omics.data.edu.au

Pathway Tools "How-to" Guides

How to use the Cellular Overview map with Pathway Tools

Cellular Overview is a feature in Pathway Tools to allow users to choose the pathway of interest by interacting with the cellular overview diagram (Figure 1) after overlaying the c data. This way the user does not need to have the prior knowledge of a pathway of intere

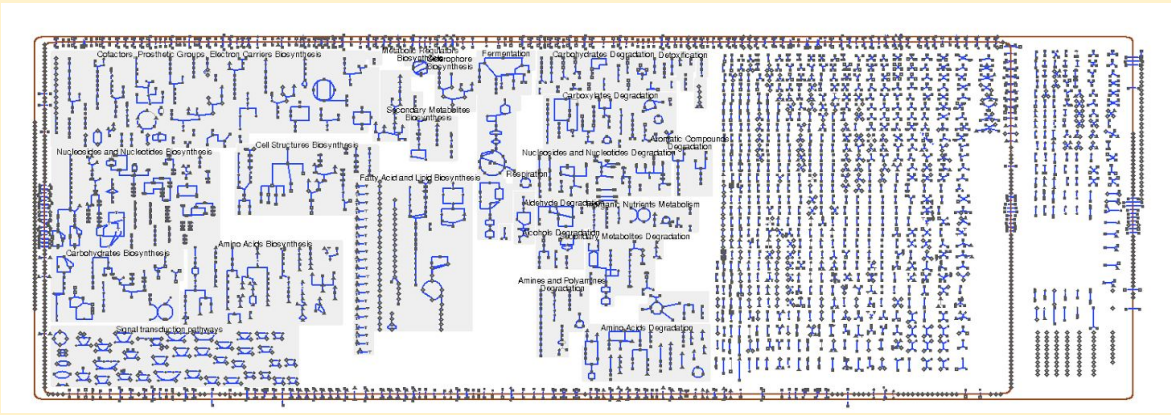


Figure 1. Cellular Overview Diagram

This tutorial is for Pathway Tools version 19.5.

1 Introduction

Pathway Tools is a software suite from BioCyc database collection. Amongst its many functionalities (see http://bioinformatics.ai.sri.com/ptools/) , it supports the visual analysis of gene-expression and metabolomics datasets, such as overlaying omics data (eg transcriptomic expression values) onto diagrams of an organism's metabolic network, as shown in Figure 2.

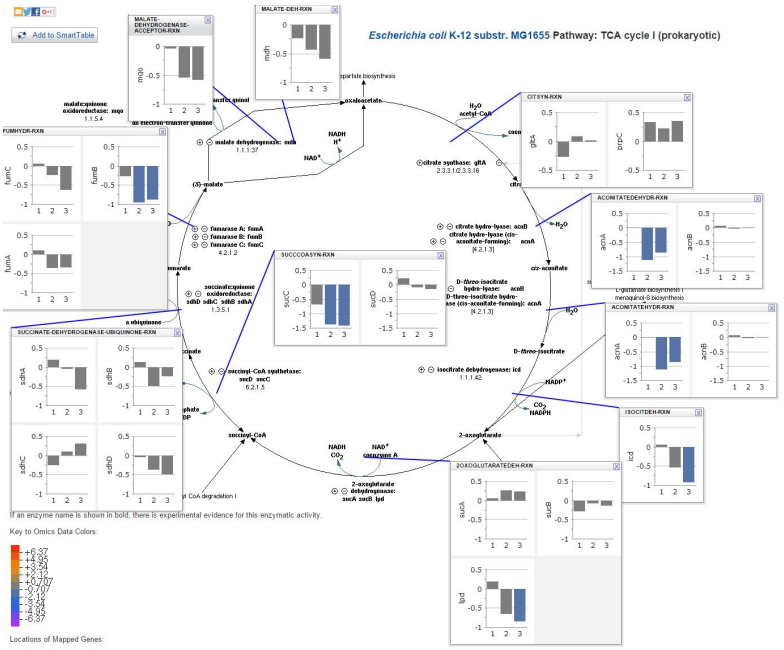


Figure 2 Overlay of experimental transcriptomic data on Metabolic Pathways. A close up of the TCA cycle in **E.coli K-12,** *with expression levels of constituent proteins shown in bar graphs across three experiments.*

By default, Pathway Tools contains four organism databases: (1) Homo sapiens, (2) Mus musculus, (3) *E.coli K-12* **and (4) MetaCyc.**

Pathway Tools also allow users to create their own organism-specific Pathway/Genome Database (PGDB) using the Pathologic tool. Creation of a customised database is covered in the online tutorial module "Pathway prediction and Annotations for new organisms".

