PACIFIC BIOSCIENCES®

# No Assembly Required – Making the Most of Iso-Seq Data

Kristin Mars, Sr. Scientist, FAS, PacBio

For Research Use Only. Not for use in diagnostic procedures. © Copyright 2019 by Pacific Biosciences of California, Inc. All rights reserved.

- What is Iso-Seq
- Who is using Iso-Seq?
- What can you do with one SMRT Cell?
- Where can I learn more?

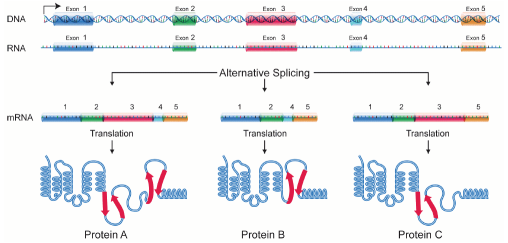
# What is Iso-Seq?

# ISO-SEQ: FULL-LENGTH RNA SEQUENCING

**RNA SEQUENCING**

## Iso-Seq is…

- Full-Length cDNA sequencing – no assembly required
- Targeted or whole transcriptome



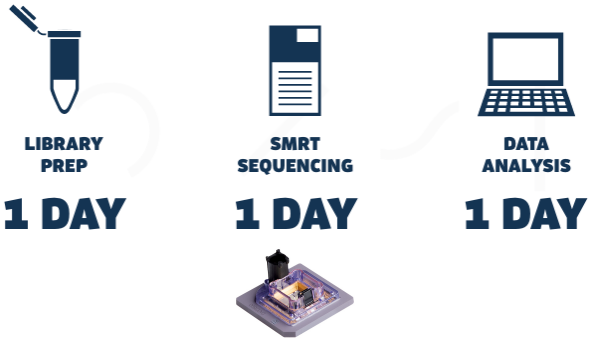

## Iso-Seq can…

- Discover novel genes and isoforms
- Improve genome annotation, with or without reference genome
- Increase the accuracy of RNA-seq quantification at isoform-level resolution

**ANNOTATION**

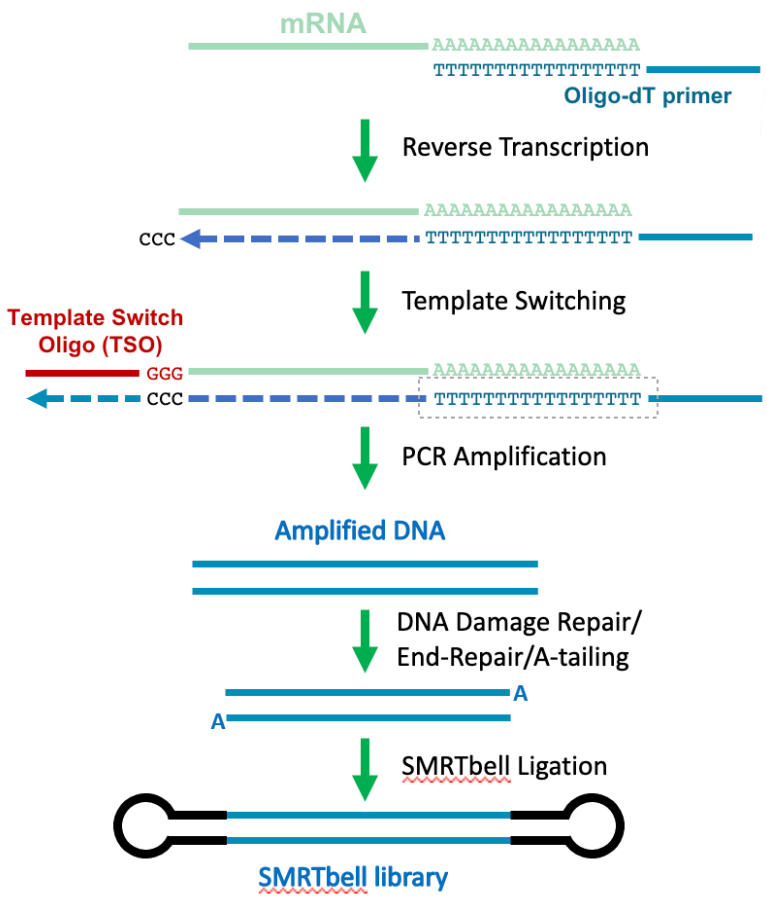
## You can do Iso-Seq with…

- 1-Day Library Prep
- 1 SMRT Cell 8M
- Full bioinformatics solution



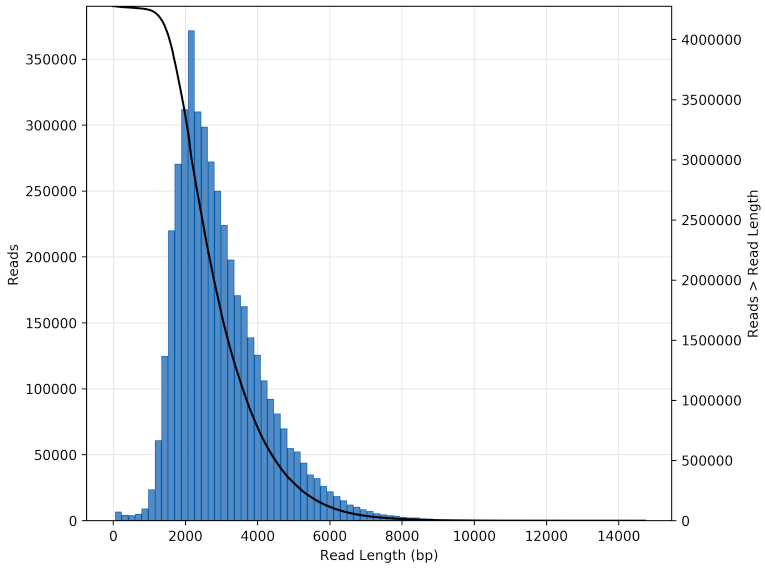
# FROM RNA TO LIBRARY IN A SINGLE DAY

## Iso-Seq Express Workflow




- **Improved formulation** for reduced input requirement at 60-300 ng
- **Significantly faster workflow** from RNA to SMRTbell library in one day.
- **Minimal handling-induced cDNA damage**
- **Capture full-length transcriptomes** in a single SMRT Cell
- **Supports multiplexing up to 12 samples**
- **No size-selection required**

# ALZHEIMER SAMPLE SEQUENCING RESULTS ON SEQUEL II SYSTEM




**CCS Read Classification**

| Value | Analysis Metric |
|---|---|
| 5,052,827 | Reads |
| 4,327,705 | Reads with 5' and 3' Primers |
| 4,292,971 | Non-Concatamer Reads with 5' and 3' Primers |
| 4,277,293 | Non-Concatamer Reads with 5' and 3' Primers and Poly-A Tail |
| 2,973 | Mean Length of Full-Length Non-Concatamer Reads |
| 1 | Unique Primers |
| 4,327,705 | Mean Reads per Primer |
| 4,327,705 | Max. Reads per Primer |
| 4,327,705 | Min. Reads per Primer |
| 725,122 | Reads without Primers |

Approx. 4 M full-length non-concatemer reads (85%)

