

BABELOMICS

Microarray Data Analysis Functional Enrichment: FatiGO

Martina Marbà

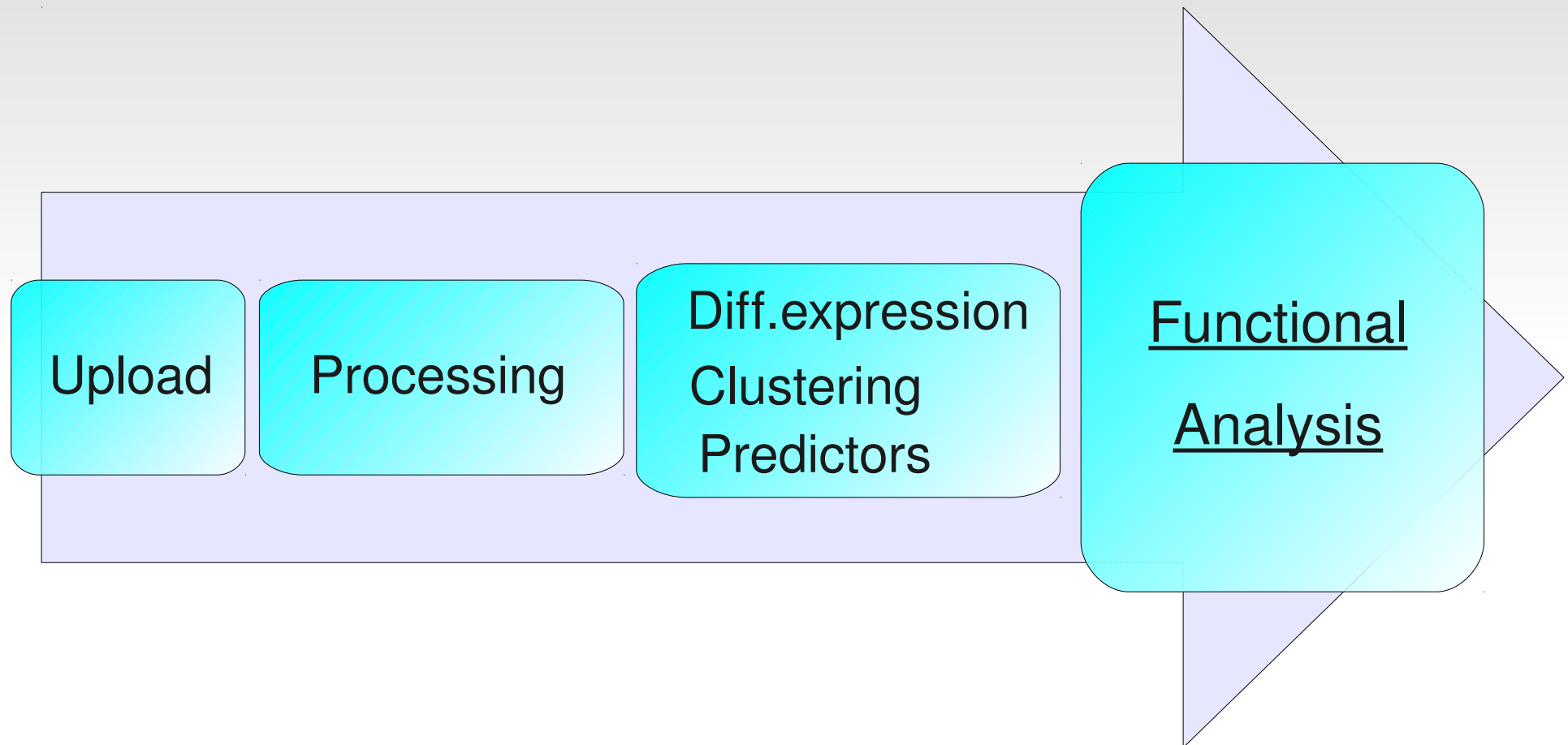
mmarba@cipf.es

Valencia, June 2010

*Bioinformatics and Genomics Department
Centro de Investigacion Principe Felipe (CIPF)
(Valencia, Spain)*



