

# Statistics for Human Genetics and Molecular Biology

## Lecture 22: Introduction to Microarray Technology

Dr. Yen-Yi Ho (yho@umn.edu)

Oct 28, 2015

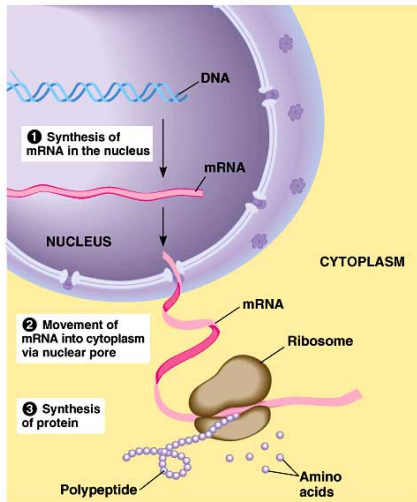
# Objectives of Lecture 22

- ▶ Types of Arrays
- ▶ Expression Microarrays
  - ▶ General Experiment Protocol
  - ▶ Hybridization, Target, Probe, Probe Set
  - ▶ Commercial Microarrays
  - ▶ Common Issues
- ▶ aCGH
- ▶ ChIP-chip

# Microarrays

- ▶ Microarrays (or just arrays) allow the simultaneous measurement of the mRNA of thousands of genes (gene expression)
- ▶ Why “micro”? Measuring things on a small scale means less mRNA is necessary to be extracted
- ▶ The simultaneous measurement is important, so that various factors/conditions are all common to the array, and relative gene expression can be measured with less variability
- ▶ Still microarray? Microarray is still widely used because of lower costs, easier experimental procedure and more established analysis methods.
- ▶ Similar problems are presented in newer technologies such as RNA-seq, and similar statistical approaches can be applied.

# Types of Arrays



## 1. DNA

- SNP arrays: genotype studies
- array CGH: DNA Copy Number
- ChIP-chip: TF binding sites

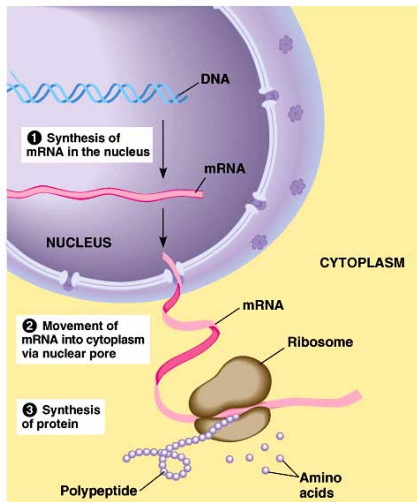
## 2. mRNA

- DNA microarrays: custom array, Affymetrix, Agilent, Illumina
- SAGE

## 3. Proteins

- 2-D electrophoresis
- Maldi-Tof mass spec

# Types of Arrays



## 1. DNA

- SNP arrays: genotype studies
- array CGH: DNA Copy Number
- ChIP-chip: TF binding sites

## 2. mRNA

- custom array, Affymetrix, Agilent, Illumina
- SAGE

## 3. Proteins

- 2-D electrophoresis
- Maldi-Tof mass spec

## **Micorarray**

Affymetrix Cost: \$425/per array

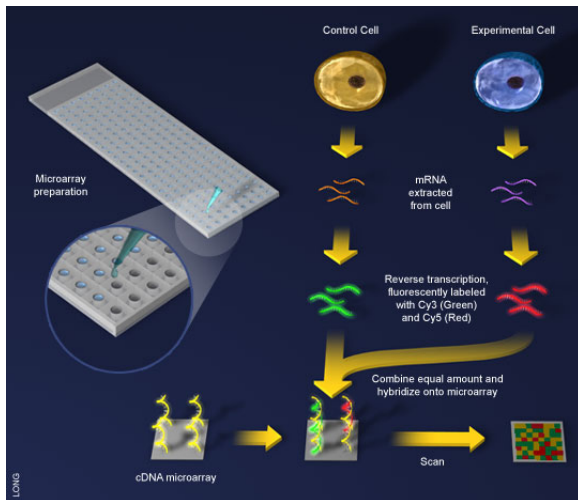
Illumina BeadChip Cost: \$ 80 ~ \$250 /per strip  $\times$  12 = \$960 ~ \$3,000 /chip

