

A Complete Solution for Full-Length Transcript Sequencing Using the PacBio Sequel II System

Jason G. Underwood, Brendan Galvin, Ting Hon and Elizabeth Tseng

PacBio, 1305 O'Brien Drive, Menlo Park, CA 94025

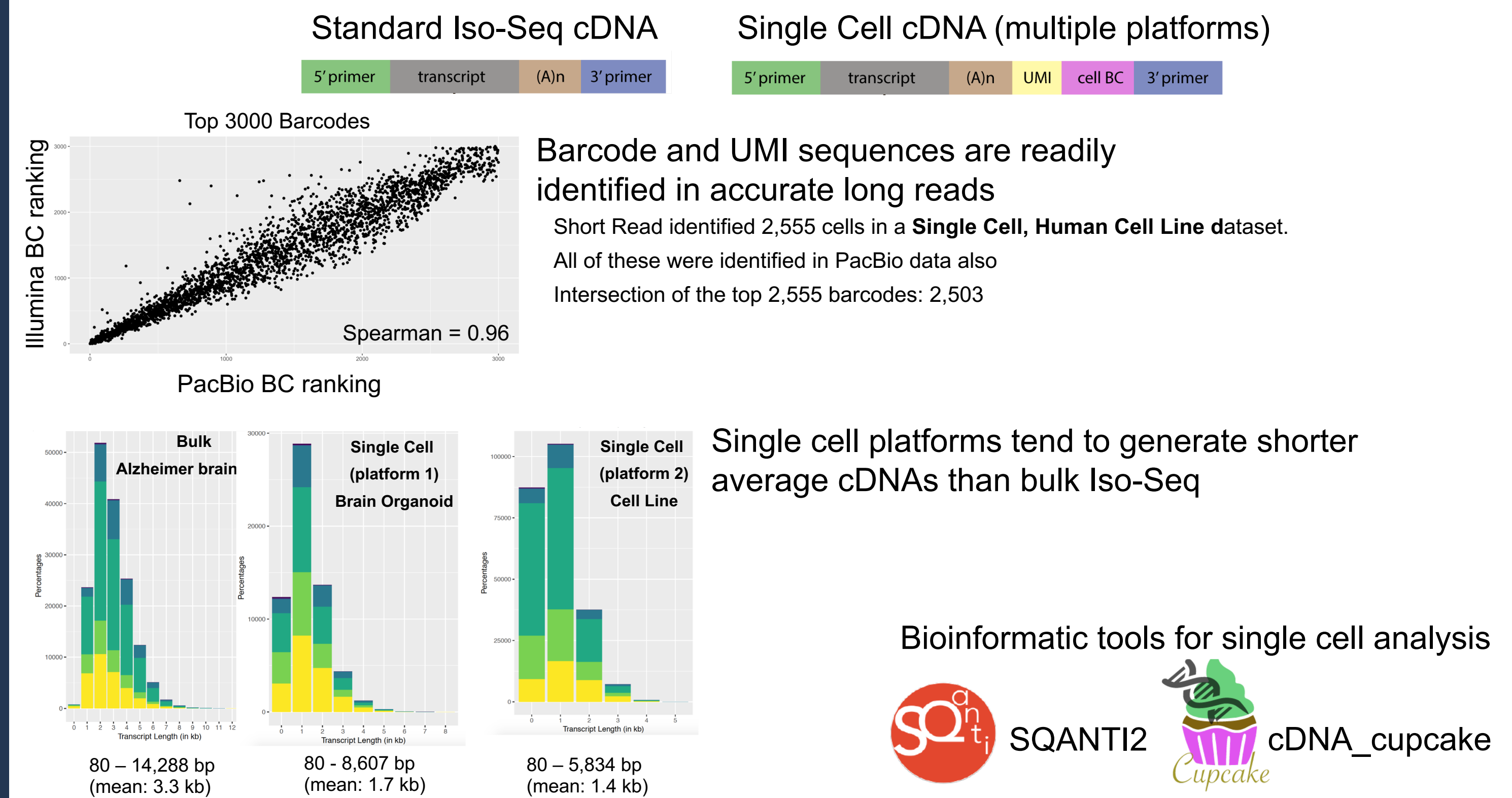


**RNA
SEQUENCING**

Iso-Seq on the Sequel II System

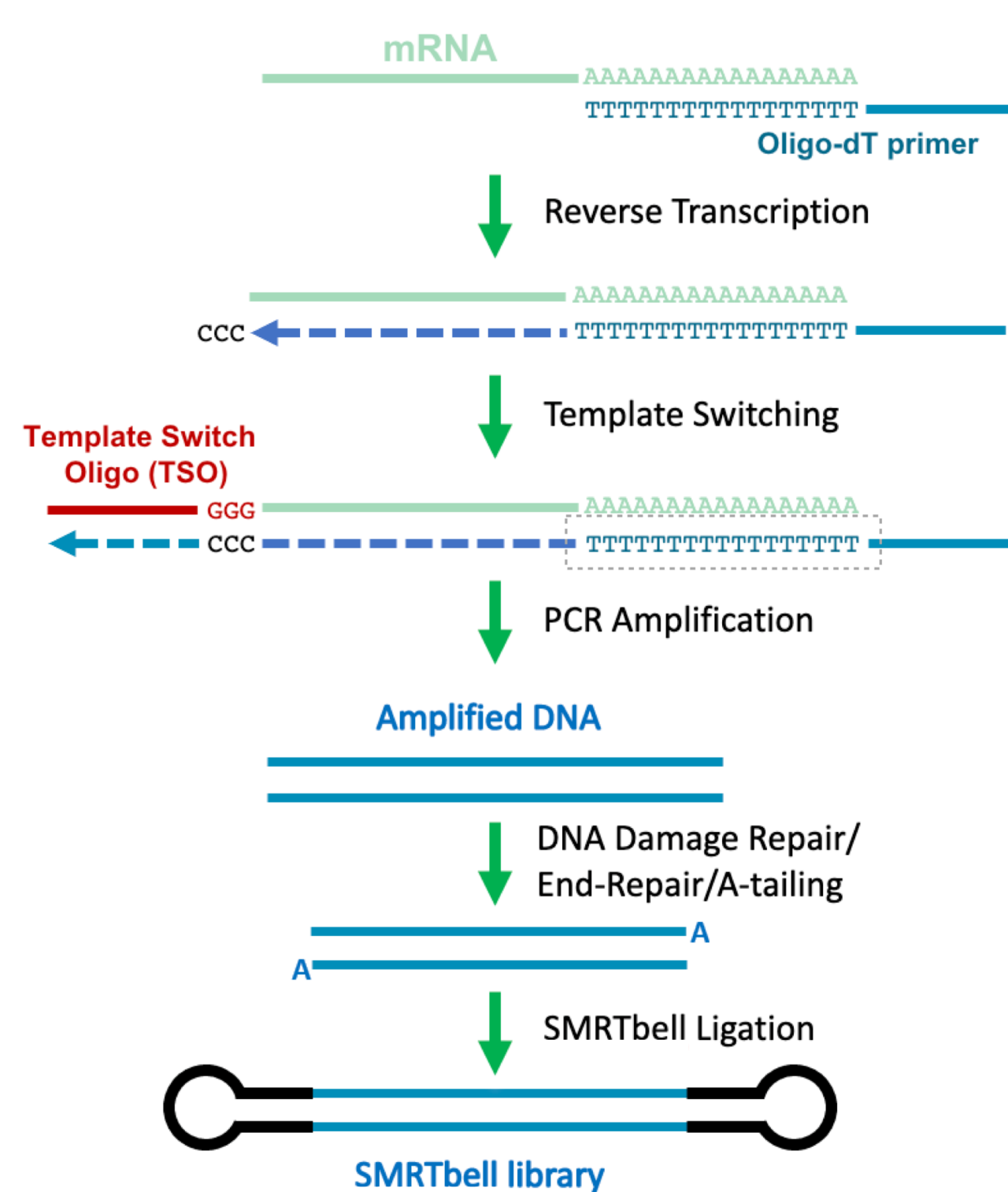
- Generate full-length transcript sequences up to ~15kb
- High accuracy (>99%) for ORF prediction
- No reference genome required
- Bioinformatics tools from raw data to functional annotation
- Many applications, including: ^[1]
 - Genome annotation
 - Novel gene and isoform discovery
 - Fusion gene detection
 - Allele-specific isoform expression analysis
 - Improving a reference for RNA-seq quantification
 - Assess genome assembly quality
 - Single cell analysis

Single Cell Iso-Seq



Iso-Seq Express Kit ^[2]

