Pathway Analysis

Niels W. Hanson
Ph.D. Candidate Bioinformatics
Tuesday, February 11 2014

Hallam Laboratory
Hydrocarbon MetaPathways Workshop
The University of British Columbia, Vancouver

https://github.com/nielshanson/mp_tutorial/wiki/Pathway-Analysis

Prerequisites

- Install Pathway Tools
  http://biocyc.org/download-bundle.shtml
- Perl v5.0 http://www.perl.org/

Downloads

- Presentation Slides: MetaPathways_Tutorial_Pathway_Analysis.pdf
- HOT Fosmid-end ePGDBs: HOT_Sanger_ePGDBs.zip
- Perl Pathway Extractor Script: extract_pathway_table_from_pgdb.pl
- ORF Abundance Tables:
  1_upper_euphotic_rxn.wide.txt, HOT_Sanger_rxn.wide.txt

Goals of Tutorial

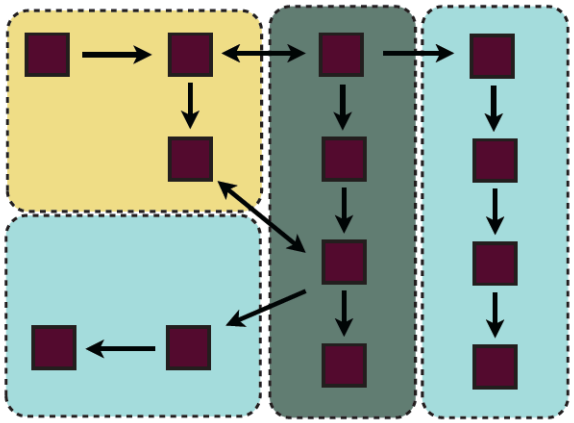
1. Give a brief overview of the Pathway Tools and the Pathologic Algorithm
2. Load processed Environmental Pathway Genome Databases (ePGDBs) into Pathway Tools
3. Explore predicted Pathways in Pathway Tools in the Cellular Overview, Pathway, and Reaction pages.
4. Highlight predicted pathways on the cellular overview and compare samples
5. Extract pathway and associated ORFs into long and wide formats.
6. Understand how the 'Omics Tools' feature to overlay quantitative metadata about pathways, e.g., ORF abundance.

1. Pathway Tools & PathoLogic

Pathway Tools

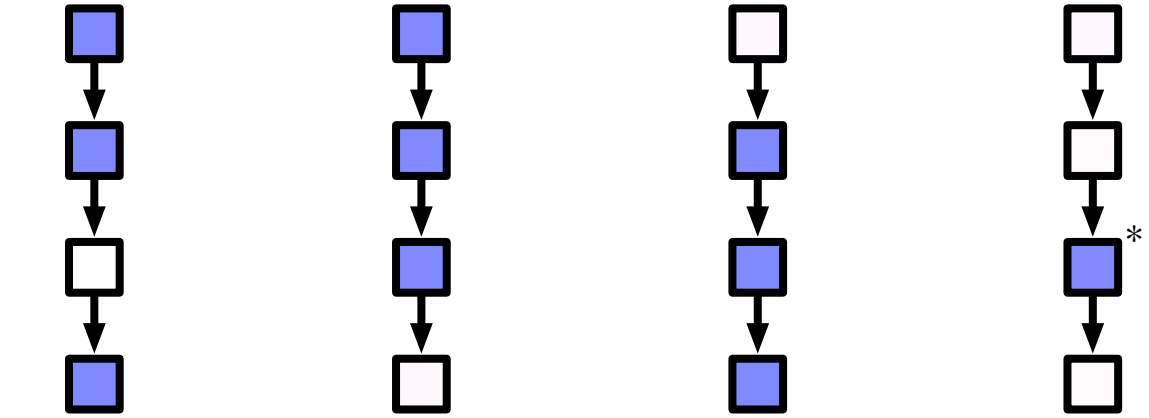
Pathway Tools

- Genes operate within structure of metabolism
- Pathway tools a software framework to integrate genomic annotations with MetaCyc pathways
- Data structure of Genes + Pathways: Pathway/Genome Database (PGDB)



