

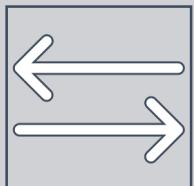
RNA-sequencing



What is RNA-seq and how does it differ from conventional sequencing?



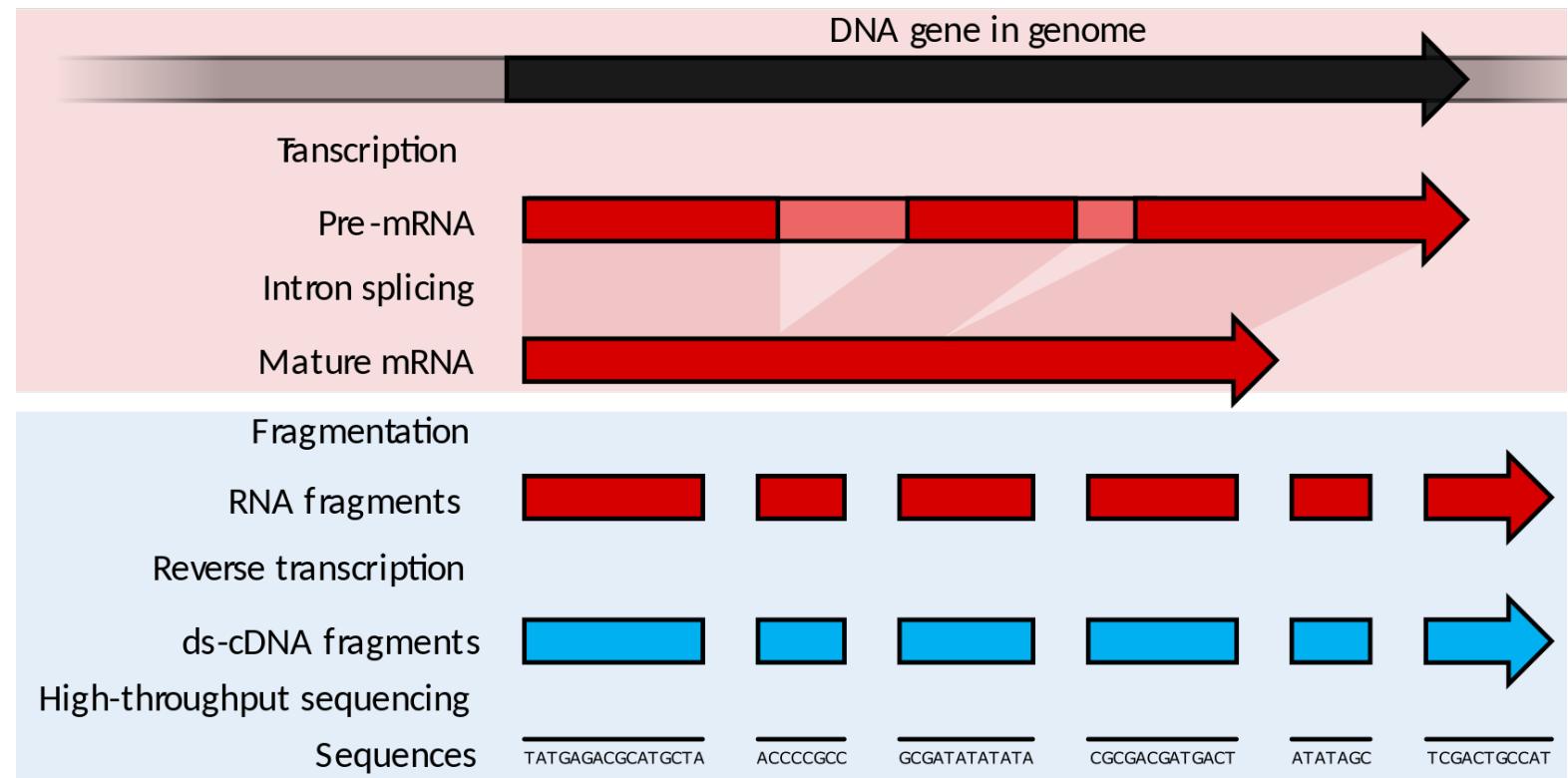
What are the goals and questions that can be addressed with RNA-seq?



Constructing RNA-seq libraries and special considerations

How is RNA different from DNA from the perspective of sequencing?

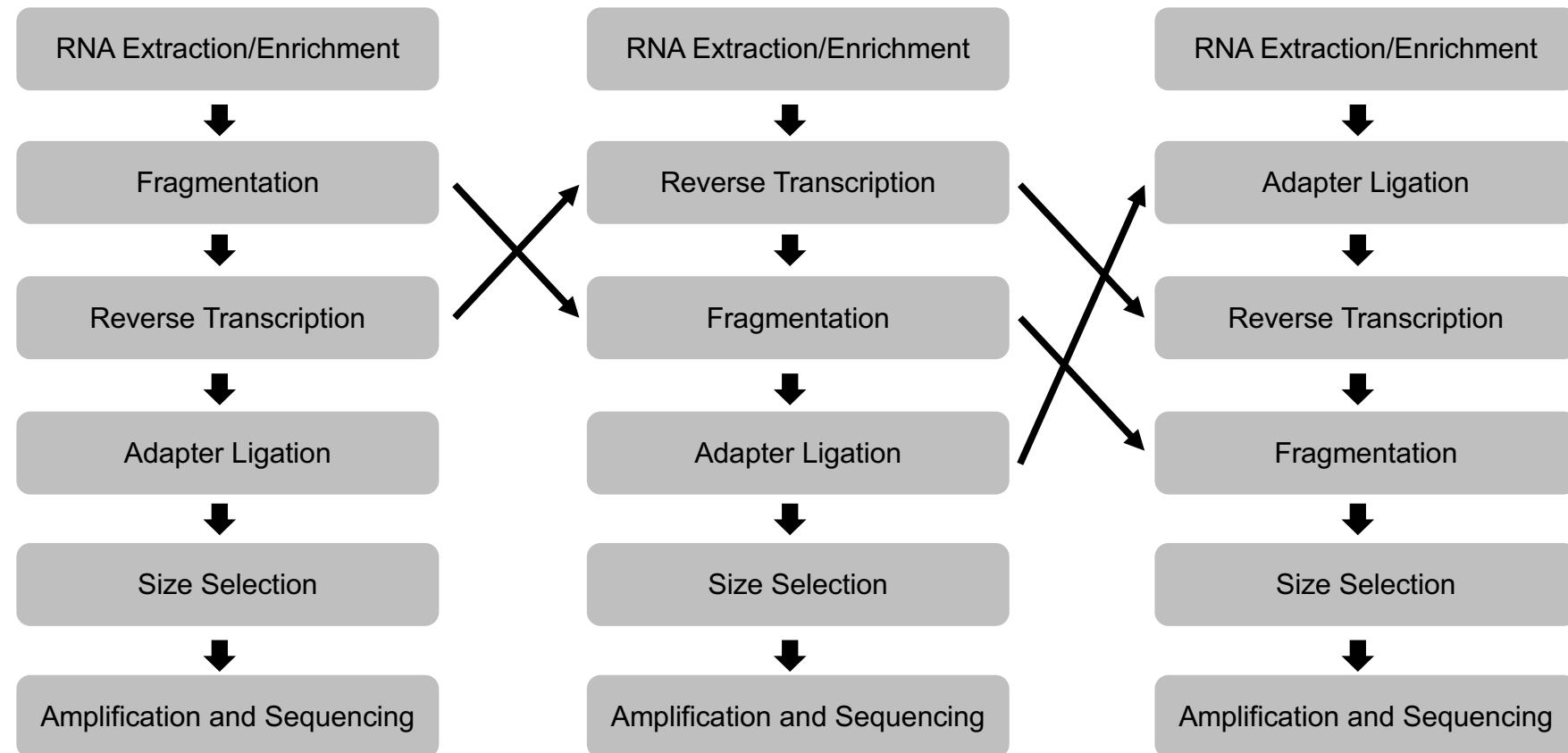
- Single-stranded
- Variable copy number
- Post transcriptionally modified
 - Intron splicing
 - RNA editing
 - Polyadenylation (Poly-A tail)



