MEDWELL PUBLISHING
Research Journal of Biological Sciences 4 (1): 77-81, 2009
ISSN: 1815-8846

Exploiting of DNA Microarray Technologies in Genome Analysis

[1]Hasan Koyun, [2]Seyrani Koncagul and [1]Abdullah Yeşilova
[1]Department of Animal Sciences, Faculty of Agricultural, University of Yüzüncü Yil,
Biometry and Genetics Unit, 65080, Van, Turkey
[2]Department of Animal Sciences, Faculty of Agricultural Faculty, Harran University,
Animal Breeding and Genetics Unit, 63100, Ş. Urfa, Turkey

Abstract: A new era has been started by discovering DNA microarray technologies in fields of genetics. In this mini review article, the basic molecular methods and statistical analyses for DNA microarrays (DNA chips) technologies have been reviewed. Types of commonly used microarrays, the target labeling for detecting and quantifying of gene expression levels and approaches of DNA microarrays were described. Moreover, basic image and statistical analyses of microarray data such as cluster analysis and a general linear model (GLM) were also recited and summarized.

Key words: DNA microarrays (chips), target labeling, cluster analysis, General Linear Model (GLM)

INTRODUCTION

In a recent decade entire genome sequencing of many organisms such as human, mouse, farm animals (pig, sheep, bovine and poultry) genomes have created the need for high throughput analysis of gene expression patterns. As a result, a new era has been started by discovering DNA microarray technologies in the fields of genetics.

DNA microarrays or so called DNA (gene) chips are a new technology recently used in Human Genome Project (HGP) as well as in other genomes to estimate mRNA or gene expression levels in specific cells or tissue samples for many genes at once. They also allow one can to monitor the occurrence of polymorphisms in genomic DNA. In other words, the basic assumption of the DNA microarray is that RNA samples or targets are hybridized to known cDNAs or oligo probes on the arrays (Gerhold *et al*., 1999; www.microarraystation.com).

DNA array technologies: In this revolutionary technology to functionally analyze genomes, single stranded DNAs (ssDNA) for the genes under investigation are immobilized and arranged in a regular grid-like pattern therefore one can investigate a gene or genes in question with thousands of different DNA sequences immobilized at different positions on the surface. DNA arrays are usually made up with a nylon membrane, a glass or plastic slide and a quartz wafer (Balding *et al*., 2003; Draghici, 2003).

Once mRNA is extracted from cells of interest, labeled and hybridized to the array, genes showing different levels of mRNA can be detected. Nevertheless, in this molecular process target labeling by hybridizing with probes plays a certain role in order for the ones to detect expression levels of genes of interest or polymorphisms in genomic DNA.

Labeling: Labeling is done for image visualization and quantification purposes. In microarray technologies target is mainly labeled with radioactively and fluorescently. Radioactive markers, such as P^{33} are utilized in radioactive labeling and image can be captured by a photo sensitive device (e.g., radioautograph). On the other hand, fluorescent dyes, such as phycoerythrin or water-soluable Cyanine family (Cy3 and Cy5) are often used in fluorescent labeling.

Figure 1 illustrates a brief explanation of fluorescent labeling. Once mRNA is extracted from two different types of cells (Cell A and B) mRNA of cell A is labeled Cy3 (in green color) and mRNA of cell B is labeled Cy5 (in red color). Next, mRNA is hybridized to arrays. Eventually color reading is performed based on hybridization. Each spot on the array corresponds to a different gene. In other words, the color of each spot on the array corresponds to the relative concentration of mRNA for a particular gene in two different cell types. Spots in green color on the array indicate that mRNA mainly comes f rom cell type A.

