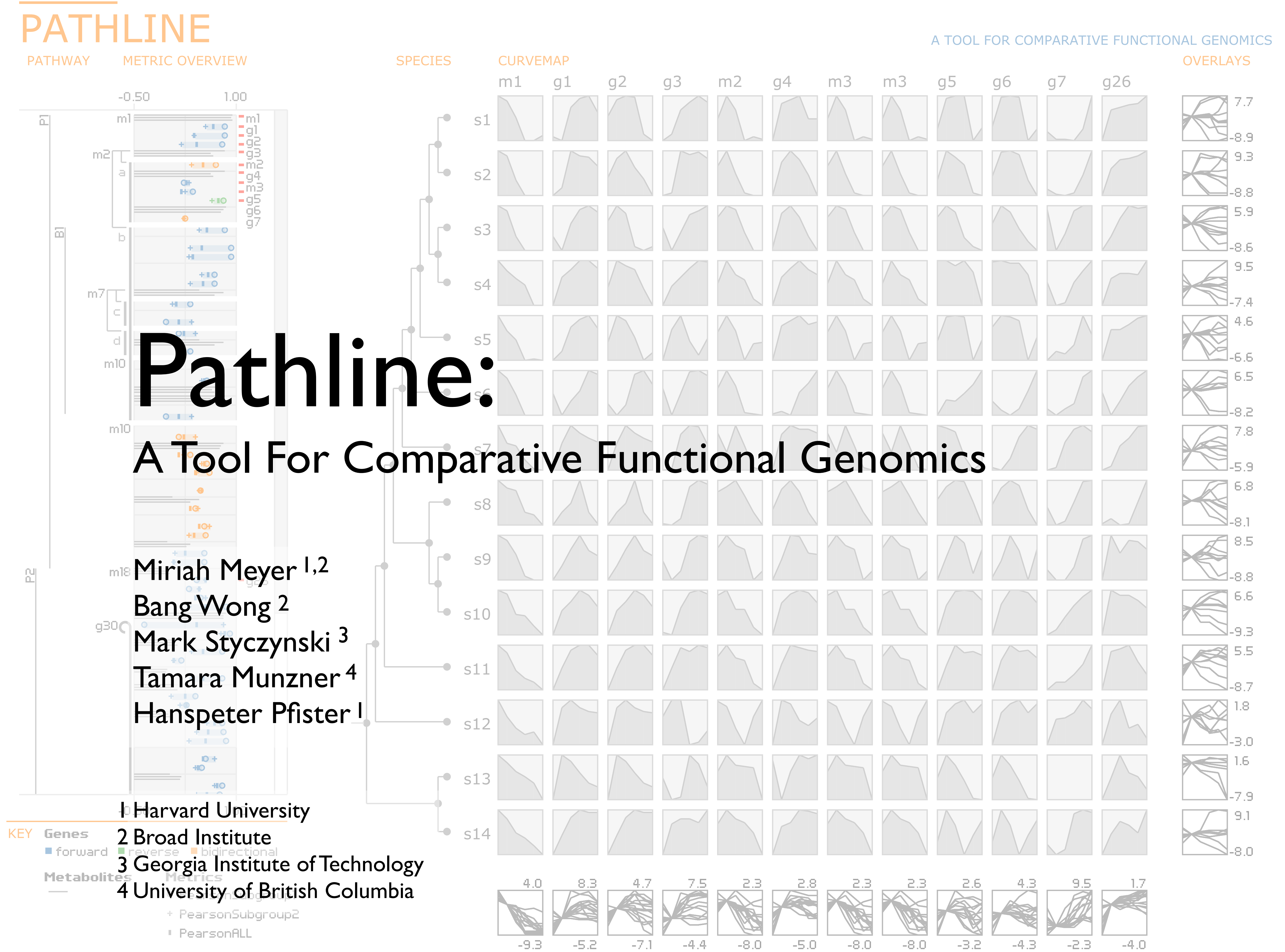


Pathline:

A Tool For Comparative Functional Genomics

Miriah Meyer [1,2]
Bang Wong [2]
Mark Styczynski [3]
Tamara Munzner [4]
Hanspeter Pfister [1]

1 Harvard University
2 Broad Institute
3 Georgia Institute of Technology
4 University of British Columbia

roadmap

- background
- data & tasks
- Pathline
- case studies
- conclusions and future work

background

functional genomics

how do genes work together to perform different functions in a cell?

functional genomics data

gene expression
molecular pathways

gene expression is …

… the measured level of how much a gene is on or off

… a single quantitative value

0.2

gene expression is …

… the measured level of how much a gene is on or off

… a single quantitative value

biologists measure it …

… for many genes

g1	0.2
g2	1.0
g3	-0.7
g4	1.0
g5	-0.5
g6	-0.7
g7	-1.0
g8	-0.5

gene expression is …

… the measured level of how much a gene is on or off

… a single quantitative value

biologists measure it …

… for many genes

… in many samples (time points, tissue types, species)

	s1	s2	s3	s4	s5
g1	0.2	0.4	1.0	1.0	1.0
g2	1.0	0.0	0.0	0.0	1.0
g3	-0.7	0.8	1.0	1.0	0.8
g4	1.0	0.0	0.2	0.5	1.0
g5	-0.5	0.8	0.5	-0.3	-0.5
g6	-0.7	0.5	0.8	-0.7	-1.0
g7	-1.0	-0.3	0.4	-1.0	-1.0
g8	-0.5	0.0	0.0	-0.7	-0.5

gene expression is ...

... the measured level of how much a gene is on or off

... a single quantitative value

biologists measure it ...

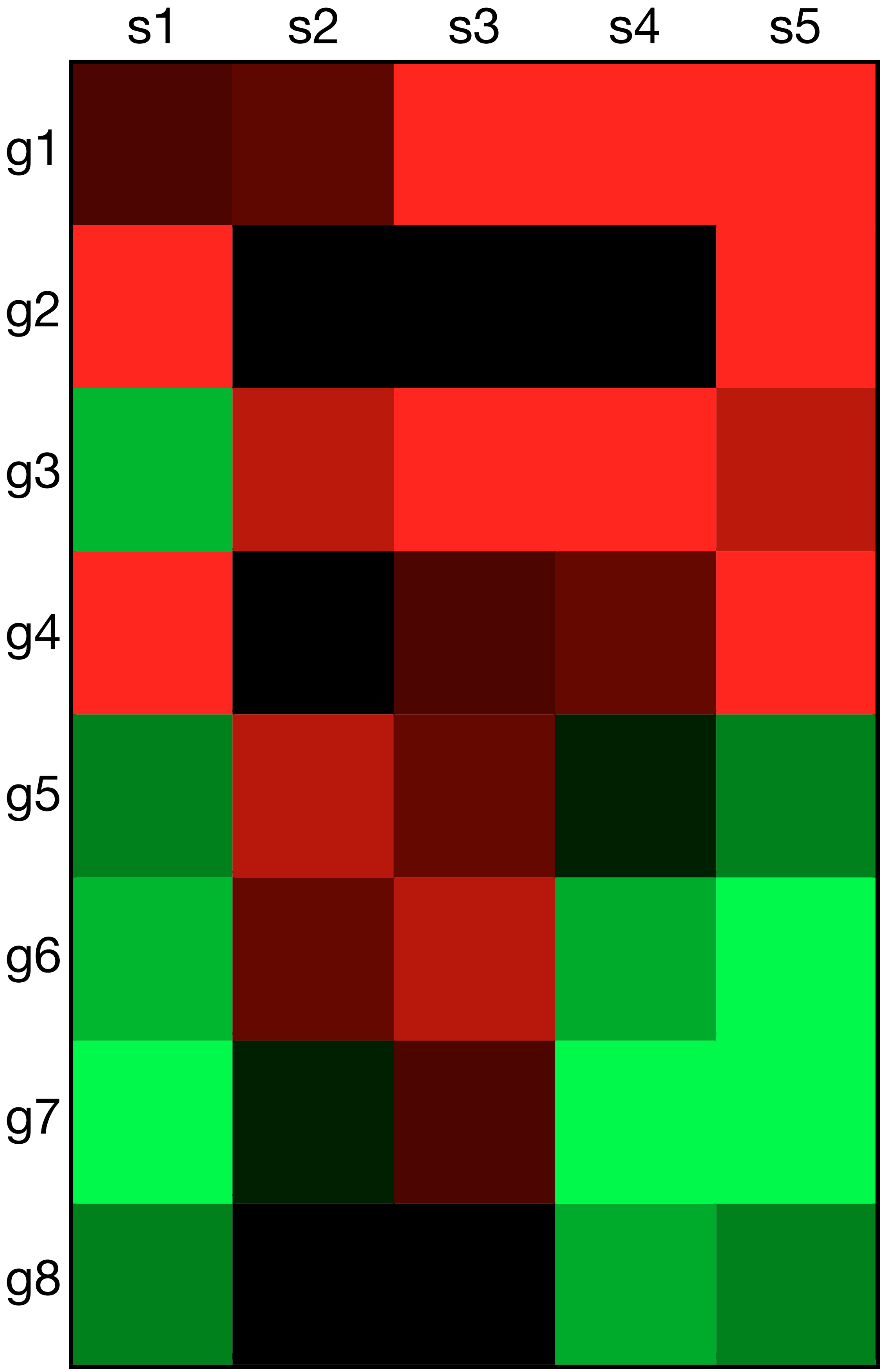
... for many genes

... in many samples (time points, tissue types, species)

visualized with heatmaps

[Wilkinson09][Saldanha04][Seo02][Eisen98]
[Gehlenborg10][Weinstein08]

encode value with color



gene expression is ...

... the measured level of how much a gene is on or off

... a single quantitative value

it is measured ...

... for many genes

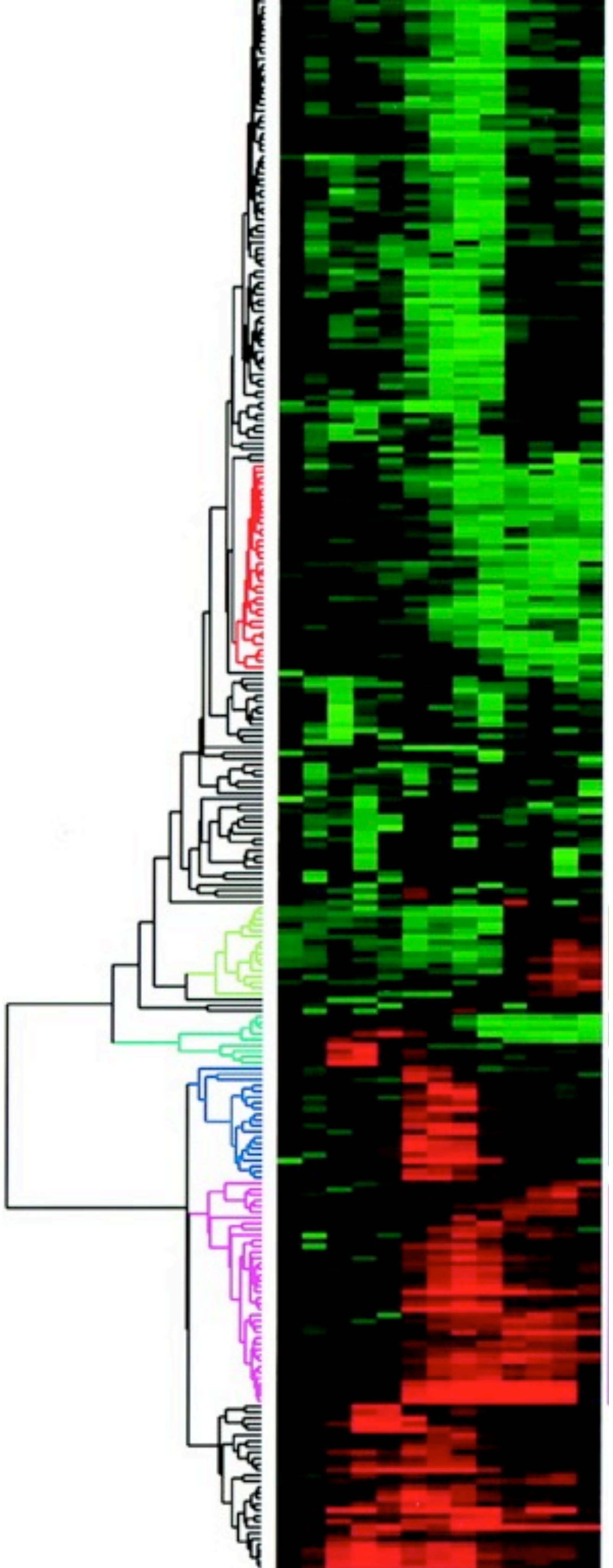
... in many samples (time points, tissue types, species)

visualized with heatmaps

[Wilkinson09] [Saldanha04] [Seo02] [Eisen98]
[Gehlenborg10] [Weinstein08]

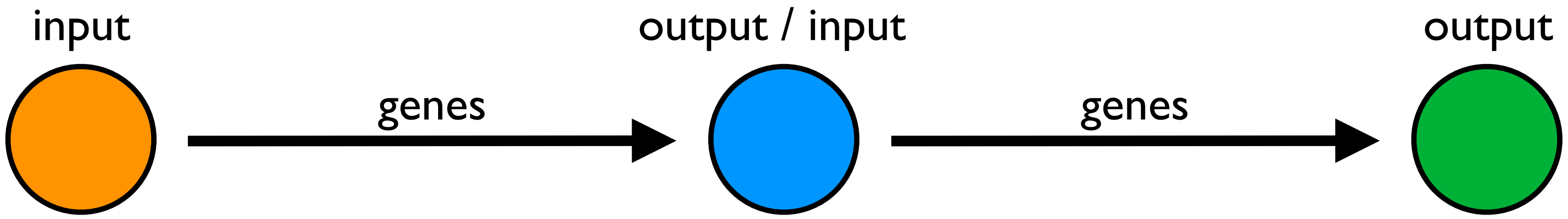
encode value with color

augmented with clustering

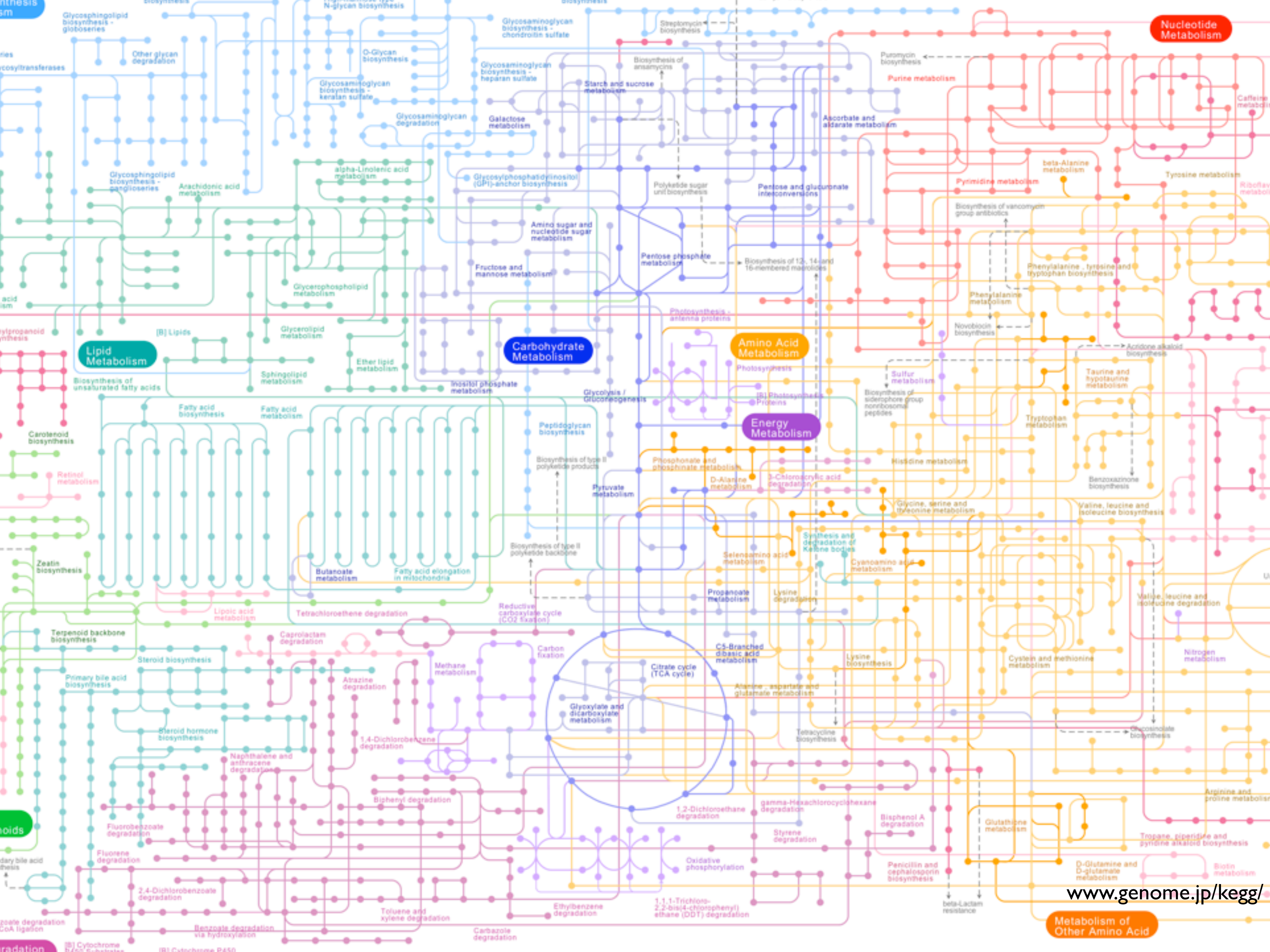


[Eisen98]

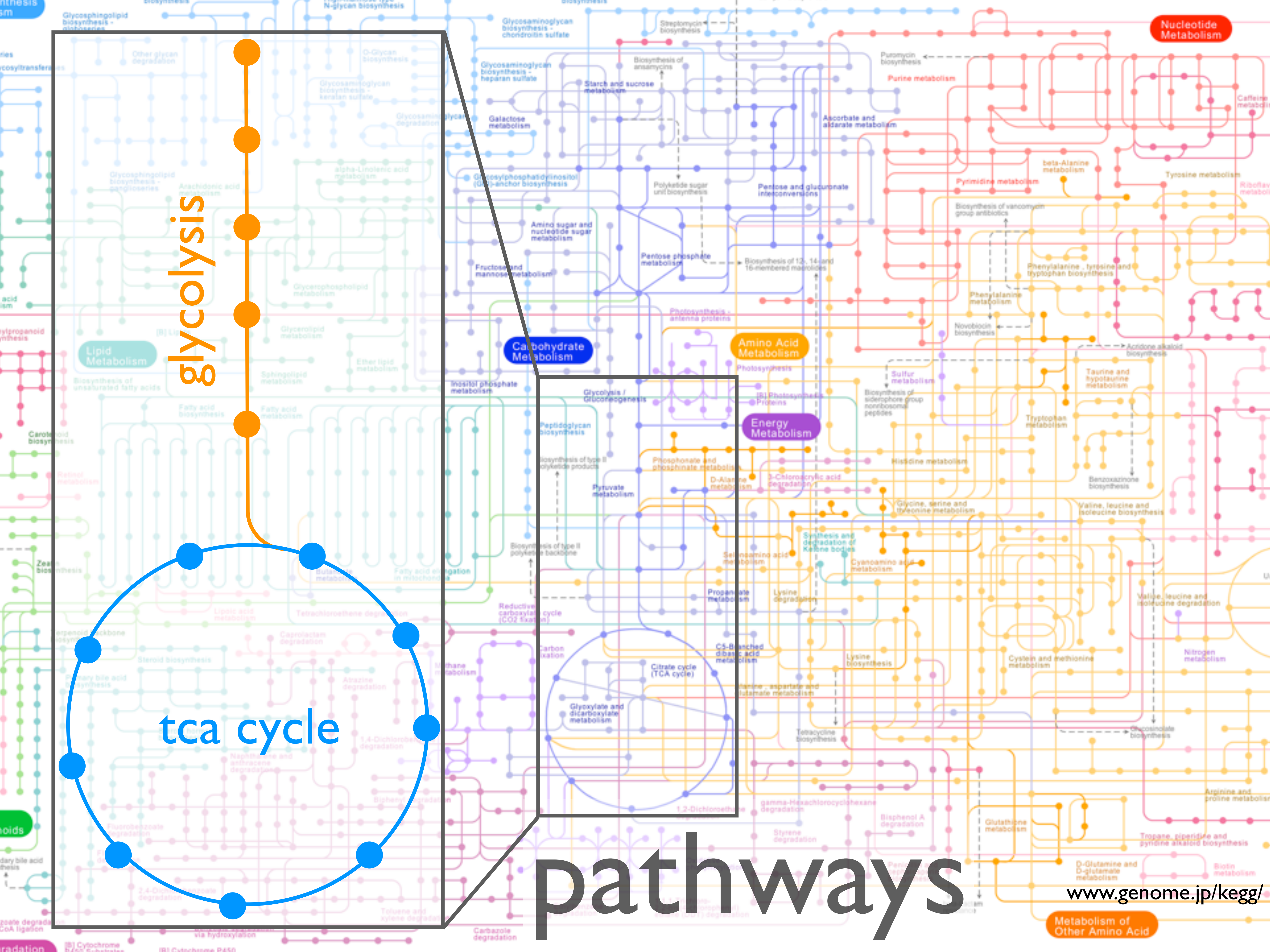
the functioning of a cell is controlled by many interrelated chemical reactions performed by genes


