

Different types of microarrays

of samples
hybridized

| Method of generating probes | | # of samples hybridized | |
|--------------------------------|-------------|---|-----------------------------------|
| | | Single-channel | Dual-channel |
| cDNA Library | | | cDNA spotted arrays (Stanford) |
| | Synthesized | Affymetrix (25-mer) NimbleGen (50-75-mer) Agilent (60-mer) Illumina (50-mer) | NimbleGen Agilent Illumina |

What Is Microarray Technology?

- Different Approaches

| | | |
|---------------------------------------|------------------------------|------------------|
| | Stanford/ Pat Brown | Affymetrix |
| How DNA sequences are laid down | Spotting | Photolithography |
| Length of DNA sequences | cDNA(Comple te sequences) | Oligonucleotides |

complementary DNA (cDNA)

- cDNA is a strand of DNA that is complementary to part of an mRNA sequence.

| | |
|------|--------------------|
| mRNA | ...CCUGAUAGAUGG... |
| cDNA | ...GGACTATCTACC... |

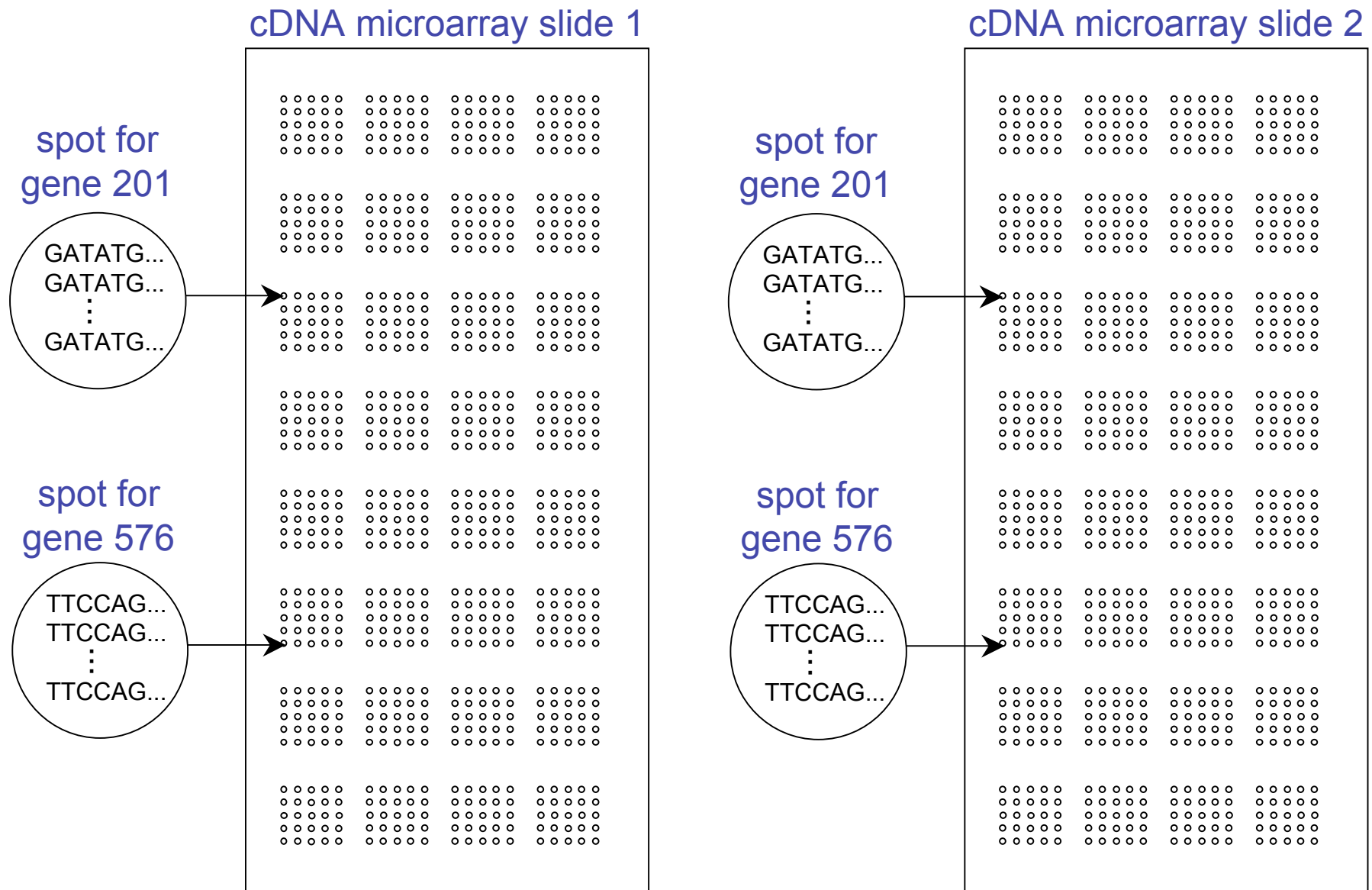
- cDNA can be formed by extracting mRNA and then using mRNA as a template for formation of cDNA
- cDNA sequences can be copied rapidly using PCR (polymerase chain reaction).
- These sequences can be spotted on glass slides to serve as microarray probes.
- Sequence length varies from a few hundred bases to a thousand or so.

cDNA Microarrays

- Glass slides or similar supports containing cDNA sequences that serve as probes for measuring mRNA levels in target samples
- cDNAs are arrayed on each slide in a grid of spots.
- Each spot contains thousands of copies of a sequence that matches a segment of a gene's coding sequence.
- A sequence and its complement are present in the same spot.

cDNA Microarray (continued)

- Different spots typically represent different genes, but some genes may be represented by multiple spots
- The spotted sequences are known (or can be determined) and their locations on the array are known.
- The sequence locations do not change from slide to slide.
- A single slide typically contains thousands of spots.



Each spot contains many copies of a sequence along with its complement (not shown).

Spotting cDNA Probes on Microarrays

- Solutions containing probes are transferred from a plate to a microarray slide by a robotic arrayer.
- The robot picks up a small amount of solution containing a probe by dipping a pin into a well on a plate.
- The robot then deposits a small drop of the solution on the microarray slide by touching the pin onto the slide.
- The pin is washed and the process is repeated for a different probe.
- Most arrayers use several pins so that multiple probes are spotted simultaneously on a slide.
- Most arrayers print multiple slides together so that probes are deposited on several slides prior to washing.

Cartoon of Printing Process (side view from the table top)

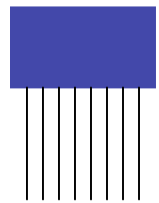


Plate with wells
containing probes



microarray slides

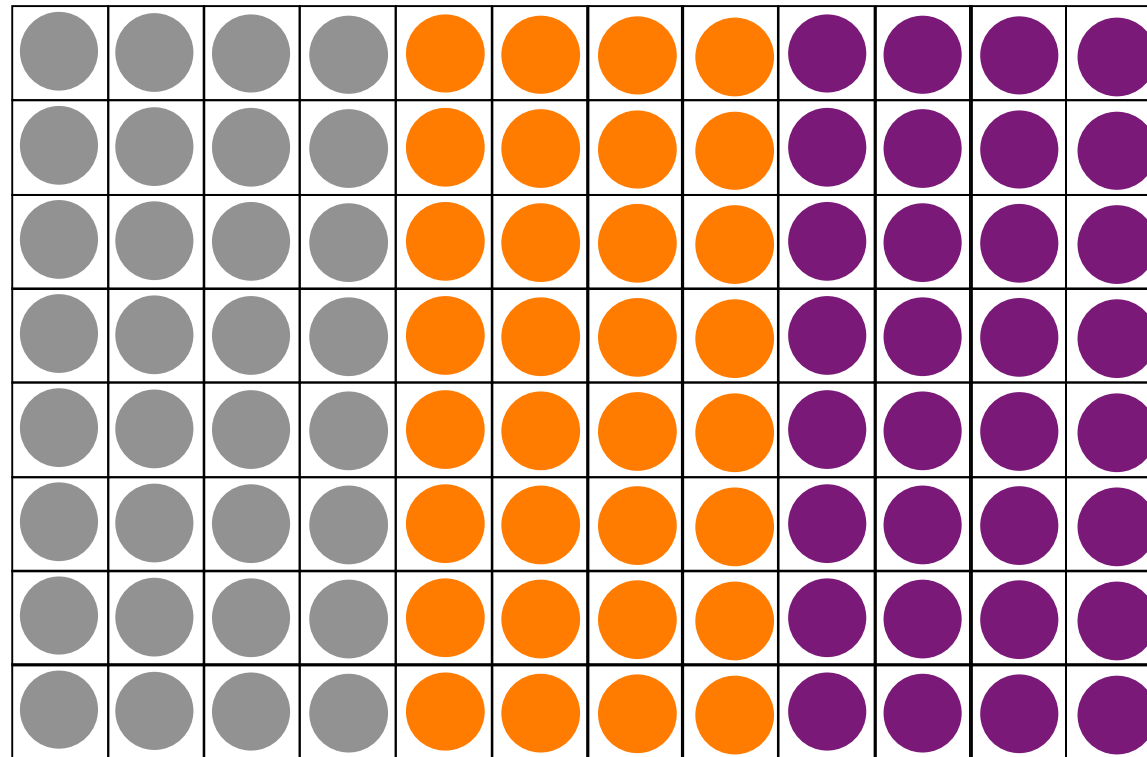


vacuum
wash
station

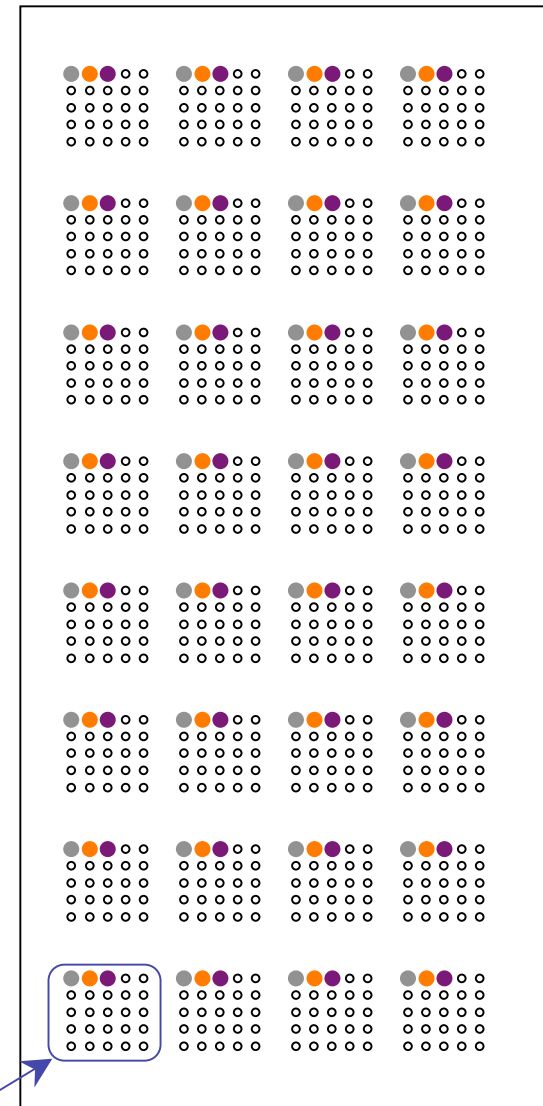
Spotting the Probes on the Microarray

8 X 4 Print Head

plate with wells holding probes in solution



microarray slide



All spots of the same color are made at the same time.

