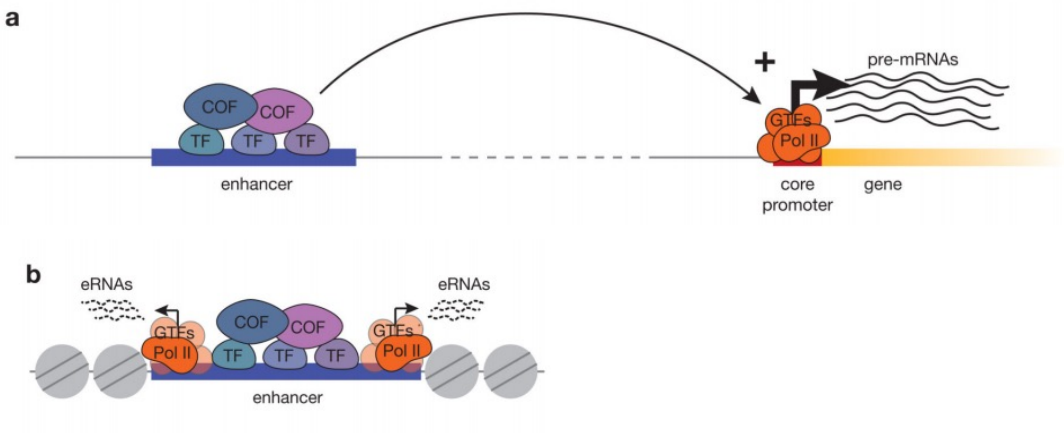
PRO-seq and other methods to study nascent transcription

HDYDI, 9th June 2020

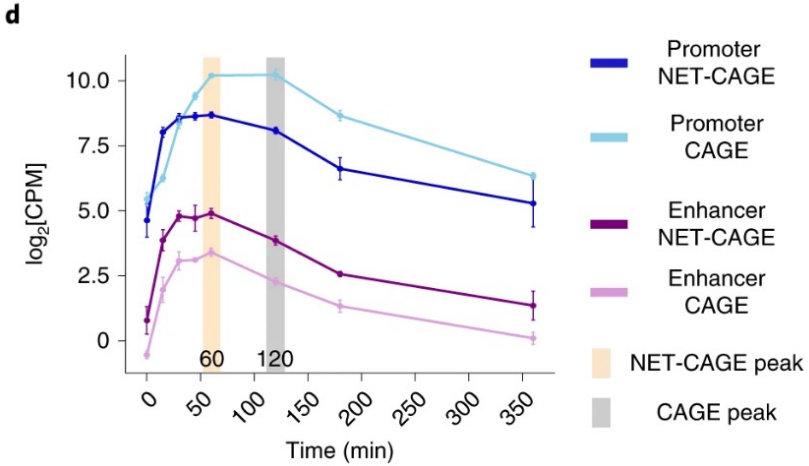
Henry F. Thomas

Why would you care?

- Detailed information about position of transcribing polymerase, TSS, splicing…
- Temporal resolution