APPLICATIONS OF RNA-Seq

- Quantify gene expression
- Determine gene sequences in non-model species (de novo transcriptome assembly)
- Identify post-transcriptional modifications: (alternative) splicing, RNA editing
- Identify functional/transcribed elements (non-coding RNAs)
- Identify targets or footprints of RNA-binding proteins

RNA-seq Protocol Outline



RNA extraction



Silica-Membrane Kits

- Does NOT retain RNAs < 200 nt
- More expensive
- Lower yield?
- Higher purity?



TRIzol

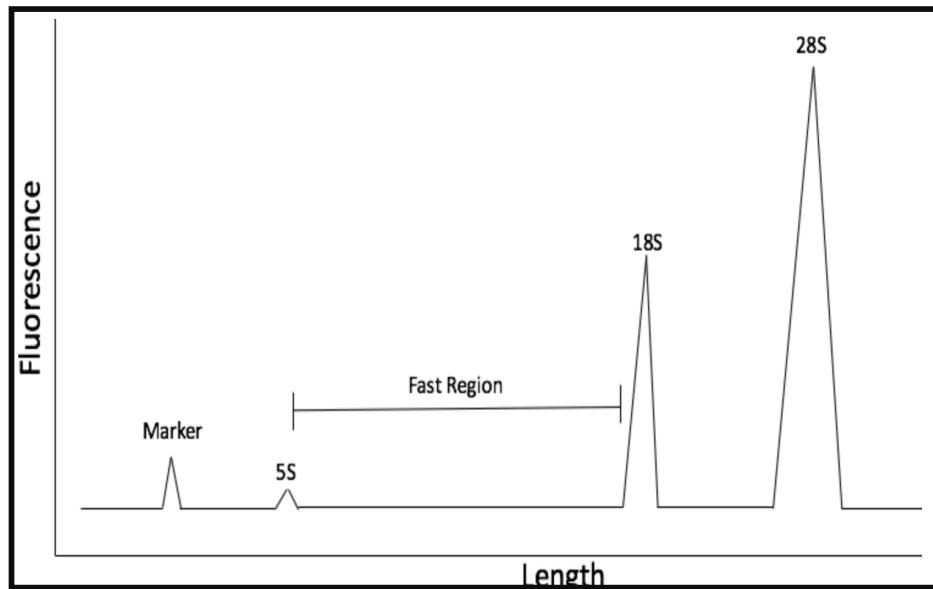
- Does retain RNAs < 200 nt
- Less expensive
- Higher yield?
- Lower purity?

RNA Quantity, Quality and RIN

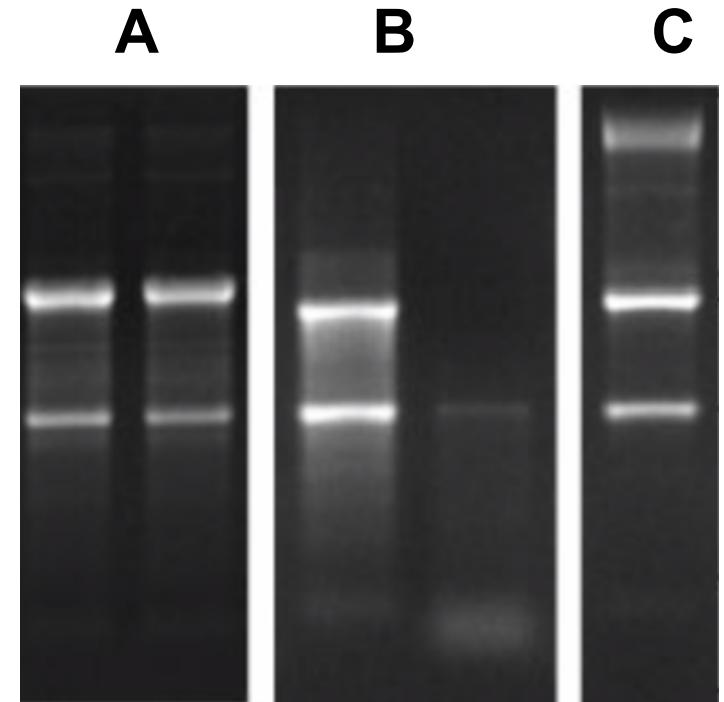
- **DNA contamination** (DNase treatment)
- **Measure concentration** (NanoDrop, Qubit/fluorometry, Agilent BioAnalyzer/TapeStation)
- **Verify purity** (NanoDrop, Qubit/fluorometry, agarose gel, PCR)

- **Verify size/integrity**

(agarose gel, Agilent BioAnalyzer/TapeStation)



An idealized TapeStation trace showing a RIN Score of 10



- A. Relatively non-degraded RNA showing a 2:1 28S and 18S bands
- B. Two degraded RNA samples, the right being severely degraded.
- C. DNA contamination

