Single molecule RNA sequencing of *C. elegans* transcriptome

RACHAEL WORKMAN

TIMP LAB

JOHNS HOPKINS UNIVERSITY

Why compare direct RNA vs cDNA sequencing

Many advantages to direct RNA:

- Poly-A profiling
- Modification detection
- Simplified library preparation could reduce bias and artifacts

But necessary to first understand differences between the data types

What to compare: Direct RNA vs cDNA sequencing

Library preparation
Data quality
Transcript detection
Abundance
Splice variant resolution
Homopolymer calling

Direct RNA and cDNA library prep comparison

cDNA
LSK-108 strand switching
RT: Maxima H Minus, 50C

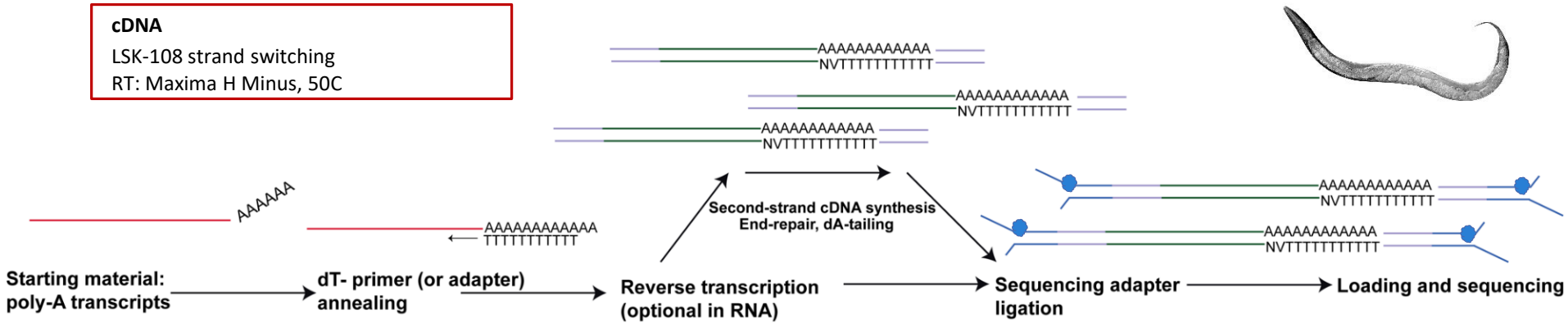
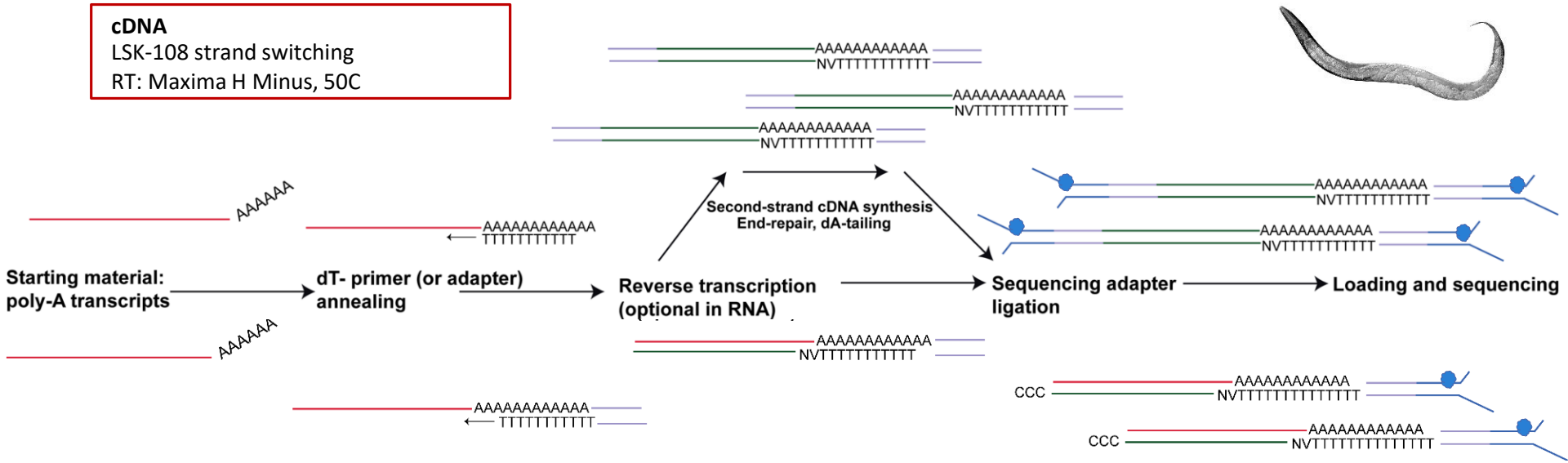