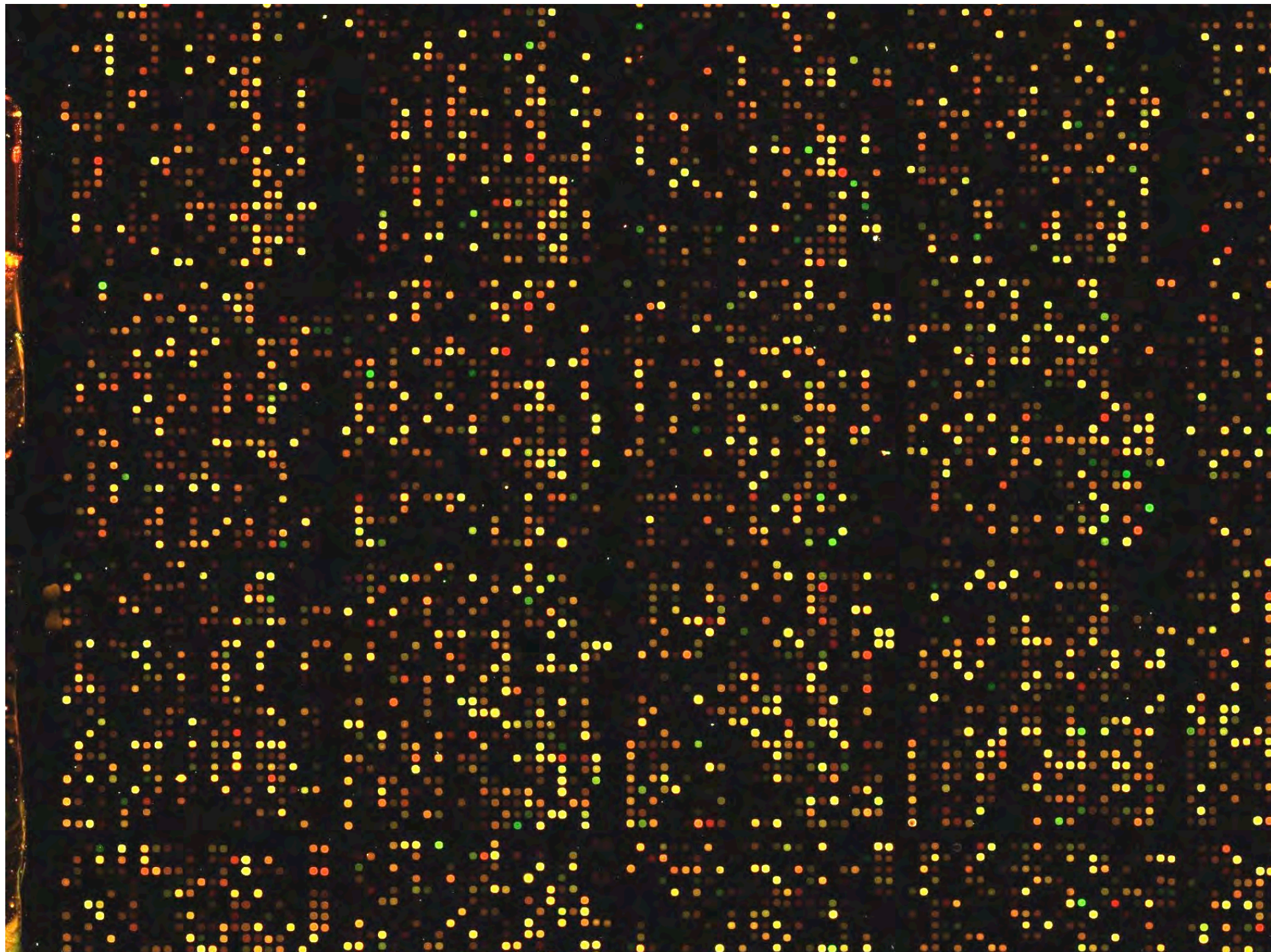
All spots in the same sector are made by the same pin.

Using cDNA Microarrays to Measure mRNA Levels

- RNA is extracted from a target sample of interest.
- mRNA are reverse transcribed into cDNA.
- The resulting cDNA are labeled with a fluorescent dye.
- The dyed cDNA are placed on a microarray slide.
- Dyed cDNA sequences hybridize to complementary probes spotted on the array.
- A laser excites the dye and a scanner records an image of the slide.
- The image is quantified to obtain measures of fluorescence intensity for each pixel.
- Pixel values are processed to obtain measures of mRNA abundance for each probe spotted on the array.

Using cDNA Microarrays to Measure mRNA Levels (ctd.)

- Usually two samples, dyed with different dyes, are hybridized to a single slide.
- The dyes fluoresce at different wavelengths so it is possible to get separate images for each dye.
- Cyanine 3 (Cy3) and Cyanine 5 (Cy5) are currently the two most commonly used dyes.
- Images from the scanner are black and white, but it is typical to display Cy3 images as green and Cy5 images are displayed as red.
- It is common to superimpose the two images using yellow to indicate a mixture of green and red.



There are many ways to obtain a labeled target sample.
Here's a simplified version of one method.



viral enzyme *reverse transcriptase*
recognizes poly-T bound to poly-A
and begins to add complementary
DNA nucleotides. The C nucleotides
are dyed.

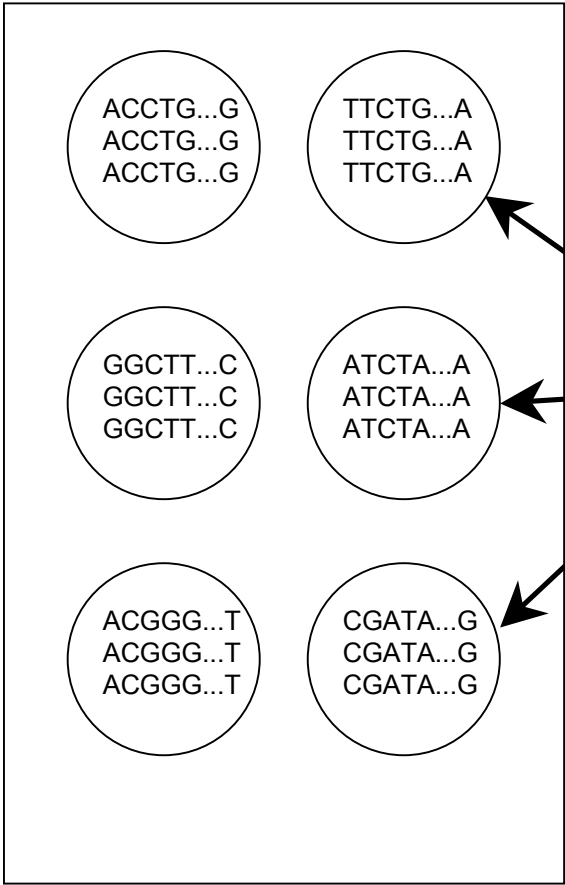
poly-T primer

Difficult to Make Meaningful Comparisons between Genes

- The measures of mRNA levels are affected by several factors that are partly or completely confounded with genes (e.g., cDNA source plate, cDNA well, print pin, slide position, length of mRNA sequence, base composition of mRNA sequence, specificity of probe sequence, etc.).
- Within-gene comparisons of multiple cell types or across multiple treatment conditions are much more meaningful.

Using cDNA Microarrays to Measure mRNA Levels

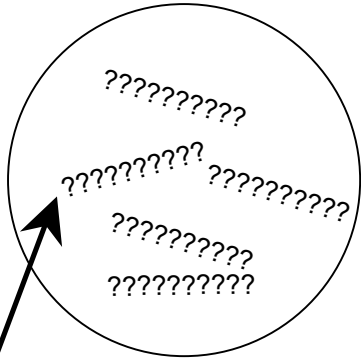
Microarray Slide



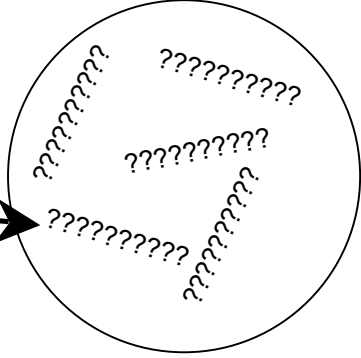
Spots
(Probes)

Unknown
mRNA
Sequences
(Target)

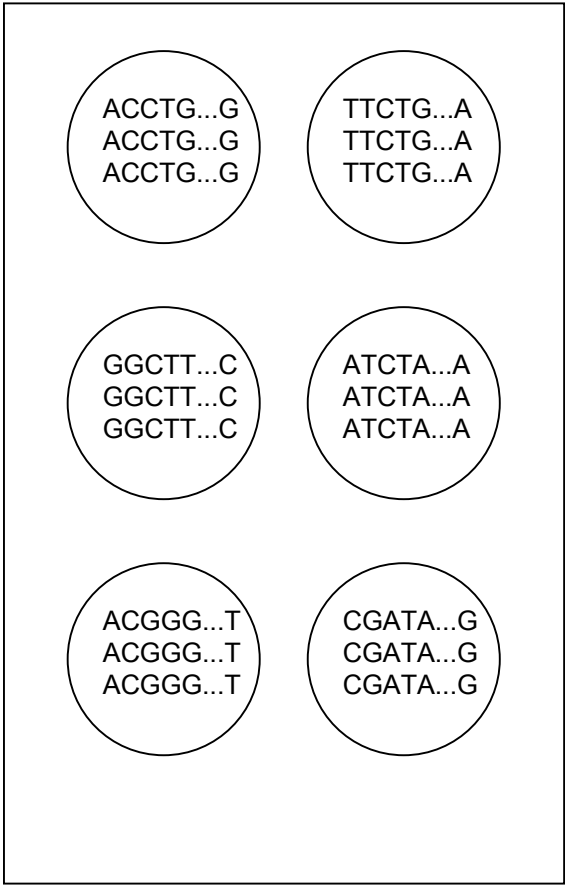
Sample 1



Sample 2



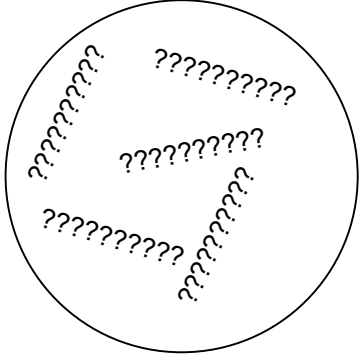
Extract mRNA



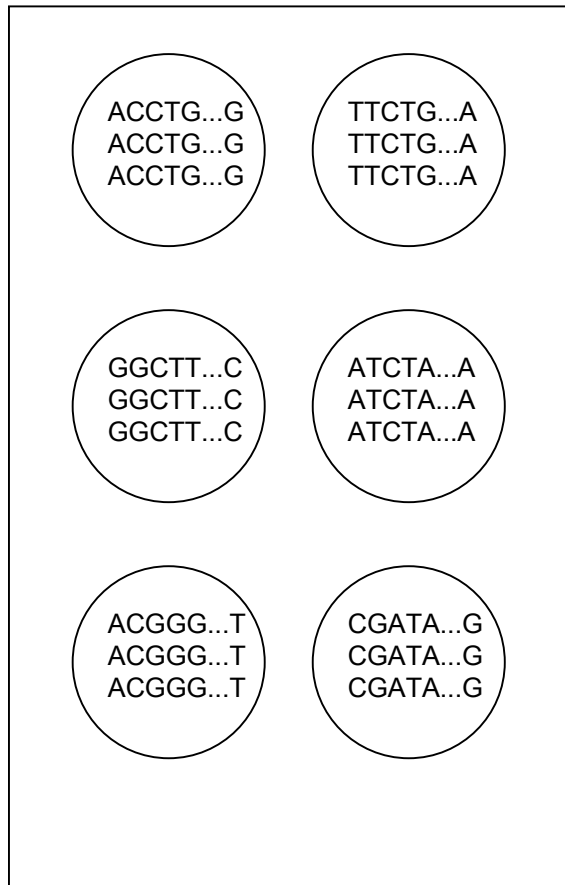
Sample 1



Sample 2

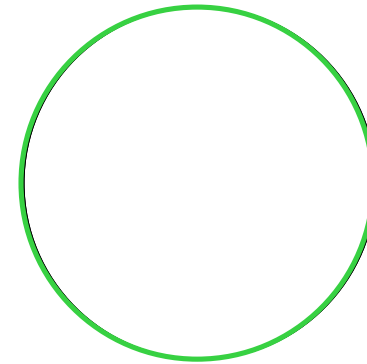


Convert to cDNA and Label with Fluorescent Dyes

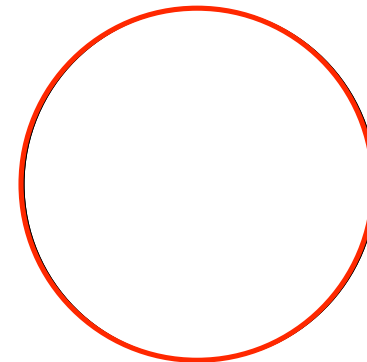


?????????
?????????
?????????
?????????
?????????

