

BIOL 343

Applied Bioinformatics I

Intro to RNA sequencing

Dr. Nic Wheeler

Introduction



Class introductions

Name (and pronouns if desired), year of study, major

Why you took this class

One thing you hope to learn in this class

One thing you like about yourself



Course introduction

Welcome to the University of Wisconsin-Eau Claire
Bioinformatics Core Facility!

Links: GitHub organization profile and BIOL 343 repository



Weekly agenda

Lecture Monday

Content delivery on a particular aspect of RNA sequencing

Tool Tuesday

Introduction of RNA-seq bioinformatics tools using example data and Jupyter notebook

Dry Lab Thursday

In-class homework using Tuesday's tools on data selected by the student

Semester agenda

Weeks 1-8

RNA-seq bootcamp on
published data

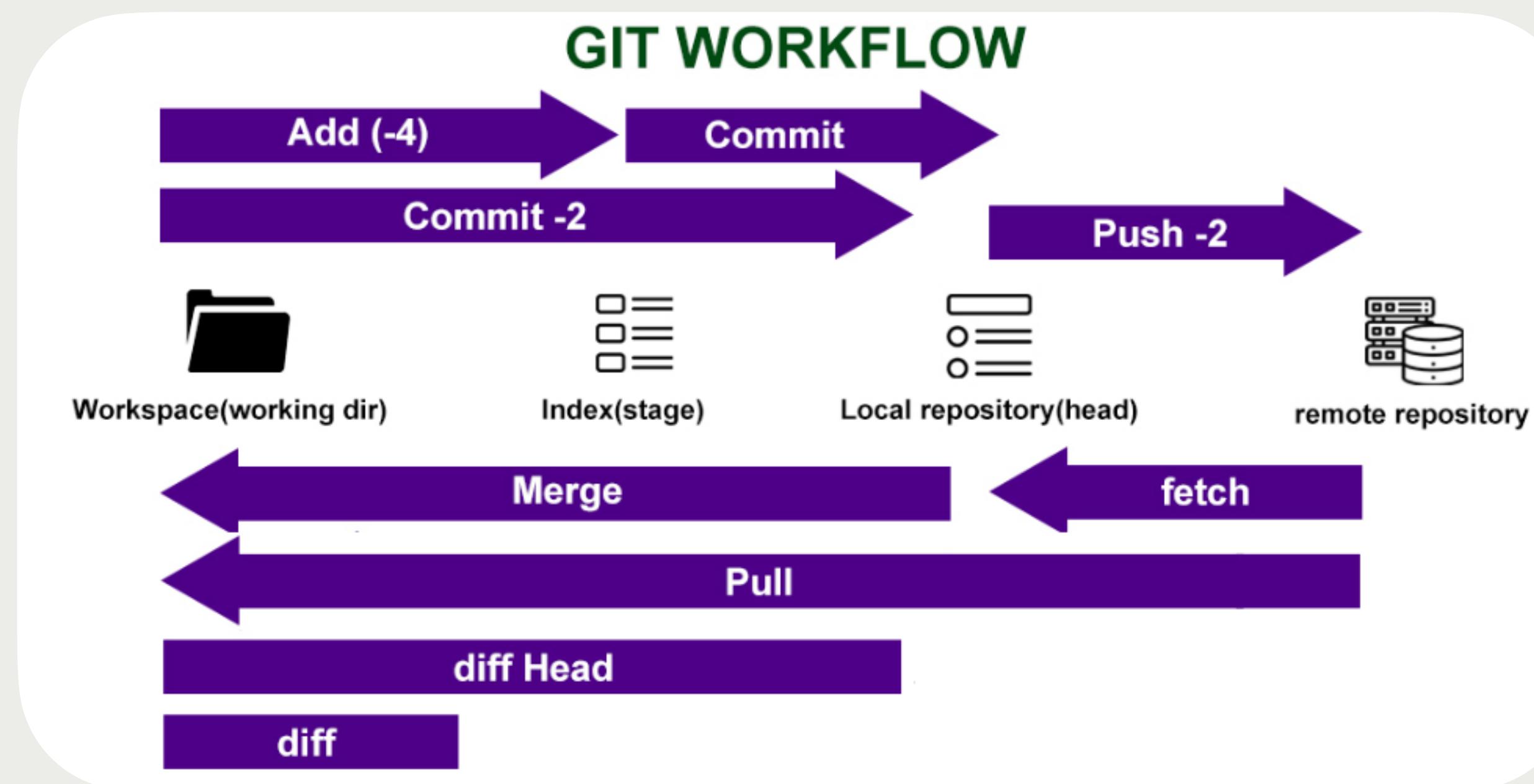
Weeks 9-14

Independent project on novel
data

Syllabus quick hitters

1. Grading - specs grading
 1. Attendance
 2. Drop-in hours
 3. GitHub

GitHub and version control



GitHub and version control

1. Make GitHub profile
2. Join UWEC Bioinformatics organization
3. Setup SSH key for repo cloning
 1. Generate key on BOSE (no passphrase)
 2. Add key to GitHub account

GitHub and version control

1. Clone BIOL343 repo into BOSE home dir and move into it: `cd BIOL343`
2. Make a new branch with your name: `git checkout -b {name}`
3. Make a new file: `nano temp.txt`
4. Edit the file
5. Stage: `git add .`
6. Commit: `git commit -m "{message}"`
7. Push: `git push`
8. Switch to main: `git checkout main`
9. Pull: `git pull`
10. Switch to your branch: `git checkout {name}`
11. Merge: `git merge main`

Questions?

Break if time...

Learning Objectives

You will be able to:

1. Explain why RNA quantification is important to biology research
2. Identify real-world uses of RNA sequencing in biology research and biomedical applications and explain in detail the use-case for the *Winners vs. Losers* study
3. Describe the technology behind historical approaches to RNA quantification and list their disadvantage(s)

Applied bioinformatics

Using sequences to tell us something about biology

Bioinformatics is the computational analysis of sequence data

- DNA
 - Species genetic differences
 - Population/individual genetic differences
 - Causes of genetic diseases
 - ***Quality (the actual sequence and variation in it)***
- RNA (typically mRNA)
 - Cellular function
 - Organismal/disease status
 - ***Quantity***

Tophat Questions

Join code: 684418

Types of RNA

mRNA

Messenger

Will be translated to protein

Alternatively spliced

Regulated (degradation)

Highly dynamic quantities

rRNA

Ribosomal

Never translated

Creates ribosomes (along
with proteins)

Highly stable

Highly expressed

tRNA

Transfer

Never translated

Binds to amino acids

Recognizes codons during
translation

Highly stable

Lowly expressed

Small RNA

Many types (miRNA,
snRNA, snoRNA, piRNA,
etc.)

Often regulate mRNA levels

Types of RNA

mRNA

Messenger
Will be translated to protein
Alternatively spliced
Regulated (degradation)
Highly dynamic quantities

rRNA

Ribosomal
Never translated
Creates ribosomes (along
with proteins)
Highly stable
Highly expressed

tRNA

Transfer
Never translated
Binds to amino acids
Recognizes codons during
translation
Highly stable
Lowly expressed

