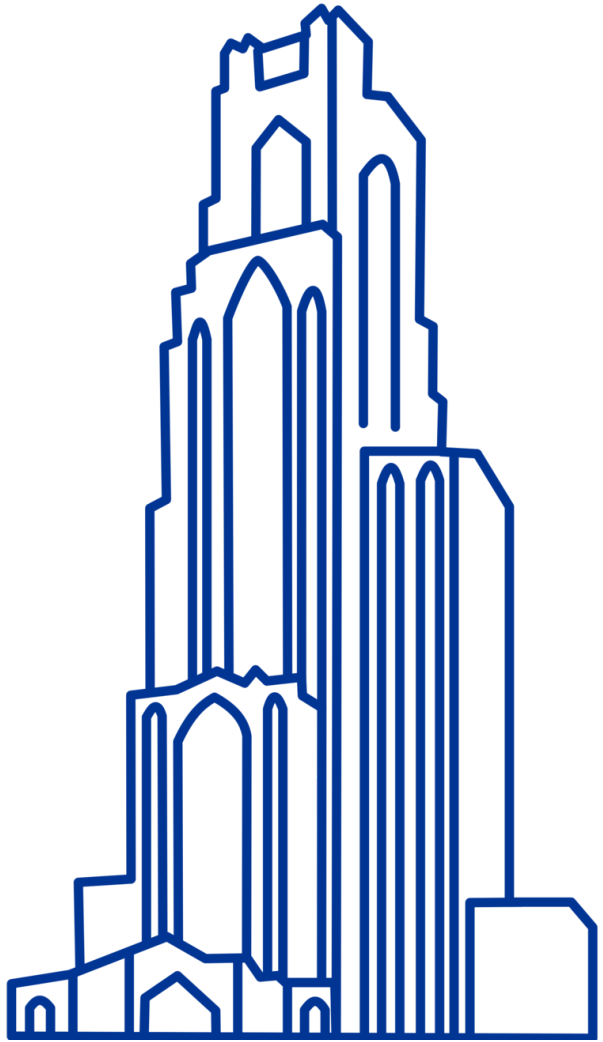


Computational Biology

(BIOSC 1540)

Lecture 06A

Read mapping

Foundations

Feb 11, 2025

University of Pittsburgh

Announcements

Assignments

- Assignment P01D is due Friday (Feb 14)

Quizzes

- Quiz 02 is on Feb 18 and will cover lectures 04A to 06A

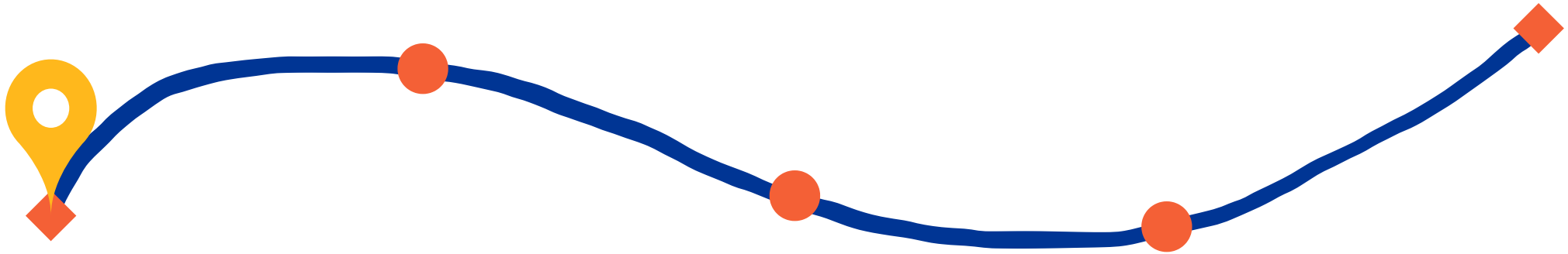
CBytes

- CByte 03 **expires** on Feb 15
- CByte 04 **expires** on Feb 28

Next reward: Checkpoint Submission Feedback

ATP until the next reward: 653

After today, you should have a better understanding of



How transcriptomics extends beyond genomics

DNA

Genomics helps answer key biological questions

Genomics tells us **what's possible for an organism to do** but not when or how it does it.

Questions genomics can answer:

- **What genes are present?** (e.g., Does a bacterium have antibiotic resistance genes?)
- **How are species related?** (e.g., Evolutionary trees based on genome sequences.)
- **What mutations exist?** (e.g., Cancer-causing genetic changes.)

A genomic sequence alone doesn't tell us what genes are active

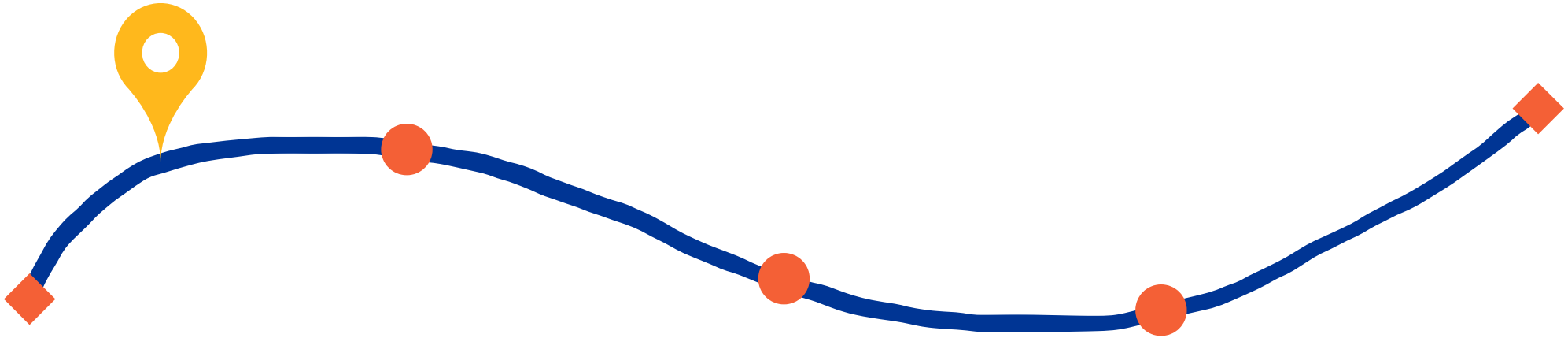
DNA is like a book of instructions—just because a gene exists doesn't mean it's being used.

Examples:

- Every cell in your body has the same genome, but a **neuron** and a **liver cell** express **different genes**.
- In cancer, certain genes are **turned on or off incorrectly**—but genomics alone can't detect this.

Key insight: To understand cellular function, we need to know **which genes are active and when**.

