BABELOMICS:
Microarray Data Analysis

Barcelona, 23 November 2010

Martina Marbà
mmarba@cipf.es

Bioinformatics and Genomics Department
Centro de Investigación Príncipe Felipe (CIPF)
(Valencia, Spain)
DNA Microarrays

- Paradigm of High Throughput Technologies
- Yield concentration measurements for: genes, SNP, exons, mRNA ...
- Measure cells in different biological conditions
- In a genomic scale
- Allow us conducting biological experiments

So... How do they work?
Central Dogma of Molecular Biology

For a cell, at a particular time, thousands of mRNA are created and sent out of the nucleus to be translated into proteins.

Protein concentration regulates biological systems.
We want to know which genes are expressed under particular biological conditions.

We can extract all mRNA molecules that are being translated within the cells and provide an expression level indicator of its concentration in the biological sample.
RNA Extraction

**BREAKING CELLS AND TISSUES**
The first step in the purification of most proteins is to disrupt tissues and cells in a controlled fashion. Using gentle mechanical procedures, called homogenization, the plasma membranes of cells can be ruptured so that the cell contents are released. Four commonly used procedures are shown here.

1. **Break cells with high-frequency sound**
2. **Use a mild detergent to break holes in the plasma membrane**
3. **Force cells through a small hole using high pressure**
4. **Shear cells between a close-fitting rotating plunger and the thick walls of a glass vessel**

**Differential Centrifugation**
Repeated centrifugation at progressively higher speeds will fractionate cell homogenates into their components.

- **Low-speed centrifugation**
  - Pellet 1: whole cells, nuclei, and chromosomes
- **Medium-speed centrifugation**
  - Supernatant: microsomal fractions,
  - Pellet 2: mitochondria, lysosomes, and peroxisomes
- **High-speed centrifugation**
  - Supernatant: other small vesicles
  - Pellet 3: rough and microsomal membranes
- **Very high-speed centrifugation**
  - Supernatant 4: ribosomes, monosomes, large macromolecules

**Velocity Sedimentation**
Subcellular components sediment at different speeds according to their size when carefully layered over a dilute salt solution. In order to stabilize the sedimenting components against convective mixing in the tube, the solution contains a continuous shallow gradient of sucrose that increases in concentration toward the bottom of the tube. This is typically 5-20% sucrose. When sedimented through such a dilute sucrose gradient, different cell components separate into distinct bands that can, after an appropriate time, be collected individually.

**Equilibrium Sedimentation**
The ultracentrifuge can also be used to separate cellular components on the basis of their buoyant density, independently of their size or shape. The sample is usually either layered on top of, or dispersed within, a steep density gradient that contains a very high concentration of sucrose or cesium chloride. Each subcellular component will move up or down when centrifuged until it reaches a position where its density matches its surroundings and then will move no further.

A series of distinct bands will eventually be produced, with those nearest the bottom of the tube containing the components of highest buoyant density. This method is also called density gradient centrifugation.

**The Centrifuge**
A centrifuge consists of a rotor, containing one or more buckets in which samples are placed, and a swinging arm which connects the buckets to the shaft of the rotor. The buckets hold the tubes that contain the sample and are free to swing outward as the rotor spins.

At such speeds, centrifuge chambers must be refrigerated and evacuated so that friction does not heat up the homogenate. The centrifuge is surrounded by thick armor plating, since an unbalanced rotor can shatter an explosive release of energy. A fixed-angle rotor can hold

**Equilibrium**
At equilibrium, components have migrated to a region in the gradient that matches their own density.
Labelling the Sample

Fluorescent Dye
Hybridization
Expression Measurement

If the fluorescent label is attached to one spot we know that the particular complementary gene transcript was present in our cell sample.

The greater the fluorescence the greater the concentration of the transcript.
Scanning the Microarray

Measuring the intensity of the fluorescence in each spot

We get a measurement of the hybridization in each spot of the microarray
The Data

For each biological sample (individual)

We get intensity measurements for thousands of genetic transcripts.

