

Practical Bioinformatics

Mark Voorhies

4/28/2017

Pearson similarity

$$s(x, y) = \frac{\sum_i^N (x_i - x_{offset})(y_i - y_{offset})}{\sqrt{\sum_i^N (x_i - x_{offset})^2} \sqrt{\sum_i^N (y_i - y_{offset})^2}}$$

Pearson similarity

$$s(x, y) = \frac{\sum_i^N (x_i - x_{offset})(y_i - y_{offset})}{\sqrt{\sum_i^N (x_i - x_{offset})^2} \sqrt{\sum_i^N (y_i - y_{offset})^2}}$$

Pearson distance

$$d(x, y) = 1 - s(x, y)$$

Pearson similarity

$$s(x, y) = \frac{\sum_i^N (x_i - x_{offset})(y_i - y_{offset})}{\sqrt{\sum_i^N (x_i - x_{offset})^2} \sqrt{\sum_i^N (y_i - y_{offset})^2}}$$

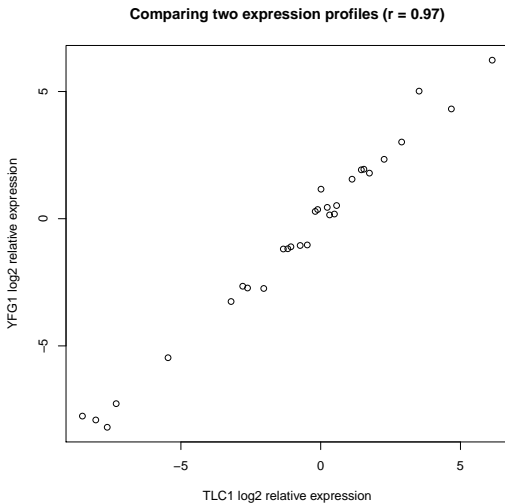
Pearson distance

$$d(x, y) = 1 - s(x, y)$$

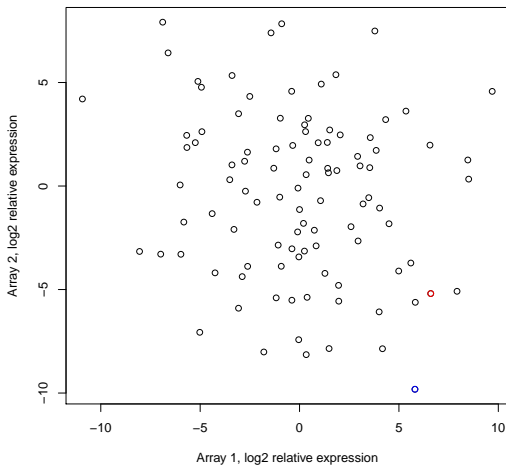
Euclidean distance

$$\frac{\sum_i^N (x_i - y_i)^2}{N}$$

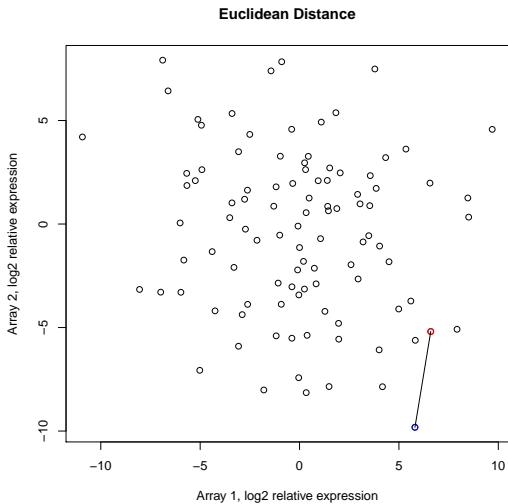
Comparing all measurements for two genes



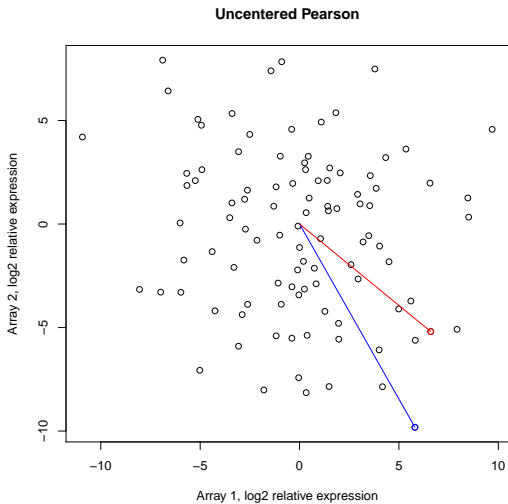
Comparing all genes for two measurements



