

# Supplementary material for: MetaNetter: inference and visualization of high-resolution metabolomic networks.

**Fabien Jourdan**, UMR1089 Xénobiotiques INRA-ENVT, Toulouse, France

**Rainer Breitling**, Groningen Bioinformatics Centre, University of Groningen, the Netherlands

**Michael P. Barrett**, Division of Infection and Immunity, Institute of Biomedical and Life Sciences, University of Glasgow

**David Gilbert** Bioinformatics Research Centre, University of Glasgow

1. Transformation List.....	1
2. Glycolysis example .....	3

## 1. Transformation List

Table 1 contains a sample transformation list based on a bibliographic study. This list is under development and can be directly edited and expanded using the MetaNetter plugin.

<b>Transformation label</b>	<b>Formula</b>	<b>Mass</b>
Isomeric		0
Alanine	C3H5NO	71.0371138
Arginine	C6H12N4O	156.101111
Asparagine	C4H6N2O2	114.042928
Aspartic Acid	C4H5NO3	115.026943
Cysteine	C3H5NOS	103.009186
Cystine	C6H10N2O3S2	222.013286
Glutamic Acid	C5H7NO3	129.042593
Glutamine	C5H8N2O2	128.058578
Glycine	C2H3NO	57.0214638
Histidine	C6H7N3O	137.058912
Isoleucine	C6H11NO	113.084064
Leucine	C6H11NO	113.084064
Lysine	C6H12N2O	128.094963
Methionine	C5H9NOS	131.040486
Phenylalanine	C9H9NO	147.068414
Proline	C5H7NO	97.0527639
Serine	C3H5NO2	87.0320285
Threonine	C4H7NO2	101.047679
Tryptophan	C11H10N2O	186.079313
Tyrosine	C9H9NO2	163.063329
Valine	C5H9NO	99.068414
acetotacetate (-H2O)	C4H4O2	84.0211294
acetone (-H)	C3H5O	57.0340398
adenylate (-H2O)	C10H12N5O6P	329.052522
biotinyl (-H)	C10H15N2O3S	243.080339
biotinyl (-H2O)	C10H14N2O2S	226.0776
carbamoyl P transfer (-H2PO4)	CH2ON	44.0136387
co-enzyme A (-H)	C21H34N7O16P3S	765.099566
co-enzyme A (-H2O)	C21H33N7O15P3S	748.096826
glutathione (-H2O)	C10H15N3O5S	289.073243

isoprene addition (-H)	C5H7	67.0547753
malonyl group (-H2O)	C3H2O3	86.000394
palmitoylation (-H2O)	C16H30O	238.229666
pyridoxal phosphate (-H2O)	C8H8NO5P	229.014011
urea addition (-H)	CH3N2O	59.0245378
adenine (-H)	C5H4N5	134.04667
adenosine (-H2O)	C10H11N5O3	249.086189
Adenosine 5'-diphosphate (-H2O)	C10H13N5O9P2	409.018854
Adenosine 5' monophosphate (-H2O)	C10H12N5O6P	329.052522
cytidine 5' diphosphate (-H2O)	C9H13N3O10P2	385.007621
cytidine 5' monophosphate (-H2O)	C9H12N3O7P	305.041288
cytosine (-H)	C4H4N3O	110.035437
Guanosine 5- diphosphate (-H2O)	C10H13N5O10P2	425.013769
Guanosine 5- monophosphate (-H2O)	C10H12N5O7P	345.047436
guanine (-H)	C5H4N5O	150.041585
guanosine (-H2O)	C10H11N5O4	265.081104
deoxythymidine 5' diphosphate (-H2O)	C10H14N2O10P2	384.012372
thymidine (-H2O)	C10H12N2O4	224.079707
thymine (-H)	C5H5N2O2	125.035102
thymidine 5' monophosphate (-H2O)	C10H13N2O7P	304.046039
uridine 5' diphosphate (-H2O)	C9H12N2O11P2	385.991636
uridine 5' monophosphate (-H2O)	C9H11N2O8P	306.025304
uracil (-H)	C4H3N2O2	111.019452
uridine (-H2O)	C9H10N2O5	226.058972
acetylation (-H)	C2H3O2	59.0133044
acetylation (-H2O)	C2H2O	42.0105647
C2H2	C2H2	26.0156501
Carboxylation	CO2	43.9898293
CHO2	CHO2	44.9976543
condensation/dehydration	H2O	18.0105647
diphosphate	H3O6P2	160.94049
ethyl addition (-H2O)	C2H4	28.0313001
Formic Acid (-H2O)	CO	27.9949146
glyoxylate (-H2O)	C2O2	55.9898293
hydrogenation/dehydrogenation	H2	2.01565007
hydroxylation (-H)	O	15.9949146
Inorganic Phosphate	P	30.9737634
ketol group (-H2O)	C2H2O	42.0105647
methanol (-H2O)	CH2	14.0156501
phosphate	HPO3	79.9663324
primary amine	NH2	16.0187241
pyrophosphate	PP	61.9475268
secondary amine	NH	15.010899
sulfate (-H2O)	SO3	79.9568157
tertiary amine	N	14.003074
C6H10O5	C6H10O5	162.052824
C6H10O6	C6H10O6	178.047738
D-Ribose (-H2O) (ribosylation)	C5H8O4	132.042259
disaccharide (-H2O)	C12H20O11	340.100562
glucose-N-Phosphate (-H2O)	C6H11O8P	242.019156

