

Expression Profiling

Mark Voorhies

4/3/2012

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

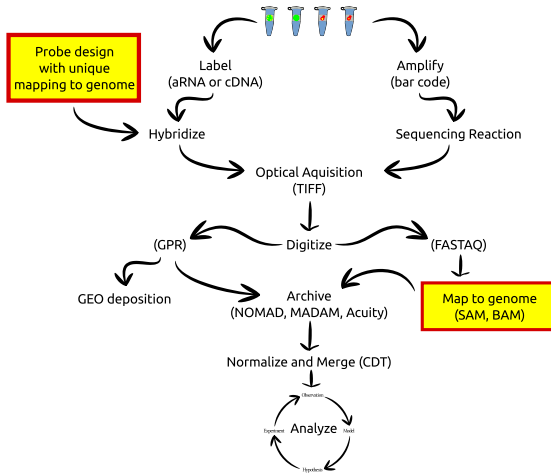
<http://stat.bell-labs.com/project/trellis/interview.html>

Why profile *transcription*?

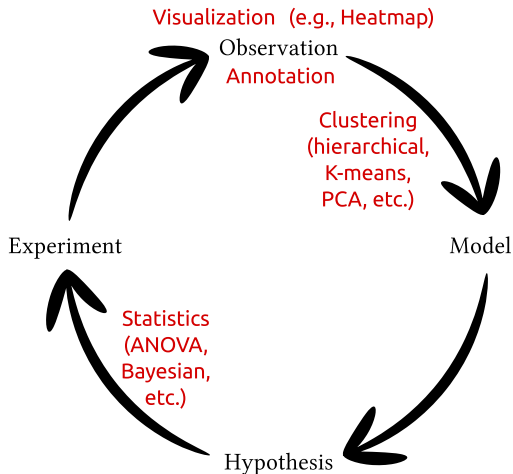
Why profile *transcription*?

- Major mode of regulation
- Due to feedback, “shadows” other modes of regulation
- Thanks to Watson-Crick base pairing, we can assay arbitrary nucleic acids in a uniform way

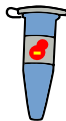
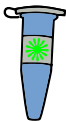
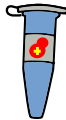
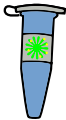
Expression Profiling Workflow