Sample 1

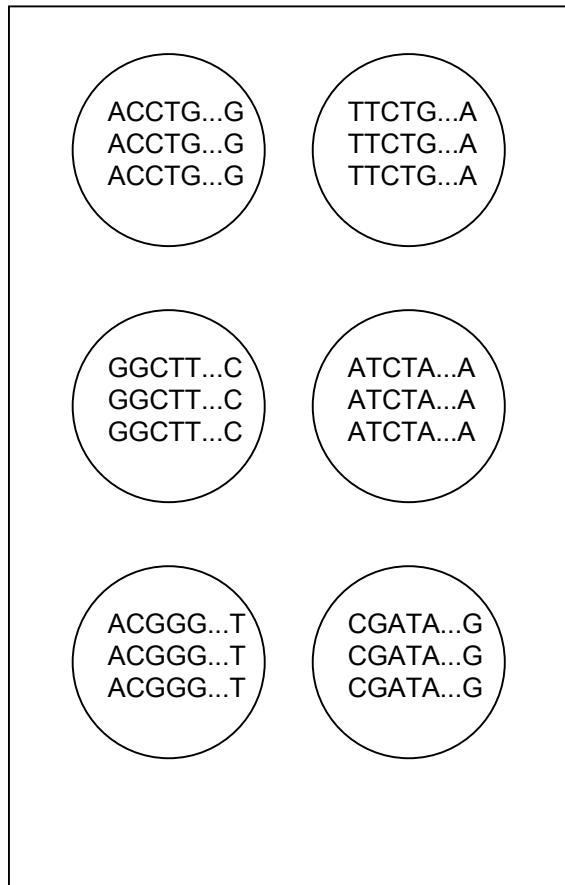


Sample 2



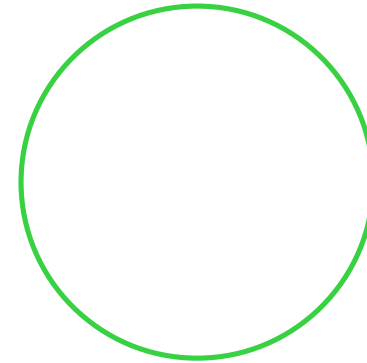
?????????
?????????
?????????
?????????
?????????

Mix Labeled cDNA



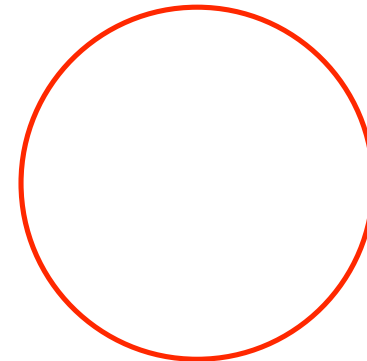
iiiiiiiiii
iiiiiiiiii
iiiiiiiiii
iiiiiiiiii
iiiiiiiiii

Sample 1

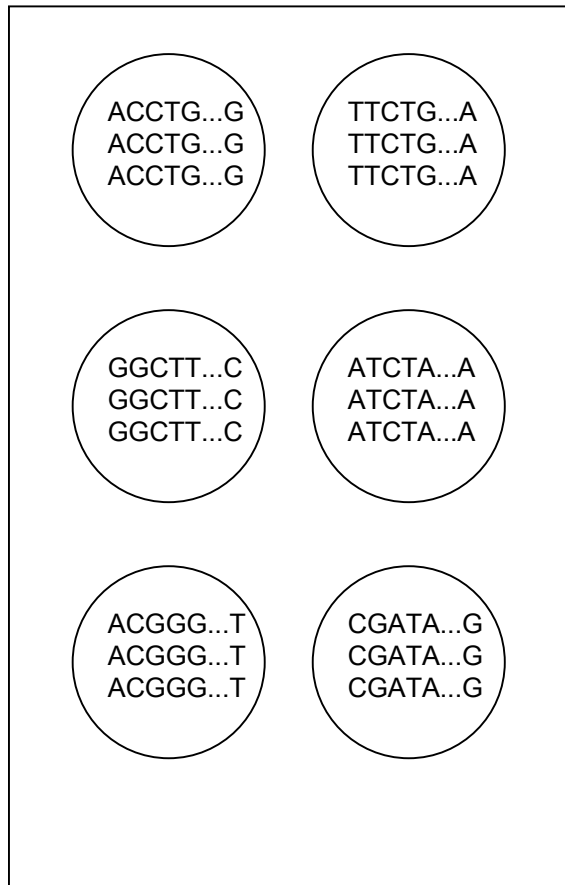


Sample 2

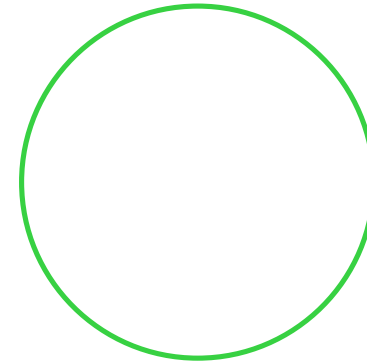
iiiiiiiiii
iiiiiiiiii
iiiiiiiiii
iiiiiiiiii
iiiiiiiiii



