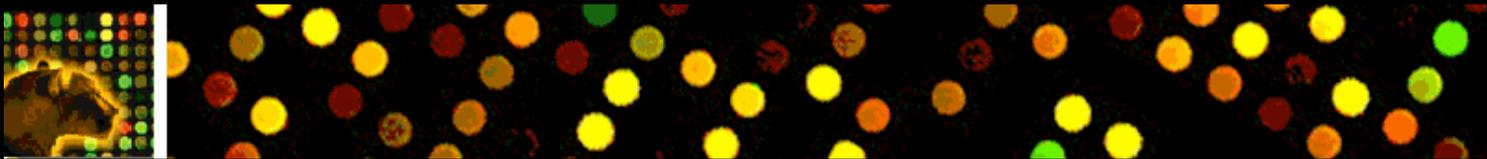
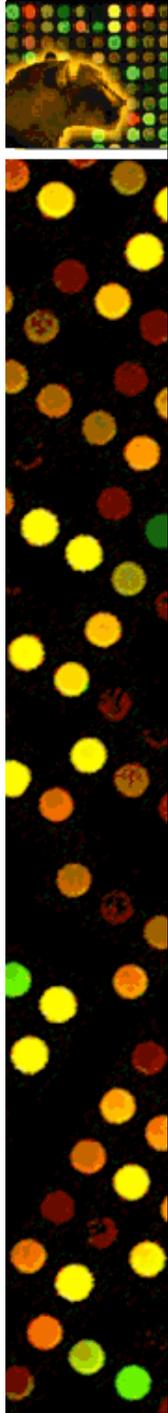


# Database Repository and Tools



John Matese

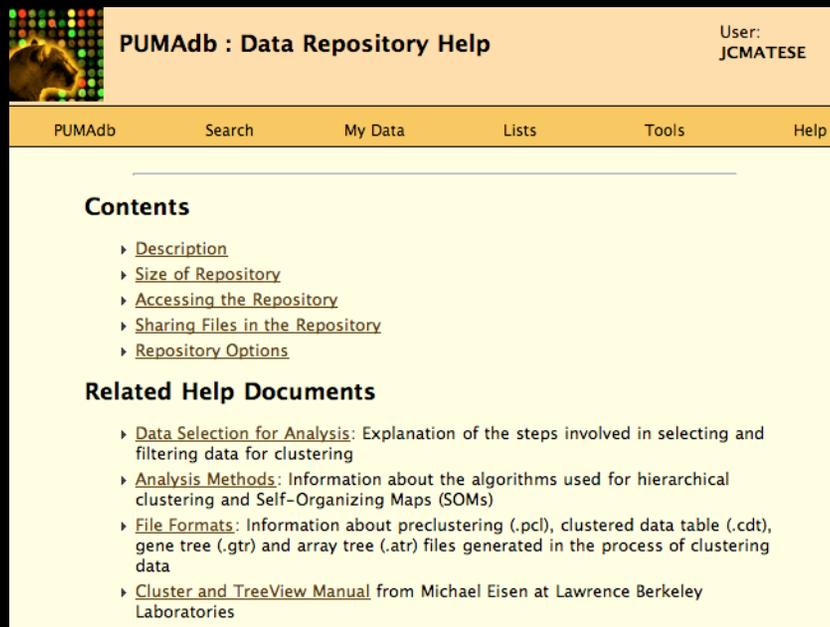
May 9, 2008



# What is the “Repository”?

- Save and exchange retrieved and analyzed **datafiles**
- Perform datafile manipulations (averaging and annotations)
- Run specialized algorithms (imputation, SVD)

# Data Repository Help



**PUMAdb : Data Repository Help** User: JCMATESE

PUMAdb Search My Data Lists Tools Help

**Contents**

- › [Description](#)
- › [Size of Repository](#)
- › [Accessing the Repository](#)
- › [Sharing Files in the Repository](#)
- › [Repository Options](#)

**Related Help Documents**

- › [Data Selection for Analysis](#): Explanation of the steps involved in selecting and filtering data for clustering
- › [Analysis Methods](#): Information about the algorithms used for hierarchical clustering and Self-Organizing Maps (SOMs)
- › [File Formats](#): Information about preclustering (.pcl), clustered data table (.cdt), gene tree (.gtr) and array tree (.atr) files generated in the process of clustering data
- › [Cluster and TreeView Manual](#) from Michael Eisen at Lawrence Berkeley Laboratories

- How to use the tools
- Sharing data
- Options
- Links to help for
  - analysis methods,
  - data file formats
  - data retrieval and clustering

<http://puma.princeton.edu/help/repository.shtml>

# File Format Help



PUMAdb : File Format Help

User:  
JCMATESE

PUMAdb

Search

My Data

Lists

Tools

Help

## Contents

- ▶ Experiment Results Files
  - xls file format
  - pcl file format
  - cdt file format
  - gtr file format
  - atr file format
- ▶ Printlist Files
  - gdl file format
  - gal file format

<http://puma.princeton.edu/help/formats.shtml>

# File Formats: Pre-clustering (PCL) File

UID is the Unique Identifier for the Spot/Reporter

NAME sequence label for the Spot/Reporter

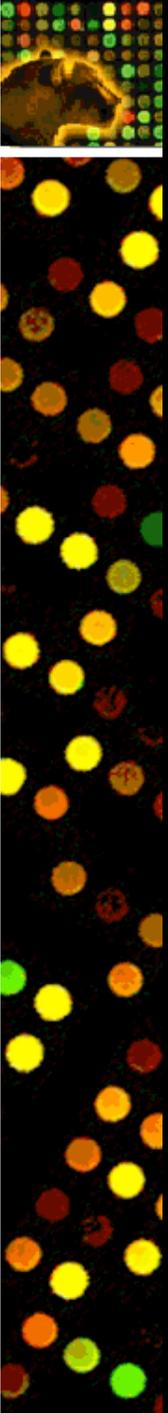
EWEIGHT indicates the weight the Array/ Assay is given in clustering

GWEIGHT indicates the weight the Spot/ Reporter is given in clustering

Names and orders of assays (if assays are not clustered)

Matrix values are for each spot/reporter on each array (usually log ratios)

sample.txt										
	A	B	C	D	E	F	G	H	I	J
1	UID	NAME	GWEIGHT	spo0	spo30	spo2	spo5	spo7	spo9	spo11
2	EWEIGHT									
3	YAL003W	EF		0.23	-1.79	-1.29	-1.56		-0.27	
4				0.41	-0.38			-1.6	-1.84	-1.6
5		SS		0.61	-0.07			-2	-1.84	-2.25
6		MI		0.16	-0.15			-1.89	-1.74	-1.6
7		CY		0.03	1.39			-2.84	-2.47	-2.4
8		NT		-0.18	-0.18			-1.69	-1.43	-1.79
9		YF		-0.51	-0.62			4.54	3.22	4.33
10		MARKS		-0.14	-3.32			-2.4	-1.03	-0.6
11	YAL034C	FUN19	1	0.19	-0.03			-1.84	-1.94	-1.74
12	YAL035W	FUN12	1	0.01	-1.47	-1.15	-0.69	-1.36	-1.64	-1.29
13	YAL036C	FUN11	1	-0.15	-2.74	-1.79	-1.32	-2.12	0.3	-0.89
14	YAL038W	CDC19	1	-0.06	-1.89	-1.69	-2.32	-2.4	-0.81	-1.6
15	YAL040C	CLN3	1	-0.17	-2.25	-1.69	-2.25	-2.56	-0.3	-2.4
16	YAL054C	ACS1	1	0.51	2.6	1.9		1.35	-0.03	-0.23
17	YAL055W	YAL055W	1	-0.32	0.83				2.05	2.24
18	YAL062W	GDH3	1	0.3	2.59				0.34	1.36
19	YAL067C	SEO1	1	-0.17	3.44				1.61	2.8
20	YAR003W	YAR003W	1	-0.29	0.54				1.86	1.42
21	YAR007C	RFA1	1	-0.14	1.74				0.57	0.84
22	YAR015W	ADE1	1	0.11	-1.51				-1.89	-2
23	YAR027W	YAR027W	1	0.24	-1.06	-1.36		-1.25	-0.94	-1.36
24	YBL009W	YBL009W	1	-0.01	0.62	1.04	1.3	2.52	2.15	2.24
25	YBL010C	YBL010C	1	0.01	0.21	0.7	1.45	2.25	1.77	1.24
26	YBL015W	ACH1	1	0.52	1.01	1.49	1.75	1.49	0.58	0.19
27	YBL027W	RPL19A	1	0.01	-1.84	-0.97	-1.47	-1.79	-1	-0.51



# What is the Repository?

- A method to save data sets to prevent repeatedly performing the same data retrieval
- A method to share processed data with others
- A way we can provide you with web access to new and/or computationally-intensive tools

# Accessing the Repository



The screenshot shows the PUMAdb website navigation menu. The 'My Data' menu is open, showing options like 'Enter My Data', 'Display My Data', 'Select My Data', 'My Repository', 'Add Protocol', and 'Procedural Info'. The 'List' menu is also open, showing a long list of options including 'Users', 'User Group', 'Organism List', 'Sequence Types', 'SUID Retrieval', 'Projects', 'Plate Storage Locations', 'Printers', 'Print config', 'Categories', 'SubCategories', 'Normalizations', 'Plate List', 'Print List', 'Experiment List', 'Procedures', 'Protocol', 'Parameters', 'Experiment Types', 'Experiment Set Types', 'View Clinical Data', 'Repository List', 'Publication List', 'Meta Data List', and 'Ontology Browser'. An orange callout box points to the 'Repository List' option in the 'List' menu, containing the text 'from either navigation menu or lists'.

PUMAdb Search My Data Lists Tools Help

**Search**

- ▶ [Advanced Search](#)
- ▶ [Basic Search](#)
- ▶ [Gene Search](#)

**EnterData**

- ▶ [Experiments and Results](#)
- ▶ [Replace Proxy Image](#)
- ▶ [Patient and Clinical Data](#)
- ▶ [Procedural Information](#)
- ▶ [Protocol](#)

**List**

- ▶ [Users](#)
- ▶ [User Group](#)
- ▶ [Organism List](#)
- ▶ [Sequence Types](#)
- ▶ [SUID Retrieval](#)
- ▶ [Projects](#)
- ▶ [Plate Storage Locations](#)
- ▶ [Printers](#)
- ▶ [Print config](#)
- ▶ [Categories](#)
- ▶ [SubCategories](#)
- ▶ [Normalizations](#)
- ▶ [Plate List](#)
- ▶ [Print List](#)
- ▶ [Experiment List](#)
- ▶ [Procedures](#)
- ▶ [Protocol](#)
- ▶ [Parameters](#)
- ▶ [Experiment Types](#)
- ▶ [Experiment Set Types](#)
- ▶ [View Clinical Data](#)
- ▶ [Repository List](#)
- ▶ [Publication List](#)
- ▶ [Meta Data List](#)
- ▶ [Ontology Browser](#)

from either navigation menu or lists

# Repository overview

## Links

- "Home",
- Upload,
- Delete

[MY REPOSITORY](#) | [UPLOAD](#) | [BULK DELETE](#)

Microarray Repository for John Matese. Contains **50 entries**.

Click on a header to sort by that column.

