

# Pathway Analysis

Arpan Kumar Basak  
Post-doctoral Researcher  
Max-Planck Institute for Plant Breeding Research,  
Cologne, Germany

Internship | Functional Bioinformatics | 01-31 July 2023

**Bioakademy**  
A UNIT OF AGT BIOSCIENCES (OPC) PVT. LTD.



Dept. of Botany, Gauhati University

## **Disclaimer**

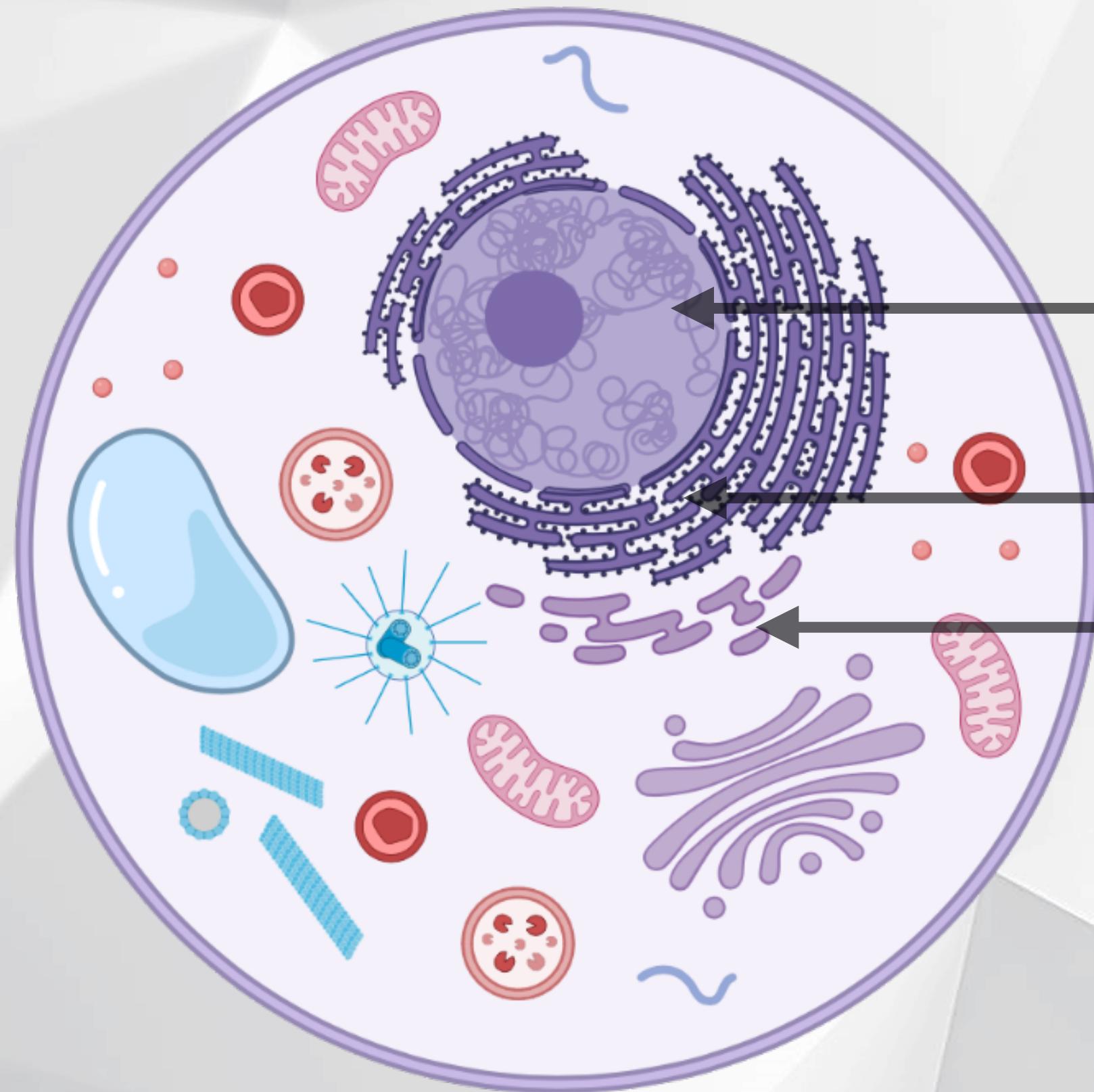
The organizers (AGT Biosciences and Dept. of Botany, Gauhati University) are not responsible for the accuracy or reliability of the information provided in this presentation. It is the responsibility of the presenter, and the presenter grants the exclusive rights to AGT Biosciences (OPC) Pvt. Ltd. to record, store, distribute and display the presentation in print, electronic and digital forms.



# Learning outcome

- Concept of cellular pathway
- What questions can we derive by understanding cell pathway?
- Statistical summary and inference
- Why do we need to understand the pathway?
- Some example of pathway analysis
- Commonly used databases and caveats
- Using bioconductoR packages for pathway analysis

