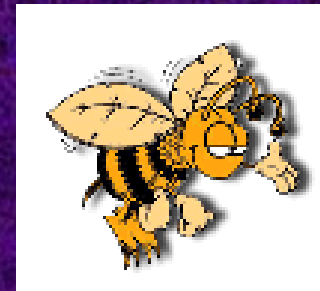


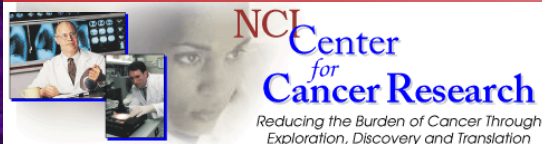
# Introduction to mAdb

Esther Asaki, Kathleen Meyer, John Powell

- I. Introducing the mAdb system
- II. Managing projects in mAdb
- III. Putting your data in mAdb
- IV. Evaluating array quality
- V. Getting started with analysis
- VI. Managing your data



October 1, 2008



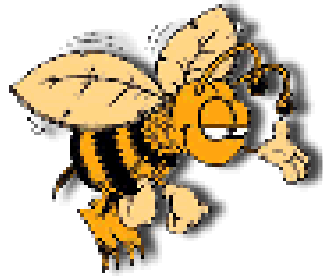
# Logging into the Training Server

- Point your browser at <http://madb-training.cit.nih.gov> – for use in class only!
- Your username is on the card on your desk
- Today's Password is on whiteboard near door
- Don't request a mAdb account on the training server!! – request at [madb.nci.nih.gov](http://madb.nci.nih.gov) or [madb.niaid.nih.gov](http://madb.niaid.nih.gov)
- Do not maximize your browser; leave room to see and click on other windows

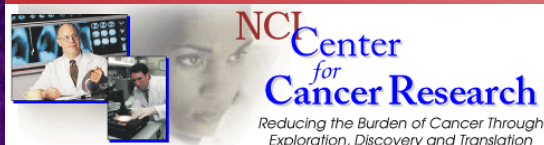
# I. Introducing the mAdb system

# mAdb Bioinformatics Project

## Goals:



- Provide an integrated set of web-based analysis tools and a data management system for storing and analyzing cDNA/oligo/Affy Gene Expression data using open systems design.
- Support spotted arrays produced by the NCI, NIAID and FDA Microarray Centers as well as some commercially produced arrays.
- Support various image analysis programs (some upon request)
  - Axon GenePix
  - Perkin-Elmer QuantArray
  - Arraysuite II / IP Lab
  - Agilent Feature Extraction
  - Affymetrix MAS5/GCOS - (mouse, human, rat)
  - Illumina Bead Studio
  - NimbleGen



# mAdb Home Page URLs

<http://madb.nci.nih.gov>

<http://madb.niaid.nih.gov>

The screenshot shows the mAdB home page for the National Cancer Institute. The header includes the NCI logo and the text "Center for Cancer Research". Navigation links include "mAdB Home", "Analysis Gateway", "Upload Status", "Forums", "Reference Info", "Program Downloads", and "GeneCards". The page is dated Monday, 16-Dec-2002 15:14:10 EST. A welcome message states: "Welcome to the mAdB (aka Mad Bee) Home page. CIT/BIMAS is collaborating with NCI/CCR in the development of the BioInformatics to manage, access and analyze cDNA  $\mu$ Array data generated by the NCI/CCR  $\mu$ Array Center." A list of links is provided:

- [Gateway - Data Upload, Access and Analysis Tools](#) (Note: Must be a registered mAdB user - Login/Password required)
- [Array Ordering/Tracking WEB site](#) (Note: Different WEBSITE, requires it's own Login/Password)
- [mAdB Account Request](#) - Request a new user account.
- [mAdB Training via CIT](#)
  - [mAdB Basic Informatics Course](#) - Description and Sign up for 2 hour intro
  - [NEW mAdB Intermediate Informatics Course](#) - Description and Sign up for
  - [Statistical Analysis of Microarray Data](#) - Description and Sign up for advanced design with afternoon hands-on lab with BRB ArrayTools
- NCI/CDT/BRB's [Announcement](#) on BRB ArrayTools release.
- Retrieve/Compare - [Mouse sets](#) or [Human sets](#)
- [GAL](#) (Gene Array List) files or [Gene Lists](#) (GIPO/Comprehensive) (Including Affymetrix Hs U133 and Mm U74 sets)
- mAdB [Feature Report](#) by Feature, Accession, GID or Well ID (Including features from Affymetrix Hs U133 and Mm U74 sets)
- mAdB [Tools](#) for mining our local copy of NCBI's [UniGene](#) database
- [MedMmer](#), NCI/LMP/Genomics & Bioinformatics Group's Text-Mining tool.

Navigation links at the bottom: [mAdB Home](#) | [Analysis Gateway](#) | [Upload Status](#) | [Forums](#) | [Reference Info](#) | [Program Downloads](#) | [GeneCards](#)

The screenshot shows the mAdB home page for the National Institute of Allergy and Infectious Diseases (NIAID). The header includes the NIAID logo and the text "RESEARCH TECHNOLOGIES BRANCH". Navigation links include "mAdB Home Page", "mAdB Gateway", "Upload Status", "Training/Reference", "Program Downloads", and "GeneCards". The page is dated Wednesday, 21-Jul-2004 17:08:24 EDT. A welcome message states: "In collaboration with the Microarray Research Facility at NIAID and the Advanced Technology Center at NCI, the Bioinformatics and Molecular Analysis Section (BIMAS), NIH Center for Information Technology offers the mAdB microarray data analysis system." A list of links is provided:

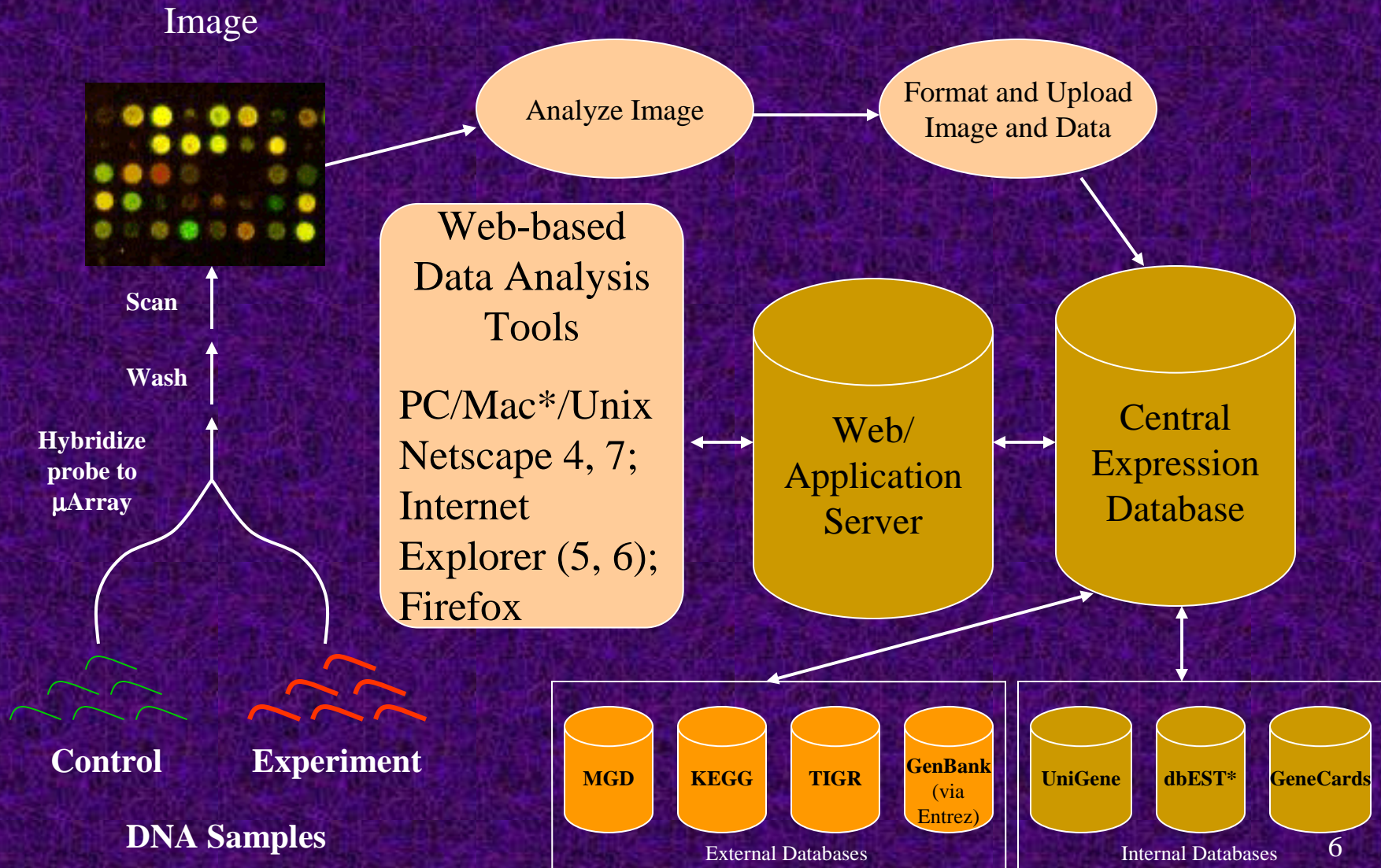
- [Register for a mAdB Account](#)
- [Start mAdB session \(requires mAdB account\)](#)
- [mAdB Training/Reference Information](#)
- [Gene Array List \("GAL"\) files for NIAID MRF Arrays](#)
- [Comprehensive Gene Lists](#)
- [Lookup mAdB Features](#)

Navigation links at the bottom: [mAdB Home Page](#) | [mAdB Gateway](#) | [Upload Status](#) | [Training/Reference](#) | [Program Downloads](#) | [GeneCards](#)

NIH Bioinformatics support provided by [BIMAS/CBEI/CIT](#). We can be contacted by [email](mailto:madb_support@bimas.cit.nih.gov).

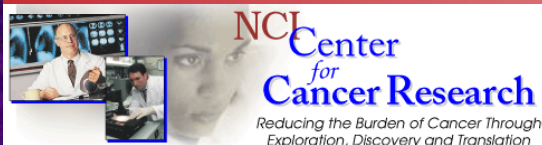
For support, please e-mail: [madb\\_support@bimas.cit.nih.gov](mailto:madb_support@bimas.cit.nih.gov)

# Architecture for $\mu$ Array Informatics



# mAdb Quick Facts

- Over 80,000 arrays uploaded since Feb. 2000
- Over 1,500 registered users (NIH and collaborators)
- Among the largest collections of microarray data in the world, although data sharing is determined by each investigator – no one has access to all the data
- MIAME capable format – available upon request
- Provide assistance in submitting your data into public repositories – GEO (NCBI), ArrayExpress (EBI)



# mAdb System Features

- Gene Discovery
  - Outlier detection
  - Scatter plots
  - Ad hoc keyword queries
  - Multiple array viewer
- Class Comparison
  - t-test; Wilcoxon; ANOVA; Kruskal-Wallis; SAM
- Class Prediction
  - PAM classifier
- Class Discovery (unsupervised)
  - Clustering – Hierarchical, K-means, SOMs
  - Multidimensional Scaling
  - Principal Components Analysis
- Pathway summary – GO, KEGG, BioCarta
- Boolean comparison of data

**Class #412 -  
Analyzing  
Microarray  
Data using  
the mAdb  
System**



# Live Demo

# Home Page Notes

- Account Requests link
- Analysis Gateway link
- Training signup/Reference documents link
- mAdb support e-mail link

# mAdb Training/Reference Page

## mAdb Training/Reference Information

Page Updated: Tuesday, 03-Oct-2006 16:55:47 EDT

- ◆ mAdb Training Classes via [CIT](#)
  - ◊ [Introduction to mAdb](#) - Description and Sign up for a 3 hour introductory class on using mAdb.
  - ◊ [Analyzing Microarray Data with the mAdb System](#) - Description and Sign up for a two half-day hands-on class using mAdb.
  - ◊ [Statistical Analysis of Microarray Data](#) - Description and Sign up for an overview of statistical issues and experimental design with a hands-on lab with BRB ArrayTools.
  
- ◆ mAdb Training Documentation
  - ◊ Introduction to mAdb (CIT class #411) Training Slides with Labs: [PowerPoint](#) or [PDF](#)  
Updated Thursday, 30-Mar-2006 11:53:45 EST
  
  - ◊ Analyzing Microarray Data with the mAdb System (CIT class #412) Training Slides
    - Lecture Slides Day 1: [PowerPoint](#) or [PDF](#)  
Updated Monday, 20-Mar-2006 17:30:03 EST
    - Lecture Slides Day 2: [PowerPoint](#) or [PDF](#)  
Updated Wednesday, 22-Mar-2006 11:39:57 EST
    - Hands-on Labs: [PowerPoint](#) or [PDF](#)  
Updated Tuesday, 21-Mar-2006 10:26:03 EST
  
- ◆ mAdb Reference Documentation
  - ◊ Increasing Upload Speed with Internet Explorer on the PC: [Word](#) or [PDF](#)  
Updated Wednesday, 14-May-2003 10:23:06 EDT
  
  -  ◊ Uploading Affymetrix Data to mAdb: [PDF](#)  
Updated Thursday, 27-May-2004 16:04:17 EDT
  
  - NOTE: You must request permission from [mAdb support](#) before uploading Affymetrix Data.
  
- ◆ Check [Java Version](#) running on your browser.

 **Must request Affy privileges be turned on for your account**

## II. Managing Projects in mAdb

# Setting up your mAdb area

- Login to mAdb Gateway page
  - change password if first-time user (case sensitive)
- Create project - logical organization for arrays
- Grant project access to others (if desired)
- Return to gateway and use Upload Array data link
- Select type of array for project
  - Spotted OR
  - Affymetrix (need to request permission via e-mail for first usage)
- Copy or move arrays to your project

# mAdb Gateway- link for User Profile Management

## *mAdb Gateway*

Access previously extracted data located in **ncidemo**'s:

[Permanent](#) area

or Choose one or more Projects, select a Tool and Continue

**Projects:**

- XX guest - Time Course Demo Set #1
- XX guest - Time Course Demo Set #2
- XX guest - Repeats and Reciprocal Retests Demo Set #3
- XX guest - Multiple Types Demo Set #4
- AU ncidemo - my project
- AU ncidemo - Oligo and cDNA

**Note:** Tools marked with "XX" only support selection of one project

**Tool:**

### Uploading Links

- ♦ [Upload Array](#) data
- ♦ [Status](#) of Uploads
- ♦ [Upload Identifier](#) lists
- ♦ [Manage](#) Identifier lists



### Management Tools

- ♦ [Create/Manage](#) Projects
- ♦ [Manage](#) User Profile



### Additional mAdb Tools/Resources:

[Feature Report](#) - provides additional details for restricted feature sets

[Extract](#) Dataset for a mAdb Print

[Access](#) Training/Public Datasets

[Access](#) Additional Public Datasets

# User Profile Management

## Managing User Profile

---

[Change](#) Your Password

[Update](#) Your User Profile

Profile for "ncidemo" last modified on Sep 03, 2004 at 15:03:08

**Title** Mr.

**First Name** DEMO

**Middle Initial**

**Last Name** NCI

**E-mail** jip@helix.nih.gov

**Position**

**Affiliation**

**NIH Address** 12A/2033 Bethesda, MD 20892

**Work Phone**

**Fax**

You have chosen to NOT Subscribe to the E-Newsletter

# mAdb Gateway- link for Project Creation & Management

## *mAdb Gateway*

Access previously extracted data located in **ncidemo**'s:

[Permanent](#) area

or Choose one or more Projects, select a Tool and Continue

**Projects:**

- XX guest - Time Course Demo Set #1
- XX guest - Time Course Demo Set #2
- XX guest - Repeats and Reciprocal Retests Demo Set #3
- XX guest - Multiple Types Demo Set #4
- AU ncidemo - my project
- AU ncidemo - Oligo and cDNA

**Note:** Tools marked with "XX" only support selection of one project

**Tool:**

### Uploading Links

- [Upload Array](#) data
- [Status](#) of Uploads
- [Upload Identifier](#) lists
- [Manage](#) Identifier lists



### Management Tools

- [Create/Manage](#) Projects
- [Manage](#) User Profile



### Additional mAdb Tools/Resources:

[Feature Report](#) - provides additional details for restricted feature sets

[Extract](#) Dataset for a mAdb Print

[Access](#) Training/Public Datasets

[Access](#) Additional Public Datasets



# Managing Projects

## Managing Projects

[Create](#) New Project

Filter for the list of projects displayed below.

Created by

Time Period

any user

within

any time

Update

List of Projects below.