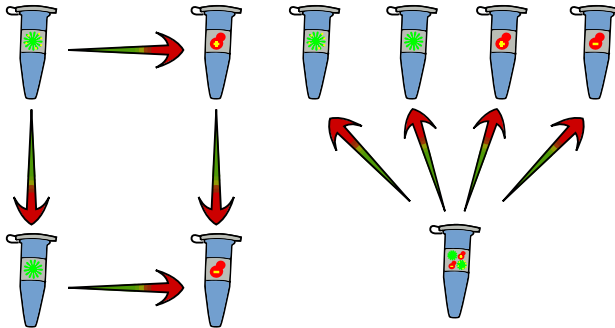
Expression Profiling Analysis



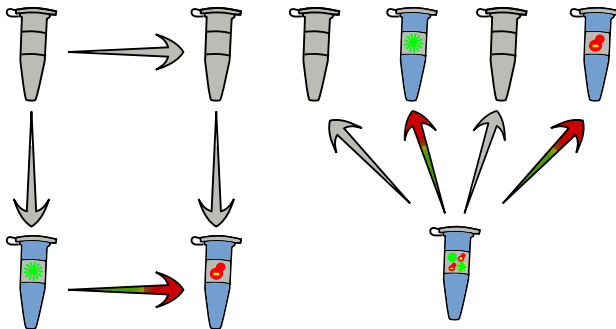
Sample Preparation



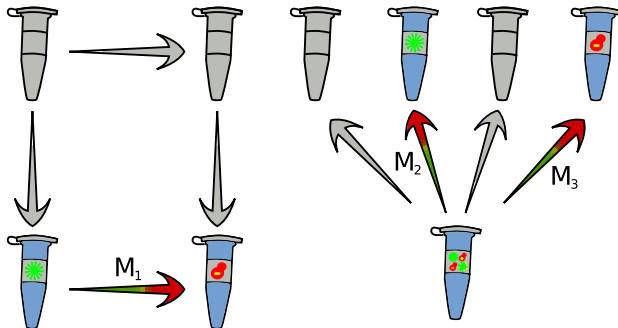
Transforming Ratios



Transforming Ratios

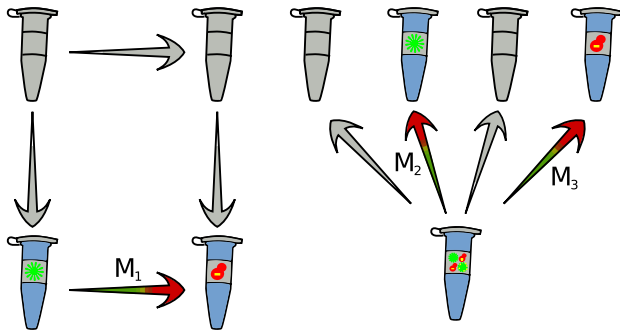


Transforming Ratios



$$M_1 = M_3 / M_2$$

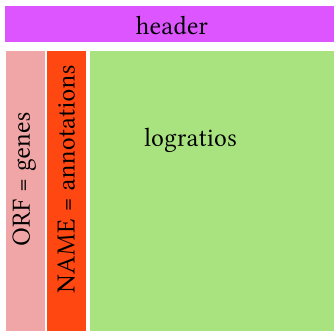
Transforming Ratios



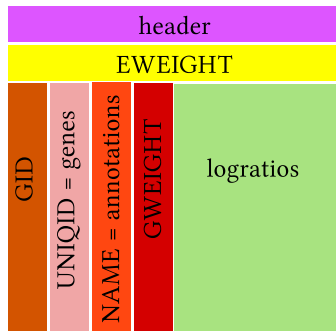
$$\log_2 M_1 = \log_2 M_3 - \log_2 M_2$$

The CDT file format

Minimal CLUSTER input

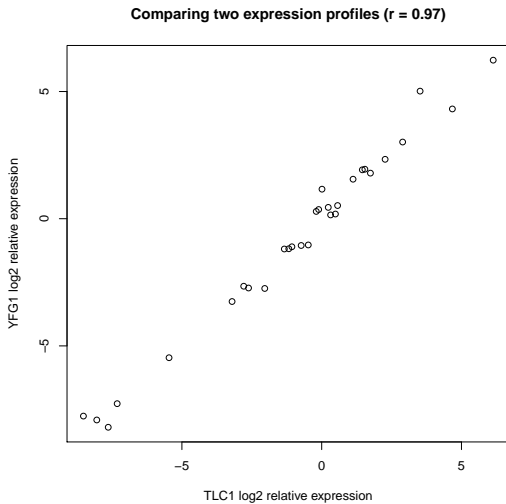


Cluster3 CDT output

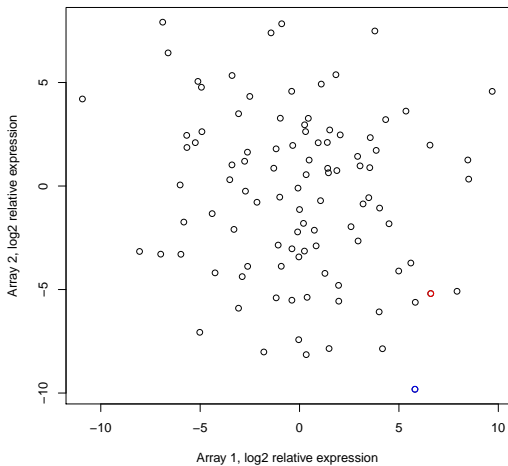


- Tab delimited (`\t`)
- UNIX newlines (`\n`)
- Missing values \rightarrow empty cells

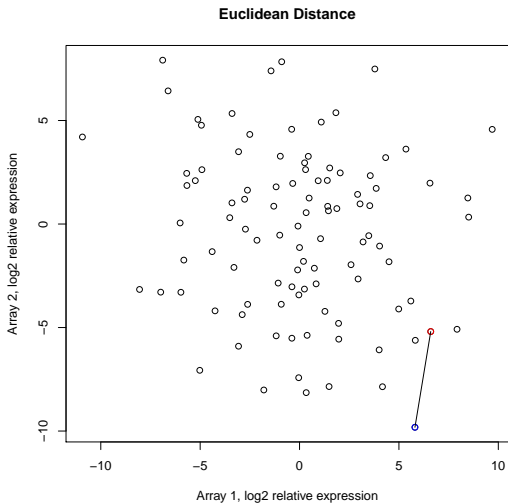
Comparing all measurements for two genes