1 2 3 >

Name	Organism	Date	Type	Genes	Expts.	Size	Options (help)
barcode cluster	<i>Saccharomyces cerevisiae</i>	04/25/06	CDT	20036	14	77740 kB	<a href="#">View</a> <a href="#">Delete</a> <a href="#">Edit</a> <a href="#">G</a> <a href="#">X</a> <a href="#">Filter</a> <a href="#">SVD</a> <a href="#">Synth</a>
Clustered data : Fig 3A of Whitfield et al : human cell cycle	<i>Homo sapiens</i>	04/21/04	CDT	1134	118	3929 kB	<a href="#">View</a> <a href="#">Delete</a> <a href="#">Edit</a> <a href="#">G</a> <a href="#">X</a> <a href="#">Filter</a> <a href="#">SVD</a> <a href="#">Synth</a>
derisi pathways	<i>Saccharomyces cerevisiae</i>	04/24/04	CDT	59	7	73 kB	<a href="#">View</a> <a href="#">Delete</a> <a href="#">Edit</a> <a href="#">G</a> <a href="#">X</a> <a href="#">Filter</a> <a href="#">SVD</a> <a href="#">Synth</a>
diauxic-genesonly-cluster	<i>Saccharomyces cerevisiae</i>	08/06/04	CDT	319	7	381 kB	<a href="#">View</a> <a href="#">Delete</a> <a href="#">Edit</a> <a href="#">G</a> <a href="#">X</a> <a href="#">Filter</a> <a href="#">SVD</a> <a href="#">Synth</a>
HessAlteredCluster	<i>Saccharomyces cerevisiae</i>	11/19/04	CDT	656	11	886 kB	<a href="#">View</a> <a href="#">Delete</a> <a href="#">Edit</a> <a href="#">G</a> <a href="#">X</a> <a href="#">Filter</a> <a href="#">SVD</a> <a href="#">Synth</a>
malarial phaseogram cluster	<i>Plasmodium falciparum</i>	01/06/06	CDT	1283	46	3540 kB	<a href="#">View</a> <a href="#">Delete</a> <a href="#">Edit</a> <a href="#">G</a> <a href="#">X</a> <a href="#">Filter</a> <a href="#">SVD</a> <a href="#">Synth</a>
33MIN averaged	<i>Plasmodium falciparum</i>	11/20/06	PCL	3250	6	177 kB	<a href="#">View</a> <a href="#">Delete</a> <a href="#">Edit</a> <a href="#">Filter</a> <a href="#">SVD</a> <a href="#">Synth</a>
act_low_phos	<i>Saccharomyces cerevisiae</i>	01/24/05	PCL	5657	5	970 kB	<a href="#">View</a> <a href="#">Delete</a> <a href="#">Edit</a> <a href="#">Filter</a> <a href="#">SVD</a> <a href="#">Synth</a>
Agilent identifiers	<i>Saccharomyces cerevisiae</i>	11/15/04	PCL	6060	3	176 kB	<a href="#">View</a> <a href="#">Delete</a> <a href="#">Edit</a> <a href="#">Filter</a> <a href="#">SVD</a> <a href="#">Synth</a>
barcode preliminary	<i>Saccharomyces cerevisiae</i>	04/25/06	PCL	20036	14	9189 kB	<a href="#">View</a> <a href="#">Delete</a> <a href="#">Edit</a> <a href="#">Filter</a> <a href="#">SVD</a> <a href="#">Synth</a>
barcode survey	<i>Saccharomyces cerevisiae</i>	04/27/06	PCL	20036	20	9311 kB	<a href="#">View</a> <a href="#">Delete</a> <a href="#">Edit</a> <a href="#">Filter</a> <a href="#">SVD</a> <a href="#">Synth</a>
barcode survey centered	<i>Saccharomyces cerevisiae</i>	04/27/06	PCL	20036	20	9516 kB	<a href="#">View</a> <a href="#">Delete</a> <a href="#">Edit</a> <a href="#">Filter</a> <a href="#">SVD</a> <a href="#">Synth</a>
batch 1 & 2 imputed	<i>Acyrtosiphon pisum</i>	10/21/06	PCL	1129	24	212 kB	<a href="#">View</a> <a href="#">Delete</a> <a href="#">Edit</a> <a href="#">Filter</a> <a href="#">SVD</a> <a href="#">Synth</a>
ch1 intensity	<i>Acyrtosiphon pisum</i>	06/14/06	PCL	7589	15	1793 kB	<a href="#">View</a> <a href="#">Delete</a> <a href="#">Edit</a> <a href="#">Filter</a> <a href="#">SVD</a> <a href="#">Synth</a>
diauxic-geneonly	<i>Saccharomyces cerevisiae</i>	08/06/04	PCL	319	7	17 kB	<a href="#">View</a> <a href="#">Delete</a> <a href="#">Edit</a> <a href="#">Filter</a> <a href="#">SVD</a> <a href="#">Synth</a>
diauxic-mostannotation	<i>Saccharomyces cerevisiae</i>	08/06/04	PCL	319	7	36 kB	<a href="#">View</a> <a href="#">Delete</a> <a href="#">Edit</a> <a href="#">Filter</a> <a href="#">SVD</a> <a href="#">Synth</a>
glucose pcl	<i>Saccharomyces cerevisiae</i>	01/25/05	PCL	2875	4	466 kB	<a href="#">View</a> <a href="#">Delete</a> <a href="#">Edit</a> <a href="#">Filter</a> <a href="#">SVD</a> <a href="#">Synth</a>
growth rate pcl	<i>Saccharomyces cerevisiae</i>	09/28/07	PCL	1469	36	1441 kB	<a href="#">View</a> <a href="#">Delete</a> <a href="#">Edit</a> <a href="#">Filter</a> <a href="#">SVD</a> <a href="#">Synth</a>
GSE1639 conversion	<i>Saccharomyces cerevisiae</i>	01/14/08	PCL	9335	18	1665 kB	<a href="#">View</a> <a href="#">Delete</a> <a href="#">Edit</a> <a href="#">Filter</a> <a href="#">SVD</a> <a href="#">Synth</a>
imputed	<i>Acyrtosiphon pisum</i>	06/14/06	PCL	7589	15	1130 kB	<a href="#">View</a> <a href="#">Delete</a> <a href="#">Edit</a> <a href="#">Filter</a> <a href="#">SVD</a> <a href="#">Synth</a>

1 2 3 >

[MY REPOSITORY](#) | [UPLOAD](#) | [BULK DELETE](#)

View the repository of

You are currently using 168.80 MB of space, of an available 400 MB.

# Repository overview

- Which repository?
- How many deposits?
- How much space?

MY REPOSITORY | [UPLOAD](#) | [BULK DELETE](#)

Microarray Repository for John Matese. Contains **50 entries**.  
Click on a header to sort by that column.

1 2 3 >

Name	Organism	Date	Type	Genes	Expts.	Size	Options (help)
barcode cluster	<i>Saccharomyces cerevisiae</i>	04/25/06	CDT	20036	14	77740 kB	<a href="#">View</a> <a href="#">Delete</a> <a href="#">Edit</a> <a href="#">Filter</a> <a href="#">SVD</a> <a href="#">Synth</a>
Clustered data : Fig 3A of Whitfield et al : human cell cycle	<i>Homo sapiens</i>	04/21/04	CDT	1134	118	3929 kB	<a href="#">View</a> <a href="#">Delete</a> <a href="#">Edit</a> <a href="#">Filter</a> <a href="#">SVD</a> <a href="#">Synth</a>
derisi pathways	<i>Saccharomyces cerevisiae</i>	04/24/04	CDT	59	7	73 kB	<a href="#">View</a> <a href="#">Delete</a> <a href="#">Edit</a> <a href="#">Filter</a> <a href="#">SVD</a> <a href="#">Synth</a>
diauxic-genesonly-cluster	<i>Saccharomyces cerevisiae</i>	08/06/04	CDT	319	7	381 kB	<a href="#">View</a> <a href="#">Delete</a> <a href="#">Edit</a> <a href="#">Filter</a> <a href="#">SVD</a> <a href="#">Synth</a>
HessAlteredCluster	<i>Saccharomyces cerevisiae</i>	11/19/04	CDT	656	11	886 kB	<a href="#">View</a> <a href="#">Delete</a> <a href="#">Edit</a> <a href="#">Filter</a> <a href="#">SVD</a> <a href="#">Synth</a>
malarial phaseogram cluster	<i>Plasmodium falciparum</i>	01/06/06	CDT	1283	46	3540 kB	<a href="#">View</a> <a href="#">Delete</a> <a href="#">Edit</a> <a href="#">Filter</a> <a href="#">SVD</a> <a href="#">Synth</a>
33MIN averaged	<i>Plasmodium falciparum</i>	11/20/06	PCL	3250	6	177 kB	<a href="#">View</a> <a href="#">Delete</a> <a href="#">Edit</a> <a href="#">Filter</a> <a href="#">SVD</a> <a href="#">Synth</a>
act_low_phos	<i>Saccharomyces cerevisiae</i>	01/24/05	PCL	5657	5	970 kB	<a href="#">View</a> <a href="#">Delete</a> <a href="#">Edit</a> <a href="#">Filter</a> <a href="#">SVD</a> <a href="#">Synth</a>
Agilent identifiers	<i>Saccharomyces cerevisiae</i>	11/15/04	PCL	6060	3	176 kB	<a href="#">View</a> <a href="#">Delete</a> <a href="#">Edit</a> <a href="#">Filter</a> <a href="#">SVD</a> <a href="#">Synth</a>
barcode preliminary	<i>Saccharomyces cerevisiae</i>	04/25/06	PCL	20036	14	9189 kB	<a href="#">View</a> <a href="#">Delete</a> <a href="#">Edit</a> <a href="#">Filter</a> <a href="#">SVD</a> <a href="#">Synth</a>
barcode survey	<i>Saccharomyces cerevisiae</i>	04/27/06	PCL	20036	20	9311 kB	<a href="#">View</a> <a href="#">Delete</a> <a href="#">Edit</a> <a href="#">Filter</a> <a href="#">SVD</a> <a href="#">Synth</a>
barcode survey centered	<i>Saccharomyces cerevisiae</i>	04/27/06	PCL	20036	20	9516 kB	<a href="#">View</a> <a href="#">Delete</a> <a href="#">Edit</a> <a href="#">Filter</a> <a href="#">SVD</a> <a href="#">Synth</a>
batch 1 & 2 imputed	<i>Acyrtosiphon pisum</i>	10/21/06	PCL	1129	24	212 kB	<a href="#">View</a> <a href="#">Delete</a> <a href="#">Edit</a> <a href="#">Filter</a> <a href="#">SVD</a> <a href="#">Synth</a>
ch1 intensity	<i>Acyrtosiphon pisum</i>	06/14/06	PCL	7589	15	1793 kB	<a href="#">View</a> <a href="#">Delete</a> <a href="#">Edit</a> <a href="#">Filter</a> <a href="#">SVD</a> <a href="#">Synth</a>
diauxic-geneonly	<i>Saccharomyces cerevisiae</i>	08/06/04	PCL	319	7	17 kB	<a href="#">View</a> <a href="#">Delete</a> <a href="#">Edit</a> <a href="#">Filter</a> <a href="#">SVD</a> <a href="#">Synth</a>
diauxic-mostannotation	<i>Saccharomyces cerevisiae</i>	08/06/04	PCL	319	7	36 kB	<a href="#">View</a> <a href="#">Delete</a> <a href="#">Edit</a> <a href="#">Filter</a> <a href="#">SVD</a> <a href="#">Synth</a>
glucose pcl	<i>Saccharomyces cerevisiae</i>	01/25/05	PCL	2875	4	466 kB	<a href="#">View</a> <a href="#">Delete</a> <a href="#">Edit</a> <a href="#">Filter</a> <a href="#">SVD</a> <a href="#">Synth</a>
growth rate pcl	<i>Saccharomyces cerevisiae</i>	09/28/07	PCL	1469	36	1441 kB	<a href="#">View</a> <a href="#">Delete</a> <a href="#">Edit</a> <a href="#">Filter</a> <a href="#">SVD</a> <a href="#">Synth</a>
GSE1639 conversion	<i>Saccharomyces cerevisiae</i>	01/14/08	PCL	9335	18	1665 kB	<a href="#">View</a> <a href="#">Delete</a> <a href="#">Edit</a> <a href="#">Filter</a> <a href="#">SVD</a> <a href="#">Synth</a>
imputed	<i>Acyrtosiphon pisum</i>	06/14/06	PCL	7589	15	1130 kB	<a href="#">View</a> <a href="#">Delete</a> <a href="#">Edit</a> <a href="#">Filter</a> <a href="#">SVD</a> <a href="#">Synth</a>

1 2 3 >

MY REPOSITORY | [UPLOAD](#) | [BULK DELETE](#)

View the repository of

You are currently using 168.80 MB of space, of an available 400 MB.

