



PACIFIC
BIOSCIENCES®

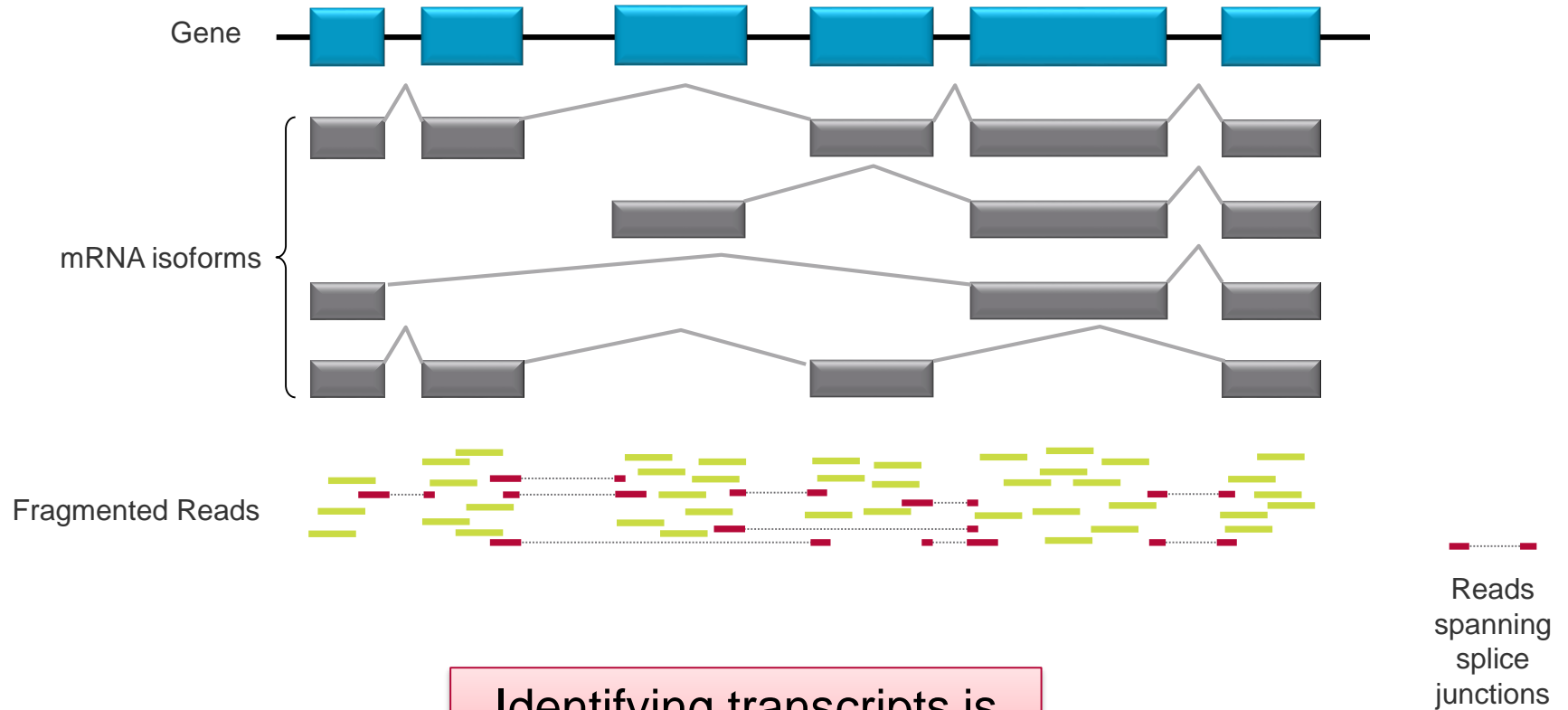
Accurate Characterization of Complex Isoforms in Bulk and Single Cell Transcriptome

Elizabeth Tseng, Principal Scientist, PacBio

TALK OVERVIEW

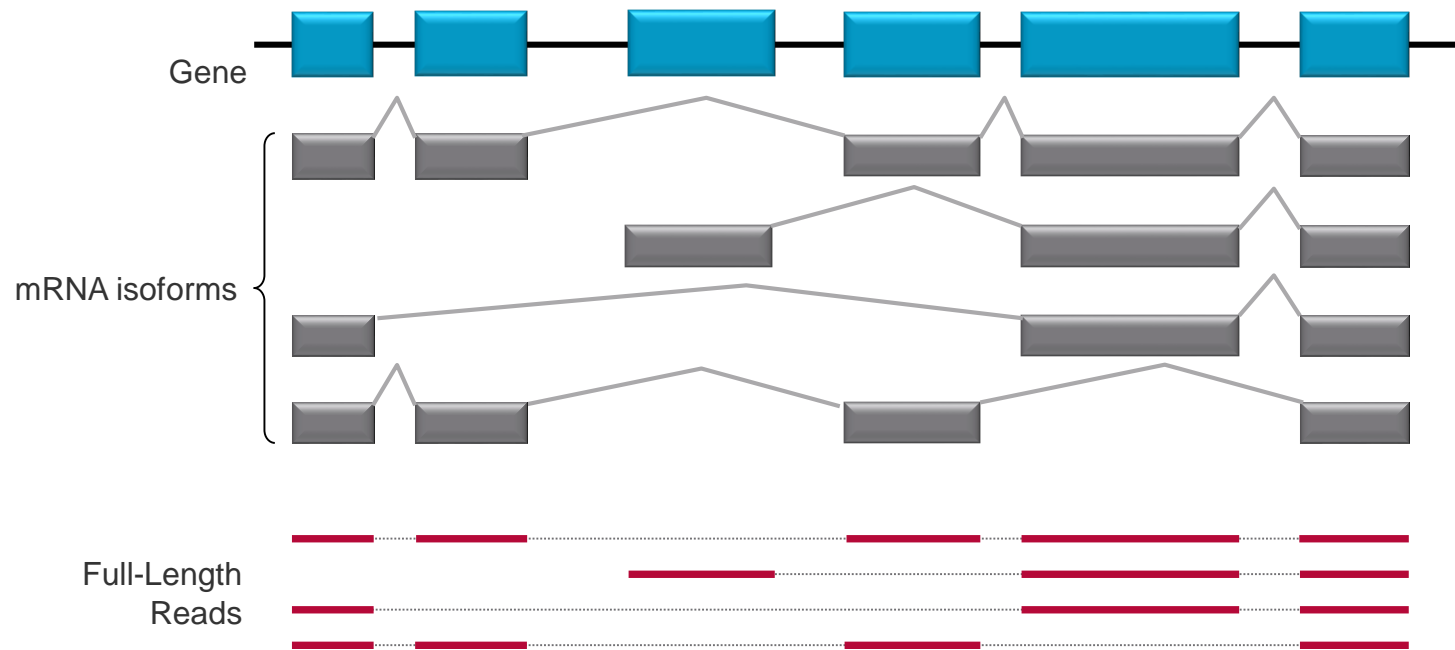
- Iso-Seq Introduction
- Iso-Seq for Whole Transcriptome Sequencing
- Single Cell Iso-Seq

TRADITIONAL RNA-SEQ



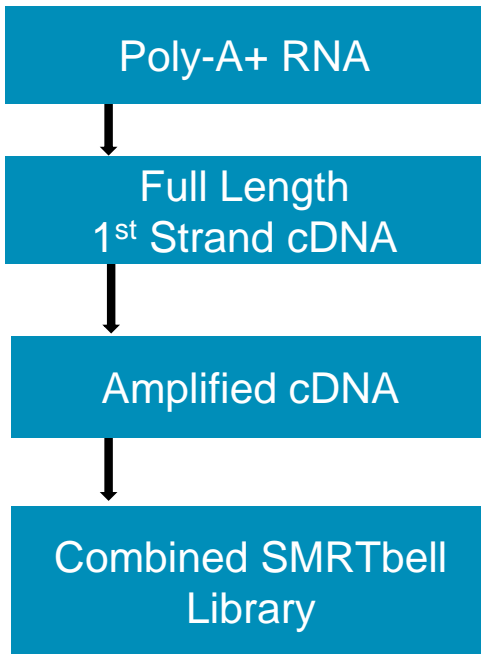
Identifying transcripts is an **assembly** problem

PACBIO ISO-SEQ: FULL-LENGTH RNA-SEQ



Identifying transcripts is a **clustering** problem

PACBIO ISO-SEQ: FULL-LENGTH RNA-SEQ



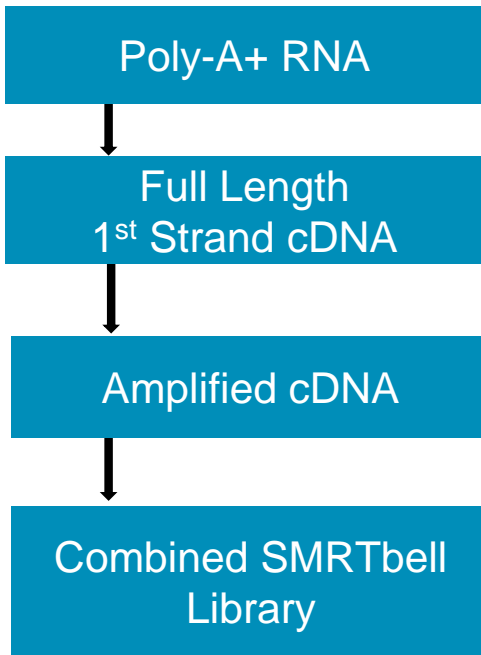
Sample Prep:

- Full-Length cDNA
- One read = one molecule

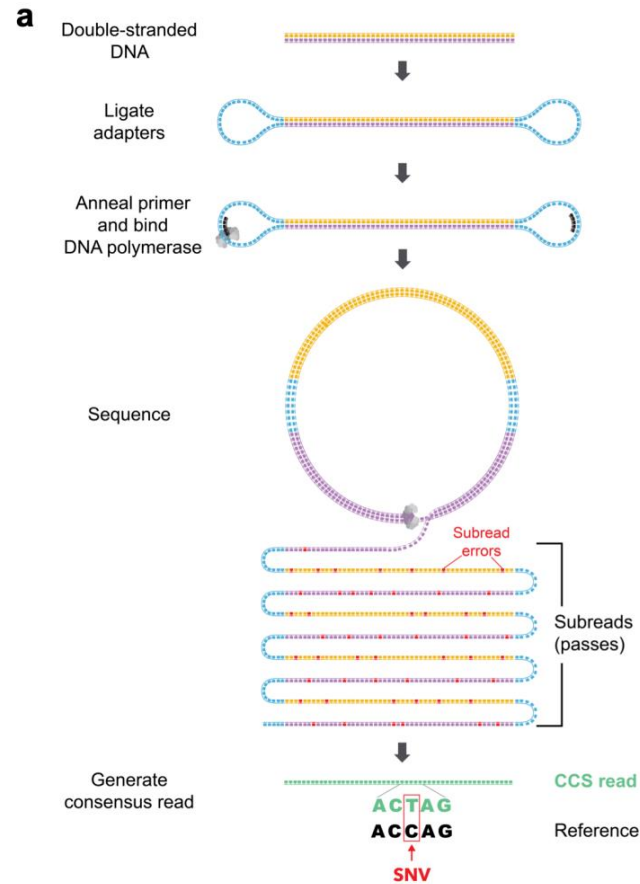
Analysis:

- Group full-length reads at transcript-level

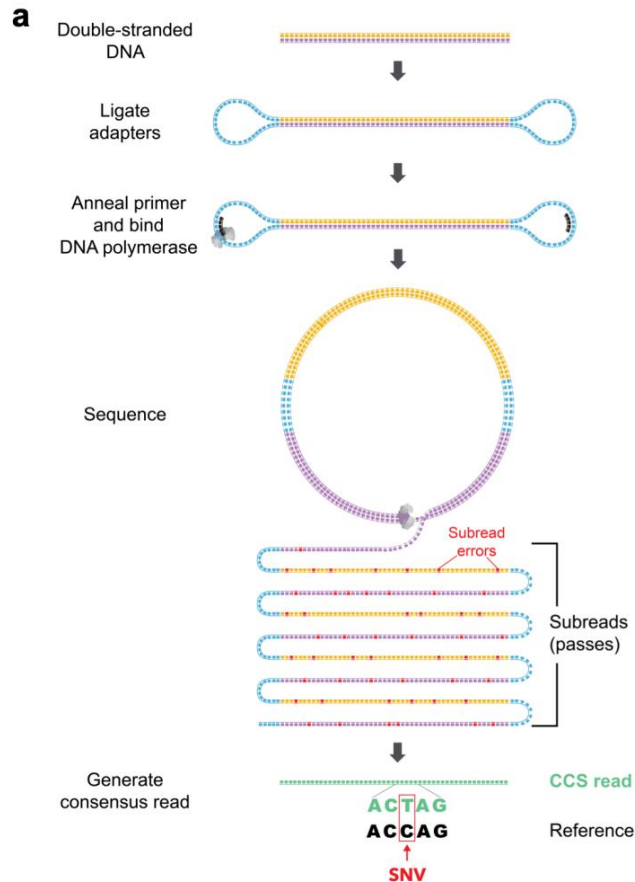
ISO-SEQ READS ARE HIGHLY ACCURATE



Generate CCS reads

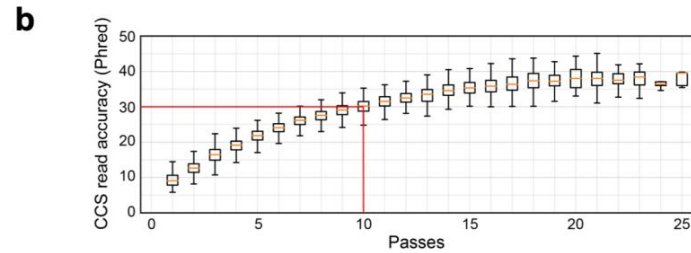
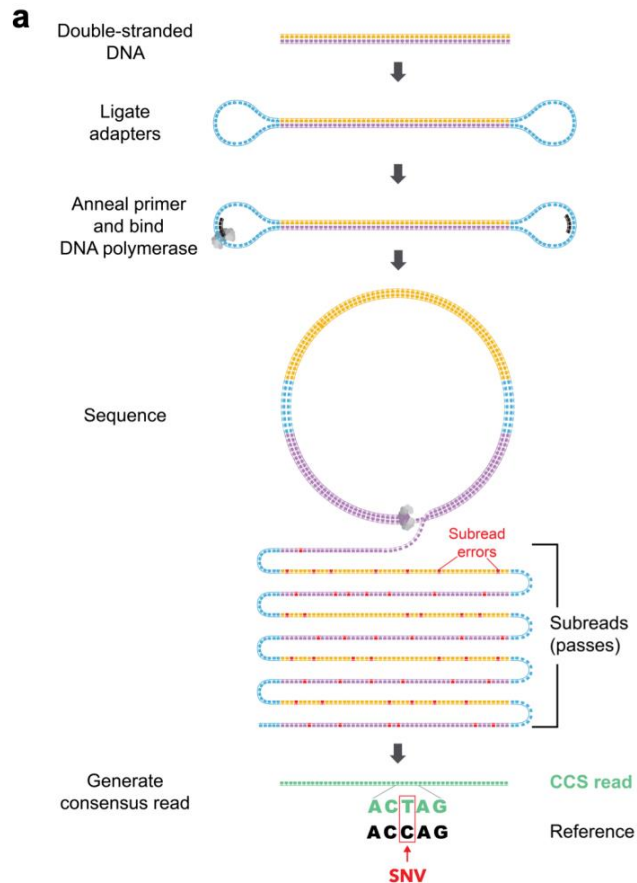


CIRCULAR CONSENSUS SEQUENCING GENERATES HIGH FIDELITY READS



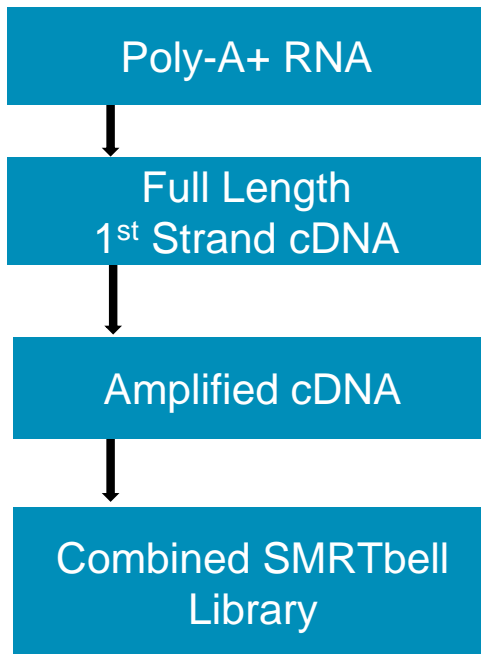
- Circular Consensus Sequencing (CCS) derives a consensus read from multiple passes
- Errors are randomly distributed → more passes, errors cancel out

CIRCULAR CONSENSUS SEQUENCING GENERATES HIGH FIDELITY READS



- At 10-pass, reads reach QV30 (99.9%)

ISO-SEQ ADVANTAGES



- *de novo* (genome not required)
- No assembly
- Full-length (5' to 3')
- High accuracy (>99%)

ISO-SEQ: SEQUEL II REDUCES SEQUENCING NEED

Iso-Seq on Sequel

Cell	Pol Reads	Pol Reads ReadLength	Full-Length Reads	Genes	Isoforms
1	692,920	46 kb	430,257	9,795	21,655
2	708,612	39 kb	420,709	9,574	20,997