After today, you should have a better understanding of

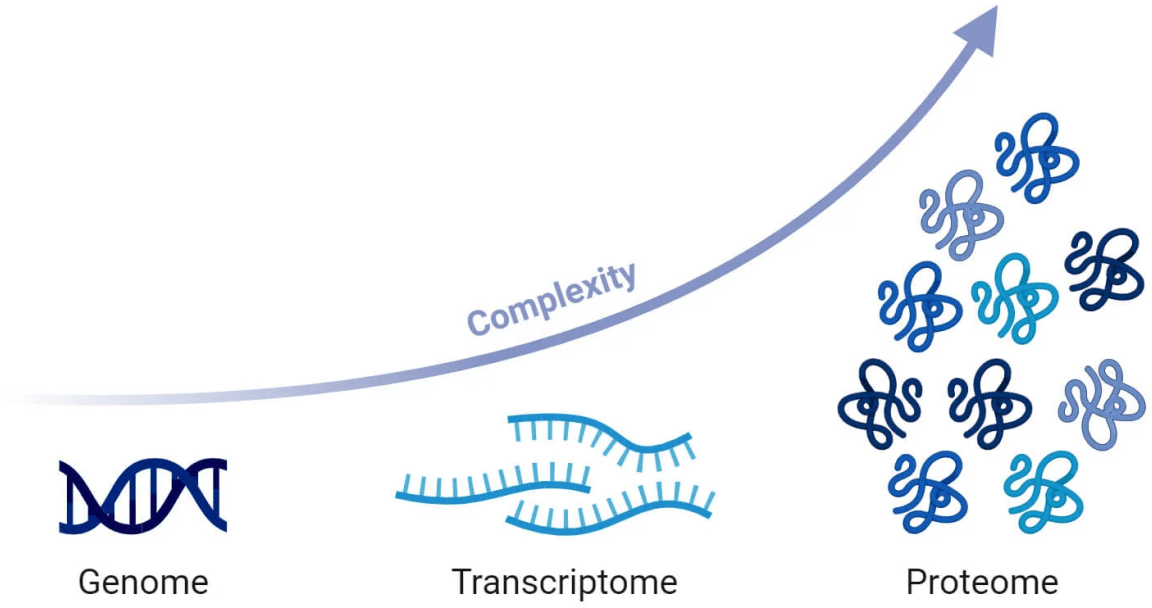


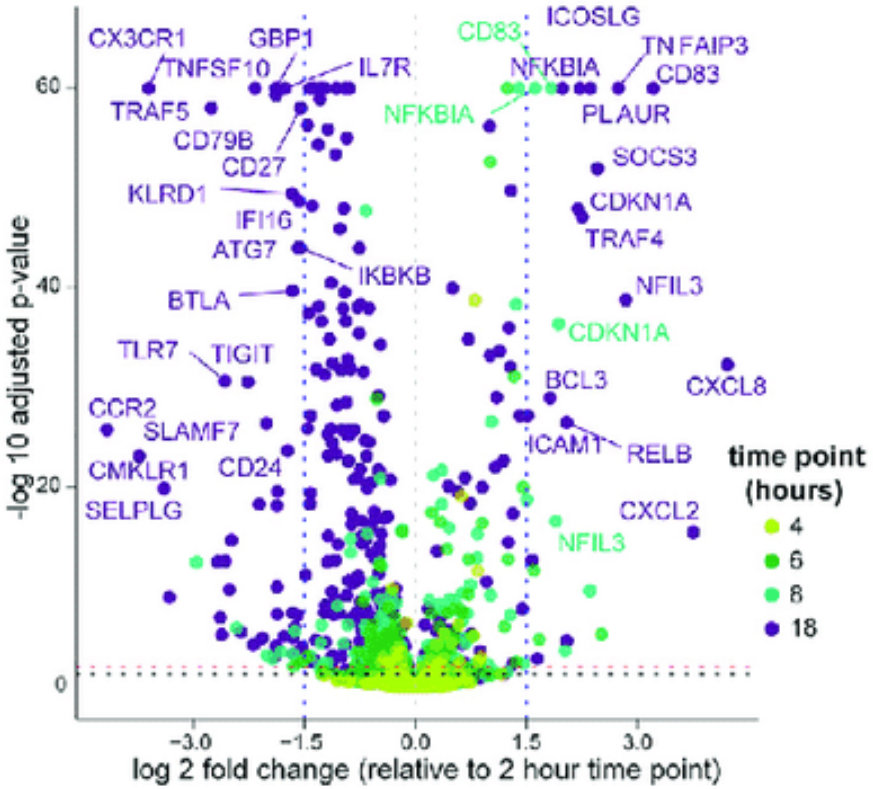
How transcriptomics extends beyond genomics

RNA

Transcriptomics: A real-time microscope

Transcriptomics allows us to see precisely what genes are active at a given moment



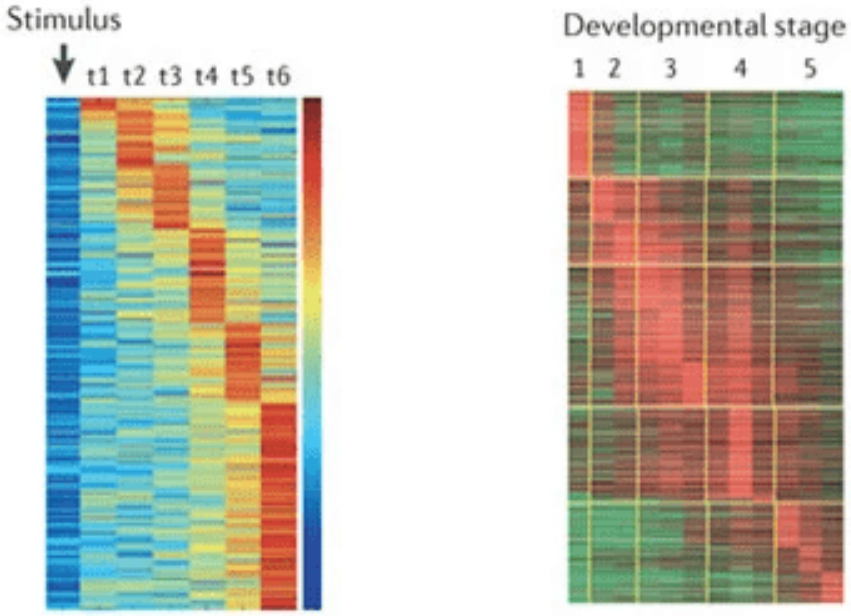


We can see gene expression changes over time

Genomics Provides a Static Blueprint, but Transcriptomics Captures Dynamic Activity

The transcriptome is constantly changing and **captures the cell's response to its environment and internal signals**

- **Environmental conditions:** Cells respond to stress, nutrients, or pathogens by changing gene expression
- **Developmental stage:** The genes active in an embryo differ from those in an adult
- **Cell type:** A neuron will have a different gene expression profile than a liver cell



