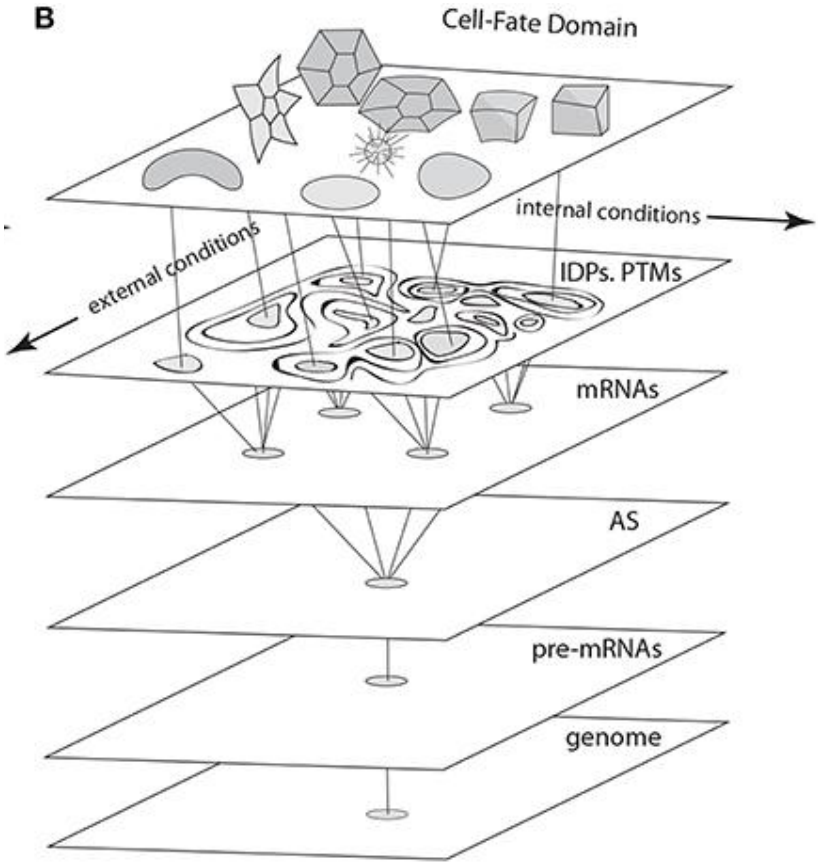
Gene Ontology

Genotype to phenotype

http://www.uniprot.org/uniprot/P42345

MLGTGPAAAT TAATTSSNVS VLQQFASGLK
SRNEETRAKA AKELQHYVTM ELREMSQEES
TRFYDQLNHH IFELVSSSDA NERKGGILAI
ASLIGVEGGN ATRIGRFANY LRNLLPSNDP
VVMEMASKAI GRLAMAGDTF TAEYVEFEVK
RALEWLGADR NEGRRHAAVL VLRELAISVP
TFFFQQVQPF FDNIFVAVWD PKQAIREGAV
AALRACLILT TQREPKEMQK PQWYRHTFEE
AEKGFDETLA KEKGMNRDDR IHGALLILNE
LVRISSMEGE RLREEMEEIT QQQLVHDKYC
KDLMGFGTKP RHITPFTSFQ AVQPQQSNAL
VGLLGYSSHQ GLMGFGTSPS PAKSTLVESR
CCRDLMEEKF DQVCQWVLKC NSKNSLIQM



Gene Ontology

Universe of concepts relating to **gene functions** (‘GO terms’), and how these functions are related to each other (‘relations’).

Ontology updates are made collaboratively between the GOC ontology team and scientists who request the updates. Most requests come from scientists making GO annotations and from domain experts in particular areas of biology

Submit requests for either new terms or new relations in the ontology!

Gene Ontology

Describe function with respect to three aspects:

- **molecular function**: molecular-level activities performed by gene products
- **cellular component**: the locations relative to cellular structures in which a gene product performs a function
- **biological process**: the larger processes, or ‘biological programs’ accomplished by multiple molecular activities

Commonly used relationships: is a (is a subtype of); part of; has part; regulates, negatively regulates and positively regulates