# A typical cell and its compartments



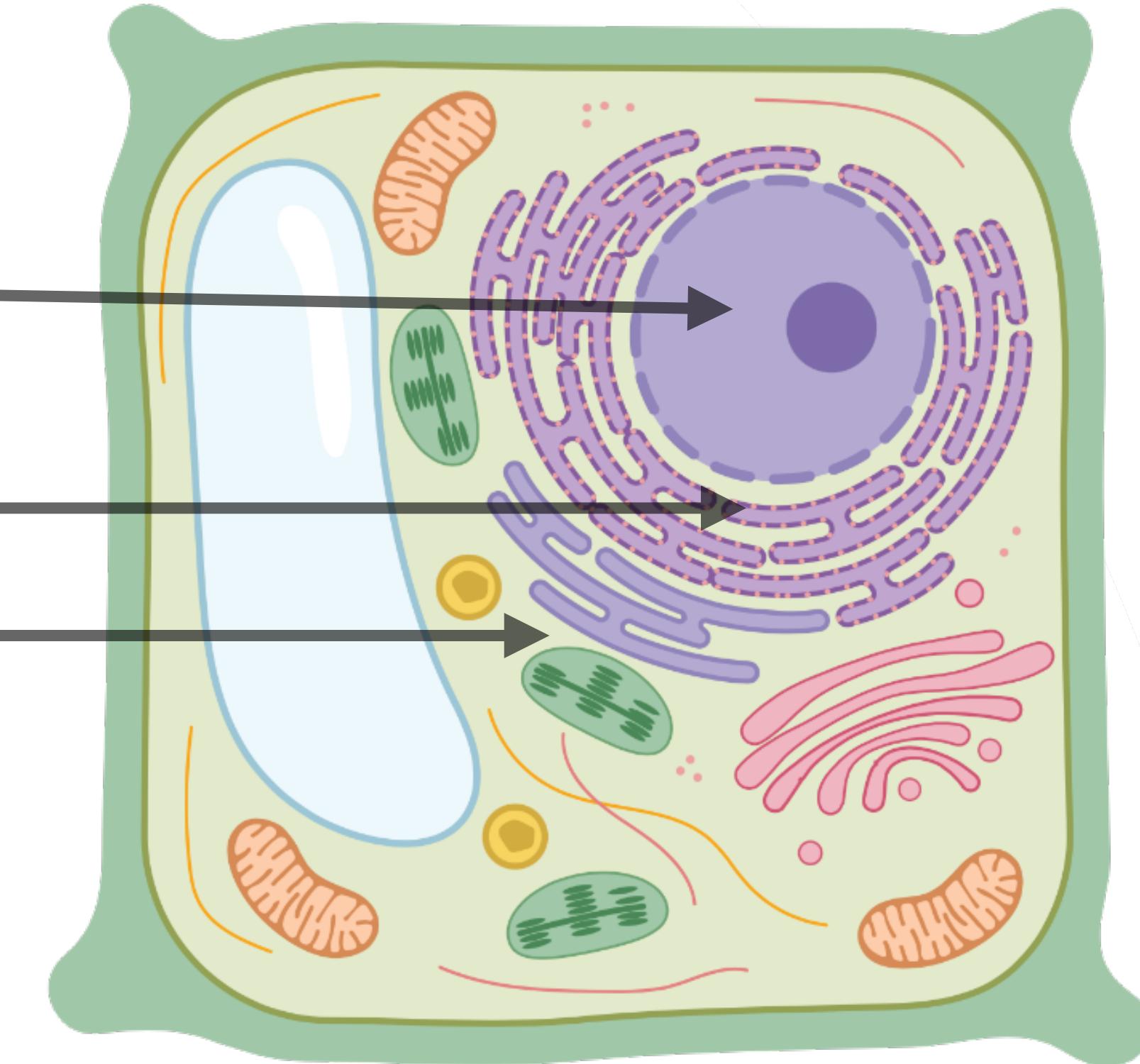
DNA replication

Transcription

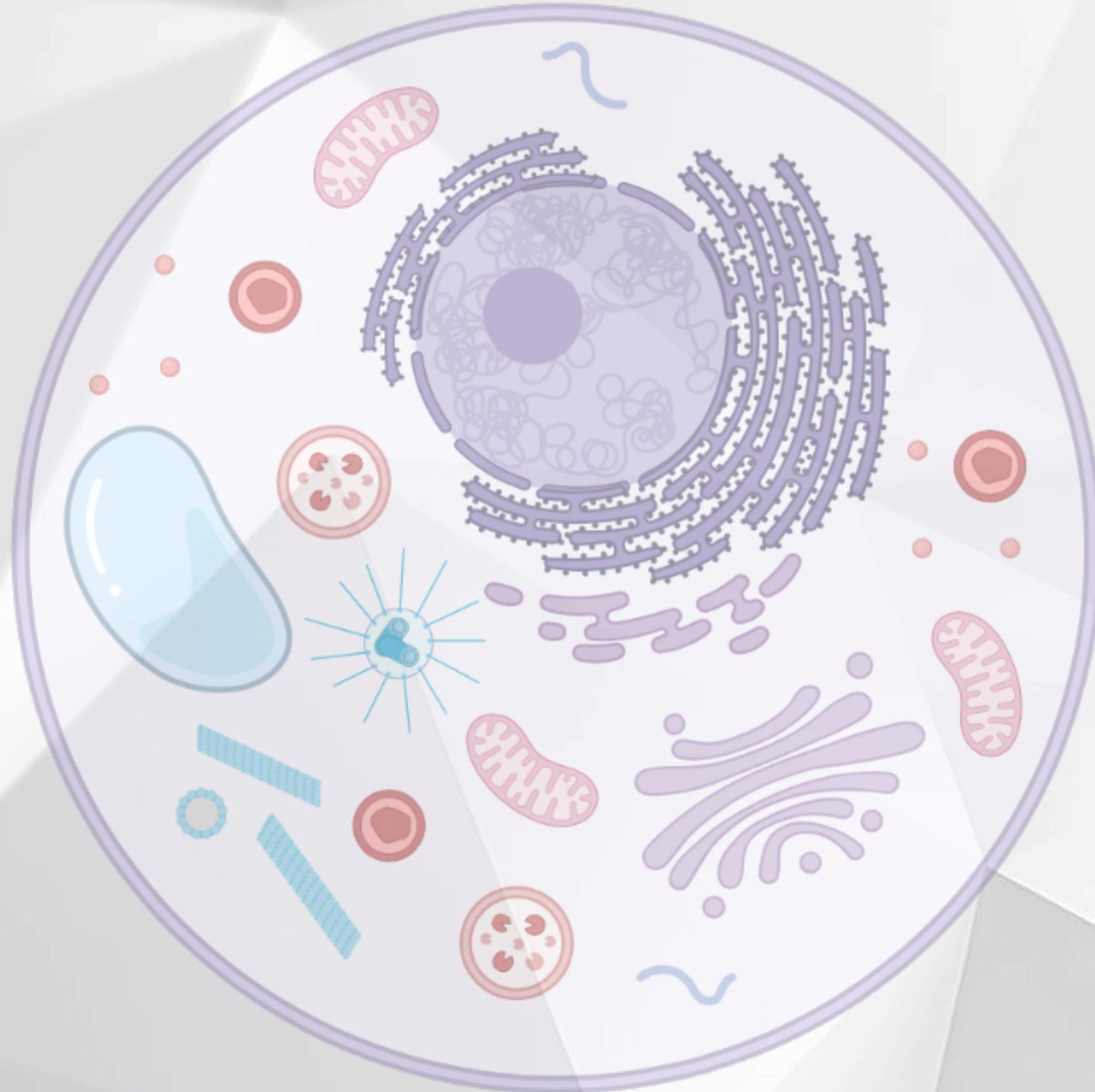
Protein synthesis

Protein modification

Metabolic function



# How does cell orchestrate a pathway?



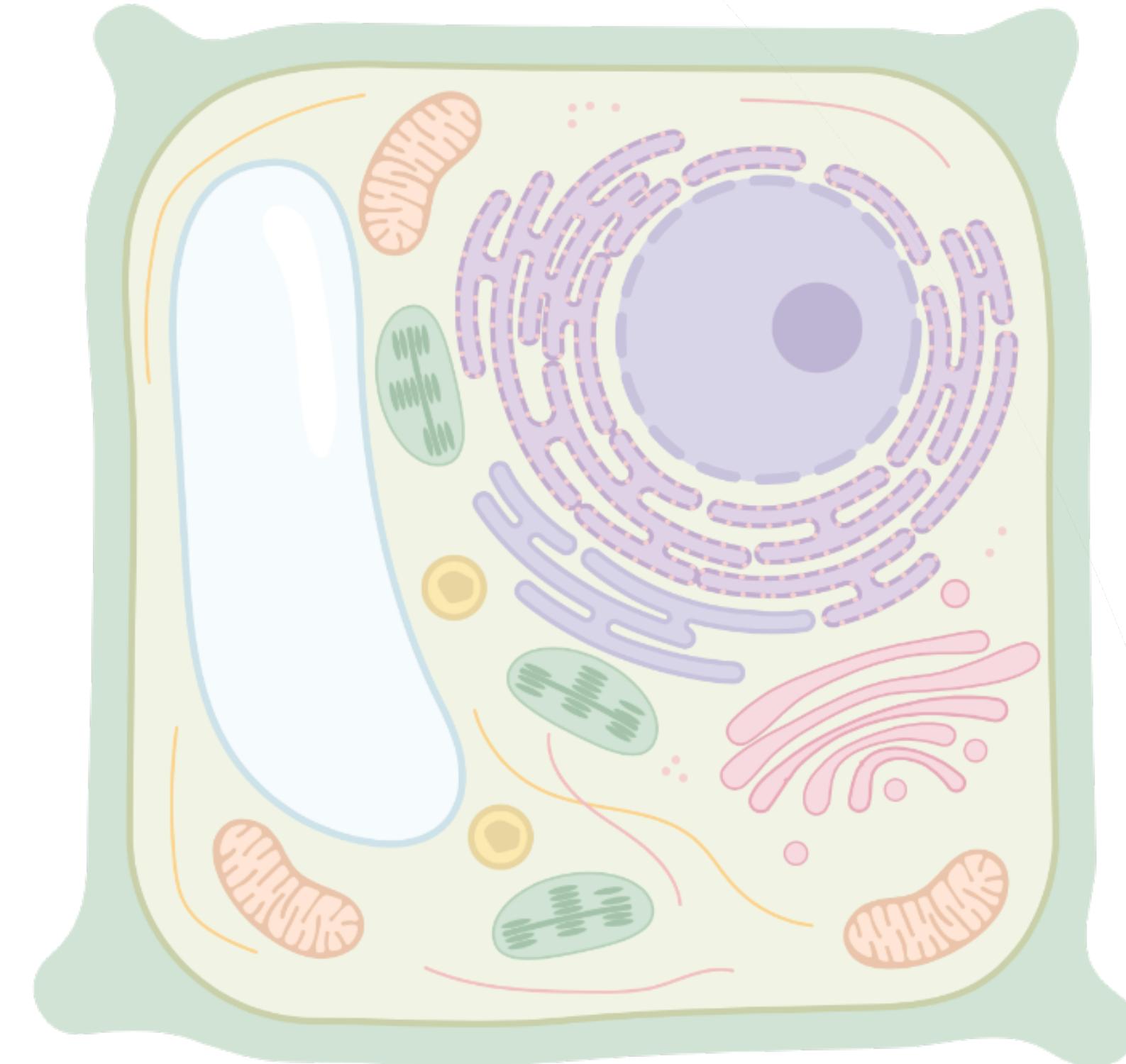
**Protein Complex A**

**Protein A  
Protein B  
Protein C**

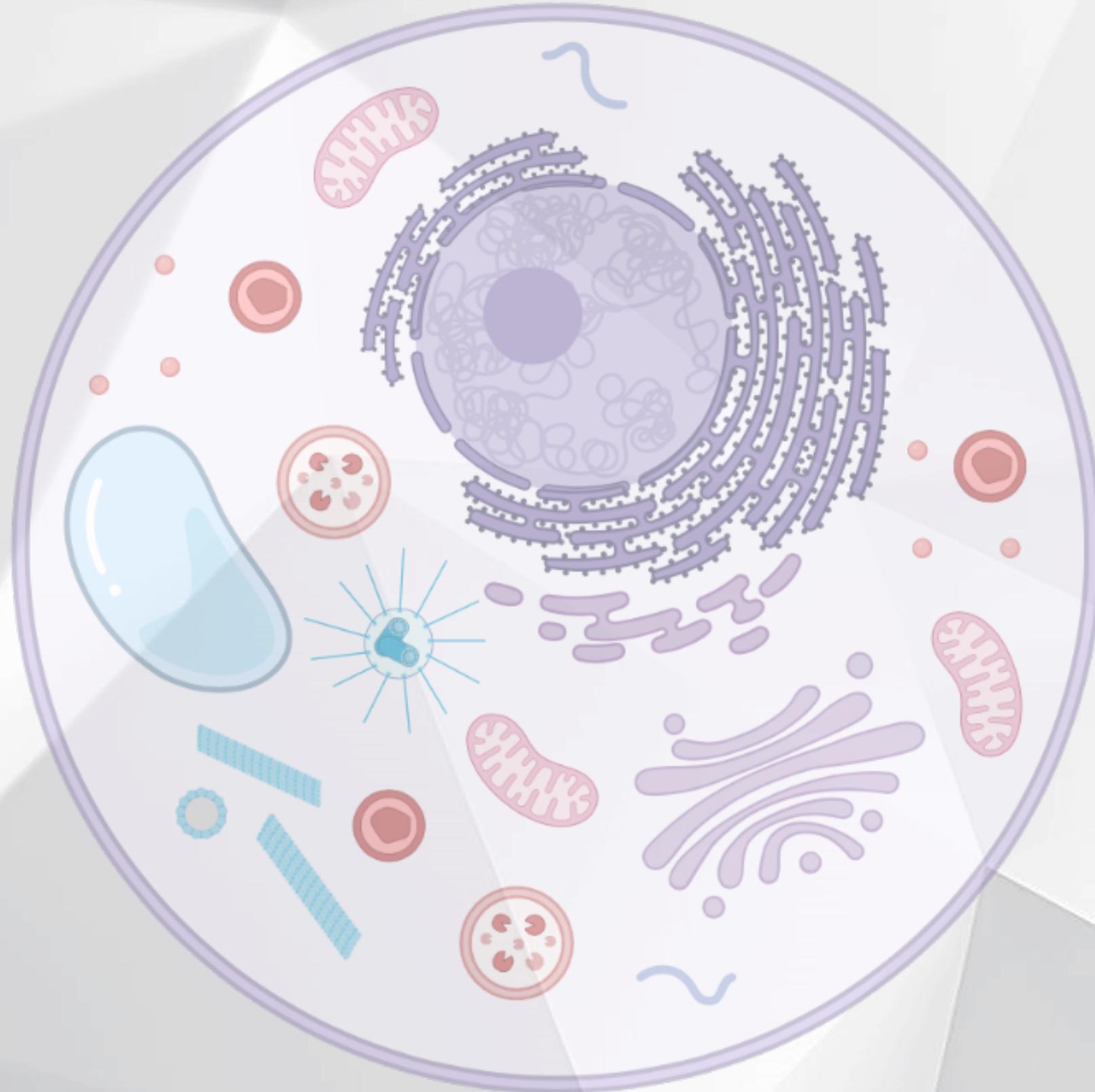
**Modulating the protein-protein interactions**