2 Aim

This how-to guide will provide step-by-step instructions on how to use the overlay function of Pathway Tools using the web-browser.

This guide uses the metabolic pathways from the model organism *Escherichia coli K-12* **and the data from the published study Varas** *et al*, **2017,** *Multi-level evaluation of Escherichia coli polyphosphate related mutants using global transcriptomic, proteomic and phenomic analyses*, **Biochimica et Biophysica Acta 1861:871-883. In this study, the expression of all genes and proteins of** *E.coli K-12* **was measured using microarray and LC-ESI-MS/MS methods. The expressed were measured across three mutant strains (** *ppk1, ppx, and polyP),* **which lack enzymes related to inorganic polyphosphate metabolism.**

3 Data

The transcriptomic data files are sourced from the Genome Expression Omnibus (GEO) database on NCBI. The following samples from series GSE29954 were downloaded, these contained the gene expression levels relative to wild-type *E.coli* **K-12 for:**

- *ppk1* **(GSM741272)**
- *ppx* **(GSM741273)**
- *polyP* **(GSM741274)**

and the metadata file GSE29954_raw_data_IDs_ALL_columns.txt.gz with information about the gene IDs and symbols used in the study. The gene symbols are used in the pathway overlay feature to map the genes (and proteins) in the expression data files to the metabolic pathway.

Pathway Tools simply takes in any data with 2 or more columns, where the first column contains a gene symbol and subsequent columns are the expression level values for different experimental conditions. As such, while the example data used in this guide was derived from microarray experiments, the following step-by-step guide to overlay the gene expression onto the metabolic pathways can also be applied to NGS-derived data.

We have already performed the data transformation required to match the format required by pathway tools. The input file for this guide is "all_mutants_transcriptomic_modified.txt". Below in Box 3.1 describes the steps used for transforming and concatenating the GSM datasets into one file.

3.1 Data transformation

As noted above, Pathway Tools takes in data with 2 or more columns, where the first colu contains a gene symbol and subsequent columns are the expression values. Here we desc how the data transformation was done for this guide.

Each of the raw files (GSM741272, GSM741273 and GSM741274) have two columns: ID_REF and log2 value. The ID_REF values are project specific gene identifiers assigned by Varas *et* (e.g. 4.3.1.17) and need to be converted to a gene symbol since gene symbols are used witl Pathway Tools to denote genes/proteins. In this case, the gene annotations were supplied the accompanying GSE29954_raw_data_IDs_ALL_columns.txt file, which was used to map t ID_REF values to their corresponding gene symbol.

Following this conversion, the three modified 'raw' files (each with two columns: Gene Syr and gene expression level, were then combined into one master file "all_mutants_transcriptomic_modified.txt" with the following four columns:

1. **gene symbol**
2. **log2(wt_ *ppk1*) – relative gene expression level of wild type vs ppk1 mutant**
3. **log2(wt_ *ppx*) - relative gene expression level of wild type vs ppx mutant**
4. **log2(wt_ *polyP*) – relative gene expression level of wild type vs (ppk1 + ppx) mutar**

This file "all_mutants_transcriptomic_modified.txt" is used as input for this guide.

