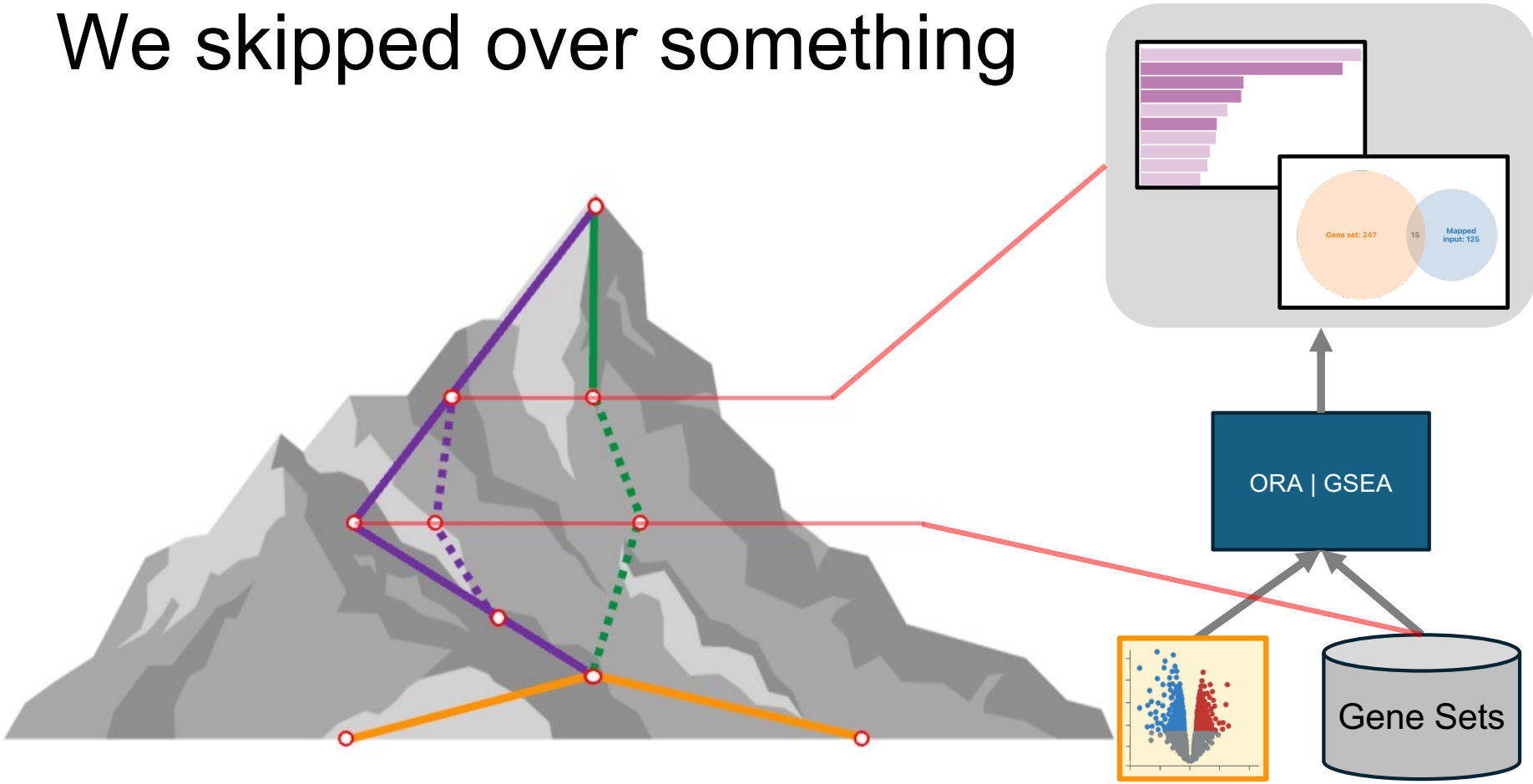
GO, KEGG, and Friends

- Whence gene sets?
- Attributes of a GO gene set
- Other knowledge bases
- 2025 functional analysis ecosystem
- A short digression on topology analysis

