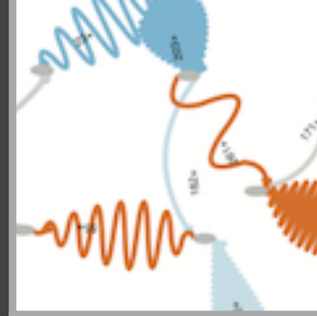
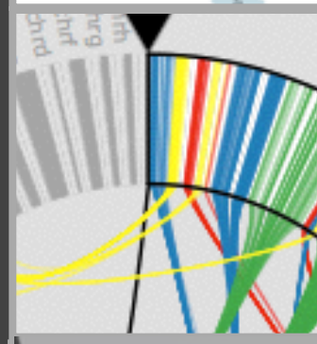
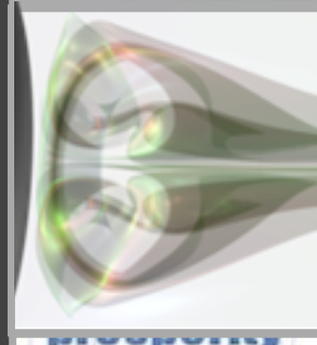
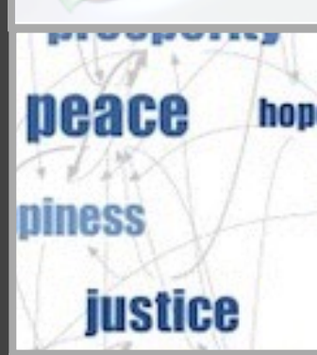
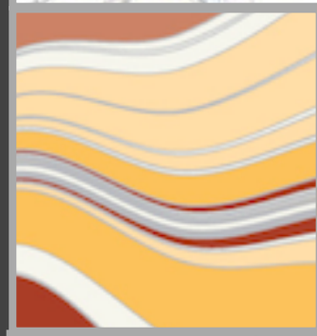
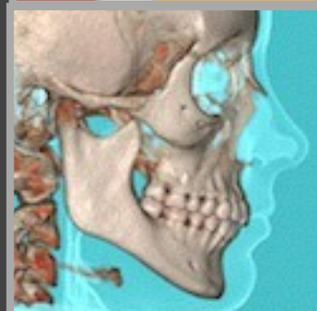
cs6630 | December 2 2014

DESIGN STUDIES

Miriah Meyer
University of Utah

administrivia . . .

- parallel coordinates grades out
- questions about transfer functions?
- exam next Thursday

last time . . .

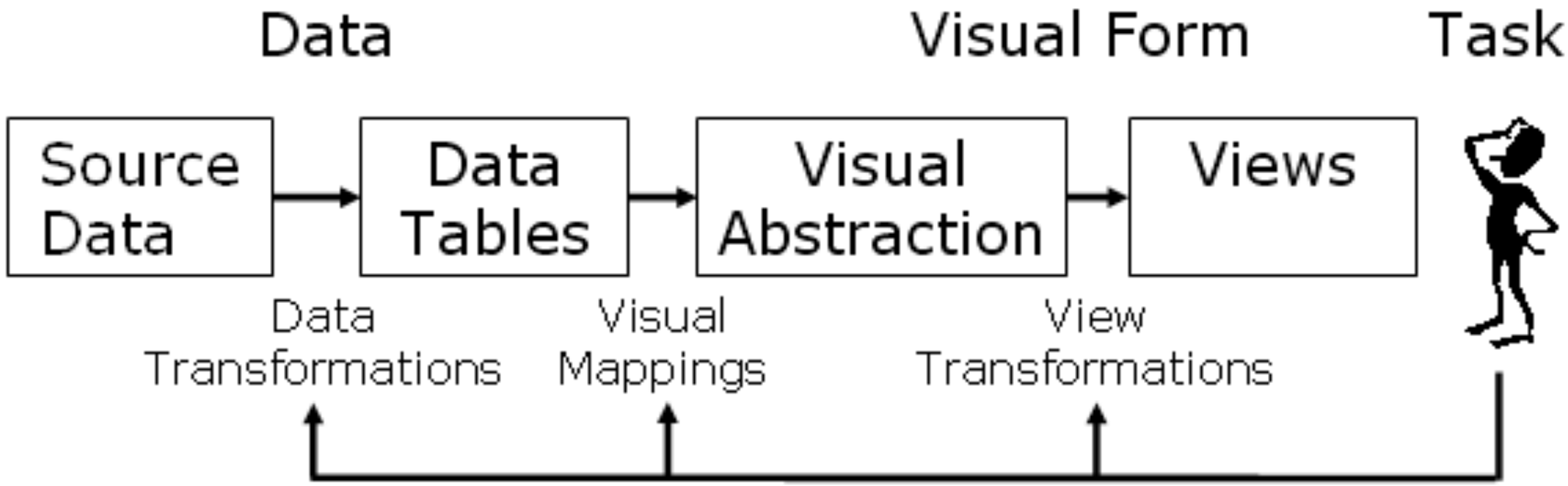
- **software architecture models**
  - *focus on the structure of a software system in terms of its programmatic components*

- software architecture models
  - *focus on the structure of a software system in terms of its programmatic components*
- design decision models
  - *describe and capture design decisions*

- **software architecture models**
  - *focus on the structure of a software system in terms of its programmatic components*
- **design decision models**
  - *describe and capture design decisions*
- **process models**
  - *describe stages with concrete actions a designer should engage in*

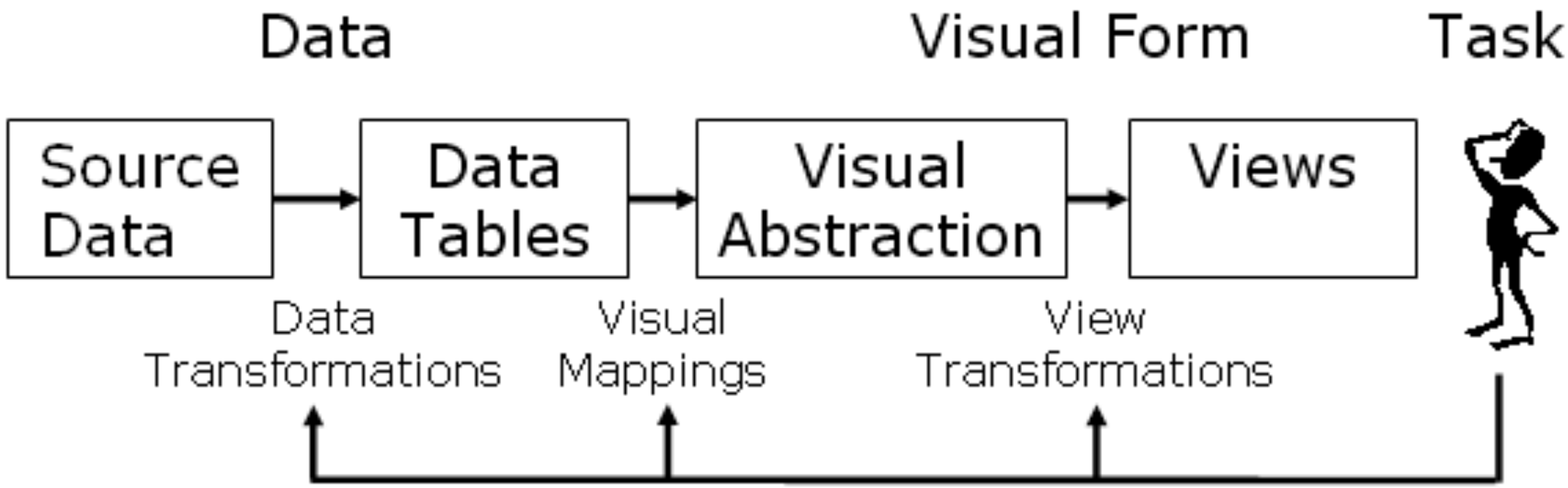
reference model

- **software architecture pattern**
  - *breaks up visualization (user) process into a series of discrete steps*



reference model

- **software architecture pattern**
  - *breaks up visualization (user) process into a series of discrete steps*

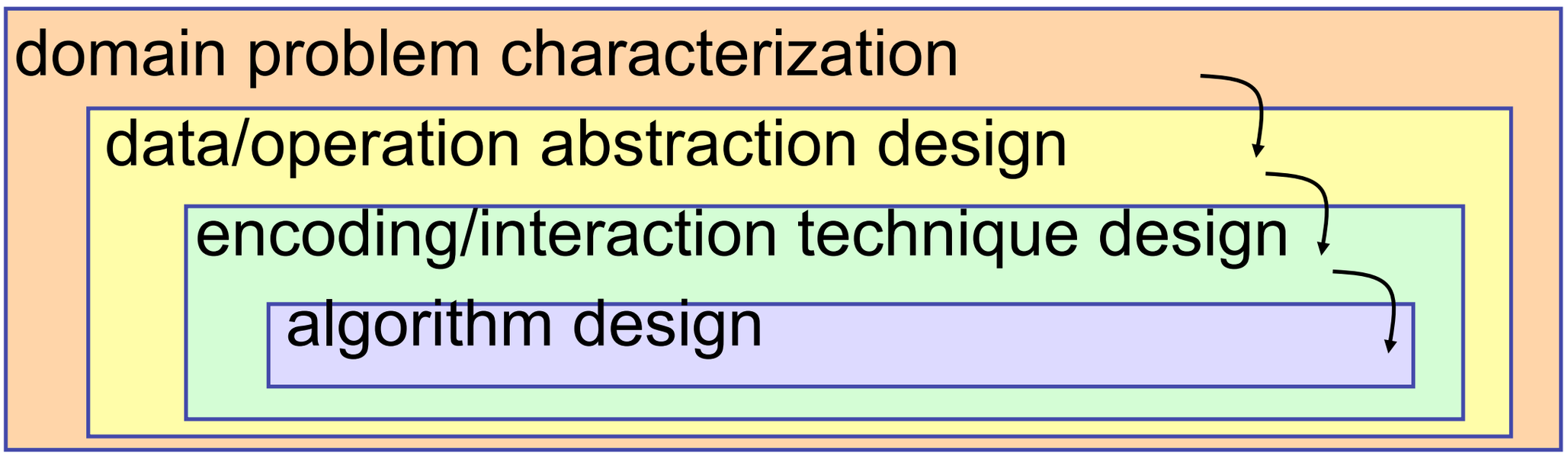


originally developed by Ed Chi as part of PhD dissertation, called the data state model; showed equivalence to data flow model used in existing toolkits like VTK

later interpreted by Card, Mackinlay, and Shneiderman, dubbing it the information visualization reference model

Nested levels in model

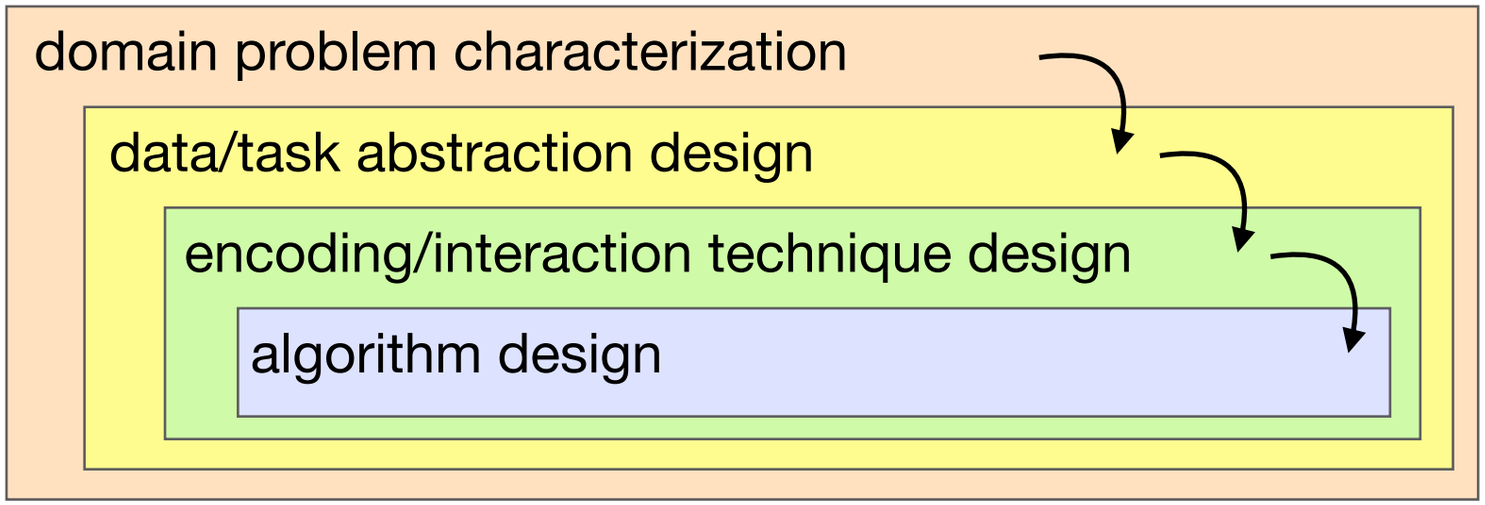
- output of **upstream** level → input to **downstream** level
- challenge: upstream errors inevitably cascade
  - if poor abstraction choice made, even perfect technique and algorithm design will not solve intended problem



NESTED BLOCKS AND GUIDELINES

[Meyer 2013]

NESTED MODEL

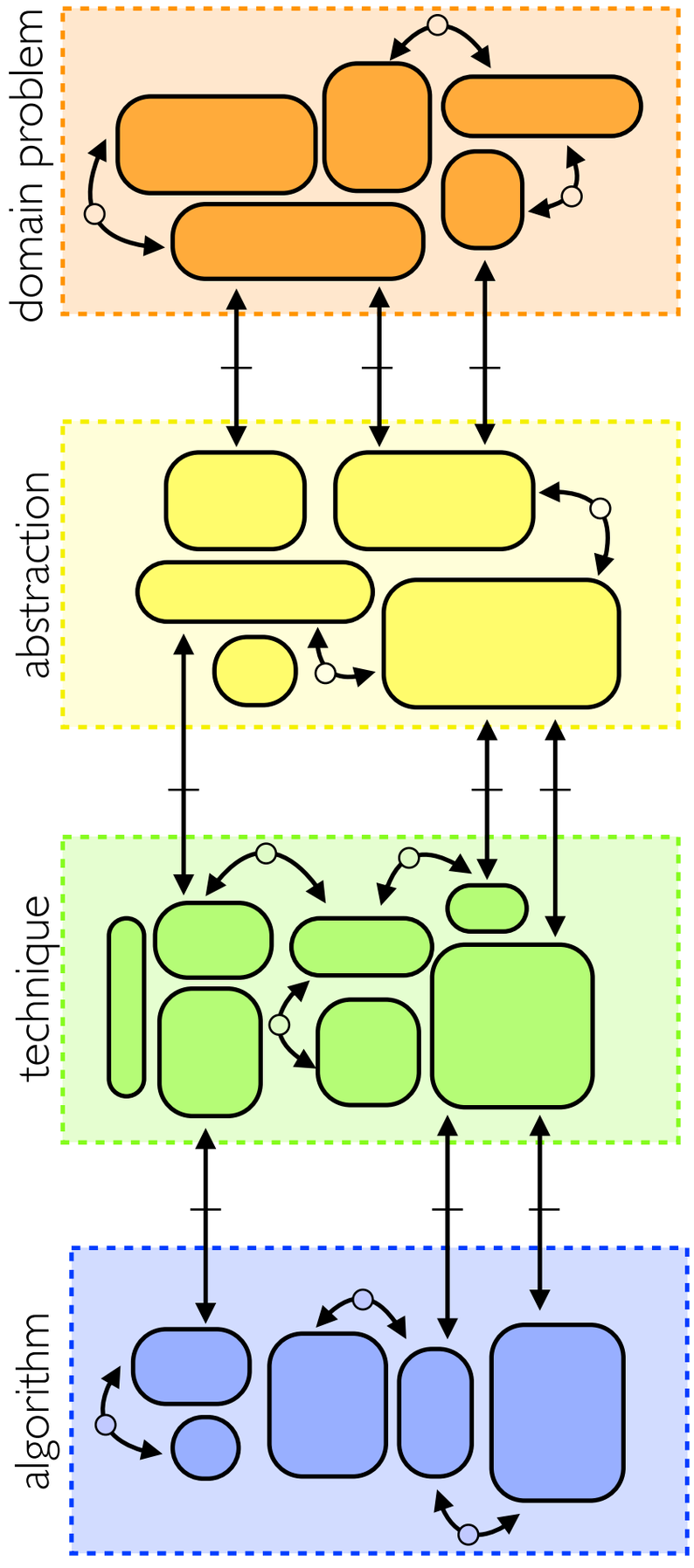


Munzner 2009

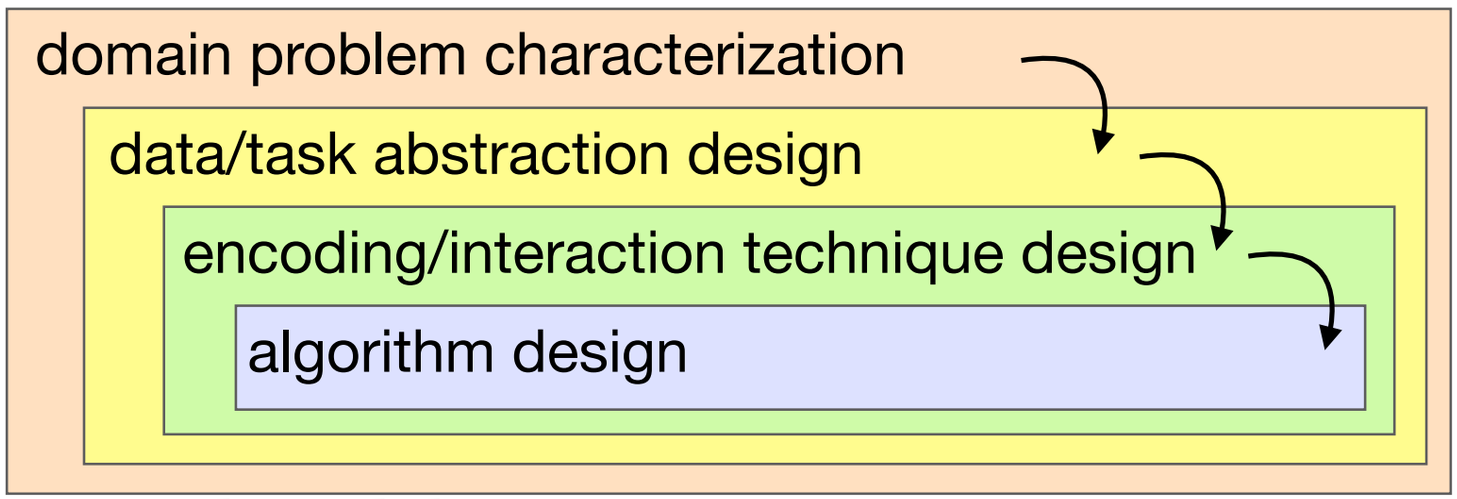
blocks guidelines

between-level guideline

within-level guideline



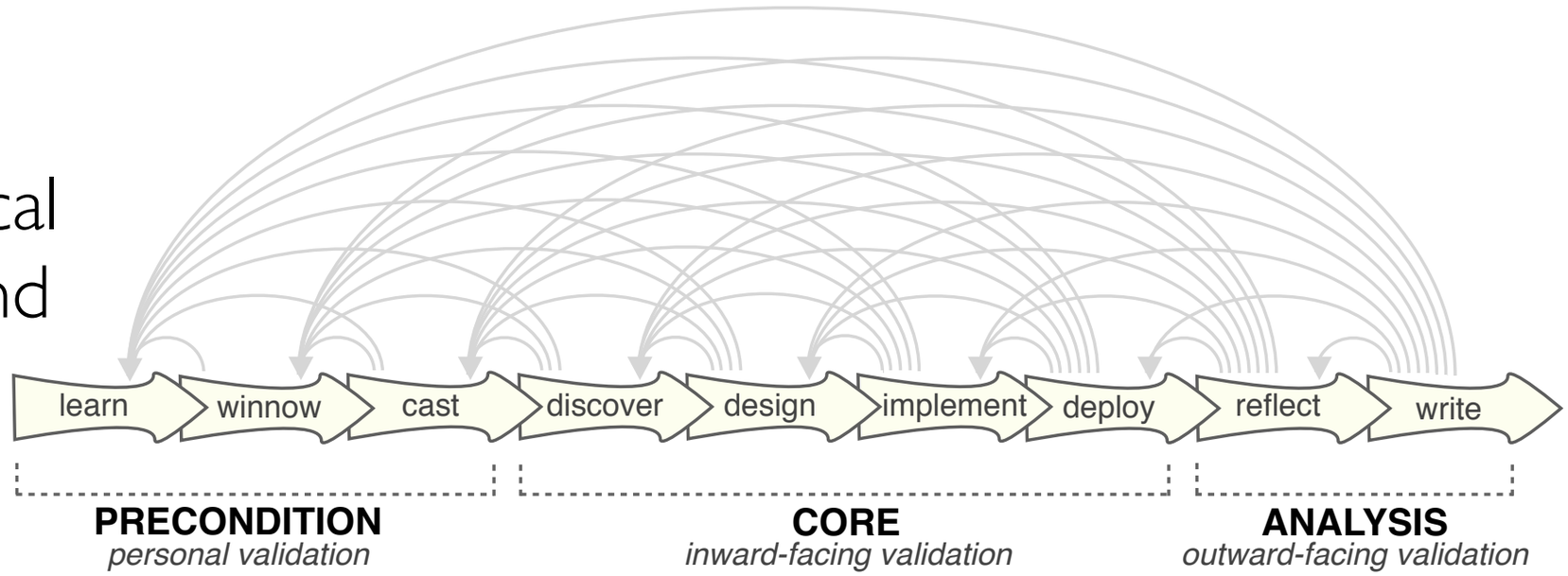
design decision models *vs* process models



nested model

design decision model: describes levels of design inherent to, and should be considered in, the creation of a tool

process model: gives practical advice in how to design and develop a tool



9-stage framework

today . . .

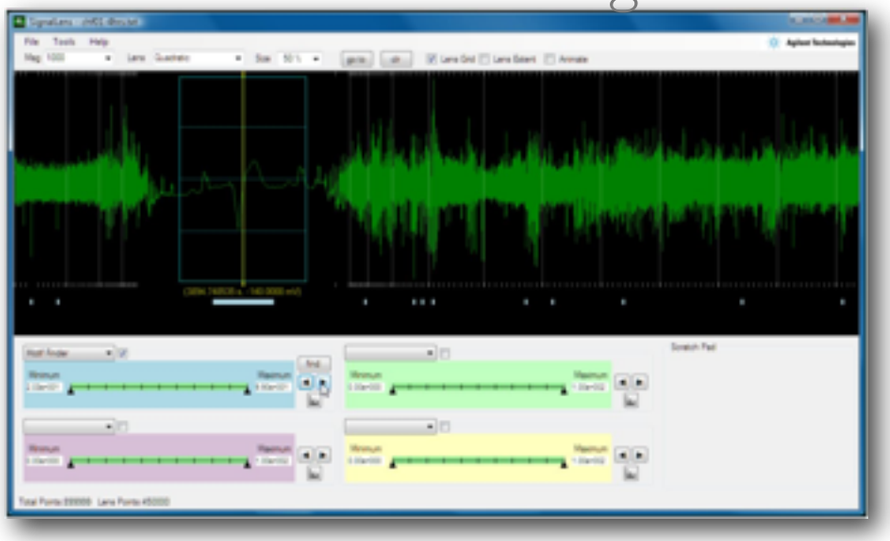
DESIGN STUDIES

- what is a design study?
- data and task axes
- nine-stage framework
- Pathline
- MizBee
- selected pitfalls

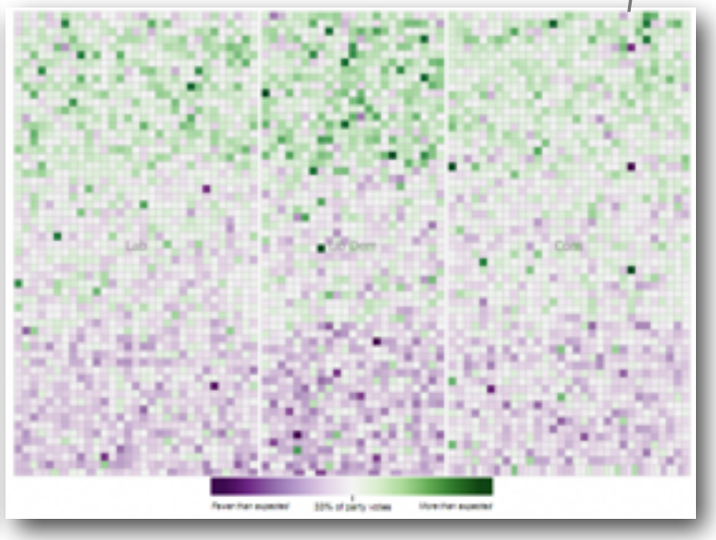
Design Studies

Popular

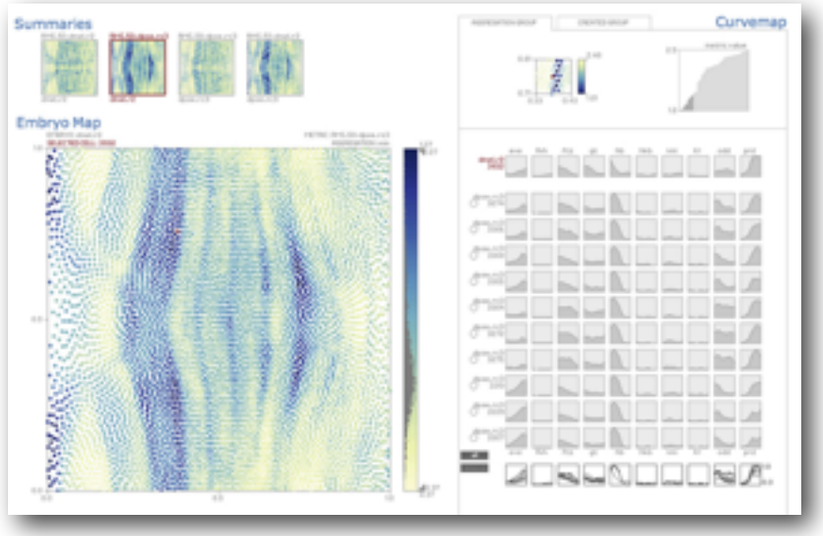
SignalLens



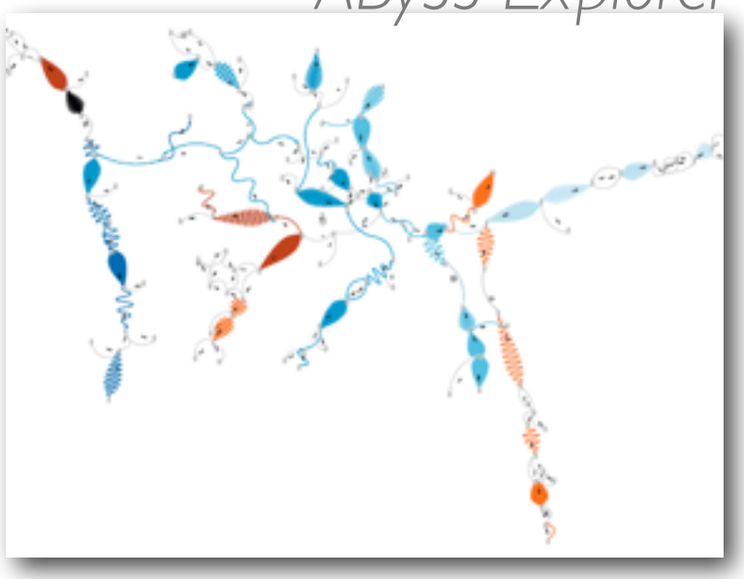
BallotMaps



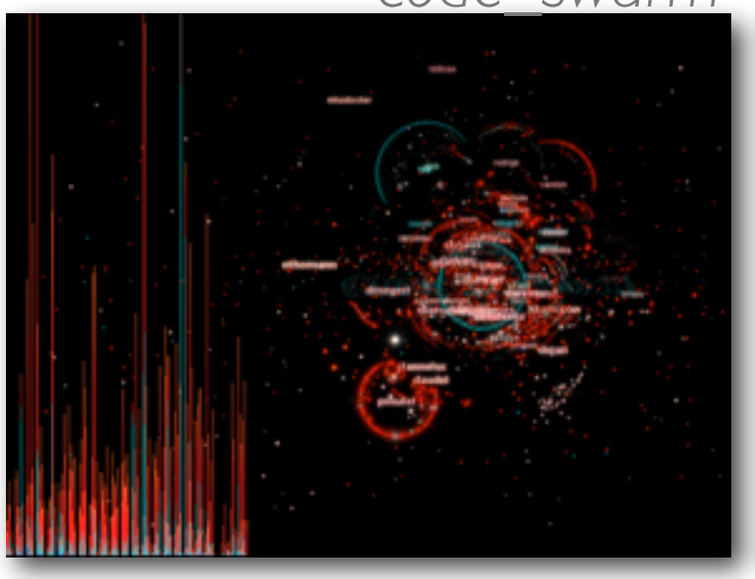
MulteeSum



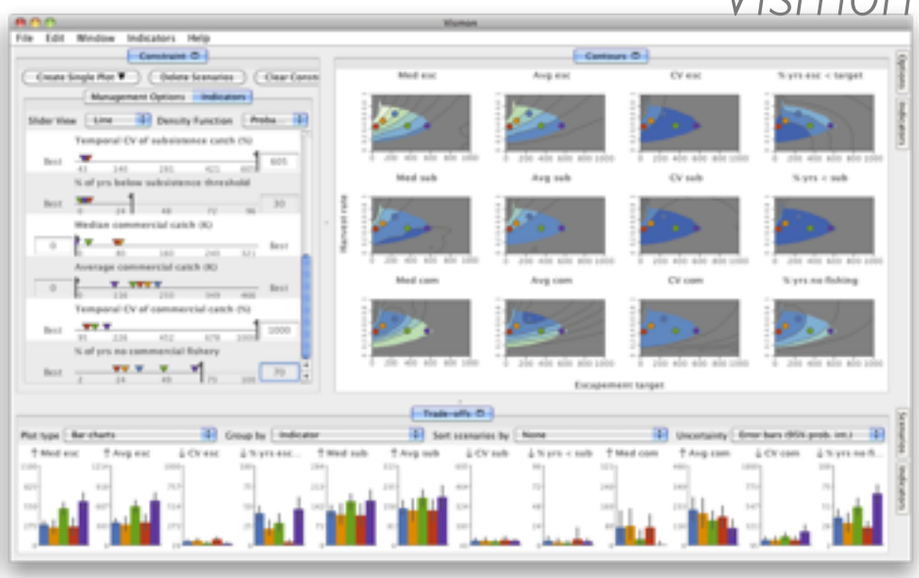
ABySS-Explorer



code_swarm



Vismon

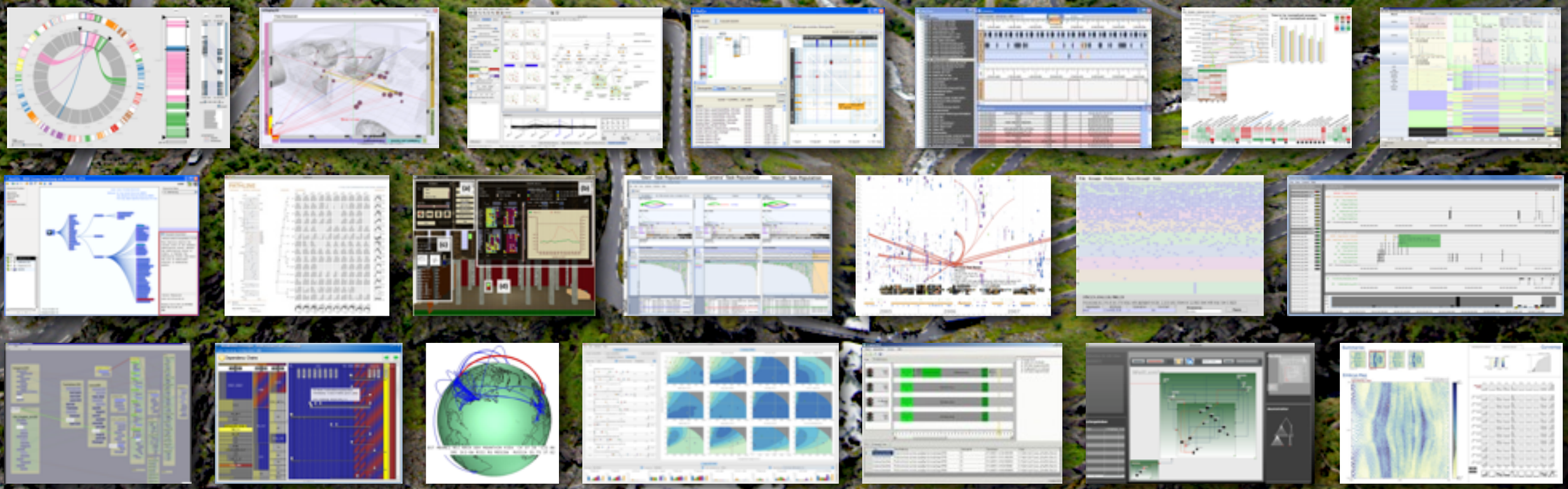


Design Studies

Hard!

Design Studies

Hard!



Design Studies

How to?

Methods

Methodology

http://www.air-n-water.com/blog/quick-summer-meal/
http://manwifeanddog.com/2012/05/05/a-homemade-recipe-for-a-happy-wife/

Design Studies

How to?

Methods

Methodology

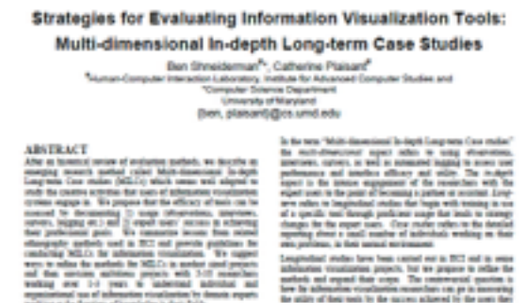
http://www.air-n-water.com/blog/quick-summer-meal/
http://manwifeanddog.com/2012/05/05/a-homemade-recipe-for-a-happy-wife/

Design Studies

How to: Methods

Data Sketches

[Lloyd and Dykes, InfoVis 2011]



Strategies for Evaluating Information Visualization Tools:
Multi-dimensional In-depth Long-term Case Studies

Ben Shneiderman[1]*, Catherine Plaisant[2]
[1]Human-Computer Interaction Laboratory, Institute for Advanced Computer Studies and
[2]Computer Science Department
University of Maryland
{ben, plaisant}@cs.umd.edu

ABSTRACT

MILCs

[Shneiderman and Plaisant, BELIV 2006]

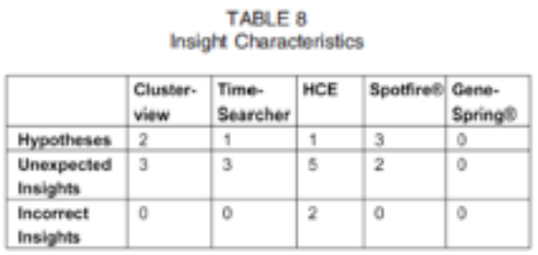


TABLE 8
Insight Characteristics

	Cluster-view	Time-Searcher	HCE	Spotfire®	Gene-Spring®
Hypotheses	2	1	1	3	0
Unexpected Insights	3	3	5	2	0
Incorrect Insights	0	0	2	0	0

Insight-Based

[Sarayia et al., TVCG 2005]

Design Studies

How to: Methodology

Methods

Methodology

http://www.air-n-water.com/blog/quick-summer-meal/
http://manwifeanddog.com/2012/05/05/a-homemade-recipe-for-a-happy-wife/

Design Studies

How to: Methodology

Good ingredients

Methods

... but no recipes

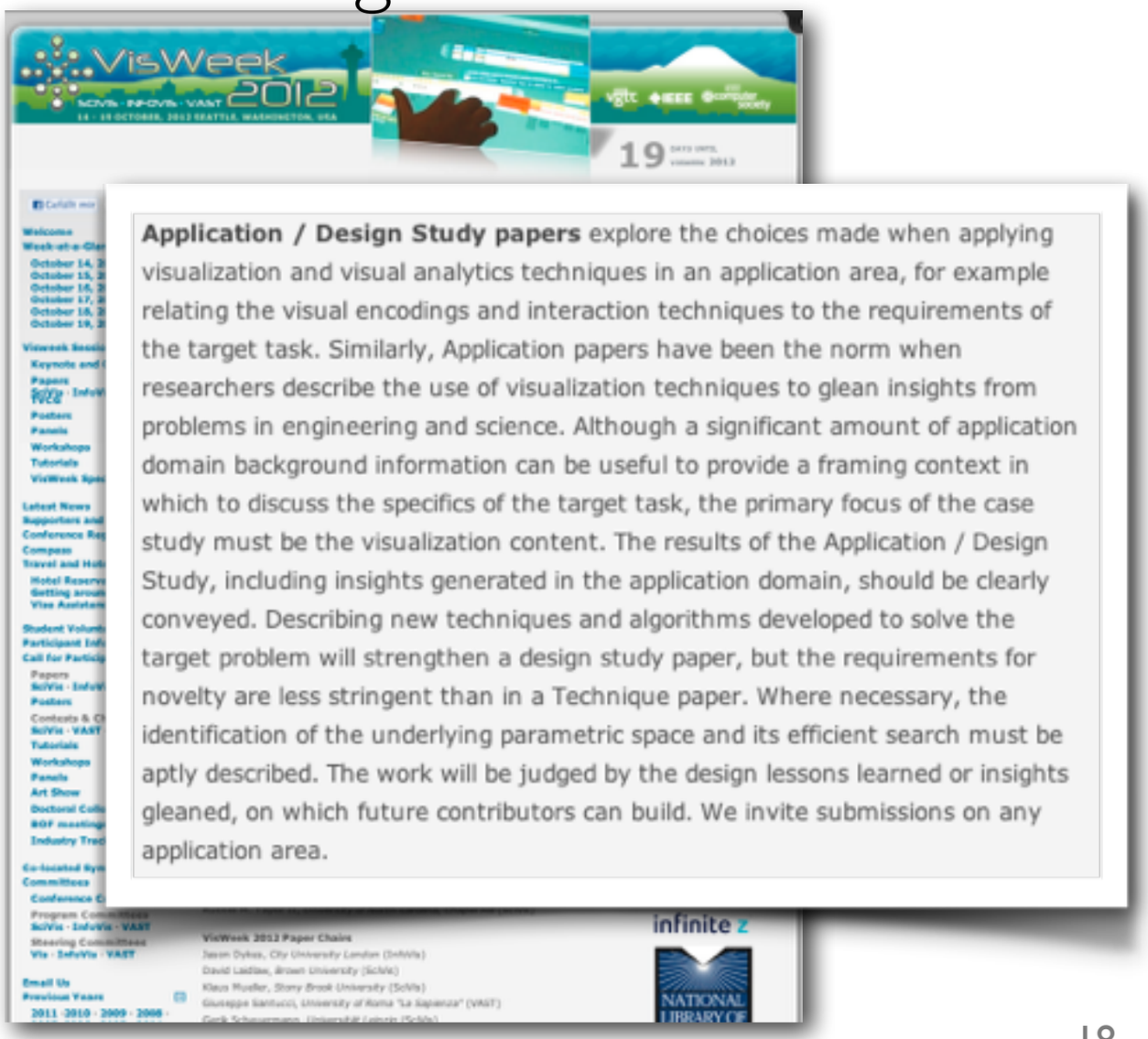
Methodology

http://www.air-n-water.com/blog/quick-summer-meal/
http://manwifeanddog.com/2012/05/05/a-homemade-recipe-for-a-happy-wife/

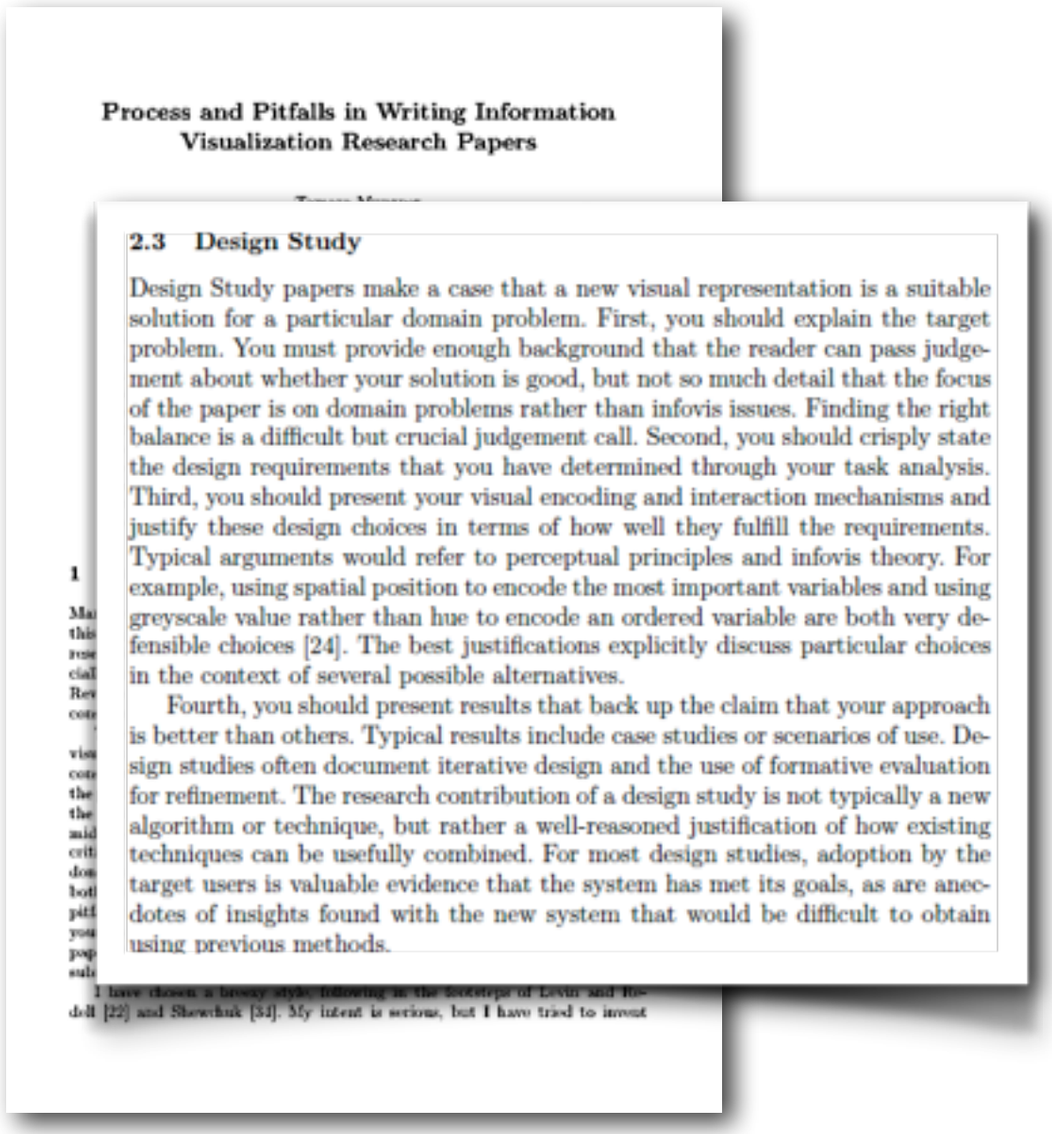
Design Studies

Three paragraphs!

visweek.org



Munzner 2008



Goal

Design Study Methodology

design study

- a project
- analyze a real-world problem
- design a visualization system
- validate the design
- reflect about lessons learned

Design Study Methodology: Reflections for the Trenches and the Stacks.
M. Sedlmair, M. Meyer, T. Munzner, IEEE TVCG (Proceedings of InfoVis 2012).

design study

- a project
- analyze a **real-world** problem
- design a visualization system
- validate the design
- reflect about lessons learned

Design Study Methodology: Reflections for the Trenches and the Stacks.
M. Sedlmair, M. Meyer, T. Munzner, IEEE TVCG (Proceedings of InfoVis 2012).

design study

- a project
- analyze a real-world problem
- **design** a visualization system
- validate the design
- reflect about lessons learned

Design Study Methodology: Reflections for the Trenches and the Stacks.
M. Sedlmair, M. Meyer, T. Munzner, IEEE TVCG (Proceedings of InfoVis 2012).

design study

- a project
- analyze a real-world problem
- design a visualization system
- **validate** the design
- reflect about lessons learned

Design Study Methodology: Reflections for the Trenches and the Stacks.
M. Sedlmair, M. Meyer, T. Munzner, IEEE TVCG (Proceedings of InfoVis 2012).

design study

- a project
- analyze a real-world problem
- design a visualization system
- validate the design
- **reflect** about lessons learned

Design Study Methodology: Reflections for the Trenches and the Stacks.
M. Sedlmair, M. Meyer, T. Munzner, IEEE TVCG (Proceedings of InfoVis 2012).

AXES

AXES

AXES

AXES

TASK CLARITY

INFORMATION LOCATION

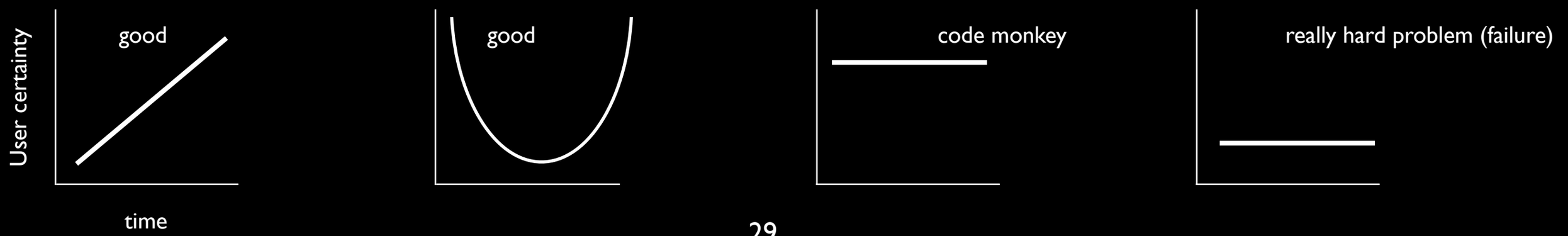
- **tasks in vis are usually rather complex**
  - not just: *buy a train ticket*
  - instead: *wicked problems*
- **what is a wicked problem?**

- **tasks in vis are usually rather complex**
  - not just: *buy a train ticket*
  - instead: *wicked problems*
- **what is a wicked problem?**

10 properties of a wicked problem

(1) *Wicked problems* have no definitive formulation, but every formulation of a *wicked problem* corresponds to the formulation of a solution.
(2) *Wicked problems* have no stopping rules.
(3) Solutions to *wicked problems* cannot be true or false, only good or bad.
(4) In solving *wicked problems* there is no exhaustive list of admissible operations.
(5) For every *wicked problem* there is always more than one possible explanation, with explanations depending on the *Weltanschauung* of the designer.[39]
(6) Every *wicked problem* is a symptom of another, "higher level," problem.[40]
(7) No formulation and solution of a *wicked problem* has a definitive test.
(8) Solving a *wicked problem* is a "one shot" operation, with no room for trial and error.[41]
(9) Every *wicked problem* is unique.
(10) The *wicked problem* solver has no right to be wrong—they are fully responsible for their actions.

- **SubAxes**
  - task scope
    - *broad vs. narrow*
    - *task decomposition*
  - task characterization
    - *shared understanding*
  - task stability
    - *designer influence is disruptive*
      - **abstraction and tools change user's needs: CONTRIBUTION**
    - *user's needs change during project*
      - **DANGER**



TASK CLARITY

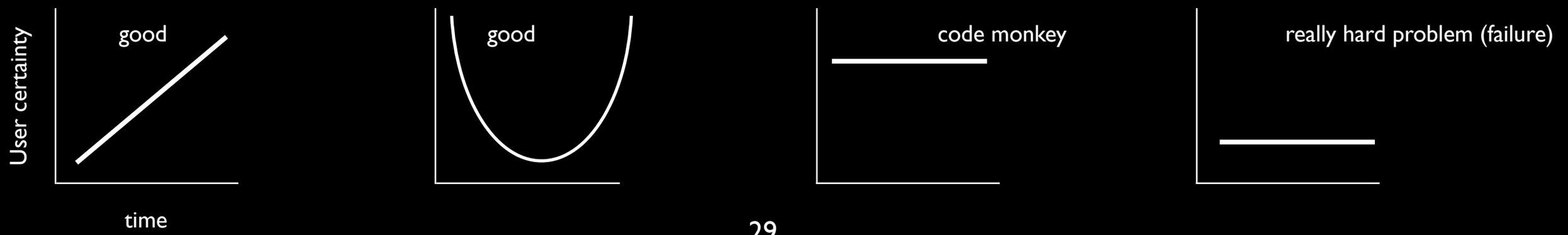
fuzzy — *crisp*

- **SubAxes**
  - task scope
    - *broad vs. narrow*
    - *task decomposition*
  - task characterization
    - *shared understanding*
  - task stability
    - *designer influence is disruptive*
      - **abstraction and tools change user's needs: CONTRIBUTION**
    - *user's needs change during project*
      - **DANGER**

broad — *narrow*

confusion — *shared understanding*

unstable — *stable*

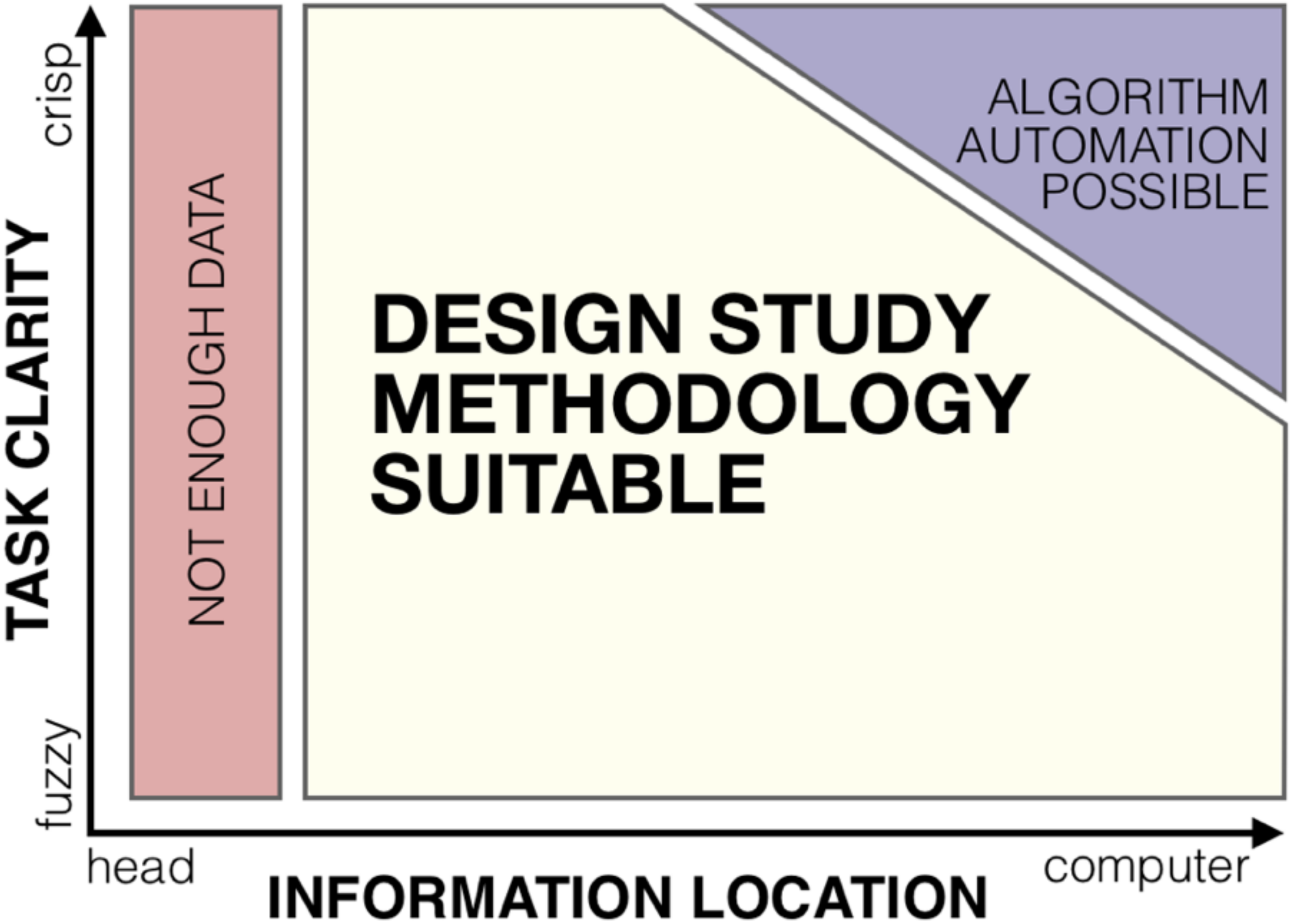


be aware...

- **changing user practice**
  - researcher is actively intervening: *change can be good (contribution!), but might be hard to track*
- **demand characteristics | experimenter bias**
  - the system your are studying is changing by the fact that you are studying it
    - *users change behavior because they know they are being studied*
    - *unconscious bias by experimenters that effect subjects*
      - **Pitfall: "But they liked it ..."**
        - necessary but not sufficient

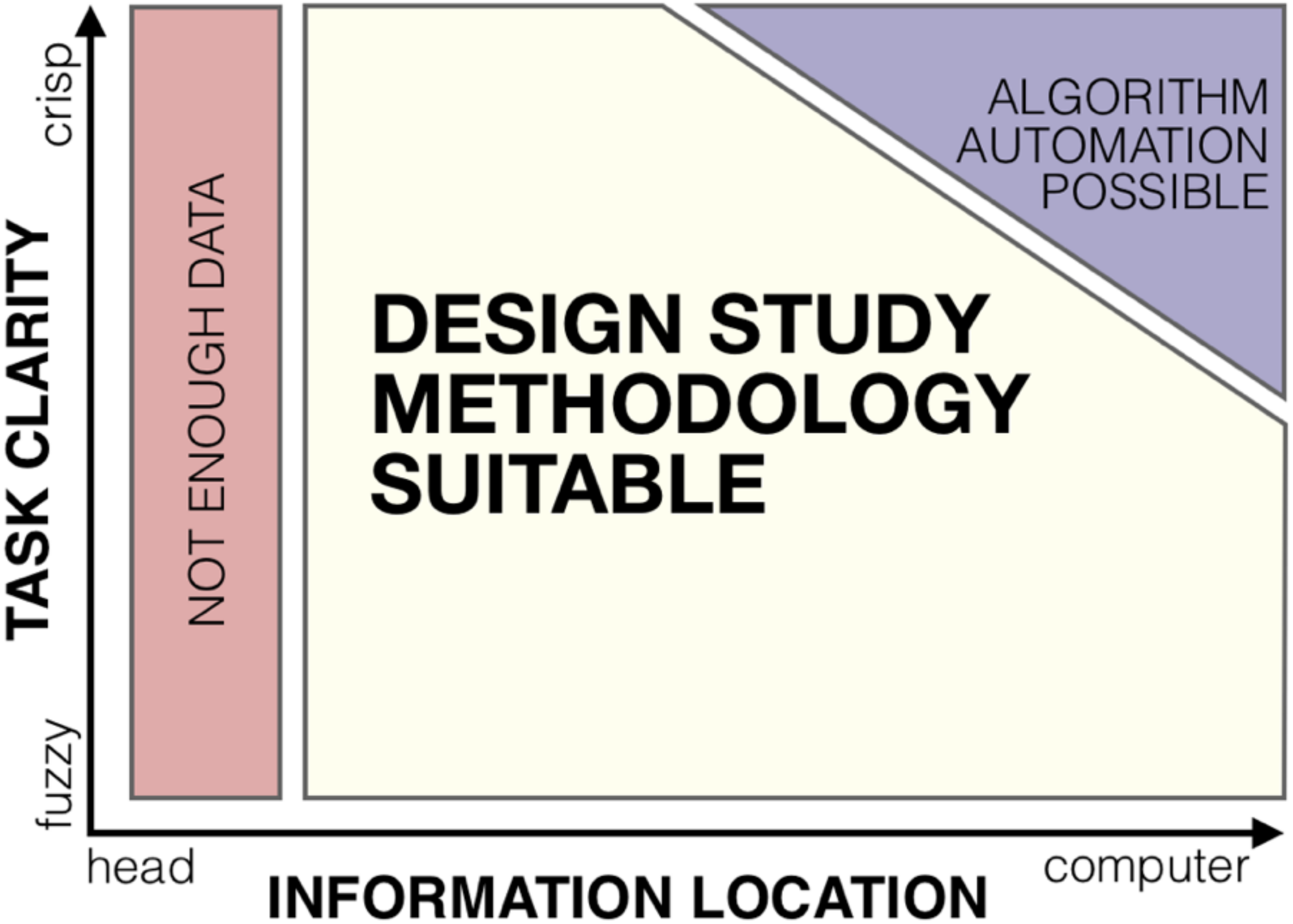
- **how much information is made explicit in digital form vs implicit in user head**
  - more than just ‘data’
    - *metadata*
    - *surrounding knowledge and context*

- **how much information is made explicit in digital form vs implicit in user head**
  - more than just ‘data’
    - *metadata*
    - *surrounding knowledge and context*



You have to be at least this far in order to start designing a visualization solution.

Now it's so well-defined that we can write an algorithm to solve the problem.



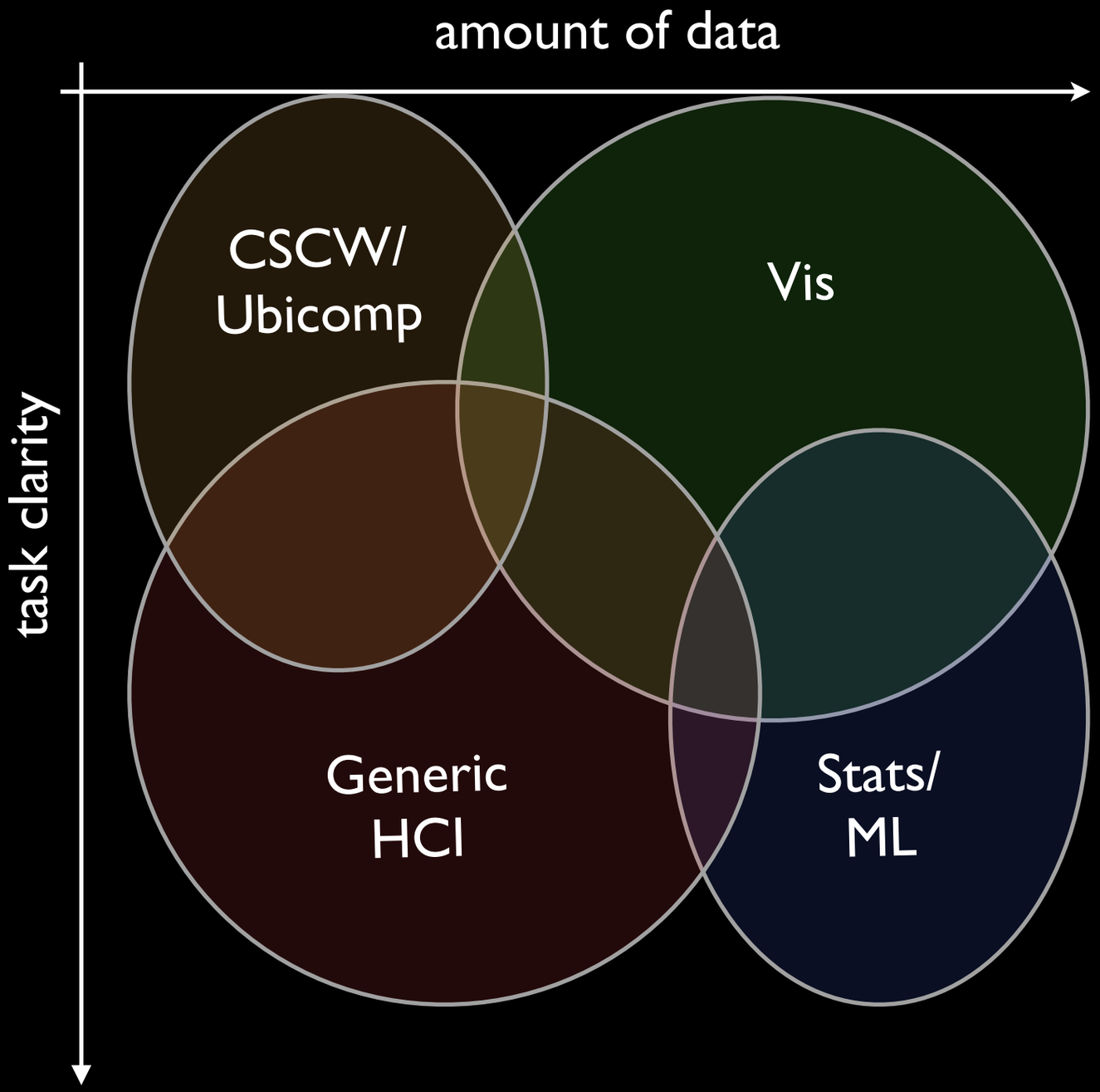
You have to be at least this far in order to start designing a visualization solution.

Now it's so well-defined that we can write an algorithm to solve the problem.

- **movement along the axes**
  - back and forth along task
  - usually only forward along information
- **movement along one axis often causes movement along the other**
  - increased task clarity facilitates understanding of derived data needs
  - increased information articulation facilitates understanding of analysis needs
- **forward movement along the axes as a problem characterization/abstraction contribution**
  - vs. technique driven: you are at a specific point on these axes

- **using task axis to compare vis to other areas**
  - in vis we wanna build tools for ill-defined hard problems (wicked problems), that's what makes us different form Stats or ML
  - we share this with other communities: CSCW, UbiComp
    - *but in these areas: no or little data involved*

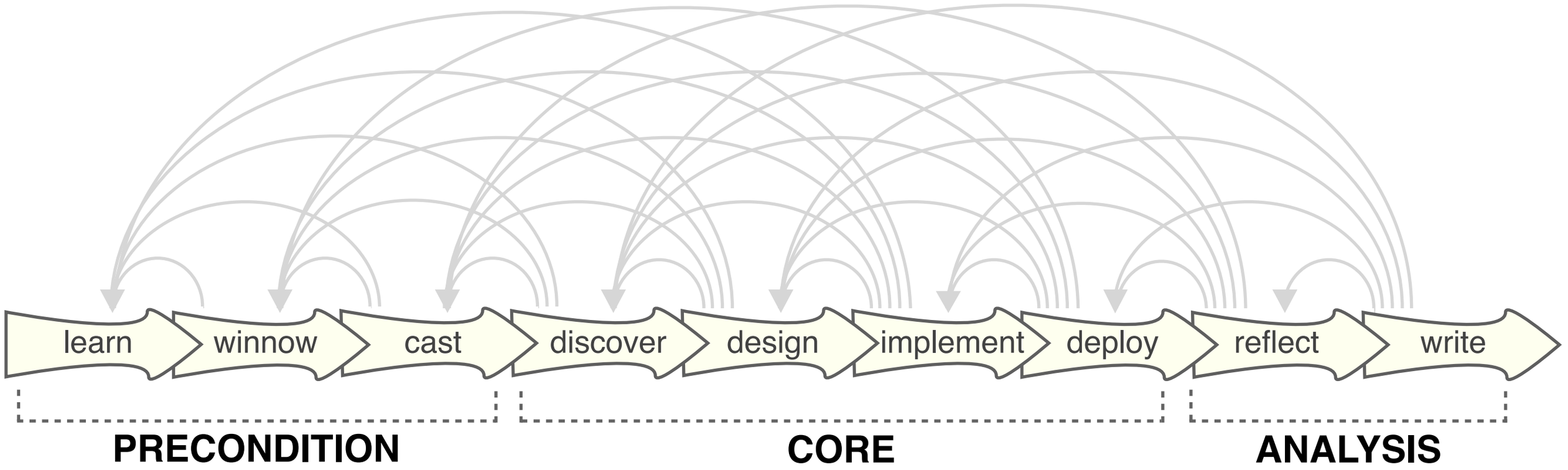
Vis: Relation to other areas/fields

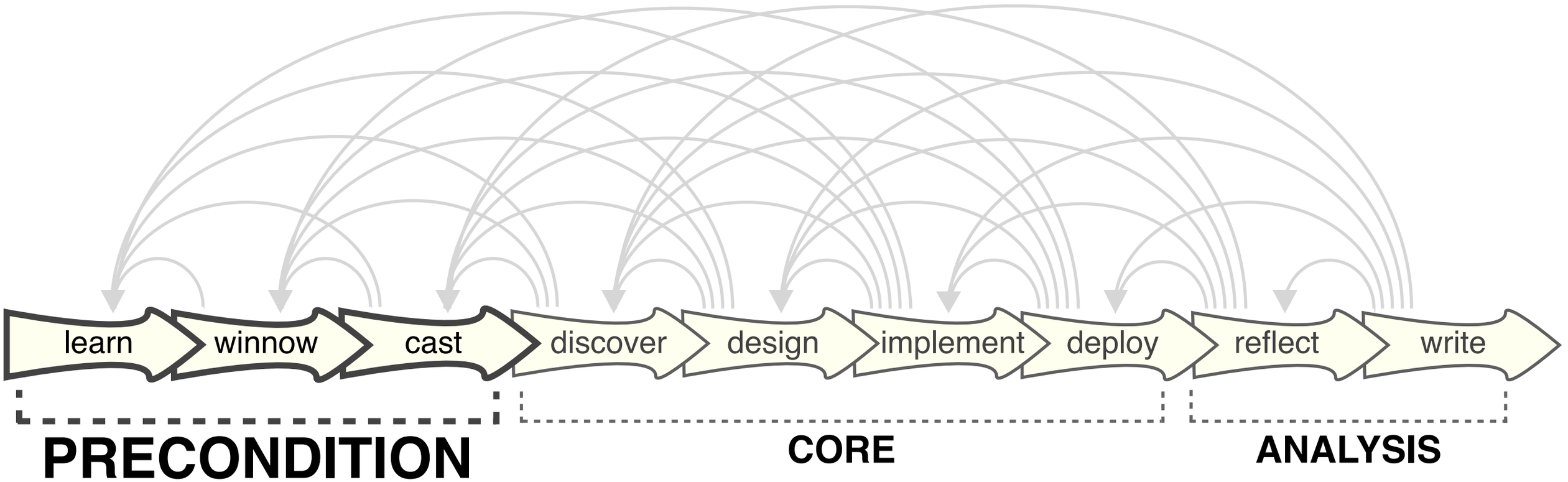


Vis vs.

