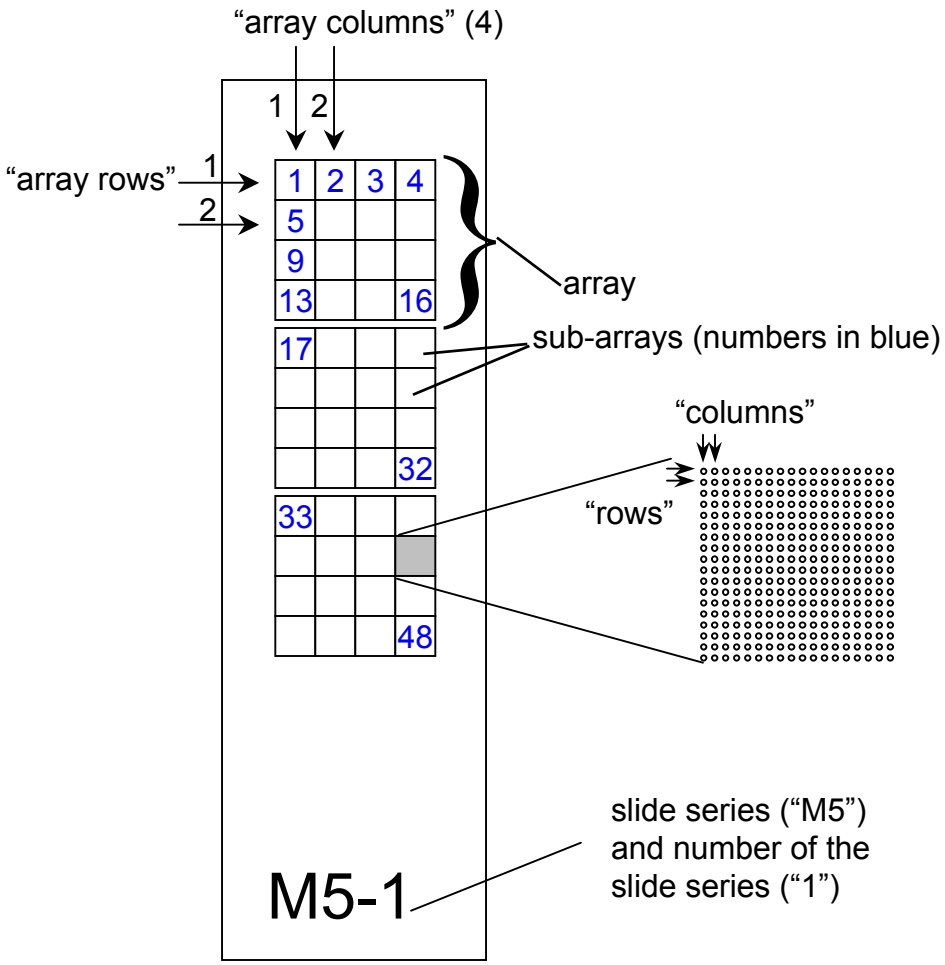
Part I:

How is the microarray slide set up?

How do I equalize dye channels in a microarray slide in ScanArray?

How do I generate raw data using QuantArray?

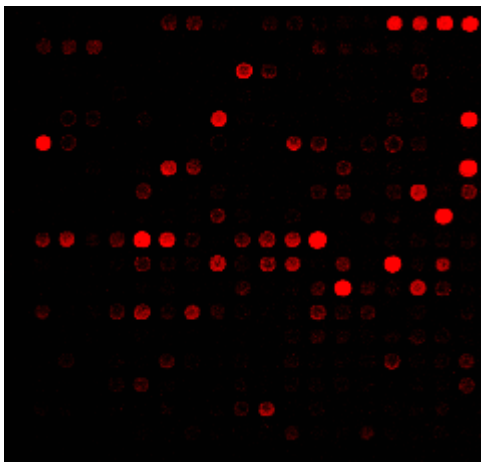
Microarray Slide Layout



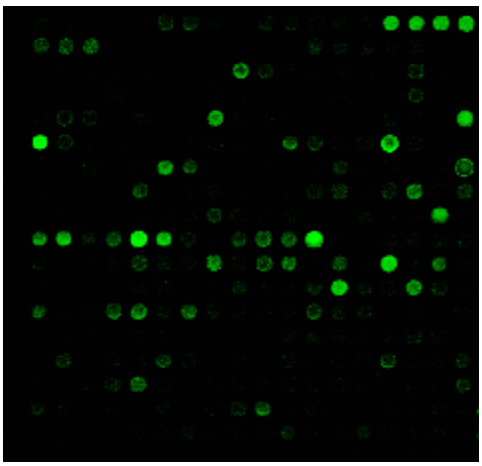
- Spots are also given a unique number or index. Sub-arrays are numbered first by row, then column, in the order of sub-arrays given on the diagram.
- We use 16 pins to print the microarrays; each pin corresponds to a single sub-array
- Mouse and human oligo slides have 3 arrays; hypothalamus slides have two duplicate arrays

Microarray Images

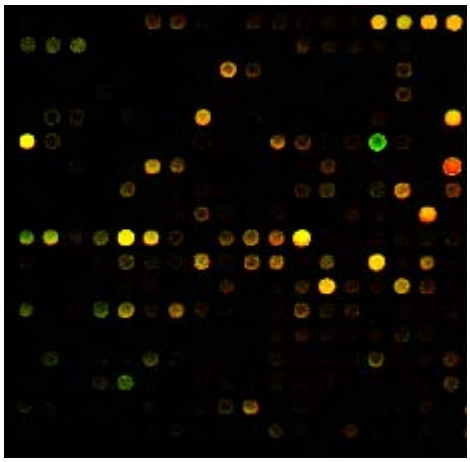
Cy5 (biotin; red)



Cy3 (fluorescein; green)

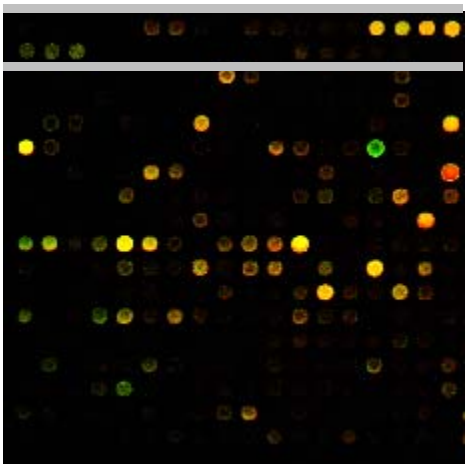


Composite image



- Note: QuantArray (QA) assigns the name “Channel 1” or “Channel 2” to whichever image (Cy3 or Cy5) was opened first. This can be manually changed by right-clicking one of the opened images in QA and choosing “set as control image”. Channel is synonymous to dye, but is an arbitrary parameter and may change from slide to slide.

Sub-array Properties



Controls are located in the first two rows of a sub-array. Control spots presently include: luciferase, pooled mouse oligo, blanks, positive and negative controls, and salmon sperm DNA

- The number of rows in a sub-array are 18 (mouse oligo), 20 (human oligo), or 20 (hypothalamus)
- The number of columns in a sub-array is 18 (mouse oligo), 22 (human oligo), or 22 (hypothalamus)

Step I-a: Scan in the microarray slide (ScanArray)

1. Turn on the Scanner using the button on the right-hand side of the instrument.
2. The first window to pop up after double-clicking on the ScanArray Icon is shown in Figure 1. Double-click on the “SA 5000” icon. When the two lights on the scanner are green and not flashing, turn on lasers 1 and 3 by pressing the buttons in the top of the window. It takes 15 minutes for the lasers to warm up.

3. Place a slide in the scanner and press the insert slide button. 
4. Press the play button to begin.  Alternatively, you can go to the <Acquire> heading and select “Start”.
5. The next window to pop up is shown in Figure 2. If you do not have your own protocol, then press the <New> button. If the protocol is new, you will have to <Add> the Cy3 and Cy5 fluorophores to the “Fluorophore Selection” grouping. Make sure you save the protocol by pressing <SaveAs>.
6. In the “Acquisition Area” grouping, use the mouse to select the region of the slide that you want to scan. You can also type values into the boxes.
7. Choose Quick Scan for the initial scan. Note: if you want to change the laser power or gain for the initial scan, press the <Settings…> button. A suggested laser power and PMT gain to start is 80%.
8. Press <Acquire> for the scanning to begin. Note that you can stop the scan at any time by pressing the Stop button. 