We skipped over something



Set the wayback machine for 1995

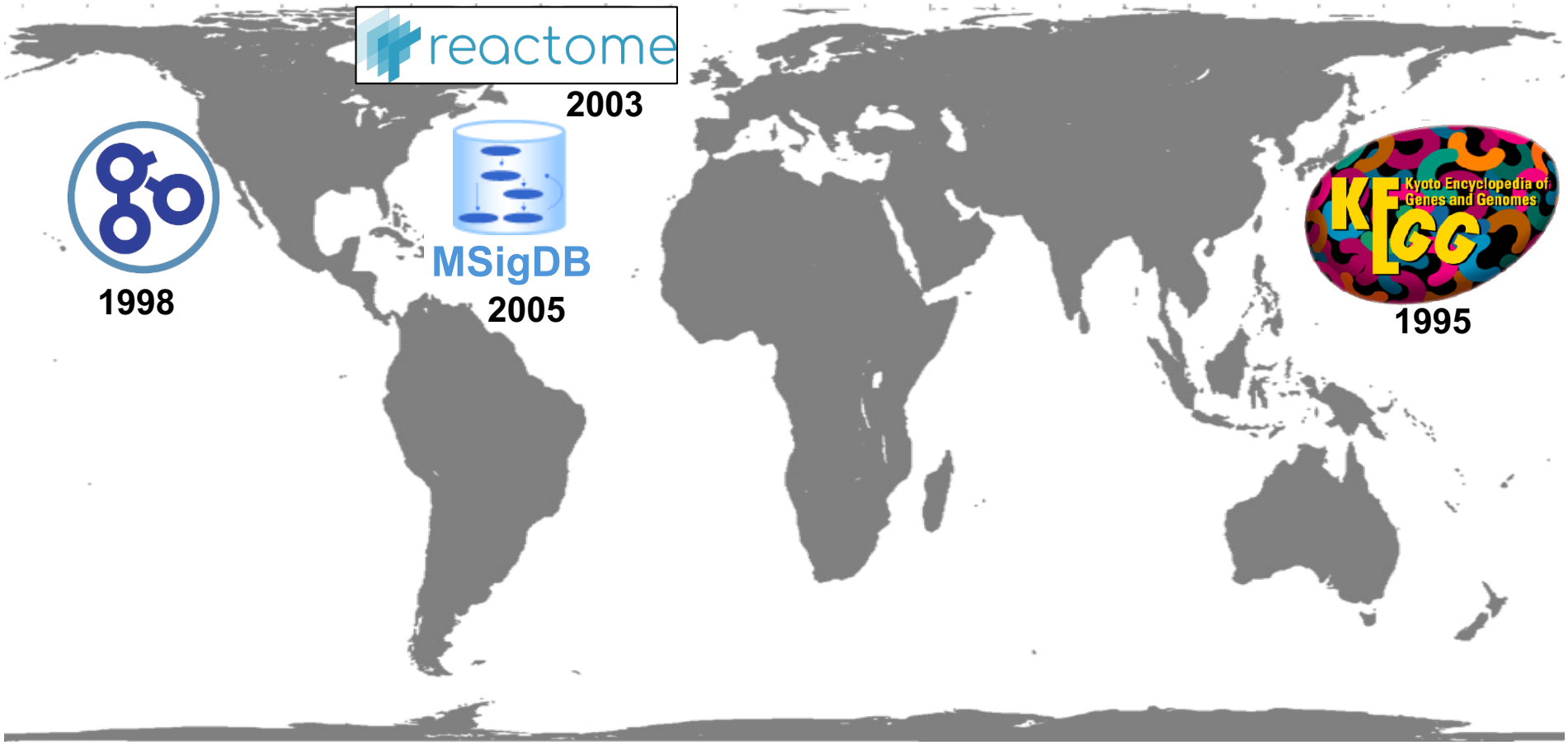
Set the wayback machine for 1995

Let's Share What We Know

World Wide Web

Microsoft Windows 95

“What if we had biological gene sets?”



