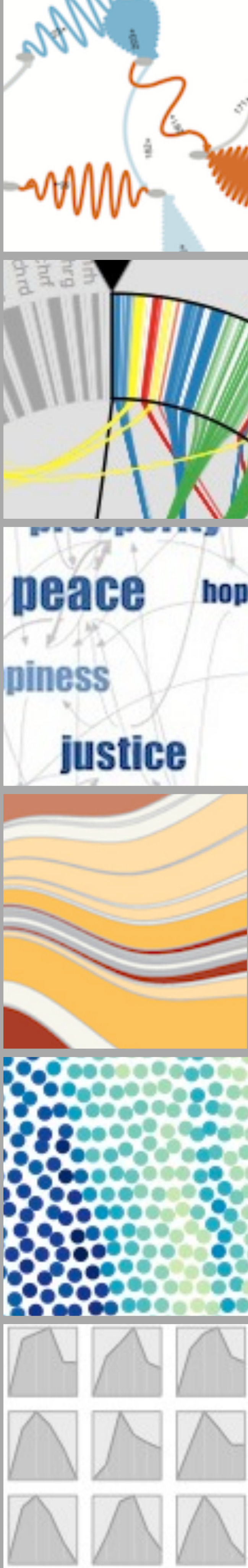
cs6964 | January 17 2012

PROCESS

Miriah Meyer
University of Utah



ADMINISTRIVIA

LAST TIME

Tufte's integrity principles

Clear, detailed, and thorough labeling should be used to defeat graphical distortion and ambiguity.

The representation of numbers, as physically measured on the surface of the graphic itself, should be directly proportional to the numerical quantities represented. *aka The Lie Factor*

Show data variation, not design variation.

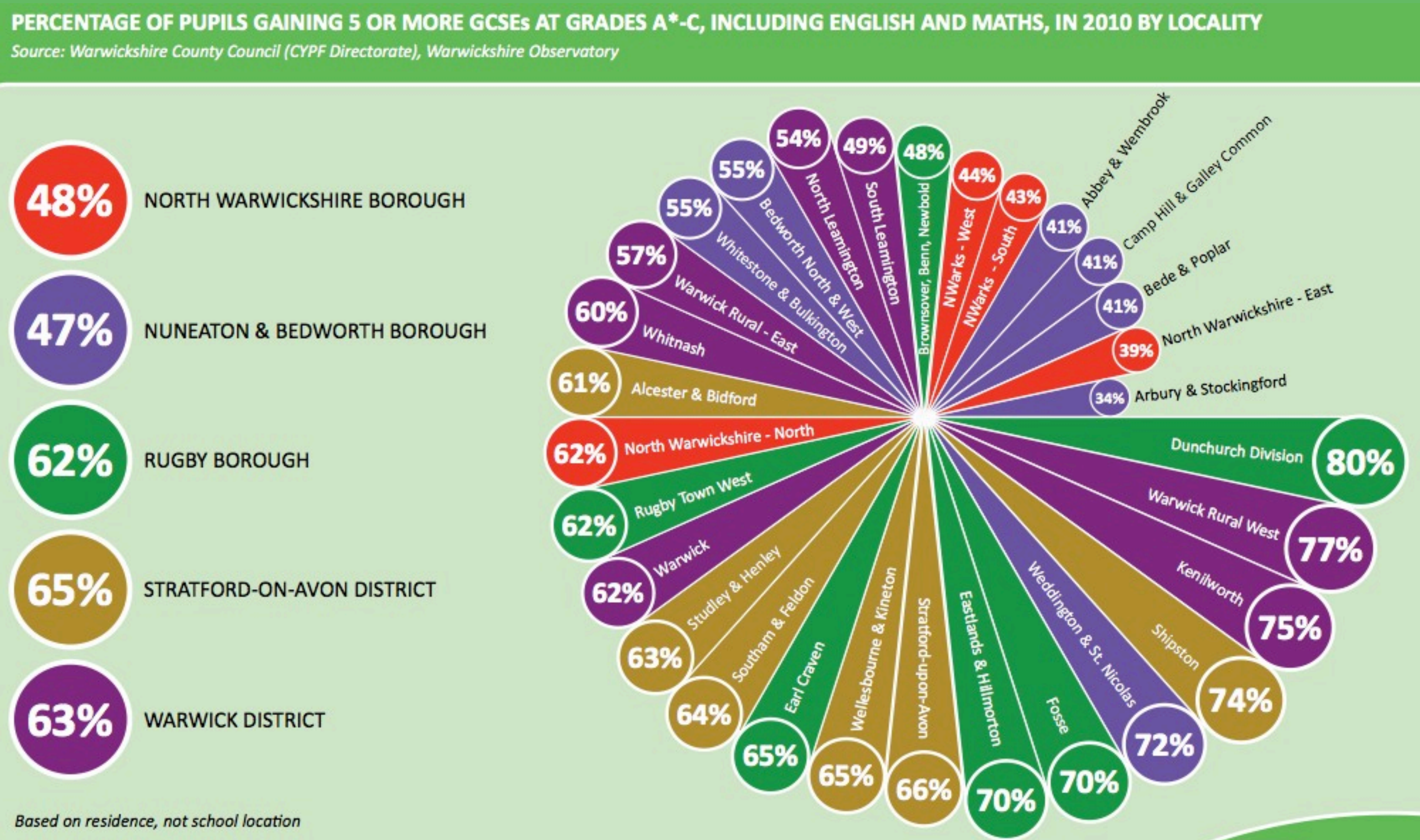
Tufte's design principles

- **maximize the data-ink ratio**
- **avoid chart junk (*sometimes*)**
- **use multifunctioning elements**
- **layer information**
- **maximize the data density**
  - shrink the graphics
  - maximize the amount of data shown (***sometimes***)

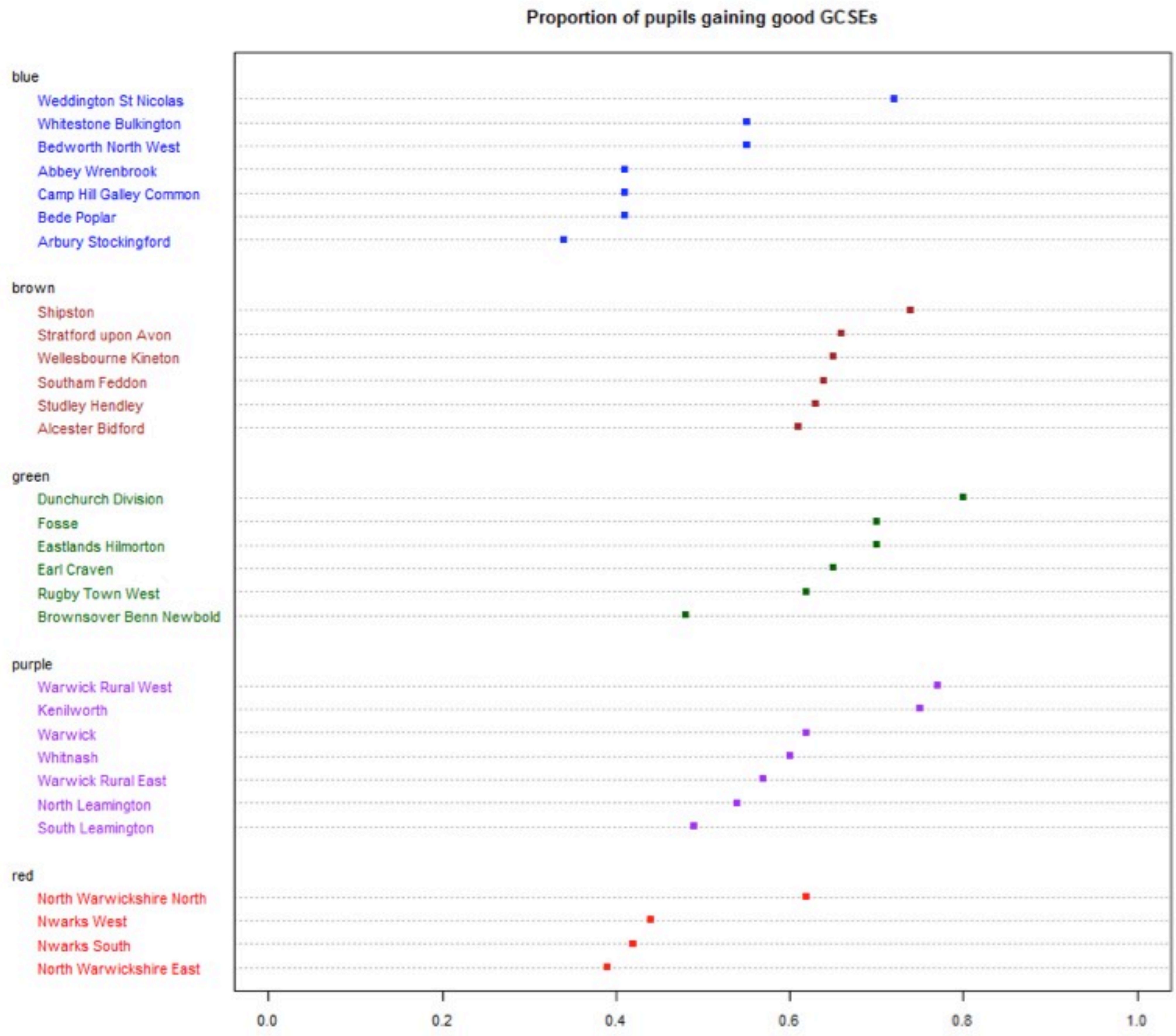
Williams's design principles

Contrast
Repetition
Alignment
Proximity

critique & redesign exercise . . .



http://junkcharts.typepad.com/junk_charts/2011/11/ornaments-or-fireworks-for-christmas.html



http://junkcharts.typepad.com/junk_charts/2011/11/ornaments-or-fireworks-for-christmas.html

- visualization design process
- types of research contributions

COMMENTS ON READINGS

- **visualization design process**
- **types of research contributions**

target

target a . . .

. . . specific domain

. . . set of users

learn about goals of users and kinds of data

acquire and clean data

How do you select and foster a good collaboration?

target

↓

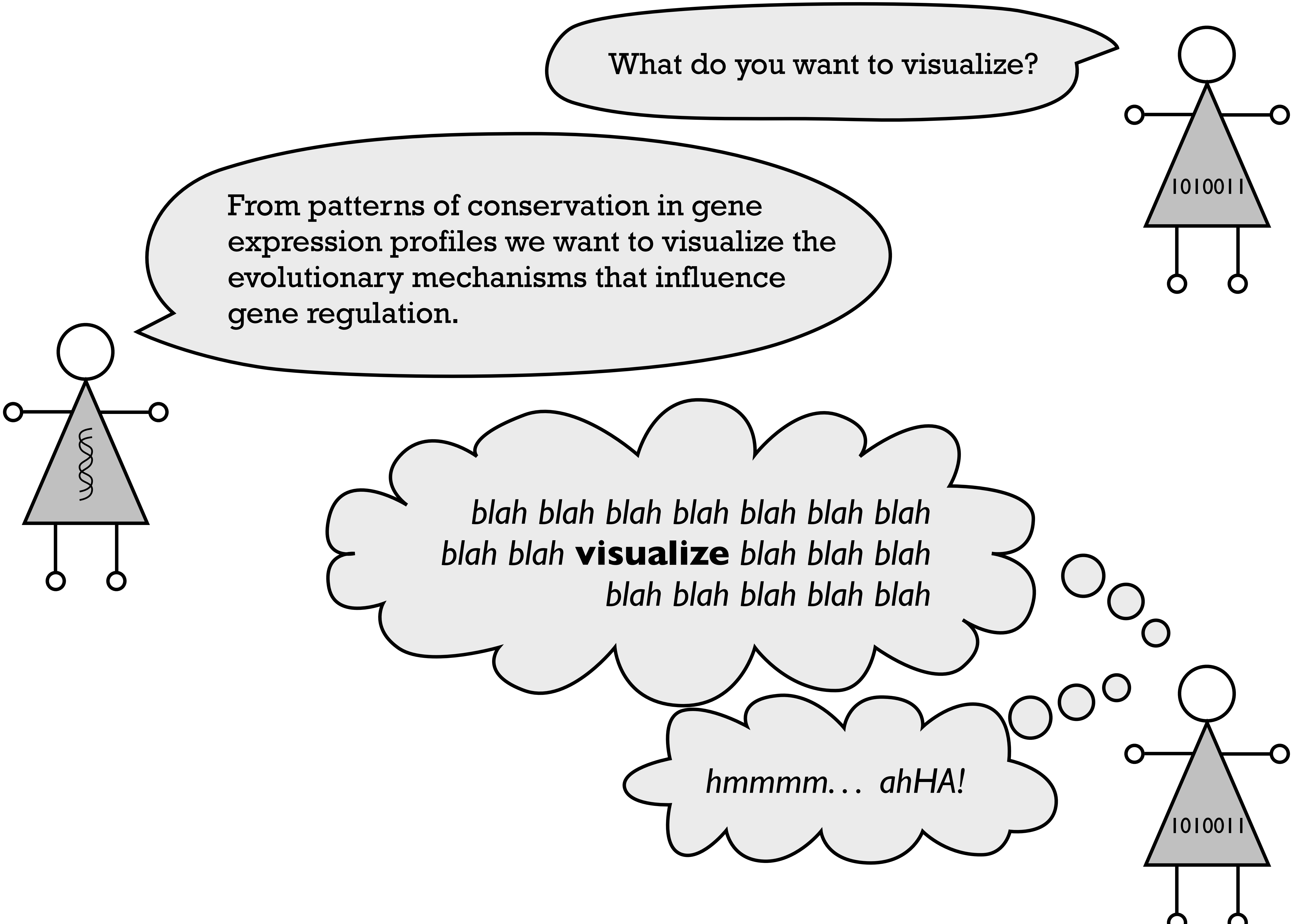
translate

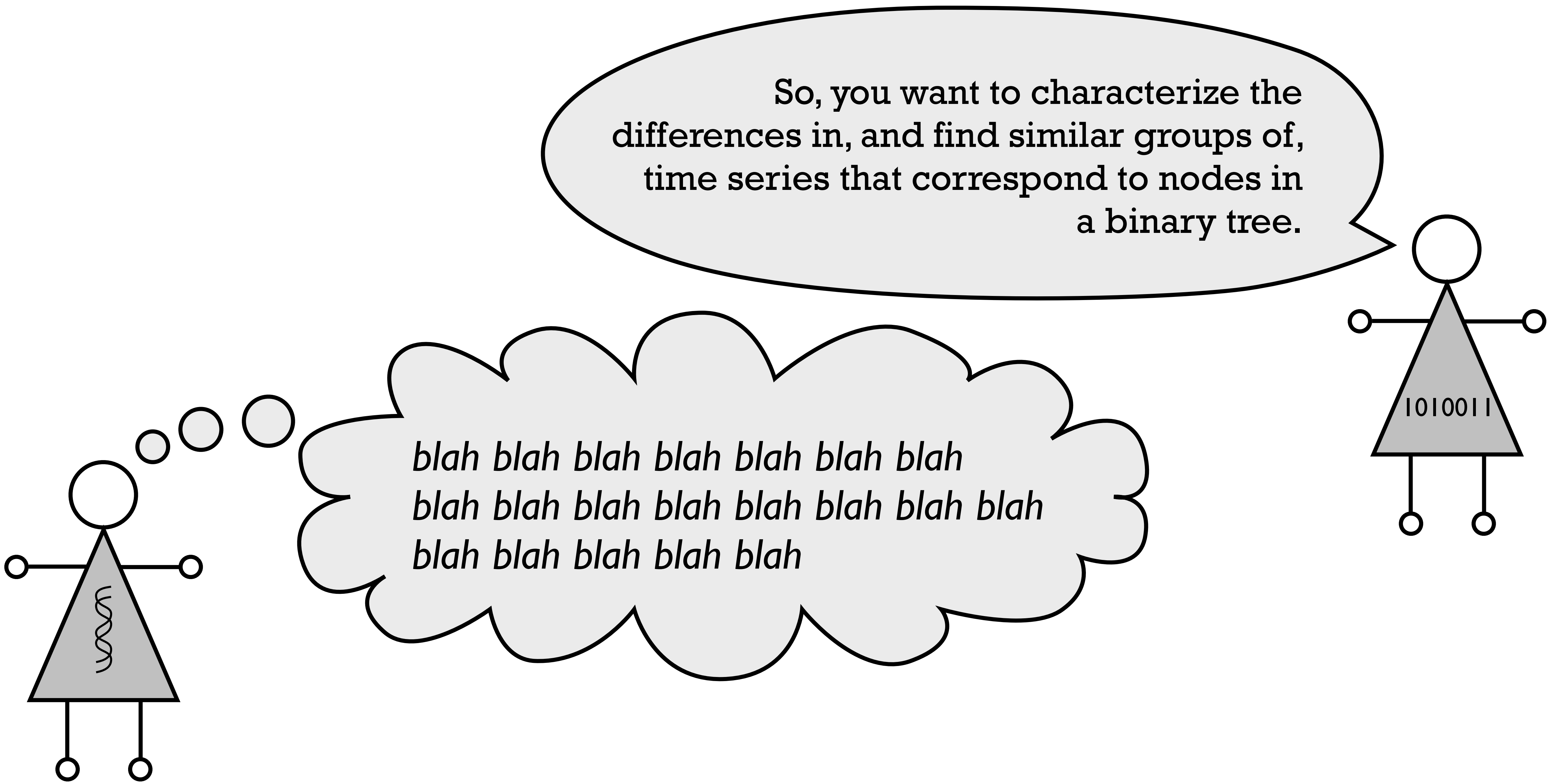
translate goals into data analysis tasks

structure and characterize data

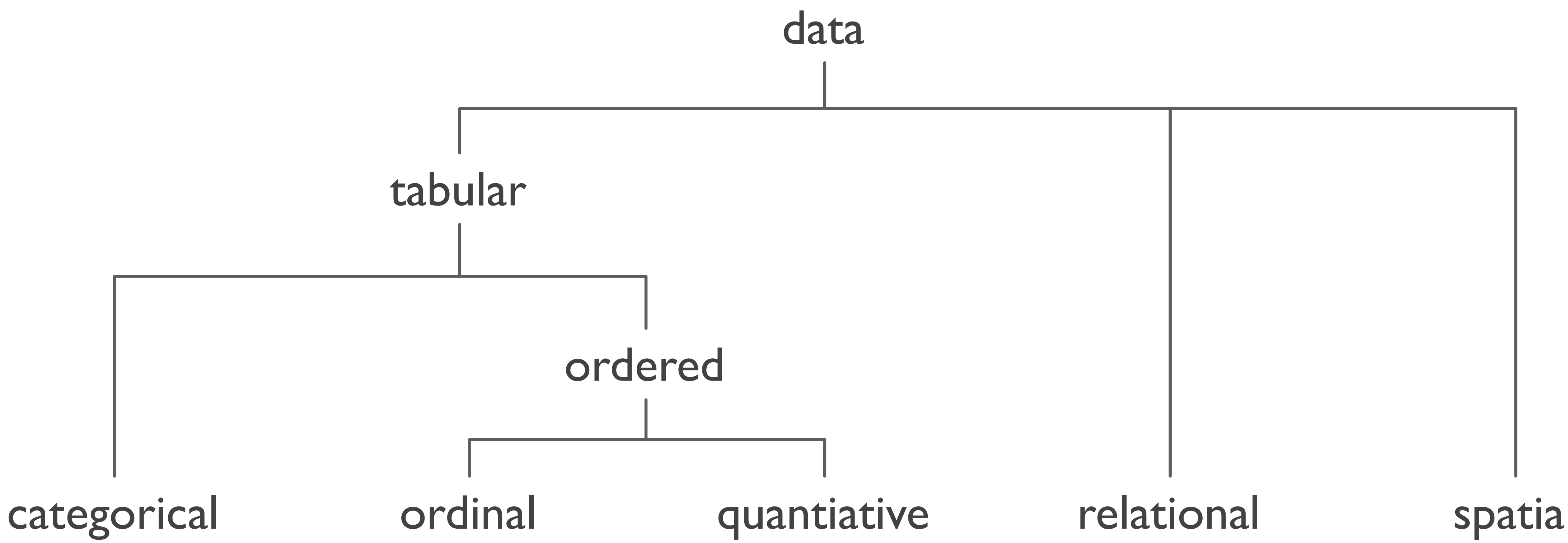
create abstraction of problem

from The Far Side by Gary Larsen

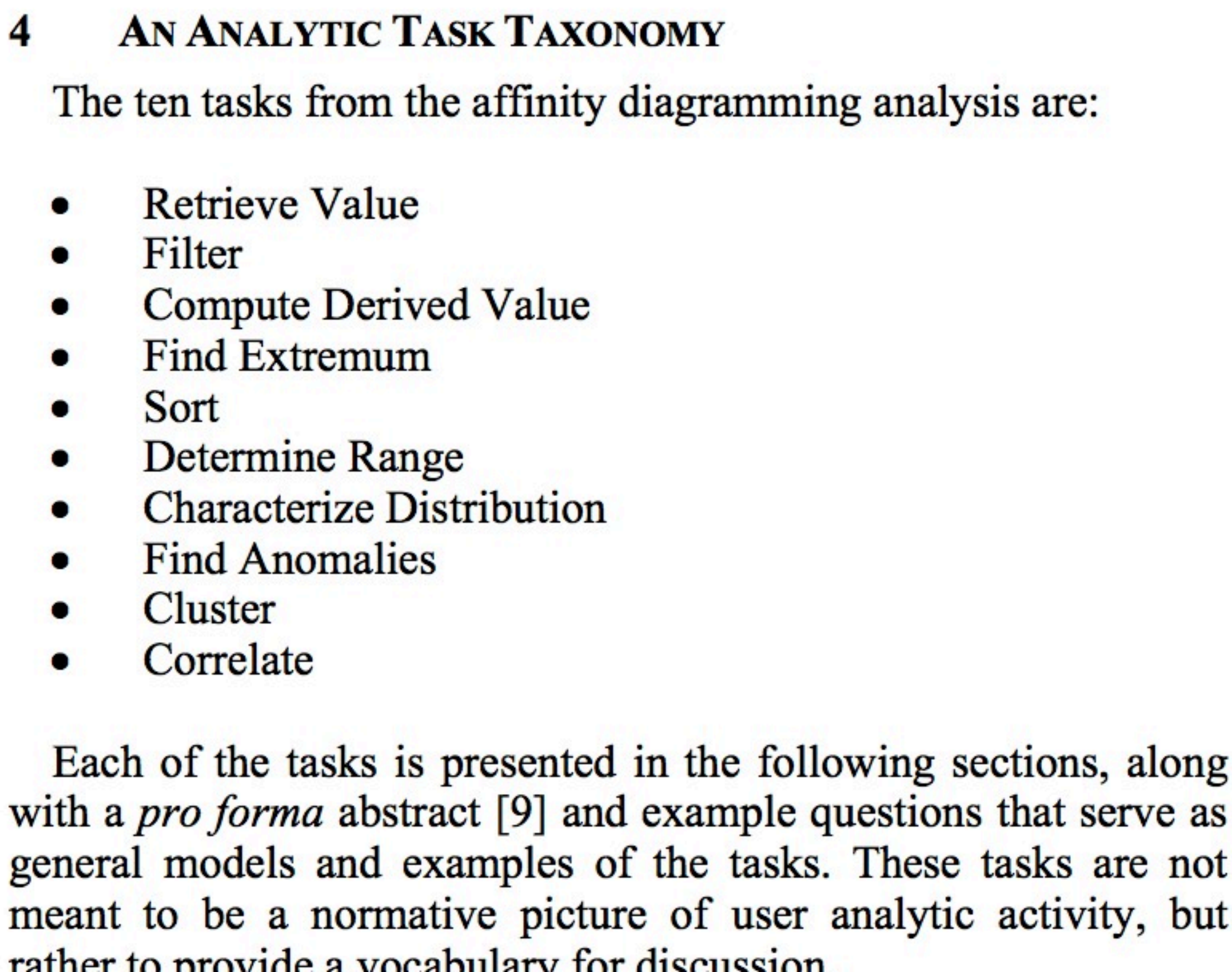

What do you want to visualize?
1010011
From patterns of conservation in gene expression profiles we want to visualize the evolutionary mechanisms that influence gene regulation.
blah blah blah blah blah blah blah blah blah **visualize** blah blah blah blah blah blah blah blah
hmmmm… ahHA!
1010011


So, you want to characterize the differences in, and find similar groups of, time series that correspond to nodes in a binary tree.
blah blah blah blah blah blah blah blah blah blah blah blah blah blah blah blah blah blah blah blah
1010011

this Thursday



next Thursday


4 AN ANALYTIC TASK TAXONOMY

The ten tasks from the affinity diagramming analysis are:

- Retrieve Value
- Filter
- Compute Derived Value
- Find Extremum
- Sort
- Determine Range
- Characterize Distribution
- Find Anomalies
- Cluster
- Correlate

Each of the tasks is presented in the following sections, along with a *pro forma* abstract [9] and example questions that serve as general models and examples of the tasks. These tasks are not meant to be a normative picture of user analytic activity, but rather to provide a vocabulary for discussion.

