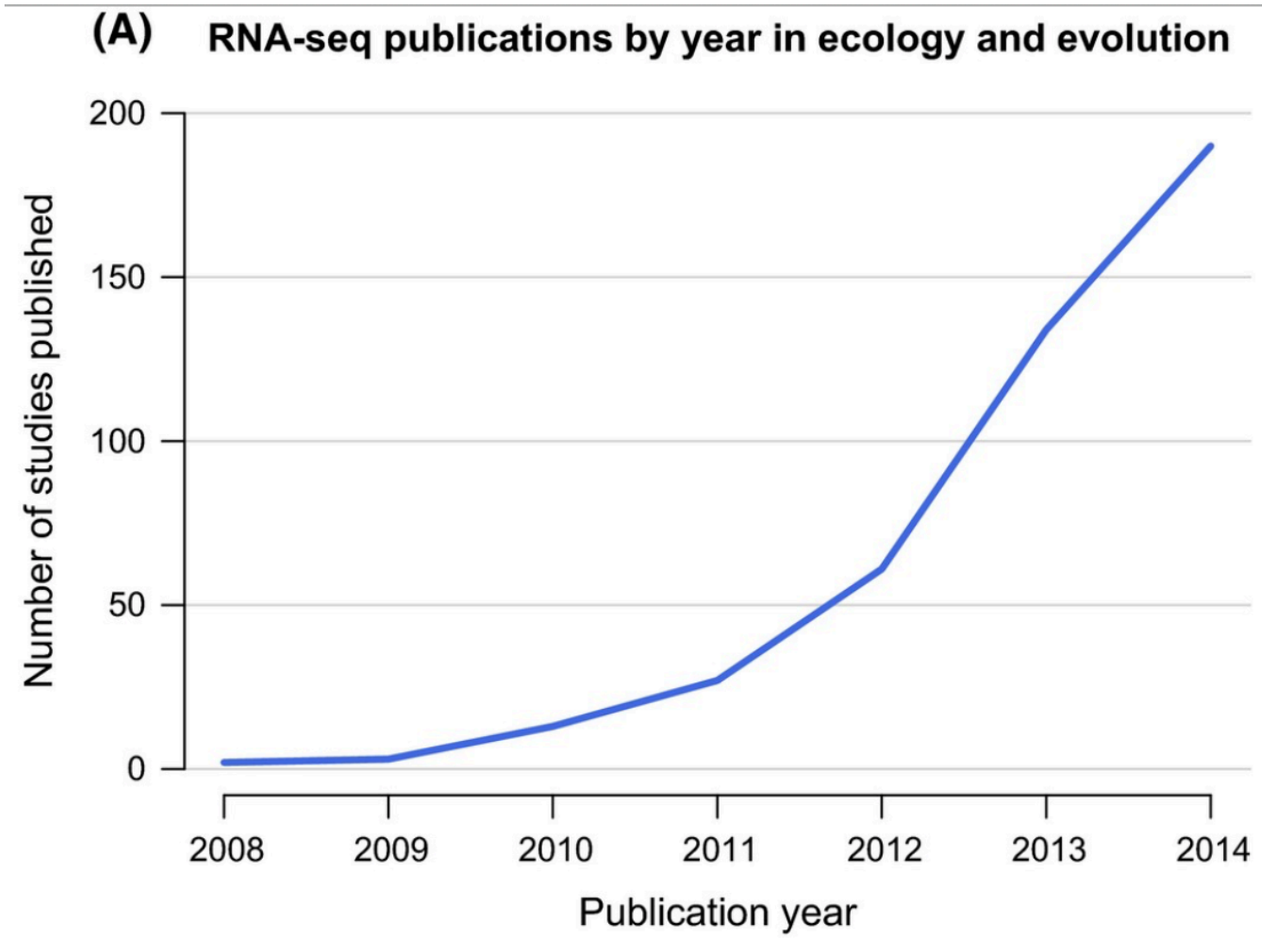
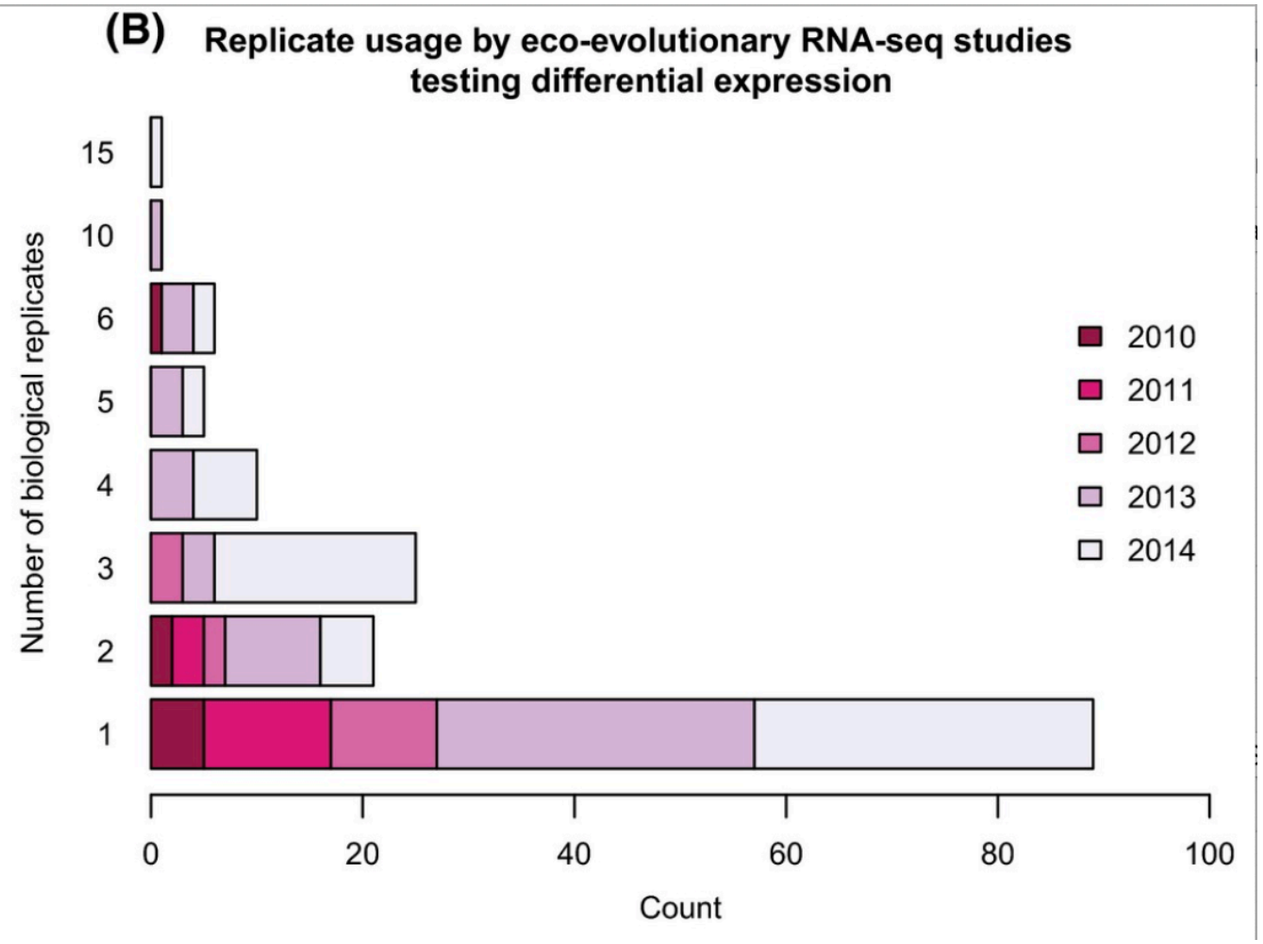
TAG-Seq

Dhivya Arasappan

(Thanks for Some Slides from Misha Matz,
and Emily Tallman)

Whole RNA-Seq typically means small number of biological replicates





Todd, E. et al, Molecular Ecology, 2016

3’ TAGSEQ- An Alternative to Whole RNA-Seq

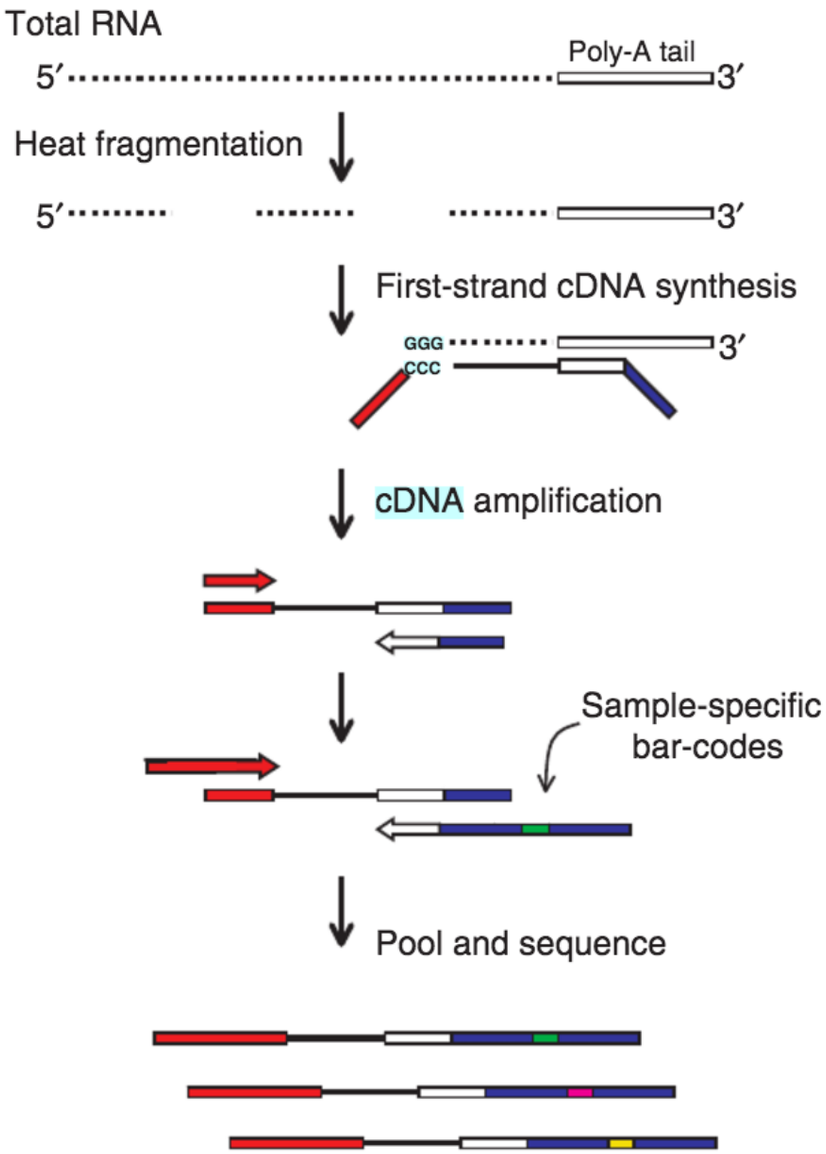


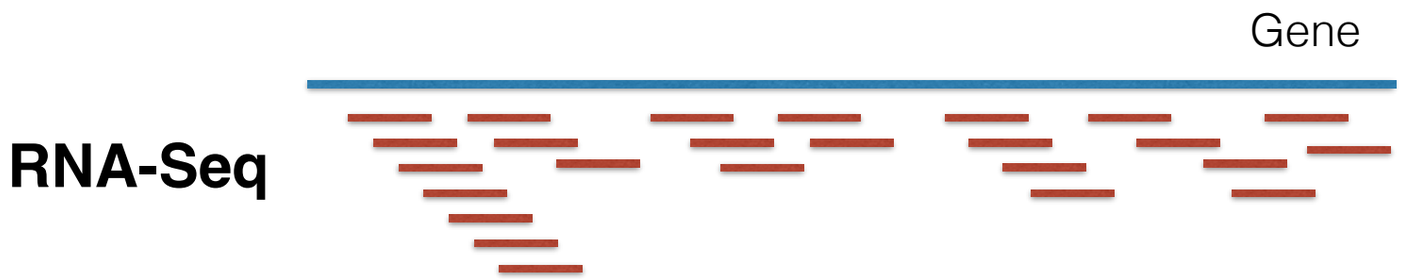
Fig. 1 Overview of the protocol used to prepare 3′ cDNA tag libraries from total RNA. RNA was fragmented at the beginning to eliminate biases resulting from differences in transcript lengths. First-strand cDNA was primed with a modified oligo-dT containing primer to target 3′ ends. Each sample was prepared with a sample-specific oligonucleotide barcode, then quantified and pooled prior to sequencing.