input
genes
output / input
genes
output

= cell function

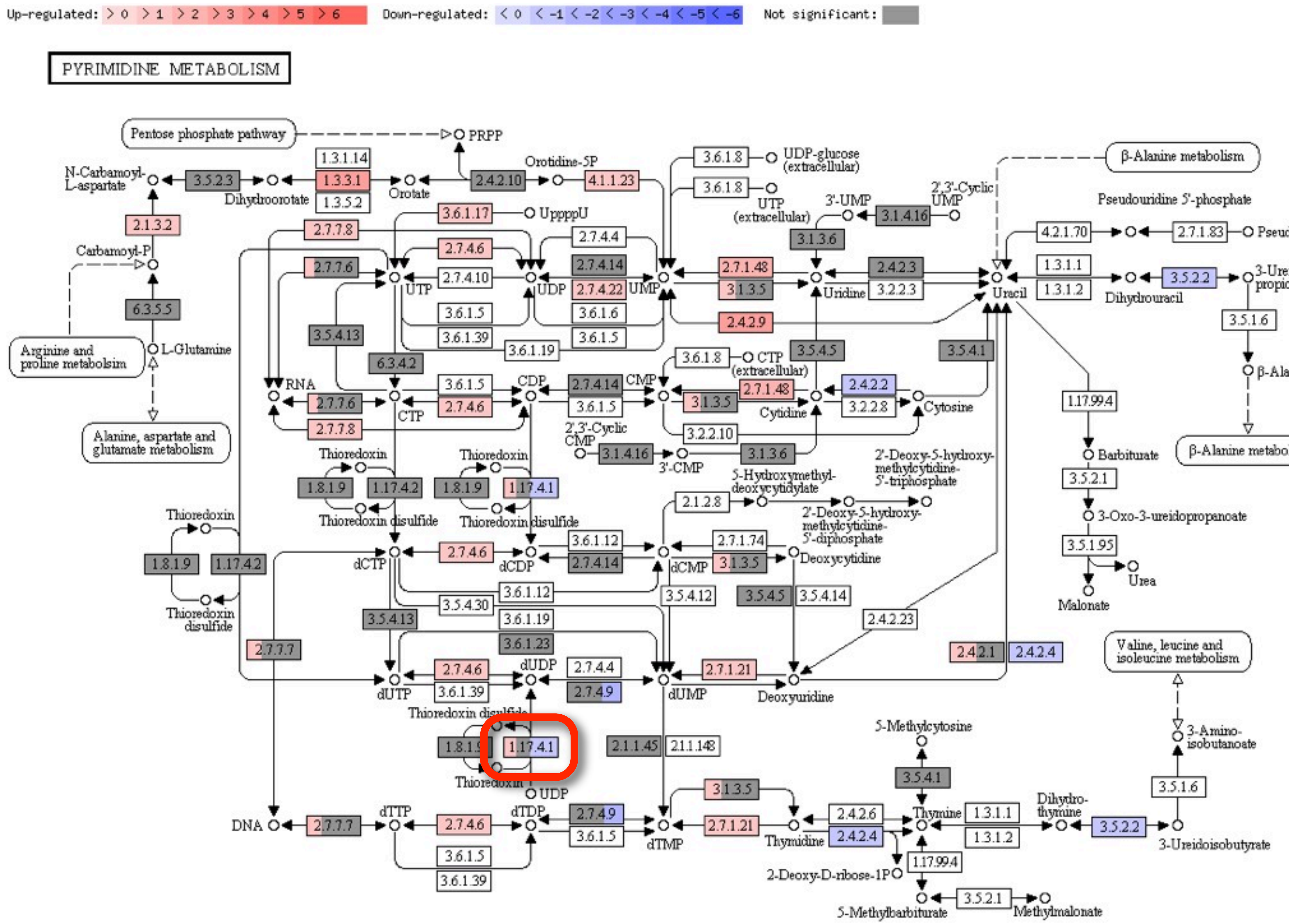


www.genome.jp/kegg/



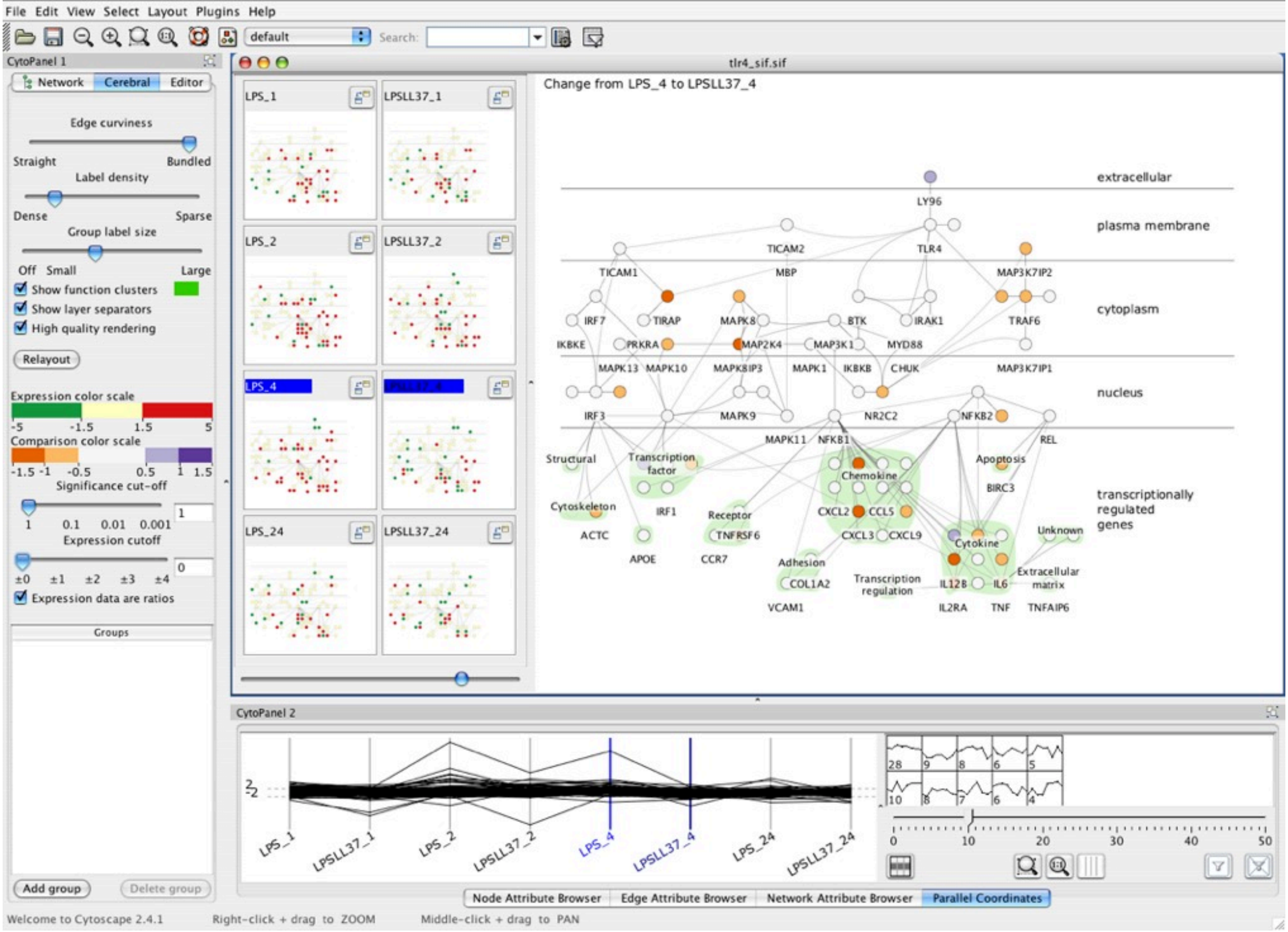
pathways

www.genome.jp/kegg/



[Kim09] [Bourqui09] [Barsky08] [Letunic08] [Junker06] **[Alm05]** [Mlecnik05] [Shannon03]

www.microbesonline.org



☑ Show function clusters 🟩

☑ Show layer separators

☑ High quality rendering

☑ Expression data are ratios

Significance cut-off: 1

Expression cutoff: 0

Change from LPS_4 to LPSLL37_4

Welcome to Cytoscape 2.4.1

[Kim09] [Bourqui09] **[Barsky08]** [Letunic08] [Junker06] [Alm05] [Mlecnik05] [Shannon03]

www.pathogenomics.ca/cerebral/

Rat, At injury site (T9), Mild Injury, 7 days after injury, Normalized to time 0, using Plier

Top 10 Pathways

Rank	Pathway
1	Cholesterol_Biosynthesis
2	Calcium_regulation_in_cardiac_cells
3	Electron_Transport_Chain
4	Proteasome_Degradation
5	Cell_cycle_KEGG
6	EGFR1_NetPath_4
7	GPCRDB_Other
8	Delta-Notch_NetPath_3
9	Monoamine_GPCRs
10	Regulation_of_Actin_Cytoskeleton_KEGG

17 genes in this pathway: Cholesterol_Biosynthesis

Probeset	Gene Name	Cell Type
rc_AI230255_at	Cyp51	neurons
AB016800_at	Dhcr7	astrocytes
D45252_s_at	Lss	neurons
M89945mRNA_at	Fdps	neurons
M89945mRNA_g_at	Fdps	neurons
rc_AI172293_at	Sc4mol	astrocytes
rc_AI177004_s_at	Hmgcs1	neurons
M29472_at	Mvk	neurons
U31352_at	Lss	neurons
X52625_at	Hmgcs1	neurons
D37920_at	Sqle	
AF003835_at	Idi1	neurons
rc_H33426_at	Fdft1	astrocytes

bioinformatics.cnmcresearch.org/GeneShelf

[Kim09] [Bourqui09] [Barsky08] [Letunic08] [Junker06] [Alm05] [Mlecnik05] [Shannon03]

functional genomics

how do genes work together to perform different functions in a cell?

comparative functional genomics

how do the gene interactions vary across different species?

collaborators: Regev Lab at the Broad Institute

biology: metabolism in yeast

data: multiple genes
multiple time points
multiple related species
multiple pathways

problem: *existing tools can only look at a subset of this data*

comparative functional genomics

how do the gene interactions vary across different species?

contributions

Pathline

first interactive tool for visualizing multiple genes, time points, species, and pathways

linearized pathway representation

for comparing quantitative data along a pathway

curvemap

visual encoding of temporal gene expression

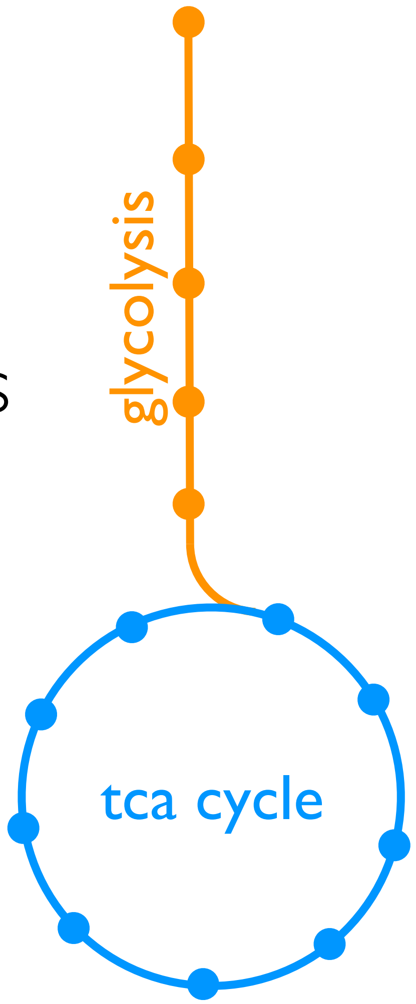
validation

case studies describing efficiency gains and new biological findings

data & tasks

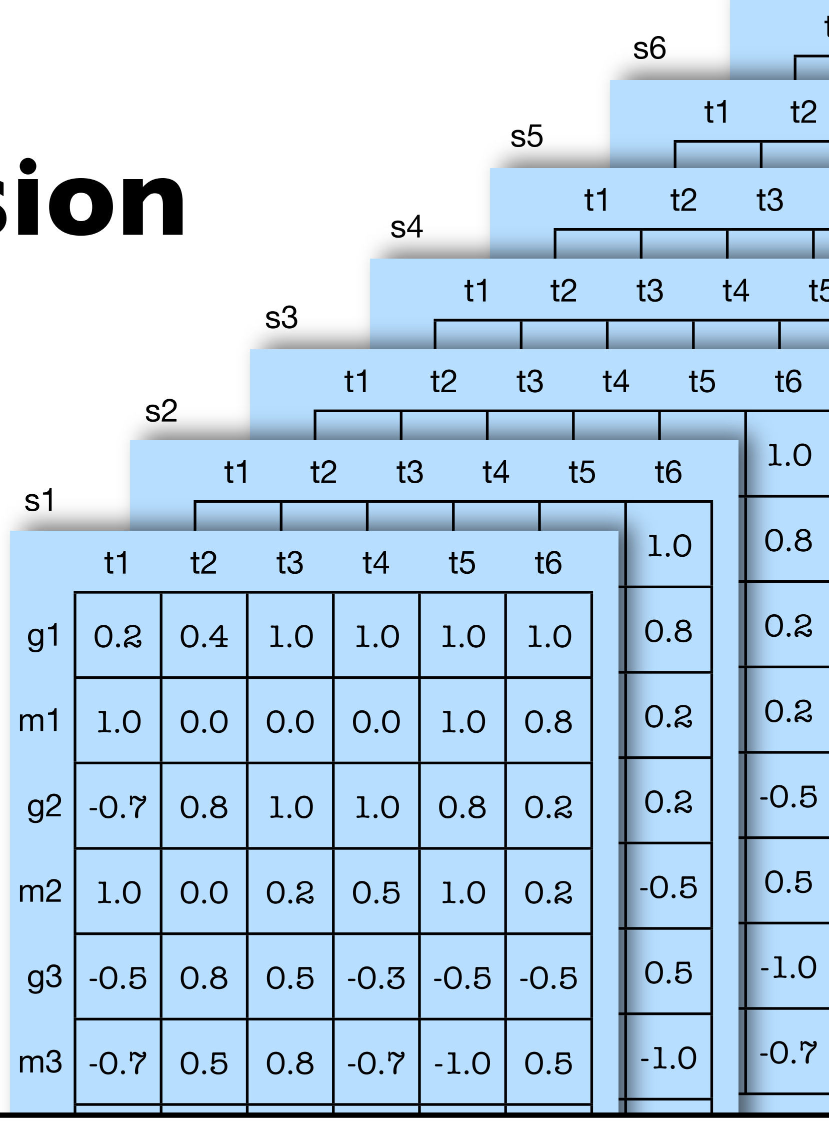
metabolic pathways

- 10 to 50 pathways of interest
- inputs/outputs called metabolites



gene expression

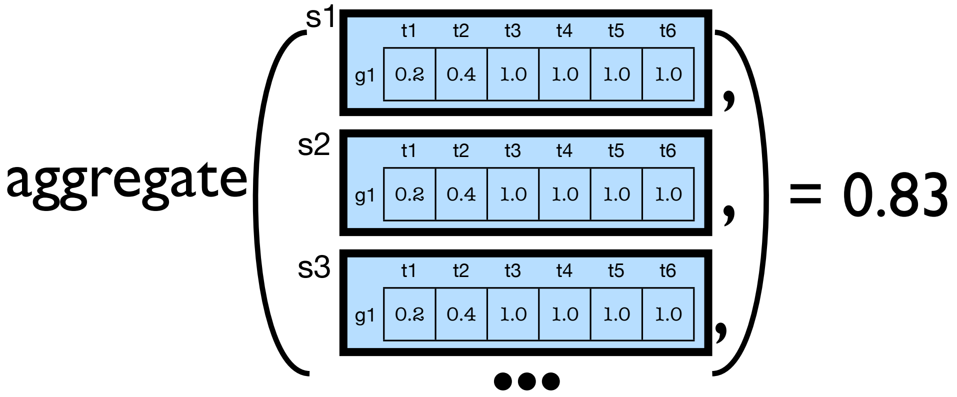
- 6000 genes and 140 metabolites
- 6 time points
- 14 species of yeast



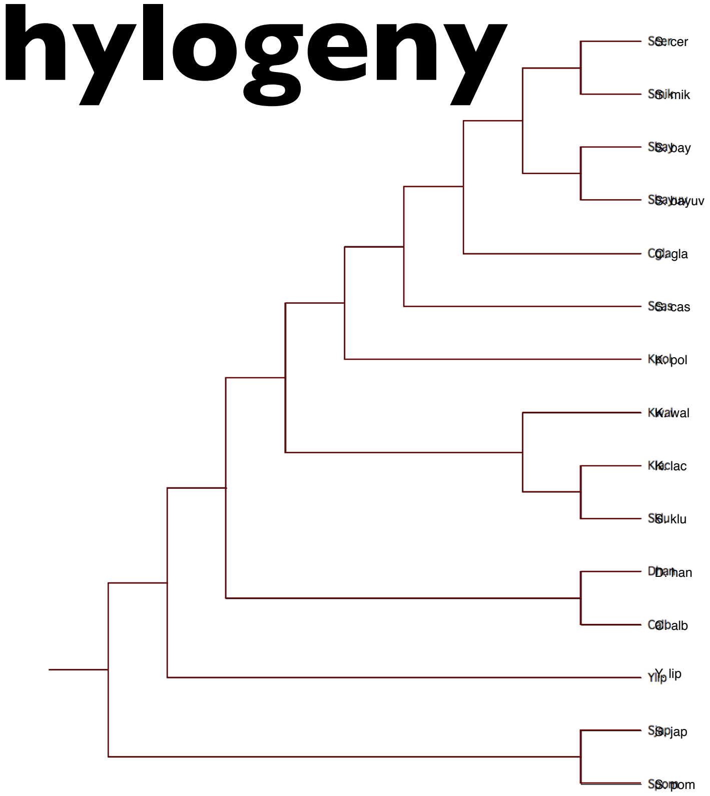
	t1	t2	t3	t4	t5	t6
g1	0.2	0.4	1.0	1.0	1.0	1.0
m1	1.0	0.0	0.0	0.0	1.0	0.8
g2	-0.7	0.8	1.0	1.0	0.8	0.2
m2	1.0	0.0	0.2	0.5	1.0	0.2
g3	-0.5	0.8	0.5	-0.3	-0.5	-0.5
m3	-0.7	0.5	0.8	-0.7	-1.0	0.5

similarity scores

- aggregate time series for a gene/metabolite over species
- similarity of expression across species
- aggregate: Pearson, Spearman, others