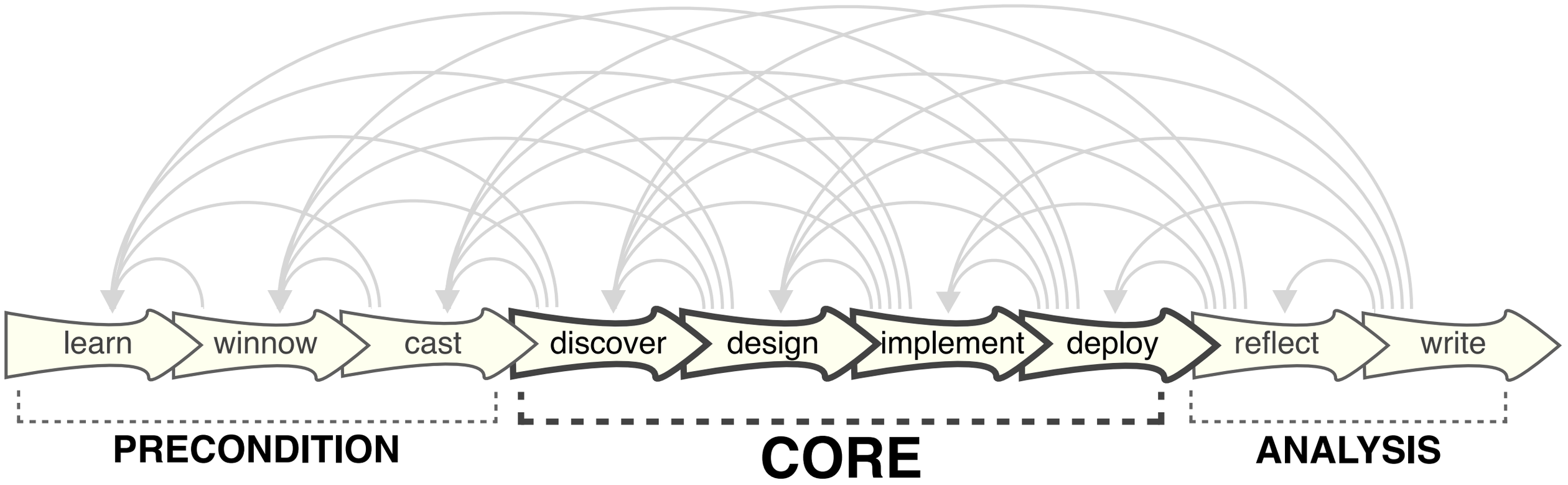
- HCI: The user task is larger, more complex; more data
- CSCW, Ubicomp: Share the squishy task, but it's not about data analysis
- Stats/ML: Data Analysis but crisp task

NINE-STAGE FRAMEWORK

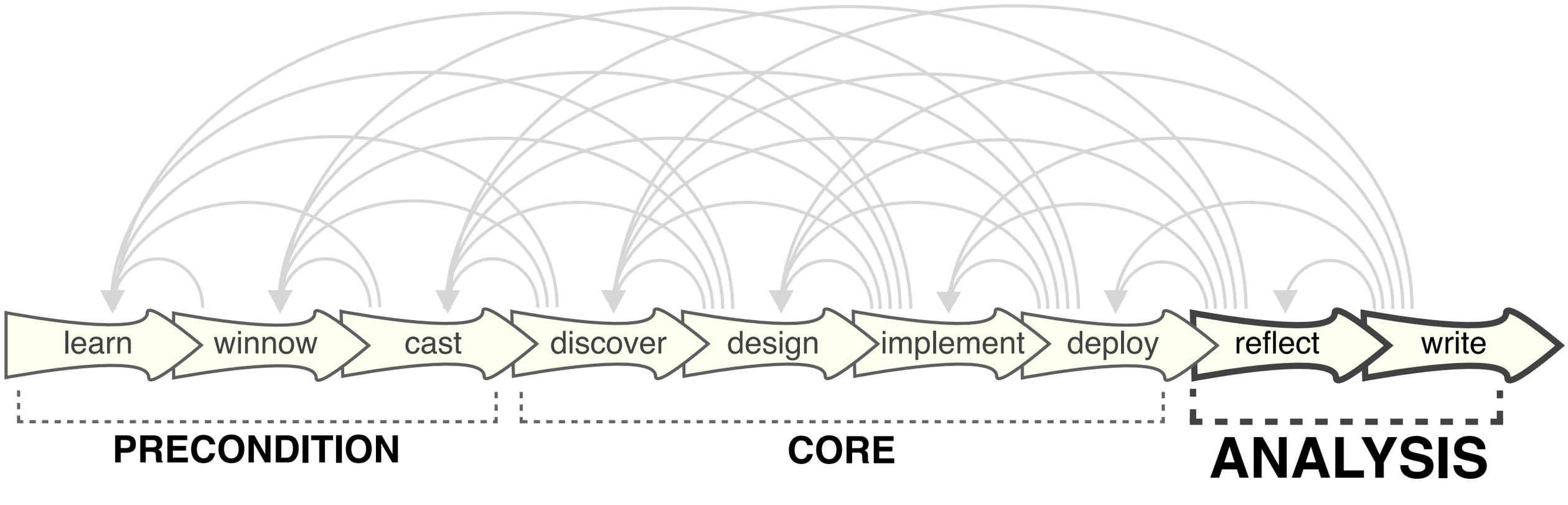

learn
winnow
cast
discover
design
implement
deploy
reflect
write
PRECONDITION
CORE
ANALYSIS



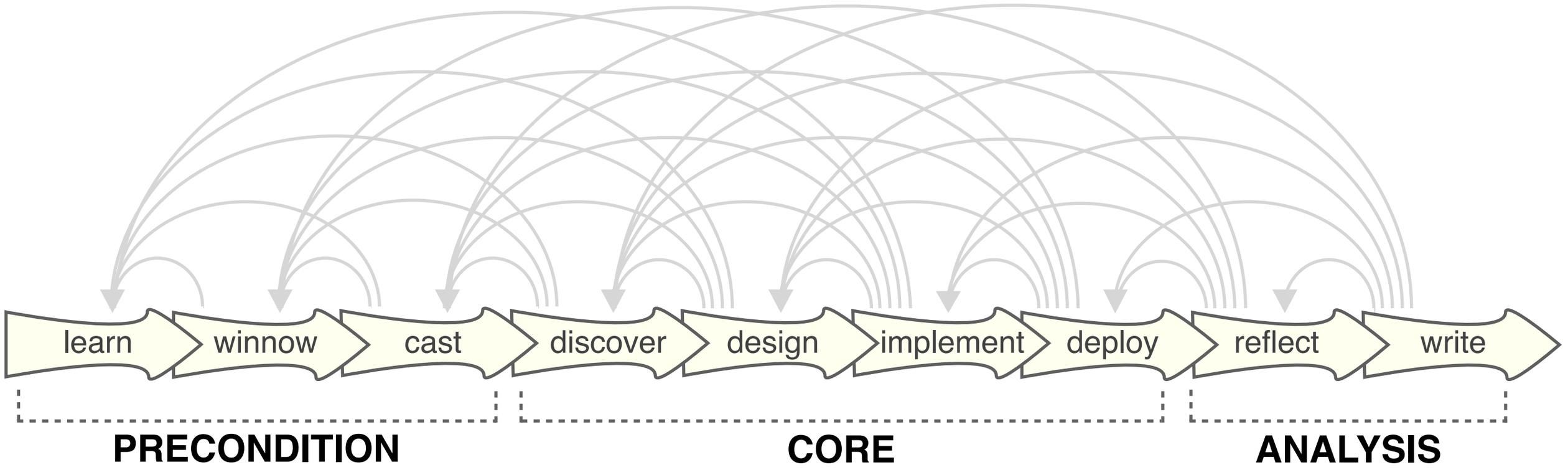
what must be done before starting a project

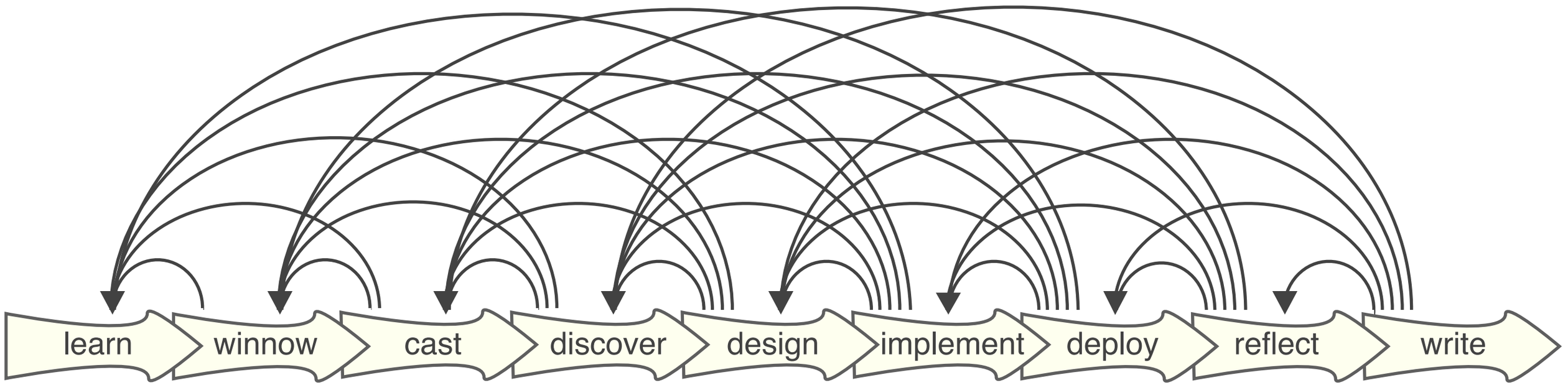


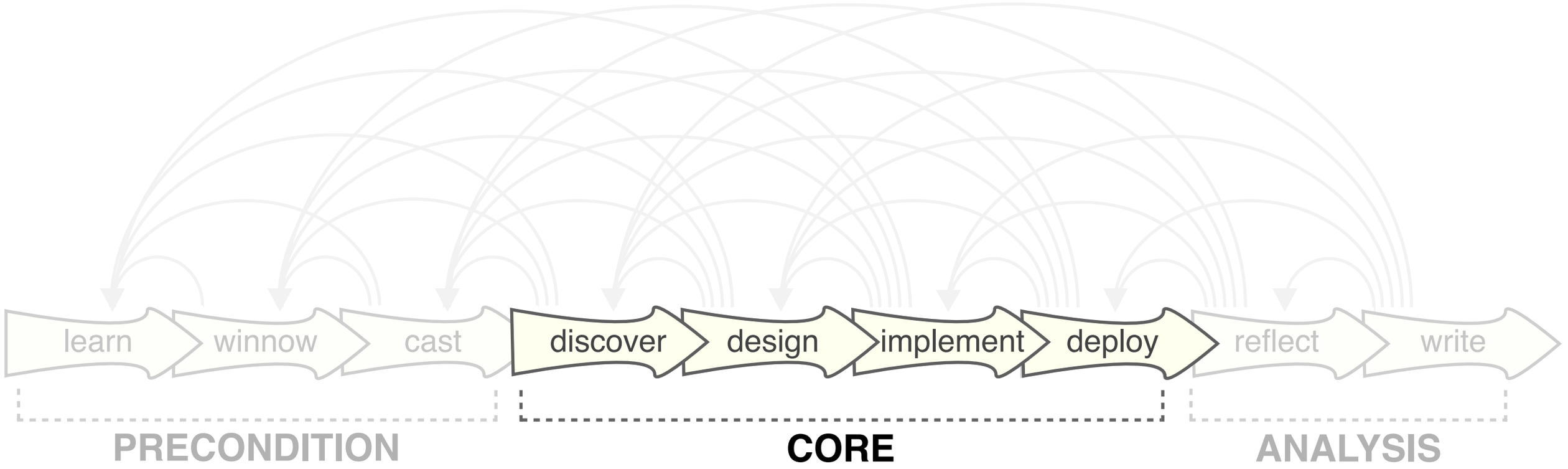
main steps of a design study



analytical reasoning at the end


learn
winnow
cast
discover
design
implement
deploy
reflect
write
PRECONDITION
CORE
ANALYSIS


learn
winnow
cast
discover
design
implement
deploy
reflect
write
PRECONDITION
CORE
ANALYSIS


learn
winnow
cast
discover
design
implement
deploy
reflect
write
PRECONDITION
CORE
ANALYSIS

Eurographics/ IEEE-VGTC Symposium on Visualization 2010
G. Melançon, T. Munzner, and D. Weiskopf
(Guest Editors)

Volume 29 (2010), Number 3

Pathline: A Tool For Comparative Functional Genomics

M. Meyer[1,2], B. Wong[2], M. Styczynski[3], T. Munzner[4], and H. Pfister[1]

[1]Harvard University, USA
[2]Broad Institute, USA
[3]Georgia Institute of Technology, USA
[4]University of British Columbia, Canada

Abstract

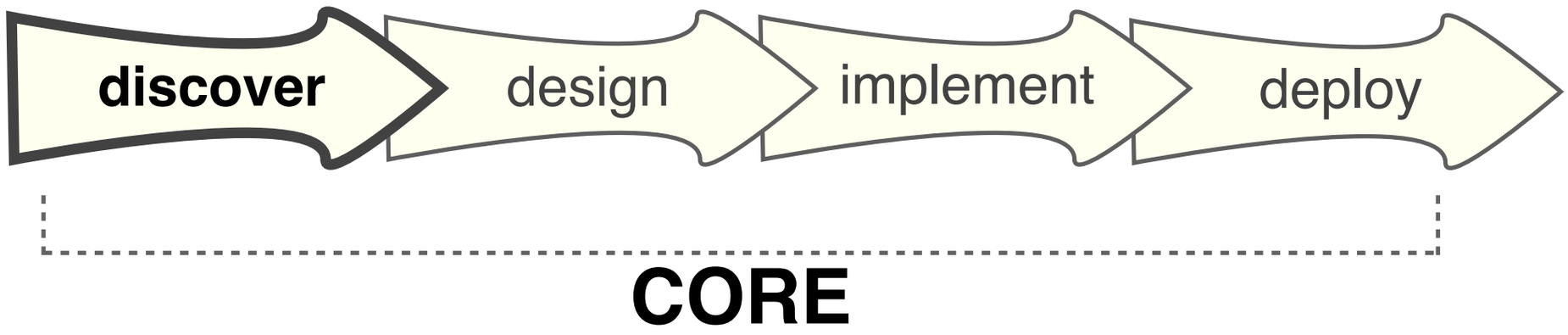
Biologists pioneering the new field of comparative functional genomics attempt to infer the mechanisms of gene regulation by looking for similarities and differences of gene activity over time across multiple species. They use three kinds of data: functional data such as gene activity measurements, pathway data that represent a series of reactions within a cellular process, and phylogenetic relationship data that describe the relatedness of species. No existing visualization tool can visually encode the biologically interesting relationships between multiple pathways, multiple genes, and multiple species. We tackle the challenge of visualizing all aspects of this comparative functional genomics dataset with a new interactive tool called Pathline. In addition to the overall characterization of the problem and design of Pathline, our contributions include two new visual encoding techniques. One is a new method for linearizing metabolic pathways that provides appropriate topological information and supports the comparison of quantitative data along the pathway. The second is the curvemap view, a depiction of time series data for comparison of gene activity and metabolite levels across multiple species. Pathline was developed in close collaboration with a team of genomic scientists. We validate our approach with case studies of the biologists' use of Pathline and report on how they use the tool to confirm existing findings and to discover new scientific insights.

Categories and Subject Descriptors (according to ACM CCS): I.3.3 [Computer Graphics]: Picture/Image Generation—Line and curve generation

1. Introduction

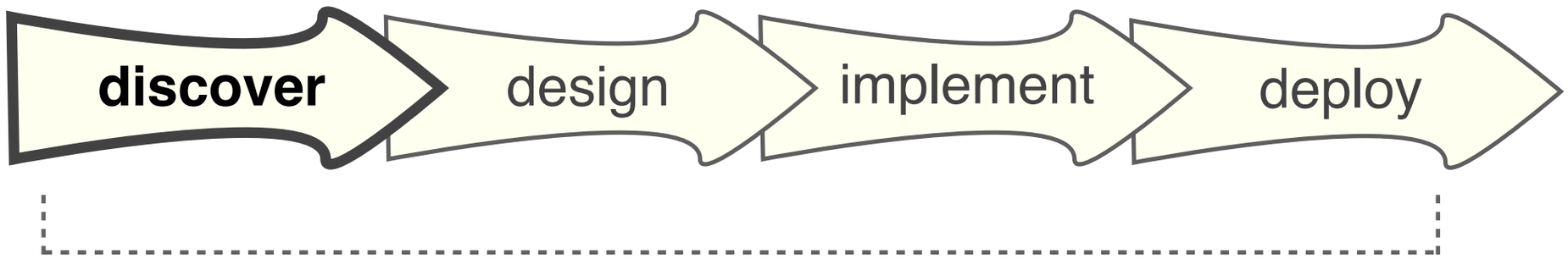
Biologists conduct comparative functional genomics studies to understand the mechanisms of gene regulation. They need to analyze the levels of gene activity and metabolites belonging to multiple pathways over time and across multiple species. Their visualization needs were not met

problem characterization & abstraction



problem characterization & abstraction

“data counseling”



functional genomics

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

functional genomics data

gene expression
molecular pathways

functional genomics data

gene expression

molecular pathways

gene expression is ...

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

gene expression is ...

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

... a single quantitative value

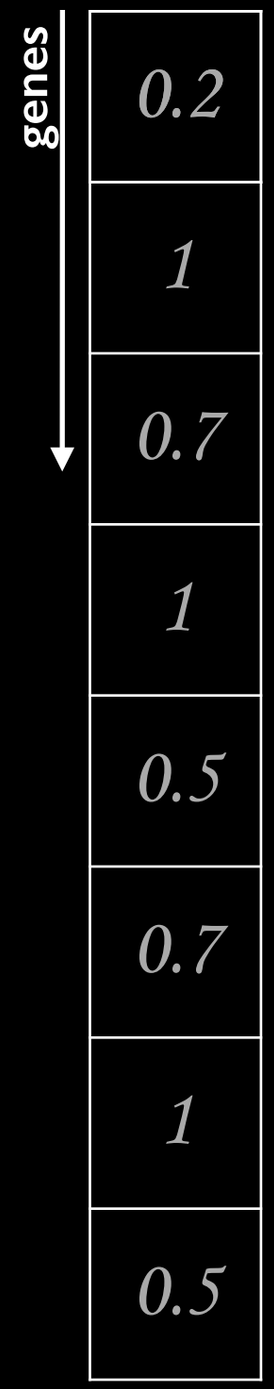
gene expression is ...

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

... a single quantitative value

biologists measure it ...

... for many genes



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)

samples

genes

0.2	0.4	1	1	1	0.8
1	0	0	0	1	1
0.7	0.8	1	1	0.8	0.6
1	0	0.2	0.5	1	1
0.5	0.8	0.5	0.3	0.5	0.8
0.7	0.5	0.8	0.7	1	1
1	0.3	0.4	1	1	1
0.5	0	0	0.7	0.5	0.3

gene expression is ...

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

... a single quantitative value

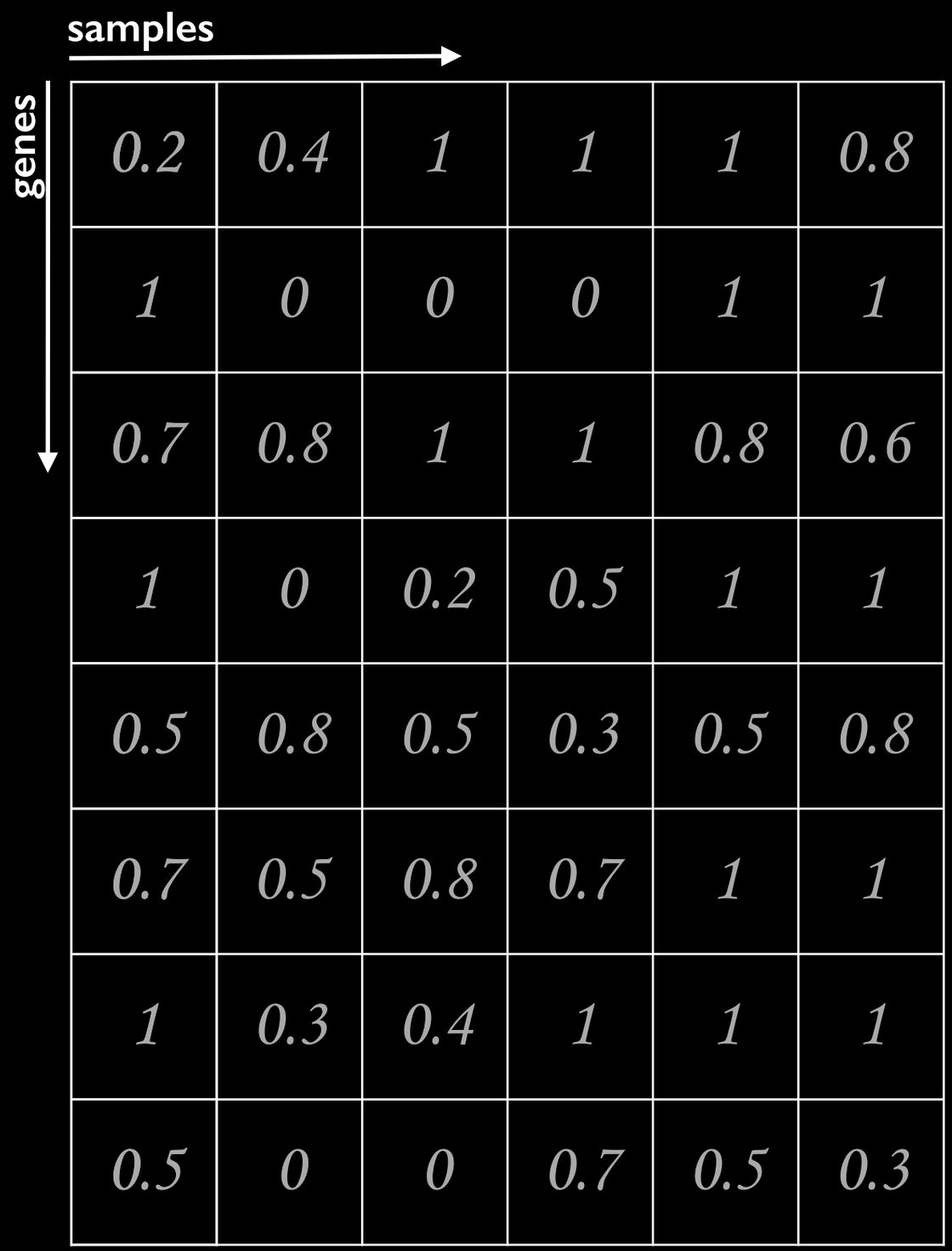
biologists measure it ...

... for many genes

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

visualized with heatmaps

encode value with color



gene expression is ...

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

... a single quantitative value

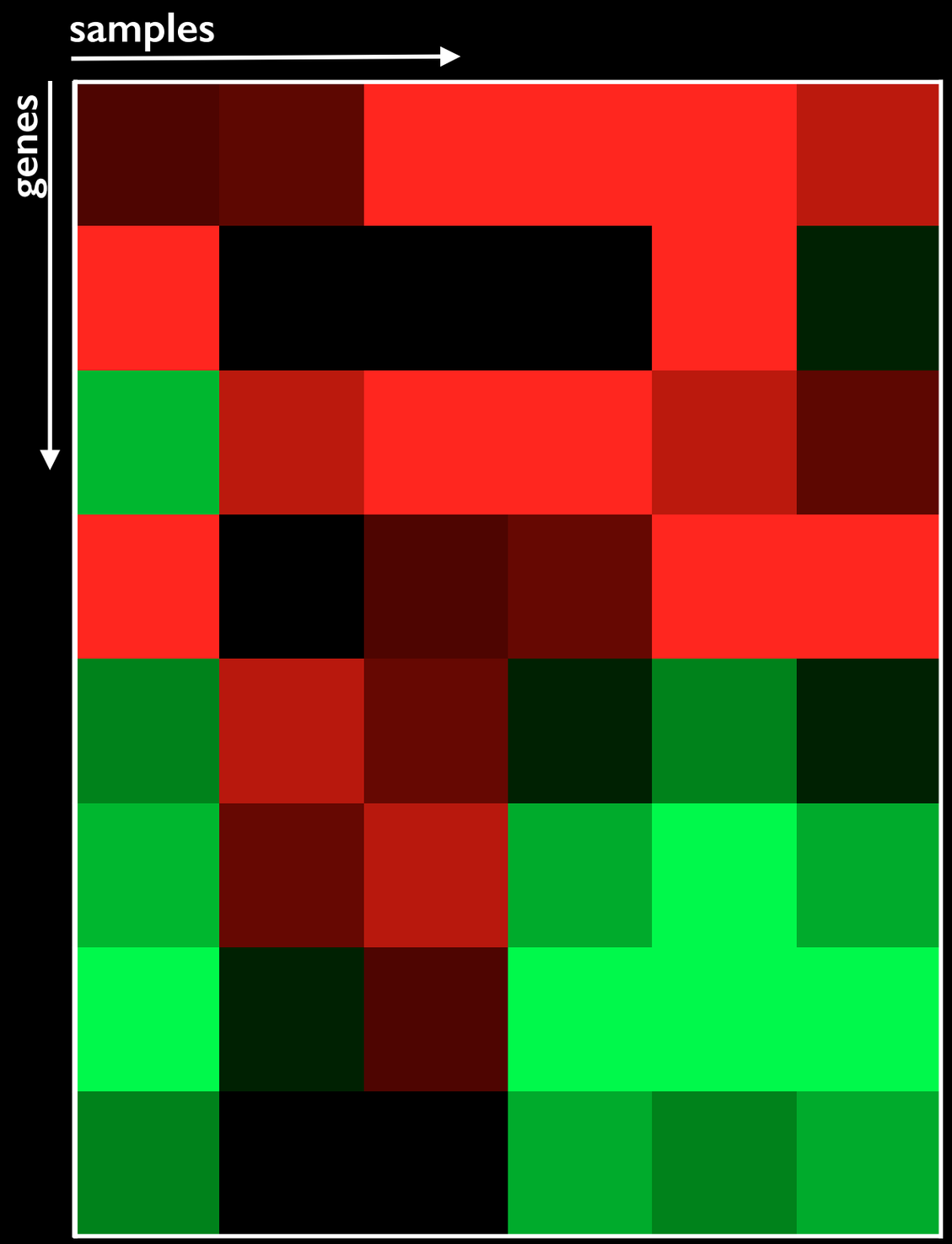
biologists measure it ...

... for many genes

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

visualized with heatmaps

encode value with color



gene expression is ...

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

... a single quantitative value

biologists measure it ...

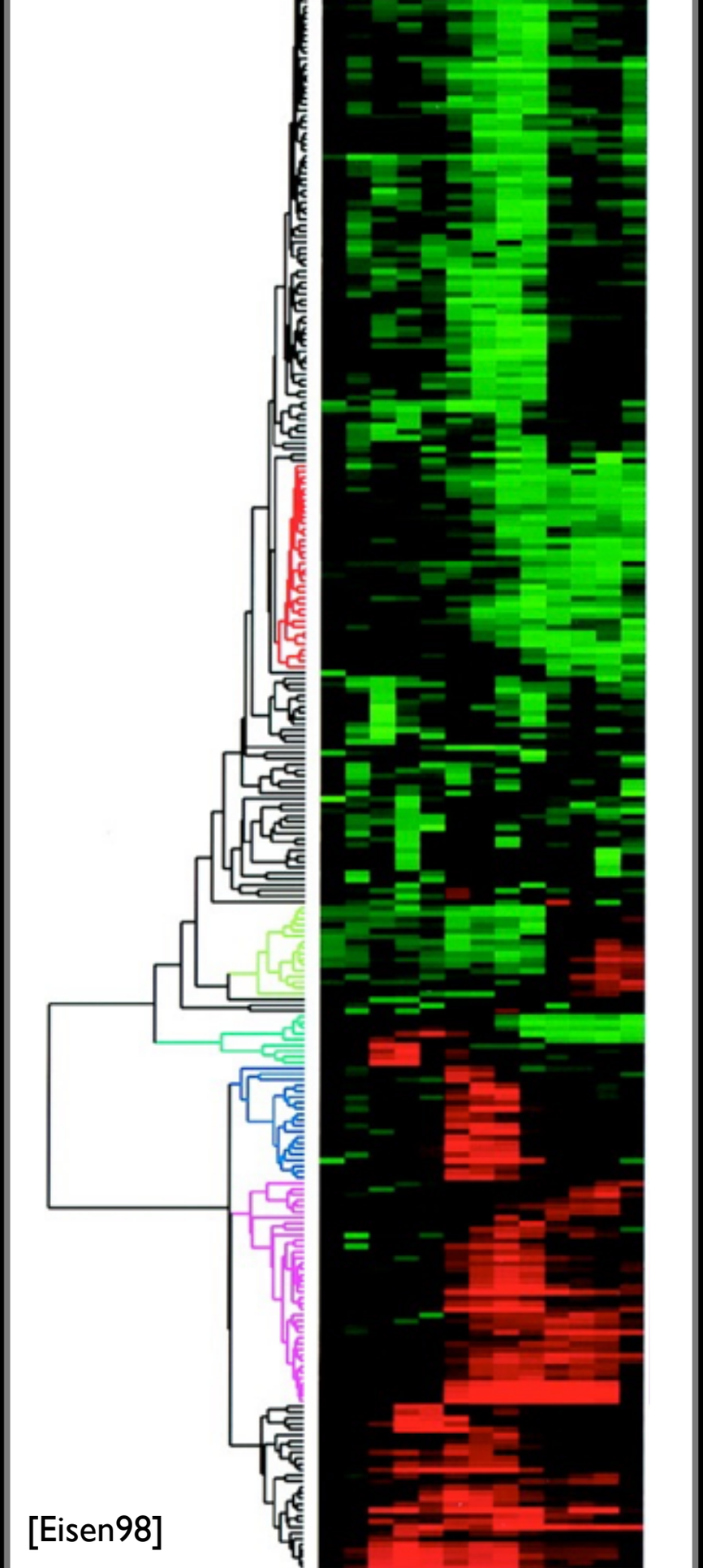
... for many genes

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

visualized with heatmaps

encode value with color

augmented with clustering



functional genomics data

gene expression

molecular pathways

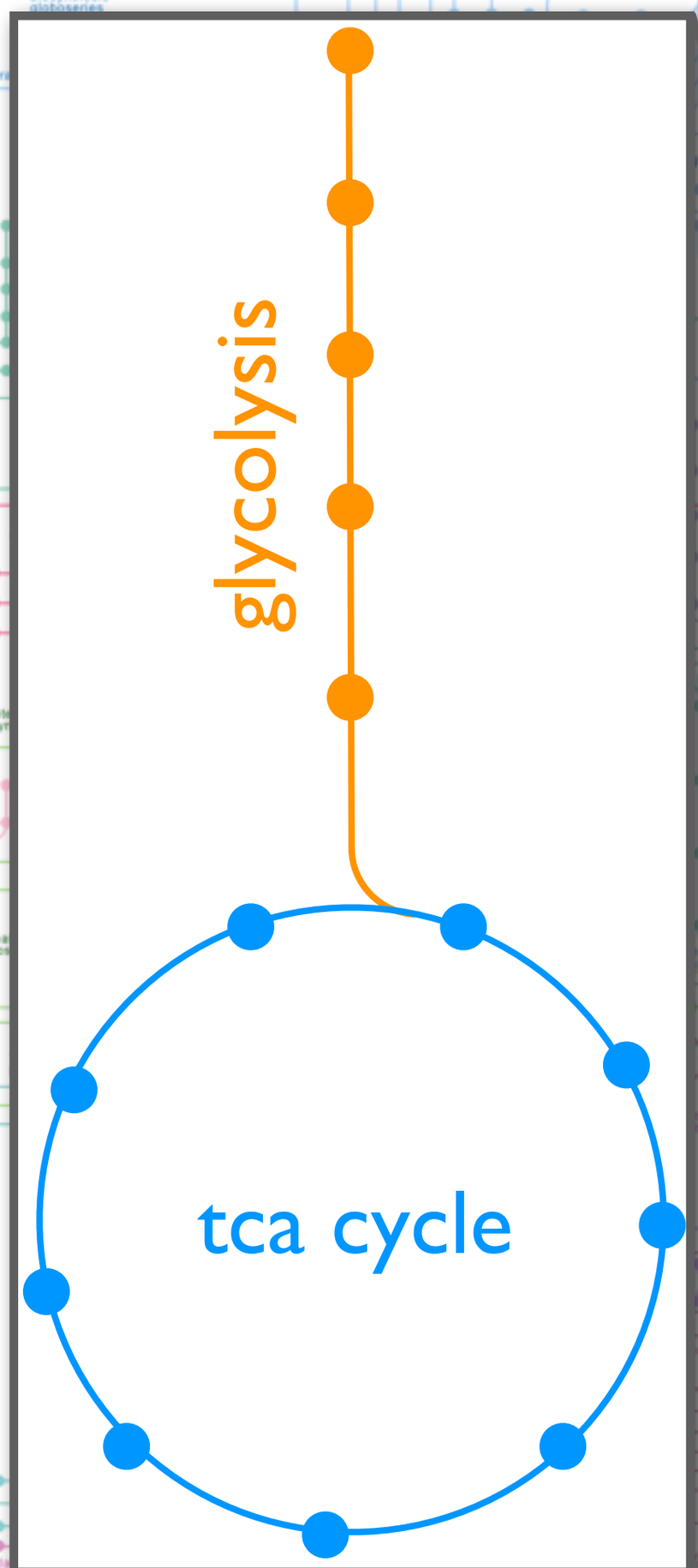
functional genomics data

gene expression

molecular pathways

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

Nucleotide Metabolism
Carbohydrate Metabolism
Amino Acid Metabolism
Lipid Metabolism
Energy Metabolism
Metabolism of Other Amino Acid
Glycosphingolipid biosynthesis - globoseries
Other glycan degradation
N-glycan biosynthesis
Glycosaminoglycan biosynthesis - chondroitin sulfate
O-Glycan biosynthesis
Glycosaminoglycan biosynthesis - heparan sulfate
Glycosaminoglycan biosynthesis - keratan sulfate
Glycosaminoglycan degradation
Galactose metabolism
Starch and sucrose metabolism
Streptomycin biosynthesis
Biosynthesis of ansamycins
Puromycin biosynthesis
Purine metabolism
Ascorbate and aldarate metabolism
Caffeine metabolism
Glycosphingolipid biosynthesis - ganglioseries
Arachidonic acid metabolism
alpha-Linolenic acid metabolism
Glycosylphosphatidylinositol (GPI)-anchor biosynthesis
Polyketide sugar unit biosynthesis
Pentose and glucuronate interconversions
beta-Alanine metabolism
Pyrimidine metabolism
Tyrosine metabolism
Riboflavin metabolism
Biosynthesis of vancomycin group antibiotics
Amino sugar and nucleotide sugar metabolism
Fructose and mannose metabolism
Pentose phosphate metabolism
Biosynthesis of 12-, 14- and 16-membered macrolides
Phenylalanine, tyrosine and tryptophan biosynthesis
Glycerophospholipid metabolism
Phenylalanine metabolism
Photosynthesis - antenna proteins
Novobiocin biosynthesis
Acridone alkaloid biosynthesis
[B] Lipids
Glycerolipid metabolism
Ether lipid metabolism
Sphingolipid metabolism
Insitol phosphate metabolism
Glycolysis / Gluconeogenesis
Photosynthesis
Sulfur metabolism
Taurine and hypotaurine metabolism
Biosynthesis of unsaturated fatty acids
[B] Photosynthesis Proteins
Biosynthesis of siderophore group nonribosomal peptides
Tryptophan metabolism
Fatty acid biosynthesis
Fatty acid metabolism
Peptidoglycan biosynthesis
Carotenoid biosynthesis
Phosphonate and phosphinate metabolism
Histidine metabolism
Retinol metabolism
D-Alanine metabolism
3-Chloroacrylic acid degradation
Benzoxazinone biosynthesis
Biosynthesis of type II polyketide products
Pyruvate metabolism
Glycine, serine and threonine metabolism
Valine, leucine and isoleucine biosynthesis
Biosynthesis of type II polyketide backbone
Synthesis and degradation of Ketone bodies
Selenoamino acid metabolism
Cyanoamino acid metabolism
Zeatin biosynthesis
Butanoate metabolism
Fatty acid elongation in mitochondria
Propanoate metabolism
Lysine degradation
Valine, leucine and isoleucine degradation
Lipoic acid metabolism
Tetrachloroethene degradation
Reductive carboxylate cycle (CO2 fixation)
Terpenoid backbone biosynthesis
Caprolactam degradation
Carbon fixation
C5-Branched dibasic acid metabolism
Lysine biosynthesis
Cysteine and methionine metabolism
Nitrogen metabolism
Steroid biosynthesis
Methane metabolism
Citrate cycle (TCA cycle)
Primary bile acid biosynthesis
Atrazine degradation
Alanine, aspartate and glutamate metabolism
Steroid hormone biosynthesis
Glyoxylate and dicarboxylate metabolism
Tetracycline biosynthesis
Glycosaminate biosynthesis
1,4-Dichlorobenzene degradation
Naphthalene and anthracene degradation
Arginine and proline metabolism
Biphenyl degradation
1,2-Dichloroethane degradation
gamma-Hexachlorocyclohexane degradation
Bisphenol A degradation
Glutathione metabolism
Fluorobenzoate degradation
Styrene degradation
Tropane, piperidine and pyridine alkaloid biosynthesis
Fluorene degradation
Oxidative phosphorylation
Penicillin and cephalosporin biosynthesis
D-Glutamine and D-glutamate metabolism
Biotin metabolism
2,4-Dichlorobenzoate degradation
Toluene and xylene degradation
Ethylbenzene degradation
1,1,1-Trichloro-2,2-bis(4-chlorophenyl) ethane (DDT) degradation
beta-Lactam resistance
Benzoate degradation via hydroxylation
Carbazole degradation
www.genome.jp/kegg/



pathways

www.genome.jp/kegg/

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

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

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?

process

- semistructured & contextual interviews

- 4 biologists
  - 2 experimental, 2 computational

- 3-4 hours per week for a month

- in parallel with design stage

metabolic pathways

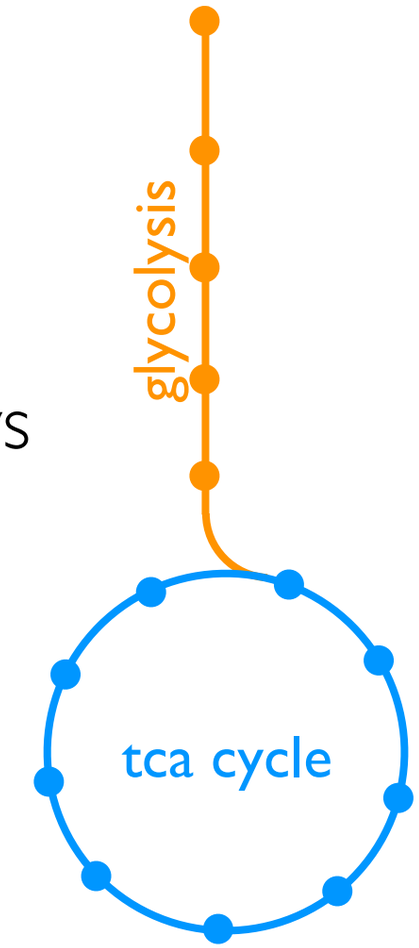
gene expression

similarity scores

phylogeny

metabolic pathways

- 10 to 50 pathways of interest



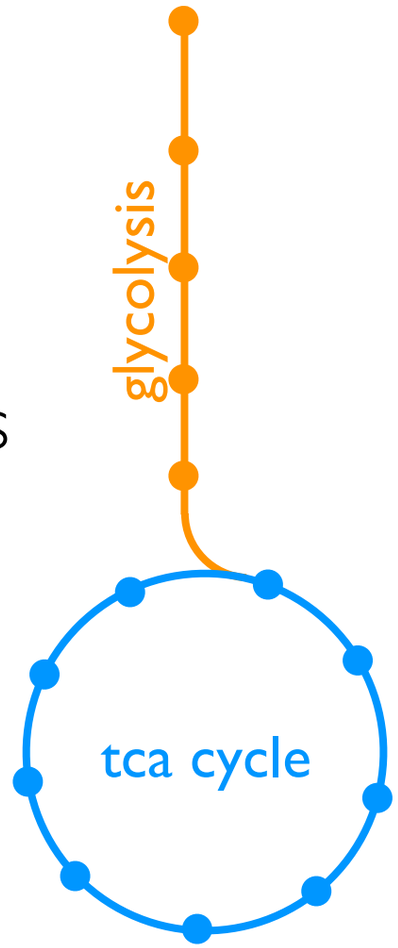
gene expression

similarity scores

phylogeny

metabolic pathways

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



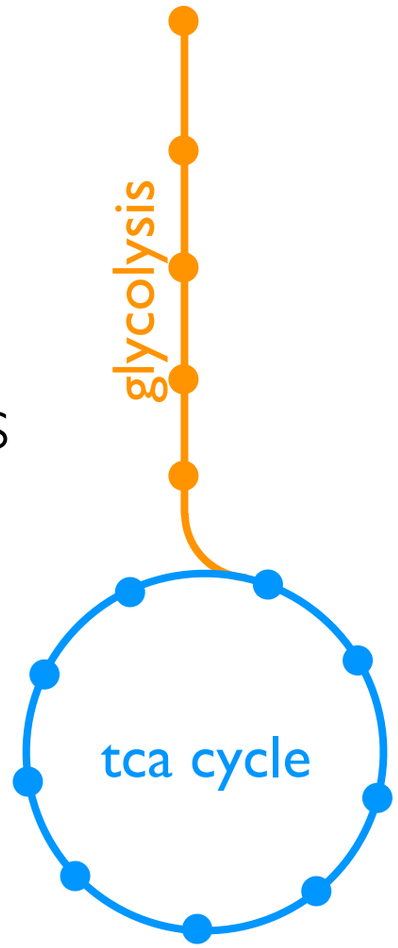
gene expression

similarity scores

phylogeny

metabolic pathways

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



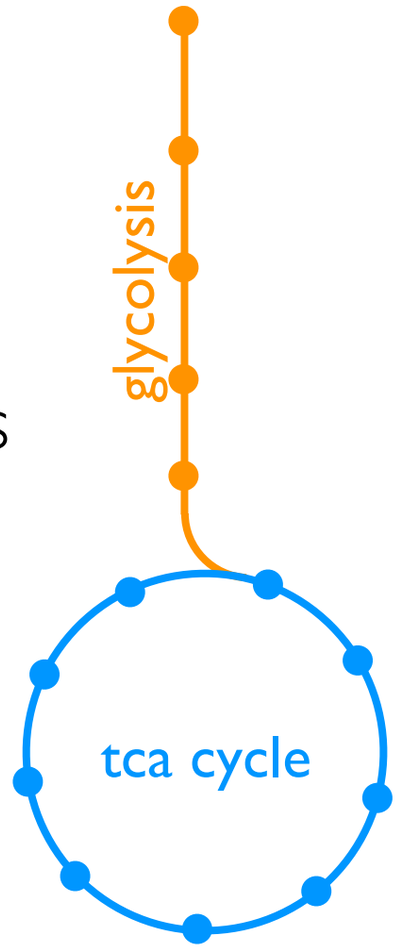
gene expression

similarity scores

phylogeny

metabolic pathways

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



gene expression

- 6000 genes and 140 metabolites

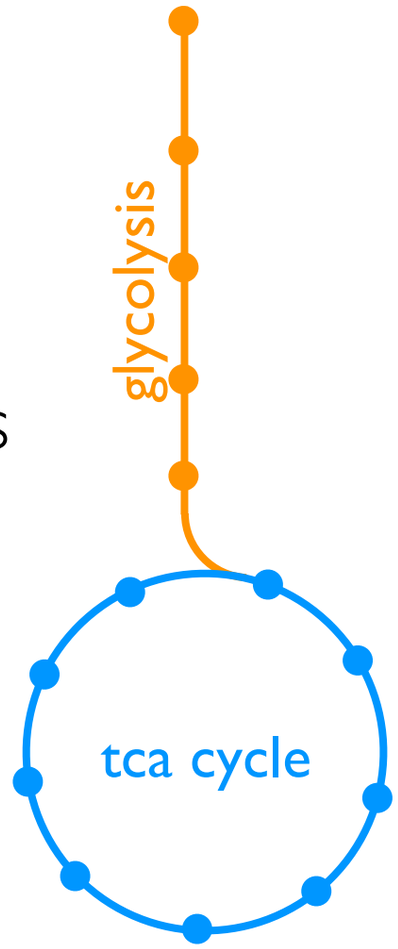
	t1	t2	t3	t4	t5	t6
g1	0.2	0.4	1	1	1	1
m1	1	0	0	0	1	0.8
g2	-0.7	0.8	1	1	0.8	0.2
m2	1	0	0.2	0.5	1	0.2
g3	-0.5	0.8	0.5	-0.3	-0.5	-0.5
m3	-0.7	0.5	0.8	-0.7	-1	0.5

similarity scores

phylogeny

metabolic pathways

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



gene expression

- 6000 genes and 140 metabolites
- 6 time points

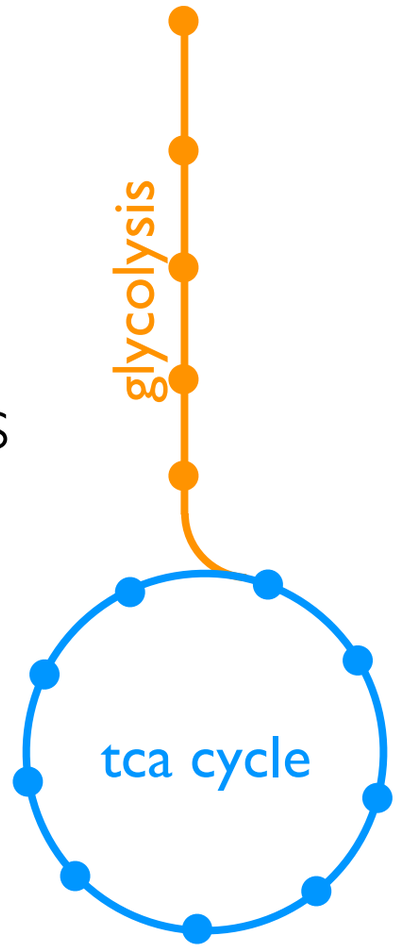
	t1	t2	t3	t4	t5	t6
g1	0.2	0.4	1	1	1	1
m1	1	0	0	0	1	0.8
g2	-0.7	0.8	1	1	0.8	0.2
m2	1	0	0.2	0.5	1	0.2
g3	-0.5	0.8	0.5	-0.3	-0.5	-0.5
m3	-0.7	0.5	0.8	-0.7	-1	0.5

similarity scores

phylogeny

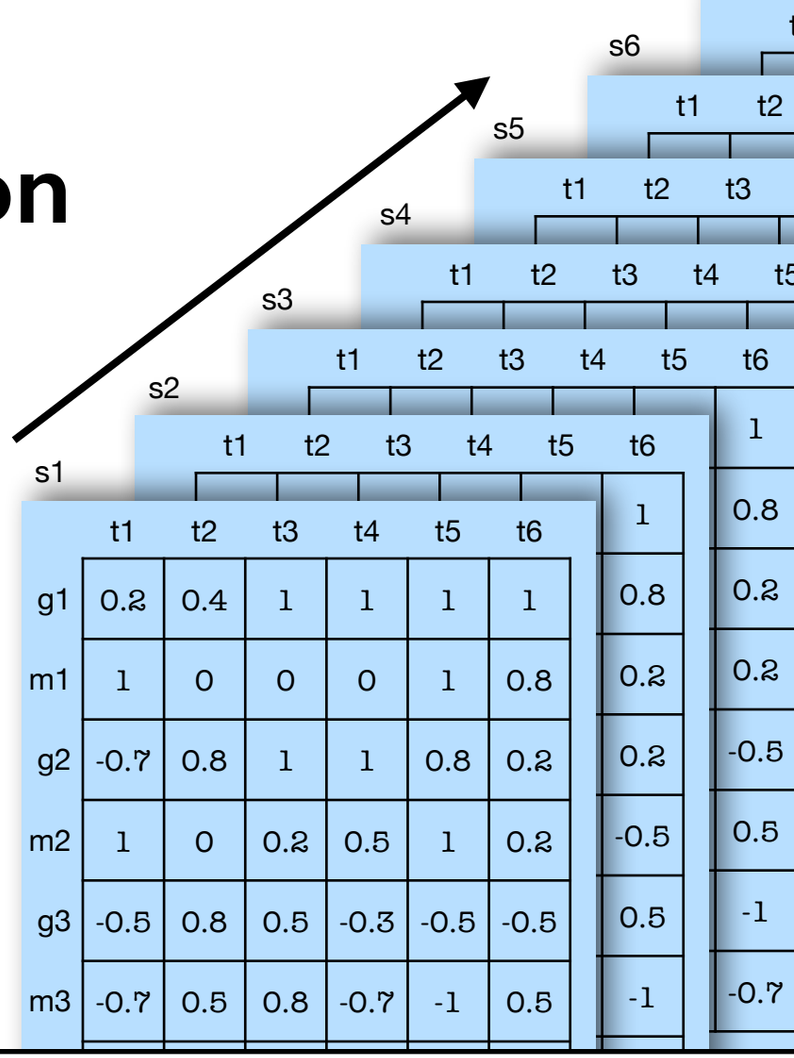
metabolic pathways

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



gene expression

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

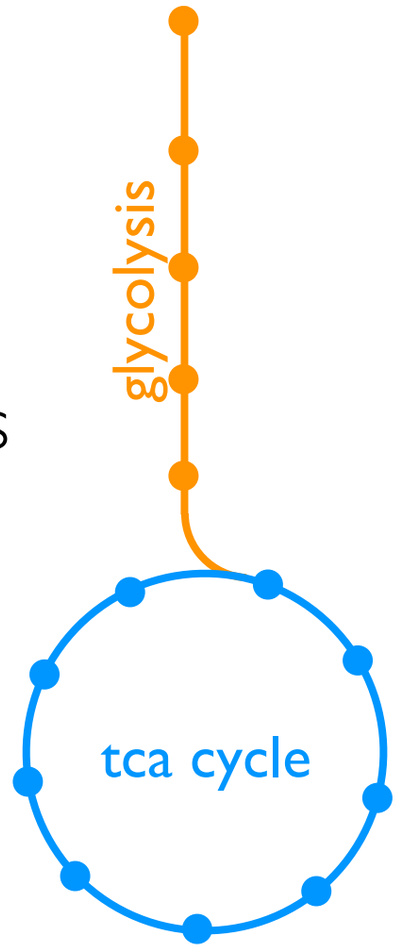


similarity scores

phylogeny

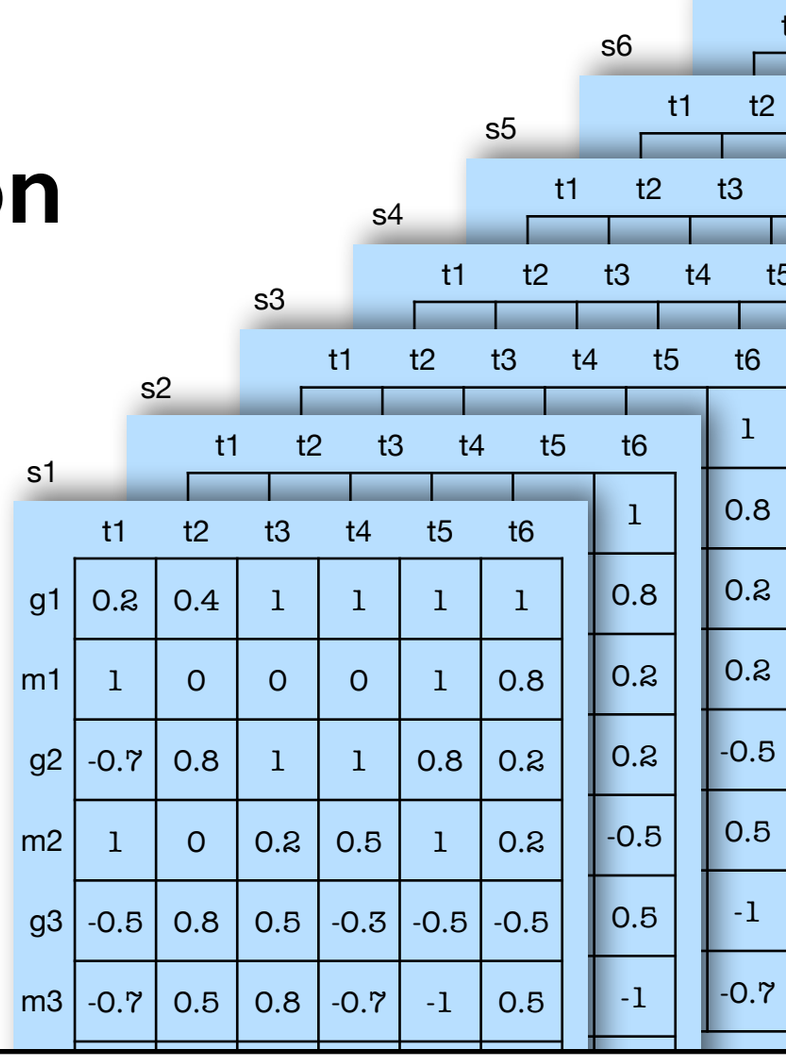
metabolic pathways

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



gene expression

- 6000 genes and 140 metabolites
- 6 time points
- 14 species of yeast
- **3D table**

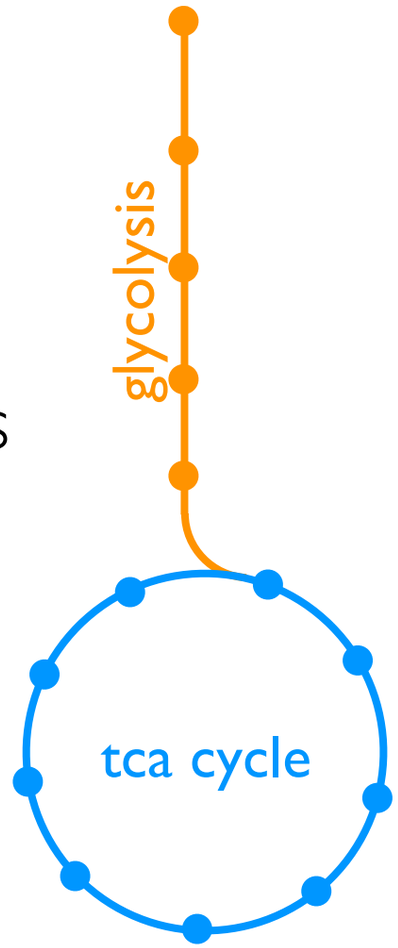


similarity scores

phylogeny

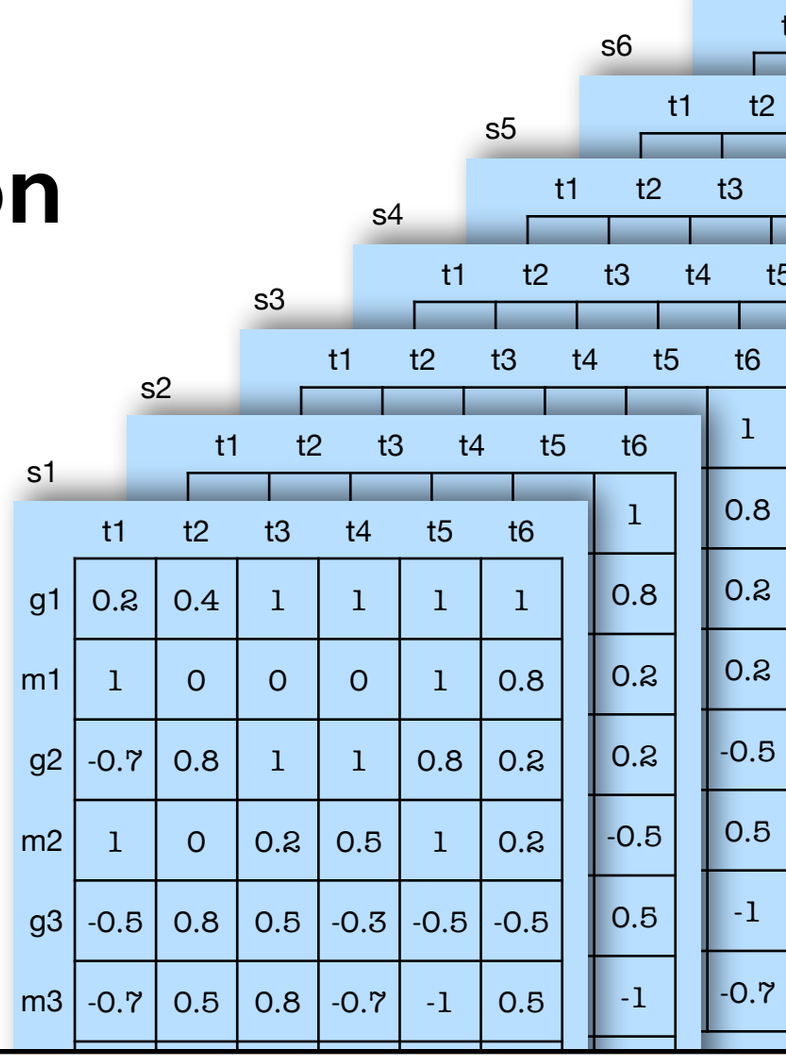
metabolic pathways

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



gene expression

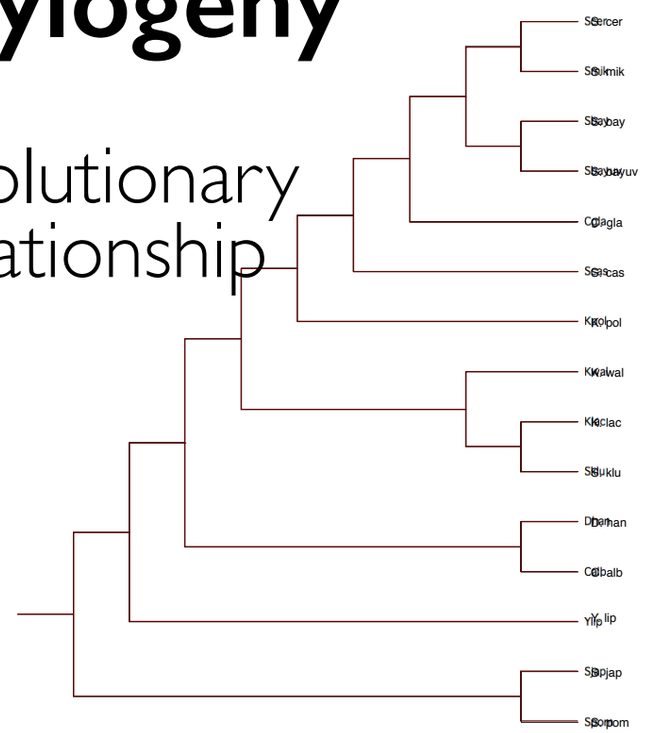
- 6000 genes and 140 metabolites
- 6 time points
- 14 species of yeast
- **3D table**



similarity scores

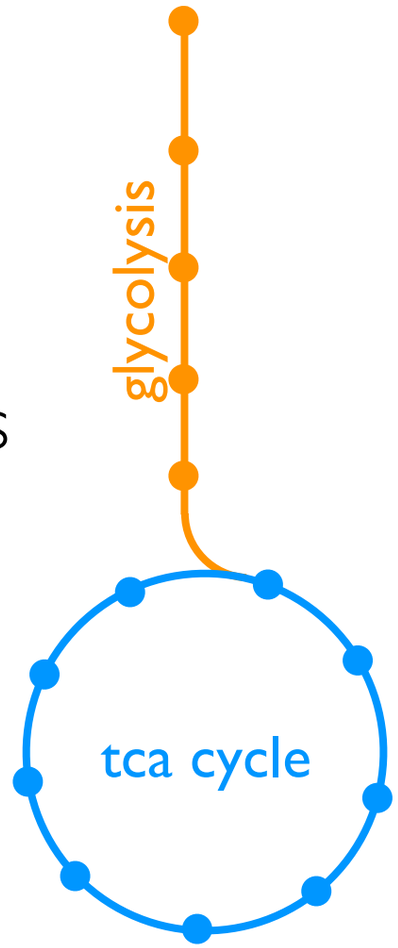
phylogeny

- evolutionary relationship



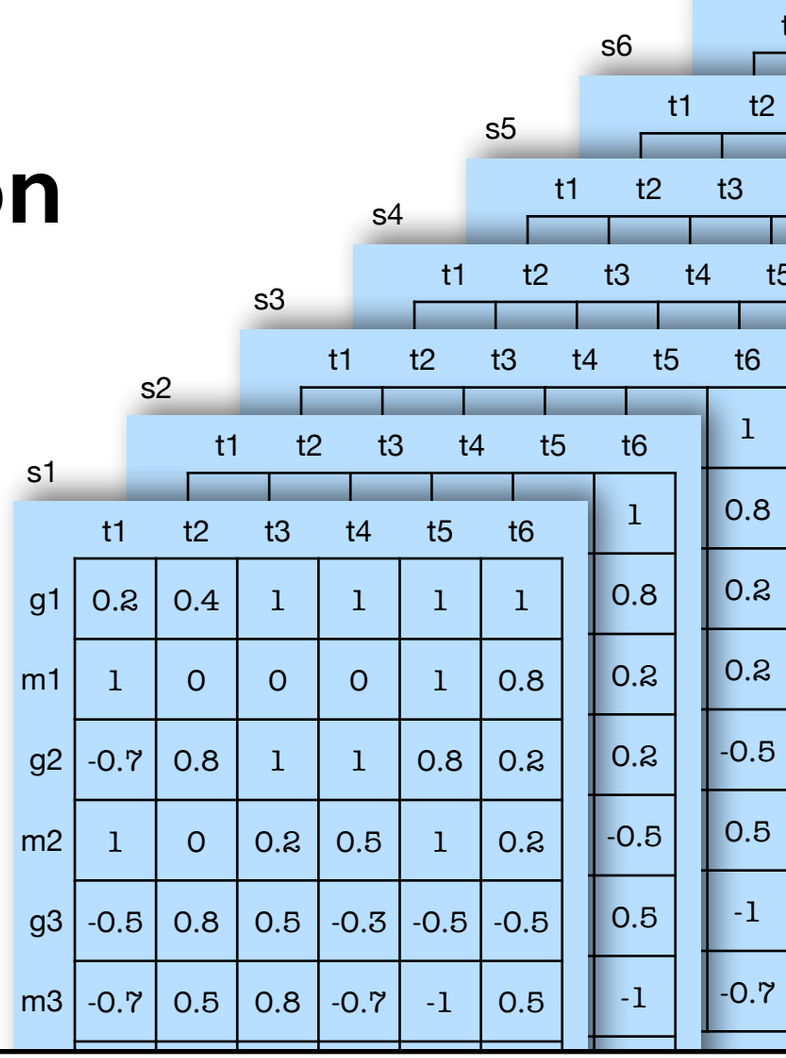
metabolic pathways

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



gene expression

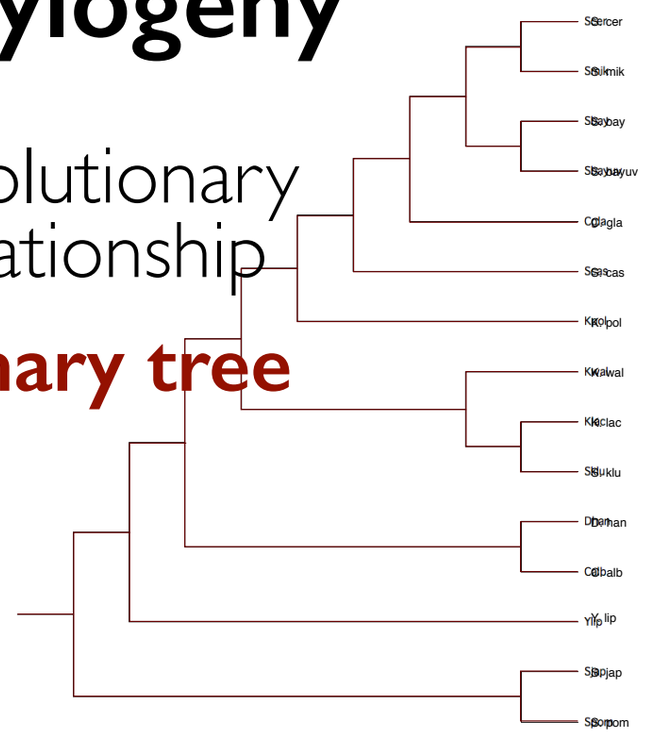
- 6000 genes and 140 metabolites
- 6 time points
- 14 species of yeast
- **3D table**



similarity scores

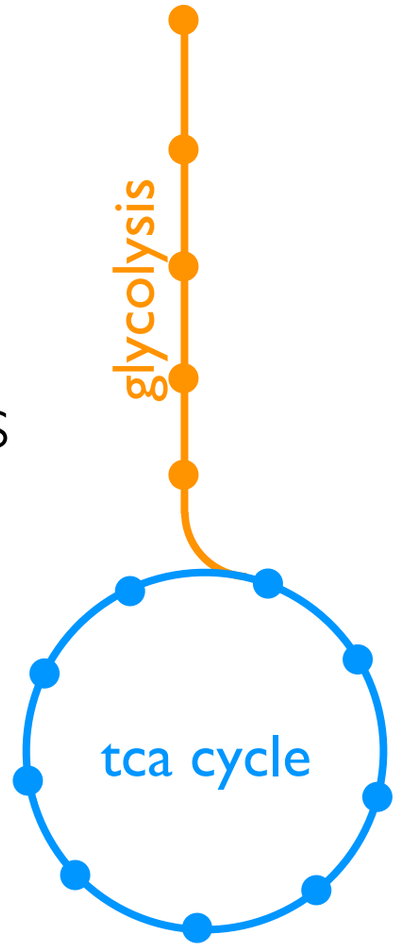
phylogeny

- evolutionary relationship
- **binary tree**



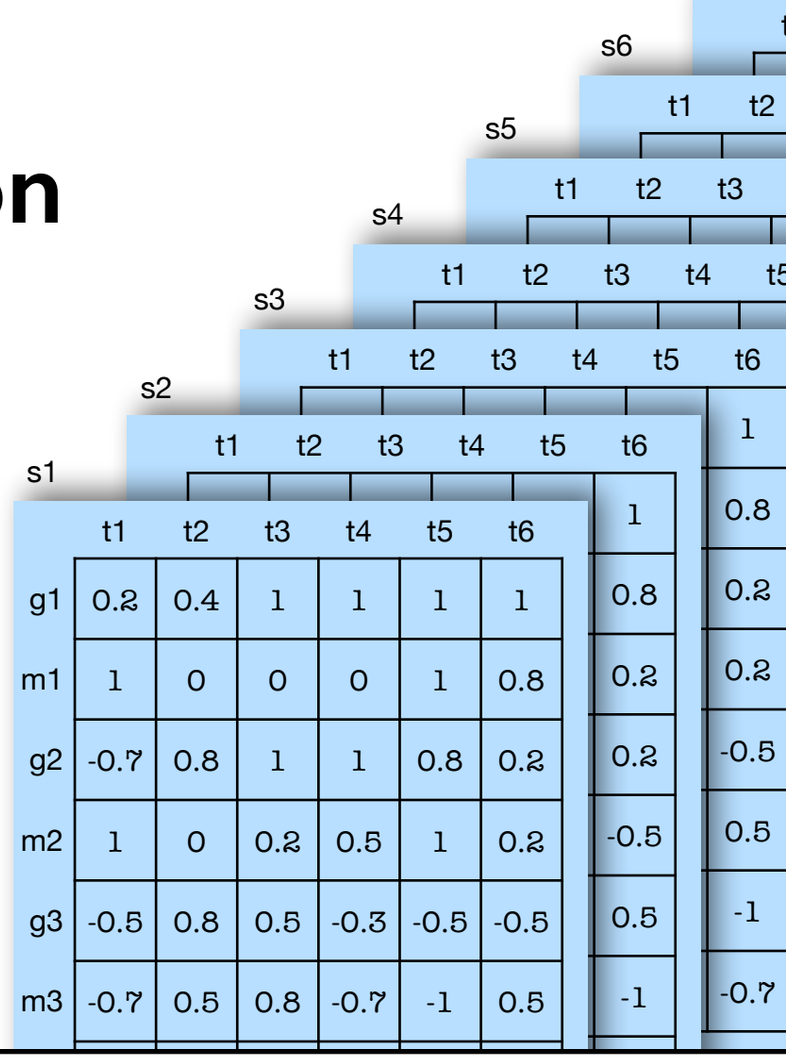
metabolic pathways

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



gene expression

- 6000 genes and 140 metabolites
- 6 time points
- 14 species of yeast
- **3D table**



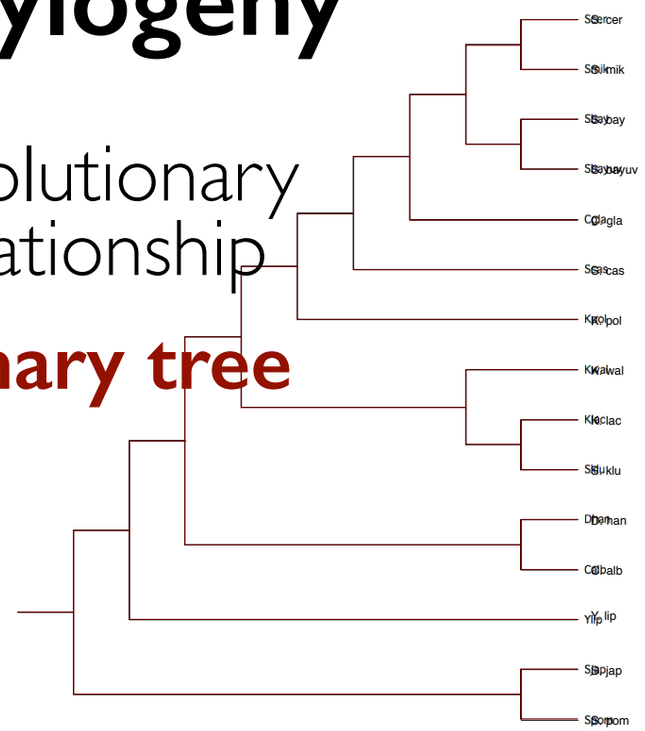
similarity scores

- aggregate time series for a gene/metabolite over species

aggregate( )

