RNA-Seq
Now What?

BIS180L
Professor Maloof

We have differentially expressed genes, what do we want to know about them?

We have differentially expressed genes, what do we want to know about them?

- What biological processes are affected by:
    - treatment
    - genotype
    - genotype x treatment interaction

We have differentially expressed genes, what do we want to know about them?

- What biological processes are affected by:
    - treatment
    - genotype
    - genotype x treatment interaction
- What transcription factors might control expression of these genes?

We have differentially expressed genes, what do we want to know about them?

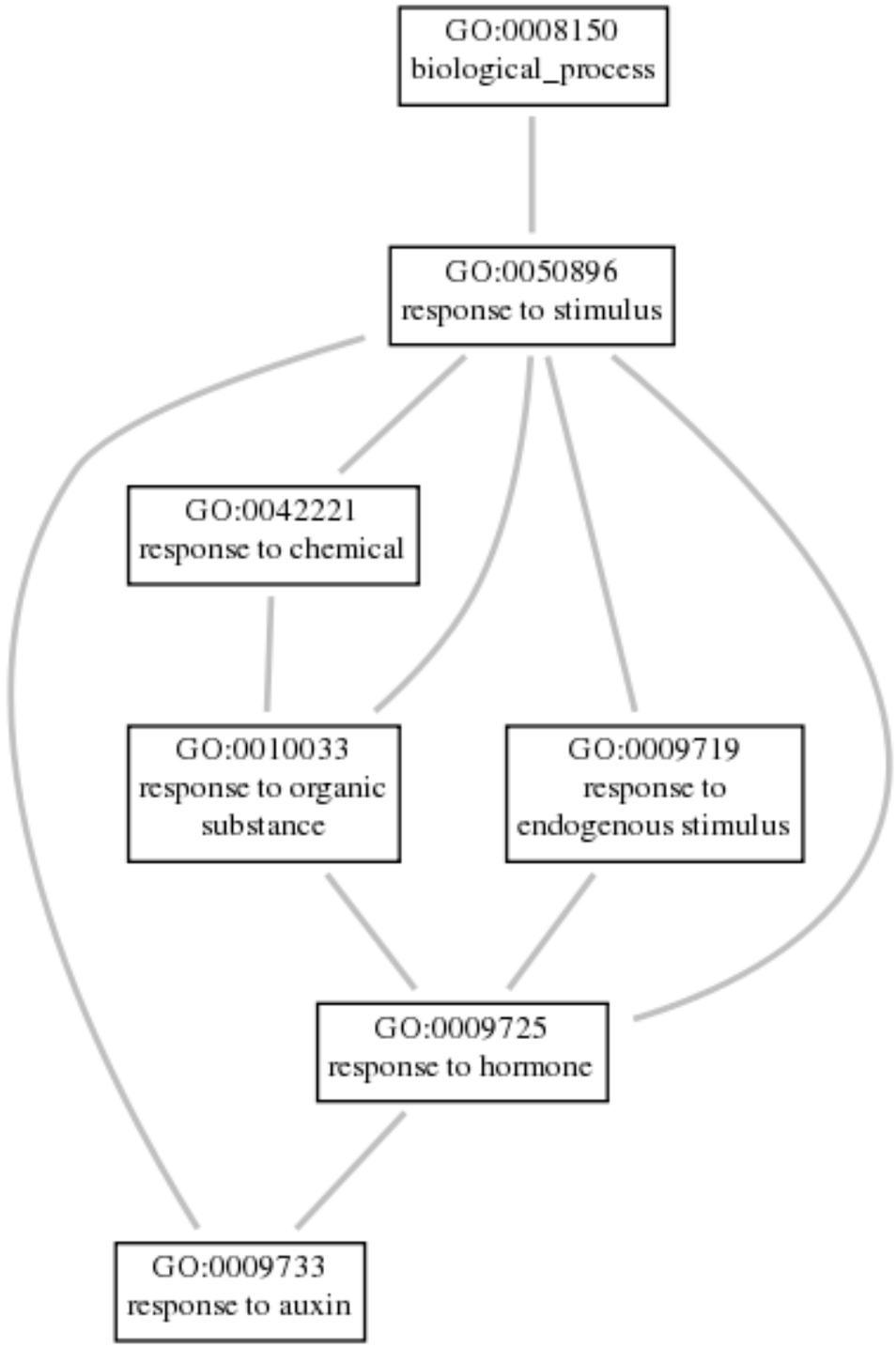
- What biological processes are affected by:
    - treatment
    - genotype
    - genotype x treatment interaction
- What transcription factors might control expression of these genes?
- We have hundreds of differentially expressed genes, how do we figure this out?