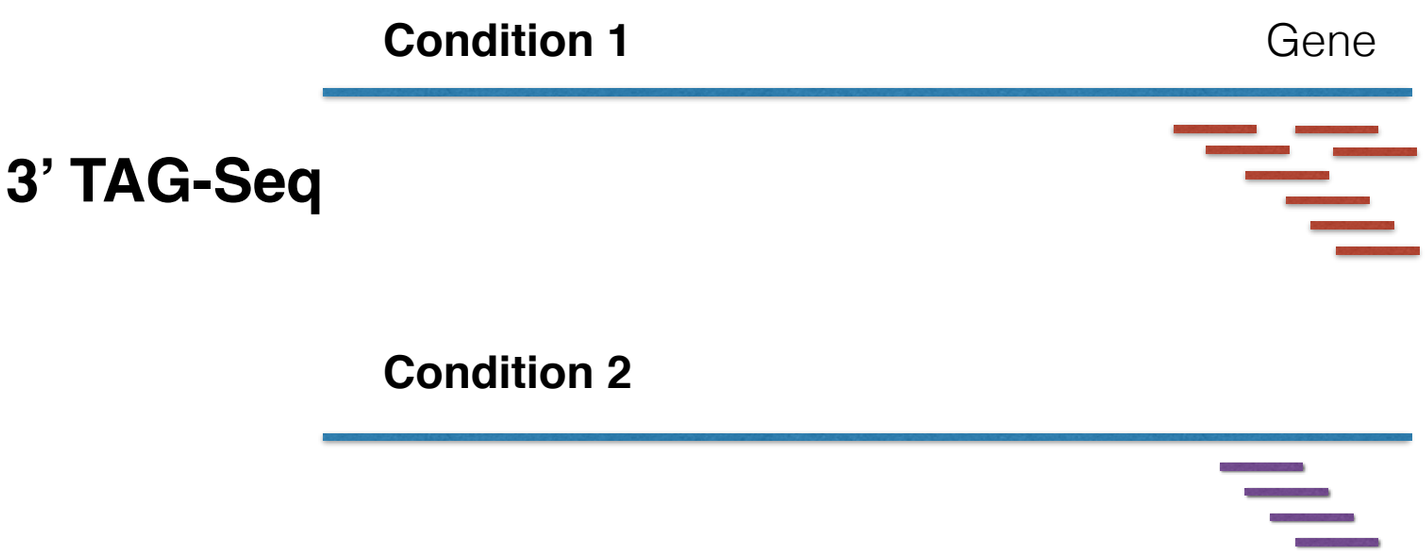
Targeting the 3’ prime end of RNA

Meyer, E. et al, Molecular Ecology, 2011

WHY TAGSEQ?

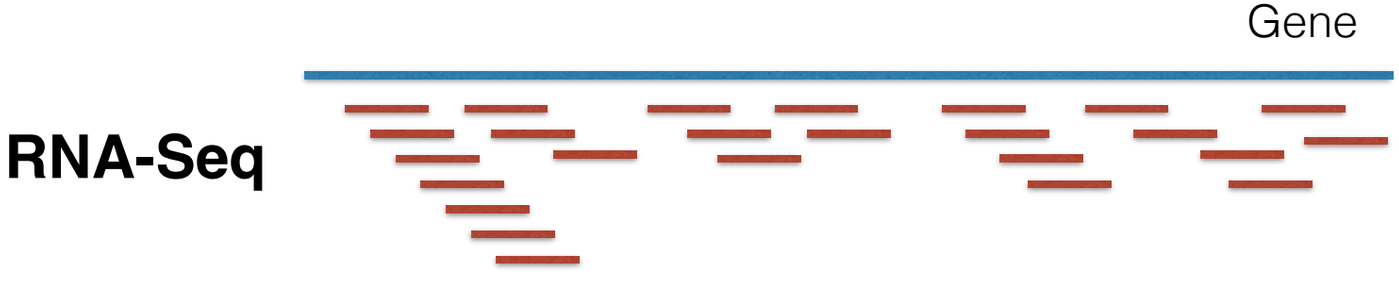
- Cheaper to sequence 3' end instead of the entire RNA.
  - Fewer reads need to be sequenced.
  - Lib prep is 10 times cheaper
- Amount of input RNA required is less.
- You can still identify differential expression.

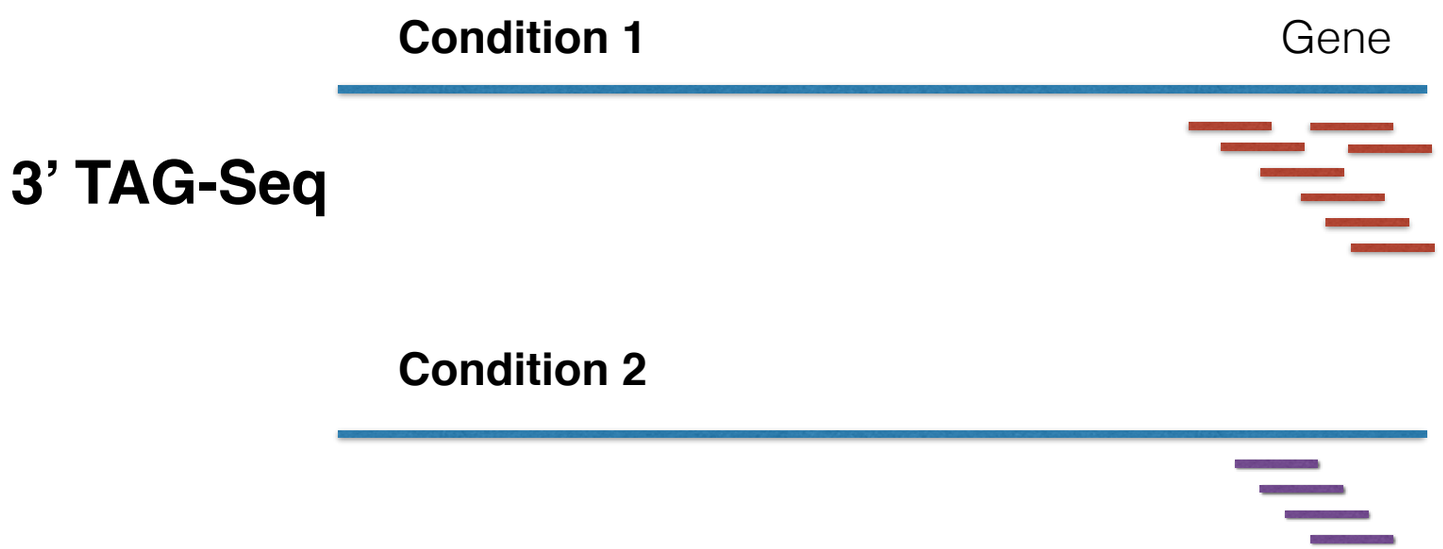




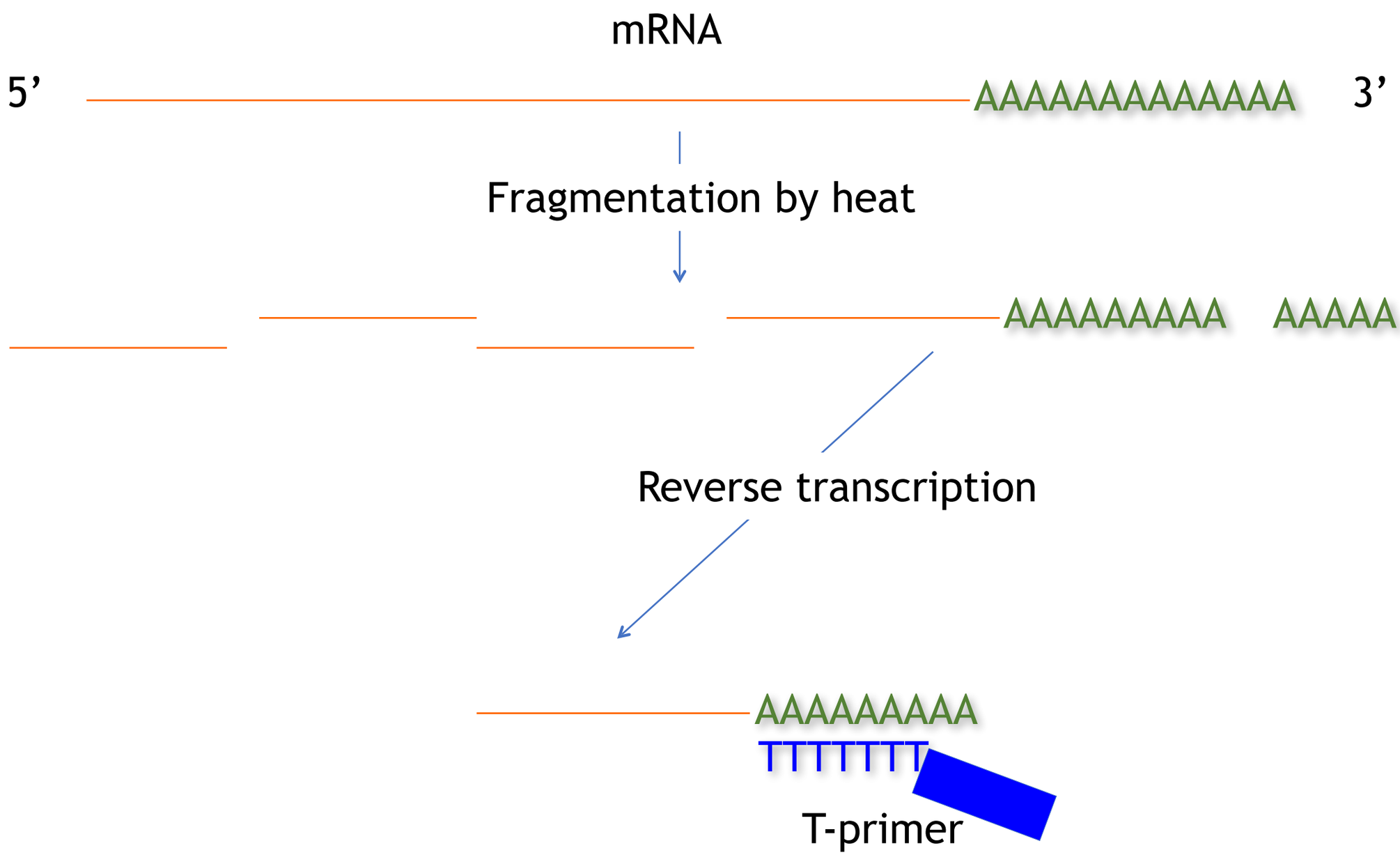
WHY NOT TAGSEQ?

