Functional Analysis and Over-representation Analysis (ORA)

- Functional analysis good because …?
- Pattern matching : two questions
- Over-representation Analysis (ORA) revealed
- Starting from a specific experiment

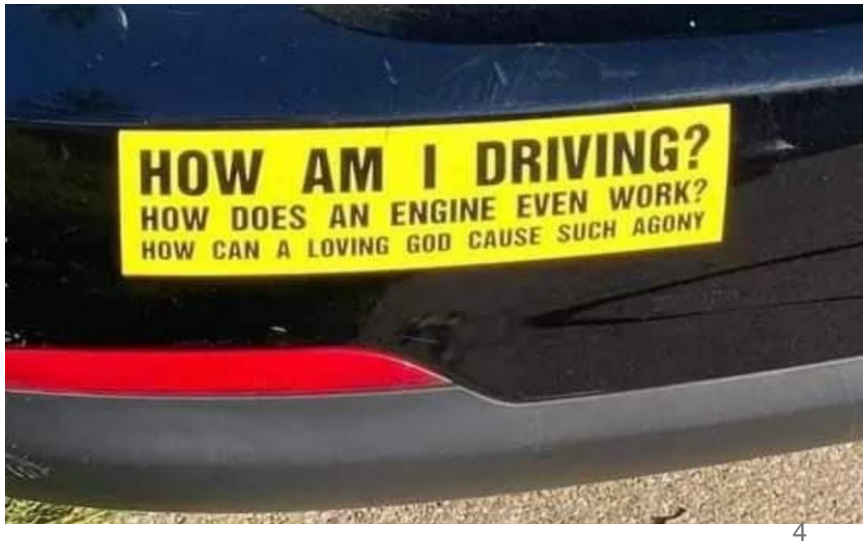
Why do we do functional analysis?

~~Why do we do functional analysis?~~

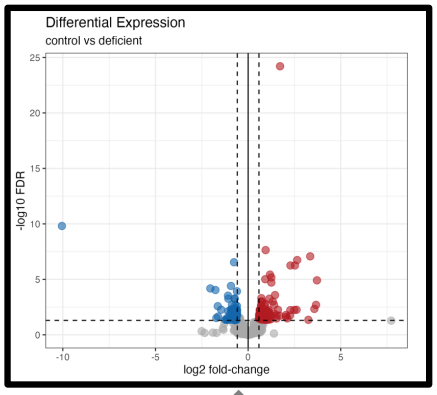
- Well, why do we do differential expression analysis?

~~Why do we do functional analysis?~~

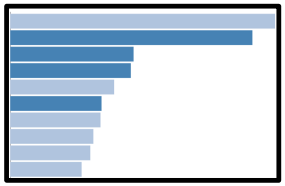
- ~~Well, why do we do differential expression analysis?~~
  - Wait - why did we run the experiment in the first place?



Why do we do functional analysis?



Biological gene set	Enrichment Ratio	P value
DNA mismatch repair	1	1
Aerobic respiration	0.2	0.021
Anaerobic respiration	1.6	0.093
Glycolytic process	3.6	2.51e-12
...		



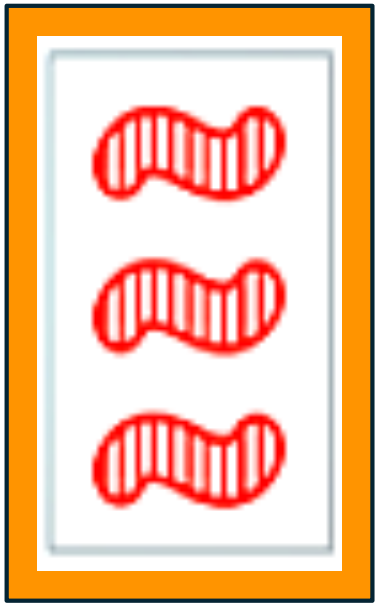
Matches observed DE patterns with biological phenomena

1) Connects results to the research question
2) Simplifies interpretation
3) Reveals coordinated changes across genes
4) Enables comparison across DE experiments

Functional Analysis matches observed DE patterns with biological phenomena

- Let’s do some pattern matching

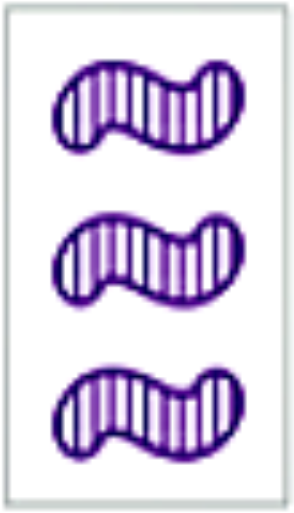
Pattern matching (1)



Is my card of interest related to the other cards?

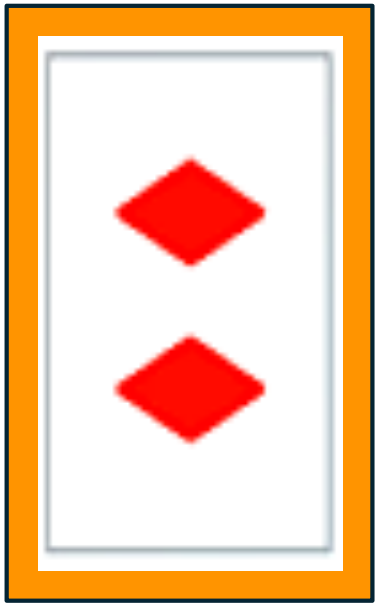
A

B



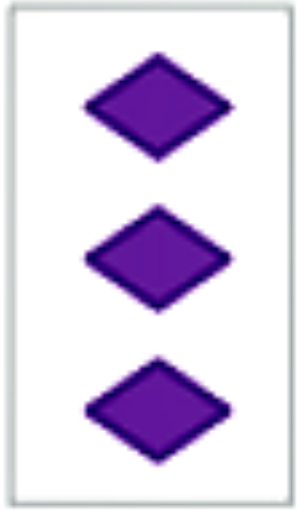
C

Pattern matching (2)



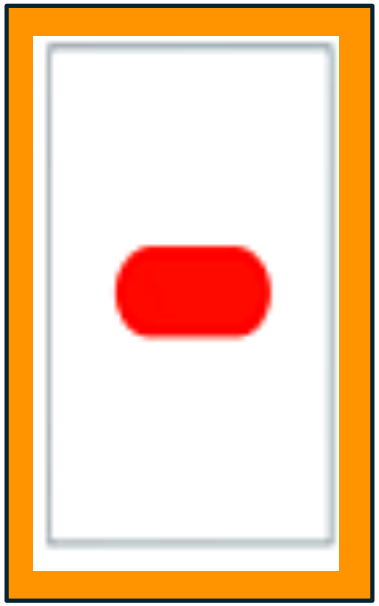
A

B

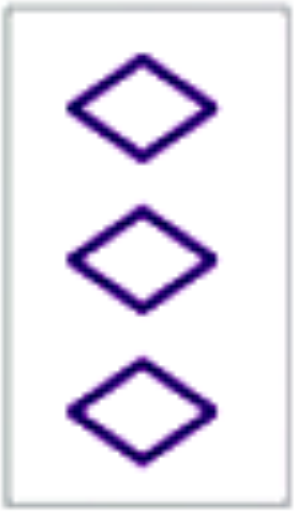


C

Pattern matching (3)



