ELSEVIER

NeuroToxicology 24 (2003) 321–332

NeuroToxicology

Review

Use of Microarray Technologies in Toxicology Research

Kent E. Vrana*, Willard M. Freeman†, Michael Aschner

Department of Physiology and Pharmacology, Wake Forest University School of Medicine, Winston-Salem, NC 27157-1083, USA

Received 8 August 2002; accepted 3 October 2002

Abstract

Microarray technology provides a unique tool for the determination of gene expression at the level of messenger RNA (mRNA). The simultaneous measurement of the entire human genome (thousands of genes) will facilitate the uncovering of specific gene expression patterns that are associated with disease. One important application of microarray technology, within the context of neurotoxicological studies, is its use as a screening tool for the identification of molecular mechanisms of toxicity. Such approaches enable researchers to identify those genes and their products (either single or whole pathways) that are involved in conferring resistance or sensitivity to toxic substances. This review addresses: (1) the potential uses of array data; (2) the various array platforms, highlighting both their advantages and disadvantages; (3) insights into data analysis and presentation strategies; and (4) concrete examples of DNA array studies in neurotoxicological research.

Keywords: Gene–environment interaction; Microarray technologies; Neurotoxicological research

INTRODUCTION ON ROLE OF GENE EXPRESSION IN TOXICOLOGY

Neurotoxicological research into gene–environment interactions stands at the beginning of a new era. The tremendous increase in genetic information—knowledge of both sequence and function—requires that traditional methods of examining one gene at a time be supplemented with high throughput screening tools. To take full advantage of the increases in genomic information, methods have been developed that examine hundreds or thousands of genes simultaneously. As a research community, we need to appropriately embrace these new technologies, like microarrays. These new methods will not replace traditional hypothesis-based testing of single-gene products, but serve as discovery tools that allow entire genomes to be examined for the role of gene products in neurotoxicology.

The flow of genetic information is depicted in Fig. 1. The sequencing and organization of the genome is referred to as genomics. Technologies in this area focus on DNA sequencing and the association of DNA polymorphisms (nucleotide variations) with health and disease. Functional genomics is concentrated on the study of expressed gene sequences, and therefore, utilizes technological tools that assess mRNA levels. The intent of functional genomic studies is to better understand how patterns of gene expression (tissue-specific identity and levels of expression) determine normal and abnormal physiological states. Concerted efforts are underway to develop and popularize technologies that will directly assess global protein expression by two-dimensional electrophoresis and mass spectrometry, in a new field termed proteomics. This latter field is important in its own right, but will not be addressed in the present review. Global examination of modified proteins (cf. Conrads et al., 2002) has also been postulated but at present this is not an easily achieved technological reality.

* Corresponding author.
E-mail address: kvrana@wfubmc.edu (K.E. Vrana).
† Present address: Vollum Institute for Advanced Biomedical Research, Oregon Health & Science University, L474, 3181 SW Sam Jackson Park Road, Portland, OR 97239, USA.

0161-813X/02/$ – see front matter
PII: S0161-813X(02)00193-6

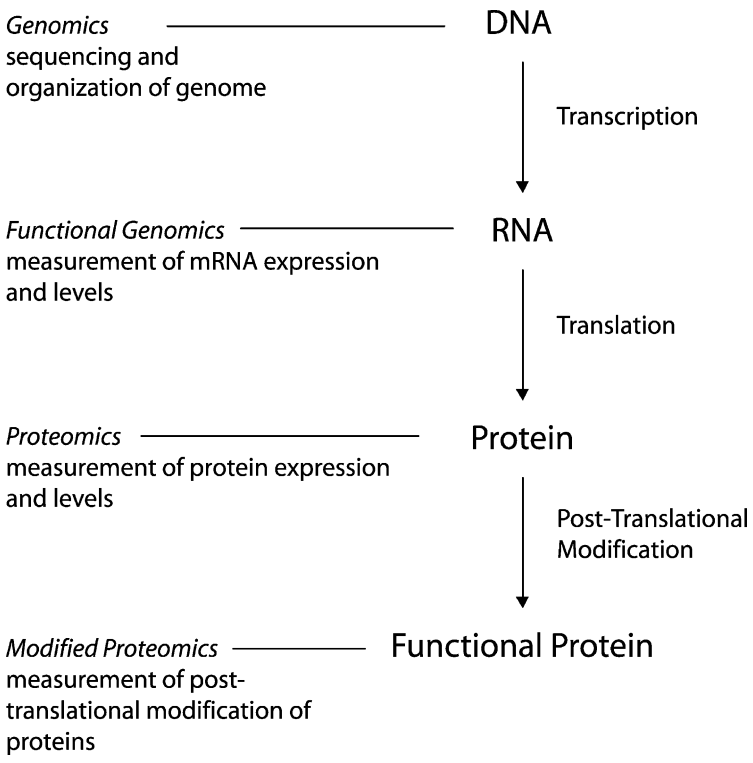

Fig. 1. The flow of genetic information. Genetic information flows from DNA into mRNA through transcription and then from mRNA to protein through translation. Proteins can be further modified post-translationally to alter their function. The technology for assessing the full complement of nucleic acids or proteins at that level is given in italics.

Microarray technology provides a unique tool for the determination of gene expression at the level of messenger RNA (mRNA). The simultaneous measurement of large fractions of the genome (thousands of genes) facilitates the uncovering of specific gene expression patterns that are associated with disease. Another important application of microarray technology is in ''toxicogenomics'', which is predicated on the premise that each individual possesses unique patterns of gene expression that, in turn, exhibit individual responses to a particular toxic substance. Finally, within the context of neurotoxicological studies, the technique can provide a screening tool for the identification of molecular mechanisms of toxicity, differentiating between cell-specific responses and enabling the researcher to identify those genes and their products (either single genes or entire pathways) that are involved both in conferring resistance or sensitivity to toxic substances.