Consider a GO term (gene set)

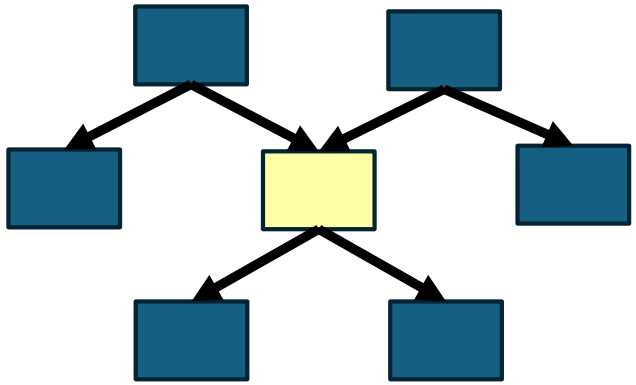
Accession	**GO:0008286**
Name	**insulin receptor signaling pathway**
Definition	**The series of molecular signals generated as a consequence of the insulin receptor binding to insulin. …**

Annotations (Genes)

IRS1
IRS2
IGF1R
INSRR
…

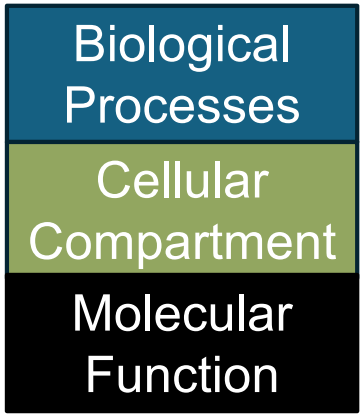
Evidence

Ontology



GO term ontology

3 “Aspects”



