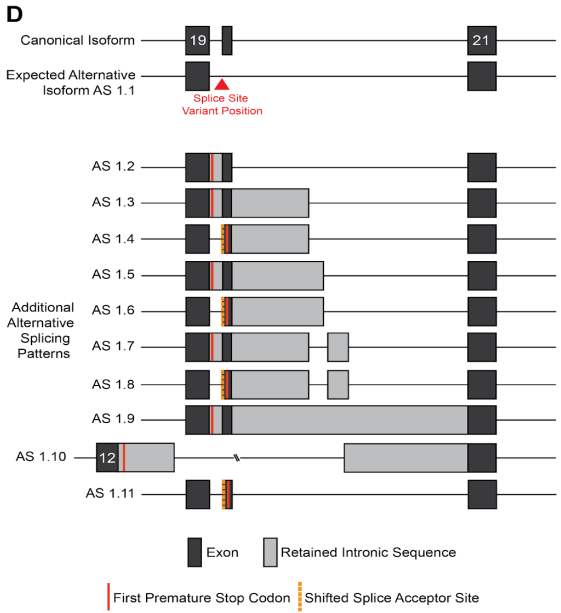
GENETIC DIAGNOSIS FOR RARE DISEASES

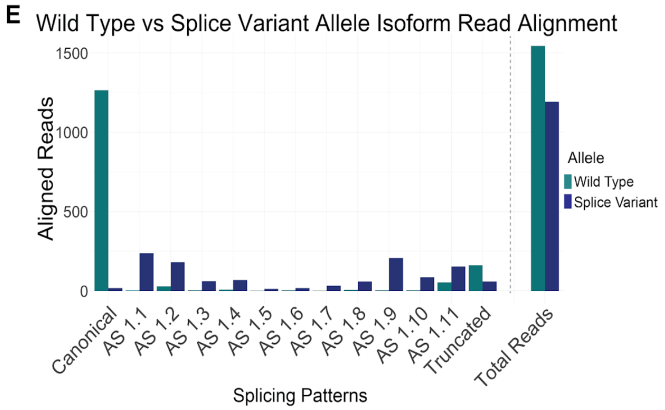




Dainis et al., "Targeted Long-Read RNA Sequencing Demonstrates Transcriptional Diversity Driven by Splice-Site Variation in MYBPC3." *Circulation. Genomic and Precision Medicine* (2019)

GENETIC DIAGNOSIS FOR RARE DISEASES





Dainis et al., "Targeted Long-Read RNA Sequencing Demonstrates Transcriptional Diversity Driven by Splice-Site Variation in MYBPC3." *Circulation. Genomic and Precision Medicine* (2019)

