UDC:
xxx
xxx
10.18485/juuntics240101011d

Genetic Chip Technology and Applications

Dragana Đokić[1], Vladimir Đokić[1], Željko Stanković[2], Dalibor Drljača[3], Siniša Tomić[3]*

[1] Doctoral student, University "Union – Nikola Tesla", Belgrade, Serbia,
E-mail: draganadjokic@unionnikolatesla.edu.rs, vladimirdjokic@unionnikolatesla.edu.rs
[2]University "Union – Nikola Tesla", Faculty of Informatics and Computer Science, Belgrade, Serbia,
E-mail: zstankovic@unionnikolatesla.edu.rs
[3]Faculty of Information Technologies, Pan-European Aperion University, Banja Luka, Bosnia and Herzegovina,
E-mail: dalibor.p.drljaca@apeiron-edu.eu, sinisa.m.tomic@apeiron-edu.eu

Abstract: Gene chips are widely used in animal and plant science. Generally, microarrays are of two types, cDNA or oligonucleotide, cDNA was developed at Stanford University, while oligonucleotides were developed by Affymetrix. We will explore genetic chips, the technology used for their production and reading, and their applications. Gene chips, which are becoming increasingly popular throughout the field of molecular biology and biotechnology, are a recently developed technique that provides efficient access to genetic information using miniaturized high-density arrays of DNA or oligonucleotide probes. Both fundamental and applied research benefit greatly from the specific advantages of this method. There are several steps in the design and implementation of DNA chip technology: probe generation, chip fabrication, sample preparation, analysis, readout, and software. At present, the main application of large-scale microarrays is comparative expression analysis. Microarray technology allows for the analysis of expression profiles for thousands of genes in parallel. One of the most significant applications of this technique is gene expression and profiling on a genome-wide scale. Genechip technology is a new technique that will undoubtedly significantly increase the speed of molecular biology research.

Keywords: *Genetic chips, DNA, RNA, biotechnology.*

1. INTRODUCTION

Gene chips are widely used in animal and plant science. Generally, microarrays are of two types, cDNA or oligonucleotide, cDNA was developed at Stanford University, while oligonucleotides were developed by Affymetrix. The construction of cDNA or oligonucleotide helps to compare the gene expression level of treated and control samples by labeling mRNA with green (Cy3) and red (Cy5) dyes. The hybridized gene chip emits fluorescence whose intensity and color can be measured. RNA labeling can be done directly or indirectly. The indirect method involves amino allyl modified dUTP in place of a previously labeled nucleotide. Hybridization of the gene chip generally occurs in the minimum possible volume to ensure the formation of a heteroduplex and ten times more DNA is observed on it. Confocal or semi-confocal laser technologies in combination with a CCD camera are used to acquire images. Genes are used for standardization to detect cDNA gene chips that are not present in treated or control samples. Statistical analysis (image analysis) and cluster analysis software were developed by Stanford University. Gene chip technology has many applications such as expression analysis, gene expression (molecular phenotypes) and regulatory element coexpression. Gene chips, which are becoming increasingly popular throughout the field of molecular biology and biotechnology, are a recently developed technique that provides efficient access to genetic information using miniaturized high-density arrays of DNA or oligonucleotide probes. Both fundamental and applied research have benefited greatly from the specific advantages of this method. Originally developed to enhance the genome sequencing project, the technology has evolved rapidly and has been adapted to a wide range of applications. DNA chips are powerful tools for studying the molecular basis of interactions on a scale that would be impossible using conventional analysis. Recent developments in microarray technology have greatly accelerated research into gene regulation. Arrays are mainly used to identify which genes are turned on or off in a cell or tissue, as well as to assess the level of gene expression under different conditions. This technology has successful ap-

*Corresponding author: draganadjokic@unionnikolatesla.edu.rs

plications for studying the simultaneous expression of thousands of genes and for detecting mutations or polymorphisms, as well as for their mapping and sequencing. We will explore DNA chips, the technology used for their fabrication and reading, and their applications.

