BIOC4004 - Industrial Biochemistry
Lecture 17 - Mon Mar 08, 04

Topics for the Day:

- DNA Microarrays (cont.)
  - data visualization
  - analysis
  - data-mining
- Novel applications of the microarray concept

DNA microarray hybridization experiments

Comparative Genomic Hybridization

Cells

↓ DNA isolation

Genomic DNA

↓ Probe labeling

Genomic DNA probe

Hybridization to microarray

control Cy5

test Cy3

Transcript Profiling

Cells

↓ RNA isolation

RNA

↓ • cDNA synthesis
• Probe labeling

cDNA probe

Hybridization to microarray

control Cy3

test Cy5

↓ Laser Scanning

Cy3 / Cy5 Images

↓ Image Analysis

Spot Intensities

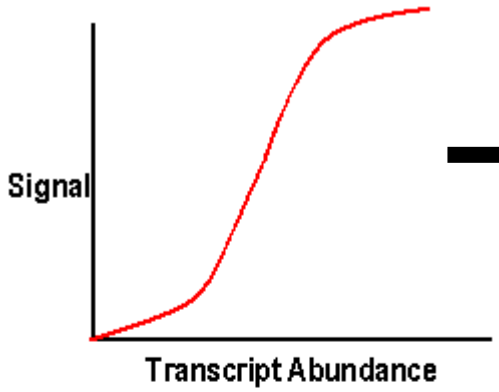
Post Analysis:

- "grouping" genes that behave similarly
- "grouping" samples that behave similarly

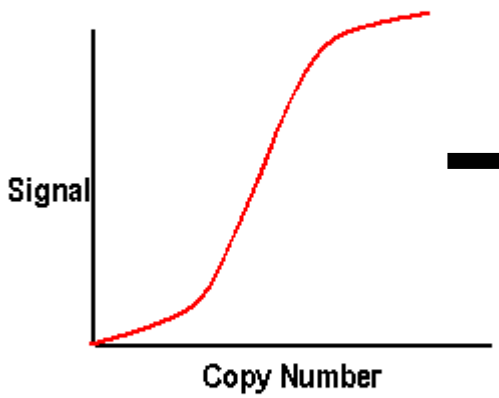
How is microarray-based hybridization used ?

- Signal intensity reflects steady state equilibrium of the hybridization

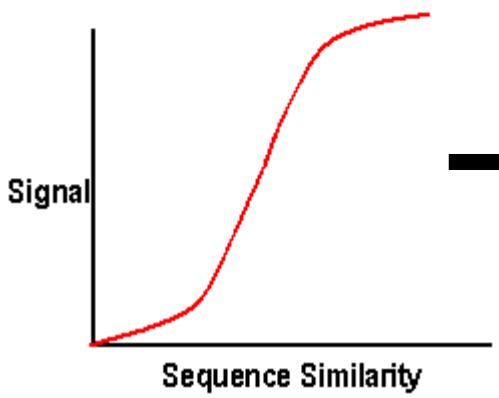
- Affected by :
  - probe abundance
  - sequence similarity between probe and target



Transcript profiling
- tissue specific differences
- temporal differences
- stimulus-induced differences
- treatment-induced differences

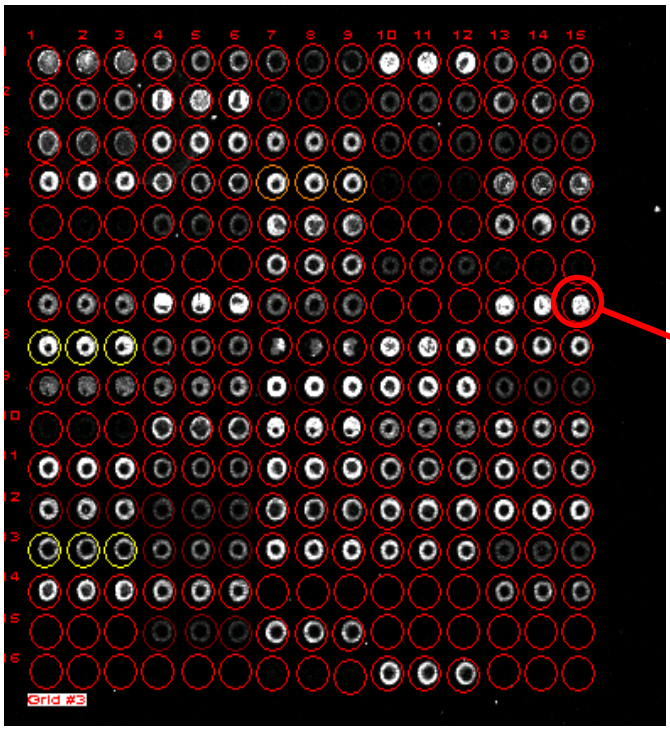


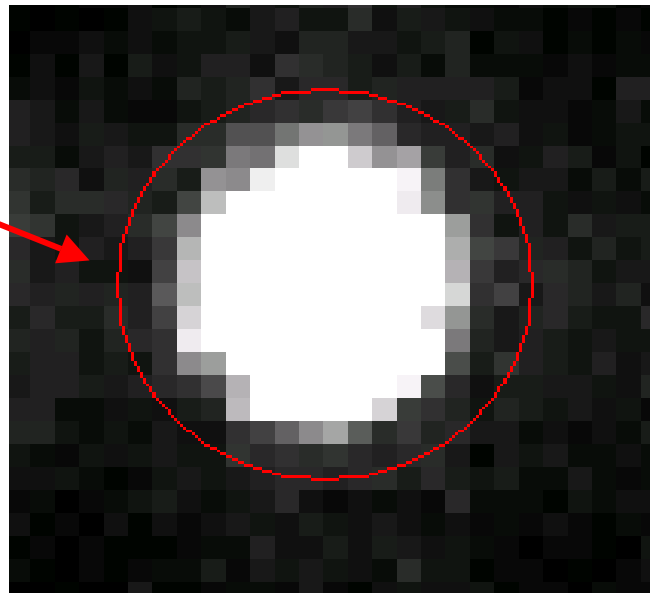
Comparative genomics
- gene copy number differences (due to genetic instability)
  - observed in certain cancers
  - bacterial strains
  - plant cultivars