# Who is using Iso-Seq?

# SELECTED ISO-SEQ PUBLICATIONS

HUMAN BIOMEDICAL RESEARCH

PLANT AND ANIMAL SCIENCES


- Armitage, S. A. O. et al. (2014) Quantitative profiling of Drosophila melanogaster Dscam1 isoforms reveals no changes in splicing after bacterial exposure. *PLoS One*. 9 (10), e108660.
- Brinzevich, D. et al. (2014) HIV-1 interacts with human endogenous retrovirus K (HML-2) envelopes derived from human primary lymphocytes. *Journal of Virology*. 88 (11), 6213–6223.
- Cavelier, L. et al. (2015) Clonal distribution of BCR-ABL1 mutations and splice isoforms by single-molecule long-read RNA sequencing. *BMC Cancer*. 15, 45.
- Chen, J.-Y. et al. (2015) Emergence, retention and selection: A trilogy of origination for functional de novo proteins from ancestral lncRNAs in primates. *PLoS Genetics*. 11 (7), e1005391.
- Chen, L. et al. (2014) Transcriptional diversity during lineage commitment of human blood progenitors. *Science* . 345 (6204), 1251033.
- Dasari, S. et al. (2015) Proteomic detection of immunoglobulin light chain variable region peptides from amyloidosis patient biopsies. *J. Proteome Res.* 14 (4), 1957–1967.
- Davidson, N. M. et al. (2015) JAFFA: High sensitivity transcriptome-focused fusion gene detection. *Genome Medicine*. 7 (1), 43.
- Dong, L. et al. (2015) Single-molecule real-time transcript sequencing facilitates common wheat genome annotation and grain transcriptome research. *BMC Genomics*. 16 (1), 1039.
- Gao, Q. et al. (2015) Predominant contribution of cis-regulatory divergence in the evolution of mouse alternative splicing. *Molecular Systems Biology*. 11 (7), 816.
- Gordon, S. P. et al. (2015) Widespread polycistronic transcripts in fungi revealed by single-molecule mRNA sequencing. *PLoS One*. 10 (7), e0132628.
- Gross, S. M. et al. (2013) De novo transcriptome assembly of drought tolerant CAM plants, Agave deserti and Agave tequilana. *BMC Genomics*. 14, 563.
- Ichihashi, N. et al. (2015) Periodic pattern of genetic and fitness diversity during evolution of an artificial cell-like system. *Molecular Biology and Evolution*. 32 (12), 3205–3214.
- Kleinman, C. L. et al. (2014) Fusion of TTYH1 with the C19MC microRNA cluster drives expression of a brain-specific DNMT3B isoform in the embryonal brain tumor ETMR. *Nature Genetics*. 46 (1), 39–44.
- Larsen, P. A. et al. (2014) Next-generation approaches to advancing eco-immunogenomic research in critically endangered primates. *Molecular Ecology Resources*. 14 (6), 1198–1209.
- Larsen, P. A. & Smith, T. P. L. (2012) Application of circular consensus sequencing and network analysis to characterize the bovine IgG repertoire. *BMC Immunology*. 13, 52.
- Li, A. et al. (2014) PLEK: a tool for predicting long non-coding RNAs and messenger RNAs based on an improved k-mer scheme. *BMC Bioinformatics*. 15, 311.
- Lu, T. et al. (2015) Transcriptome-wide investigation of circular RNAs in rice. *RNA*. 21 (12), 2076–2087.
- Macaulay, I. C. et al. (2015) G&T-seq: parallel sequencing of single-cell genomes and transcriptomes. *Nat. Methods*. 12 (6), 519–522.
- Minoche, A. E. et al. (2015) Exploiting single-molecule transcript sequencing for eukaryotic gene prediction. *Genome Biology*. 16, 184.
- Ocwieja, K. E. et al. (2012) Dynamic regulation of HIV-1 mRNA populations analyzed by single-molecule enrichment and long-read sequencing. *Nucleic Acids Res.* 40 (20), 10345–10355.
- Pretto, D. I. et al. (2015) Differential increases of specific FMR1 mRNA isoforms in premutation carriers. *Journal of Medical Genetics*. 52 (1), 42–52.
- Salton, M. et al. (2014) Identification by high-throughput imaging of the histone methyltransferase EHMT2 as an epigenetic regulator of VEGFA alternative splicing. *Nucleic Acids Res.* 42 (22), 13662–13673.
- Schreiner, D. et al. (2014) Targeted combinatorial alternative splicing generates brain region-specific repertoires of neurexins. *Neuron*. 84 (2), 386–398.
- Simpson, L. et al. (2015) Comparison of the mitochondrial genomes and steady state transcriptomes of two strains of the trypanosomatid parasite, Leishmania tarentolae. *PLoS Negl. Trop. Dis.* 9 (7), e0003841.
- Singh, N. et al. (2016) IsoSeq analysis and functional annotation of the infratentorial ependymoma tumor tissue on PacBio RSII platform. *Meta Gene*. 7, 70-75.
- Tevz, G. et al. (2016) Identification of a novel fusion transcript between human relaxin-1 (RLN1) and human relaxin-2 (RLN2) in prostate cancer. *Molecular and Cellular Endocrinology*. 420, 159–168.
- Thomas, S. et al. (2014) Long-read sequencing of chicken transcripts and identification of new transcript isoforms. *PLoS One*. 9 (4), e94650.
- Tombácz, D. et al. (2015) Characterization of novel transcripts in pseudorabies virus. *Viruses*. 7 (5), 2727–2744.
- Treutlein, B. et al. (2014) Cartography of neurexin alternative splicing mapped by single-molecule long-read mRNA sequencing. *Proceedings of the National Academy of Sciences*. 111 (13), E1291–E1299.
- Verma, A. et al. (2015) Transcriptome sequencing reveals thousands of novel long non-coding RNAs in B cell lymphoma. *Genome Medicine*. 7 (1), 110.
- Voit, R. A. et al. (2014) Nuclease-mediated gene editing by homologous recombination of the human globin locus. *Nucleic Acids Res.* 42 (2), 1365–1378.
- Vollmers, C. et al. (2015) Novel exons and splice variants in the human antibody heavy chain identified by single cell and single molecule sequencing. *PLoS One*. 10 (1), e0117050.
- Westbrook, C. J. et al. (2015) No assembly required: Full-length MHC class I allele discovery by PacBio circular consensus sequencing. *Human Immunology*. 76(12), 891-896.
- Windmüller, N. et al. (2015) Transcriptional adaptations during long-term persistence of Staphylococcus aureus in the airways of a cystic fibrosis patient. *International Journal of Medical Microbiology*. 305 (1), 38–46.
- Xu, Z. et al. (2015) Full-length transcriptome sequences and splice variants obtained by a combination of sequencing platforms applied to different root tissues of Salvia miltiorrhiza and tanshinone biosynthesis. *The Plant Journal*. 82 (6), 951–961.
- You, X. et al. (2015) Neural circular RNAs are derived from synaptic genes and regulated by development and plasticity. *Nature Neuroscience*. 18 (4), 603–610.
- Zhang, W. et al. (2014) PacBio sequencing of gene families - a case study with wheat gluten genes. *Gene*. 533 (2), 541–546.

Circulation: Genomic and Precision Medicine

RESEARCH LETTER

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

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

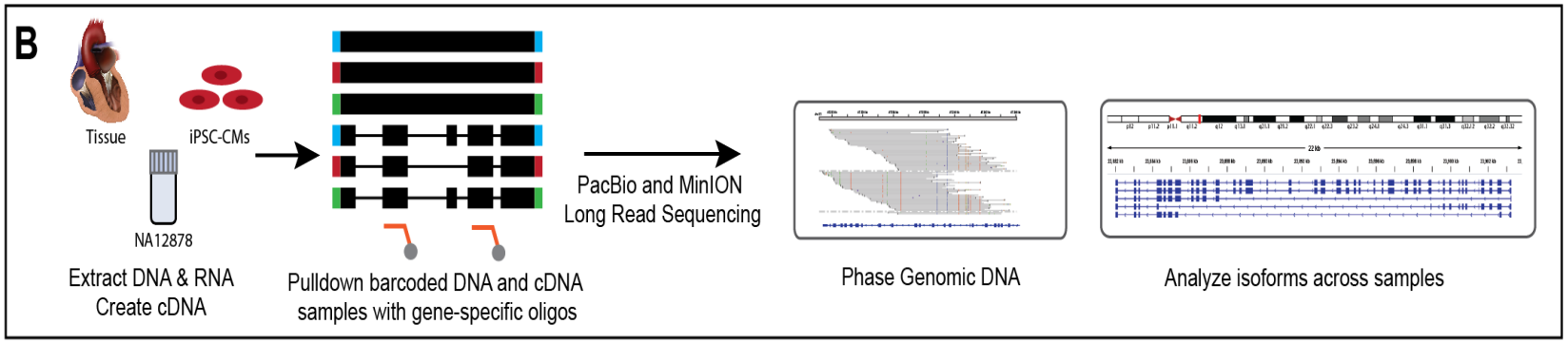
To date, clinical sequencing has focused on genomic DNA using targeted panels and exome sequencing. Sequencing of a large hypertrophic cardiomyopathy (HCM) cohort revealed that positive identification of a disease-associated variant was returned in only 32% of patients, with an additional 15% receiving 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]

