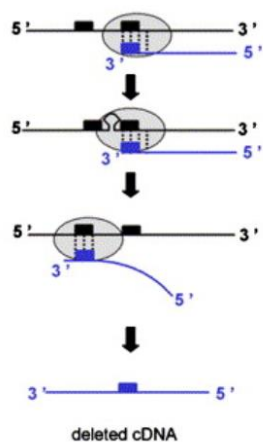
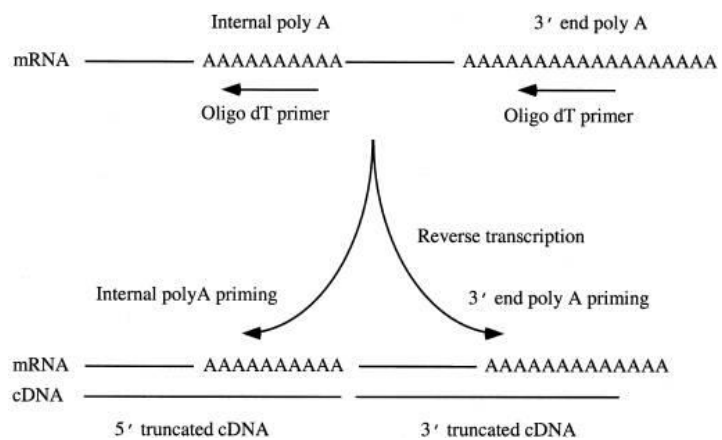