The measured intensity is used as an indicator of gene expression.

<table>
<thead>
<tr>
<th>Gene ID</th>
<th>Intensity</th>
</tr>
</thead>
<tbody>
<tr>
<td>200000_s_at</td>
<td>134.4</td>
</tr>
<tr>
<td>200001_at</td>
<td>586.5</td>
</tr>
<tr>
<td>200002_at</td>
<td>1868.4</td>
</tr>
<tr>
<td>200003_s_at</td>
<td>1232.7</td>
</tr>
<tr>
<td>200004_at</td>
<td>1071.6</td>
</tr>
<tr>
<td>200005_at</td>
<td>312.8</td>
</tr>
<tr>
<td>200006_at</td>
<td>1712.6</td>
</tr>
<tr>
<td>200007_at</td>
<td>606.5</td>
</tr>
<tr>
<td>200008_s_at</td>
<td>421.9</td>
</tr>
<tr>
<td>200009_at</td>
<td>395.6</td>
</tr>
<tr>
<td>200010_at</td>
<td>1228.6</td>
</tr>
<tr>
<td>200011_s_at</td>
<td>132.5</td>
</tr>
<tr>
<td>200012_x_at</td>
<td>2606.3</td>
</tr>
<tr>
<td>200013_at</td>
<td>1572.9</td>
</tr>
<tr>
<td>200014_s_at</td>
<td>138.7</td>
</tr>
<tr>
<td>200015_s_at</td>
<td>124.1</td>
</tr>
<tr>
<td>200016_x_at</td>
<td>1058.7</td>
</tr>
<tr>
<td>200017_at</td>
<td>889.4</td>
</tr>
<tr>
<td>200018_at</td>
<td>3964.2</td>
</tr>
<tr>
<td>200019_s_at</td>
<td>1069.9</td>
</tr>
<tr>
<td>200020_at</td>
<td>212.1</td>
</tr>
<tr>
<td>200021_at</td>
<td>1018.1</td>
</tr>
<tr>
<td>200022_at</td>
<td>1254.8</td>
</tr>
<tr>
<td>200023_s_at</td>
<td>1202.8</td>
</tr>
<tr>
<td>200024_at</td>
<td>2460.6</td>
</tr>
</tbody>
</table>
Several Microarrays

<p>| | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>31307_at</td>
<td>5.53</td>
<td>31307_at</td>
</tr>
<tr>
<td>31308_at</td>
<td>7.07</td>
<td>31308_at</td>
</tr>
<tr>
<td>31309_r_at</td>
<td>6.05</td>
<td>31309_r_at</td>
</tr>
<tr>
<td>31310_at</td>
<td>7.42</td>
<td>31310_at</td>
</tr>
<tr>
<td>31311_at</td>
<td>7.77</td>
<td>31311_at</td>
</tr>
<tr>
<td>31312_at</td>
<td>9.47</td>
<td>31312_at</td>
</tr>
<tr>
<td>31313_at</td>
<td>8.58</td>
<td>31313_at</td>
</tr>
<tr>
<td>31314_at</td>
<td>7.14</td>
<td>31314_at</td>
</tr>
<tr>
<td>31315_at</td>
<td>9.62</td>
<td>31315_at</td>
</tr>
<tr>
<td>31316_at</td>
<td>5.45</td>
<td>31316_at</td>
</tr>
<tr>
<td>31317_r_at</td>
<td>10.27</td>
<td>31317_r_at</td>
</tr>
<tr>
<td>31318_at</td>
<td>5.7</td>
<td>31318_at</td>
</tr>
<tr>
<td>31319_at</td>
<td>9.25</td>
<td>31319_at</td>
</tr>
<tr>
<td>31320_at</td>
<td>11.5</td>
<td>31320_at</td>
</tr>
<tr>
<td>31321_at</td>
<td>7.79</td>
<td>31321_at</td>
</tr>
<tr>
<td>31322_at</td>
<td>6.98</td>
<td>31322_at</td>
</tr>
<tr>
<td>31323_r_at</td>
<td>11.18</td>
<td>31323_r_at</td>
</tr>
<tr>
<td>31324_at</td>
<td>7.97</td>
<td>31324_at</td>
</tr>
<tr>
<td>31325_at</td>
<td>9.53</td>
<td>31325_at</td>
</tr>
<tr>
<td>31326_at</td>
<td>9.67</td>
<td>31326_at</td>
</tr>
<tr>
<td>31327_at</td>
<td>6.48</td>
<td>31327_at</td>
</tr>
<tr>
<td>31328_at</td>
<td>8.92</td>
<td>31328_at</td>
</tr>
<tr>
<td>31329_at</td>
<td>6.11</td>
<td>31329_at</td>
</tr>
<tr>
<td>31330_at</td>
<td>14.44</td>
<td>31330_at</td>
</tr>
<tr>
<td>31331_at</td>
<td>6.3</td>
<td>31331_at</td>
</tr>
<tr>
<td>31331_at</td>
<td>6.42</td>
<td></td>
</tr>
</tbody>
</table>
# Data Matrix