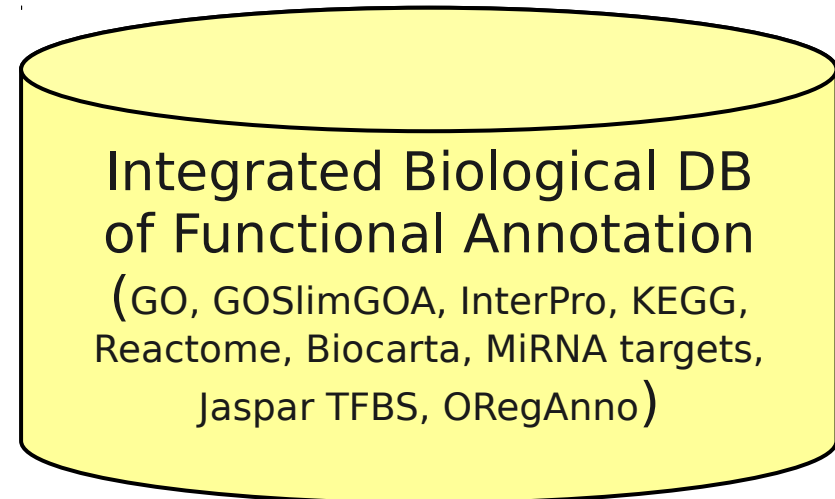
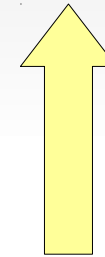
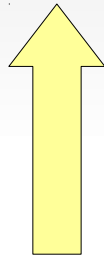
Data analysis workflow



Fatigo

Schema

FatiGO is a web tool for:
statistical test, multiple test corrections, filtering ...



FatiGO

Questions that Functional enrichment analysis try to answer

- Is there any significant functional enrichment in my gene list?
- Are these genes involved in same pathways?
- Are they sharing a specific microRNA regulation?
- Are they involved in the same disease?

FatiGO

- A web-based tool for the functional profiling of genome-scale experiments

The screenshot displays the Babelomics 4 web interface. At the top, the logo 'BABELOMICS 4' is shown with the subtitle 'gene expression and functional profiling analysis suite'. Below the logo is a navigation bar with several tabs: 'Upload data', 'Processing data', 'Expression', 'Genomic', 'Functional analysis', and 'Utilities'. The 'Functional analysis' tab is highlighted with a red circle. Below the navigation bar, a status bar indicates the user 'mmarba@cipf.es' is working on a project named 'Pre-processing Agilent' with 91.30 Mb of 1.00 Gb (8.92%) used and no active sessions. A green message box contains the text: 'Welcome to the new Babelomics 4, you can still use Babelomics 3 at: <http://babelomics3>'. Below this, the 'Functional analysis' section is titled. Underneath, there is a 'Single enrichment analysis' section with a list of tools. The 'FatiGO' tool is highlighted with a red circle and has a red arrow pointing to it from the 'Functional analysis' tab. The description for FatiGO is: 'Resource to show significant over-representation of GO terms.' Below it, the 'Marmite' tool is listed with the description: 'Extracts blocks of related genes from an ordered list of genes by an associated value to the Marmite tool'.

BABELOMICS 4
gene expression and functional profiling analysis suite

Upload data Processing data Expression Genomic **Functional analysis** Utilities

mmarba@cipf.es working on project *Pre-processing Agilent* 91.30 Mb of 1.00 Gb (8.92%) no active

Welcome to the new Babelomics 4, you can still use Babelomics 3 at: <http://babelomics3>

Functional analysis

Single enrichment analysis

- **FatiGO**
Resource to show significant over-representation of GO terms.
- Marmite
Extracts blocks of related genes from an ordered list of genes by an associated value to the Marmite tool

FatiGO features

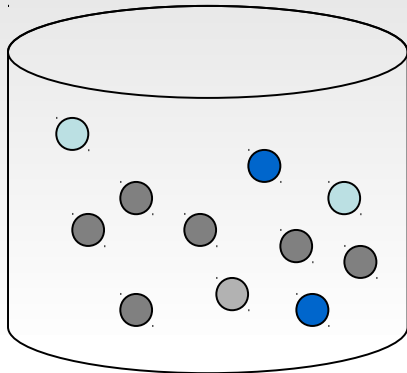
- It allows us to compare functional annotation of:
 - **Two** list of genes
 - **One** list against the rest of genome
 - Lists of genes with user **submitted annotations**
- One statistical test for each Functional **Block** of annotation
 - Fisher's exact test
 - Multiple testing context (hundreds of annotation)
 - Filtering of annotation is convenient (the less tests the best correction)