- If you want to assemble a transcriptome
- If you want to look at differential splicing
- If you want to identify polymorphisms in gene sequences

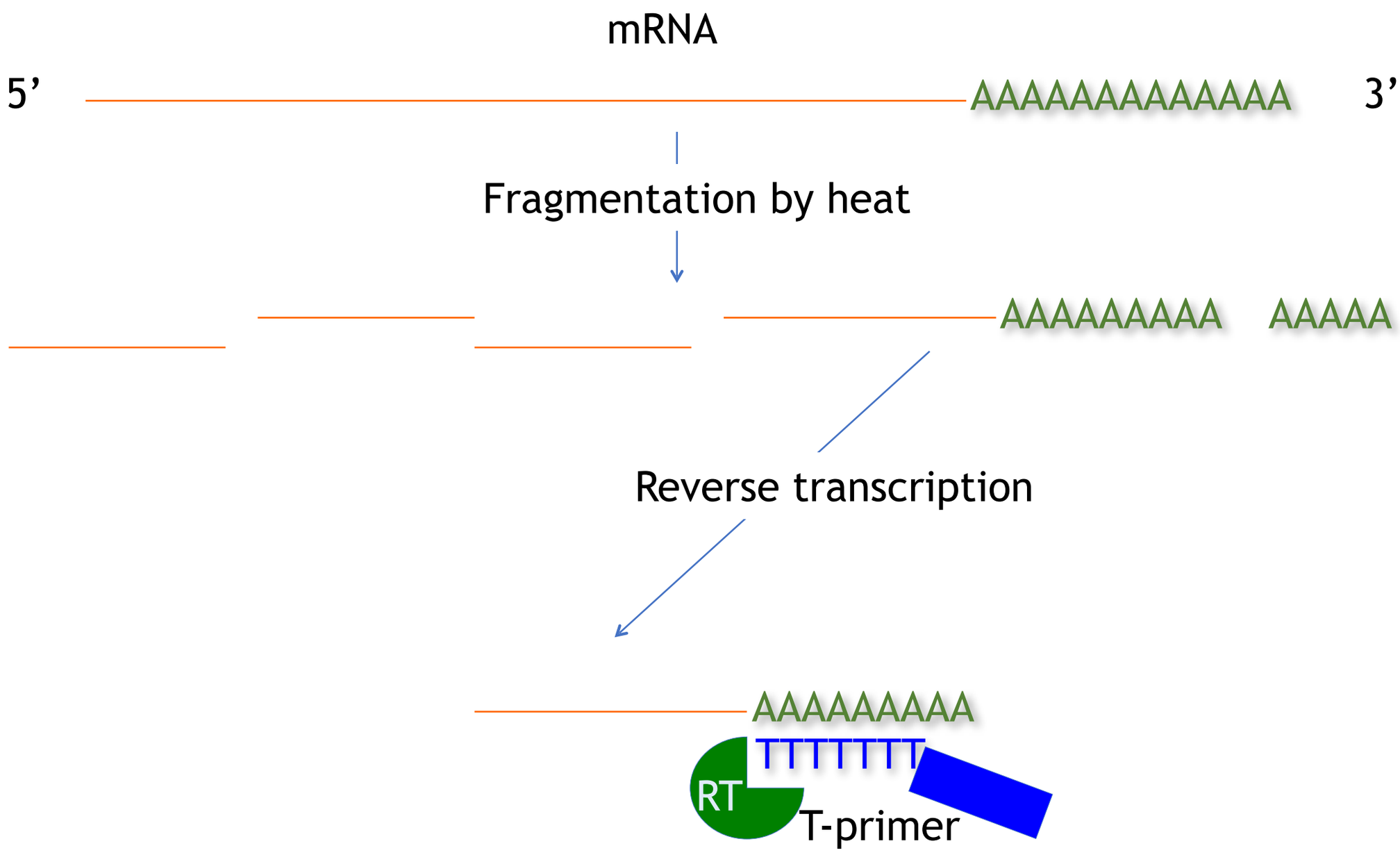




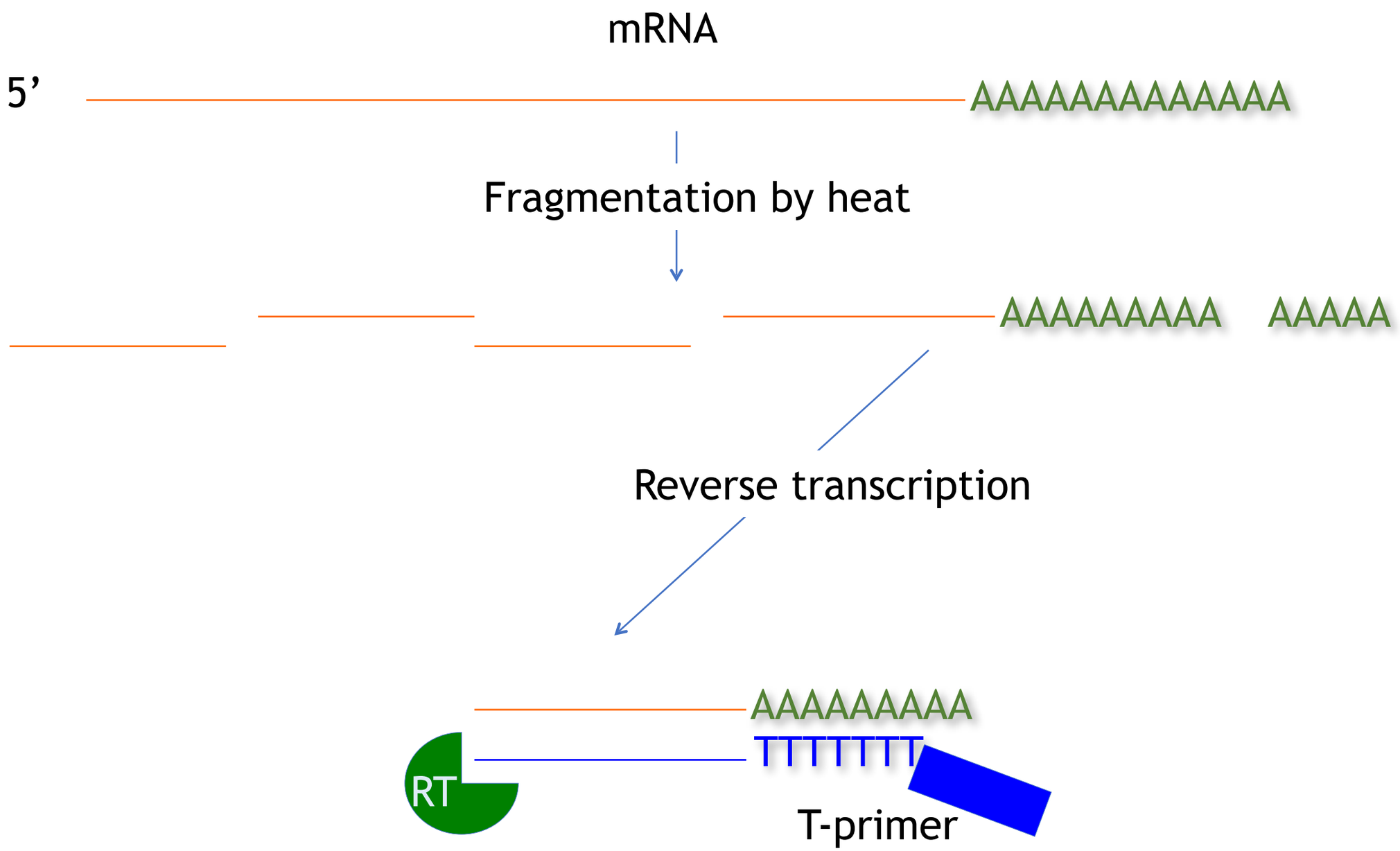
TAG-Seq Template Switching



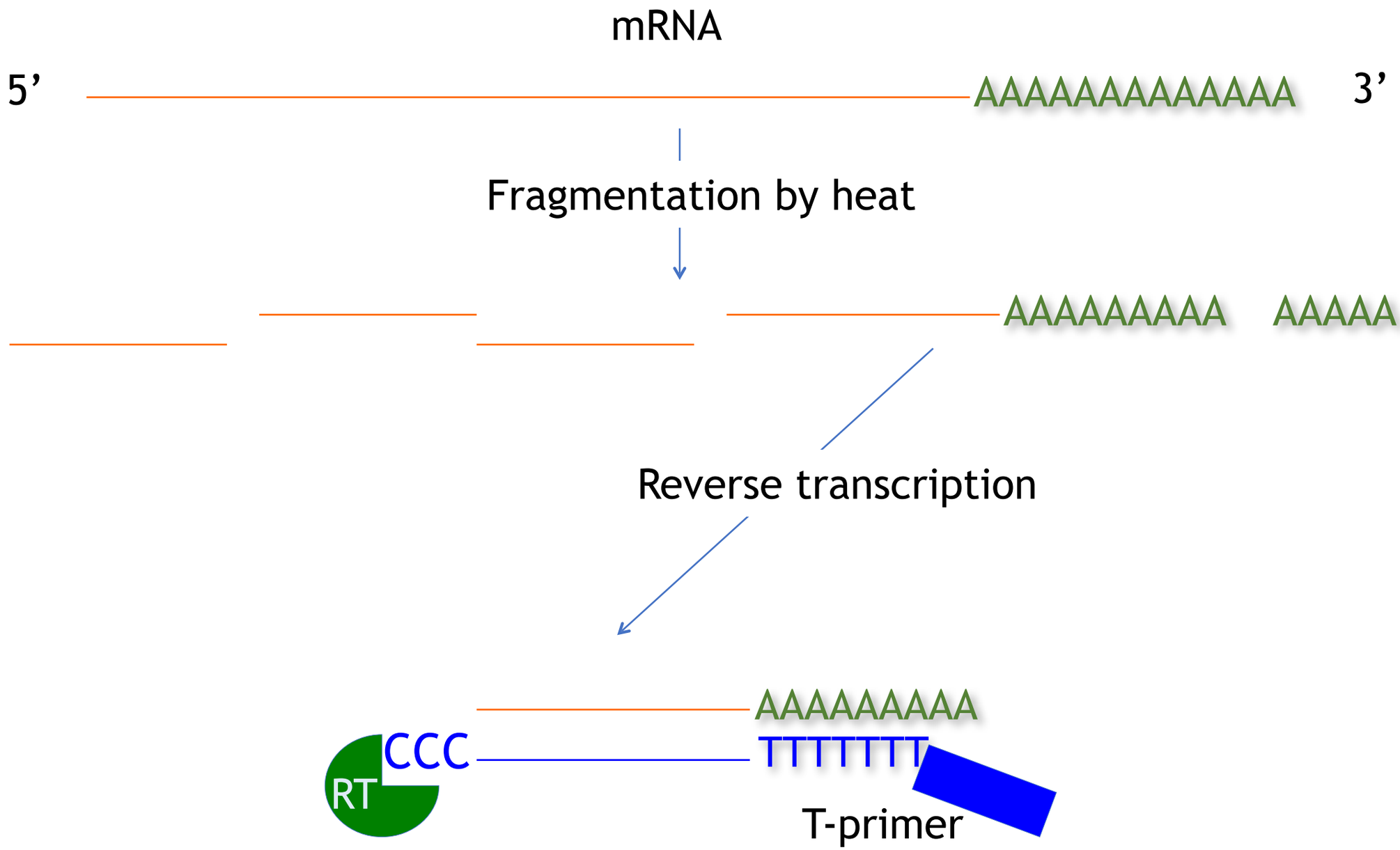
Tag-Seq Template Switching



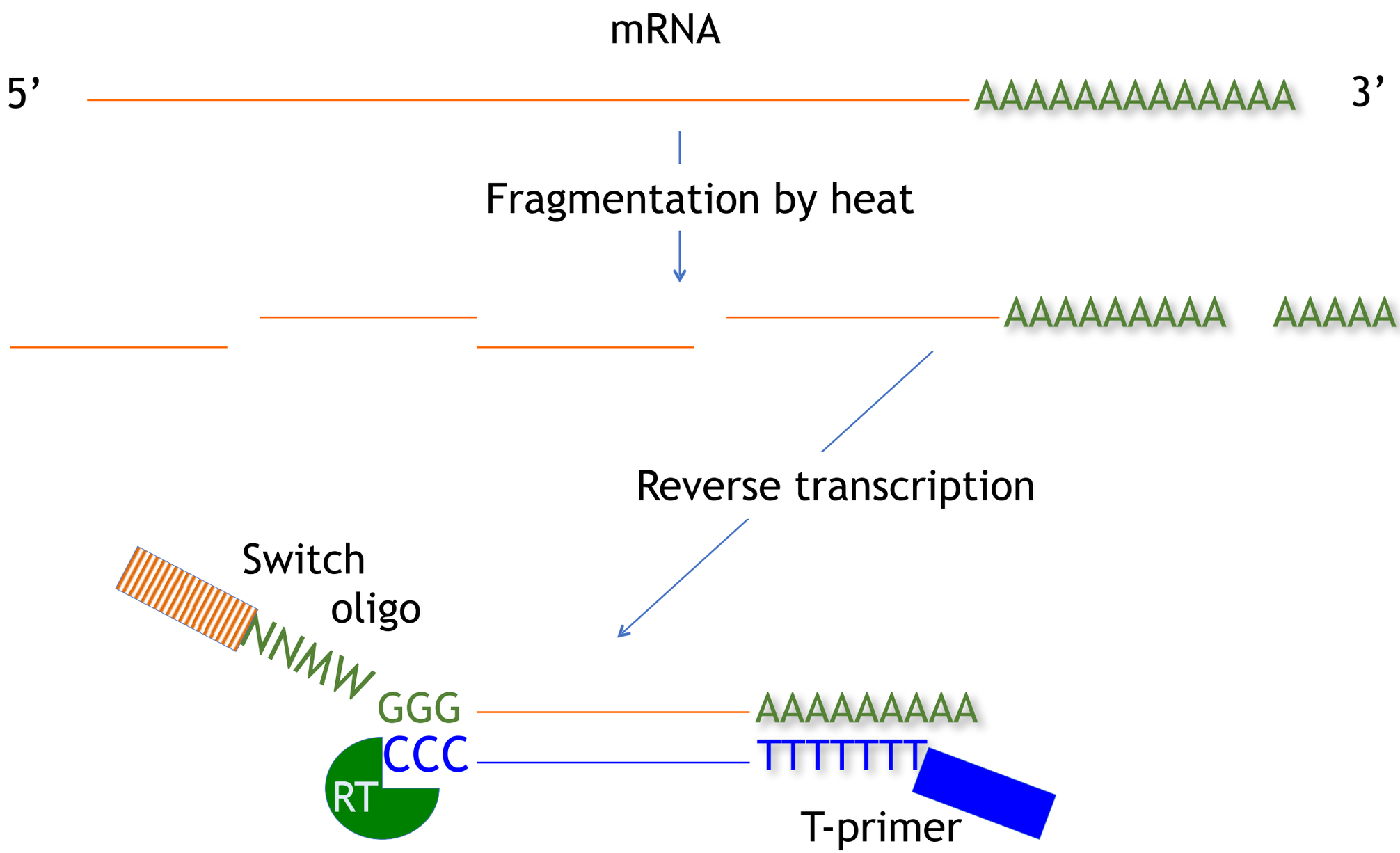
Tag-Seq Template Switching



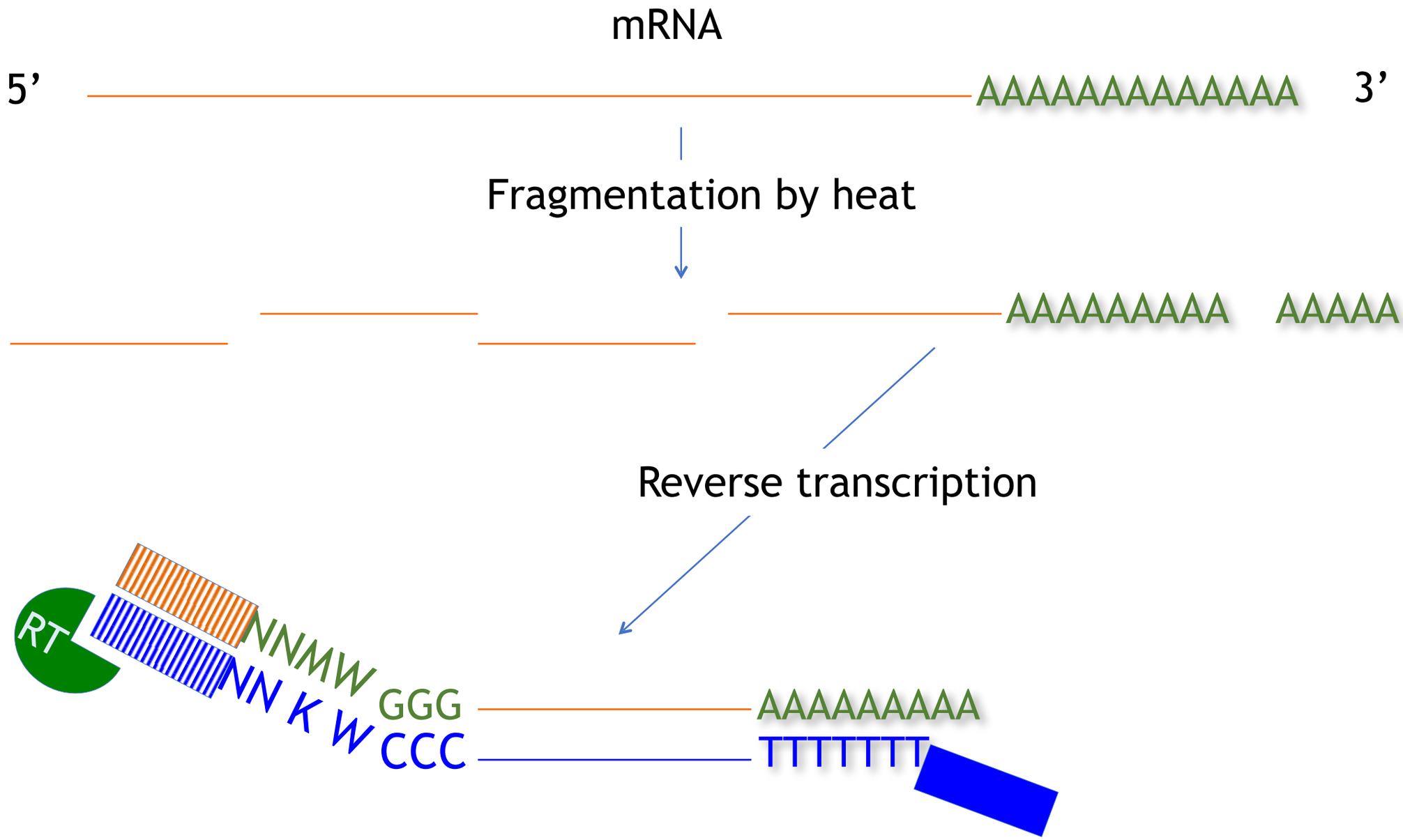
Tag-Seq Template Switching



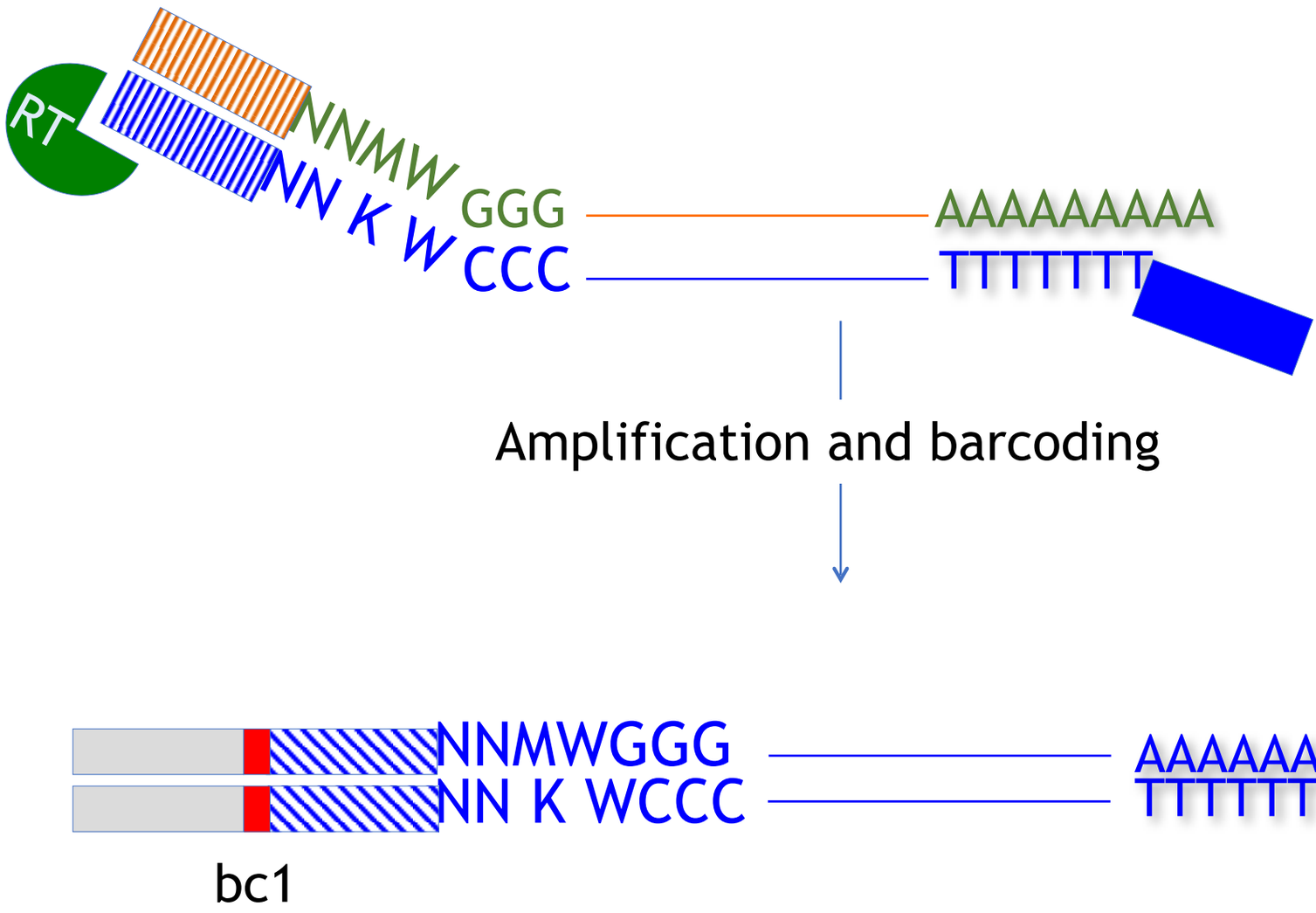
