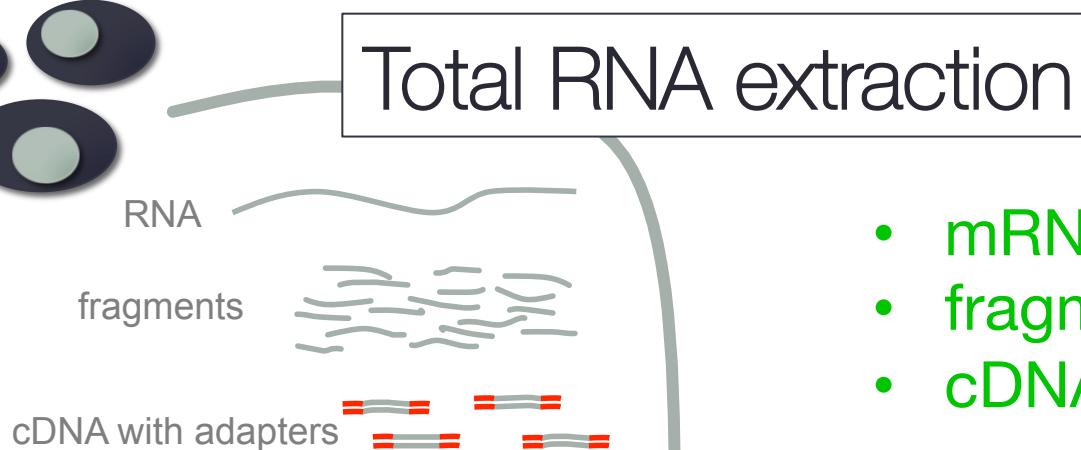
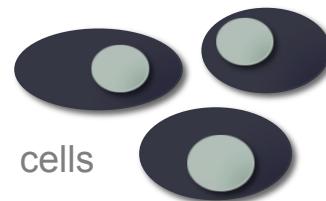


