


# Galaxy

## RNA-seq Analysis: Mapping

[www.galaxyproject.org](http://www.galaxyproject.org)

# RNA-Seq Exercise

Create new history

 (cog) → Create New

Get some data

Shared Data → Data Libraries

→ Demonstration Datasets

→ Human RNA-seq: CHB ENCODE Exercise

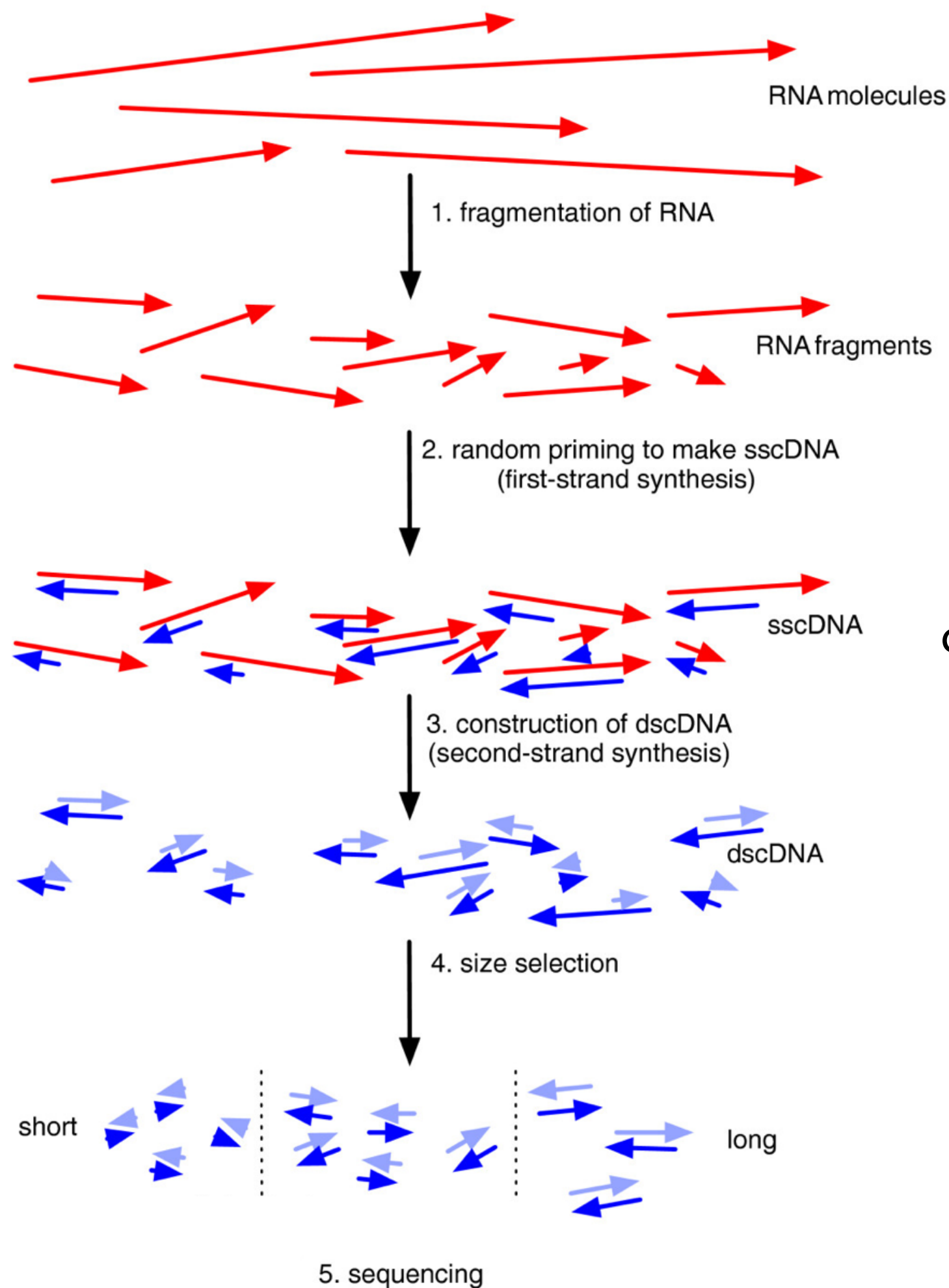
→ Select All Datasets in folder  
and then Import to current history

(We're ignoring quality control,  
in practice this would be a good time for FASTQC)

# **RNA-seq data generation**

Enrich for RNA  
population of  
interest upfront

(e.g. with Poly-dT  
beads for mRNA)