2. GENETIC CHIP

The emergence of DNA chips in the second half of the 1990s has become a new biotechnical revolution that is equivalent in importance to the decoding of DNA structures in the 1950s and the investigation of the fundamental laws of molecular genetics, such as the genetic code, and the main assertion of molecular biology in the 1960s, i.e. the discovery of reverse transcription, the creation of the first recombinant constructs in the 1970s and the investigation of enzymatic methods of manipulating genetic material in vitro. This technology has long been has found its way, since the first official meeting of scientists working in this field, held in Moscow in 1991, which was attended by about 50 scientists from Europe and the USA. At that time, the initial goal was to develop a technology for high-throughput sequencing of the human genome, called hybridization sequencing. However, the field has rapidly diversified in its methodology and applications. Today, DNA chips seem to be a common part of biological science. DNA chips have been very successful in various genomic analyses (Debouck C., Goodfellow P., 1999), ranging from SNP (Wang D.G. et all, 1998) detection to functional genomics (Lockhart D.J. et al., 1996). There are different names for chips, such as DNA/RNA chips, Biochip, Genechip or DNA arrays. An array, described as a macroarray, can be defined as an ordered collection of macrodots, or microdots, where each dot contains a single defined type of nucleic acid. Large sets of these nucleic acid probe sequences are immobilized at defined locations on the surface of a substrate capable of accessing large amounts of genetic information from biological samples in a single hybridization assay. Each dot represents the equivalent of a conventional assay performed in a test tube. Dot arrays applied to a biochip reproduce each of these assays several times, up to a total of three or four, thereby increasing the reliability of the assay results. Using the "lock and key principle," complementary hybridization is used in biochip assays. In a typical application, DNA or RNA target sequences are isolated from a biological sample using standard molecular biology protocols. The sequences are fragmented and labeled with detection molecules, and a mixture of labeled sequences ("key") is applied to the array, under controlled conditions, for hybridization with surface probes ("lock"). Sequence complementarity leads to hybridization between two single-stranded nucleic acid molecules, one of which is immobilized on a matrix (Southern E., Mir K., Shchepinov M., 1999). The array is then imaged with a reader to locate and measure the binding of target sequences from the sample to the complementary probe on the array, and software reconstructs the sequence data and presents it in a defined format. In this way, highly parallel DNA hybridization occurs, and assays are possible on miniaturized flat substrates or "chips". There are two variants of chips, in terms of the properties of the aligned nucleic acid sequences of known identity: Format I - cDNA microarrays and Format II - oligonucleotide arrays. Although both cDNA and oligonucleotide chips can be used to analyze gene expression patterns, there are fundamental differences between these methods. Microarrays are subsequently exposed to a labeled sample, hybridized, and largely determine the complementary sequences. Furthermore, there are two basic approaches to creating gene chips. Historically, the first approach was to deposit pre-synthesized oligonucleotides or cDNA fragments onto the chip. The second approach was in situ synthesis using photolithography or piezoelectric printing, which uses technology similar to inkjet printing.

The study of gene expression based on hybridization of mRNA to high-density arrays of immobilized target sequences that can correspond to specific genes is called gene chip technology. cDNA gene chips were developed by Patrick Brown and Ronald Davison at Stanford University (Schena M. et al., 1996). The system is based on standard transcription and PCR amplification followed by cloning of the cDNA for each gene of interest. The cDNA clones are then generally "cloned" onto glass slides, printing technologies have been developed that allow printing of 100,000 clones on a single slide, this system allows for custom arrays (Rose D., 2000). mRNA obtained from biological samples is reverse transcribed and end-labeled using a fluorescent (or occasionally radioactive) laser labeled with nucleotides and hybridized to the array. The fluorescence or radioactive emission from each "spot" is proportional to the relative abundance of a specific transcript in the cDNA pool hybridized to the cDNA spot. The high abundance of the specific transcript in the sample is compared to the high abundance of that transcript in the control sample. This technique, although laborious, provides a fully customizable high-density system with high detection specificity. Oligonucleotide gene chips were developed by Fodor et al. (1991) and Lipshutz et al. (1995). In these systems, oligonucleotide probes are directly synthesized on a solid surface using chemical synthesis supported by lithophotographic techniques. Originally, this strategy was developed to detect DNA single nucleotide polymorphisms (SNPs) (Lipshutz R.J. et al., 1995). It was later applied to measure