Gene Ontology (GO) terms

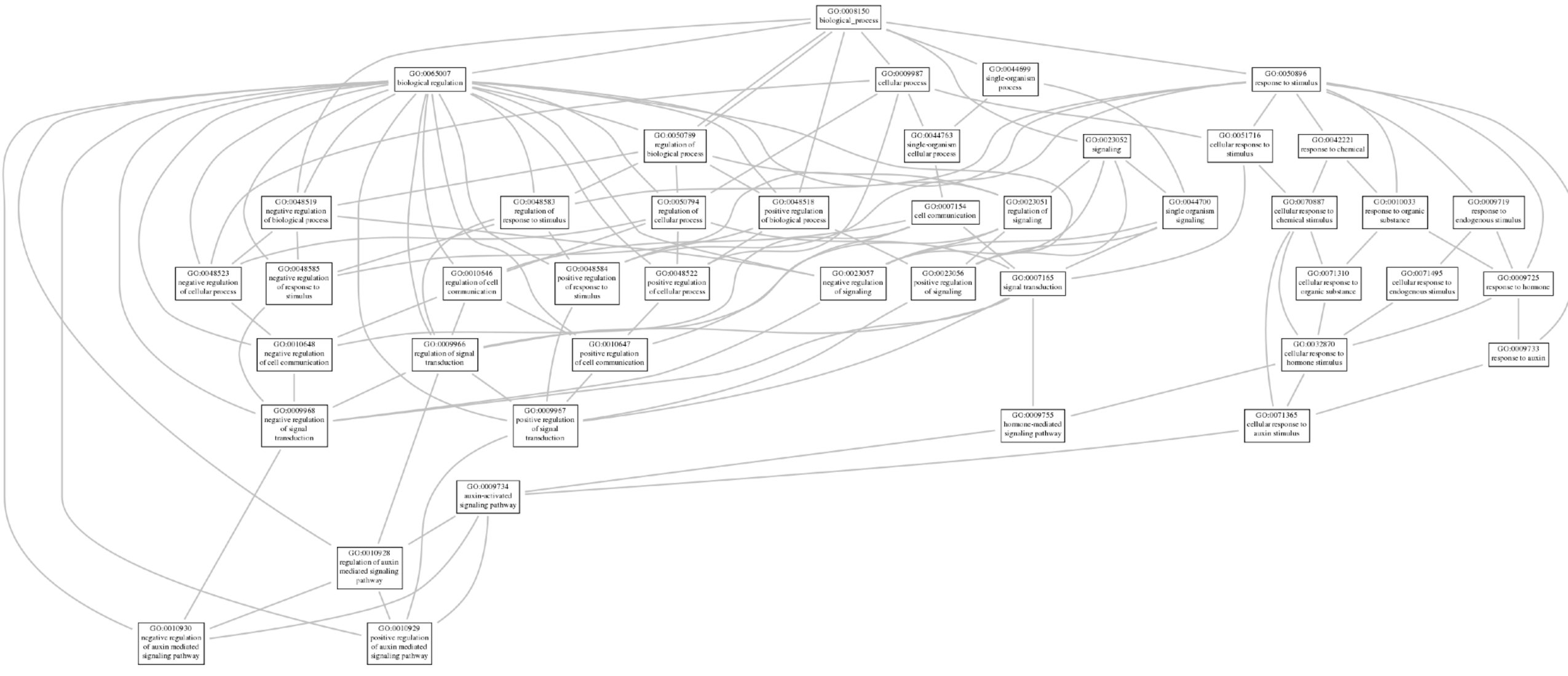
- GO terms are a defined hierarchical vocabulary for describing gene functions