Tag-Seq Template Switching



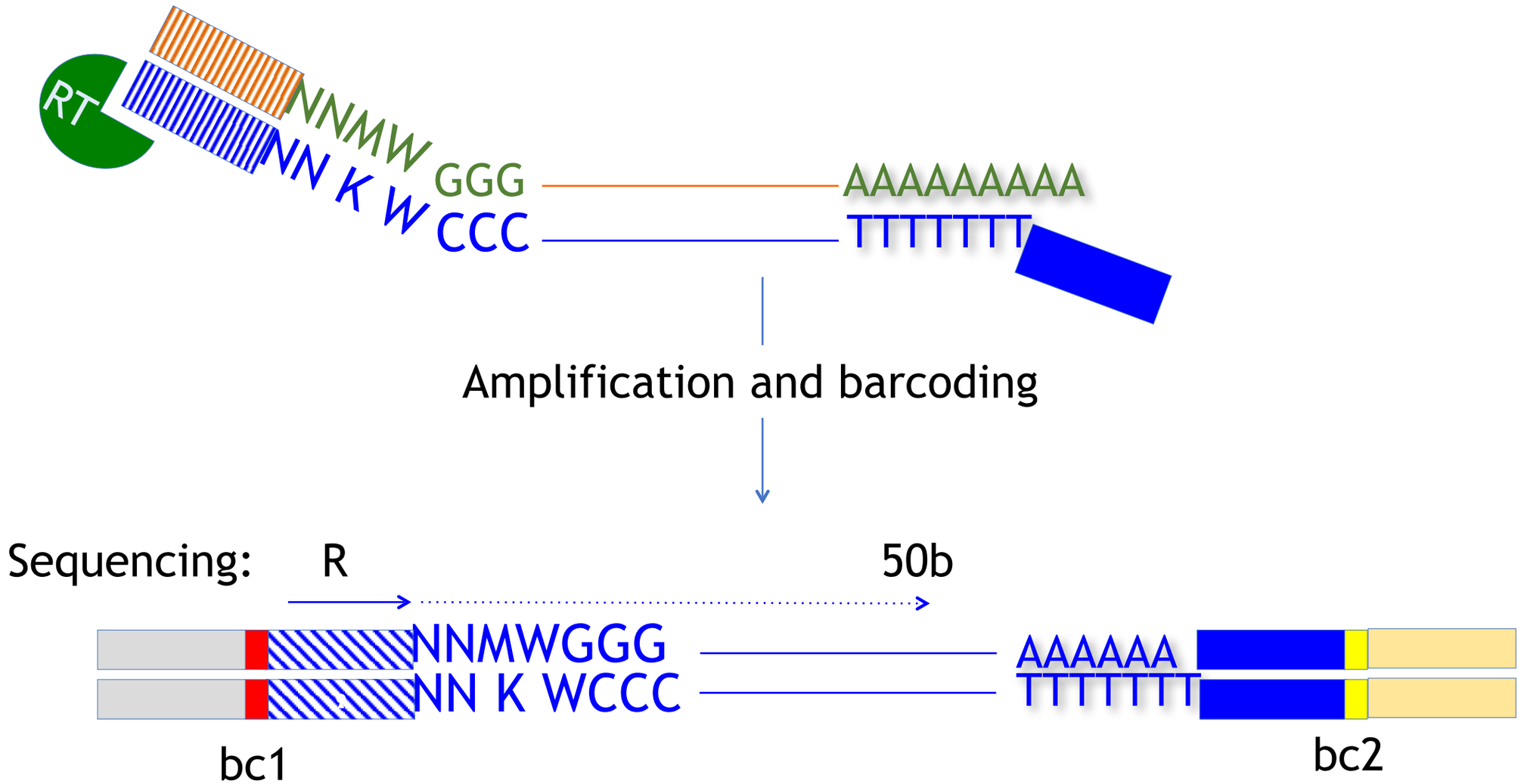
Tag-Seq Template Switching

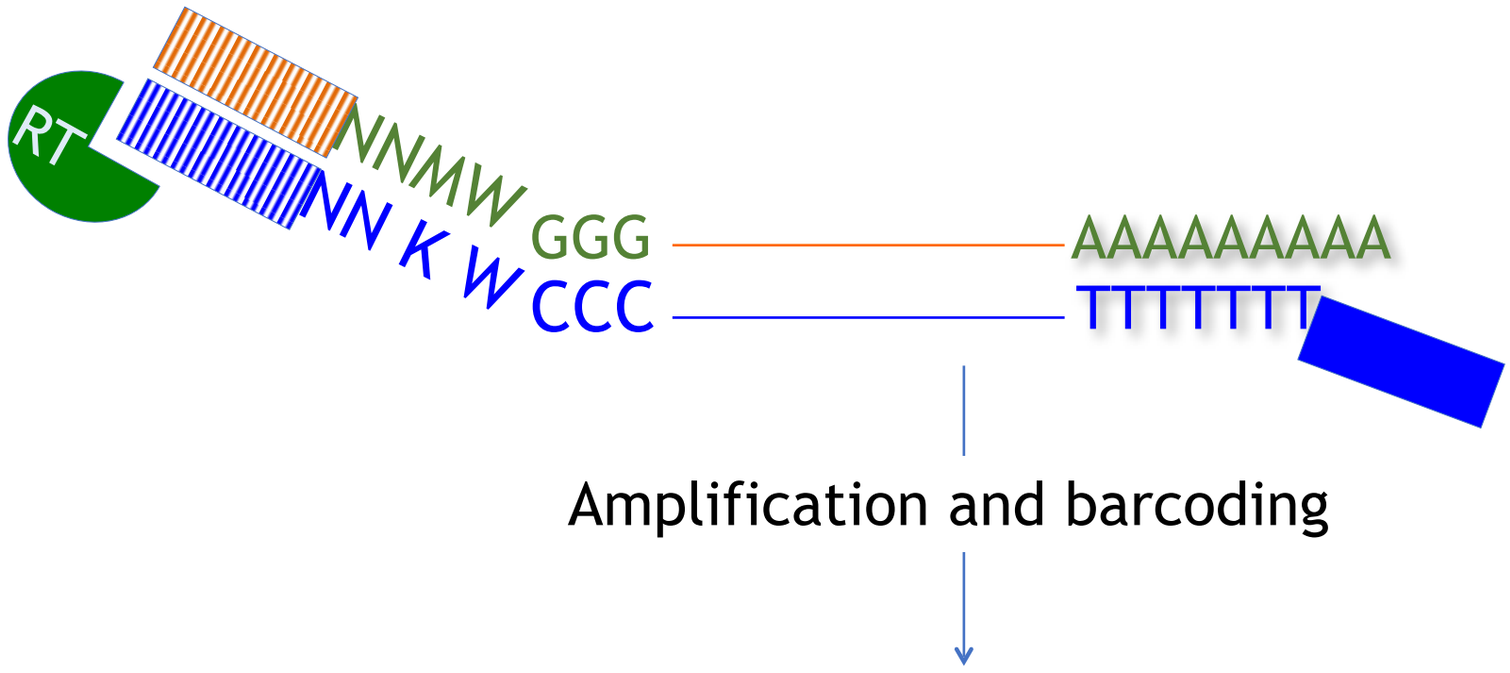


Tag-Seq Template Switching



Tag-Seq Template Switching



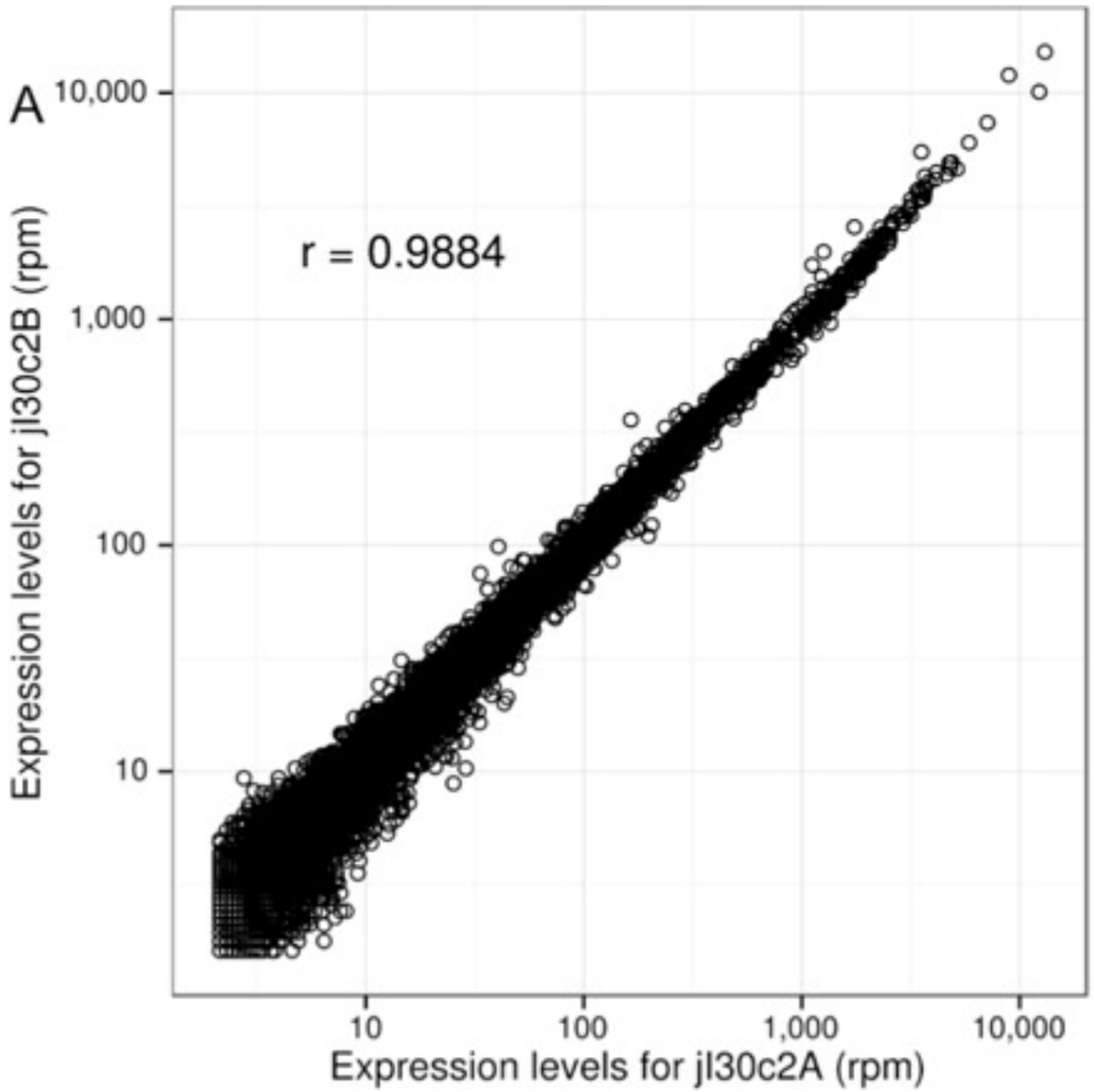


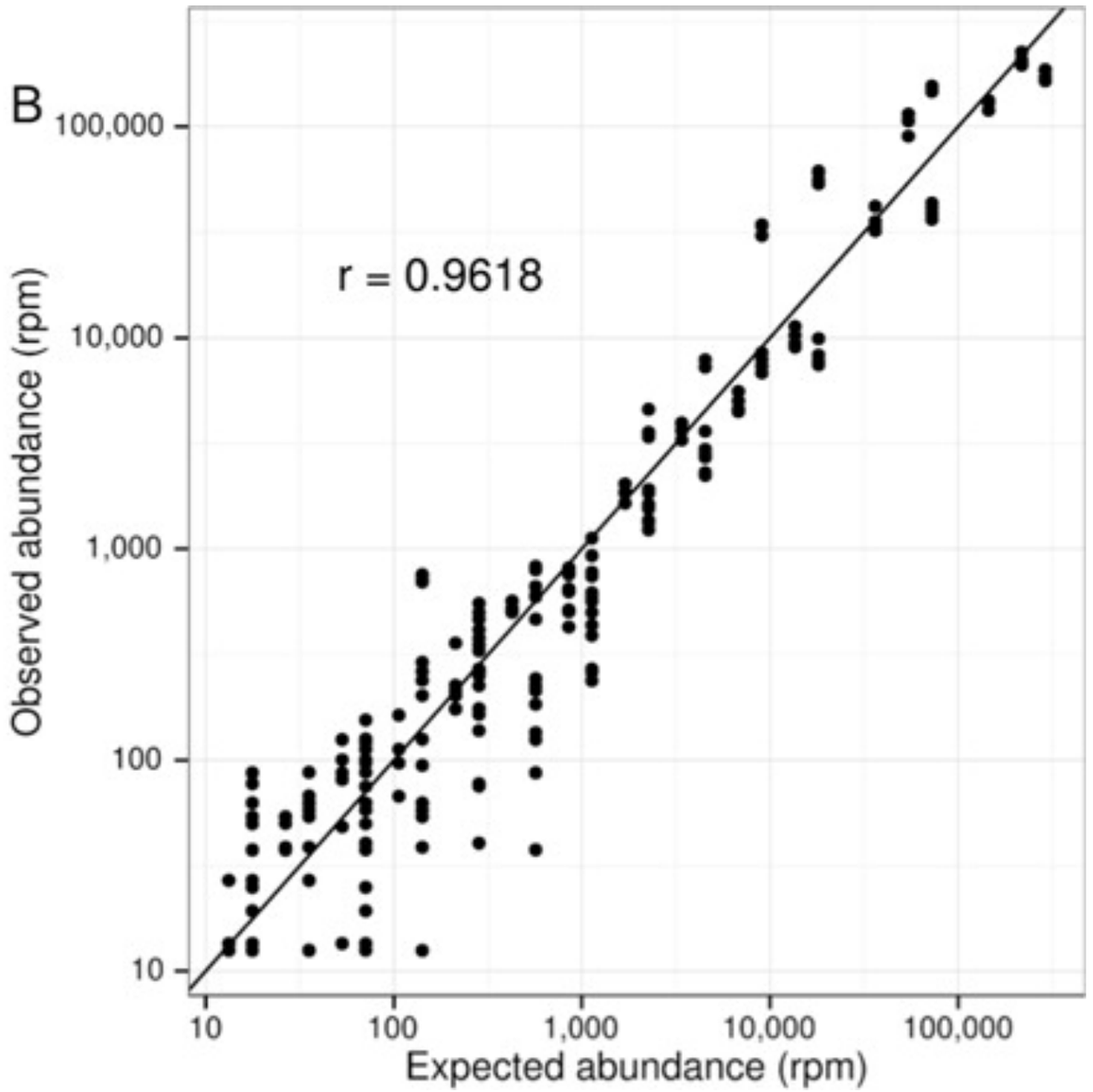
One tag per mRNA
Unique:

- NNMW (64-fold)+
- First 20 bases (Frag point)

TCATGGGGAAAAAGCAATTGATAATGACAGCCTGGTGTTTTATTGGAAA
TACAGGGGGGAAAGATGTGAATTTCATCAGAGTAGCACTTTATGTACTT
TCCTGGGGAGACATAGAAAACAACACCTTTCATGCAAAAACAATGGGAA
ACCTGGGGGAGATATGCAGATGAGCATTTGCTAACAGAATTACAAGATC