FatiGO features

- It allows us to compare functional annotation of:
 - **Two** list of genes
 - **One** list against the rest of genome
 - Lists of genes with user **submitted annotations**
- One statistical test for each Functional **Block** of annotation
 - Fisher's exact test
 - Multiple testing context (hundreds of annotation)
 - Filtering of annotation is convenient (the less tests the best correction)

FatiGO test

One Gene List (A)

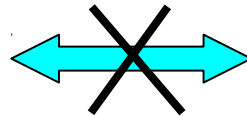
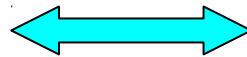


Biosynthesis 60% ●

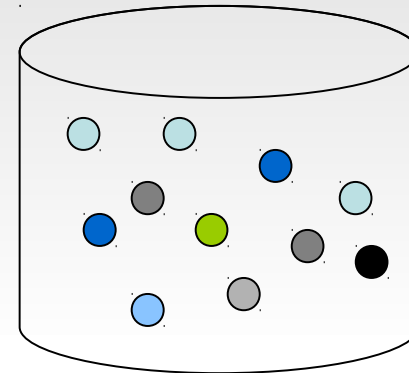
Sporulation 20% ●

Genes in group A have significantly to do with biosynthesis, but not with sporulation.

Are this two groups of genes carrying out different biological roles?



The other list (B)



Biosynthesis 20% ●

Sporulation 20% ●

	A	B
Biosynthesis	6	2
No biosynthesis	4	8

We do this for each term (GO, miRNA, Interpro , ...)!!!

FatiGO form

Define your comparison

- Id list vs Id list
- Id List vs Rest of genome
- Id List vs Rest of ids contained in your annotations (complementary list)

Select your data

List 1 :

no data selected.
Or go to Upload Data form: [Upload \[idlist\]](#)

List 2 :

no data selected.
Or go to Upload Data form: [Upload \[idlist\]](#)

Options

Fisher exact test

Remove duplicates?

Databases

Organism

GO biological process [\[options\]](#)

FatiGO form

Do you want to compare 2 conditions or one vs the rest of genome ?

Define your comparison

Id list vs Id list
 Id List vs Rest of genome
 Id List vs Rest of ids contained in your annotations (c

Select your data

List 1 : no data selected.
Or go to Upload Data form: [Upload \[idlist\]](#)

List 2 : no data selected.
Or go to Upload Data form: [Upload \[idlist\]](#)

Options

Fisher exact test
Remove duplicates?

Databases

Organism
 GO biological process [\[options\]](#)

eg. Compare 2 tissues or responder genes vs. non-responders

eg: genes that respond to one treatment against the genome

FatiGO form

Define your comparison

- Id list vs Id list
- Id List vs Rest of genome
- Id List vs Rest of ids contained in your annotations (c

Upload first at Data Upload

Select your data

List 1 : no data selected.
Or go to Upload Data form: [Upload \[idlist\]](#)

List 2 : no data selected.
Or go to Upload Data form: [Upload \[idlist\]](#)

“txt” file with
gene lists:

```
gene1  
gene2  
gene3  
...
```

Data
selection

Options

Fisher exact test

Remove duplicates?

Databases

Organism

GO biological process [\[options\]](#)

FatiGO form

Define your comparison

- Id list vs Id list
- Id List vs Rest of genome
- Id List vs Rest of ids contained in your annotations (c

Select your data

List 1 :

no data selected.
Or go to Upload Data form: [Upload \[idlist\]](#)

List 2 :

no data selected.
Or go to Upload Data form: [Upload \[idlist\]](#)

Options

Fisher exact test

Remove duplicates?

Databases

Organism

GO biological process [\[options\]](#)

HINT:

- *Two tailed* for 2 lists
- *One tailed* for 1list vs rest of genome (or your annotations)

Algorithm options

Removing duplicates:

- Choose one or other option depends on from where gene lists come from.