Small RNA

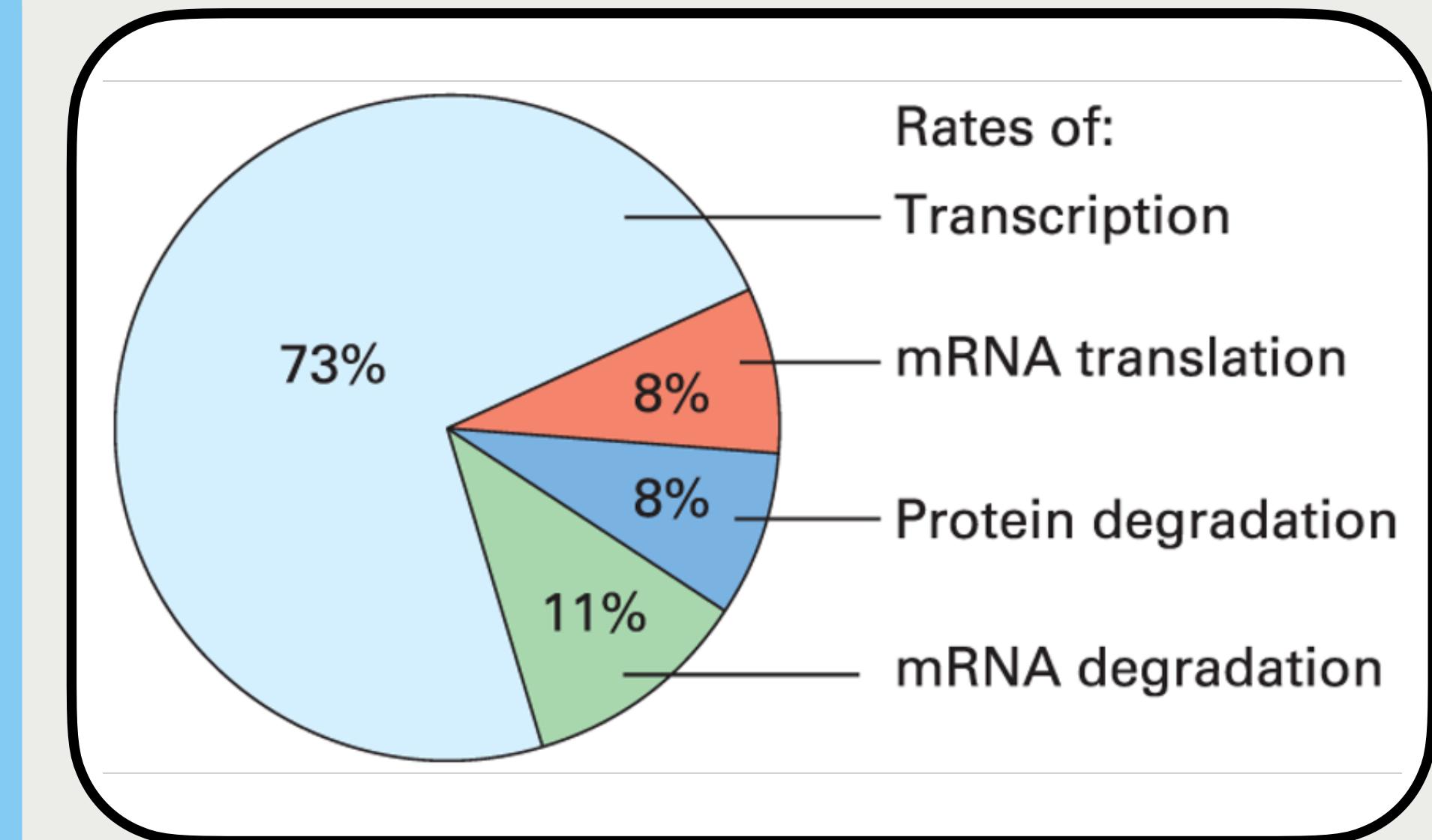
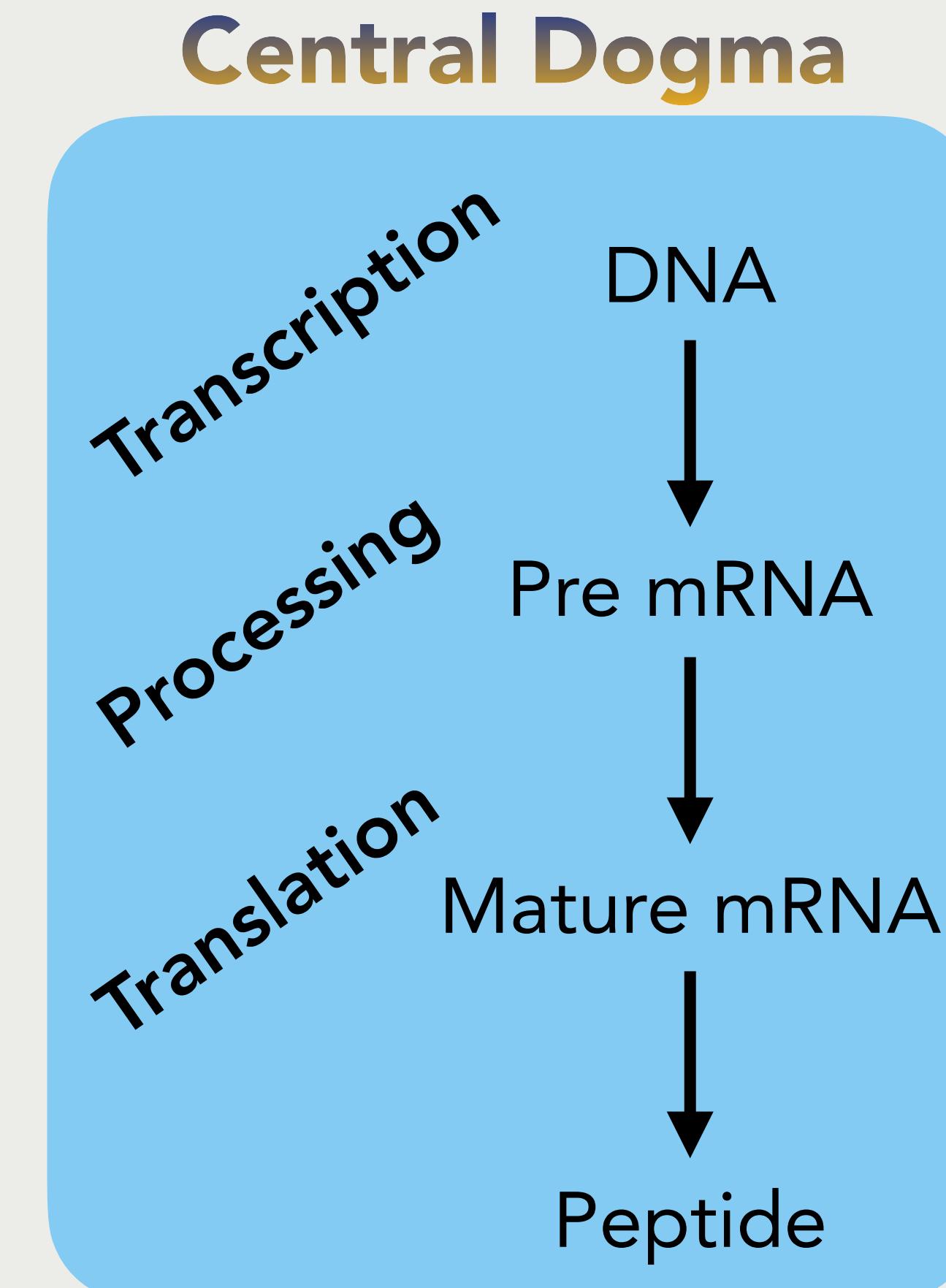
Many types (miRNA,
snRNA, snoRNA, piRNA,
etc.)
Often regulate mRNA levels

Sometimes...

Why quantify mRNA expression?

mRNA quantity is correlated with protein quantity

- Differential protein expression determines cellular function
- Four opportunities for regulation
 1. Transcription of mRNA
 2. Degradation of mRNA
 3. Translation of protein
 4. Degradation of protein



Why quantify mRNA expression?

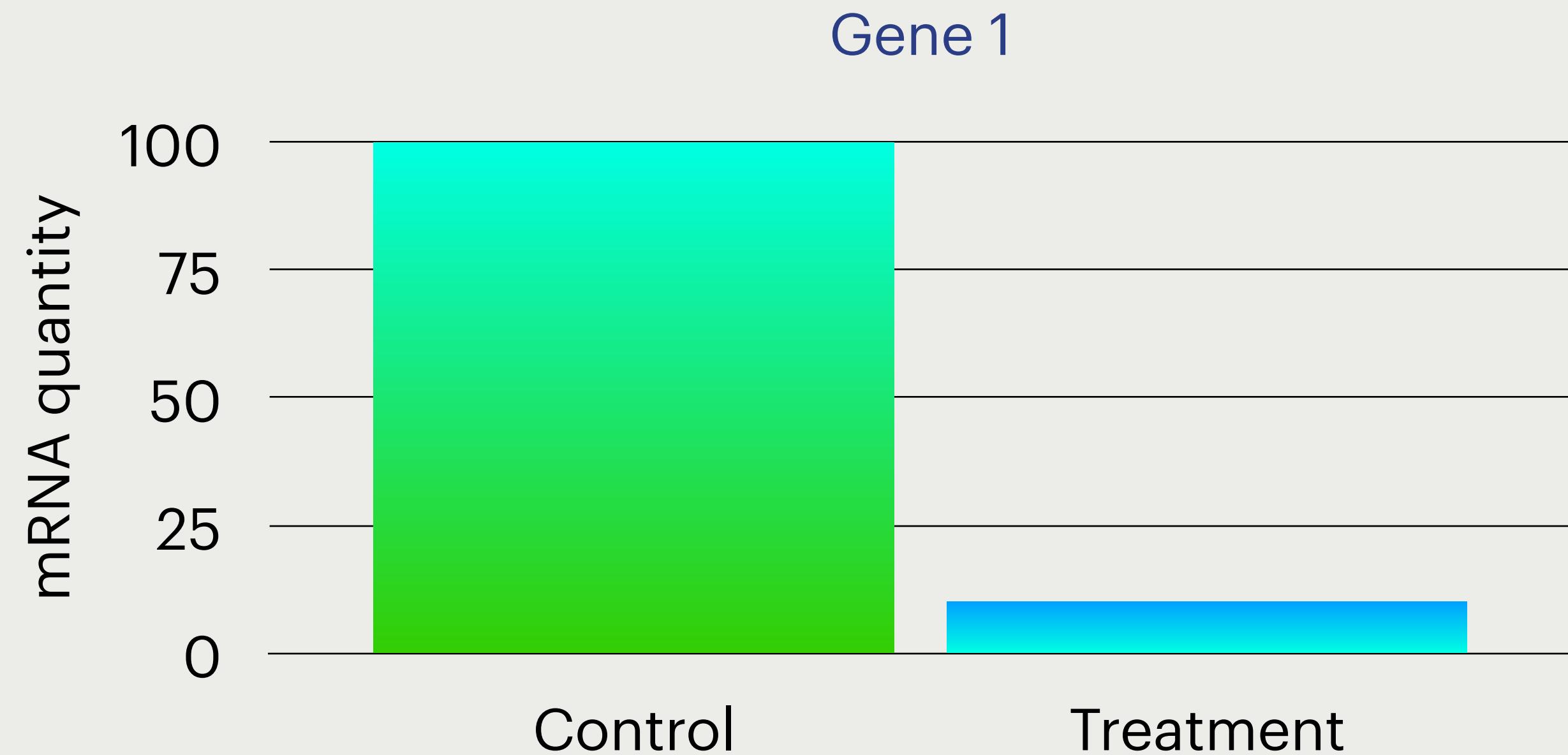
mRNA quantity is indicative of cellular function or status

For example...

Suppose an animal has 1 protein-coding gene:

Gene 1 - living

What does the following data tell you about the function and status of the cell?



Why quantify mRNA expression?

mRNA quantity is indicative of cellular function or status

For example...