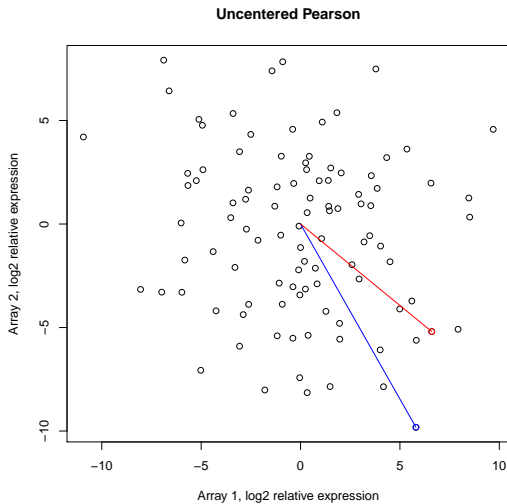
Comparing all genes for two measurements



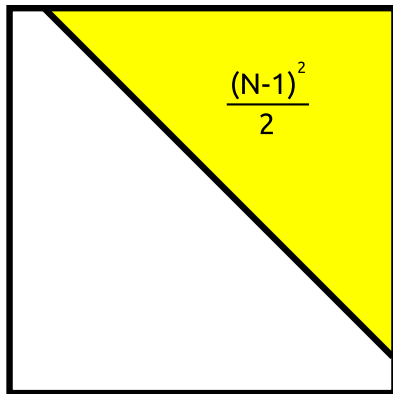
Comparing all genes for two measurements



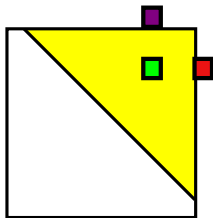
Comparing all genes for two measurements



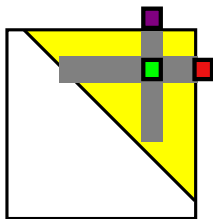
Measure all pairwise distances under distance metric



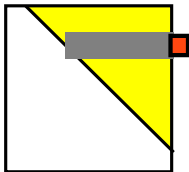
Hierarchical Clustering



Hierarchical Clustering



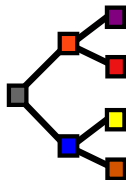
Hierarchical Clustering



Hierarchical Clustering



Hierarchical Clustering



Using the Cluster3 GUI

Gene Cluster 3.0

File Help

File loaded

Job name

Data set has Rows Columns

Filter Genes

- % Present \geq 80
- SD (Gene Vector) 2,0
- At least 1 observations with $\text{abs}(\text{Val}) \geq$ 2,0
- $\text{MaxVal} - \text{MinVal} \geq$ 2,0