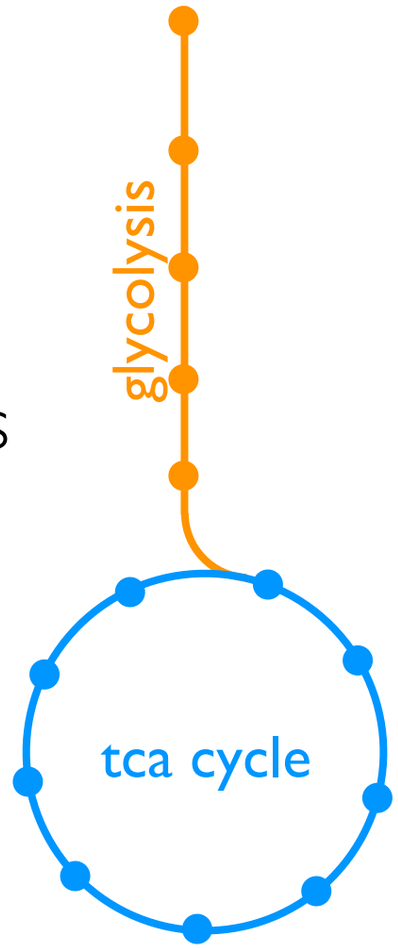
phylogeny

- evolutionary relationship
- **binary tree**



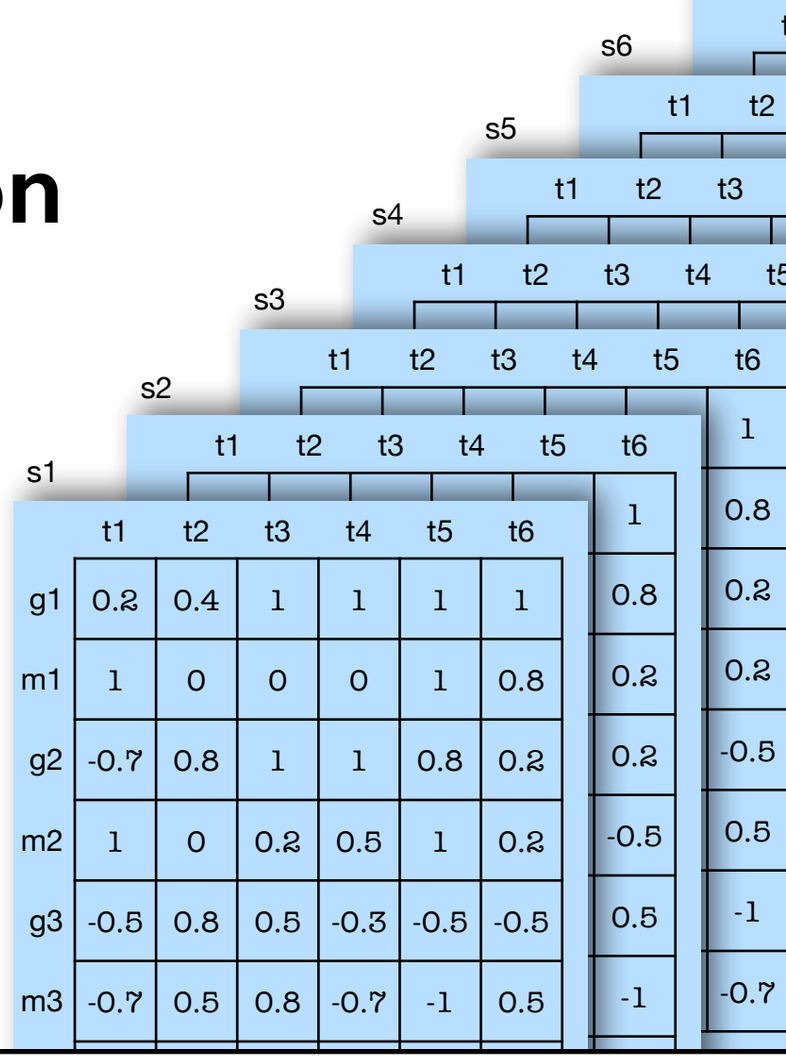
metabolic pathways

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



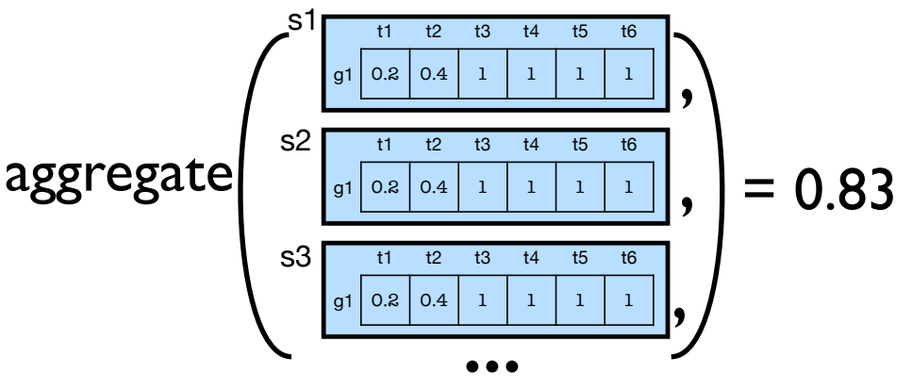
gene expression

- 6000 genes and 140 metabolites
- 6 time points
- 14 species of yeast
- **3D table**



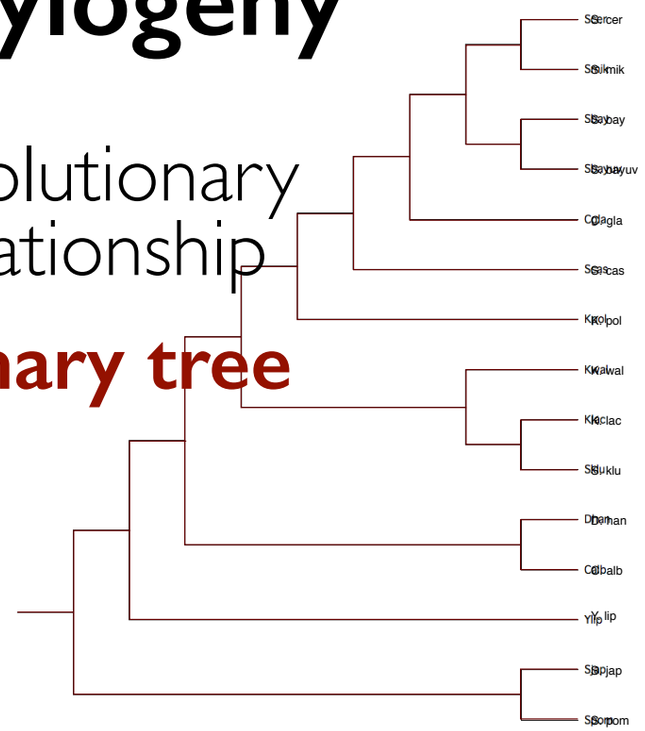
similarity scores

- aggregate time series for a gene/metabolite over species



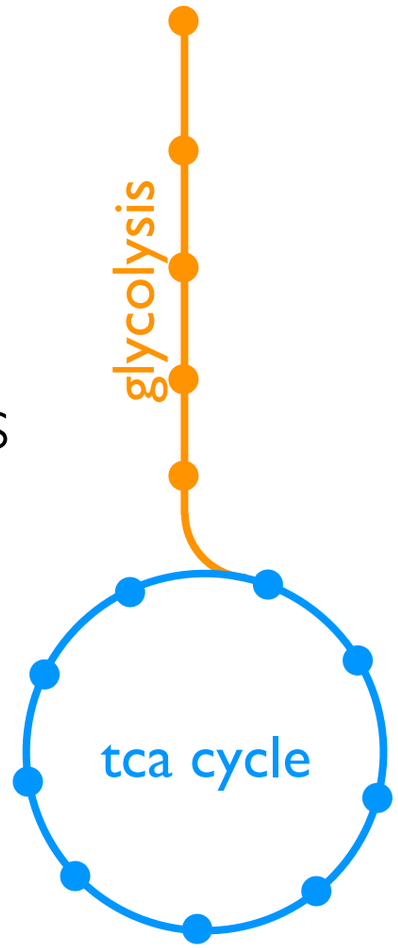
phylogeny

- evolutionary relationship
- **binary tree**



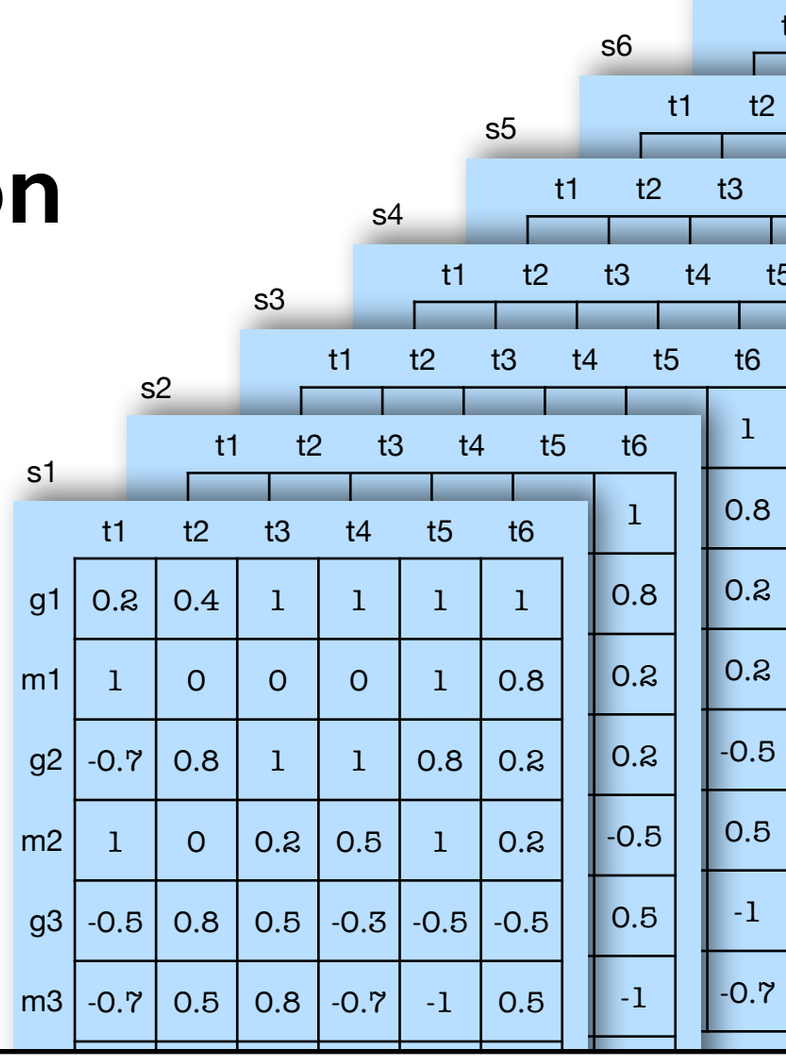
metabolic pathways

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

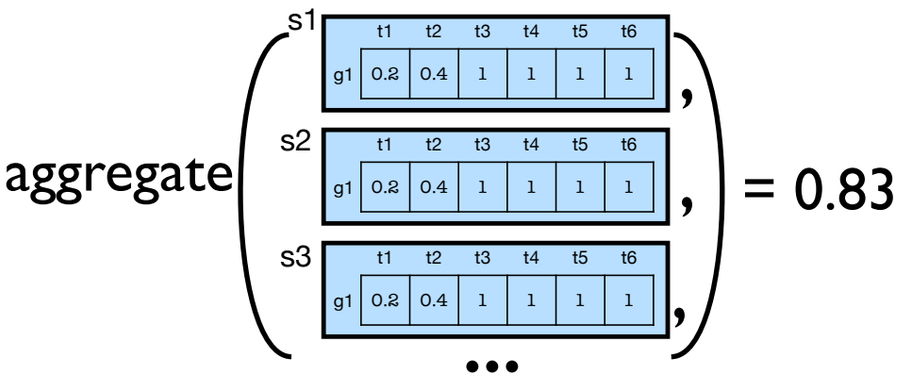


gene expression

- 6000 genes and 140 metabolites
- 6 time points
- 14 species of yeast
- **3D table**



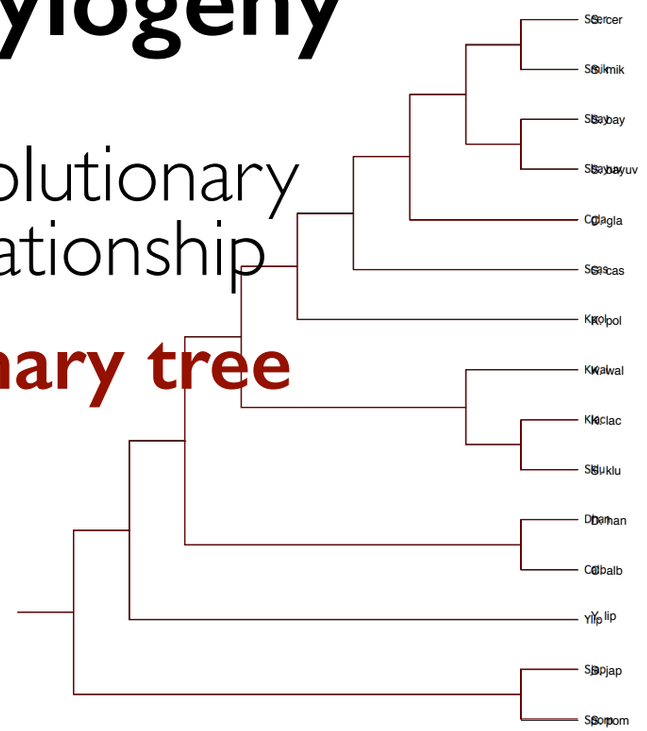
similarity scores



- aggregate time series for a gene/metabolite over species
- similarity of expression across species

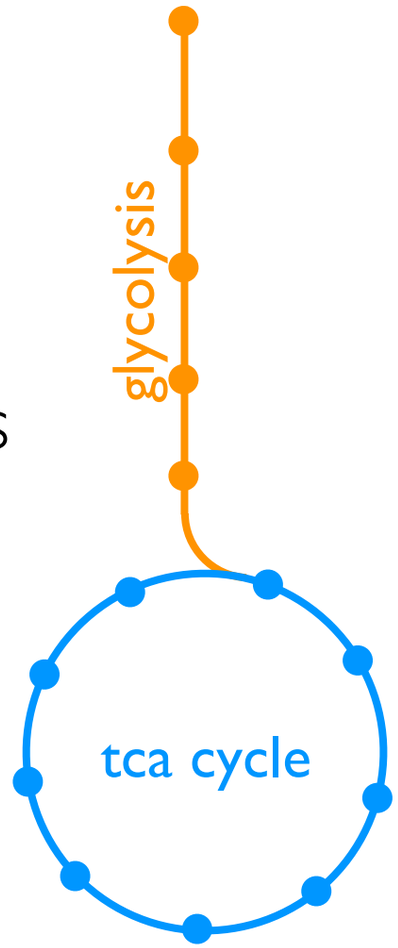
phylogeny

- evolutionary relationship
- **binary tree**



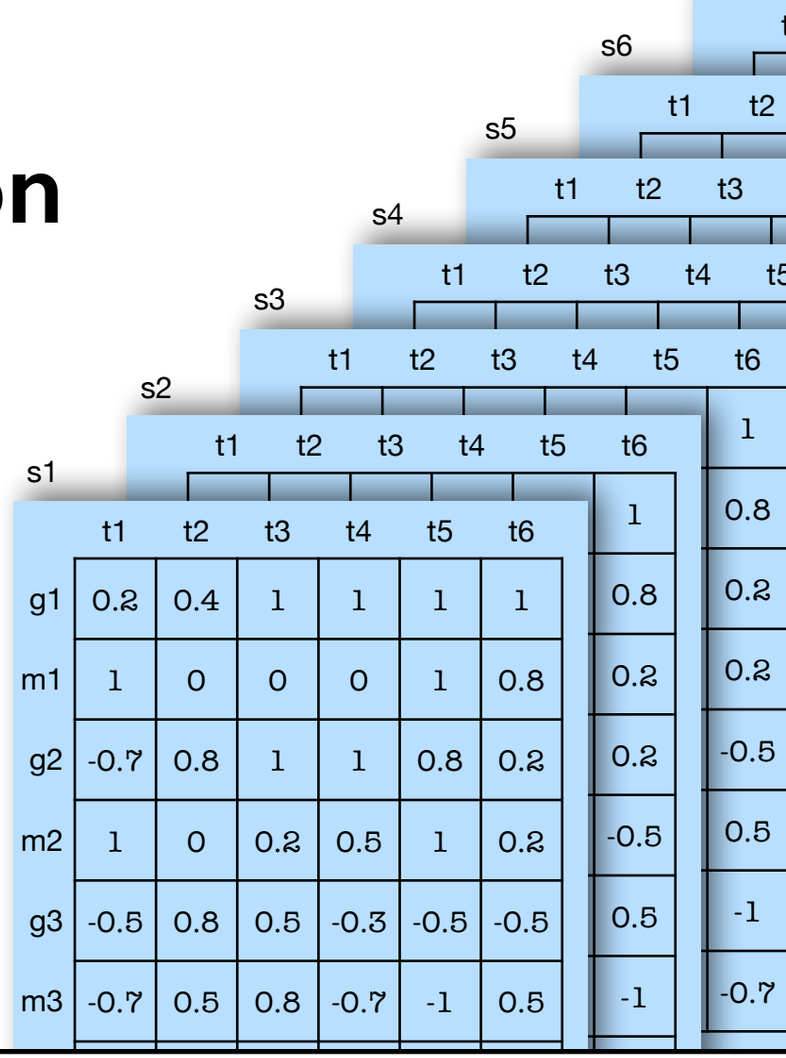
metabolic pathways

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

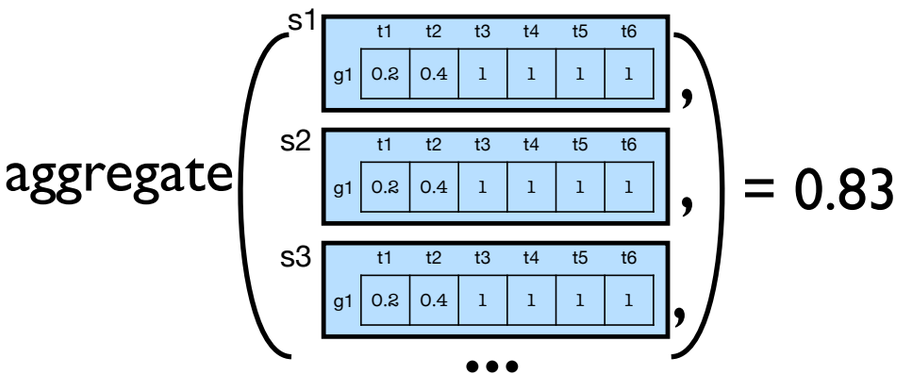


gene expression

- 6000 genes and 140 metabolites
- 6 time points
- 14 species of yeast
- **3D table**



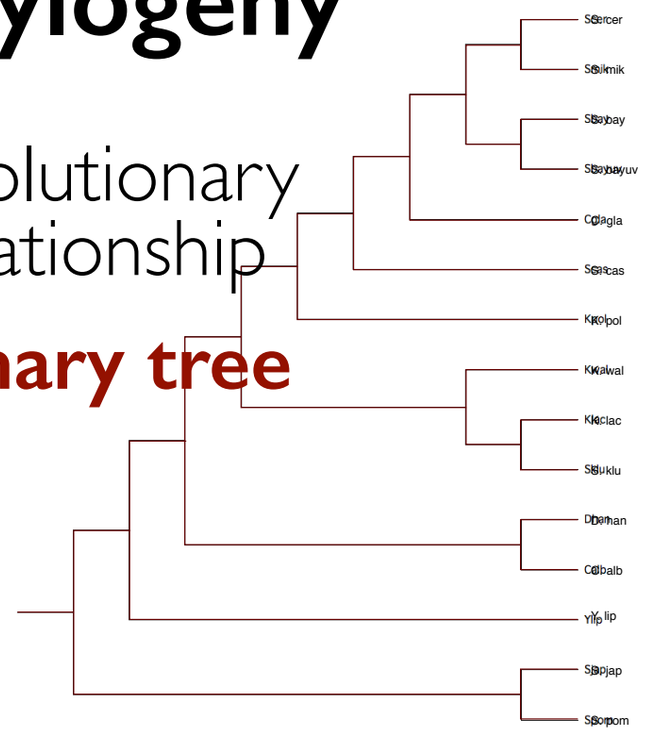
similarity scores



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

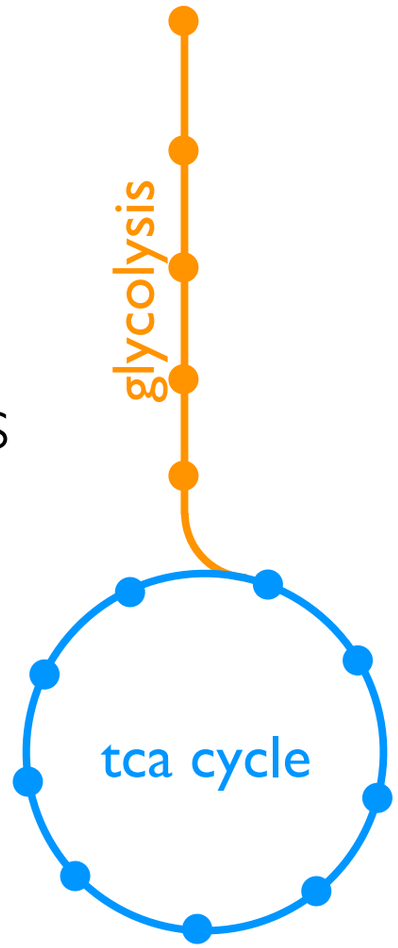
phylogeny

- evolutionary relationship
- **binary tree**



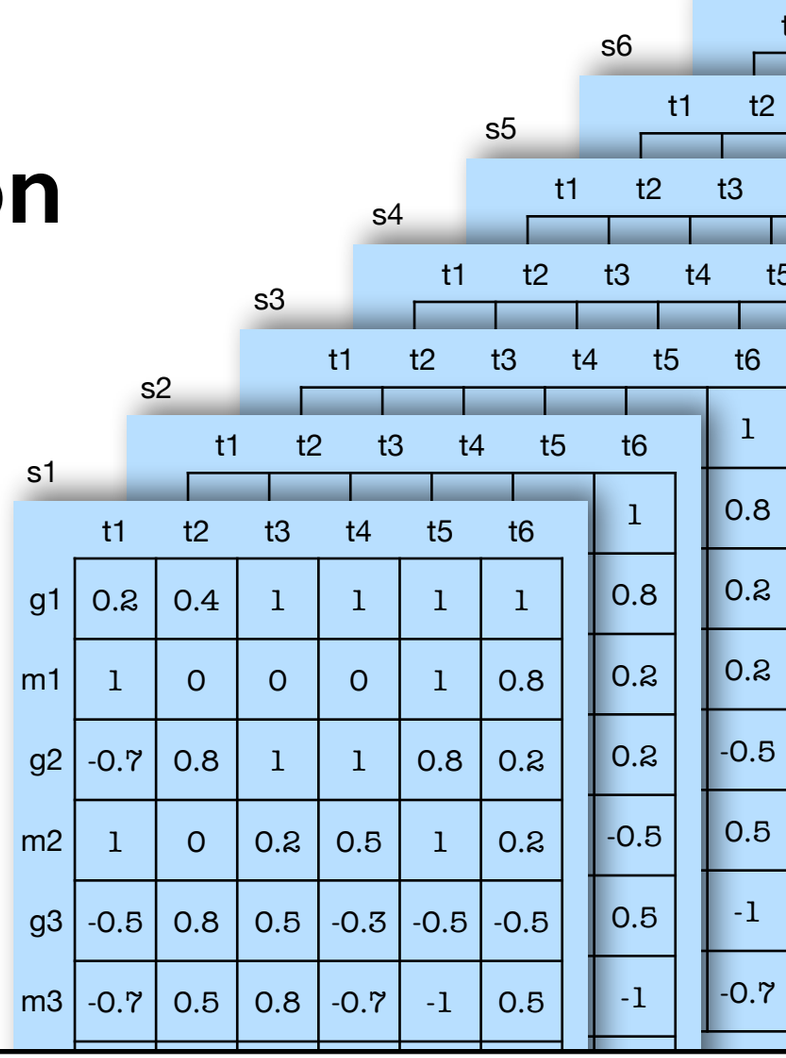
metabolic pathways

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

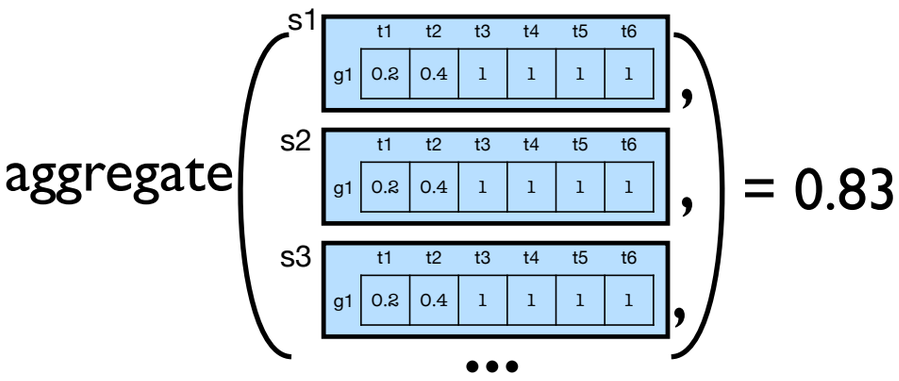


gene expression

- 6000 genes and 140 metabolites
- 6 time points
- 14 species of yeast
- **3D table**



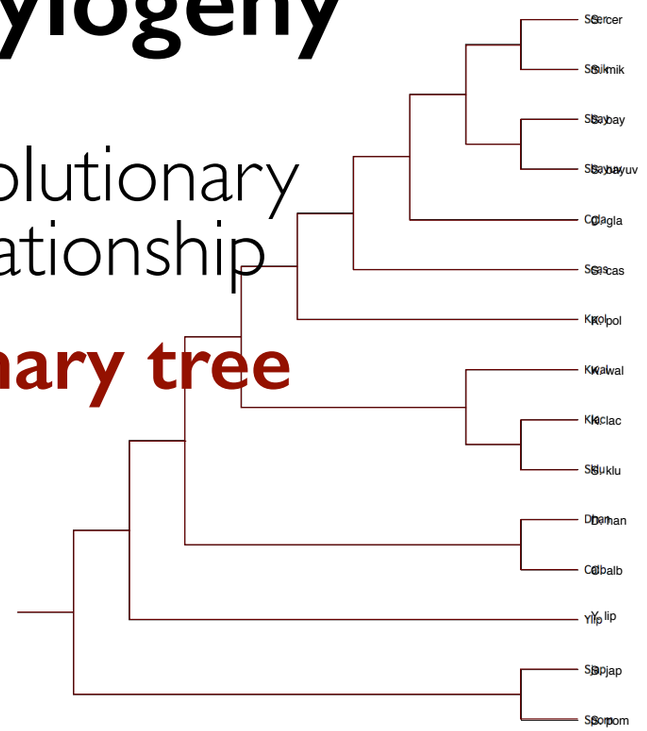
similarity scores



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

phylogeny

- evolutionary relationship
- **binary tree**



tasks

- study expression data as a time series

- compare a limited number of time series

- compare similarity scores along a pathway(s)

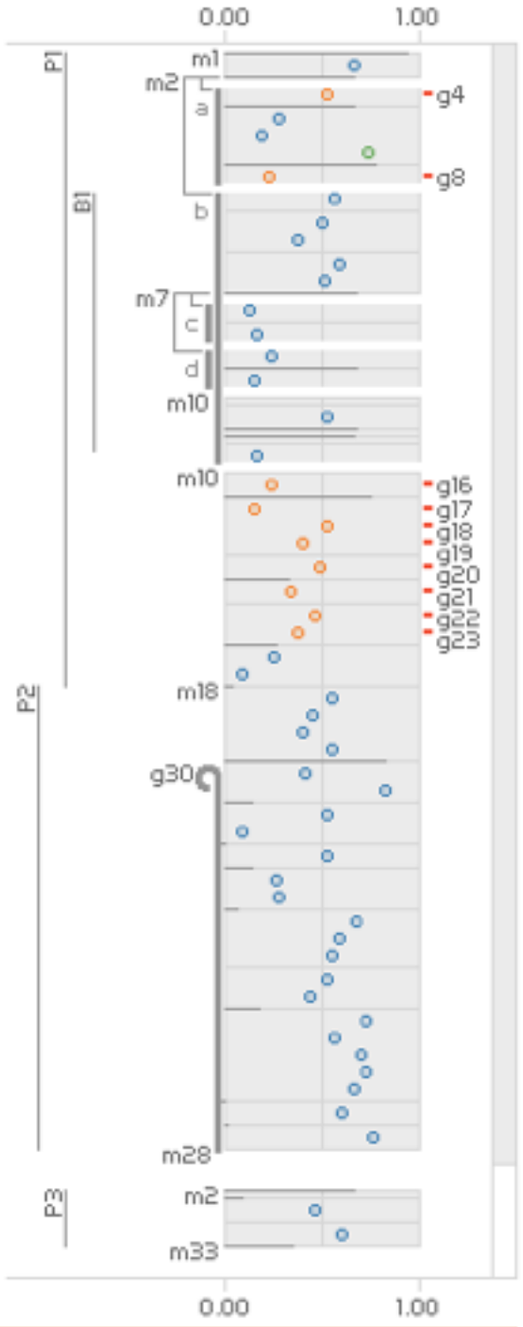
- comparison of multiple similarity scores