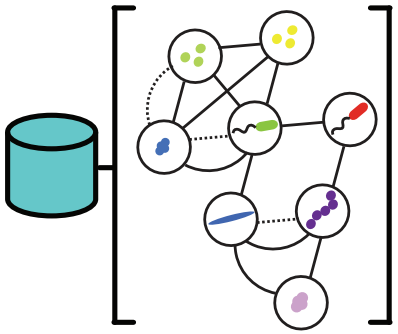
Pathologic

- rule-based model on on pathways class and completeness

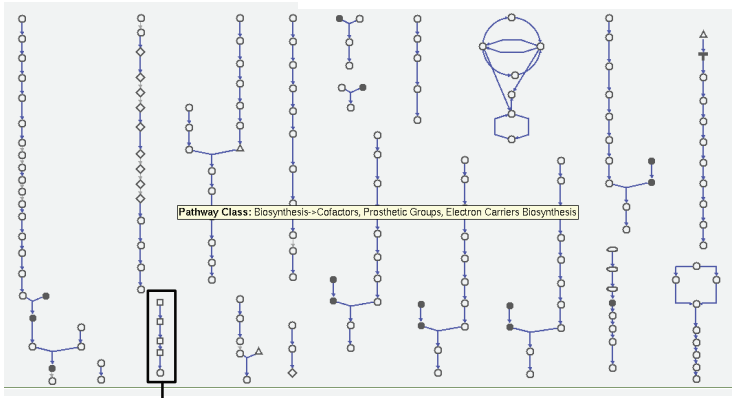


1. Pathway Tools & PathoLogic

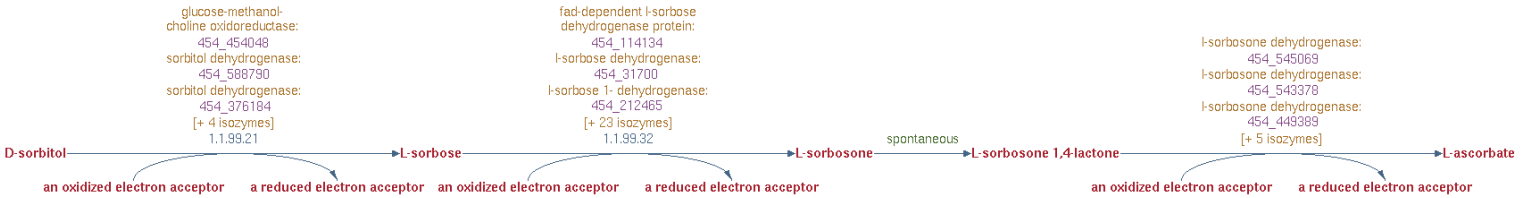
Pathway Tools



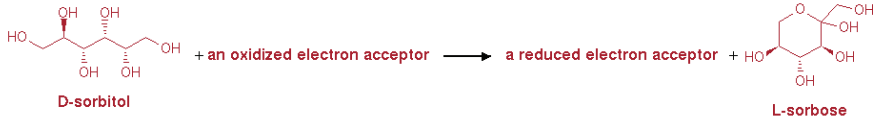
ePGDB



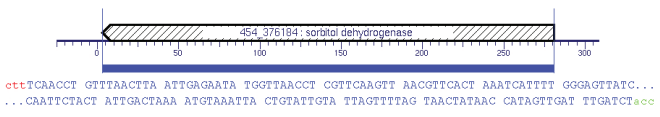
Cellular Overview



Metabolic Pathway



Reaction



Open Reading Frame

2. Loading ePGDBs into Pathway Tools

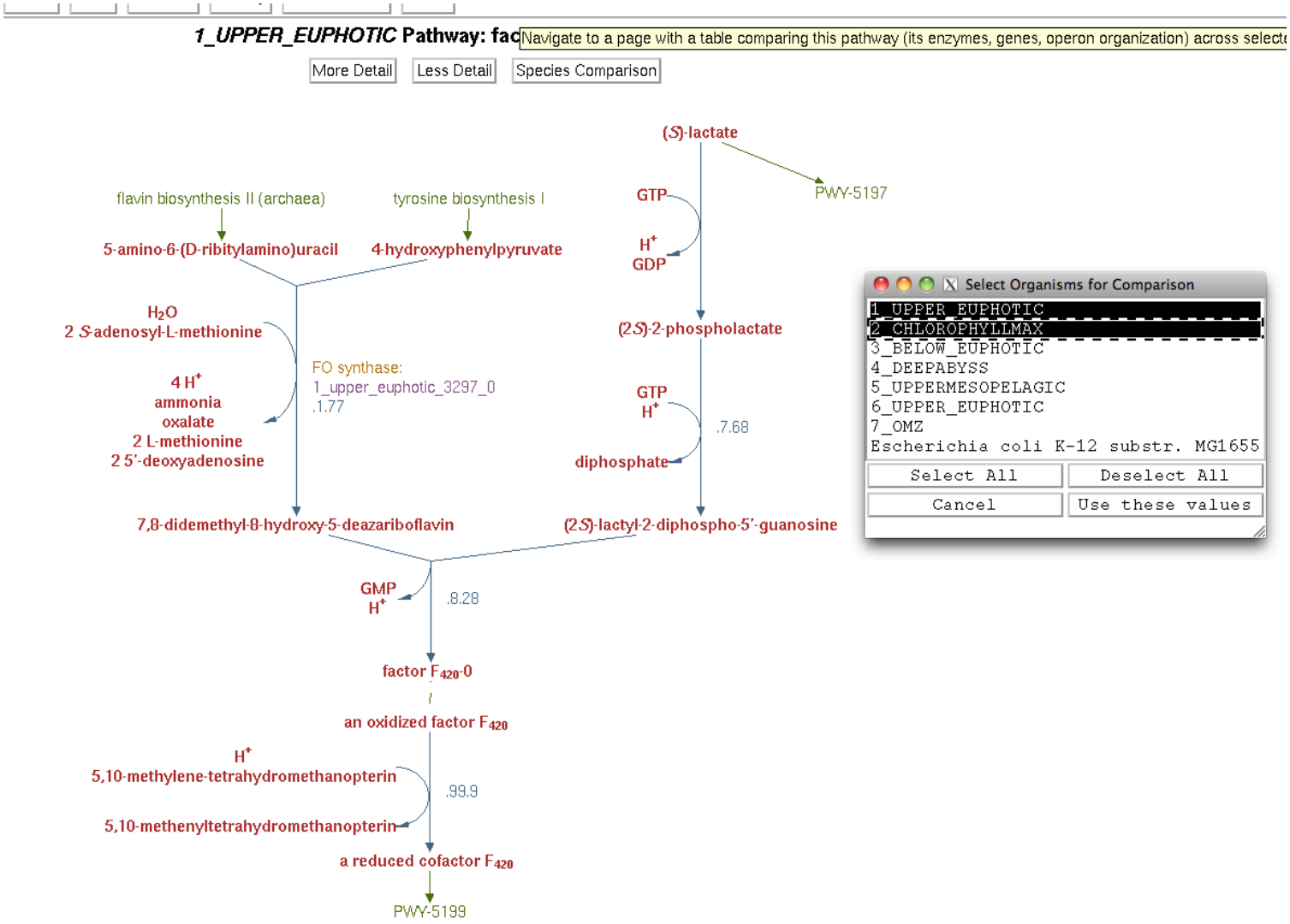
Two Operations:

1. Place completed ePGDB `<sample>cyc/` folder into `ptools-local/pgdbs/user/`
2. Update organism.dat in each ePGDB `<sample>cyc/1.0/input/organism.dat`

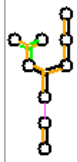
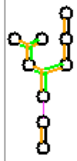
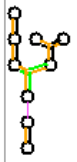
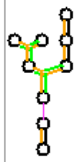
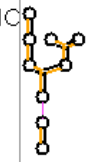
```
28
29  ID  1_UPPER_EUPHOTIC
30  STORAGE  FILE
31  NAME  unclassified sequences
32  ABBREV-NAME  u. sequences
33
```

```
28
29  ID  1_UPPER_EUPHOTIC
30  STORAGE  FILE
31  NAME  1_UPPER_EUPHOTIC
32  ABBREV-NAME  1_UPPER_EUPHOTIC
33
```

3. Cellular Overview, Pathway, and Reaction Pages



3. Cellular Overview, Pathway, and Reaction Pages

Organism	Evidence Glyph	Enzymes and Genes for factor 420 biosynthesis
1_UPPER_EUPHOTIC		RXN-8076: None EC .1.77: FO synthase: 1_upper_euphotic_3297_0 EC .7.68: None EC .8.28: None EC .99.9: None
2_CHLOROPHYLLMAX		RXN-8076: None EC .1.77: FO synthase: 2_chlorophyllmax_3549_0 EC .7.68: None EC .8.28: LPPG:FO 2-phospho-L-lactate transferase: 2_chlorophyllmax_594_0 2-phospho-L-lactate transferase: 2_chlorophyllmax_5766_1 EC .99.9: None
3_BELOW_EUPHOTIC		RXN-8076: None EC .1.77: None EC .7.68: None EC .8.28: 2-phospho-L-lactate transferase: 3_below_euphotic_1432_1 EC .99.9: None
4_DEEPABYSS		RXN-8076: None EC .1.77: FO synthase subunit 1: 4_deepabyss_3687_1 FO synthase subunit 1: 4_deepabyss_4601_1 EC .7.68: None EC .8.28: 2-phospho-L-lactate transferase: 4_deepabyss_4262_1 2-phospho-L-lactate transferase: 4_deepabyss_3652_0 LPPG:FO 2-phospho-L-lactate transferase: 4_deepabyss_3332_1 2-phospho-L-lactate transferase: 4_deepabyss_10739_0 EC .99.9: None
5_UPPERMESOPELAGIC		**This pathway is not marked as present in this organism.** No enzymes or genes have been identified for this pathway

3. Cellular Overview, Pathway, and Reaction Pages

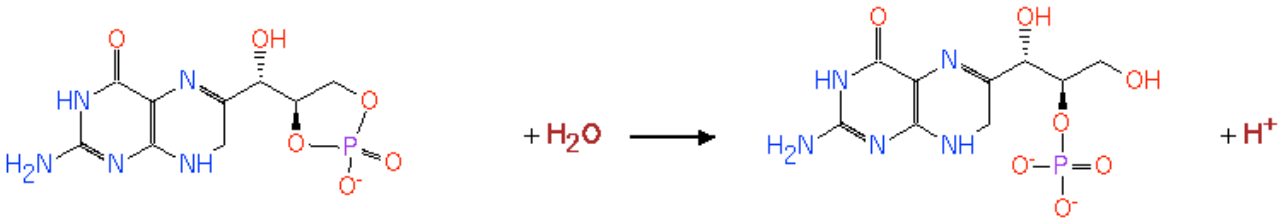
***1_UPPER_EUPHOTIC* Reaction: [no EC number assigned]**

Species Comparison

Superclasses: **Reactions-Classified-By-Conversion-Type** -> **Simple-Reactions** -> **Chemical-Reactions**
Reactions-Classified-By-Substrate -> **Small-Molecule-Reactions**

In Pathway: **6-hydroxymethyl-dihydropterin diphosphate biosynthesis II (archaea)**

Atom Mapping: None found for this reaction.



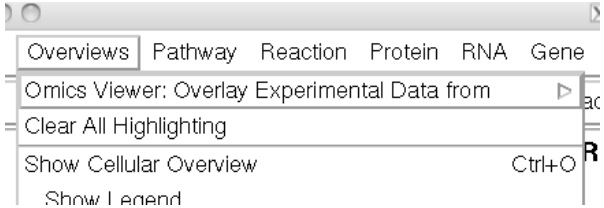
7,8-dihydroneopterin 2',3'-cyclic phosphate

