

Measuring Gene Expression Part 3

David Wishart

Bioinformatics 301

david.wishart@ualberta.ca

Objectives*

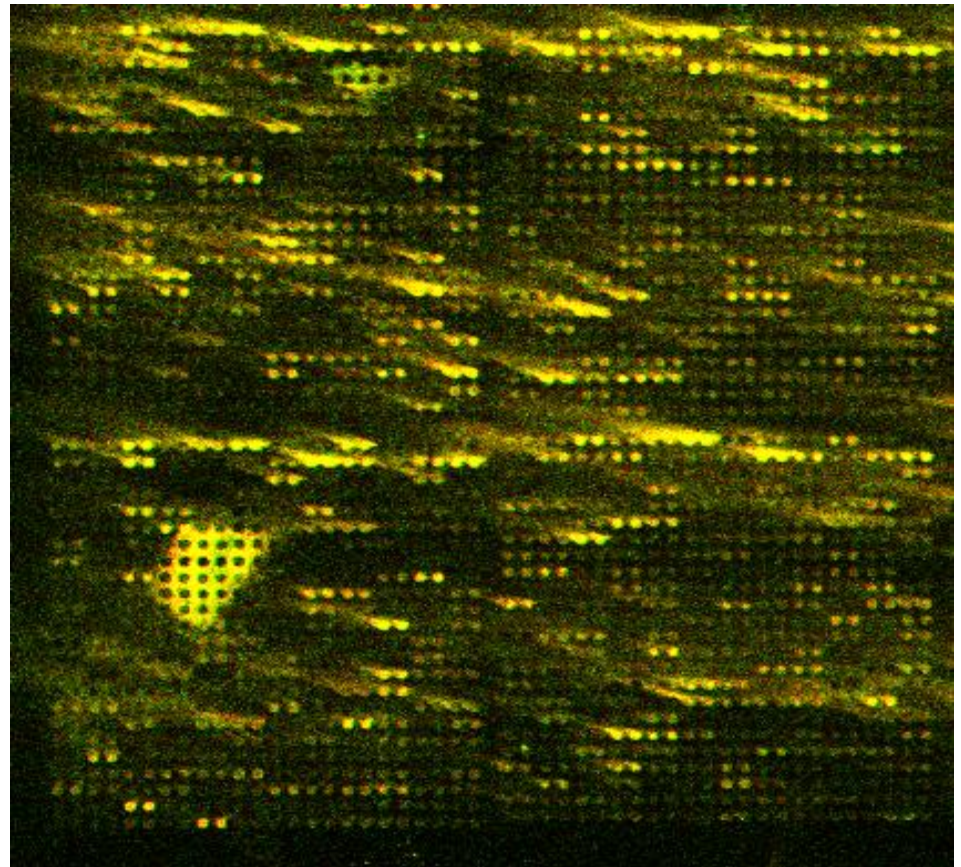
- **Become aware of some of the causes of low quality microarray data**
- **Become familiar with gridding, spot picking, intensity determination & quality control issues**
- **Become familiar with normalization, curve fitting and correlation**
- **Understand how microarray data is analyzed**

Key Steps in Microarray Analysis*

- **Quality Control (checking microarrays for errors or problems)**
- **Image Processing**
 - **Gridding**
 - **Segmentation (peak picking)**
 - **Data Extraction (intensity, QC)**
- **Data Analysis and Data Mining**

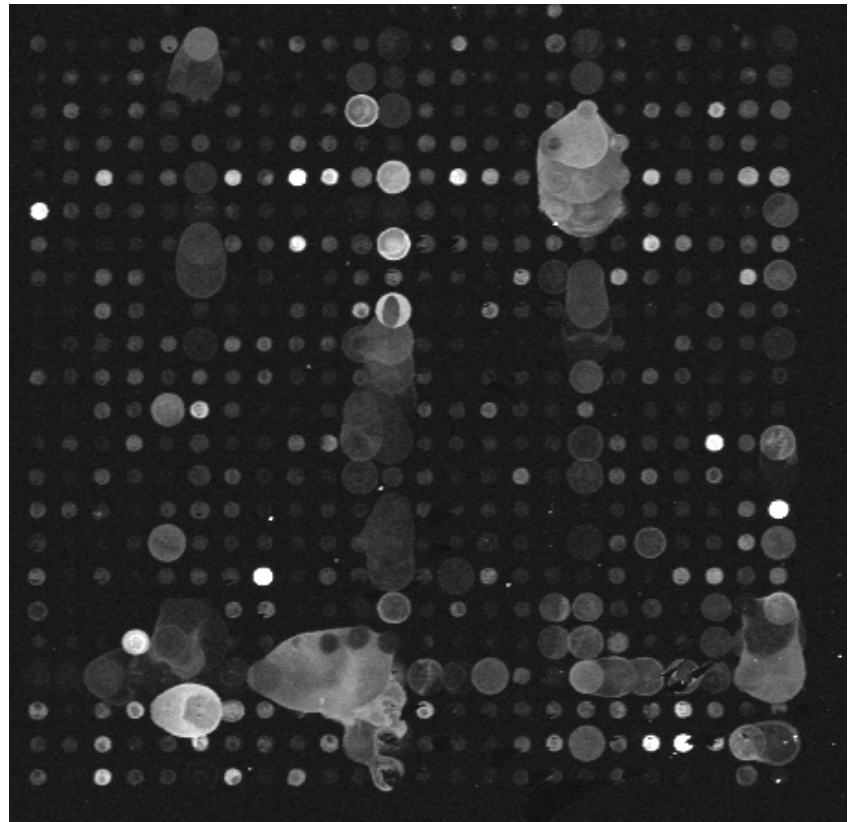
Comet Tailing*

- Often caused by insufficiently rapid immersion of the slides in the succinic anhydride blocking solution.

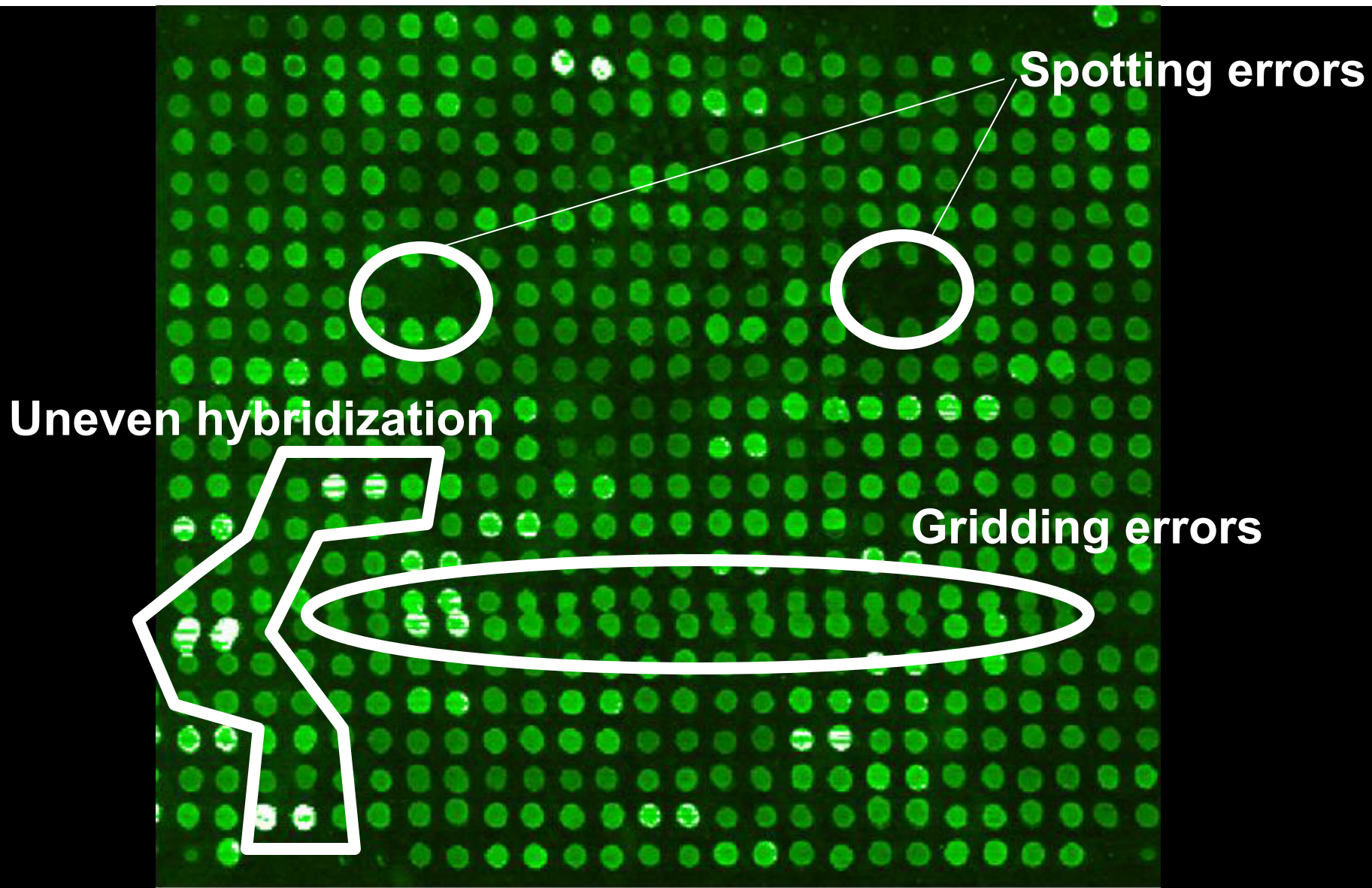


Uneven Spotting/Blotting

- **Problems with print tips or with overly viscous solution**
- **Problems with humidity in spotting chamber**



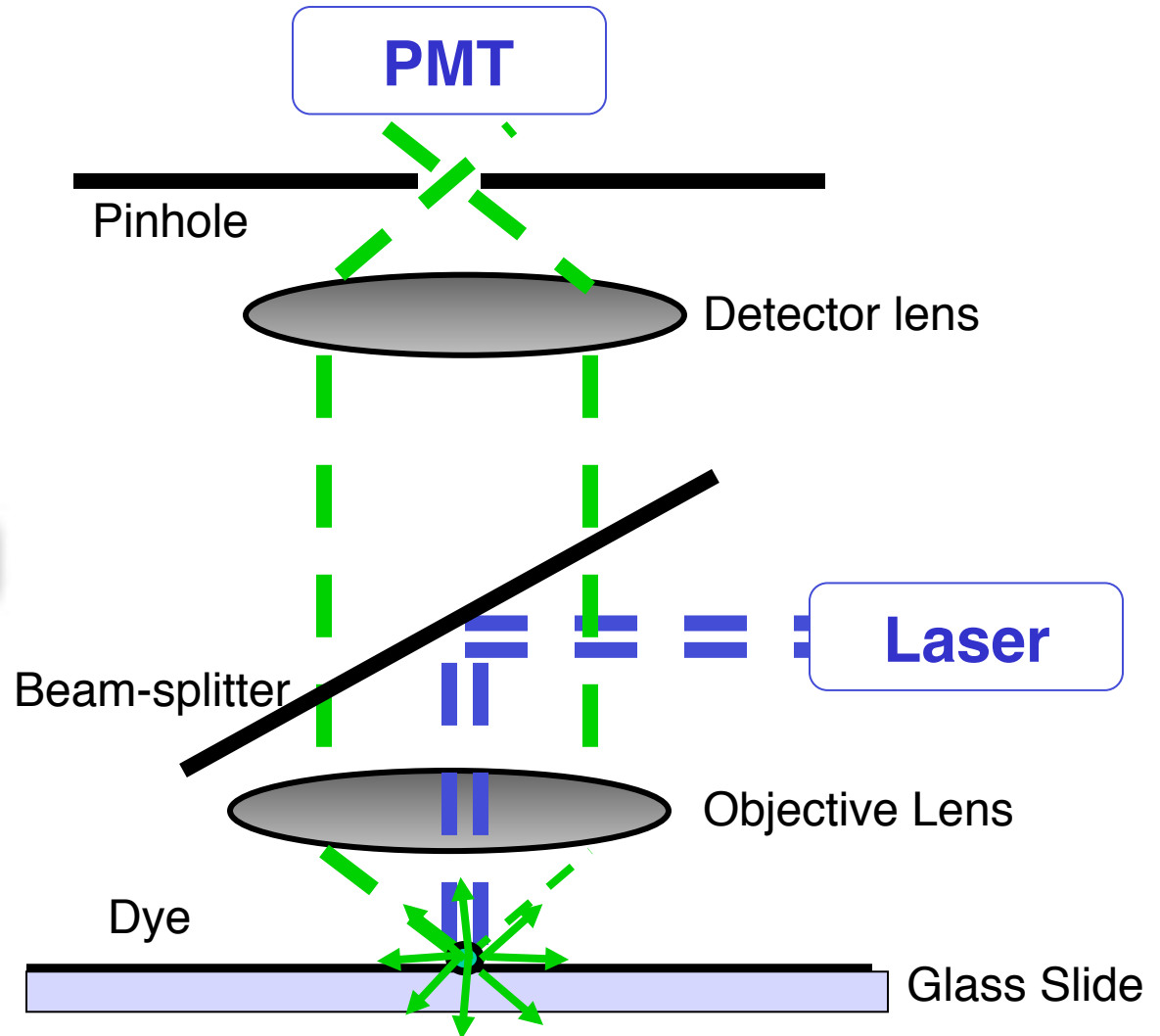
Gridding Errors



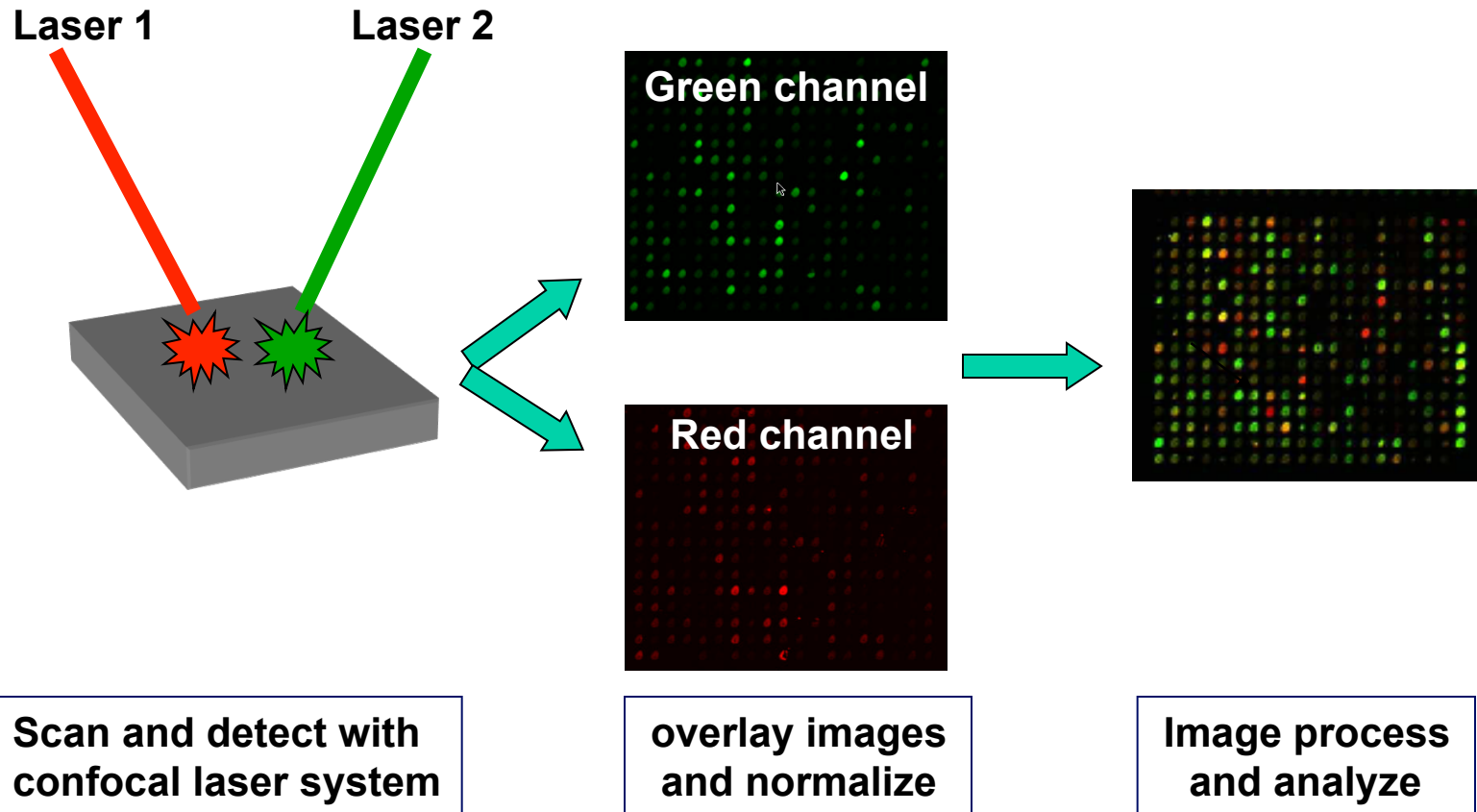
Key Steps in Microarray Analysis

- **Quality Control (checking microarrays for errors or problems)**
- **Image Processing**
 - **Gridding**
 - **Segmentation (spot picking)**
 - **Data Extraction (intensity, QC)**
- **Data Analysis and Data Mining**

