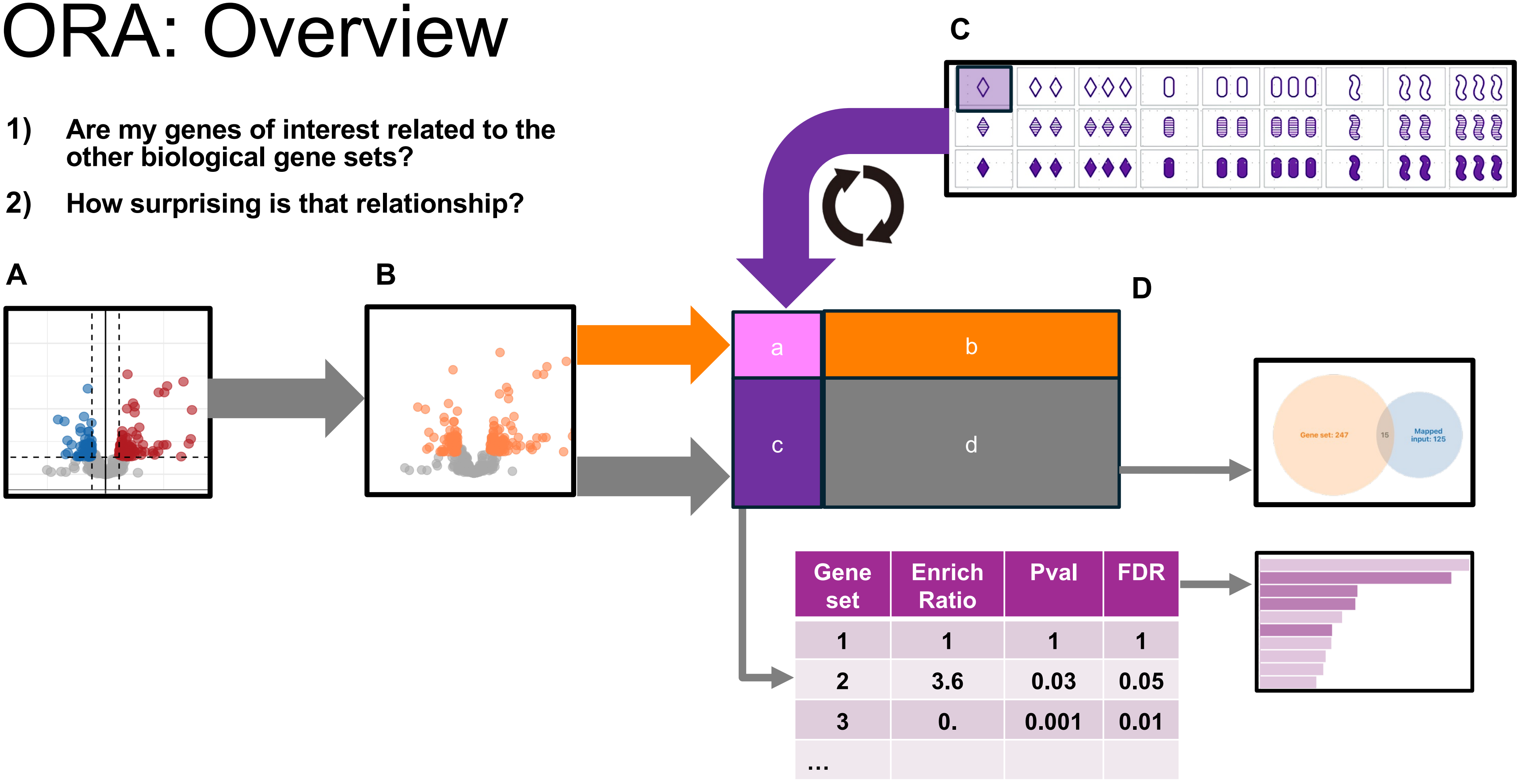
Reviewing ORA and Introducing GSEA

- ORA Review
- ORA pros & cons
- Introducing GSEA
- Consider a scRNA-Seq experiment

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
…			

ORA Pros vs. Cons

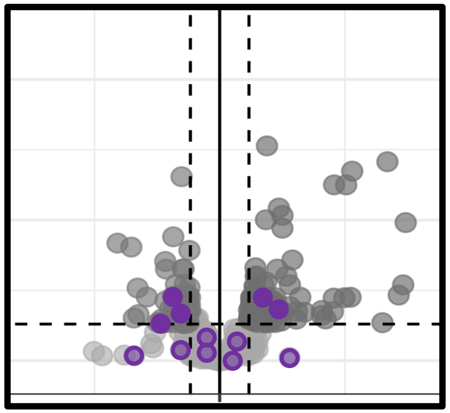
Pros

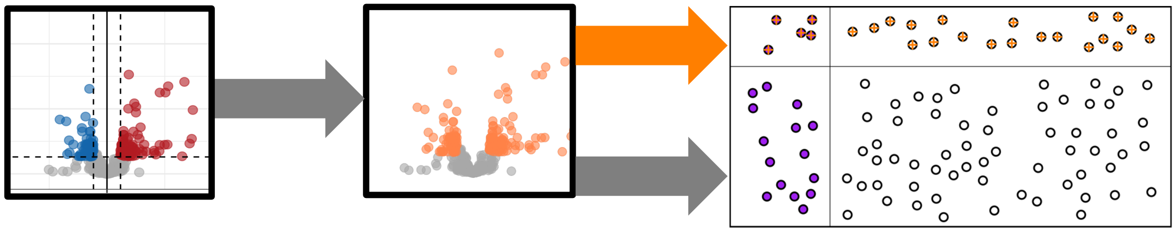
- Very mature & broadly adopted
- Mathematically straightforward
- Computationally light & very fast
- Few assumptions of inputs or biological gene sets
  - Biological gene sets can include many genes
  - There are many compatible biological gene sets