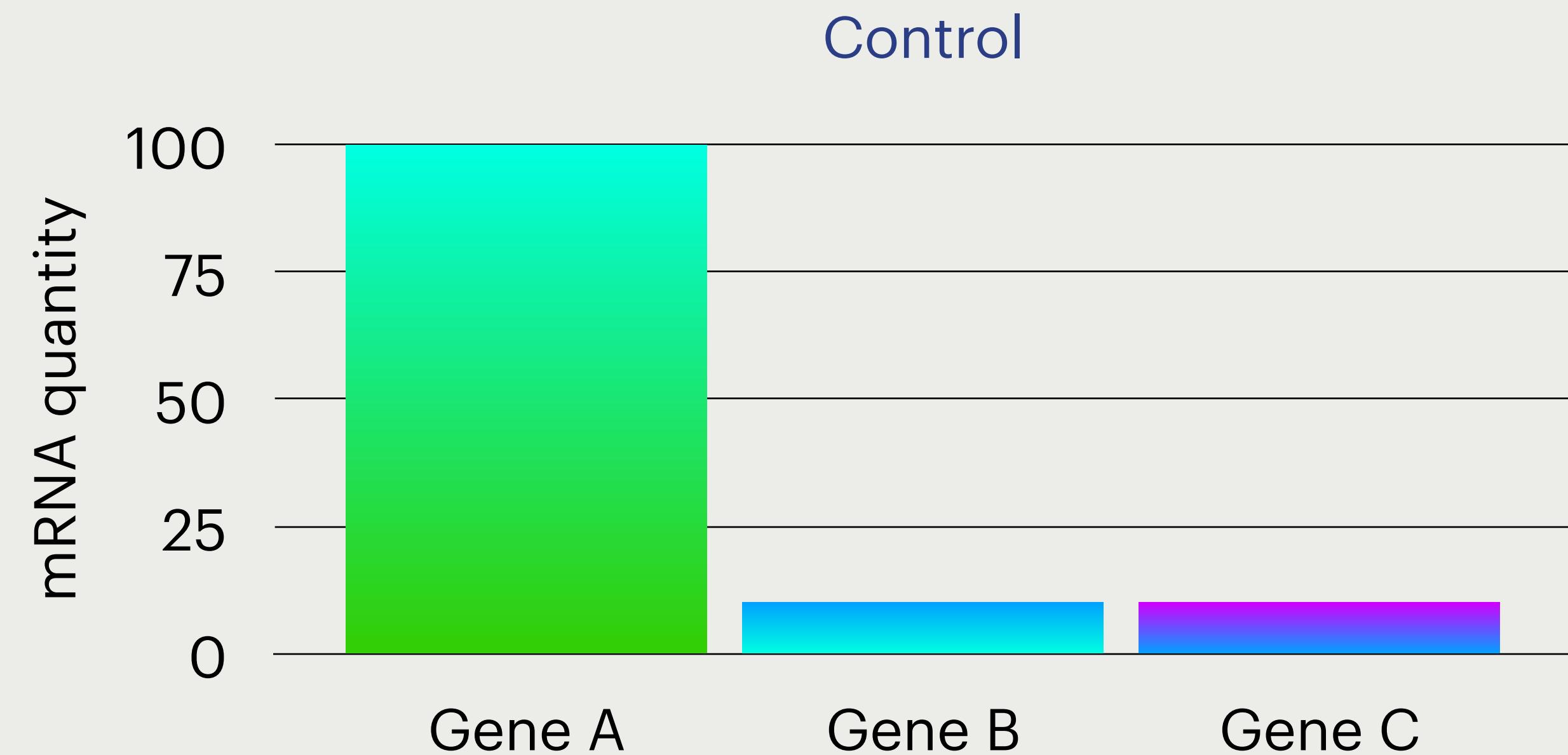
Suppose an animal has 3 protein-coding genes:

Gene A - metabolism

Gene B - cell division

Gene C - immunity

What does the following data tell you about the function and status of the cell?



Why quantify mRNA expression?

mRNA quantity is indicative of cellular function or status

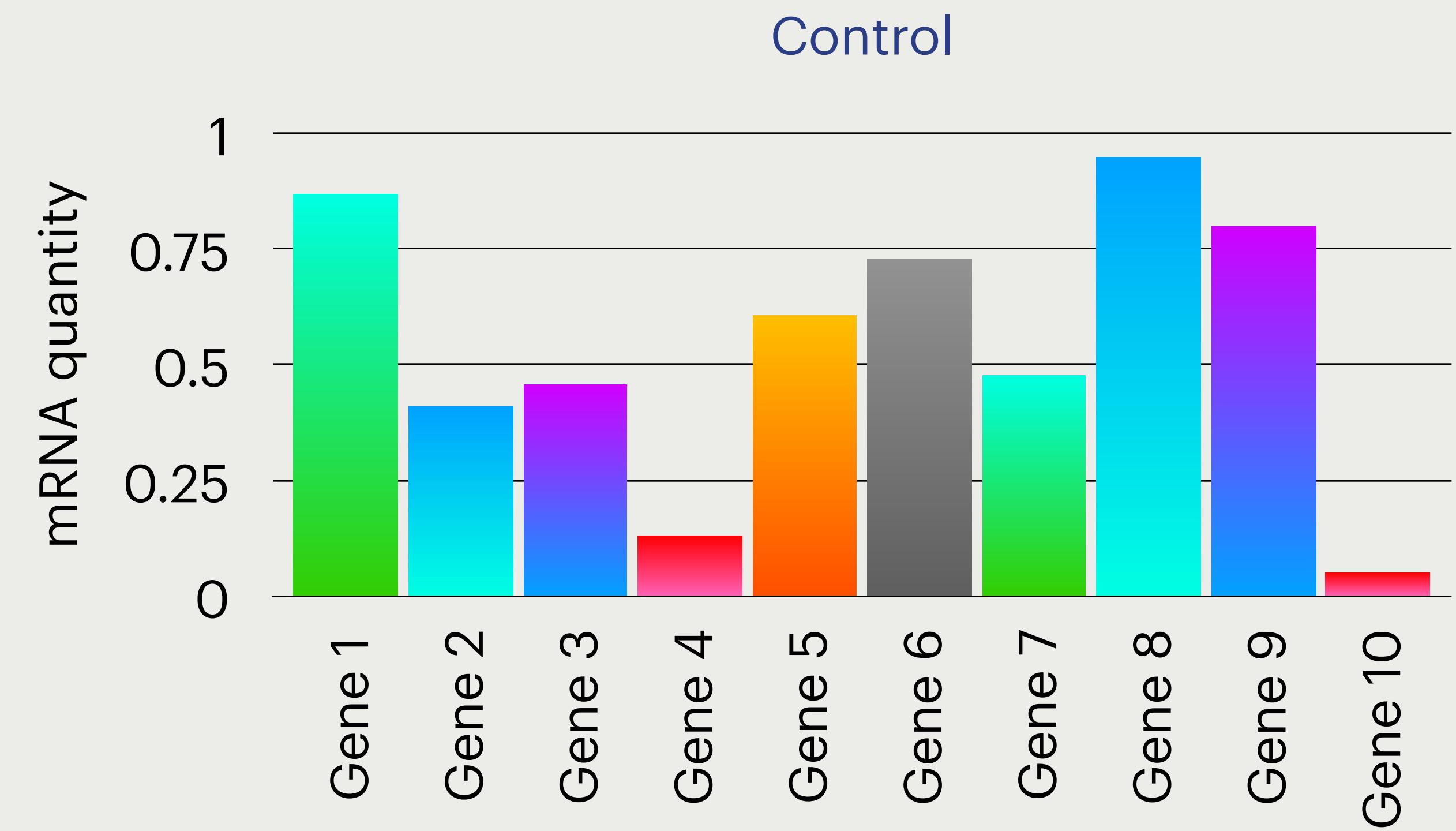
For example...

Suppose an animal has 10 protein-coding genes:

Genes 1-5 - happiness

Genes 6-10 - sadness

What does the following data tell you about the function and status of the cell?



Why quantify mRNA expression?

mRNA quantity is indicative of cellular function or status

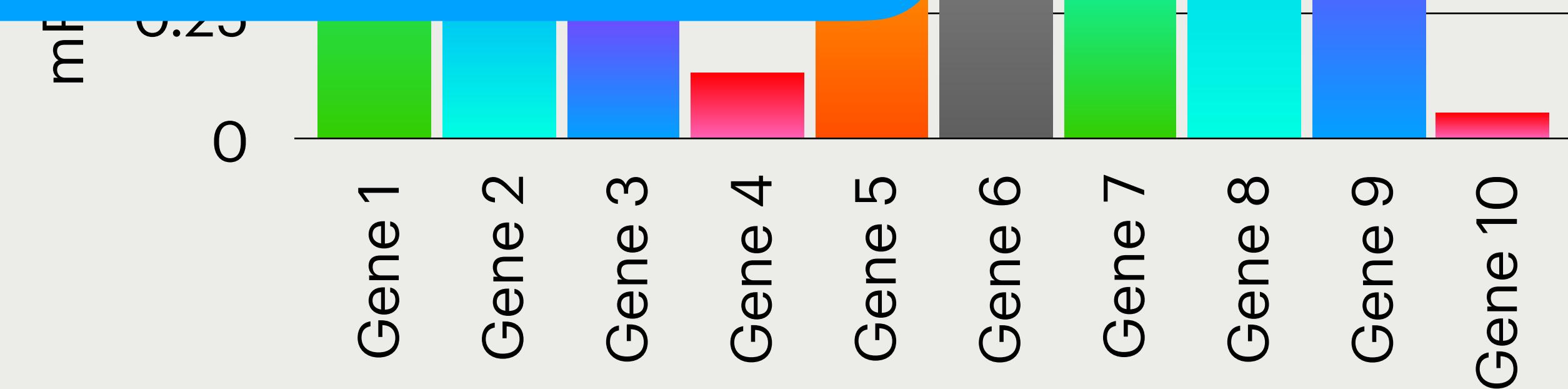
For example...

Suppose an animal has 15,000 protein-coding genes.