Amar et al. 2005

target

translate

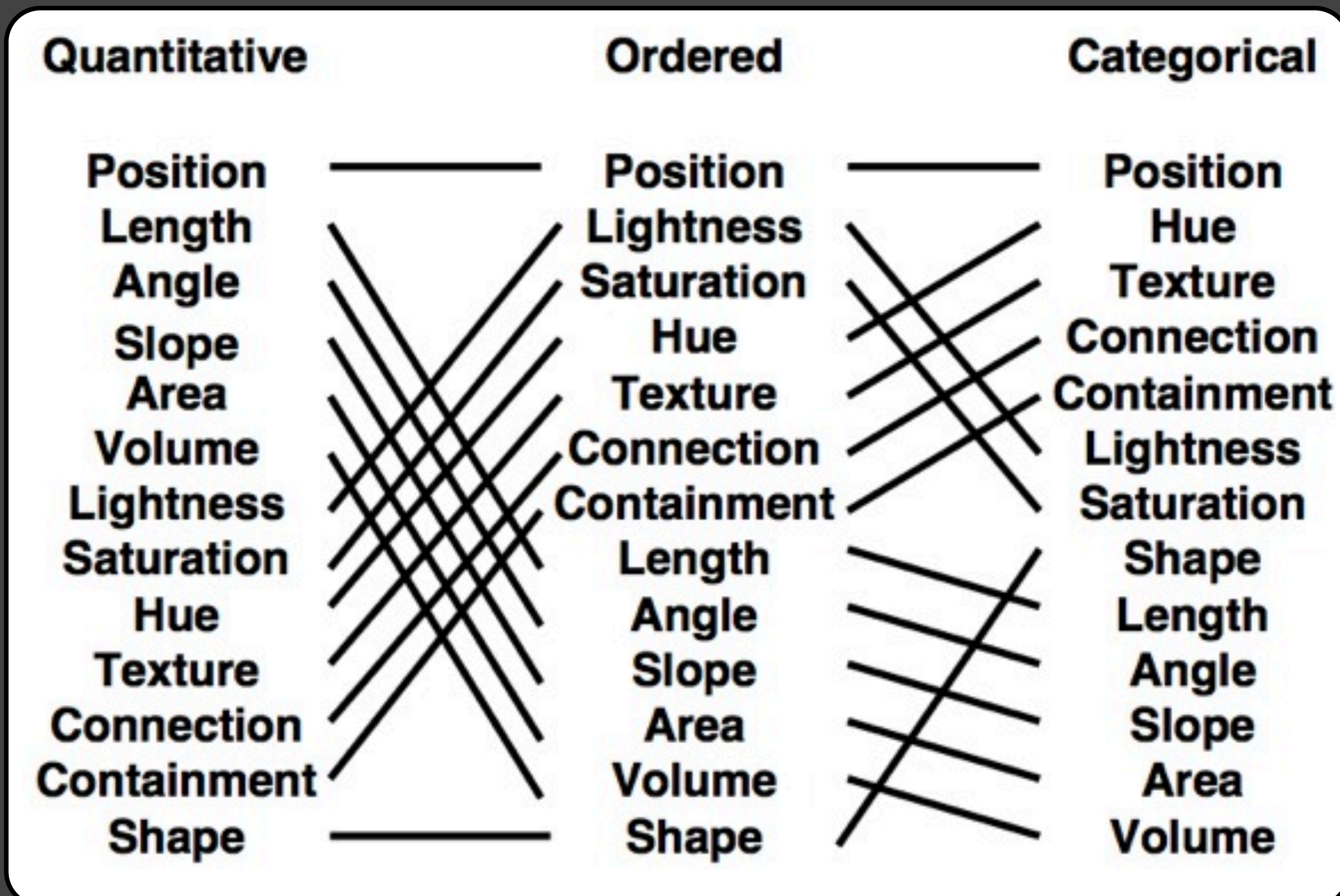
design

design visual encodings and interaction mechanisms to support data and task abstraction

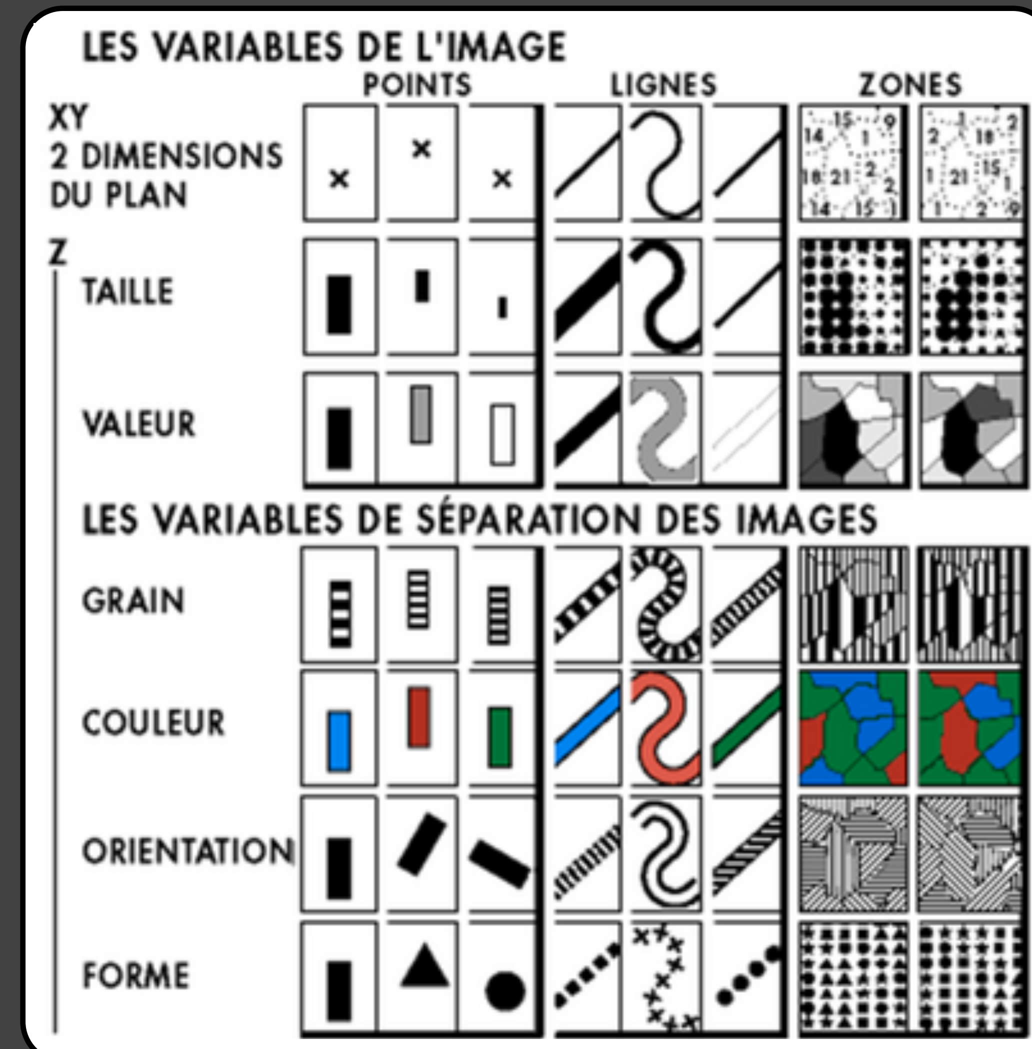
transform data with appropriate computational methods

try many ideas!

next Tuesday

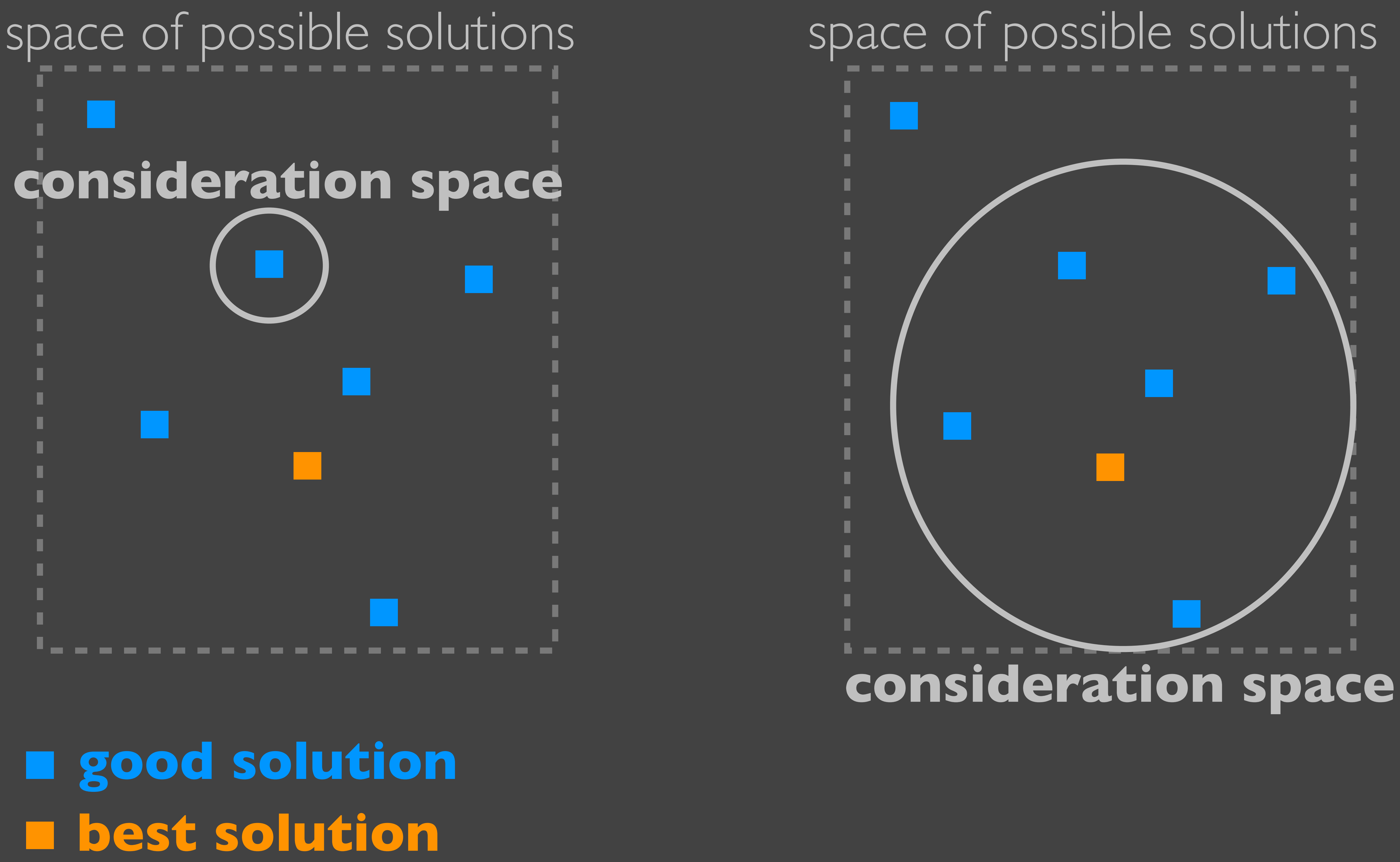


Automating the Design of Graphical Presentations of Relational Information
MacKinlay, 1986



Semiology of Graphics
Bertin, 1967

try many ideas



RAPID PROTOTYPING

Michael Sedlmair

ideation exercise . . .

target

↓

translate

↓

design

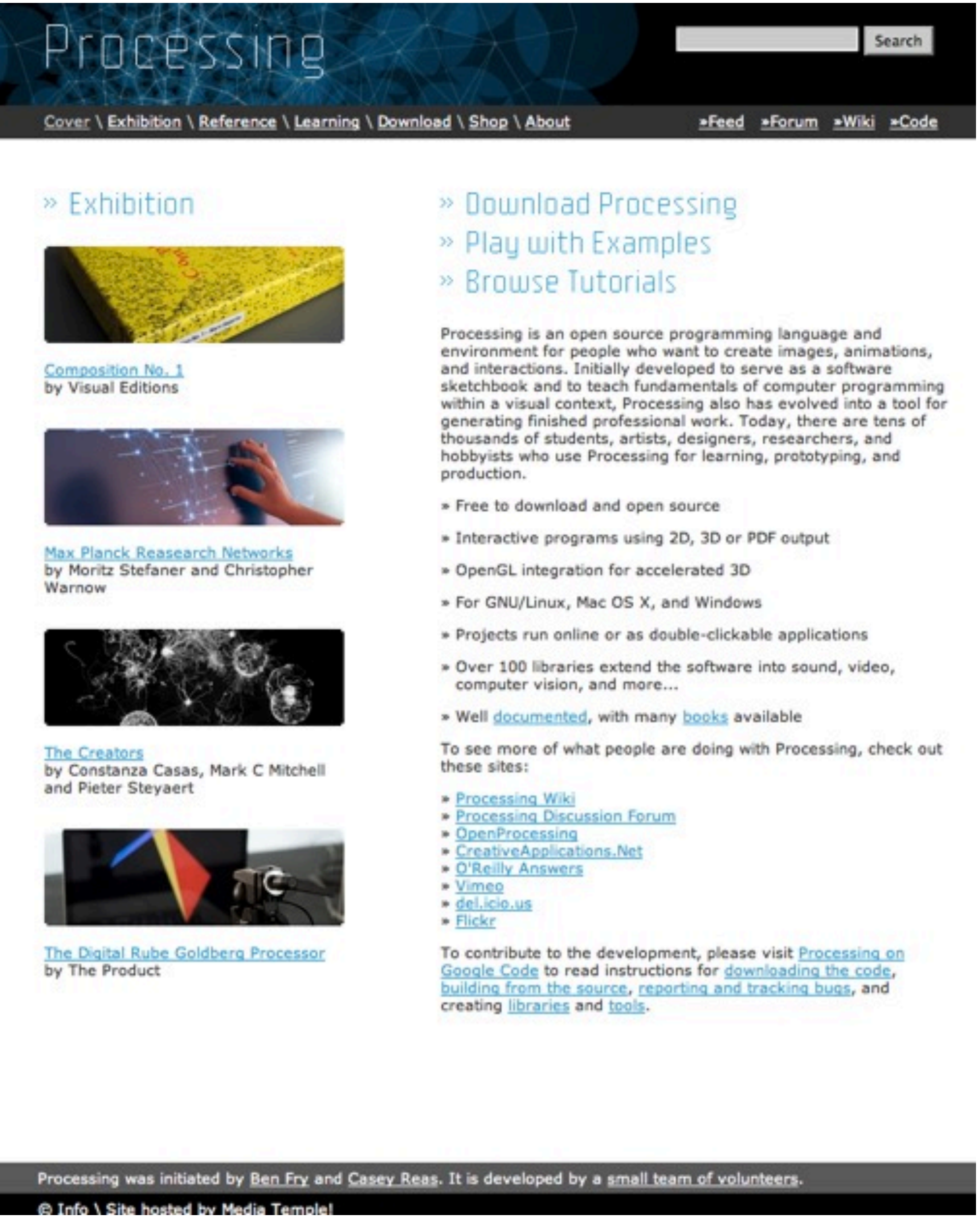
↓

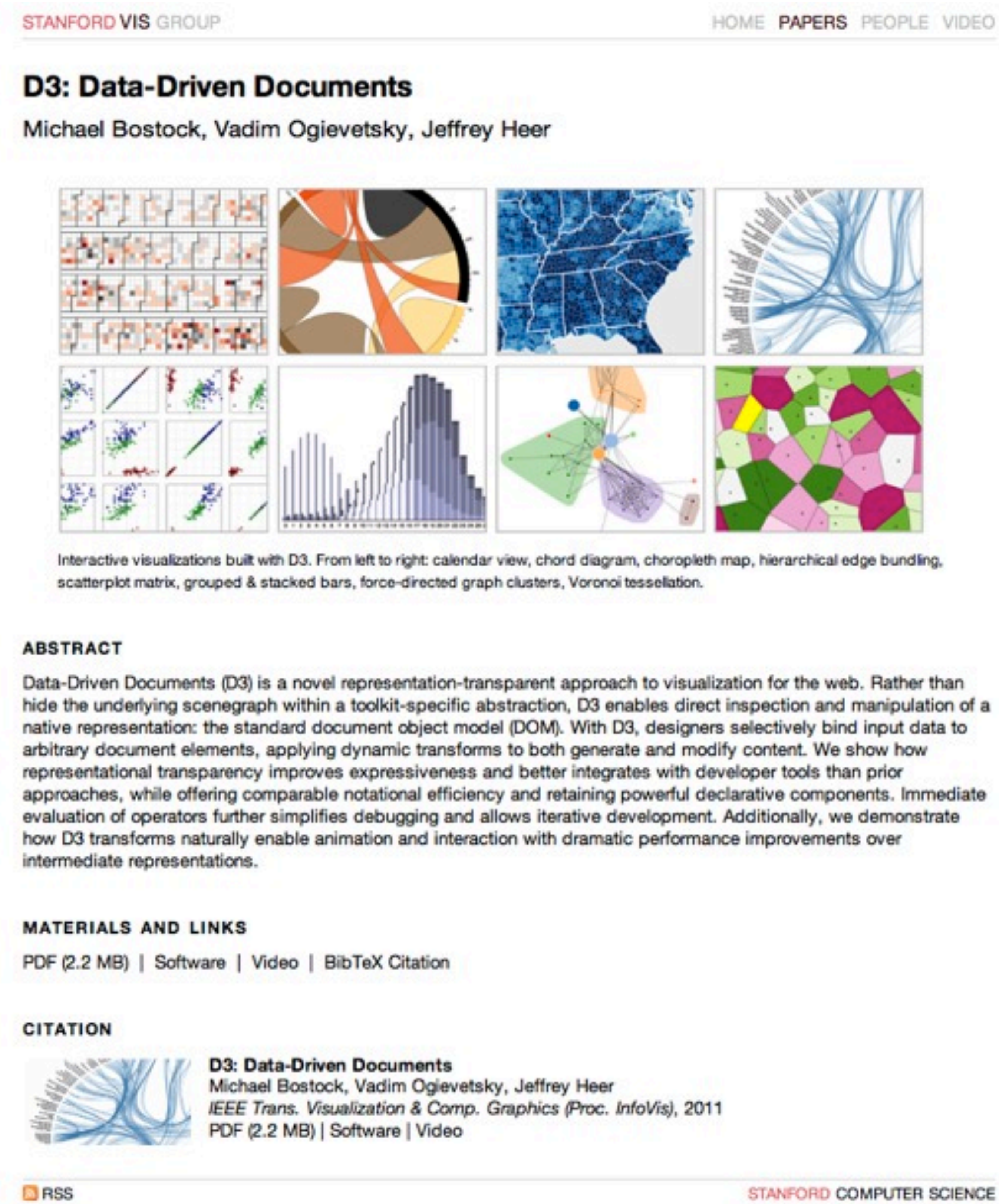
implement

bring your design to life

optimize algorithms

in early March





http://processing.org/

http://vis.stanford.edu/papers/d3

target

translate

problem characterization
and abstraction

80%

design

visualization design

20%

implement

target

↓

translate

↓

design

↓

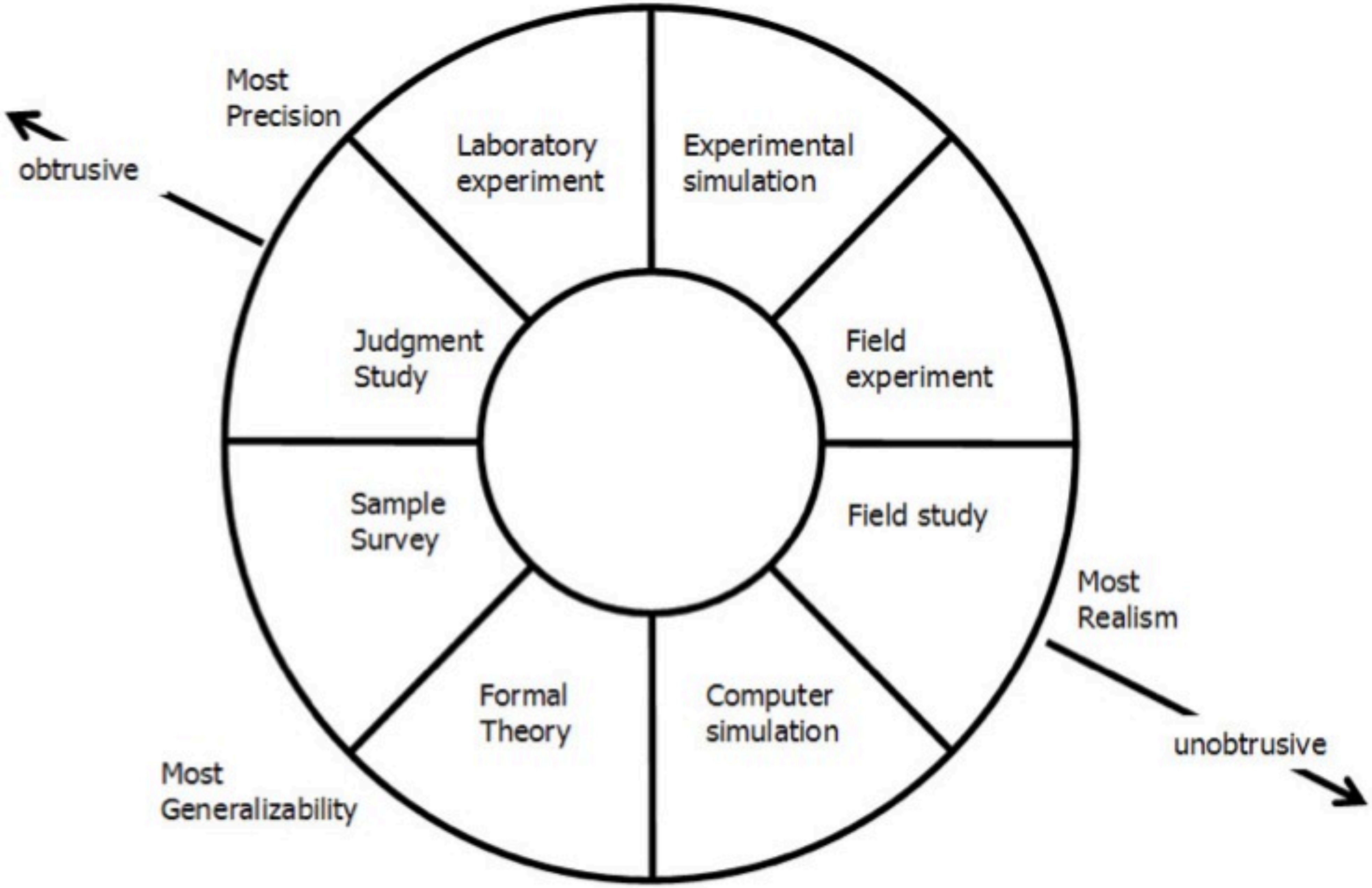
implement

↓

validate

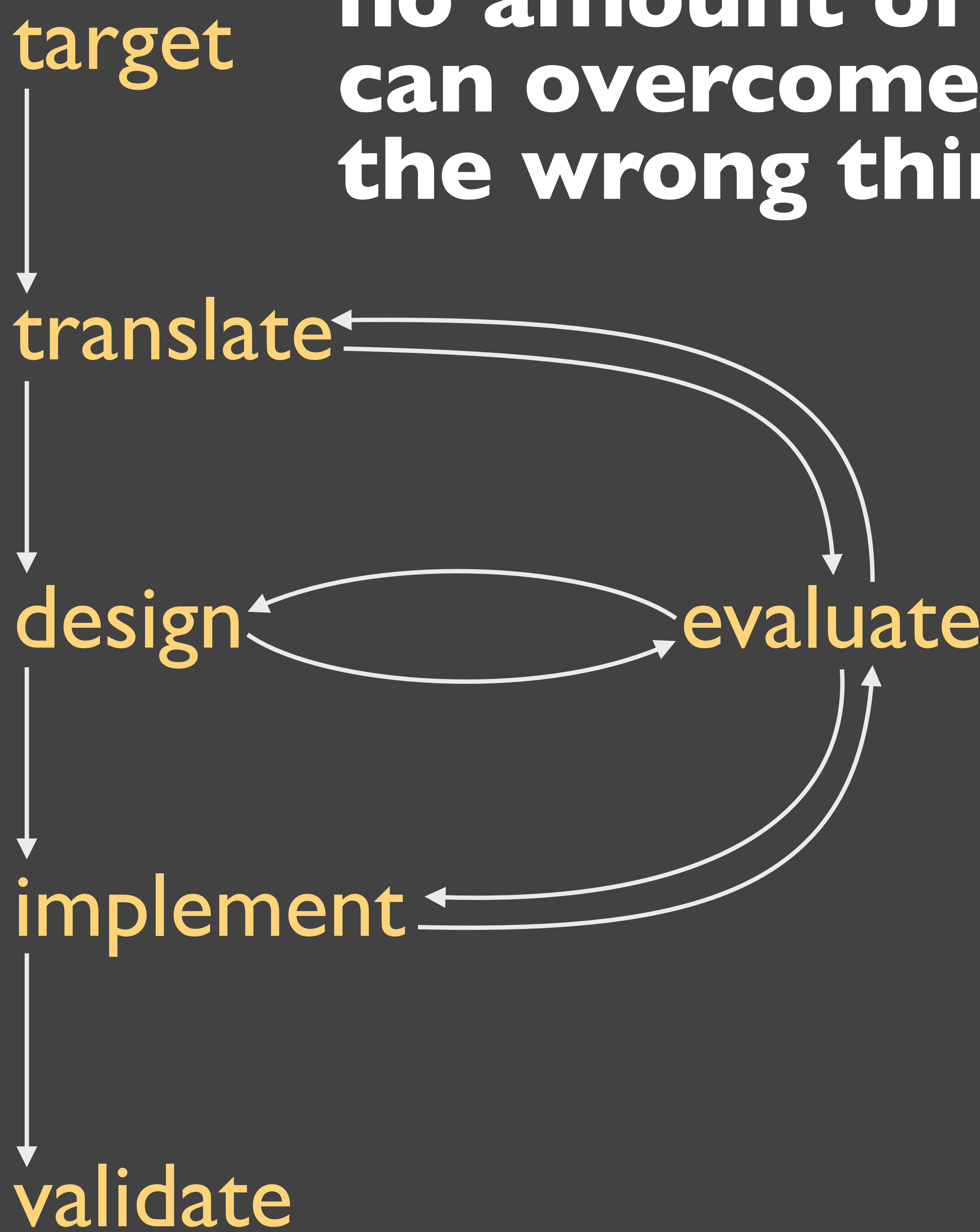
ensure visualization system supports target users' goals

in late March

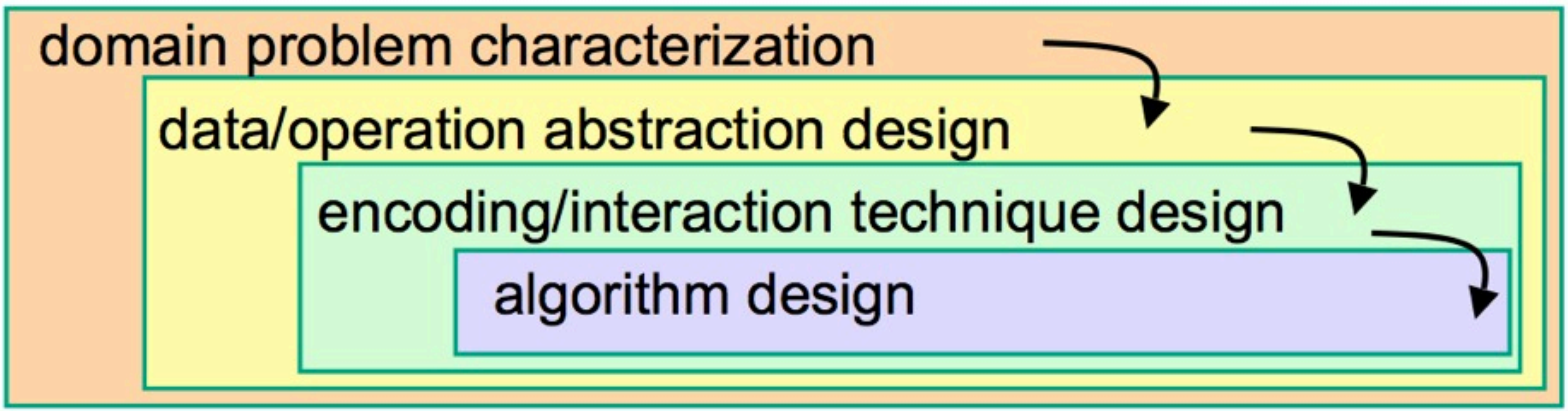


Carpendale 2008

no amount of brilliant design can overcome designing for the wrong thing

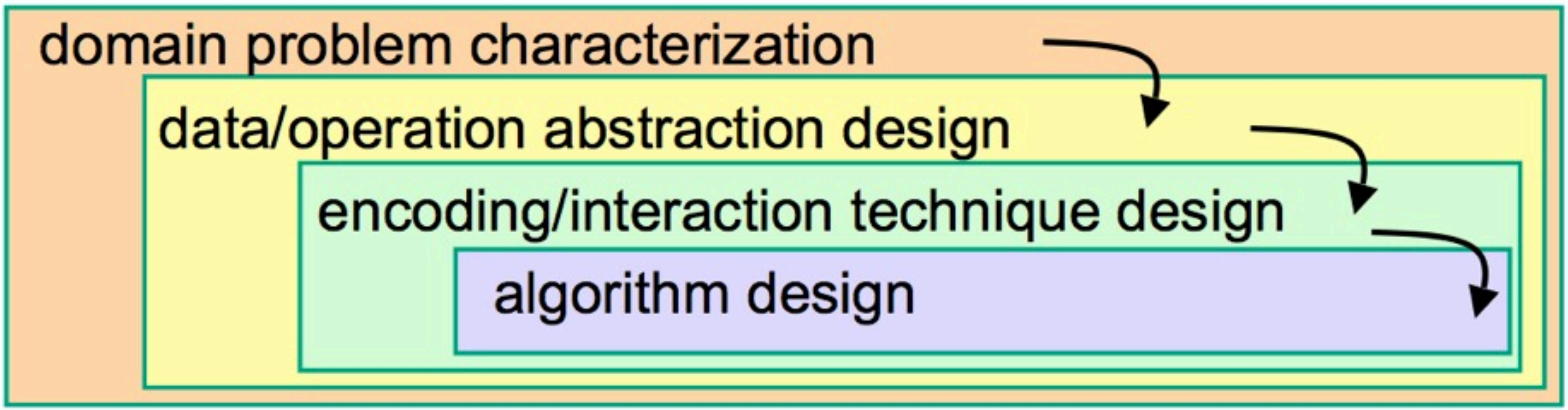


NESTED MODEL



what can go wrong, and how do we validate?