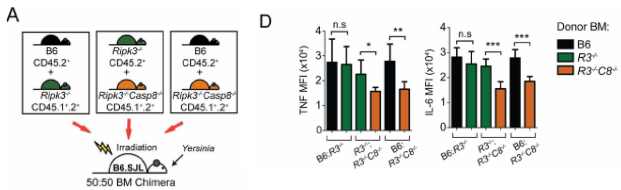
Comparing all genes for two measurements



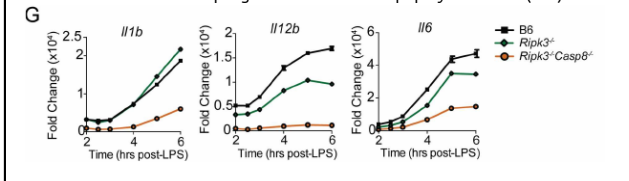
Comparing all genes for two measurements



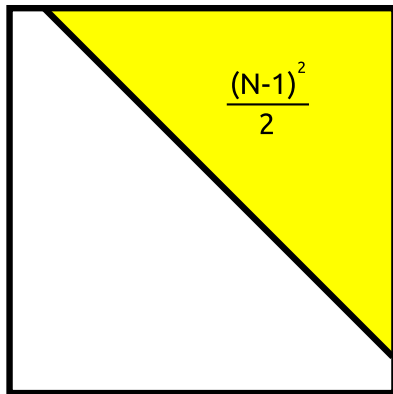
1) *in vivo* Lower inflammatory response for *Ripk3*^{-/-}/*Casp8*^{-/-} transplanted macrophages in mice infected with *Yersinia pestis*



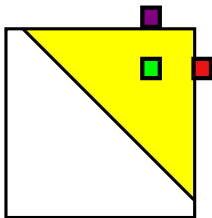
2) *ex vivo* Lower inflammatory response *at the mRNA level* in *Ripk3*^{-/-}/*Casp8*^{-/-} macrophages stimulated with lipopolysaccharide (LPS)



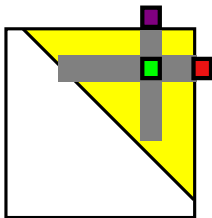
Measure all pairwise distances under distance metric



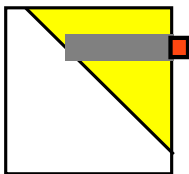
Hierarchical Clustering



Hierarchical Clustering



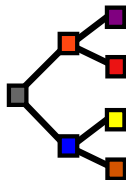
Hierarchical Clustering



Hierarchical Clustering



Hierarchical Clustering

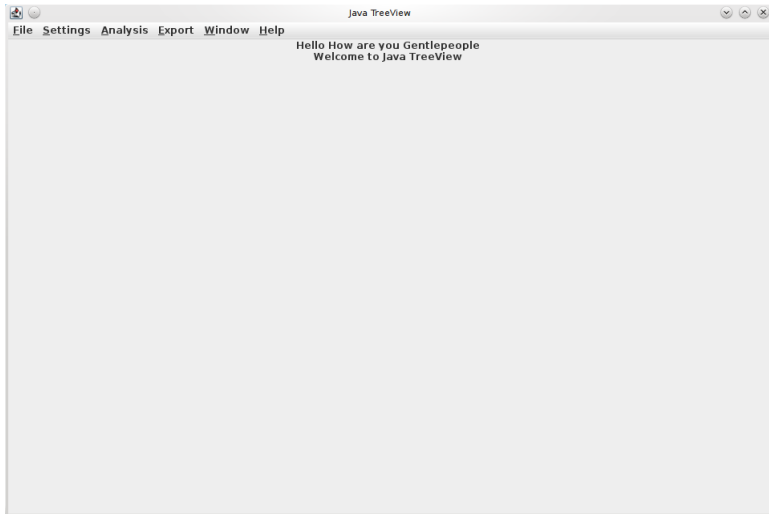


It's hard work at times, but you have to be realistic. If you have a large database with many variables and your goal is to get a good understanding of the interrelationships, then, unless you get lucky, this complex structure is bound to require some hard work to understand.