Genes 1-5 - have high mRNA levels

Genes 6-10 - same

What about when an animal has 15,000 protein-coding genes?



Global analysis of RNA expression: *transcriptomics*

Transcriptome analysis is a powerful tool for comparing molecular biology

- Real-world examples of transcriptomics use-cases:
 - Identify genes that are highly/lowly expressed in metastatic tumors in comparison to non-metastatic tumors or healthy tissue
 - Identify marker genes for a disease state (i.e., Alzheimer)
 - Analyze changes in gene expression during embryonic development or wound healing
 - Prioritize genes expressed in the etiological agents of infectious diseases for study as potential druggable targets
 - Compare/contrast expression differences between different but related populations of animals
 - What else??

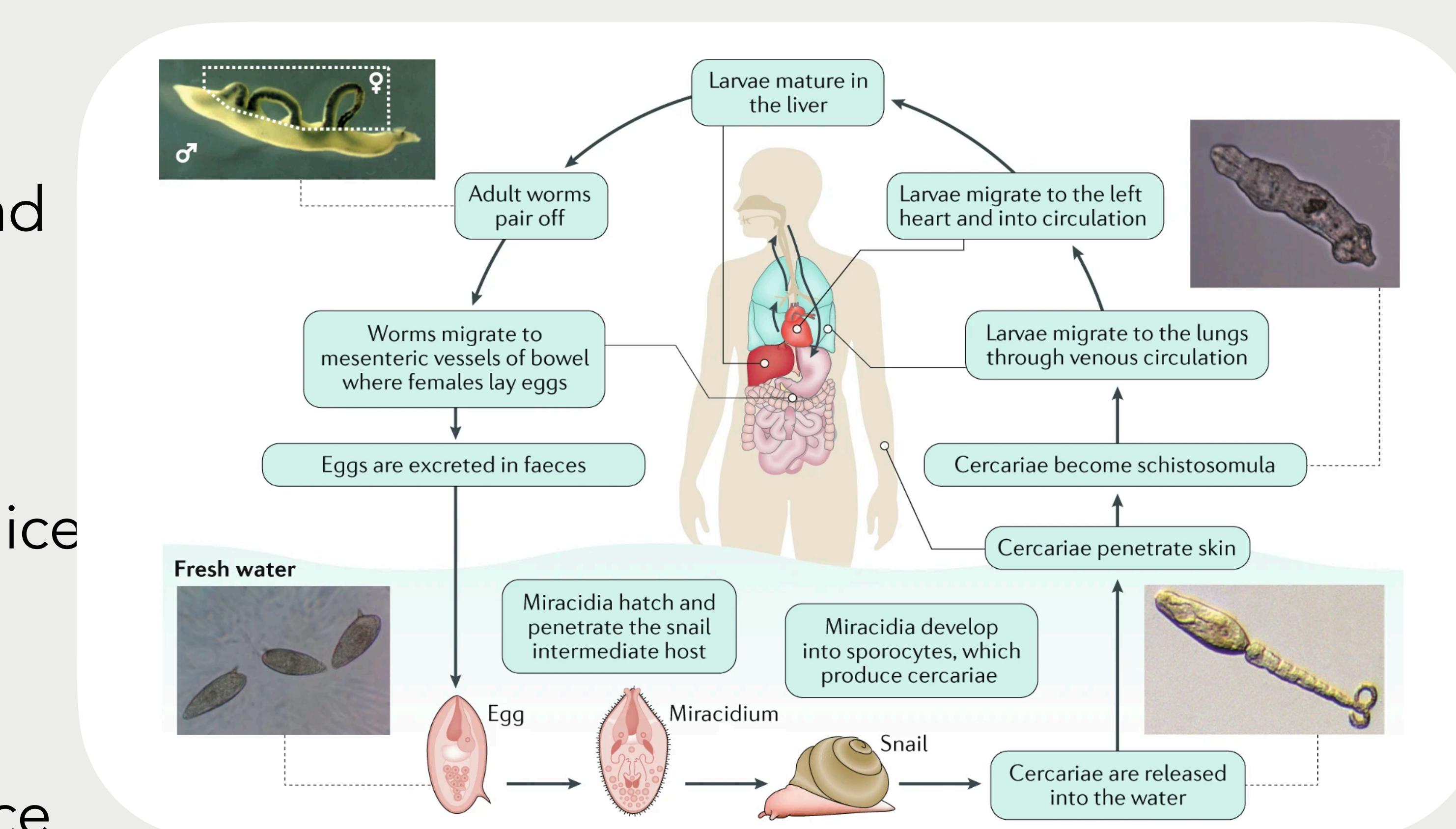
Generic goal: identify ***differentially expressed genes (DEGs)***

Why quantify mRNA expression?

Winners vs. losers: *Schistosoma mansoni* intestinal and liver eggs exhibit striking differences in gene expression and immunogenicity

Schistosomiasis

- Parasitic flatworm
- Infects >350 million people around the world
- Complex life cycle
- Laboratory strain maintained in mice
- Worm eggs extracted from livers, miracidia hatched and used for experiments/life cycle maintenance

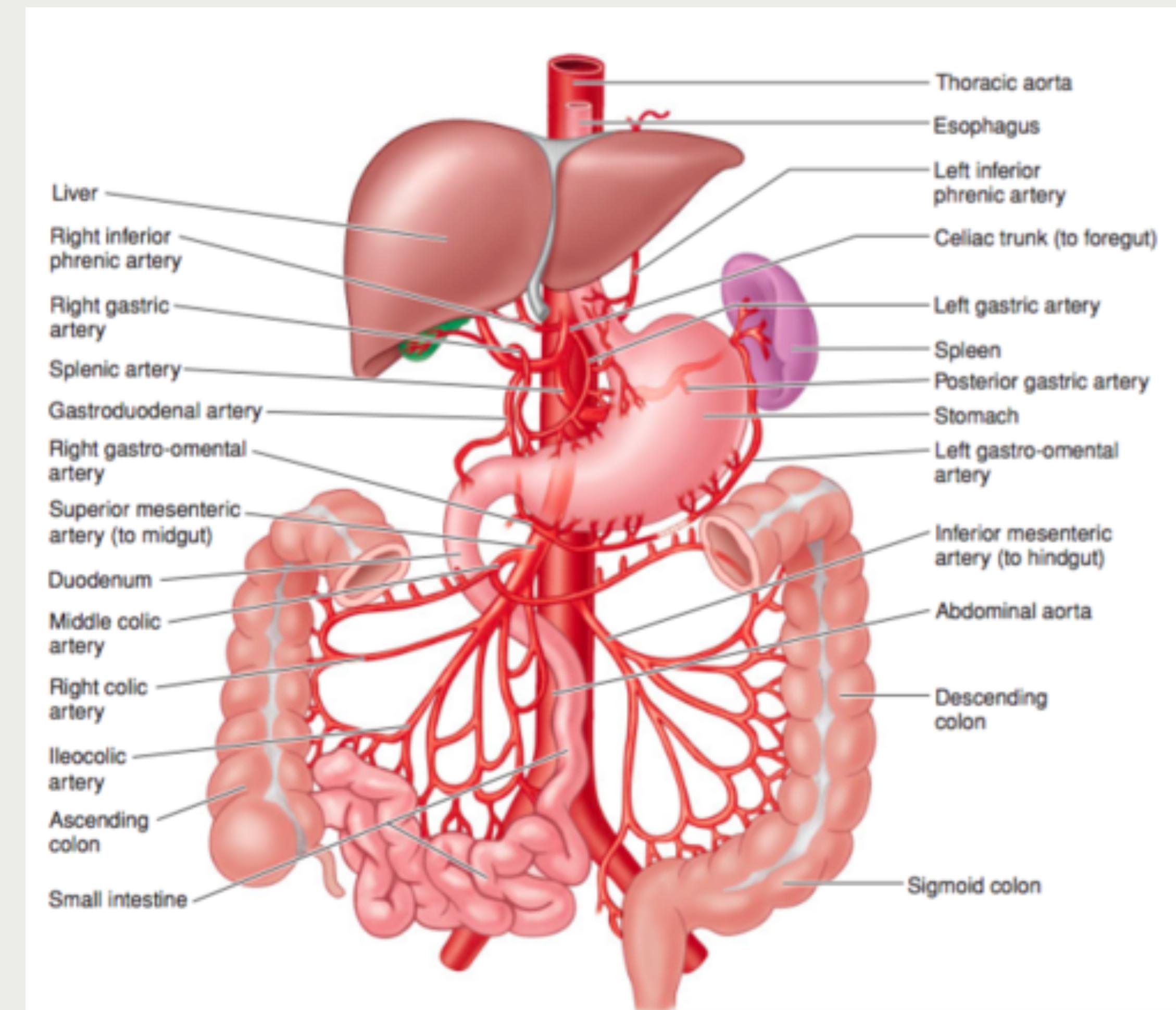


Why quantify mRNA expression?