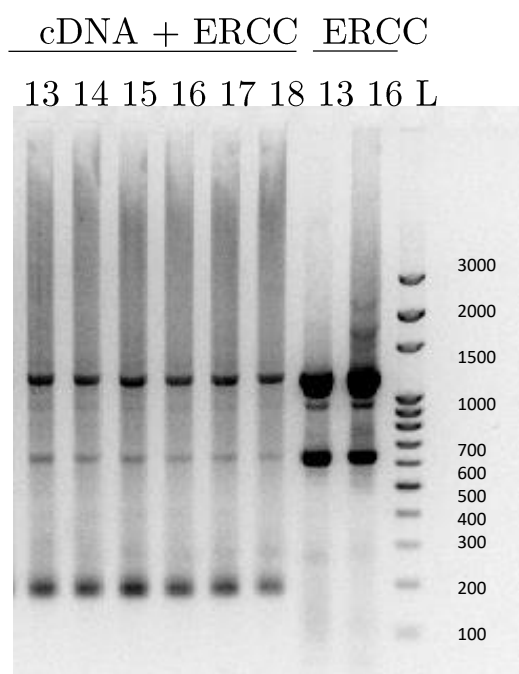
# A



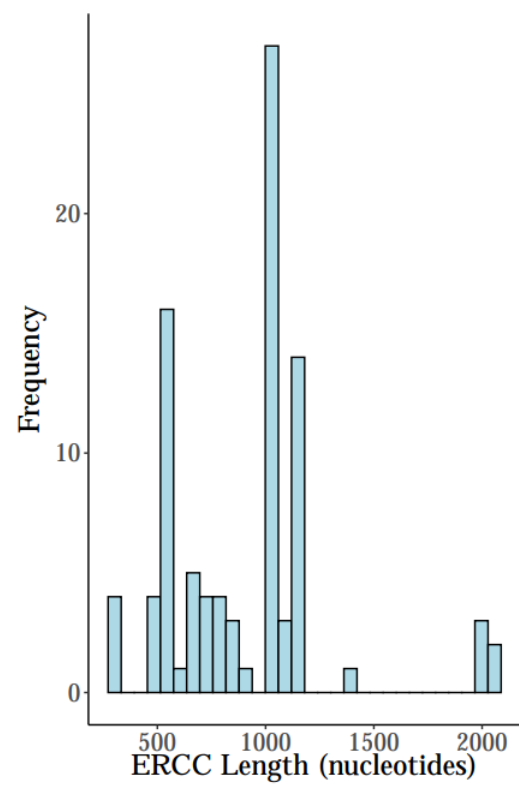
# B



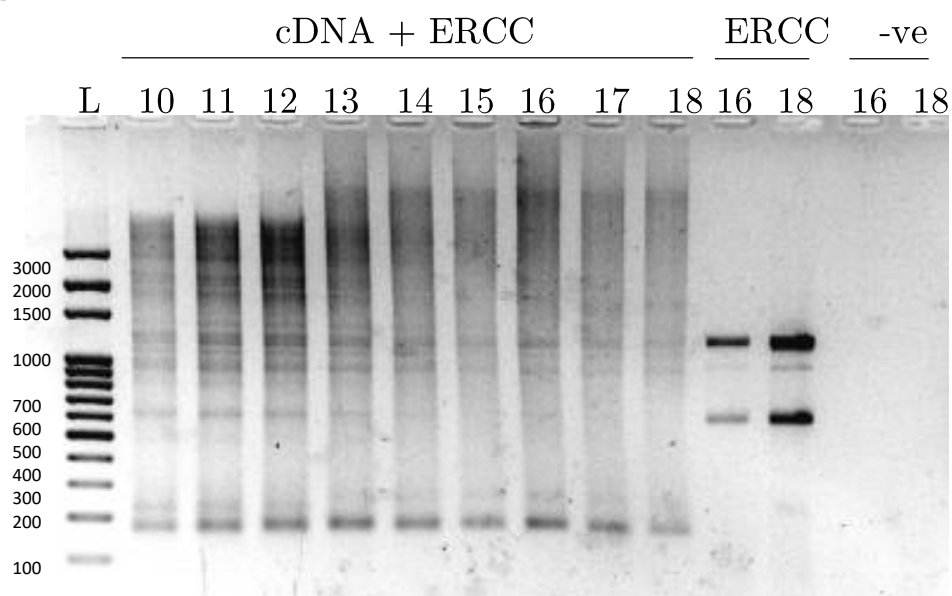
a

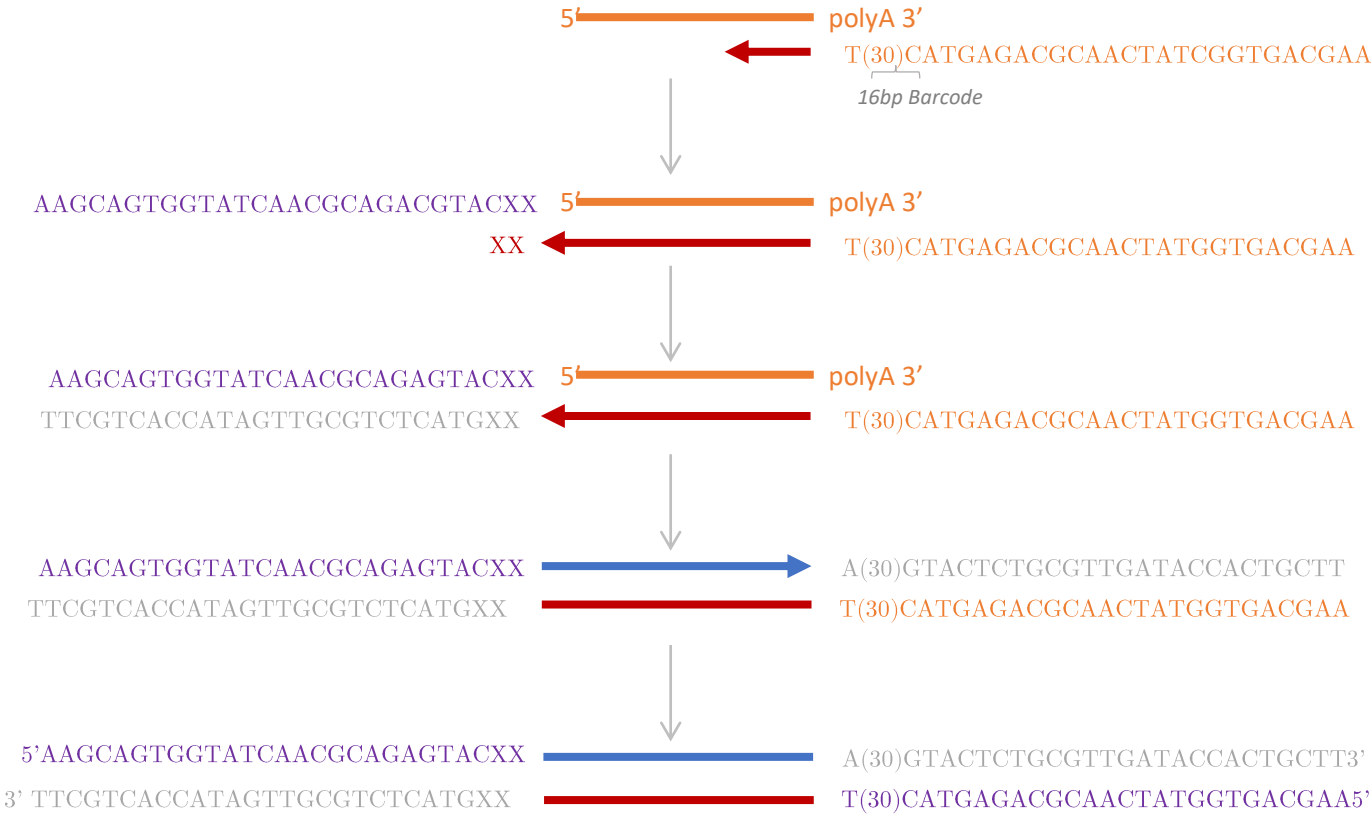


b



c





A

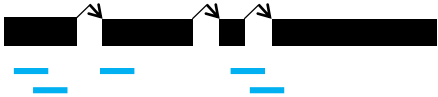
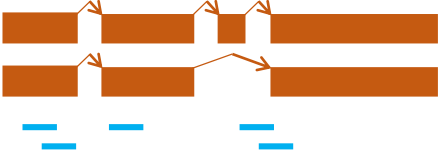



B Without Multiplexing

Strand	Sequence
Plus strand start	AATGTACTTCGTTCA GTTACGTATTGCTAAGCAGTGGTATCAACGCAGAGTACATGGG
Plus strand end	AAAAAAAAAGTACTCTGCGTTGATACCACTGCTT
Minus strand start	AATGTACTTCGTTCA GTTACGTATTGCTAAGCAGTGGTATCAACGCAGAGTACTTTTTTT T
Minus strand end	CCCATGTACTCTGCGTTGATACCACTGCTT

C With Multiplexing

Strand	Sequence
Plus strand start	AATGTACTTCGTTCA GTTACGTATTGCTAAGCAGTGGTATCAACGCAGAGTACATGGG
Plus strand end	AAAAAAAA<Barcode>GTACTCTGCGTTGATACCACTGCTT
Minus strand start	AATGTACTTCGTTCA GTTACGTATTGCT<Barcode>AAGCAGTGGTATCAACGCAGAGT ACTTTTTTTT
Minus strand end	CCCATGTACTCTGCGTTGATACCACTGCTT

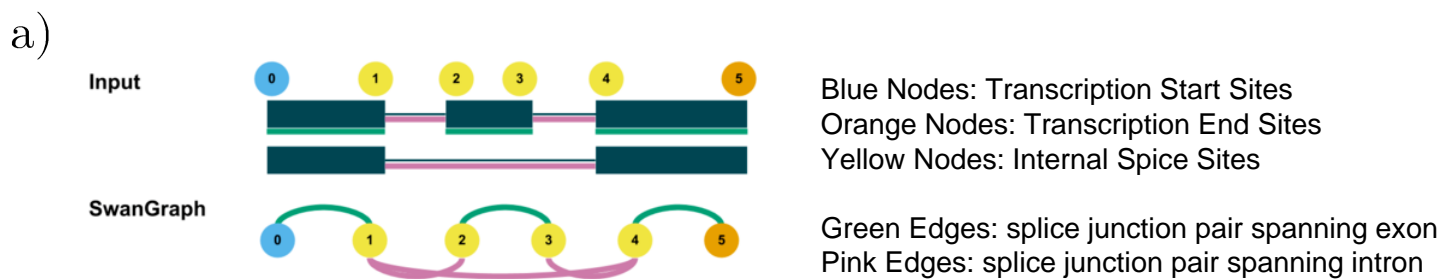
	Analysis	Annotation	Expression
A		Reference genome	RNA-Seq reads
B		Long-read transcriptome	RNA-Seq reads
C		Long-read transcriptome	Long reads



Isoform	Associated Gene	Structural Category	Associated Transcript	Length	FL Sample 1	FL Sample 2
PB.1.1	Gene 1	FSM	ENMUST1	3000	10	10
PB.1.2	Gene 1	ISM	ENMUST1	1500	100	90
PB.1.3	Gene 1	ISM	ENMUST2	2400	20	20
PB.1.4	Gene 1	ISM	ENMUST3	2900	5	4
PB.1.5	Gene 1	ISM	ENMUST3	2190	5	2
PB.1.6	Gene 1	FSM	ENMUST4	2420	1000	1250
PB.1.7	Gene 1	FSM	ENMUST4	2560	1	4



