

# BioinformaticsLibrary Manual

## Table of Contents

auto_extract_intersection_from_enrichment.....	1
extractIntersection.....	1
plot_PANTHER_enrich.....	2

## Index

---

auto\_extract\_intersection\_from\_enrichment

---

### Description

Will loop through all tabs of a GO enrichment .xlsx file to extract the “intersection” genes from a second, gene annotation file. Note each sheet name from the enrich\_file will be used as the treatment description. I.e., ensure that the sheet names in your enrich\_file are named something specific to your treatments (ex: LeafFeEx).

```
auto_extract_intersection_from_enrichment(  
    gene_file,  
    enrich_file,  
    output_name=None)
```

### Arguments

gene_file	Accepts a .csv or .xlsx file. The first column must be the gene identifiers that match the gene identifiers found in the “intersection” column of the “enrich_file”. If .xlsx file is provided, the first sheet will be used. The first column must provide the Gene_ID.
enrich_file	Standard GO enrichment output from any gProfiler enrichment function used in this Bioinformatics library.
output_name	User specified output file name. If no name is provided, the default file name, “Compiled_GO_Enrich_Genes” will be used.

### Examples

```
auto_extract_intersection_from_enrichment(gene_file='Updated_Sbi_20400  
tim_genes_TPM_samplemean_annotations.xlsx',  
Enrich_file='09022022_Sorghum_Cluster_GO_Enrichments.xlsx',  
  
Output_name='Sorghum_Intersection_Genes_from_Clus  
ter_Enrichment')
```

---

extractIntersection

---

### Description

Will extract a dataframe of gene information based on the “intersection” column that is an output from the gProfiler enrichment functions.

```
extractIntersection (  
    geneDF,
```

```
goEnrichDF)
```

### Arguments

geneDF	Dataframe of gene annotation information. The first column must contain Gene_ID information and must match the IDs in the “intersection” column of the goEnrichDF dataframe.
goEnrichDF	Dataframe of standard GO enrichment output from any gProfiler enrichment function used in this Bioinformatics library.

### Examples

```
extractIntersection(geneDF=genes,  
                   goEnrichDF=enrichment)
```

---

```
plot_PANTHER_enrich
```

---

### Description

Generate dotplot from standard PANTHER overenrichment output test file.

#### Usage

```
plot_PANTHER_enrich(  
    file,  
    title='PANTHER GO Enrichment',  
    total_query_genes=None,  
    sort_by='Name',  
    output_name=None,  
    plot_width=8,  
    plot_height=10)
```

### Arguments

file	Standard output .txt file from PANTHER overenrichment test.
title	Name of the figure title to be added to the top of the figure. Default is 'PANTHER GO enrichment'.
total_query_genes	Number of genes used in the GO enrichment query. This input number can be different than the number identified by PANTHER. Accepts integer values. If left blank, it will extract the number provided by the output file.
sort_by	Categories that can be used to sort the row order of the dot plot. Acceptable arguments are: 'Name', 'Pvalue', 'Counts', 'Ratio'.
output_name	Specifies the name of the output file image. The file extension will dictate the file format of the saved image. Acceptable file extensions are '.pdf', '.png', '.svg', '.eps'. If an acceptable file extension is not provided in the file name, a pdf will be generated with a default name and an error message.
plot_width	Specifies the width of the final image. Accepts integer values. Default is 8.
plot_height	Specifies the height of the final image. Accepts integer values. Default is 10.

**Examples**

```
plot_PANTHER_enrich(file='GOenrichment_dgcr_selected.txt',  
                    title='Example Dot Plot',  
                    total_query_genes=218,  
                    sort_by='Ratio',  
                    output_name='DGCR_Selected_Enrichment_Dotplot.png',  
                    plot_width=4,  
                    plot_height=5)
```