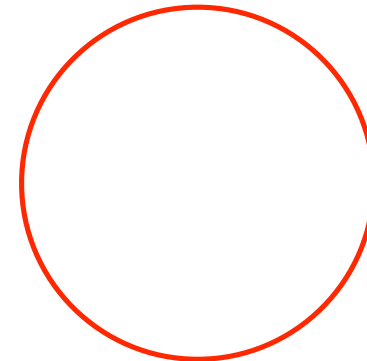
Hybridize cDNA to the Slide



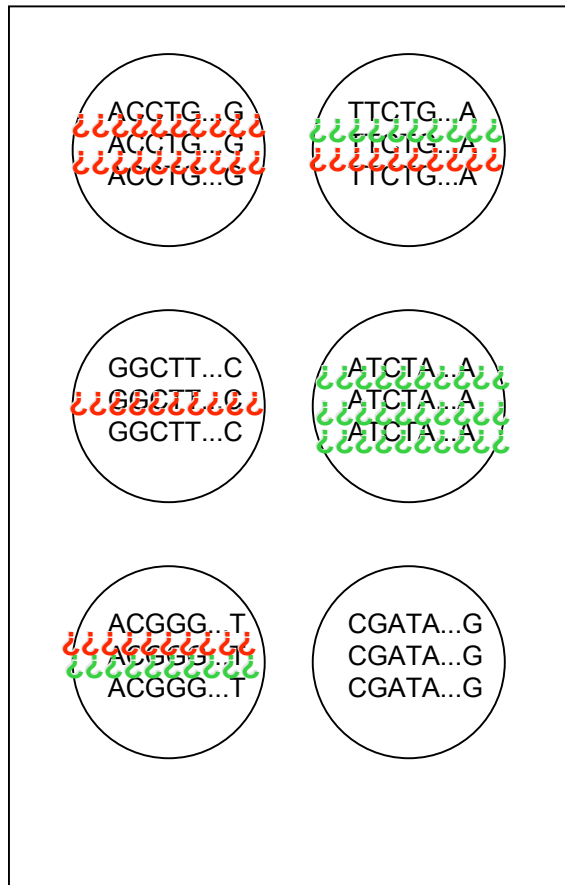
Sample 1



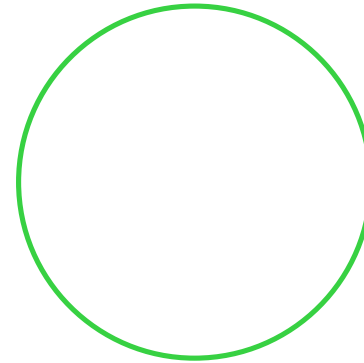
Sample 2



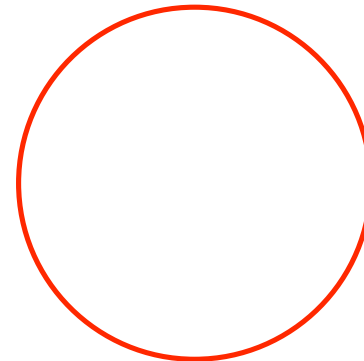
Excite Dyes with Laser



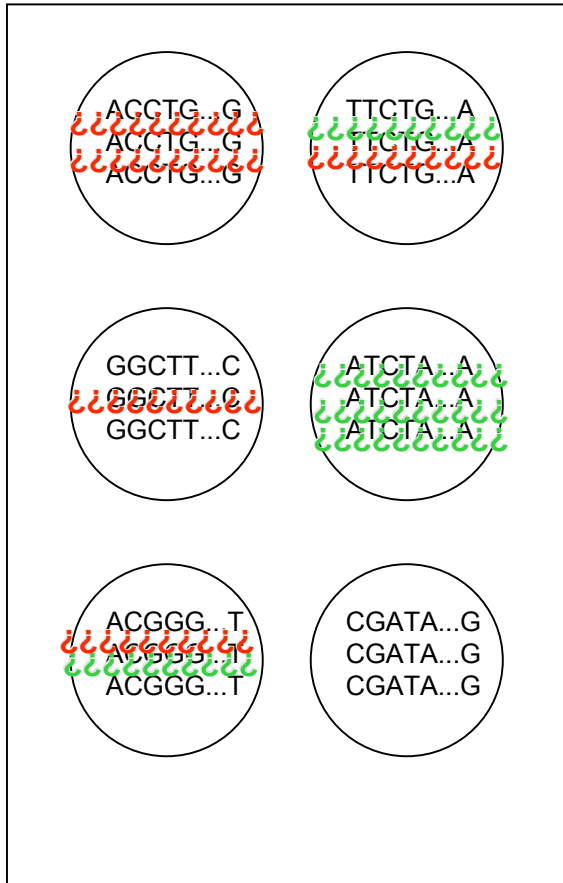
Sample 1



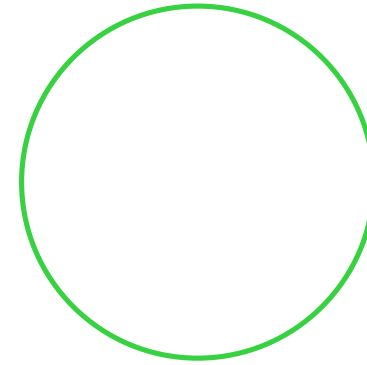
Sample 2



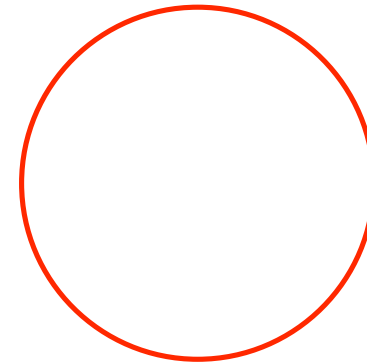
Scan



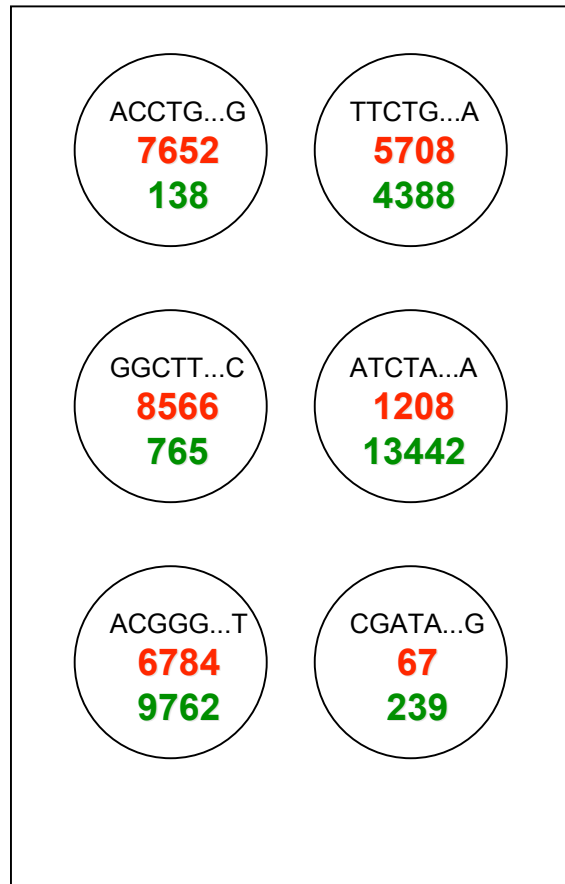
Sample 1



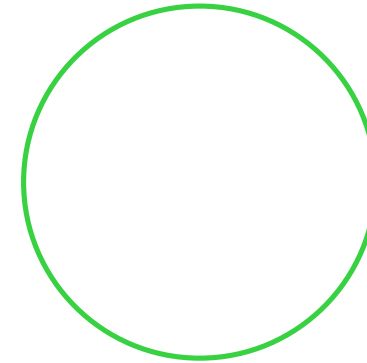
Sample 2



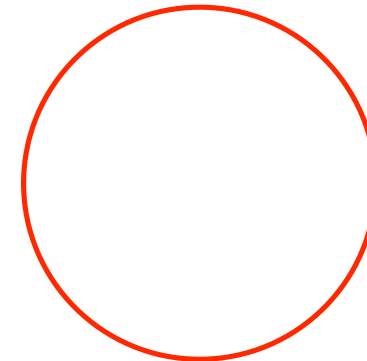
Quantify Signals



Sample 1



Sample 2



cDNA Arrays: Advantages

- Non-redundant clone sets are available for numerous organisms (humans, mouse, rats, *drosophila*, yeast, *c.elegans*, *arabidopsis*)
- Prior knowledge of gene sequence is not necessary: good choice for gene discovery
- Large cDNA size is great for hybridization
- Glass or membrane spotting technology is readily available

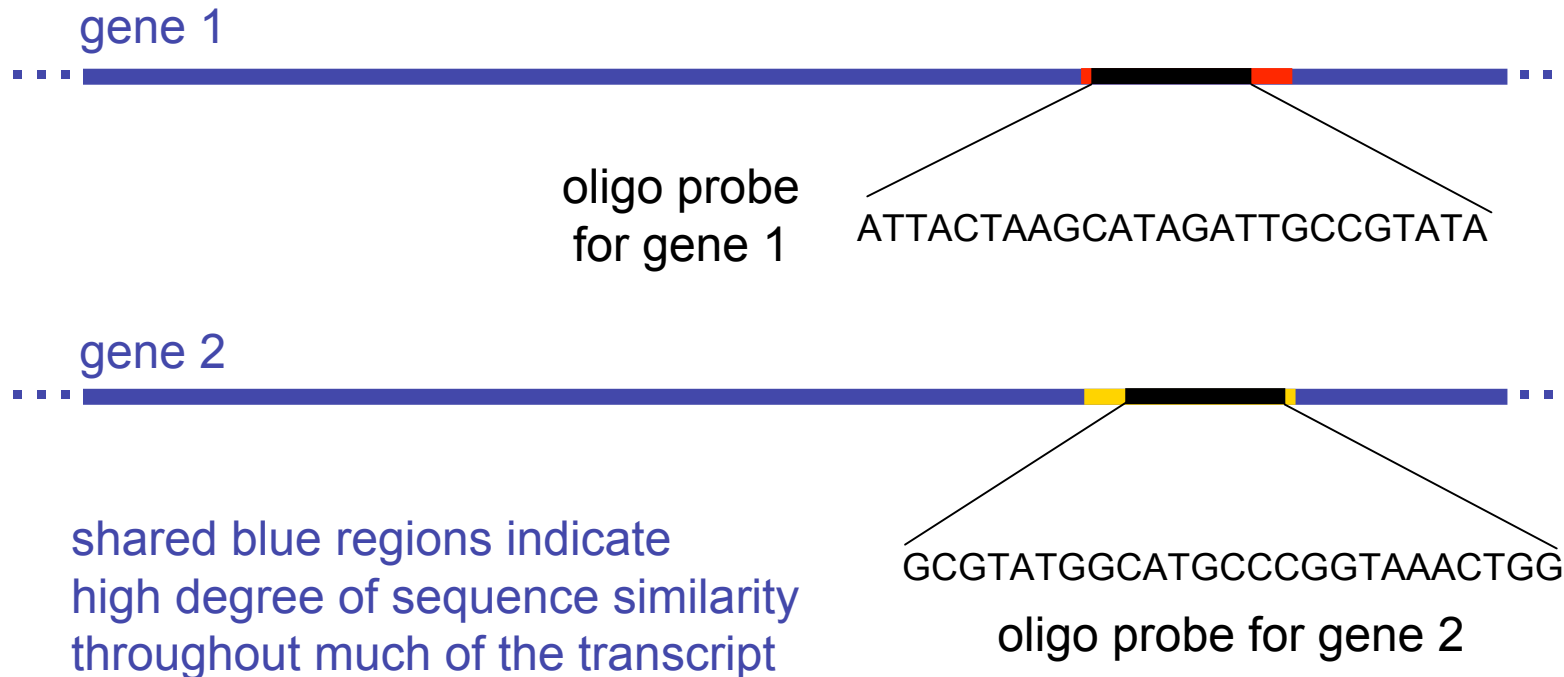
cDNA Arrays: Disadvantages

- Processing cDNAs to generate “spotting-ready” material is cumbersome
- Low density compared to oligonucleotide arrays
- cDNAs may contain repetitive sequences (like *Alu* in humans)
- Common sequences from gene families (ex: zinc fingers) are present in all cDNAs from these genes: potential for cross-hybridization
- Clone authentication can be difficult

Oligonucleotides

- An oligonucleotide is a short sequence of nucleotides. (oligonucleotide=oligo for short)
- An oligonucleotide microarray is a microarray whose probes consist of synthetically created DNA oligonucleotides.
- Probes sequences are chosen to have good and relatively uniform hybridization characteristics.
- A probe is chosen to match a portion of its target mRNA transcript that is unique to that sequence.
- Oligo probes can distinguish among multiple mRNA transcripts with similar sequences.

Simplified Example



Oligo Microarray Fabrication

- Oligos can be synthesized and stored in solution for spotting as is done with cDNA microarrays.
- Oligo sequences can be synthesized on a slide or chip using various commercial technologies.
- In one approach, sequences are synthesized on a slide using ink-jet technology similar to that used in color printers. Separate cartridges for the four bases (A, C, G, T) are used to build nucleotides on a slide.
- The company Affymetrix uses a photolithographic approach which we will describe briefly.

Affymetrix GeneChips

- Affymetrix (www.affymetrix.com) is a company that manufactures GeneChips.
- GeneChips are oligonucleotide arrays.
- Each gene (more accurately sequence of interest or *feature*) is represented by multiple short (25-nucleotide) oligo probes.
- Some GeneChips include probes for around 60,000 genes.
- mRNA that has been extracted from a biological sample can be labeled (dyed) and hybridized to a GeneChip in a manner similar to that described for cDNA microarrays.
- Only one sample is hybridized to each GeneChip rather than two as in the case of cDNA microarrays.

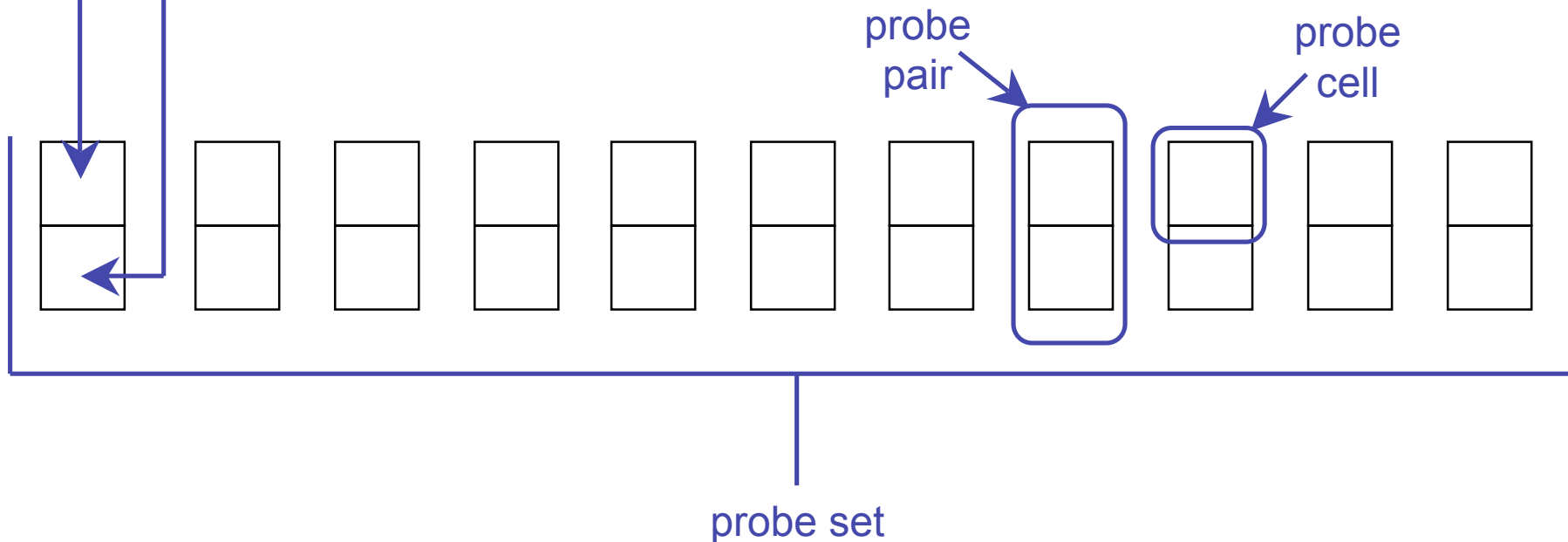
Affymetrix Probe Sets

- A *probe set* is used to measure mRNA levels of a single gene.
- Each probe set consists of multiple *probe cells*.
- Each probe cell contains millions of copies of one oligo.
- Each oligo is intended to be 25 nucleotides in length.
- Probe cells in a probe set are arranged in *probe pairs*.
- Each probe pair contains a *perfect match* (PM) probe cell and a *mismatch* (MM) probe cell.
- A PM oligo perfectly matches part of a gene sequence.
- A MM oligo is identical to a PM oligo except that the middle nucleotide (13th of 25) is intentionally replaced by its complementary nucleotide.