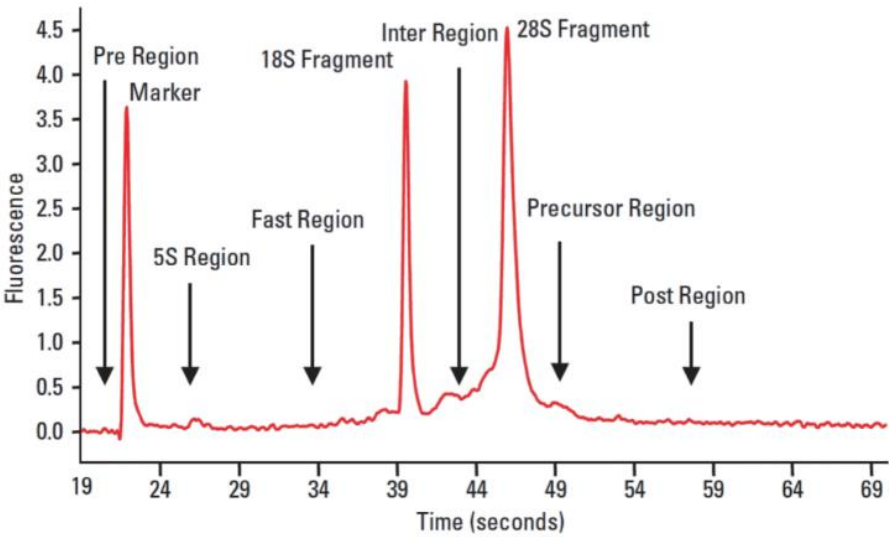
Isoform	Associated Gene	Structural Category	Associated Transcript	Length	FL Sample 1	FL Sample 2
PB.1.1	Gene 1	FSM	ENMUST1	3000	110	100
PB.1.3	Gene 1	ISM	ENMUST2	2400	20	20
PB.1.4	Gene 1	ISM	ENMUST3	2900	10	6
PB.1.7	Gene 1	FSM	ENMUST4	2560	1001	1254



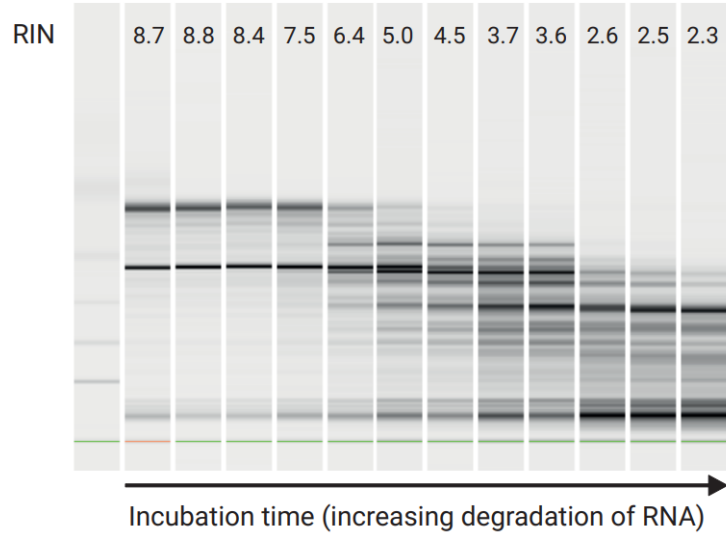
b)

ENMUST0000014268.4	Associated Transcript	Mean Expression, Length	Structure
	PB.6948.1 (FSM)	0.08FL, 3939bp	
	PB.6948.7 (FSM)	645FL, 1600bp	
	PB.6948.16 0 (ISM)	0.083FL, 1079bp	
	PB.6948.35 (ISM)	46.7FL, 1000bp	
	PB.6948.81 (ISM)	69.2FL, 812bp	
	PB.6948.11 5 (ISM)	32.0FL, 663bp	
	PB.6948.15 (ISM)	0.083FL, 847bp	

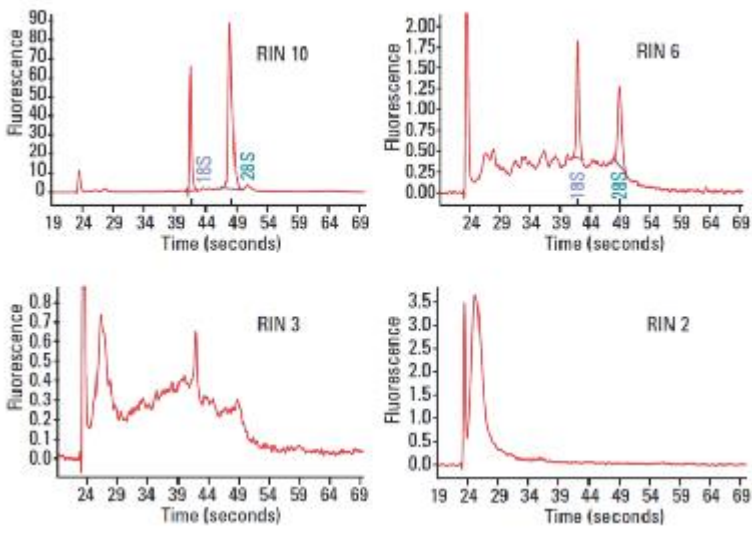
a)



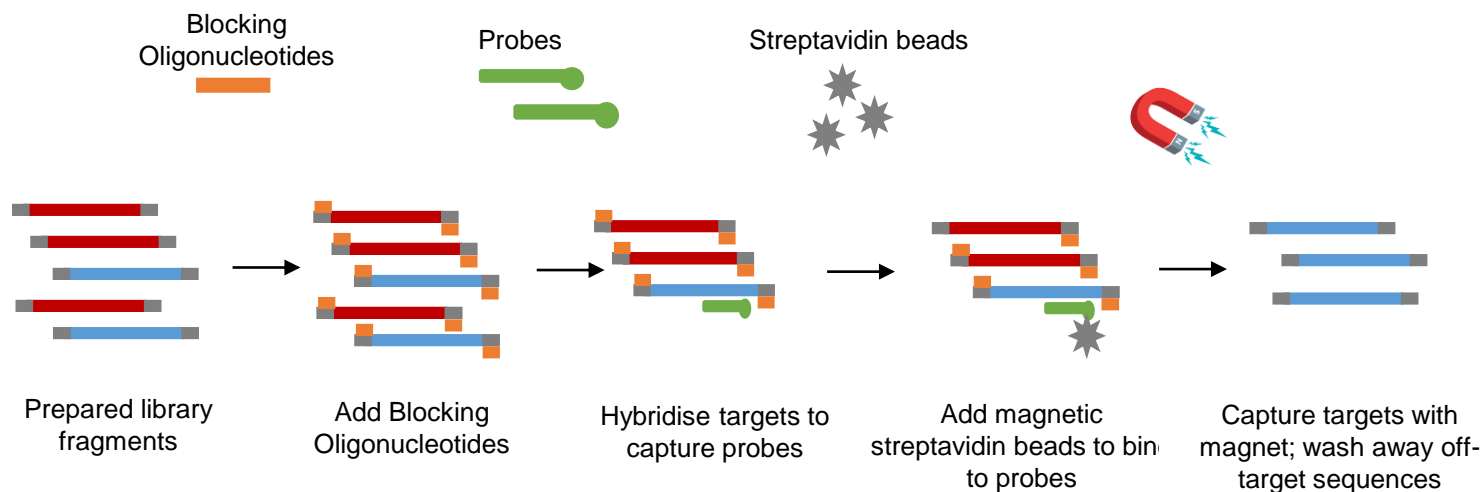
b)



c)