<table>
<thead>
<tr>
<th>Gene</th>
<th>Array 1</th>
<th>Array 2</th>
<th>Array 3</th>
<th>Array 4</th>
<th>Array 5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gene 1</td>
<td>5.53</td>
<td>5.66</td>
<td>5.52</td>
<td>5.69</td>
<td>5.62</td>
</tr>
<tr>
<td>Gene 2</td>
<td>7.07</td>
<td>7.14</td>
<td>7.05</td>
<td>7.19</td>
<td>7.05</td>
</tr>
<tr>
<td>Gene 3</td>
<td>6.05</td>
<td>5.33</td>
<td>5.35</td>
<td>5.07</td>
<td>5.29</td>
</tr>
<tr>
<td>Gene 4</td>
<td>7.42</td>
<td>7.02</td>
<td>7.02</td>
<td>7.04</td>
<td>7.22</td>
</tr>
<tr>
<td>Gene 5</td>
<td>7.77</td>
<td>7.83</td>
<td>7.79</td>
<td>7.75</td>
<td>7.77</td>
</tr>
<tr>
<td>Gene 7</td>
<td>8.58</td>
<td>8.67</td>
<td>8.52</td>
<td>8.42</td>
<td>8.52</td>
</tr>
<tr>
<td>Gene 8</td>
<td>7.14</td>
<td>7.3</td>
<td>7.19</td>
<td>7.27</td>
<td>7.32</td>
</tr>
<tr>
<td>Gene 10</td>
<td>5.45</td>
<td>5.53</td>
<td>5.3</td>
<td>5.35</td>
<td>5.44</td>
</tr>
<tr>
<td>Gene 11</td>
<td>10.27</td>
<td>10.75</td>
<td>10.41</td>
<td>10.45</td>
<td>10.3</td>
</tr>
<tr>
<td>Gene 12</td>
<td>5.7</td>
<td>5.53</td>
<td>5.59</td>
<td>5.58</td>
<td>5.67</td>
</tr>
<tr>
<td>Gene 14</td>
<td>11.5</td>
<td>11.51</td>
<td>11.35</td>
<td>11.36</td>
<td>11.25</td>
</tr>
<tr>
<td>Gene 15</td>
<td>7.79</td>
<td>7.91</td>
<td>7.76</td>
<td>7.82</td>
<td>7.74</td>
</tr>
<tr>
<td>Gene 16</td>
<td>6.98</td>
<td>6.93</td>
<td>6.91</td>
<td>7.04</td>
<td>6.8</td>
</tr>
</tbody>
</table>
How to treat the data?