Winners vs. losers: *Schistosoma mansoni* intestinal and liver eggs exhibit striking differences in gene expression and immunogenicity

Schistosomiasis

- Adult worms live in the mesenteric veins
- Blood travels through the veins into the portal system (liver)
- Eggs get stuck in the liver and cause pathology (never hatch)
- To continue the life cycle, eggs need to travel against the flow of blood into the intestinal tract to be excreted



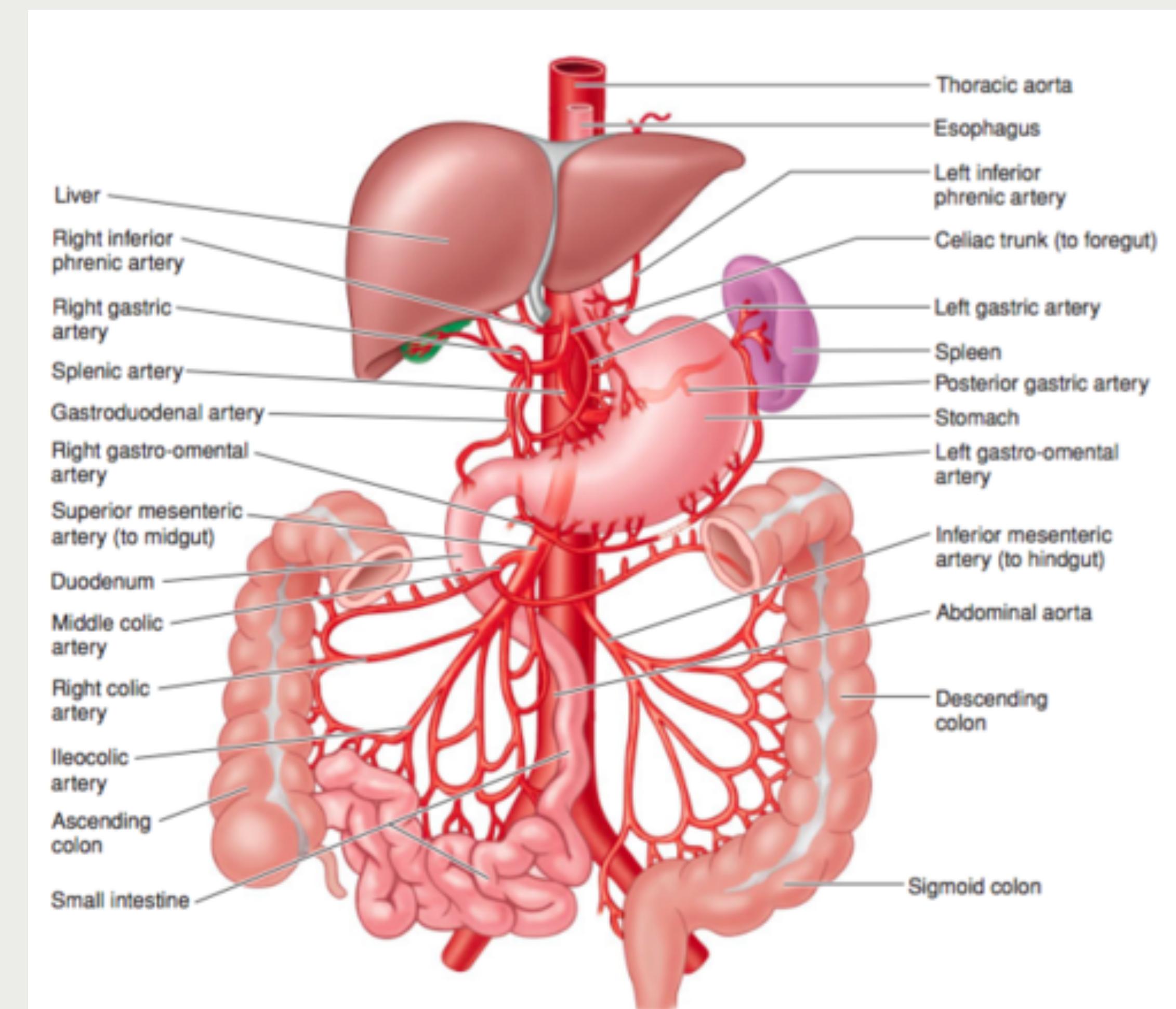
Why quantify mRNA expression?

Winners vs. losers: *Schistosoma mansoni* intestinal and liver eggs exhibit striking differences in gene expression and immunogenicity

Major research question:

Are eggs found in the mouse liver different than those found in the intestine?

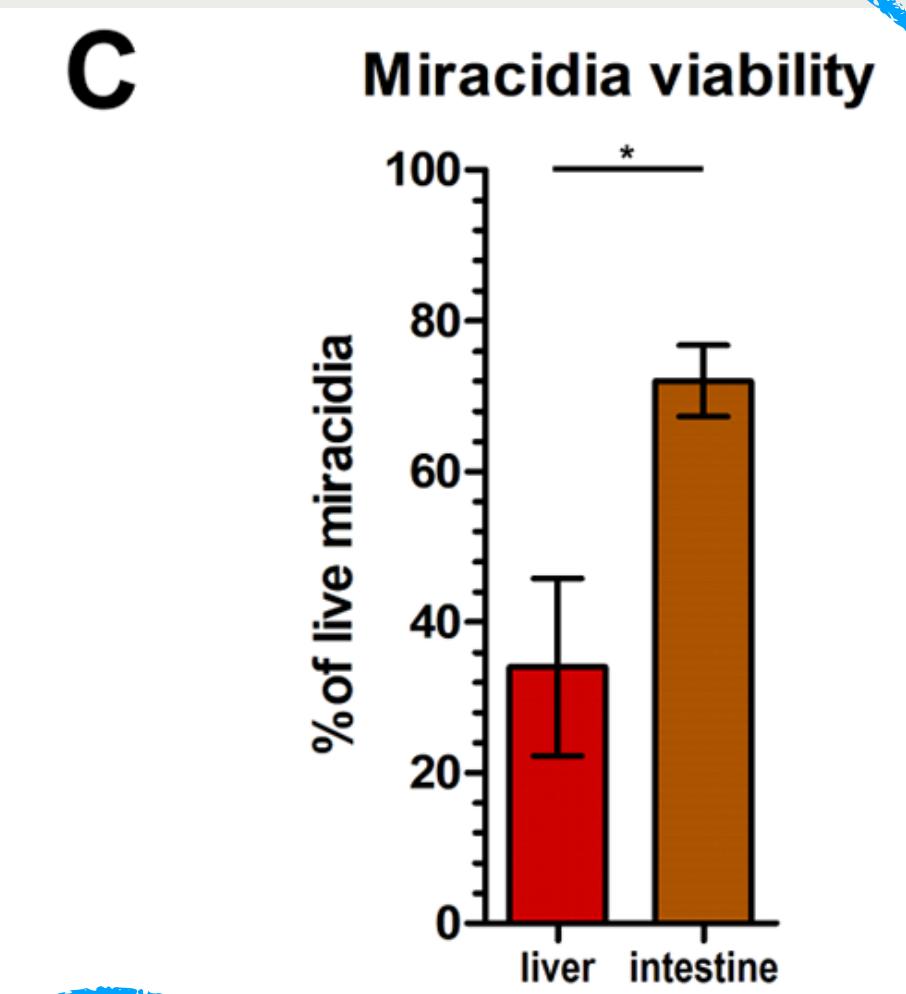
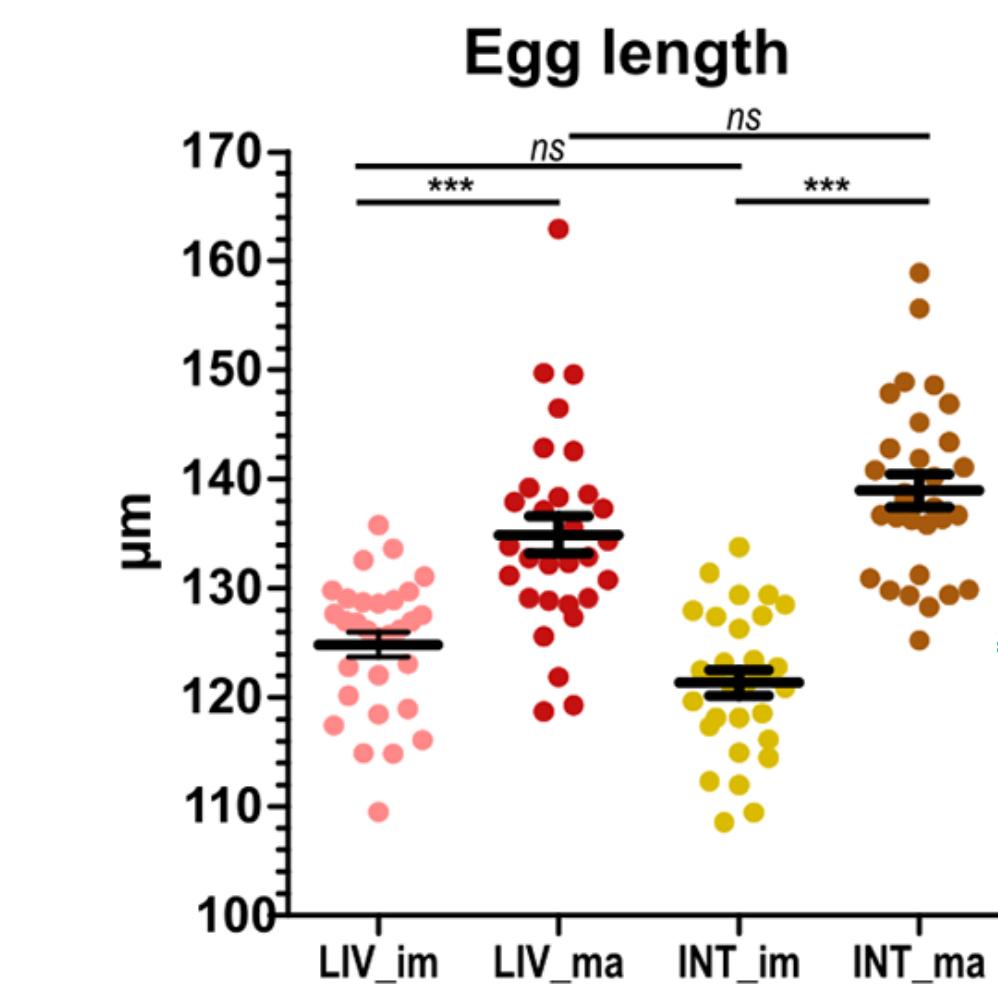
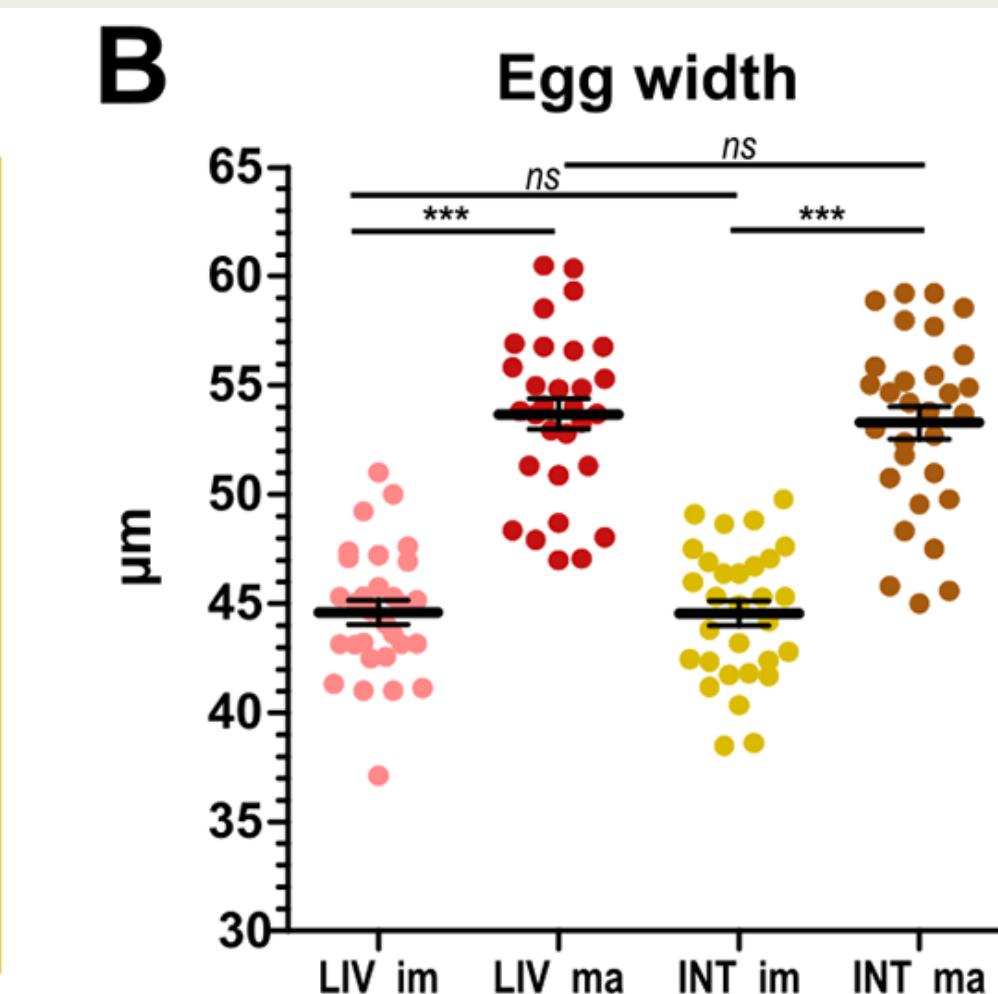
- Morphology (Do they look the same?)
- Immunogenicity (Do they cause the same immune response in mice?)
- Transcriptomically (Do they express the same mRNAs at the same quantities?)