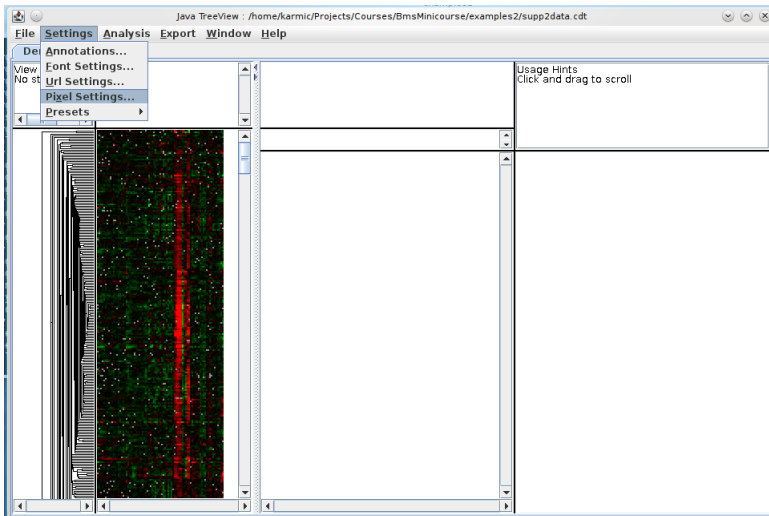
Bill Cleveland and Rick Becker

<http://stat.bell-labs.com/project/trellis/interview.html>

Using JavaTreeView



Adjust pixel settings for global view



Adjust pixel settings for global view

The screenshot shows the Java TreeView application window. The main view displays a heatmap with a dendrogram on the left. A 'Pixel Settings' dialog box is open in the foreground, allowing for adjustments to the heatmap's appearance. The dialog includes the following controls:

- Global:** Radio buttons for 'Fixed Scale' (with input fields for X: 481012658227 and Y: 663964329145) and 'Fill' (selected).
- Zoom:** Radio buttons for 'Fixed Scale' (with input fields for X: 12.0 and Y: 12.0) and 'Fill'.
- Contrast:** A slider with a 'Value' of 3.0.
- LogScale:** A checkbox for 'Log (base 2)' and a 'Center' input field set to 1.0.
- Colors:** Four color selection buttons: 'Positive' (red), 'Zero' (black), 'Negative' (green), and 'Missing' (grey). Below these are 'Load...', 'Save...', and 'Make Preset' buttons, and a dropdown menu currently showing 'RedGreen' and 'YellowBlue' options.
- A 'Close' button at the bottom of the dialog.

Select annotation columns

The screenshot shows the Java TreeView application interface. The title bar reads "Java TreeView : /home/karmic/Projects/Courses/BmsMinicourse/examples2/supp2.data.txt". The menu bar includes "File", "Settings", "Analysis", "Export", "Window", and "Help". The "Settings" menu is open, showing options like "Annotations...", "Font Settings...", "Url Settings...", "Pixel Settings...", and "Presets".

The main window is divided into three panes:

- Left Pane:** A dendrogram showing hierarchical clustering of samples. A red vertical bar highlights a specific cluster of samples.
- Middle Pane:** A heatmap visualization where rows represent genes and columns represent samples. The color scale ranges from black (low expression) to red (high expression).
- Right Pane:** A list of gene annotations. The top of this pane includes "Usage Hints" and "Click and drag to scroll". The list of genes and their associated biological processes is as follows:

| Gene ID | Gene Name | Biological Process |
|---------|--|-------------------------|
| GDH3 | GLUTAMATE BIOSYNTHESIS | NADP-GLUTAMAT |
| GDH1 | GLUTAMATE BIOSYNTHESIS | GLUTAMATE DEH |
| SEC18 | SECRETION | NSF; VESICLE FUSION |
| ABF2 | MITOCHONDRIAL GENOME MAI (PUTATIVE) HM | |
| RH03 | CYTOSKELETON | GTP-BINDING PROTEI |
| TFE1 | TRANSCRIPTION | TFIIH 75 KD SUBUNI |
| TAFL45 | TRANSCRIPTION | TFIID 145 KD SUBUN |
| INF52 | ENDOCYTOSIS (PUTATIVE) | INOSITOL POLY |
| POB3 | DNA REPLICATION (PUTATIV | BINDS DNA POL |
| PH08 | PHOSPHATE METABOLISM | VACUOLAR ALKA |
| GAT1 | NITROGEN CATABOLISM | TRANSCRIPTION F |
| DPP1 | PHOSPHOLIPID METABOLISM | DIACYLGLYCERO |
| MRF20 | PROTEIN SYNTHESIS | RIBOSOMAL PROTEI |
| DRS2 | TRANSPORT | CA(2+)-TRANSPORTING A |
| ECM13 | CELL WALL BIOGENESIS | UNKNOWN |
| BUB2 | CELL CYCLE, CHECKPOINT | UNKNOWN |
| CTK2 | CELL CYCLE | CYCLIN-LIKE |
| GCN5 | CHROMATIN STRUCTURE | HISTONE ACETYLT |
| MNN4 | PROTEIN GLYCOSYLATION | PHOSPHATIDYLI |
| TFCS | TRANSCRIPTION | TFIIIB 90 KD SUBUN |
| SNF2 | TRANSCRIPTION | COMPONENT OF SWI/S |
| SEC2 | SECRETION | GDP/GTP EXCHANGE FACT |
| UPE1 | SECRETION | ER MEMBRANE T-SNARE |
| NUP42 | NUCLEAR PROTEIN TARGETIN | NUCLEAR PORE |
| WHI4 | CELL SIZE | PUTATIVE RNA BINDING |
| USS1 | MRNA SPLICING | U5 SNRNP PROTEIN |
| REF2 | MRNA 3'-END PROCESSING | UNKNOWN |
| GLE2 | NUCLEAR PROTEIN TARGETIN | NUCLEAR PORE |
| BAT1 | BRANCHED CHAIN AMINO ACI | TRANSAMINASE |
| MOT2 | MATING | TRANSCRIPTIONAL REGULAT |
| KG02 | TCA CYCLE | 2-OXOGLUTARATE DEHYDR |
| COD4 | UBIQUITINOME BIOSYNTHESIS | UNKNOWN |
| CP1 | OXIDATIVE STRESS RESPONSI | CYTOCHROME-C |
| PDX1 | GLYCOLYSIS | PYRUVATE DEHYDROGENAS |
| ECM37 | CELL WALL BIOGENESIS | UNKNOWN |
| ECM27 | CELL WALL BIOGENESIS | UNKNOWN |

Select annotation columns

The screenshot shows the Java TreeView application interface. The main window displays a dendrogram on the left and a heatmap in the center. A dialog box titled 'Annotation Settings' is open, showing a list of columns to be included in the annotation table. The columns listed are: **GID**, **ORF**, **NAME**, and **GWEIGHT**. The dialog also has tabs for 'Array Tree' and 'Gene Tree', and a 'Close' button.

The annotation table on the right side of the window lists the following columns: Gene, ORF, NAME, GWEIGHT, and Usage Hints. The Usage Hints column contains various biological processes such as GLUTAMATE BIOSYNTHESIS, SECRETION, MITOCHONDRIAL GENOME MAI, CYTOSKELETON, TRANSCRIPTION, ENDOCYTOSIS, DNA REPLICATION, PHOSPHATE METABOLISM, NITROGEN CATABOLISM, PHOSPHOLIPID METABOLISM, PROTEIN SYNTHESIS, TRANSPORT, CELL WALL BIOGENESIS, CELL CYCLE, CYCLIN-LIKE, CHROMATIN STRUCTURE, HISTON, PROTEIN GLYCOSYLATION, TRANSCRIPTION, TRANSCRIPTION, SECRETION, NUCLEAR PROTEIN TARGETIN, CELL SIZE, PUTATIVE RNA, MRNA SPLICING, MRNA 3'-END PROCESSING, NUCLEAR PROTEIN TARGETIN, BRANCHED CHAIN AMINO ACI TRAI, MATING, TCA CYCLE, 2-OXOGLUTARU, UBIQUINONE BIOSYNTHESIS, OXIDATIVE STRESS RESPONSI, PYRUVATE DEI, CELL WALL BIOGENESIS, and CELL WALL BIOGENESIS.