# EXPERIMENTAL DESIGN

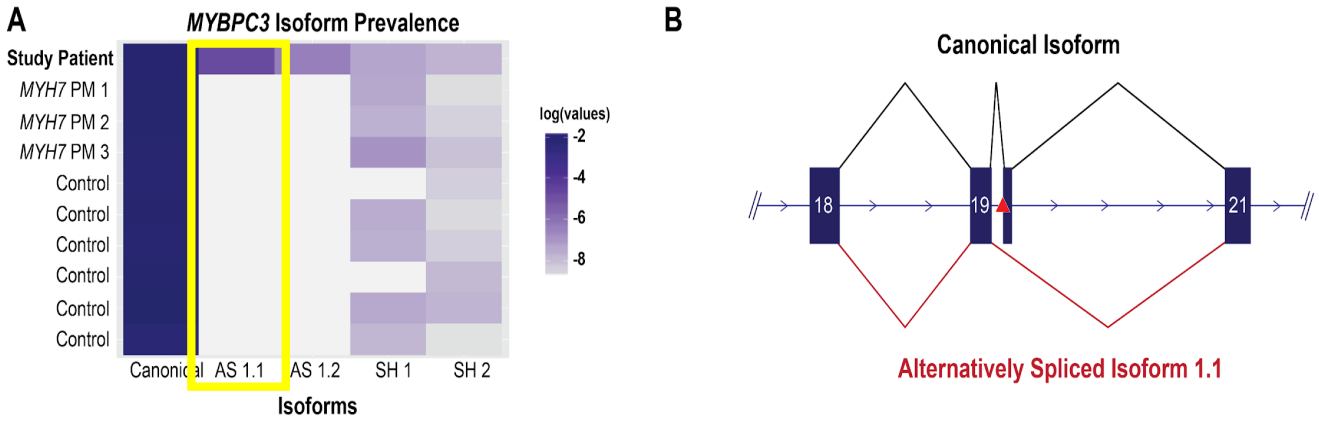
- 21y female patient with severe hypertrophic cardiomyopathy (HCM)
- HCM panel identified single base (c.1898-1G>A) mutation in *MYBPC3*
- Mutation expected to affect splicing between exon 19-20

**Sequencing:**

- Targeted Iso-Seq (probe-based) from control & HCM heart samples

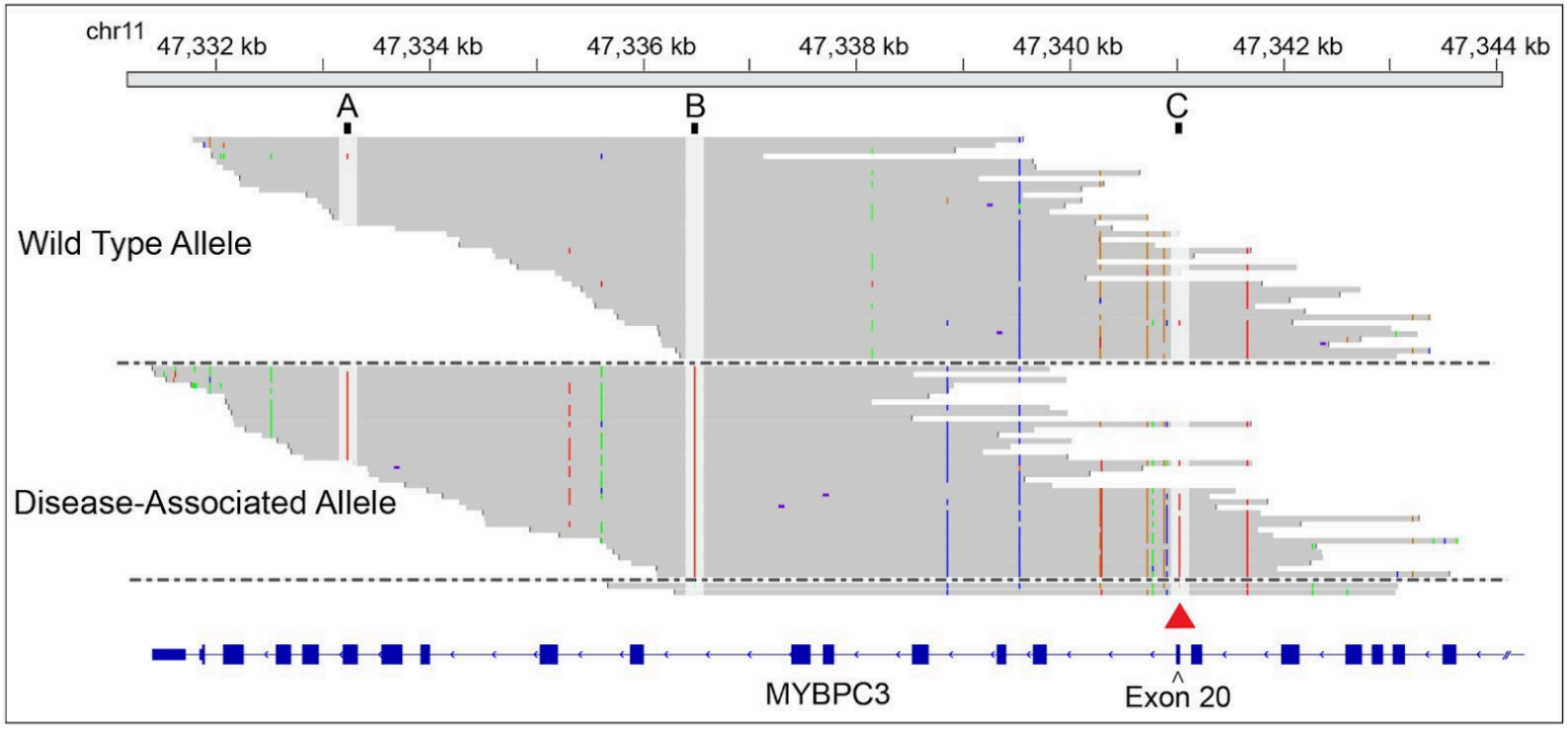


# HCM PATIENT SHOWS ALT ISOFORM THAT SKIPS EXON 20



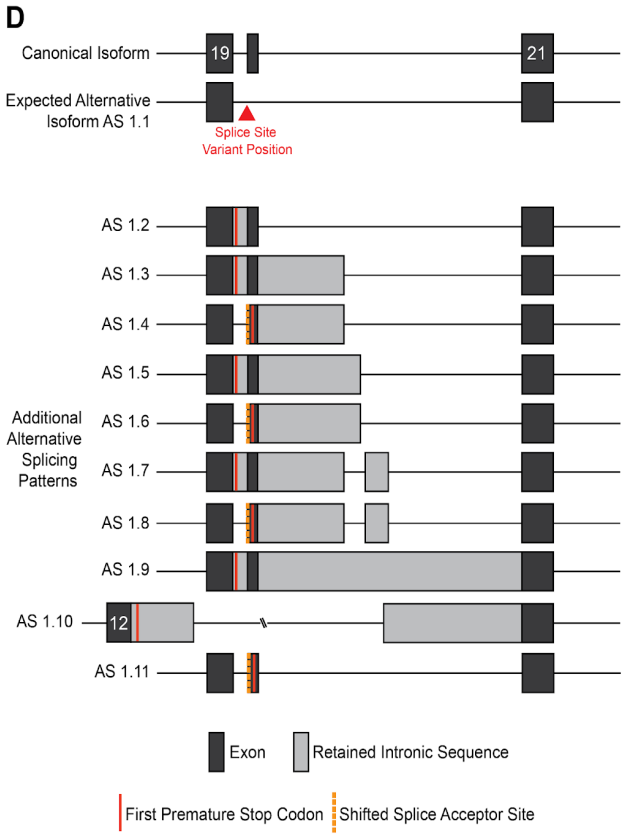

- Study patient dominant isoform PB.1.1 skips exon 20
  - Retains coding frame

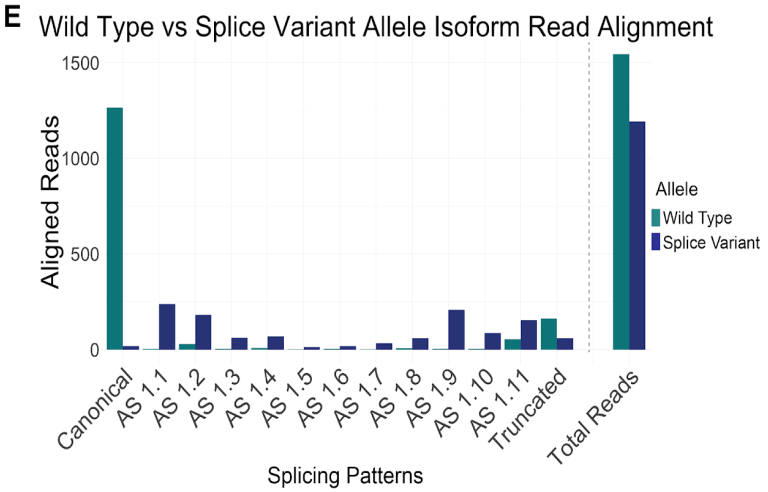
# TARGETED gDNA + cDNA USING LONG READS




- Using targeted enrichment with long read sequencing on both gDNA and cDNA, the author was able to haplotype the different MYBPC3 isoforms

# MYBPC3 MUTATION CAUSES SPLICING HAVOC




- For the 21y HCM patient, the mutation caused many alternative splicing events for the MYBPC3 gene (AS.1.1 through AS 1.11) that is seen only on the mutated allele

- **The confirms the genetic mutation to be the disease-causing mechanism**

# CONCLUSION

- Targeted Iso-Seq revealed complex alternative splicing events for the MYBPC3 gene in a HCM patient
- Probe-based enrichment on gDNA and cDNA data allowed allelic-specific isoform expression analysis
- Confirmed for a 21y HCM patient that a point mutation (c.1898-1G>A) resulted in aberrant splicing that included exon 20-skipping event and isoforms with intron retention seen only from the mutated allele
- The patient had no family data and no existing literature to support the mutation. The long-read data provided the only genetic diagnosis. Given the mutation is dominant, there's a 50% chance of passing on the mutation.