# Repository overview

[MY REPOSITORY](#) | [UPLOAD](#) | [BULK DELETE](#)

Microarray Repository for John Matese. Contains **50 entries**.

Click on a header to sort by that column.

1 2 3 >

Name	Organism	Date	Type	Genes	Expts.	Size	Options (help)
barcode cluster	<i>Saccharomyces cerevisiae</i>	04/25/06	CDT	20036	14	77740 kB	<a href="#">View</a> <a href="#">Delete</a> <a href="#">Edit</a> <a href="#">G</a> <a href="#">X</a> <a href="#">Filter</a> <a href="#">SVD</a> <a href="#">Synth</a>
Clustered data : Fig 3A of Whitfield et al : human cell cycle	<i>Homo sapiens</i>	04/21/04	CDT	1134	118	3929 kB	<a href="#">View</a> <a href="#">Delete</a> <a href="#">Edit</a> <a href="#">G</a> <a href="#">X</a> <a href="#">Filter</a> <a href="#">SVD</a> <a href="#">Synth</a>
derisi pathways	<i>Saccharomyces cerevisiae</i>	04/24/04	CDT	59	7	73 kB	<a href="#">View</a> <a href="#">Delete</a> <a href="#">Edit</a> <a href="#">G</a> <a href="#">X</a> <a href="#">Filter</a> <a href="#">SVD</a> <a href="#">Synth</a>
diauxic-genesonly-cluster	<i>Saccharomyces cerevisiae</i>	08/06/04	CDT	319	7	381 kB	<a href="#">View</a> <a href="#">Delete</a> <a href="#">Edit</a> <a href="#">G</a> <a href="#">X</a> <a href="#">Filter</a> <a href="#">SVD</a> <a href="#">Synth</a>
HessAlteredCluster	<i>Saccharomyces cerevisiae</i>	11/19/04	CDT	656	11	886 kB	<a href="#">View</a> <a href="#">Delete</a> <a href="#">Edit</a> <a href="#">G</a> <a href="#">X</a> <a href="#">Filter</a> <a href="#">SVD</a> <a href="#">Synth</a>
malarial phaseogram cluster	<i>Plasmodium falciparum</i>	01/06/06	CDT	1283	46	3540 kB	<a href="#">View</a> <a href="#">Delete</a> <a href="#">Edit</a> <a href="#">G</a> <a href="#">X</a> <a href="#">Filter</a> <a href="#">SVD</a> <a href="#">Synth</a>
33MIN averaged	<i>Plasmodium falciparum</i>	11/20/06	PCL	3250	6	177 kB	<a href="#">View</a> <a href="#">Delete</a> <a href="#">Edit</a> <a href="#">Filter</a> <a href="#">SVD</a> <a href="#">Synth</a>
act_low_phos	<i>Saccharomyces cerevisiae</i>	01/24/05	PCL	5657	5	970 kB	<a href="#">View</a> <a href="#">Delete</a> <a href="#">Edit</a> <a href="#">Filter</a> <a href="#">SVD</a> <a href="#">Synth</a>
Agilent identifiers	<i>Saccharomyces cerevisiae</i>	11/15/04	PCL	6060	3	176 kB	<a href="#">View</a> <a href="#">Delete</a> <a href="#">Edit</a> <a href="#">Filter</a> <a href="#">SVD</a> <a href="#">Synth</a>
barcode preliminary	<i>Saccharomyces cerevisiae</i>	04/25/06	PCL	20036	14	9189 kB	<a href="#">View</a> <a href="#">Delete</a> <a href="#">Edit</a> <a href="#">Filter</a> <a href="#">SVD</a> <a href="#">Synth</a>
barcode survey	<i>Saccharomyces cerevisiae</i>	04/27/06	PCL	20036	20	9311 kB	<a href="#">View</a> <a href="#">Delete</a> <a href="#">Edit</a> <a href="#">Filter</a> <a href="#">SVD</a> <a href="#">Synth</a>
barcode survey centered	<i>Saccharomyces cerevisiae</i>	04/27/06	PCL	20036	20	9516 kB	<a href="#">View</a> <a href="#">Delete</a> <a href="#">Edit</a> <a href="#">Filter</a> <a href="#">SVD</a> <a href="#">Synth</a>
batch 1 & 2 imputed	<i>Acyrtosiphon pisum</i>	10/21/06	PCL	1129	24	212 kB	<a href="#">View</a> <a href="#">Delete</a> <a href="#">Edit</a> <a href="#">Filter</a> <a href="#">SVD</a> <a href="#">Synth</a>
ch1 intensity	<i>Acyrtosiphon pisum</i>	06/14/06	PCL	7589	15	1793 kB	<a href="#">View</a> <a href="#">Delete</a> <a href="#">Edit</a> <a href="#">Filter</a> <a href="#">SVD</a> <a href="#">Synth</a>
diauxic-geneonly	<i>Saccharomyces cerevisiae</i>	08/06/04	PCL	319	7	17 kB	<a href="#">View</a> <a href="#">Delete</a> <a href="#">Edit</a> <a href="#">Filter</a> <a href="#">SVD</a> <a href="#">Synth</a>
diauxic-mostannotation	<i>Saccharomyces cerevisiae</i>	08/06/04	PCL	319	7	36 kB	<a href="#">View</a> <a href="#">Delete</a> <a href="#">Edit</a> <a href="#">Filter</a> <a href="#">SVD</a> <a href="#">Synth</a>
glucose pcl	<i>Saccharomyces cerevisiae</i>	01/25/05	PCL	2875	4	466 kB	<a href="#">View</a> <a href="#">Delete</a> <a href="#">Edit</a> <a href="#">Filter</a> <a href="#">SVD</a> <a href="#">Synth</a>
growth rate pcl	<i>Saccharomyces cerevisiae</i>	09/28/07	PCL	1469	36	1441 kB	<a href="#">View</a> <a href="#">Delete</a> <a href="#">Edit</a> <a href="#">Filter</a> <a href="#">SVD</a> <a href="#">Synth</a>
GSE1639 conversion	<i>Saccharomyces cerevisiae</i>	01/14/08	PCL	9335	18	1665 kB	<a href="#">View</a> <a href="#">Delete</a> <a href="#">Edit</a> <a href="#">Filter</a> <a href="#">SVD</a> <a href="#">Synth</a>
imputed	<i>Acyrtosiphon pisum</i>	06/14/06	PCL	7589	15	1130 kB	<a href="#">View</a> <a href="#">Delete</a> <a href="#">Edit</a> <a href="#">Filter</a> <a href="#">SVD</a> <a href="#">Synth</a>

1 2 3 >

[MY REPOSITORY](#) | [UPLOAD](#) | [BULK DELETE](#)

View the repository of

You are currently using 168.80 MB of space, of an available 400 MB.

Page Navigation

# Repository overview

Deposit  
Sorting

MY REPOSITORY | [UPLOAD](#) | [BULK DELETE](#)

Microarray Repository for John Matese. Contains **50 entries**.

Click on a header to sort by that column.

1 2 3 >

Name	Organism	Date	Type	Genes	Expts.	Size	Options (help)
barcode cluster	<i>Saccharomyces cerevisiae</i>	04/25/06	CDT	20036	14	77740 kB	<a href="#">View</a> <a href="#">Delete</a> <a href="#">Edit</a> <a href="#">G</a> <a href="#">X</a> <a href="#">Filter</a> <a href="#">SVD</a> <a href="#">Synth</a>
Clustered data : Fig 3A of Whitfield et al : human cell cycle	<i>Homo sapiens</i>	04/21/04	CDT	1134	118	3929 kB	<a href="#">View</a> <a href="#">Delete</a> <a href="#">Edit</a> <a href="#">G</a> <a href="#">X</a> <a href="#">Filter</a> <a href="#">SVD</a> <a href="#">Synth</a>
derisi pathways	<i>Saccharomyces cerevisiae</i>	04/24/04	CDT	59	7	73 kB	<a href="#">View</a> <a href="#">Delete</a> <a href="#">Edit</a> <a href="#">G</a> <a href="#">X</a> <a href="#">Filter</a> <a href="#">SVD</a> <a href="#">Synth</a>
diauxic-genesonly-cluster	<i>Saccharomyces cerevisiae</i>	08/06/04	CDT	319	7	381 kB	<a href="#">View</a> <a href="#">Delete</a> <a href="#">Edit</a> <a href="#">G</a> <a href="#">X</a> <a href="#">Filter</a> <a href="#">SVD</a> <a href="#">Synth</a>
HessAlteredCluster	<i>Saccharomyces cerevisiae</i>	11/19/04	CDT	656	11	886 kB	<a href="#">View</a> <a href="#">Delete</a> <a href="#">Edit</a> <a href="#">G</a> <a href="#">X</a> <a href="#">Filter</a> <a href="#">SVD</a> <a href="#">Synth</a>
malarial phaseogram cluster	<i>Plasmodium falciparum</i>	01/06/06	CDT	1283	46	3540 kB	<a href="#">View</a> <a href="#">Delete</a> <a href="#">Edit</a> <a href="#">G</a> <a href="#">X</a> <a href="#">Filter</a> <a href="#">SVD</a> <a href="#">Synth</a>
33MIN averaged	<i>Plasmodium falciparum</i>	11/20/06	PCL	3250	6	177 kB	<a href="#">View</a> <a href="#">Delete</a> <a href="#">Edit</a> <a href="#">Filter</a> <a href="#">SVD</a> <a href="#">Synth</a>
act_low_phos	<i>Saccharomyces cerevisiae</i>	01/24/05	PCL	5657	5	970 kB	<a href="#">View</a> <a href="#">Delete</a> <a href="#">Edit</a> <a href="#">Filter</a> <a href="#">SVD</a> <a href="#">Synth</a>
Agilent identifiers	<i>Saccharomyces cerevisiae</i>	11/15/04	PCL	6060	3	176 kB	<a href="#">View</a> <a href="#">Delete</a> <a href="#">Edit</a> <a href="#">Filter</a> <a href="#">SVD</a> <a href="#">Synth</a>
barcode preliminary	<i>Saccharomyces cerevisiae</i>	04/25/06	PCL	20036	14	9189 kB	<a href="#">View</a> <a href="#">Delete</a> <a href="#">Edit</a> <a href="#">Filter</a> <a href="#">SVD</a> <a href="#">Synth</a>
barcode survey	<i>Saccharomyces cerevisiae</i>	04/27/06	PCL	20036	20	9311 kB	<a href="#">View</a> <a href="#">Delete</a> <a href="#">Edit</a> <a href="#">Filter</a> <a href="#">SVD</a> <a href="#">Synth</a>
barcode survey centered	<i>Saccharomyces cerevisiae</i>	04/27/06	PCL	20036	20	9516 kB	<a href="#">View</a> <a href="#">Delete</a> <a href="#">Edit</a> <a href="#">Filter</a> <a href="#">SVD</a> <a href="#">Synth</a>
batch 1 & 2 imputed	<i>Acyrtosiphon pisum</i>	10/21/06	PCL	1129	24	212 kB	<a href="#">View</a> <a href="#">Delete</a> <a href="#">Edit</a> <a href="#">Filter</a> <a href="#">SVD</a> <a href="#">Synth</a>
ch1 intensity	<i>Acyrtosiphon pisum</i>	06/14/06	PCL	7589	15	1793 kB	<a href="#">View</a> <a href="#">Delete</a> <a href="#">Edit</a> <a href="#">Filter</a> <a href="#">SVD</a> <a href="#">Synth</a>
diauxic-geneonly	<i>Saccharomyces cerevisiae</i>	08/06/04	PCL	319	7	17 kB	<a href="#">View</a> <a href="#">Delete</a> <a href="#">Edit</a> <a href="#">Filter</a> <a href="#">SVD</a> <a href="#">Synth</a>
diauxic-mostannotation	<i>Saccharomyces cerevisiae</i>	08/06/04	PCL	319	7	36 kB	<a href="#">View</a> <a href="#">Delete</a> <a href="#">Edit</a> <a href="#">Filter</a> <a href="#">SVD</a> <a href="#">Synth</a>
glucose pcl	<i>Saccharomyces cerevisiae</i>	01/25/05	PCL	2875	4	466 kB	<a href="#">View</a> <a href="#">Delete</a> <a href="#">Edit</a> <a href="#">Filter</a> <a href="#">SVD</a> <a href="#">Synth</a>
growth rate pcl	<i>Saccharomyces cerevisiae</i>	09/28/07	PCL	1469	36	1441 kB	<a href="#">View</a> <a href="#">Delete</a> <a href="#">Edit</a> <a href="#">Filter</a> <a href="#">SVD</a> <a href="#">Synth</a>
GSE1639 conversion	<i>Saccharomyces cerevisiae</i>	01/14/08	PCL	9335	18	1665 kB	<a href="#">View</a> <a href="#">Delete</a> <a href="#">Edit</a> <a href="#">Filter</a> <a href="#">SVD</a> <a href="#">Synth</a>
imputed	<i>Acyrtosiphon pisum</i>	06/14/06	PCL	7589	15	1130 kB	<a href="#">View</a> <a href="#">Delete</a> <a href="#">Edit</a> <a href="#">Filter</a> <a href="#">SVD</a> <a href="#">Synth</a>