**Metabolic function**



# How does cell orchestrate a pathway?



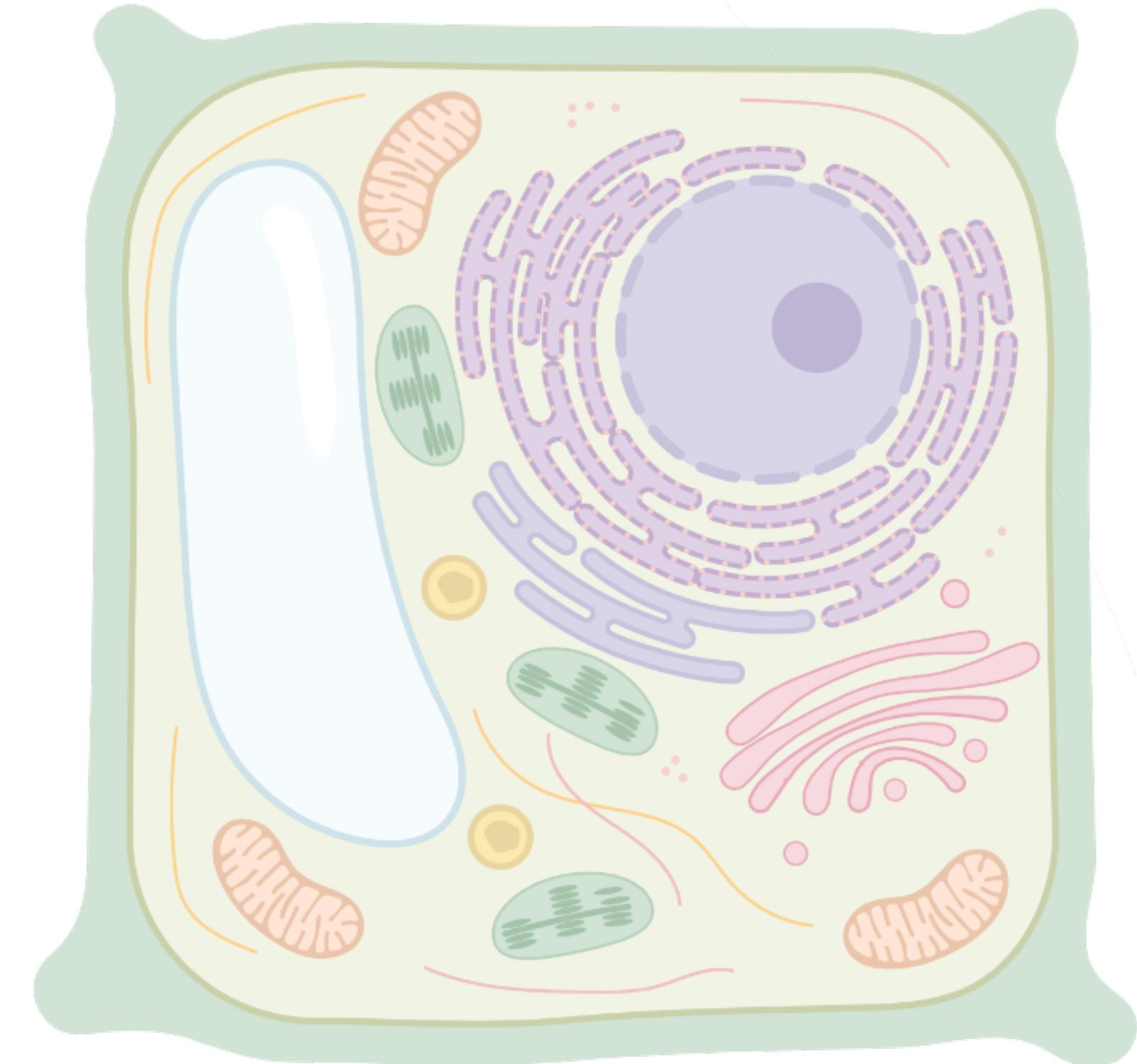
**Protein A**

**Gene A  
Gene B  
Gene C**

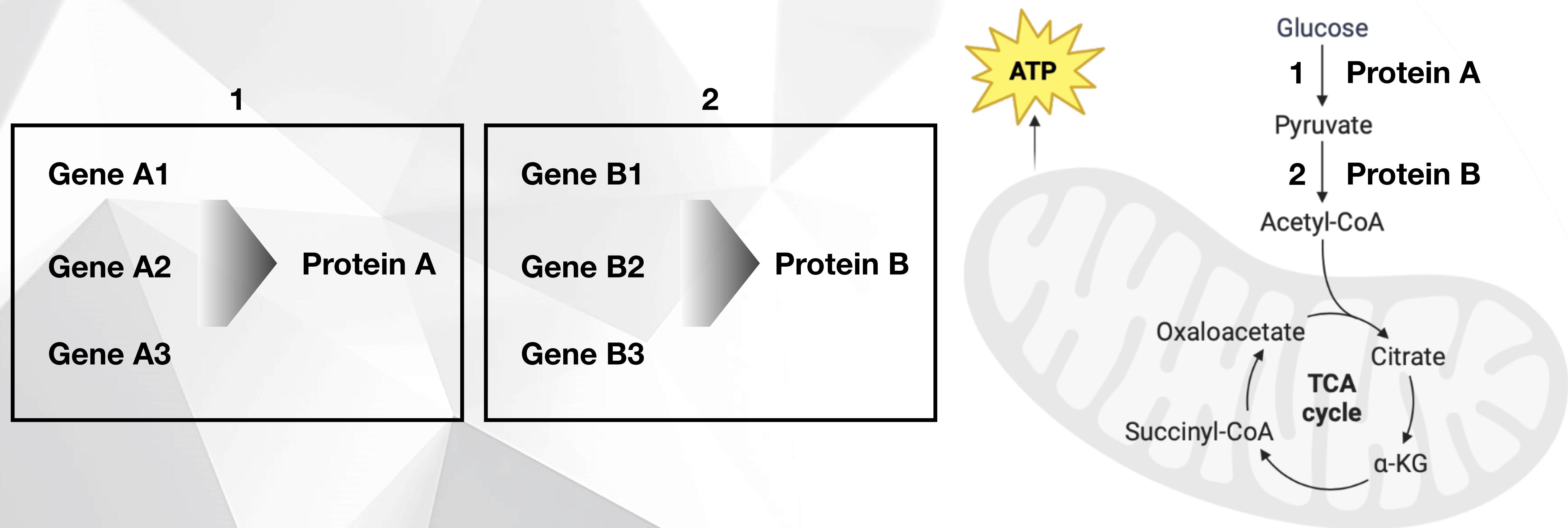
**Modulating the gene expression**



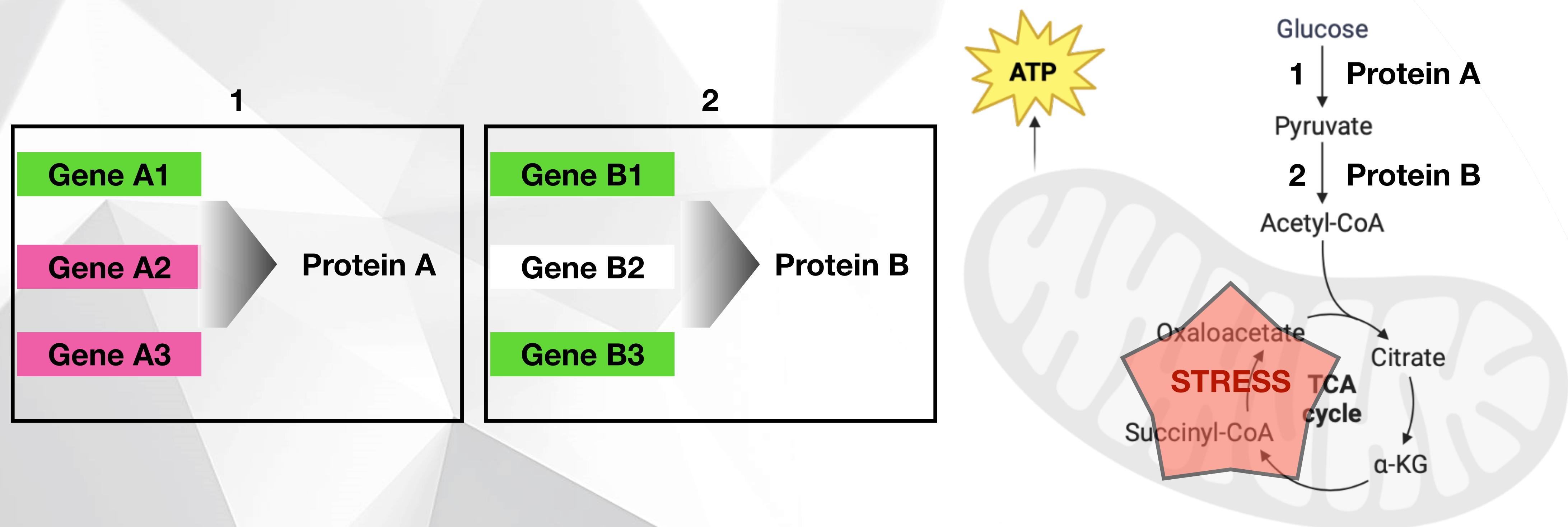