Image courtesy of: Universiteit Utrecht, Developmental Biology

Direct RNA and cDNA library prep comparison

cDNA
LSK-108 strand switching
RT: Maxima H Minus, 50C



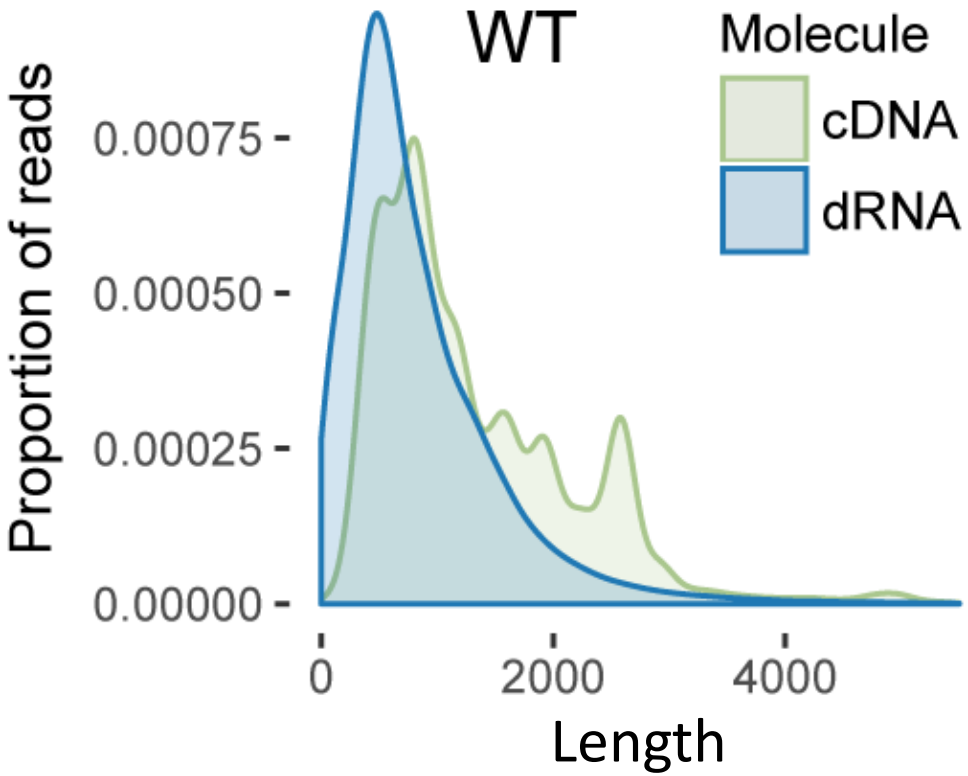
RNA
RNA-001
RT: SuperScript IV, 55C

Image courtesy of: Universiteit Utrecht, Developmental Biology

What to compare: Direct RNA vs cDNA sequencing

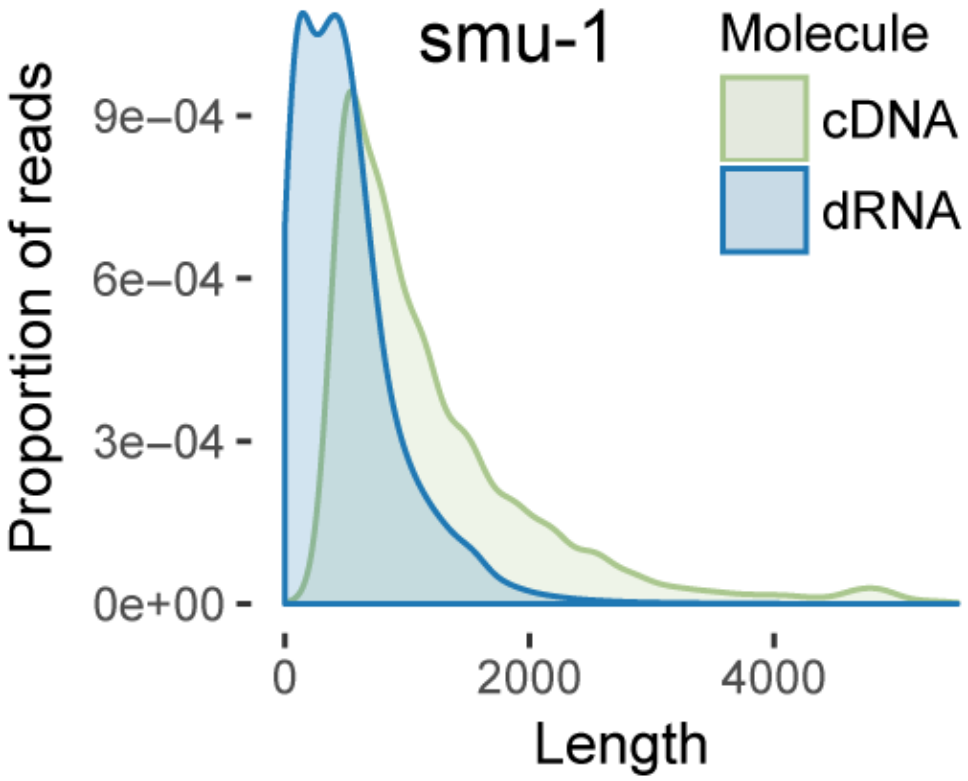
Library preparation
Data quality
Transcript detection
Abundance
Splice variant resolution
Homopolymer calling

Shorter read lengths in RNA, lower yield



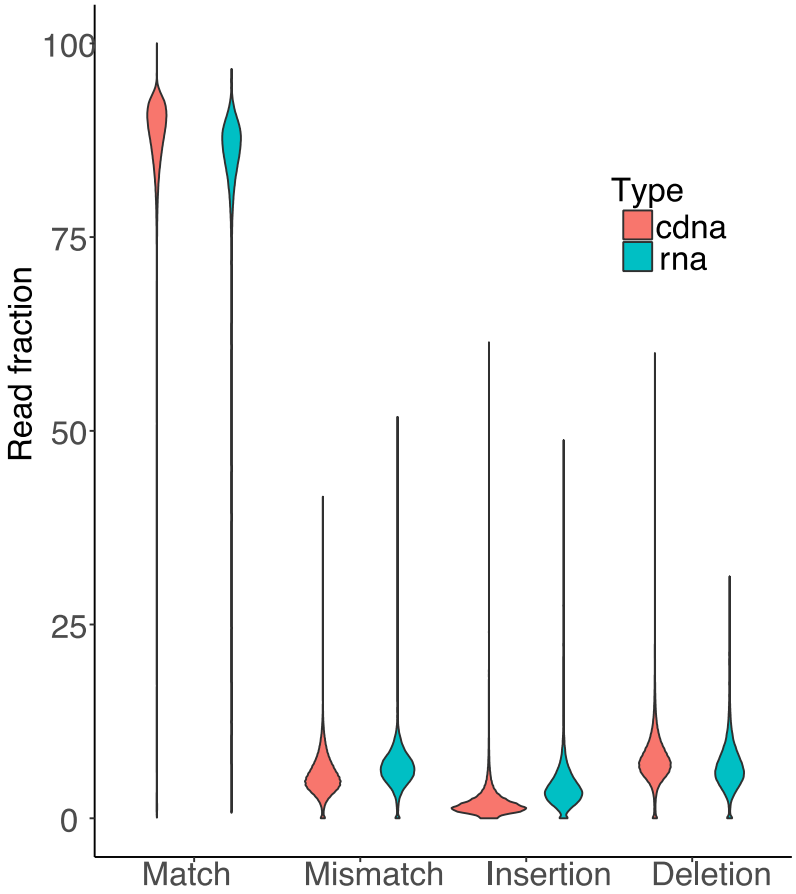
	RNA	cDNA
Reads	240K	2400K
Yield	0.2Gb	3.23Gb
Mean read length	652bp	1340bp

