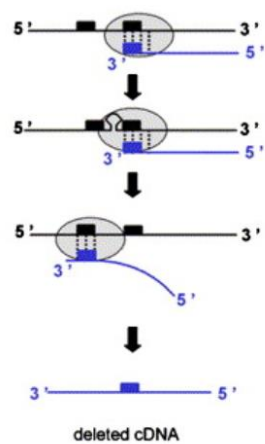
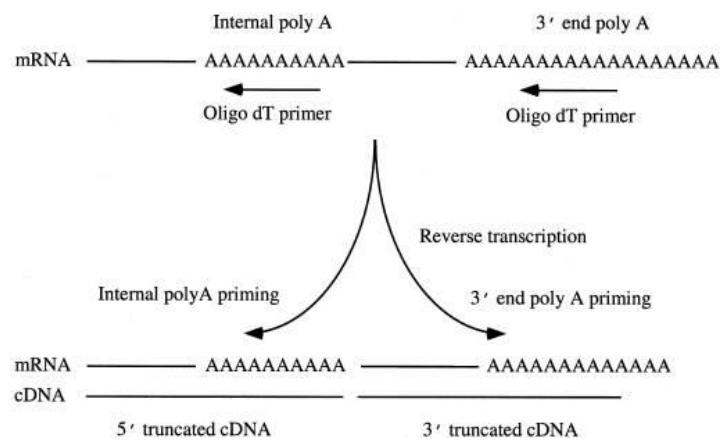