Figure 1

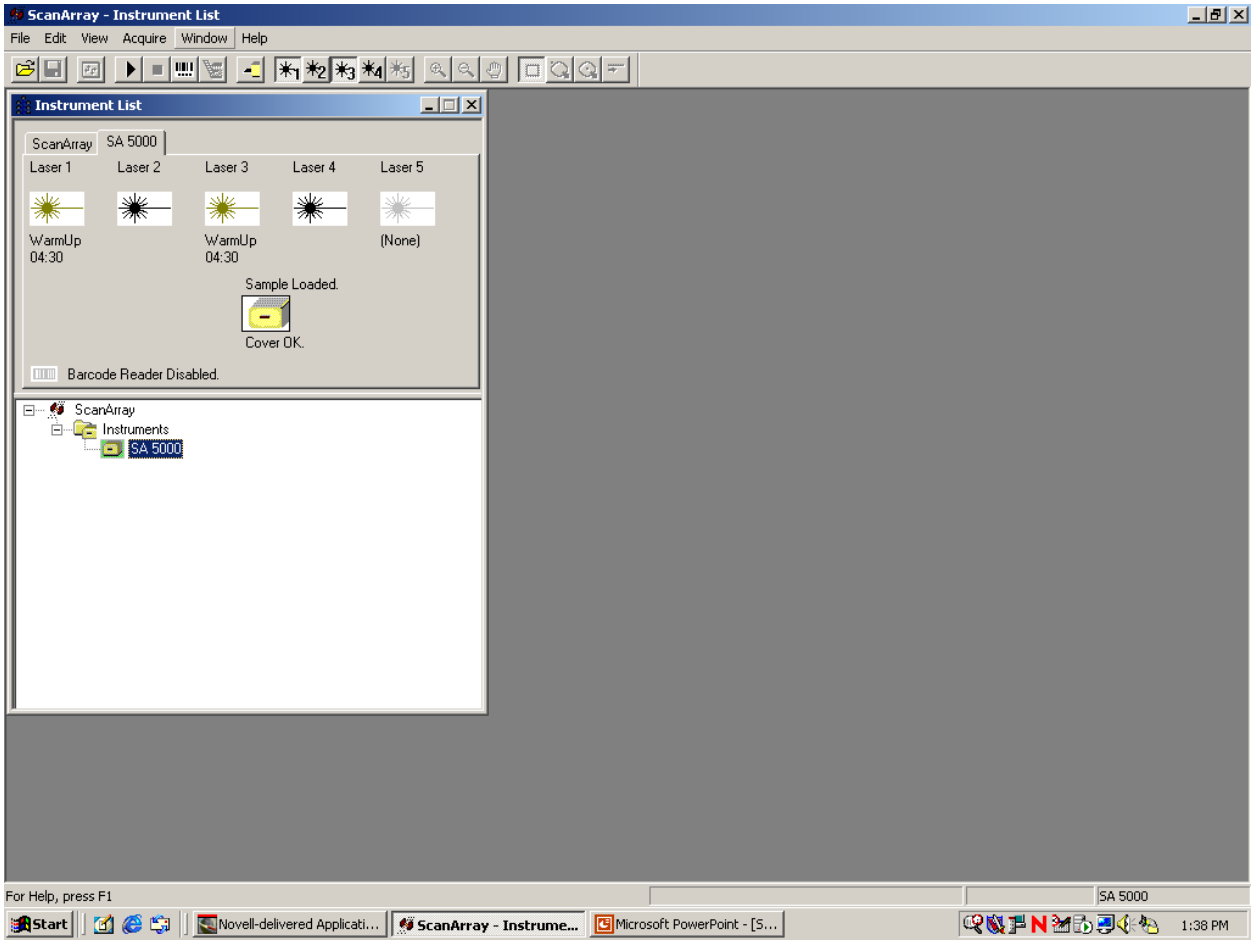
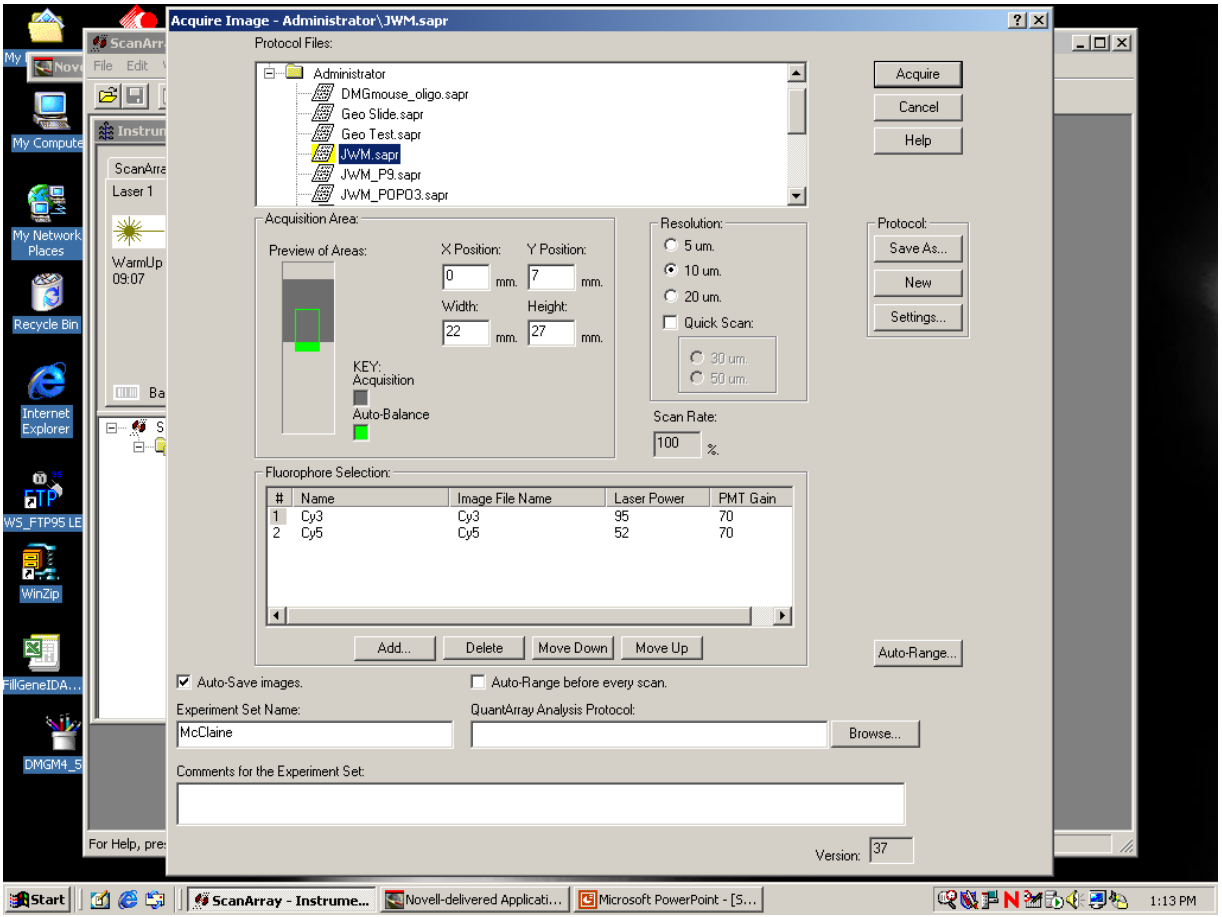
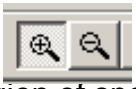

ScanArray - Instrument List
File Edit View Acquire Window Help
Instrument List
ScanArray
SA 5000
Laser 1
Laser 2
Laser 3
Laser 4
Laser 5
WarmUp
04:30
WarmUp
04:30
(None)
Sample Loaded.
Cover OK.
Barcode Reader Disabled.
ScanArray
Instruments
SA 5000
For Help, press F1
SA 5000
Start
Novell-delivered Applicati...
ScanArray - Instrume...
Microsoft PowerPoint - [S...
1:38 PM

Figure 2


Acquire Image - Administrator\JWM.sapr
Protocol Files:
Administrator
DMGmouse_oligo.sapr
Geo Slide.sapr
Geo Test.sapr
JWM.sapr
JWM_P9.sapr
JWM_POPO3.sapr
Acquire
Cancel
Help
Acquisition Area:
Preview of Areas:
X Position:
Y Position:
Width:
Height:
KEY:
Acquisition
Auto-Balance
Resolution:
5 um.
10 um.
20 um.
Quick Scan:
30 um.
50 um.
Scan Rate:
Protocol:
Save As...
New
Settings...
Fluorophore Selection:
Name Image File Name Laser Power PMT Gain
1 Cy3 Cy3 95 70
2 Cy5 Cy5 52 70
Add...
Delete
Move Down
Move Up
Auto-Range...
Auto-Save images.
Auto-Range before every scan.
Experiment Set Name:
McClaine
QuantArray Analysis Protocol:
Browse...
Comments for the Experiment Set:
Version:
37
Start
ScanArray - Instrume...
Novell-delivered Applicati...
Microsoft PowerPoint - [S...
1:13 PM

I-b: Equalize the fluorophores

9. Figure 3 shows the Cy3 image that is scanned in. Use the Zoom in and/or Zoom out buttons to make this image as magnified as you would like. 

10. Select a region of spots that you would like to use to match the channels (4-8 spots). Press the line scan button 

    and draw a line in the center of the spots. Note that you can only move upwards from where the line is first drawn, so start low. Once you let go of the mouse-click, the window in Figure 4 is shown.

11. Select one of the dyes by clicking on the # 1 or 2. Then press the <Start Scan> button. Change the focus by adjusting the position using your mouse until the “Current Average Signal” is maximized. Adjustments of the focus should always be done before changing laser power or gain (for both dyes), and only has to be done once (for a given slide).

    NOTE: Performing this step can begin to bleach out the fluorescence in the region being scanned. You can safely run “line scan” ~10 times in the same region, if you are still having problems balancing the channels, you may want to consider scanning a different region.

12. Adjust the laser power (and possibly the PMT Gain) until the “Current Average Signal” is the same for both dyes. Note: if possible (meaning if the background is low), try to get your spot intensity to be between 50 and 75 % on the y-axis when using ‘control spots’ for the line scan. When using spots other than controls, your spot intensity may be significantly lower than 50-75%. You do not want to have too many white spots, because this indicates that the intensities are at a maximum.

13. You may want to scan the slide in at more than one intensity to see the affect on results. For example, if spots within a particular region are light, you may want to do a line scan in that region and scan the slide, then perform a second line scan in a region where spots are darker and perform a second scan.

I-c: Perform the final scan

14. Once you have equalized the strength of the dyes, then press <OK>, and <OK> on the next screen that comes up. Your changes will be reflected in your protocol (Figure 2).
15. Change the resolution to 10 microns.
16. Make sure the entire region of the slide containing the array is selected in the “Preview of Area” grouping.
17. Press <Acquire> to begin scanning.
18. When the two fluorophores have been scanned in, click on each scan and go to the “File” pulldown menu and click on “Save As” to save each of your scans in your folder. Save your scans as both a TIFF and BITMAP file.