But still sufficient for our analysis



	RNA	cDNA
Reads	913K	987K
Yield	0.53Gb	0.89Gb
Mean read length	431bp	935bp

Alignment quality similar between runs

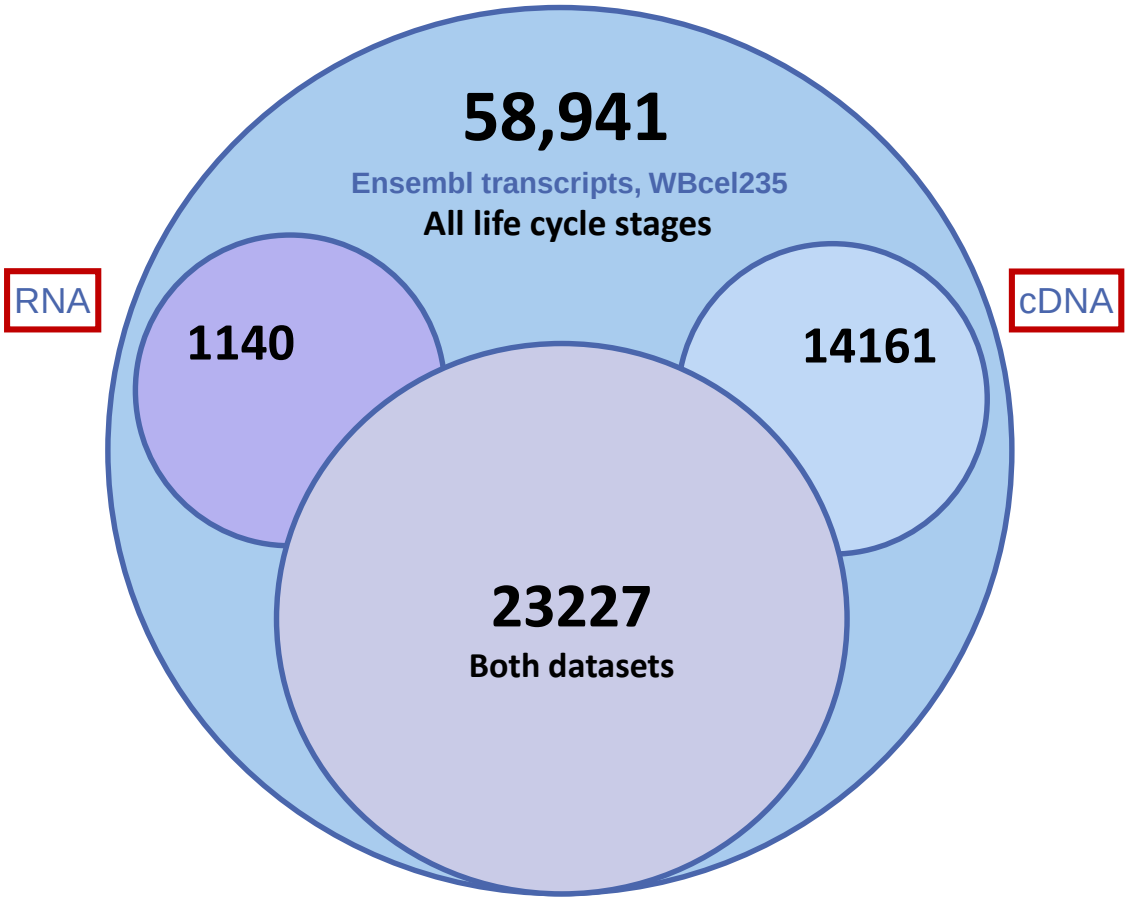


	RNA	cDNA
Alignment	65%	85%
Mapq >10	77K	545K
Mean match len	752bp	1130bp
Median match fraq	82%	87%
% Accuracy	83%	85%

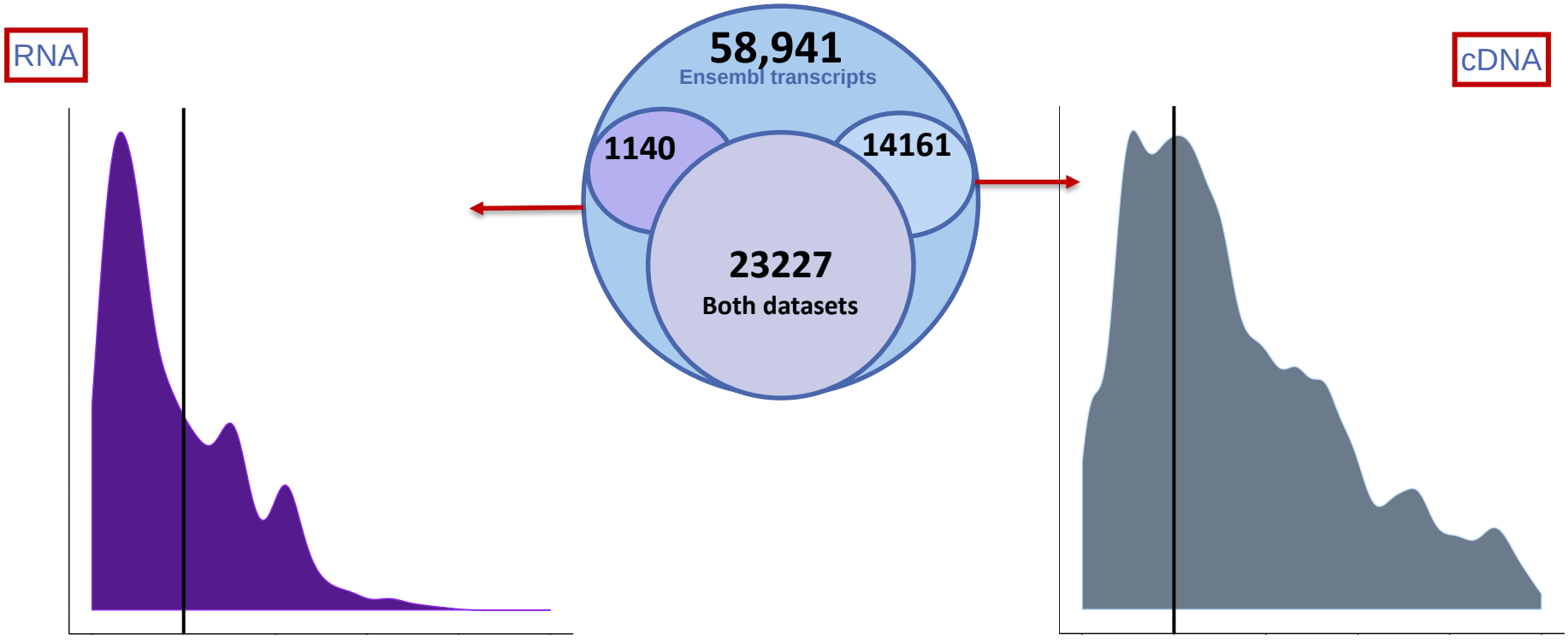
What to compare: Direct RNA vs cDNA sequencing