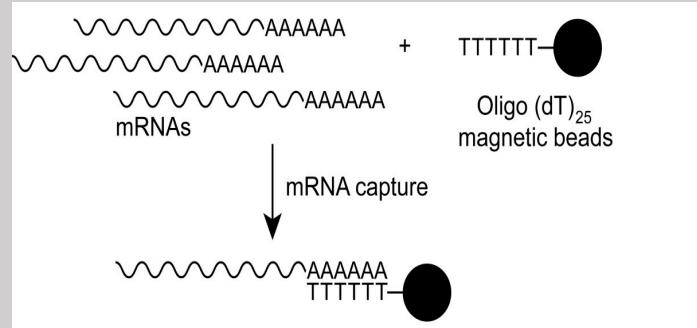
Capturing RNA molecules of interest

- ~80% of total RNA is rRNA
- ~15% is tRNA
- ~4% is mRNA**
- ~1% is other ncRNA

Ribosomal RNA (rRNA) Depletion

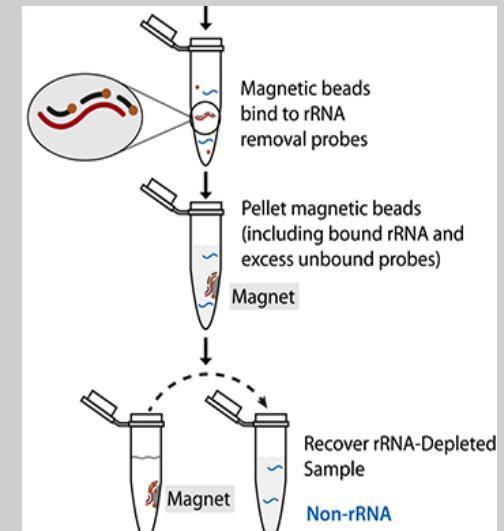
Positive Selection

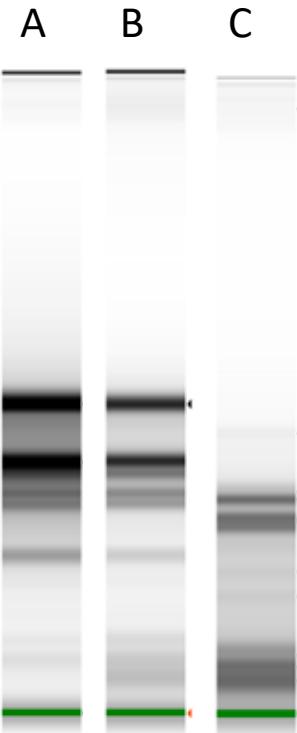
- Oligo-dT magnetic beads to capture polyA+ transcripts
- Inexpensive
- mRNAs only
- Biased representation (e.g., organellar RNAs, transcript maturation, AT-rich sequences, 3' ends)



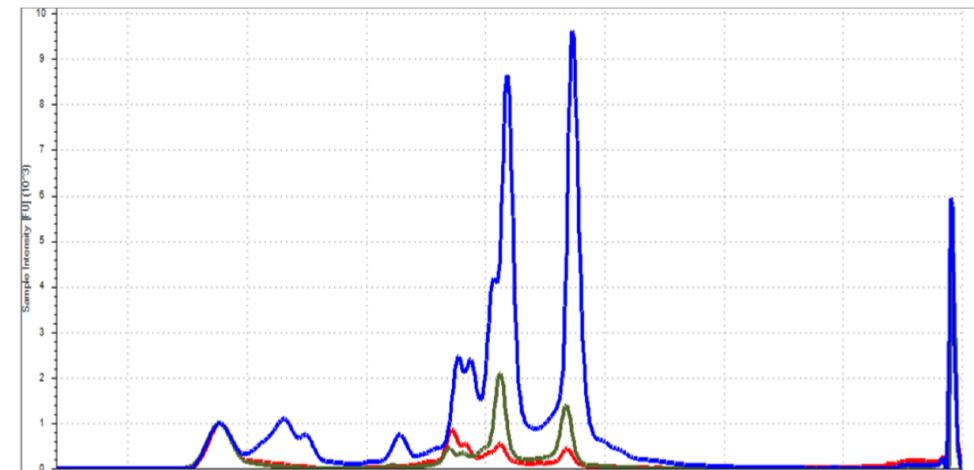
Negative Selection

- Targeted probes and magnetic beads for rRNA
- More costly
- Everything except rRNA
- Taxon-specific (not available for all organisms)





The trace below shows three different RNA samples on a TapeStation. Which colored line corresponds to 1. total cellular RNA, 2. captured rRNA, and 3. rRNA-depleted RNA?

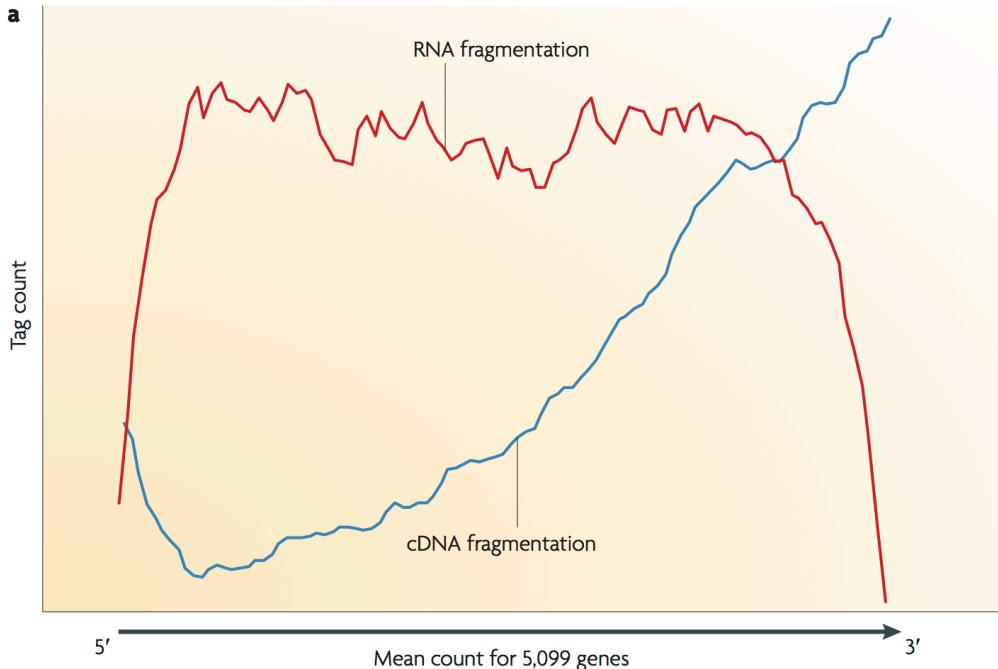


Fragmentation

Size limitation on Illumina sequencers is <600 bp.

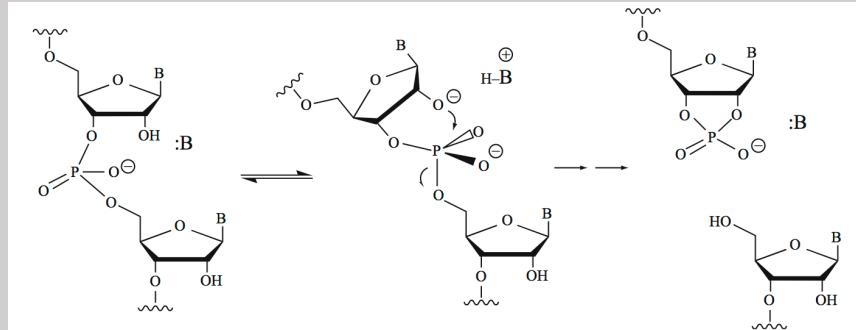
The average size of an animal mRNA is 2.5 kb.