Projects created by "any user" within "any time" for which "ncidemo" is an administrator.  
Projects are ordered first by the Creator and then by the Creation Date  
In the Access List, **Bold** indicates a user with administrative access

### Management Options

mAdb ID# 160 created by "ncidemo" on Jun 26, 2000 at 15:47:00 contains 10 Arrays

**Project Title:** my project

**Description:** Description by jip. Altered @1:00pm on 8/31/2004 and altered again by "easaki" on 9/1/2004

**Comments:** Comments by jip. Altered 8/31/2004

**Access List:** easaki, **jmgreene**, jpowell, **ncidemo**

### Management Options

mAdb ID# 1195 created by "ncidemo" on May 30, 2002 at 13:53:50 contains 10 Arrays

**Project Title:** Oligo and cDNA

**Description:** mixture of oligo and cDNA arrays

**Comments:** for IM class

**Access List:** easaki, **ncidemo**

# Create New Project

**Create New Project**

---

created by ncidemo

**Project Title:**

**Description:**

**Comments:**

- **A project is a logical grouping of your arrays**
- **Arrays can be copied/moved between projects**

# Project Management Options

## Project Management Options

mAdb ID# 160 created by "ncidemo" on Jun 26, 2000 at 15:47:00 contains 10 Arrays

**Project Title:** my project

**Description:** Description by jip. Altered @1:00pm on 8/31/2004 and altered again by "easaki" on 9/1/2004

**Comments:** Comments by jip. Altered 8/31/2004

**Access List:** **easaki, jmgreene, jpowell, ncidemo**

**Click** **Options available for this Project**



Can not be deleted - contains 10 Arrays

[Edit](#) To modify the Project Information (Title, Description, Comments)

[Add](#) To Add user(s) to the Access List for this Project

[Remove](#) To Remove user(s) from the Access List for this Project

[Privileges](#) To Grant or Revoke User(s) Administrative/Upload privileges for this Project

[Return](#) to Managing Projects

**Bold names on access list indicate administrative privileges for account**

# Project Access

## Add User(s)

mAdb ID# 160 created by "ncidemo" on Jun 26, 2000 at 15:47:00 contains 10 Arrays

**Project Title:** my project

**Description:** Description by jip. Altered @1:00pm on 8/31/2004 and altered again by "easaki" on 9/1/2004

**Comments:** Comments by jip. Altered 8/31/2004

**Access List:** easaki, jmgreene, jpowell, ncidemo

The List below includes **ALL mAdb users** not already having access to this project.

Add User(s)

Reset Form

Cancel

### Check to select User(s) to add to this project

▼ Last name, First name ( Login )

Abdool, Karen ( abdoolk )

Abul-Hassan, Khaled ( hassank )

Ajay, Dr ( ajay\_dr )

Akagi, Keiko ( akagik )

Aksamit, Robert ( aksamit )

Al-Timimi, Ali ( altimima )

Albert, Paul ( albertp )

Aleman, Claudina ( alemanc )

Alexander, H. Richard ( ralexander )

Alizadeh, Ash ( alizadeh )

Alkharouf, Nawal ( nalkhar )

Amornphimoltham, Panomwat ( pa79w )

Amundson, Sally ( amundson )

Anderson, Soni ( andersso )

Andersson, John ( jandersson )

Andreola, Fausto ( andreolf )

▼ Last name, First name ( Login )

Mazzanti, Chiara ( chiara )

McCarty, Tom ( tmccarty )

McConnell, Melanie ( melanie.mcconnell )

McDonald, Shannon ( slmcdonald )

McKee, Marian ( mmckee )

McNeil, Nicole ( mcneihn )

McNeill, Megan ( mmcneill )

McShane, Lisa ( mcshanel )

Medjahed, Djamel ( medjahed )

Mejido, Josef ( mejido )

Melani, Raffaella ( rmelani )

Meletiadis, Joseph ( meletiaj )

Melillo, Giovanni ( melillo )

Meltzer, Stephen ( umddemo )

Memon, Sarfraz ( memonsa )

Menard, Cynthia ( menardc )

Adding a user allows that mAdb account holder to view your arrays in a project and work with the data to create filtered datasets

# User Access Levels

## Change User(s) Privileges

mAdb ID# 160 created by "ncidemo" on Jun 26, 2000 at 15:47:00 contains 10 Arrays

**Project Title:** my project

**Description:** Description by jip. Altered @1:00pm on 8/31/2004 and altered again by "easaki" on 9/1/2004

**Comments:** Comments by jip. Altered 8/31/2004

Check/UnCheck as appropriate to select privileges

Admin Upload

<input type="checkbox"/>	<input type="checkbox"/>	Last name, First name ( Login )
<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<b>AU</b> Asaki, Esther ( easaki )
<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<b>AU</b> Greene, John ( jmgreene )
<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<b>AU</b> NCI, DEMO ( ncidemo )
<input type="checkbox"/>	<input type="checkbox"/>	-- Powell, John ( jpowell )

Record Changes

Reset Form

Cancel

Access levels allow user to:

- View data
- Upload Arrays
- Administer access to arrays and edit project/array descriptions

# III. Putting your data in mAdb

# Setting up your mAdb area

- Login to mAdb Gateway page
  - change password if first-time user (case sensitive)
- Create project - logical organization for arrays
- Grant project access to others (if desired)
- Return to gateway and use Upload Array data link
- Select **type** of array for project
  - Spotted OR
  - Affymetrix (need to request permission via e-mail for first usage)
- Copy or move arrays to your project

# mAdb Tool Gateway- link for uploading

## *mAdb Gateway*

Access previously extracted data located in **ncidemo**'s:

[Permanent](#) area

or Choose one or more Projects, select a Tool and Continue

**Projects:**

- XX guest - Time Course Demo Set #1
- XX guest - Time Course Demo Set #2
- XX guest - Repeats and Reciprocal Retests Demo Set #3
- XX guest - Multiple Types Demo Set #4
- AU ncidemo - my project
- AU ncidemo - Oligo and cDNA

**Note:** Tools marked with "X" only support selection of one project

**Tool:**

### Uploading Links

- ◆ [Upload Array](#) data
- ◆ [Status](#) of Uploads
- ◆ [Upload Identifier](#) lists
- ◆ [Manage](#) Identifier lists



### Management Tools

- ◆ [Create/Manage](#) Projects
- ◆ [Manage](#) User Profile



### Additional mAdb Tools/Resources:

[Feature Report](#) - provides additional details for restricted feature sets

[Extract](#) Dataset for a mAdb Print

[Access](#) Training/Public Datasets

[Access](#) Additional Public Datasets




# Select Array Type

## Data Uploading

Select the type of Array Analysis Software (Affymetrix MAS5/GCOS, Agilent Feature Extraction, GenePix Pro or ArraySuite II, Illumina BeadStudio Gene Expression, NimbleGen CGH Software) used and a project to upload your Array data into, then press the "Continue" button.

To create a new project, use this [link](#).

Array Analyzed with



- Affymetrix MAS5/GCOS
- Agilent Feature Extraction
- GenePix Pro or ArraySuite II
- Illumina BeadStudio Gene Expression
- NimbleGen CGH Software

Project

Continue

# Spotted Array Data Upload

- Fill in experimental info for each array
  - Pick Print Set
  - Select image file of array
  - Select data file for array
- Submit and confirm upload
- Check upload status page to display progress
- Close browser when finished (for security)

# mAdb GAL files

## Current (2002) NCI Production Gene Array List Files (GAL files) ( blocks x columns x rows)

- **NEW** [Earlier NCI production printings](#)
- [Custom printings](#)
- [NIAID printings](#)
- **NEW** [FDA printings](#)
- [Mini-lymphochip GAL files](#) (restricted to registered users)

Human Array Sets			
GAL File	Array Sets		
<a href="#">Hs-UniGEM2-v2px-32Bx18Cx18R.gal</a> Generated Tuesday, 21-May-2002 09:21:39 EDT Note: Also use for 2.1px, 2.3px, 2.4px, 2.5px, 2.6px, 5.0px See below for special 3.5px gal file <b>NEW</b> See below for special 4.0px gal file <b>NEW</b> See below for special 4.1px gal file <b>NEW</b> See below for special 4.2px gal file	Hs-UniGEM2-v2.4p1	Hs-UniGEM2-v2.4p2	Hs-UniGEM2-v2.4p3
	Hs-UniGEM2-v2.4p4	Hs-UniGEM2-v2.4p5	Hs-UniGEM2-v2.4p8
	Hs-UniGEM2-v2.4p9		
	Hs-UniGEM2-v2.5p3	Hs-UniGEM2-v2.5p4	Hs-UniGEM2-v2.5p5
	Hs-UniGEM2-v2.5p6	Hs-UniGEM2-v2.5p7	Hs-UniGEM2-v2.5p11
	Hs-UniGEM2-v2.6p2	Hs-UniGEM2-v2.6p3	Hs-UniGEM2-v2.6p6
	Hs-UniGEM2-v2.6p7	Hs-UniGEM2-v2.6p8	Hs-UniGEM2-v2.6p9
	Hs-UniGEM2-v2.6p10		
	Hs-UniGEM2-v5.0p1	Hs-UniGEM2-v5.0p2	Hs-UniGEM2-v5.0p3
	Hs-UniGEM2-v5.0p4	Hs-UniGEM2-v5.0p5	Hs-UniGEM2-v5.0p6
	Hs-UniGEM2-v5.0p7	Hs-UniGEM2-v5.0p8	Hs-UniGEM2-v5.0p9
	Hs-UniGEM2-v5.0p10	Hs-UniGEM2-v5.0p12	Hs-UniGEM2-v5.0p13
	Hs-UniGEM2-v5.0p14	Hs-UniGEM2-v5.0p15	Hs-UniGEM2-v5.0p16
	Hs-UniGEM2-v5.0p17	Hs-UniGEM2-v5.0p18	
<a href="#">Hs-UniGEM2-v3.5px-32Bx19x17R.gal</a> Generated Tuesday, 21-May-2002 09:33:10 EDT	Hs-UniGEM2-v3.5p1	Hs-UniGEM2-v3.5p2	
<a href="#">Hs-UniGEM2-4.0px-32Bx18Cx18R.gal</a> Generated Monday, 25-Nov-2002 15:03:35 EST	Hs-UniGEM2-v4.0p2	Hs-UniGEM2-v4.0p4	Hs-UniGEM2-v4.0p5
	Hs-UniGEM2-v4.0p6	Hs-UniGEM2-v4.0p7	Hs-UniGEM2-v4.0p8
	Hs-UniGEM2-v4.0p9	Hs-UniGEM2-v4.0p10	Hs-UniGEM2-v4.0p11
<a href="#">Hs-UniGEM2-4.1px-32Bx18Cx18R.gal</a> Generated Monday, 25-Nov-2002 15:34:59 EST	Hs-UniGEM2-v4.1p1		

• Shows the actual GAL (Gene Array list) files – link block, row, column to what DNA is spotted there

• One printset layout is usually used for many lots of slides

• Please e-mail mAdb support if you cannot find your GAL file listed

# Affymetrix Data Upload

- Select:
  - Data File (Metrics - .txt file)
  - CEL file
- Fill in Experiment data
- Submit and confirm upload
- Check upload status page to display progress
- Close browser when finished (for security)

# Uploading Spotted Arrays

Upload to Project #3751: my test project

Use this portion of the Form to Control the Print Choices below.

Organism: Human  from Facility/Vendor: NCI  printed Time Period: in the past 180 days

Print Choices below.

Don't see a print in the drop down list?

NCI Human Print: Hs-OperonV3.0-v1p23-121905  since Aug 04, 2005 (past 180 days).  
Try changing the Print Choice options above and click Update.

Array Name: Hs-OperonV3-45 Suggested form: HsOC2p13-45

Short Description: 4 hours

Long Description:  
(Optional)

Channel A (generally Cy3 tagged)

Sample: control

Sample Label: Cy3

Channel B (generally Cy5 tagged)

treated

Cy5

Composite Image & Arraysuite Sample Intensities or GenePix GPR Files

Image File: myImage.jpg

Data File: myData.gpr

# Confirming Upload

## NCI/NIH *mAdb* Data Loading Gateway

---

### Upload Confirmation:

Details from a preliminary inspection of the Intensity and Image files are provided below.  
You may Confirm or Cancel the uploading process.

### Data File:

C:\Documents and Settings\greenej1.NIH\Desktop\DataFile.txt

### Image File:

C:\Documents and Settings\greenej1.NIH\Desktop\ImageFile.img

Data file appears to be: Axon Text Format (GenePix Pro 3/4 Results)

Number of Data Values appears to be : 8837

Image Format: JPEG

[Return to Data Loading Page](#)


[Return to MicroArray Home Page](#)

[mAdb Home](#) | [Analysis Gateway](#) | [Upload Status](#)  
[Forums](#) | [Reference Info](#) | [Program Downloads](#) | [GeneCards](#)

You should check that the image and file type appear correct and that the file line count is roughly equal to the number of spots on the array

# Adding Affy Arrays

## Upload MAS5 Analysis Data to:my project

Note the  marks the link which lead to detailed help on required Affymetrix file format

### Affymetrix Files for Upload

<b>Data File:</b>	<input type="text"/>	<input type="button" value="Browse..."/>
<b>Cel File:</b>	<input type="text"/>	<input type="button" value="Browse..."/>

- Browse to Metrics (\*.txt) file for the Data File box
- Browse to the corresponding .CEL file in second box

# Adding Affy Arrays

## Confirm Affymetrix Genechip Data

### Experiment Information

You have uploaded Absolute Analysis data for a Human Genome Array U95A genechip.

The Data have not been scaled in your analysis.

Please check/complete the information on this page. Click the Confirm button to complete the upload process or use the Cancel button to abort and start again.

**Uploaded Data File:** C:\GeneChip\TESTDATA\Gene Logic Spike\92453hgu95a11\_test.txt

**Uploaded CEL File:** C:\GeneChip\TESTDATA\Gene Logic Spike\92454hgu95a11.cel

Fields labeled with \*\* are mandatoray.

**Array Print Set:** U95A

**Array Name: \*\***

**Sample Type:**

**Sample Description:**

**Comments:**

Confirm

Cancel



# Affymetrix – CHP file

	Analysis Name	Probe Set Name	Stat Pairs	Stat Pairs Used	Signal	Detection	Detection p-value	Stat Common P
1	92453hgu95a11_test	AFFX-MurIL2_at	20	20	0.7	A	0.949771	
2	92453hgu95a11_test	AFFX-MurIL10_at	20	20	0.3	A	0.989683	
3	92453hgu95a11_test	AFFX-MurIL4_at	20	20	0.3	A	0.997133	
4	92453hgu95a11_test	AFFX-MurFAS_at	20	20	2.7	A	0.529760	
5	92453hgu95a11_test	AFFX-BioB-5_at	20	20	399.3	P	0.000297	
6	92453hgu95a11_test	AFFX-BioB-M_at	20	20	247.2	P	0.000081	
7	92453hgu95a11_test	AFFX-BioB-3_at						
8	92453hgu95a11_test	AFFX-BioC-5_at						
9	92453hgu95a11_test	AFFX-BioC-3_at						
10	92453hgu95a11_test	AFFX-BioDn-5_at						
11	92453hgu95a11_test	AFFX-BioDn-3_at						
12	92453hgu95a11_test	AFFX-CreX-5_at						
13	92453hgu95a11_test	AFFX-CreX-3_at						
14	92453hgu95a11_test	AFFX-BioB-5_st						
15	92453hgu95a11_test	AFFX-BioB-M_st						
16	92453hgu95a11_test	AFFX-BioB-3_st						
17	92453hgu95a11_test	AFFX-BioC-5_st						
18	92453hgu95a11_test	AFFX-BioC-3_st						
19	92453hgu95a11_test	AFFX-BioDn-5_st						
20	92453hgu95a11_test	AFFX-BioDn-3_st						
21	92453hgu95a11_test	AFFX-CreX-5_st						
22	92453hgu95a11_test	AFFX-CreX-3_st						
23	92453hgu95a11_test	AFFX-hum_alu_at						
24	92453hgu95a11_test	AFFX-DapX-5_at						
25	92453hgu95a11_test	AFFX-DapX-M_at						
26	92453hgu95a11_test	AFFX-DapX-3_at						
27	92453hgu95a11_test	AFFX-LysX-5_at						
28	92453hgu95a11_test	AFFX-LysX-M_at						
29	92453hgu95a11_test	AFFX-LysX-3_at						
30	92453hgu95a11_test	AFFX-PheX-5_at						
31	92453hgu95a11_test	AFFX-PheX-M_at						

Analysis Options

Scatter Graph | Series Graph | Intensity Graph | Pivot | Metrics

Preference

- Save Analysis Info
- Save All Metric Results
- Save all analyses to one file.
- Save each analysis to a separate file.