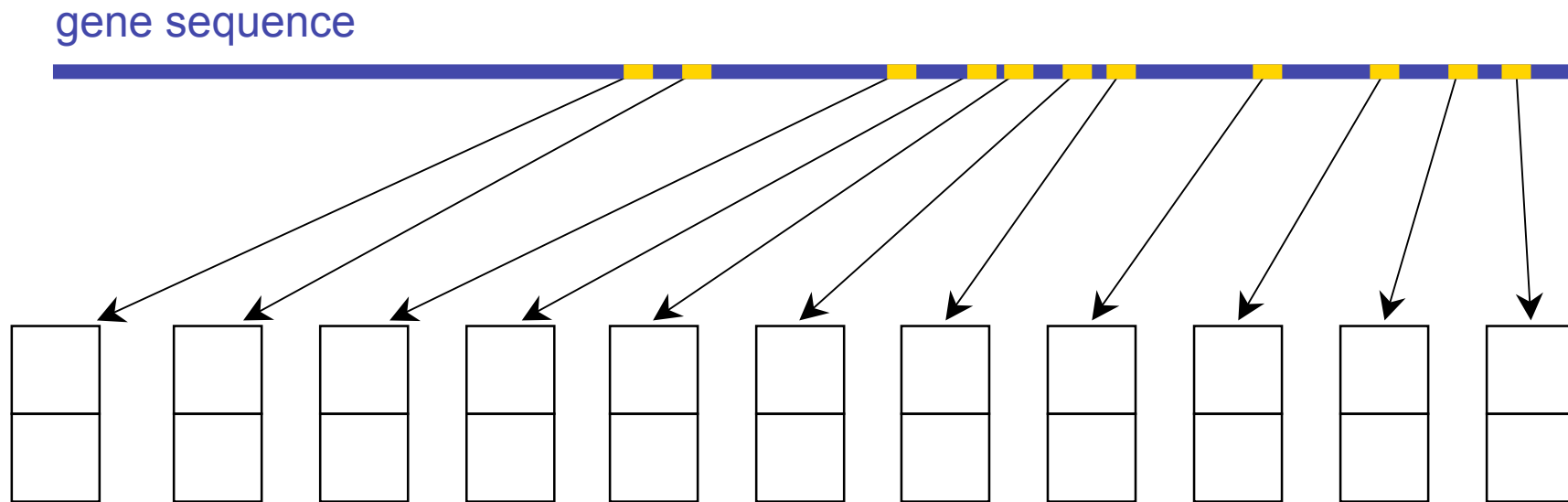
A Probe Set for Measuring Expression

Level of a Particular Gene

...TGCAATGGGTCAGAA**G**GACTCCTATGTGCCT... ← gene sequence
AATGGGTCAGAA**G**GACTCCTATGTG ← perfect match sequence
AATGGGTCAGAA**C**GACTCCTATGTG ← mismatch sequence

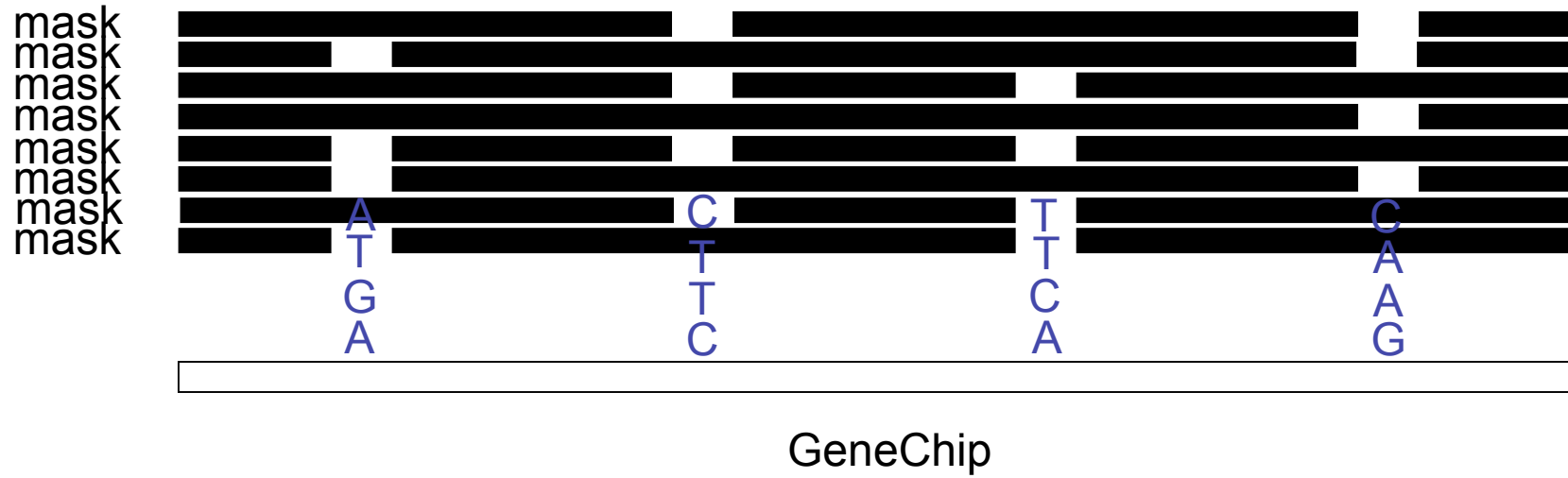


Different Probe Pairs Represent Different Parts of the Same Gene



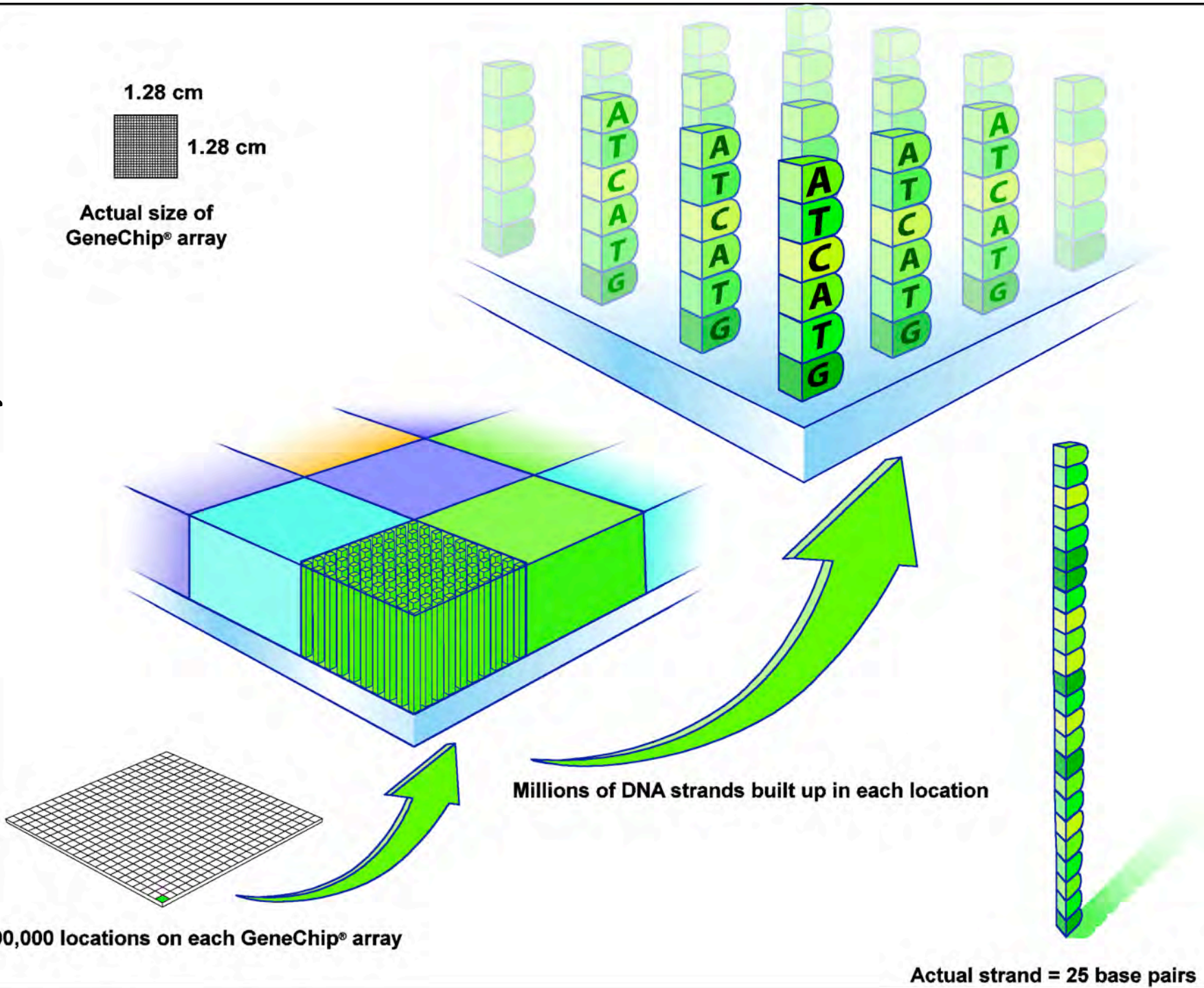
Probes are selected to be specific to the target gene and have good hybridization characteristics.