40k terms

174 “deep” species
5500 total

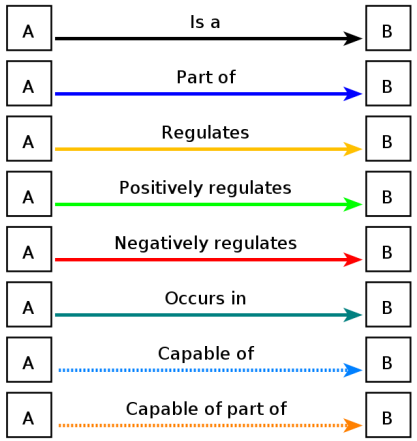
GO term ontology

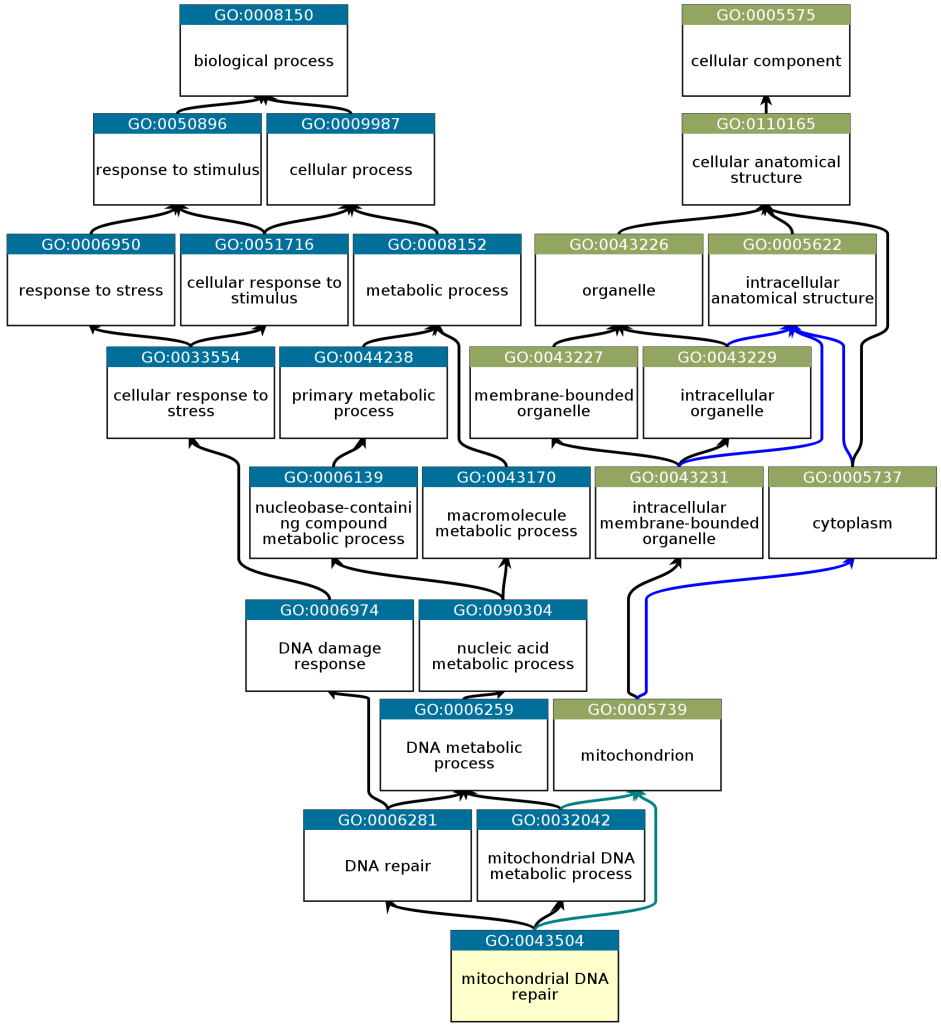
3 Aspects

40k terms

- Biological Processes
- Cellular Compartment
- Molecular Function

174 "deep" species
5500 total





GO term ontology

3 Aspects

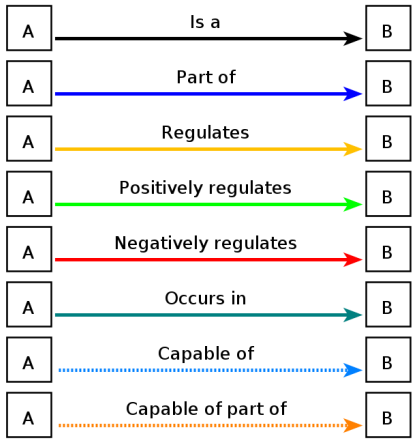
40k terms

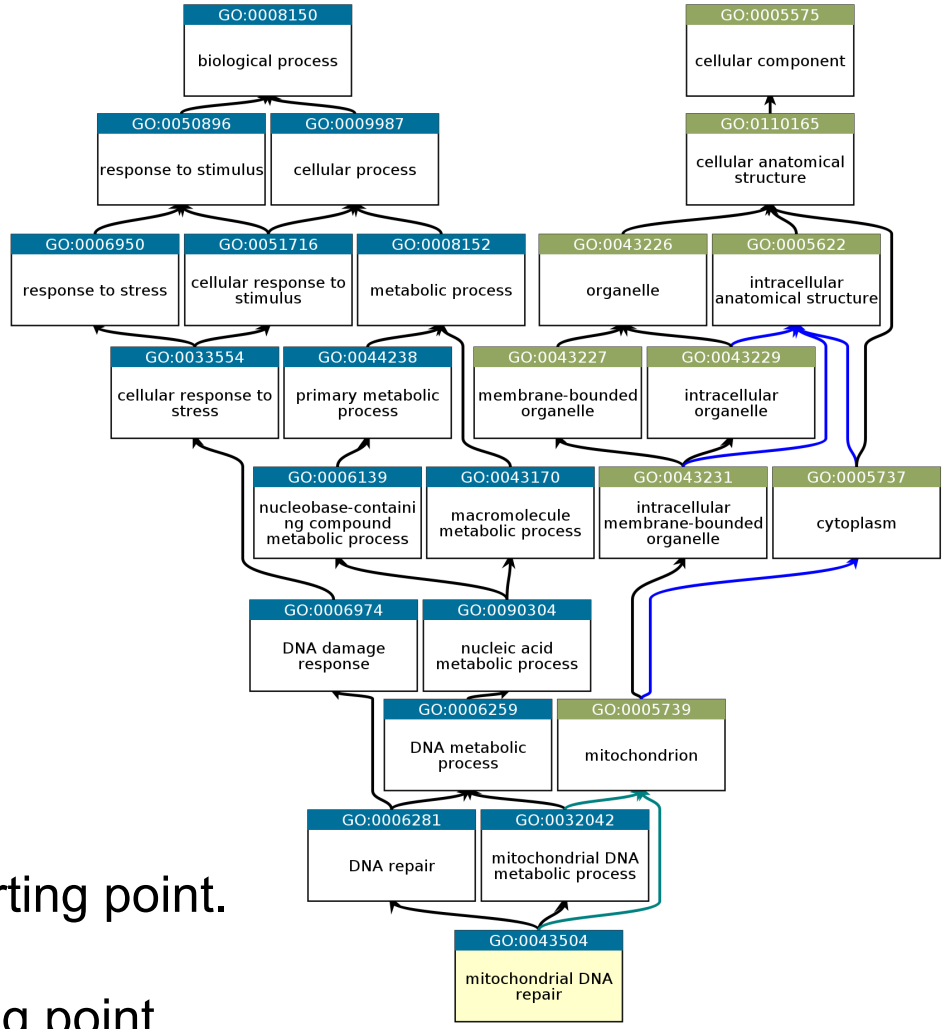
Biological Processes