Pathline

PATHLINE

A TOOL FOR COMPARATIVE FUNCTIONAL GENOMICS

PATHWAY METRIC OVERVIEW



KEY **Genes**

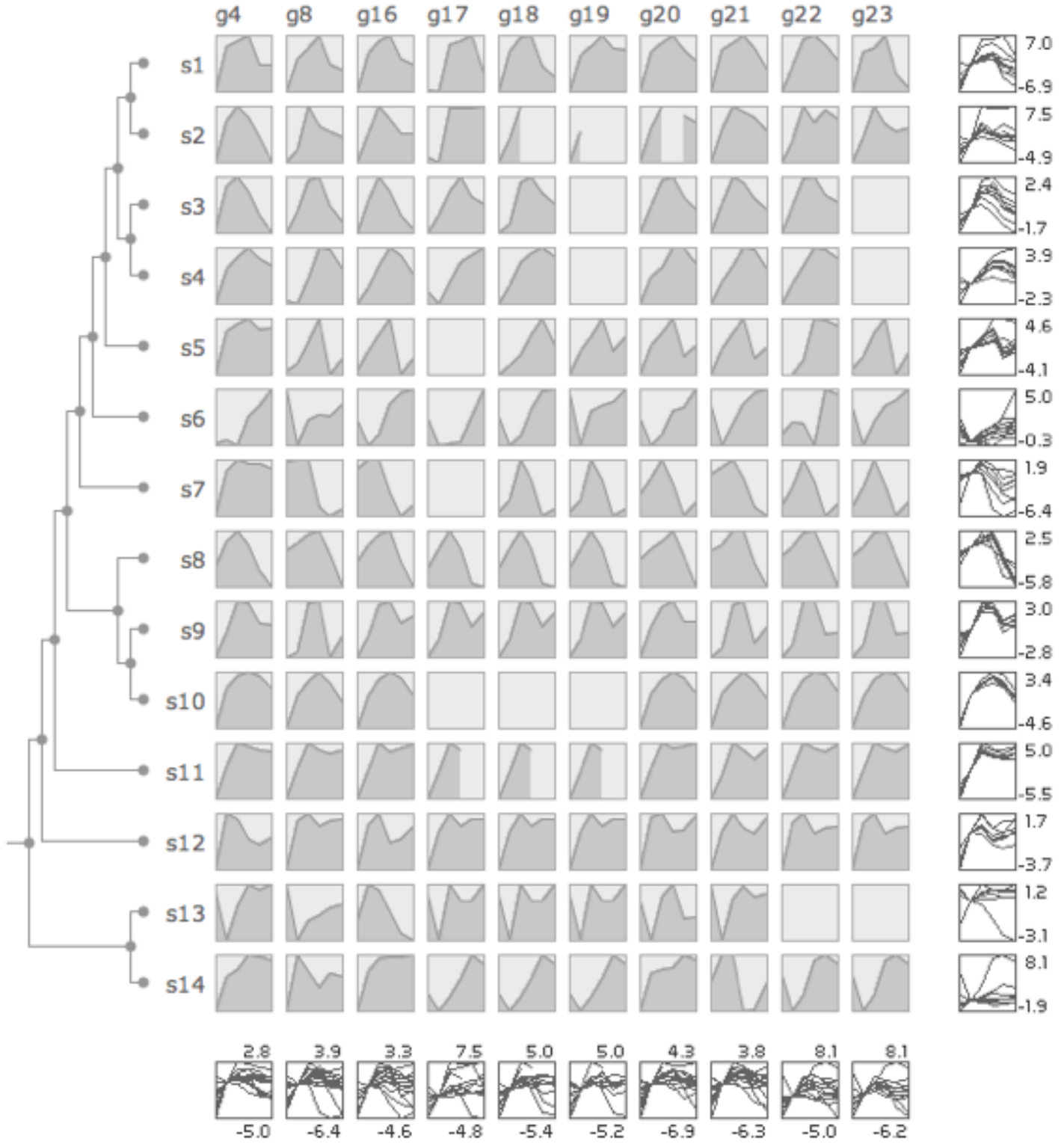
■ forward ■ reverse ■ bidirectional

Metabolites

Metrics

○ PearsonALL

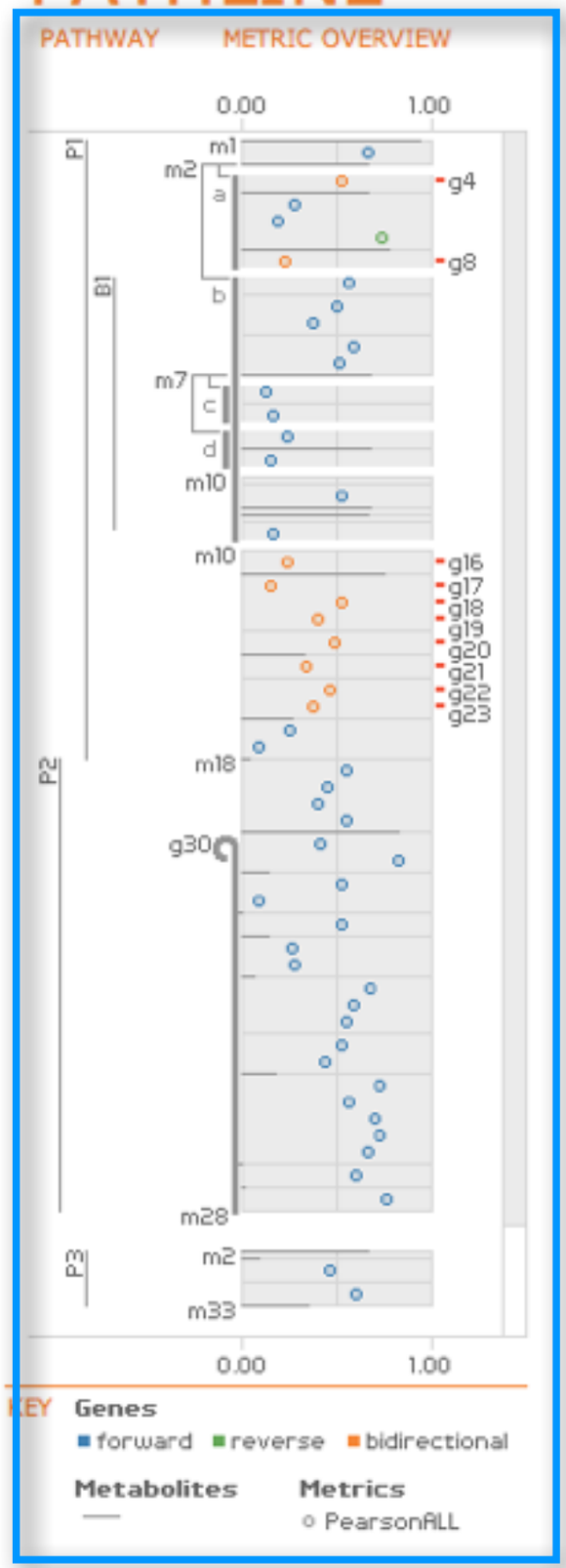
SPECIES CURVEMAP OVERLAYS

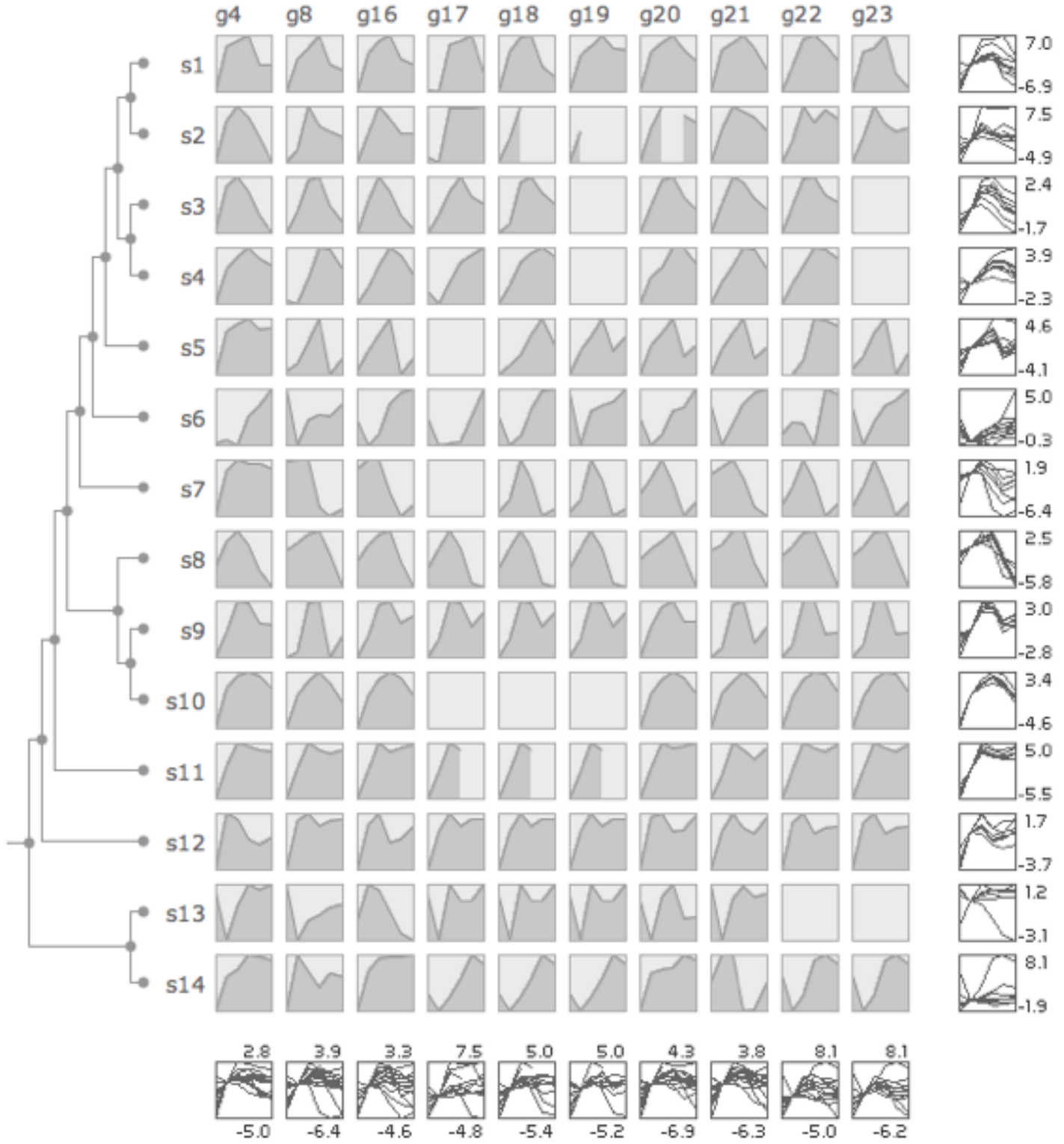


Pathline

PATHLINE

A TOOL FOR COMPARATIVE FUNCTIONAL GENOMICS



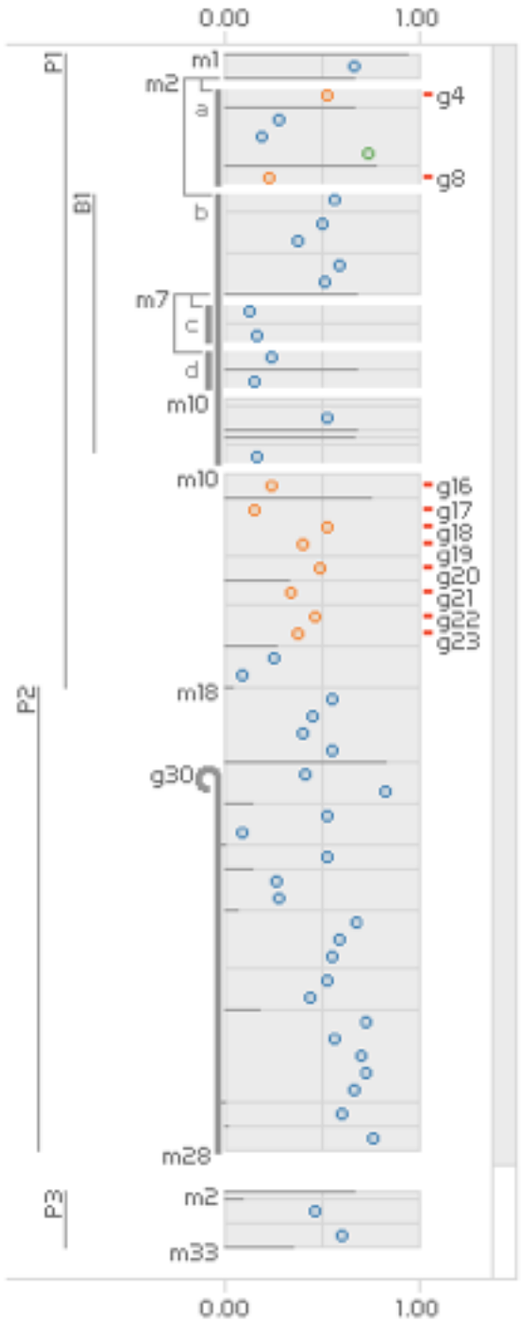


Pathline

PATHLINE

A TOOL FOR COMPARATIVE FUNCTIONAL GENOMICS

PATHWAY METRIC OVERVIEW



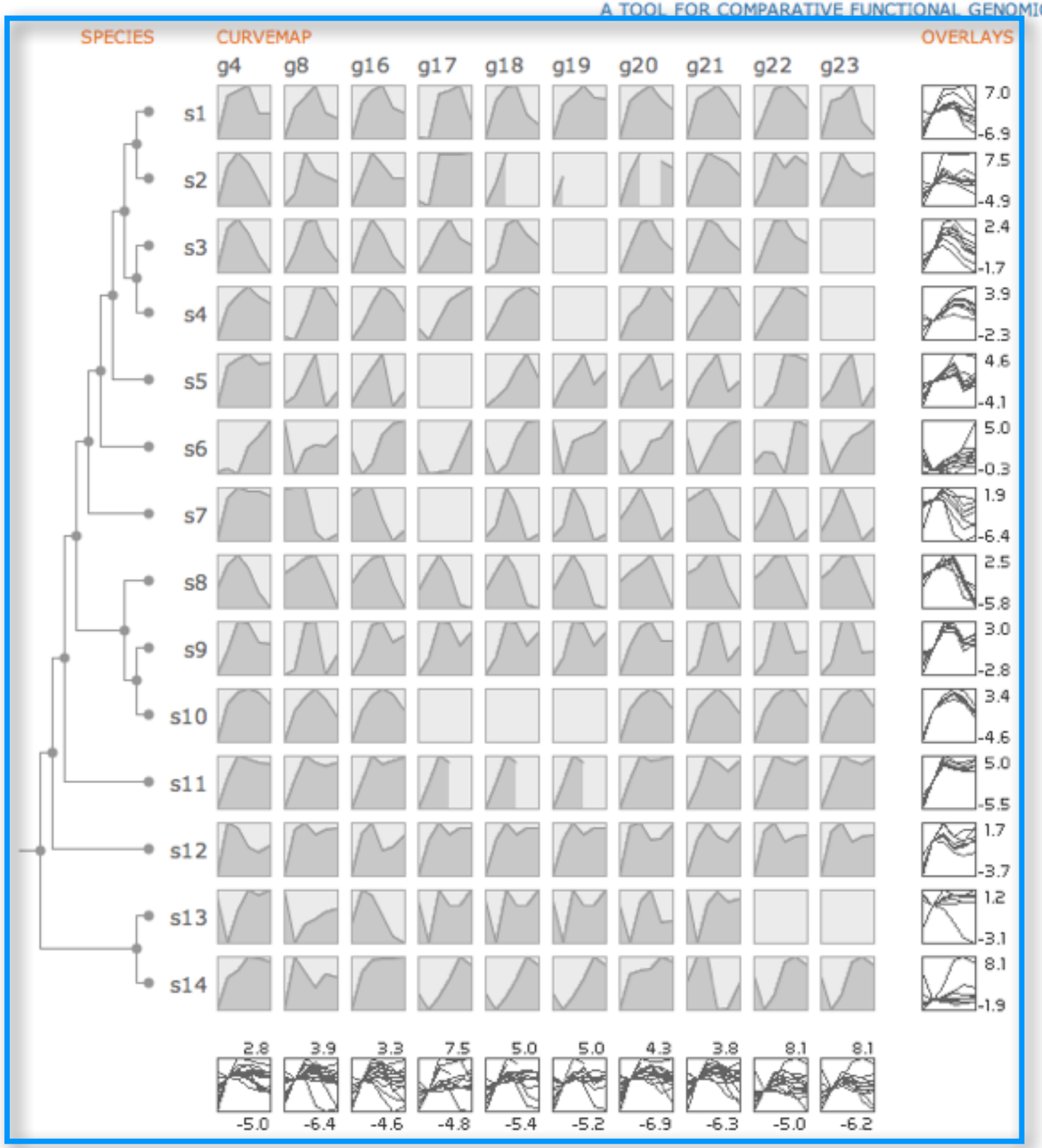
KEY **Genes**

■ forward ■ reverse ■ bidirectional

Metabolites

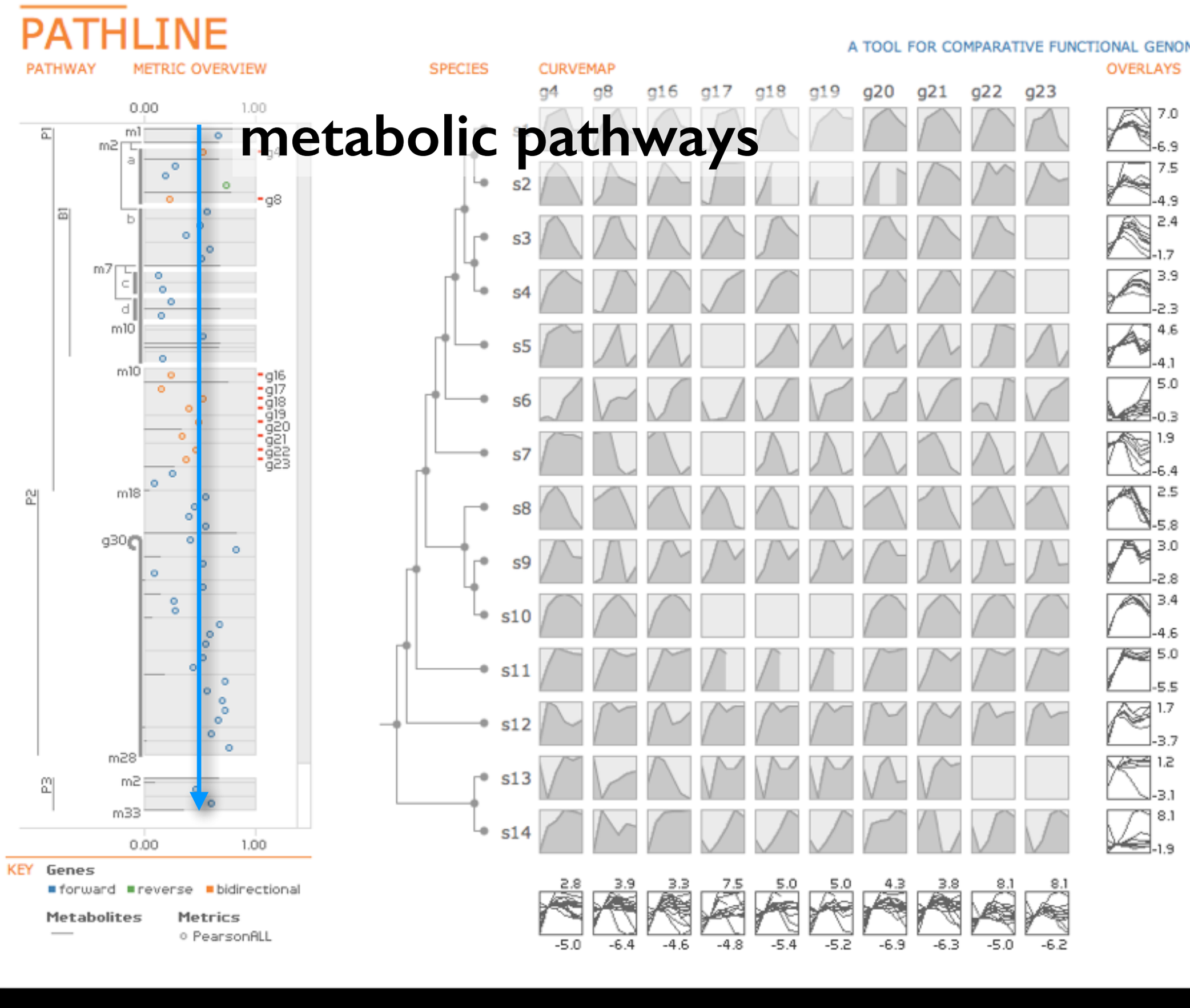
Metrics

○ PearsonALL



Pathline

metabolic pathways



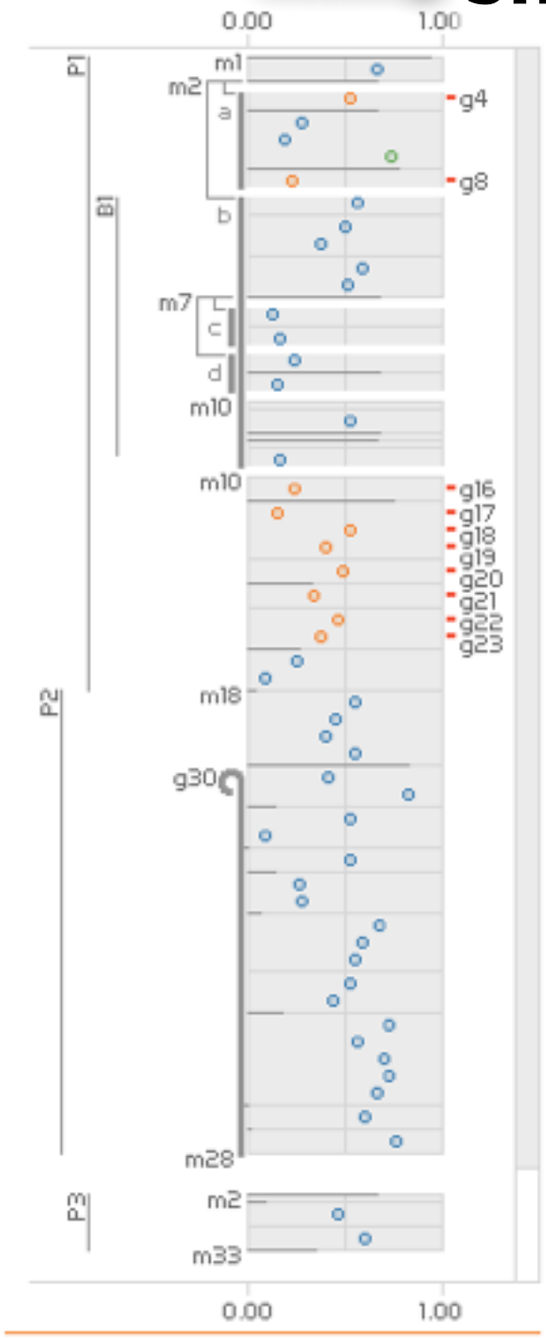
Pathline

PATHLINE

A TOOL FOR COMPARATIVE FUNCTIONAL GENOMICS

PATHWAY METRIC OVERVIEW SERIES CURVEMAP OVERLAYS

similarity scores



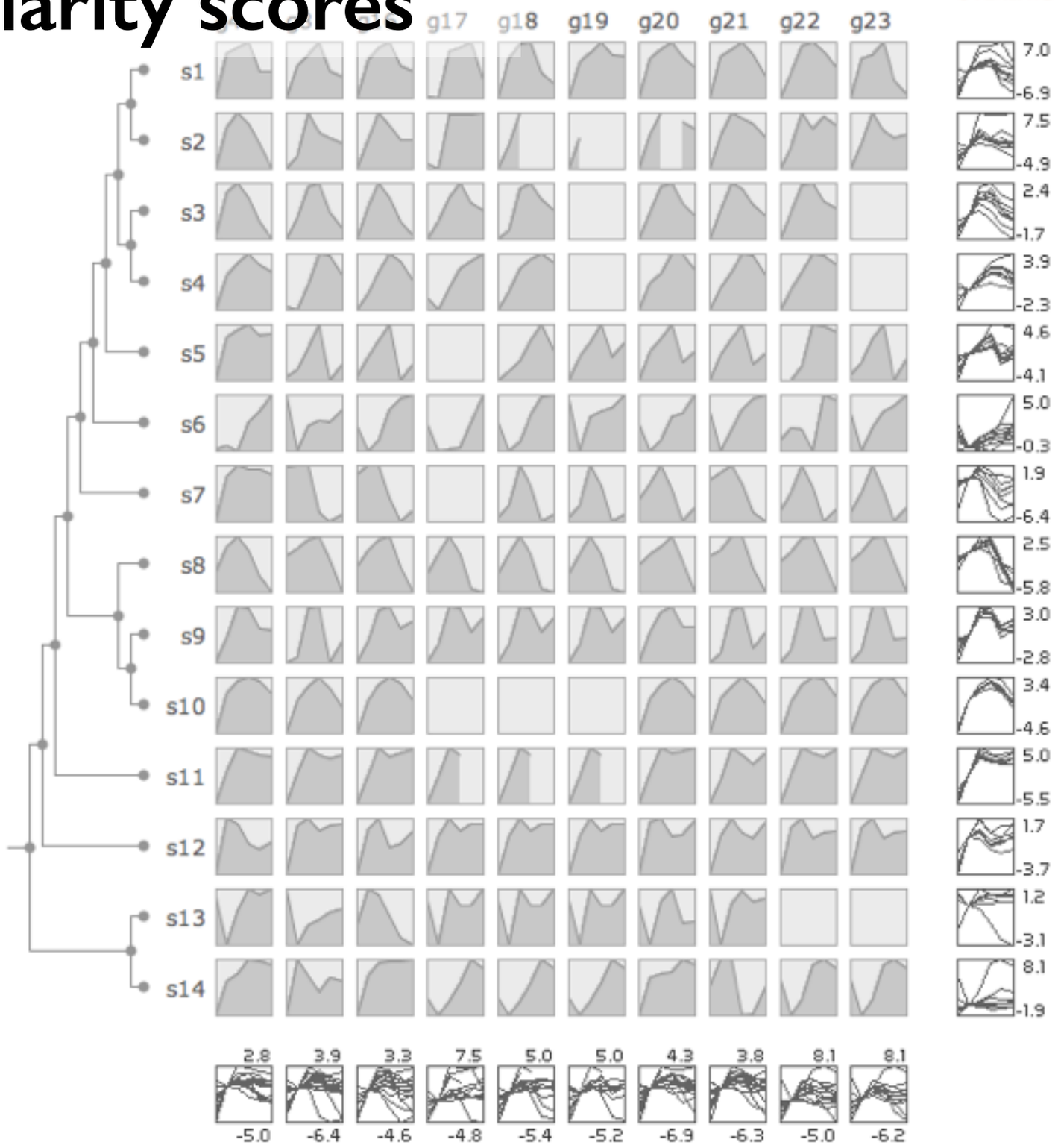
KEY **Genes**

■ forward ■ reverse ■ bidirectional

Metabolites

Metrics

○ PearsonALL



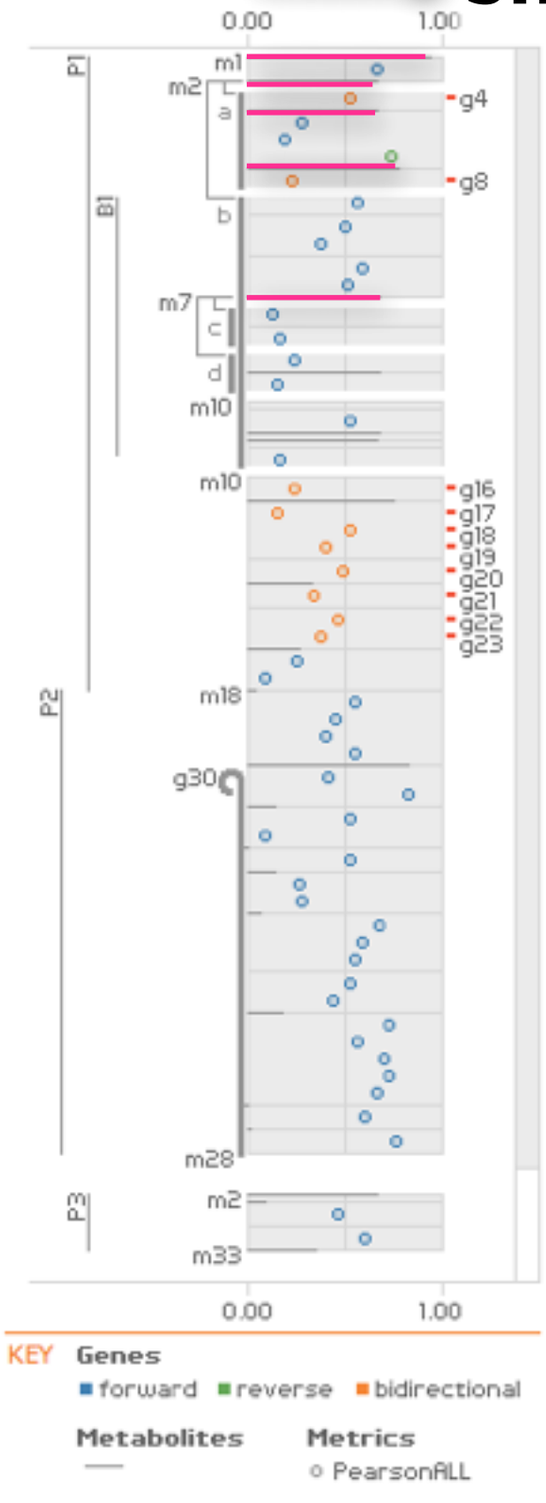
Pathline

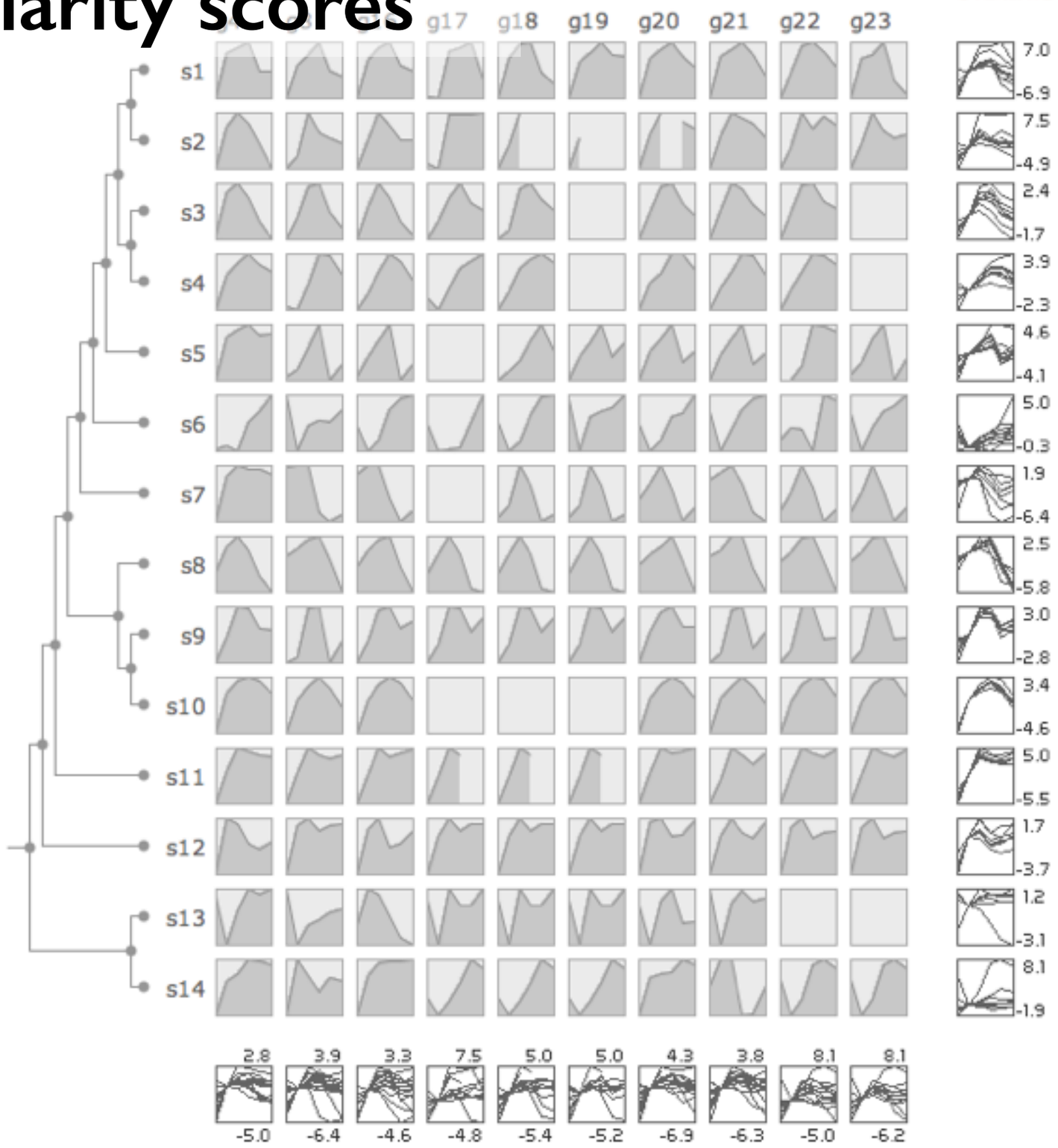
PATHLINE

A TOOL FOR COMPARATIVE FUNCTIONAL GENOMICS

PATHWAY METRIC OVERVIEW SERIES CURVEMAP OVERLAYS

similarity scores





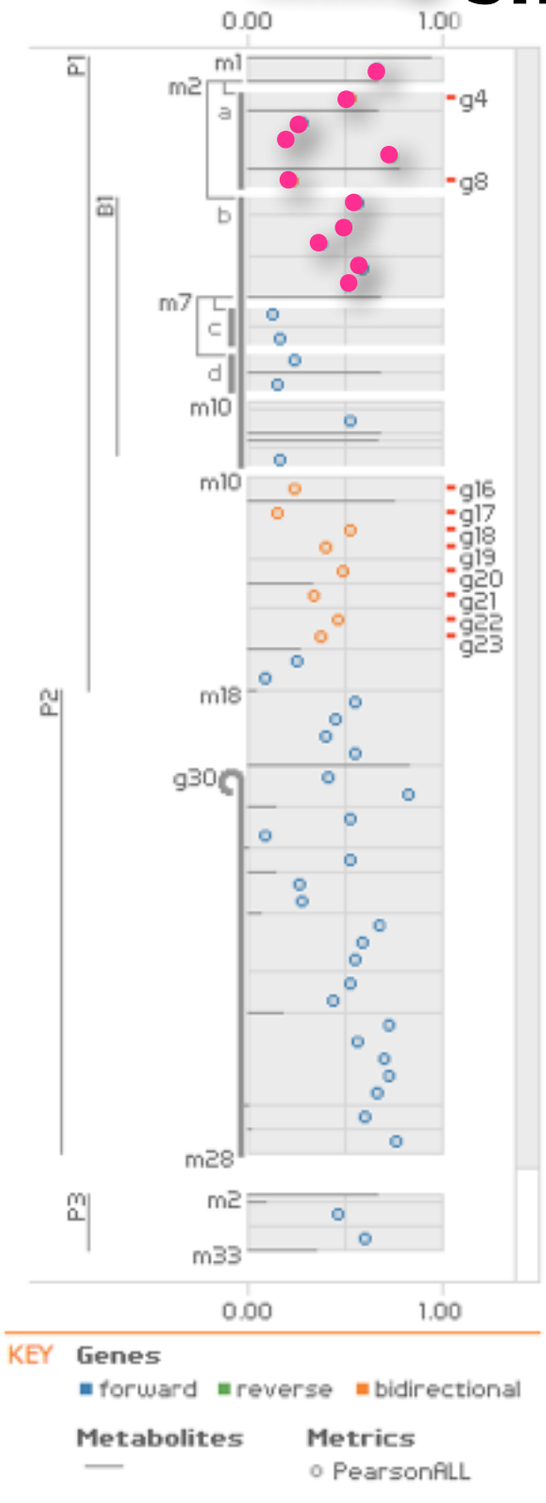
Pathline

PATHLINE

A TOOL FOR COMPARATIVE FUNCTIONAL GENOMICS

PATHWAY METRIC OVERVIEW SERIES CURVEMAP OVERLAYS

similarity scores



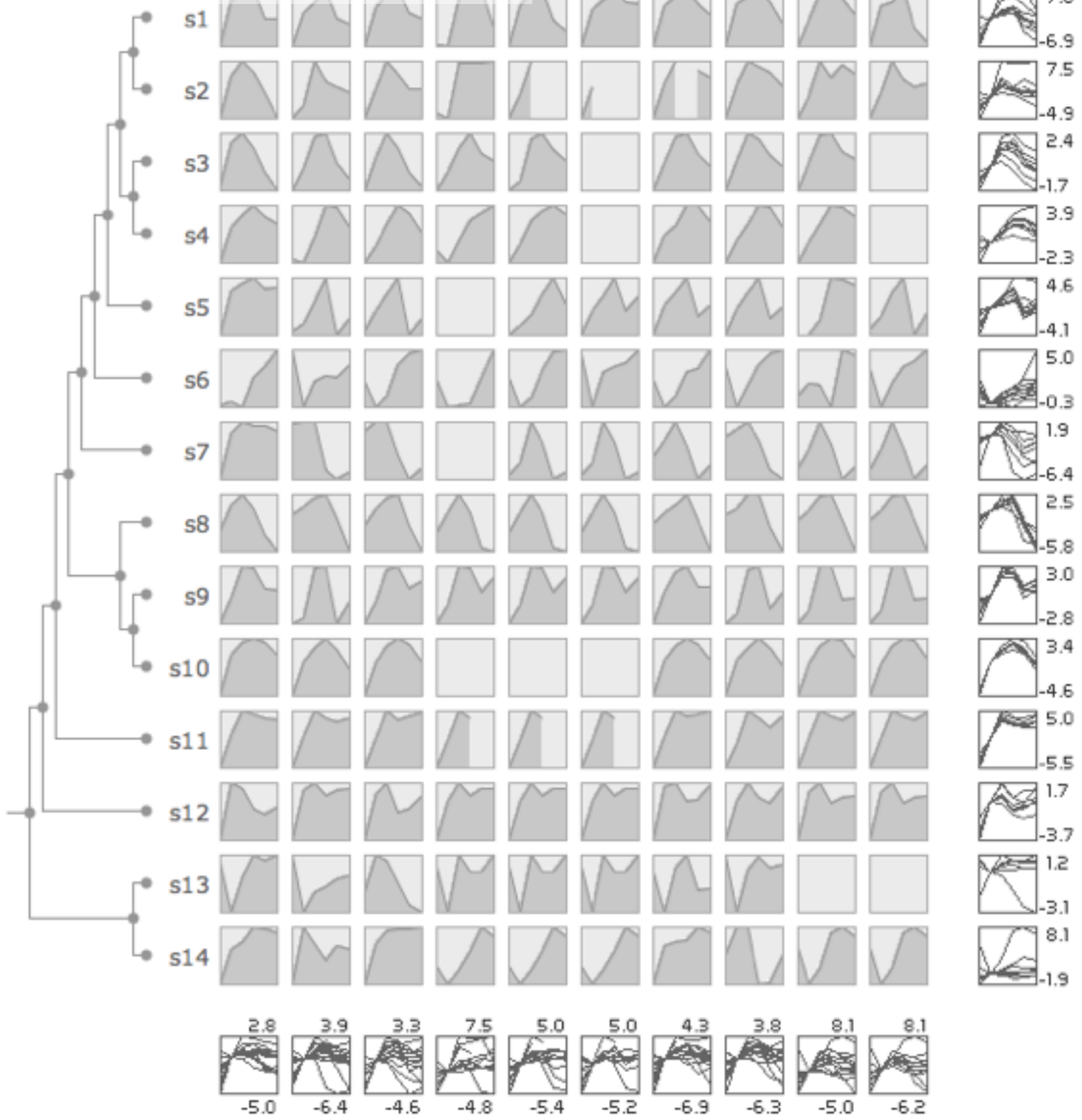
KEY Genes

■ forward ■ reverse ■ bidirectional

Metabolites

Metrics

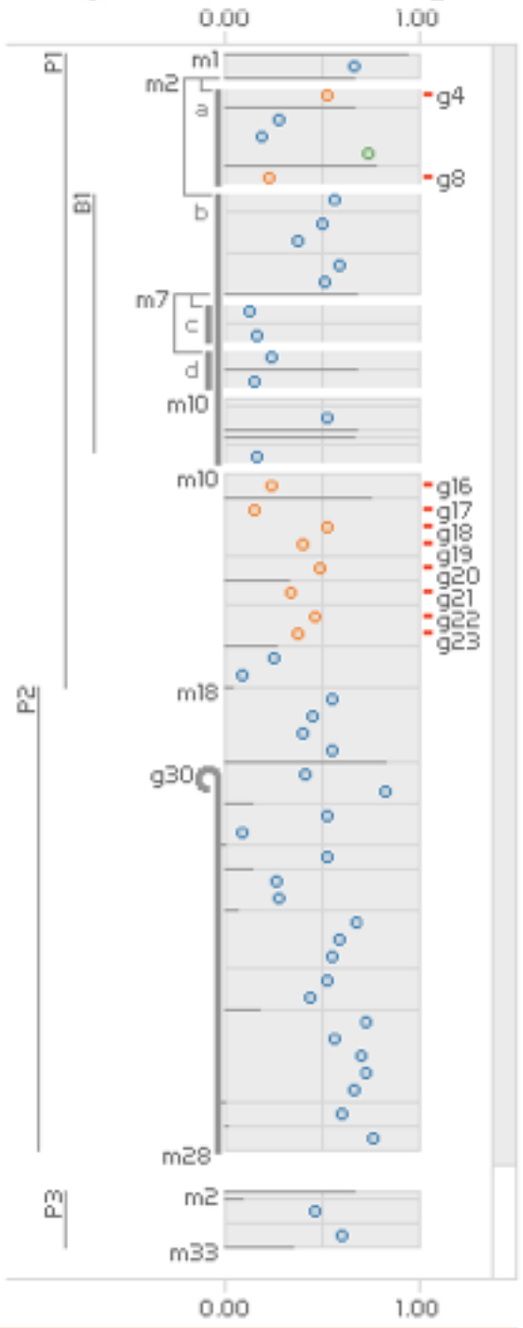
○ PearsonALL



Pathline

PATHLINE

gene expression



KEY **Genes**

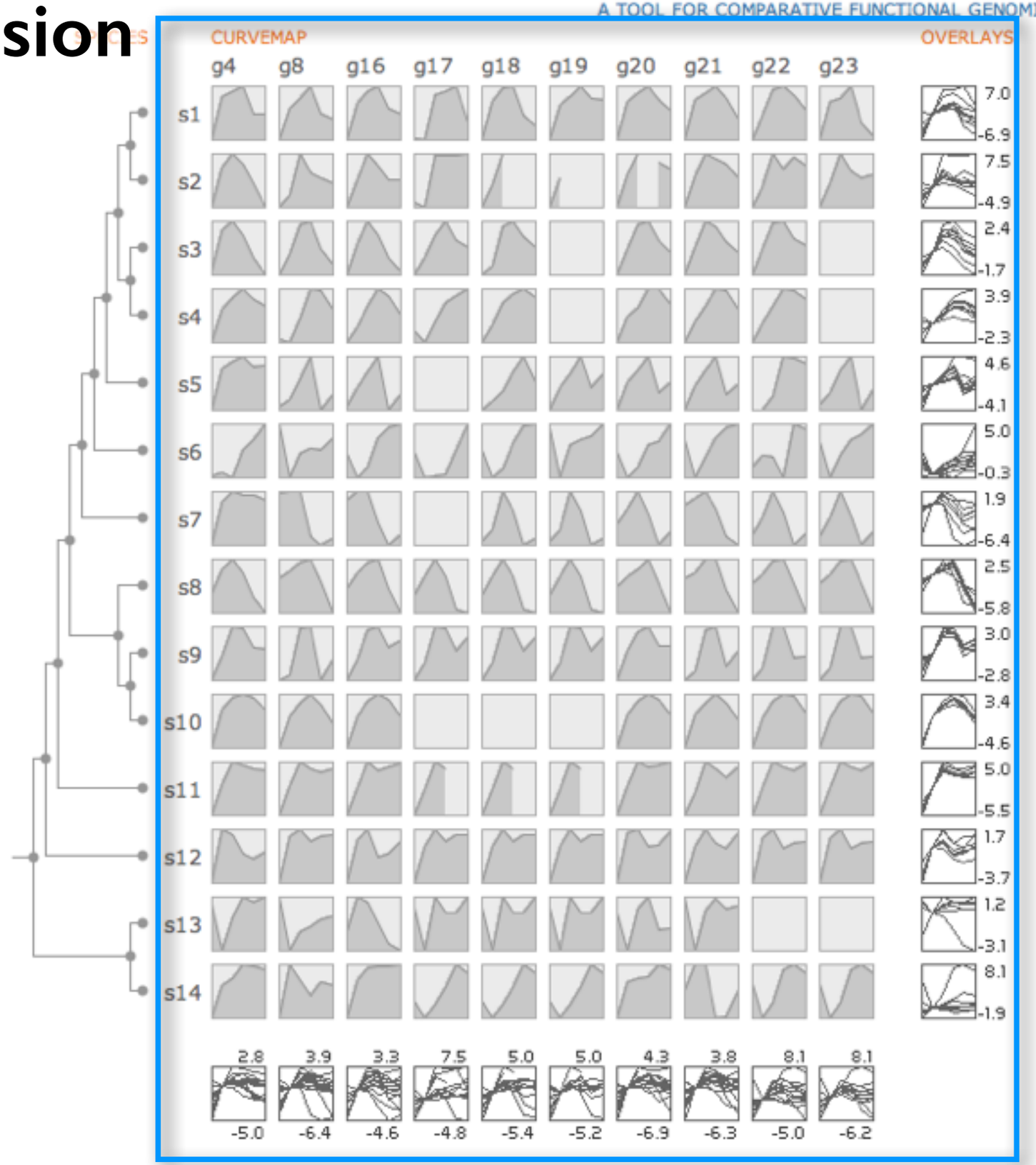
■ forward ■ reverse ■ bidirectional

Metabolites

Metrics

○ PearsonALL

A TOOL FOR COMPARATIVE FUNCTIONAL GENOMICS



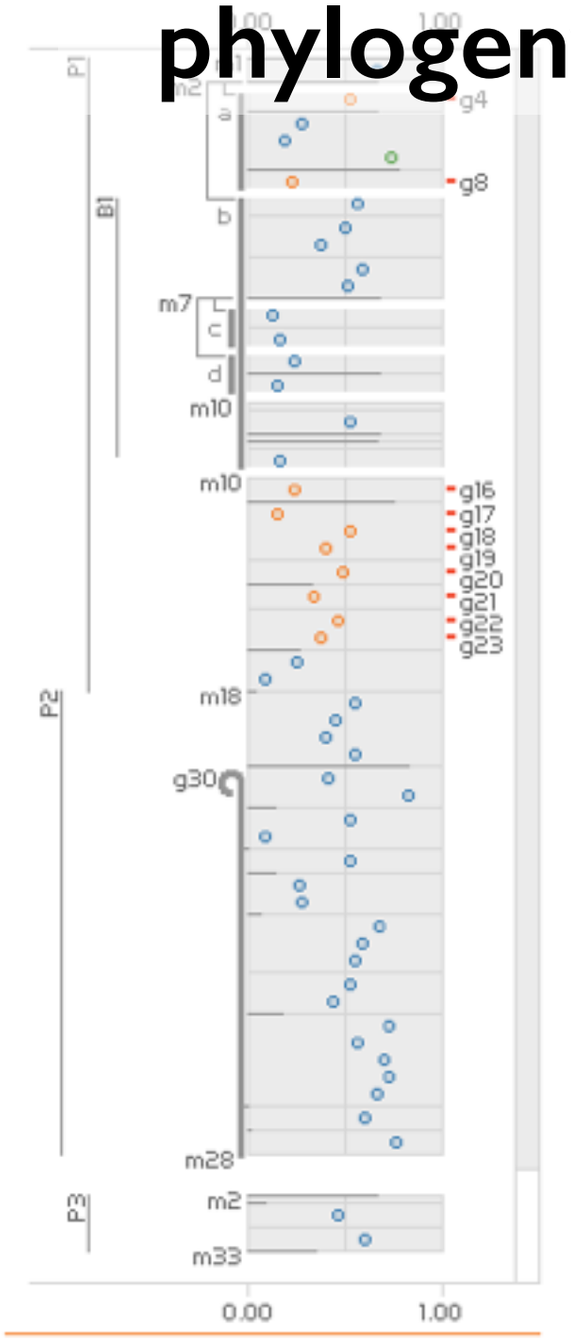
Pathline

PATHLINE

A TOOL FOR COMPARATIVE FUNCTIONAL GENOMICS

PATHWAY METRIC OVERVIEW SPECIES CURVEMAP OVERLAYS

phylogeny

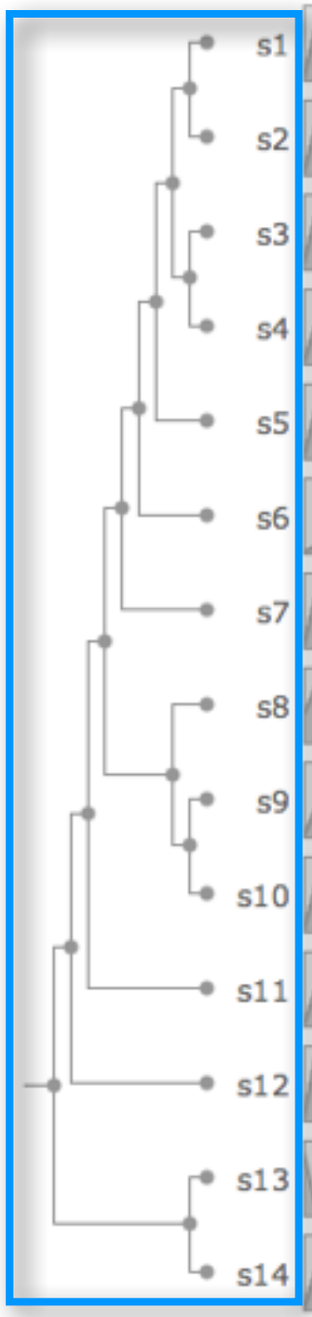


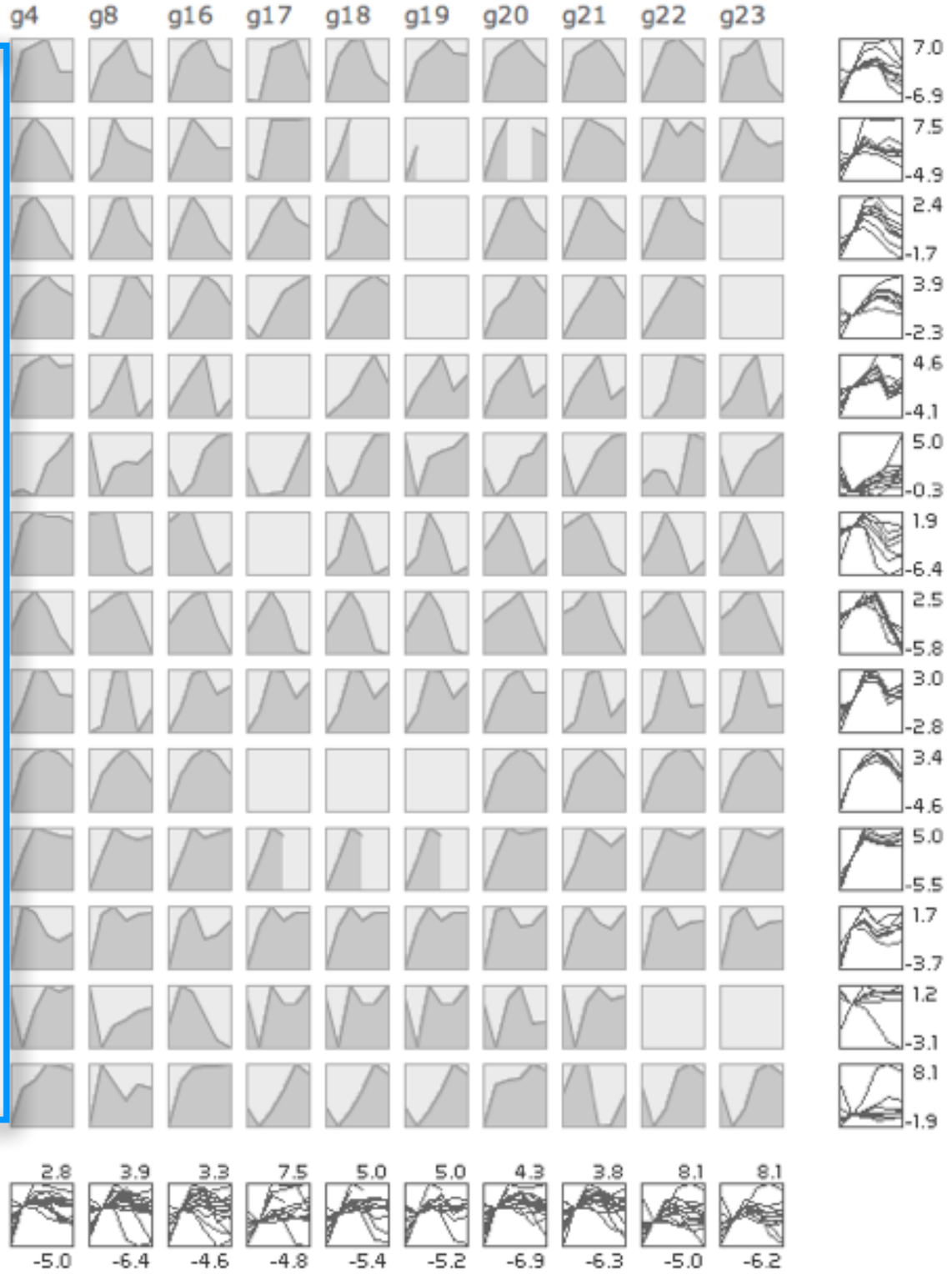
KEY **Genes**

forward reverse bidirectional

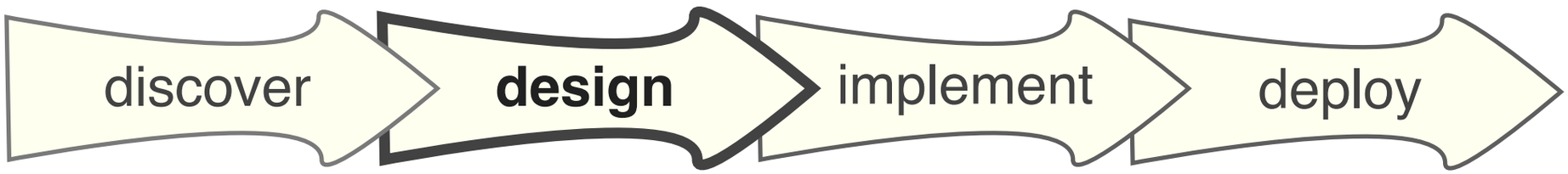
Metabolites **Metrics**

PearsonALL





broad consideration space ⟶ narrow proposal space



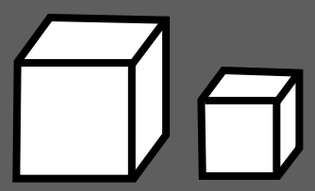
CORE

less accurate *more accurate*

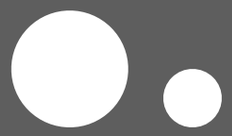
color

density

volume



area



angle

slope

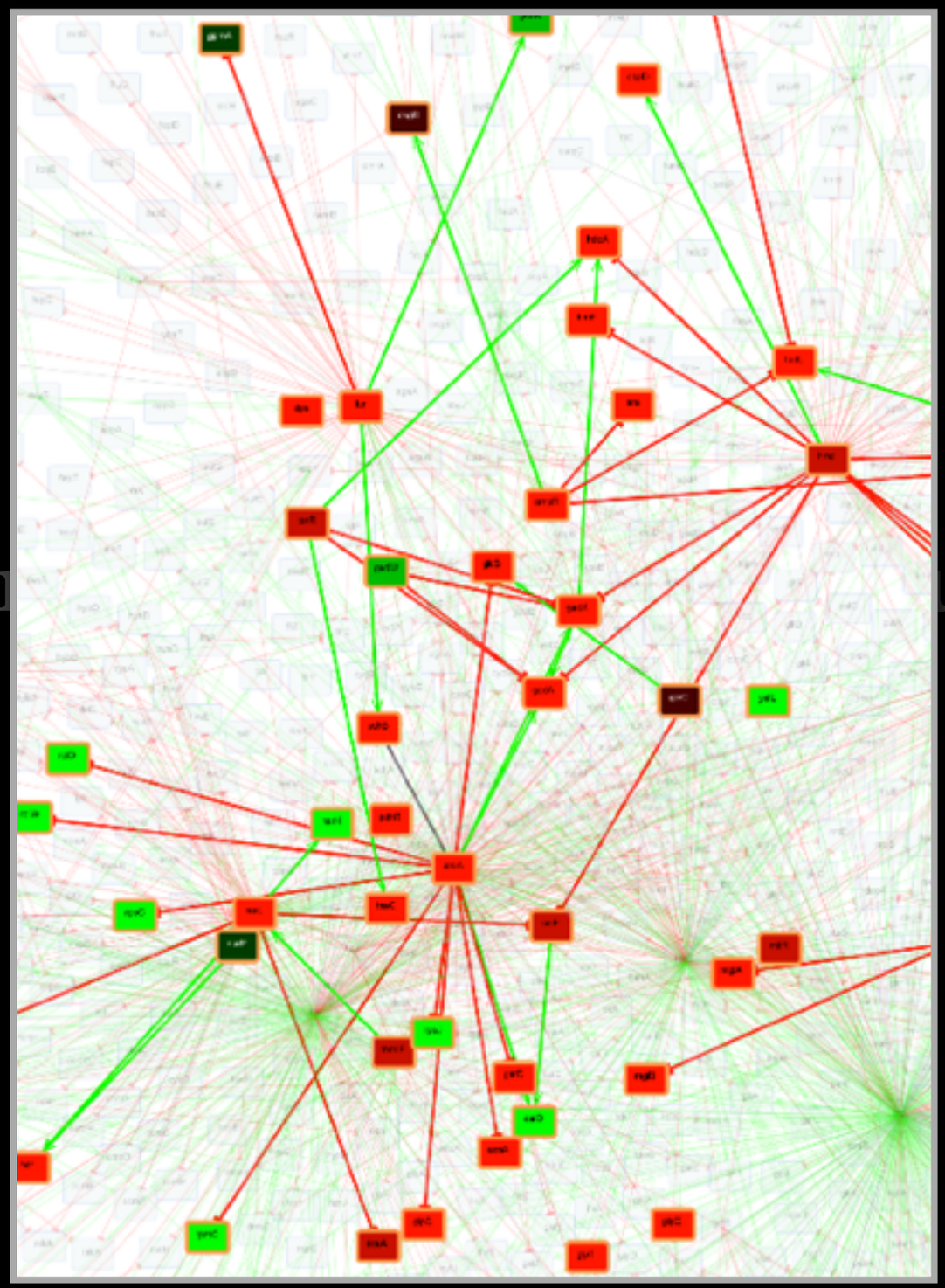
length

position

Graphical Perception: Theory, Experimentation, and Application to the Development of Graphical Methods.
W. Cleveland and R. McGill, Journal of the American Statistical Association, 1984.

encode quantitative values with spatial position

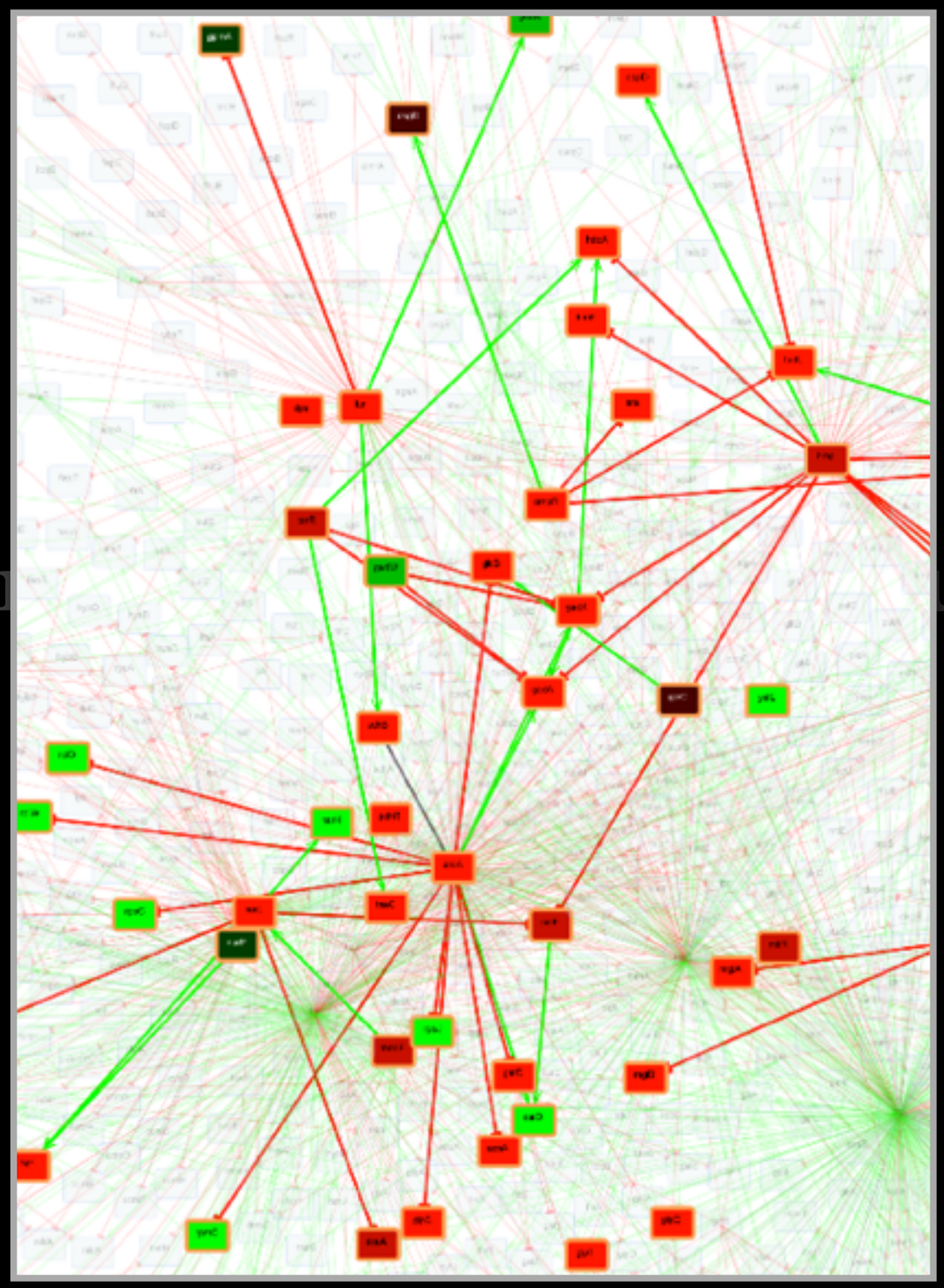
topological layout



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

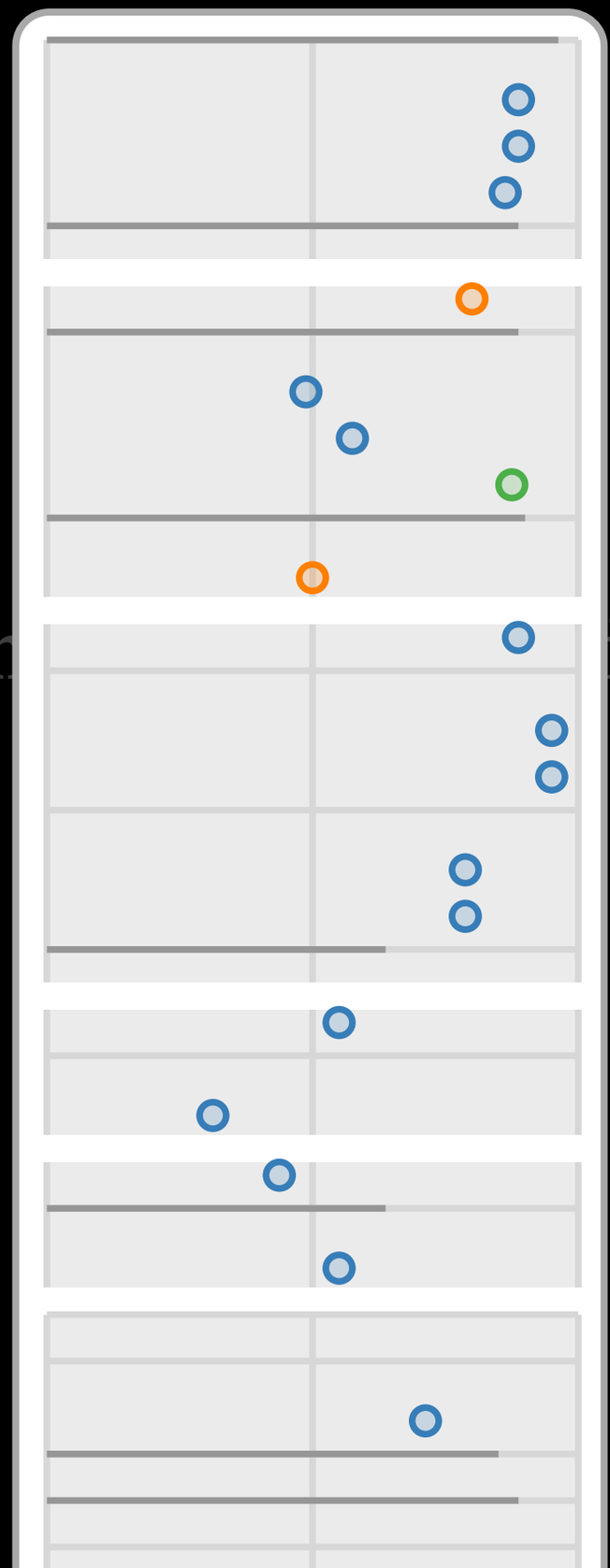
ues with spatial position

topological layout



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

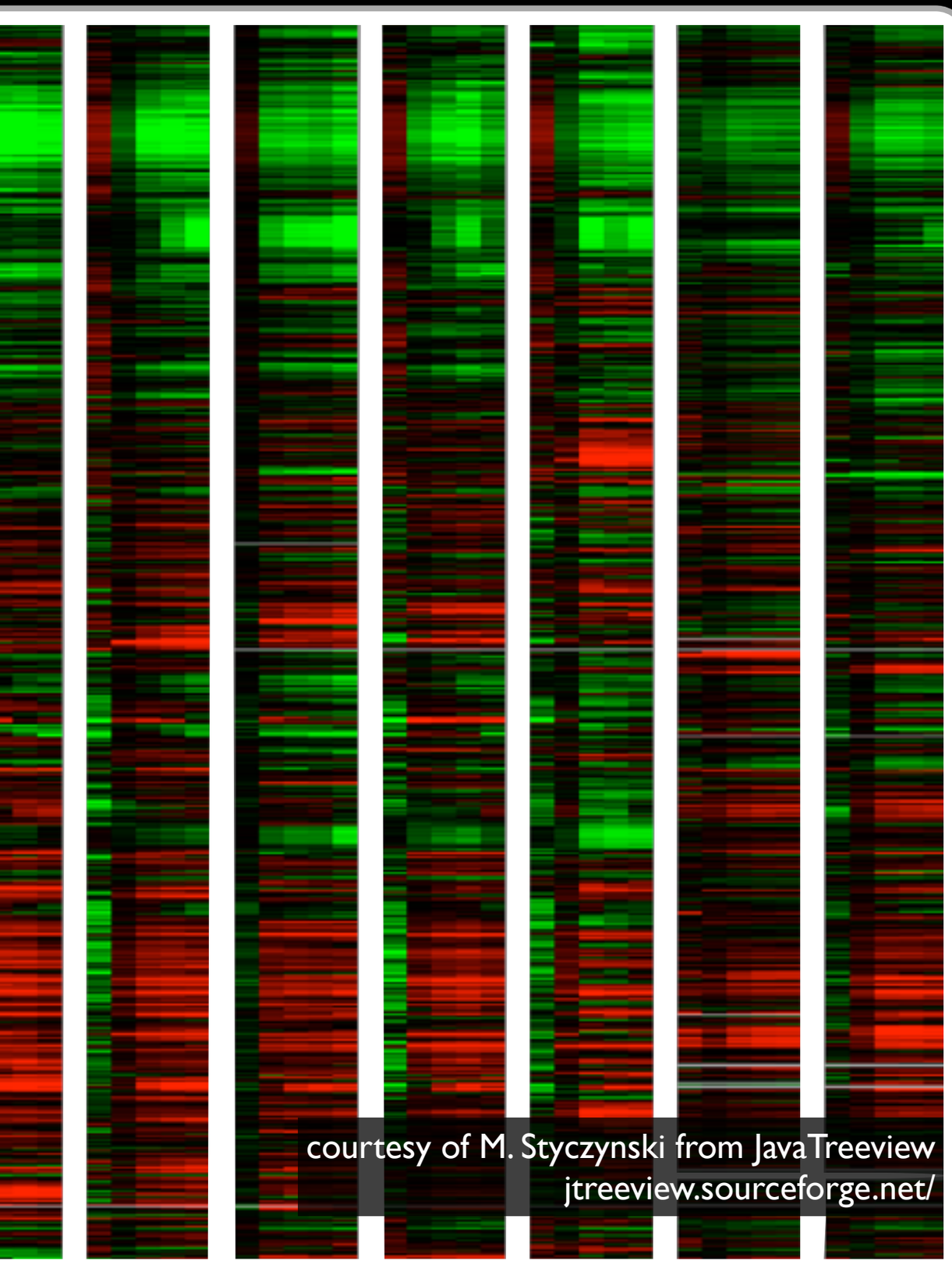
linearized pathway



encode quantitative values with spatial position

encode quantitative values with spatial position

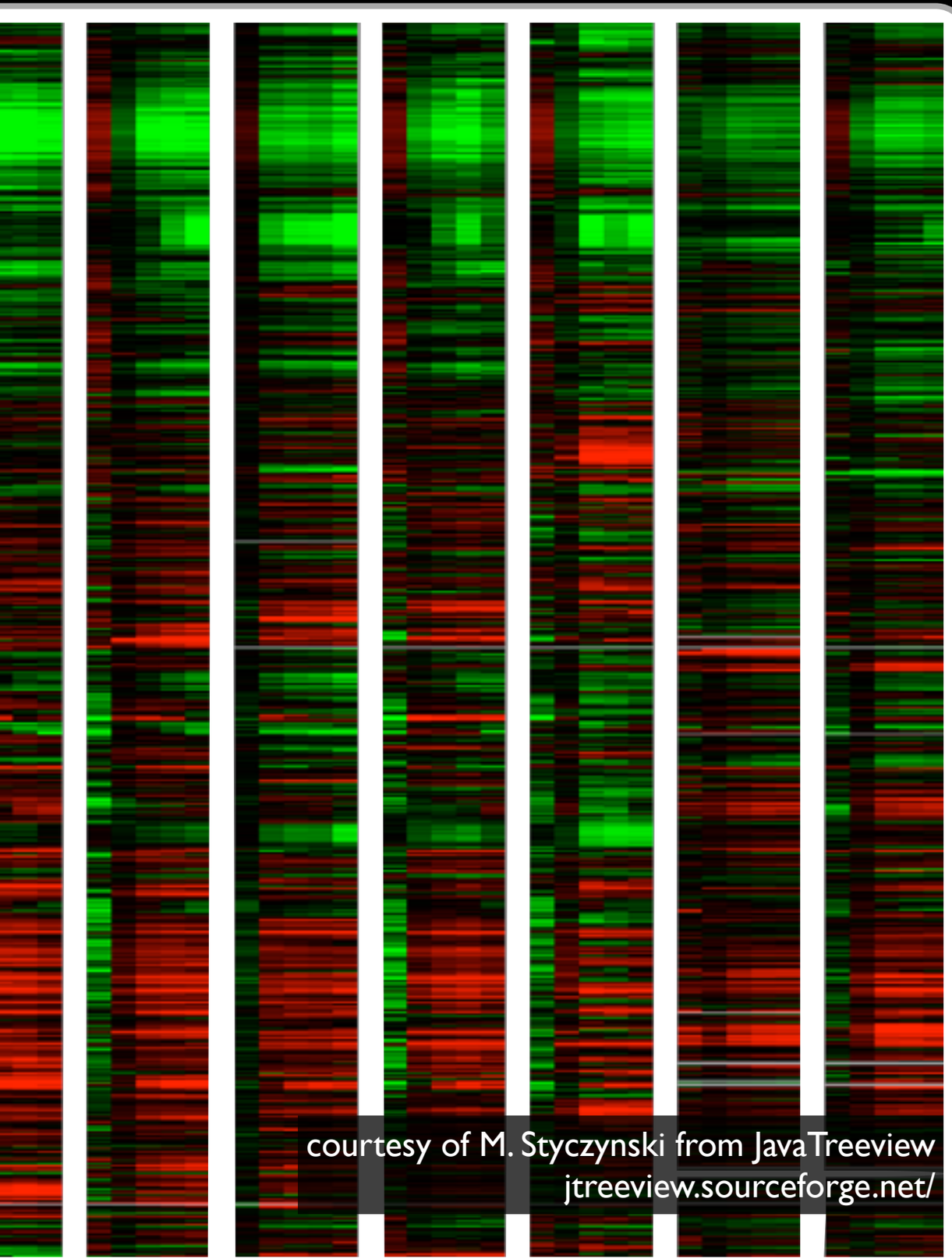
heatmap



courtesy of M. Styczynski from JavaTreeview
jtreeview.sourceforge.net/

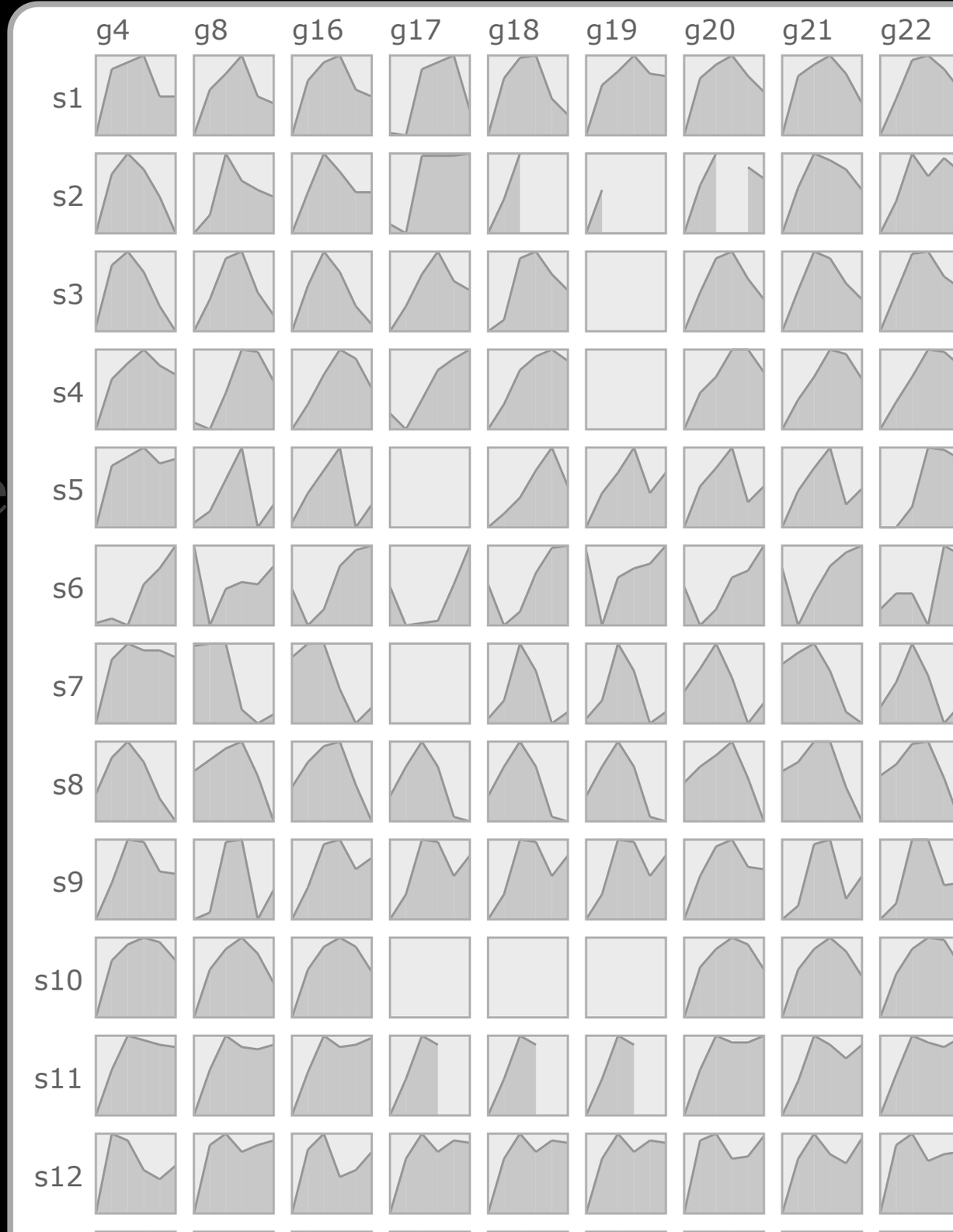
ues with spatial position

heatmap



courtesy of M. Styczynski from JavaTreeview
jtreeview.sourceforge.net/

curvemap



Pathline

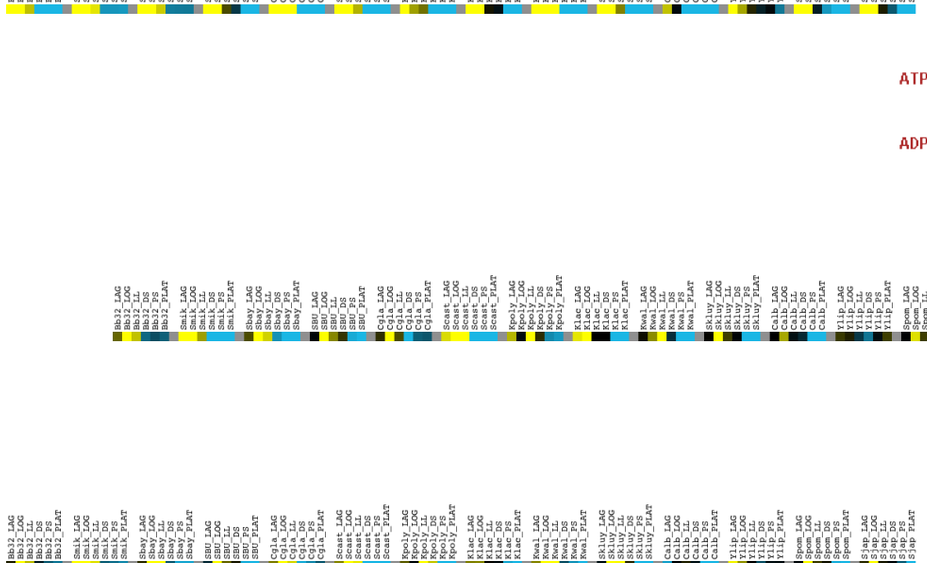
linearized pathway representation

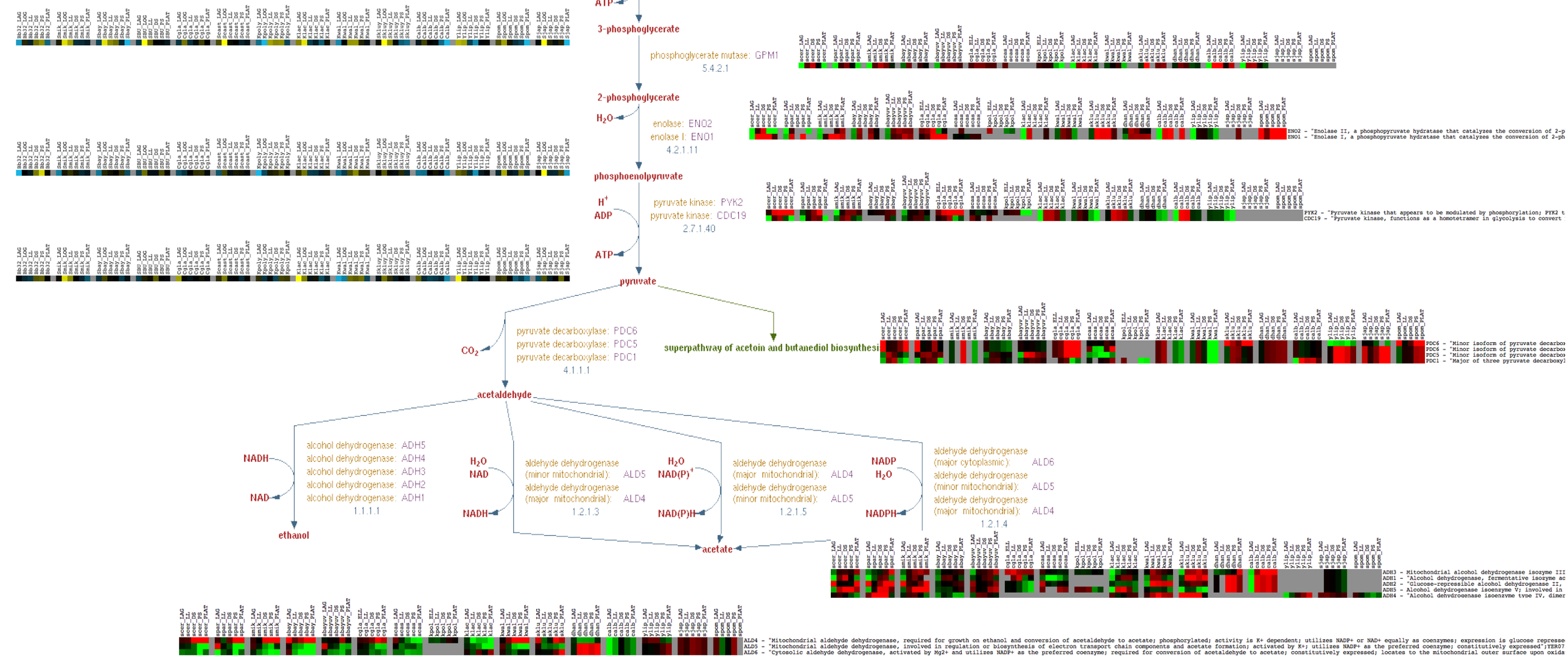
STARTING POINT

MET: yellow-up
blue-down
Sat:2x

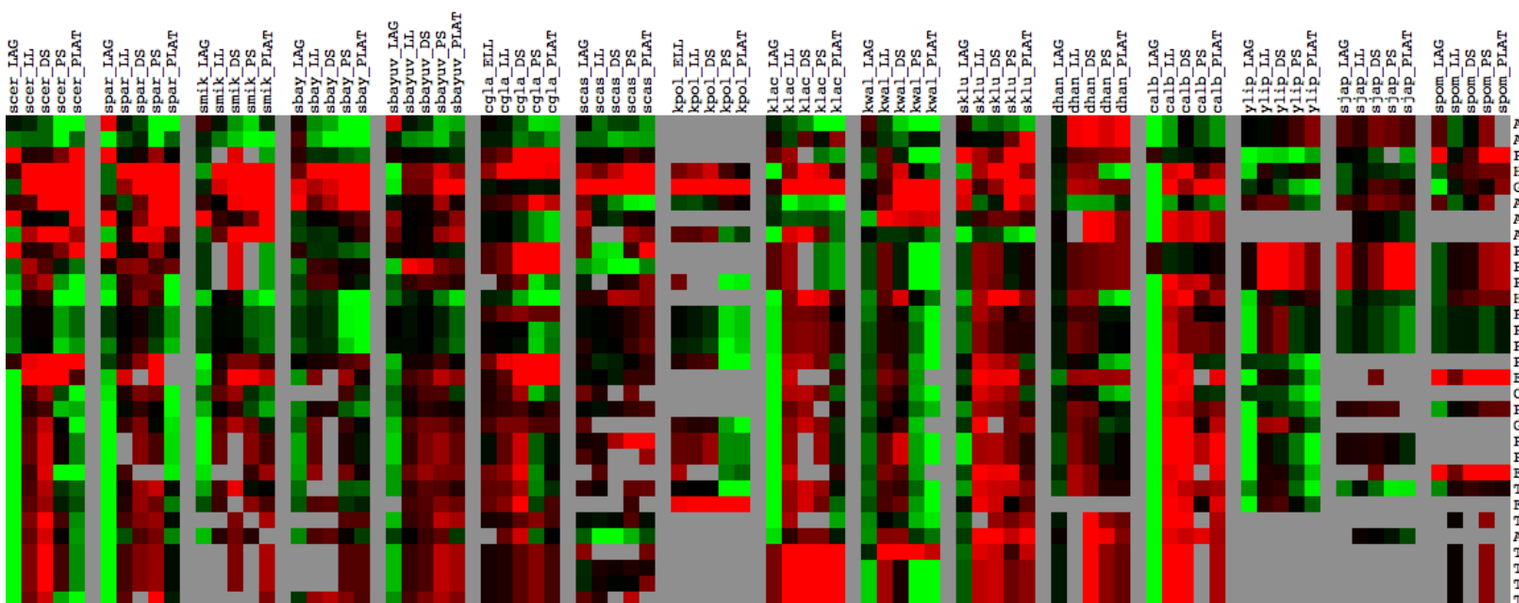
Glycolysis

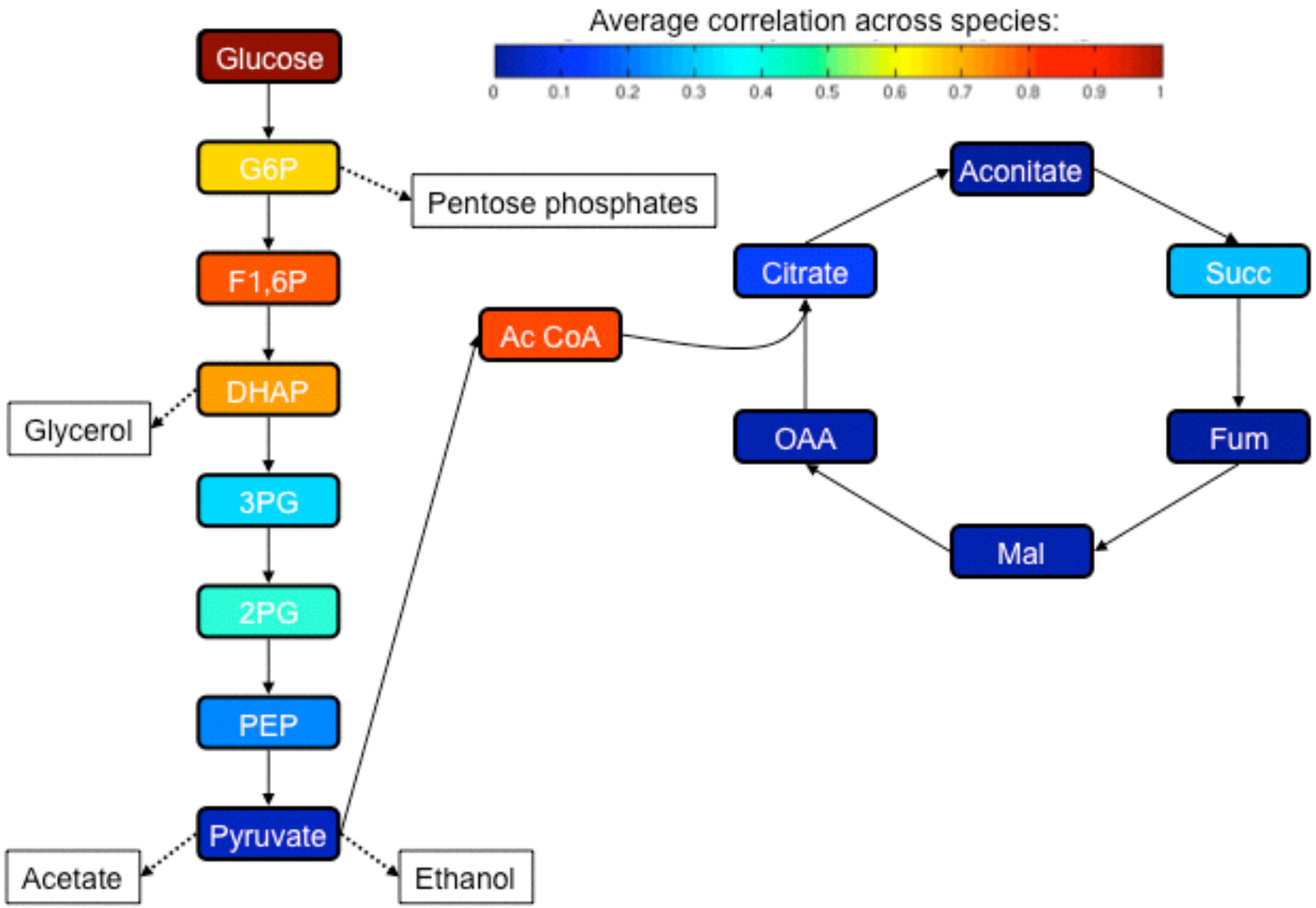
GE: red-up
green-down
Sat:3x



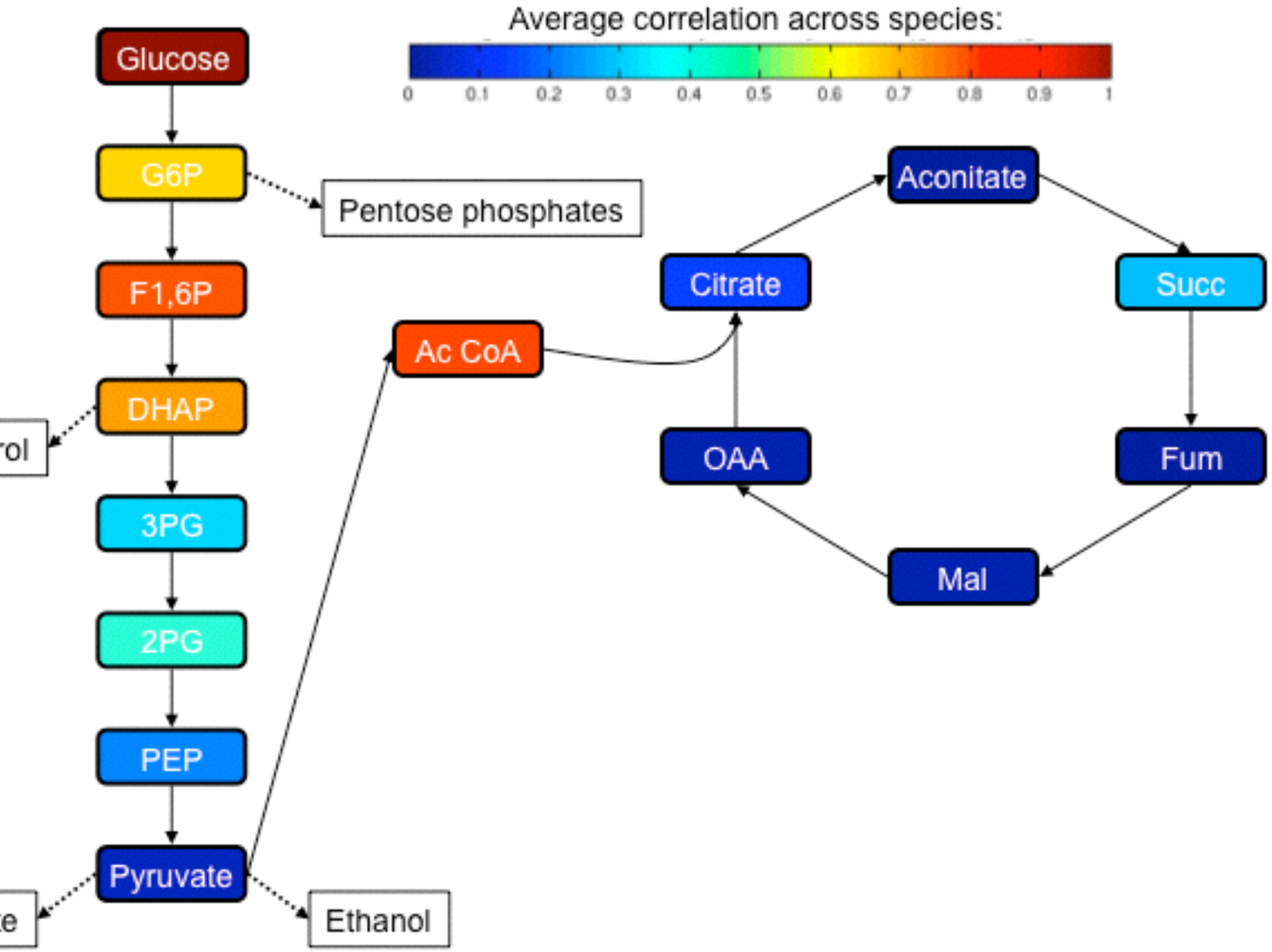

ATP
3-phosphoglycerate
phosphoglycerate mutase: GPM1
5.4.2.1
2-phosphoglycerate
H2O
enolase: ENO2
enolase I: ENO1
4.2.1.11
phosphoenolpyruvate
H+
ADP
ATP
pyruvate kinase: PYK2
pyruvate kinase: CDC19
2.7.1.40
pyruvate
CO2
pyruvate decarboxylase: PDC6
pyruvate decarboxylase: PDC5
pyruvate decarboxylase: PDC1
4.1.1.1
superpathway of acetoin and butanediol biosynthesis
acetaldehyde
NADH
NAD
alcohol dehydrogenase: ADH5
alcohol dehydrogenase: ADH4
alcohol dehydrogenase: ADH3
alcohol dehydrogenase: ADH2
alcohol dehydrogenase: ADH1
1.1.1.1
ethanol
H2O
NAD
NADH
aldehyde dehydrogenase (minor mitochondrial): ALD5
aldehyde dehydrogenase (major mitochondrial): ALD4
1.2.1.3
H2O
NAD(P)+
NAD(P)H
aldehyde dehydrogenase (major mitochondrial): ALD4
aldehyde dehydrogenase (minor mitochondrial): ALD5
1.2.1.5
NADP
H2O
NADPH
aldehyde dehydrogenase (major cytoplasmic): ALD6
aldehyde dehydrogenase (minor mitochondrial): ALD5
aldehyde dehydrogenase (major mitochondrial): ALD4
1.2.1.4
acetate
ENO2 - "Enolase II, a phosphopyruvate hydratase that catalyzes the conversion of 2-p
ENO1 - "Enolase I, a phosphopyruvate hydratase that catalyzes the conversion of 2-ph
PYK2 - "Pyruvate kinase that appears to be modulated by phosphorylation; PYK2 t
CDC19 - "Pyruvate kinase, functions as a homotetramer in glycolysis to convert
PDC6 - "Minor isoform of pyruvate decarbox
PDC6 - "Minor isoform of pyruvate decarbox
PDC5 - "Minor isoform of pyruvate decarbox
PDC1 - "Major of three pyruvate decarboxyl
ADH3 - Mitochondrial alcohol dehydrogenase isozyme III
ADH1 - "Alcohol dehydrogenase, fermentative isozyme ac
ADH2 - "Glucose-repressible alcohol dehydrogenase II,
ADH5 - Alcohol dehydrogenase isoenzyme V; involved in
ADH4 - "Alcohol dehydrogenase isoenzyme type IV, dimer

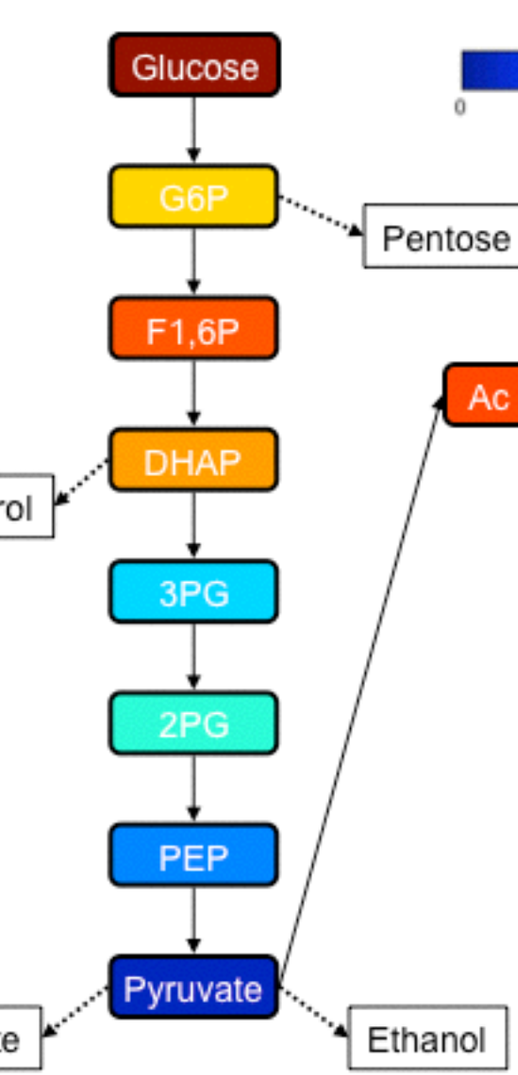
ALD4 - "Mitochondrial aldehyde dehydrogenase, required for growth on ethanol and conversion of acetaldehyde to acetate; phosphorylated; activity is K+ dependent; utilizes NADP+ or NAD+ equally as coenzymes; expression is glucose represse
ALD5 - "Mitochondrial aldehyde dehydrogenase, involved in regulation or biosynthesis of electron transport chain components and acetate formation; activated by K+; utilizes NADP+ as the preferred coenzyme; constitutively expressed";YER07
ALD6 - "Cytosolic aldehyde dehydrogenase, activated by Mg2+ and utilizes NADP+ as the preferred coenzyme; required for conversion of acetaldehyde to acetate; constitutively expressed; locates to the mitochondrial outer surface upon oxida

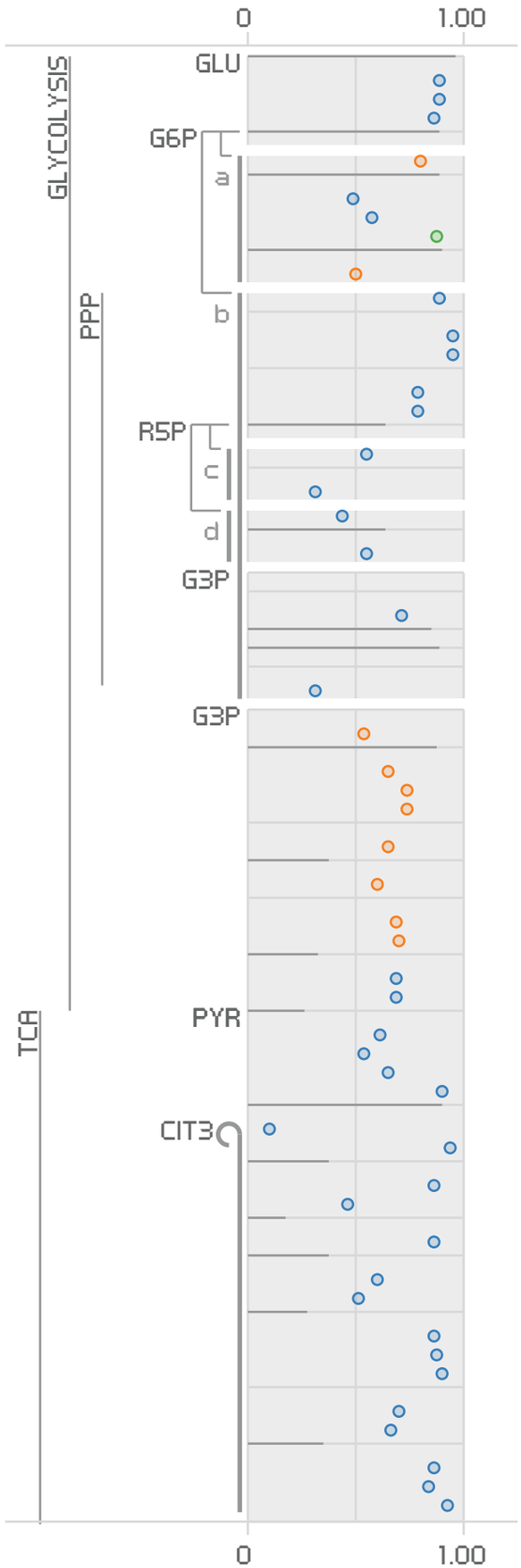

ALD5 - "Mitochondrial aldehyde dehydrogenase, involved in regulati
ALD6 - "Cytosolic aldehyde dehydrogenase, activated by Mg2+ and ut
PDC6 - "Minor isoform of pyruvate decarboxylase, decarboxylates py
HXK1 - "Hexokinase isoenzyme 1, a cytosolic protein that catalyzes
GLK1 - "Glucokinase, catalyzes the phosphorylation of glucose at C
ALD4 - "Mitochondrial aldehyde dehydrogenase, required for growth
ADH2 - "Glucose-repressible alcohol dehydrogenase II, catalyzes th
ADH5 - Alcohol dehydrogenase isoenzyme V; involved in ethanol prod
PDC6 - "Minor isoform of pyruvate decarboxylase, decarboxylates py
PDC5 - "Minor isoform of pyruvate decarboxylase, key enzyme in alc
PDC1 - "Major of three pyruvate decarboxylase isozymes, key enzyme
HXK2 - Hexokinase isoenzyme 2 that catalyzes phosphorylation of gl
PFK1 - "Alpha subunit of heterooctameric phosphofructokinase invol
PFK1 - "Alpha subunit of heterooctameric phosphofructokinase invol
PFK2 - "Beta subunit of heterooctameric phosphofructokinase involv
PYK2 - "Pyruvate kinase that appears to be modulated by phosphoryl
ENO1 - "Enolase I, a phosphopyruvate hydratase that catalyzes the
CDC19 - "Pyruvate kinase, functions as a homotetramer in glycolysi
PGI1 - "Glycolytic enzyme phosphoglucose isomerase, catalyzes the
GPM1 - "Tetrameric phosphoglycerate mutase, mediates the conversio
PGK1 - "3-phosphoglycerate kinase, catalyzes transfer of high-ener
PGK1 - "3-phosphoglycerate kinase, catalyzes transfer of high-ener
ENO2 - "Enolase II, a phosphopyruvate hydratase that catalyzes the
TPI1 - "Triose phosphate isomerase, abundant glycolytic enzyme; mR
FBA1 - "Fructose 1,6-bisphosphate aldolase, required for glycolysi
TDH3 - "Glyceraldehyde-3-phosphate dehydrogenase, isozyme 3, invol
ADH1 - "Alcohol dehydrogenase, fermentative isozyme active as homo
TDH3 - "Glyceraldehyde-3-phosphate dehydrogenase, isozyme 3, invol
TDH3 - "Glyceraldehyde-3-phosphate dehydrogenase, isozyme 3, invol
TDH3 - "Glyceraldehyde-3-phosphate dehydrogenase, isozyme 3, invol
TDH2 - "Glyceraldehyde-3-phosphate dehydrogenase, isozyme 2, invol
TDH1 - "Glyceraldehyde-3-phosphate dehydrogenase, isozyme 1, invol
ADH3 - Mitochondrial alcohol dehydrogenase isozyme III; involved i
ADH4 - "Alcohol dehydrogenase isoenzyme type IV, dimeric enzyme de

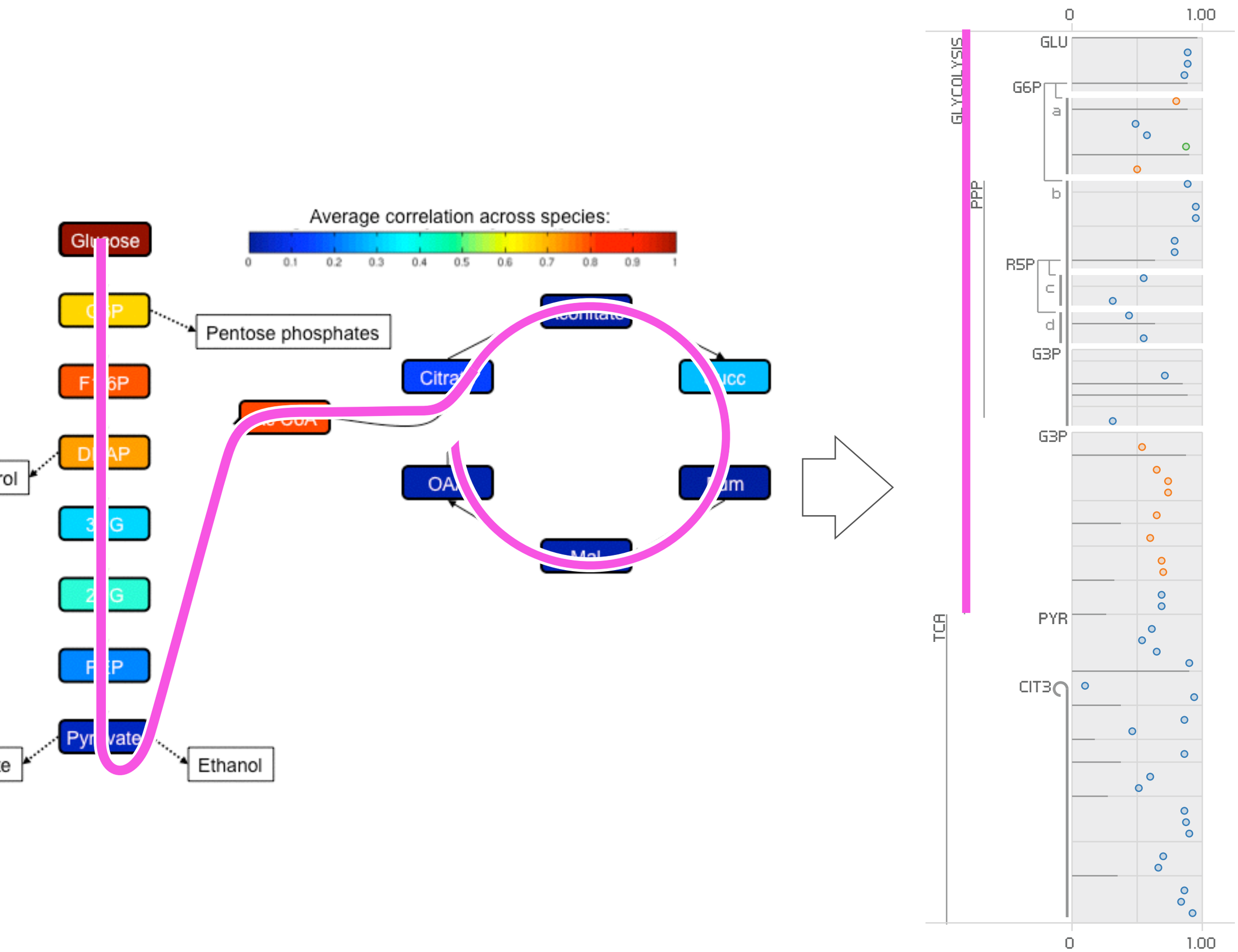


Rainbow Color Maps (Still) Considered Harmful.
D. Borland and R. Taylor, Computer Graphics and Applications, 2007.