<table>
<thead>
<tr>
<th>Gene</th>
<th>Array 1</th>
<th>Array 2</th>
<th>Array 3</th>
<th>Array 4</th>
<th>Array 5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gene 1</td>
<td>5.53</td>
<td>5.66</td>
<td>5.52</td>
<td>5.69</td>
<td>5.62</td>
</tr>
<tr>
<td>Gene 2</td>
<td>7.07</td>
<td>7.14</td>
<td>7.05</td>
<td>7.19</td>
<td>7.05</td>
</tr>
<tr>
<td>Gene 3</td>
<td>6.05</td>
<td>5.33</td>
<td>5.35</td>
<td>5.07</td>
<td>5.29</td>
</tr>
<tr>
<td>Gene 4</td>
<td>7.42</td>
<td>7.02</td>
<td>7.12</td>
<td>7.04</td>
<td>7.22</td>
</tr>
<tr>
<td>Gene 5</td>
<td>7.77</td>
<td>7.05</td>
<td>7.18</td>
<td>7.75</td>
<td>7.77</td>
</tr>
<tr>
<td>Gene 7</td>
<td>8.58</td>
<td>8.67</td>
<td>8.67</td>
<td>8.42</td>
<td>8.52</td>
</tr>
<tr>
<td>Gene 8</td>
<td>7.14</td>
<td>7.3</td>
<td>7.19</td>
<td>7.27</td>
<td>7.32</td>
</tr>
<tr>
<td>Gene 10</td>
<td>5.45</td>
<td>5.53</td>
<td>5.3</td>
<td>5.35</td>
<td>5.44</td>
</tr>
<tr>
<td>Gene 11</td>
<td>10.27</td>
<td>10.75</td>
<td>10.41</td>
<td>10.45</td>
<td>10.3</td>
</tr>
<tr>
<td>Gene 12</td>
<td>5.7</td>
<td>5.53</td>
<td>5.59</td>
<td>5.58</td>
<td>5.67</td>
</tr>
<tr>
<td>Gene 14</td>
<td>11.5</td>
<td>11.51</td>
<td>11.35</td>
<td>11.36</td>
<td>11.25</td>
</tr>
<tr>
<td>Gene 15</td>
<td>7.79</td>
<td>7.91</td>
<td>7.76</td>
<td>7.82</td>
<td>7.74</td>
</tr>
<tr>
<td>Gene 16</td>
<td>6.98</td>
<td>6.93</td>
<td>6.91</td>
<td>7.04</td>
<td>6.8</td>
</tr>
</tbody>
</table>
Babelomics Pipeline

Herrero et al., 2003, 2004; Vaquerizas et al., 2005 NAR; Montaner et al., 2006 NAR; Al-Shahrour et al., 2005, 2006 NAR; 2005 Bioinformatics
Definitions

Babelomics

- One of the most complete integrated packages of tools for *microarray data analysis* available over the web
- A complete suite of web tools for *functional analysis* of genome-scale experiments
Babelomics Web

Overview

Babelomics is an integrative platform for the analysis of transcriptomics, proteomics and genomic data with advanced functional profiling. The new version of Babelomics integrates primary (normalization, calls, etc.) and secondary (signatures, predictors, associations, TDT, clustering, etc.) analysis tools within an environment that allows relating genomic data and/or interpreting them by means of different functional enrichment or gene set methods. Such interpretation is made not only using functional definitions (GO, KEGG, Biocarta, etc.) but also regulatory information (prom Transfac, Jaccar, etc.) and other levels of regulation such as mRNA-mediated interference, protein-protein interactions, text-mining modules definitions, and the possibility of producing de novo annotations through the Blas2GO system.

Babelomics has been extensively re-engineered and now it includes the use of web services and Web 2.0 technology features, a new user interface with persistent sessions and a new extended database of gene identifiers. Babelomics is available at http://babelomics.bioinfo.cipf.es

In this release GEPIAS and Babelomics have integrated into a unique web application with many new features and improvements:

- **Data input**: import and quality control for the most common microarray formats
- **Normalization and baseline calling**: for the most common expression, tiling and SNP microarrays (Affymetrix and Agilent)
- **Transcriptomics**: diverse analysis options that include whole well established as well as novel algorithms for normalization, gene selection, class prediction, clustering and time series analysis
- **Genotyping**: stratification analysis, association, TDT
- **Functional profiling**: functional enrichment and gene set enrichment analysis with functional terms (GO, KEGG, Biocarta, etc.), regulatory (Transfac, Jaccar, mTRPAs, etc.), text-mining derived bioactivities, protein-protein interaction analysis
- **Integrative analysis**: different variables can be related to each other (e.g. gene expression to genomic copy number) and the results subjected to functional analysis