## **Next-Generation Sequencing**

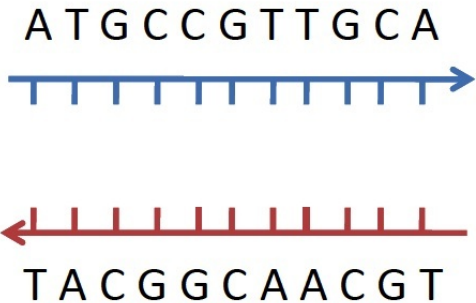
Illumina ~\$2,000 per samples (for 1-7 samples)

Source: University of Minnesota Genomics Center

# Steps in Obtaining Gene Expression Data Using Microarray

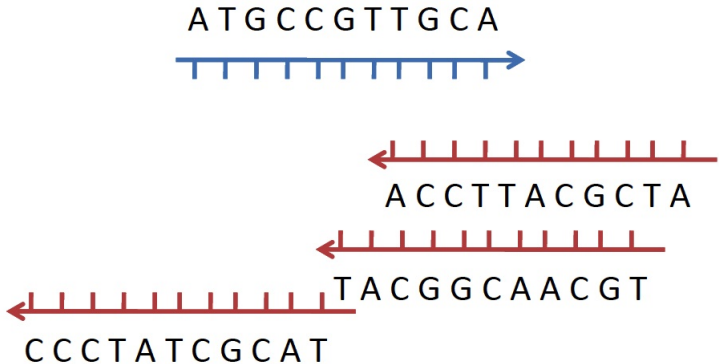


# Denaturation





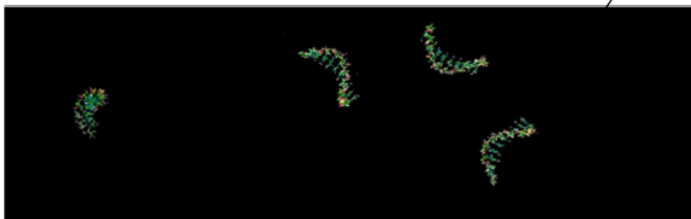
# Hybridization



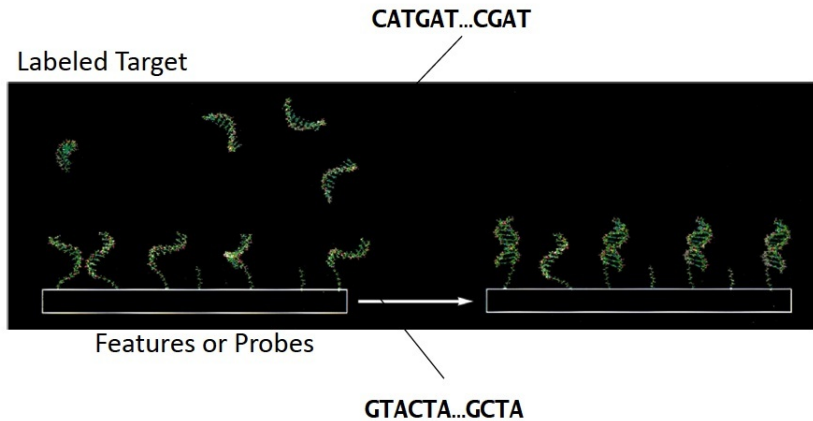
# Target

Target

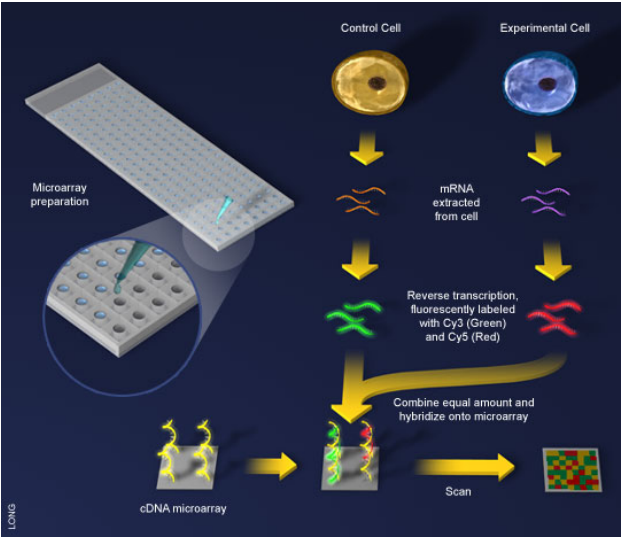