The present review: (1) addresses the potential uses of array data; (2) describes the various array platforms, highlighting both their advantages and disadvantages; (3) provides insight into data analysis and presentation strategies; and (4) provides concrete examples of DNA array studies in neurotoxicological research. Throughout the review, wherever possible, examples will be drawn from genome-wide studies in neurotoxicology, though it must be stated from the outset that the toxicology field has not fully embraced this powerful technique. This has occurred for a number of different reasons—because of the expense of this experimental approach, unfamiliarity with the technique, and/or because of its perception as being a non-hypothesis driven ''fishing expedition''. Accordingly, a major aim of this review is to describe concepts necessary to readily understand gene array technology, with the intent and hope that the technique will be adapted for neurotoxicological applications and become a routine tool for research in this field.

A key component of future toxicology research will be the study of allostatic gene expression (environment-induced changes in gene expression) indicative of disease and/or pharmacological or environmental exposure. This concept is explained in Fig. 2 (see below). Keep in mind that every cell (with a few exceptions such as germ and red blood cells) in the body contains the same genetic information (genetic legacy). The identity of a cell or organ is determined by the subset of that genetic heritage (the cell-specific gene set), which is expressed in a tissue-specific fashion. Similarly, whether a particular cell/tissue type (e.g. liver tissue) is in a state of health or disease can depend on which genes are being expressed and at what levels.

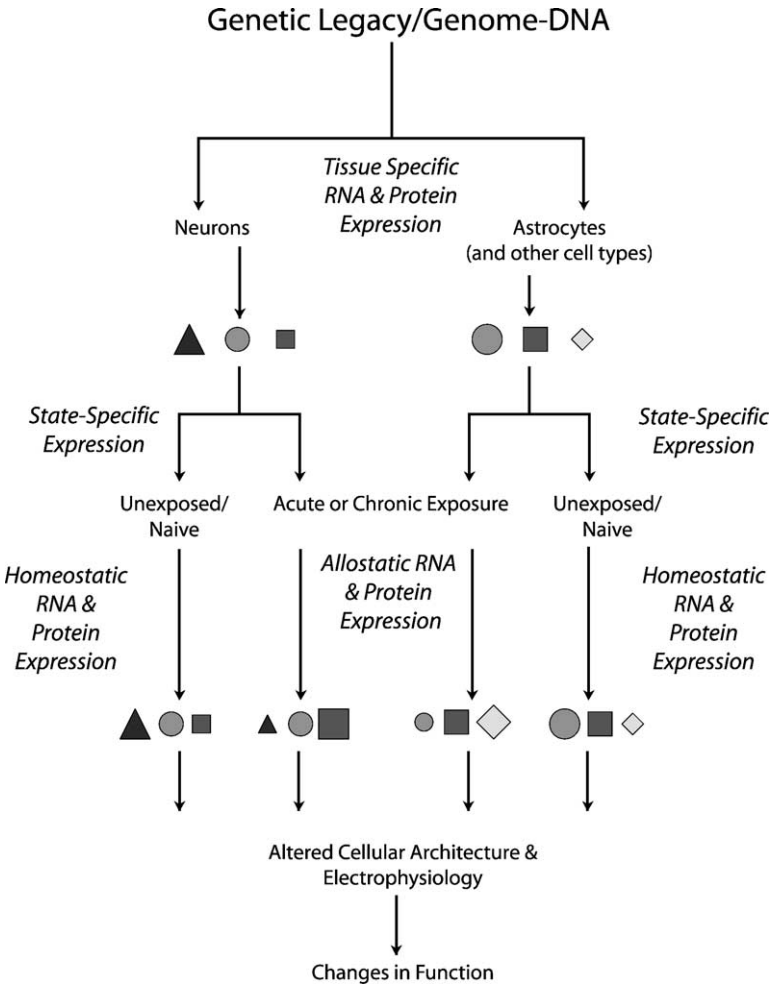

Fig. 2. State-specific gene expression. The portion of the genome used not only defines the nature of a cell, but also its response to toxic agents. Homeostatic gene expression is the ''normal'' pattern of gene expression within a cell that determines its identity and function. Allostatic gene expression is the new pattern of gene expression in response to a toxic agent. The changes in this pattern of gene expression, as compared to the homeostatic state, contribute to the altered function/dysfunction of the cell. Shaded shapes represent different transcripts, and their relative size indicates expression level.

The causes of deleterious gene expression patterns can be internal or external (e.g. resulting from internal stress hormones or following external toxin exposure). If there is a normal homeostatic set point that is typical of health, then there is likely to be an altered pattern that is typical of disease or toxicological insult. Allostatic gene expression is defined as drug- or environment-induced changes in gene expression that is indicative of disease and/or pharmacological or environmental exposure. This concept of allostasis refers to the observation that a physiological system, when repeatedly perturbed from its homeostatic ''normal'' condition, will respond to the new condition and subsequently adopt a new allostatic condition.

How, then, does an investigator identify this altered pattern or allostatic gene expression? We have already established that it will be inefficient to take a single-gene approach. Moreover, many of these epigenetic effects will result from altered expression *patterns* of constellations of genes, as opposed to changes affecting single genes. Under traditional approaches, this pattern recognition is not possible. Now, however, several technology platforms permit discrimination of gene expression patterns. These platforms are helping to generate compendia of gene expression data that can aid researchers in identifying pathways and uncharacterized genes.