Figure 3

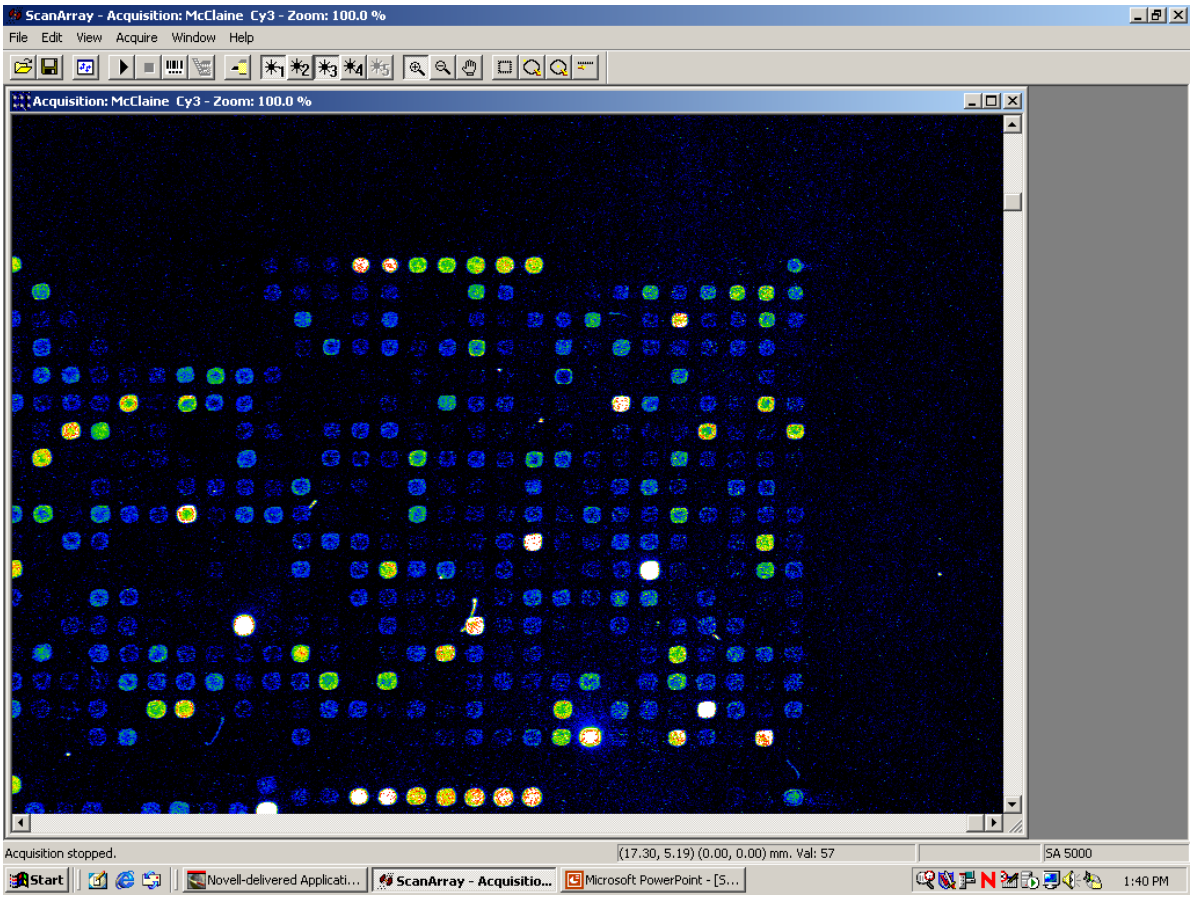
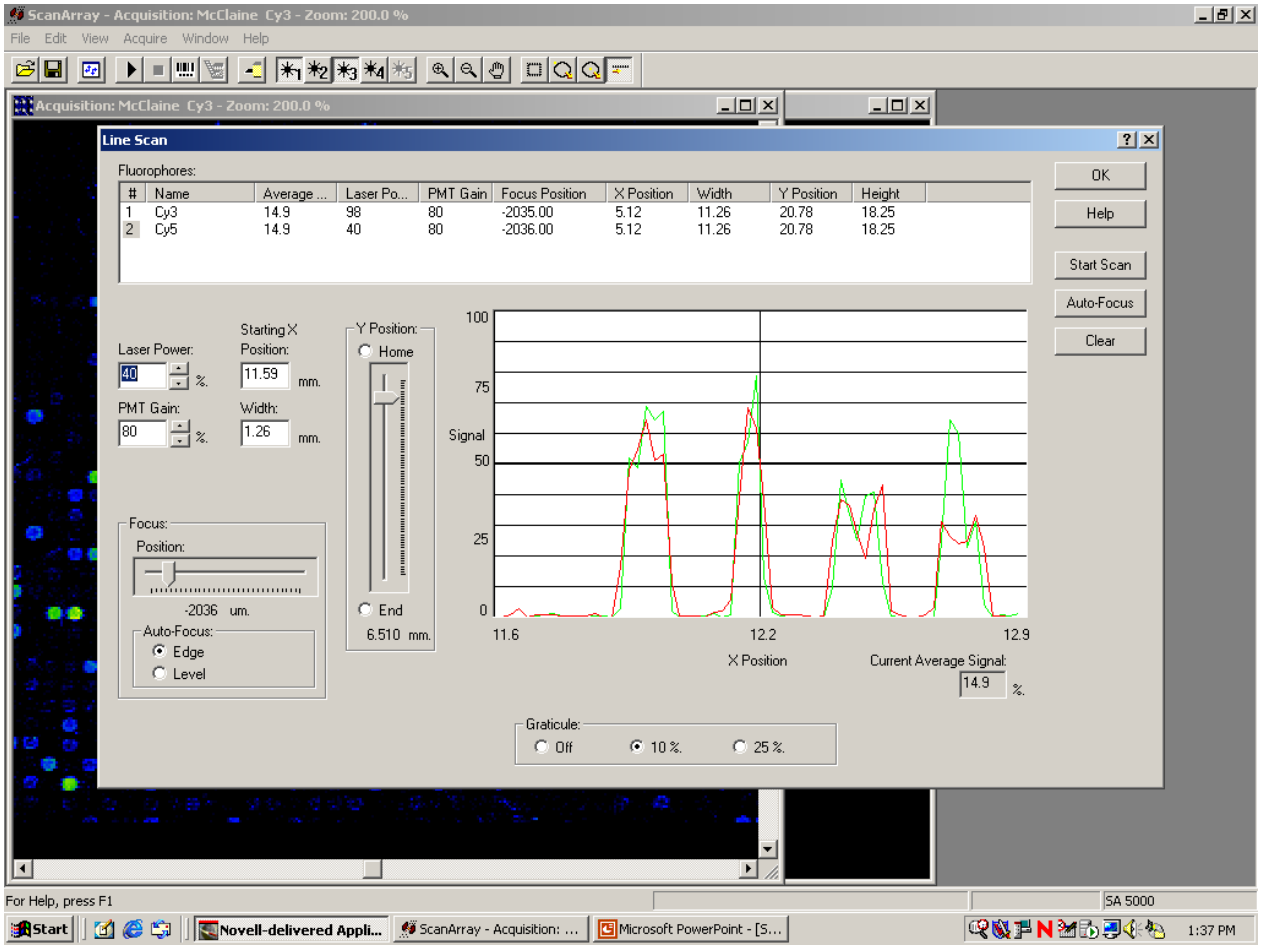

ScanArray - Acquisition: McClaine Cy3 - Zoom: 100.0 %
File Edit View Acquire Window Help
Acquisition: McClaine Cy3 - Zoom: 100.0 %
Acquisition stopped.
(17.30, 5.19) (0.00, 0.00) mm. Val: 57
SA 5000
Start
Novell-delivered Applicati...
ScanArray - Acquisitio...
Microsoft PowerPoint - [S...
1:40 PM

Figure 4

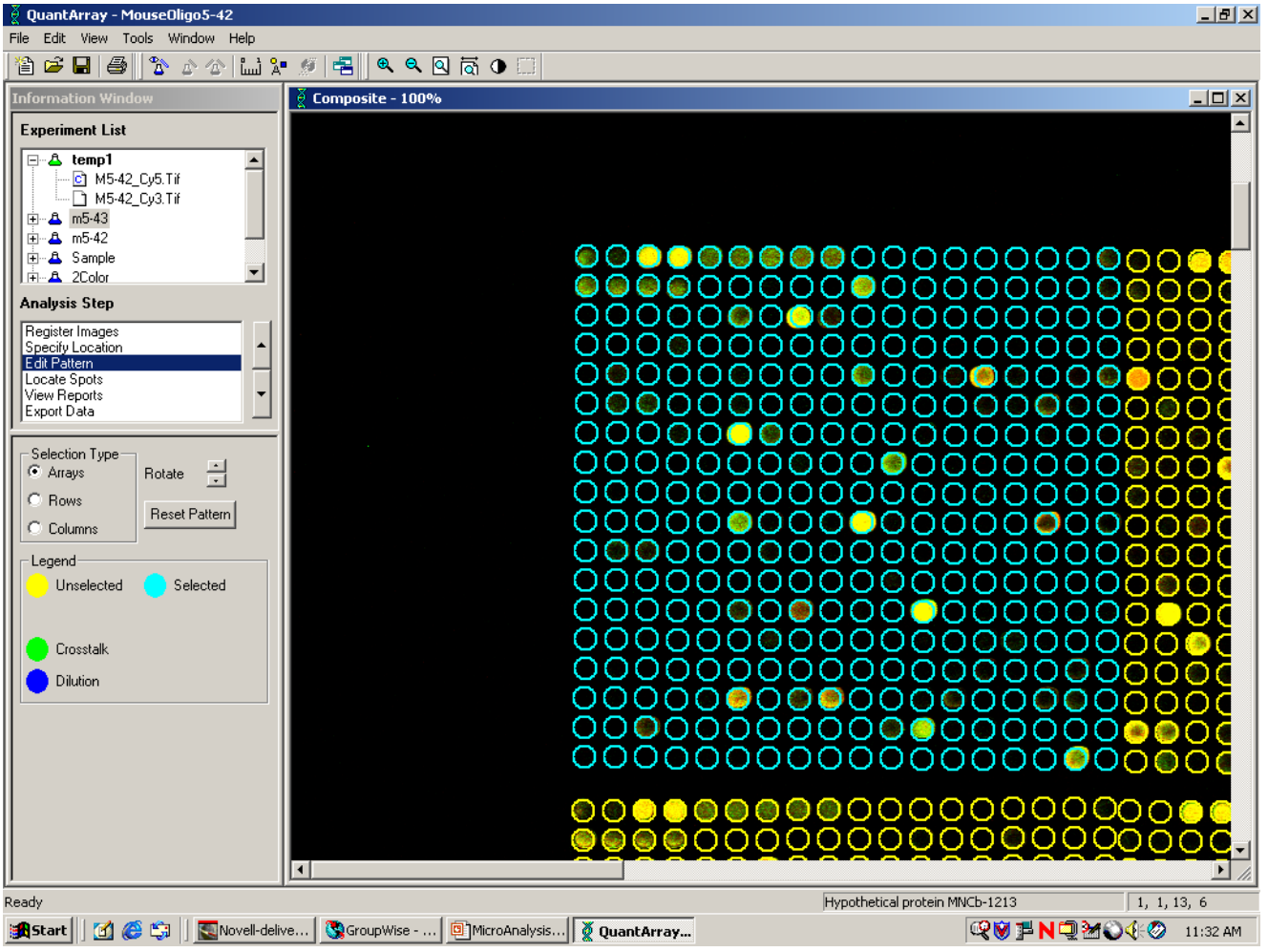

ScanArray - Acquisition: McClaine Cy3 - Zoom: 200.0 %
File Edit View Acquire Window Help
Acquisition: McClaine Cy3 - Zoom: 200.0 %
Line Scan
Fluorophores:
Name Average ... Laser Po... PMT Gain Focus Position X Position Width Y Position Height
1 Cy3 14.9 98 80 -2035.00 5.12 11.26 20.78 18.25
2 Cy5 14.9 40 80 -2036.00 5.12 11.26 20.78 18.25
OK
Help
Start Scan
Auto-Focus
Clear
Laser Power:
40 %.
Starting X Position:
11.59 mm.
PMT Gain:
80 %.
Width:
1.26 mm.
Y Position:
Home
End
6.510 mm.
Focus:
Position:
-2036 um.
Auto-Focus:
Edge
Level
100
75
Signal
50
25
0
11.6
12.2
12.9
X Position
Current Average Signal:
14.9 %.
Graticule:
Off
10 %.
25 %.
For Help, press F1
SA 5000
Start
Novell-delivered Appli...
ScanArray - Acquisition: ...
Microsoft PowerPoint - [S...
1:37 PM

Step I-d: Use QuantArray to Locate Spots

1. Double-click on the QuantArray icon. Go to “File”, “Open protocol”, and open the appropriate protocol. Note: The protocols are created after printing for each slide series and are located in the “QuantArray protocols” folder within the “McroArry” folder on the T drive. (They can also be copied onto your computer if you don’t yet have access to this folder.)
2. In the “File” menu, select “Open Image” and open both your Cy3 and Cy5 images for a given slide. You will have to give a unique name to this experiment when the window pops up. It is a good idea to have the slide series and number in the name of the experiment. Click <OK>.
3. Go to the “View” menu and select “Composite Image”. This allows you to see the composite image of both Cy3 and Cy5 scans.
4. Use the Zoom in and Zoom out buttons to change the magnification of the slide. 
5. In the “Analysis Step” grouping, choose “Specify Location”. Put the crosshairs in the center of the upper left-most spot in the upper left-most sub-array of the slide and press the mouse click. Note: you will have to turn the zoom buttons off before this can be done.
6. Choose “Edit Pattern” in the “Analysis Step” grouping. This will overlay a grid pattern on your slide. If you put the cursor on a spot the gene name will be given in the bottom right of the window (Figure 5). The array row, array column, row and column are also given (in this order) in the bottom right corner of the window.
7. Use the contrast key  to enhance the contrast of the slide.