Iso-Seq on Sequel II

Cell	Pol Reads	Pol Reads ReadLength	Full-Length Reads	Genes	Isoforms
1	4,526,555	52 kb	2,673,823	14,955	84,096
2	6,438,820	50 kb	4,077,813	16,314	106,358

SEQUEL II DATA RELEASE

[Data Release Link](#)

Sequel II System Data Release: Universal Human Reference (UHR) Iso Seq

tkerelska edited this page 2 hours ago · 9 revisions

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



Bulk Transcriptome Sequencing

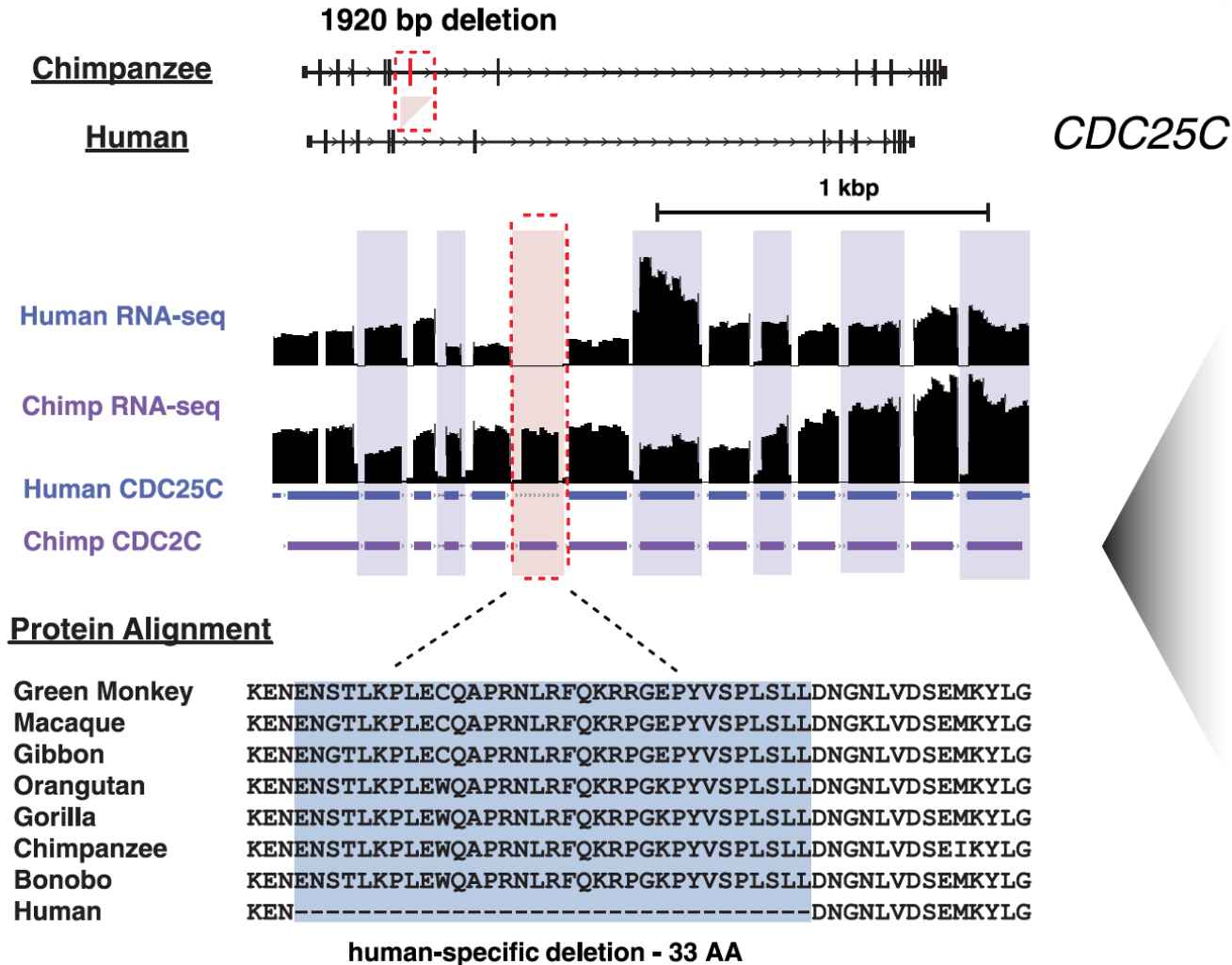
COMPARATIVE GENOME + TRANSCRIPTOME SEQUENCING



- 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 (~600k 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](#)



SEQUENCING THE CANCER GENOME & TRANSCRIPTOME

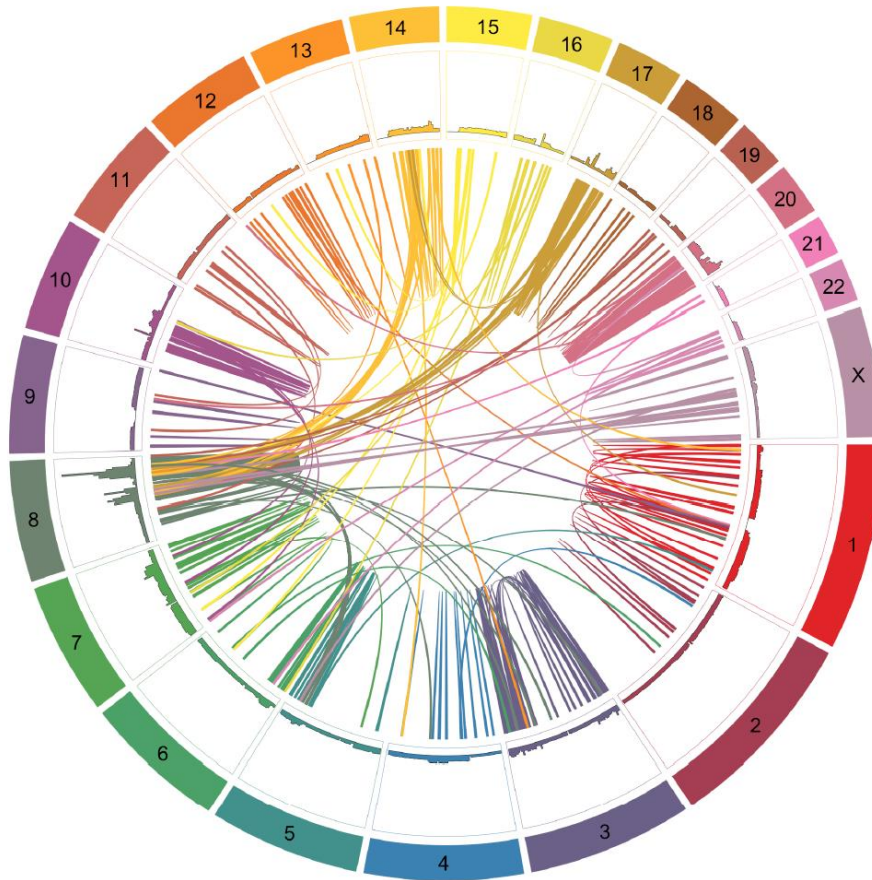


Complex rearrangements and oncogene amplifications revealed by long-read DNA and RNA sequencing of a breast cancer cell line

Maria Nattestad¹, Sara Goodwin¹, Karen Ng², Timour Baslan¹, Fritz Sedlazeck³, Philipp Rescheneder⁴, Tyler Garvin¹, Han Fang¹, James Gurtowski¹, Elizabeth Hutton¹, Elizabeth Tseng⁵, Jason Chin⁵, Timothy Beck², Yogi Sundaravadanam², Melissa Kramer¹, Eric Antoniou¹, John McPherson², James Hicks¹, W. Richard McCombie¹ and Michael C Schatz^{6,7}

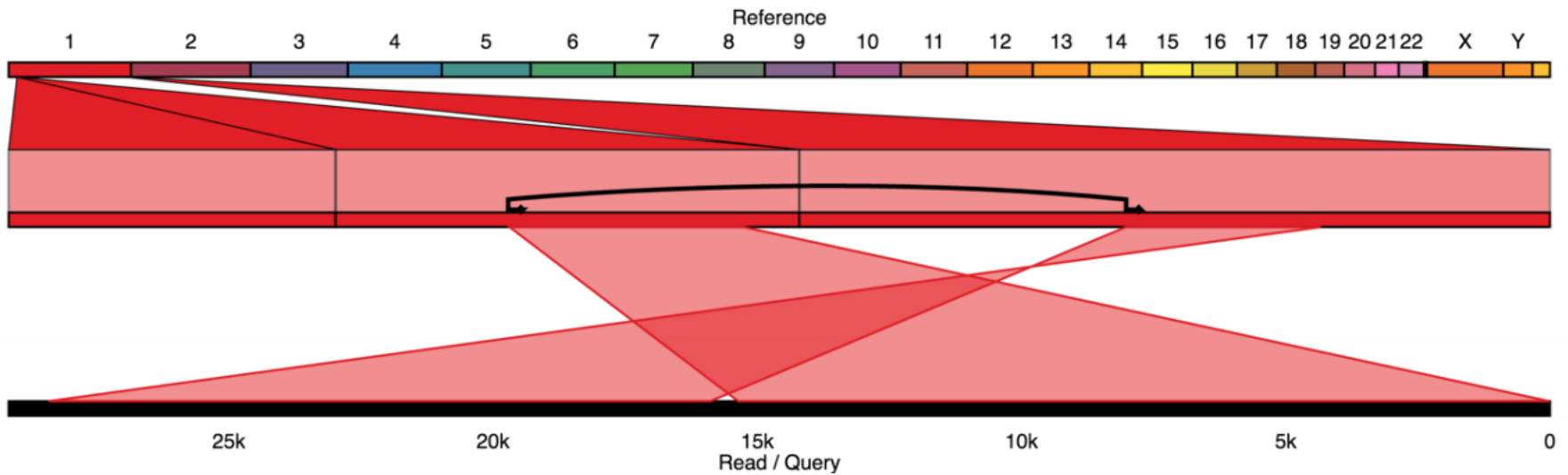
Blog: [Visualizing the Chaos of Cancer](#)

SK-BR3 REARRANGEMENTS, VISUALIZED



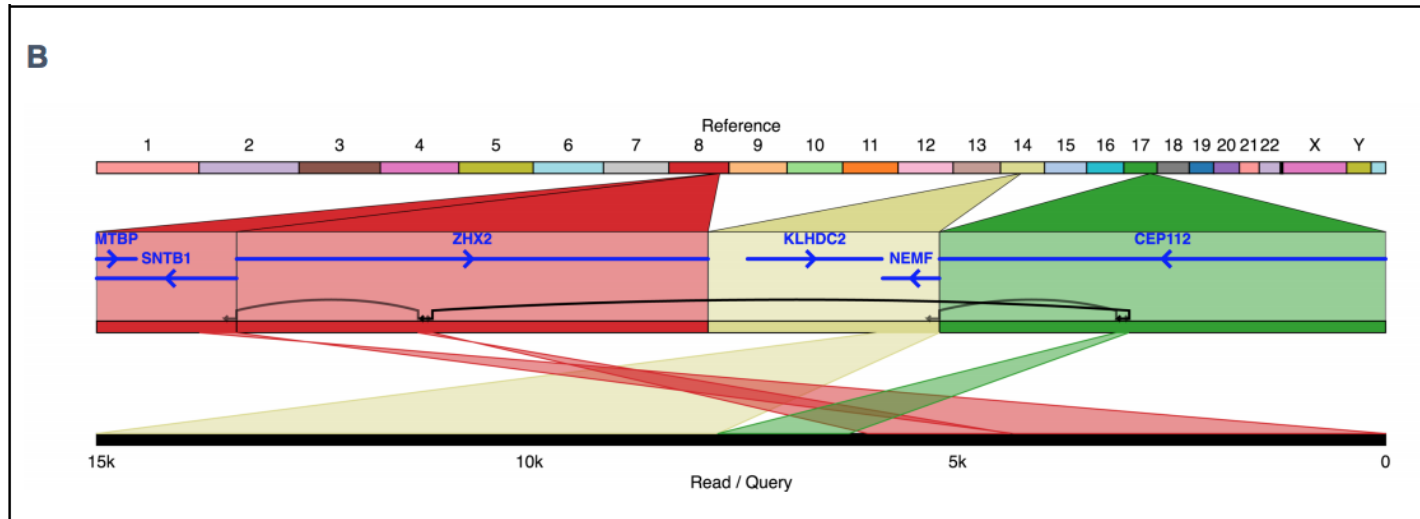
Blog: [Visualizing the Chaos of Cancer](#)

SK-BR3 REARRANGEMENTS, VISUALIZED



Blog: [Visualizing the Chaos of Cancer](#)

ISO-SEQ CONFIRMS A NEW “3-HOP” GENE



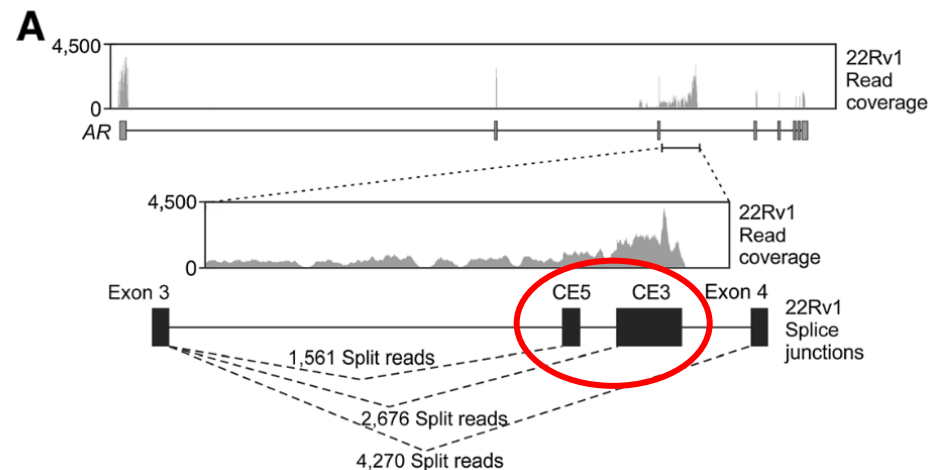
Supplementary Figure 18. Ribbon plot of “3-hop” KLHDC2-SNTB1 gene fusion captured by long reads. This is a “3-hop” gene fusion in SK-BR-3 created by a series of three variants (A). These variants are captured together in several individual SMRT sequencing reads, one of which is shown in (B).

- In total, Iso-Seq identified 15 gene fusions with genomic evidence

Blog: [Visualizing the Chaos of Cancer](#)

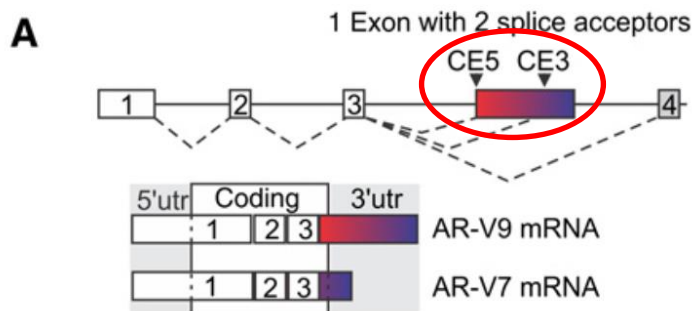
ISO-SEQ FINDS NOVEL AR ISOFORM IN PROSTATE CANCER

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



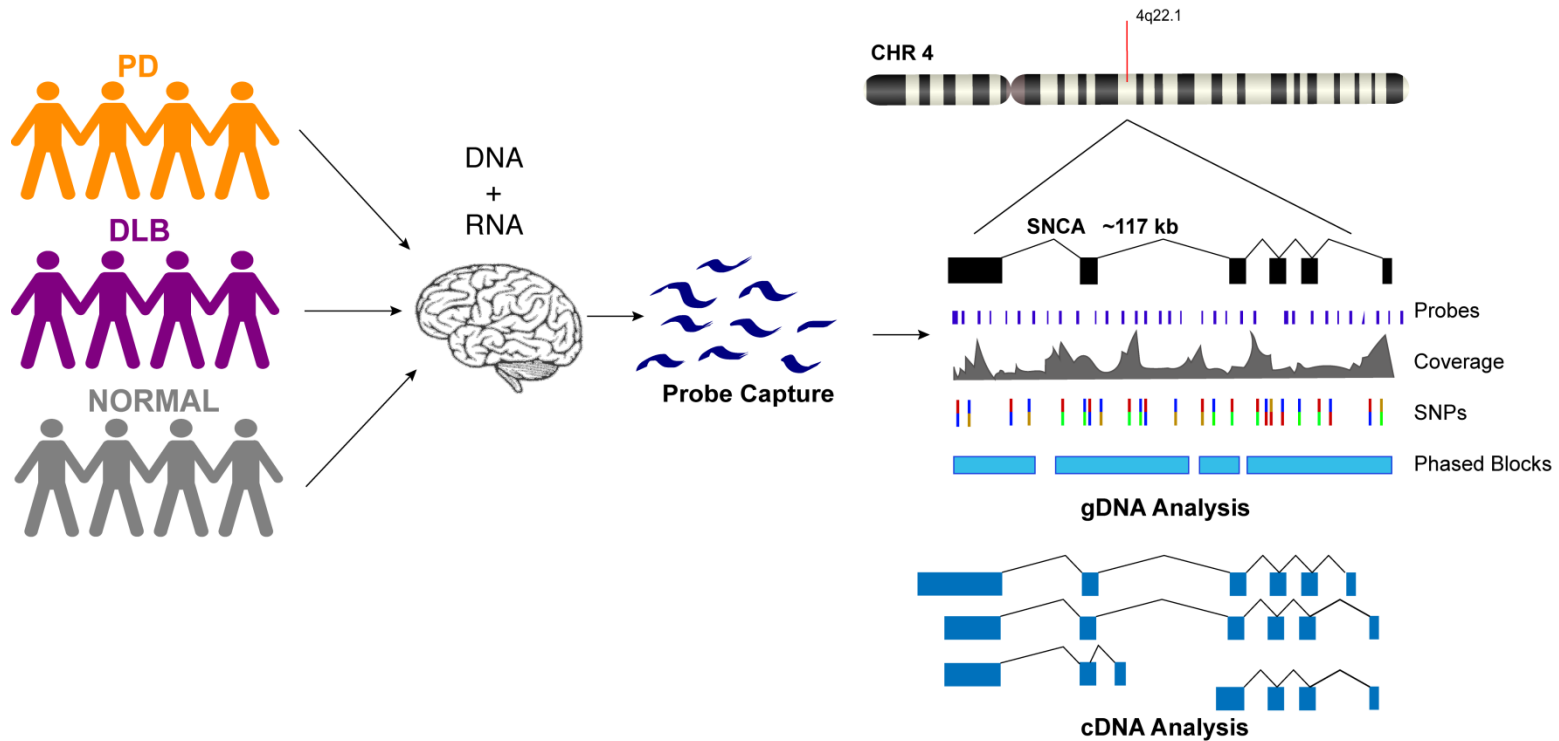
ISO-SEQ FINDS NOVEL AR ISOFORM IN PROSTATE CANCER