Average correlation across species:
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1
Glucose
G6P
Pentose phosphates
F1,6P
DHAP
rol
3PG
2PG
PEP
Pyruvate
te
Ethanol
Ac CoA
Citrate
Aconitate
Succ
OAA
Fum
Mal

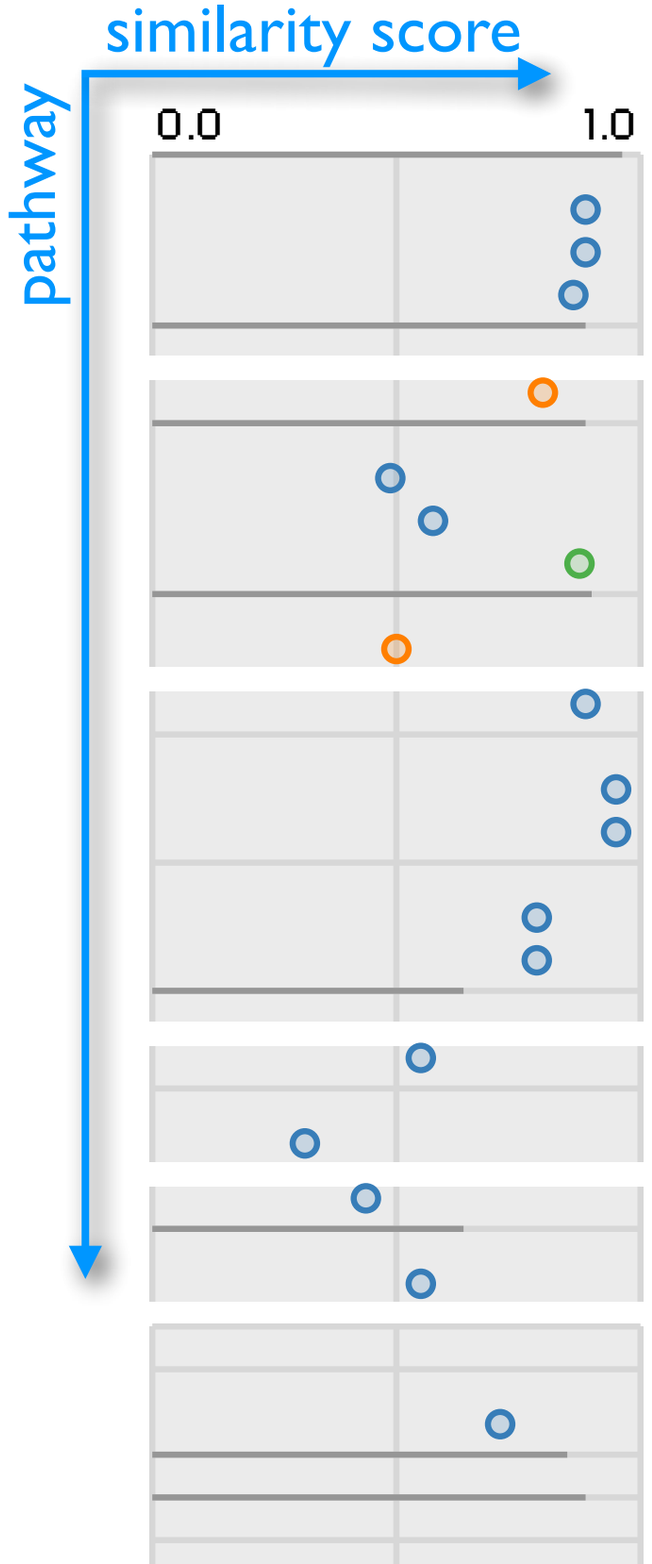

Glucose
G6P
Pentose phosphates
F1,6P
DHAP
3PG
2PG
PEP
Pyruvate
Ethanol
Ac CoA
Citrate
Aconitate
Succ
Fum
Mal
OAA
Average correlation across species:
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1


0
1.00
GLYCOLYSIS
PPP
TCA
GLU
G6P
a
b
R5P
c
d
G3P
G3P
PYR
CIT3
0
1.00

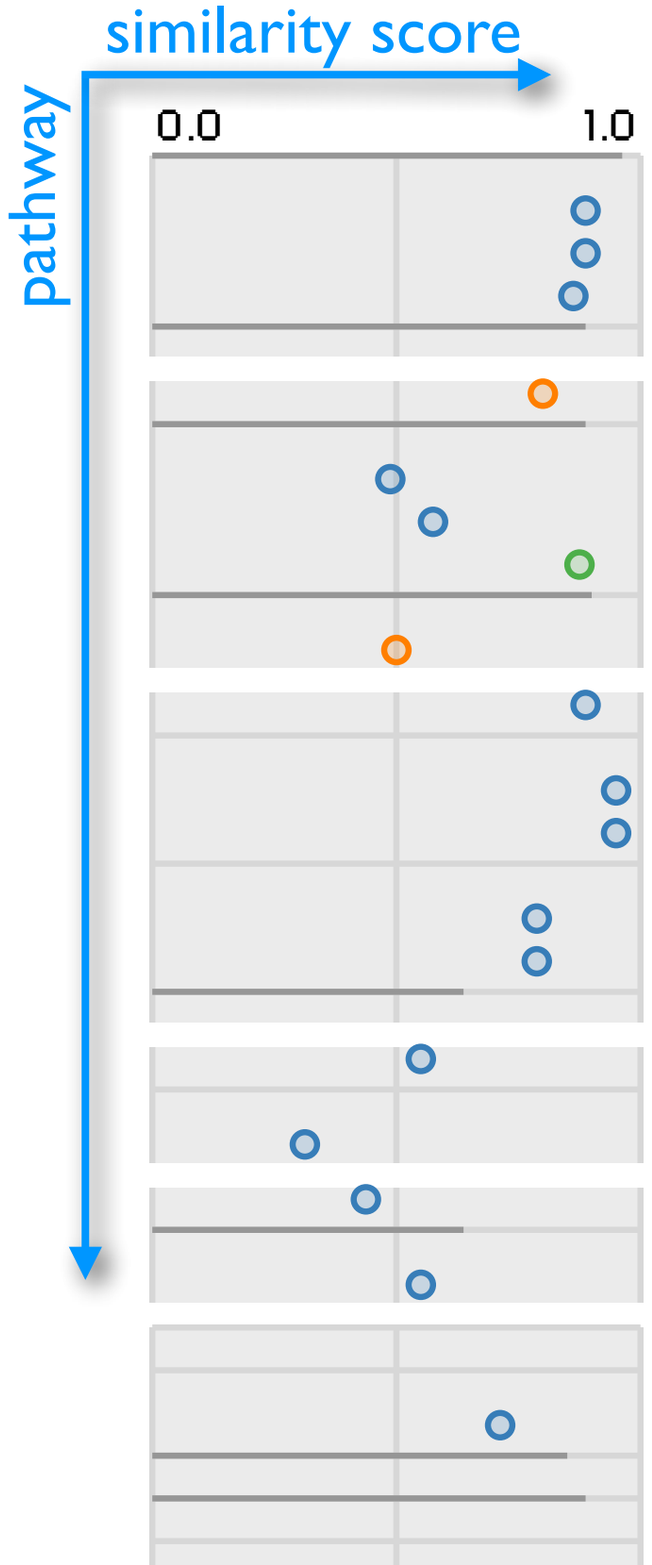

Average correlation across species:
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1
Glucose
Pentose phosphates
Ethanol
GLYCOLYSIS
PPP
TCA
GLU
G6P
a
b
R5P
c
d
G3P
G3P
PYR
CIT3
0
1.00
0
1.00

linearized pathway representation

common axes to compare similarity scores



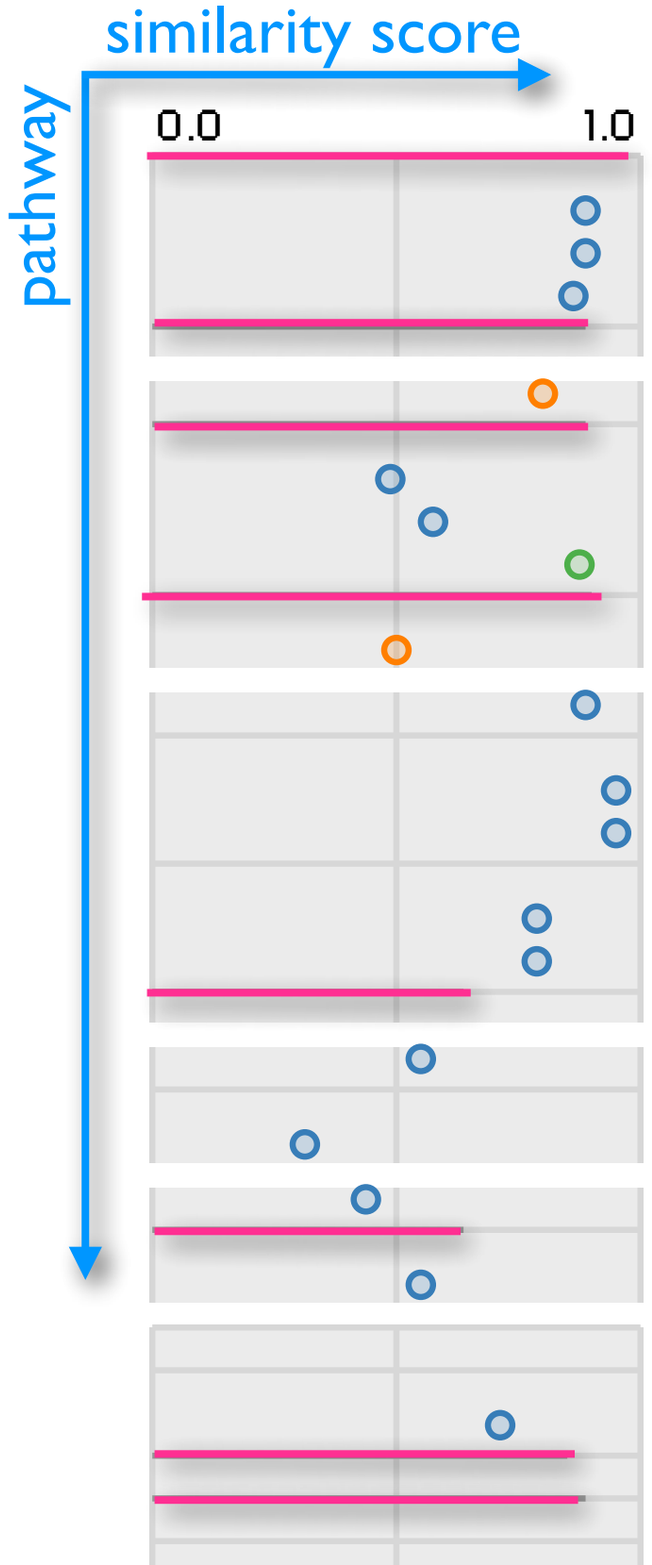
linearized pathway representation



common axes to compare similarity scores

- bars and circles
  - visual layers for selective attention
  - color-code gene direction

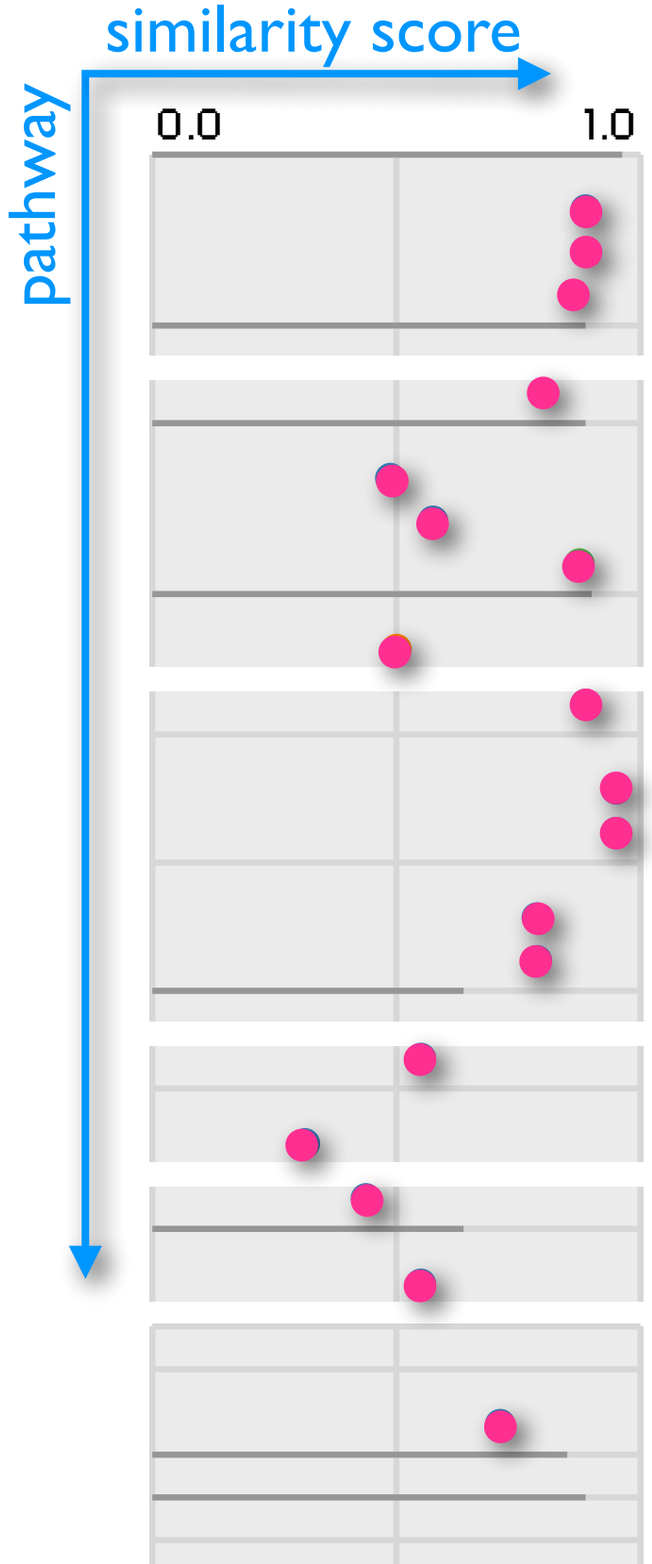
linearized pathway representation



common axes to compare similarity scores

- bars and circles
  - visual layers for selective attention
  - color-code gene direction

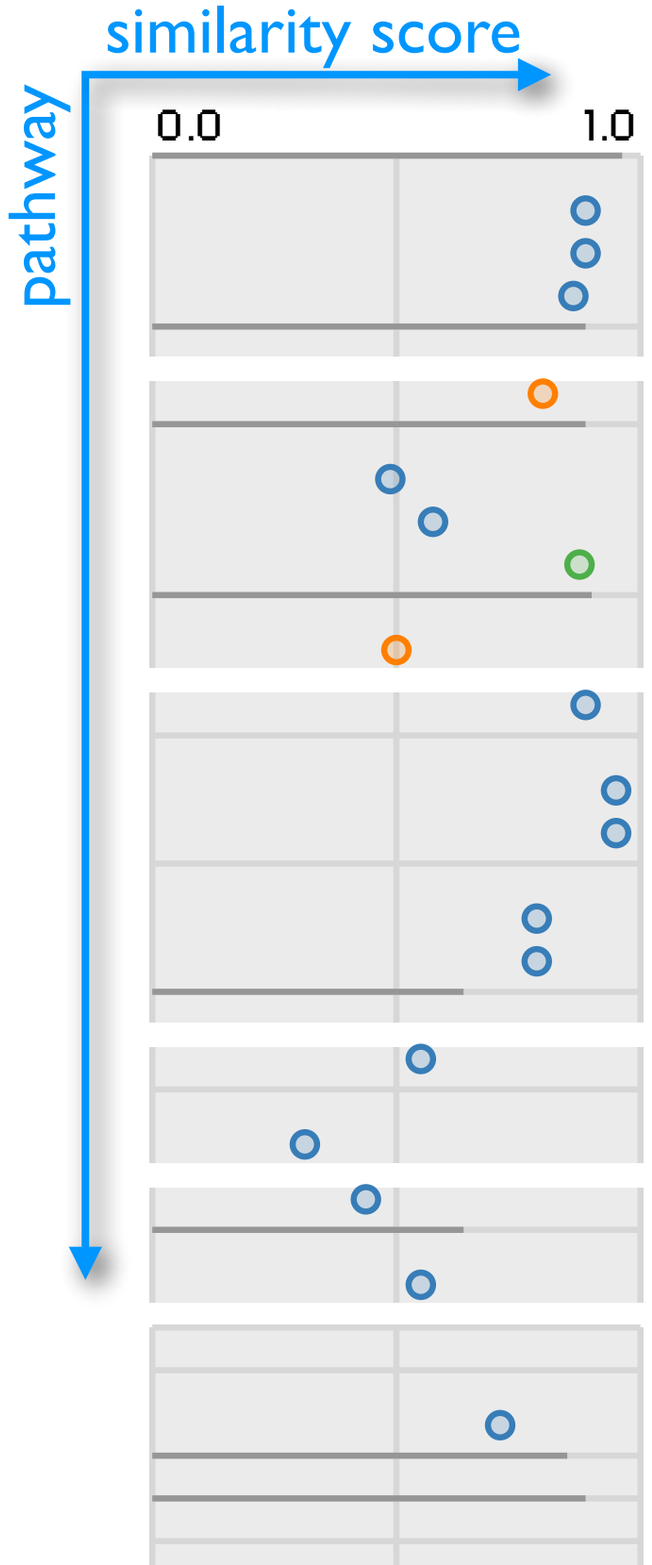
linearized pathway representation



common axes to compare similarity scores

- bars and circles
  - visual layers for selective attention
  - color-code gene direction

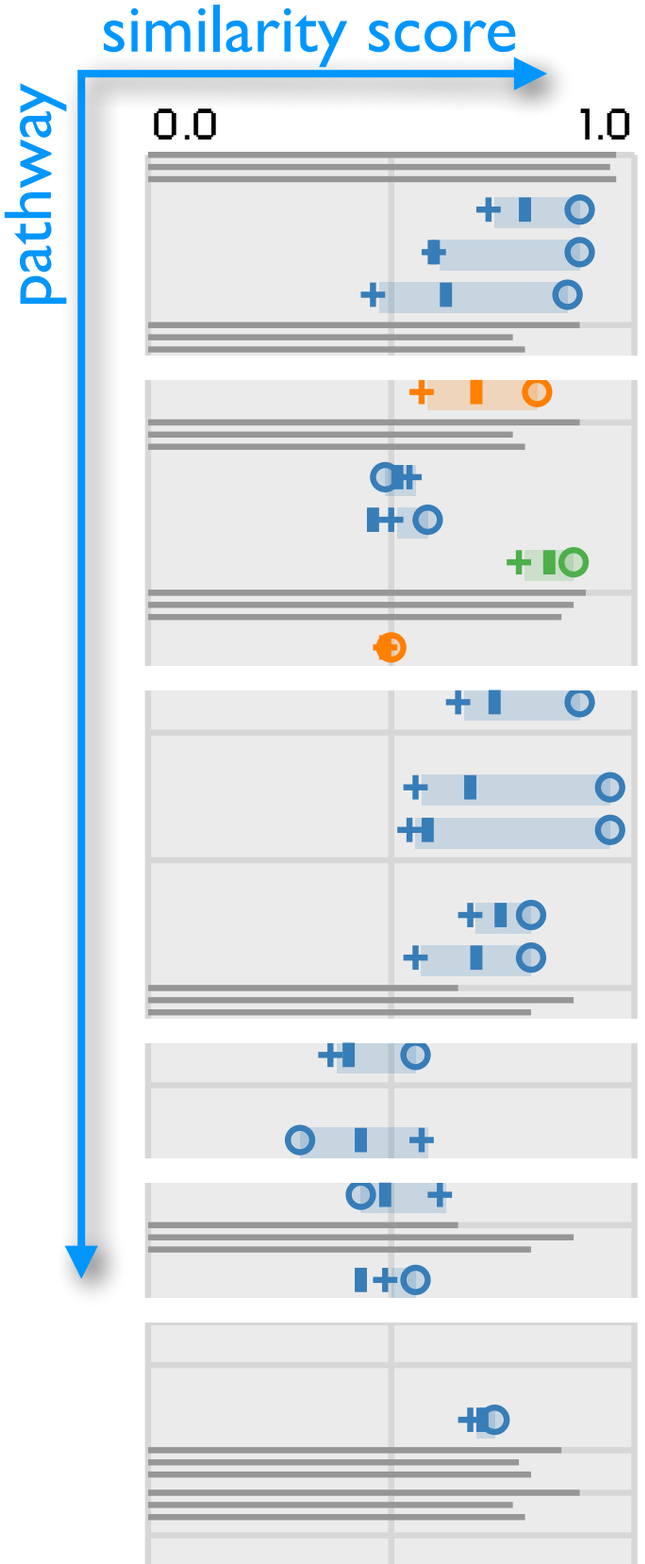
linearized pathway representation



common axes to compare similarity scores

- bars and circles
  - visual layers for selective attention
  - color-code gene direction

linearized pathway representation



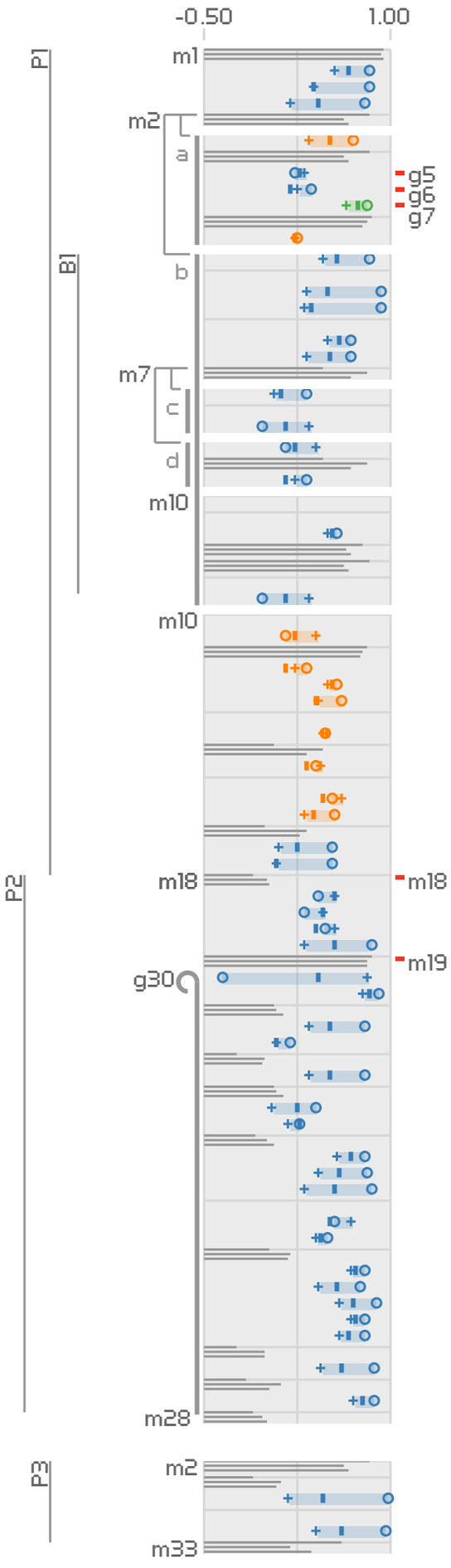
common axes to compare similarity scores

- bars and circles
    - visual layers for selective attention
    - color-code gene direction
- multiple similarity scores

linearized pathway representation

common axes to compare similarity scores

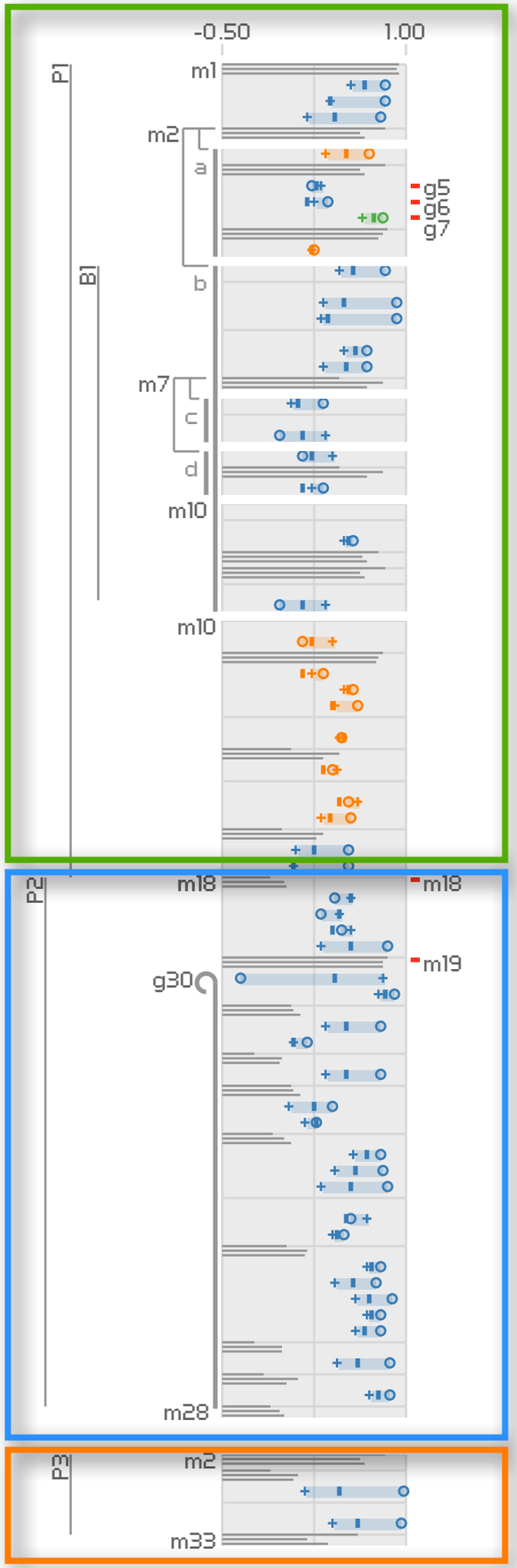
- bars and circles
  - visual layers for selective attention
  - color-code gene direction
- multiple similarity scores
- multiple pathways



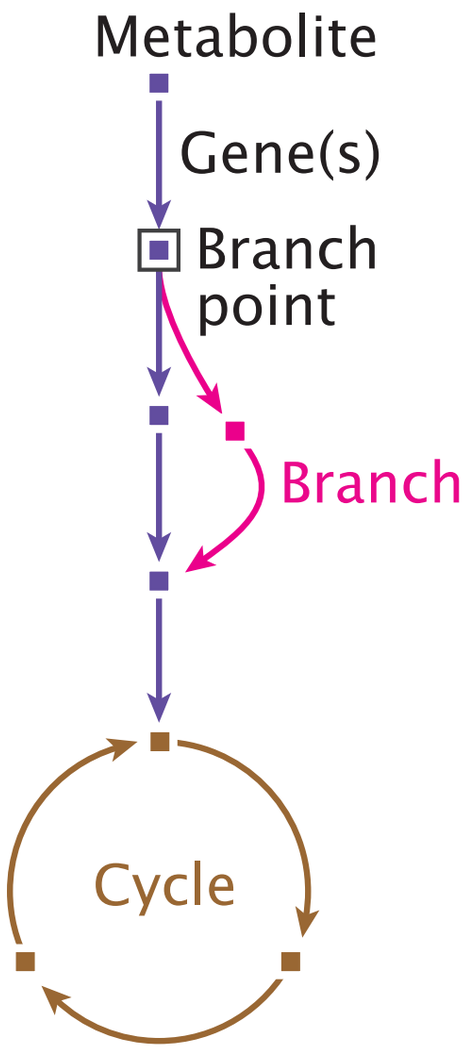
linearized pathway representation

common axes to compare similarity scores

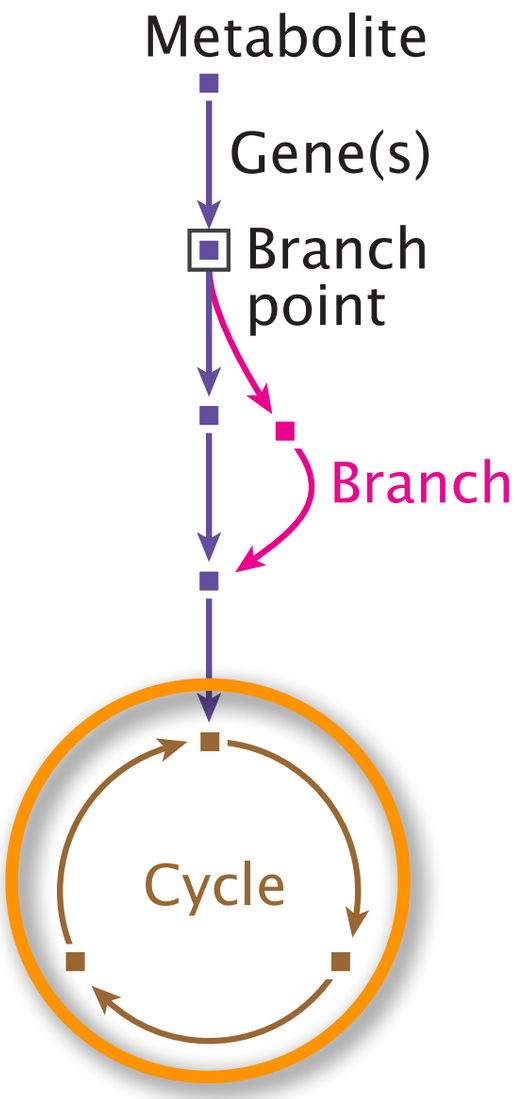
- bars and circles
    - visual layers for selective attention
    - color-code gene direction
- multiple similarity scores
- multiple pathways



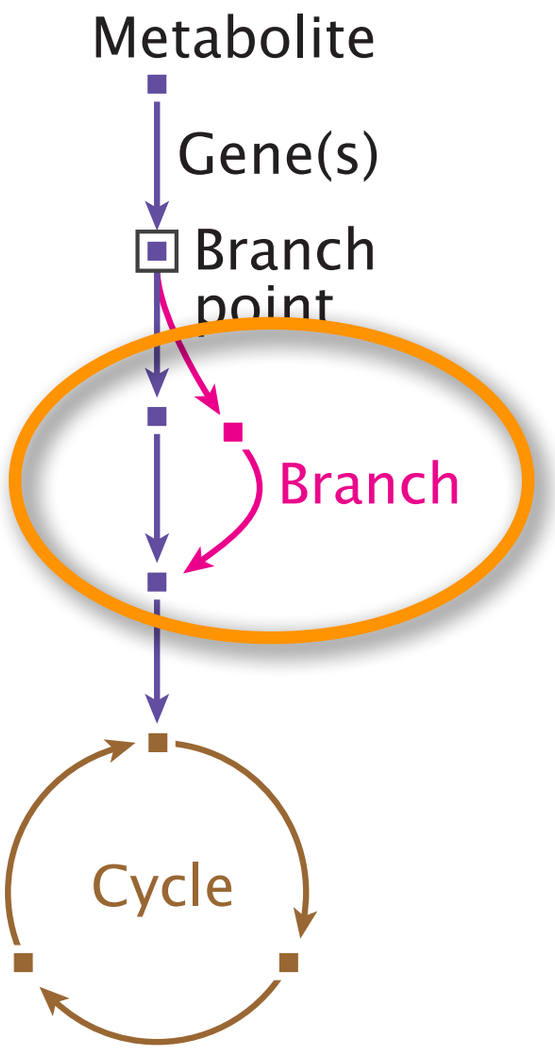
pathway to ordered list of nodes



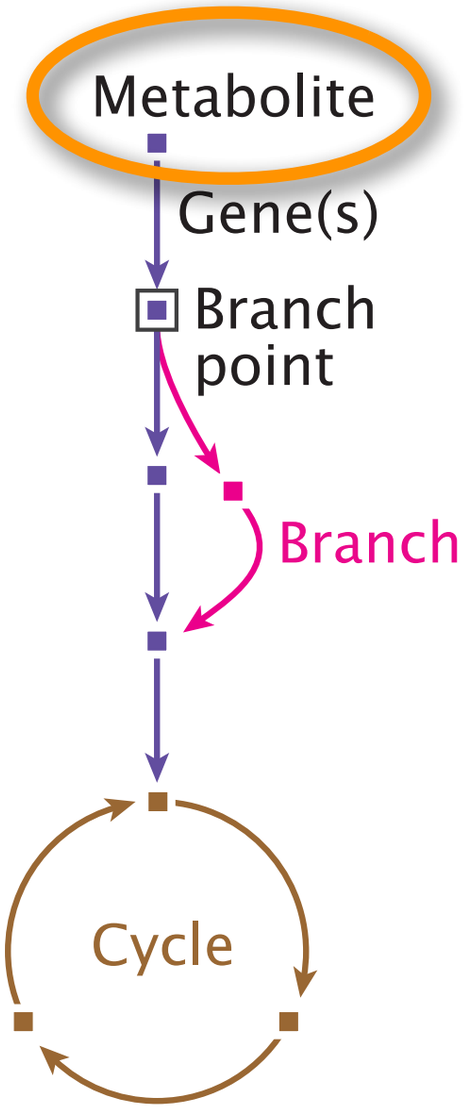
pathway to ordered list of nodes



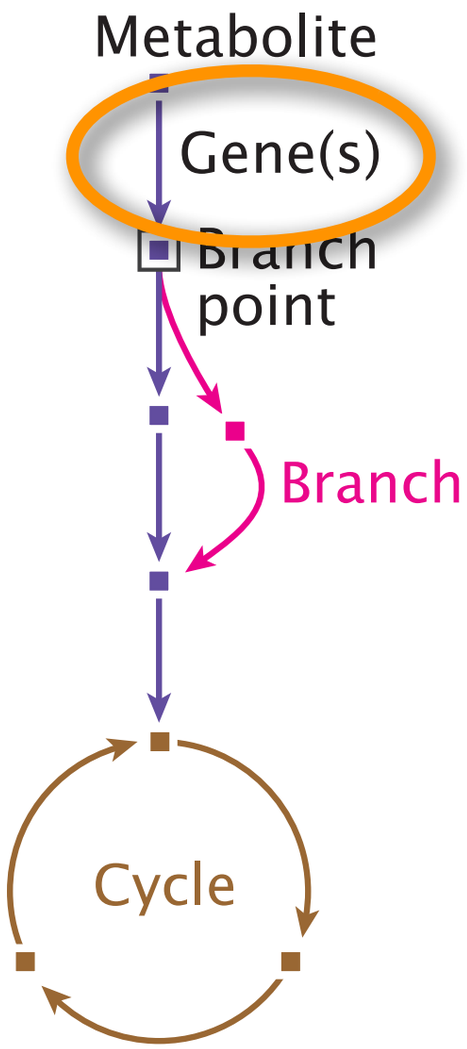
pathway to ordered list of nodes



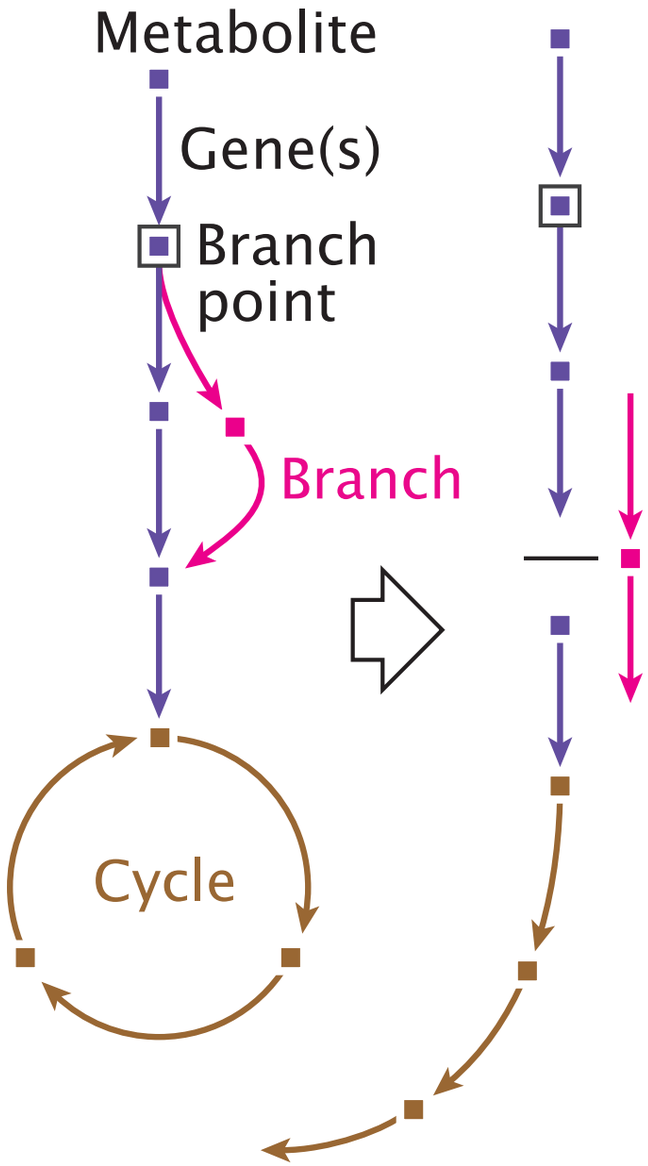
pathway to ordered list of nodes



pathway to ordered list of nodes

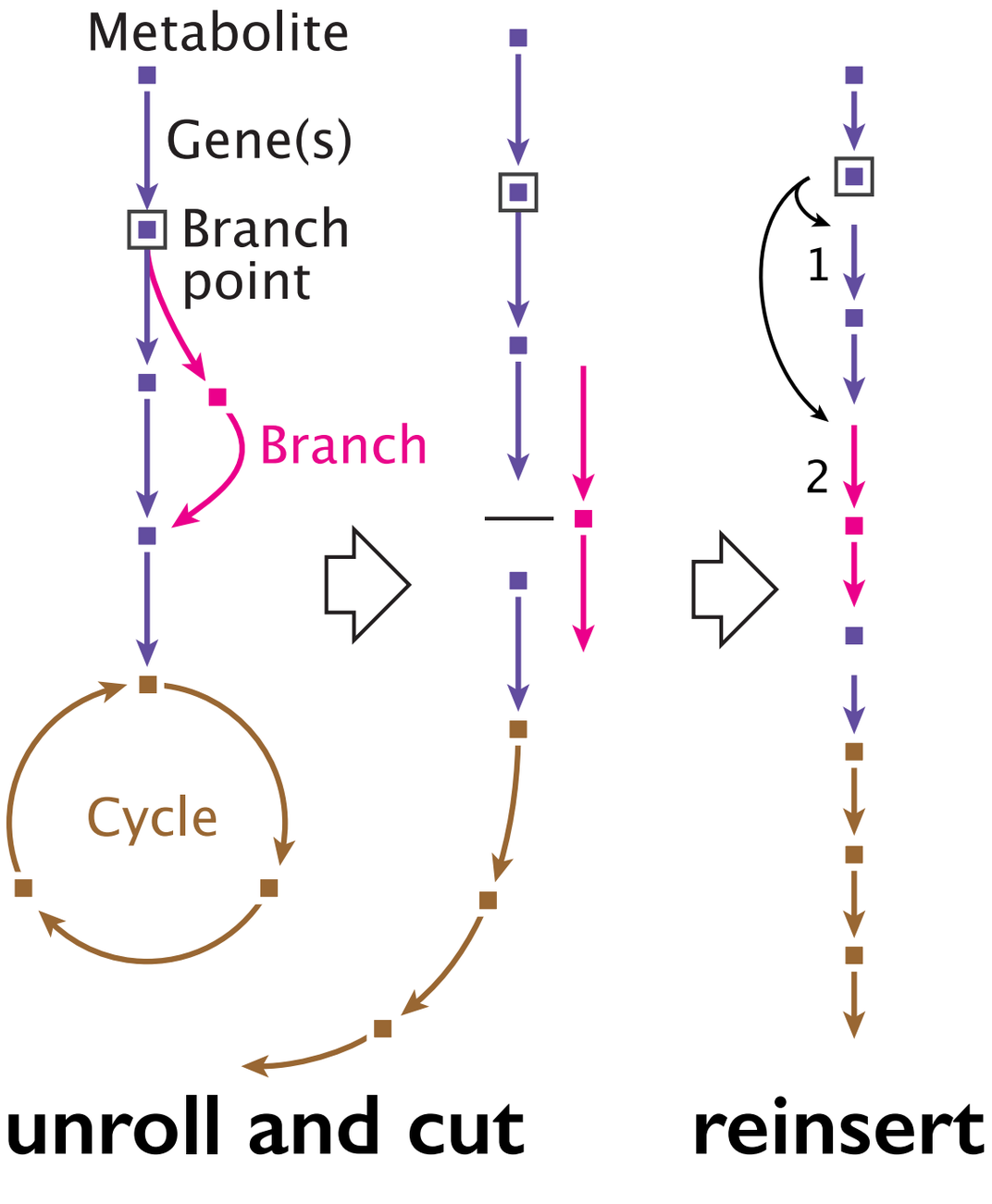


pathway to ordered list of nodes

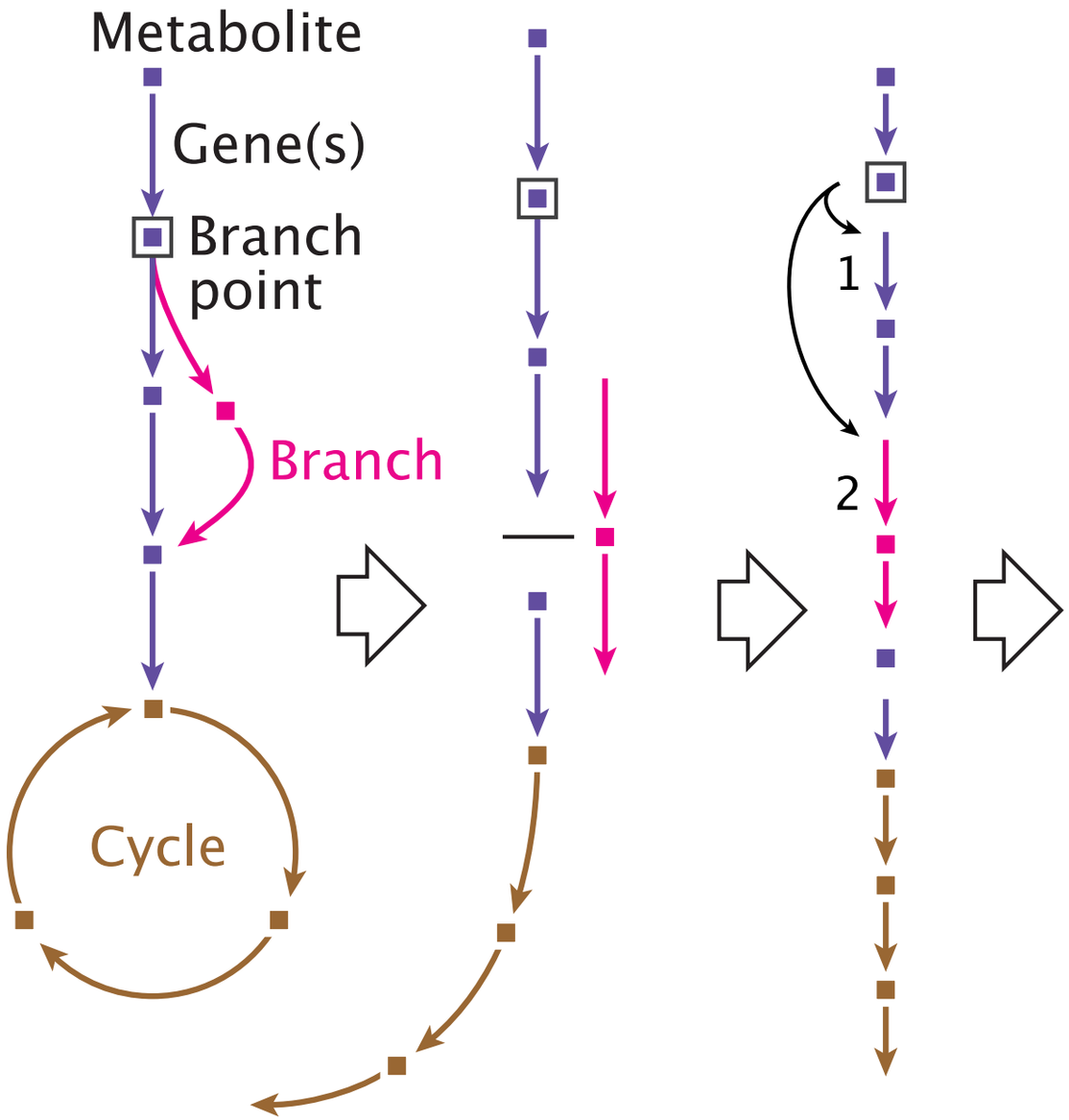


unroll and cut

pathway to ordered list of nodes

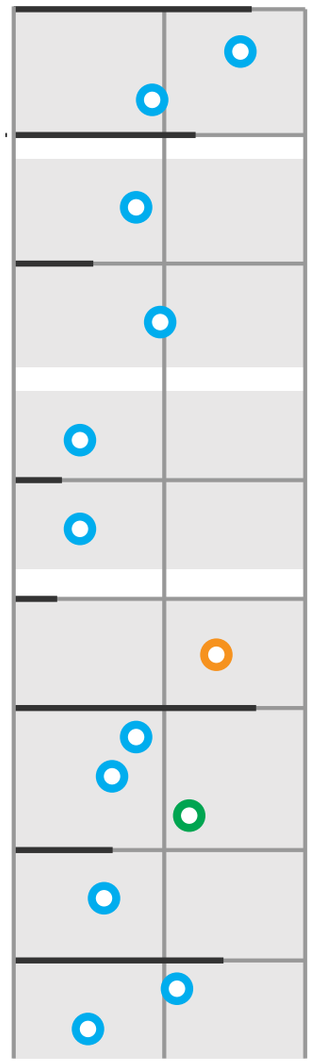


pathway to ordered list of nodes



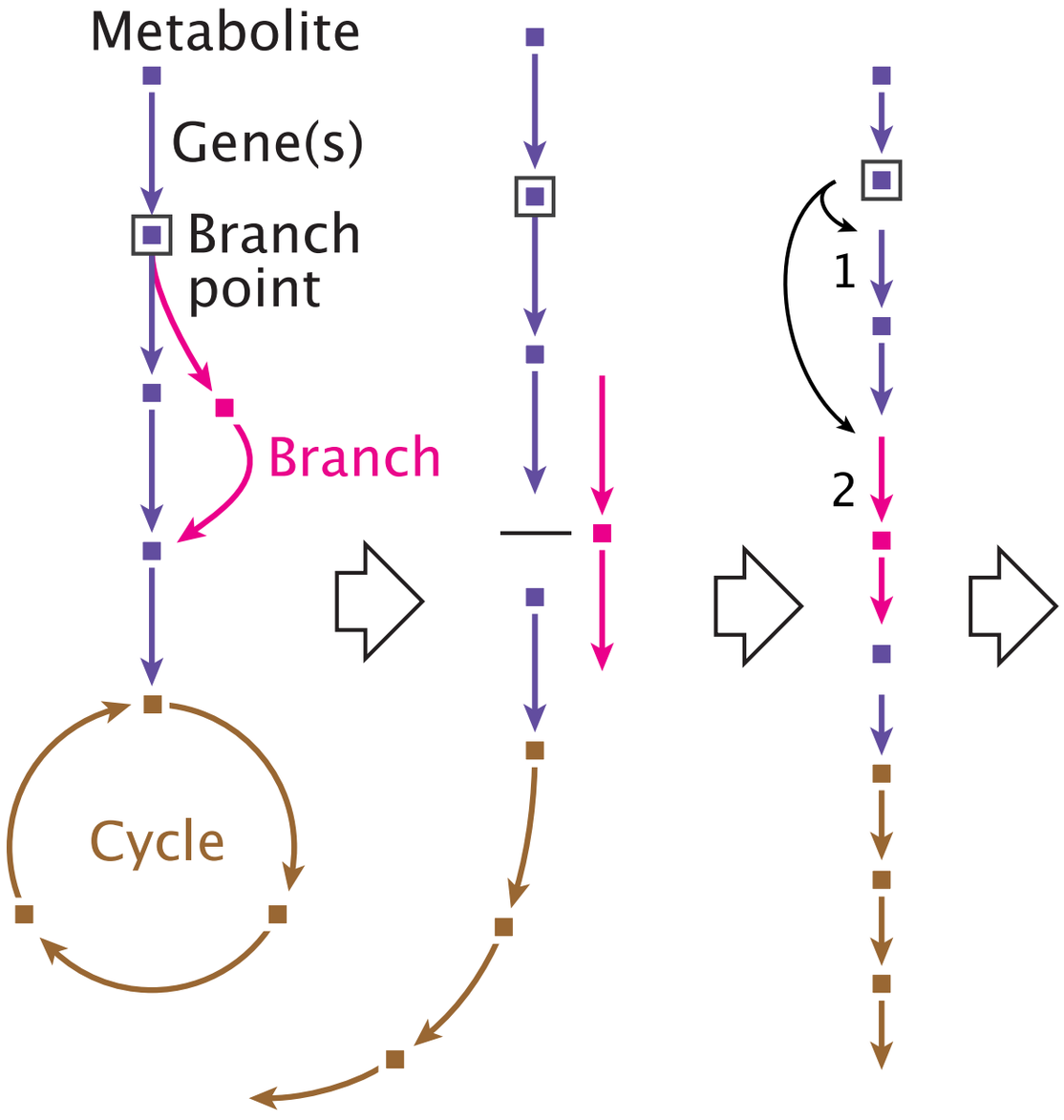
unroll and cut

reinsert



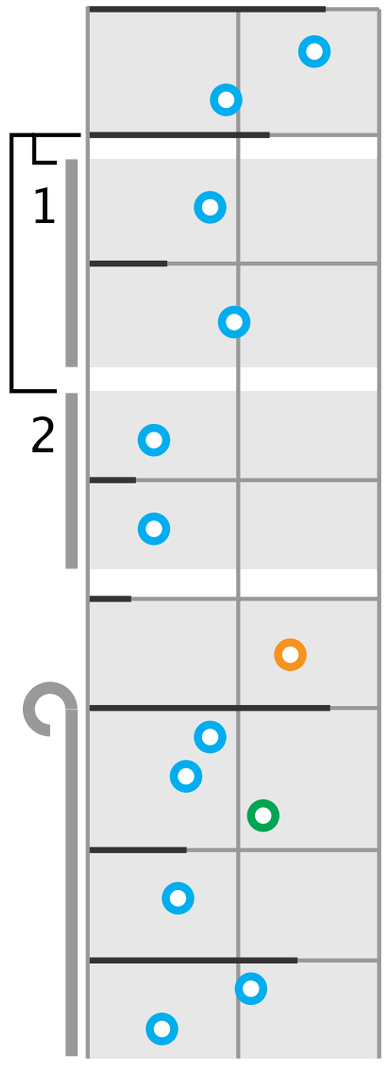
shared coordinate frame and

pathway to ordered list of nodes



unroll and cut

reinsert

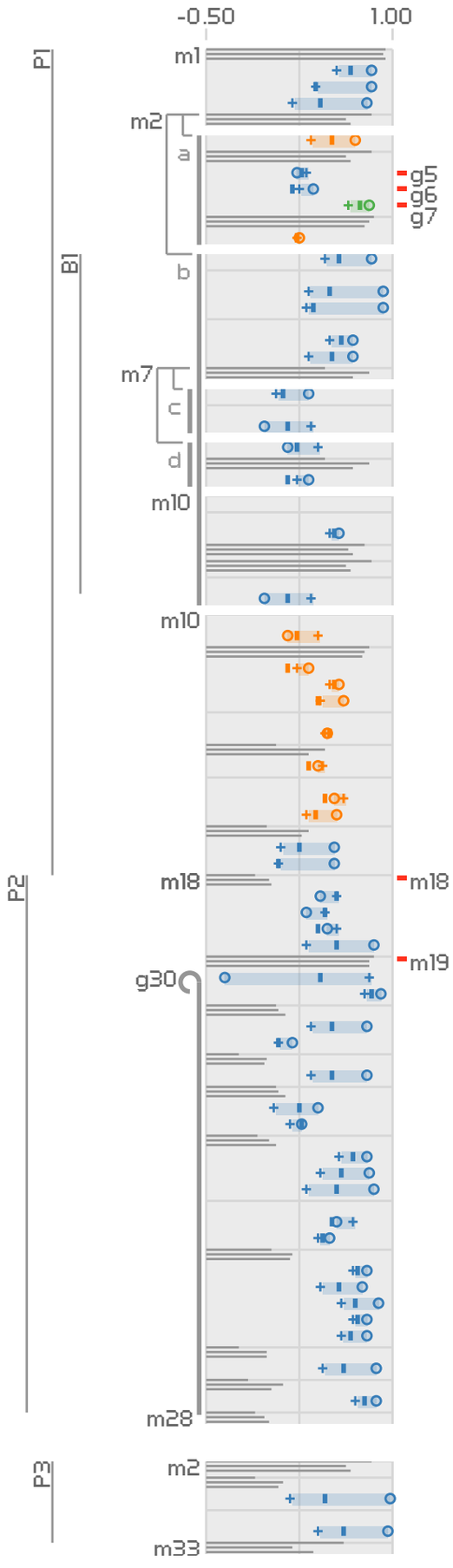


shared coordinate frame and stylized marks

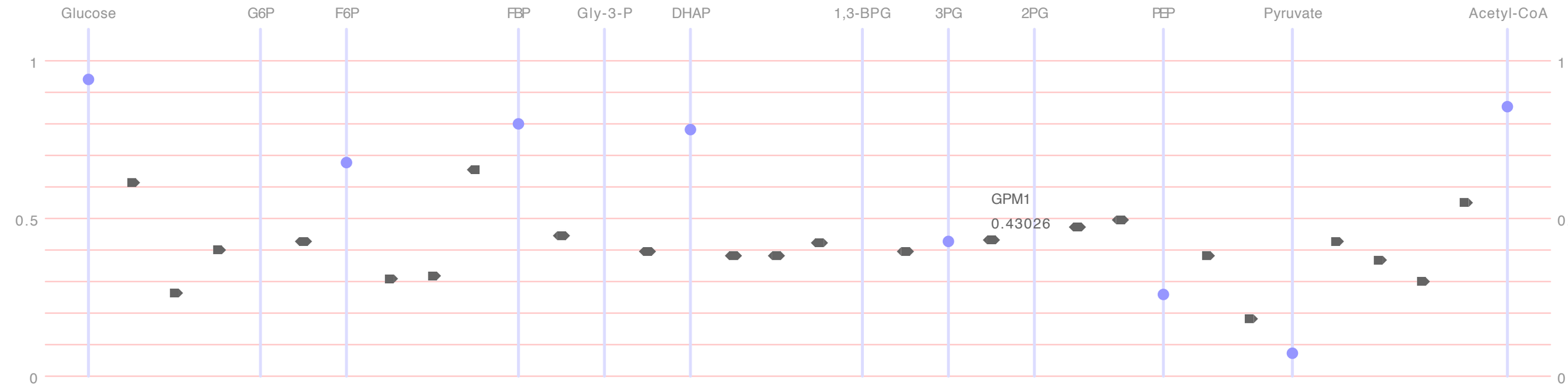
linearized pathway representation

putting it together . . .

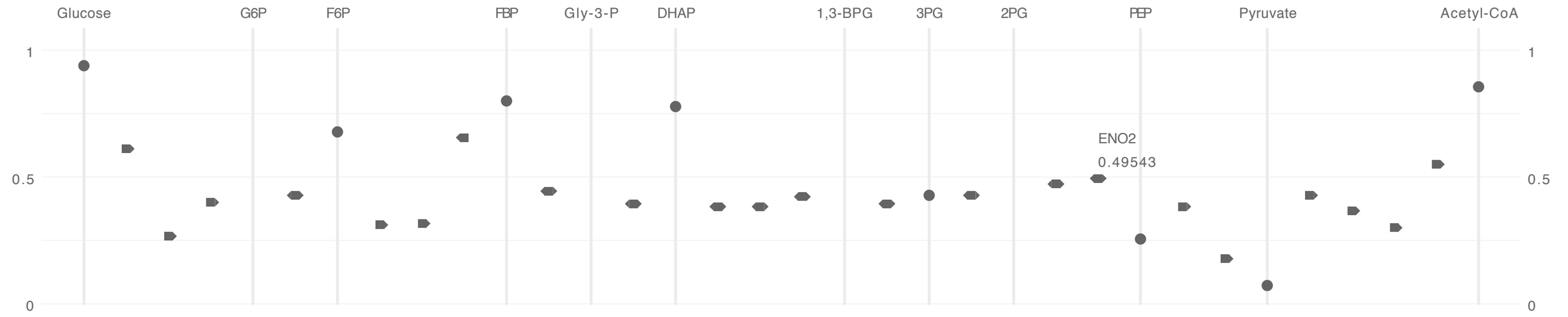
- use spatial position for similarity scores
- topology is secondary



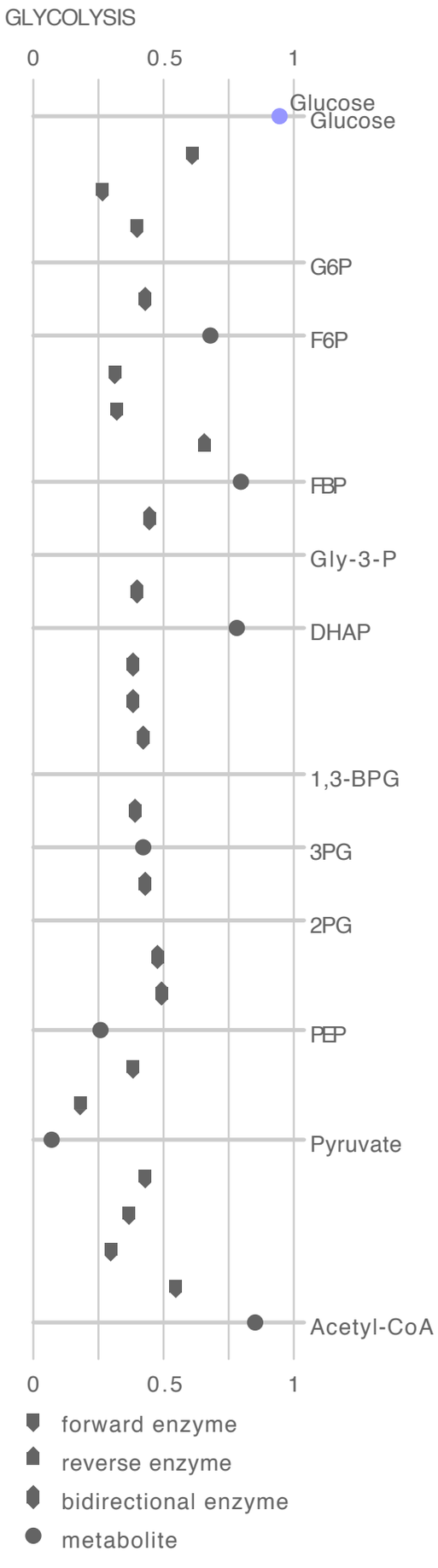
PAPER PROTOTYPES



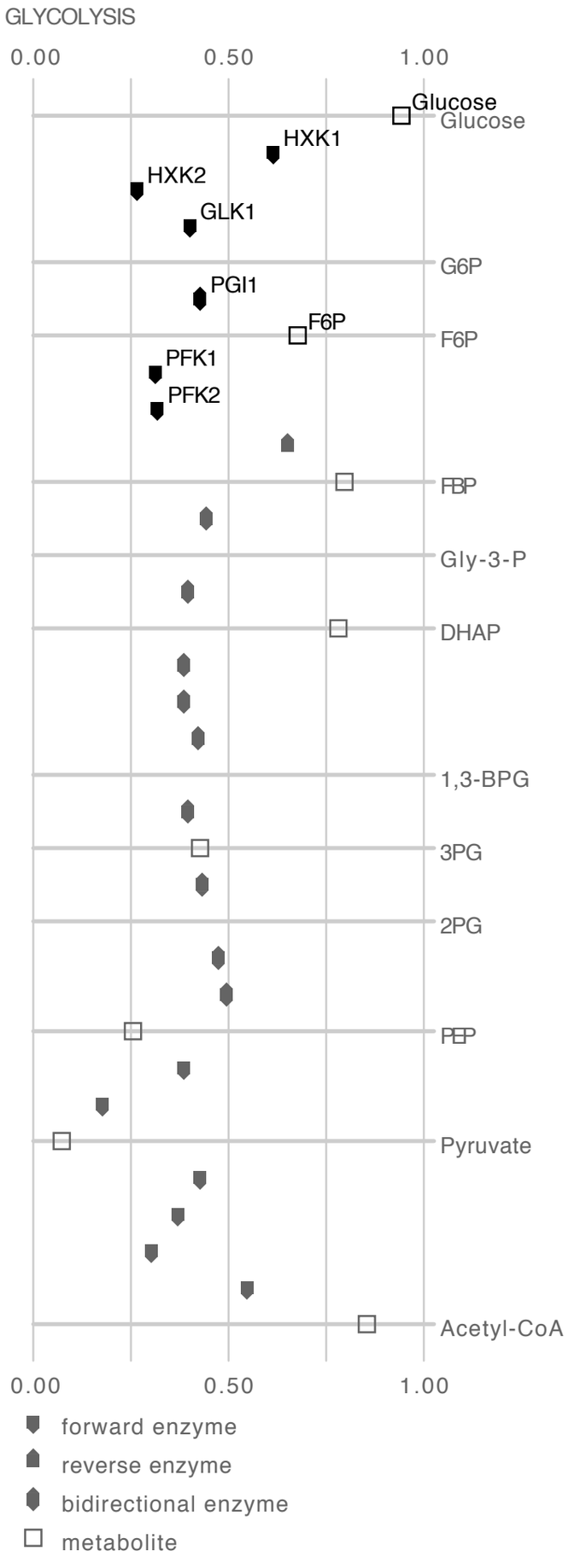
orientation & marks



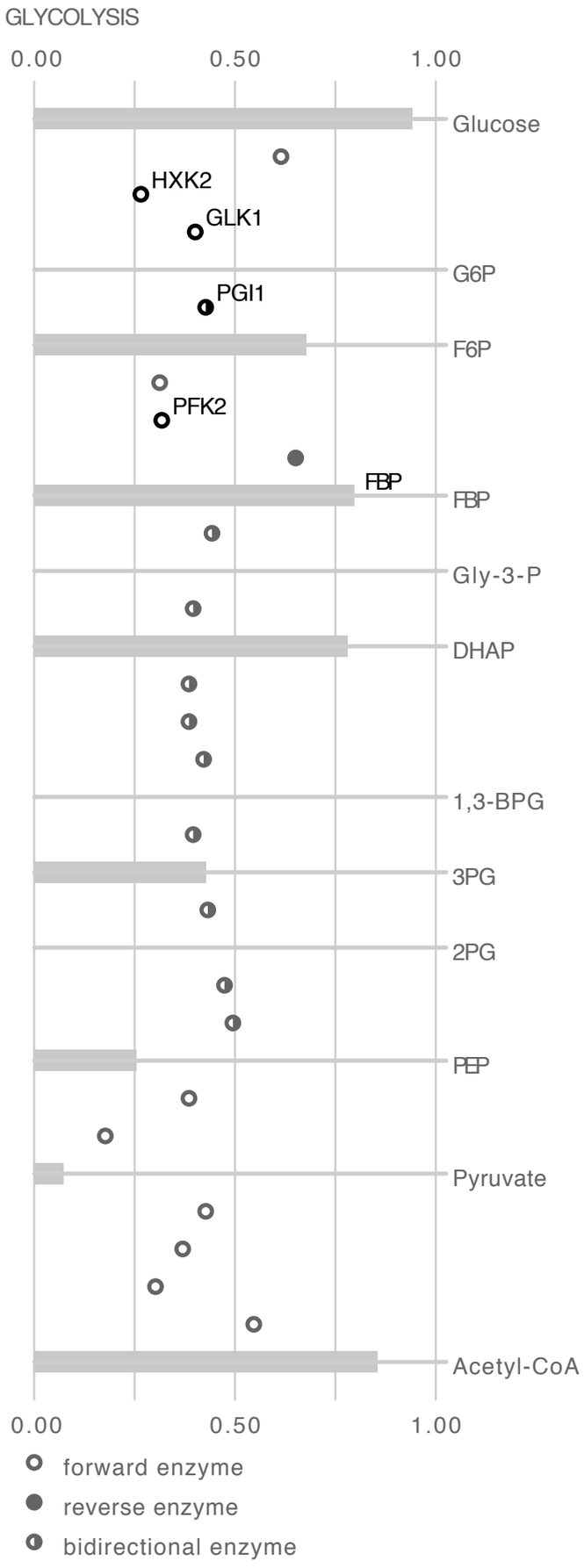
orientation & marks



orientation & marks

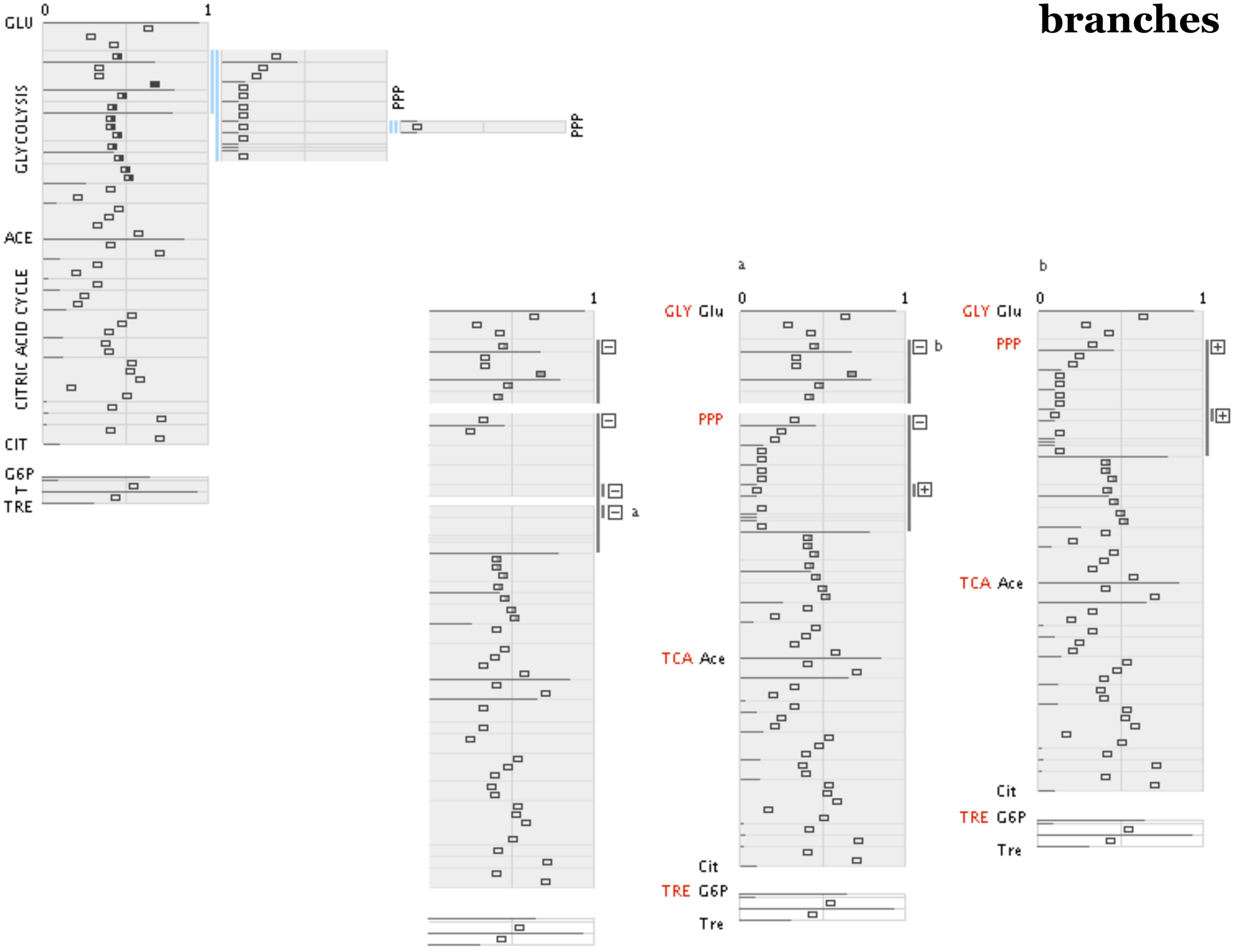


orientation & marks

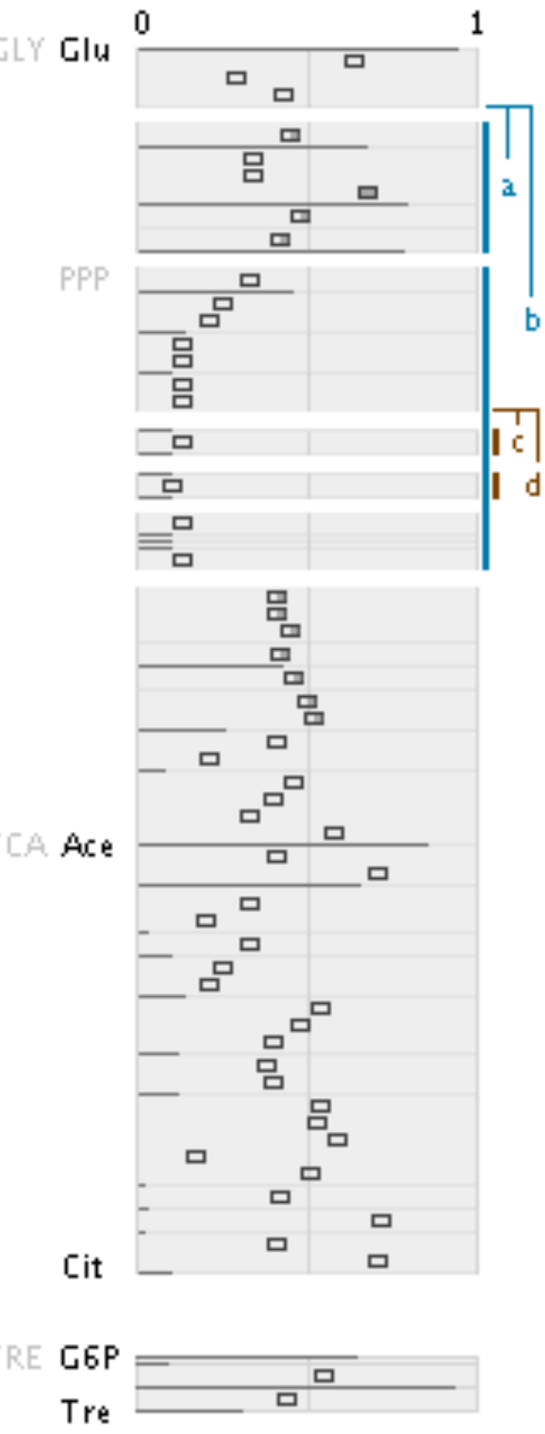


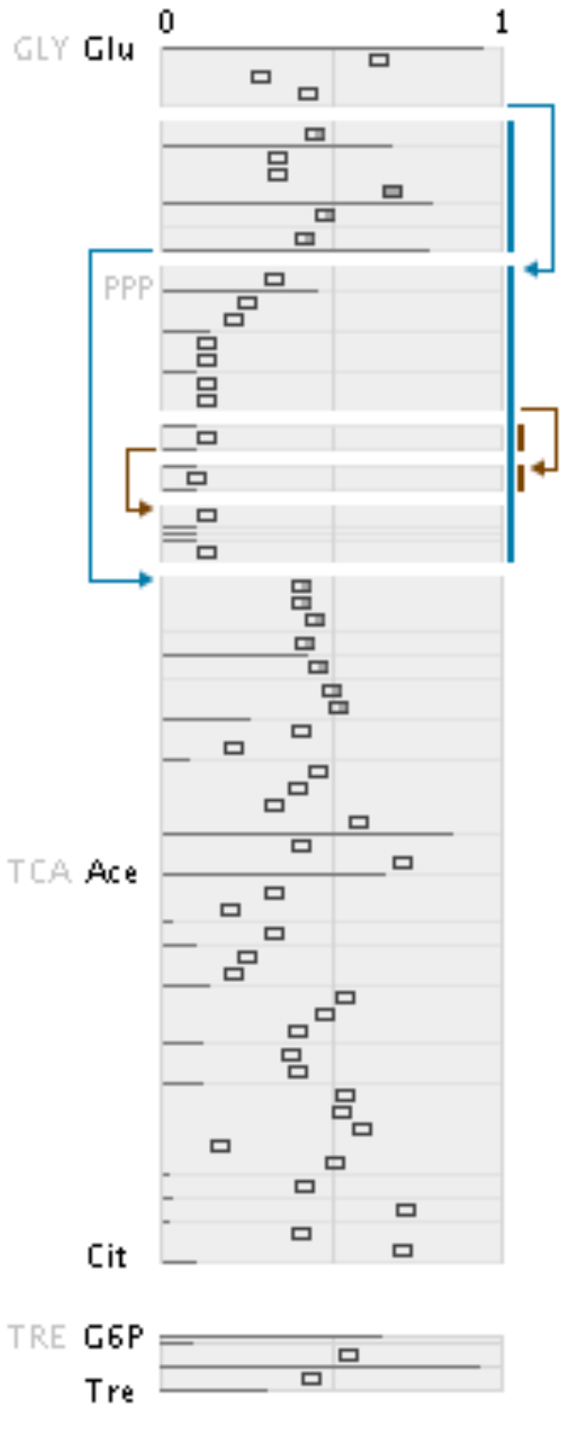
orientation & marks

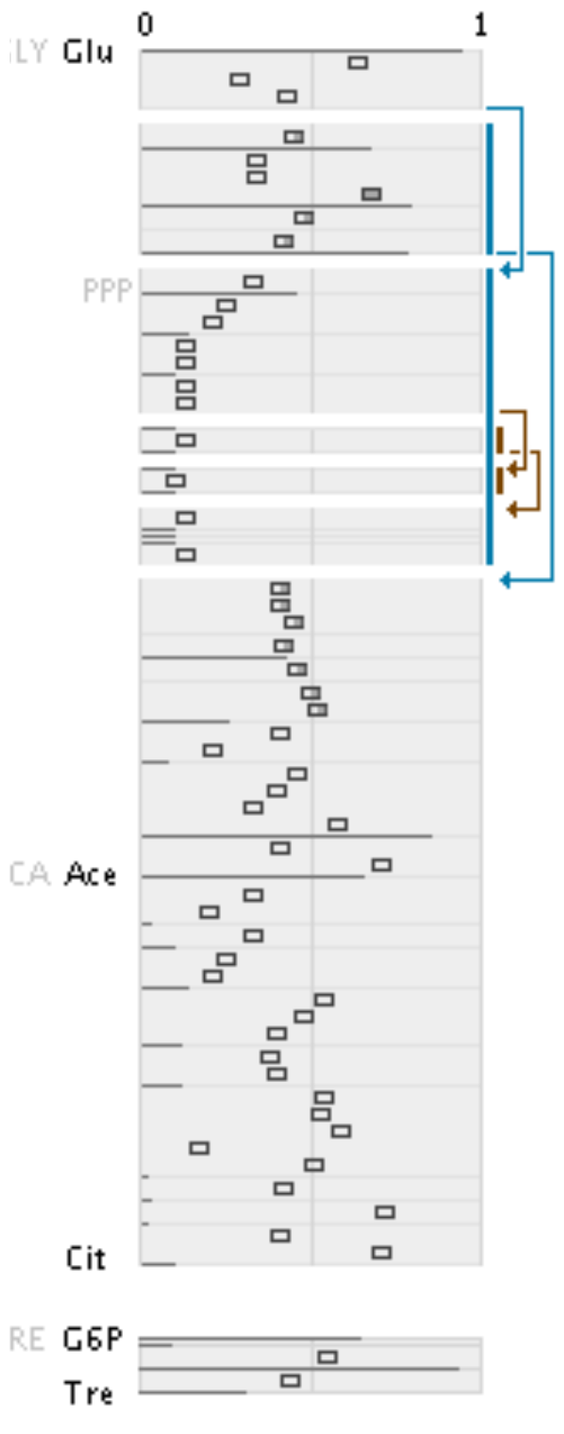
branches

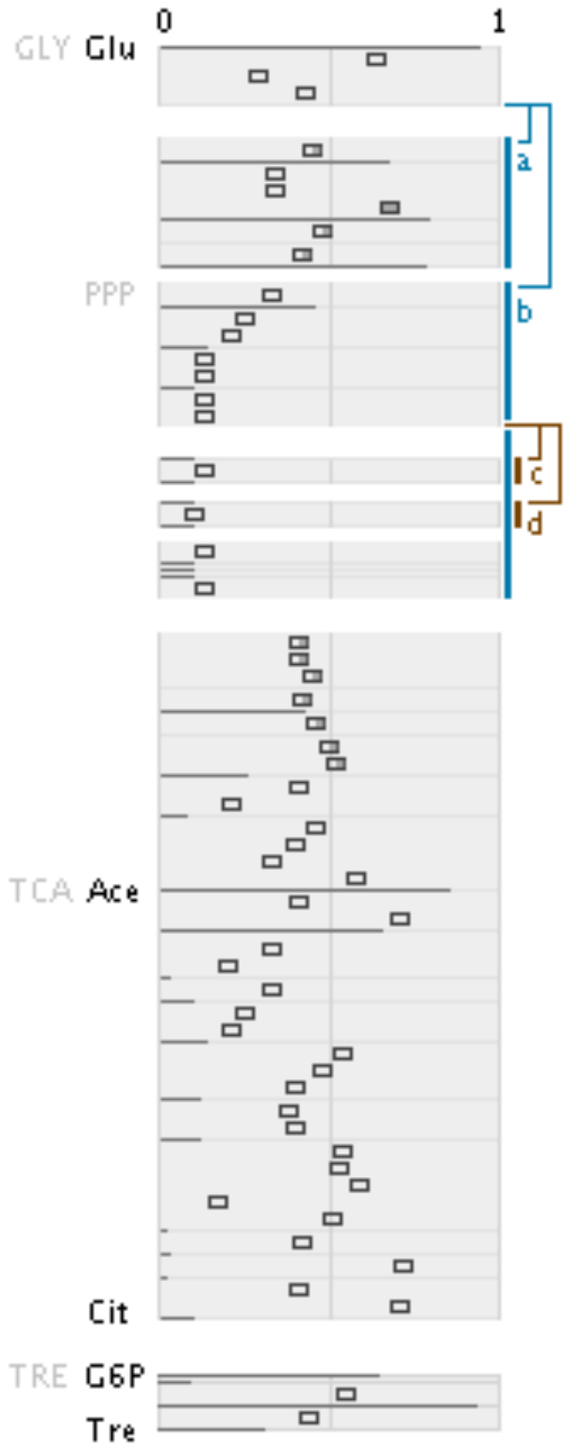

GLU
GLYCOLYSIS
ACE
CITRIC ACID CYCLE
CIT
G6P
TRE
PPP
PPP
a
b
GLY Glu
PPP
TCA Ace
Cit
TRE G6P
Tre
GLY Glu
PPP
TCA Ace
Cit
TRE G6P
Tre

branches

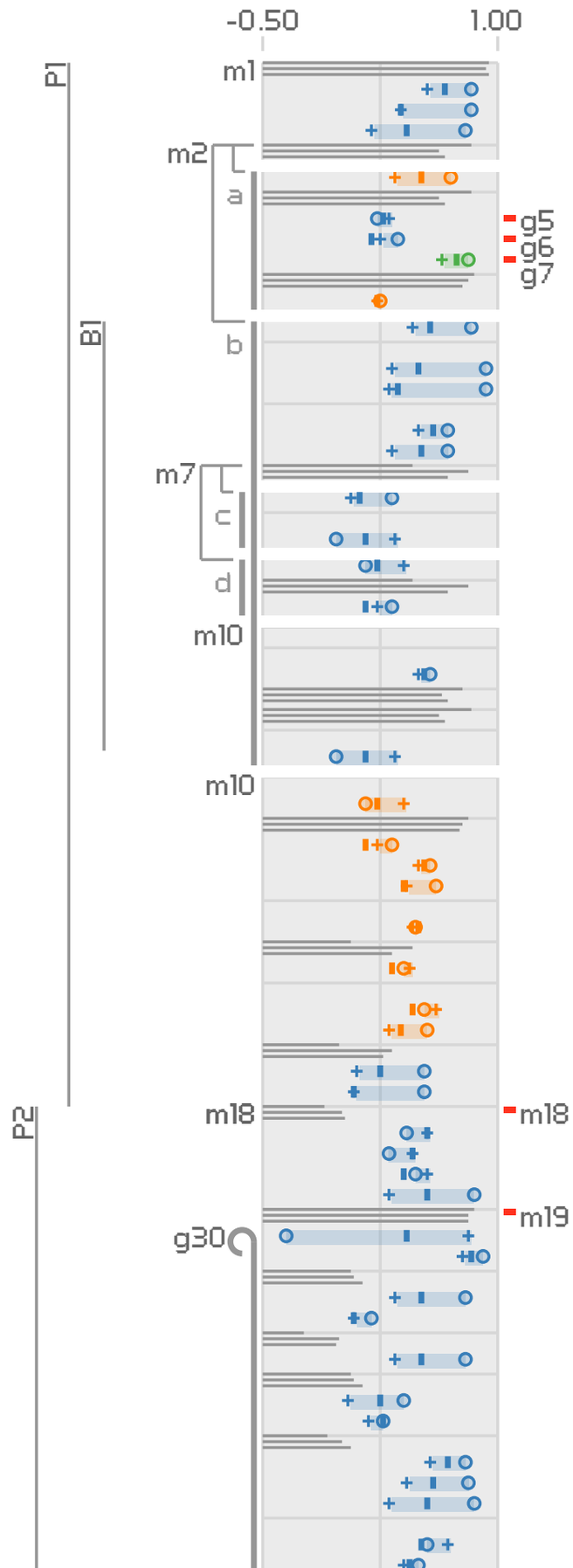

GLY Glu
0
1
a
b
c
d
PPP
TCA Ace
Cit
TRE G6P
Tre


GLY Glu
0
1
PPP
TCA Ace
Cit
TRE G6P
Tre


LY Glu
0
1
PPP
CA Ace
Cit
RE G6P
Tre


GLY Glu
0
1
a
b
c
d
PPP
TCA Ace
Cit
TRE G6P
Tre

PATHWAY METRIC OVERVIEW



Pathline

curvemap

STARTING POINT

species

time

genes

scer_LAG scer_LL scer_DS scer_PS scer_PLAT
spar_LAG spar_LL spar_DS spar_PS spar_PLAT
smik_LAG smik_LL smik_DS smik_PS smik_PLAT
sbay_LAG sbay_LL sbay_DS sbay_PS sbay_PLAT
sbayuv_LAG sbayuv_LL sbayuv_DS sbayuv_PS sbayuv_PLAT
cgla_ELL cgla_LL cgla_DS cgla_PS cgla_PLAT
scas_LAG scas_LL scas_DS scas_PS scas_PLAT
kpol_ELL kpol_LL kpol_DS kpol_PS kpol_PLAT
klac_LAG klac_LL klac_DS klac_PS klac_PLAT
kwal_LAG kwal_LL kwal_DS kwal_PS kwal_PLAT
sklu_LAG sklu_LL sklu_DS sklu_PS sklu_PLAT
dhan_LAG dhan_LL dhan_DS dhan_PS dhan_PLAT
calb_LAG calb_LL calb_DS calb_PS calb_PLAT
ylip_LAG ylip_LL ylip_DS ylip_PS ylip_PLAT
sjap_LAG sjap_LL sjap_DS sjap_PS sjap_PLAT
spom_LAG spom_LL spom_DS spom_PS spom_PLAT