Select URL for gene annotations

The screenshot shows the Java TreeView application window titled "Java TreeView : /home/karmic/Projects/Courses/BmsMinicourse/examples2/supp2data.cdt". The "File" menu is open, and the "Gene Url Presets..." option is selected. A secondary menu is displayed, listing various preset options:

- Gene Url Presets... (Ctrl-P)
- Array Url Presets...
- Dendrogram Color Presets...
- KnnDendrogram Color Presets...
- Karyoscope Color...
- Karyoscope Coordinates...
- Scatterplot Color...

The main window displays a dendrogram on the left and a heatmap on the right. The heatmap has a color scale from 0 (black) to 100 (red). The gene names are listed on the left side of the heatmap, and the corresponding gene annotations are listed on the right side of the heatmap.

| Gene | Annotation |
|---------|-------------------------------------|
| YAL062W | GDH3 GLUTAMATE BIOSYNTHESIS NADI |
| YOR375C | GDH1 GLUTAMATE BIOSYNTHESIS GLU |
| YBR080C | SEC18 SECRETION NSF; VESICLI |
| YMR072W | ABF2 MITOCHONDRIAL GENOME MAI (PU |
| YTL118W | RHO3 CYTOSKELETON GTP-BIND; |
| YOR311W | TFB1 TRANSCRIPTION TFIIH 75 |
| YGR274C | TAF145 TRANSCRIPTION TFIIID 14; |
| YNL106C | INP52 ENDOCYTOSIS (PUTATIVE) INO |
| YML069W | POB3 DNA REPLICATION (PUTATIV BINI |
| YDR481C | PHO8 PHOSPHATE METABOLISM VACI |
| YFL021W | GAT1 NITROGEN CATABOLISM TRANSI |
| YDR284C | DPP1 PHOSPHOLIPID METABOLISM DIAI |
| YDR405W | MFP20 PROTEIN SYNTHESIS RIBOSOM |
| YAL028C | DPS2 TRANSPORT CA (2+) TRAN |
| YBL043W | ECM13 CELL WALL BIOGENESIS UNKI |
| YMR055C | BUB2 CELL CYCLE CHECKPOINT UNKI |
| YJL006C | CTK2 CELL CYCLE CYCLIN-LIKE |
| YGR252W | GCN5 CHROMATIN STRUCTURE HISTOF |
| YKL201C | MNN4 PROTEIN GLYCOSYLATION PHO |
| YNL039W | TF15 TRANSCRIPTION TFIIIB 94 |
| YOR290C | SNF2 TRANSCRIPTION COMPONENT |
| YNL272C | SEC2 SECRETION GDP/GTP EXCI |
| YOR075W | LEF1 SECRETION ER MEMBRANE |
| YDR192C | NUP42 NUCLEAR PROTEIN TARGETIN NUCL |
| YDL224C | WHI4 CELL SIZE PUTATIVE RN |
| YER112W | USS1 mRNA SPLICING U6 SNRNP |
| YOR109W | REF2 MRNA 3' END PROCESSING UNKI |
| YER107C | GLE2 NUCLEAR PROTEIN TARGETIN NUCL |
| YHR208W | BAT1 BRANCHED CHAIN AMINO ACI TRAI |
| YER069W | MOT2 MATING TRANSCRIPTION; |
| YDR149C | KG02 TCA CYCLE 2-OXOGLUTAR; |
| YDR204W | COO4 UBIQUINONE BIOSYNTHESIS UNKI |
| YKR069C | CP1 OXIDATIVE STRESS RESPON CYTI |
| YGR193C | POX1 GLYCOLYSIS PYRUVATE DEI |
| YTL146C | ECM37 CELL WALL BIOGENESIS UNKI |
| YJL109W | ECM27 CELL WALL BIOGENESIS UNKI |

http://www.ensembl.org/Mus_musculus/Gene/Summary?g=HEADER

Select URL for gene annotations

The screenshot shows the Java TreeView application interface. At the top, the title bar reads "java TreeView - /home/karmic/Projects/Courses/BmsMinicourse/examples2/supp2data.cdt". The menu bar includes "File", "Settings", "Analysis", "Export", "Window", and "Help".

