

BMg works with DNA Chips

A.T. Navarro Muñoz M. Esparza Manzano S. Götz
alnavar@itaca.upv.es miesman@upvnet.upv.es stefang@fis.upv.es

J.M. García Gómez M. Robles Viejo
juanmig@die.upv.es mrobles@die.upv.es

*Biomedical Computer Science
ITACA
Universidad Politécnica de Valencia*

Abstract

From the co partnership in European eTumour Project, the Biomedical Mining Group (BMg) of the BET-IBM got deeply into DNA chips data analysis. Likewise, the greeting collaboration with IVIA has been important too.

Brief overview of DNA microarrays principles and a review of our work in this area is made.

1. Introduction

In specific experiment, DNA-chips or microarrays allow us to discover genetic cells behaviour by gene expression in the samples. In the late 80's a scientific group (Stephen Fodor, Michael Pirrung, Leighton Read and Lubert Stryer) settled the bases for the creation of the first microarray platform, named GeneChip®, manufactured by Affymetrix Corporation. The new technology would allow to determinate and quantify the genetic contents within a sample, by the possibility to analyze multiple genes at the same time. [1]

Nowadays, microarray technology and studies can be found in all disciplines related to molecular biology.

BMg works in the high-level analysis of microarray data applied to health research. At present, we participate in two european projects (eTumour and HealthAgents) that try to caraterize brain tumours using combined analysis of microarrays and magnetic resonance spectra. Simultaneously, BMg works on functional gene annotation methods in order to help the interpretation of the results in experiments.

2. Applications

The use of biochips has been extended to several areas in science. For example, they allow the agronomics engineers to analyze and study agricultural species and new stumps and learn about the biologic plant processes. In animal studies biochips are used to understand relationships between species. But its greater application can be found in health care related studies like genetic expression analysis, mutation and polymorphism detection, personalization medicine, pharmacogenomic products, molecular diagnostic or illness prognosis among others.

Wherever we can find genetic material, we can to use biochips for its analysis.

3. Microarrays manufacturing

A microarray is based on genetic material arranged on solid slide which is distributed in a two-dimensional array of spots or cells. This genetic material could be either short sequences, called oligonucleotids, or longer sequences like cDNA (complementary DNA) or PCR products (in vitro replication of DNA sequences from polymerasa reaction). Simplifying, we can say that those kinds of genetic material are RNA sequences contained in a biochip. [7]

All biochips experiments go through the same process of planning, experimentation and analysis as shown in Figure 4.

In the first place, we must determine what is the goal of the experiment and collect the samples that will be used.

After, depending on the genes we want to study, the biochip is designed and manufactured with the RNA probe sequences appropriated.

From samples, their complementary RNA is extracted and labeled. This RNA sequences are called probe targets.

The probe targets are stored in biochip, generating the hybridization process. In this process, several probe targets are joined with its complementary RNA in the biochip, allowing the DNA identification and quantification contained in each sample.

In continuation, washed, stained and scanned processes are made to obtain a microarray image of the spots/cells.

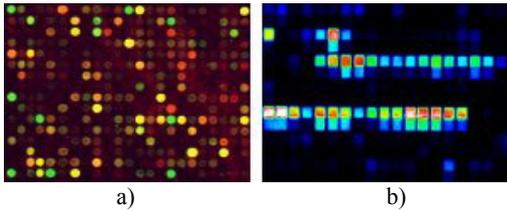


Figure 1. Biochips scanned image.
a) Agilent, b) Affymetrix

The main manufacturers of microarrays are Affymetrix and Agilent Technologies.

Microarrays build by Affymetrix are based on synthesis in situ. This means that probe sequences are synthesized directly on slide using photolithographic techniques for making oligonucleotide sequences.

The GeneChip technology, invented in the late 80's, was the basis of creation Affymetrix company, as division of Affymax in 1991. Affymetrix began operating independently in 1992. The world's first microarray prototype built using microscope slide was in 1989. [1]



Figure 2. Affymetrix GeneChip®

Agilent Technologies in collaboration with Incyte Genomics manufacture cDNA biochips, where the probe sequences are clone of cDNA fragments or completely gene sequence placed on slide, distributed in spots on slide.