AGATGGGGACACCAGTGTCCCCAGTGGATTGAACGACATCGATACGAAG
ATAAGGGGTATTTTAGCTTTGTGCTGGGACACTCAGTGACAGATATAAT
TACTGGGGTAGTTGCTTGGCTCTTGTTGCCGCTCCATGTCCATAGGATC
CCCAGGGGGAACAATATGGCAAAAAAAAAAAAAAAAAAAAAAAAAAAA

Is Tag-Seq Accurate?





Rozenberg, A. et al, BioTechniques, 2018

Is Tag-Seq comparable to whole RNA-Seq?

Tag-Seq recovers known concentrations of mRNA (ERCC controls) with more accuracy than whole mRNASeq

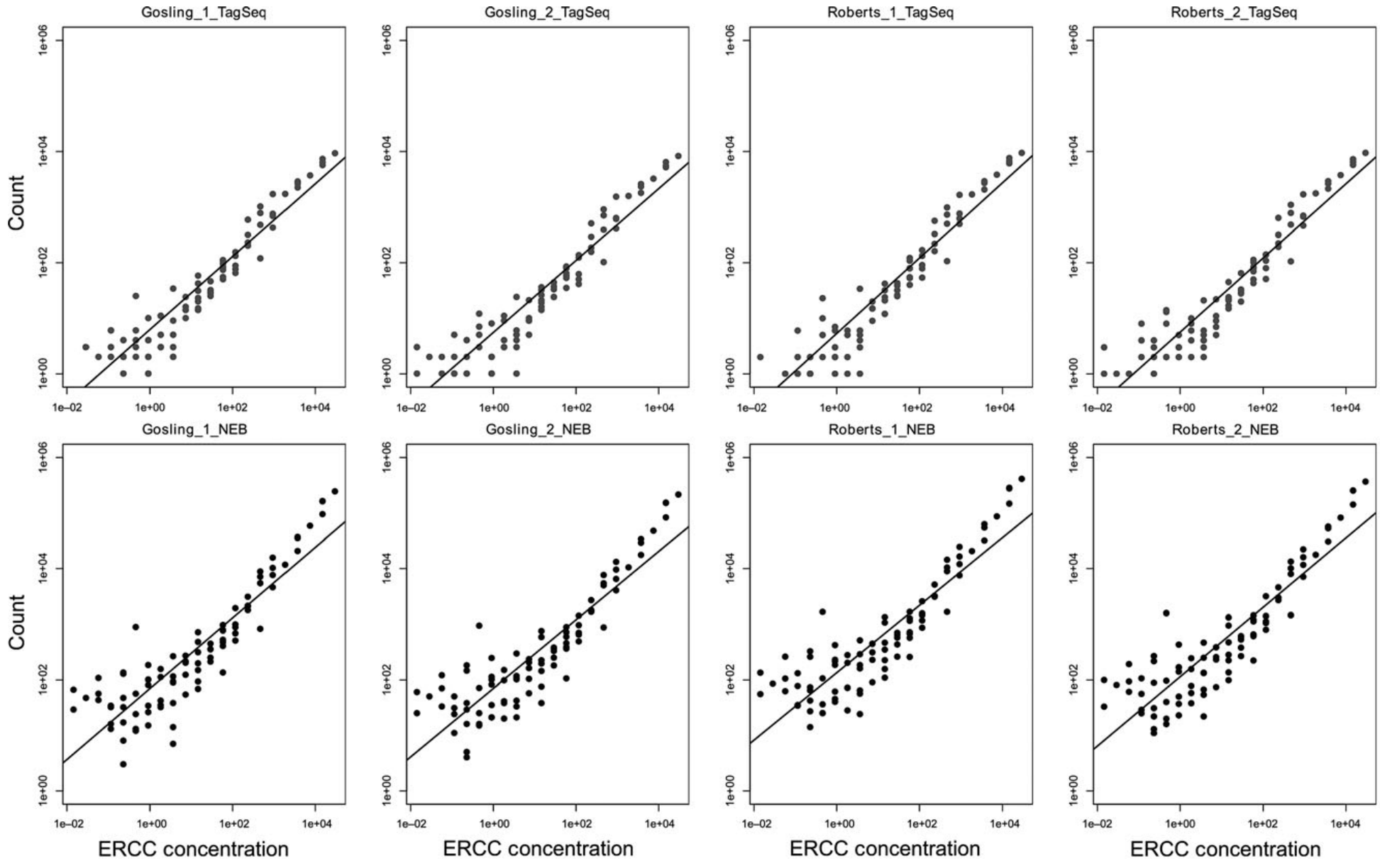