Microarray Scanning*



Microarray Principles*



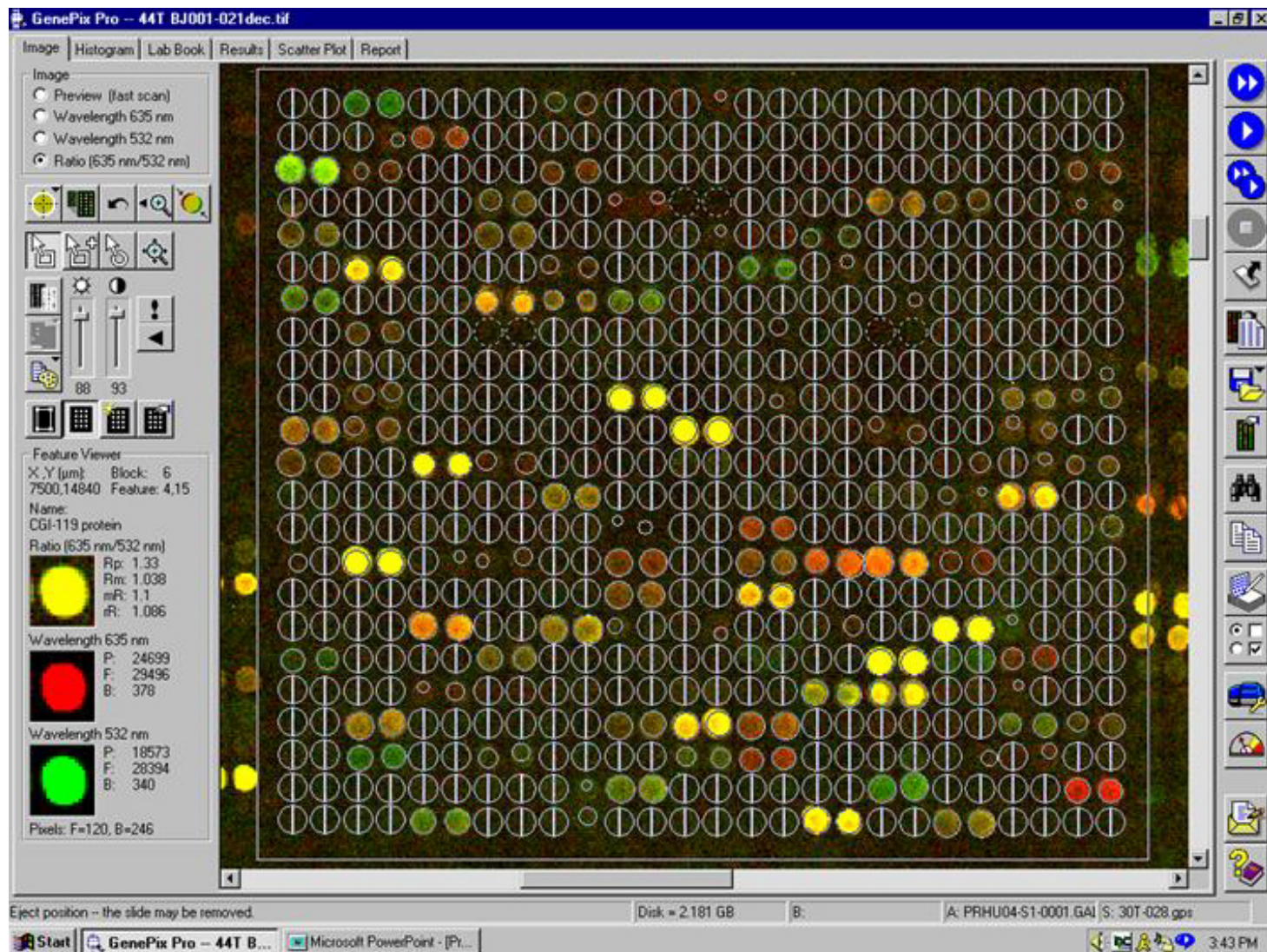
Microarray Images

- **Resolution**
 - standard $10\mu\text{m}$ [currently, max $5\mu\text{m}$]
 - $100\mu\text{m}$ spot on chip = 10 pixels in diameter
- **Image format**
 - TIFF (tagged image file format) 16 bit (64K grey levels)
 - 1cm x 1cm image at 16 bit = 2Mb (uncompressed)
 - other formats exist i.e. SCN (Stanford University)
- **Separate image for each fluorescent sample**
 - channel 1, channel 2, etc.

Image Processing*

- **Addressing or gridding**
 - Assigning coordinates to each of the spots
- **Segmentation or spot picking**
 - Classifying pixels either as foreground or as background
- **Intensity extraction (for each spot)**
 - Foreground fluorescence intensity pairs (R, G)
 - Background intensities
 - Quality measures

Gridding

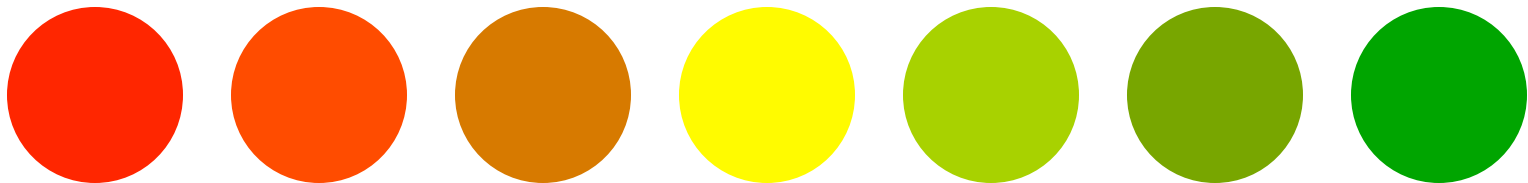


Gridding Considerations*

- Separation between rows and columns of grids
- Individual translation of grids
- Separation between rows and columns of spots within each grid
- Small individual translation of spots
- Overall position of the array in the image
- *Automated & manual methods available*

Spot Picking

- **Classification of pixels as foreground or background (fluorescence intensities determined for each spot are a measure of transcript abundance)**
- **Large selection of methods available, each has strengths & weaknesses**

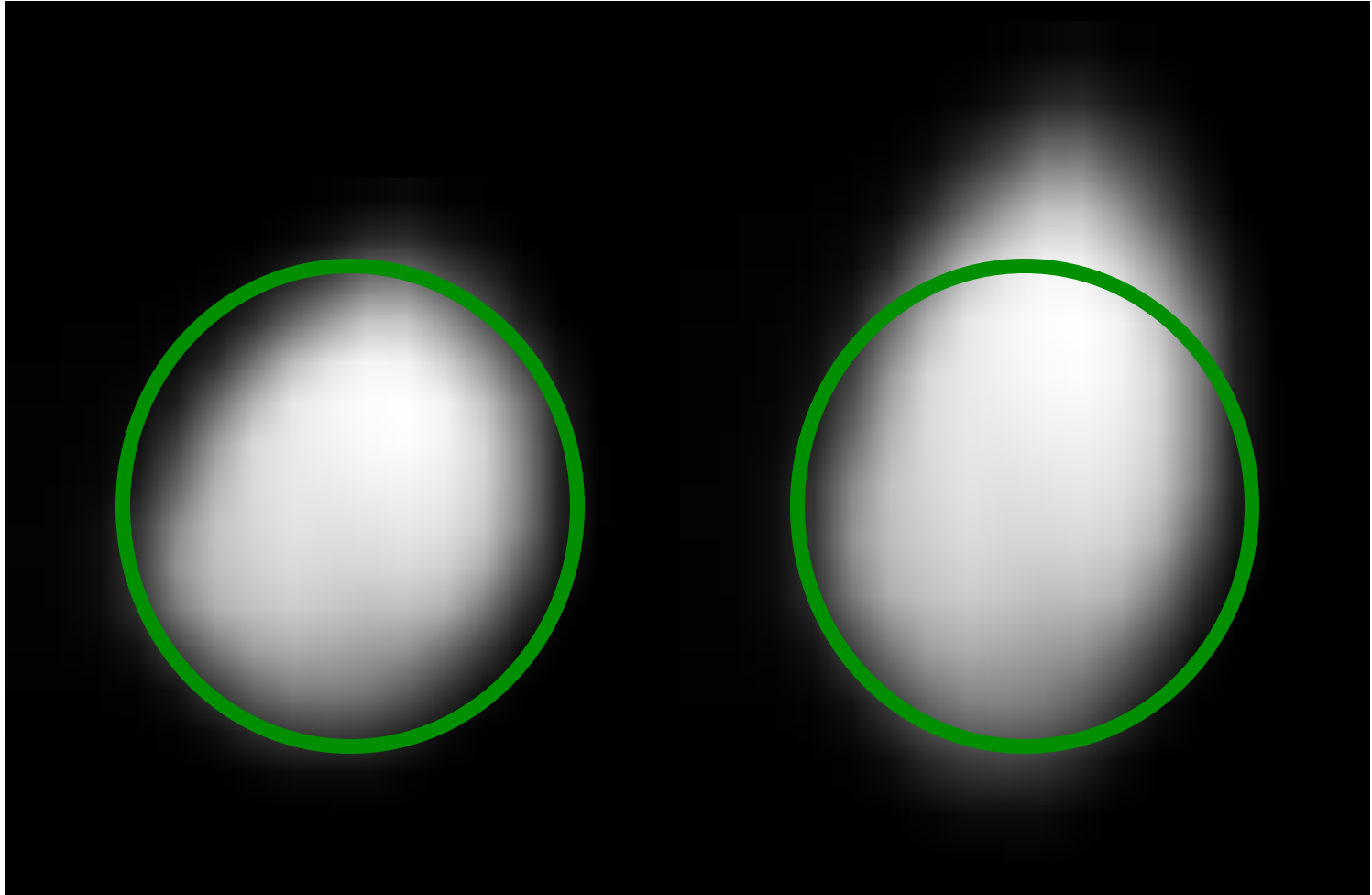


Spot Picking*

- **Segmentation/spot picking methods:**
 - **Fixed circle segmentation**
 - **Adaptive circle segmentation**
 - **Adaptive shape segmentation**
 - **Histogram segmentation**

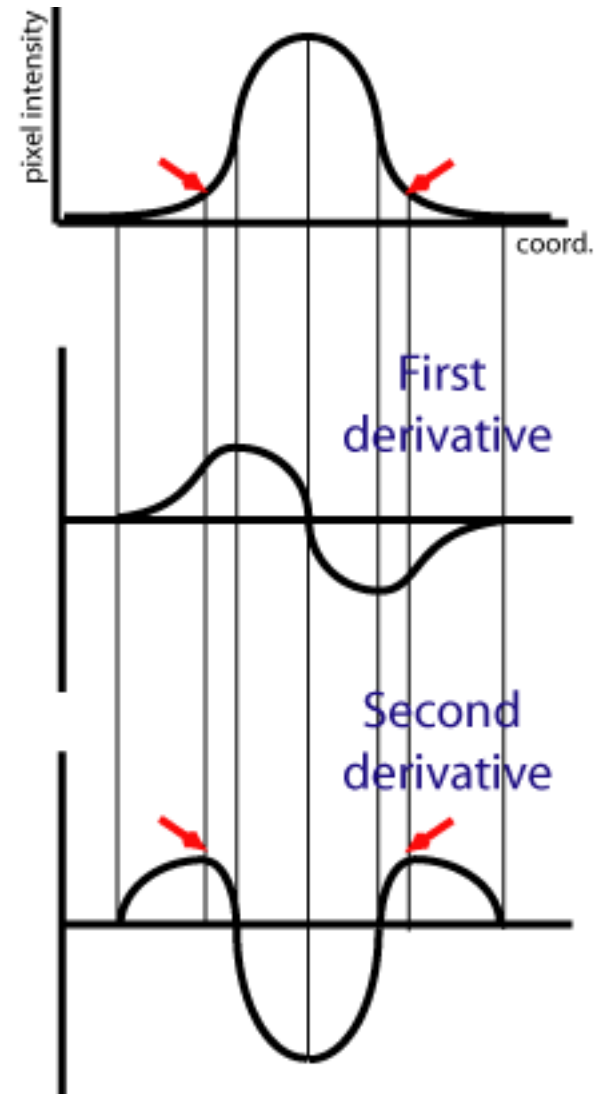
Fixed circle	ScanAlyze, GenePix, QuantArray
Adaptive circle	GenePix, Dapple
Adaptive shape	Spot, region growing and watershed
Histogram method	ImaGene, QuantArraym DeArray and adaptive thresholding

Fixed Circle Segmentation*

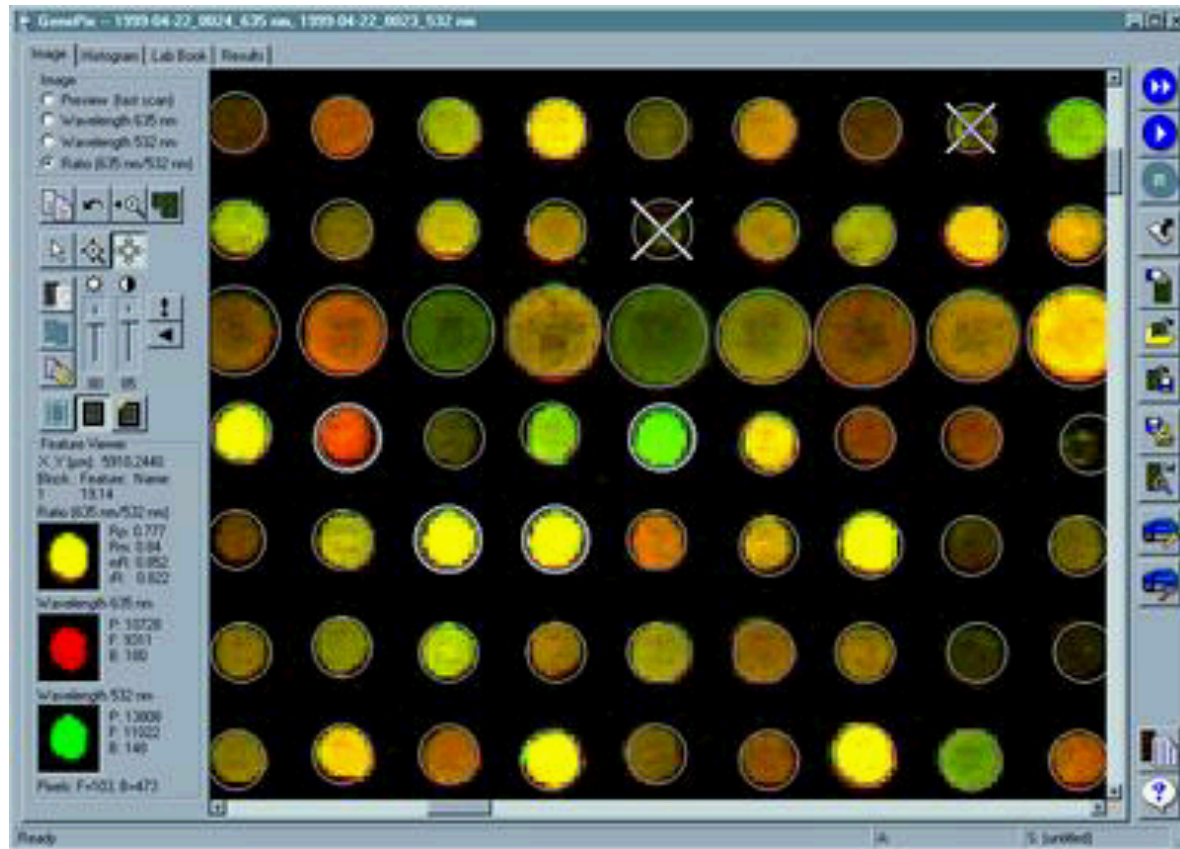


Adaptive Circle Segmentation*

- The circle diameter is estimated separately for each spot
- *GenePix* finds spots by detecting edges of spots (second derivative)
- Problematic if spot exhibits oval shapes

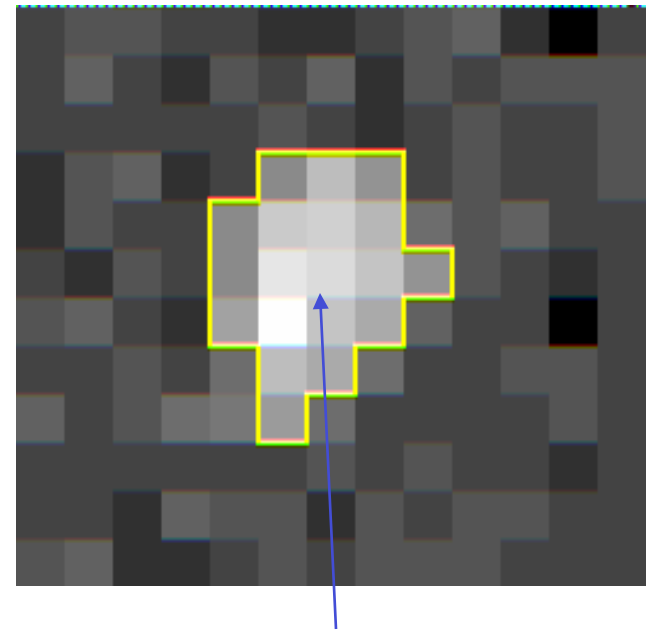


Adaptive Circle Segmentation



Information Extraction

- **Spot Intensities**
 - mean (pixel intensities)
 - median (pixel intensities)
- **Background values**
 - Local Background
 - Morphological opening
 - Constant (global)
- **Quality Information**

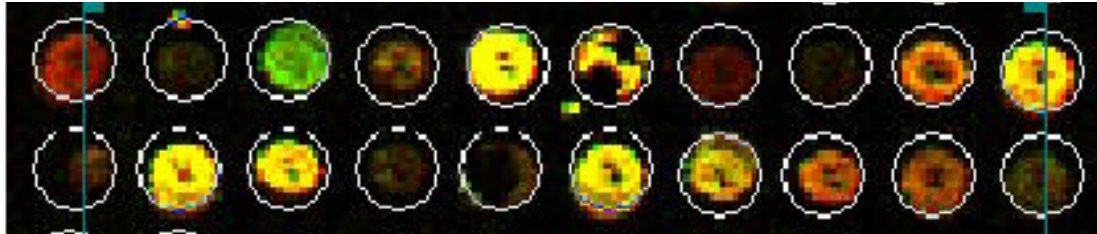


Take the average

Spot Intensity*

- The total amount of hybridization for a spot is proportional to the *total fluorescence* at the spot
- Spot intensity = sum of pixel intensities within the spot mask
- Since later calculations are based on **ratios** between cy5 and cy3, we compute the average* pixel value over the spot mask
- **Can use ratios of medians instead of means**

Means vs. Medians*

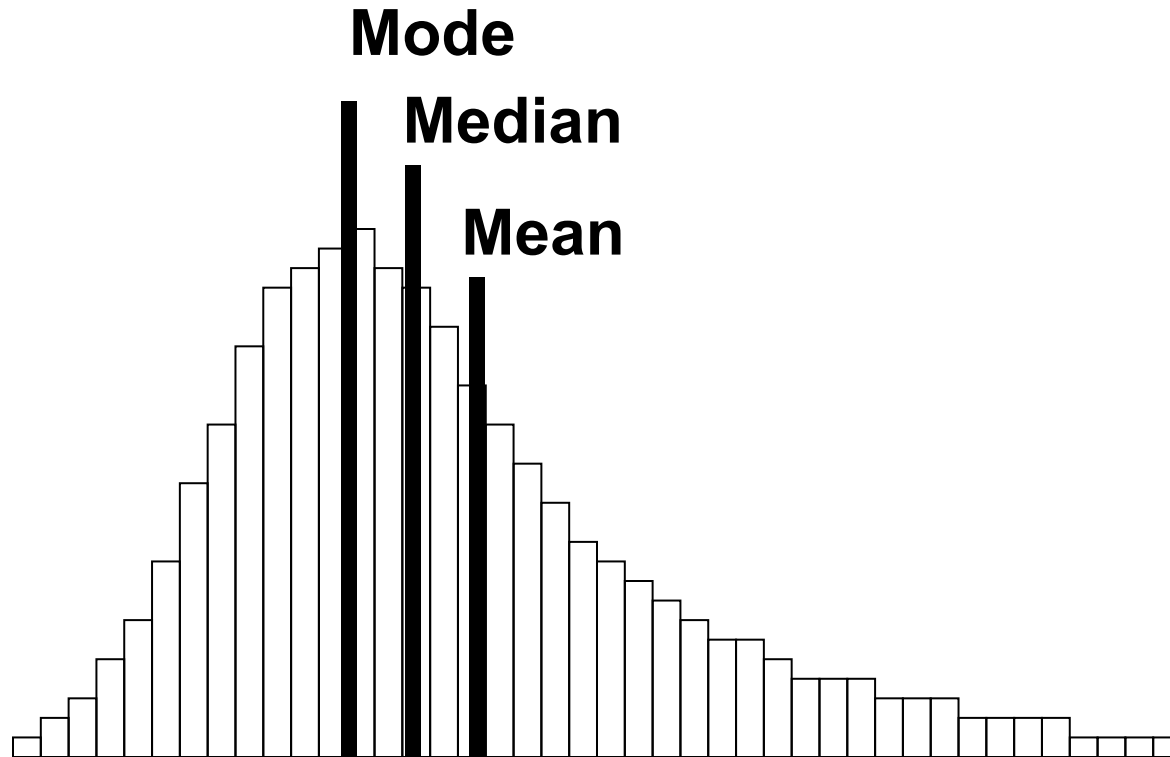


row	col	ch1_sig_mea	ch2_sig_mea	ch1_sig_med	ch2_sig_med
1	1	56000	2000	58000	1900
1	2	1000	600	600	800
1	3	2000	60000	3000	59000



etc.