CATGAT...CGAT



# Probes



# Microarray Experiment Procedure

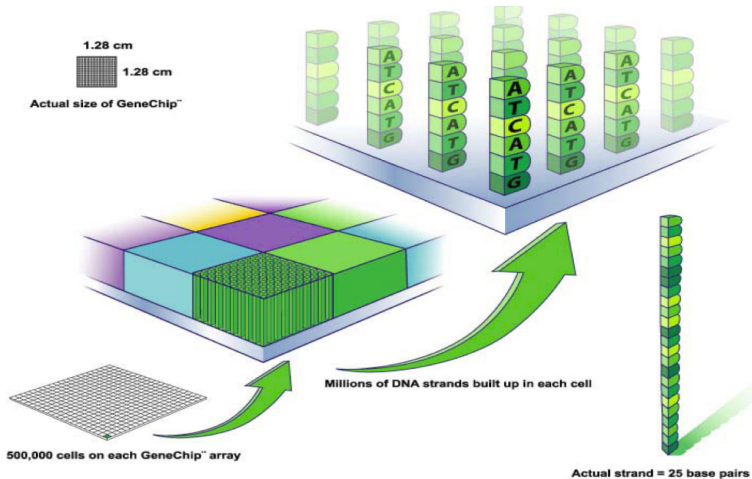


LONG

# Platforms that dominate market

- ▶ Affymetrix
- ▶ Agilent
- ▶ Illumina
- ▶ NimbleGen

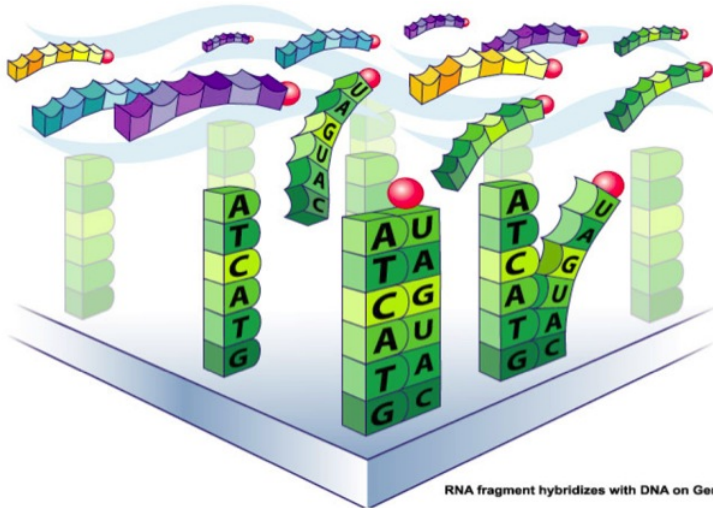
# Affymetrix GeneChip®



source: Affymetrix

# Affymetrix GeneChip®

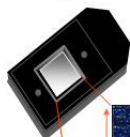
RNA fragments with fluorescent tags from sample to be tested