Munzner 2009

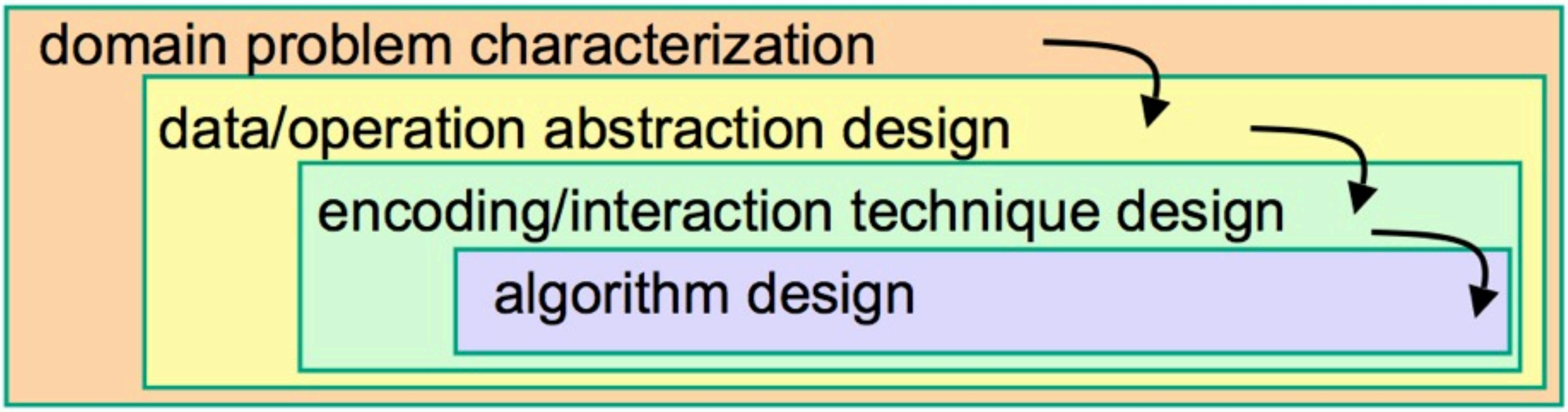


problem

threat: they don't do that

validation: *immediate*, observe and interview users
downstream, notice adoption rates

Munzner 2009

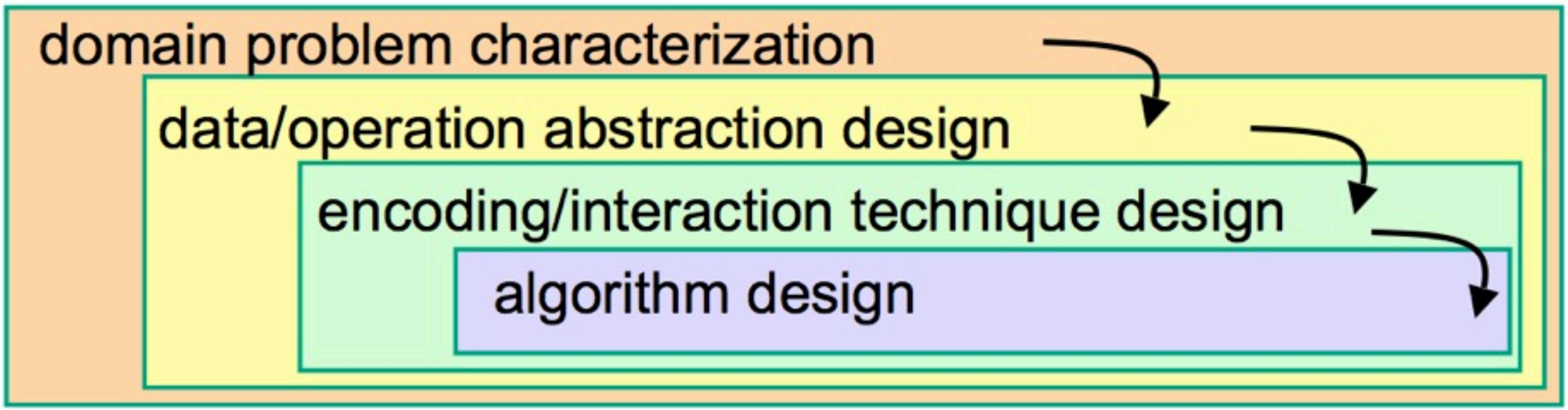


abstraction

threat: you're showing them the wrong thing

validation: *downstream,* deploy and observe usage

Munzner 2009



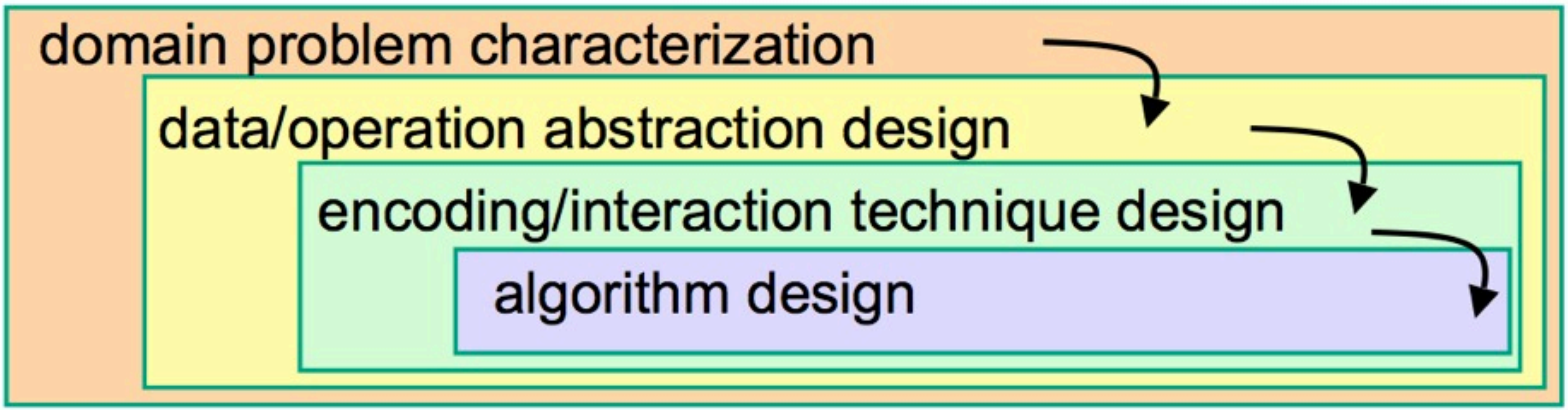
design

threat: they way you show it doesn't work

validation: *immediate*, justification with known principles

downstream, qualitative or quantitative analysis of results; lab study measuring time and error

Munzner 2009



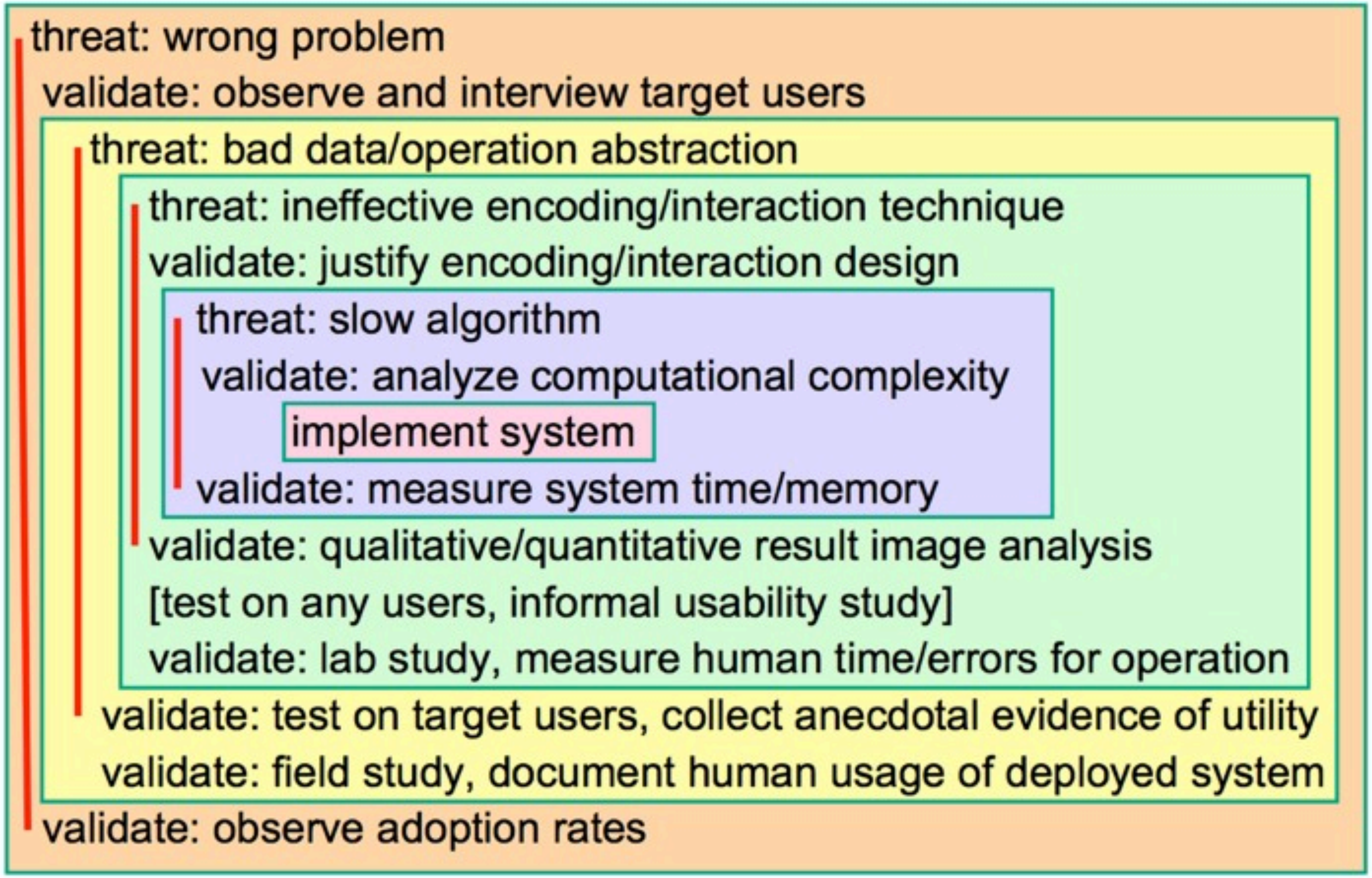
algorithm

threat: you're code is too slow

validation: *immediate*, complexity analysis

downstream, benchmarks for system time and memory

Munzner 2009

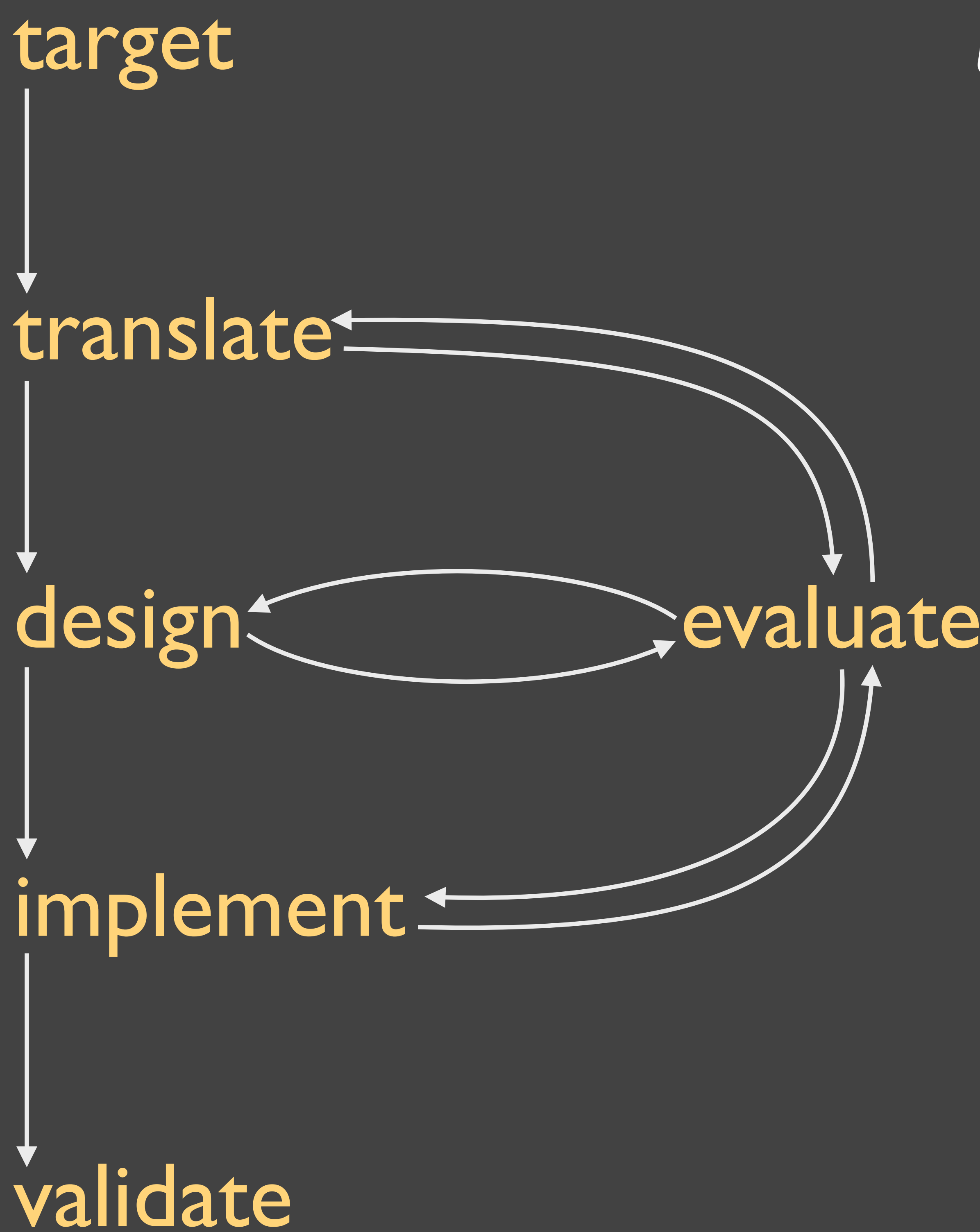


avoid validation mismatch

-cannot validate encoding with system timings

-cannot validate abstraction with lab studies

Munzner 2009

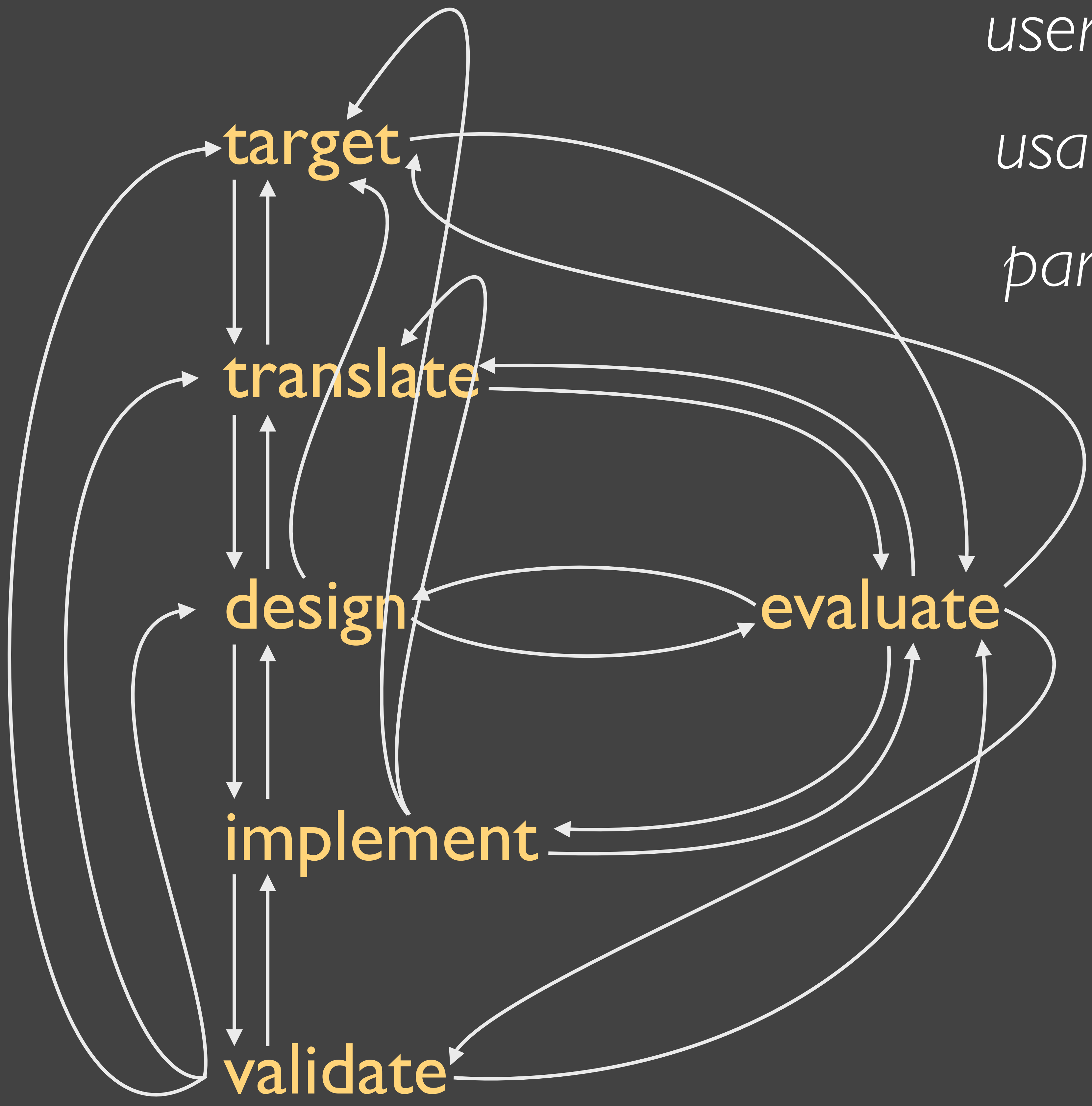


user-centered design

usability engineering

participatory design

design thinking

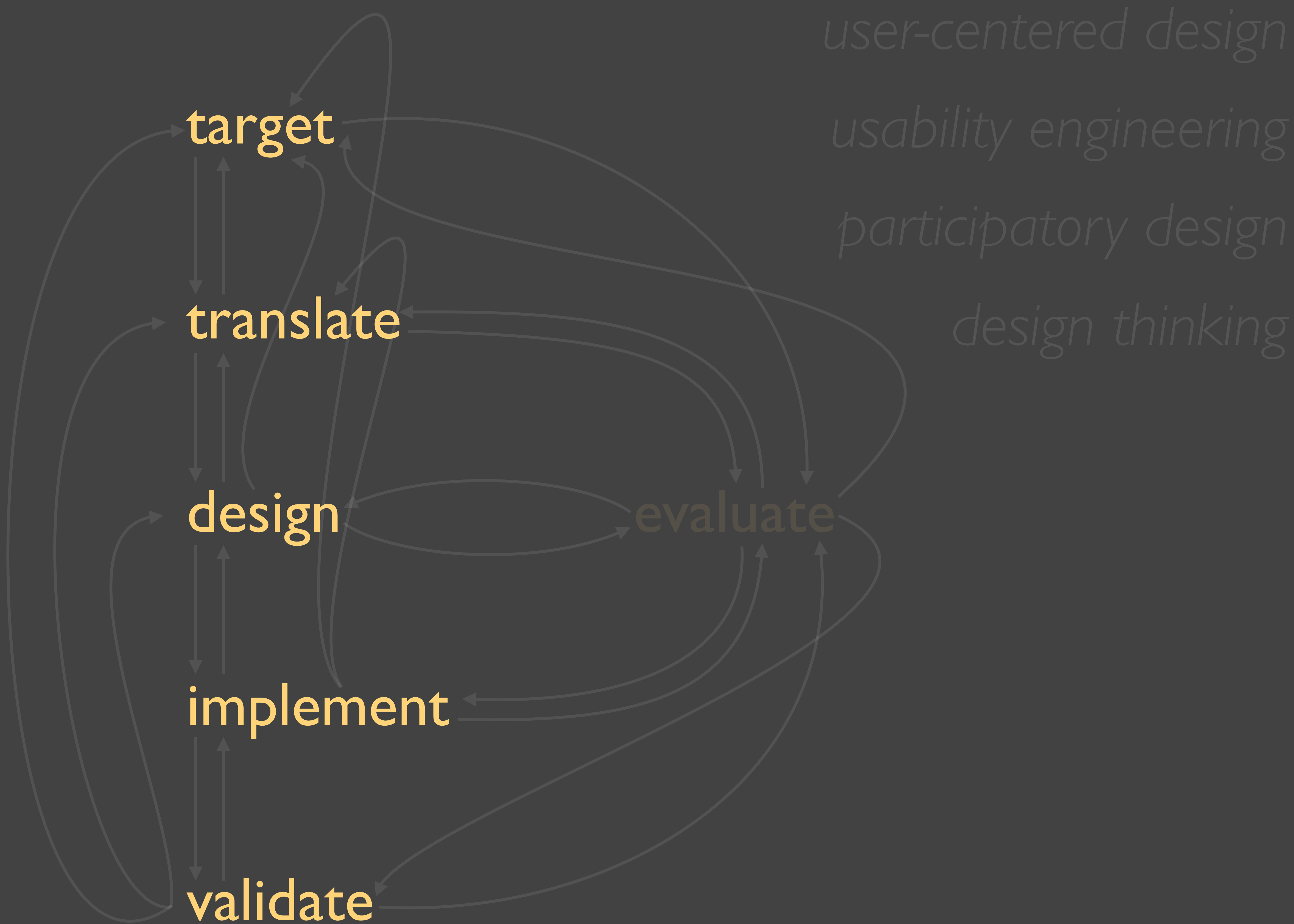


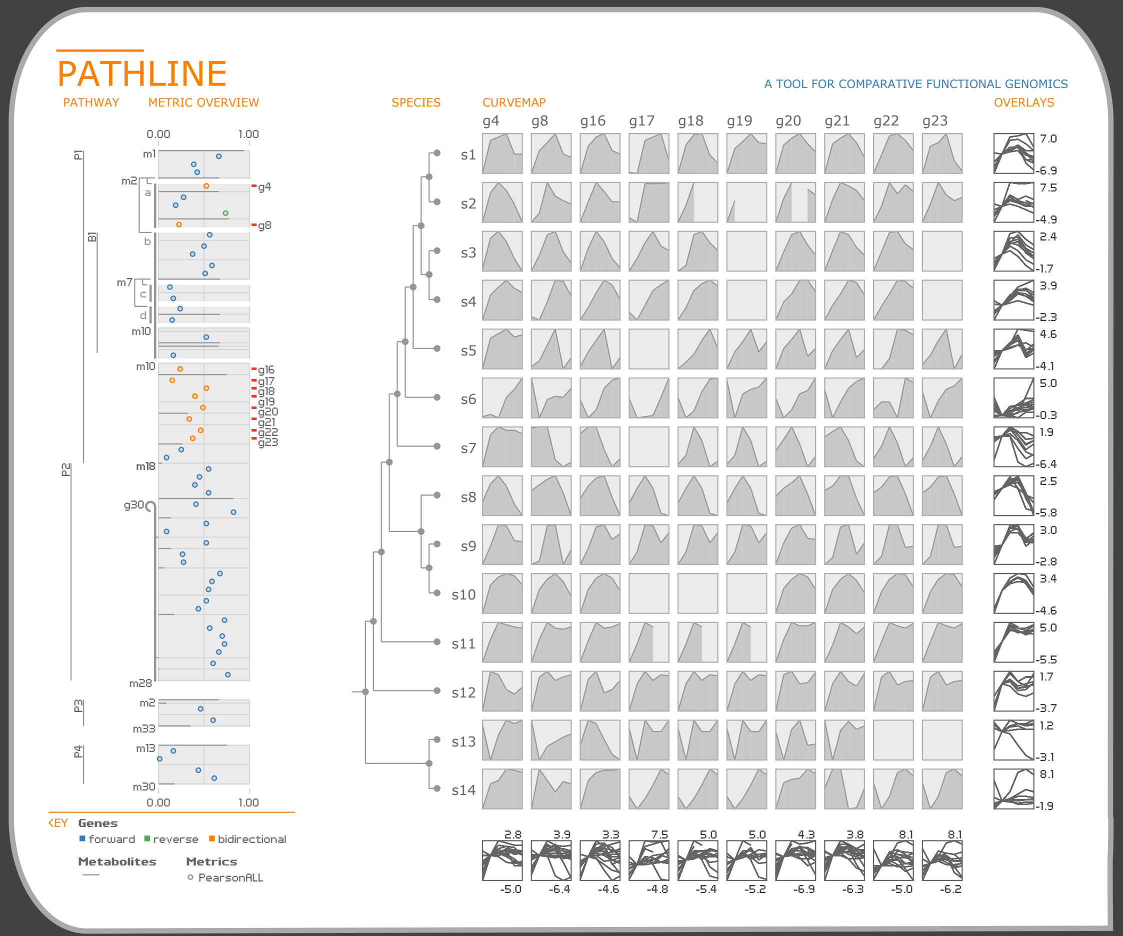
user-centered design

usability engineering

participatory design

design thinking


target
translate
design
evaluate
implement
validate
user-centered design
usability engineering
participatory design
design thinking



Pathline

A Tool for Comparative Functional Genomics Data

joint work with:
Bang Wong, Mark Styczynski, Tamara Munzner, Hanspeter Pfister

Pathline: A Tool for Comparative Functional Genomics
M. Meyer et al., IEEE/Eurographics EuroVis 2010.

target

translate

design

implement

validate

functional genomics

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

functional genomics data

gene expression

molecular pathways

gene expression is ...