Fig. 1 Regression of observed vs. expected ERCC transcripts shows TagSeq has higher adjusted R^2 values for four different biological samples prepared with both methods (paired *t*-test, $t = 18.63$, d.f. = 3, $P < 0.001$).

Lohman et al, Molecular Ecology Resources, 2016

Is Tag-Seq comparable to whole RNA-Seq?

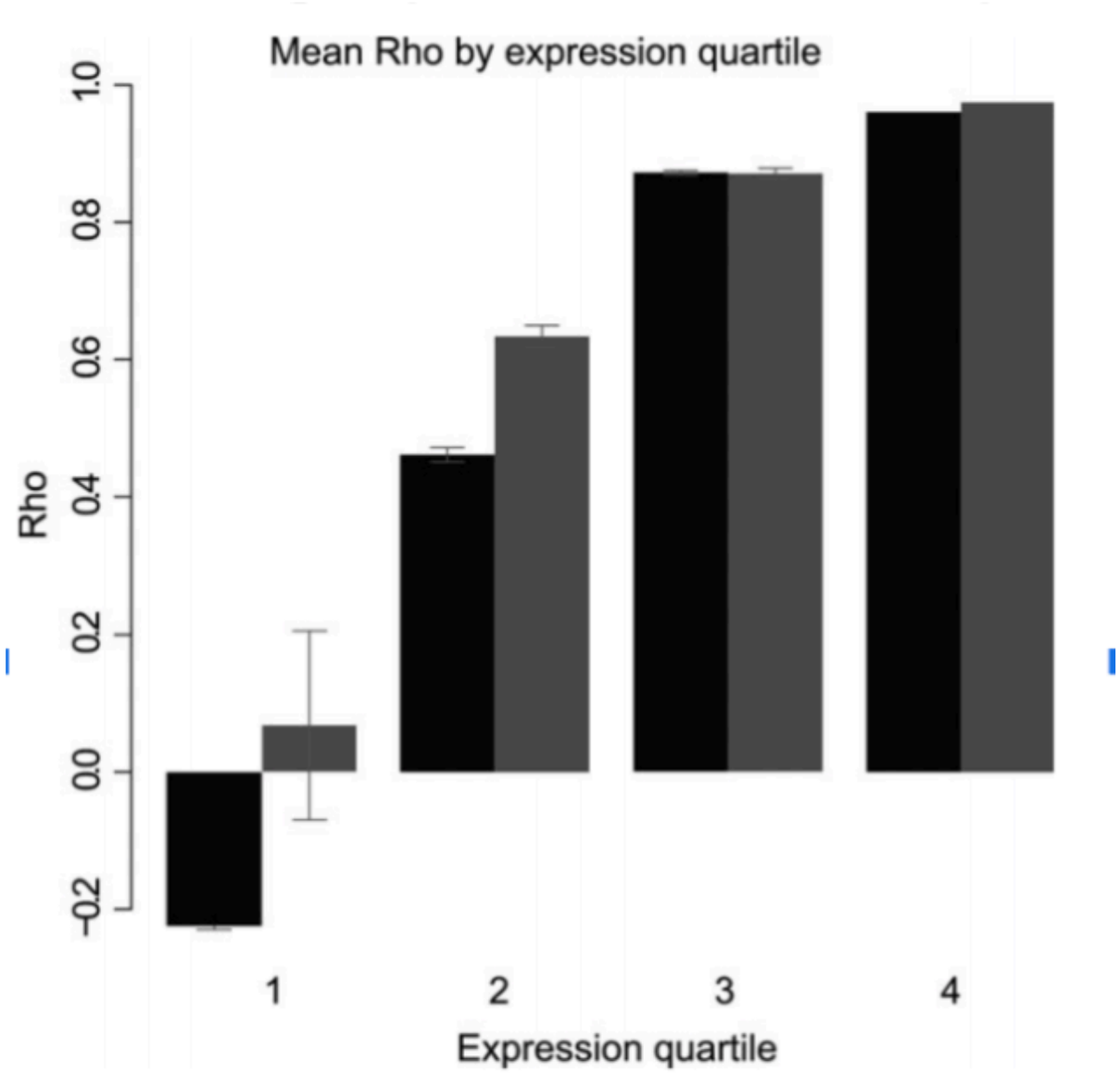


Fig. 3 Breakdown of control mRNAs by abundance class shows that TagSeq recovers mRNAs better than TruSeq, especially at lower abundances. Light grey bars are TagSeq, dark grey bars are whole mRNAseq. Fences indicate standard error.

