

BIOL 343

Applied Bioinformatics I

Library preparation

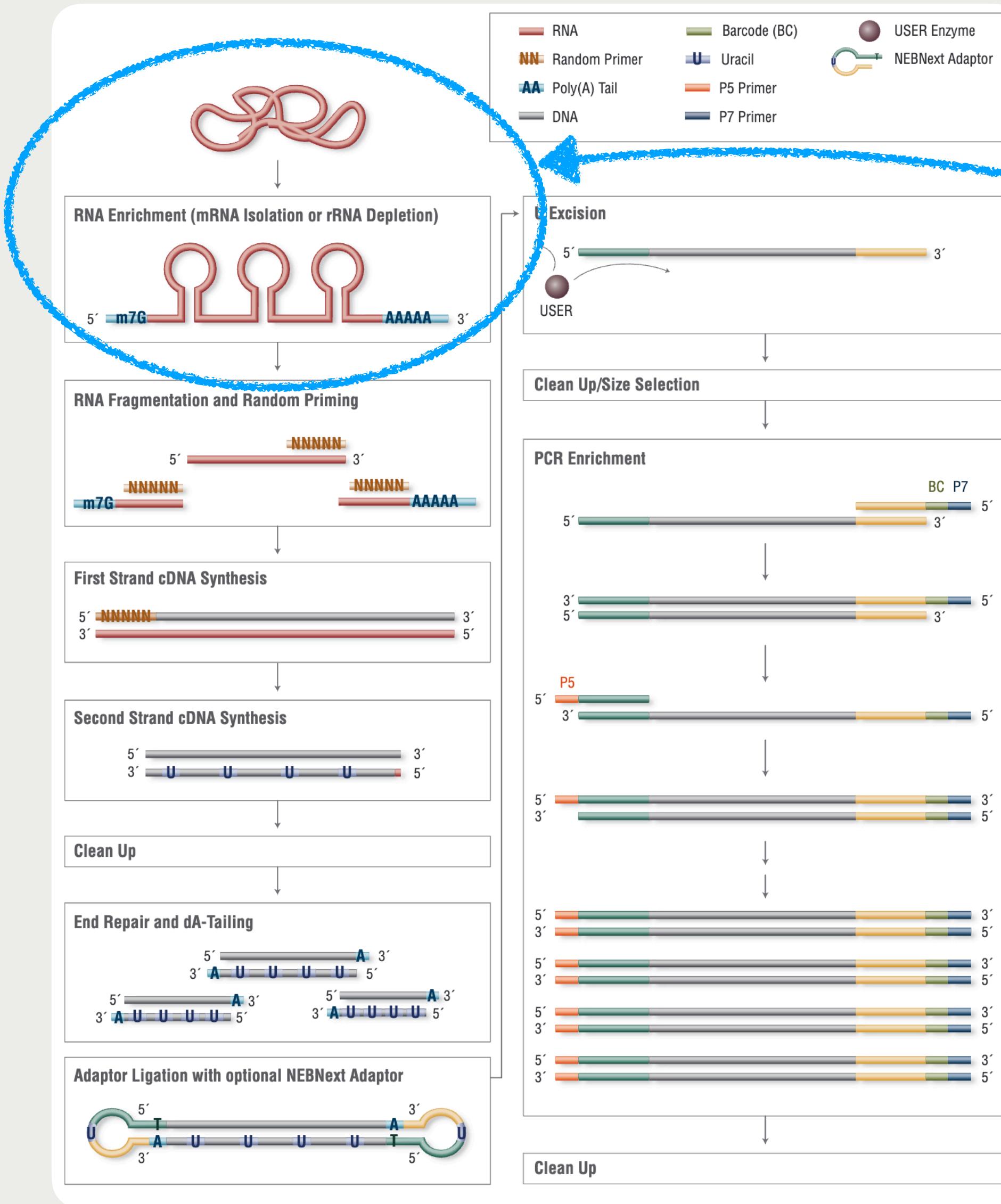
Dr. Nic Wheeler

Learning Objectives

You will be able to:

1. Describe the process of preparing an RNA sequencing library
2. Understand how library prep choices such as read length, adaptors, barcodes, UMIs, etc. contribute to downstream steps in data analysis
3. Use bioanalyzer traces to assess steps of the library prep process

We have high quality RNA... ...now what?

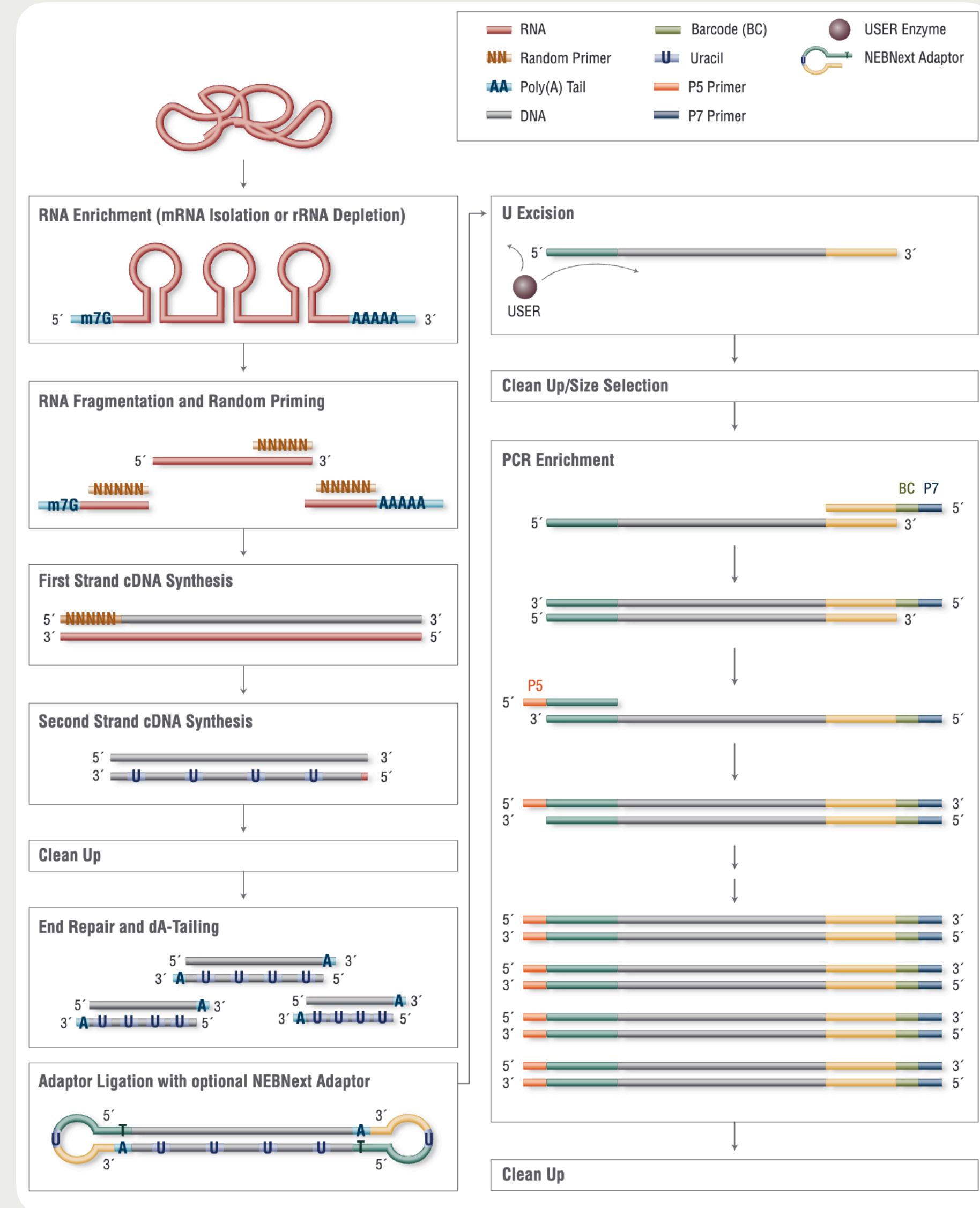


Recall last week...

1. Different ways to extract RNA
(guanidium thiocyanate-phenol; silica column)
2. Poly(A) enrichment and/or rRNA depletion
3. RNA QC

We have high quality RNA...

...now what?



Generally, RNA is not directly sequenced:

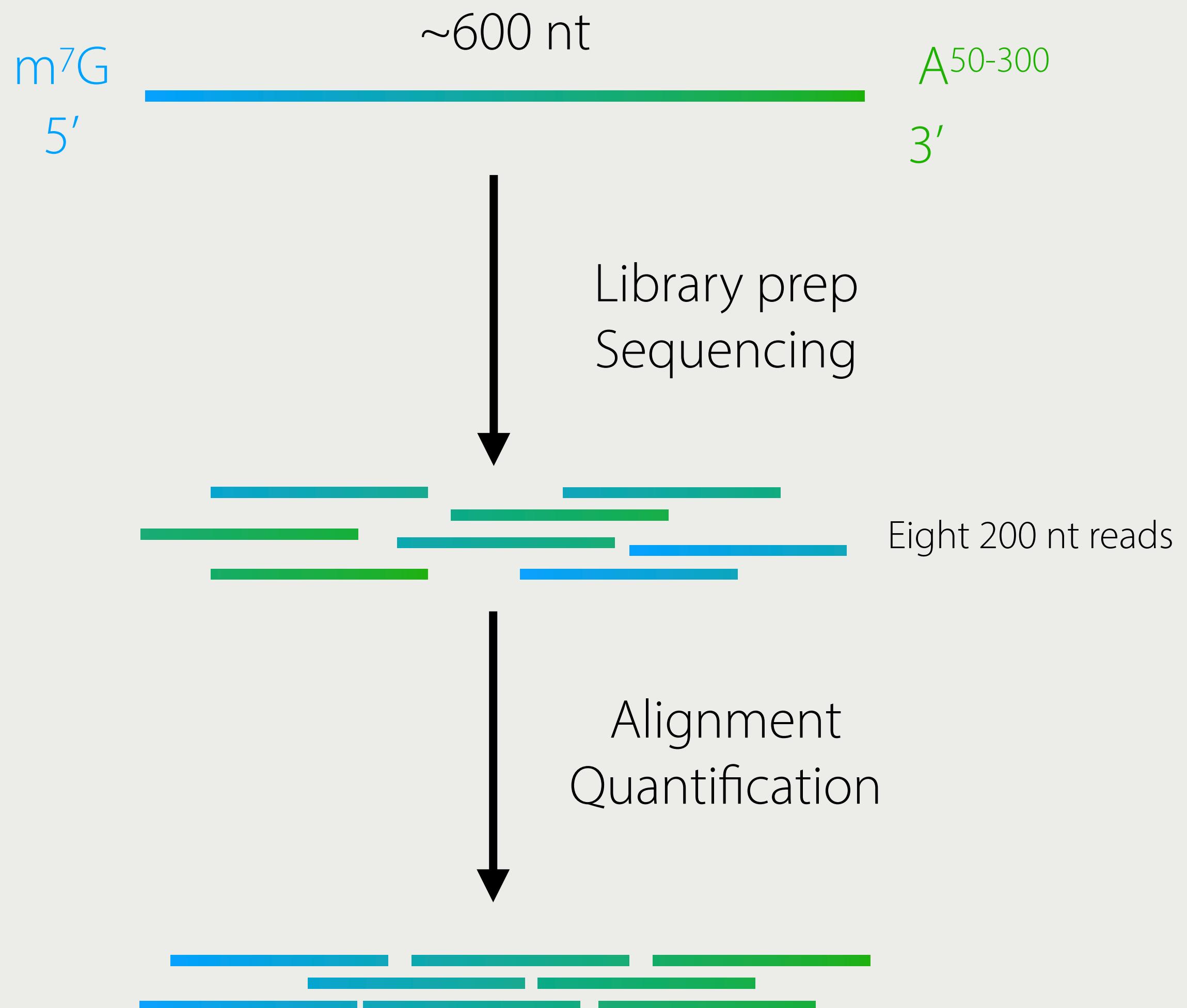
1. RNA is very unstable
2. RNA sequencers are only able to read specific lengths or nucleic acids (less than the length of a typical mRNA)
3. (Sometimes) low abundance RNA needs to be amplified prior to sequencing
4. Sequencers produce more data than necessary, so samples are pooled and need to be able to be deconvoluted

mRNA is converted to a **sequencing library**

A bit about reads

mRNAs are sequenced as “reads” that are then “aligned”

- A **read** is an individual unit of sequencing data
- Read lengths are dictated by the library prep kit and the sequencer
- The absolute number of reads generated is dictated by the sequencer (millions to billions)
- Reads can be **paired-end** or **single-end**
- The job of the bioinformatician is to convert reads into actionable data (i.e., relative mRNA quantities)



