Pathview: a pathway based data integration and visualization tool

Weijun Luo*, Cory Brouwerer
Email: luo_weijun@yahoo.com

UNC Charlotte
Bioinformatics Service Division
Dept. Bioinformatics & Genomics
Agenda

• Overview

• Main features
  – Basic visualization
  – Data integration
  – Integrated pathway analysis

• Implementation & impact
Overview
Pathview package

Luo et al. Bioinformatics 2013

• Pathway visualization
  – Informative and readable: KEGG based
  – Publication quality: full graphics

• Strong data integration
  – different data types, formats, attributes
  – numerous ID types, species etc.

• Automated/integrated pathway analysis
Design

```
pathview(gene.data = gse16873.d, pathway.id = "04110")
```
Basic visualization
Data integration
KEGG view: multiple samples

Data on KEGG graph
Rendered by Pathview
Graphviz view: multiple samples

Node types:
- gene (protein/enzyme)
- group (complex)
- compound (metabolite/glycan)
- map (pathway)

Pathway name
Data attributes
## Examples of 3000 species supported

<table>
<thead>
<tr>
<th>KEGG Code</th>
<th>Scientific Name</th>
<th>Common Name</th>
<th>KEGG GeneID</th>
<th>Entrez GeneID</th>
</tr>
</thead>
<tbody>
<tr>
<td>hsa</td>
<td><em>Homo sapiens</em></td>
<td>human</td>
<td>100</td>
<td>100</td>
</tr>
<tr>
<td>mmu</td>
<td><em>Mus musculus</em></td>
<td>mouse</td>
<td>100042069</td>
<td>100042069</td>
</tr>
<tr>
<td>rno</td>
<td><em>Rattus norvegicus</em></td>
<td>rat</td>
<td>100359539</td>
<td>100359539</td>
</tr>
<tr>
<td>xla</td>
<td><em>Xenopus laevis</em></td>
<td>African clawed frog</td>
<td>100037030</td>
<td>100037030</td>
</tr>
<tr>
<td>dre</td>
<td><em>Danio rerio</em></td>
<td>zebrafish</td>
<td>1000000978</td>
<td>1000000978</td>
</tr>
<tr>
<td>dme</td>
<td><em>Drosophila melanogaster</em></td>
<td>fruit fly</td>
<td>Dmel_CG10078</td>
<td>38753</td>
</tr>
<tr>
<td>cel</td>
<td><em>Caenorhabditis elegans</em></td>
<td>nematode</td>
<td>CELE_AH6.1</td>
<td>191639</td>
</tr>
<tr>
<td>ath</td>
<td><em>Arabidopsis thaliana</em></td>
<td>thale cress</td>
<td>AT1G01210</td>
<td>839481</td>
</tr>
<tr>
<td>sce</td>
<td><em>Saccharomyces cerevisiae</em></td>
<td>budding yeast</td>
<td>YAL038W</td>
<td>851193</td>
</tr>
<tr>
<td>eco</td>
<td><em>Escherichia coli</em> K-12 MG1655</td>
<td></td>
<td>b0049</td>
<td>944770</td>
</tr>
</tbody>
</table>
KEGG Orthology with metagenomics
Integrated workflows for pathway analysis
Pathway analysis workflows (with Pathview)

- Transcriptomics
  - Microarray
  - RNA-Seq
- Genomics
  - GWAS
  - CNV
- Metabolomics
- Proteomics
- Epigenomics

- Related tools:
  - GAGE
  - GSEA
  - clusterProfiler
  - opm
  - CompGO
  - GeneProf
  - ...
RNA-Seq workflows with GAGE/Pathview

-Tutorial in BioC gage package
-Google: rna-seq pathway

![Diagram showing RNA-Seq workflows with GAGE/Pathview]

Key steps:
- Read mapping (TopHat2)
- Read counting
- Preprocessing
  - Normalization
  - Diff. expression
- Gene set test (GAGE)
  - Pathways
  - GO terms
  - Coexp. genes
- Visualization (GAGE)
  - Heatmap
  - Scatterplot
- Pathview
  - Pathway graph
Implementation & Impact
Software package: pathview

• Status
  – Current release/devel version: 1.4.1/1.5.2
  – Initial release: BioC 2.12 (R-3.0) (04/2013)
  – Multiple platforms, complete doc and help
  – Actively maintained and improved

• Available:
  – Google: pathview bioconductor
Global Impact

• Bioinformatics Most-Read (June-October, 2013).
• Related tools:
  – R/BioC: clusterProfiler, CompGO, gage, opm
  – Web: GeneProf
• 10,000 downloads past year
• Inquiries/refers:
  – emails, BioC list
  – SEQanswers, Biostars
  – online blogs, tutorials
Course adoptions

Techniques in Glycobiology
Prof. Jennifer Van Eyk, USA

Environmental Genomics
Prof. John Colbourne, USA

Cancer Bioinformatics
Prof. Ying Xu, China

Functional genomics
Prof. Koji Kadota, Japan
Multiplexed Cancer Pathway Analysis
nCounter PanCancer Pathways Panel for Gene Expression

Differential Gene Expression Mapped to Proteins in the Cell Cycle Pathway

Data on KEGG graph
Rendered by Pathview
Copyright © 1995-2014 Konomia Laboratories

http://www.nanostring.com/forms/pancancer/download
Future Development

• R/Bioc interface improvement
• Multiple user interfaces
  – Web
  – Galaxy
• More node/edge attributes
• More pathway databases besides KEGG
Acknowledgements

UNC Charlotte
Bioinformatics Service Division
Dept. Bioinformatics & Genomics

R/Bioconductor teams
GAGE/Pathview users

All of you!

Questions/comments?
luo_weijun@yahoo.com
## Data integration/processing capacity

<table>
<thead>
<tr>
<th>Data Attribute</th>
<th>Pathview Coverage</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Type</strong></td>
<td>Gene, protein, metabolite, genetics, literature, and others</td>
</tr>
<tr>
<td><strong>ID type</strong></td>
<td>12 gene, 21 compound</td>
</tr>
<tr>
<td><strong>Species</strong></td>
<td>3000 KEGG species and ortholog</td>
</tr>
<tr>
<td><strong>Value</strong></td>
<td>Continuous, discrete, 1 or 2 directions</td>
</tr>
<tr>
<td><strong>Format</strong></td>
<td>Matrix, vector</td>
</tr>
</tbody>
</table>
Pathway analysis visualized

Before

Now
Conclusion
Pathview

Luo et al. Bioinformatics 2013

• Informative, readable, publishable graphs

• Strong data integration

• Integrated pathway analysis for omics data

• Global impact, widely adopted