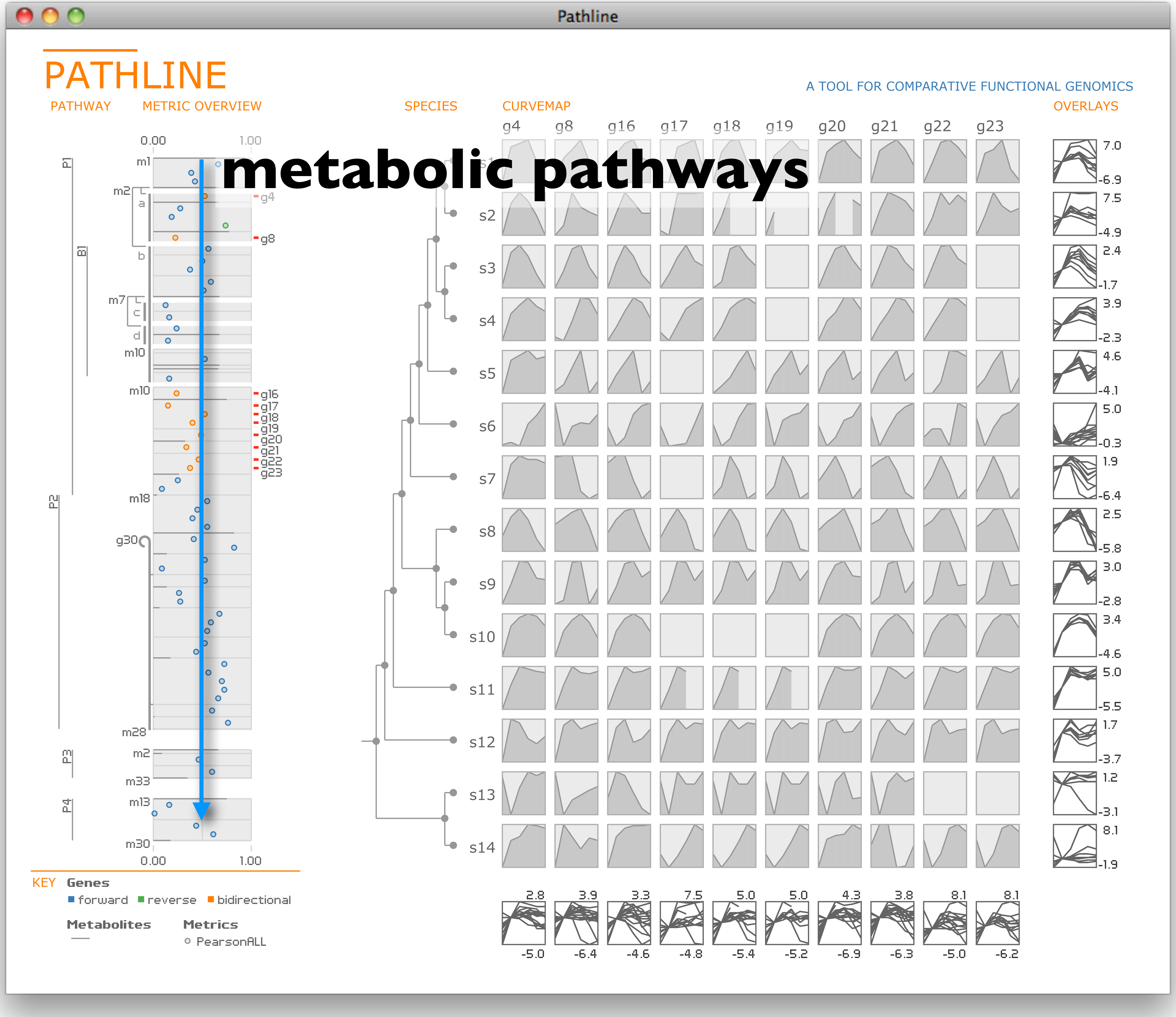
phylogeny



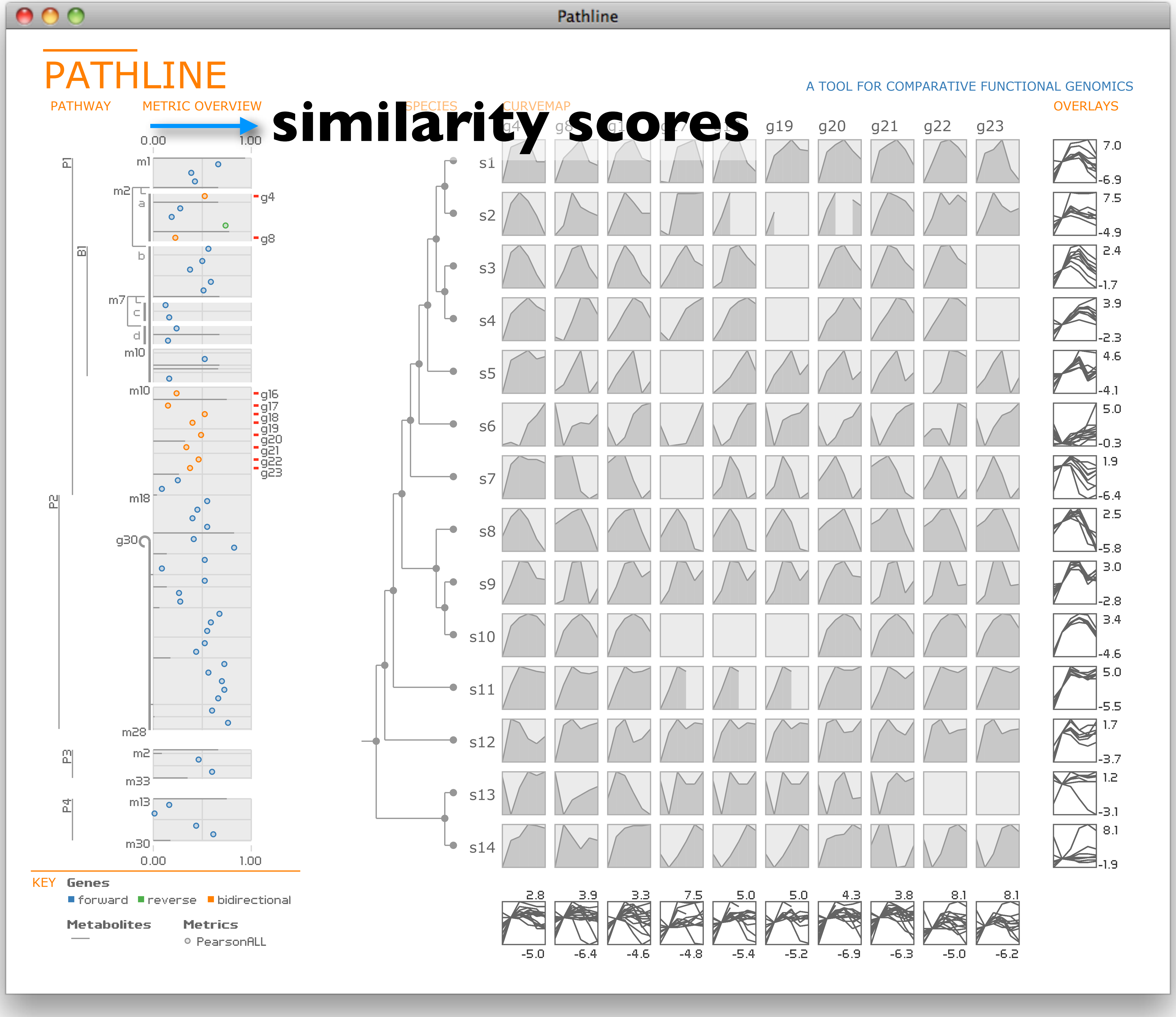
tasks

1. study expression data as a time series
    - *look for peaks, valleys, time shifts*
2. detailed comparison of a limited number of time series
    - *filter using pathways*
    - *filter again using genes or species*
3. comparison of similarity scores of genes along a pathway(s)
4. comparison of multiple similarity scores

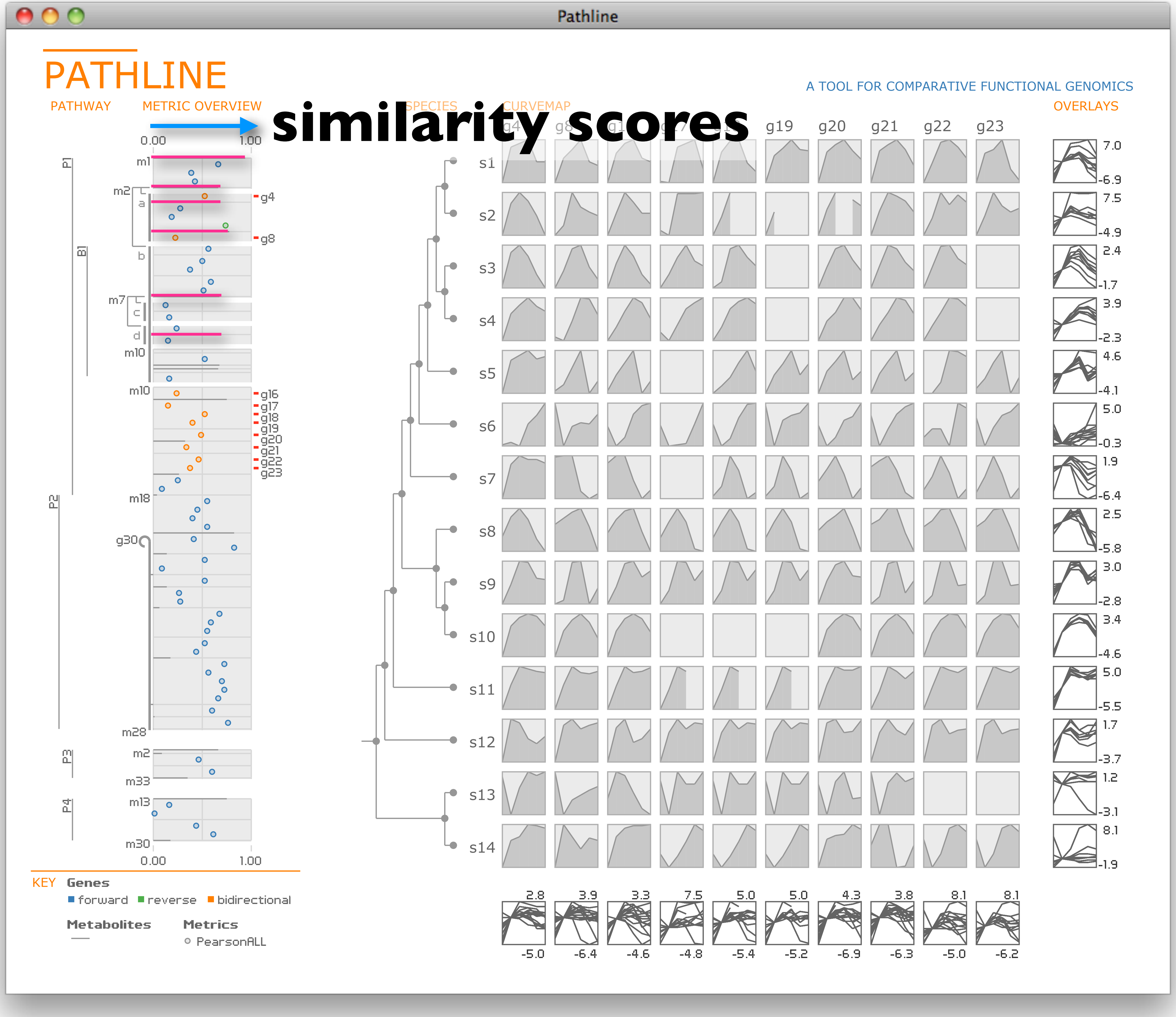
Pathline



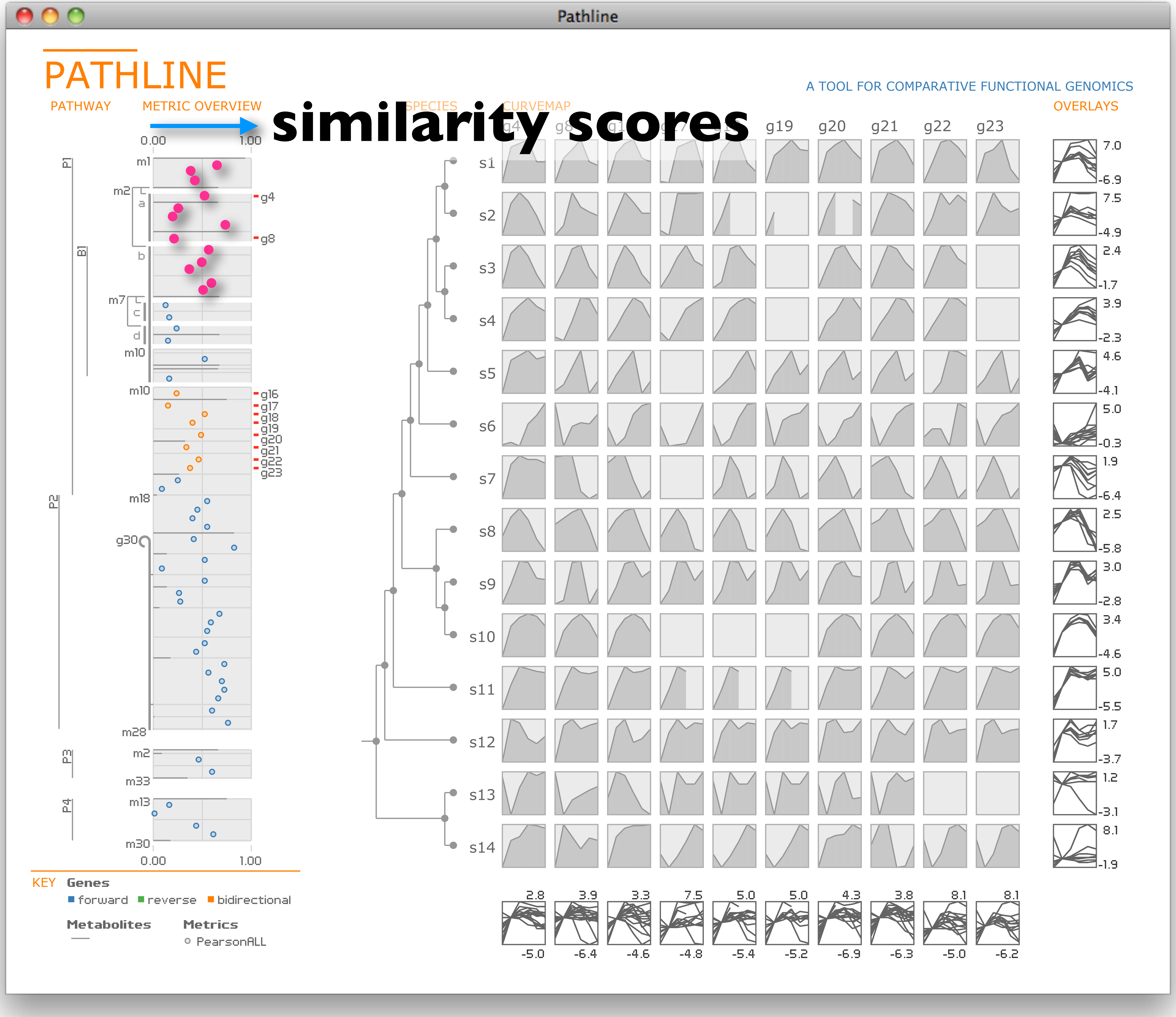
metabolic pathways



similarity scores

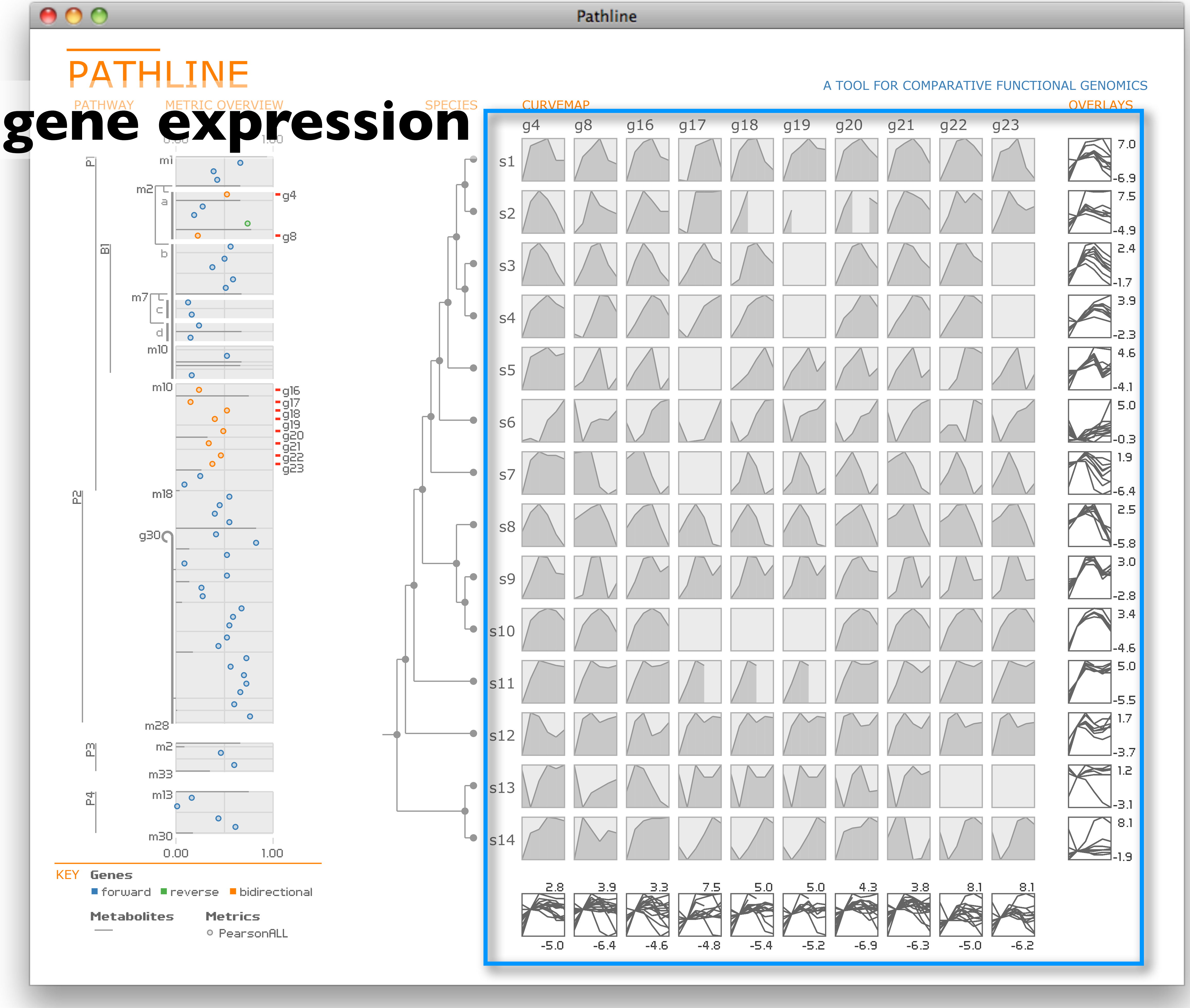


similarity scores

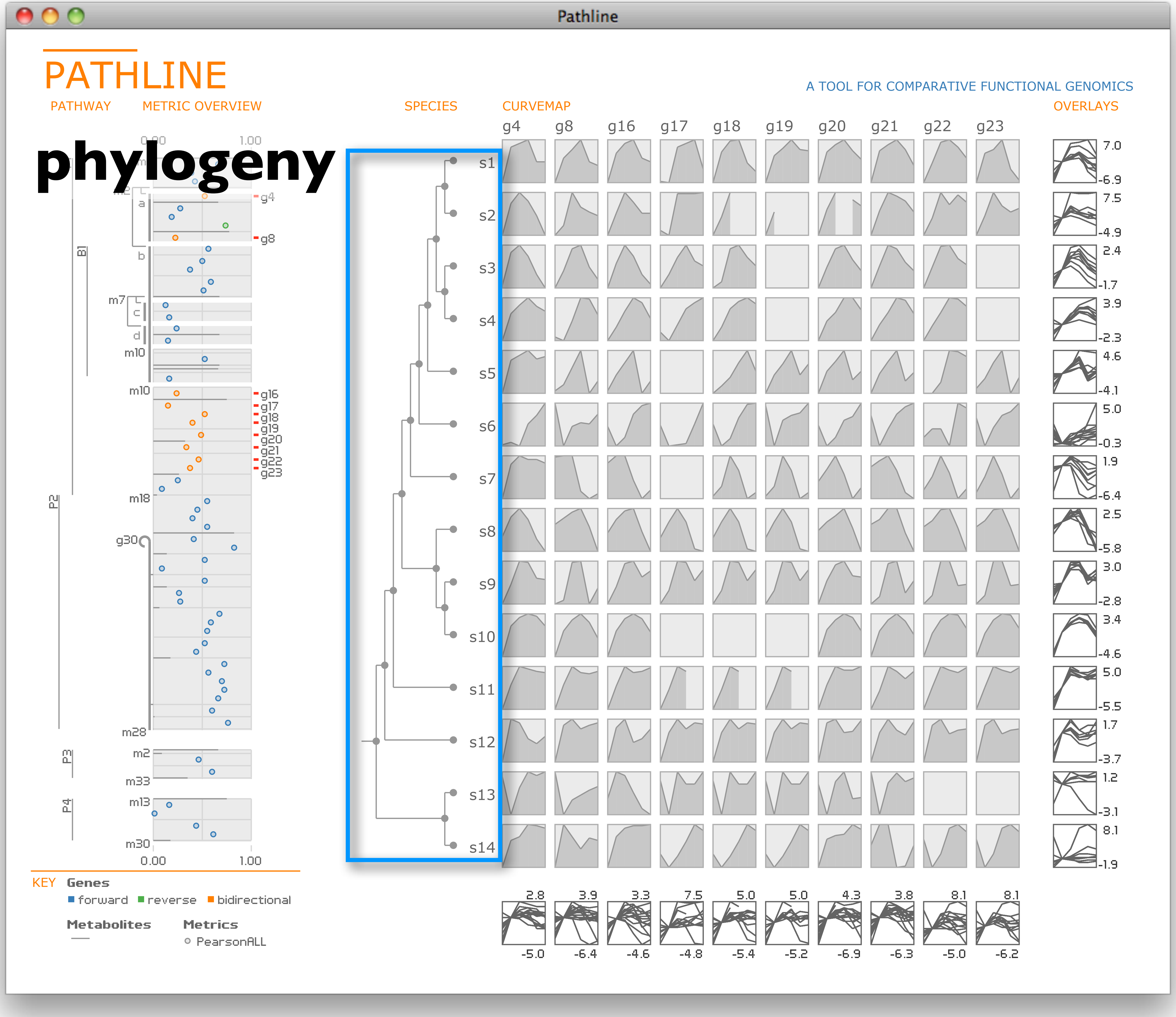


similarity scores

gene expression



phylogeny


Pathline
PATHLINE
A TOOL FOR COMPARATIVE FUNCTIONAL GENOMICS
PATHWAY
METRIC OVERVIEW
SPECIES
CURVEMAP
OVERLAYS
0.00
1.00
g4
g8
g16
g17
g18
g19
g20
g21
g22
g23
g30
B1
P2
P3
P4
a
b
c
d
m2
m7
m10
m13
m18
m28
m30
m33
s1
s2
s3
s4
s5
s6
s7
s8
s9
s10
s11
s12
s13
s14
7.0
-6.9
7.5
-4.9
2.4
-1.7
3.9
-2.3
4.6
-4.1
5.0
-0.3
1.9
-6.4
2.5
-5.8
3.0
-2.8
3.4
-4.6
5.0
-5.5
1.7
-3.7
1.2
-3.1
8.1
-1.9
2.8
3.9
3.3
7.5
5.0
5.0
4.3
3.8
8.1
8.1
-5.0
-6.4
-4.6
-4.8
-5.4
-5.2
-6.9
-6.3
-5.0
-6.2
KEY
Genes
forward
reverse
bidirectional
Metabolites
Metrics
PearsonALL