Mean, Median & Mode



Mean, Median, Mode*

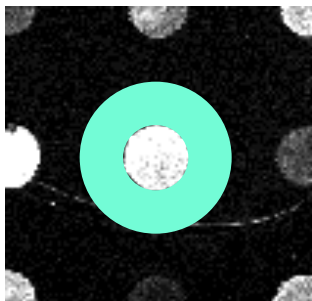
- In a Normal Distribution the mean, mode and median are all equal
- In skewed distributions they are unequal
- **Mean** - average value, affected by extreme values in the distribution
- **Median** - the “middlemost” value, usually half way between the mode and the mean
- **Mode** - most common value

Background Intensity

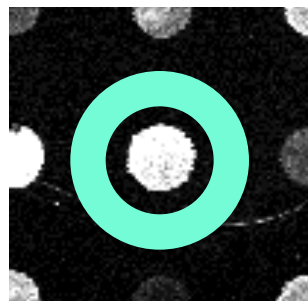
- **A spot's measured intensity includes a contribution of non-specific hybridization and other chemicals on the glass**
- **Fluorescence intensity from regions not occupied by DNA can be different from regions occupied by DNA**

Local Background Methods*

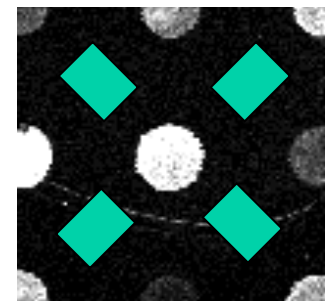
- Focuses on small regions around spot mask
- Determine median pixel values in this region
- Most common approach



ScanAlyze



ImaGene



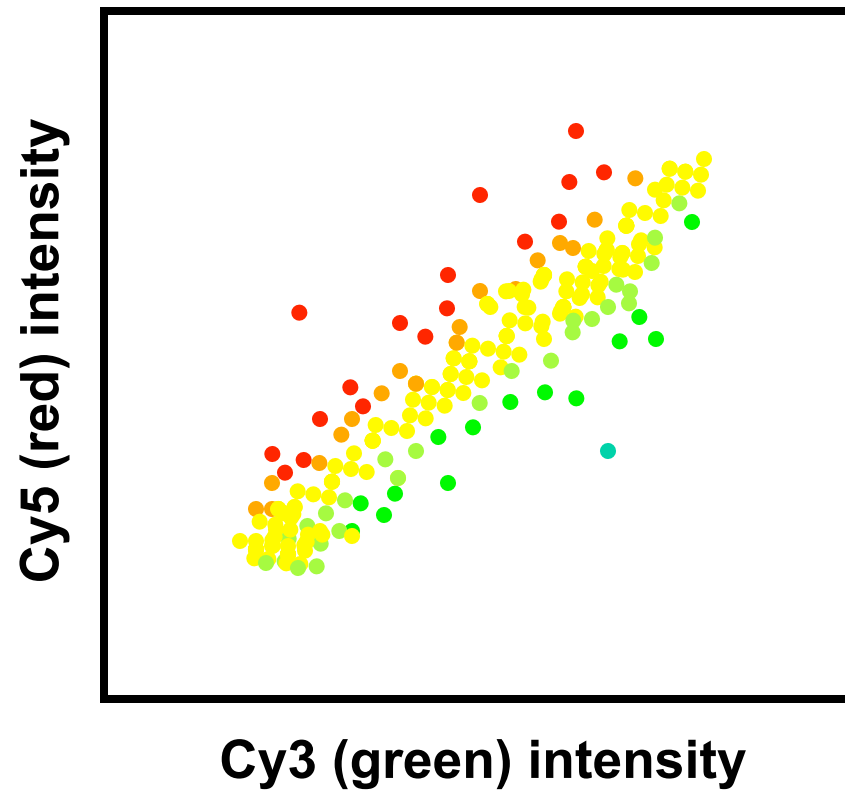
Spot, GenePix

- ***By not considering the pixels immediately surrounding the spots, the background estimate is less sensitive to the performance of the segmentation procedure***

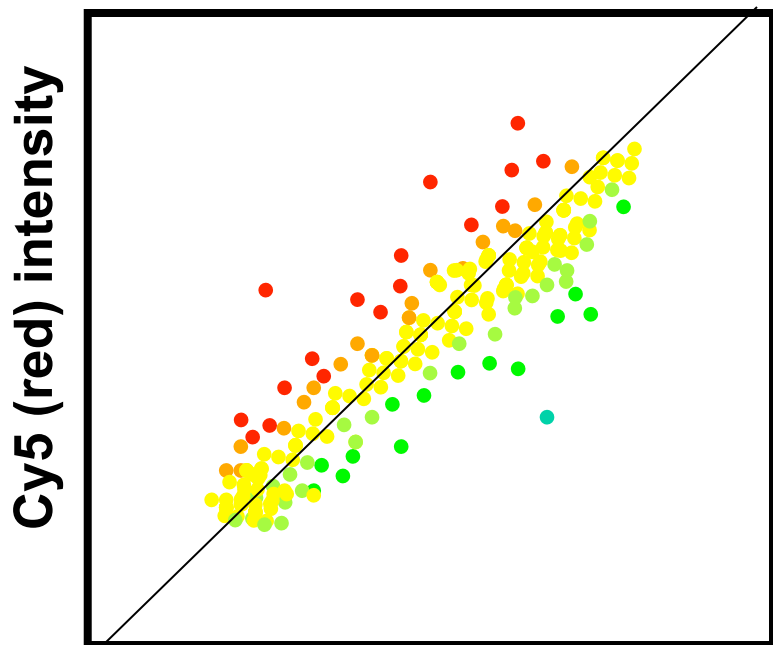
Quality Measurements*

- **Array**
 - *Correlation between spot intensities*
 - Percentage of spots with no signals
 - Distribution of spot signal area
 - Inter-array consistency
- **Spot**
 - Signal / Noise ratio
 - Variation in pixel intensities
 - ID of “bad spots” (spots with no signal)

A Microarray Scatter Plot

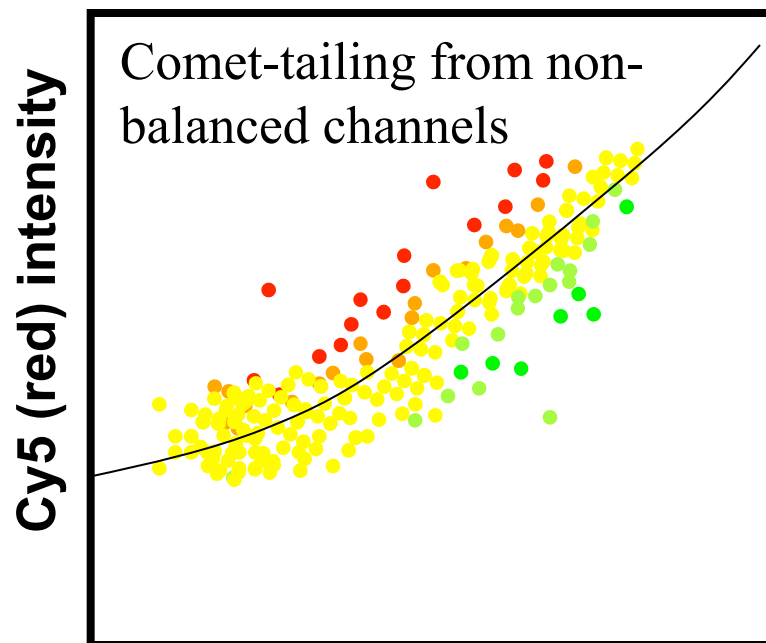


Correlation*



Cy3 (green) intensity

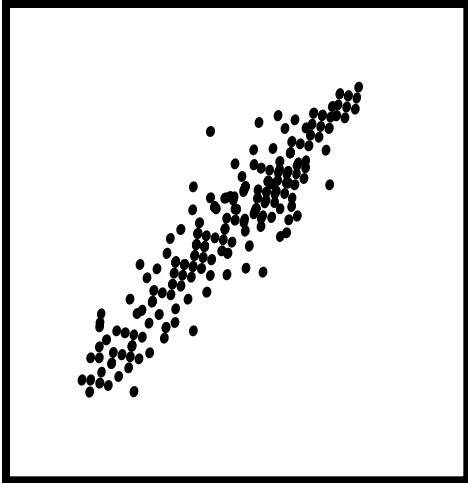
Linear



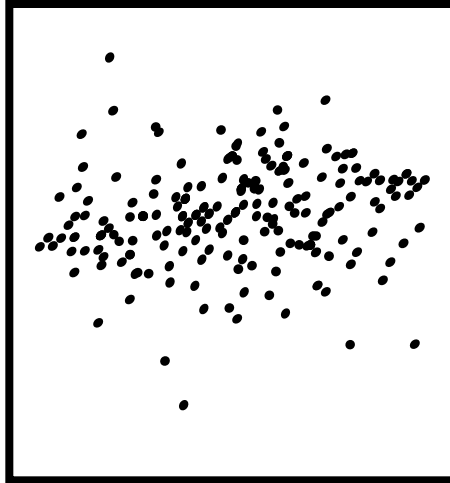
Cy3 (green) intensity

Non-linear

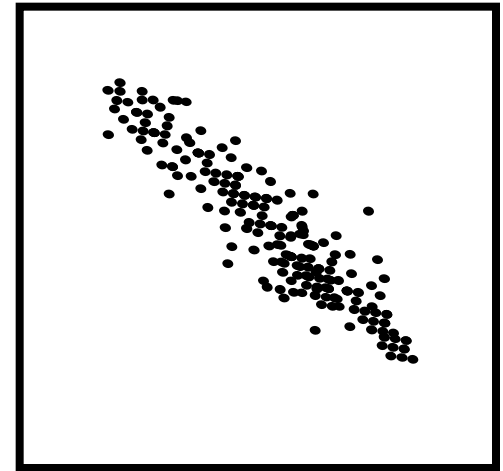
Correlation



“+” correlation

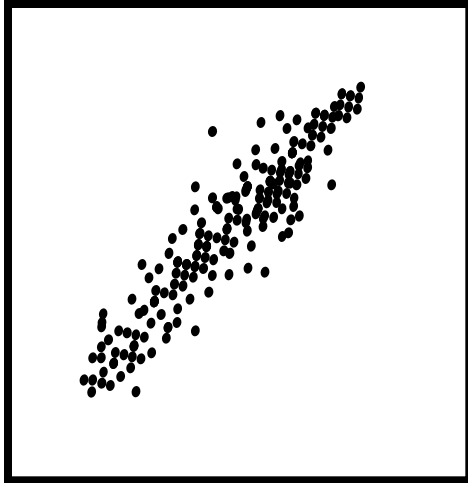


Uncorrelated

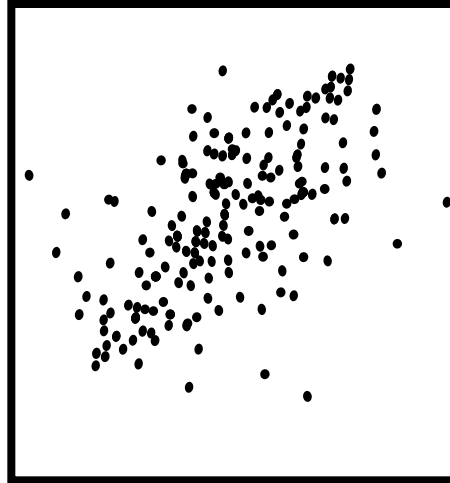


“-” correlation

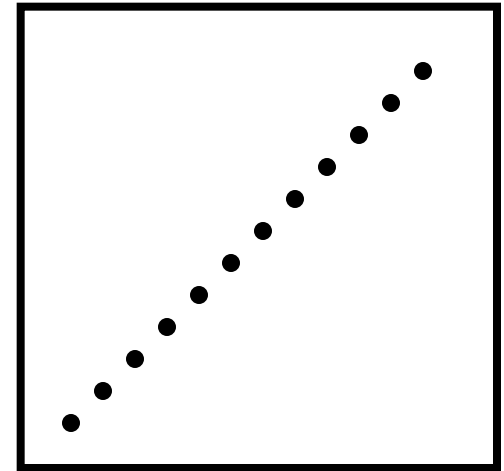
Correlation



**High
correlation**



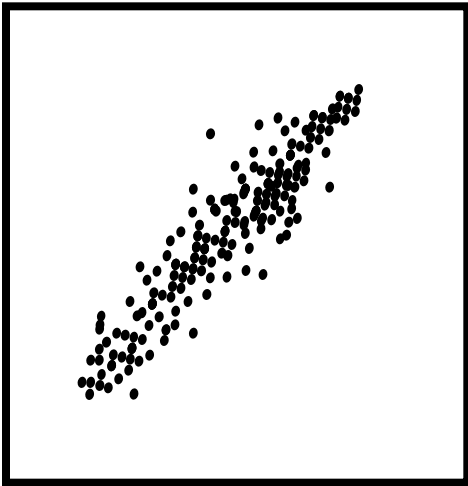
**Low
correlation**



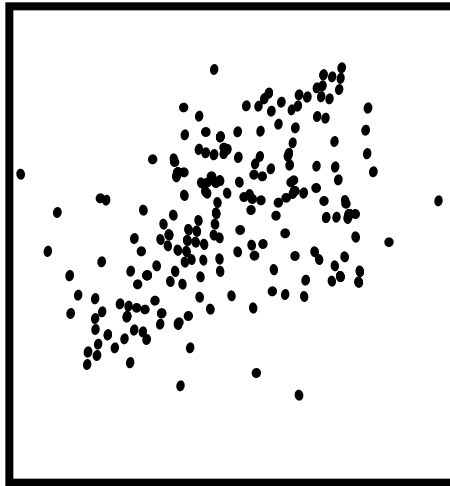
**Perfect
correlation**

Correlation Coefficient*

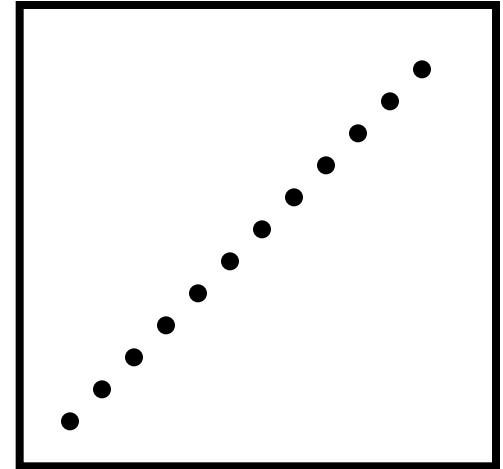
$$r = \frac{\sum(x_i - \mu_x)(y_i - \mu_y)}{\sqrt{\sum(x_i - \mu_x)^2 \sum(y_i - \mu_y)^2}}$$



$r = 0.85$



$r = 0.4$

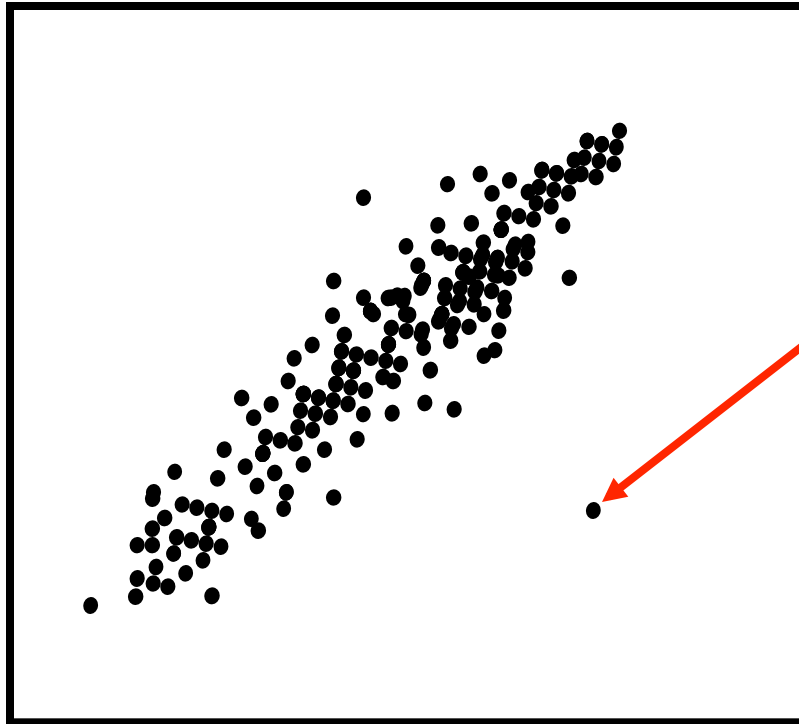


$r = 1.0$

Correlation Coefficient

- **Sometimes called coefficient of linear correlation or Pearson product-moment correlation coefficient**
- **A quantitative way of determining what model (or equation or type of line) best fits a set of data**
- **Commonly used to assess most kinds of predictions or simulations**

Correlation and Outliers



Experimental error or something important?