ORA Pros vs. Cons

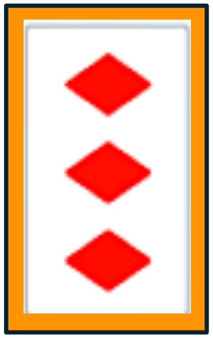
Cons

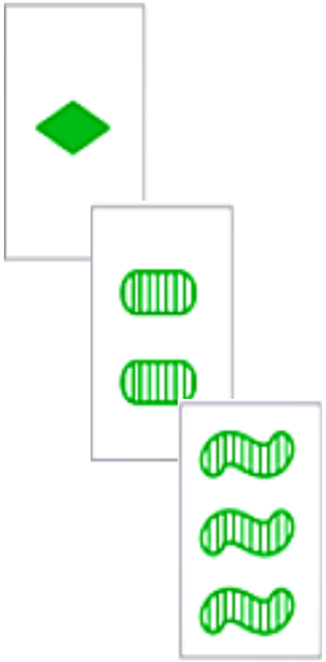
- Have to set the “background” correctly
- Have to pick a cutoff
  - too few interesting genes
  - too many interesting genes
  - biological gene sets straddle the cutoff
- Much info is discarded

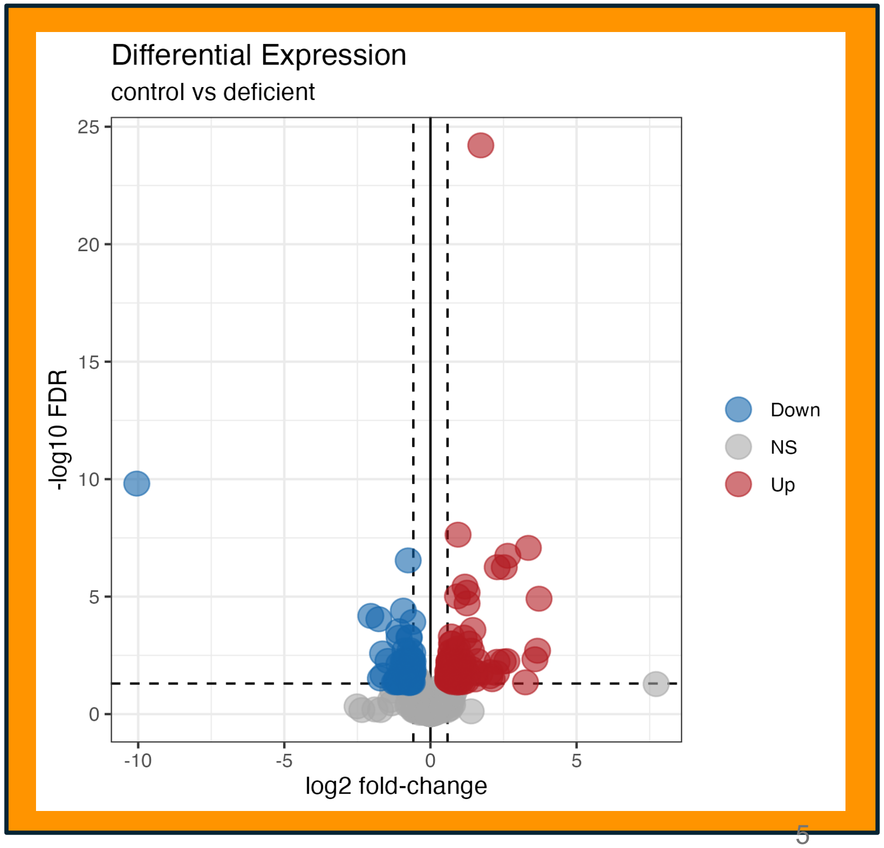




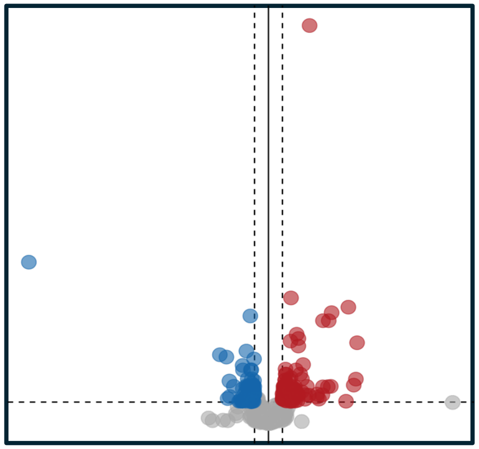
Gene Set Enrichment Analysis (GSEA)

