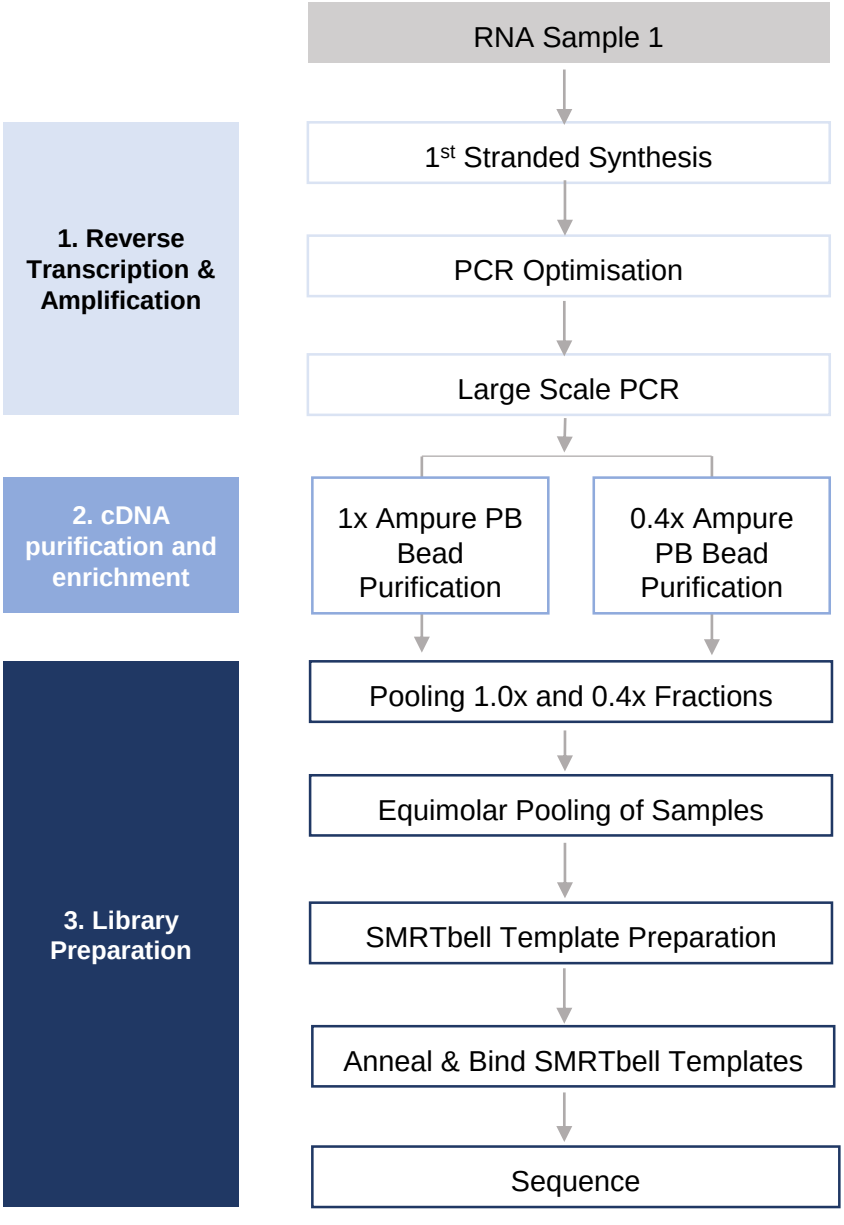
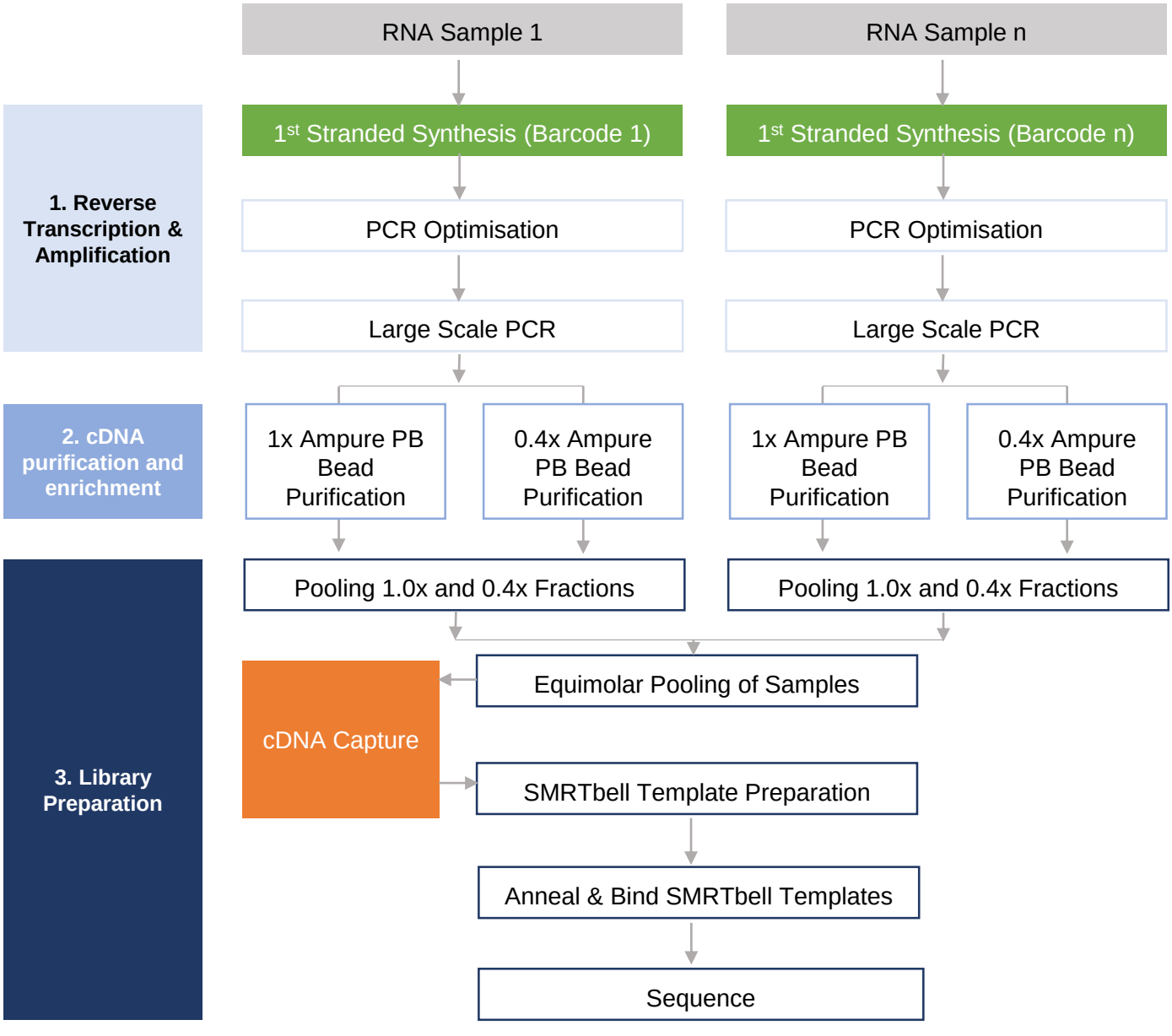
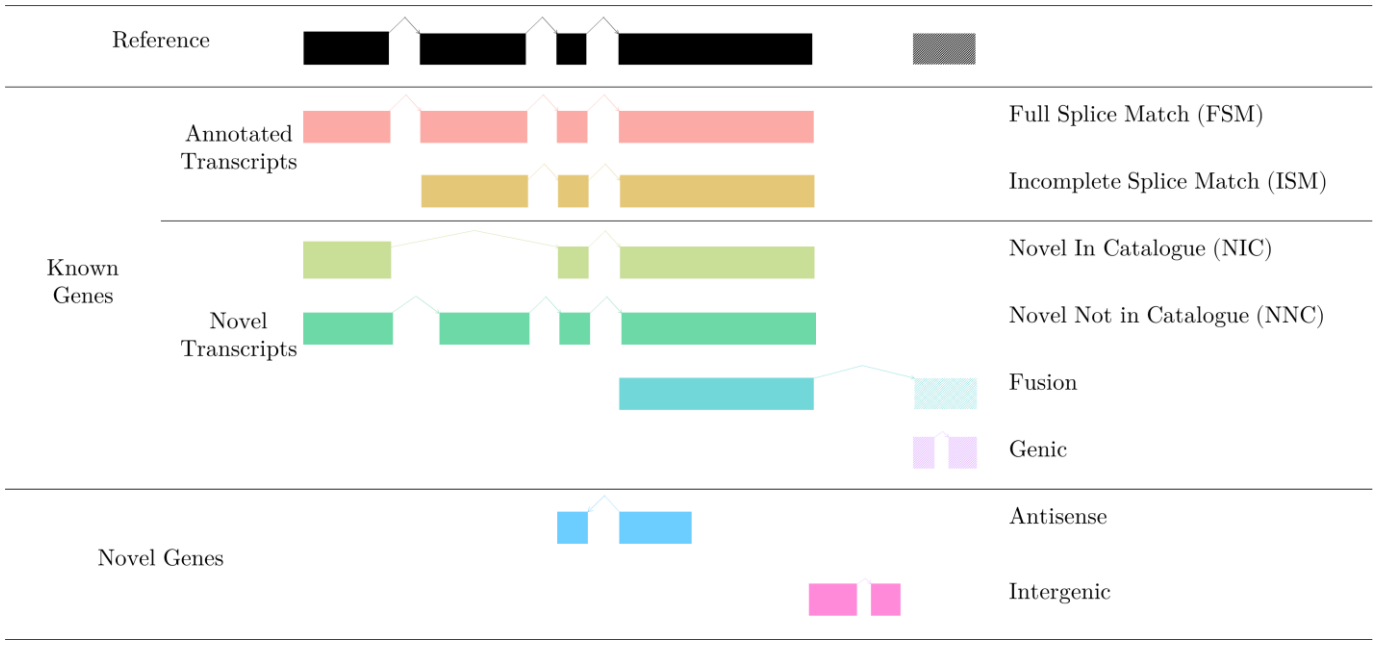
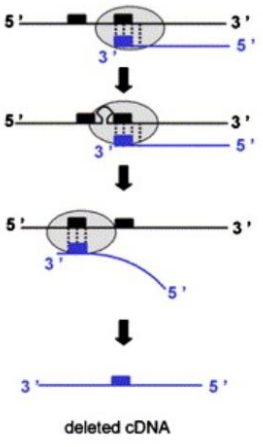