ALTERED GENE EXPRESSION IN TOXICOLOGY

One relatively simple example of the toxicological interaction of varying cell types is represented by neuron–glial relationships. The understanding of molecular signals that regulate neuron–glial interactions has

increased greatly with the advent of molecular and cellular biological techniques, as well as genetically modified mice. Studies in which cell ablations are genetically targeted with ectopic gene expression, and gene knockouts performed with single-cell specificity, have established the distinct roles played by different cell types and genes during development. For example, a primary genetic disorder of astrocytes has been invoked in the etiology of Alexander disease, which is characterized by astrocyte cytoplasmic inclusions referred to as Rosenthal fibers, in association with increased expression of heat-shock proteins (Brenner et al., 2001). Sequence analysis of DNA samples from patients with Alexander disease phenotypes point to non-conservative mutations in the coding region of glial fibrillary acidic protein (GFAP), establishing a primary genetic disorder that is linked to astrocytes. In the mature CNS, the astrocyte is an active participant in a variety of homeostatic functions (Sonnewald et al., 2002), and is perhaps involved in the processing of information in the CNS (Vesce et al., 1999), a staggering departure from the old dogma that the astrocyte serves only as physical support for juxtaposed neurons. Furthermore, toxins that uniquely afflict astrocytes play an important role in the pathogenesis of neurodegeneration, as exemplified by altered glutamate metabolism and glutamatergic neurotransmission in hyperammonemia and hepatic encephalopathy (HE) (Norenberg, 1995). Thus, characterization of changes in cell-specific gene expression with cDNA microarray techniques offers a unique means for identifying single- or multiple-gene products associated with cell-specific neurotoxicity.

Fig. 2 depicts the genetic legacy and neuron–astrocyte interactions that mediate homeostatic interactions within the CNS, with different genes being represented by the colored shapes. Under ‘‘normal’’ conditions (parallel outside tracks in Fig. 2), cell-specific RNA and protein expression in astrocytes and neurons vary both in type (represented by symbols) and abundance (size of symbols). Some genes are uniquely expressed in a specific cell type, for example, GFAP and tyrosine hydroxylase are uniquely expressed in astrocytes and catecholamine neurons, respectively (triangle and diamond in Fig. 2). Other genes are expressed in multiple cell types but at different abundances (denoted by the circle and rectangle), as exemplified by lower levels of the tripeptide glutathione (GSH) in neurons compared to astrocytes (Sagara et al., 1993). Microarray technology can be profitably used in the development of a global understanding of gene expression abnormalities that contribute to neurotoxicity and progression of neurodegenerative changes, for differentiating between responsive genes in acute and chronic toxin exposure, and in delineating cell-specific gene expression patterns (parallel inside tracks in Fig. 2) that contribute to perturbed interactions between neurons and astrocytes. These genes include those regulating cell differentiation (during development), oxidative stress and energetics, inflammatory responses, neurotransmitter transport, neurotrophic factors, cell death pathways, and cytoskeletal proteins, just to name a few. Obviously, however, this mixed cell identity and function also creates the need for neuroanatomical discrimination within some experiments (differentiating neuron from astrocyte or differentiating gene expression in varying neuron types).

A recent example of microarray use in neurotoxicological research combines toxic drug administration with a genetic knockout model (Krasnova et al., 2002). Transgenic mice overexpressing Cu/Zn superoxide dismutase (SOD) and wild-type mice were treated with neurotoxic doses of amphetamine and gene expression in the striatum was assayed by macroarrays at a series of time points. The subset changes seen by microarray were confirmed by quantitative RT-PCR. This allowed for standard statistical comparisons between transgenic SOD and wild-type mice over the 24 h after amphetamine administration. For example, the induction of Activin A was shown to be significantly lower in transgenic SOD mice, suggesting a greater oxidative load in wild-type mice. This paper combines many of the aspects discussed next of experimental design, data analysis and post hoc confirmation.

GENERAL APPROACH TO DNA ARRAY TECHNOLOGIES

With the completion of the human and other genome projects (and the identification of 30,000+ potential human genes), traditional models of examining one gene at a time are being supplemented by large-scale screening technologies. DNA hybridization arrays are a common form of screening technology and allow the analysis of hundreds to thousands of genes in parallel (Lockhart and Barlow, 2001). In the past, Northern blotting, dot blots, in situ hybridization, and quantitative reverse transcription-polymerase chain reaction (QRT-PCR) were the common methods for investigating changes in gene expression. While these approaches remain in common use, they suffer a pervading problem of low throughput.

Hybridization array technology, on the other hand, offers to bypass many of the limitations of these techniques by simultaneously creating labeled copies

of multiple RNAs and then hybridizing them to many different, gene-specific, DNA molecules. The nomenclature has developed whereby the labeled sample RNA is termed the target and the individual gene sequences placed on the array are termed probes.

Although arrays are increasingly used for gene expression analysis, there are caveats to their use. One limitation to this technology is that arrays only measure relative levels of mRNA expression (Sample A has 100% more of the specific RNA than Sample B), and not absolute amounts (Sample A has 1000 copies of the RNA and Sample B has 500 copies of the transcript). Moreover, most hybridization arrays are not designed to differentiate between alternatively spliced transcripts of the same gene and, in some cases, between highly homologous members of a gene family. Finally, a change in messenger RNA does not necessarily correlate with a change in protein expression (Anderson and Seilhamer, 1997), and the translated protein often requires further modification to realize its full activity. These latter two points are a common, and legitimate criticism of array technology because it measures an intermediate step (mRNA levels) and not functional product (active protein). However, until proteomic technologies (Fig. 1) become universally accessible to the research community and dramatically increase in sensitivity and reproducibility, hybridization arrays are the best opportunity for studying gene expression on a genomic scale.

ARRAY PLATFORMS

Several different hybridization array formats have been developed. Current array formats can be categorized into three groups: macroarrays, microarrays, and high-density oligonucleotide arrays (GeneChips). While terms like microarray and GeneChips are sometimes used interchangeably, and microarray is often used to describe the technology in general, we also use the term microarray to describe a distinct DNA array format. The varying hybridization array platforms differ according to the material used to construct the platform (matrix), type of probe on the array, probe number/density, array size, and type of label. An understanding of the strengths and weaknesses of each platform is necessary to decide which is appropriate for an individual investigator's research aims.