Fragment at the cDNA or RNA stage?



Fragmenting RNA rather than cDNA results in less bias in coverage

Fragmentation methods



Chemical fragmentation of RNA with heat and salts (divalent cations)

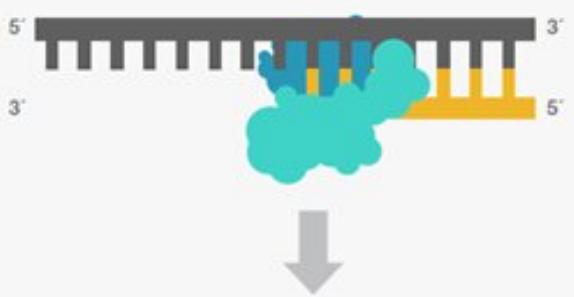
- Quick, cheap, and easy
- Fine-tunable: adjust incubation time

Enzymatic fragmentation (RNase III)

- Quick, cheap, and easy
- Has a preference for double-stranded RNA, can result in sequencing bias

Small RNA-seq → No fragmentation necessary!

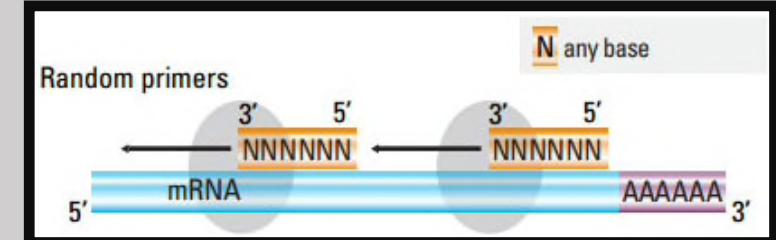
Reverse transcription



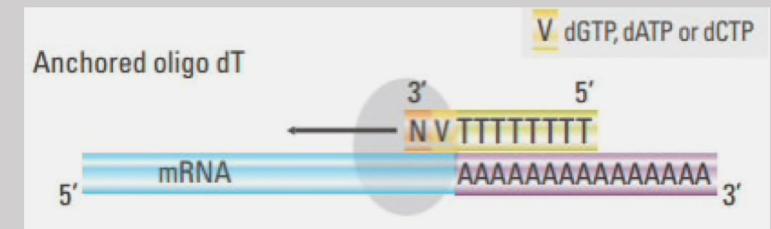
<https://www.thermofisher.com/us/en/home/life-science/cloning/cloning-learning-center/invitrogen-school-of-molecular-biology/rt-education/reverse-transcriptase-attributes.html>

[https://www.idtdna.com/pages/education/decoded/article/use-of-template-switching-oligos-\(ts-oligos-tsos\)-for-efficient-cdna-library-construction](https://www.idtdna.com/pages/education/decoded/article/use-of-template-switching-oligos-(ts-oligos-tsos)-for-efficient-cdna-library-construction)

- **Random primers (random hexamer): (preferable to oligo-dT primers)**



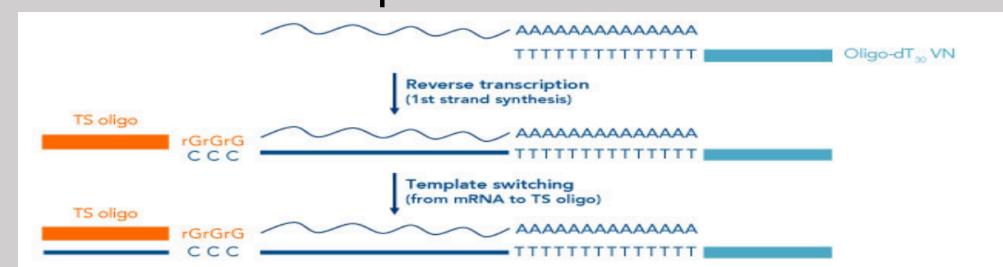
- Creates DNA that is complementary to RNA transcript → cDNA



