# MSDA tutorial: How to extract GO ontologies for long lists of genes/proteins
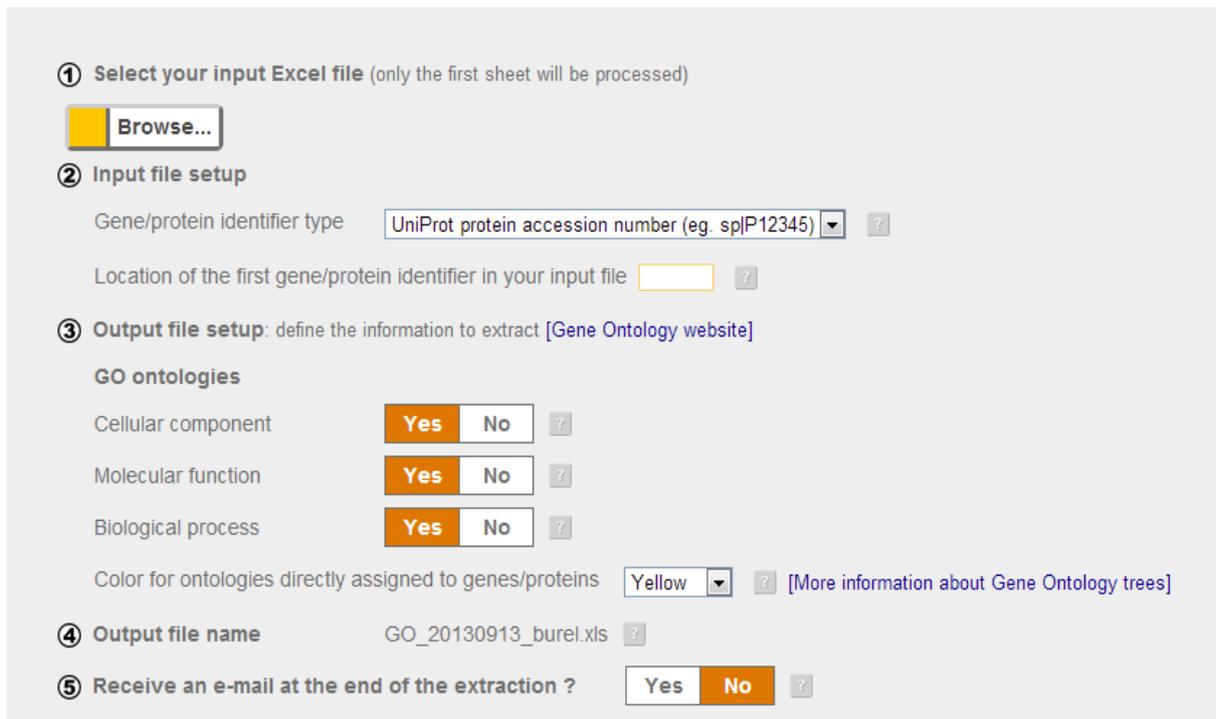
To extract GO ontologies from the GO database (http://www.geneontology.org/), select "Gene Ontology" in the "Annotation Explorer" menu.

Gene/protein identifiers (see below) should be listed in a sole column of an input Excel file. Multiple identifiers for a given gene/protein may be present in an Excel cell if they are separated by a semicolon.



## Step 1
Select your input Excel file using the browse button.

## Step 2
Specify the type of gene/protein identifier it contains and the location of the first cell containing an identifier. The extraction of GO ontologies can be performed from different identifiers:

- UniProtKB protein accession numbers (such as "sp|P12345" or "P12345")
- UniProtKB protein names (such as "1A43_HUMAN")
- NCBInr protein accession numbers (such as "gi|123456")
- Gene symbols (such as "Serpina3k")
  - Gene symbols are not unique identifiers and the studied taxonomy ID is also required in that case. If "Gene symbol" is chosen then the Taxonomy ID is asked.

### Step 3

Specify the GO ontologies (i.e. GO terms) you wish to extract among the following:

- Cellular component ontology
- Molecular function ontology
- Biological process ontology

You are also invited to choose a color for those GO ontologies that are directly assigned to gene/protein entries in the GO database. This can be useful when visualizing the GO graphs that contain also all the GO term ancestries (parental lineages)

### Step 4

Here you can find the name of the output Excel file you will obtain.

### Step 5

Choose whether or not you wish to be informed by an Email of the completion of your extraction.

### Step 6

At the end of the extraction, you will be able to download the output Excel file. This output file will be stored for 7 days and will remain available during that time from the "My GO files" page you can find in the "Annotation Explorer" menu.