Macroarrays

Macroarrays (Fig. 3) are generally defined by the deposition of probes onto membranes or plastic and by

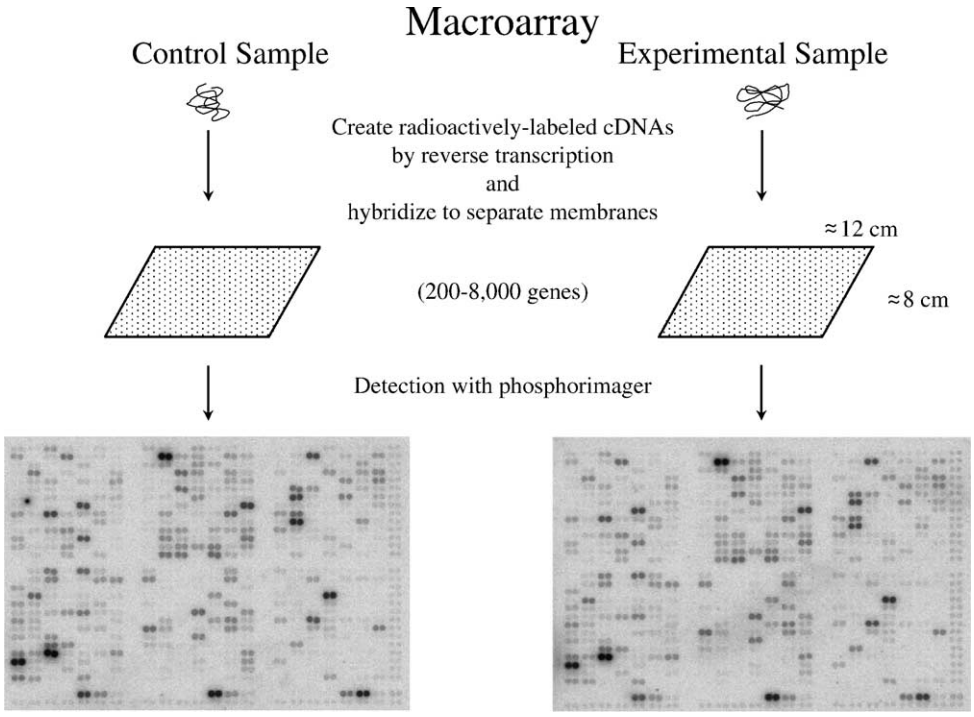

Fig. 3. Macroarray. Macroarrays are the most basic form of hybridization array. The predominant features of macroarrays are that they use radioactivity as a label, probes are deposited on either a membrane or plastic base, and experiments can typically be performed without specialized equipment.

the use of radioactivity for detection. The term macroarray, as opposed to microarray (discussed in the next section), also refers to the lower probe density on these arrays. While density varies among arrays, the term macroarray is useful because of other inherent differences of these membrane-based arrays. Currently, cDNA clones, PCR amplicons, or oligonucleotides are spotted onto membranes using spotting robots or ink-jet like printers. Macroarrays are unique among hybridization arrays in that they generally use radioactive target labeling. After labeling the target, control and experimental samples are hybridized to individual and separate arrays and phosphorimagers detect the bound target-probe complexes. These arrays, typically containing between 200 and 8000 genes, are commercially available for a wide variety of organisms and genes, and can be obtained from a number of companies. ''Custom'' macroarrays can also be constructed in-house and may contain as few as a dozen or as many as thousands of genes.

A major advantage of the macroarray is that any laboratory that has conducted Northern blots or Southern blots can easily conduct this type of experiment. Moreover, the macroarray does not require access to any special equipment, except for the relatively ubiquitous phosphorimager. The macroarray experiment requires that at least two different samples of RNA be radioactivity-labeled and then used to query identical, but separate, membranes. This has an obvious disadvantage because the use of two different membranes provides an opportunity for hybridization error and therefore experimental variability. The advantages of macroarrays make it the prime approach for initial forays into functional genomics before moving on to other, higher throughput, formats.

Microarrays

Microarrays (Fig. 4) can be differentiated from macroarrays (Fig. 3) in three ways. First, microarrays generally use glass microscope slides as a matrix and, second, they use fluorescent dye-labeling detection. Third, microarrays tend to have a larger number and higher density of probes than macroarrays. As with macroarrays, probes are made from clones, PCR amplicons, or oligonucleotides and spotted robotically onto the matrix surface. Targets are labeled with different fluorescent dyes (typically Cy3 or Cy5). This allows a competitive hybridization scheme to be used where both samples are hybridized to the same glass slide. By using competitive hybridization of separate targets to the same array, a major source of variability, hybridization, is eliminated. For experimental designs incorporating multiple different samples, only one dye and sample is used on each array to facilitate data analysis later. Machines which allow for simultaneous

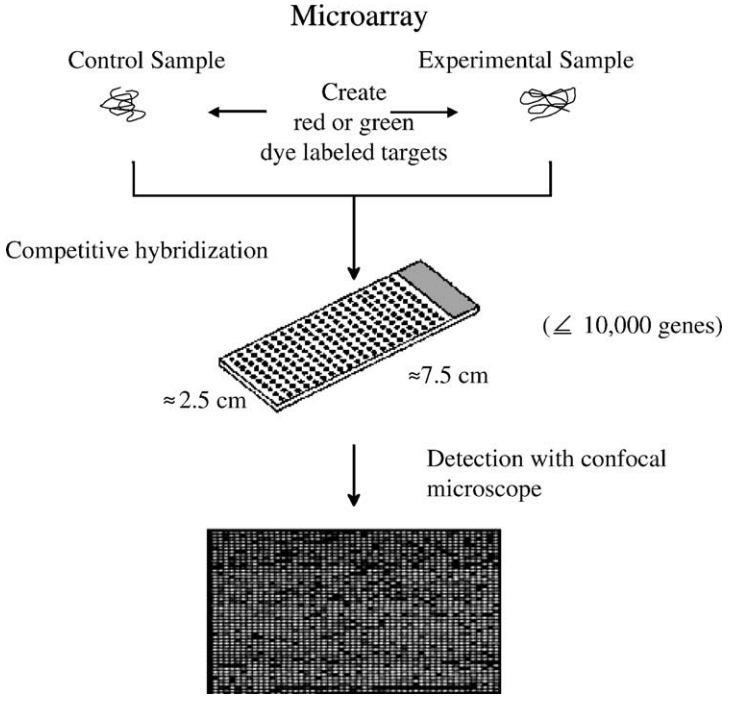

Fig. 4. Microarray. Microarrays consist of probes spotted onto microscope slides. Targets are labeled with different fluorescent dyes and competitively hybridized to the same array. Final imaging requires a fluorescent scanner that can excite and detect at multiple wavelengths.