RNA Sample 1
1st Stranded Synthesis
PCR Optimisation
Large Scale PCR
1x Ampure PB Bead Purification
0.4x Ampure PB Bead Purification
Pooling 1.0x and 0.4x Fractions
Equimolar Pooling of Samples
SMRTbell Template Preparation
Anneal & Bind SMRTbell Templates
Sequence
1. Reverse Transcription & Amplification
2. cDNA purification and enrichment
3. Library Preparation

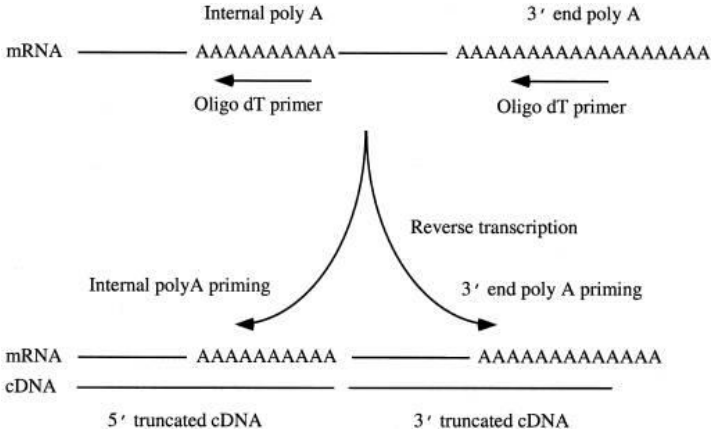

1. Reverse Transcription & Amplification
2. cDNA purification and enrichment
3. Library Preparation
RNA Sample 1
RNA Sample n
1st Stranded Synthesis (Barcode 1)
1st Stranded Synthesis (Barcode n)
PCR Optimisation
PCR Optimisation
Large Scale PCR
Large Scale PCR
1x Ampure PB Bead Purification
0.4x Ampure PB Bead Purification
1x Ampure PB Bead Purification
0.4x Ampure PB Bead Purification
Pooling 1.0x and 0.4x Fractions
Pooling 1.0x and 0.4x Fractions
Equimolar Pooling of Samples
cDNA Capture
SMRTbell Template Preparation
Anneal & Bind SMRTbell Templates
Sequence


Reference
Annotated Transcripts
Known Genes
Novel Transcripts
Novel Genes
Full Splice Match (FSM)
Incomplete Splice Match (ISM)
Novel In Catalogue (NIC)
Novel Not in Catalogue (NNC)
Fusion
Genic
Antisense
Intergenic

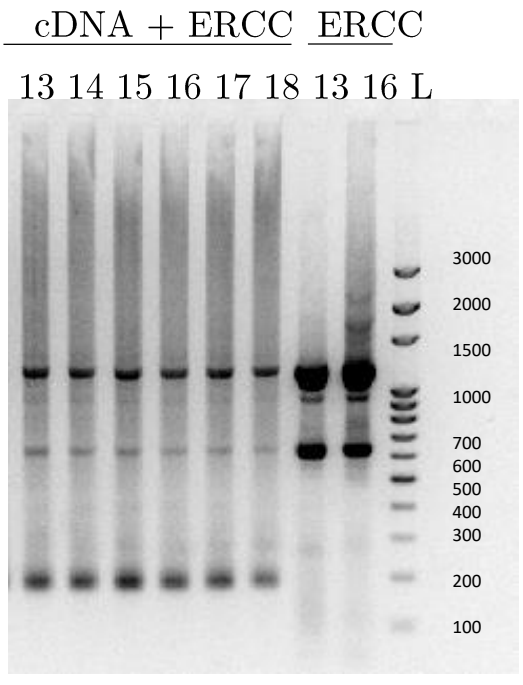
A


5'
3'
3'
5'
deleted cDNA

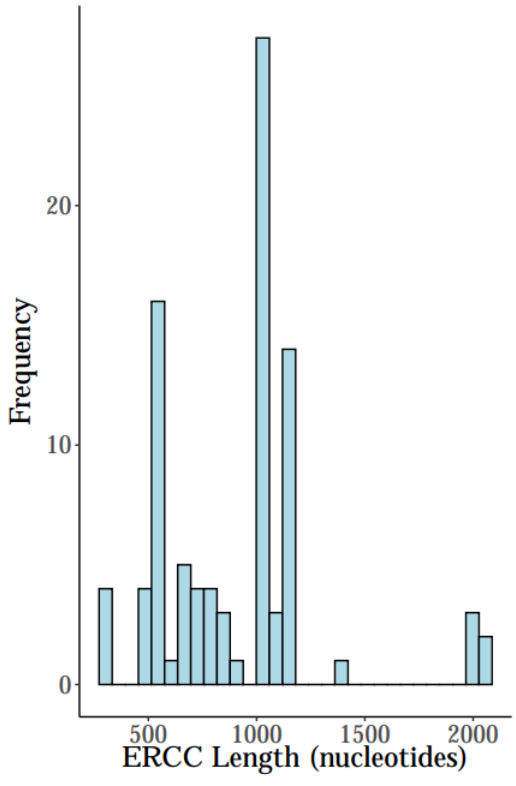
B


Internal poly A
3′ end poly A
mRNA
AAAAAAAAAA
AAAAAAAAAAAAAAAAAA
Oligo dT primer
Oligo dT primer
Reverse transcription
Internal polyA priming
3′ end poly A priming
mRNA
AAAAAAAAAA
AAAAAAAAAAAAA
cDNA
5′ truncated cDNA
3′ truncated cDNA