1 2 3 >

MY REPOSITORY | [UPLOAD](#) | [BULK DELETE](#)

View the repository of

You are currently using 168.80 MB of space, of an available 400 MB.

# Repository overview

MY REPOSITORY | [UPLOAD](#) | [BULK DELETE](#)

Microarray Repository for John Matese. Contains **50 entries**.

Click on a header to sort by that column.

1 2 3 >

Name	Organism	Date	Type	Genes	Expts.	Size	Options (help)
barcode cluster	<i>Saccharomyces cerevisiae</i>	04/25/06	CDT	20036	14	77740 kB	View PA De-Delete Edit g X
Clustered data : Fig 3A of Whitfield et al : human cell cycle	<i>Homo sapiens</i>	04/21/04	CDT	1134	118	3929 kB	View PA De-Delete Edit g X
derisi pathways	<i>Saccharomyces cerevisiae</i>	04/24/04	CDT	59	7	73 kB	View PA De-Delete Edit g X
diauxic-genesonly-cluster	<i>Saccharomyces cerevisiae</i>	08/06/04	CDT	319	7	381 kB	View PA De-Delete Edit g X
HessAlteredCluster	<i>Saccharomyces cerevisiae</i>	11/19/04	CDT	656	11	886 kB	View PA De-Delete Edit g X
malarial phaseogram cluster	<i>Plasmodium falciparum</i>	01/06/06	CDT	1283	46	3540 kB	View PA De-Delete Edit g X
33MIN averaged	<i>Plasmodium falciparum</i>	11/20/06	PCL	3250	6	177 kB	View PA De-Delete Edit Filter SVD Synth
act_low_phos	<i>Saccharomyces cerevisiae</i>	01/24/05	PCL	5657	5	970 kB	View PA De-Delete Edit Filter SVD Synth
Agilent identifiers	<i>Saccharomyces cerevisiae</i>	11/15/04	PCL	6060	3	176 kB	View PA De-Delete Edit Filter SVD Synth
barcode preliminary	<i>Saccharomyces cerevisiae</i>	04/25/06	PCL	20036	14	9189 kB	View PA De-Delete Edit Filter SVD Synth
barcode survey	<i>Saccharomyces cerevisiae</i>	04/27/06	PCL	20036	20	9311 kB	View PA De-Delete Edit Filter SVD Synth
barcode survey centered	<i>Saccharomyces cerevisiae</i>	04/27/06	PCL	20036	20	9516 kB	View PA De-Delete Edit Filter SVD Synth
batch 1 & 2 imputed	<i>Acyrtosiphon pisum</i>	10/21/06	PCL	1129	24	212 kB	View PA De-Delete Edit Filter SVD Synth
ch1 intensity	<i>Acyrtosiphon pisum</i>	06/14/06	PCL	7589	15	1793 kB	View PA De-Delete Edit Filter SVD Synth
diauxic-geneonly	<i>Saccharomyces cerevisiae</i>	08/06/04	PCL	319	7	17 kB	View PA De-Delete Edit Filter SVD Synth
diauxic-mostannotation	<i>Saccharomyces cerevisiae</i>	08/06/04	PCL	319	7	36 kB	View PA De-Delete Edit Filter SVD Synth
glucose pcl	<i>Saccharomyces cerevisiae</i>	01/25/05	PCL	2875	4	466 kB	View PA De-Delete Edit Filter SVD Synth
growth rate pcl	<i>Saccharomyces cerevisiae</i>	09/28/07	PCL	1469	36	1441 kB	View PA De-Delete Edit Filter SVD Synth
GSE1639 conversion	<i>Saccharomyces cerevisiae</i>	01/14/08	PCL	9335	18	1665 kB	View PA De-Delete Edit Filter SVD Synth
imputed	<i>Acyrtosiphon pisum</i>	06/14/06	PCL	7589	15	1130 kB	View PA De-Delete Edit Filter SVD Synth

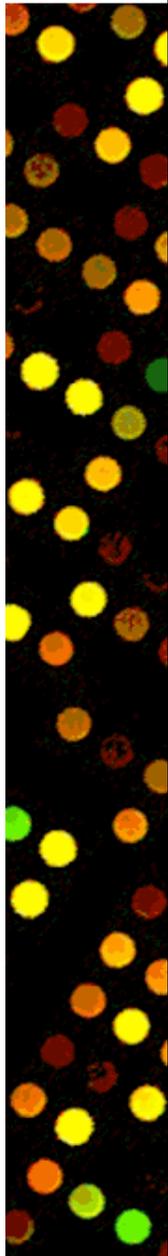
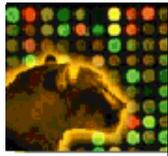
1 2 3 >

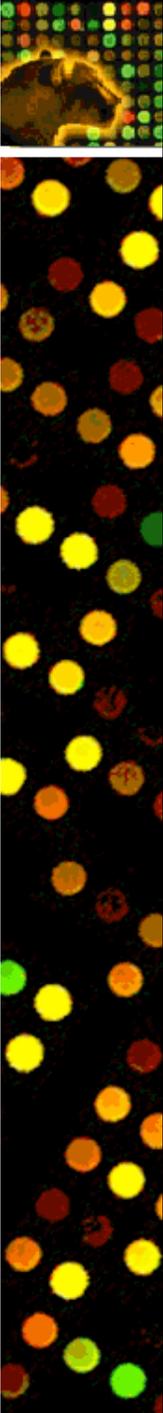
MY REPOSITORY | [UPLOAD](#) | [BULK DELETE](#)

Repository of ALEGESSE (Aster Legesse-Miller) Submit

You are currently using 168.80 MB of space, of an available 400 MB.

Other  
Repositories





# Depositing Data into your Repository

[Download PreClustering File](#)  
[View data retrieval summary report](#)  
[Add this preclustering file to your repository](#)

Proceed to Gene Filtering

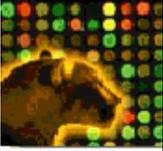
- Deposit, from data retrieval “pipeline”

[MY REPOSITORY](#) [UPLOAD](#) [BULK DELETE](#)

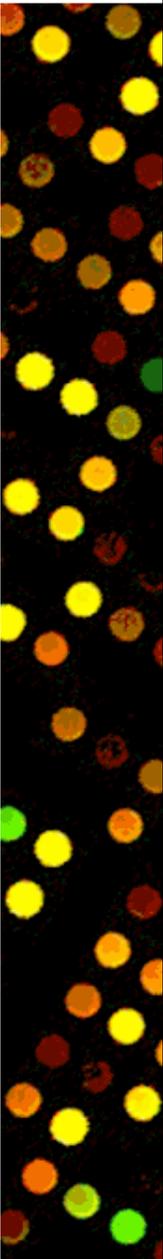
Microarray Repository for John Matese.