Homepage: http://babelomics.bioinfo.cipf.es
Questions

- Is there any significant difference in gene expression between tumor and healthy cells?
- Can we detect group of genes with similar expression profiles?
- Can we classify new biological samples based on gene expression patterns?
- Is there any significant functional enrichment in my gene list?
- Are these genes involved in the same disease?
Modules

- Gene Expression Analysis
  - Normalization
  - Differential Expression
  - Clustering
  - Predictors
- Functional Analysis
  - FatiGO
  - Fatiscan
Modules / Analysis progress

Upload data → Processing data → Expression → Genomic → Functional analysis

Data → Normalization → Diff. Expression Clustering Predictors → FatiGO Fatiscan

Gene Expression ANALYSIS → Functional ANALYSIS
Gene Expression Analysis

Herrero et al., 2003, 2004; Vaquerriz et al., 2005 NAR; Montaro et al., 2006 NAR; Al-Shahrour et al., 2005, 2006 NAR; 2005 Bioinformatics
Normalization

Processing data

- Edit
- Modify your uploaded data.
- Normalize
  - Expression
    - One-channel
      - Affymetrix
    - Normalization
  - Agilent

Affy normalization

- Online examples (test the form with example data)

Select your data

browse server: no data selected.
Or go to Upload Data form: Upload [affymetrix:expression:microarray]

Analysis

- RMA
- Plier
- Present-absent calls

Job

Job name: sample name
Job description

Run
Differential expression

- Differential expression
  - Class comparison:
    - Study differential
  - Correlation
  - Survival
  - Study the relation:
    - Time/dosage series
  - The module finds:
  - Predictors
    - Class prediction
    - Builds prediction:
  - Clustering

Differential expression: class comparison

- Online examples (test the form with example data)

Select your data

- [browse server]
  - no data selected.
  - Or go to Upload Data form: Upload [datamatrix]

Select the class to analyse

- Class name: [No classes available]

Specify the class values to test:

Select test

One-class (for log ratios)

- Limma

Two-classes

- T-test
  - Limma
  - Fold-change

Multi-classes
Predictors
Clustering

Expression

- Differential expression
  - Class comparison
  - Correlation
- Survival
- Study the relation
  - Time/dosage series
  - The module finds

- Predictors
  - Class prediction
  - Builds prediction
- Clustering

Clustering

- Online examples (test the form with example data)

Select your data

- browse server: no data selected.
- Or go to Upload Data form: Upload [datamatrix]

Select type of clustering: samples and/or genes

- Clustering of samples
- Clustering of genes

Select method

- UPGMA
  - SOTA
  - K-means
  - Number of sample-clusters (k-value): 5
  - Number of gene-clusters (k-value): 15

Select distance

- Euclidean (normal)
- Euclidean (square)
Functional Analysis

Herrero et al., 2003, 2004; Vaquerizas et al., 2005 NAR; Montano et al., 2006 NAR; Al-Shahour et al., 2003, 2006 NAR; 2005 Bioinformatics
From data to functional interpretation

**Gene Expression Analysis**

- Differential expression
- Predictors
- Clustering

**Microarray Data**

- Preprocessing (Normalization, Scaling, ...)
- Tab matrix file

**Functional Analysis**

- Genes differentially expressed
- Predicting genes
- Genes with same expression patterns
Functional Analysis

Babelomics, a suite of web tools for: statistical test, multiple test corrections, blast, ...

List of genes or ids, ie: differentially expressed genes, ...

