COS 597c: Topics in Computational Molecular Biology

Lecture 19a: December 1, 1999
Lecturer: Robert Phillips
Scribe: Robert Osada

DNA arrays

Before exploring the details of DNA chips, let's take a step back and look at the larger significance of this technology. DNA chips are transforming the way scientists think of cells. Because of this technology, scientists can now view a cell as an array of expressed genes. This is very different from the way cells were previously viewed and this incredible amount of detail – at the very gene level – gives much more information about the internal processes of a cell than was ever previously possible.

From this new point of view, a patient can be thought of as a collection of genetic mutations. Our knowledge of specific mutation can help in a prognosis, and in further drug selection and design. More importantly, we can think of disease as a result of inappropriate gene expression. In this way, DNA chips can be instructive in detecting diseases such as cancer and finding drugs targets for treatment.

Background

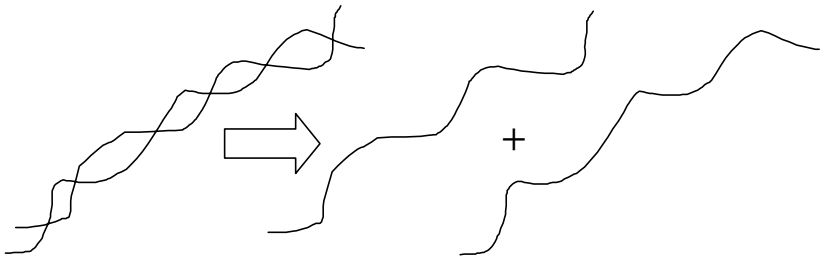
We will start by first going into a short background of gene expression. Cells have a genome, which contains genes (made up of DNA). For example, the human genome consists of 3 billion base pairs of DNA, and is thought to contain 75,000-100,000 genes. From the DNA, expressed genes are transcribed into messages (RNA). It is important to realize that although each cell contains a copy of the entire genome, in any given cell, at any given time point, not all genes are necessarily expressed.

Experimentally, the RNA messages can be isolated from a cell and reversed-transcribed back into DNA, called cDNA. A collection of cDNAs from cellular RNA makes up a cDNA library of the cell, and two different cDNA libraries can be compared using a technique called differential hybridization.

There are many possible sources of cDNA libraries. Two different cell populations, such as a normal cell and cancer cell, can be used as probes – another term for cDNA libraries. Such a comparison could yield valuable insight into the differences of the two populations. Another option is to look at the same cell as it develops, in this way seeing which genes are active over time during the different stages of the cell. This can give clues to the gene function within a cell. Finally, cDNA libraries can be made of a cell as it responds to therapy, which can be used to measure the effectiveness of a drug.

cDNA hybridization

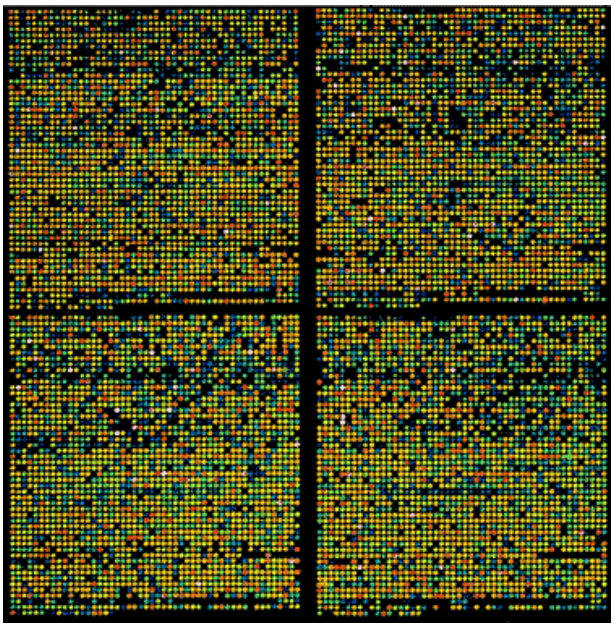
The mechanism behind hybridization is the basis of DNA chips. DNA molecules are double-stranded, but these two strands can melt apart at a characteristic melting temperature, Tm, which is usually above 65°C (see diagram below). As the temperature is reduced and held below the melting temperature, single-stranded molecules hybridize to their counterparts. In the same way, a RNA molecule (single-stranded, part of probe) can hybridize to a melted cDNA molecule.



Hybridization protocol

DNA chips are two-dimensional arrays of prepared cDNA molecules. To use such an array, we first take a pool of cDNA molecules and transcribe them into RNA molecules. These RNA molecules are then labeled with a fluorescent dye (fluorochrome), a different color for each of the two different pools we wish to compare. The two colors currently used to label molecules are red and green. Because we use two different colored dyes, we can mix the two different populations and compare their genetic expression at the same time. This is important, because we want to have nearly the same environment for both populations throughout the measuring process.