Glucuronic Acid (-H <sub>2</sub> O)	C <sub>6</sub> H <sub>8</sub> O <sub>6</sub>	176.032088
monosaccharide (-H <sub>2</sub> O)	C <sub>6</sub> H <sub>10</sub> O <sub>5</sub>	162.052824
trisaccharide (-H <sub>2</sub> O)	C <sub>18</sub> H <sub>30</sub> O <sub>15</sub>	486.158471
erythrose (-H <sub>2</sub> O)	C <sub>4</sub> H <sub>8</sub> O <sub>4</sub>	102.031695
transamination (-O)	NH <sub>4</sub>	2.039459

**Table 1:** Sample transformation list based on text-books (e.g. J.G. Salway, 'Metabolism at a Glance') or databases (e.g. KEGG). Originally developed by Phenomenome Discoveries (<http://www.phenomenome.com>) as used in Breitling et al. 2006 Metabolomics 2, 155-64).

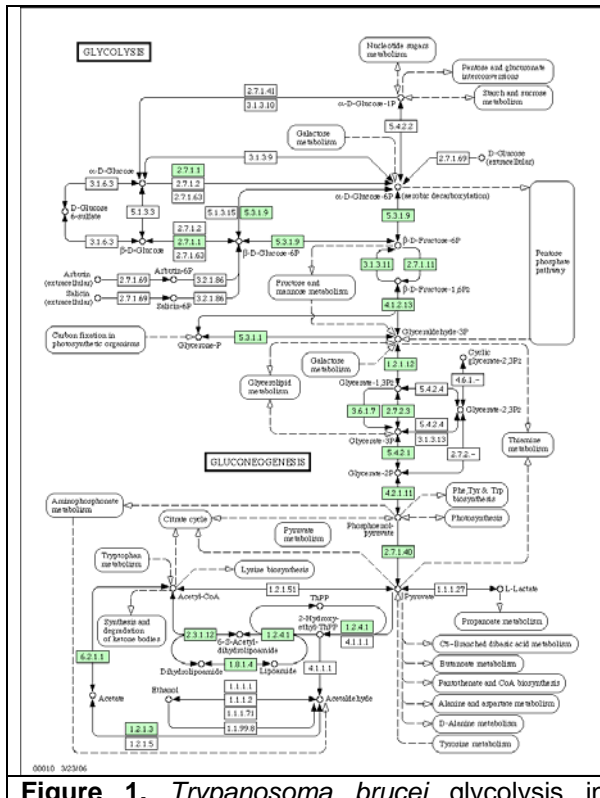
## 2. Glycolysis example

### • Materials and methods

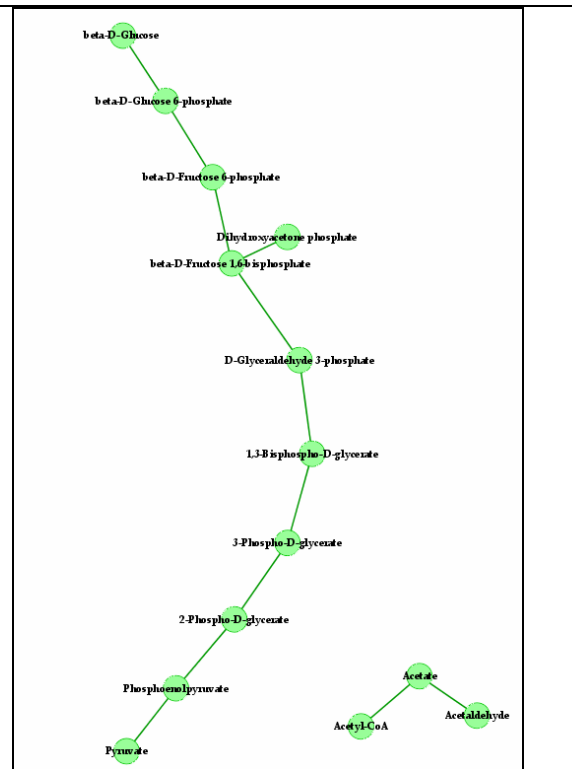
In order to illustrate the inference of an *ab initio* network, we used an artificial test case. Instead of using masses measured by a metabolomics mass spectrometry experiment, we used exact masses from a known metabolic pathway. From the KEGG database, we selected the