Library preparation
Data quality
Transcript detection
Abundance
Splice variant resolution
Homopolymer calling

Large portion curated transcripts detected



Large portion curated transcripts detected



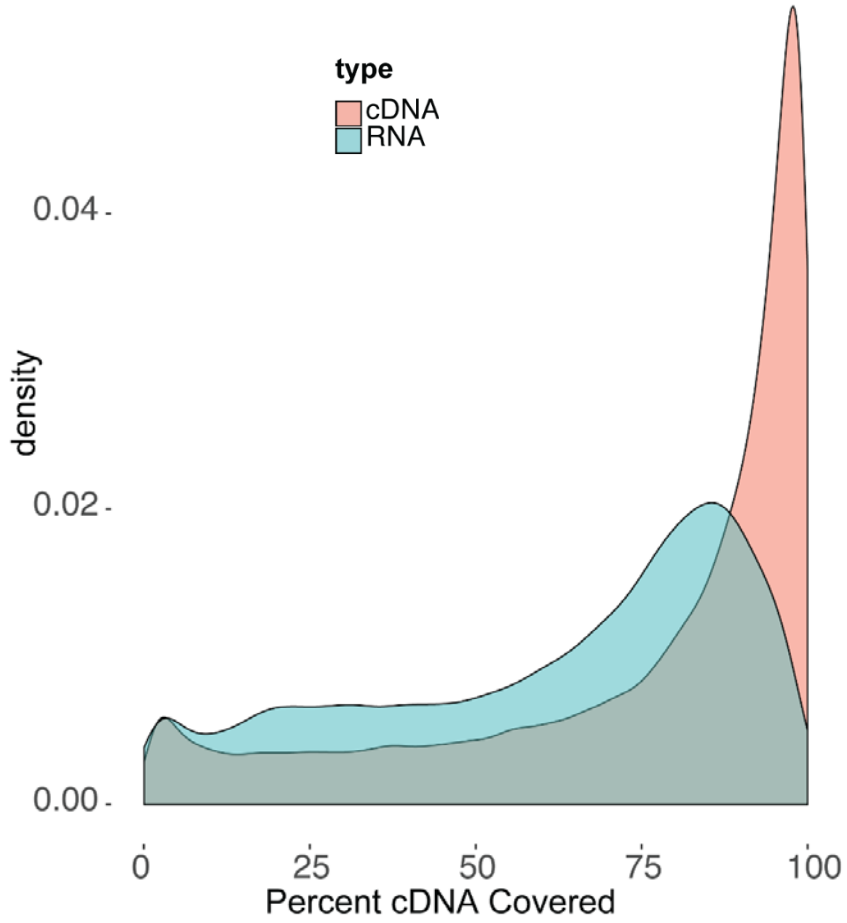
Higher % full length transcripts in cDNA

Pileup of percent transcript covered by each read

More degradation in RNA run, respectable lengths in both

Removing RT step may reduce degradation

Non-full length reads- preparatory degradation, aligner clipping



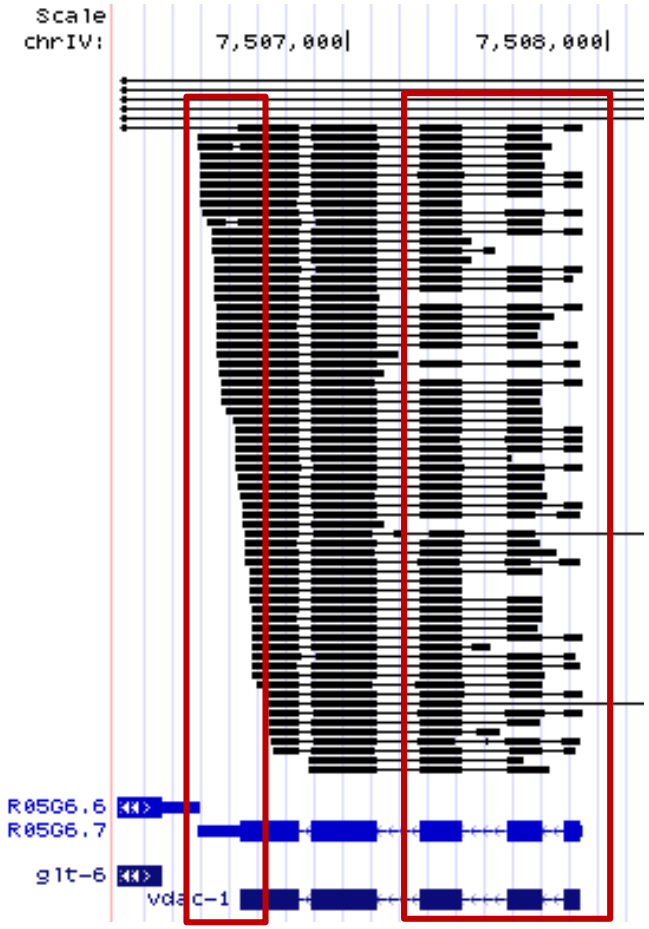
Non-full length reads due to preparation and alignment