hybridization of multiple arrays have been developed making large experiments more practical. Like macroarrays, an ever-expanding number of microarrays are commercially available. Many research institutions are currently investing heavily in the equipment to produce custom microarrays in-house; however, there are relatively few of these facilities that are generating publication quality, reproducible reagents for general use. In this regard, the NIEHS has established a National Center for Toxicogenomics (NCT; can be viewed at http://www.niehs.nih.gov/nct/org.htm). NIEHS and its Microarray Center provide resources and support for Intramural investigators. Moreover, a Toxicoinformatics Group develops the NCT's National Toxicoinformatics Database and provides liaison and coordination with other genomic-related databases that are developing nationally and internationally.

High-Density Oligonucleotide Arrays (Affymetrix)

High-density oligonucleotide arrays, also called GeneChips (Fig. 5) differ from other formats in that the probe is generated in situ on the surface of the matrix. The leader in this type of array is Affymetrix (Santa Clara, CA), a company that uses a unique combinatorial synthesis method. This method makes use of a process called photolithography to construct probes on the array surface by building oligonucleotides one base at a time directly onto the chip surface. Because the combinatorial synthesis scheme has a finite efficiency at each step, synthesis of oligonucleotides longer than 25 bases is problematic. As a result of using these 25-mer oligonucleotides for gene expression analysis, mismatches and spurious target-probe binding can take place because of the limited specificity and binding affinity for a 25-residue oligonucleotide. To overcome this problem, a series of oligonucleotides that differ by a one base mismatch from the gene-specific probe are also included on the array and can be used to determine the amount of mismatch hybridization, which can then be subtracted from the signal. These arrays, which are available only from Affymetrix, contain up to 40,000 genes and 1,000,000 oligonucleotide features (including multiple mismatch controls for each gene) and provide the highest density of probes of any array. GeneChips have found a great deal of use in academic research settings. The major limitations of this technology is that the arrays themselves can only be made by Affymetrix and that dedicated hybridization and imaging equipment are required, making this a comparatively expensive approach.

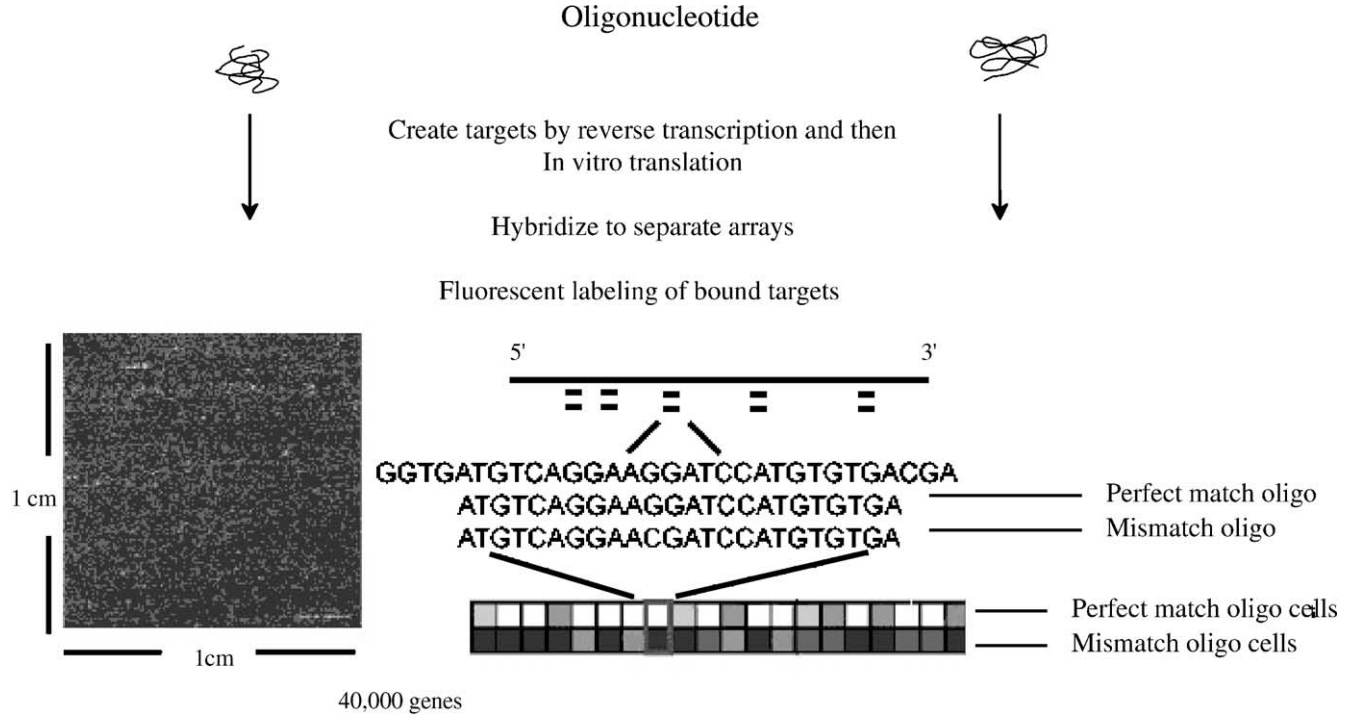

Fig. 5. GeneChips. GeneChips, or high-density oligonucleotide arrays, are made by in situ construction of oligonucleotide probes. This type of array uses multiple probes per gene as well as mismatch probes to determine the specificity of target-probe binding.

