Heatmaps – The Gene Expression Edition

Assoc. Prof. Dr. Ilker BUYUK
What are Heat Maps?

The data on the left is the same as that on the right—but one is much easier to understand.
When (and Why) to use Heat Maps?

Heatmaps are used to show relationships between two variables, one plotted on each axis. By observing how cell colors change across each axis, you can observe if there are any patterns in value for one or both variables.
Heat maps is thought to have originated in the 19th century, where manual gray-scale shading was used to depict data patterns in matrices and tables.

AN EARLY HEAT MAP SHOWING POPULATION DENSITY IN 19TH CENTURY PARIS DISTRICTS.
The example heatmap depicts the daily precipitation distribution, grouped by month, and recorded over eleven years in Seattle, Washington.*

<table>
<thead>
<tr>
<th>Month</th>
<th>&lt; 0.1</th>
<th>0.1 - 4.0</th>
<th>4.1 - 10.0</th>
<th>10.1 - 20.0</th>
<th>&gt; 20.0</th>
</tr>
</thead>
<tbody>
<tr>
<td>January</td>
<td>255</td>
<td>167</td>
<td>123</td>
<td>65</td>
<td>41</td>
</tr>
<tr>
<td>February</td>
<td>244</td>
<td>196</td>
<td>89</td>
<td>52</td>
<td>12</td>
</tr>
<tr>
<td>March</td>
<td>268</td>
<td>198</td>
<td>119</td>
<td>46</td>
<td>20</td>
</tr>
<tr>
<td>April</td>
<td>321</td>
<td>179</td>
<td>88</td>
<td>31</td>
<td>11</td>
</tr>
<tr>
<td>May</td>
<td>413</td>
<td>152</td>
<td>51</td>
<td>31</td>
<td>4</td>
</tr>
<tr>
<td>June</td>
<td>432</td>
<td>137</td>
<td>42</td>
<td>15</td>
<td>4</td>
</tr>
<tr>
<td>July</td>
<td>567</td>
<td>63</td>
<td>14</td>
<td>7</td>
<td>6</td>
</tr>
<tr>
<td>August</td>
<td>550</td>
<td>63</td>
<td>24</td>
<td>8</td>
<td>6</td>
</tr>
<tr>
<td>September</td>
<td>454</td>
<td>114</td>
<td>39</td>
<td>16</td>
<td>7</td>
</tr>
<tr>
<td>October</td>
<td>337</td>
<td>165</td>
<td>81</td>
<td>44</td>
<td>24</td>
</tr>
<tr>
<td>November</td>
<td>238</td>
<td>162</td>
<td>110</td>
<td>65</td>
<td>55</td>
</tr>
<tr>
<td>December</td>
<td>253</td>
<td>176</td>
<td>113</td>
<td>64</td>
<td>35</td>
</tr>
</tbody>
</table>

*Seattle precipitation by month, 1998-2018
Best practices for using a heatmap

Choose an appropriate color palette
Best practices for using a heatmap

Include a legend
Show values in cell

- There is a lack of precision for mapping color to value, especially compared to other encodings like position or length. Where possible, it is a good idea to add cell value annotations to the heatmap as a double encoding of value.
Sort levels by similarity or value

The right-side heatmap is sorted by the last column value.
Illustration of gene expression data

- A heat map is a well-received approach to illustrate gene expression data.
- It is an impressive visual exhibit that addresses explosive amounts of NGS data.
- It’s packed with closely set patches in shades of colors, pomping the gene expression data of multifarious high-throughput tryouts.
• The individual tiles or rectangles in a heat map are scaled with a range of colors proportionate to gene expression values.

• The outcome makes a pitch to check upon the rows, columns, and joint structural patterns. It’s a tool used widely by the statisticians and bioinformatics scientists to make sense of large multi-dimensional datasets. It helps one to pry out groups among the genes and the samples handled by setting up associations.
- The gene sequences correspond to the rows of the matrix and the chips/samples correspond to the columns.

- A colored matrix display represents the matrix of values as a grid; the number of rows is equal to the number of genes being analyzed, and the number of columns is equal to the number of chips.