Haberle&Stark, 2018



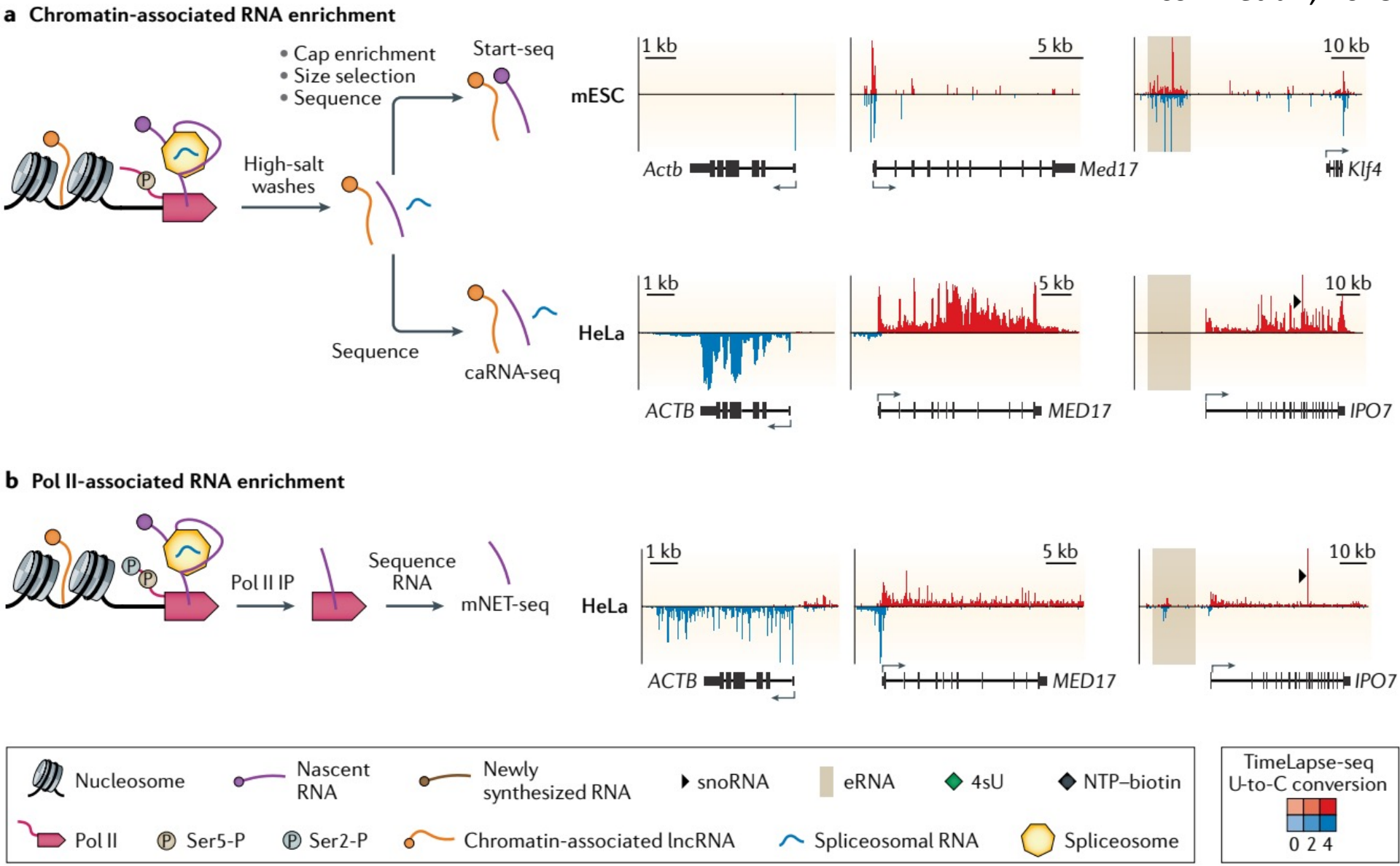
Hirabayashi *et al*., 2019

Why would you care?

- Detailed information about position of transcribing polymerase, TSS, splicing…
- Temporal resolution
- Transcriptional vs post-transcriptional regulation