Integrated Biological DB of Functional Annotation
  (GO, KEGG, InterPro, Ensembl, SwisProt, Transcription Factors, MicroRNA, Cisred, Biocarta, Bioentities Literature)
FatiGO

Functional analysis

- Single enrichment analysis
  - FatiGO
    - Provides significant
  - Marmite
- Set enrichment analysis
  - Gene set analysis
    - Finds gene-sets with gene-set analysis
  - MarmiteScan
    - Implements gene-set analysis
  - GeSBAP
    - Gene set analysis (a with SNPs or CNVs)

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:
- Browse server: no data selected.
- Or go to Upload Data form: Upload [idlist]

List 2:
- Browse server: no data selected.
- Or go to Upload Data form: Upload [idlist]

Options

- Fisher exact test: Two tailed
- Remove duplicates? Never
FatiGO features

- Compares two lists of genes.
- Compares one list of genes against the rest of the genome.

- One statistical test (Fisher’s exact) for each Block of annotation.
  - Multiple testing context.
  - Filtering of annotation is convenient (the less tests the best correction).
FatiGO test

One Gene List (A)

- Biosynthesis 6/10
- Sporulation 2/10

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

The other list (B)

- Biosynthesis 2/10
- Sporulation 2/10

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

<table>
<thead>
<tr>
<th></th>
<th>A</th>
<th>B</th>
</tr>
</thead>
<tbody>
<tr>
<td>Biosynthesis</td>
<td>6</td>
<td>2</td>
</tr>
<tr>
<td>No biosynthesis</td>
<td>4</td>
<td>8</td>
</tr>
</tbody>
</table>

We do this for each GO, miRNA, Interpro, ... !!!
FatiScan

Functional analysis

- Single enrichment analysis
  - Fatigo
    - Provides significant
  - Marmite
    - Single enrichment

- Set enrichment analysis
  - Gene set analysis
    - Finds gene-sets with gene-set analysis
  - MarmiteScan
    - Implements gene-
  - GeSBAP
    - Gene set analysis (a with SNPs or CNVs)

Gene set analysis

Select your ranked list
- browse server: no data selected.
  Or go to Upload Data form: Upload [idlist:ranked]

Options
- Logistic model
- Fatiscan
  - Fisher exact test: Two tailed
  - Remove duplicates: Never

Databases
- Organism: Select an organism
  - GO biological process [options]
  - GO molecular function [options]
  - GO cellular component [options]
FatiScan

- Interpret a **ranked list of genes**.
- There is not need for choosing a cut-off. All information is included.

- One statistical test for each **Block** of annotation.
  - Multiple testing context (hundreds of annotation).
  - Filtering of annotation is convenient (the less tests the best correction).
FatiScan
Testing along an ordered list

- Index ranking genes according to some biological aspect under study.
- Database that stores gene class membership information.
- **FatiScan** searches over the whole ordered list, trying to find runs of functionally related genes.

**List of genes**

- **Annotation label A**
- **Annotation label B**
- **Annotation label C**

**Block of genes enriched in the annotation A**

**Annotation C** is homogeneously distributed along the list

**Block of genes enriched in the annotation B**
# FatiScan results

## Significant results:

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

**Enriched class**

Annotated genes par GO from each list
Babelomics

Running an Example...
Experiment

Experiment description:
Rheumatoid Arthritis and Osteoarthritis

- 15 Affymetrix (HG-U133A), 38,500 genes
- Human
- 3 classes: control, osteoarthritis and rheumatoid arthritis
Example

Rheumatoid Arthritis and Osteoarthritis

Analysis:

- Normalization
- Differential expression
- FatiScan
Normalization

- Select your data > browser server > select your data: affy_arthritis
- Analysis: RMA
- Job name: normalization_arthritis
Welcome to the new Babelomics 4, you can still use Babelomics 3 at: [http://babelomics3.bioinfo.cipf.es](http://babelomics3.bioinfo.cipf.es)

**Affy normalization**

- Online examples (test the form with example data)

**Select your data**

- browse server: affy_arthritis

Or go to Upload Data form: Upload [affymetrix:expression: microarray]

**Analysis**

- RMA
- Plier
- Present-absent calls

**Job**

- Job name: normalization_arthritis

**all fields are correct**
Welcome to the new Babelomics 4, you can still use Babelomics 3 at: http://babelomics3.bioinfo.cipf.es