glycolysis pathway and transferred all compounds that are metabolized by enzymes shown to be present in *T. brucei* (based on the genome annotation) into a table along with its mass (given in KEGG to 0.001 AMU). This table of metabolites was then used as the metabolite input to MetaNetter and used to generate a network using the biochemical connections in Table 1. This test case, where we know all metabolites and their correct connectivities, serves to present the general concept and limitations of *ab initio* network reconstruction as discussed below.

### • Result

In the first step, we imported the description of the glycolytic pathway provided by KEGG for *Trypanosoma brucei* (Figure 1 and Figure 2) into the MetaNetter plug-in (Figure 3).



**Figure 1.** *Trypanosoma brucei* glycolysis in



**Figure 2.** KEGG glycolytic pathway imported in

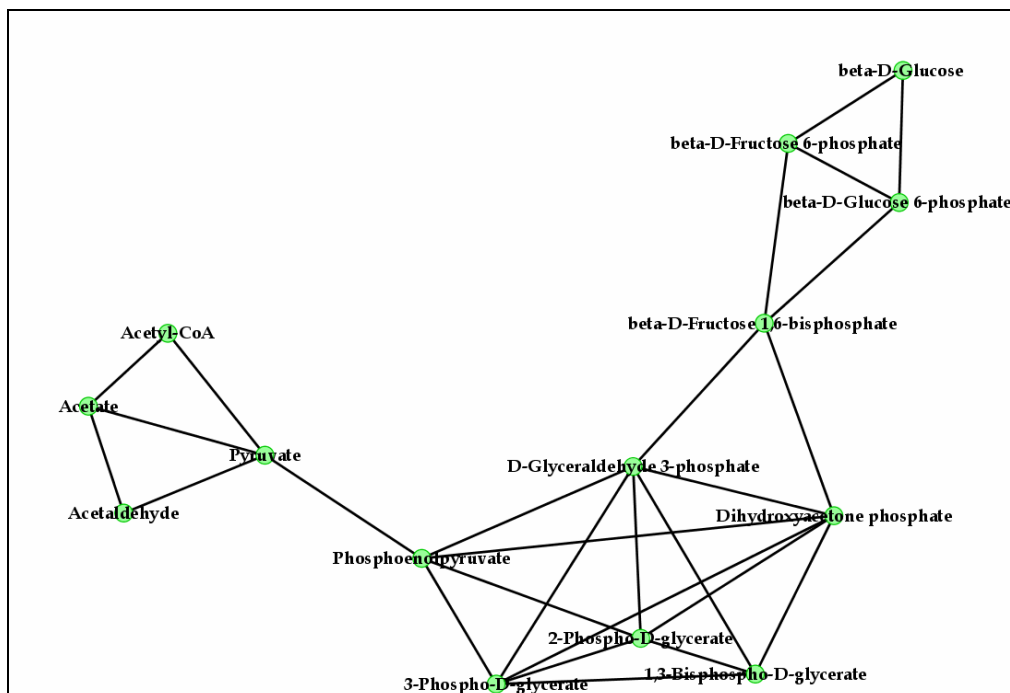
KEGG. Those enzymes found in the *T. brucei* genome annotation are marked in green. No consideration was given to life-cycle expression for the purposes of this example.

Cytoscape.