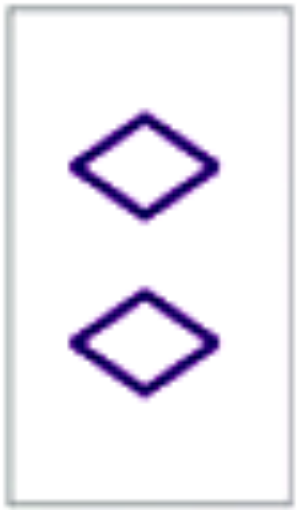
A



B

C

Pattern matching (4)

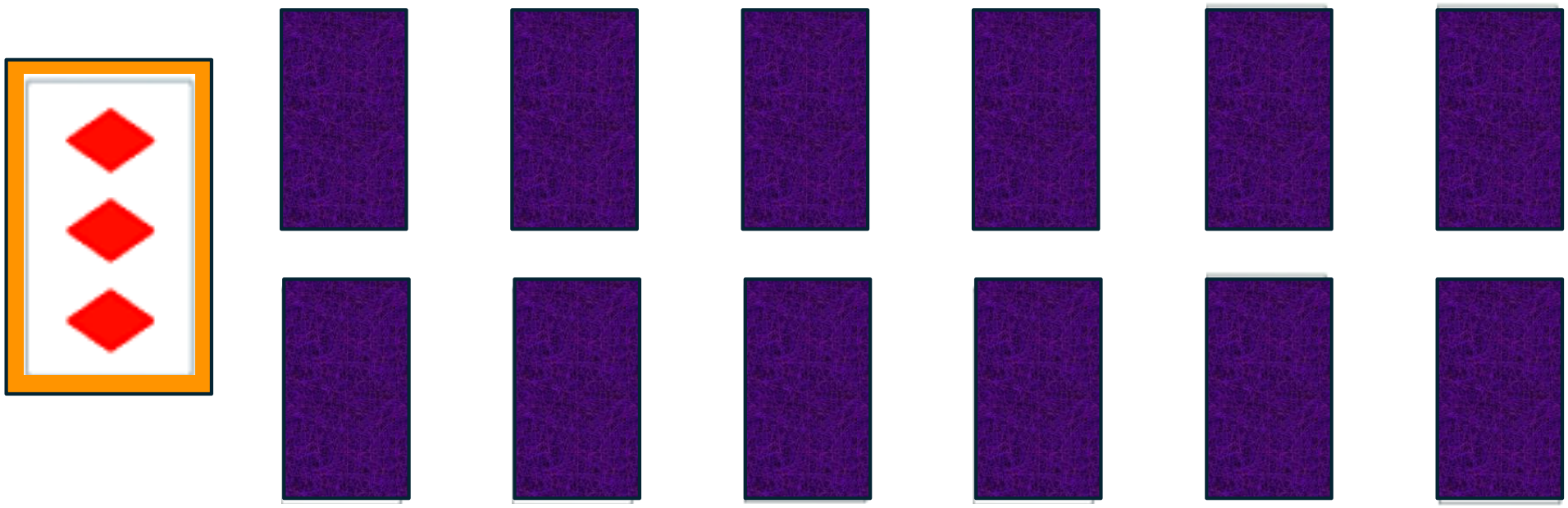


A

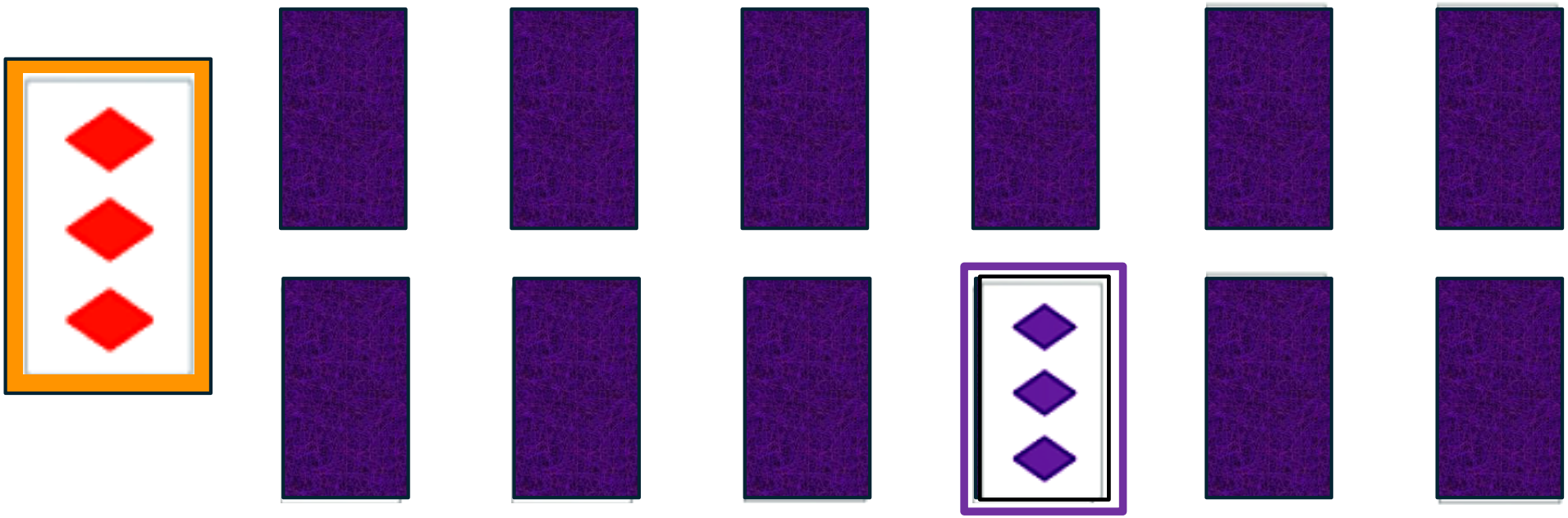
B

C

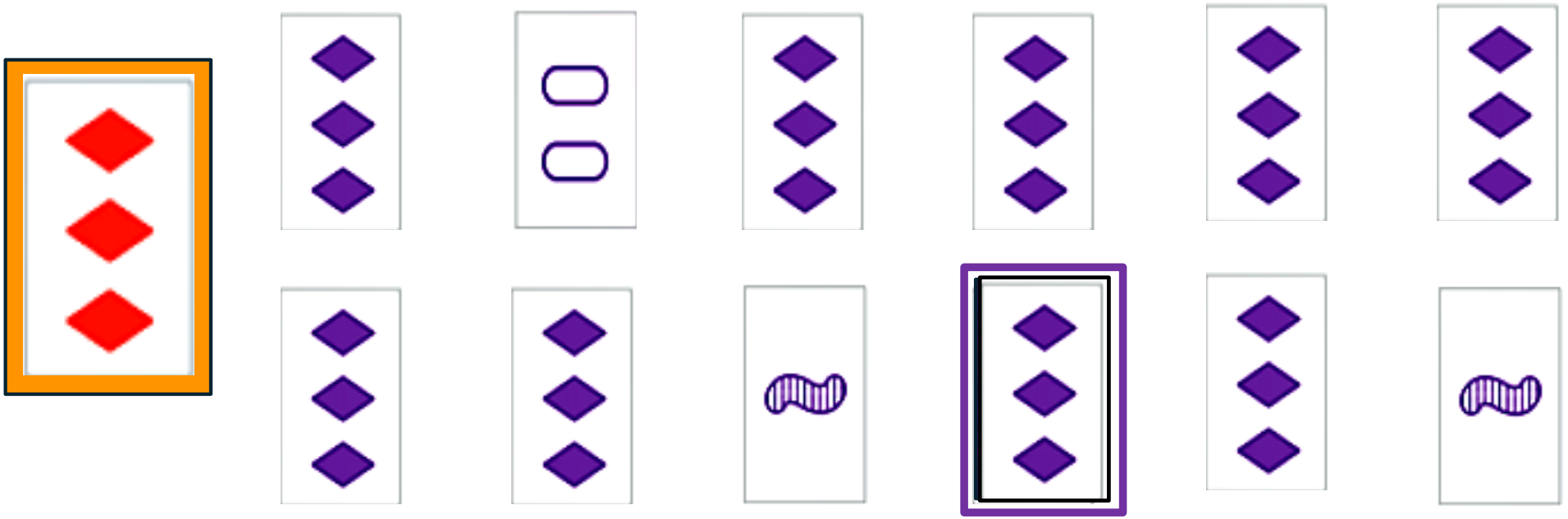
Pattern matching (5)



Pattern matching (5)