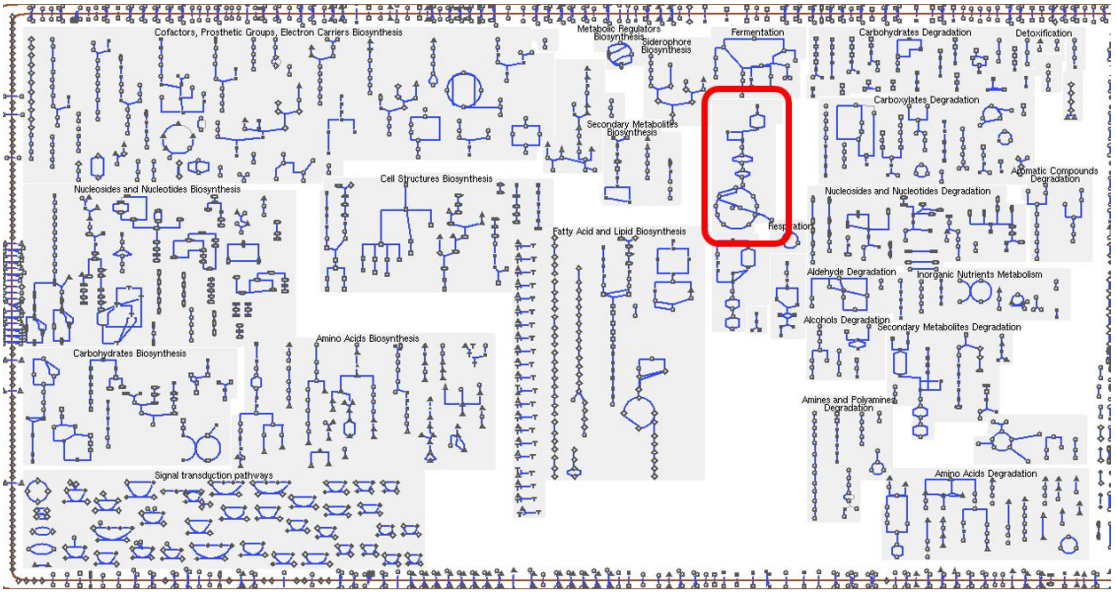
4. Step-by-step guide

In this example, the superpathway of glycolysis, pyruvate dehydrogenase, TCA, and glycolylate bypass (Figure 3) is used for demonstration.

Step	Instruction	Note(s)
1	Go to http://abrpi.genome.edu.au:9655/	Note: this is a private hosted instance of Pathway Tools that is part of the larger OMICs platform (http://omics.data.edu.au/use/).

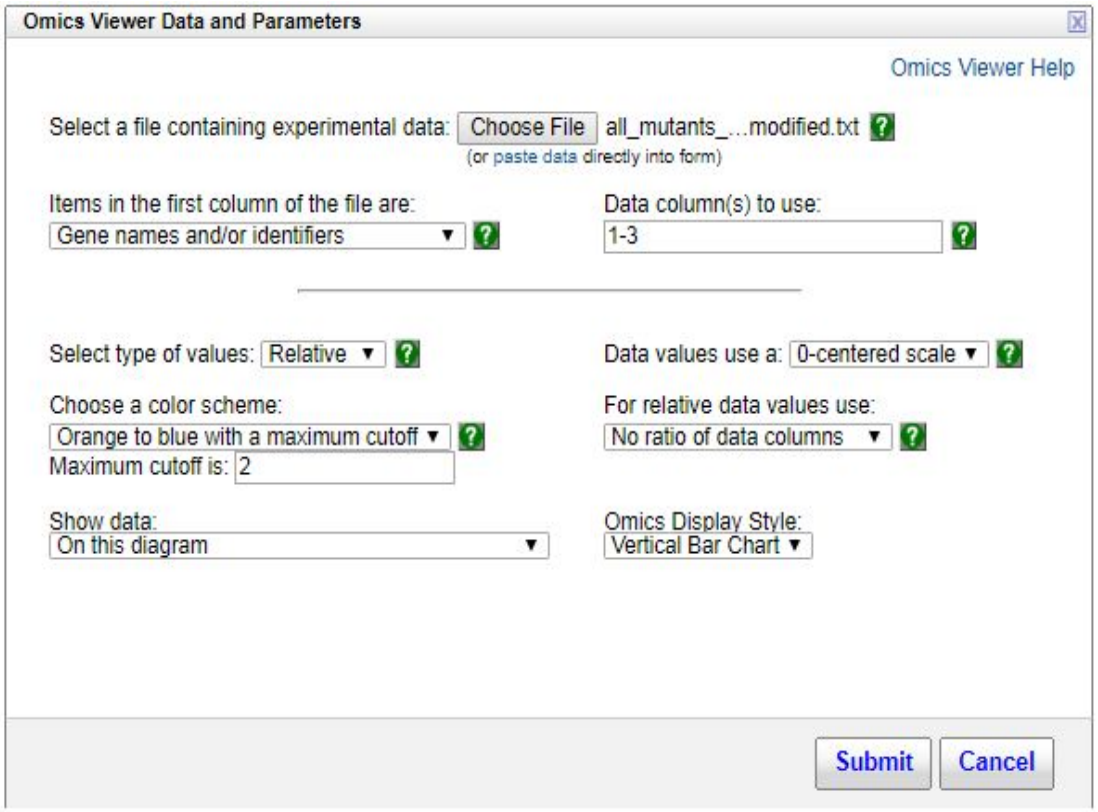
2	Select “Escherichia coli K-12 substr. MG1655 (EcoCyc)” as the organism database by clicking the “change organism database” under the search column at the top right corner of the webpage.	The E.coli K12 MG1655 database contains the gene identifier and metabolic pathway information for E.coli Strain K1 This has been prepared previously by [EcoCyc Staff]().
3	Click on “Metabolism” on the top menu on the webpage	Navigate to “superpathway of glycolysis, pyruvate dehydrogenase, TCA, and glycoxylate bypass”
4	Click on “ Cellular Overview” from the drop down menu	
5	Zoom in the “superpathway of glycolysis, pyruvate dehydrogenase, TCA, and glycoxylate bypass”. Use the zooming tool on the left hand side of the cellular diagram zoom into the pathway.	