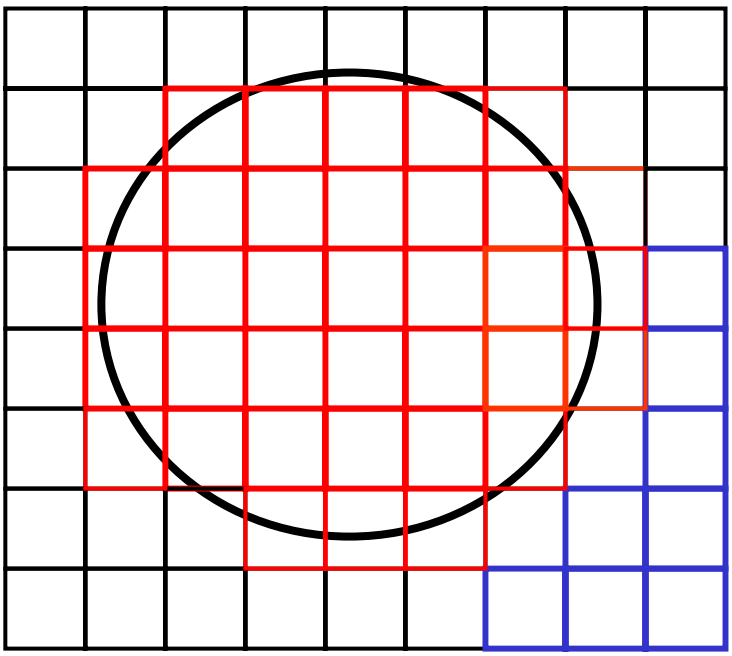


Comparative genomics
- sequence differences
  - molecular typing
  - evolutionary studies

About Calculating Spot Intensities…



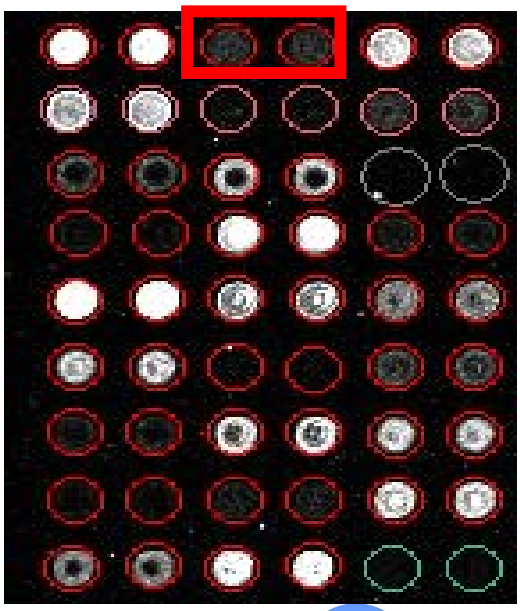


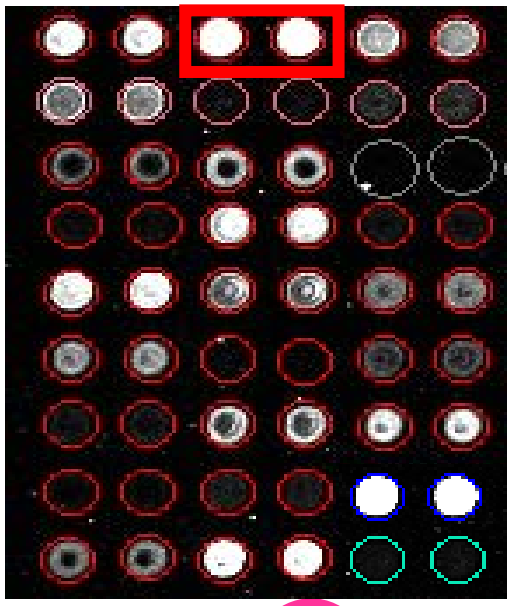


- **Each pixel is 16 bit, ie. it can have 65535 shades of grey**
- **The intensity of each pixel contained in a spot is calculated**
- **Signal from surrounding neighboring pixels is used to determine background**
- **Signal: [mean (or median) pixel intensity] - [Background]**

- We are interested in comparing the signal obtained from one sample (Cy3) vs the other sample (Cy5) --> the ratio of one to the other

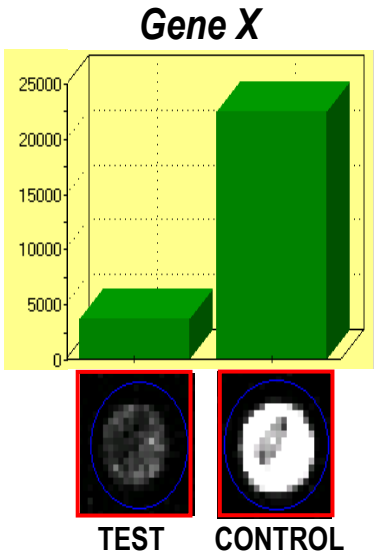
What does the Microarray data mean ?

