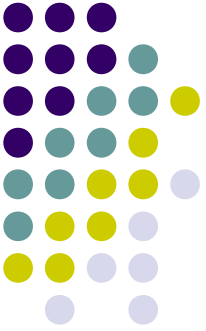
Lecture #1

Introduction to microarray technology



Outline

- General purpose
- Microarray assay concept
- cDNA arrays
- Oligonucleotide arrays
- Array design
- MIAME standards
- Shortcomings of the microarray technology

Purpose

- Understand the processes and associations both within and between genes (functional genomics)
  - Genetic diseases (many disorders are multifactorial)
    - disregulation
    - splice variants
    - SNPs
  - Pathogens
  - Drug discovery
- Gene interactions are complex in nature, such that it is necessary to assay many simultaneously
- Requires high-throughput technology

Microarray Concept

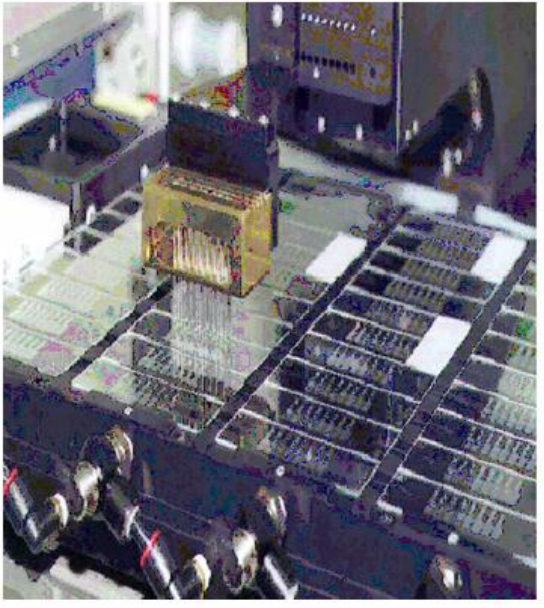
- Quantitative measure of mRNA
  - Since most changes in cell states are associated with mRNA
- General approach
  - Solid surface material (plate, slide, chip, etc.) with DNA sequence complementary to EST or gene of interest attached (probe)
  - DNA or RNA from sample is extracted, fragmented, and tagged (label)
  - Sample DNA/RNA is run over surface and sequences specific to probes hybridize
  - Laser is applied to label to cause fluorescence, providing quantitative measure of mRNA abundance
  - Expression signal is associated with gene expression value

cDNA Arrays

- cDNA arrays
  - Less expensive technology
  - Complete sequence is attached to chip (probe)
  - Two-color hybridization used for each probe
    - Internal control
    - cDNA is PCR'd with random 6mer primers and dCTP-dye conjugates
      - **Cy5** abs=650 nm; emm=667 nm
      - **Cy3** abs=552 nm; emm=568 nm
  - Utilizes spotting technology to attach probes to chip
    - Printers/robots