-

GO term example: response to auxin



More of the auxin GO hierarchy

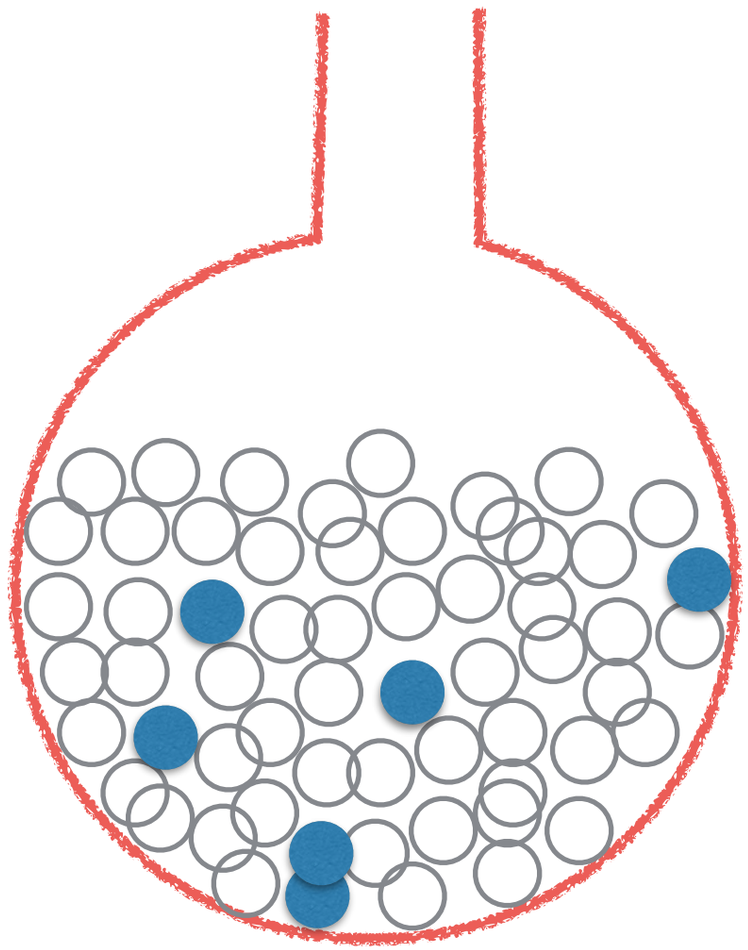


GO and differential expression analysis

- Alternative to analyzing individual genes

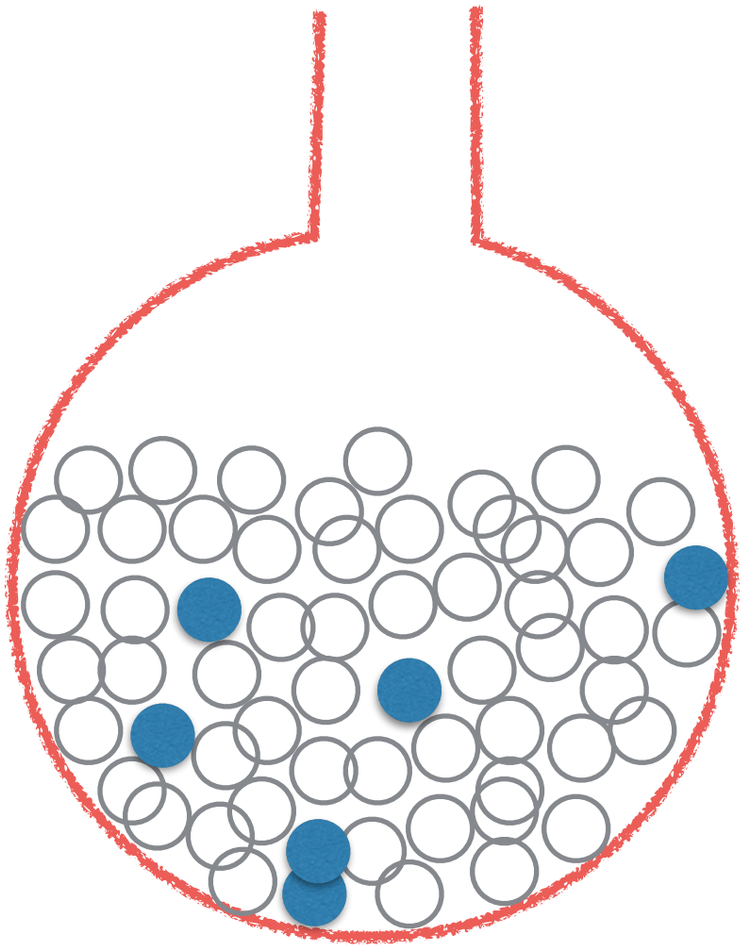
- Determine which GO terms are over-represented among the differentially expressed genes

Over-representation analysis



Jar with 1000 marbles
800 white, 200 blue

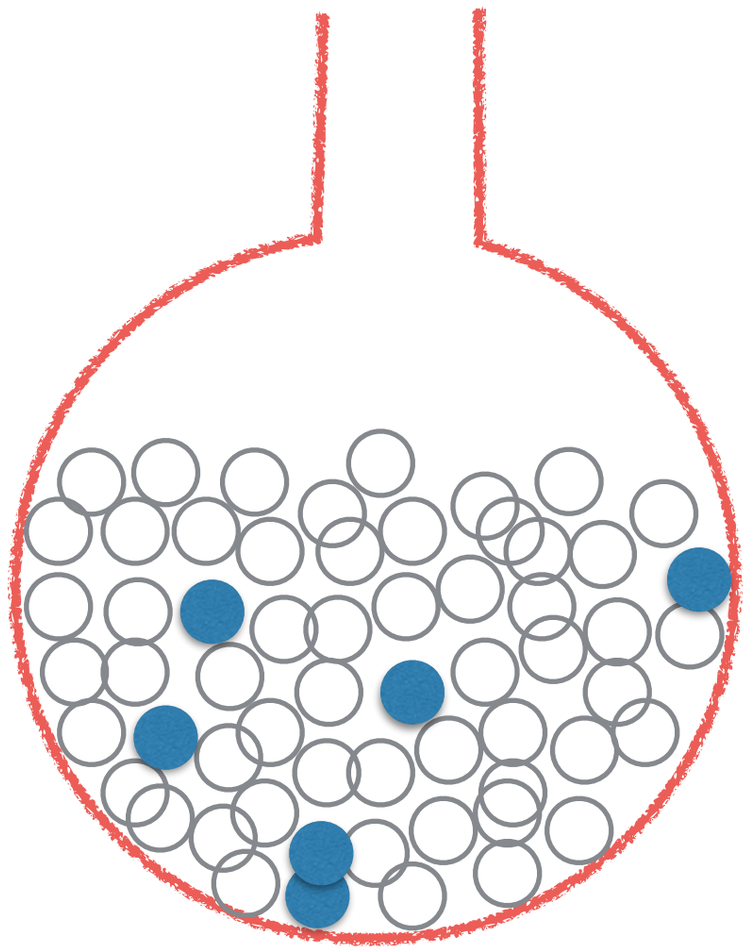
Over-representation analysis



Jar with 1000 marbles
800 white, 200 blue

Remove 20
How many white and blue
do you expect?

