Review Article

Indian Journal of Genetics and Molecular Research
Volume 4 Number 2, July - December 2015
DOI: https://dx.doi.org/10.21088/ijgmr.2319.4782.4215.4

Principle, Types and Applications of Microarrays: An Overview

Sudhakar N.*, Karphaga Niveditha K.S., Mohan R.****

Abstract

Gene expression is the conversion of DNA into messenger RNA (mRNA), which are translated into proteins that perform critical functions of the cells. Microarray is a high-throughput technique for studying global gene expression profiling. Microarray mainly works based on the principle of hybridization. Two major types of microarray chips are cDNA arrays and Oligonucleotide arrays. The principle, methodology, types and applications of microarrays are discussed in this review. Apart from studying gene expression, microarrays are also used in studying Single Nucleotide Polymorphisms (SNPs), DNA methylation, micro RNA, copy number variations, etc. in the genome. Though there is a recent boom and excitement in genomic analysis using Next Generation Sequencing (NGS) methods and studying the whole genome expression profiling using RNAseq, it is difficult to substitute the microarrays in terms of diverse applications and cost effectiveness of the latter.

Keywords: Gene Expression; Gene Expression Profiling; Microarrays.

Gene Expression – An Outline

Gene expression is the term used to describe the conversion of DNA into messenger RNA (mRNA) molecules, which are translated into proteins that perform critical functions of the cells. Gene expression is a tightly regulated process, which act as both an "on/off" switch that results in increase or decrease in the expression of certain genes. Scientists study the amount of mRNA produced by a cell at a given point to learn which genes are expressed or suppressed, that in turn can act as gene expression signature or a biomarker.

The methods used for studying gene expression are Northern blotting [1], Reverse-Transcriptase PCR (RT-PCR) [2], Differential Display RT-PCR [3], Serial analysis of gene expression (SAGE) [4] Real-time Quantitative PCR [5,6], etc. Such techniques were limited only to examine few genes at a time.

Microarrays

Microarray is a high-throughput technique, which helps in studying the expression of thousands of genes at a time. Microarray otherwise called Gene chip or DNA chip consists of a small membrane or glass slide containing sequences of thousands of genes spotted in a uniform pattern. The microarray spots can be DNA, cDNA or synthetic oligonuleotides. Initially microarray was developed to study the differential gene expression but refinement of the technique led to detect copy number imbalances and gene amplification in DNA [7], Single Nucleotide Polymorphisms (SNPs), mutation detection and methylation profile etc [8,9,10].

A microarray works by the ability of a given mRNA molecule from the sample, which can hybridize to the complementary DNA sequence. Each single-stranded DNA sequence is made up of four different nucleotides, adenine (A), thymine (T), guanine (G), and cytosine (C) that are connected by phosphodiester bond. Adenine is the complement of thymine, and guanine is the complement of cytosine. Therefore, the complementary sequence to C-T-A-G-C-A will be G-A-T-C-G-T. When two complementary

Author's Affiliation: *Associate Professor, ** Department of Biotechnology, Dr. M.G.R. Educational and Research Institute, Maduravoyal, Chennai - 600 095. India.

Corresponding Author: N. Sudhakar, Associate Professor, Department of Biotechnology, Dr. M.G.R. Educational and Research Institute Maduravoyal, Chennai-600 095 India.
E-mail: nsudha79@gmail.com

sequences find each other, such as the immobilized target DNA and the mobile probe DNA, cDNA, or mRNA, they will hybridize together [11]. By using an array containing many DNA spots, Scientists unravel the expression level of thousands of genes within a cell by measuring the amount of mRNA bound to each site on the array. With the help of a computer, the amount of mRNA bound to the spots on the microarray is precisely measured, generating a profile of gene expression in the cell.