Defaults

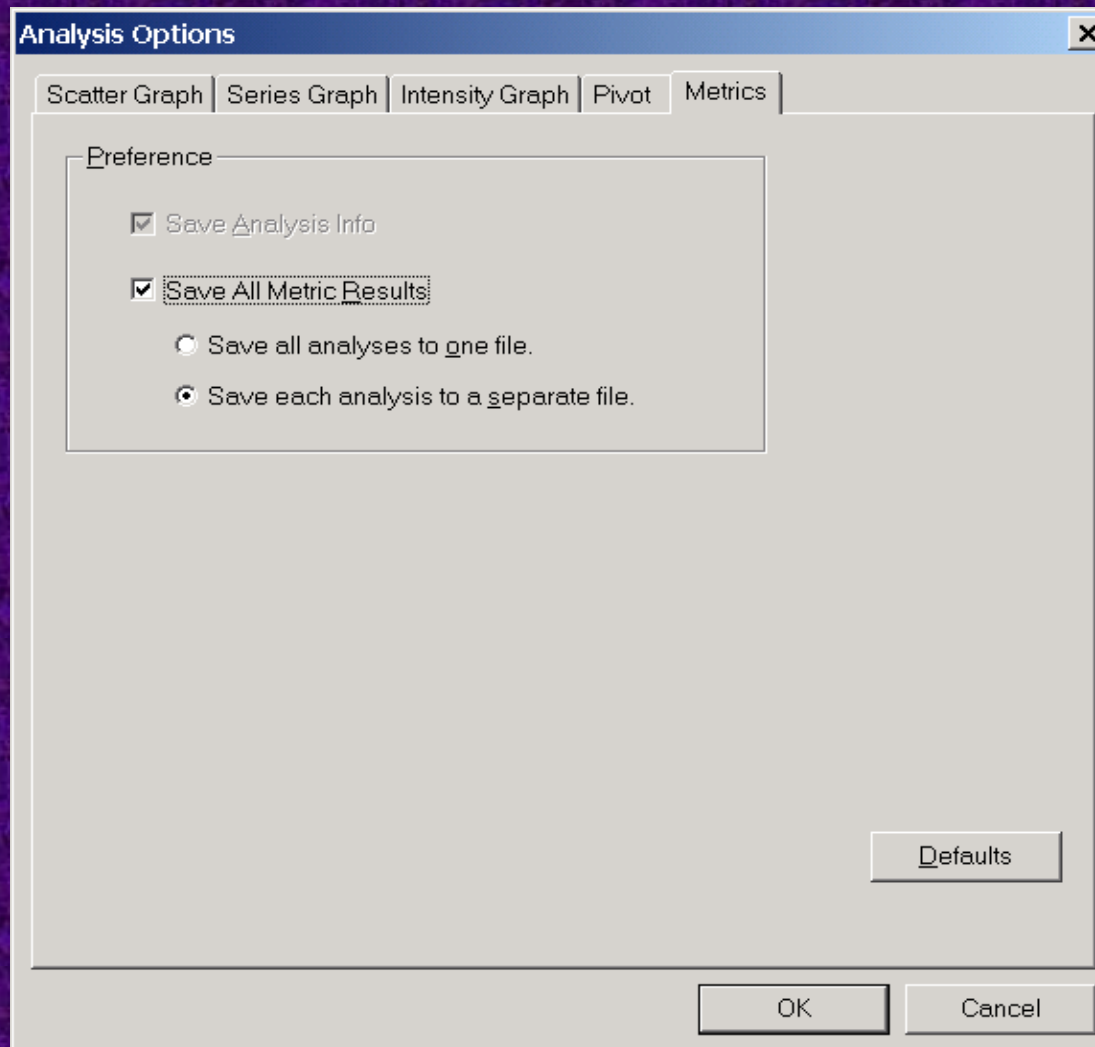
OK Cancel

Set Metrics options:

- Save all Metric Results
- Save each analysis to a separate file

Select Metrics tab before saving

# Affymetrix – CHP file Metrics options



# Upload Status

- Shows your arrays and totals for all users
- Two step process:
  - Data is parsed and entered into Sybase db
  - Image is processed and stored
- You can work with data without waiting for image processing to finish

## mAdb WEB Upload Status Report

Status Updated: Tue Sep 28 10:54:29 EDT 2004  
(This page refreshes every 10 minutes)

### Other mAdb WEB Upload Reports:

Graphical summary by [month](#) (past 12 months) or by [day](#) (past 90 days)

Details of arrays [queued](#) for processing

Details of arrays uploaded within the past [24 hours](#), [7 days](#), [30 days](#) or [all](#)

### mAdb login      Arrays      Status

ncidemo            0            Queued for/or loading into mAdb

**Total all Users** 0            Queued for/or loading into mAdb

ncidemo            0            Loaded; Queued for/or Image processing

**Total all Users** 0            Loaded; Queued for/or Image processing

### Activity for the past 30 days

ncidemo            0            Processing completed

**Total all Users** 1037      Processing completed (332 Affymetrix, 705 Spotted)

ncidemo            0            Canceled, UnConfirmed, Bad Files/Rejected Submissions;

**Total all Users** 44            Canceled, UnConfirmed, Bad Files/Rejected Submissions

# GenePix Analysis Notes

- Download correct GAL file from mAdb
- Carefully grid each block
  - Do not delete any blocks – Mark “bad” instead
- Allow program to “Find spots” and adjust spot size
- Set option to “Analyze absent spots”
- Adjust JPEG for desired contrast/brightness
- Analyze spots

# Common Spotted Uploading Errors

- Choosing wrong print set
  - If you don't see your print in the drop down list, then adjust the search parameters and press "Show" button
  - If you still don't see the print, then contact [madb-support@bimas.cit.nih.gov](mailto:madb-support@bimas.cit.nih.gov)
- Loading GAL file, Excel file, or Set Up file in place of GenePix data (.gpr) file
- Loading multi-image TIFF file instead of composite, single image JPEG or PICT file

# Affymetrix Analysis Notes

- Run chip through fluidics station to get CEL file
- Analyze CEL file (usually scale all spots to 500)
- With CHP file open, set analysis options on metrics tab as:
  - “Save All Metric Results”
  - “Save each analysis to a separate file”
- Click on Metric tab
- Save file as .... Xxxx.txt
- Note: If uploading comparison data, then upload absolute baseline data first.

# Copy or move arrays between projects

- Accessible from the Gateway Tool menu

- Need administrative access to both projects

- Create a “trash” project to “delete” unwanted arrays

**mAddb Copy/Move Arrays**

---

**Options**

Selected Arrays

To Project:

**Arrays from**  
**Project 1038:** Multiple Types Demo Set #4  
**Created on:** Mar 5 2002 9:02AM  
**Description:** Example of repeats of different types (for example tissue, cell lines, animal strain)

**Array Selection**

	A	mAddbID: Array Name & Short Description
<input checked="" type="radio"/>	<input type="radio"/>	28733: Mm-Incyte-v1p1-1 Sample 1/Type A
<input checked="" type="radio"/>	<input type="radio"/>	28742: Mm-Incyte-v1p1-10 Sample 5/Type B
<input checked="" type="radio"/>	<input type="radio"/>	28734: Mm-Incyte-v1p1-2 Sample 2/Type A
<input checked="" type="radio"/>	<input type="radio"/>	28735: Mm-Incyte-v1p1-3 Sample 3/Type A
<input checked="" type="radio"/>	<input type="radio"/>	28736: Mm-Incyte-v1p1-4 Sample 4/Type A
<input checked="" type="radio"/>	<input type="radio"/>	28737: Mm-Incyte-v1p1-5 Sample 5/Type A
<input checked="" type="radio"/>	<input type="radio"/>	28738: Mm-Incyte-v1p1-6 Sample 1/Type B
<input checked="" type="radio"/>	<input type="radio"/>	28739: Mm-Incyte-v1p1-7 Sample 2/Type B
<input checked="" type="radio"/>	<input type="radio"/>	28740: Mm-Incyte-v1p1-8 Sample 3/Type B
<input checked="" type="radio"/>	<input type="radio"/>	28741: Mm-Incyte-v1p1-9 Sample 4/Type B

# Re-order arrays within a project

## Order Arrays within Project

Note: This tool changes the order designation for arrays within this project. All users who have access to this project will see this order designation.

**Arrays**

↑  
Change  
Array  
order.  
↓

Mm-Incyte-v1p1-6 Sample 1/Type B
Mm-Incyte-v1p1-7 Sample 2/Type B
Mm-Incyte-v1p1-8 Sample 3/Type B
Mm-Incyte-v1p1-9 Sample 4/Type B
Mm-Incyte-v1p1-10 Sample 5/Type B
Mm-Incyte-v1p1-1 Sample 1/Type A
Mm-Incyte-v1p1-2 Sample 2/Type A
Mm-Incyte-v1p1-3 Sample 3/Type A
Mm-Incyte-v1p1-4 Sample 4/Type A
Mm-Incyte-v1p1-5 Sample 5/Type A

Submit Cancel

**Change Array Order** by highlighting an array name and using the change array order up and down arrows.  
Click the **Submit** button when finished or the **Cancel** button to return to the Analysis Gateway.

From the mAdb Gateway page, select a project and the “Order Arrays Within a Project” Tool and hit “Continue”



# IV. Evaluating Array Quality

- Signal definition
- Normalization
- Use of log base 2
- Project Summary Report
- Comprehensive Graphical Quality Report

# mAdb Definitions

- Signal - refers to background corrected values (i.e. Target Intensity - Background Intensity).
- Defaults:
  - MEAN Intensity – MEDIAN background (for GenePix)
  - MEAN Intensity – MEAN background (for ArraySuite)
- Normalization factor – initially calculated so that the median overall ratio (Cy5 Signal/ Cy3 Signal) is adjusted to 1.0 (linear space) or 0.0 (log base 2) for each array. Spots with an extremely low signal are excluded from this calculation.

# Need for Normalization of Ratios

- Unequal incorporation of labels (green Cy3 incorporates better than red Cy5)
- Unequal amounts of samples
- Unequal PMT voltage settings
- Different backgrounds
- Total brightness may differ between chips

# Why use ratios converted to log base 2?

- Makes variation of ratios more independent of absolute magnitude
- Symmetrical graphing – otherwise upregulated genes plotted from 1 to  $\infty$  ; downregulated genes compressed between 0 and 1
- Clearer interpretation – negative numbers are downregulated genes; positive numbers are upregulated genes

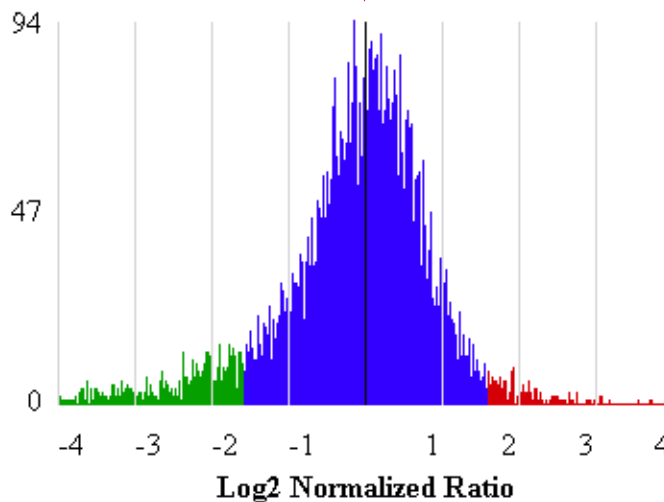
## mAdb Array Histogram

[Comprehensive Graphical Report](#) (Be Patient!)

Array: Mm-Incyte-v1p1-1

Short Description: Sample 1/Type A

Long Description: Add a description **re-centered**



Empty wells and flagged spots filtered out

Green: Ratio < 1/3, Red: Ratio > 3

Mean Signal		Median Bkg		Sgl/Bkg		Not Found	Normal. Factor**
Ch A	Ch B	Ch A	Ch B	Ch A	Ch B		
326	455	110	84	3.0	5.4	30%	0.617

Normalization factor is calculated and multiplied against each ratio to re-center array distribution around 1 (linear), equal to 0 in log base 2

# Project Summary

## mAdb Project Summaries 1.0

Retrieve Array Summaries formatted for

[Edit](#) Project #1038: Multiple Types Demo Set #4

Created on: Mar 05, 2002

Description: Example of repeats of different types (for example tissue, cell lines, animal strain)

### Summary Statistics

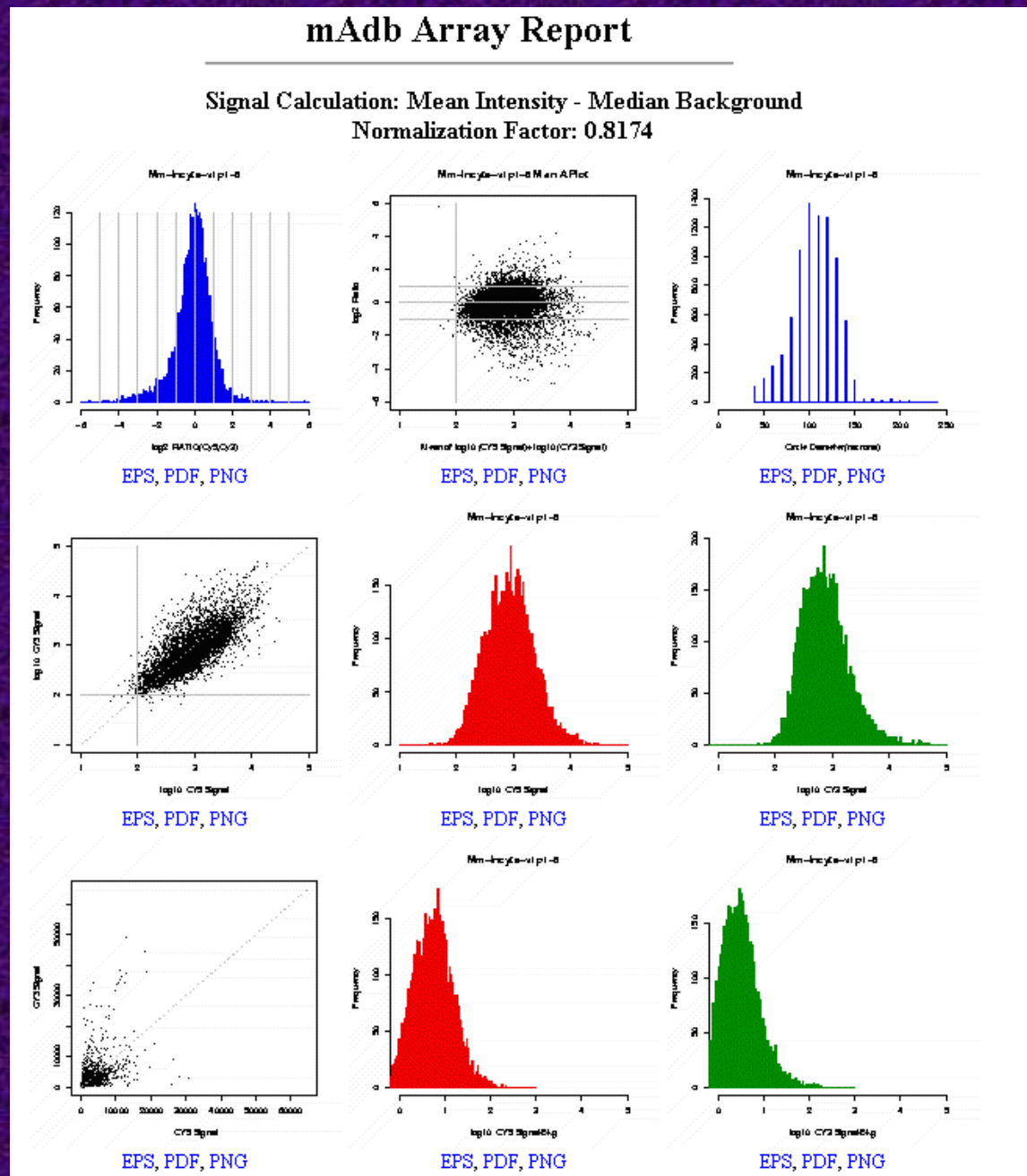
### Array Information

	Mean Signal		Median Bkg		Sgl/Bkg		% Found	Normal. Factor	mAdb ID	Uploaded	Array Print	Array	Probe A	Probe B	Short
	Ch A	Ch B	Ch A	Ch B	Ch A	Ch B									
<a href="#">Edit</a> 1.	326	455	110	84	3.0	5.4	70%	0.626	28733	Mar 5 2002 9:07AM	Mm-Incyte-v1p1-090600	Mm-Incyte-v1p1-1	Control	Sample 1	Samp
<a href="#">Edit</a> 2.	1677	2088	241	160	7.0	13.1	93%	0.769	28742	Mar 5 2002 9:24AM	Mm-Incyte-v1p1-090600	Mm-Incyte-v1p1-10	Control	Sample 5/B	Samp
<a href="#">Edit</a> 3.	880	673	200	364	4.4	1.8	84%	1.055	28734	Mar 5 2002 9:10AM	Mm-Incyte-v1p1-090600	Mm-Incyte-v1p1-2	Control	Sample 2	Samp
<a href="#">Edit</a> 4.	1056	1473	259	154	4.1	9.6	93%	0.658	28735	Mar 5 2002 9:11AM	Mm-Incyte-v1p1-090600	Mm-Incyte-v1p1-3	Control	Sample 3	Samp
<a href="#">Edit</a> 5.	297	493	117	87	2.5	5.7	84%	0.542	28736	Mar 5 2002 9:13AM	Mm-Incyte-v1p1-090600	Mm-Incyte-v1p1-4	Control	Sample 4	Samp
<a href="#">Edit</a> 6.	443	543	123	89	3.6	6.1	83%	0.708	28737	Mar 5 2002 9:15AM	Mm-Incyte-v1p1-090600	Mm-Incyte-v1p1-5	Control	Sample 5	Samp
<a href="#">Edit</a> 7.	499	541	120	101	4.2	5.4	84%	0.858	28738	Mar 5 2002 9:17AM	Mm-Incyte-v1p1-090600	Mm-Incyte-v1p1-6	Control	Sample 1/B	Samp
<a href="#">Edit</a> 8.	626	717	146	113	4.3	6.3	85%	0.890	28739	Mar 5 2002 9:21AM	Mm-Incyte-v1p1-090600	Mm-Incyte-v1p1-7	Control	Sample 2/B	Samp
<a href="#">Edit</a> 9.	1280	1399	272	190	4.7	7.4	93%	0.830	28740	Mar 5 2002 9:22AM	Mm-Incyte-v1p1-090600	Mm-Incyte-v1p1-8	Control	Sample 3/B	Samp
<a href="#">Edit</a> 10.	1113	1371	261	156	4.3	8.8	91%	0.779	28741	Mar 5 2002 9:23AM	Mm-Incyte-v1p1-090600	Mm-Incyte-v1p1-9	Control	Sample 4/B	Samp