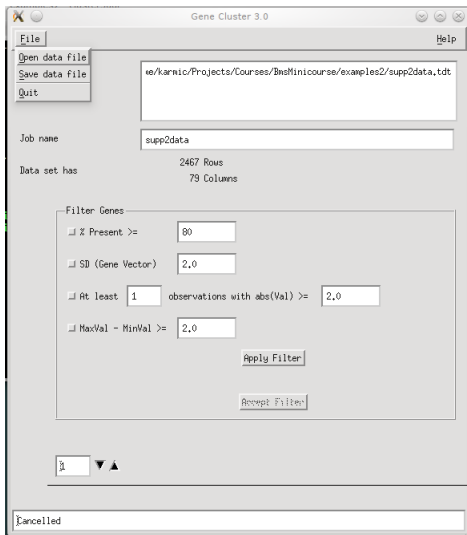
Apply Filter

Accept Filter

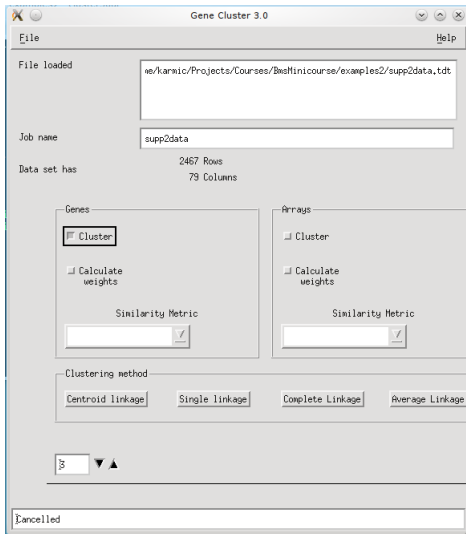
1

1

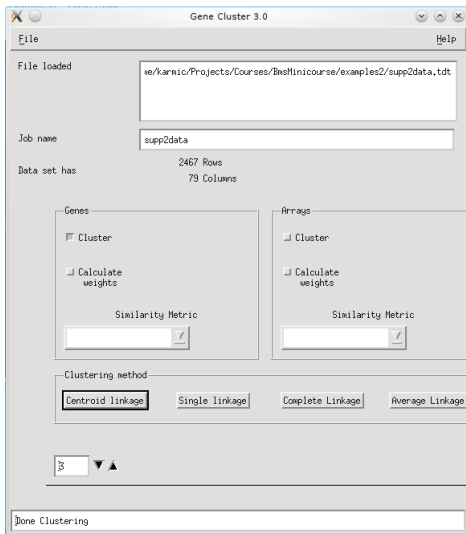
Load your data



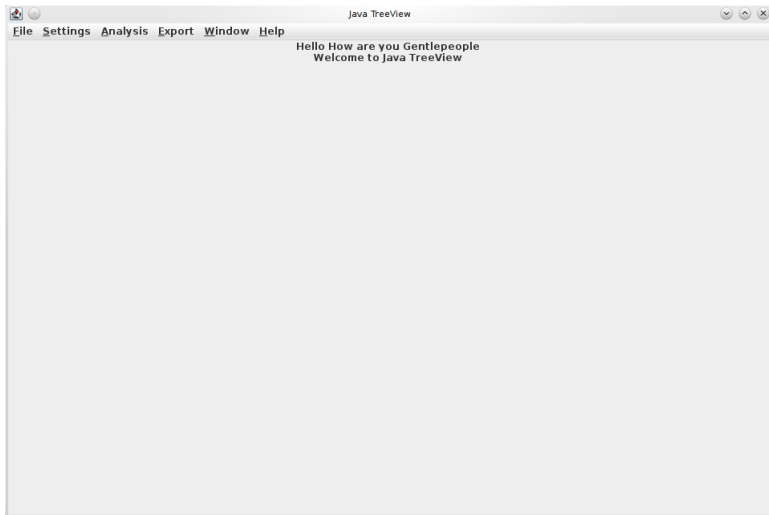
Choose distance function



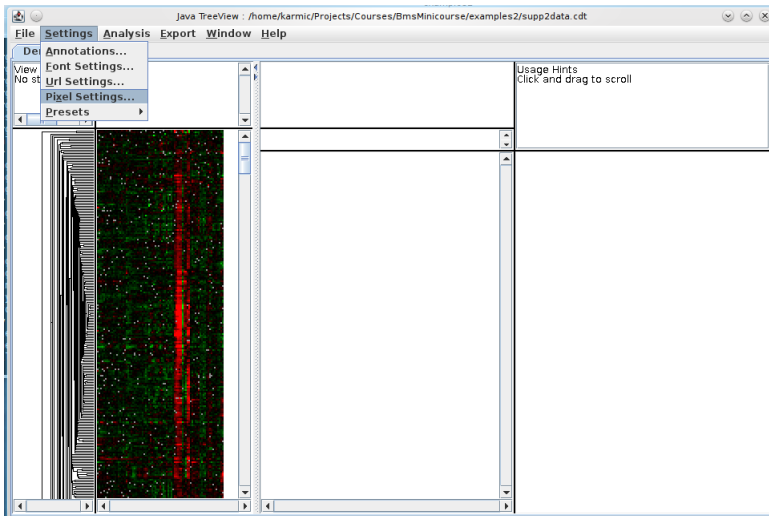
Choose linking method



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 dendrogram on the left and a heatmap on the right. The heatmap uses a color scale where red indicates positive values and green indicates negative values. A 'Pixel Settings' dialog box is open in the foreground, allowing for adjustments to the heatmap's appearance. The dialog includes sections for Global, Zoom, Contrast, and LogScale settings, as well as color selection options.

