Visualization Approaches for Biomolecular Data

Alexander Lex

UMass Lowell, November 06, 2013
bio research requires understanding data

but there is so much of it...
methylatio

levels

mRNA

expression

Gene

Protein

expression

The Cancer Genome Atlas

understanding genomics
to improve cancer care

clinical parameters

pathways

methylation levels

mRNA expression

copy number status

mutation status

microRNA expression
Data Visualization

... makes data accessible
... combines strengths of humans and computers
... enables insight
... communicates
## World University Ranking

<table>
<thead>
<tr>
<th>Rank</th>
<th>School Name</th>
<th>Country</th>
<th>Academic reputation</th>
<th>Employer repu</th>
<th>Faculty/stud</th>
<th>Citation</th>
<th>Citations per pub</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Massachusetts Insti</td>
<td>United States</td>
<td>40.00%</td>
<td>17.99%</td>
<td>93.7 (0.94)</td>
<td>19.63%</td>
<td>19.63%</td>
</tr>
<tr>
<td>2</td>
<td>Harvard University</td>
<td>United States</td>
<td>40.00%</td>
<td>32.94%</td>
<td></td>
<td>19.63%</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>University of Camb</td>
<td>United Kingdom</td>
<td>40.00%</td>
<td>32.94%</td>
<td></td>
<td>19.63%</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>Imperial College L</td>
<td>United Kingdom</td>
<td>40.00%</td>
<td>32.94%</td>
<td></td>
<td>19.63%</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>University of Oxf</td>
<td>United Kingdom</td>
<td>40.00%</td>
<td>32.94%</td>
<td></td>
<td>19.63%</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>UCL (University Col)</td>
<td>United Kingdom</td>
<td>40.00%</td>
<td>32.94%</td>
<td></td>
<td>19.63%</td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>Stanford University</td>
<td>United States</td>
<td>40.00%</td>
<td>32.94%</td>
<td></td>
<td>19.63%</td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>Yale University</td>
<td>United States</td>
<td>40.00%</td>
<td>32.94%</td>
<td></td>
<td>19.63%</td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>Princeton Universit</td>
<td>United States</td>
<td>40.00%</td>
<td>32.94%</td>
<td></td>
<td>19.63%</td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>University of Chica</td>
<td>United States</td>
<td>40.00%</td>
<td>32.94%</td>
<td></td>
<td>19.63%</td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>ETH Zurich (Swiss F)</td>
<td>Switzerland</td>
<td>40.00%</td>
<td>32.94%</td>
<td></td>
<td>19.63%</td>
<td></td>
</tr>
<tr>
<td>12</td>
<td>Columbia Universit</td>
<td>United States</td>
<td>40.00%</td>
<td>32.94%</td>
<td></td>
<td>19.63%</td>
<td></td>
</tr>
<tr>
<td>13</td>
<td>University of Penns</td>
<td>United States</td>
<td>40.00%</td>
<td>32.94%</td>
<td></td>
<td>19.63%</td>
<td></td>
</tr>
<tr>
<td>14</td>
<td>Cornell University</td>
<td>United States</td>
<td>40.00%</td>
<td>32.94%</td>
<td></td>
<td>19.63%</td>
<td></td>
</tr>
<tr>
<td>15</td>
<td>University of Edinb</td>
<td>United Kingdom</td>
<td>40.00%</td>
<td>32.94%</td>
<td></td>
<td>19.63%</td>
<td></td>
</tr>
<tr>
<td>16</td>
<td>Ecole Polytechniqu</td>
<td>Switzerland</td>
<td>40.00%</td>
<td>32.94%</td>
<td></td>
<td>19.63%</td>
<td></td>
</tr>
<tr>
<td>17</td>
<td>King's College Lond</td>
<td>United Kingdom</td>
<td>40.00%</td>
<td>32.94%</td>
<td></td>
<td>19.63%</td>
<td></td>
</tr>
<tr>
<td>18</td>
<td>University of Toron</td>
<td>Canada</td>
<td>40.00%</td>
<td>32.94%</td>
<td></td>
<td>19.63%</td>
<td></td>
</tr>
<tr>
<td>19</td>
<td>McGill University</td>
<td>Canada</td>
<td>40.00%</td>
<td>32.94%</td>
<td></td>
<td>19.63%</td>
<td></td>
</tr>
<tr>
<td>20</td>
<td>National University</td>
<td>Singapore</td>
<td>40.00%</td>
<td>32.94%</td>
<td></td>
<td>19.63%</td>
<td></td>
</tr>
<tr>
<td>21</td>
<td>University of Michi</td>
<td>United States</td>
<td>40.00%</td>
<td>32.94%</td>
<td></td>
<td>19.63%</td>
<td></td>
</tr>
<tr>
<td>22</td>
<td>University of Califo</td>
<td>United States</td>
<td>40.00%</td>
<td>32.94%</td>
<td></td>
<td>19.63%</td>
<td></td>
</tr>
<tr>
<td>23</td>
<td>California Institute</td>
<td>United States</td>
<td>40.00%</td>
<td>32.94%</td>
<td></td>
<td>19.63%</td>
<td></td>
</tr>
<tr>
<td>24</td>
<td>University of Bristol</td>
<td>United Kingdom</td>
<td>40.00%</td>
<td>32.94%</td>
<td></td>
<td>19.63%</td>
<td></td>
</tr>
<tr>
<td>25</td>
<td>Duke University</td>
<td>United States</td>
<td>40.00%</td>
<td>32.94%</td>
<td></td>
<td>19.63%</td>
<td></td>
</tr>
</tbody>
</table>
Who am I?