Types of Microarray Chips

Two main methods used to make microarrays are cDNA arrays and oligonucleotide arrays. In cDNA arrays, which are also called spotted arrays, the spots are purified polymerase chain reaction products generated from cDNA libraries or clone collections. They are generally gene fragments greater than several hundred base pairs long [12]. These probes are deposited onto a solid surface in defined locations by an xyz robot. Spot size in an array ranges from 80 to 150 µm in diameter and the array can contain 200 to 30,000 spots / genes in a chip. Gene sequences to be arrayed are selected from several public databases, which contain information on well-characterized genes and expressed sequence tags (ESTs), which represent genes of unknown function. In spotted microarrays it is standard practice to compare the gene expression of two biological samples on one chip. The mRNA is prepared in such a way that the expression of two different samples can be measured simultaneously using a single chip. This method of hybridization is called dual color labeling [13].

On the other hand, high density oligonucleotide microarrays are constructed to an extremely high density and accuracy using short synthetic oligonucleotides with a length of between 20 and 25 nucleotides as the probes. Two leading companies producing these high density arrays are Affymetrix and Agilent Technologies. Especially the Affymetrix Inc Genechips™ are produced by synthesizing tens of thousands of short oligonucleotides *in situ* onto glass wafers, one nucleotide at a time, using a modification of semiconductor photolithography technology [12].

Principle of Microarray Technology

In the case of cDNA microarrays, a reference sample (normal tissue) is studied along with the experimental sample (e.g. tumor sample). Both total and mRNA can be used; Preparation of probe DNA usually includes use of fluorescence dyes such as Cy3 (green color) and Cy5 (red color). For example, one can use Cy3 to label cDNA from normal cells and Cy5 for tumor cells. For microarray experiment, 25µg of total RNA or 5µg of mRNA from experimental and reference samples are reverse transcribed to cDNAs, which are labeled with different fluorescent dyes (Cy dyes – Amersham Biosciences) and mixed. The mixture is then hybridized overnight to the microarray chip and scanned at two different wavelengths. The bound cDNAs are detected using laser scanner and the data is stored in a computer. The intensities of the fluorescence are precisely measured for each spot on the microarray and the ratio of fluorescent intensity reveals the quantity of RNA expressed by the experimental sample in relation to the reference [14,15].

Each spot on an array is associated with a particular gene. Depending on the type of array used, the location and intensity of a color will tell us whether the gene is present in either reference and / or sample DNA. For example, GREEN spot represents control DNA, where either DNA or cDNA derived from normal tissue is hybridized to the target DNA. RED represents tumor DNA, where either DNA or cDNA derived from tumor tissue hybridized to the target DNA. YELLOW represents a combination of control and tumor DNA (GREEN + RED →YELLOW), where both hybridized equally to the target DNA. BLACK represents areas where neither the control nor tumor DNA hybridized to the target DNA [16].

Alternatively, the Affymetrix oligonucleotide arrays uses a single color fluorescent label, where experimental mRNA is enzymatically amplified, biotin labeled for detection, hybridized to the wafer, and detected through the binding of a fluorescent compound (streptavidin-phycoerythrin) [17]. Affymetrix chips contain two types of probes, one perfect match (PM) and another mismatch (MM) probe to measure the gene expression. PM probes in the array have a 25-mer sequence that exactly matches with a gene sequence. MM probes are made in such a way to have a single base mismatch at the binding of a gene to a 25-mer sequence. The purpose of MM probe in the array is to serve as a negative control for binding of a gene sequence to a PM probe [18,19].

The main advantage of Affymetrix GeneChips is their ability to measure the absolute expression of genes in cells or tissues. Their disadvantage, in addition to their higher costs, includes their current inability to simultaneously compare, the degree of expression of two related biological samples on the same array. In addition, oligonucleotide-based microarrays require a prior knowledge of the gene sequences and complex computational manipulation to convert the signals into an actual expression value [20].

Advantages and Drawbacks of Spotted Arrays vs. Oligonucleotide Arrays

Spotted arrays offer the advantage of studying the expression analysis of two biological samples, such as experimental sample and reference sample simultaneously in a single array. In addition, it is offered at a lower price and flexibility in design. This direct comparison of expression profile of two biological samples, such as untreated cells vs. treated cells or normal tissue compared with a cancer tissue, is an enormous advantage for any pair wise analysis. Furthermore, because these arrays can be spotted with thousands of expressed genes and ESTs of unknown function, they offer the potential for the discovery of new genes and defining their role in disease [20]. One disadvantage of spotted arrays is that they provide information only on the relative gene expression between specific cells or tissue samples as opposed to direct quantification of mRNA expression in the case of oligonucleotide arrays.