Lohman et al, Molecular Ecology Resources, 2016

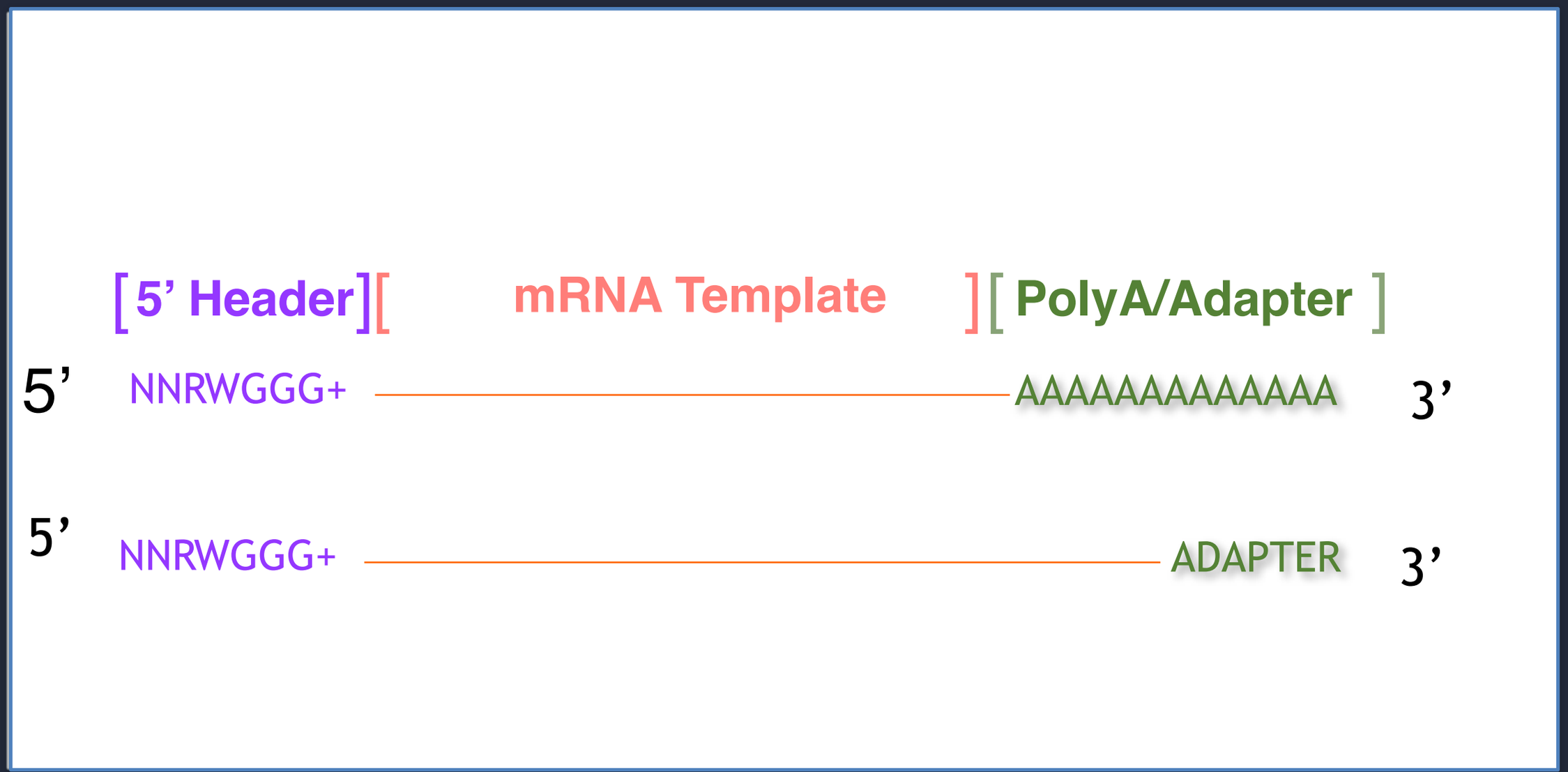
How do we analyze RNA-Seq data?

- **STEP 1**: PREPROCESS RAW DATA
- **STEP 2**: EVALUATE QUALITY
- **STEP 3**: MAP TO REFERENCE, ASSESS RESULTS
- **STEP 4**: QUANTIFY TRANSCRIPTS
- **STEP 5**: TEST FOR DIFFERENTIAL EXPRESSION
- **STEP 6**: VISUALIZE AND PERFORM OTHER DOWNSTREAM ANALYSIS

STEP 1 - Preprocess Raw Data

https://github.com/z0on/tag-based_RNAseq

Tag-Seq Read Structure



STEP 1- Preprocess Tag-Seq Data

Trim Header

Degenerate Base: More than one base possibility

N = A, T, G, or C

M = A or C

W = A or T

Before

[NNMWGGGN]AAAAAATTTGGGCCT

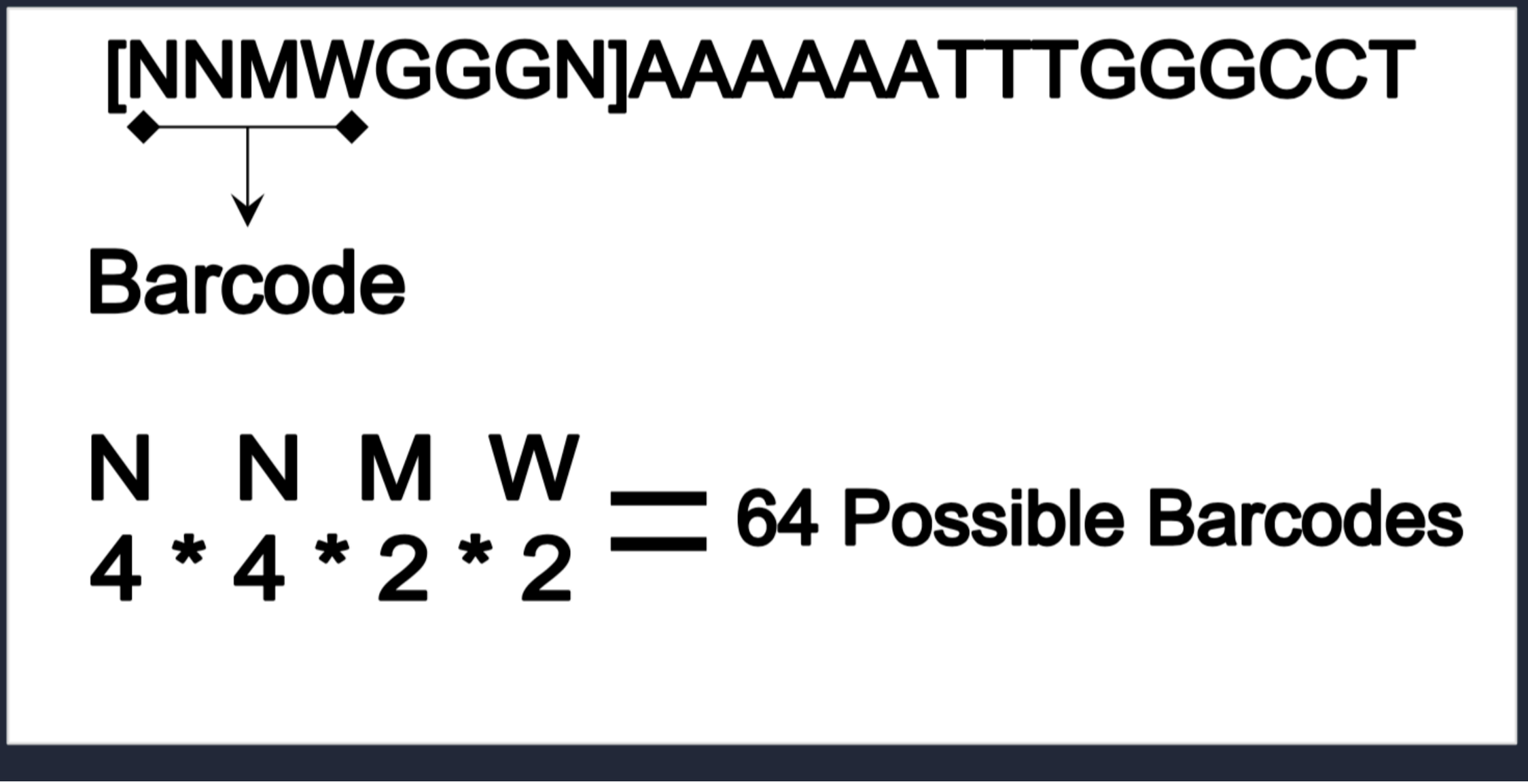
After

AAAAAATTTGGGCCT

STEP 1- Preprocess Tag-Seq Data

Remove Duplicates

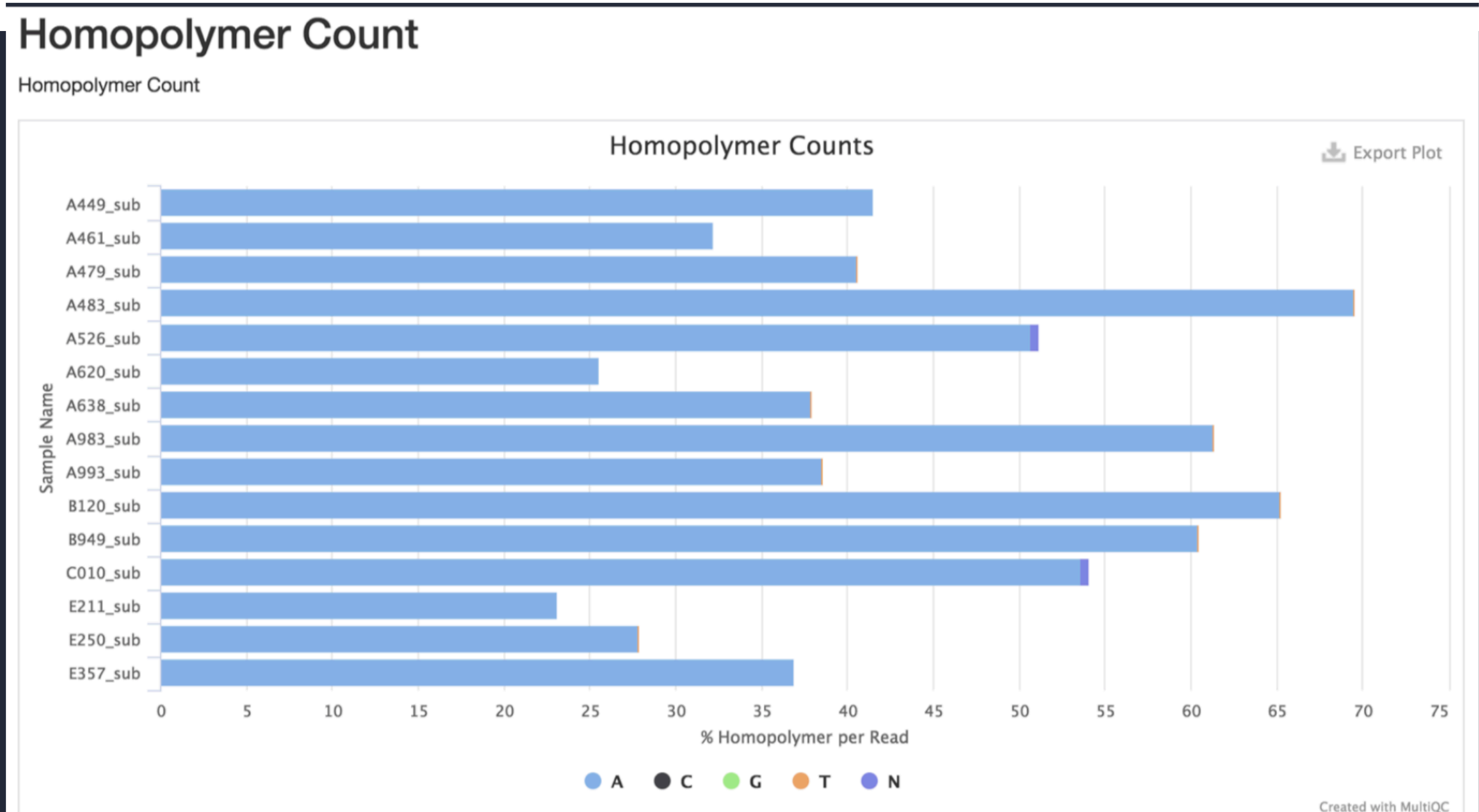
Reads that share the same degenerate header and the first 20 bases of the sequence.