PostDoc @ Harvard, Hanspeter Pfister’s Group
PhD from TU Graz, Austria
Co-leader of Caleydo Project
What is CALEYDO?

Software for analyzing molecular biology data
Software for doing research in visualization
  developed in academic setting
  platform for trying out radically new visualization ideas

Quest for compromise between academic prototyping and ready-to-use software
What is Caleydo?

Open source platform for developing visualization and data analysis techniques

easily extendible due to plug-in architecture

you can create your own views

you can plug-in your own algorithms

http://caleydo.org
The Team

Marc Streit
Johannes Kepler University Linz, AT

Christian Partl
Graz University of Technology, AT

Samuel Gratzl
Johannes Kepler University Linz, AT

Nils Gehlenborg
Harvard Medical School, Boston, USA

Dieter Schmalstieg
Graz University of Technology, AT

Hanspeter Pfister
Harvard University, Cambridge, USA
Rankings are omnipresent
Goal

Intuitive
Interactive
Multi-Attribute
Ranking Visualization
To Create
Refine
Explore
10 Requirements
University

MIT, USA
Harvard, USA
Princeton, USA
Cambridge, UK
Oxford, UK
Encode Rank
10 Requirements

1. Encode Rank
<table>
<thead>
<tr>
<th>Rank</th>
<th>University</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>MIT, USA</td>
</tr>
<tr>
<td>2.</td>
<td>Harvard, USA</td>
</tr>
<tr>
<td>3.</td>
<td>Princeton, USA</td>
</tr>
<tr>
<td>4.</td>
<td>Cambridge, UK</td>
</tr>
<tr>
<td>5.</td>
<td>Oxford, UK</td>
</tr>
</tbody>
</table>
10 Requirements

1. Encode Rank

Encode Cause of Rank
10 Requirements

1. Encode Rank
2. Encode Cause of Rank

Encode Cause of Rank
<table>
<thead>
<tr>
<th>Rank</th>
<th>University</th>
<th>Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>MIT, USA</td>
<td></td>
</tr>
<tr>
<td>2.</td>
<td>Harvard, U</td>
<td></td>
</tr>
<tr>
<td>3.</td>
<td>Princeton,</td>
<td></td>
</tr>
<tr>
<td>4.</td>
<td>Cambridge</td>
<td></td>
</tr>
<tr>
<td>5.</td>
<td>Oxford, UK</td>
<td>4.0</td>
</tr>
</tbody>
</table>
10 Requirements