- Upload, from local workstation within repository itself



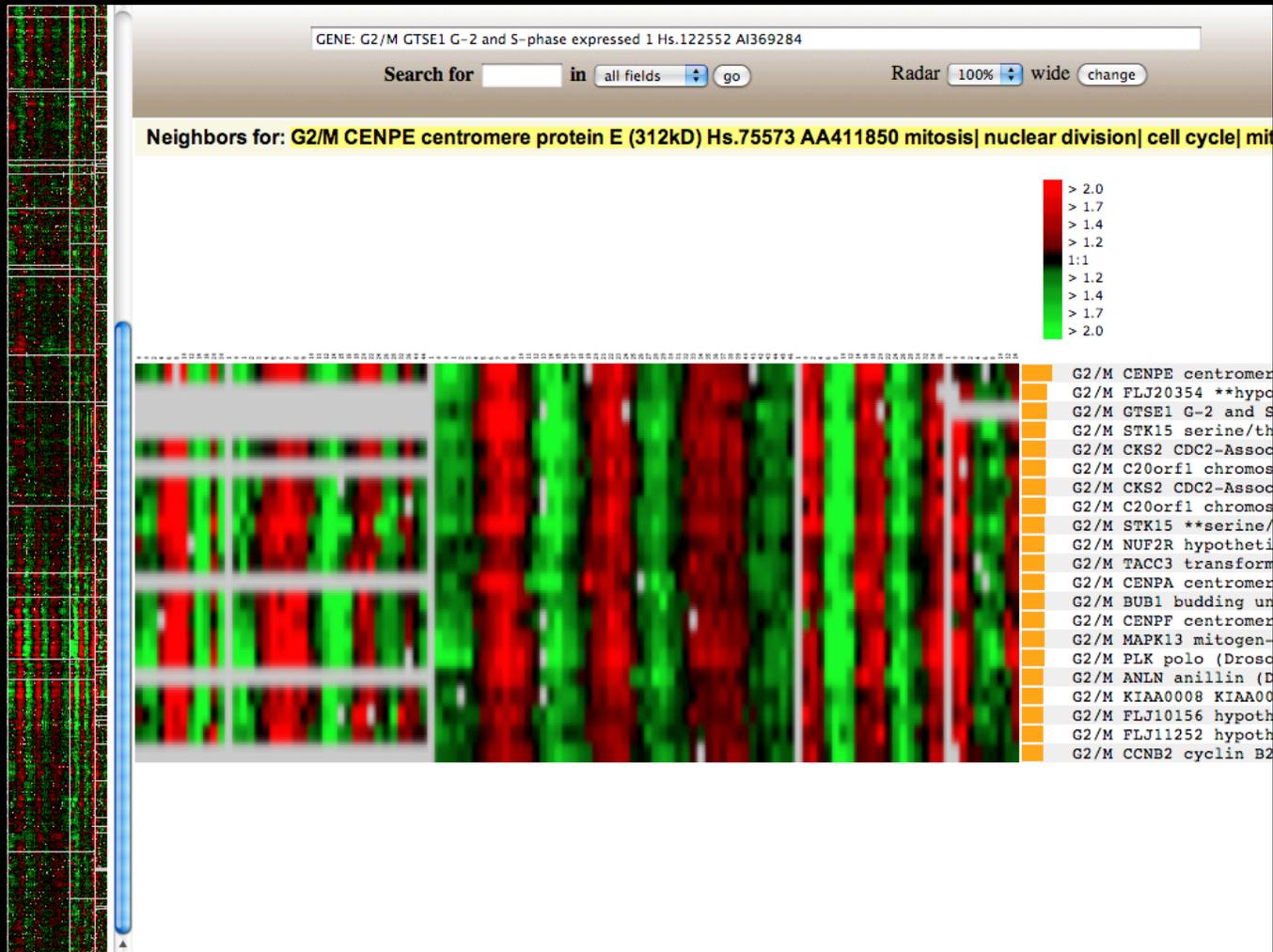
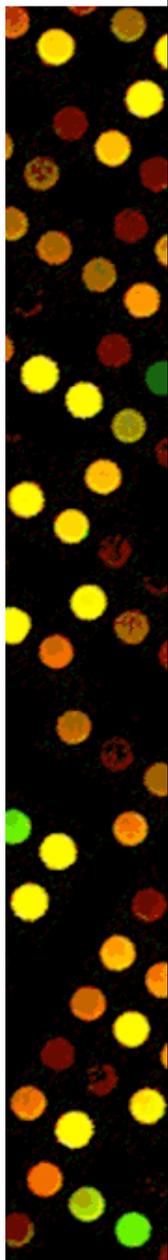
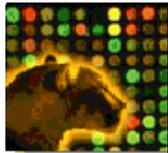


# CDT Deposit Options



-  View retrieval report
-  Download datafiles
-  GeneXplorer (web browser)
-  Java Treeview (web applet)
-  Cluster Images (from pipeline)
-  Clustered Spot Images (from pipeline)
-  Combo Images (from pipeline)
-  Edit deposit name, description, or access
-  Delete deposit





## GeneXplorer : correlation view

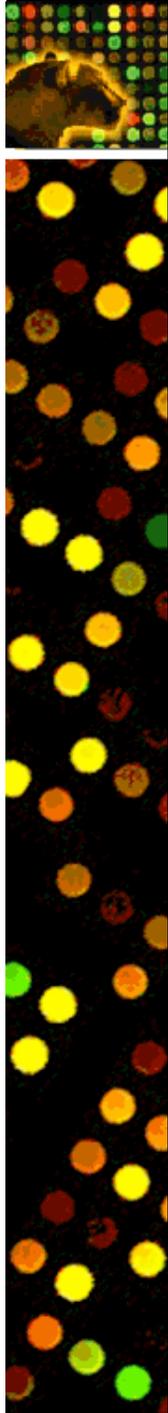
Clicking on a vector within the zoom view display a list of vectors having the best correlations.



# Repository Deposits : PCL

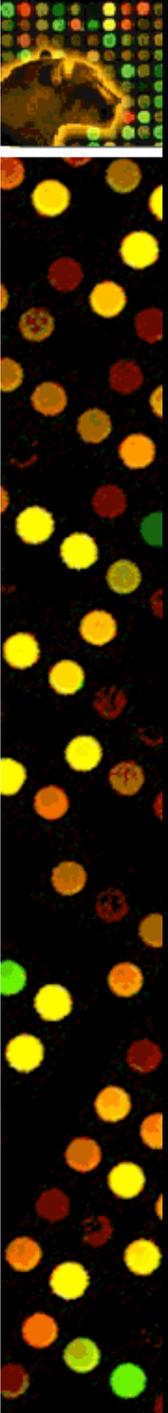
Name	Organism	Date	Type ↓	Genes	Expts.	Size	Options (help)
Ty red	<i>Saccharomyces cerevisiae</i>	06/06/07	PCL	41775	13	6783 kB	View  Filter  Synth
YZRNADNA both median centered	<i>Saccharomyces cerevisiae</i>	05/04/08	PCL	4348	8	1234 kB	View  Filter  Synth
YZRNADNA median centered arrays	<i>Saccharomyces cerevisiae</i>	05/04/08	PCL	4348	8	1199 kB	View  Filter  Synth
all except first 2 Sb processed and clustered	<i>Saccharomyces bayanus</i>	01/02/08	PCL_UPLOAD	4840	169	4239 kB	View  Filter  Synth
HO data from Abram	<i>Saccharomyces cerevisiae</i>	02/11/08	PCL_UPLOAD	7926	12	1089 kB	View  Filter  Synth
May1 RNADNA orthologs	<i>Saccharomyces cerevisiae</i>	05/01/08	PCL_UPLOAD	6277	56	3120 kB	View  Filter  Synth

- Deposit the results of an analysis
- Re-enter data retrieval “pipeline” (filter, cluster), potentially avoiding data retrieval procedures that can be both repetitive and tedious!
- Use Singular Value Decomposition (SVD) tools
- Average data by “synthetic genes”
- View retrieval report
- Assign access



# PCL Deposit Options

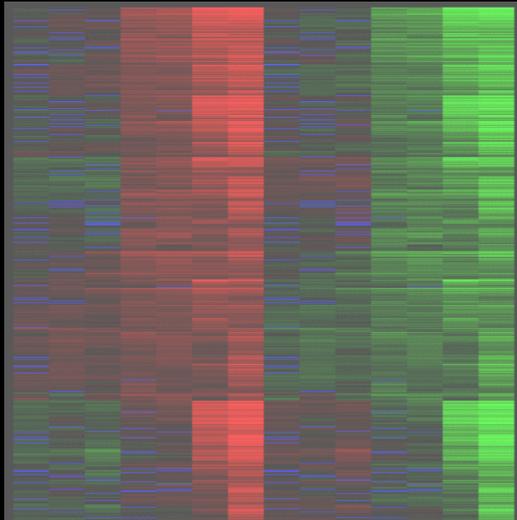
-  View retrieval report
-  Download datafiles
-  Filter by gene expression pattern (pipeline re-entry)
-  Cluster and visualize (pipeline re-entry)
-  Singular value decomposition (SVD) tools
-  Average data by “synthetic genes”
-  Impute missing data
-  Edit deposit name, description, or access
-  Delete deposit



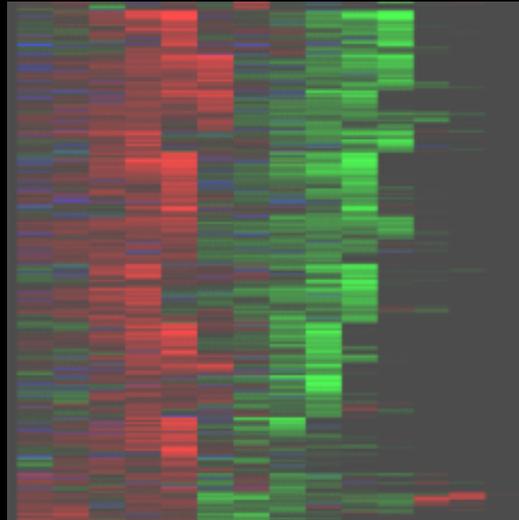
# KNNImpute: The Missing Values Problem

- Microarrays can have systematic or random missing values
- Some algorithms aren't robust to missing values
- Large literature on parameter estimation exists
- What's best to do for microarrays?

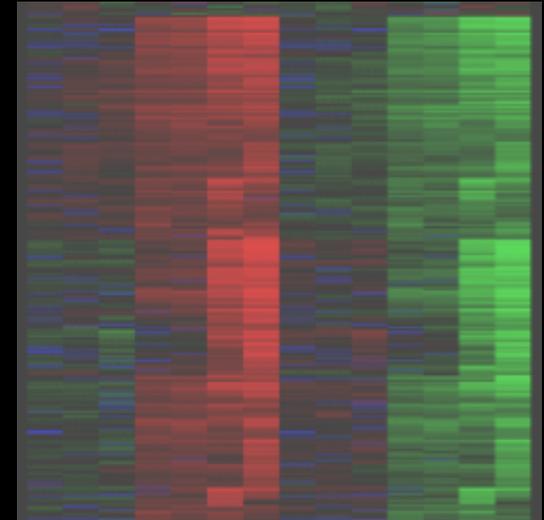
# Why Estimate Missing Values?



Complete data set



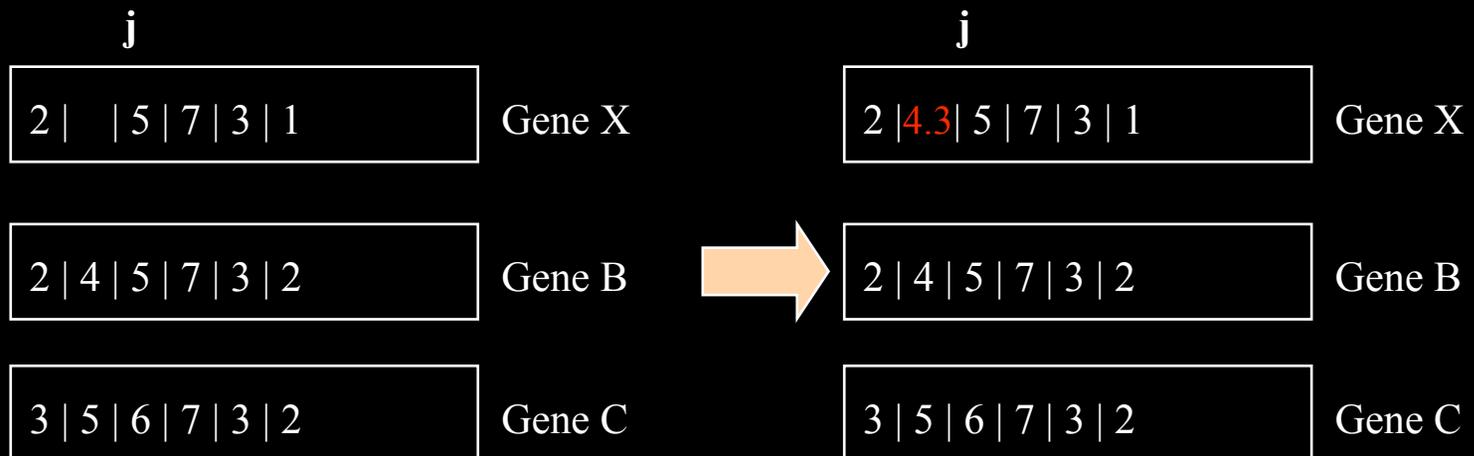
Data set with 30%  
entries missing  
(missing values appear  
black)