**Metabolic function**



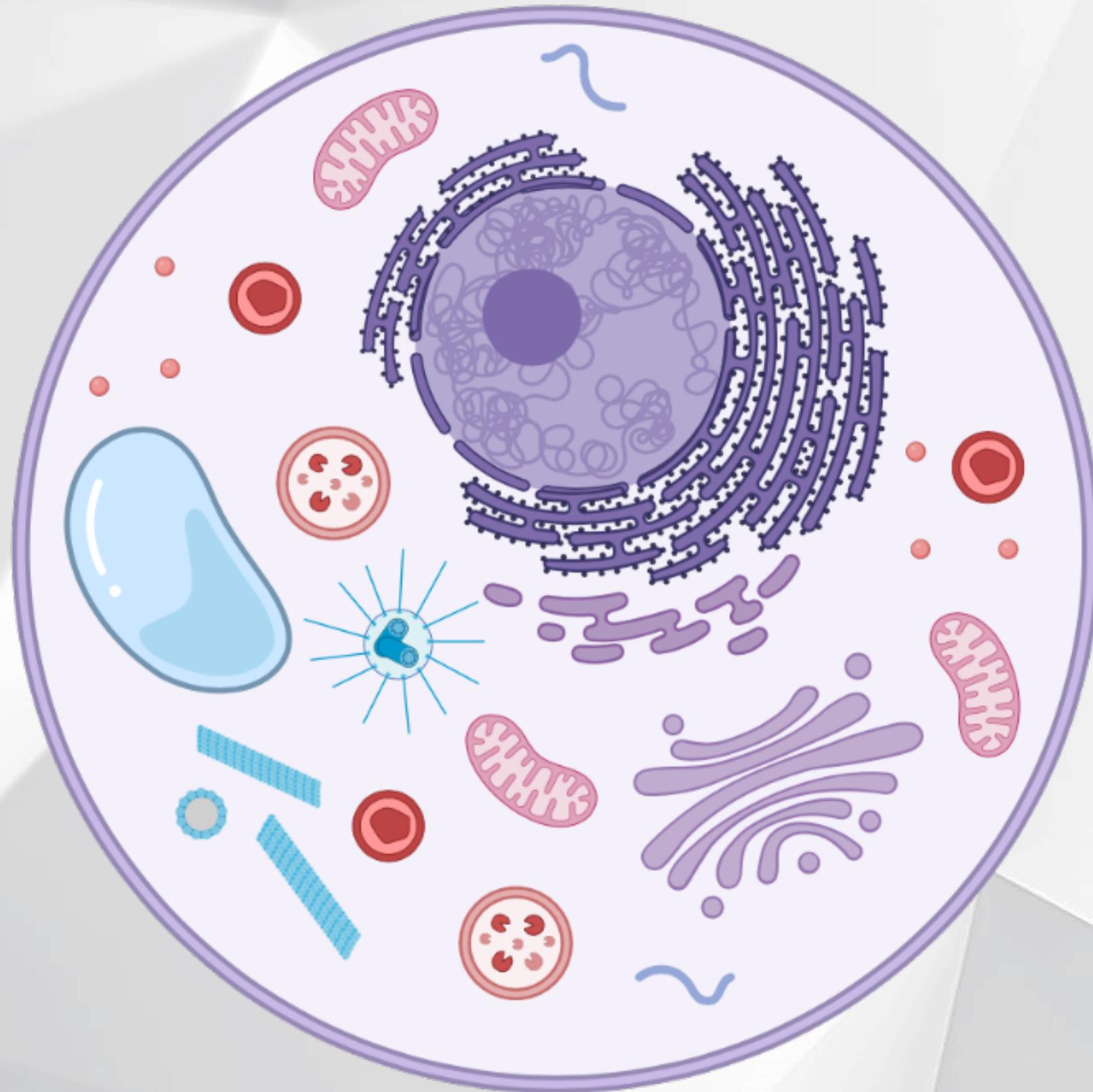
# How does cell orchestrate a pathway?



# How does cell orchestrate a pathway?



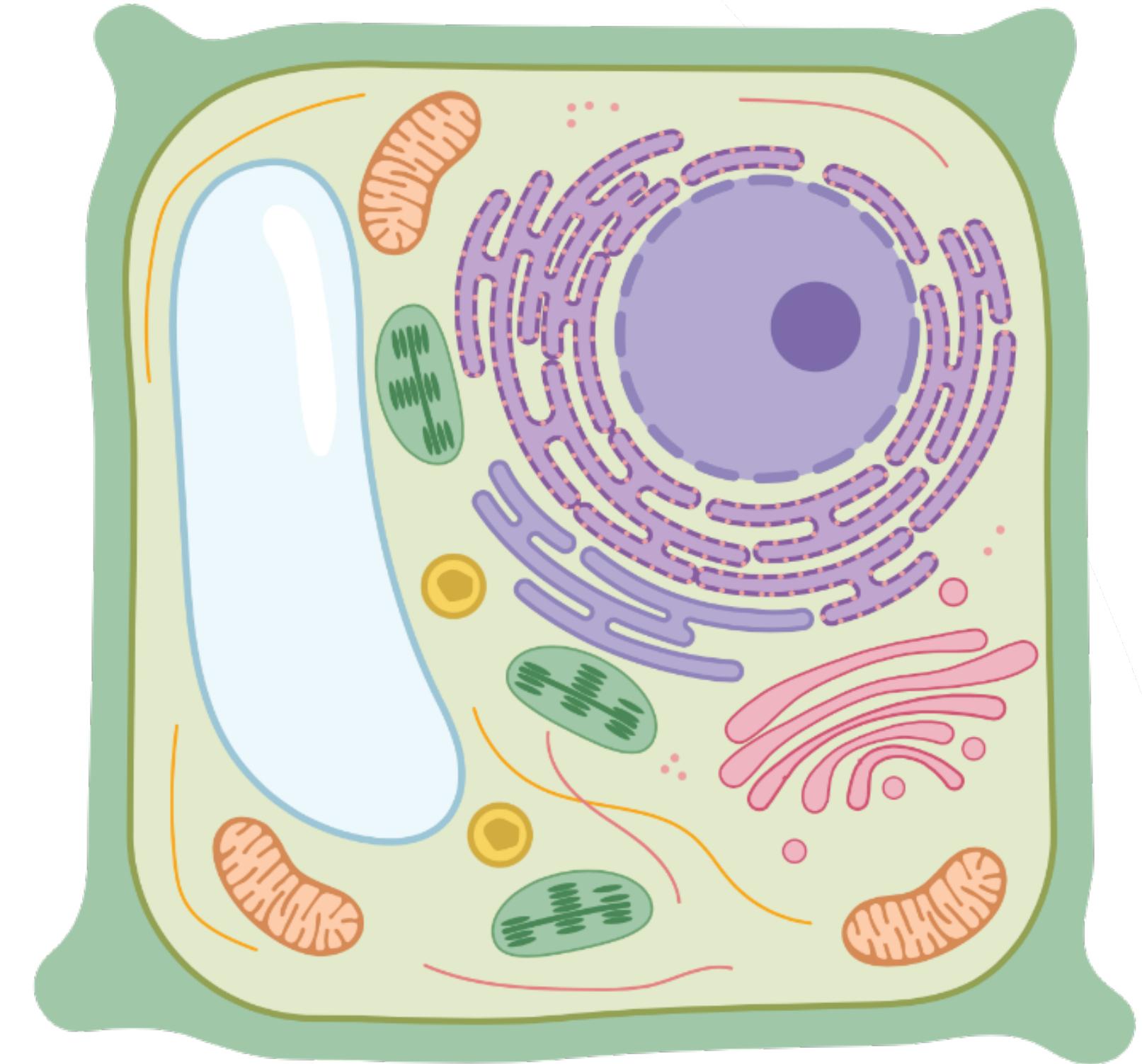
# What if you want to capture the complete gene expression pattern of not one but for all cellular pathways?