Library preparation prepares RNA for sequencing

Different protocols for different objectives and technologies

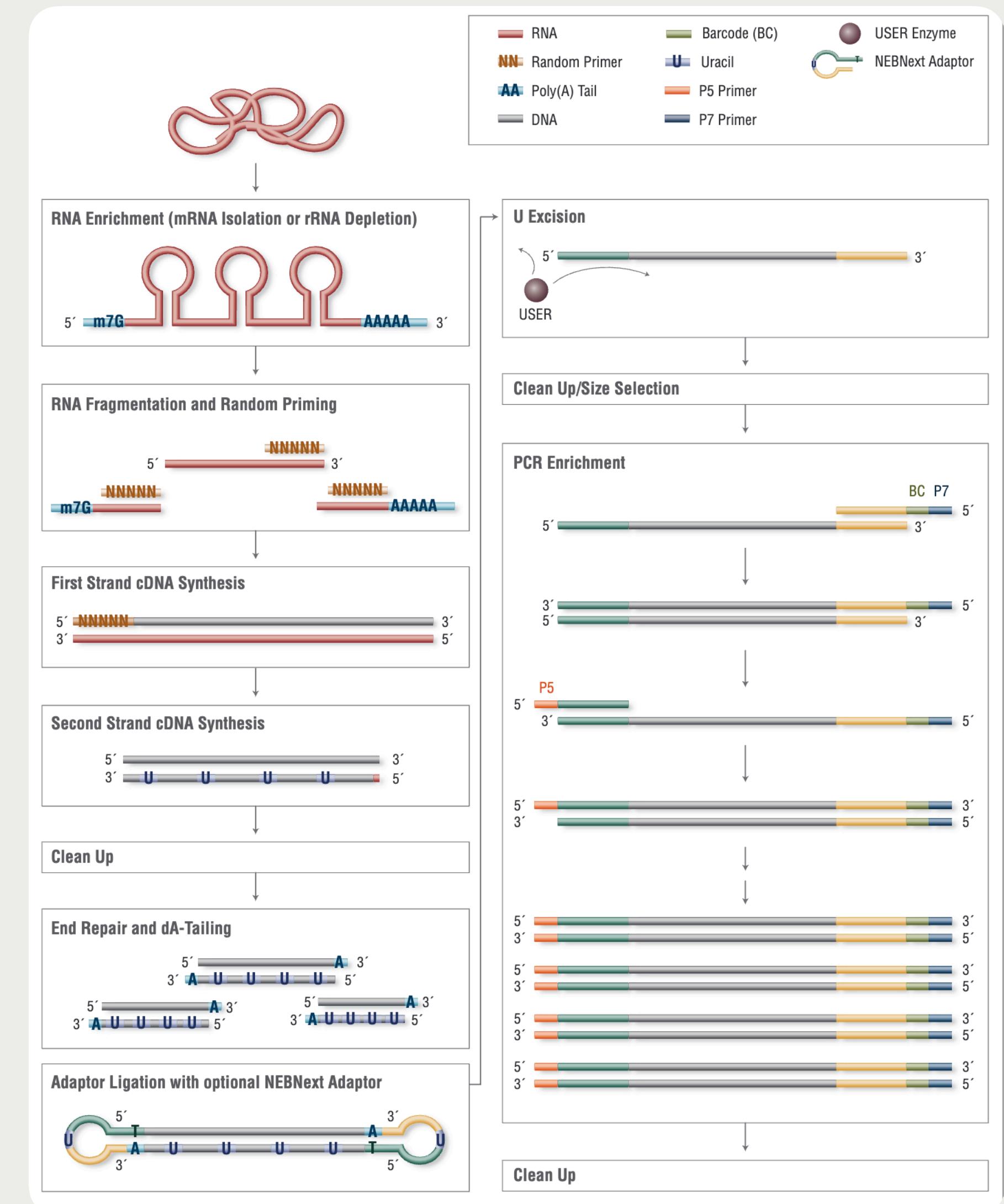
Three types of RNA sequencing:

1. Short-read (Illumina)
2. Long-read (PacBio)
3. Long/direct (Oxford Nanopore Tech)

General protocol for Illumina-compatible library prep

NEBNext Ultra II Directional RNA Workflow

1. RNA extraction and enrichment
2. RNA fragmentation
3. Priming and first strand cDNA synthesis
4. Second strand cDNA synthesis
5. End repair and A-tailing
 - a. U excision
6. Size selection
7. PCR enrichment/amplification
 - a. (Index PCR)



General protocol for Illumina-compatible library prep

NEBNext Ultra II Directional RNA Workflow

A word on biotech companies...

illumina

Primary manufacturer of short-read sequencers

Makes kits for these sequencers; specific chemistry



Popular manufacturer for mol bio kits/reagents

Makes kits that are Illumina-compatible



Monopoly on biology sales

Sells/distributes Illumina and NEB reagents/kits

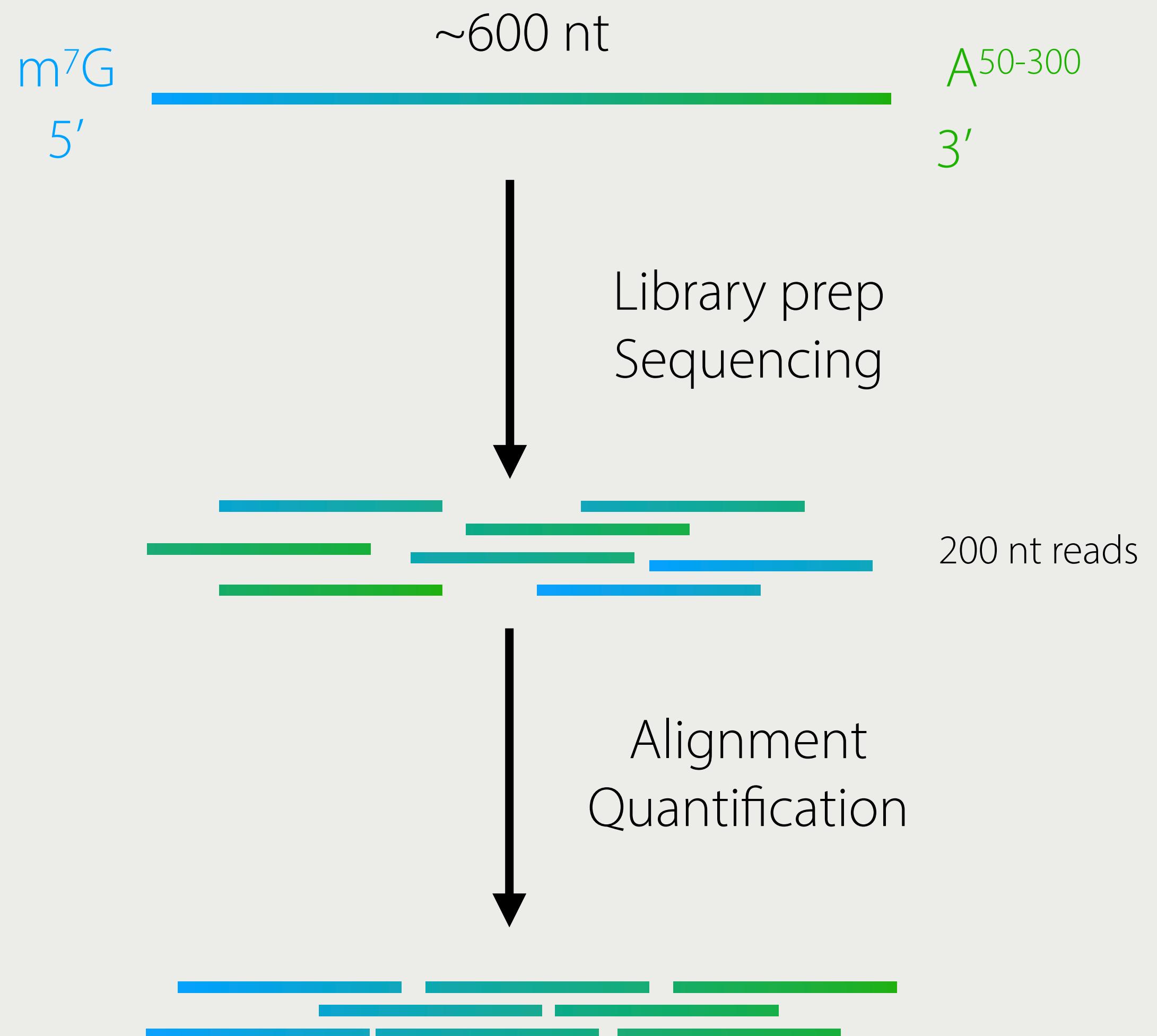
2. RNA fragmentation

Optimized to insert/read length

Remember...

1. mRNAs are often >1000 nt long (or much shorter/longer)
2. Illumina (short-read) technology read lengths are 50-300 nt
3. Alignment will put reads back together to generate the parental mRNA sequences

Read length < mRNA length

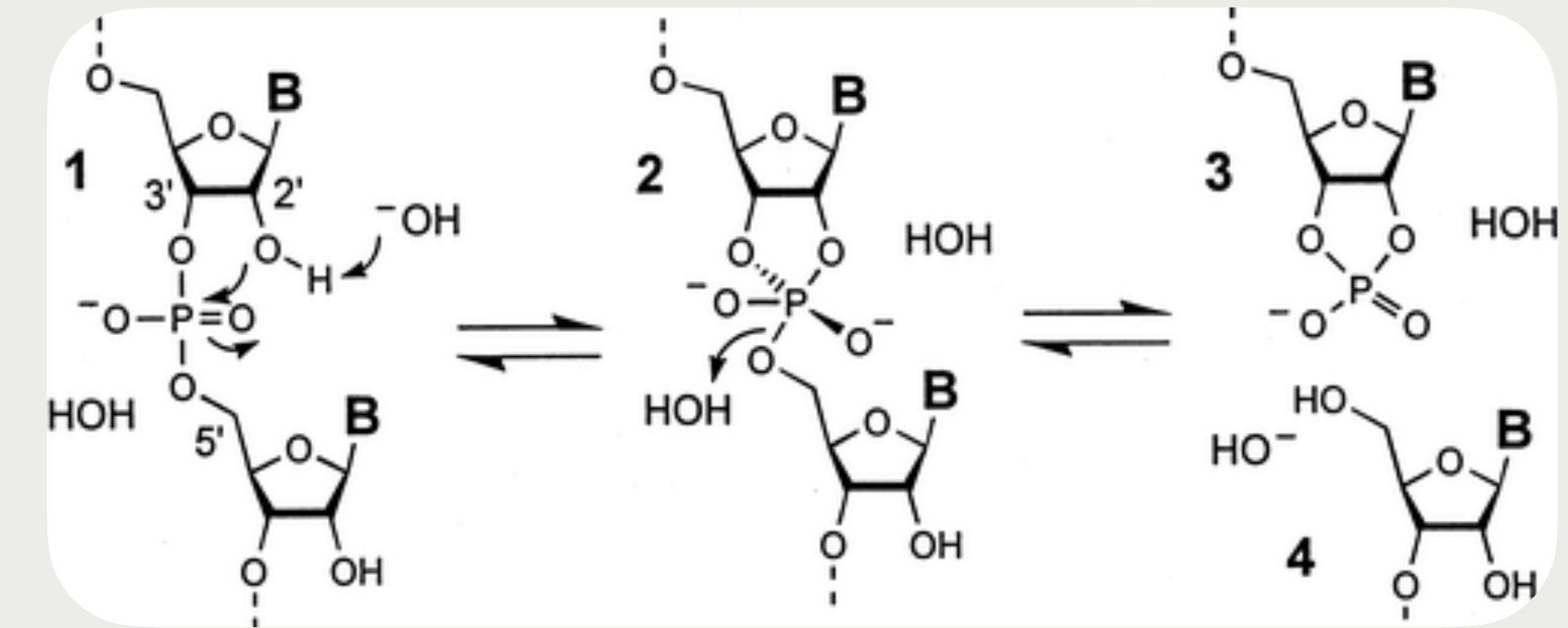


2. RNA fragmentation

Optimized to insert/read length

Multiple ways to fragment RNA:

- Enzymes
- Sonication or “acoustic shearing”
- **Heat + divalent cations (Mg^{2+})**
- Base-catalyzed cleavage
(transesterification)



Difference between RNA and DNA...

2. RNA fragmentation

Optimized to insert/read length

NEBNext Ultra II Directional

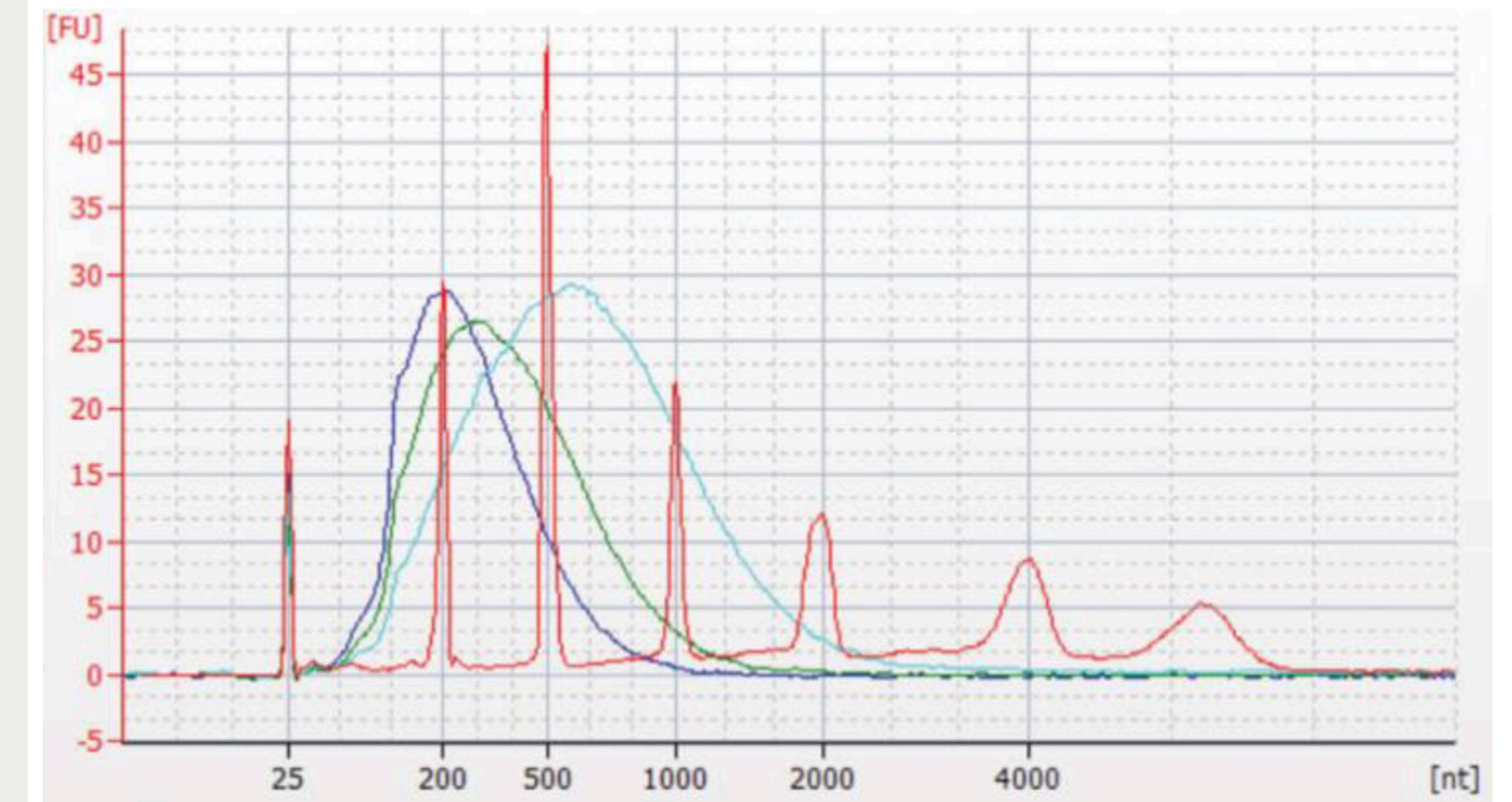
1.2.37 (page 7) - Incubate the sample in a thermal cycler with the heated lid set at 105°C as follows:

1. 15 minutes at 94°C[†]
2. Hold at 4°C*

* Immediately transfer the tube to ice for 1 minute as soon as it is cool enough to handle (~65°C)

† Optimized for 200 nt fragments (based on desired read length)

Figure 6.1. Modified fragmentation times for longer RNA inserts.