- Aid to QC – overall array statistics, links to histogram, array image
- If you have admin access to a project, can edit project and array descriptions from “Edit” links here

# Comprehensive Graphical Quality Report

- Accessed from histogram display
- More QC parameters, including:
  - M versus A plot
  - spot size distribution
  - log and linear plots of each channel
  - signal intensity distribution
  - signal/background distribution





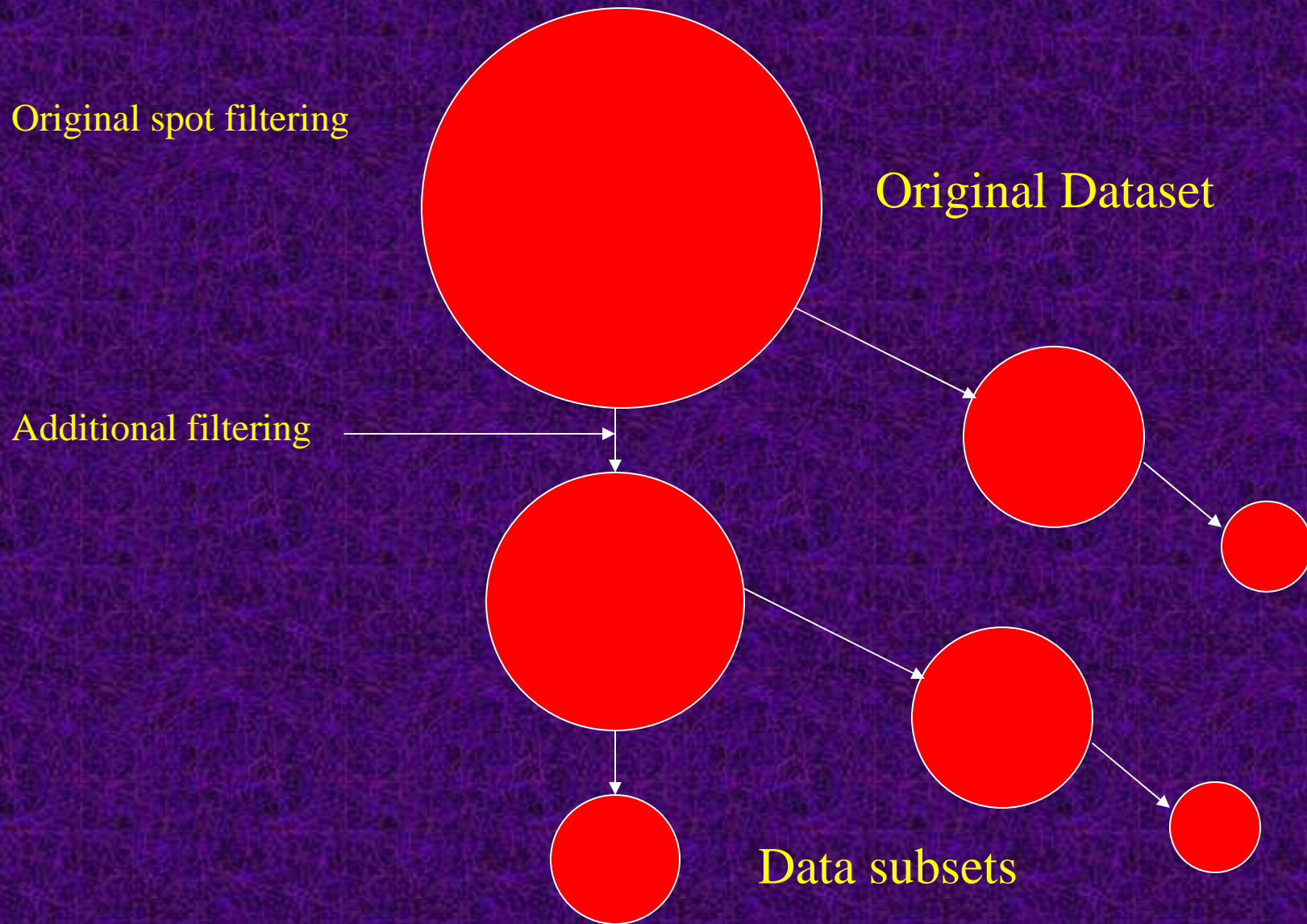


# V. Getting started with analysis

# mAdb Analysis Paradigm

1. **Create project; Upload arrays to that project**
2. **Quality control – Project Summary and Graphical Reports**
3. **Create a filtered dataset:**
  - **Select arrays for analysis**
  - **Define quality parameters (minimum signal values, S/N, etc.)**
  - **Select normalization method, so different arrays can be compared**
  - **Align genes from different array layouts (based on well IDs)**
4. **Apply Data/Gene criteria filters, if desired, to create subset dataset(s)**
5. **Apply appropriate Analysis/Visualization Tools to the dataset(s)**
6. **Repeat Steps 3, 4, and 5 as desired**
7. **Interpret Datasets/Results**

# Dataset Structure -Filtering hierarchy /tree structure



# Lab 1 – Creating a filtered dataset

Goal: To start analyzing arrays using only high quality/reliable spots

Do NOT maximize the browser window, so multiple windows can be distinguished on the monitor.

# Lab 1. Choosing Project and Extended Dataset Extraction Tool

[Home Page](#) | [mAdb Gateway](#) | [Upload Status](#)  
[Forums](#) | [Reference Info](#) | [Program Downloads](#) | [GeneCards](#)

## *mAdb Gateway*

**NEW** [Upload](#) lists of identifiers such as Clone, Gene Symbol, LocusLink ID, UniGene ID and Well ID. These lists can be used as filters with the Feature Properties Filtering tool.

Choose one or more Projects, select a Tool and Continue or access previously extracted data located in **ncidemo**'s: [Permanent](#) area

**Projects:**

- AX guest - Time Course Demo Set #1
- AX guest - Time Course Demo Set #2
- AX guest - Repeats and Reciprocal Retests Demo Set #3
- AU guest - Multiple Types Demo Set #4**
- AU ncidemo - my project
- AU ncidemo - Oligo and cDNA

**Note:** Tools marked with "\*" only support selection of one project

**Tool:** Extended Dataset Extraction

③

④

⑤

1. Open a web browser and type the URL for the mAdb home page, <http://madb-training.cit.nih.gov>.

2. Click the first bullet on the home page, to access the **mAdb Gateway**, web page, shown at left. You will need to login the mAdb Gateway with the mAdb account as instructed.

3. On the mAdb Gateway Web page, in the **Projects:** list, select the “**guest – Multiple Types Demo Set #4**” project  
NOTE: You can select multiple projects by holding down the **Ctrl** key when you click on a project

4. On the **Tools:** menu just below, select “**Extended Dataset Extraction**”

5. Press the **Continue** button

# Lab 1. Selecting Filtering Options

**GenePix Extraction**

Note the ⓘ marks items which lead to additional help when clicked

**Signal, Normalization & Ratio Options ⓘ**

Signal Calculation: Median Int - Median Bkg ⓘ

Normalization Method: 50th Percentile (Median) ⓘ

Default Ratio: ChanB/ChanA (Cy5/Cy3) ⓘ

Limit Normalization to HouseKeeping Genes

Caution: Most array prints do not have an identified set of HouseKeeping Genes

Include Control Features in the extracted set

1

1. In the **Signal, Normalization, & Ratio Options** panel, choose **Signal Calculation: Median Int – Median Bkg**, **Normalization Method: 50<sup>th</sup> Percentile (Median)**, and **Default Ratio: ChanB/ChanA**. Leave the checkboxes empty. Using this Normalization method, the output is re-normalized based on the spots which pass the filters.

**Spot Filter Options ⓘ**

Check boxes on the left to activate specific criteria

Exclude any Spots Indicated as Bad or Not Found ⓘ

Target diameter is between 50 μm and 300 μm

Target Pixels Saturated <= 50 % and 50 %

---

	Chan A (Cy3)		Chan B (Cy5)
<input type="checkbox"/> Target Pixels 1 SD above Bkg >=	30 %	and	30 %
<input type="checkbox"/> Signal Above Background >=	0 SDs	and	0 SDs
<input type="checkbox"/> Signal/Background Ratio >=	2	and	2
<input checked="" type="checkbox"/> Signal >=	200	and	200
<input type="checkbox"/> Override if Chan B Signal >=			5000
<input checked="" type="checkbox"/> Override if Chan A Signal >=	5000		
<input type="checkbox"/> Set Signal Floor Value =	100		100

2

2. In the **Spot Filter Options** panel, check the boxes on the left to activate the appropriate filter(s), and choose appropriate values by typing in numbers into the form elements to the right of each filter checkbox. For the purposes of this exercise, check:
  - Exclude any Spots indicated as **Bad or Not Found**
  - Signal >= **200** and **200**
  - Override if Chan B Signal >= **5000**
  - Override if Chan A Signal >= **5000**

2

3. Go to next page of lab to choose arrays

## Lab 1. Selecting Dataset Properties and Arrays

**Dataset Properties**

Rows Ordered by: Average(Log2 Ratio) Descending

Dataset Location: Transient Area

Dataset Label: My Type A data - qual filtered

**Array Selection**

A

	A	1/R	mAdbID: Array Name & Short Description
<input checked="" type="radio"/>	<input checked="" type="radio"/>	<input type="checkbox"/>	28738: Mm-Incyte-v1p1-6 Sample 1/Type B
<input checked="" type="radio"/>	<input checked="" type="radio"/>	<input type="checkbox"/>	28739: Mm-Incyte-v1p1-7 Sample 2/Type B
<input checked="" type="radio"/>	<input checked="" type="radio"/>	<input type="checkbox"/>	28740: Mm-Incyte-v1p1-8 Sample 3/Type B
<input checked="" type="radio"/>	<input checked="" type="radio"/>	<input type="checkbox"/>	28741: Mm-Incyte-v1p1-9 Sample 4/Type B
<input checked="" type="radio"/>	<input checked="" type="radio"/>	<input type="checkbox"/>	28742: Mm-Incyte-v1p1-10 Sample 5/Type B
<input type="radio"/>	<input checked="" type="radio"/>	<input type="checkbox"/>	28733: Mm-Incyte-v1p1-1 Sample 1/Type A
<input type="radio"/>	<input checked="" type="radio"/>	<input type="checkbox"/>	28734: Mm-Incyte-v1p1-2 Sample 2/Type A
<input type="radio"/>	<input checked="" type="radio"/>	<input type="checkbox"/>	28735: Mm-Incyte-v1p1-3 Sample 3/Type A
<input type="radio"/>	<input checked="" type="radio"/>	<input type="checkbox"/>	28736: Mm-Incyte-v1p1-4 Sample 4/Type A
<input type="radio"/>	<input checked="" type="radio"/>	<input type="checkbox"/>	28737: Mm-Incyte-v1p1-5 Sample 5/Type A

1. In the **Dataset Properties** panel, choose **Rows Ordered by: Average(Log2 Ratio) and Descending; Dataset Location: Transient Area**, and **Dataset Label: “My Type A data – qual filtered”**.
2. In the **Array Selection** panel, choose just the Type A arrays using the radio buttons under **A**. **N.B.** If a dye swap or reverse fluor, check the **1/R** box to take the reciprocal value of the ratio for direct comparison.
3. Press **Submit**

## Lab 1. Waiting for Data Extraction ...

This page monitors the progress and allows you to continue when the results are available.

**Please wait for completion.**

Waiting ...

**Done! Please click**

**NOTE:** The dataset has been stored in your **Temporary** area. Datasets stored in the Temporary area are automatically deleted when 14 days expire with no access to the data. Accessing (that is "opening") the original set or a derived filtered/adjusted subset resets the "14 day clock". The mAdb Dataset management tool allows you to delete datasets from this area.

[Home](#) | [Analysis Tools](#) | [Forums](#) | [Reference Info](#) | [Program Downloads](#) | [GeneCards](#)

Intermediate screen which monitors the data extraction process. When the creation of the working dataset is complete, the user can continue to the Data Display page.



## Extended Tool: Signal, Normalization & Ratio Options:

- **Signal Calculation**
  - Mean Intensity – Median Background
  - Median Intensity – Median Background
  - Median or Mean Intensity (with no Background subtraction)
- **Normalization**
  - None
  - 50<sup>th</sup> Percentile (Median)
    - Applied to extracted spots (spots passing filter)
    - All spots or only Housekeeping spots (on limited prints)
  - Pre-calculated 50<sup>th</sup> percentile (based on all spots)
  - Loess non-linear normalization
- **Default Ratio**

Chan B/Chan A (CY5/CY3),  
but for reverse fluor can choose Chan A/Chan B (CY3/CY5)

## Spot Filter Options:

**Important** - Check box to Activate!

- Exclude any Spots Flagged as *Bad Or Not Found, Bad*
- Target diameter is between *xx and yy microns*
- Target Pixels Saturated
- Target Pixels 1 Standard Deviation above background  $\geq N \%$
- Signal above background  $\geq N$  SDs (*standard deviations*)
- Signal/Background Ratio  $\geq N$
- Signal  $\geq N$  (*raw signals*)
- Override bracketed criteria ( in yellow above) if Chan B and /or A  
Signal  $\geq N$

# Signal Floor

- When one channel has a very low signal and the other has a moderate or high signal, the resulting ratio value could be misleading (i.e. very high/low)
- To adjust such a highly skewed ratio, mAdb allows the user to set a floor (e.g. 100) for signals below a threshold
- Compare  $50000/1$  vs  $50000/100$

# Lab 1. Main mAdb Dataset Display – Part 1

1. The listing at the top shows the array group, a link to the array image, a link to a histogram display, the re-calculated normalization factor (based on those spots which passed the quality filters), the array name, and the short description for all of the chosen arrays to be filtered
2. After the Dataset name (which can be **edited** with the link to the left), is the history of what was done in the preceding filtering step.
3. Go to the next page of the lab and scroll down to the bottom of the Web page.