Why quantify mRNA expression?

Winners vs. losers: *Schistosoma mansoni* intestinal and liver eggs exhibit striking differences in gene expression and immunogenicity

Eggs show significant morphological and physiological differences based on tissue source

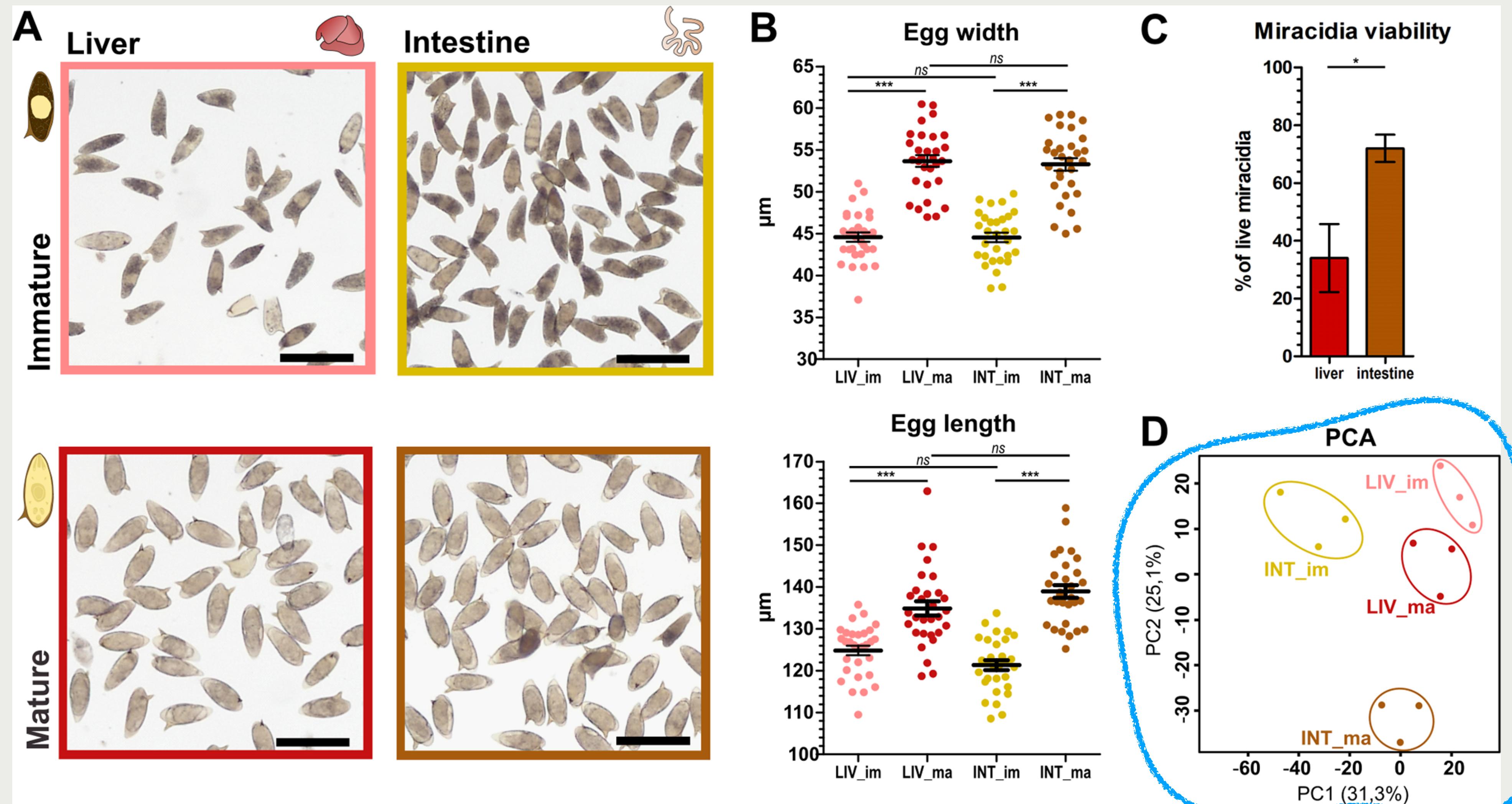


What about molecular differences?

Why quantify mRNA expression?

Winners vs. losers: *Schistosoma mansoni* intestinal and liver eggs exhibit striking differences in gene expression and immunogenicity

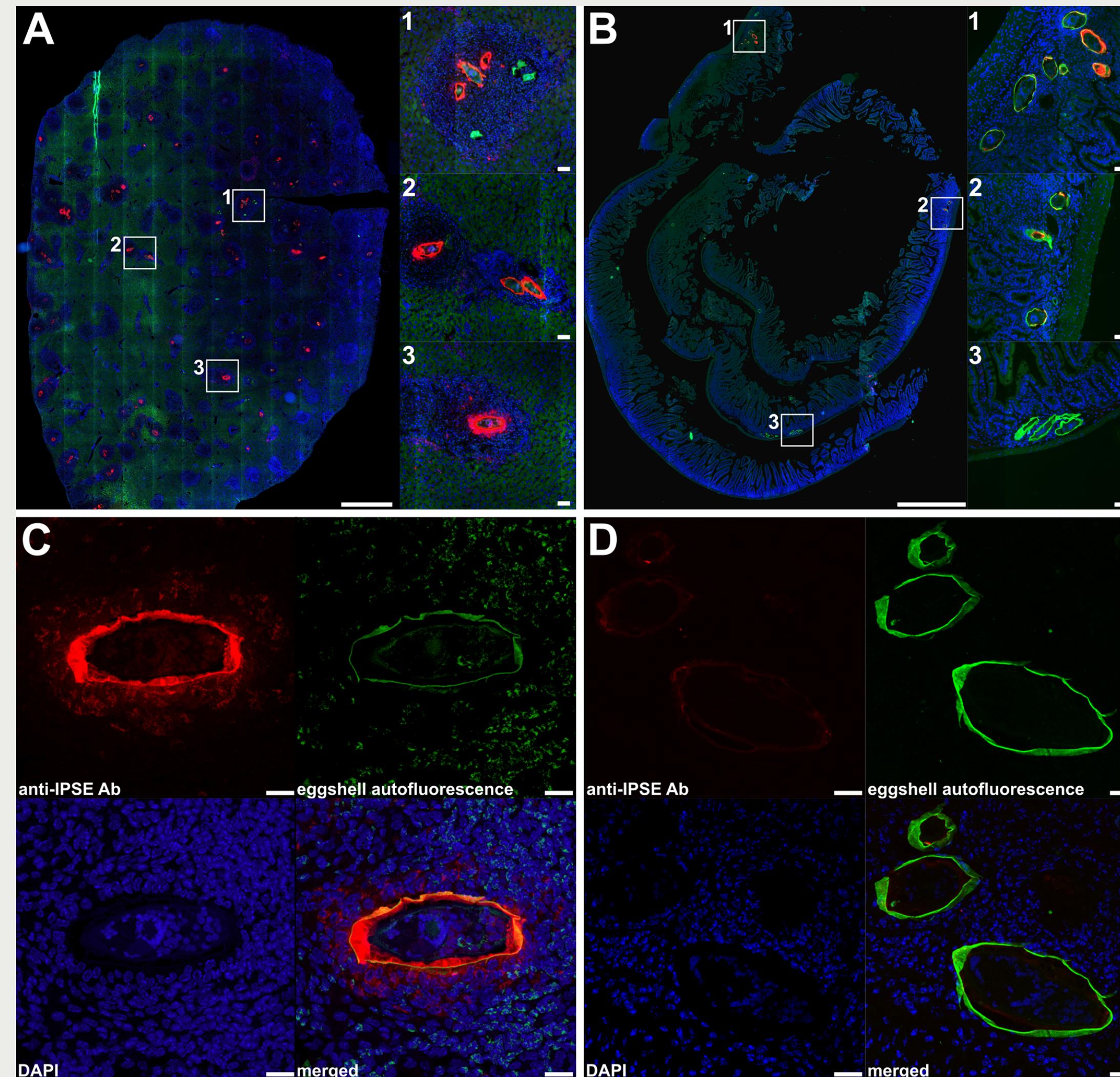
Egg transcriptomes cluster differently based on tissue source (global transcriptomic differences)
Are there specific genes that matter most?