RNAseq: Experimental Design and Wet Lab Considerations

28-Jan-2020

RNA-seq workflow overview



- mRNA enrichment
- fragmentation
- cDNA library preparation



Sequencing

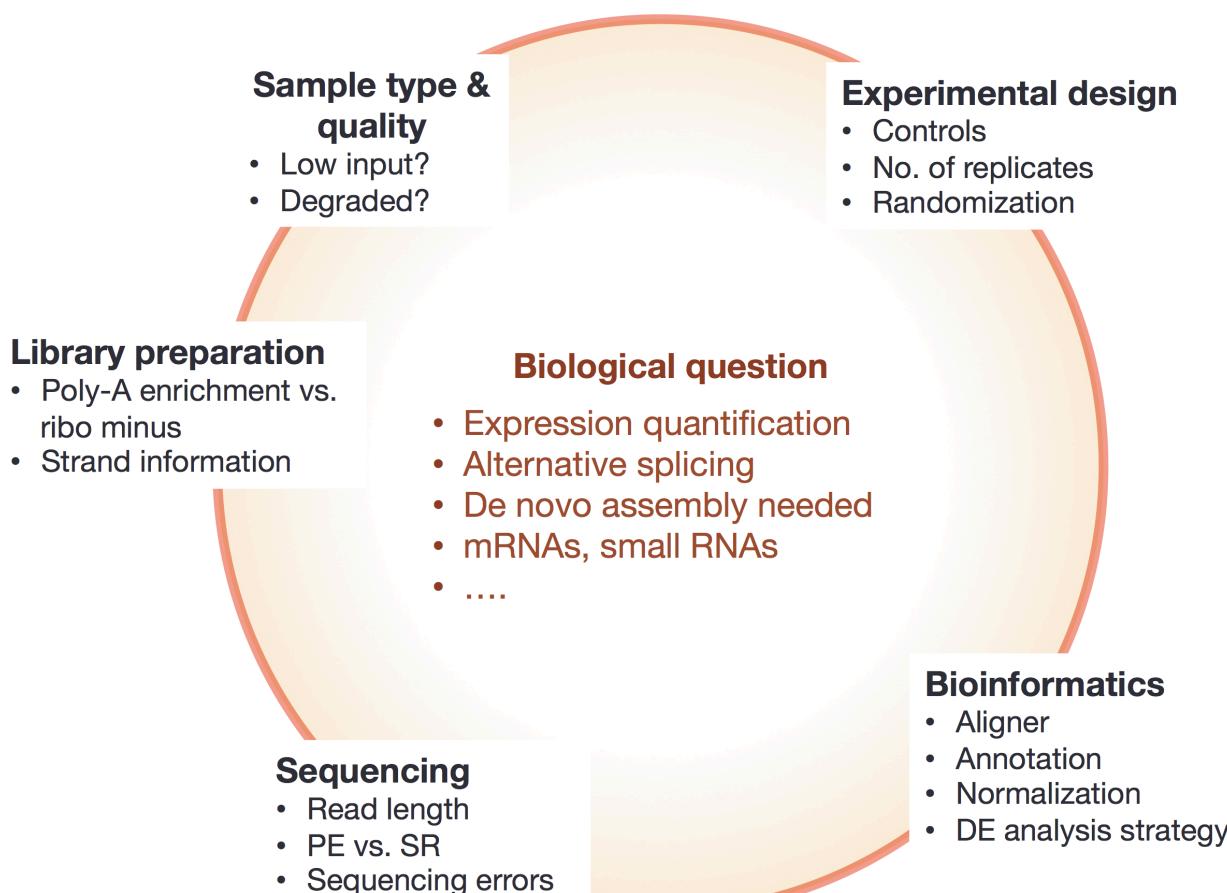


- cluster generation
- sequencing by synthesis
- image acquisition

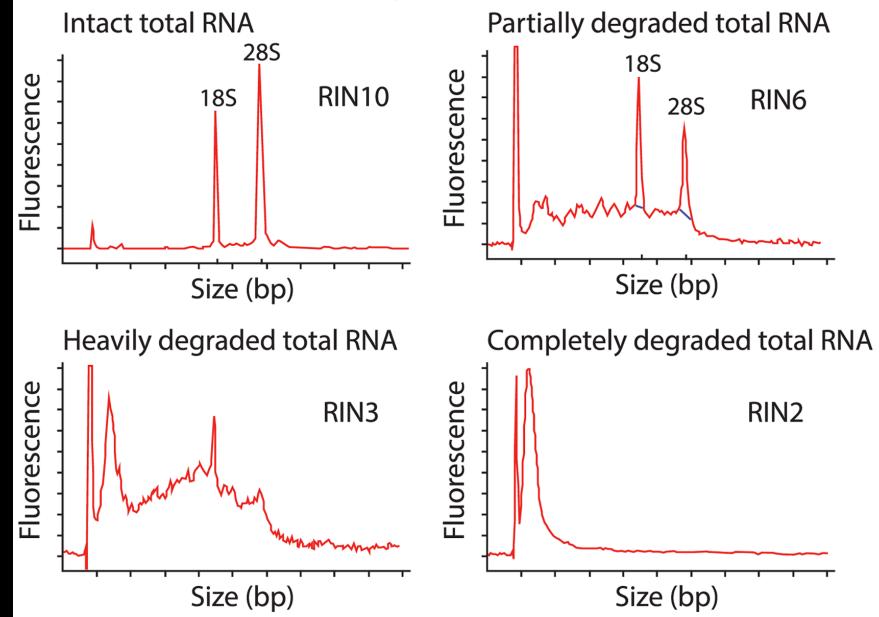
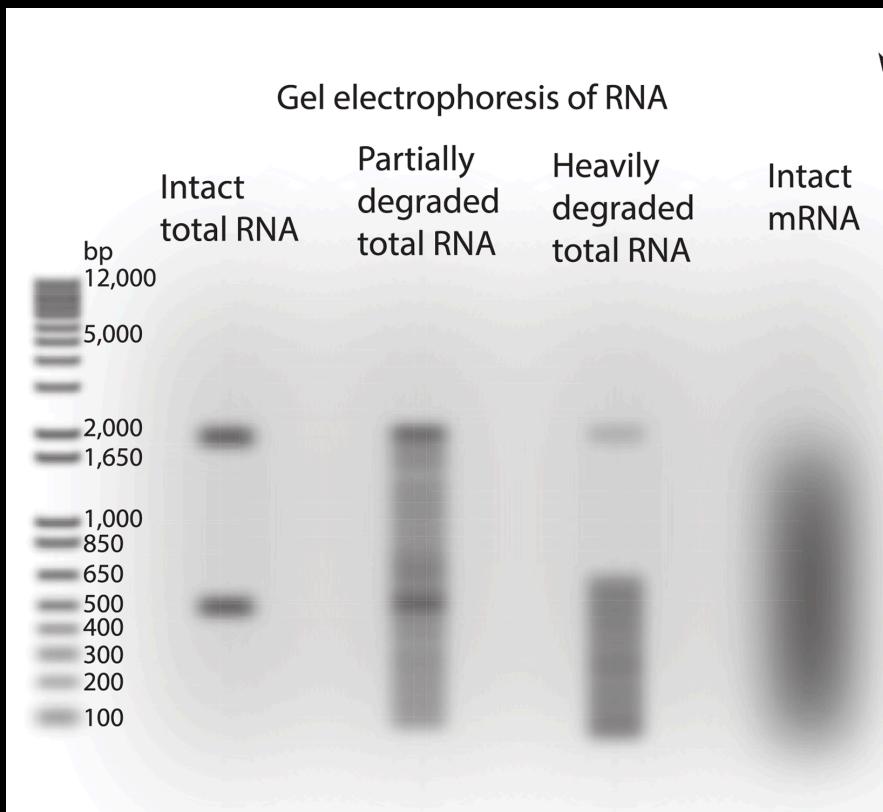
Bioinformatics