FatiGO form

Define your comparison

- Id list vs Id list
- Id List vs Rest of genome
- Id List vs Rest of ids contained in your annotations (c

Select your data

List 1 : no data selected.
Or go to Upload Data form: [Upload \[idlist\]](#)

List 2 : no data selected.
Or go to Upload Data form: [Upload \[idlist\]](#)

Options

Fisher exact test

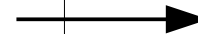
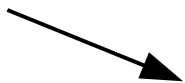
Remove duplicates?

Databases

Organism

GO biological process [\[options\]](#)

Which type of functional information?



FatiGO form

Which type of functional information?

The screenshot shows the 'Databases' section of the FatiGO form. It includes a dropdown menu for 'Organism' and a list of database options, each with a radio button and a link to 'options'. At the bottom, there is a 'browse server' button, a 'Your annotations' checkbox, and a link to 'Upload [annotation]'. Two callout boxes with arrows point to the database list and the 'Your annotations' option, respectively.

Databases

Organism ▼

- GO biological process [\[options\]](#)
- GO molecular function [\[options\]](#)
- GO cellular component [\[options\]](#)
- GOSlim GOA [\[options\]](#)
- Interpro [\[options\]](#)
- KEGG pathways [\[options\]](#)
- Reactome [\[options\]](#)
- Biocarta [\[options\]](#)
- miRNA targets [\[options\]](#)
- Jaspar TFBS [\[options\]](#)
- ORegAnno [\[options\]](#)

Your annotations no data selected.
Or go to Upload Data form: [Upload \[annotation\]](#)

Use one or more of the given databases

If it is not in the databases, use your annotations option.

FatiGO form

A.

Databases

Organism: Human (homo sapiens)

GO biological process [options]

GO molecular function [options]

GO cellular component [options]

First select an organism

OPTIONS:

Test all the GO or only annotated terms

Discard functions with too few or too many genes?

If you have an hypothesis, better test this first!!!!!!

GO biological process options

GO parameters

Select annotation through ontology levels

Propagate annotation to upper levels

Direct annotation

GO level must be among levels and

Filter terms by number of annotated ids in DB

Minimum (typically 5-20)

Maximum (typically 500-Inf)

Number of annotated ids is computed from

Genome

Your input ids

Filter terms by keywords

Keywords (e.g. metabolism cancer)

Your search must match

all keywords

any keyword

Add children of selected terms

FatiGO form

B.

Which type of functional information?

Databases

Organism

- GO biological process [\[options\]](#)
- GO molecular function [\[options\]](#)
- GO cellular component [\[options\]](#)
- GOSlim GOA [\[options\]](#)
- Interpro [\[options\]](#)
- KEGG pathways [\[options\]](#)
- Reactome [\[options\]](#)
- Biocarta [\[options\]](#)
- miRNA targets [\[options\]](#)
- Jaspar TFBS [\[options\]](#)
- ORegAnno [\[options\]](#)
- Your annotations no data selected.
Or go to Upload Data form: [Upload \[annotation\]](#)

Job

job name:

job description:

Your annotations: useful when you work with your own annotations OR with an organism that is not in Babelomics

Upload first at Data Upload

Example (your annotations):

38969_at	GO:0003677
37639_at	GO:0006306
37149_s_at	GO:0004674
37149_s_at	GO:0005525
37639_at	GO:0006306
37149_s_at	GO:0004674
...	...

FatiGO form

Databases

Organism

- GO biological process [\[options\]](#)
- GO molecular function [\[options\]](#)
- GO cellular component [\[options\]](#)
- GOSlim GOA [\[options\]](#)
- Interpro [\[options\]](#)
- KEGG pathways [\[options\]](#)
- Reactome [\[options\]](#)
- Biocarta [\[options\]](#)
- miRNA targets [\[options\]](#)
- Jaspas TFBS [\[options\]](#)
- ORegAnno [\[options\]](#)

Your annotations no data selected.
Or go to Upload Data form: [Upload \[annoti](#)

What's
your
job name?

Job

job name:

job description:

Set up a job name
and optionally,
give a description.

FatiGO

Running an example ...

FatiGO

Example:

Description

Molecular Apocrine Breast Cancer dataset:

- 49 Affymetrix (HG-U133A), 14,500 genes
- Human
- 3 tumor classes:
apocrine, basal and luminal.