- 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
- AR-V9 expression predictive of therapy resistance

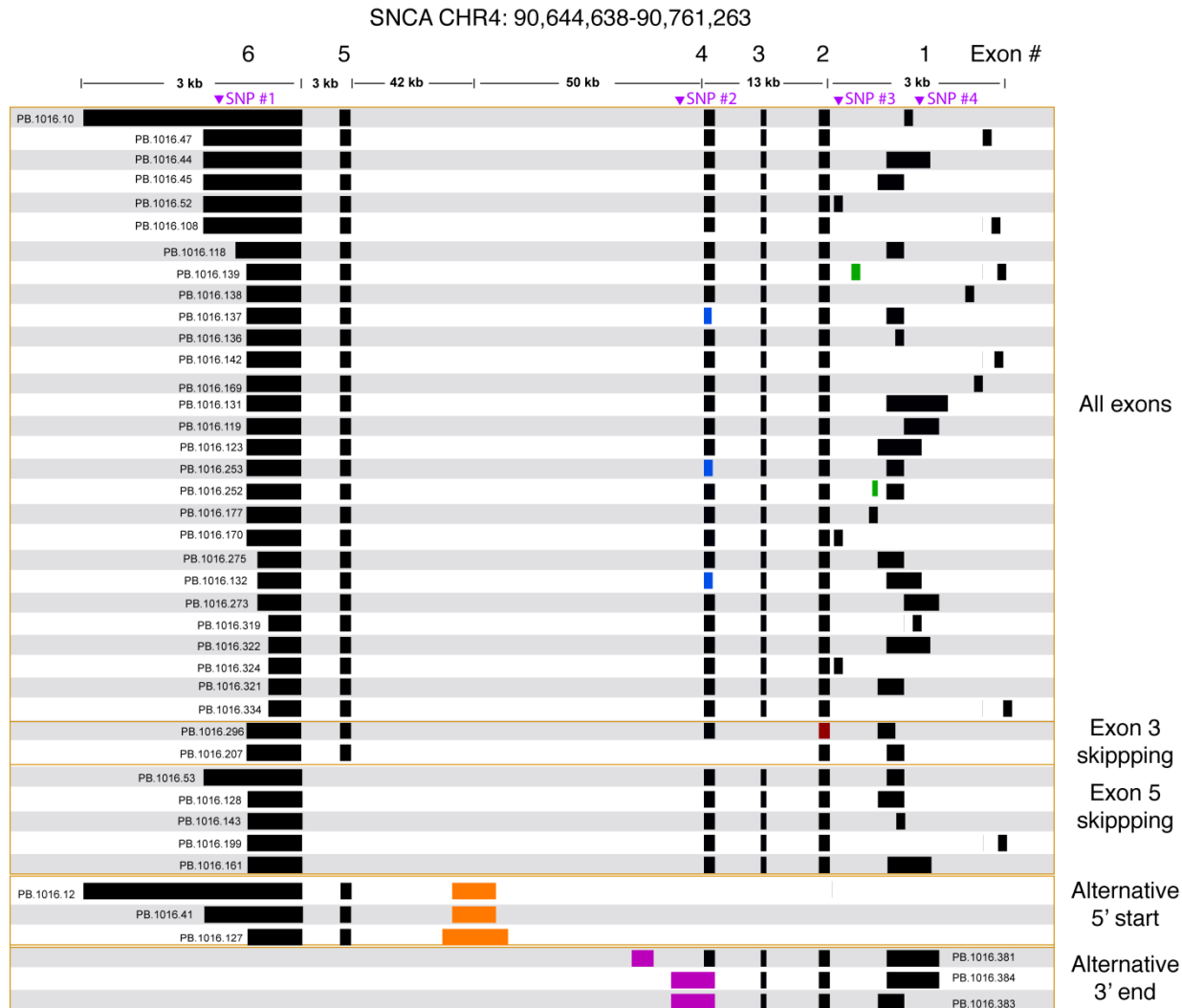


TARGETED DNA + RNA CAPTURE OF SNCA GENE

Collaboration with Ornit Chiba-Falek (Duke), [preprint online](#)



TARGETED DNA + RNA CAPTURE OF SNCA GENE

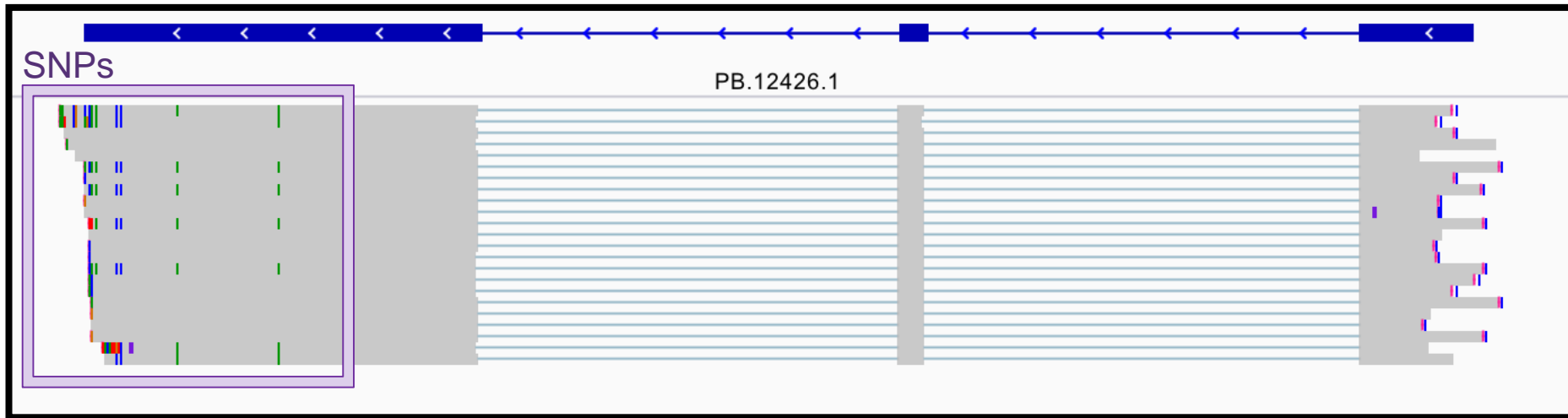




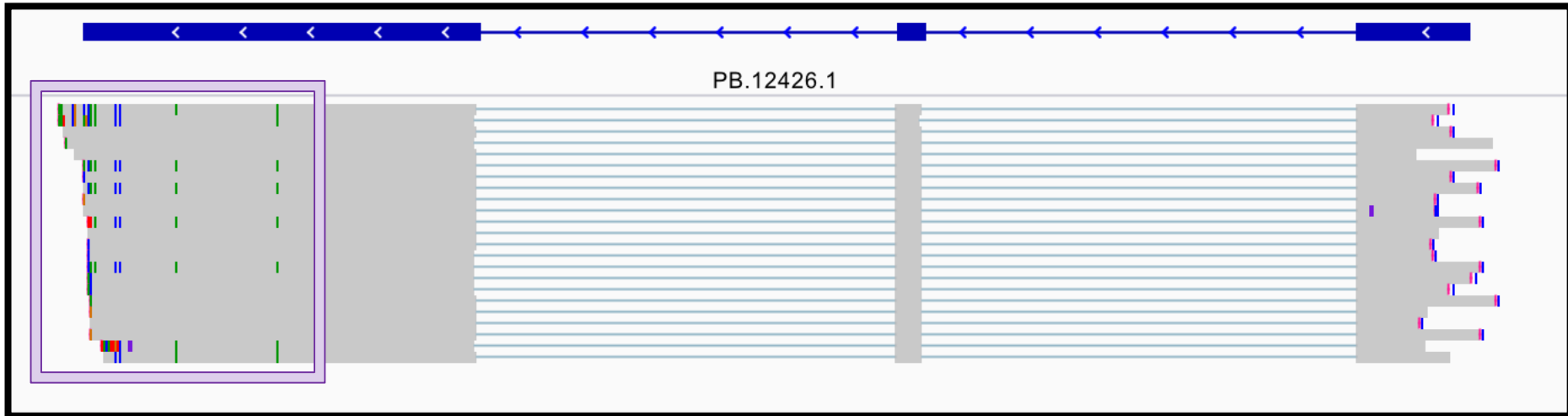
Isoform-level Haplotyping

ISOFORM-LEVEL PHASING IN ISO-SEQ

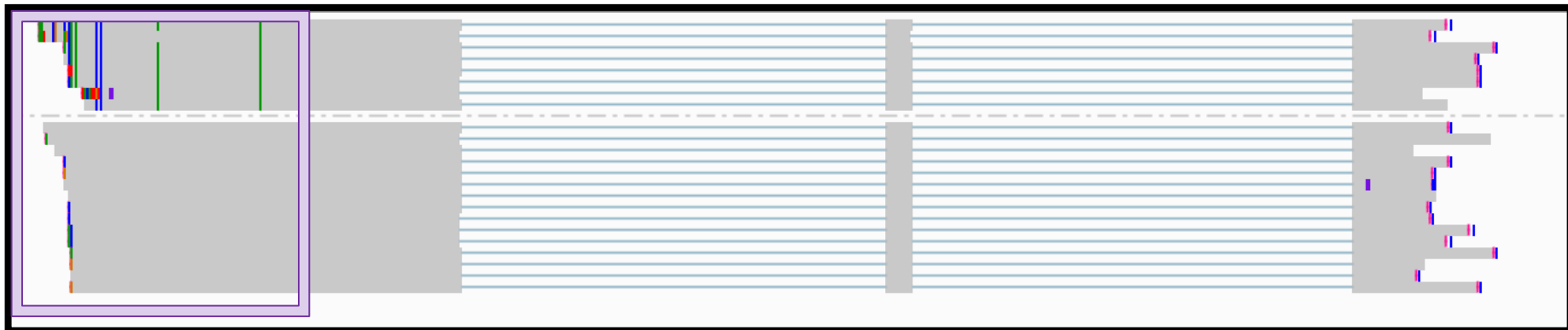
The full-length and single-molecule nature of Iso-Seq data enables isoform-level phasing.



ISOFORM-LEVEL PHASING IN ISO-SEQ

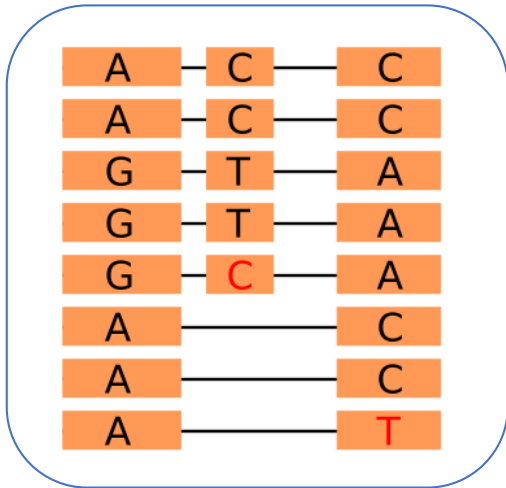


Sort alignment in IGV by SNP

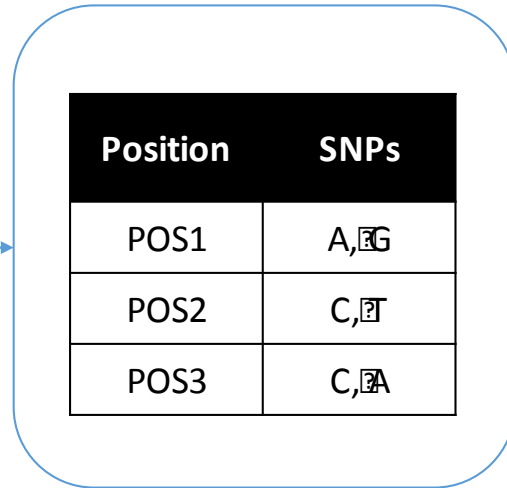


ISO-PHASE: ISOFORM PHASING USING ISO-SEQ DATA

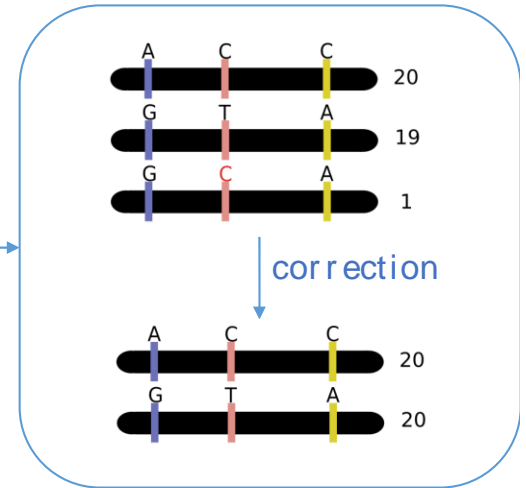
ALIGNMENT



SNP CALLING



PHASING

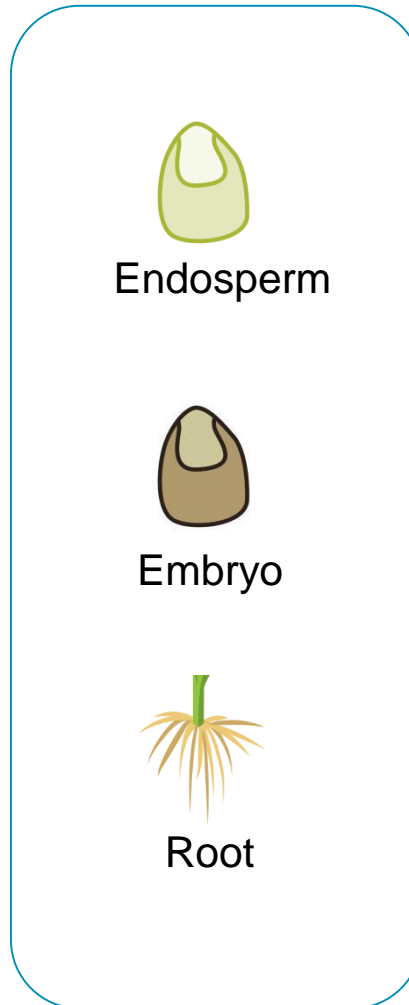
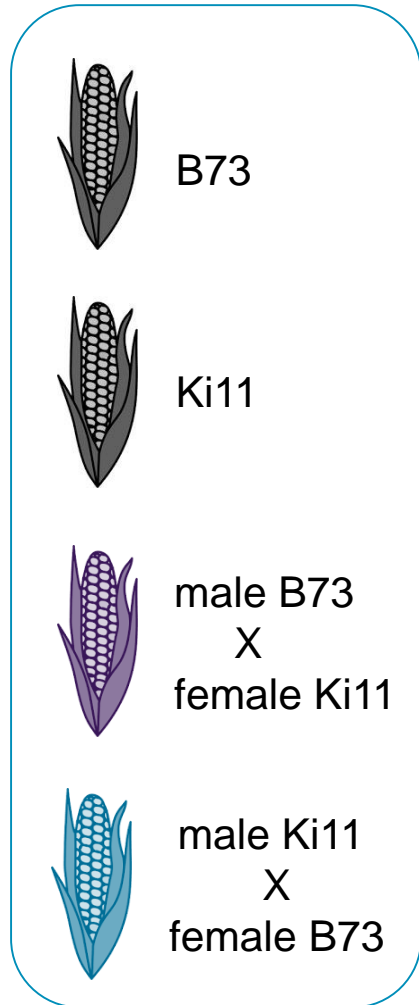


VCF OUTPUT

Can optionally include RNA-seq as input for SNP calling

```
##fileformat=VCFv4.2
#CHROM POS ID REF ALT QUAL FILTER INFO FORMAT ISOFORM1 ISOFORM2
chr1 105 . A G . PASS DP=40;AF=0.50 GT:HQ 0|1:20,20 0:15
chr1 190 . C T . PASS DP=40;AF=0.50 GT:HQ 0|1:20,20 0:15
chr1 336 . C A . PASS DP=40;AF=0.50 GT:HQ 0|1:20,20 0:15
```