source: Affymetrix

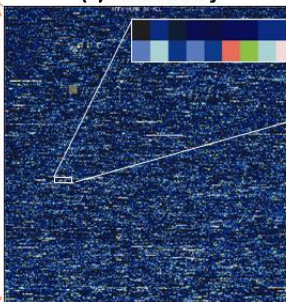
# Affymetrix GeneChip®

## Human Genome U133A GeneChip® Array



1.28cm

### (1) Probe Array



### (2) Probe Set

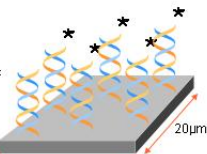
Each Probe Set contains  
11 Probe Pairs (PM:MM)  
of different probes

### (3) Probe Pair

Each Perfect Match  
(PM) and MisMatch  
(MM) Probe Cells are  
associated by pairs

### (4) Probe Cell

Each Probe Cell contains  
 $\sim 40 \times 10^7$  copies of a specific  
probe  
complementary to genetic  
information of interest  
probe: single stranded,  
sense, fluorescently labeled  
oligonucleotide (25 mers)

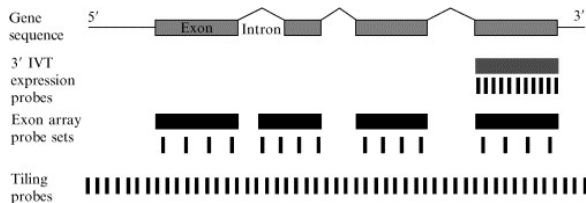


The Human Genome U133 A  
GeneChip® array represents  
more than 22,000 full-length  
genes and EST clusters.



# Probe Set

A probe set is a collection of probes designed to interrogate a given sequence.



# PM versus MM

## GeneChip® Expression Array Design

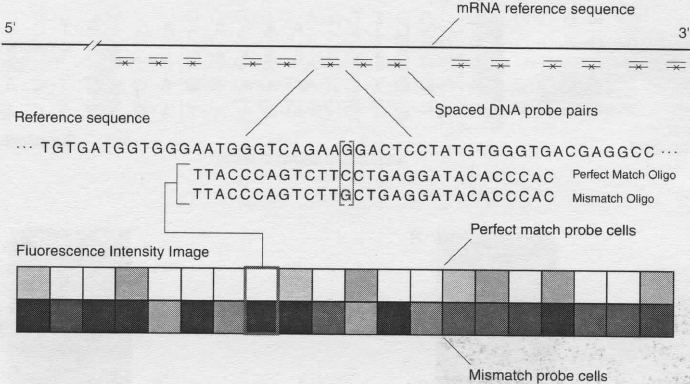
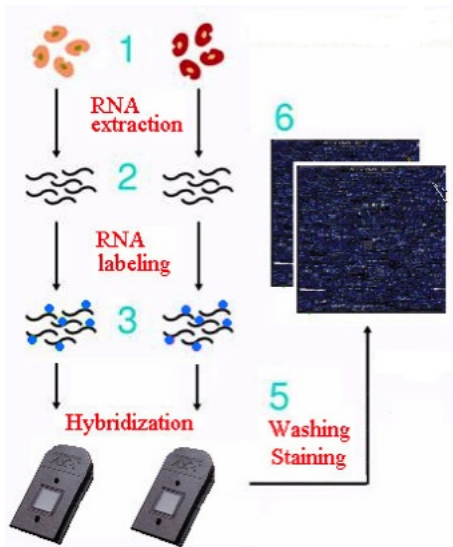
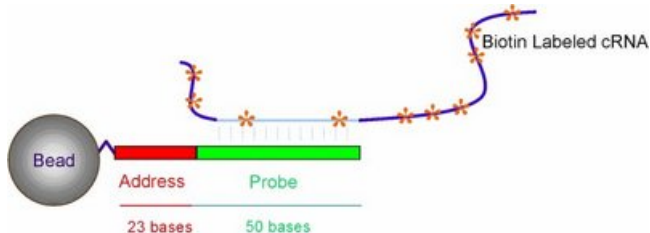