Key Considerations

Everything's connected...



Quality of RNA



- RIN: RNA Integrity Number
 - Ratio of 28S:18S
 - 10 ideal; above 8 is acceptable.
 - Below 8: varies by application

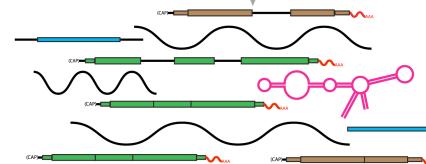
Influence of the RNA enrichment strategy

mostly
rRNA & tRNA
< 2% mRNA!

which transcripts are
you interested in?

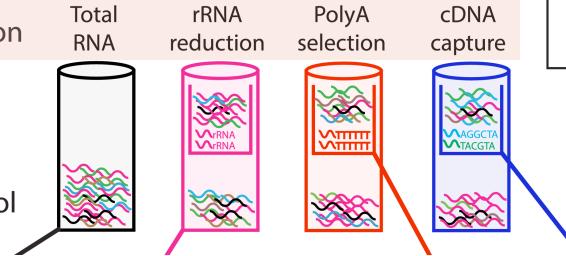
what type of noise
can you tolerate?

Initial RNA pool



Selection/depletion

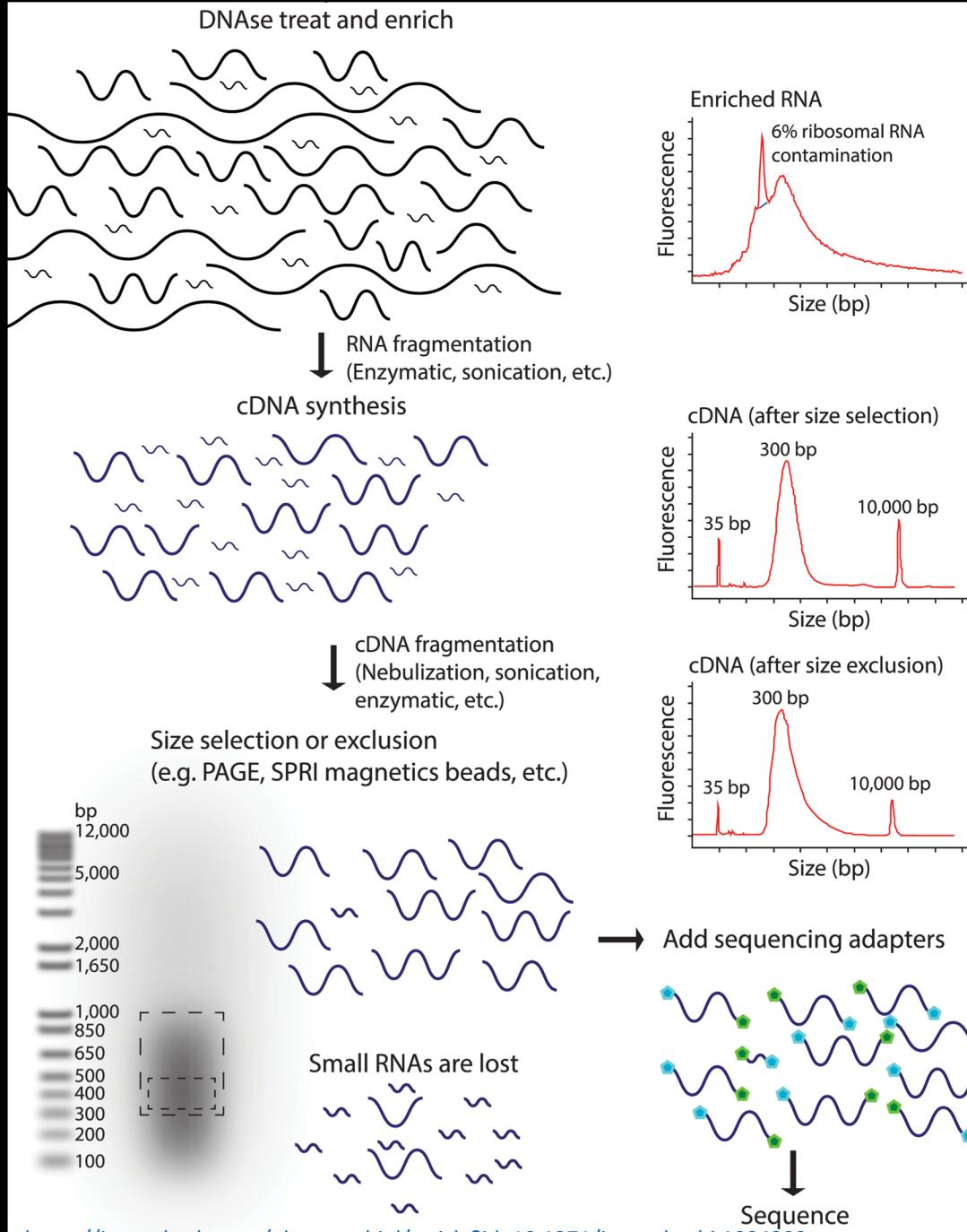
Resulting RNA pool



Legend	
wavy line	genomic DNA
brown line with cap	immature RNA
brown line	mature RNA
blue line	non-coding RNA
pink loop	ribosomal RNA
dashed line	paired end reads

Fragmentation & Size Selection

- Over vs under shearing will impact your data
- Small RNAs are lost in many standard protocols
- Size selection:
 - Can set upper and lower cutoffs:
 - Gel extraction
 - Beads
 - Columns – only exclude small RNAs



RNAseq – Best Practices in the lab

- Work in a dedicated RNA space to reduce contamination and degradation.
- Keep it simple
- Include biological replicates when applicable
- Be wary of batch effects → include technical replicates

General Recommendations for RNAseq options based upon experimental objectives.

Criteria	Annotation	Differential Gene Expression
Biological replicates	Not necessary but can be useful	Essential
Coverage across the transcript	Important for de Novo transcript assembly and identifying transcriptional isoforms	Not as important; however the only reads that can be used are those that are uniquely mappable.