Corresponding Author: Hasan Koyun, Department of Animal Sciences, Faculty of Agricultural, Yüzüncü Yil University, Biometry and Genetics Unit, 65080 Van, Turkey

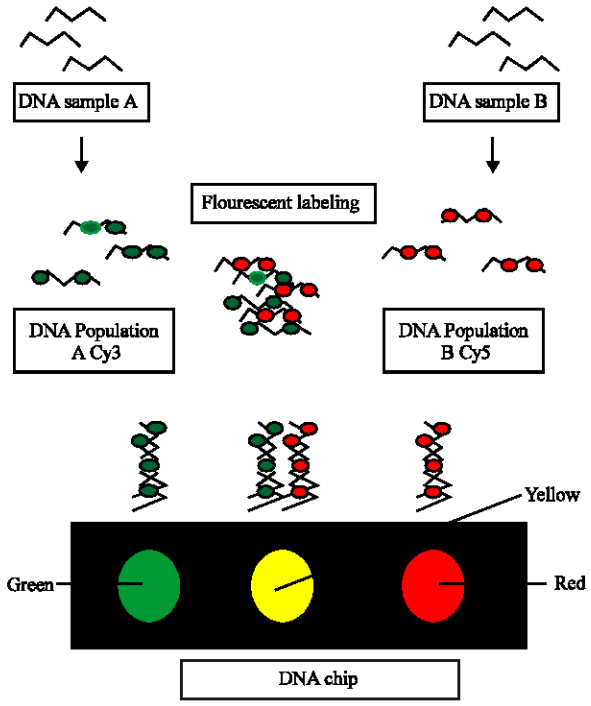

Fig. 1: Fluorescent labeling in DNA microarray technologies

Similarly, spots in red color imply that mRNA mainly comes from cell type B. Spots in yellow or yellow-off color, rather than clear green and red colors point that mRNA is equally mixed from both cells.

In fluorescent labeling, image detection can be performed based on signal intensity at the matching probe sites using two samples labeled with different fluorescent dyes therefore detecting at two corresponding wavelengths. Fluorescent pattern is captured by laser. The intensity images are then scanned by a special detector at a high resolution in order to display each probe spot by many pixels (Shi, 2002; Balding *et al*., 2003).

Types of microarrays

Synthetic oligonucleotide arrays: They have also been known as fabricated arrays or *in situ* synthesis. This method includes a direct photochemical synthesizing of oligonucleotide sequences (20-80 mer oligos) with probes designed based on genomic sequence information on slide, chip or membrane using photolithography during array fabrication. These kinds of arrays do not require any spotting, cloning and PCR processes (Lipshutz *et al*., 1999; Shi, 2002; Draghici, 2003).

cDNA arrays: Unlike synthetic oligonucleotide arrays or *in situ* synthesis, in this technique, also known as spotted cDNA arrays, DNA is prepared away from the chip and small sequences (0.5-5.0 kb) are used as a cDNA probe. Basically, PCR products from clones of genes of interest are spotted on a glass or a plastic slide using a robot dipping thin pins into the solutions containing the desired DNA material. Extracted cellular mRNA is reverse-transcribed into cDNAs for hybridization (Shi, 2002; Draghici, 2003).

There are no basic advantages when to compare both techniques. On the other hand, due to the fact that *in situ* synthesis introduces less noise in DNA system therefore, requiring less molecular genetics and analysis work in addition to being cheaper than those of cDNA arrays (Shi, 2002; Draghici, 2003).

Applications to DNA microarrays: There are two main application forms for DNA technologies, which are to identify genes and gene mutations and to determine expression levels of genes (Shi, 2002). Moreover, DNA chips can be used as Variant Detector Arrays (VDAs) to look for DNA sequences that differ by single nucleotide polymorphisms (SNPs) (Carr, 2007). SNP chips are, a type of microarray, designed with an array of user defined oligonucleotides attached to the substrate to identify genetic variants or differences in an individual (Shi, 2002; McCann *et al*., 2007). DNA microarray technologies allow SNPs to serve as genetic markers frequently abunded along with chromosomes. Hundreds of different alleles at different loci of SNPs can be tested and screened simultaneously in the same experiment. Consequently, one of the main advantages of the SNP chip (DNA microarrays) over traditional genotyping methods (microsatellites) requires smaller DNA samples and is much faster (Gunning and Behlke, 2007; Wallace *et al*., 2007; http://www.affymetrix.com). Currently, SNP chips are commonly used in almost every eukoryatic genome research studies (Lee and Hossner, 2002; Cougburn 2003; Kim *et al*., 2003; Hu *et al*., 2005; Jander and Barth, 2007; Lehert *et al*., 2007; Walia *et al*., 2007; www.imgenex.com/tissue_micro_arrays.php) for the purposes of more genetic discoveries to improve plant and livestock productions as well as using model organisms such as mice and pigs for studying complex human diseases like diabetes, obesity and cancer types.