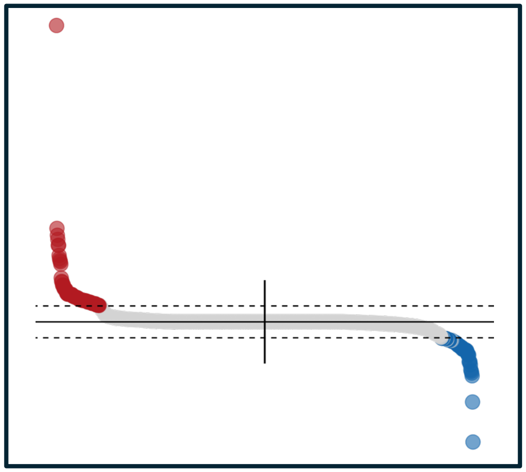




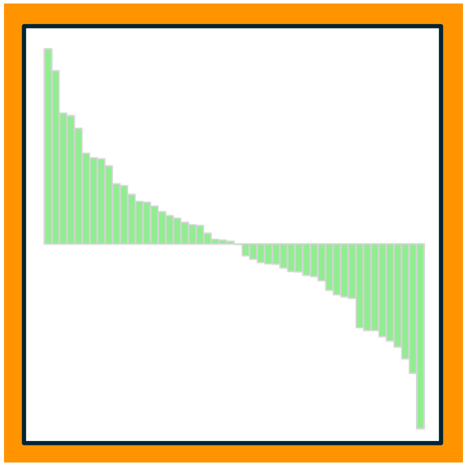


GSEA: Transform DE results to ranked list





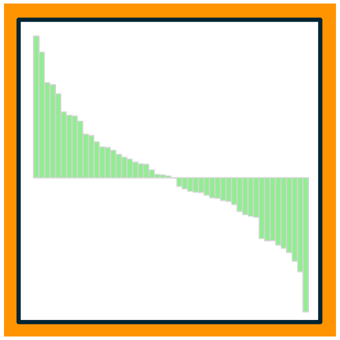
Log2foldchange
adj p value
→ **rank metric**

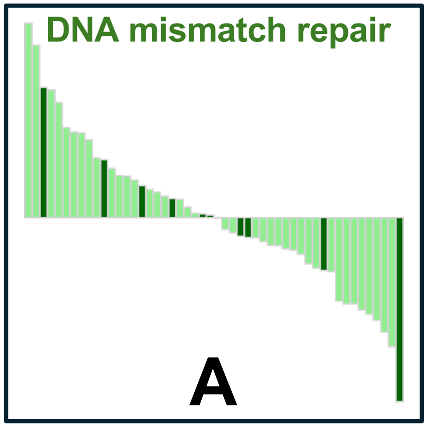


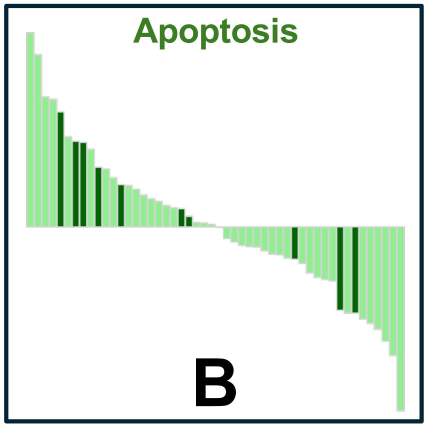
Consider all genes

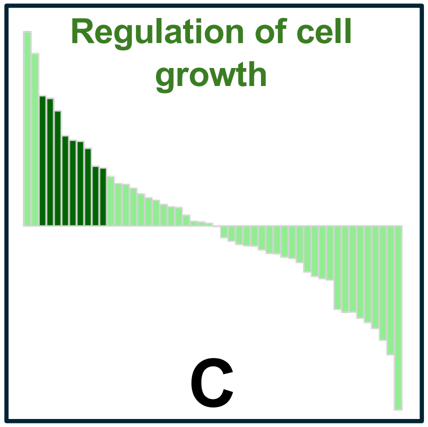
GSEA: Consider where bio gene sets fall on my ranked list

Is my gene expression pattern related to other biological gene sets?