mRNA expression. In Affymetrix Arabidopsis arrays, each gene is represented on oligonucleotide probes, while in the central part of the probe, a base is usually replaced by a complementary one. The probe sequences obtained in most cases cover the 3' parts of the coding sequence to increase the sensitivity of the test and the accuracy for SNPs. The RNA obtained from the biological samples is processed before hybridization. Double-stranded cDNA is obtained from total or mRNA (treated) using reverse transcription. cRNA is then obtained using in vitro transcription with biotinylated nucleotides, and then the fragment is hybridized to an oligonucleotide array. If the biotinylated nucleotides are used to synthesize cRNA, the array is visualized by staining with a streptavidin-phycoerythrin conjugate. A value is calculated for the expression of each gene based on the fluorescence derived from each part of the probe.

3. INTEGRATED CIRCUIT TECHNOLOGY

An integrated circuit (IC), also known as a chip or microchip in electronics, is an electronic circuit built on a semiconductor substrate, usually one of the crystals of which is silicon. The circuit is packaged in a hermetically sealed box or non-hermetic plastic capsule, with leads extending from it to the input, output, and power connections, as well as to other connections that may be necessary when the device is put into use. There are also different scales of integration of such circuits. In the case of very large-scale integration, an IC can contain more than 1000 transistors. Intel founder Gordon Moore observed that the number of transistors per semiconductor chip doubles every 18-24 months. Current microprocessor chips (used in personal computers) contain more than 100 million transistors, each a few square centimeters in size. Such a huge degree of integration, which has almost reached the limits of manufacturing technology, helps modern computers to be very fast, compact and relatively inexpensive. A comparable phenomenon has been observed in molecular biology. The miniaturization of certain tools suitable for the manufacture of smart and portable devices, the spotted array system, offers the pharmaceutical, biotechnology and agricultural industries more efficient and economical solutions. DNA microarrays built using photolithography were on a similar path to integrated circuits. In 1998, the Affimetric array contained fewer than 1000 genes, by 2000 it contained 12,000. In the process of manufacturing biochips, semiconductor materials are widely used and widespread. A semiconductor is a solid material whose electrical conductivity at room temperature is between that of a conductor and that of an insulator. At high temperatures, its conductivity approaches that of metals, and at low temperatures it acts as an insulator. The first materials used for semiconductors were the elements germanium, silicon, and tin. Although other materials can be used, chips are usually manufactured on wafers of single-crystal silicon, i.e. silicon in which the orientation of all the crystals is the same. The manufacturing steps include image formation, impurity doping, photolithography, and packaging. Silicon oxide, also called silicon dioxide, is grown on the surface during the image formation phase. Photolithographic methods are used to selectively remove the oxide from the silicon. A layer of photoresist is added to the oxide layer and then exposed to ultraviolet light through a mask. After exposure, the silicon wafer is developed, a process that removes unwanted oxide areas. Impurities add charge carriers to the silicon, and this process produces the unique electronic properties of the semiconductor. Additional layers of silicon are deposited to create, for example, bipolar transistors.

4. MICROCHIP DESIGN

There are several steps in the design and implementation of DNA chip technology: probe creation, chip fabrication, sample preparation, analysis, readout, and software. Many strategies have been explored for each of these steps.