1. Encode Rank
2. Encode Cause of Rank

Support Multiple Attributes
10 Requirements

1. Encode Rank
2. Encode Cause of Rank
3. Support Multiple Attributes

Support Multiple Attributes
Score = $f(A, B, C)$

<table>
<thead>
<tr>
<th>Rank</th>
<th>University</th>
<th>A</th>
<th>B</th>
<th>C</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>MIT, USA</td>
<td>![MIT A bar]</td>
<td>![MIT B bar]</td>
<td>![MIT C bar]</td>
</tr>
</tbody>
</table>
Combiner functions: \( f(A, B, C) \)

(Weighted) sum
\[
Score = w_a A + w_b B + w_c C
\]

Maximum
\[
Score = \max(A, B, C)
\]

Product

Nesting

\( \rightarrow \) Serial

\( \rightarrow \) Parallel

\( \rightarrow \) Complex Combiners

…
Serial Combiner (as Stacked Bar)

<table>
<thead>
<tr>
<th>Rank</th>
<th>University</th>
<th>A</th>
<th>B</th>
<th>C</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>MIT, USA</td>
<td>Red</td>
<td>Green</td>
<td>Purple</td>
</tr>
<tr>
<td>2.</td>
<td>Harvard, USA</td>
<td>Red</td>
<td>Green</td>
<td>Purple</td>
</tr>
<tr>
<td>3.</td>
<td>Princeton, USA</td>
<td>Red</td>
<td>Green</td>
<td>Purple</td>
</tr>
<tr>
<td>4.</td>
<td>Cambridge, UK</td>
<td>Red</td>
<td>Green</td>
<td>Purple</td>
</tr>
<tr>
<td>5.</td>
<td>Oxford, UK</td>
<td>Red</td>
<td>Green</td>
<td>Purple</td>
</tr>
</tbody>
</table>

The diagram represents the Serial Combiner with weights $w_a, w_b, w_c$ corresponding to the universities MIT, Harvard, Princeton, Cambridge, and Oxford, respectively.
Serial Combiner (as Stacked Bar)

\[ w_a A + w_b B + w_c C \]

<table>
<thead>
<tr>
<th>Rank</th>
<th>University</th>
<th>A</th>
<th>B</th>
<th>C</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>MIT, USA</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2.</td>
<td>Harvard, USA</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3.</td>
<td>Princeton, USA</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4.</td>
<td>Cambridge, UK</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5.</td>
<td>Oxford, UK</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
## Serial Combiner (as Stacked Bar)

The Serial Combiner is represented as a stacked bar chart with weights $w_a A$, $w_b B$, and $w_c C$.

<table>
<thead>
<tr>
<th>Rank</th>
<th>University</th>
<th>A</th>
<th>B</th>
<th>C</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>MIT, USA</td>
<td><img src="#" alt="Bar" /></td>
<td><img src="#" alt="Bar" /></td>
<td><img src="#" alt="Bar" /></td>
</tr>
<tr>
<td>2.</td>
<td>Harvard, USA</td>
<td><img src="#" alt="Bar" /></td>
<td><img src="#" alt="Bar" /></td>
<td><img src="#" alt="Bar" /></td>
</tr>
<tr>
<td>3.</td>
<td>Princeton, USA</td>
<td><img src="#" alt="Bar" /></td>
<td><img src="#" alt="Bar" /></td>
<td><img src="#" alt="Bar" /></td>
</tr>
<tr>
<td>4.</td>
<td>Cambridge, UK</td>
<td><img src="#" alt="Bar" /></td>
<td><img src="#" alt="Bar" /></td>
<td><img src="#" alt="Bar" /></td>
</tr>
<tr>
<td>5.</td>
<td>Oxford, UK</td>
<td><img src="#" alt="Bar" /></td>
<td><img src="#" alt="Bar" /></td>
<td><img src="#" alt="Bar" /></td>
</tr>
</tbody>
</table>
### Parallel Combiner (as Multi Bar)