Why quantify mRNA expression?

Winners vs. losers: *Schistosoma mansoni* intestinal and liver eggs exhibit striking differences in gene expression and immunogenicity

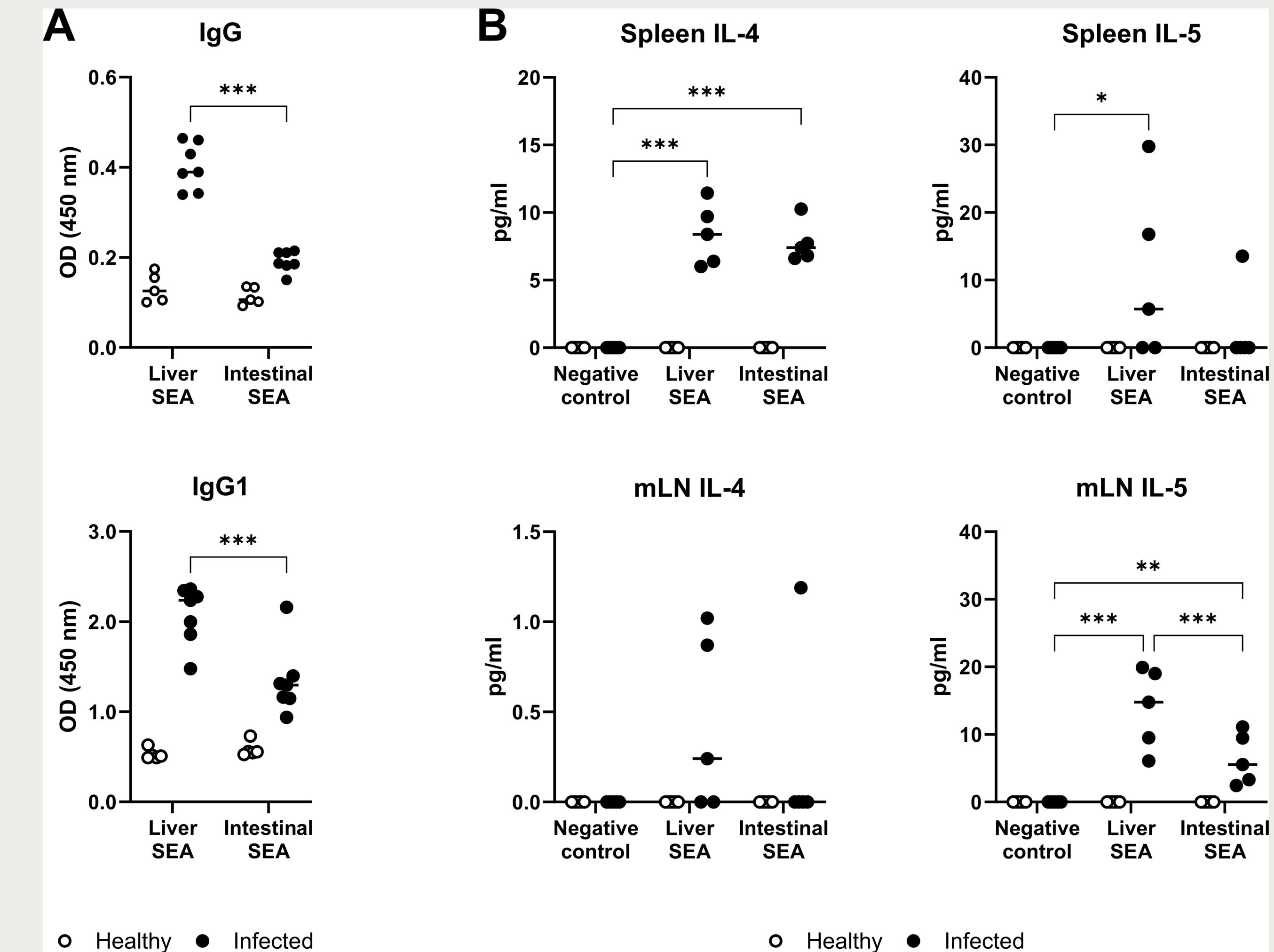
Liver eggs have significantly increased expression of immunomodulatory proteins IPSE/alpha-1 and omega-1



Why quantify mRNA expression?

Winners vs. losers: *Schistosoma mansoni* intestinal and liver eggs exhibit striking differences in gene expression and immunogenicity

Eggs from different tissues cause a significantly different host immune response



Ways to measure mRNA quantity

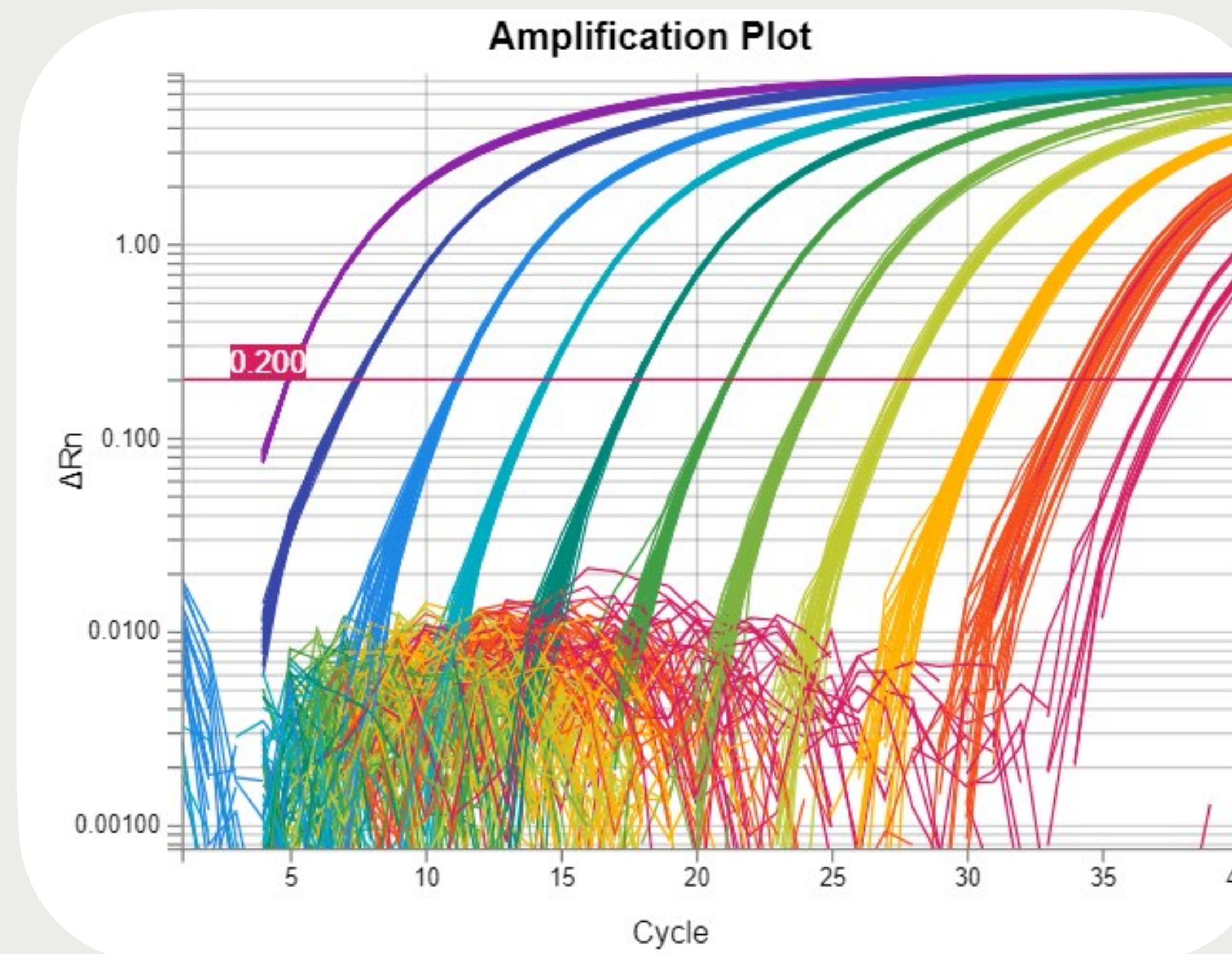
RT-qPCR: one to dozens transcripts at a time