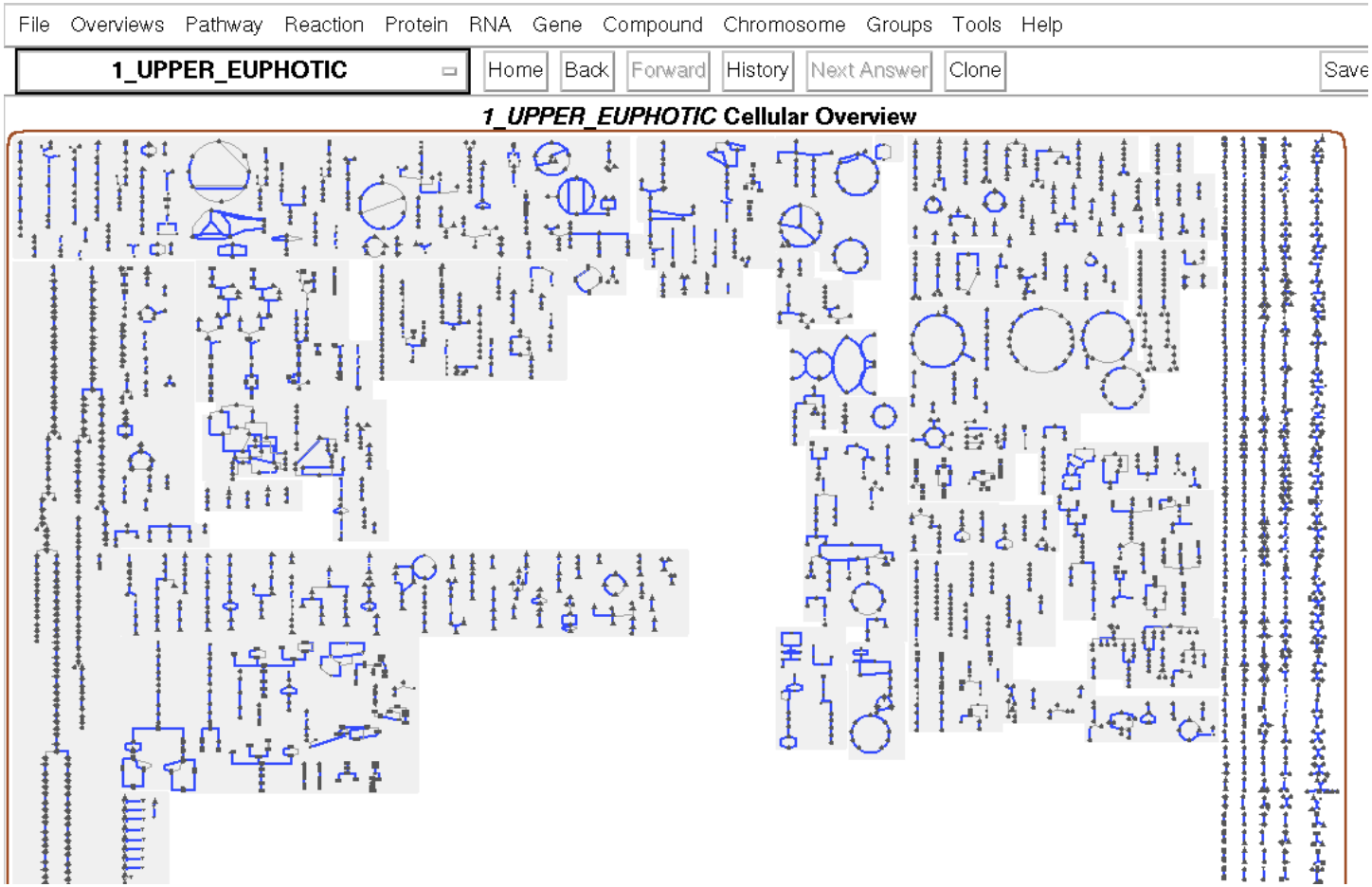
7,8-dihydroneopterin 2'-phosphate

- The reaction direction shown, that is, A + B <--> C + D versus C + D <--> A + B, is in accordance with the direction in which it was curated.
- Most BioCyc compounds have been protonated to a reference pH value of 7.3, and some reactions have been computationally balanced for hydrogen by adding free protons. Please see the PGDB Concepts Guide for more information.
- Mass balance status: Balanced.

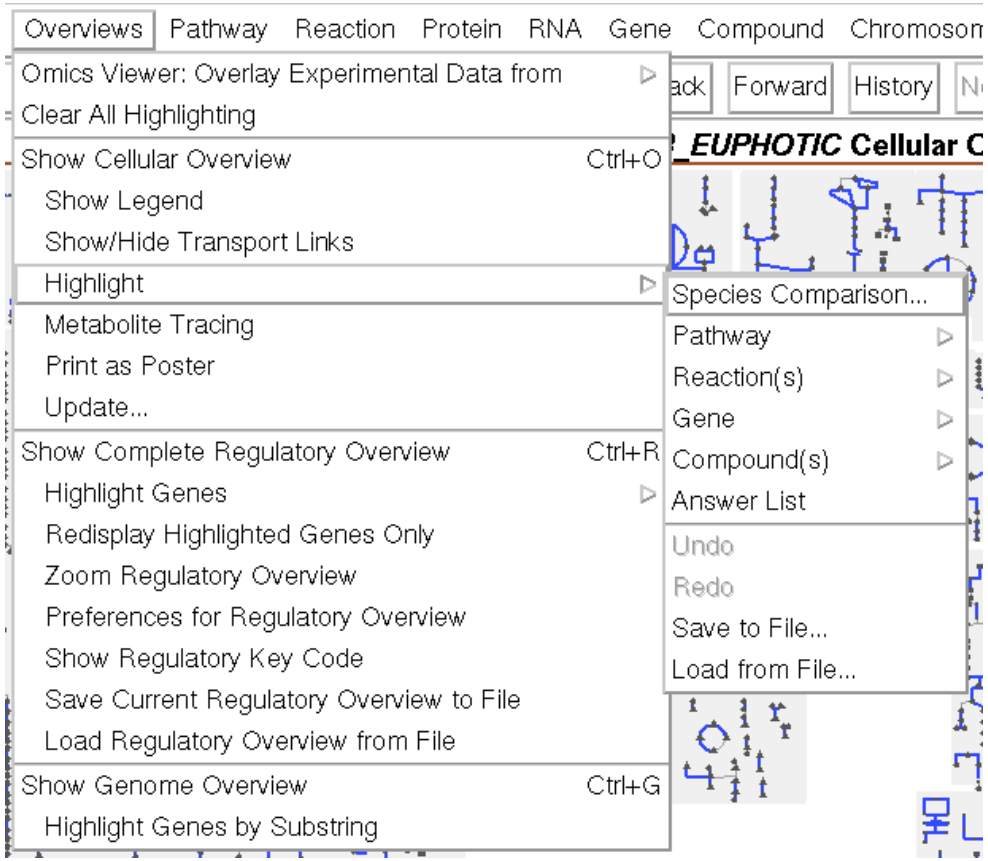
Credits: Created in MetaCyc 26-Apr-2011 by **Caspi R, SRI International**
Imported **from MetaCyc** 13-Jan-2014 by

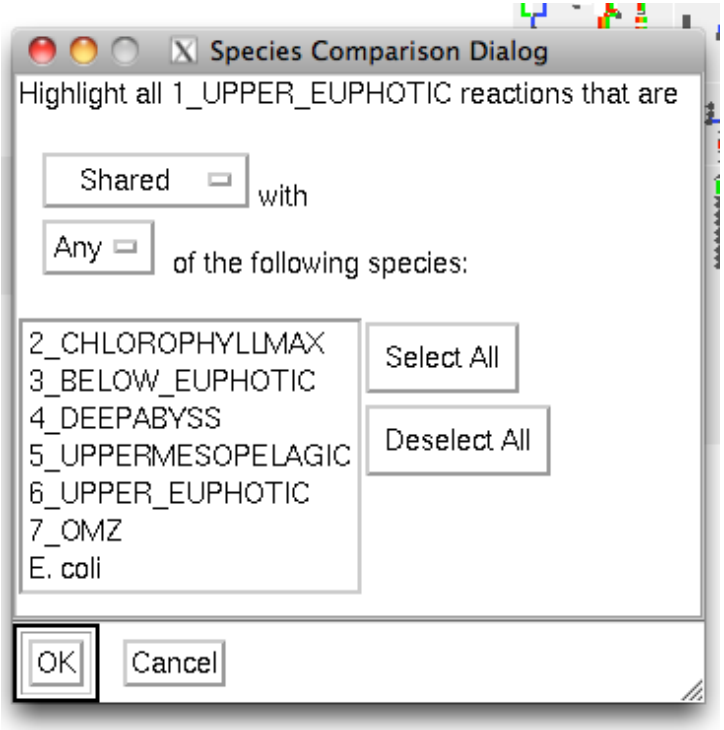
4. Highlight predicted pathways to compare



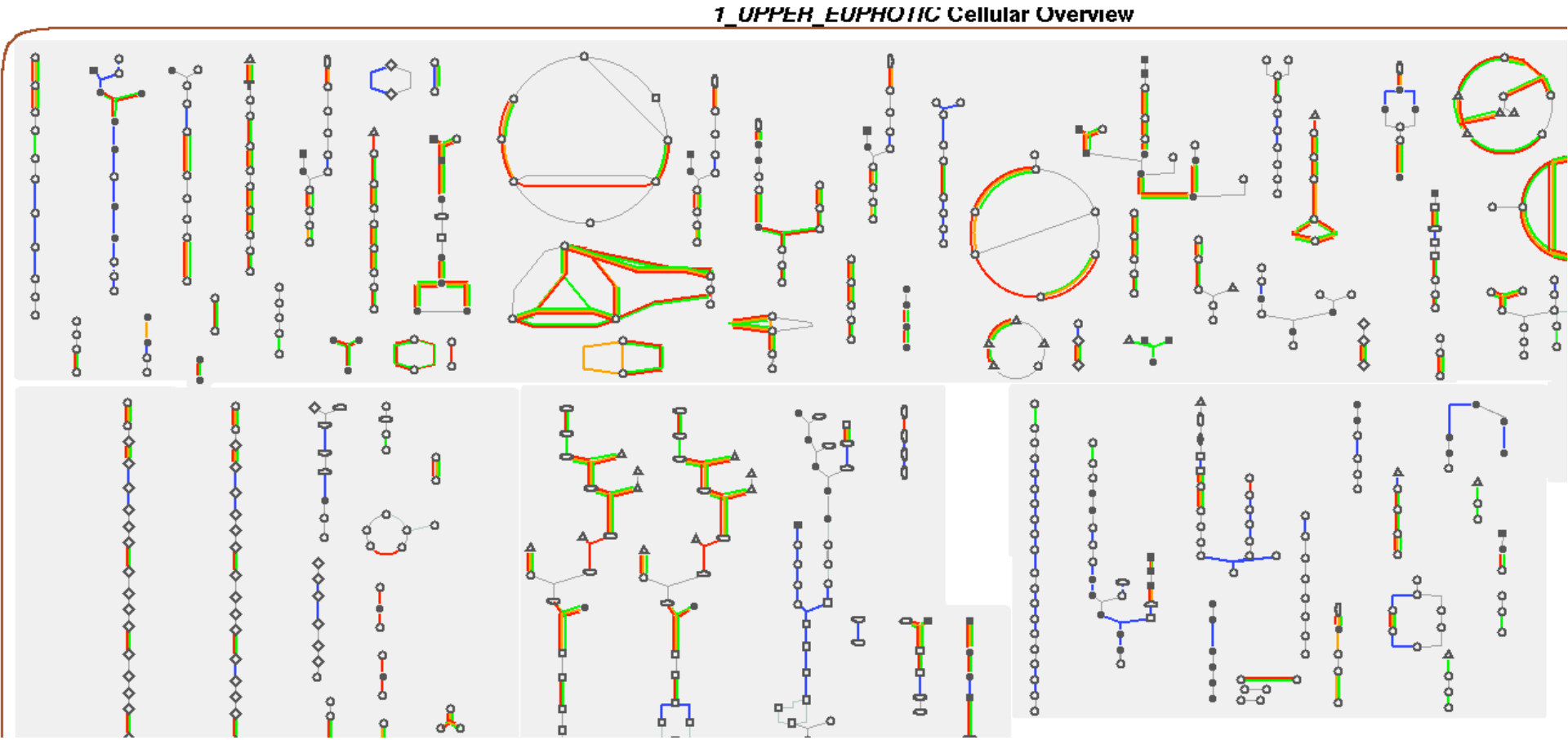


4. Highlight predicted pathways to compare





4. Highlight predicted pathways to compare



5. Extract pathways and ORFs

extract_pathway_table_from_pgdb.pl

1. Start Pathway Tools in –api mode:
   pathway-tools/pathway-tools –api
2. In **another shell** run extract_pathway_table_from_pgdb.pl to extract pathways:

```
# List Available ePGDBs in Pathway Tools
perl extract_pathway_table_from_pgdb.pl -l

# Extract pathways from pathway tools
perl extract_pathway_table_from_pgdb.pl
            -f [list of samples]
            -out [output file]
            -c [pathway coverage] <0.0-1.0>
            -s [pathway support] <1-100>
            -t [type of output table] <lookup, long, or wide>
            -orfs [output orfs in lookup mode]
            -rxn [output reactions in wide mode]
```