Red Ladder

Blue 150–300 bp, mRNA fragmented for 15 minutes at 94°C

Green 200–500 bp mRNA fragmented for 10 minutes at 94°C

Cyan 400–1,000 bp mRNA fragmented for 5 minutes at 94°C

Section 6, Appendix A, Figure 6.1

2. RNA fragmentation

Optimized to insert/read length

RNA prior to rRNA depletion (18S/28S peaks)
RNA after rRNA depletion (median length of ~1500 nt)

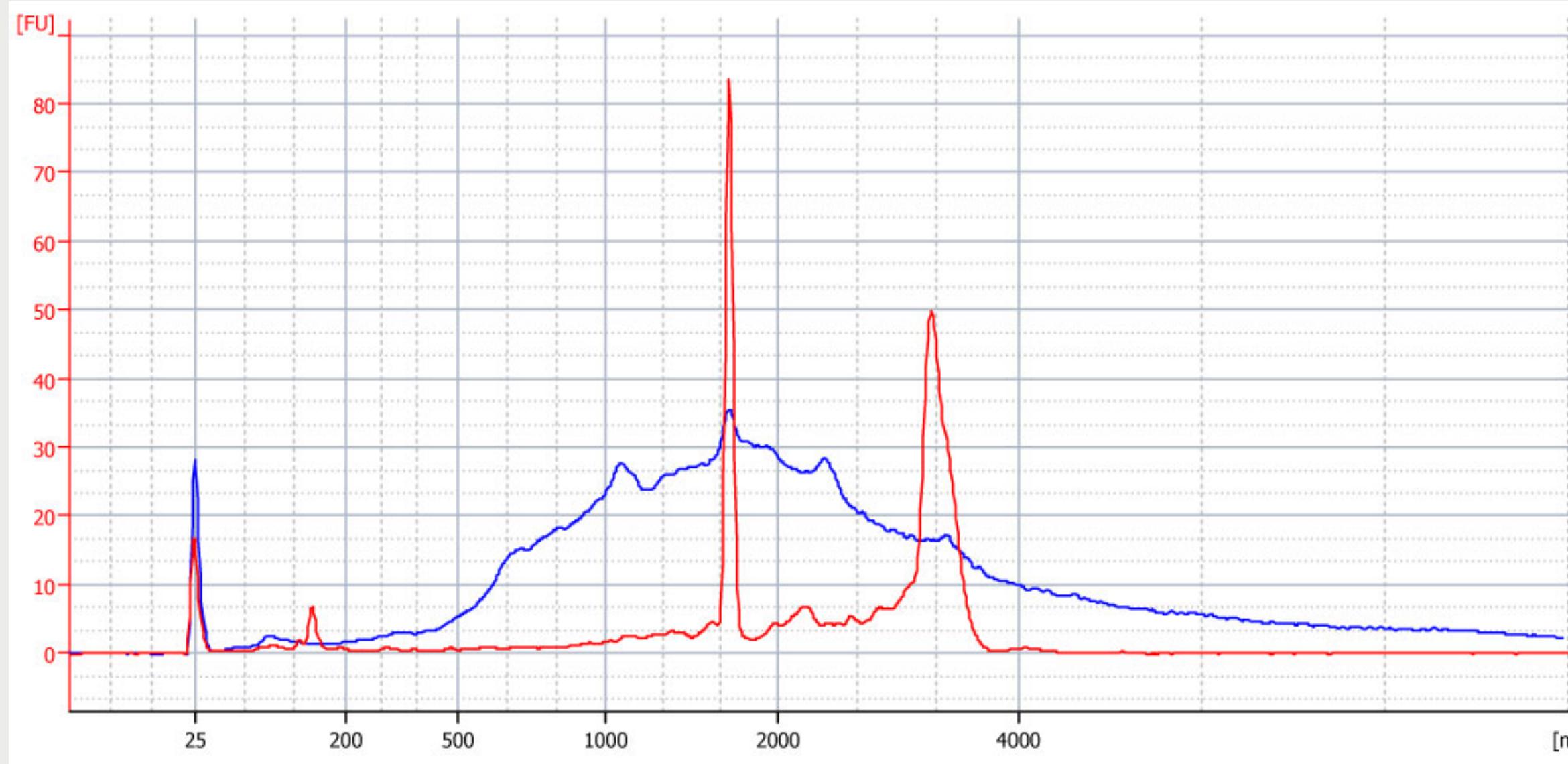
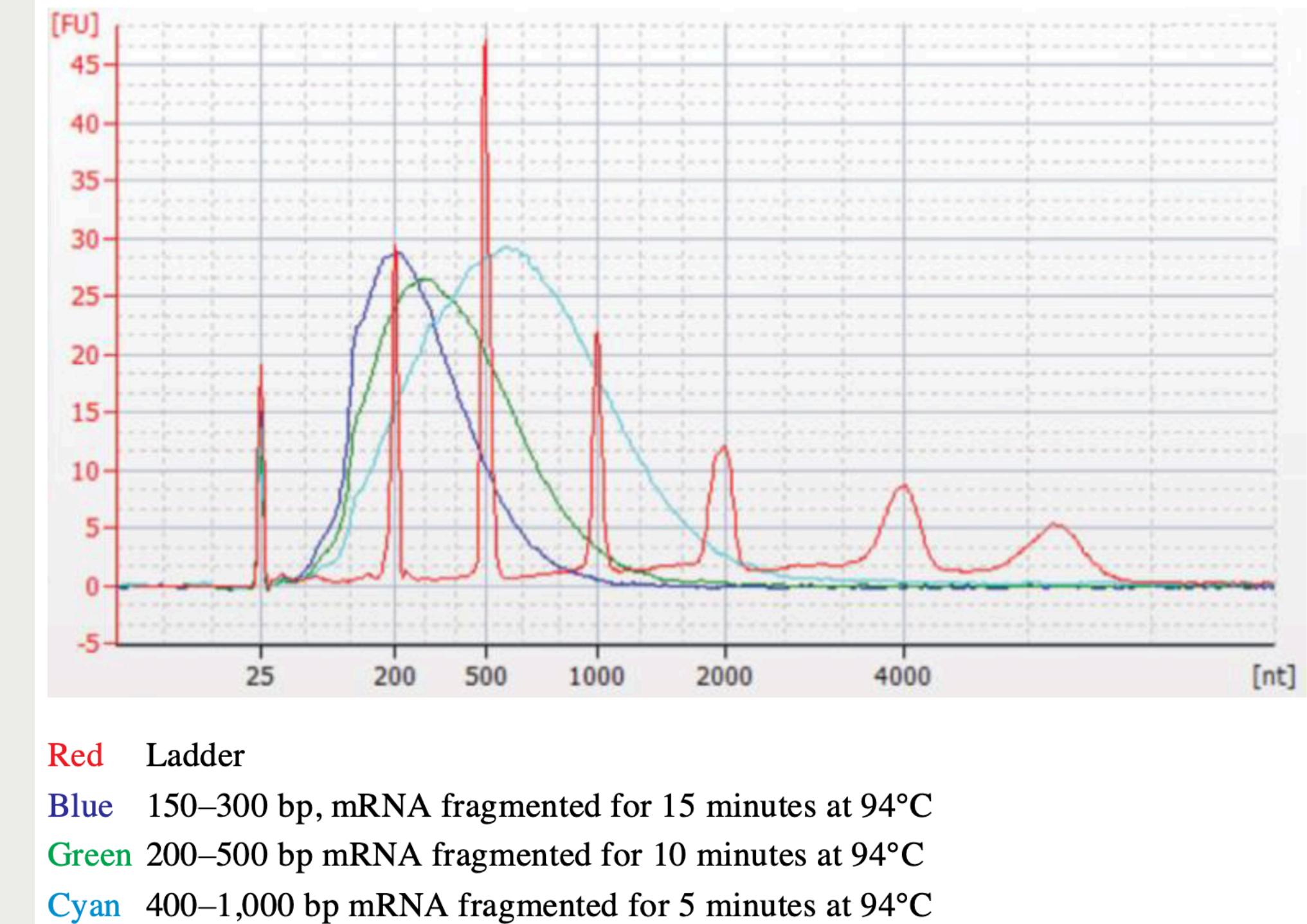
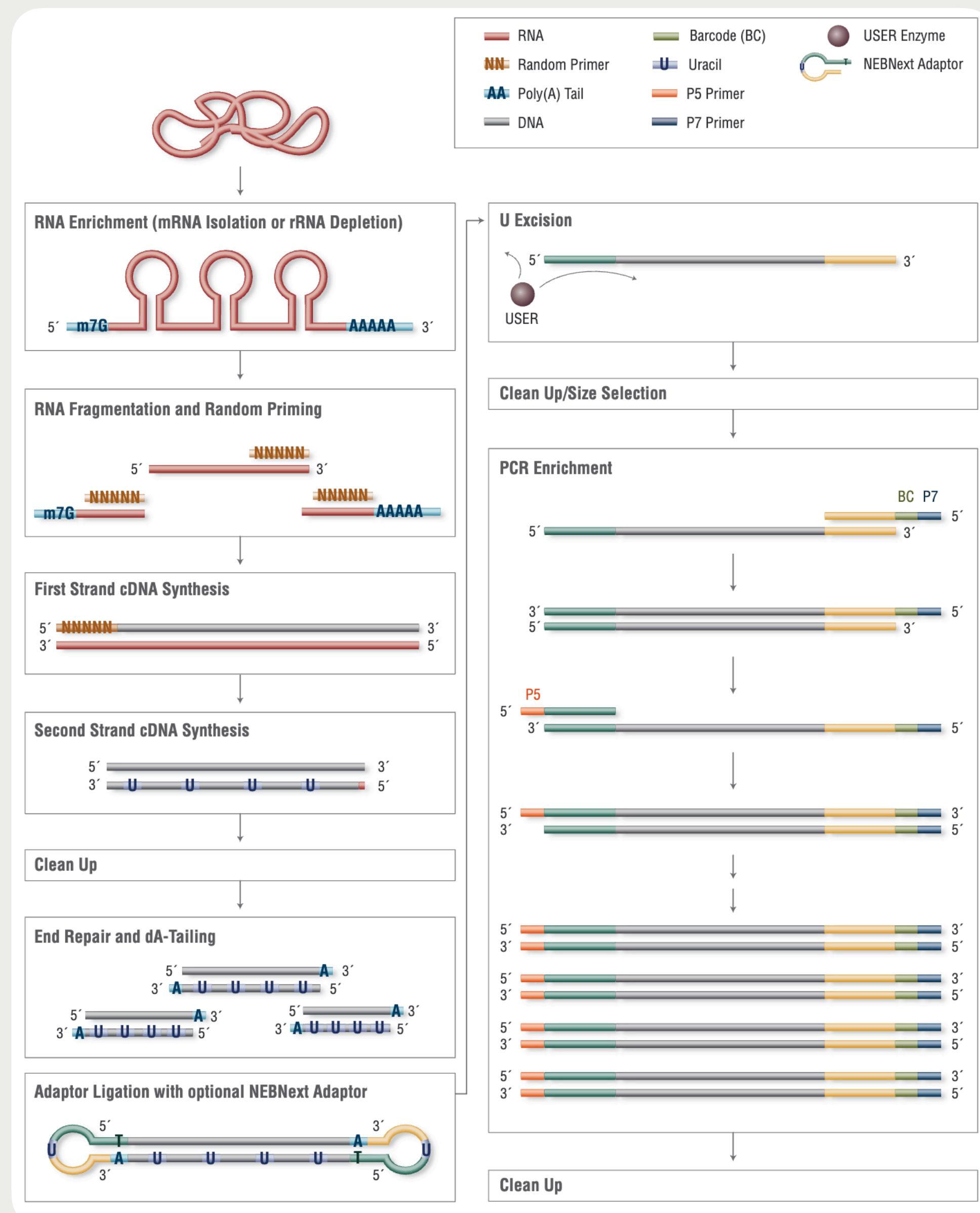


Figure 6.1. Modified fragmentation times for longer RNA inserts.