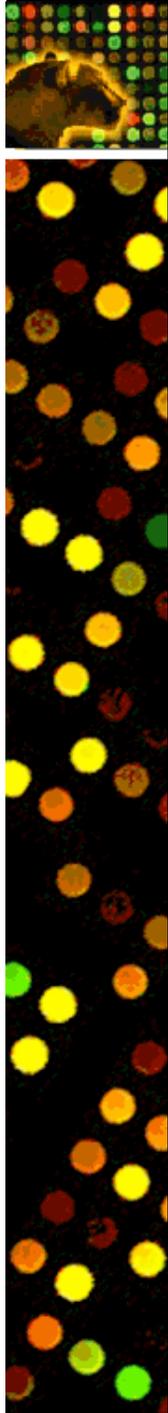
Data set with missing  
values estimated by  
KNNimpute algorithm

# KNNimpute Algorithm

- Idea: use genes with similar expression profiles to estimate missing values



Troyanskaya *et al.* Missing value estimation methods for DNA microarrays. *Bioinformatics* (2001) vol. 17 (6) pp. 520-5

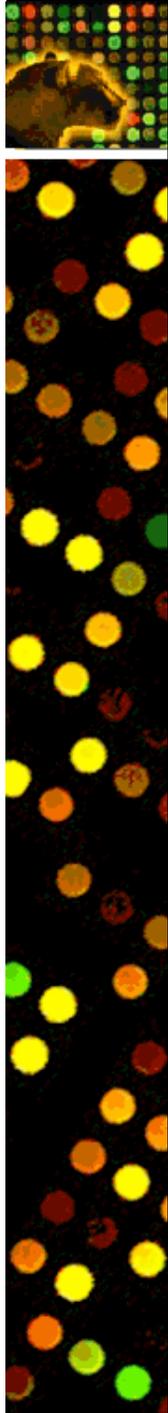


# KNNImpute implementation

## Background

Because some data analysis algorithms either require (i.e. SVD) or are sensitive to missing data, data imputation is sometimes necessary. The interface provided here initiates an implementation of K-nearest neighbors data imputation (KNNimpute), as described in [Troyanskaya et al. Bioinformatics 17 \(6\): 520. \(2001\)](#) , and places it within the job queue. The resulting imputed dataset will be deposited to your repository.

Impute Parameters	
Number of nearest neighbors :	<input type="text" value="15"/>
Distance metric :	<input type="text" value="euclidean"/> <small>note: only euclidean has been implemented</small>
Resulting Deposit Name :	<input type="text"/>
Resulting Deposit Description:	<input type="text" value="describe the new imputed pcl file"/>



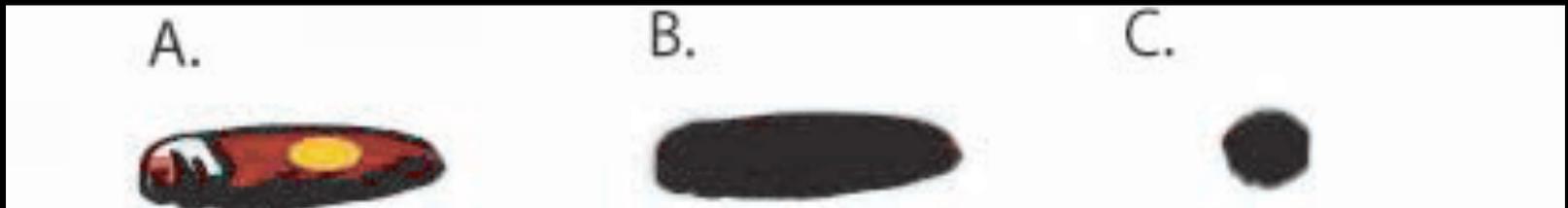
# Initial Dataset Characterization

- What is the structure of the data?
- What are the patterns within the noise?
- Reduce a dataset's dimensionality, but capturing its variance
  - Principle component analysis (**PCA**)
  - Singular Value Decomposition (**SVD**)

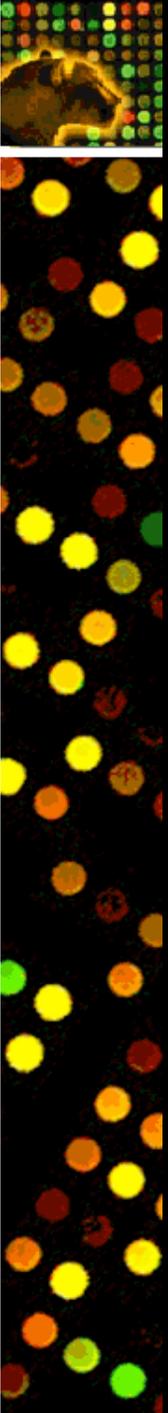
<http://public.lanl.gov/mewall/kluwer2002.html>

“Singular value decomposition and principal component analysis” in *A Practical Approach to Microarray Data Analysis* (D.P. Berrar, W. Dubitzky, M. Granzow, eds.) Kluwer: Norwell, MA, 2003. pp. 91-109. LANL LA-UR-02-4001

# Reducing dimensionality



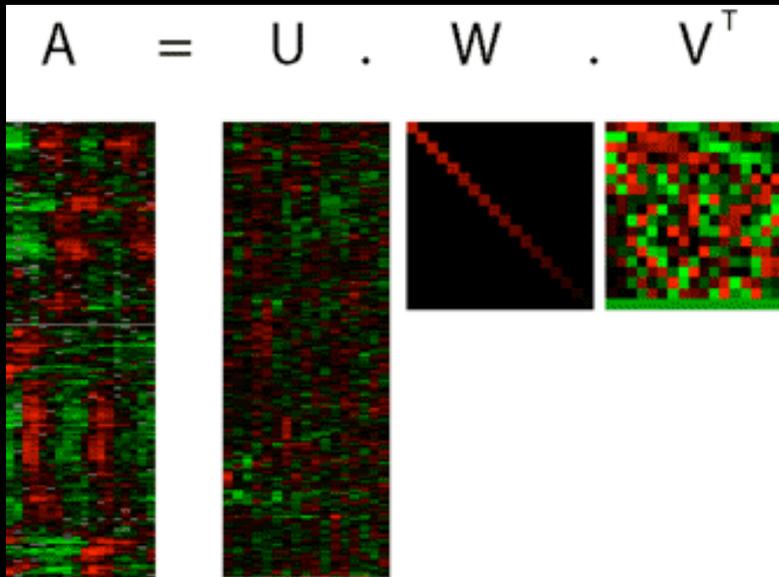
- Let's imagine we have a three-dimensional cigar, as shown in **A**
- We can represent this in one dimension, by looking at its lengthwise shadow (**B**)
- Looking at its cross-wise shadow (**C**), we get an orthogonal view of the cigar that tells us more about the three-dimensional object than **B** alone.



# SVD: Singular Value Decomposition

- SVD determines a set of patterns that describe the greatest amount of variance in a dataset
- SVD determines unique orthogonal (or uncorrelated) gene and corresponding array expression patterns ("eigengenes" and "eigenarrays," respectively)
- These patterns often are correlated with...
  - biological processes
  - technical artifacts

# SVD : Algorithm

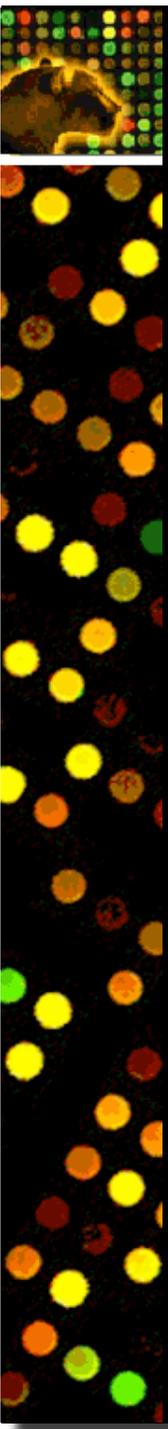


Singular value decomposition of a matrix of gene expression values (matrix **A**) results in three matrices, **U** (the eigenarrays matrix), **W** (the eigenexpression levels matrix) and **V<sup>T</sup>** (the eigengenes matrix). Matrix **V<sup>T</sup>** contains the eigengenes, matrix **W** contains the eigenvalues, matrix **U** contains the coefficients for the genes for each eigengene.

SVD is a linear transformation that uses a theorem from linear algebra to decompose a matrix to the product of three other matrices. For microarray data, the mathematical variables of SVD are the "eigengenes," "eigenarrays" and "eigenexpression levels." The eigengenes are gene patterns of expression. The eigenarrays are array patterns of expression. And, the eigenexpression levels indicate the dominance, and therefore also the significance, of these gene and array patterns of expression in the data. These matrices are shown. For more detail please see:

<http://puma.princeton.edu/help/svd.shtml>

Alter *et al.* Singular value decomposition for genome-wide expression data processing and modeling. Proc Natl Acad Sci USA (2000) vol. 97 (18) pp. 10101-6



# SVD: Missing Data Estimation

## Estimation of Missing Data

Some algorithms (such as SVD) cannot be performed on datasets with missing data.

Unless you replace the missing data, genes with missing data will be discarded from your dataset.

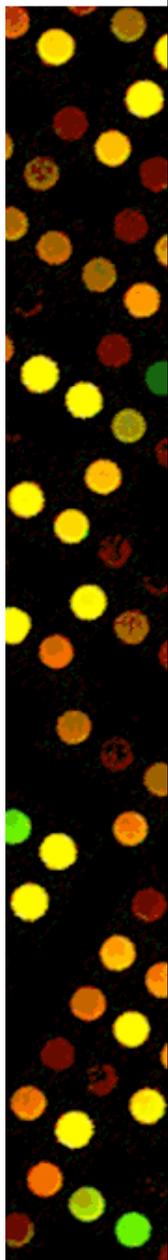
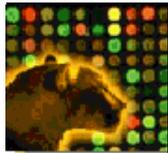
Genes in dataset : **3719**

Genes with missing data : **612**

Replace missing data using row averaging for genes with at least

% data present.