A single “bad” point can destroy a good correlation

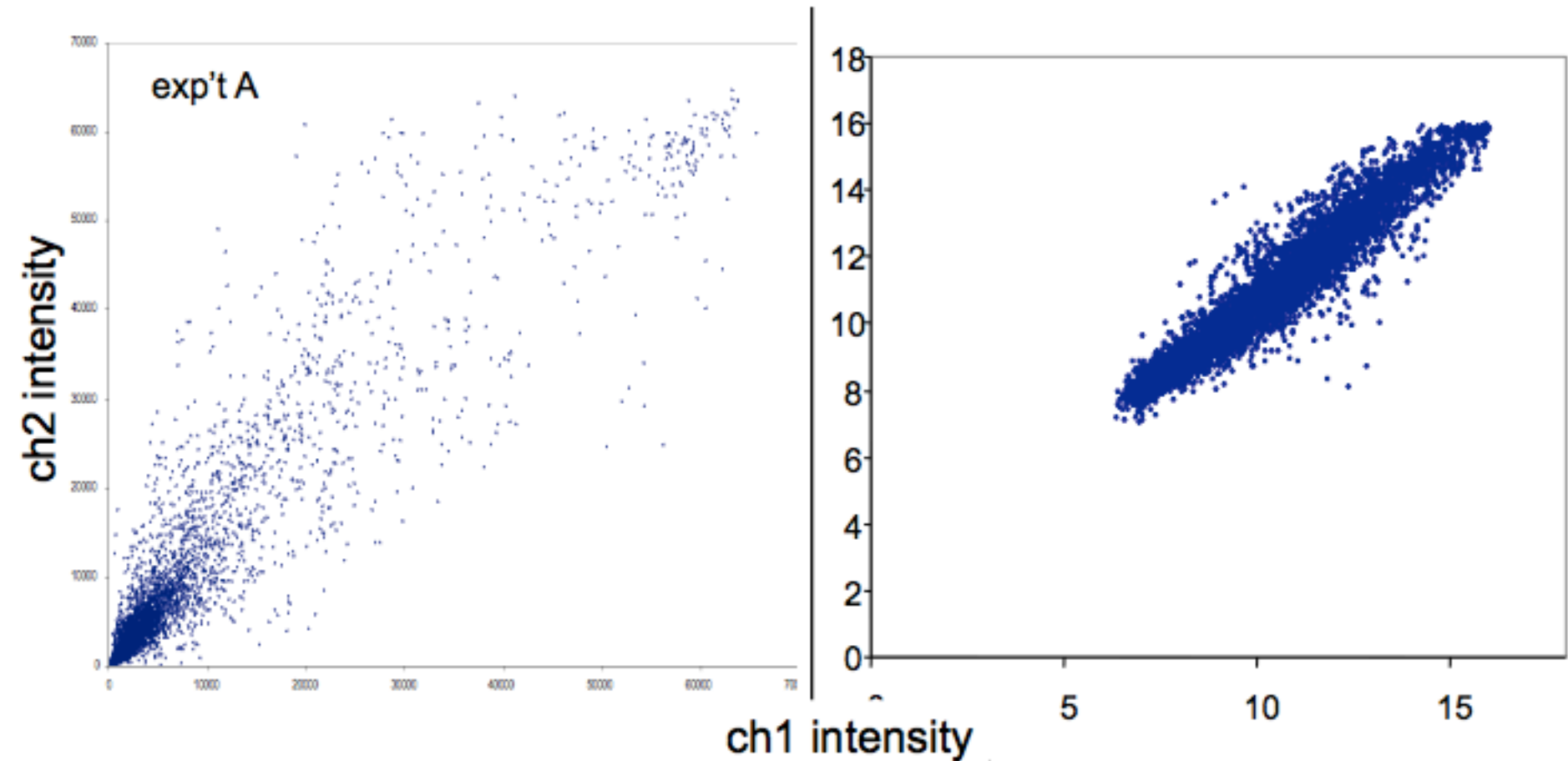
Outliers*

- Can be both “good” and “bad”
- When modeling data -- you don't like to see outliers (suggests the model is bad)
- Often a good indicator of experimental or measurement errors -- only you can know!
- When plotting gel or microarray expression data you do like to see outliers
- A good indicator of something **significant**

Log Transformation*

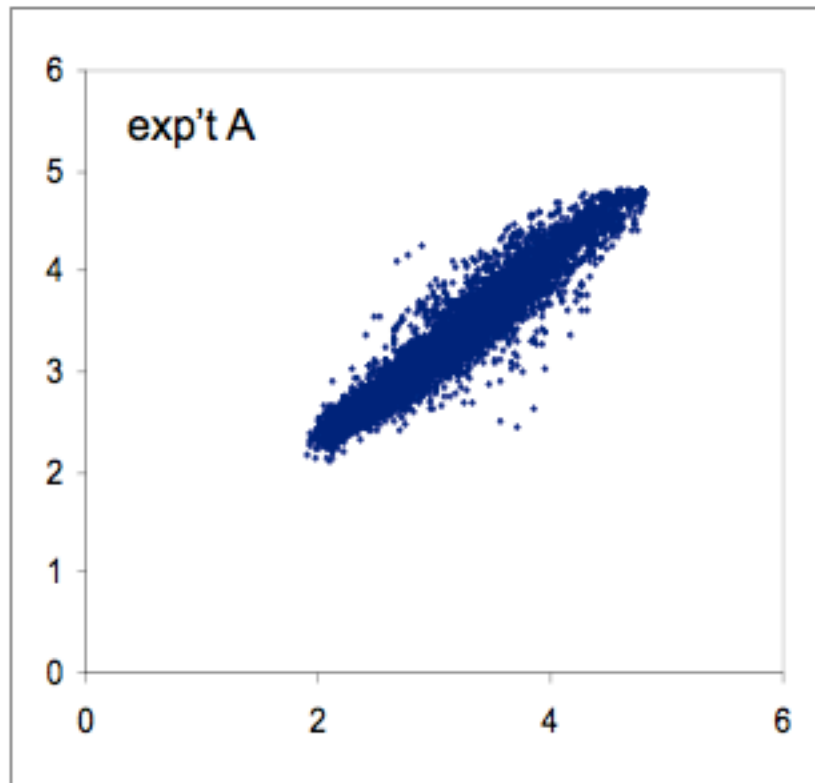
linear scale

\log_2 scale

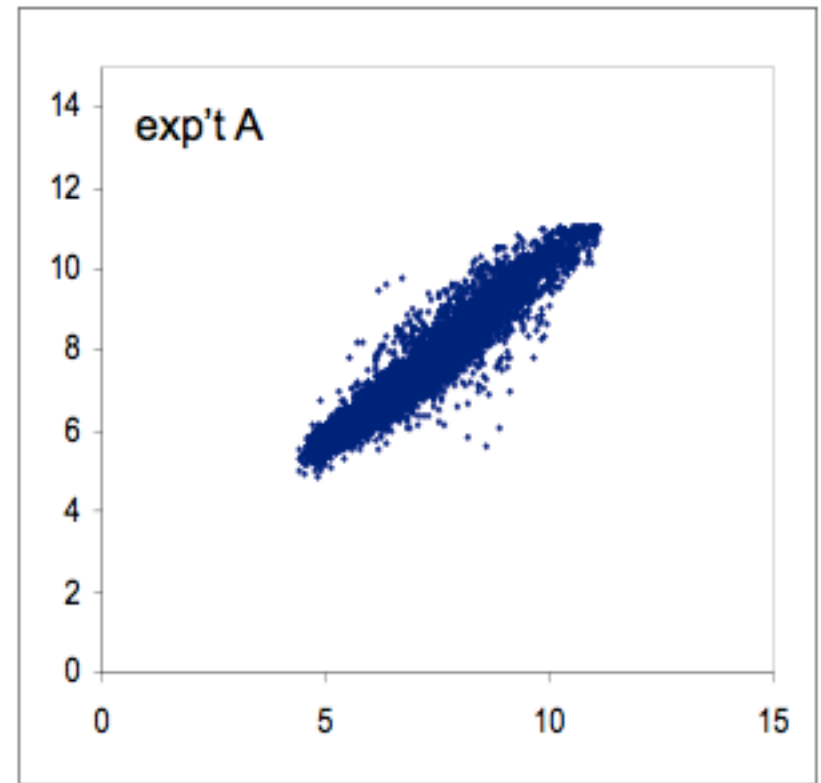


Choice of Base is Not Important

\log_{10}



\ln



Why Log_2 Transformation?*

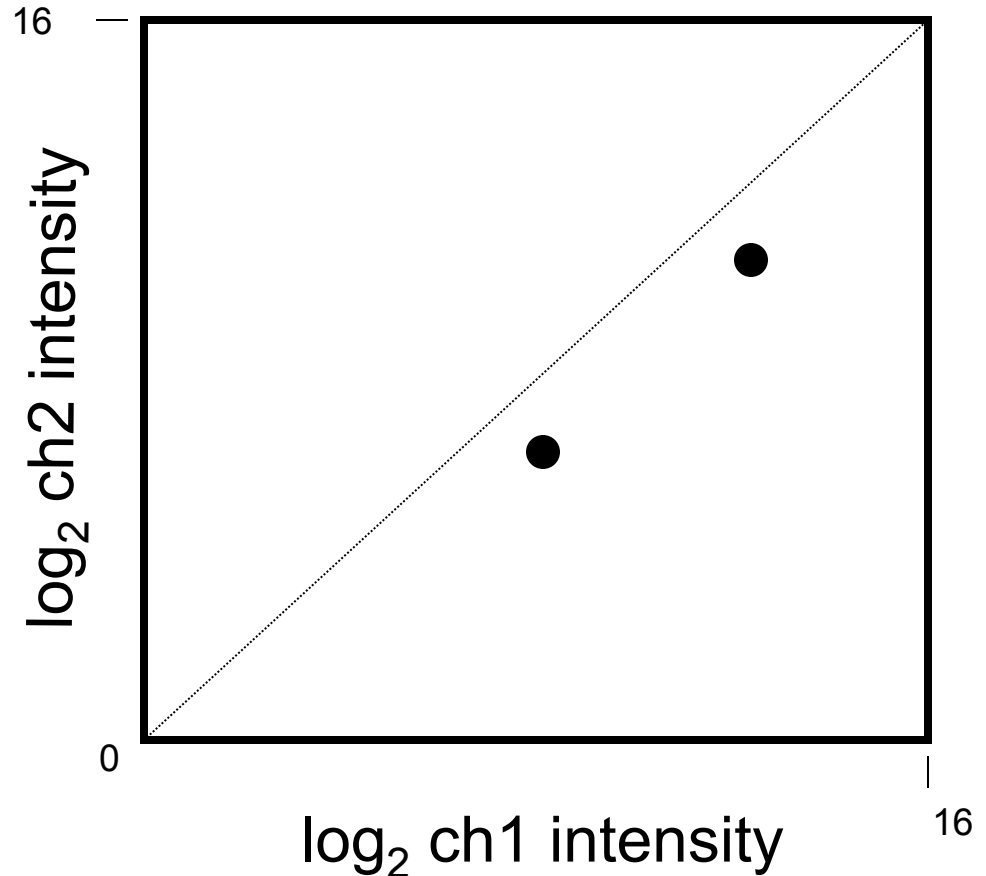
- **Makes variation of intensities and ratios of intensities more independent of absolute magnitude**
- **Makes normalization additive**
- **Evens out highly skewed distributions**
- **Gives more realistic sense of variation**
- **Approximates normal distribution**
- **Treats up- and down- regulated genes symmetrically**

Log Transformations

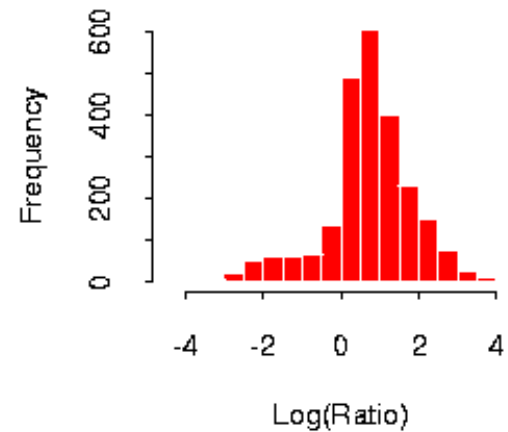
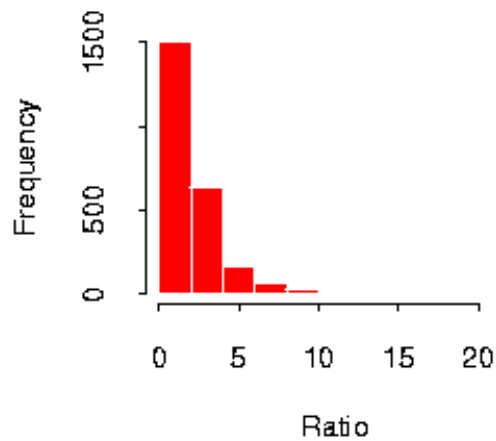
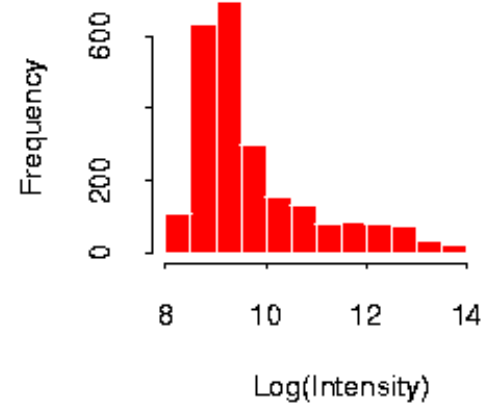
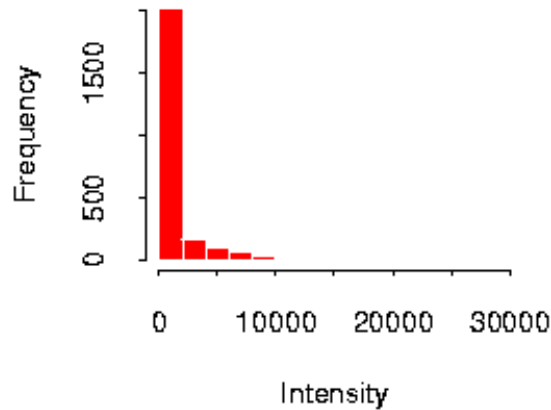
Applying a **log transformation** makes the variance and offset more proportionate along the entire graph

<u>ch1</u>	<u>ch2</u>	<u>ch1/ch2</u>
60 000	40 000	1.5
3000	2000	1.5

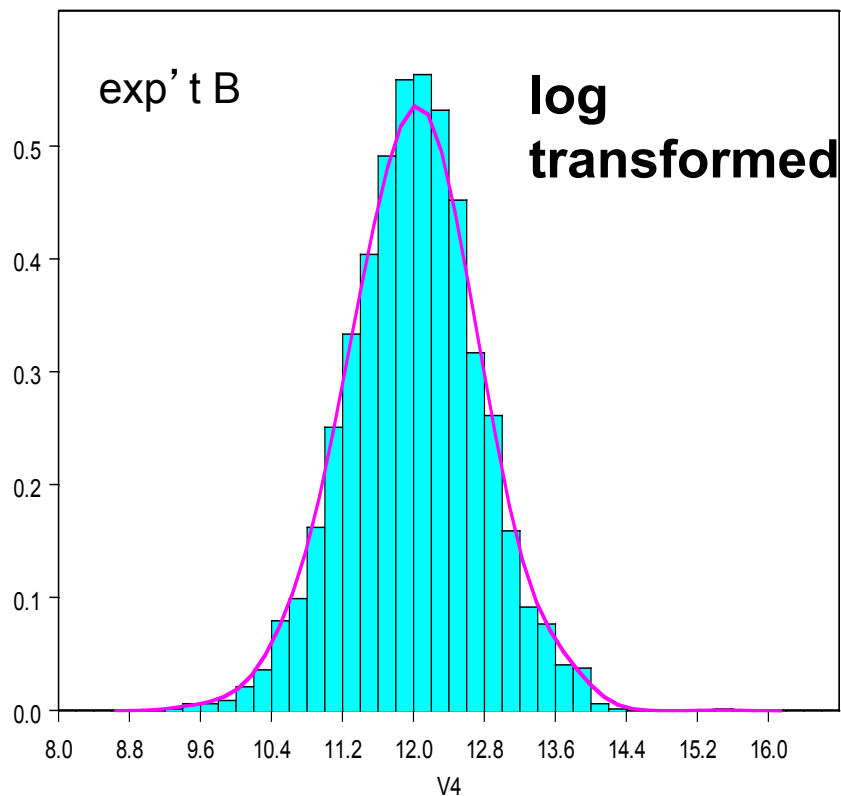
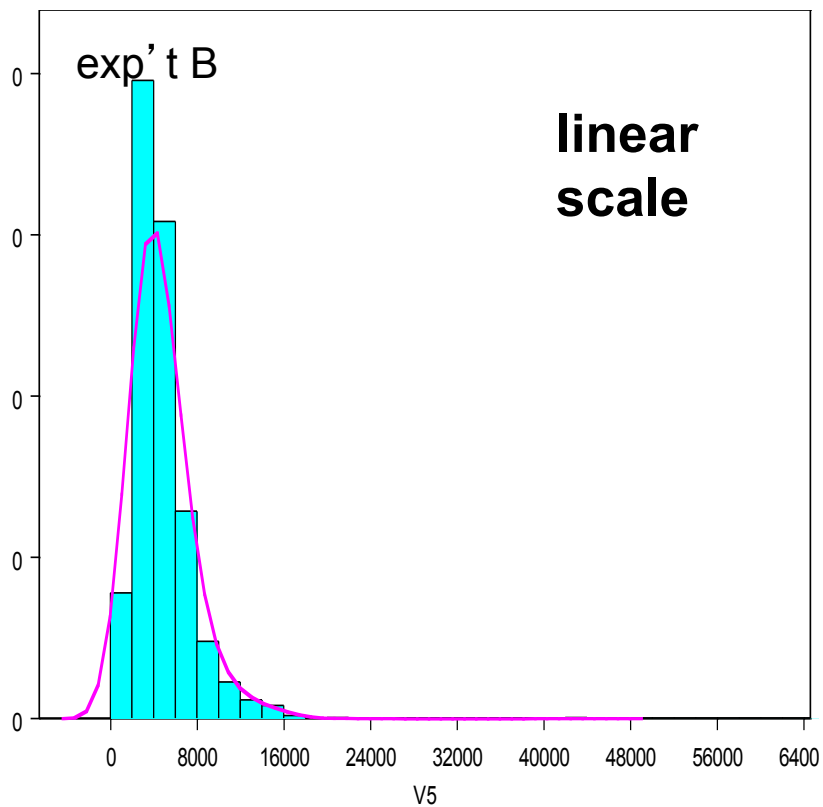
<u>log₂ ch1</u>	<u>log₂ ch2</u>	<u>log₂ ratio</u>
15.87	15.29	0.58
11.55	10.97	0.58



