Pathline:
A Tool For Comparative Functional Genomics

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

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

Pathway:
Metric Overview
Species
Curvemap
Overlays

Key:
- Genes
  - forward
  - reverse
  - bidirectional
- Metabolites
  - PearsonSubgroup2
  - PearsonALL
roadmap

• background
• data & tasks
• Pathline
• case studies
• conclusions and future work
background
functional genomics

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

gene expression

molecular pathways
gene expression is ...

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

... a single quantitative value
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)
gene expression is ...

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

biologists measure it ...

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

visualized with heatmaps

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

encode value with color
gene expression is ...

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

it is measured ...

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

visualized with heatmaps

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

encode value with color
augmented with clustering
the functioning of a cell is controlled by many interrelated chemical reactions performed by genes

input \rightarrow \text{genes} \rightarrow \text{output / input} \rightarrow \text{genes} \rightarrow \text{output} = \text{cell function}
glycolysis

tca cycle

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?
contributions

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

**linearized pathway representation**
*for comparing quantitative data along a pathway*

**curvemap**
*visual encoding of temporal gene expression*

**validation**
*case studies describing efficiency gains and new biological findings*
data & tasks
**metabolic pathways**
- 10 to 50 pathways of interest
- inputs/outputs called metabolites

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

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

**phylogeny**
tasks

1. study expression data as a time series
   • look for peaks, valleys, time shifts

2. detailed comparison of a limited number of time series
   • filter using pathways
   • filter again using genes or species

3. comparison of similarity scores of genes along a pathway(s)

4. comparison of multiple similarity scores
metabolic pathways
similarity scores
gene expression
phylogeny
Pathline

design decisions
encode quantitative values with spatial position

[Cleveland84] [Lam07]
encode quantitative values with spatial position

[Cleveland84] [Lam07]

instead of a ..

**topological layout**

**linearized pathway**

www.win.tue.nl/~mwestenb/genevis/
encode quantitative values with spatial position

instead of a .

conditioned heatmap
curvemap

courtesy of M. Styczynski from JavaTreeview
jtreeview.sourceforge.net/
Pathline
linearized pathway representation
linearized pathway representation

common axes to compare similarity scores
linearized pathway representation

common axes to compare similarity scores

• bars and circles
linearized pathway representation

common axes to compare similarity scores

• bars and circles
linearized pathway representation

common axes to compare similarity scores

• bars and circles
  • visual layer for attenuation
  • color-code gene direction
linearized pathway representation

common axes to compare similarity scores

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

• multiple similarity scores
linearized pathway representation

common axes to compare similarity scores

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

• multiple similarity scores
• multiple pathways
unroll from pathway to ordered list of nodes

unroll and cut

reinsert

stylized marks and shared coordinate frame
linearized pathway representation

putting it together . . .

• use spatial position for similarity scores instead of topology

• topology is secondary
Pathline
curvemap
curvemap

inspired by heatmaps

- base visual unit is a curve
curvemap

inspired by heatmaps

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

inspired by heatmaps

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

• rows are species
• columns are genes/metabolites
curvemap

inspired by heatmaps

• *base visual unit is a curve*

• *filled, framed line charts to enhance shape perception*

• rows are species

• columns are genes/metabolites

• overlays to enhance trend perception
case studies
whole genome duplication
whole genome duplication
whole genome duplication

both genes
one gene
whole genome duplication
gene-level relationships

PATHWAY  METRIC OVERVIEW
gene-level relationships
gene-level relationships
gene-level relationships
gene-level relationships
gene-level relationships
conclusions and future work
conclusions

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

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

• case studies: efficiency gains and new discoveries
future work

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

questions?

acknowledgements

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

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

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