**mAdb Dataset Display**

[View](#) Array Summaries

A [0.644](#) 1. Mm-Incyte-v1p1-1 Sample 1/Type A

A [1.056](#) 2. Mm-Incyte-v1p1-2 Sample 2/Type A

A [0.627](#) 3. Mm-Incyte-v1p1-3 Sample 3/Type A

A [0.551](#) 4. Mm-Incyte-v1p1-4 Sample 4/Type A

A [0.727](#) 5. Mm-Incyte-v1p1-5 Sample 5/Type A

[Edit](#) Data for Dataset: My Type A data - qual filtered

5 Arrays and 5276 Expression Rows extracted.  
Default Ratio: ChanB/ChanA (Cy5/Cy3)  
Signal calculation: Median Intensity minus Median Background  
Any Features designated Control were excluded.  
Normalization method: 50th Percentile (Median) using all spot filtered Genes  
Spot Filter Options:  
Include Spots not flagged BAD or Not Found  
AND Target diameter >= 50 um AND Target diameter <= 300 um  
AND Both Chan A and Chan B Signal/Background Ratios >= 2.000  
Override other Chan A & B criteria and Include if Chan A Signal >= 5000 OR Chan  
Data was extracted and aligned by the Inventory Well ID  
Any multiple occurrences of Well ID were reduced to a single instance  
by selecting the one with the strongest signal (Chan A + Chan B)

## Lab 1. Main mAdb Dataset Display – Part 2

Records 1 to 25 of 5276 total records

#1	#2	#3	#4	#5	Well ID	Feature ID	Description
		4.2019			616842	IMAGE:481151	procollagen, type IX, alpha 1
	4.2293	4.1005			617147	IMAGE:493658	lipocalin 2
		4.0493		3.7699	614212	IMAGE:402800	Mus musculus transcribed sequences
		3.8949			614066	IMAGE:374725	RIKEN cDNA 2310047E01 gene
		3.0624			613588	IMAGE:333418	protein tyrosine phosphatase, receptor type, l
	3.7330	3.9396	1.7578	2.7487	617076	IMAGE:571759	RIKEN cDNA 9530006B08 gene
3.2349	2.7053	3.0574	2.8567	3.3126	614354	IMAGE:403453	protein tyrosine phosphatase, receptor type, l
	2.8860	2.8628	2.9509	3.3728	619013	IMAGE:832158	extracellular proteinase inhibitor

1. This is the main page to display expression data, and as we will see on the next page, is highly customizable. Each column represents an array, each row a gene feature. Gray boxes are either missing values or data that was filtered out due to low quality. You can page through the data using the **arrow** just above the columns of data.
2. The mAdb **Well ID** uniquely identifies the piece of DNA used on that feature, and the **Feature ID** is an external identifier. The **Well ID** is a hyperlink to a montage of the spot images and raw signal values, whereas the **Feature ID** is a **Hyperlink to a Feature Report**, integrating information about the gene related to the feature and its function(s).
3. There is a brief description of the feature on the right hand side of the display. Note that each column can be sorted in either ascending or descending order using the **grey arrows** above each column.

## *mAdb* Feature Report

---

**Clone** [IMAGE:402800](#)  
**Library Source** Soares mouse embryo NbME13.5 14.5  
**5' Sequence** [W80005](#) BLAST Results: [NT](#) [NR](#) [RefSeq](#) [Genome](#)

---

**mAdb Annotation Source** GenBank Accession W80005 Maps to UniGene Cluster

**mAdb Mapping**      **Mapped**    **Str** **Chr**      **Start:Stop BP**      **Cytoband**    **Genome Assembly**  
 Mm.279310            3    132819050:132887928    3 G3    NCBI B36 (UCSC mm8)

**CytoGenetic Map** 3 G3

**Entrez GeneID** [114249](#)

**Gene Symbol** Npnt

**UniGene Cluster** [rp Mm.279310](#) CGAP's Gene Stanford's S.O.U.R.C.E.

**RefSeqs in Cluster** [NM\\_001029836](#) [NM\\_033525](#)

**UG Title** Nephronectin

**KEGG Pathways** [ECM-receptor interaction](#)

GO <sup>TM</sup> Annotations	Evidence	Source	Pub
<b>Function</b>			
<a href="#">calcium ion binding</a>	RCA	MGI	<a href="#">PM</a>
<a href="#">integrin binding</a>	IPI	MGI	<a href="#">PM</a>
<b>Process</b>			
<a href="#">cell-matrix adhesion</a>	IDA	MGI	<a href="#">PM</a>
<a href="#">cell-matrix adhesion</a>	IPI	MGI	<a href="#">PM</a>
<b>Component</b>			
<a href="#">extracellular matrix (sensu Metazoa)</a>	IDA	MGI	<a href="#">PM</a>
<a href="#">extracellular matrix (sensu Metazoa)</a>	IPI	MGI	<a href="#">PM</a>
<a href="#">extracellular space</a>	RCA	MGI	<a href="#">PM</a>
<a href="#">membrane</a>	RCA	MGI	<a href="#">PM</a>

---

## Lab 1. Main mAdb Dataset Display – Part 3

Dataset Retrieval & Display Options

Retrieve Dataset formatted for Eisen Cluster

Redisplay

Show Array Details at the top of the page

Background Color Red/Yellow/Green Contrast 3

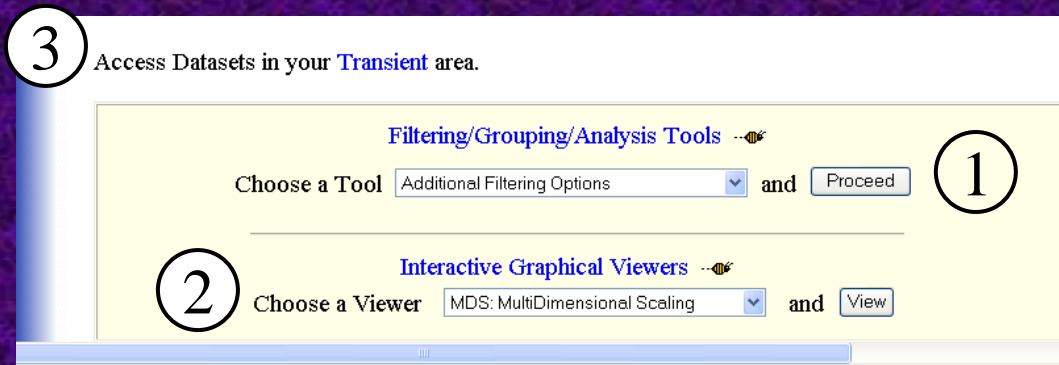
Limiting display to to 25 genes

<input type="checkbox"/> Show Data Values	<input type="checkbox"/> Use Names in Column Heading
<input checked="" type="checkbox"/> Apply log2 transform	<input type="checkbox"/> Use Description in Column Heading
<input type="checkbox"/> Show Spot Images	<input type="checkbox"/> <b>NEW</b> Show Physical Mapping
<input checked="" type="checkbox"/> Show UniGene Cluster	<input type="checkbox"/> Show Entrez GeneID
<input type="checkbox"/> Show Locus Tags	<input checked="" type="checkbox"/> Show Gene Symbols
<input checked="" type="checkbox"/> Show BioCarta Pathways	<input checked="" type="checkbox"/> Show KEGG Pathways
<input type="checkbox"/> Show GO Tier 2 Component	<input type="checkbox"/> Show GO Tier 3 Component
<input type="checkbox"/> Show GO Tier 2 Function	<input type="checkbox"/> Show GO Tier 3 Function
<input type="checkbox"/> Show GO Tier 2 Process	<input type="checkbox"/> Show GO Tier 3 Process
<input checked="" type="checkbox"/> Show Gene Description	<input type="checkbox"/> Show GO Terms

[Save](#) a Feature Property List (used with the Feature Properties Filtering tool).

1. Here is where the data display on the preceding page can be customized, by checking or unchecking the checkboxes next to each column name. One can include numerical data (**Log2 Ratio**); pathways (**KEGG, BioCarta**); Gene Ontology (**GO**) classifications; and display individual **Spot Images**, among others. One can also change or eliminate the **Background Color** on the table of data values, adjust its **Contrast** (the point where max red and green are reached), and also adjust how many genes are displayed in the table on a Web page (the default is 25). Once the choices are made, push the **Redisplay** button to refresh the page with your desired changes.
2. You can also retrieve the dataset for MS-Excel, the Eisen Cluster program format, or in tab-delimited files for the Macintosh, PC, or UNIX platforms.

## Lab 1. Main mAdb Dataset Display – Part 4




1. Once the data is filtered by quality, the most likely next step is to do additional filtering and create a subset of this parent dataset. Under *Filtering/Grouping/Analysis Tools*, choose the default pulldown option of **Additional Filtering Options** and press **Proceed**.
2. Alternately, one could access *Interactive Graphical Viewers* from here,
3. Also, you could **Access other Datasets in your Transient Area** from here with the link above the yellow panels.



# Affy Extraction Tool (for Absolute data)

## Affymetrix GeneChip Analysis Options


Note the  marks items which lead to additional help when clicked

### Affy GeneChip Expression Analysis Options

Probeset Analysis:

GeneChip Type:

## Affymetrix GeneChip MAS5 Options and Array Selection

Note the  marks items which lead to additional help when clicked

### Data Transformation Options

Trimmed Mean Scaling Target:

Transformation:

Signal Floor =

### Filter Options

Check boxes on the left to activate specific criteria

Exclude All Present (P) Calls

Exclude All Marginal (M) Calls

Exclude All Absent (A) Calls

Present (P) Call AND Signal  $\geq$

Marginal (M) Call AND Signal  $\geq$

Absent (A) Call AND Signal  $\geq$

# Sample Analysis Questions

- How can I evaluate the consistency of the arrays across my biological repeats?
- Which genes have enough data points to give confidence in the results?
- Which genes have values that are less consistent across the arrays?
- How can I keep track of these genes that seem to have unreliable values?
- Which genes are most differentially expressed?
- Are any of these genes in my “unreliable” list?

# Lab 2 – Assessing array correlation

Goal: To evaluate the consistency of data values across a set of arrays and determine which genes are not well correlated based on a minimal number of data points

# Evaluating correlation across all pairs of arrays

Filtering/Grouping/Analysis Tools

Choose a Tool: Additional Filtering Options and Proceed

Choose a View: View

Retrieve Data

Redisplay

2


Column Heading  
on in Column Heading

Show Spot Images  Show Gene Symbols

From the mAdb Dataset Display Page, select the “Correlation Summary Report” Tool and hit the “Proceed” button

# Correlation Summary Report

(How can I evaluate the consistency of the arrays across my biological repeats?)










Redisplay      Background Color Scheme      Green/White/Red 

Color Saturation Max/Mid/Min      1      .85      .75

Note: For proper coloring Max > Mid > Min


**Note:** Click on the Correlation values to display the corresponding ScatterPlot

## Correlations

	A	A	A	A	A					
	#1	#2	#3	#4	#5	Grp		Array Name	Array Description	
1.	#1 A	0.855	0.927	0.917	0.912	A	 	1. Mm-Incyte-v1p1-1	Sample 1/Type A	
2.		#2 A	0.802	0.844	0.831	A	 	2. Mm-Incyte-v1p1-2	Sample 2/Type A	
3.			#3 A	0.948	0.935	A	 	3. Mm-Incyte-v1p1-3	Sample 3/Type A	
4.				#4 A	0.940	A	 	4. Mm-Incyte-v1p1-4	Sample 4/Type A	
5.					#5 A	A	 	5. Mm-Incyte-v1p1-5	Sample 5/Type A	


Allows pair wise comparison of all arrays in a project – useful for comparing replicates and reverse fluors

# Evaluating correlation between two arrays

Filtering/Grouping/Analysis Tools 

Choose a Tool  and

---

Interactive Graphical Viewers 

Choose a Viewer  and

- MDS: MultiDimensional Scaling
- MDS: MultiDimensional Scaling
- PCA: Principal Components Analysis
- Multi-Array Viewer
- Scatter Plot - log Ratios**

Dataset formatted for

---

Show Array Details at the top of the page

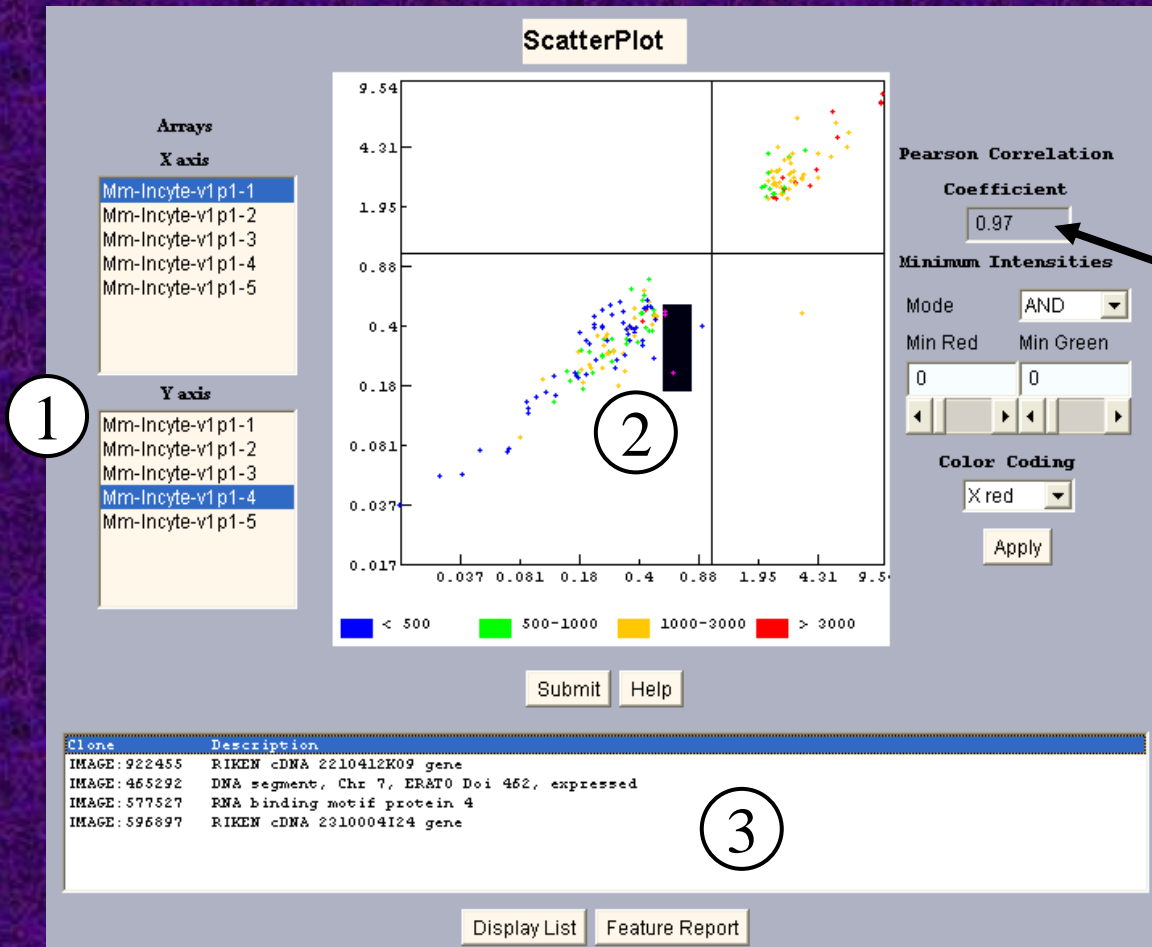
Background Color  Contrast

Limiting display to

<input checked="" type="checkbox"/> Show Data Values	<input type="checkbox"/> Use Names in Column Heading
<input checked="" type="checkbox"/> Apply log2 transform	<input type="checkbox"/> Use Description in Column Heading
<input type="checkbox"/> Show Spot Images	<input checked="" type="checkbox"/> Show Gene Symbols

From the mAdb Dataset Display Page, select the “Scatter Plot – log Ratios” Tool and hit the “View” button

# Visualization Tools – Interactive Scatter Plot Applet



- Replicate experiments should be on a 45° angle (slope of 1) and the Pearson Correlation Coefficient should be approaching 1

- Reverse fluor experiments should have a Pearson Correlation Coefficient approaching -1

Access from *Interactive Graphical Viewers* Menu on main **mAdb Dataset Display** page:

1. Choose Arrays to be compared on X and Y axes
2. Can select outlying spots with mouse – genes will be shown in window below plot
3. Can get **Feature Report** by clicking on gene name in lower display box

## Selecting spots based on value characteristics

The screenshot shows the 'Filtering/Grouping/Analysis Tools' window. At the top, it says 'Filtering/Grouping/Analysis Tools' with a speaker icon. Below this, there are several sections:

- Choose a Tool:** A dropdown menu is open, showing a list of tools. The first item, 'Additional Filtering Options', is highlighted in blue. Other items include 'Ad Hoc Query/Filtering Options', 'Feature Property Filtering Options', 'Array Order Designation/Filtering', 'Array Group Assignment/Filtering', 'Filter/Group by Array Properties', 'Average Arrays within Groups', 'Group Statistics (mean, median, stddev...)', 'Group Comparison (t-test, ANOVA, Wilcoxon, ...)', 'SAM: Significance Analysis of Microarrays', 'PRE-BETA Missing Value Imputation', 'PAM: Prediction Analysis for Microarrays', 'Boolean Comparison with another Set', 'Clustering: Hierarchical', 'Clustering: Kmeans', 'Clustering: SOM', 'Correlation Summary Report', 'Gene Ontology Summary Report', 'Pathways Summary Report', 'Save As a New Dataset', and 'Show Spot Images'.
- Choose a View:** A 'View' button is visible.
- Data:** A 'Retrieve' button is visible.
- List:** A 'Redisplay' button is visible.
- Other elements:** An 'and' label, a 'Proceed' button, a '2' in a text box, and labels for 'Column Heading' and 'on in Column Heading' are also present.