Pathline

design decisions

encode quantitative values with spatial position

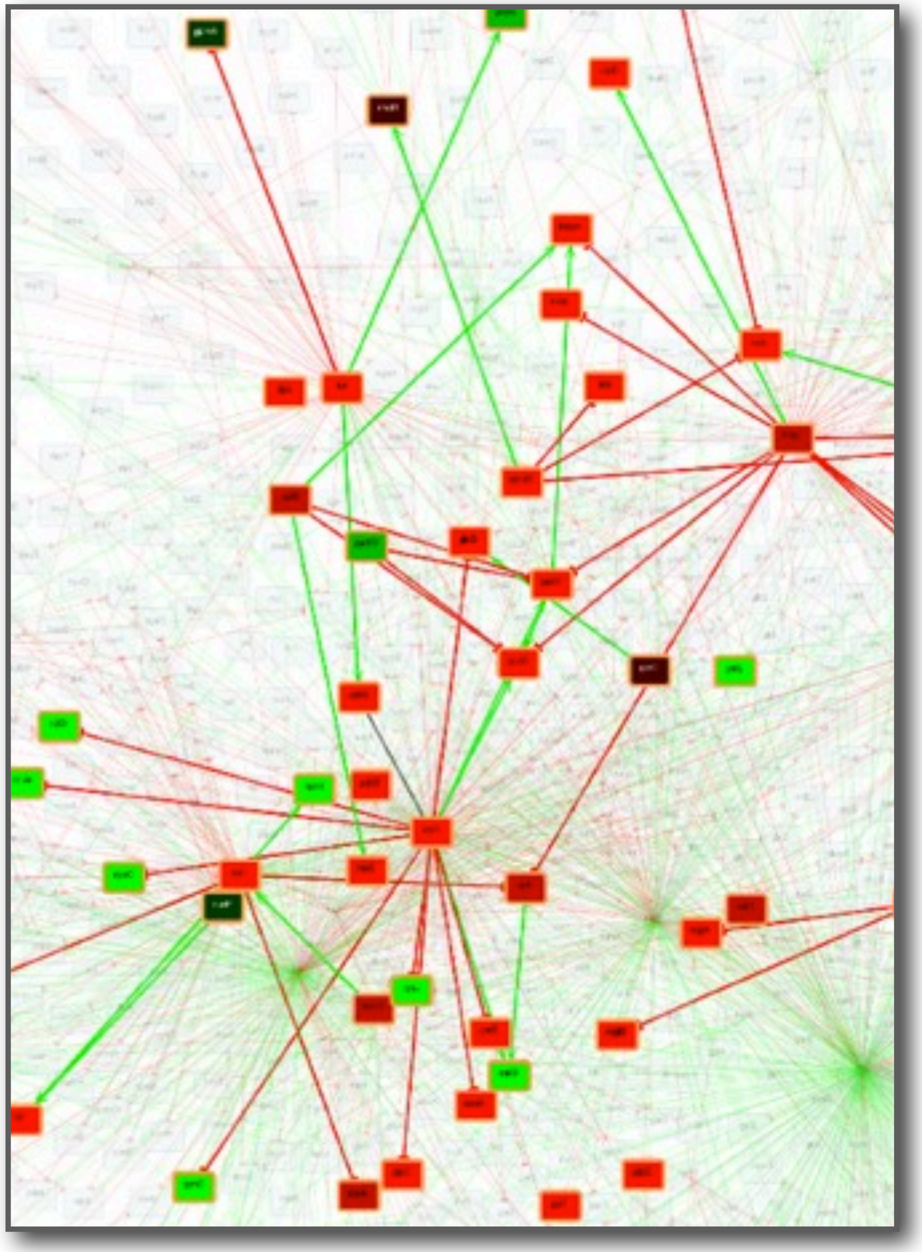
[Cleveland84] [Lam07]

encode quantitative values with spatial position

[Cleveland84] [Lam07]

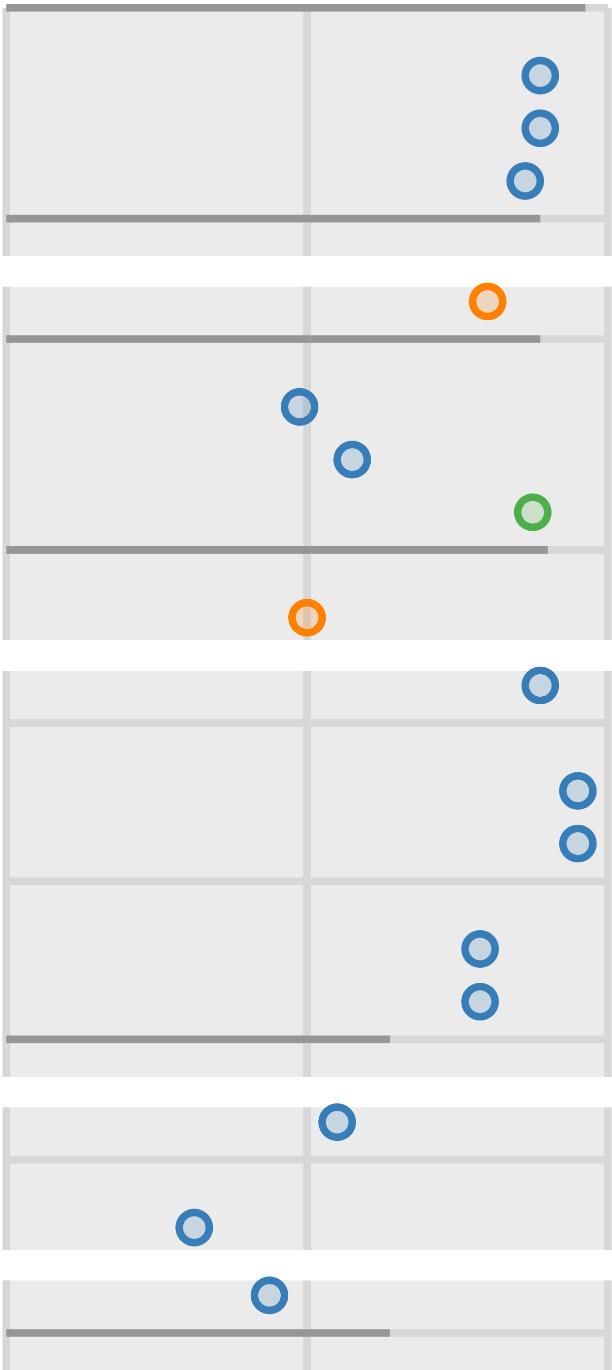
instead of a . .

topological layout



www.win.tue.nl/~mwestenb/genevis/

linearized pathway

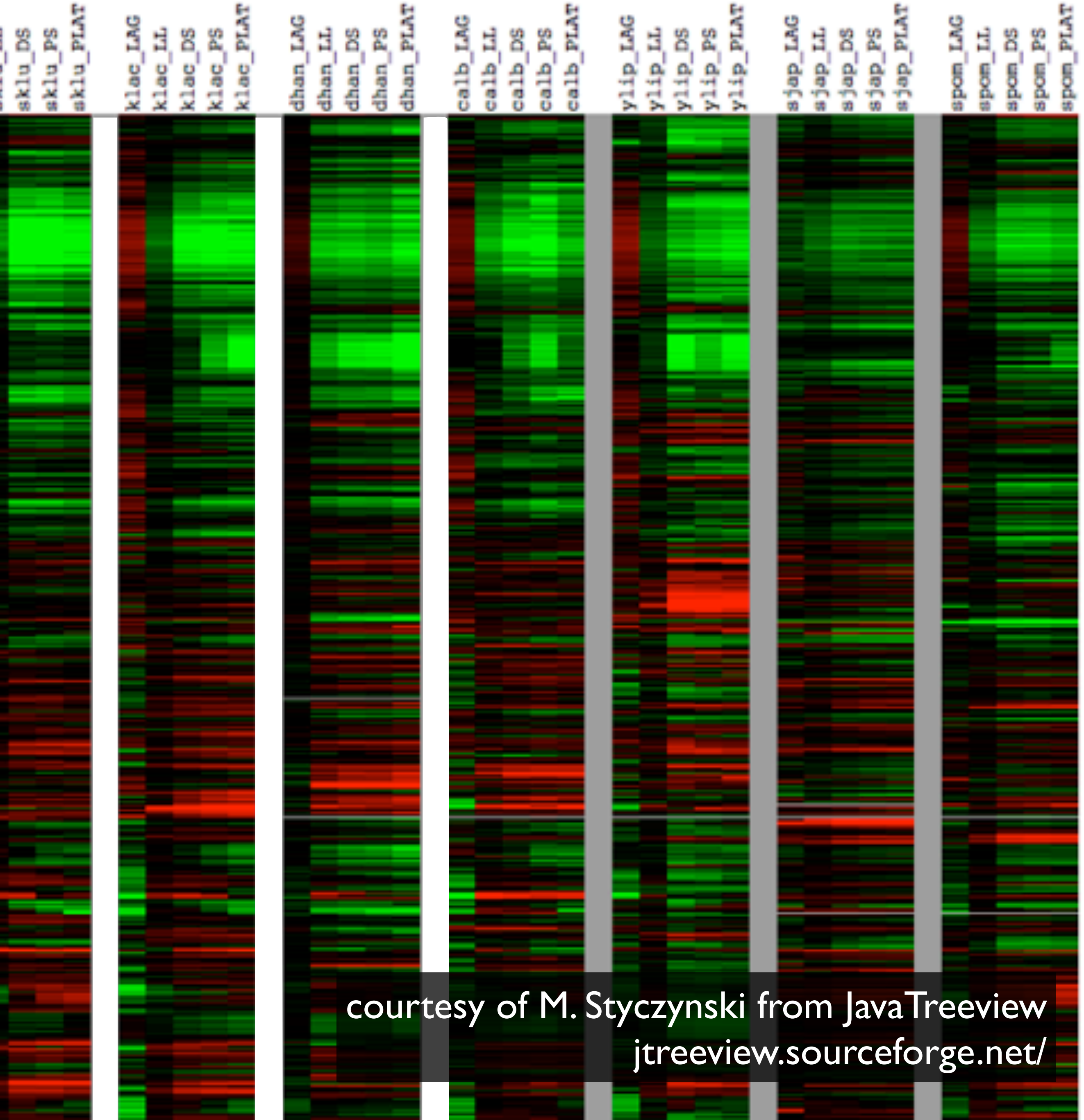


encode quantitative values with spatial position

[Cleveland84] [Lam07]

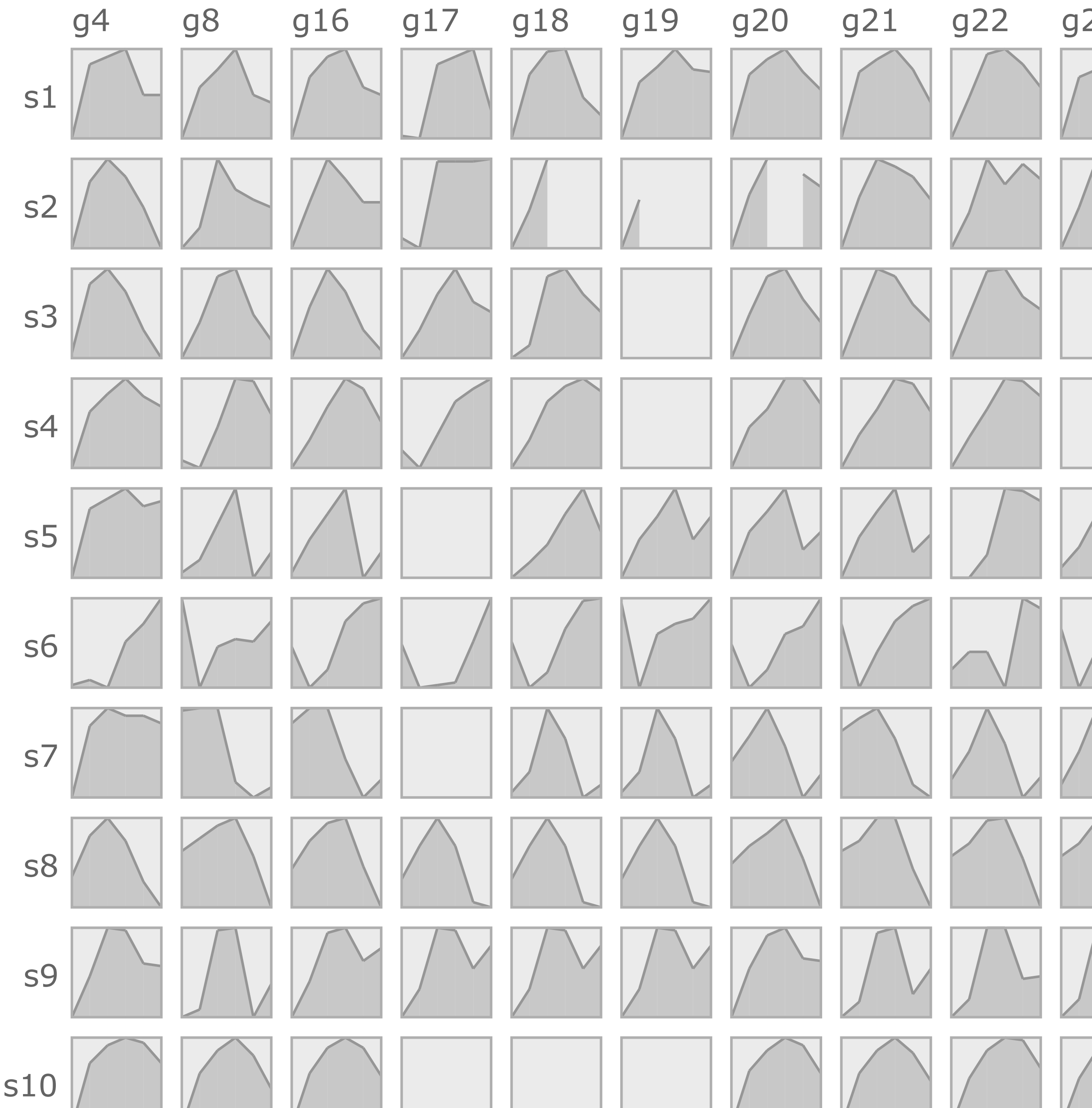
instead of a . .

conditioned heatmap



courtesy of M. Styczynski from Java Treeview
jtreeview.sourceforge.net/

curvemap

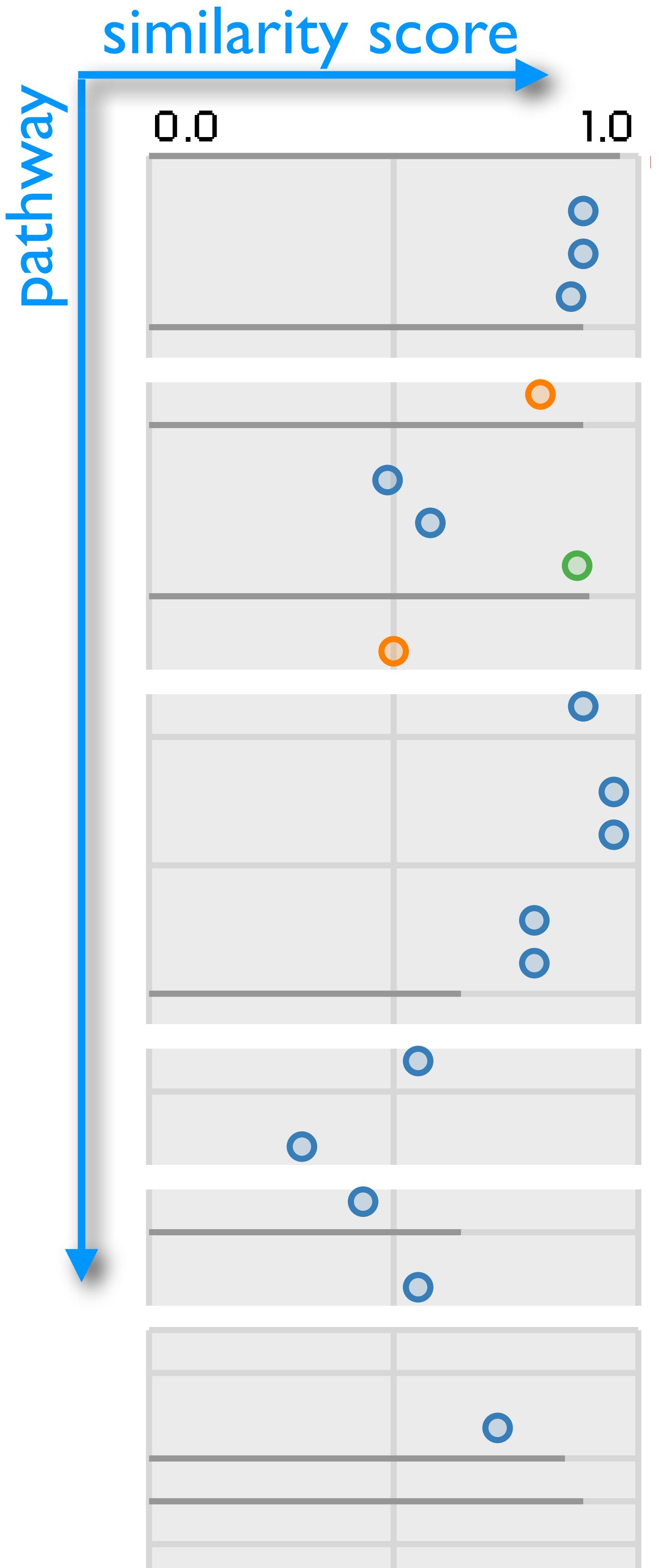


Pathline

linearized pathway representation

linearized pathway representation

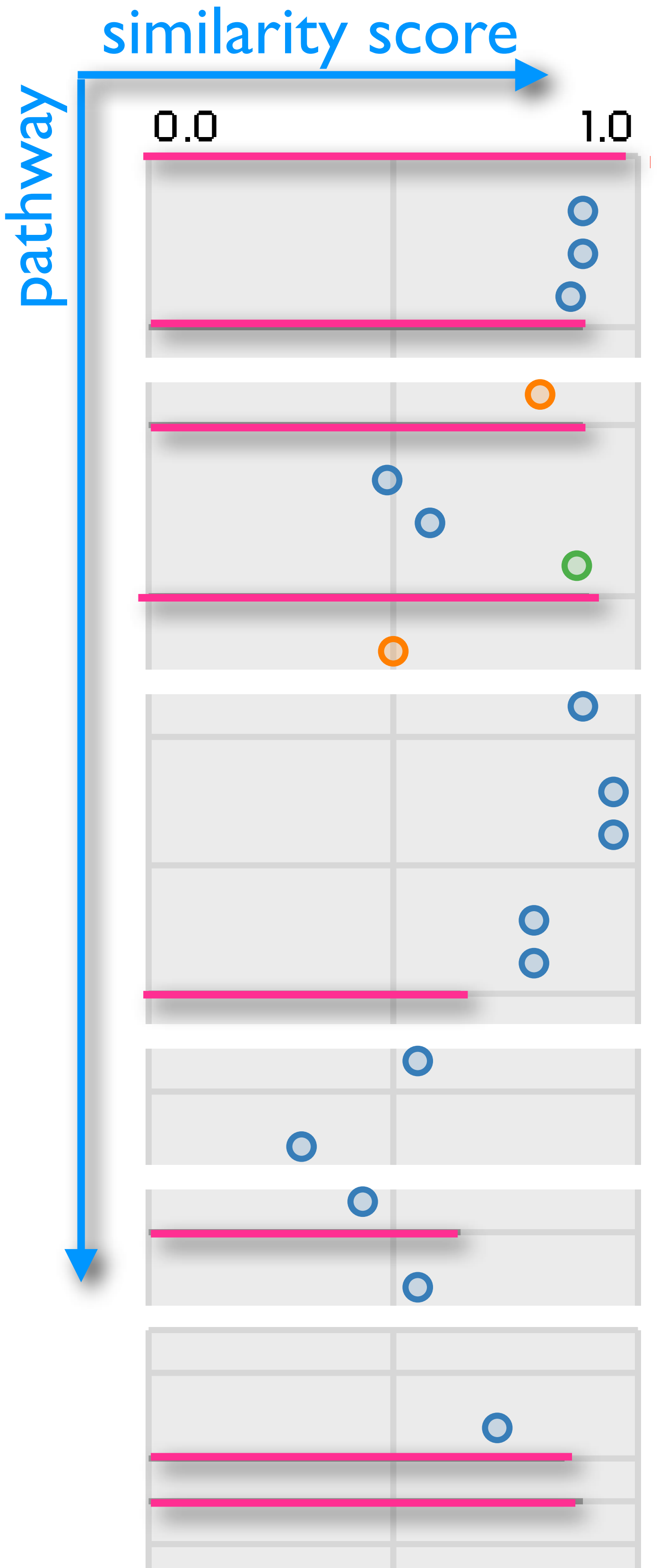
common axes to compare similarity scores