Section 6, Appendix A, Figure 6.1

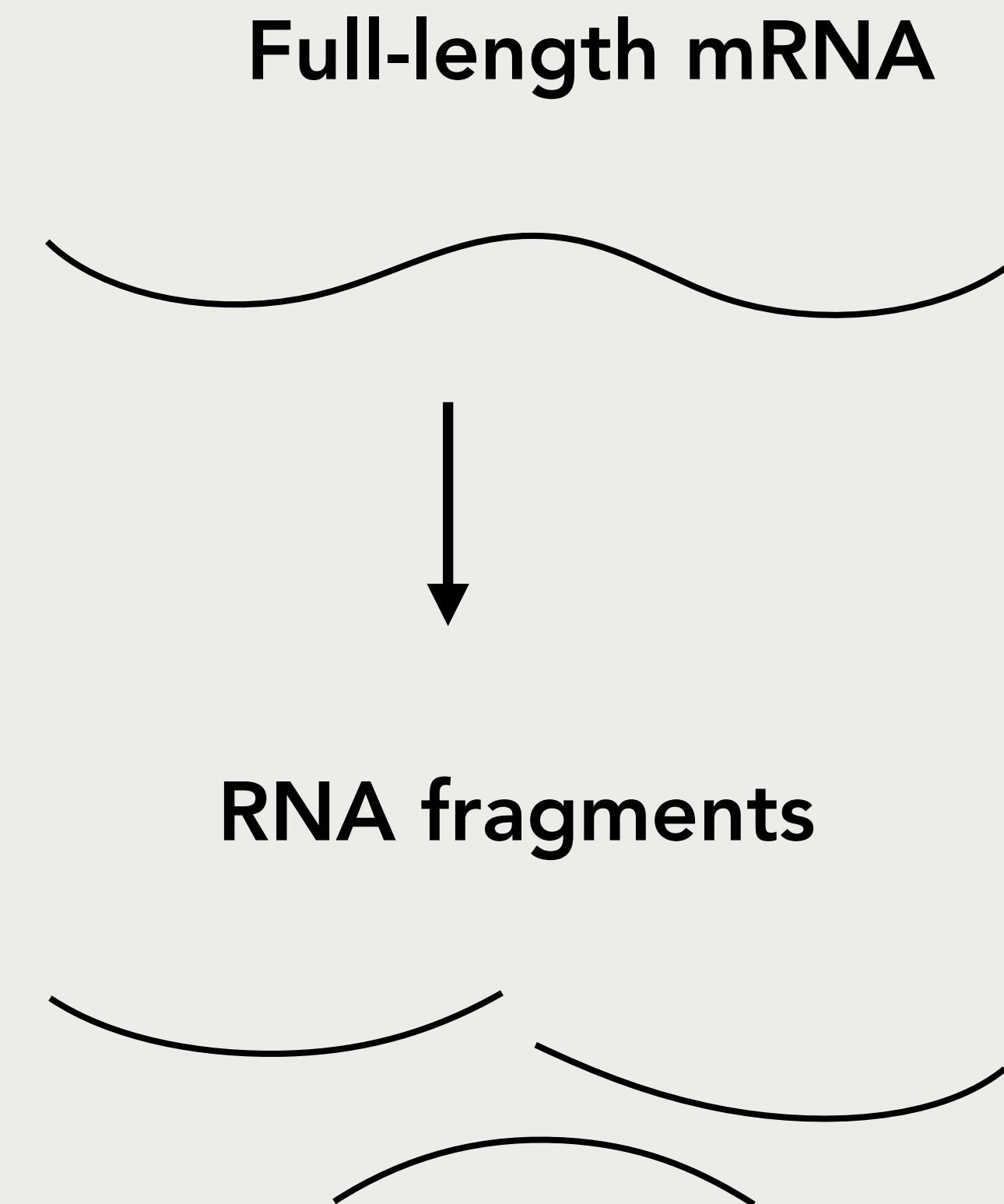
3. First strand cDNA synthesis



3. First strand cDNA synthesis

RNA to dsDNA to stabilize the template to be sequenced

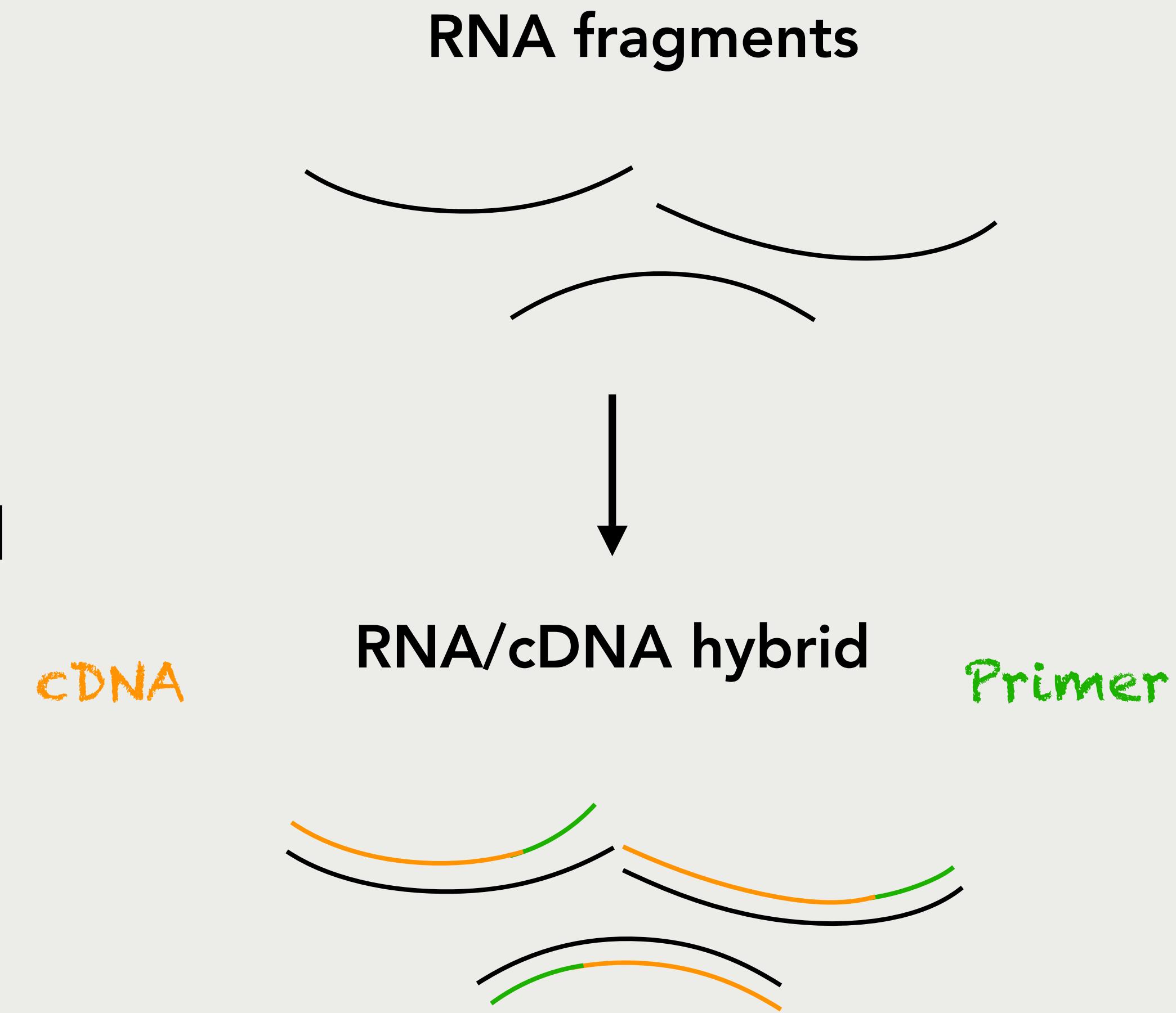
- RNA fragments remain labile (readily break down or change) because of the 2' OH and single strandedness
- Converting RNA to cDNA to dsDNA stabilizes and protects the fragments
- Two reactions:
 - ***First-strand synthesis*** (RNA/cDNA hybrid)
 - ***Second-strand synthesis*** (dsDNA)



3. First strand cDNA synthesis

RNA to dsDNA to stabilize the template to be sequenced

- Normally, DNA is **transcribed** to RNA (RNA polymerase)
 - Producing DNA from RNA: **reverse transcription (reverse transcriptase)**
- Like transcription, requires a primer and dNTPs (nucleic acid monomers)
- Result is one strand of RNA (template) and **complementary** strand of DNA



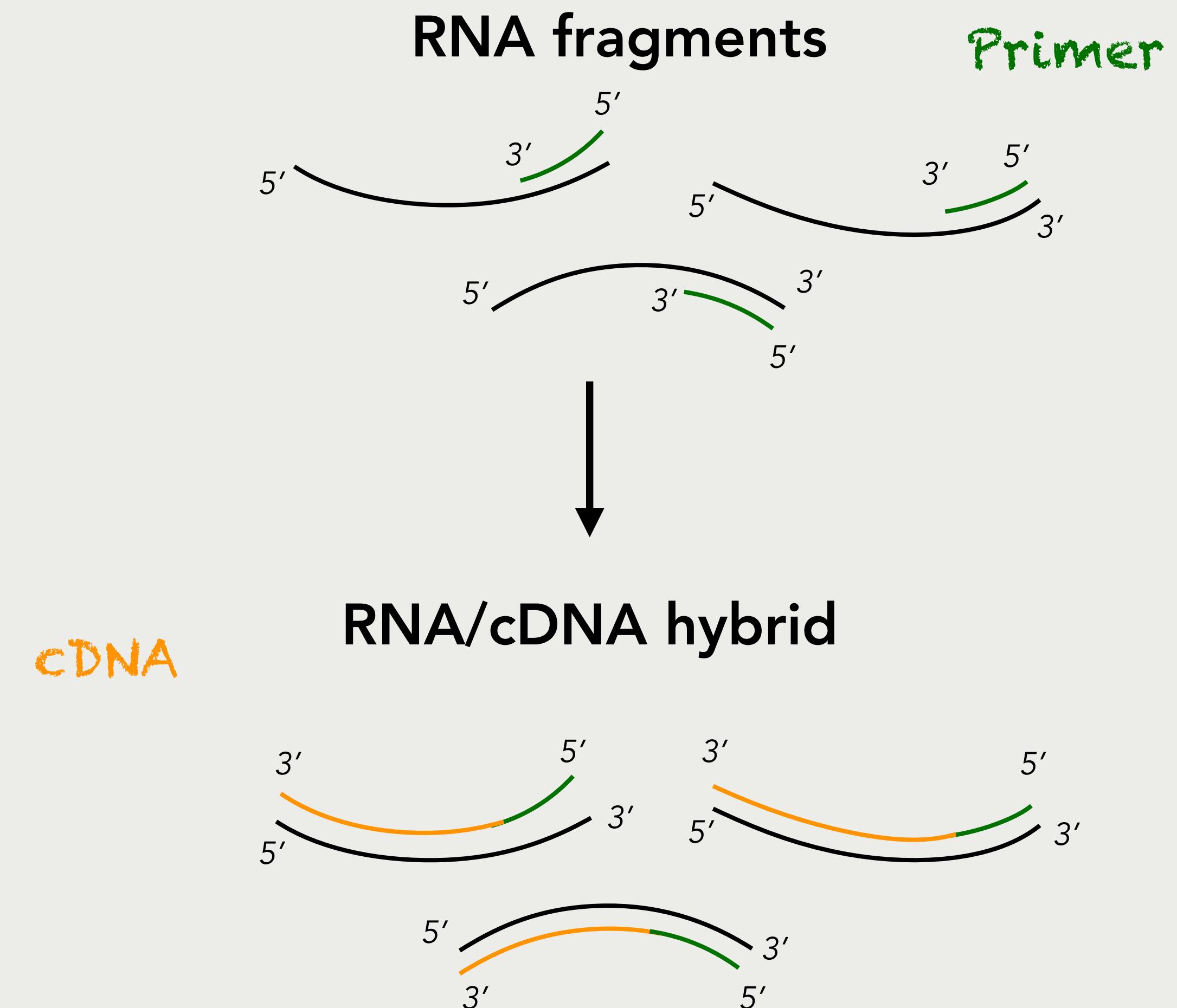
3. First strand cDNA synthesis

RNA to dsDNA to stabilize the template to be sequenced

NEBNext Ultra II Directional

1.2.36 - 1.3.4 (pages 7-8)

- RNA primed with *random primers* during fragmentation
- Buffer and RT is added
 1. 10 minutes at 25°C
 2. 15 minutes at 42°C
 3. 15 minutes at 70°C



3. First strand cDNA synthesis

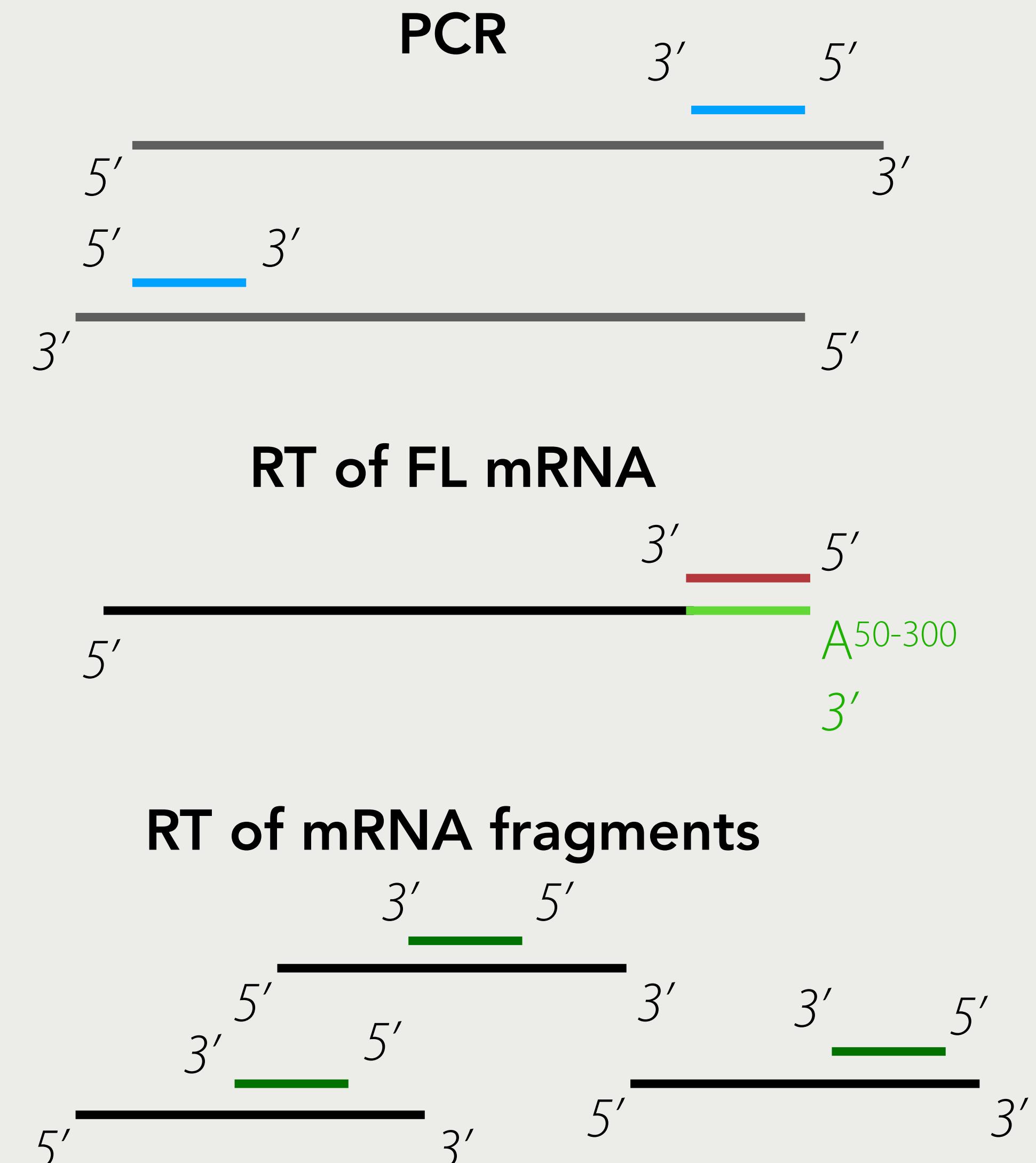
A bit about priming for RT reactions

- All nucleic acid polymerases (DNA polymerases, RNA polymerases, reverse transcriptase) require **priming**
- A short stretch of **duplex** (two complementary strands) with a free 3' OH to extend
- Types of primers:

Gene specific primers (i.e., in PCR)

Oligo(dT) (i.e., in RT reactions of FL mRNA)

Random (i.e., in RT reactions of RNA fragments)



3. First strand cDNA synthesis

A bit about priming for RT reactions

- All nucleic acid polymerases (DNA polymerases, RNA polymerases, reverse transcriptase) require **priming**
 - A short stretch of **duplex** (two complementary strands) with a free 3' OH to extend
- Types of primers:

Gene specific primers (i.e., in PCR)

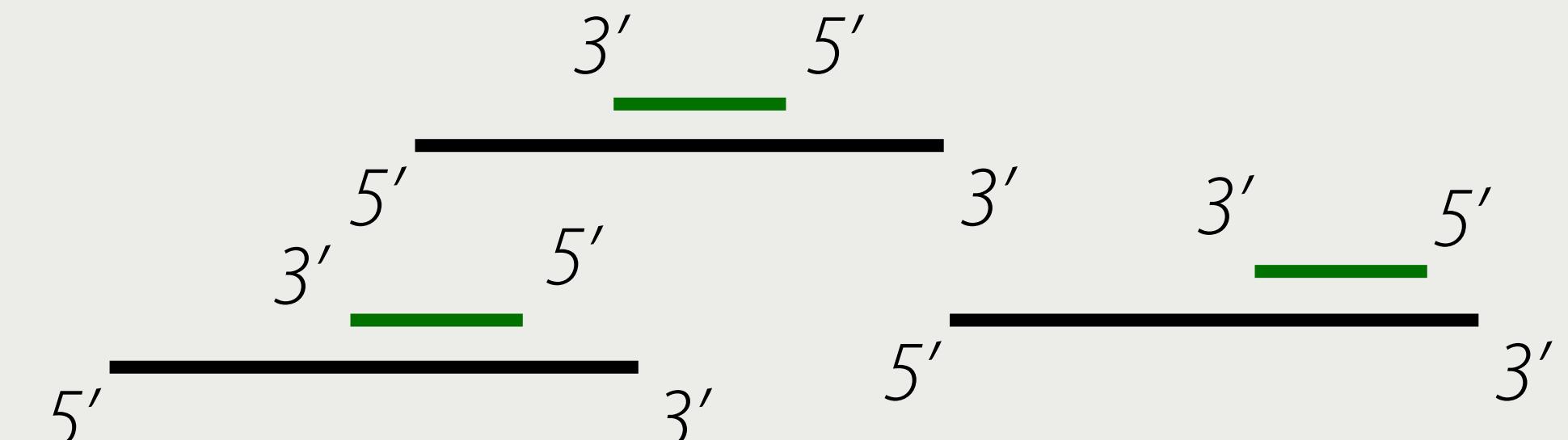
Oligo(dT) (i.e., in RT reactions of FL mRNA)

Random (i.e., in RT reactions of RNA fragments)

Random primers will bind to their reverse complement...**anywhere** in the 200 nt fragment.

Thus, the cDNA size distribution will be on average shorter than the fragmented RNA size distribution.

RT of mRNA fragments



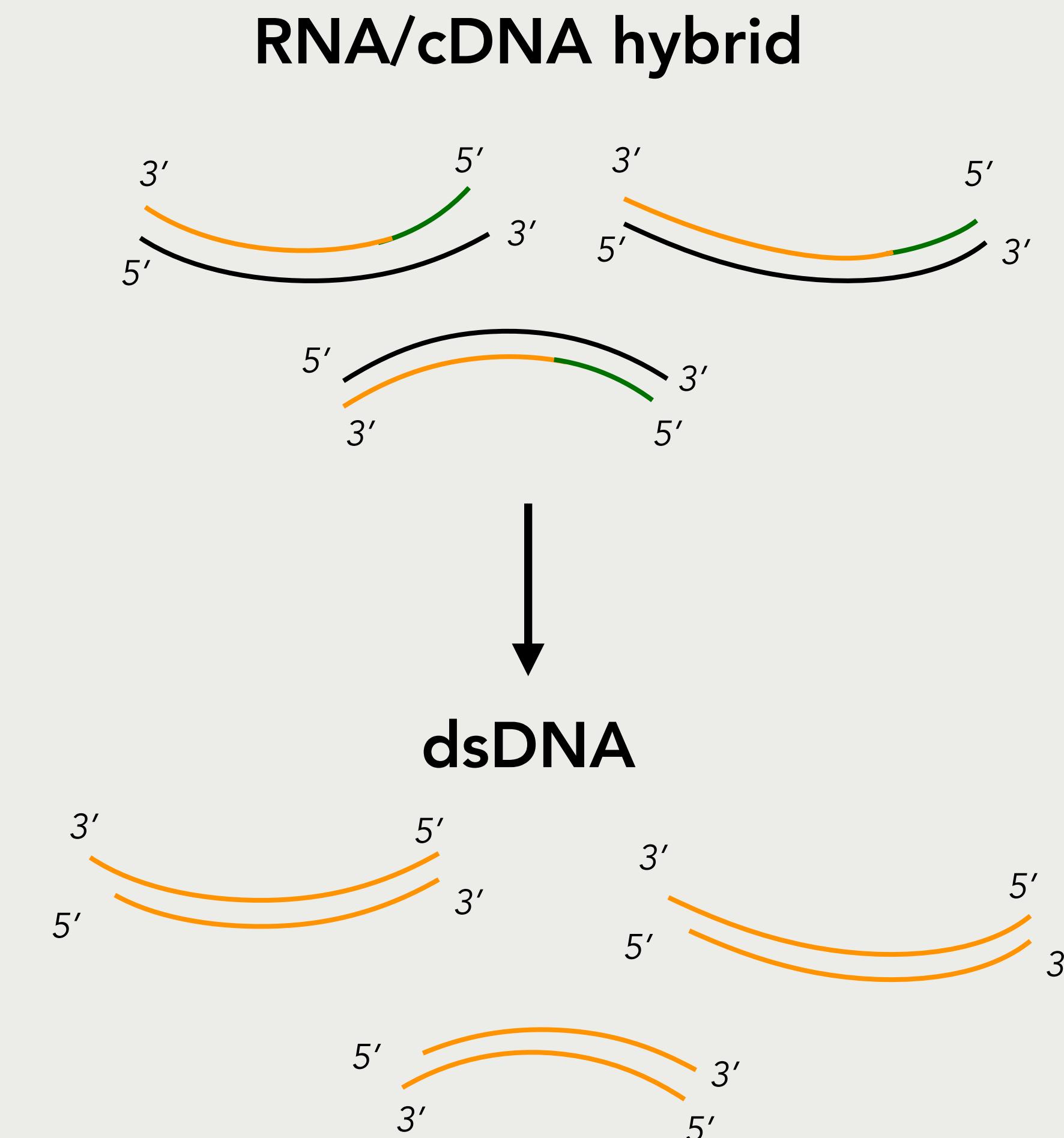
4. Second strand cDNA synthesis

RNA to dsDNA to stabilize the template to be sequenced

NEBNext Ultra II Directional

1.4 (page 8)

- Three enzymes included in the enzyme mix:
 - RNase H
 - DNA polymerase
 - DNA ligase
1. Incubate for 1 hour at 16°C



Primers from first-strand synthesis not shown for simplicity

4. Second strand cDNA synthesis

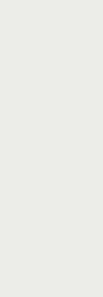
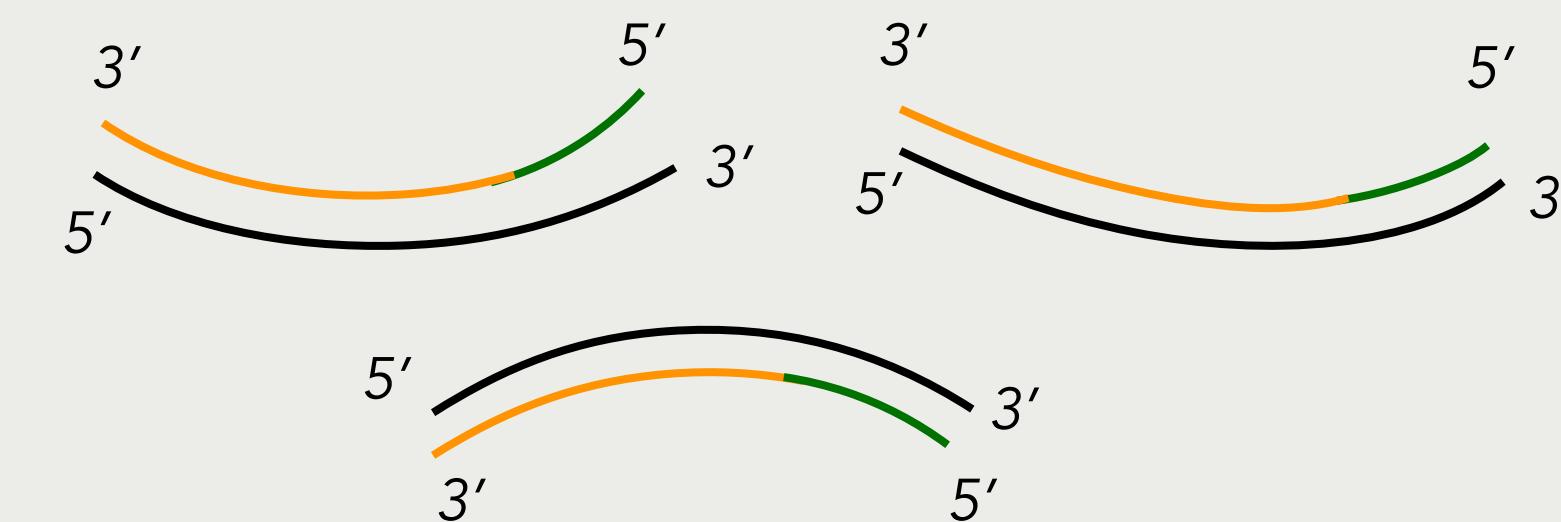
RNA to dsDNA to stabilize the template to be sequenced

NEBNext Ultra II Directional

1.4 (page 8)

- Three enzymes included in the enzyme mix:
 - RNase H
 - DNA polymerase
 - DNA ligase
1. Incubate for 1 hour at 16°C

RNA/cDNA hybrid



dsDNA

5. End repair and dA-tailing

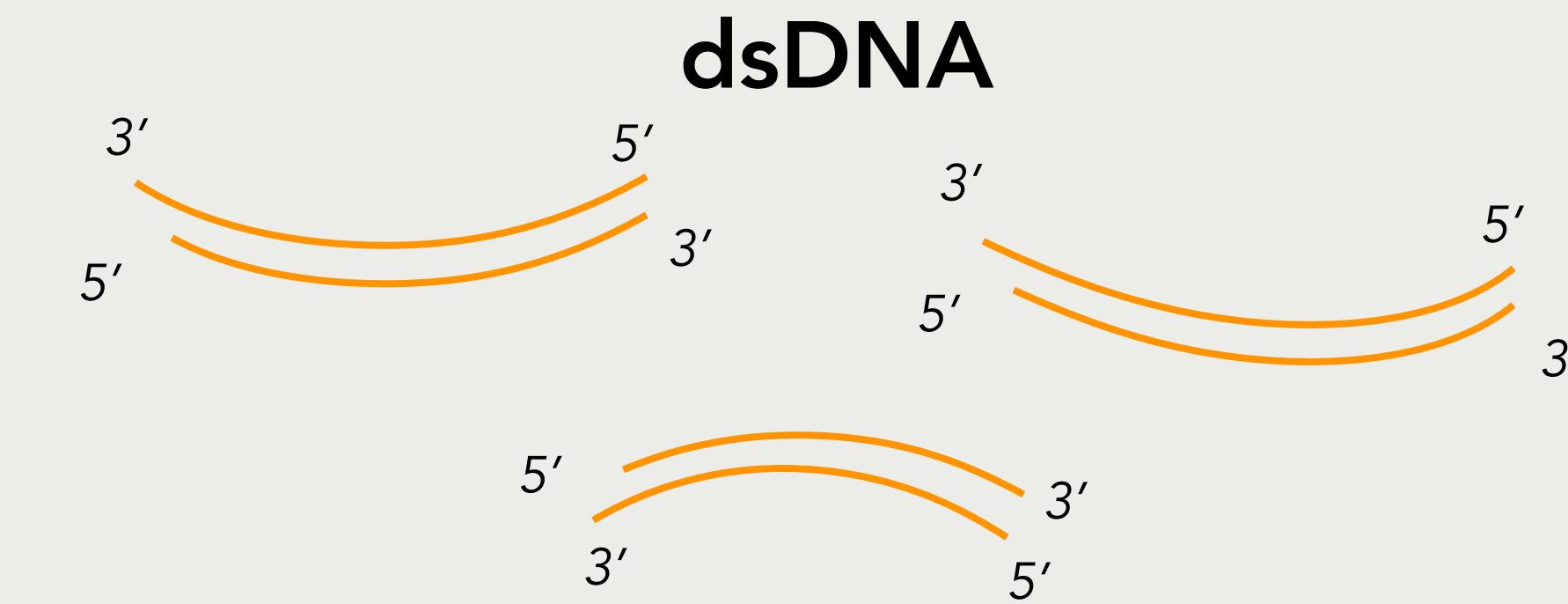
Fix DNA ends and add some glue...

