



Tutorial 1.2: Import Modules From Biomart

Import a module file that maps all human genes to Gene Ontology Biological processes terms and the annotations of these terms

- New
- Open
- Save
- Save As ...
- Close
- Import
- Export
- Exit

IntOGen

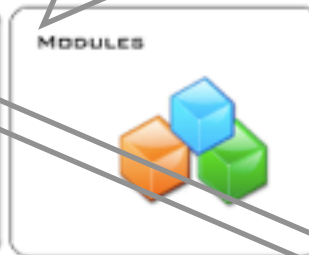
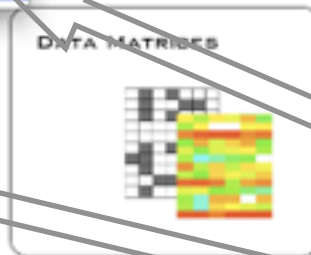
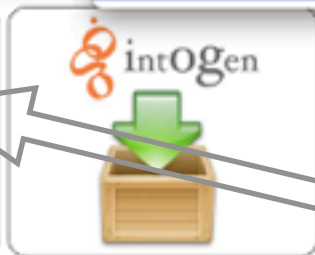
BioMart

Table ...

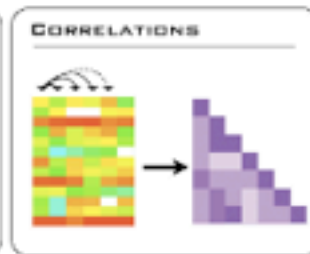
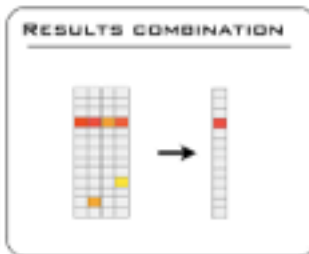
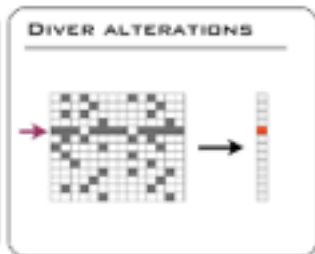
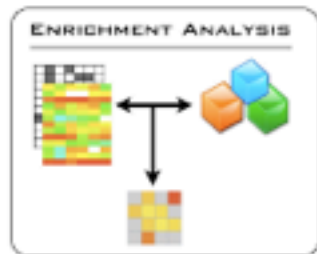
Modules ...

Data formats

These buttons will open a box that explains the different types of data used in Gitools



Start an analysis

**STEP 1**

Click Import Data
from Biomart or
select Import>
BioMart>Modules



Impo

bio

Start



Import Biomart data

What kind of data do you want to import ?

☐ Table

Select this option to download Biomart datasets.

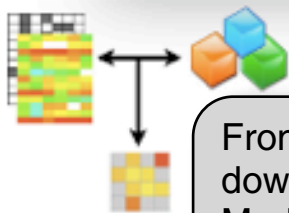
☒ Modules

Select this option to download modules (gene sets).

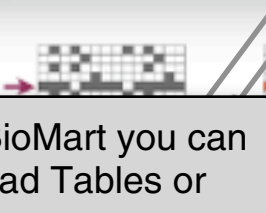
OK

Cancel

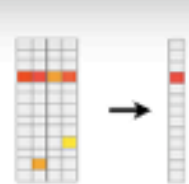
ENRICHMENT ANALYSIS



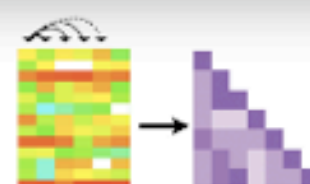
DIVER ALTERATIONS



RESULTS COMBINATION



CORRELATIONS



From BioMart you can
download Tables or
Modules.

STEP 2

Select Modules
and click OK

Import modules...

Select destination file



Name

Folder

Format

Generated file

STEP 3

Give a name to the file that will be downloaded and indicate a Folder where to place it. Click Next.