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

... a single quantitative value

biologists measure it ...

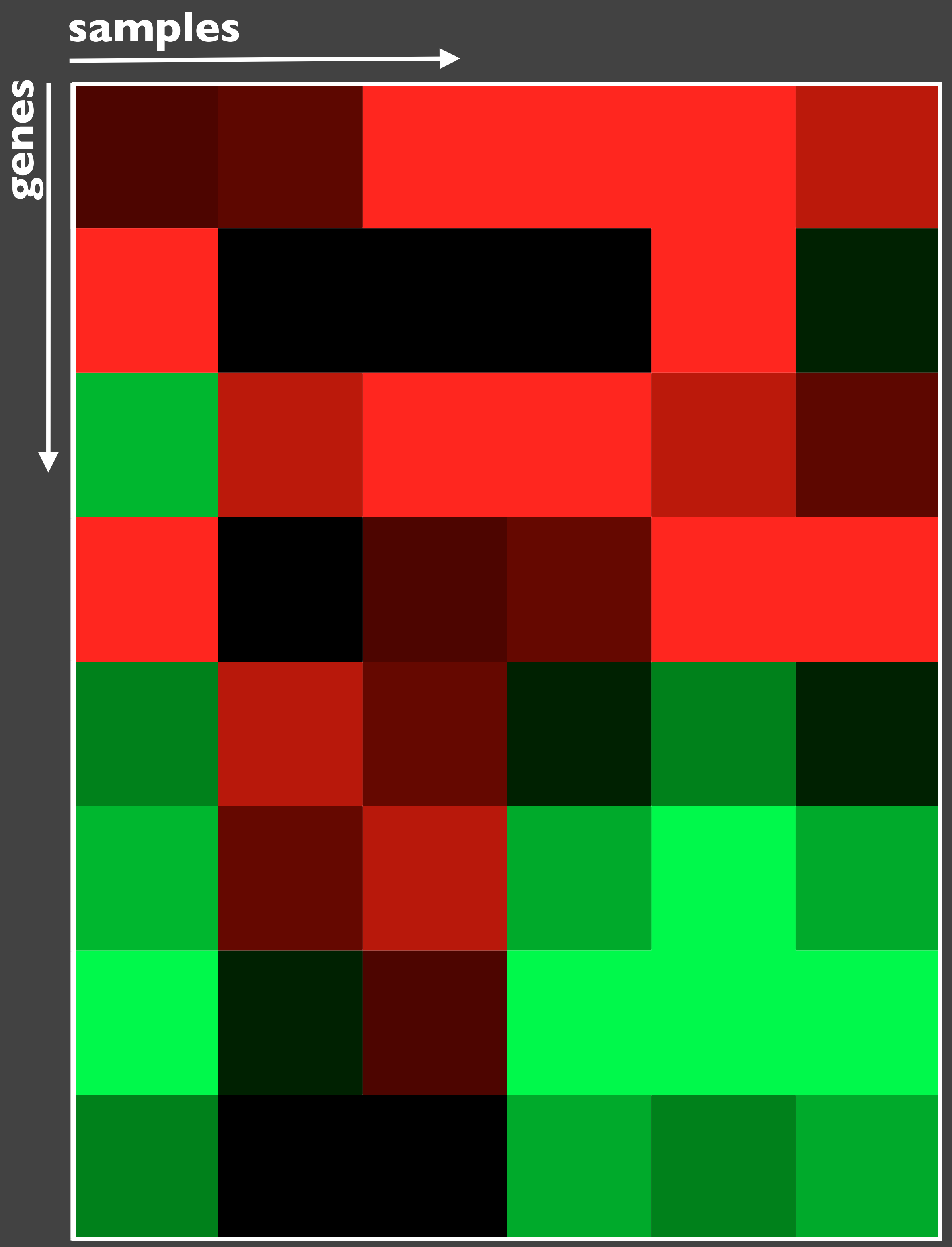
... for many genes

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

visualized with heatmaps

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

encode value with color



gene expression is …

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

… a single quantitative value

biologists measure it …

… for many genes

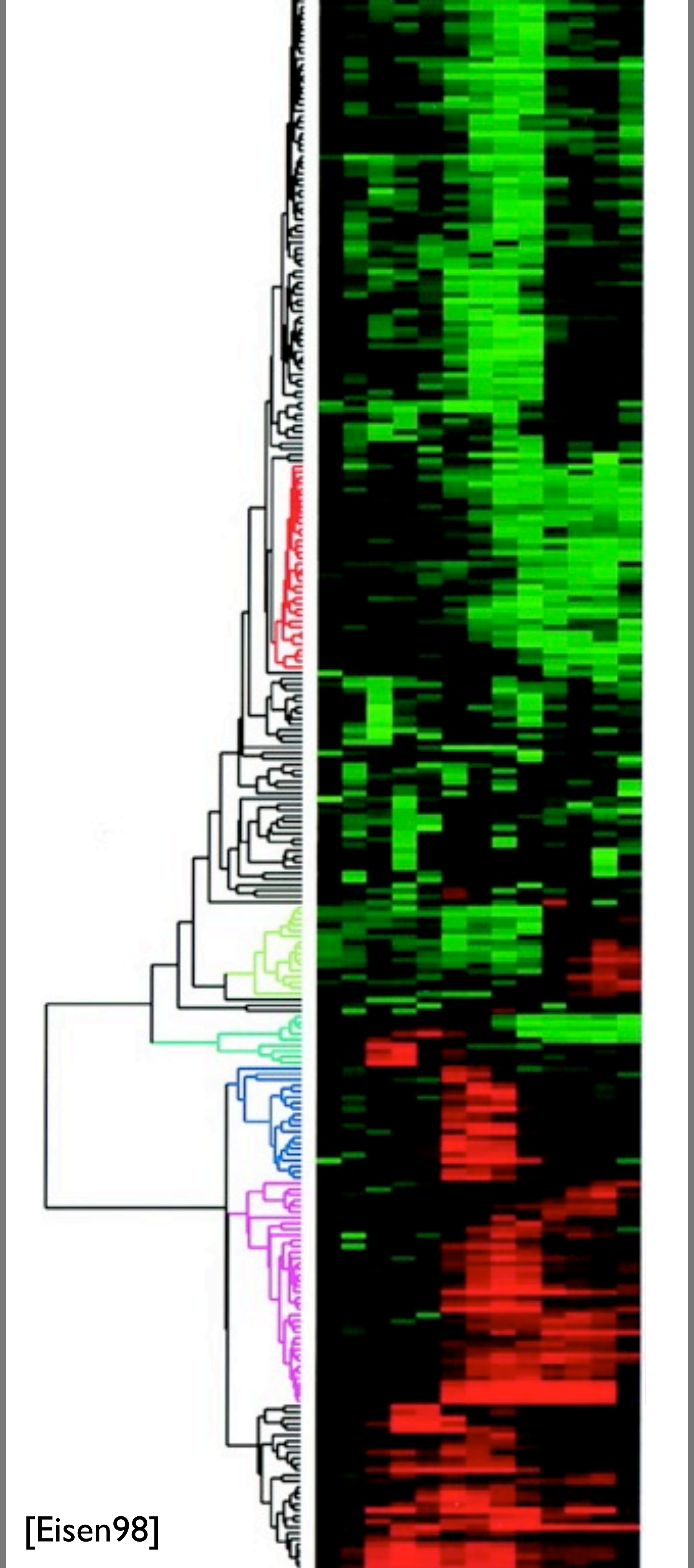
… in many samples (time points, tissue types, species)

visualized with heatmaps

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

encode value with color

augmented with clustering



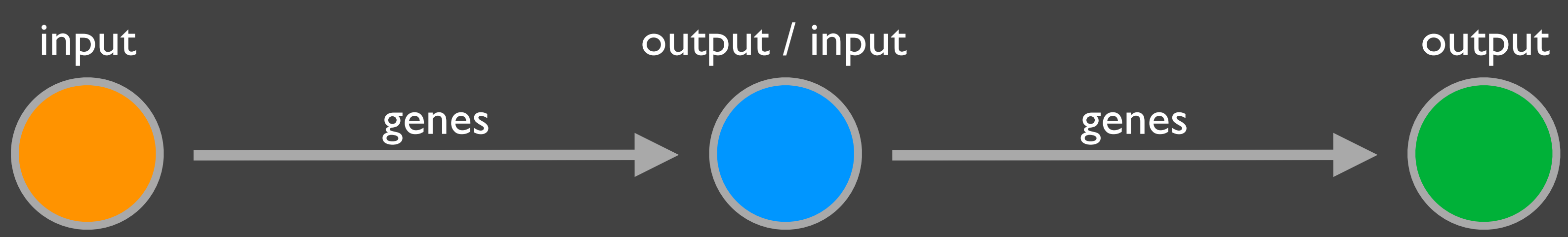
[Eisen98]

functional genomics data

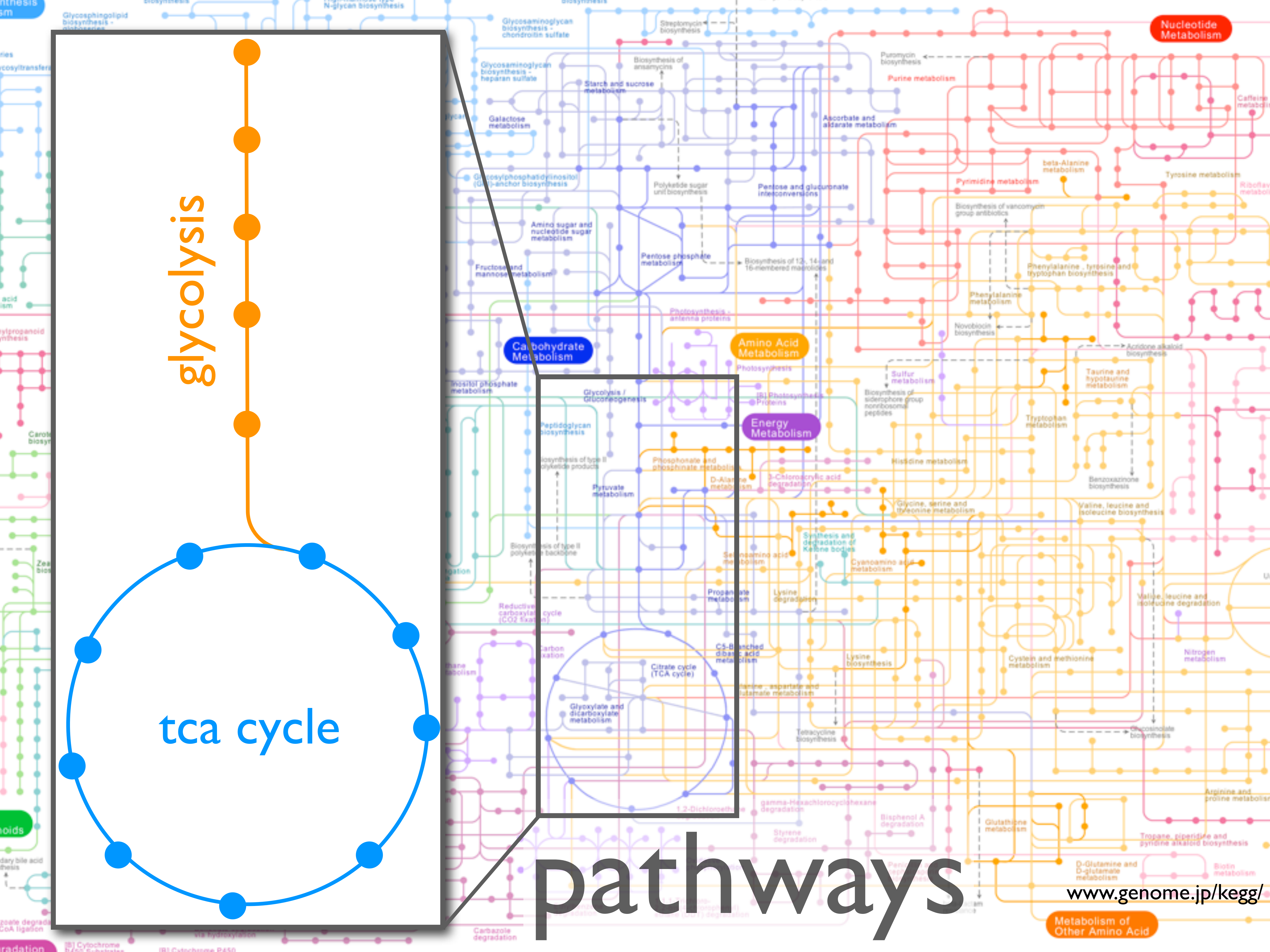
gene expression

molecular pathways

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



= cell function


glycolysis
tca cycle
Carbohydrate Metabolism
Amino Acid Metabolism
Energy Metabolism
Nucleotide Metabolism
Metabolism of Other Amino Acid
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

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?

target

translate

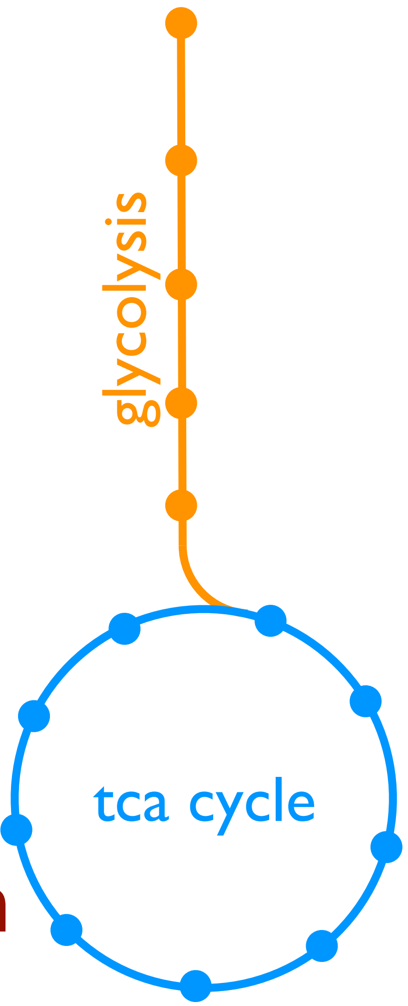
design

implement

validate

metabolic pathways

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

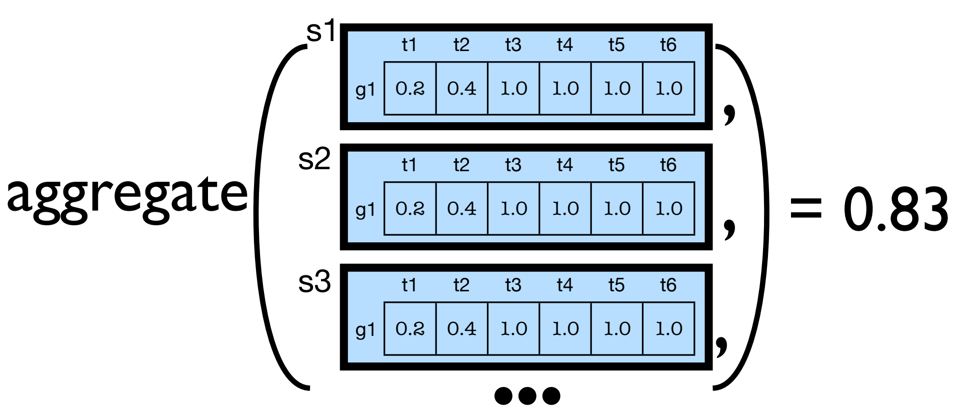


gene expression

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

s1	t1	t2	t3	t4	t5	t6
g1	0.2	0.4	1.0	1.0	1.0	1.0
m1	1.0	0.0	0.0	0.0	1.0	0.8
g2	-0.7	0.8	1.0	1.0	0.8	0.2
m2	1.0	0.0	0.2	0.5	1.0	0.2
g3	-0.5	0.8	0.5	-0.3	-0.5	-0.5
m3	-0.7	0.5	0.8	-0.7	-1.0	0.5

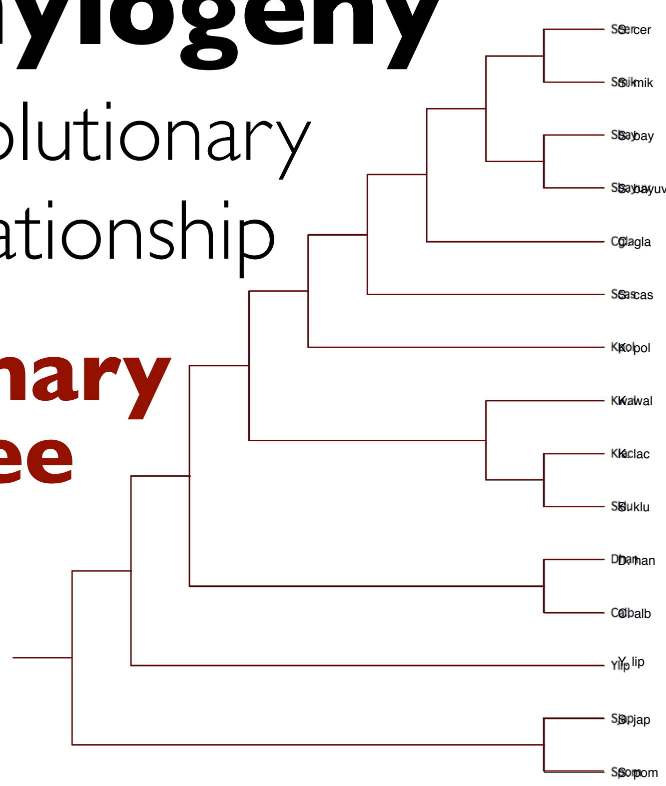
similarity scores



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

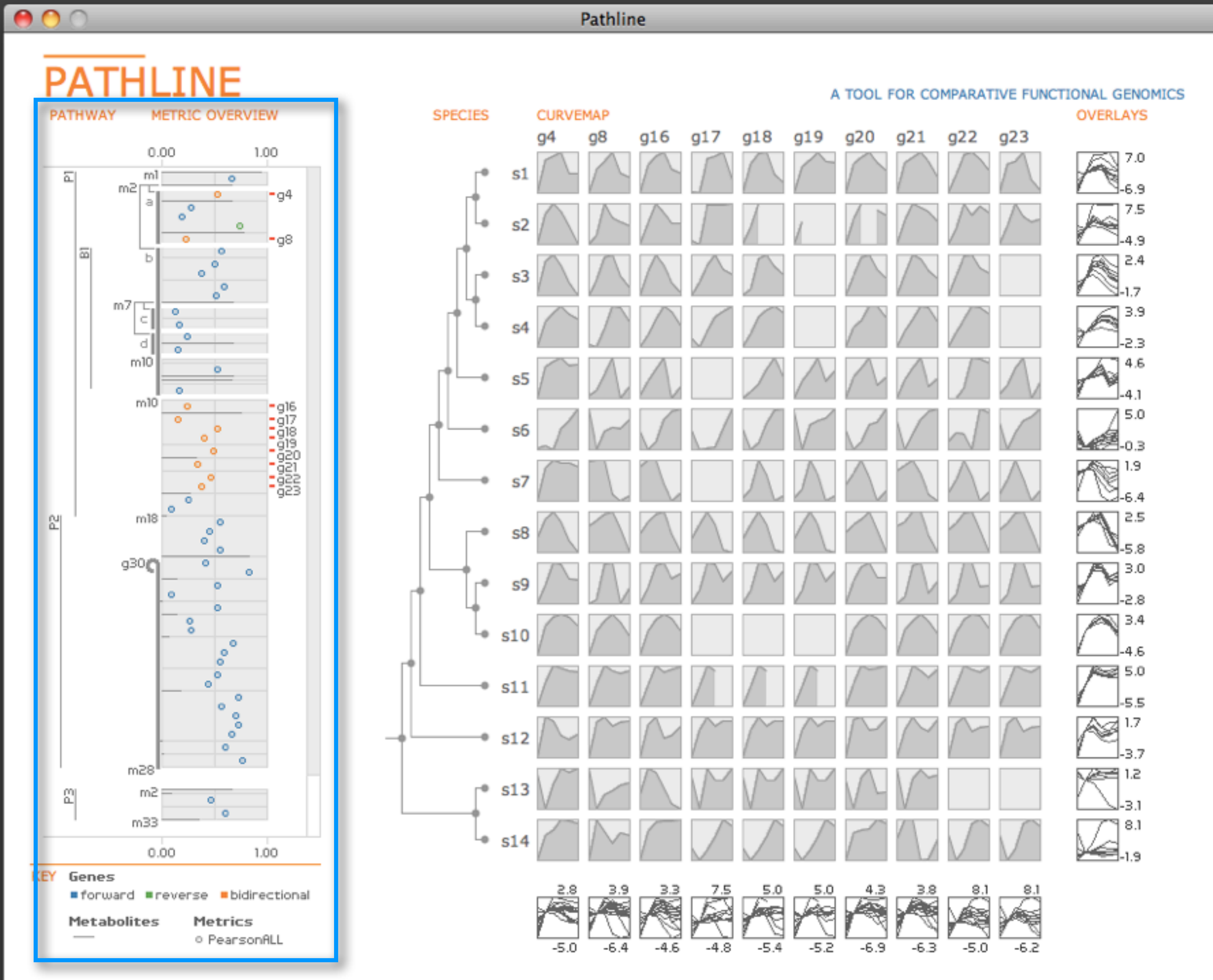
phylogeny

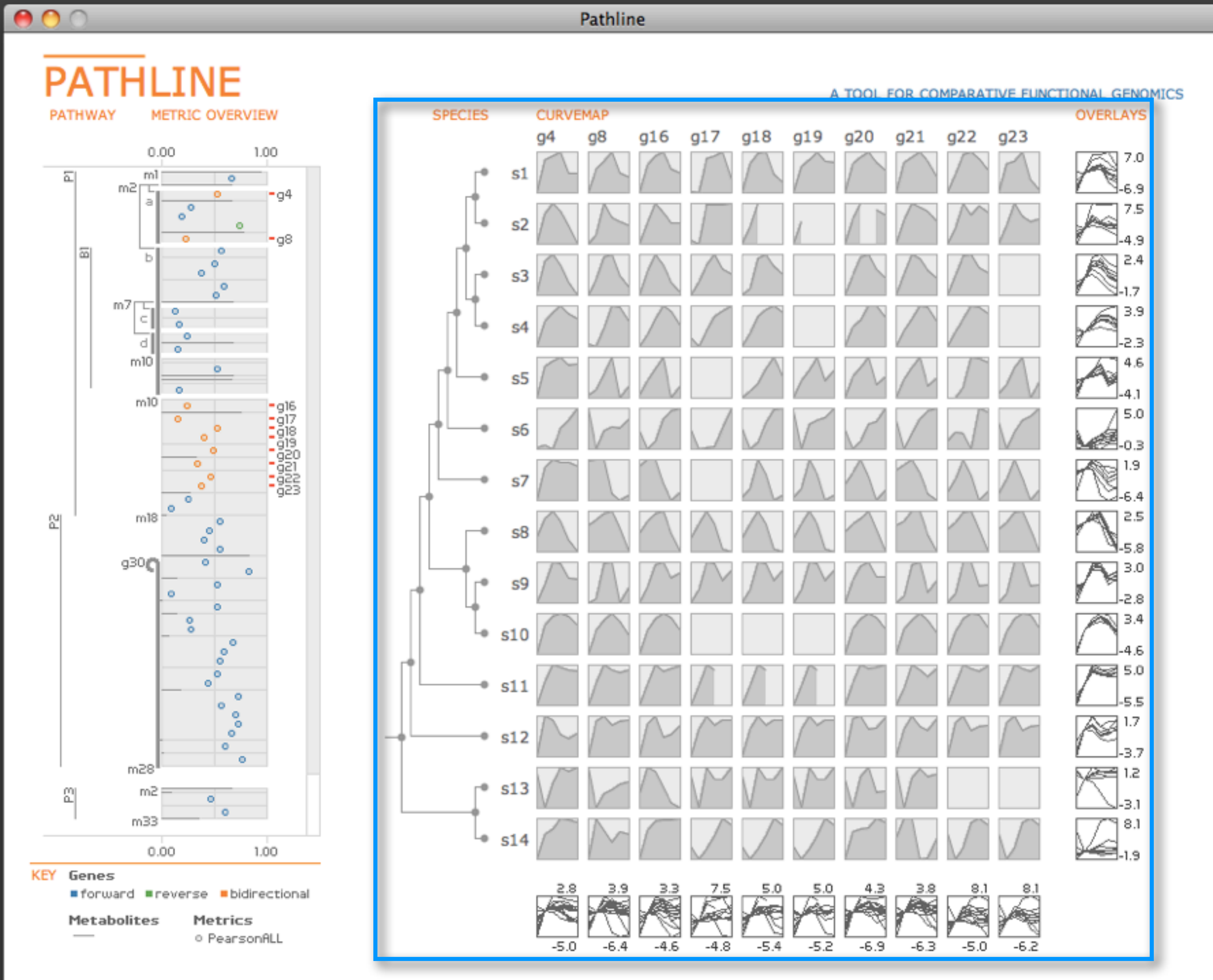
- evolutionary relationship
- **binary tree**