5. Extract pathways and ORFs

```
perl extract_pathway_table_from_pgdb.pl -f 1_upper_euphotic -out
1_upper_euphotic.lookup.txt -t lookup
```

```
head 1_upper_euphotic.lookup.txt
SAMPLE   PWY_NAME PWY_COMMON_NAME   NUM_REACTIONS    NUM_COVERED_REACTIONS
    ORF_COUNT
1_upper_euphotic PWY-5913 TCA cycle VI (obligate autotrophs) 11   10   65
1_upper_euphotic REDCITCYC    TCA cycle III (helicobacter)  9    8    33
1_upper_euphotic PWY-5690 TCA cycle II (eukaryotic) 9    8    42
1_upper_euphotic TCA  TCA cycle I (prokaryotic) 10   9    43
1_upper_euphotic ANARESP1-PWY respiration (anaerobic)   13   10   44
```

Only output pathways with more than 50% reactions covered and at least 7 ORFs:

```
perl extract_pathway_table_from_pgdb.pl -f 1_upper_euphotic -out
1_upper_euphotic.lookup.c05.s7.txt -t lookup -c 0.5 -s 7
```

5. Extract pathways and ORFs

the **`-t long`** table format displays each each ORF in each pathway:

```
perl extract_pathway_table_from_pgdb.pl -f 1_upper_euphotic -out
1_upper_euphotic.long.txt -t long
```

```
head 1_upper_euphotic.long.txt
SAMPLE      PWY_NAME    PWY_COMMON_NAME   NUM_REACTIONS     NUM_COVERED_REACTIONS   ORF_COUNT   ORF
1_upper_euphotic  PWY-5913    TCA cycle VI (obligate autotrophs)  11    10    65    1_upper_euphotic_3417_1
1_upper_euphotic  PWY-5913    TCA cycle VI (obligate autotrophs)  11    10    65    1_upper_euphotic_5953_0
1_upper_euphotic  PWY-5913    TCA cycle VI (obligate autotrophs)  11    10    65    1_upper_euphotic_7270_0
1_upper_euphotic  PWY-5913    TCA cycle VI (obligate autotrophs)  11    10    65    1_upper_euphotic_643_1
1_upper_euphotic  PWY-5913    TCA cycle VI (obligate autotrophs)  11    10    65    1_upper_euphotic_14_0
1_upper_euphotic  PWY-5913    TCA cycle VI (obligate autotrophs)  11    10    65    1_upper_euphotic_670_0
```

the **`-t wide`** with multiple samples creates a “master” table:

```
# - wide table format of pathways from multiple samples
perl extract_pathway_table_from_pgdb.pl -f 1_upper_euphotic 6_upper_euphotic
2_chlorophyllmax 3_below_euphotic 5_uppermesopelagic 7_omz 4_deepabyss -out
HOT_Sanger_pwy.wide.txt -t wide
```

```
head 1_upper_euphotic.long.txt
PWY   1_upper_euphotic  6_upper_euphotic  2_chlorophyllmax  3_below_euphotic  5_uppermesopelagic
      7_omz 4_deepabyss
SUCSYN-PWY  0     16    10    8     17    13    16
PWY-6733    0     1     0     0     0     0     0
PWY-5274    0     0     0     1     0     0     0
PWY-6728    0     0     0     46    0     59    0
PWY-241     12    6     8     6     7     12    7
```

5. Extract pathways and ORFs

the **-t wide -rxn** options produce a list of reactions and abundance:

```
perl extract_pathway_table_from_pgdb.pl -f 1_upper_euphotic -out
1_upper_euphotic_rxn.wide.txt -t wide –rxn
```

```
head 1_upper_euphotic_rxn.wide.txt
RXN   1_upper_euphotic
RXN1G-617   1
RXN-6641    1
RXN0-2381   2
DTDPGLUCOSEPP-RXN 2
RXN0-6479   3
DADPKIN-RXN 4
```

the **-t wide -rxn** with multiple samples creates a “master” rxn table:

```
perl extract_pathway_table_from_pgdb.pl -f 1_upper_euphotic 6_upper_euphotic
2_chlorophyllmax 3_below_euphotic 5_uppermesopelagic 7_omz 4_deepabyss -out
HOT_Sanger_rxn.wide.txt -t wide -rxn
```

```
head HOT_Sanger_rxn.wide.txt
PWY   1_upper_euphotic  6_upper_euphotic  2_chlorophyllmax  3_below_euphotic  5_uppermesopelagic
      7_omz 4_deepabyss
SUCSYN-PWY  0     16    10    8     17    13    16
PWY-6733    0     1     0     0     0     0     0
PWY-5274    0     0     0     1     0     0     0
PWY-6728    0     0     0     46    0     59    0
PWY-241     12    6     8     6     7     12    7
```

5. Extract pathways and ORFs

the **-t wide -rxn** options produce a list of reactions and abundance:

```
perl extract_pathway_table_from_pgdb.pl -f 1_upper_euphotic -out
1_upper_euphotic_rxn.wide.txt -t wide –rxn
```

```
head 1_upper_euphotic_rxn.wide.txt
RXN   1_upper_euphotic
RXN1G-617   1
RXN-6641    1
RXN0-2381   2
DTDPGLUCOSEPP-RXN 2
RXN0-6479   3
DADPKIN-RXN 4
```

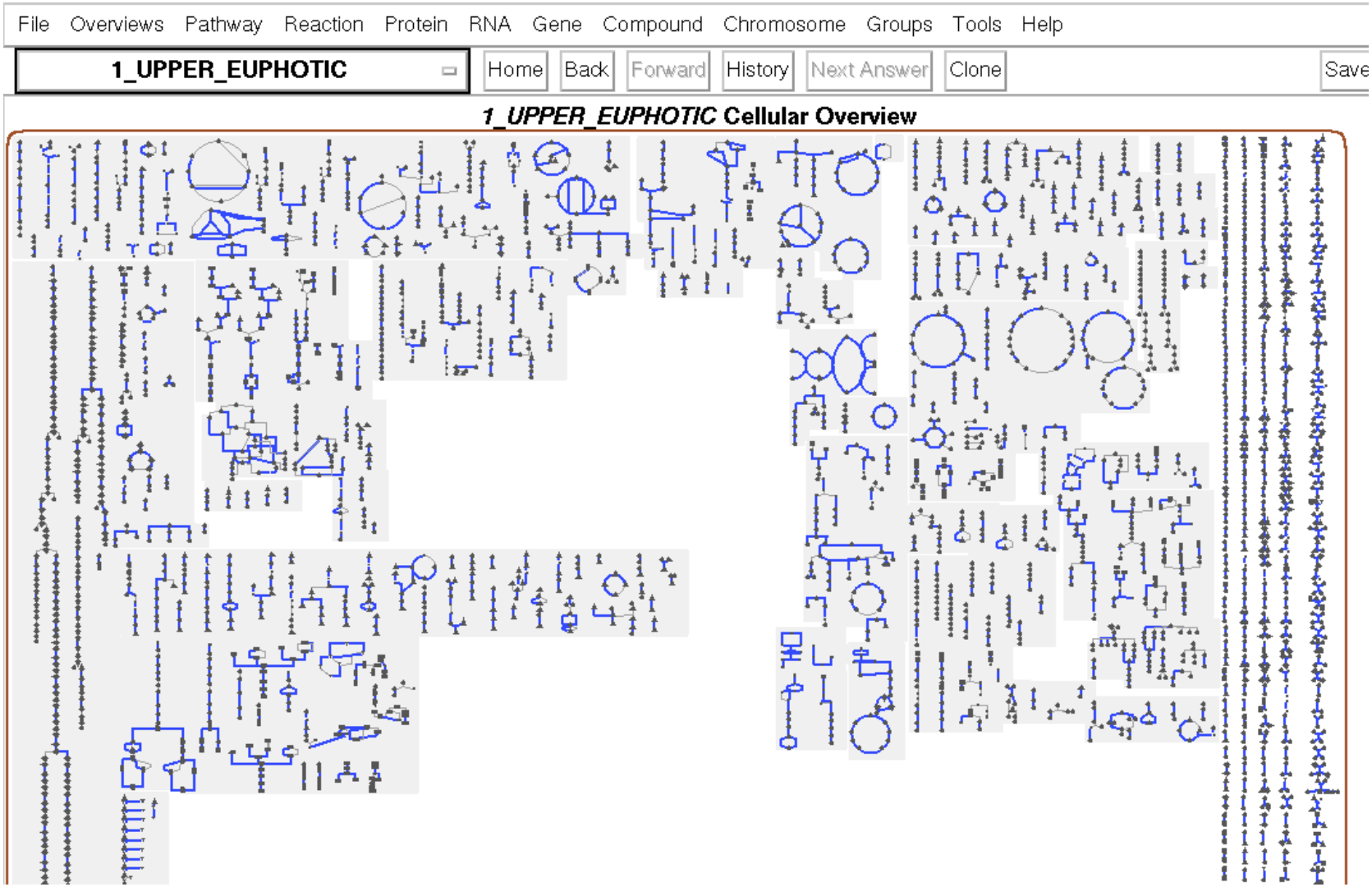
****Use these two to highlight ORF counts on the Cellular Overview****