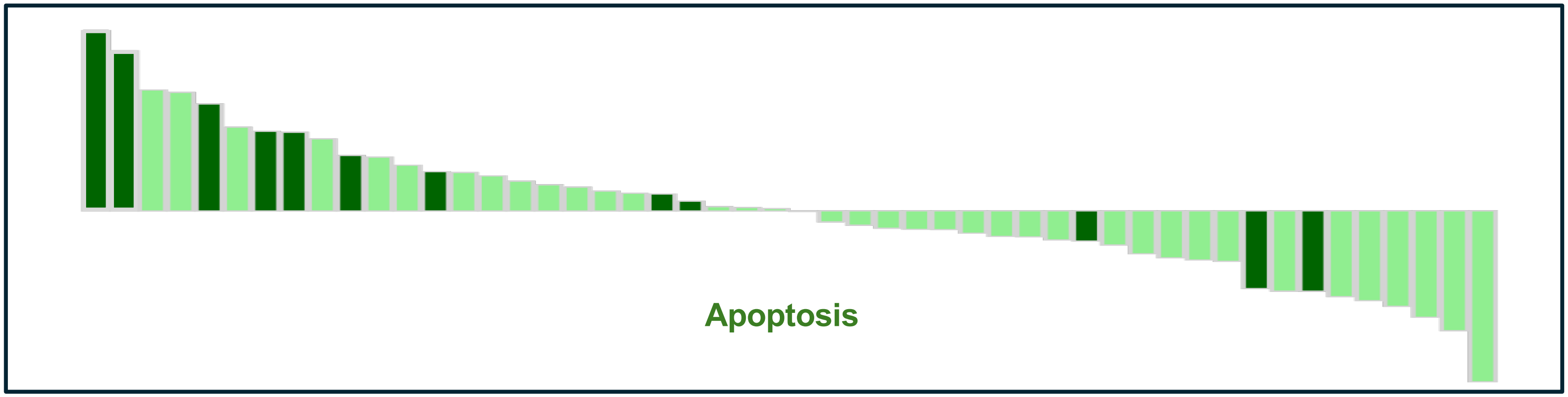




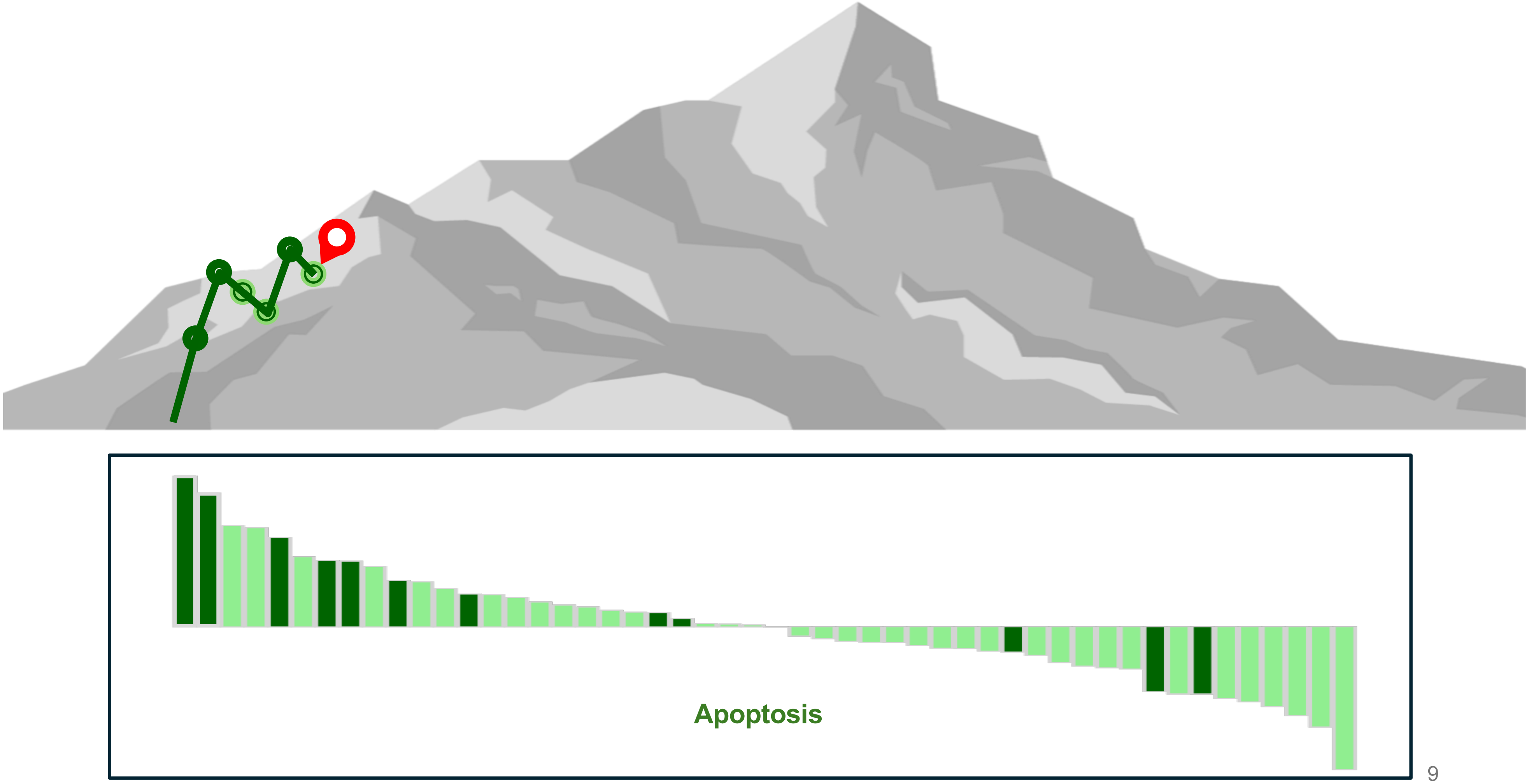




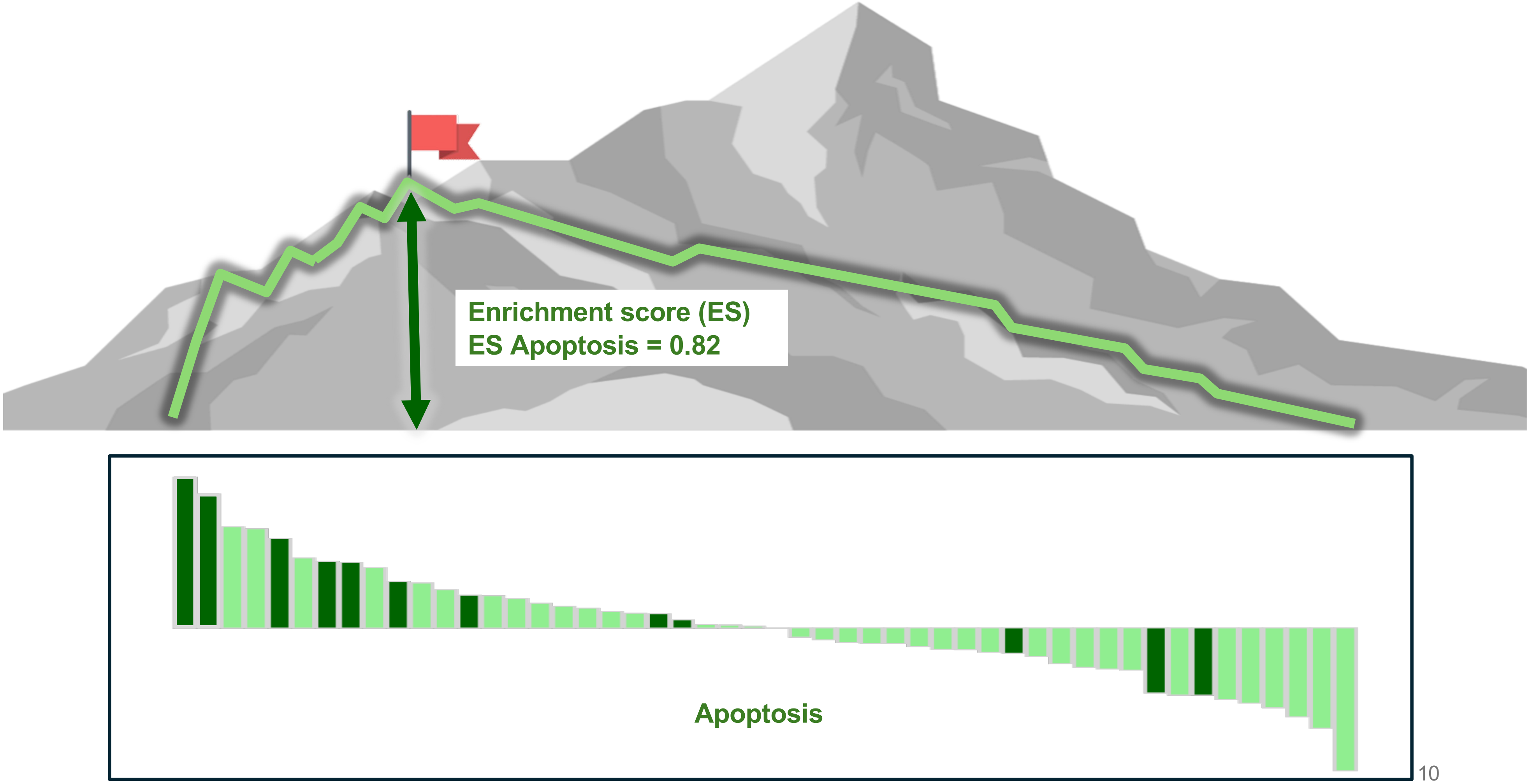
Calculate an enrichment score



Calculate an enrichment score



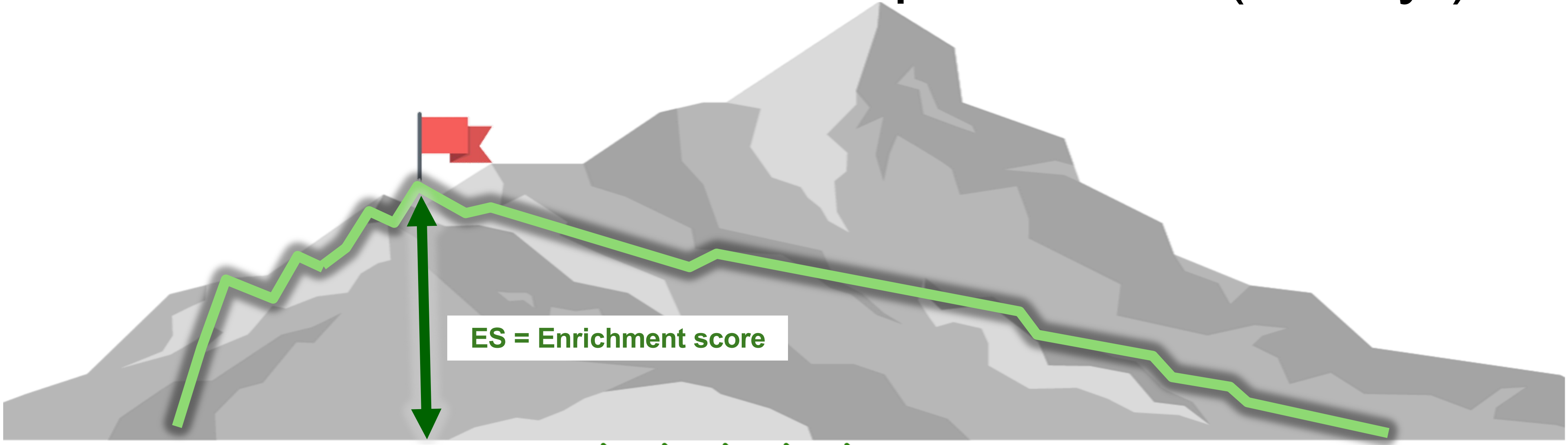
Calculate an enrichment score



Enrichment score (ES)
ES Apoptosis = 0.82

Apoptosis