Affymetrix's Photolithographic Approach



Source: www.affymetrix.com

1.28 cm
1.28 cm
Actual size of
GeneChip® array

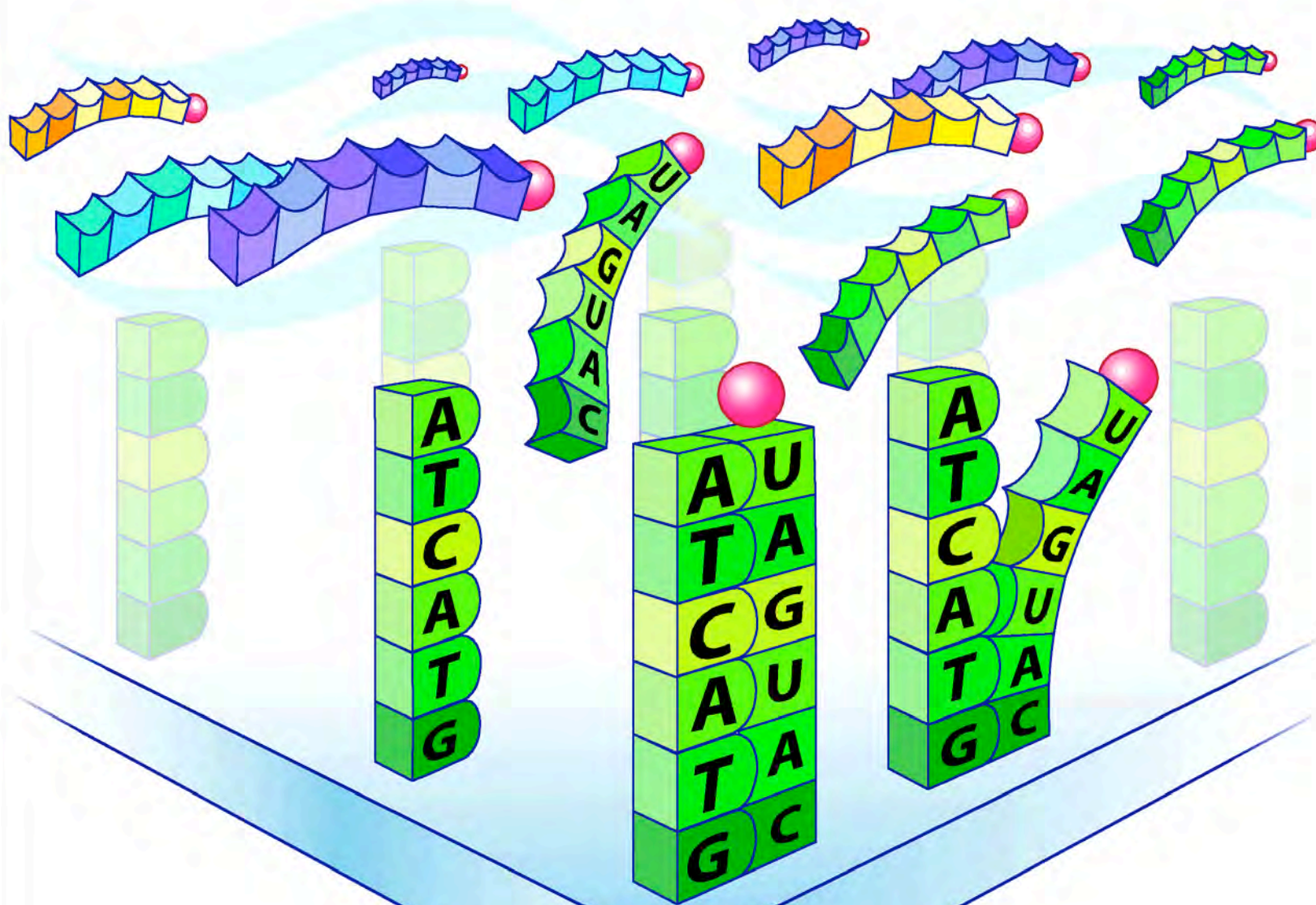


500,000 locations on each GeneChip® array

Millions of DNA strands built up in each location

Actual strand = 25 base pairs

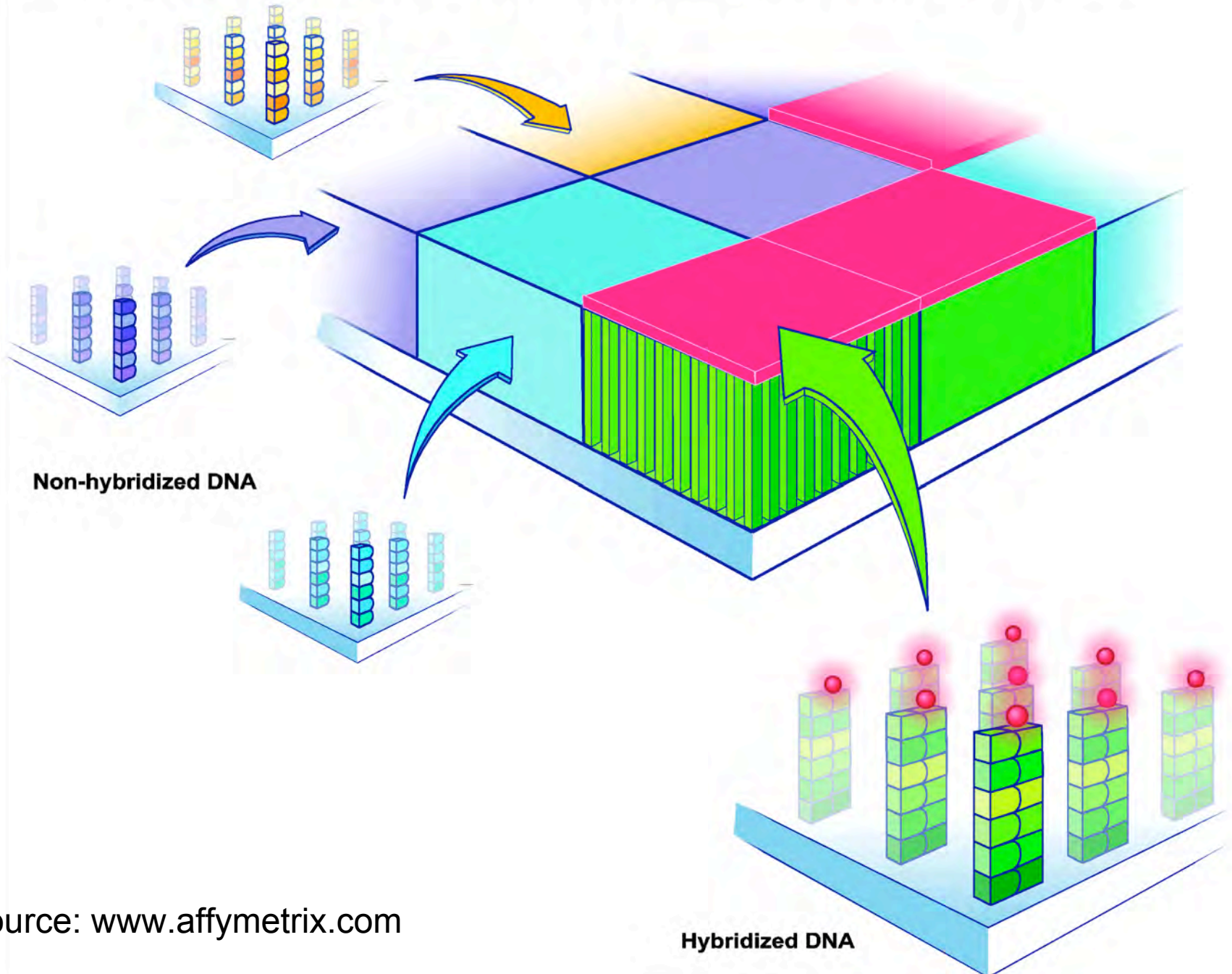
RNA fragments with fluorescent tags from sample to be tested



Source: www.affymetrix.com

RNA fragment hybridizes with DNA on GeneChip® array

Shining a laser light at GeneChip® array causes tagged DNA fragments that hybridized to glow



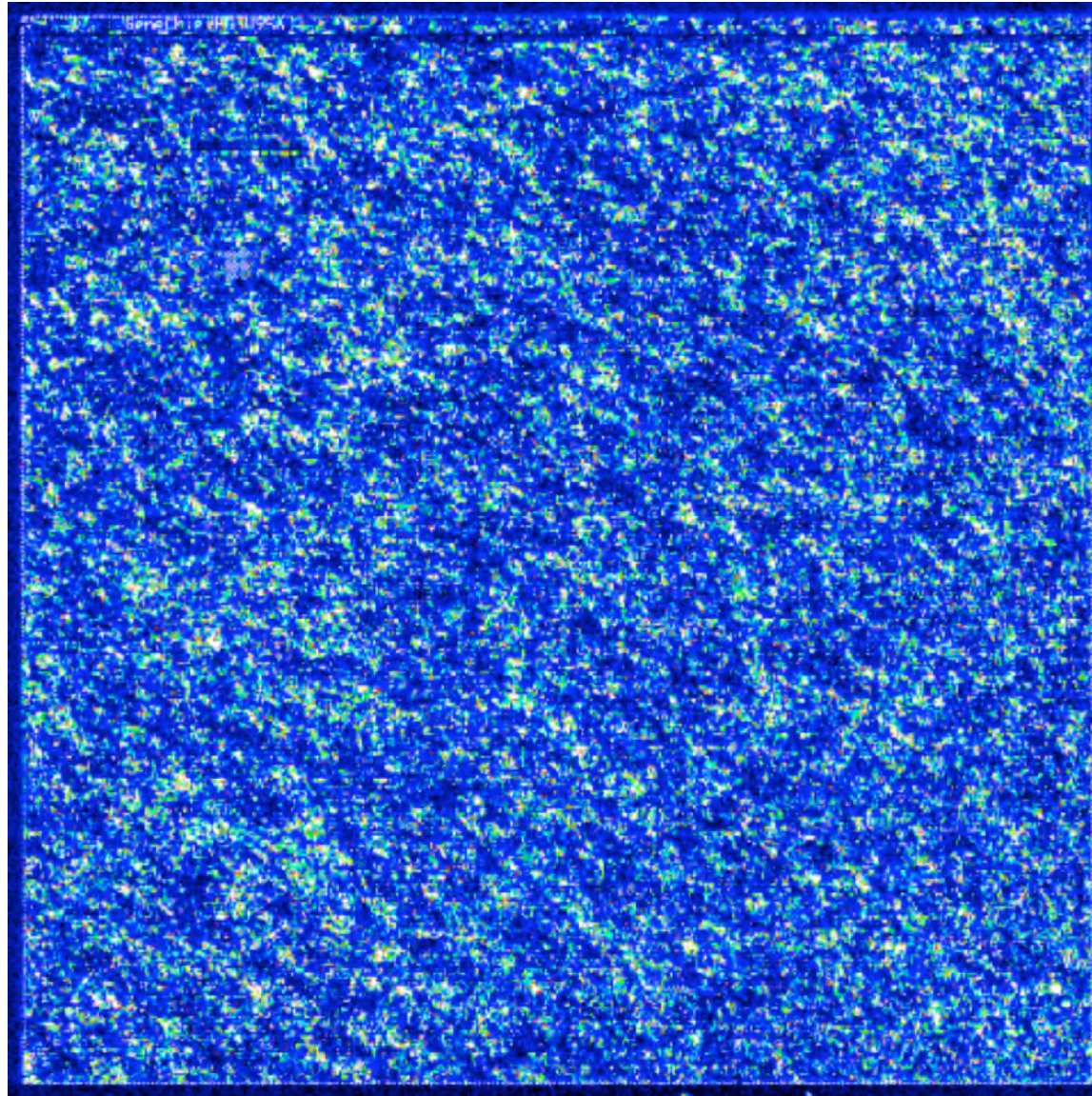
Source: www.affymetrix.com

Obtaining Labeled Target

1. RNA → single strand cDNA
2. single strand cDNA → double strand cDNA
3. double strand cDNA → labeled single strand cRNA complementary to coding sequence

Number of copies of each sequence gets amplified in conversion to cRNA.