Depth of sequencing	High enough to maximize coverage of rare transcripts and transcriptional isoforms	High enough to infer accurate statistics
Role of sequencing depth	Obtain reads that overlap along the length of the transcript	Get enough counts of each transcript such that statistical inferences can be made
DSN	Useful for removing abundant transcripts so that more reads come from rarer transcripts	Not recommended since it can skew counts
Stranded library prep	Important for de Novo transcript assembly and identifying true anti-sense transcripts	Not generally required especially if there is a reference genome
Long reads (>80 bp)	Important for de Novo transcript assembly and identifying transcriptional isoforms	Not generally required especially if there is a reference genome
Paired-end reads	Important for de Novo transcript assembly and identifying transcriptional isoforms	Not important

Recommended Sequencing Depth

- Varies by application
- Some general guidelines:
 - Differential Expression: 10-25M
 - Alternative Splicing: 50-100M
 - Allele-specific expression: 50-100M
 - *De novo* assembly: >100M
 - Small RNA sequencing (microRNA):
 - DE: 1-2M
 - Discovery: 5-8M

Typical biases of Illumina sequencing

- sequencing errors
- miscalled bases

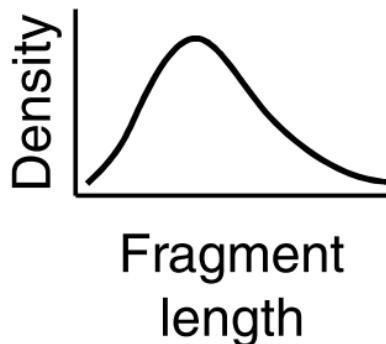
specific for the sequencing platform/
machine

• PCR artifacts (library preparation)

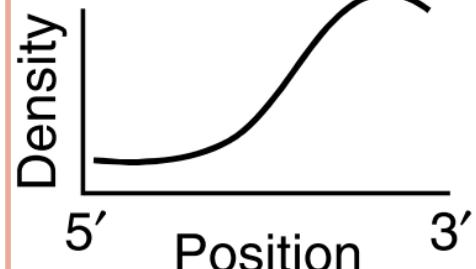
- duplicates (due to low amounts of starting material)
- length bias
- GC bias

sample-
specific
problems!