Cellular Compartment

Molecular Function

174 "deep" species
5500 total





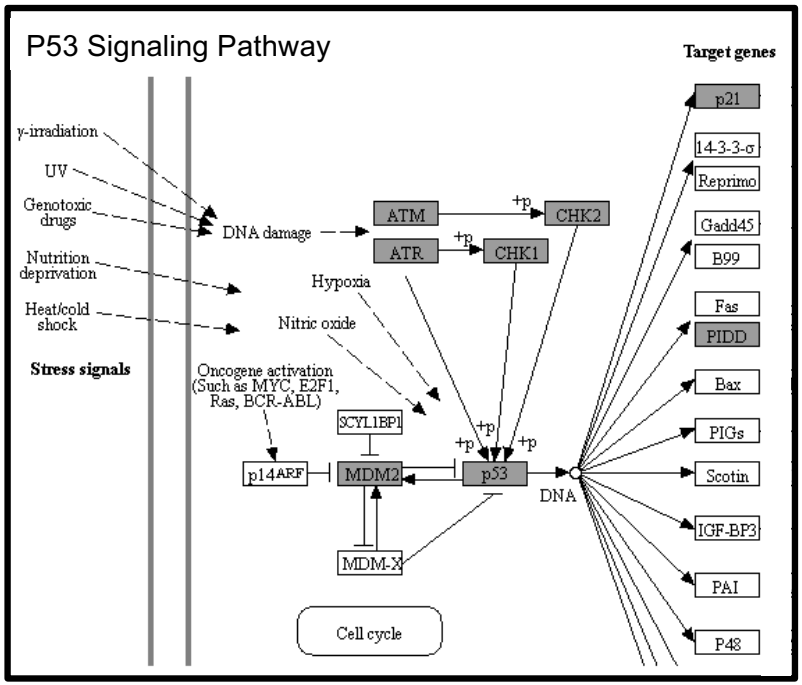
- Normal biological processes
- Rich hierarchy creates redundancy
- Reduced subset (GOslims) good starting point.
- Usually one "aspect" at a time.
  Biological Processes common starting point.

Pathway knowledge bases

1400 pathways

- Systems
- Genomic
- Chemical
- Health

(many species)