Log Transformations*



Log Transformation

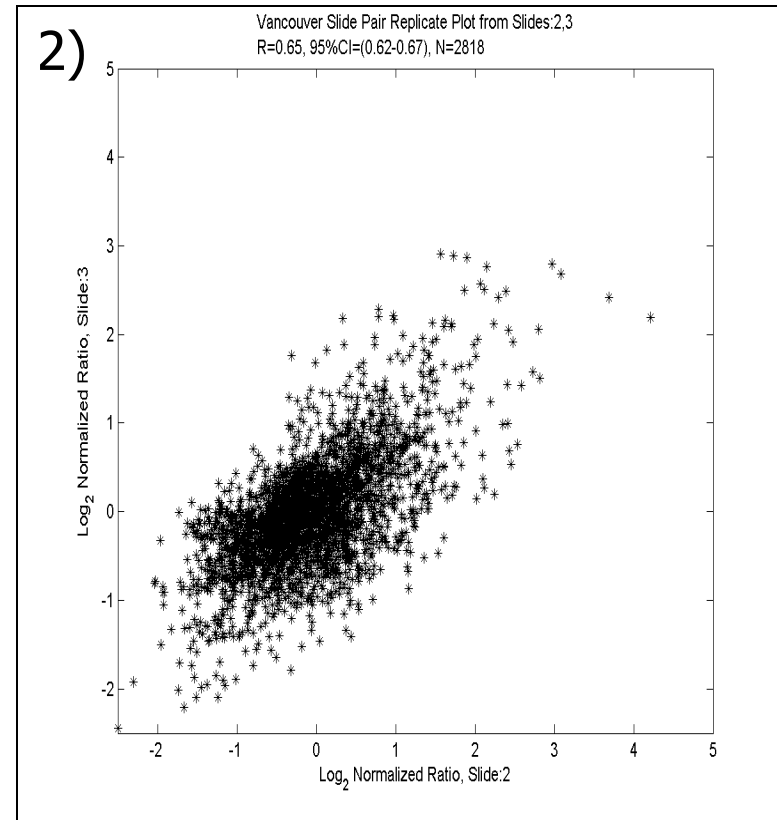
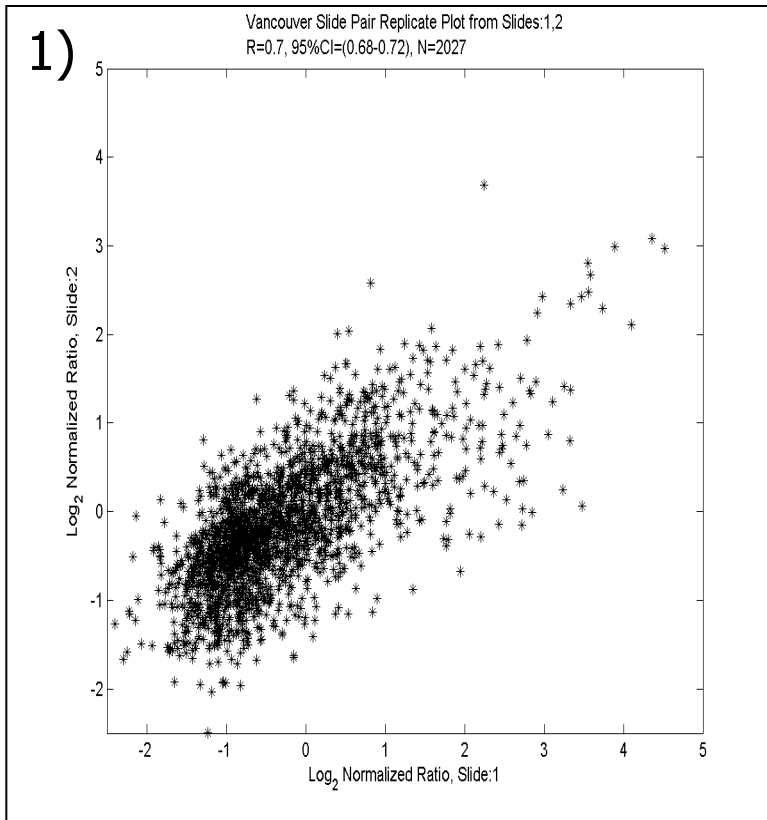


Normalization*

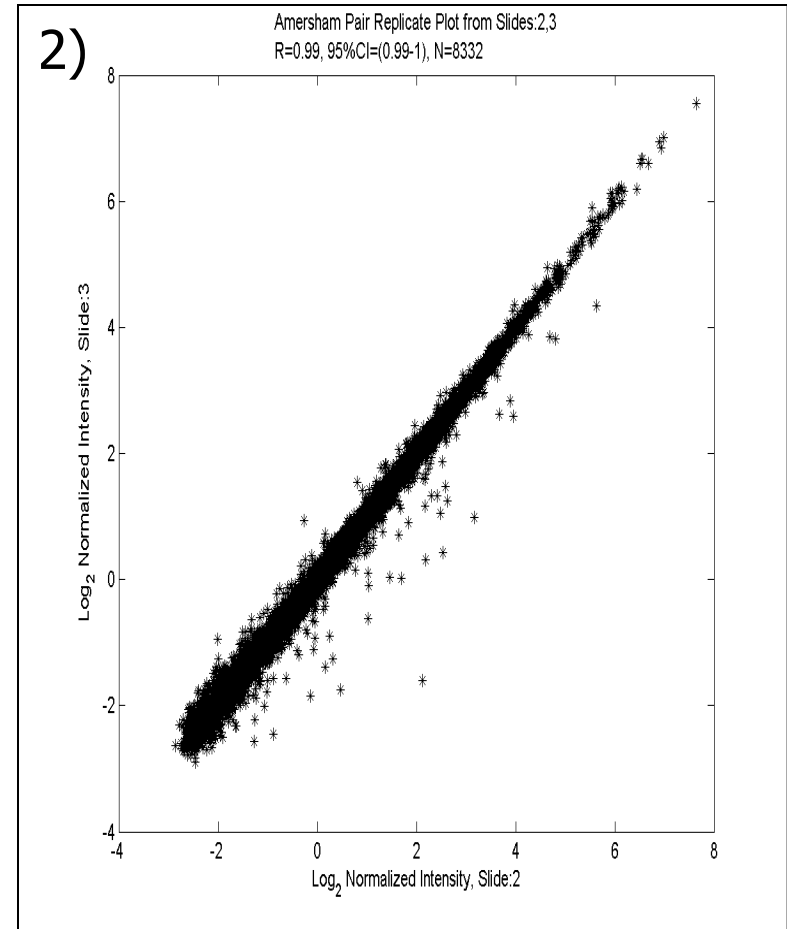
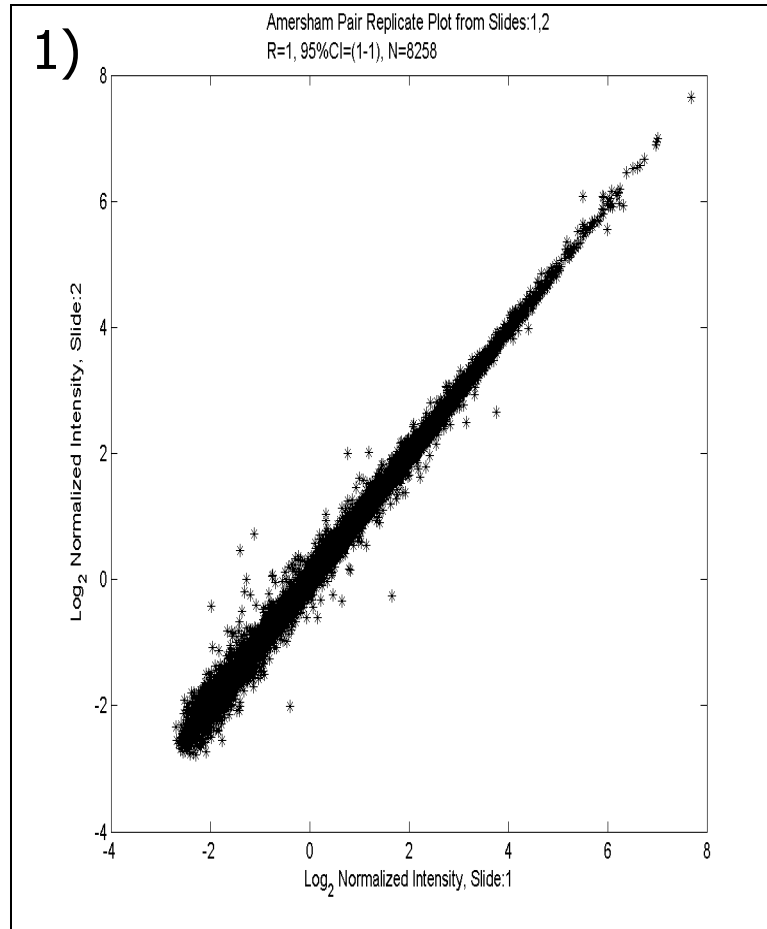
- **Reduces systematic (multiplicative) differences between two channels of a single hybridization or differences between hybridizations**
- **Several Methods:**
 - **Global mean method**
 - **(Iterative) linear regression method**
 - **Curvilinear methods (e.g. loess)**
 - **Variance model methods**

Try to get a slope ~ 1 and a correlation of ~ 1

Example Where Normalization is Needed



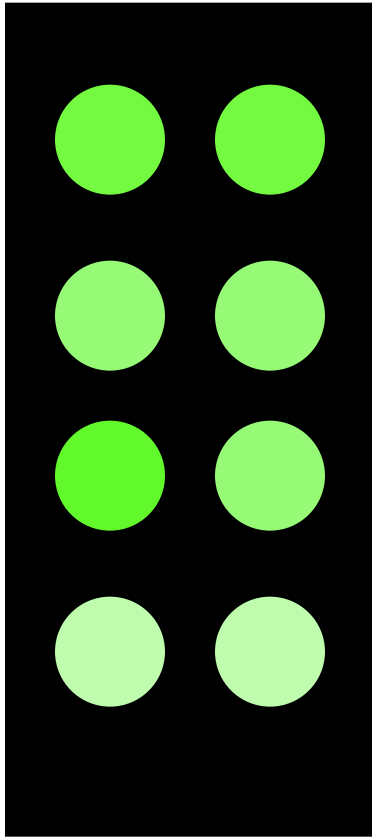
Example Where Normalization is Not Needed



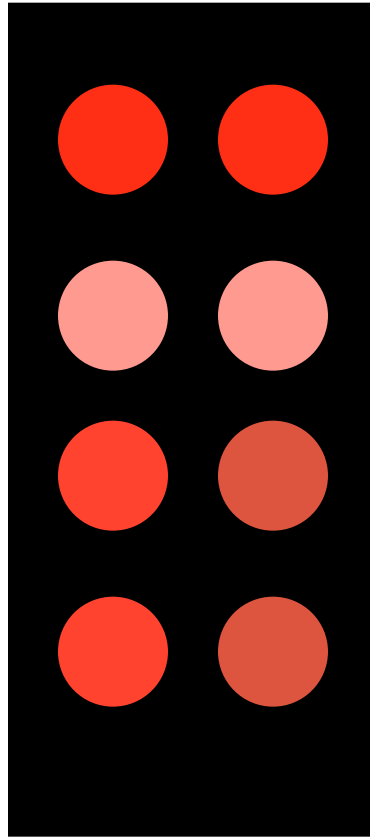
Normalization to a Global Mean*

- **Calculate mean intensity of all spots in ch1 and ch2**
 - e.g. $\mu_{ch2} = 25\ 000$ $\mu_{ch2}/\mu_{ch1} = 1.25$
 - $\mu_{ch1} = 20\ 000$
- **On average, spots in ch2 are 1.25X brighter than spots in ch1**
- **To normalize, multiply spots in ch1 by 1.25**

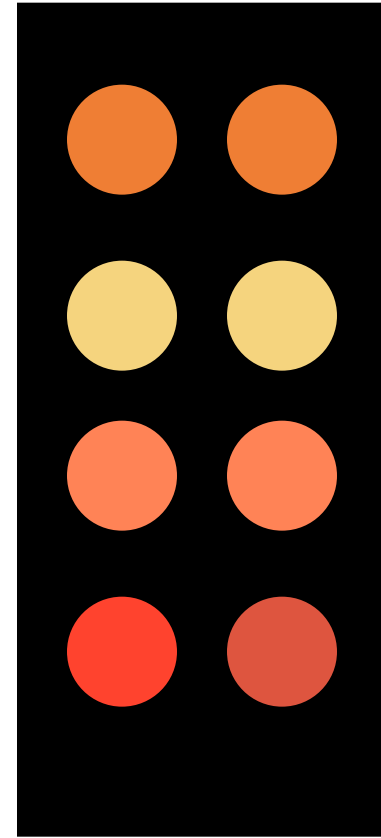
Visual Example: Ch2 is too Strong



Ch 1

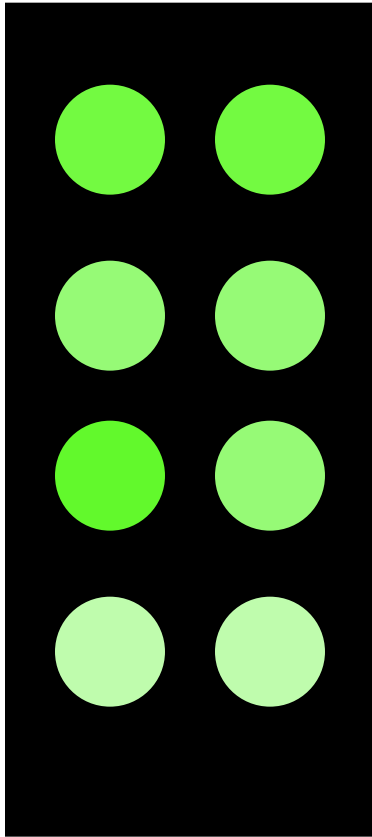


Ch 2

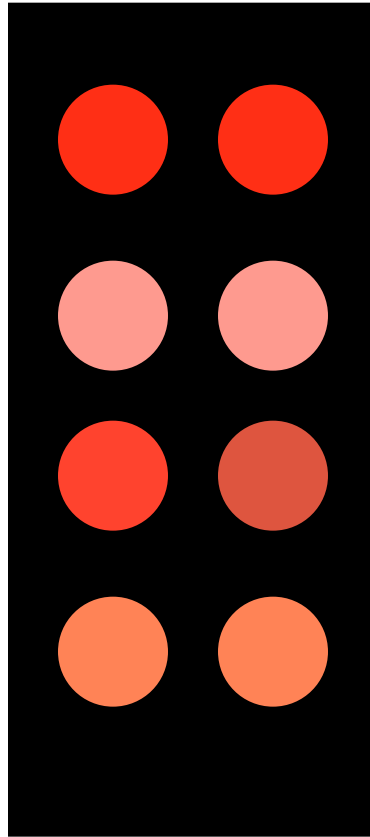


Ch1 + Ch2

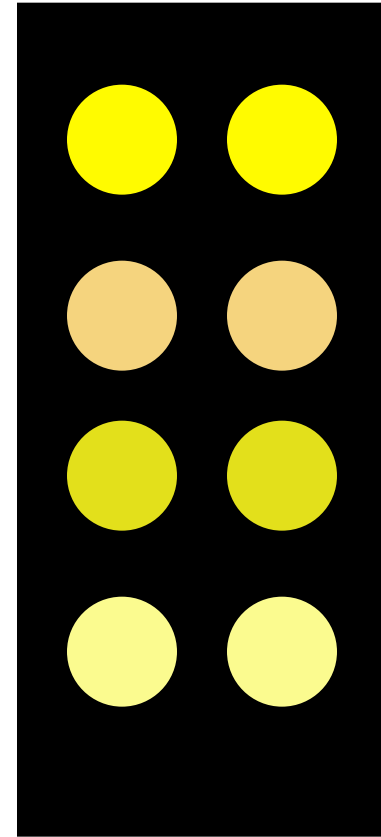
Visual Example: Ch2 and Ch1 are Balanced