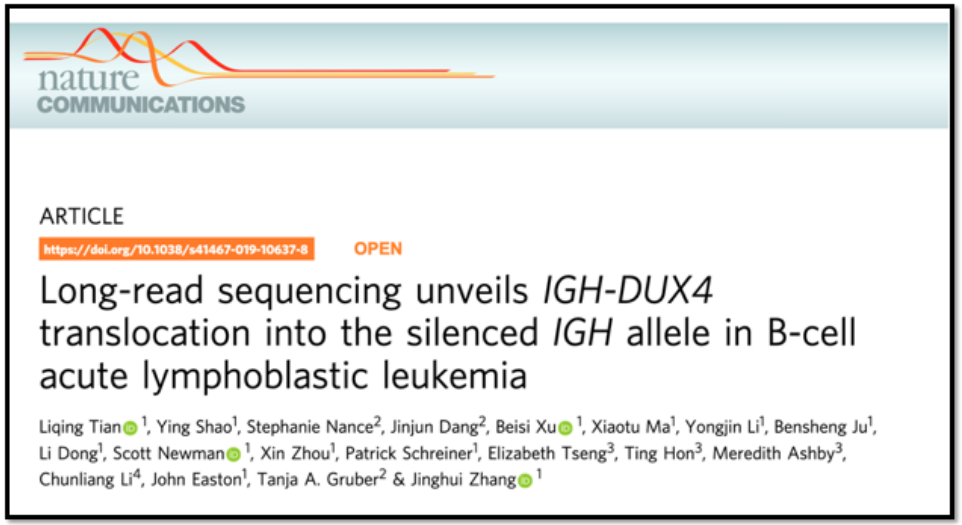
GENETIC DIAGNOSIS FOR RARE DISEASES

- 21y female patient with severe hypertrophic cardiomyopathy (HCM)
- HCM panel identified single base (c.1898-1G>A) mutation in MYBPC3
- Mutation expected to affect splicing between exon 19-20

- Iso-Seq method identified patient-specific alternative splice variants (AS 1.1-1.11) linked to the c.1898-1G>A mutation that are not expressed on the WT allele

Dainis et al., “Targeted Long-Read RNA Sequencing Demonstrates Transcriptional Diversity Driven by Splice-Site Variation in MYBPC3.” *Circulation. Genomic and Precision Medicine* (2019)

FUSION TRANSCRIPTS IN CANCER



nature COMMUNICATIONS

ARTICLE

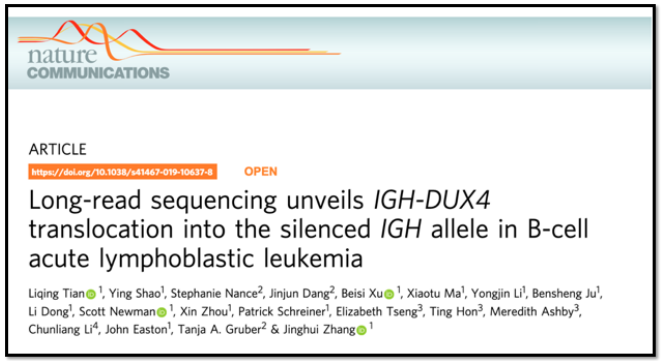
https://doi.org/10.1038/s41467-019-10637-8 OPEN

Long-read sequencing unveils *IGH-DUX4* translocation into the silenced *IGH* allele in B-cell acute lymphoblastic leukemia

Liqing Tian[1], Ying Shao[1], Stephanie Nance[2], Jinjun Dang[2], Beisi Xu[1], Xiaotu Ma[1], Yongjin Li[1], Bensheng Ju[1], Li Dong[1], Scott Newman[1], Xin Zhou[1], Patrick Schreiner[1], Elizabeth Tseng[3], Ting Hon[3], Meredith Ashby[3], Chunliang Li[4], John Easton[1], Tanja A. Gruber[2] & Jinghui Zhang[1]

Tian et al., "Long-read sequencing unveils IGH-DUX4 translocation into the silenced IGH allele in B-cell acute lymphoblastic leukemia", *Nat Comm* (2019)