<table>
<thead>
<tr>
<th>Rank</th>
<th>University</th>
<th>MAX(A,B,C)</th>
<th>B</th>
<th>C</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>MIT, USA</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2.</td>
<td>Harvard, USA</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3.</td>
<td>Princeton, USA</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4.</td>
<td>Cambridge, UK</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5.</td>
<td>Oxford, UK</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Parallel Combiner (as Multi Bar)

<table>
<thead>
<tr>
<th>Rank</th>
<th>University</th>
<th>MAX(A,B,C)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>MIT, USA</td>
<td>![Bar Chart]</td>
</tr>
<tr>
<td>2.</td>
<td>Harvard, USA</td>
<td>![Bar Chart]</td>
</tr>
<tr>
<td>3.</td>
<td>Princeton, USA</td>
<td>![Bar Chart]</td>
</tr>
<tr>
<td>4.</td>
<td>Cambridge, UK</td>
<td>![Bar Chart]</td>
</tr>
<tr>
<td>5.</td>
<td>Oxford, UK</td>
<td>![Bar Chart]</td>
</tr>
</tbody>
</table>
Parallel Combiner (as Multi Bar)

<table>
<thead>
<tr>
<th>Rank</th>
<th>University</th>
<th>MAX(A,B,C)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>MIT, USA</td>
<td></td>
</tr>
<tr>
<td>2.</td>
<td>Harvard, USA</td>
<td></td>
</tr>
<tr>
<td>3.</td>
<td>Princeton, USA</td>
<td></td>
</tr>
<tr>
<td>4.</td>
<td>Cambridge, UK</td>
<td></td>
</tr>
<tr>
<td>5.</td>
<td>Oxford, UK</td>
<td></td>
</tr>
</tbody>
</table>
10 Requirements

1. Encode Rank
2. Encode Cause of Rank
3. Support Multiple Attributes

Interactive Refinement
and Visual Feedback
10 Requirements

1. Encode Rank
2. Encode Cause of Rank
3. Support Multiple Attributes
4. Interactive Refinement and Visual Feedback

Interactive Refinement and Visual Feedback
<table>
<thead>
<tr>
<th>Rank</th>
<th>School Name</th>
<th>Country</th>
<th>Faculty/student ratio</th>
<th>Employer reputation</th>
<th>Citations per faculty</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>American University</td>
<td>United States</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>Arizona State University</td>
<td>United States</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>Aston University</td>
<td>United Kingdom</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>Birkbeck College, University of L</td>
<td>United Kingdom</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>Boston College</td>
<td>United States</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>Boston University</td>
<td>United States</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>Brandeis University</td>
<td>United States</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>Brown University</td>
<td>United States</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>Brunel University</td>
<td>United Kingdom</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>California Institute of Technology</td>
<td>United States</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>Cardiff University</td>
<td>United Kingdom</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>12</td>
<td>Case Western Reserve University</td>
<td>United States</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>13</td>
<td>City University London</td>
<td>United Kingdom</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>14</td>
<td>College of William &amp; Mary</td>
<td>United States</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>15</td>
<td>Colorado State University</td>
<td>United States</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>16</td>
<td>Columbia University</td>
<td>United States</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>17</td>
<td>Cornell University</td>
<td>United States</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>18</td>
<td>Cranfield University</td>
<td>United Kingdom</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>19</td>
<td>Dartmouth College</td>
<td>United States</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>20</td>
<td>Drexel University</td>
<td>United States</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>21</td>
<td>Duke University</td>
<td>United States</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>22</td>
<td>Durham University</td>
<td>United Kingdom</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Flexible Mapping of Attributes to Scores
10 Requirements