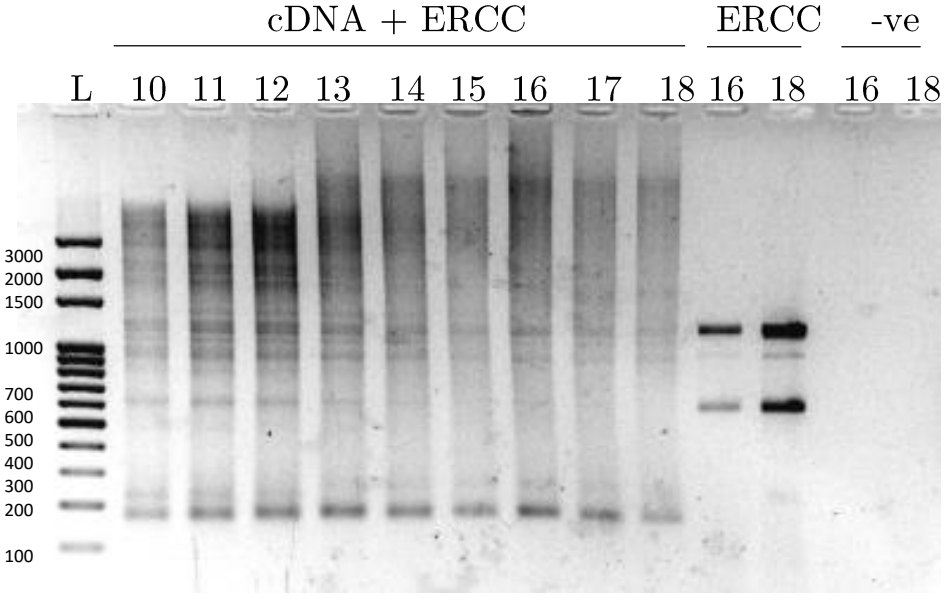
a

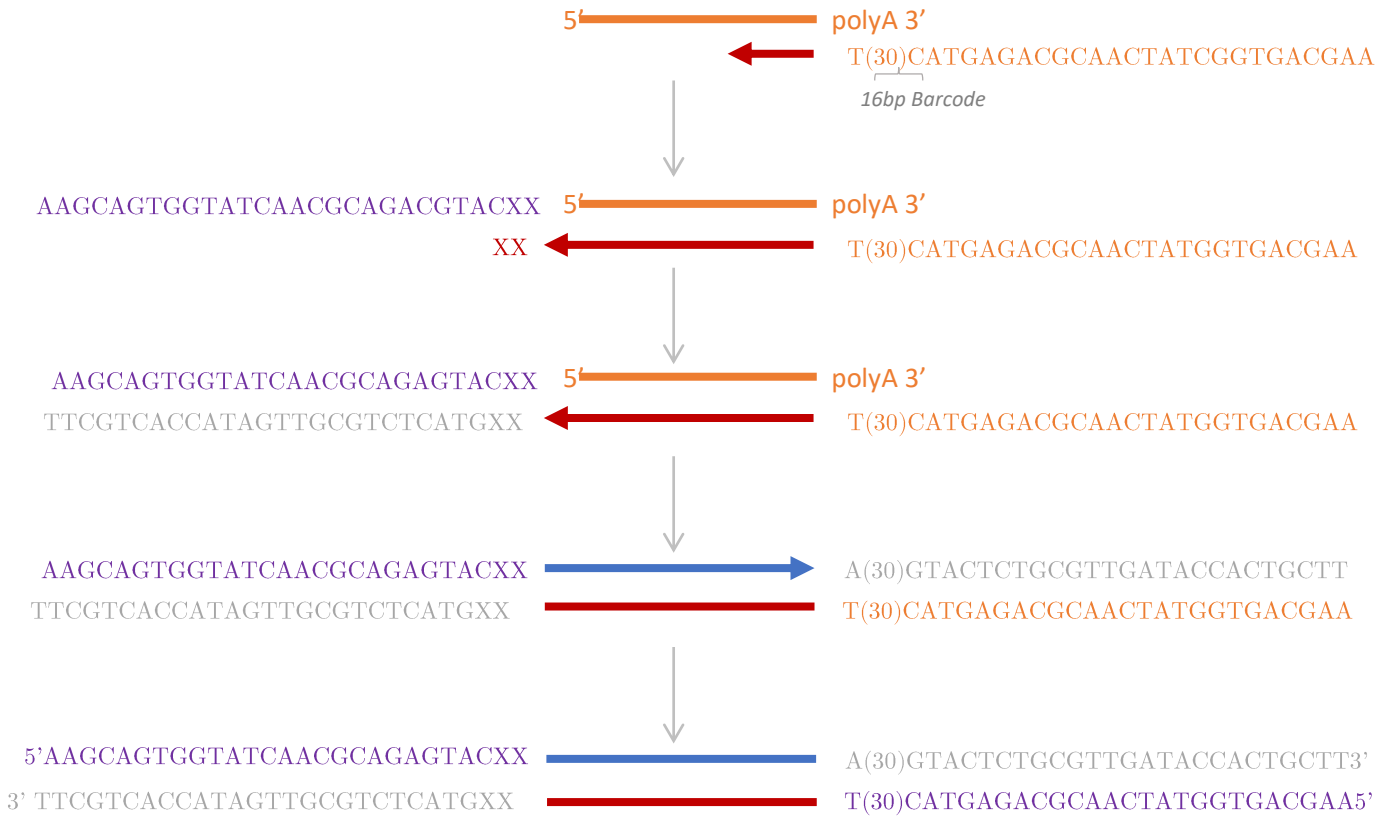

cDNA + ERCC
ERCC
13 14 15 16 17 18 13 16 L
3000
2000
1500
1000
700
600
500
400
300
200
100

b


Frequency
20
10
0
500
1000
1500
2000
ERCC Length (nucleotides)

c


cDNA + ERCC
ERCC
-ve
L 10 11 12 13 14 15 16 17 18 16 18 16 18
3000
2000
1500
1000
700
600
500
400
300
200
100


5'
polyA 3'
T(30)CATGAGACGCAACTATCGGTGACGAA
16bp Barcode
AAGCAGTGGTATCAACGCAGACGTACXX
5'
polyA 3'
XX
T(30)CATGAGACGCAACTATGGTGACGAA
AAGCAGTGGTATCAACGCAGAGTACXX
5'
polyA 3'
TTCGTCACCATAGTTGCGTCTCATGXX
T(30)CATGAGACGCAACTATGGTGACGAA
AAGCAGTGGTATCAACGCAGAGTACXX
A(30)GTACTCTGCGTTGATACCACTGCTT
TTCGTCACCATAGTTGCGTCTCATGXX
T(30)CATGAGACGCAACTATGGTGACGAA
5'AAGCAGTGGTATCAACGCAGAGTACXX
A(30)GTACTCTGCGTTGATACCACTGCTT3'
3' TTCGTCACCATAGTTGCGTCTCATGXX
T(30)CATGAGACGCAACTATGGTGACGAA5'

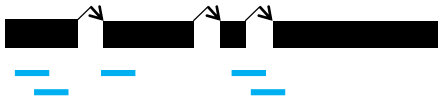
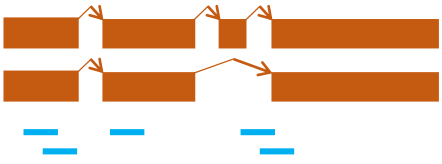
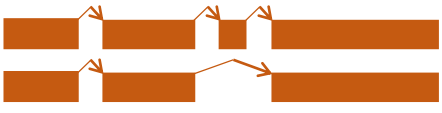
A

B Without Multiplexing

Strand	Sequence
Plus strand start	AATGTACTTCGTTCAGTTACGTATTGCTAAGCAGTGGTATCAACGCAGAGTACATGGG
Plus strand end	AAAAAAAAGTACTCTGCGTTGATACCACTGCTT
Minus strand start	AATGTACTTCGTTCAGTTACGTATTGCTAAGCAGTGGTATCAACGCAGAGTACTTTTTTTTT
Minus strand end	CCCATGTACTCTGCGTTGATACCACTGCTT

C With Multiplexing

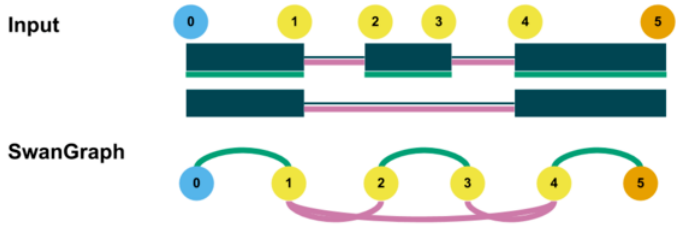
Strand	Sequence
Plus strand start	AATGTACTTCGTTCAGTTACGTATTGCTAAGCAGTGGTATCAACGCAGAGTACATGGG
Plus strand end	AAAAAAAA<Barcode>GTACTCTGCGTTGATACCACTGCTT
Minus strand start	AATGTACTTCGTTCAGTTACGTATTGCT<Barcode>AAGCAGTGGTATCAACGCAGAGTACTTTTTTTT
Minus strand end	CCCATGTACTCTGCGTTGATACCACTGCTT

	Analysis	Annotation	Expression
A		Reference genome	RNA-Seq reads
B		Iso-Seq defined transcriptome	RNA-Seq reads
C		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

a)

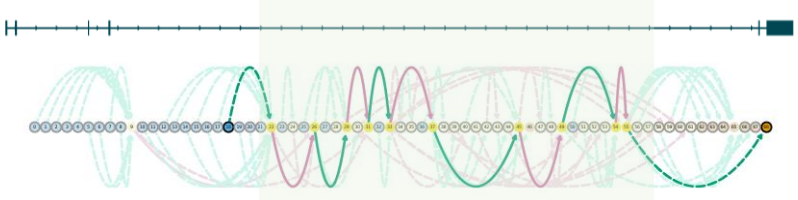
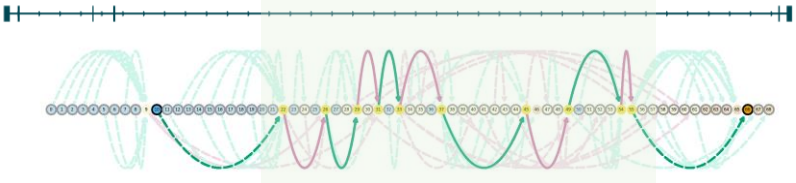
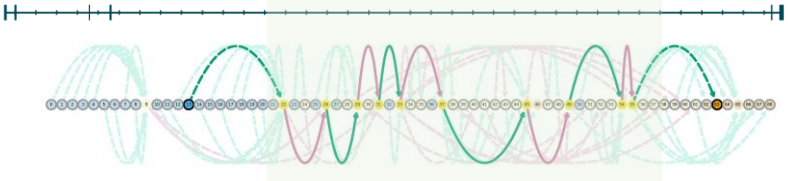
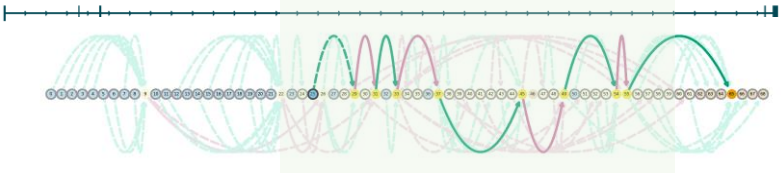
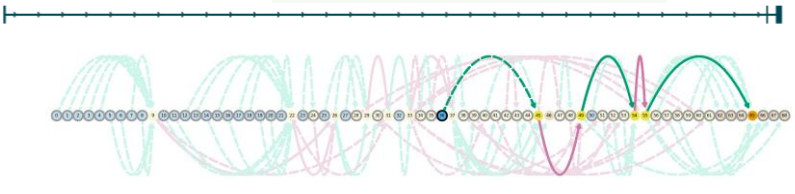
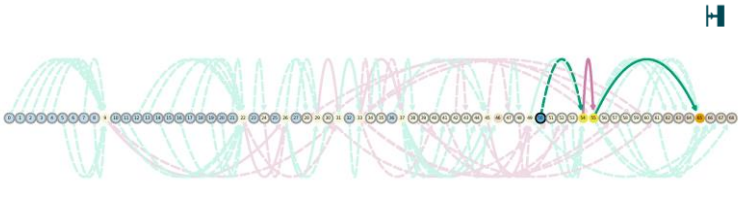
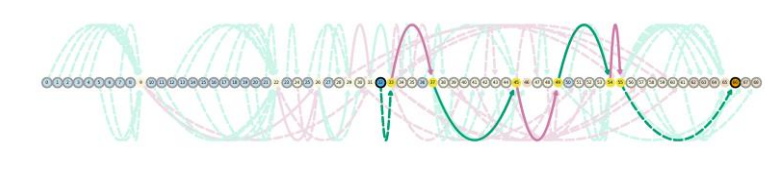