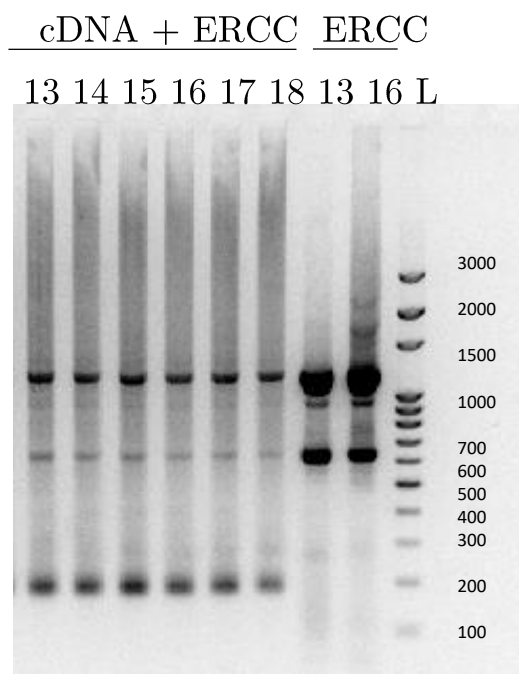
a



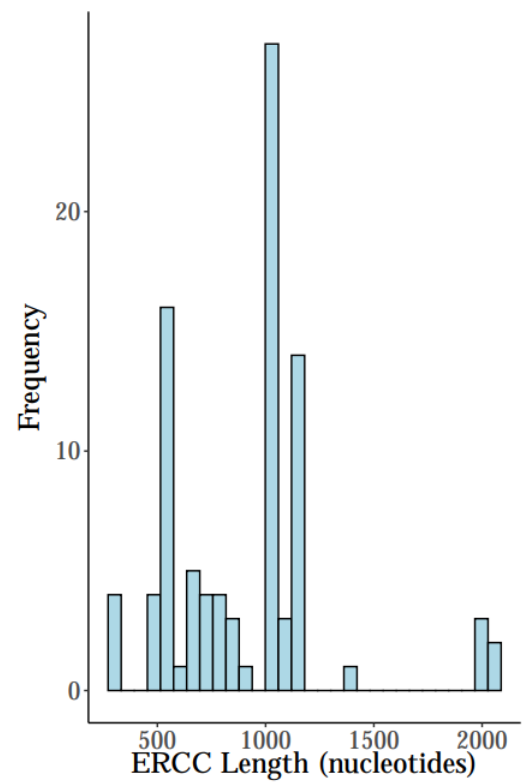
b



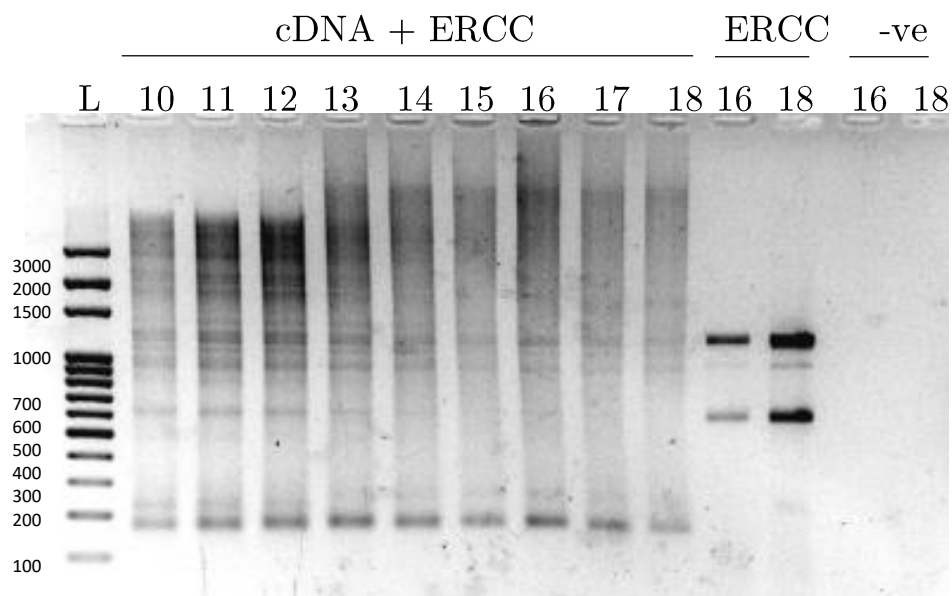
a

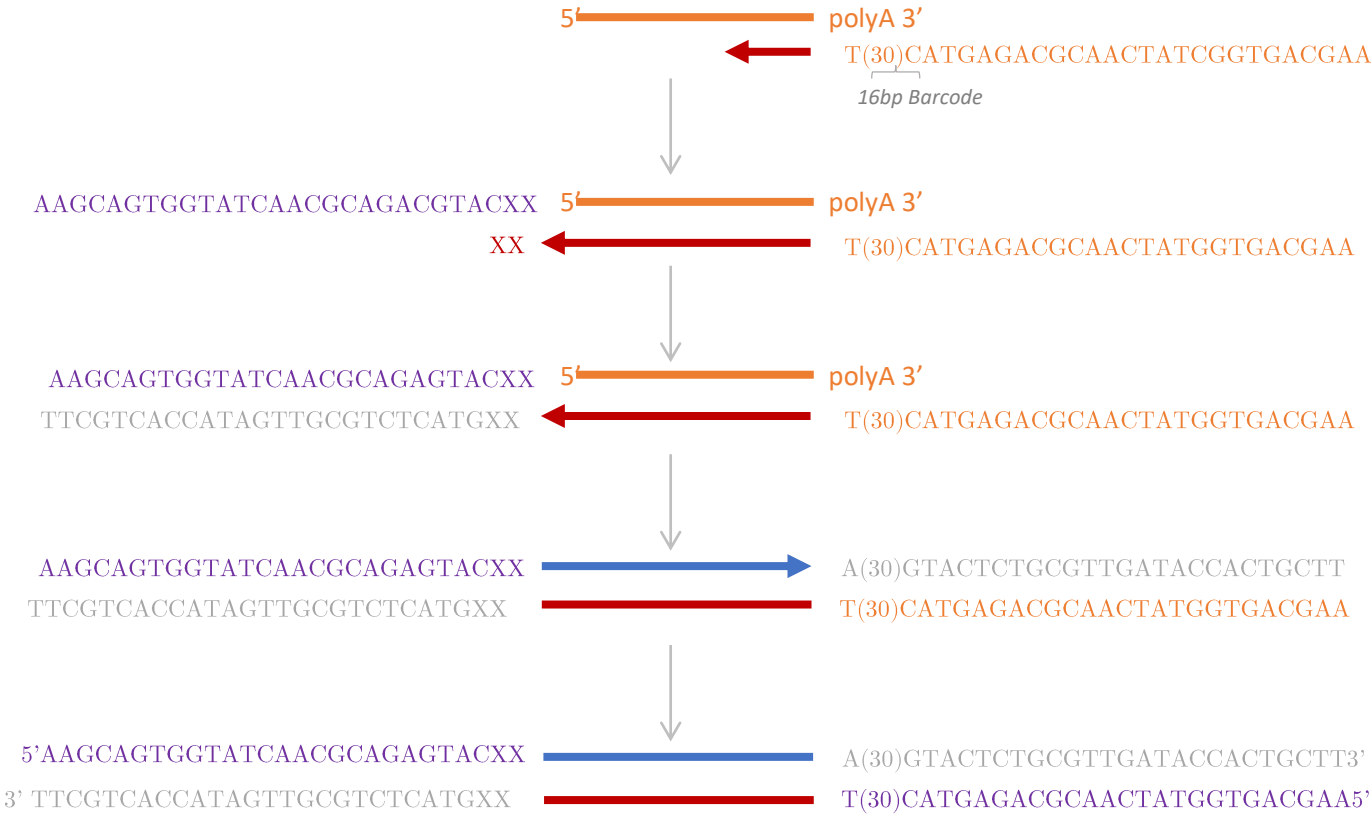


b



c





a)

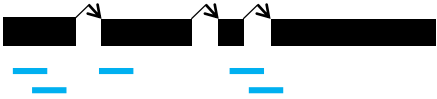
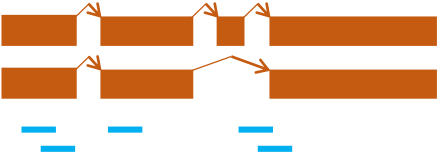



b) Without Multiplexing

Strand	Sequence
Plus strand start	AATGTACTTCGTTTCAGTTACGTATTGCTAAGCAGTGGTATCAACGCAGAGTACATGGG
Plus strand end	AAAAAAAAAGTACTCTGCGTTGATACCACTGCTT
Minus strand start	AATGTACTTCGTTTCAGTTACGTATTGCTAAGCAGTGGTATCAACGCAGAGTACTTTTTTTT
Minus strand end	CCCATGTACTCTGCGTTGATACCACTGCTT

c) With Multiplexing

Strand	Sequence
Plus strand start	AATGTACTTCGTTTCAGTTACGTATTGCTAAGCAGTGGTATCAACGCAGAGTACATGGG
Plus strand end	AAAAAAAA<Barcode>GTACTCTGCGTTGATACCACTGCTT
Minus strand start	AATGTACTTCGTTTCAGTTACGTATTGCT<Barcode>AAGCAGTGGTATCAACGCAGAGTACTTTTTT TTT
Minus strand end	CCCATGTACTCTGCGTTGATACCACTGCTT