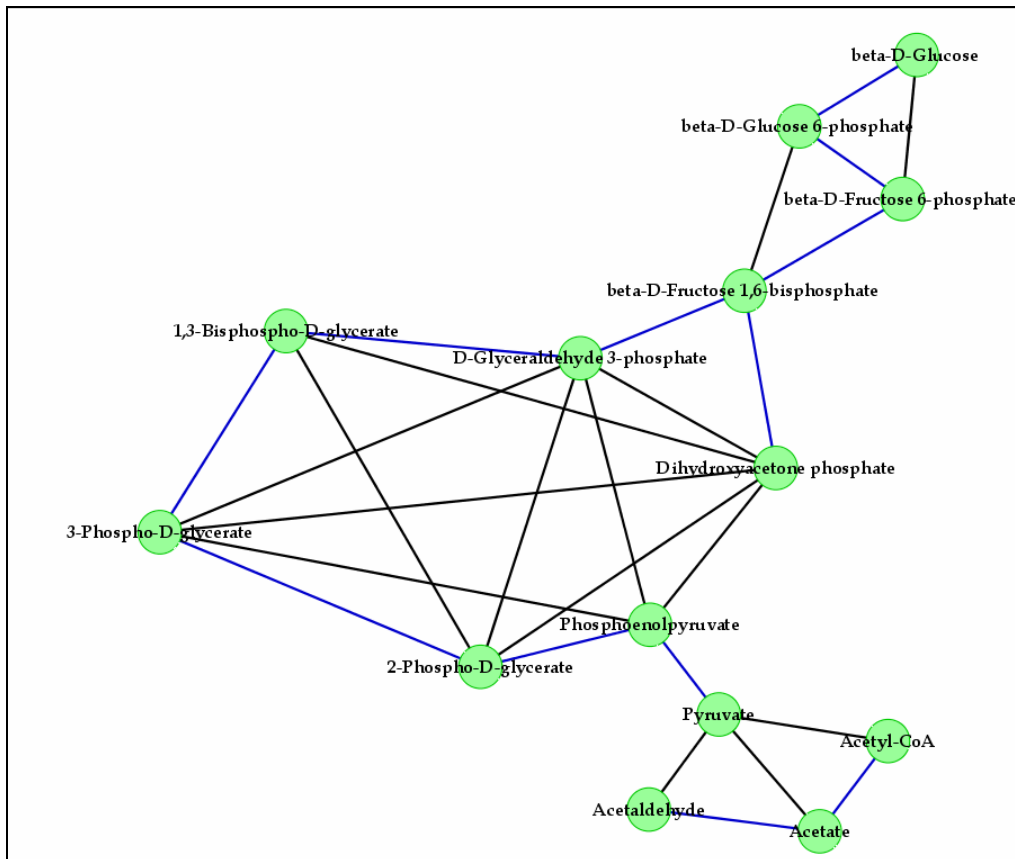
For each compound KEGG also provides the high accuracy mass (Table 2 shows a list of compounds). Figure 3 shows the *ab initio* network obtained with the metabolite mass data from Table 2. We added to the transformation list the isomeric transformation where the difference between two masses is  $0 \pm 2$  ppm. Figure 3 shows KEGG edges added to this network. Two masses are linked by a KEGG edge if they are involved in the same reaction.

Compound	Mass
beta-D-Fructose 1,6-bisphosphate	339.996
Dihydroxyacetone phosphate	169.998
D-Glyceraldehyde 3-phosphate	169.9981
beta-D-Glucose 6-phosphate	260.0298
Acetyl-CoA	809.1256
beta-D-Fructose 6-phosphate	260.03
beta-D-Glucose	180.0634
Pyruvate	88.016
2-Phospho-D-glycerate	185.993
1,3-Bisphospho-D-glycerate	265.9593
3-Phospho-D-glycerate	185.9931
Acetate	60.0211
Phosphoenolpyruvate	167.9824
Acetaldehyde	44.0262

**Table 2.** Compounds (with their masses) that are involved in the *Trypanosoma brucei* glycolytic pathway. Note that isomers are associated with masses varying from only 0.0001.



**Figure 3.** *Ab initio* pathways based on masses in Table 2.



**Figure 4.** Merging the *ab initio* and the KEGG network. All the blue edges are in both networks. There is no edges in KEGG that are not retrieved by the *ab initio* process.

- **Conclusion**

It is notable that the KEGG pathway is recreated more or less as according to classical text book representations. The *ab initio* reconstruction contains a number of edges (potential biochemical connections) that are not present in the known pathway (for example the mass of dihydroxyacetone phosphate could link to the mass of 1,3-bisphosphoglycerate when mass alone is considered). Such false positives are an inevitable consequence of the method, which reports on *all possible* transformations between metabolites within a data set.