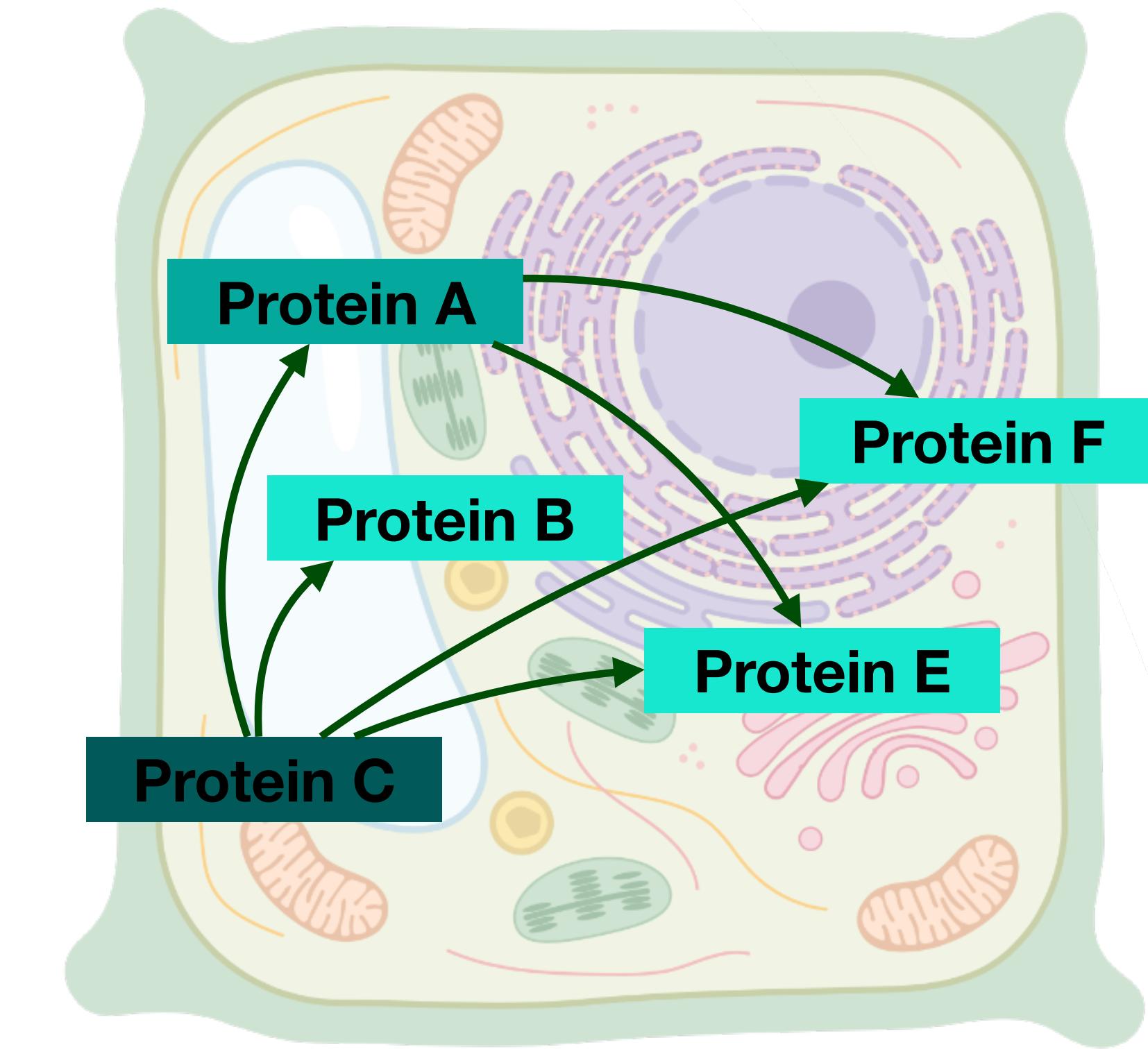
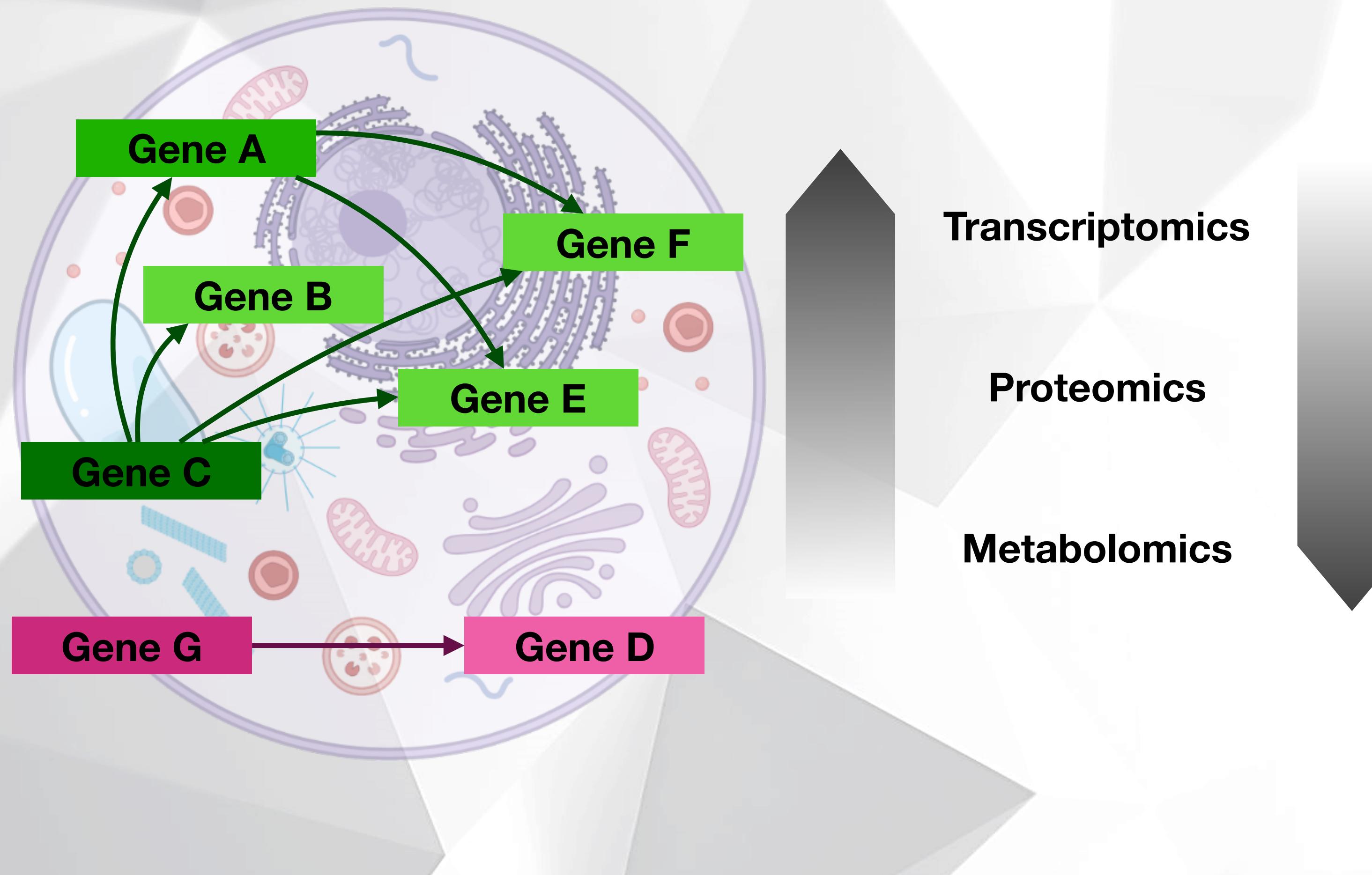
Transcriptomics

Proteomics

Metabolomics

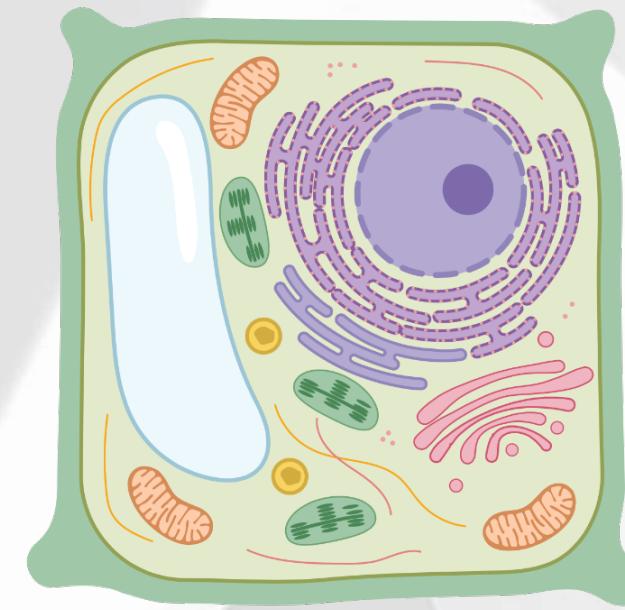


# What if you want to capture the complete gene expression pattern of not one but for all cellular pathways?



## What about function?

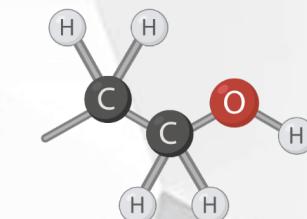
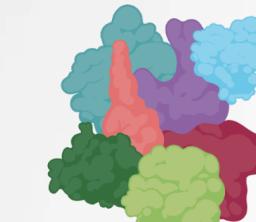
# Components of metabolic pathway



**HT-Sequencing**



**Mass Spectrometry**



Mapping the spectrum

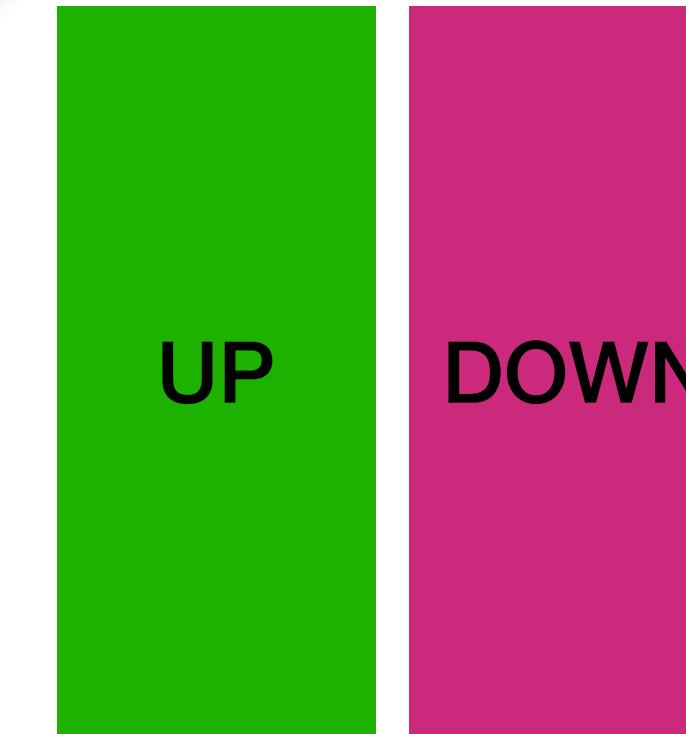
ATGCGATTCCA...