Microarray fabrication begins with the selection of probes to be printed on the array. In most cases, they are selected directly from databases including GeneBank (Benson D.A. et al., 1997) and UniGene (Schuler G.D. et al., 1996). The two most commonly used types of arrays, Format I and Format II, differ in the size of the arrayed nucleic acids. In Format I, cDNA probes, which are relatively long, DNA molecules (500–5000 bases in length) are immobilized on a solid surface such as a membrane, glass, or silicon chip. The probes are often single-stranded DNA fragments processed using the polymerase chain reaction (PCR). These PCR products are amplified with gene-specific primers generated using chromosomal DNA as a template and then purified by precipitation or gel filtration, or both. Spotting cDNA or PCR products representing specific genes on the array produces DNA arrays. Deposition of a few nanoliters of purified material typically at 100-500 mg=ml generates each spot on the array. Printing is done either by a high-precision robotic system, i.e., by a high-speed robot (array) or by using inkjet technology that spots a sample of each gene product onto a number of matrices in a batch operation or prints. These types of arrays are mainly used for large-scale screening and expression studies. In Format II, an array

of oligonucleotide probes is produced either by in situ light-directed chemical synthesis or conventional synthesis followed by on-chip immobilization . Those with short nucleic acids (oligonucleotides up to 25 nt) are useful for mutation detection and expression monitoring, gene discovery, and mapping.

DNA chip technology has evolved in two main ways. In one method, nucleic acids (previously chemically synthesized oligonucleotides or single-stranded DNA fragments, i.e. cDNA) are immobilized on the surface of a chip in sequence to form capture probes. Using either printing needles or inkjet printing, many copies of the nucleic acid sequence can be attached to the substrate (Figure 1).

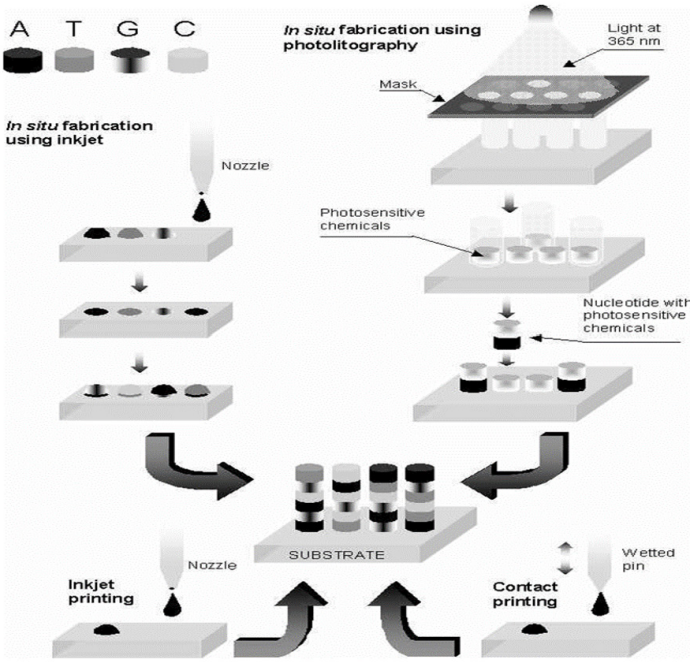

Figure 1. Methods for creating DNA arrays (Gabig-Ciminska M.,& Cminski A., 2001)

Contact printing involves wetting a printing stylus with a nucleic acid solution and tapping it onto the surface of the microarray. Inkjet printing ejects uniform droplets of the solution onto the substrate. An alternative, known as in situ fabrication, builds the oligonucleotide sequence at each site one nucleotide at a time. This is done using inkjet or photolithography (Figure 1). Construction of such chips begins with a glass substrate that has been chemically prepared with sites ready for nucleotide binding. In inkjet printing, nucleotide solutions are ejected from a nozzle onto the substrate and then chemically fixed to the surface. The next set of nucleotides is dropped onto the first and chemically fixed. The process is repeated until the desired set of nucleic acids is complete. In photolithography, a photosensitive chemical that is released under illumination closes the sites. Light at 365 nm shines through a patterned mask onto the chip, causing the capping chemical to detach from the areas it hits, revealing the primed sites. A solution containing one of four types of nucleotides (each molecule of which is itself bound to a capping molecule) is then washed over the chip. The nucleotides bind only to the areas that have been exposed and add a capping layer to themselves. Since the process can be repeated with a different mask and a different nucleotide, different DNA sequences can be built on the chip. Multiple probe arrays are synthesized simultaneously on a large wafer. This parallel process improves reproducibility and helps achieve high synthesis rates (Southern, Mir, & Shchepinov, 1999). Production chips can pack 400,000 probes into 20 mm patches. One drawback of the current photolithography method is that a new set of masks must be produced for each new array type. There is a maskless technique that uses an array of micromirrors that reflect to appropriate locations on the chip.

