Painting omics data in biological pathways

Version 2.0

User Guide

Fernando Garcia-Alcalde and Ana Conesa

Contents

1 What is Paintomics?

Paintomics is a web tool for the integration and visualization of transcriptomics and metabolomics data.

Basically, **Paintomics** takes two data matrices containing gene expression and metabolite levels for the same set of samples, finds associated KEGG pathways and paints these on a two-color scheme indicating the over or under abundance of corresponding pathway elements.

Currently **Paintomics** supports integrated visualization for over one hundred species of different biological kingdoms and offers users the possibility to request any other organism present in the KEGG database. **Paintomics** database is totally updated on a monthly basis.

For any question on **Paintomics**, an e-mail can be sent to paintomics@cipf.es.

2 Data upload

There are 4 data files **Paintomics** will regularly accept:

1. A tab-delimited file with *transcriptomics data*, containing genes in rows and samples in columns (Table 1). This is the only required parameter of the application. The first row must contain sample names, preferabily as short compact labels, and the first column must indicate gene identifiers. As a general rule **Paintomics** accepts EntrezGene IDs as gene identifiers, although for some model species (Table 2) other IDs such as *ensembl* IDs, *refseq* IDs, *unigene* IDs, as well as some *affymetrix*, *agilent* and *illumina* chips are also supported. Gene expression data should be in the log scale, positive values indicating over-expression and negative values indicating under-expression.

name	**treatment_1**	**treatment_2**
APOBEC3B	-0.82	-0.1
AA085955	-0.28	-0.32
ATP11B	0.31	0.97
AK092846	0.72	1.43
DNAJA1	-1.68	-1.01
THC2450799	0.19	-0.34
EHMT2	0.15	0.45
RPL23	0.09	0.32
T12590	-0.18	0.59
A_24_P704878	0.13	0.58
RPS13	0.01	-0.13
AK021474	0.25	-0.28
HDDC3	-0.43	-1.53
—	—	—

Table 1: Transcriptomics data input format

Name
Arabidopsis thaliana
Bos taurus
Caenorhabditis elegans
Danio rerio
Drosophila melanogaster
Homo sapiens
Mus musculus
Rattus norvegicus
Saccharomyces cerevisiae

Table 2: Species with authomatic gene ID convertion

2. A list of *significant genes* (optional). This is a ONE column text file with the names of a selection of genes from the previous file. Note that the type of gene indentifier used for the transcriptomics data and the significant list must coincide.

3. A tab-delimited file with *metabolite measurements* (optional) for the same samples as in the transcriptomics dataset (Table 3). Metabolite names should be placed in rows and identical sample ID must label columns. Data should be also provided in log scale, negative values indicating reduced metabolite levels and positive values indicating increased metabolite levels.

name	**treatment_1**	**treatment_2**
isoleucine	0.14	-0.04
leucine	0.02	0.02
valine	0.02	0.02
phenylalanine	0.02	-0.01
alanine	-0.02	-0.02
lactate	0.33	0.05
acetate	0.01	-0.12
pyruvate	0.09	-0.06
citrate	-0.01	-0.06
—	—	—

Table 3: Metabolomics data input format

4. A list of *significant metabolites* (optional). Similarly to the gene data, this is a ONE column text file containing a subset of metabolites names from the metabolomics dataset.

5. Organism. Currently **Paintomics** supports 101 species (see Appendix A). If your species is not available at the drop-down menu, please contact with the user support service at paintomics@cipf.es.

Once data is uploaded and the species of your choice is selected the application can be lanuched by cliking on the `Run` buttom.

3 Metabolite Selection

Once omics data has been submitted **Paintomics** parses input files to match identifiers to the KEGG database. Metabolite assignments can be ambiguous in some cases and need to be specified by the user. For each matched metabolite you can find the default assigned KEGG compound names (Figure 1). By clicking on *Show more matches* next to each metabolite name a list with additional possible assignments is shown. You can manually indicate the precise compound the metabolite in your dataset refers to, or let **Paintomics** to choose by default. In the default option **Paintomics** searches input metabolite names in the KEGG database and selects the closest entry, including all possible isomers. For example, if 'Glucose' is present in the user data, **Paintomics** will select 'L-Glucose' and 'D-Glucose'.