- Some algorithms (such as SVD) cannot operate with missing data
- You can use this simple method or you can use KNNImpute to estimate missing data



# SVD Display

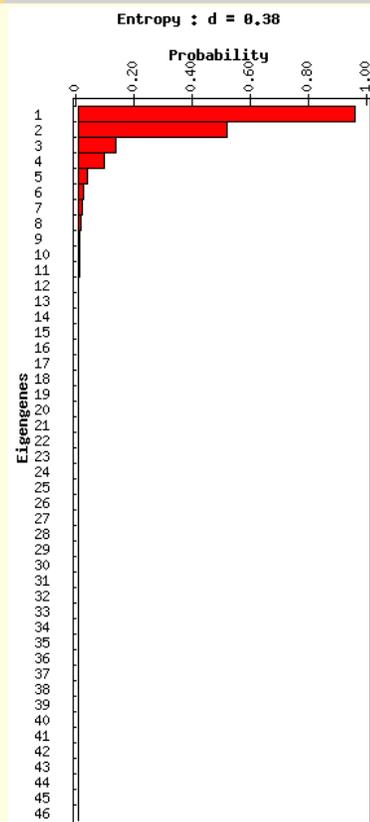
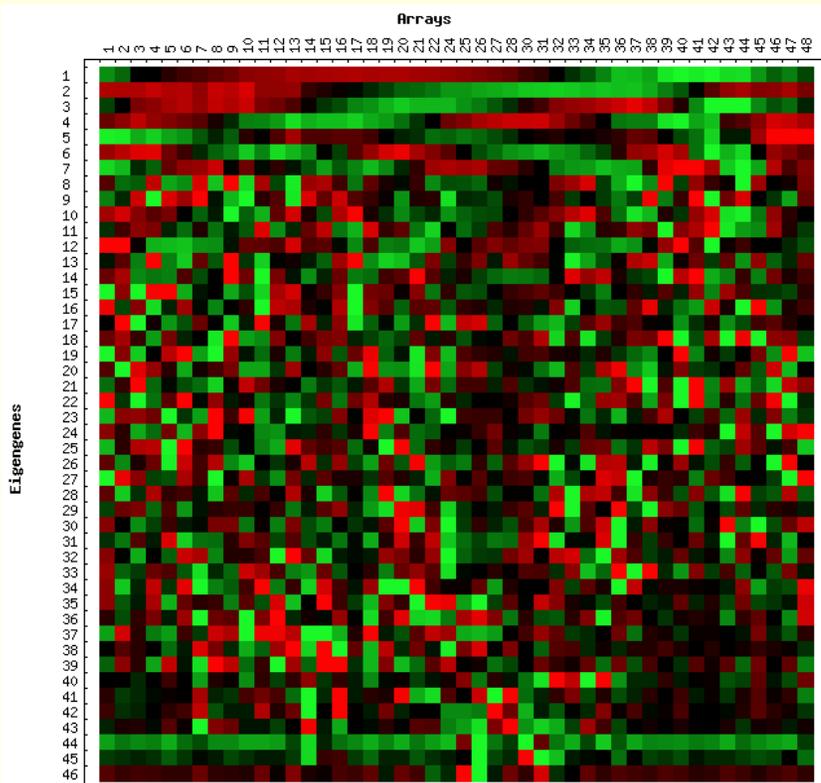
## Eigengenes Raster Display

Click on a single eigengene to see projections of all genes within the eigengene

[Download vector file](#)

Select Eigengene (s)  
To Remove or To Plot

Eigenexpression Probabilities  
(fraction of eigenexpression for each eigengene)

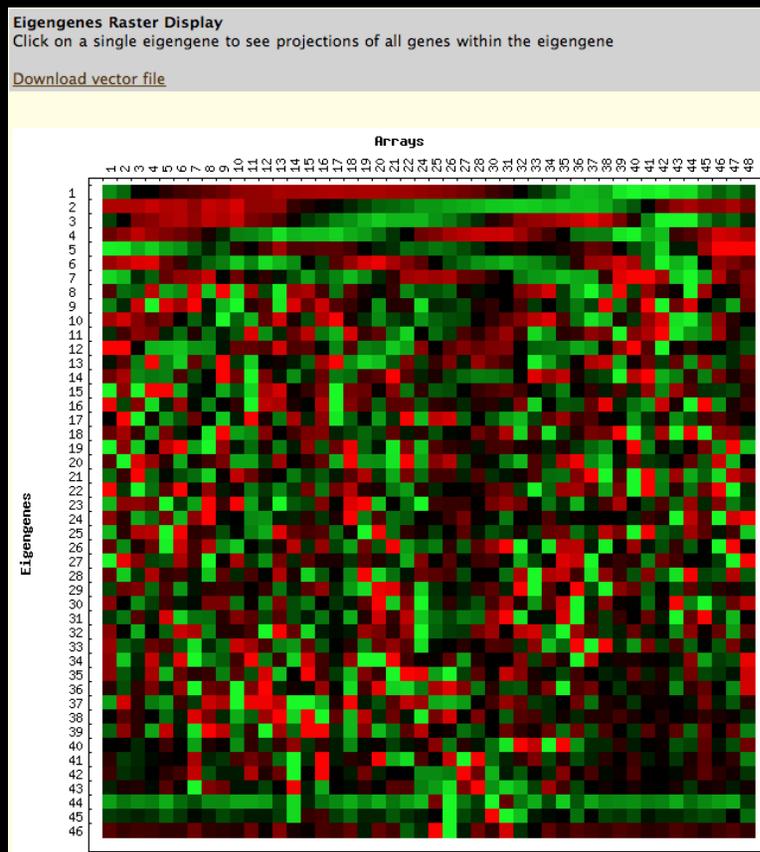


Remove selected eigengene (s)

Plot selected eigengene (s)

Reset

# SVD: Raster Display

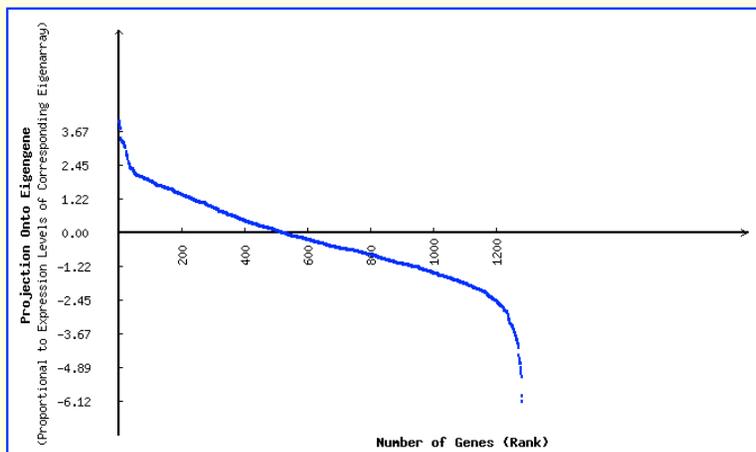


- Each row represents an “eigengene” -- an orthogonal representation of the genes in the dataset
- The topmost eigengene contributes the most to the data set

# SVD: View Projection

## Graph Display of Genes Sorted by Projections Onto The Selected Eigengene

[Download projection onto and correlation with eigengene](#)



### Selecting Subsets of Data:

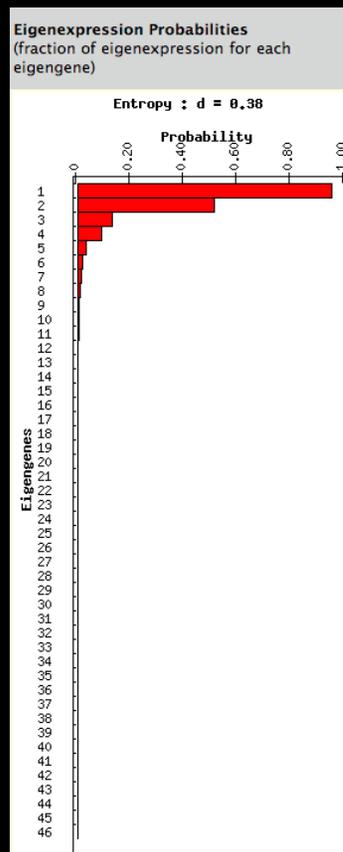
Select those genes whose projections fall within certain values. After selecting a subset of genes, you can download a .pcl file of the subset for further analysis.

Projection >

Projection <

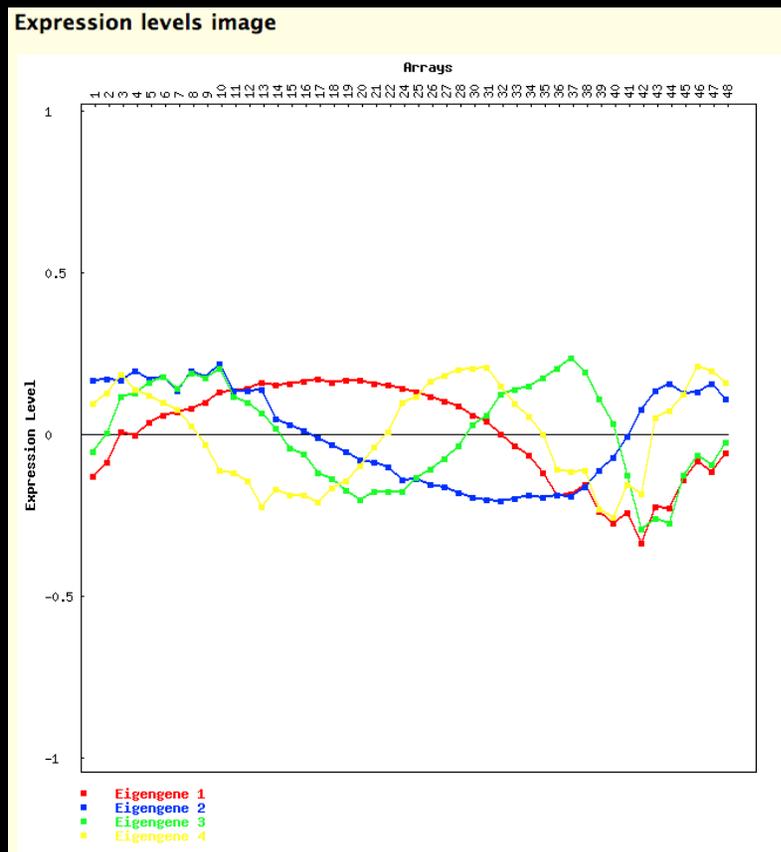
- Clicking on a row in the Raster Display brings you the Projection View
- You can select genes that have high and low contributions from an eigengene and download them in a PCL file
- In this way, you might use SVD to help classify subtypes

# SVD: Eigenexpression



- Each bar show the probability of expression of each eigengene
- You can compare the probabilities to see which eigengenes contribute more to the overall “view” of the data

# SVD: Plot selected eigengenes



- You can plot as many or as few eigengenes as you like
- This plot gives you an easy-to-understand view of the behavior of each eigengene

# Access to Synthetic Gene Tools

Name	Organism	Date	Type ↓	Genes	Expts.	Size	Options (help)
Ty red	<i>Saccharomyces cerevisiae</i>	06/06/07	PCL	41775	13	6783 kB	View
YZRNADNA both median centered	<i>Saccharomyces cerevisiae</i>	05/04/08	PCL	4348	8	1234 kB	View
YZRNADNA median centered arrays	<i>Saccharomyces cerevisiae</i>	05/04/08	PCL	4348	8	1199 kB	View
all except first 2 Sb processed and clustered	<i>Saccharomyces bayanus</i>	01/02/08	PCL_UPLOAD	4840	169	4239 kB	View
HO data from Abram	<i>Saccharomyces cerevisiae</i>	02/11/08	PCL_UPLOAD	7926	12	1089 kB	View
May1 RNADNA orthologs	<i>Saccharomyces cerevisiae</i>	05/01/08	PCL_UPLOAD	6277	56	3120 kB	View

- Put data in your repository.
- Click “Synth” icon.



# Synthetic Gene Options

## Synthetic Gene Options

Calculated synthetic gene vectors will be added to the .pcl file.

**PLEASE NOTE** that the curated synthetic gene lists are highly provisional and in flux!

- ▶ Choose any number of sets of curated synthetic gene lists.

- ▶ Choose any number of your own genelists.

- ▶ Choose an operation.

Average expression vectors by synthetic gene producing a pre-clustering (pcl) file of averaged vectors.

Choose how much original data to retain in the processed file.

- Retain all original data.
- Remove data used in creation of synthetic gene vectors.
- Remove all original data leaving only synthetic gene vectors.