Pixel Settings

Global:

- X: Fixed Scale 481012658227 Fill
- Y: Fixed Scale 663964329145 Fill

Zoom:

- X: Fixed Scale 12.0 Fill
- Y: Fixed Scale 12.0 Fill

Contrast: Value: 3.0

LogScale: Log (base 2) Center: 1.0

Colors:

- Positive
- Zero
- Negative
- Missing

Buttons: Load... Save... Make Preset

Color Presets: RedGreen YellowBlue

Close

Select annotation columns

The screenshot shows the Java TreeView application interface. The title bar reads "Java TreeView : /home/karmic/Projects/Courses/BmsMinicourse/examples2/supp2data.dat". 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 of gene expression data. The color scale ranges from green (low expression) to red (high expression). A yellow box highlights a specific region of the heatmap.
- Right Pane:** A list of gene annotations. The list includes gene names and their associated biological processes. The visible portion of the list is as follows:

Gene	Annotation
GDH3	GLUTAMATE BIOSYNTHESIS NADP-GLUTAMAT
GDH1	GLUTAMATE BIOSYNTHESIS GLUTAMATE DEH
SEC18	SECRETION NSF; VESICLE FUSION
ABF2	MITOCHONDRIAL GENOME MAI (PUTATIVE) HM
RHO3	CYTOSKELETON GTP-BINDING PROTEIN
TFE1	TRANSCRIPTION TFIIH 75 KD SUBUNI
TAFL145	TRANSCRIPTION TFIIID 145 KD SUBUNI
INP52	ENDOCYTOSIS (PUTATIVE) INOSITOL POLY
POB3	DNA REPLICATION (PUTATIV BINDS DNA POL
PHO8	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 SUBUNI
SNF2	TRANSCRIPTION COMPONENT OF SWI/S
SEC2	SECRETION GTP/GTP EXCHANGE FACT
UPE1	SECRETION ER MEMBRANE T-SNARE
NUP42	NUCLEAR PROTEIN TARGETIN NUCLEAR PORE
WHI4	CELL SIZE PUTATIVE RNA BINDING
USS1	MRNA SPLICING U6 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
POX1	GLYCOLYSIS PYRUVATE DEHYDROGENAS
ECM37	CELL WALL BIOGENESIS UNKNOWN
ECM27	CELL WALL BIOGENESIS UNKNOWN

Select annotation columns

The screenshot shows the Java TreeView application window. The title bar reads "java TreeView : /home/karmic/Projects/Courses/BmsMinicourse/examples/2/supp2data.dwt". The menu bar includes "File", "Settings", "Analysis", "Export", "Window", and "Help". The "Dendrogram" tab is active, showing a hierarchical tree on the left and a heatmap on the right. The heatmap has a red vertical bar on the left side. An "Annotation Settings" dialog box is open in the center, with the "Gene" tab selected. The "Headers to include" list contains: "GID", "ORF", "NAME", and "GWEIGHT". The "Close" button is visible at the bottom of the dialog. In the background, a list of gene IDs and their corresponding annotations is visible, such as "YAL062W GOH3 GLUTAMATE BIOSYNTHESIS NADH".