FUSION TRANSCRIPTS IN CANCER



nature COMMUNICATIONS

ARTICLE

https://doi.org/10.1038/s41467-019-10637-8 OPEN

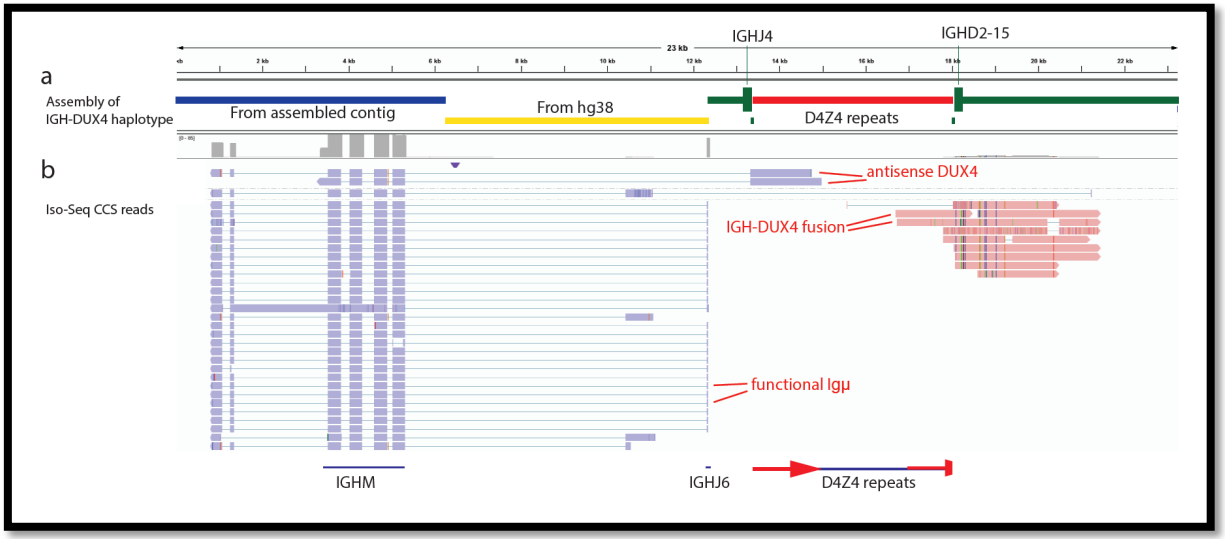
Long-read sequencing unveils *IGH-DUX4* translocation into the silenced *IGH* allele in B-cell acute lymphoblastic leukemia

Liqing Tian[1], Ying Shao[1], Stephanie Nance[2], Jinjun Dang[2], Beisi Xu[1], Xiaotu Ma[1], Yongjin Li[1], Bensheng Ju[1], Li Dong[1], Scott Newman[1], Xin Zhou[1], Patrick Schreiner[1], Elizabeth Tseng[3], Ting Hon[3], Meredith Ashby[3], Chunliang Li[4], John Easton[1], Tanja A. Gruber[2] & Jinghui Zhang[1]

- In B-cell development, one IGH allele is silenced
- IGH-DUX4 translocation is subtype of B-cell acute lymphoblastic leukemia (B-ALL)
- Accounts for ~7% of pediatric B-ALL
- DUX4 is a transcription factor located in a GC-rich region
- Short reads difficult to map accurately IGH-DUX4 translocation

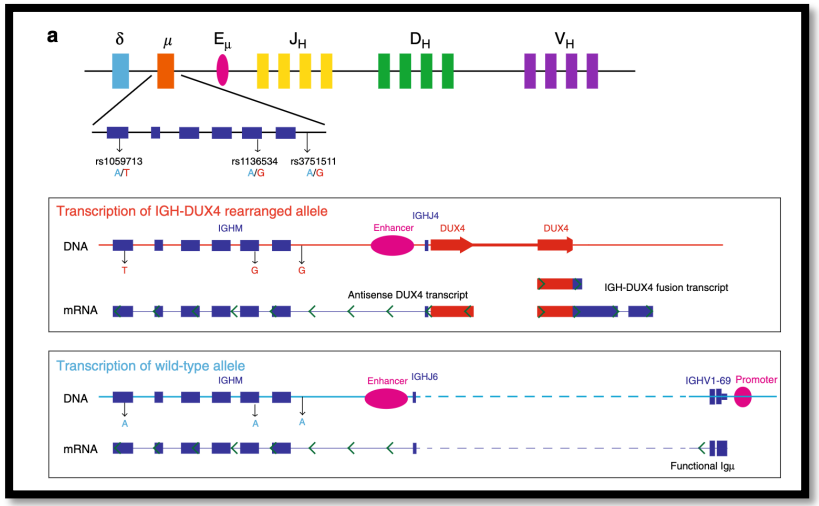
Tian et al., "Long-read sequencing unveils IGH-DUX4 translocation into the silenced IGH allele in B-cell acute lymphoblastic leukemia", *Nat Comm* (2019)