Over-representation analysis

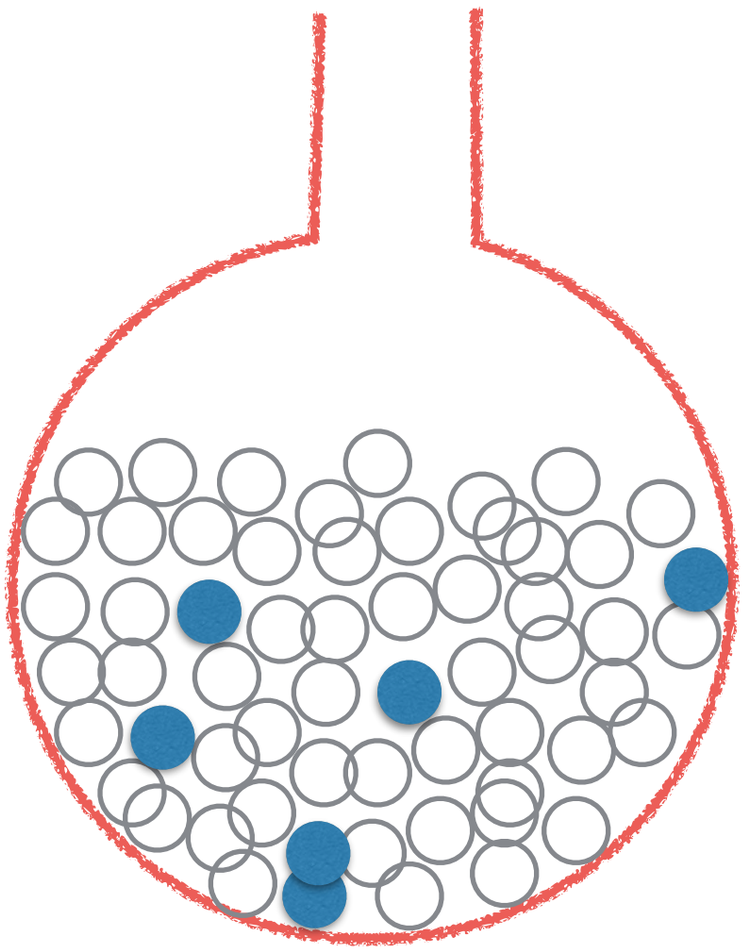


Jar with 1000 marbles
800 white, 200 blue

What if you find
10 white and 10 blue?

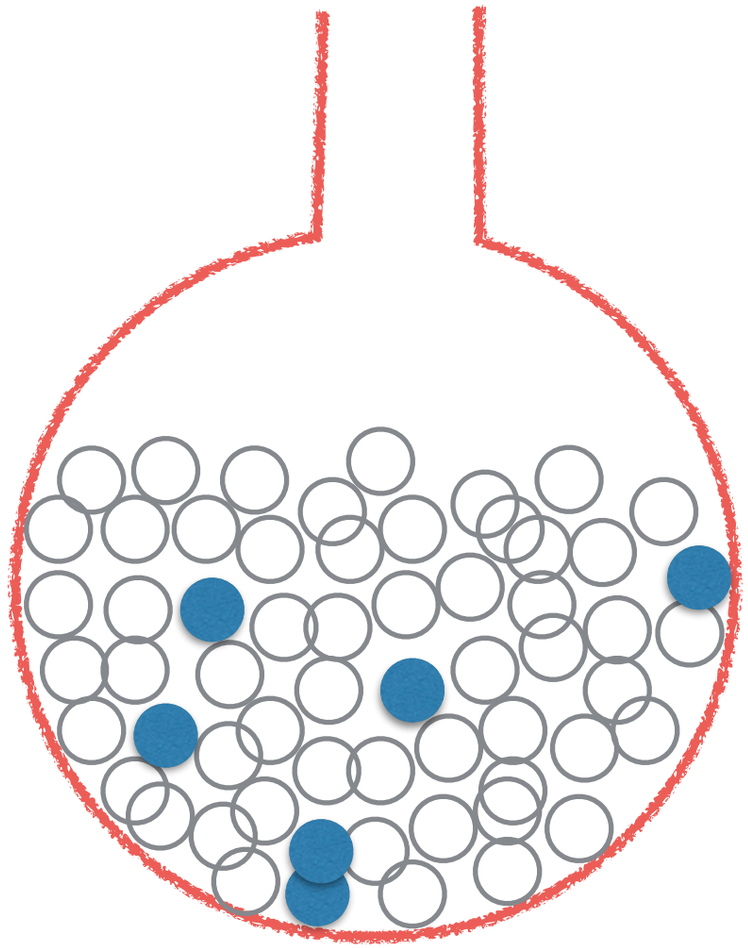
Blues are **over-represented**
in your sample

Over-representation analysis



Genome with 34,000 genes
340 have GO term:
“Cell wall modification”