EXPERIMENTAL DESIGN

Experimental design in neurotoxicological studies of gene expression remains critical for production of high quality and relevant data. The simplest studies involve comparing a control group of animals to a group that has received some sort of toxicological insult. While this can be a useful approach, more complex experimental designs can be useful in determining not only which genes change in expression but, those that are directly linked to toxicological effects. Two strategies that can be used are multiple groups and correlation to other data. For multiple groups a series of toxicological doses are a typical example. Genes are then profiled by microarrays and those exhibiting a dose–response relationship can be identified. Combining microarray data with other biochemical or anatomical measurements of toxicity can be profitable in providing more direct linkages between a change in gene expression and the toxic end result. This style of experiment has seen a great deal of use in the cancer field in attempts to provide prognostic molecular markers of tumor progression (e.g. Beer et al., 2002). Other traditional experimental models such as time-courses, and transgenic animals can also be used to help refine microarray studies.

In combination with careful model selection, two critical parameters of experimental design are the timing of sample collection and the variability of array results. For example, when studying a toxicological insult that results in cell death, RNA samples must be collected before cell death and degradation in order to assess mechanisms of toxicity rather than the end results of cellular death.

One inherent aspect of hybridization arrays that affects experimental design is the variability of the array results. Thousands of genes are measured for their relative expression levels on a single array, resulting in a large number of dependent variables with only one measurement. This can lead to a large number of false-positive and false-negative results. The first approach to deal with this variability is to perform replicate experiments. While this increases the costs of already expensive experiments, it is absolutely required. Replicate experiments can either be performed with the same RNA or with RNA from different cells or animals. This allows assessment of technical and biological variability, respectively. Lastly, while replicate arrays and careful data analysis (described in the next section) can produce genes that have a low probability of being false positives, post hoc confirmation of gene expression changes, either at the level of mRNA or protein, must be included in an experimental design. Studies which report only a list of genes are, correctly, becoming regarded as only preliminary results and not worthy of publication in and of themselves. Confirmation of array results by other means also allows for standard statistical assessment of the gene expression changes discovered with the hybridization array. It is in the confirmation role that one-gene-at-a-time approaches serve to complement the discovery role of screening methodologies like hybridization arrays.

DATA ANALYSIS

Data analysis is a critical component of hybridization array experiments and poses a number of challenges due to the large amount of data generated even in a single experiment. The most basic goal of array data analysis is to identify genes that are differentially regulated. As stated previously, hybridization arrays are prone to false positives and therefore analysis strategies attempt to decrease this error rate without overly increasing false negatives. A large number of analysis methods are being developed (Hess et al., 2001; Miles, 2001; Brazma and Vilo, 2000). There is also a great deal of information available on microarray data analysis on the Internet, two good portal sites are: http://www.sciencemag.org/feature/plus/sfg/index.shtml, and http://linkage.rockefeller.edu/wli/microarray/. While data analysis is a complex subject that cannot be fully addressed here, the basic steps of array data analysis: image analysis, normalization, presence calling, differential expression calling and data presentation are described below. For the sake of simplicity, the most basic, empirical approach to data analysis is described.

Image Analysis

The initial steps in data analysis are background subtraction and normalization. The principals of both are similar to the techniques used with conventional nucleic acid or protein blotting. Background subtraction corrects for non-specific background noise and permits comparison of specific signals. For illustration, if the signal intensities for the control and experimental spots are 400 and 600, respectively, it would appear that the experimental value is 50% higher. However, if a background of 200 is subtracted from both signal intensities, the experimental value is actually 100% higher than control (200 versus 400). Background is often taken from the blank areas on the array.

A complication to background subtraction is that differences in background across the array can affect some spots more than others. An alternative is to use either a local background for the area around each spot or designate spots with the lowest signal intensities for background determination. The latter may be a more accurate determination of non-specific background because it represents the non-specific binding of targets to probe. Background intensities from blank areas (no nucleic acids) do not contain probe, and therefore are arguably a different form of background.

Normalization

Normalization is the process by which differences between separate arrays are accounted for. All hybridization array experiments require the use of normalization for accurate comparisons. For example, when a pair of macroarrays representing control and treated samples show a difference in overall or total signal intensity, such differences can arise from unequal starting amounts of RNA or cDNAs, from labeling reactions of different efficiencies, or from differences in hybridization. Any of these factors can skew the results. One method of normalization is to use housekeeping gene(s)—a gene thought to be invariant under experimental conditions—for comparison. If the signal for this gene is higher on one array than the other, it can be used to normalize the data. Housekeeping genes may be problematic because they themselves vary under some experimental conditions (Lee et al., 2002). To overcome the variability of these genes, researchers turn to a ''sum'' approach for normalization. This strategy is based on the precept that the total amount of labeled target should be the same in all samples. That is, even though individual genes will have selected increases and decreases, on balance, the total hybridization signal should be constant. Therefore, equilibrating the sum of the intensities for all control and experimental spots can be used to normalize arrays. In a similar vein, the median value of signal intensities can be used. This value is less susceptible to distortions caused by outlying signals. More complex methods of normalization are being developed such as those that apply Lowess intensity-dependent normalization, which is even less likely to be skewed by outliers (Yang et al., 2002).

Gene Calling

The next step in data analysis is to determine which genes were detected by the array analysis. When using large-scale arrays, there will be probes for genes that are not expressed in the sample, or are expressed below the level of detection. In fact, arrays are inherently much less sensitive than RT-PCR. This means that there must be some method for ''calling'' a gene as being detected by the array. For array experiments, the approach to calling a gene as present can be a set threshold, expressed as a percentage above background. According to our experience, a 50% above background threshold has worked well for macroarrays, though this is purely an empirical judgment. High-density oligonucleotide arrays take advantage of the multiple oligonucleotide pairs (mismatch and match) for each gene to compute the presence of a gene. Not all users of hybridization arrays perform this gene-calling step. However, implementation of a gene-calling scheme can increase the quality of the end results by eliminating low intensity signal noise, which represents fluctuations in background signal and not biological events. Gene-calling methods can be adjusted to suit the needs and tolerance to variability of the individual experiment.

