Functional Annotation of Gene Lists

Garrett Dancik, PhD

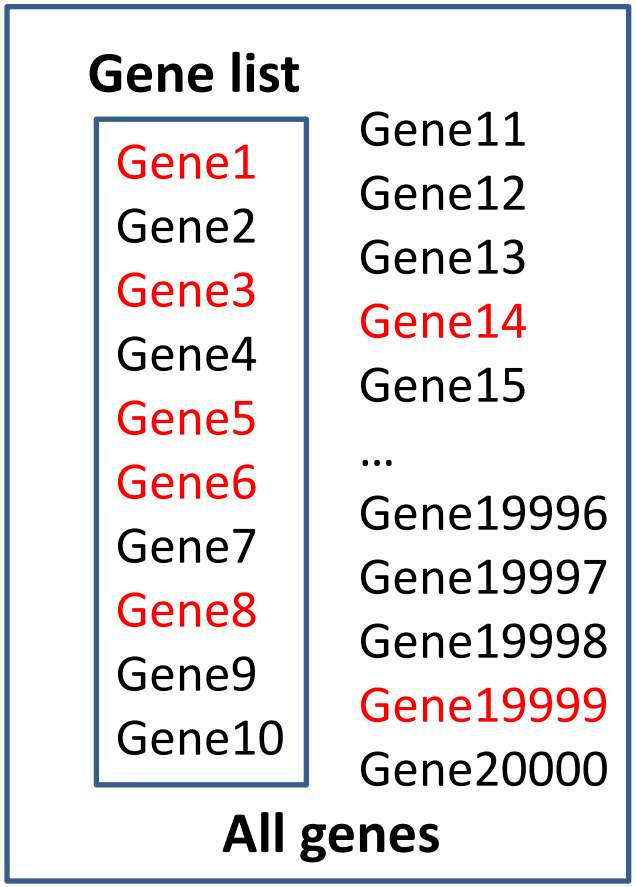
Functional Annotation

- After a bioinformatics analysis identifies a list of relevant probes, the follow-up questions are
  - What genes do the probes correspond to?
    - Look at GPL (GEO platform) data
  - What are the functions of the specific genes?
    - Look at, e.g., http://www.genecards.org
  - What biological processes or pathways are associated with these genes?
    - Use, e.g., DAVID (https://david.ncifcrf.gov)to perform a gene set enrichment analysis

Gene set enrichment analysis

In a list of interest, 50% of genes are related to a functional process, such as *cell cycle* (Gene Ontology, KEGG databases)

Out of all possible genes (background), 20% are related to the same functional process



The gene list is 2.5 times as likely to contain a cell cycle related gene as is the background

P-values (and adjusted p-values) determine whether the gene list is significantly enriched in cell cycle related genes?

Gene Ontology (GO)

- A controlled vocabularly for
  - Biological Processes (BP)
  - Molecular Functions (MF)
  - Cellular Components (CC)
- Relationships between terms leads to a "tree" structure
- http://geneontology.org
- Let's search TP53, look at Ontology, click on GO Term, and look at Graph Views

KEGG Pathways

- Molecular interactions and reactions related to metabolism, cellular processes, diseases, and others
- http://www.genome.jp/kegg/pathway.html
- Let's look at "pathways in cancer"