linearized pathway representation

common axes to compare similarity scores

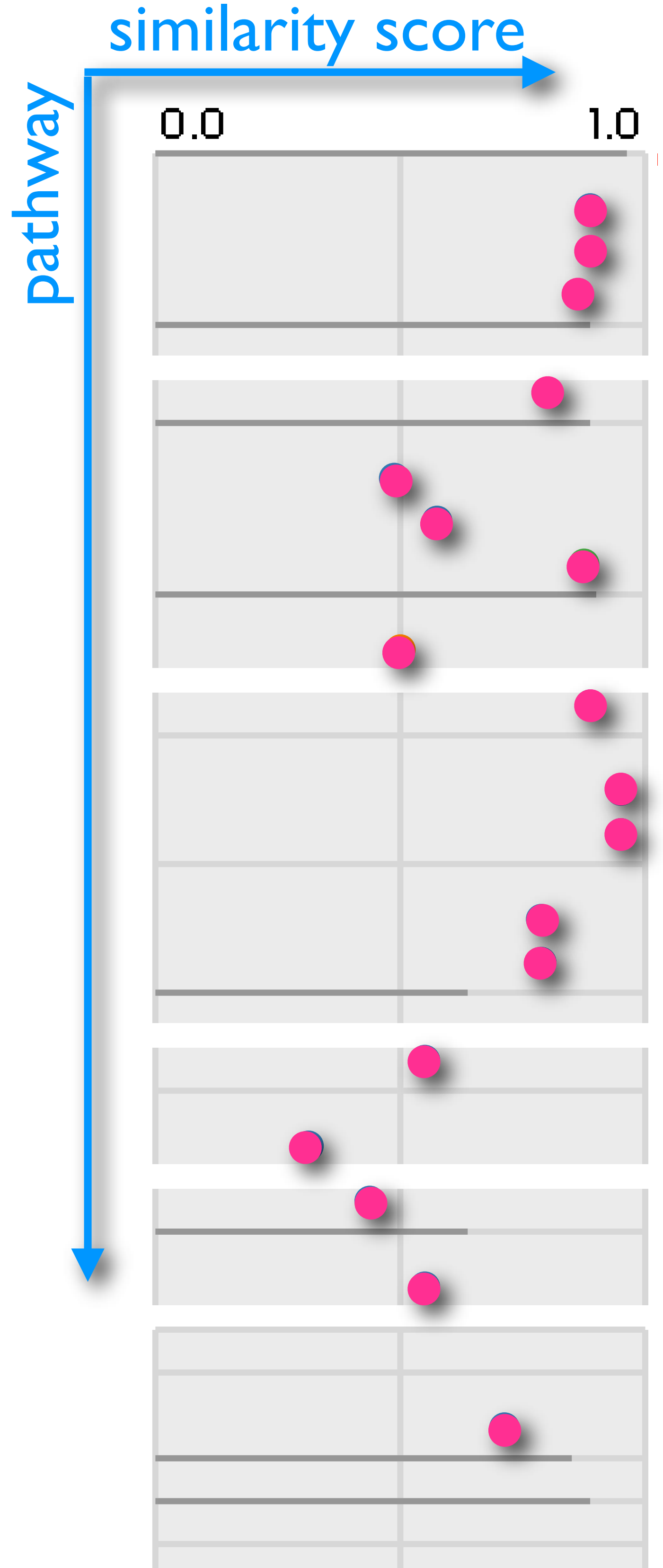
- *bars and circles*



linearized pathway representation

common axes to compare similarity scores

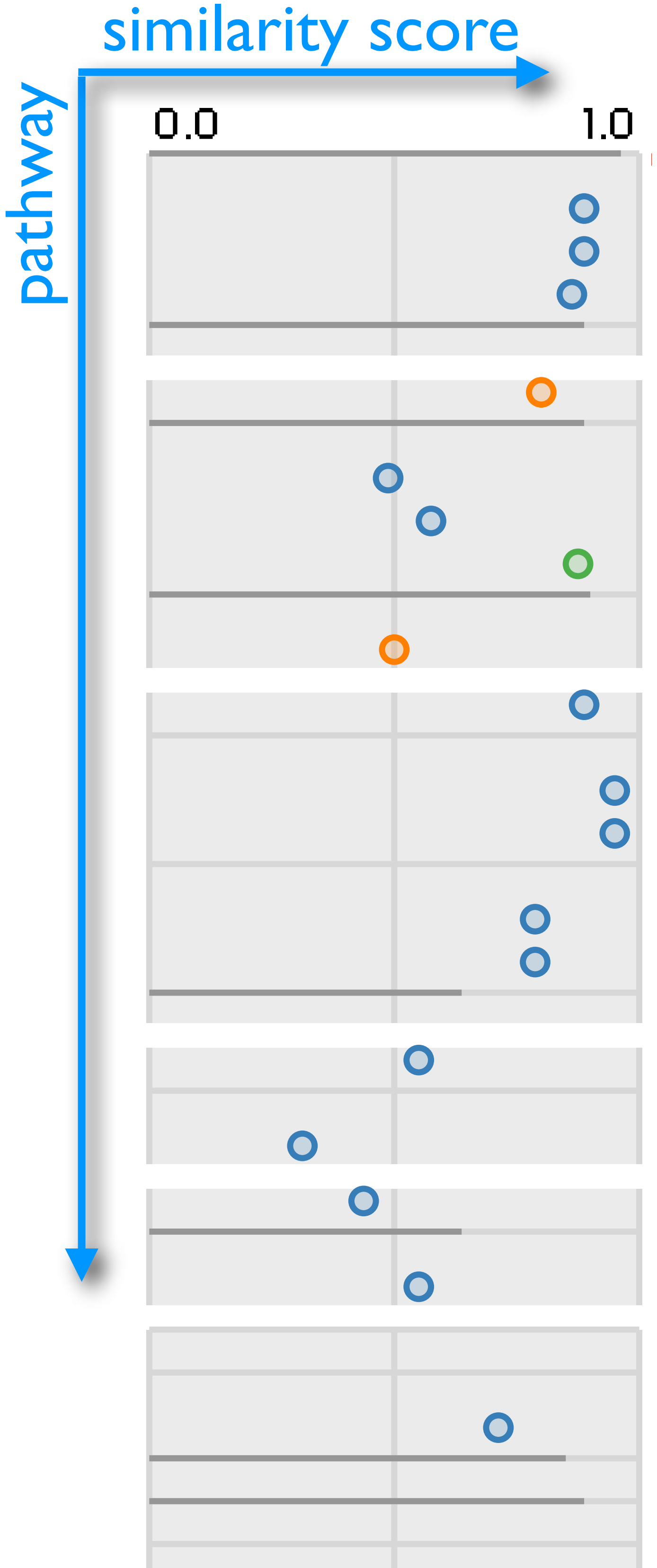
- *bars and circles*



linearized pathway representation

common axes to compare similarity scores

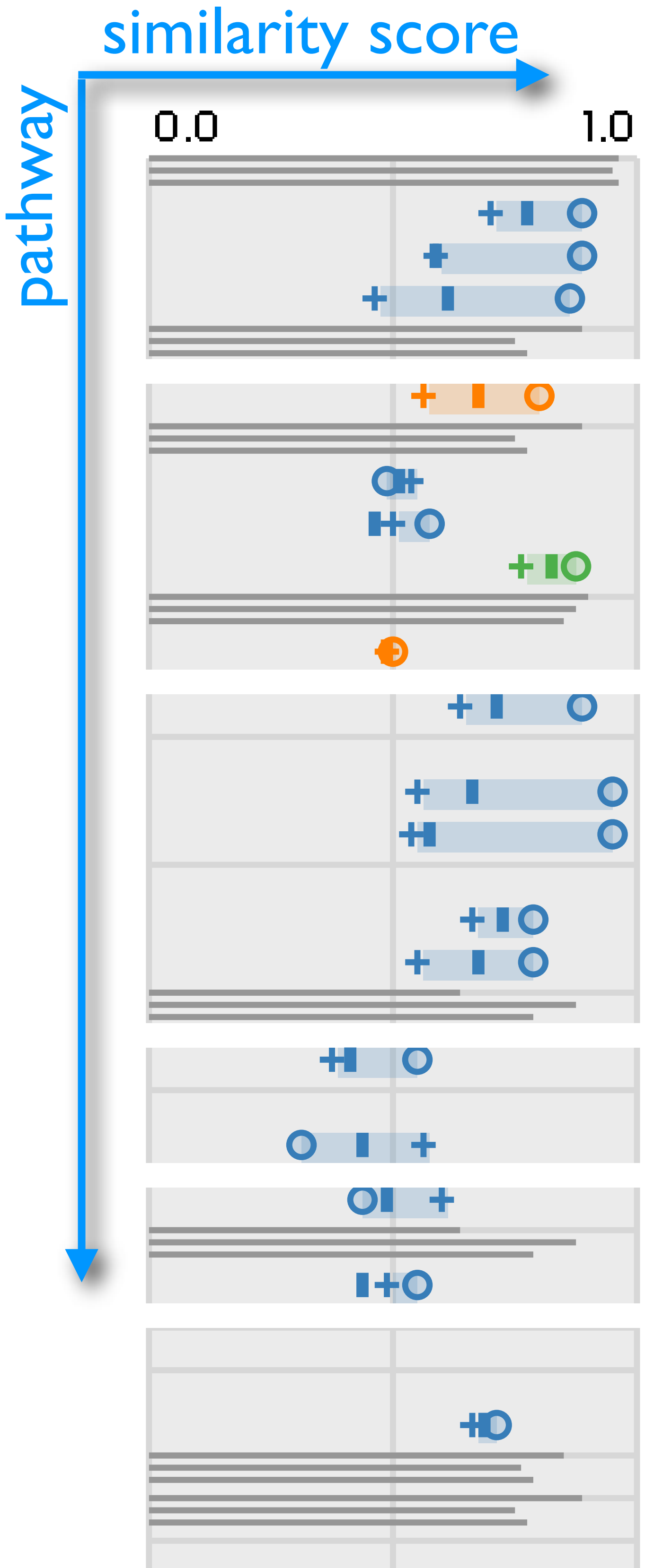
- *bars and circles*
  - *visual layer for attenuation*
  - *color-code gene direction*



linearized pathway representation

common axes to compare similarity scores

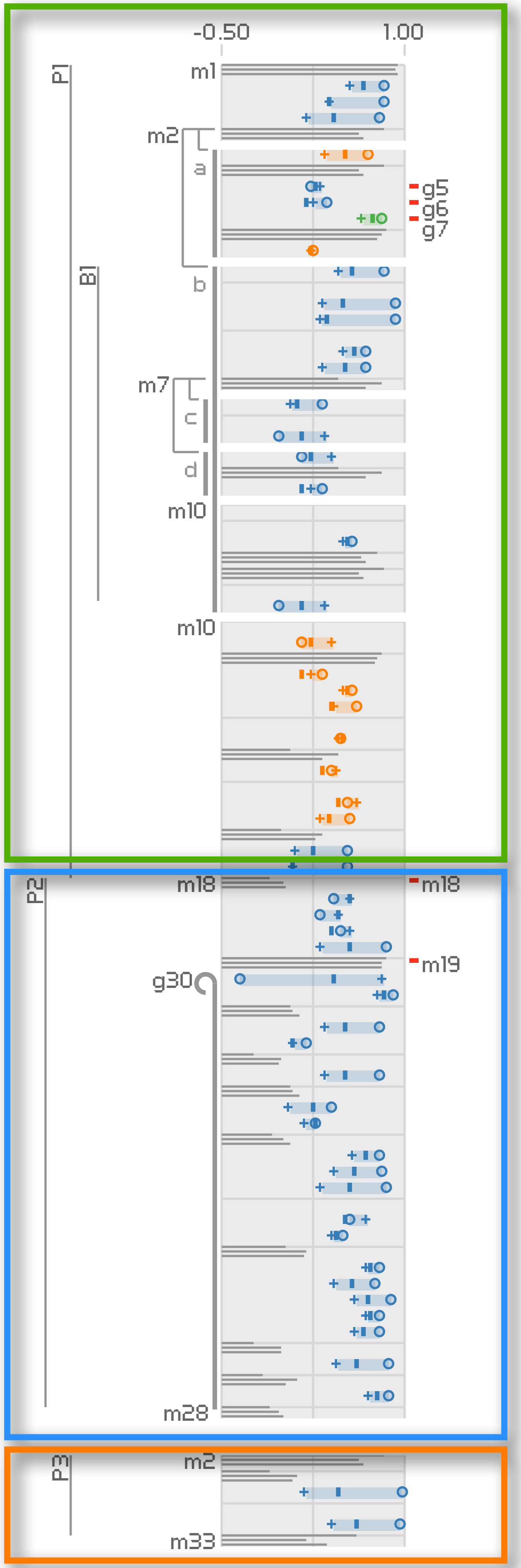
- *bars and circles*
    - *visual layer for attenuation*
    - *color-code gene direction*
- *multiple similarity scores*



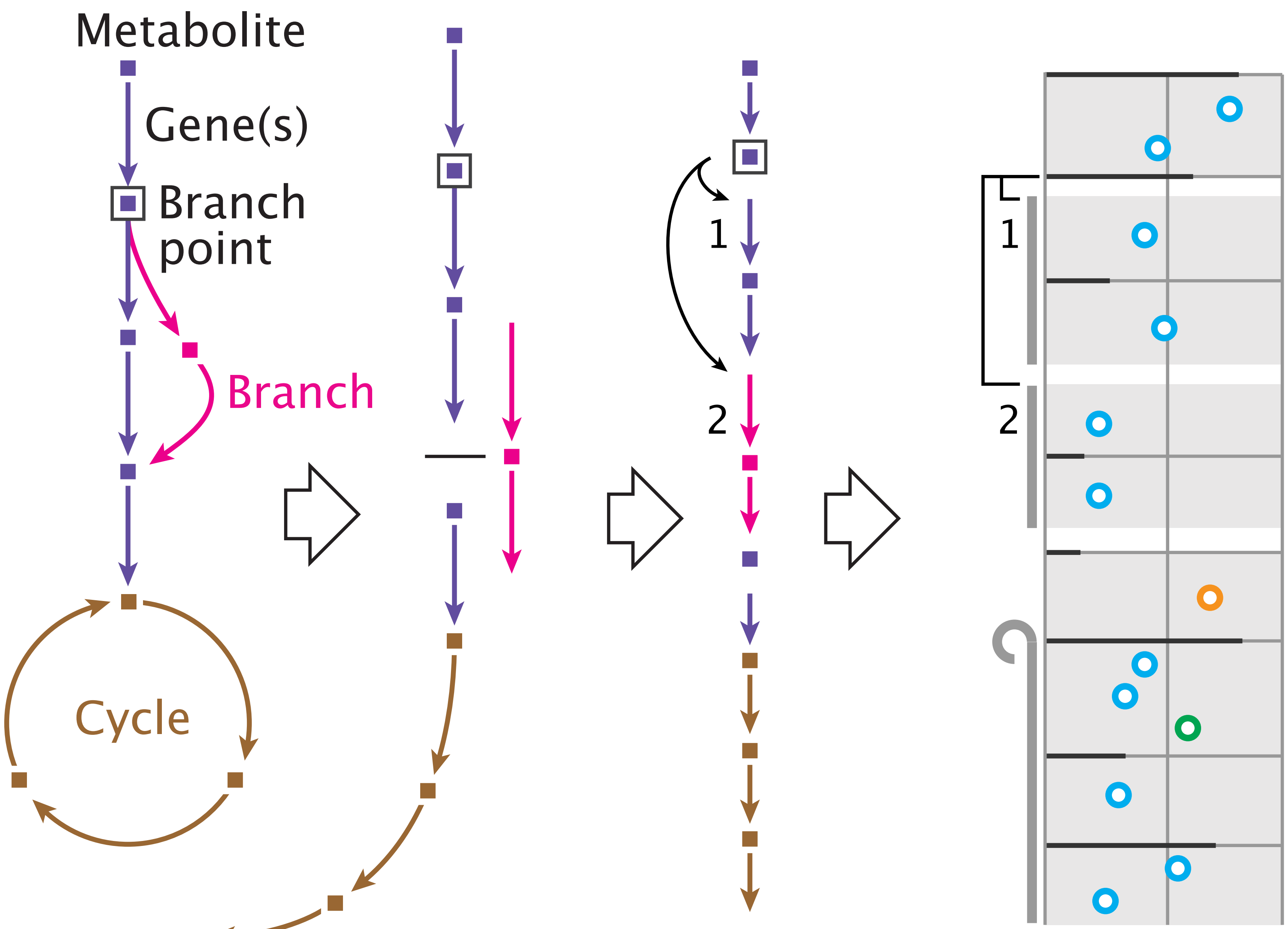
linearized pathway representation

common axes to compare similarity scores

- *bars and circles*
    - *visual layer for attenuation*
    - *color-code gene direction*
- *multiple similarity scores*
- *multiple pathways*



unroll from pathway to ordered list of nodes



unroll and cut

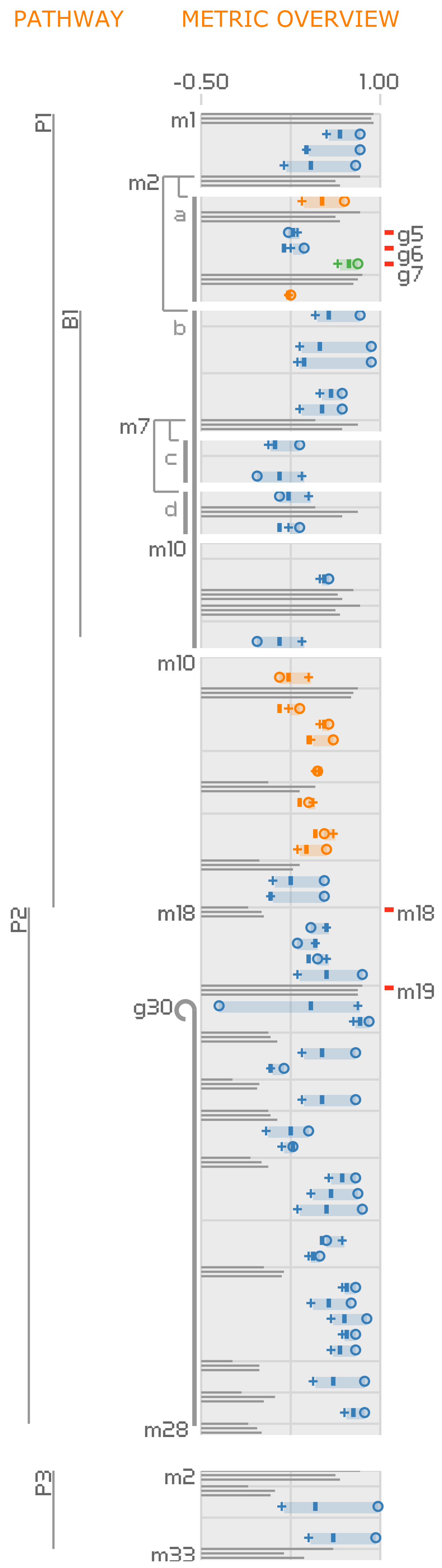
reinsert

stylized marks and shared coordinate frame

linearized pathway representation

putting it together . . .

- *use spatial position for similarity scores instead of topology*
- *topology is secondary*



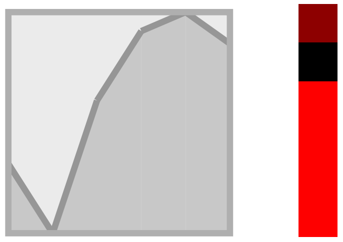
Pathline

curvemap

curvemap

inspired by heatmaps

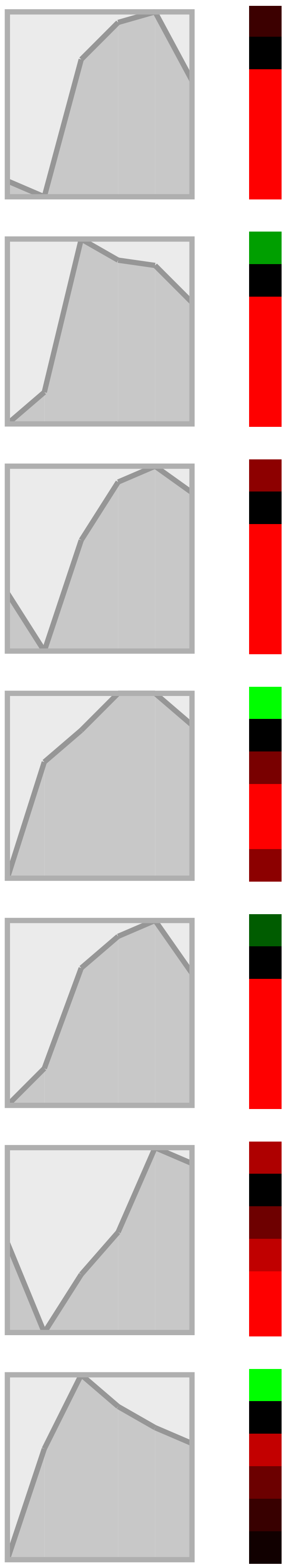
- *base visual unit is a curve*



curvemap

inspired by heatmaps

- *base visual unit is a curve*
- *filled, framed line charts to enhance shape perception*

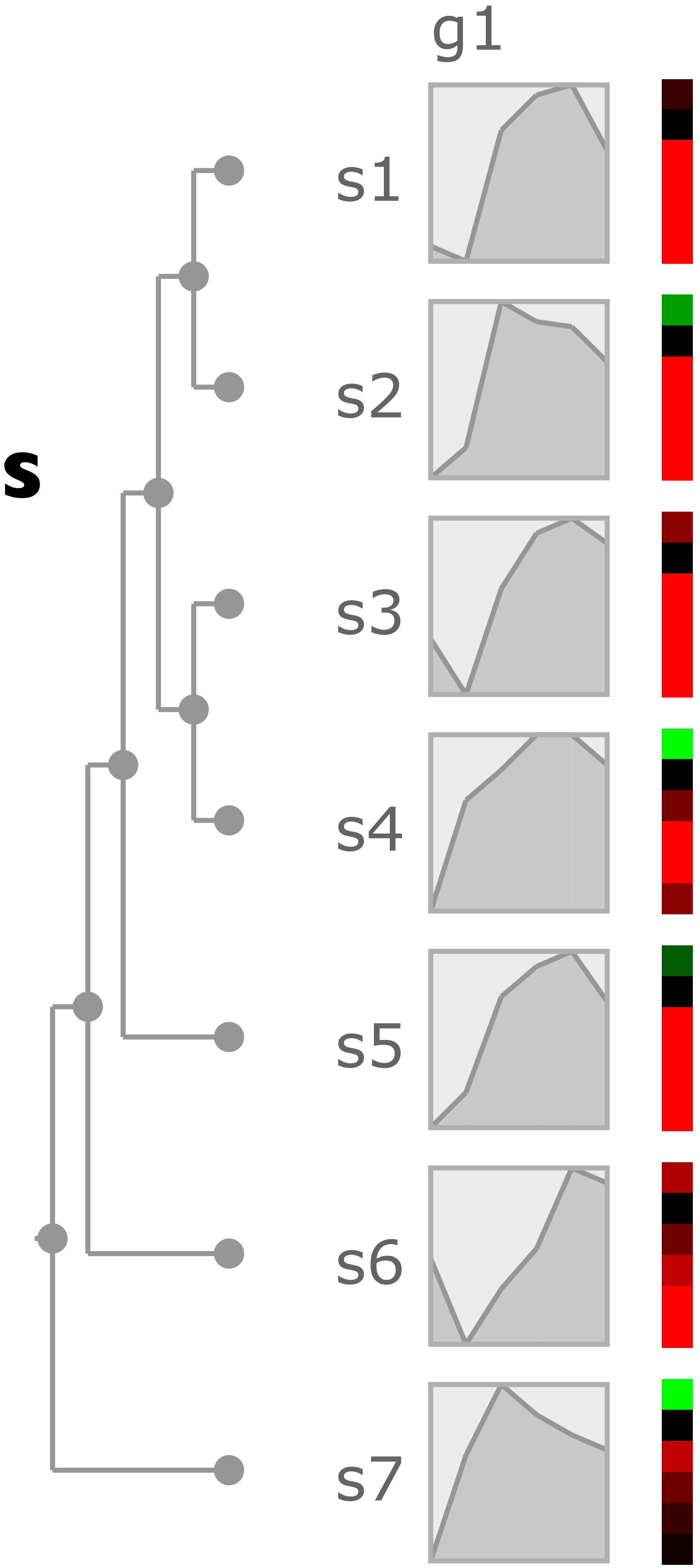


curvemap

inspired by heatmaps

  - *base visual unit is a curve*
  - *filled, framed line charts to enhance shape perception*