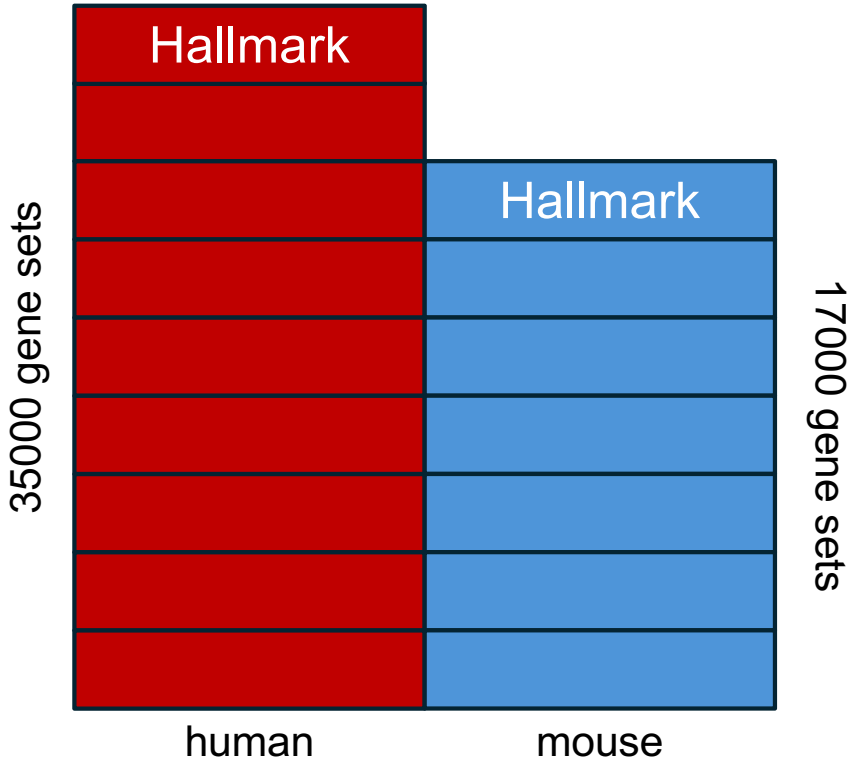


reactome

2800 pathways

(human)

Molecular Signatures Database (MSigDB)



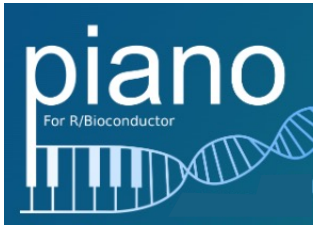
MSigDB

- Many curated from source
- *Also aggregated* from GO, KEGG, Reactome, etc.
- Built alongside GSEA
- **Hallmark** subsets are clean starting point