Blue Nodes: Transcription Start Sites
Orange Nodes: Transcription End Sites
Yellow Nodes: Internal Spice Sites

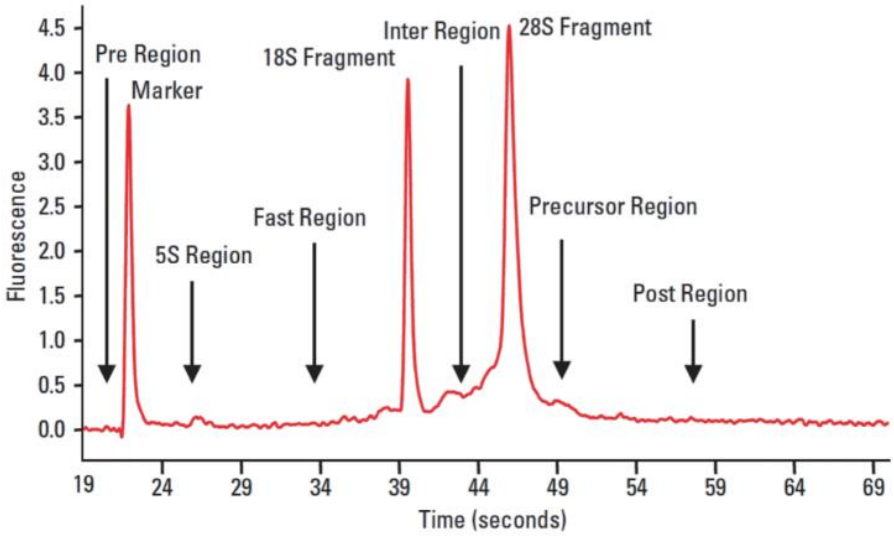
Green Edges: splice junction pair spanning exon
Pink Edges: splice junction pair spanning intron

b)

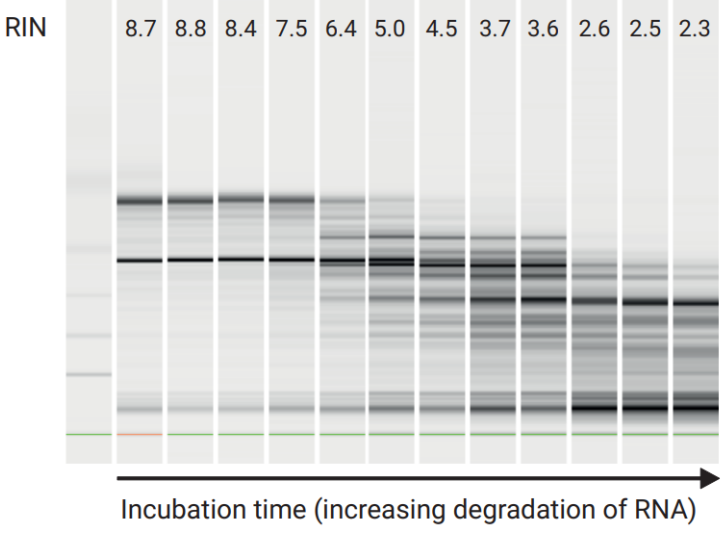
ENMUST00000014268.4

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	

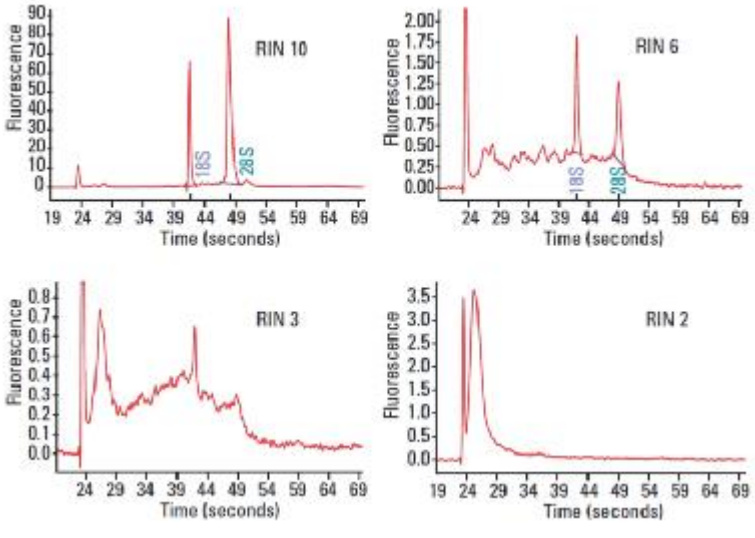
a)


Pre Region
Marker
5S Region
18S Fragment
Fast Region
Inter Region
28S Fragment
Precursor Region
Post Region
Fluorescence
Time (seconds)

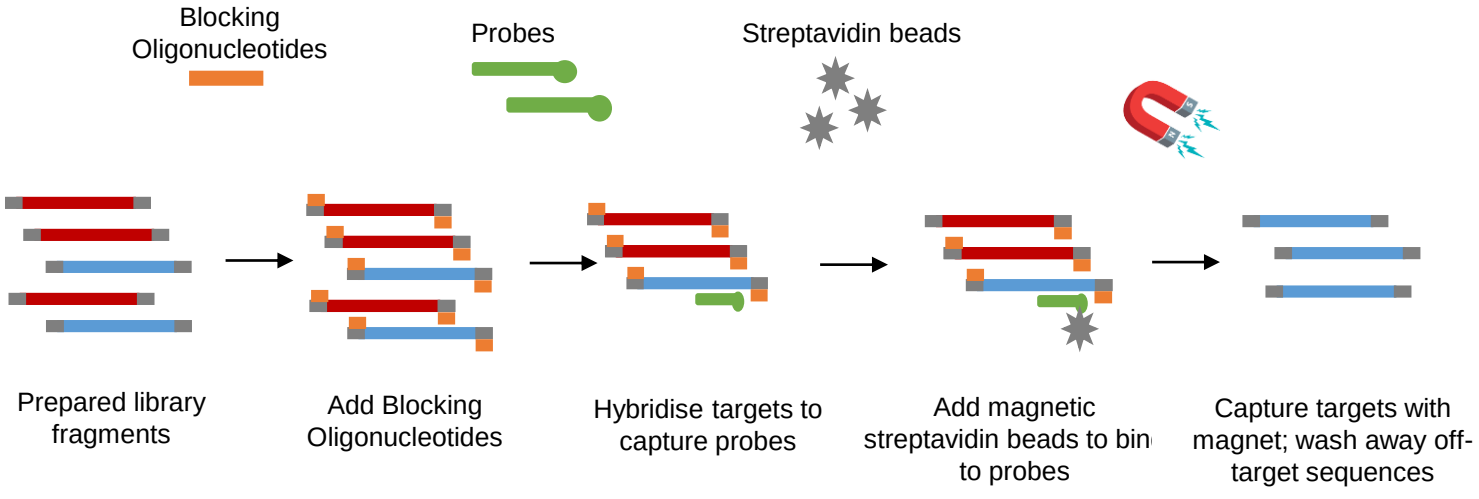
b)


RIN
8.7 8.8 8.4 7.5 6.4 5.0 4.5 3.7 3.6 2.6 2.5 2.3
Incubation time (increasing degradation of RNA)

c)

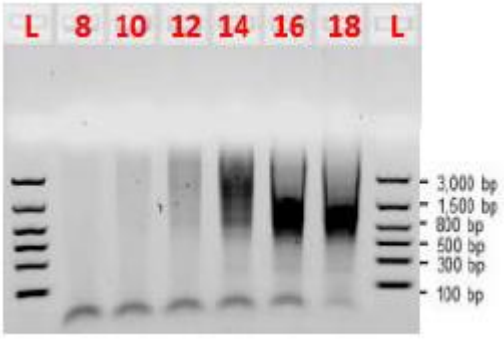

RIN 10
RIN 6
RIN 3
RIN 2
18S
28S
Fluorescence
Time (seconds)

A

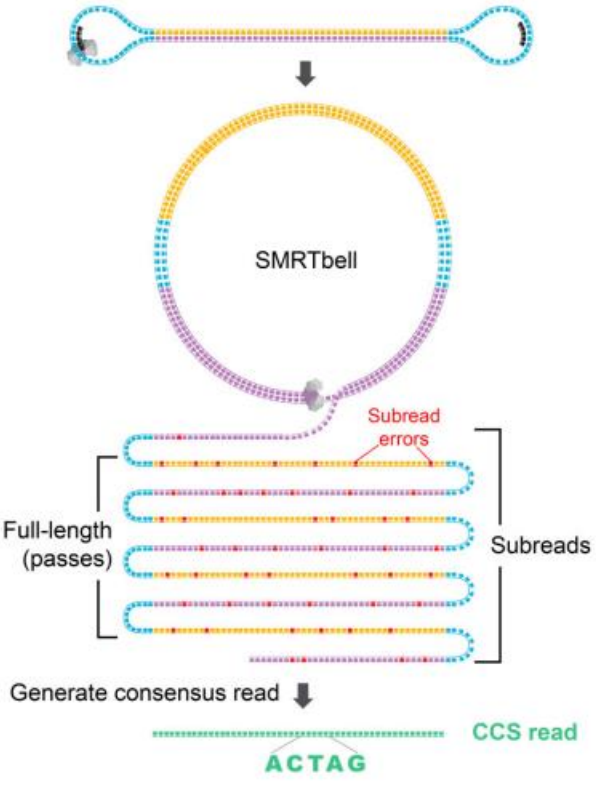

Blocking Oligonucleotides
Probes
Streptavidin beads
Prepared library fragments
Add Blocking Oligonucleotides
Hybridise targets to capture probes
Add magnetic streptavidin beads to bin to probes
Capture targets with magnet; wash away off-target sequences