Differential Expression

Once genes have been designated as being detected by the array, the next step is to determine which genes were changed in their expression by the experimental condition. Differential expression calls are more problematic than simply determining whether or not a gene is present. There are many methods available for this task, but there are no universally accepted standards (for example, Tusher et al., 2001; Zhang et al., 2002; Zhou and Abagyan, 2002). For illustration, we will describe the most basic non-statistical method. The normalized signal for each gene is converted into a ratio of the treated signal intensity over the control signal intensities, the expression ratio. An arbitrary expression ratio cutoff can then be set (e.g. 50% increase or 33% decrease) which a gene must meet in all of the replicate arrays. While this approach was useful in early hybridization array studies, it is not based on statistical principles and is dependent on the original signal intensity. It is for this reason that statistical methods of differential gene calling are being developed (Miles, 2001). Researchers using hybridization arrays find this part of data analysis one of the most daunting aspects of the experiments. Many methods and software packages are being developed for this purpose. Until the field comes to accept a common standard, however, confirming changes observed on hybridization arrays by other means

(quantitative RT-PCR or immunoblotting, for example) remains the standard for definitively demonstrating changes in gene expression.

Pattern Recognition/Clustering

A secondary goal of array experiments can be to look for groups of genes that behave similarly across a series of treatments (i.e. clustering analysis). Again, there are a number of clustering methods, but none has been definitively shown to be superior to others. *k*-means, self-organizing maps, hierarchical techniques, and principle component analysis are examples of these types of analysis (Alon et al., 1999; Ben Dor et al., 1999; Raychaudhuri et al., 2000; Tamayo et al., 1999). The key point is to use the wealth of information for the many independent gene determinations to uncover novel patterns of biological effects. In fact, neurotoxicology research will be a particularly fruitful arena for the application of clustering analysis. For instance, in the field of toxicogenomics, *patterns* of altered gene expression will prove to be diagnostic of classes of toxicants. Therefore, routine screening of compounds may ultimately involve application of clustering techniques to simple surrogate systems.

Data Presentation and Archiving

After all the work in obtaining and analyzing microarray data, only the steps of presentation and archiving are left. Data presentation standards are still evolving and the key to successful presentation is clarity. Data is often presented in scatterplots and various clustering plots. While there are no traditional modes of presentation, the data must be presented in a clear manner that highlights the points of interest and does not swamp the reader with the shear weight of these large datasets. Efforts are underway to provide public repositories of gene expression data, along the general lines of GeneBank. The Gene Expression Omnibus (GEO) is the National Center for Biotechnology Information program for providing a publicly housed repository of large-scale gene expression data (Edgar et al., 2002). Unlike genomic data, expression data is highly dependent on the experimental and data analysis methods used to generate it. In order to develop a standard nomenclature and set of information to fully describe a microarray experiment, the minimum information about a microarray experiment (MIAME) standard has been promulgated (Brazma et al., 2001).

PRACTICAL EXAMPLES OF THE USE OF DNA ARRAYS IN NEUROTOXICOLOGY RESEARCH

Microarray technology has been used to assess toxin-induced gene expression abnormalities in cancer biology (Clarke et al., 2001; Nuwaysir et al., 1999), and hepatotoxicity (Baker et al., 2001; Gerhold et al., 2001; Bartosiewicz et al., 2001). It has also been widely used in drug abuse studies. For example, microarray technology was instrumental in identifying a diverse number of genes that contribute to methamphetamine-induced dopaminergic neurotoxicity in the ventral mid-brain (Barrett et al., 2001; Xie et al., 2002). These include genes associated with energy metabolism (cytochrome c oxidase subunit 1, reduced nicotinamide adenine dinucleotide ubiquinone oxidoreductase chain 2, and phosphoglycerate mutase B), ion regulation (members of sodium/hydrogen exchanger and sodium/bile acid cotransporter family), signal transduction (adenylyl cyclase III), and cell differentiation and degeneration (*N*-myc downstream-regulated gene 3 and tau protein), as well as free radicals and transcription factors (Cadet et al., 2001).

Dopaminergic neurodegeneration associated with the exposure to the heroin analog designer drug *N*-methyl-4-phenyl-1,2,3,6-tetrahydropyridine (MPTP) is associated with aberrant expression of genes regulating oxidative stress (oxidative stress-induced protein A 170, cytochrome P450 1A1 and Osp94), inflammation (cytotoxic cytokines such as IL-1, IL-6, TNF-α), protective cytokines (IL-10), glutamate receptors (NMDA but not AMPA receptors), neurotrophic factors (GDNF, EGF), nitric oxide synthase and transferrin receptor (Mandel et al., 2000), as well as activation of JNK1/2 MAP kinases followed by induction of apoptosis regulating genes, such as caspase-1 and caspase-3 (Chun et al., 2001). Myelin-related genes have been shown to be down-regulated in chronic alcoholism consistent with the reduction in cerebral white matter in alcoholics (Lewohl et al., 2001). A number of distinct cocaine exposure models, from our laboratories, have illuminated a constellation of genes noteworthy for their regulatory roles (Freeman et al., 2001a,b; 2002a,b). Specifically, members of key signal transduction pathways are altered by cocaine exposure (e.g. PYK2, MEK1, PKA, PKC, etc.). Interestingly, these genes are not universally regulated. That is, different genes are induced (or reduced) depending on the brain region, the animal model, or the mode of cocaine administration. The tremendous throughput power of the DNA array will allow us to accumulate

the large amounts of information necessary to make sense of these rather disparate patterns of gene expression in response to the same drug.