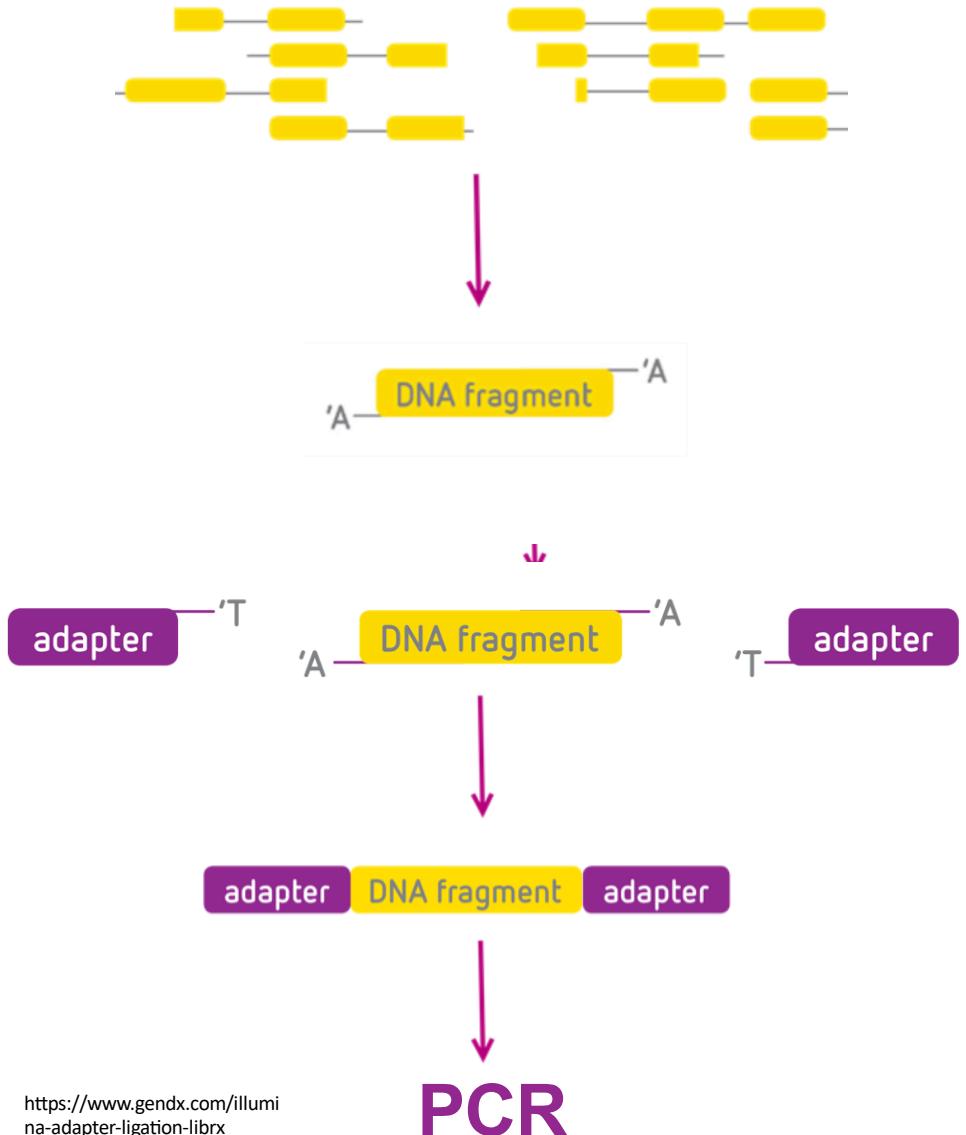
- First- and second-strand synthesis

Specialized RT methods

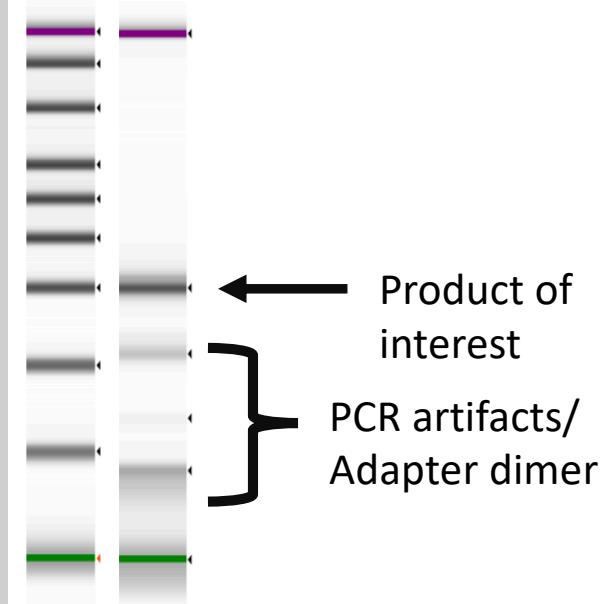
- Not So Random (NSR) primers. Avoids rRNA but species-specific, common in prokaryotic studies with small inputs
- Template-switching RT:



Adapter ligation, Amplification, Size Selection, Sequencing



- The ends of the DNA fragments are repaired to create blunt ends and then a dA-tail is added.
- The adapters are then ligated to the DNA fragment.
- PCR amplification: Normally 8–12 cycles are used during PCR.
- Size selection can be done to avoid sequencing PCR artifacts and adapter dimers.



Small RNA-seq libraries

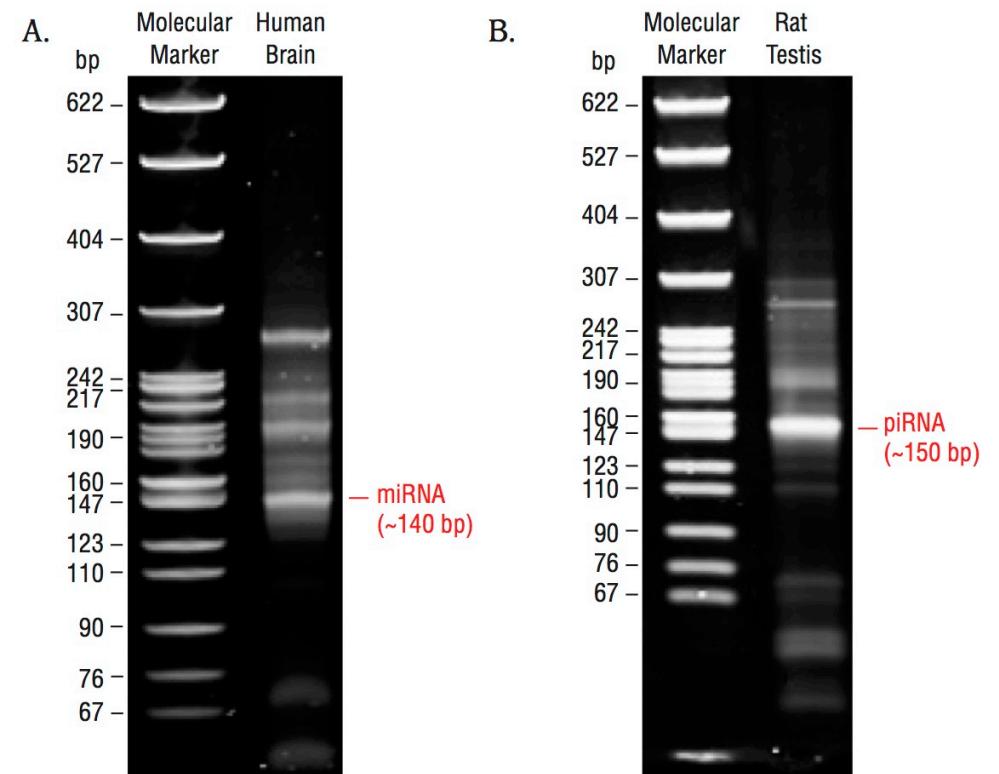
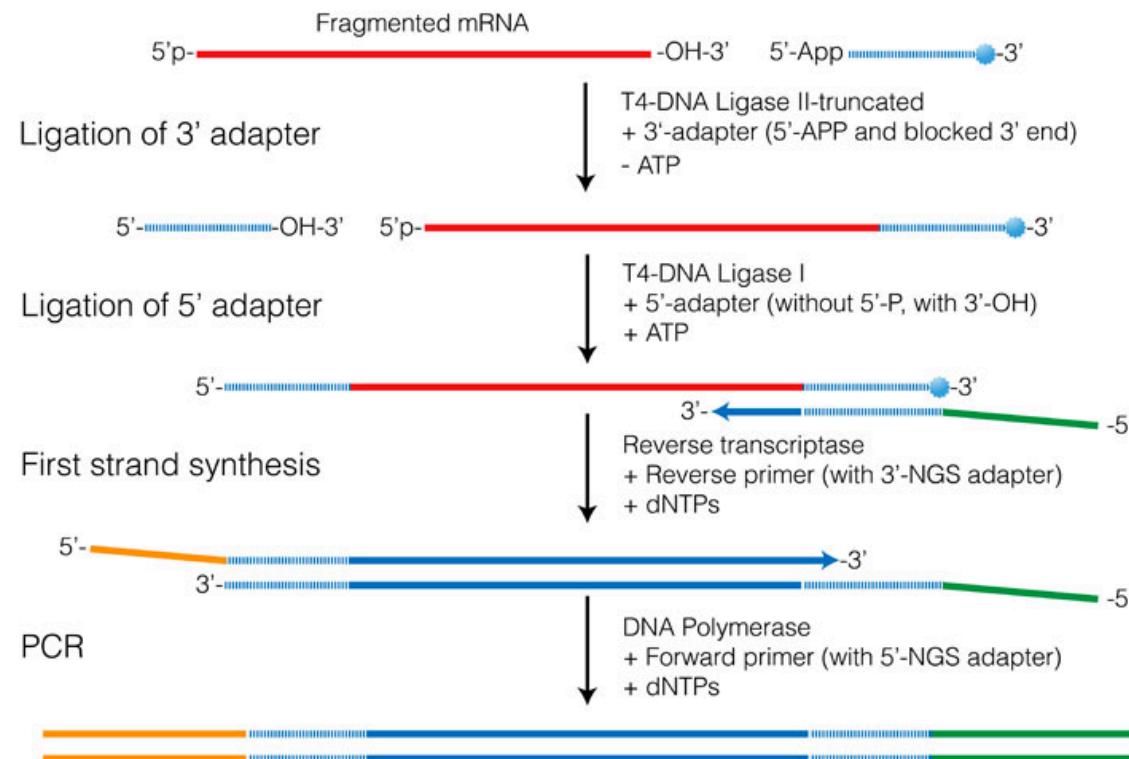
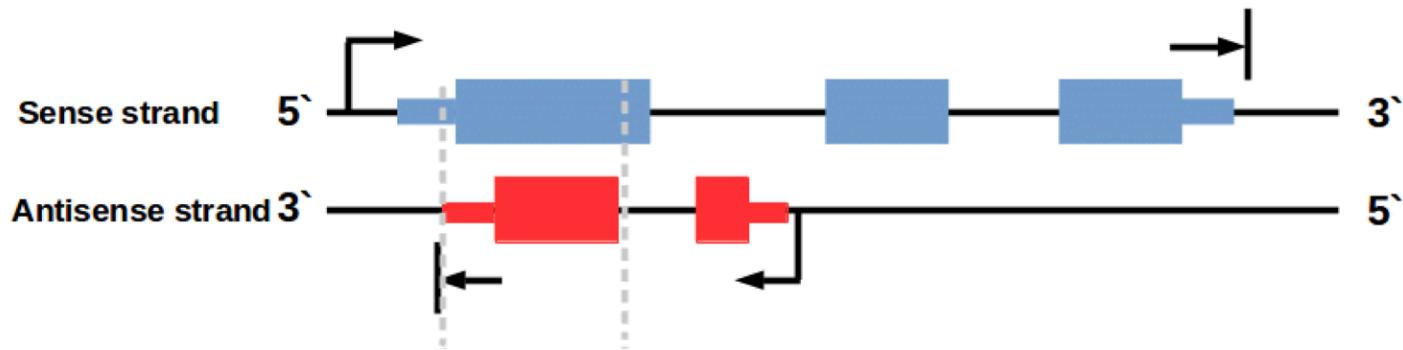