The main window is divided into several panes:

- Dendrogram:** Located at the top left, it shows a hierarchical tree structure of nodes.
- View Status:** Below the dendrogram, it says "Select Node to view annotator".
- Heatmaps:** Two heatmaps are visible, showing gene expression data with red and green colors.
- Usage Hints:** A text box on the right says "Click to select node - use arrow keys to navigate tree".
- Presets:** A dialog box titled "Modify Url Presets" is open in the foreground. It has tabs for "Gene" and "Array". The table below lists the presets:

| Enabled | Header | Name | Template | Default? |
|--------------------------|--------|----------------|---|----------------------------------|
| <input type="checkbox"/> | * | SGD | http://genome-www4.stanford.edu/cgi-bin/SGD/locus.pl?locus=HEADER | <input checked="" type="radio"/> |
| <input type="checkbox"/> | * | YPD | http://www.proteome.com/databases/YPD/reports/HEADER.html | <input type="radio"/> |
| <input type="checkbox"/> | * | WormBase | http://www.wormbase.org/cgi-bin/locate.pl?locus=HEADER&start=0&start=0&ie=utf-8&oe=utf-8 | <input type="radio"/> |
| <input type="checkbox"/> | * | Source CloneID | http://genome-www4.stanford.edu/cgi-bin/SMD/source/sourceResult?option=CloneID | <input type="radio"/> |
| <input type="checkbox"/> | * | FlyBase | http://flybase.bio.indiana.edu/bin/fbgenq.html?HEADER | <input type="radio"/> |
| <input type="checkbox"/> | * | MouseGD | http://www.jax.org/avaw/Servlet/SearchTool?query=HEADER&selectedQuery=Genes+and+Markers | <input type="radio"/> |
| <input type="checkbox"/> | * | GenomeNetEcoli | http://www.genome.ad.jp/dbget-bin/www_bget?eco:HEADER | <input type="radio"/> |
| <input type="checkbox"/> | | None | | <input type="radio"/> |

At the bottom of the main window, there are more heatmaps and a list of gene names with their associated biological processes, such as "YER107C", "YHR208B", "YER060W", "YDR148C", "YDR204W", "YKR866C", "YCR190C", "YTL146C", "YJR106W", "MET2", "GLI2", "BAT1", "MOT2", "KGD2", "COQ4", "COF1", "PDX1", "ECM37", "ECM27", "PWR1", "PWR2", "PWR3", "ERU1", "NUCLEAR PROTEIN TARGETIN", "BRANCHED CHAIN AMINO ACI", "MATING", "TRANSCRIPTION", "TCA CYCLE", "2-OXOGLUTAR", "UBIQUINONE BIOSYNTHESIS", "OXIDATIVE STRESS RESPON", "GLYCOLYSIS", "PYRUVATE DE", "CELL WALL BIOGENESIS", "CELL WALL BIOGENESIS".

Activate and detach annotation window

The screenshot shows the Java TreeView application window titled "java TreeView : /home/karmac/Projects/Courses/BmsMinicourse/examples2/supp2data.cdt". The interface includes a menu bar (File, Settings, Analysis, Export, Window, Help) and a toolbar. The "Analysis" menu is open, showing options like "Find Genes...", "Find Arrays...", "Stats...", "Flip Array Tree Node", "Align to Tree...", "Compare to...", "Remove comparison", "Summary Window...", "Dendrogram", "Alignment", "KnnDendrogram", "Karyoscope", "Scatterplot", "ArrayTreeAnno", "GeneTreeAnno", "Remove Current", and "Detach Current".