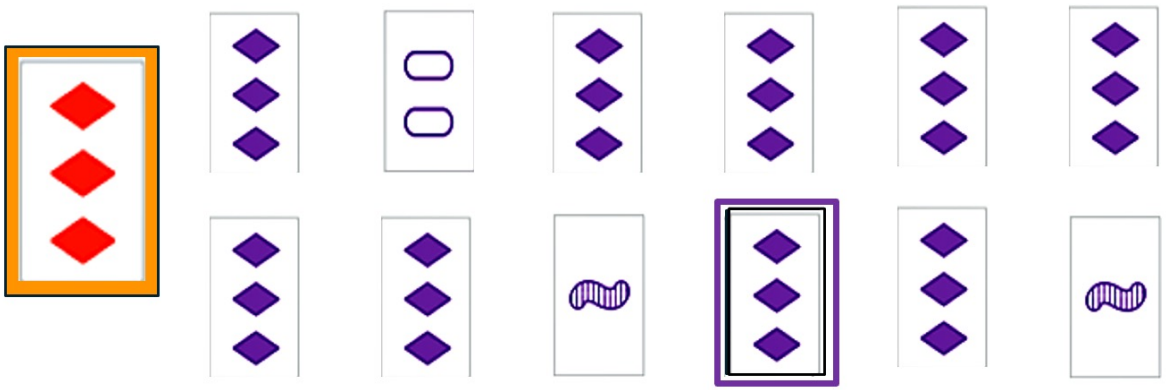


Pattern matching (5)



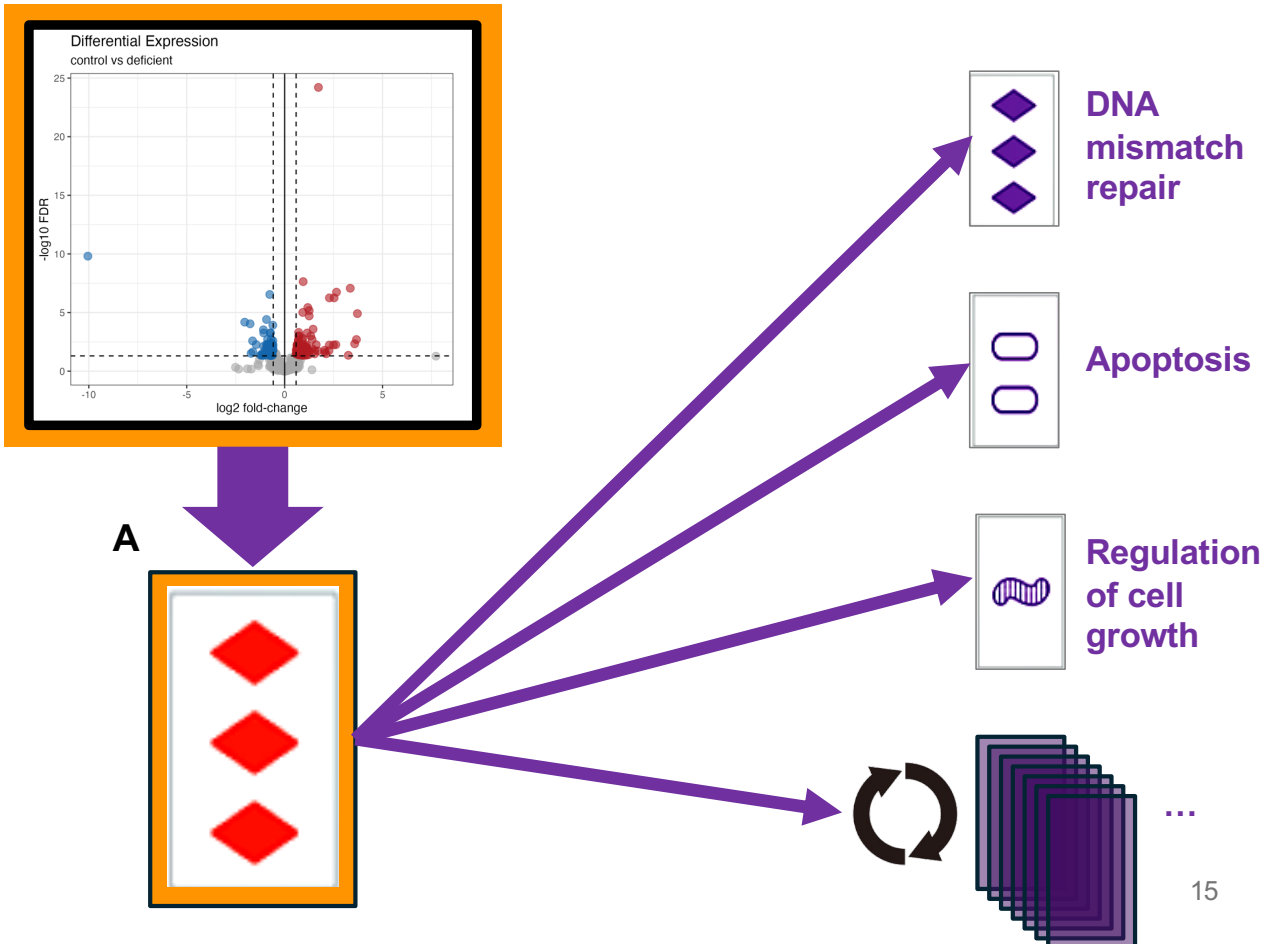
Pattern matching: Two questions

1) **Is my card of interest related to the other cards?**
2) **How surprising is that relationship?**