1. Encode Rank
2. Encode Cause of Rank
3. Support Multiple Attributes
4. Interactive Refinement and Visual Feedback
5. Flexible Mapping of Attributes to Scores

Flexible Mapping of Attributes to Scores
<table>
<thead>
<tr>
<th>Rank</th>
<th>School Name</th>
<th>Country</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>University of Oxford</td>
<td>United Kingdom</td>
</tr>
<tr>
<td>2.</td>
<td>University of Cambridge</td>
<td>United Kingdom</td>
</tr>
<tr>
<td>3.</td>
<td>Harvard University</td>
<td>United States</td>
</tr>
<tr>
<td>4.</td>
<td>Stanford University</td>
<td>United States</td>
</tr>
<tr>
<td>5.</td>
<td>Massachusetts Institute of Technology (MIT)</td>
<td>United States</td>
</tr>
</tbody>
</table>
10 Requirements

1. Encode Rank
2. Encode Cause of Rank
3. Support Multiple Attributes
4. Interactive Refinement and Visual Feedback
5. Flexible Mapping of Attributes to Scores

Compare Rankings
10 Requirements

1. Encode Rank
2. Encode Cause of Rank
3. Support Multiple Attributes
4. Interactive Refinement and Visual Feedback
5. Flexible Mapping of Attributes to Scores
6. Compare Rankings

Compare Rankings
<table>
<thead>
<tr>
<th>Rank</th>
<th>University</th>
<th>Score</th>
<th>Change</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>MIT, USA</td>
<td></td>
<td>(+1)</td>
</tr>
<tr>
<td>2</td>
<td>Harvard, USA</td>
<td>(−1)</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>Princeton, USA</td>
<td>(−1)</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>Cambridge, UK</td>
<td>(−2)</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>Oxford, UK</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
### Bump Charts

<table>
<thead>
<tr>
<th>Rank</th>
<th>University</th>
<th>Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>MIT, USA</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>Harvard, USA</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>Princeton, USA</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>Cambridge, UK</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>Oxford, UK</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Rank</th>
<th>Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td></td>
</tr>
</tbody>
</table>
Video showing:
• Creating snapshot for comparison
• Play with weights
• Show delta
• Select by clicking on slopegraph
10 Requirements

1. Encode Rank
2. Encode Cause of Rank
3. Support Multiple Attributes
4. Interactive Refinement and Visual Feedback
5. Flexible Mapping of Attributes to Scores
6. Compare Rankings
7. Scalability
8. Filtering
9. Handle Missing Values
10. Optimization
Demos, Videos & More:
http://lineup.caleydo.org
Cancer Subtypes

Cancer is not homogeneous
- different histology
- different molecular alterations

Subtypes have serious implications
- different treatment for subtypes
- prognosis varies between subtypes

StratomeX
Cancer Subtype Analysis

many different types of data

for

large numbers of patients.
Stratification of a Single Dataset

Subtypes are identified by stratifying datasets, e.g., based on an expression pattern, a mutation status, a copy number alteration, or a combination of these.
Patients

Genes

Candidate Subtype / Heat Map

Header / Summary of Whole Stratification
Stratification of Multiple Datasets

Cluster A1

Cluster A2

Cluster A3

Tabular, e.g., mRNA

B1

B2

Categorical, e.g., mutation status
Stratification of Multiple Datasets

Cluster A1
Cluster A2
Cluster A3

B1
B2

Tabular
e.g., mRNA

Categorical,
e.g., mutation status
Clustering of mRNA Data

Stratification on Copy Number Status

Many shared Patients
Stratification of Multiple Datasets

Cluster A1

Cluster A2

Cluster A3

Tabular
e.g., mRNA

Categorical,
e.g., mutation status

B1

B2
Stratification of Multiple Datasets

Cluster A1
Cluster A2
Cluster A3

B1
B2

Dep. C1
Dep. C2

Tabular, e.g., mRNA
Categorical, e.g., mutation status
Dependent Data, e.g. clinical data
Survival data in Kaplan Meier plots
Dependent Pathway
Stratification based on clinical variable (gender)
How to Choose Stratifications?