Allows us to see **which annotated genes are actually being used**

Transcriptomics works with the complete set of RNA transcripts

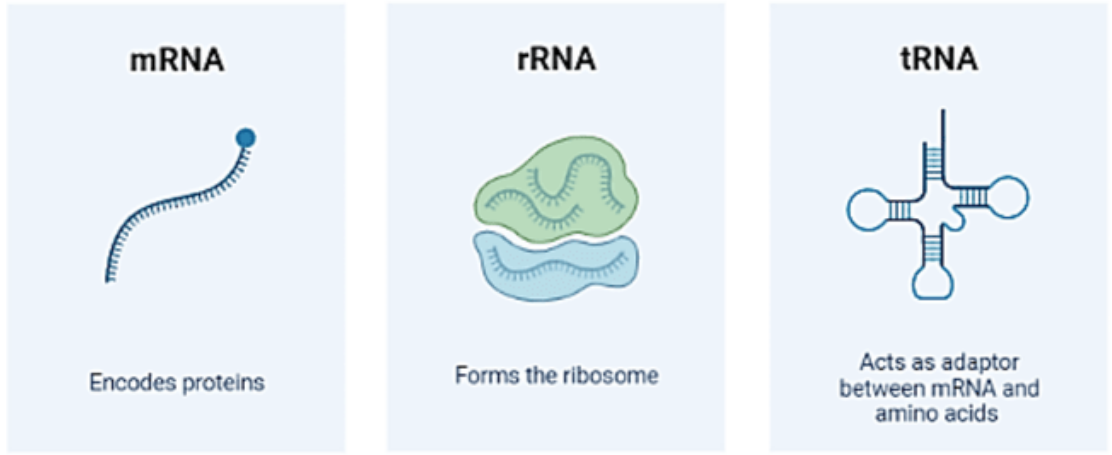
This includes

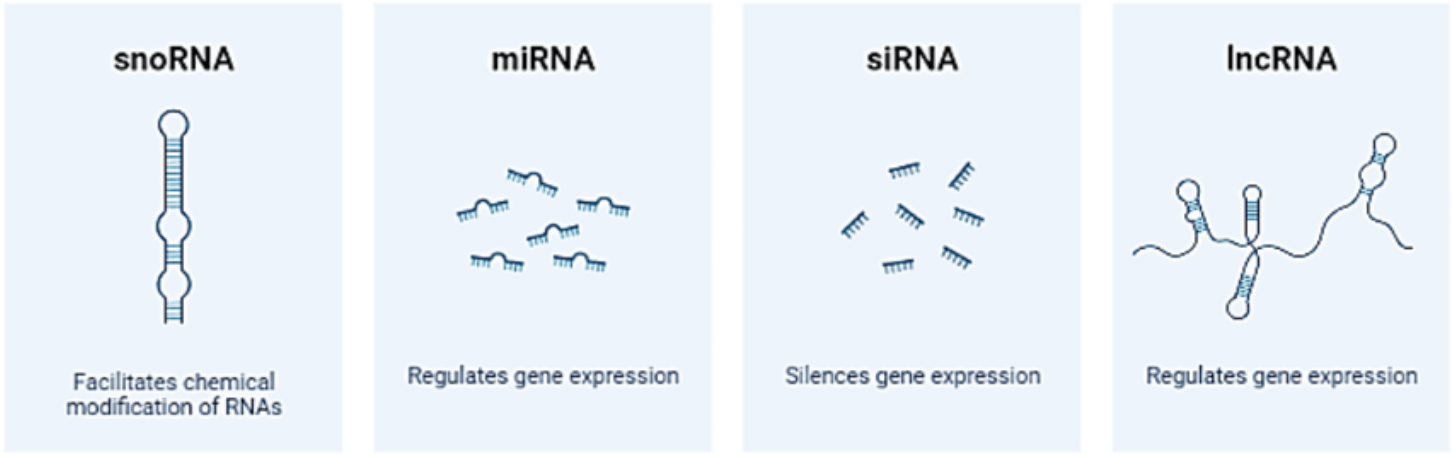
mRNA: instructions for protein synthesis

rRNA: forms part of the ribosome structure

tRNA: helps translate the genetic code into proteins

Non-coding RNAs: play regulatory roles in the cell





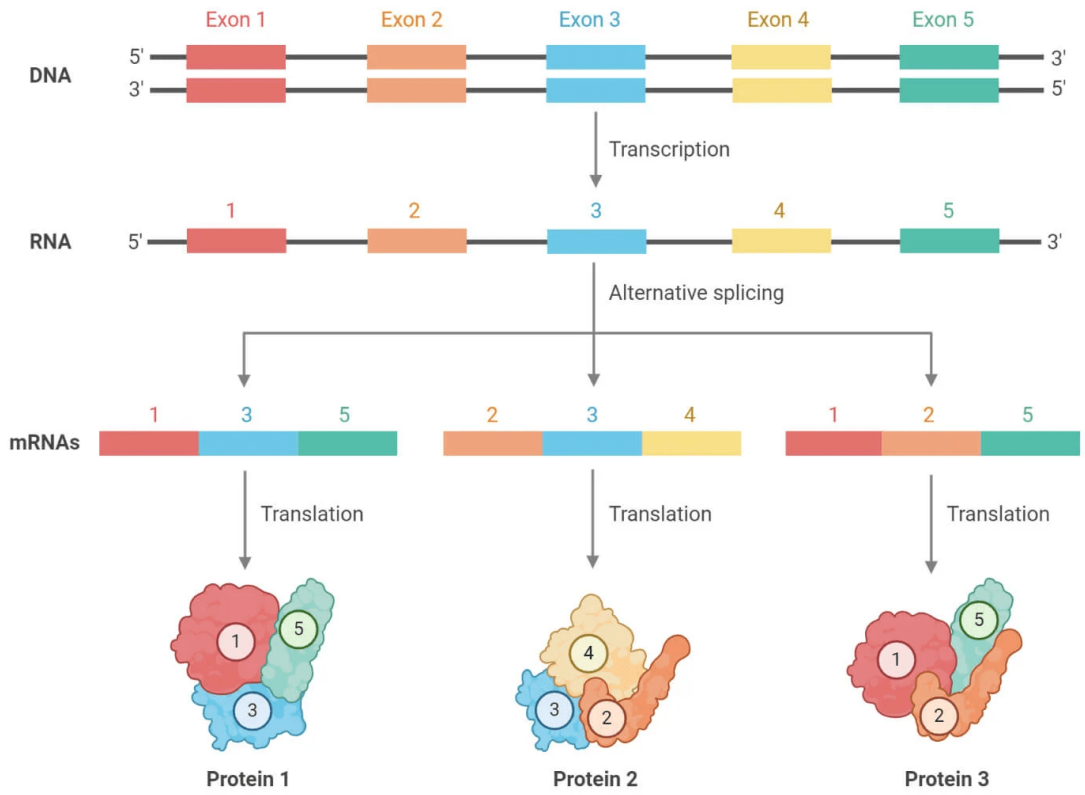
(And more)

Transcriptomics reveals alternative splicing and isoforms

A single gene can produce multiple mRNA transcripts, which we call **isoforms**

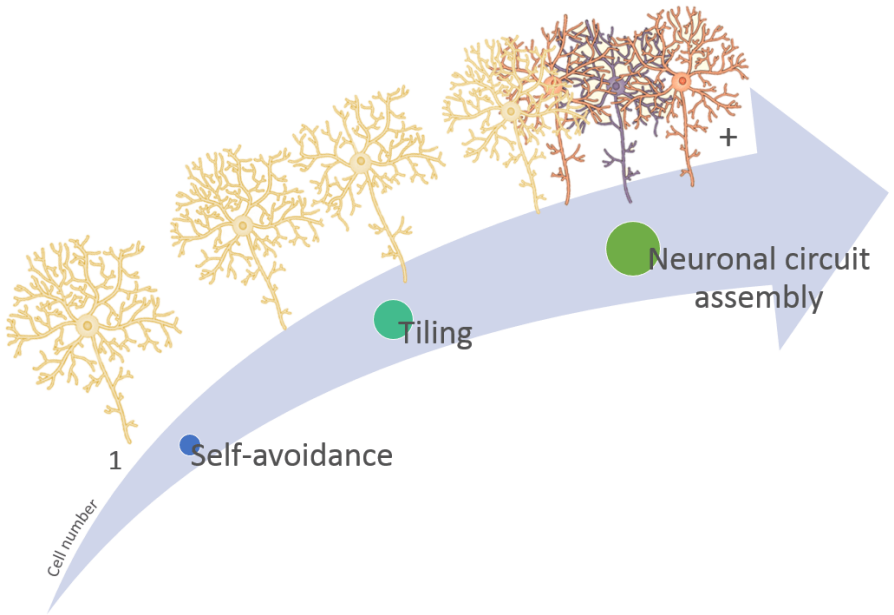
One of the main ways organisms can **increase protein diversity without increasing the number of genes**