   This will make it easier to match the grid up to the spots.

Figure 5

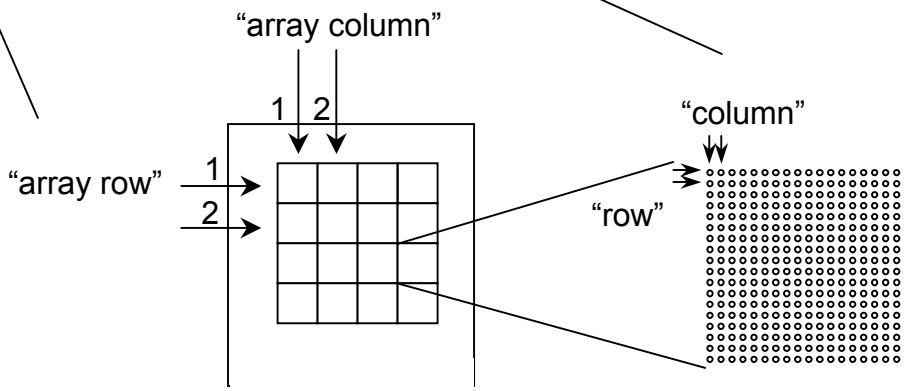

QuantArray - MouseOligo5-42
File Edit View Tools Window Help
Information Window
Experiment List
temp1
M5-42_Cy5.Tif
M5-42_Cy3.Tif
m5-43
m5-42
Sample
2Color
Analysis Step
Register Images
Specify Location
Edit Pattern
Locate Spots
View Reports
Export Data
Selection Type
Arrays
Rows
Columns
Rotate
Reset Pattern
Legend
Unselected
Selected
Crosstalk
Dilution
Composite - 100%
Ready
Hypothetical protein MNCb-1213
1, 1, 13, 6
Start
Novell-delive...
GroupWise - ...
MicroAnalysis...
QuantArray...
11:32 AM

QuantArray continued

8. Use the mouse or array keys to match the grid to the spots within a particular sub-array. You can choose to move arrays, rows, or columns of a sub-array. We recommend using that you move the entire sub-array. To do this, select “Arrays” in the Selection Type” grouping. The selected array will appear blue (and the rest of the grid is yellow). You may also have to use the rotate keys: up corresponds to counterclockwise and down corresponds to clockwise.
   NOTE: There may be one or two spots that just do not fit into the grid. these can be manually moved later after you run “Locate Spots”
9. Save a protocol for each of your slides (with the slide series and number in the name), so that if you ever have to go back to this stage, you don’t have to realign your spots.
10. When all sub-arrays are aligned with the grid, choose “Locate Spots” and select “Use Nominal Locations”. The press <Start Locate>. Check your spot locations to make sure that the cross hairs are in the center of the spot (or very near the center). If you need to move a spot, click and drag the square until the cross hairs are now in the center of the spot.
11. Select “Export Data”, and press the <Export> button. Note: this often gives an error message for an unknown reason, and may have to be done more than once before the file is successfully saved.
12. Immediately afterwards, click <Open Last Exported File With Excel>. Make sure that you save this file into a known folder as an Excel workbook. It is a good idea to have the slide series and number in the name of the file.

QuantArray Excel Output

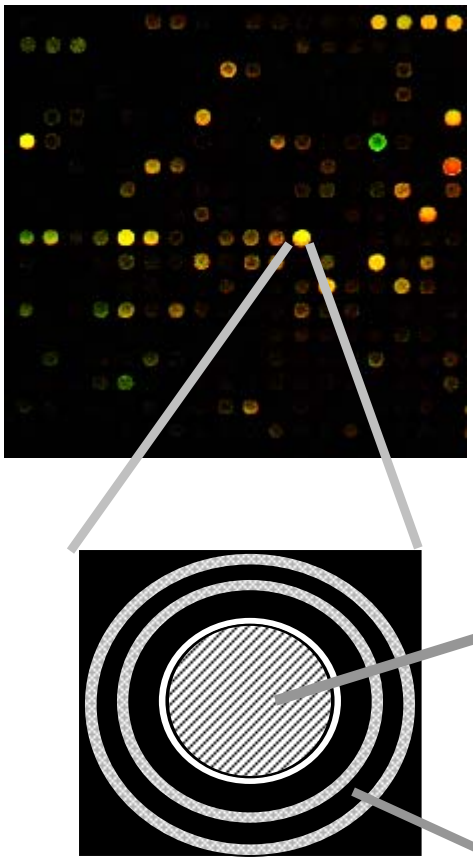
Begin Data											
Number	Array Row	Array Column	Row	Column	Name		ch1 Intensity	ch1 Background		ch2 Intensity	ch2 Background
6481	6	1	1	1	Myosin, heavy polypeptide 2, skeletal muscle, adult		424.26	0.805556		1200.61	0
6482	6	1	1	2	Midline 1		503.63	0		290.06	0
6483	6	1	1	3	Vomeronasal organ family 2, receptor, 15		148.02	0		43.63	0
6484	6	1	1	4	Nuclear factor of activated T-cells 5		169.31	0		44.48	0
6485	6	1	1	5	Carbonic anhydrase 14		469.02	0		210.27	0



Part II:

How is my data normalized using the MANv1.0 (*M*icro*A*rray *N*ormalization) web interface?

Step II: Normalization of data: ***II-a*** subtract background



Spot intensity is measured for pixels within a specified diameter.

Background intensity is measured for pixels between a specified inner and outer diameter.

The mean (or median) background is subtracted from the mean (or median) spot intensity (default output from QA is the mean value; the protocol can be changed to give the median). This will give a corrected spot intensity (“corrected” for background).

Step II-b: normalize to control (PO or Luc) for each sub-array

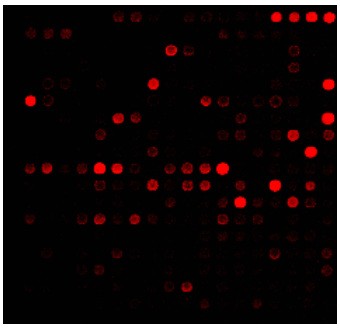
- Calculate the trimmed mean or median value of control spots (usually pooled mouse oligo or luciferase) for the two channels (Cy3 and Cy5) in each sub-array
- For each gene within a given sub-array, divide corrected channel 1 intensity by the mean or median control channel 1 intensity in that sub-array
- For each gene within a given sub-array, divide corrected channel 2 intensity by the mean or median control channel 2 intensity in that sub-array
- Note: In the McroArray folder on the T drive, there is a sample file with all normalization calculations if you want to test the calculations by hand.

Step II-c: remove unwanted genes

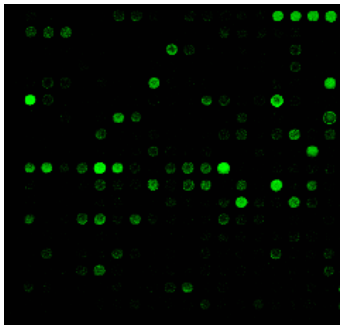
- Remove salmon sperm DNA, blanks, luciferase, and negative and positive controls. These controls will skew average intensity values.

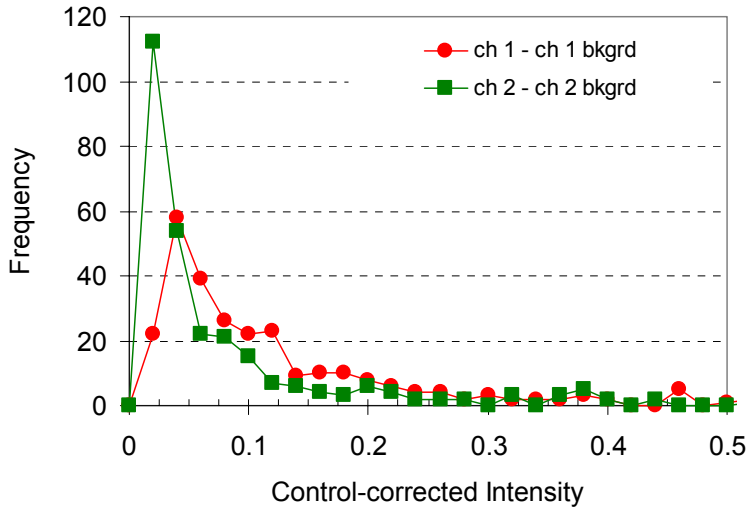
Step II-d: normalize to trimmed mean or median intensity per sub-array

Cy5 (biotin)



Cy3 (fluorescein)



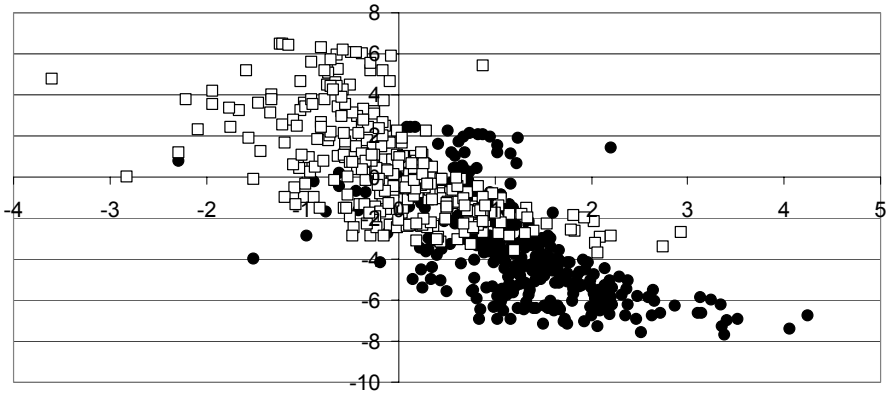


- The mean (or median) for the two channels is usually slightly different, even after equalizing channels in ScanArray
- Because the frequency distribution is not symmetric, the trimmed mean and median values for a given channel will be different.