Gene chip sample preparation can be used to measure gene expression levels in a variety of ways. One of the most popular applications of gene chip comparison is to compare gene expression levels in two different samples, e.g. the same cell or cell type under two different conditions. This is based on labeling mRNA extracted from one condition with one dye, e.g. green, and from the other condition with a different dye, e.g. red. The hybridized gene chip is ignited by a laser and scanned at wavelengths suitable for the detection of red and green, as shown in Figure 2.

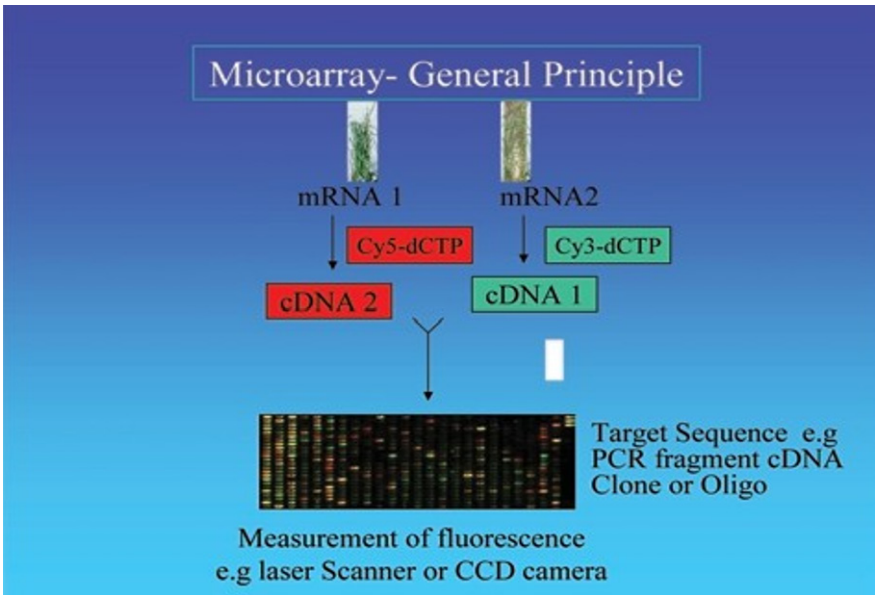

Figure 2. Microarray: General functional principle

The amount of fluorescence emitted by the laser pulse corresponds to the amount of nucleic acid bound to each spot. If nucleic acid from the sample in state 1 is present in large quantities, the spot will be green, while if nucleic acid from the sample in state 2 is present in large quantities, it will be red. If both are equal, then it will be yellow. If neither is present, it will not fluoresce and will then appear black. The intensity of fluorescence and color for each spot determines the relative expression levels as genes in both samples are expressed and can be assessed. In this way, thousands of data points, each providing information on the expression of a particular transcript, can be obtained from a single experiment. The complexity of gene array analysis means that tissue sampling becomes a crucial factor in the data produced. As gene chips have been used with increasing frequency in recent years, the amount of variation in gene expression between samples, even from the same tissue in the same person, can exist. Precise sampling (including factors such as sampling at the same time of day and month) and the ability to sample homogeneous tissues, using techniques such as laser capture (Simone N.L., et al., 1998) microdissection, are key steps in obtaining accurate analysis. At the same time, the amount of tissue required can be a problem if relatively large amounts of RNA are needed. However, newer techniques have been developed to amplify RNA that allow extraction from small tissue (Hertzberg M., et al., 2001) samples. Brown's group at Stanford University described how a pool of RNA obtained from 11 different human tumor cell lines became a kind of de facto universal human reference RNA. Recently, a labeled oligonucleotide complementary to each function on the array has proven to be an efficient reference, without the complications associated with mRNA-derived references (Dudley A.M. et al., 2002). Ultimately, the development of gene chip technology or other processes that allow for parallel high-throughput measurements of absolute large amounts of RNA are needed to provide robust characterization of transcriptomes of specific cell lineages, developmental stages and diseases.