SAMPLE SETUP



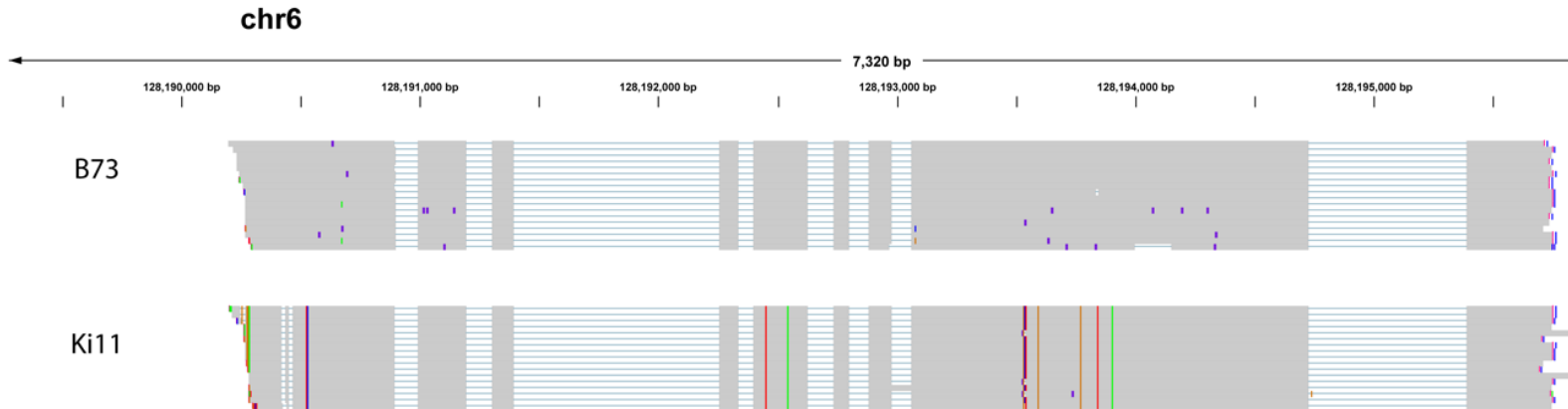
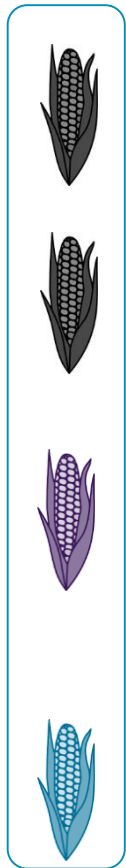
12 barcoded cDNA samples

Pooled into 4 libraries

Sequenced for each lib

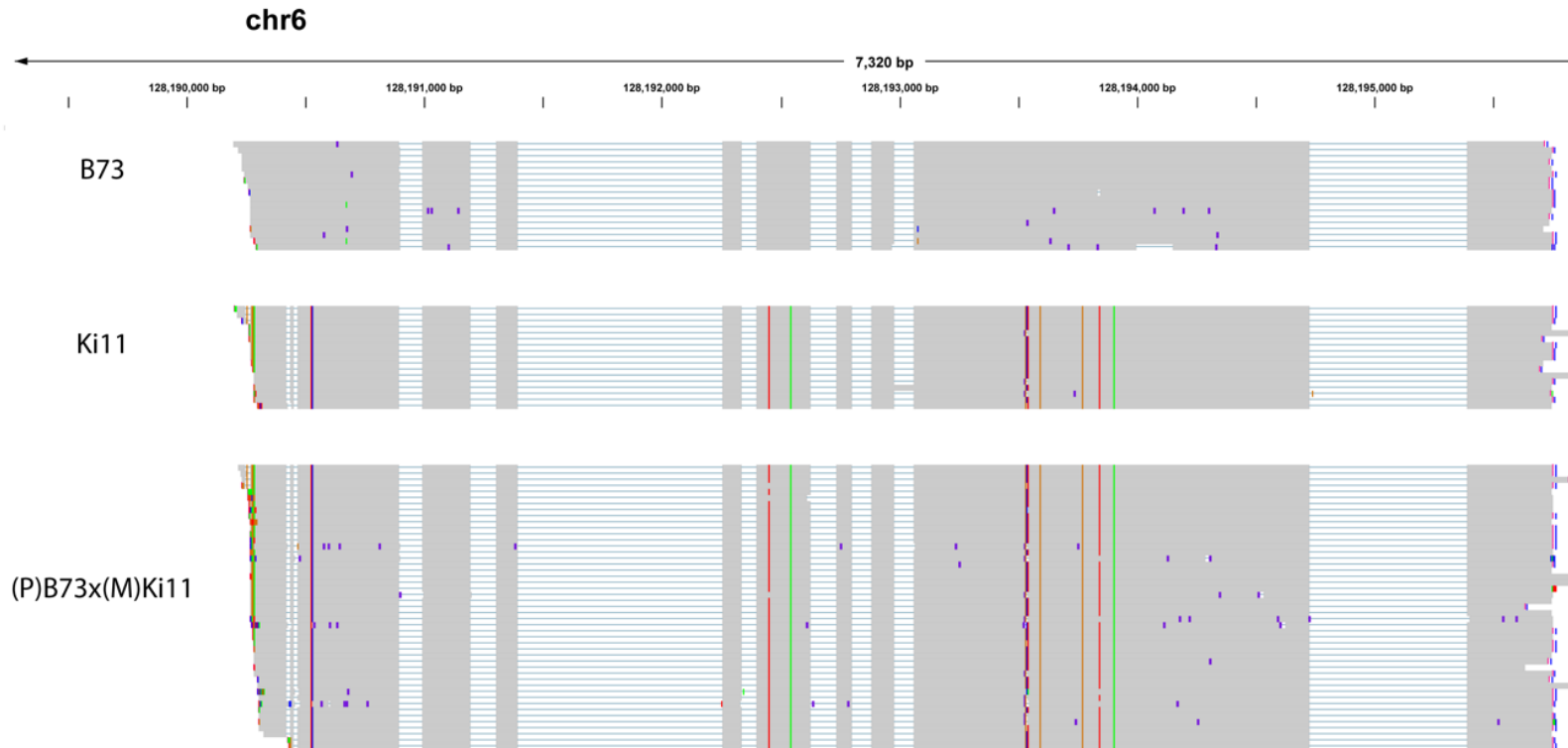
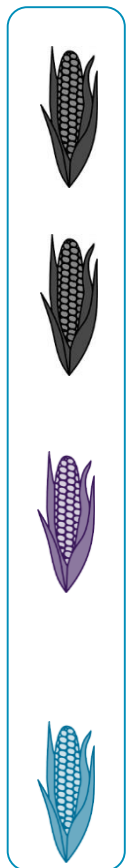
ALLELE-SPECIFIC EXPRESSION

- B73xKi11 only expresses the Ki11 (female) allele
- Ki11xB73 only expresses the B73 (female) allele
- Short read data confirms this allele-specific expression



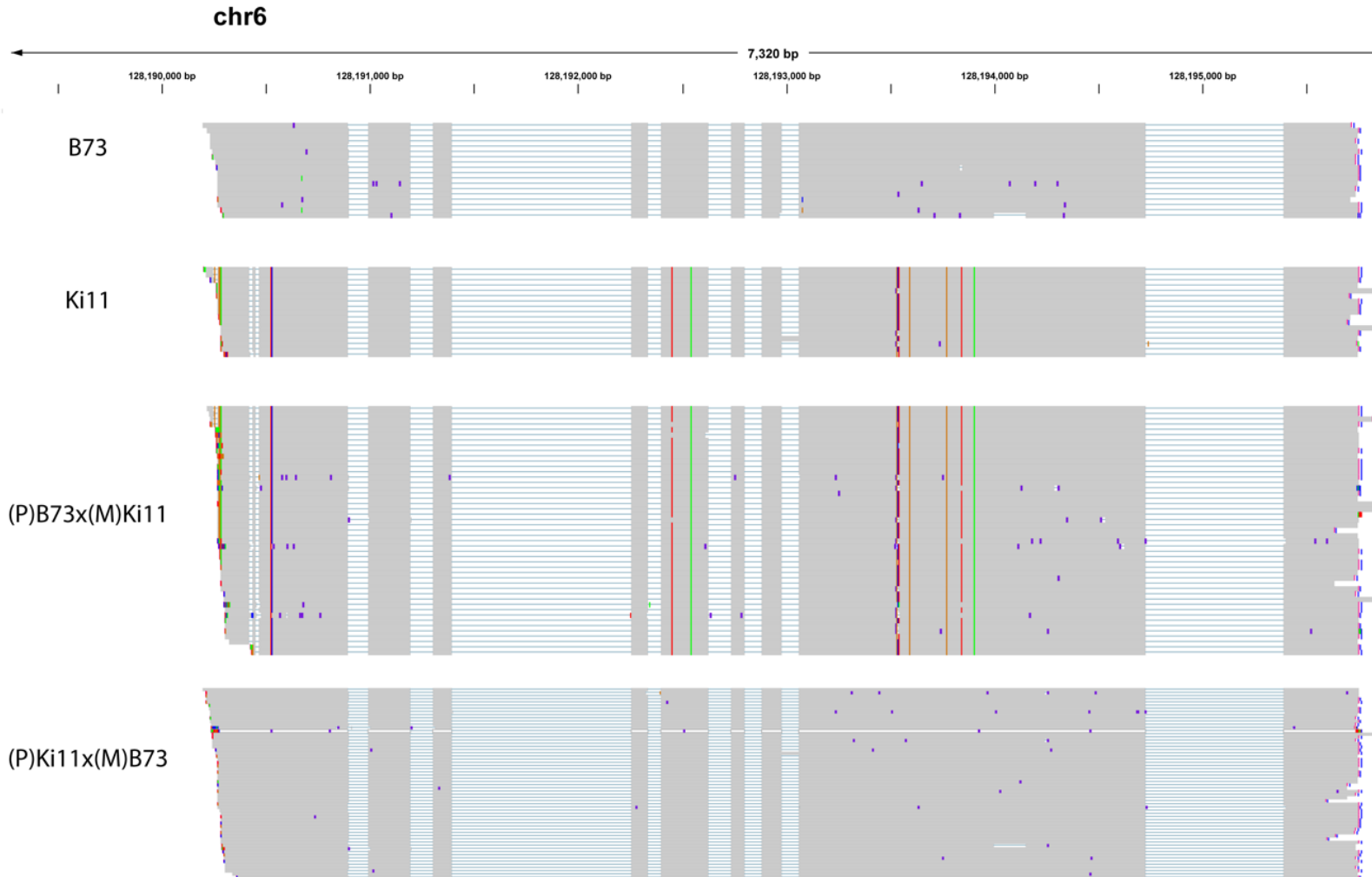
ALLELE-SPECIFIC EXPRESSION

- B73xKi11 only expresses the Ki11 (female) allele
- Ki11xB73 only expresses the B73 (female) allele
- Short read data confirms this allele-specific expression

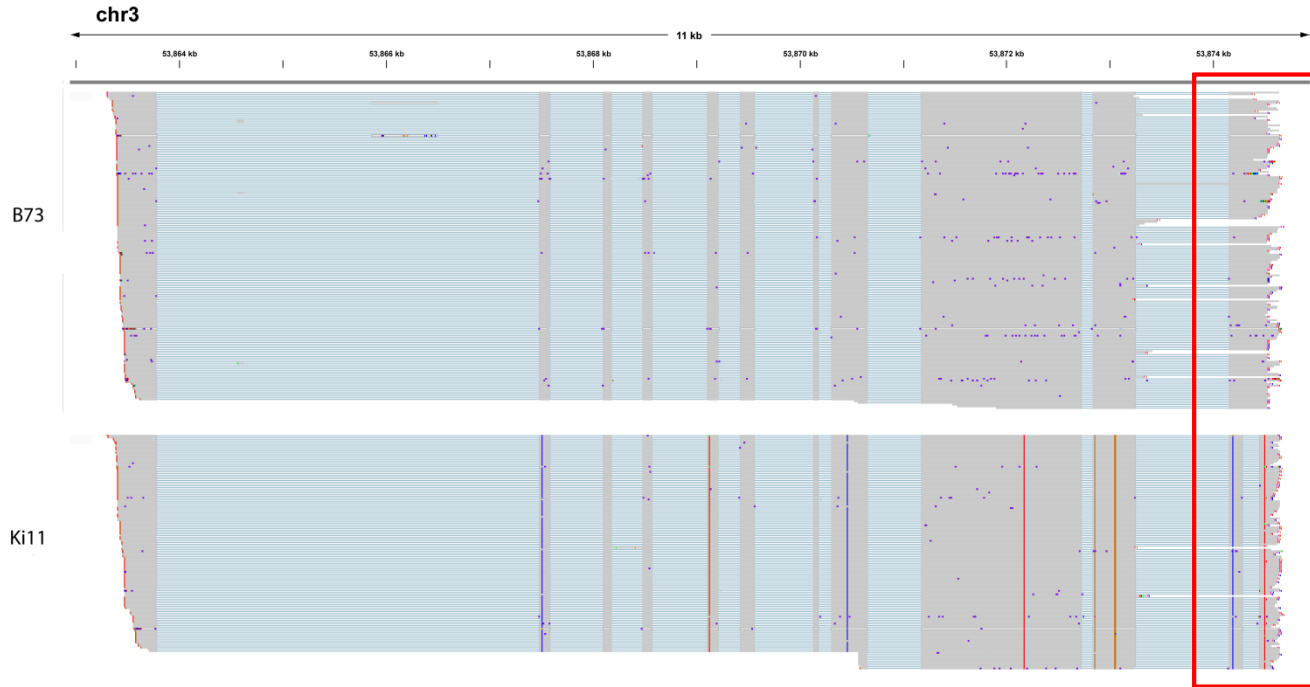


ALLELE-SPECIFIC EXPRESSION

- B73xKi11 only expresses the Ki11 (female) allele
- Ki11xB73 only expresses the B73 (female) allele
- Short read data confirms this allele-specific expression



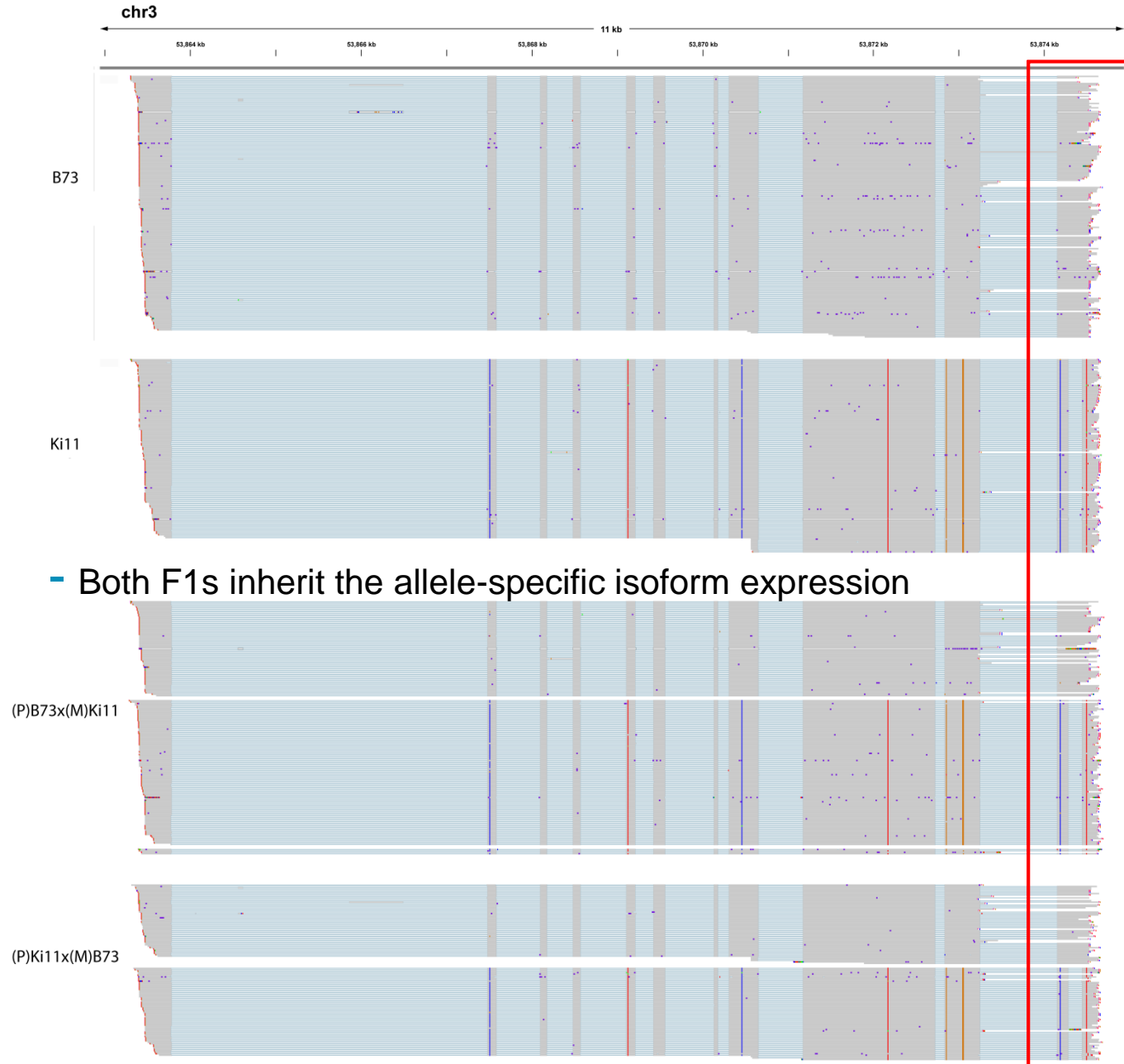
ALLELE-SPECIFIC ISOFORM EXPRESSION



- Two parents express different isoforms (3' exon difference)



ALLELE-SPECIFIC ISOFORM EXPRESSION



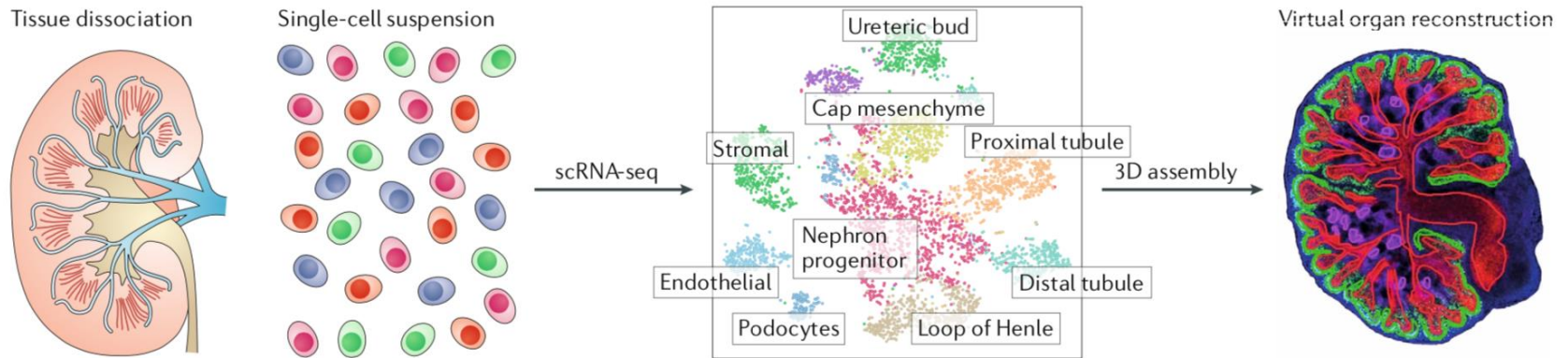
THE PACBIO ISO-SEQ METHOD CAN BE USED TO...