FatiGO input

- 1) Run *differential expression* (using ANOVA test) because we are comparing 3 conditions
- 2) Send results to FatiGO:



- 3) Input FatiGO form:
 - Comparison: list vs rest of genome
 - Test: one tailed test

FatiGO results

Summary results:

■ Id annotations per DB :

<i>DB</i>	<i>List1</i>	<i>Genome</i>
GO biological process (levels from 3 to 9)	350 of 500 (70%) 11.26 annotations/id	11716 of 23198 (2343.2%) 5.08 annotations/id
GO molecular function (levels from 3 to 9)	344 of 500 (68.8%) 3.05 annotations/id	11370 of 23198 (2274%) 1.92 annotations/id

Tables significant terms:

▼ Significant Results

■ Number of significant terms per DB :

<i>DB</i>	<i>Number of significant terms</i>
GO biological process (levels from 3 to 9)	142
GO molecular function (levels from 3 to 9)	30

FatiGO results

Significant results:

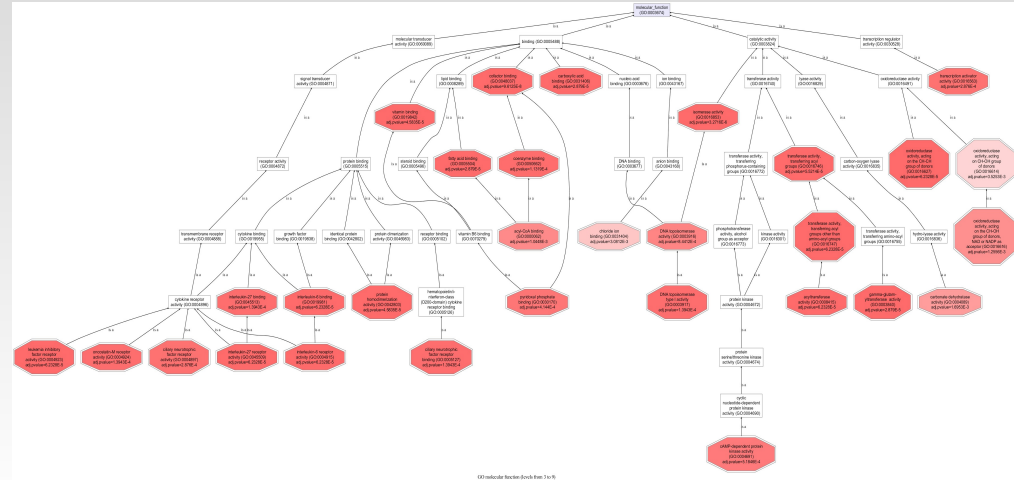
Term	Term size	Term size (in genome)	Term annotation % per list	Annotated ids	Odds ratio (log _e)	pvalue	Adjusted pvalue ▲
negative regulation of apoptosis (GO:0043066)	412	403	list 1: 7.2% list 2: 1.62%	list 1: 205225_at,20979... list 2: ENSG00000001084,ENSG...	1.5495	7.006e-13	7.65e-10
negative regulation of programmed cell death (GO:0043069)	418	409	list 1: 7.2% list 2: 1.65%	list 1: 205225_at,20979... list 2: ENSG00000001084,ENSG...	1.5334	1.074e-12	7.65e-10
cellular amino acid derivative metabolic process (GO:0006575)	182	173	list 1: 4.8% list 2: 0.68%	list 1: 209604_s_at,209... list 2: ENSG00000001084,ENSG...	1.995	9.24e-13	7.65e-10
cellular amino acid and derivative metabolic process (GO:0006519)	447	447	list 1: 7.4% list 2: 1.77%	list 1: 209604_s_at,209... list 2: ENSG00000001084,ENSG...	1.491	1.7e-12	9.082e-10

↑
Enriched class

↑
Annotated genes per GO from each list

FatiGO results

Graphical results:



More results:

▼ Other actions

- Open input form
- Change p value

0.1 0.05 0.01 **0.005**

FatiGO

Exercises

- Go to the tutorial:
http://bioinfo.cipf.es/babelomicstutorial/enrichment_analysis
- Run worked examples
- Repeat examples modifying parameters
- Run FatiGO exercise (from the tutorial)