B

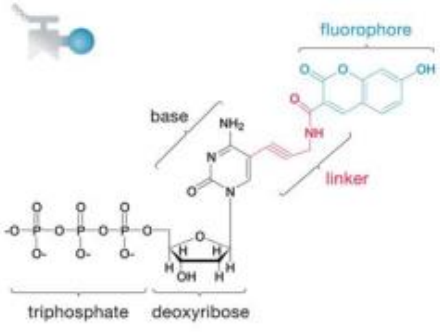

1
Combine DNA with blockers
Dry down DNA
Perform hybridization reaction
Total time: 15 minutes
Total time: Variable
Total time: 4–16 hours
2
Prepare buffers
Total time: 15 minutes*
3
Wash streptavidin beads
Total time: 15 minutes*
4
Perform bead capture
Total time: 45 minutes
5
Perform washes
Total time: 30 minutes
6
Perform post-capture PCR
Total time: 30 minutes
7
Purify post-capture PCR fragments
Total time: 30 minutes


L
8
10
12
14
16
18
L
3,000 bp
1,500 bp
800 bp
500 bp
300 bp
100 bp

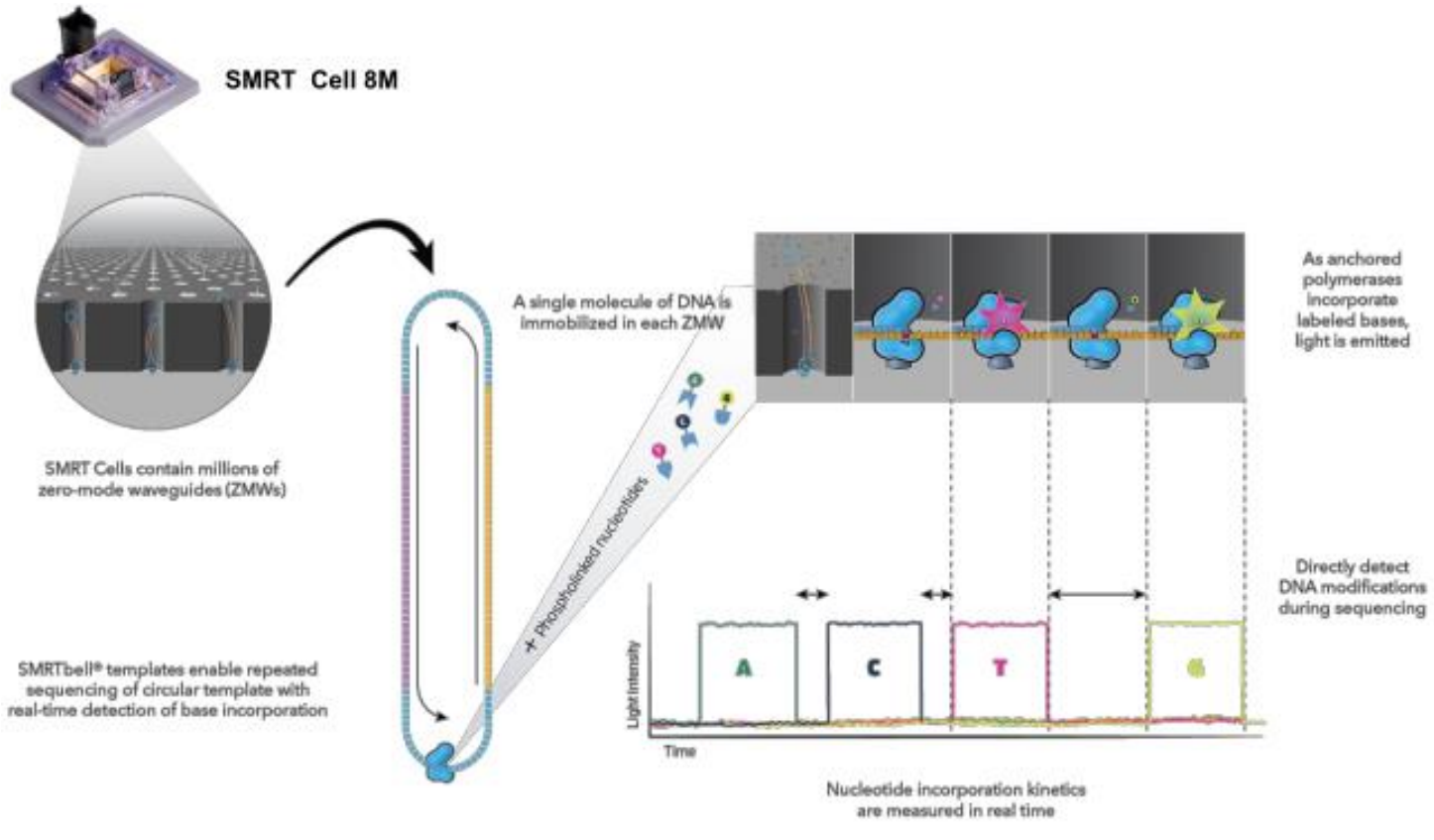
A

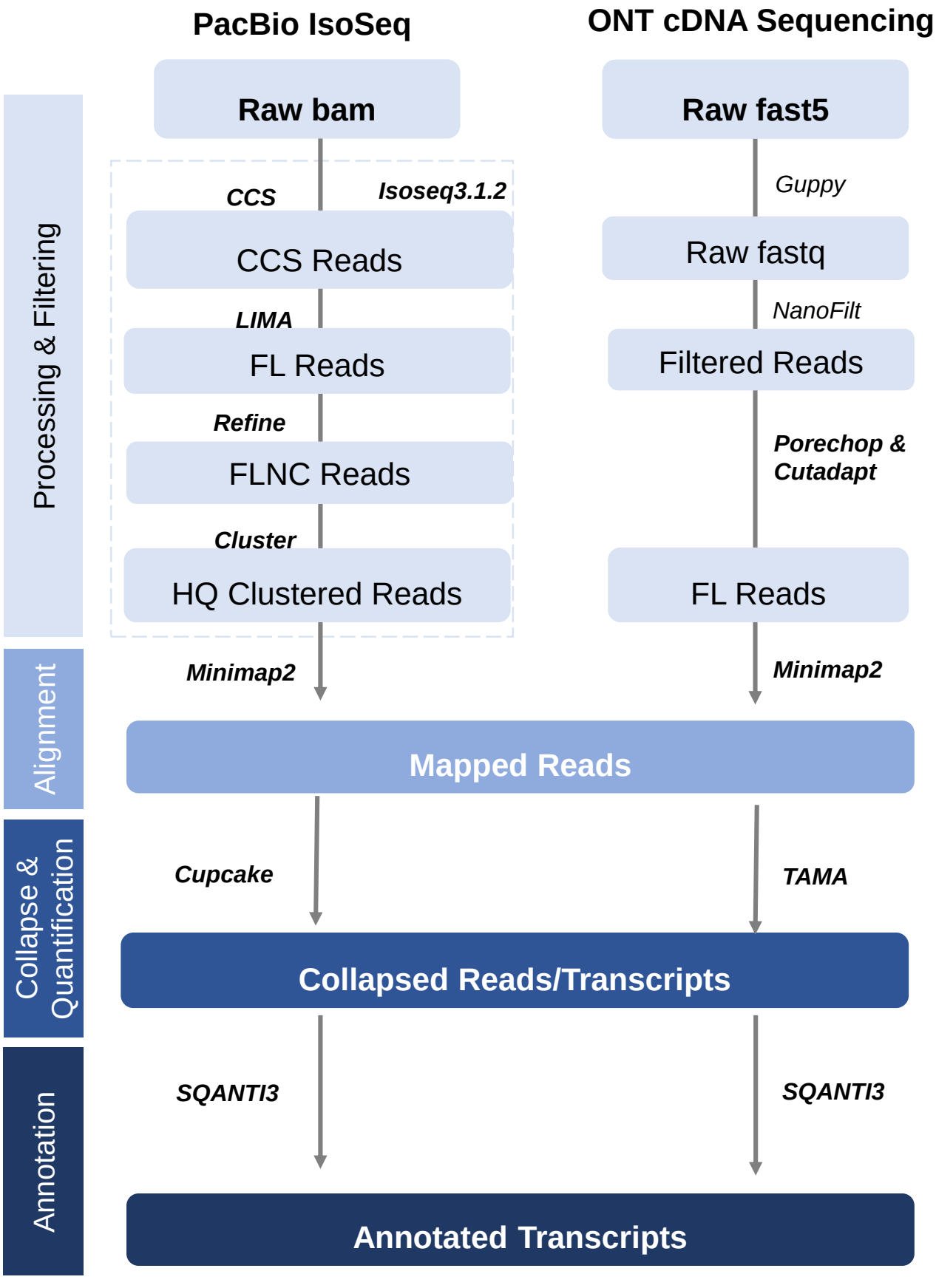

SMRTbell
Subread errors
Full-length (passes)
Subreads
Generate consensus read
CCS read
ACTAG