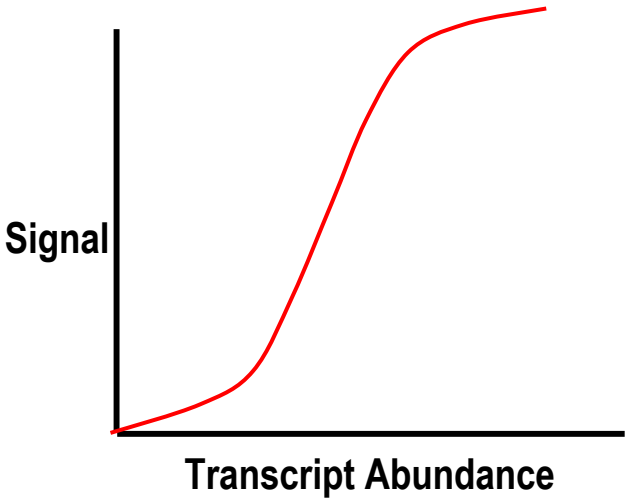




test Cy5

control Cy3





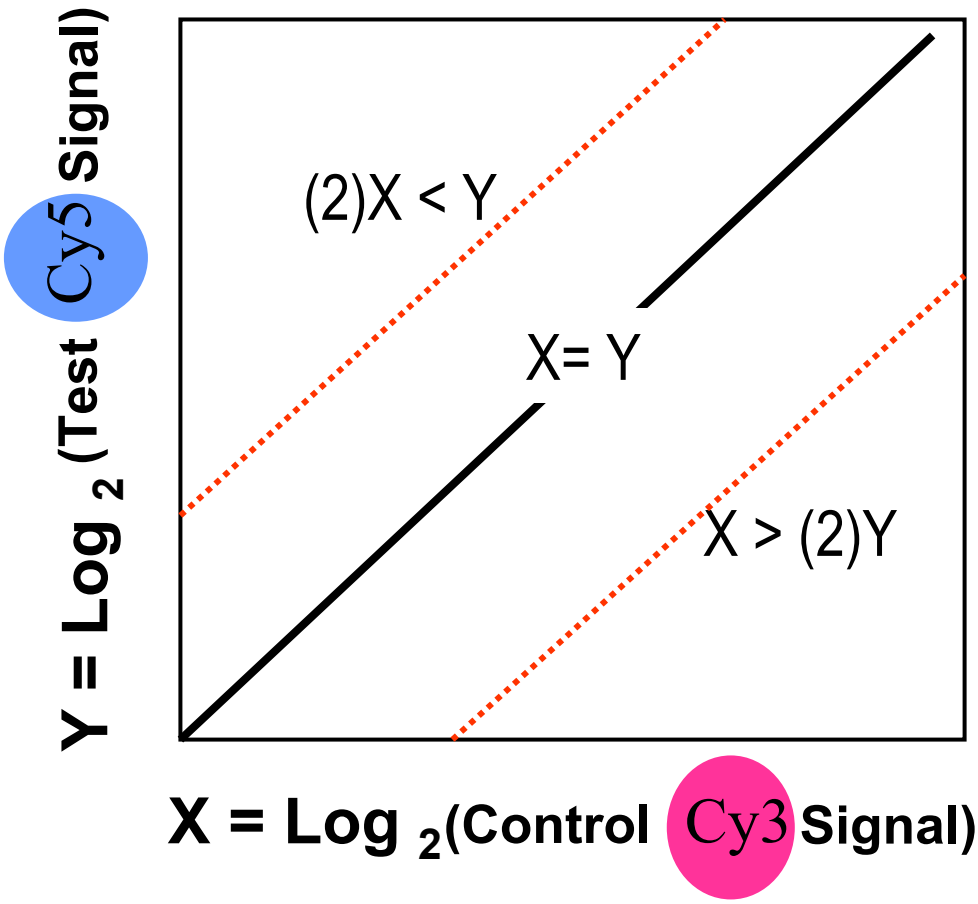
GOAL: inference of "relative transcript levels"

TEST has 5-fold lower signal than CONTROL

Gene X transcript is **"5-times less abundant"** in TEST than in CONTROL

Visualizing Microarray Data using Scatter-plots

- An easy way to visualize Test/Control Signal Ratios is the scatterplot
- Plot Signal of Test on Y-axis vs Signal of Control on X-axis
    - plotting log[Signal] makes visualization more intuitive
- A similar signal intensity on both samples: spot close to 45 degree line
    - X = Y is the 1:1 ratio line
    - Signals that are stronger in Y are above the 1:1 line
    - Signals that are stronger in X are below the 1:1 line

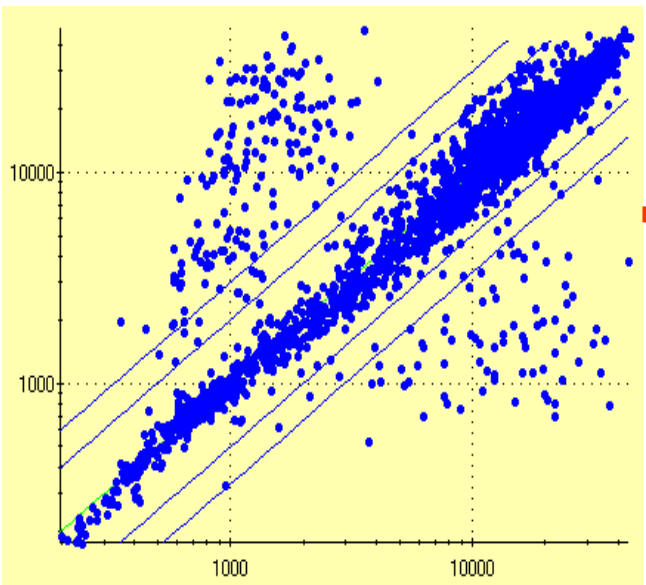


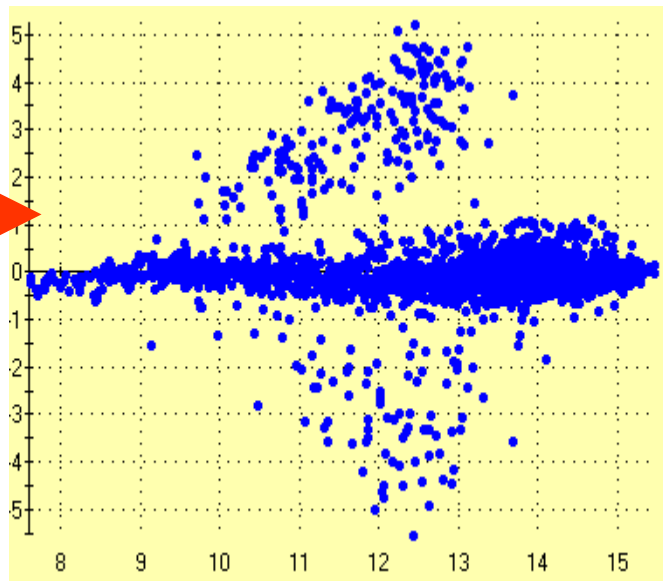
- Spots far from X=Y are called “OUTLIERS”: these are the interesting ones
- The farther the spots are from X=Y, the more differential the signal