2025 Functional Analysis Ecosystem

A limited excerpt of tools

Ingenuity Pathway Analysis

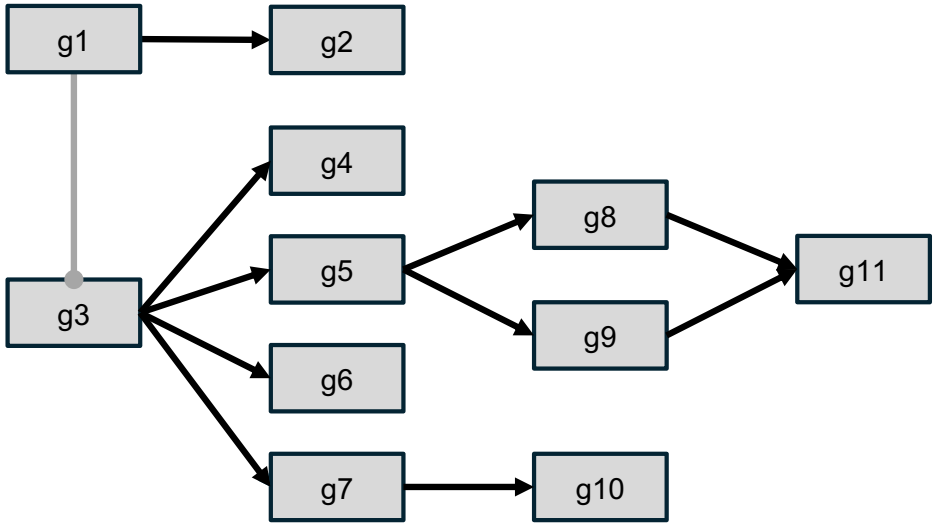


fGSEA

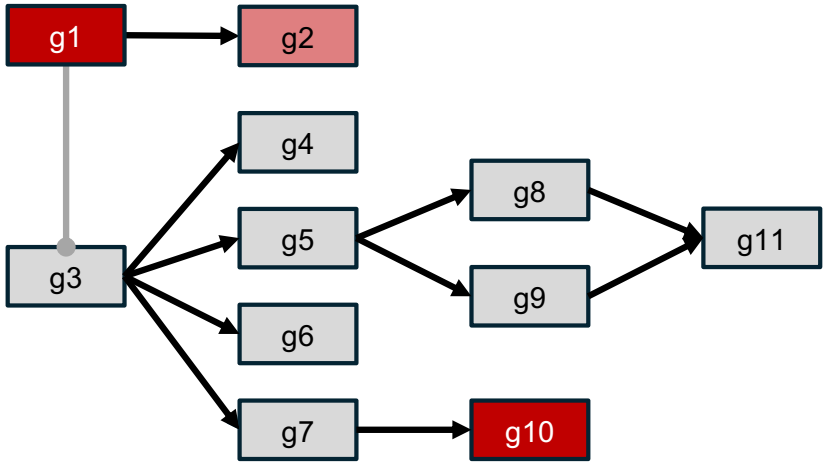
Knowledge bases

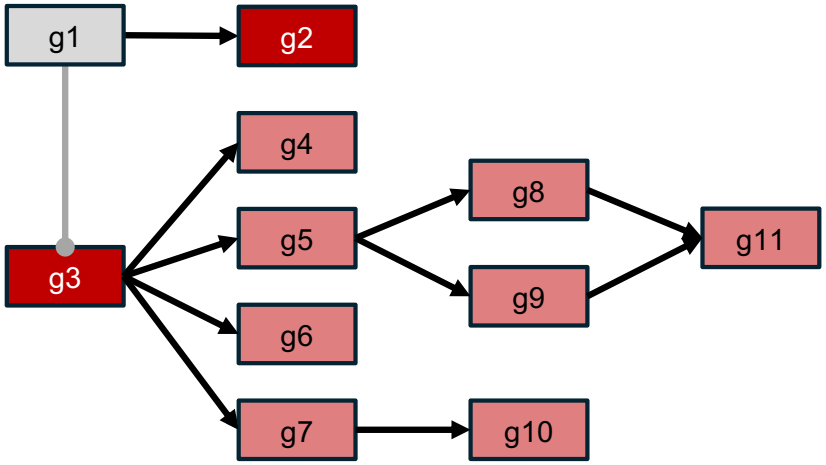
+ many more

A digression on Pathway Topology Analysis

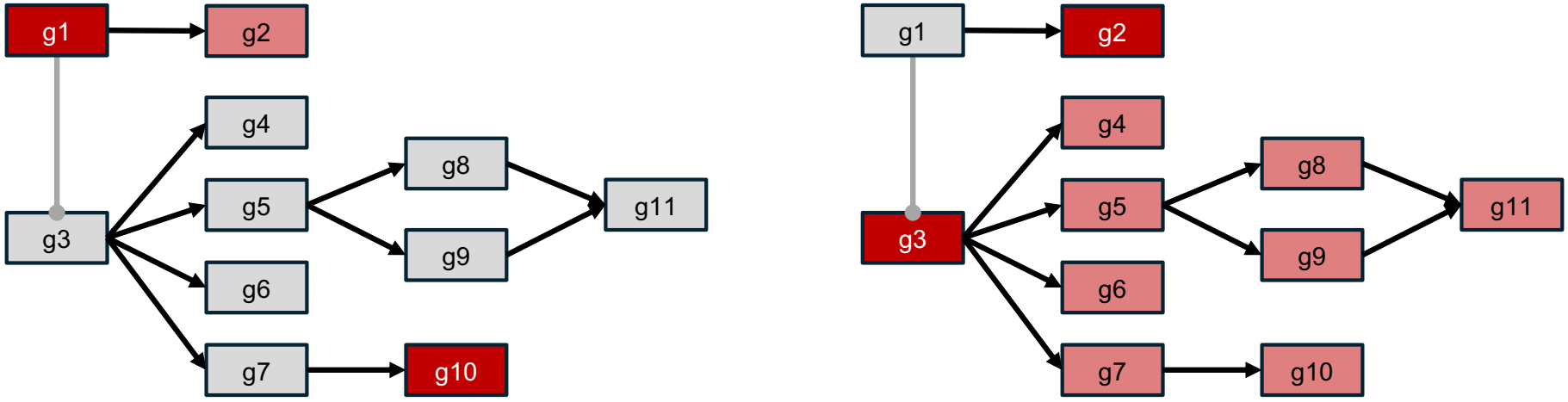