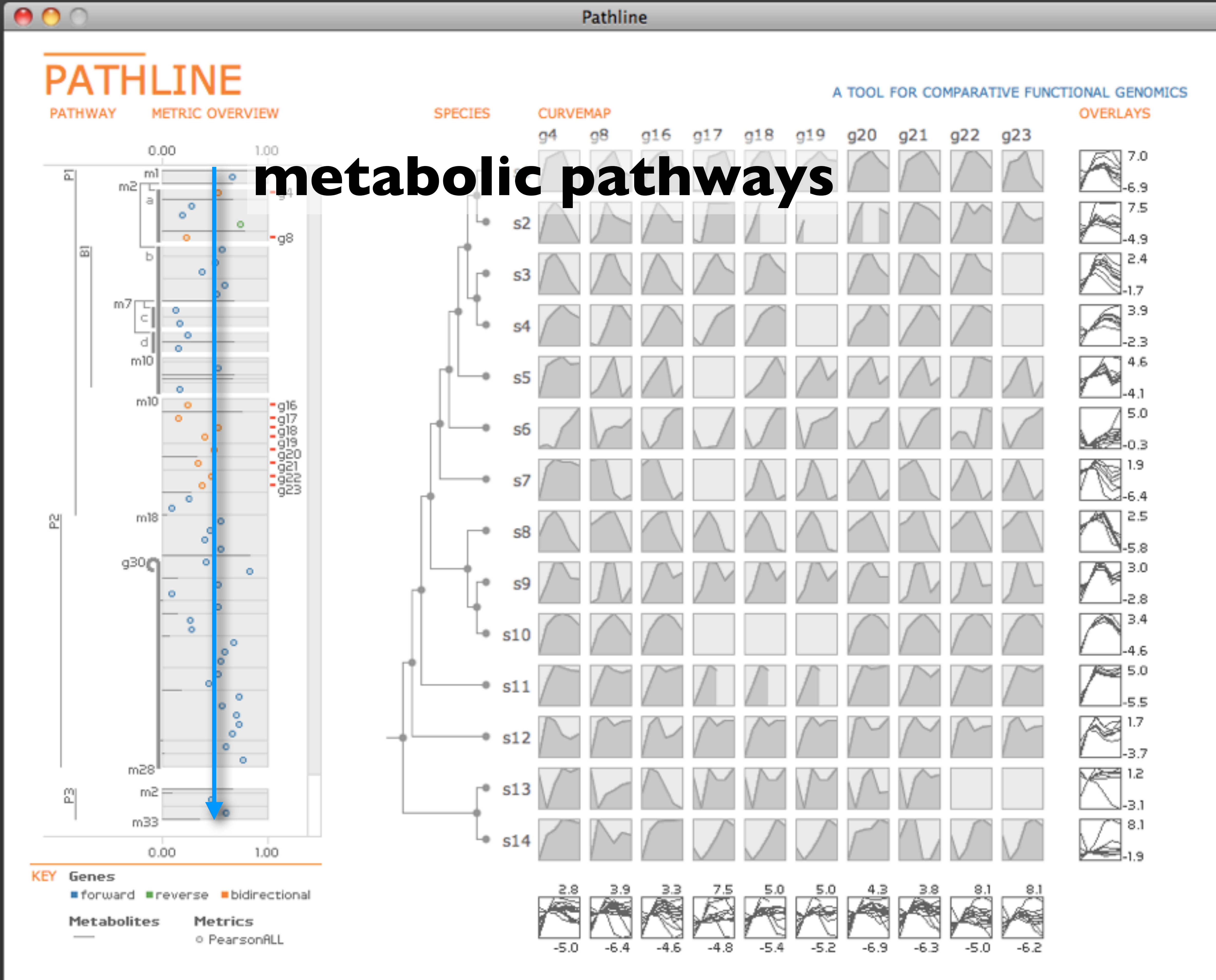


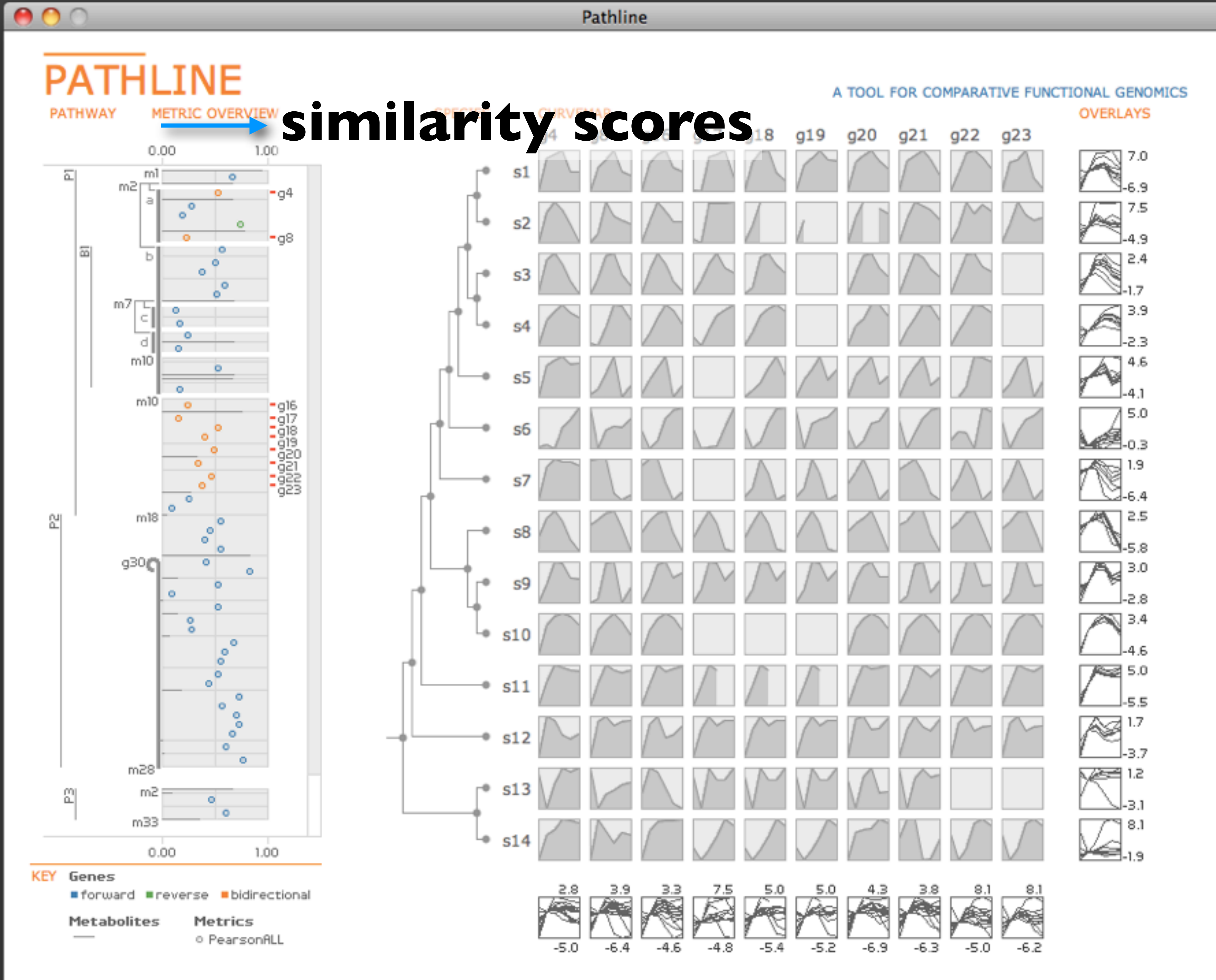
tasks

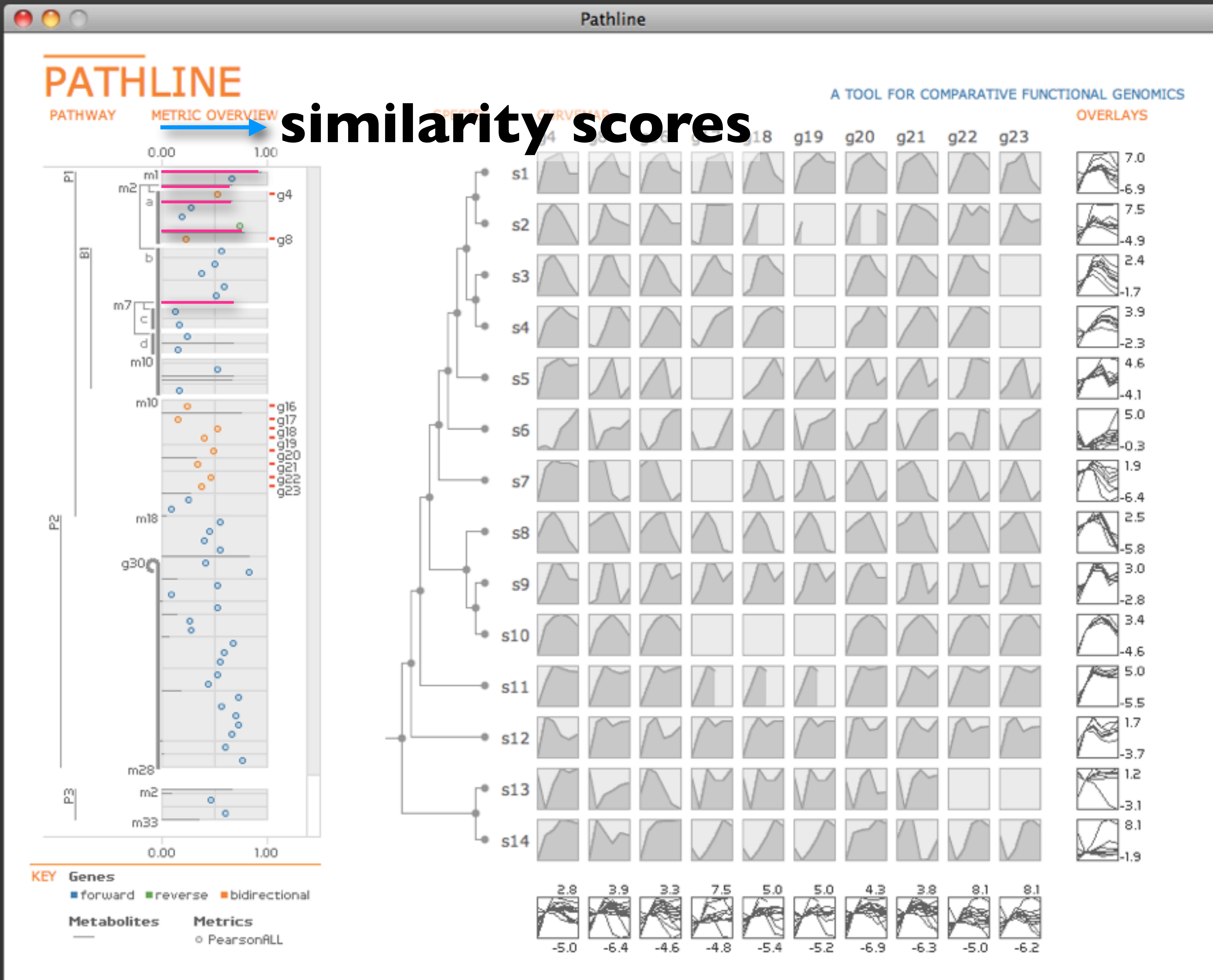
- study expression data as a time series
- compare a limited number of time series
- compare similarity scores along a pathway(s)
- comparison of multiple similarity scores

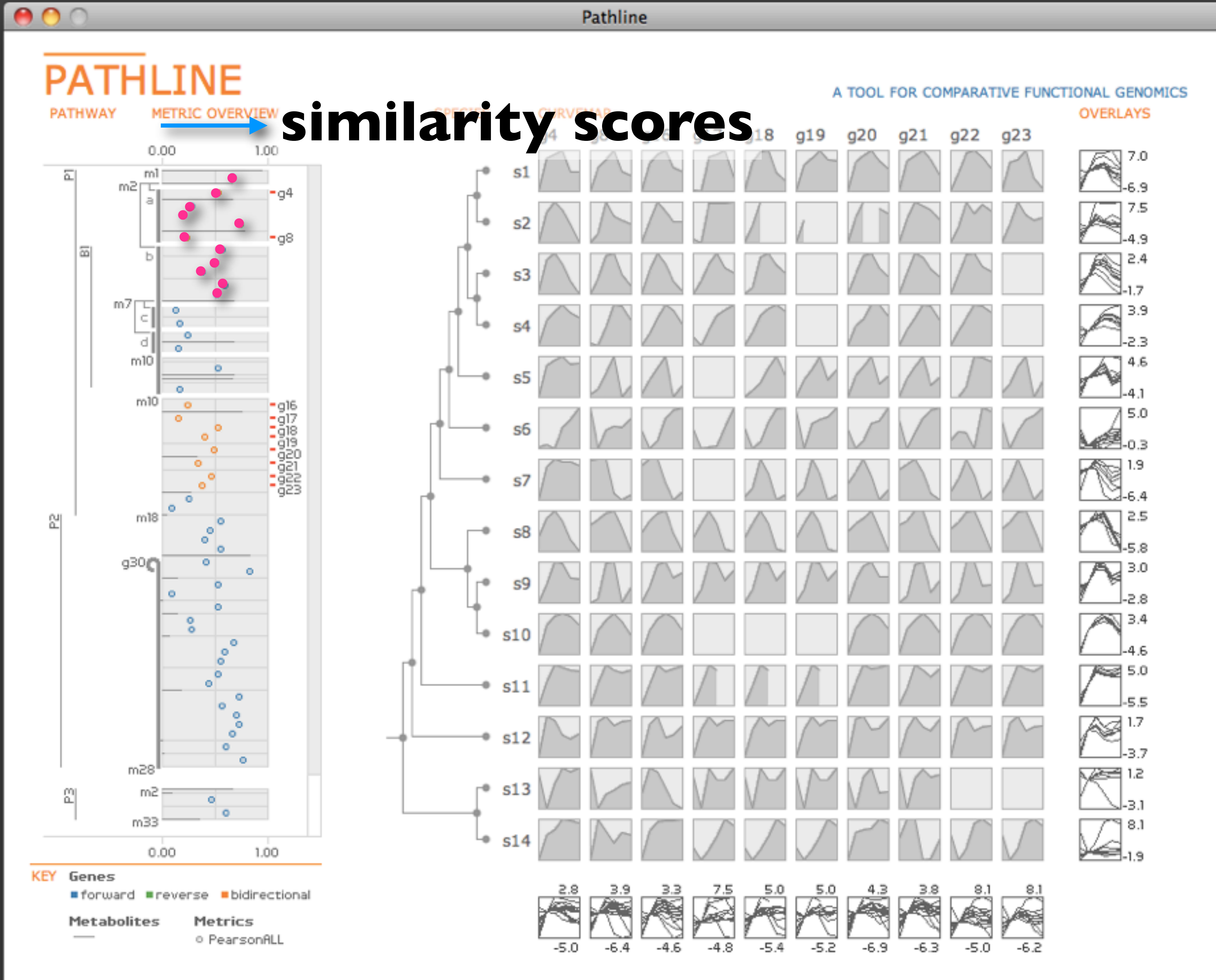

Pathline
PATHLINE
A TOOL FOR COMPARATIVE FUNCTIONAL GENOMICS
PATHWAY
METRIC OVERVIEW
SPECIES
CURVEMAP
OVERLAYS
0.00
1.00
P1
P2
P3
B1
m1
m2
m7
m10
m18
m28
m33
g30
g4
g8
g16
g17
g18
g19
g20
g21
g22
g23
s1
s2
s3
s4
s5
s6
s7
s8
s9
s10
s11
s12
s13
s14
KEY
Genes
forward
reverse
bidirectional
Metabolites
Metrics
PearsonALL


Pathline
PATHLINE
PATHWAY
METRIC OVERVIEW
A TOOL FOR COMPARATIVE FUNCTIONAL GENOMICS
SPECIES
CURVEMAP
OVERLAYS
g4 g8 g16 g17 g18 g19 g20 g21 g22 g23
s1 s2 s3 s4 s5 s6 s7 s8 s9 s10 s11 s12 s13 s14
KEY
Genes
forward
reverse
bidirectional
Metabolites
Metrics
PearsonALL


Pathline
PATHLINE
A TOOL FOR COMPARATIVE FUNCTIONAL GENOMICS
PATHWAY
METRIC OVERVIEW
SPECIES
CURVEMAP
OVERLAYS
g4 g8 g16 g17 g18 g19 g20 g21 g22 g23
s2 s3 s4 s5 s6 s7 s8 s9 s10 s11 s12 s13 s14
metabolic pathways
KEY Genes
forward reverse bidirectional
Metabolites
Metrics
PearsonALL

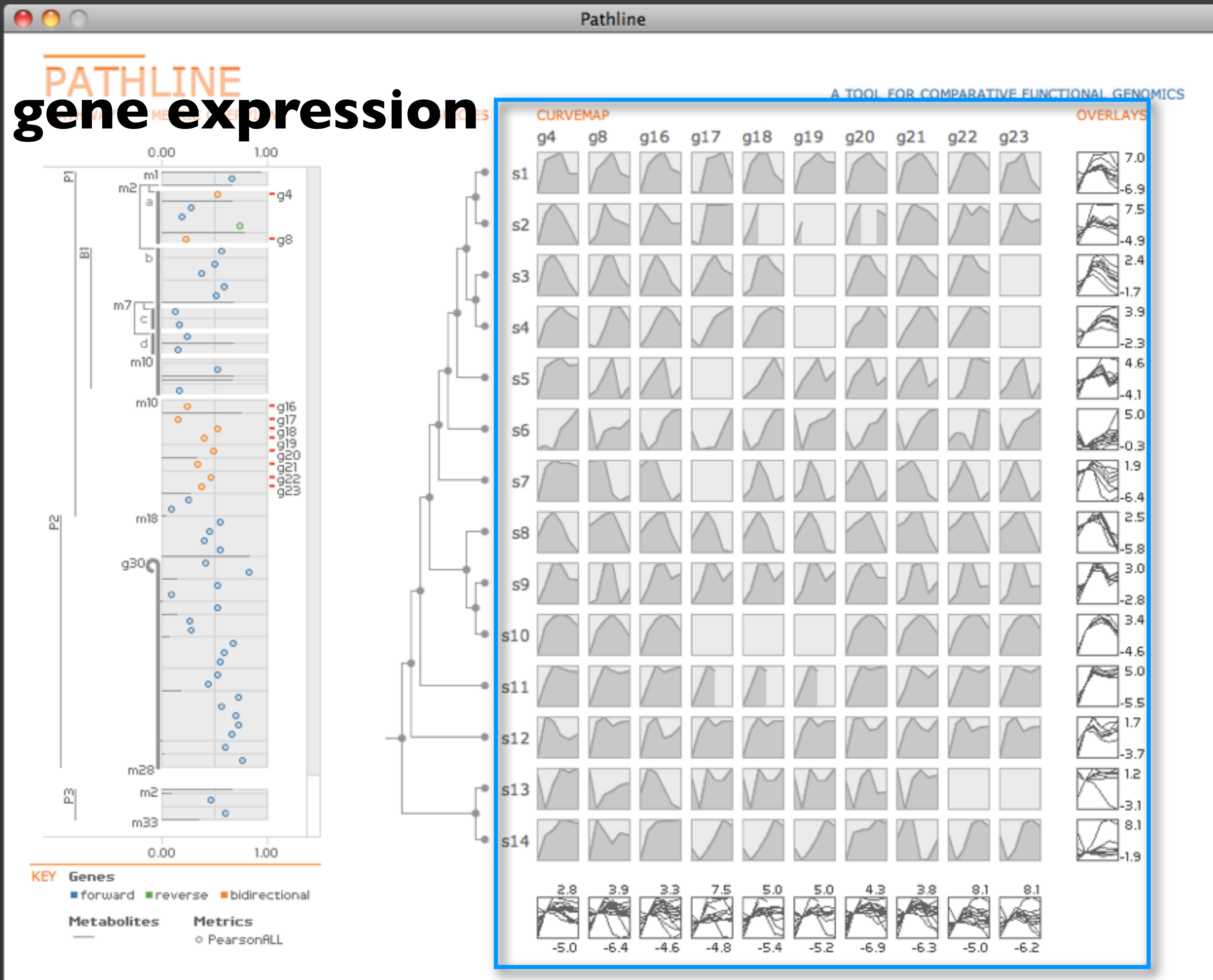

Pathline
PATHLINE
PATHWAY
METRIC OVERVIEW
similarity scores
A TOOL FOR COMPARATIVE FUNCTIONAL GENOMICS
OVERLAYS
g19
g20
g21
g22
g23
s1
s2
s3
s4
s5
s6
s7
s8
s9
s10
s11
s12
s13
s14
0.00
1.00
KEY
Genes
forward
reverse
bidirectional
Metabolites
Metrics
PearsonALL

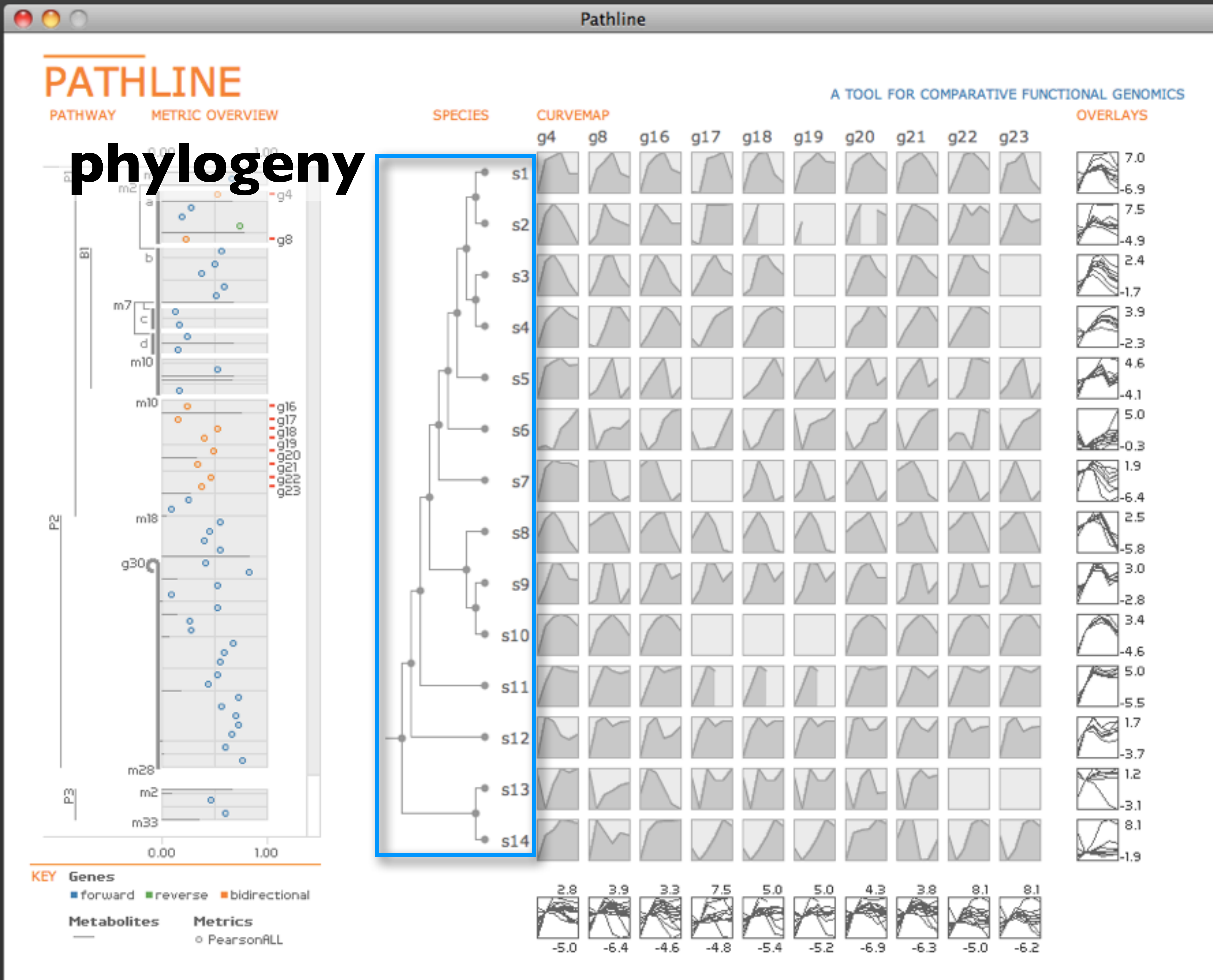

Pathline
PATHLINE
PATHWAY
METRIC OVERVIEW
similarity scores
A TOOL FOR COMPARATIVE FUNCTIONAL GENOMICS
OVERLAYS
g19 g20 g21 g22 g23
s1 s2 s3 s4 s5 s6 s7 s8 s9 s10 s11 s12 s13 s14
0.00 1.00
KEY Genes
forward reverse bidirectional
Metabolites
Metrics
PearsonALL



Pathline
PATHLINE
A TOOL FOR COMPARATIVE FUNCTIONAL GENOMICS
PATHWAY
METRIC OVERVIEW
OVERLAYS
similarity scores
g19
g20
g21
g22
g23
s1
s2
s3
s4
s5
s6
s7
s8
s9
s10
s11
s12
s13
s14
KEY
Genes
forward
reverse
bidirectional
Metabolites
Metrics
PearsonALL

gene expression




Pathline
PATHLINE
A TOOL FOR COMPARATIVE FUNCTIONAL GENOMICS
PATHWAY
METRIC OVERVIEW
SPECIES
CURVEMAP
OVERLAYS
phylogeny
g4 g8 g16 g17 g18 g19 g20 g21 g22 g23
s1 s2 s3 s4 s5 s6 s7 s8 s9 s10 s11 s12 s13 s14
KEY
Genes
forward
reverse
bidirectional
Metabolites
Metrics
PearsonALL