It's estimated that over 90% of human genes undergo alternative splicing

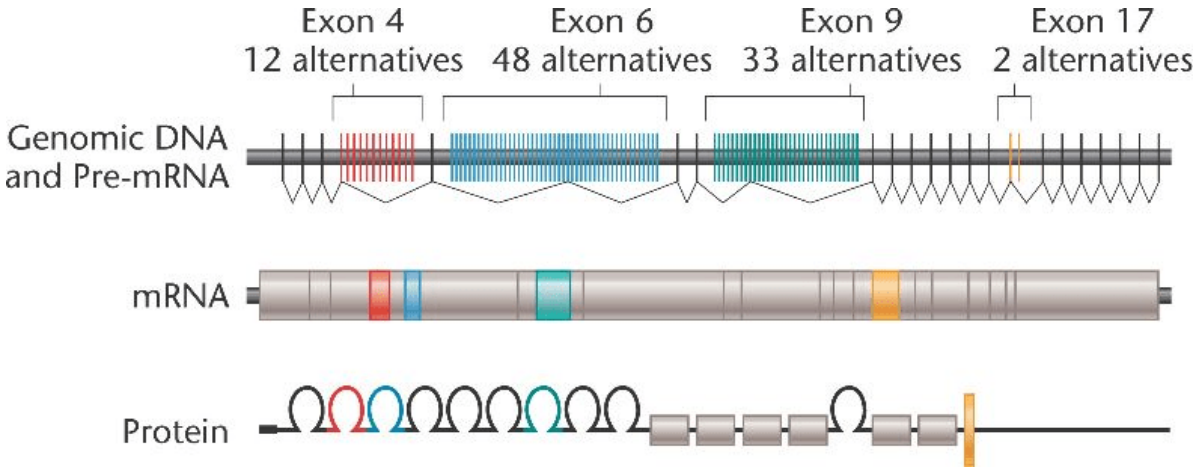


Example: *Dscam* in *Drosophila*

Dscam (Down syndrome cell adhesion molecule) is involved in neural development



Drosophila melanogaster has over 38,000 isoforms from this one gene



Functional insights

Genomics

- Identifies potential functional elements
- Predicts disease risk

Transcriptomics

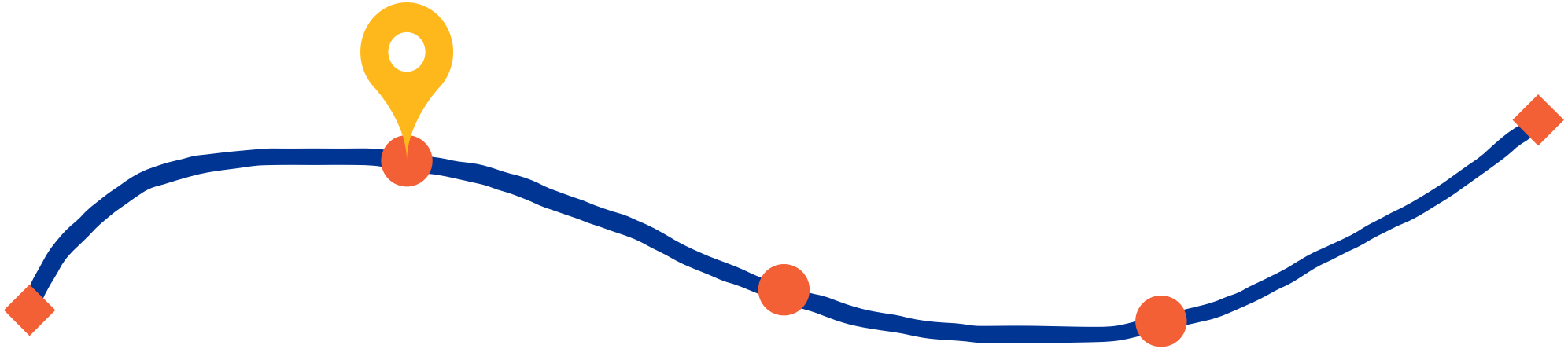
- Reveals which elements are active
- Shows diseases state

Temporal insights

- Requires one-time sampling
- Reveals evolutionary history

- Captures real-time cellular responses

After today, you should have a better understanding of



The role of RNA-seq in modern transcriptomics

Sample collection

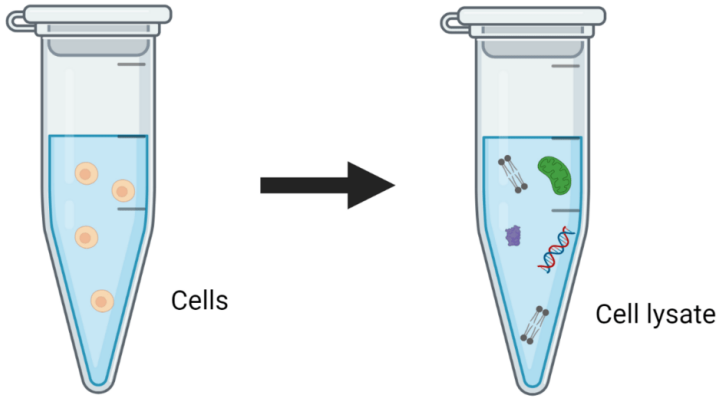
Separate cells from media

The first step is always to centrifuge and separate our cells and media

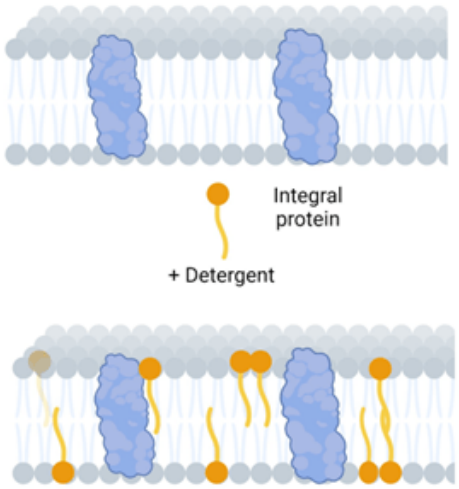
Keep the part that has our **component of interest** (RNA)

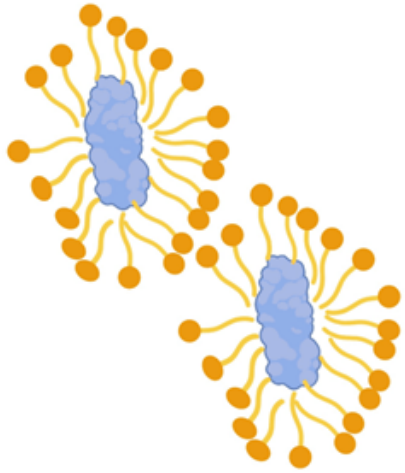
Great! We have our cells, but how can we extract our RNA?

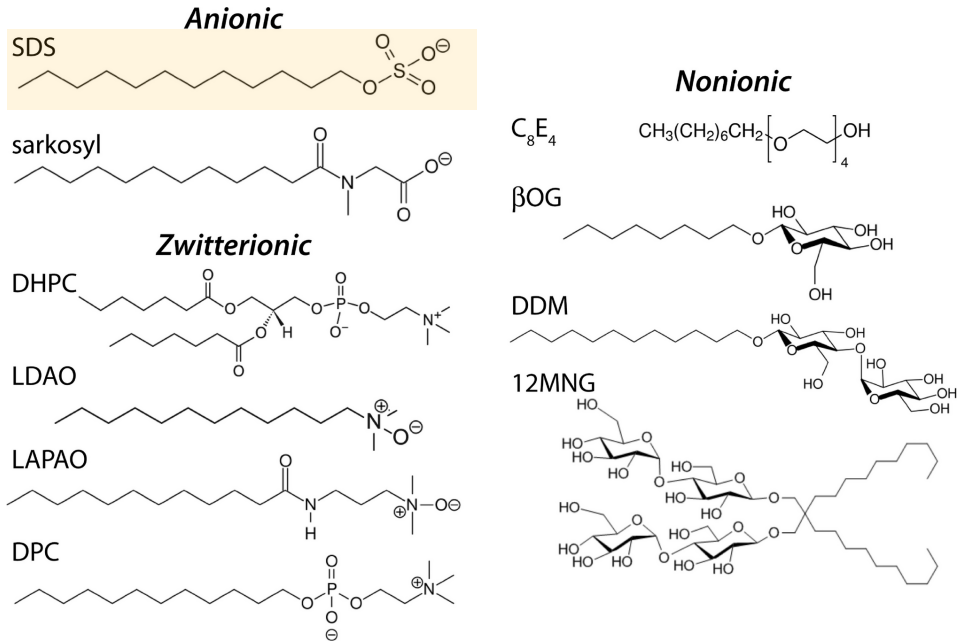
We break open our cells by lysing them



Chemical lysis destabilizes the lipid bilayer and denatures proteins



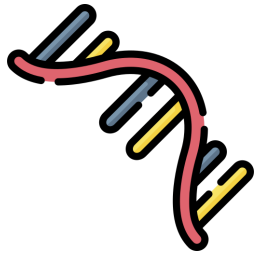




Surfactants have a hydrophilic head and hydrophobic tail

Phenol-chloroform extraction exploits solubility and density differences

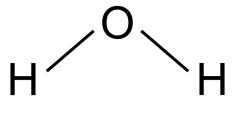
DNA



RNA

Protein

Lipids

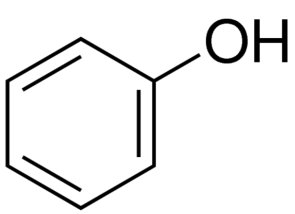


Water

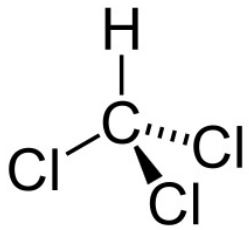
Phosphate backbone (negative charged)