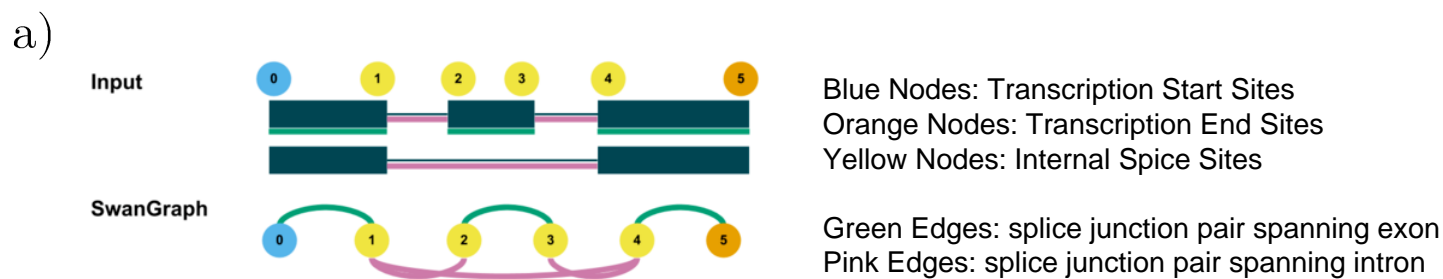
	Analysis	Annotation	Expression
1.		Reference genome	RNA-Seq reads
2.		Iso-Seq defined transcriptome	RNA-Seq reads
3.		Iso-Seq defined transcriptome	Iso-Seq reads



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



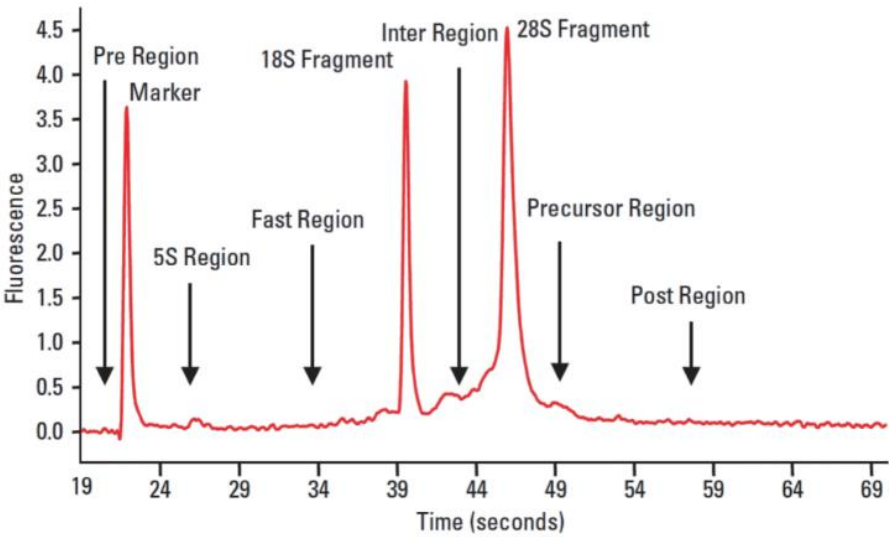
Isoform	Associated Gene	Structural Category	Associated Transcript	Length	FL Sample 1	FL Sample 2
PB.1.1	Gene 1	FSM	ENMUST1	3000	110	100
PB.1.3	Gene 1	ISM	ENMUST2	2400	20	20
PB.1.4	Gene 1	ISM	ENMUST3	2900	10	6
PB.1.7	Gene 1	FSM	ENMUST4	2560	1001	1254



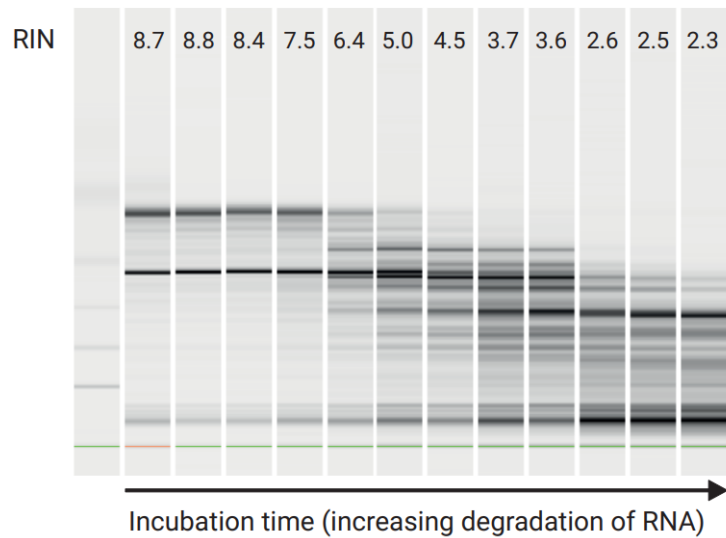
b)