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



**LIBRARY
PREP
1 DAY**

**SMRT
SEQUENCING
1 DAY**






**DATA
ANALYSIS
1 DAY**

Sequel II System

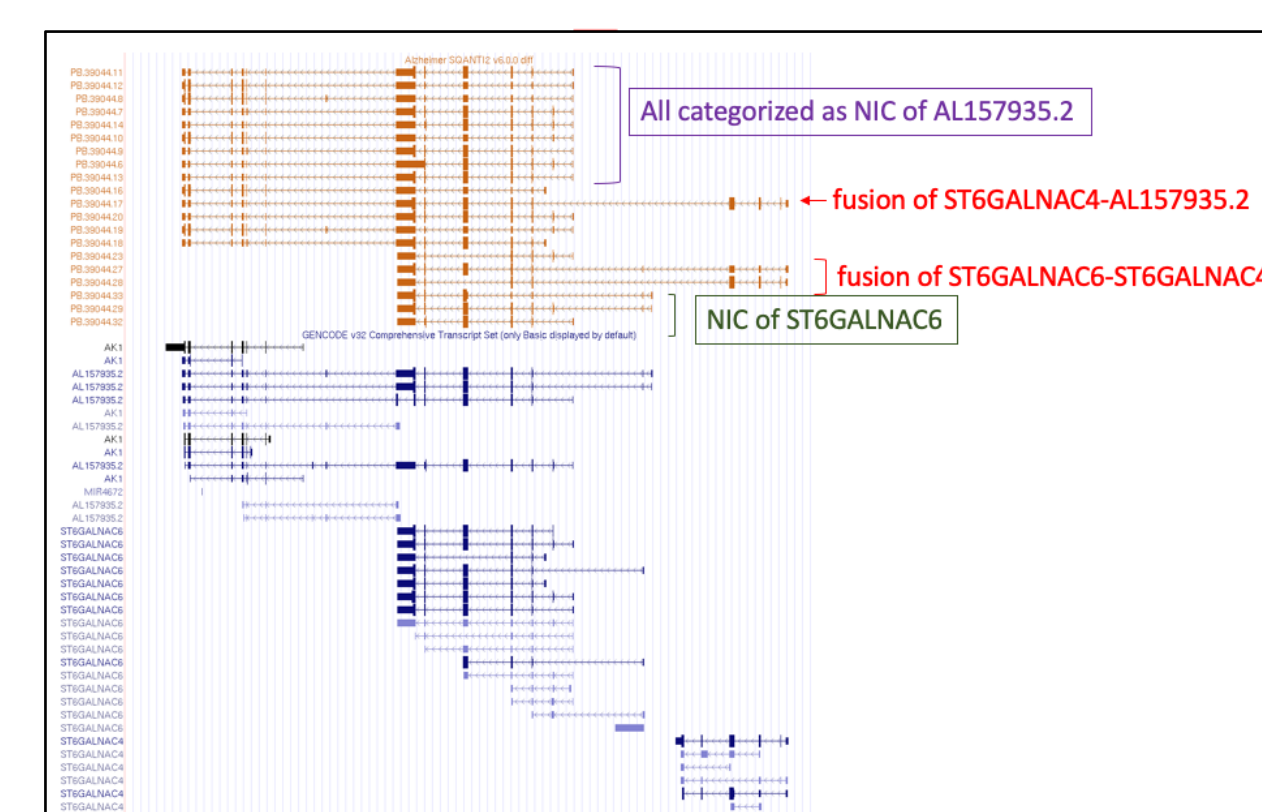
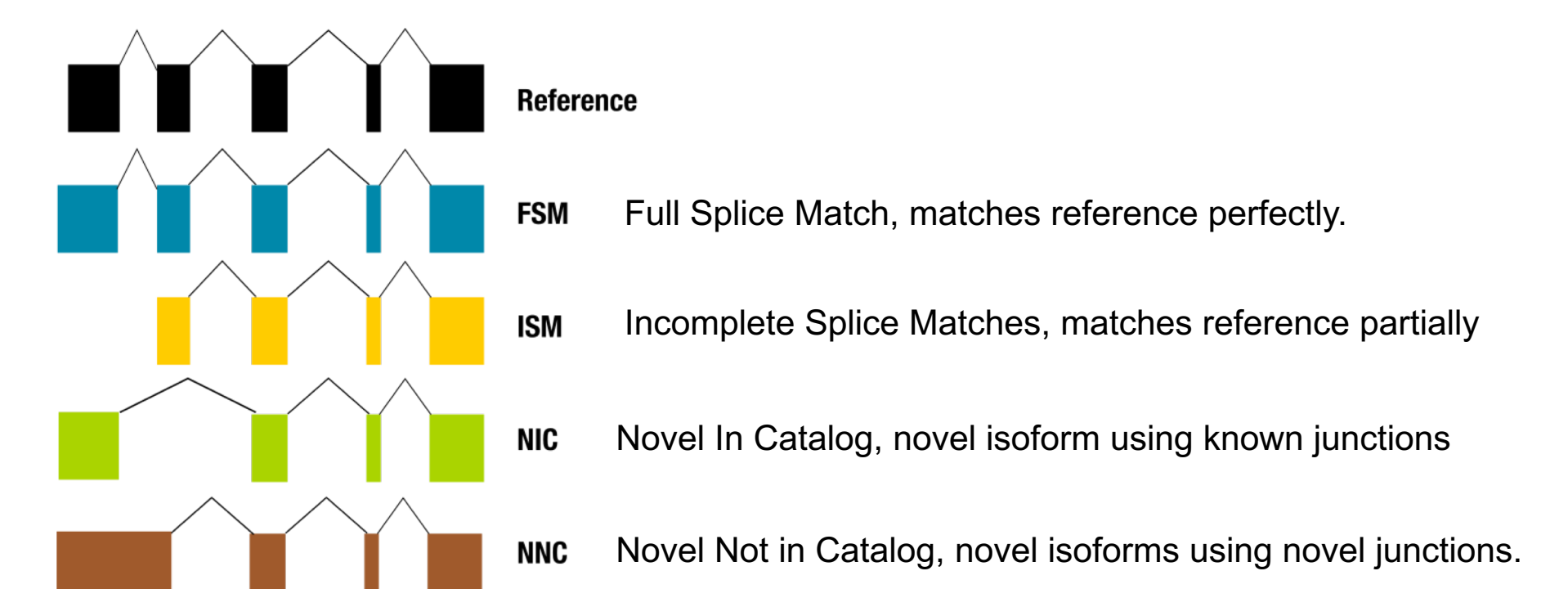
- 1 SMRT Cell 8M for whole transcriptome
- Up to 4 million full-length reads
- Accuracy 99-99.9%