Agilent is a spin-off of Hewlett-Packard Company set up in 1999. Agilent Technologies doesn't work only with biochips but also provides core electronic and bioanalytical measurement tools to advance in the electronics, communications, life science research, environmental and petrochemical industries. [2]



Figure 3. Agilent biochip

4. Microarray data analysis

When a microarray image is obtained, a digital process is made that consist in an automatic segmentation to obtain a single intensity value of each cell.

At this point, the necessary data has been obtained to start the analysis process.

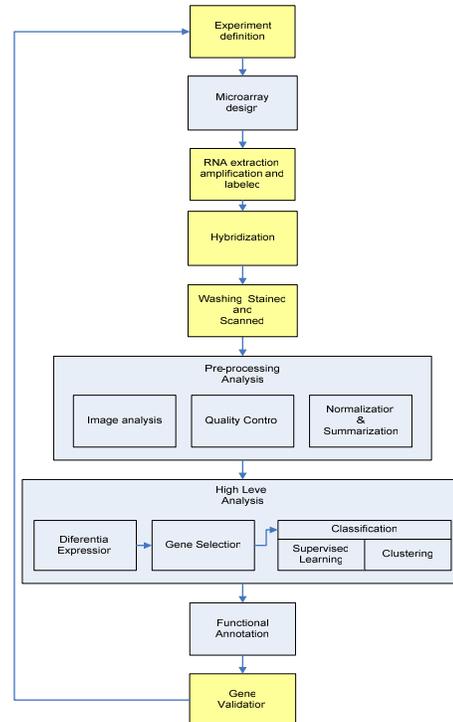


Figure 4. Experiment design and analysis with microarrays

For a right use of the intensity data it is necessary to apply a pre-process to correct and decrease systematic and accidental errors produced in the manufacture process. Therefore, the microarray is submitted to a quality control process using control probe sequences introduced in biochip design step. Moreover, the microarray image is visualized using intensity data to detect scratches and possible contamination. [10]

Finally, the DNA replications and gene transcriptions distributed by the biochip are fitted in a single gene expression value of each sample using statistical techniques.

4.1. High-level Analysis

In the moment that data is preprocessed, we can to resolve of the biologic problem, the final goal of the experiment.

One of the main objectives in microarrays analysis is to determine the genes expressed in each sample. The differential expression analysis discriminate the genes between samples, allowing to distinguish the representative genes from one sample against the genes of the other sample. In consequence, we obtain a list of representative genes for the experiment. [3][4]

Later, the classification process allow to establish rules of grouping between the genes selected, using supervised learning techniques or clustering patterns.

Finally, the interpretation of the results is very important to validate of the experiment design.

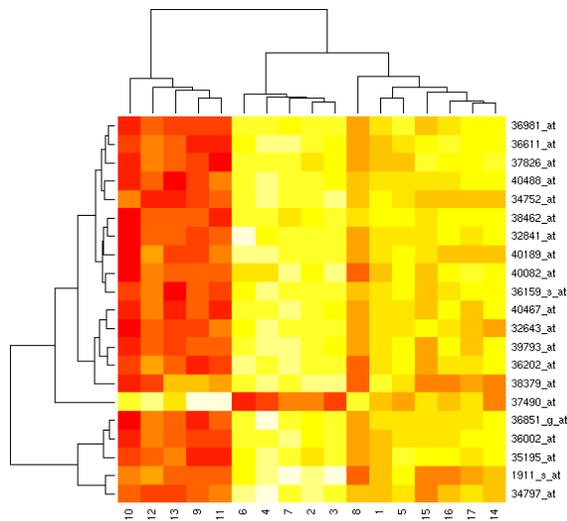


Figure 5. Clustering of a microarray experiment

5. Data bases and gene ontologies

Many biochip experiments around the world are made. All information extracted is used by the specialists community in their studies. The data bases are repositories of genetic expression data from transcriptomic projects. These databases compile the raw data of the experiments, and in other cases the normalized and images files are included. There exists specific databases associated to determined gene expression platform, others are more generals and compile data from several sources.