Select biomart source

biomart



Portal Ensembl Last release Portal

Databases

Datasets

Ensembl Genes 57
Ensembl Variation 57
Ensembl Functional Genomics 57
Vega 3
Reactome
WormBase (CSHL US)
PRIDE (EBI UK)

Ornithorhynchus anatinus genes (OANA5)
Taeniopygia guttata genes (taeGut3.2.4)
Cavia porcellus genes (cavPor3)
Gasterosteus aculeatus genes (BROADS1)
Loxodonta africana genes (loxAfr3)
Myotis lucifugus genes (myoLuc1)
Homo sapiens genes (GRCh37)
Choloepus hoffmanni genes (choHof1)
Ciona savignyi genes (CSAV2.0)
Felis catus genes (CAT)
Rattus norvegicus genes (RGSC3.4)
Gallus gallus genes (WASHUC2)
Tupaia belangeri genes (tupBel1)
Kenopus tropicalis genes (JGI4.1)
Equus caballus genes (EquCab2)

STEP 4

Select the Biomart database from which you want to import data. In this case Ensembl Last release Portal

STEP 6

Select the dataset. In this case Homo sapiens genes.

STEP 5

In this case Ensembl Genes 57

< Back

Next >

Cancel

Finish

Help

Import modules...

Select attribute for items

bio::mart



Clear

Features

▼ GENE

▼ Ensembl

- ☒ Ensembl Gene ID
- ☐ Ensembl Transcript ID
- ☐ Ensembl Protein ID
- ☐ Canonical transcript stable ID(s)
- ☐ Description
- ☐ Chromosome Name
- ☐ Gene Start (bp)
- ☐ Gene End (bp)
- ☐ Strand
- ☐ Band
- ☐ Transcript Start (bp)
- ☐ Transcript End (bp)
- ☐ Associated Gene Name

STEP 7

Select the attribute you want for each item. This must correspond with the attribute of the rows in you data matrix. In this case since the attribute used in IntOGen matrix are Ensembl Gene IDs, we will choose this for attributes of items in the module file. Click Next.

Expand all

Collapse all

< Back

Next >

Cancel

Finish

Help



Import modules...

Select attribute for modules

Clear

Features

GENE

Ensembl

EXTERNAL

go biological process (bp)

GO Term Accession (bp)

GO Term Name (bp)

GO Term Definition (bp)

GO Term Evidence Code (bp)

go cellular component (cc)

GO Term Accession (cc)

GO Term Name (cc)

GO Term Definition (cc)

GO Term Evidence Code (cc)

go molecular function (mf)

GO Term Accession

Expand all

Collapse all

< Back

Next >

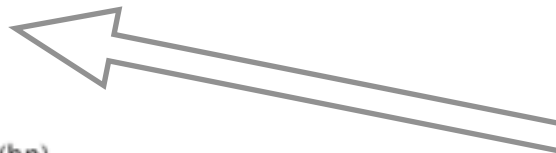
Cancel

Finish

Help

STEP 8

Select the attribute you want for each Module. In this case we will select GO Term Accession for GO Biological Process. And click Finish.



▼ modulefiles
gene2GOprocess.tsv

Gene ID	GO Term ID
ENSG00000215754	GO:0006355
ENSG00000230380	GO:0006355
ENSG00000233794	GO:0006355
ENSG00000212884	GO:0008152
ENSG00000212884	GO:0050872
ENSG00000212884	GO:0045892
ENSG00000215756	GO:0006355
ENSG00000229617	GO:0006355
ENSG00000215750	GO:0000226
ENSG00000229450	GO:0007049
ENSG00000215700	GO:0045449
ENSG00000215700	GO:0031087
ENSG00000215700	GO:0000184
ENSG00000215615	GO:0000226
ENSG00000204879	GO:0006355
ENSG00000231089	GO:0006355
ENSG00000218497	GO:0006355
ENSG00000217642	GO:0006355
ENSG00000184895	GO:0007548
ENSG00000184895	GO:0030154
ENSG00000184895	GO:0045449
ENSG00000184895	GO:0030238
ENSG00000224558	GO:0006355
ENSG00000185602	GO:0007165
ENSG00000233698	GO:0006355
ENSG00000228710	GO:0006355
ENSG00000233294	GO:0006355
ENSG00000236575	GO:0006355
ENSG00000215600	GO:0000244

Once the download has been completed you will have a new file in the directory that you indicated.

This file is a two columns mapping. The first column indicates the name of the gene or item and the second column the ID of the module.

In this case the ID for modules is GO term Accession, but it would be useful to also have the GO description, to recognize easily each of the biological processes. For that we are going to download a table that contains annotations for GO terms.