Associated Transcript	Mean Expression, Length	Structure
PB.6948.1 (FSM)	0.08FL, 3939bp	
PB.6948.7 (FSM)	645FL, 1600bp	
PB.6948.160 (ISM)	0.083FL, 1079bp	
PB.6948.35 (ISM)	46.7FL, 1000bp	
PB.6948.81 (ISM)	69.2FL, 812bp	
PB.6948.115 (ISM)	32.0FL, 663bp	
PB.6948.15 (ISM)	0.083FL, 847bp	

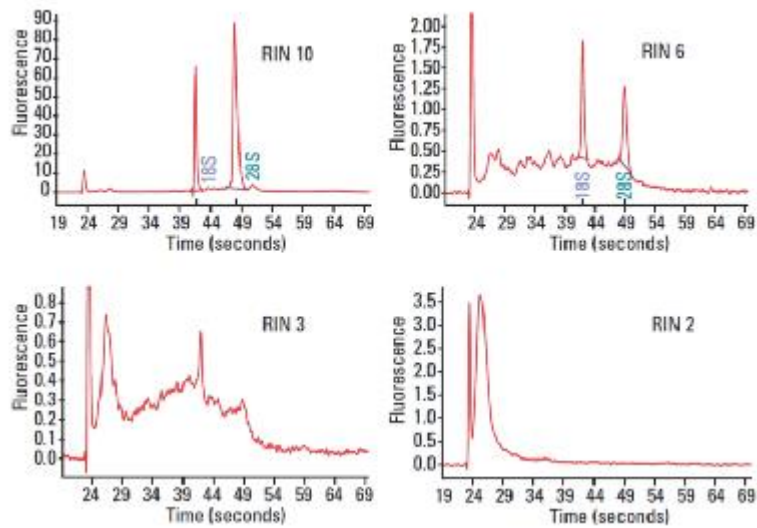
a)



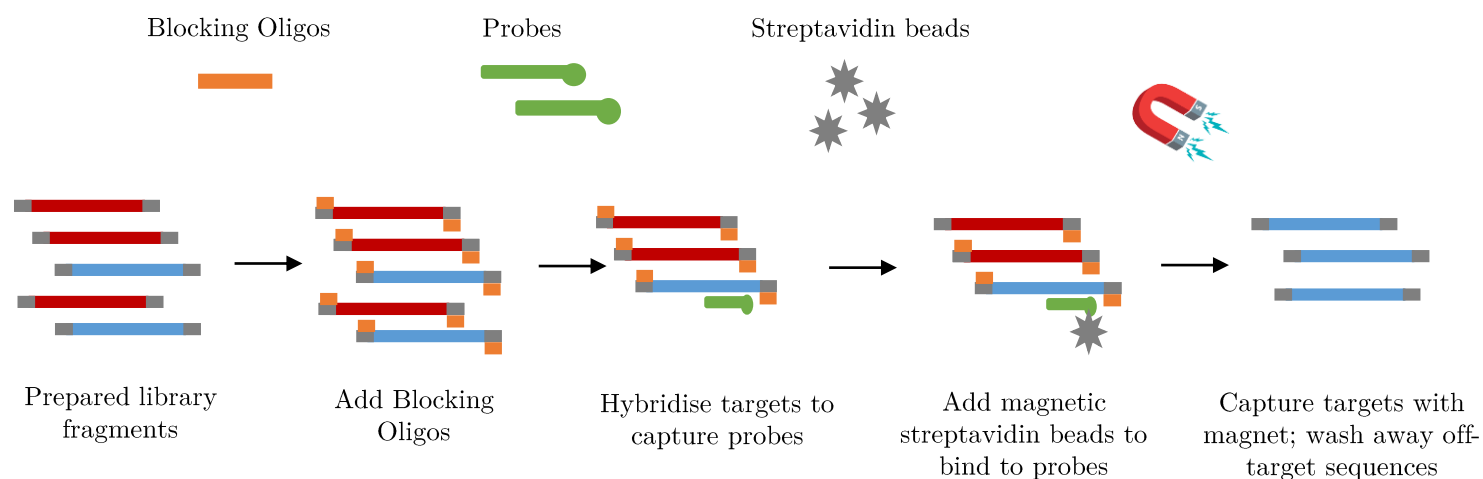
b)



c)