Enrichment score →NES, pVal, FDR (oh my!)



Permutation test ⤨⤨⤨⤨⤨ … x1000

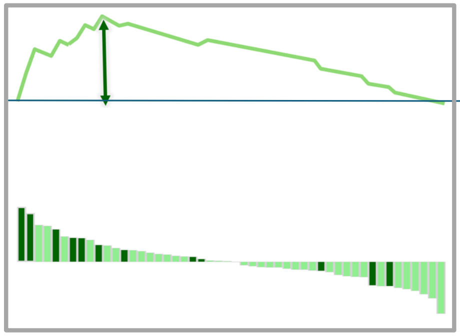
Normalized Enrichment Score (NES)

Nominal p-value → FDR

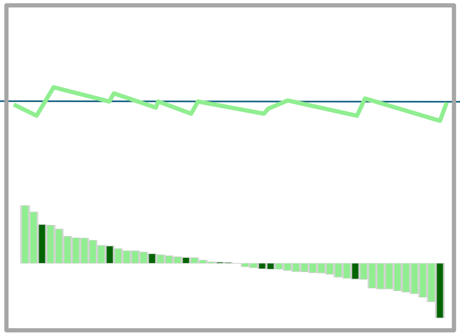
Is my gene expression pattern related to a biological gene set?

How surprising is that relationship?