Real-time quantitative polymerase chain reaction

1. Begin with RNA template (i.e., all RNA from an animal)
2. Convert to complementary DNA (**cDNA**) with **reverse transcriptase**
3. Amplify cDNA with Taq polymerase and **fluorescent dNTPs**
4. Thermocycler cycles through melt (denature), anneal, and extension temperatures
 - a. Sensor detects the fluorescence of intercalating DNA dye

Ways to measure mRNA quantity

RT-qPCR: one to dozens transcripts at a time



Drawbacks:

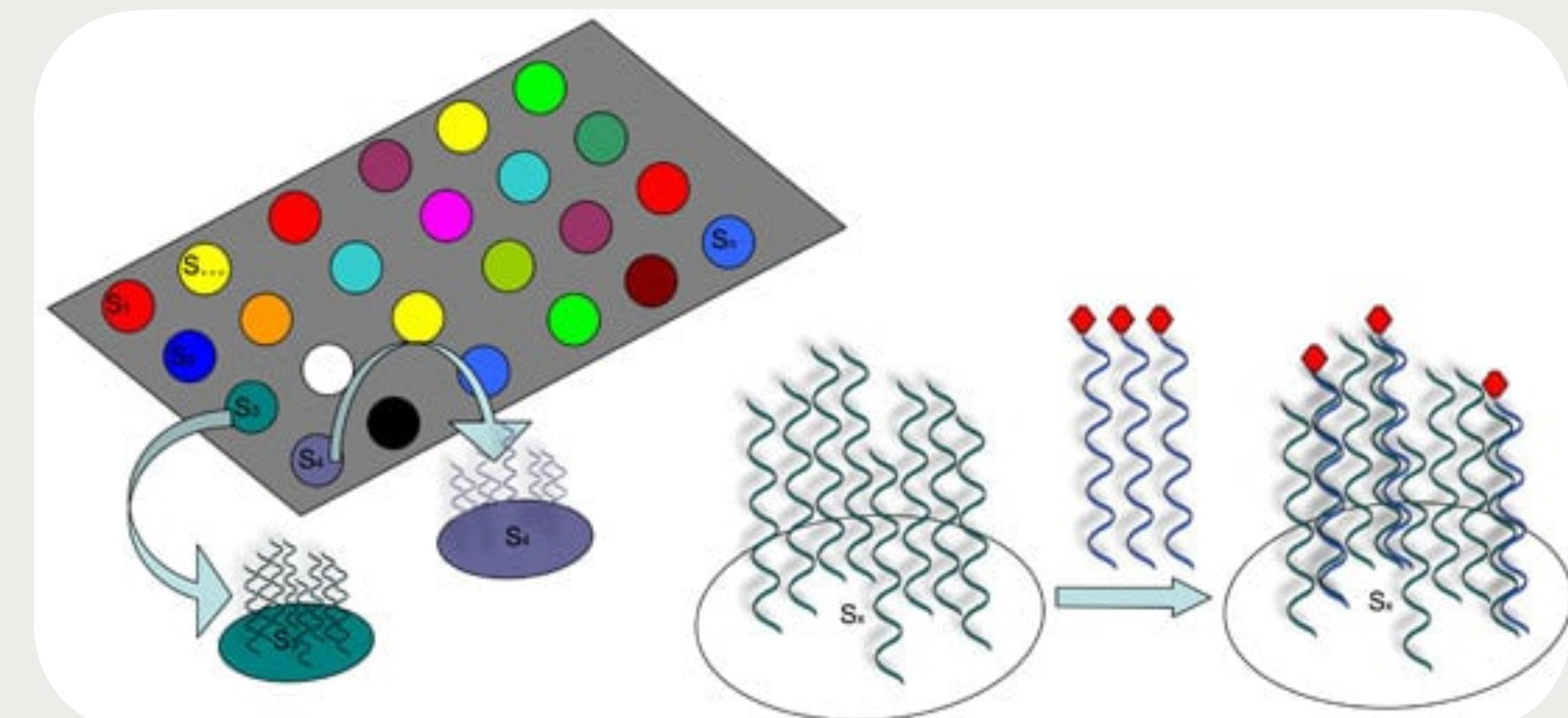
- 1) Required to design primers for each gene of interest
- 2) Number of genes of interest limited by technology (thermocyclers, well plates)

Ways to measure mRNA quantity

Microarray: hundreds to thousands of transcripts at a time

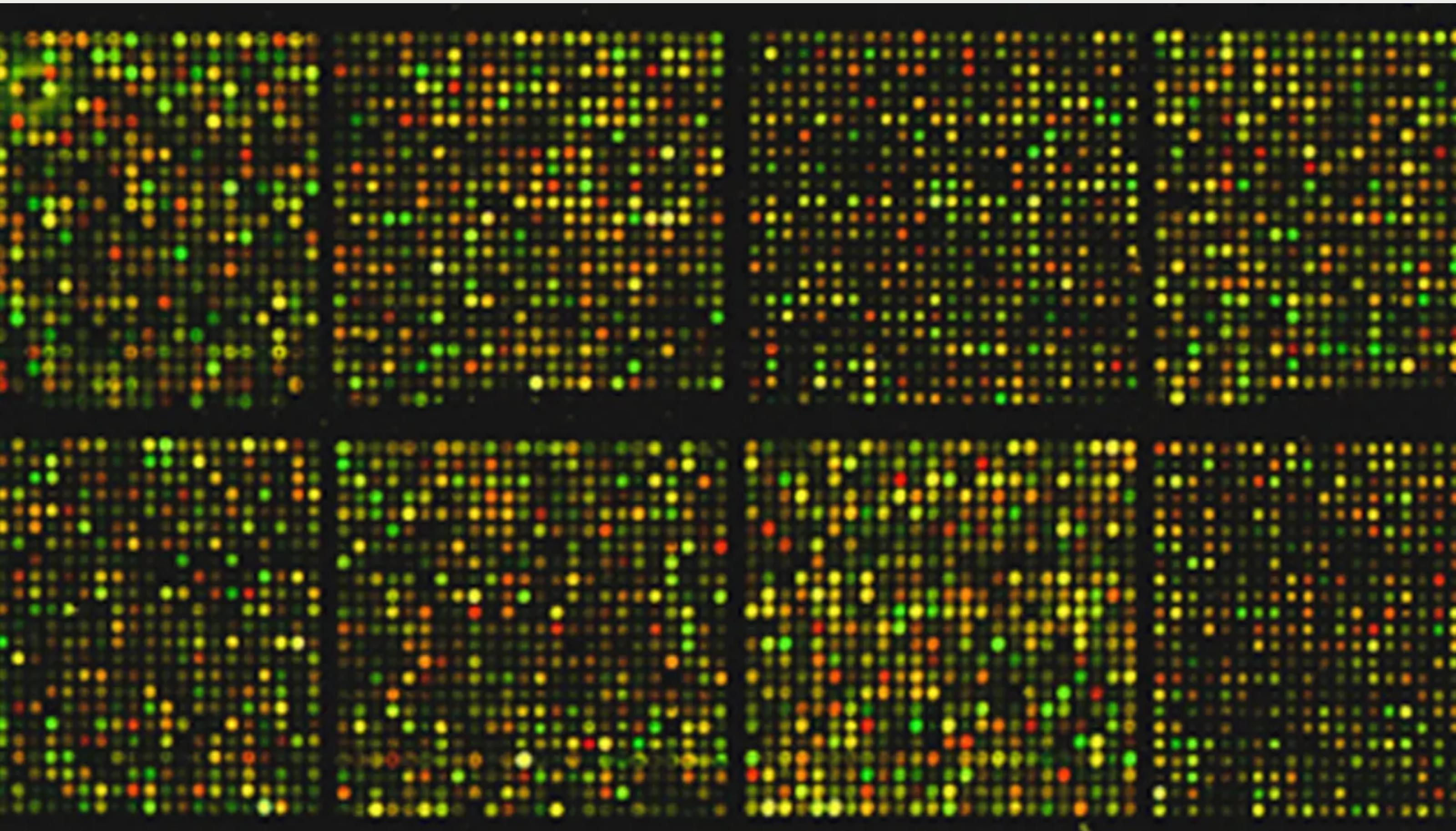
Microarray

1. Begin with RNA template (i.e., all RNA from an animal)
2. Convert to complementary DNA (**cDNA**) with **reverse transcriptase** and **fluorescent dNTPs**
3. Design a microchip with an array of probes
4. Melt cDNA and **hybridize** to the chip
5. Detect fluorescent spots on chip



Ways to measure mRNA quantity

Microarray: hundreds to thousands of transcripts at a time



Drawback:

- 1) Required to design probes for each gene of interest
- 2) Lacks sensitivity (lower than RT-qPCR)

Ways to measure mRNA quantity

Expressed sequence tags: hundreds to thousands of transcripts at a time (non-quantitative)

ESTs

1. Begin with RNA template (i.e., all RNA from an animal)
2. Convert to complementary DNA (**cDNA**) with **reverse transcriptase**
3. Clone cDNAs (<1000 bp) into plasmids
4. Sequence each clone (old school tech)

Drawback:

- 1) Only one read per clone ("single-pass;" non-quantitative)

Ways to measure mRNA quantity

RNA sequencing: unlimited number of transcripts

RNA-seq

1. Begin with RNA template (i.e., all RNA from an animal)
2. Convert to complementary DNA (**cDNA**) with **reverse transcriptase**
 - a. Optional: PCR amplify to increase sensitivity
3. Prepare **sequencing libraries**
4. Sequence

Drawbacks:

- 1) Significant expertise needed for data quality control and analysis

Ways to measure mRNA quantity

RNA sequencing: unlimited number of transcripts