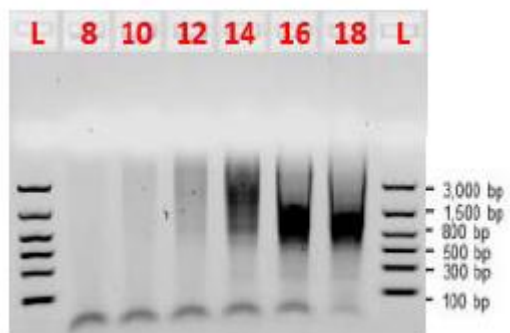


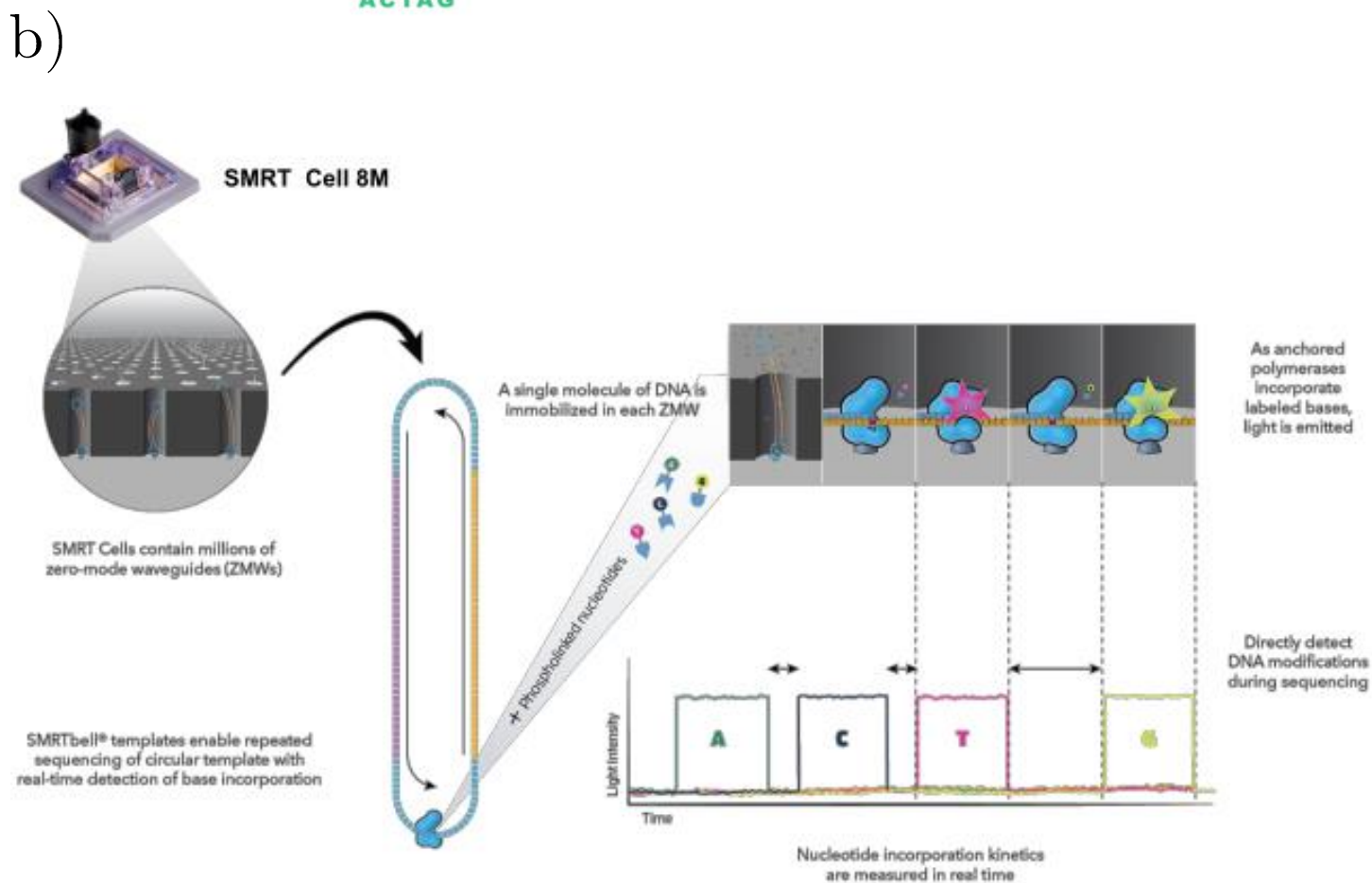
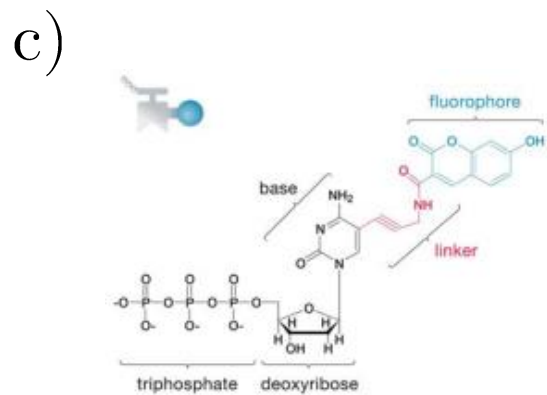
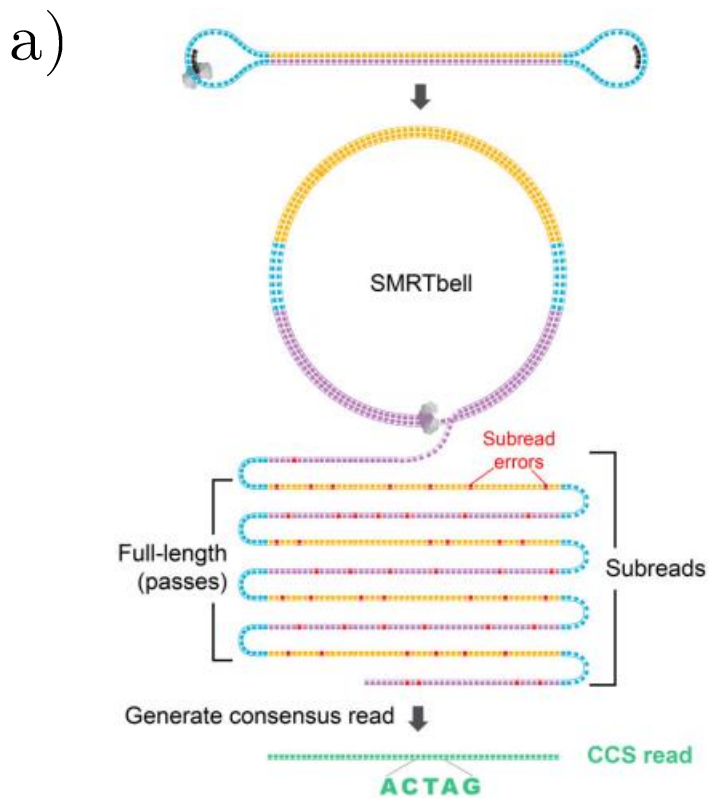
a)



b)