Pileup of percent transcript covered by each read

More degradation in RNA run, respectable lengths in both

Removing RT step may reduce degradation

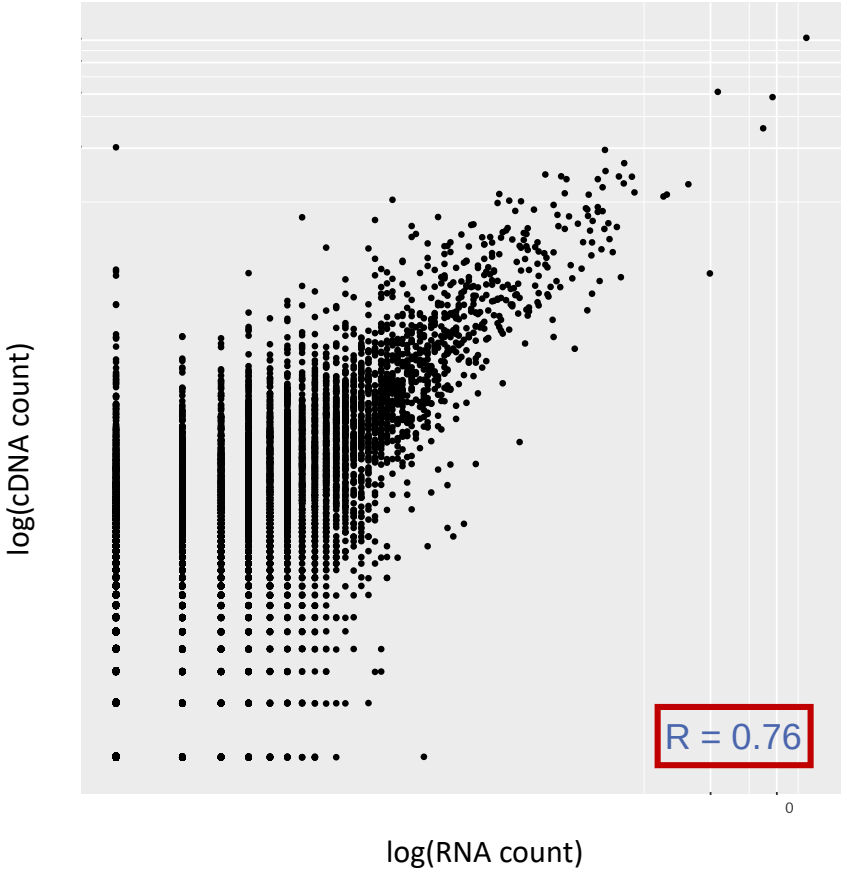
Non-full length reads- preparatory degradation, aligner clipping

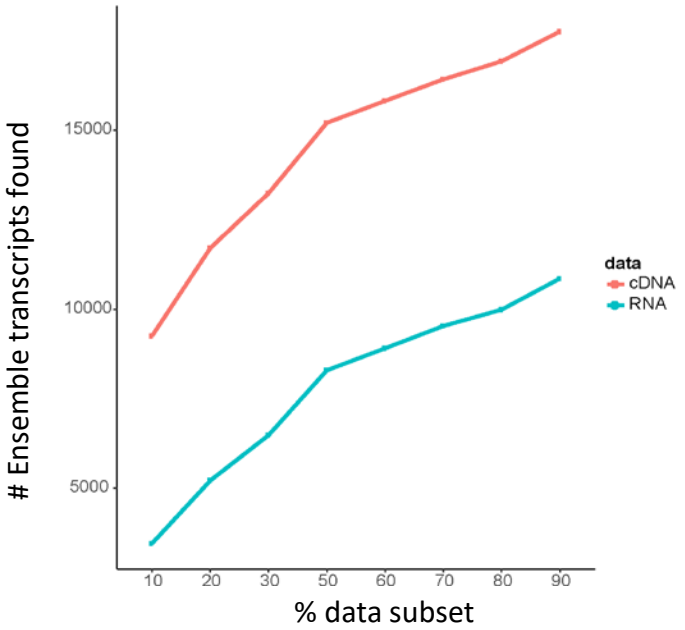


What to compare: Direct RNA vs cDNA sequencing

Library preparation
Data quality
Transcript detection
Abundance
Splice variant resolution
Homopolymer calling

Transcript abundance consistent between cDNA and RNA runs





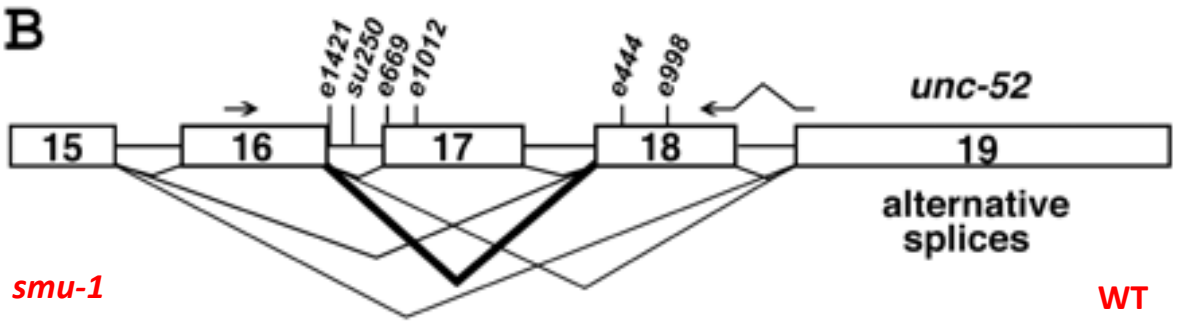
Abundance between runs well correlated, data subsetting shows trending towards saturation mirrored in both types of sequencing

What to compare: Direct RNA vs cDNA sequencing

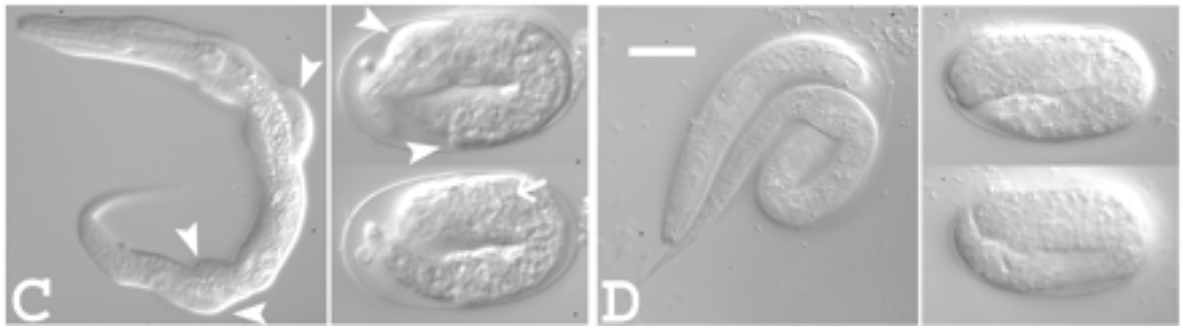
Library preparation
Data quality
Transcript detection
Abundance
Splice variant resolution
Homopolymer calling

Splice mutant *smu-1*

smu-1 gene enhances exon 17 skipping of unc-52, which encodes a set of perlecan homologs (basement membrane proteins)
-- Homolog of spliceosome-associated protein fSAP57
--Leads to mechanosensory and chemosensory defects
-- Spike et al found 3.5X increase in 16-18-19 isoform abundance