ALD5 - "Mitochond
ALD6 - "Cytosolic
PDC6 - "Minor iso
HXK1 - "Hexokinas
GLK1 - "Glucokina
ALD4 - "Mitochond
ADH2 - "Glucose-r
ADH5 - Alcohol de
PDC6 - "Minor iso
PDC5 - "Minor iso
PDC1 - "Major of
HXK2 - Hexokinase
PFK1 - "Alpha sub
PFK1 - "Alpha sub
PFK2 - "Beta subu
PYK2 - "Pyruvate
ENO1 - "Enolase I
CDC19 - "Pyruvate
PGI1 - "Glycolyti
GPM1 - "Tetrameri
PGK1 - "3-phospho
PGK1 - "3-phospho
ENO2 - "Enolase I
TPI1 - "Triose ph
FBA1 - "Fructose
TDH3 - "Glyceral
ADH1 - "Alcohol d
TDH3 - "Glyceral
TDH3 - "Glyceral
TDH3 - "Glyceral
TDH2 - "Glyceral
TDH1 - "Glyceral
ADH3 - Mitochondr
ADH4 - "Alcohol d

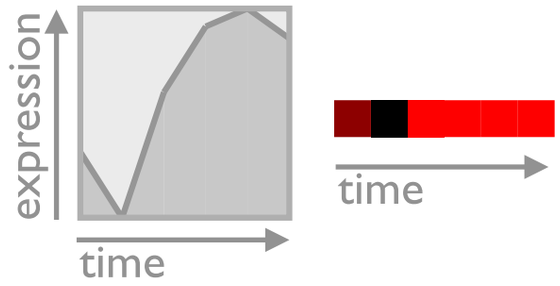
curvemap

inspired by heatmaps

curvemap

inspired by heatmaps

- base visual unit is a curve

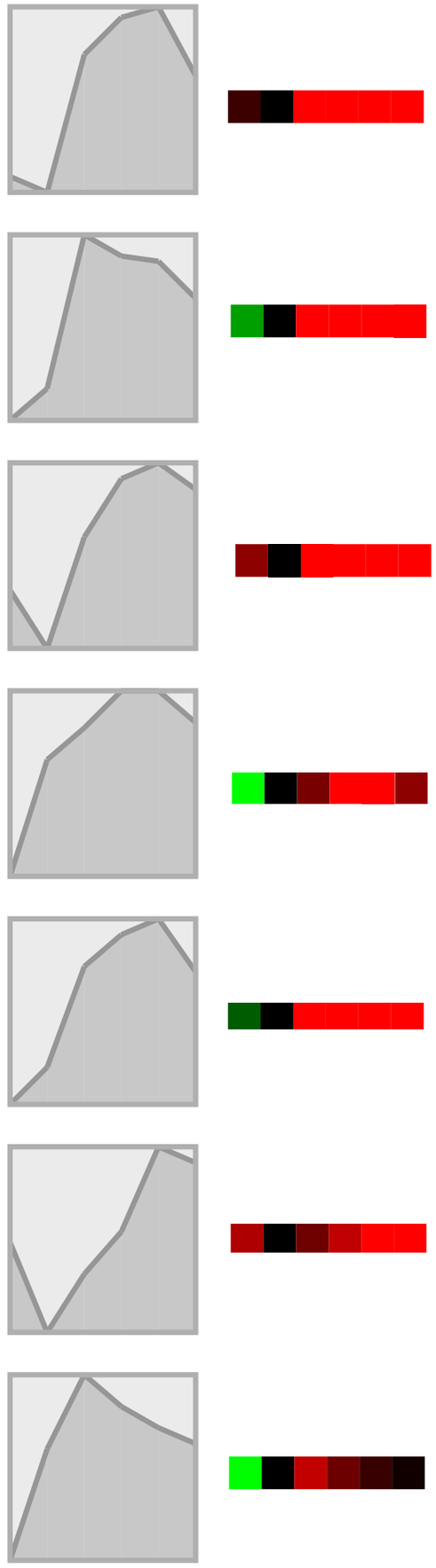


curvemap

inspired by heatmaps

- base visual unit is a curve

- filled, framed line charts to enhance shape perception

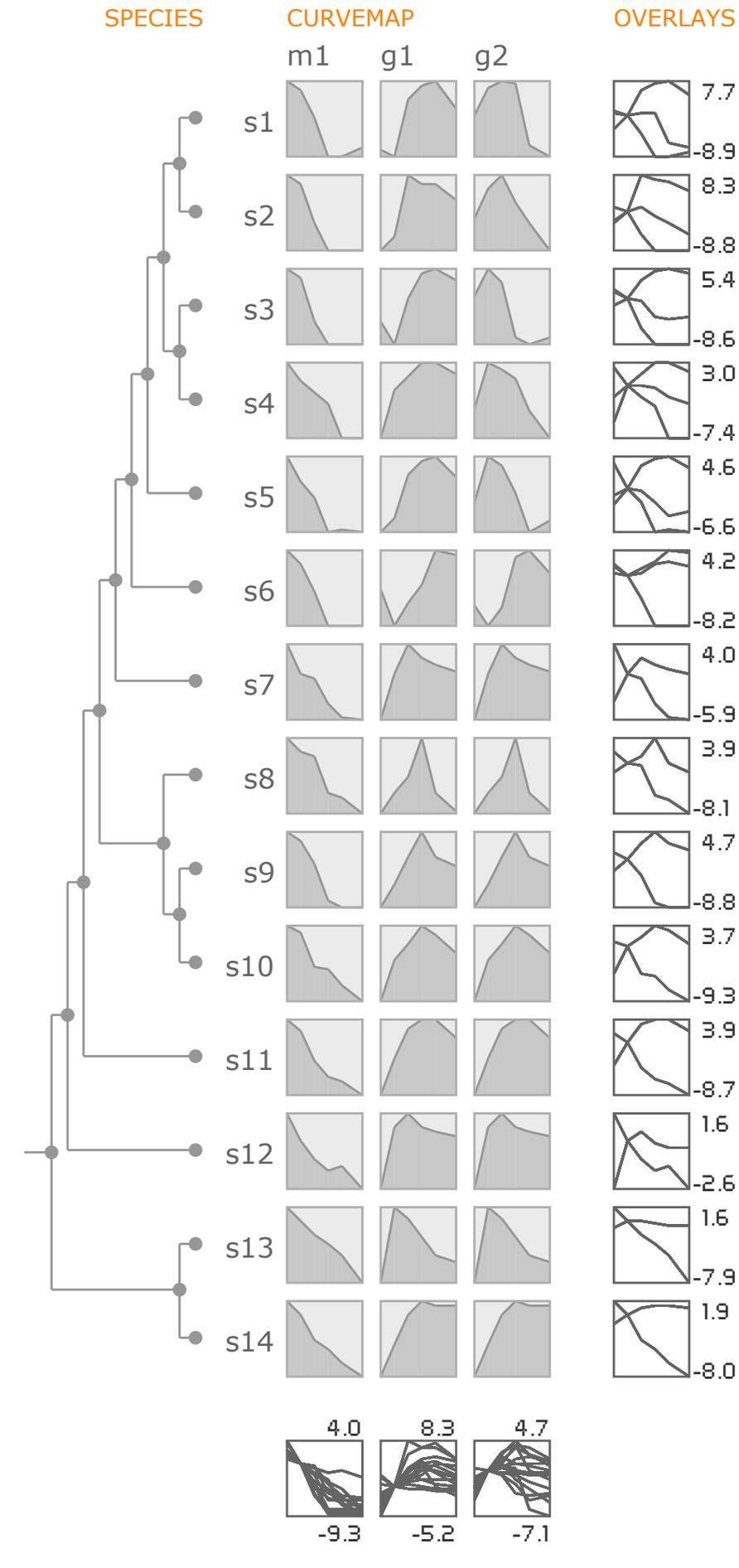


curvemap

inspired by heatmaps

- base visual unit is a curve

- filled, framed line charts to enhance shape perception



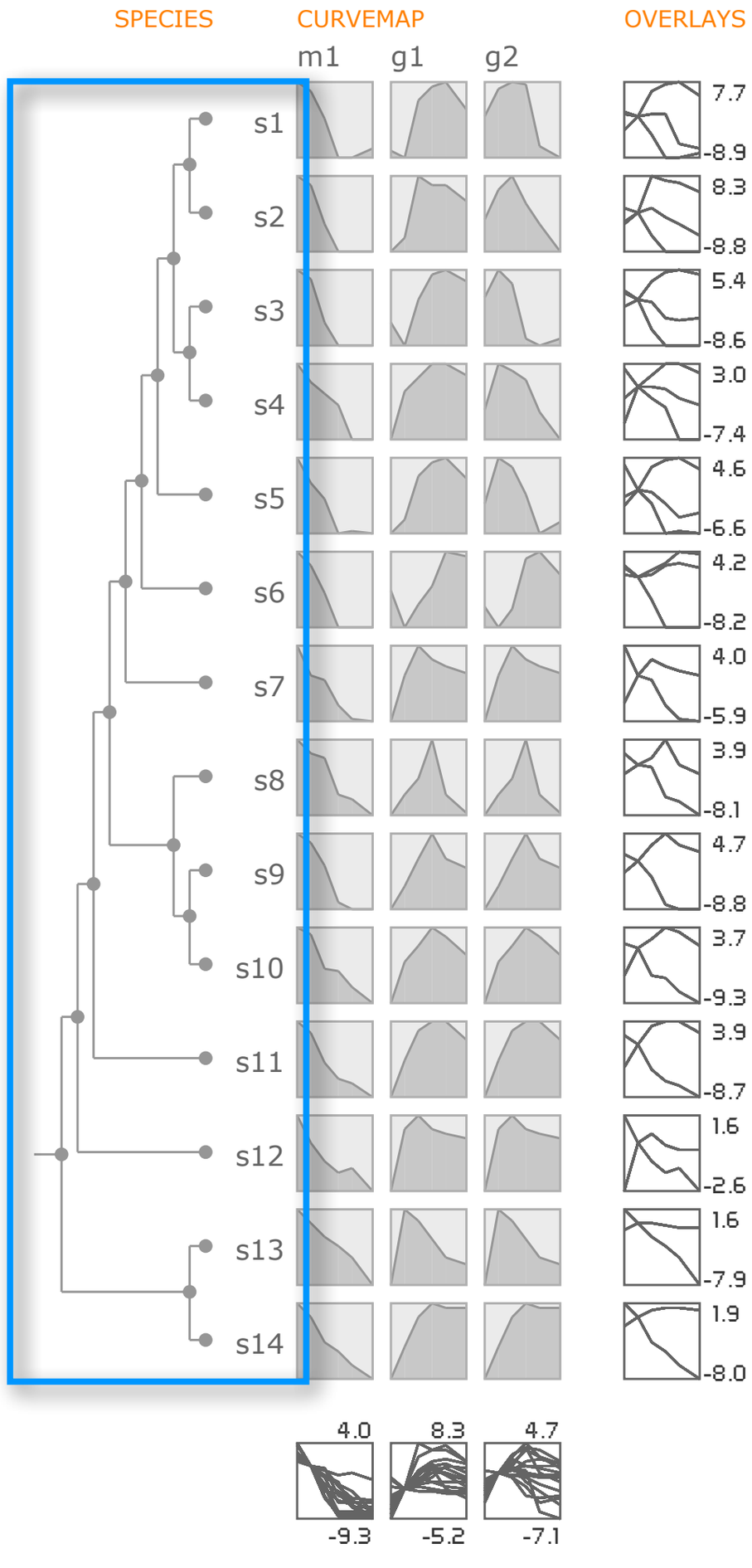
curvemap

inspired by heatmaps

- base visual unit is a curve

- filled, framed line charts to enhance shape perception

- rows are species



curvemap

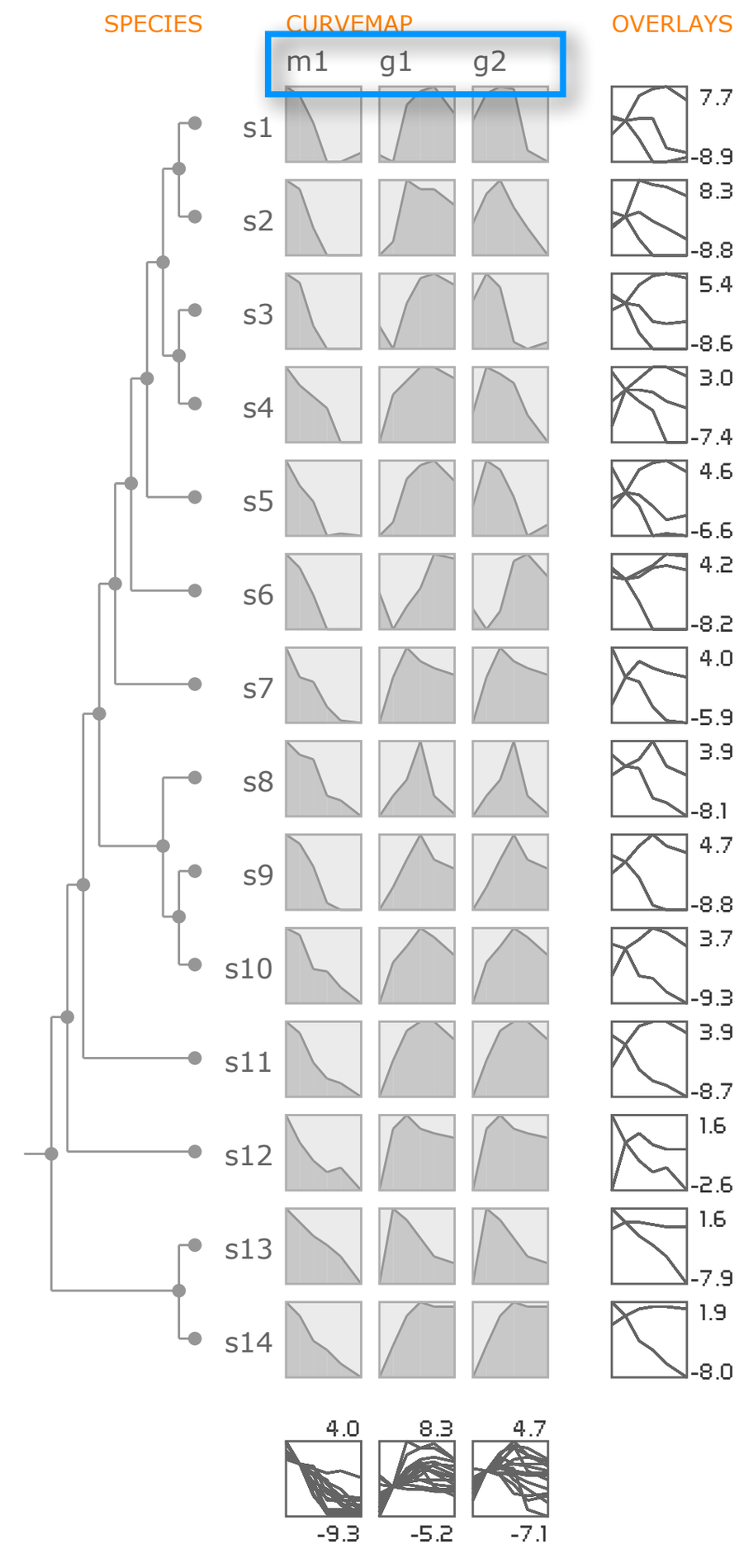
inspired by heatmaps

- base visual unit is a curve

- filled, framed line charts to enhance shape perception

- rows are species

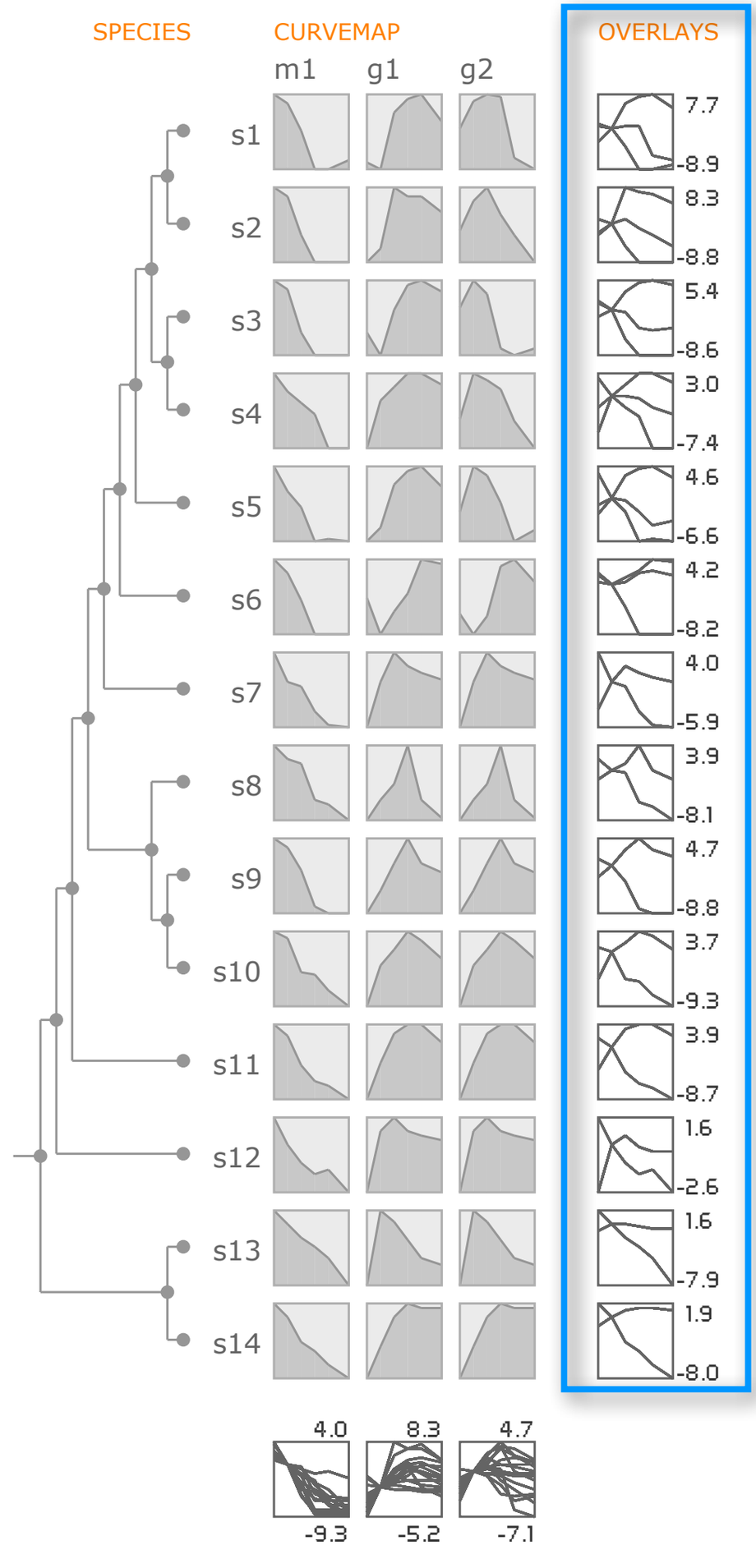
- columns are genes



curvemap

inspired by heatmaps

- base visual unit is a curve
- filled, framed line charts to enhance shape perception
- rows are species
- columns are genes
- overlays to enhance trends



curvemap

inspired by heatmaps

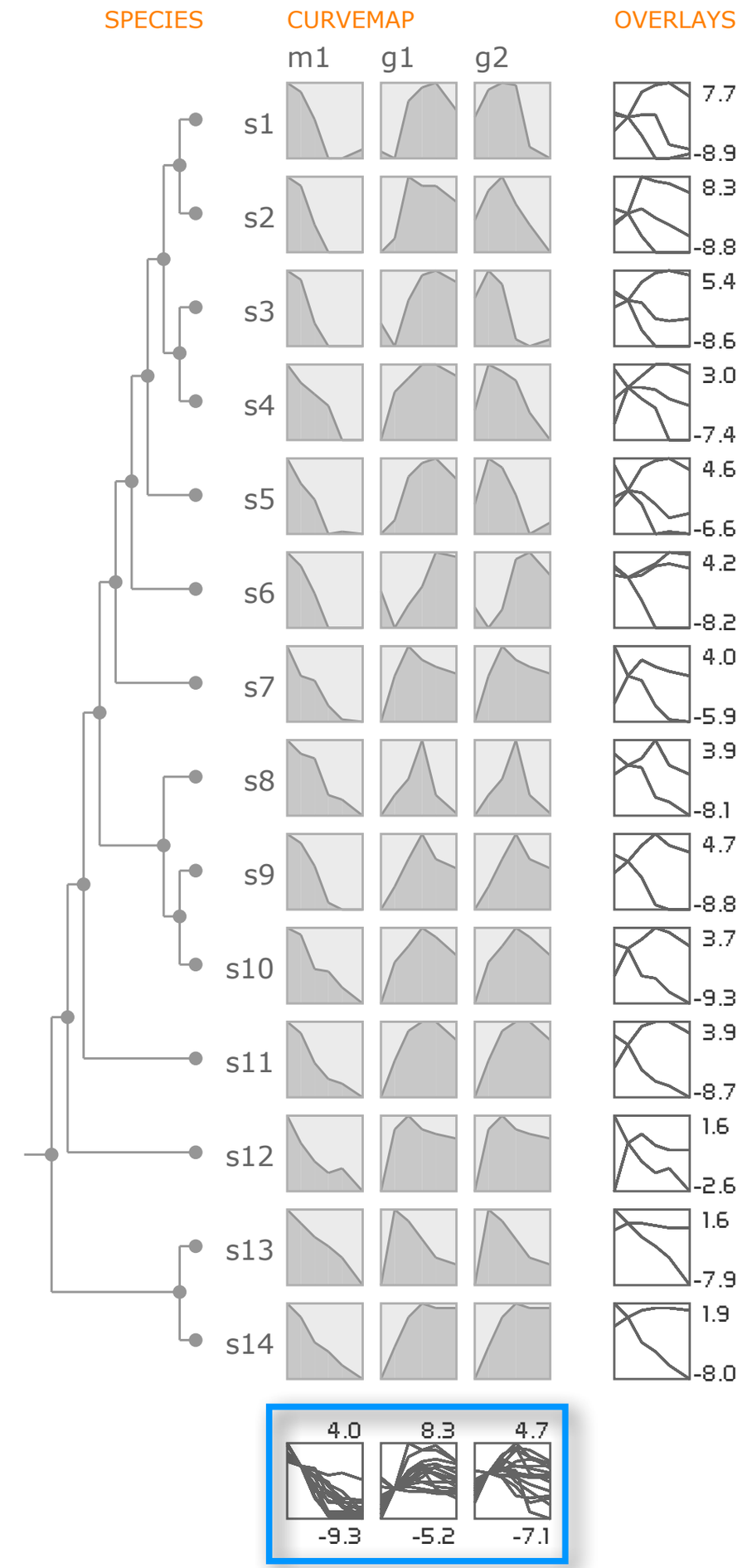
- base visual unit is a curve

- filled, framed line charts to enhance shape perception

- rows are species

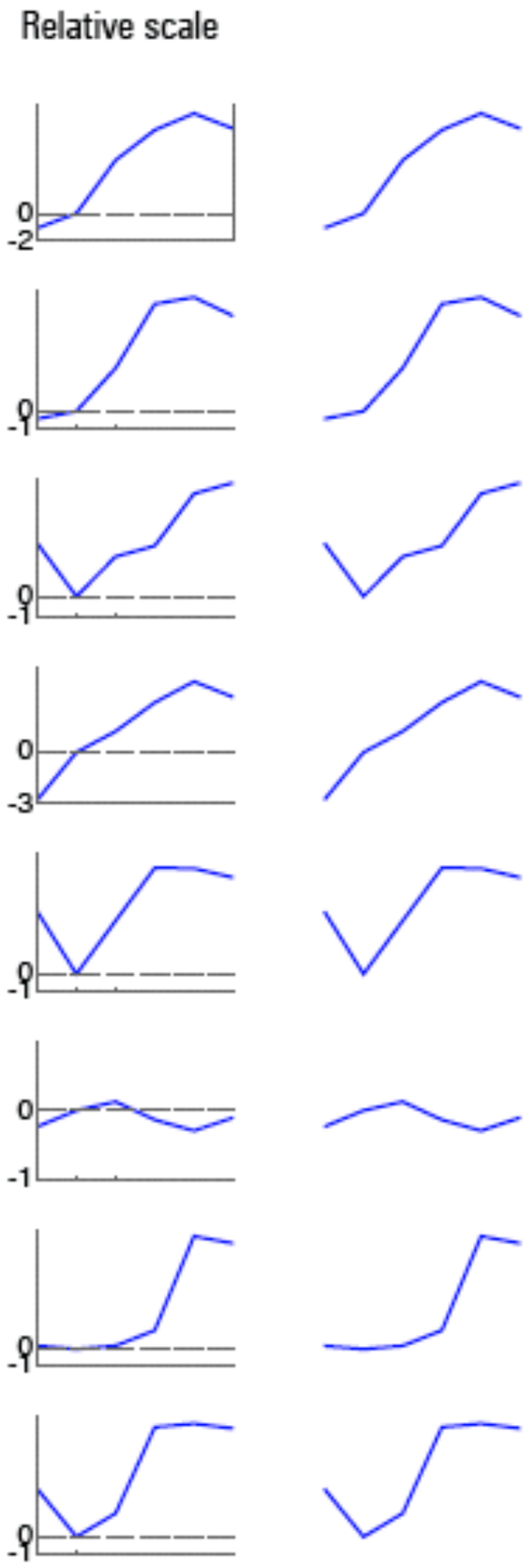
- columns are genes

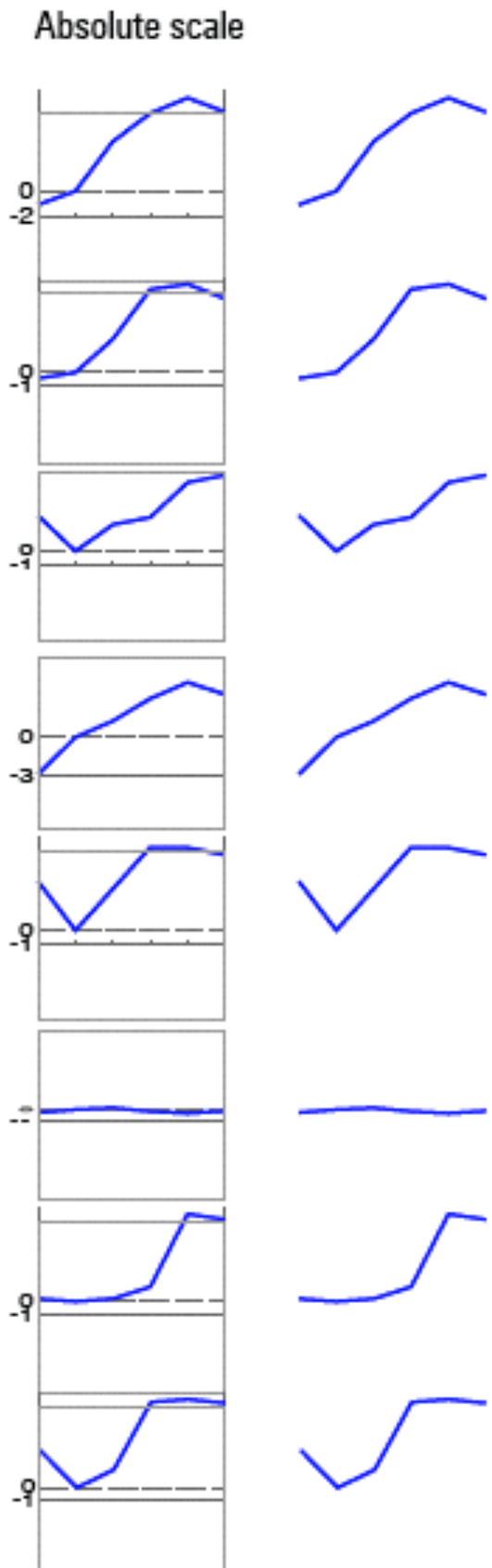
- overlays to enhance trends

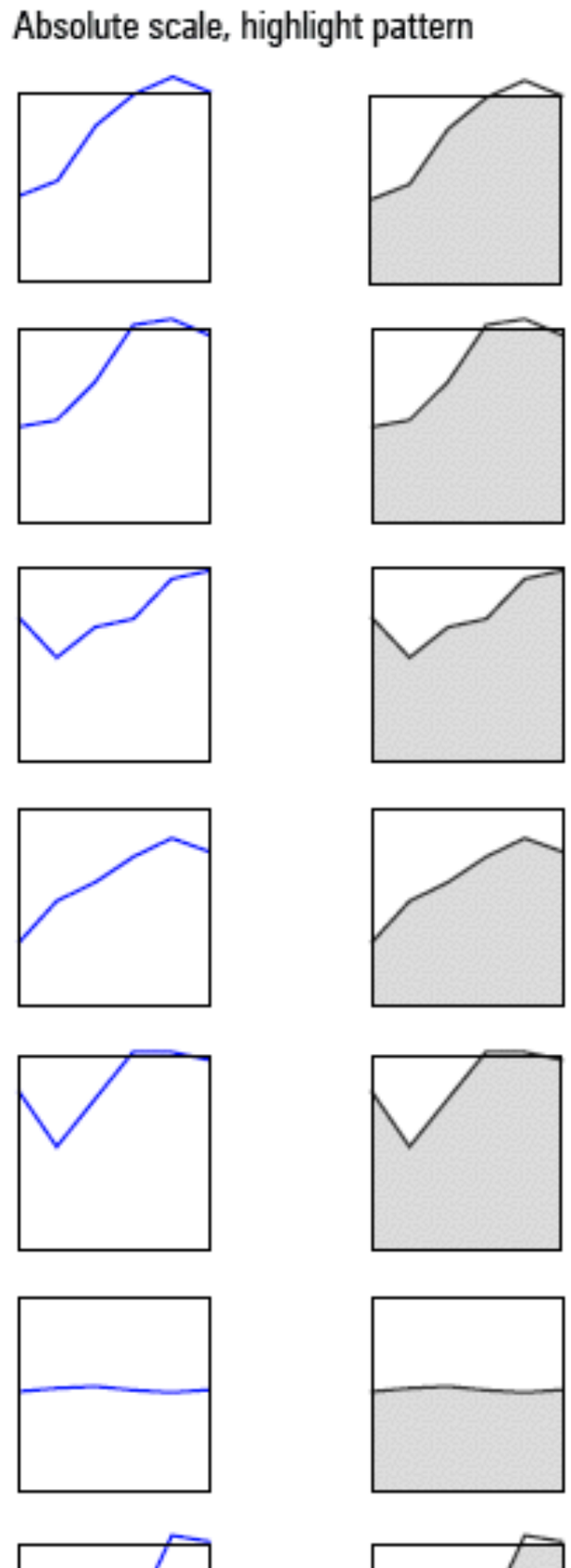


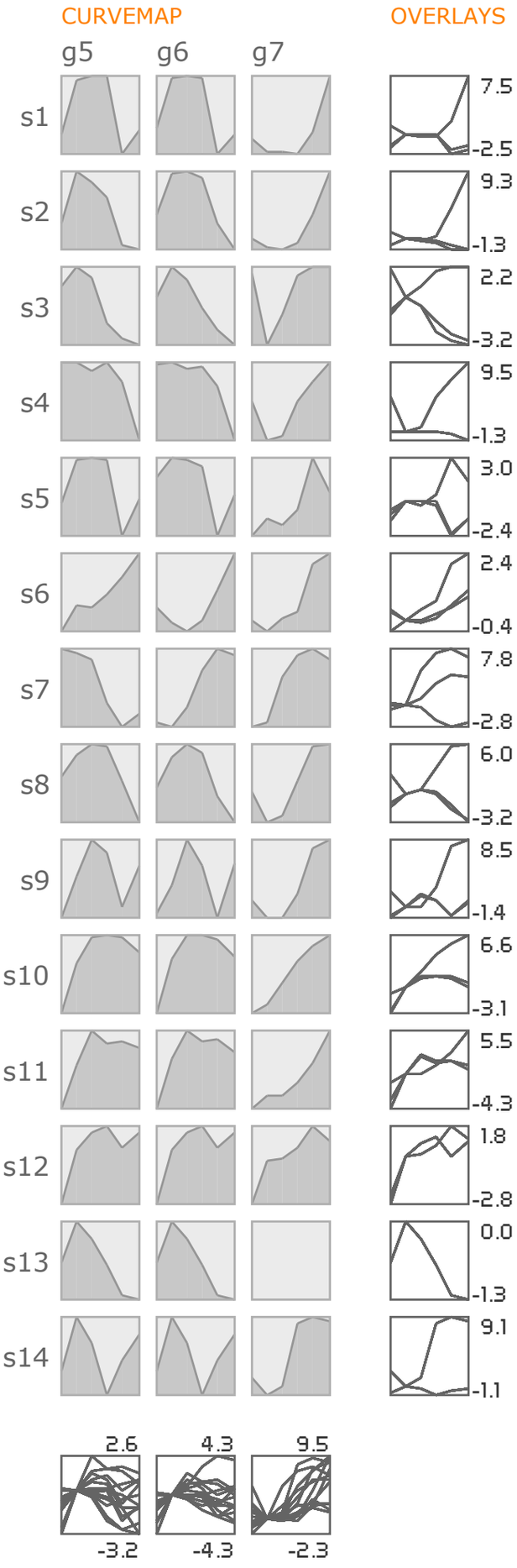
PAPER PROTOTYPES

time series

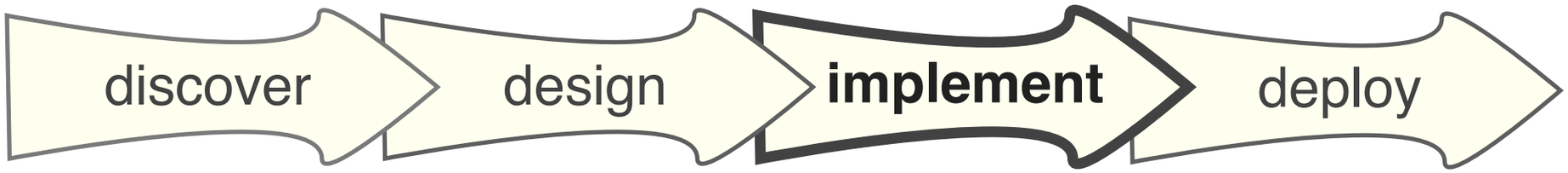

Relative scale


Absolute scale


Absolute scale, highlight pattern


CURVEMAP
OVERLAYS
g5
g6
g7
s1
s2
s3
s4
s5
s6
s7
s8
s9
s10
s11
s12
s13
s14
7.5
-2.5
9.3
-1.3
2.2
-3.2
9.5
-1.3
3.0
-2.4
2.4
-0.4
7.8
-2.8
6.0
-3.2
8.5
-1.4
6.6
-3.1
5.5
-4.3
1.8
-2.8
0.0
-1.3
9.1
-1.1
2.6
-3.2
4.3
-4.3
9.5
-2.3

rapid software prototyping

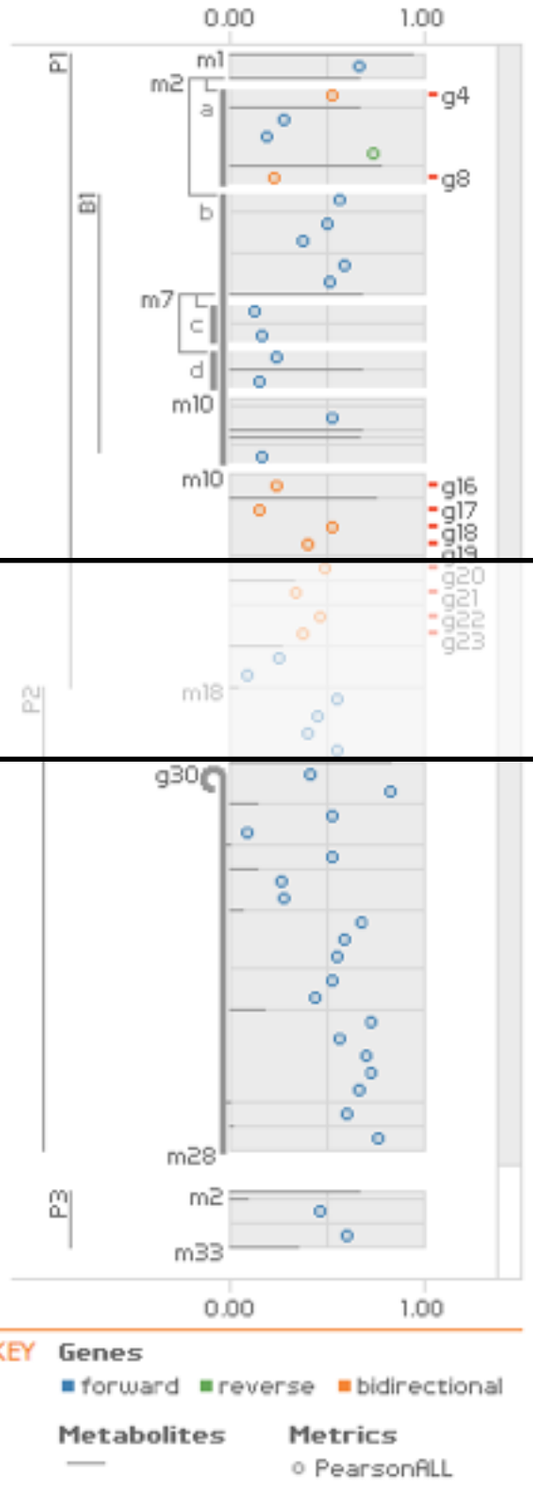


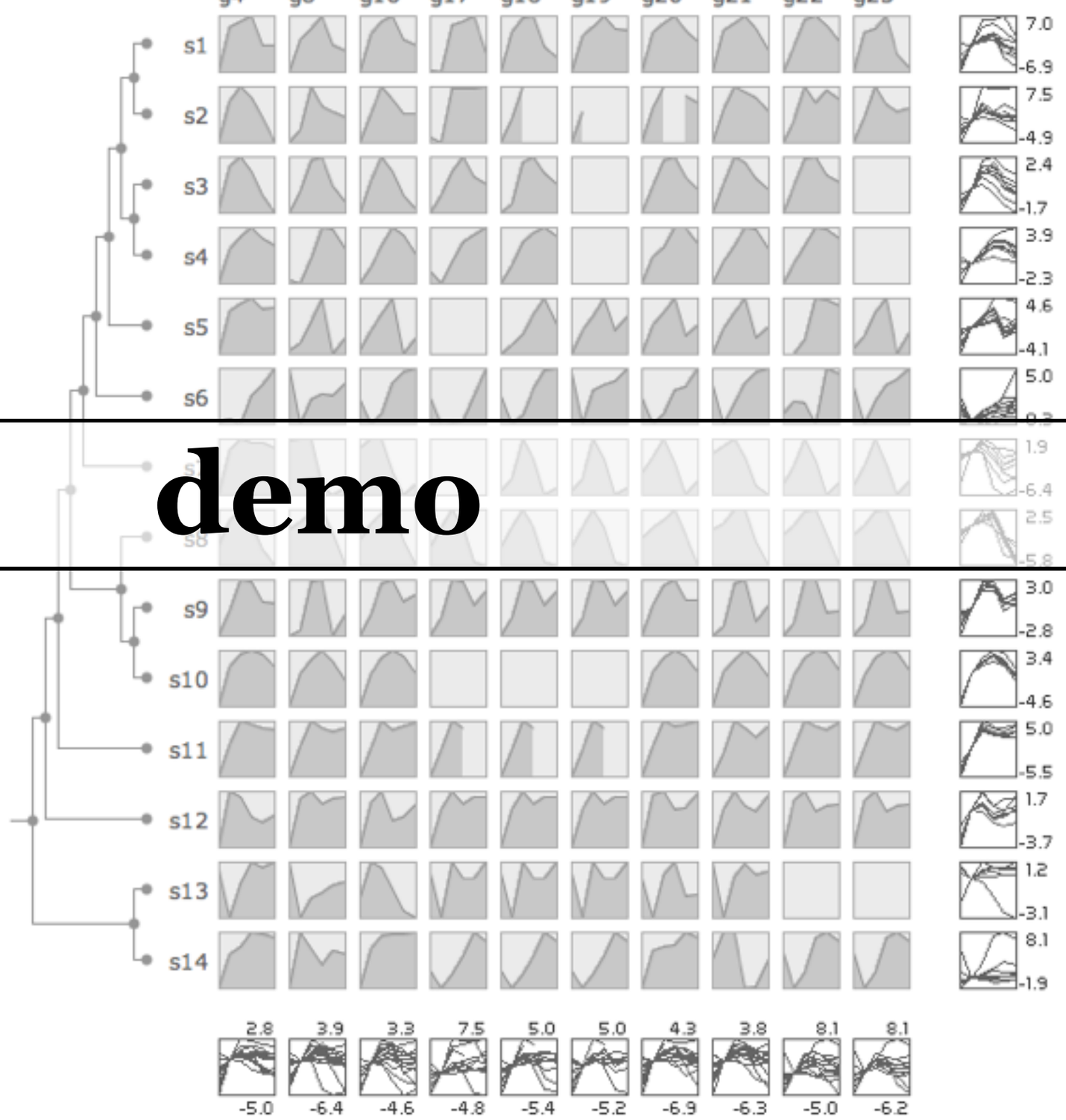
CORE

Pathline

PATHLINE

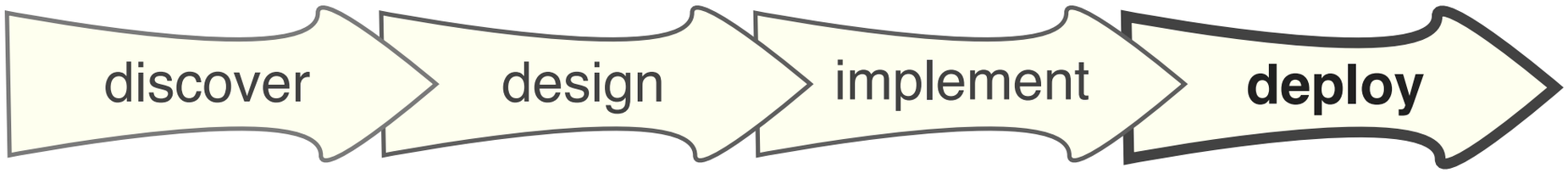
A TOOL FOR COMPARATIVE FUNCTIONAL GENOMICS





demo

release & gather feedback “in the wild”

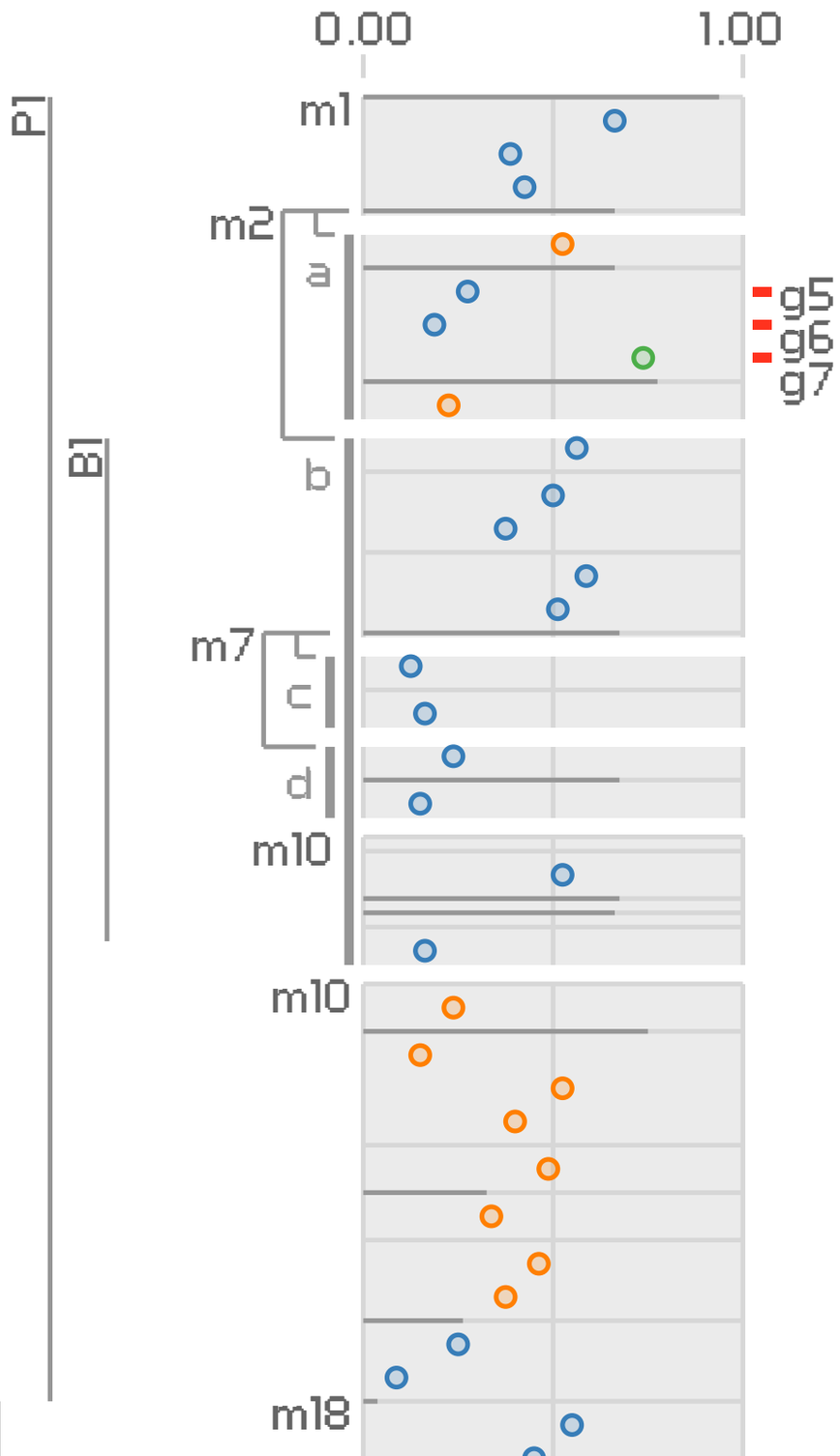


CORE

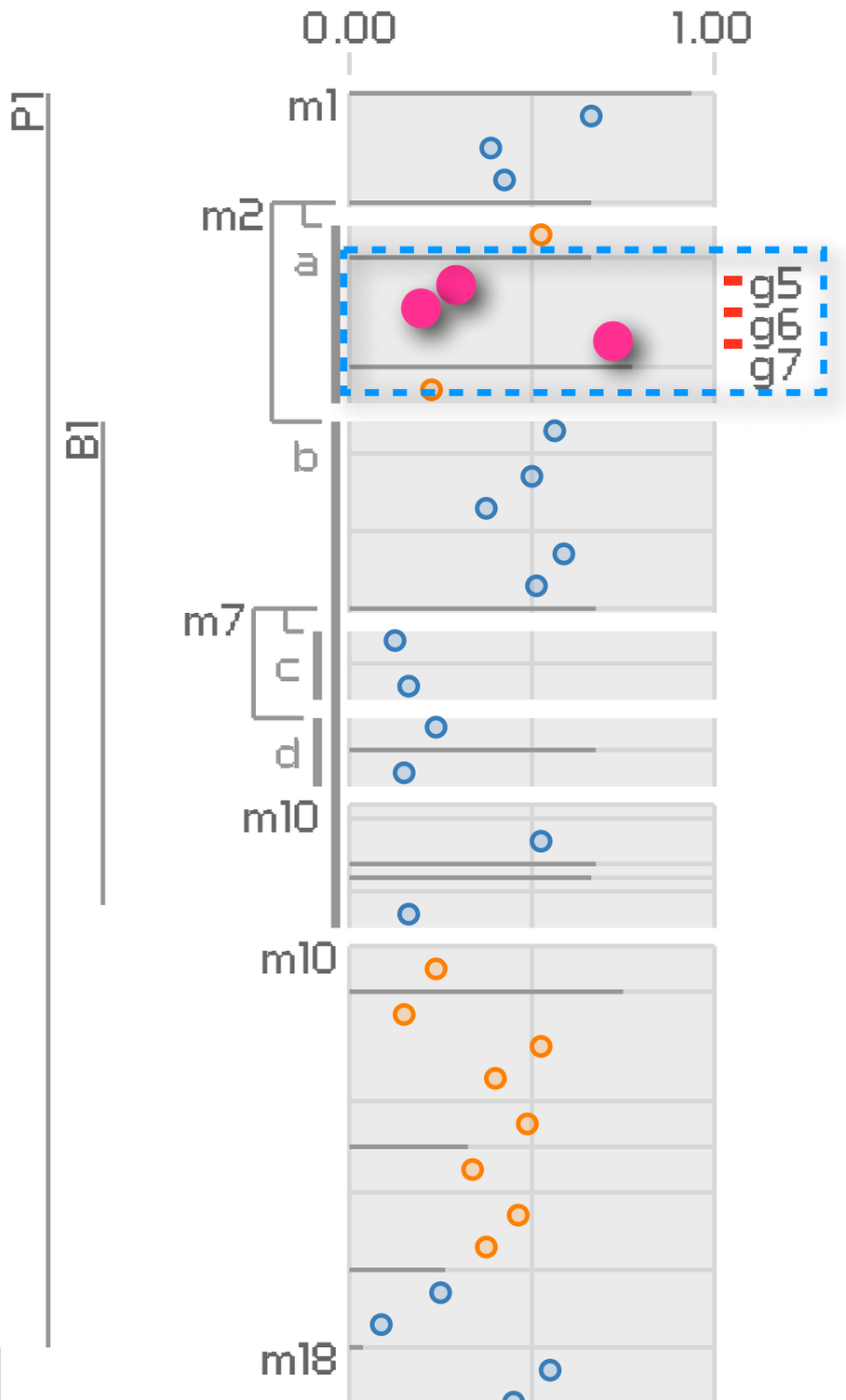
case study

- qualitative research method
- in-depth study of individual or group
- real-world setting
- description and interpretation

