DNA microarrays and pharmacogenomics

Rolando Meloni*, Olfa Khalfallah, Nicole Faucon Biguet

Laboratoire de Génétique de la Neurotransmission et des Processus Neurodégénératifs, C.N.R.S. UMR 7091, Bât C.E.R.VI,
Hôpital de la Pitié-Salpêtrière, 83 Boulevard de l’Hôpital, 75013 Paris, France

Accepted 26 June 2003

Abstract

The DNA microarrays have proven to be a state of the art technique for high throughput comprehensive analysis of thousand of genes in parallel. The application of a DNA microarray to compare normal and pathological cells, tissues or organs may allow, along with classical positional cloning techniques, to speed up the discovery of genes and gene pathways implicated in several diseases. This in turn will result in further application of the DNA microarray technique in the field of pharmacogenomics in order to characterize and validate new therapeutic targets, their mechanism of action, metabolic pathways and unwanted secondary effects. However, the lack of standardized criteria for the analysis and interpretation of the huge amount of data generated by DNA microarrays may hamper its utilization on a routine basis in the human health domain.

Keywords: DNA microarrays; Expression profiling; Pharmacogenomics; Drug target; Gene discovery

1. Introduction

The different sequencing projects that have elucidated the human [1,2] and other key model organism genomes open the challenge to determine the exact role of all the genes in the development and functioning of the living beings. The emerging discipline of functional genomics addresses the description of global gene expression patterns from two classical point of view: what is the specific tissue expression of genes and how the gene expression pattern is modified in physiological, as well as in disease conditions. In this perspective, the DNA microarrays technique coupled with the availability of genome information enables performing the global analyses of the expression of thousands of genes in a single assay in a specific cell, tissue or organ [3].

In classical molecular biology cloning studies the DNA was purified, cut and ligated in a vector, in order to transform bacteria and obtain a bank of unknown clones. This bank was then screened with a specific probe for finding a particular gene. DNA array technology turns upside down this approach: instead of screening an array of unknown clones with a defined probe, an array of defined DNA fragments is probed with a sample of unknown composition [4–7].

The primary concept of microarray analysis is based on the precise positioning of tethered fragments of DNA (oligonucleotides, genomic fragments, cDNAs) representing the microarray probes. These probes corresponding to specific genes are deposed or synthesized in a known configuration and at high density on a solid support matrix such as a nylon membrane or a glass slide. Then, the RNAs extracted from the sample of interest are labeled using either radioactivity or fluorescent markers in order to generate the targets that are hybridized to these arrayed probes. The corresponding probe will bind a labeled target with an affinity that is proportional to the degree of complementarity between the two sequences. The probes is always in excess so that there is no limit in the quantity of target it can bind. In this way, the signal generated by the probe-target complex and detected by a scanning apparatus is proportional to the quantity of the target and allow establishing the identity and relative abundance of the RNAs present in the sample (Fig. 1). Thus, the exploitation of microarrays is based on the standard molecular biology technique of complementary hybridization of nucleic acids [8], adapted to the handling of minute volumes, yielding high throughput and high quality data [9].

The utilization of DNA microarrays has increased exponentially over time, as shown by the number of papers published relating the results obtained with this technique. A Medline search yielded 1 hit in 1995 [4] up to 731 hits in
2. Expression profiling using DNA microarrays

The paradigmatic utilization of DNA microarrays is based on the expression profiling, that is the relative quantification of transcripts by comparing the mRNAs extracted from two samples using single or dual labeling. The samples are, for example, microorganisms, cell lines, tissues, organs or pluricellular organisms that differ for a given treatment, developmental stage, disease status etc. Different protocols based on different approaches and materials are utilized in expression profiling, each of them with peculiar advantages and drawbacks [10,11]. A major care in these experiments is devoted to the quality both of the microarrayed probes and the RNAs targets. The RNAs extracted from two different samples to be compared must be extracted in optimal standard condition in order to reduce to a minimum inter-individual variability, sample heterogeneity and sampling errors. The pooling of multiple samples and the performance of multiple replicates may also greatly contribute to reduce the introduction of spurious variations and increase the reproducibility between different experiences.

The probes, corresponding to series of known DNA sequences that hybridize specifically and unambiguously to their target, are arrayed using usually a robotic device on a solid support such as a nylon filter or a glass slide in order to constitute the array platform.

Nylon membranes may be hybridized to more sensitive radioactive targets and can be reused up to four or five times, while glass is a durable material that sustains high temperatures and washes of high ionic strength. DNA samples can be covalently attached to a glass matrix coated with the widely used poly-l-lysine or aminosilane. Since glass is nonporous, the hybridization volume is minimal, enhancing the hybridization kinetics. Glass has also a low inherent fluorescence and significantly reduces the background “noise”.