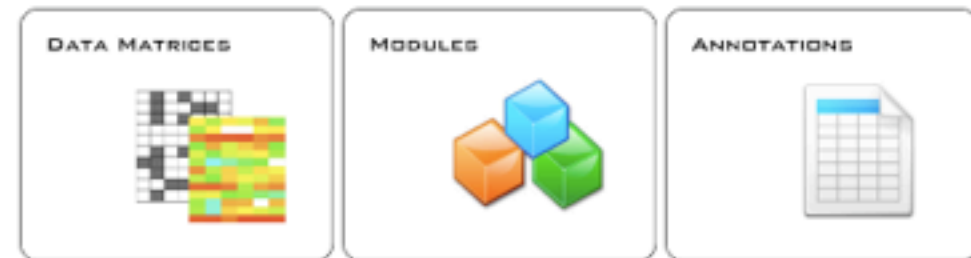
Welcome X Demo X



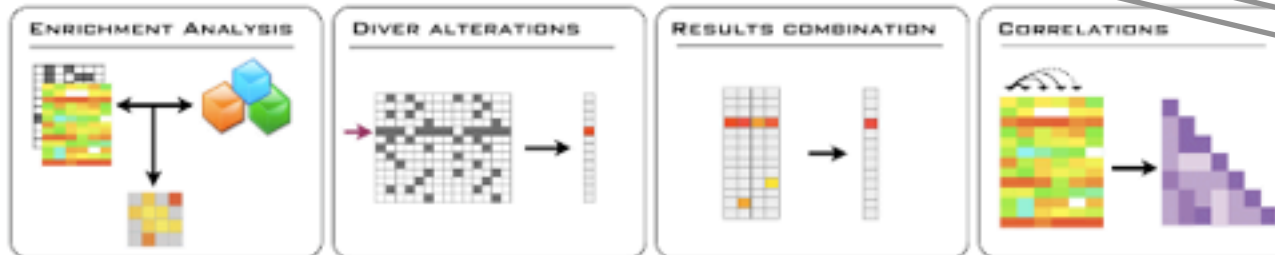
Import data



Data types



Start an analysis



STEP 9

Click Import Data
from Biomart



Welcome x Demo x

Import Biomart data

What kind of data do you want to import ?

☒ Table

Select this option to download Biomart datasets.

☐ Modules

Select this option to download modules (gene sets).

OK

Cancel



Import data



Start an analysis

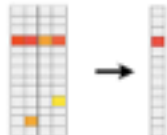
ENRICHMENT ANALYSIS



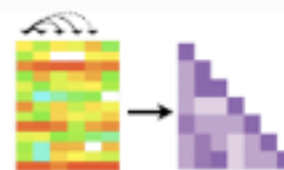
DIVER ALTERATIONS



RESULTS COMBINATION



CORRELATIONS

**STEP 10**Select Table and
click OK



Import table ...

Select destination file



Name

Folder

Format

Generated file

STEP 11

Give a name to the file that will be downloaded and indicate a Folder where to place it. Click Next.

Select biomart source

biomart



Portal Ensembl Last release Portal

Databases

Datasets

Ensembl Genes 57
Ensembl Variation 57
Ensembl Functional Genomics 57
Vega 3
Reactome
WormBase (CSHL US)
PRIDE (EBI UK)

Ornithorhynchus anatinus genes (OANA5)
Taeniopygia guttata genes (taeGut3.2.4)
Cavia porcellus genes (cavPor3)
Gasterosteus aculeatus genes (BROADS1)
Loxodonta africana genes (loxAfr3)
Myotis lucifugus genes (myoLuc1)
Homo sapiens genes (GRCh37)
Choloepus hoffmanni genes (choHof1)
Ciona savignyi genes (CSAV2.0)
Felis catus genes (CAT)
Rattus norvegicus genes (RGSC3.4)
Gallus gallus genes (WASHUC2)
Tupaia belangeri genes (tupBel1)
Xenopus tropicalis genes (JGI4.1)
Equus caballus genes (EquCab2)

STEP 12

Select the Biomart database from which you want to import data. In this case Ensembl Last

STEP 13

In this case Ensembl Genes 57

STEP 14

Select the dataset. In this case Homo sapiens genes.

< Back

Next >

Cancel

Finish

Help



Import table ...

Select attributes

Press [Add...] button to select attributes

blo+mart



Add...

Remove

Up

Down

Load...

Save...