Applications of Microarrays – An outline

Microarrays in Gene Expression Profiling

Microarrays are used in numerous studies in different diseases including cancer. For an example: through the seminal studies done by the Stanford group using microarray analysis of breast cancer enlightened the molecular subtypes of breast cancer as luminal, normal breast-like, HER2 and basal-like breast cancer based on gene expression pattern [21,22]. The study of breast cancer using microarrays undisputedly changed the perception of the disease and the role of tumor heterogeneity in breast cancer [23, 24].

Microarray also been Used in Other areas of Genomics

- Researchers can use SNP array technology to test an individual for a disease expression pattern to determine whether an individual is susceptible to (at risk of developing) a particular disease or not. Also, SNP array would help in personalized treatment or pharmacogenomics [25].
- Microarray Comparative Genomic Hybridization (CGH) can be used to analyze for genomic gains and losses or for a change in the number of copies of a particular gene involved in a disease state [7, 25].
- A protein microarray (protein chip) is a miniaturized assay system consisting of small amounts of antibodies, proteins, protein fragments or peptides. Three types of protein arrays are currently used to study the activities of proteins: analytical microarrays, functional microarrays and reverse phase microarrays (RPA). Analytical microarrays are typically used to profile a complex mixture of proteins in order to measure binding affinity, specificity and protein expression level in the mixture of proteins. Functional protein arrays are mainly used to study protein-protein, protein-lipid, protein-DNA, protein-peptide and protein-drug interactions to identify enzyme substrates and to profile immune responses. RPAs are used for studying the expression of a target protein in a large number of biological samples [26,27].
- In CpG arrays, CG rich sequences derived from human CpG is land genomic library are arrayed onto solid supports. Amplicons derived from pools of methylated CpG DNA are hybridized to the arrays to detect the genomic DNA methylation. Stefansson *et al* identified the DNA methylation based signatures associated basal-like and luminal-B breast cancer subtypes [28].
- A microRNA (miRNA) microarray is used for studying the expression of 1500 small noncoding RNAs known as miRNA at a time. Thus, genome wide analysis of miRNA expression signatures between normal tissue samples and disease samples, such as cancer can be distinguished to use as a marker for diagnosis and prognosis [29].

Use of Gene Expression Data

Microarray datasets are freely accessible to the public and these datasets can be viewed, downloaded and reanalyzed to come out with a meaningful result. The gene expression data are accessible through Gene Expression Omnibus (GEO), a public repository at the National Center for Biotechnology Information (NCBI). The GEO database stores multiple data submitted by scientific community.

Microarray datasets organized in different sections, a platform (GPLxxx) defines the array template that contain sequence identity tracking information, a sample record (GSMxxx) containing the measured hybridization data along with a description of biological source, a series record (GSExxx) containing experimentally related samples. The datasets are retrieved and downloaded (CEL files) from GEO database by keyword or accession number access (1). BRB-Array tool integrated with data import wizard that import raw CEL files and perform pre-processing steps (filtering, normalization and computing probe-set expression summaries) to process the raw data. Class comparison Tool is used to identify

differentially expressed Genes between two or more experimental conditions. The output result provides list of significant genes with annotation and links to websites containing additional information. In addition, BRB-Array tool directly evaluate differential expression of Gene Ontology categories, Kegg Pathway or Biocarta pathway [30,31].

Endnote

Though there is a recent boom and excitement in genomic analysis using Next Generation Sequencing (NGS) methods and studying the whole genome expression profiling using RNA seq, it is difficult to substitute the microarrays in terms of diverse applications and cost effectiveness of the latter.

In near future, use of DNA microarray will become a part of routine laboratory diagnostics. Complete SNP profile of individuals would be available in database and the clinicians could access the database before prescribing a drug to the patient. In this way, conventional method of clinical diagnostics of a disease coupled with comprehensive molecular level of screening using microarray might revolutionize the disease diagnosis and foster the clinical decision making process precisely for betterment of therapy.