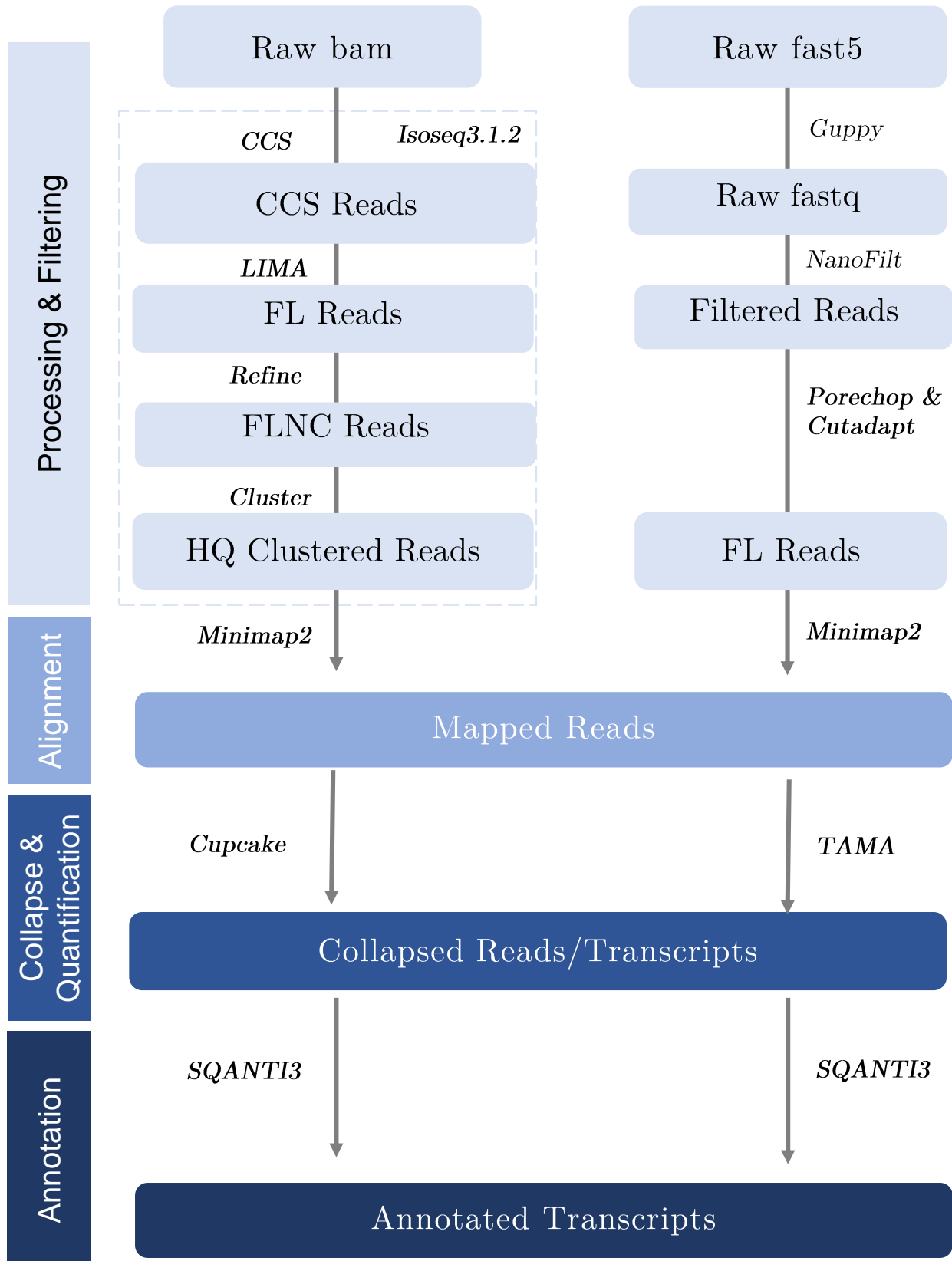


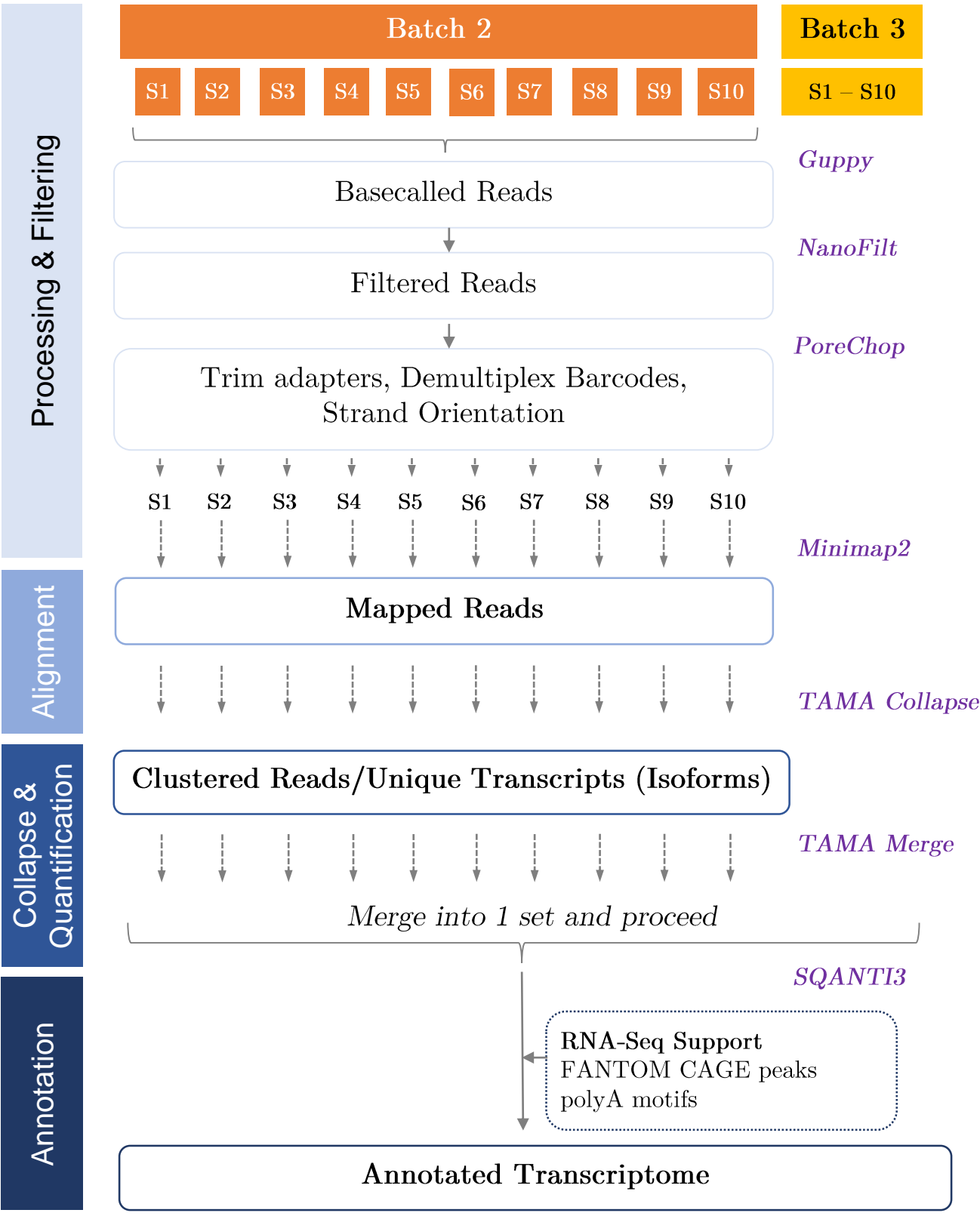




## PacBio IsoSeq

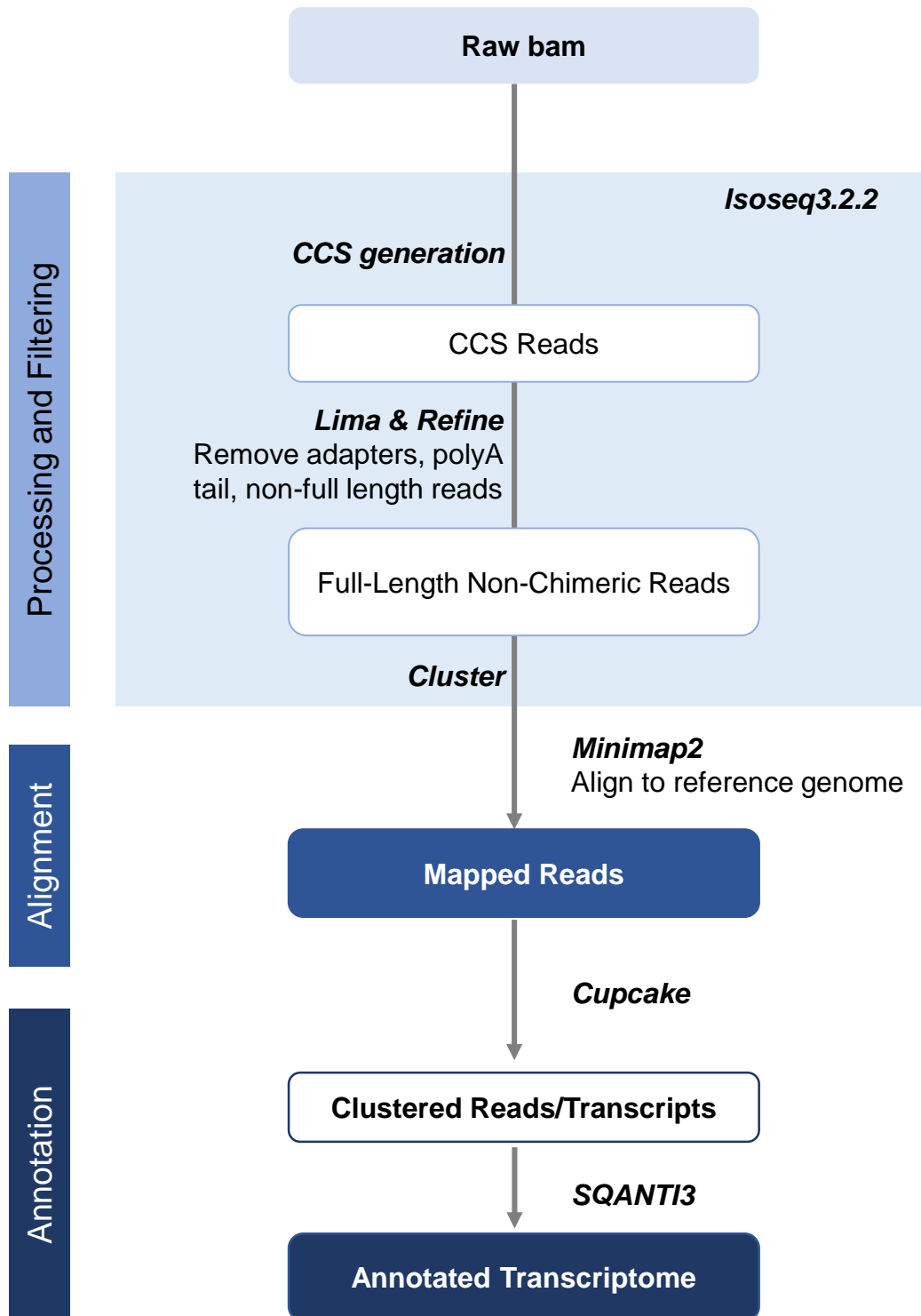
## ONT cDNA Sequencing



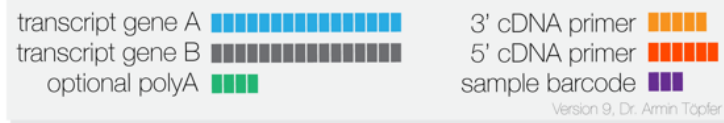
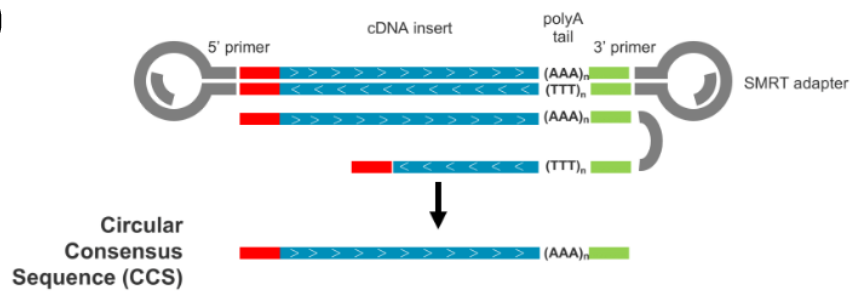




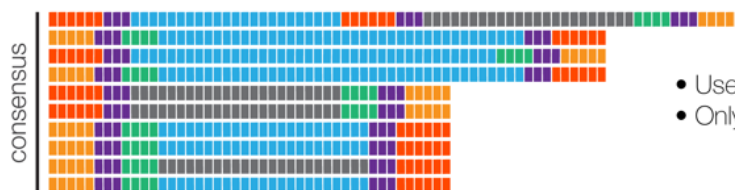
# PacBio Iso-Seq Bioinformatics Pipeline



a)

