MetPA: a web-based metabolomics tool for pathway analysis and visualization
Jianguo Xia\textsuperscript{1} and David S. Wishart\textsuperscript{1,2,3, *}
\textsuperscript{1} Department of Biological Sciences, \textsuperscript{2}Department of Computing Sciences and \textsuperscript{3}National Research Council, National Institute for Nanotechnology (NINT), University of Alberta, Canada

Reading Project of Machine Learning & Bioinformatics
Report Date: 2010/11/10

\textbf{Student name & ID:} 林方偉 \textbf{N26991653}
白逸翰 \textbf{N26991700}
姚宗儒 \textbf{N26991718}
ABSTRACT

- MetPA (Metabolomics Pathway Analysis) is a user-friendly, web-based tool dedicated to the analysis and visualization of metabolomic data within the biological context of metabolic pathways.

- The results are presented in a Google-map style. supports intuitive and interactive data exploration through point-and-click, dragging and lossless zooming.
A comprehensive compound library for metabolite name conversion.

MetPA currently enables analysis and visualization of 874 metabolic pathways, covering 11 common model organisms.
INTRODUCTION

- pathway analysis has emerged as an invaluable aid to understanding the data generated from various ‘omics’ technologies.
a number of robust software tools have been developed to support pathway analysis for genomics and proteomics studies.

- The first pathway analysis tools were typically designed to take a list of differentially expressed genes and compare the number of differentially expressed genes detected in each pathway.

- Second-generation pathway analysis tools typically use normalized gene expression data to calculate the expression of biological pathways in association with phenotypes.
MetPA (Metabolomics Pathway Analysis), a dedicated pathway analysis and visualization tool to facilitate the use of these relatively new and powerful methods in metabolomic studies.
Method

- 2.1 Pathway analysis
- 2.2 Pathway library construction and visualization
- 2.3 Implementation of web application
2.1 Pathway analysis

Pathway analyses in MetPA are conducted through three routes.
(1) Pathway enrichment analysis
(2) Overrepresentation analysis
(3) Pathway topological analysis

Pathway impact = \( \frac{\text{(the sum of the importance measures of the matched metabolites normalized)}}{\text{(the sum of the importance measures of all metabolites in each pathway)}} \)
Pathway enrichment analysis usually refers to quantitative enrichment analysis directly based on the compound concentration values as compared to the compound lists used by over representation analysis.

It is usually more sensitive than over-representation analysis and has the potential to discover "subtle but consistent" changes among compounds within the same biological pathway.
(2) Overrepresentation analysis

- Over-representation analysis is to test if a particular group of compounds is represented more than expected by chance within the user uploaded compound list.

- In the context of pathway analysis, they are testing if compounds involved in a particular pathway is enriched compared by random hits. The most common methods for such analysis is Fishers' exact test and hypergeometric test.
MetPA’s pathway topological analysis is based on the centrality measures of a metabolite in a given metabolic network.

Centrality is a local quantitative measure of the position of a node relative to the other nodes, and is often used to estimate a node’s relative importance or role in network organization.

Since metabolic networks are directed graphs, MetPA uses relative betweenness centrality and out degree centrality measures to calculate compound importance.
2.2 Pathway library construction and visualization

- The current library contains 874 metabolic pathways from 11 model organisms including humans, mouse, Drosophila, Arabidopsis, Escherichia coli, etc.

- This visualization system supports lossless zooming, dragging and linking operations based on Ajax(Asynchronous JavaScript with XML) technology.
2.3-1 Implementation of web application

MetPA’s web interface was implemented using the JSF or Java Server Faces (http://java.sun.com/javaee/javaserverfaces) framework.

The pathway analysis algorithms were implemented in the R programming language (http://www.r-project.org/).