Denatures and aggregates at interface

Nonpolar



Phenol



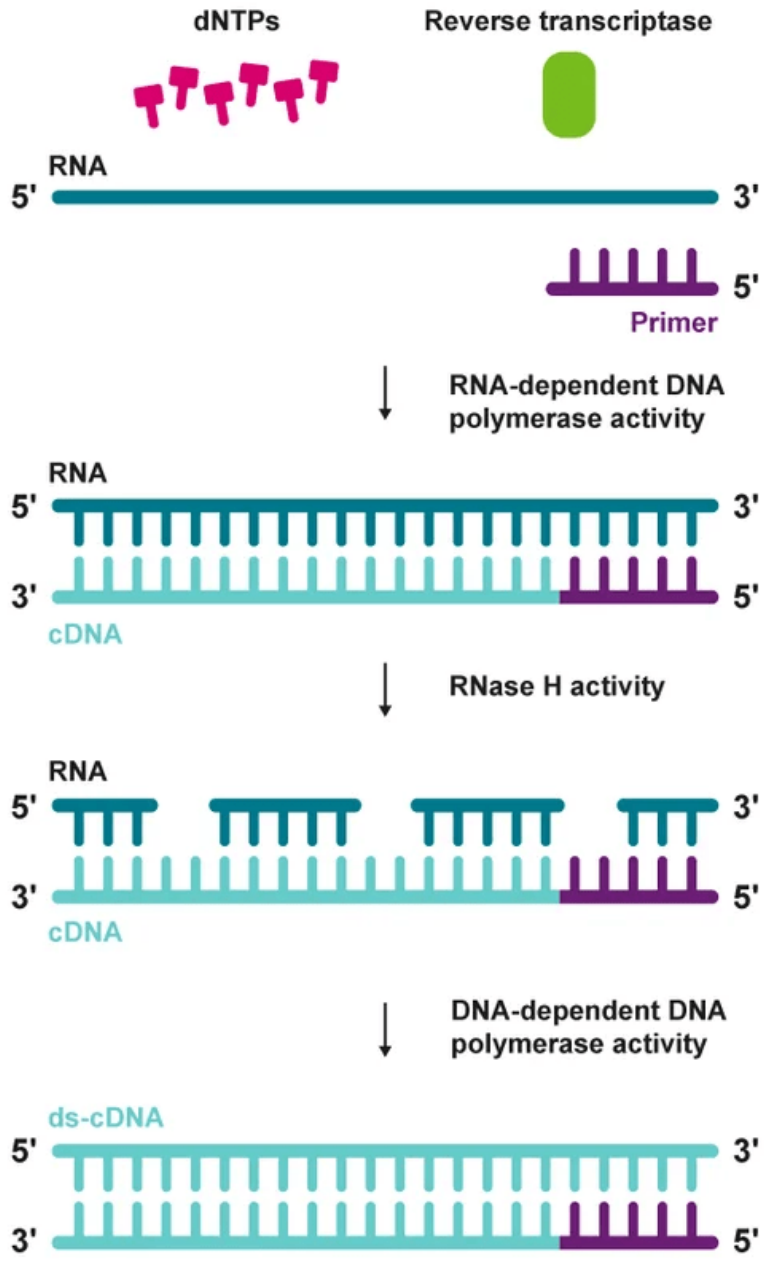
Chloroform

Collecting our aqueous phase selects only DNA and RNA

Reverse transcription introduces unique challenges

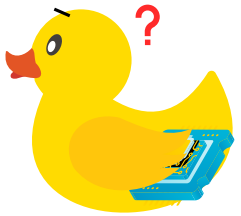
RNA is converted to cDNA using **reverse transcriptase**

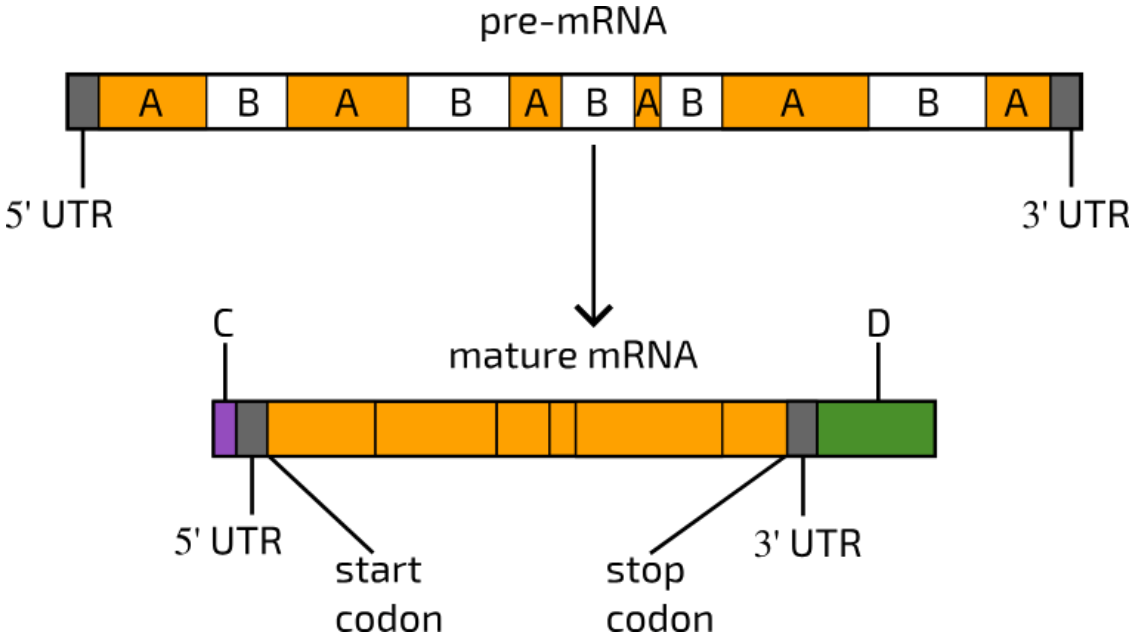
- Random or oligo(dT) primers influence transcript representation
- Second-strand synthesis method can preserve strand information



mRNA enrichment focuses sequencing on protein-coding transcripts

How could we filter our sample for only mRNA?



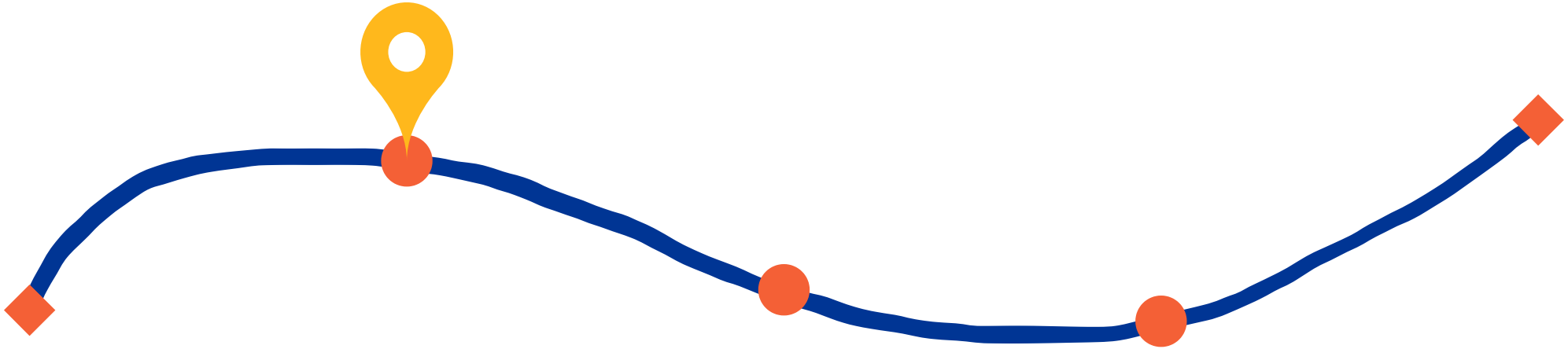


Poly(A) selection captures mature mRNAs

Enrichment method affects

- Gene expression measurements
- Detection of non-coding RNAs
- Identification of immature transcripts