the **-t wide -rxn** with multiple samples creates a “master” rxn table:

```
perl extract_pathway_table_from_pgdb.pl -f 1_upper_euphotic 6_upper_euphotic
2_chlorophyllmax 3_below_euphotic 5_uppermesopelagic 7_omz 4_deepabyss -out
HOT_Sanger_rxn.wide.txt -t wide -rxn
```

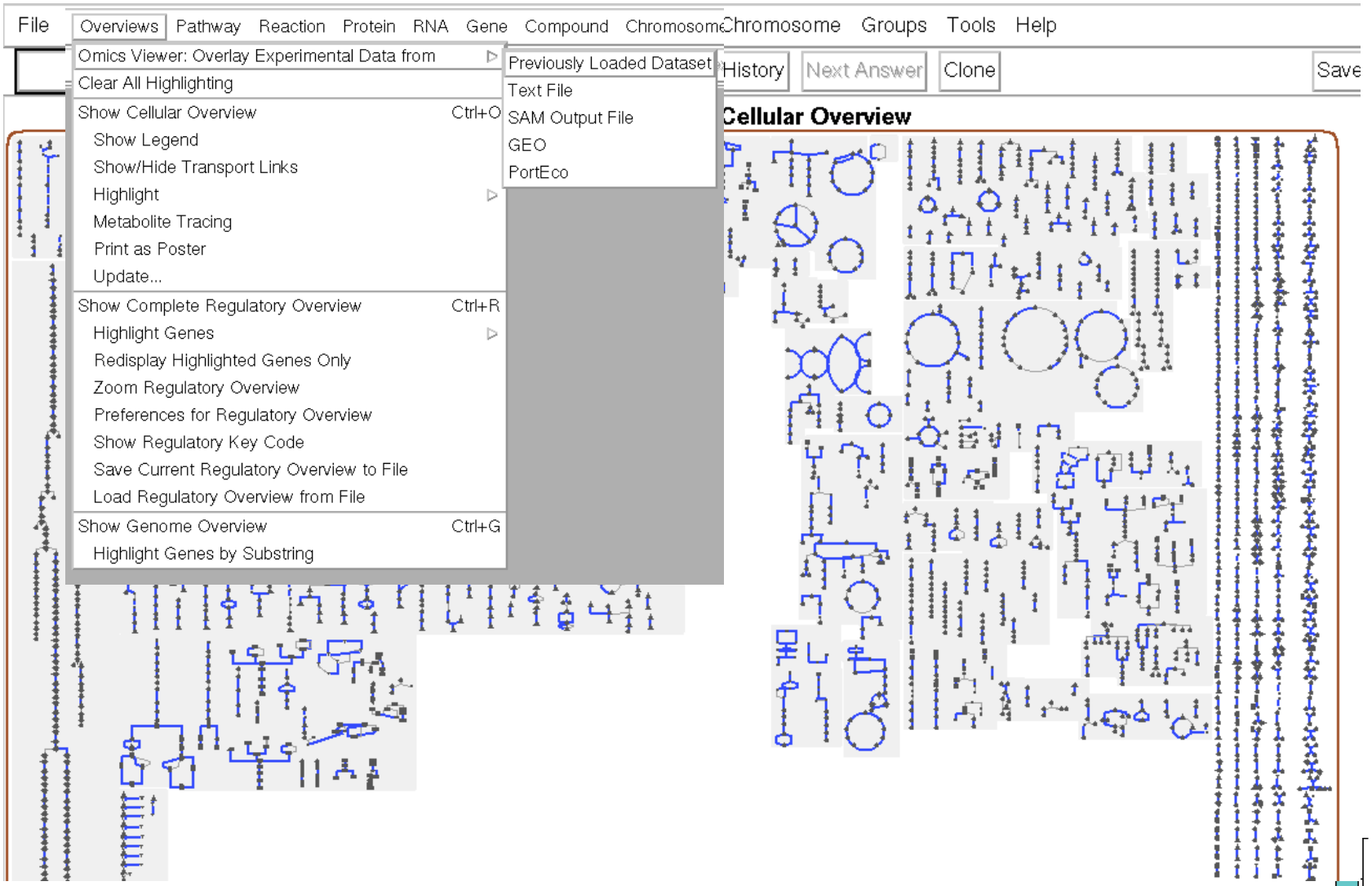
```
head HOT_Sanger_rxn.wide.txt
RXN   1_upper_euphotic  6_upper_euphotic  2_chlorophyllmax  3_below_euphotic  5_uppermesopelagic
      7_omz 4_deepabyss
1.7.7.2-RXN 0     0     2     0     0     1     1
PPGPPSYN-RXN      0     1     0     0     0     0     0
RXN-6641    1     0     1     0     1     3     2
RXN3DJ-170  1     0     0     0     0     0     2
```