PATHWAY METRIC OVERVIEW

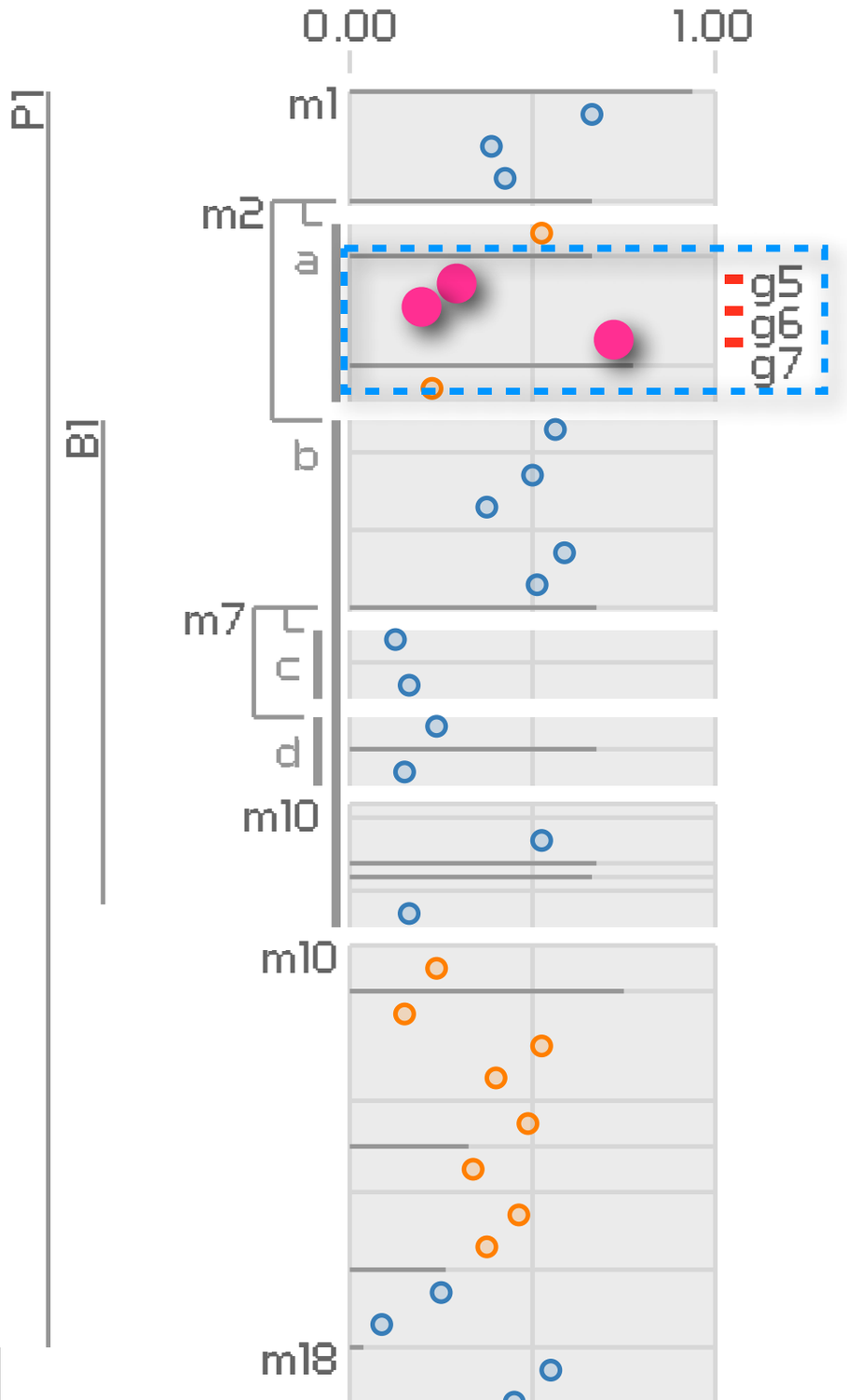


PATHWAY METRIC OVERVIEW



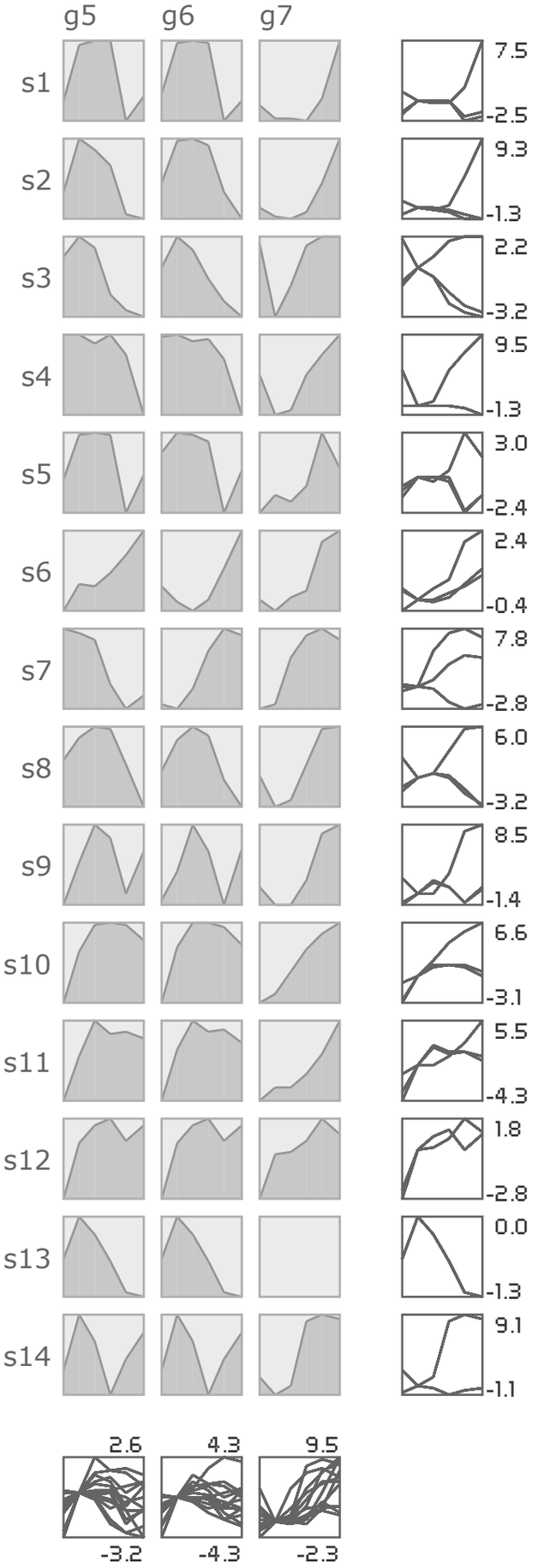
PATHWAY

METRIC OVERVIEW



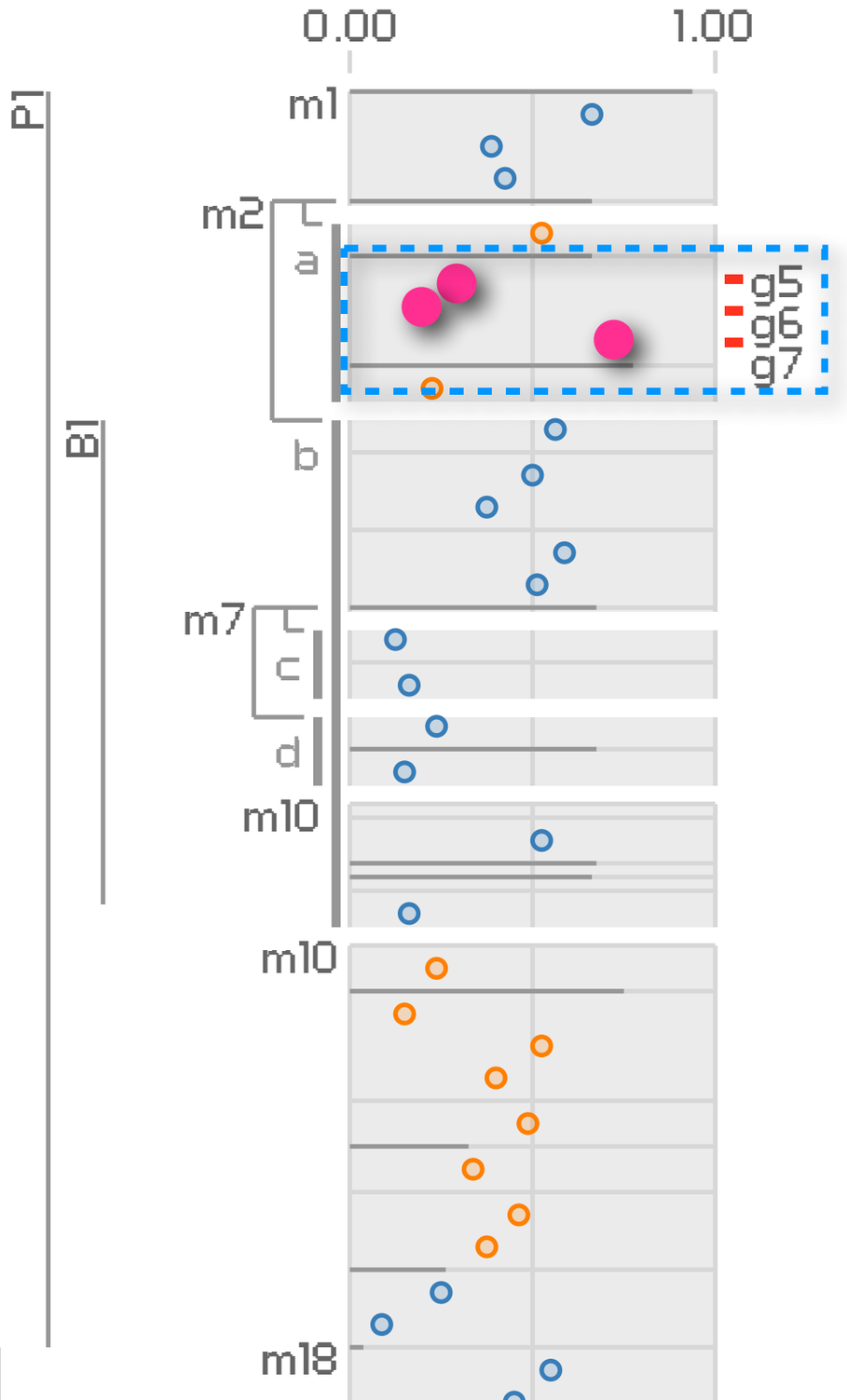
CURVEMAP

OVERLAYS



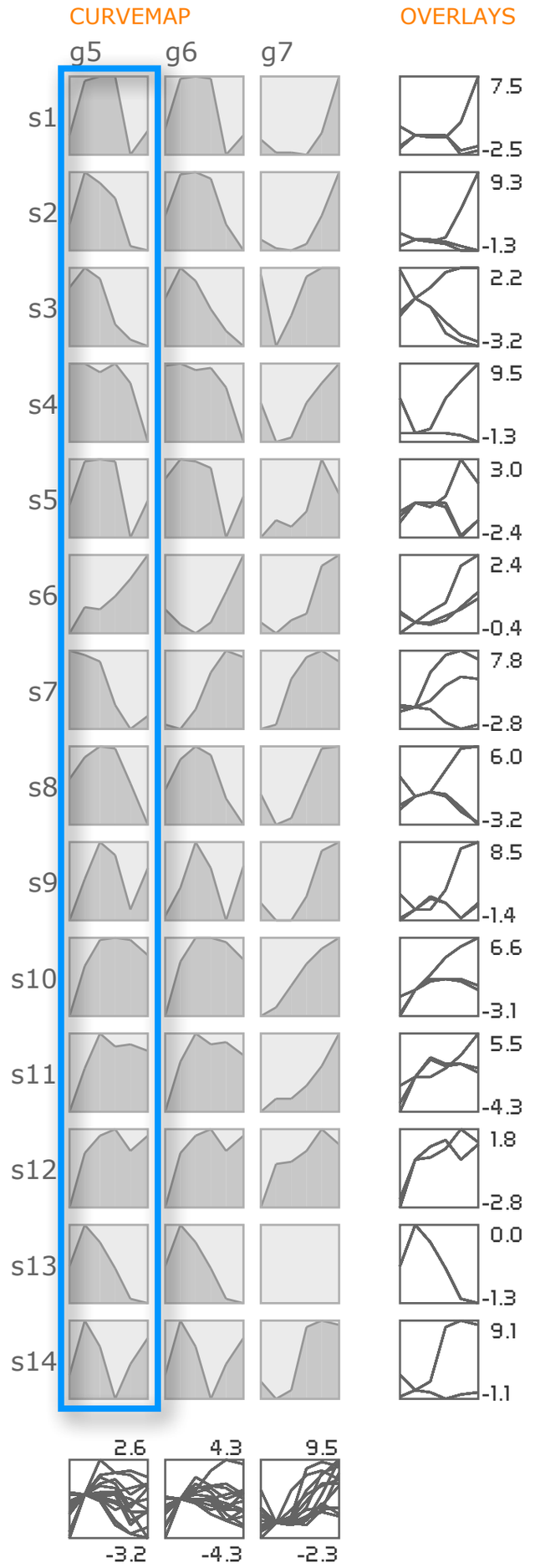
PATHWAY

METRIC OVERVIEW



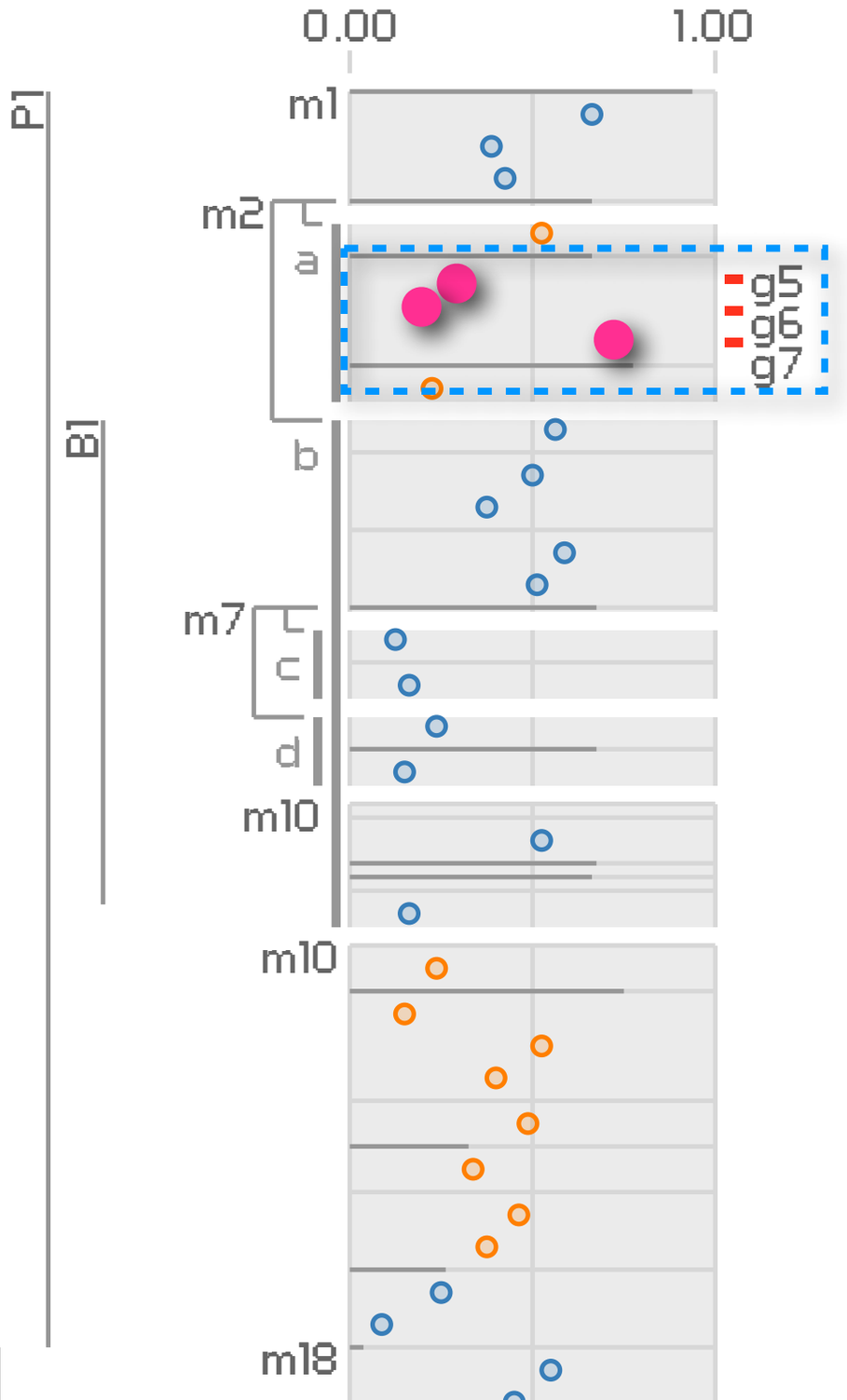
CURVEMAP

OVERLAYS



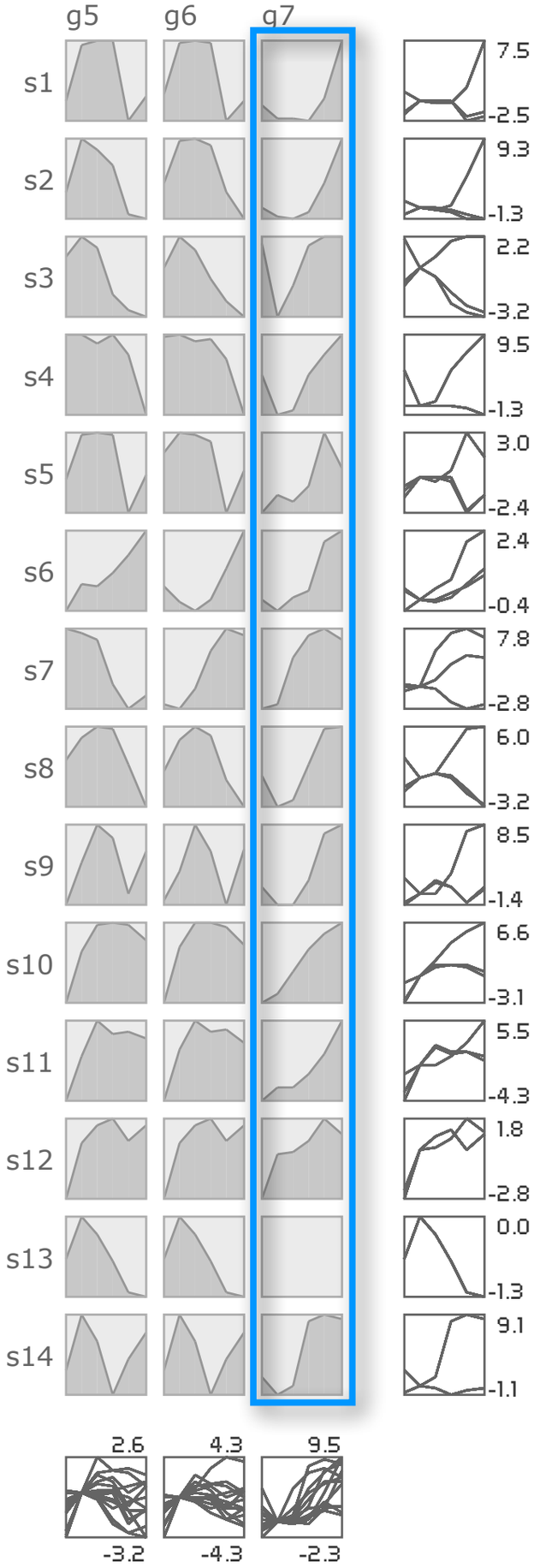
PATHWAY

METRIC OVERVIEW



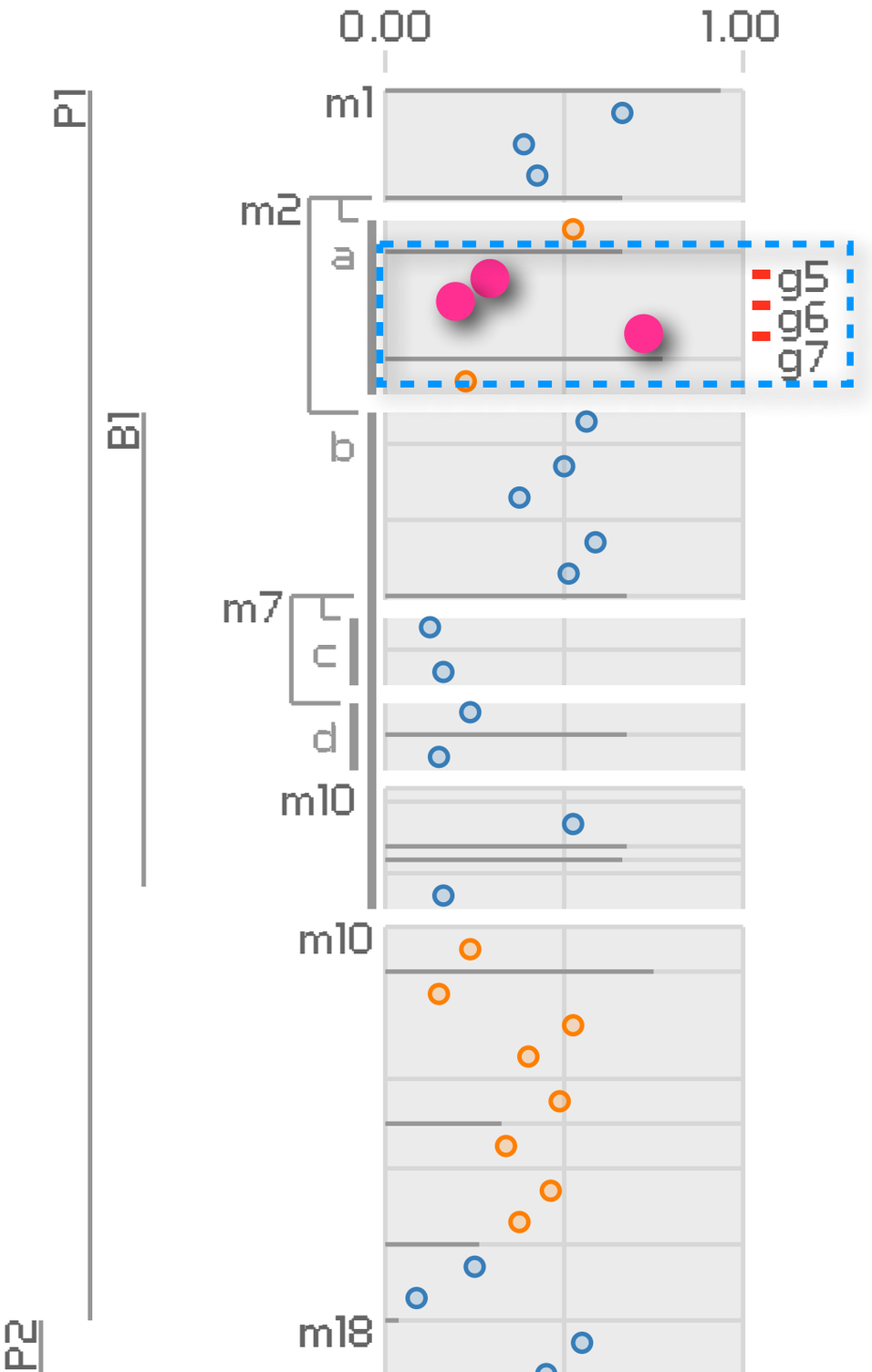
CURVEMAP

OVERLAYS



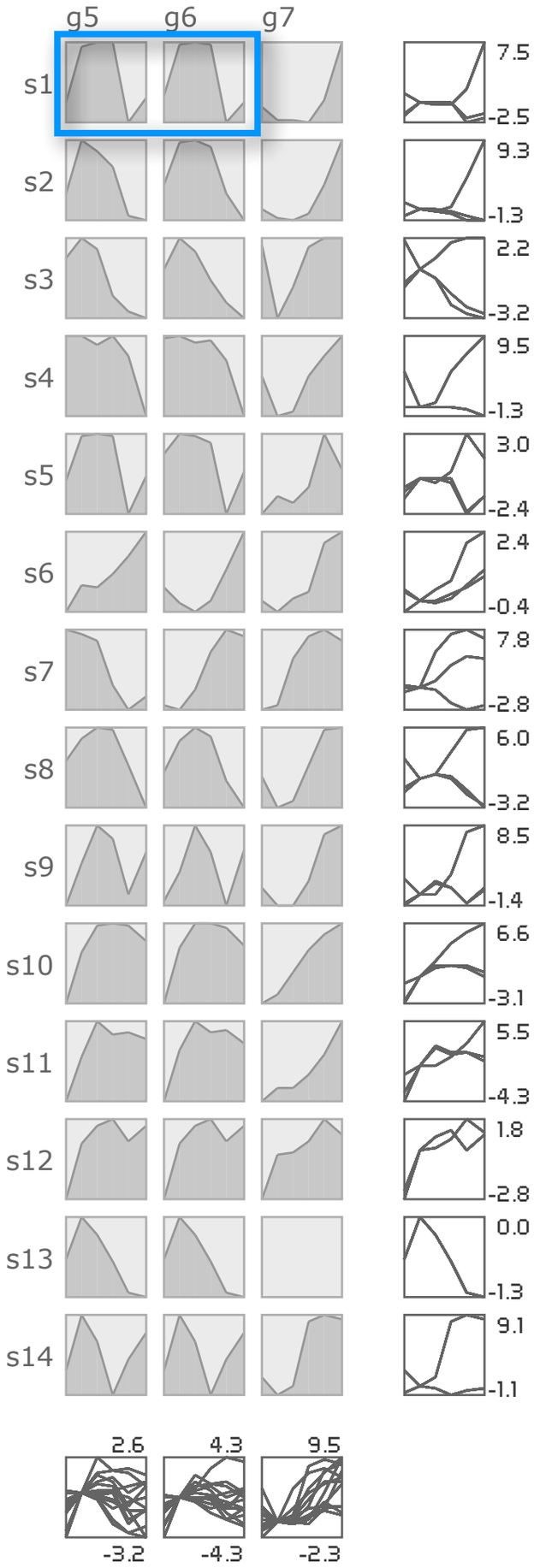
PATHWAY

METRIC OVERVIEW



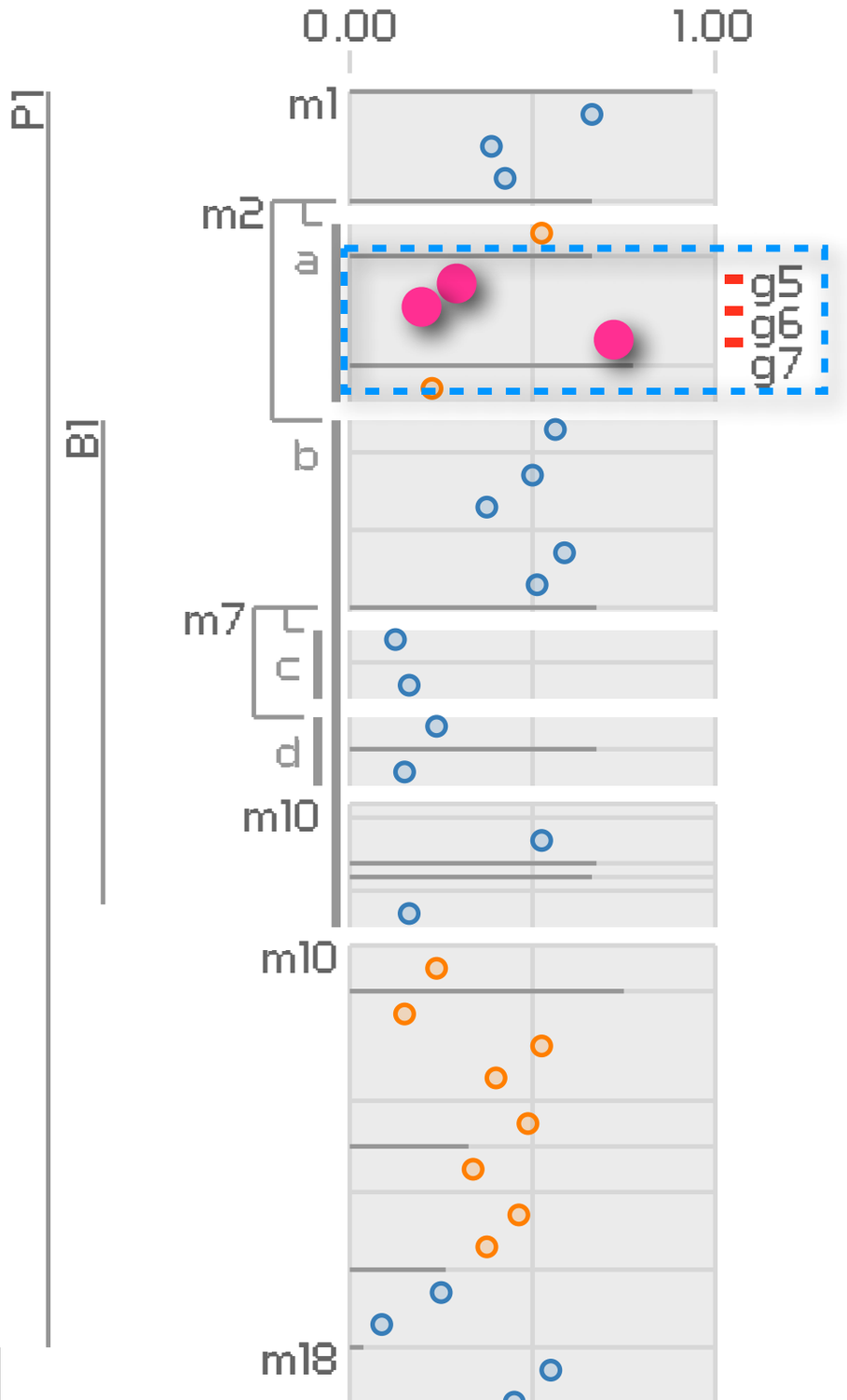
CURVEMAP

OVERLAYS



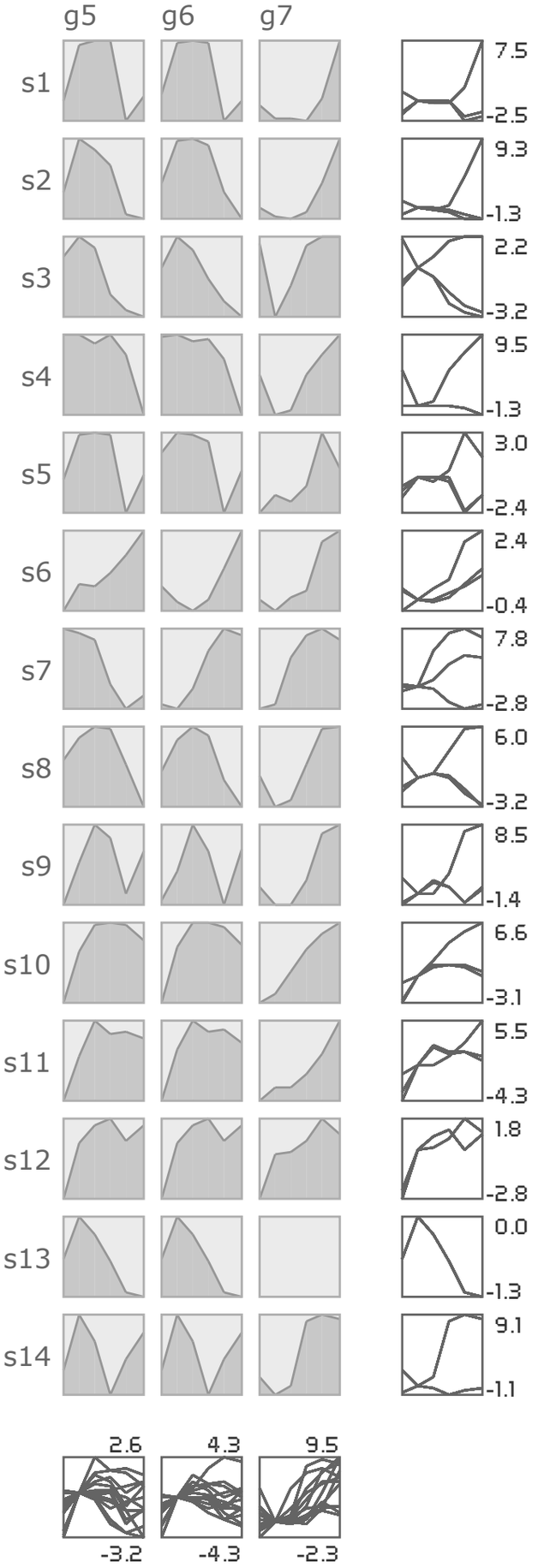
PATHWAY

METRIC OVERVIEW



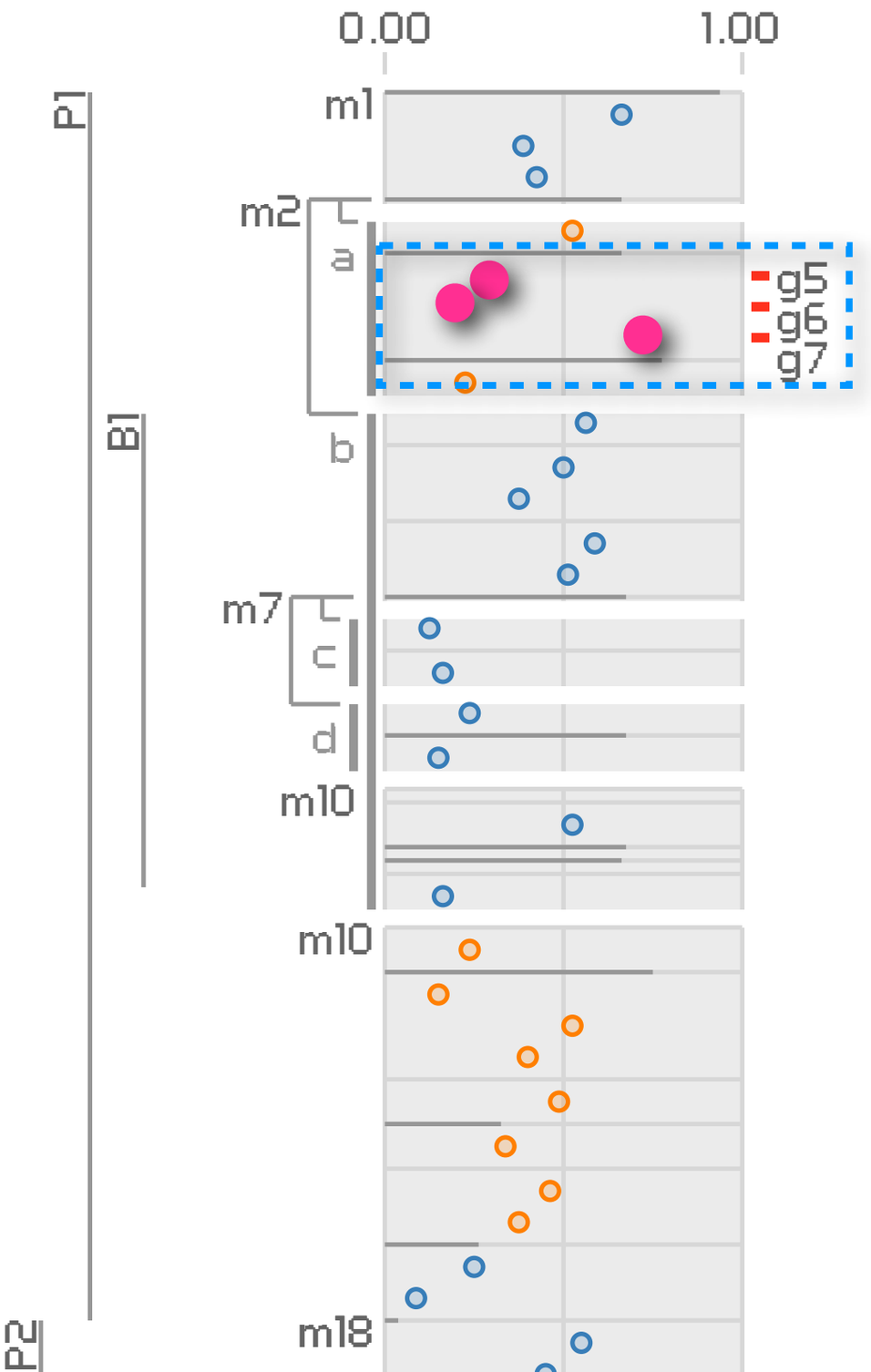
CURVEMAP

OVERLAYS



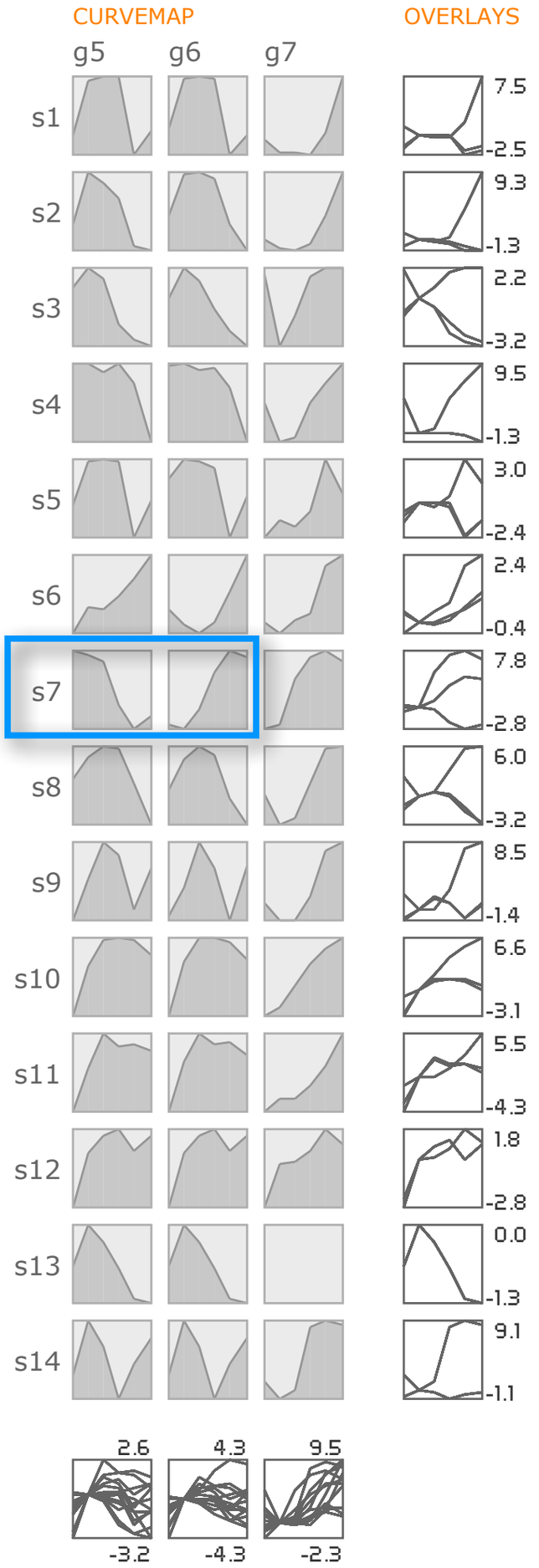
PATHWAY

METRIC OVERVIEW



CURVEMAP

OVERLAYS



MizBee: A Multiscale Synteny Browser

Miriah Meyer, Tamara Munzner, *Member, IEEE*, and Hanspeter Pfister, *Senior Member, IEEE*

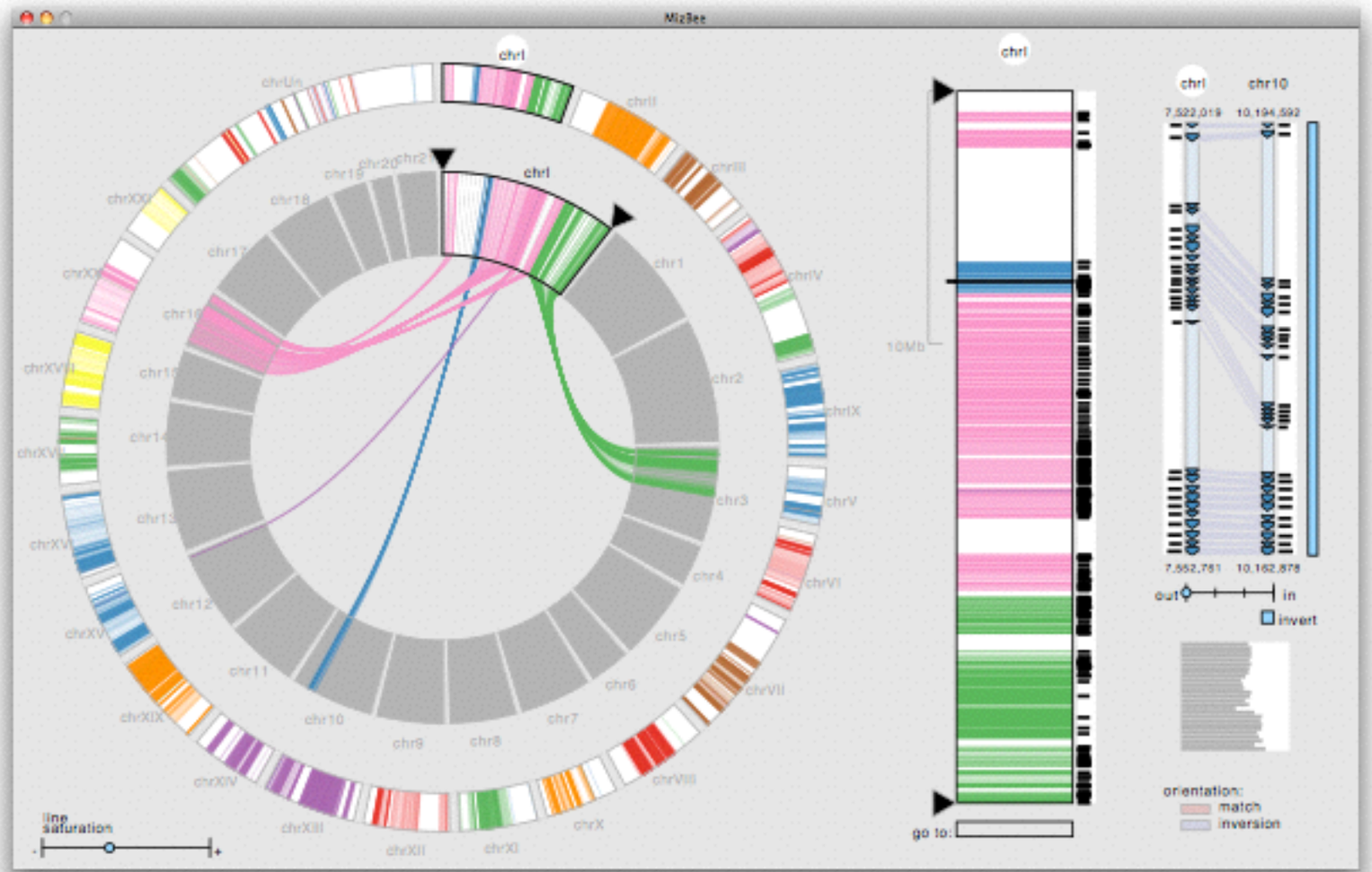


Fig. 1. The multiscale MizBee browser allows biologists to explore many kinds of conserved synteny relationships with linked views at the genome, chromosome, and block levels. Here we compare the genomes of two fish, the stickleback and the pufferfish.

Abstract—In the field of comparative genomics, scientists seek to answer questions about evolution and genomic function by comparing the genomes of species to find regions of shared sequences. Conserved syntenic blocks are an important biological data abstraction for indicating regions of shared sequences. The goal of this work is to show multiple types of relationships at multiple scales in a way that is visually comprehensible in accordance with known perceptual principles. We present a task analysis for this domain where the fundamental questions asked by biologists can be understood by a characterization of relationships into the four types of proximity/location, size, orientation, and similarity/strength, and the four scales of genome, chromosome, block, and genomic feature. We also propose a new taxonomy of the design space for visually encoding conservation data. We present MizBee, a multiscale synteny browser with the unique property of providing interactive side-by-side views of the data across the range of scales supporting exploration of all of these relationship types. We conclude with case studies from two biologists who used MizBee to augment their previous automatic analysis work flow, providing anecdotal evidence about the efficacy of the system for the visualization

- comparative genomics

- interviews with two biologists

- validate, analyze, and communicate computational results

biology concepts

biology concepts

- compare **genomes**

biology concepts

- compare **genomes**

- genomes made of **chromosomes**

biology concepts

- compare **genomes**

- genomes made of **chromosomes**

- contiguous features (genes) grouped into **blocks**

biology concepts

- compare **genomes**

- genomes made of **chromosomes**

- contiguous features (genes) grouped into **blocks**

- similar blocks on different chromosomes implies **conservation**

high level biology questions

low level data-centric questions

high level biology questions

evolution: How long ago did two species share a common ancestor?

function: Which segment of the genome is responsible for a specific function in the cell?

low level data-centric questions

high level biology questions

evolution: How long ago did two species share a common ancestor?

function: Which segment of the genome is responsible for a specific function in the cell?

low level data-centric questions

Are the paired features within a block contiguous?

Which chromosomes share conserved blocks?

Are similarity scores alike within a block?

high level biology questions

evolution: How long ago did two species share a common ancestor?

function: Which segment of the genome is responsible for a specific function in the cell?

low level data-centric questions

1. Are the paired features within a block contiguous?
2. Which chromosomes share conserved blocks?
3. Are similarity scores alike within a block?

...

14.

- **domain**
  - comparative genomics
- **data**
  - multiscale
    - *genome*
    - *chromosome*
    - *block*
    - *feature*
- **task**
  - syntenic relationship: features on the same chromosome
    - *proximity and location*
    - *size*
    - *orientation*
    - *similarity*

MizBee

source: Human
destination: Lizard

chrX chrY chr1 chr2 chr3 chr4 chr5 chr6 chr7 chr8 chr9 chr10 chr11 chr12 chr13 chr14 chr15 chr16 chr17 chr18 chr19 chr20 chr21 chr22

chra chrb chrc chrd chre chrf chrg chrh

chr3

chr3

10Mb

chr3 chr2

51280143 152008850

51709189 152345239

out in

invert

line saturation

\- +

go to:

orientation:

match

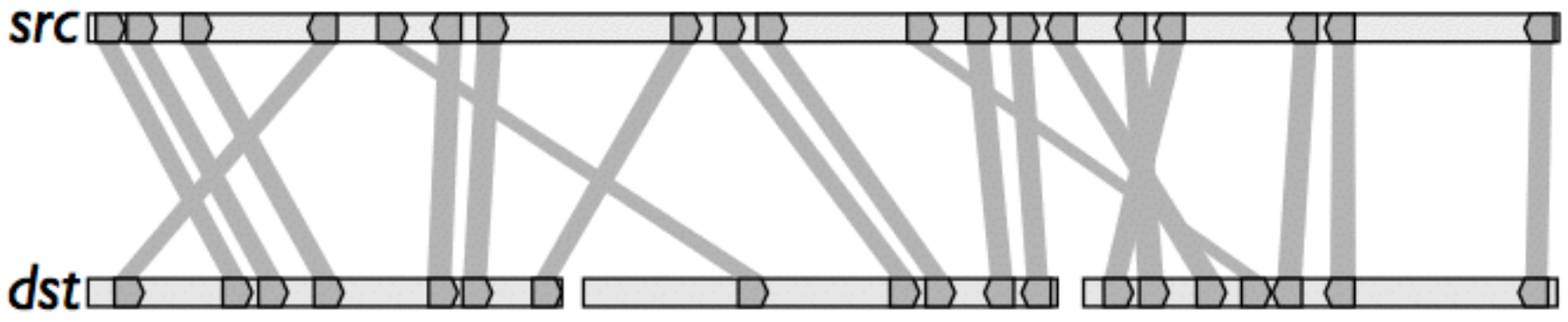
inversion

DEMO

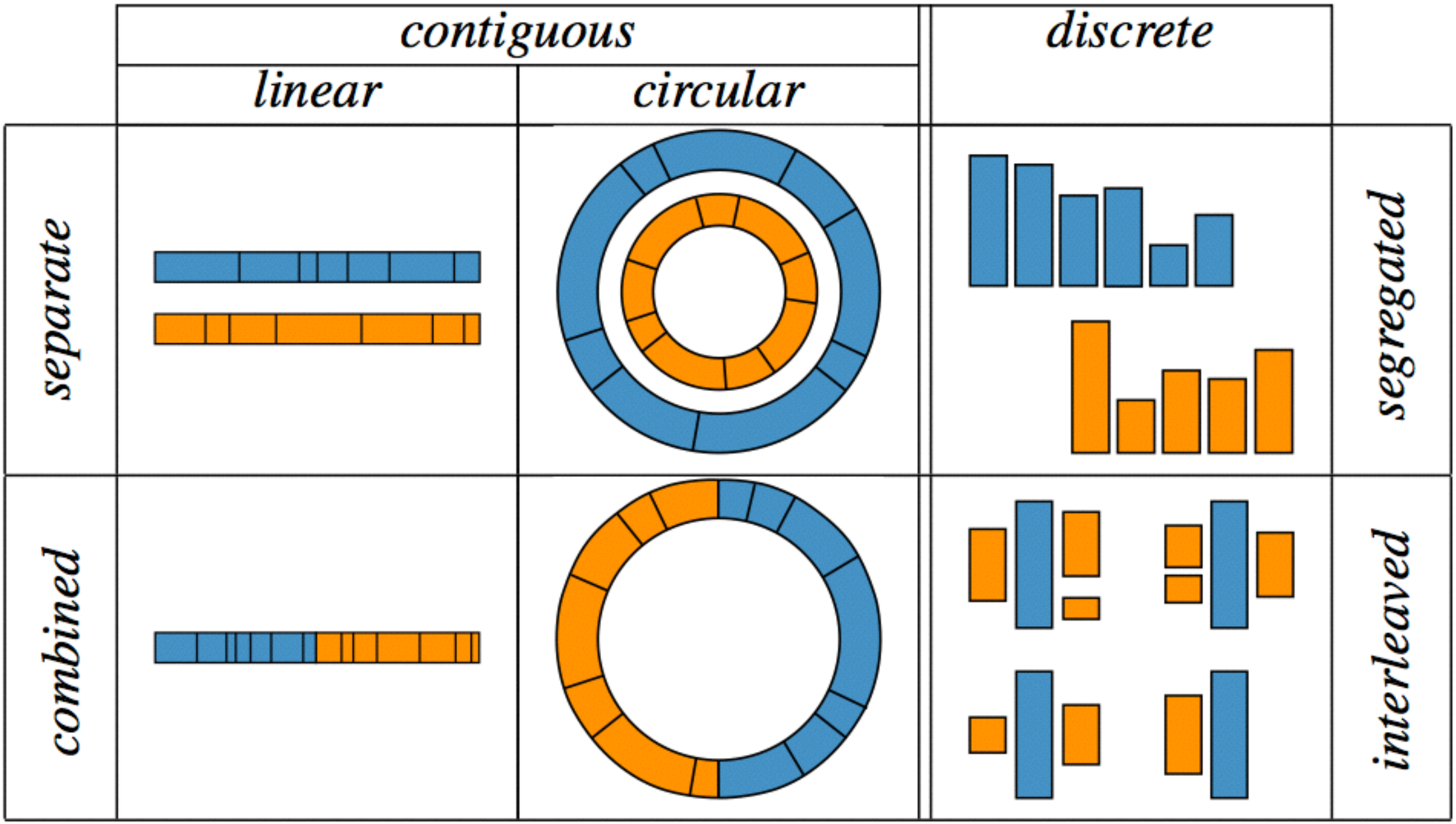
VISUAL ENCODING

- **color limits**
  - no info about destination
  - <12 distinguishable colors

- **connection limits**
  - visual clutter



TAXONOMY


contiguous
discrete
linear
circular
separate
segregated
combined
interleaved

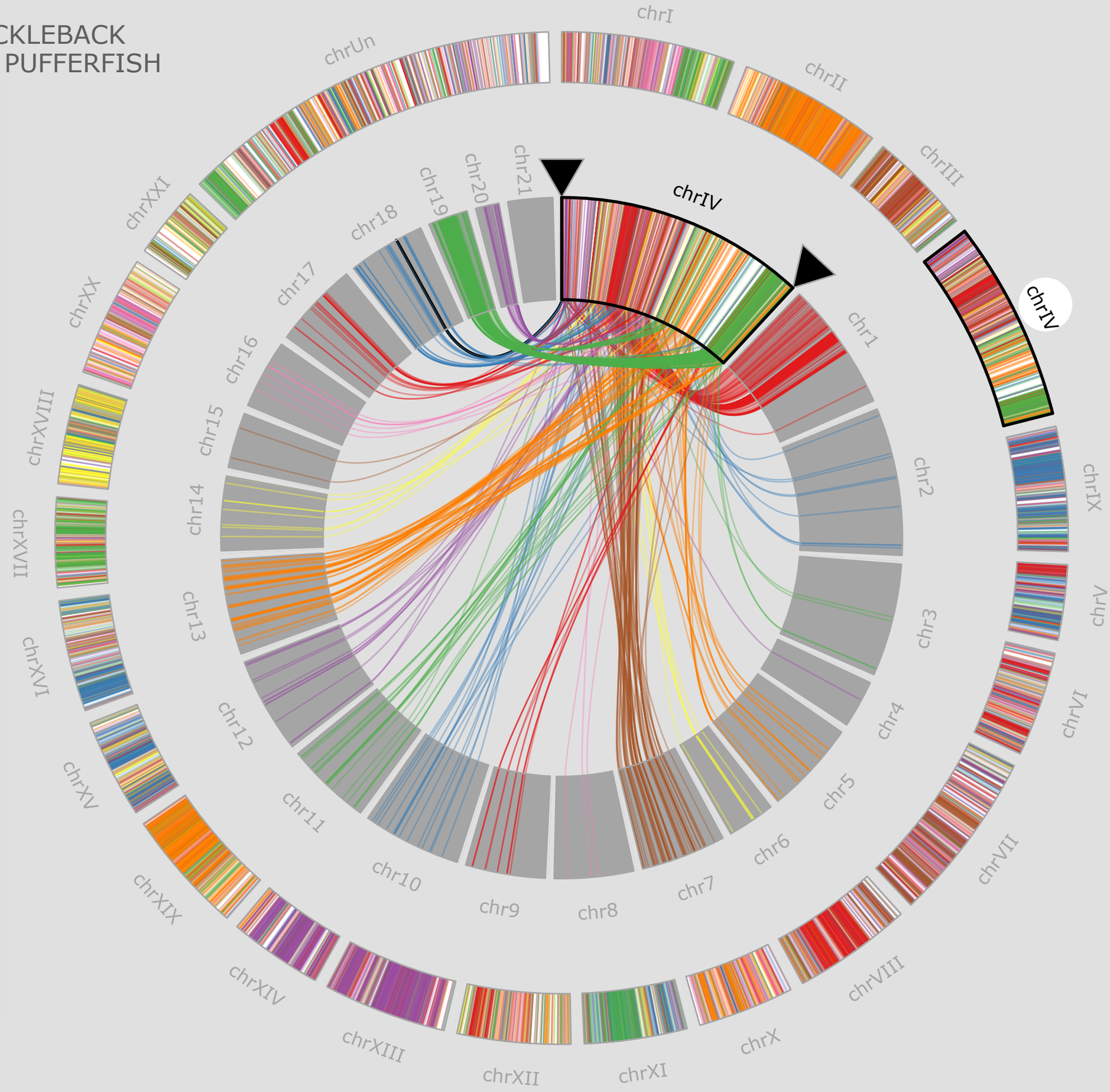
TECHNIQUES

- **multiple linked views**
- **overview + detail: 3 levels**
  - genome: separate-circular, color and connection
    - *edge-bundling*
  - chromosome: rectangular, color
    - *more screenspace for details*
    - *histograms for block stats*
    - *annotations for marking feature positions*
  - block: connection
    - *separate + contiguous histograms for feature stats*

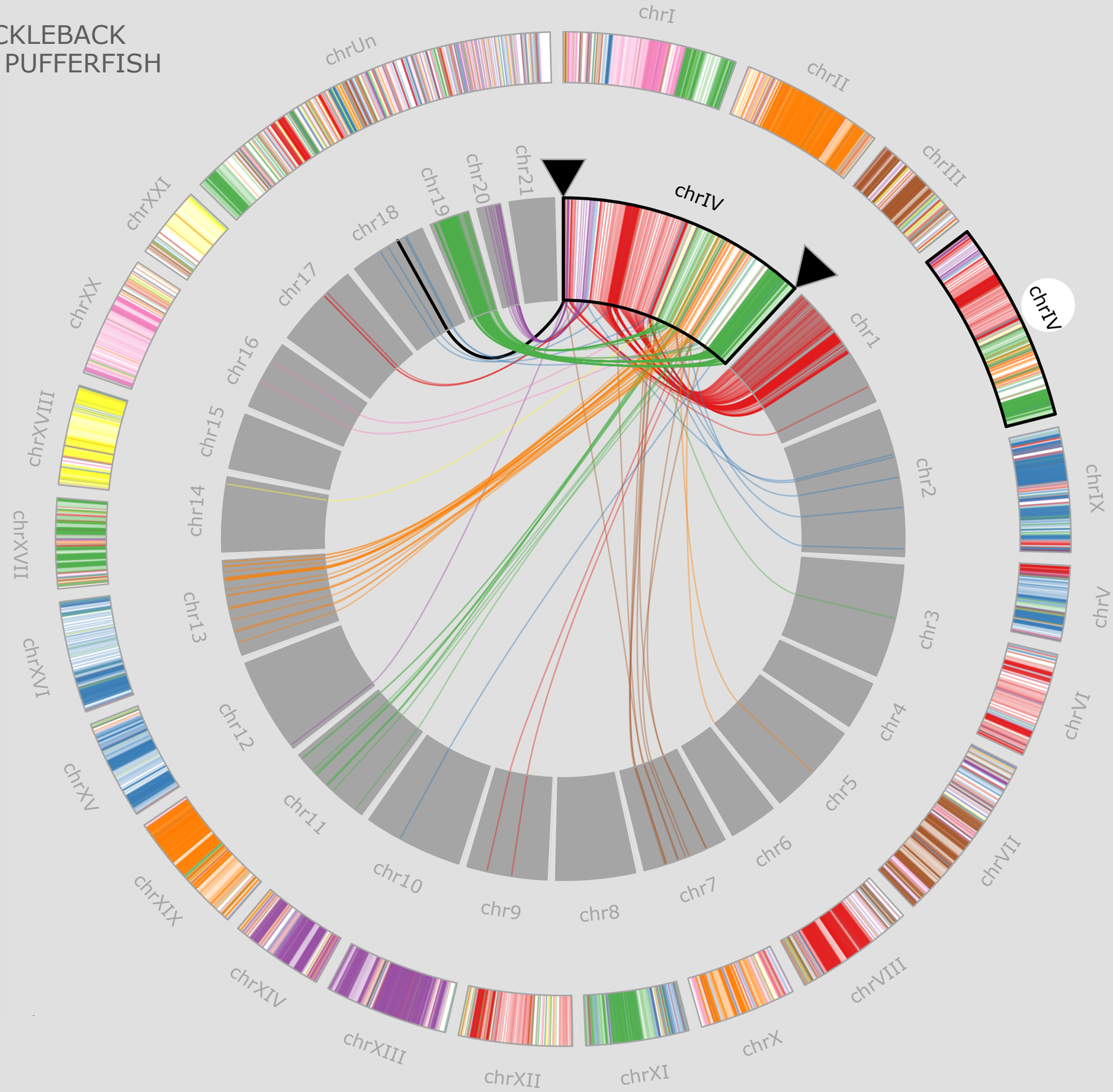
CASE STUDY

photo courtesy of Daniel Berner

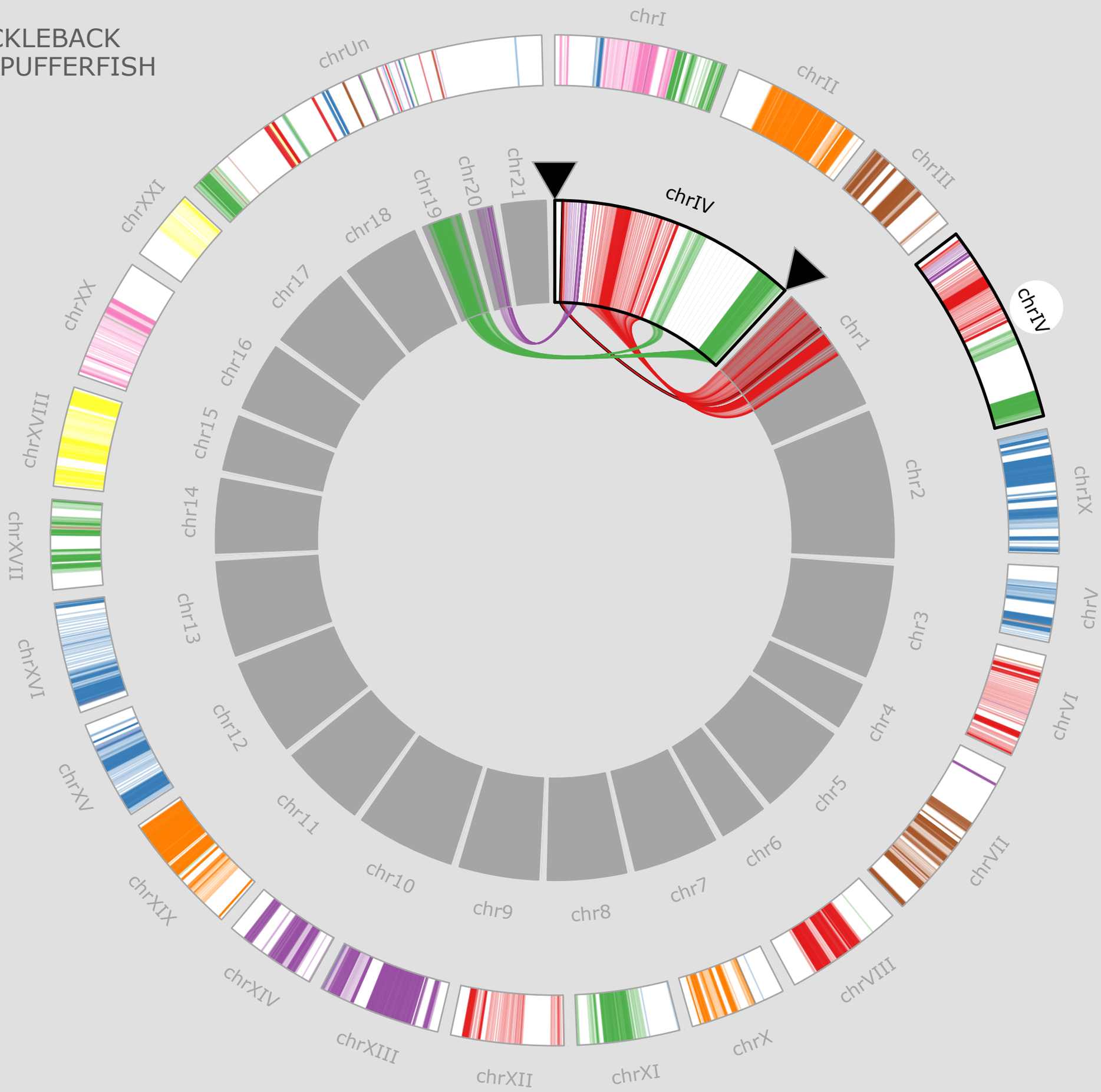
source: STICKLEBACK
destination: PUFFERFISH



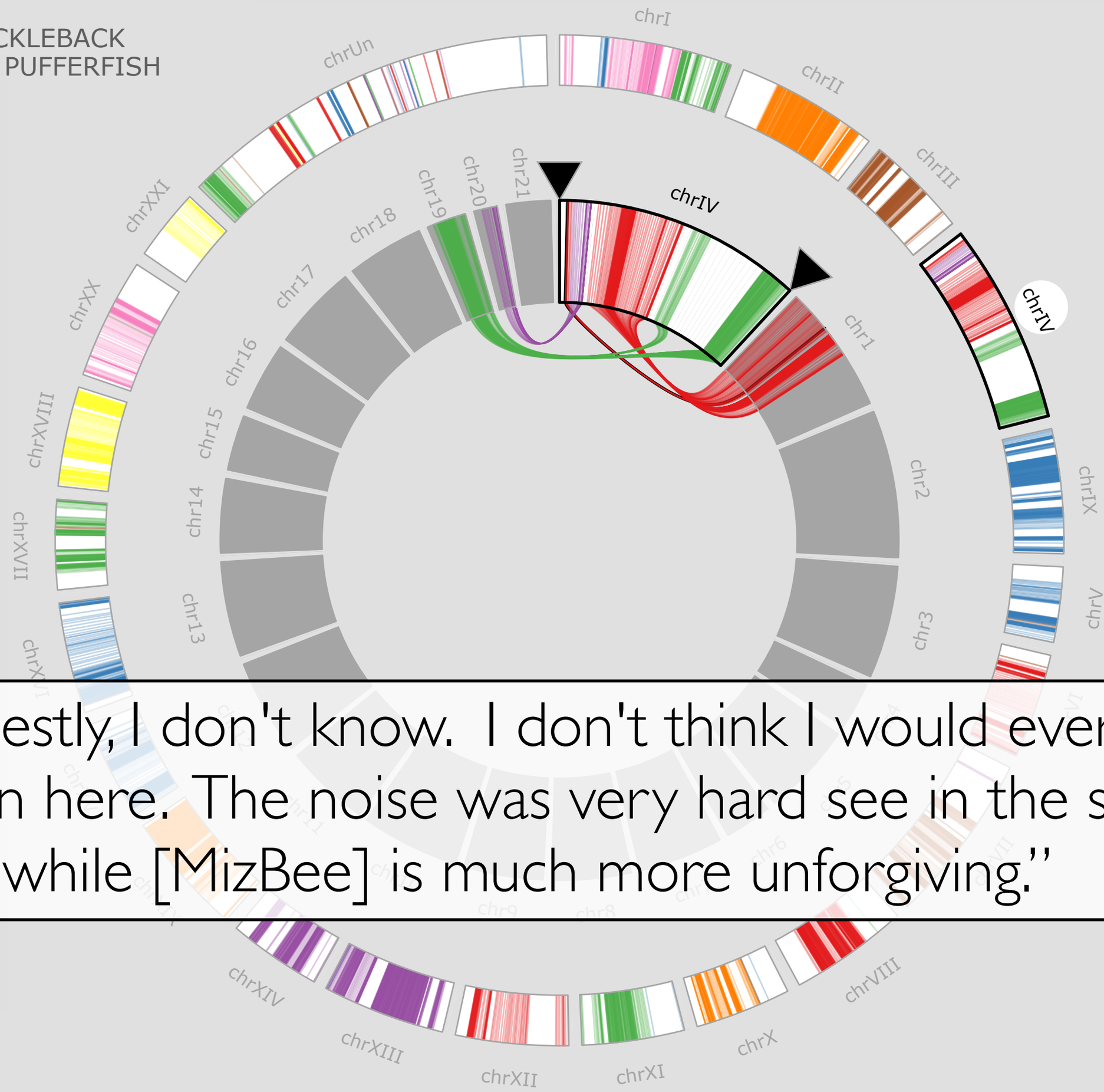
source: STICKLEBACK
destination: PUFFERFISH



source: STICKLEBACK
destination: PUFFERFISH



source: STICKLEBACK
destination: PUFFERFISH



"Honestly, I don't know. I don't think I would ever have gotten here. The noise was very hard see in the scatter plots while [MizBee] is much more unforgiving."

KEY IDEAS

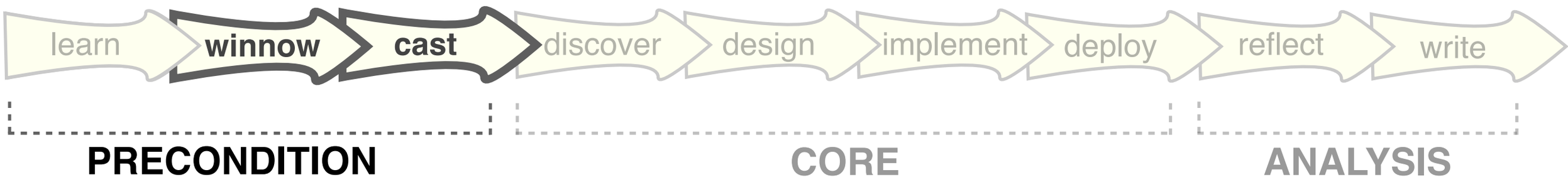
- **power of linked views for multiscale**
- **abstraction from domain to generic problems**
- **visual encoding choices according to known limitations**
- **clutter reduction via edge bundles**
- **two levels of task**
  - block reliability vs higher level science

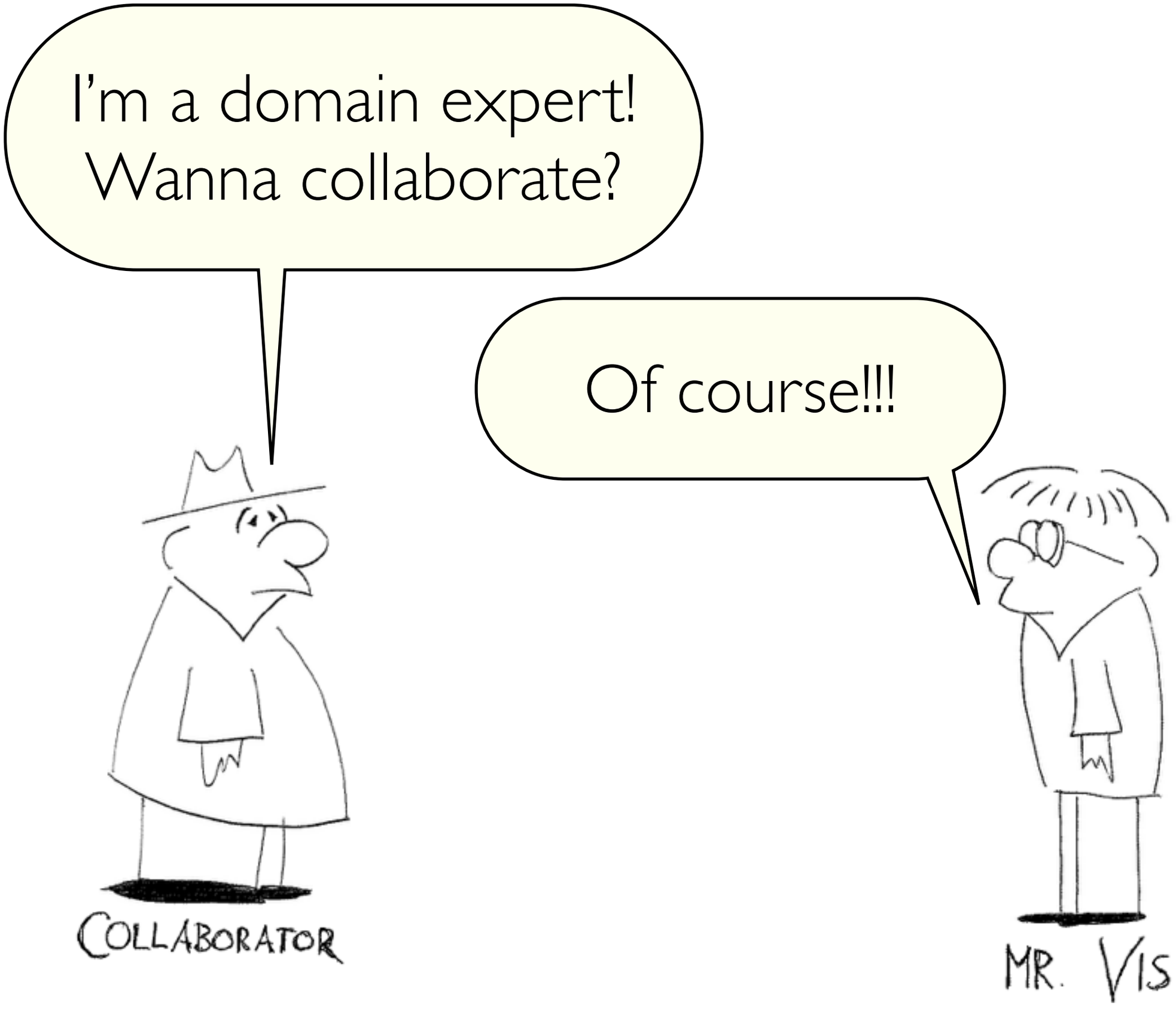
Selected Pitfalls

What to avoid?

PITFALL

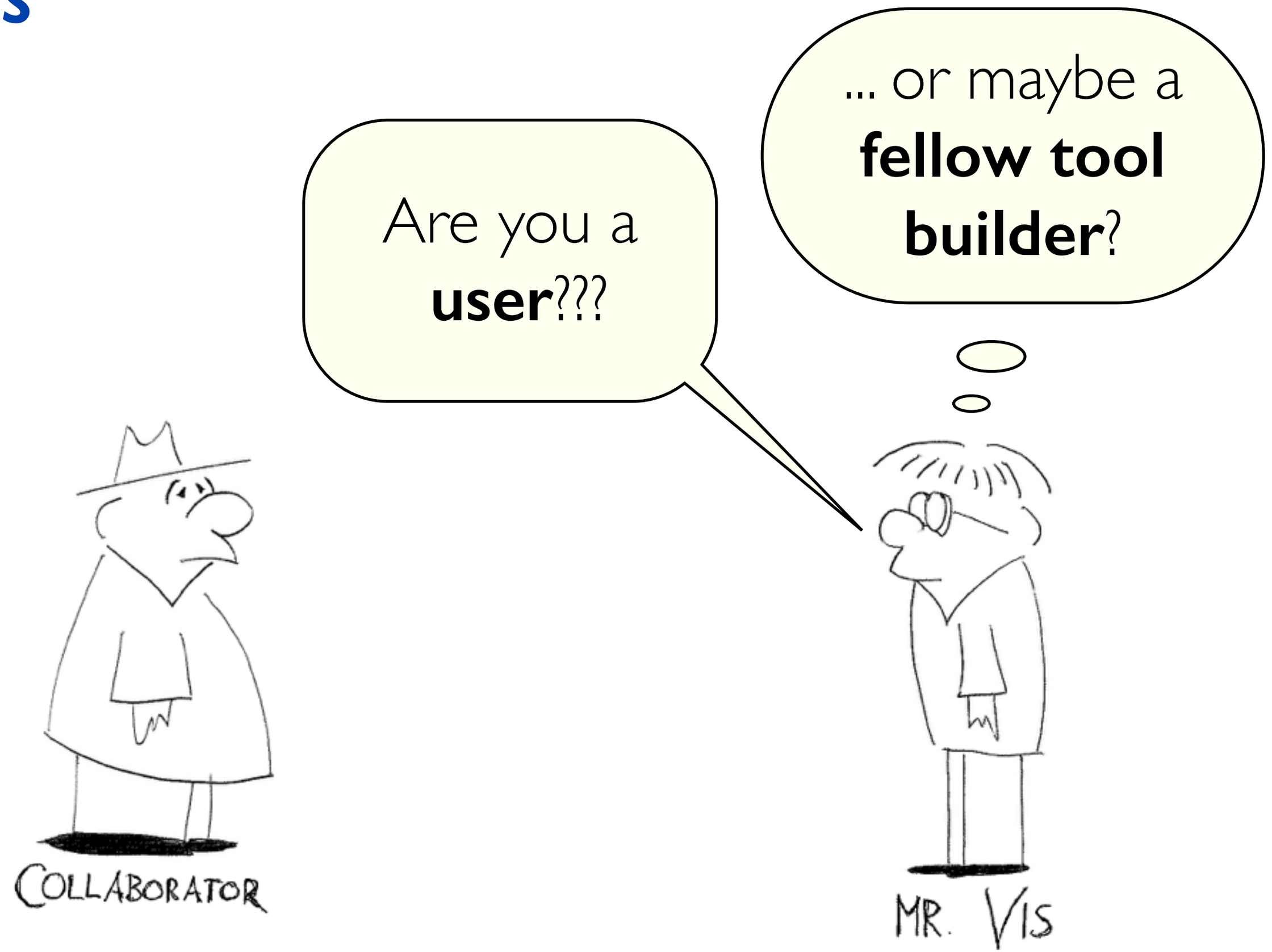
PREMATURE COLLABORATION




I'm a domain expert!
Wanna collaborate?
Of course!!!
COLLABORATOR
MR. VIS

considerations

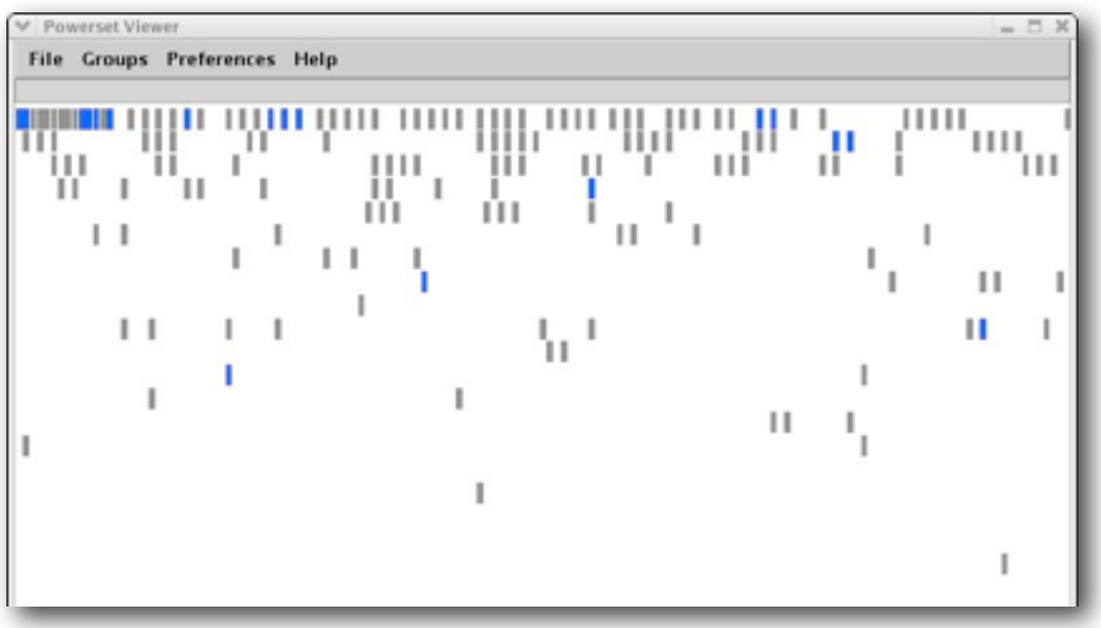
roles


Are you a **user**???
... or maybe a **fellow tool builder**?
COLLABORATOR
MR. VIS

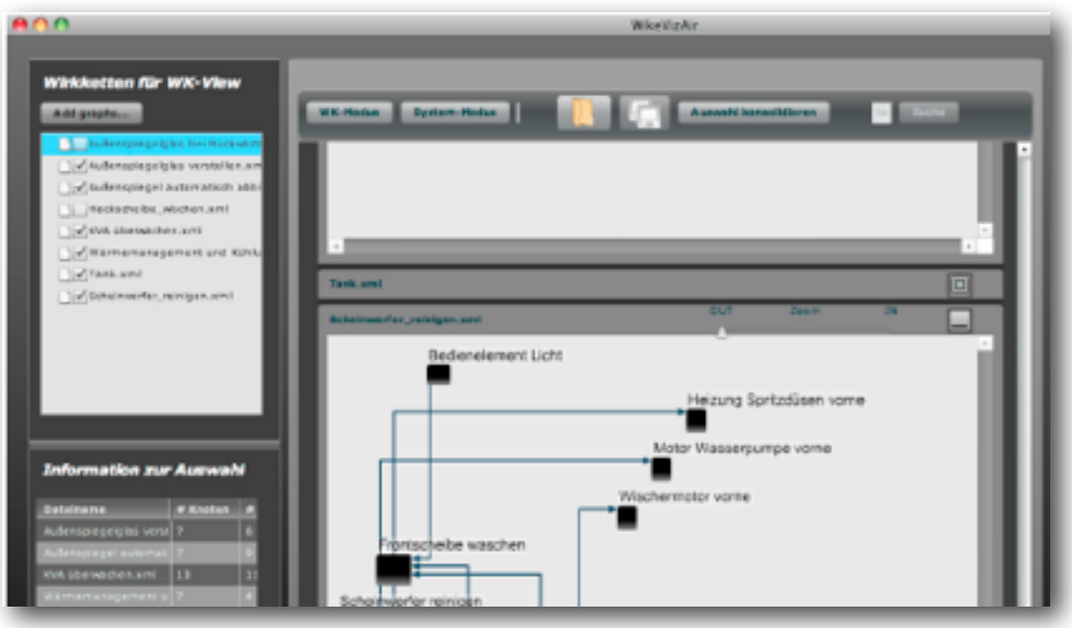
Example From The Trenches

Premature Collaboration!