STEP 15

Click Add button to
start adding fields in
the table

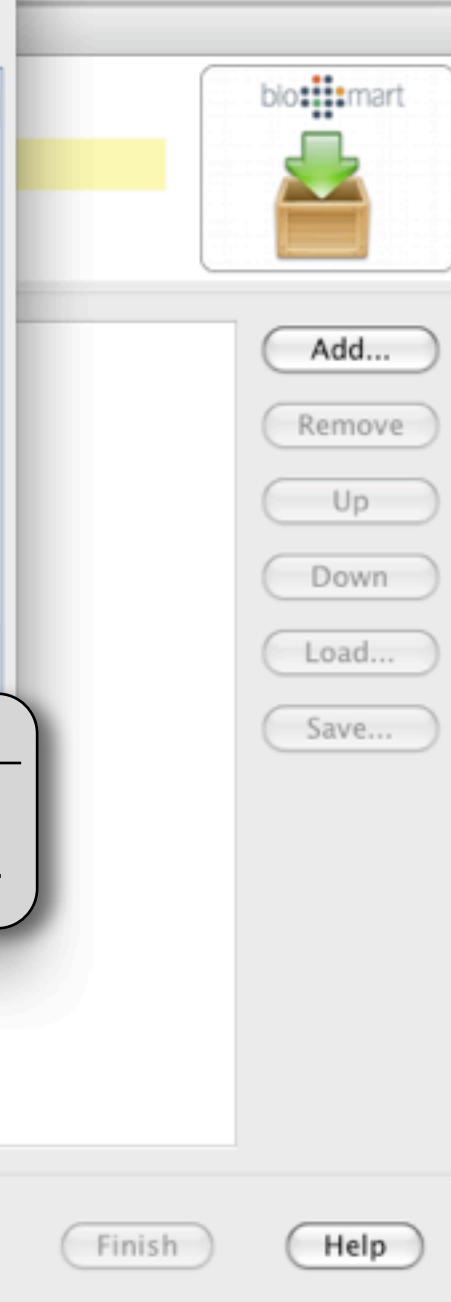
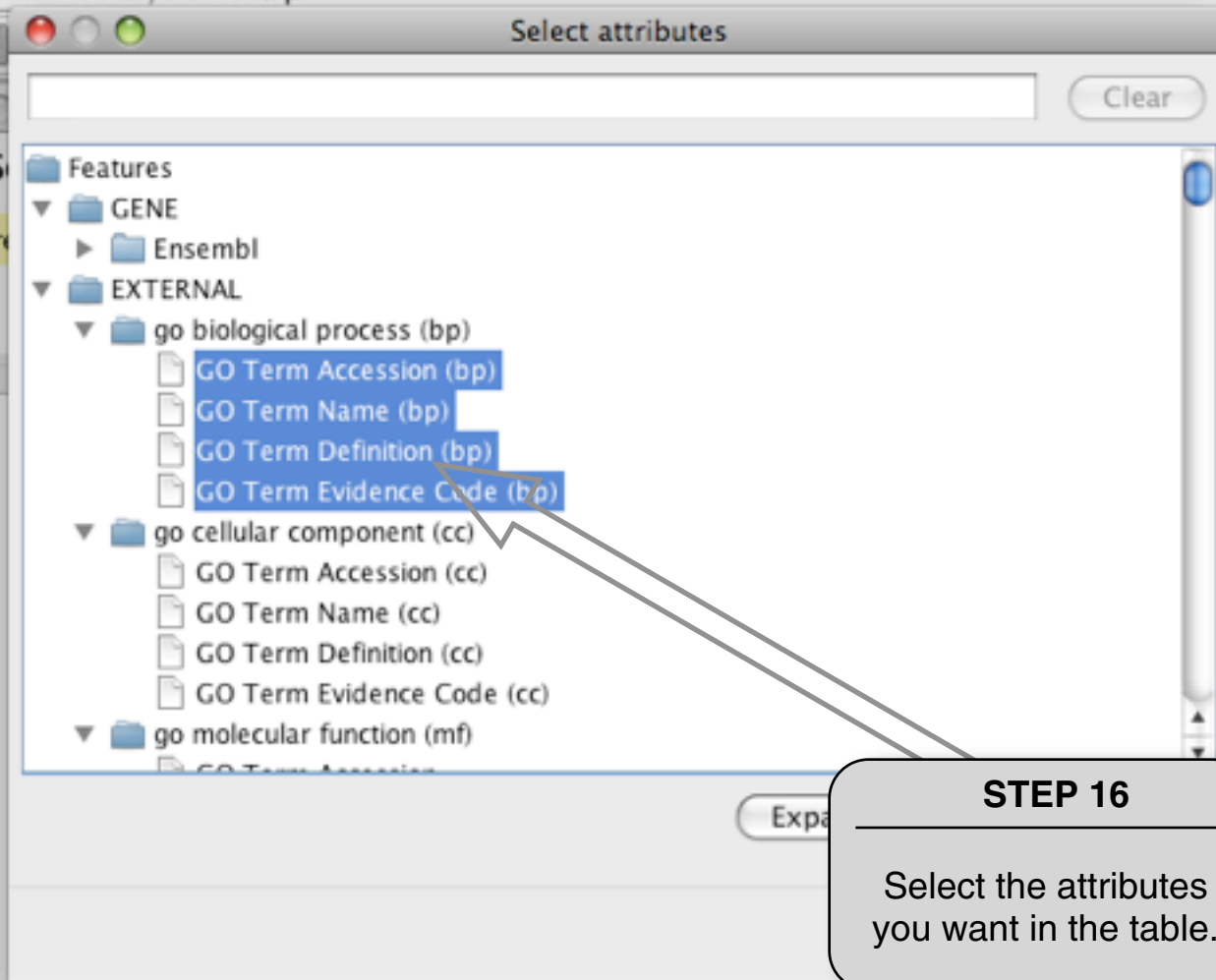
< Back