The communication between R and Java was established through the Rserve TCP/IP server (http://www.rforge.net/Rserve/).
The web application is platform independent and has been successfully tested on Mozilla Firefox 3.0+, Safari 4.0+, Google-Chrome 5.0+, Opera 10.0+ and Internet Explorer 8.0.
Example analysis

- MetPA provides a Google-map style network visualization system which implements a comprehensive three-level view
  - metabolome view
  - pathway view
  - compound view
What are the input requirements of MetPA?

MetPA accepts either a list of compound labels (common names, HMDB IDs or KEGG IDs) with one compound per row, or a compound concentration table with samples in rows and compounds in columns. The second column must be phenotype labels (binary, multi-group, or continuous). The table is uploaded as comma separated values (.csv).
An illustration is shown below:

<table>
<thead>
<tr>
<th>Patient ID</th>
<th>Muscle loss</th>
<th>Acetic acid</th>
<th>Acetone</th>
<th>Adipic acid</th>
<th>L-Alanine</th>
<th>L-Asparagine</th>
<th>Betaine</th>
<th>L-Carnitine</th>
</tr>
</thead>
<tbody>
<tr>
<td>Patient_1</td>
<td>1</td>
<td>126.61</td>
<td>9.5</td>
<td>38.11</td>
<td>313.78</td>
<td>158.75</td>
<td>110.06</td>
<td>264.06</td>
</tr>
<tr>
<td>Patient_2</td>
<td>1</td>
<td>212.24</td>
<td>11.87</td>
<td>325.61</td>
<td>872.16</td>
<td>157.76</td>
<td>245.13</td>
<td>120.61</td>
</tr>
<tr>
<td>Patient_3</td>
<td>1</td>
<td>314.33</td>
<td>4.44</td>
<td>131.62</td>
<td>465.9</td>
<td>89.47</td>
<td>116.36</td>
<td>25.07</td>
</tr>
<tr>
<td>Patient</td>
<td>1</td>
<td>37.32</td>
<td>206.47</td>
<td>143.32</td>
<td>592.16</td>
<td>272.93</td>
<td>220.1</td>
<td></td>
</tr>
<tr>
<td>Patient</td>
<td>1</td>
<td>409.4</td>
<td>44.1</td>
<td>15.02</td>
<td>1113.64</td>
<td>83.6</td>
<td>85.17</td>
<td></td>
</tr>
<tr>
<td>Control_1</td>
<td>0</td>
<td>51.47</td>
<td>3.25</td>
<td>8.42</td>
<td>71.51</td>
<td>150.51</td>
<td>127.21</td>
<td></td>
</tr>
<tr>
<td>Control_2</td>
<td>0</td>
<td>96.01</td>
<td>5.72</td>
<td>8.7</td>
<td>34.55</td>
<td>42.61</td>
<td>22.89</td>
<td></td>
</tr>
<tr>
<td>Control_3</td>
<td>0</td>
<td>7.48</td>
<td>2.79</td>
<td>69.53</td>
<td>13.89</td>
<td>15.27</td>
<td>9.85</td>
<td></td>
</tr>
<tr>
<td>Control_4</td>
<td>0</td>
<td>9.94</td>
<td>NA</td>
<td>8.23</td>
<td>102.62</td>
<td>32.87</td>
<td>31.17</td>
<td>7.33</td>
</tr>
<tr>
<td>Control_5</td>
<td>0</td>
<td>100.34</td>
<td>9.16</td>
<td>14.62</td>
<td>961.2</td>
<td>222.41</td>
<td>149.29</td>
<td>488.08</td>
</tr>
<tr>
<td>Control_6</td>
<td>0</td>
<td>27.96</td>
<td>6.52</td>
<td>18.47</td>
<td>163.55</td>
<td>32.1</td>
<td>218.86</td>
<td>230.61</td>
</tr>
</tbody>
</table>
Demo

http://metpa.metabolomics.ca/MetPA/faces/Home.jsp
Conclusion

- MetPA is a full-featured, easy-to-use pathway analysis and visualization environment that combines advanced statistical enrichment analysis with pathway topological characteristics to help researchers identify the most relevant pathways involved in the conditions under study.
Thanks for your attention