Microarrays

CAP 5937-01 Bioinformatics
Fall 2004
Amar Mukherjee

Molecular Biology Laboratory Techniques :Hybridization

- Given a short (8- to 30-nuclreotides) synthetic fragment DNA, called ***probe***, a target single stranded DNA molecule (produced by denaturing) will ***hybridize*** or bind to the probe if there is a substring in the target sequence that is complementary to the probe.
- For example, a target DNA sequence *CCCTGGCACCTA* will hybridize to a probe *ACCGTGGA* since the complementary sequence *TGGCACCT* is present in the target.
- In a mix of DNA sequences, the presence of a particular DNA can be tested by making the probe fluorescent or radioactive. This idea has lead to the development of DNA chips or ***microarrays*** that allows rapid DNA sequencing.

DNA Microchip: Microarrays- A Large-scale Biological Tool

- An excellent presentations about microarray:
  - http://www.ncbi.nlm.nih.gov/About/primer/microarrays.html
- We will discuss here only the principle and show how a DNA sequence can be derived by using a microarray.

More Tutorials on Microarrays

2. **http://www.accessexcellence.org/AB/GG/**
Graphics gallery of biotechnology related topics. Also contains a detailed explanation of every figure.
3. **http://www.bio.davidson.edu/courses/genomics/chip/chip.html**
This is a super comprehensive flash animation of the microarray methodology. A must see web site for bioinformatics enthusiastic.
4. **http://brainarray.mhri.med.umich.edu/Brainarray/analysis.html**
This site contains a comprehensive guide to some of the wet lab procedures/protocols underlying microarray experiments.
5. **http://www.indstate.edu/thcme/mwking/home.html**
This site contains dozens of tutorials on diverse biochemistry topics. Good source for understanding gene regulation, protein synthesis, DNA replication, DNA hybridization, etc.
6. **http://www.affymetrix.com**
The Affymetrix website. Good source for understanding the commercially available options and packages.
7. **http://www.ncbi.nlm.nih.gov/About/primer/microarrays.html**
Another comprehensive tutorial on microarrays.

Microarray

- The more important application of microarray is to determine the **expression level** of different genes rapidly and simultaneously.
- It can give information about gene expression of 30,000 or more genes in one experiment.
- It has also been used for proteomics and study of biological pathways. Another important application is diagnosis of a disease.

Microarray

- Microarray technology uses nanoscale machining and robotics.
- The data generated by microarrays have to be analyzed using mathematical techniques, pattern recognition, supervised learning, signal processing etc .
- The development microarrays has been an interdisciplinary project.
- The device is simply an array of small circular spots on a silicon surface. The spots contain “probes” which may be a DNA fragment, proteins or small molecules.

Two Basic Techniques (1)

- **Sequencing DNA using microarrays**
  - The microarray is loaded with all “probes” (possible DNA sequences) of length *l* in its array of spots.
  - Generate a hybridization solution containing many copies of fluorescently labeled DNA target fragment. The DNA fragment hybridizes with those probes that are complementary to substrings of length *l* of the fragment.
  - Detect probes that hybridize with the DNA fragment. This is usually done by using lasers.
  - Apply a combinatorial algorithm like shortest superstring problem to reconstruct the DNA sequence.

Example

***l*=4. The array looks like**

	AA	AT	AG	AC	TA	TT	TG	TC	GA	GT	GG	GC	CA	CT	CG	CC
AA																
AT		**atag**														
AG																
AC													**acgc**			
TA												**tagg**				
TT																
TG																
TC																
GA																
GT																
GG																**ggca**
GC		**gcaa**														
CA		**caaa**														
CT																
CG																
CC																

Target DNA TATCCGTTT (complement of ATAGGCAAA)

```
A T A G G C A A A
A T A G
  T A G G
    A G G C
      G G C A
        G C A A
          C A A A
```

- The array provides only information about the *-l--mers or l-grams* present but does not provide any clue where these *l-mers* are present in a sequence.
- Obtaining the above alignment is done by a **shortest superstring algorithm**.
- The shortest superstring problem is NP-complete but there is a greedy algorithm that is at most four times optimal.

Gene Expression Determination using DNA Micropchip

- Although each cell of human body contains the same genetic material, not all cells produce the same proteins.
- For example the proteins synthesized by a muscle cells are quite different from those that create growth of hair.
- If a gene is active in a cell, it is **expressed**. By studying the expression levels of different genes in a cell the biologists can understand the function of a cell.
- Microarrays can determine the expression levels of thousands of genes simultaneously. The set up is as follows.

Determining Gene Expression using Microarray

- The micorarry spots are filled up with functional DNA.
- Each spot may contain an oligonucleotide or a cDNA fragment which is characteristic of an mRNA.
- Then, mRNA from a cell or cell population are labeled with fluorescent tags and allowed to hybridize with the cDNAs.
- If the cDNA is complimentary to a substring of the mRNA, the mRNA will hybridize to that spot.

Determine the Expression Using Microarray

- A strong intensity of fluorescent light indicates that a high level of mRNA hybridizes to that spot and therefore the gene which is characterized by that spot is very active in that cell or cell population.
- Conversely, if the spot is dark it means that the gene for that spot is inactive and a moderate amount of intensity means that the gene is somehow active.
- The level of intensity is measured by a laser beam and the intensity levels are pre-calibrated for the activity level.
- For further details, visit the website mentioned earlier.

Cancer Detection Using Microchip

- Cancers that are caused by mutations of genes (such as BRCA1 and BRCA2 which are responsible of 60% of breast and ovarian cancer) can be identified by a microchip.
- Since a large gene has several possible places where mutations may occur to cause diseases other than cancer, it is a difficult task to pinpoint exactly which mutation is responsible for cancer by trial and elimination.

- First, the DNA chip is filled with synthetic single stranded DNA sequences that are found in a targeted gene such as BRCA1.
- To determine whether an individual has a specific BRCA1 cancer mutation, a blood sample is collected from the patient and also from a normal person that does not have any of BRCA1 and BRCA2 mutations.
- Single stranded DNAs are derived from both of these samples, cut into pieces of appropriate size and labeled with colored dyes, the patient's DNA is dyed green and the normal person's DNA is dyed red.
- The leveled sets of DNA are then mixed together and allowed to hybridize with the array.

- If the patient does not have any mutation on the BRCA1 gene (it is a normal gene) then both green and red DNA fragments will hybridize in equal proportion to the spots in the microarray.
- If the patient has cancer, the normal DNA (red) will still hybridize but the green DNA will not hybridize in locations of spots where fragments contain the regions of mutations.
- The biologist can then further investigate these regions of the gene of the patient for further analysis of risk factors.

- Microarrays have been used for assessing risk factors for diseases like cancer, heart disease and diabetics.
- Further use of DNA molecules includes protein microarays, study of biological pathways, and macromolecular interactions.