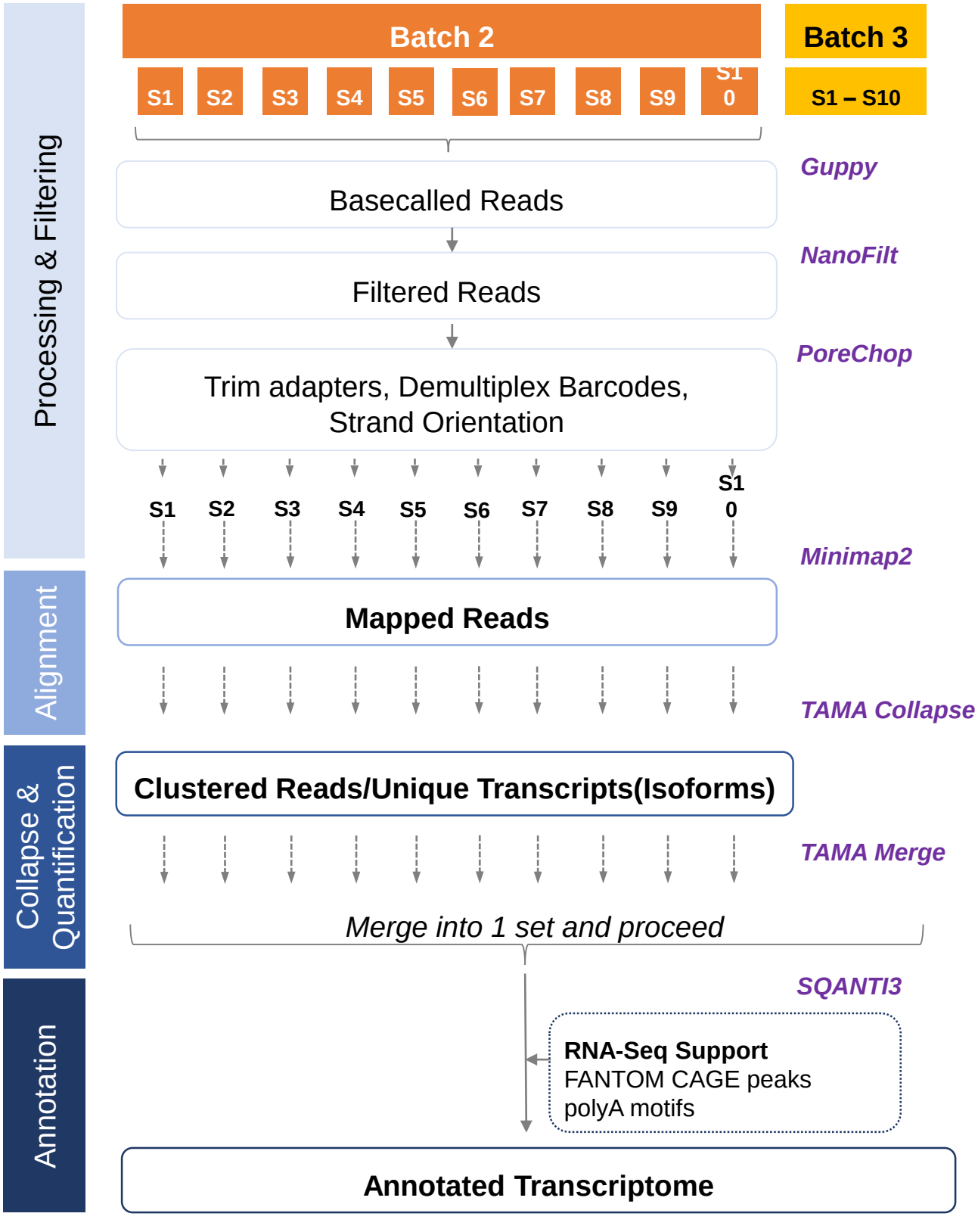
B


fluorophore
base
linker
triphosphate
deoxyribose

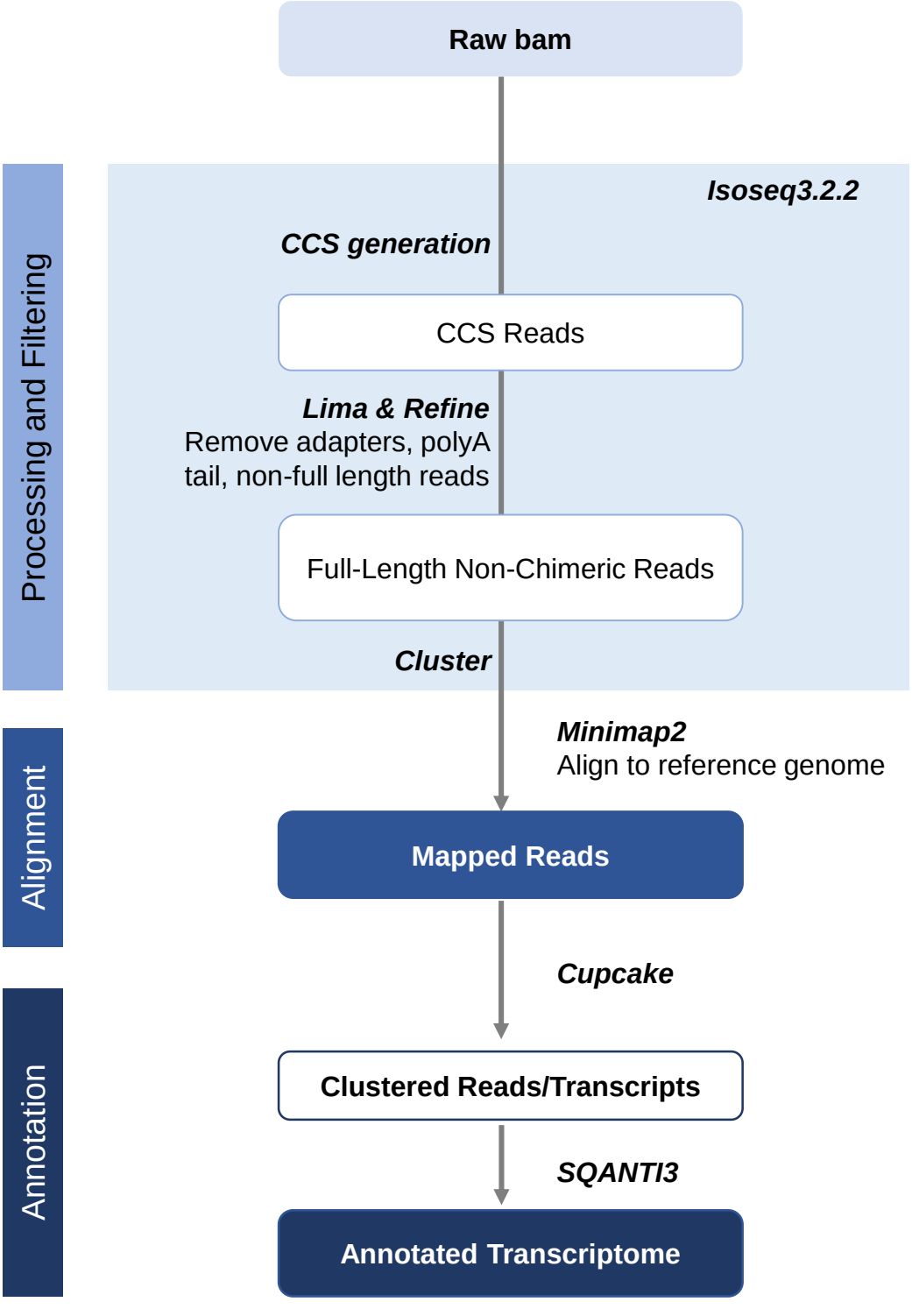
C

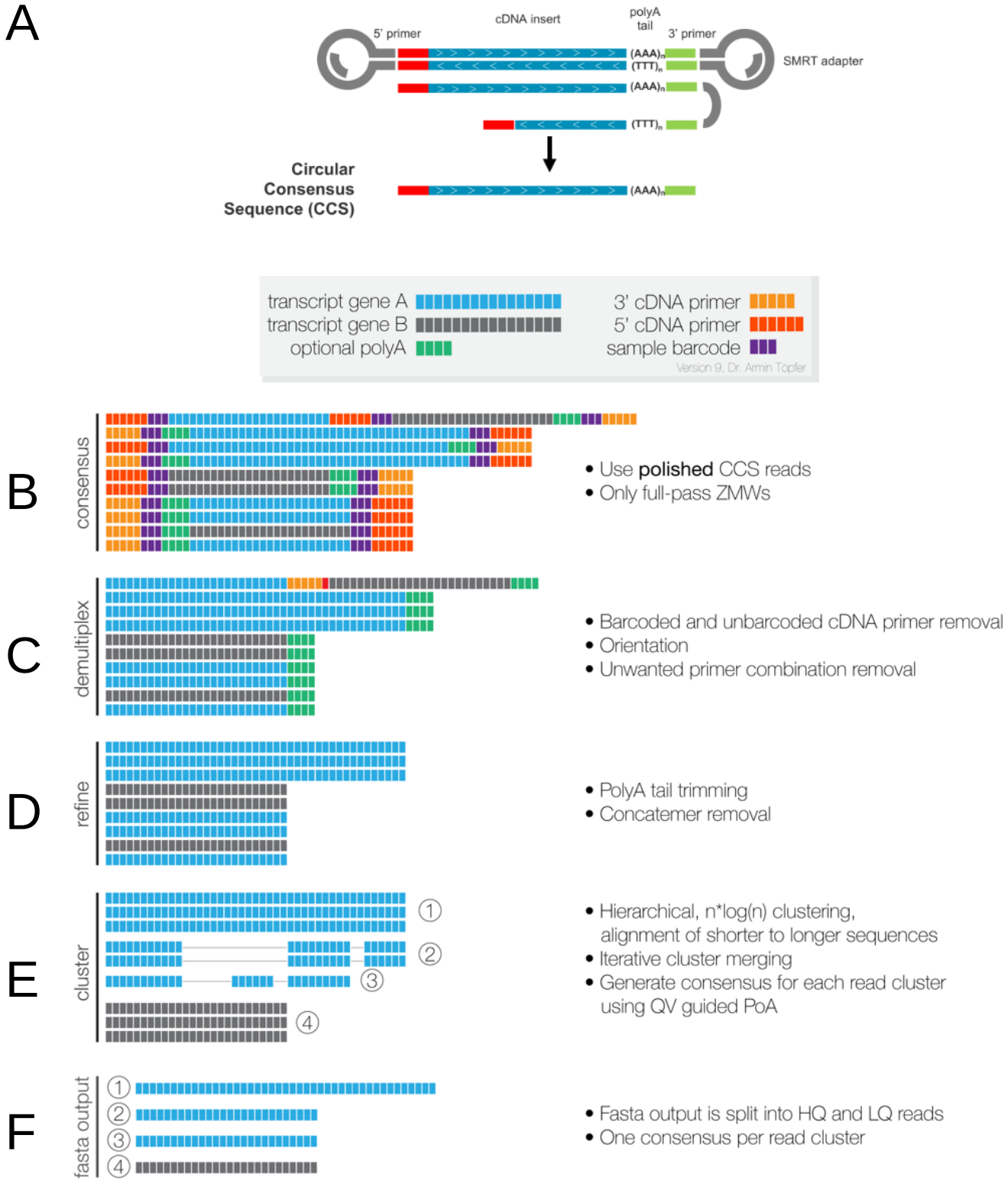

SMRT Cell 8M
SMRT Cells contain millions of zero-mode waveguides (ZMWs)
A single molecule of DNA is immobilized in each ZMW
Phospholinked nucleotides
SMRTbell® templates enable repeated sequencing of circular template with real-time detection of base incorporation
As anchored polymerases incorporate labeled bases, light is emitted
Directly detect DNA modifications during sequencing
Light Intensity
Time
A
C
T
G
Nucleotide incorporation kinetics are measured in real time


PacBio IsoSeq
ONT cDNA Sequencing
Raw bam
Raw fast5
Processing & Filtering
CCS
Isoseq3.1.2
Guppy
CCS Reads
Raw fastq
LIMA
NanoFilt
FL Reads
Filtered Reads
Refine
Porechop &
Cutadapt
FLNC Reads
Cluster
HQ Clustered Reads
FL Reads
Minimap2
Minimap2
Alignment
Mapped Reads
Collapse &
Quantification
Cupcake
TAMA
Collapsed Reads/Transcripts
Annotation
SQANTI3
SQANTI3
Annotated Transcripts


Processing & Filtering
Alignment
Collapse & Quantification
Annotation
Batch 2