OK

your job has been successfully sent to the server
Normalization Results

```plaintext
| NUMBER_FEATURES | 12625 |
| NUMBER_SAMPLES | 15 |

<table>
<thead>
<tr>
<th>VARIABLE</th>
<th>filename</th>
<th>STRING VALUES</th>
</tr>
</thead>
<tbody>
<tr>
<td>GSM34379.CEL, GSM34383.CEL, GSM34385.CEL, GSM34391.CEL, GSM34393.CEL, GSM34394.CEL, GSM34395.CEL, GSM34396.CEL, GSM34397.CEL, GSM34398.CEL</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>VARIABLE</th>
<th>class</th>
<th>CATEGORICAL</th>
<th>VALUES</th>
<th>STRING VALUES</th>
</tr>
</thead>
<tbody>
<tr>
<td>id</td>
<td>ND_1, ND_2, ND_3, ND_4, ND_5, OA_A, OA_B, OA_x, OA_y, OA_z, RA_A, RA_B, RA_x, RA_y, RA_z</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>disease</td>
<td>normal_donor, osteoarthritis, rheumatoid_arthritis</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>VARIABLE</th>
<th>arthritis</th>
<th>CATEGORICAL</th>
<th>VALUES</th>
<th>STRING VALUES</th>
</tr>
</thead>
<tbody>
<tr>
<td>arthritis</td>
<td>0, 0, 0, 1, 1, 1, 1, 1, 1, 1, 1, 1</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>VARIABLE</th>
<th>filename</th>
<th>STRING VALUES</th>
</tr>
</thead>
<tbody>
<tr>
<td>GSM3437</td>
<td></td>
<td></td>
</tr>
<tr>
<td>GSM3439</td>
<td></td>
<td></td>
</tr>
<tr>
<td>GSM3439</td>
<td></td>
<td></td>
</tr>
<tr>
<td>GSM3439</td>
<td></td>
<td></td>
</tr>
<tr>
<td>GSM3439</td>
<td></td>
<td></td>
</tr>
<tr>
<td>GSM3439</td>
<td></td>
<td></td>
</tr>
<tr>
<td>GSM3439</td>
<td></td>
<td></td>
</tr>
<tr>
<td>GSM3439</td>
<td></td>
<td></td>
</tr>
<tr>
<td>GSM3439</td>
<td></td>
<td></td>
</tr>
<tr>
<td>GSM3439</td>
<td></td>
<td></td>
</tr>
<tr>
<td>GSM3439</td>
<td></td>
<td></td>
</tr>
<tr>
<td>GSM3439</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
```
Normalization Results

Summary

RMA summary: rma.summary.txt
Send to Preprocessing tool...

Box-plot

RMA box-plot:
Differential expression (ND vs RA)

- Dataset file name: rma.summary.txt
- Class: CLASS [ND, RA]
- Test: t
- Multiple-test correction: fdr
- Adjusted p-value: 0.05
- Job name: Diffexpr_2class_arthritis
Differential expression: class comparison

Online examples (test the form with example data)

Select your data

browse server  rma.summary
Or go to Upload Data form: Upload [datamatrix]

Select the class to analyse

Class name:  CLASS

Specify the class values to test:

- ND
- none
- RA

Select test

One-class (for log ratios)

- Limma

Two-classes

- T-test
- Limma
- Fold-change
Diff Expression Results
Diff Expression Results
FatiScan

- Species: hsa
- Duplicates management: Never
- Fisher exact test: Over represented terms in list 2
- Organism: Homo sapiens
- Data bases: GO biological process
- Job name: fatiscan_bottomlist_arteritis
Gene set analysis

Online examples (test the form with example data)

Select your ranked list

- browse server: t_ranked_list.txt (from job Diffexpr_2class_arthritis)
- Or go to Upload Data form: Upload [id: list: ranked]