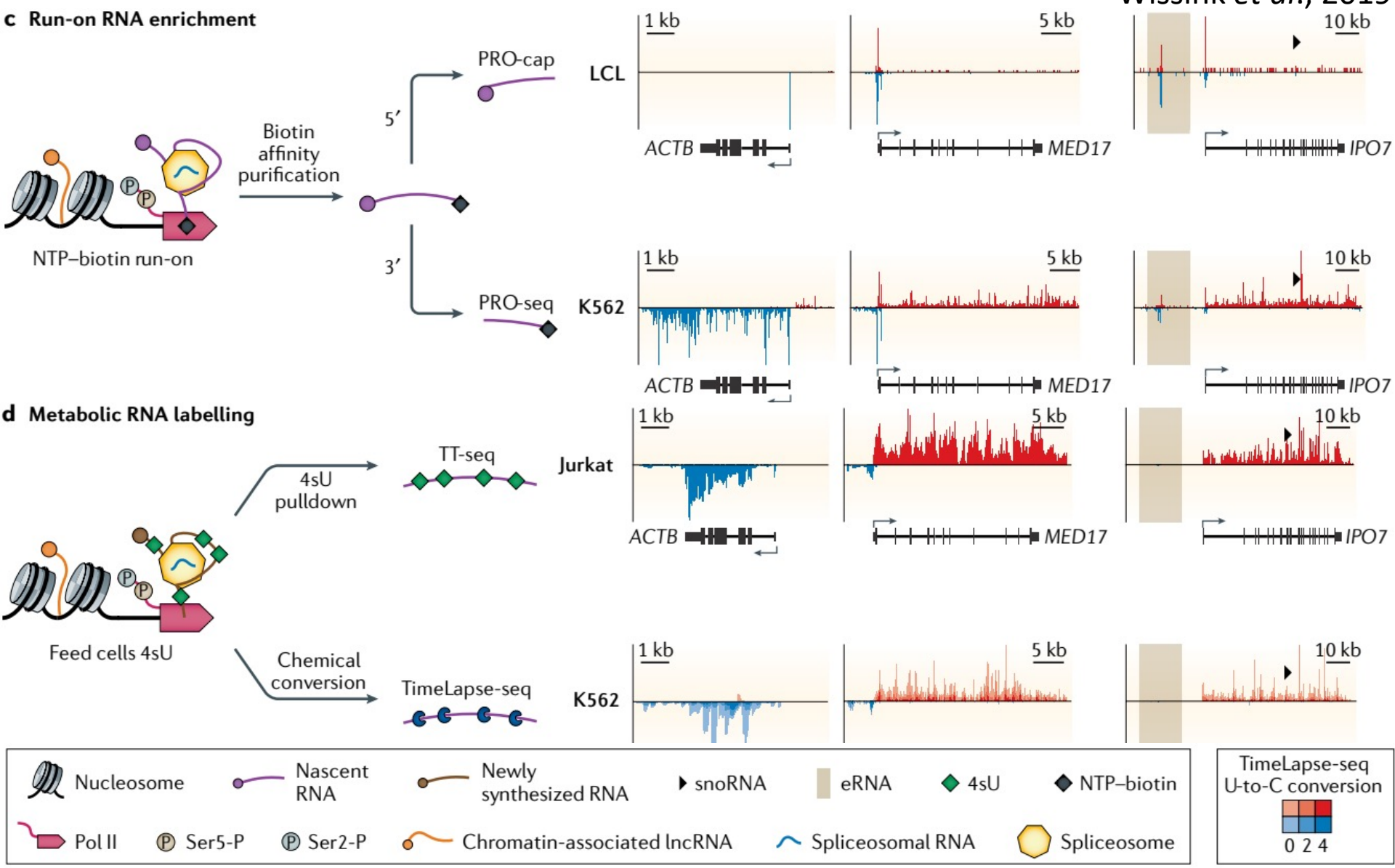
Principles of nascent RNA isolation

Wissink *et al.*, 2019



Principles of nascent RNA isolation

Wissink *et al.*, 2019



Comparison of methods

Table 1 | **Strengths and limitations of nascent RNA methods**

Method	Advantages	Considerations
caRNA-seq	• Can be used to isolate all chromatin-associated RNA species • Can be combined with methods that assay co-transcriptional processes, including RNA methylation and editing	Also sequences non-nascent RNAs that stably associate with chromatin
Start-seq	• Simultaneously identifies initiation and pausing sites • Allows de novo calling of putative enhancers	Does not report transcription beyond the first ~100 nucleotides
Yeast NET-seq	• Is Pol II specific (antibody enrichment) • Identifies Pol II positions at nucleotide resolution genome-wide	Is limited to cells with epitope-tagged Pol II
mNET-seq	• Is Pol II specific (antibody enrichment) • Identifies Pol II positions at nucleotide resolution genome-wide • Can isolate Pol II with different post-translational modifications	• Includes RNAs that are stably associated with Pol II • Does not currently include RNA <30 nucleotides in length • Has detected eRNA transcription from previously called enhancers
PRO-cap	• Identifies transcription initiation sites • Allows de novo calling of putative enhancers	Does not report transcription beyond the first ~100 nucleotides
PRO-seq	• Captures RNAs from transcriptionally competent polymerases • Identifies positions of active transcription at nucleotide resolution genome-wide • Allows de novo calling of putative enhancers	• Does not measure polymerase backtracking • Also captures RNAs being transcribed from Pol I and Pol III
CoPRO	• Simultaneously identifies initiation and pausing sites • Measures RNA capping status	Does not measure transcription beyond promoter-proximal pause site
SMIT-seq	Measures splicing status during transcription	Limited to species with short introns
TT-seq	• Captures RNAs from actively transcribing polymerases • Can be used to determine RNA stability • Identifies transcription termination sites	• Does not detect Pol II pausing • Has detected eRNA transcription from previously called enhancers
SLAM-seq and TimeLapse-seq	• Captures RNAs from actively transcribing polymerases • Can be used to determine RNA stability	• Requires deep sequencing to measure chemical conversion rate • Long labelling times do not capture newly synthesized RNA
Intron sequential FISH	• Detects transcription of thousands of genes in single cells • Contains positional information of transcribed genes in the 3D space of the nucleus	• Does not report chromosomal positions of active Pol II complexes • Does not distinguish different steps of transcription • Requires a library of intron-targeting probes and series of hybridizations

caRNA-seq, chromatin-associated RNA sequencing; CoPRO, coordinated precision run-on and sequencing; eRNA enhancer RNA; FISH, fluorescence in situ hybridization; mNET-seq, mammalian native elongating transcript sequencing; NET-seq, native elongating transcript sequencing; Pol, RNA polymerase; PRO-cap, precision run-on with cap selection; PRO-seq, precision run-on sequencing; SLAM-seq, thiol (SH)-linked alkylation for the metabolic sequencing of RNA; SMIT-seq, single-molecule intron tracking sequencing; TT-seq, transient transcriptome sequencing.

Wissink *et al.*, 2019

Principles of nascent RNA isolation

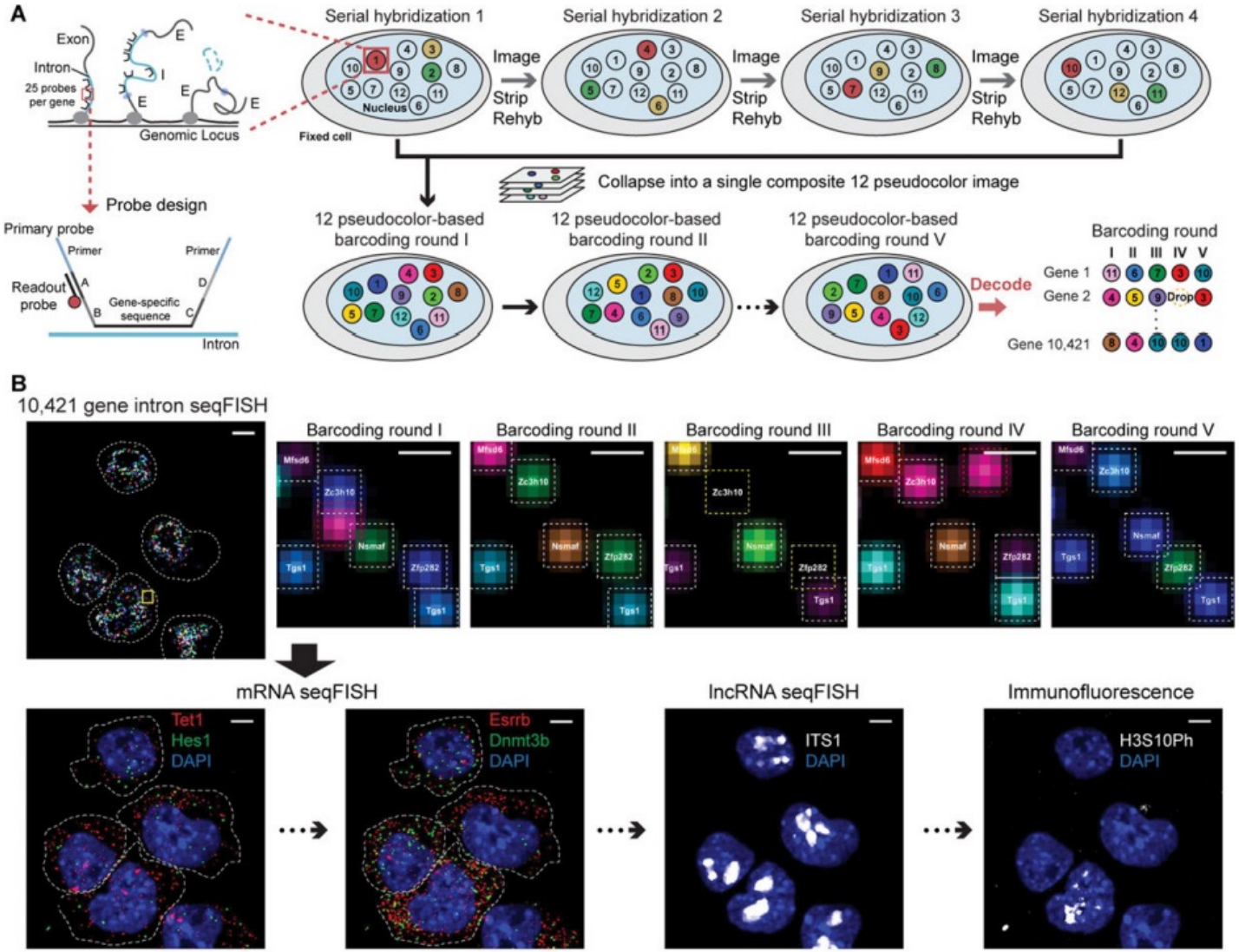
Table 2 | **Methods used to investigate different steps of transcription**