~ 15 clusterings per matrix
~ 15,000 stratifications for copy number & mutations
~ 500 pathways
~ 20 clinical variables

Calculating scores for matches
Ranking the results
Algorithms for finding...

- matching stratification
- matching subtype
- mutual exclusivity
- relevant pathway
- stratification with effect in survival
- high/low structural variation
Live-Demo!

http://stratomex.aleydo.org
Entourage
[Lex, InfoVis ‘13]

enRoute
[Partl, BioVis ‘12]
[Partl, BMC Bioinformatics ‘13]

many pathways
pathway cross-connections

experimental data on pathways
Background
Background
Challenges

How to visualize multiple pathways at the same time?

How to visualize pathway relationships?
How to visualize experimental data on pathways?

Challenges

Experimental data analysis
How to visualize multiple pathways at the same time?

How to visualize pathway relationships?

How to visualize experimental data on pathways?
Approaches

Whole Network

[Kono2009]

Connected Pathways

[Klukas2006]
Contextual Subsets

Focus Pathway

Pathway A

B ────── C
  |  |
  |  v
  E ────── D

  F

Context Pathway

Pathway B

G ────── H
  |  |
  |  v
  E ────── A

  J ────── I

PW B

G

Contextual Subset

J

H

I
Contextual Subsets

Focus Pathway

Pathway A

A

B

C

D

E

F

Context Pathways

PW C

M

E

K

L

PW D

B

P

E

N
Levels of Detail

High

Medium

Low
How to visualize multiple pathways at the same time?

How to visualize pathway relationships?

How to visualize experimental data on pathways?
Finding Related Pathways
Finding Related Pathways
Visualizing Relationships
Visualizing Relationships
Visualizing Relationships

[Collins2009]
Visualizing Relationships
How to visualize multiple pathways at the same time?

How to visualize pathway relationships?

How to visualize experimental data on pathways?
Cannot account for variation found in real-world data
Branches can be (in)activated due to mutation, changed gene expression, modulation due to drug treatment, etc.

Experiment Data and Pathways enRoute [Partl, BioVis ‘12]
Good Old Color Coding

<table>
<thead>
<tr>
<th></th>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
<th>E</th>
<th>F</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>-3.4</td>
<td>2.8</td>
<td>3.1</td>
<td>-3</td>
<td>0.5</td>
<td>0.3</td>
</tr>
</tbody>
</table>
Challenge: Data Scale

Large number of experiments
  Large datasets have more than 500 experiments
Multiple groups/conditions
Challenge: Data Heterogeneity

Different types of data, e.g.,

- mRNA expression: numerical
- Mutation status: categorical
- Copy number variation: ordered categorical
- Metabolite concentration: numerical

Require different visualization techniques
Challenge: Supporting Multiple Tasks

Two central tasks:

- Explore **topology** of pathway
- Explore the **attributes** of the nodes (experimental data)

Need to support both!
Pathway View

enRoute View

Concept
Experimental Data Representation

Gene Expression Data (Numerical)

Copy Number Data (Ordered Categorical)

Mutation Data
Live-Demo!

http://entourage.caleydo.org
http://enroute.caleydo.org
Visualization Approaches for Biomolecular Data

Alexander Lex, Harvard University
alex@seas.harvard.edu
http://alexander-lex.com
@alexander_lex

Credits:
Marc Streit, Nils Gehlenborg, Hanspeter Pfister, Anne Mai Wasserman, Mark Borowsky, Christian Partl, Denis Kalkofen, Samuel Gratzl, Dieter Schmalstieg

Harvard
School of Engineering and Applied Sciences
On-Node Mapping

Path highlighting with Bubble Sets [Collins2009]
enRoute View

Path Representation
Experimental Data Representation

enRoute View

Diagram of experimental data representation with groups and datasets.
Case Study