Batch 3
S1
S2
S3
S4
S5
S6
S7
S8
S9
S10
S1 – S10
Guppy
Basecalled Reads
NanoFilt
Filtered Reads
PoreChop
Trim adapters, Demultiplex Barcodes, Strand Orientation
Minimap2
Mapped Reads
TAMA Collapse
Clustered Reads/Unique Transcripts(Isoforms)
TAMA Merge
Merge into 1 set and proceed
SQANTI3
RNA-Seq Support
FANTOM CAGE peaks
polyA motifs
Annotated Transcriptome

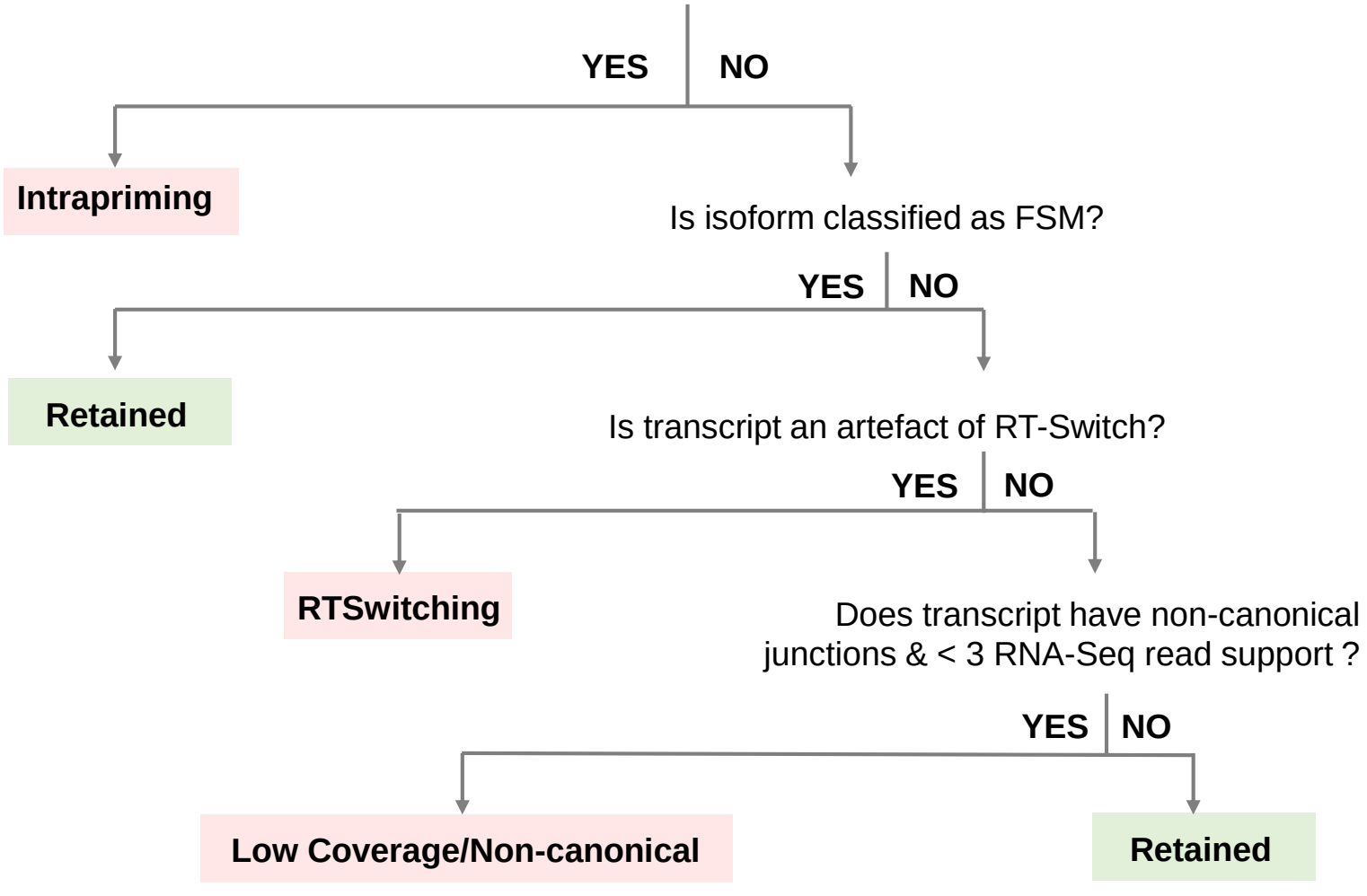
PacBio Iso-Seq Bioinformatics Pipeline




A
5' primer
cDNA insert
polyA tail
3' primer
SMRT adapter
Circular Consensus Sequence (CCS)
transcript gene A
transcript gene B
optional polyA
3' cDNA primer
5' cDNA primer
sample barcode
Version 9, Dr. Armin Töpfer
Iso-Seq3
B
consensus
• Use polished CCS reads
• Only full-pass ZMWs
C
demultiplex
• Barcoded and unbarcoded cDNA primer removal
• Orientation
• Unwanted primer combination removal
D
refine
• PolyA tail trimming
• Concatemer removal
E
cluster
• Hierarchical, n*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
• Fasta output is split into HQ and LQ reads
• One consensus per read cluster

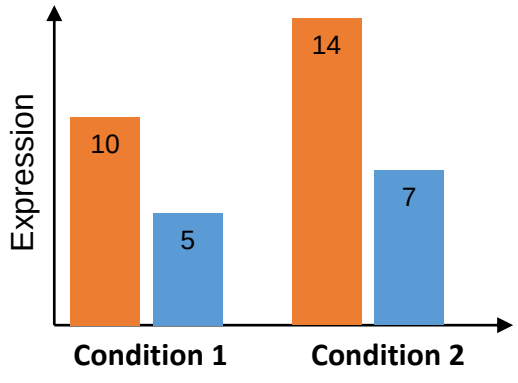
Does the isoform have:

- >60% genomic As in 20bp window downstream of TTS **&**
- unknown distance of query isoform 5'end and reference TSS &
- no detected polyA motif **&**
- **d**istance between query isoform 3' end and reference TTS >50bp



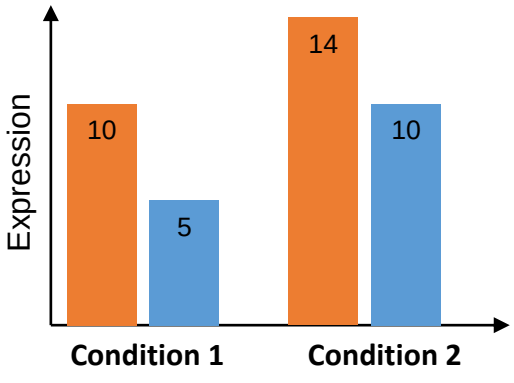
A

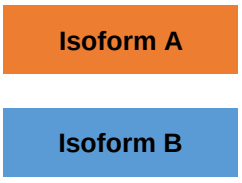
Differential Transcript Expression



B

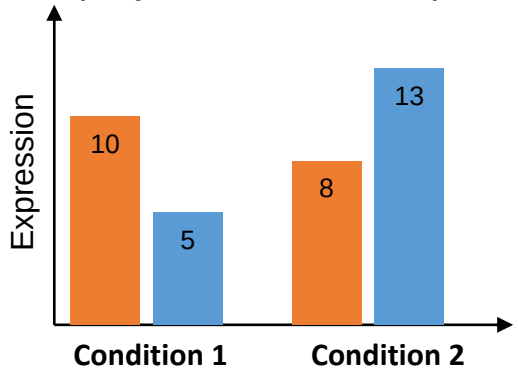
Differential Transcript Usage





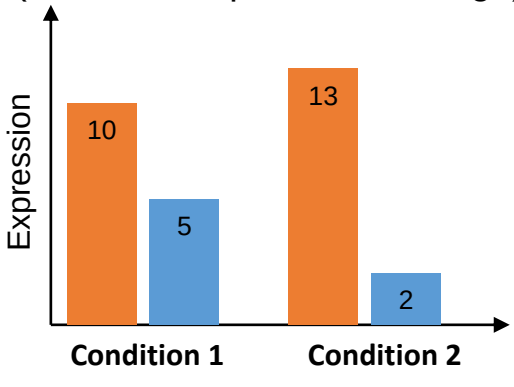
C

Differential Transcript Usage
(Major Isoform Switch)

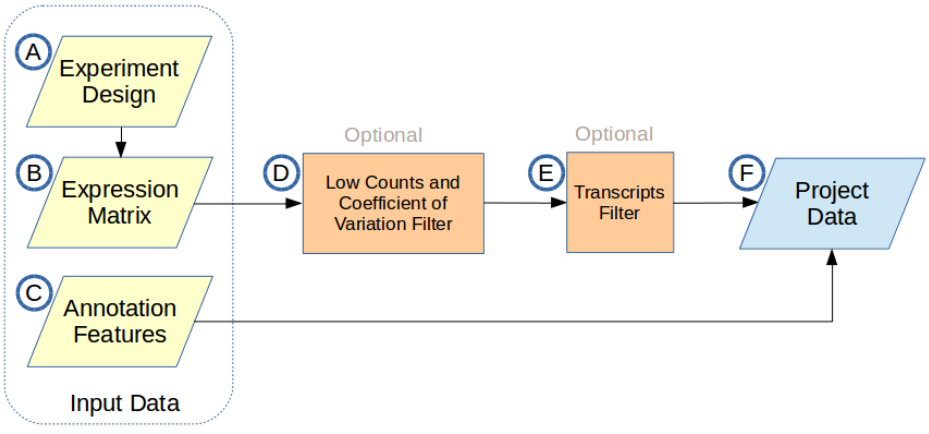


D

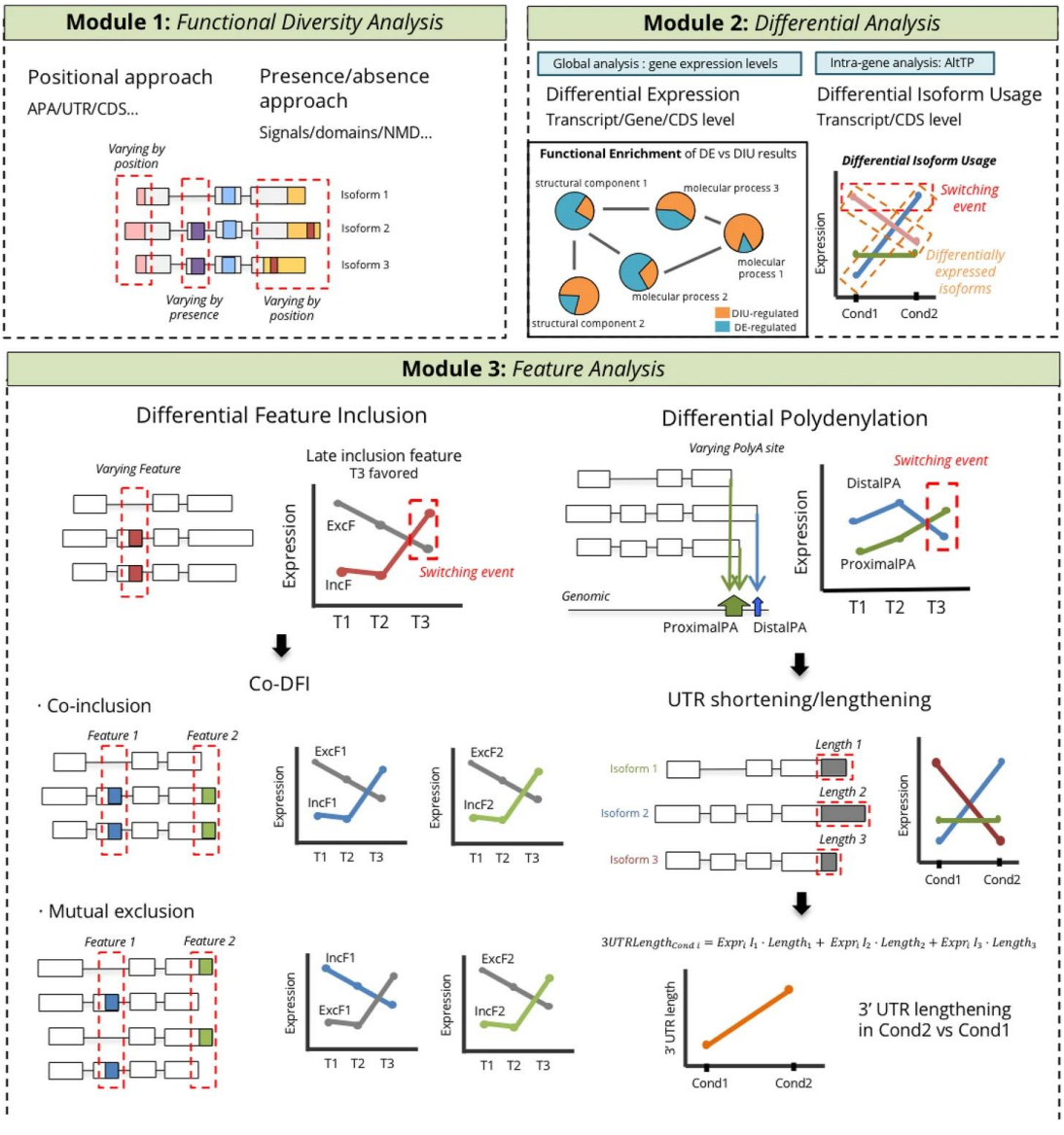
Differential Transcript Usage
(No Gene Expression Change)



A


A
Experiment Design
B
Expression Matrix
C
Annotation Features
Input Data
Optional
D
Low Counts and Coefficient of Variation Filter
Optional
E
Transcripts Filter
F
Project Data

B


Module 1: Functional Diversity Analysis
Positional approach
APA/UTR/CDS...
Presence/absence approach
Signals/domains/NMD...
Varying by position
Isoform 1
Isoform 2
Isoform 3
Varying by presence
Varying by position
Module 2: Differential Analysis
Global analysis : gene expression levels
Intra-gene analysis: AltTP
Differential Expression
Transcript/Gene/CDS level
Differential Isoform Usage
Transcript/CDS level
Functional Enrichment of DE vs DIU results
structural component 1
molecular process 3
molecular process 1
molecular process 2
structural component 2
DIU-regulated
DE-regulated
Differential Isoform Usage
Switching event
Differentially expressed isoforms
Expression
Cond1
Cond2
Module 3: Feature Analysis
Differential Feature Inclusion
Varying Feature
Late inclusion feature
T3 favored
ExcF
IncF
Switching event
Expression
T1 T2 T3
Co-DFI
· Co-inclusion
Feature 1
Feature 2
ExcF1
IncF1
ExcF2
IncF2
· Mutual exclusion
IncF1
ExcF1
ExcF2
IncF2
Differential Polydenylation
Varying PolyA site
Genomic
ProximalPA
DistalPA
Switching event
UTR shortening/lengthening
Length 1
Length 2
Length 3
Isoform 1
Isoform 2
Isoform 3
$3UTRLength_{Cond\,i} = Expr_i\,I_1 \cdot Length_1 + Expr_i\,I_2 \cdot Length_2 + Expr_i\,I_3 \cdot Length_3$
3' UTR length
3' UTR lengthening in Cond2 vs Cond1