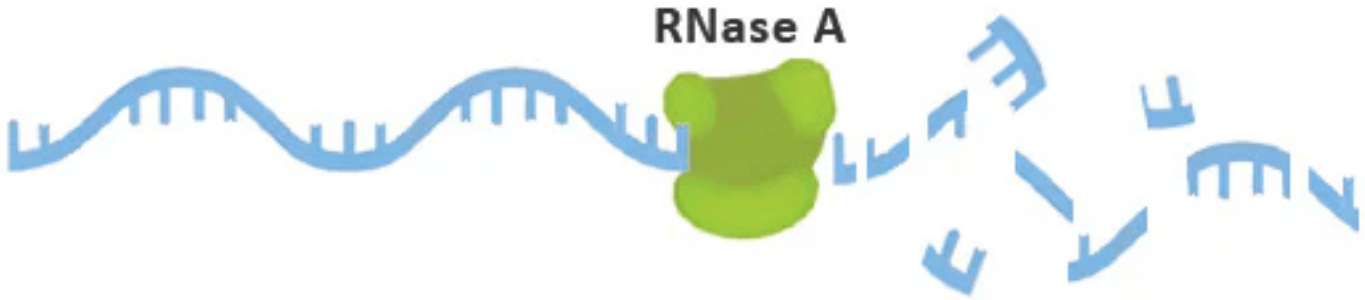
After today, you should have a better understanding of



The role of RNA-seq in modern transcriptomics

Sample quality

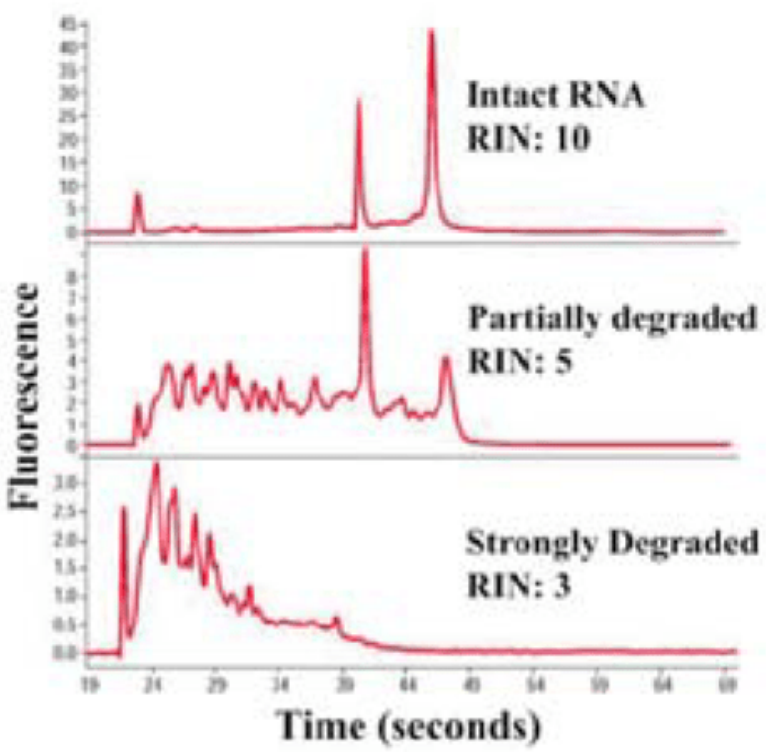
RNase starts degrading RNA rapidly

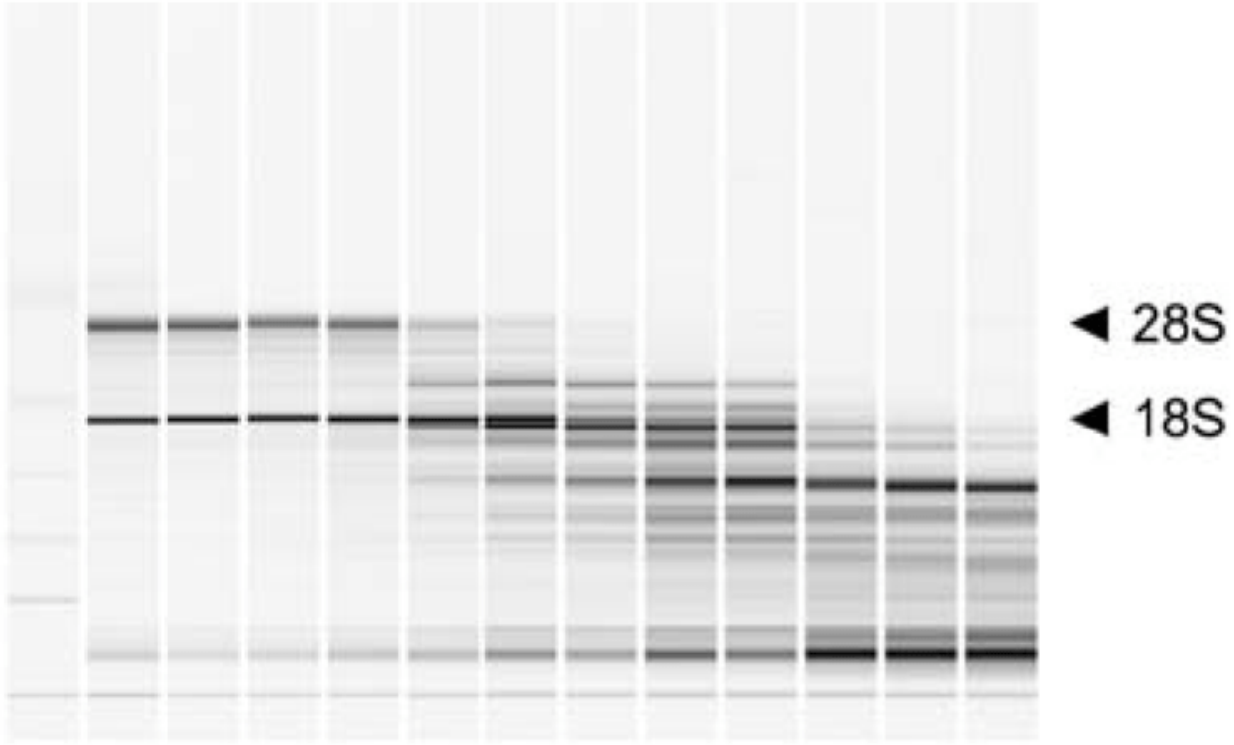


RNA quality is critical for successful sequencing

Assess RNA integrity (RNA Integrity Number)

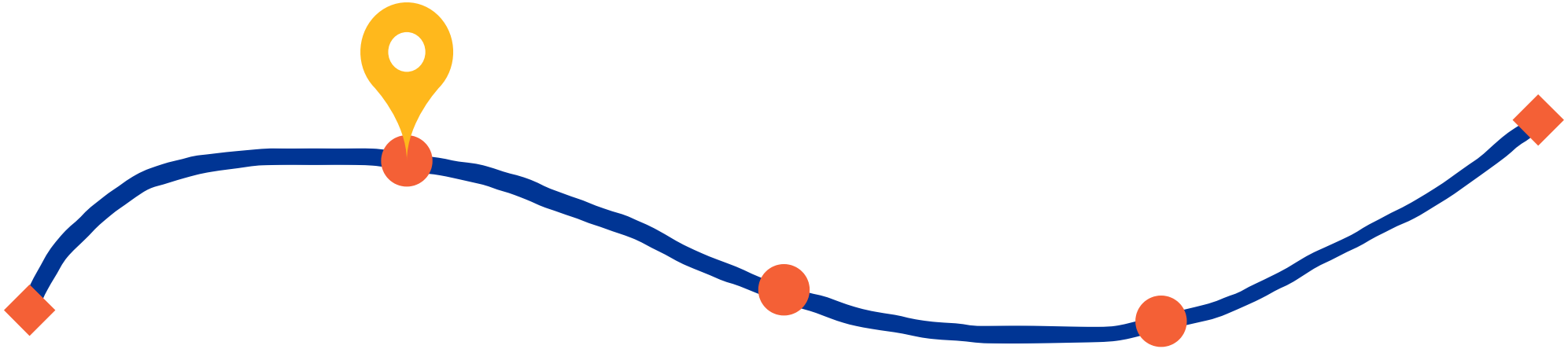
- rRNA makes up a large (~85%) of our RNA
- Based on the ratio of 28S and 18S rRNA vs. all RNA