Mapping

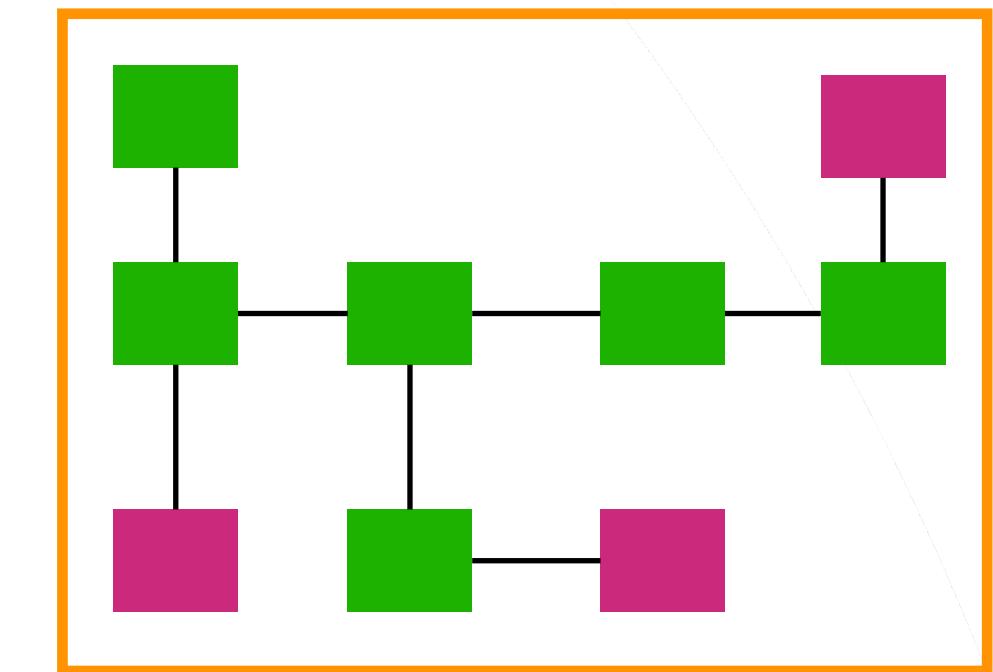
MDESNKKL...

CC(=O)C(O)=O

**Moderated T-test or GLM**



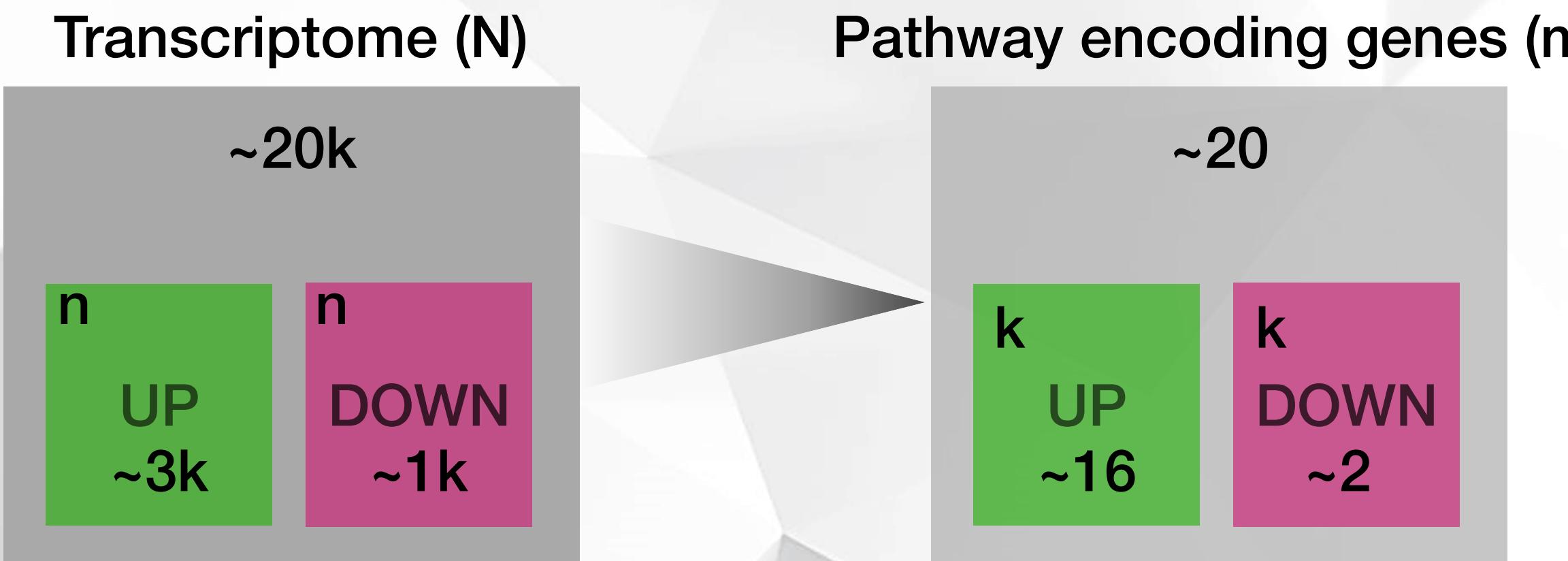
**Pathway Database**



**Enriched**

**Enrichment analysis by Hypergeometric test**

# Statistical inferences in pathway analysis



$N$  is the total number of genes considered (gene universe).  
 $n$  is the number of DE genes.  
 $m$  is the number of genes in the gene set.  
 $k$  is the number of DE genes in the gene set.

$$\Pr(X \geq k) = \sum_{x=k}^n \frac{\binom{m}{x} \binom{N-m}{n-x}}{\binom{N}{n}}$$

# Statistical inferences in pathway analysis

- For multi-omics: Canonical correlation analysis
- Linear or Mixed Discriminant analysis
- Machine Learning models for functional prediction (Random forest or Support Vector Machine)
- Network graph analysis

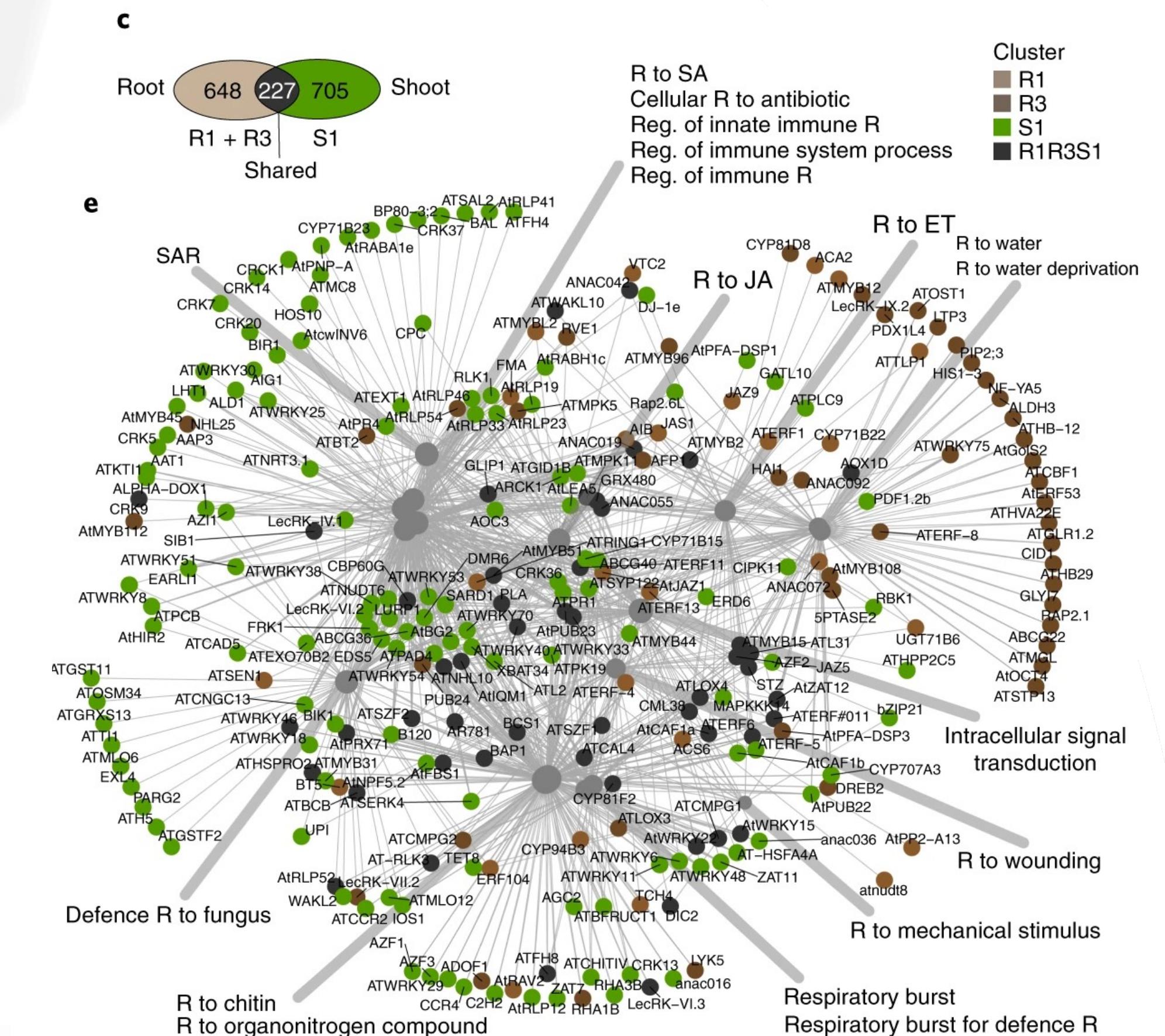
# Why pathway analysis?