Ch 1

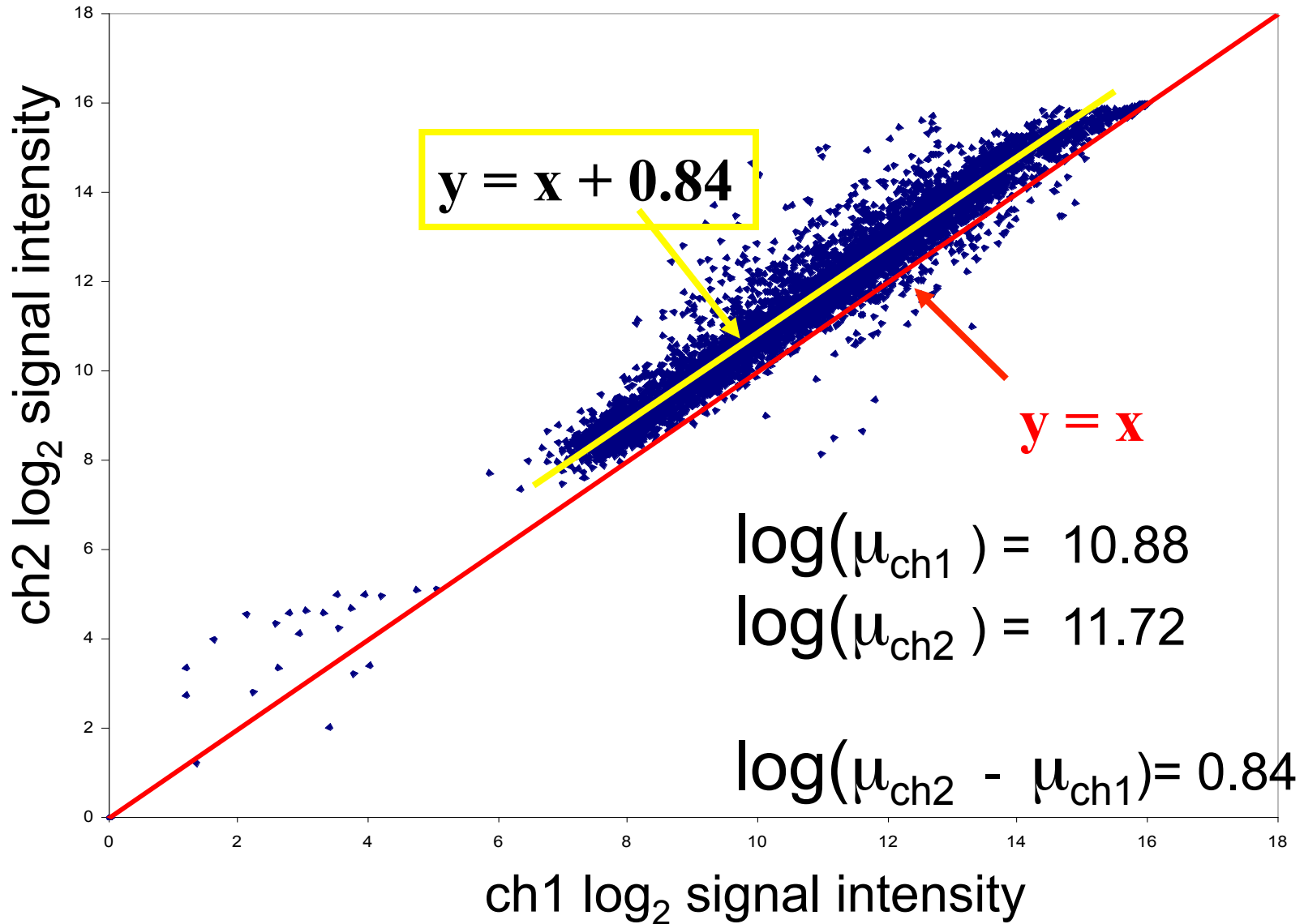


Ch 2

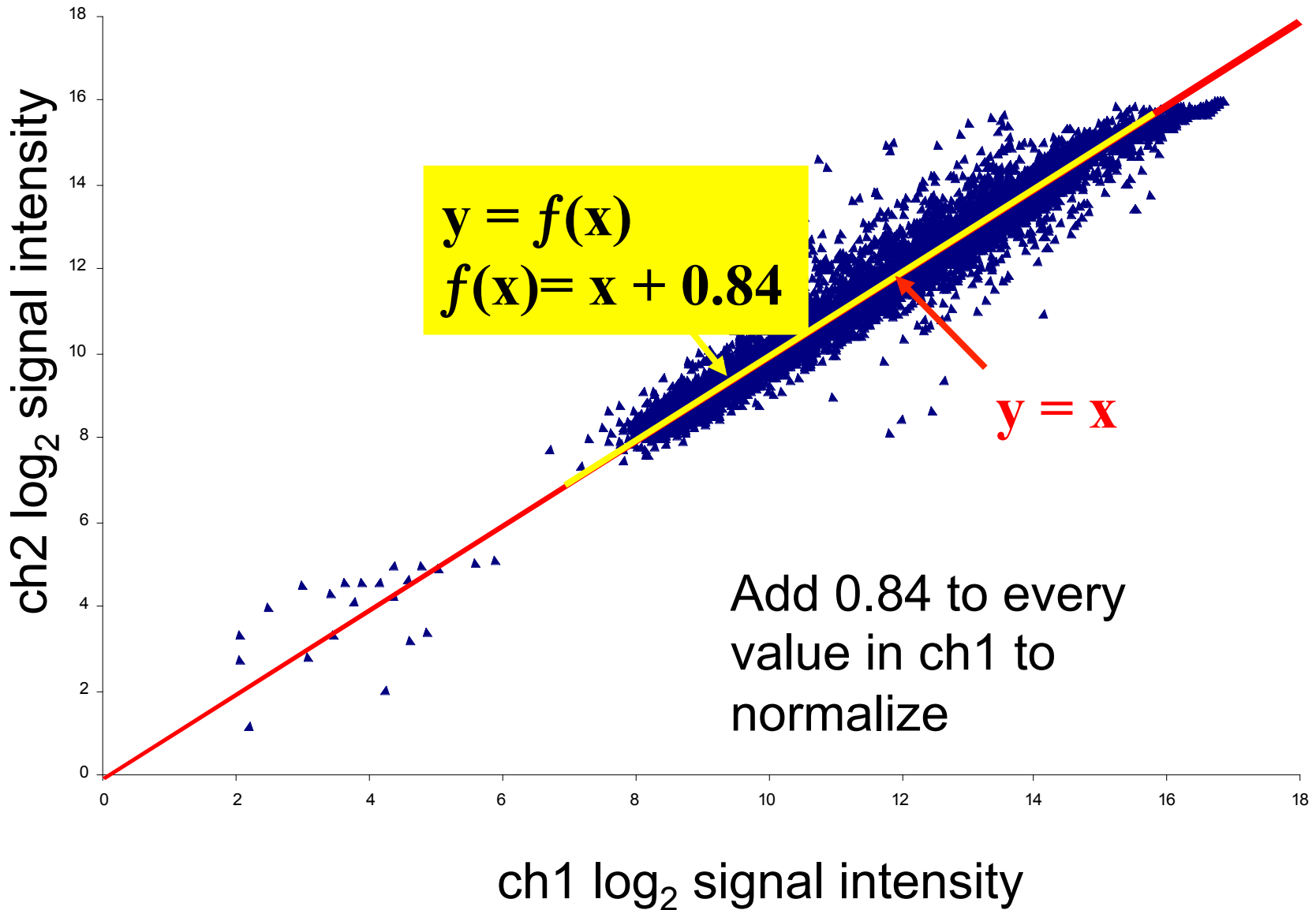


Ch1 + Ch2

Pre-normalized Data



Normalized Microarray Data

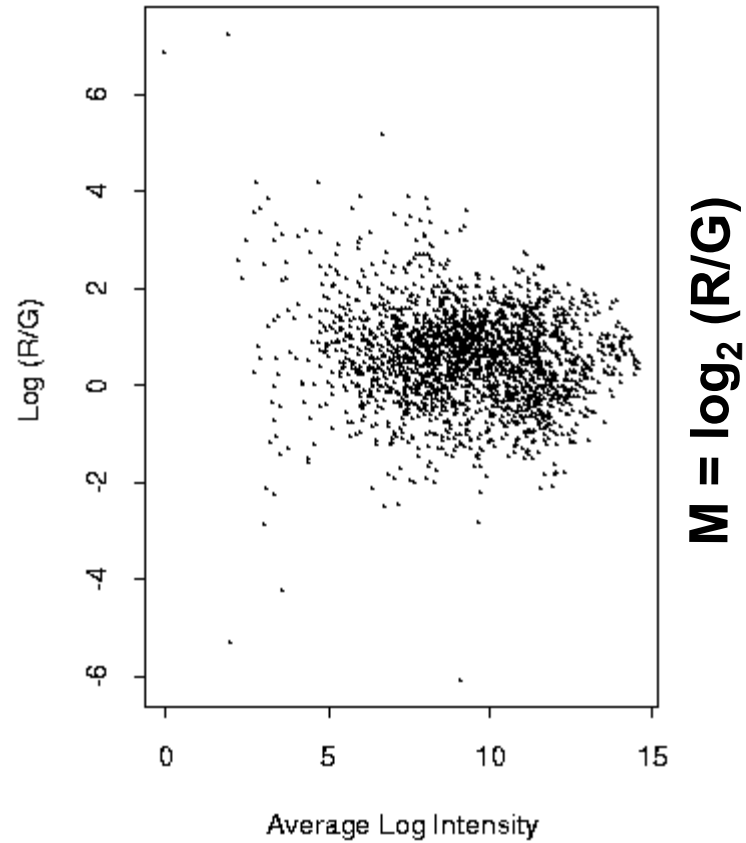
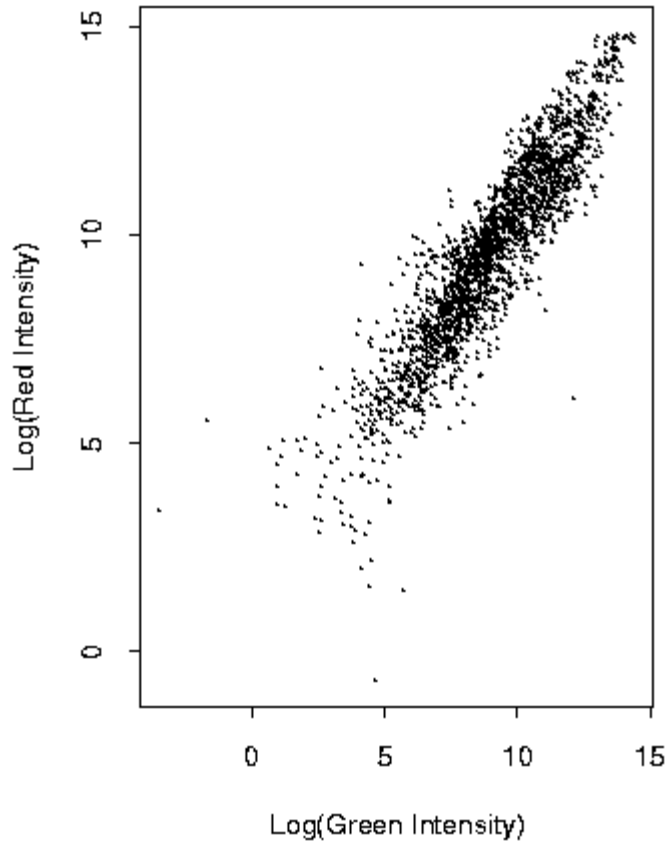


Normalization to Loess Curve*

- **A curvilinear form of normalization**
- **For each spot, plot ratio vs. mean (ch1,ch2) signal in log scale (A vs. M)**
- **Use statistical programs (e.g. S-plus, SAS, or R) to fit a loess curve (local regression) through the data**
- **Offset from this curve is the normalized expression ratio**