From the mAdb Dataset Display Page, select the “Additional Filtering Options” Tool and hit the “Proceed” button



## Filtering based on missing values


**Data Filtering Options**

Check boxes on the left to activate specific filters  
▼

**Missing Value Filters**  1

Genes: Require values in  $\geq$   Arrays

Arrays: Require values in  $\geq$   % of Genes

**Gene Filters** 

Ratio  $\geq$   in  $\geq$   % of Arrays   
 Apply Symmetrically

Ratio  $\geq$   in  $\geq$   Arrays  OR  
Ratio  $\leq$   in  $\geq$   Arrays

Average Ratio  $\geq$    
 Apply Symmetrically

Max (Ratio) / Min (Ratio)  $\geq$

Variance (Gene Vector) percentile  $\geq$   %

Subset Label:  2

3

1. Filter the rows of data from the parent dataset for missing values, requiring genes in  $\geq 3$  Arrays. Alternately, it is possible to filter out Arrays by requiring values in  $\geq 60\%$  of genes.
2. Label the subset “value required in 60% of arrays”
3. Press the **Filter** button to continue and create the desired subset.

# Filtering based on missing values

(Which genes have enough data points to give confidence in the results?)

Edit Data for Subset: value required in 60% of arrays  
from Dataset: Extracted type A

The filter input data set contained 5 arrays and 7525 genes.  
The filtered output data set contains 5 arrays and 3771 genes.  
3754 genes excluded for being present in less than 60% (3) arrays.

View the complete [History](#).

[Expand](#) this Dataset.

Access Datasets in your [Temporary](#) area.

1

2

3

Records 1 to 25 of 3771 total records displayed.

A	A	A	A	A	Well ID	Feature ID	Gene
#1	#2	#3	#4	#5			
-0.5580	-0.5632	-0.1758	-0.4063	-0.4641	621790	IMAGE:651430	Mkrm2
	-0.5094	-0.6059		-0.2902	621785	IMAGE:523795	Ptdss1
0.5718	-0.0435	0.4337	0.6030	0.4537	621781	IMAGE:533862	Ccl3
	0.3440	0.5024		0.2958	621779	IMAGE:533299	Xpo1
	2.7681	1.9809		1.4138	621777	IMAGE:522713	

1. Note that in the returned dataset, there are many fewer missing values – see the history log for how many genes were filtered out to create this subset.
2. This is a data subset – you can view the complete History of the dataset via this link.
3. You can also **Expand this Dataset** to show the parent and all children, or again **Access Datasets in your Temporary Area** via these links.

## Notes:

- Applies selected filtering options to the dataset based on values in the data and creates a new subset.
- For gene filters, ratios are expressed as fold changes and all calculations are done in log space

# Calculating Group Statistics

The screenshot shows a software interface titled "Filtering/Grouping/Analysis Tools". It features a dropdown menu for selecting a tool, currently set to "Additional Filtering Options". The dropdown list includes various analysis tools, with "Group Statistics (mean, median, stddev...)" highlighted. Other tools listed include "Ad Hoc Query/Filtering Options", "Feature Property Filtering Options", "Array Order Designation/Filtering", "Array Group Assignment/Filtering", "Filter/Group by Array Properties", "Average Arrays within Groups", "Group Comparison (t-test, ANOVA, Wilcoxon, ...)", "SAM: Significance Analysis of Microarrays", "PRE-BETA Missing Value Imputation", "PAM: Prediction Analysis for Microarrays", "Boolean Comparison with another Set", "Clustering: Hierarchical", "Clustering: Kmeans", "Clustering: SOM", "Correlation Summary Report", "Gene Ontology Summary Report", "Pathways Summary Report", "Save As a New Dataset", and "Show Spot Images". There are also checkboxes for "Show Gene Symbols" and "Show Gene Symbols". The interface includes buttons for "Retrieve", "Redisplay", "View", and "Proceed".

From the mAdb Dataset Display Page, select the “Group Statistics” Tool and hit the “Proceed” button

# Filtering on Group Statistics

(Which genes have values that are less consistent across the arrays?)

Filtering/Grouping/Analysis Tools

Choose a Tool: **Statistics Results Filtering** and **Proceed**

Choose a View: **Statistics Results Filtering** and **View**

Retrieve Data

Redisplay

Library

- Array Order Designation/Filtering
- Array Group Assignment/Filtering
- Filter/Group by Array Properties
- Average Arrays within Groups
- Group Statistics (mean, median, Group Comparison (t-test, ANOVA)
- SAM: Significance Analysis of Microarrays
- PRE-BETA Missing Value Imputation
- PAM: Prediction Analysis for Microarrays
- Boolean Comparison with another Dataset
- Clustering: Hierarchical
- Clustering: Kmeans
- Clustering: SOM
- Correlation Summary Report
- Gene Ontology Summary Report
- Pathways Summary Report
- Save As a New Dataset

New tool appears when statistical results are present in the dataset

## Statistics Results Filtering Options

Check boxes on the left to activate specific filters

- Group A Mean  $\geq$  0
- Group A Median  $\geq$  0
- Group A StdDev  $\geq$  1

Subset Label: group standard deviation  $\geq$  1

Filter

Cancel

# Sorting on Group Statistics

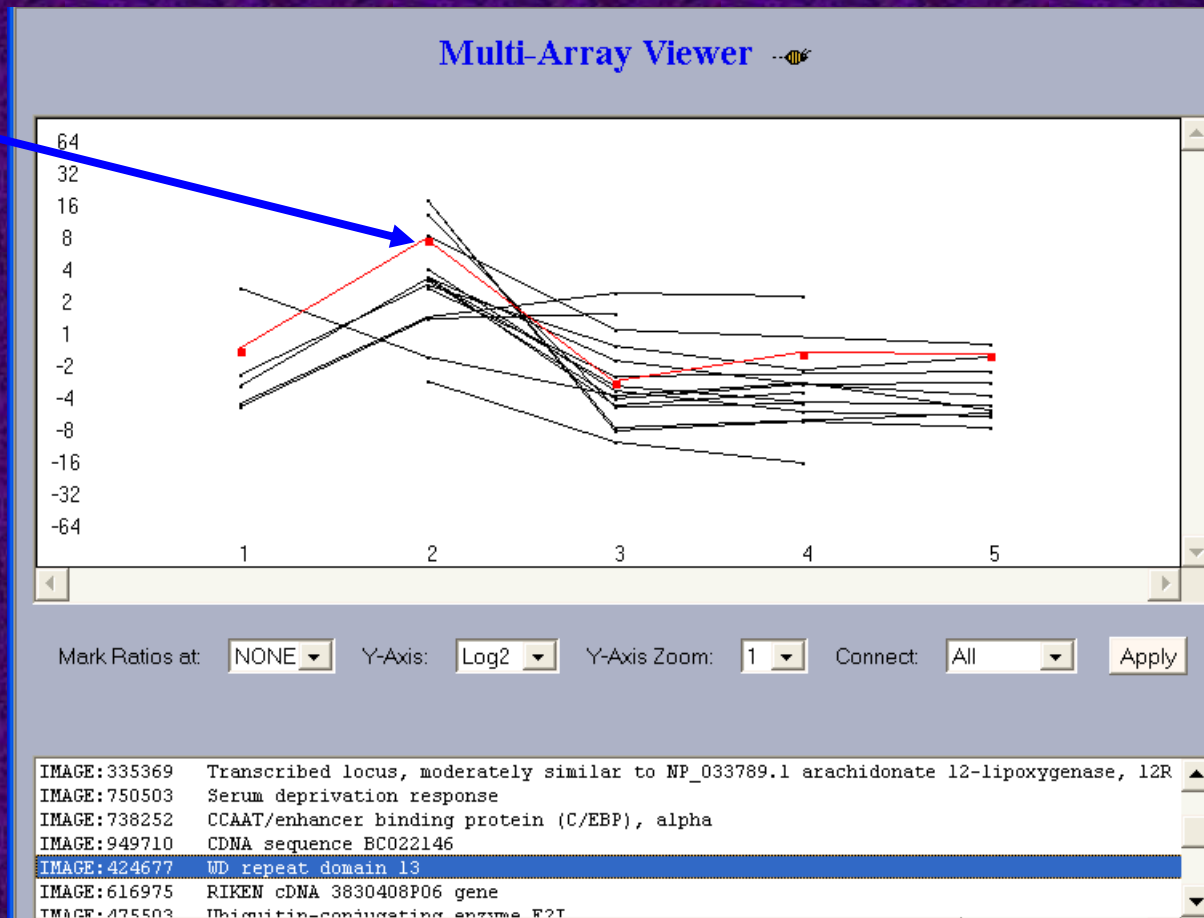
- Group Means
  Group Medians  
 Group StdDevs

Save a Feature Property List (used with the Feature Properties Filtering tool).

Records 1 to 16 of 16 total records displayed.

A	A	A	A	A	↓ ↑	↓ ↑	↓ ↑	↓ ↑	↓ ↑
#1	#2	#3	#4	#5	Group A Mean	Group A StdDev	Well ID	Feature ID	Gene
	4.4465	-2.6394		-2.2019	-0.1316	3.242	616653	IMAGE:480196	
	4.4261	-2.7721	-2.4251	-2.6466	-0.8544	3.051	613650	IMAGE:336497	Slc39a4
	3.9956	-2.0067	-1.9103		0.0262	2.807	613408	IMAGE:331681	Pcbd
	2.3167	-1.5024	-2.1655	-2.3175	-0.9172	1.892	621404	IMAGE:777580	
	2.0528	-1.9512	-1.5359		-0.4781	1.798	620614	IMAGE:738252	Cebpa
-0.1835	3.2536	-1.2083	-0.2851	-0.3586	0.2436	1.549	615066	IMAGE:424677	Wdr13
	3.3436	0.3888		-0.0690	1.2212	1.512	618887	IMAGE:643725	Ramp2
	1.9644	-1.7476	-1.2831	-1.2732	-0.5849	1.484	613517	IMAGE:330336	H2-K1
-1.3422	2.0174	-1.3657	-1.8360	-1.9434	-0.8940	1.476	621175	IMAGE:803488	Sfrp1
	2.0624	-0.5530	-1.2824	-1.6303	-0.3508	1.447	621320	IMAGE:790857	Tex261
-1.9145	0.8009	1.5663	1.4701		0.4807	1.414	616332	IMAGE:475503	Ube2i
1.6904	-0.4691	-1.6724	-1.2459	-2.0954	-0.7585	1.337	619489	IMAGE:949710	BC022146
-2.0005	0.7449	0.9103			-0.1151	1.335	617168	IMAGE:573075	Bag2
	1.6819	-1.0463		-0.8763	-0.0802	1.248	613545	IMAGE:335369	
	-1.2203	-3.0920	-3.7503		-2.6875	1.072	620919	IMAGE:750503	Sdpr
-1.0227	1.8079	-0.1077	-0.8707	-0.4519	-0.1290	1.02	617980	IMAGE:616975	3830408P0

User can sort rows by clicking on up/down arrows above columns



Access from *Interactive Graphical Viewers* Menu on main **mAdb Dataset Display** page :

1. Can choose a point on graphical window to display a graph of that gene's expression which passes through that point
2. Can select a gene name on lower list and graph will appear in plot above
3. Can get **Feature Report** by clicking on gene name in lower display box

# Save a Feature Property List

(How can I keep track of these genes that seem to have unreliable values?)

Group Means
  Group Medians

Group StdDevs

Save a Feature Property List (used with the Feature Properties Filtering tool).

Records 1 to 16 of 16 total records displayed.

#1	#2	#3	#4	#5	Group A Mean	Group A StdDev	Well ID	Feature ID	Gene
	4.4465	-2.6394		-2.2019	-0.1316	3.242	616653		
	4.4261	-2.7721	-2.4251	-2.6466	-0.8544	3.051	613650		
	3.9956	-2.0067	-1.9103		0.0262	2.807	613408		
	2.3167	-1.5024	-2.1655	-2.3175	-0.9172	1.892	621404		
	2.0528	-1.9512	-1.5359		-0.4781	1.798	620614		
-0.1835	3.2536	-1.2083	-0.2851	-0.3586	0.2436	1.549	615066		
	3.3436	0.3888		-0.0690	1.2212	1.512	618887		
	1.9644	-1.7476	-1.2831	-1.2732	-0.5849	1.484	613517		
-1.3422	2.0174	-1.3657	-1.8360	-1.9434	-0.8940	1.476	621175		
	2.0624	-0.5530	-1.2824	-1.6303	-0.3508	1.447	621320		
-1.9145	0.8009	1.5663	1.4701		0.4807	1.414	616332		
1.6904	-0.4691	-1.6724	-1.2459	-2.0954	-0.7585	1.337	619489		
-2.0005	0.7449	0.9103			-0.1151	1.335	617168		
	1.6819	-1.0463		-0.8763	-0.0802	1.248	613545		
	-1.2203	-3.0920	-3.7503		-2.6875	1.072	620919		
-1.0227	1.8079	-0.1077	-0.8707	-0.4519	-0.1290	1.02	617980		

**mAdb: Save a Feature Property List**

---

Save a List of: mAdb Well IDs

Store the List as: Global (Available in all Datasets)

List Label: high std dev genes

Overwrite any existing list with the same label

---

Save

- Can save a list of well IDs, clone/feature identifiers, gene symbols, UniGene identifiers from the dataset display page
- List can be stored as local to the dataset or globally available to all datasets

# Dataset History

History for Subset: **group standard deviation**  $\geq 1$   
from Dataset: **Extracted type A**

5 Arrays and 7525 Expression Rows extracted.  
Default Ratio: ChanB/ChanA (Cy5/Cy3)  
Signal calculation: Median Intensity minus Median Background  
Any Features designated Control were excluded.  
Normalization method: 50th Percentile (Median) using all spot filtered Genes  
Spot Filter Options:  
Include Spots not flagged BAD or Not Found  
AND Both Chan A and Chan B Signal  $\geq 200$   
Override other Chan A & B criteria and Include if Chan A Signal  $\geq 5000$  OR Chan B Signal  $\geq 5000$   
Data was extracted and aligned by the Inventory Well ID  
Any multiple occurrences of Well ID were reduced to a single instance  
by selecting the one with the strongest signal (Chan A + Chan B)  
Note: For all GenePix results from Axon scanned arrays Chan A is CY3 and Chan B is CY5.  
Rows ordered by Average(Log Ratio) descending.

-----  
Fri Aug 19 11:29:03 EDT 2005

5 arrays, 7525 genes in the [original Dataset](#)  
3771 Genes and 5 arrays passed filters  
3754 genes excluded for being present in less than 60% (3) arrays.

-----  
Fri Aug 19 11:29:35 EDT 2005

[input Dataset](#)  
Group Statistic calculations performed for each Group

-----  
Fri Aug 19 11:35:07 EDT 2005

3771 genes in the [input Dataset](#)  
The filtered output data set contains 16 genes  
3755 genes excluded by **Group A StdDev**  $\geq 1$

Link to the [output Dataset](#)

A log is maintained for each dataset tracing the analysis history.  
When the history is displayed, links are provided to allow the user to  
recall any dataset in the analysis chain.

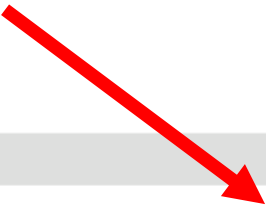


# Lab 3 – Examining differentially expressed genes

Goal: To find differentially expressed genes and evaluate the reliability of values

# Opening earlier subset

Active Subsets		<a href="#">Need Help?</a>	Containing	
Label			Arrays	Genes
<a href="#">Rename</a>	Extracted type A	<a href="#">Open</a>	5	7525
<a href="#">Rename</a>	value required in 60% of arrays	<a href="#">Open</a> <a href="#">History</a>	5	3771
<a href="#">Rename</a>	Group Statistics for value required in 60% of arrays	<a href="#">Open</a> <a href="#">History</a>	5	3771
<a href="#">Rename</a>	group standard deviation $\geq 1$	<a href="#">Open</a> <a href="#">History</a>	5	16



1. From the mAdb Dataset display page, click on the “Expand this Dataset” link to view all subsets
  1. “Open” subset named “value required in 60% of arrays”

# Refining spot selection criteria

The screenshot shows a software interface titled "Filtering/Grouping/Analysis Tools". On the left, there are sections for "Choose a Tool" and "Choose a View". The "Choose a Tool" section has a dropdown menu currently displaying "Additional Filtering Options", with a "Proceed" button to its right. The "Choose a View" section has a "View" button. Below these are "Retrieve" and "Redisplay" buttons. A central list of tools is shown, including "Additional Filtering Options", "Ad Hoc Query/Filtering Options", "Feature Property Filtering Options", "Array Order Designation/Filtering", "Array Group Assignment/Filtering", "Filter/Group by Array Properties", "Average Arrays within Groups", "Group Statistics (mean, median, stddev...)", "Group Comparison (t-test, ANOVA, Wilcoxon, ...)", "SAM: Significance Analysis of Microarrays", "PRE-BETA Missing Value Imputation", "PAM: Prediction Analysis for Microarrays", "Boolean Comparison with another Set", "Clustering: Hierarchical", "Clustering: Kmeans", "Clustering: SOM", "Correlation Summary Report", "Gene Ontology Summary Report", "Pathways Summary Report", "Save As a New Dataset", and "Show Spot Images". On the right side, there are checkboxes for "Show Gene Symbols" (checked) and "Show Spot Images" (unchecked). Other visible elements include a "Data:" label, a "2" in a box, and "Column Heading" text.