High RIN **Low RIN**

After today, you should have a better understanding of

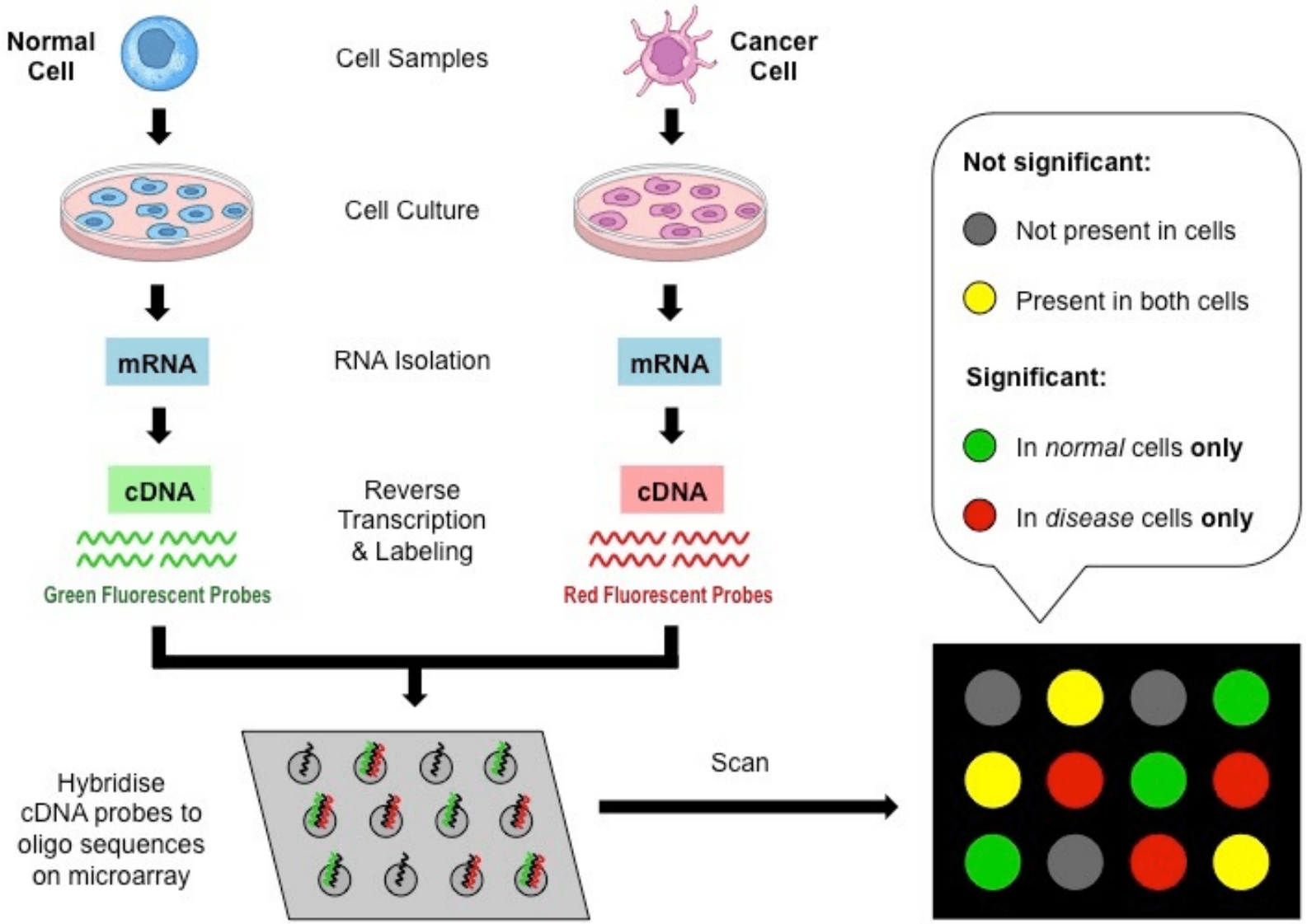


The role of RNA-seq in modern transcriptomics

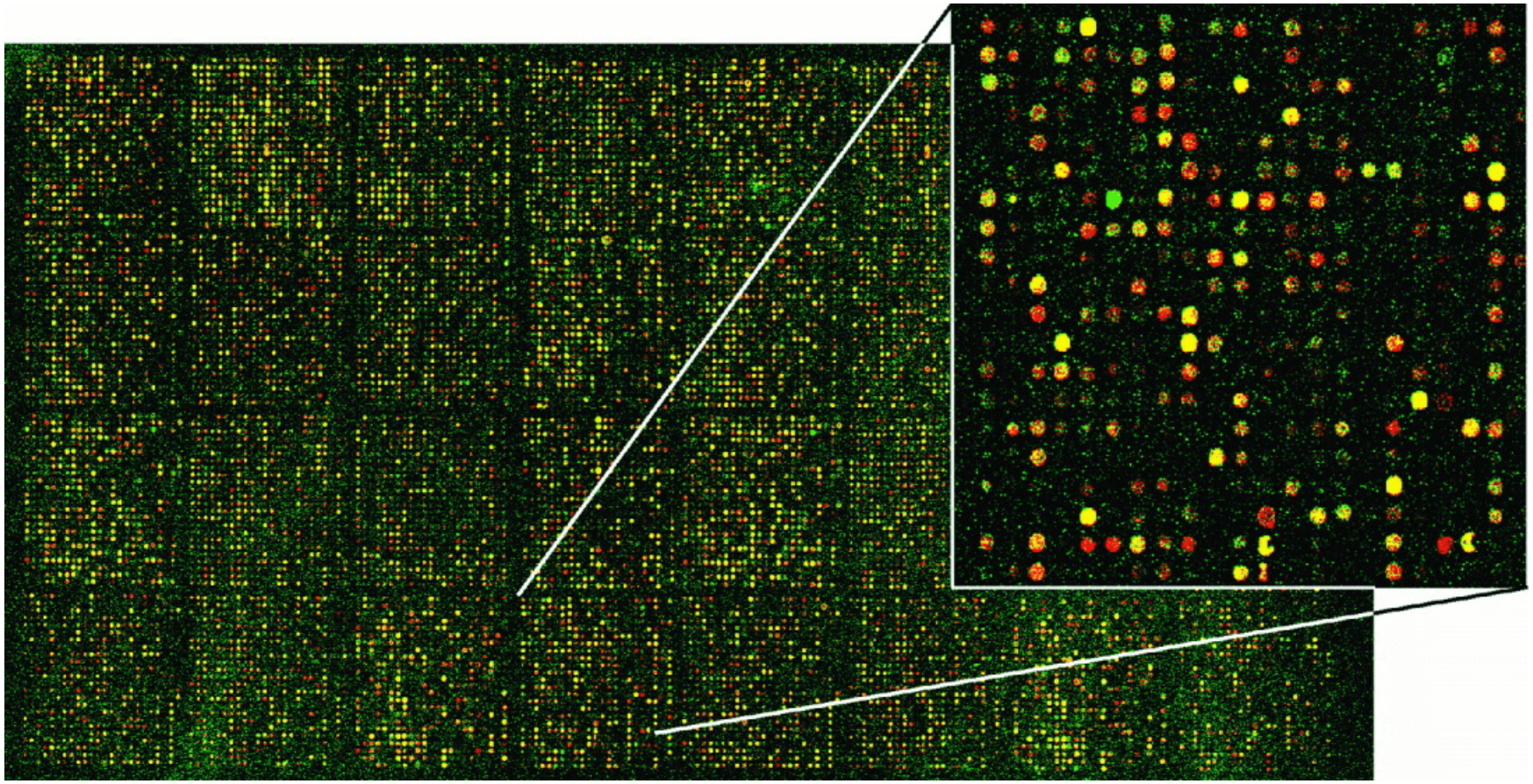
Microarrays

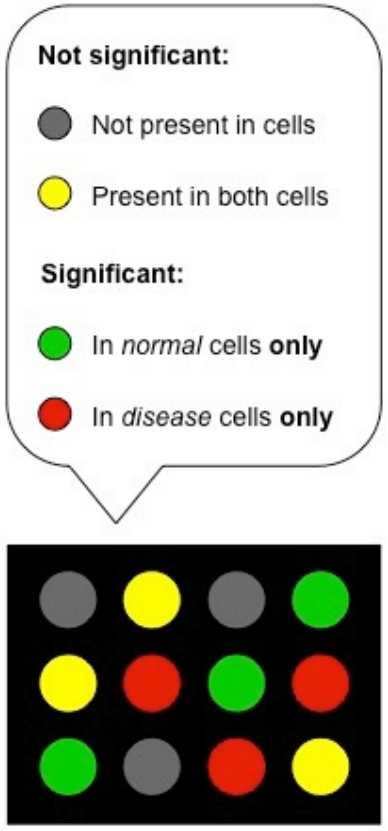
Once upon a time, we had **microarrays**