- rows are species
- columns are genes/metabolites

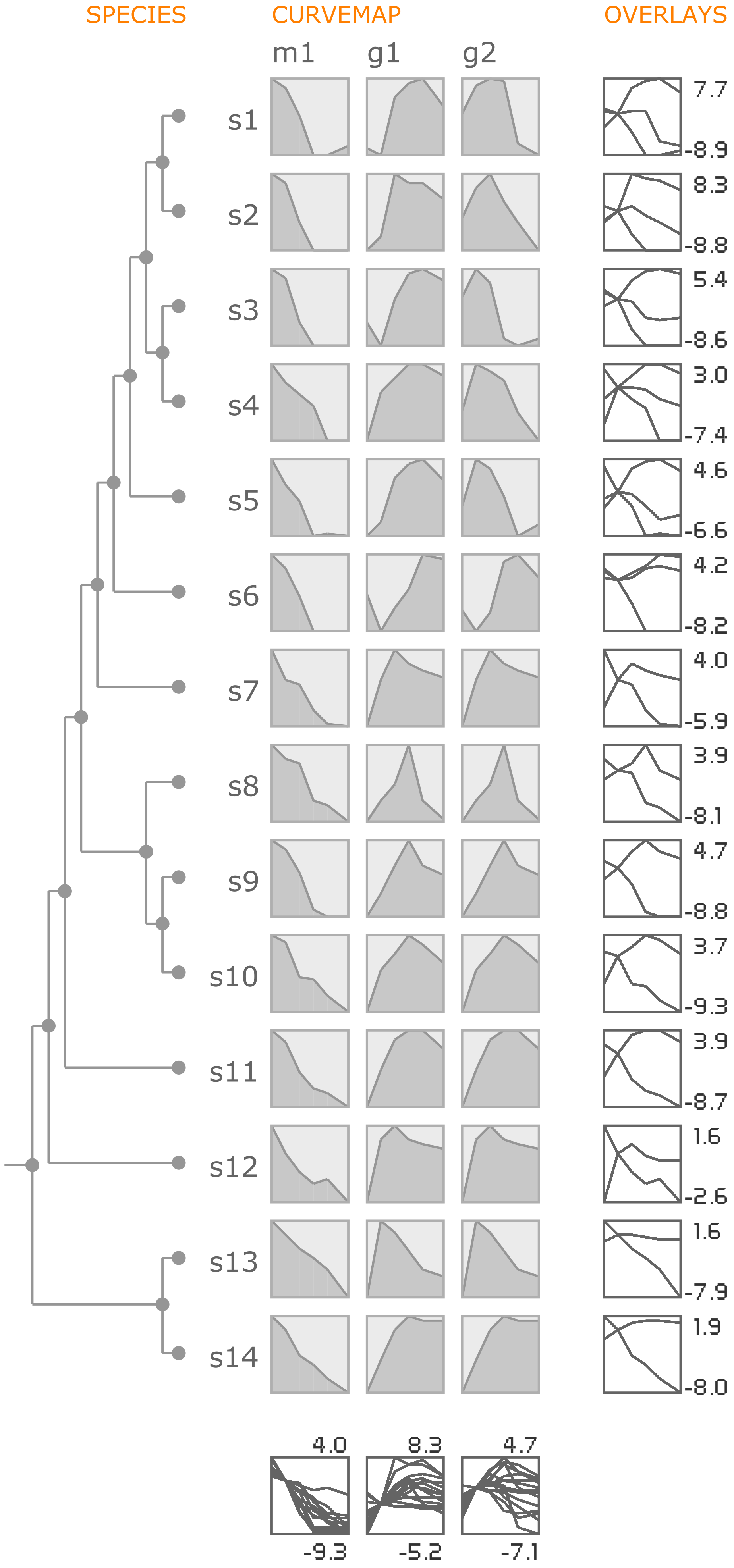


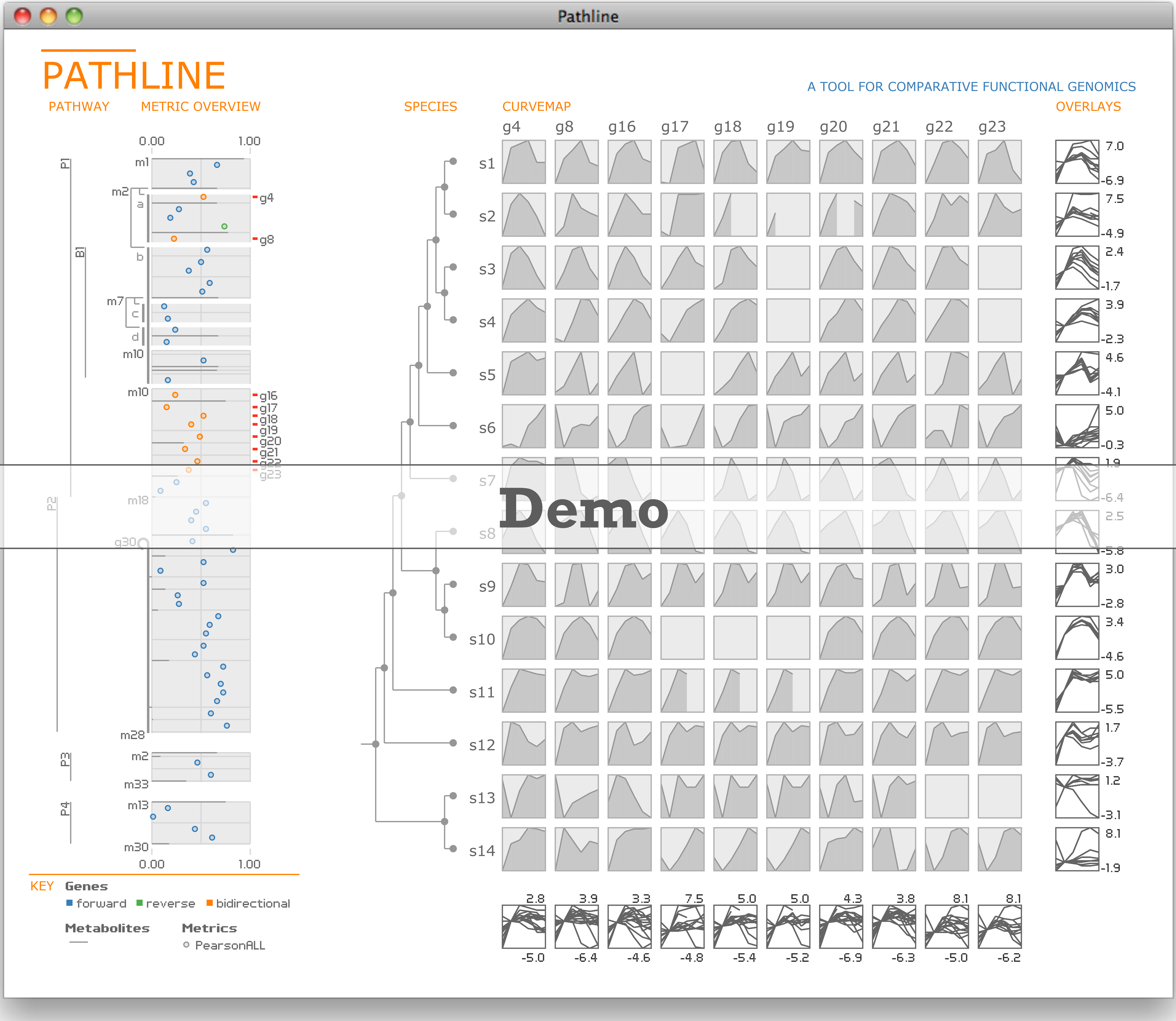
curvemap

inspired by heatmaps

  - *base visual unit is a curve*
  - *filled, framed line charts to enhance shape perception*

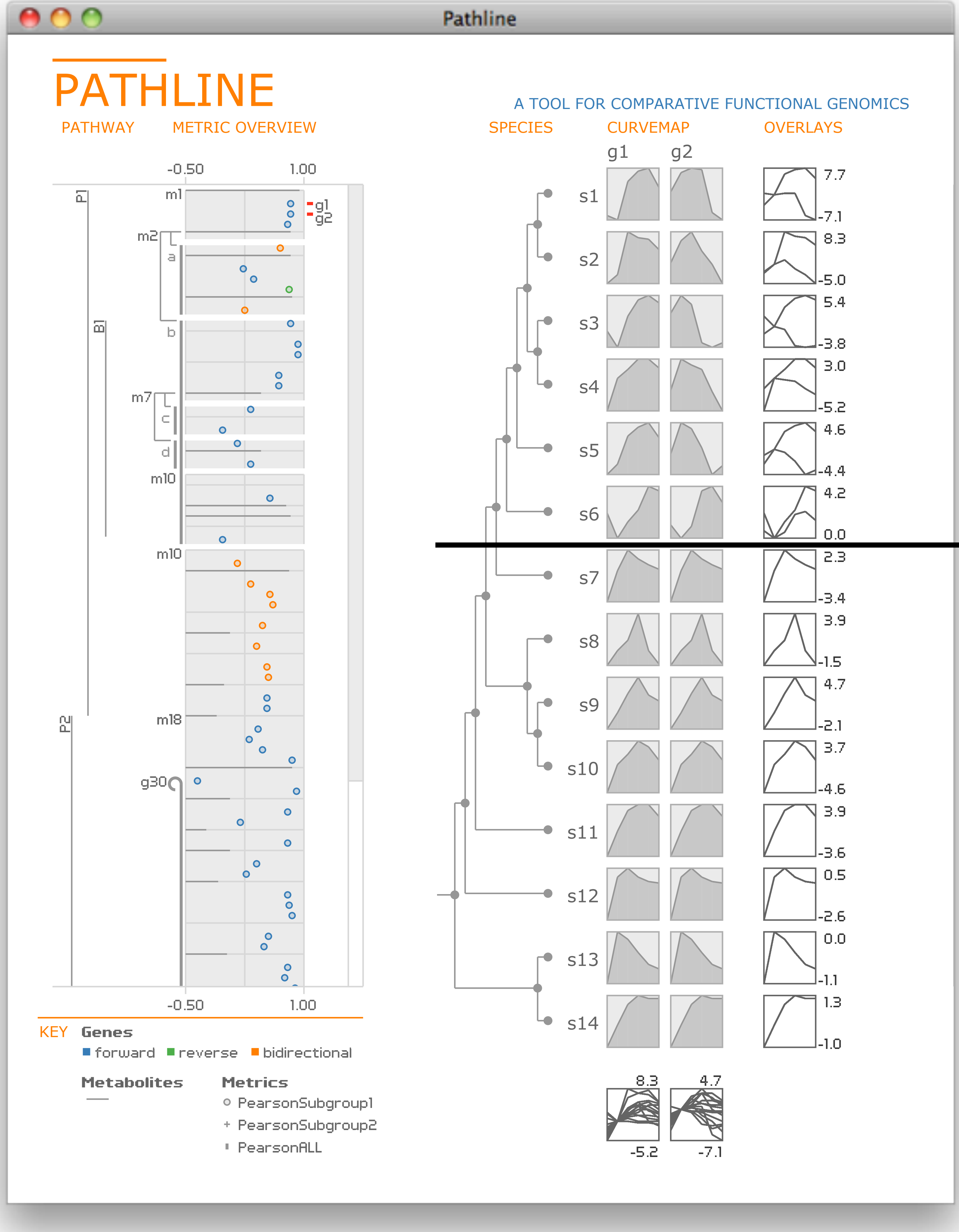
- rows are species
- columns are genes/metabolites
- overlays to enhance trend perception




Pathline
PATHLINE
A TOOL FOR COMPARATIVE FUNCTIONAL GENOMICS
PATHWAY
METRIC OVERVIEW
SPECIES
CURVEMAP
OVERLAYS
0.00
1.00
P1
B1
P2
P3
P4
m1
m2
a
b
m7
c
d
m10
m18
g30
m28
m33
m13
m30
g4
g8
g16
g17
g18
g19
g20
g21
g22
g23
s1
s2
s3
s4
s5
s6
s7
s8
s9
s10
s11
s12
s13
s14
7.0 -6.9
7.5 -4.9
2.4 -1.7
3.9 -2.3
4.6 -4.1
5.0 -0.3
1.9 -6.4
2.5 -5.8
3.0 -2.8
3.4 -4.6
5.0 -5.5
1.7 -3.7
1.2 -3.1
8.1 -1.9
2.8 -5.0
3.9 -6.4
3.3 -4.6
7.5 -4.8
5.0 -5.4
5.0 -5.2
4.3 -6.9
3.8 -6.3
8.1 -5.0
8.1 -6.2
Demo
KEY
Genes
■ forward
■ reverse
■ bidirectional
Metabolites
Metrics
○ PearsonALL