Analysis of microarray data: Microarrays may generate tens of thousands of data points for every experiment performed. A study may be composed of many experiments generating a million data points (Gaasterland and Bekiranov, 2000). Microarrays analysis involves identifying all the genes that are turned on in a

particular cell and generating a list of those genes. As for analysis in gene expression studies there are some main analytical ways to be considered.

Image processing and normalization: In order to make a clear visual analysis image quantifications are done based on image intensity. Each probe spot is introduced by many pixels and overall intensity value of each probe is estimated with corresponding pixels (Balding *et al*., 2003). Normalization is performed in order to adjust and combine systematic differences across different data sets and eliminate experimental artifacts such as noise occurring during the molecular process of DNA micrroarrays (Draghici, 2003).

Commercial software programs such as Spot (www.csiro.au), Able image analyser (www.mulabs.com), AIDA Array Metrix (Raytest GmbH; www.raytest.de), ArrayFox (ImaxiaCorp.www.imaxia.com) for DNA microarray image analysis are available. Also, software programs for image analysis can freely be downloaded (www.biocompare.com; http://download.chip.eu).

Statistical analysis: In general, statistical analysis involves performing DNA (SNP) chip data analysis for whole genome association studies. Identifying patterns of gene expression and grouping genes into expression classes may provide much greater insight into their biological functions. The main goal fulfilling statistical analysis is to promptly determinate gene expression levels, phenotypic-genotypic associations and predictive biomarkers.

A large group of statistical methods, generally referred to as cluster analysis, have been developed to identify genes that behave similarly across a range of experimental conditions (Li *et al*., 2006).

Clustering and classification: Classification is a way to find genes changing in mRNA expression level predict phenotype. If two genes are functionally related, they may be expressed in the same way. Therefore, it is necessary to detect groups of co-expressed genes (Nakatani *et al*., 2006). In order for the genes to group into clusters, one needs to define some measure of distance between them. The most commonly used one is the geometric, Euclidean distance between the 2 points (expression vectors) i and j, the square of which is defined as (Svrakic *et al*., 2003):

$$D_{i,\,j}{}^2 = \Sigma\,(E_i, k\_E_j, k)^2 \qquad (1)$$

Statistical algorithms can be divided into two classes, depending upon whether they are based on 'similarity' measures or not. Methods based on 'similarity' measures rely on defining a distance (or 'dissimilarity') between gene expression vectors (Li *et al*., 2006). Clustering and classification analysis is also performed using the Pearson correlation coefficient (Li *et al*., 2006) and traditional regression methods (Park *et al*., 2007).

ANOVA models for microarray data: Concerning more systematic analysis of multiple slide experiments, a statistical approach based on the ANOVA technique was proposed by Kerr *et al*. (2000).

Comparisons of multiple samples from a microarray experiment, every measurement in a microarray experiment is associated with a particular combination of an array in the experiment, a dye (red or green), a variety and a gene. To rationalize the multiple sources of variation in a microarray experiment, the General Linear Model is:

$$\text{Log}\,(y_{ijkg}) = \mu + A_i + D_j + V_k + G_g + (AG)_{ig} + (VG)_{kg} + \varepsilon^2_{ijkg} \qquad (2)$$

where:

- y_{ijkg} = Denotes the measurement from the ith array, jth dye, kth variety and gth gene.
- μ = The overall average signal.
- A_i = The effect of the ith array.
- D_j = The effect of the j th dye.
- V_k = The effect of the kth variety.
- G_g = The effect of the gth gene.
- $(AG)_{ig}$ = A combination of array i and gene g (i.e., a particular spot on a particular array.
- $(VG)_{kg}$ = The interaction between the kth variety and the gth gene.