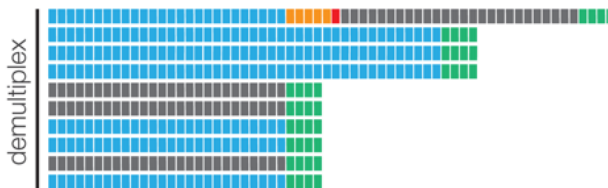


b)



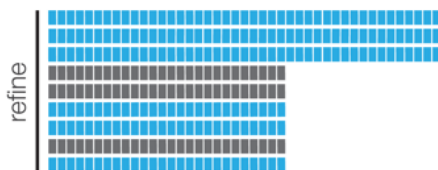
- Use **polished** CCS reads
- Only full-pass ZMWs

c)



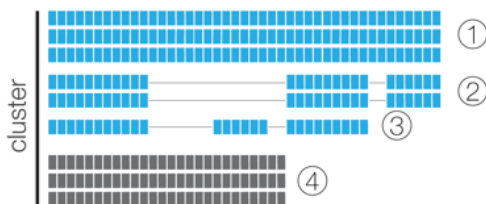
- Barcoded and unbarcoded cDNA primer removal
- Orientation
- Unwanted primer combination removal

d)



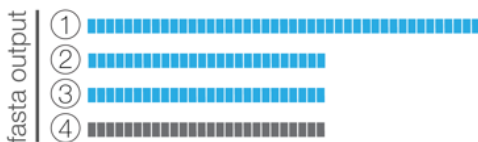
- PolyA tail trimming
- Concatemer removal

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 As in 20bp window downstream of TTS &
- unknown distance of query isoform 5'end and reference TSS &
- no detected polyA motif &
- distance between query isoform 3' end and reference TTS >50bp