Over-representation analysis

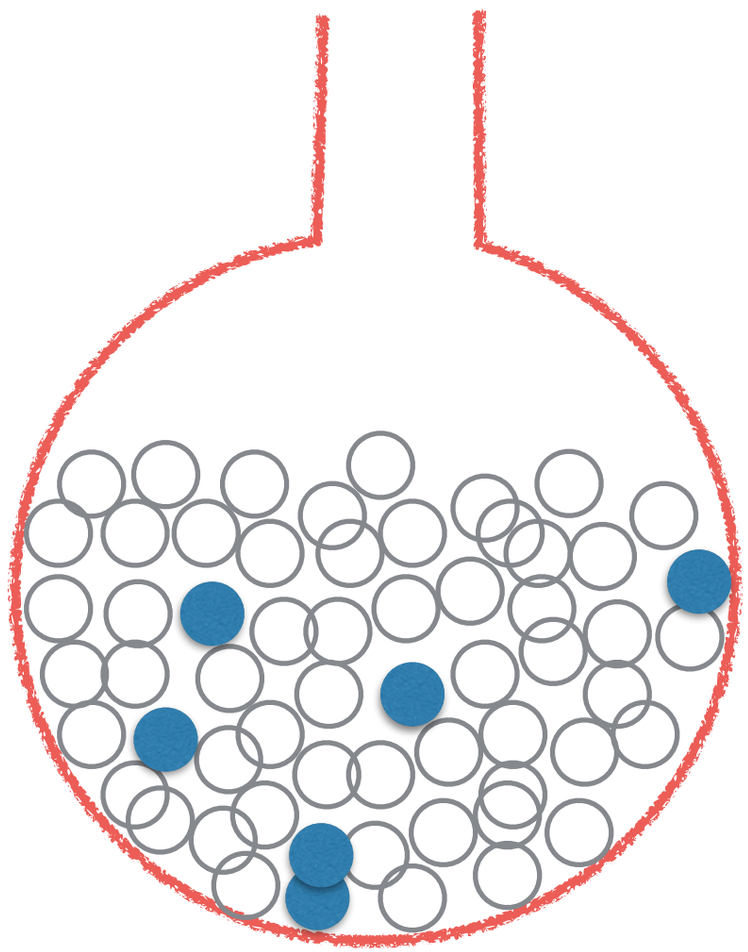


Genome with 34,000 genes
340 have GO term:
“Cell wall modification”

Find 1,000 genes differentially expressed in response to shade.

How many do you expect to have the GO term “cell wall modification”?

Over-representation analysis



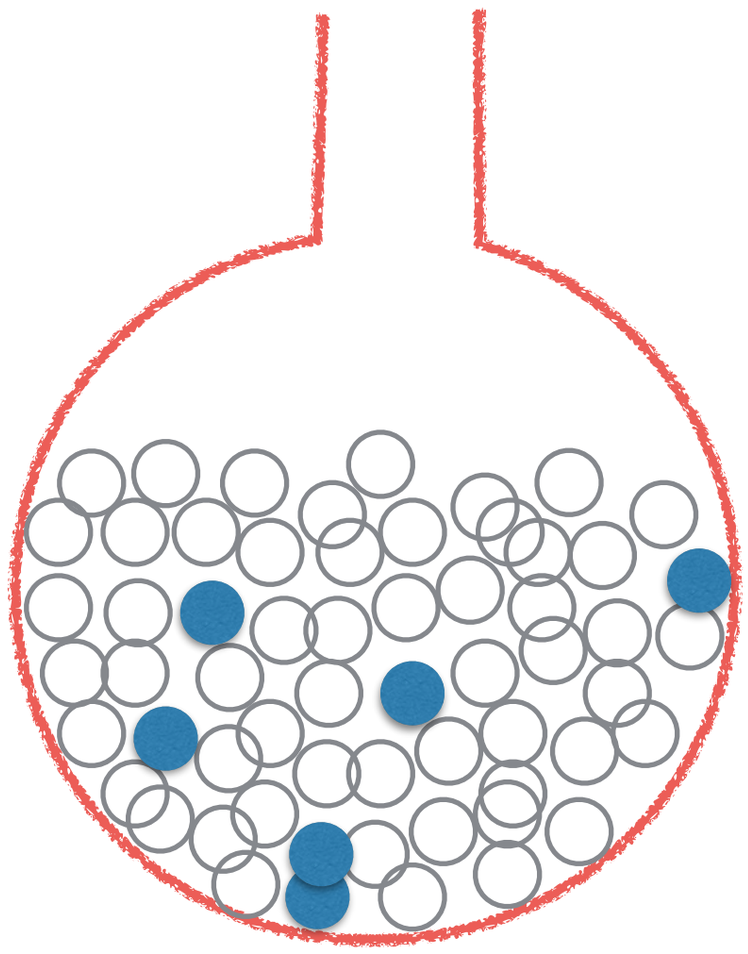
Genome with 34,000 genes
340 have GO term:
“Cell wall modification”

Find 1,000 genes differentially expressed in response to shade.

How many do you expect to have the GO term “cell wall modification”?

