Protein Networks and Systems Biology

ELIXIR 2015, Bologna, Italy

**Gene Ontology**

Luana Licata

Bioinformatics and Computational Biology Unit

Molecular Genetics Laboratory

University of Rome Tor Vergata, Italy

**Introduction**

**Gene Ontology** ([http://www.geneontology.org](http://www.geneontology.org" \o "http://www.geneontology.org" \t "_blank))

The Gene Ontology Consortium (GOC) is a worldwide group collaborating closely on the development of the Gene Ontology and gene function annotations.

The Gene Ontology project has developed ontologies for over 40,000 biological notions that are regularly revised to reflect new discoveries.

The Gene Ontology project offers:

* Structured controlled vocabularies (ontologies) that describes **biological processes**, **cellular components** and **molecular functions** associated togene products
* Gene products annotation using this ontology

**Step1: Ontology Search in IntAct**

Go to the IntAct advanced search page: <http://www.ebi.ac.uk/intact/search>.

Go down at the Ontology Search and write the GO term biological process “nitrogen compound metabolic process” (*"GO:0006807").*

**Q1:** How many binary interactions are associated to this GO term?

**Step2: GO quick search**

Go to the Amigo2 page: <http://amigo.geneontology.org/amigo/landing>.

Search for “muscle differentiation”

**Q2:** How many gene ontology terms are associated to this search? How many Genes and gene products associated with these GO terms?

# Click the ontology button; refine the search by filtering the entities that contain also the term “cardiomyocyte”. Then click on the term “cardiac muscle cell differentiation”. Go to found entities and refine search by choosing only the “homo sapiens” genes.

# Q3: How many genes are associated to these searches (muscle differentiation + cardiomyocyte + homo sapiens)?

Now refine moreover the search, by selecting only the entities “assigned by” UniProt.

**Q4:** How many genes are associated to this search? What GO\_REF:0000024 stand for?

**Step 3: Go enrichment analysis**

# Go to GO Enrichment Analysis Tool at: [www.geneontology.org/page/go-enrichment-analysis](http://www.geneontology.org/page/go-enrichment-analysis). Use the following list of genes, involved in a “well-known pathway”, to do a Go enrichment analysis.

**Genes list**:

|  |
| --- |
| AMBRA1 |
| AMPK |
| ATG10 |
| ATG12/5/16L1 |
| Atg13 |
| ATG14 |
| ATG3 |
| ATG4B |
| ATG7 |
| BECN1 |
| Fip200 |
| GABARAP |
| GABARAPL1 |
| GABARAPL2 |
| MAP1LC3C |
| mTORC1 |
| ULK1 |
| UVRAG |

Paste the Gene name list in the enrichment analysis window, select “cellular component” and “Homo sapiens”, then click search.

**Q5:** Looking at the result page, which are the 5 terms more enriched? How many genes of your list belong to that terms? Which ones have the highest P-values?

Paste the Gene name list in the enrichment analysis window, select “biological process” and “Homo sapiens”, then click search.

**Q6:** Looking at the result page, which are the 5 terms more enriched? How many genes of your list belong to that terms? Which ones have the highest P-values?

**Q7:** Can you make any biological conclusion analyzing these two results? To which pathway the genes belong?

If you finish earlier, you can select a list of genes of your interest related with your research project and do again an enrichment analysis using “cellular component” and “the taxid of your interest” and/or “biological process” and “the taxid of your interest”.

**Q8:** Which are the GO term enriched, if any? How many genes of your list belong to that terms? It is possible to make any speculation or conclusion analyzing your result?