Step II-d in more detail

- Calculate trimmed mean or median value of ch 1 and ch 2 control-corrected intensities in each sub-array
- For each gene within a given sub-array, divide control-corrected channel 1 intensity by the mean or median channel 1 intensity in that sub-array
- For each gene within a given sub-array, divide control-corrected channel 2 intensity by the mean or median channel 2 intensity in that sub-array

MA plot: x-axis $(1/2)\log_2(Cy5*Cy3)$
y-axis $\log_2(Cy5/Cy3)$



- control-corrected intensities
- control- and median intensity-corrected

After normalization, the two channels have approximately equal means or medians, and the median intensity ratio is shifted closer to zero (on log scale)

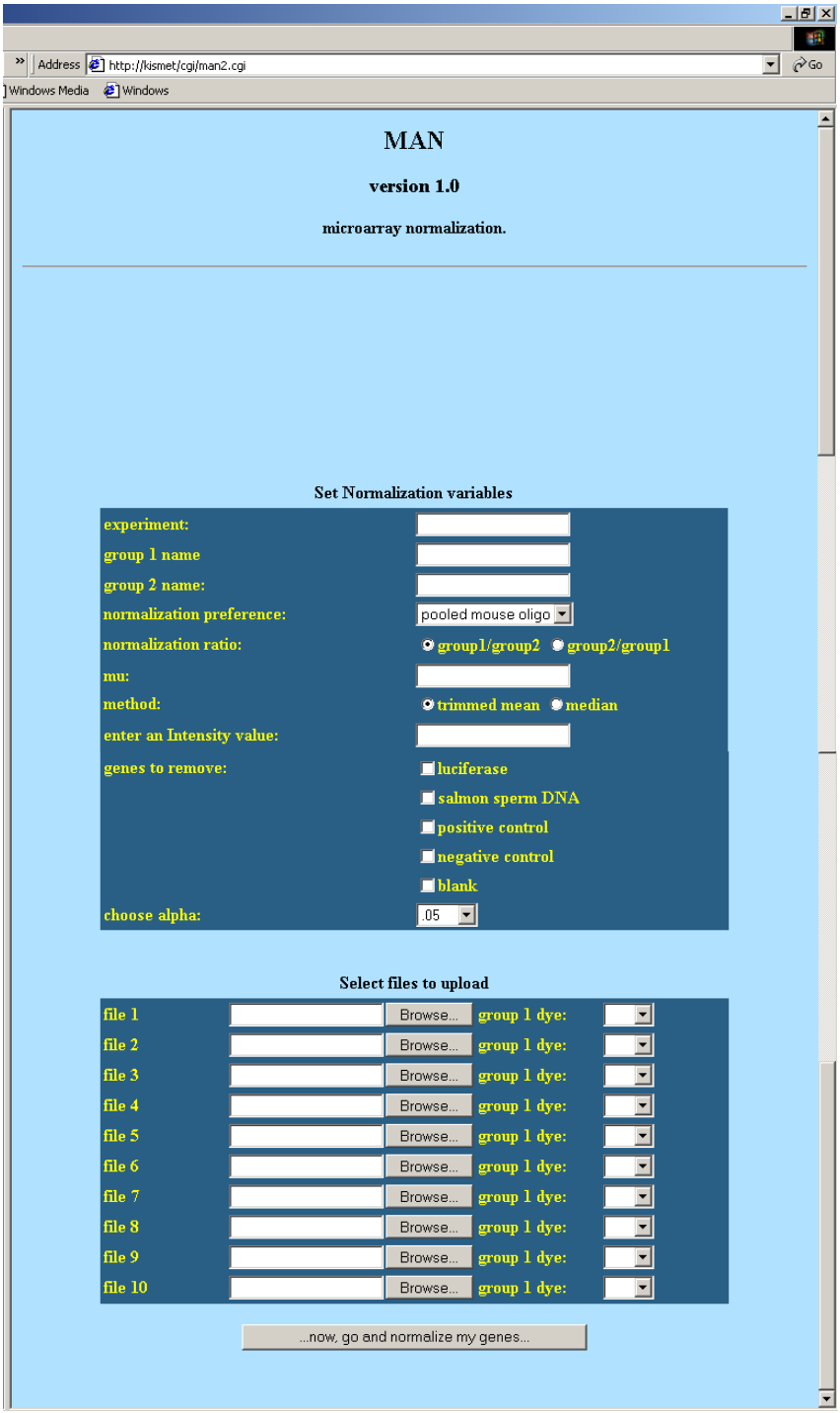
Step II-e: calculate normalized ratio

- Divide control- and intensity-corrected ch 1 values with control- and intensity-corrected ch 2 values (or vice versa, depending on which groups were labeled with what dye) to get the intensity ratio.

Step II-f: perform relevant statistical calculations and tests

- Convert to log, base 2
- After all slides have been processed, calculate a mean and standard deviation for each gene.
- Perform statistics (t-test, confidence interval); user must define α for the confidence interval calculation.

*M*icro*A*rray *N*ormalization (*MAN*) user interface:



Required input from user

- Name of experiment
- The name of the two groups tested in this set of microarrays. For example: control group and treatment group
- Choose the reference spots used for normalization, either Luciferase or pooled oligo for the mouse oligo slides.
- Choose how you want the output shown, group 1 / group 2 or vice versa
- Enter a μ value. This is a value that you want to test your average ratio against in a t-test to see if it is significantly different. For now, μ should be equal to 0 (because our values are log base 2 and 1 to the log base 2 is equal to 0).
- Choose whether to normalize to trimmed mean or median
- Enter a minimum intensity value, such that intensity values lower than this are excluded from calculation.
- Check which genes to remove.
- Choose α value for confidence intervals.
- Upload files; current capability is 10 files.
- For each slide, choose the dye that was used to label group one. Remember, Cy3 is fluorescein and Cy5 is biotin.
- Press the <now, go and normalize my genes> button to begin.

Sample Output

RATIO 1	RATIO 2	RATIO 3	RATIO 4	RATIO 5	RATIO 6	GENE	SUB-ARRAY	NUMBER	MEAN	D.F.	S.D.	T-TEST	P-VALUE	T-Critical	CI left	CI right
0.45	0.64	0.39	0.85	0.88	NA	riken cdna 15000(	43	13870	0.64	4	0.22	6.45	0.00	2.78	0.37	0.92
1.13	1.09	-0.19	3.76	3.04	NA	riken cdna 12000	47	14956	1.76	4	1.60	2.46	0.07	2.78	-0.23	3.76
0.25	1.10	-0.58	-0.01	0.13	NA	lymphocyte antige	3	689	0.18	4	0.61	0.66	0.54	2.78	-0.57	0.93
-1.96	0.16	-0.78	-1.26	-1.93	NA	karyopherin (impo	16	5022	-1.16	4	0.89	-2.92	0.04	2.78	-2.26	-0.06
0.01	0.37	NA	NA	NA	NA	riken cdna 23100	26	8295	0.19	1	0.25	1.07	0.48	12.71	-2.07	2.45
0.06	1.01	-0.32	1.28	0.47	1.64	h6 homeo box 1	14	4387	0.69	5	0.75	2.26	0.07	2.57	-0.10	1.48
-1.15	0.27	NA	NA	NA	NA	riken cdna 28104(	27	8745	-0.44	1	1.00	-0.62	0.65	12.71	-9.44	8.57
-1.70	0.15	0.42	-1.38	-3.35	NA	riken cdna 20100(	37	11775	-1.17	4	1.53	-1.72	0.16	2.78	-3.07	0.73
-0.37	-0.13	-1.25	-1.07	0.39	0.28	catalase 1	14	4325	-0.36	5	0.68	-1.29	0.25	2.57	-1.07	0.36
-0.38	-0.46	-0.90	0.33	NA	NA	riken cdna 31100	24	7658	-0.35	3	0.51	-1.37	0.26	3.18	-1.16	0.46
-1.40	-0.57	-0.70	-0.28	-0.73	-2.09	alpha-2-hs-glycop	14	4379	-0.96	5	0.66	-3.54	0.02	2.57	-1.66	-0.26
1.97	1.46	0.76	0.65	1.18	2.15	myxovirus (influen	5	1611	1.36	5	0.62	5.40	0.00	2.57	0.71	2.01
-0.25	0.13	0.35	-0.08	0.78	0.79	cyclin g2	1	101	0.29	5	0.43	1.62	0.17	2.57	-0.17	0.74
0.27	0.05	0.32	-0.34	0.77	NA	expressed in non-	10	3109	0.21	4	0.41	1.18	0.30	2.78	-0.29	0.72
-0.73	-1.40	NA	NA	NA	NA	urocortin 2	27	8517	-1.06	1	0.47	-3.19	0.19	12.71	-5.30	3.17
0.20	0.28	0.56	0.56	-1.27	NA	riken cdna 39304(	22	7022	0.06	4	0.76	0.19	0.86	2.78	-0.88	1.01
-0.47	0.00	-0.58	-0.12	1.05	NA	found in inflammat	18	5634	-0.02	4	0.65	-0.07	0.94	2.78	-0.82	0.78
-0.56	0.07	-1.09	-0.58	-1.08	NA	riken cdna 20100(	42	13336	-0.65	4	0.48	-3.04	0.04	2.78	-1.24	-0.06
0.27	0.12	-0.37	-0.30	-1.23	NA	keratin complex 2,	22	6992	-0.30	4	0.59	-1.14	0.32	2.78	-1.03	0.43
1.94	1.12	2.12	1.54	1.94	1.83	midkine	4	1015	1.75	5	0.36	11.80	0.00	2.57	1.37	2.13
-0.78	-0.41	NA	NA	NA	NA	riken cdna 49324	29	9391	-0.59	1	0.27	-3.15	0.20	12.71	-2.99	1.80
0.51	1.08	-0.07	0.25	0.25	1.12	branched chain ar	14	4337	0.52	5	0.48	2.65	0.05	2.57	0.02	1.03
-0.07	0.05	0.17	0.17	NA	NA	riken cdna 24000(	24	7618	0.08	3	0.12	1.39	0.26	3.18	-0.10	0.26
0.35	0.05	NA	NA	NA	NA	atonal homolog 7	26	8141	0.20	1	0.21	1.31	0.41	12.71	-1.72	2.12
-1.09	-0.23	-1.48	-1.08	NA	NA	mago-nashi homo	29	9242	-0.97	3	0.52	-3.70	0.03	3.18	-1.80	-0.14
-1.10	0.13	0.10	-0.53	-2.02	-1.41	steroidogenic acu	13	3999	-0.80	5	0.86	-2.29	0.07	2.57	-1.71	0.10
-1.41	-1.25	-0.25	-1.33	0.13	NA	riken cdna 49305	17	5400	-0.82	4	0.71	-2.58	0.06	2.78	-1.70	0.06

Part III:

Now that you have your normalized data results –

How was the normalization process and analysis performed? (as described in Part II and in the flow chart shown on the next page)

How well did the normalization process work? Which method of linear normalization should I choose?

What do the 'numbers' in the output file mean?

What do I do next?

What is the future of microarray analysis at PBRC?

Microarray Analysis and Normalization v_1.0 [median/mouse oligo]

DMG/JWM/AAP/RAK/JB
Updated 02/24/03

Step 1: Scan slides using QuantArray software and assign USER DEFINED treatments to Cy3 or Cy5*

**example channels: High gainers/Low gainers, High exercisers/Low exercisers, High fat diet/Low fat diet*

Step 2: Get USER DEFINED choice for reference gene. Calculate raw intensity values and reference medians per sub-array.
(pooled oligo (default) or luciferase)

Step 3: Normalize channel intensities (per sub-array [48]) with USER DEFINED choice(s).
(pooled oligo (default) or luciferase)

Step 4: Remove unwanted genes and genes with intensity < critical intensity (default = 0):

Step 5: Calculate the corrected ratio of all genes per sub-array using USER DEFINED method;
(median or trimmed mean for ~324 genes, 48 sub-arrays/slide)
Pending: Option to use 3 arrays (T-test alert if inappropriate)

Step 6: Calculate and retrieve the normalized ratio per gene
(slide)(array)(sub-array) = total genes
(1X3)(4X4)(18X18) = 15552 genes

All slides Processed?