GO Annotations

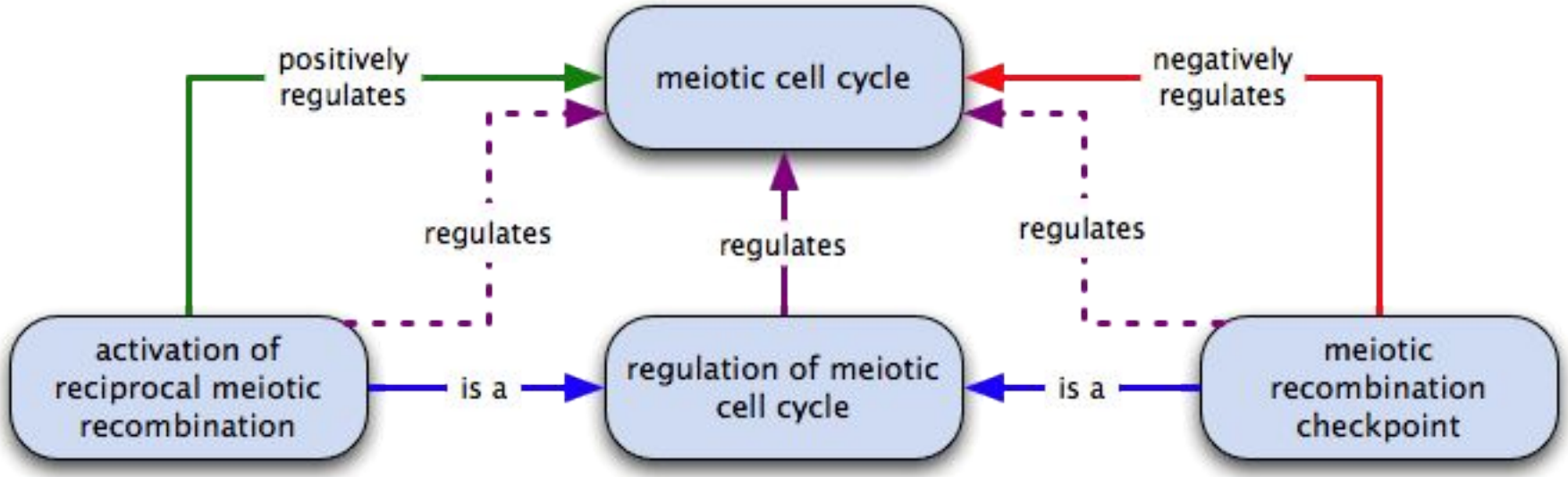
Statement about the function of a particular gene

GO evidence codes

- Experimentally-supported annotations
- Phylogenetically-inferred annotations
- Computationally-inferred annotations

The GO Consortium implements a number of automated queries to check the quality of the annotations submitted to the GO database

GO annotations and relationships



GO slim

Cut-down versions of the GO ontologies containing a subset of the terms in the whole GO

Available GO slims: http://www.geneontology.org/page/go-slim-and-subset-guide

Ten Quick Tips for Using the Gene Ontology:
http://journals.plos.org/ploscompbiol/article?id=10.1371/journal.pcbi.1003343

KEGG PATHWAY

Collection of manually drawn pathway maps representing our knowledge on the molecular interaction, reaction and relation networks for:

1. Metabolism	5. Organismal Systems
2. Genetic Information Processing	6. Human Diseases
3. Environmental Information Processing	7. Drug Development
4. Cellular Processes	

Gene Set Enrichment Analysis

Technique for highlighting biological processes is gene category over-representation analysis.

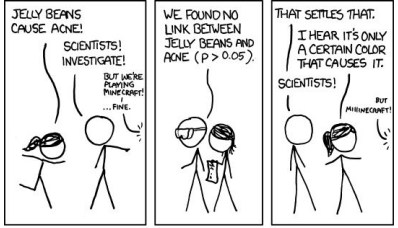
Genes are grouped into *a priori* categories (GO, biological pathway, location) and then tested to find categories that are over represented amongst the differentially expressed genes.

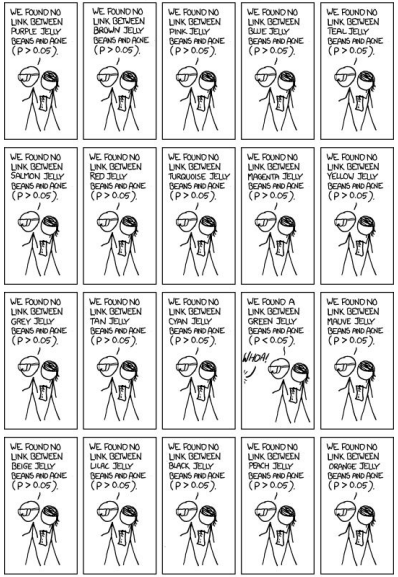
Tools available for performing GO analysis: GOseq (R), EasyGO, GOminer, GOstat, GOToolBox, topGO, GSEA, DAVID, AMIGO/PANTHER