Fragment length
(size selection)



Positional bias
(degradation)



RNA-seq-specific

sequence bias
(PCR amplification)

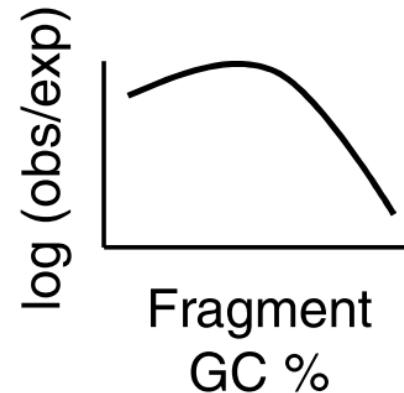
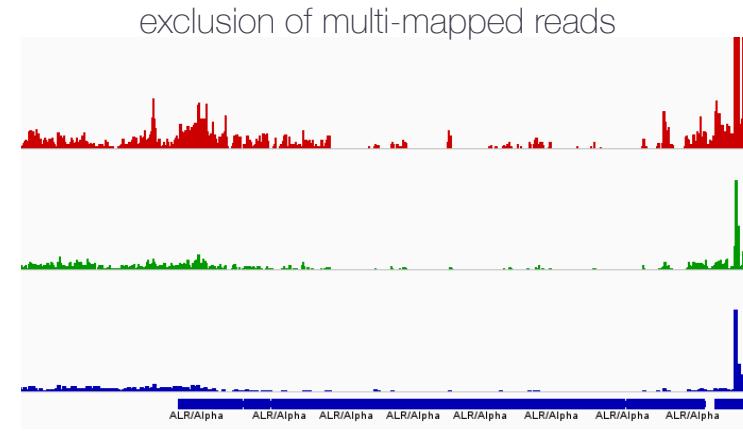
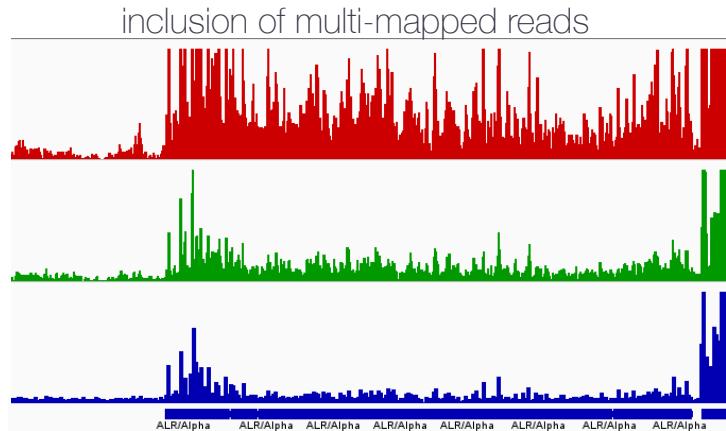


Figure from Love et al. (2016). Nat Biotech, 34(12). More details & refs in course notes (esp. Table 6).

General sources of biases (not inherently sample-specific)

- issues with the **reference**
 - CNV
 - mappability
- inappropriate **data processing**



Resources available at LAB & NMNH

- Dedicated RNA space w/ thermocycler and Qubit
- In-House sequencing available:
 - Illumina MiSeq
 - ONT MinION
- Able to facilitate outsourced sequencing (via Genohub)
- Experienced personnel – training available for many protocols
- Bioinformatics – Hydra & support from OCIO



Additional Resources available at LAB

- Main lab at NMNH:
 - Qubit
 - iD3 Plate Reader (Quant-iT)
 - Covaris ME220 Ultrasonicator
 - Qsonica Q800R Sonicator
 - BluePippin
 - TapeStation 2200
 - Bioanalyzer 2100
 - ViiA7 Real-Time PCR system (qPCR)
- At MSC:
 - TapeStation 4200 (WRBU)
 - ViiA7 Real-Time PCR system



Helpful Links

- <https://journals.plos.org/ploscompbiol/article/file?type=supplementary&id=info:doi/10.1371/journal.pcbi.1004393.s009>
- <https://genohub.com/recommended-sequencing-coverage-by-application/>
<https://www.illumina.com/library-prep-array-kit-selector.html>
- <https://rnaseq.uoregon.edu/>
- <https://chagall.med.cornell.edu/RNASEQcourse/>
- <https://bmcgenomics.biomedcentral.com/articles/10.1186/s12864-019-5953-1>

**Questions on extraction, library prep,
and/or sequencing considerations?**