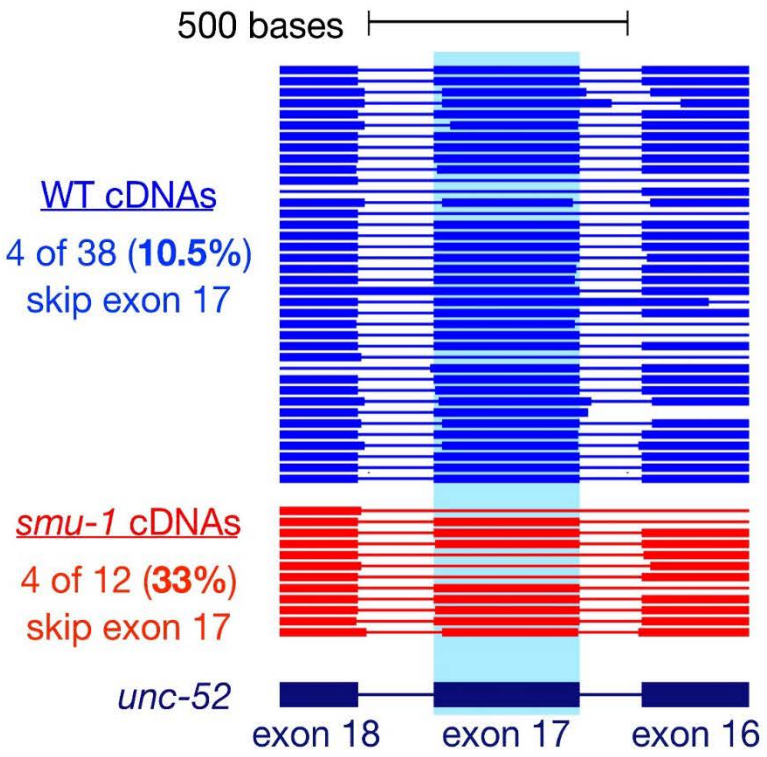


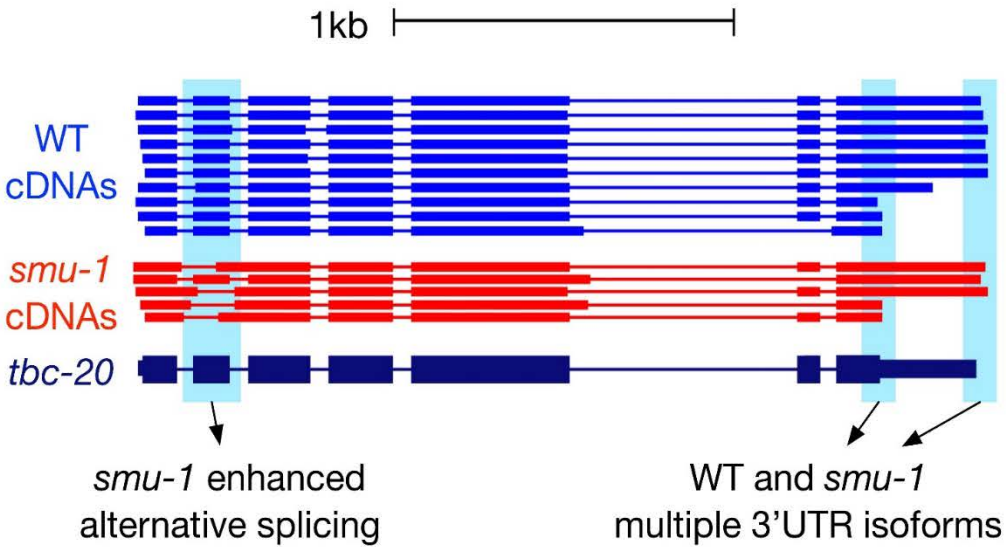
Hypodermal bulges
Early pharynx development



Spike et al 2001, Mol. And Cell Bio
Spartz et al 2004, Mol Cell Bio

Evidence of enhanced exon 17 skipping in both direct RNA and cDNA preparations



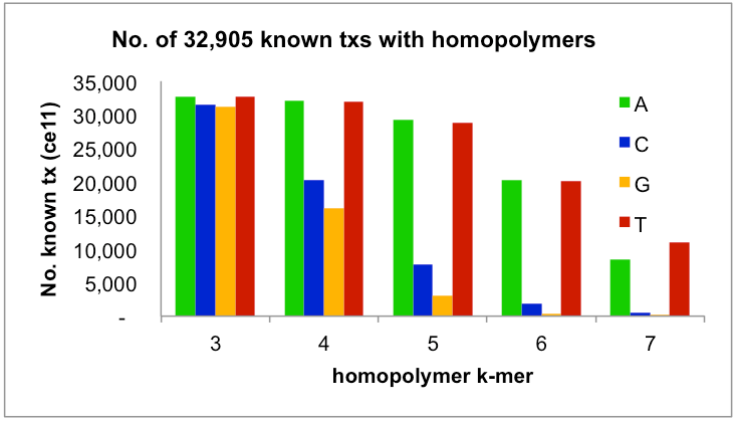


unc-52 exon 17 skipping increase by 3.14X, additional putative splice changes detected

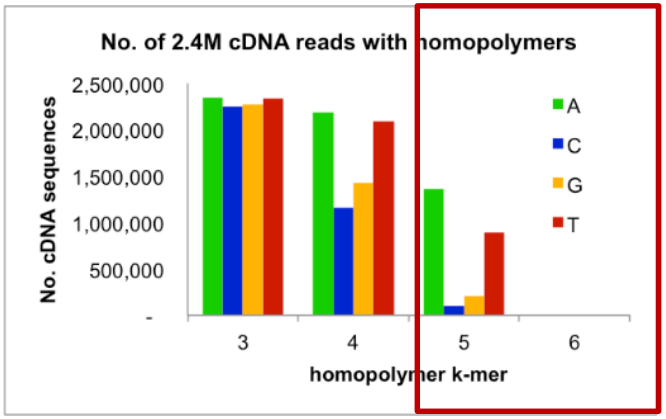
What to compare: Direct RNA vs cDNA sequencing

Library preparation
Data quality
Transcript detection
Abundance
Splice variant resolution
Homopolymer calling

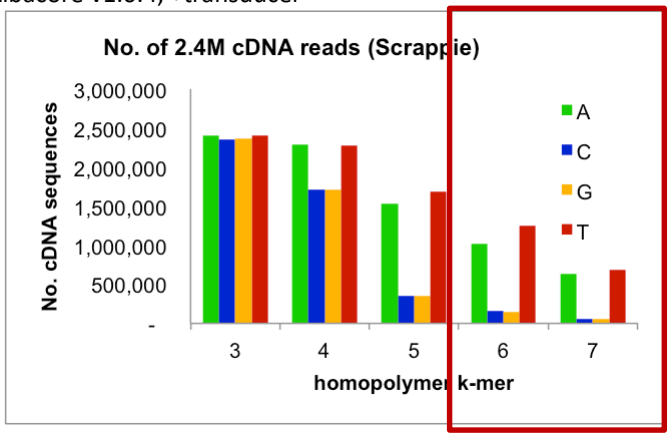
Updated basecallers improve homopolymer recovery