View Status
Row: 7 (YGR2)
Column: 23 (Elu)
Value: -0.06

Usage Hints
Mouse over to get info

Annotation Settings

Array Tree Gene Tree

Gene Array

Headers to include

- GID
- ORF
- NAME
- GWEIGHT

Close

YAL062W GOH3 GLUTAMATE BIOSYNTHESIS NADH
YOR375C GOH1 GLUTAMATE BIOSYNTHESIS GLU
YBR080C SEC18 SECRETION NSF; VESICLE
YMR072W ABF2 MITOCHONDRIAL GENOME MAI (PU
YIL119W RH03 CYTOSKELETON GTP-BIND;
YDR311W TFB1 TRANSCRIPTION TFIH 75
YGR274C TAF145 TRANSCRIPTION TFIID 145
YNL106C INP52 ENDOCYTOSIS (PUTATIVE) INO
YML069W POB3 DNA REPLICATION (PUTATIV BINE
YDR481C PH08 PHOSPHATE METABOLISM VACI
YFL021W GAT1 NITROGEN CATABOLISM TRANS
YDR284C DPP1 PHOSPHOLIPID METABOLISM DIA
YDR405W MFP20 PROTEIN SYNTHESIS RIBOSOM
YAL028C DRS2 TRANSPORT CA(2+) TRAN
YBL043W ECM13 CELL WALL BIOGENESIS UNK
YMR055C BUB2 CELL CYCLE CHECKPOINT UNK
YJL006C CTK2 CELL CYCLE CYCLIN-LIKE
YGR252W GCN5 CHROMATIN STRUCTURE HISTO
YKL201C MNH4 PROTEIN GLYCOSYLATION PHO
YNL035W TFC5 TRANSCRIPTION TFIIB 94
YOF280C SMF2 TRANSCRIPTION COMPONENT
YNL272C SEC2 SECRETION GDP/GTP EXC
YOR075W LEF1 SECRETION ER MEMBRANE
YDR192C NUP42 NUCLEAR PROTEIN TARGETIN NUC
YDL224C WHI4 CELL SIZE PUTATIVE RNA
YER112W USS1 MRNA SPLICING U6 SNRNP
YDR195W REF2 MRNA 3'-END PROCESSING UNK
YER107C GLE2 NUCLEAR PROTEIN TARGETIN NUC
YHF208W BAT1 BRANCHED CHAIN AMINO ACI TRAI
YER068W MOT2 MATING TRANSCRIPTION;
YDR149C KGD2 TCA CYCLE 2-OXOGLUTAR
YDR204W COO4 UBIQUINONE BIOSYNTHESIS UNK
YKR068C OCP1 OXIDATIVE STRESS RESPON CTT
YGR193C FOX1 GLYCOLYSIS PYRUVATE DEI
YIL146C ECM37 CELL WALL BIOGENESIS UNK
YJL106W ECM27 CELL WALL BIOGENESIS UNK

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 application is displaying a gene expression heatmap with a dendrogram on the left. A menu is open under the "Settings" menu, showing options for "Gene Url Presets...", "Array Url Presets...", "Dendrogram Color Presets...", "KnnDendrogram Color Presets...", "Karyoscope Color...", "Karyoscope Coordinates...", and "Scatterplot Color...". The heatmap shows a grid of colored cells (red, green, black) representing gene expression levels. The gene names are listed on the left side of the heatmap, and the gene annotations are listed on the right side. The annotations include terms like "GLUTAMATE BIOSYNTHESIS", "SECRETION", "MITOCHONDRIAL GENOME MAI", "CYTOSKELETON", "TRANSCRIPTION", "ENDOCYTOSIS (PUTATIVE)", "DNA REPLICATION (PUTATIV BINI", "PHOSPHATE METABOLISM", "NITROGEN CATABOLISM", "PROTEIN SYNTHESIS", "TRANSPORT", "CELL WALL BIOGENESIS", "CELL CYCLE, CHECKPOINT", "CYCLIN-LIKE", "CHROMATIN STRUCTURE", "PROTEIN GLYCOSYLATION", "TRANSCRIPTION COMPONENT", "SECRETION", "NUCLEAR PROTEIN TARGETIN", "CELL SIZE", "MRNA SPLICING", "MRNA 3' END PROCESSING", "NUCLEAR PROTEIN TARGETIN", "BRANCHED CHAIN AMINO ACI TRAI", "MATING", "TCA CYCLE", "UBIQUINONE BIOSYNTHESIS", "OXIDATIVE STRESS RESPONSI CYTI", "GLYCOLYSIS", "PYRUVATE DEI", "CELL WALL BIOGENESIS", and "CELL WALL BIOGENESIS".

File Settings Analysis Export Window Help

Annotations...
Font Settings...
Url Settings...
Pixel Settings...
Presets
Gene Url Presets...
Array Url Presets...
Dendrogram Color Presets...
KnnDendrogram Color Presets...
Karyoscope Color...
Karyoscope Coordinates...
Scatterplot Color...