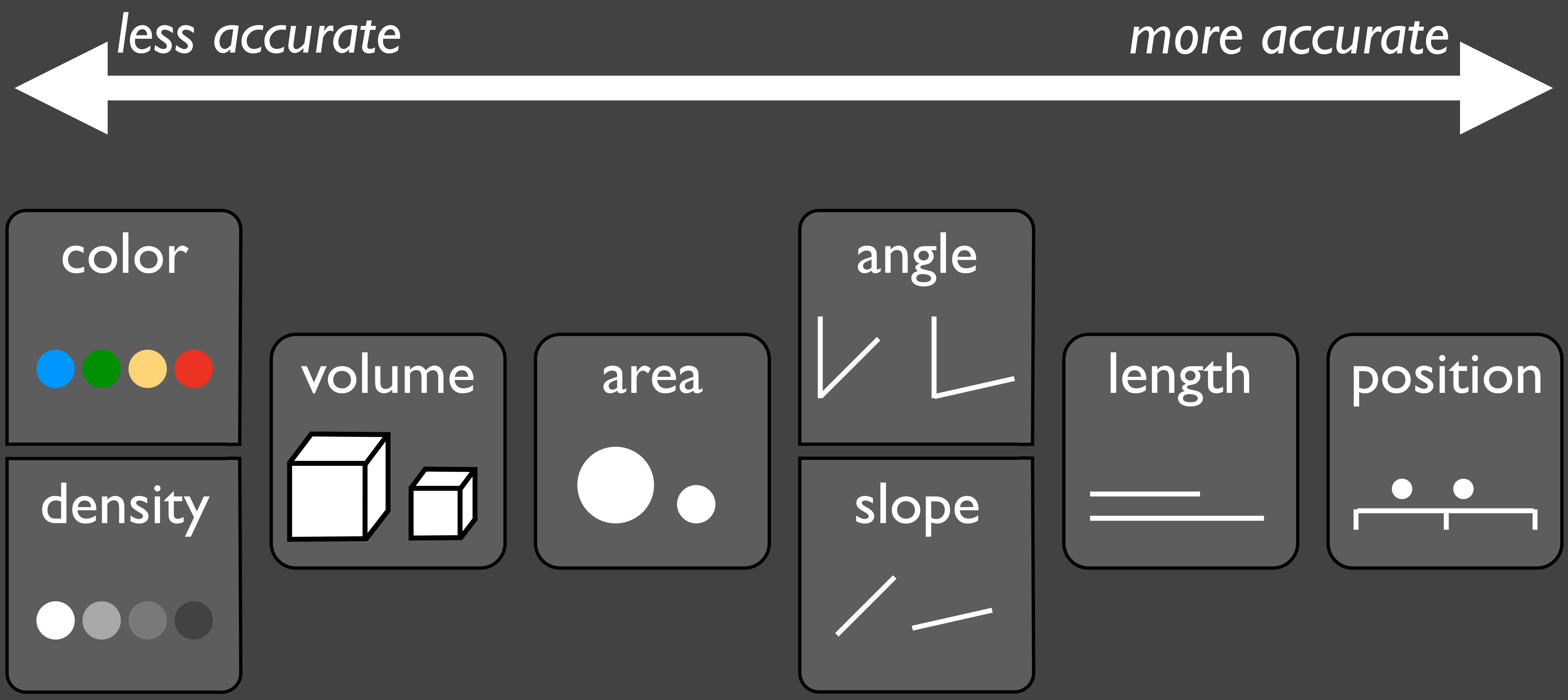
target

translate

design

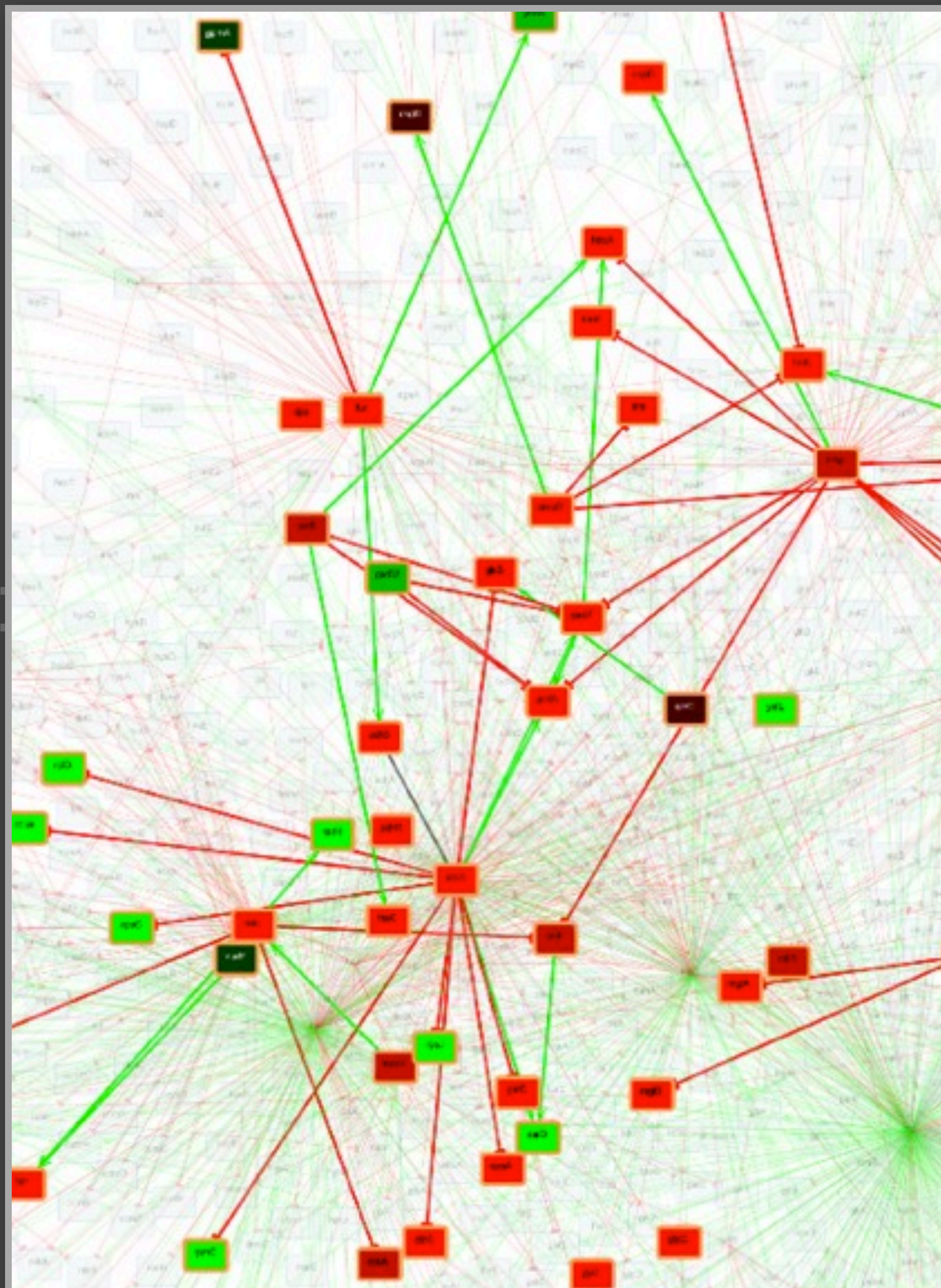
implement

validate


less accurate
more accurate
color
density
volume
area
angle
slope
length
position

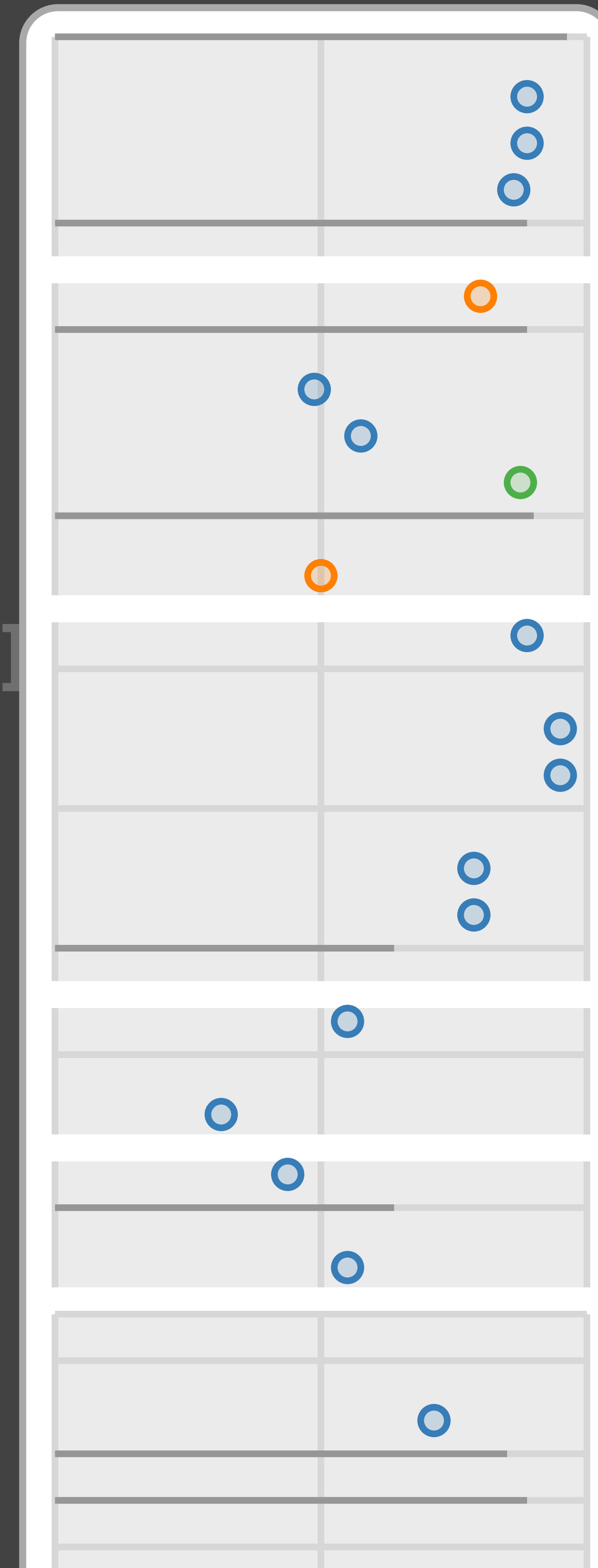
encode quantitative values with spatial position

topological layout

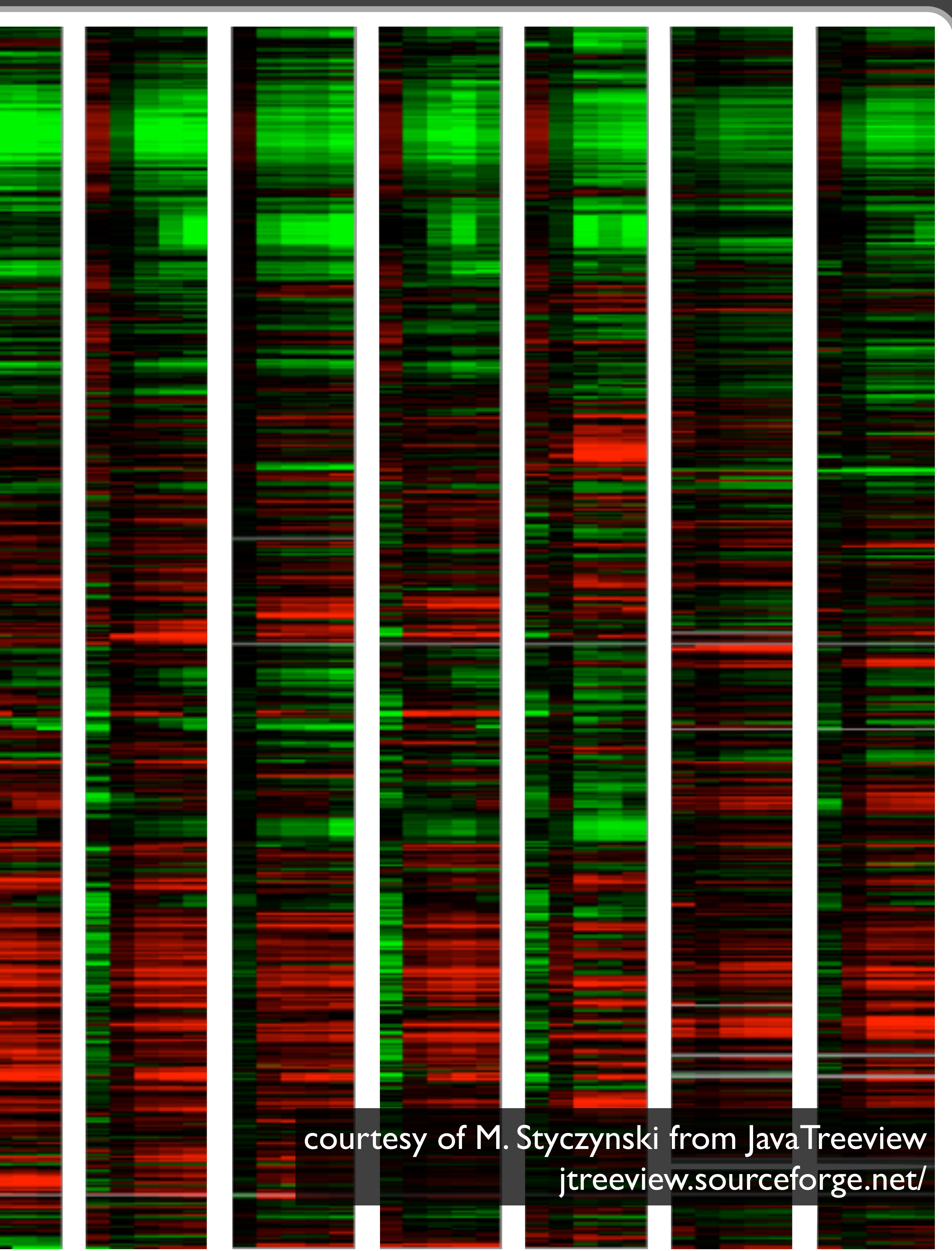


www.win.tue.nl/~mwestenb/genevis/

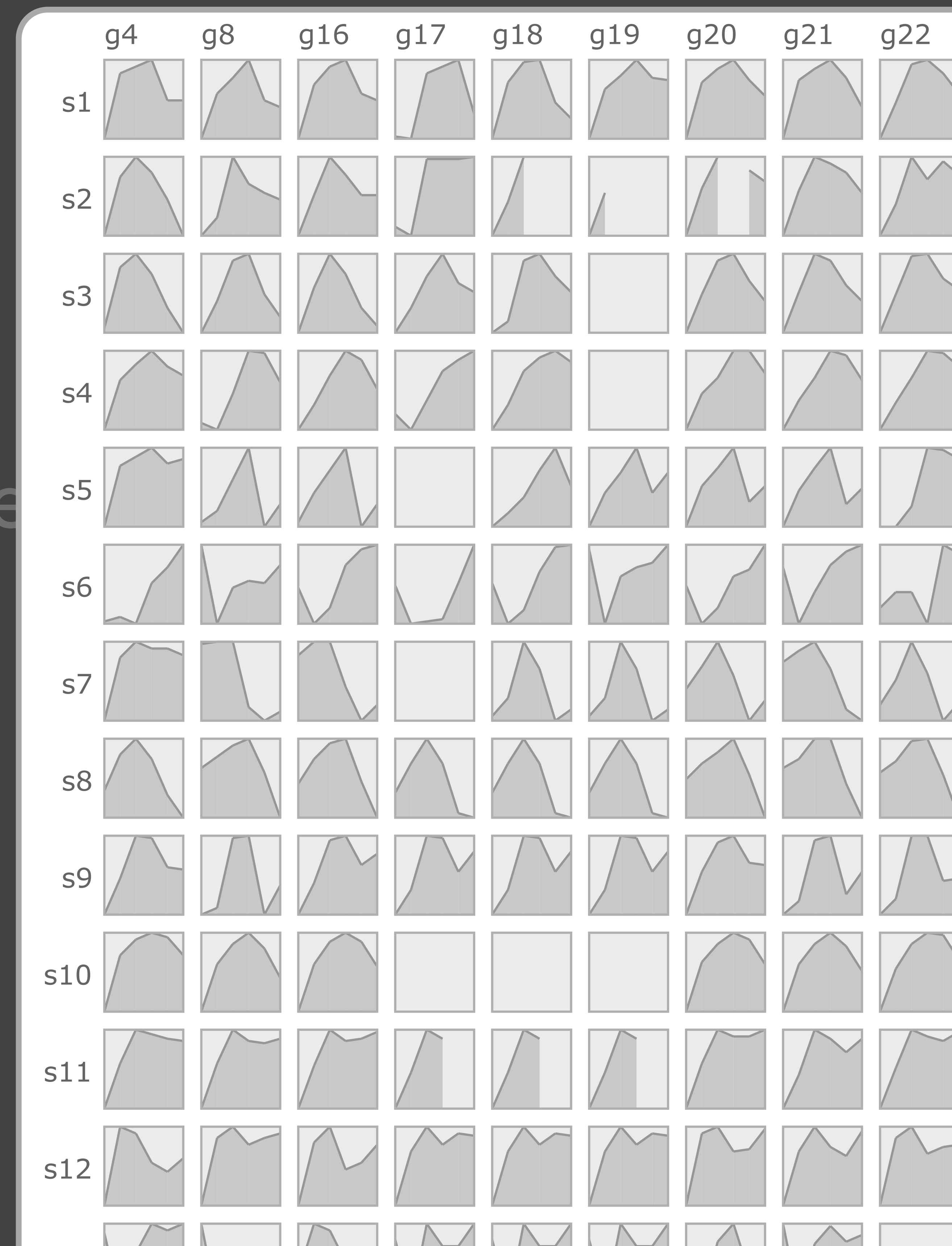
linearized pathway



heatmap



curvemap



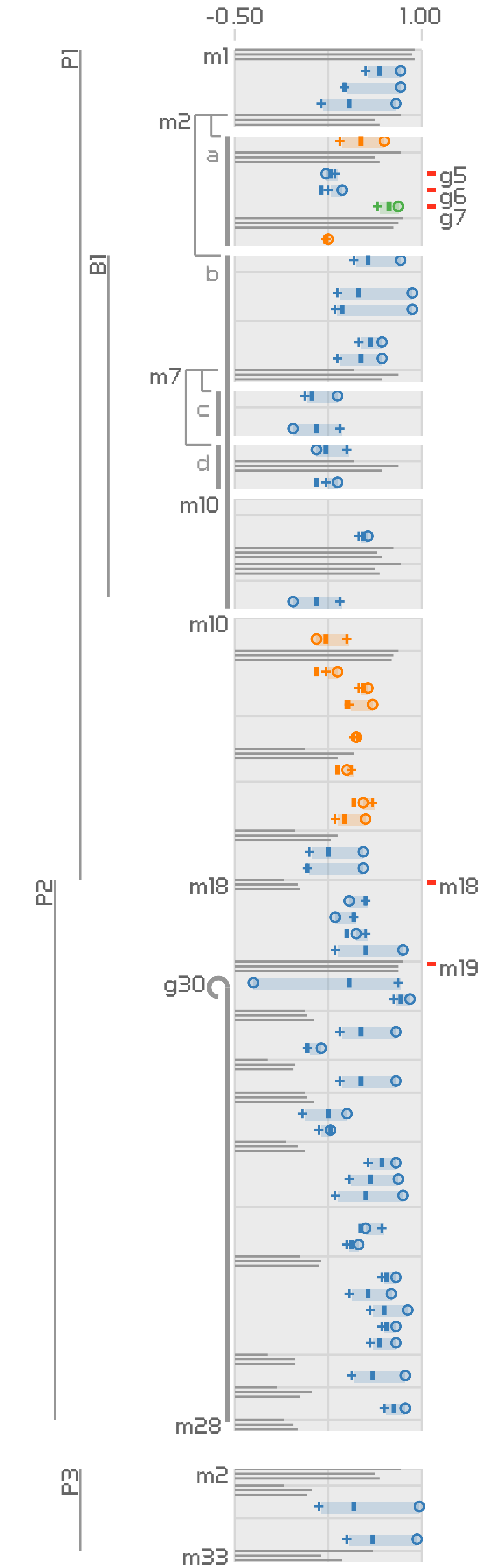
Pathline

linearized pathway representation

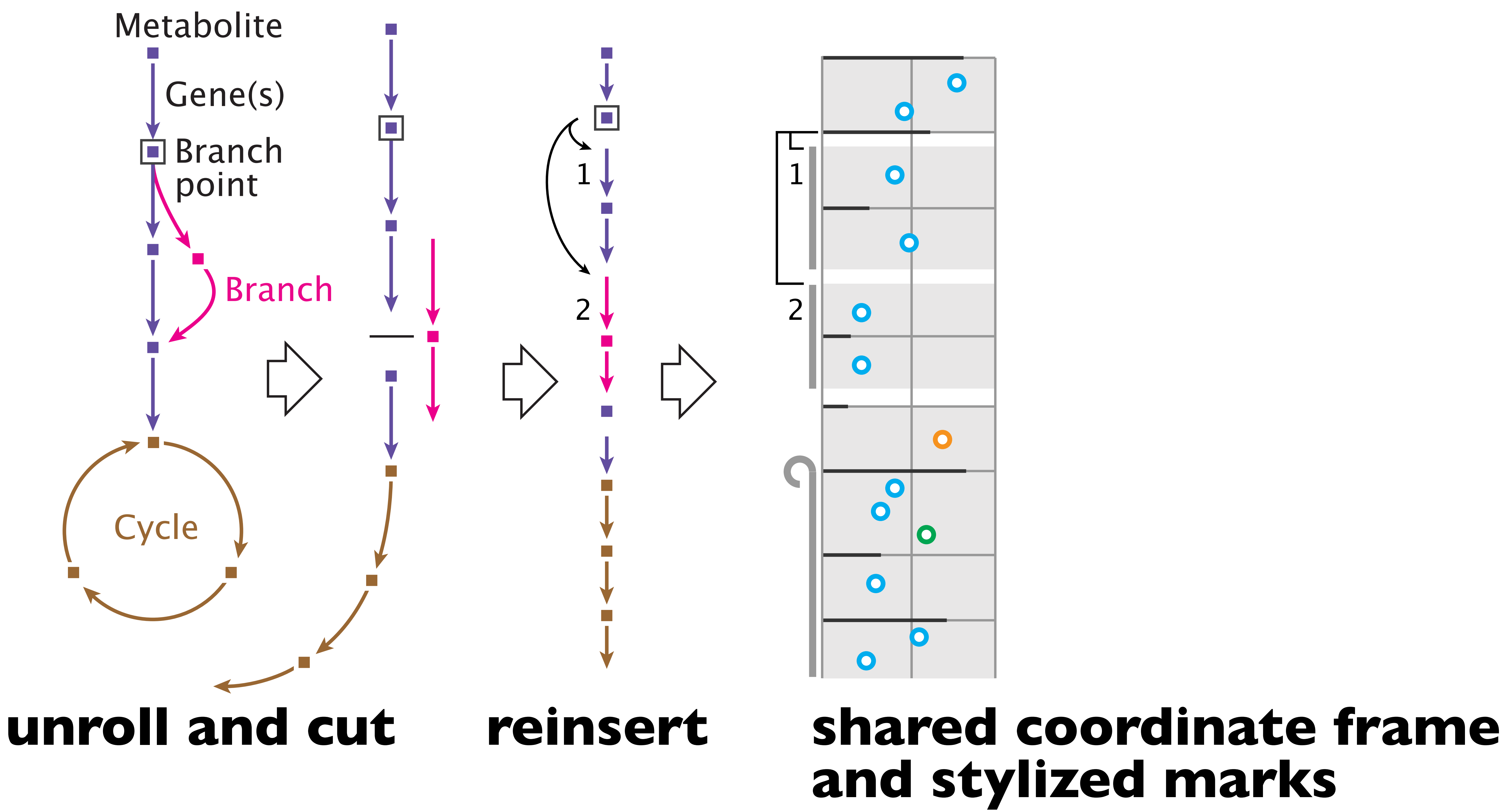
linearized pathway representation

common axes to compare similarity scores

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



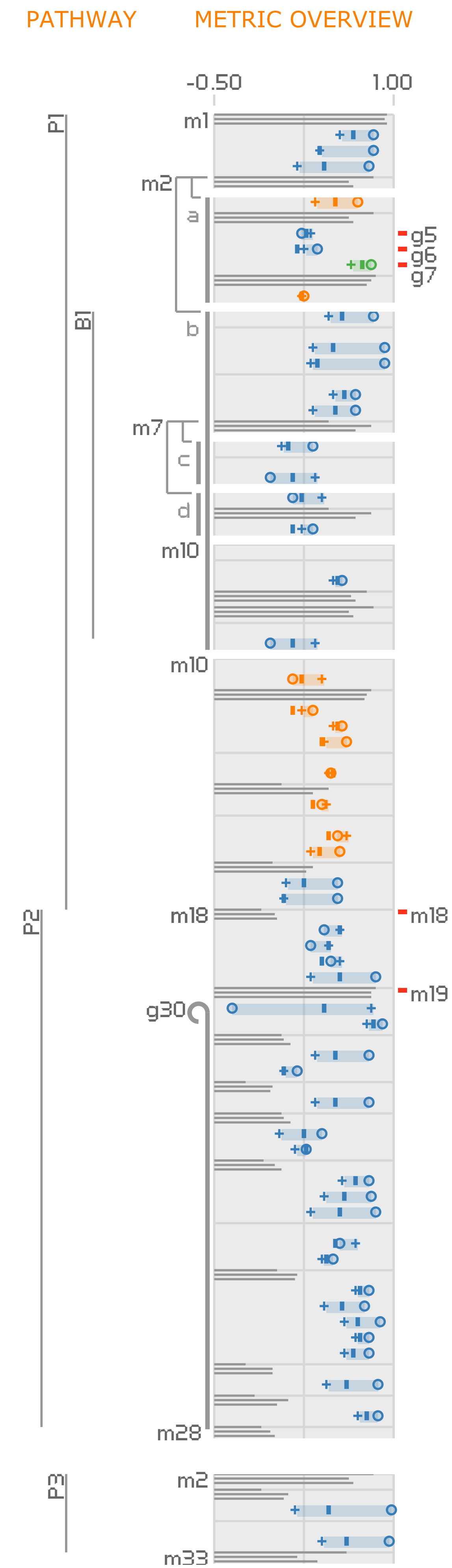
pathway to ordered list of nodes



linearized pathway representation

putting it together . . .