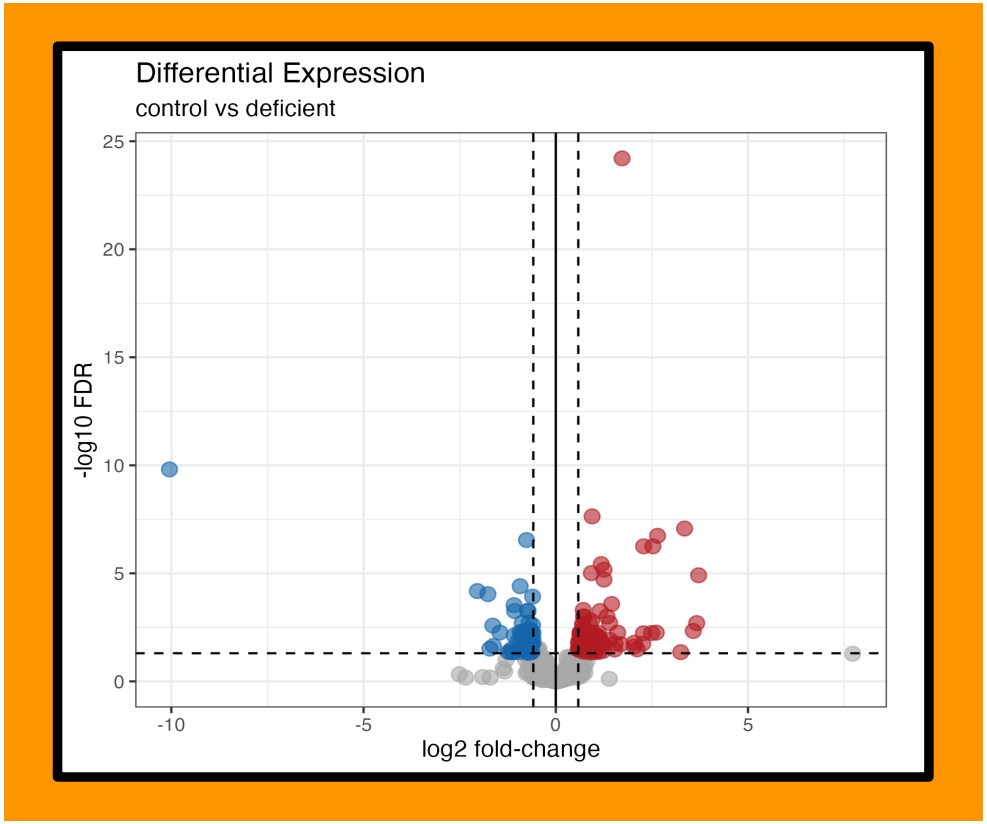


Functional Analysis is pattern matching

1) Are my DE genes of interest related to the other biological gene sets?
2) How surprising is that relationship?