Options

- Logistic model
- Fatiscan

Fisher exact test: Over-represented terms in list 2
Remove duplicates: Never

Databases

- Organism: Human (homo sapiens)
- GO biological process
- GO molecular function
- GO cellular component
- GO Slim GOA
- Interpro
- KEGG pathways
- Reactome
- Biocarto

Job list

Active Jobs

- Diffexpr_2class_arthritis
  - 2010-06-21 02:47:45.9
- Normalization_arthritis
  - 2010-06-21 02:02:10.0

Data list
**FatiScan Results**

<table>
<thead>
<tr>
<th>Term</th>
<th>Term size</th>
<th>Term size (in genome)</th>
<th>Term annotation % per list</th>
<th>Annotated Ids</th>
<th>Odds ratio (log e)</th>
<th>pvalue</th>
<th>Adjusted pvalue</th>
</tr>
</thead>
<tbody>
<tr>
<td>cell activation (GO:0001775)</td>
<td>443</td>
<td>425</td>
<td>2.51%</td>
<td>list 1:</td>
<td>0.8580</td>
<td>1.610e-18</td>
<td>3.5e-15</td>
</tr>
<tr>
<td>regulation of immune response (GO:000778)</td>
<td>220</td>
<td>259</td>
<td>1.30%</td>
<td>list 1:</td>
<td>-1.3397</td>
<td>5.182e-18</td>
<td>5.602e-15</td>
</tr>
<tr>
<td>regulation of immune system process (GO:0002682)</td>
<td>440</td>
<td>436</td>
<td>2.69%</td>
<td>list 1:</td>
<td>-1.0327</td>
<td>4.704e-18</td>
<td>5.602e-15</td>
</tr>
<tr>
<td>positive regulation of immune system process (GO:0002604)</td>
<td>263</td>
<td>266</td>
<td>1.62%</td>
<td>list 1:</td>
<td>-1.2452</td>
<td>2.376e-17</td>
<td>1.284e-14</td>
</tr>
<tr>
<td>leukocyte activation (GO:0045321)</td>
<td>395</td>
<td>366</td>
<td>2.21%</td>
<td>list 1:</td>
<td>-0.8179</td>
<td>1.718e-17</td>
<td>1.823e-14</td>
</tr>
<tr>
<td>antigen processing and presentation (GO:001852)</td>
<td>68</td>
<td>237</td>
<td>0.28%</td>
<td>list 1:</td>
<td>2.1608</td>
<td>7.013e-17</td>
<td>3.02e-14</td>
</tr>
<tr>
<td>immune response-regulating cell surface receptor signaling pathway (GO:0002768)</td>
<td>59</td>
<td>54</td>
<td>0.23%</td>
<td>list 1:</td>
<td>-2.2080</td>
<td>1.167e-16</td>
<td>4.206e-14</td>
</tr>
<tr>
<td>immune response-activating cell surface receptor signaling pathway (GO:0002429)</td>
<td>56</td>
<td>51</td>
<td>2.19%</td>
<td>list 1:</td>
<td>-2.2680</td>
<td>1.034e-15</td>
<td>3.523e-13</td>
</tr>
<tr>
<td>positive regulation of immune response (GO:0050778)</td>
<td>149</td>
<td>180</td>
<td>0.84%</td>
<td>list 1:</td>
<td>-1.4836</td>
<td>2.058e-15</td>
<td>4.944e-13</td>
</tr>
<tr>
<td>lymphocyte activation (GO:0046649)</td>
<td>339</td>
<td>314</td>
<td>2.2%</td>
<td>list 1:</td>
<td>-1.06</td>
<td>2.56e-15</td>
<td>5.577e-13</td>
</tr>
</tbody>
</table>
Acknowledgements

Supervisor:
- Joaquín Dopazo

The Babelomics team:
- Fatima Al-Shahrour
- Eva Alloza
- José Carbonell
- Ana Conesa
- Pablo Escobar
- Francisco García
- Stefan Göetz
- Martina Marbà
- Ignacio Medina
- Pablo Mínguez
- David Montaner
- Luis Pulido
- Javier Santoyo
- Joaquín Tárraga

http://bioinfo.cipf.es