The superpathway of glycolysis, pyruvate dehydrogenase, TCA, and glycoxylate bypass is red box



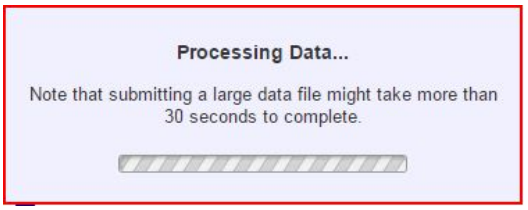
7	Click on the “Upload Data from File” on the menu at the right hand side on the webpage.	

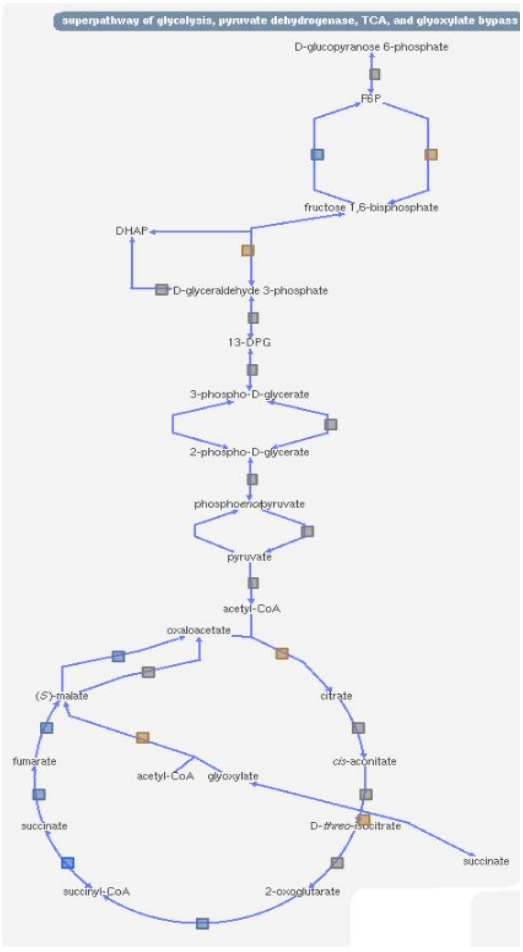
A new window will appear:



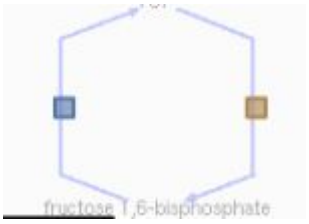
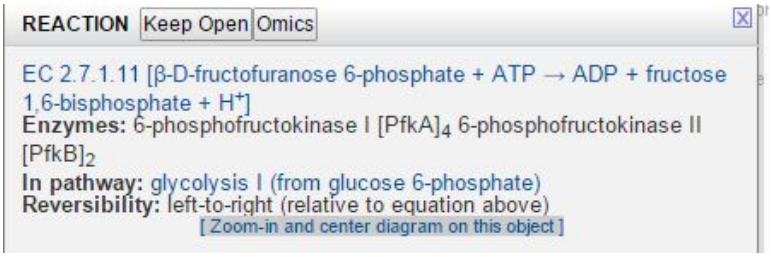
The following are mandatory settings:

8	**Select the "Choose File" in the figure to upload the transcriptomic data**	**In this example, use the all_mutants_transcriptomic_modified.txt file. Navigate to a local copy and upload it.**
10	**Items in the First Column of the File are: Gene names and/or Identifiers**	**The first column (column "0") of this file contain gene symbol whereas columns 1-3 contain gene expression levels of the 3 mutants relative to wildtype E.coli K12**
11	**Data Column(s) to Use: 1-3**	
12	**Select type of values: Relative**	**The numerical values in the data file might be considered absolute or relative** **ŏ absolute - all negative values in your data file will be skipped** **ŏ relative - allows you to specify ratios of columns whereas absolute does not.**
13	**Data Values Use a: 0-centered scale**	**Implies that the numerical data of your file can contain positive and negative values. The value 0 is considered to be the centre of the numerical values provided in your data file.**
14	**Choose a color scheme: Orange to blue with a maximum cutoff**	