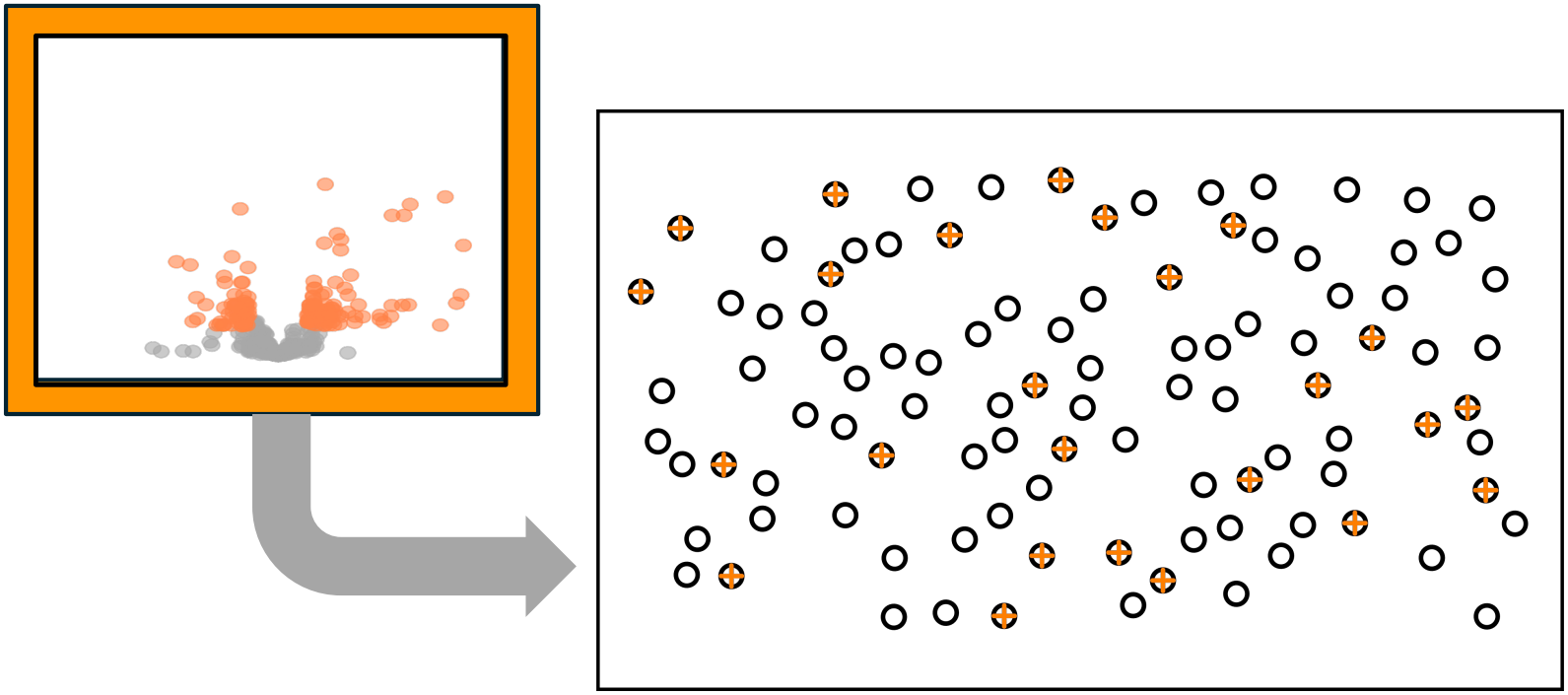


Over-representation Analysis (ORA)
Step 1: Transform observed DE pattern

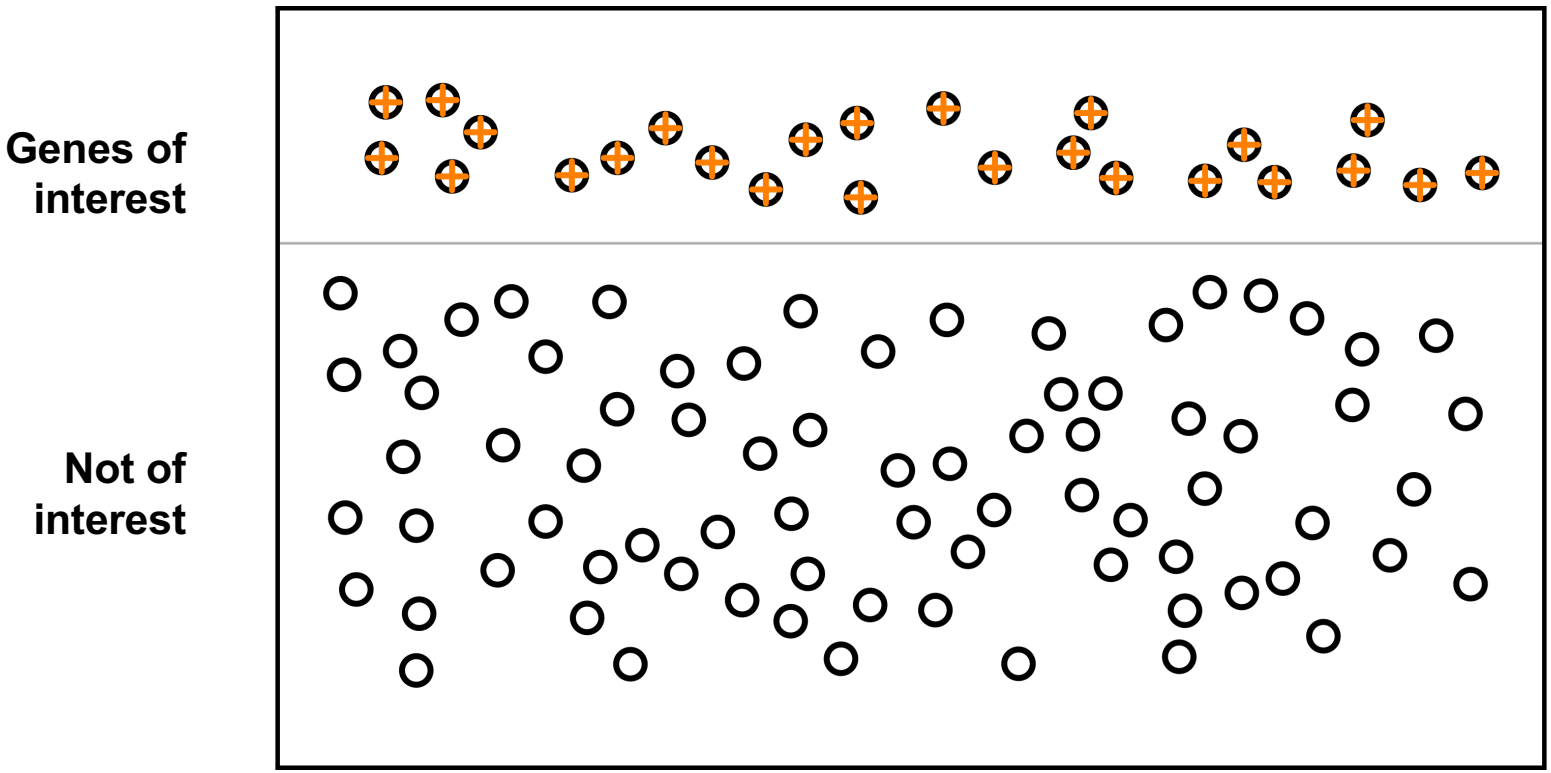


ORA: Transform observed DE pattern

A) Mark DE genes (aka “interesting genes”).
B) Dump all genes in a box

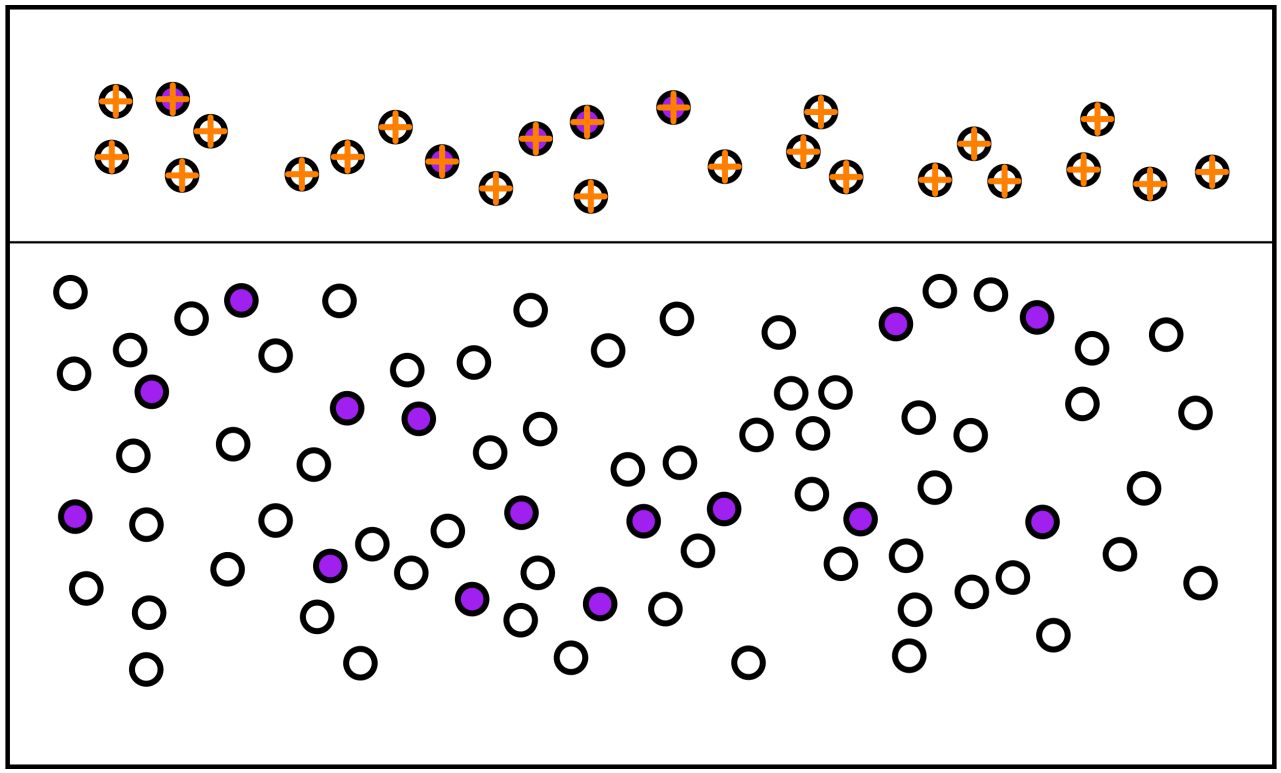


Float genes of interest to the top



DNA mismatch repair

Mark a biological gene set



DNA mismatch repair

Slide bio gene set left

	In biological gene set	Not in biological gene set
Gene set of interest		
Not of interest		