Figure 1-3 Expression tiling strategy

# Affymetrix GeneChip® Experiment Protocol



# Illumina BeadArrays



- Each silica bead is 3  $\mu$  in diameter
- 700,000 copies of same probe sequence attached to each bead
- May have more than one bead for a particular gene

# Illumina BeadArrays

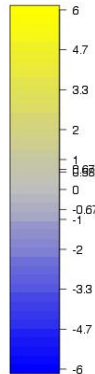
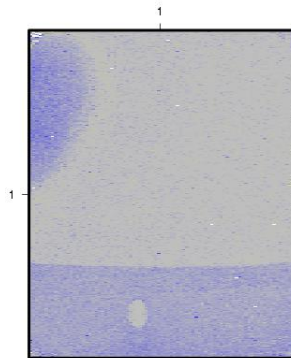
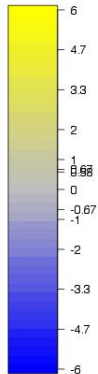
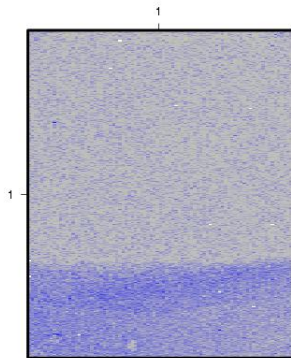


- RefSeq BeadChip (left) 8 arrays per chip, 1 strip= 1 array
- Whole Genome (right) 6 arrays per chip, 2 strip= 1 array

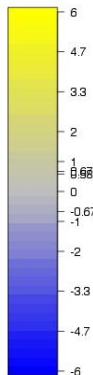
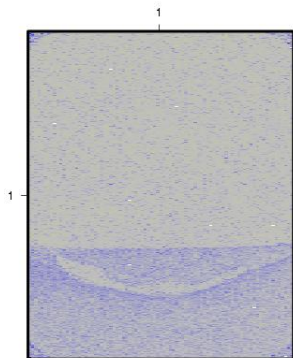
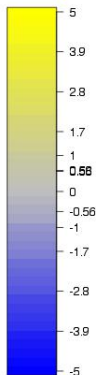
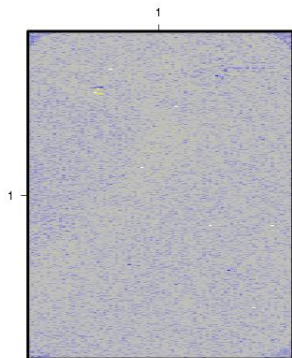
# Commons Issues

- ▶ Background: print-tip, plate, print-order, spatial effects
- ▶ Between arrays: batches, plates, cross platform comparison, experiment protocols
- ▶ Within arrays: background noise, intensity dependent effect