A better Visualization Tool: The MA Plot

M = Log_2(Test Signal) – Log_2 (Control Signal)
Or
Log_2(Test Signal / Control Signal)

A = [Log_2(Test Signal) + Log_2 (Control Signal)] / 2



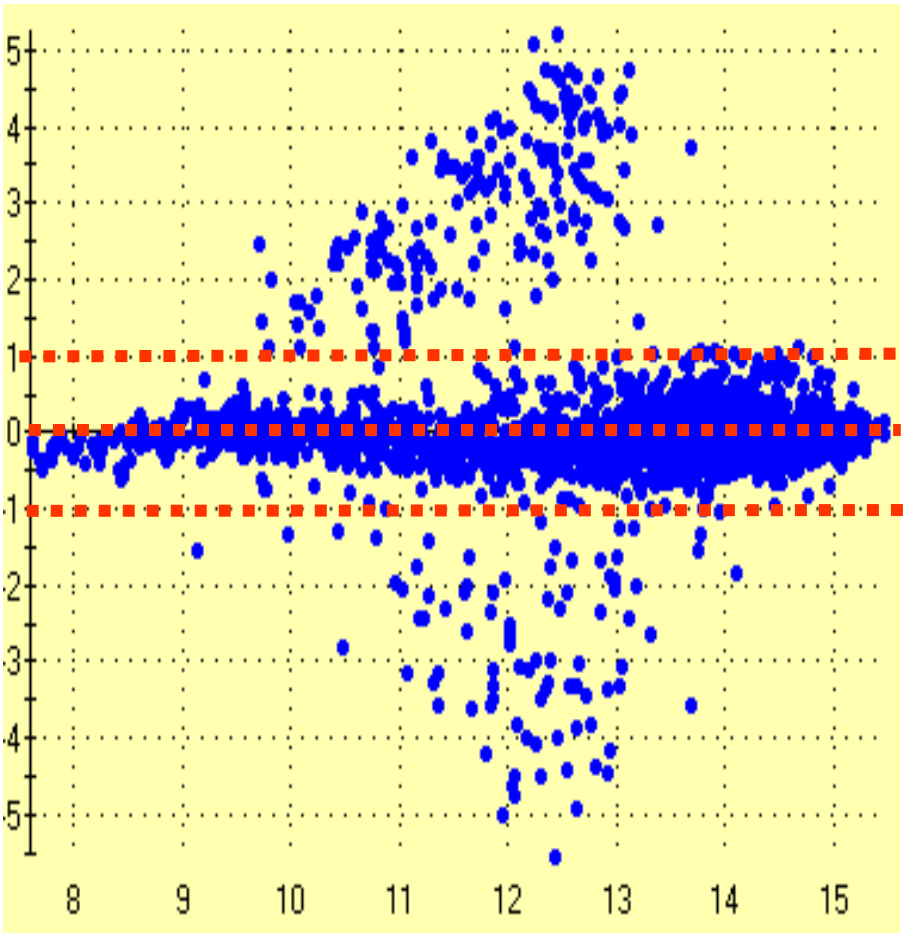


Y = log_2 signal Test
X = log_2 signal Control

Y = log_2 ratio (or log difference)
X = log_2 average signal

- *more intuitive:*
  - *+ve log ratio = up-regulated*
  - *–ve log ratio = down-regulated*
  - *log ratio = 0 = no change*
- *data is calibrated according to intensity*
  - *low intensity data can't be trusted as much*