Expression labeling protocols are based on reverse transcription of mRNA, either from highly purified poly(A) mRNA or total RNA extract. Extensive purification of RNA is necessary to remove all contaminating proteins, polysaccharides and other organic materials, especially RNases. Many protocols developed for high-quality RNA extraction use a variety of in-house and commercial reagents. Initial target labeling protocols, where reverse transcription from mRNA is prepared, used poly(dT) primers in the presence of fluorescently labeled nucleotides, usually cy3- or cy5-conjugated (dCTP or dUTP). Cy3- or cy5-conjugated nucleotides are bulky, however, and their use makes their absorption by standard enzymes very inefficient. In addition, the rate of absorption can vary between dyes, potentially leading to color bias. In an alternative method, an amino allyl modified dUTP is used in place of a pre-labeled nucleotide. After reverse transcription, the free amine group on the amino allyl modifies the dUTP and can be coupled to the reactive N-hydroxysuccinimidyl ester of a fluorescent dye. Although this technique is more time-consuming than direct labeling, its advantages include better sensitivity, less color bias, and reduced cost.

5. ANALYSIS

Although gene chip technology exists, the basic methods of analysis are well established. An image analysis program, in most cases, is used to calculate the quantitative ratio of gene expression levels between the two samples to be compared. Before the software is able to interpret the intensity value for each probe as a grid, it must be superimposed on the gene chip image to map the location of the pixels representing each dot. A software package for initial image analysis was developed at Stanford University and is available to academic users. Most scanners are sold with the option of an image analysis software package such as Imagene (Biodiscovery, Los Angeles, CA, USA) Quantarray (GSI Lumonics) or GenePix 2.0 (Axon instruments, Foster City, CA, USA). Some of these packages use cluster analysis algorithms to group genes in a sequence according to their expression profiles over a number of experiments. A version of this type of program, called Cluster, is also available from Stanford University. Bilban et al. (2002) introduced a method for filtering false positive and false negative DNA gene array experiments. This was achieved by evaluating a set of positive and negative controls by receiver operating characteristic (ROC) analysis. The advantage of this approach is that users can define thresholds based on the sensitivity and specificity of the observations. The area under the ROC curve allows for quality control of gene array hybridization. This method was applied to custom gene arrays developed for the analysis of invasive melanomas where tumor cells were examined with fewer misclassified genes. Provided that appropriate positive and negative controls are included in the set on the gene array, ROC analysis eliminates the problem of arbitrary selection of threshold levels in gene array experiments. The proposed method is also applicable to adapted to commercially available cDNA gene chips and can help improve the predictive reliability of gene chip experiments.

6. CONCLUSION

We can conclude that gene chips are becoming increasingly popular throughout the field of biotechnology. Recently, a technique has been developed that provides efficient access to genetic information, and software development allows for insight into the complete scanned sequence.

The readout of the image from the gene chip is recorded as an image obtained by a scanner for the detection of a fluorescent signal or for the detection of radioactive signals. Scanners currently available use either scanning confocal laser technology or charge-coupled device (CCD) cameras.