	Unique Genes	Unique Transcripts	Unique ORFs
Single Cell, Human Brain Organoid	14,737	60,815	34,697
Single Cell, Human Cell Line	17,767	237,951	89,399
Bulk, UHRR ^[3]	16,328	183,689	60,649
Bulk, Alzheimer Brain ^[4]	17,670	162,290	80,539

Main Bioinformatics Tools

	Input	Output
Sequencing		subreads .bam
Iso-Seq Analysis		subreads.bam or ccs.bam
Transcript Classification		Unique transcripts Reference genome Annotation (GTF) CAGE Peak Junction data...
Functional Annotation		SQANTI output
Differential Analysis		Experimental design Annotated GTFs

SQANTI2 Transcript Classification



SQANTI classifies isoforms from an Alzheimer brain Iso-Seq dataset at a complex locus.

Novel isoforms (NIC) and readthrough transcripts of multiple genes identified.

Supporting Bioinformatics Tools



- collapse redundant transcripts
- merge multi-sample output
- saturation curve
- file format conversion
- single cell analysis



- gene family finding
- genome reconstruction
- evaluate assembly



- collapse redundant transcripts
- merge multi-sample output
- NMD/ORF prediction
- transcript filtering

TALON

- long read processing & annotation pipeline developed independently by ENCODE4

^[1] Low et al., "Haplotype-Resolved Cattle Genomes Provide Insights Into Structural Variation and Adaptation", *bioRxiv* (2019)

Beiki et al., "Improved annotation of the domestic pig genome through integration of Iso-Seq and RNA-seq data", *BMC Genomics* (2019)

Wang et al., "Reviving the Transcriptome Studies: An Insight Into the Emergence of Single-Molecule Transcriptome Sequencing", *Front Genet* (2019)

^[2] Iso-Seq Express: <https://www.pacb.com/wp-content/uploads/Procedure-Checklist-Iso-Seq-Express-Template-Preparation-for-Sequel-and-Sequel-II-Systems.pdf>

^[3] UHRR: [https://github.com/PacificBiosciences/DevNet/wiki/Sequel-II-System-Data-Release:-Universal-Human-Reference-\(UHR\)-Iso-Seq](https://github.com/PacificBiosciences/DevNet/wiki/Sequel-II-System-Data-Release:-Universal-Human-Reference-(UHR)-Iso-Seq)

^[4] Alzheimer brain: https://downloads.paccloud.com/public/dataset/Alzheimer2019_IsoSeq/