case studies

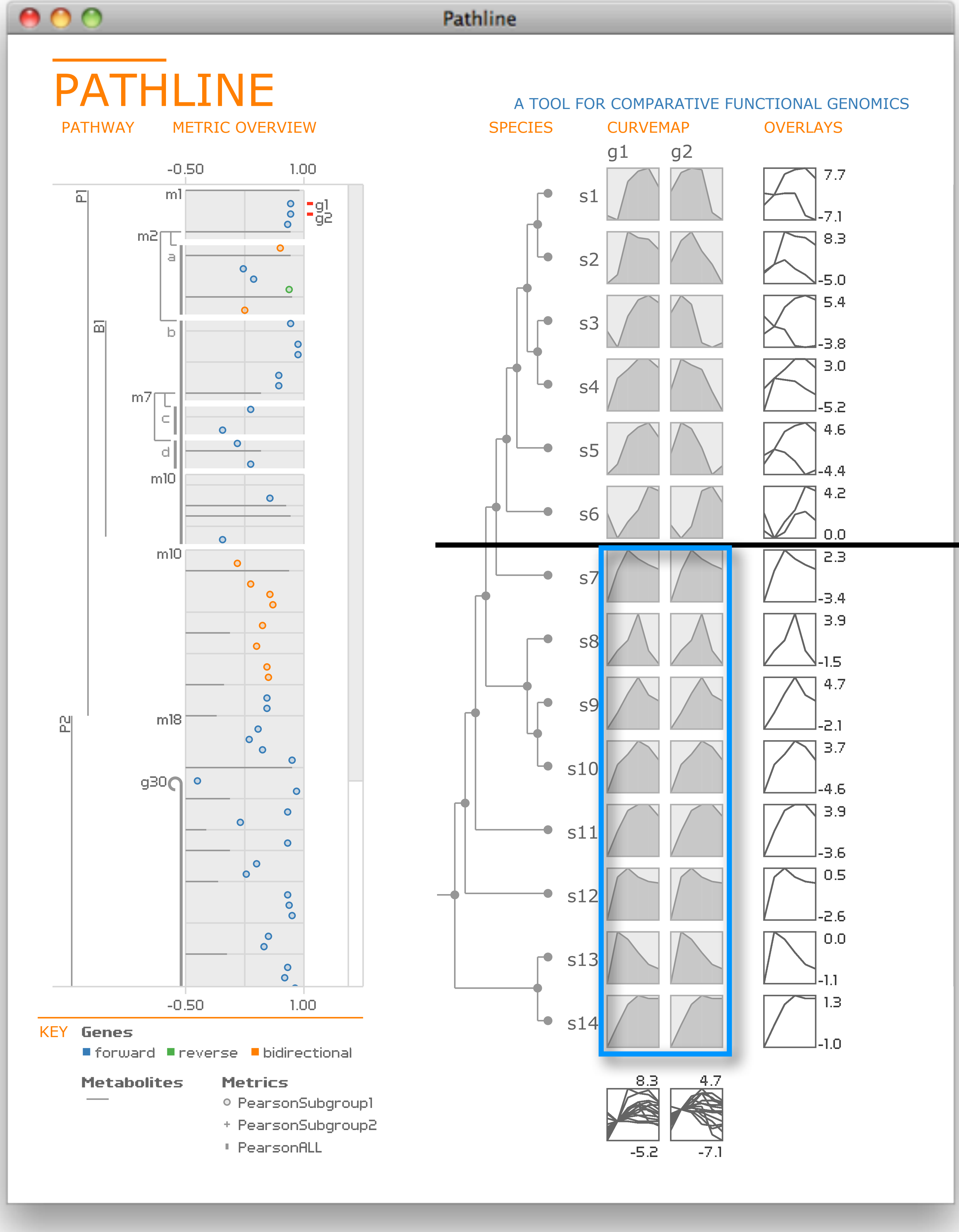
whole genome duplication



both genes

one gene

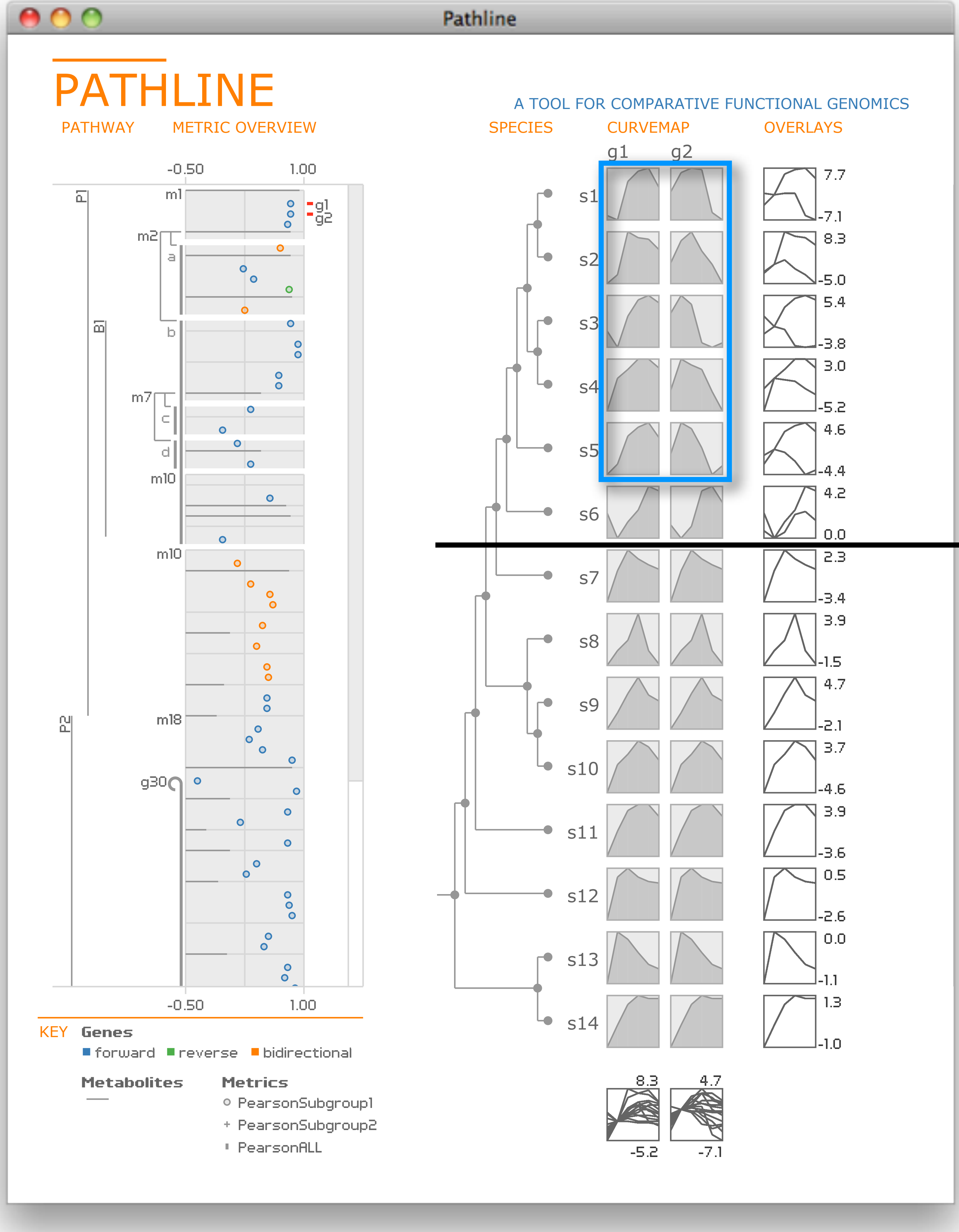
whole genome duplication



both genes

one gene

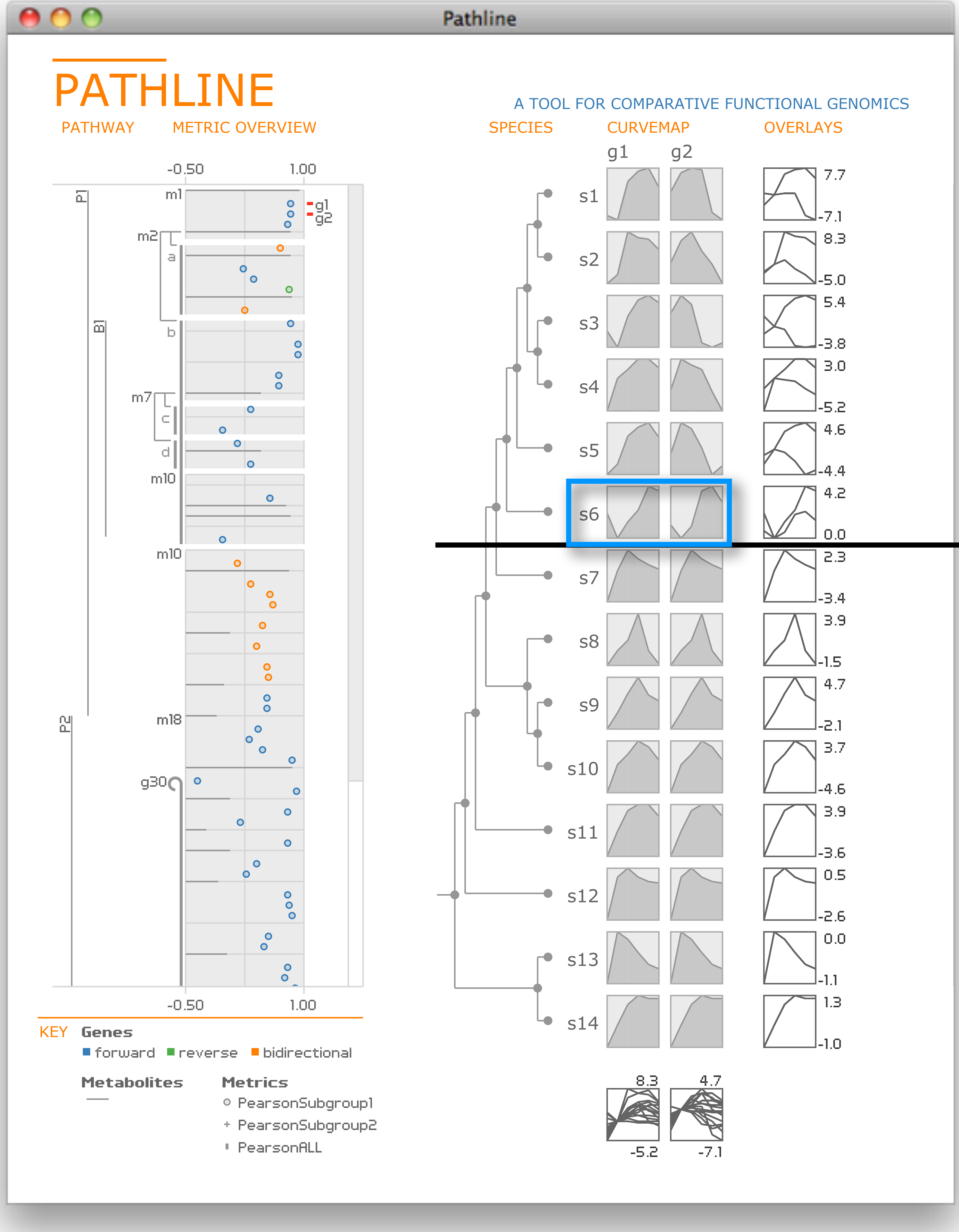
whole genome duplication



both genes

one gene

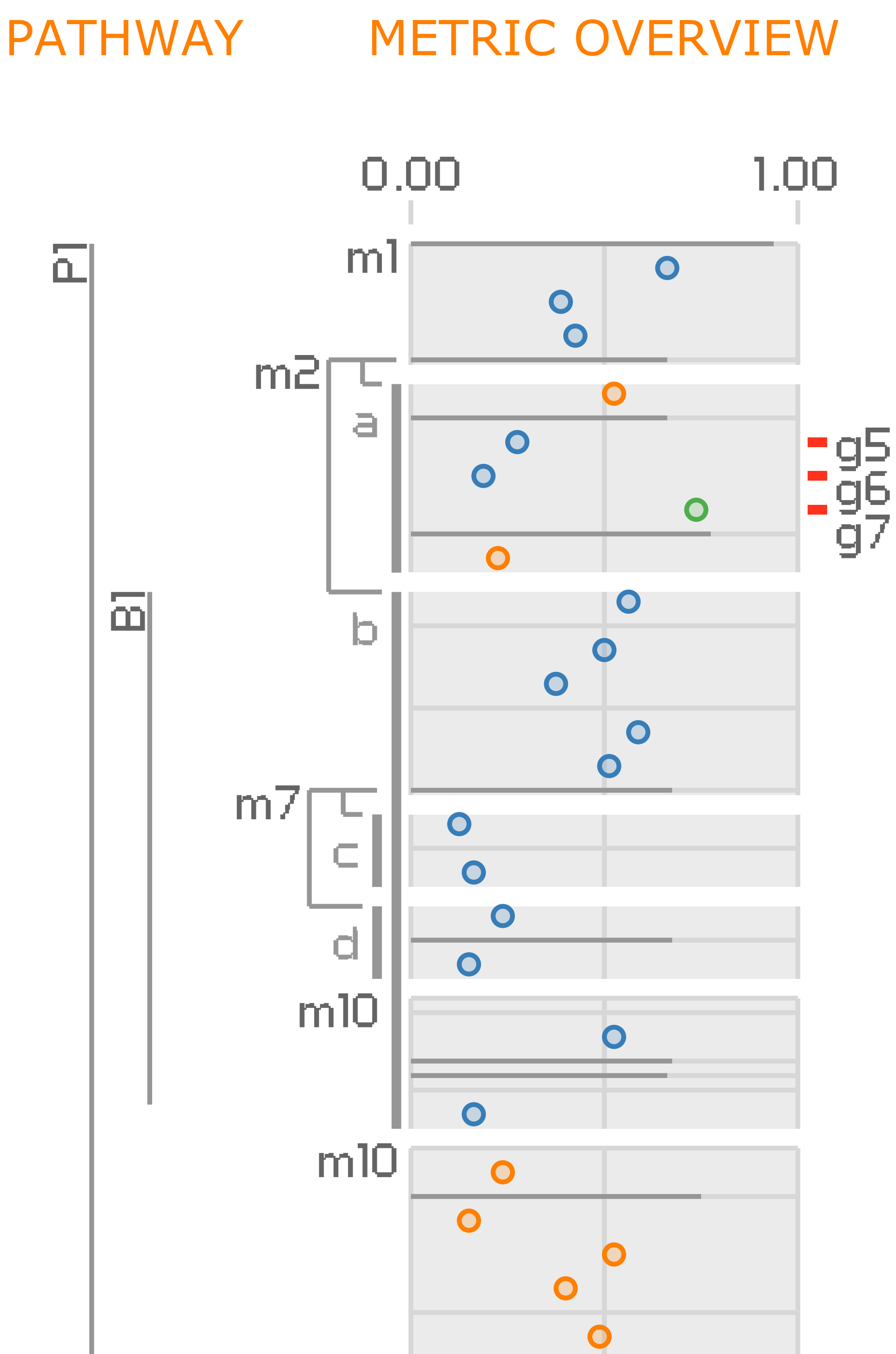
whole genome duplication



both genes

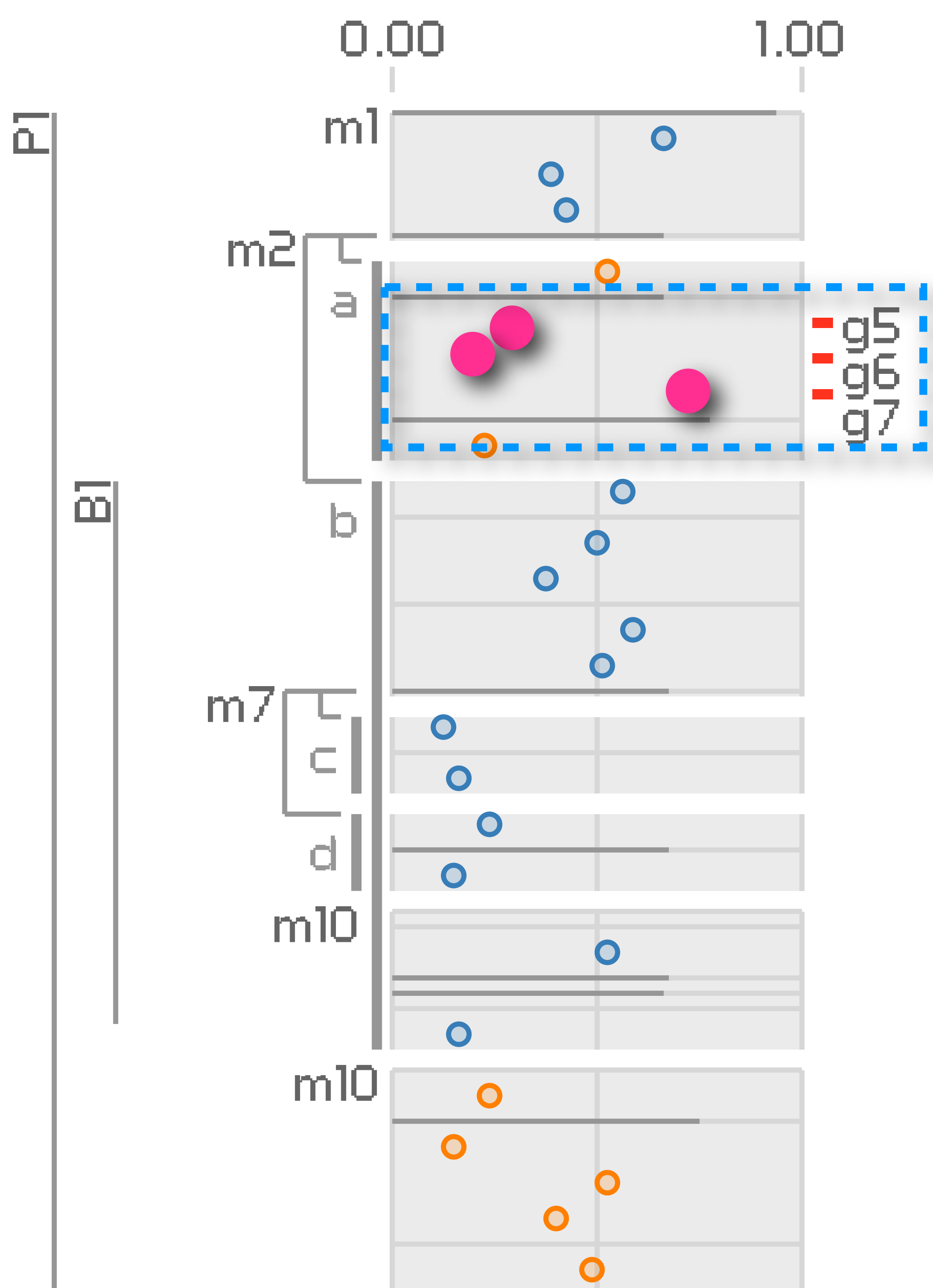
one gene

gene-level relationships

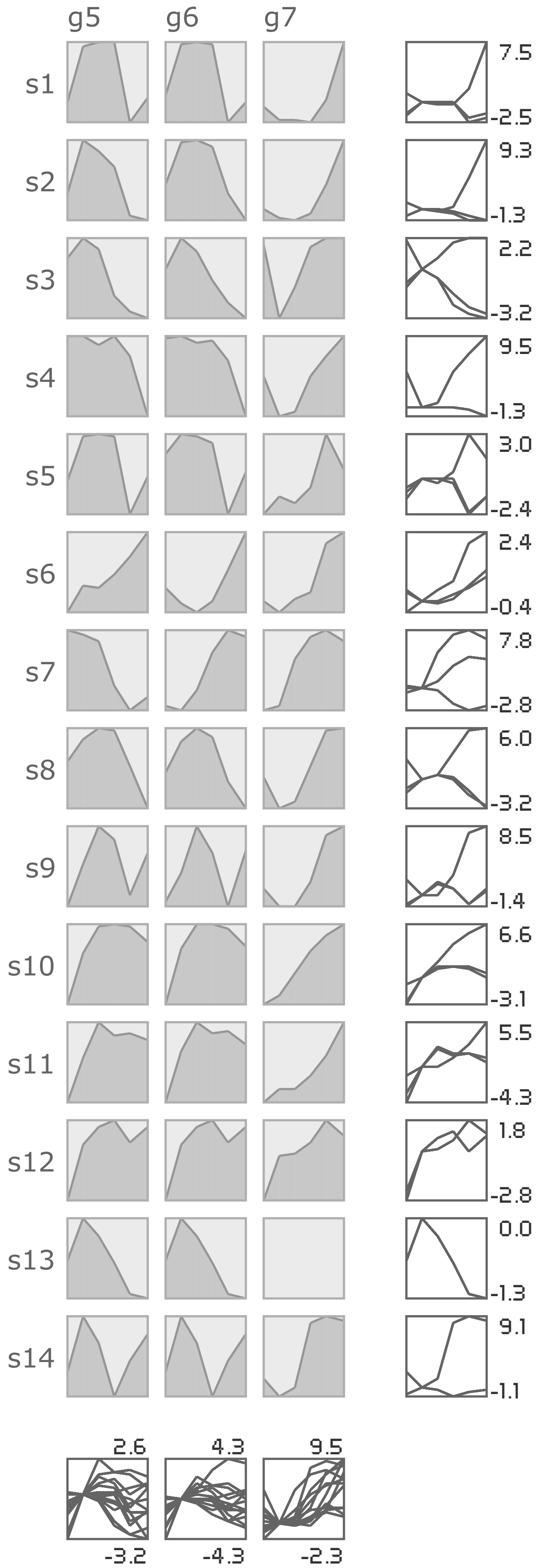


gene-level relationships

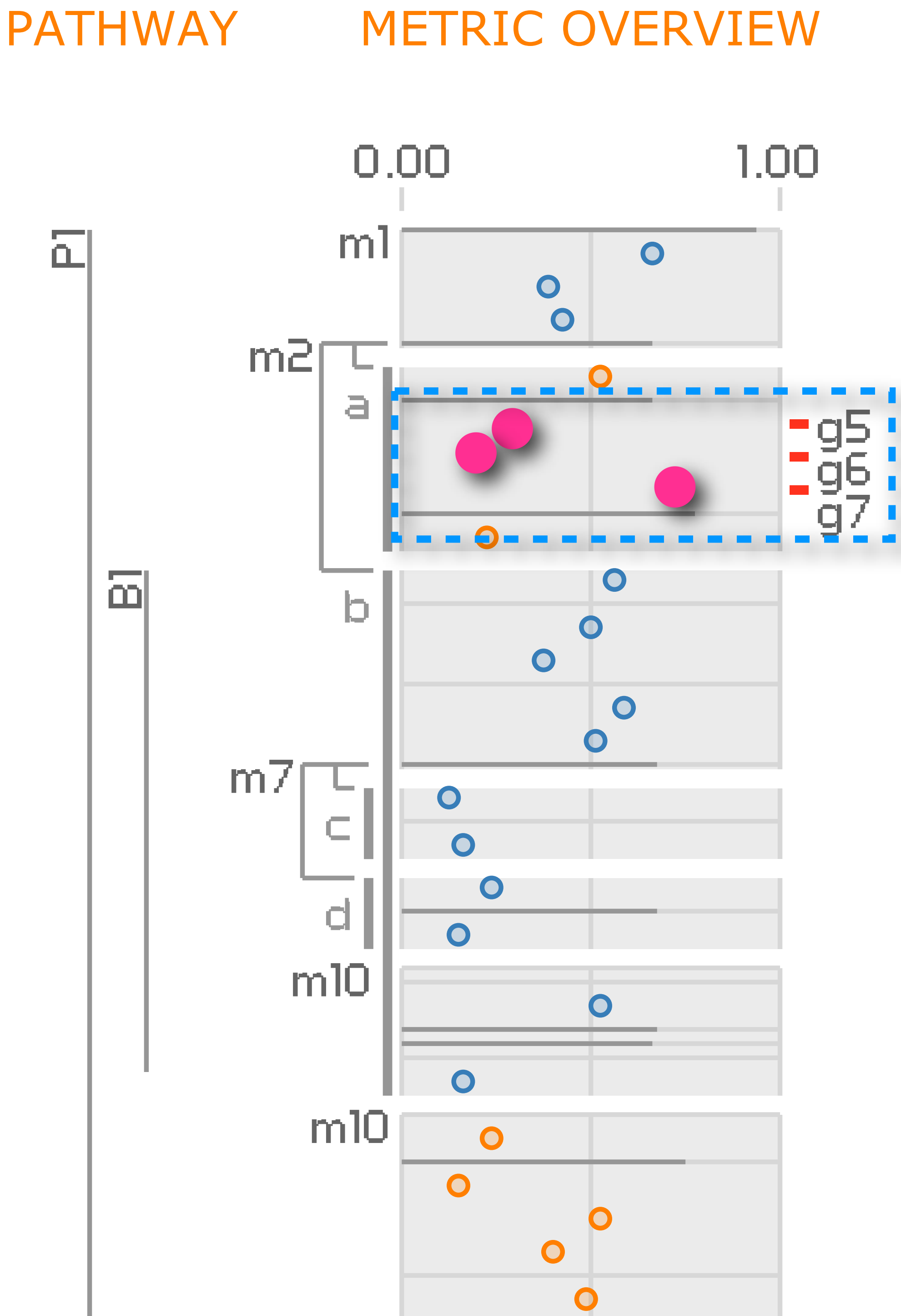
PATHWAY METRIC OVERVIEW

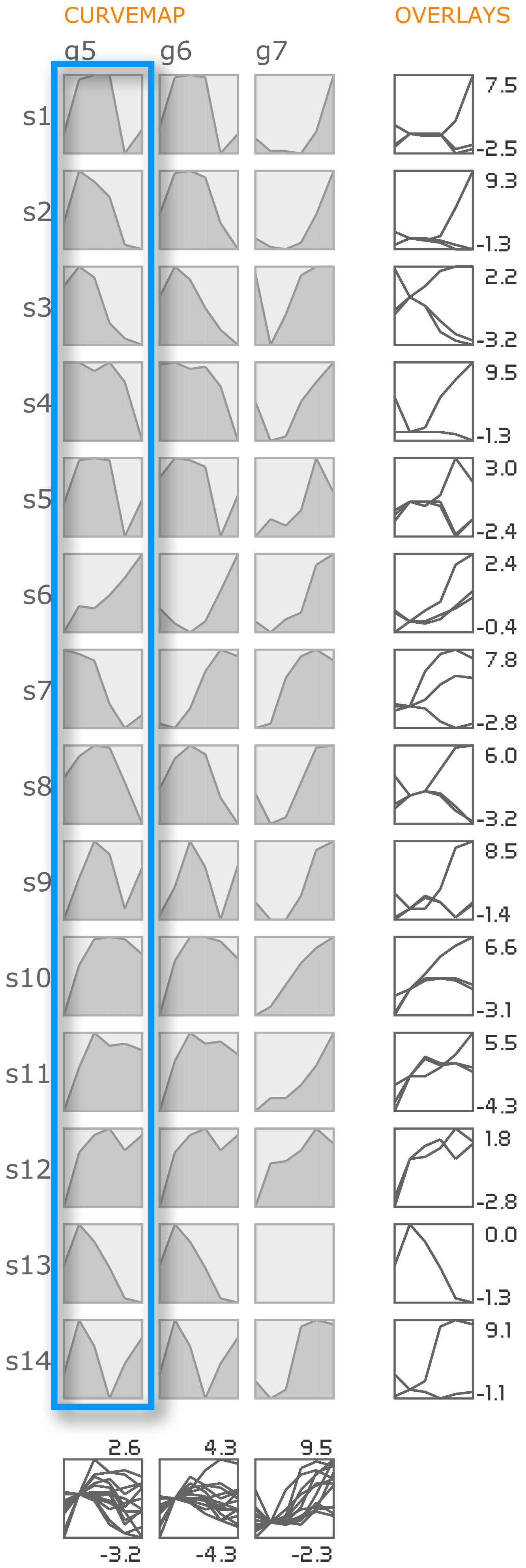


CURVEMAP OVERLAYS

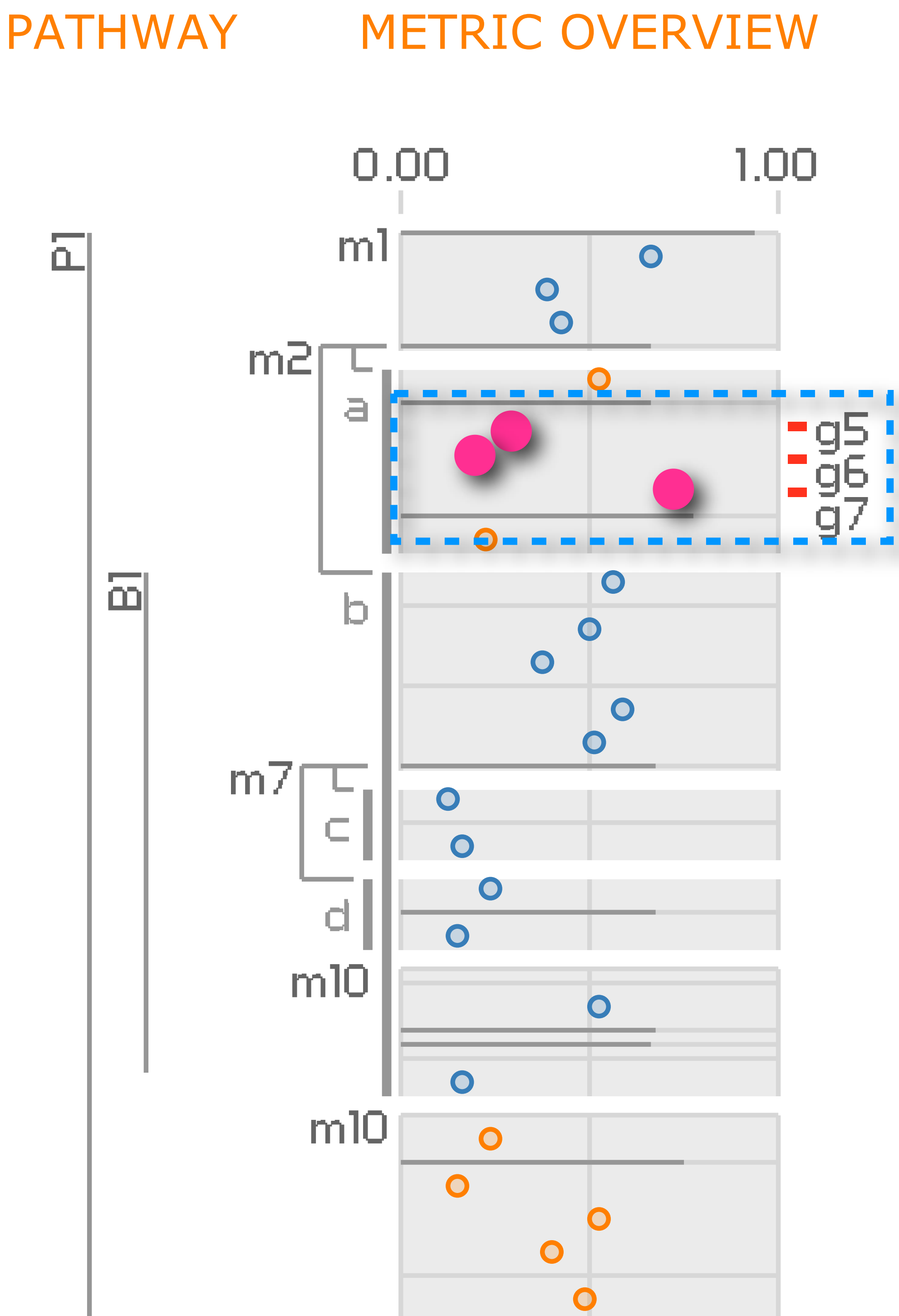


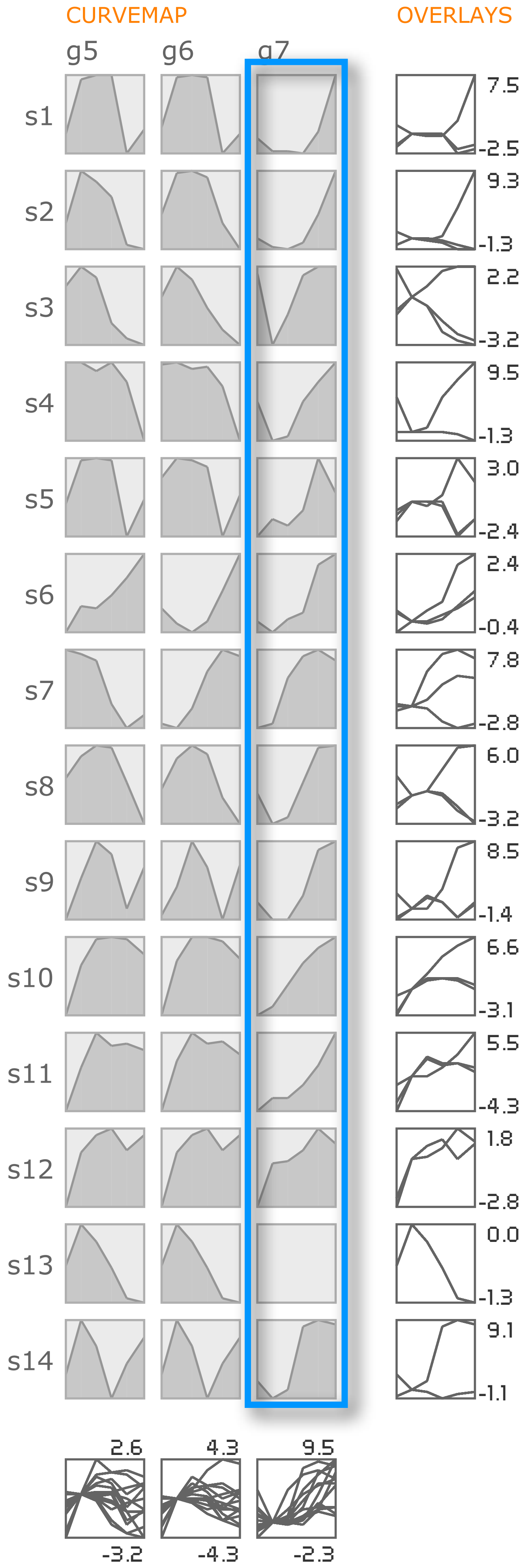
gene-level relationships





gene-level relationships

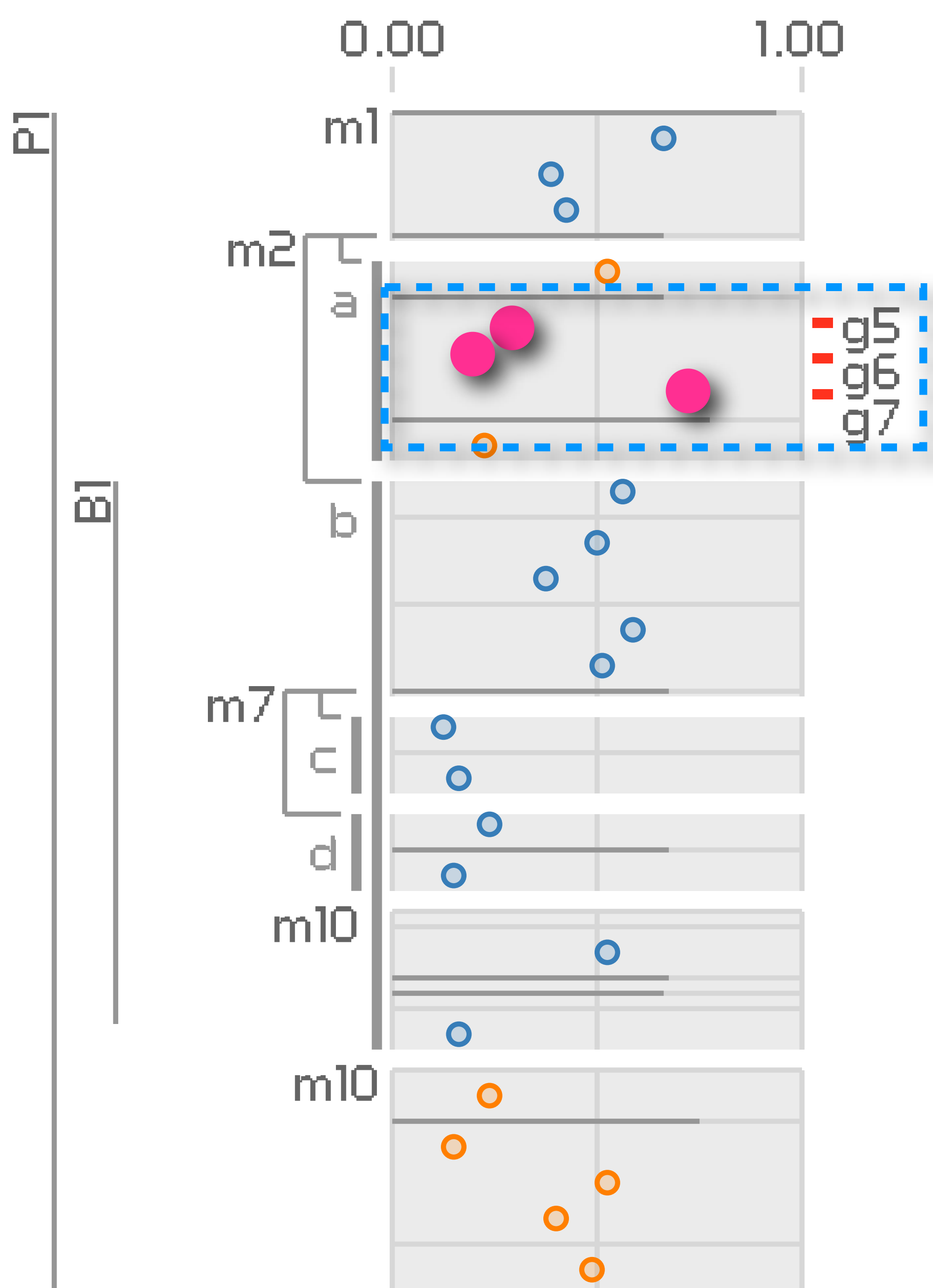




gene-level relationships

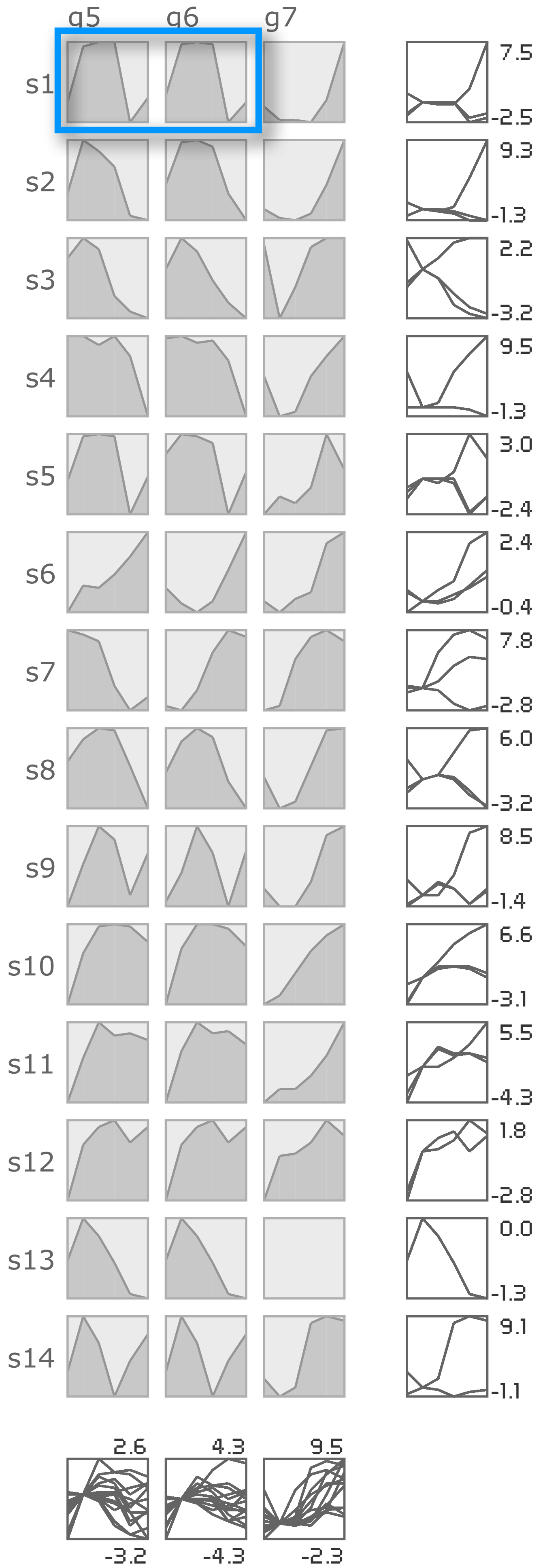
PATHWAY

METRIC OVERVIEW

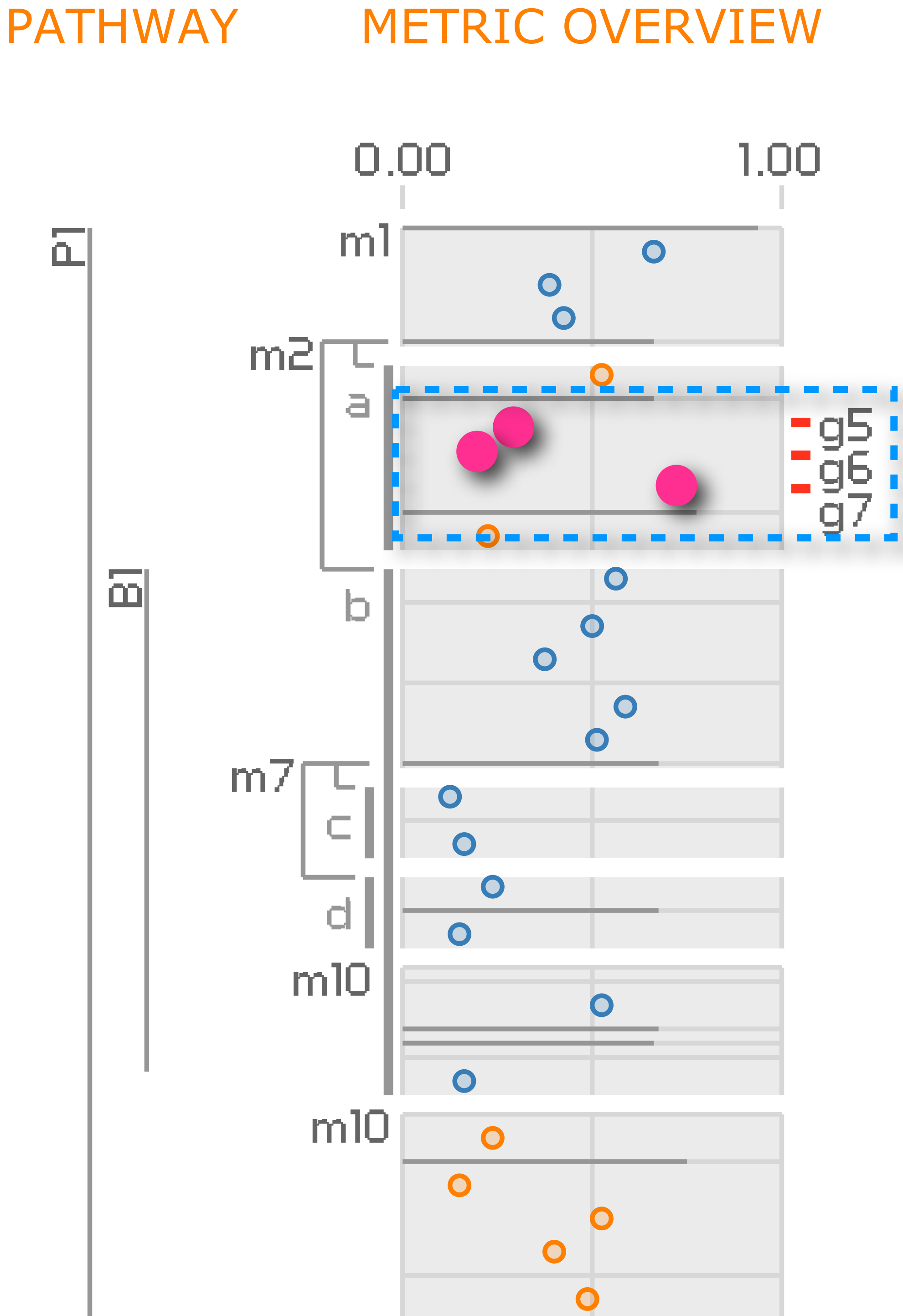


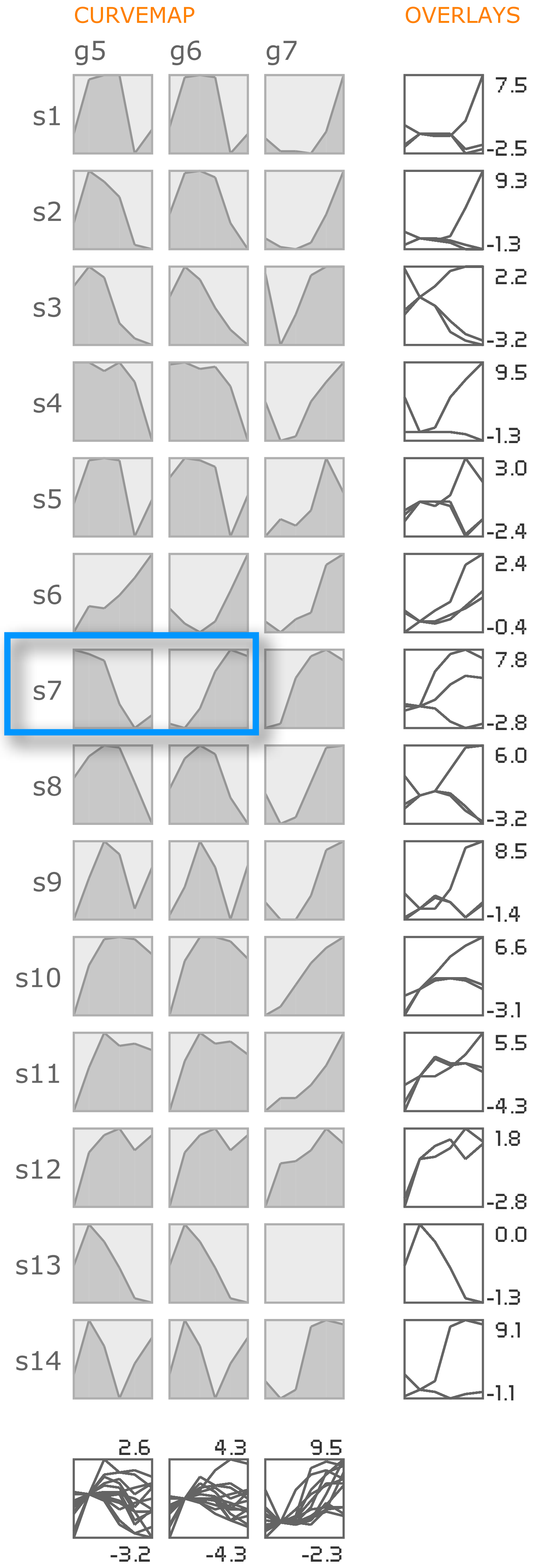
CURVEMAP

OVERLAYS



gene-level relationships