- Global perspective on the cell pathway
- Understanding the interdependencies between the expression patterns of the gene
- Data interpretation in the context of biological processes
- Data integration of multi-omics to identify function of the pathway

# Advantages of learning metabolic pathways

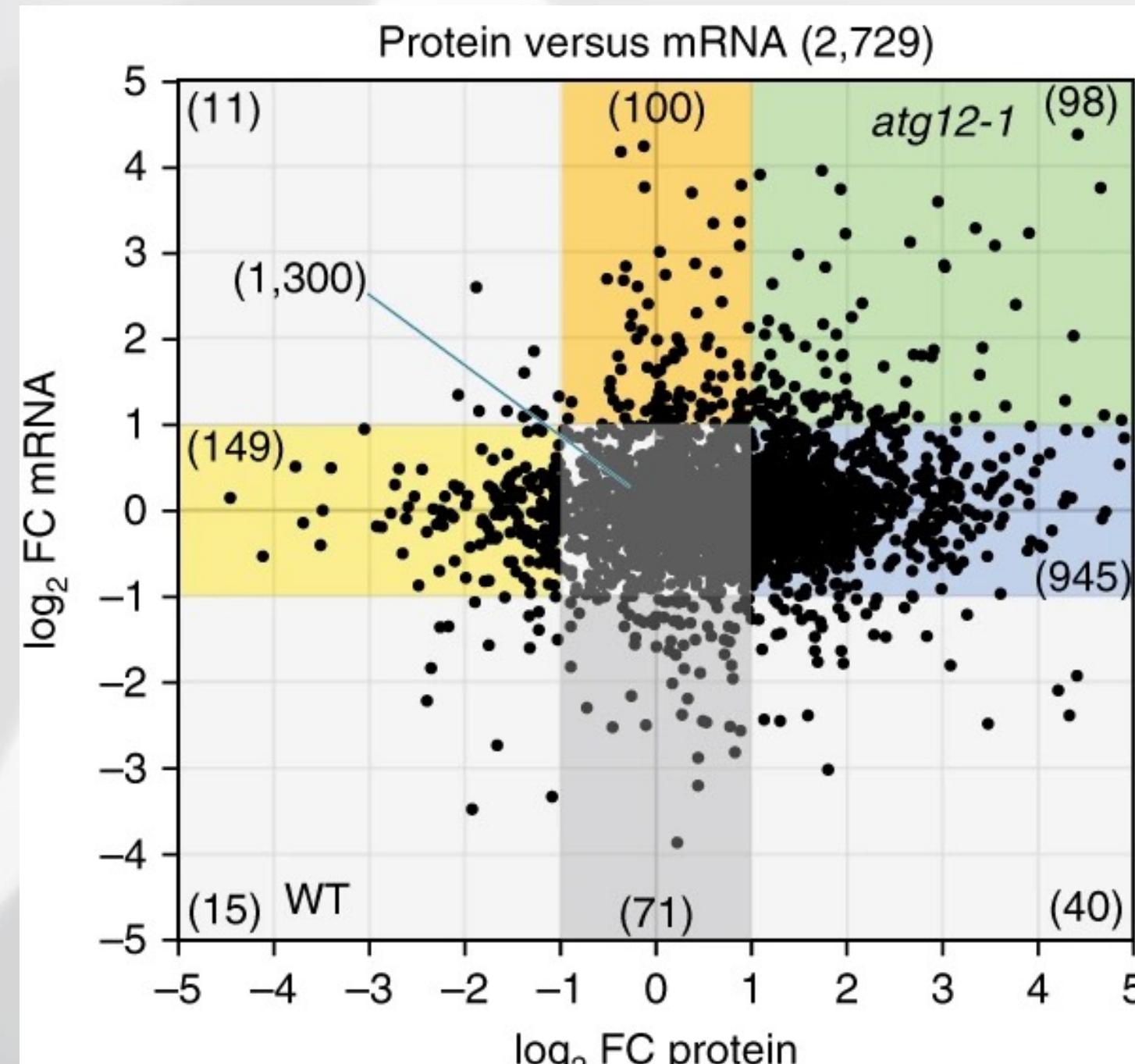
- Prediction of drug targets
- Identification of disease causing genes and its regulatory mechanism
- Impact of experimental perturbation (stress, drug) on cellular pathway

# Gene based analysis of pathway

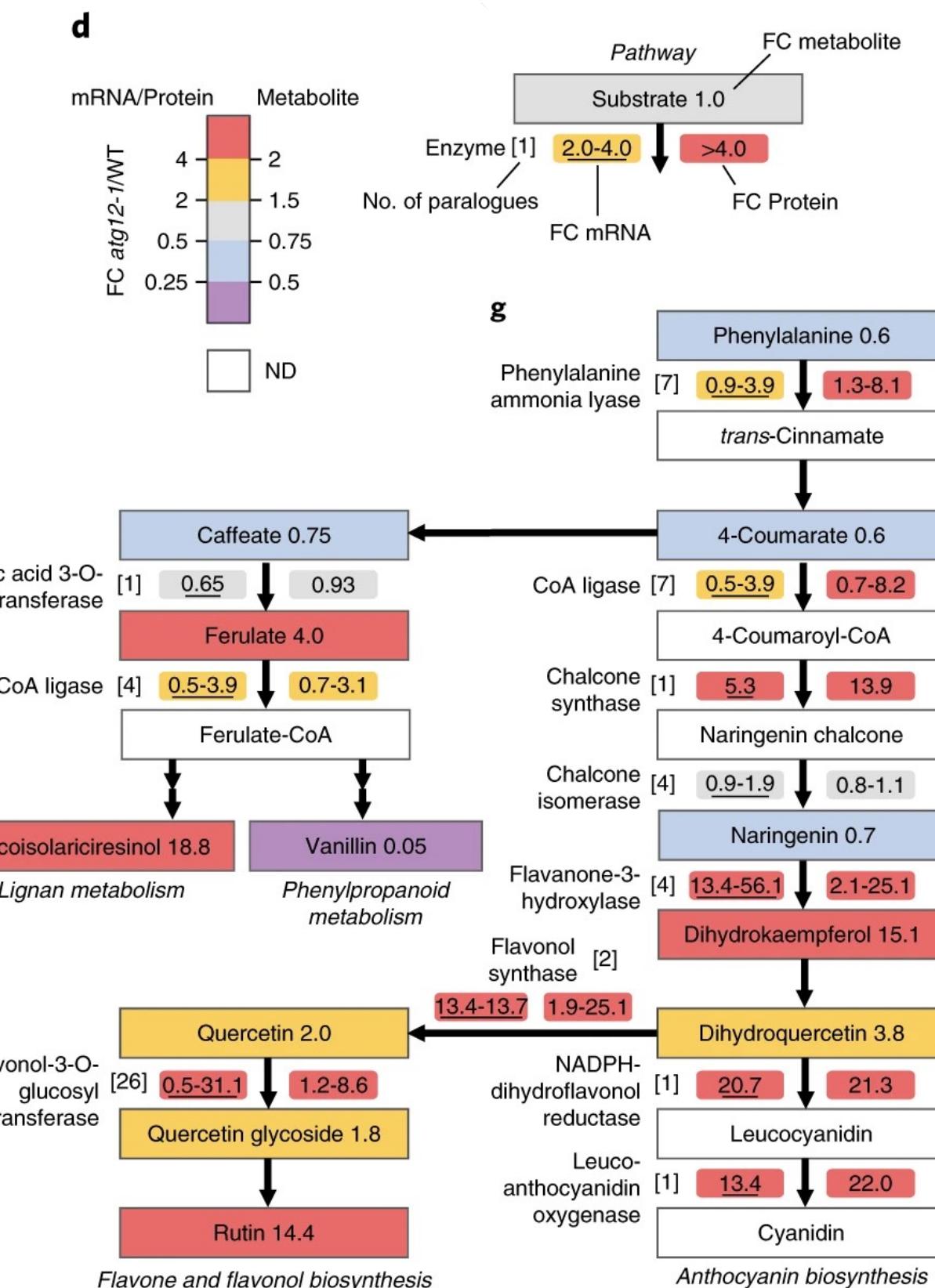
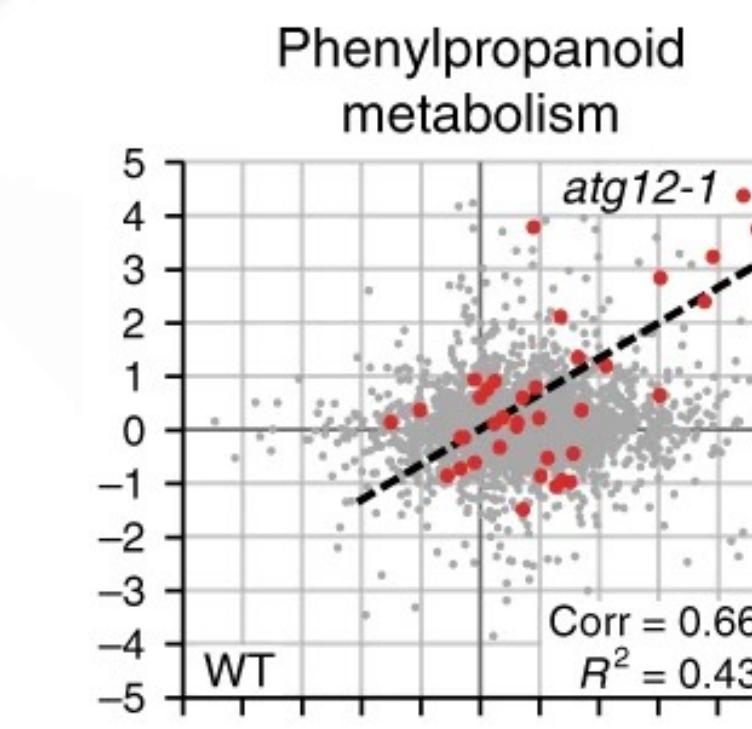