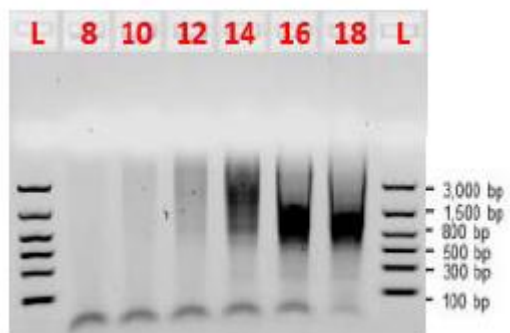


# A

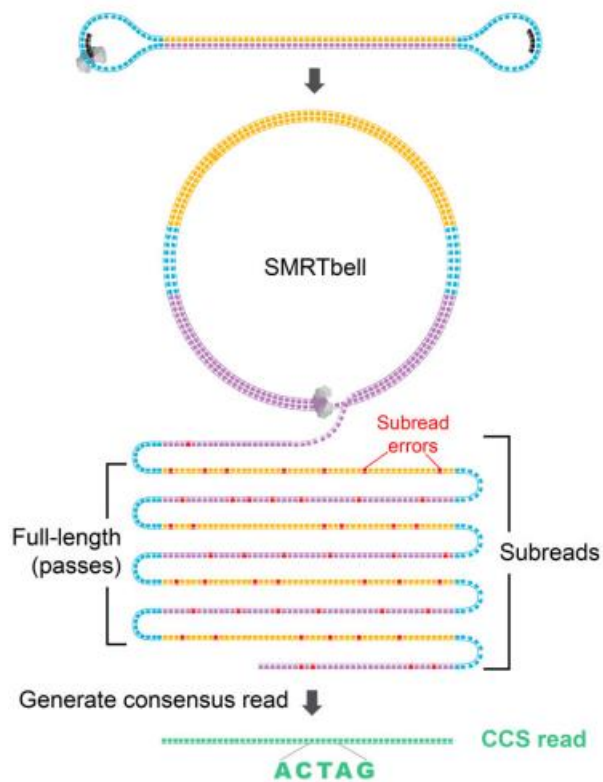


# B

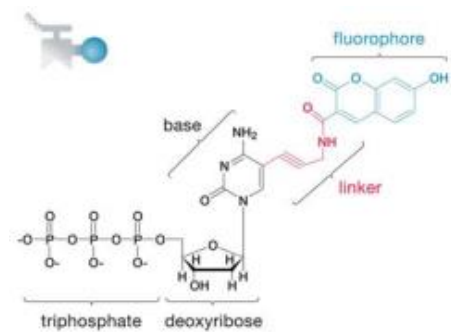




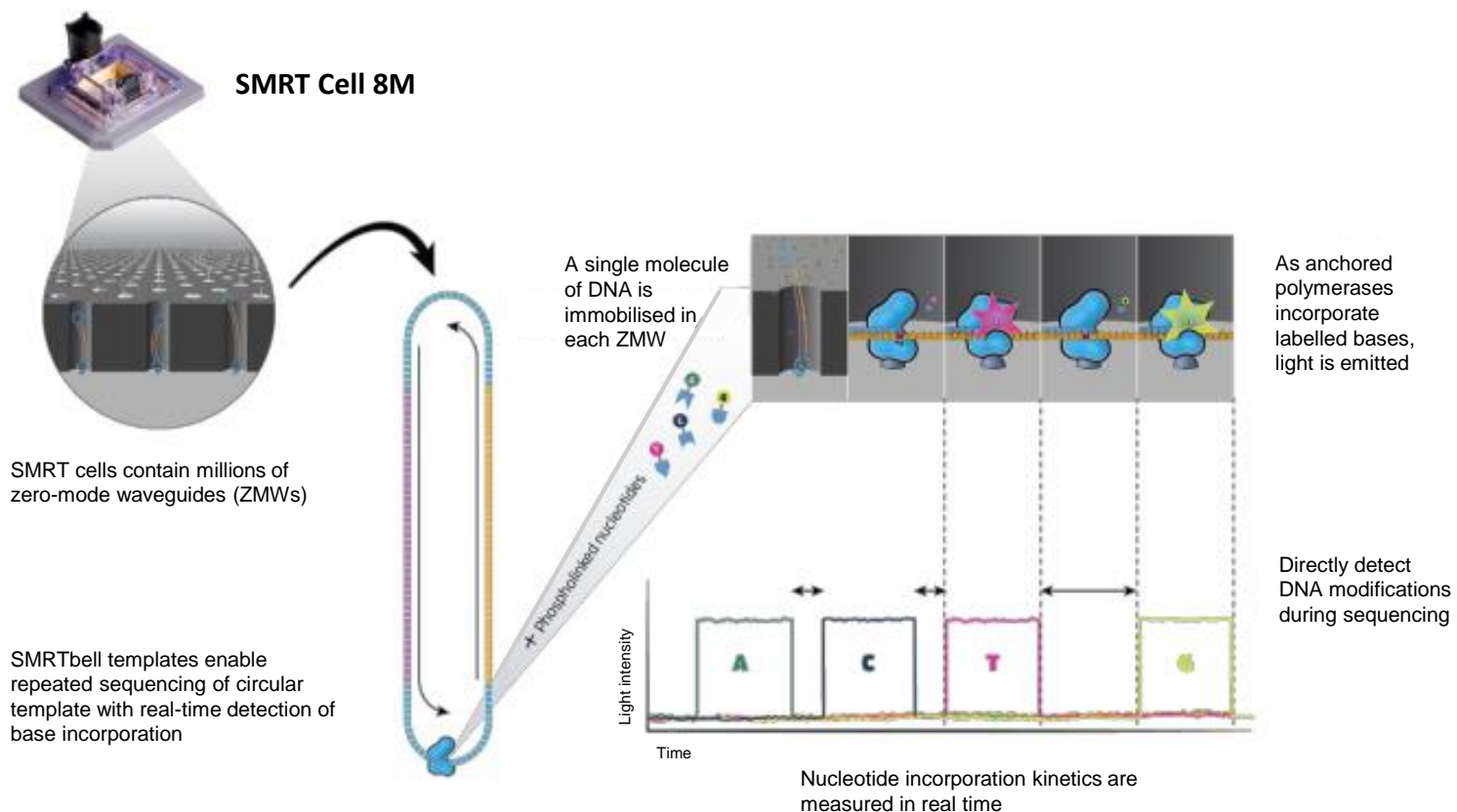
A

