Getting started with: Next-generation sequencing applications

Morgan A. Sammons, PhD

Assistant Professor of Biology

Life Sciences 2075

masammons@albany.edu

Guidelines ≠ Guarantees

Goal: Discuss best practices and tips for getting started with next-generation sequencing

Caveat: There is no one-size-fits-all approach for all experiments.

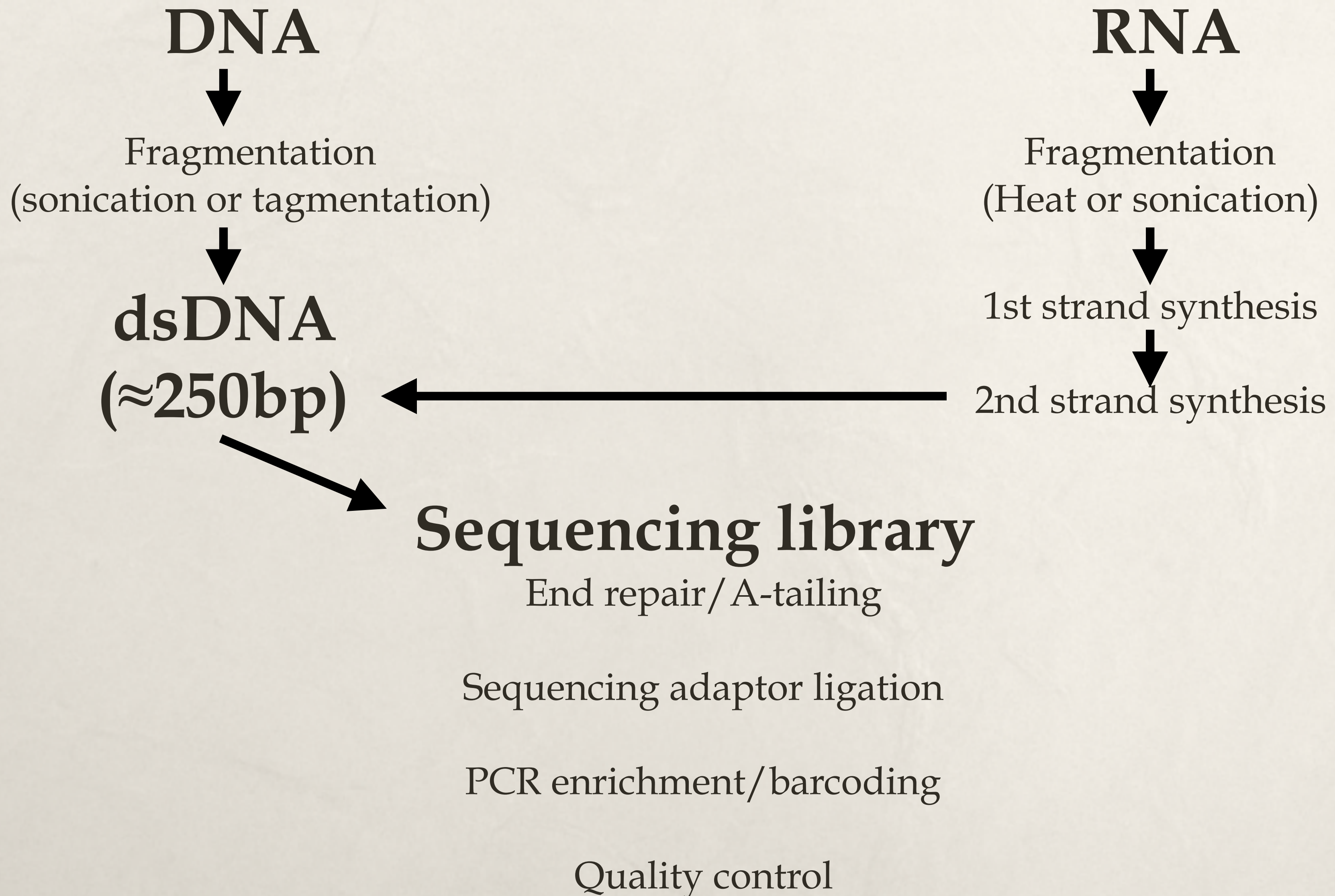
If you can isolate a nucleic acid,
you can *sequence* it