Some important gene data bases on web are:

- ArrayExpress is the public repository for microarray data, which has the objective to store information with consistent annotation, in agree to the MGED recommendations. The MGED (Microarray Gene Expression Data Society) have developed the MIAMI system (Minimal Information about a Microarray), this is a group of rules about the minimal annotation that must satisfy every experiment of microarrays.

- Standford Microarray Database (SMD) store all kind of data, like raw data and normalized data. Moreover, SMD provide interfaces to search, analysis and visualization of expression data.

- NASCArrays is a database from Affymetrix generated by the NASC (Nottingham Arabidopsis Stock Center) transcriptomic service.

Nowadays, the data information has experimented a standarization with the objective to obtain functional gene annotation. This standarization is made by the gene ontology consorcium.

The Gene Ontology (GO) project is a collaborative effort to address the need for consistent descriptions of gene products in different databases. The GO collaborators are developing three structured, controlled vocabularies (ontologies) that describe gene products in a species-independent manner. The current ontologies of the GO project are molecular function, biological process, and cellular component. [5]

6. The software resources

Since microarray data analysis like image treatment, statistics and classifier development treats with big amounts of data, like gene expression data and functional annotation, new computational methods and tools were developed to support biochips technologies.

Some tools and projects supporting the study of microarray experiments are:

- MicroArray Suite by Affymetrix: It drives system operation, instrument control and data analysis for the entire GeneChip platform.
- Stratagen Company: Offers systems biology-focused pathway analysis, microarray data analysis and enterprise-level data management solutions. One of his highlighted products is the GeneTraffic software.
- Bioconductor: An open source and open development software project for the analysis and comprehension of genomic data.[3]
- Blast2GO: An universal Gene Ontology annotation, visualization and analysis tool for functional genomics research.[6]

7. Acknowledge

eTumour Project: an European project to develop a web accessible molecular resonance (MR) decision support system for brain tumour diagnosis and prognosis.

HealthAgents Project: Agent-based architecture distributed decision support system for brain tumour diagnosis and prognosis.

IVIA: Valencian Institute of Agrarian Investigations

8: References

- [1] Affymetrix Corporation technical documentation, <http://www.affymetrix.com/support/technical/index.affx>
- [2] Agilent Corporation technical support, <http://www.agilent.com>
- [3] R.Gentleman, V.Carey, D.Bates, B.Bolstad et al. Bioconductor: Open software development for computational biology and bioinformatics. *Bioconductor Project Working Papers*. Working Paper 1. 2004
- [4] A.Heydebreck, W.Huber and R.Gentleman, Differential Expression with the Bioconductor Project. *Bioconductor Project Working Papers*. Working Paper 7. 2004. <http://www.bepress.com/bioconductor/paper7>
- [5] The Gene Ontology Project. <http://www.geneontology.org>
- [6] A.Conesa, S.Götz, J.M.García, J.Terol, M.Talón, M.Robles, Blast2GO, *Bioinformatics* 21, 3674-3676, 2005. <http://www.blast2go.de>
- [7] M.López, P. Mallorquín, M.Vega, *Microarrays y Biochips de ADN. Informe de Vigilancia Tecnológica.*, Genoma España, 2002.
- [8] A.T.Navarro. *Affymetrix GeneChip Data Analysis. Technical Report*. BCS. ITACA, 2006.

[9] J.M.García. *Bioinformática. Minería de datos y búsqueda de patrones biológicos. Technical Report*. Física Aplicada. UPV, 2006.

[10] P.Baldi, G.Wesley Hatfield. *DNA Microarrays and gene expression. From experiments to data analysis and modeling*. Cambridge. 2002

[11] T.Pasanen, J.Saarela, T.Toivanen, M.Tolvanen, M.Vihinen, G.Wong. *DNA Microarrays Data Analysis*, CSC.Scientific Computing Ltd., 2003.

[12] S.Dudoit, J.Fridlyand, and T.P. Speed. comparison of discrimination methods for the classification of tumors using gene expression data. *Journal of the American Statistical Association*, 97(457):77-87, 2002.