# DNA chips in Bioinformatics

Ana Madevska-Bogdanova

Institute of Informatics
Faculty of Natural Sciences and Mathematics
ana@ii.edu.mk

**Abstract.** DNA chip technology is an improved and highly scaled-up version of a 25 year old method to reveal very small changes in several hundreds or even thousands of genes in **one** step rather than searching one gene at a time. This technique can also estimate the activity of many genes. This way one can see what genes are doing inside the cells, or seeing genes in action. The outcome of this method is can be used in pattern recognition problems in bioinformatics.

## 1 Introduction

Tthousands of genes and their products (i.e., RNA and proteins) in a given living organism **function** in a complicated and orchestrated way. Traditional methods in molecular biology are "one gene in one experiment" basis, and the outcome is very limited, so the "whole picture" of gene function is hard to obtain.

In the past decade, a new technology has raisin, called **DNA microarray**, and it caused tremendous interests among biologists.

This technology promises to monitor the whole genome on a single chip. It allows us to take a 'photograph' of genes and catch them in action.

In the literature there are items that describe this technology, but they are not limited to:

- biochip,
- DNA chip,
- DNA microarray,
- gene array.

The main principle is the base-pairing (i.e., A-T and G-C for DNA; A-U and G-C for RNA).

Hhybridisation is also the underlining principle of DNA microarray, that is the NA / RNA hybridisation is used in this technology.

Array is an orderly arrangement of samples. It provides a medium for matching known and unknown DNA samples based on base-pairing rules and automating the process of identifying the unknowns. Microarrays however are the sample spot sizes, typically less than 200 microns in diameter. They usually contain thousands of spots and require specialized robotics and imaging equipment.

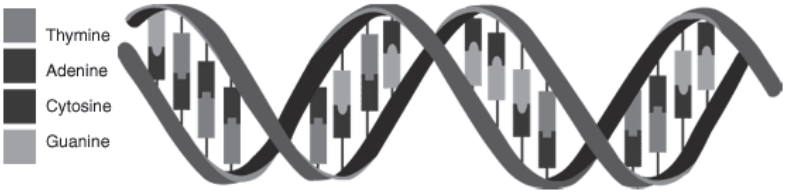


**Figure 1**. DNA double helix

DNA microarray, or DNA chips are fabricated by high-speed robotics, they are generally on glass. Probes with known identity are used to determine complementary binding.

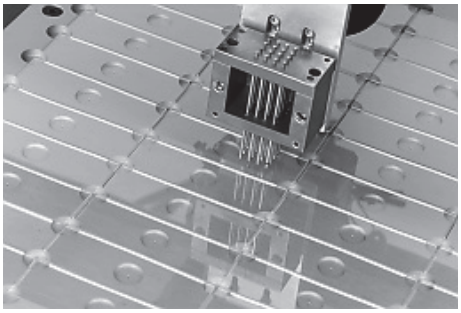

**Figure 2**. DNA chip

There is a slight difference between the terms: "*probes*" and "*targets*". The "probe" is the nucleic acid with known sequence and "target" - the free nucleic acid sample whose identity/abundance is being detected.

"DNA microarray(s)" and "DNA chip(s)" are used interchangeably. One should be aware of the technical difference. They share the same scientific principle. They may be used with similar TARGETS, but they differ in construction and type of the probe.

## 2 Design of a DNA Microarray System

There are several steps in the design and implementation of a DNA microarray experiment:

1. Choosing Cell Populations
2. mRNA Extraction and Reverse Transcription
3. Fluorescent Labelling of cDNA's
4. Hybridization to a DNA Microarray
5. Scanning the Hybridized Array
6. Interpreting the Scanned Image

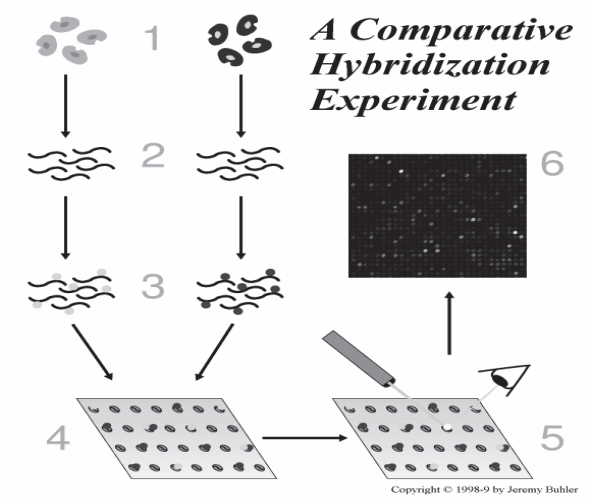


**Figure 3**. The six steps of building DNA microarray System

DNA microarrays are perfectly suited for comparing **gene expression** in different populations of cells.

### 2.1 Demonstration-how DNA microarray are performed

One can use DNA chips to compare what happens to yeast genes when they are grown in aerobic ($O_2$) vs. anaerobic conditions (without $O_2$)

The task is to determine which **genes are activated** and which **genes are repressed when** two populations of cells are compared.

Step 1: Choosing Cell Populations. There are two tubes where yeast cells are grown, one tube is open (with $O_2$) and the other one is closed, so there is no oxygen ($O_2$). The cells are grown and adjusted which genes are needed to be activated or repressed in order to survive.

Step 2: Second step is the isolation of the mRNA of the two tubes (they are spinned). After that, the cells are in pellets. When the liquid is removed, the mRNA is extracted from the cells.

Step 3: The fluorescent labelling of cDNA's. In the process, fluorescent cDNA (complementary DNA) is bounded to the mRNA. cDNA is representing the invisible mRNA in the way that green cDNA is for mRNA of the $O_2$ genes and red cDNA - mRNA of anaerobe genes.