Garbage in, Garbage out

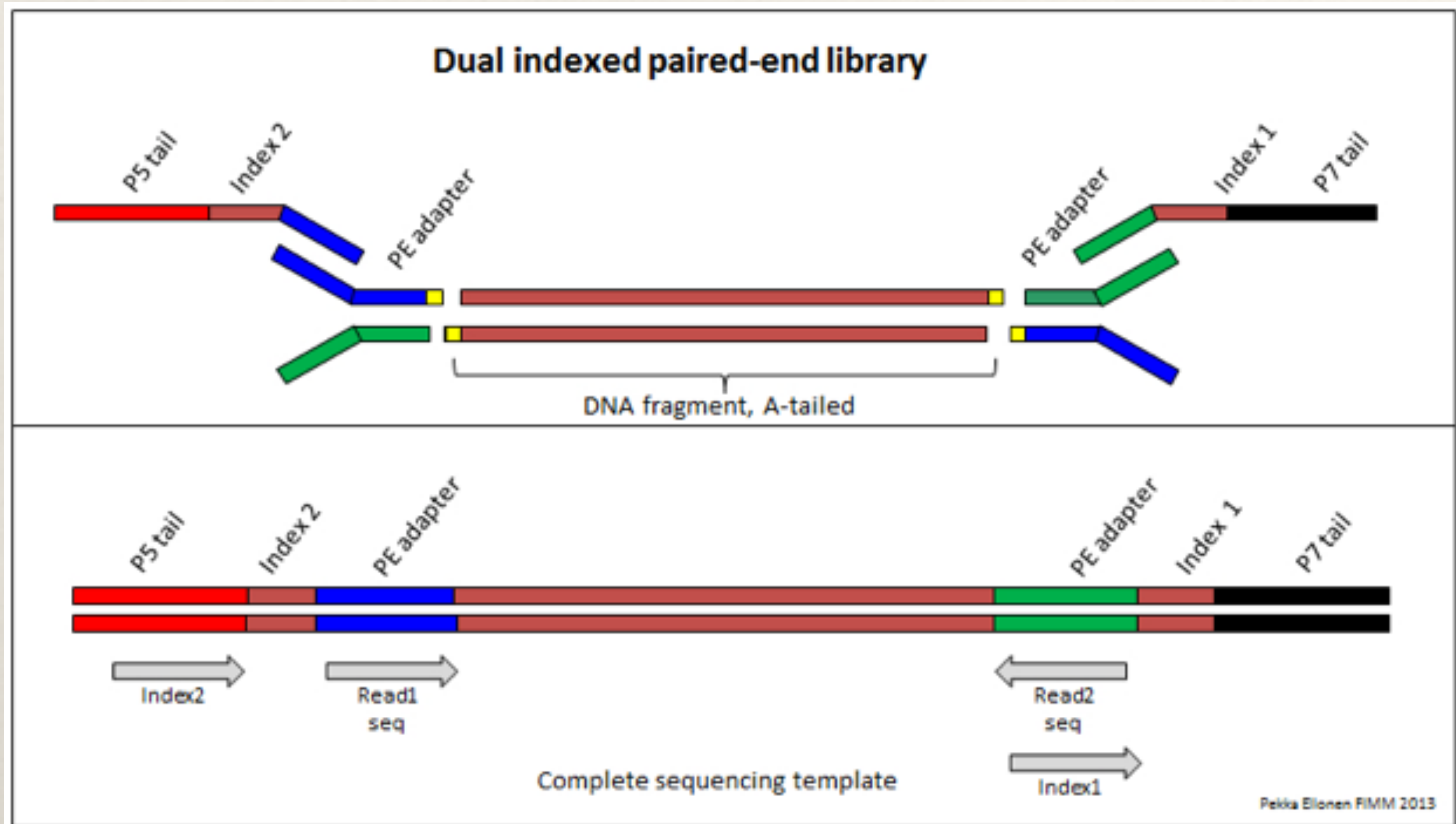
Remember: Good starting material will more likely result in “good” data.

Remember: Data that do not fit expected results may reflect less-than-optimal starting material.

