

SciLifeLab

Pathway Analysis

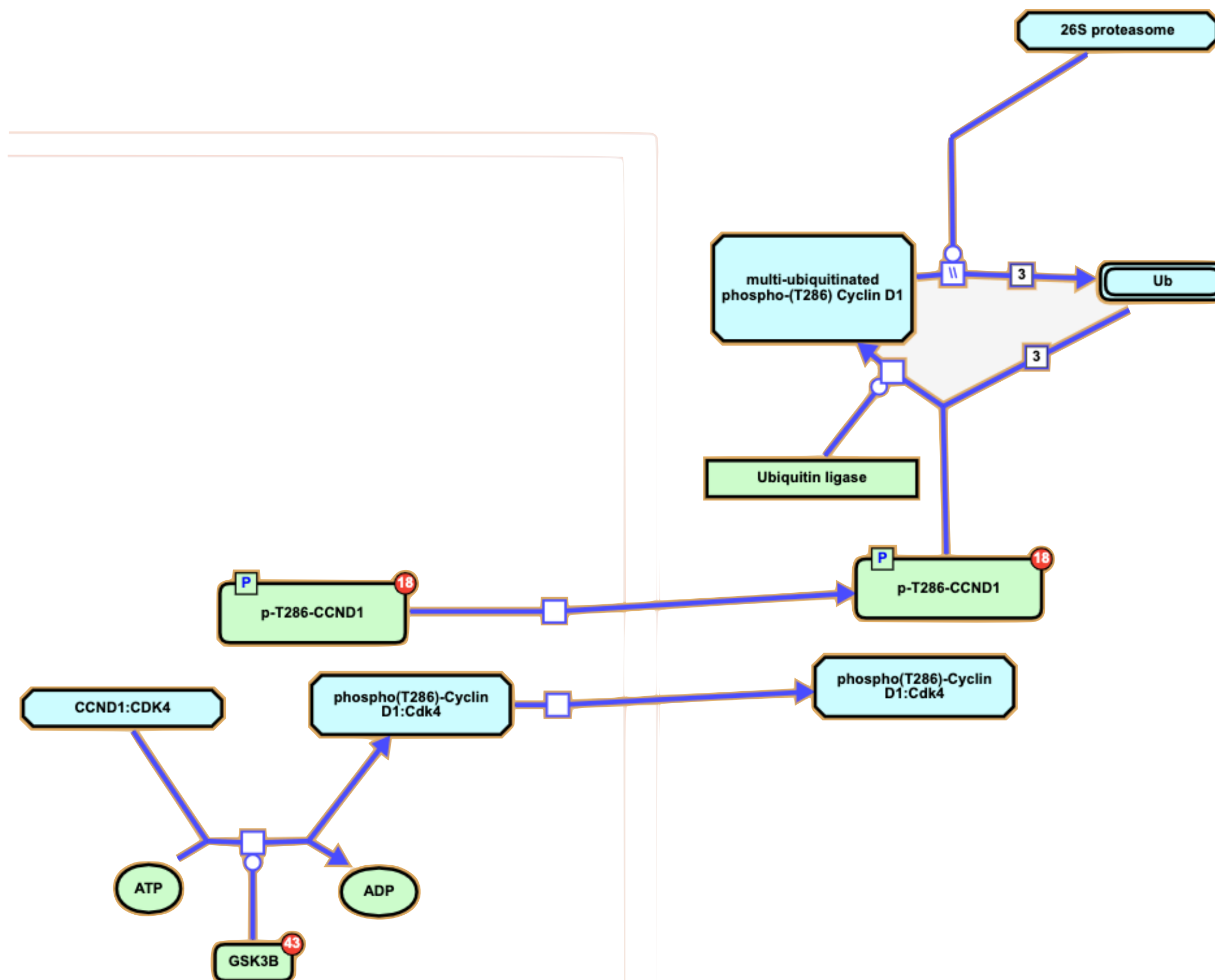
CB2030

Lukas Käll, KTH



Pathway Analysis

Pathway analysis is the summarisation of the measurements of related analyses into one integrated metric of pathway activity



Pathway Analysis

- For the purpose of the lecture, a pathway are any set of analyses that are thought to be co-regulated, i.e. they belong to the same metabolic pathway, they interact, they belong to the same protein complex, or sit in the same cellular compartment.