# WHOLE TRANSCRIPTOME ISO-SEQ ANALYSIS FOR CANCER FUSION

nature COMMUNICATIONS

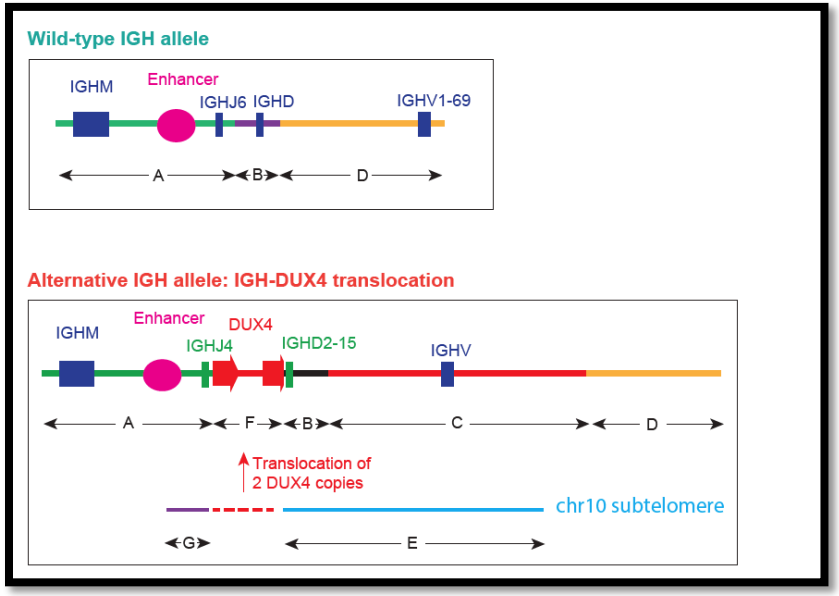
ARTICLE

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]

# ISO-SEQ ANALYSIS OF CANCER FUSION IGH-DUX4



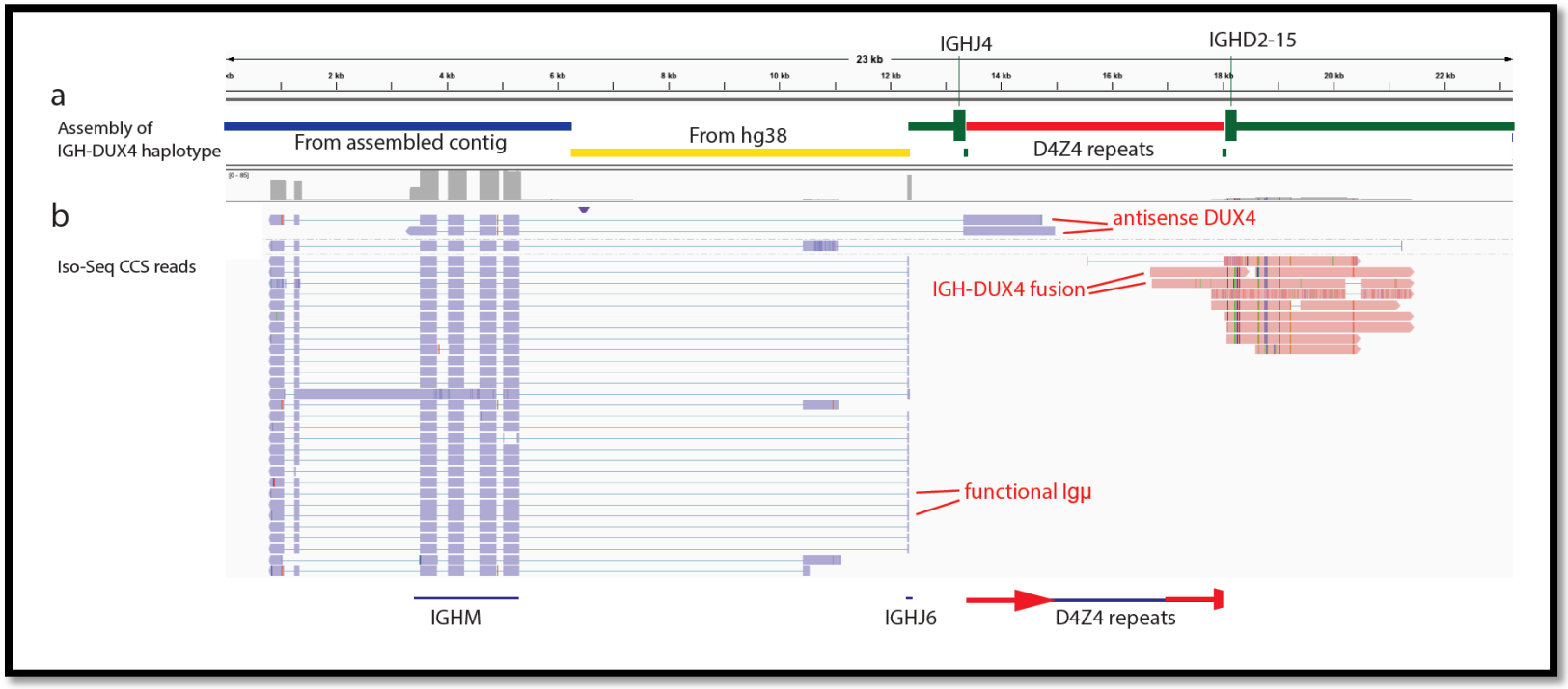

- In B-cell development, one IGH allele is silenced
- IGH-DUX4 translocation is subtype of B-cell acute lymphoblastic leukemia (B-ALL)
- Accounts for ~7% of pediatric B-ALL
- DUX4 is a transcription factor located in a GC-rich region
- Short reads difficult to map accurately IGH-DUX4 translocation

**Study Design:**

- Nalm6 B-ALL cell line known to harbor IGH-DUX4
- Whole transcriptome Iso-Seq on Sequel 1M – 2 SMRT Cells sequenced

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

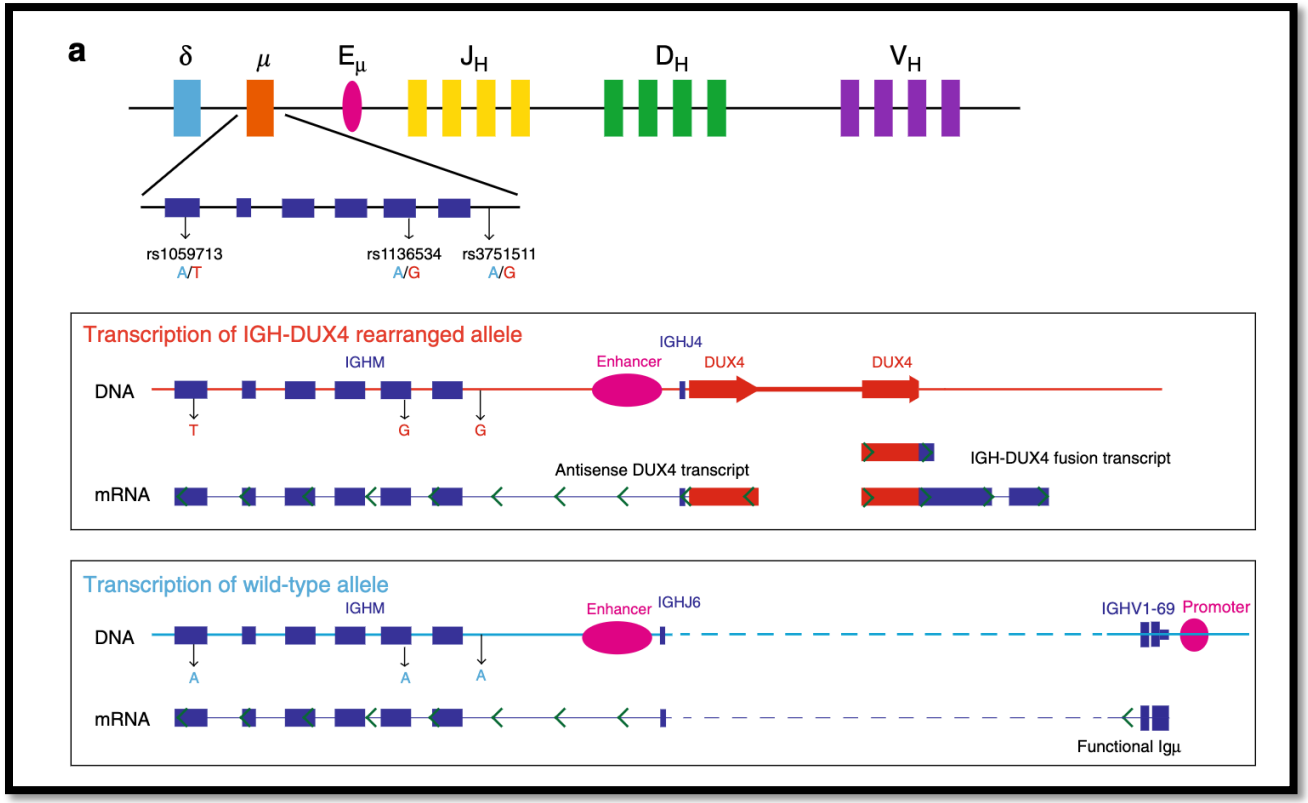
# ISO-SEQ ANALYSIS REVEALS FULL-LENGTH IGH-DUX4 TRANSLOCATION




- Mapped Iso-Seq reads reveal
  - functional Igμ
  - IGH-DUX4 fusion
  - antisense DUX4