Workflow for sequencing



I. What is a sequencing library?



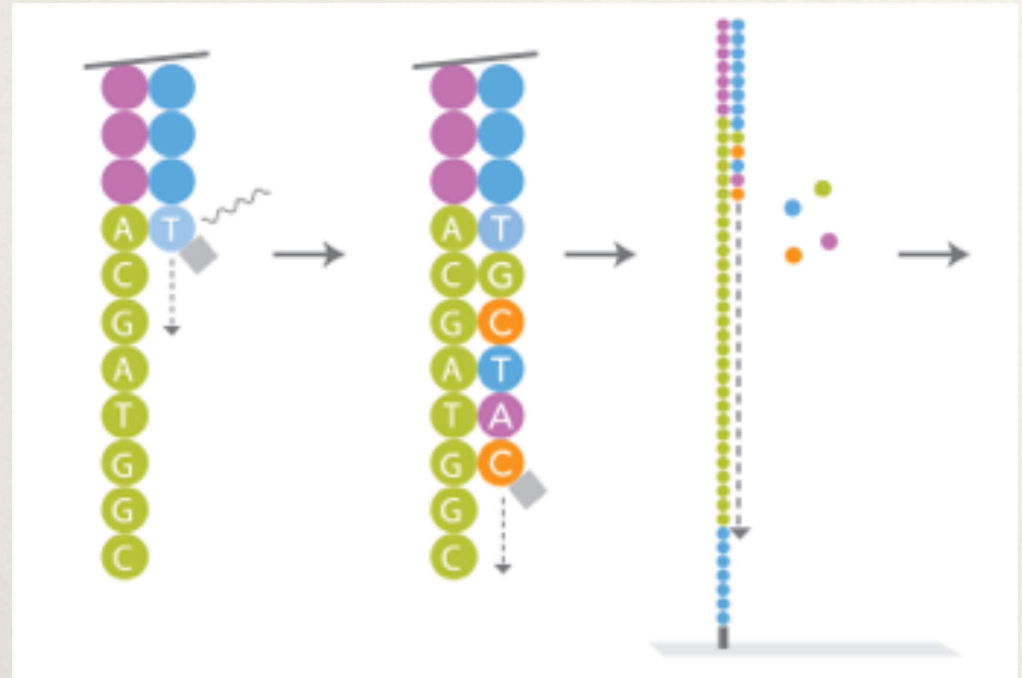
Sequencing library components

5′ and 3′ ends compatible with the flowcell/ sequencing instrument

Primer binding sites to sequence your sample (from both ends)

Primer binding sites to sequence index/barcode (single or double)

Your DNA of interest



Making sequencing libraries for:

RNA-Seq

Small RNA-Seq

ChIP-Seq

DNA-seq

RNA-Seq - Advantages

RNA isolation is straightforward

Low sample requirements (as low as 10pg…1 cell)

Unbiased view of the transcriptome (no prior knowledge)

Robust data analysis/statistical pipelines available

Mature technology

RNA-Seq - Disadvantages

"Relatively" expensive

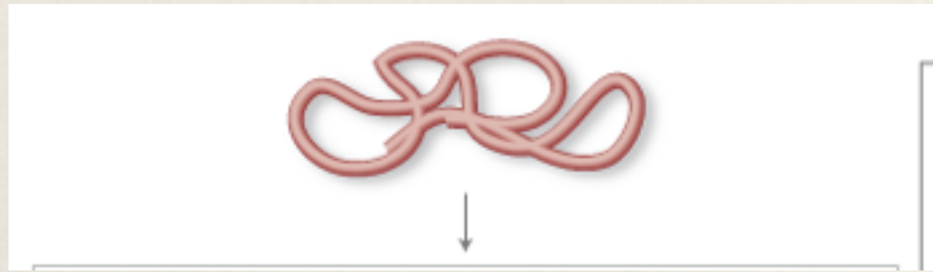
High knowledge barrier to entry (many many many tools/ software packages

EVERYTHING is observed (no more willful ignorance)

Normalization is not easy (Target/GAPDH doesn't cut it)

Practical Considerations for *RNAseq*

Starting RNA pool



RNA can come from almost unlimited number of sources.

RNA quality is critical.

Practical Considerations for *RNAseq*

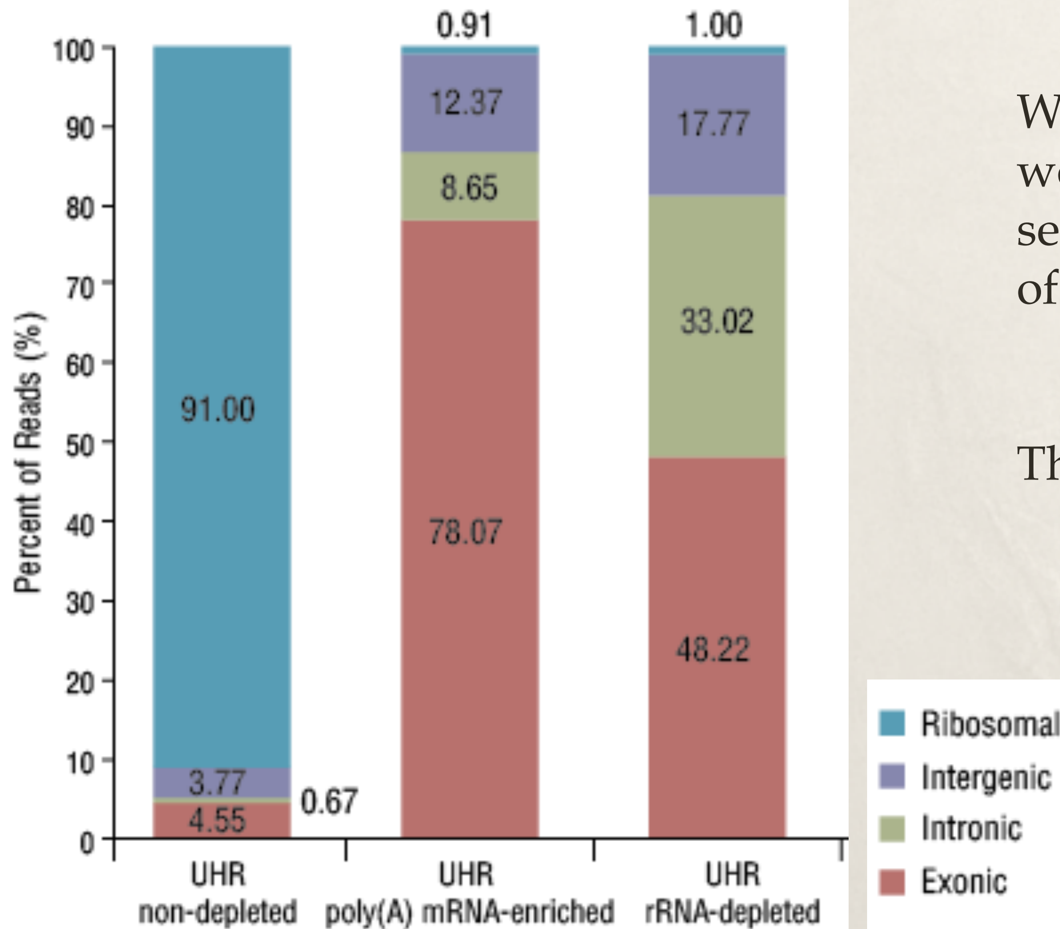
rRNA depletion
More complex transcriptome
lncRNA, miRNA, tRNA, eRNA…