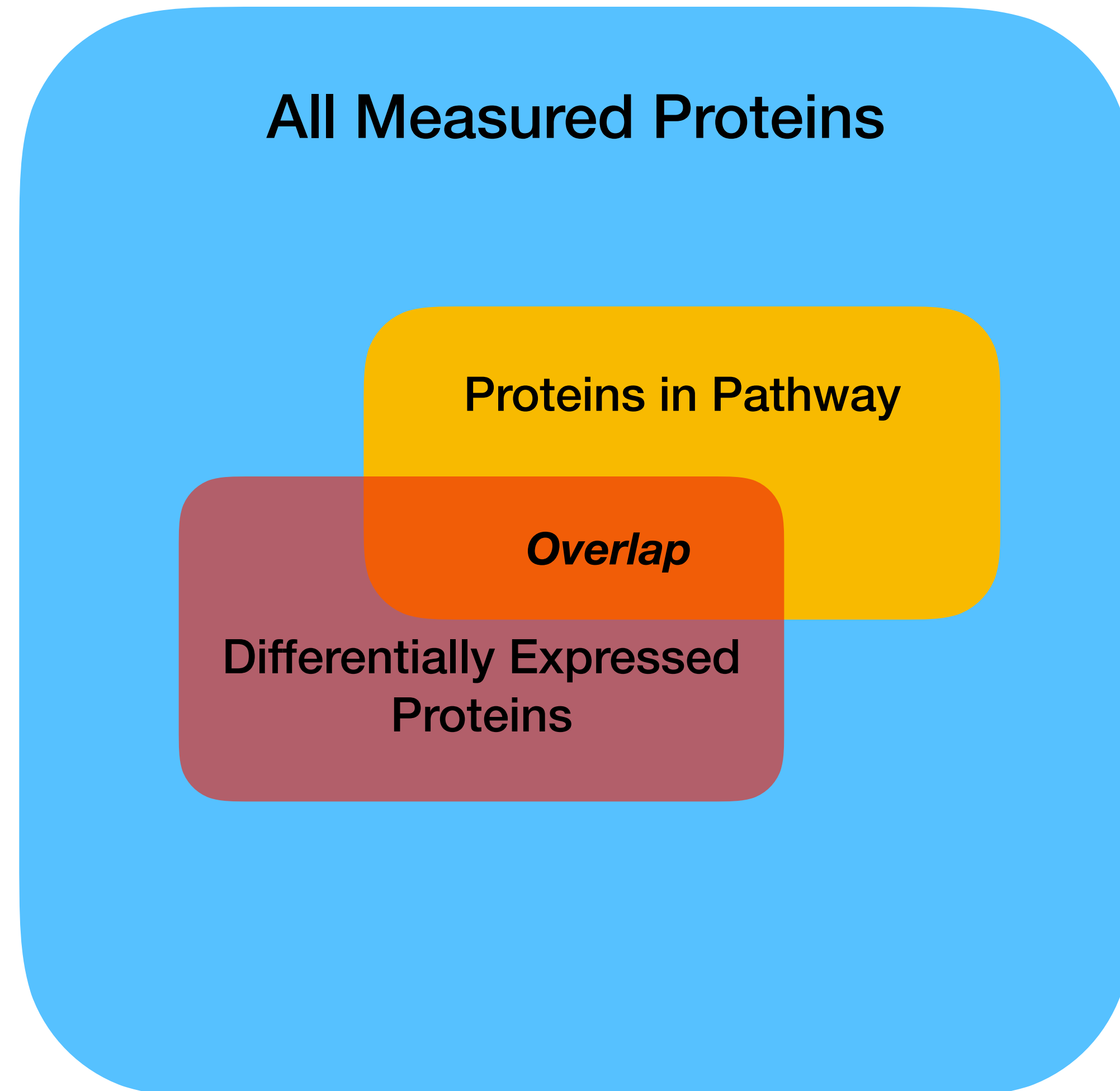
Rational behind Pathway Analysis

- Most genes and proteins are acting in concert, hence it can make sense to measure their activity on group level instead of the individual protein
- Co-variation is a better signal than just individual measurements of analyses
- Fewer hypothesis => Better signal-to-noise ratio

There are a many types of “pathway” databases

- Some common pathway databases:
 - Metabolic Pathways (Proteins + Metabolites): KEGG, Reactome, Pathway Commons
 - Interactions (Proteins): CORUM, IntAct, STRING
 - Gene Ontology (Gene Products): Molecular Function, Cellular Component, Biological Process