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

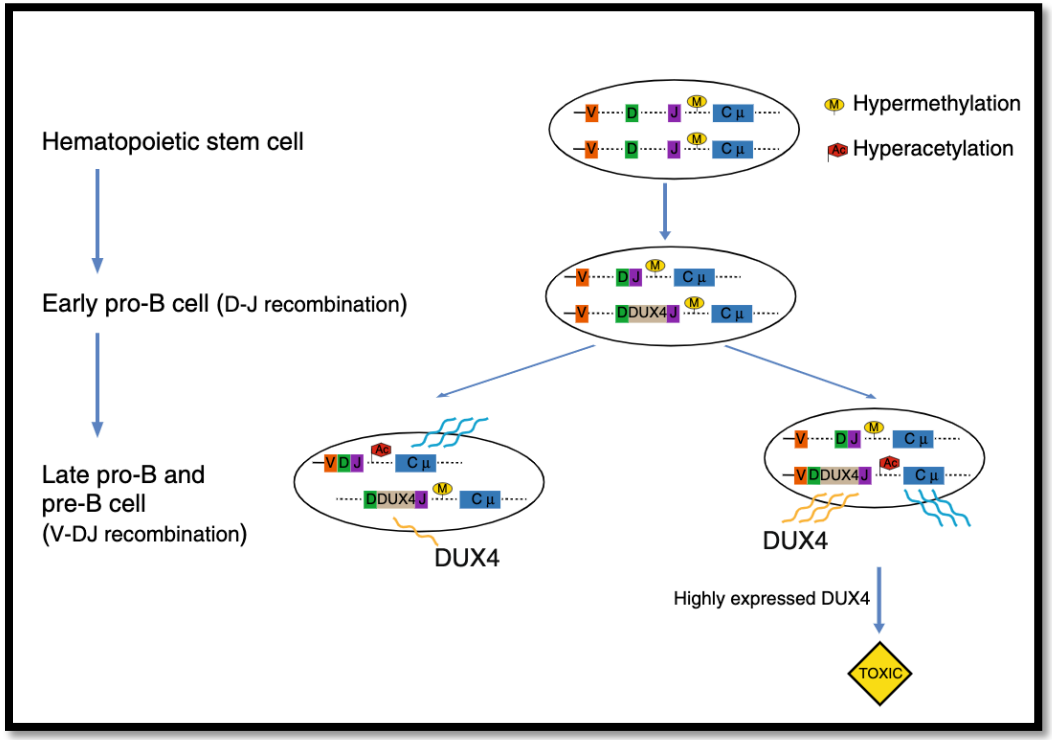
# ALLELIC IMBALANCE IN IGH-DUX4 TRANSLOCATION




- 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)

# ISO-SEQ ANALYSIS HELPS FORM HYPOTHESIS FOR IGH-DUX4 TRANSLOCATION EVENT DURING B-CELL DEVELOPMENT




Hypothesis formed by combining WGS, ATAC-Seq, Hi-C, RNA-seq, and Iso-Seq analysis:

- During early pro-B cell development, IGH-DUX4 breakpoint occurs
- In later stage, if IGH-DUX4 translocation allele is activated, DUX4 is overexpressed and leads to apoptosis
- Therefore, only wildtype IGH 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)

# What can you do with one SMRT Cell?

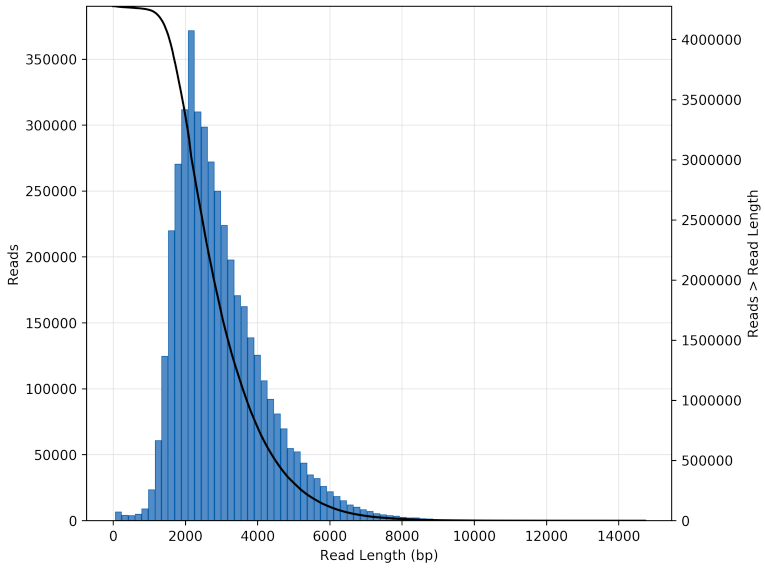
# WHAT CAN YOU DO WITH ONE SMRT CELL 8M?

**RNA SEQUENCING**

| SMRT Sequencing Applications | Number of SMRT Cells 8M* | PacBio Consumable Estimated Costs (US List Price)† |
|---|---|---|
| **Whole Transcriptome**<br>Characterize alternative splicing with full-length transcripts | **1** | $1300/sample |
| **Genome Annotation**<br>Sequence full-length transcripts and multiplex up to 8 tissues | **1** | $185/tissue |

*Study design, sample type, and level of multiplexing may affect the number of SMRT Cells 8M required. †All prices are listed in USD and cost may vary by region. Pricing includes library and sequencing reagents run on your Sequel II System and does not include instrument amortization or other reagents.