Method	Transcription step						
	TSS[a]	RNA capping	Promoter-proximal pausing	Co-transcriptional RNA processing	Transcription termination	Pol II CTD modification	Transcription bursting
Chromatin isolation-based methods							
caRNA-seq	No	No	No	Yes[42,105–107]	No	No	No
Start-seq	Yes[43]	No	Yes[43]	No	No	No	No
mNET-seq	No	No	Yes[41,73]	Yes[41,63,64]	Yes[41]	Yes[41,63]	No
SMIT-seq	No	No	No	Yes[159,160]	No	No	No
Run-on methods							
GRO-cap and PRO-cap	Yes[4,42]	No	No	No	No	No	No
GRO-seq, PRO-seq and ChRO–seq	No	No	Yes[42,48,74]	Yes[166]	Yes[42]	No	No
CoPRO	Yes[49]	Yes[49]	Yes[49]	No	No	No	No
Metabolic labelling methods							
TT-seq	No	No	No	No	Yes[47]	No	No
Imaging-based methods							
Intron sequential FISH	No	No	No	No	No	No	Yes[55]

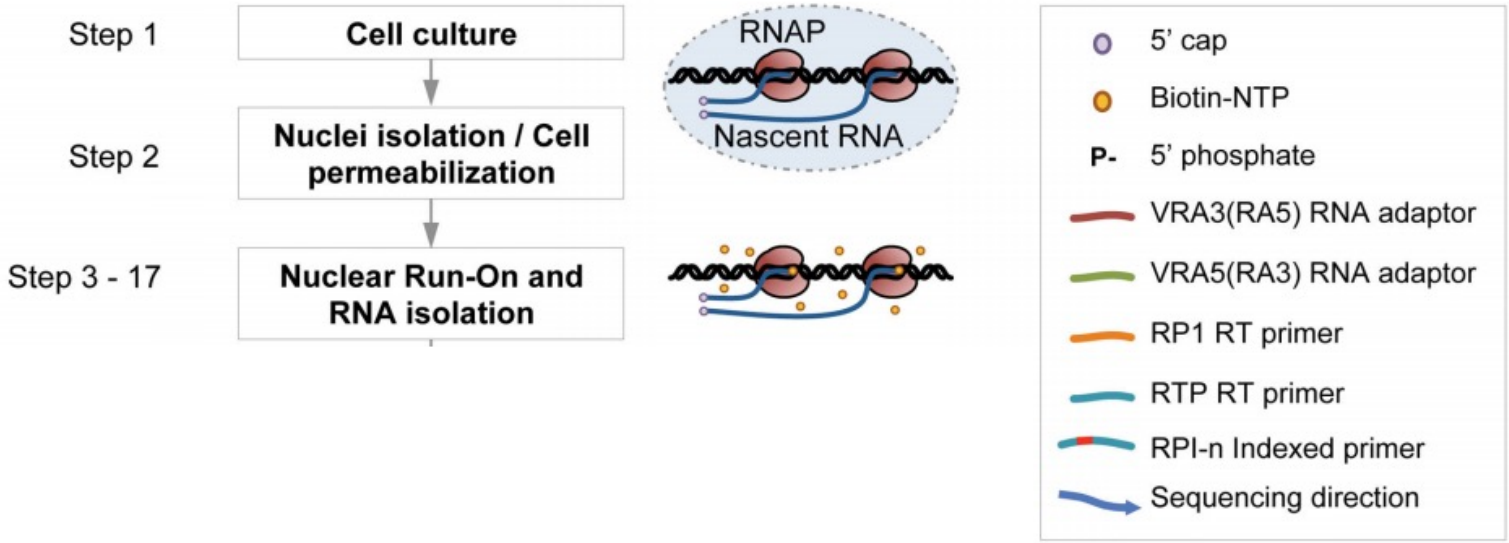
caRNA-seq, chromatin-associated RNA sequencing; ChRO–seq, chromatin run-on and sequencing; CoPRO, coordinated precision run-on and sequencing; CTD, C-terminal domain; FISH, fluorescence in situ hybridization; GRO-cap, global run-on with cap selection; GRO-seq, global run-on sequencing; mNET-seq, mammalian native elongating transcript sequencing; Pol II, RNA polymerase II; PRO-cap, precision run-on with cap selection; PRO-seq, precision run-on sequencing; SMIT-seq, single-molecule intron tracking sequencing; TSS, transcription start site; TT-seq, transient transcriptome sequencing. [a]'Yes' indicates that the method was designed specifically to detect TSSs, but GRO-seq can infer TSSs, as can PRO-seq and mNET-seq with sufficiently long reads.