Moreover, the major advantage of glass over filter arrays is that the glass DNA microarrays are more adapted to fluorescent detection and can utilize double labeling while, in contrast, nylon microarrays are generally probed in serial or parallel hybridization reactions. Therefore, for reasons of sensitivity, handling, resistance and convenience the glass supports have almost completely supplanted the filters in the current microarray research.

3. DNA microarray platforms

Two main microarray platforms are commonly used: cDNA microarrays, which utilize cloned probe molecules corresponding to characterized expressed sequences, and oligonucleotide microarrays, made of synthetic probe sequences based on database information [11]. A cDNA microarray comprises a collection of gene sequences, usually pure PCR products ranging in size from 100 to 2000 bp, derived from cDNA and expressed sequence tag (EST) clones. Oligonucleotide microarrays contain instead an assembly of single-stranded probes synthesized according to sequence information gathered in databases.

The oligonucleotides probes are much shorter than cDNA microarrays, which utilize cloned probe molecules and require special care in their design, since their base composition may strongly influence the conditions of hybridization. However, long oligonucleotides of about 60 bp length offer very good performances in terms of specificity and sensitivity [12].

4. Fabrication of DNA microarrays

Usually, cDNAs probes are positioned at fixed spots on the support by automatic robots, while oligonucleotide probes are better suited to be directly synthesized in situ. Practically, each element in a cDNA microarray is produced by the deposition of a few nanoliters of a purified PCR product at a concentration of 100–500 μg/ml. Spots are typically 100 μm in diameter and can be deposited at a density of up to 20,000 features/cm². Contact (mechanical microspotting) or noncontact (ink jetting) methods are used for spotting. In mechanical microspotting, the DNA sample is loaded into a spotting pin by capillary action, and a small
Oligonucleotide microarrays are made by directly synthesizing in situ hundreds of thousands of different oligonucleotides using photolithography-directed combinatorial chemistry. Alternatively, oligonucleotide may be synthesized and then arrayed by ink jet spotting. An ink-jet printing method has also been developed for oligonucleotide synthesis in situ using standard phosphoramidite chemistry.[12]

Microarray density, defined as the number of elements per square cm, has increased over the years, up to tens of thousands targets allowing to analyze whole genomes on single slides or in chambers. Both have the characteristic of reducing the signal intensity[23,24]. However, while the complexity of the arrays may be stepped up by increasing the number of clones, for example, from 5000 to 20–30,000, the maintaining of high quality controls is not proportional, but more likely exponential when increasing the number of targets. In parallel with the development of whole genome chips, the development of focused DNA microarrays or thematic DNA microarrays, intended to detailed analysis of selected gene subsets, such as pathways or disease markers, has allowed generating a target in which sequence representation is skewed compared with the original mRNA pool. In order to overcome this problem, RNA amplification based on the Eberwine mRNA amplification procedure[16] has been proposed[17–19]. Moreover, the amplification protocol may be coupled with the microdissection laser capture method to the same level of high throughput performance with as few cells as possible down to the single cell level, some form of amplification is needed for targets obtained from such samples. The PCR technique may offer a highly efficient method for exponentially amplifying a population of single stranded cDNAs. However, the nonlinear amplification results in a target in which sequence representation is skewed compared with the original mRNA pool. In order to overcome this problem, RNA amplification based on the Eberwine mRNA amplification procedure[16] has been proposed[17–19]. Moreover, the amplification protocol may be coupled with the method for the retrieval and the microarray profiling analysis of specific cell types that are scarce and/or dispersed in a non-homogeneous tissue[21,22]. An alternative to the amplification of scarce mRNA material is represented by Tyramide signal amplification that increase substantially the signal intensity[23,24].

The fluorescent labeling of the RNAs is obtained either directly or indirectly using a fluorescent-labeled nucleotide such as dUTP conjugated with the *N*-hydroxysuccinimidyl fluorochromes Cy3 or Cy5. These fluorochromes have high incorporation efficiencies with reverse transcriptase, good photostability and their respective signals are easily separated. The direct labeling of the RNA is obtained by incorporating in the reverse transcription products primed with a poly(dT) the Cy3-dUTP and Cy5-dUTP modified nucleotides. In order to improve this method, which is prone to dye bias, the incorporation may also be performed with an amino allyl modified dUTP that is coupled after reverse transcription to the *N*-hydroxysuccinimidyl dye. This method of indirect labeling requires an increase in work time that is largely compensated by an enhanced sensitivity, lack of dye bias and lower costs.