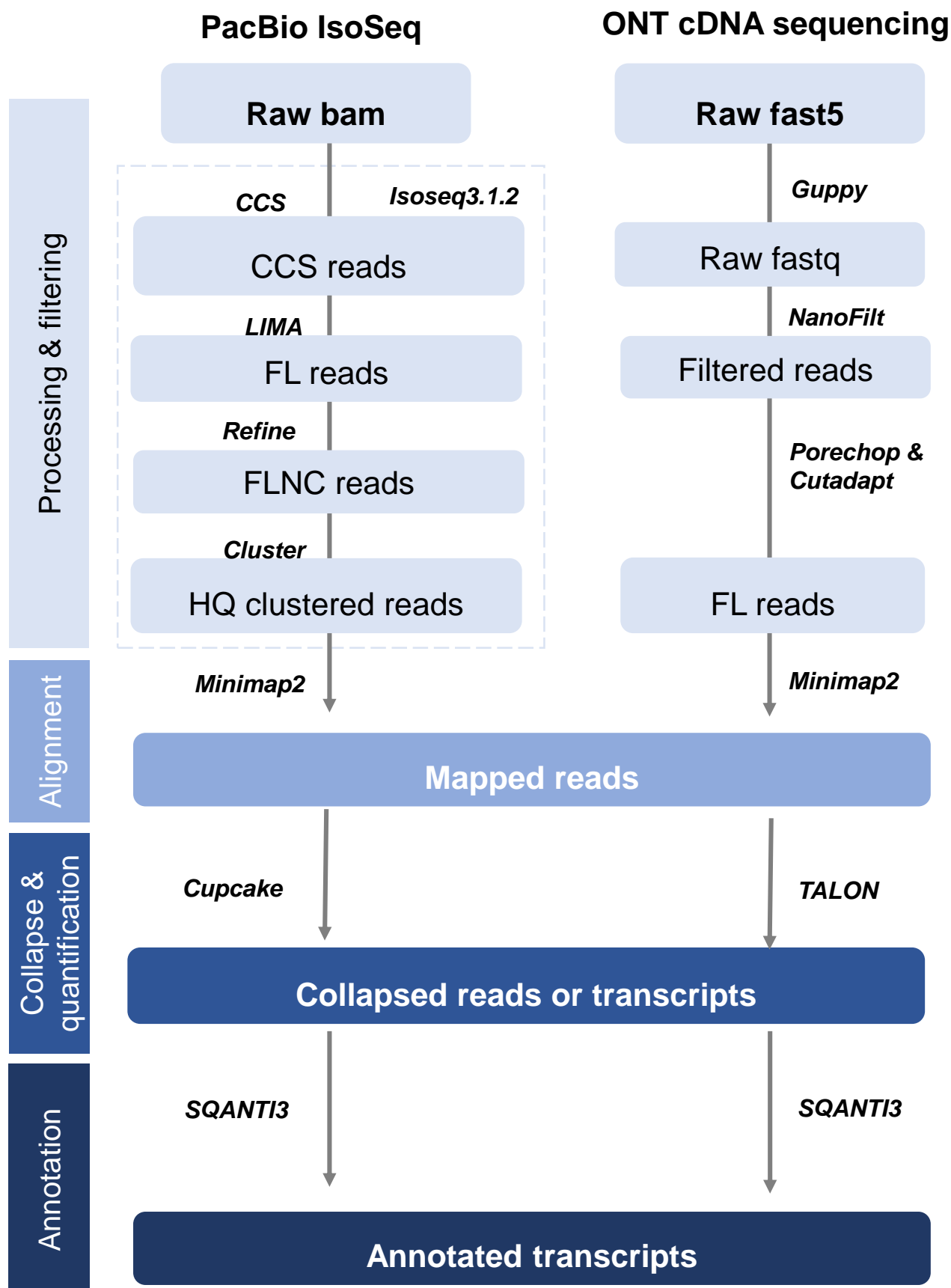


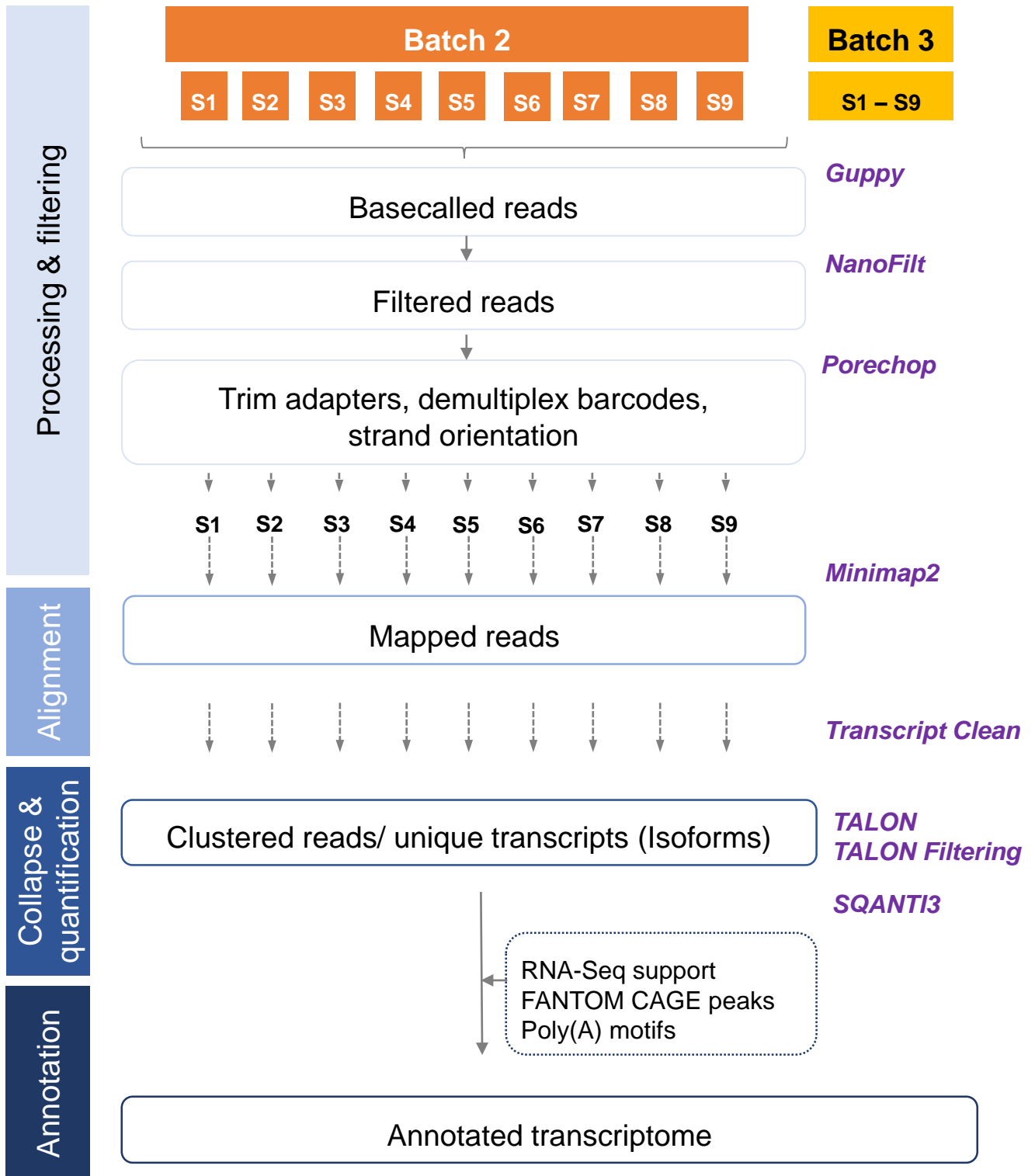
C



B

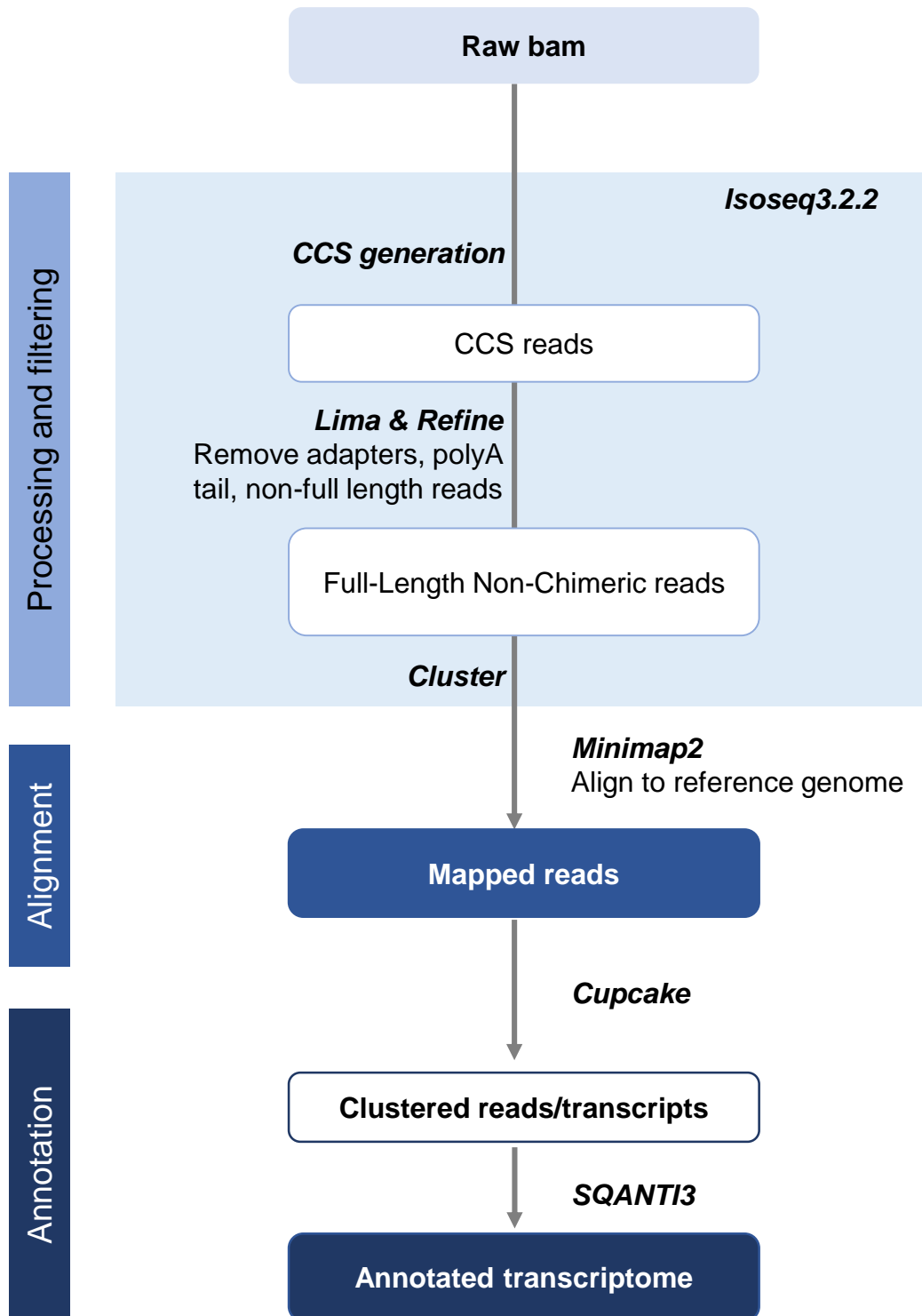




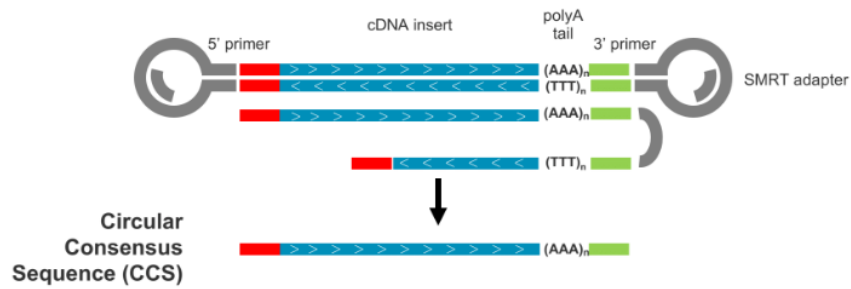




# PacBio Iso-Seq Bioinformatics Pipeline



A

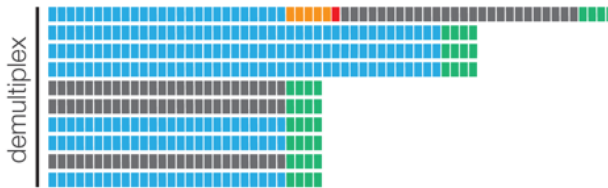


B



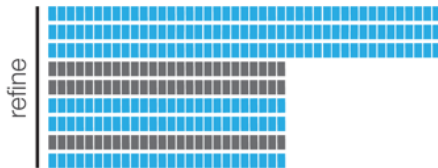
- Use **polished** CCS reads
- Only full-pass ZMWs