More than a decade of rapid progress in biology has swept an avalanche of genetic information among scientists. But the analysis of such a huge number of novelties, leads to the conclusion that studying the inner workings of cells or diagnosing diseases would not be practical without high-throughput technologies, such as DNA microarrays. They allow scientists to search for the presence, production, or sequence of thousands of genes at the same time. Currently, the power and breadth of DNA microarrays have become a dominant factor in genomics, transcriptomics, pharmacogenetics, and biology, while also becoming increasingly important in preclinical research and even clinical studies. Innovative efforts, combining fundamental biological and chemical science with technological advances in the fields of micromachining and microfabrication, should lead to even more powerful devices that will accelerate the realization of large-scale genetic research testing. GeneChip offers extraordinary tools for genetic analysis, appropriate applications for these new devices, and the intellectual contribution that is still needed.

At present, the main application of large-scale microarrays is comparative expression analysis. Microarray technology allows the analysis of expression profiles for thousands of genes in parallel. One of the most significant applications of this technique is gene expression and their profiling on a genome-wide scale. Genechip technology is a new technique that will undoubtedly significantly increase the speed of biotechnology research.

REFERENCE

Benson D.A., Boguski M.S., Lipman D.J., Ostell J. (1997). GenBank. Nucleic Acids Res 25: 1–6.

Bilban M., Buehler L.K., Head S., Desoye G. and Quaranta V. (2002). Defining signal thresholds in DNA microarrays: exemplary application for invasive cancer. BMC Genomics 3:19.

Debouck C., Goodfellow P. (1999). DNA microarrays in drug discovery and development. Nat Genet 21: 48–50.

Dudley A.M., John A., Martin A.S., George M.C. (2002). Measuring absolute expression with microarrays with a calibrated reference sample and an extended signal intensity range. PNAS, USA 99:7554-7559.

Fodor S.P.A., Read J.L., Pirrung M.C., Stryer L., Lu A.T., Solas D. (1991). Light-directed, spatially addressable parallel chemical synthesis. Science 251:767-773.

Gabig-Ciminska M., Cminski A. (2001). An Introduction to DNA Chips. ResarchGete.

Hertzberg M., Sievertzon M., Aspeborg H., Nilsson P., Sandberg G., Lundeberg J. (2001). cDNA microarray analysis of small plant tissue samples using a cDNA tag-target amplification protocol. Plant J. 25:585-591.

Lipshutz R.J., Morris D., Chee M., Hubbell E., Kozal M.J., Shah N., Shen N., Yang R., Fodor, S.P.A. (1995). Using oligonucleotide probe arrays to access genetic diversity. Biotechniques 19:442-447.

Lockhart D.J., Dong H., Byrne M.C., Follettie M.T. (1996). Expression monitoring by hybridisation to high density oligonucleotide arrays. Nat Bitechnol 14: 1675–1680.

McGall G., Labadie J., Brook P., Wallraff G. (1996). Ligh-directed synthesis of high-density oligonucleotide arrays using semiconductor photoresists. Proc Natl Acad Sci USA 91: 5022–5026.

Rose D. (2000). Micro fluidic technologies and instrumentation for printing DNA microarrays. In: Microarray Biochip Technology. Ed. Schena, M., pp.19-38, Eaton Publishing.

Schena M., Shalon D., Heller R., Chai A., Brown P.O., Davis R.W. (1996). Parallel human genome analysis: microarray-based expression monitoring of 1000 genes. PNAS, USA. 93:10614-10619.

Schuler G.D., Boguski M.S., Stewart E.A., Stein L.D. (1996). A gene map of the human genome. Science 274: 540–546.

Simone N.L., Bonner R.F., Gillespie J.W., EmmertBuck M.R., Liotta L.A. (1998). Laser-capture microdissection: opening the microscopic frontier to molecular analysis. Trends Genet. 14:272-276.

Southern E., Mir K., Shchepinov M. (1999). Molecular interactions on microarrays. Nat Genet 21: 5–9.

Southern E., Mir K., Shchepinov M. (1999). Molecular interactions on microarrays. Nat Genet 21: 5–9.

Wang D.G., Fan J., Siao C., Berno A. (1998). Large-scale identification, mapping and genotyping of single-nucleotide polymorphisms in the human genome. Science 280: 1077–1082.