NEBNext Ultra II Directional

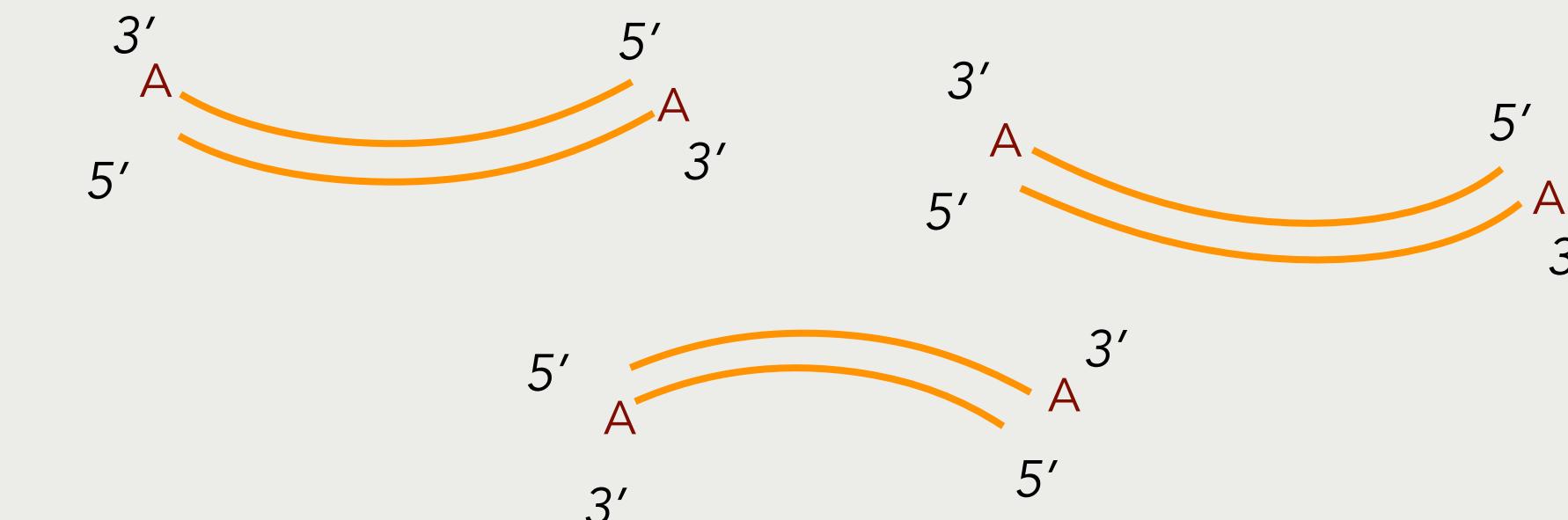
1.6 (page 9)

- dsDNA has 3' overhangs
- Make blunt and add dA-tails

1. Incubate for 30 minutes at 20°C
2. Incubate for 30 minutes at 65°C



Blunt dsDNA + tails



Tophat Questions

Join code: 684418

6. Adaptor ligation and U-excision

Prepare for Illumina-chemistry and/or barcodes, UMIs, etc.

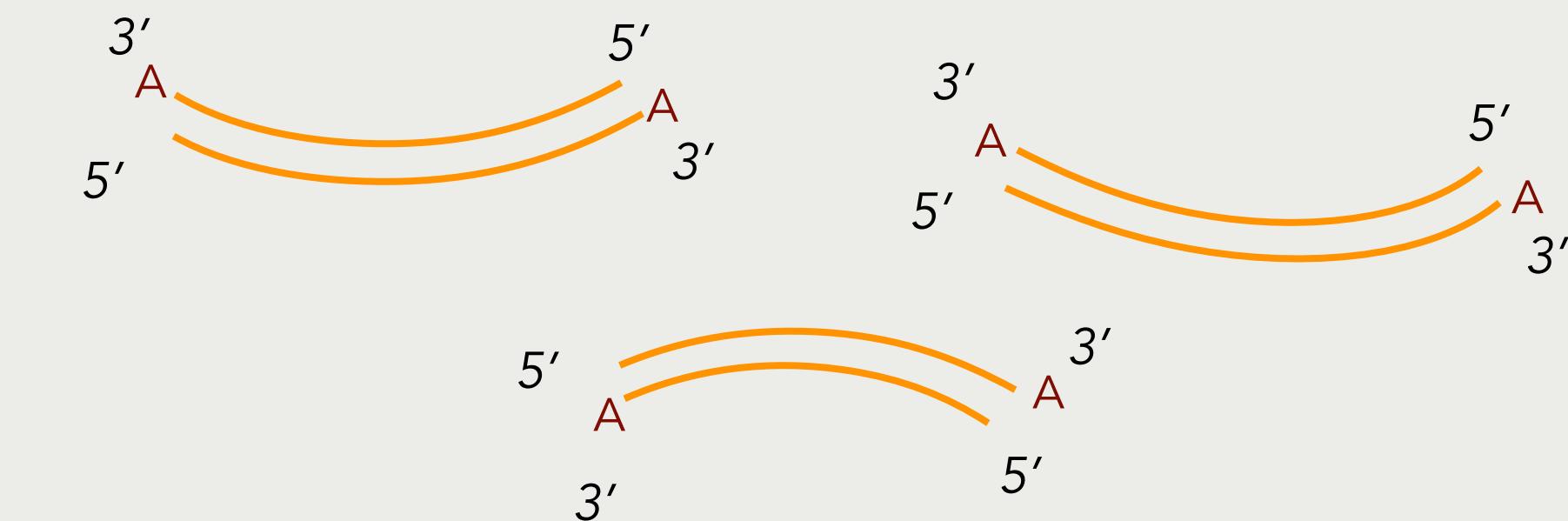
NEBNext Ultra II Directional

1.7 (pages 9-10)

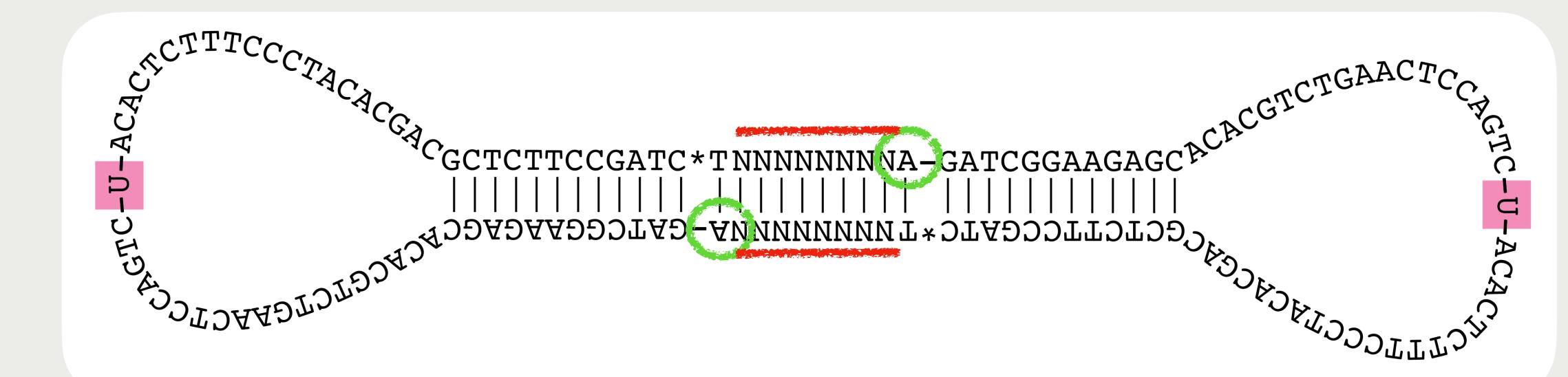
- **Adapters** adapt...allow for DNA to connect to other pieces of DNA
 - Known sequence of short dsDNA
 - NEBNext - hairpin-loop with T overhangs

1. Incubate for 15 minutes at 20°C (ligation)
 2. Add USER Enzyme
 3. Incubate for 15 minutes at 37°C (U-excision)

Blunt dsDNA + tails



dsDNA with circular adapters



A's from dA-tailing

dsDNA “insert”

6. Adaptor ligation and U-excise

Prepare for Illumina-chemistry and/or barcodes, UMIs, etc.

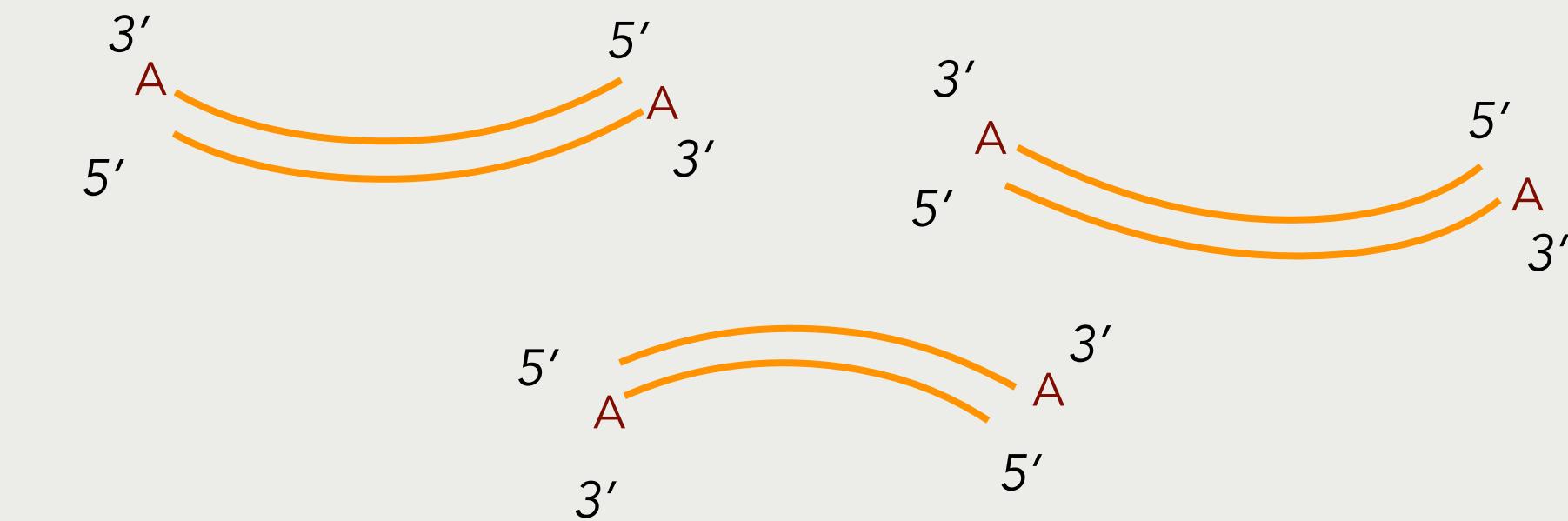
NEBNext Ultra II Directional

1.7 (pages 9-10)

- **Adaptors** adapt...allow for DNA to connect to other pieces of DNA
- Known sequence of short dsDNA
 - NEBNext - hairpin-loop with T overhangs

1. Incubate for 15 minutes at 20°C (ligation)
2. Add USER Enzyme
3. Incubate for 15 minutes at 37°C (U-excision)

Blunt dsDNA + tails



dsDNA with circular adapters



6. Adaptor ligation and U-excise

Prepare for Illumina-chemistry and/or barcodes, UMIs, etc.

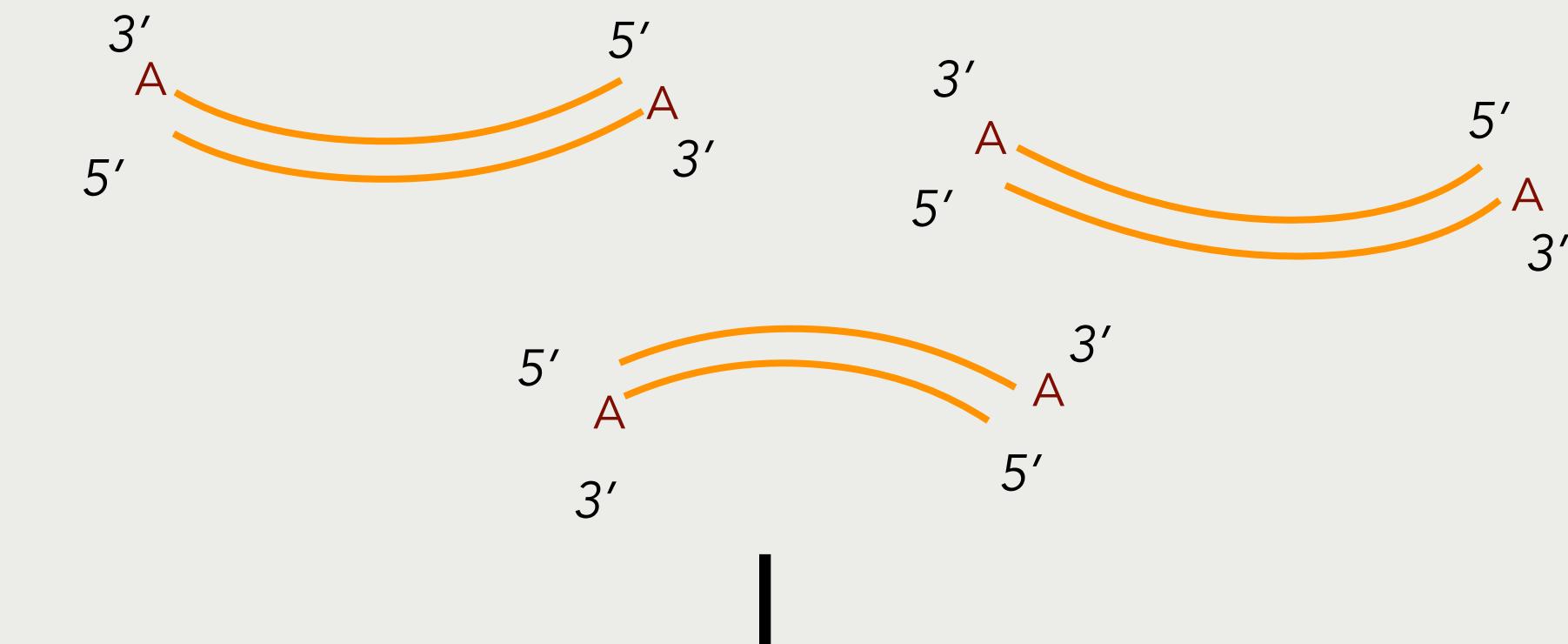
NEBNext Ultra II Directional

1.7 (pages 9-10)

- **Adaptors** adapt...allow for DNA to connect to other pieces of DNA
- Known sequence of short dsDNA
 - NEBNext - hairpin-loop with T overhangs