A limitation of cDNA microarray technology is the large amount, in the range of tens of microgram, of RNA required to produce an adequate signal over noise, which becomes a fundamental issue with low-abundance transcripts or when limited quantities of material are available. Fluorescence detection requires about 10 μg of total RNA (equivalent to about a million cells), whereas radioactive detection enables detection with as little as 0.1 μg of starting total RNA (10,000 cells). However, since the ultimate goal is to reduce variability in order to carry out expression profiling with as few cells as possible down to the single cell level, some form of amplification is needed for targets obtained from such samples. The PCR technique may offer a highly efficient method for exponentially amplifying a population of single stranded cDNAs. However, the nonlinear amplification results in a target in which sequence representation is skewed compared with the original mRNA pool. In order to overcome this problem, RNA amplification based on the Eberwine mRNA amplification procedure[16] has been proposed[17–19]. Moreover, the amplification protocol may be coupled with the microdissection laser capture method[20] for the retrieval and the microarray profiling analysis of specific cell types that are scarce and/or dispersed in a non-homogeneous tissue[21,22]. An alternative to the amplification of scarce mRNA material is represented by Tyramide signal amplification that increase substantially the signal intensity[23,24].

6. Hybridization

The hybridization step may be conducted under cover slips or in chambers. Both have the characteristic of reducing...
considerably the volumes to carry on the procedure. The
problems inherent to the even distribution of the labeled
target over the probes, which may create spatial bias in
the signal, have been greatly reduced with the utilization
of automated apparatuses. These specific systems assure a
uniform handling of the solutions, mixing, shaking at con-
trolled temperature for highly reproducible experimental
conditions and results.

7. Analysis of microarray data

A scanning apparatus, using the confocal laser or the
charge coupled device technique, detects the binding re-
action of the target to the probe. The data generated by
the scanning of the microarray is usually extracted using
commercial softwares such as GenPix and Quant-Array for
being stored and normalized before being analyzed. The
amount of data that a single microarray run can produce may
range between 100,000 and a million data points, with tens
or hundreds of runs for a typical experiment [25]. A first
low-level analysis for data collected from a microarray ex-
periment aims to background elimination, filtration, and nor-
malization, for the removal of systematic variations. Then,
high-level microarray analysis is implemented for data min-
ing in order to identify relevant patterns related to the domain
investigated.

The programs used for this type of data analysis use vari-
ious statistical techniques. These techniques are based either
on an exploratory approach without an “a priori” hypothesis
that is best suited for a clustering analysis [26], or a hypothe-
sis driven data analysis that utilizes a computational method
for recognizing reproducible patterns in the microarray [27].

8. Applications of DNA microarray

The major application of DNA microarrays is the mea-
surement of gene expression in a wide range of experimental
situations that include but are not limited to:

• gene discovery for the global description of genes po-
tentially involved in developmental, physiological, and
pathological processes [28],

• tumor profiling [29],

• gene regulation for the description of regulatory networks,
assuming that genes regulated in parallel share common
control mechanisms [30],

• diagnosis for the identification of patterns of gene expres-
sion related to disease states that may be used as diagnost-
ic and prognostic indicators [31],

• drug discovery and toxicology aiming at the identification
of disease-related genes that may become targets for drug
therapy, as well as indicators for drug response and play a
relevant role in the development of personalized medicine
[32,33].

Additional applications of the DNA microarray technique
other than differential expression screening include mutation
detection [34] and genotyping [35].

9. DNA microarrays and pharmacogenomics

Although the molecular biology revolution has introduced
several innovations in the processes of drug discovery, the
limited number of cloned genes initially reduced the number
of potential targets. Moreover, the validation of such targets,
that is linking the gene with the possibility of therapeutic
exploitation, required the positioning of the target gene in
the pathway leading to the pathophysiological processes.
The majority of human genes are now available as targets
and gene expression profiling by DNA microarrays has the
potential of providing a rapid access to the identification and
validation of novel therapeutic targets.

In this perspective the DNA microarrays may be used to
explore either variations in gene expression determining the
individual phenotype or variation in DNA sequence that rep-
resent the individual genotype [36]. The study of expression
phenotype may help in the discovery by positional cloning of
genes that play a role in complex polygenic human diseases
such as diabetes, hypertension, psychiatric diseases that are
refractory to the classical approaches used for monogenic
diseases.

The determination of variation in DNA sequences by
DNA microarray represent the other approach for determin-
ing the genetic blueprint underlying the predisposition to
complex disease. The combination of these two approaches
allows establishing the correlation between the phenotype
and the genotype that may be instrumental for pharmacoge-
nomic studies aiming to implement a personalized medicine
and provides the pharmacological industry with new and
more effective targets for drug development.