Microarray Data Interpretation



Tester > 2 X Control

Tester = 2 X Control

Tester = Control

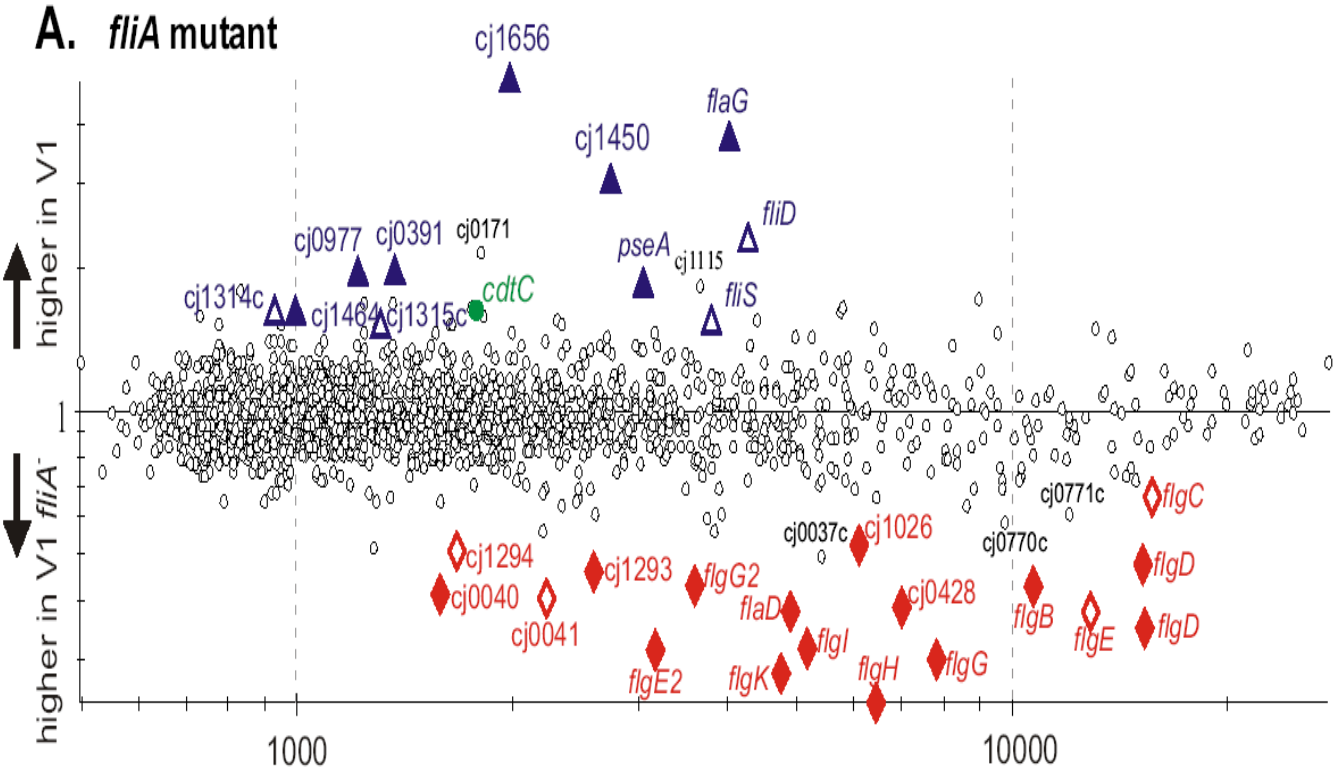
2 X Tester = Control

2 X Tester < Control

- Assume most of the genes will give similar signal (ie. most spots should end up close to the 1:1 ratio line)
  - look for a tight scatter around 1:1 for most of the points
  - Too many outliers may mean technical problems:
    - probe preparation
    - hybridization
    - washing
    - scanning
- Most researchers only consider outliers that deviate from 1:1 line by factor of 2 or more:
  - ie. Log_2(Test/Control) > 1.0 or < -1.0
  - smaller deviations can be due to technical reasons
  - Replicates are ALWAYS performed
  - Confirmation using other methods (eg. qPCR)

Real Gene Expression Microarray Data

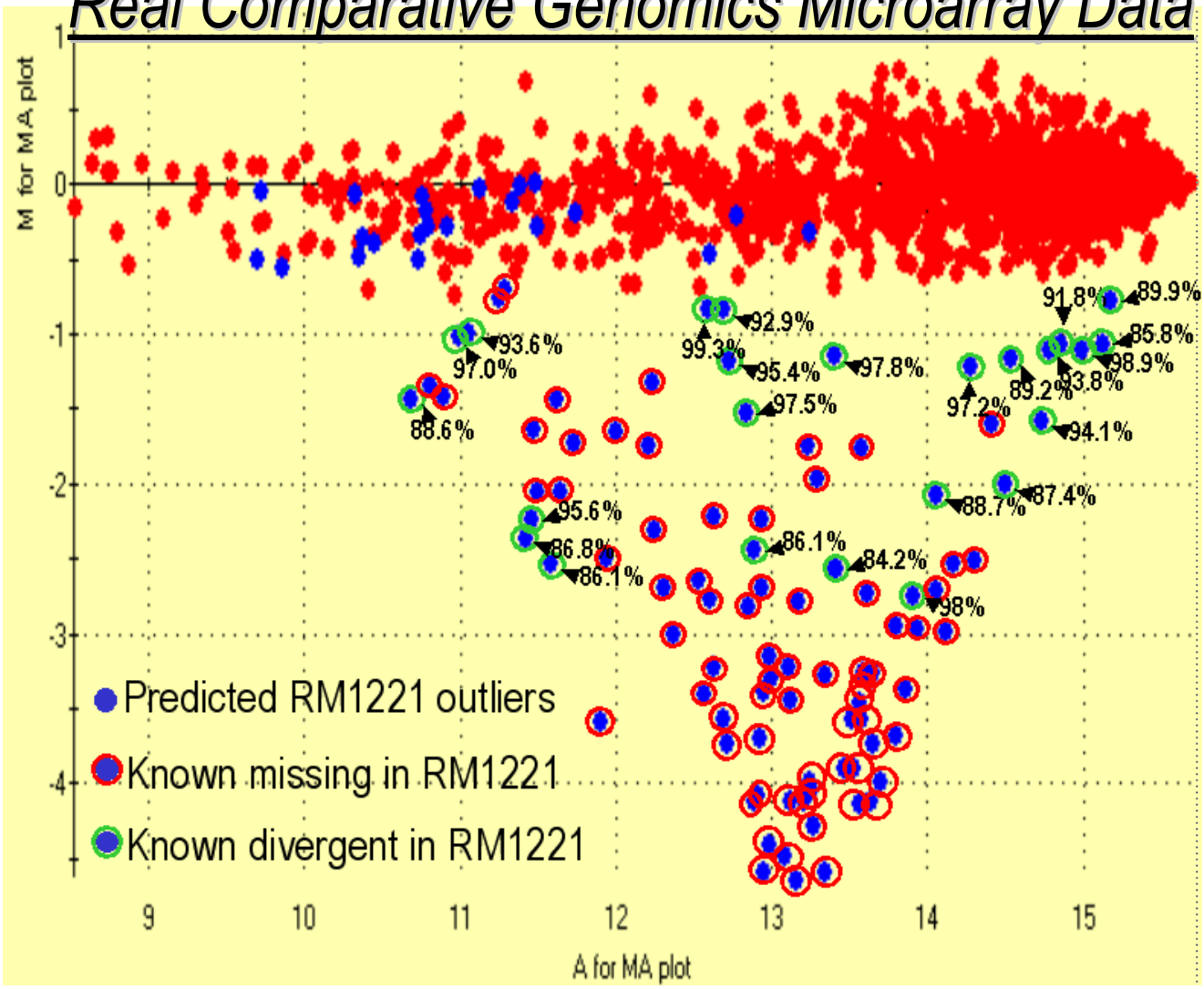
Carrillo *et al*., (2004) JBC, in press



fliA is a sigma factor for some flagellar genes (sigma-28)

- *knock-out mutant has less expression of these (blue triangles)*
  - *ie. higher in the wild type*
- *knock-out has more expression of sigma-54 regulated genes (red diamonds):*
  - *ie. sigma-28 (in)directly represses sigma-54 regulated genes*