# ALZHEIMER SAMPLE SEQUENCING RESULTS ON SEQUEL II SYSTEM

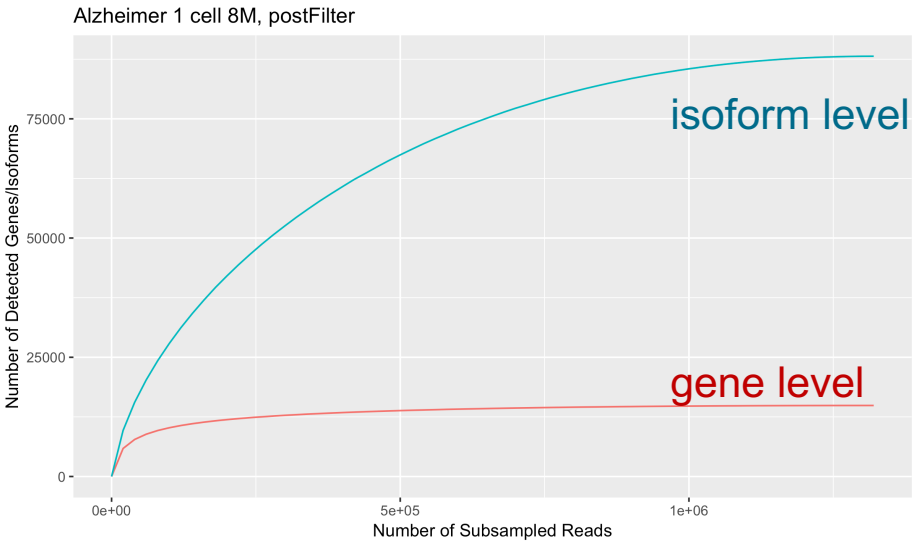


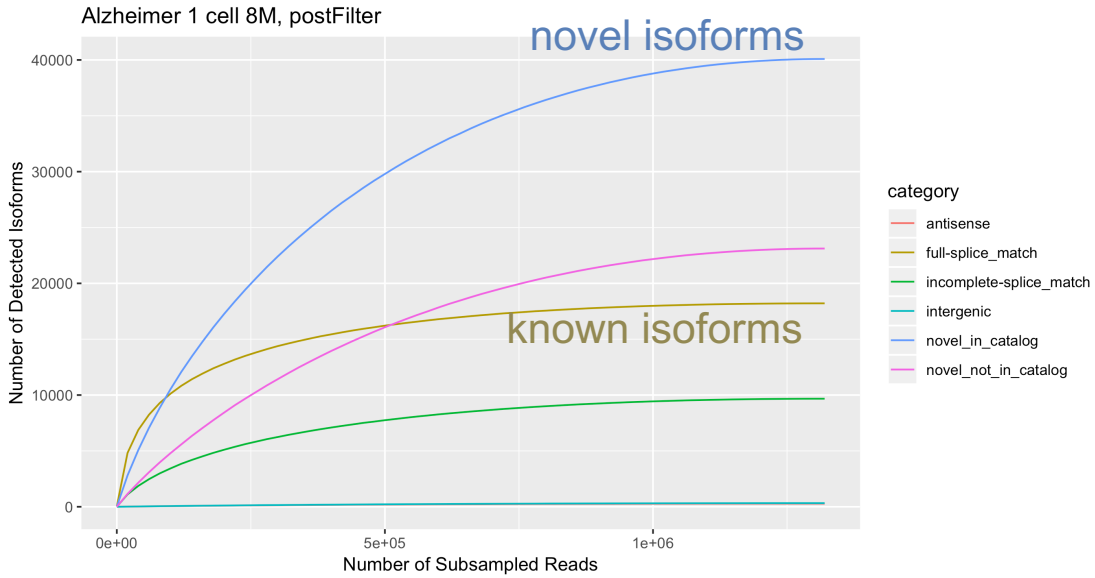

**CCS Read Classification**

| Value | Analysis Metric |
|---|---|
| 5,052,827 | Reads |
| 4,327,705 | Reads with 5' and 3' Primers |
| 4,292,971 | Non-Concatamer Reads with 5' and 3' Primers |
| 4,277,293 | Non-Concatamer Reads with 5' and 3' Primers and Poly-A Tail |
| 2,973 | Mean Length of Full-Length Non-Concatamer Reads |
| 1 | Unique Primers |
| 4,327,705 | Mean Reads per Primer |
| 4,327,705 | Max. Reads per Primer |
| 4,327,705 | Min. Reads per Primer |
| 725,122 | Reads without Primers |

Approx. 4 M full-length non-concatemer reads (85%)

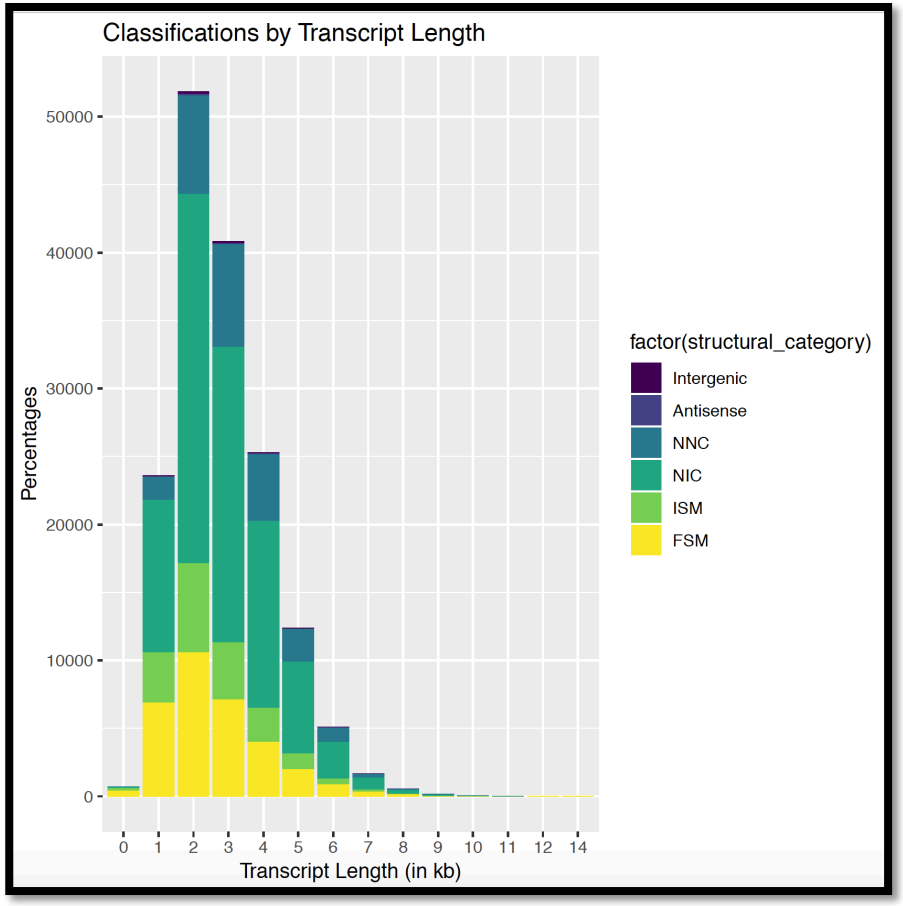
# 1 SMRT CELL 8M SATURATES KNOWN GENES & ISOFORMS