6. 'Omics Tools' feature for Metadata



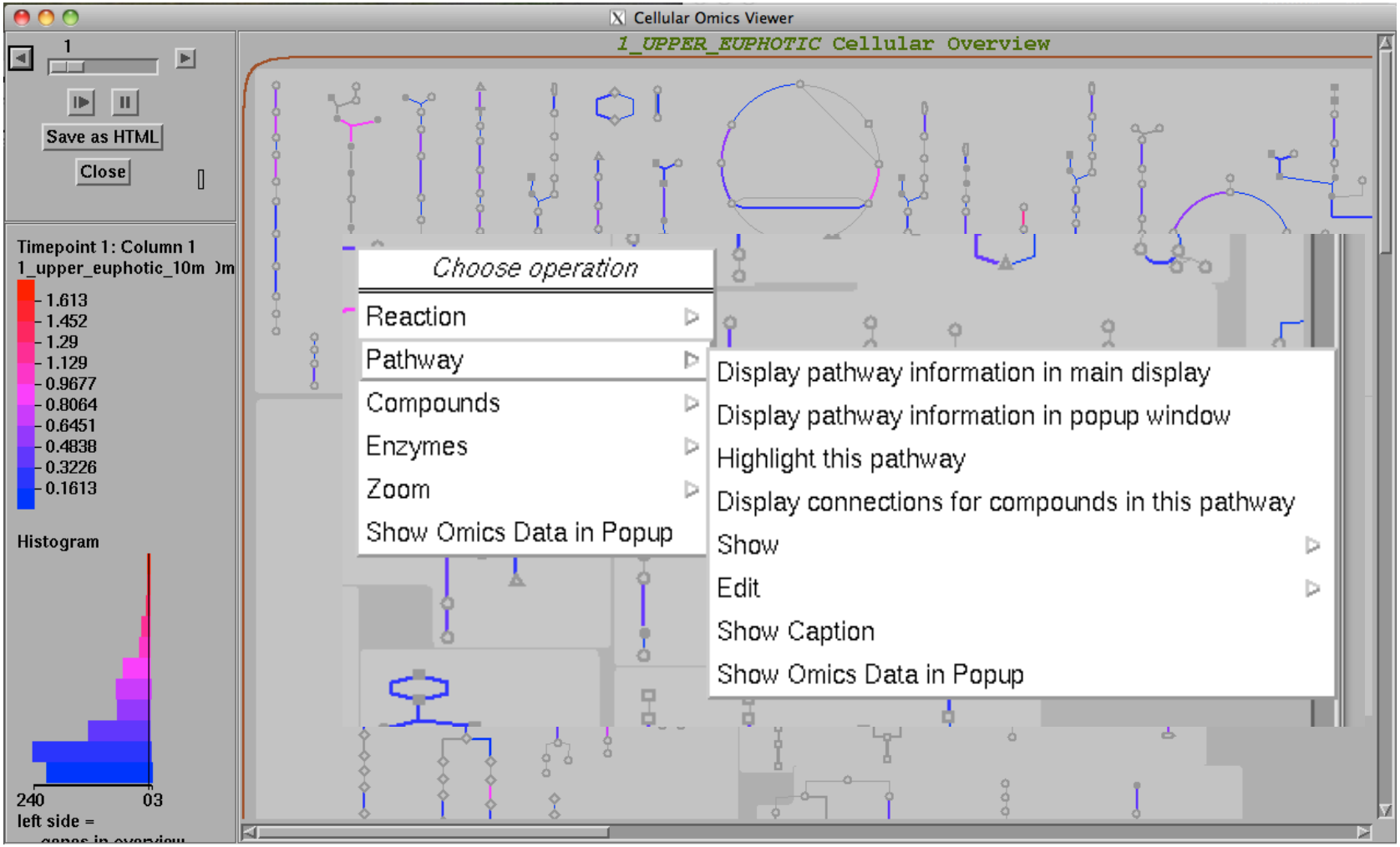
6. 'Omics Tools' feature for Metadata

Load `1_upper_euphotic_rxn.wide.txt`

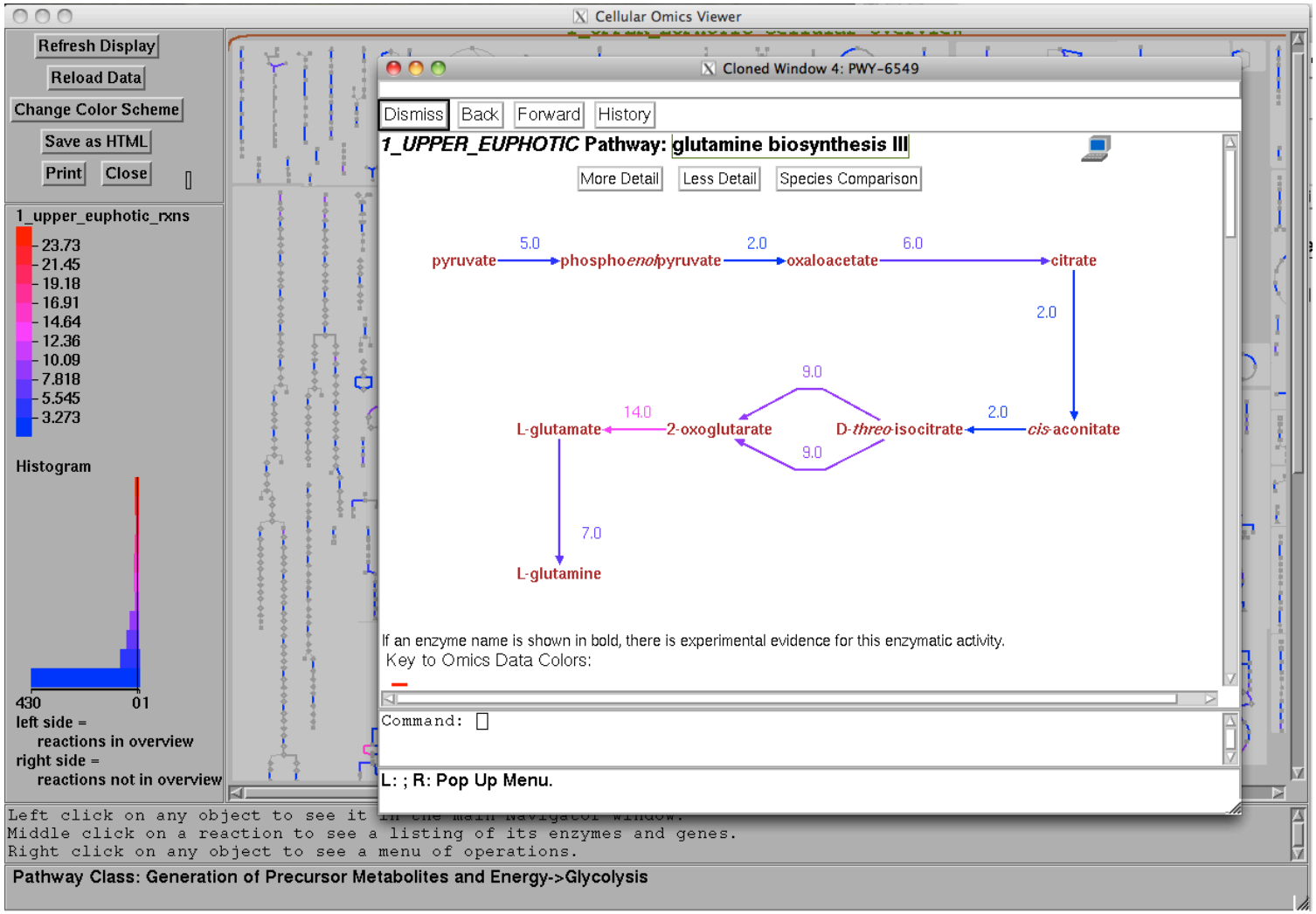


6. 'Omics Tools' feature for Metadata

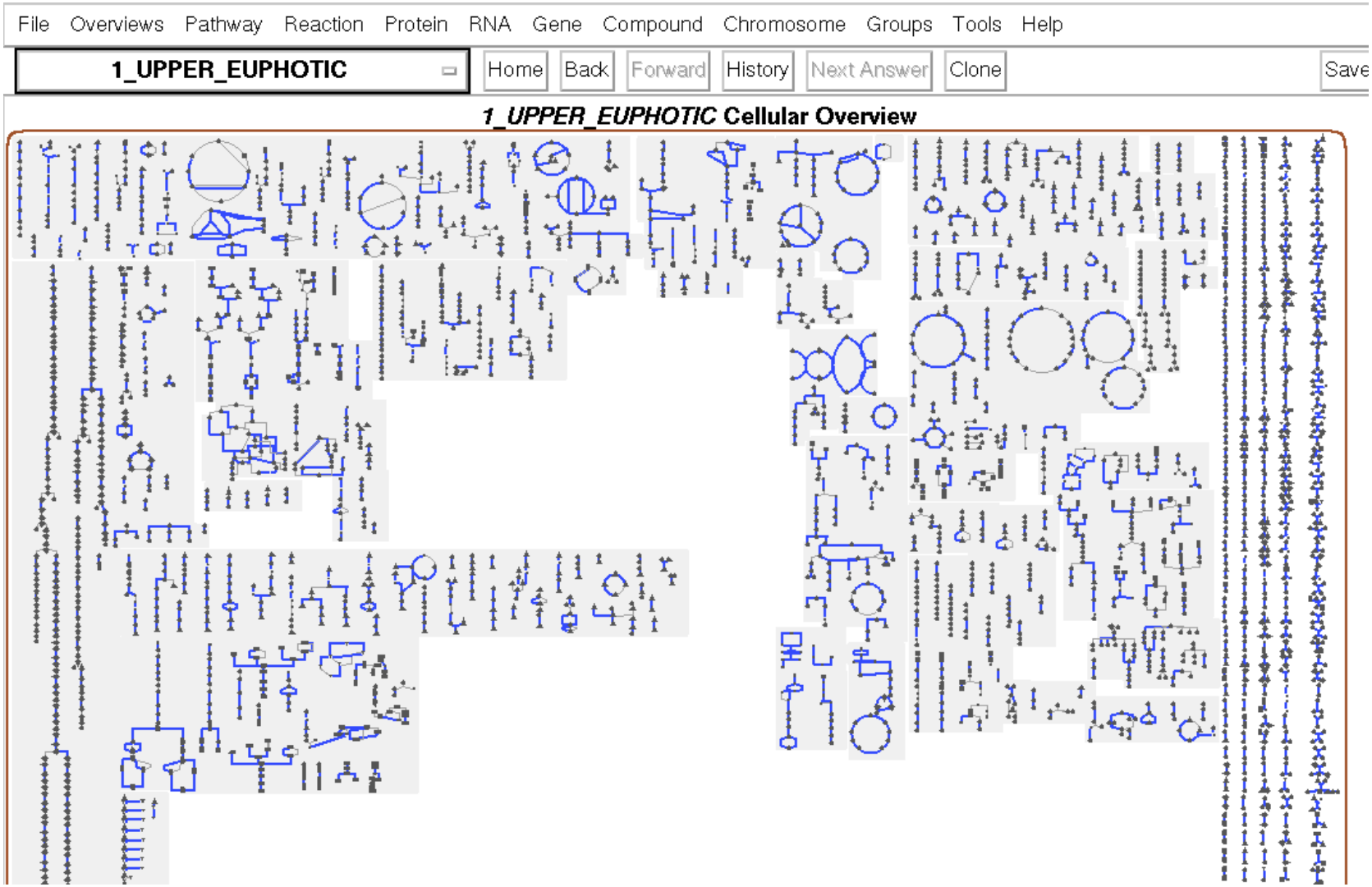