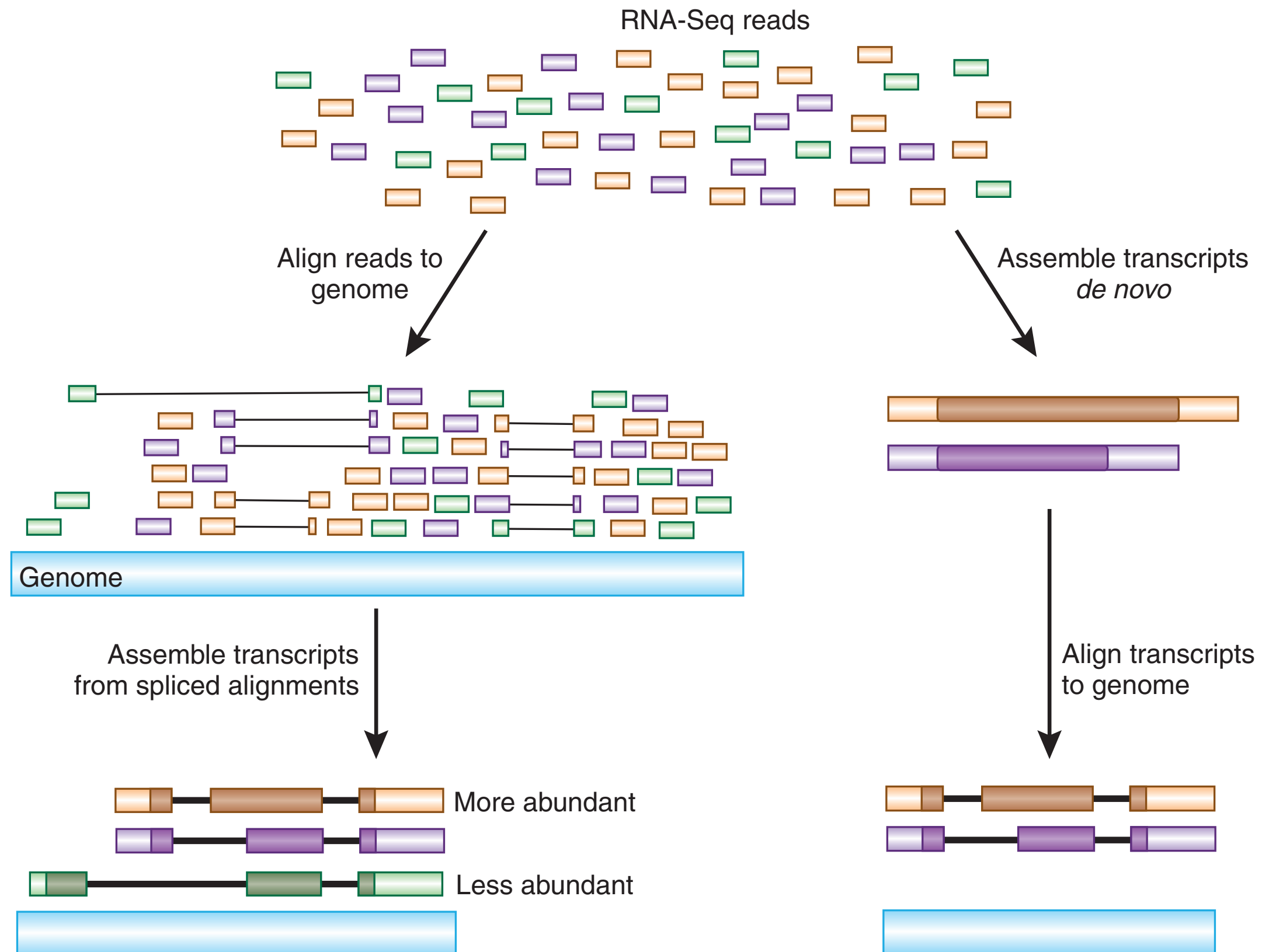


RNA to cDNA  
conversion using  
random primers

# RNA-seq data analysis

- Can be analyzed in many different ways depending on goals of the experiment, what other data is available, et cetera

# Align-then-assemble or *de novo*?



- Align-then-assemble: potentially more sensitive, but requires a reference genome, confounded by structural variation
- *de novo*: likely to only capture highly expressed transcripts, but does not require a reference genome, robust to variation



# RNA-seq Exercise: Mapping with Tophat

Tophat is a spliced read aligner for aligning RNA-seq data to a reference genome

# Mapping with Tophat: Use Existing Annotations?

You can bias Tophat towards known annotations

- Use Own Junctions → Yes
  - Use Gene Annotation → Yes
  - Gene Model Annotation → Gene Annotations chr19

# Mapping with Tophat: Use Existing Annotations?

You can bias Tophat towards known annotations

- Use Own Junctions → Yes
  - Use Gene Annotation → Yes
  - Gene Model Annotation → Gene Annotations chr19

(This is going to take about 10 minutes)

## Summary

RNA-seq analysis using a reference genome requires an aligner that is splicing aware (can handle long “deletions” in the reads)

Tophat2 is one such aligner, based on bowtie, and available in Galaxy