Hou et.al., 2021

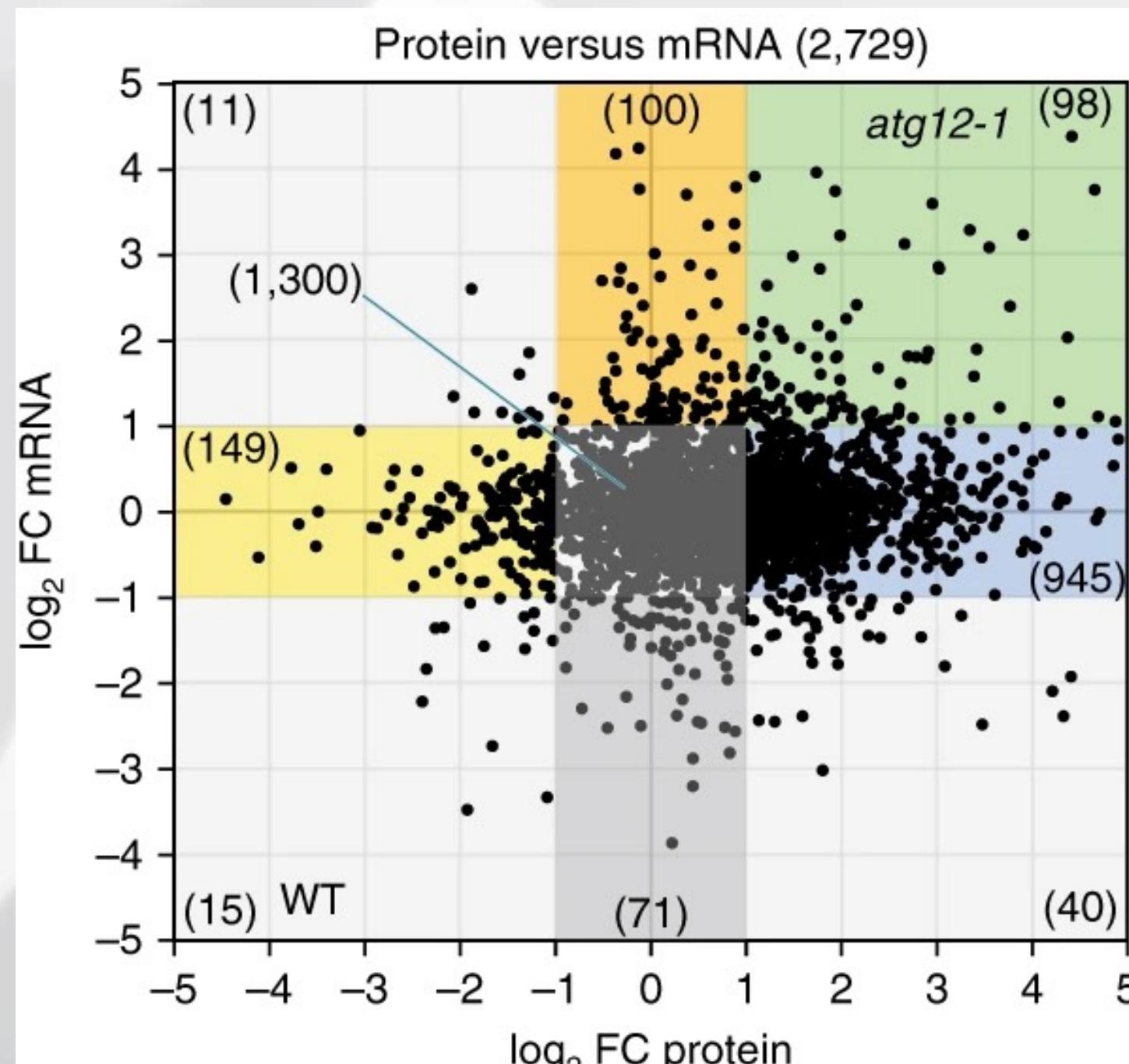
# Combination of both and more multi-omics approach



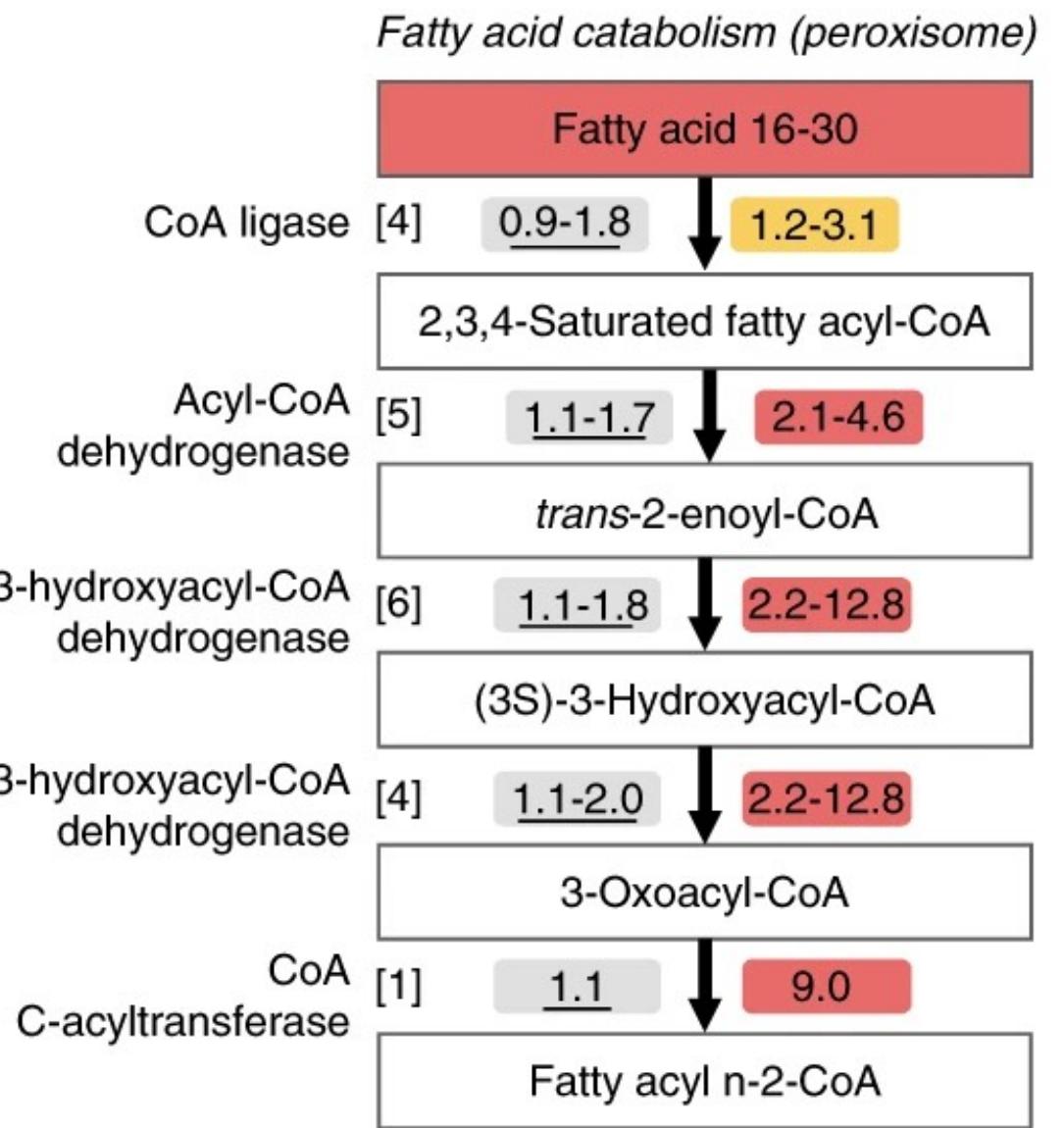
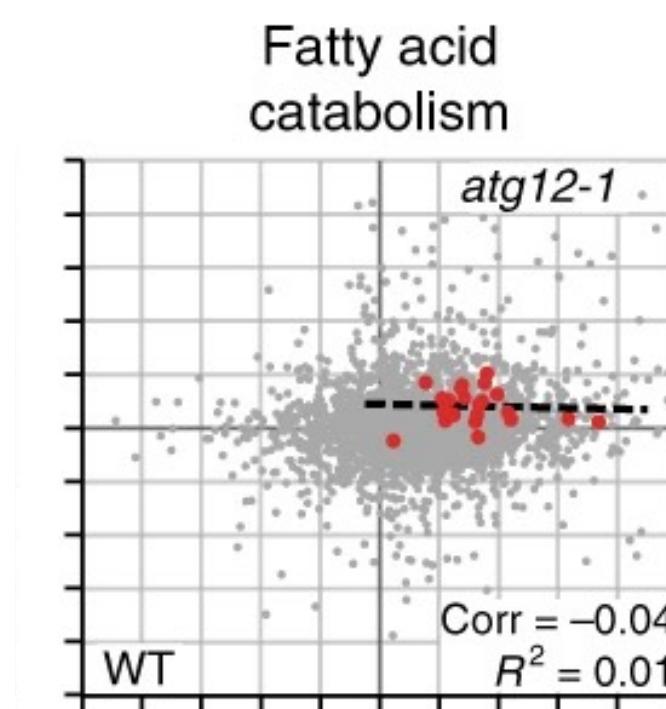
McLoughlin et.al., 2018



# Combination of both and more multi-omics approach



McLoughlin et.al., 2018



# Choose your database that fits into your question?

- SIB databases (UNIPROT)
- KEGG
- GO
- REACTOME
- BioGrid
- DIP
- CAZYme
- BioCyc
- Databases are evolving...

**Next step experiment**

# What have we learned so far:

- Pathway analysis provide an overview of understanding how cells respond to condition X
- Modulation of which components of the pathway among Y components explain condition X
- Pathway enrichment analysis as a statistical tool to test the interdependencies of X and Y, thereby providing a functional overview
- Choose the database based on your question or hypothesis, followed by statistical inference

# Lets move to Practical

Internship | Functional Bioinformatics | 01-31 July 2023

**Bioakademy**  
A UNIT OF AGT BIOSCIENCES (OPC) PVT. LTD.



Dept. of Botany, Gauhati University