# Spatial Effect

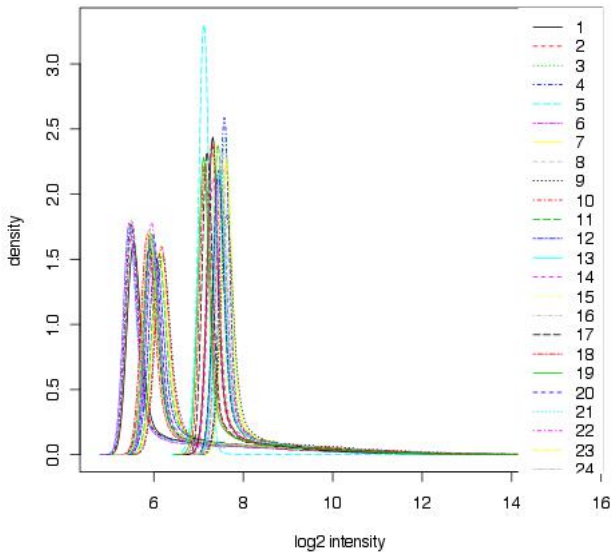


# Spatial Effect

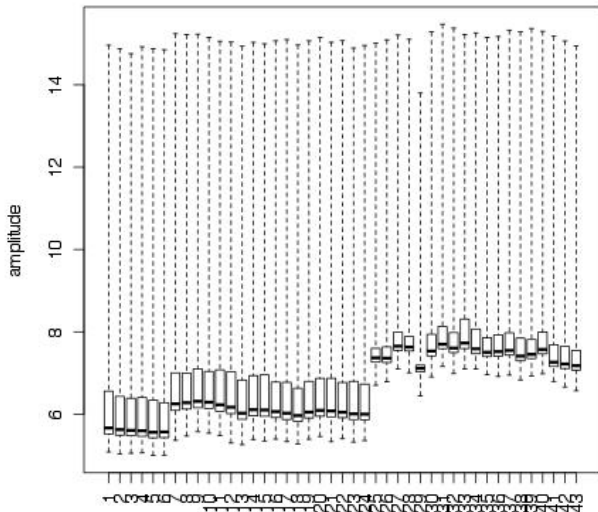




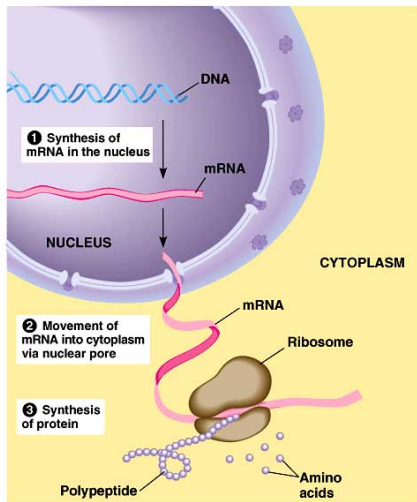
## Batch Effect



# Batch Effect



# Types of Arrays



## 1. DNA

- SNP arrays: genotype studies
- array CGH: DNA Copy Number
- ChIP-chip: TF binding sites