- Characterize full-length transcripts for whole transcriptome and targeted genes with multiplexing capacity.
- Identify novel isoforms
- Validate cancer fusion genes
- Discover allelic specific isoform expression (parental data not required)



Single Cell Iso-Seq

SINGLE CELL SEQUENCING REVEALS CELL HETEROGENEITY



PACBIO WORKS ACROSS SINGLE CELL PLATFORMS

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

Iain C Macaulay¹, Wilfried Haerty^{2,10}, Parveen Kumar^{3,10}, Yang I Li^{2,9}, Tim Xiaoming Hu², Mabel J Teng⁴, Mubeen Goolam⁵, Nathalie Saurat⁶, Paul Coupland⁷, Lesley M Shirley⁷, Miriam Smith⁷, Niels Van der Aa³, Ruby Banerjee⁸, Peter D Ellis⁷, Michael A Quail⁷, Harold P Swerdlow^{7,9}, Magdalena Zernicka-Goetz⁵, Frederick J Livesey⁶, Chris P Ponting^{1,2,11} & Thierry Voet^{1,3,11}



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 Karlsson¹ and Sten Linnarsson^{2*} 

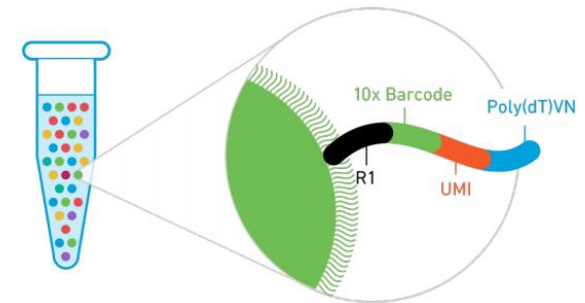
PACBIO WORKS ACROSS SINGLE CELL PLATFORMS

nature
biotechnology

LETTERS

Single-cell isoform RNA sequencing characterizes isoforms in thousands of cerebellar cells

Ishaan Gupta^{1,9}, Paul G Collier^{1,9}, Bettina Haase², Ahmed Mahfouz^{1,3,4}, Anoushka Joglekar¹, Taylor Floyd¹, Frank Koopmans⁵, Ben Barres^{6,8}, August B Smit⁵, Steven A Sloan⁶, Wenjie Luo⁷, Olivier Fedrigo², M Elizabeth Ross¹ & Hagen U Tilgner¹



Single-cell virus sequencing of influenza infections that trigger innate immunity

Alistair B. Russell¹, Jacob R. Kowalsky¹, Jesse D. Bloom^{1,2,3*}

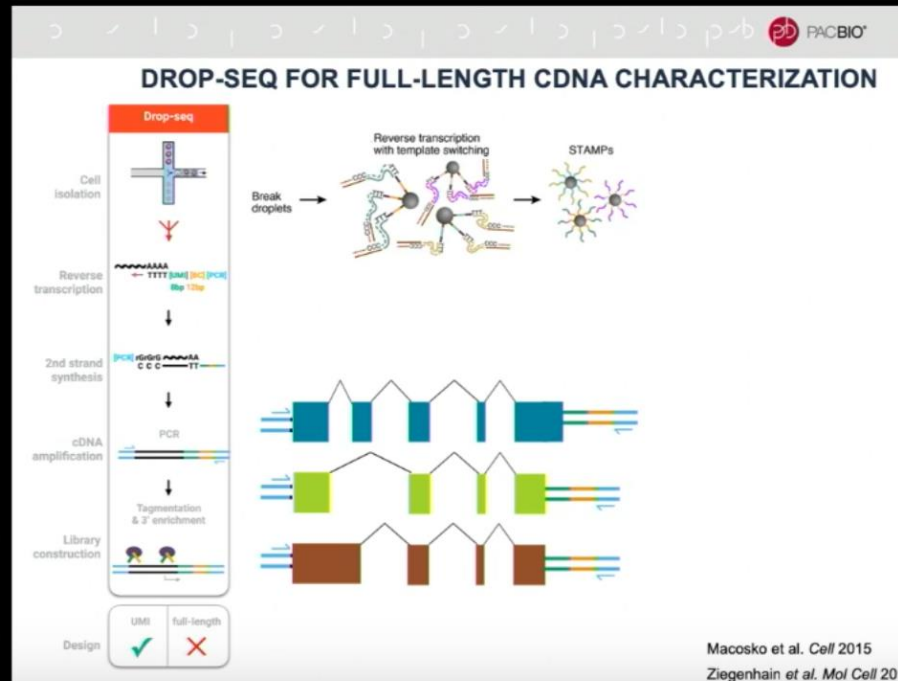
¹Basic Sciences Division and Computational Biology Program, Fred Hutchinson Cancer Research Center

²Department of Genome Sciences, University of Washington

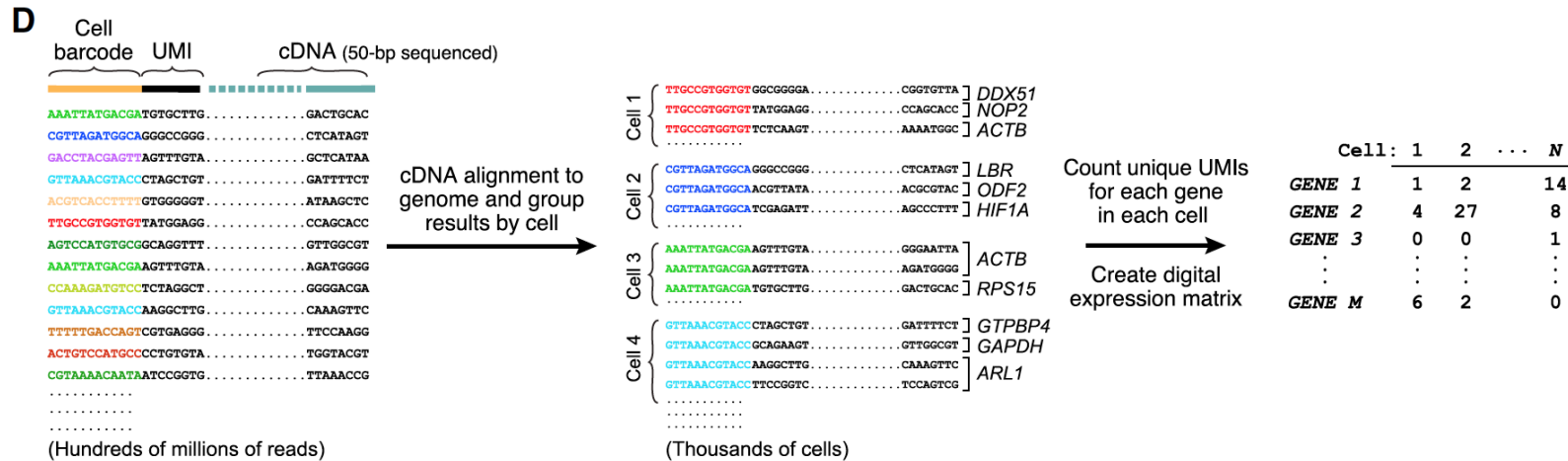
³Howard Hughes Medical Institute
Seattle, WA 98109, USA

*Lead contact for correspondence: jbloom@fredhutch.org

PACBIO WORKS ACROSS SINGLE CELL PLATFORMS



WHY LONG READS FOR SINGLE CELL SEQUENCING?



- Short reads only review gene-level information



PACBIO®

Single Cell Flu Viral Sequencing

Single-cell virus sequencing of influenza infections that trigger innate immunity

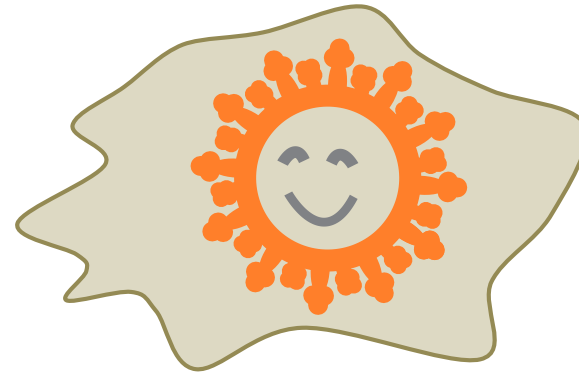
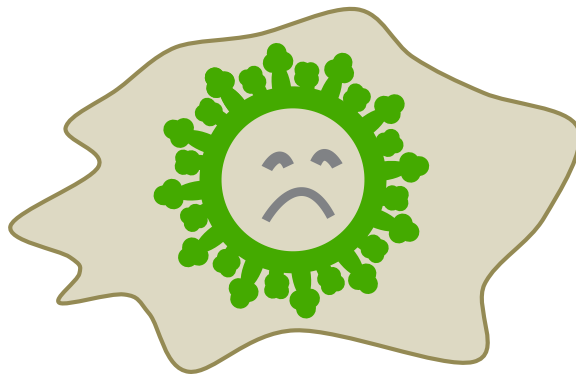
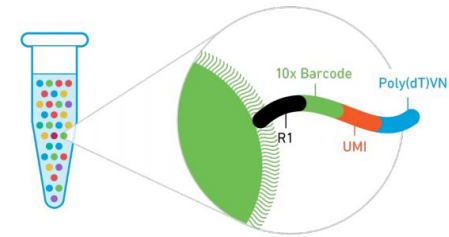
Alistair B. Russell¹, Jacob R. Kowalsky¹, Jesse D. Bloom^{1,2,3*}

¹Basic Sciences Division and Computational Biology Program, Fred Hutchinson Cancer Research Center

²Department of Genome Sciences, University of Washington

³Howard Hughes Medical Institute
Seattle, WA 98109, USA

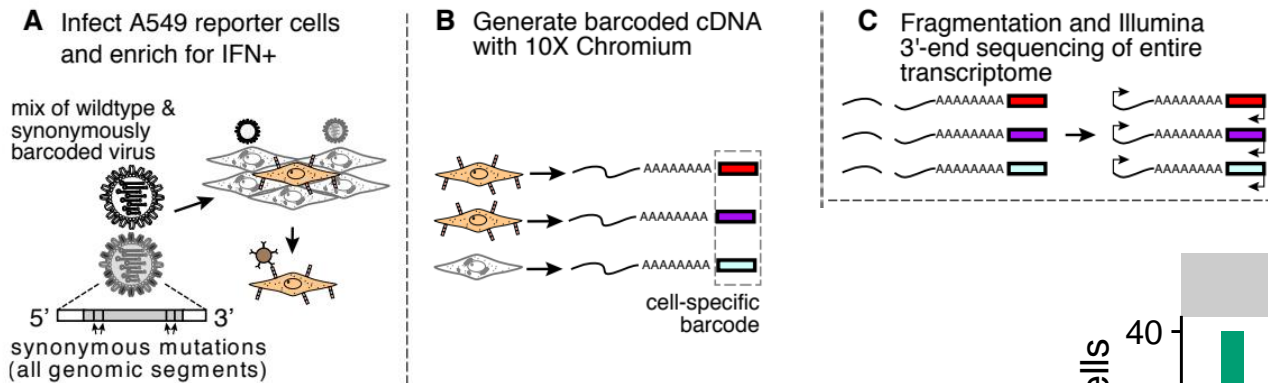
*Lead contact for correspondence: jbloom@fredhutch.org



<https://www.biorxiv.org/content/biorxiv/early/2018/12/28/437277.full.pdf>

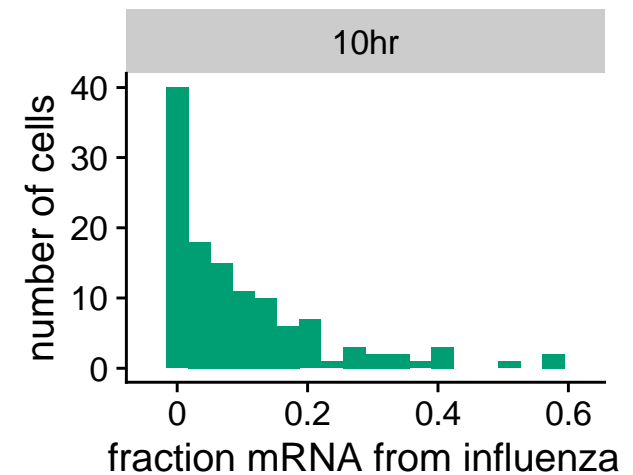
Russell, Alistair B., Jacob R. Kowalsky, and Jesse D. Bloom. 2018. "Single-Cell Virus Sequencing of Influenza Infections That Trigger Innate Immunity." *bioRxiv*. <https://doi.org/10.1101/437277>.

SINGLE CELL DATA REVEALED EXTREME VARIATION IN VIRAL RESPONSE



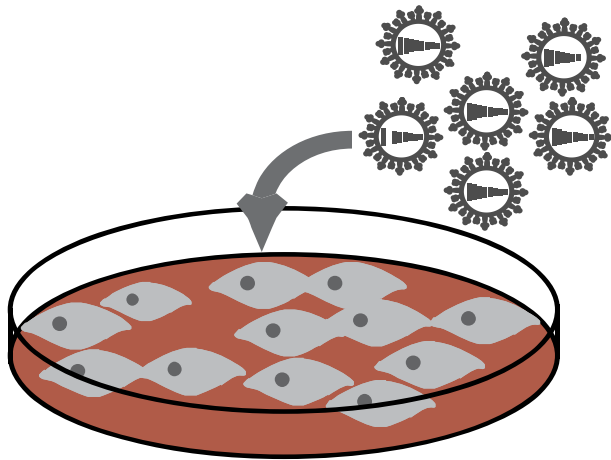
Variation in viral response was NOT explained by

- Initial viral dosage
- Gene expression differences
- Immune response (IFN expression)

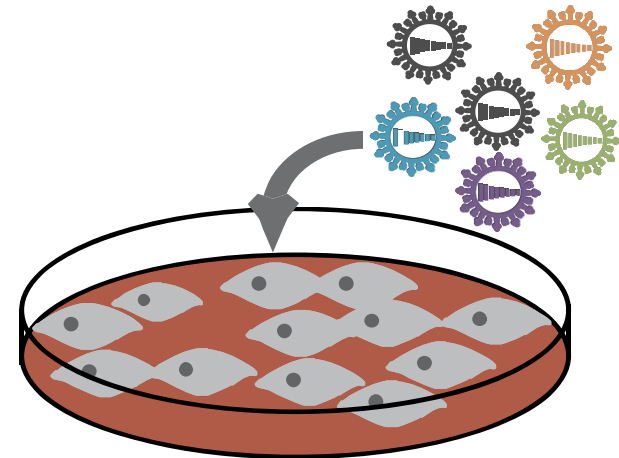


HYPOTHESIS: MUTATIONS IN THE "WILDTYPE" VIRUS IS RESPONSIBLE FOR THE VARIATION

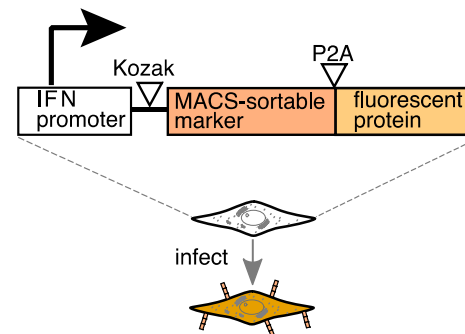
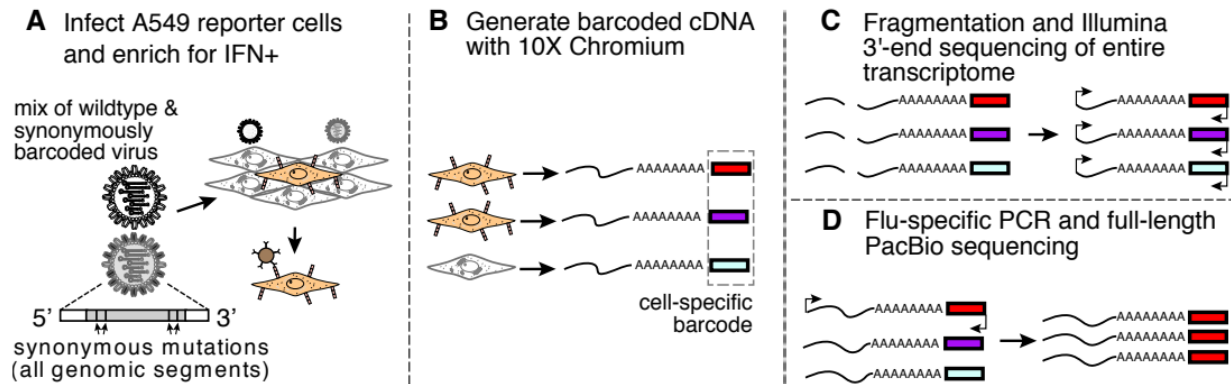
IDEAL