ISO-SEQ SEQUENCING OF NALM6 B-ALL CELL LINE



Tian et al., "Long-read sequencing unveils IGH-DUX4 translocation into the silenced IGH allele in B-cell acute lymphoblastic leukemia", *Nat Comm* (2019)

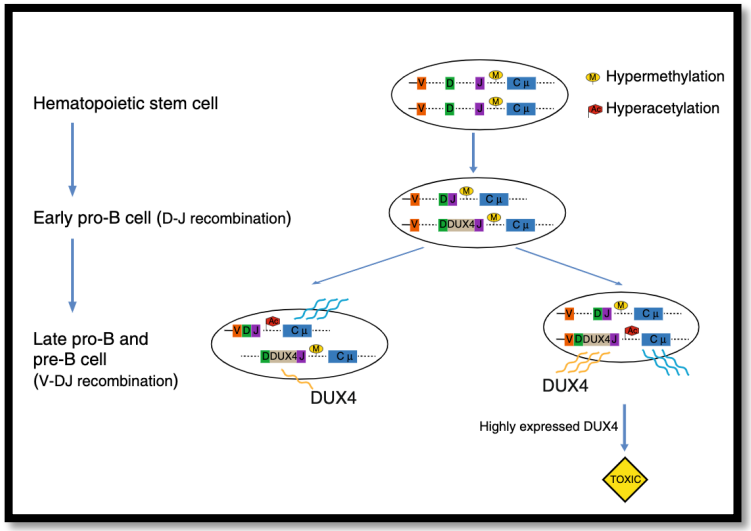
FUSION TRANSCRIPTS IN CANCER



- IGHM locus is diploid with three SNPs (A-A-A vs T-G-G)
- Iso-Seq analysis shows functional Igμ comes exclusively from A-A-A wildtype
- Iso-Seq analysis shows antisense DUX4 comes exclusively from T-G-G fusion allele --> proves that IGH-DUX4 translation comes from the silenced IGH allele

Tian et al., "Long-read sequencing unveils IGH-DUX4 translocation into the silenced IGH allele in B-cell acute lymphoblastic leukemia", *Nat Comm* (2019)

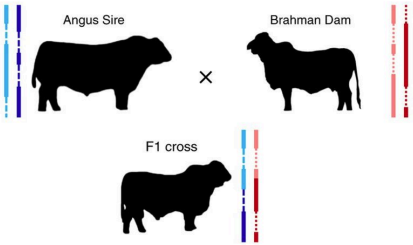
FUSION TRANSCRIPTS IN CANCER



- Activation of the IGH-DUX4 is toxic (due to overexpression of DUX4)
- Therefore, only the wildtype IGH allele is viable

Tian et al., "Long-read sequencing unveils IGH-DUX4 translocation into the silenced IGH allele in B-cell acute lymphoblastic leukemia", *Nat Comm* (2019)

ISO-SEQ METHOD FOR PLANT & ANIMAL



Brahman x Angus F1 Cattle

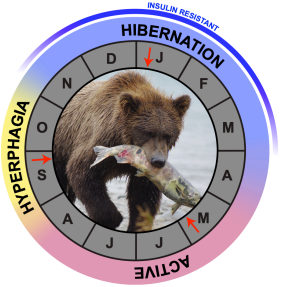
- Allele-specific isoform expression
- Tissue-specific isoform expression

Low et al (2020)

Cannabis

- Tissue-specific transcripts associated w/ THC & CBD synthesis
  - Chr Y gene annotation

McKernan, webinar



Grizzly Bear

- Tissue-specific alternative splicing
- Hibernation vs active state

Trojahn, Kelley, et al. (WSU)

LEARN MORE ABOUT AND THE ISO-SEQ METHOD

pacb.com/applications/rna-sequencing/

tinyurl.com/isoseq-google

@PacBio @Magdoll

github.com/PacificBiosciences
github.com/Magdoll

PACBIO®

www.pacb.com