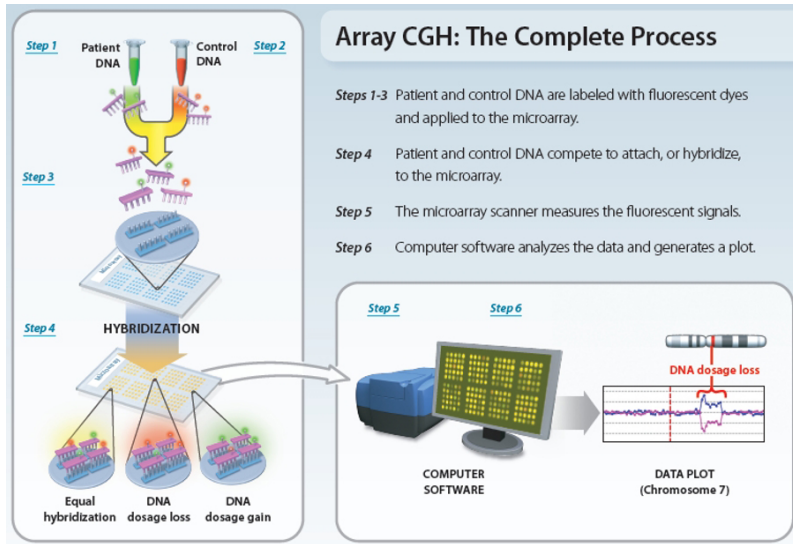
## 2. mRNA

- DNA microarrays: cDNA, Affymetrix, Agilent
- SAGE

## 3. Proteins

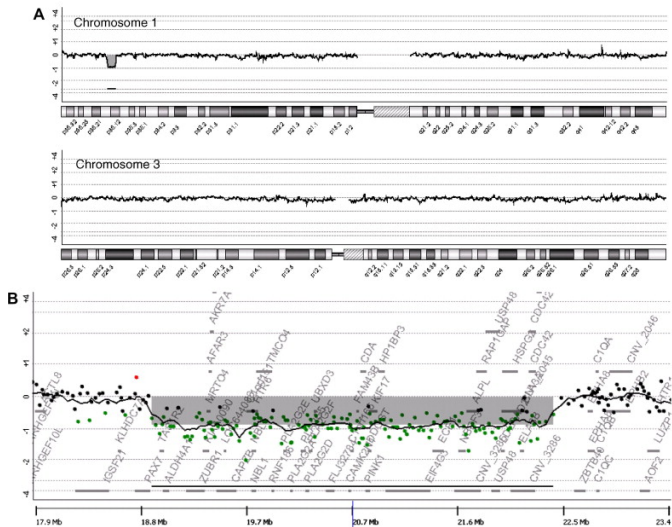
- 2-D electrophoresis
- Maldi-Tof mass spec

# Array-comparative genomic hybridization (aCGH)



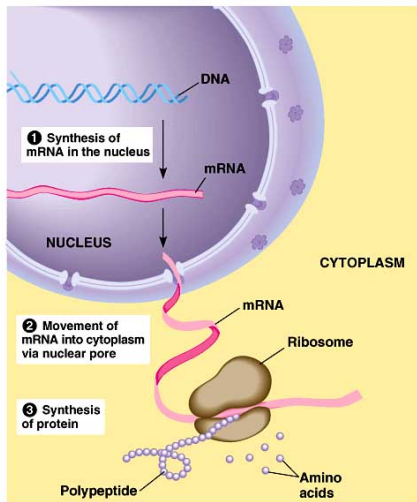
Nature Education1(1). Microarray-based Comparative Genomic Hybridization (aCGH)

# array CGH



Brian and Development 31(6):629:633.

# Types of Arrays



## 1. DNA

- SNP arrays: genotype studies
- array CGH: DNA Copy Number
- ChIP-chip: TF binding sites

## 2. mRNA

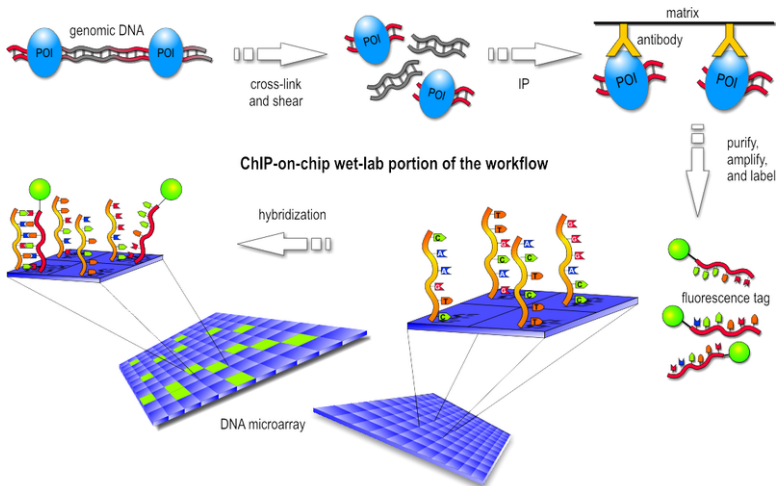
- DNA microarrays: cDNA, Affymetrix, Agilent
- SAGE

## 3. Proteins

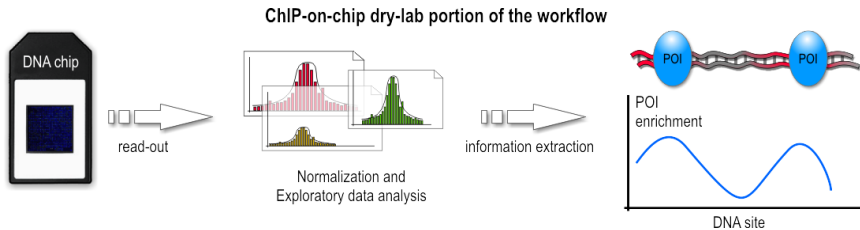
- 2-D electrophoresis
- Maldi-Tof mass spec