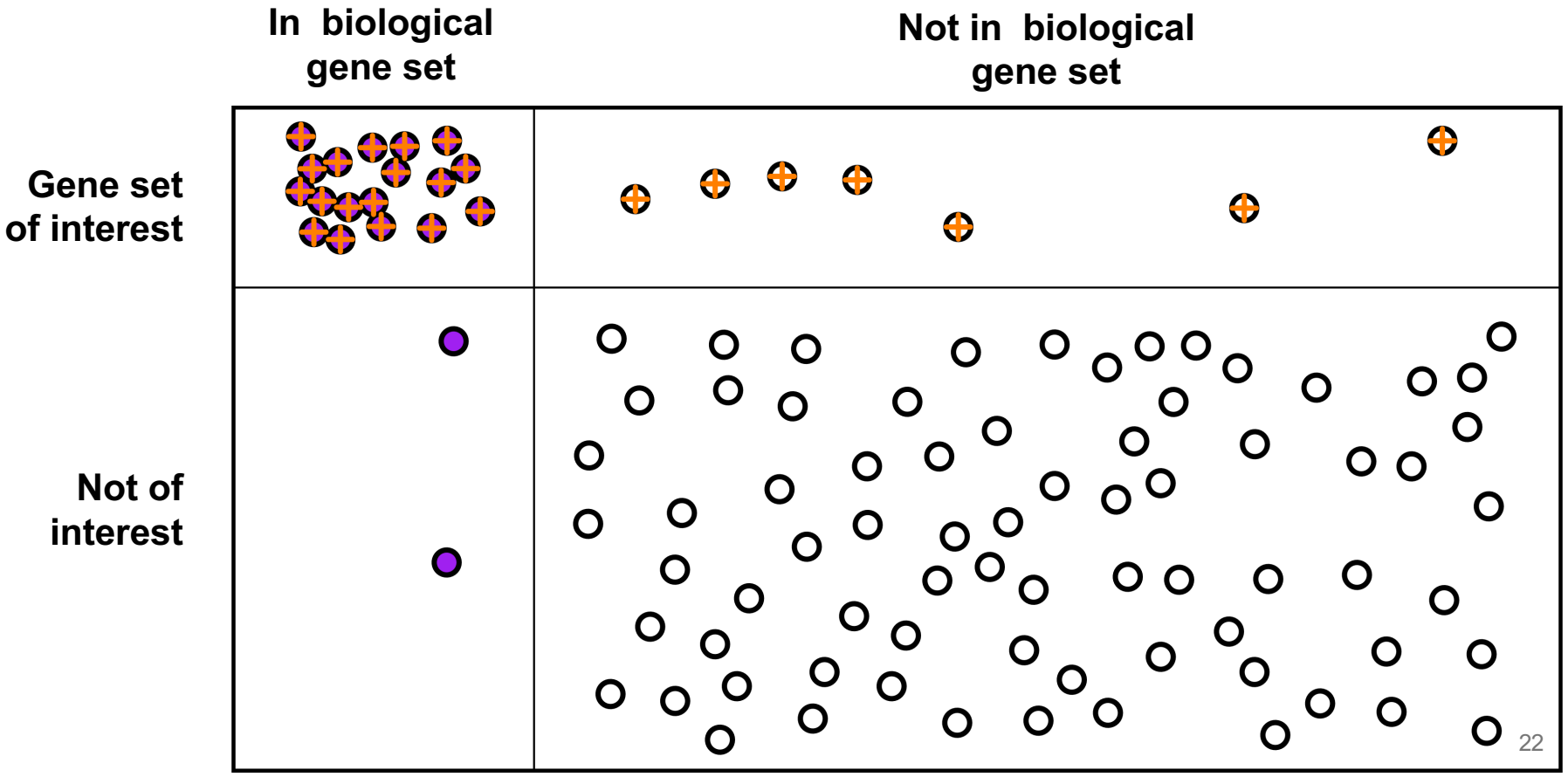
DNA mismatch repair

GSOI related to Bio GS? (1)

Gene set of interest (GSOI)	In biological gene set (BioGS)	Not in biological gene set
Not of interest		

Glycolytic process

GSOI related to Bio GS? (2)



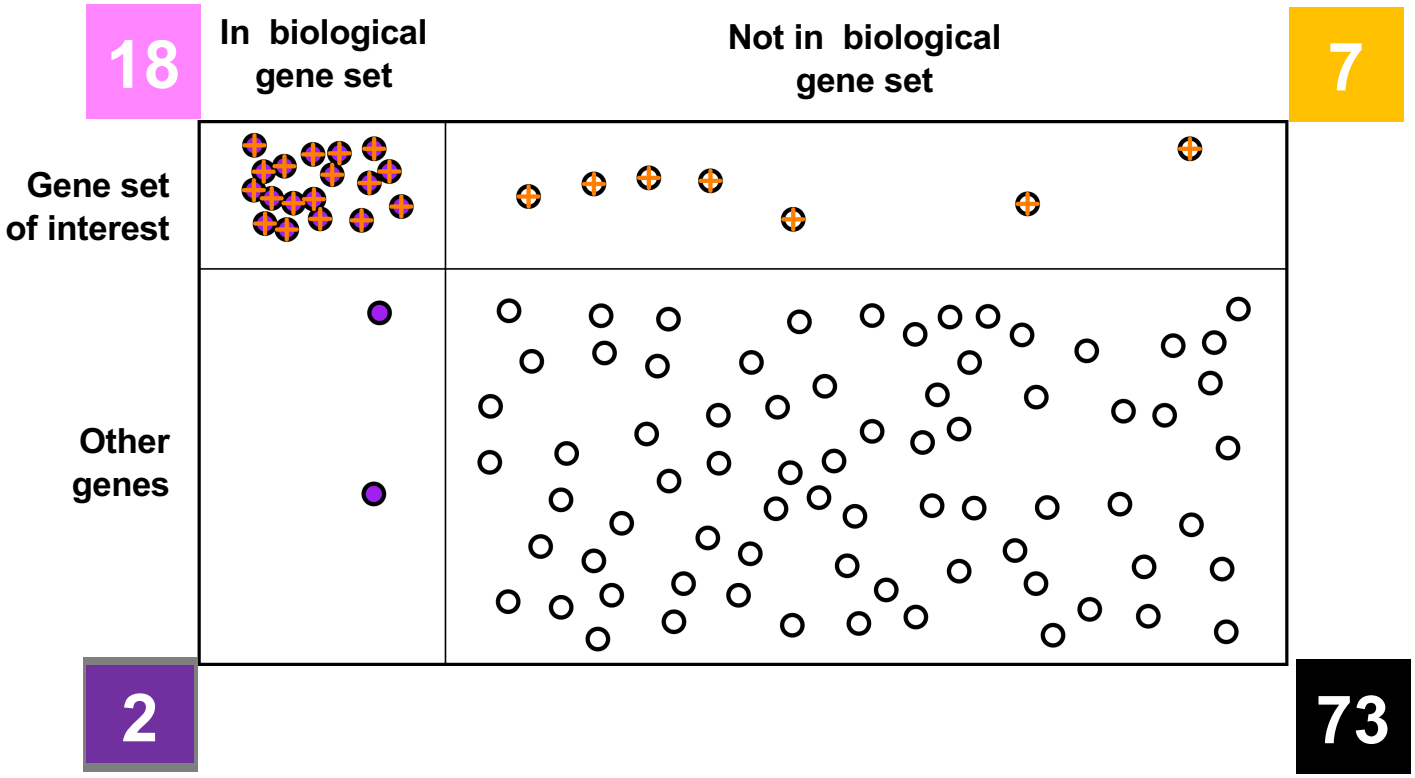
Aerobic respiration

GSOI related to Bio GS? (3)

	In biological gene set	Not in biological gene set
Gene set of interest		
Not of interest		

ORA is based on 4 numbers

Glycolytic process



4 numbers (aka “Contingency table”)

Glycolytic process			
	In biological gene set	Not in biological gene set	Total
Genes of interest	18	7	25
Not of interest	2	73	75
Total	20	80	100

What's the expected proportion?