Albacore v0.8.4, -transducer



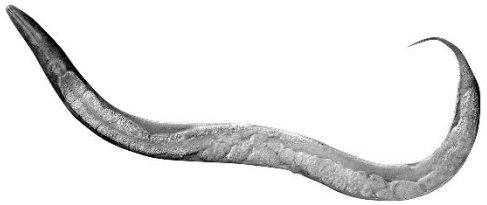
Albacore v1.0.4, +transducer



Conclusions

Library preparation

- Robust in both, simpler in RNA, mRNA lengths better preserved in cDNA
- Primary limitation in RNA is input and throughput



Quality, transcript detection and abundance

- Comparable when taking into account yield differences

Homopolymer calling improved with implemented transducer model

- Future analysis: pA tail detection, 3’ UTR and PAS profiling, catalog global splicing differences for *smu-1* mutant

Image courtesy of: Universiteit Utrecht, Developmental Biology

Acknowledgements

Timp Lab

Winston Timp

Taylor Lab

James Taylor

Mallory Freeberg

Kim Lab

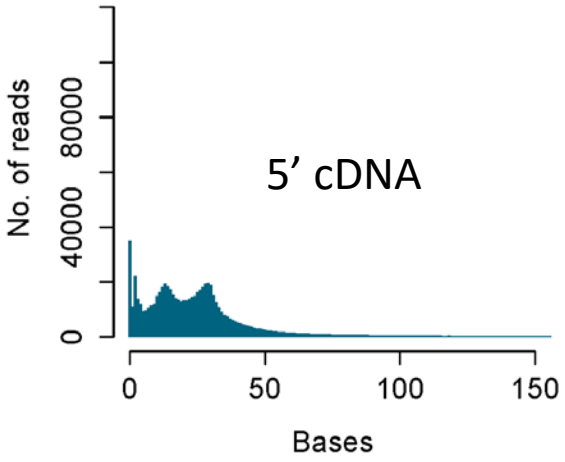
John Kim

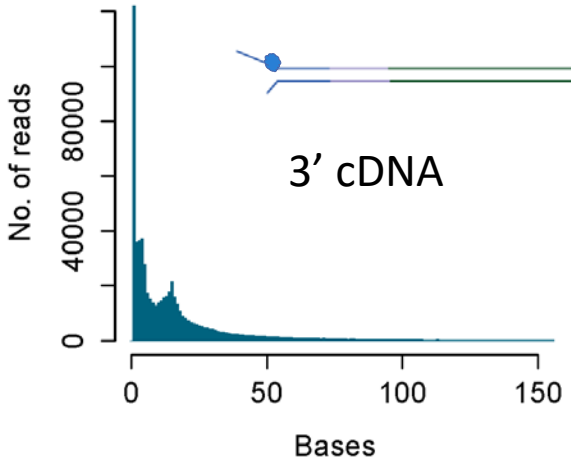
Amelia Alessi

Funding/Reagent Support

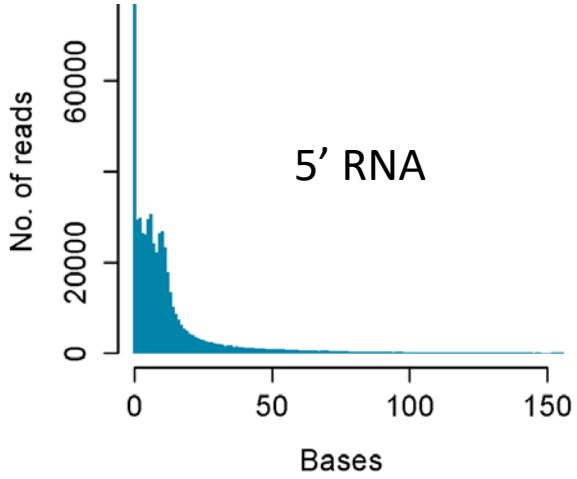
HFSP