N

Y

Step 7: Analyze normalized ratios per gene.

a.) calculate raw intensity channel values X_i X_k per sub-array:

$X_{i\ (\text{pooled oligo})}$ = (intensity-background)
$X_{k\ (\text{pooled oligo})}$ = (intensity-background)

b.) sort $X_{i\ (\text{pooled oligo})}$ $X_{k\ (\text{pooled oligo})}$ per sub-array.

c.) get medians $m_{i\ (\text{pooled oligo})}$, $m_{k\ (\text{pooled oligo})}$ of X_i X_k

a.) Normalize each channel intensity C_i C_j per sub-array.

$$N_i = \frac{C_i}{m_{i\ (\text{pooled oligo})}} \qquad N_k = \frac{C_k}{m_{k\ (\text{pooled oligo})}}$$

Remove USER DEFINED genes , ie: salmon sperm, pos/neg controls, etc. (default = salmon Sperm DNA (Cy5), 'blanks', neg/pos).

Remove intensities lower than a default of 0, or lower than users choice.
NOTE: luciferase may be used as an alternate reference gene

a.) sort norm channels N_i N_k per sub-array.

b.) calculate medians $m_{i(\text{norm})}$ $m_{k(\text{norm})}$ per channel N_i N_k per sub-array.

(NOTE: this provides potentially different normalizing factors for genes located in each of the 48 sub-arrays):

$$CR_k = N_k / m_k \qquad CR_i = N_i / m_i$$

NOTE: the trimmed mean may also be used for normalization

a.) get users ratio: $group_i / group_j$ || $group_j / group_i$ **

$$R_{\text{Normalized ratio spot}} = \frac{group_i}{group_k} = \frac{CR_k}{CR_i}$$

b.) output gene normalized $R_{\text{Normalized ratio spot}}$ ratio to list.

(**Note: this may be ch1/ch 2 OR ch2/ch1 depending on scan method.)

a.) get list of $R_{\text{Normalized ratio spot}}$ per gene per sub-array per number.
b.) convert $R_{\text{Normalized ratio spot}}$ ratios to $\log_2$
c.) get mean & S.D.
d.) (option: get USER DEFINED μ), default $\mu = 0$
e.) perform 2-sided Students T-test (test of mu, small n)
f.) get p-value for informational reference.
g.) get USER DEFINED α
h.) determine Confidence Interval for μ:

$$t = \frac{(\bar{X}_R - \mu)}{(s/\sqrt{n})}$$

$$\bar{X} - t_{\alpha/2,\,(n-1)\text{d.f.}}(s/\sqrt{n}) < \mu < \bar{X} + t_{\alpha/2,\,(n-1)\text{d.f.}}(s/\sqrt{n})$$

i.) return p-value, t-score, and Confidence Intervals to the user for analysis.

T-Test : (2-sided)
H_o: $\mu = 0$
H_a: $\mu \neq 0$
$(\log_2(1) = 0)$

How well did the normalization process work for dye-based differences?

We can assess the quality of the normalization process on 'Cy'-dye incorporation/ development by generating an MA plot.

MA plot: [R = 'red'/biotin/Cy5; G = 'green'/fluorescein/Cy3]

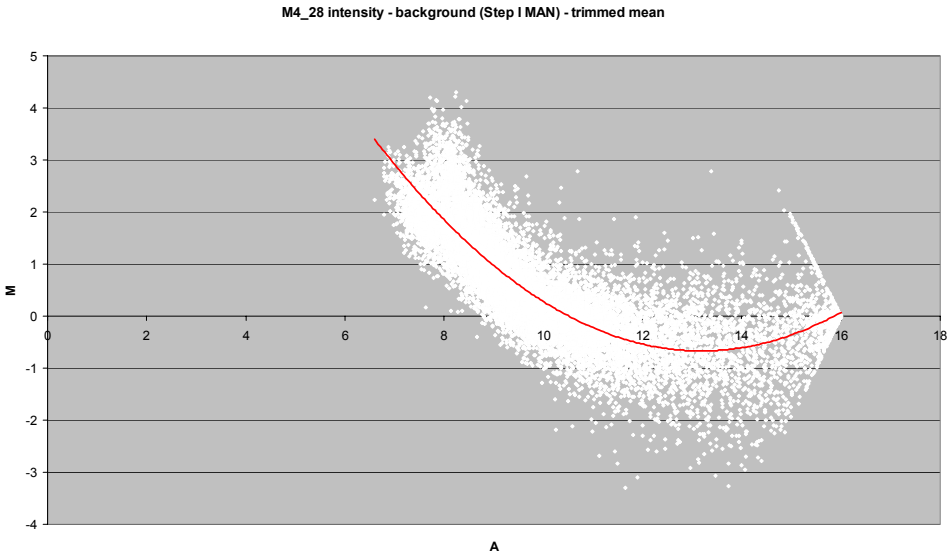
$M = \log_2(R/G)$

$A = \log_2(\mathrm{sqrt}(R*G))$

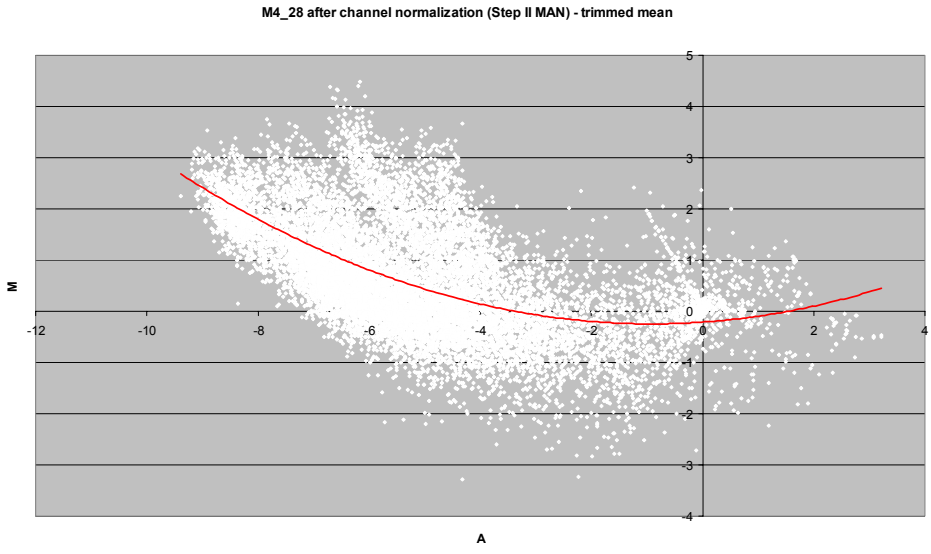
Note: under ideal conditions, the line fitted to the data would be linear with a slope of zero.

The following two slides show MA plots for the data generated during 3 of the steps coded in the M.A.N. script and illustrate how the data distribution changes during normalization. The M.A.N. script was coded in house by John Bearden, incorporating input from both the HEF bioinformatics group (J. Bearden, C. Faulk, D. Graunke, R. Koza, J. McClaine, and A. Ptitsyn), the PBRC bioinformatics group and the PBRC biostatistics group.

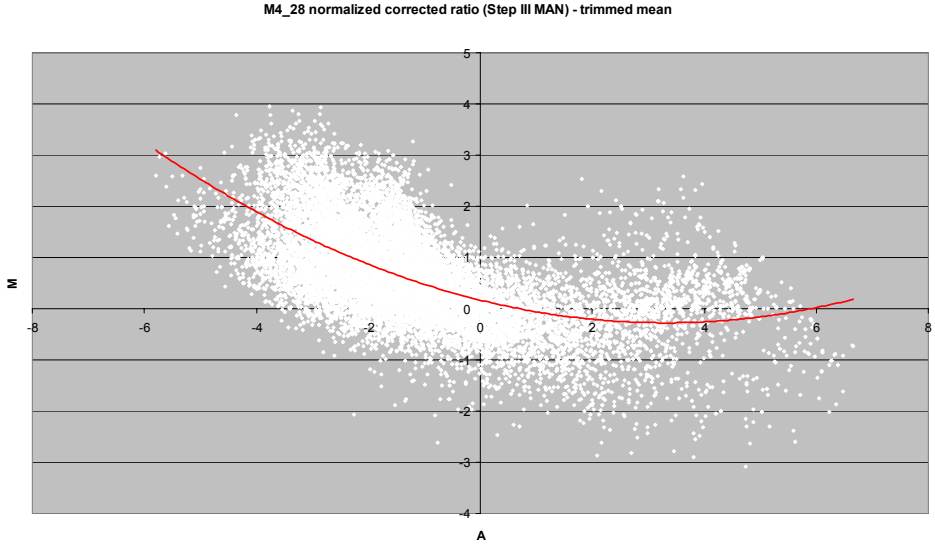
The effect of using trimmed mean during normalization vs median:
Trimmed mean



MA plot for each spot (intensity – background)

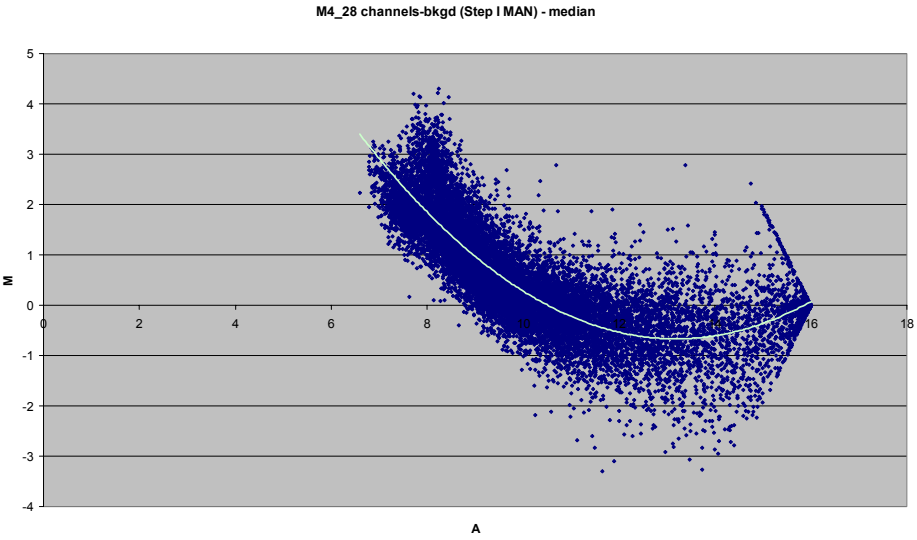


MA plot for each spot (following normalization to the trimmed mean value of *pooled oligo* or luciferase spots – performed per sub-array)