Usage Hints
Click and drag to scroll

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	TFIID 14;
YNL106C	INP52	ENDOCYTOSIS (PUTATIVE)	INO
YML069W	POB3	DNA REPLICATION (PUTATIV BINI	
YDR481C	PHO8	PHOSPHATE METABOLISM	VACI
YFL021W	GAT1	NITROGEN CATABOLISM	TRANS
YDR284C	DPP1	PHOSPHOLIPID METABOLISM	DIAP
YOR405W	MFP20	PROTEIN SYNTHESIS	RIBOSOM
YAL028C	DPS2	TRANSPORT	CA(2+)
YBL043W	ECM13	CELL WALL BIOGENESIS	UNKI
YMR055C	BUB2	CELL CYCLE, CHECKPOINT	UNKI
YJL006C	CTK2	CELL CYCLE	CYCLIN-LIKE
YGR252W	GCN5	CHROMATIN STRUCTURE	HISTO
YKL201C	MNN4	PROTEIN GLYCOSYLATION	PHO
YNL039W	TF15	TRANSCRIPTION	TFIIIB 9;
YOR290C	SNF2	TRANSCRIPTION	COMPONENT
YNL272C	SEC2	SECRETION	GDP/GTP EXCI
YOR075W	UPE1	SECRETION	ER MEMBRANE
YDR192C	NUP42	NUCLEAR PROTEIN TARGETIN	NUCI
YDL224C	WHI4	CELL SIZE	PUTATIVE RN
YER112W	USS1	MRNA SPLICING	UG SNRNP
YOR109W	REF2	MRNA 3' END PROCESSING	UNKI
YER107C	GLE2	NUCLEAR PROTEIN TARGETIN	NUCI
YHR208W	BAT1	BRANCHED CHAIN AMINO ACI TRAI	
YER069W	MOT2	MATING	TRANSCRIPTION;
YDR149C	KG02	TCA CYCLE	2-OXOGLUTAR;
YDR204W	COO4	UBIQUINONE BIOSYNTHESIS	UNKI
YKR069C	COF1	OXIDATIVE STRESS RESPONSI CYTI	
YGR193C	POX1	GLYCOLYSIS	PYRUVATE DEI
YTL146C	ECM37	CELL WALL BIOGENESIS	UNKI
YJL109W	ECM27	CELL WALL BIOGENESIS	UNKI

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 sections:

- Dendrogram:** Located at the top left, it shows a hierarchical tree structure of nodes.
- View Status:** Below the dendrogram, it says "View Status" and "Select Node to view annotator".
- Heatmap:** A large central area displaying a heatmap with red and green cells, representing gene expression data.
- Usage Hints:** A text box on the right side of the heatmap area, containing the text: "Usage Hints", "Click to select node", and "- use arrow keys to navigate tree".
- Presets:** A dialog box titled "Presets" is open in the foreground. It has tabs for "Gene" and "Array". The "Gene" tab is active, showing a table titled "Modify Url Presets".

The "Modify Url Presets" table has the following columns: "Enabled", "Header", "Name", "Template", and "Default?".

Enabled	Header	Name	Template	Default?
<input type="checkbox"/>	*	SGD	http://genome-www4.stanford.edu/cqi-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/lookup/feature?feature=HEADER&source=genome&start=0&end=0&ie=utf-8&oe=utf-8	<input type="radio"/>
<input type="checkbox"/>	*	Source CloneID	http://genome-www4.stanford.edu/cqi-bin/SMD/source/sourceResult?option=CloneID	<input type="radio"/>
<input type="checkbox"/>	*	FlyBase	http://flybase.bio.indiana.edu/bin/fbqgenq.html?HEADER	<input type="radio"/>
<input type="checkbox"/>	*	MouseGD	cs.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_bqet?eco:HEADER	<input type="radio"/>
<input type="checkbox"/>		None		<input type="radio"/>

At the bottom of the application, there are several panels showing gene names and their corresponding expression patterns in the heatmap. Some visible gene names include YER107C, YHR206W, YER066W, YDR148C, YDR204W, YKR866C, YCR190C, YTL146C, and YJR106W. The expression patterns are shown as colored squares (red, green, black) in a grid format.

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 main area is divided into three panes:

- Left Pane:** A vertical dendrogram showing hierarchical clustering of samples. A red rectangular box highlights a specific cluster of samples.
- Middle Pane:** A heatmap visualization where rows represent genes and columns represent samples. The color scale ranges from green (low expression) to red (high expression). A red box highlights the same cluster of samples as in the dendrogram.
- Right Pane:** A table of gene annotations. The first column lists gene IDs, and the second column lists their associated biological processes. The same gene IDs as in the heatmap are listed here.