Real Comparative Genomics Microarray Data

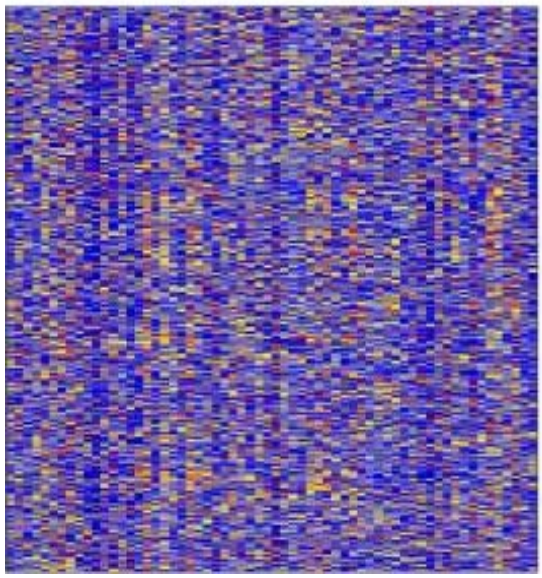


Two sequenced genomes of the enteric pathogen C. jejuni

- *Strain 11168 has 80 genes not found in strain RM1221*
- *Strain RM1221 has ~ 200 genes not found in strain 11168*
- *A number of genes are in both, but not 100% conserved*

- *A good correlation between microarray Log Ratio and gene conservation*
- *no "false positives": all outliers make sense*
- *some false negatives: not outliers but should have been*
    - *mostly low signal spots*

Making sense of microarray data: clustering

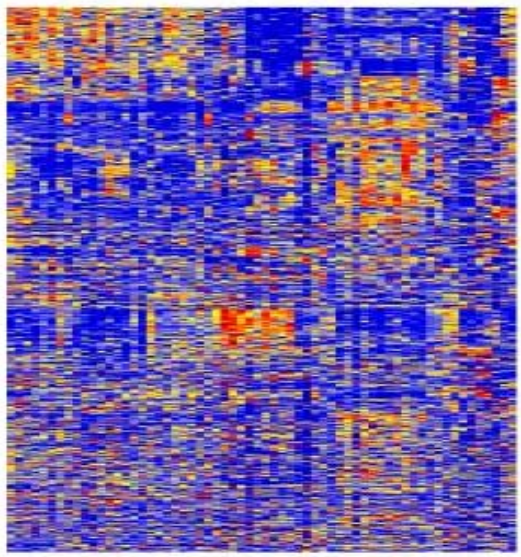


No discernible pattern observed

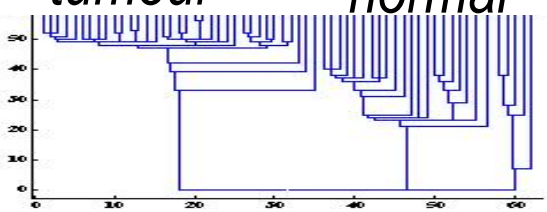
Non-clustered

Samples that behave similarly

Genes that behave similarly



tumour *normal*



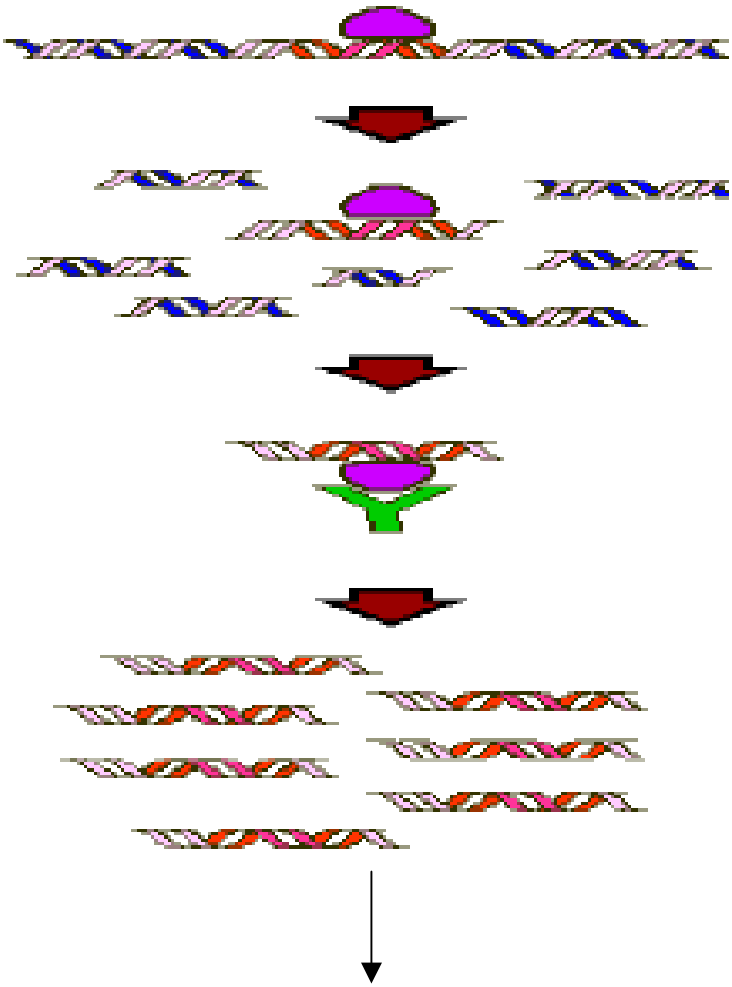
Two-dimensional clustering

- samples grouped based on overall similarity of expression profiles
- genes grouped based on their expression behaviour across the group of samples

Use clustering algorithms to "group" samples or genes with similar behaviour

Novel Applications of Microarrays : ChIP on CHIP

ChIP = Chromatin Immuno-Precipitation



Protein bound to DNA

- x-link with formaldehyde (reversible)