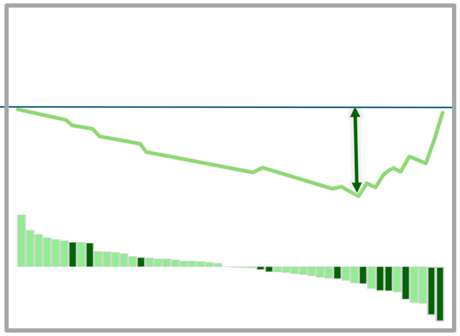
The biological gene set is enriched “where”?



ES > 0
gs enriched at top



ES ~= 0
not enriched

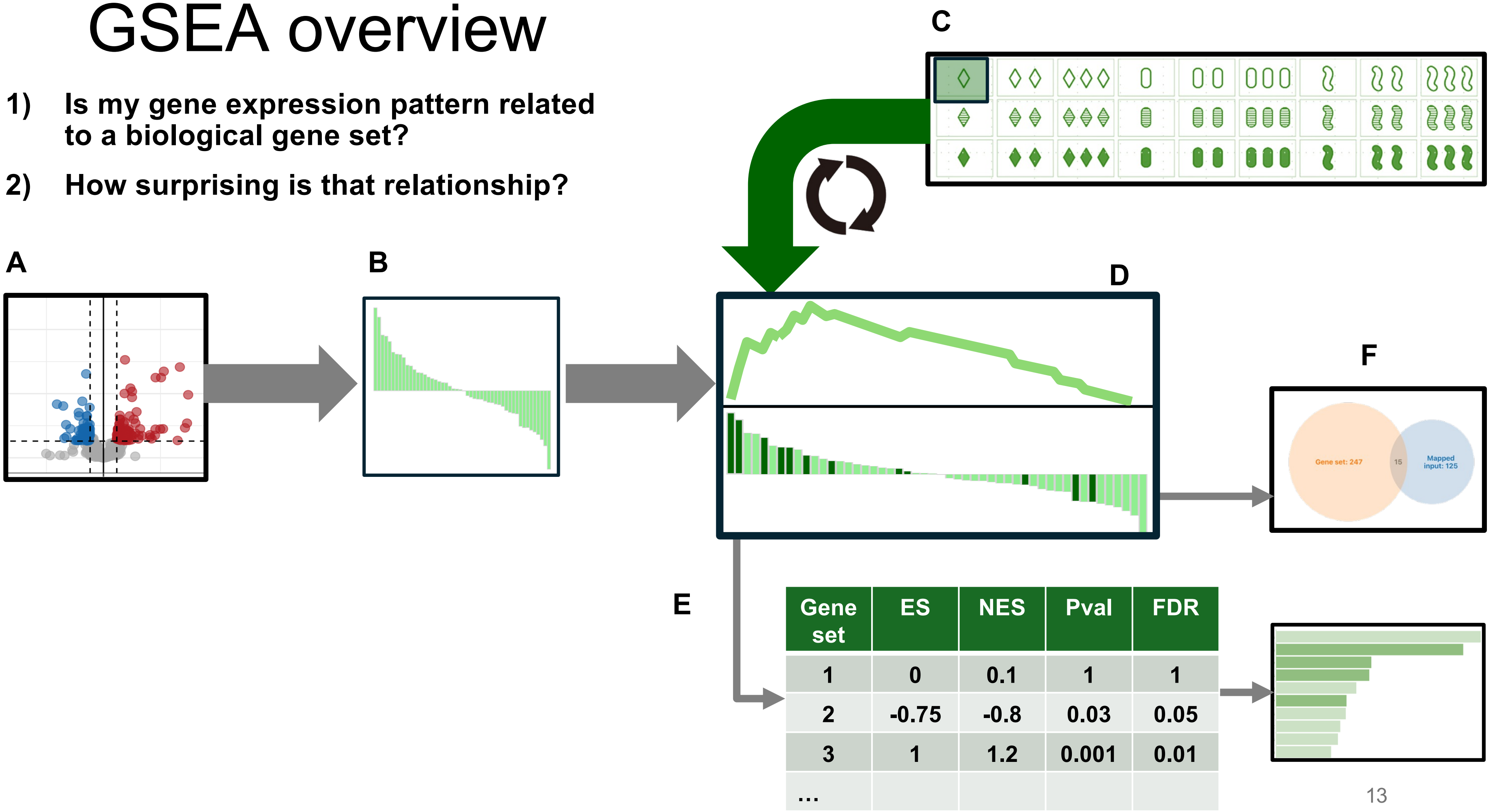


ES < 0
gs enriched at bottom

