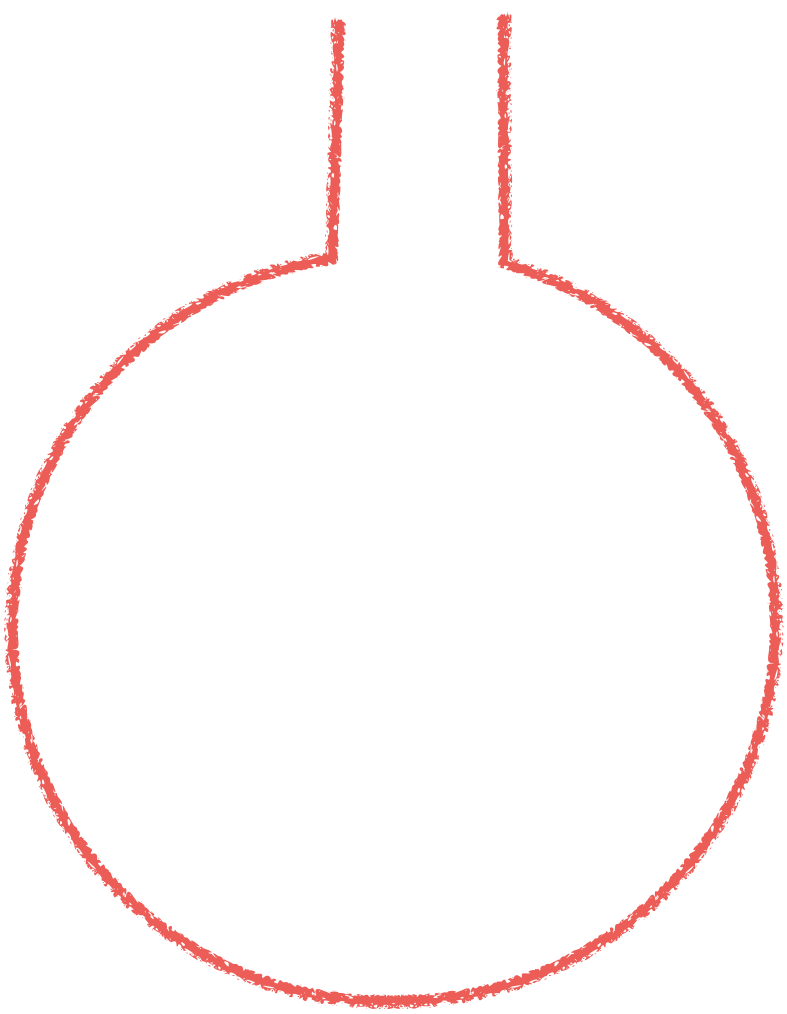


Over-representation analysis: additional considerations













































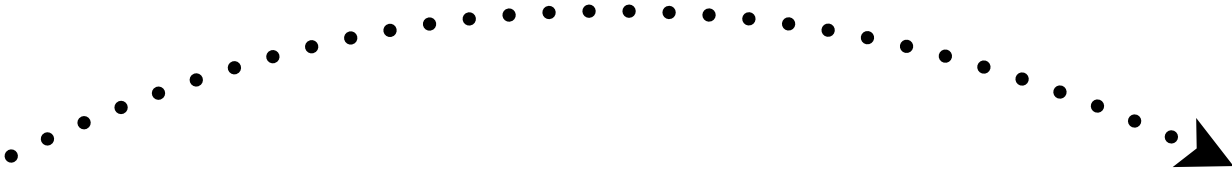






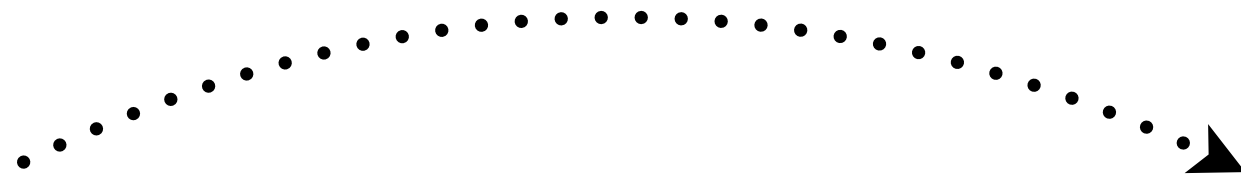


1) The Jar (aka Universe) should only contain **expressed genes**. (Why?)

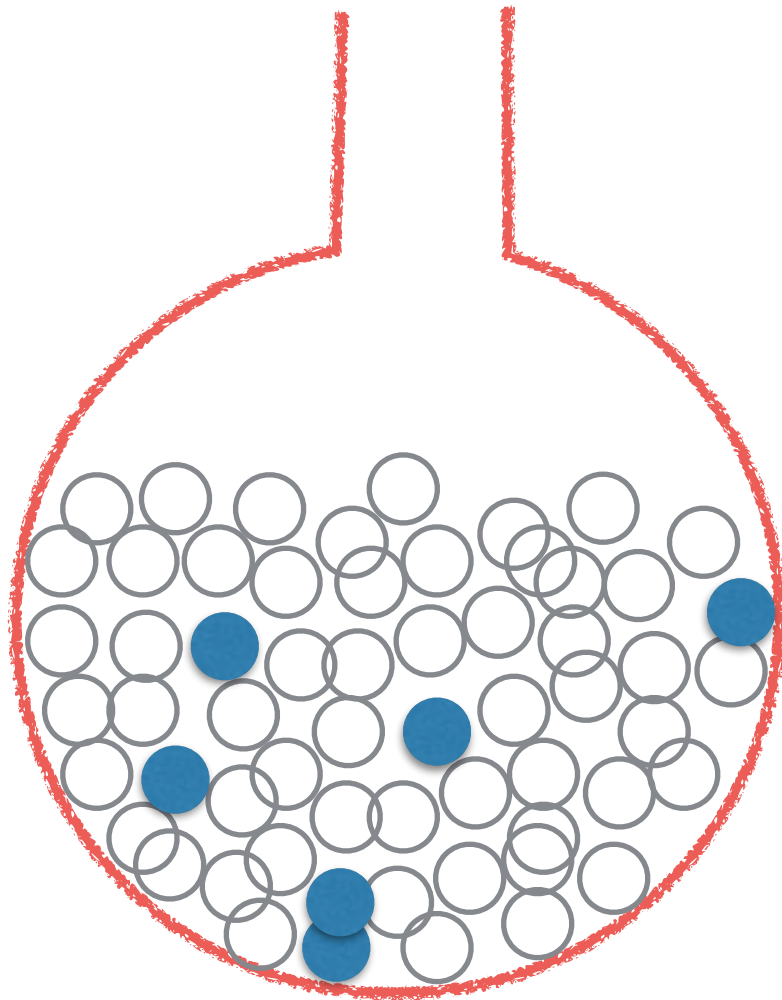


2) Sequence length bases in detecting differentially expressed genes should be considered

Over-representation analysis: additional considerations



2) Sequence length bases 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

		Differentially Expressed	
		NO	YES
GO term “cell wall modification” present?	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)?