C



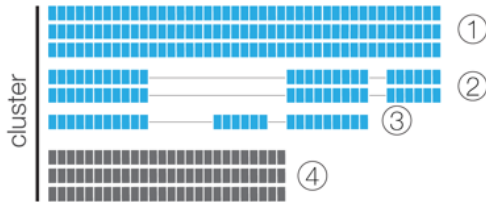
- Barcoded and unbarcoded cDNA primer removal
- Orientation
- Unwanted primer combination removal

D



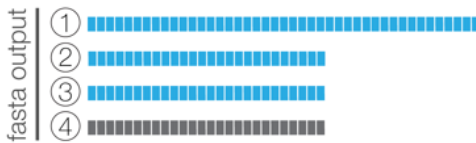
- PolyA tail trimming
- Concatemer removal

E



- Hierarchical,  $n \cdot \log(n)$  clustering, alignment of shorter to longer sequences
- Iterative cluster merging
- Generate consensus for each read cluster using QV guided PoA

F

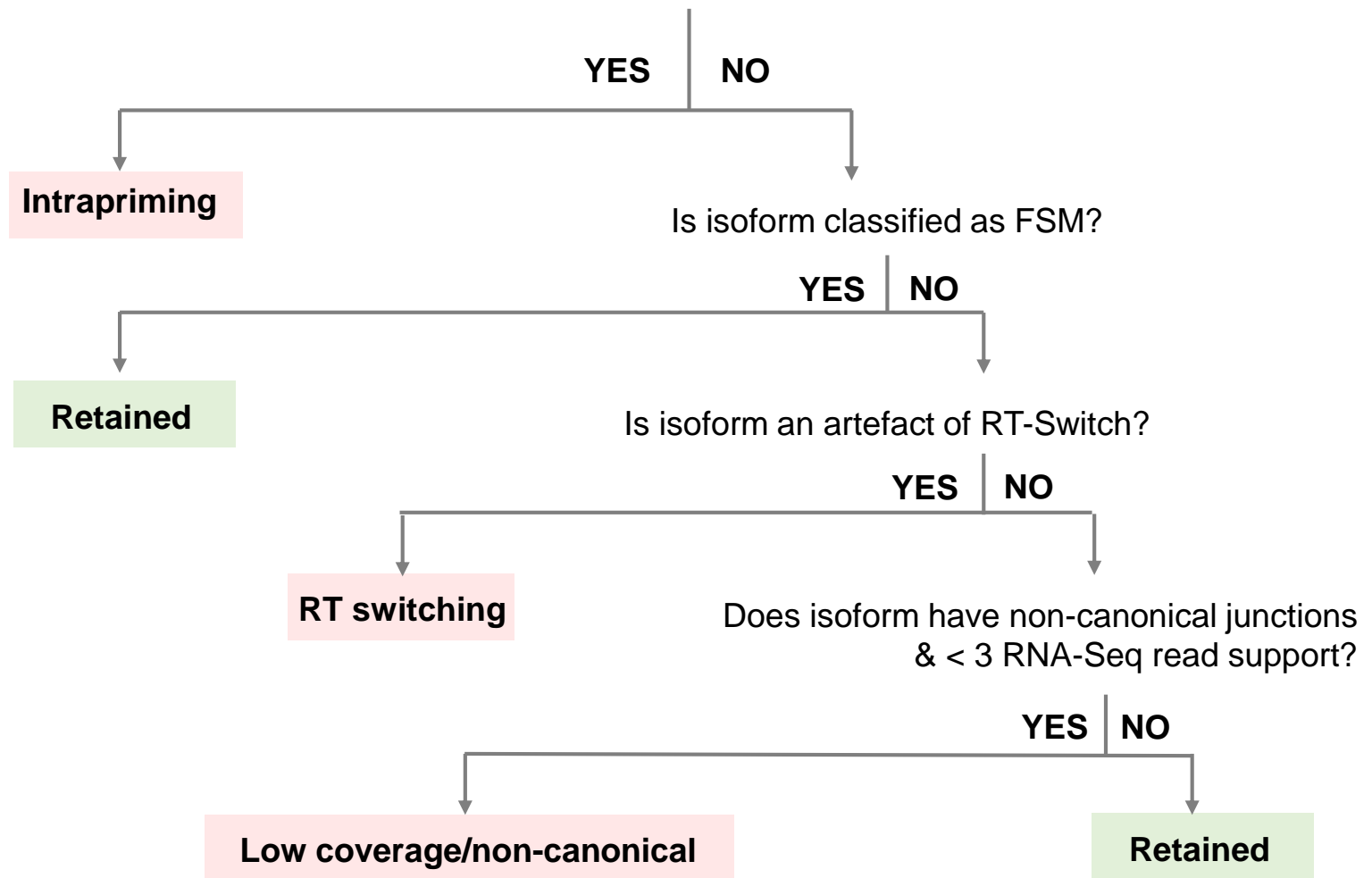


- Fasta output is split into HQ and LQ reads
- One consensus per read cluster

Iso-Seq3

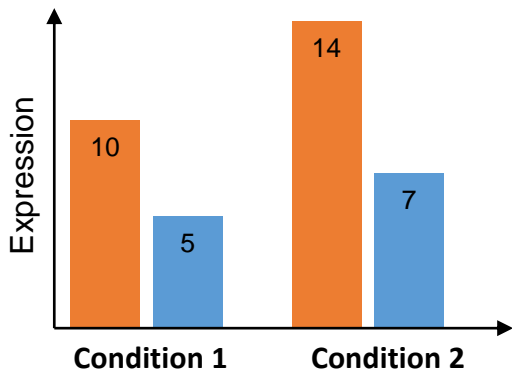
Does the isoform have:

- > 60% genomic A's in 20bp window downstream of TTS &
- unknown distance of query isoform 5' end and reference TSS &
- no detected poly(A) motif &
- distance between query isoform 3' end and reference TTS > 50bp