Figure 3: Shows typical results from Human Brain (A) and Rat Testis (B) Total RNA libraries. The 140 and 150 bp bands correspond to miRNAs (21 nt) and piRNAs (30 nt), respectively.

Adapters are ligated at the RNA stage

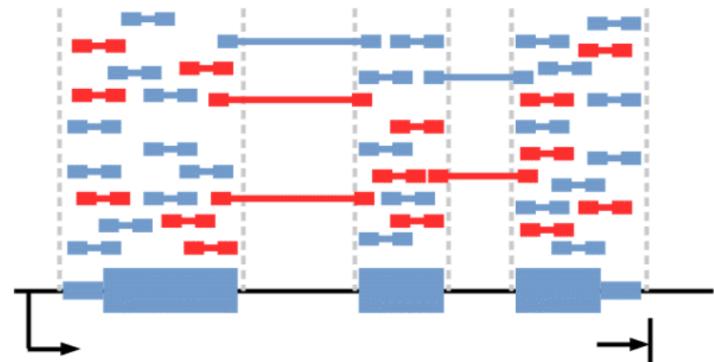
Strand-Specific RNA Sequencing

Sense vs. antisense matters!

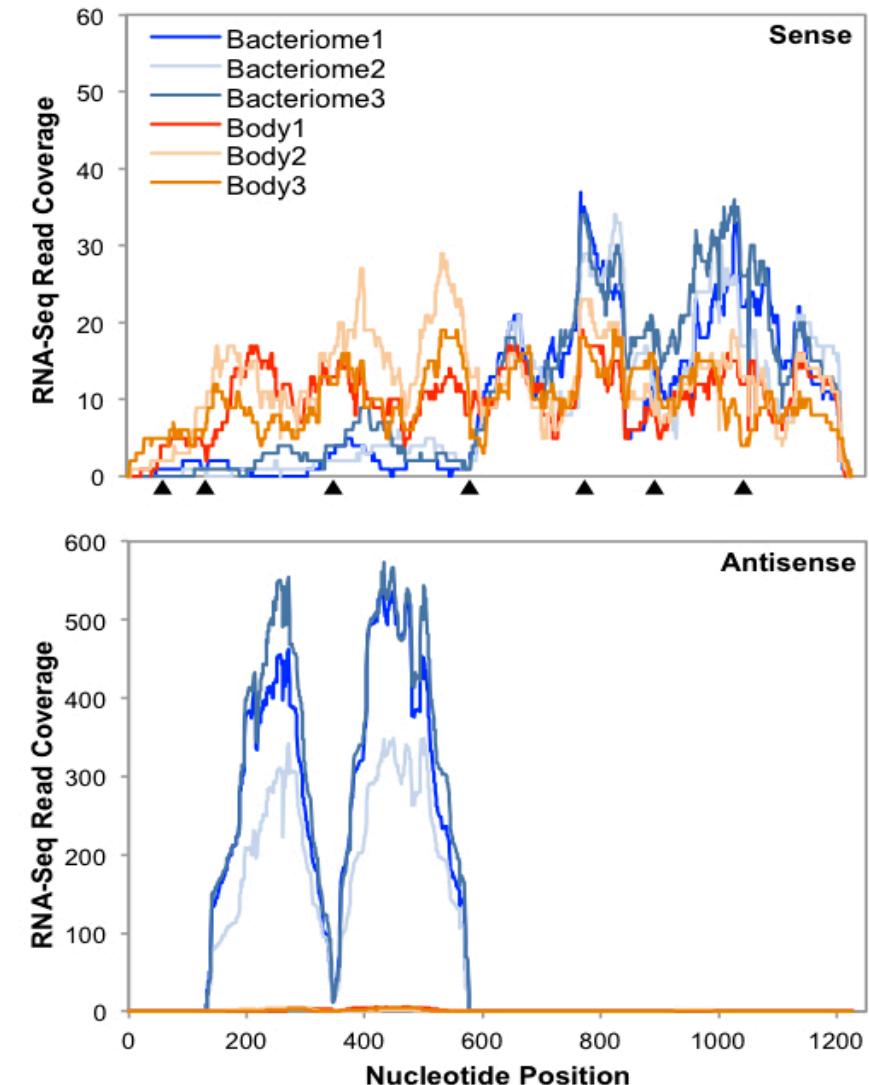


Which strand is transcribed?