We expect **10** genes to have the GO term “cell wall modification”

Over-representation analysis

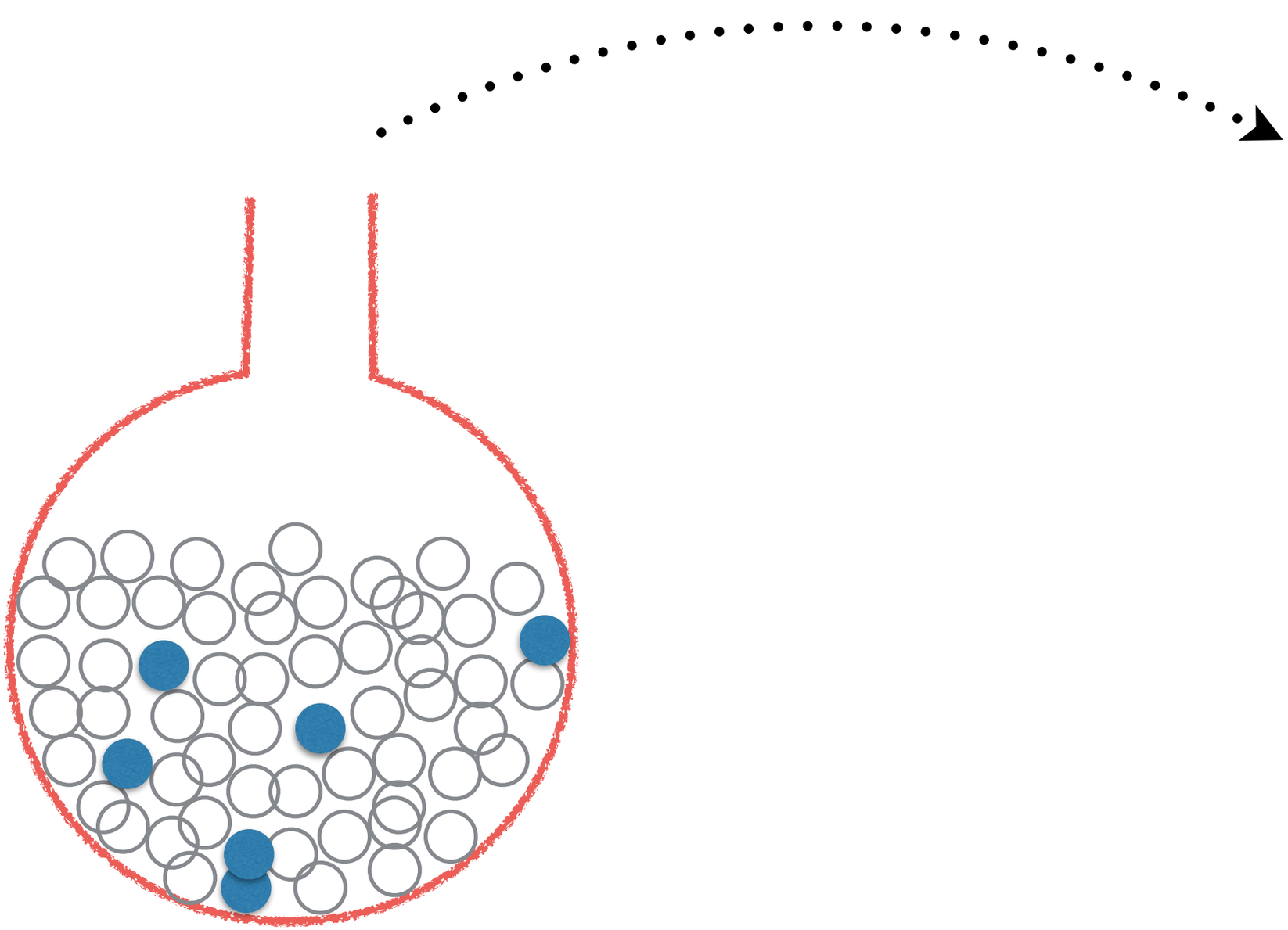


Find 1,000 genes differentially expressed in response to shade.

If instead of expectation, you find that 50 of these 1,000 have the GO term “Cell wall modification”, then what do you conclude?