REALITY



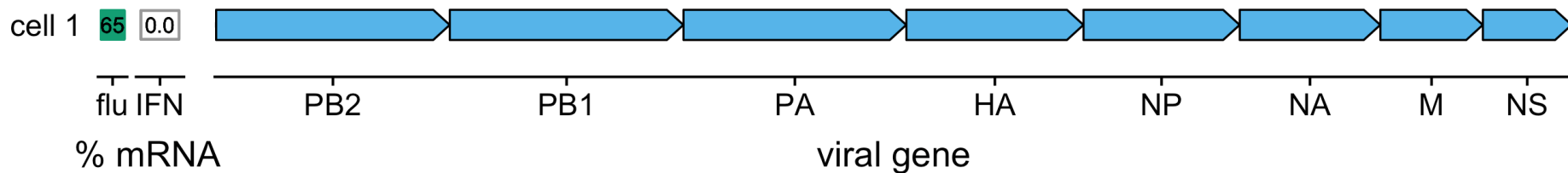
PACBIO FOR FULL-LENGTH VIRAL SEQUENCING



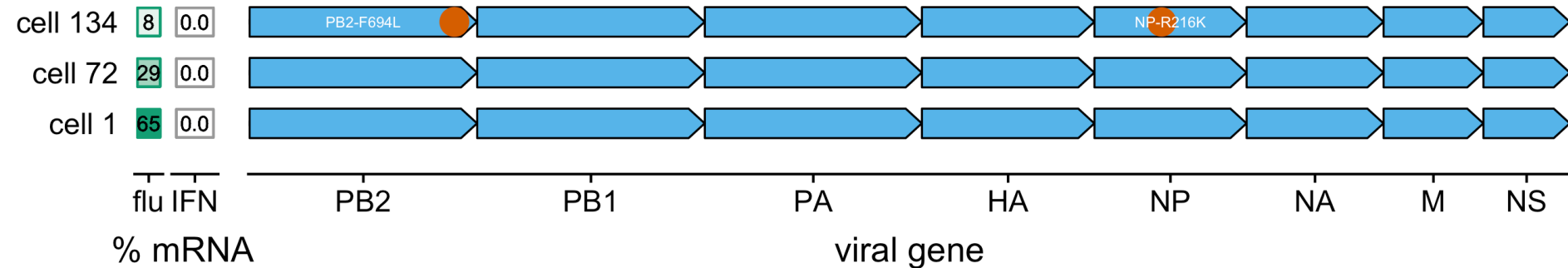
<https://www.biorxiv.org/content/biorxiv/early/2018/12/28/437277.full.pdf>

Russell, Alistair B., Jacob R. Kowalsky, and Jesse D. Bloom. 2018. "Single-Cell Virus Sequencing of Influenza Infections That Trigger Innate Immunity." *bioRxiv*. <https://doi.org/10.1101/437277>.

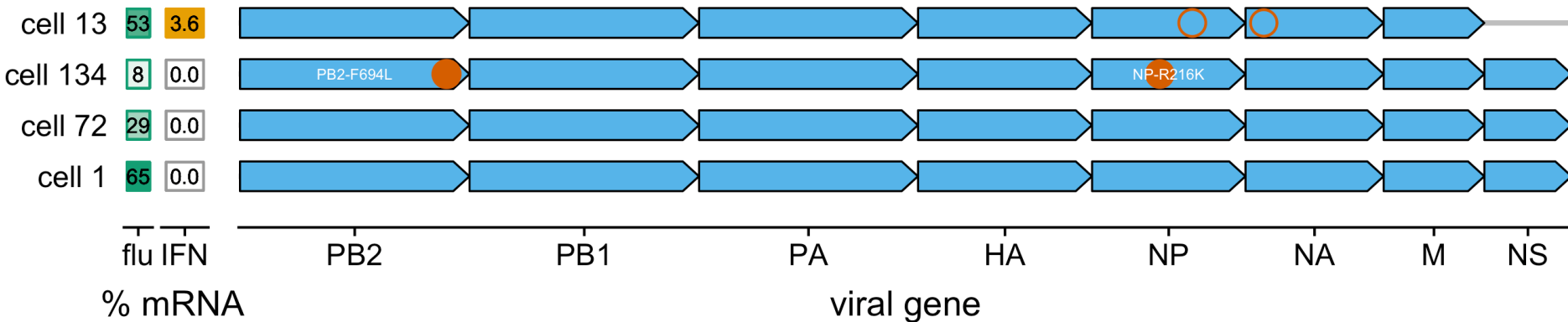
Cells infected by wildtype virions often produce lots of viral mRNA



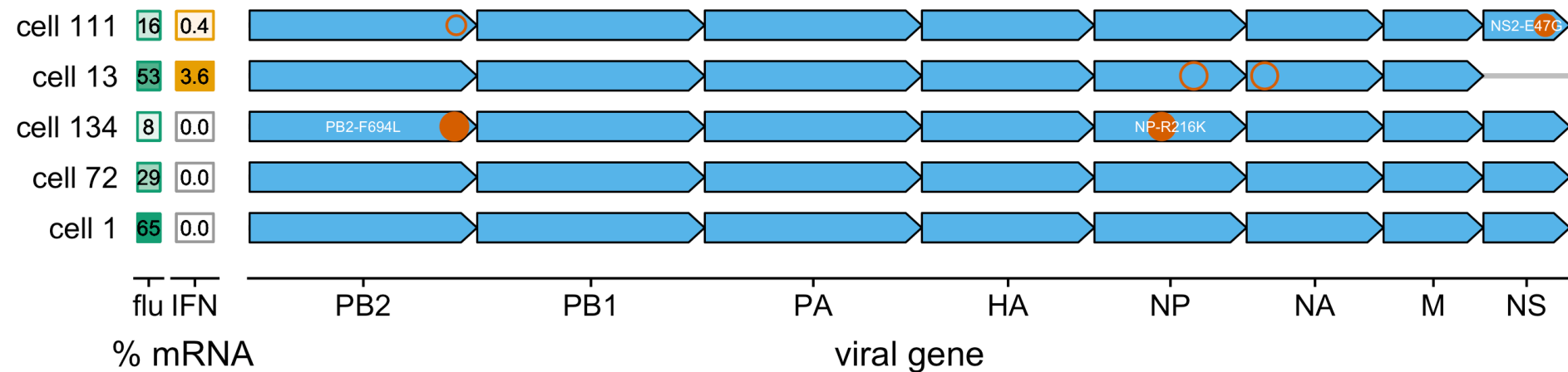
Virions with mutations sometimes produce little viral mRNA



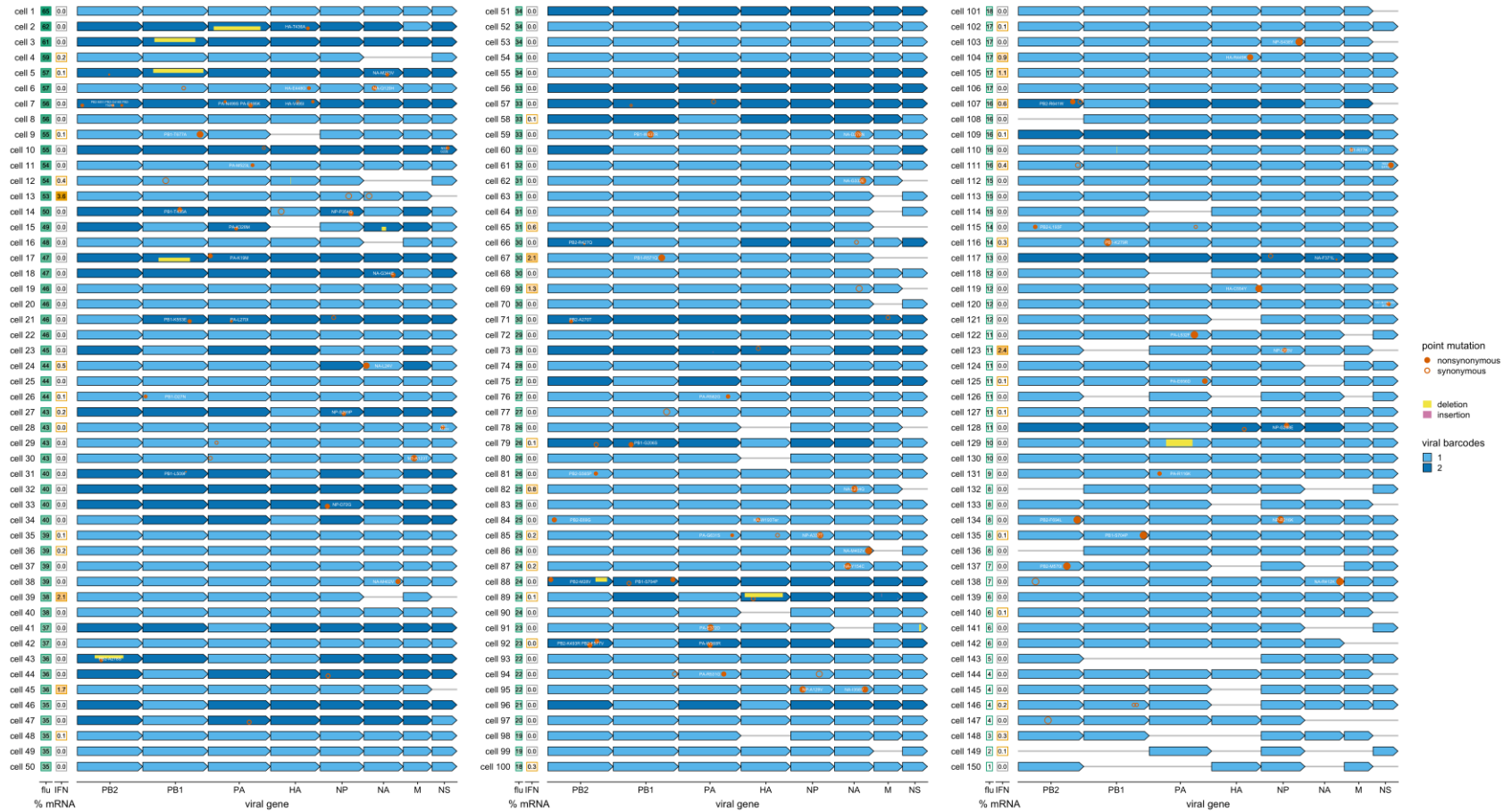
Virions with defects sometimes produce IFN: fails to express NS

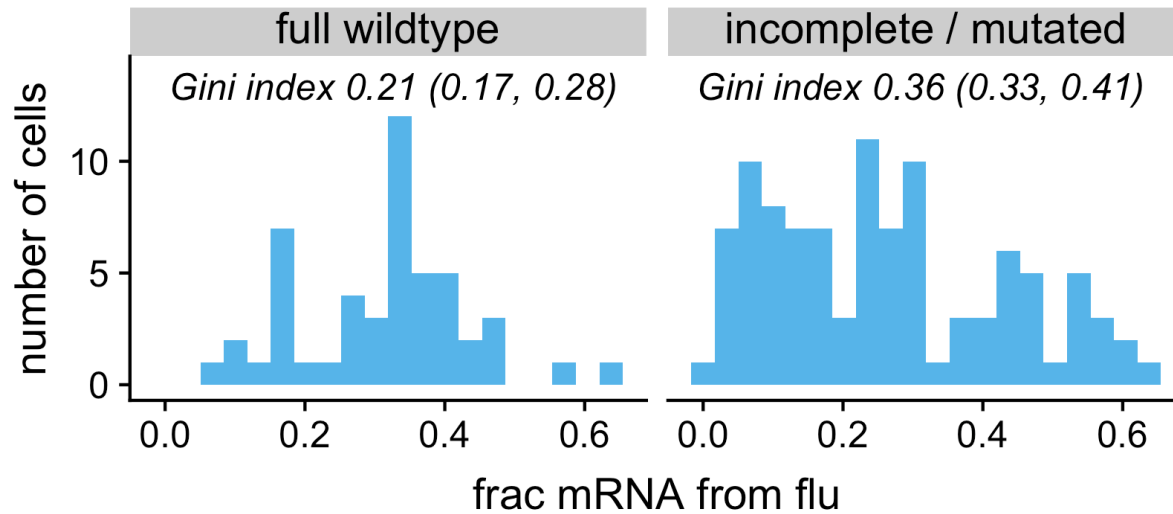


Virions with defects sometimes produce IFN: point mutation in NS

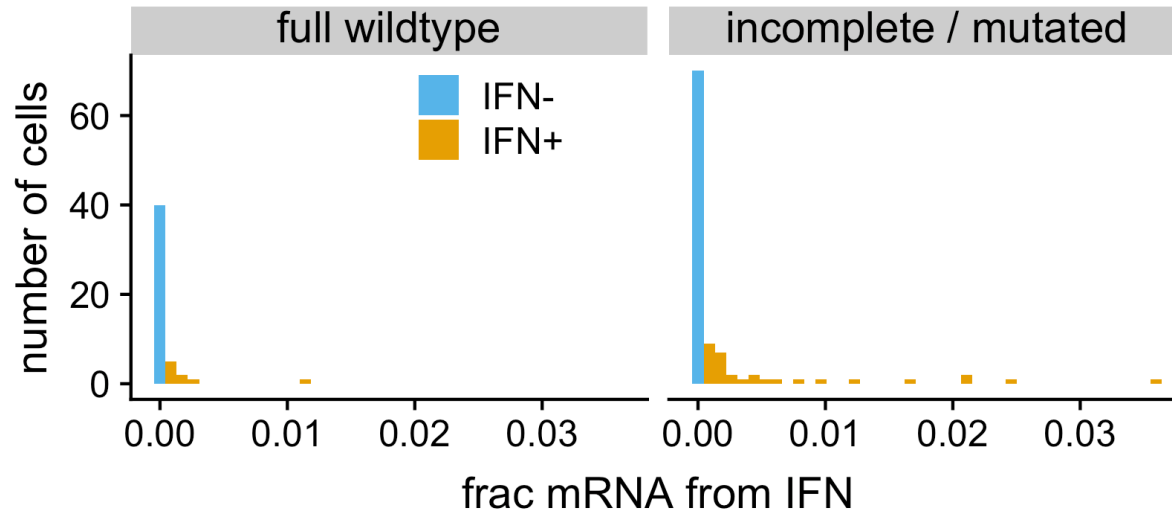


150 cells infected, only 49 by wildtype virions





Viral genetic variation is a significant contributor to heterogeneity in viral burden



Viral mutations contribute to IFN induction



Single Cell Iso-Seq on the Sequel II platform

SINGLE CELL ISO-SEQ FOR HUMAN ORGANOIDS




Single cell isoform sequencing (scIso-Seq) identifies novel full-length mRNAs and cell type-specific expression

Jason G. Underwood, Ph.D
Principal Scientist, PacBio

Genome Sciences
UNIVERSITY OF WASHINGTON

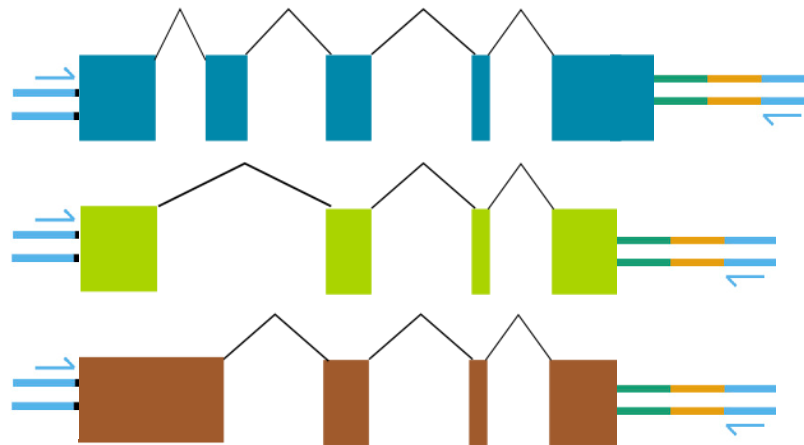
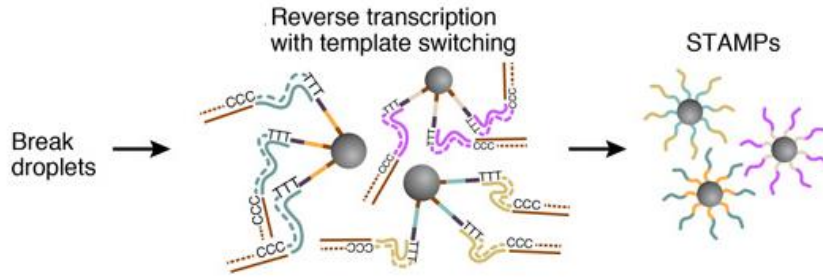
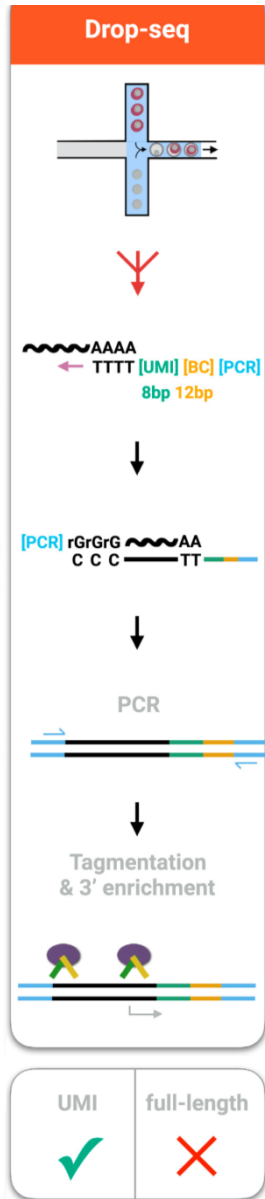
hhmi

