BMMB 852: Applied Bioinformatics

Week 1, Lecture 3

István Albert

Bioinformatics Consulting Center
Penn State, 2015
What do the terms mean?

ialbert@grit ~/appbio/lec02
$ cat sc.tab | cut -f 1,2,4 | grep repeat | head -3
S000028864 telomeric_repeat
S000028866 X_element_combinatorial_repeat
S000006787 long_terminal_repeat YALWdelta1

ialbert@grit ~/appbio/lec02
$
Ontologies

• Structured vocabularies that offer both a definition and a classification (taxonomy)

• Intended to remove ambiguity in the terminology

• There may be multiple ontologies describing the same domain of knowledge. Each attempts to capture some dimension.
Let’s build an ontology

- Milky Way Galaxy
- PSU Old Main
- United States
- Earth
- Delta Quadrant
- Sun
- Halley Comet
- Voting District 5
- Happy Valley

- We could build a network of interrelating definitions
- We could also build a network of interrelating attributes
Defining biological phenomena

• **Sequence Ontology (SO)** → deals with the **definition** of ALL biological terms

• **Gene Ontology (GO)** → deals with the **functional characterization** of genes
Welcome to MISO, the Sequence Ontology Browser

- Search for a SO term by entering a SO term name or synonym in the query box above;
- Explore the structure of SO and browse for SO terms using the expandable, cascading tree on the left;
- Go to the detail page for a term where you can:
  - Get details about a term, its definition and relationships;
  - See graphical views of a term’s place in the ontology and link to its neighbors;
  - Export details about a term in a variety of formats;
  - Access and contribute detailed documentation about a term and its history by linking through to the SO wiki
In the image below graph nodes link to the appropriate terms. Clicking the image background will toggle the image between large and small formats.

Searching for

X_element_combinatorial_repeat
Another type of ontology

Gene Ontology
What is the Gene Ontology GO

An approach to describe gene products in terms of their associated:

1. biological processes
2. cellular components
3. molecular functions

in a *species-independent* manner.

Uses a controlled and structured vocabulary
GO Domains

• **Cellular Component** (localization) → endoplasmic reticulum, ribosome, protein dimer

• **Biological Process** (events) → signal transduction, alpha-glycoside transport

• **Molecular Function** (activity) → transporter activity, toll-like receptor binding activity
GO is directed graph
Annotation associations

• GO terms are abstract concepts in relation to one-another

• Association is a correspondence between a gene product and a GO term. Reflects the deposited knowledge on an organism.

• Both GO and associations change over time.

• What does one do if their gene is not annotated → various workarounds → related species → pick most similar annotated gene etc.
How to use the GO...

Ten Quick Tips for Using the Gene Ontology

Submitted by Monica Munoz-Torres on Tue, 11/26/2013 - 08:22

The Gene Ontology (GO) provides core biological knowledge representation for modern biologists, whether computationally or experimentally based. GO resources include biomedical ontologies that cover molecular domains of all life forms as well as extensive compilations of gene product annotations to these ontologies that provide largely species-neutral, comprehensive statements about what gene products do. Although extensively used in data analysis workflows, and widely incorporated into numerous data analysis platforms and applications, the general user of GO resources often misses fundamental distinctions about GO structures, GO annotations, and what can and can not be extrapolated from GO resources. The article referenced here offers ten quick tips for using the Gene Ontology.

To read more, visit http://www.ploscompbiol.org/article/info:doi/10.1371/journal.pcbi.100334...

The tips are mostly: make sure to read the documentation ...

The sites are not all that usable – takes quite a bit of trial and error
Gene Set Enrichment Analyses

• A group of feature names (typically genes) → Gene Set

• Find out what is “special/unexpected” about these.

• Characterize common features in terms of GO domains
Enrichment Analysis

• Given a selected set of genes and their annotations are there functional roles common to most of these genes?

• Enrichment analysis is typically this is one of the last steps of any analysis → Making sense of the results.

• There are many tools to do enrichment – may produce different results.
Enrichment search
Question: Why does each GO enrichment method give different results?

pixienon • 70 wrote:

I'm new to GO terms. In the beginning it was fun, as long as I stuck to one algorithm. But then I found that there are many out there, each with its own advantages and caveats (the quality of graphic representation, for instance).
Key concepts to remember

• There is **Sequence Ontology (SO)** – what type is a feature and how are types interrelated?

• And there is **Gene Ontology (GO)** – what does a feature do? how does it do it? and where does it do it?
Homework 3

1. From your feature file pick a **feature type** that is unfamiliar to you and look it up in the **Sequence Ontology**, paste the explanation into the homework.

2. Select (by any means) a group of gene names from your file. Perform a **Gene Ontology** (GO) enrichment analysis with Panther (on the GO main site).
   - What are some of the common annotations across the genes? Investigate all three classes of terms (biological process, molecular function, cellular component).
   - Are any of your terms statistically significant?

Hint: you may need to expand the results of the selection to see all entries.