- The boxes of the grid are colored according to the numerical value in the corresponding matrix cell.

- Genes are represented in rows of the matrix and chips/samples in the columns.

- What you get is a framework in a color matrix.

- This will look like a grid of boxes, colored to the gene expression values.
• You will be able to pick genes based on their expression levels under different conditions.

• Many do not change, either because they are not on or are need to be on all the time.

• Those that do change are of the most interest. These represent the signature tune of gene expression affiliated to a particular condition.
Online Tools

• CIMminer
• HeatmapGenerator
• MultiExperiment Viewer
• Bioconductor
• Morpheus
• MeV
• GenePatterns
• ExpressionProfiler
• geWorkbench
• DAVID
• Babelomics
• GoMiner
• ConceptGen
• GSEA
• EXPANDER
• ClustVis

Databases

• GEO
• ArrayExpress
• Expression Atlas
CIMminer generates color-coded Clustered Image Maps (CIMs) ("heat maps") to represent "high-dimensional" data sets such as gene expression profiles. We introduced CIMs in the mid-1990's for data on drug activity, target expression, gene expression, and proteomic profiles. Clustering of the axes brings like together with like to create patterns of color. (Weinstein, et al., Science 1997; 275:343-349). To learn more details, look at the following articles.
Step 1: Select input file
(Detailed description of the input file is available.)
Input should be a text file with '.txt' extension (example file). CIMminer only accepts tab delimited text files.
(limit 1000 rows)

Step 2: Your E-mail Address

Step 3: Fine tune image
 ● Use defaults  ○ Advanced options
CIMminer Success Result - Refld 1219264670

webadmin@discover.nci.nih.gov

ACL: ben

CIMminer processing is completed. Your Reference Id is: 1219264670.

To view your CIM, click this: https://discover.nci.nih.gov/cimMiner/cimMinerResultLoad.do?Id=1219264670&Dir=cimminerdata.

To download your CIM, click this: https://discover.nci.nih.gov/cimminerdata/1219264670/download.zip.

Your result will be available for next one week. After that it will be deleted from our system.
Organism: Phaseolus vulgaris

Transcript Name: Phvul.002G116000.1 (primary)

Location: Chr02:24790146..24795421 reverse

Alias: Phvul.002G116000.v1.0  Phvul.002G116000.1.v1.0

Description: (1 of 1) PTHR32002:SF13 - PROTEIN RKD5

Links: B  M

<table>
<thead>
<tr>
<th>Experiment Group</th>
<th>Experiment Name</th>
<th>FPKM</th>
<th>Locus DE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tissue Sample</td>
<td>Leaves</td>
<td>0.844</td>
<td></td>
</tr>
<tr>
<td>Tissue Sample</td>
<td>Nodules</td>
<td>0.744</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Flower Buds</td>
<td>Flowers</td>
<td>Green Mature Pods</td>
</tr>
<tr>
<td>-------</td>
<td>-------------</td>
<td>---------</td>
<td>-------------------</td>
</tr>
<tr>
<td>Pvul-NIN-1</td>
<td>2.299</td>
<td>0.797</td>
<td>2.722</td>
</tr>
<tr>
<td>Pvul-NIN-2</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Pvul-NIN-4</td>
<td>1.149</td>
<td>1.999</td>
<td>3.249</td>
</tr>
<tr>
<td>Pvul-NIN-6</td>
<td>0.035</td>
<td>0.019</td>
<td>0.01</td>
</tr>
<tr>
<td>Pvul-NIN-8</td>
<td>0.959</td>
<td>0.248</td>
<td>0.198</td>
</tr>
<tr>
<td>Pvul-NIN-9</td>
<td>0.241</td>
<td>0.304</td>
<td>0</td>
</tr>
<tr>
<td>Pvul-NIN-10</td>
<td>0.092</td>
<td>0.117</td>
<td>0.072</td>
</tr>
<tr>
<td>Pvul-NIN-11</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Pvul-NIN-12</td>
<td>0.111</td>
<td>0.153</td>
<td>0.038</td>
</tr>
</tbody>
</table>
Thanks for listening