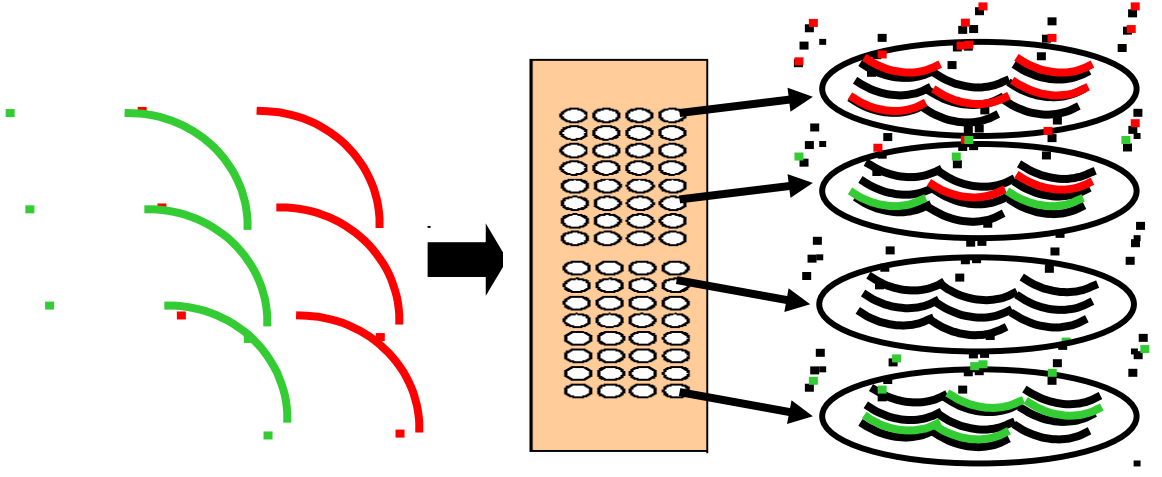
cDNA Technology

A glass slide, membrane, or polymer that has been spotted with non-labeled DNA probes designed to hybridize specific complementary DNA's of interest (cDNA).



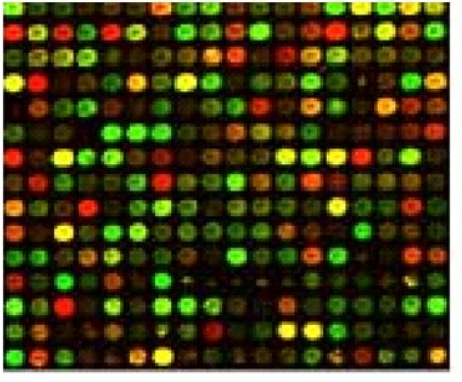
cDNA Technology (cont.)

Samples are prepared from both an experimental sample, (i.e. malignant tumor) and a control sample, (i.e.normal tissue) and are then over laid on the array and allowed to hybridize to each spotted probe:



cDNA Technology (cont.)

Following hybridization, the array is scanned and the resulting gene expression information for each spotted probe on the array is reported



A green intensity = Control Only expressed gene
A red intensity = Experiment Only expressed gene
A mixed color = gene expressed in both Control & Experiment

cDNA Technology (cont.)

The gene expression information that is actually reported are ratios of the amount of experiment sample to control sample that has hybridized to each spotted probe on the array

$$\text{Probe 1}\left[\frac{\text{Amount of experiment sample hybridized}}{\text{Amount of control sample hybridized}}\right] = \text{Probe 1 Expression Ratio}$$

$$\text{Probe 2}\left[\frac{\text{Amount of experiment sample hybridized}}{\text{Amount of control sample hybridized}}\right] = \text{Probe 2 Expression Ratio}$$

$$\text{Probe n}\left[\frac{\text{Amount of experiment sample hybridized}}{\text{Amount of control sample hybridized}}\right] = \text{Probe n Expression Ratio}$$

Oligonucleotide Arrays

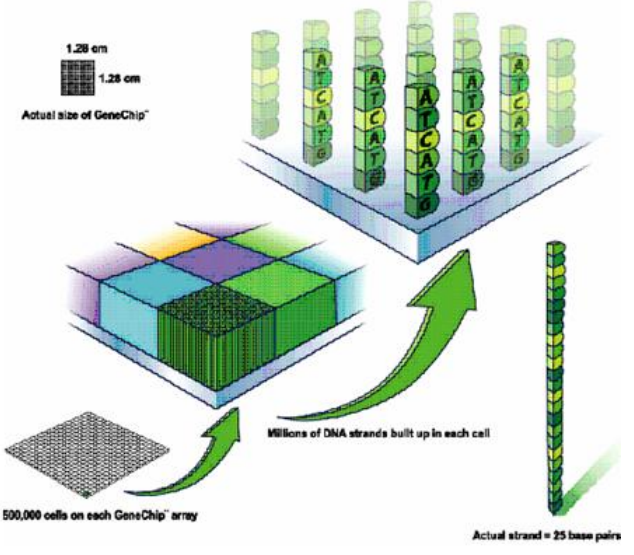
- Oligonucleotide array - Affymetrix
  - More expensive technology
  - Small (11-25 mers) or large (50-70 mers) sequence is attached to chip (probe)[2]
  - Allows for non-repetitive or unique probe design for a particular gene
  - Multiple probes represent same gene/EST with overlap method
  - Each probe has a mismatch complement with single bp mutation
    - Cross-hybridization
    - Background correction
  - Utilizes photolithography technology to attach probes to chip

Oligo Technology (cont.)