Genes impact pathways differently





ORA vs. GSEA vs. Pathway Topology



ORA

GSEA

Pathway Topology

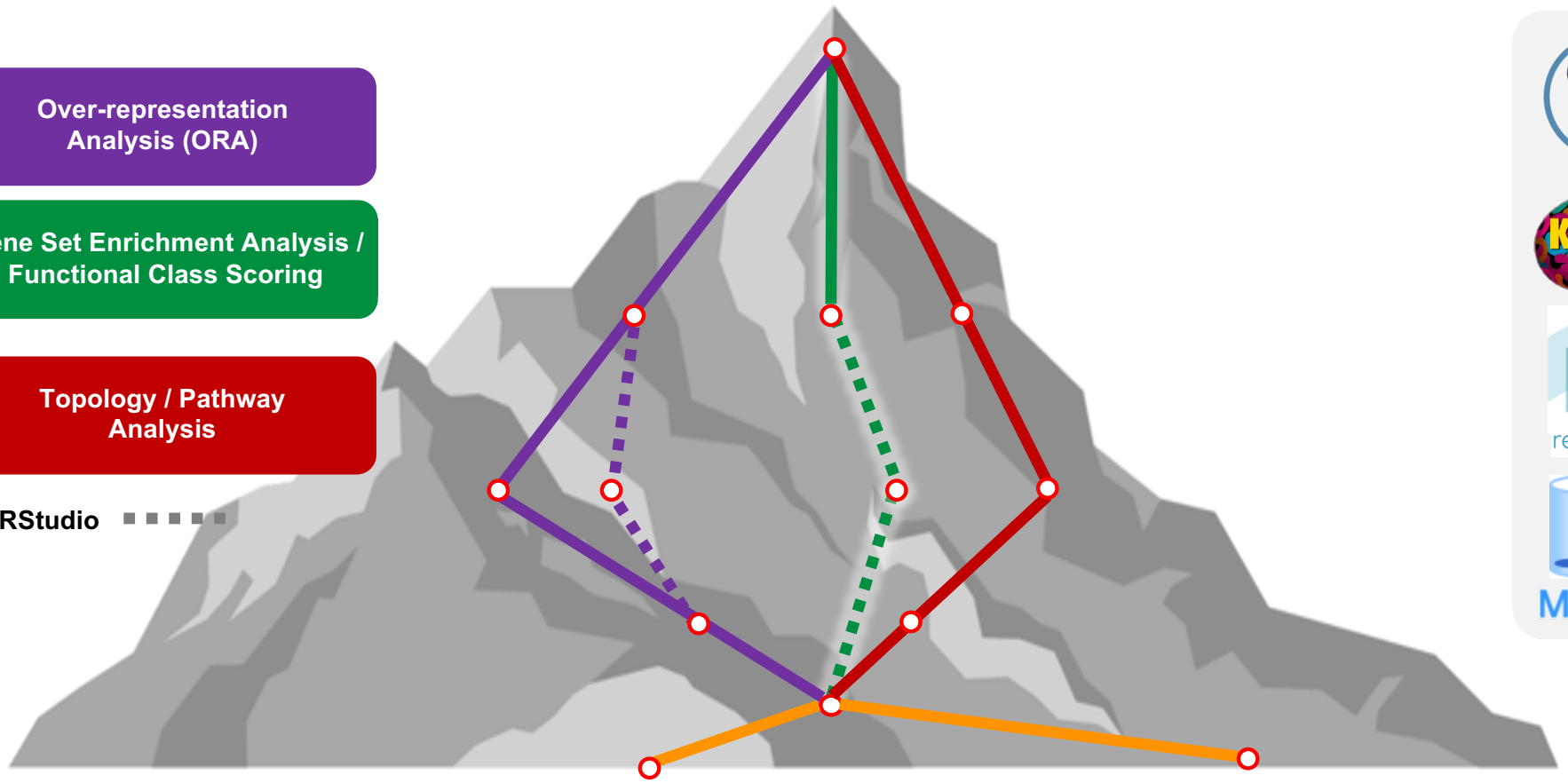
Algorithmic Complexity
Computational Burden/Time

Over-representation Analysis (ORA)

Gene Set Enrichment Analysis / Functional Class Scoring

Topology / Pathway Analysis

R/ RStudio



reactome

MSigDB