(Now obsolete)



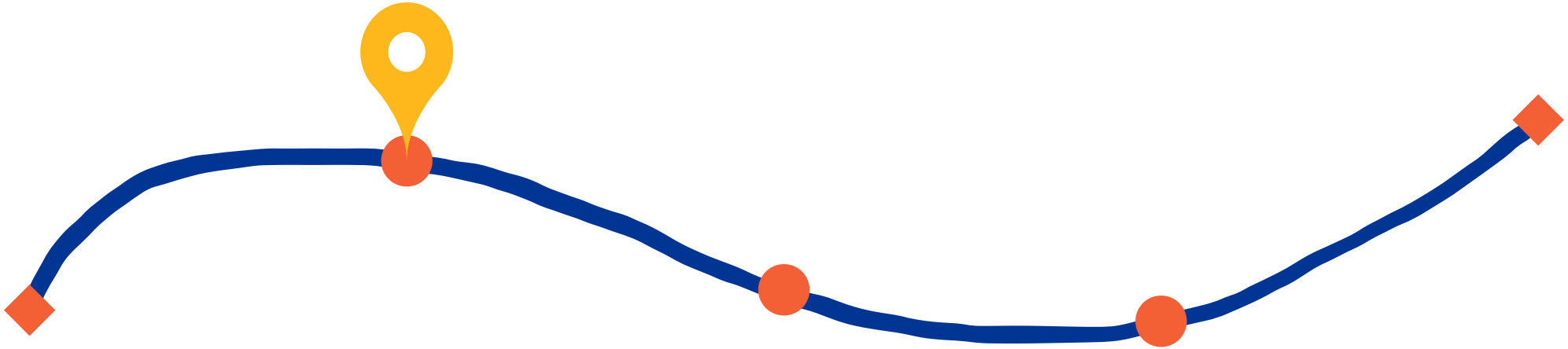
Microarrays have some caveats





- **Limited to known sequences:** Can only detect pre-defined sequences
- **Cross-hybridization:** Similar sequences may cause false positives
- **Limited dynamic range:** May miss very low or high abundance transcripts
- **Normalization challenges:** Complex process, potential for bias

After today, you should have a better understanding of

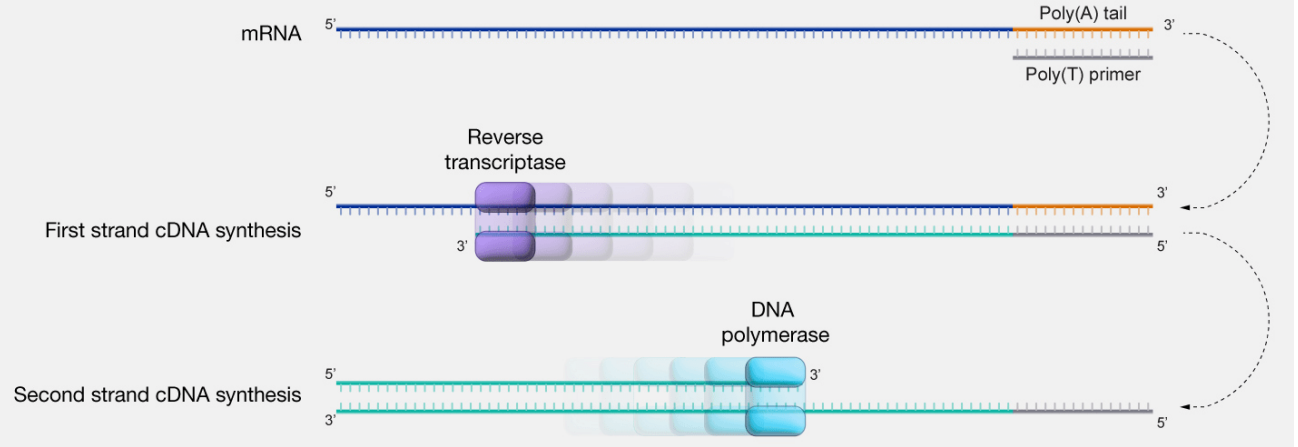


The role of RNA-seq in modern transcriptomics

RNA-seq

RNA sequencing changed the game

Now we just sequence the cDNA

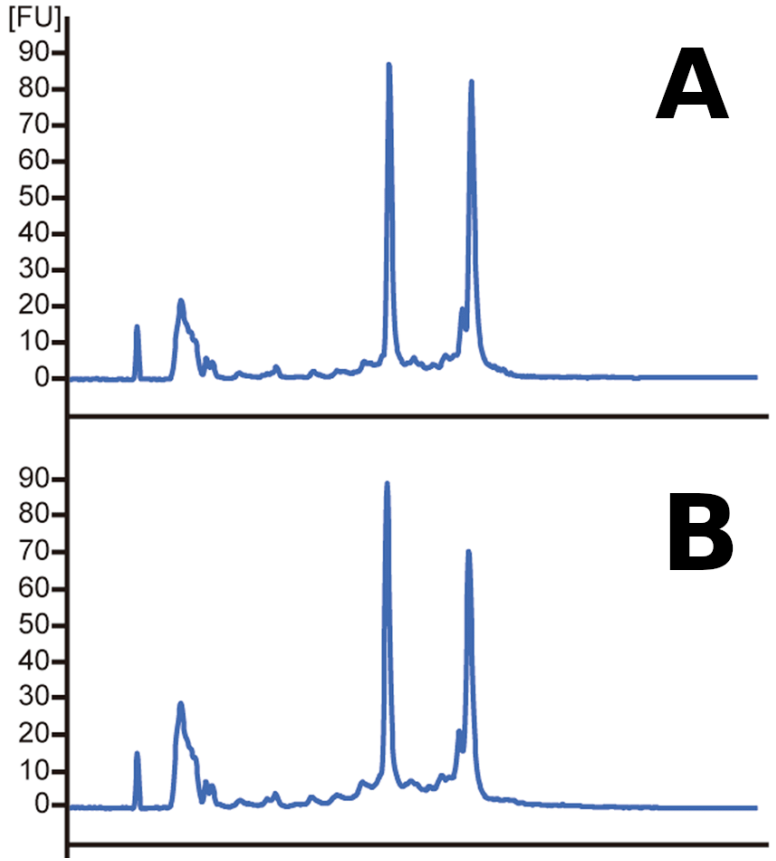


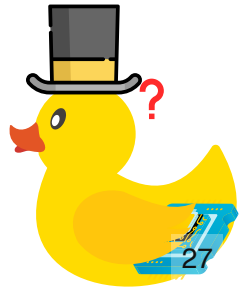
- RNA-seq doesn't require prior knowledge of sequences
- Enables discovery of novel transcripts and isoforms
- Provides absolute quantification rather than relative concentration

TopHat questions

What is the primary advantage of RNA-seq over microarray technology?

Which sample has a higher RIN?





Before the next class, you should

Lecture 06A:

Read mapping - Foundations

Lecture 06B:

Read mapping - Methodology

Today

Thursday

- Work on P01D (due Feb 14)
- Study for Quiz 02 (on Feb 17)