The main window displays a dendrogram on the left and a heatmap on the right. The heatmap has a grid of colored cells (green, red, black) representing data points. A red rectangular highlight is visible on the dendrogram. The heatmap is annotated with gene names on the left and biological processes on the right. The gene names listed are: YAL063W, YOR375C, YBR080C, YMR072W, YIL119W, YOR311W, YOR274C, YNL106C, YML069W, YDR481C, YFL021W, YDR284C, YDR495W, YAL029C, YBL043W, YMR055C, YJL006C, YGR252W, YKL201C, YML039W, YOR290C, YML272C, YOR075W, YDR192C, YDL224C, YER112W, YOR185W, YER107C, YHR206W, YER066W, YDR148C, YDR204W, YKR066C, YGR183C, YJL146C, YJRI06W, YAL063W, YOR375C, YBR080C, YMR072W, YIL119W, YOR311W, YOR274C, YNL106C, YML069W, YDR481C, YFL021W, YDR284C, YDR495W, YAL029C, YBL043W, YMR055C, YJL006C, YGR252W, YKL201C, YML039W, YOR290C, YML272C, YOR075W, YDR192C, YDL224C, YER112W, YOR185W, YER107C, YHR206W, YER066W, YDR148C, YDR204W, YKR066C, YGR183C, YJL146C, YJRI06W, YAL063W, YOR375C, YBR080C, YMR072W, YIL119W, YOR311W, YOR274C, YNL106C, YML069W, YDR481C, YFL021W, YDR284C, YDR495W, YAL029C, YBL043W, YMR055C, YJL006C, YGR252W, YKL201C, YML039W, YOR290C, YML272C, YOR075W, YDR192C, YDL224C, YER112W, YOR185W, YER107C, YHR206W, YER066W, YDR148C, YDR204W, YKR066C, YGR183C, YJL146C, YJRI06W, YAL063W, YOR375C, YBR080C, YMR072W, YIL119W, YOR311W, YOR274C, YNL106C, YML069W, YDR481C, YFL021W, YDR284C, YDR495W, YAL029C, YBL043W, YMR055C, YJL006C, YGR252W, YKL201C, YML039W, YOR290C, YML272C, YOR075W, YDR192C, YDL224C, YER112W, YOR185W, YER107C, YHR206W, YER066W, YDR148C, YDR204W, YKR066C, YGR183C, YJL146C, YJRI06W.

The biological processes listed on the right are: GLUTAMATE BIOSYNTHESIS, GLUTAMATE BIOSYNTHESIS, SECRETION, MITOCHONDRIAL GENOME MAI (PU, CYTOSKELETON, TRANSCRIPTION, ENDOCYTOSIS (PUTATIVE), DNA REPLICATION (PUTATIV BINI, PHOSPHATE METABOLISM, NITROGEN CATABOLISM, PHOSPHOLIPID METABOLISM, PROTEIN SYNTHESIS, TRANSPORT, CELL WALL BIOGENESIS, CELL CYCLE, CHECKPOINT, CELL CYCLE, CYCLIN-LIKE, CHROMATIN STRUCTURE, HISTO, PROTEIN GLYCOSYLATION, TRANSCRIPTION, SECRETION, GDP/GTP EXO, SECRETION, ER MEMBRANE, NUCLEAR PROTEIN TARGETIN NUCL, CELL SIZE, PUTATIVE RN, MRNA SPLICING, MRNA 3'-END PROCESSING, NUCLEAR PROTEIN TARGETIN NUCL, BRANCHED CHAIN AMINO ACI TRAI, MATING, TCA CYCLE, 2-OXOGLUTAR, UBIQUINONE BIOSYNTHESIS, OXIDATIVE STRESS RESPONSA CYT, PYRUVATE DE, CELL WALL BIOGENESIS, CELL WALL BIOGENESIS.

Usage Hints: Click and drag to scroll

Activate and detach annotation window

The screenshot shows the Java TreeView application window. The title bar reads "Java TreeView : /home/karmic/Projects/Courses/BmsMinicourse/examples2/supp2data.cdt". The menu bar includes "File", "Settings", "Analysis", "Export", "Window", and "Help". The "Analysis" menu is open, showing options: "Find Genes..." (Ctrl-G), "Find Arrays..." (Ctrl-A), "Stats..." (Ctrl-S), "Dendrogram", "Alignment", "KnnDendrogram", "Karyoscope", "Scatterplot", "ArrayTreeAnno", "GeneTreeAnno", "Remove Current", and "Detach Current".

Below the menu, there are input fields for "Name" and "Annotation". The main area contains a table with the following columns: NODEID, LEFT, RIGHT, CORRELAT..., NAME, and ANNOTATI... The table lists various nodes and their associated gene IDs and correlation values.