NES correction will push the limits a teensy bit, so you may see NES < -1 or > 1

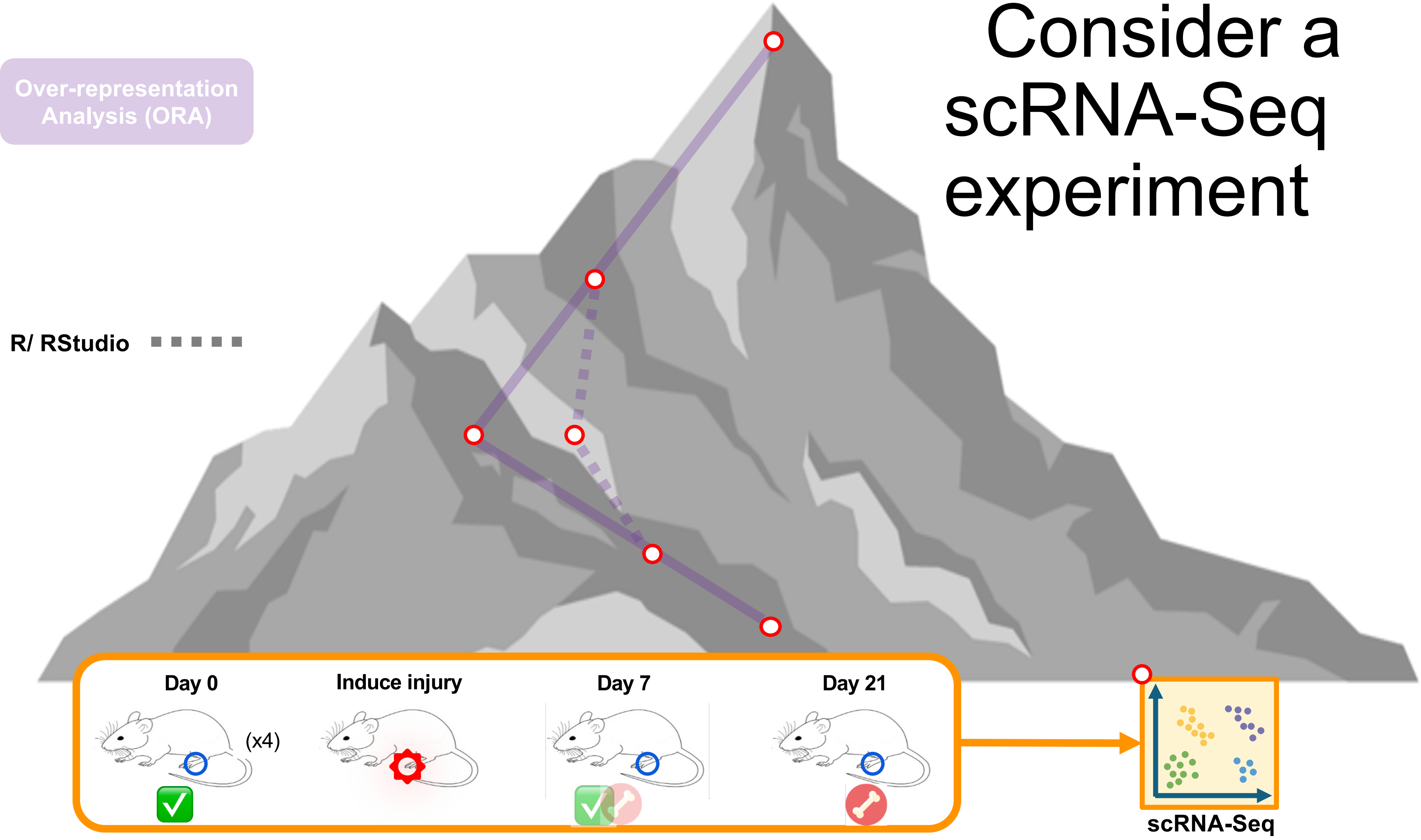
GSEA overview

1) **Is my gene expression pattern related to a biological gene set?**
2) **How surprising is that relationship?**