Glycolytic process			
	In biological gene set	Not in biological gene set	Total
Genes of interest	18	7	25
Not of interest	2	73	75
Total	20	80	100

	Proportion	Count
Expected	20 / 100 = 0.20	0.20 * 25 = 5

What's the observed proportion?

Glycolytic process			
	In biological gene set	Not in biological gene set	Total
Genes of interest	18	7	25
Not of interest	2	73	75
Total	20	80	100

	Proportion	Count
Expected	20 / 100 = 0.20	0.2 * 25 = 5
Observed	18 / 25 = 0.72	18

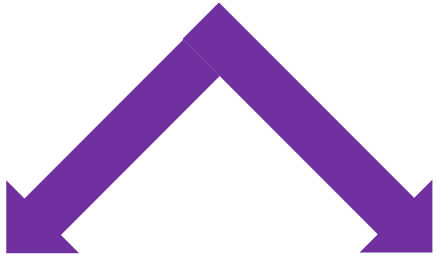
Enrichment Ratio = Observed/Expected

Glycolytic process			
	In biological gene set	Not in biological gene set	Total
Genes of interest	18	7	25
Not of interest	2	73	75
Total	20	80	100

	Proportion	Count
Expected	20 / 100 = 0.20	0.2 * 25 = 5
Observed	18 / 25 = 0.72	18
Enrichment Ratio	0.72 / 0.20 = 3.6	

Enrichment Ratio and p value

Glycolytic process			
	In biological gene set	Not in biological gene set	Total
Genes of interest	18 a	7 b	25
Not of interest	2 c	73 d	75
Total	20	80	100 n



Fisher's exact test

$$p = \frac{\binom{a+b}{a}\binom{c+d}{c}}{\binom{n}{a+c}}$$

Enrichment Ratio = 3.6

Are my genes of interest related to another other biological gene set?

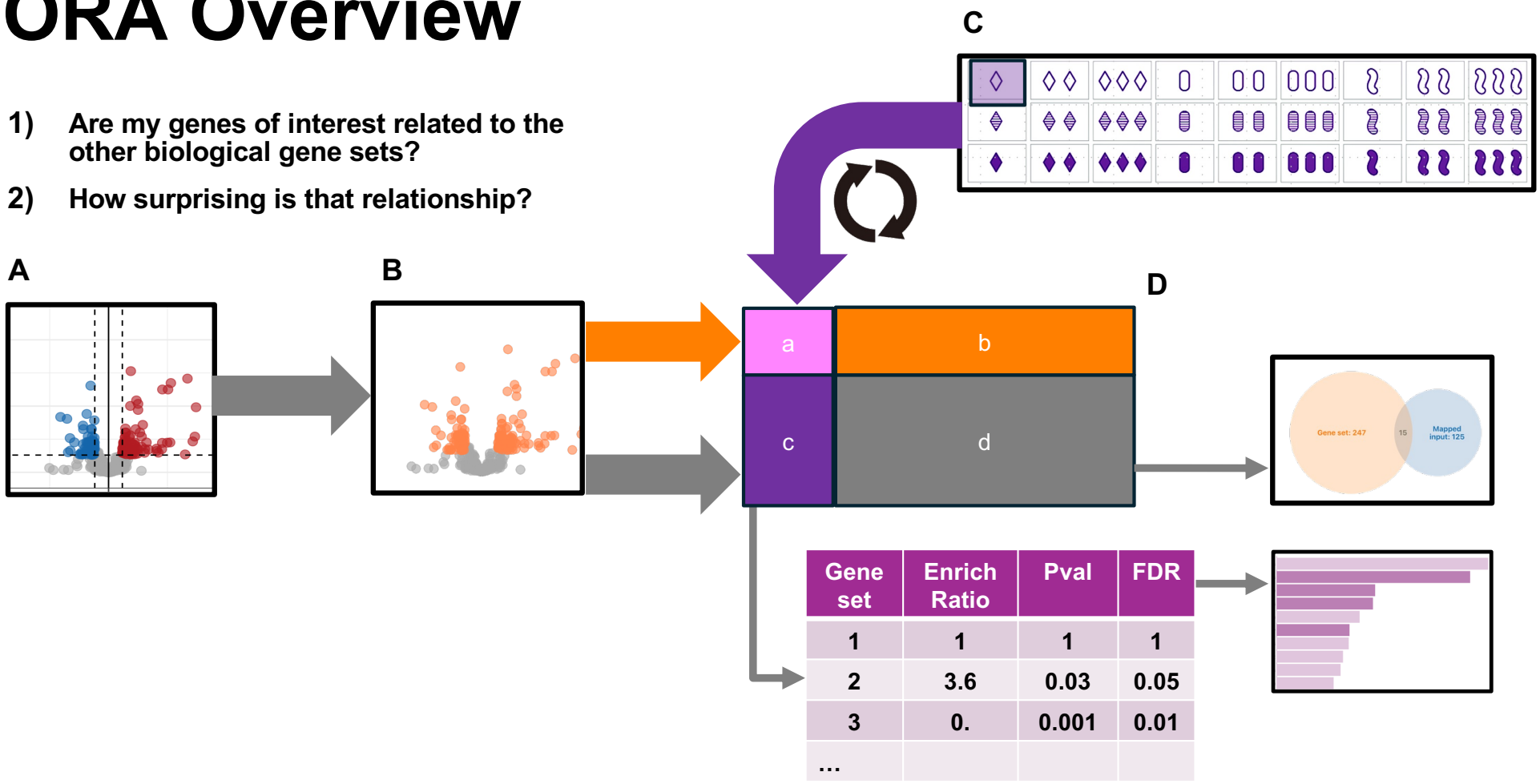
p value = 2.51e-12

How surprising is that relationship?