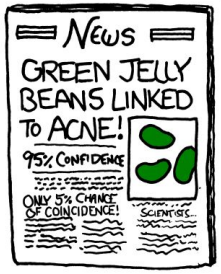
Gene Set Enrichment Analysis

0. Identify significantly DE genes between conditions (RNAseq and microarray data have different statistical suppositions)

1. Calculate the enrichment score (ES) that represents the amount to which the genes in the set are over-represented at either the top or bottom of the list.
2. Estimate the statistical significance of the ES. This calculation is done by a phenotypic-based permutation test in order to produce a null distribution for the ES.
3. Adjust for multiple hypothesis testing for when a large number of gene sets are being analyzed at one time. The enrichment scores for each set are normalized and a false discovery rate is calculated.


JELLY BEANS CAUSE ACNE!
SCIENTISTS! INVESTIGATE!
BUT WE'RE PLAYING MINECRAFT!
... FINE.
WE FOUND NO LINK BETWEEN JELLY BEANS AND ACNE (P > 0.05).
THAT SETTLES THAT.
I HEAR IT'S ONLY A CERTAIN COLOR THAT CAUSES IT.
SCIENTISTS!
BUT MINECRAFT!


WE FOUND NO LINK BETWEEN PURPLE JELLY BEANS AND ACNE (P > 0.05).
WE FOUND NO LINK BETWEEN BROWN JELLY BEANS AND ACNE (P > 0.05).
WE FOUND NO LINK BETWEEN PINK JELLY BEANS AND ACNE (P > 0.05).
WE FOUND NO LINK BETWEEN BLUE JELLY BEANS AND ACNE (P > 0.05).
WE FOUND NO LINK BETWEEN TEAL JELLY BEANS AND ACNE (P > 0.05).
WE FOUND NO LINK BETWEEN SALMON JELLY BEANS AND ACNE (P > 0.05).
WE FOUND NO LINK BETWEEN RED JELLY BEANS AND ACNE (P > 0.05).
WE FOUND NO LINK BETWEEN TURQUOISE JELLY BEANS AND ACNE (P > 0.05).
WE FOUND NO LINK BETWEEN MAGENTA JELLY BEANS AND ACNE (P > 0.05).
WE FOUND NO LINK BETWEEN YELLOW JELLY BEANS AND ACNE (P > 0.05).
WE FOUND NO LINK BETWEEN GREY JELLY BEANS AND ACNE (P > 0.05).
WE FOUND NO LINK BETWEEN TAN JELLY BEANS AND ACNE (P > 0.05).
WE FOUND NO LINK BETWEEN CYAN JELLY BEANS AND ACNE (P > 0.05).
WE FOUND A LINK BETWEEN GREEN JELLY BEANS AND ACNE (P < 0.05).
WHOA!
WE FOUND NO LINK BETWEEN MAUVE JELLY BEANS AND ACNE (P > 0.05).
WE FOUND NO LINK BETWEEN BEIGE JELLY BEANS AND ACNE (P > 0.05).
WE FOUND NO LINK BETWEEN LILAC JELLY BEANS AND ACNE (P > 0.05).
WE FOUND NO LINK BETWEEN BLACK JELLY BEANS AND ACNE (P > 0.05).
WE FOUND NO LINK BETWEEN PEACH JELLY BEANS AND ACNE (P > 0.05).
WE FOUND NO LINK BETWEEN ORANGE JELLY BEANS AND ACNE (P > 0.05).


News
GREEN JELLY BEANS LINKED TO ACNE!
95% CONFIDENCE
ONLY 5% CHANCE OF COINCIDENCE!
SCIENTISTS

Gene Set Enrichment Analysis

1. Load upregulated/downregulated gene list
2. Select knowledge base (GO function, KEGG, etc)
3. Select species
4. Select background ('any gene that COULD HAVE been positive')

- P-value is the probability or chance of seeing at least x number of genes out of the total n genes in the list annotated to a particular GO term, given the proportion of genes in the whole genome that are annotated to that GO Term. The closer the p-value is to zero, the more significant
- Background frequency is the number of genes annotated to a GO term in the entire background set
- Sample frequency is the number of genes annotated to that GO term in the input list.

Example: https://david.ncifcrf.gov/summary.jsp