



Lunch & Learn: Science, Statistics, and Getting it Right
ASIP 2018 Annual Meeting at Experimental Biology

Vignette 5

In order to understand the toxicity of rattlesnake venom, you are exposing human endothelial cells to vehicle or rattlesnake venom, and performing gene expression profiling using microarrays that assess the expression of 20,000 genes to identify those that are differentially expressed. Across triplicate experiments, you identify a set of 200 genes that are overexpressed greater than 2 fold (on average) in cells exposed to venom vs. vehicle, and a set of 50 genes that are under-expressed greater than 2 fold (on average).

In a previous experiment, you exposed human endothelial cells to *Salmonella*, which identified a set of 1000 overexpressed genes, 7 of which were also in the set of 200 overexpressed genes in the rattlesnake venom experiment.

Examining the list of under-expressed genes, you recognize several that are associated with protein ubiquitination based on your reading of the literature. Using Gene Ontology (<http://amigo.geneontology.org/>), you identify a curated set of 35 human genes associated with protein ubiquitination (GO:0016567), 3 of which were in the set of 50 under-expressed genes in your rattlesnake venom experiment.

- 1) You wish to know whether there is more overlap in the overexpressed gene sets from the rattlesnake venom and *Salmonella* microarray experiments than you would expect by chance. What statistical approaches can you use to perform this assessment?
- 2) Likewise, what additional information do you need to know about the gene ontology database to determine whether there is more overlap in the set of genes under-expressed in rattlesnake venom and the GO protein ubiquitination gene set?
- 3) You find a database of 10,000 microarray experiments performed on the same platform as yours. What considerations must be taken into account if you are going to compare your differentially expressed gene sets to differentially expressed gene sets from this database of 10,000 experiments?