UCSF



AGBT™

DROP-SEQ FOR FULL-LENGTH CDNA CHARACTERIZATION



Post PCR

PacBio
Iso-Seq



Shear

3' RNA-Seq



CEREBRAL ORGANOIDS AS A MODEL FOR PRIMATE BRAIN DEVELOPMENT



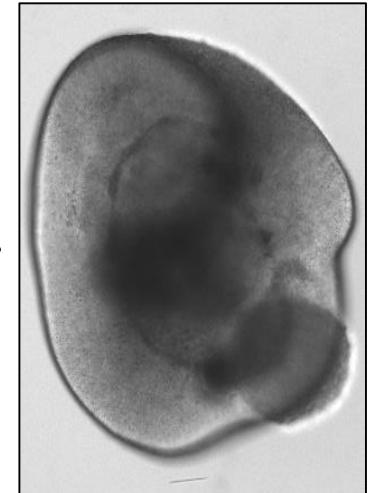
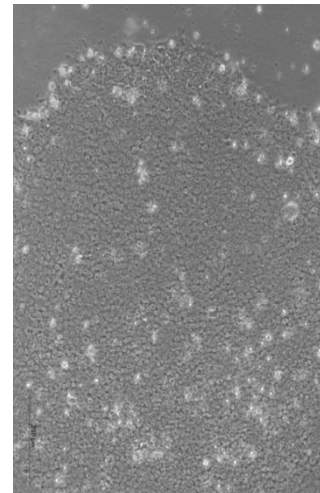
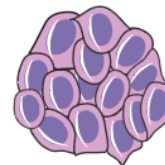
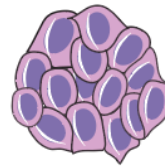
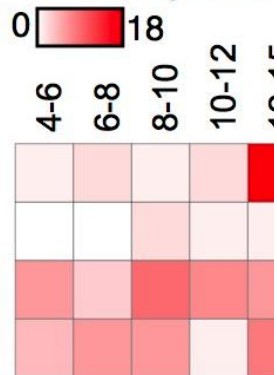
PCW 9 - 17

Available
iPS Lines

Replicates

Samples

Individuals Samples



Astrocytes
Neurons
Radial Glia
Cycling (G2M)

Cellular Composition



Kronenberg *et al.* *Science* 2018



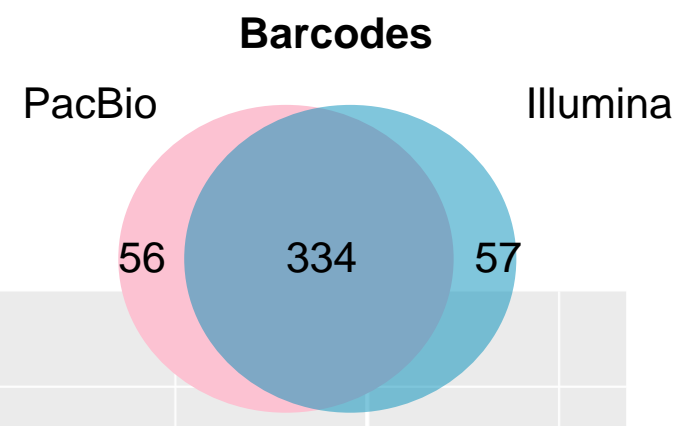
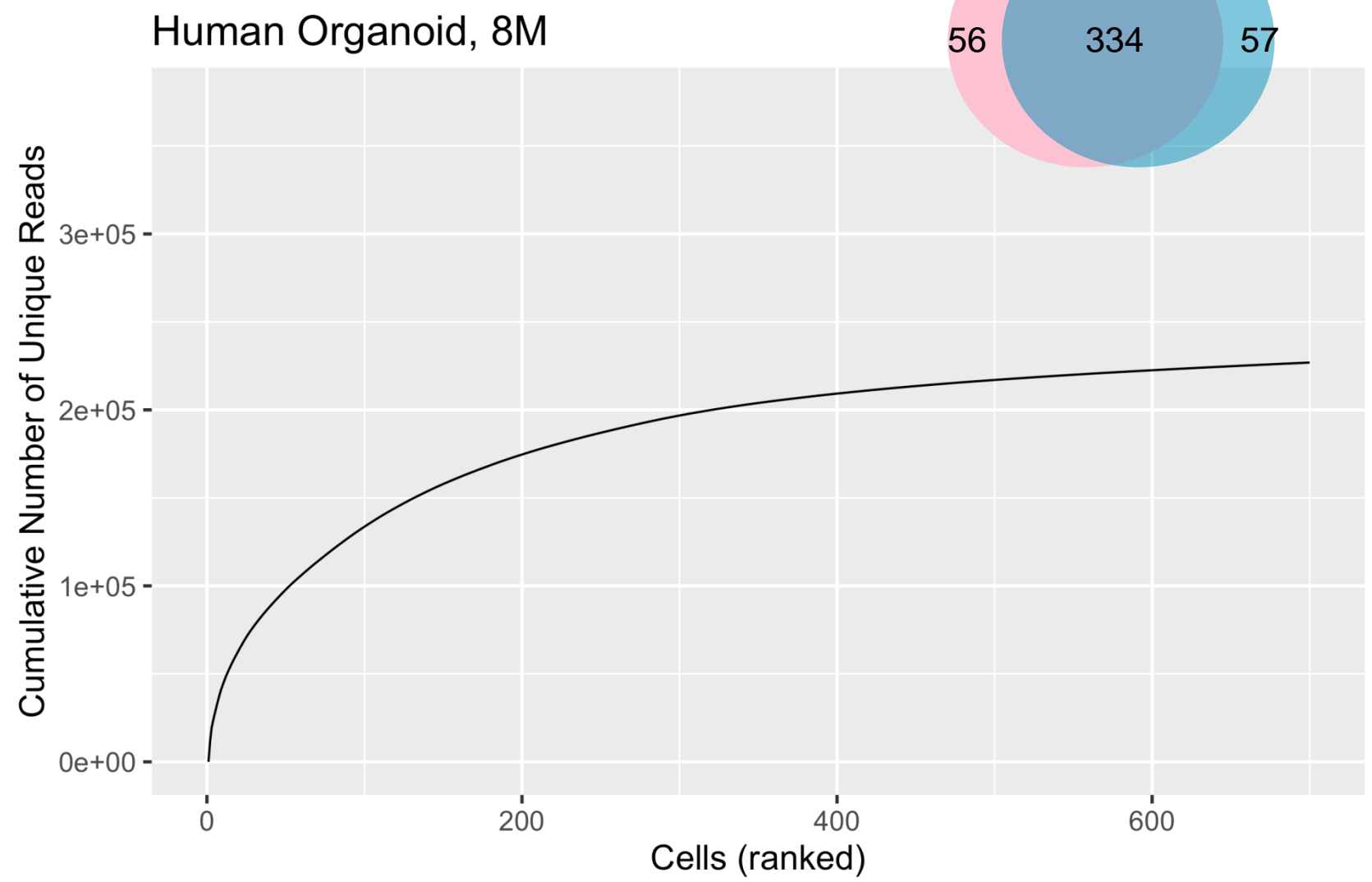
Pollen *et al.* *Cell* 2019

SINGLE CELL ISO-SEQ ON THE SEQUEL II PLATFORM

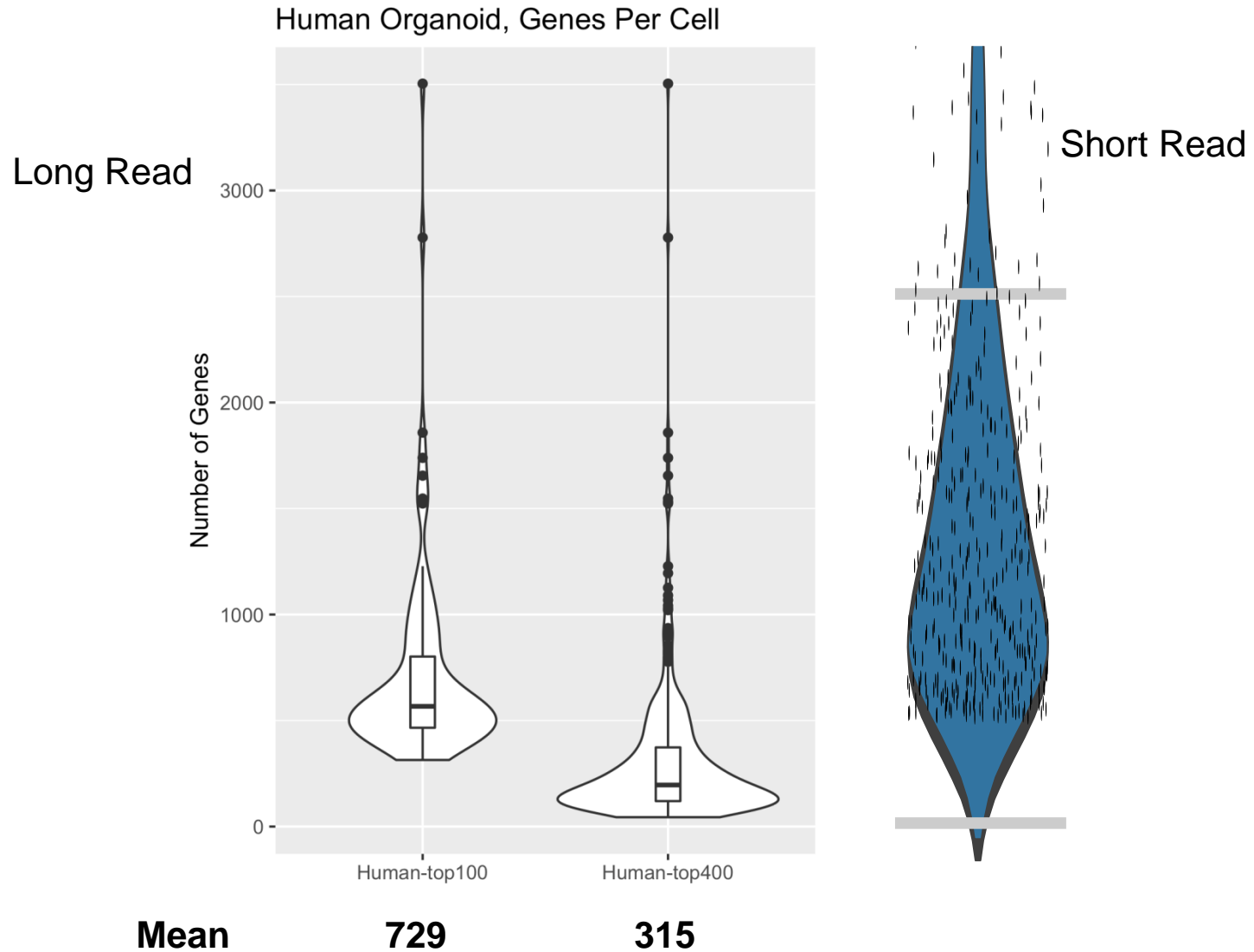
*Mapped to hg38. Compared to Gencode v29.

SAMPLE	FLNC (post-filter)	UNIQUE READS	UNIQUE GENES*	UNIQUE ISOFORMS*
Chimp Organoid	2,303,267	418,542	14,049	58,892
Human Organoid	2,291,947	382,734	14,737	60,815

CUMULATIVE READ PLOT



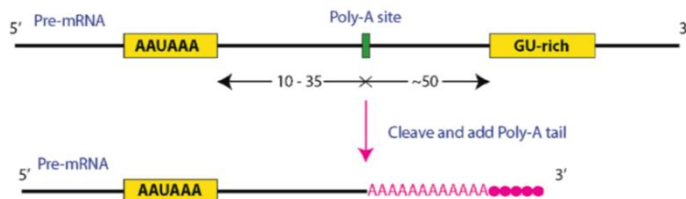
GENES DETECTED PER CELL: HUMAN



POLYA SIGNALS IN SINGLE CELL ISO-SEQ

PacBio Human Organoid Single Cell

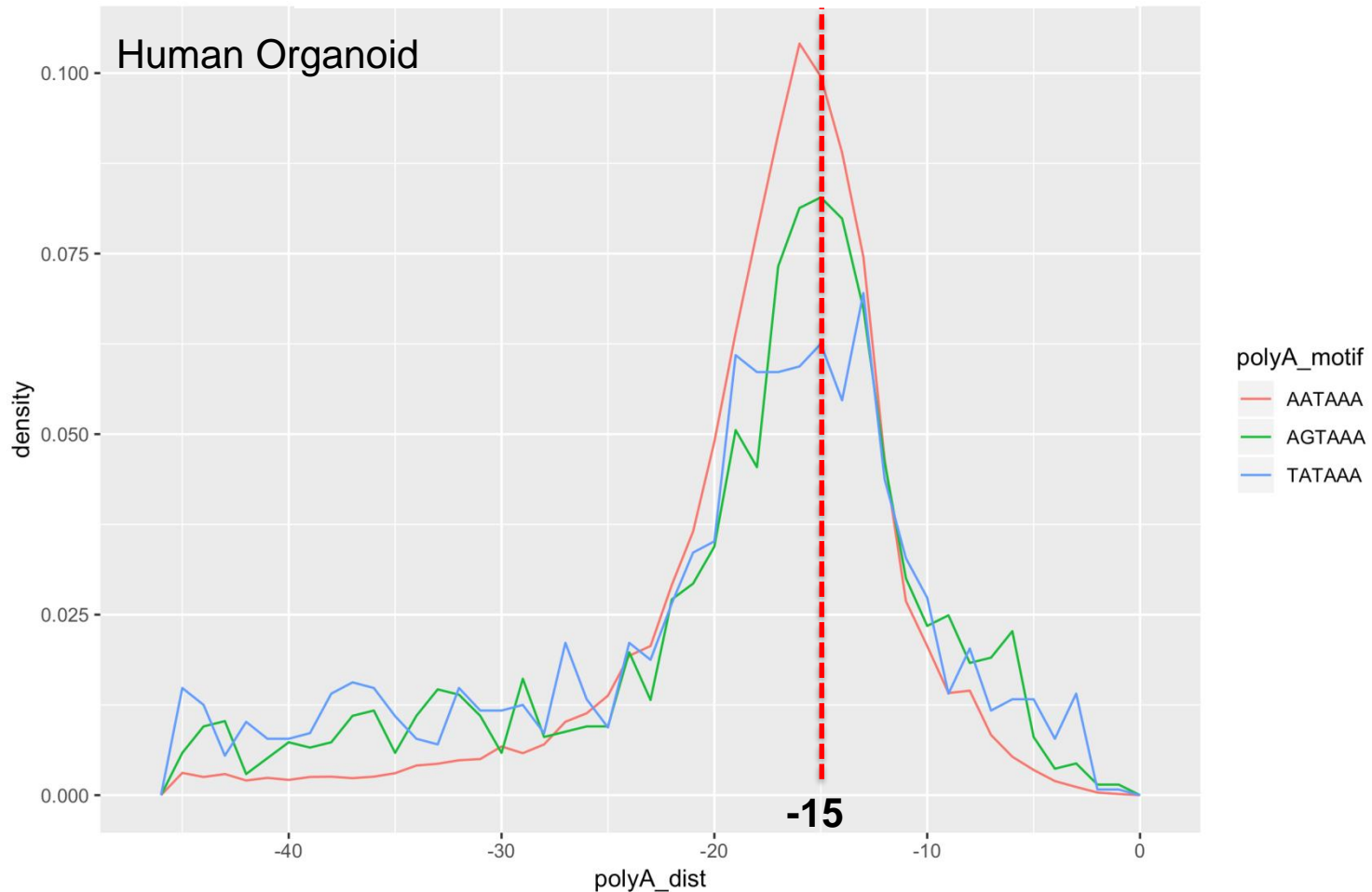
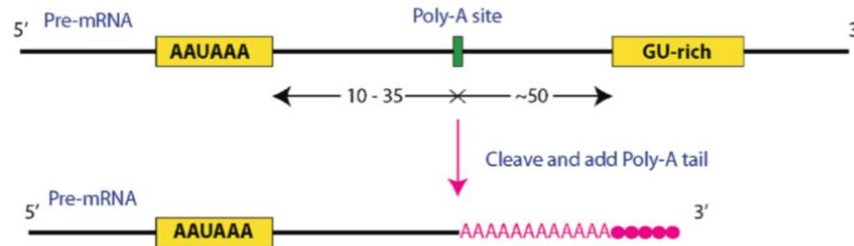
	# of Obs	% of Obs
AAUAAA	24,596	59.3%
AUUAAA	6,606	15.9%
AGUAAA	1,365	3.3%
UAUAAA	1,280	3.1%
AAGAAA	1,028	2.4%
AAAAAG	932	2.2%
...		