# 1 SMRT CELL 8M DETECTS SHORT & LONG TRANSCRIPTS

Alzheimer brain, 1 SMRT Cell 8M using Iso-Seq Express kit




163,815 transcripts

Min: 80 bp

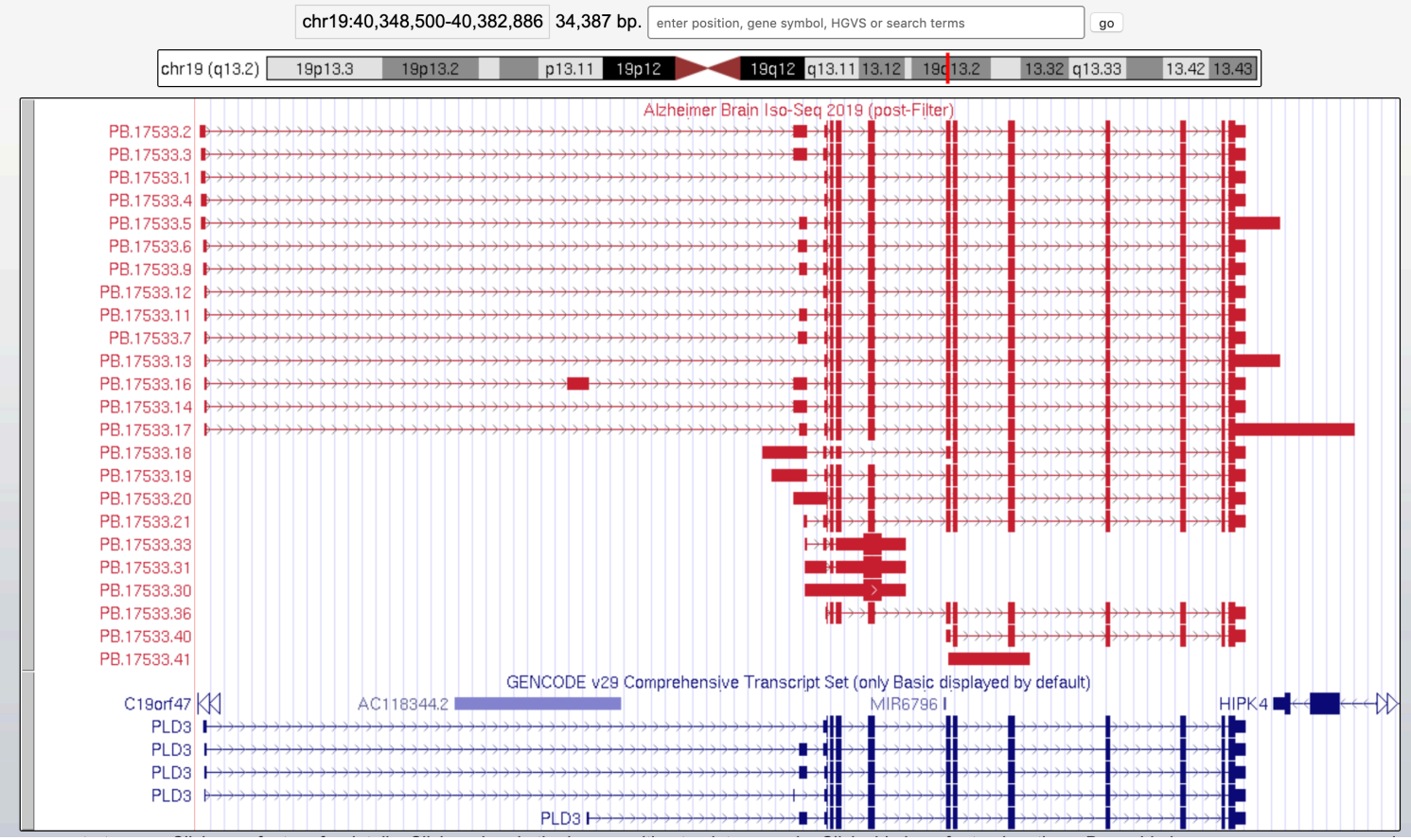
Max: 14,288 bp

Mean: 3,351 bp

# 1 SMRT CELL 8M DETECTS KNOWN & NOVEL TRANSCRIPTS

Alzheimer brain, 1 SMRT cell 8M

Post-processed using SQANTI2



# PLANNING YOUR HUMAN TRANSCRIPTOME PROJECT

| WORKFLOW STEP | DETAILS | COST |
|---|---|---|
| SAMPLE PREP | Prepare full-length transcripts from 300 ng of total RNA | NEBNext® kit – **$27** |
| LIBRARY PREP | Use the SMRTbell Express Template Prep Kit 2.0 to prepare libraries in 1 day | Express Template Prep Kit 2.0 – **$85.64** |
| SMRT SEQUENCING | Sequence 1 SMRT Cell 8M per sample on Sequel II System. 24-hour movie with 2-hour pre-extension | 1 SMRT Cell 8M – **$1,203.98** SMRT Cell run reagents/cons. – **$22.99** |
| DATA ANALYSIS | Generate up to 4 million full-length, reads per SMRT Cell 8M Fully-supported Iso-Seq Analysis in SMRT Link | Software license cost - **$0** |

***PacBio consumables: $1312.61 for 1 SMRT Cell 8M per human transcriptome***

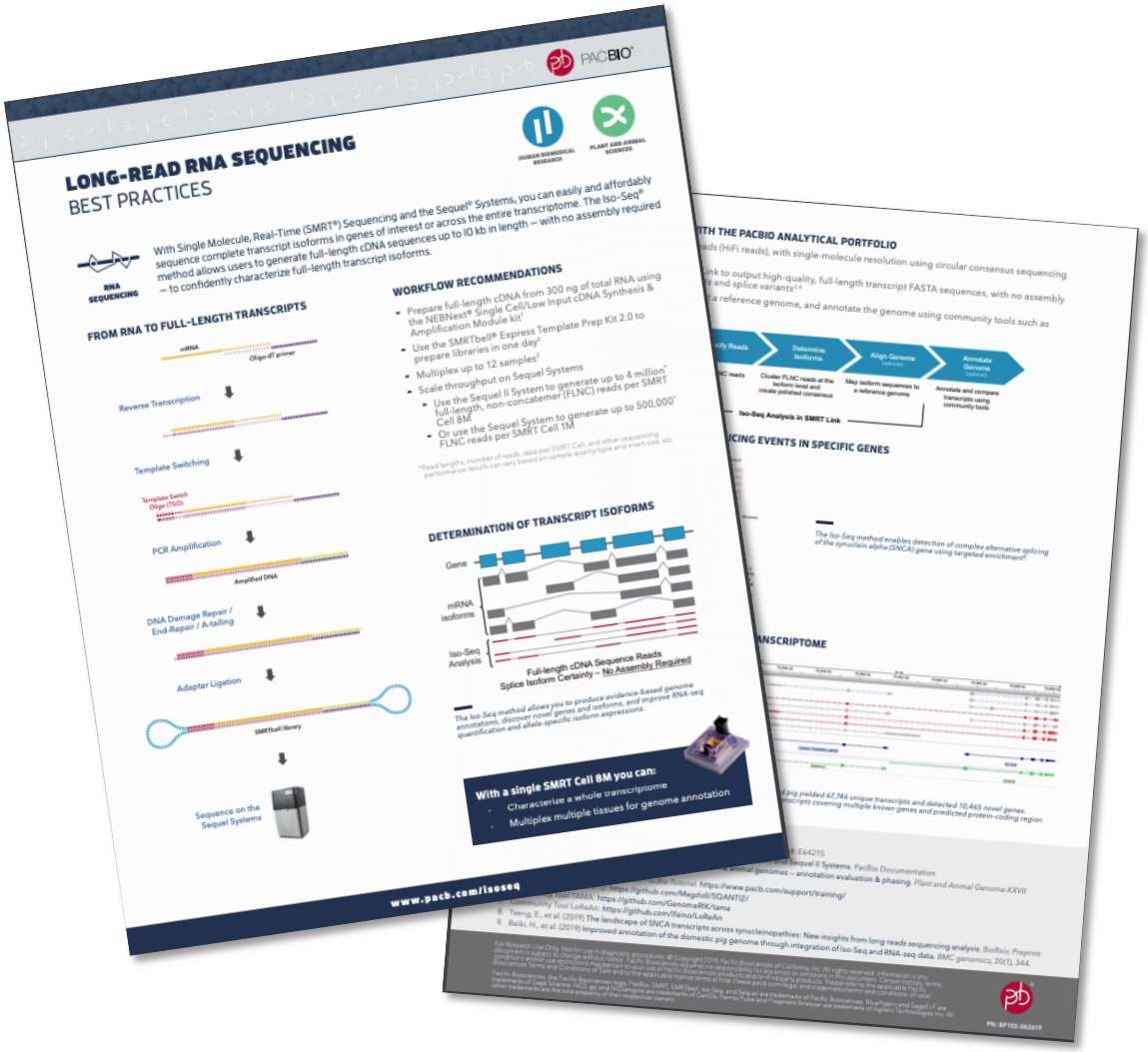
# RECOMMENDATIONS FOR ANNOTATION PROJECTS

***The answer is always 1 SMRT Cell 8M!***

- Go deeper for better reference gene set
- Use tools to correct genome

- Multiplex up to 8 tissues currently per SMRT Cell 8M

# LONG-READ RNA SEQUENCING: BEST PRACTICE GUIDE




- Use the SMRTbell® Express Template Prep Kit 2.0 to prepare libraries in one day

[Procedure & Checklist - Iso-Seq Express Template Preparation for Sequel and Sequel II Systems.]

[Application Brief:] Learn more about these best practices for long-read RNA sequencing

# Where can I learn more?

# BLOG / MEDIUM

Medium

## Liz Tseng

Following

All things RNA. Food lover of Taiwanese origin. Trains in circus arts (rope and straps). Opinions are my own.

Latest

Liz Tseng in PacBio
Jun 17 · 6 min read

**A Tale of Mutants: Sequencing the Full-Length Influenza Virus at the Single Cell Level**

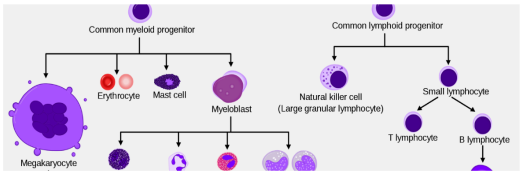
12

Liz Tseng in PacBio
May 9 · 7 min read

**SMRT Leiden 2019 Day 3: Bioinformatics for Transcriptome, Genome Analysis, and Targeted Sequencing**

4

Liz Tseng in PacBio
Apr 15 · 7 min read




**Finding Differentiating Isoforms in Bone Marrow Subpopulations: an interview with Anne**

51

https://medium.com/@Magdoll

**RNA SEQUENCING**

Human RNA Sequencing

RNA Sequencing for Plant and Animal Sciences

## BRING CERTAINTY TO ISOFORM DETECTION

Changes in isoform variants and relative abundancies are an important clue in deciphering how genomic variants drive the phenotypic differences between health and disease. Only long-read sequencing can provide unbiased, direct detection of all the isoforms present in your sample.

## OBTAIN A MORE COMPREHENSIVE SET OF TRANSCRIPT ISOFORMS

The PacBio Isoform Sequence (Iso-Seq) method generates full-length cDNA sequences – from 5’ to the poly-A tail – to confidently characterize the full transcriptome.

# HUMAN BIOMEDICAL RESEARCH

## SMRT Reads

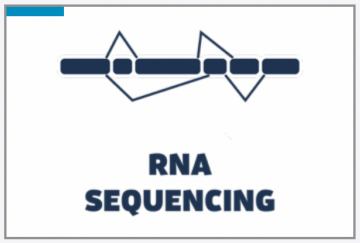
PACBIO®

Whether you are profiling whole transcriptomes or screening gene isoforms as potential biomarkers, PacBio RNA Sequencing provides comprehensive and accurate full-length transcript detection. Explore our new, faster Iso-Seq Express workflow and hear how RNA Sequencing is advancing understanding of human health and disease.