The A versus M Plot*

More Informative Graph



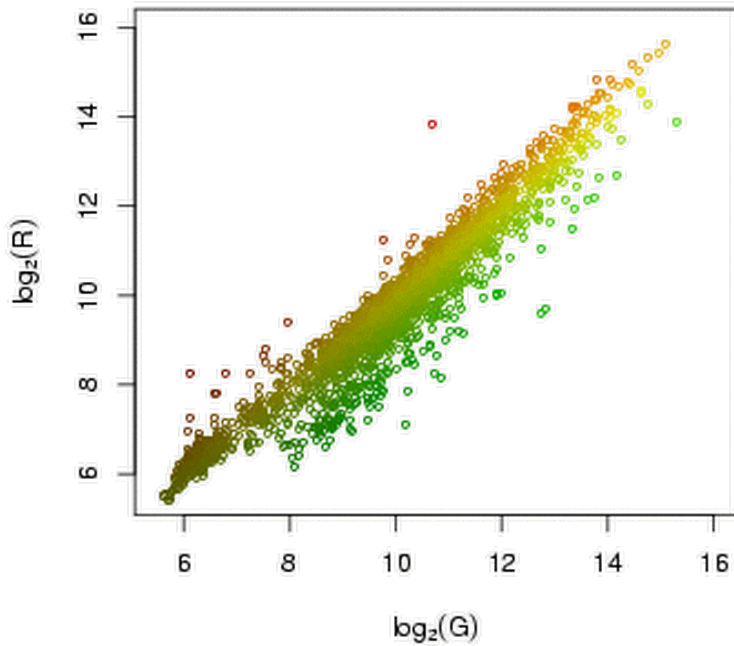
$$A = 1/2 \log_2 (R * G)$$

$$M = \log_2 (R/G)$$

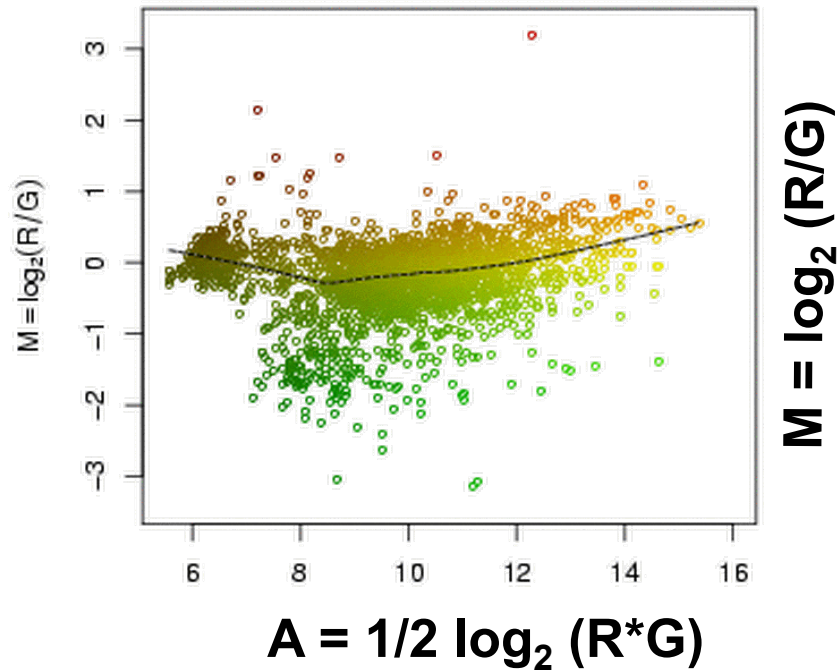
A vs. M Plot

More Informative Graph

dUDG414



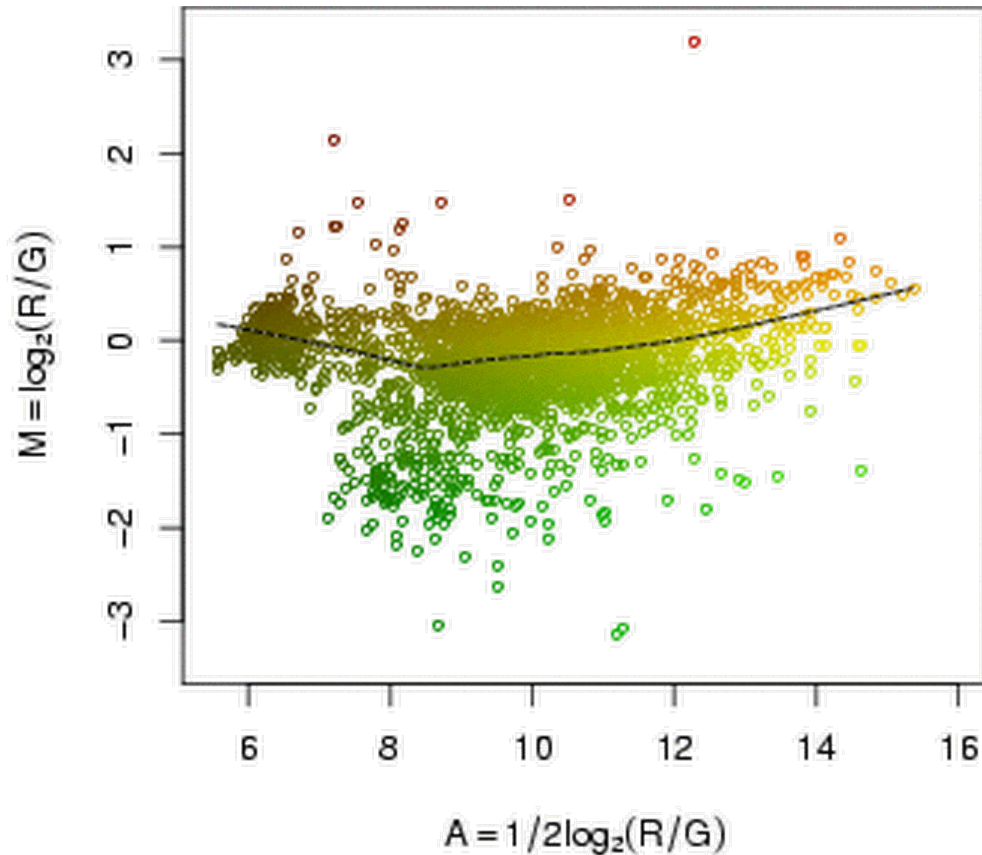
dUDG414



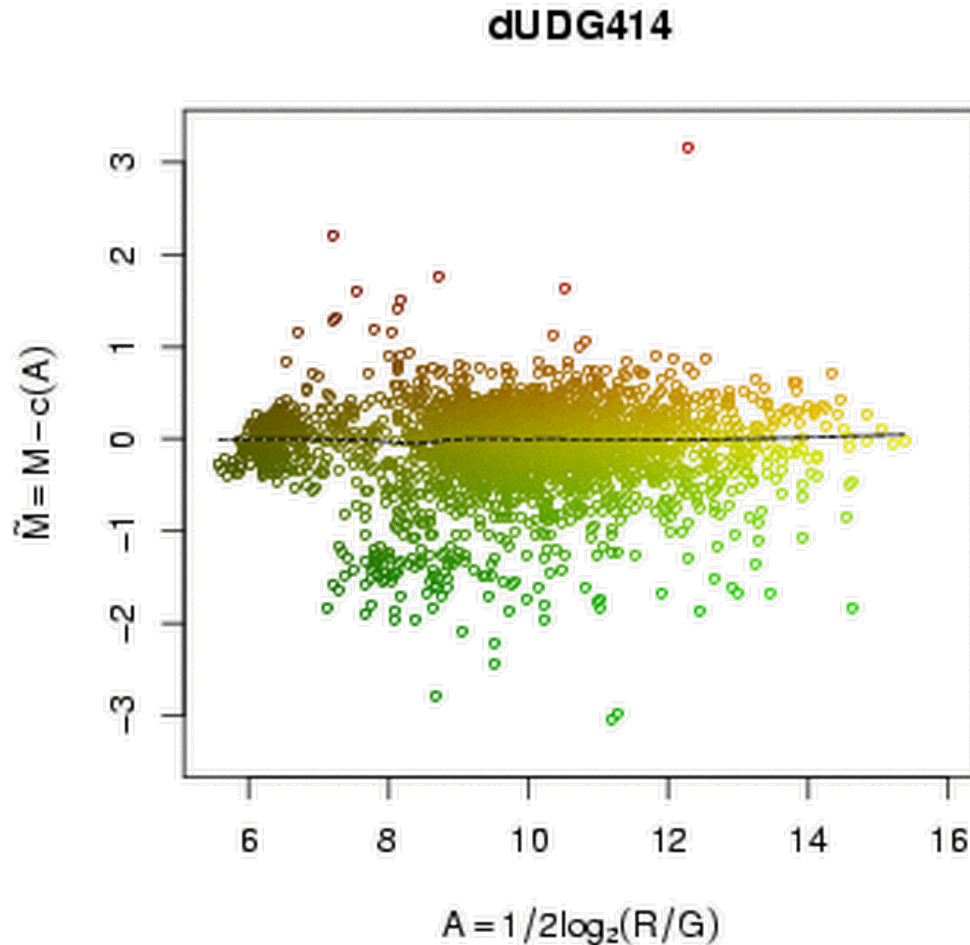
Prior To Normalization

Non-normalized data $\{(M, A)\}_{n=1..5184}$

$$M = \log_2(R/G)$$



Global (Loess) Normalization



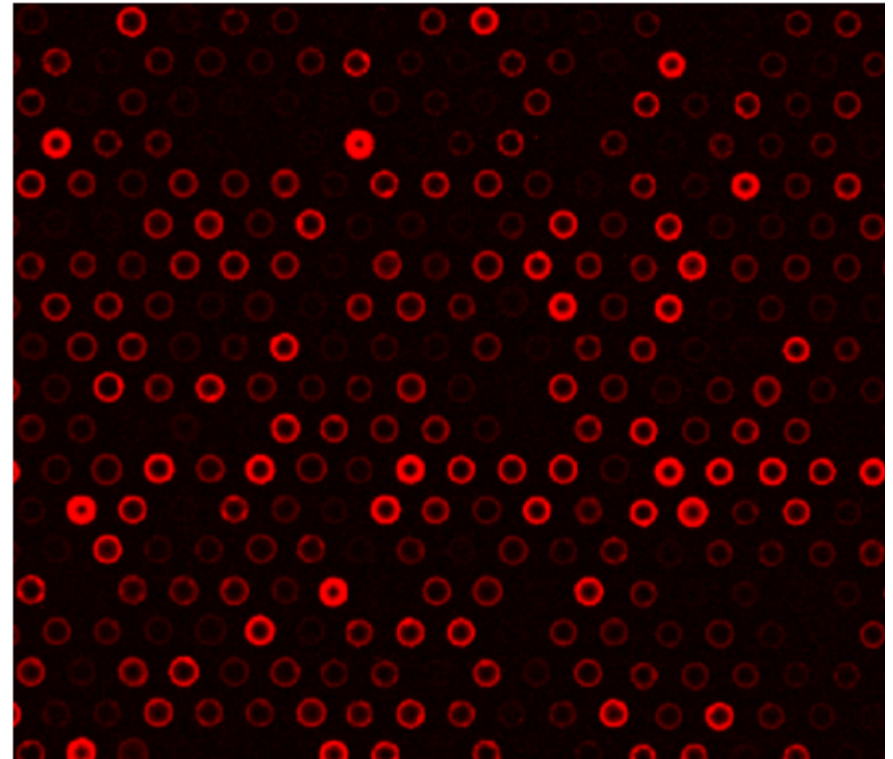
Quality Measurements

- **Array**
 - *Correlation between spot intensities*
 - Percentage of spots with no signals
 - Distribution of spot signal area
 - Inter-array consistency
- **Spot**
 - Signal / Noise ratio
 - Variation in pixel intensities
 - ID of “bad spots” (spots with no signal)

Quality Assessment



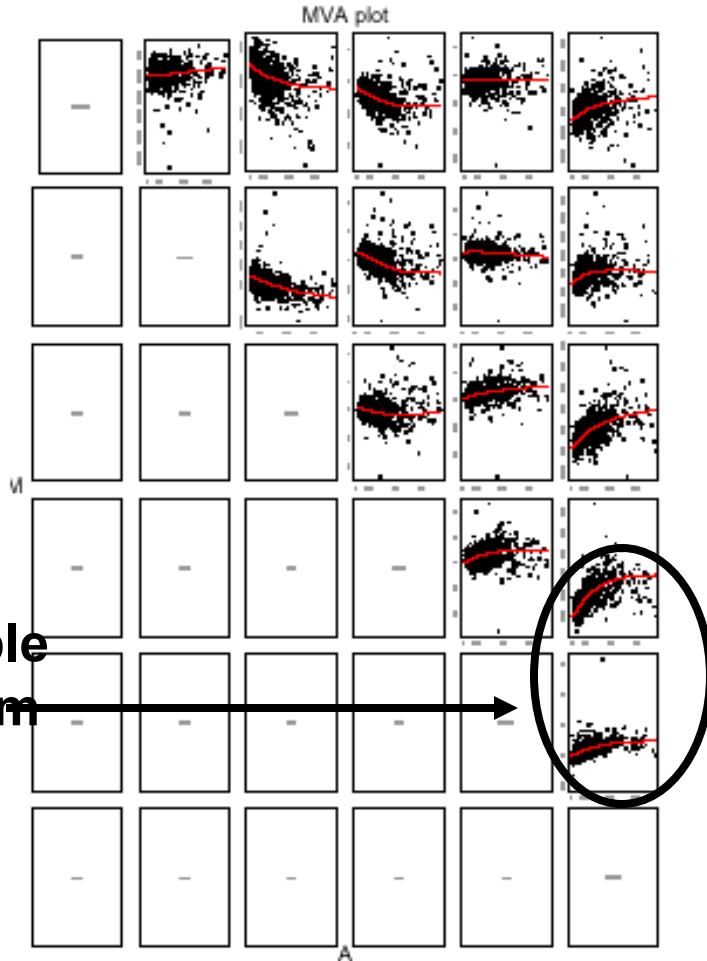
OK quality



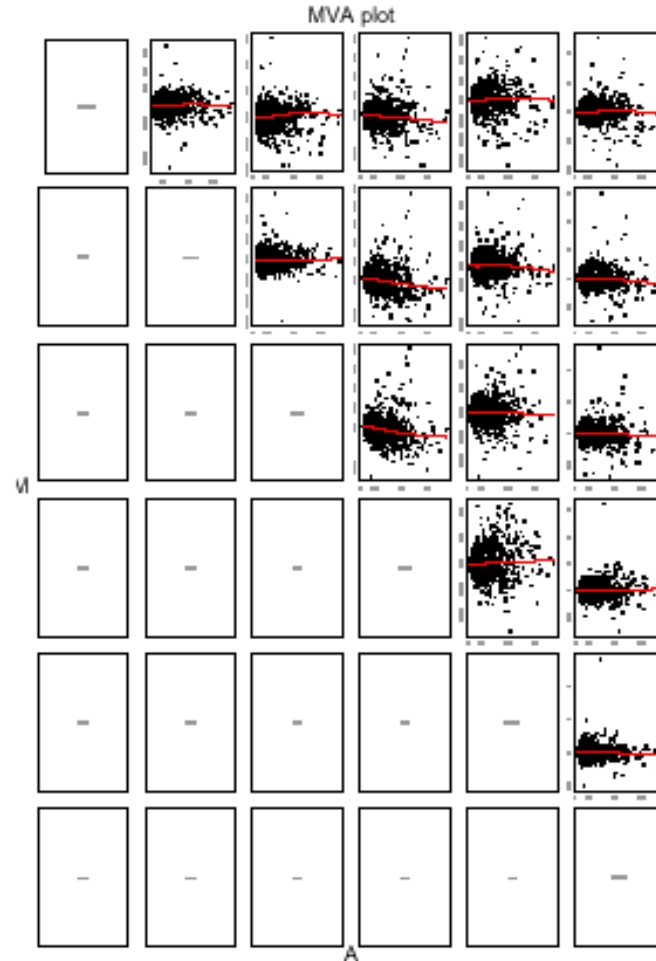
High quality

Inter-Array Consistency*

Pre-normalized



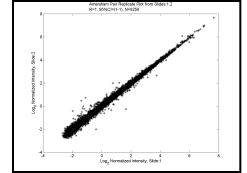
Normalized



Quality Assessment

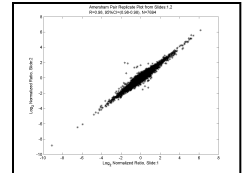
High Quality Array

1)	$R=1$	95% CI=(1-1)	N=8258
2)	$R=0.99$	95% CI=(0.99-1)	N=8332
3)	$R=0.99$	95% CI=(0.99-0.99)	N=8290



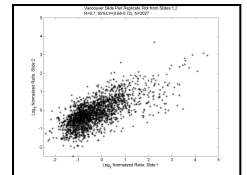
High Quality Array

1)	$R=0.98$	95% CI=(0.98-0.98)	N=7694
2)	$R=0.97$	95% CI=(0.97-0.98)	N=7873
3)	$R=0.97$	95% CI=(0.97-0.97)	N=7694



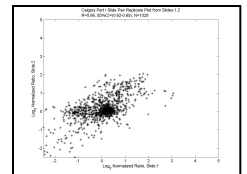
Good Quality Array

1)	$R=0.7$	95% CI=(0.68-0.72)	N=2027
2)	$R=0.65$	95% CI=(0.62-0.67)	N=2818
3)	$R=0.61$	95% CI=(0.59-0.64)	N=2001



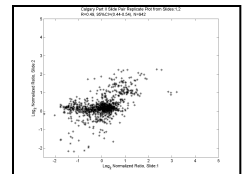
Poor Quality Array

1)	$R=0.66$	95% CI=(0.62-0.69)	N=1028
2)	$R=0.86$	95% CI=(0.85-0.87)	N=1925
3)	$R=0.64$	95% CI=(0.61-0.68)	N=1040



Poor Quality Array

1)	$R=0.49$	95% CI=(0.44-0.54)	N=942
2)	$R=0.81$	95% CI=(0.8-0.83)	N=1700
3)	$R=0.57$	95% CI=(0.52-0.61)	N=973



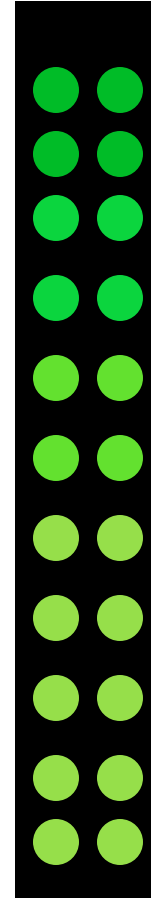
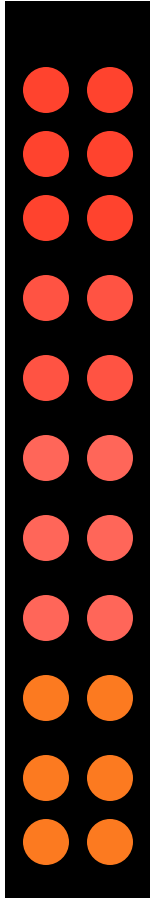
Final Result

Highly Exp

Trx	16.8
Enh1	13.2
Hin2	11.8
P53	8.4
Calm	7.3
Ned3	5.6
P21	5.5
Antp	5.4
Gad2	5.2
Gad3	5.1
Erp3	5.0

Reduced Exp

GPD	0.11
Shn2	0.13
Alp4	0.22
OncB	0.23
Nrd1	0.25
LamR	0.26
SetH	0.30
LinK	0.32
Mrd2	0.32
Mrd3	0.33
TshR	0.34



Fold change

Key Steps in Microarray Analysis

- **Quality Control (checking microarrays for errors or problems)**
- **Image Processing**
 - **Gridding**
 - **Segmentation (peak picking)**
 - **Data Extraction (intensity, QC)**
- **Data Analysis and Data Mining (Differential gene expression)**

Identifying Patterns of Gene Expression*

- **Key Goal:** identify differentially & co-regulated groups of genes via clustering
- **This leads to:**
 - inferences about physiological responses
 - generalizations about large data sets
 - identification of regulatory cascades
 - assignment of possible function to uncharacterized genes
 - identification of shared regulatory motifs

Clustering Applications in Bioinformatics*

- **Microarray or GeneChip Analysis**
- **2D Gel or ProteinChip Analysis**
- **Protein Interaction Analysis**
- **Phylogenetic and Evolutionary Analysis**
- **Structural Classification of Proteins**
- **Protein Sequence Families**

Clustering*

- **Definition - a process by which objects that are logically similar in characteristics are grouped together.**
- **Clustering is different than Classification**
- **In classification the objects are assigned to pre-defined classes, in clustering the classes are yet to be defined**
- **Clustering helps in classification**

Clustering Requires...

- **A method to measure similarity (a similarity matrix) or dissimilarity (a dissimilarity coefficient) between objects**
- **A threshold value with which to decide whether an object belongs with a cluster**
- **A way of measuring the “distance” between two clusters**
- **A cluster seed (an object to begin the clustering process)**

Clustering Algorithms*

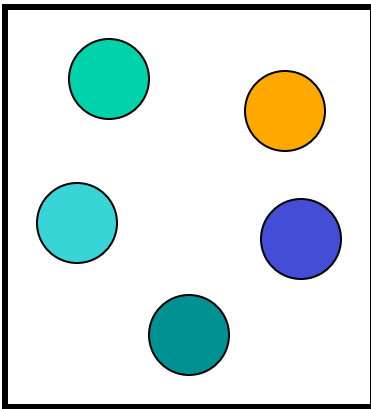
- **K-means or Partitioning Methods** - divides a set of N objects into M clusters -- with or without overlap
- **Hierarchical Methods** - produces a set of nested clusters in which each pair of objects is progressively nested into a larger cluster until only one cluster remains
- **Self-Organizing Feature Maps** - produces a cluster set through iterative “training”

Hierarchical Clustering*

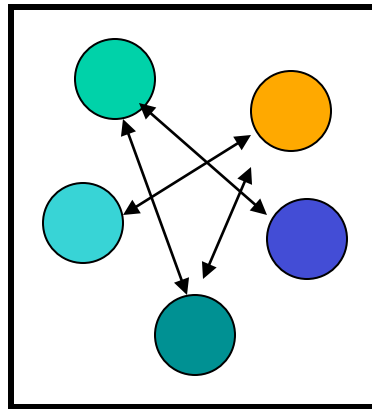
- *Find the two closest objects and merge them into a cluster*
- *Find and merge the next two closest objects (or an object and a cluster, or two clusters) using some similarity measure and a predefined threshold*
- *If more than one cluster remains return to step 2 until finished*

Hierarchical Clustering*

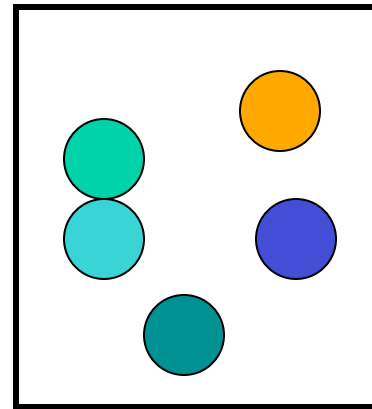
Initial cluster



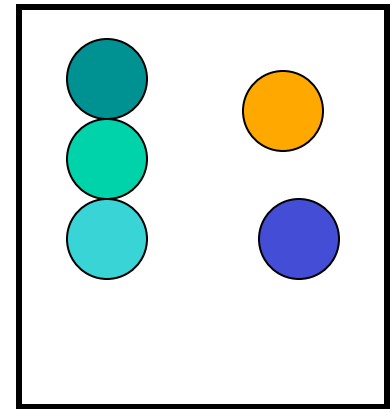
pairwise compare



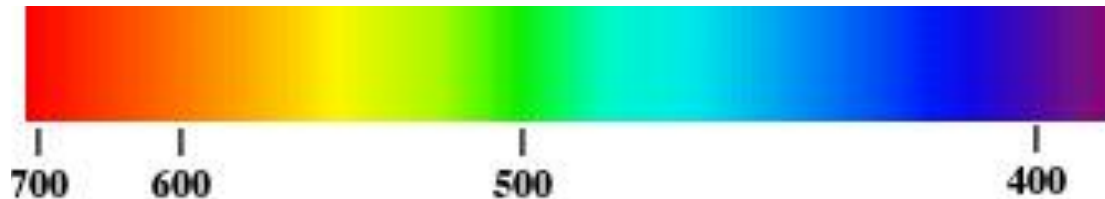
select closest



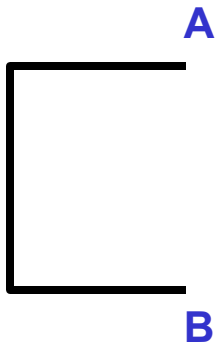
select next closest



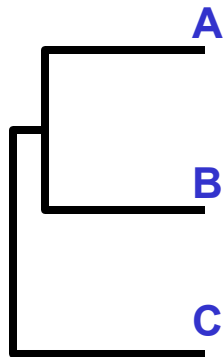
$$\text{Rule: } \lambda_T = \lambda_{\text{obs}} \pm 50 \text{ nm}$$