STEP 1- Preprocess Tag-Seq Data

Filter polyA and low quality reads

Reads that share the same degenerate header and the first 20 bases of the sequence.



STEP 1- Preprocess Tag-Seq Data

Filter polyA and low quality reads

Reads that share the same degenerate header and the first 20 bases of the sequence.

PolyAs:

fastx_clipper -a AAAAAAAA -l 20 -Q33

Low Quality Reads:

fastq_quality_filter -Q33 -q 20 -p 90

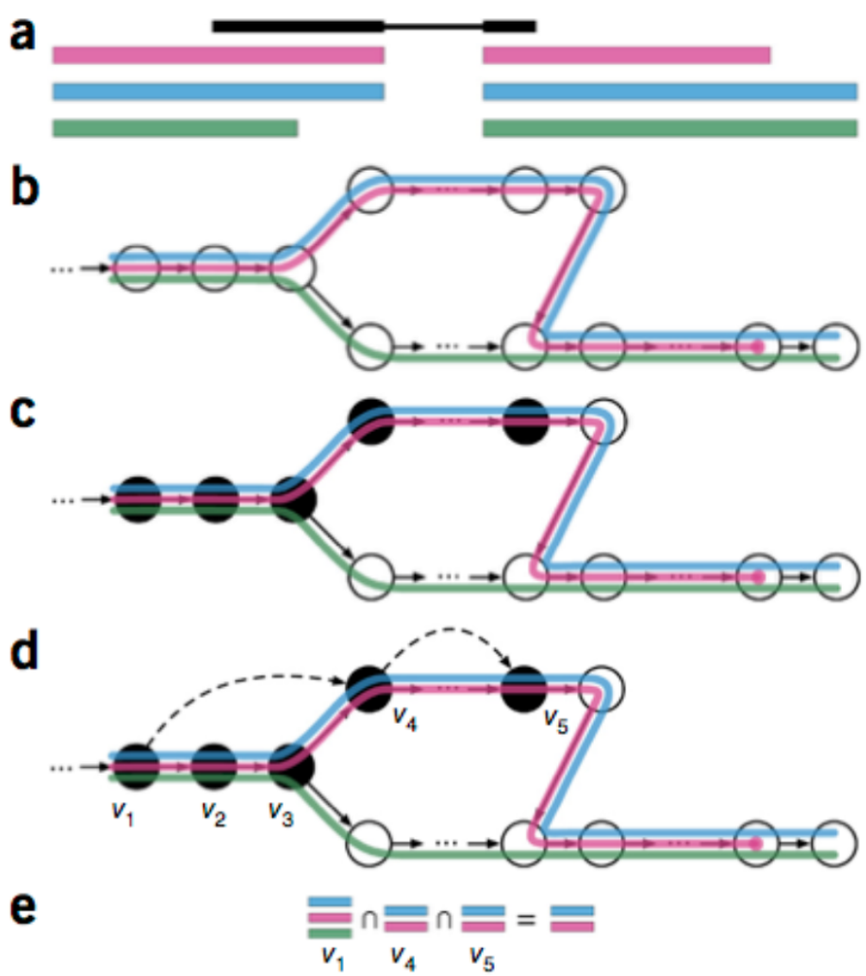
How do we analyze RNA-Seq data?

- **STEP 1**: PREPROCESS RAW DATA
- **STEP 2**: EVALUATE QUALITY
- **STEP 3**: MAP TO REFERENCE, ASSESS RESULTS
- **STEP 4**: QUANTIFY TRANSCRIPTS
- **STEP 5**: TEST FOR DIFFERENTIAL EXPRESSION
- **STEP 6**: VISUALIZE AND PERFORM OTHER DOWNSTREAM ANALYSIS

Pseudoaligners don’t seem to perform as well with Tag-Seq Data

Sample	Kallisto Mapping %	BWA Mapping %
A1-WLD1_S28	61.2%	77.75%
A2-WLD2_S36	60.4%	77.68%
A3-WLD3_S44	59.5%	77.10%
A4-WLD4_S52	68.4%	79.17%

Kallisto



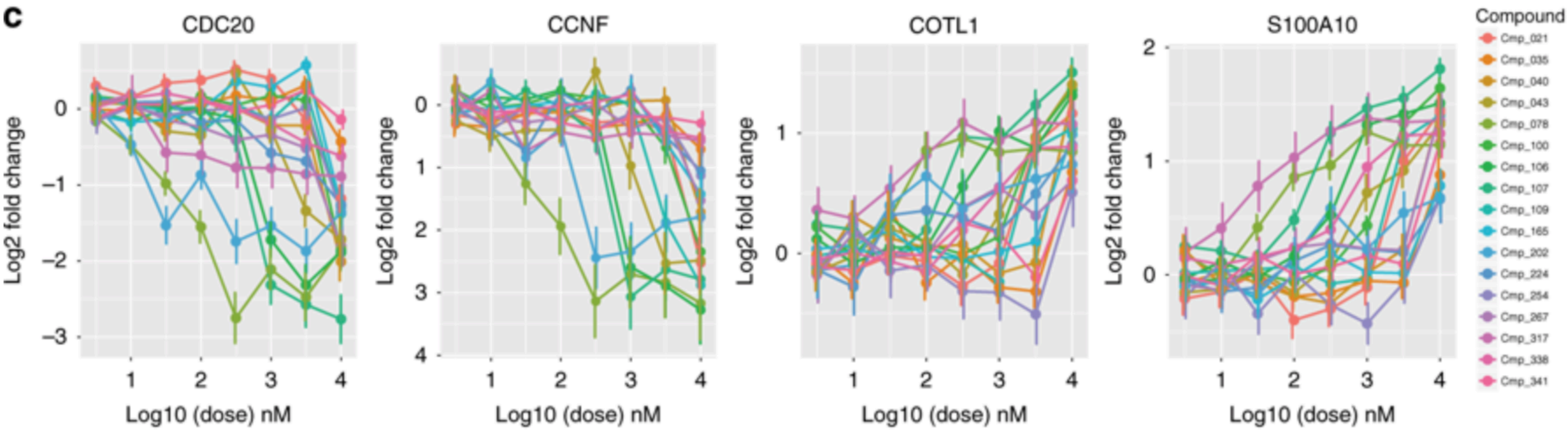
Bray, N. et al, Nature Biotechnology, 2016

Mapping Summary

- For Tag-Seq, unspliced and spliced mappers would both work fine.
- Most often, you will map to the transcriptome instead of genome.
- For Tag-Seq, Avoid pseudoaligners like Kallisto.

Where is Tag-Seq Being Used?

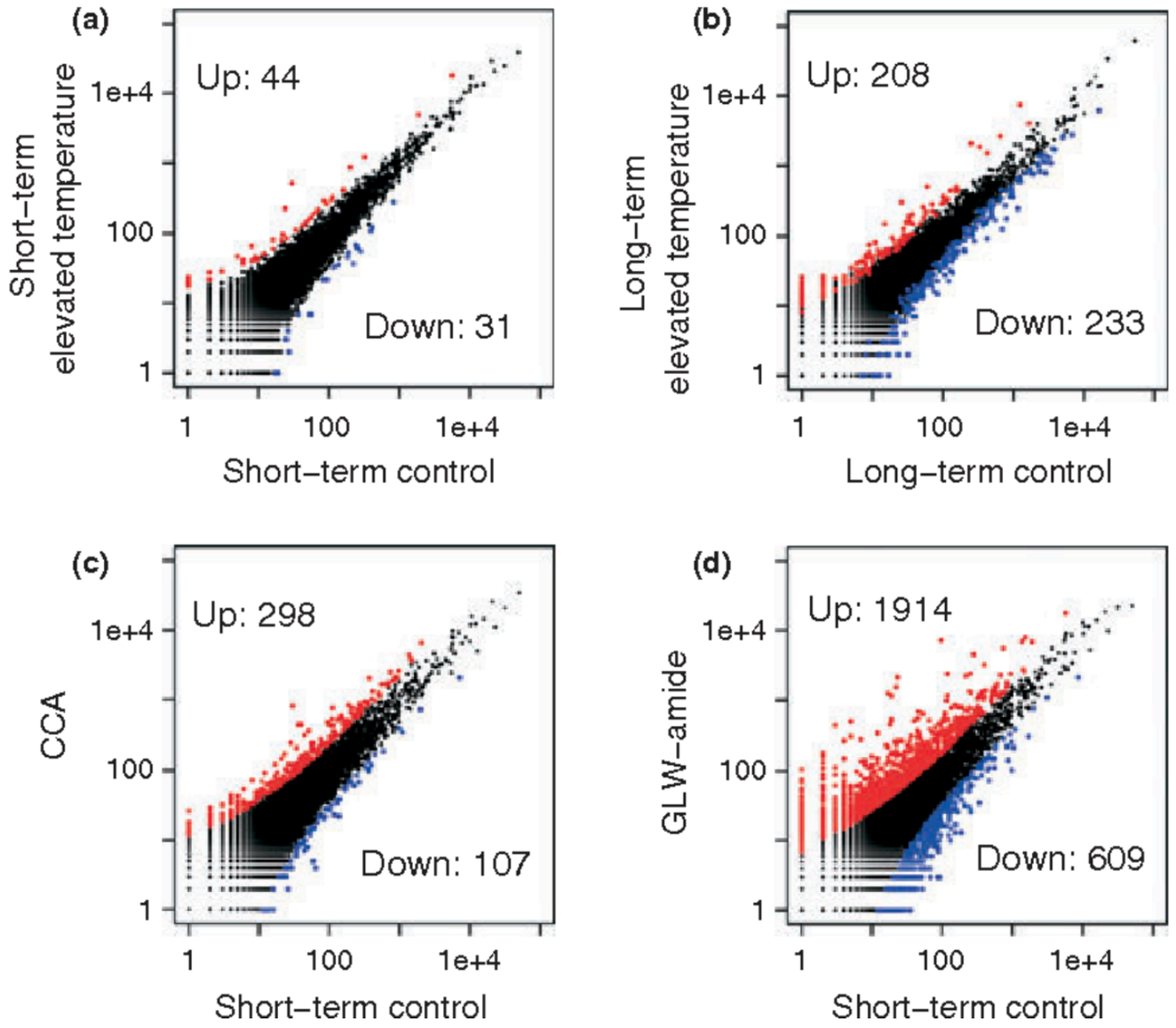
Identifying Gene Expression Signatures for Drug Compounds



Ye, C. et al, Nature Communications, 2018

Where is Tag-Seq Being Used?

Studying Gene Expression Responses of Coral Larvae to Temperature Changes



Meyer, E. et al, Molecular Ecology, 2011

Where is Tag-Seq Being Used?

- Whole RNA-Seq for assembling a transcriptome.
- Tag-Seq for large number of samples for gene expression profiling.