Hear the Latest

Introducing Iso-Seq Express for faster RNA Sequencing

Learn the Method

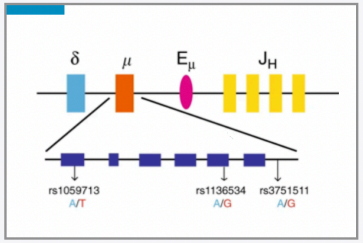



Generate full-length cDNA sequences up to 10 kb in length — no assembly required

Read a Publication

Targeted Iso-Seq analysis reveals new insights on SNCA gene and its role in complex disease

Explore an Article




Long-read sequencing unveils IGH-DUX4 translocation in B-cell acute lymphoblastic leukemia

# 2019 SMRT GRANT PROGRAMS

PACBIO®

**Sequence with Confidence**

Resolve the Genetic Complexity of Neurological Disease with SMRT® Sequencing

## Upcoming SMRT Grant Programs

Discover the benefits for your research and explore the range of Single Molecule, Real-Time (SMRT) Sequencing Applications

| Research Focus | Application | Grant Opens |
|---|---|---|
| Neuroscience Research | Multiple Applications | Now Open |
| Human Biomedical Research | Whole Genome Sequencing | October |
| Celebrate Advances in SMRT Sequencing | Multiple Applications | November |

# ISO-SEQ SEQUEL II DATA

## Universal Human Reference (UHRR)

## 2 SMRT Cells 8M

**SAMPLE**

Universal Human Reference RNA (Agilent) + SIRV Isoform Mix E0 (Lexogen)

**METHODS**

- Library prep Iso-Seq Template Preparation for Sequel Systems (PN 101-070-200)
- Sequencing Sequel System II with "Early Access" binding kit (101-490-800) and chemistry (101-490-900)
- Run time:
- ◦ 4 hrs. pre-extension
- ◦ 15 hrs. run time per SMRT Cell
- Reference hs37d5 (GRCh37 with decoy)
- Analysis SMRTlink 7.0 "IsoSeq With Mapping protocol" with hg38+SIRV combined reference genome

**FILES**

Users are encouraged to work with the mapped, unique transcripts in GFF and FASTQ format provided in the "PolishedMappedTranscripts" subfolder. The two raw movie files are in "subreads" subfolder. The intermediate data, full-length reads, are available in "FullLengthReads" folder.

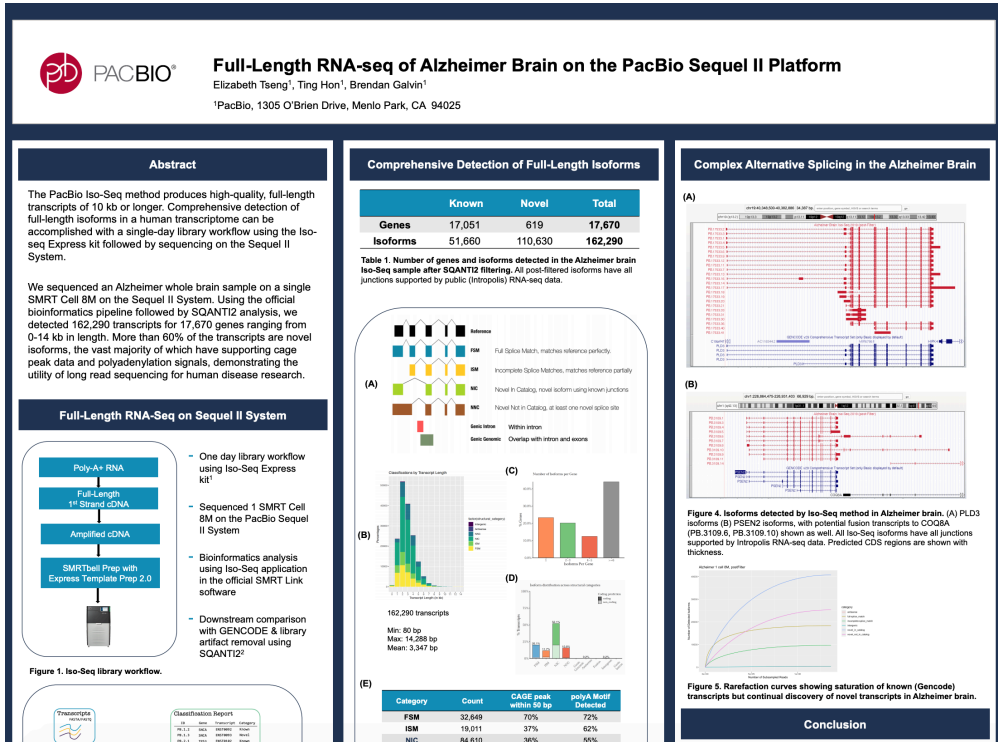
```
PolishedMappedTranscripts/
                      - out.abundance.txt
                      - out.fastq
                      - out.gff

FullLengthReads/
                     - flnc.bam

Subreads/
                     - m64012_181221_231243.subreads.bam
                     - m64012_181221_231243.subreads.bam.pbi
                     - m64012_181222_192540.subreads.bam
                     - m64012_181222_192540.subreads.bam.pbi
```

Download

**Download from China**: https://downloads-ap.pacbcloud.com/public/dataset/UHR_IsoSeq/

## Alzheimer Whole Brain

## 1 SMRT Cell 8M (data release pending)



PACBIO

**Full-Length RNA-seq of Alzheimer Brain on the PacBio Sequel II Platform**

Elizabeth Tseng[1], Ting Hon[1], Brendan Galvin[1]

[1]PacBio, 1305 O'Brien Drive, Menlo Park, CA 94025

**Abstract**

The PacBio Iso-Seq method produces high-quality, full-length transcripts of 10 kb or longer. Comprehensive detection of full-length isoforms in a human transcriptome can be accomplished with a single-day library workflow using the Iso-seq Express kit followed by sequencing on the Sequel II System.

We sequenced an Alzheimer whole brain sample on a single SMRT Cell 8M on the Sequel II System. Using the official bioinformatics pipeline followed by SQANTI2 analysis, we detected 162,290 transcripts for 17,670 genes ranging from 0-14 kb in length. More than 60% of the transcripts are novel isoforms, the vast majority of which have supporting cage peak data and polyadenylation signals, demonstrating the utility of long read sequencing for human disease research.

**Full-Length RNA-Seq on Sequel II System**

Poly-A+ RNA → Full-Length 1st Strand cDNA → Amplified cDNA → SMRTbell Prep with Express Template Prep 2.0

- One day library workflow using Iso-Seq Express kit[1]
- Sequenced 1 SMRT Cell 8M on the PacBio Sequel II System
- Bioinformatics analysis using Iso-Seq application in the official SMRT Link software
- Downstream comparison with GENCODE & library artifact removal using SQANTI2[2]

Figure 1. Iso-Seq library workflow.

**Comprehensive Detection of Full-Length Isoforms**

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

Table 1. Number of genes and isoforms detected in the Alzheimer brain Iso-Seq sample after SQANTI2 filtering. All post-filtered isoforms have all junctions supported by public (Intropolis) RNA-seq data.

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, at least one novel splice site; Genic Intron: Within intron; Genic Genome: Overlap with intron and exons

162,290 transcripts

Min: 80 bp
Max: 14,288 bp
Mean: 3,347 bp

| Category | Count | CAGE peak within 50 bp | polyA Motif Detected |
|---|---|---|---|
| FSM | 32,649 | 70% | 72% |
| ISM | 19,011 | 37% | 62% |
| NIC | 84,610 | 38% | 65% |

**Complex Alternative Splicing in the Alzheimer Brain**

Figure 4. Isoforms detected by Iso-Seq method in Alzheimer brain. (A) PLD3 isoforms (B) PSEN2 isoforms, with potential fusion transcripts to COQ8A (PB.3109.6, PB.3109.10) shown as well. All Iso-Seq isoforms have all junctions supported by Intropolis RNA-seq data. Predicted CDS regions are shown with thickness.

Figure 5. Rarefaction curves showing saturation of known (Gencode) transcripts but continual discovery of novel transcripts in Alzheimer brain.

**Conclusion**

AGBT-PH2019 poster

PACBIO®

www.pacb.com

For Research Use Only. Not for use in diagnostic procedures. © Copyright 2019 by Pacific Biosciences of California, Inc. All rights reserved. Pacific Biosciences, the Pacific Biosciences logo, PacBio, SMRT, SMRTbell, Iso-Seq, and Sequel are trademarks of Pacific Biosciences. BluePippin and SageELF are trademarks of Sage Science. NGS-go and NGSengine are trademarks of GenDx. FEMTO Pulse and Fragment Analyzer are trademarks of Advanced Analytical Technologies.

All other trademarks are the sole property of their respective owners.