Expensive (>$50/sample)

PolyA isolation
Less complex transcriptome
Only mRNA (-/+ a few things)

Cheap (≈$3/sample)

Why enrich for non-rRNA transcriptome?

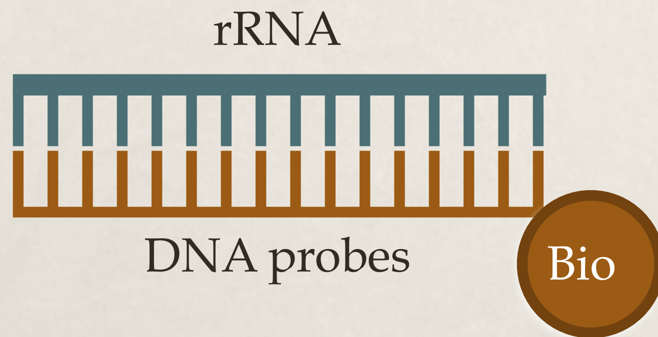


Without rRNA depletion, you would need 15-20X more sequencing to get the same # of desired reads.

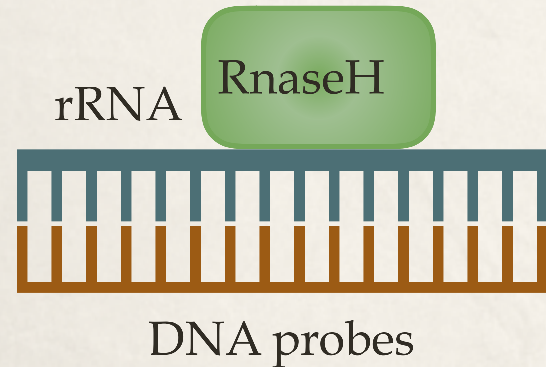
That means 15-20X more $$$

rRNA depletion

Affinity-mediated depletion



RNAse-mediated degradation



5′ phosphate-dependent exonuclease

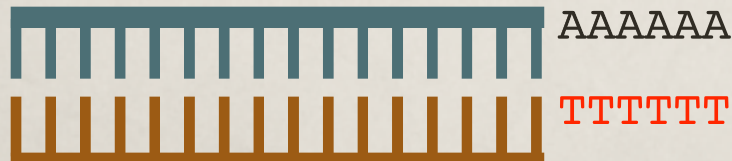
PolyA Isolation Strategies for RNA-seq

Isolation of polyA with anchored poly dT beads

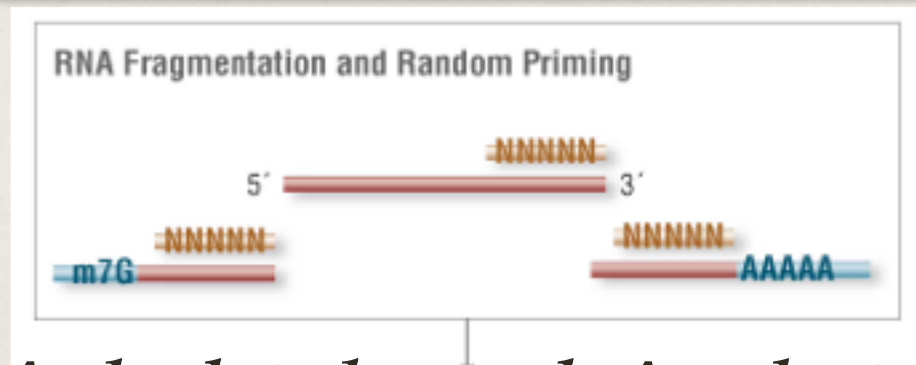
Most Vendors (*Illumina, NEB, etc*)