Every microarray has up to 500,000 individual probe-cells, each 18µm across and containing millions of identical DNA molecules.[2]

The human U133A array, for example contains over 260,000 different probes that together measure the expression of 22,283 different transcripts at once.[2]

Chips exist for a variety of organisms including human, mouse, yeast, arabidopis, and rat



Oligo Technology (cont.)

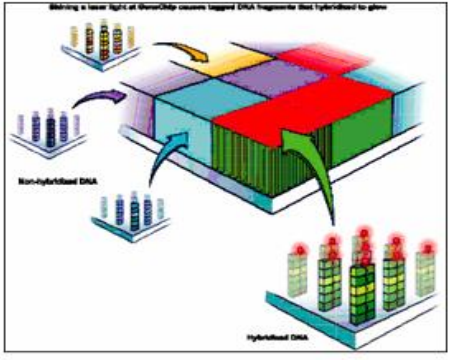
Fragmented RNA is labeled with a fluorescent tag and run over the chip.[2]

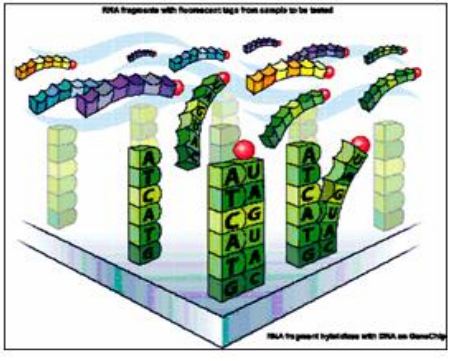
Wherever there is a complementary probe sequence on the chip, the RNA can hybridize to it.[2]

Since there are millions of oligos for each probe-sequence, the amount of labeled RNA that sticks corresponds to the amount in solution.[2]

When the chip is scanned by a laser, the tagged fragments fluoresce, producing spots with a brightness proportional to the amount of RNA that has hybridized.[2]

This is recorded by a camera and the array image processed by computer to produce expression levels for the different genes.[2]





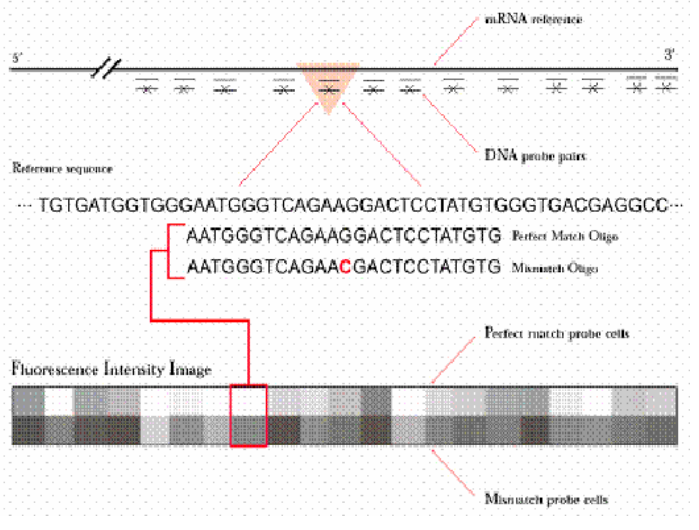
Oligo Technology (cont.)

The chips are designed so that every transcript is represented by between 11 to 20 probes that match different parts of the 3' end of the mRNA sequence.[2]

Every chip probe consists of a pair 25 base oligos, one a perfect match (PM) to the transcript, the other a mismatch (MM) in which the middle residue has been changed.[2]

This probe-pairing strategy helps minimize the effects of non-specific hybridization and background signal.[2]