The resulting labeled RNA probe is put in contact with the array of targets on the DNA chip in a thin liquid layer for several hours at Tm, 65°C. Finally, when the labeled RNA molecules have had time to hybridize, the chip is washed in a solution several times and then scanned. The scanning process uses a laser and a computer to measure the level of excitation of the fluorochromes on the DNA chip target array. The laser can detect both wavelength and intensity of the excitation, and produces an image where each pixel corresponds to a single cDNA molecule (see below). The color of the pixel indicates the level of hybridization intensity ratio between the two probes. For example, a red color indicates higher hybridization of probe 1, while green indicates a higher hybridization of probe 2.



Because of the measuring technique used, gene expression has a magnitude. In addition, there is a threshold to response before a gene actually registers as being expressed.

Gene chip manufacturers

There are two main types of DNA chips: oligonucleotide arrays and cDNA arrays. Oligonucleotide arrays are produced by a process similar to the production of computer chips – layers of photolithographically controlled solid-phase chemicals are placed on the chip surface, resulting in the synthesis of 25-nucleotide oligomers on the chip. A typical chip can contain ~65,000 wells. This chip is very expensive and difficult to product, but offers extreme sensitivity in studying single base mutations. This type of chip is manufactured by companies such as Affymetrix and Synteni.

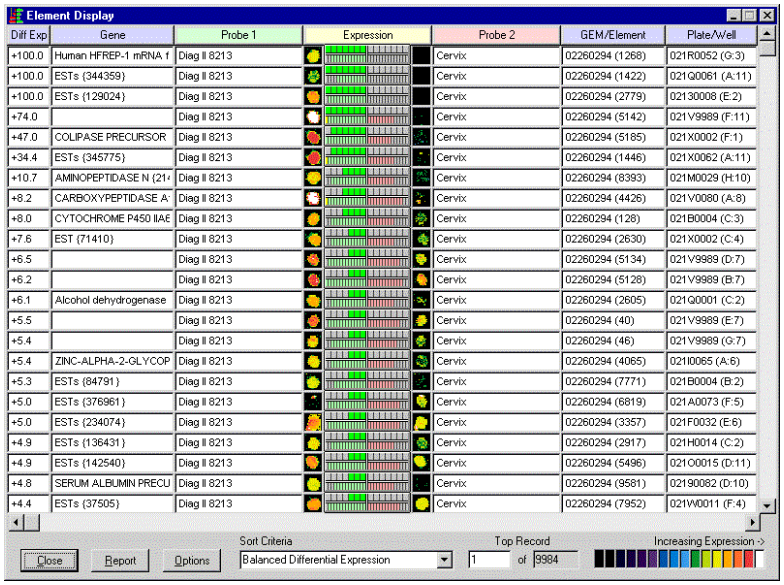
cDNA arrays are produced by depositing and drying microdrops of DNA containing liquid onto treated glass. Once the DNA liquid is assembled, the manufacturing process of these arrays is very inexpensive and quick. In addition, the resulting arrays are very dense, with current technology allowing approximately 10,000 genes per chip on a printed surface less than 2 cm^2. Manufactures of these types of chips include Genome Systems and Research Genetics, and claim the ability to detect 1 gene in a pool of 100,000 – although real results depend on the signal to noise ratio of the data and the statistical analysis software used.

cDNA are quickly gaining popularity because they are very inexpensive to produce and allow the simultaneous measurement of many genes.

Analysis software

The result of using a cDNA chip is the gene expression magnitudes of two probes in approximately 10,000 genes. Interpreting the meaning and significance of 20,000 such numbers is a very big challenge, which becomes even more overwhelming because such a measurement is usually performed several times. The task of analyzing this amount of data is made easier by software packages. These packages are usually sold by the same company that manufactures the DNA chips, and are typically quite expensive.

The following is a display of a characteristic DNA chip software package. From left to right, the columns are the hybridization intensity ratio, the gene name or cDNA reference, probe 1 identity, graphic representation of expression, probe 2 identity, array address, and location of the cDNA in the freezer.



Inferring Functions for Unknown Genes

Many genes listed do not have a known name or function, but seeing the variation of these genes in different populations is still interesting and may yield clues to the function of the gene. For example, if the expression levels of a number of genes are correlated, then this gives evidence that these genes have similar functions. If one of these genes has a known function, then this suggests the function of the other genes. Thus, clustering gene expression

data is a useful method for analyzing array data. DNA chip experiment results need a great deal of data analysis, and there is some software available publicly at www.genome.stanford.

Detecting Single Nucleotide Polymorphisms

Single Nucleotide Polymorphisms (SNPs) are commonly found mutations in human genes. Thousands of oligonucleotides representing known SNPs can be screened at once on a single gene chip, which allows for the identification of SNPs responsible for a given genetic disease. For example, after measuring the gene expression of 100 normal patients relative to the gene expression of 100 multiple-myeloma patients, we can correlate each SNP with the disease. This way we can isolate disease-associated SNPs, which will allow for diagnosis and treatment of future patients. DNA chips have given us the ability to look at SNP levels, which is akin to being able to look at the very beginning of a tumor, even before it actually begins.