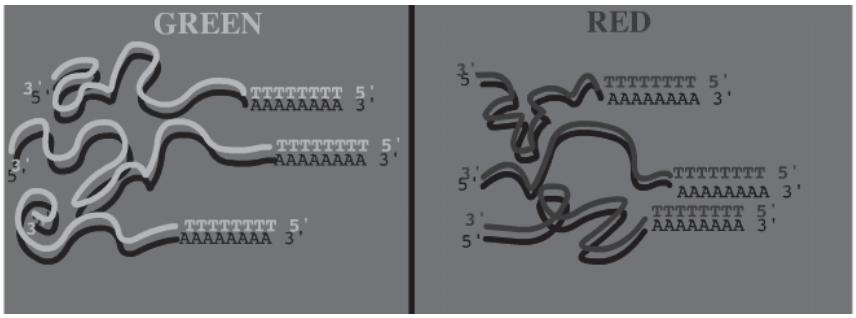


**Figure 4**. The red and green cDNA are combined

Step 4: the hybridization to a DNA Microarray. Some of the labelled cDNA has bound to the DNA chip spots. Then, the unbound cDNA is washed off and it can be seen what has bounded to the microarray. Nothing is visible at the DNA chip yet. The microscope slide containing the microarray is placed into a darkened box. Inside the box, it is scanned with green and then with the red laser to detect the **bounded** cDNA. At the end, the retrieving and merging is done.

Step 5: scanning the hybrized array. The green and the red image are retrieved and the merged image is produced.

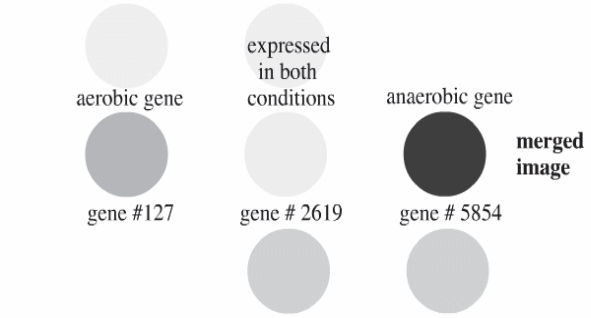


**Figure 5**. Example of 3 genes and their expression level

The product is a coloured DNA chip spot, i.e. the variations of green, red and yellow colours, and the ddifferences in the intensity represent the differences in gene expression level! This analysis is to be conducted on all 6000 genes at the same time.

Step 6: Image is processed and data analyzed with the help of bioinformatics.

The next table represents the different gene expression levels in one experiment, represented with numbers. This table is only a part of the whole DNA chip. The sign '-' means that the corresponding gene was repressed.

**Table 1**. Different gene expression levels in one experiment

| | | |
|---|---|---|
| **0,685** | **-0,040** | **0,046** |
| **0,084** | **-0,840** | **-0,641** |
| **-0,091** | **-0,673** | **0,456** |
| **-0,167** | **0,611** | **0,370** |
| **-0,917** | **1934533,000** | **10245,000** |
| **-0,169** | **-522628,000** | **19952,000** |
| **131,000** | **-0,807** | **-2763,000** |
| **0,002** | **-0,524** | **-0,790** |
| **-0,365** | **-1324013,0** | **-0,922** |
| **-0,175** | **0,550** | **0,789** |
| **0,861** | **-1383755,0** | **-13931,00** |
| **-0,573** | **-168291,000** | **-0,186** |
| **-0,466** | **-0,138** | **-0,022** |
| **0,398** | **0,974** | **0,322** |
| **-189,00** | **1069293,000** | **0,275** |

The analysis of the DNA chip outcomes can bring all sorts of new knowledge. Bionformatics plays central role in processing this data. Even the methods of pattern recognition can be used.

## 3 DNA chip products used in pattern recognition problem in Bioinformatics

The main goal is to identify unknown genes of similar function from gene expression data. Similar gene expression level means similar functionality of the genes.

The input data are obtained in different experiments with yeast, like high temperature, very low temperature, pesticides etc. The number of used DNA chips is equal to the number of experiments, i.e. each DNA chip provides data for one of the experiments.

Each input object in this pattern recognition problem is a gene from the yeast genome. The dimensionality is the number of experiments and the values are the **level of the gene expression** in each experiment (obtained by the used DNA chips). This data is used as a training set to create a SVM machine for the given task. SVM's are dealing greatly with this kind of problem.

## 4 Some Applications of DNA Microarray Technology

DNA Microarray technology has many other applications, such as:

- Identifying drug targets and validation of new drugs
- Gene expression in pathogens
- Viral gene expression during latency and infection
- Population genetics: study of species diversity
- Homogeneous vs. heterogeneous diseases
- Prognosis and preventive measures

## 5 Conclusion

The DNA chip technology provides huge amount of data from different experiments. Fluorescence imaging coupled with robust bioinformatics analysis provides raw data on gene expression profiling, genotyping for polygenic traits, infectious and other disease patient management (HIV, TB) and also for different pattern recognition problems.

## References


1. P. Baldi and S. Brunak, *Bioinformatics, the Machine Learning Approach*, MIT Press, 1998
2. Troyanskaya OG, Garber ME, Brown PO, Botstein D: Nonparametric methods for identifying differentially expressed genes in microarray data. *Bioinformatics*. 2002 Nov; **18**(11): 1454-61.
3. M. P. S. Brown, W. Grundy , D.Lin, N.Cristianini, C.W. Sugnet, T.S. Furey, M.Ares Jr., and D.Haussler , Knowledge-based analysis of microarray gene expression data by using support vector machines, Proceedings of the National Academy of Sciences. Vol. 97, Issue 1, 262-267, January 4, 2000