Wissink *et al.*, 2019

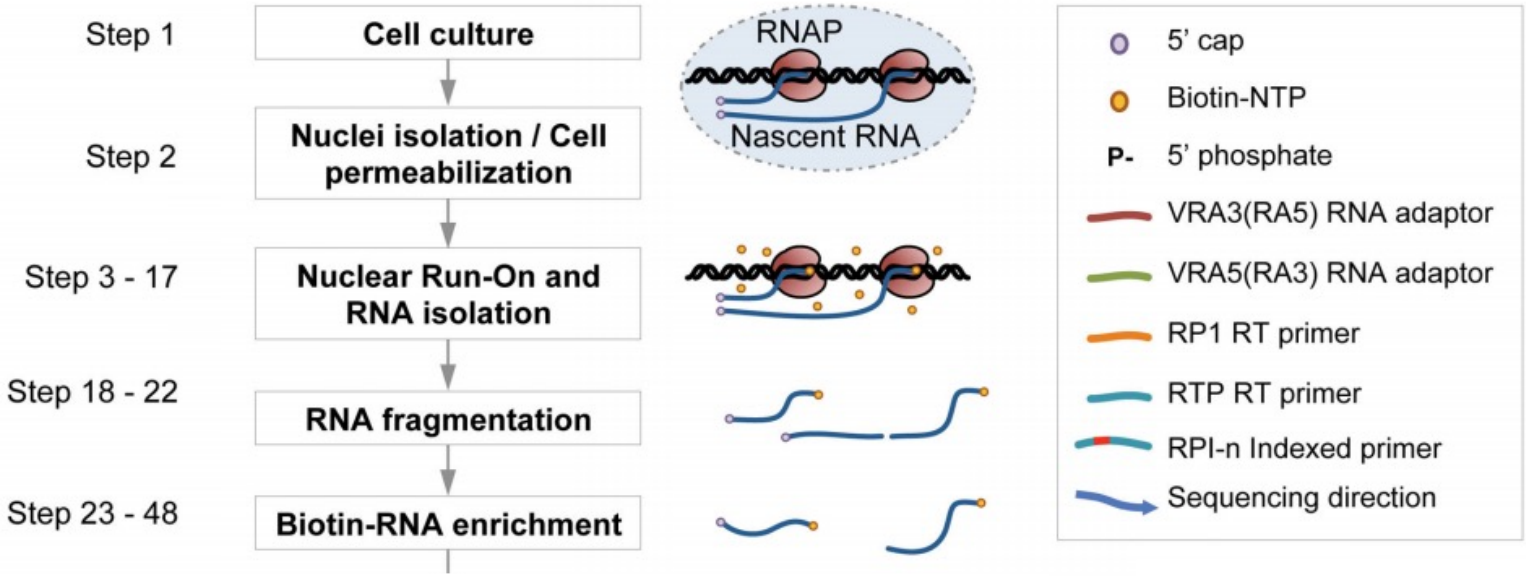
Intron seqFISH



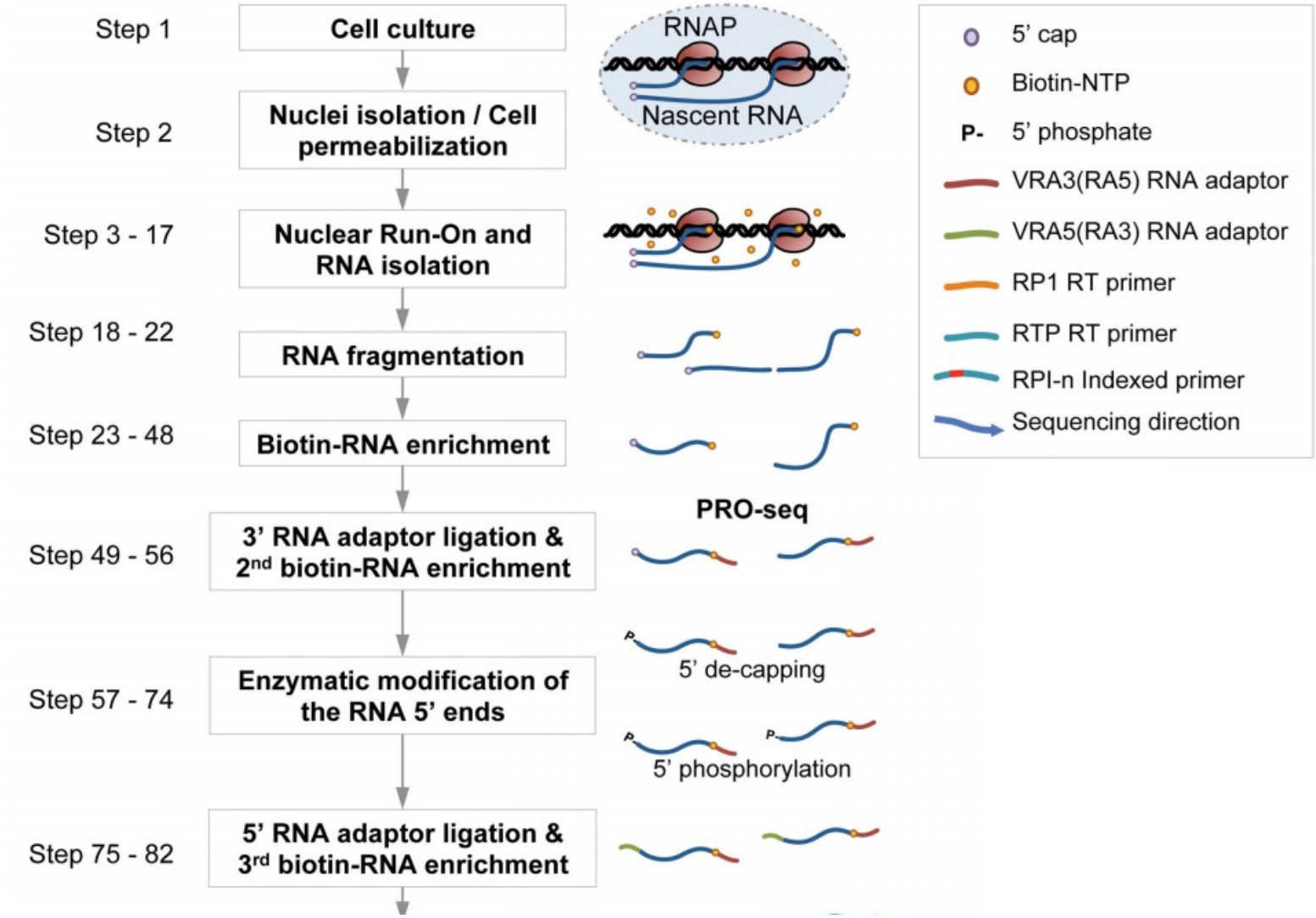
Shah *et al*., 2018



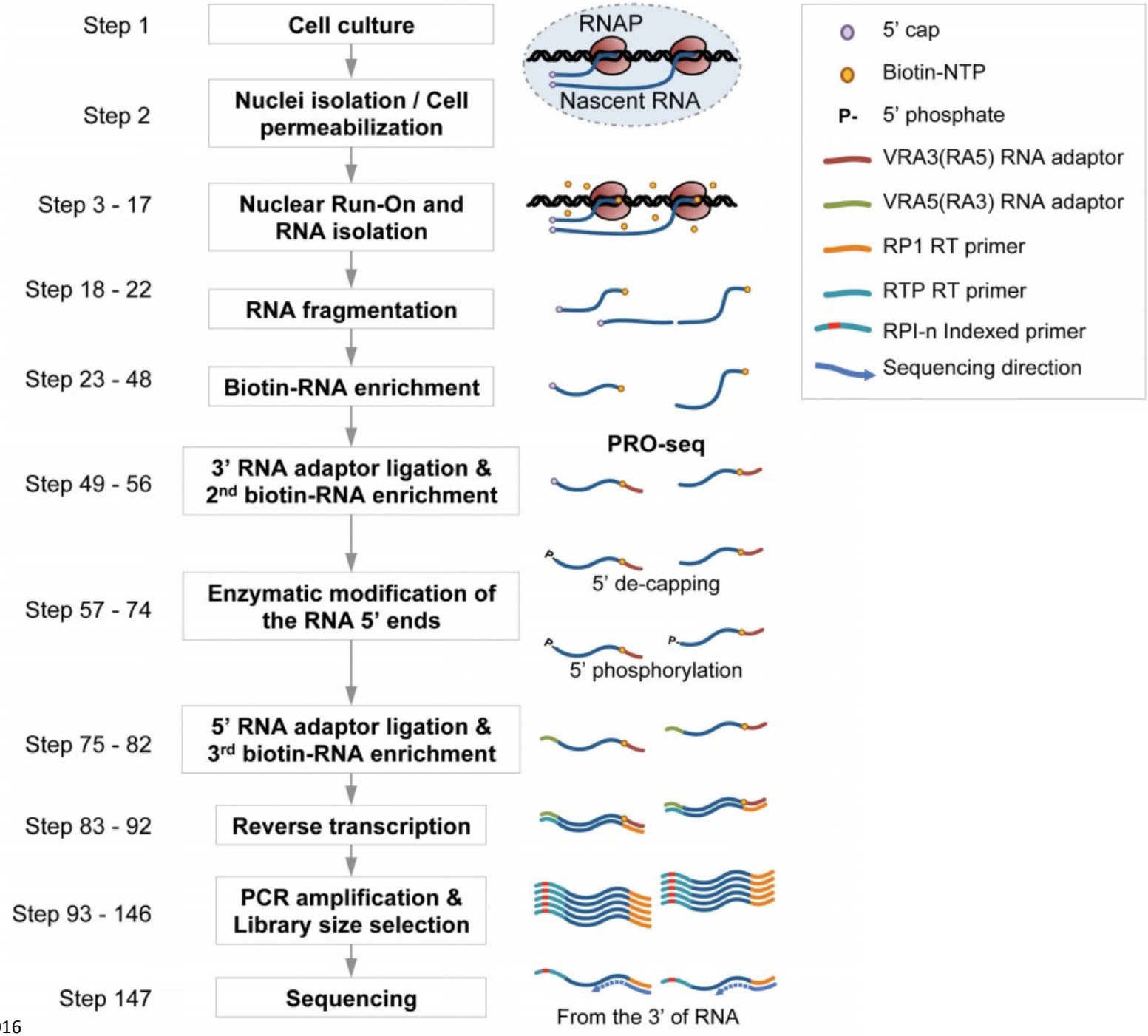
Mahat *et al.*, 2016



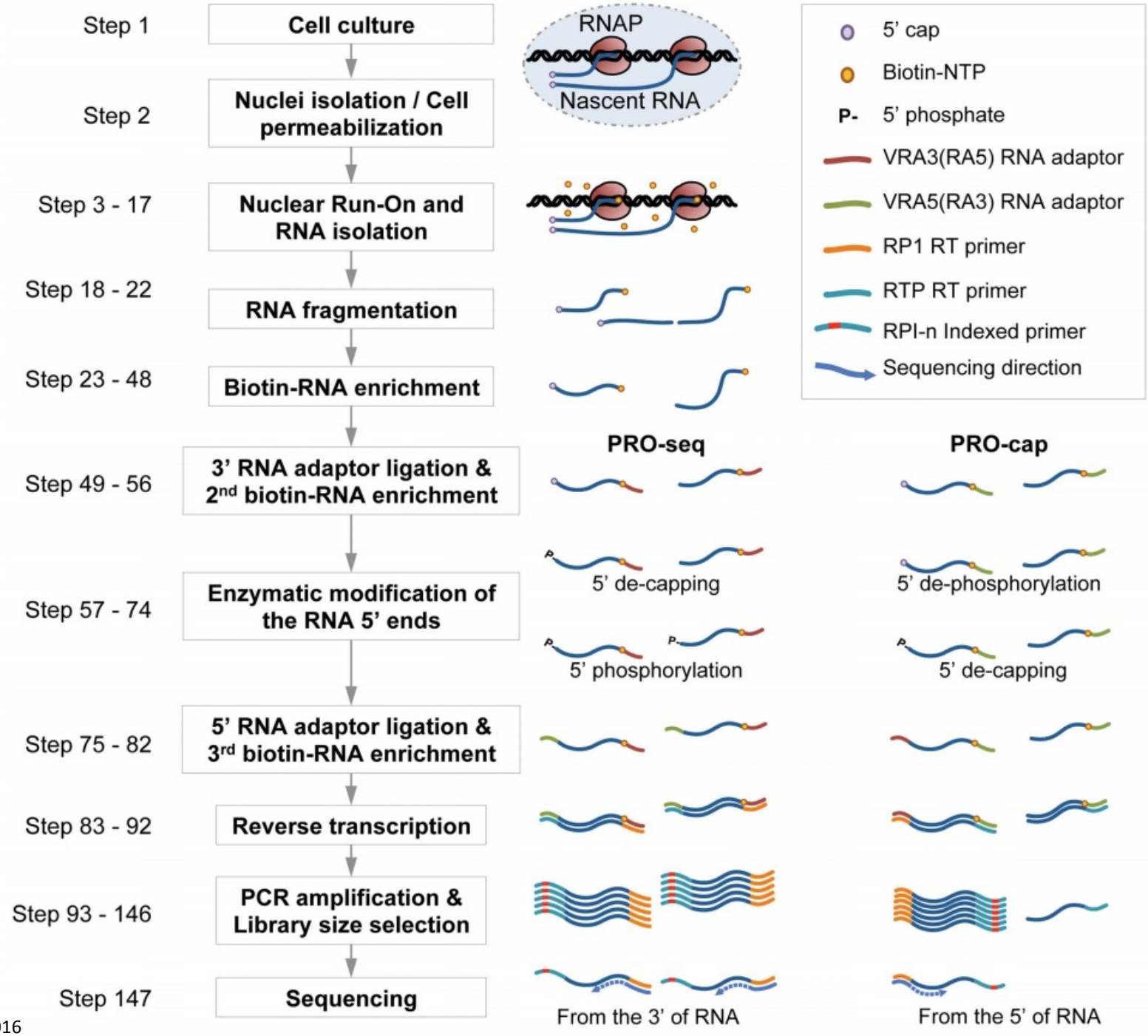
Mahat *et al.*, 2016



Mahat *et al.*, 2016



Mahat *et al.*, 2016

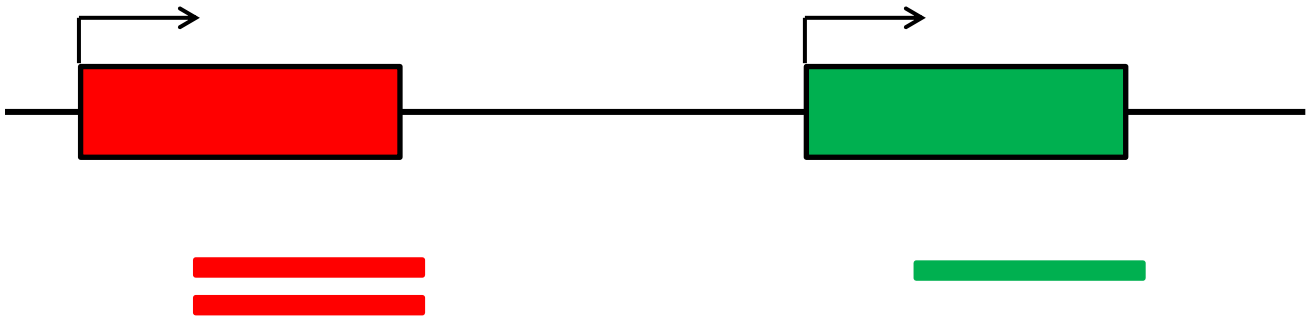