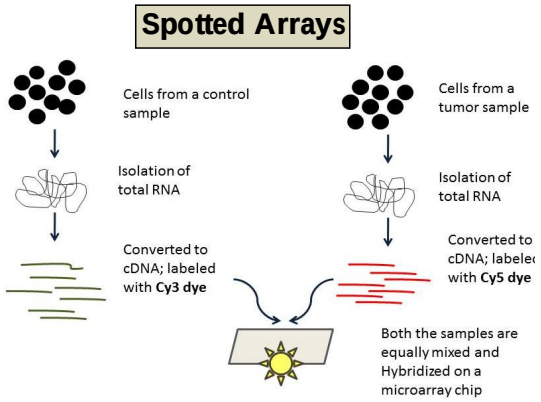

Fig 1a. Spotted arrays

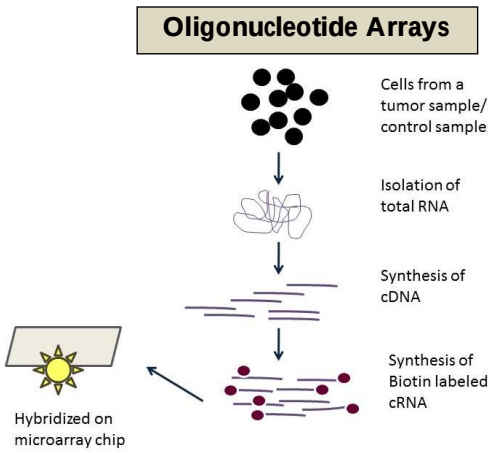

Fig. 1b: Oligonucleotide arrays. The methodology of spotted arrays and oligonucleotide arrays are illustrated in figure

Acknowledgement

We sincerely thank Dr. M.G.R. Educational and Research Institute for providing all the necessary support for our research and preparation of this manuscript. We express our deep sense of gratitude to Dr. Venil N Sumantran, Visiting Faculty, Dr. M.G.R. E & R.I for sharing her experience in analysis of microarray data.

Conflicts of Interest: None

References

1. Yang H, McLeese J, Weisbart M, Dionne JL, Lemaire I, Aubin RA. Simplified high throughput protocol for northern hybridization. *Nucleic Acids Res.* 1993 Jul 11; 21(14): 3337-8.
2. Gause WC, Adamovicz J. "The use of the PCR to quantitate gene expression". *PCR Methods Appl.* 1994; 3(6): S123–35.
3. Liang, P. &Pardee, A.B. Differential display of eukaryotic messenger RNA by means of the polymerase chain reaction. *Science.* 1992; 257: 967–971.
4. Velculescu VE, Zhang L, Vogelstein B, Kinzler KW. Serial analysis of gene expression. *Science.* 1995; 270(5235): 484–7.
5. Livak KJ, Schmittgen TD. "Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta DeltaC(T)) Method". *Methods.* 2001; 25(4): 402–8.
6. Schmittgen TD, Livak KJ. Analyzing real-time PCR data by the comparative C(T) method. *Nat Protoc.* 2008; 3(6): 1101-8.
7. Pollack JR, Perou CM, Alizadeh AA, Eisen MB, Pergamenschikov A, Williams CF, *et al.* Genome-wide analysis of DNA copy-number changes using cDNA microarrays. *Nat Genet.* 1999 Sep; 23(1): 41-6.
8. Mao X, Young BD, Lu YJ.The application of single nucleotide polymorphism microarrays in cancer research. *Curr Genomics.* 2007 Jun; 8(4): 219-28.
9. Baaj Y, Magdelaine C, Ubertelli V, Valat C, Talini

L, Soussaline F, *et al*. A highly specific microarray method for point mutation detection. *Biotechniques*. 2008 Jan; 44(1): 119-26.