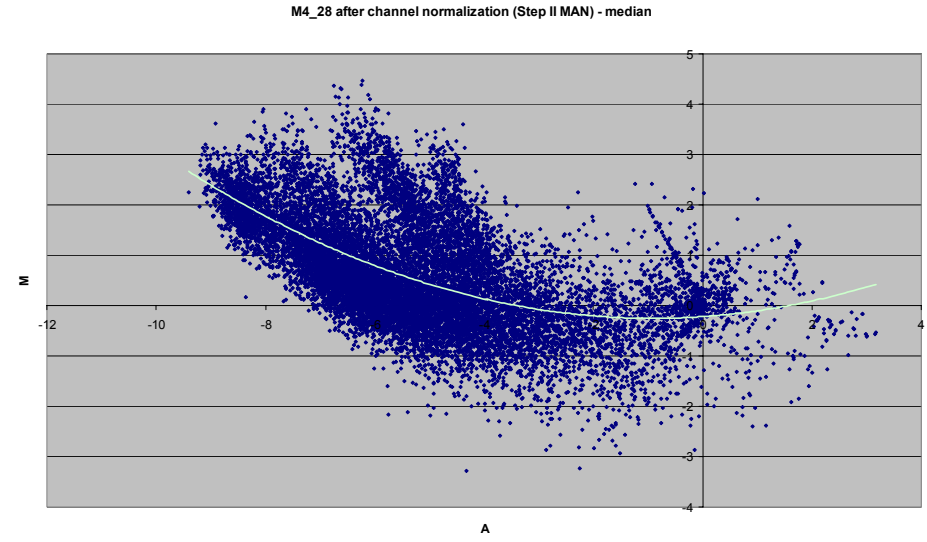


MA plot for each spot (following normalization to correction factor [*trimmed mean*] – performed per sub-array)

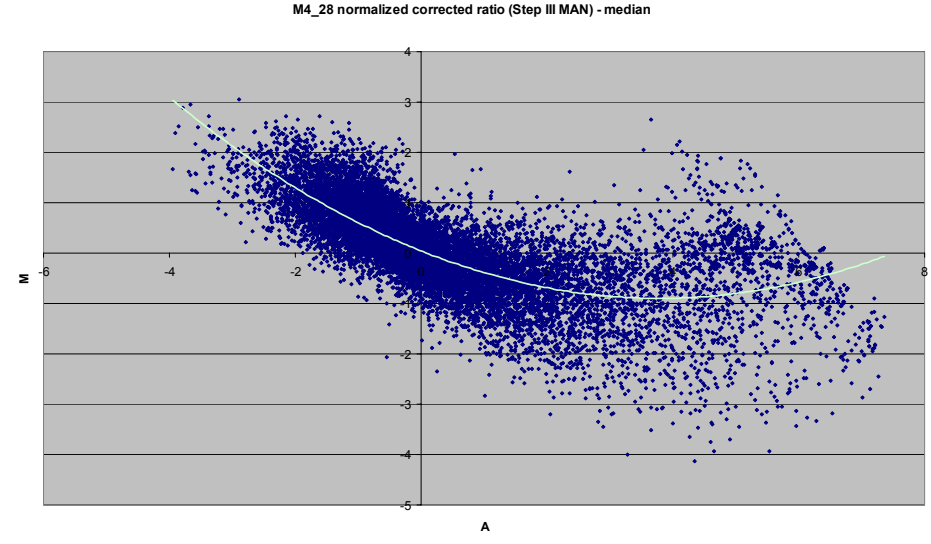
The effect of using trimmed mean during normalization vs median:
Median



MA plot for each spot (intensity – background)



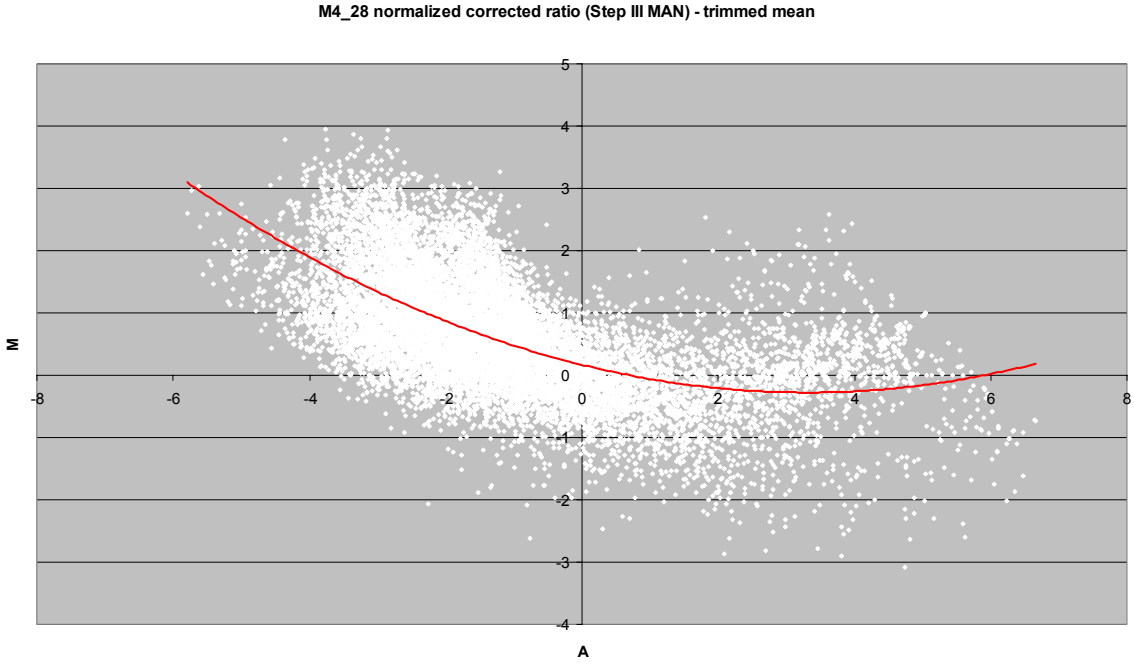
MA plot for each spot (following normalization to the median value of *pooled oligo* or luciferase spots – performed per sub-array)

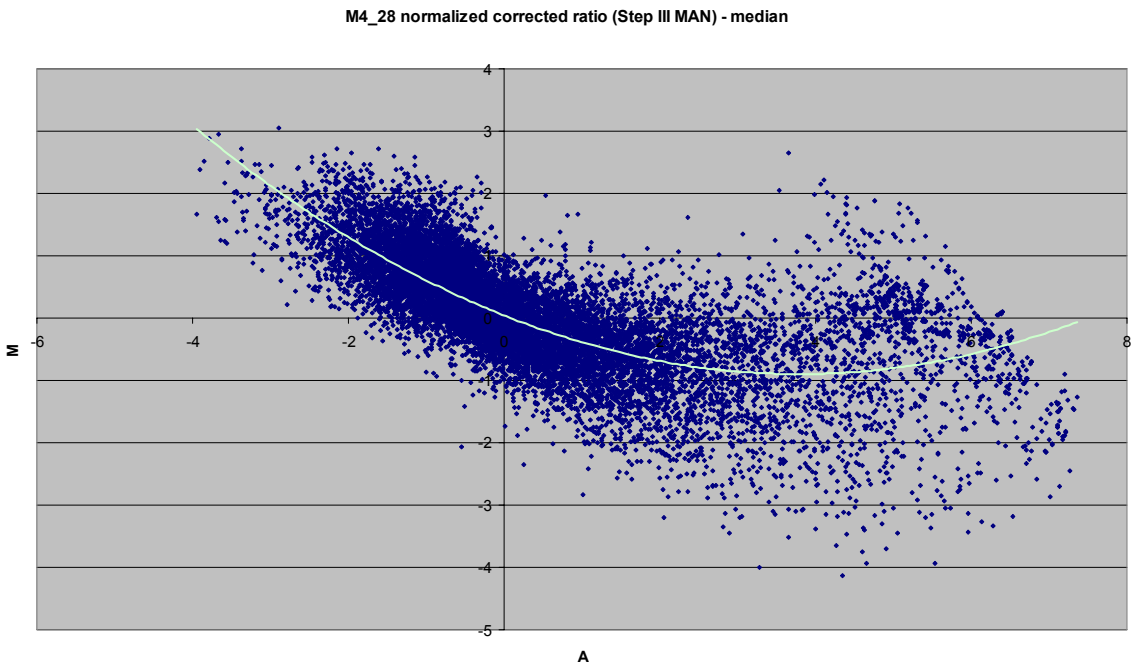


MA plot for each spot (following normalization to correction factor [*median*] – performed per sub-array)

Do I choose to use the median or trimmed mean values during normalization?

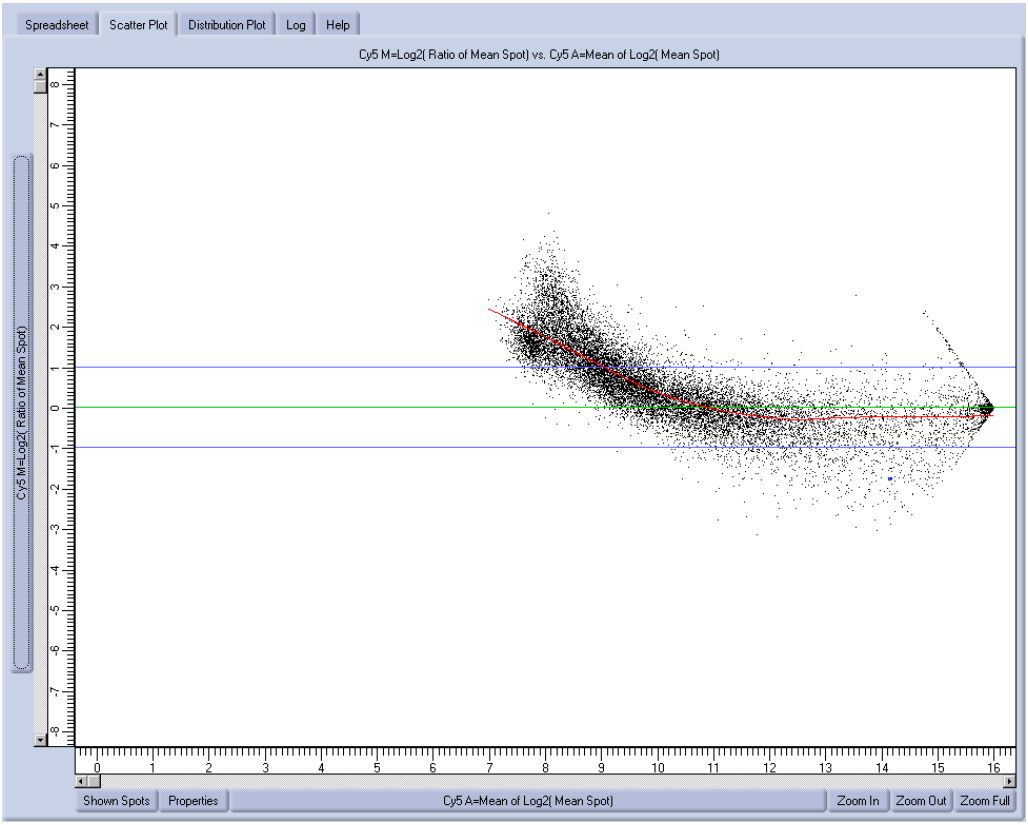
Both yield similar results although the trimmed mean is 'more' linear after normalizing this particular slide. Results may vary from experiment to experiment although you should use one method consistently.