A. Mapped reads from an unstranded library

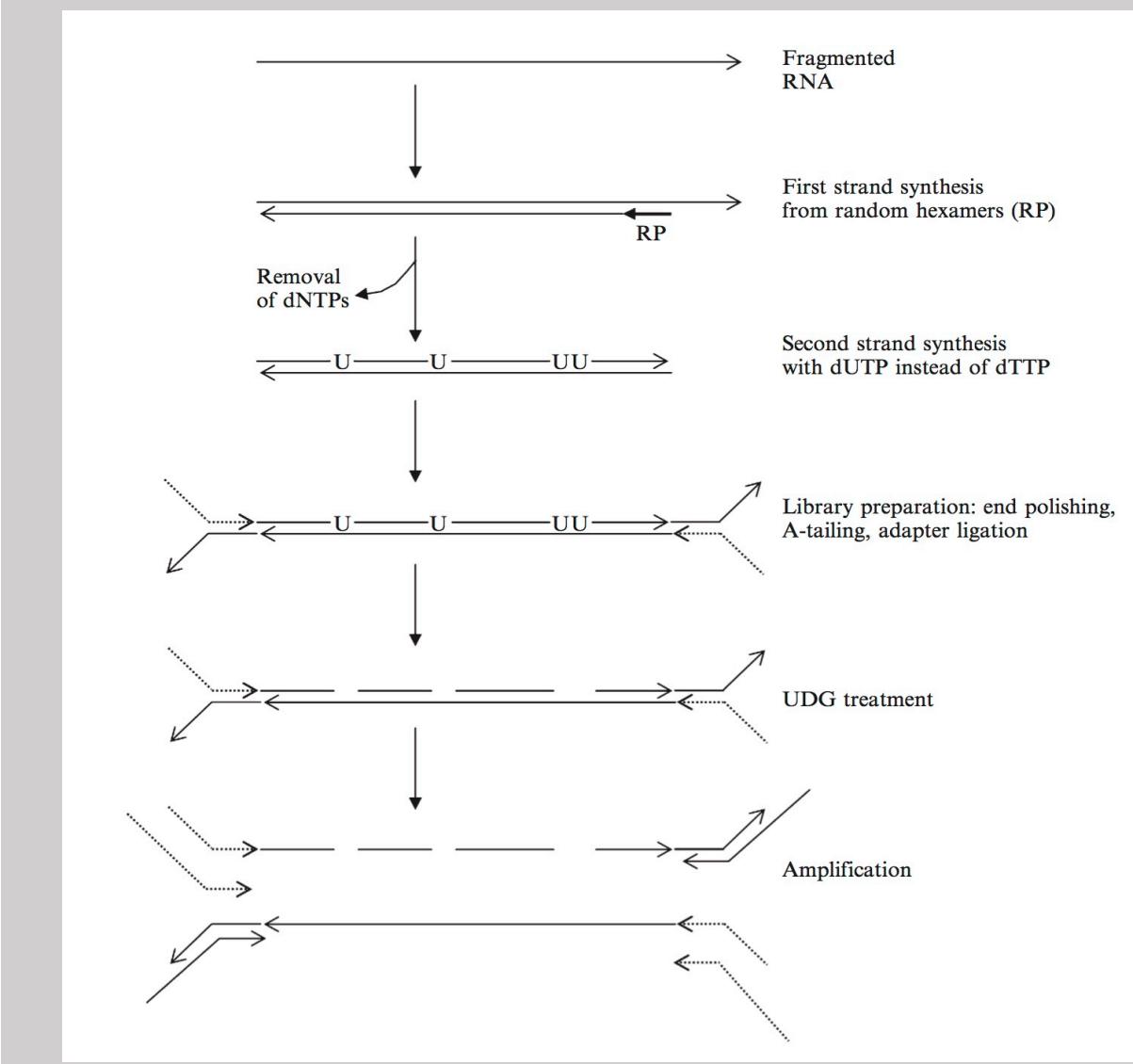


<https://www.ecseq.com/>



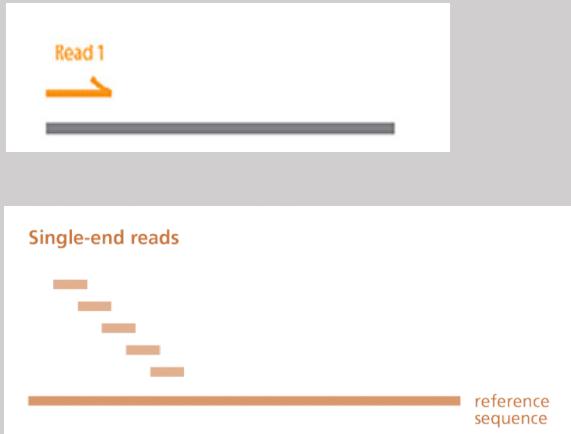
Strand-Specific RNA Sequencing

- Multiple alternative methods using RNA adapter ligation (reviewed in Levin et al. 2010 Nature Methods. 7:709–715).
- Template switching, 3' adapter sequence is added to the cDNA molecule, and first strand cDNA molecule can be PCR amplified directly without second strand synthesis.
- **The “strand-marking” dUTP/UDG method has emerged as the standard.**
- If using an RNA-seq kit, check to make sure it directional/strand specific.



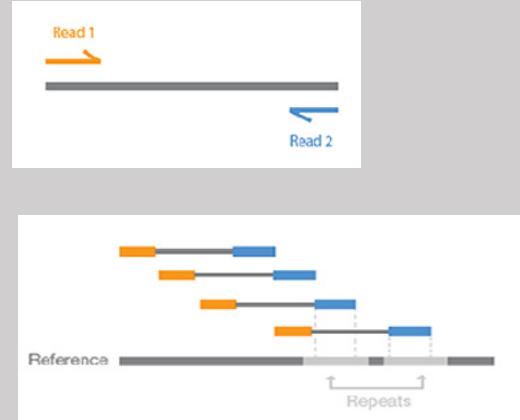
Paired-End vs. Single Read

Single-end reads



Gene expression: independent read counts rather than total sequence are the relevant measure of sample size. Get more independent reads per dollar with single-read sequencing.