PowerSet Viewer
2 years / 4 researchers



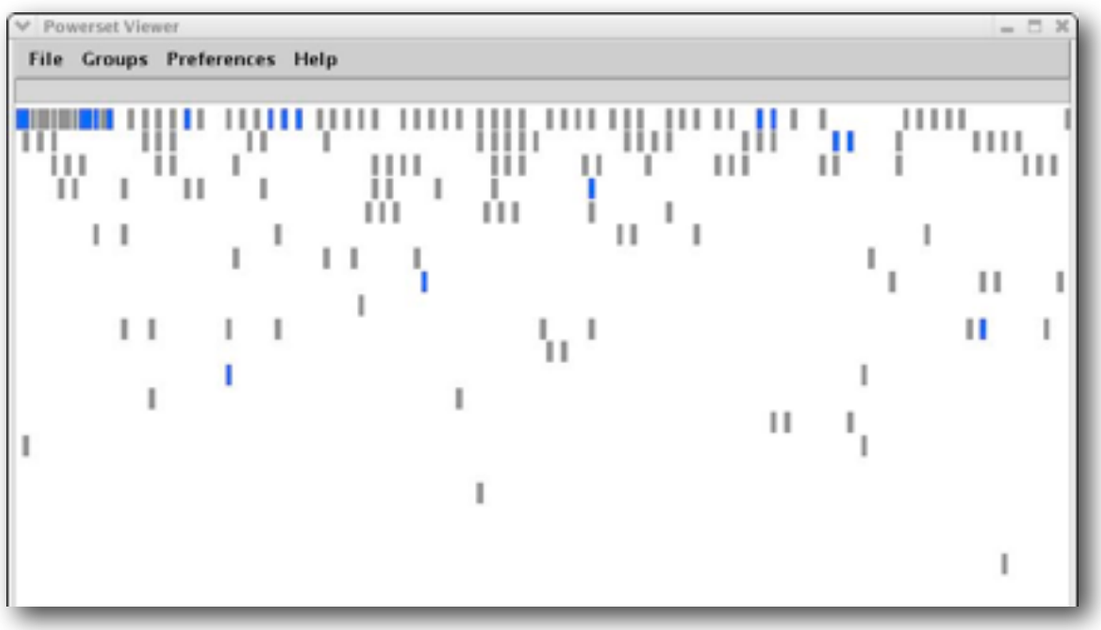
WikeVis
0.5 years / 2 researchers



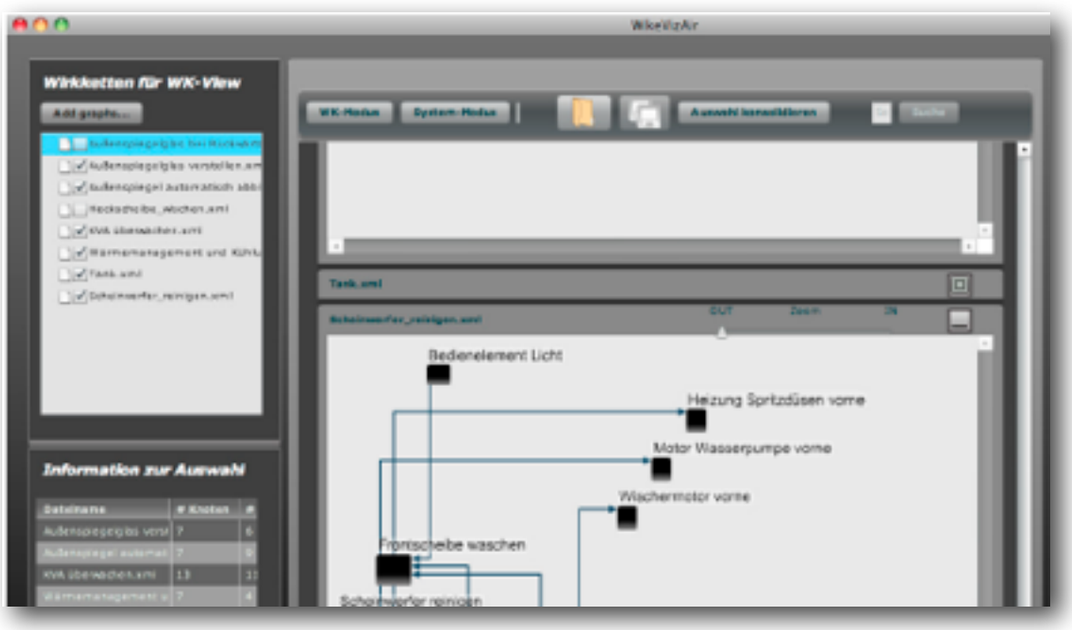
Example From The Trenches

Premature Collaboration!

PowerSet Viewer
2 years / 4 researchers



WikeVis
0.5 years / 2 researchers



Example From The Trenches

Premature Collaboration!

PowerSet Viewer
2 years / 4 researchers

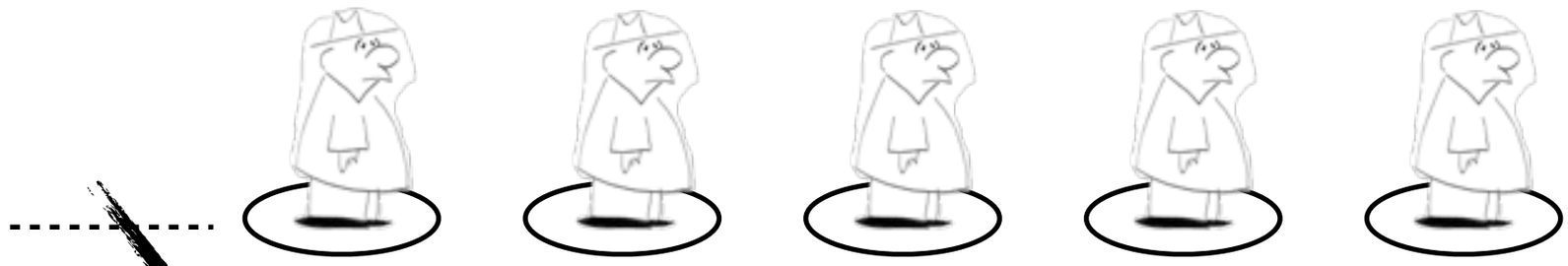
WikeVis
0.5 years / 2 researchers

- Fellow tool builders
- Data promised

Metaphor
Winnowing

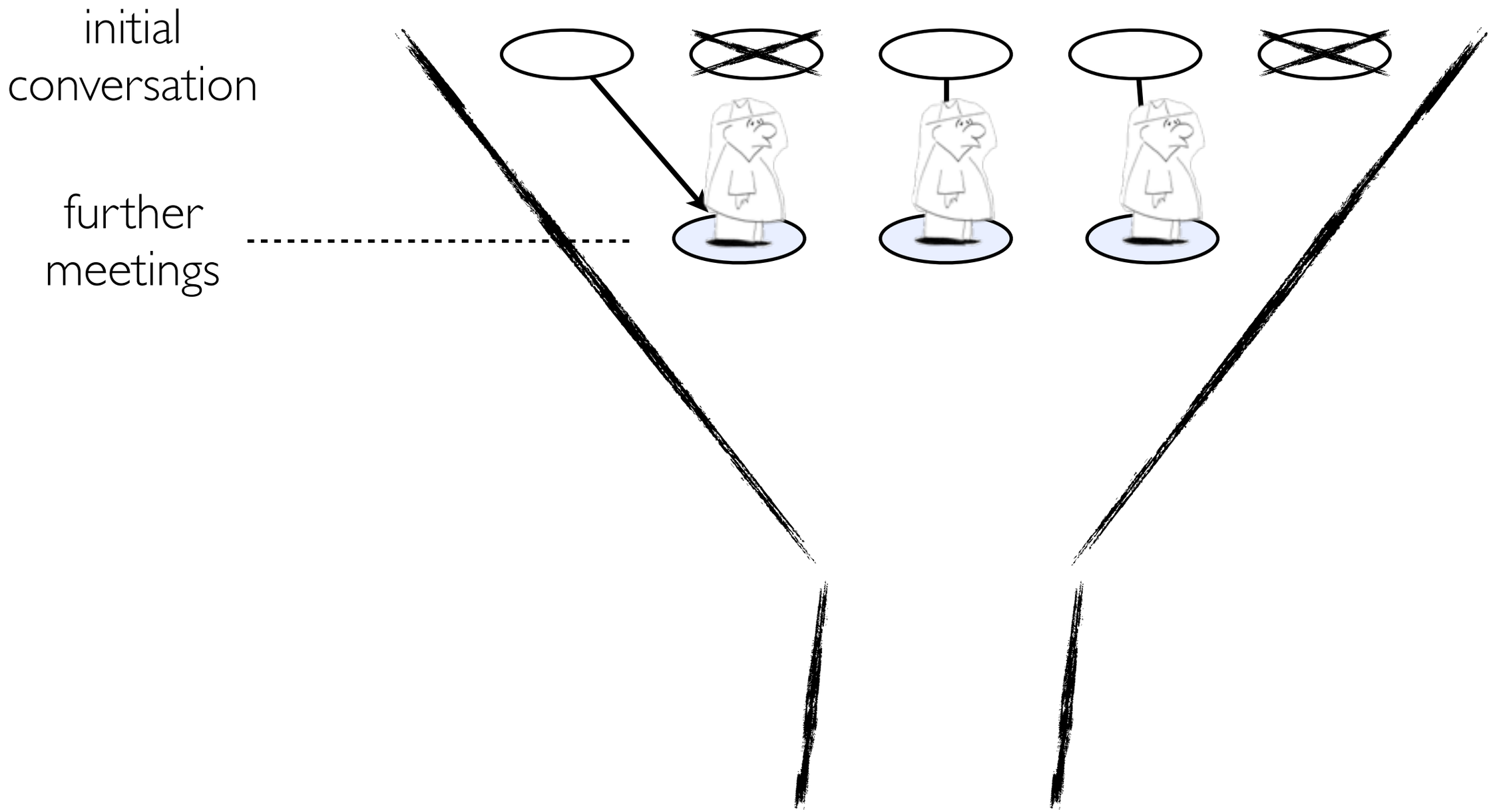
Collaborator Winnowing

initial conversation



(potential collaborators)

COLLABORATOR WINNOWING

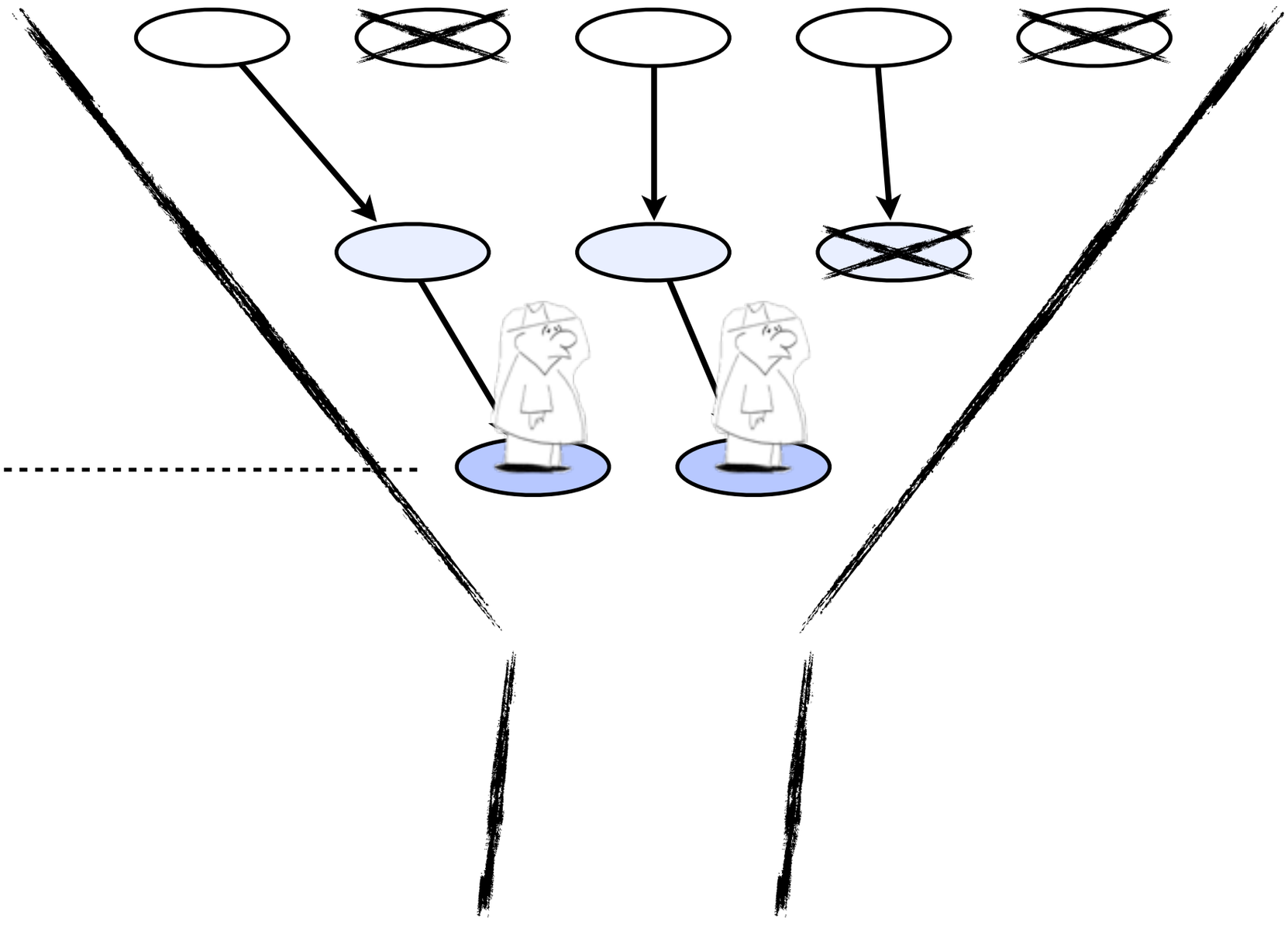


COLLABORATOR WINNOWING

initial conversation

further meetings

prototyping



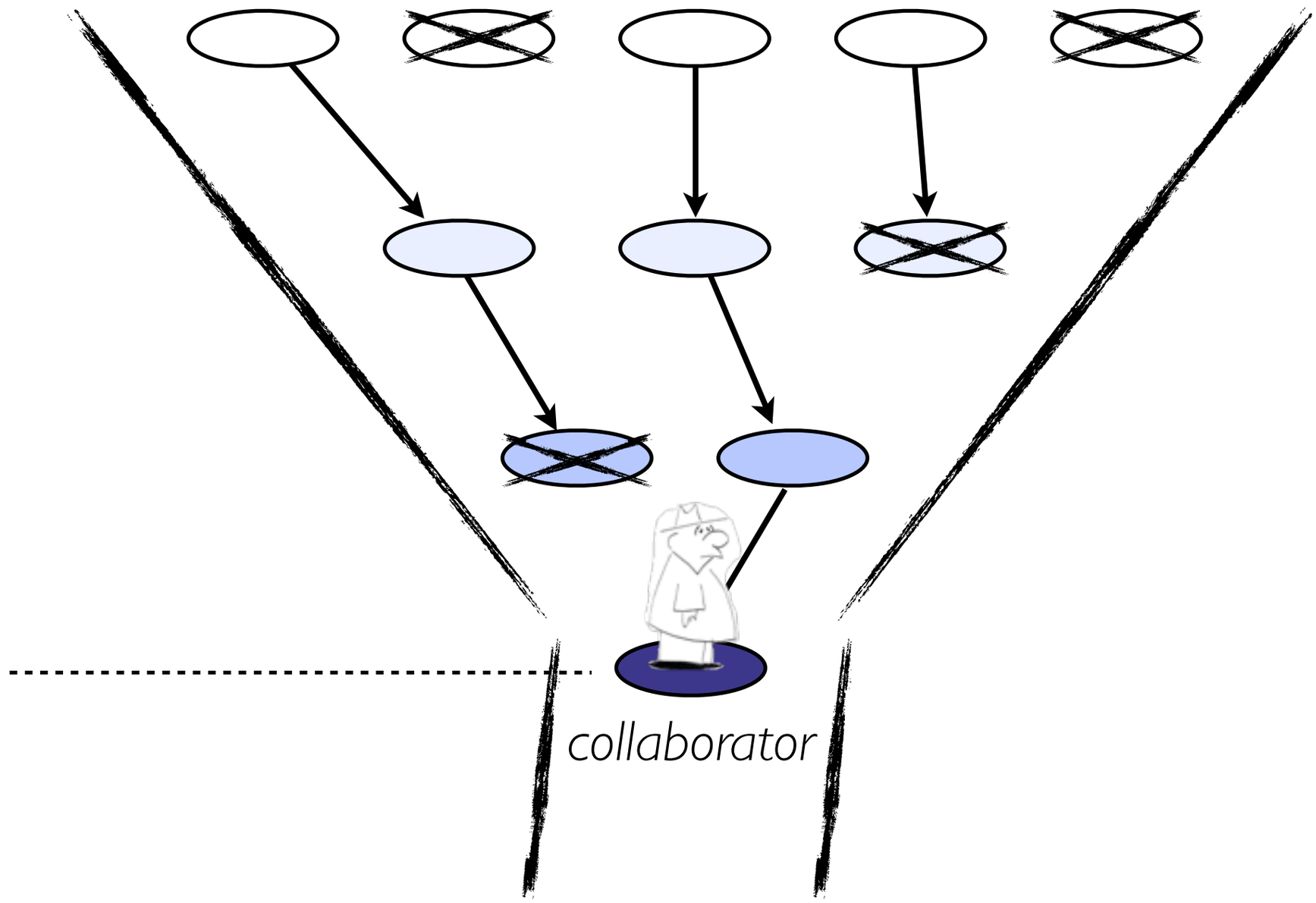
COLLABORATOR WINNOWING

initial conversation

further meetings

prototyping

full collaboration



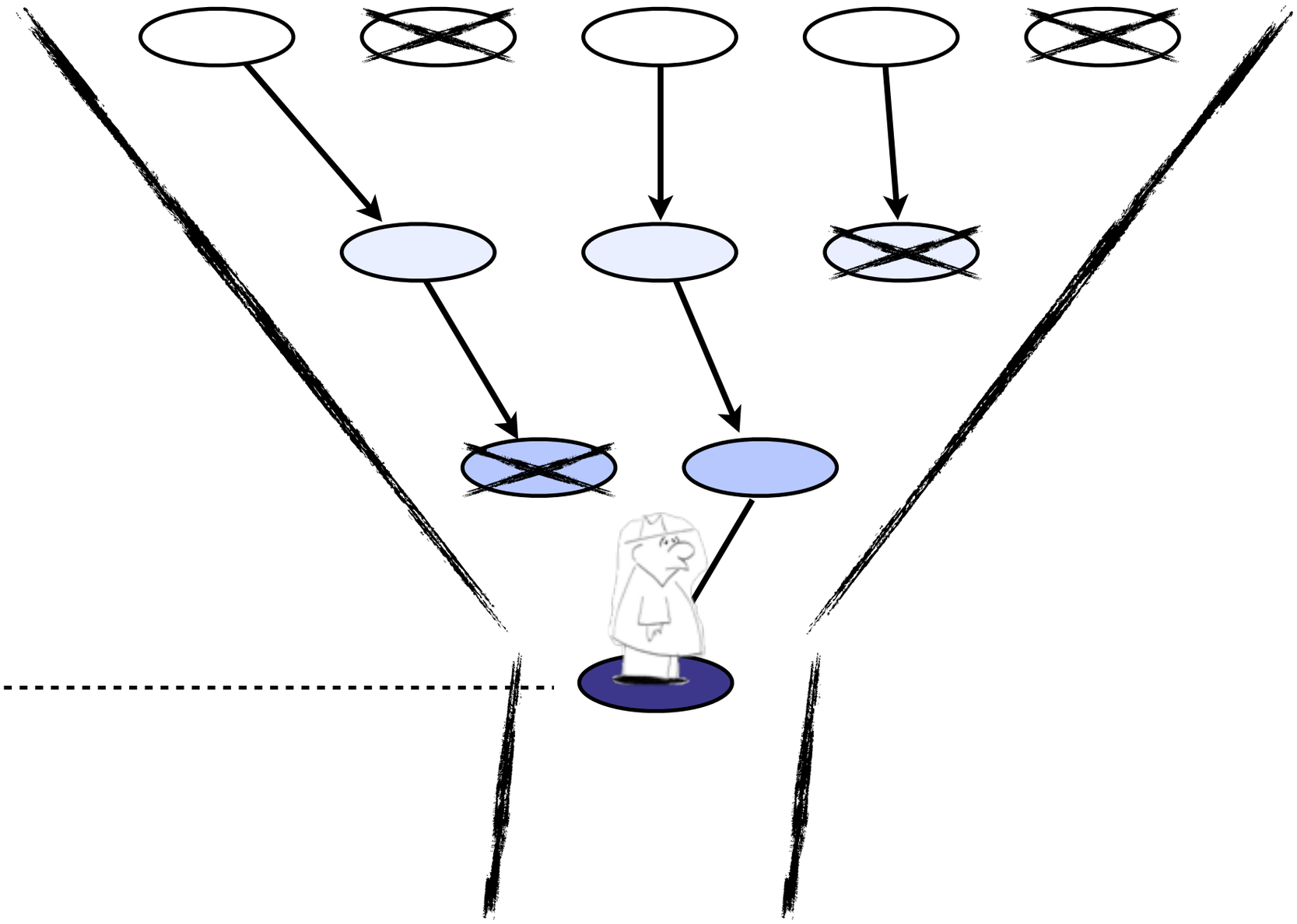
COLLABORATOR WINNOWING

initial conversation

further meetings

prototyping

full collaboration

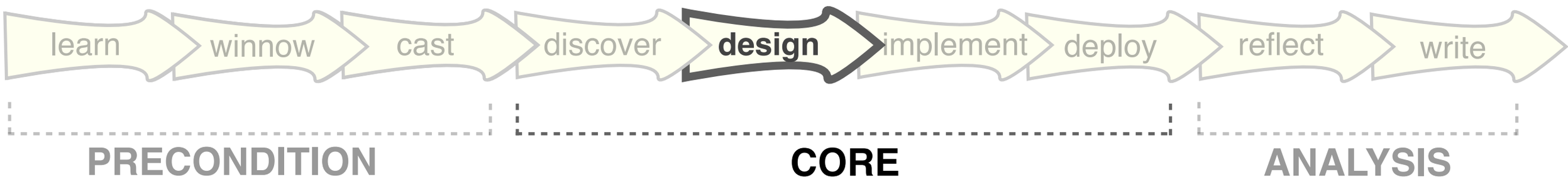


COLLABORATOR WINNOWING

Talk with many, stay with few!

Pitfall

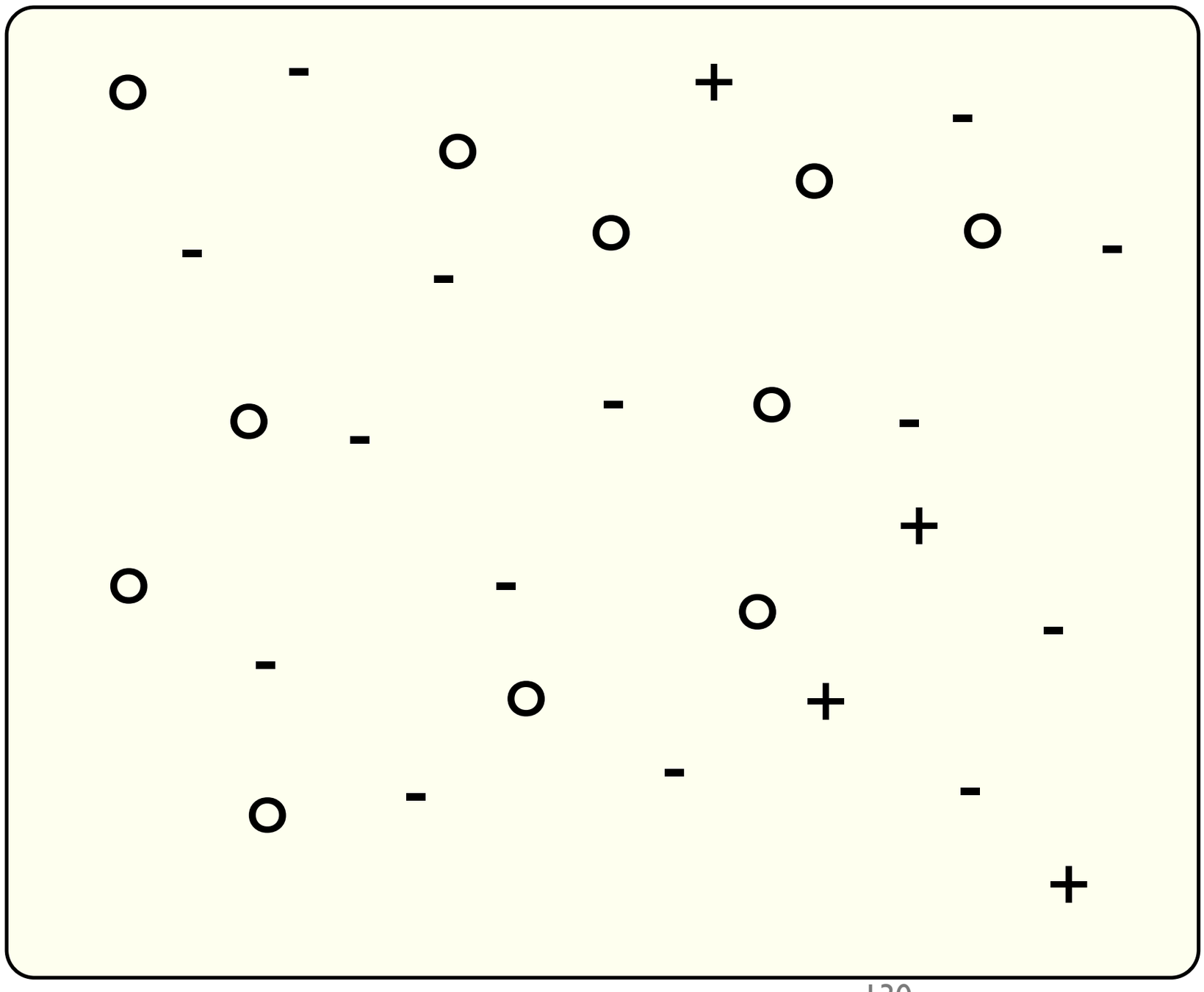
Premature Design Commitment



Of course they need the cool **technique** I built last year!

METAPHOR

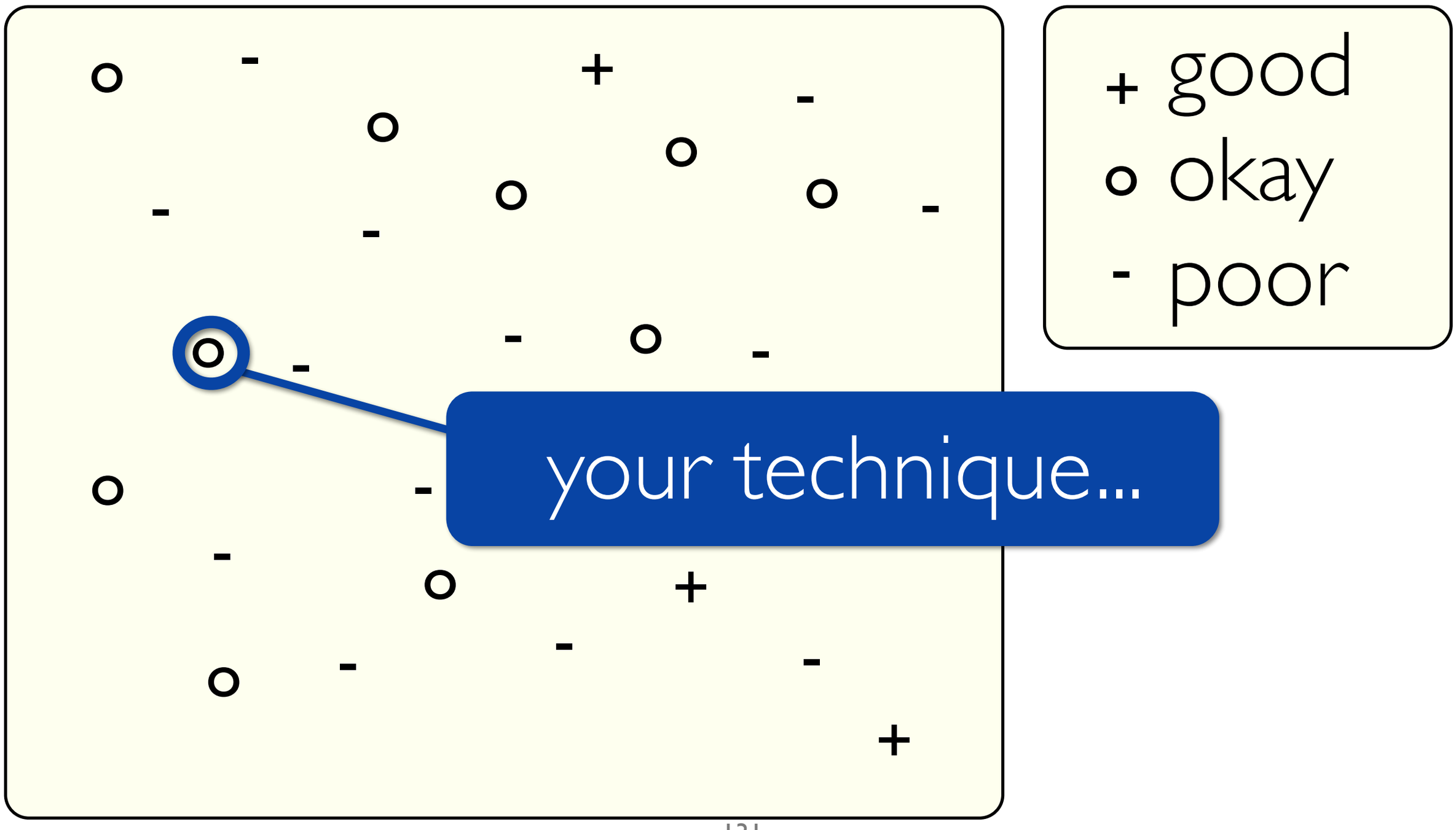
Design Space



+ good
o okay
- poor

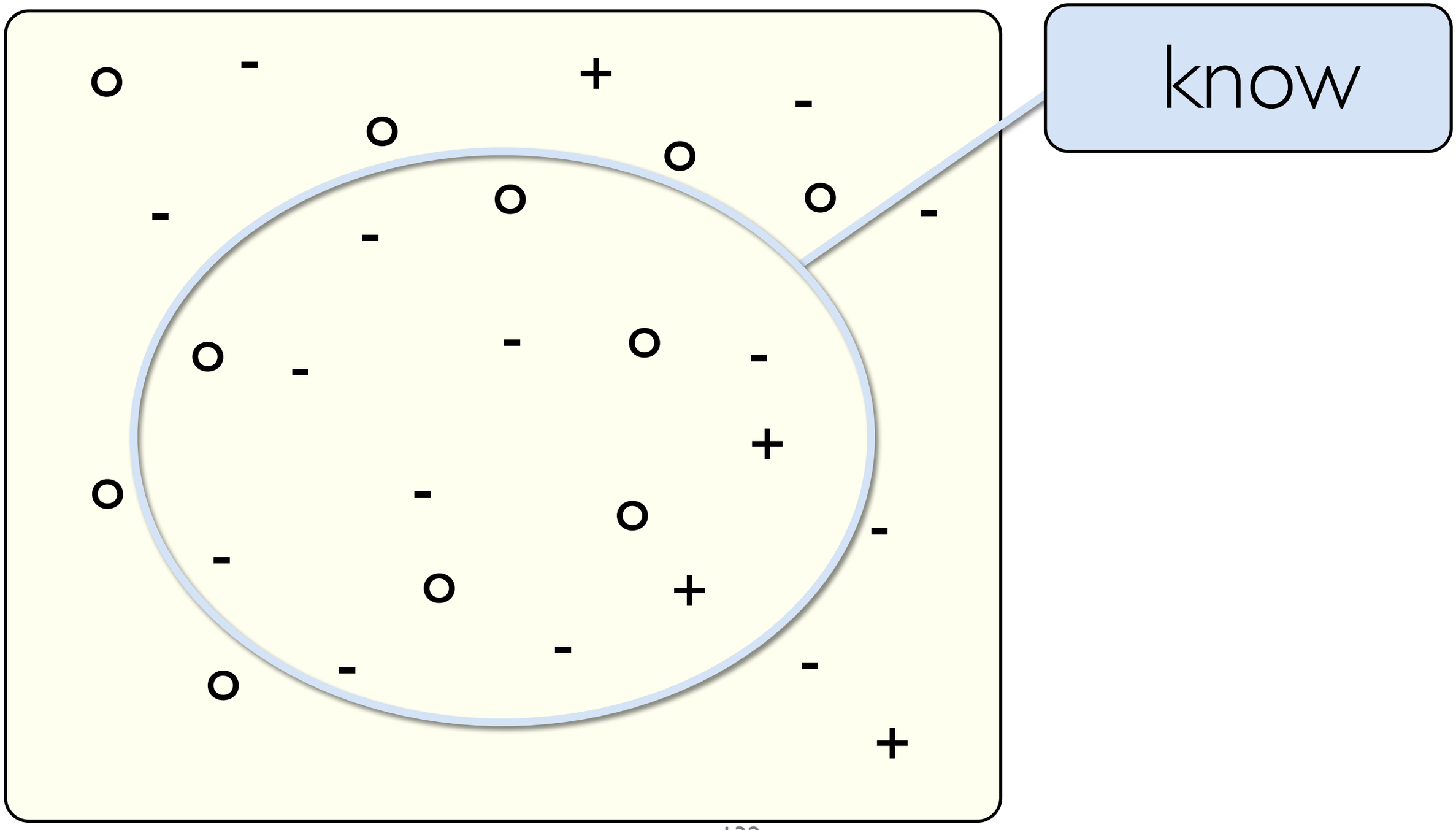
METAPHOR

Design Space



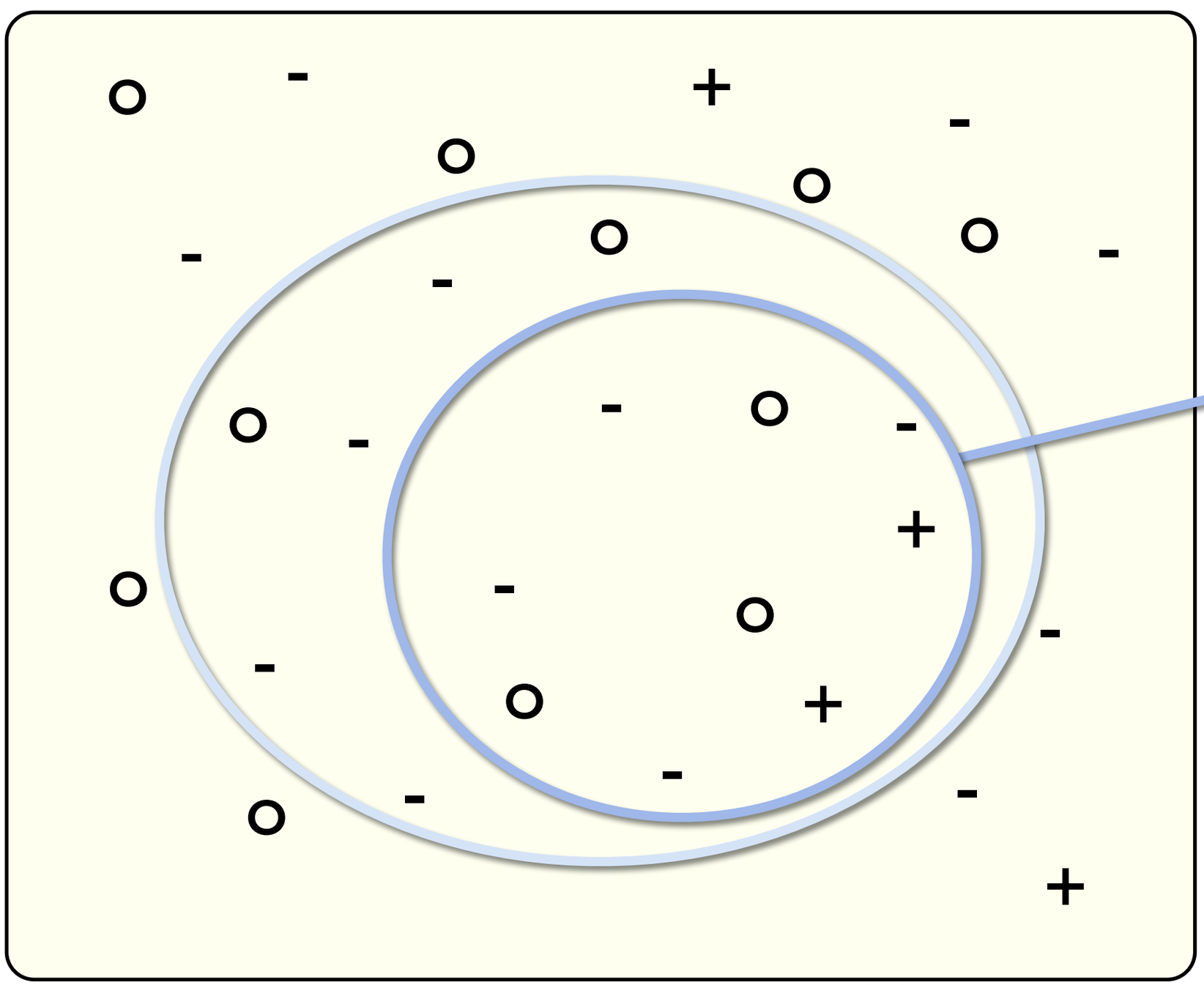
METAPHOR

Design Space



METAPHOR

Design Space

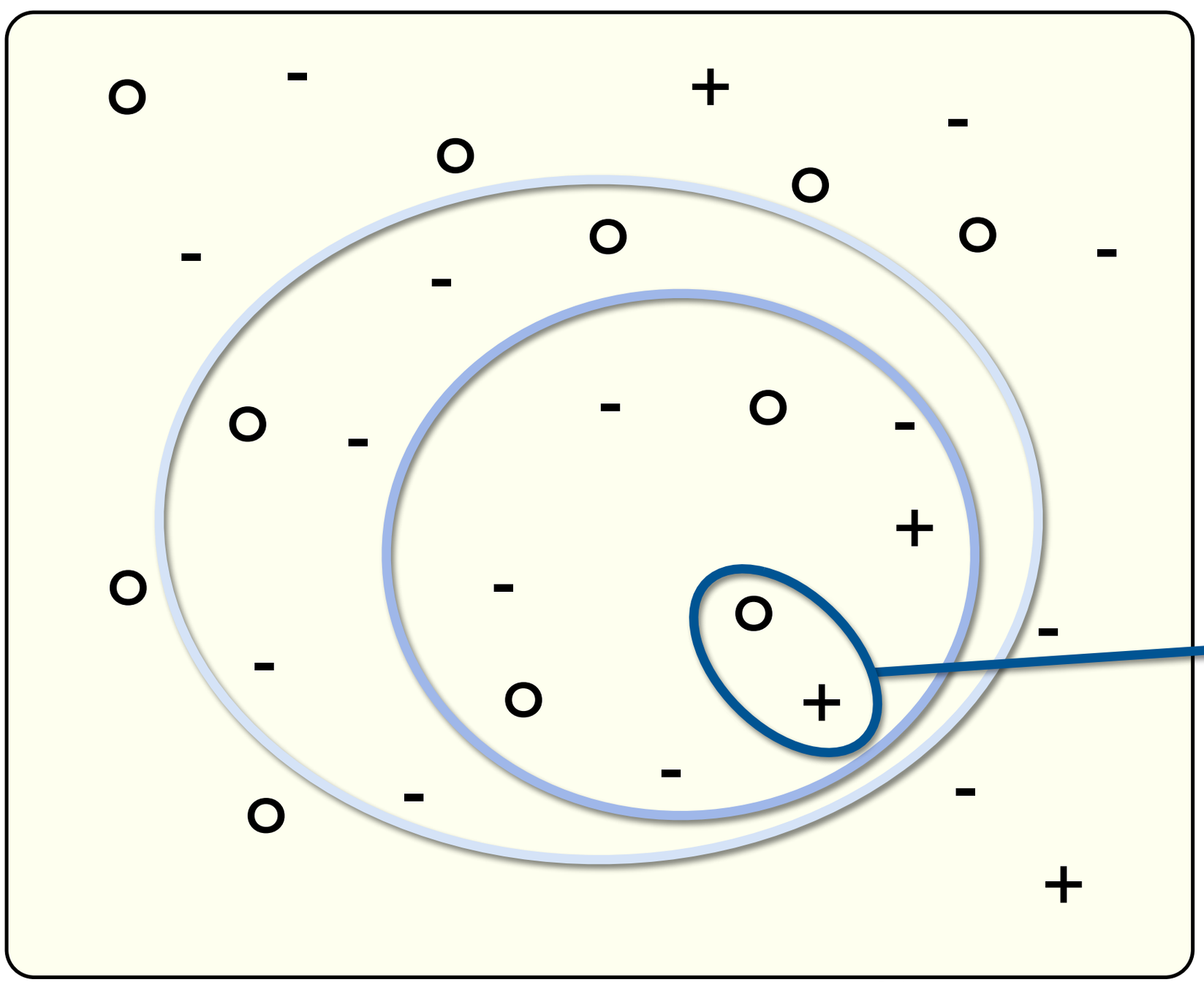


know

consider

METAPHOR

Design Space



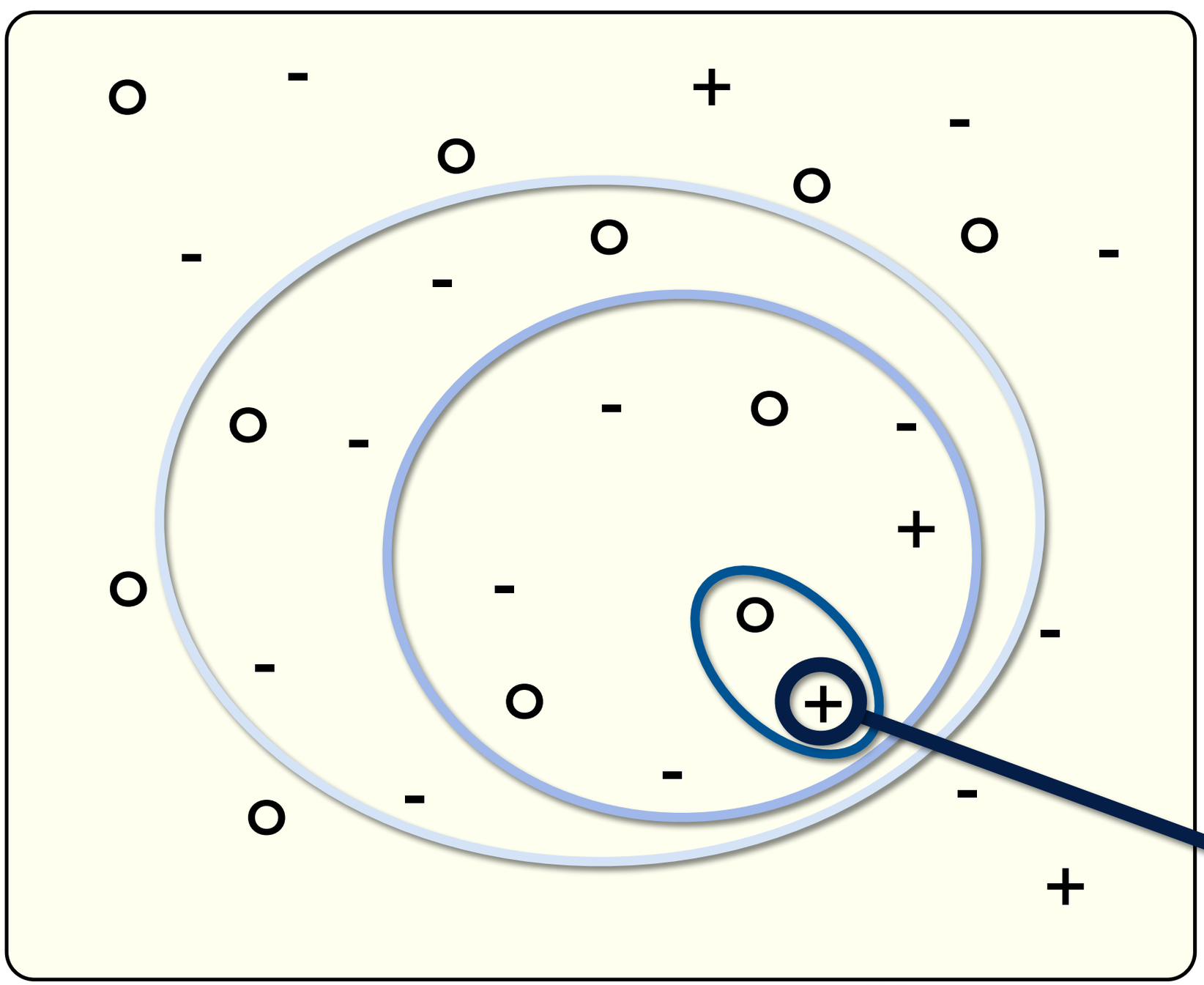
know

consider

propose

METAPHOR

Design Space



know

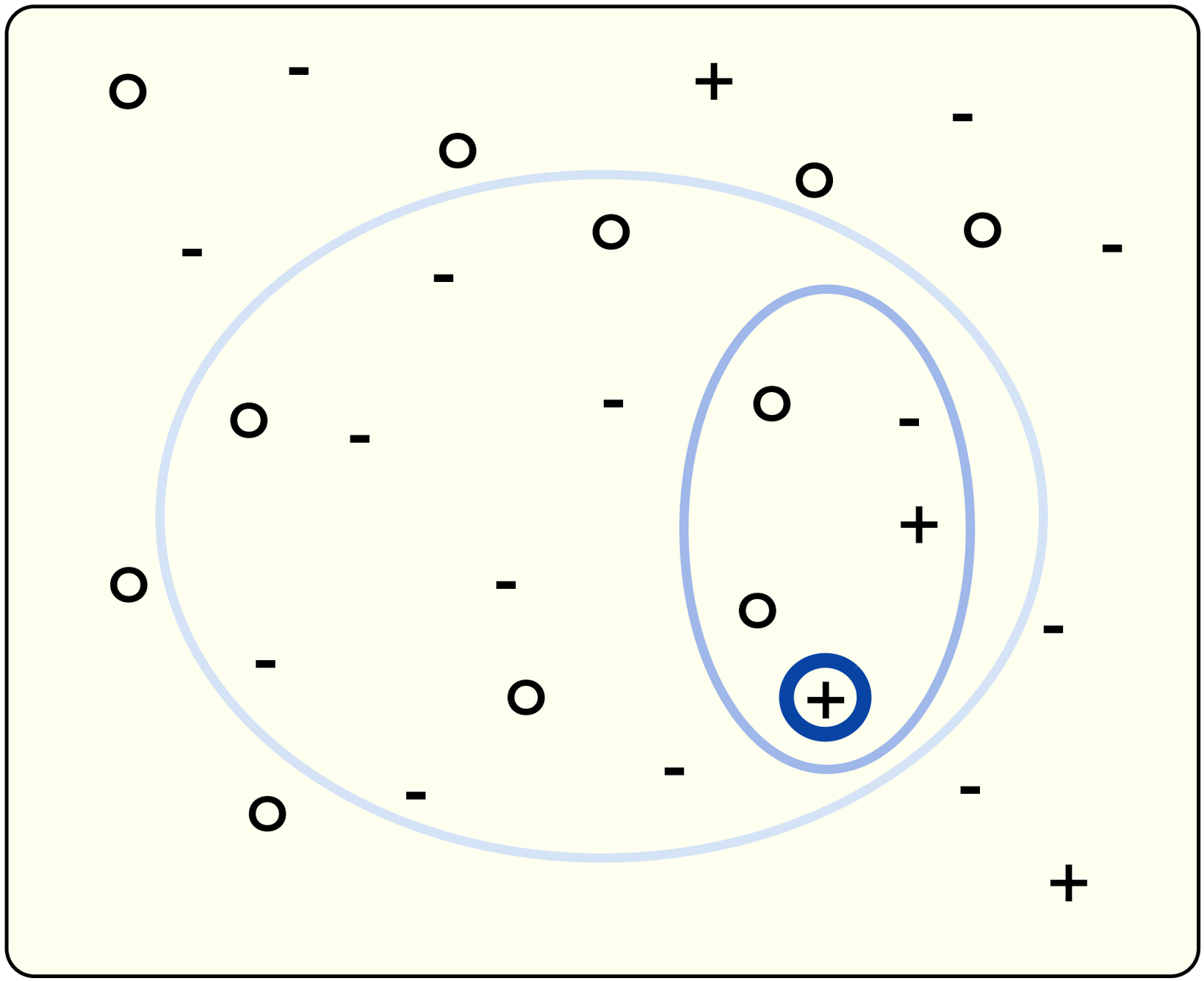
consider

propose

select

METAPHOR

Design Space



+ good
o okay
- poor

consider

propose

select

Metaphor

Design Space

Think broad!

+ good
o okay
- poor

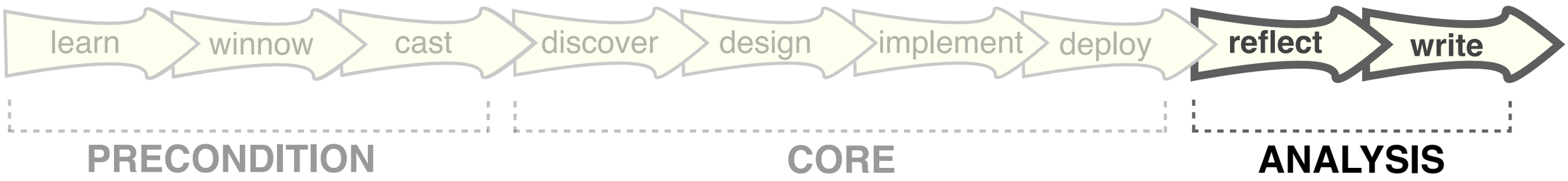
consider

propose

select

Pitfall

Premature Publishing



I can write a design study **paper** in a week!

I can write a design study **paper** in a week!

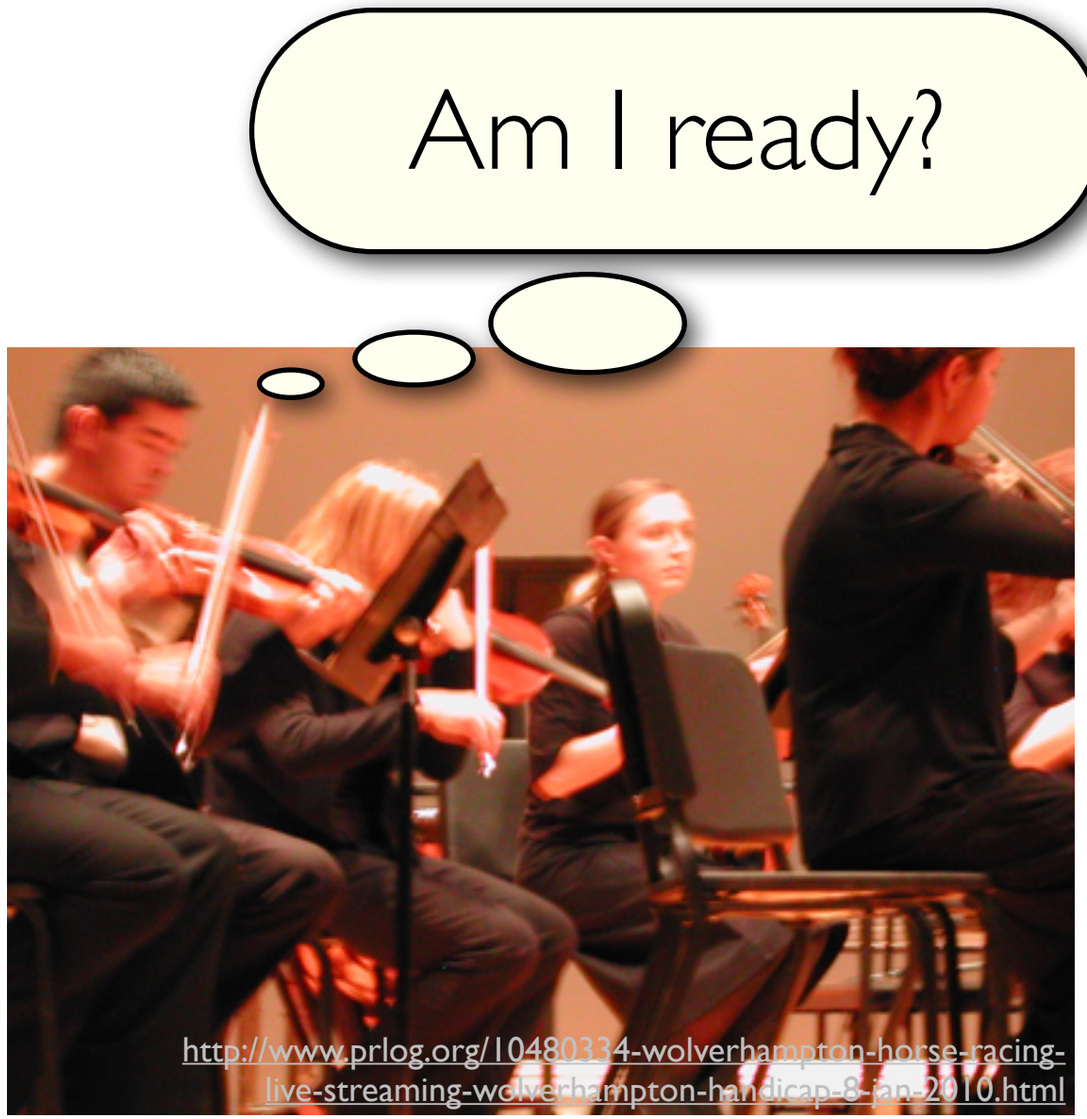
"writing is research"

[Wolcott: Writing up qualitative research, 2009]

METAPHOR

Horse Race vs. Music Debut

technique-driven

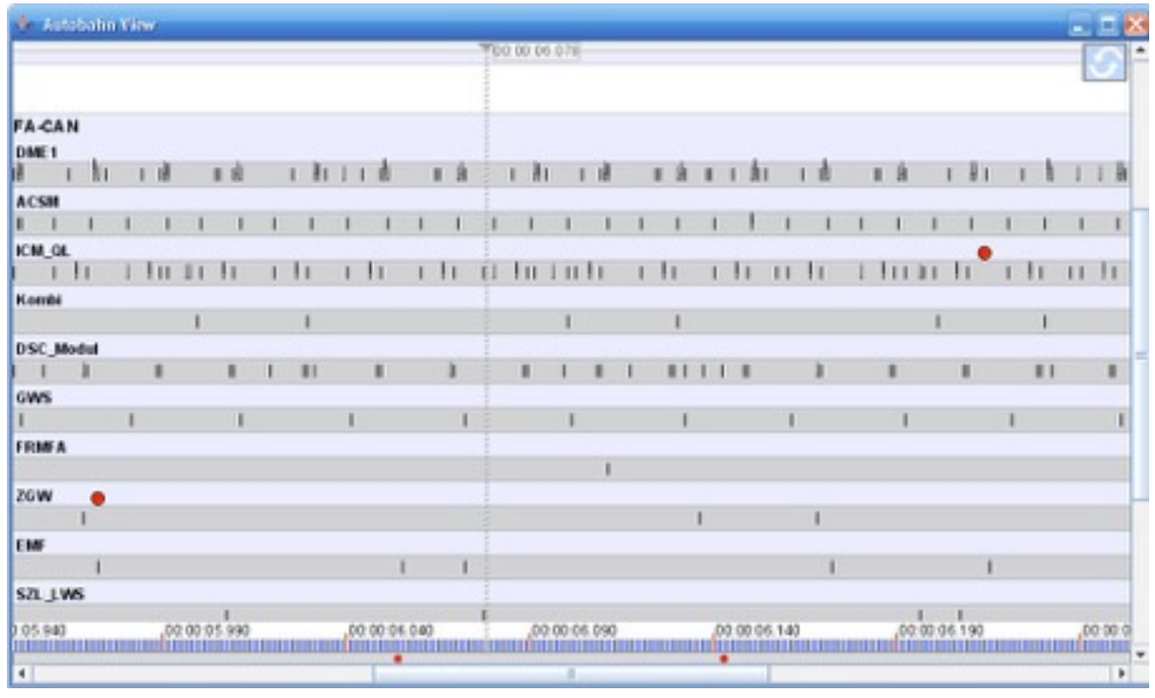


problem-driven

Example From The Trenches

Don’t step on your own toes!

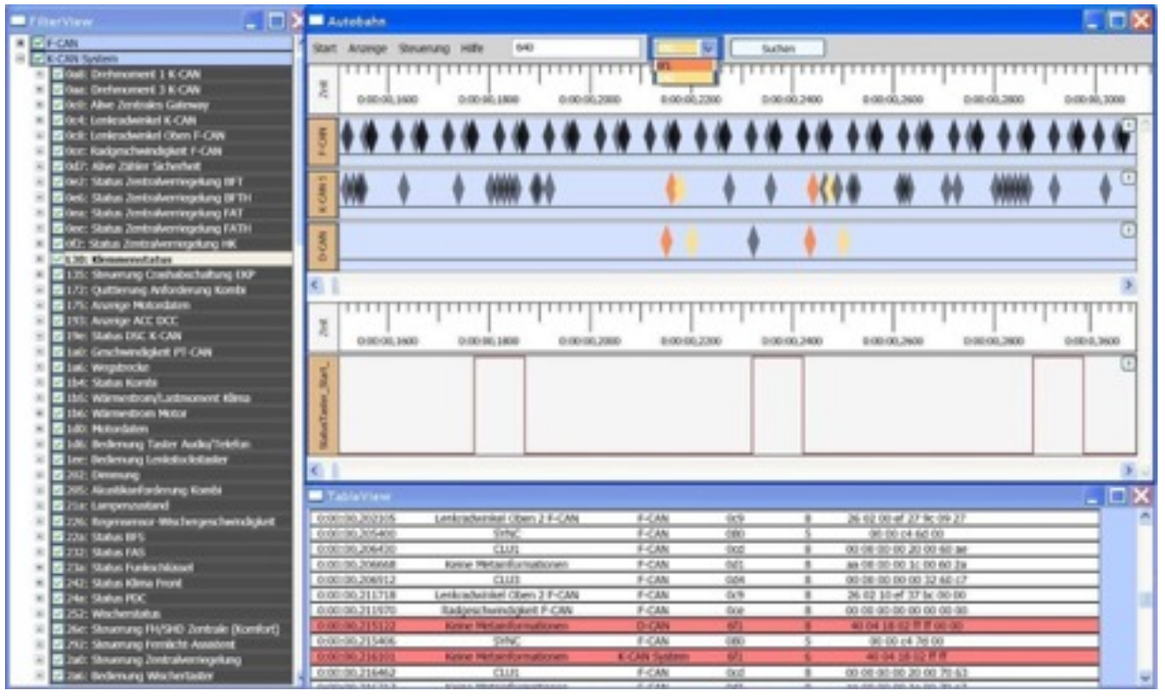
First design round published



AutobahnVis 1.0

[Sedlmair et al., Smart Graphics, 2009]

Subsequent work not stand-alone paper



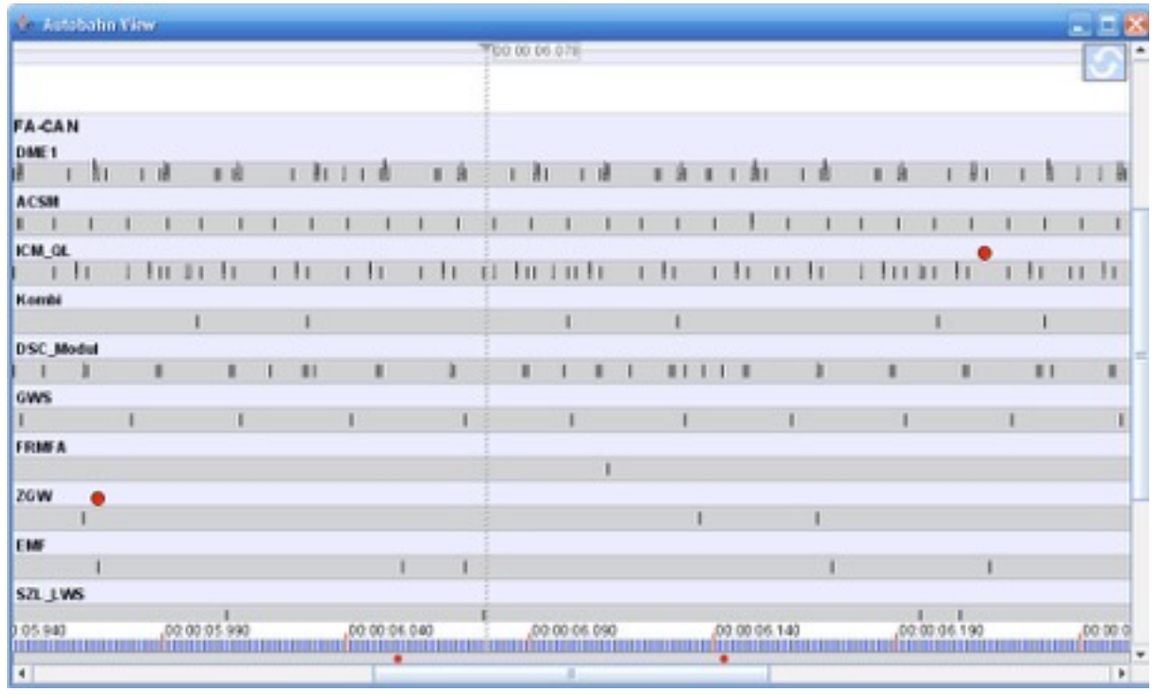
AutobahnVis 2.0

[Sedlmair et al., Information Visualization 10(3), 2011]

Example From The Trenches

Don’t step on your own toes!

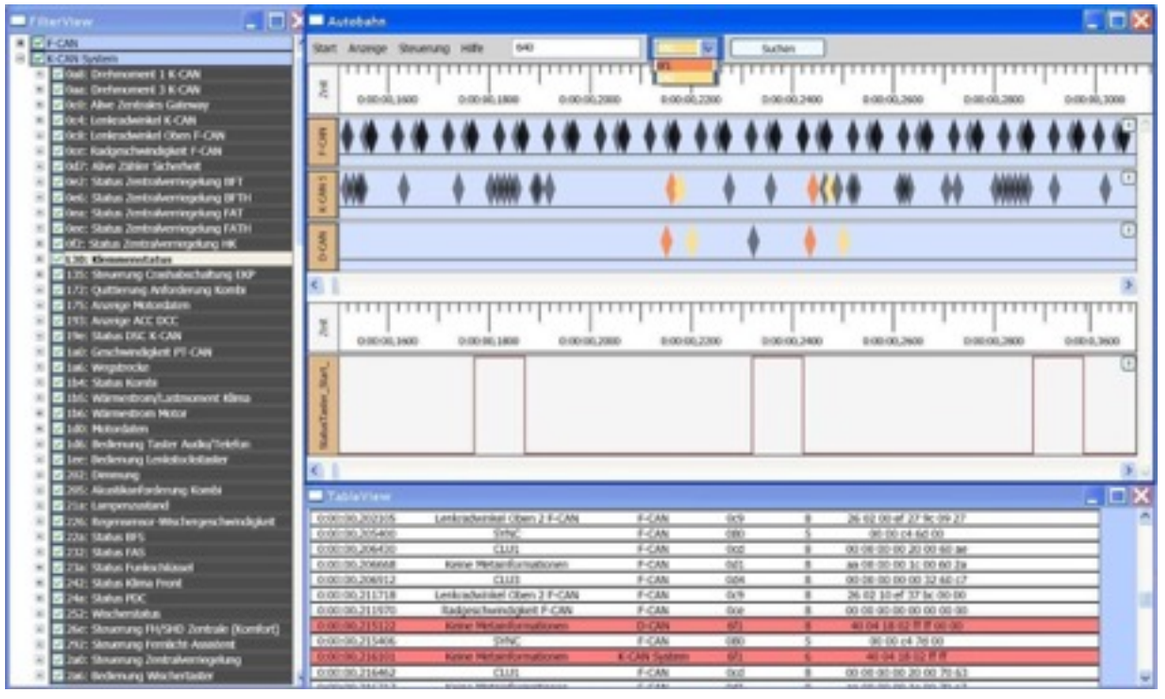
First design round published



AutobahnVis 1.0

[Sedlmair et al., Smart Graphics, 2009]

Subsequent work not stand-alone paper



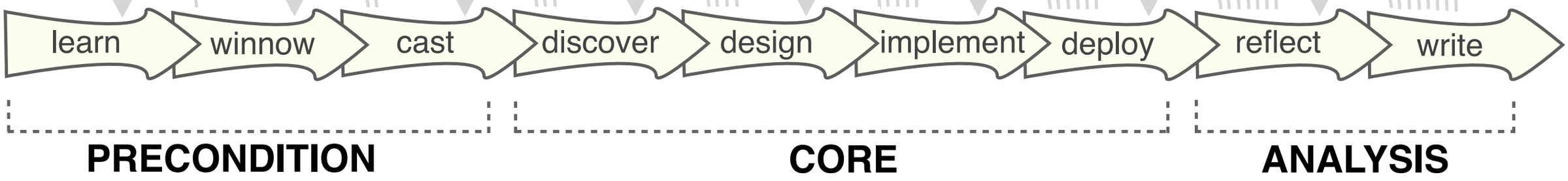
AutobahnVis 2.0

[Sedlmair et al., Information Visualization 10(3), 2011]

Future Work

A Start, not an End!

Not the only way!



L25: Molecular animation

REQUIRED READING

Opinion

Cell PRESS

Special issue – CellBio-X

Animating the model figure

Janet H. Iwasa

Department of Cell Biology, Harvard Medical School, 240 Longwood Avenue, Boston, MA 02115, USA

In all branches of scientific inquiry, researchers build models that enable them to visualize, formulate and communicate their hypotheses to others. In cell biology, our conceptual understanding of a process is typically embodied in a model figure. These visual models should ideally represent pre-existing knowledge of molecular interactions, movement, structure and localization but, in reality, they often fall short. Cell biologists have begun to look to the use of three-dimensional animation to visualize and describe complex molecular and cellular events. In addition to aiding teaching and communication, animation is emerging as a powerful tool for providing researchers with insight into the processes that they study. Two case studies focusing on the structure/function of the motor protein dynein and the structure of the centriole are discussed.

Molecular Animation as a Teaching, Communication, and Discovery Tool

Over the past several years, there has been a steep increase in the use of animation to communicate dynamic molecular processes to a wide range of audiences. Biology students can view animations on numerous educational websites and in media packaged with their textbooks, and are increasingly presented with biological animations in classrooms and lecture halls. Studies in high school and graduate-level biology courses have shown that the use of animations in teaching has a positive impact; students who have viewed animations as part of their curriculum report a higher level of interest in the course material, and have biochemical and genetics assays. These visualizations can communicate a specific hypothesis for how a molecular process proceeds, and often can do so in a much more efficient and intuitive manner than a written description and with more accuracy and detail than a simplistic diagram or illustration.

An example of this type of dynamic molecular model is shown in Figure 1. In collaboration with Tomas Kirchhausen (Harvard Medical School), I have created an animation that illustrates the process of clathrin-mediated endocytosis, focusing on the assembly and disassembly of the clathrin cage around a newly formed vesicle. A majority of the proteins shown in the animation are derived from crystal structures and the animation shows the progress of endocytosis in "real time" (based on light microscopy), such that the formation of the clathrin cage takes approximately one minute, and disassembly follows rapidly, spanning just a few seconds [3].

Historically, physical 3D models of molecules have been used as thinking tools and have aided in scientific discovery (Box 1). In some cases, these models were created as an educational device, but were later brought into the laboratory and used to help researchers visualize and solve a problem. I believe that molecular animation will follow a similar trajectory, and that animations will increasingly become tools that enable thinking and discovery, in addition to aiding teaching and communication.

The Making of a Model Figure

Cell biologists often employ a model figure when presenting