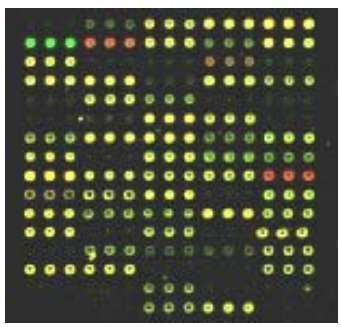
Sheared DNA

Antibody binding

- eg. Anti-p53 mAb

protein release

- reverse x-link
- fluorescent labelling of chromatin

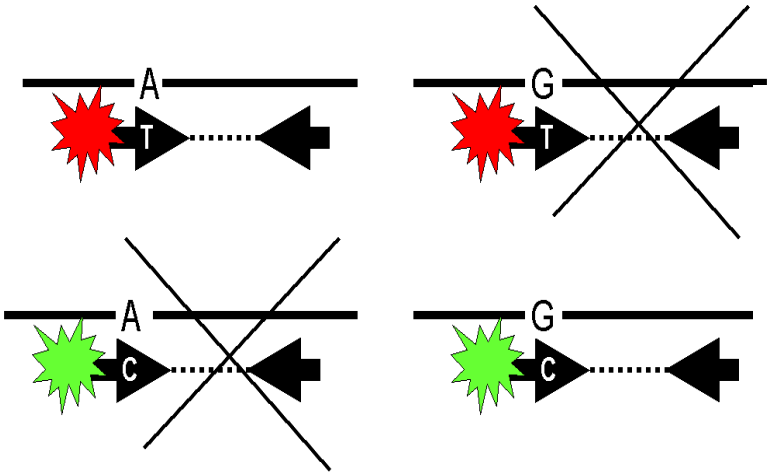


Hybridize to chip with promoter sequences from different genes

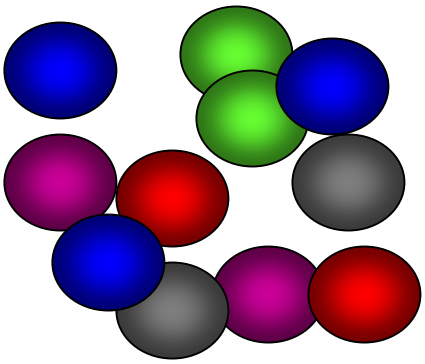
- compare p53-bound “enriched chromatin” to “non-enriched”
- Will tell you if p53 binds to a given gene’s promoter

Novel applications of microarrays: micro-bead arrays (Pt.1)

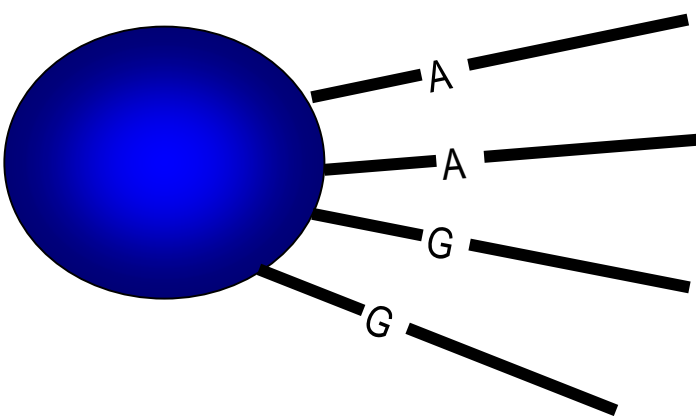
SNPs are single-nucleotide polymorphisms
• useful as a diagnostic tool (we'll talk about them later in class)



• primers are designed so that they can only amplify one SNP allele and not others
• each allele-specific primer is labeled with a different fluorescent tag



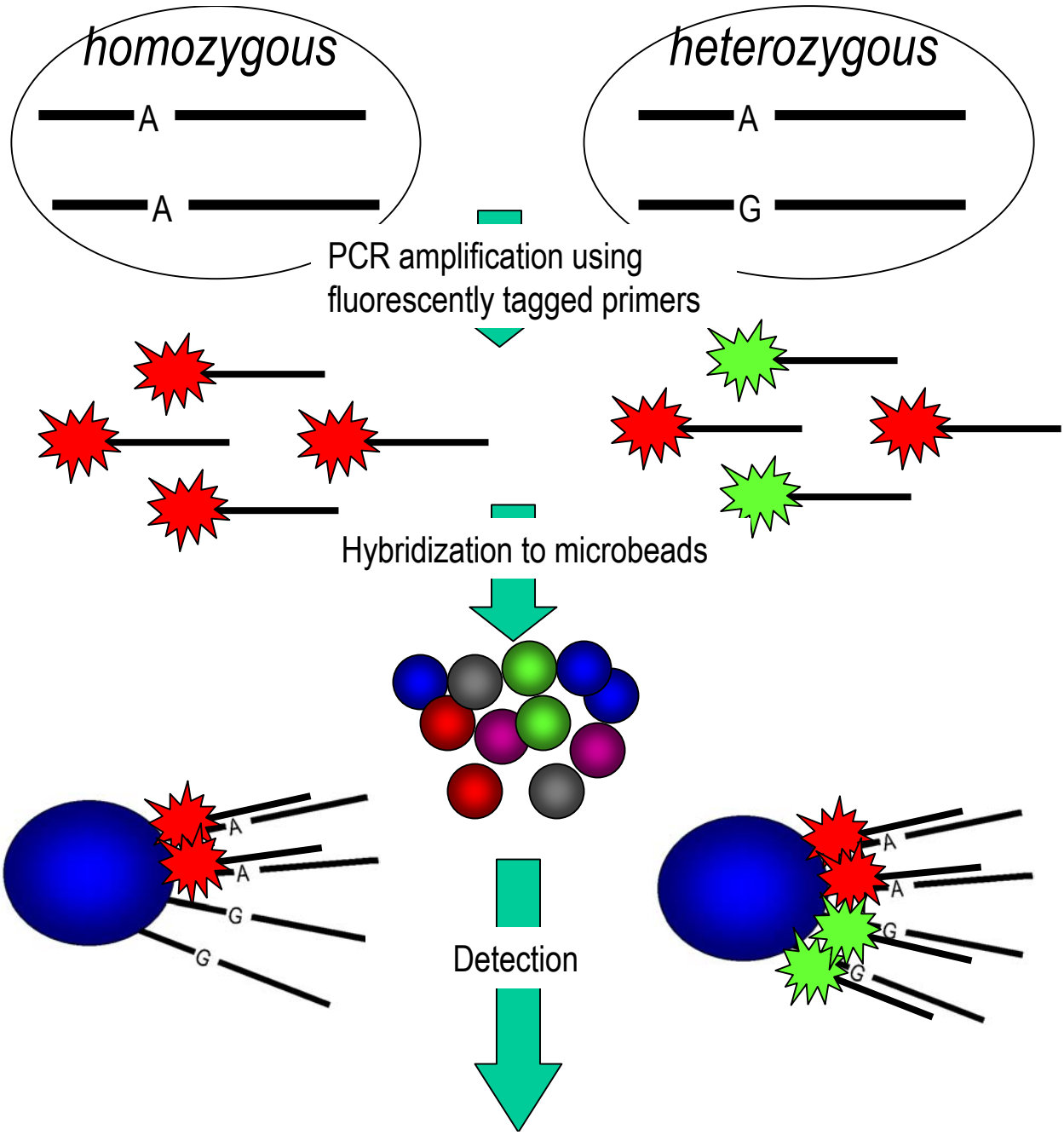
About the microbeads:
• each type of microbead has its own detectable colour
• the microbead can be ID'd by its colour as it passes through a detector