Paired-end reads



Applications of longer reads and paired-end sequencing for RNA-seq

- *De novo* transcriptome assembly
- Mapping to repetitive regions

Normalization

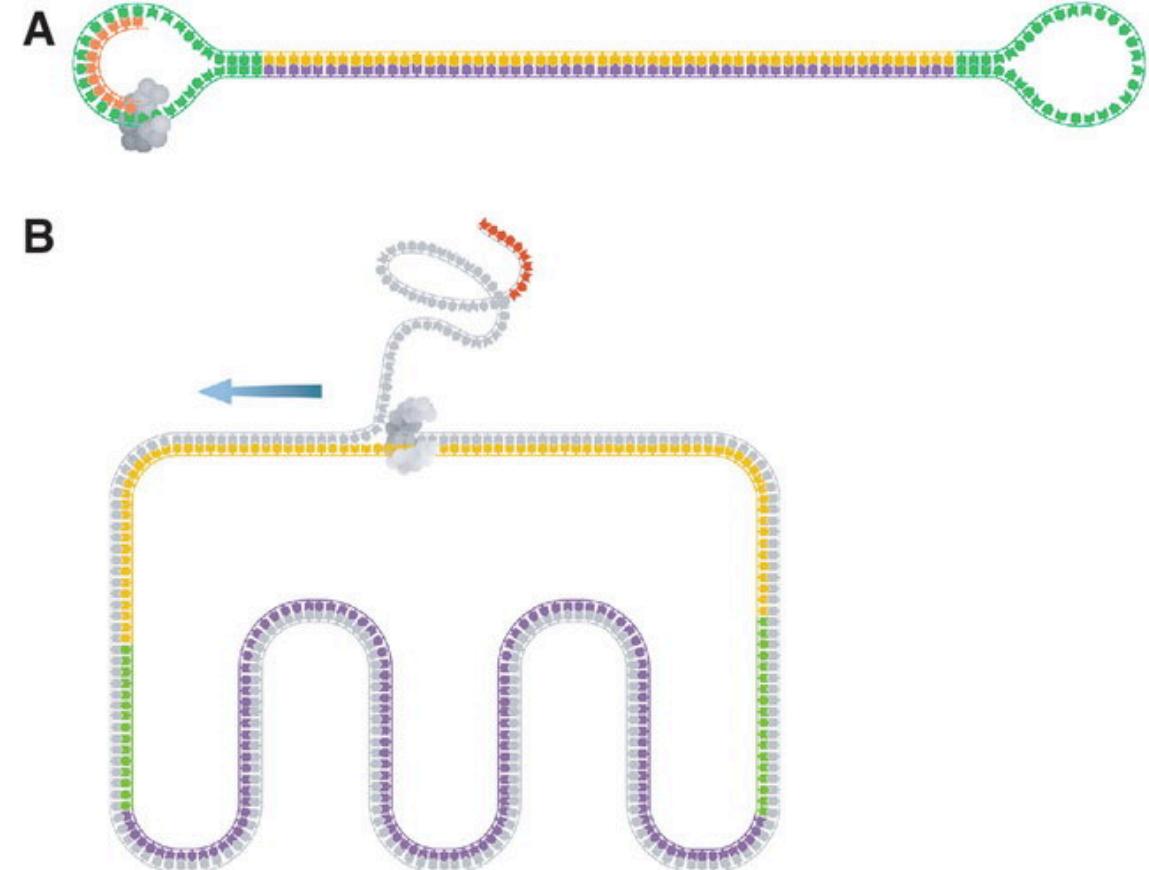
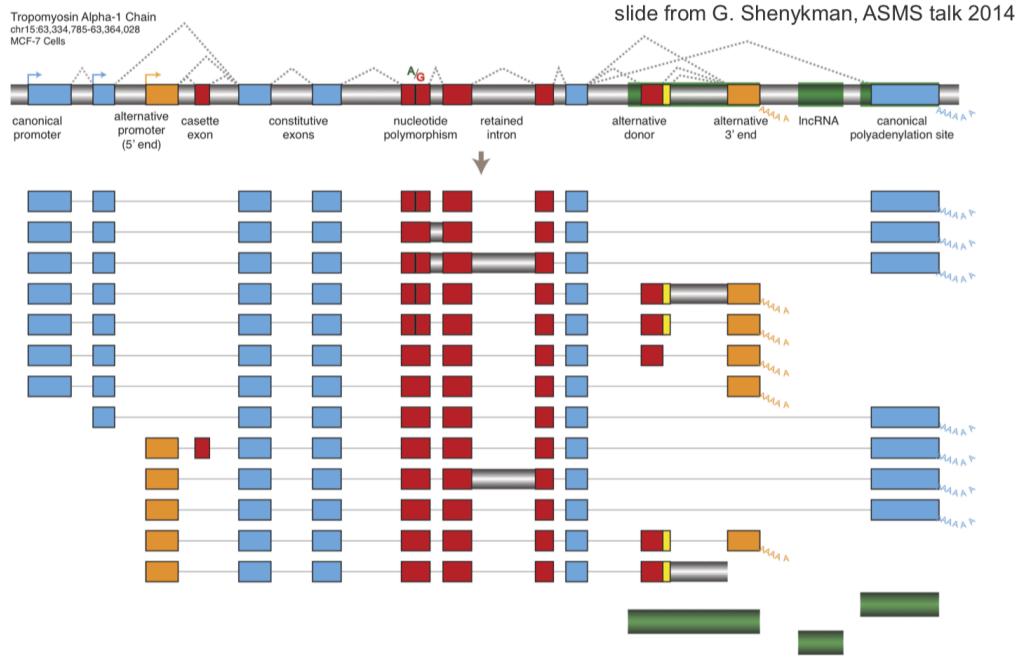
Gene expression levels span more than five orders of magnitude

- Reducing high-copy sequences and enriching for rare transcripts
- Duplex-specific nuclease (DSN) was isolated from the Kamchatka Crab and digests double-stranded but not single-stranded DNA.
- Re-annealing kinetics: after melting, high-copy sequences reform double-stranded complexes faster than low-copy sequences, so subsequent DSN treatment will preferentially degrade high-copy sequences.
- Requires subsequent re-amplification step (lowers library complexity)



As sequencing costs have declined, less important for saving money; largely replaced by digital normalization (khmer; Titus Brown) to improve computational efficiency.

PacBio- Full-Length cDNAs (Iso-Seq)



- Generate full-length cDNA sequences — no assembly required — to characterize transcript isoform
- Determine relative timing of different splicing and/or RNA editing events

You want to sequence the protein-coding transcripts (mRNA) from a newly discovered marine annelid that can survive high temperatures. Design an RNA-seq experiment, considering which method you would use for each step:

- 1. RNA extraction**
- 2. Check quality/contamination of RNA**
- 3. rRNA depletion**
- 4. Single or Paired-end reads**