Full-length, 1st strand synthesis from a total RNA sample

Clontech SMRT-seq kits



Fragmentation and Priming for RNA-seq



rRNA-depleted or polyA-selected RNA

RNA fragmentation = heat + MgCl2

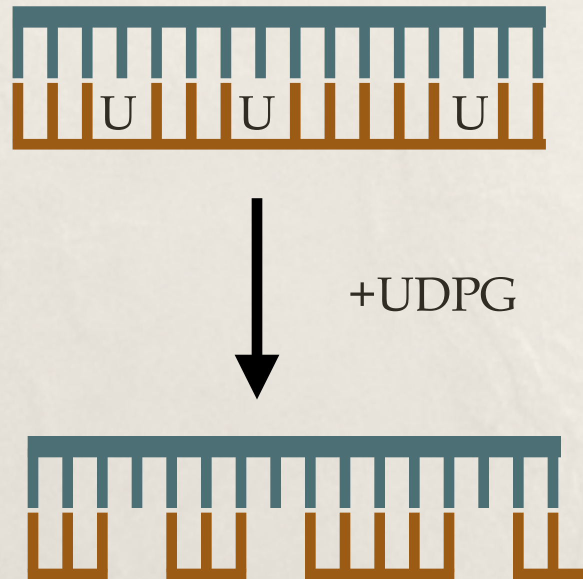
For Clontech-style kits, fragmentation happens after 2nd strand synthesis

Aiming for ≈200bp fragments
Remember, we're doing *short read* sequencing

1st strand synthesis with random hexamers/RT

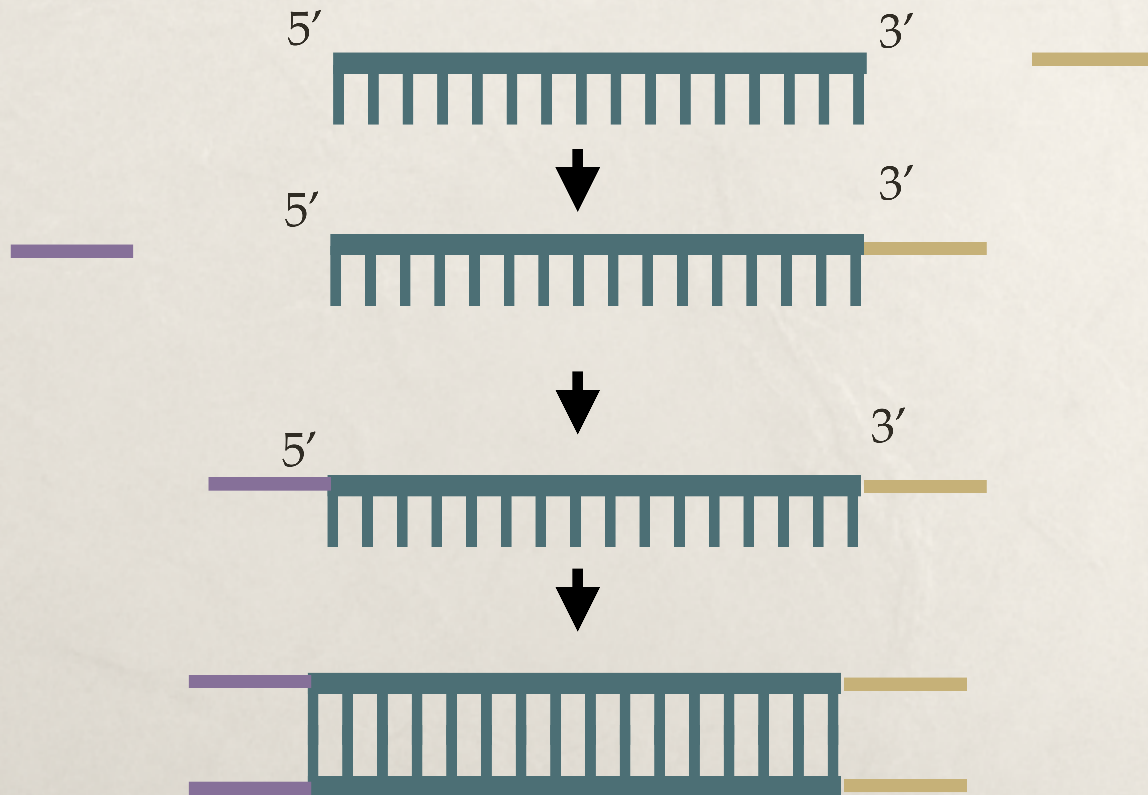
Defining RNA strandedness during 2nd strand synthesis

***dUTP* method allows determination of strandedness**

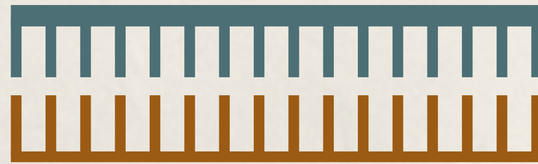


Defining RNA strandedness through ligation

*Adaptors are added to either **small RNA fragments** or to larger fragments for strandedness*