6. 'Omics Tools' feature for Metadata



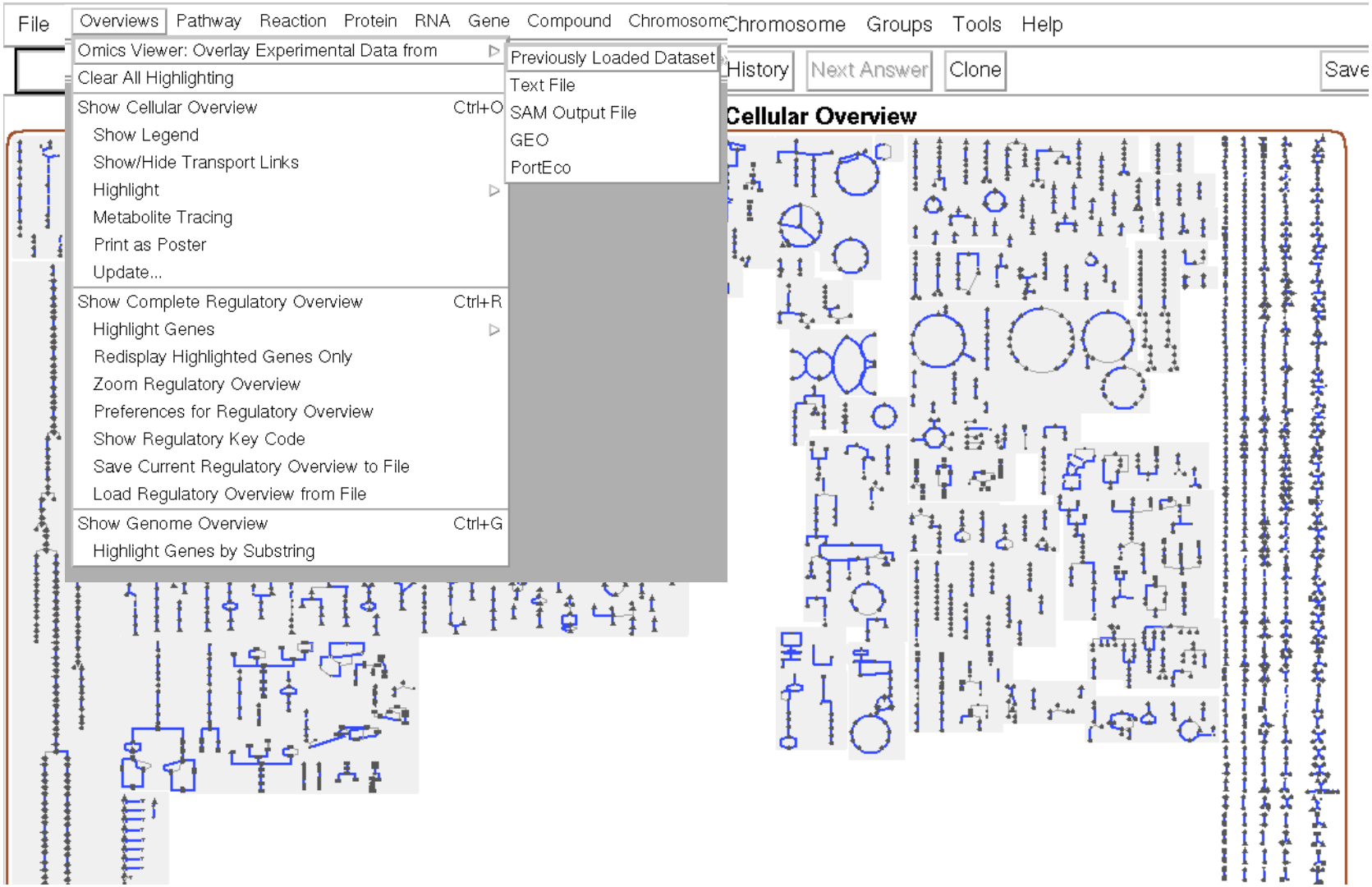
6. 'Omics Tools' feature for Metadata



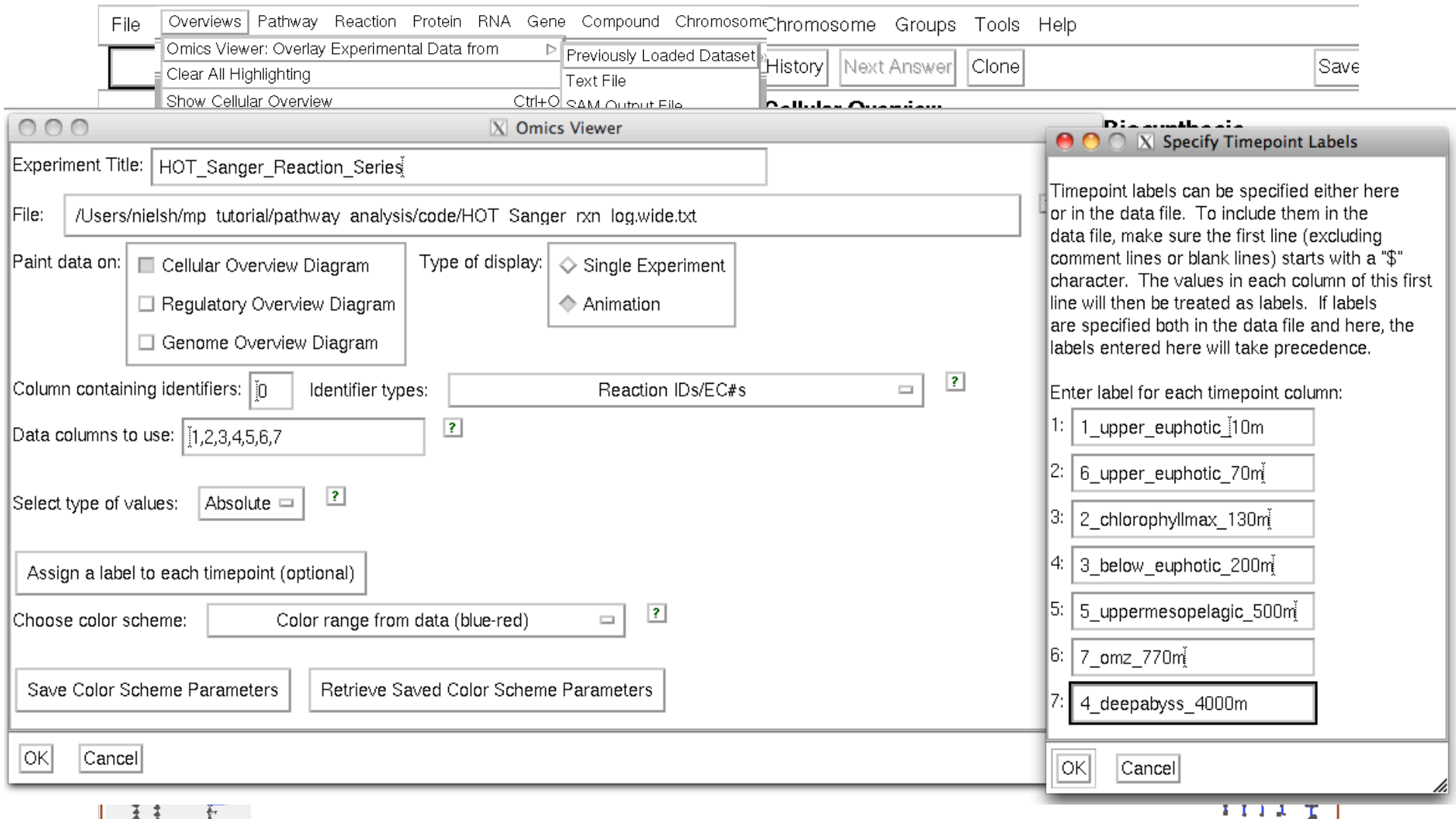
6. 'Omics Tools' feature for Metadata



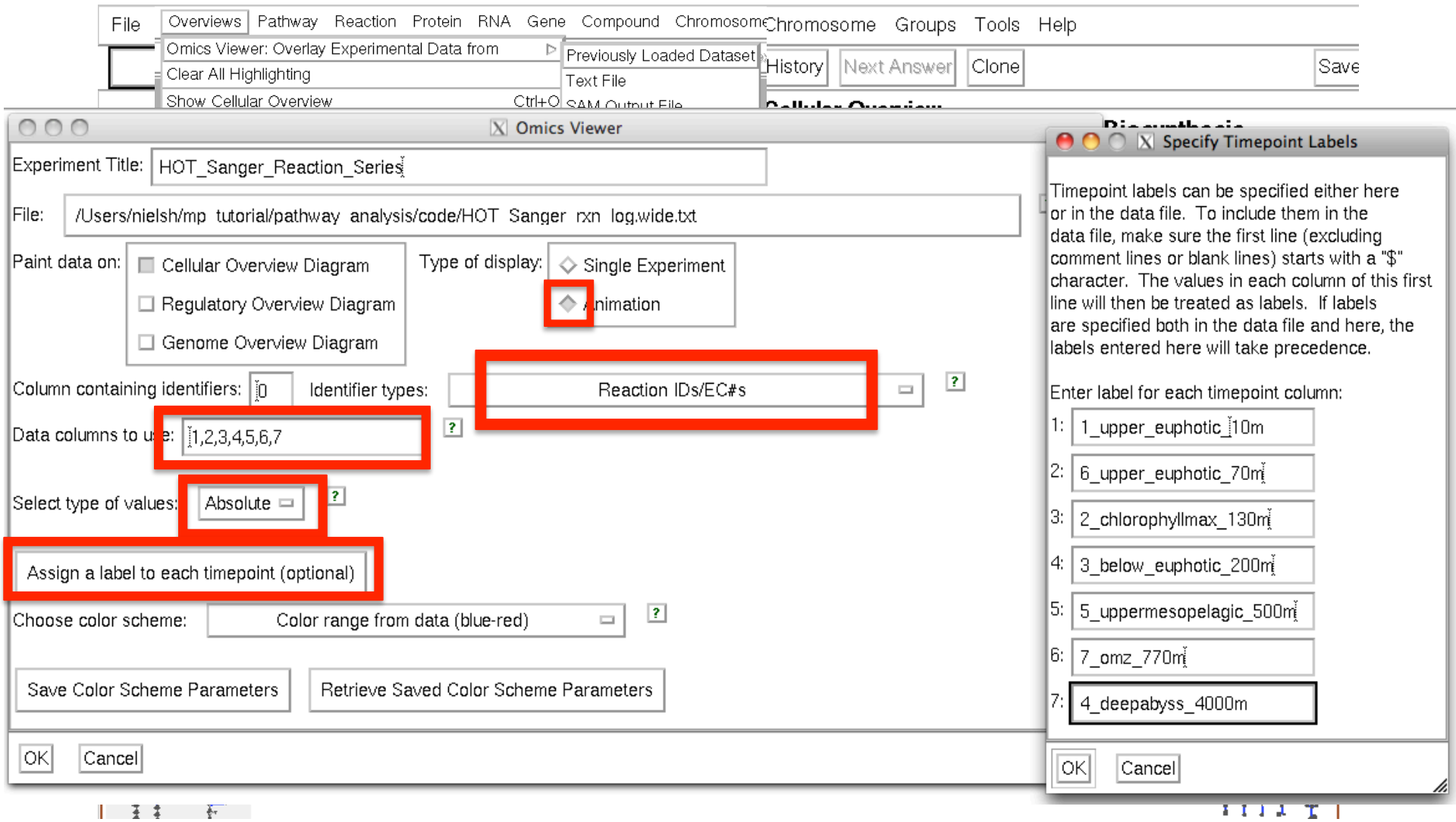
6. 'Omics Tools' feature for Metadata