The error terms ε^2_{ijkg} are assumed to be independent and identically distributed with mean 0. The array effects Ai account for differences between arrays averaged over all genes, dyes and varieties.

DISCUSSION

Now a days, DNA microarrays or chips have been becoming widely used for detection of DNA polymorphisms and for gene expression analyses. Commercial and academic efforts have resulted in DNA chips for linkage analysis. Companies such as Affymetrix (GeneChip Mapping arrays; www.affymatrix.com) and Illumina (BeadChips; www.illumina.com/dna) as well as HapMap (www.hapmap.org), PrettyBase (www.math works.com), Pedigree formats (www.omicsoft.com) and academic groups are collecting SNPs and building arrays to allow rapid linkage analyses by DNA chip hybridization for biological (gene expression levels), agricultural

(Quantitative Trait Loci (QTL) mapping for better plant and livestock productions and medical (drug effects and metabolism and clinical diagnosis) purposes (www.ncbi.nlm.nih.gov).

Needless to say that SNP chips will become major tools for genotyping human beings in the next decades until full genome sequencing becomes cheap. In the near future, probably high density microarrays of genomes which come from different organisms will have been developed. Not only will this trigger intense improvements, which are primarily based on the number of SNPs on the chips, SNP selection criteria and genome wide coverage, but also generate and update Bioinformatics (integrating expression data and gene data-bases). Consequently, it will be possible to profoundly study and search for Genome-Wide Associations (GWAS) and make the global analysis of gene expression from different genomes.

Throughout, the global analysis one will be able to identify mutations and the cellular function of genes in question, the nature and regulation of biochemical pathways and the regulatory mechanisms at play during certain signaling conditions or diseases as well as detecting underlying functional QTL that of course include QTN (Quantitative Trait Nucloetides) thus obtaining improved agricultural productions.

REFERENCES

Balding, D.J., M. Bishop and C. Cannings, 2003. Handbook of Statistical Genetics. 2nd Edn. John Wiley and Sons, Chichester, UK, 1: 162-167. ISBN: 0-470-84829.

Carr, S.M., 2007. Principle of a DNA microarray chip: Use as Variant Detector Arrays (VDAs). http://www.mun.ca/biology/scarr/DNA_Chips.html.

Draghici, S., 2003. Data Analysis Tools for DNA Microarrays. 1st Edn. CRP Pres, UK, pp: 17-43. ISBN: 1-58488-315-4 (alk. Paper).

Gaasterland, T. and S. Bekiranov, 2000. Making the most of microarray data (news and views). Nat. Genet., 24 (3): 204-206. DOI: 10.1038/73392. http://www.nature.com/ng/journal/v24/n3/pdf/ng0300_204.pdf.

Gerhold, D., T. Rushmore and C.T. Caskey, 1999. DNA chips: Promising toys have become powerful tools. Techniques. TIBS, 24: 168-173. http://www.ucd.ie/indmicro/lecture/TIBS%2024,168.pdf.

Gunning, K.B. and M.A. Behlke, 2007. SNP-Chips with True On/Off Base Discrimination. Integrated DNATechnologies, 1710 Commercial Park, Coralville, IA 52241 800-328-2661. kgunning@idtdna.com. http://www.idtdna.com.

Hu, N., C. Wang, Y. Hu, H.H. Yang, Giffen, Z.Z. Tang, X.Y. Han, A.M. Goldstein, M.R. Emmert-Buck, K.H. Buetow, P.R. Taylor and P.M.P. Lee, 2005. Genome-Wide Association Study in Esophageal Cancer Using GeneChip Mapping 10 k Array. Cancer Res., 65 (7): 2542-2546. http://cancerres.aacrjournals.org/cgi/reprint/65/7/2542.

Jander, G. and C. Barth, 2007. Tandem gene arrays: A challenge for functional genomics. Trends Plant Sci., 12 (5): 203-210. DOI: 10.1016/j.tplants.2007.03.008. http://www.as.wvu.edu/~cbarth/Jander%20and%20Barth%202007%20TIPS%2013%20p203_210.pdf.

Kerr, M.K., M. Martin and G.A. Churchill, 2000. Analysis of Variance for Gene Expression Microarray Data. J. Comput. Biol., 7 (6): 819-837. DOI: 10.1089/106652700 50514954.