| NODEID | LEFT | RIGHT | CORRELAT... | NAME | ANNOTATI... |
|------------|------------|------------|-------------|------|-------------|
| NODE243... | GENE182... | NODE239... | 0.347965 | | |
| NODE244... | NODE242... | NODE243... | 0.347965 | | |
| NODE244... | GENE550X | NODE239... | 0.344607 | | |
| NODE244... | NODE243... | NODE244... | 0.342251 | | |
| NODE244... | NODE244... | GENE4X | 0.334454 | | |
| NODE244... | NODE240... | NODE239... | 0.333461 | | |
| NODE244... | NODE244... | NODE243... | 0.331585 | | |
| NODE244... | NODE244... | NODE238... | 0.328813 | | |
| NODE244... | NODE244... | GENE229... | 0.305824 | | |
| NODE244... | GENE495X | GENE217... | 0.304111 | | |
| NODE244... | GENE219... | GENE218... | 0.303188 | | |
| NODE245... | NODE244... | GENE215X | 0.301587 | | |
| NODE245... | NODE244... | NODE242... | 0.298323 | | |
| NODE245... | NODE240... | NODE244... | 0.289436 | | |
| NODE245... | NODE242... | GENE219... | 0.287138 | | |
| NODE245... | NODE245... | NODE243... | 0.284232 | | |
| NODE245... | NODE245... | GENE527X | 0.277872 | | |
| NODE245... | NODE245... | NODE234... | 0.27761 | | |
| NODE245... | NODE245... | NODE244... | 0.271103 | | |
| NODE245... | NODE233... | NODE245... | 0.260487 | | |
| NODE245... | NODE243... | NODE245... | 0.220385 | | |
| NODE246... | NODE244... | NODE245... | 0.197665 | | |
| NODE246... | NODE245... | NODE243... | 0.180953 | | |
| NODE246... | NODE246... | GENE182... | 0.161919 | | |
| NODE246... | NODE246... | NODE119... | 0.126461 | | |
| NODE246... | NODE246... | NODE245... | 0.098323 | | |
| NODE246... | NODE245... | NODE246... | -0.087409 | | |
| NODE246... | NODE246... | NODE246... | -0.354391 | | |

Activate and detach annotation window

The screenshot displays the Java TreeView application interface. The main window is titled "java TreeView : /home/karmic/Projects/Courses/BmsMinicourse/examples2/supp2data.cdt". It features a menu bar with "File", "Settings", "Analysis", "Export", "Window", and "Help".

The interface is divided into several sections:

- Dendrogram:** Located at the top left, it shows a hierarchical tree structure of samples. The "View Status" panel indicates "Row: 115 (Y)", "Column: 49 (S)", and "Value: 1.34".
- Heatmap:** The central area displays a color-coded heatmap where red indicates high expression and green indicates low expression. A vertical red bar highlights a specific cluster of genes.
- Gene List:** On the right side, a list of genes is shown with their corresponding expression values in the heatmap. The genes include YFR028C, YML069W, YIL139C, YNL318C, YFR023W, YHR015W, YDR263C, YLR045C, YOR033C, YIL159W, YKL042W, YNL225C, YCR092C, YLR210W, YLR314C, YBR045C, YDL159W, YDR118W, YDR253C, YMR190W, and YDR113C.
- Usage Hints:** A small box at the top right says "Usage Hints" and "Mouse over to get info".
- GeneTreeAnno:** A separate window at the bottom shows a table of gene annotations for "Sporulation".

The "GeneTreeAnno" window contains the following table:

| Name | Annotation | | | | |
|-------------|----------------------------------|------------|-------------|-------------|-------------|
| Sporulation | Genes upregulated in sporulation | | | | |
| NODEID | LEFT | RIGHT | CORRELAT... | NAME | ANNOTATI... |
| NODE184... | NODE184... | NODE152... | 0.627369 | Sporulation | Genes up... |
| NODE184... | NODE184... | GENE56X | 0.627369 | | |
| NODE184... | NODE184... | NODE178... | 0.627369 | | |
| NODE184... | NODE150... | GENE177... | 0.627287 | | |

Buttons for "Dock" and "Close" are visible at the bottom of the GeneTreeAnno window.

- 1 For a small expression profiling matrix (1000 genes):
 - Cluster the genes
 - Calculate the correlation matrix
 - Write a CDT file of the clustered gene matrix with the correlation matrix appended
 - Visualize the CDT+GTR files in JavaTreeView – how well did the clustering work?
- 2 Repeat the previous exercise, exploring difference clustering methods and/or distance methods
- 3 Read the supplemental RnaSeq methods for PLoS Pathogens 12:e1005910 (Text S2, exported from RStudio). To what extent is this a reproducible method? Is there additional data that would make it more reproducible?