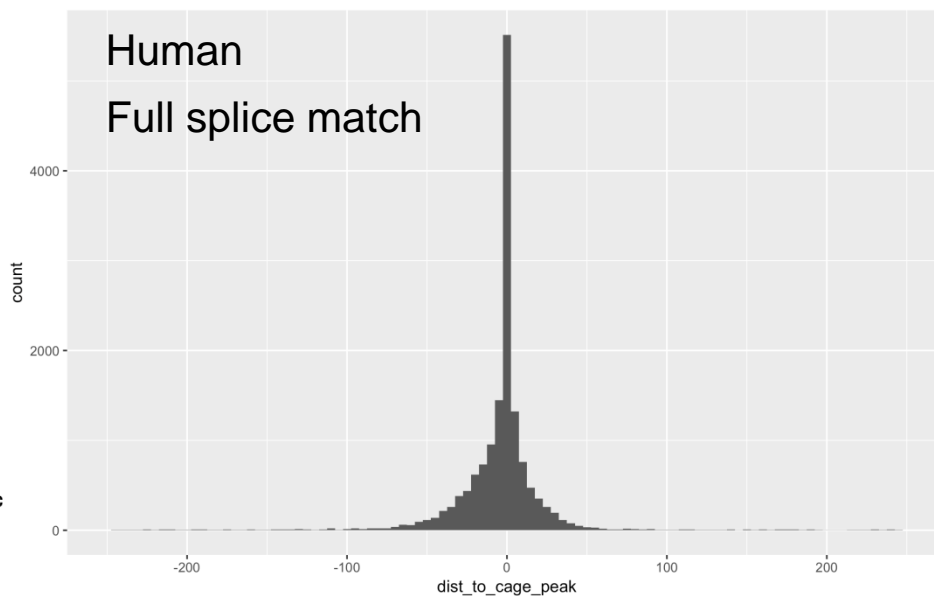
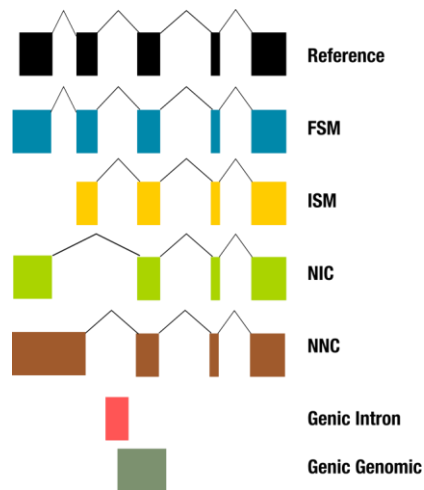
Hexamer	Observed (expected) ^a	% sites	p^b	Position average \pm SD	Location ^c
AAUAAA	3286 (317)	58.2	0	-16 ± 4.7	
AUUAAA	843 (112)	14.9	0	-17 ± 5.3	
AGUAAA	156 (32)	2.7	6×10^{-57}	-16 ± 5.9	
UAUAAA	180 (53)	3.2	4×10^{-45}	-18 ± 7.8	
CAUAAA	76 (23)	1.3	1×10^{-18}	-17 ± 5.9	
GAUAAA	72 (21)	1.3	2×10^{-18}	-18 ± 6.9	
AAUUA	96 (33)	1.7	2×10^{-19}	-18 ± 6.9	
AAUACA	70 (16)	1.2	5×10^{-23}	-18 ± 8.7	
AAUAGA	43 (14)	0.7	1×10^{-9}	-18 ± 6.3	
AAAAAG	49 (11)	0.8	5×10^{-17}	-18 ± 8.9	

pA signals in humans: Beaudoin *et al.* 2000

POLYA SIGNAL LOCATION IN SINGLE CELL ISO-SEQ

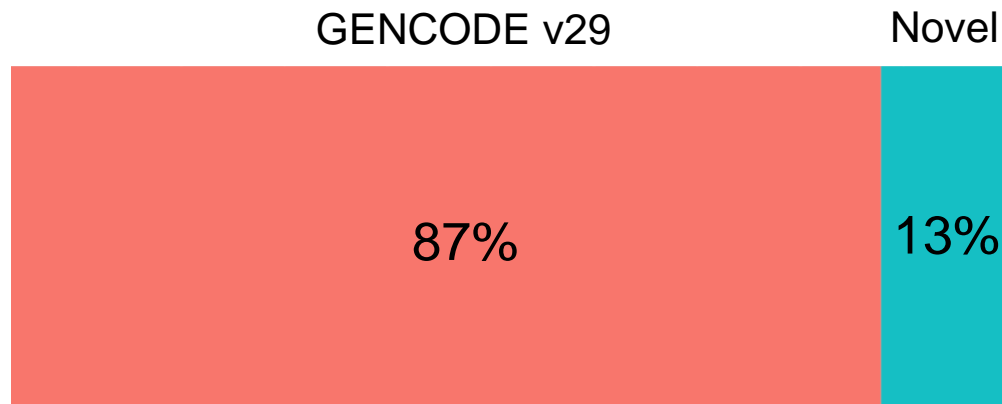


MATCHING TSS IN SINGLE CELL ISO-SEQ TO CAGE DATA

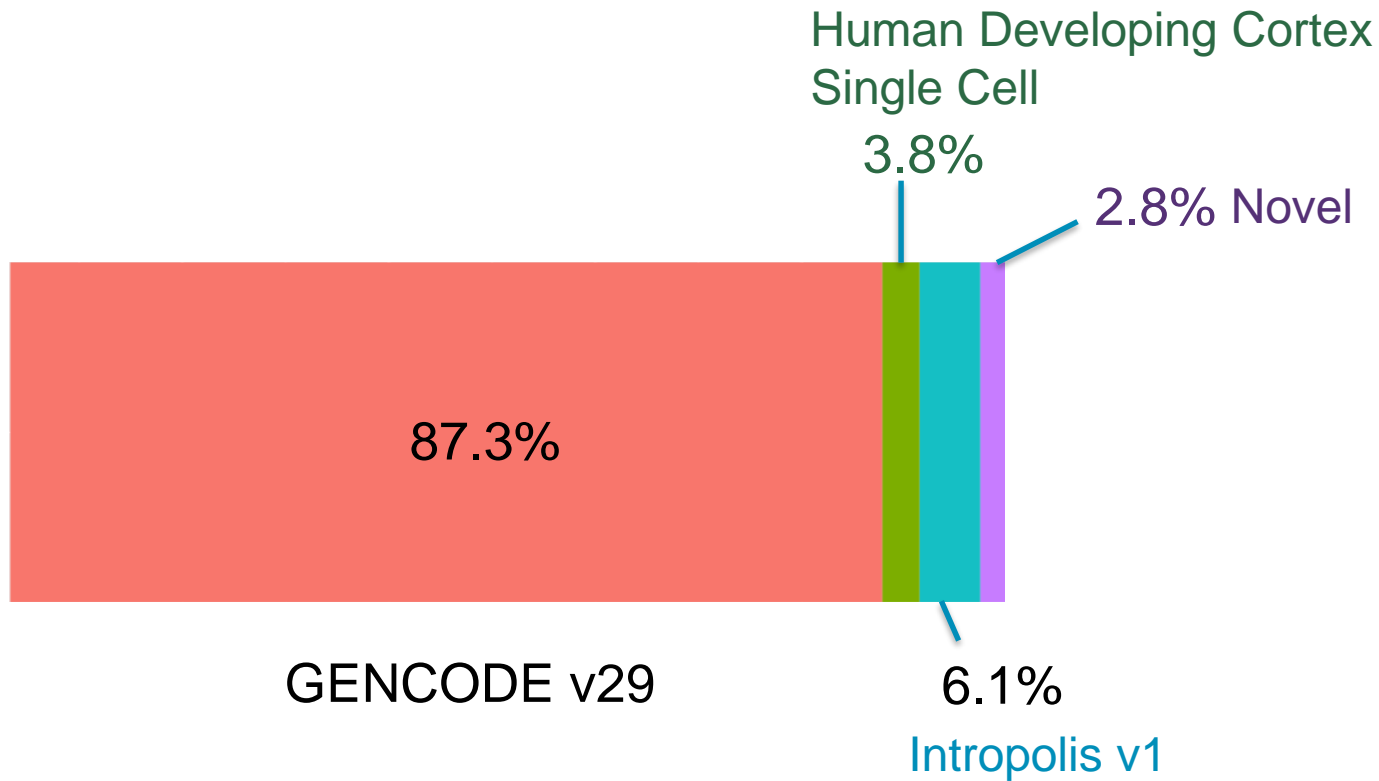


Class	# Isoforms	% with CAGE Peak ≤ 50 bp
Full Splice Match	18,344	78%
Incomplete Splice Match	13,802	37%
Novel In Catalog	19,033	44%
Novel Not In Catalog	9,197	67%
Intergenic	245	29%

SPLICE JUNCTION VALIDATION



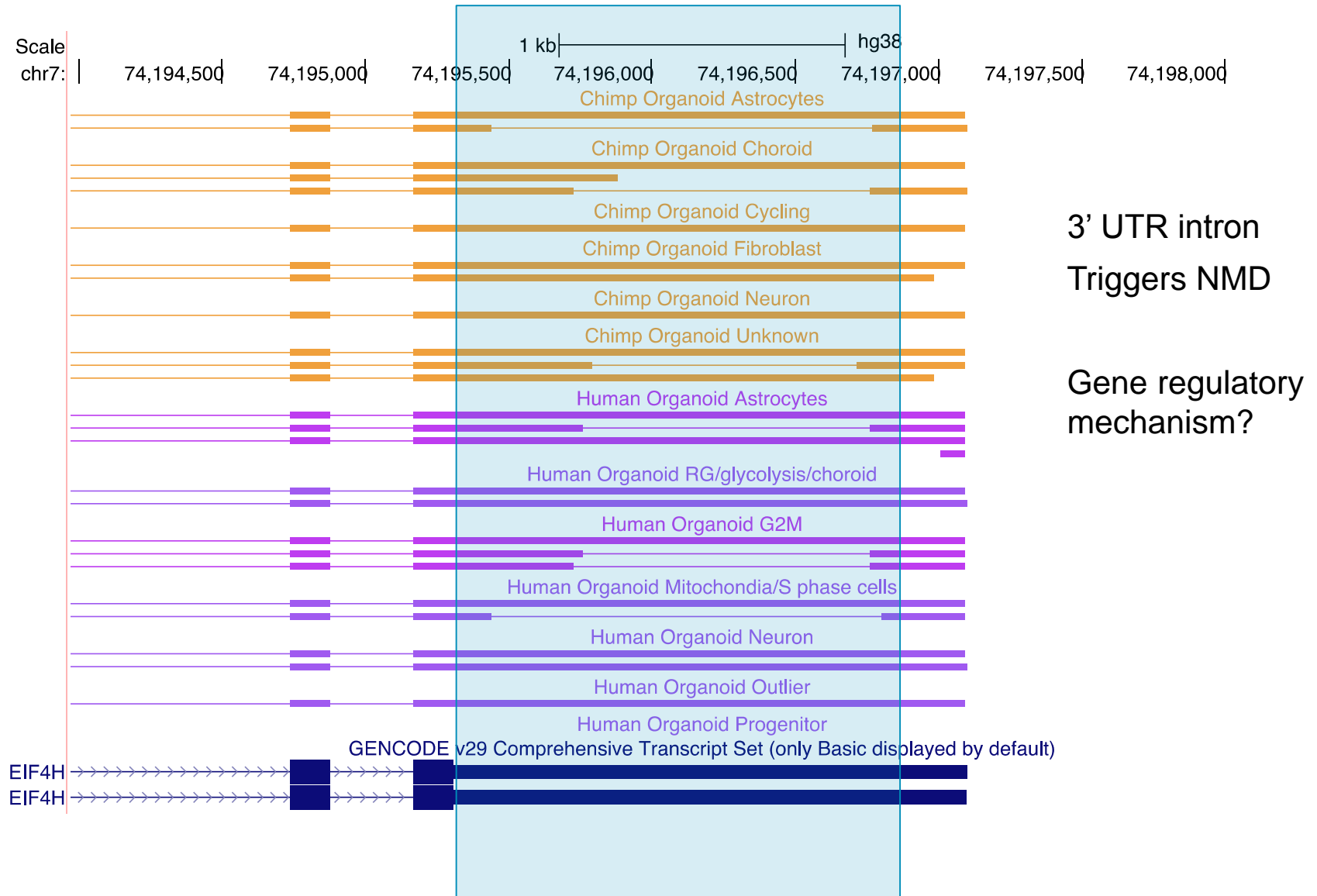
SPLICE JUNCTION VALIDATION



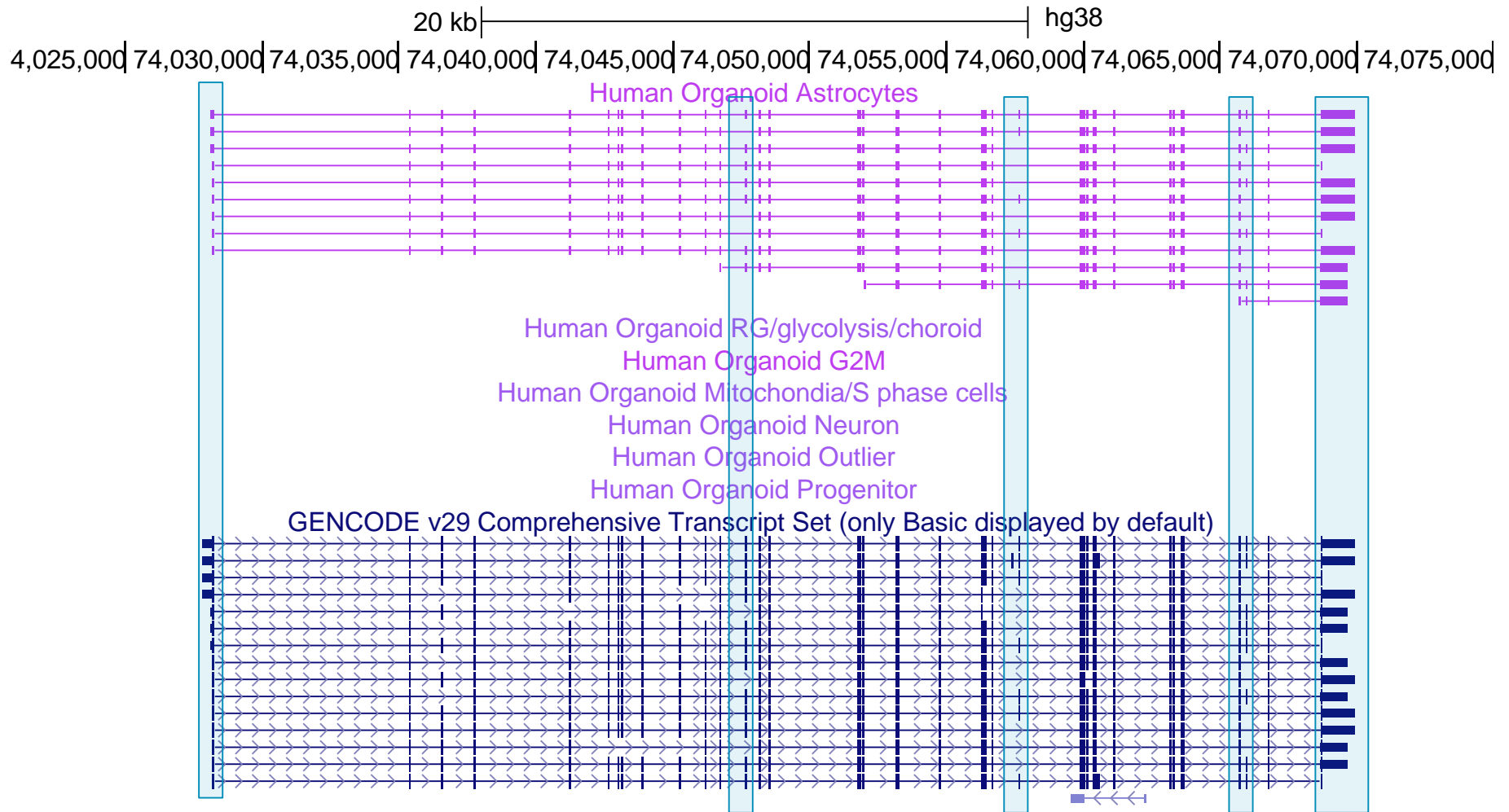
DevCortex- ~4000 cells C1: Nowakowski *et al. Science* 2017

21,500 RNAseq datasets: Intropolis: Nellore *et al. Genome Biol.* 2016

POST-TRANSCRIPTIONAL GENE REGULATION ON DISPLAY



MULTIPLE TROPOELASTIN ISOFORMS ARE EXPRESSED IN ASTROCYTES



SINGLE CELL ISO-SEQ

- Single Cell Iso-Seq delivers full-length transcript information
- Highly accurate CCS (HiFi) reads yield accurate cell barcode and UMI information that can be matched back to short read data



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.