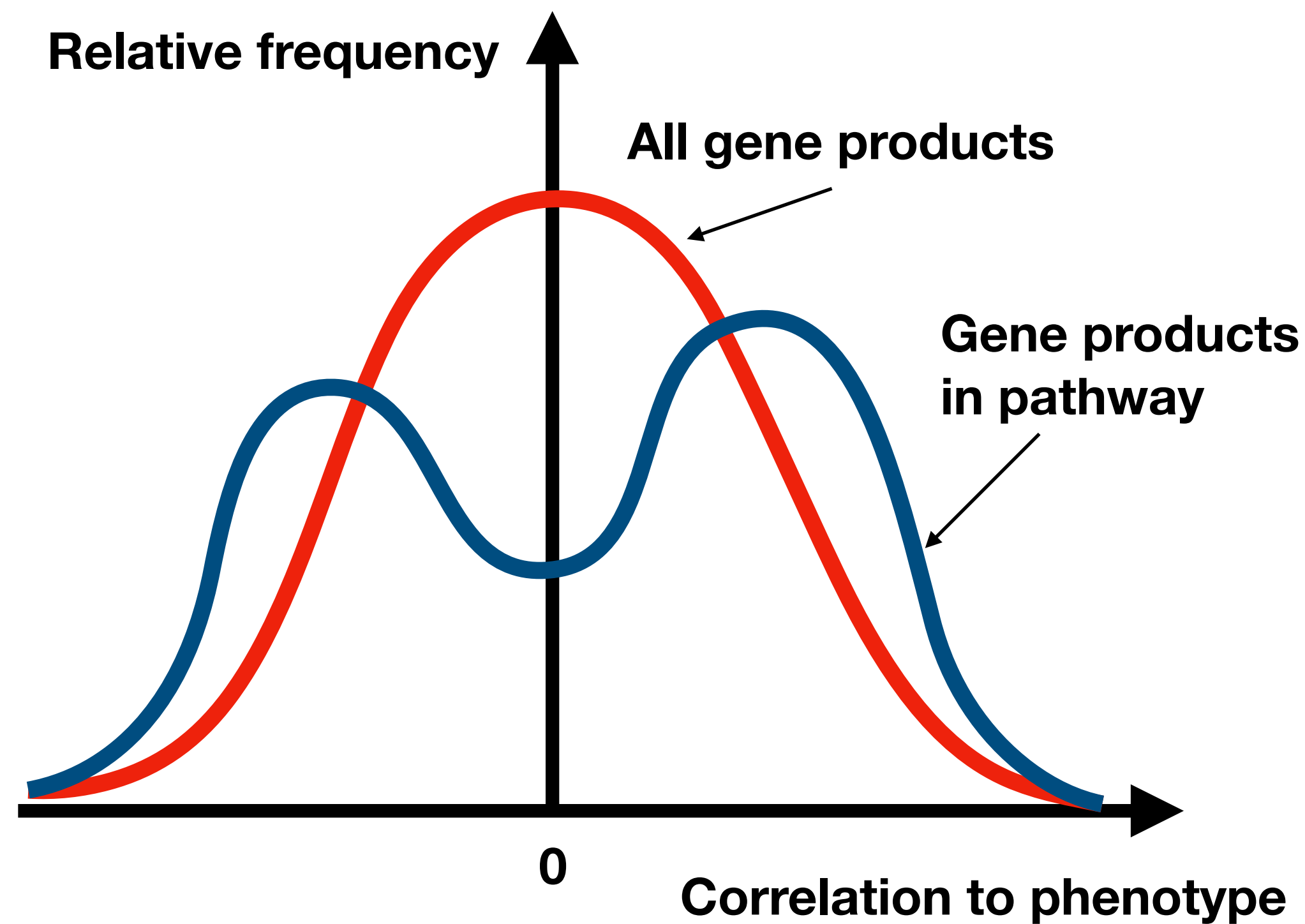
Over-representation Analysis



- Compare the lists of “genes” in the pathway to the list of differentially expressed gene products
- Check if the overlap is significantly larger than expected by chance
- Significance through hypergeometric tests or Fisher’s exact test

$$H_0: \frac{\text{yellow}}{\text{blue}} = \frac{\text{orange}}{\text{maroon}}$$

Gene-set enrichment analysis



- Compare the distribution of all measured to the pathway's gene product's correlation to the phenotype (e.g. Healthy=1; Disease=-1)
- If distributions are significantly different there is differential pathway activity.
- Current implementations of GSEA only measures up or down regulation, not both on the same time.