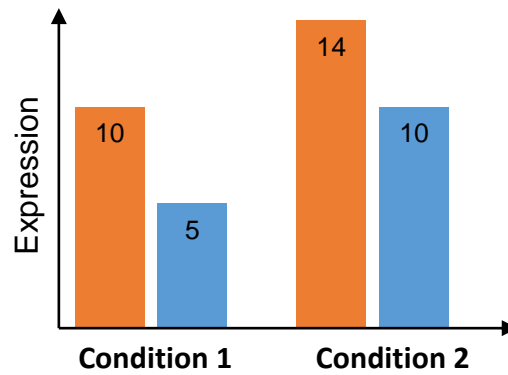
A

Differential transcript expression



B

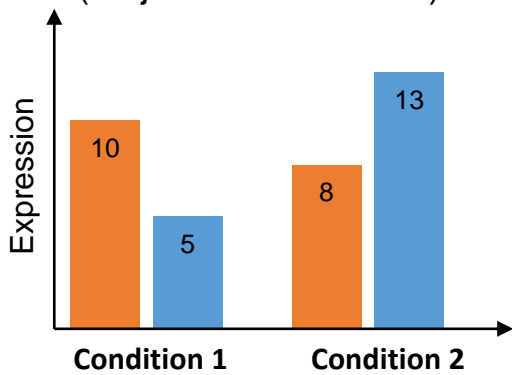
Differential transcript usage



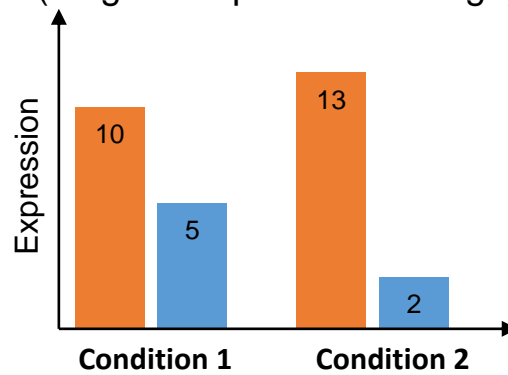
Isoform A

Isoform B

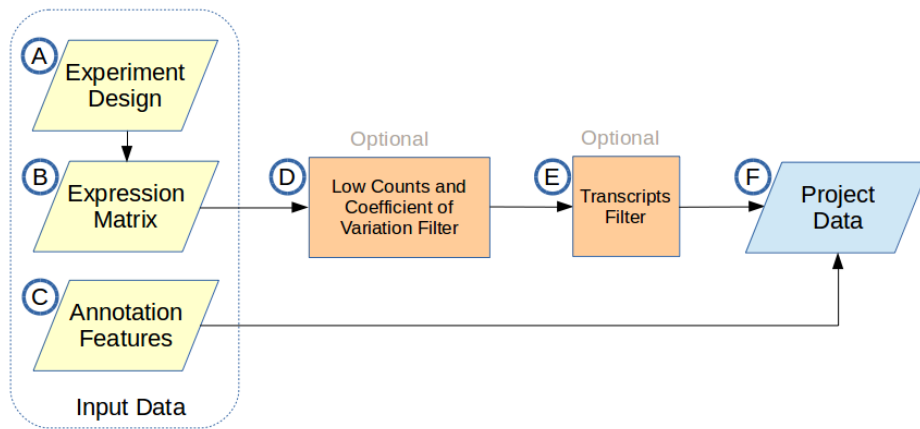
C

Differential transcript usage  
(major isoform switch)

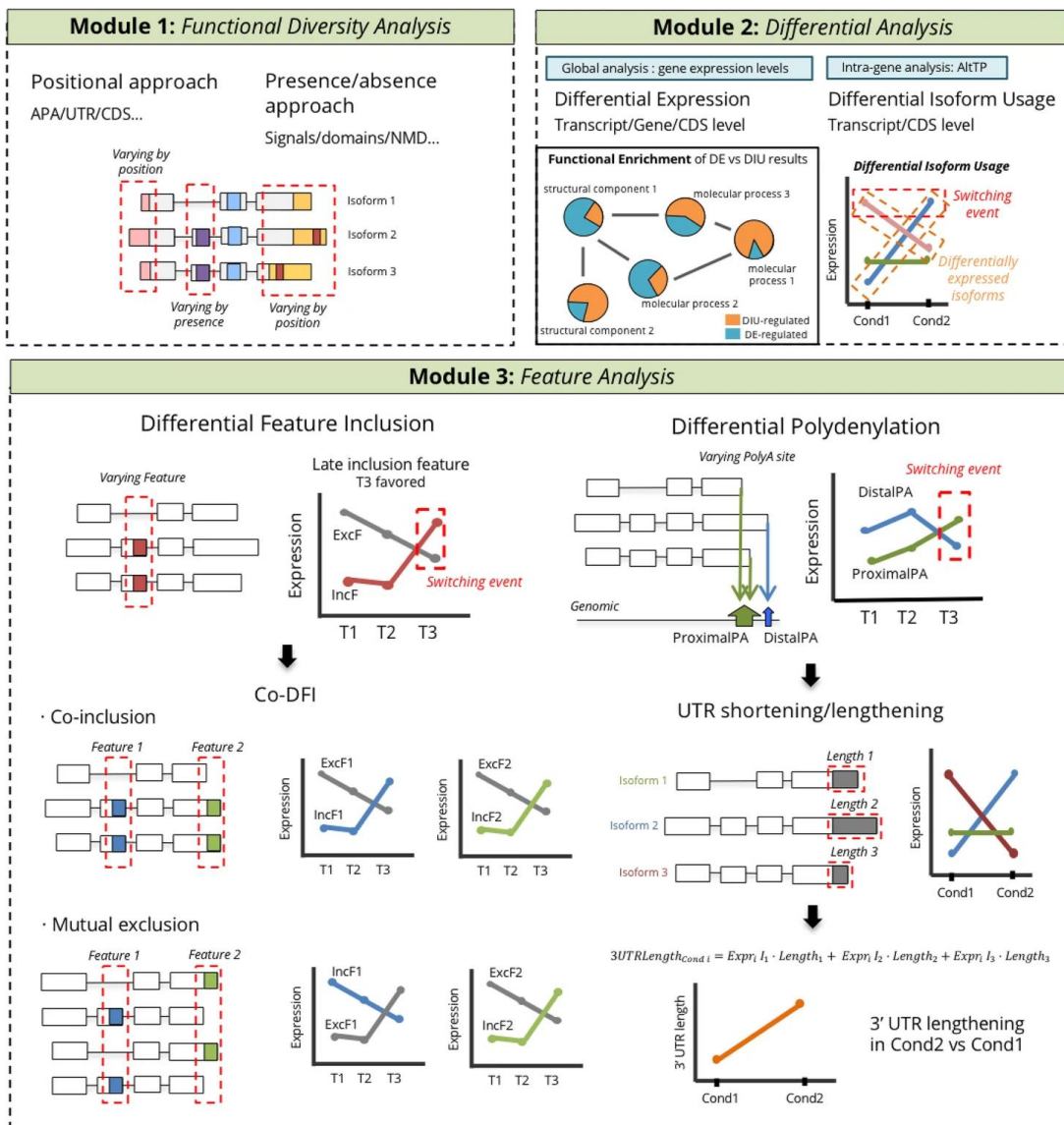
D

Differential transcript usage  
(no gene expression change)

A



B



A



B

Strand	Sequence
Plus start	AATGTACTTCGTTCACTTACGTATTGCTAAGCAGTGGTATCAACGCAGAGTACATGGG
Plus end	AAAAAAAA<B>GTACTCTGCGTTGATACCACTGCTT
Minus start	AATGTACTTCGTTCACTTACGTATTGCTAAGCAGTGGTATCAACGCAGAGTAC<B>TTTTTTTTT
Minus end	CCCATGTACTCTGCGTTGATACCACTGCTT