Gene Annotation Table:

Gene ID	Biological Process
YAL063W	GLUTAMATE BIOSYNTHESIS
YOR375C	GLUTAMATE BIOSYNTHESIS
YBR080C	SECRETION
YMR072W	MITOCHONDRIAL GENOME MAI (PU
YIL119W	CYTOSKELETON
YDR311W	TRANSCRIPTION
YCR274C	TRANSCRIPTION
YML106C	ENDOCYTOSIS (PUTATIVE)
YML069W	DNA REPLICATION (PUTATIV BINI
YDR481C	PHOSPHATE METABOLISM
YFL021W	NITROGEN CATABOLISM
YDR284C	PHOSPHOLIPID METABOLISM
YDR495W	PROTEIN SYNTHESIS
YAL029C	TRANSPORT
YBL043W	CELL WALL BIOGENESIS
YMR055C	CELL CYCLE, CHECKPOINT
YJL006C	CELL CYCLE
YCR252W	CHROMATIN STRUCTURE
YKL201C	PROTEIN GLYCOSYLATION
YML039W	TRANSCRIPTION
YOR290C	SECRETION
YML272C	SECRETION
YOR075W	SECRETION
YDR192C	NUCLEAR PROTEIN TARGETIN NUCL
YDL224C	CELL SIZE
YER112W	MRNA SPLICING
YDR185W	MRNA 3'-END PROCESSING
YER107C	NUCLEAR PROTEIN TARGETIN NUCL
YHR208W	BRANCHED CHAIN AMINO ACI TRAI
YER069W	MATING
YDR148C	TCA CYCLE
YDR204W	UBIQUINONE BIOSYNTHESIS
YKR066C	OXIDATIVE STRESS RESPONSA
YCR183C	GLYCOLYSIS
YIL146C	CELL WALL BIOGENESIS
YJR106W	CELL WALL BIOGENESIS

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 is a "Dendrogram" panel, which is currently empty. To the right is a table with columns: "Name", "Annotation", "NODEID", "LEFT", "RIGHT", "CORRELAT...", "NAME", and "ANNOTATI...". The table contains 24 rows of data.

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

Java TreeView : /home/karmic/Projects/Courses/BmsMinicourse/examples2/supp2data.cdt

File Settings Analysis Export Window Help

Dendrogram

View Status
Row: 115 (Y)
Column: 49 (S)
Value: 1.34

Usage Hints
Mouse over to get info

cdcl5_170
cdcl5_170
cdcl5_210
cdcl5_210
cdcl5_250
cdcl5_250
cdcl5_270
cdcl5_290
spo_9
spo_9
spo_7
spo_9
spo_11
spo5_2
spo5_11
spo_early
spo_mid
heat_0
heat_10
heat_20
heat_30
heat_60
heat_100

YFR028C CDC14 MITOSIS PROTEIN PHOS
YML069W ORC1 DNA REPLICATION ORIGIN F
YIL139C REV7 DNA REPAIR DNA POLYMEF
YNL318C NONE TRANSPORT HEXOSE PERM
YFR023W PES4 DNA REPLICATION UNKNOWN:
YHR015W MIP6 mRNA EXPORT, PUTATIVE RNA
YDR263C DLW7 DNA REPAIR (PUTATIVE) DNA
YLR045C STU2 CYTOSKELETON SPINDLE
YOR033C DHS1 DNA REPAIR EXONUCLEASE
YIL159W BNR1 CYTOSKELETON ACTIN FI
YKL042W SPC42 CYTOSKELETON SPINDLE
YML225C CNM67 CYTOSKELETON SPINDLE
YCR092C CDC10 CYTOKINESIS GTP BINDING
YLR210W CLB4 CELL CYCLE G2/M CYCLIN
YLR314C CDC3 CYTOKINESIS SEPTIN
YBR045C GIP1 GLUCOSE REPRESSION (PUTA
YDL159W CLB3 CELL CYCLE G2/M CYCLIN
YDR118W APC4 CELL CYCLE ANAPHASE-PF
YDR253C MET32 METHIONINE METABOLISM TRF
YML190W CLK1 CYTOSKELETON SPINDLE
YDR113C PDS1 CELL CYCLE ANAPHASE-T

GeneTreeAnno: /home/karmic/Projects/Courses/BmsMinicourse/examples2/supp2data.cdt

Sporulation

Name Sporulation Annotation 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		

Dock Close

- Compare the effects of different distance metrics and clustering algorithms on the data from the Eisen paper (note that the GORDER column for the human data will make comparison easier).
- Practice annotating clusters in JavaTreeView. Try to find the annotated yeast clusters from the paper. Follow the links to SGD to see if the annotations for these genes have changed in the past decade.
- Read Bioinformatics 20:3710

Reminder: we are in HSW-532 tomorrow!