Gene set	ES	NES	Pval	FDR
1	0	0.1	1	1
2	-0.75	-0.8	0.03	0.05
3	1	1.2	0.001	0.01
…				

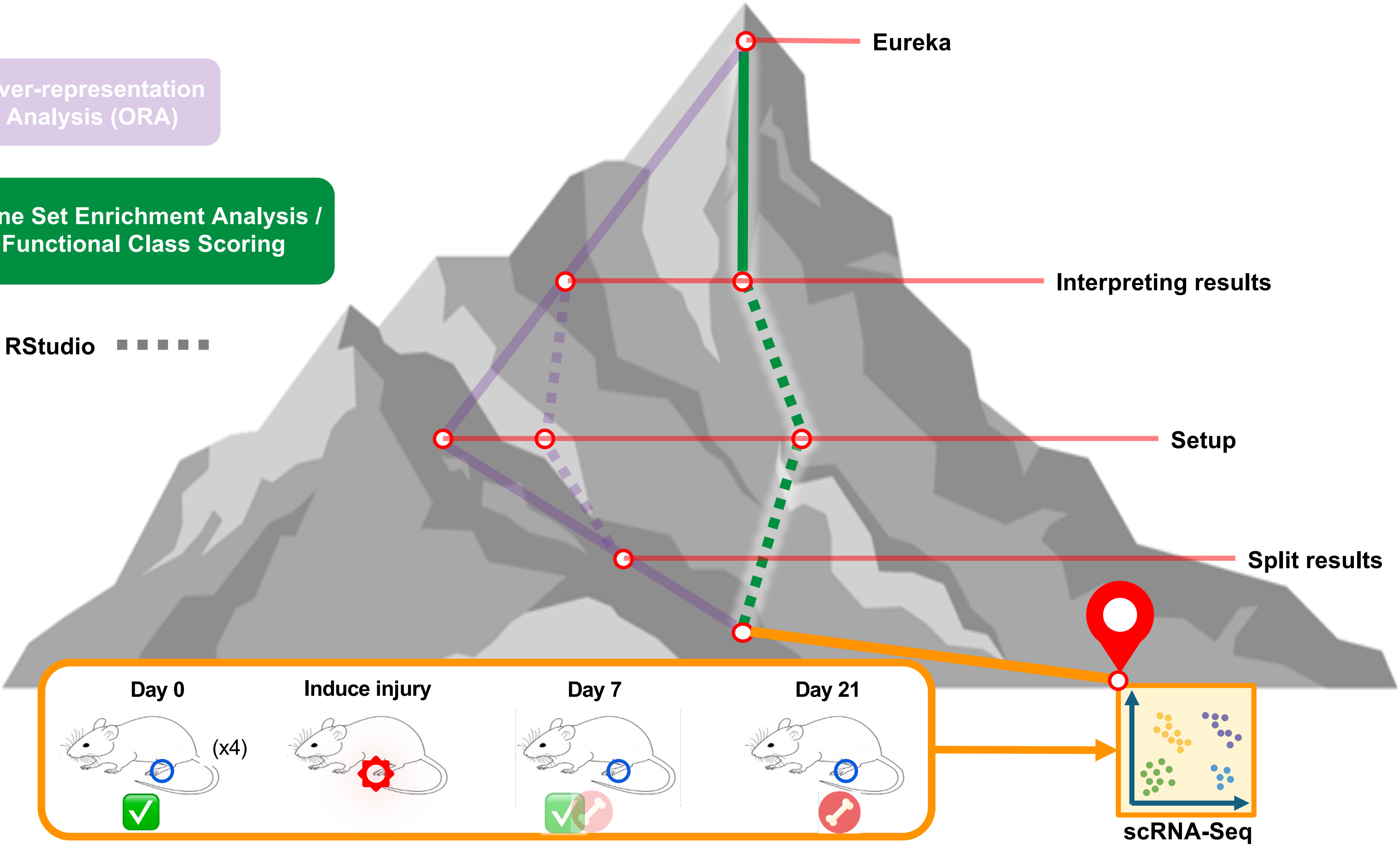
Consider a scRNA-Seq experiment



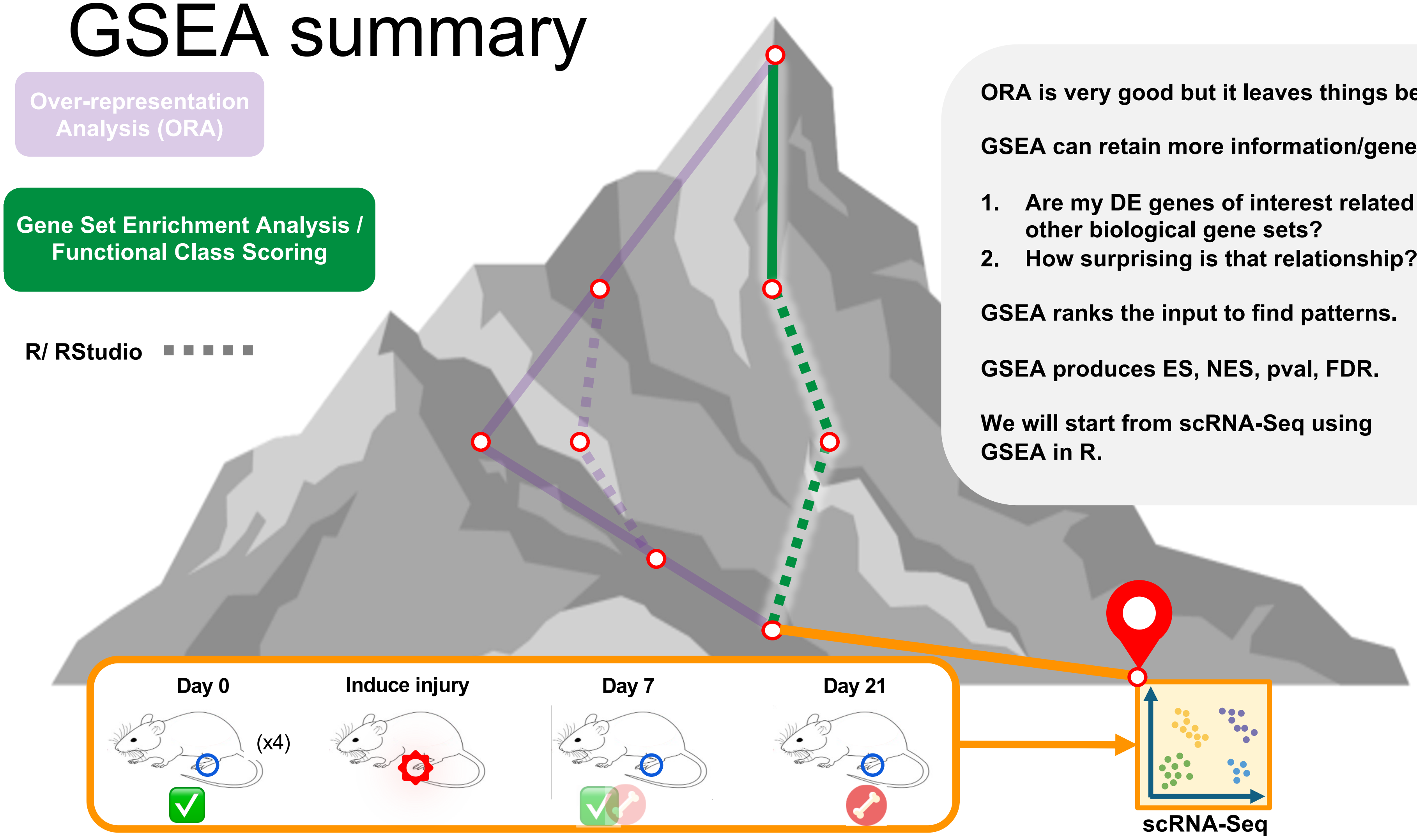
Over-representation Analysis (ORA)

Gene Set Enrichment Analysis / Functional Class Scoring

R/ RStudio ■ ■ ■ ■ ■



GSEA summary



ORA is very good but it leaves things behind.

GSEA can retain more information/genes.

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

GSEA ranks the input to find patterns.

GSEA produces ES, NES, pval, FDR.

We will start from scRNA-Seq using GSEA in R.