Mahat *et al.*, 2016

Data analysis

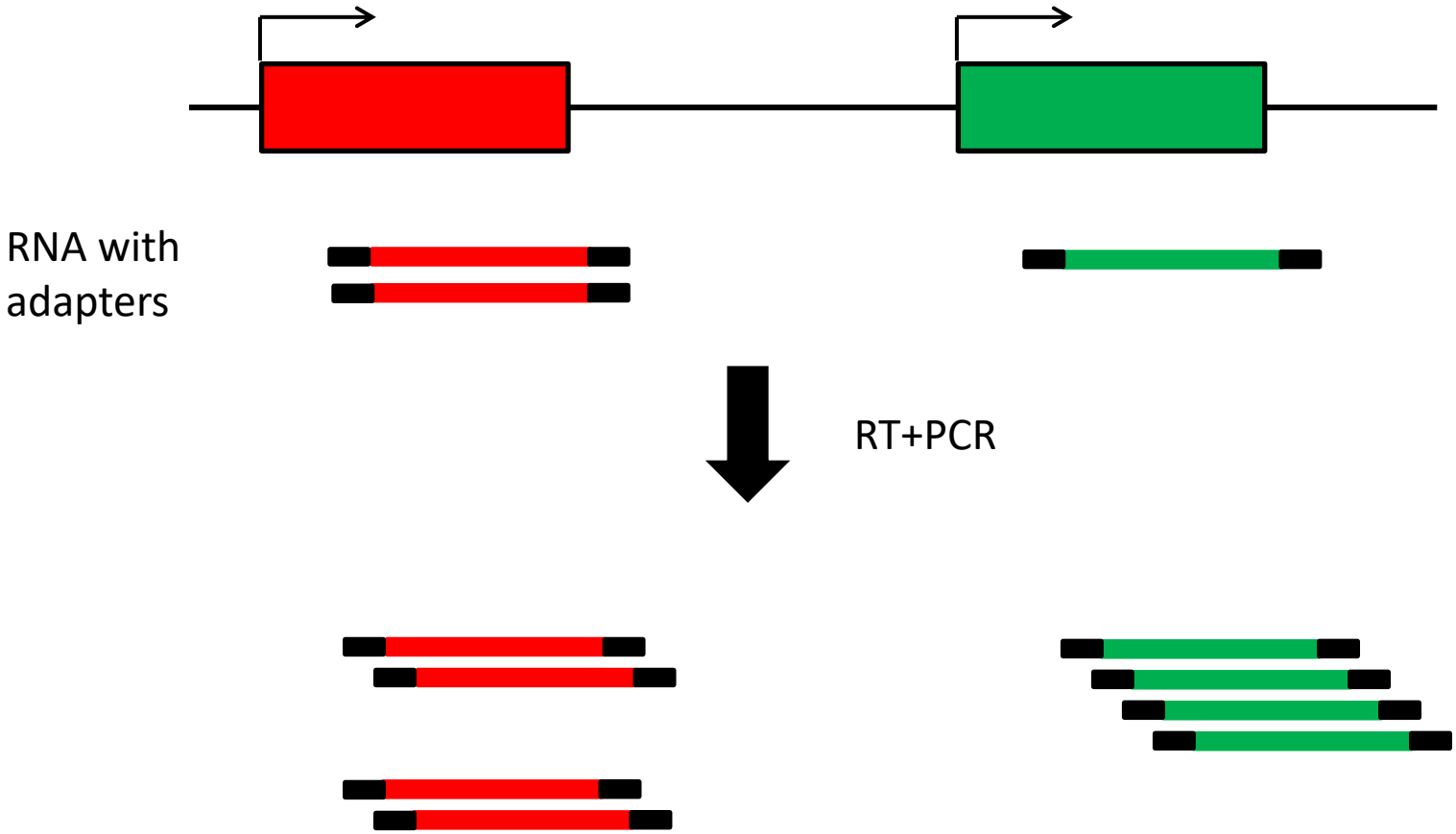
- Spike-Ins
- Random nucleotides

Why random nucleotides?



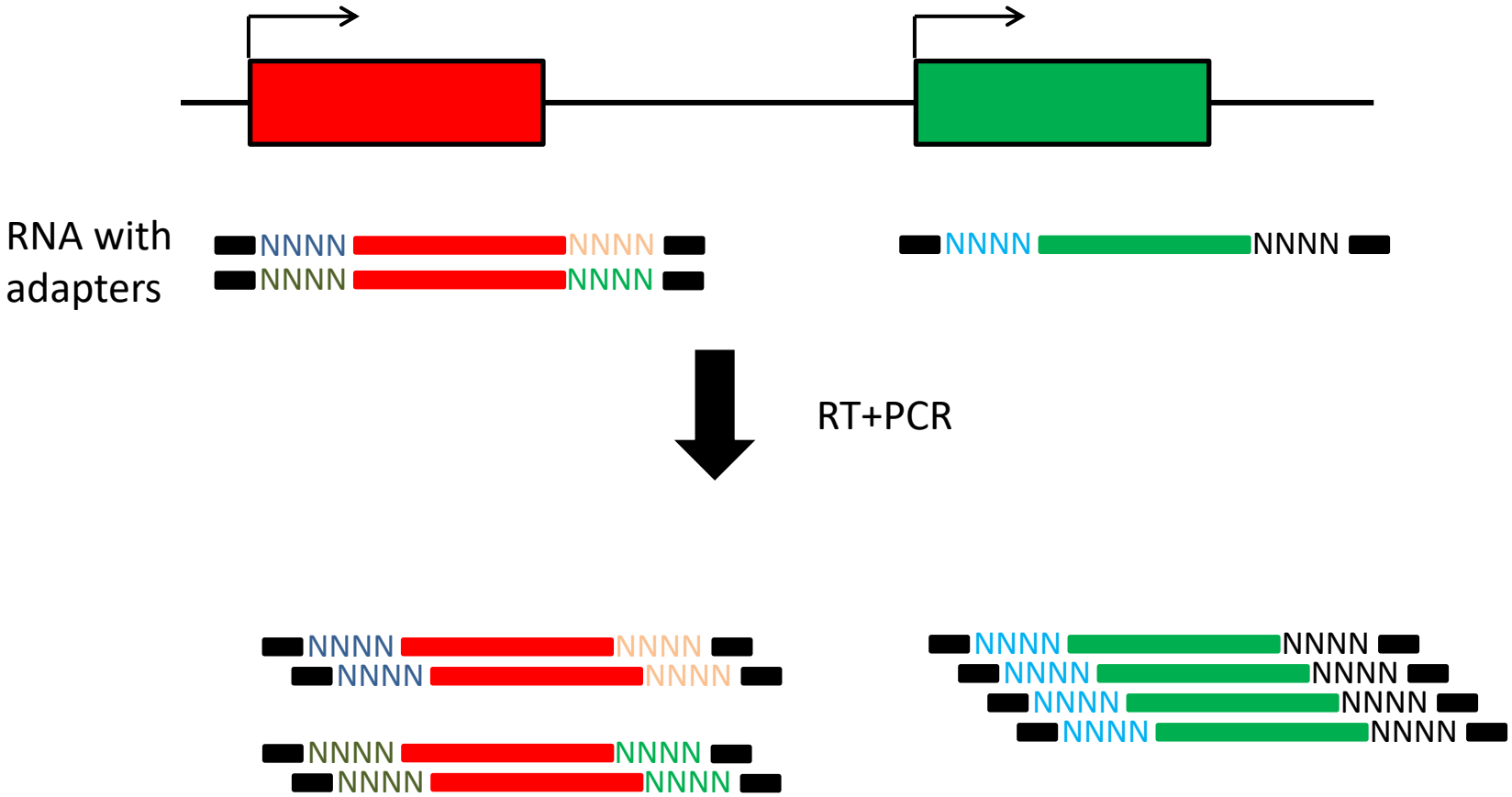
Why random nucleotides?

- Adapter ligation



Why random nucleotides?

- Adapter ligation



Data analysis work flow

- Eliminate duplicates (dedupe)

Data analysis work flow

- Eliminate duplicates (dedupe)
- Eliminate adaptors and random nmers (trim_galore)
- Map reads to genome (bowtie2; merged genome from 2 species!)
- Separate reads by which genome and which strand they map to (samtools view –b and –f/-F 16; keep in mind that reads are reverse complement of actual trasncripts)
- (optional: clip all but first base pair of every read, gatkClipReads)
- Normalise to Spike-In sequencing depth while transforming to .bedgraph file (genomeCoverageBed)
- Transform to .bw files (bedGraphToBigWig), visualise on genome browser
- Use featureCounts function in R to count reads in every gene and do e. g. differential expression analysis (manually divide resulting counts by Spike-In sequencing depth)