10. Discovering therapeutic targets

Several studies have already been conducted in order
to discover genes differentially modulated in pathological
conditions. These studies have evaluated gene variation
concomitant, for example, with monogenic diseases such as
cystic fibrosis [37], for which the causative gene has been
already identified. However, the major interest of the DNA
microarray approach resides in the study of gene variations
associated in common and genetically complex diseases
such as schizophrenia [38] for which it may generate a lead
for the discovery of susceptibility genes that remain still
unknown. In general, since the utilization of DNA microar-
rays generates huge amounts of data, target discovery with
this method requires the implementation of strong exclu-
sion criteria to define a good target from the thousands of
genes on the array. Although DNA microarray research is
not hypothesis driven, rigorous but unbiased data gathering
can narrow the search for genes that represent the best targets for future hypothesis-driven studies. Selection criteria for establishing a shortlist of high-priority candidates that show a particular expression pattern in a disease setting typically include an increase or decrease in expression, disease specificity, and tissue or cell-type selectivity. Once the number of candidate targets obtained from a primary screen has been reduced to a manageable number, a secondary screen may be needed for further assessing the role of these genes in the disease. These investigations may utilize time courses, dose–response curves to drug treatments, alternative tissue samples and animal models [39]. However, for genes with unknown functions subsequent analysis may be also required with methods such as quantitative RT–PCR, in situ hybridization and protein expression analysis.

An utilization of DNA microarray related to target discovery reverse the approach used for this purpose. Instead of comparing differences in gene expression between normal and pathological samples, mRNA changes induced by drugs or other bioactive compounds are assessed in order to deduce previously unknown actions of these products. This approach, called "forward pharmacology", may highlight a potential new therapeutic effect of a common drug for diseases in which the genes and pathway modifications by the drug are implicated [40].

11. Therapeutic target validation

Several approaches can be taken to validate and to prioritize candidate therapeutic targets once a shortlist has been identified. In this endeavor, differential mRNA expression is aimed at highlighting, using more advanced bioinformatic analysis, the biological pathways in which the potential target genes are implicated. This is the so called ‘guilt by association’ approach, since the genes that are coordinately regulated are assumed to participate in the same pathway and may constitute, in turn, several new targets for therapeutic intervention. Moreover, the discovery of potential pathways implicated in the disease may allow proposing new models for understanding the etiopathogenesis of the disease investigated.

The strategies coupled with DNA microarray analysis that are instrumental for target validation and related pathways discovery are the gene knockdown and knock-in strategies in cells, model organisms and mice. These techniques are the most powerful not only for highlighting the potential actions of drugs, identifying phenotypic changes and potential side-effects but also for recognizing pleiotropic effects of the target. A paradigmatic example of this approach in a mammalian cell system is represented by the antisense suppression of the expression of a subunit of protein kinase A. This allowed identifying two clusters of genes that define the "proliferation-transformation" and "differentiation-reverse transformation" profiles in cancer cells and tumors [41]. In this perspective, a less cumbersome approach than the knock-out or knock-in techniques for gene inactivation is RNA interference (RNAi) [42]. Thus, RNAi gene silencing in combination with DNA microarrays offers a high efficient and innovative strategy for high throughput target validation [43].

Besides target discovery and validation, DNA microarrays offer the possibility of investigating drug metabolism. DNA microarrays may evaluate at once the whole dynamic network of genes regulated in response to the challenge with drugs and xenobiotics either in livers of drug-treated rats or cultured primary human hepatocytes. The identification of the pathways implicated in the metabolism of a specific drug allows, for example, avoiding to select a drug candidate that is metabolized by only a polymorphic gene, since this drug can accumulate to toxic concentrations in individuals carrying specific polymorphisms [39].

Finally, DNA microarray profiling of gene expression also can be used for drug safety and toxicology studies. The assessment of the immediate transcriptional events and the secondary responses to the physiological perturbation resulting from a drug treatment in the experimental model utilized may elucidate the mechanisms underlying the adverse effects of drugs. Thus, in this endeavor, the DNA microarray profiling may also be useful to identify biomarkers that predict either adverse pathological responses or the efficacy of the drug tested.

In this perspective, DNA microarrays uses may be extended in the future to the implementation and speeding up of clinical trial. However, as hinted by the complex process that may be required for the validation of biomarkers as reliable predictors, the utilization of DNA microarrays will require rigorous standardization of the procedure for experimental reproduction and data analysis evaluation before being accepted in clinical studies by regulatory agencies [44].

12. Conclusions

DNA microarrays have been proven to be a novel way of getting a comprehensive overview of phenotypic variations in physiological and pathological conditions that may significantly accelerate the pace of discovery in several domains related to human health. However, it has been pointed out that for the first time in the history of the biomedical sciences, the ability to generate data in vast quantities is overcoming the possibility of understanding them [11]. Although procedures for the interpretation of microarray data are rapidly evolving, the transition from data gathering to knowledge remains a daunting challenge in the exploitation of the DNA microarrays.

References