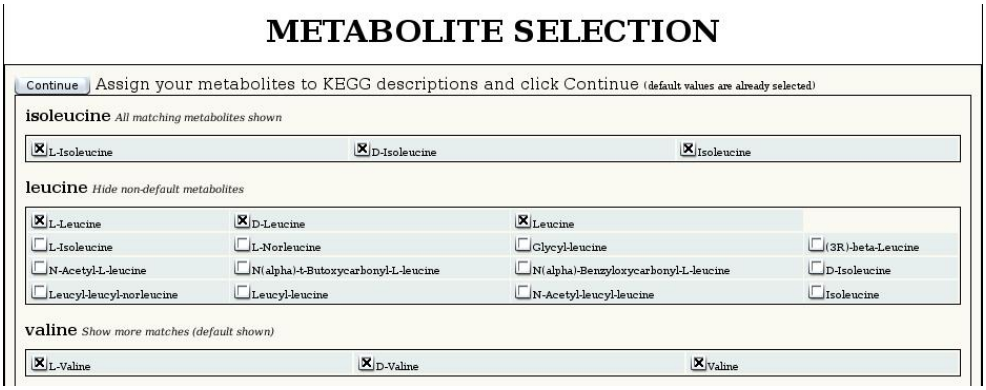

Figure 1: Metabolite selection menu

4 Data Summary and Pathway Selection

At this page you will find a summary of the intersection between your data and the KEGG database. The figures provided at the **Data Summary** section are:

1. Number of Pathways available for your organism at the KEGG database.
2. Number of Pathways found in your data.
3. Number of Genes of your organism in KEGG.
4. Total number of genes and number of significant genes present in your data.
5. Number of genes and number of significant genes of the input data found in KEGG.
6. Number of metabolites and significant metabolites present in your data.
7. Number of metabolites and significant metabolites of your input data found in KEGG.
8. Number of metabolites and significant metabolites selected in KEGG (this can be different from the previous when one-to-many mapping of your metabolite names to KEGG compounds were allowed).

At the **Pathway Selection** section you find matching data for each particular pathway:

1. Number of selected metabolites of the pathway contained in your input data
2. Number of selected metabolites of the pathway contained in your significant metabolites file.
3. Number of genes of the pathway found in in your input data.
4. Number of genes of the pathway found in in your significant genes file.
5. Total number of features (sum of 1. and 3.)
6. Total number of significant features (sum of 2. and 4.)
7. p-value of the Enrichment Analysis made by Fisher Exact's Test using the total number of features.

The **Pathway Selection** table is sortable by any of its columns and can also be exported as a text file. Select the pathawy(s) you want to paint by clicking on the corresponding check-box or alternatively select to paint *ALL pathways*. Click on `Paint` to submit painting job(s).

5 Results

Once painting is completed **Paintomics** images will be displayed at the **Results** page. For each pathway a separate image is generated of the corresponding KEGG map coloured according to the input data values. At each matched feature (significant or not), a box is drawn with as many sections as columns present in the input files, each section colored according to its correponding expression or concentration value at each particular sample. Coloring follows a gradient scheme from dark red (high positive values) to dark bue (high negative values). Additionally, significant features are highlighted by a black box line and gene names are shown for all proteins present in the indicated organism. In this way users can identify in one image species-specific reactions, all available experimental local data for the pathway and those changes that were called singificant. By clicking on any of the drawings, a full-size image is generated. Images are interactive, retaining KEGG link-outs to gene and compoud records. For browsers supporting SVG technology additional interactivity is available such as display of actual numerical values upon mouse pass-over (Figure 2).

All **Results** data can be downloaded, either as ONE wrapper zip file, or pathway by pathway. Each pathway directory contains painted images in three formats (png, pdf and svg) and an *info* file with pathway specific information. The info file lists the data values for all matched features together with the whole list of elements present in the pathway.