# ChIP-chip

Chromatin immunoprecipitation ("ChIP") with microarray technology ("chip")

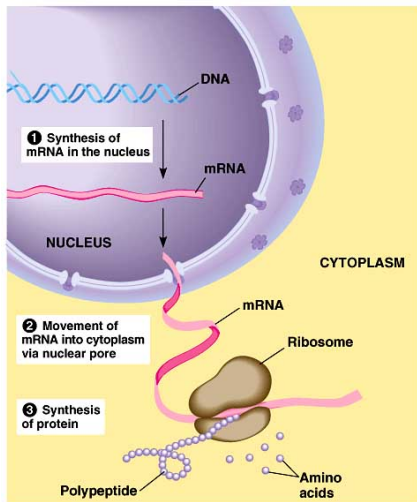


# ChIP-chip





# Types of Arrays



## 1. DNA

- SNP arrays: genotype studies
- array CGH: DNA Copy Number
- ChIP-chip: TF binding sites

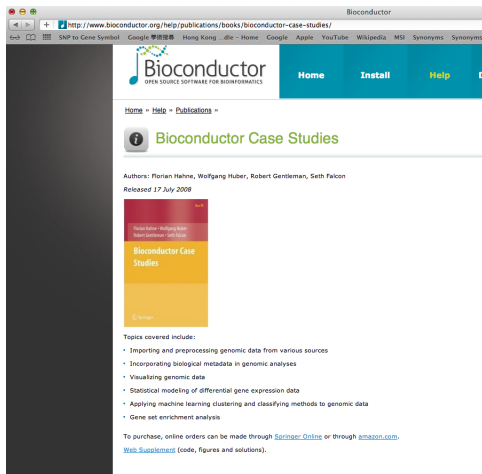
## 2. mRNA

- DNA microarrays: cDNA, Affymetrix, Agilent
- SAGE

## 3. Proteins

- 2-D electrophoresis
- Maldi-Tof mass spec

## Bioconductor Case Studies: Chapter 1, 2, 3 (page 1 -45).




The screenshot shows a web browser window displaying the Bioconductor Case Studies page. The browser's address bar shows the URL: <http://www.bioconductor.org/help/publications/books/bioconductor-case-studies/>. The page features the Bioconductor logo at the top left, which includes the text "Bioconductor" and "OPEN SOURCE SOFTWARE FOR BIOWORKFLOWS". To the right of the logo is a navigation bar with buttons for "Home", "Install", and "Help". Below the navigation bar, there is a breadcrumb trail: "Home > Help > Publications >". The main heading is "Bioconductor Case Studies" with an information icon to its left. Below the heading, the authors are listed: "Authors: Florian Hahne, Wolfgang Huber, Robert Gentleman, Seth Falcon" and the release date: "Released 17 July 2008". A book cover image is displayed, showing the title "Bioconductor Case Studies" and the authors' names. Below the book cover, the text "Topics covered include:" is followed by a bulleted list of topics. At the bottom, there is information about purchasing the book online through Springer-Online or Amazon.com, and a link to a "Web Supplement" containing code, figures, and solutions.

http://www.bioconductor.org/help/publications/books/bioconductor-case-studies/


Bioconductor  
OPEN SOURCE SOFTWARE FOR BIOWORKFLOWS

Home Install Help

Home > Help > Publications >

 Bioconductor Case Studies

Authors: Florian Hahne, Wolfgang Huber, Robert Gentleman, Seth Falcon  
Released 17 July 2008



Topics covered include:

- Importing and preprocessing genomic data from various sources
- Incorporating biological metadata in genomic analyses
- Visualizing genomic data
- Statistical modeling of differential gene expression data
- Applying machine learning clustering and classifying methods to genomic data
- Gene set enrichment analysis

To purchase, online orders can be made through [Springer-Online](#) or through [amazon.com](#).  
[Web Supplement](#) (code, figures and solutions).