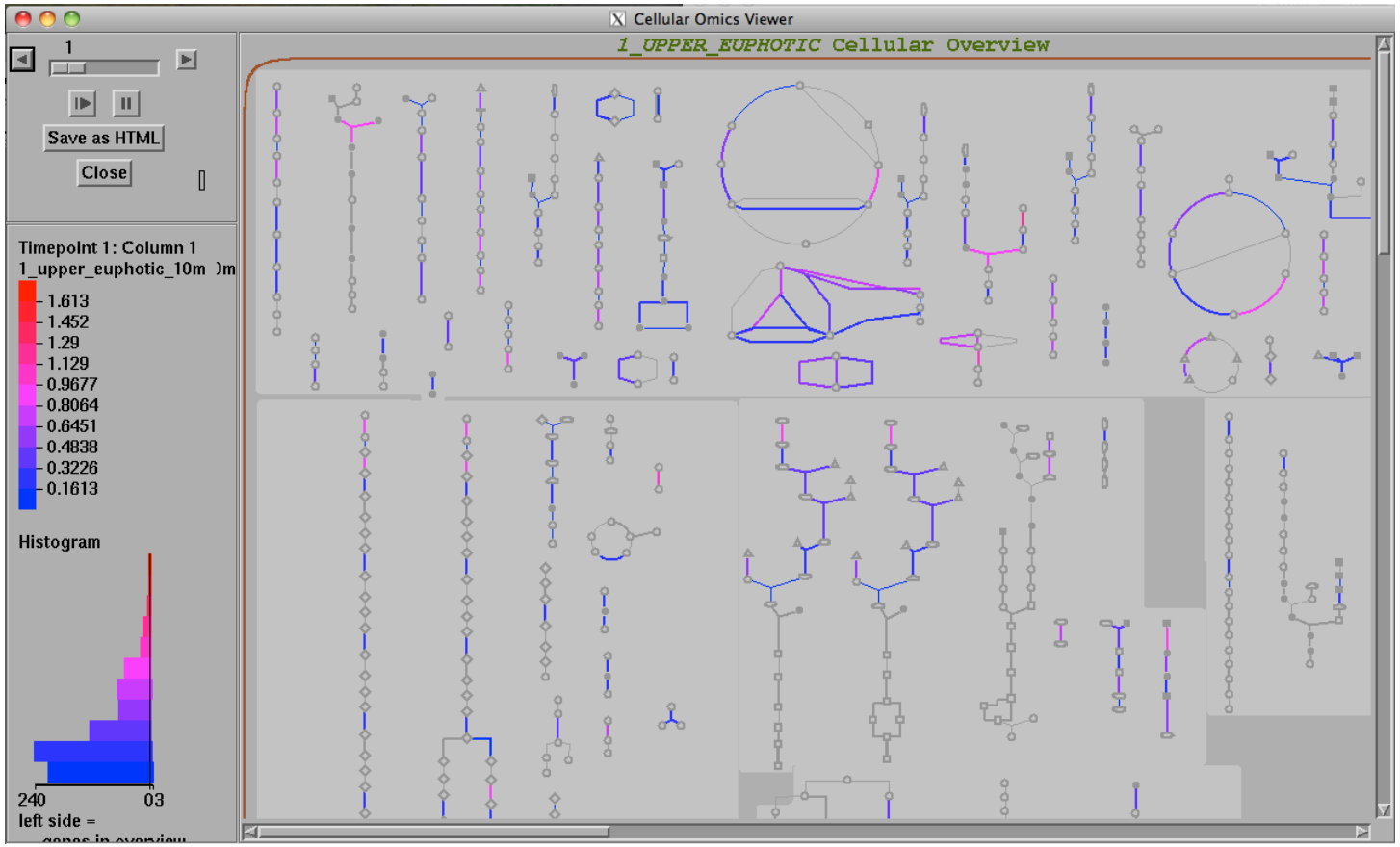
6. 'Omics Tools' feature for Metadata



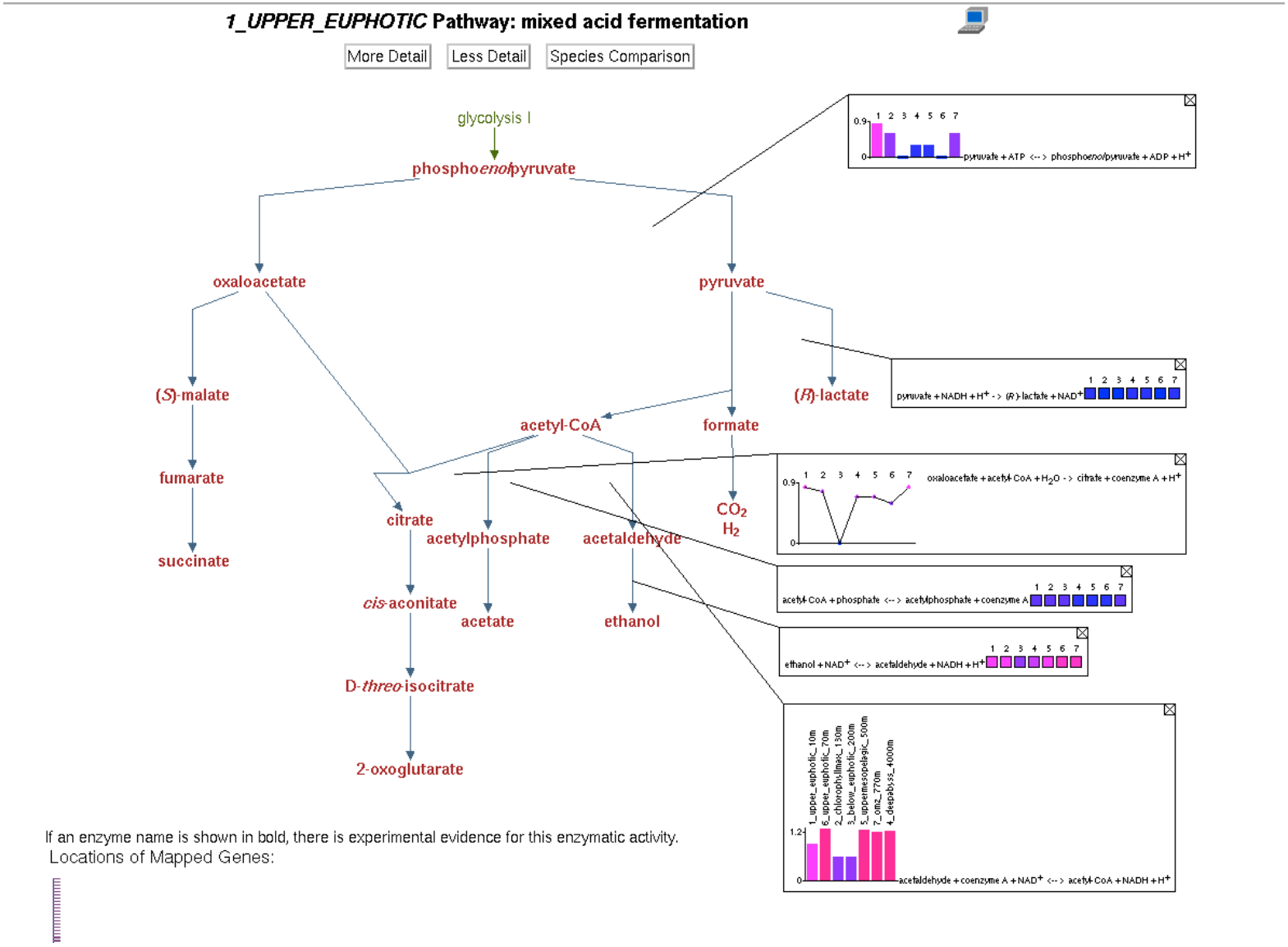
6. 'Omics Tools' feature for Metadata



6. 'Omics Tools' feature for Metadata



6. 'Omics Tools' feature for Metadata



Questions?