From the mAdb Dataset Display Page, select the “Additional Filtering Options” Tool and hit the “Proceed” button

# Filtering on data values

(Which genes are most differentially expressed?)

### Data Filtering Options

Check boxes on the left to activate specific filters

**Missing Value Filters**

Genes: Require values in  $\geq$   Arrays

Arrays: Require values in  $\geq$   % of Genes

**Gene Filters**

Ratio  $\geq$   in  $\geq$   Arrays   
 Apply Symmetrically

Ratio  $\geq$   in  $\geq$   Arrays  OR   
Ratio  $\leq$   in  $\geq$   Arrays

Average Ratio  $\geq$    
 Apply Symmetrically

Max (Ratio) / Min (Ratio)  $\geq$

Variance (Gene Vector) percentile  $\geq$   %

Subset Label:

1. Filter for at least 2-fold up in 2 or more arrays OR at least 2-fold down in 2 or more arrays.  
Other options are:
  - Filter **Ratio  $\geq 2$  in  $\geq 2$  Arrays**, with the **Apply Symmetrically** box checked to obtain genes up or down-regulated by 2-fold or more.
  - Filter for an average Ratio across the row at least two fold or more, applied symmetrically to obtain genes with an average ratio two-fold or more up or down regulated.
  - Filter for those rows showing a difference between the maximum ratio and minimum ratio on each row of 2 fold or more
  - Rank the genes by percentile of variance, and then filter for those genes in the top 10%ile of variance – ie. The genes that vary the most across the rows statistically.
  - N.B. Filters are applied in order from top to bottom – can iteratively access this tool to filter in your preferred order
2. Label the subset “2-fold up/down in 2 arrays”
3. Press the **Filter** button to continue and create the desired subset.

# Filtering by Feature Properties and/or Lists

(Are any of these genes in my “unreliable” list?)

### Feature Properties Filtering Options

Check boxes on the left to activate specific filters

Include only ▼ where Well ID is in hi std dev genes ▼

Subset Label: SUSPICIOUS - highly regulated genes

Filter

Filters any dataset so that only those identifiers matching feature properties in the selected list are included (or excluded)

Records 1 to 11 of 11 total records displayed.

A	A	A	A	A	↓ ↑	↓ ↑	↓ ↑
#1	#2	#3	#4	#5	Well ID	Feature ID	Gene
-1.9145	0.8009	1.5663	1.4701		616332	IMAGE:475503	Ube2i
	3.9956	-2.0067	-1.9103		613408	IMAGE:331681	Pcbd
	4.4465	-2.6394		-2.2019	616653	IMAGE:480196	
	2.0624	-0.5530	-1.2824	-1.6303	621320	IMAGE:790857	Tex261
	2.0528	-1.9512	-1.5359		620614	IMAGE:738252	Cebpa
	1.9644	-1.7476	-1.2831	-1.2732	613517	IMAGE:330336	H2-K1
1.6904	-0.4691	-1.6724	-1.2459	-2.0954	619489	IMAGE:949710	BC022146
	4.4261	-2.7721	-2.4251	-2.6466	613650	IMAGE:336497	Slc39a4
-1.3422	2.0174	-1.3657	-1.8360	-1.9434	621175	IMAGE:803488	Sfrp1
	2.3167	-1.5024	-2.1655	-2.3175	621404	IMAGE:777580	
	-1.2203	-3.0920	-3.7503		620919	IMAGE:750503	Sdpr

# More analysis tools

## Pathway Summary Report

Total number of features: 97

Total number of features mapped to a KEGG Pathway: 8

Total number of features mapped to a BioCarta Pathway: 5

Total number of features not mapped to any Pathway: 84

**NOTE:** Clicking on # of features creates a new subset containing only the features the mapped to the Pathway.

**NOTE:** Clicking on BioCarta Pathway ID displays the pathway.

# of Features	BioCarta Pathway	
1	<a href="#">m_cxcr4Pathway</a>	CXCR4 Signaling Pathway
1	<a href="#">m_ifngPathway</a>	IFN gamma Signaling Pathway
1	<a href="#">m_keratinocytePathway</a>	Keratinocyte Differentiation
1	<a href="#">m_etsPathway</a>	METS Affect on Macrophage Differentiation
1	<a href="#">m_ccr5Pathway</a>	Pertussis toxin-insensitive CCR5 Signaling in Macrophage
1	<a href="#">m_nktPathway</a>	Selective Expression of Chemokine Receptors during T-cell Polarization
1	<a href="#">m_malatePathway</a>	Shuttle for Transfer of Acetyl Groups from Mitochondria to the Cytosol
1	<a href="#">m_th1th2Pathway</a>	Th1/Th2 Differentiation
1	<a href="#">m_eea1Pathway</a>	The Role of FYVE-finger Proteins in Vesicle Transport

**NOTE:** Clicking on # of features creates a new subset containing only the features the mapped to the Pathway.

**NOTE:** Clicking on KEGG Pathway ID displays the pathway with features high lighted.

# of Features	KEGG Pathway	
2	<a href="#">mmu00561</a>	Glycerolipid metabolism
2	<a href="#">mmu00190</a>	Oxidative phosphorylation
1	<a href="#">mmu00193</a>	ATP synthesis
1	<a href="#">mmu00362</a>	Benzoate degradation via hydroxylation
1	<a href="#">mmu00710</a>	Carbon fixation
1	<a href="#">mmu00020</a>	Citrate cycle (TCA cycle)

From the mAdb Dataset Display Page, select “Pathways Summary Report”

1. Clicking on # of Features link creates a new dataset of just those features.
2. Clicking on BioCarta Pathway links show pathway on BioCarta Web site.
3. GO Ontology Summary Report also available <sup>87</sup>





# Ad Hoc Query Tool

Filtering/Grouping/Analysis Tools

Choose a Tool: Ad Hoc Query/Filtering Options and Proceed

Choose a View: and View

Retrieve Dataset for

Redisplay  Show Background Limiting 2

- Ad Hoc Query/Filtering Options
- Additional Filtering Options
- Ad Hoc Query/Filtering Options
- Feature Property Filtering Options
- Array Order Designation/Filtering
- Array Group Assignment/Filtering
- Filter/Group by Array Properties
- Average Arrays within Groups
- Two or more Group Comparison
- PAM: Prediction Analysis for Microarrays
- Boolean Comparison with another Set
- Clustering: Hierarchical
- Clustering: Kmeans
- Clustering: SOM
- Correlation Summary Report
- Gene Ontology Summary Report
- Pathways Summary Report
- Save As a New Dataset

From the mAdb Dataset Display Page, select the “Ad Hoc Query/Filtering Options” Tool and hit the “Proceed” button

## mAdb Ad Hoc Query

Check boxes on the left to activate additional Ad Hoc filters

1 Gene Description Contains receptor

2  and Chromosome Begins with 4

3

Subset Label: My Type A Ad Hoc Query - receptor & chr 4

Filter Cancel

Boolean Keyword search.

1. Pick from **BioCarta Pathway, Feature ID, Gene Description, Gene Symbol, GO term, KEGG Pathway, Map Location, UniGene ID, Well ID category**
2. Check box to add another term with **AND/OR** choice
3. Choose **Contains, Begins With, Equals, Does Not Contain, Does Not Begin With, Does Not Equal** for search qualifier

# Output of Ad Hoc Query

## mAdb Dataset Display

[View](#) Array Summaries

[Edit](#) Data for Subset: **My Type A Ad Hoc Query - receptor & chr 4**  
from Dataset: **test for class**

Ad Hoc Filtering

5 arrays and 340 genes in the input dataset

5 arrays and 2 genes in the output dataset.

Ad Hoc Filter:

Gene Description Contains 'receptor'

AND Chromosome Begins with '4'

Records 1 to 2 of 2 total records displayed.

A	A	A	A	A					
#1	#2	#3	#4	#5	Aver	Well ID	Feature ID	Map	Description
 3.2349	 2.7053	 3.0574	 2.8567	 3.3126	3.0334	614354	<a href="#">IMAGE:403453</a>	4 C6-D1	protein tyrosine phosphatase, receptor type, F
 -1.9201	 -2.4286	 -1.7173	 -1.9279	 -1.8618	-1.9711	620446	<a href="#">IMAGE:735186</a>	4 D2.3	nuclear receptor binding factor 1

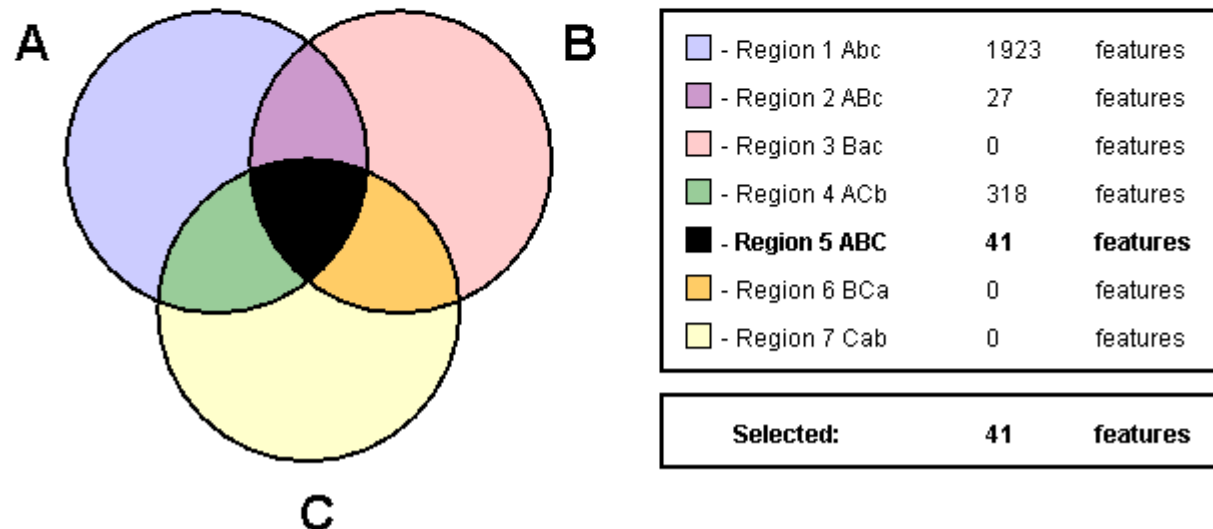
# Graphical Venn Tool

Compares subset intersections

## Boolean Comparison

	Label	Arrays	Genes	Created
Set A	Small, Round Blue Cell Tumors (SRBCTs),...	88	2309	Sep 18 2002 11:30:00am
Set B	PAM Threshold 3.928	63	68	Sep 20 2002 5:13:59pm
Set C	SAM Delta 0.800	10	359	Sep 10 2004 5:55:21pm

Click regions in the diagram to Select/Deselect



From the mAdb Dataset Display Page, select the “Boolean comparison using Venn Diagrams” Tool and hit the “Proceed” button

# Manually Create a List of Identifiers for Filtering

## mAdb Identifiers List Upload

This Form allows you to upload a list of Identifiers such as Clone, UniGene, Well ID. Uploaded lists are available as filter options in the "Feature Properties Filtering Tool".

Note; There is no need to specify the type of identifier in the "List Label". The system remembers each type of list presents your lists segregated and identified by type.

Type of List: Clone/Feature Identifier (IMAGE:12345, 12345\_at) ▼

List Label: Rab clones

Overwrite an existing list with the same label

Paste/Type in List:  
(One element/line)

IMAGE: 619501  
IMAGE: 466099  
IMAGE: 779604

Submit

Clone/Feature Identifier (IMAGE:12345, 12345\_at)  
Gene Symbol (BRCA1)  
LocusLink Identifier (12345)  
UniGene Identifier (Xx.1234)  
mAdb Well ID (12345)

From the mAdb Gateway page, use the "Upload Identifier list" link.  
Paste in list of identifier (use format as shown for specific type)

# Managing Feature Lists

## Manage Feature Identifier Lists

[Need Help?](#)

Check boxes to select Identifier lists to Delete

List (Click on a List to View/Edit)	List type
<input type="checkbox"/> <a href="#">Esther's list</a>	Clone
<input type="checkbox"/> <a href="#">my favorite genes</a>	Clone
<input type="checkbox"/> <a href="#">my interesting list</a>	Clone
<input type="checkbox"/> <a href="#">list of 340 genes 2x up down</a>	Gene
<input type="checkbox"/> <a href="#">receptors on chrom 5</a>	Gene
<input type="checkbox"/> <a href="#">oxidative phosph</a>	UniGene
<input type="checkbox"/> <a href="#">PAM-unigene</a>	UniGene
<input type="checkbox"/> <a href="#">mylist</a>	Well ID



## Feature Identifiers List

Esther's list formatted for

Type of List: **Clone**

Original List Label: **Esther's list**

List Label:

List Values:  
(1 item per line)

```
IMAGE: 697383  
IMAGE: 790571  
IMAGE: 920235  
IMAGE: 466099  
IMAGE: 316187  
IMAGE: 333232  
IMAGE: 762516  
IMAGE: 400592  
IMAGE: 467790  
IMAGE: 463386
```

List Value Order is maintained

From the mAdb Gateway page, use the “Manage Identifier list” link for existing feature lists. Click on list name to view/edit.

# VI. Managing your data

# Lab 4 – Dataset Management

Goal: To keep track of your analyses and share them with others.



## Accessing Temporary Datasets

1

**Manage** datasets located in your: [Temporary](#) or [Permanent](#) area

2

Switch to **accessing** datasets located in your: [Permanent](#) area

Temporary Datasets	Created	Containing Arrays	Genes	Need Help?	Gene Information Refreshed
<a href="#">Edit</a> hands-on qual filter	Dec 12 11:37:02am	5	5276	<a href="#">Open</a> <a href="#">Expand (1)</a> <a href="#">Refresh</a>	Dec 12 11:38:27am

3

4

5

### Dataset Access Links:

1. **Manage Transient, Temporary, or Permanent Areas**
2. **Access other dataset areas which contain data (i.e. Permanent)**
3. **Edit dataset name**
4. **Expand to see parent dataset and all children of that parent**
5. **Refresh Gene Information**

## Managing Temporary Datasets

**Access** datasets located in your: [Temporary](#) or [Permanent](#) area

Switch to **managing** datasets located in your: [Permanent](#) area

**Need Help?** 

**Check** boxes to select datasets for action

<input type="checkbox"/>	Temporary Datasets	Created	Containing Arrays	Genes	Gene Information Refreshed
<input checked="" type="checkbox"/>	hands-on qual filter	Dec 12 11:37:02am	5	5276	Dec 12 11:38:27am

Select an Action to perform on selected datasets

- Select an Action to perform on selected datasets
- Delete the selected datasets
- Move the selected datasets to your Permanent Area

[ad Status](#)  
[ads](#) | [GeneCards](#)

1

2

### Dataset Management:

1. Can delete a dataset – but must delete parent and all children!
2. Can promote datasets (Transient to Temporary or Permanent; Temporary to Permanent)

# Updating Dataset Gene Information

- Clicking the “refresh” link updates all of the gene information in the dataset (UniGene cluster, Description, Pathway info, Map info...)
- May want to “Save as a New Dataset”, and then refresh, if you want to keep previous annotation information

# Save as New Dataset

**mAdb Dataset Display**

[View](#) Array Summaries

[Edit](#) Data for Subset: **class 1/27 - 90%**  
from Dataset: **class 1/27 - 90%**

The filter input data  
The filtered output data  
3122 genes excluded from  
1814 genes excluded from

View the complete [History](#).

[Expand](#) this Dataset.  
Access Datasets in your [Ter](#)

Choose a Tool

- Additional Filtering Options
- Ad Hoc Query/Filtering Options
- Feature Property Filtering Options
- Array Order Designation/Filtering
- Array Group Assignment/Filtering
- Filter/Group by Array Properties
- Average Arrays within Groups
- Two or more Group Comparison
- PAM: Prediction Analysis for Microarrays
- Boolean Comparison with another Set
- Clustering: Hierarchical
- Clustering: Kmeans
- Clustering: SOM
- Correlation Summary Report
- Gene Ontology Summary Report
- Pathways Summary Report
- Save As a New Dataset**
- Additional Filtering Options

genes.  
genes.  
(4) arrays.  
t 80% (4) array(s)

and [Proceed](#)

At any time, researchers can save a subset as a new dataset. In effect, this starts the tree of subsets over again at the top...