- use spatial position for similarity scores
- topology is secondary



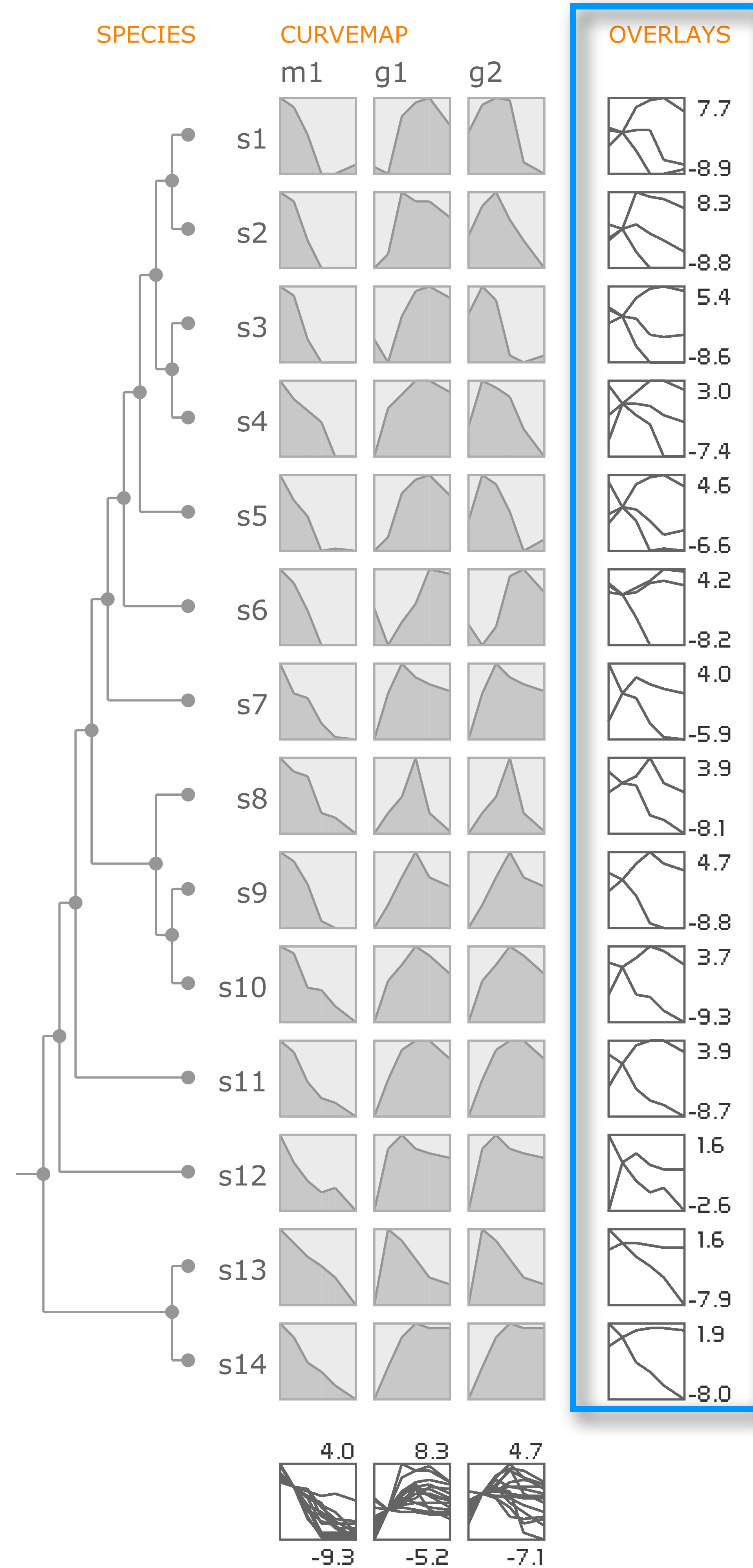
Pathline

curvemap

curvemap

inspired by heatmaps

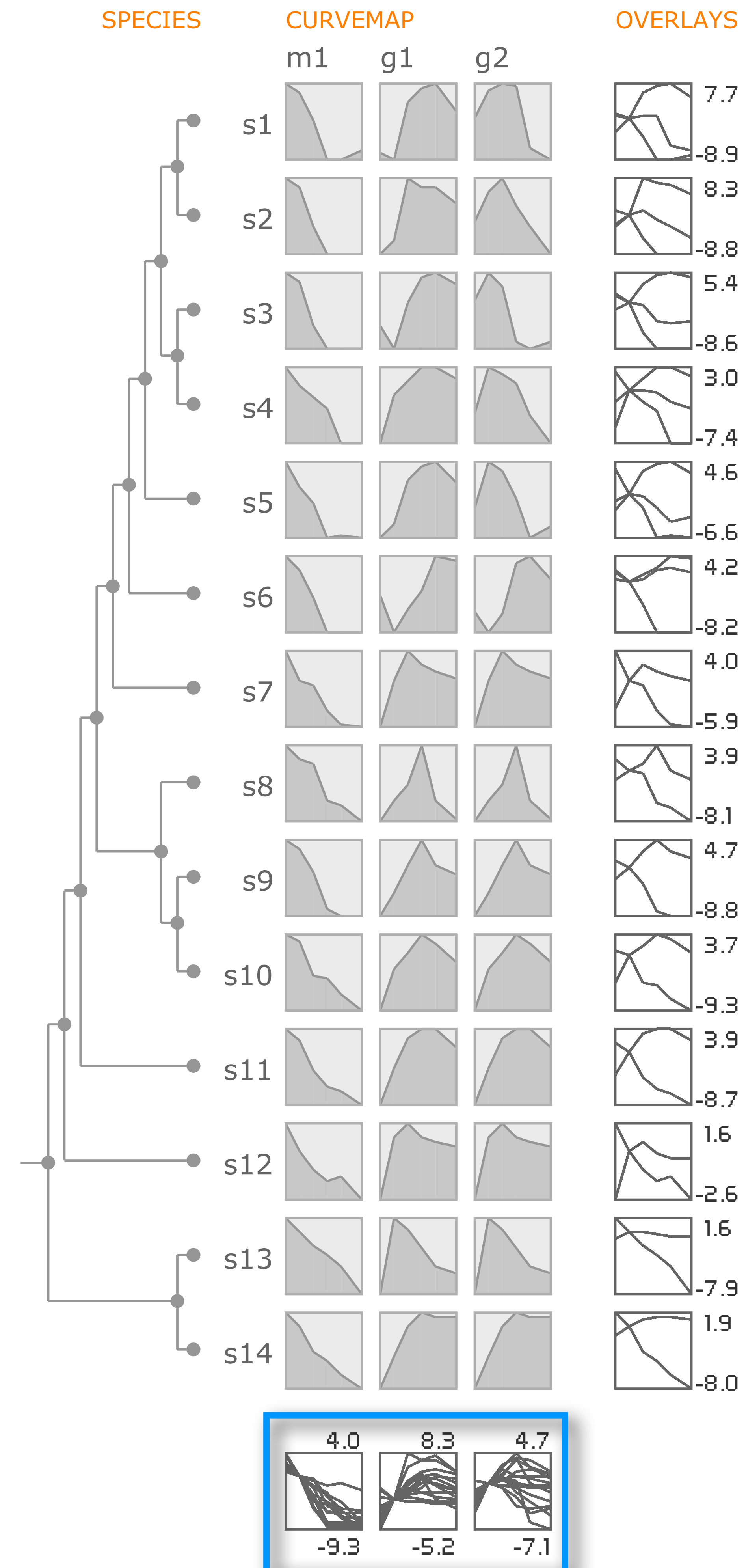
- base visual unit is a curve
- filled, framed line charts to enhance shape perception
- rows are species
- columns are genes/metabolites
- overlays to enhance trends



curvemap

inspired by heatmaps

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



target

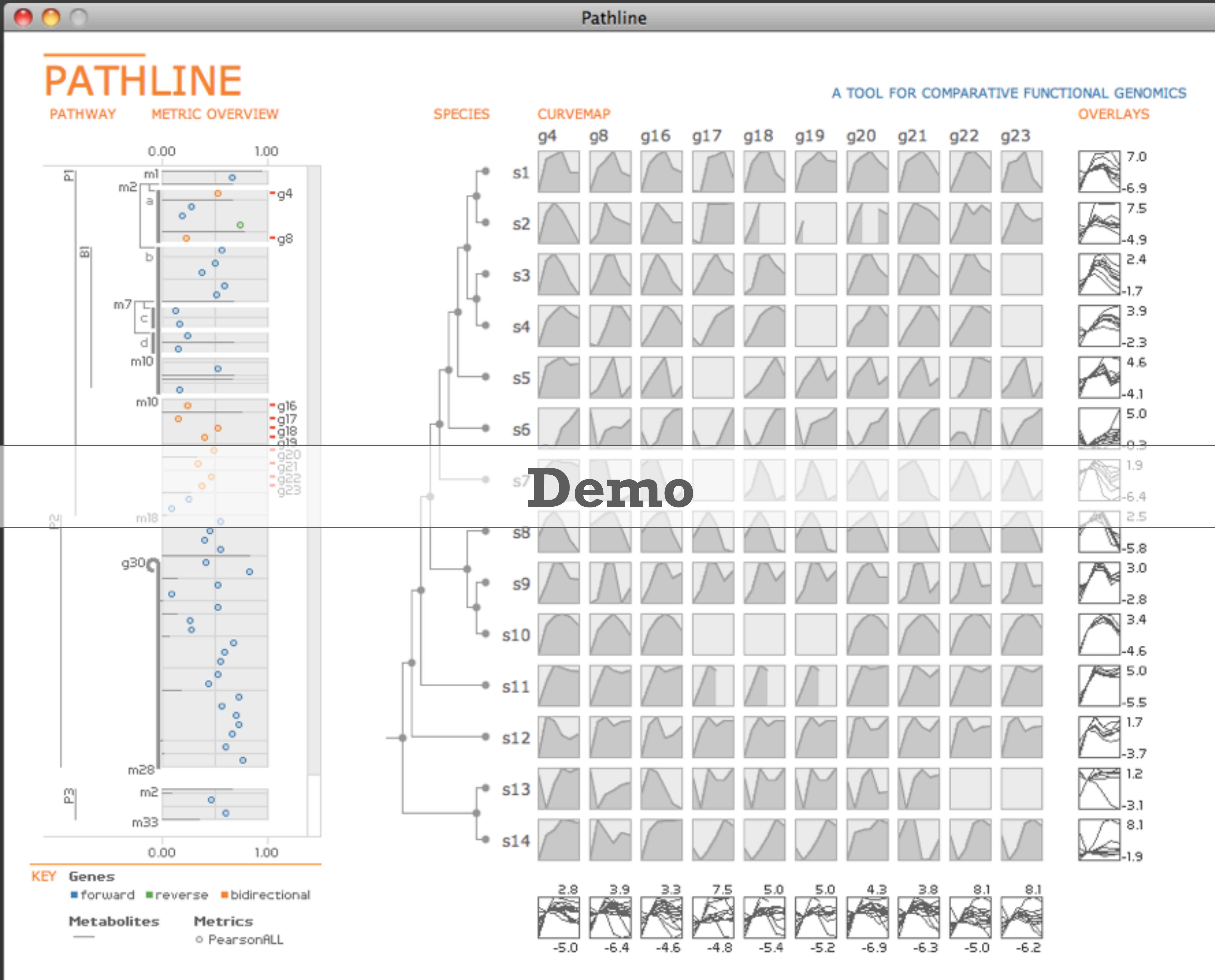
translate

design

implement

validate

Demo


Pathline
PATHLINE
A TOOL FOR COMPARATIVE FUNCTIONAL GENOMICS
PATHWAY
METRIC OVERVIEW
SPECIES
CURVEMAP
OVERLAYS
g4 g8 g16 g17 g18 g19 g20 g21 g22 g23
s1 s2 s3 s4 s5 s6 s7 s8 s9 s10 s11 s12 s13 s14
0.00 1.00
KEY Genes
forward reverse bidirectional
Metabolites
Metrics
PearsonALL

target

translate

design

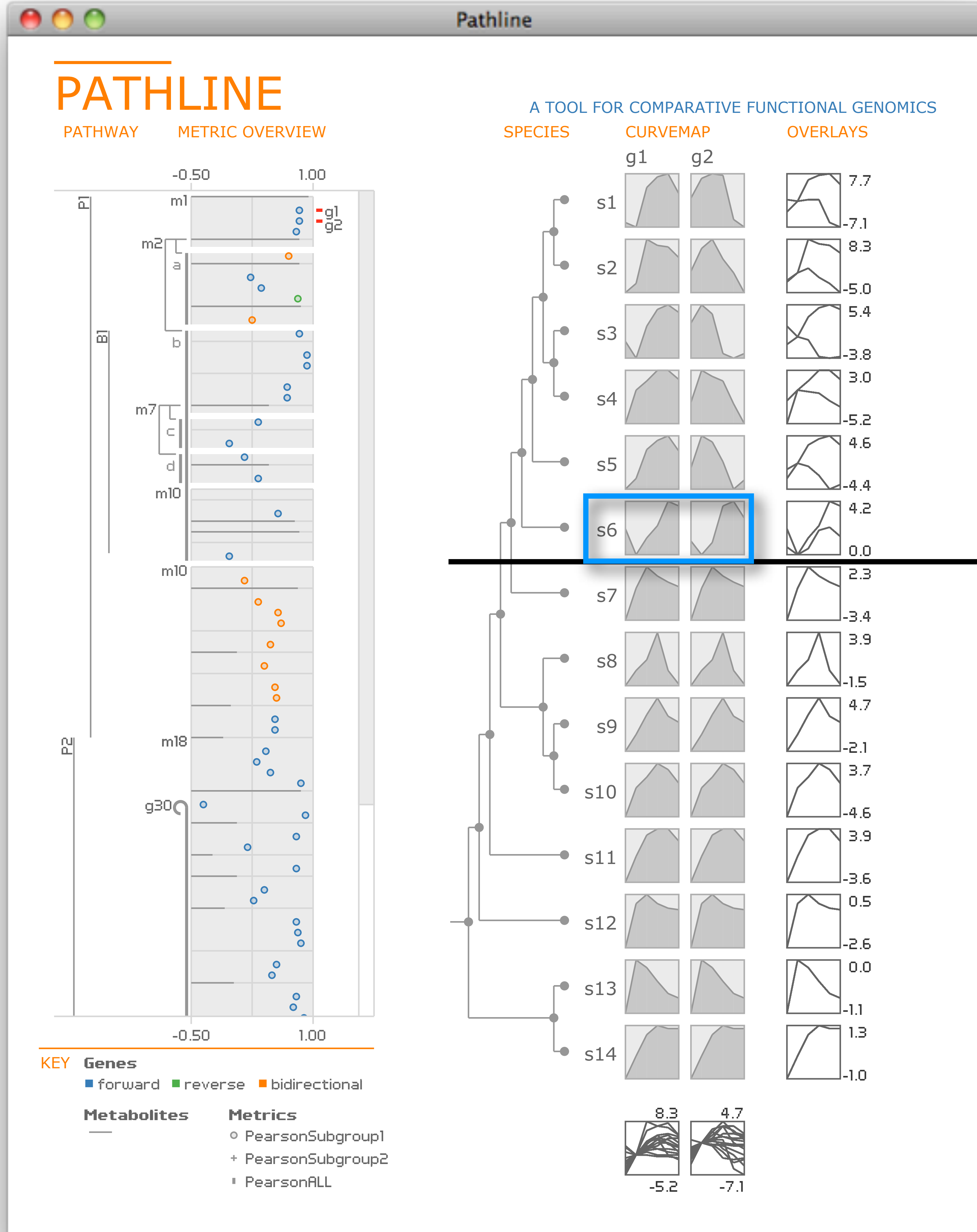
implement

validate

case study

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

whole genome duplication



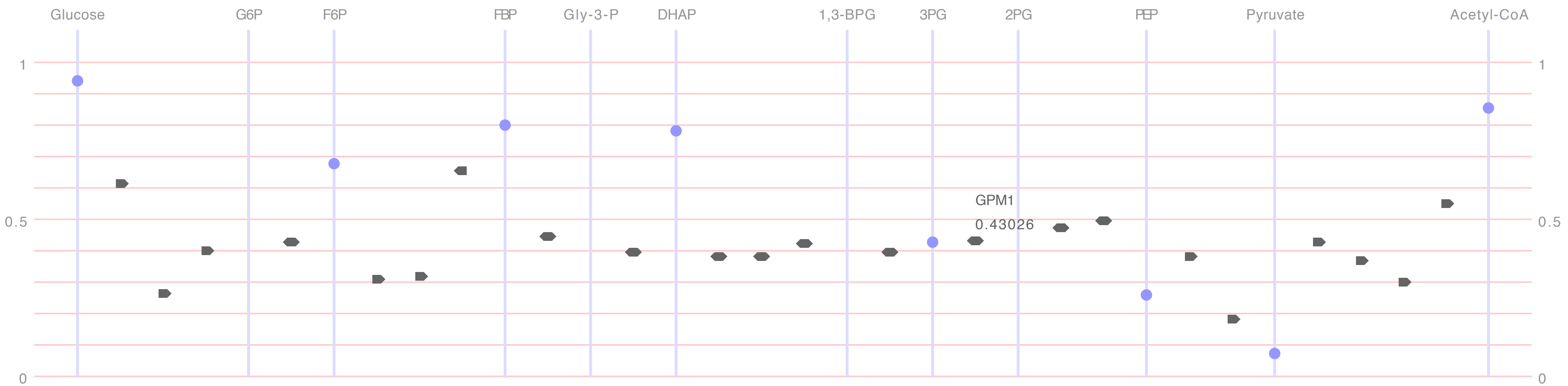
both genes

one gene

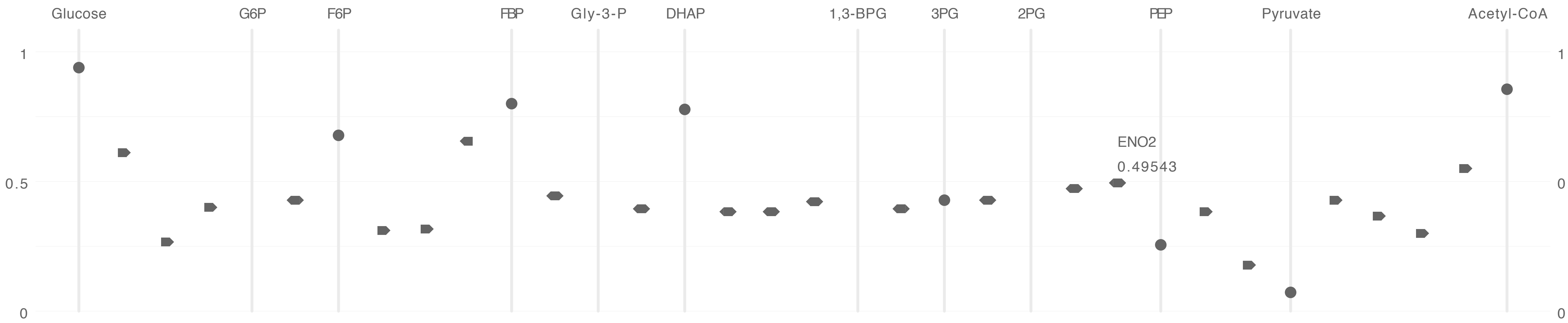
www.pathline.org

LESSONS LEARNED

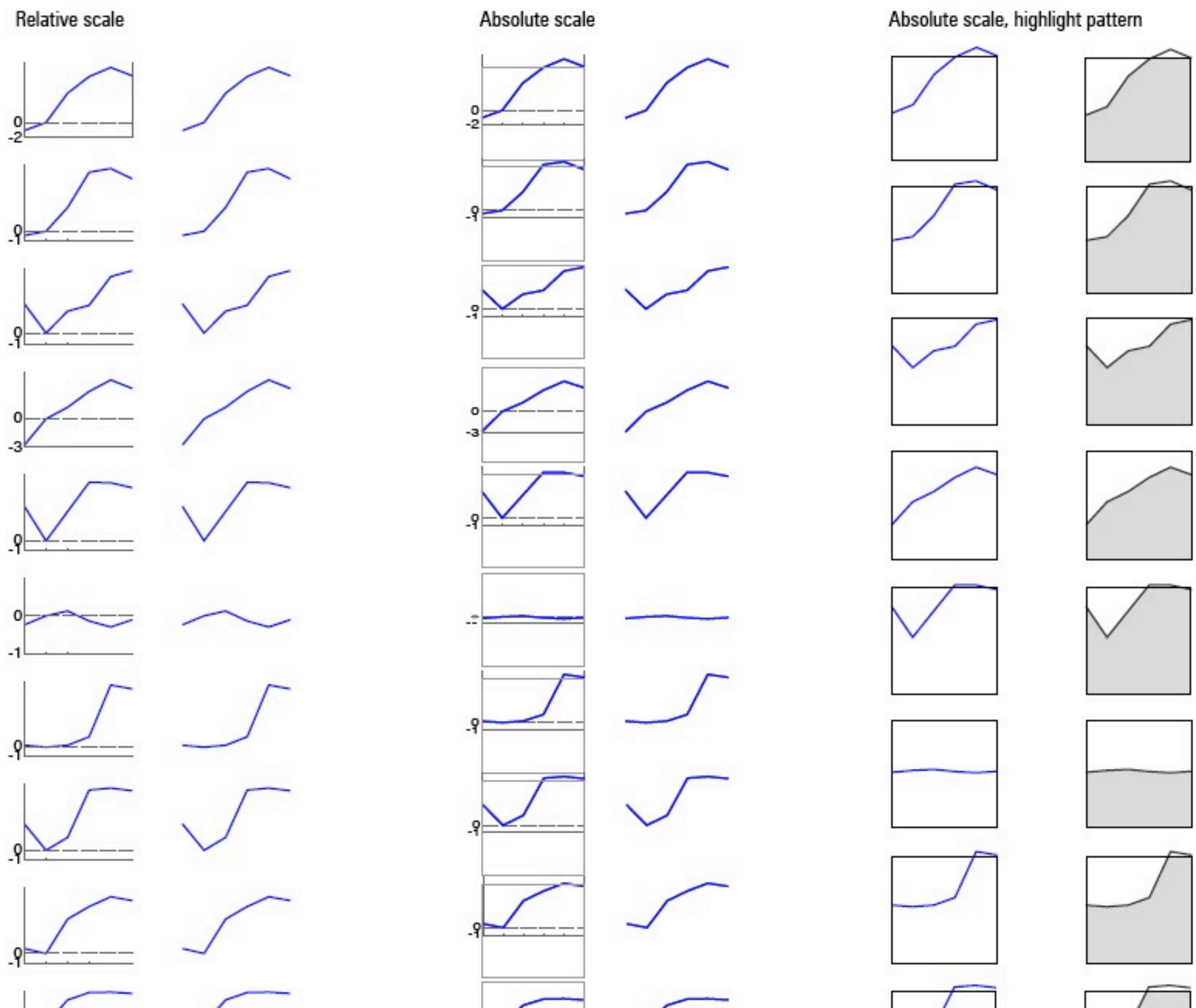
- **process supports efficient development**
- **collaborators' time commitment is front loaded**
- **rapid prototyping is essential**



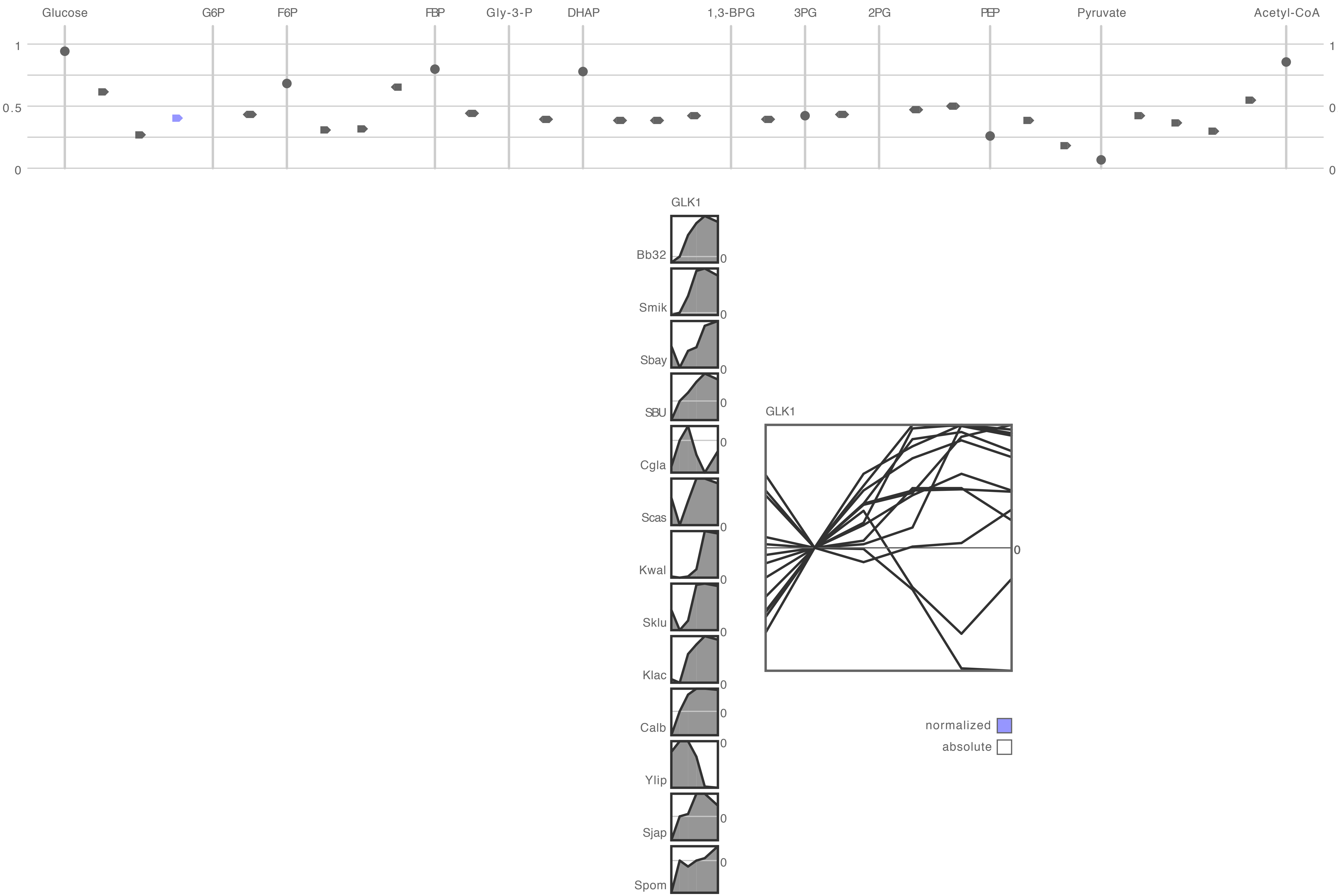
Glucose
G6P
F6P
FBP
Gly-3-P
DHAP
1,3-BPG
3PG
2PG
PEP
Pyruvate
Acetyl-CoA
1
0.5
0
GPM1
0.43026