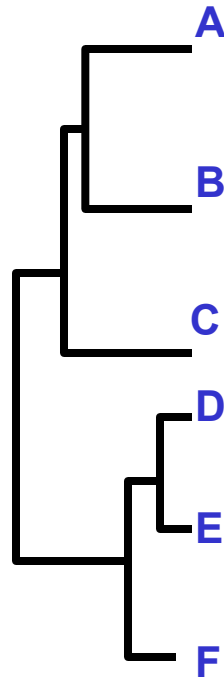
Hierarchical Clustering*



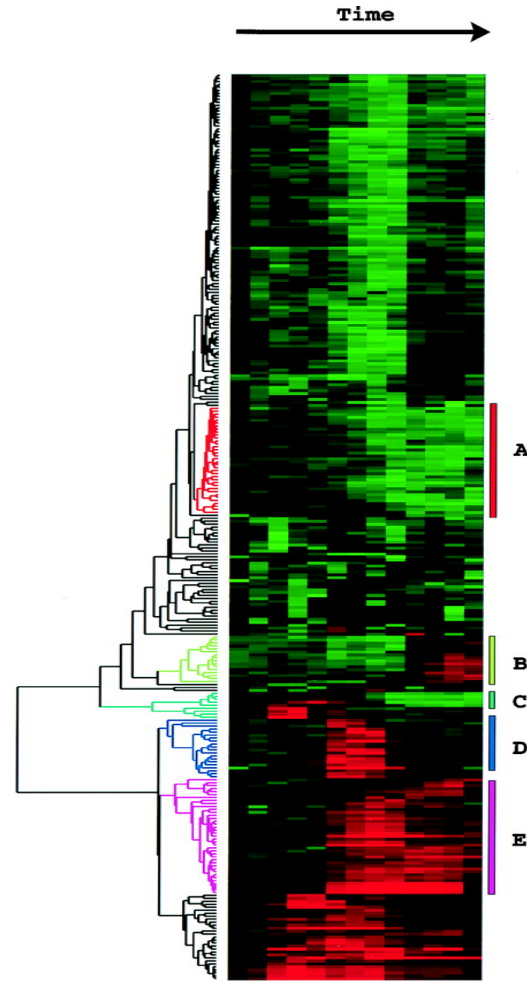
Find 2 most similar gene express levels or curves



Find the next closest pair of levels or curves

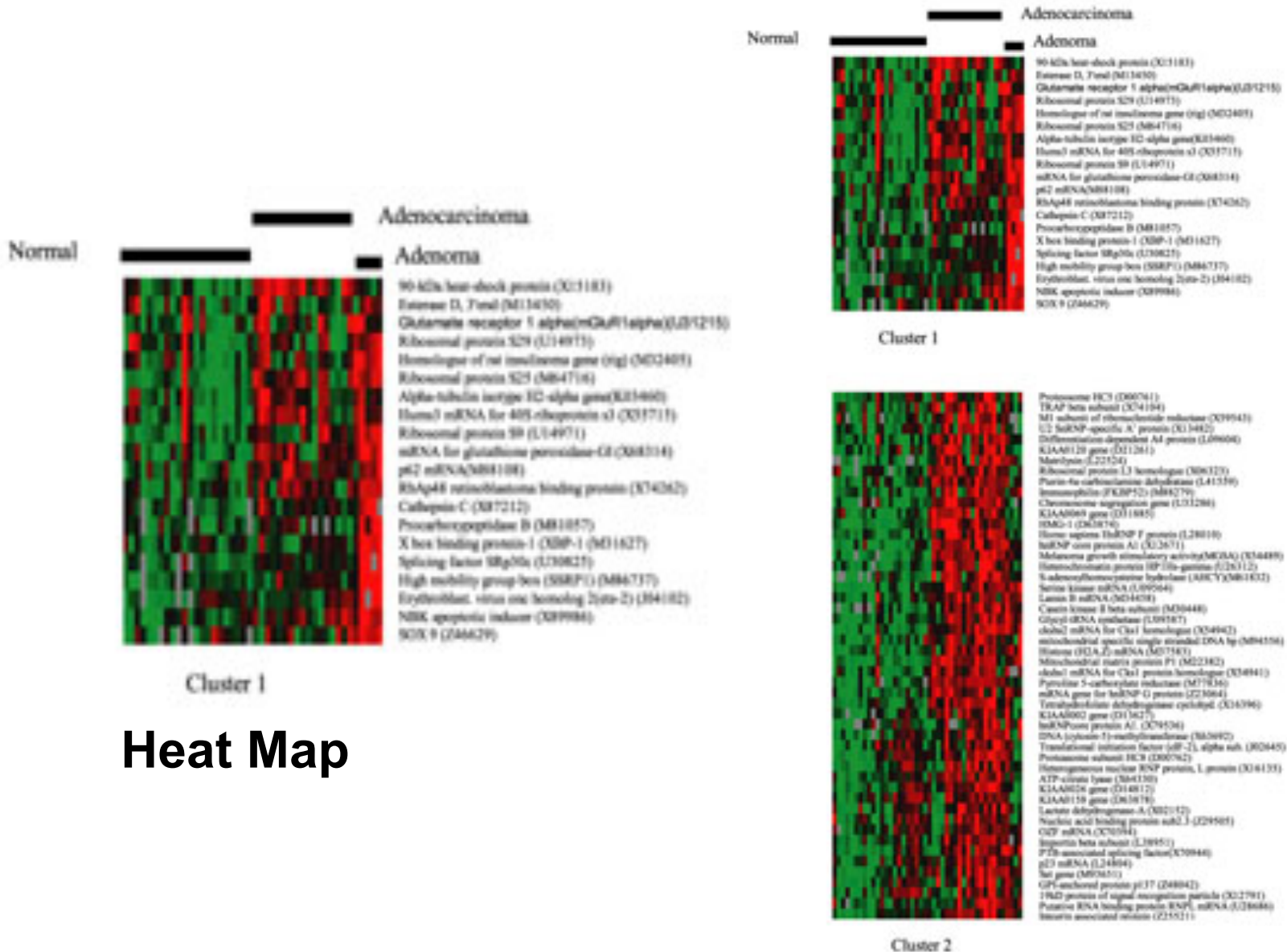


Iterate



Heat map

Results



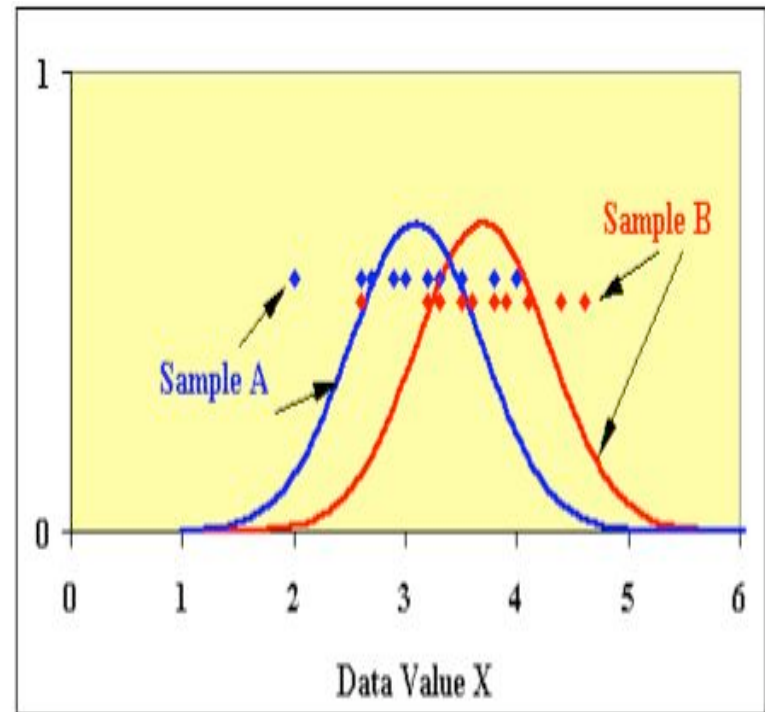
Heat Map

Putting it All Together*

- **Perform normalization**
- **Determine if experiment is a time series, a two condition or a multi-condition experiment**
- **Calculate level of differential expression and identify which genes are significantly ($p < 0.05$ using a t-test) over-expressed or under expressed (a 2 fold change or more)**
- **Use clustering methods and heat maps to identify unusual patterns or groups that associate with a disease state or conditions**
- **Interpret the results in terms of existing biological or physiological knowledge**
- **Produce a report describing the results of the analysis**

The Student's t-test*

- The Student's t-distribution was first introduced by W.S. Gossett in 1908 under the pen name Student
- Used to establish confidence limits (error bars) for the mean estimated from smaller sample sizes
- Used to test the statistical significance of a non-zero mean
- Used to test the statistical significance of the difference between means from two independent samples
- **A p value or t-stat of <0.05 is significant**



GEO2R

The screenshot shows a web browser window titled "GEO2R - GEO - NCBI". The address bar contains the URL "http://www.ncbi.nlm.nih.gov/geo/geo2r/". The browser's search bar has "Google" entered. The page header includes the NCBI logo on the left and the GEO logo (Gene Expression Omnibus) on the right. A navigation menu contains links for "GEO Publications", "FAQ", "MIAME", and "Email GEO". Below the header, the breadcrumb "NCBI » GEO » GEO2R" is visible, along with a "Login" link. The main content area features a descriptive paragraph: "Use GEO2R to compare two or more groups of Samples in order to identify genes that are differentially expressed across experimental conditions. Results are presented as a table of genes ordered by significance." This is followed by links for "Full instructions" and a "YouTube" icon. A "GEO accession" input field with a "Set" button is present. Below this is a tabbed interface with "GEO2R" selected, and other tabs for "Value distribution", "Options", "Profile graph", and "R script". A "Quick start" section contains a bulleted list of instructions: 1. Specify a GEO Series accession and a Platform if prompted. 2. Click 'Define groups' and enter names for the groups of Samples you plan to compare, e.g., test and control. 3. Assign Samples to each group. Highlight Sample rows then click the group name to assign those Samples to the group. Use the Sample metadata (title, source and characteristics) columns to help determine which Samples belong to which group. 4. Click 'Top 250' to perform the calculation with default settings. 5. Results are presented as a table of genes ordered by significance. The top 250 genes are presented and may be viewed as profile graphs. Alternatively, the complete results table may be saved. 6. You may change settings in Options tab. Below the list is a "How to use" section with a "Top 250" button and a "Save all results" link.

<http://www.ncbi.nlm.nih.gov/geo/geo2r/>

GEO2R

- **Web-based GeneChip/Microarray analysis pipeline written in R**
- **Designed to handle microarray data deposited in the GEO (Gene Expression Omnibus) database**
- **Performs relatively simple analysis of microarray data**
- **Generates lots of tables and plots**
- **Supports many different microarray platforms**
- **User-friendly, with several tutorials**

DAVID*

DAVID 2008 Functional Annotation Bioinformatics Microarray Analysis

http://david.abcc.ncifcrf.gov/

Most Visited - Getting Started - Latest Headlines

DAVID 2008 Functional Annotatio... +

DAVID Bioinformatics Resources 2008
National Institute of Allergy and Infectious Diseases (NIAID), NIH

Home | Start Analysis | Shortcut to DAVID Tools | Technical Center | Downloads & APIs | Term of Service | Why DAVID? | About Us

Shortcut to DAVID Tools

- Functional Annotation**
Gene-annotation enrichment analysis, functional annotation clustering, BioCarta & KEGG pathway mapping, gene-disease association, homologue match, ID translation, literature match and [more](#)
- Gene Functional Classification**
Provide a rapid means to reduce large lists of genes into functionally related groups of genes to help unravel the biological content captured by high throughput technologies. [More](#)
- Gene ID Conversion**
Convert list of gene ID/accessions to others of your choice with the most comprehensive gene ID mapping repository. The ambiguous accessions in the list can also be determined semi-automatically. [More](#)
- Gene Name Batch Viewer**
Display gene names for a given gene list; Search functionally related genes within your list or not in your list; Deep links to enriched detailed information. [More](#)

Recommending: [A paper published in Nature Protocols](#) describes step-by-step procedure to use DAVID!

Welcome to DAVID Bioinformatics Resources 2003 - 2009

The Database for Annotation, Visualization and Integrated Discovery (DAVID) 2008 is the [sixth version](#) of our original web-accessible programs. DAVID now provides a comprehensive set of functional annotation tools for investigators to understand biological meaning behind large list of genes. For any given gene list, DAVID tools are able to:

- Identify enriched biological themes, particularly GO terms
- Discover enriched functional-related gene groups
- Cluster redundant annotation terms
- Visualize genes on BioCarta & KEGG pathway maps
- Display related many-genes-to-many-terms on 2-D view.
- Search for other functionally related genes not in the list
- List interacting proteins
- Explore gene names in batch
- Link gene-disease associations
- Highlight protein functional domains and motifs
- Redirect to related literatures
- Convert gene identifiers from one type to another.
- And more

What's Important in DAVID 2008?

- [New requirement to cite DAVID](#)
- [IDs of Affy Exon and Gene arrays supported](#)
- [Novel Classification Algorithms](#)
- [Pre-built Affymetrix and Illumina backgrounds](#)
- [User's customized gene background](#)
- [Enhanced calculating speed](#)

Statistics of DAVID

DAVID Citations per year
Based on Google Scholar
Updated in Jan. 2009

Year	Citations
2003	8
2004	40
2005	99
2006	149
2007	246
2008	285

http://david.abcc.ncifcrf.gov/

DAVID - Output

The screenshot displays the DAVID Functional Annotation Clustering tool interface. The browser address bar shows the URL: <http://david.abcc.ncifcrf.gov/summary.jsp>. The page title is "DAVID: Functional Annotation Result Summary".

The interface includes a navigation menu with "Home", "Start Analysis", "Shortcut to DAVID Tools", and "Technical Center". The main header identifies the tool as "Functional Annotation Tool" and "DAVID Bioinformatics Resources 2008, NIAID/NIH".

On the left side, the "Gene List Manager" section shows the current gene list "demolist1" and options to "Use All Species" (Homo sapiens (171)) or "Select". Below this, there are buttons for "List Manager", "Use", "Rename", "Remove", "Combine", and "Show Gene List".

The "Annotation Summary" section displays the following selected categories:

- Main Accessions (0 selected)
- Other Accessions (0 selected)
- Gene Ontology (3 selected)
- Protein Domains (3 selected)
- Pathways (3 selected)
- General Annotations (0 selected)
- Functional Categories (3 selected)
- Protein Interactions (0 selected)
- Literature (0 selected)
- Disease (1 selected)
- Tissue Expression

The "Combined View for Selected Annotations" section includes links for "Functional Annotation Clustering", "Functional Annotation Chart", and "Functional Annotation Table".

The main content area is titled "Functional Annotation Clustering" and shows the "Current Gene List: demolist1" with "171 DAVID IDs". The "Options" section is set to "Classification Stringency" with a "Medium" stringency level. There are buttons for "Rerun using options" and "Create Sublist". A "Download File" link is also present.

The results are presented in a table with columns for "Annotation Cluster", "Enrichment Score", "Count", "P_Value", and "Benjamini".

Annotation Cluster	Enrichment Score	Count	P_Value	Benjamini
Annotation Cluster 1 Enrichment Score: 4.92				
<input type="checkbox"/> GOTERM_CC_ALL	extracellular region	34	6.2E-9	5.4E-6
<input type="checkbox"/> SP_PIR_KEYWORDS	signal	48	5.1E-7	5.5E-4
<input type="checkbox"/> SP_PIR_KEYWORDS	glycoprotein	55	2.9E-6	1.5E-3
<input type="checkbox"/> GOTERM_CC_ALL	extracellular space	17	5.1E-6	2.2E-3
<input type="checkbox"/> GOTERM_CC_ALL	extracellular region part	21	9.9E-6	2.9E-3
<input type="checkbox"/> SP_PIR_KEYWORDS	Secreted	27	5.4E-5	1.4E-2
<input type="checkbox"/> UP_SEQ_FEATURE	signal peptide	46	9.3E-5	5.9E-1
<input type="checkbox"/> UP_SEQ_FEATURE	disulfide bond	42	2.9E-4	8.4E-1
<input type="checkbox"/> UP_SEQ_FEATURE	glycosylation site:N-linked (GlcNAc...)	47	8.1E-3	1.0E0
Annotation Cluster 2 Enrichment Score: 3.26				
<input type="checkbox"/> GOTERM_BP_ALL	defense response	18	1.4E-5	6.9E-2
<input type="checkbox"/> GOTERM_BP_ALL	response to external stimulus	19	1.5E-5	3.9E-2
<input type="checkbox"/> GOTERM_BP_ALL	taxiis	8	3.6E-4	2.1E-1
<input type="checkbox"/> GOTERM_BP_ALL	chemotaxis	8	3.6E-4	2.1E-1
<input type="checkbox"/> GOTERM_BP_ALL	inflammatory response	10	1.4E-3	5.0E-1
<input type="checkbox"/> GOTERM_BP_ALL	response to stimulus	43	1.6E-3	5.0E-1
<input type="checkbox"/> GOTERM_BP_ALL	locomotory behavior	8	1.9E-3	5.1E-1
<input type="checkbox"/> GOTERM_BP ALL	response to wounding	11	5.0E-3	6.8E-1

DAVID-Annotation

- **Takes “significant” gene lists (from microarray expts or proteomic experiments) and allows users to plot heatmaps, generate graphs, identify possible pathways, common or shared functions, clusters of similar genes as well as shared gene ontologies (GO terms)**
- **Facilitates biological interpretation**

How To Do Your Assignment

- **Read the assignment instructions carefully**
- **Follow the instructions listed on the GEO2R website. If you are not clear on how to use the site, look at the YouTube video. Part of the assignment grade depends on you being able to follow instructions on your own**
- **The assignment has several tasks. Make sure to complete all tasks. Use graphs and tables to make your point or answer the questions**
- **Do not plagiarize text from the web or from papers when putting your answers together**
- **You can cut and paste tables and images from tasks you perform on webservers**

How To Do Your Assignment

- **The assignment should be assembled using your computer (cut, paste, format and edit the output or data so it is compact, meaningful and readable)**
- **No handwritten materials unless your computer/printer failed**
- **A good assignment should be 5-6 pages long and will take 4-5 hours to complete**
- **Hand-in hard copy of assignment on due date. Electronic versions are accepted only if you are on your death bed**