15	**Maximum cutoff is: 2**	
16	**Show data: On this diagram**	
17	**Omics Display Style : Vertical Bar Chart**	
18	**Click on Submit**	**The overlaying data on a cellular diagran might take up to 30 seconds to complete If so, you will see a Processing Data... window** 



To get the expression values to show as bar plots, following the next 2 steps:

17	**Click on the little colour box (e.g. light blue or light brown in our case) on the pathway.**	**The fading blue colour represents negative expression value (log scale) and the yellow colours represents positive expression values (log scale). Indeed, it is difficult to tell the differences.** 
18	**Click on "Omics" on the pop up menu to show the bar plots.**	**The bar plots show the expression value of three different experimental conditions.**
		
When the overlay function finishes, the expression values are shown in barplots over the pathway as demonstrated in Figure 3.		

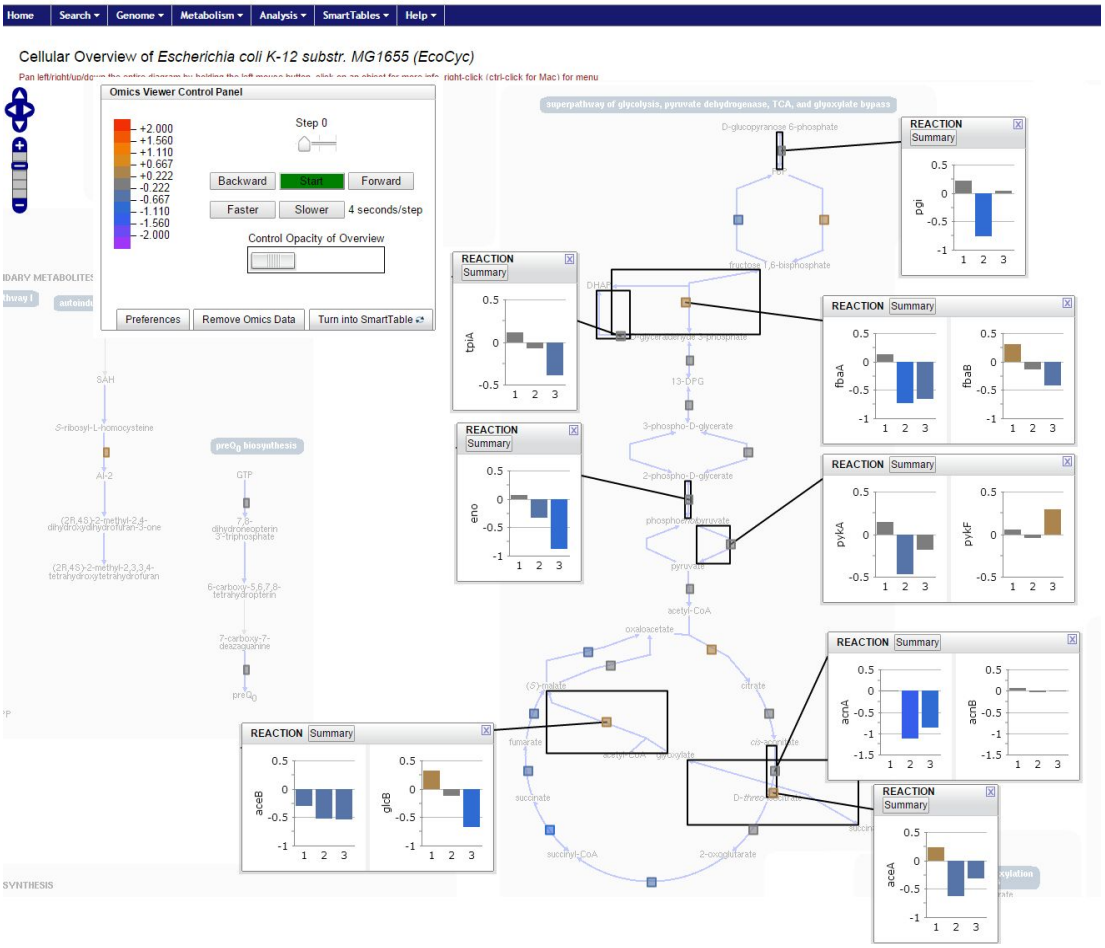


Figure 3. Transcriptomic data overlayed on Superpathway of glycolysis, pyruvate dehydrogenase, TCA, and glycolylate bypass.

If you need further assistance please contact us at omicsdataservices@lists.unimelb.edu.a