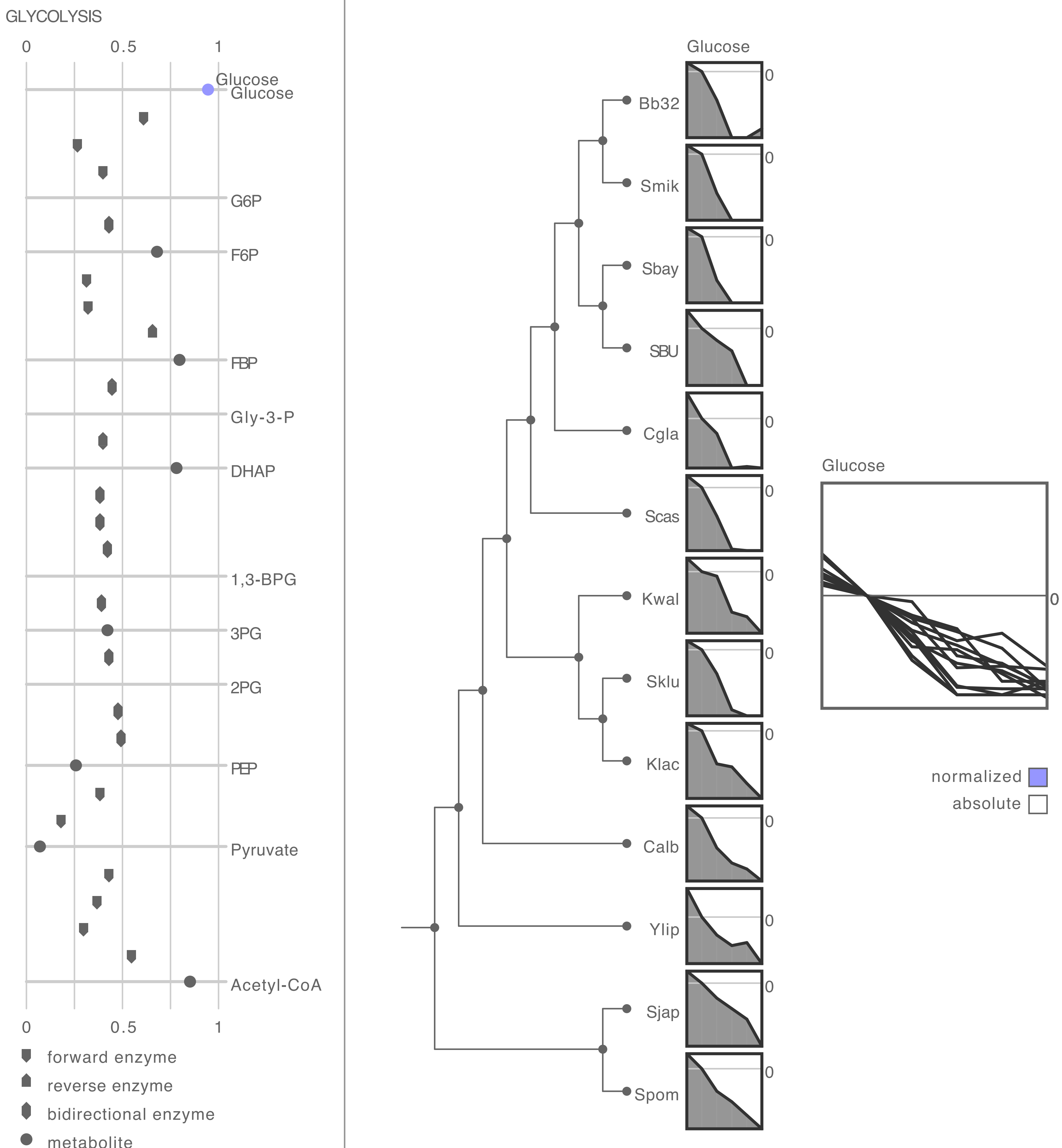
Glucose
G6P
F6P
FBP
Gly-3-P
DHAP
1,3-BPG
3PG
2PG
PEP
Pyruvate
Acetyl-CoA
ENO2
0.49543
1
0.5
0

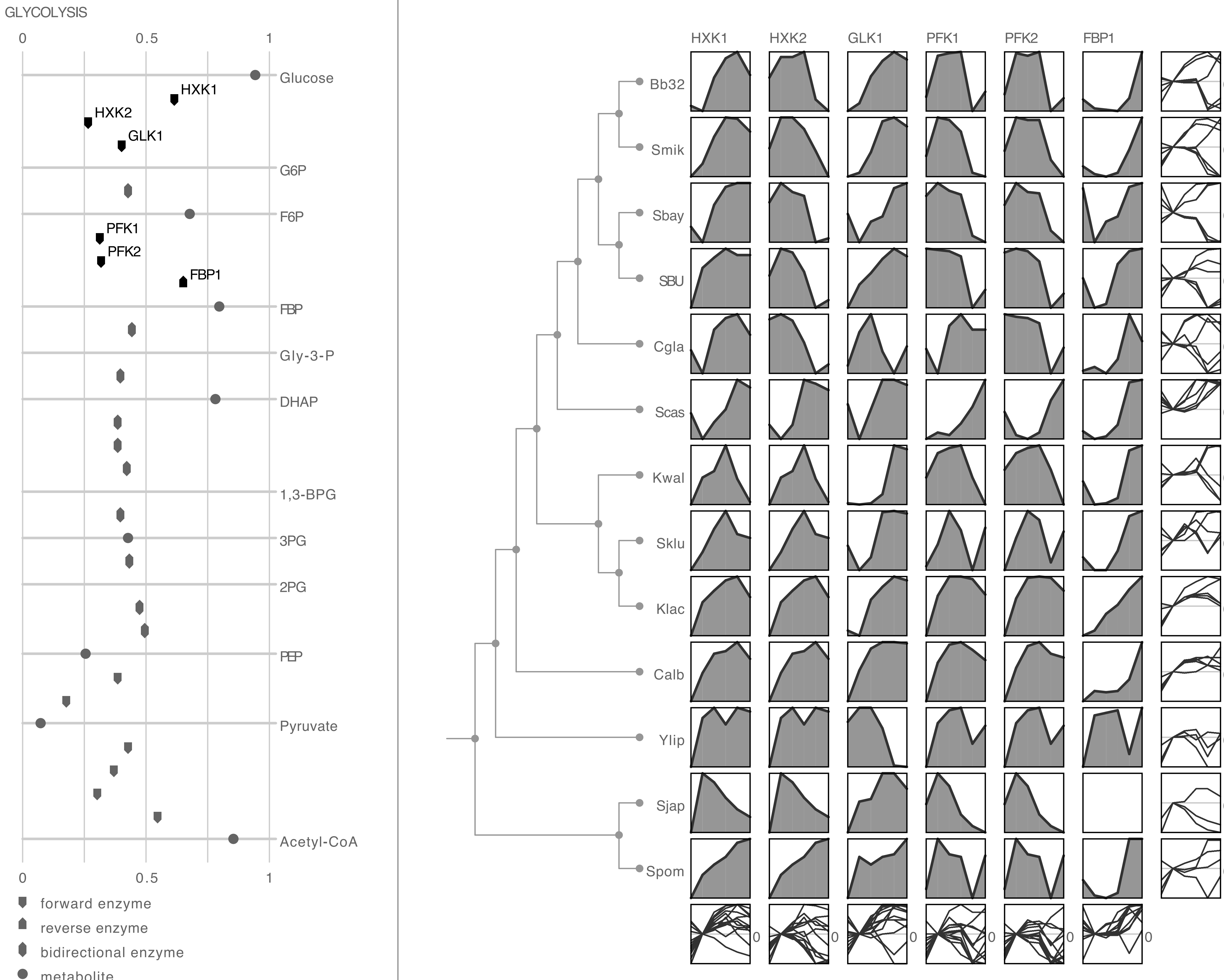


Relative scale
Absolute scale
Absolute scale, highlight pattern

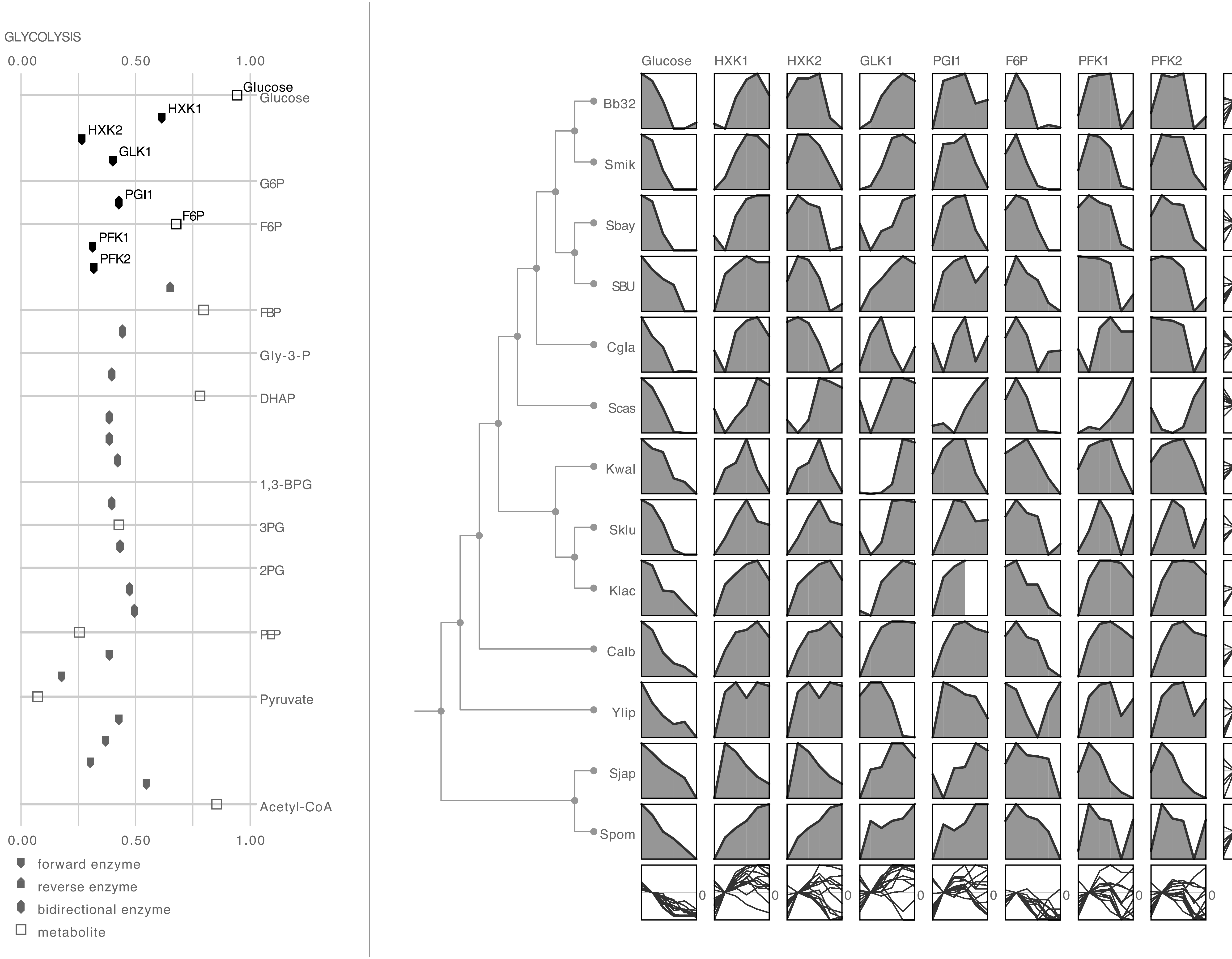


Glucose
G6P
F6P
FBP
Gly-3-P
DHAP
1,3-BPG
3PG
2PG
PEP
Pyruvate
Acetyl-CoA
1
0.5
0
GLK1
Bb32
Smik
Sbay
SBU
Cgla
Scas
Kwal
Sklu
Klac
Calb
Ylip
Sjap
Spom
GLK1
normalized
absolute

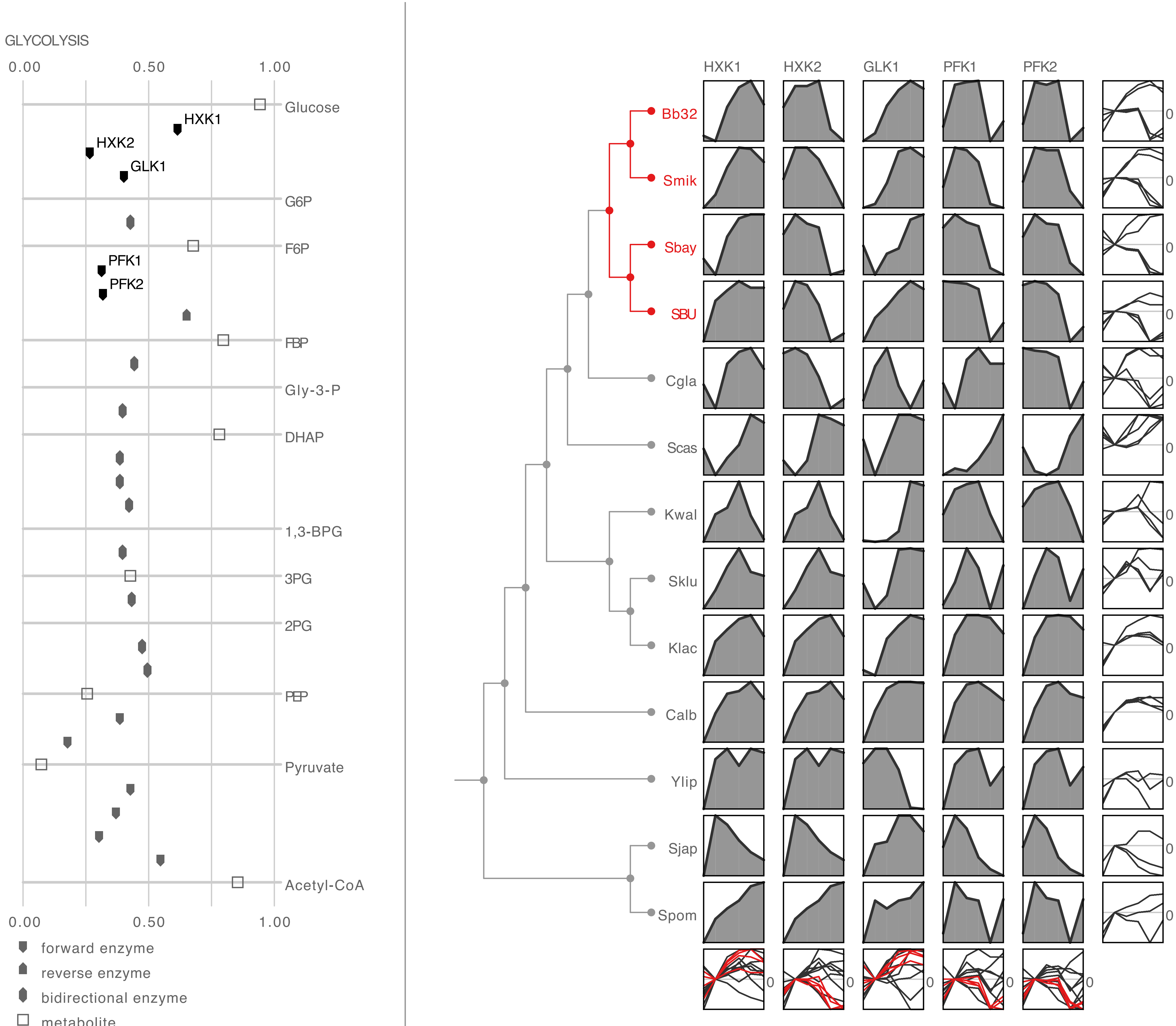

GLYCOLYSIS
0
0.5
1
Glucose
Glucose
G6P
F6P
FBP
Gly-3-P
DHAP
1,3-BPG
3PG
2PG
PEP
Pyruvate
Acetyl-CoA
forward enzyme
reverse enzyme
bidirectional enzyme
metabolite
Glucose
Bb32
Smik
Sbay
SBU
Cgla
Scas
Kwal
Sklu
Klac
Calb
Ylip
Sjap
Spom
Glucose
normalized
absolute

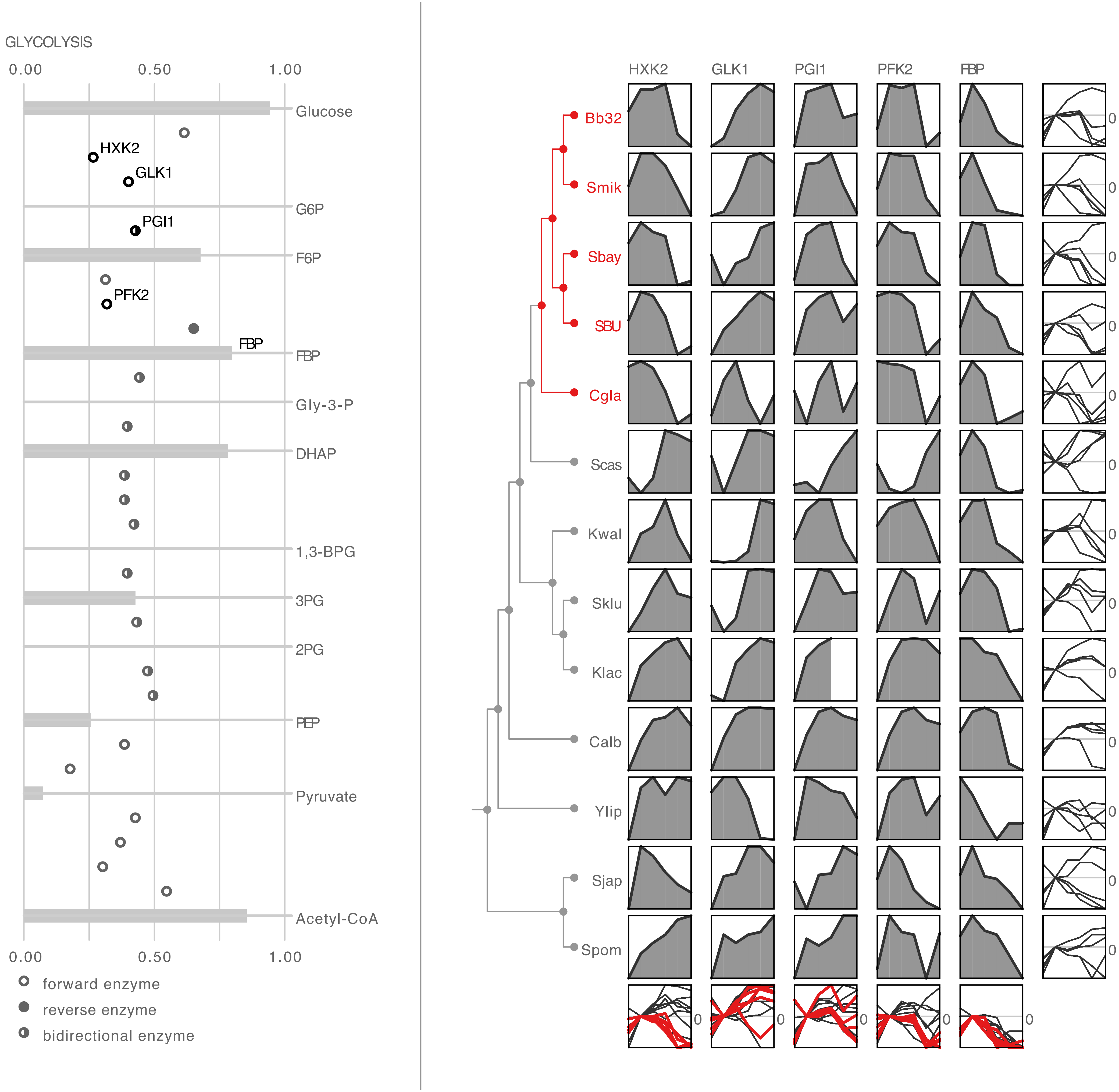


GLYCOLYSIS
0
0.5
1
Glucose
HXK1
HXK2
GLK1
G6P
F6P
PFK1
PFK2
FBP1
FBP
Gly-3-P
DHAP
1,3-BPG
3PG
2PG
PEP
Pyruvate
Acetyl-CoA
forward enzyme
reverse enzyme
bidirectional enzyme
metabolite
HXK1
HXK2
GLK1
PFK1
PFK2
FBP1
Bb32
Smik
Sbay
SBU
Cgla
Scas
Kwal
Sklu
Klac
Calb
Ylip
Sjap
Spom



GLYCOLYSIS
0.00
0.50
1.00
Glucose
HXK1
HXK2
GLK1
G6P
PGI1
F6P
PFK1
PFK2
FBP
Gly-3-P
DHAP
1,3-BPG
3PG
2PG
PEP
Pyruvate
Acetyl-CoA
forward enzyme
reverse enzyme
bidirectional enzyme
metabolite
Bb32
Smik
Sbay
SBU
Cgla
Scas
Kwal
Sklu
Klac
Calb
Ylip
Sjap
Spom
Glucose
HXK1
HXK2
GLK1
PGI1
F6P
PFK1
PFK2
0


GLYCOLYSIS
0.00
0.50
1.00
Glucose
HXK1
HXK2
GLK1
G6P
F6P
PFK1
PFK2
FBP
Gly-3-P
DHAP
1,3-BPG
3PG
2PG
PEP
Pyruvate
Acetyl-CoA
0.00
0.50
1.00
forward enzyme
reverse enzyme
bidirectional enzyme
metabolite
HXK1
HXK2
GLK1
PFK1
PFK2
Bb32
Smik
Sbay
SBU
Cgla
Scas
Kwal
Sklu
Klac
Calb
Ylip
Sjap
Spom


GLYCOLYSIS
0.00
0.50
1.00
Glucose
HXK2
GLK1
G6P
PGI1
F6P
PFK2
FBP
FBP
Gly-3-P
DHAP
1,3-BPG
3PG
2PG
PEP
Pyruvate
Acetyl-CoA
forward enzyme
reverse enzyme
bidirectional enzyme
HXK2
GLK1
PGI1
PFK2
FBP
Bb32
Smik
Sbay
SBU
Cgla
Scas
Kwal
Sklu
Klac
Calb
Ylip
Sjap
Spom
0

LESSONS LEARNED

- **process supports efficient development**
- **collaborators' time commitment is front loaded**
- **rapid prototyping is essential**
- **put off coding as long as possible**

contributions

- **Pathline**
  - multiple genes, time points, species, and pathways
- **linearized pathway representation**
- **curvemap**
- **tool deployment**
  - open source
  - used daily by several collaborators

- **visualization design process**
- **types of research contributions**

5.- PAPER TYPES

A Visweek paper typically falls into one of five categories: technique, system, design study, evaluation, or model. We briefly discuss these categories below. Although your main paper type has to be specified during the paper submission process, papers can include elements of more than one of these categories. Please see "Process and Pitfalls in Writing Information Visualization Research Papers" by Tamara Munzner for more detailed discussion on how to write a successful Visweek paper.

Technique papers introduce novel techniques or algorithms that have not previously appeared in the literature, or that significantly extend known techniques or algorithms, for example by scaling to datasets of much larger size than before or by generalizing a technique to a larger class of uses. The technique or algorithm description provided in the paper should be complete enough that a competent graduate student in visualization could implement the work, and the authors should create a prototype implementation of the methods. Relevant previous work must be referenced, and the advantage of the new methods over it should be clearly demonstrated. There should be a discussion of the tasks and datasets for which this new method is appropriate, and its limitations. Evaluation through informal or formal user studies, or other methods, will often serve to strengthen the paper, but are not mandatory.

System papers present a blend of algorithms, technical requirements, user requirements, and design that solves a major problem. The system that is described is both novel and important, and has been implemented. The rationale for significant design decisions is provided, and the system is compared to documented, best-of-breed systems already in use. The comparison includes specific discussion of how the described system differs from and is, in some significant respects, superior to those systems. For example, the described system may offer substantial advancements in the performance or usability of visualization systems, or novel capabilities. Every effort should be made to eliminate external factors (such as advances in processor performance, memory sizes or operating system features) that would affect this comparison. For further suggestions, please review "How (and How Not) to Write a Good Systems Paper" by Roy Levin and David Redell, and "Empirical Methods in CS and AI" by Toby

L4: Data

REQUIRED READING

2

Data Principles

Many aspects of a visualization design are driven by the kind of data that we have at our disposal: what kind of data do we need to look at? What information can we figure out from the data itself, versus the meanings that we must be told explicitly? What high-level concepts will allow us to split datasets apart into general and useful pieces? What kind of attributes does the data have to begin with, and what kinds of derived data might we compute in order to draw a more effective picture?

This chapter approaches these questions with a taxonomy of visualization data types and semantics that meshes well with the principles in Part II and methods in Part III. Figure 2.1 shows the big picture, which will be expanded on in the rest of the chapter.

The chapter begins with a discussion of dataset types, then makes a distinction between semantics and types, and continues with attribute types and attribute semantics. It returns to datasets with semantics. The chapter then covers derived attributes and spaces. It concludes by relating this taxonomy of data principles to the idea of the data abstraction level of the nested design model.

2.1 Dataset Types

Polaris: A System for Query, Analysis, and Visualization of Multidimensional Relational Databases

Chris Stolte, Diane Tang, and Pat Hanrahan

Abstract—In the last several years, large multidimensional databases have become common in a variety of applications such as data warehousing and scientific computing. Analysis and exploration tasks place significant demands on the interfaces to these databases. Because of the size of the data sets, dense graphical representations are more effective for exploration than spreadsheets and charts. Furthermore, because of the exploratory nature of the analysis, it must be possible for the analysts to change visualizations rapidly as they pursue a cycle involving first hypothesis and then experimentation. In this paper, we present Polaris, an interface for exploring large multidimensional databases that extends the well-known Pivot Table interface. The novel features of Polaris include an interface for constructing visual specifications of table-based graphical displays and the ability to generate a precise set of relational queries from the visual specifications. The visual specifications can be rapidly and incrementally developed, giving the analyst visual feedback as they construct complex queries and visualizations.

Index Terms—Database visualization, database analysis, visualization formalism, multidimensional databases.

1 Introduction

In the last several years, large databases have become common in a variety of applications. Corporations are creating large data warehouses of historical data on key aspects of their operations. International research projects such as the Human Genome Project [20] and Digital Sky Survey [31] are generating massive databases of scientific data.

generated from the resulting tables. Visual Insights recently released a new interface for visually exploring projections of data cubes using linked views of bar charts, scatterplots, and parallel coordinate displays [14].

In this paper, we present Polaris, an interface for the exploration of multidimensional databases that extends the Pivot Table interface to directly generate a rich, expressive