Next >

Cancel

Finish

Help



< Back

Next >

Cancel

Finish

Help



Import table ...

Select attributes

Press [Add...] button to select attributes

biomart



Features > EXTERNAL > go biological process (bp) > GO Term Accession (bp)
 Features > EXTERNAL > go biological process (bp) > GO Term Name (bp)
 Features > EXTERNAL > go biological process (bp) > GO Term Definition (bp)
 Features > EXTERNAL > go biological process (bp) > GO Term Evidence Code (bp)

Add...

Remove

Up

Down

Load...

Save...

Now you can see the list of attributes selected for the table. You could continue adding attributes by clicking Add button. You could remove or change the order of the attributes using the buttons at the right.

The first attribute in the table must correspond to the attribute in the matrix or module. In this case the attribute of the modules is GO Term Accession.

STEP 17

Click Finish

< Back

Next >

Cancel

Finish

Help

Once the download has been completed you will have a new file in the directory that you indicated.

modulefiles
GOprocessID2annotations.tsv

	A	B	C	D
1	GO Term Accession	GO Term Name (bp)	GO Term Definition (bp)	GO Term Evidence Code (bp)
2	GO:0006355	regulation of transcription, DNA-depen	Any process that modulates the frec	IEA
3	GO:0008152	metabolic process	The chemical reactions and pathway	IEA
4	GO:0050872	white fat cell differentiation	The process whereby a relatively un	IEA
5	GO:0045892	negative regulation of transcription, DN	Any process that stops, prevents or	IEA
6	GO:0000226	microtubule cytoskeleton organization	A process that is carried out at the c	IEA
7	GO:0007049	cell cycle	The progression of biochemical and	IEA
8	GO:0045449	regulation of transcription	Any process that modulates the frec	IEA
9	GO:0031087	deadenylation-independent decapping	Cleavage of the 5'-cap of a nuclear-	TAS
10	GO:0000184	nuclear-transcribed mRNA catabolic pro	The nonsense-mediated decay path	IDA
11	GO:0006355	regulation of transcription, DNA-depen	Any process that modulates the frec	NAS
12	GO:0007548	sex differentiation	The establishment of the sex of an c	IEA
13	GO:0030154	cell differentiation	The process whereby relatively unsp	IEA
14	GO:0030238	male sex determination	The specification of male sex of an i	NAS
15	GO:0007165	signal transduction	The cascade of processes by which	IEA
16	GO:0000244	assembly of spliceosomal tri-snRNP	The formation of a tri-snRNP comple	NAS
17	GO:0006376	mRNA splice site selection	Selection of a splice site by compon	IDA
18	GO:0045449	regulation of transcription	Any process that modulates the frec	NAS
19	GO:0016482	cytoplasmic transport	The directed movement of substanc	IDA
20	GO:0006406	mRNA export from nucleus	The directed movement of mRNA fr	NAS
21	GO:0048025	negative regulation of nuclear mRNA sp	Any process that stops, prevents or	IDA
22	GO:0000375	RNA splicing, via transesterification rea	Splicing of RNA via a series of two t	IDA
23	GO:0007242	Intracellular signaling cascade	A series of reactions within the cell	IEA
24	GO:0007264	small GTPase mediated signal transduc	Any series of molecular signals in w	IEA
25	GO:0015031	protein transport	The directed movement of proteins	IEA
26	GO:0016192	vesicle-mediated transport	The directed movement of substanc	IEA
27	GO:0006886	intracellular protein transport	The directed movement of proteins	IEA
28	GO:0007156	homophilic cell adhesion	The attachment of an adhesion mol	IEA
29	GO:0007155	cell adhesion	The attachment of a cell, either to a	IEA
30	GO:0007165	signal transduction	The cascade of processes by which	TAS

This file is a table that contains the fields that we have selected



THANKS FOR USING GITTOOLS

<http://www.gittools.org>