dsDNA can now be used to prepare a sequencing library

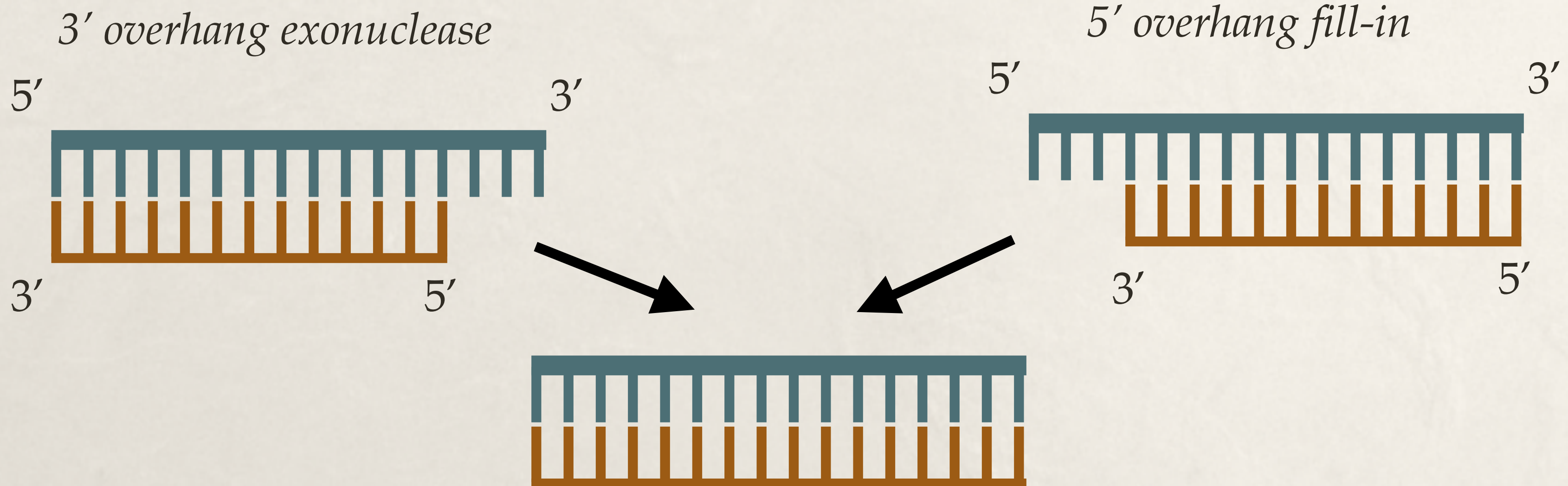


dsDNA can come from:

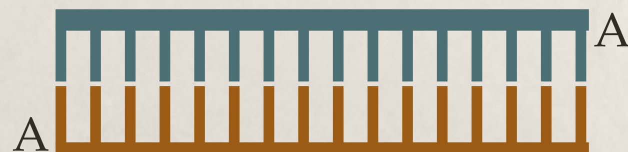
Isolated mRNA
Chromatin IP
Genomic DNA
literally anywhere….