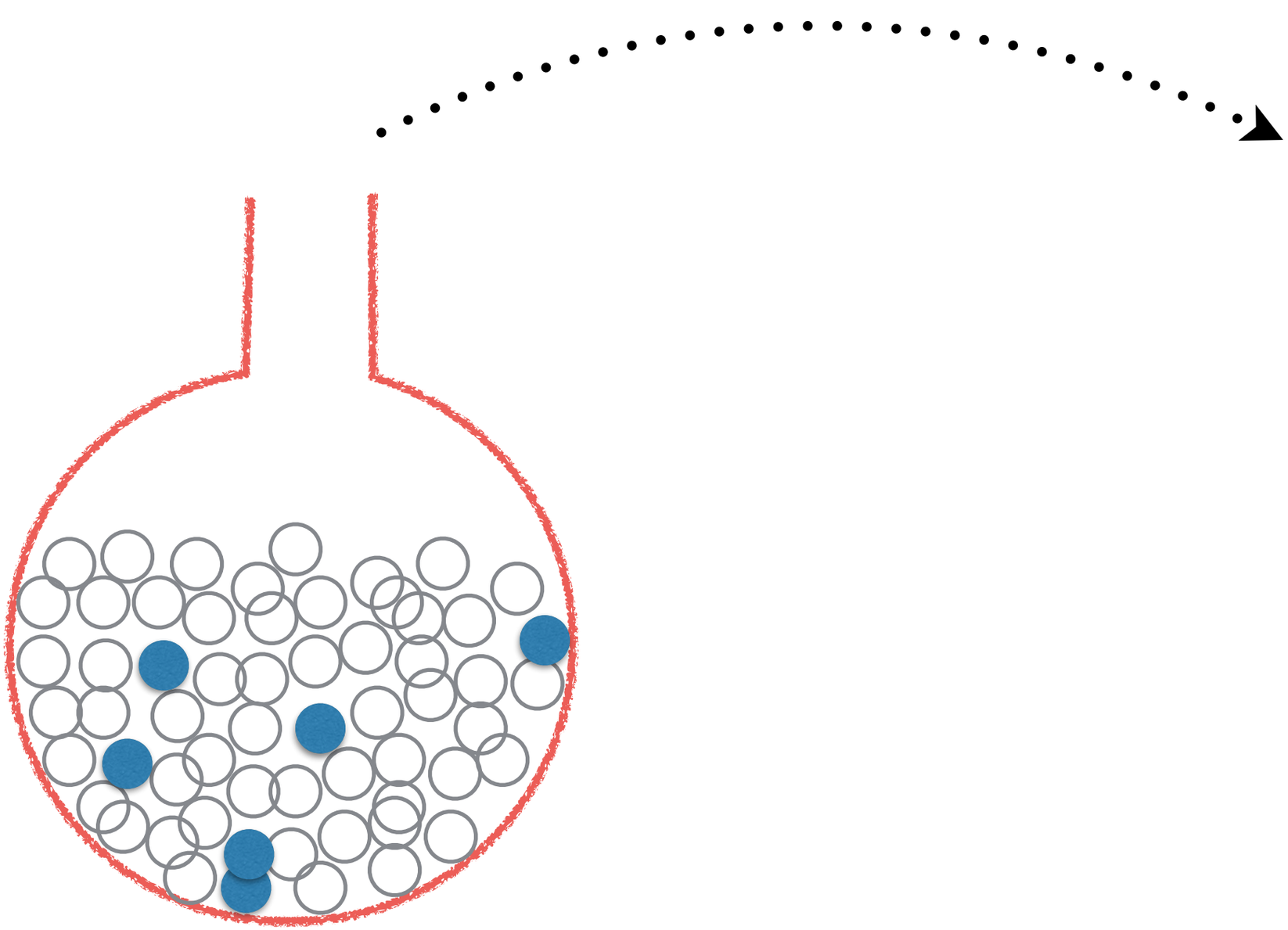
Genome with 34,000 genes
340 have GO term:
“Cell wall modification”