Kim, S., Y. Choi, F.W. Bazer and T.E. Spencer, 2003. Identification of Genes in the Ovine Endometrium Regulated by Interferon_Independent of Signal Transducer and Activator of Transcription 1. Endocr., 144 (12): 5203-5214. DOI: 10.1210/en.2003-0665. http://endo.endojournals.org/cgi/reprint/144/12/5203.

Lee, S.H. and K.L. Hossner, 2002. Coordinate regulation of ovine adipose tissue gene expression by propionate. J. Anim. Sci., 80: 2840-2849. http://jas.fass.org/cgi/reprint/80/11/2840.pdf.

Lehnert, S.A., A.A. Reverter, K.A. Byrne, Y. Wang, G.S. Nattrass, N.J. Hudson and P.L. Greenwood, 2007. Gene expression studies of developing bovine longissimus muscle from 2 different beef cattle breeds. BMC Dev., Biol., 7: 95: 1-13. DOI: 10.1186/1471-213X-7-95. http://www.biomedcentral.com/1471-213X/7/95.

Li, H., C.L. Wood, Y. Liu, T.V. Getchell, M.L. Getchell and A.J. Stromberg, 2006. Identification of gene expression patterns using planned linear contrasts. BMC Bioinfo., 7 (245): 1-11. DOI: 10.1186/1471-2105-7-245.http://www.biomedcentral.com/1471-2105/7/245.

Lipshutz, R.J., S.P.A. Fodor, T.R. Gingeras and D.J. Lockhart, 1999. High density synthetic oligonucleotide arrays. Nature America Inc. http://genetics.nature.com.

McCann, J.A., E.M. Muro, C. Palmer, G. Palidwor, C.J. Porter, M.A. Andrade-Navarro and M.A. Rudnicki, 2007. ChIP on SNP-chip for genome-wide analysis of human histone H4 hyperacetylation. BMC Genomics, 8 (322): 1-8. DOI: 10.1186/1471-2164-8-322. http://www.biomedcentral.com/1471-2105/7/245.

Nakatani, N., E. Hattori, T. Ohnishi, B. Dean, Y. Iwayama, I. Matsumoto, T. Kato, N. Osumi, T. Higuchi, S. Niwa and T. Yoshikawa, 2006. Genome-wide expression analysis detects 8 genes with robust alterations specific to bipolar I disorder: Relevance to neuronal network perturbation. Hum. Mol. Genet., 15 (12): 1949-1962. DOI: 10.1093/hmg/ddl118. http://hmg.oxfordjournals.org/cgi/reprint/15/12/1949.

Park, M.Y., T. Hastie and R. Tibshirani, 2007. Averaged gene expressions for regression. Biostat., pp: 1-16. DOI: 10.1093/biostatistics/kxl002. http://www-stat.stanford.edu/~hastie/Papers/averaged_regression.pdf.

Svrakic, N.M., O. Nesic, M.R.K. Dasu, D. Herndon and J.R. Perez-Polo, 2003. Statistical approach to DNA chip analysis recent. Prog. Horm. Res., 58: 75-93. http://rphr.endojournals.org/cgi/reprint/58/1/75.

Shi, L., 2002. Monitoring the Genome on a Chip. http://www.gene-chips.com.

Walia, H., C. Wilson, P. Condamine, A.M. Ismail, J. Xu, X. Cui and T.J. Close, 2007. Array-based genotyping and expression analysis of barley cv. Maythorpe and Golden Promise. BMC Genomics, 8 (87): 1-14. DOI: 10.1186/1471-2164-8-87. http://www.biomedcentral.com/1471-2164/8/87.

Wallace, C. R.J. Dobson, P.B. Munroe and M.J. Caulfield, 2007. Information capture using SNPs from HapMap and whole-genome chips differs in a sample of inflammatory and cardiovascular gene-centric regions from genome-wide estimates. Genome. Res., 17: 1596-1602. DOI: 10.1101/gr.5996407. http://genome.cshlp.org/cgi/content/full/17/11/1596#References.