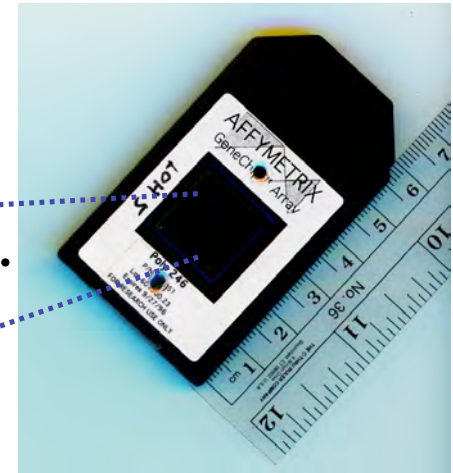
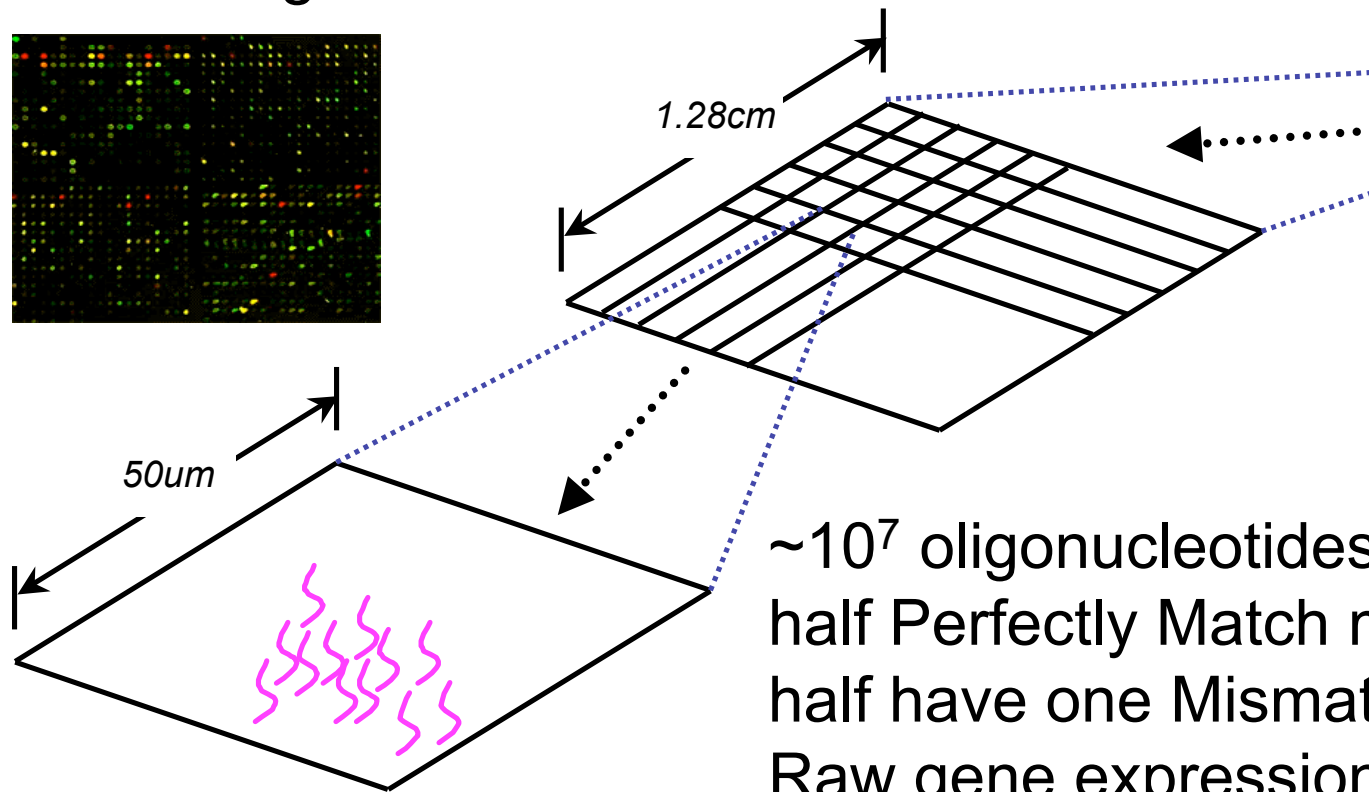
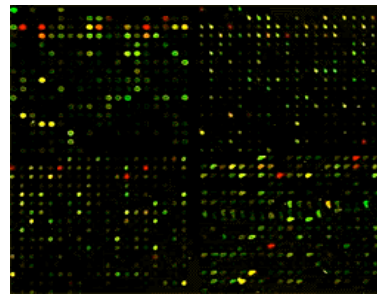
Image from Hybridized GeneChip



Source: www.affymetrix.com

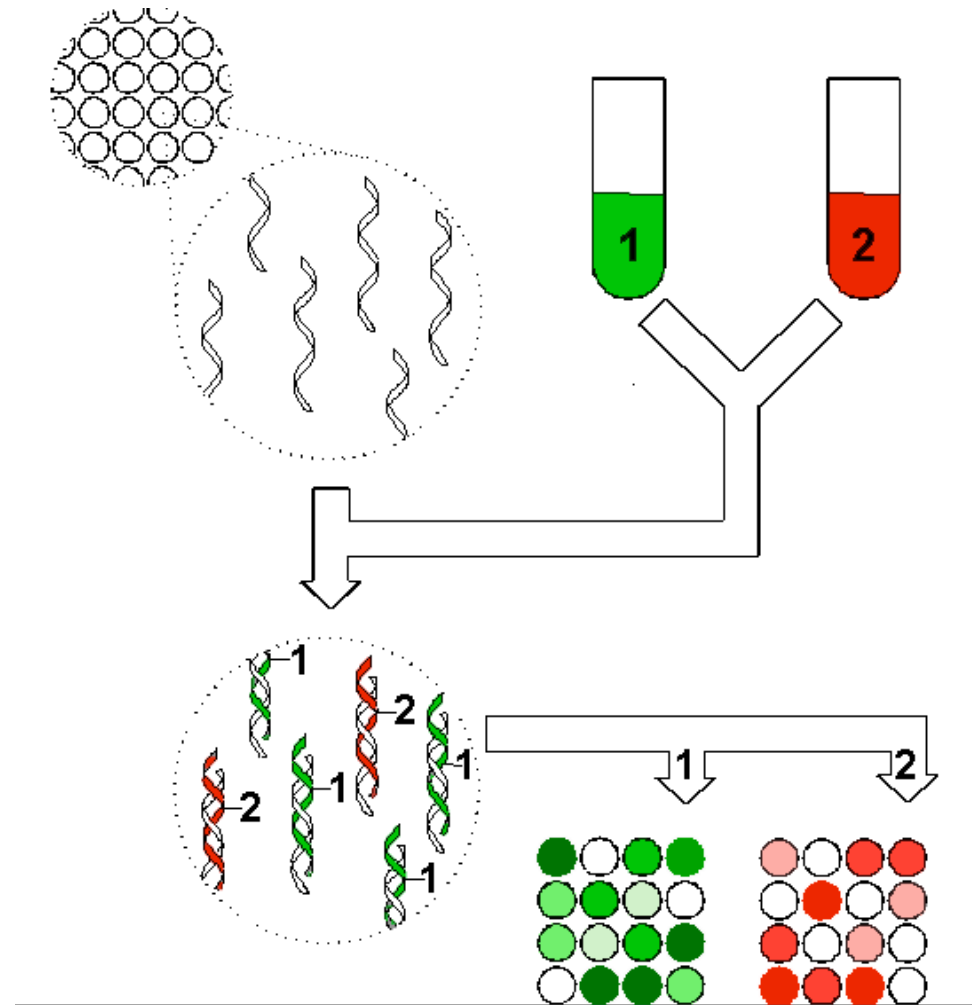
Affymetrix Microarrays

Raw image

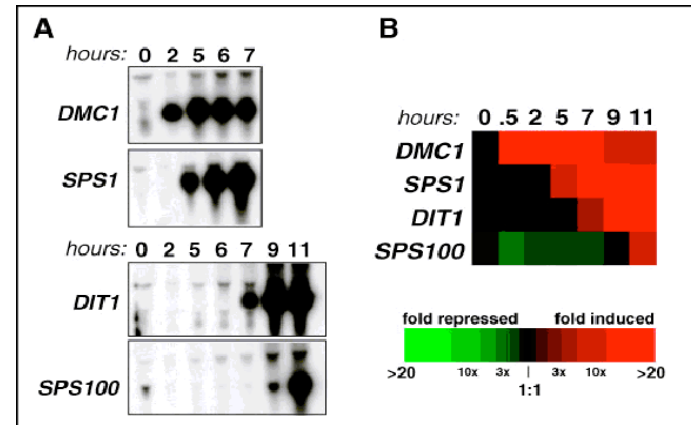
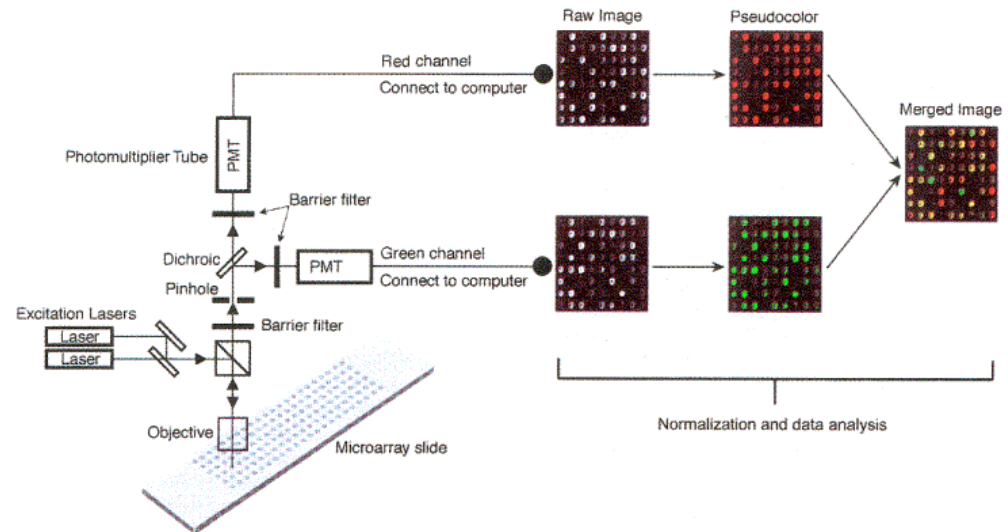
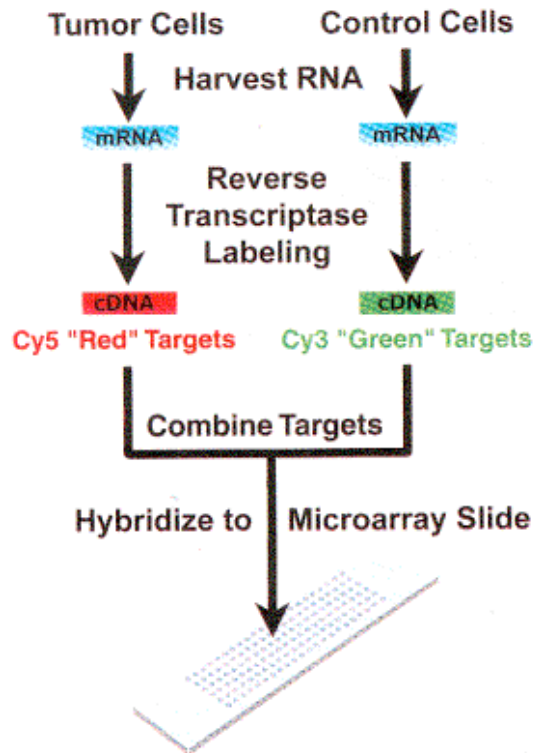


$\sim 10^7$ oligonucleotides,
half Perfectly Match mRNA (PM),
half have one Mismatch (MM)
Raw gene expression is intensity
difference: PM - MM