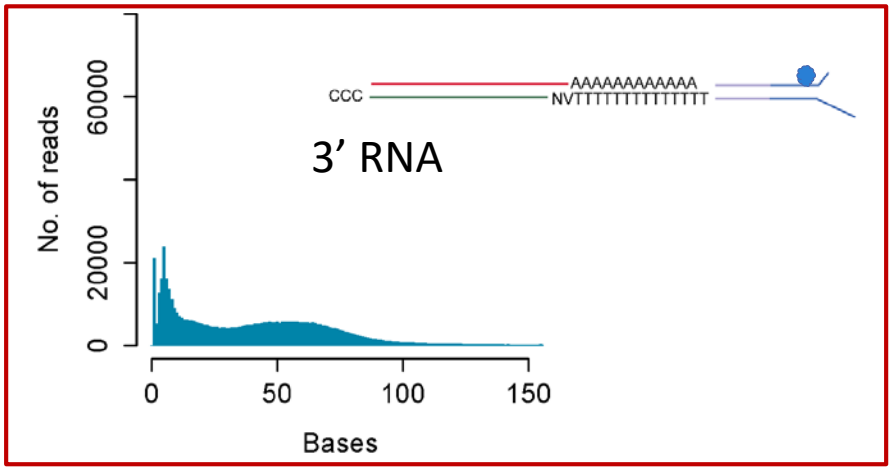
End trimming dependent on prep



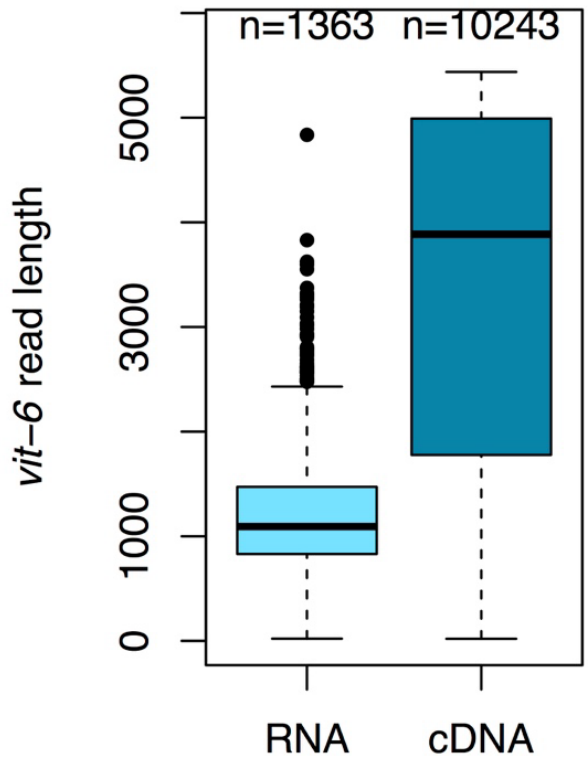


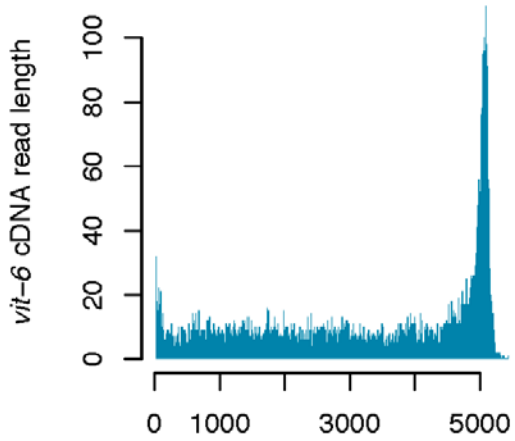
Primarily adapter clipping in cDNA ends, error in 5’ RNA, but 3’ RNA is both poly-A signal and DNA adapter.

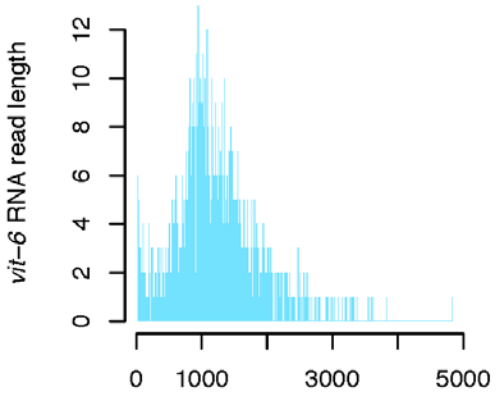




Most abundant transcript





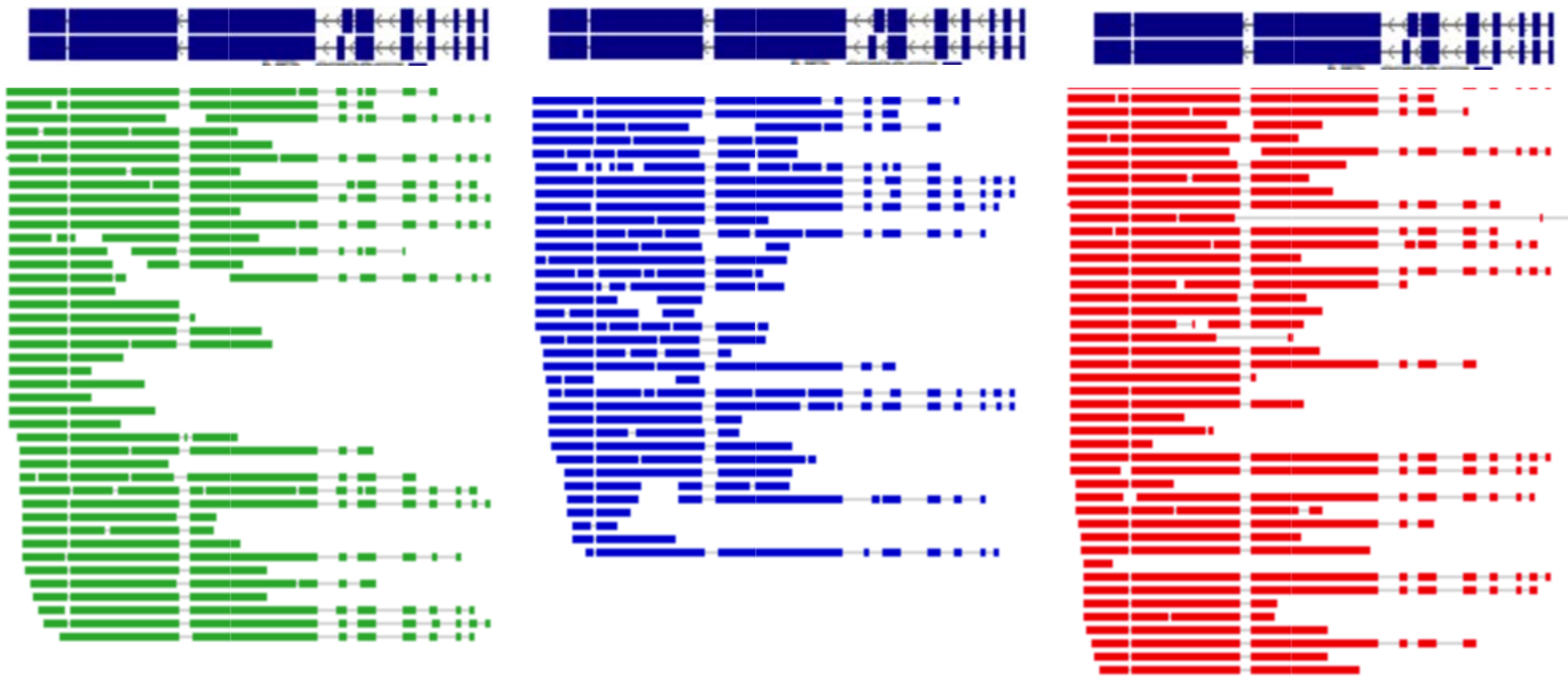


Poly-A tail potential

Eef-1A.1
Full 5' UTR 1290bp CDS Full 3' UTR 79bp poly-A tail

- **Homopolymer calling**
- Long poly-A tails aligned, likely requires further training/adapter trimming to refine

Splicing diversity captured by tested aligners



Other opts:
BLAT
STARlong
Graphmap
dds/gap2
…

GMAP LAST Exonerate

chrX:16377262-16393356, collagen alpha-2 (IV) chain

Splicing diversity captured by tested aligners

GMAP, 60% aligned

LAST

Exonerate, 80% aligned

Other opts:
BLAT
BWA-MEM
STAR
Graphmap
dds/gap2
…