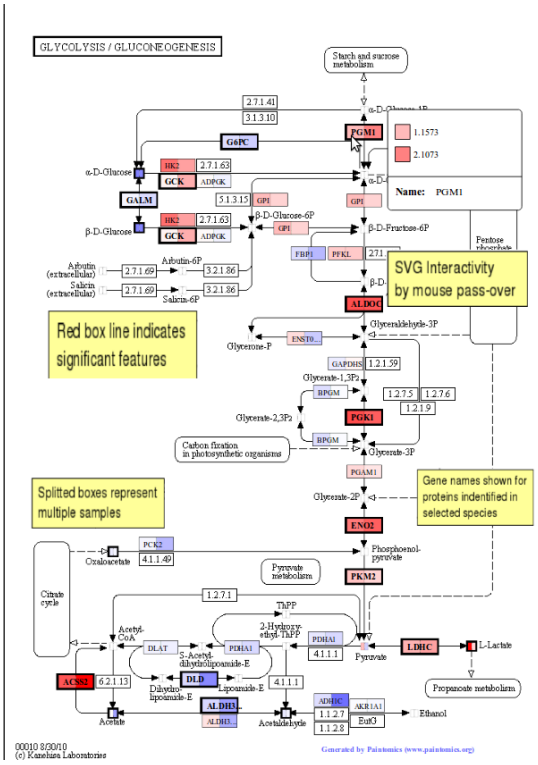

Figure 2: Interactive **Paintomics** images

A Appendix: Avalaible species

Table 4: Available species in PaintOmics

Name	
Actinobacillus pleuropneumoniae JL03 (serotype 3)	Mycobacterium tuberculosis KZN 1435
Aggregatibacter aphrophilus	Mycoplasma pneumoniae
Agrobacterium tumefaciens C58	Neisseria meningitidis alpha14
Anopheles gambiae (mosquito)	Neurospora crassa
Apis mellifera (honey bee)	Oryza sativa japonica (Japanese rice)
Arabidopsis thaliana (thale cress)	Penicillium chrysogenum
Aspergillus fumigatus	Phanerochaete chrysosporium
Aspergillus nidulans	Photorhabdus asymbiotica
Aspergillus niger	Pichia pastoris
Aspergillus oryzae	Plasmodium falciparum 3D7
Bacillus subtilis	Populus trichocarpa (black cottonwood)
Bifidobacterium dentium	Proteus mirabilis
Bifidobacterium longum subsp. longum JDM301	Pseudomonas aeruginosa PA7
Bos taurus (cow)	Pseudomonas aeruginosa PAO1
Brucella abortus S19	Pseudomonas fluorescens SBW25
Buchnera aphidicola 5A	Pseudomonas putida W619
Burkholderia mallei NCTC 10229	Pseudomonas syringae pv. phaseolicola 1448A
Caenorhabditis elegans (nematode)	Ralstonia pickettii 12D
Candida albicans	Rattus norvegicus (rat)
Chlamydia trachomatis D/UW-3/C1 (serovar D)	Rhizobium leguminosarum bv. trifolii WSM1325
Chlamydomonas reinhardtii	Rhodobacter sphaeroides ATCC 17029
Chlamydophila pneumoniae TW183	Rhodospirillum rubrum
Citrobacter rodentium	Ricinus communis (castor bean)
Clostridium acetobutylicum	Saccharomyces cerevisiae (budding yeast)
Corynebacterium glutamicum R	Salmonella enterica subsp. enterica serovar Paratyphi C
Cronobacter turicensis	Schistosoma mansoni
Cryptococcus neoformans B-3501A	Schizosaccharomyces pombe (fission yeast)
Cryptococcus neoformans JEC21	Sclerotinia sclerotiorum
Danio rerio (zebrafish)	Serratia proteamaculans
Desulfovibrio desulfuricans ATCC 27774	Shewanella violacea
Dictyostelium discoideum (cellular slime mold)	Shigella boydii CDC 3083-94
Drosophila melanogaster (fruit fly)	Sodalis glossinidius
Edwardsiella ictaluri	Sorghum bicolor (sorghum)
Enterobacter cloacae	Stenotrophomonas maltophilia R551-3
Erwinia carotovora	Streptococcus pneumoniae TCH8431/19A
Gallus gallus (chicken)	Streptomyces coelicolor
Haemophilus influenzae PittGG	Synechococcus sp. PCC7002
Helicobacter pylori B8	Taeniopygia guttata (zebra finch)
Homo sapiens (human)	Trichomonas vaginalis
Hydra magnipapillata	Trypanosoma brucei
Klebsiella pneumoniae NTUH-K2044	Tuber melanosporum
Kluyveromyces lactis	Ustilago maydis
Lactobacillus sakei	Vibrio cholerae M66-2
Lactococcus lactis subsp. lactis KF147	Vitis vinifera (wine grape)
Legionella pneumophila 2300/99 Alcoy	Wigglesworthia glossinidia
Leishmania major	Xanthomonas campestris pv. campestris B100
Leptospira biflexa serovar Patoc Patoc 1 (Paris)	Xenopus tropicalis (western clawed frog)
Listeria monocytogenes Clip81459	Xenorhabdus nematophila
Lotus japonicus (lotus)	Yersinia pestis Z176003
Moraxella catarrhalis	Zea mays (maize)
Mus musculus (mouse)	