• each bead is coated to saturation using SNP specific oligonucleotides
• the SNP probes coupled to each bead are available for hybridization
• since the SNP PCR products are fluorescently labeled, we can tell which SNP allele(s) are present in the sample derived from the group by hybridization

SNP detection using microbead arrays (Pt.2)

Detection of SNP alleles

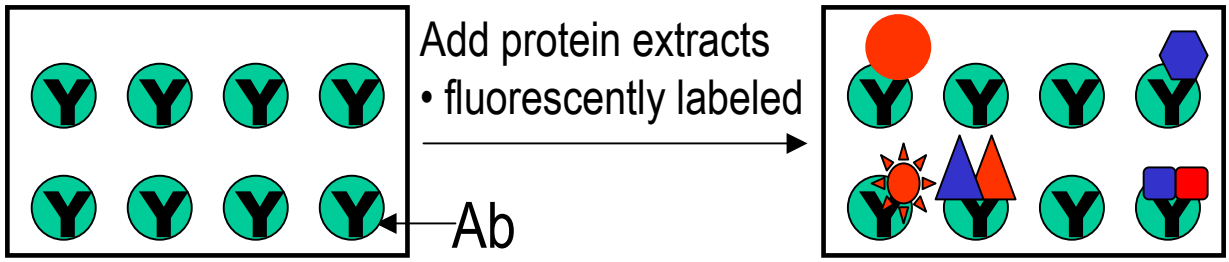


Identification of Microbead
Quantitation of fluorescent signal

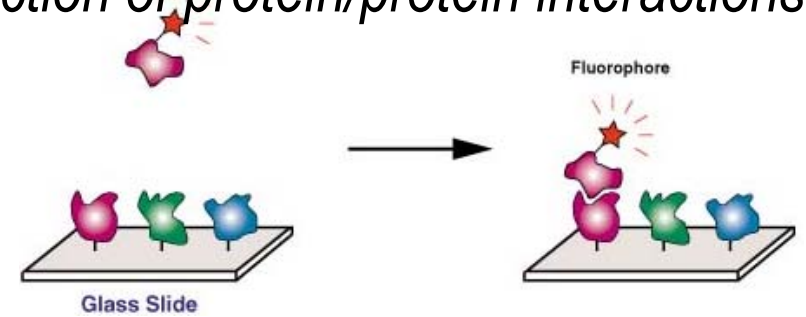
- which SNP ?
- Which alleles were present ?
- How much of each allele ?

Microarray Concept Co-opted: Protein microarrays

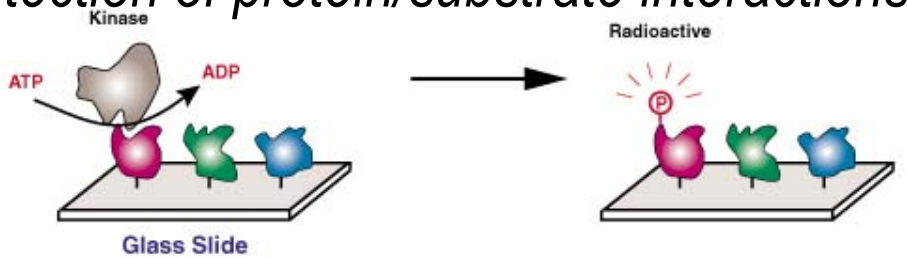
Quantitation of proteins in cell extracts: antibody arrays



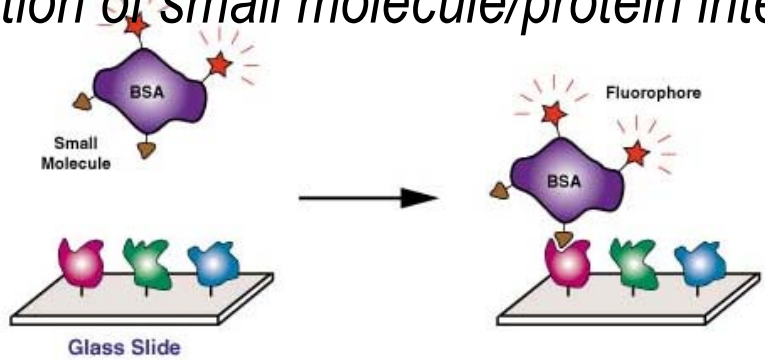
Detection of protein/protein interactions



Detection of protein/substrate interactions



Detection of small molecule/protein interactions



- *DNA retains its physical characteristics easily, proteins do not*