End repair and A-tailing of dsDNA

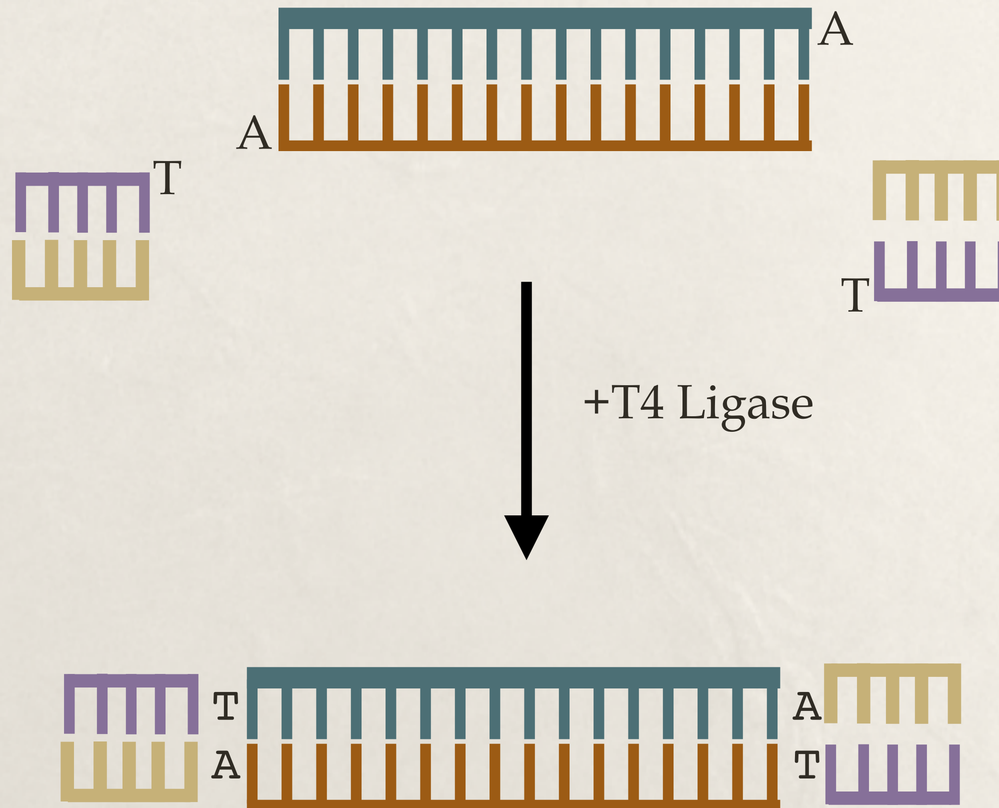
DNA is blunted with Klenow fragment



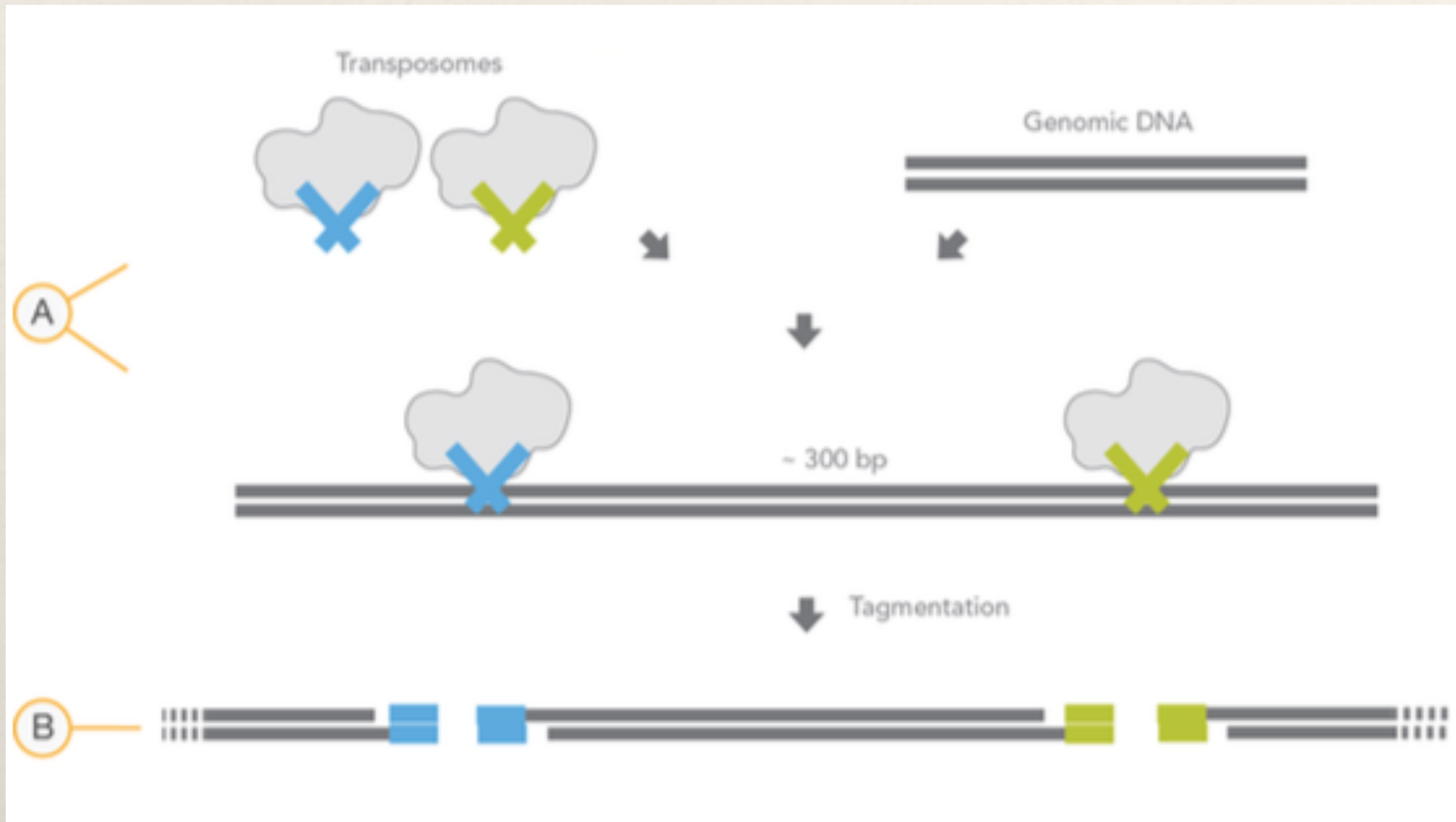
DNA is A-tailed with Taq



Ligation of sequencing adaptors



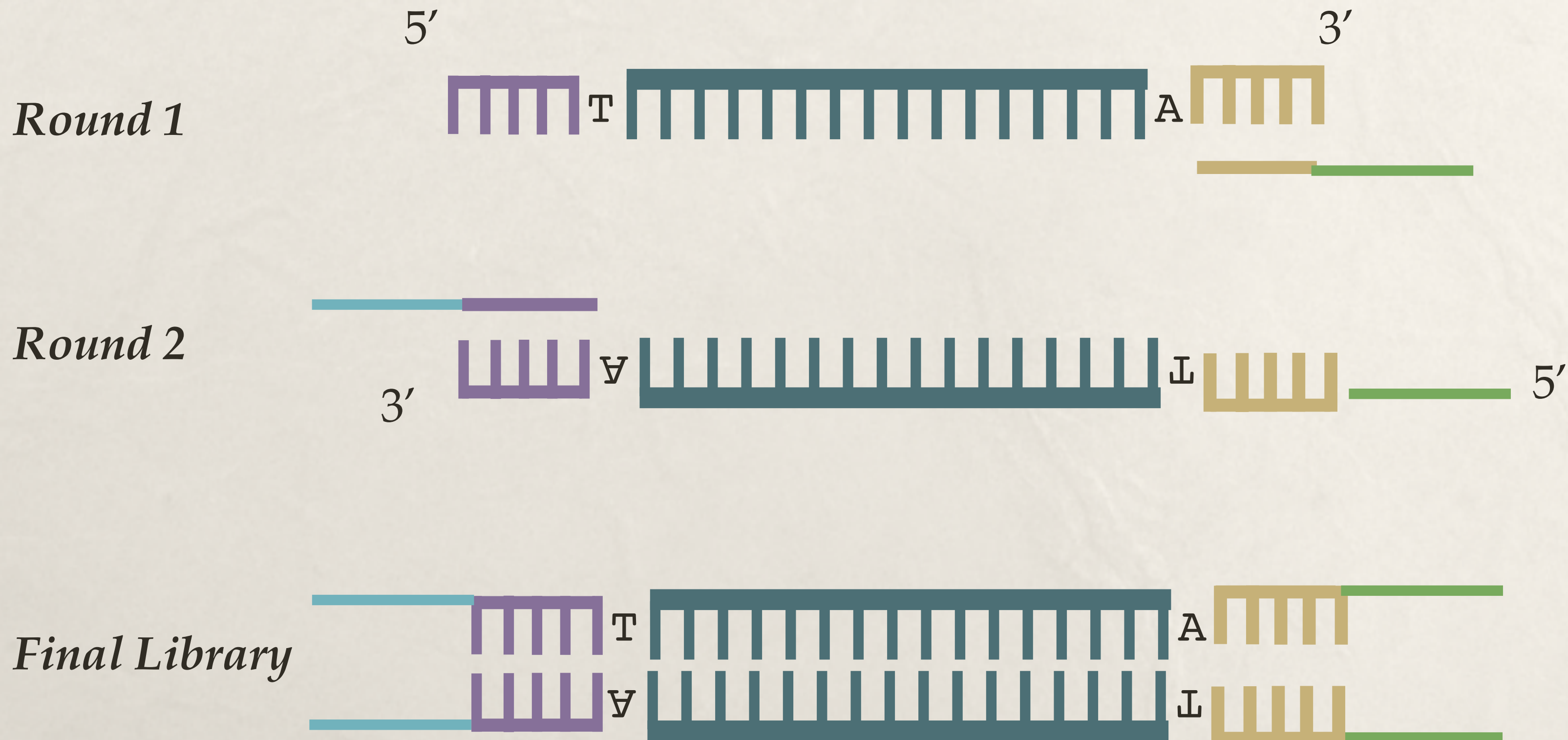
Transposase-mediated addition of sequencing adaptors (tagmentation)



This tends to be more expensive, but is very, very easy.
Tagmentation is the basis for ATAC-seq (to be discussed)

PCR-mediated amplification of your sequencing library

Indexing and addition of final sequencing-compatible ends



****If performing strand-specific RNA-seq using the dUTP method, UDPG treatment immediately precedes PCR enrichment*

What next?

- **Quality Control**
- **Sequencing**
- **Data analysis…**

Considerations: RNA-seq

- **RNA:** Did you enrich for the correct size/type RNA?
  Consider: Gel or BioA measurement of RNA size, controls

Very important

- **RNA Purity:** Do you have contaminating material?
  Consider: Methods to determine purity, efficacy of enzymes

DNA and rRNA removal critical

- **RNA-DNA Efficiency:** Do you have DNA after RT?
  Consider: Optimized RT protocols, checking enzyme efficiency, kit-recommended quality checks

Less critical if using validated reagents/kits

Considerations: ChIP-seq

- **Antibody:** Does your antibody work for ChIP?
  Consider: ChIP-western, IgG ChIP, qPCR validation

  MOST IMPORTANT

- **Sonication:** Is DNA sheared appropriately?
  Consider: Agarose gel measurement of shearing, shearing protocol optimization, alternative shearing methods

  Diagenode Bioruptor in Molecular Core
  Validate your sonication conditions EVERY TIME

- **DNA Isolation/Purification:** Do you have DNA?
  Consider: Qubit quantification of DNA, comparison of DNA to IgG or no antibody control ChIP

  NGS DNA amounts generally below Nanodrop threshold

Considerations: Topics

- **BEFORE making libraries, consider:**
  - Did your experiment "work"?
  - What strategy is best for barcoding/pooling?
  - How much sequencing *coverage* do you need?

Sequencing Coverage Resources

- Illumina:
  - http://www.illumina.com/documents/products/technotes/technote_coverage_calculation.pdf
  - http://support.illumina.com/downloads/sequencing_coverage_calculator.html
- Other:
  - https://genohub.com/recommended-sequencing-coverage-by-application/
  - https://www.encodeproject.org/data-standards/