conclusions and future work

conclusions

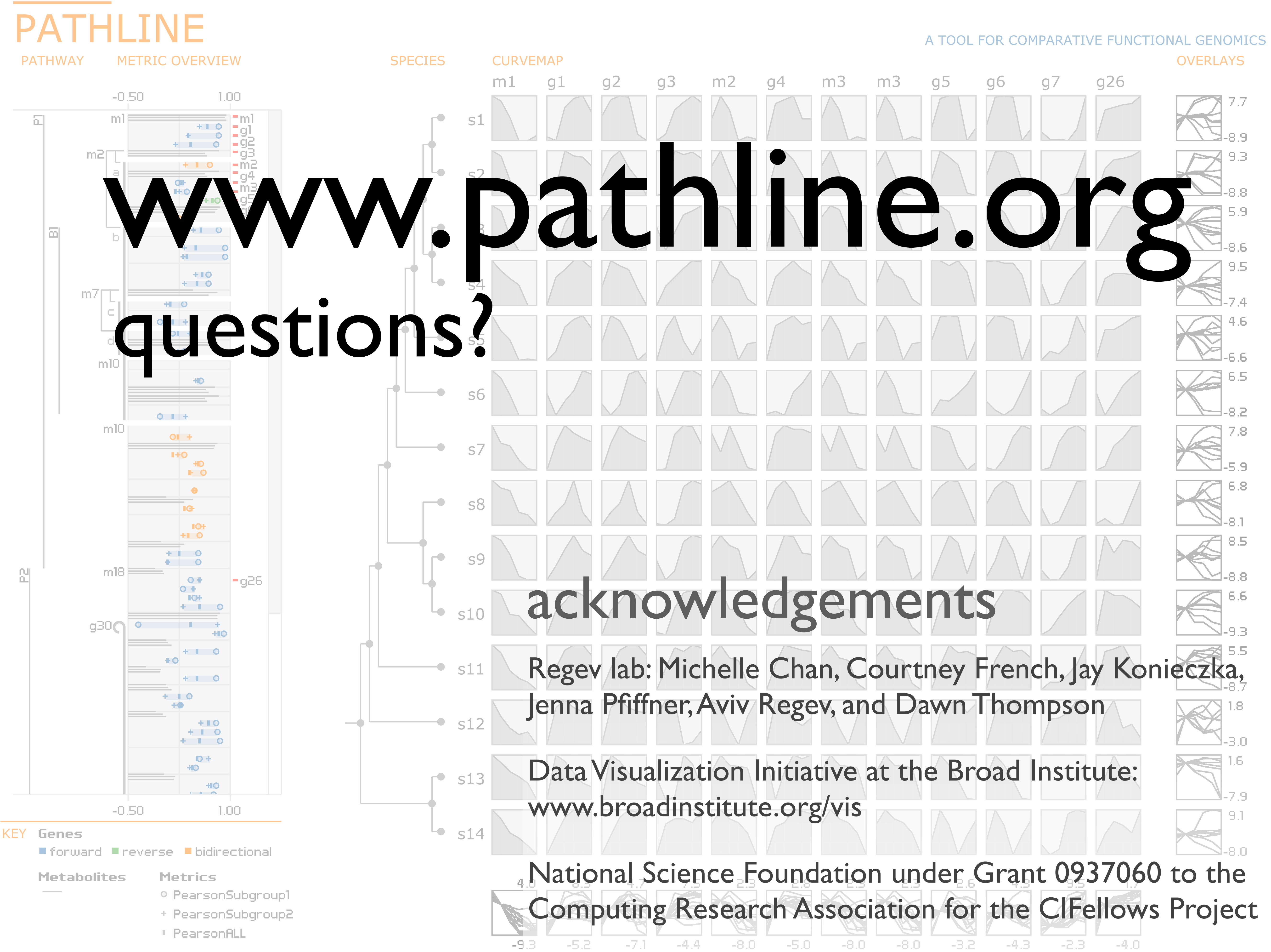
- Pathline: first interactive tool for comparative functional genomics
    - *multiple: genes, time points, species, and pathways*

- two new visual encodings
    - *curvemap for expression data with multiple dimensions*
    - *linearized pathway representation for comparing quantitative data*

- case studies: efficiency gains and new discoveries

future work

- automate pathway selection and linearization
- continue with Regev Lab: more data types
- apply ideas to other biological systems
  - *Pathline*
  - *linearized pathway representation*
  - *curvemaps*
- beyond biology: curvemap vs heatmap



www.pathline.org

questions?

acknowledgements

Regev lab: Michelle Chan, Courtney French, Jay Konieczka, Jenna Pfiffner, Aviv Regev, and Dawn Thompson

Data Visualization Initiative at the Broad Institute: www.broadinstitute.org/vis

National Science Foundation under Grant 0937060 to the Computing Research Association for the CIFellows Project