1. Incubate for 15 minutes at 20°C (ligation)
2. Add USER Enzyme
3. Incubate for 15 minutes at 37°C (U-excision)

Blunt dsDNA + tails



dsDNA with circular adapters

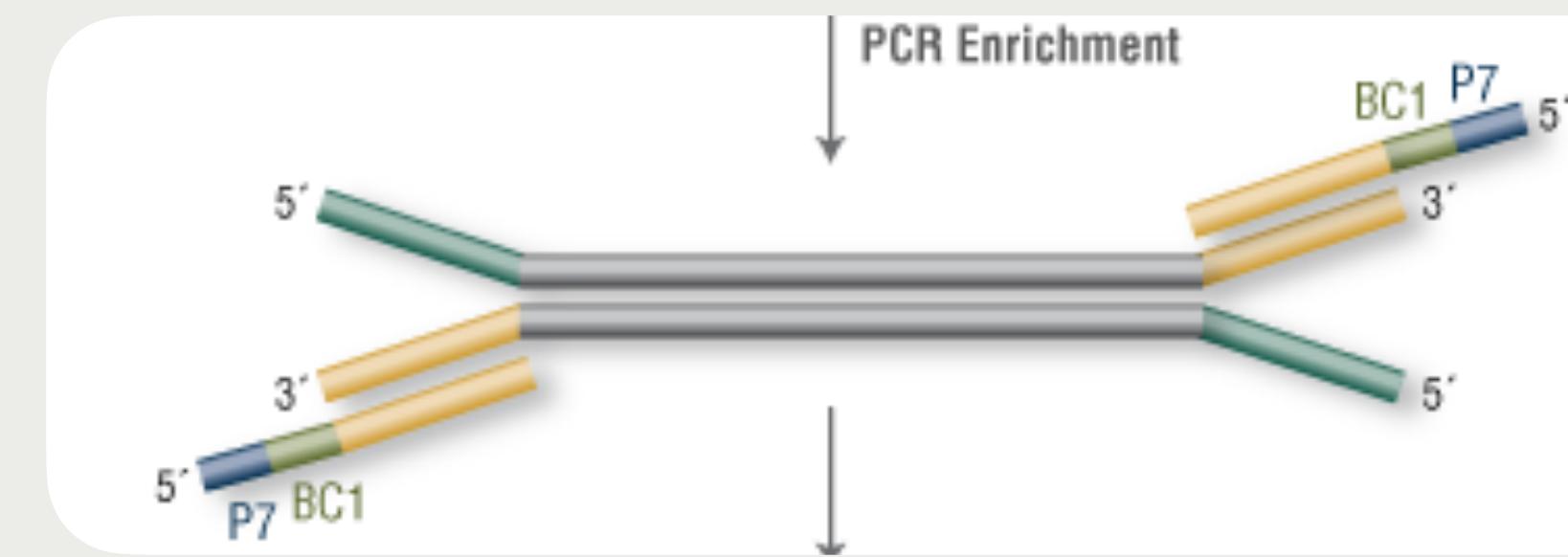


6. Adaptor ligation and U-excision

What's the point of an adaptor?

- Illumina-compatible libraries must have specific primers called P5/P7
- Adaptors allow for inclusion of other small DNA sequences
 - **Unique molecular identifiers (UMIs)** - allows distinguishing between unique RNA fragments
 - **Barcodes** - allows distinguishing between unique sequencing libraries
- UMIs/barcodes can be easily added because the adaptor sequence is known, allowing design of sequences that hybridize to the adaptor

Only adding barcodes (BC1)



Adding barcodes and UMIs)



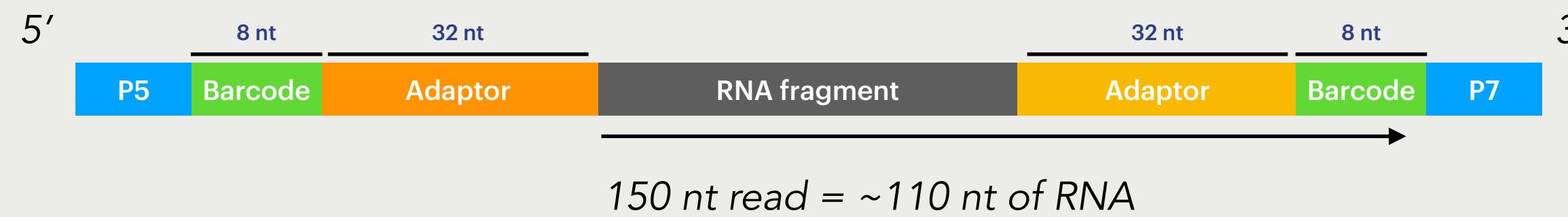
99.9% of sequencing libraries will have a barcode

Some (<10%) will have UMIs

6. Adaptor ligation and U-excision

Hold on, why do these details matter to a bioinformatician?

Consider the following scenario..



Read mapping (done by you)

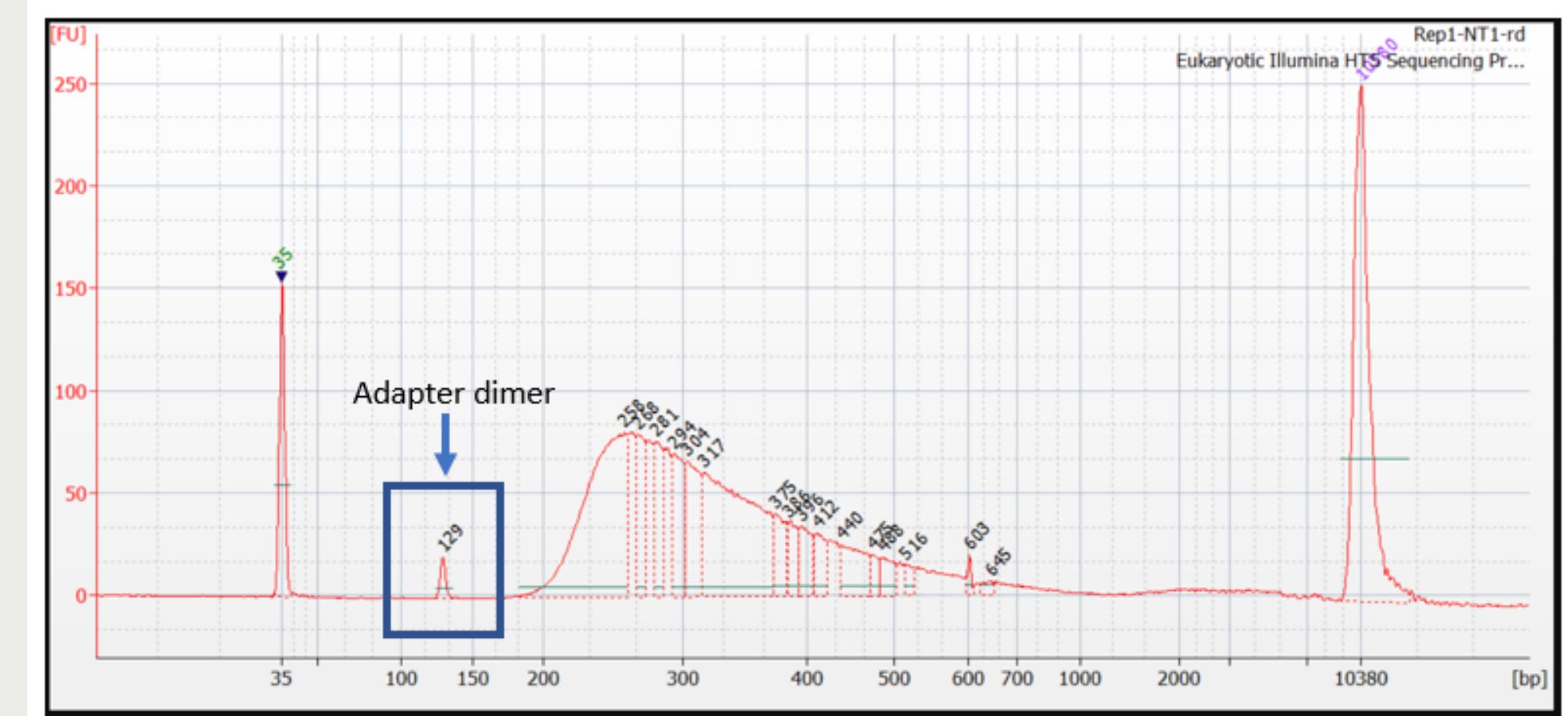


Do you ignore the adaptor + barcode? Do you remove them? Do you hope they don't have a complementary sequence in your reference genome?

7. Purification

Beads to clean your newly synthesized DNA

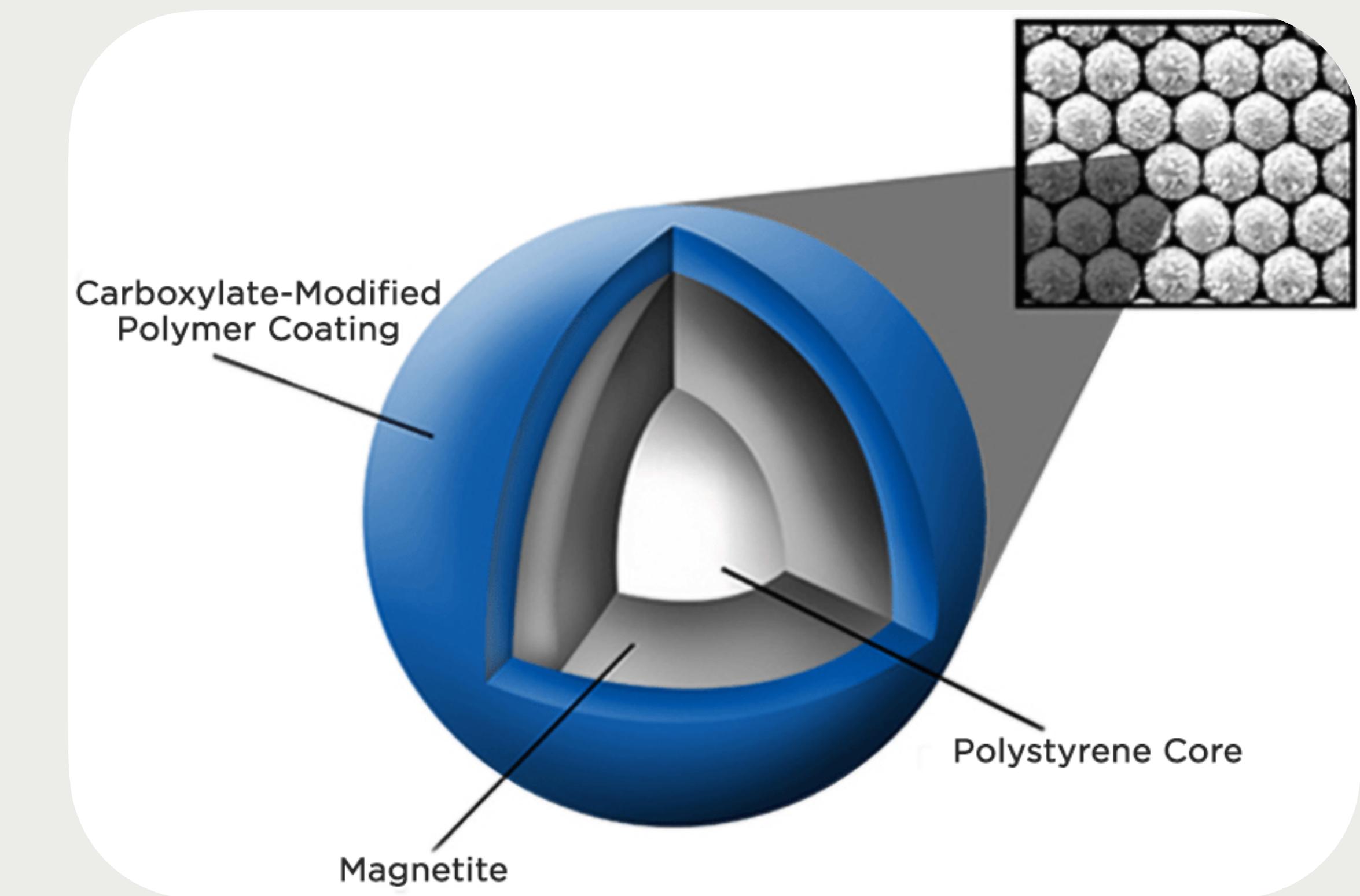
- When adding DNA, ***polymerization*** is best...
- Sometimes ***ligation*** is necessary, when adding completely new DNA sequence
 - Problem: ligation will connect any two DNA fragments with blunt-ish ends
- Adaptor dimers are recurrent contaminants in sequencing libraries
 - If not removed from the library, a significant number of reads will be from the dimer rather than from the RNA fragment