10. Bibikova M, Barnes B, Tsan C, Ho V, Klotzle B, Le JM, *et al*. High density DNA methylation array with single CpG site resolution. *Genomics*. 2011 Oct; 98(4): 288-95.
11. Gibson G, Muse SV. *A primer of genome science* (3rd ed.). Sunderland, MA: Sinauer Associates 2002.
12. Schulze A, Downward J.Navigating gene expression using microarrays—a technology review. *Nat Cell Biol*. 2001 Aug; 3(8): E190-5.
13. Shalon D; Smith SJ; Brown PO. A DNA microarray system for analyzing complex DNA samples using two-colorfluorescent probe hybridization. *Genome Res*. 1996; 6(7): 639–645.
14. Schena, M., Shalon, D., Davis, R. W. & Brown, P. O. Quantitative monitoring of gene expression patterns with a complementary DNA microarray. *Science*. 1995; 270: 467–470 .
15. DeRisi J, Penland L, Brown PO, Bittner ML, Meltzer PS, Ray M, *et al*. Use of a cDNA microarray to analyse gene expression patterns in human cancer. *Nat Genet*. 1996; 14(4): 457-60.
16. NCBI website : www.ncbi.nlm.nih.gov. Just the Facts: A Basic Introduction to the Science Underlying NCBI Resources.
17. Ramaswamy S, Golub TR. DNA microarrays in clinical oncology. *J Clin Oncol*. 2002 Apr 1; 20(7): 1932-41.
18. Relógio A, Schwager C, Richter A, Ansorge W, ValcárcelJ. Optimization of oligonucleotide-based DNA microarays. *Nucleic Acids Res*. 2002 Jun 1; 30(11): e51.
19. Irizarry RA, Bolstad BM, Collin F, Cope LM, Hobbs B, Speed TP. Summaries of Affymetrix GeneChip probe level data. *Nucleic Acids Res*. 2003 Feb 15; 31(4): e15.
20. Macgregor PF, Squire JA.Application of microarrays to the analysis of gene expression in cancer. *Clin Chem*. 2002 Aug; 48(8): 1170-7.
21. Perou CM, Sørlie T, Eisen MB, van de Rijn M, Jeffrey SS, Rees CA, *et al*. Molecular portraits of human breast tumours. *Nature*. 2000 Aug 17; 406(6797): 747-52.
22. Sorlie T, Perou CM, Tibshirani R, Aas T, Geisler S, Johnsen H, Hastie T, *et al*. Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications. *Proc Natl Acad Sci U S A*. 2001, 98: 10869-10874.
23. Weigelt B, Baehner FL, Reis-Filho JS. The contribution of gene expression profiling to breast cancer classification, prognostication and prediction: a retrospective of the last decade. *J Pathol*. 2010 Jan; 220(2): 263-80.
24. Colombo PE, Milanezi F, Weigelt B, Reis-Filho JS.Microarrays in the 2010s: Microarrays in the 2010s: the contribution of microarray-based gene expression profiling to breast cancer classification, prognostication and prediction. *Breast Cancer Res*. 2011 Jun 27; 13(3): 212.
25. Mullighan CG. Single nucleotide polymorphism microarray analysis of genetic alterations in cancer. *Methods Mol Biol*. 2011; 730: 235-58.
26. Chen CS, Zhu H. Protein microarrays. *Biotechniques*. 2006 Apr; 40(4): 423, 425, 427 passim. Review.
27. Hall DA, Ptacek J, Snyder M. Protein microarray technology *Mech Ageing Dev*. 2007 Jan; 128(1): 161-7.
28. Stefansson OA, Moran S, Gomez A, Sayols S, Arribas-Jorba C, Sandoval J, *et al*. A DNA methylation-based definition of biologically distinct breast cancer subtypes. *Mol Oncol*. 2015 Mar; 9(3): 555-68.
29. Liu CG, Calin GA, Volinia S, Croce CM. MicroRNA expression profiling using microarrays. *Nat Protoc*. 2008; 3(4): 563-78.
30. Barrett T, Edgar R. Mining microarray data at NCBI's Gene Expression Omnibus (GEO)*. Methods *Mol Biol*. 2006; 338: 175-90.
31. Simon R, Lam A, Li MC, Ngan M, Menenzes S, Zhao Y. Analysis of gene expression data using BRB-ArrayTools. Cancer Inform. 2007; 3: 11-7.