# Sharing a Dataset

A 0.697 5. Mm-Incyte-v1p1-5 Sample 5/Type A  
A 0.537 4. Mm-Incyte-v1p1-4 Sample 4/Type A  
A 0.697 5. Mm-Incyte-v1p1-5 Sample 5/Type A

[Edit](#) Data for Subset: **80% present and 2 fold up and down**  
from Dataset: **spot filter for class**

The filter input data set contained 5 arrays and 8223 genes.  
The filtered output data set contains 5 arrays and 975 genes.  
2761 genes excluded for being present in less than 80% (4) arrays.  
4487 genes excluded by ratio  $\geq 2$  or  $\leq 0.50$  in at least 50% (3) array(s).

View the complete [History](#).

[Expand](#) this Dataset.  
Access Datasets in your [Temporary](#) area.

**NEW** [Post](#) a copy of this Dataset to other mAdb users.

[Filtering/Grouping/Analysis Tools](#)

Choose a Tool  and

---

[Interactive Graphical Viewers](#)

Choose a Viewer  and

At any time, researchers can place a snapshot of their entire dataset including their analysis steps to other users.

From the mAdb Dataset Display Page, click on the “Post” link

## Interactive Array Filtering

### Arrays Included

Mm-Incyte-v1p1-1 Sample 1/Type A  
Mm-Incyte-v1p1-2 Sample 2/Type A  
Mm-Incyte-v1p1-3 Sample 3/Type A  
Mm-Incyte-v1p1-4 Sample 4/Type A  
Mm-Incyte-v1p1-5 Sample 5/Type A  
Mm-Incyte-v1p1-6 Sample 1/Type B  
Mm-Incyte-v1p1-7 Sample 2/Type B  
Mm-Incyte-v1p1-8 Sample 3/Type B  
Mm-Incyte-v1p1-9 Sample 4/Type B

↑  
Change  
Array  
order.  
↓

↓ Remove or Add Back Arrays ↑

Mm-Incyte-v1p1-10 Sample 5/Type B

### Arrays Excluded

Subset Label:   
(Optional)

Filter

Cancel

**Change Array Order** by highlighting an array name and using the change array order up and down arrows.

**Remove/Add Arrays** by highlighting an array name and using the remove or add arrows  
Enter a label in the **Subset Label** field to have it attached to the resultant subset

Click the **Filter** button when finished or the **Cancel** button to return to the Data Display.

Allows re-ordering and removal of arrays from a subset

From the mAdb Dataset Display page, select the “Array Order Designation / Filtering” Tool and hit the “Proceed” button.

# Exporting Data to Other Microarray Analysis Tools

- BRB Array tools export by well ID or by UniGene ID
- GeneSpring export

**Extraction for BRBArrayTools**

---

**Data Format/Alignment Options**

Data Alignment :

**Array Selection**

	A	mAdbID: Array Name & Short Description
<input type="radio"/>	<input checked="" type="radio"/>	28733: Mm-Incyte-v1p1-1 Sample 1/Type A
<input type="radio"/>	<input checked="" type="radio"/>	28742: Mm-Incyte-v1p1-10 Sample 5/Type B
<input type="radio"/>	<input checked="" type="radio"/>	28734: Mm-Incyte-v1p1-2 Sample 2/Type A
<input type="radio"/>	<input checked="" type="radio"/>	28735: Mm-Incyte-v1p1-3 Sample 3/Type A
<input type="radio"/>	<input checked="" type="radio"/>	28736: Mm-Incyte-v1p1-4 Sample 4/Type A
<input type="radio"/>	<input checked="" type="radio"/>	28737: Mm-Incyte-v1p1-5 Sample 5/Type A
<input type="radio"/>	<input checked="" type="radio"/>	28738: Mm-Incyte-v1p1-6 Sample 1/Type B
<input type="radio"/>	<input checked="" type="radio"/>	28739: Mm-Incyte-v1p1-7 Sample 2/Type B

From the mAdb Gateway page, select a project(s) and the “BRBArraytools Format Retrieval” Tool and hit “Continue”

# Retrieving Uploaded Data

## mAdd: Data Retrieval Form

This tool allows you to retrieve the original uploaded data files.

### Upload Retrieval Options

Package Format: ZIP file (.zip)

Include:  Image Files (Spotted Uploads only)

Array Description Files

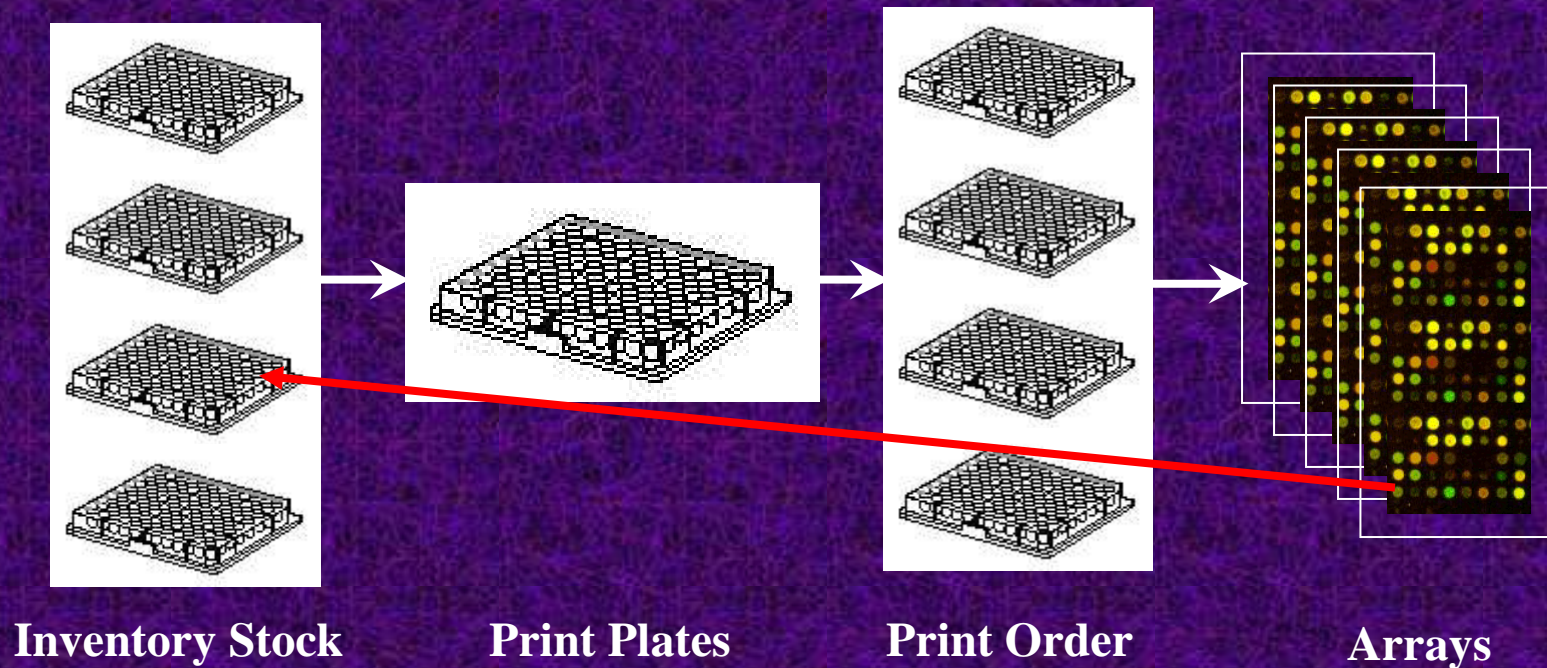
### Array Selection

	A	ID #	Array Name & Description
<input checked="" type="radio"/>	<input checked="" type="radio"/>	28733	Mm-Incyte-v1p1-1 Sample 1/Type A
<input checked="" type="radio"/>	<input checked="" type="radio"/>	28734	Mm-Incyte-v1p1-2 Sample 2/Type A
<input checked="" type="radio"/>	<input checked="" type="radio"/>	28735	Mm-Incyte-v1p1-3 Sample 3/Type A
<input checked="" type="radio"/>	<input checked="" type="radio"/>	28736	Mm-Incyte-v1p1-4 Sample 4/Type A
<input checked="" type="radio"/>	<input checked="" type="radio"/>	28737	Mm-Incyte-v1p1-5 Sample 5/Type A

From the mAdd Gateway page, select a project(s) and the “Uploaded Files Retrieval” Tool and hit “Continue”



# *mAdb* Database Design: Feature Tracking



- mAdb works with microarray facilities to track printing from arrays back to inventory plates
- Allows mAdb support staff to correct printing errors in the database

# mAdb Training/Reference Page

- Check [Java Version](#) running on your browser.

## mAdb Java Version Display

---

This page contains a Java Applet that attempts to display your Browser's Java Virtual Machine Version. If you see a pink box below and it contains a line of text - the Java Version indicated in the text is your Browser's version.

Java Version: 1.5.0\_06 from Sun Microsystems Inc.

- View Java version running on your browser

# Application Program Downloads

mAdb Program Downloads							
Page Updated: Friday, 15-Aug-2003 08:45:58 EDT							
	Program	Description	Author	Version	Updated	Download	Manual
<p><b>Axon Inc. Software</b></p> <p>This is commercial, licensed software and the GenePix application requires a "dongle" attached to the parallel port to run. The manual is accessible to all.</p> <p><a href="#">Axon's Web Site</a></p>	 <p><b>GenePix Pro 5</b></p>	<p>Fully integrated acquisition and analysis software for the GenePix 4000, 4100 &amp; 4200. Download to a folder of your choice and then run to start the installation process.</p>		<p>5.0.1.13 <a href="#">History</a></p>	<p>8/15/2003 (Posted here 8/15/2003)</p>	<p><a href="#">Download</a></p>	<p><a href="#">Users Guide &amp; Tutorial</a> (PDF)</p>
<p><b>Axon Inc. Software</b></p> <p>This is commercial, licensed software and the GenePix application requires a "dongle" attached to the parallel port to run. The manual is accessible to all.</p> <p><a href="#">Axon's Web Site</a></p>	 <p><b>GenePix Pro 4</b></p>	<p>Fully integrated acquisition and analysis software for the GenePix 4000 &amp; 4100. Download to a folder of your choice and then run to start the installation process.</p>		<p>4.0.1.17 <a href="#">History</a></p>	<p>(Posted here 3/12/2003)</p>	<p><a href="#">Download</a></p>	<p><a href="#">Manual</a> <a href="#">Axon Scanner Manual</a> (PDFs)</p>
<p><b>Axon Inc. Software</b></p> <p>This is commercial, licensed software and the GenePix application requires a "dongle" attached to the parallel port to run. The manual is accessible to all.</p> <p><a href="#">Axon's Web Site</a></p>	 <p><b>GenePix Pro 3</b></p>	<p>Fully integrated acquisition and analysis software for the GenePix 4000A. Download to a folder of your choice and then run to extract the installation files. Then run the extracted file setup.exe and follow installation instructions</p>		<p>3.0.6.89 <a href="#">History</a></p>	<p>(Posted here 02/18/2002)</p>	<p><a href="#">Download</a></p>	<p><a href="#">Manual</a> <a href="#">Axon Scanner Manual</a> (PDFs)</p>
<p><b>Stanford Genome Analysis Group Software</b></p> <p>It is available free of charge to academic and non-profit institutions.</p> <p><a href="#">Eisen Lab Download Site</a></p>	 <p><b>ScanAlyze</b></p>	<p>Image Analysis (extracts data from fluorescence images of arrays)</p>	<p><a href="#">Michael Eisen</a></p>	<p>2.44</p>	<p>11/15/99</p>	<p><a href="#">Download</a></p>	<p><a href="#">Manual</a> (PDF)</p>
	 <p><b>Cluster</b></p>	<p>Perform Hierarchical Clustering, Self-organizing Maps, k-Means Clustering, and More</p>	<p><a href="#">Michael Eisen</a></p>	<p>2.11.01</p>	<p>7/10/2000 (Posted here 10/26/2000)</p>	<p><a href="#">Download</a></p>	<p><a href="#">Manual</a> (PDF)</p>
	 <p><b>Tree View</b></p>	<p>Graphical Viewing and Browsing of Cluster Results</p>	<p><a href="#">Michael Eisen</a></p>	<p>1.5</p>	<p>04/2000 (Posted here 2/28/02)</p>	<p><a href="#">Download</a></p>	
<p><b>EASE: Expression Analysis Systematic Explorer</b></p> <p>Developed by the Laboratory of Immunopathogenesis and Bioinformatics, SAIC Frederick</p> <p><a href="#">EASE Web Site</a></p>	 <p><b>EASE</b></p>	<p>For finding "biological meaning" of gene lists via three functions: biological theme over-representation analysis, creation of annotation tables, and automated loading of genes into various online tools.</p>	<p><a href="#">Doug Hosack</a></p>	<p><a href="#">Revision history</a></p>	<p><a href="#">Current version</a></p>	<p><a href="#">Link to Download</a></p>	<p><a href="#">Online help</a> (Online)</p>
<p><b>MAExplorer</b></p> <p>Developed by and Available from LECB/FCRF/NCI.</p> <p><a href="#">MAExplorer Web Site</a></p>	 <p><b>MAExplorer</b></p>	<p>A Java data mining application for gene expression data using a variety of statistical, clustering, direct-manipulation graphical, spreadsheet and Web access methods.</p>	<p><a href="#">Peter Lemkin</a></p>	<p><a href="#">Revision History</a></p>	<p><a href="#">Current version</a></p>	<p><a href="#">Link to Download</a></p>	<p><a href="#">Manual</a> (Online) <a href="#">Use with mAdb data</a> (PDF)</p>

Various versions of GenePix are supported

Page accessible from NIH network only

Prefer GenePix updates obtained from this page – validated to work with mAdb

# Review of Basic Data Analysis Tools

- Within an extracted dataset, you can:
  - Filter for missing values and/or gene ratio levels
  - Do an *ad hoc* Keyword search
  - Filter datasets by lists of gene identifiers
  - View GO and Pathway Summaries
  - View data graphically
    - Interactive Scatter Plot
    - Correlation Summary Report
    - Multiple Array Viewer

# mAdb Tips for array analysis

- Always look at Project Summaries – Look for consistency across set of arrays. (Normalization factor for a “good” array should be between 0.5 and 2.0 if laser settings have been balanced. NOTE: Agilent scanners auto adjust settings, so normalization factors should just be consistent across set of arrays.)
- If you have replicate arrays (and you should), do a scatter plot or correlation summary report to determine the correlation between the arrays (i.e. how close the slope is to 1. For reverse fluors, how close to  $-1$ ) just for QC purposes.

# General tips for array analysis

At a recent Microarray Data Analysis conference in Washington D.C., several speakers laid out what distinguishes a good microarray experiment from a bad one:

- When possible, consult a statistician before you even design your experiment - they offer more than just analysis tools.
- Do a power analysis to determine the number of replicates (i.e. chips) you need to detect an effect. To estimate the effect size, you might want to run a pilot study first or obtain the estimate from previous similar experiments. Regardless of the power analysis results, obtain at least three replicates on different slides or chips.
- Find sources of technical variation before you embark on a hunt for biological effects and standardize your protocols.
- Randomize your variables: for example, don't run all your treatment slides on one day and all your controls on the next.
- Microarray analysis is a screening tool – confirm your observation by other methods – RT-PCR, Northern blot, protein levels
- See <http://linus.nci.nih.gov/~brb/TechReport.htm> for good references on design, analysis issues, and myths/truths

## Other microarray training

- Hands-on analysis tool mAdb class #412 – October 29-30
- Statistical Analysis of Microarray Data & BRB Array Tools (from the NCI Biometrics Research Branch) class #410; October 7-8
- Statistical Analysis of Microarray Data using The MSCL Analyst's Toolbox and JMP #423; TBA
- Partek Pro, R, GeneSpring classes and other “Seminars for Scientists” – <http://training.cit.nih.gov>
- Microarray Interest Group
  - 1<sup>st</sup> Wed. Seminar, 3<sup>rd</sup> Thu. Journal Club
  - To sign up: <http://list.nih.gov/archives/microarray-user-l.html>
- Class slides available on “Reference” page
- Sample datasets to try out the system are available from a link on the Gateway Page

### Uploading Links

- [Upload](#) Array data
- [Status](#) of Uploads
- [Upload](#) Identifier lists

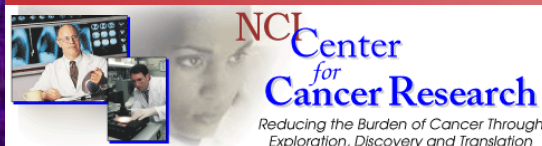


[Access](#) Training/Public Datasets

## mAdb Development and Support Team:

- John Powell, Chief, BIMAS, CIT
- Lynn Young, Ph.D.
- Wenming Xiao, Ph.D.
- Esther Asaki\*
- Yiwen He, Ph.D.\*
- Kathleen Meyer\*

\*SRA International contractor





<http://madb.nci.nih.gov>  
<http://madb.niaid.nih.gov>

**For assistance, remember:**

**[madb\\_support@bimas.cit.nih.gov](mailto:madb_support@bimas.cit.nih.gov)**

**Thank you!!**