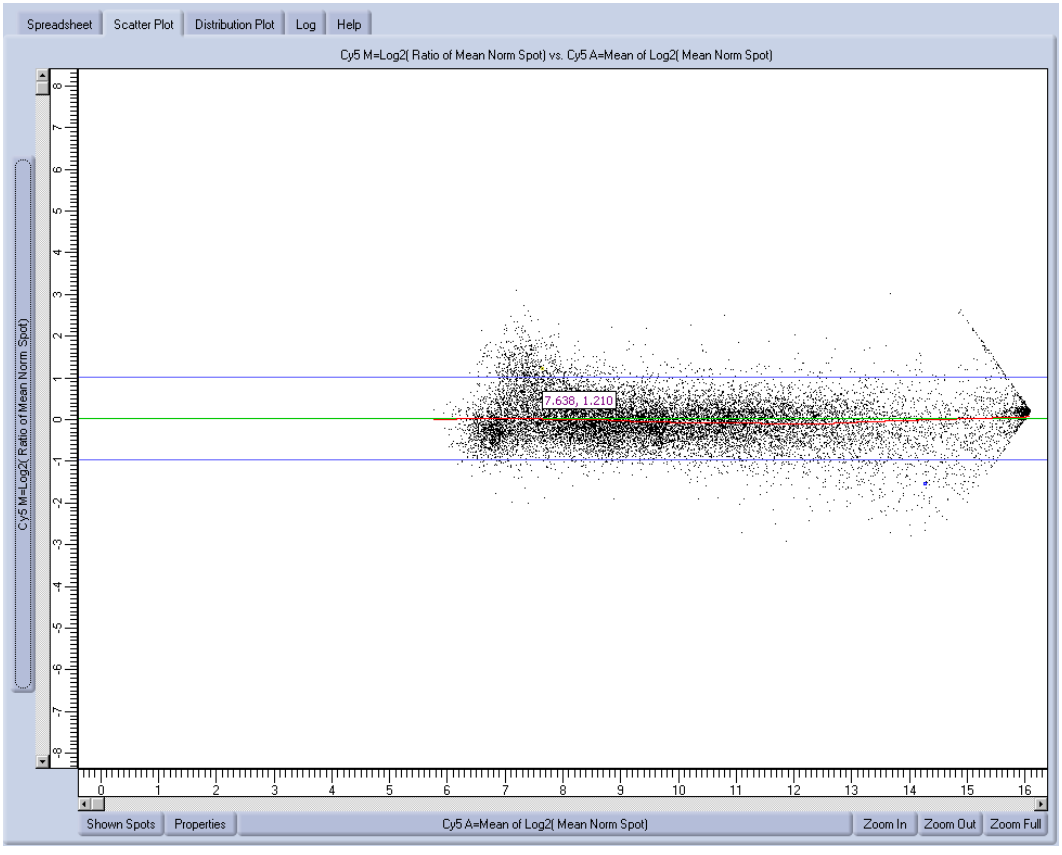
Is there another option?

We are currently developing use of a method called LOWESS (**Lo**cally **We**ighted **S**catter plot **S**moothing).



MA plot for each spot (intensity – background) prior to normalization

MA plot for each spot following *LOWESS* normalization



What do the numbers in the output file mean?

RATIO (#) – this is the $\log_2$ value of the normalized ratio for a given gene on slide #

D.F. - degrees of freedom; n – 1

n - the total number of 'good quality' normalized ratios for a given gene

S.D. - the standard deviation of all 'good quality' normalized ratios for given gene

T-Test - the calculated value of a 'statistic' which is indicative of how far your test value is from the hypothesized value

P-Value - The t-distribution is a bell-shaped curve that describes how the data is distributed. The p-value is the area under this curve beyond your calculated t-statistic value. In practical terms, a p-value estimates the potential for error; i.e. a p-value of 0.05 translates into a 5% chance of being incorrect in your conclusion that the gene is significantly up- or down-regulated.

CI left - The 'left' side of a confidence interval calculated using a given alpha. You will be asked to input a value for alpha (default is 0.05) and MANv1.0 will calculate a confidence interval within which one can be 95% (100 – 5 as defined by default above) certain that the true value of your parameter lies.

CI right - The 'right' side of the calculated confidence interval

alpha - The possibility of a false positive

non-logged average – As all values are reported as $\log_2$ values, this is easier to understand. This is the fold-difference in expression between your two experimental groups.

What do I do next?

Your first decision is whether you are interested in a specific sub-set of genes or the most significantly differentially regulated genes. We suggest that you perform an unbiased analysis using the following guidelines, however, you will have enough data to look up specific gene expression information for individual genes.

-Open output files using EXCEL (right click on text file and choose to open with EXCEL)
-Sort list using p-values (ascending)
-Copy the entries having a p-value $\leq$ 0.05 (or your chosen cutoff/alpha) onto a new worksheet
-Sort by non-logged average value (ratio) and remove (delete) the entries that have ratios between 0.5 and 2.0 (which represent a 2-fold difference) or your fold difference of choice

This is your candidate gene list and it will likely be somewhat large (>100 genes); it consists of genes that are significantly (or highly significantly if p-value < 0.01) differentially regulated by at least 2-fold.

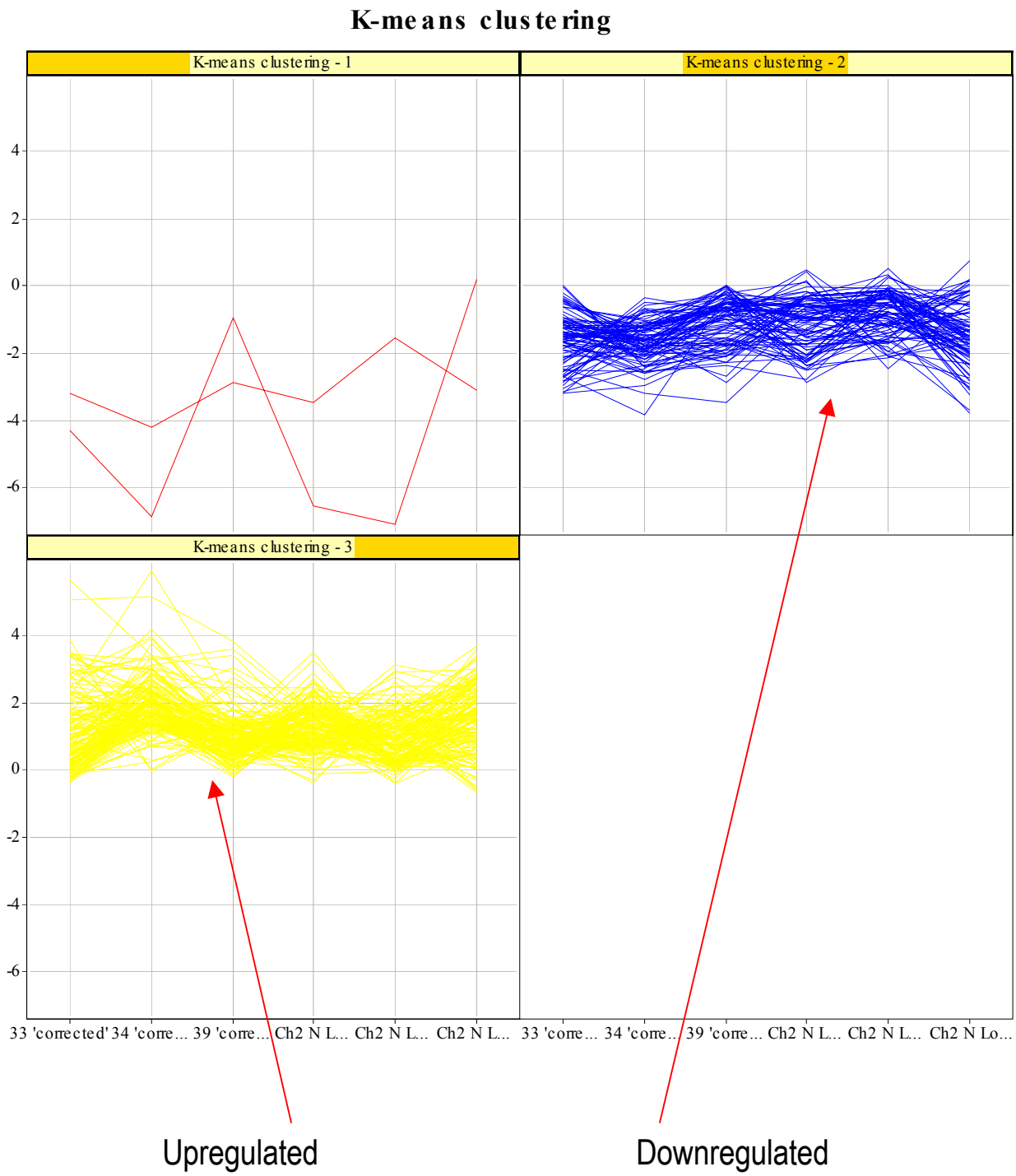
What do I do with my gene list?

We suggest that you upload the ratios for the genes in your gene list into SpotFire DecisionSite and run **K-means clustering** with a maximum cluster number set to 3. This should generate three graphs, one showing upregulated genes, one showing downregulated genes, and potentially one showing any oddities. You may want to review and remove the odd-looking profiles. See next page for sample clustering results)

Perform literature search for genes with 'logical' regulation in your experimental setting/model and your tissue of interest.

Divide genes into functional groups/pathways – GeneSpring has the capacity to aid in this step by linking to the KEGG pathway database. Sorry, this function is not currently operational in GeneSpring and must be performed manually.

Confirm gene expression using Quantitative Real Time PCR or Northern blot



Genes whose profile is shown in cluster 1 should likely be ‘thrown out’ of the analysis.

What is the future of microarray analysis at PBRC?

Everything shown in today's presentation is relevant specifically to the mouse oligonucleotide slides printed by HEF. (tentative scheduled availability is given in time after March 1)

1) Expand the analysis to be applied to human oligonucleotide arrays – this step should not be very different from the current analysis but does require additional testing and troubleshooting. (≤ 1 month)

2) Upgrade the ScanArray 5000 microarray scanner – this should occur sometime in the next two months and will require re-training for scanning as it includes new, although similar software to perform microarray scanning and quantitation. (1-2 months)

3) Expand the analysis to be applied to cDNA arrays (mouse hypothalamus) – this step should not be very different from the current analysis but does require additional testing and troubleshooting. (≤ 1 month) Also, if you have performed experiments using these slides you already know that the resulting 'gene list' is merely a list of unique identifying numbers. If you would like to learn more as to what these numbers mean and how to find out the 'genes' they correspond to, please contact Dawn Graunke (automating 'gene' identification is not possible at this time).

4) Incorporate *LOWESS* into MANv2.0 – We are currently testing various methods for Performing *LOWESS* and will develop version 2.0 for future release (3-6 months)

5) Integrate MAN into BASE (BioArray Software Environment) – (≤ 6 months) BASE is a MIAME (minimal information about microarray experiments)-compliant standard for storage and analysis of microarray data.