Figure 3: Oligonucleotide Probe Pair



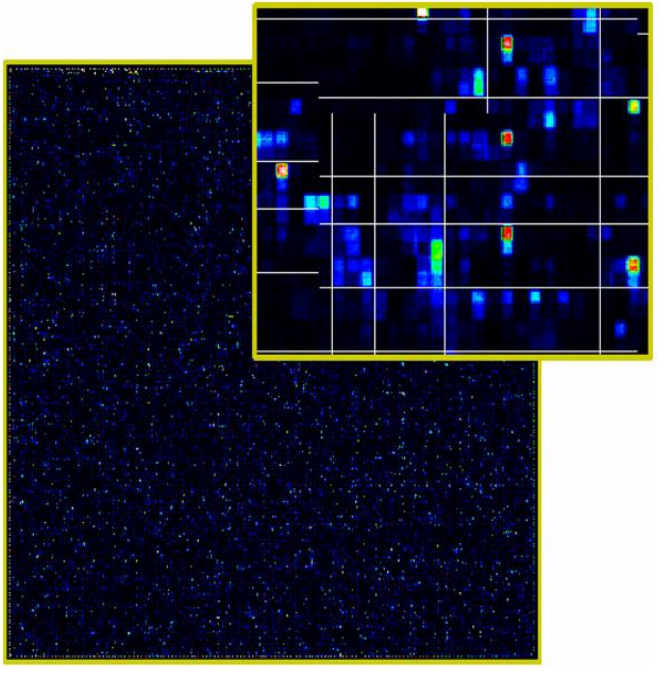
Oligo Technology (cont.)

Once the probe has hybridized, chips are scanned to generate an image (dat file).[2]

Each spot, or feature, is 20µm square and is scanned at a resolution of 3µm - giving an average of 49 pixels per spot.[2]

The array analysis software identifies individual features and overlays a grid separating each spot from its neighbors. [2]

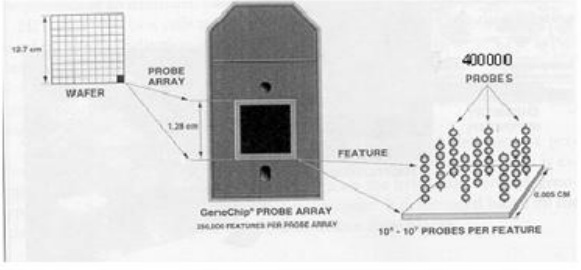
The expression level for a gene is calculated by subtracting the MM from the PM probes.[2]



Oligo Technology (cont.)

Fluidics machine, scanner, and software.

Figure 2:Manufacturing GeneChip probe array



An Affymetrix chip.

Array Design

- Density
  - Low density arrays are utilized for assaying < 100 genes
  - High density arrays are utilized for assaying 1,000's of genes
- Probe selection
  - Optimal probes for each gene minimizes background hybridization
    - More accurate measure of true expression
  - Optimal probes maximize a unique representation for each gene
    - Less probes mapped to each gene, if selection is unique
  - Continuing topic of work

MIAME

- Minimal Information About a Microarray Experiment (MIAME)
  - Organization set up to provide standards in microarray experiments and analysis
  - Provide guidelines on the minimal necessary information for interpretable results
  - Encourage depositing data into public standard repositories
    - Journals and funding agencies
- Guideline examples
  - Experimental design
  - Array design
  - Samples
  - Hybridization parameters
  - Normalization methods

Shortcomings of the Technology

- Hybridization kinetics
  - Ideally, the probe that minimizes hybridization free energy is the optimal one to represent a gene
    - However, we cannot currently compute the free energy from the sequence alone
  - The hybridization free energy for a gene depends on the concentration of that gene
    - The less expressed gene with higher free energy can give a greater signal than the more expressed gene, if it is given in greater concentration
- mRNA expression vs. protein expression
  - Gene interactions **can have** little effect on protein interactions
    - kinases, receptor-ligand binding, protein docking, etc.
- All gene expression events do not result in mRNA transcripts
  - tRNA, rRNAm snRNA
- mRNA splice variants
- Processing variability vs. biological variability

References

- [1]Li Fugen, Stormo D Gary,. (2001) Selection of optimal DNA oligos for gene expression arrays. *Bioinformatics.* **17**, 1067-1076.
- [2]The Paterson Institute: Onco-Informatics group
  - http://bioinformatics.picr.man.ac.uk/mbcf/overview_ma.jsp