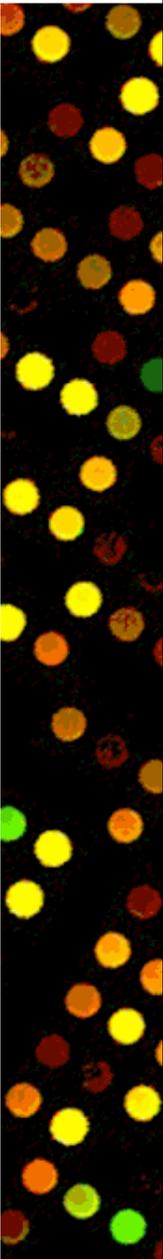
Add synthetic gene annotations to genes in the original pcl file.

Choose whether to retain the original annotations in addition to the synthetic gene annotations.

- Retain original annotations

- Select any number of public lists, and/or your own genelists.
- Choose handling of original data

# Other PCL tools



**Search**

- ▶ [Advanced Search](#)
- ▶ [Basic Search](#)
- ▶ [Gene Search](#)

**EnterData**

- ▶ [Experiments and Results](#)
- ▶ [Replace Proxy Image](#)
- ▶ [Patient and Clinical Data](#)
- ▶ [Procedural Information](#)
- ▶ [Protocol](#)

**List**

- ▶ [Users](#)
- ▶ [User Group](#)
- ▶ [Organism List](#)
- ▶ [Sequence Types](#)
- ▶ [SUD Retrieval](#)
- ▶ [Projects](#)
- ▶ [Plate Storage Locations](#)
- ▶ [Printers](#)
- ▶ [Print\\_config](#)
- ▶ [Categories](#)
- ▶ [SubCategories](#)
- ▶ [Normalizations](#)
- ▶ [Plate List](#)
- ▶ [Print List](#)
- ▶ [Experiment List](#)
- ▶ [Procedures](#)
- ▶ [Protocol](#)
- ▶ [Parameters](#)
- ▶ [Experiment Types](#)
- ▶ [Experiment Set Types](#)
- ▶ [View Clinical Data](#)
- ▶ [Repository List](#)
- ▶ [Publication List](#)
- ▶ [Meta Data List](#)
- ▶ [Ontology Browser](#)

**Edit/Delete Data**

- ▶ [Edit Experiment Details](#)
- ▶ [Batch Renormalize and/or Background Correction](#)
- ▶ [Add Experiment Access by Batch](#)

**Tools**

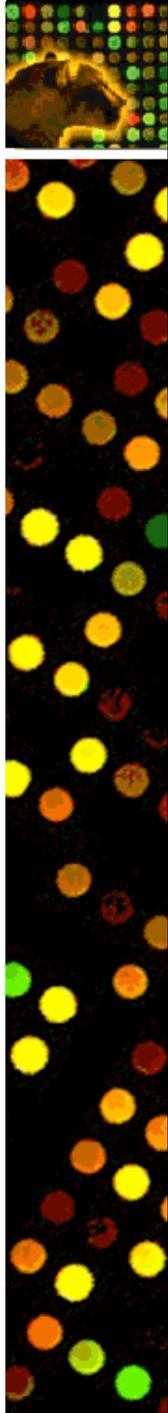
- ▶ [Convert an Arraylist to an Experiment Set](#)
- ▶ [Retrieve Experiment Meta Data for an Arraylist](#)
- ▶ [Download Raw Data for an Arraylist](#)
- ▶ [Assess array quality for an Arraylist](#)
- ▶ [Genomically align a list of data](#)
- ▶ [Validate a list of Platesamples](#)
  - [Renormalize Experiments for Genelist](#)
- ▶ [PCL file utilities](#)
  - [Merge PCL files](#)
  - [Convert NCBI GEO accessions to PCL files](#)
  - [Extract gene annotations from PCL/CDT file](#)
  - [Convert PCL files to EGR](#)
- ▶ [Lab Statistics](#)
- ▶ [Ontology Term Finder](#)

**User Info**

- ▶ [Change your Password](#)
- ▶ [Edit your Contact Information](#)
- ▶ [Manage your Form Defaults](#)

## Tools

- ▶ [Convert an Arraylist to an Experiment Set](#)
- ▶ [Retrieve Experiment Meta Data for an Arraylist](#)
- ▶ [Download Raw Data for an Arraylist](#)
- ▶ [Assess array quality for an Arraylist](#)
- ▶ [Genomically align a list of data](#)
- ▶ [Validate a list of Platesamples](#)
  - [Renormalize Experiments for Genelist](#)
- ▶ [PCL file utilities](#)
  - [Merge PCL files](#)
  - [Convert NCBI GEO accessions to PCL files](#)
  - [Extract gene annotations from PCL/CDT file](#)
  - [Convert PCL files to EGR](#)
- ▶ [Lab Statistics](#)
- ▶ [Ontology Term Finder](#)



# Merge PCL Files

- Can be used to combine 2 pcl files from different sources into a single pcl file.
- Can be used to averaged replicate assays (columns)
- Reporters that belong to the same gene can be combined into single row, based on a translation file (i.e. “synthetic genes”, discussed earlier).

# Merge PCL Files

▶ Select an organism to associate with the new file  
This is required to cluster the file, or deposit it in your repository.

▶ You may upload one or two PCL files to merge.  
**File 1:**  no file selected  
**File 2:**  no file selected

▶ You may instead (or additionally) select PCL files from your repository.  
  
33MIN averaged  
Agilent identifiers  
C. elegans developmental and sex-regulated gene expression profiles  
GSE1639 conversion

**Repository file(s):**

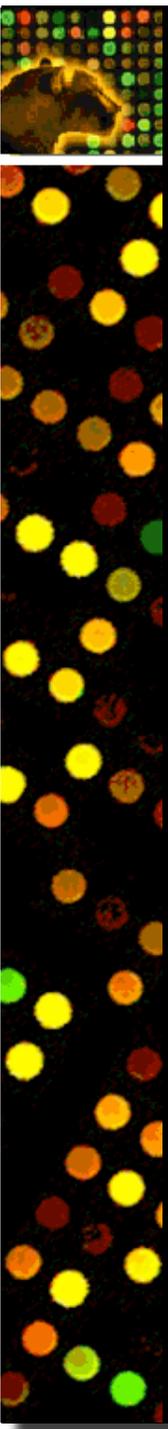
▶ Optionally provide a translation file (tab-delimited text), to map gene identifiers in the current data to identifiers in the other file(s):  
**Translation file:**  no file selected

- First column: desired final identifiers (may appear multiple times in the translation file).
- Second column: desired final annotation.
- Third and subsequent columns: identifiers that may be found in the current data and/or the other .pcl files, to be translated to the identifier in the first column.
- Identifiers not found in the translation file will be preserved unchanged.

▶ Choose whether to merge columns (arrays) with the same identifier in different files, or keep the distinct.  
 **Merge columns** with the same identifier in the various files.

▶ Choose which average (mean or median) to use for merged values.  
**Averaging method:**  Note that all values for a single gene post-translation, in a single column post-merge, will be averaged.

- Translation file:
  - Tab-delimited text file
  - First column: desired final identifier
  - Second column: desired final annotation
  - Third and subsequent columns: identifiers in the pcl files that should be collapsed to the identifier in the first column.
  - Data for identifiers not included in the translation file will not be collapsed
- Column averaging : replicate assays in the pcl files can be (mean/median)



# GEO accession to PCL

**Specify GEO Series accessions for conversion to a PCL file.**

**Step 1. Specify one or more GEO Series accessions (up to a maximum of 25).**

Enter individual GEO Series accessions, one per line, in the format GSExxx (for example, GSE715):

And/or enter a range of GEO Series accessions (for example, GSE2010 to GSE2015)

From:  To:

Please note that there can be gaps within a range of accessions. For example, within the range GSE2010 to GSE2015, only GSE2010, GSE2014 and GSE2015 exist in GEO. The log file for your job will indicate which GEO files went in to your PCL file.

**Step 2 (optional). Specify the name of a data column to join by.**

Column Name:

The column name must exist uniquely within each platform. If you do not specify a column, one will be selected, and any other possibilities will be noted in the log file.

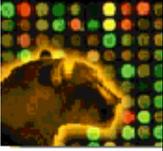
**Step 3. Annotate the file for your repository.**

Choose an organism:

Enter a brief name:

Description:  
And provide a useful description.

- Generate a PCL (pre-clustering) file from one or more NCBI GEO Series accessions
- Resulting PCL gets deposited to repository



# PCL to EGR converter

## PCL to EGR Converter

Convert a PCL file into an EGR which can be read by the Affymetrix Integrated Genome Browser (IGB). Your PCL file must contain the chromosome as well as the start and end coordinates.

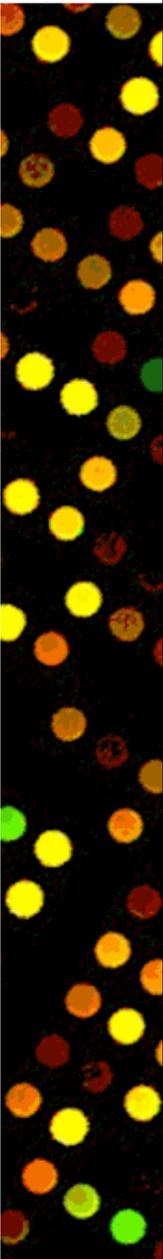
### Step 1: Specify a PCL file:

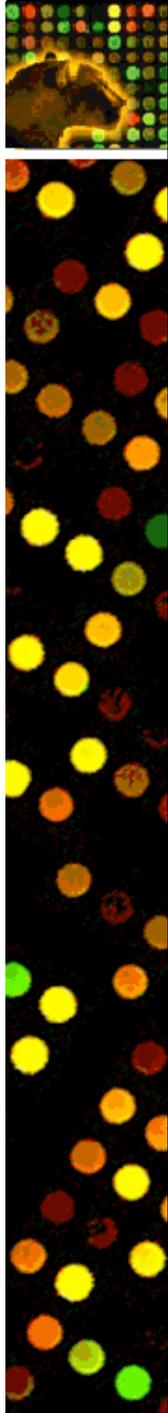
► From your repository:

► Or upload a PCL file:  no file selected

*(File will not be placed in your repository; use [repository upload](#) instead.)*

Step 2: Specify the Genome version:

- Currently only for PCL files where the chromosome, start, and end coordinates were retrieved
  - Expression GRaph (EGR) files contain data for scored intervals.
  - EGR files can be visualized using the Affymetrix Integrated Genome Browser (IGB)
- 

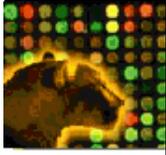


# Some References

Alter *et al.* Singular value decomposition for genome-wide expression data processing and modeling. Proc Natl Acad Sci USA (2000) vol. 97 (18) pp. 10101-6

Alter *et al.* Generalized singular value decomposition for comparative analysis of genome-scale expression data sets of two different organisms. Proc Natl Acad Sci USA (2003) vol. 100 (6) pp. 3351-6

Troyanskaya *et al.* Missing value estimation methods for DNA microarrays. Bioinformatics (2001) vol. 17 (6) pp. 520-5



# Acknowledgements

- SMD
  - <http://smd.stanford.edu>
  - Catherine Ball, Gavin Sherlock, and staff
- PUMADB, Lewis-Sigler Institute
  - <http://puma.princeton.edu>
  - Fan Kang, Laurie Kramer, Mark Schroeder, John Wiggins, Linda McMahan
- Affymetrix