7. Purification

Beads to clean your newly synthesized DNA

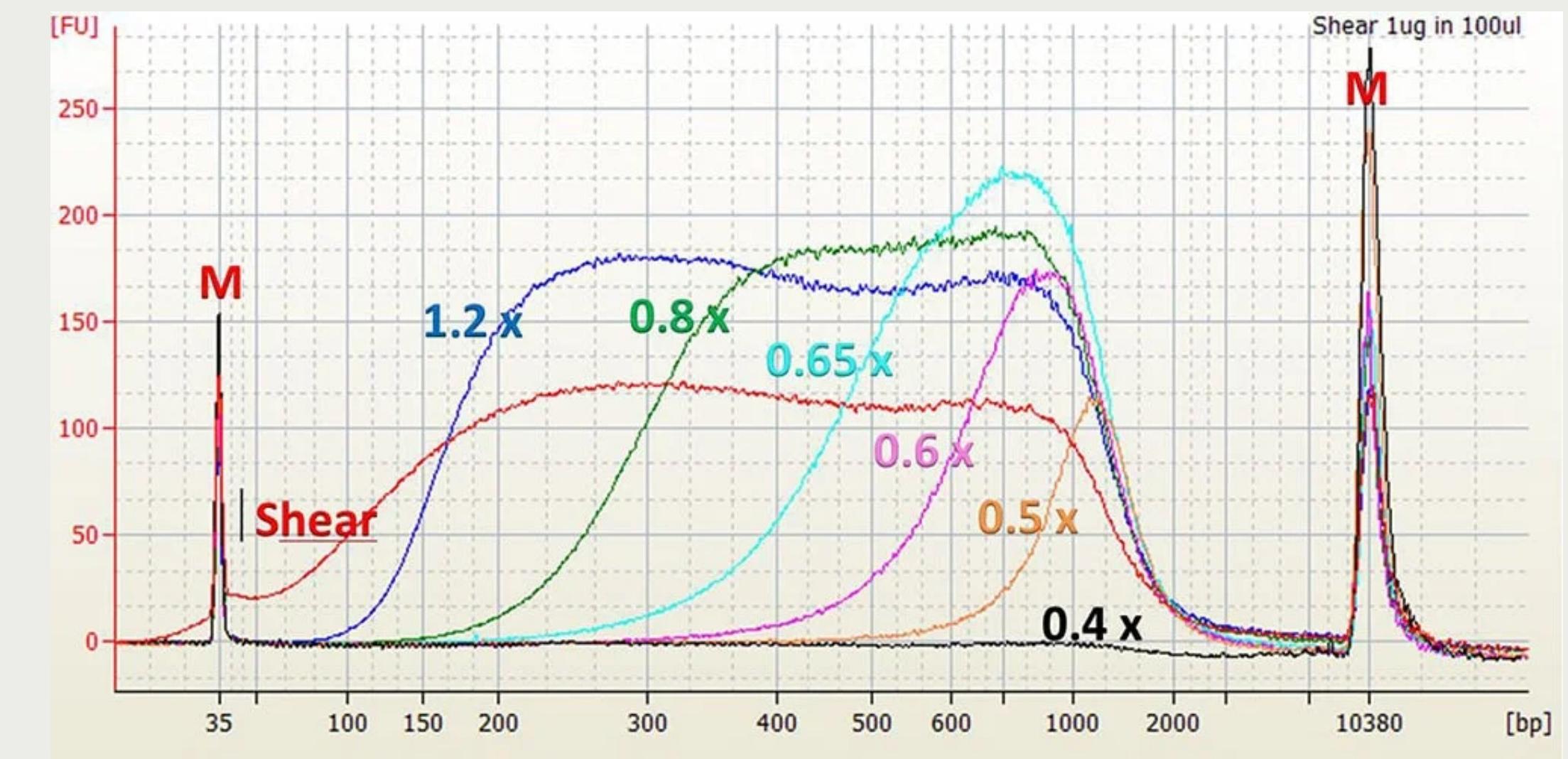
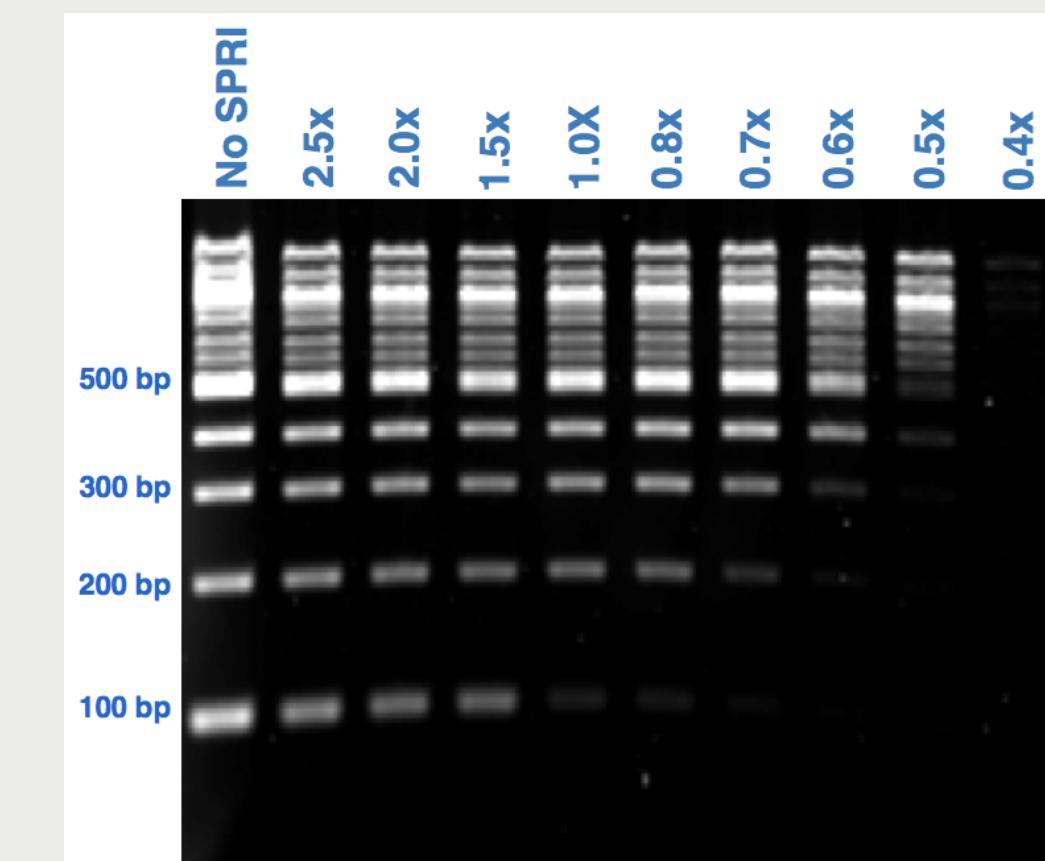
- Ways to prevent adaptor dimer contamination:
 - Clever adaptor design (hairpin loop)
 - Don't use blunt end (i.e., A-tail)
 - Clean up after ligation
- SPRI beads for cleanup and size selection
 - First used in 1.5 (page 8) at 1.8X
 - Next used 1.8 (page 10) at 0.9X
- Crowding agent to force longer DNA to the beads first



7. Purification

Beads to clean your newly synthesized DNA

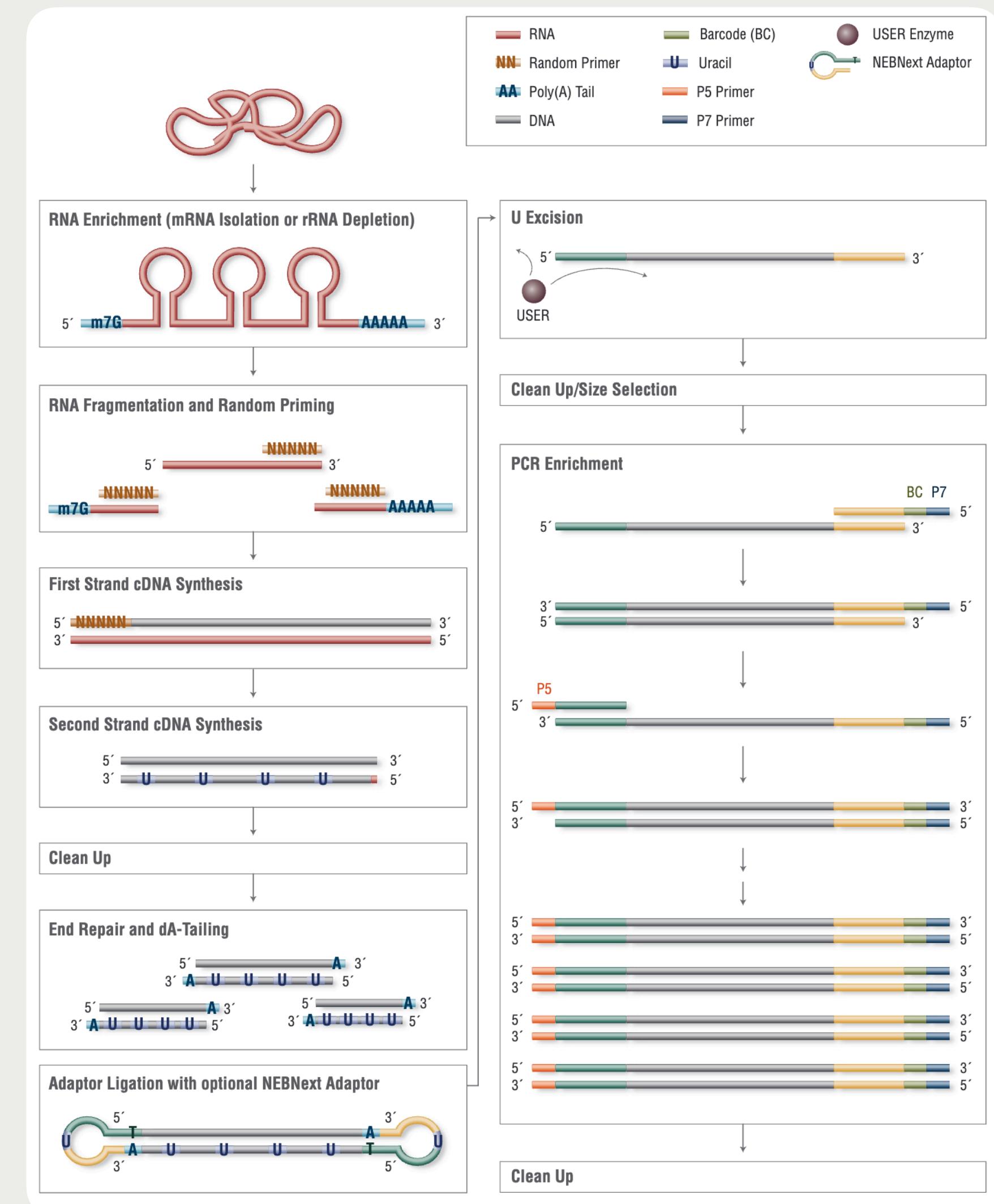
- Ways to prevent adaptor dimer contamination:
 - Clever adaptor design (hairpin loop)
 - Don't use blunt end (i.e., A-tail)
 - Clean up after ligation
- SPRI beads for cleanup and size selection
 - First used in 1.5 (page 8) at 1.8X
 - Next used 1.8 (page 10) at 0.9X
- Crowding agent to force longer DNA to the beads first



8. PCR enrichment and indexing

Amplifying dsDNA fragments to increase the amount of material

- Recap:
 - We now have single-stranded DNA that is *complementary* to the original mRNA fragment
 - Each end of the DNA has adaptor
 - Need to add barcodes (indices), UMIs (if applicable) and P5/P7 sequencing primers
 - PCR copies DNA - exponential amplification



8. PCR enrichment and indexing

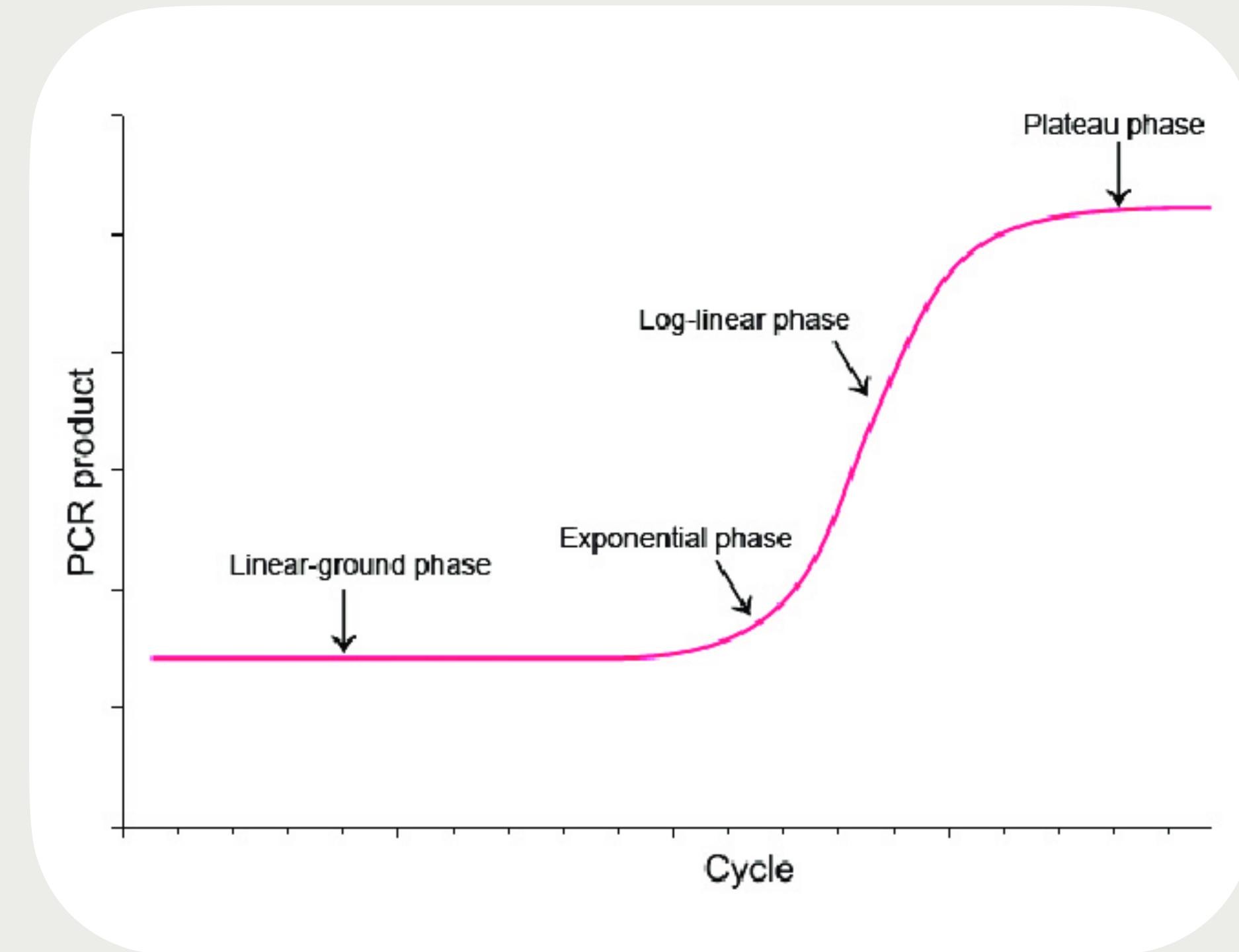
Amplifying dsDNA fragments to increase the amount of material

- Reminder:
 1. Melt/denature (get linear, single template strands)
 2. Anneal (primers anneal to template)
 - Primers are complementary to adapter and have index + P5/P7
 3. Extend (polymerase extends primers)
- Many cycles of #1-3
 - How many?

8. PCR enrichment and indexing

Amplifying dsDNA fragments to increase the amount of material

- Reminder:
 1. Melt/denature (get linear, single template strands)
 2. Anneal (primers anneal to template)
 - Primers are complementary to adapter and have index + P5/P7
 3. Extend (polymerase extends primers)
- Many cycles of #1-3
 - How many?
 - 8-16, depending on amount of starting RNA (Table 1.9.3B, page 11)



When quantitation matters, don't over-amplify.

Can't compare the number of fragments when you've reached plateau

9. Cleanup and QC

SPRI beads and bioanalyzer

NEBNext Ultra II Directional

- 1.11 (page 12)
- Clean with 0.9X SPRI beads (remove primers)
- Check for broad library size with peak near...
 - Adaptor dimers (120 - 150 nt)
 - Primers (~80 nt)

Tophat Questions

Join code: 684418

9. Cleanup and QC

SPRI beads and bioanalyzer

NEBNext Ultra II Directional

- 1.11 (page 12)
- Clean with 0.9X SPRI beads (remove primers)
- Check for broad library size with peak near...

200 nt insert +
(2 x 32 bp adaptor) +
8 bp barcode +
20 bp P5 +
24 bp P7 = 316 nt