ORA Overview

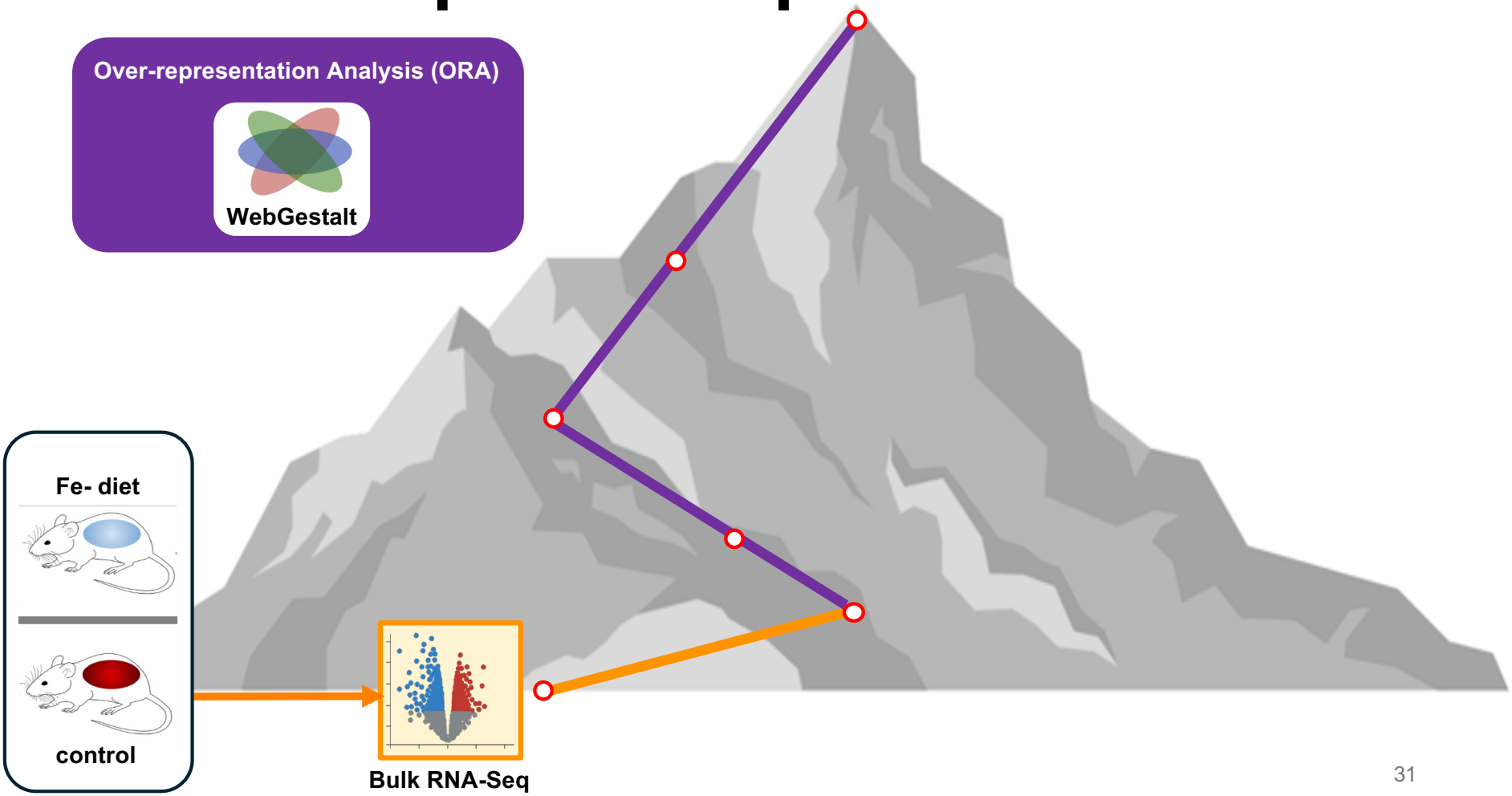
1) Are my genes of interest related to the other biological gene sets?

2) How surprising is that relationship?



Gene set	Enrich Ratio	Pval	FDR
1	1	1	1
2	3.6	0.03	0.05
3	0.	0.001	0.01
…			

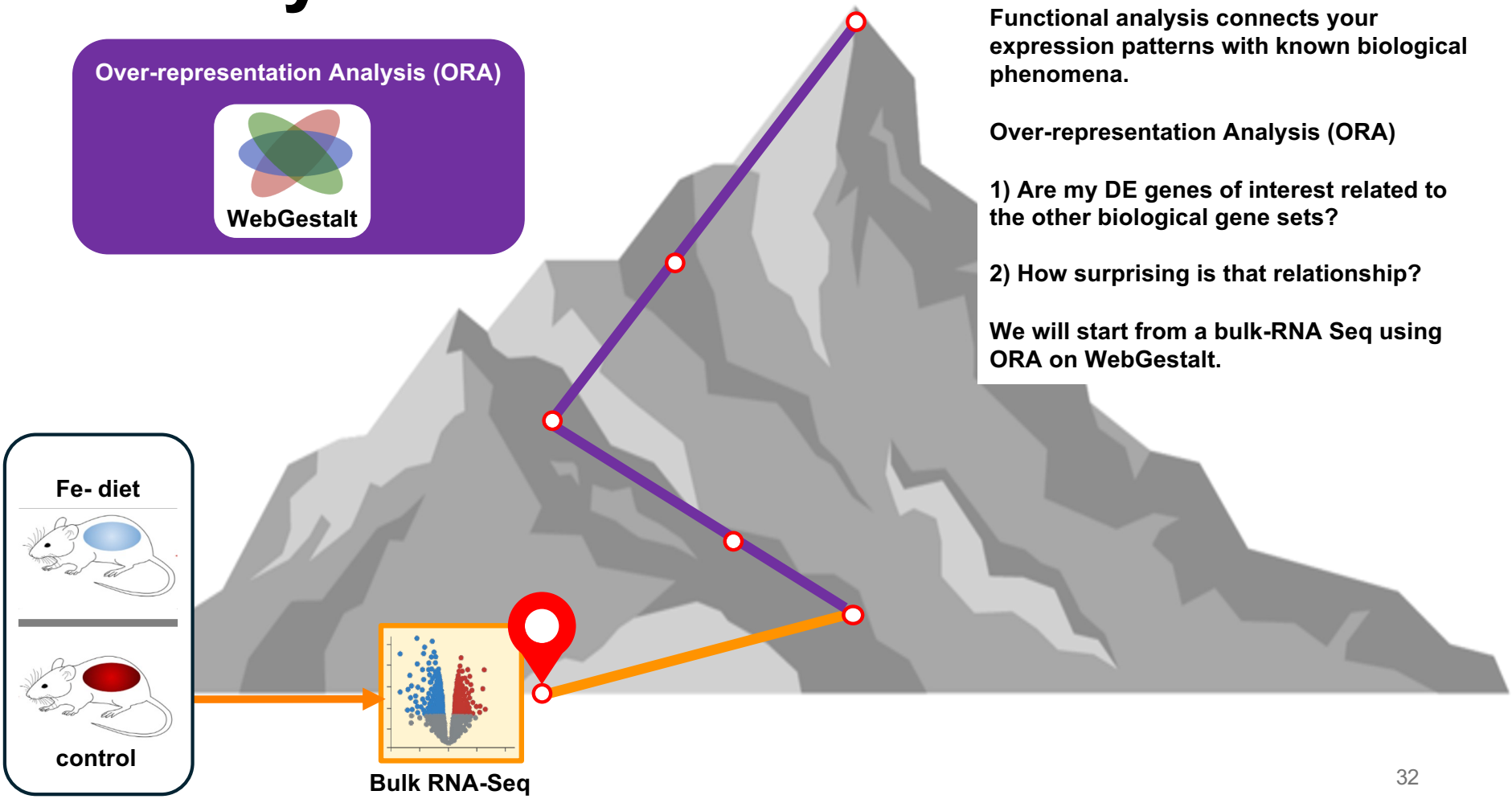
Consider a specific experiment



Summary

Over-representation Analysis (ORA)

WebGestalt



Functional analysis connects your expression patterns with known biological phenomena.

Over-representation Analysis (ORA)

1) Are my DE genes of interest related to the other biological gene sets?

2) How surprising is that relationship?

We will start from a bulk-RNA Seq using ORA on WebGestalt.