Over-representation analysis: additional considerations

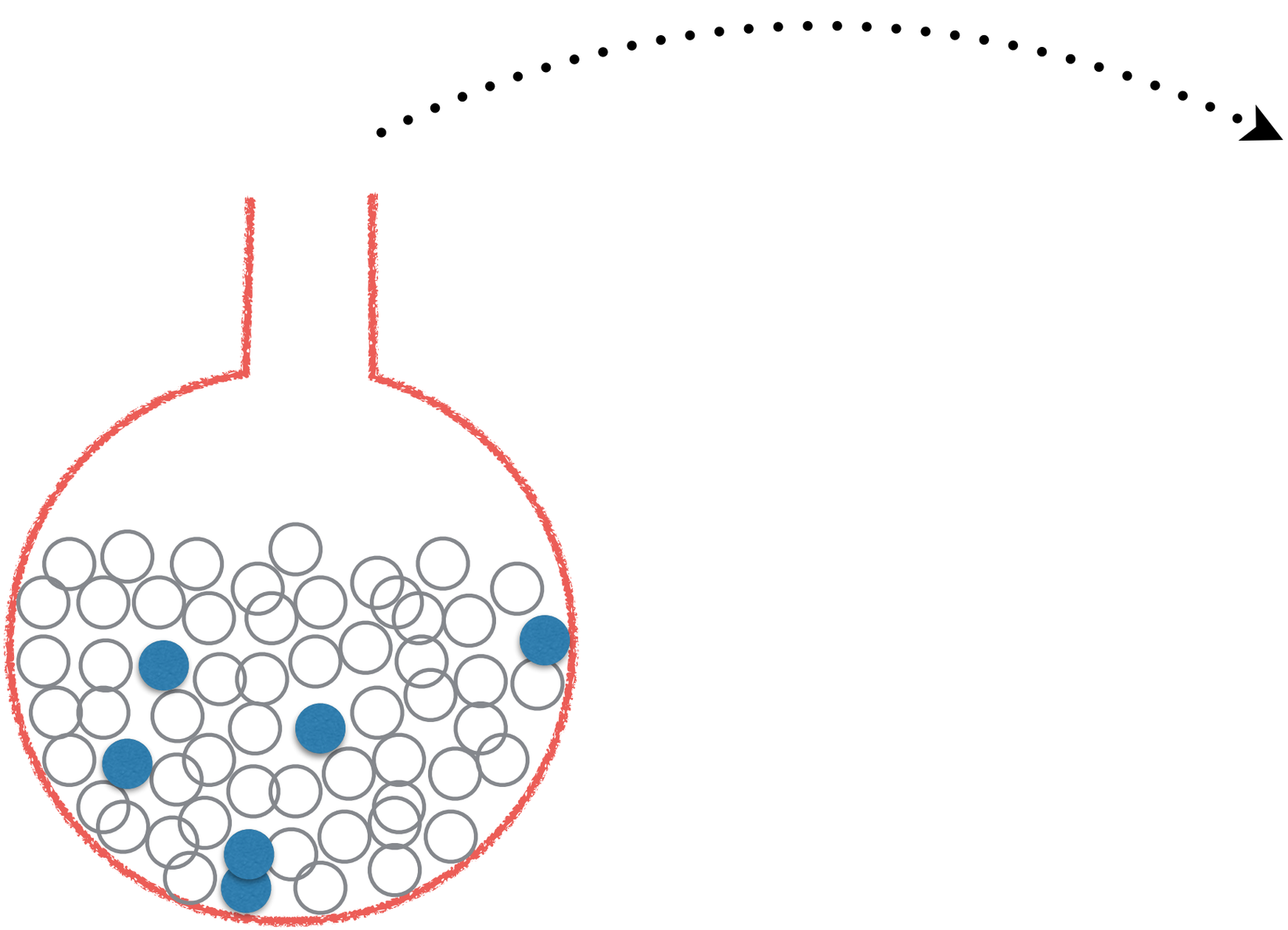


1) The Jar (aka Universe) should only contain **expressed genes.** (Why?)

Over-representation analysis: additional considerations

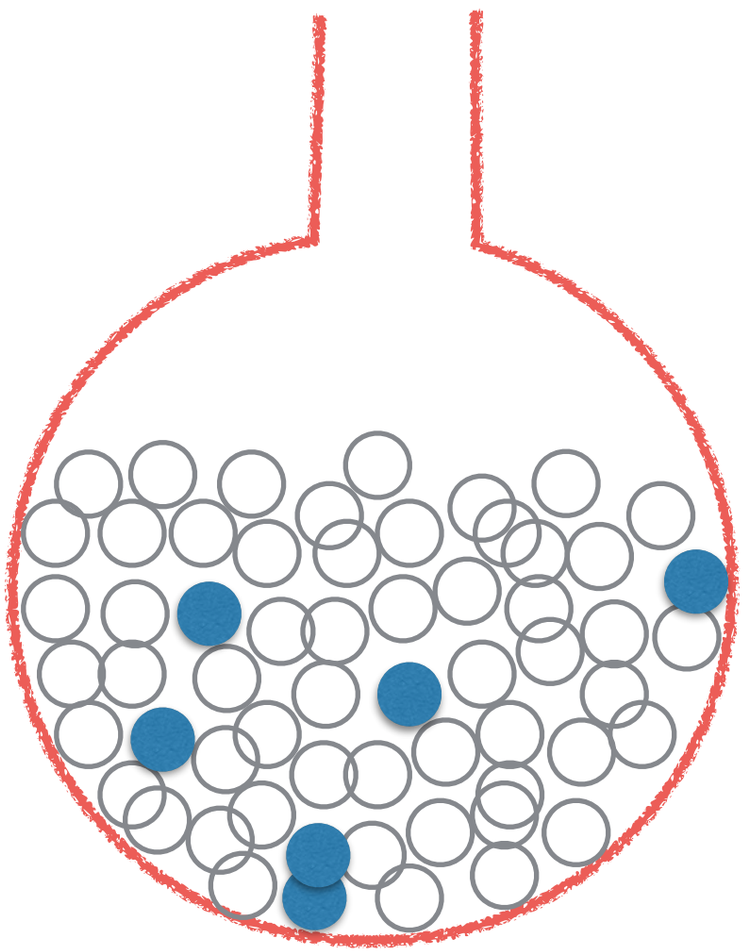


Over-representation analysis: additional considerations



1) The Jar (aka Universe) should only contain **expressed genes.** (Why?)

Over-representation analysis: additional considerations



2) Sequence length biases in detecting differentially expressed genes should be considered

1) The Jar (aka Universe) should only contain **expressed genes.** (Why?)

Statistical tests for over (or under) representation:

- Build a contingency table categorizing genes by D.E. and GO

GO term "cell wall modification" present?	Differentially Expressed: NO	Differentially Expressed: YES
NO	32,710	950
YES	290	50

- Use a (modified) Fisher's Exact Test to look for unequal ratios (is 32710:290 different from 950:50)?

Promoter Motif Analysis

Promoter Motif Analysis

- As you know transcription factors bind to specific DNA sequences or “motifs”.

Promoter Motif Analysis

- As you know transcription factors bind to specific DNA sequences or “motifs”.
- Similar to looking for enrichment of GO terms among our DE genes we can also look for enrichment of transcription factor binding motifs in the promoters of DE genes.

Promoter Motif Analysis

- As you know transcription factors bind to specific DNA sequences or “motifs”.

- Similar to looking for enrichment of GO terms among our DE genes we can also look for enrichment of transcription factor binding motifs in the promoters of DE genes.

- Why would we want to do this? (What hypotheses might be generated by such an analysis?)