Hybridization Process

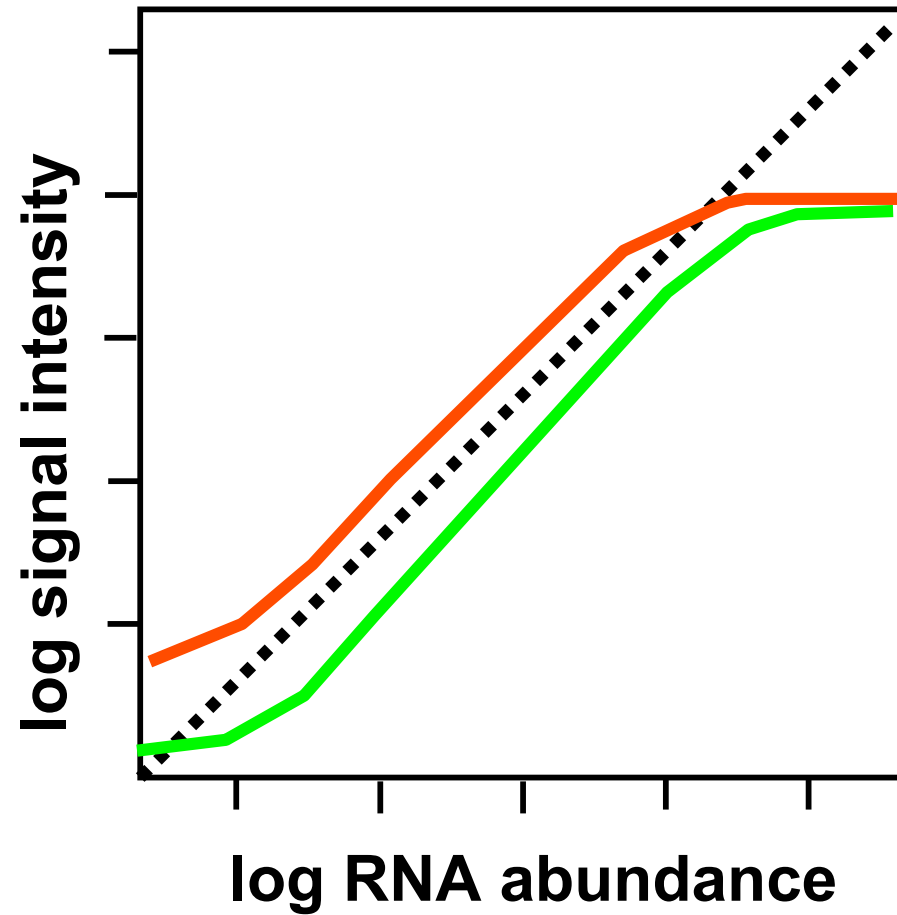


Tumor Cell Analysis



Sources of Error

- Systematic
- Random

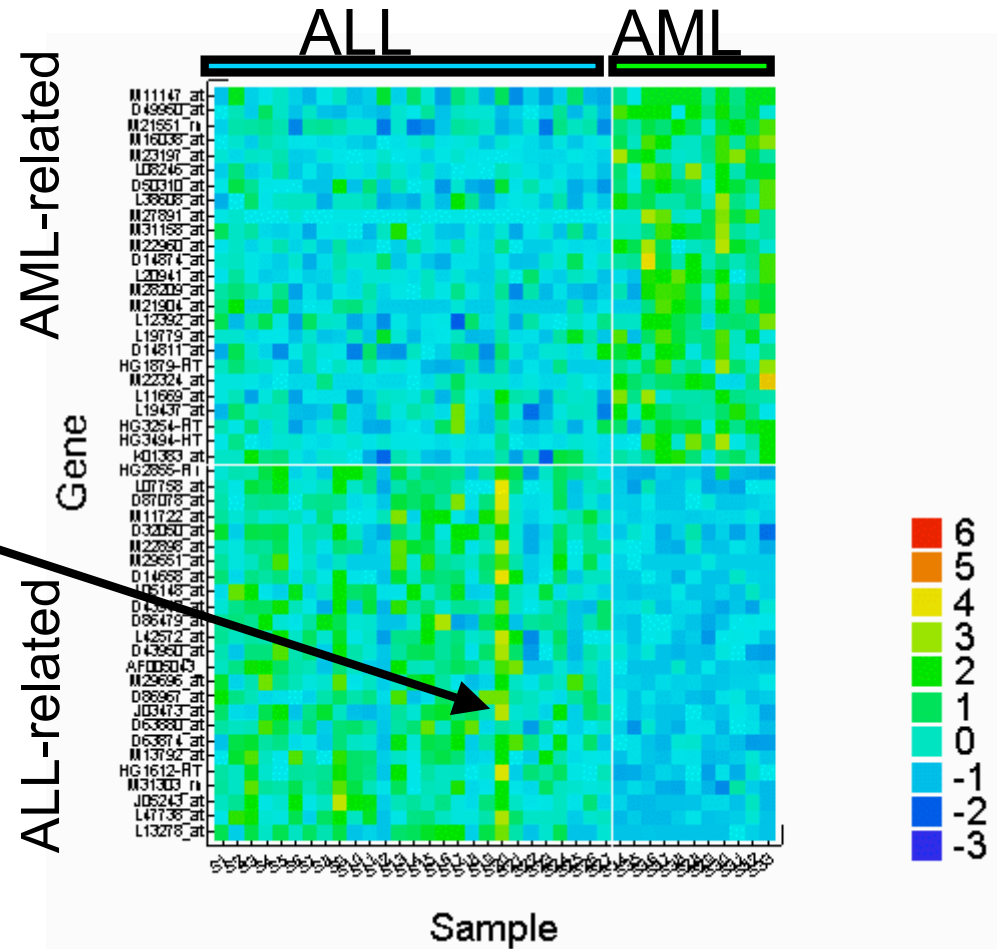
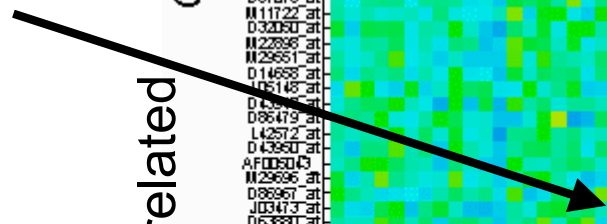


Heatmap Visualization of Selected Fields

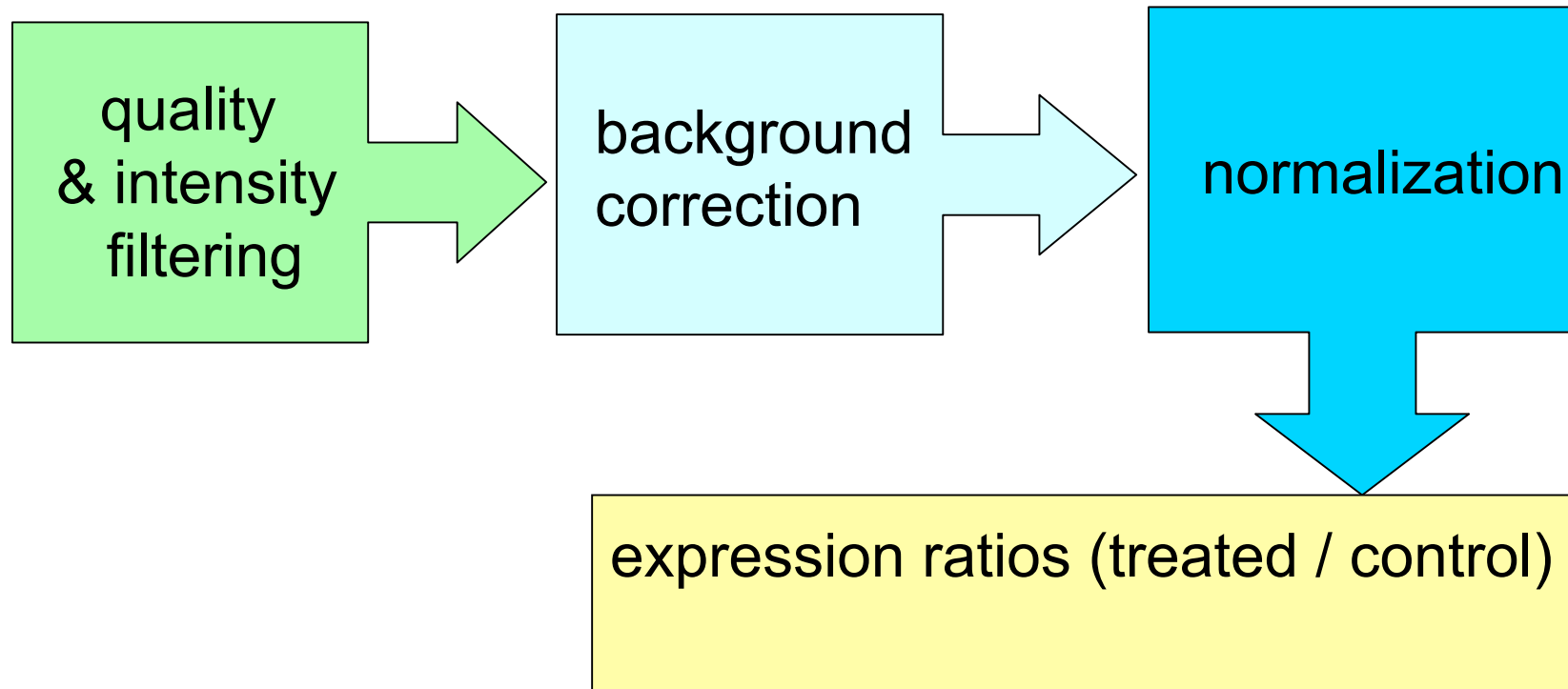
Heatmap visualization is done by normalizing each gene to mean 0, std. 1 to get a picture like this.

Good correlation overall

Possible outliers



Microarray Data Processing



Analysis Tasks

- Identify up- and down-regulated genes.
- Find groups of genes with similar expression profiles (++ / -- , fold change).
- Find groups of experiments (tissues) with similar expression profiles (++ / -- genes).
- Find genes that explain observed differences among tissues (feature selection), and new pathways.

Gene Expression

- Cells are different because of **differential gene expression**.
- About 40% of human genes are expressed at any one time.
- Gene is expressed by **transcribing** DNA exons into single-stranded mRNA
- mRNA is later **translated** into a protein
- Microarrays measure the level of mRNA expression by analyzing cDNA binding

Analysis of Gene Expression

- Examine expression during development or in different tissues
- Compare genes expressed in normal vs. diseased states
- Analyze response of cells exposed to drugs or different physiological conditions

Monitoring Changes in Genomic DNA

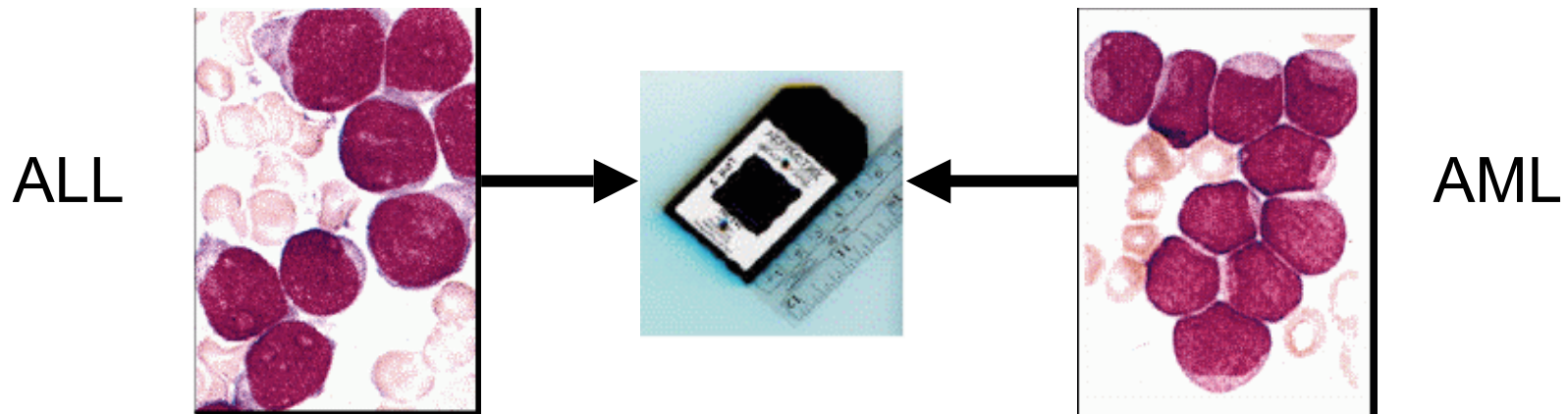
- Identify mutations
- Examine genomic instability such as in certain cancers and tumors (gene amplifications, translocations, deletions)
- Identify polymorphisms (SNPs)
- Diagnosis: chips have been designed to detect mutations in p53, HIV, and the breast cancer gene BRCA-1

Applications in Medicine

- Gene expression studies
 - Gene function for cell state change in various conditions (clustering, classification)
- Disease diagnosis (classification)
- Inferring regulatory networks
- Pathogen analysis (rapid genotyping)

Microarrays: An Example

- Leukemia: Acute Lymphoblastic (ALL) vs Acute Myeloid (AML), Golub et al, Science, v.286, 1999
 - 72 examples (38 train, 34 test), about 7,000 genes
 - well-studied (CAMDA-2000), good test example



Visually similar, but genetically very different

Applications in Drug Discovery

- Drug Discovery
 - Identify appropriate molecular targets for therapeutic intervention (small molecule / proteins)
 - Monitor changes in gene expression in response to drug treatments (up / down regulation)
 - Analyze patient populations (SNPs) and response
- Targeted Drug Treatment
 - Pharmacogenomics: individualized treatments
 - Choosing drugs with the least probable side effects