There are numerous neurotoxicological studies where total RNA is extracted from a single- or multiple-brain regions, or from cells cultured in vitro, followed by routine Northern blot preparation with standard protocols of a relatively small number of genes. However, surprisingly, wide-scale gene expression studies with DNA arrays have not been routinely used in neurotoxicology, and a search of the literature (Medline, 1996–2002) identifies only a handful of such studies. One such study was recently conducted in immortalized human fetal astrocytes in which lead (Pb) was shown by microarray technology to induce vascular endothelial growth factor (VEGF) expression via a PKC/transcription factor AP-1-dependent and hypoxia-inducible factor 1 (HIF-1)-independent signaling pathway (Hossain et al., 2000). Our own studies have used macroarrays to identify altered gene expression following dopamine-induced neurotoxicity (Stokes et al., 2002). In particular, an unbiased DNA array screen highlighted two growth arrest/DNA damage-associated genes (GADD45 and GADD153) as early response factors in this form of neurotoxicity. The results are noteworthy in that while the GADD genes have previously been associated with UV light-induced cell toxicity, this is the first time they have been linked to neuronal cell death. These studies document the feasibility and usefulness of DNA array techniques in studying sequential changes of distinctive gene expression patterns in the brain as a function of treatment, and lend credence to the utility of the technique in studies on molecular mechanisms that mediate progression of neurotoxic-induced degeneration associated with a host of compounds (metals, pesticides, food-derived toxins, etc.). In particular, these types of experiments will provide two types of gene discovery or hypothesis-generation approaches for the neurotoxicology field. First, the array-based screening of ESTs will provide important new information on these anonymous expressed genes. In addition, array studies with known genes will ascribe new functions to previously described genes and provide new insights into existing metabolic pathways. Furthermore, a compendium of global gene expression measurements from DNA microarray analysis of the CNS can be profitably used to identify gene expression signatures defining common genes that mediate neurotoxic injury, as well as molecular fingerprinting associated with specific classes of compounds (organophosphates, metals, etc.).

THE FUTURE OF MICROARRAYS IN NEUROTOXICOLOGY

Microarrays offer great promise to field of neurotoxicology. The greatest technical challenges for the future are to improve the reproducibility of array studies and for the field to develop standard methods of data analysis, presentation, and archiving. The most difficult scientific challenge ahead is to combine these large-scale gene expression study results with other experimental results and the existing literature in meaningful ways. Currently, microarray studies often result in long lists and tables. For this research to be truly powerful, this mass of data must eventually take the form of new principles and pathways that better describe the effect of toxins on the central nervous system and possible therapeutic interventions.

ACKNOWLEDGEMENTS

This review was supported in part by Public Health Service Grants from the National Institutes of Health, NIDA 06634 and 13770 (to KEV), NIEHS 07331 and 10563 (to MA), and NIDA 07246 and 07262 (to WMF).

REFERENCES

Alon U, Barkai N, Notterman DA, Gish K, Ybarra S, Mack D. Broad patterns of gene expression revealed by clustering analysis of tumor and normal colon tissues probed by oligonucleotide arrays. Proc Natl Acad Sci USA 1999;96:6745–50.

Anderson L, Seilhamer J. A comparison of selected mRNA and protein abundances in human liver. Electrophoresis 1997;18:533–7.

Baker TK, Carfagna MA, Gao H, Dow ER, Li Q, Searfoss GH et al. Temporal gene expression analysis of monolayer cultured rat hepatocytes. Chem Res Toxicol 2001;14:1218–31.

Barrett T, Xie T, Piao Y, Dillon-Carter O, Kargul GJ, Lim MK et al. A murine dopamine neuron-specific cDNA library and microarray: increased COX1 expression during methamphetamine neurotoxicity. Neurobiol Dis 2001;8:822–33.

Bartosiewicz M, Penn S, Buckpitt A. Applications of gene arrays in environmental toxicology: fingerprints of gene regulation associated with cadmium chloride, benzo(a)pyrene, and trichloroethylene. Environ Health Persp 2001;109:71–742.

Beer DG, Kardia SL, Huang CC, Giordano TJ, Levin AM, Misek DE et al. Gene-expression profiles predict survival of patients with lung adenocarcinoma. Nat Med 2002;8:816–24.

Ben Dor A, Shamir R, Yakhini Z. Clustering gene expression patterns. J Comput Biol 1999;6:281–97.

Brazma A, Vilo J. Gene expression data analysis. FEBS Lett 2000;480:17–24.

Brazma A, Hingamp P, Quackenbush J, Sherlock G, Spellman P, Stoeckert C et al. Minimum information about a microarray experiment (MIAME)-toward standards for microarray data. Nat Genet 2001;29:365–71.

Brenner M, Johnson AB, Boespflug-Tanguy O, Rodriguez D, Goldman JE, Messing A. Mutations in GFAP, encoding glial fibrillary acidic protein, are associated with Alexander disease. Nat Genet 2001;27:117–20.

Cadet JL, Jayanthi S, McCoy MT, Vawter M, Ladenheim B. Temporal profiling of methamphetamine-induced changes in gene expression in the mouse brain: evidence from cDNA array. Synapse 2001;41:40–8.

Chun HS, Gibson GE, DeGiorgio LA, Zhang H, Kidd VJ, Son JH. Dopaminergic cell death induced by MPP(+), oxidant and specific neurotoxicants shares the common molecular mechanism. J Neurochem 2001;76:1010–21.

Clarke PA, te Poele R, Wooster R, Workman P. Gene expression microarray analysis in cancer biology, pharmacology, and drug development: progress and potential. Biochem Pharmacol 2001;62:1311–36.

Conrads TP, Issaq HJ, Veenstra TD. New tools for quantitative phosphoproteome analysis. Biochem Biophys Res Commun 2002;290:885–90.

Edgar R, Domrachev M, Lash AE. Gene expression omnibus: NCBI gene expression and hybridization array data repository. Nucleic Acids Res 2002;30:207–10.

Freeman WM, Brebner K, Lynch WJ, Robertson DJ, Roberts DCS, Vrana KE. Cocaine-responsive gene expression changes in rat hippocampus. Neuroscience 2001a;108:371–80.

Freeman WM, Nader MA, Nader SH, Robertson DJ, Gioia L, Mitchell SM et al. Chronic cocaine-mediated changes in non-human primate nucleus accumbens gene expression. J Neurochem 2001;77:542–849.

Freeman WM, Brebner K, Lynch WJ, Robertson DJ, Patel KM, Roberts DCS et al. Changes in rat frontal cortex gene expression following chronic cocaine. Mol Brain Res 2002a;104:11–20.

Freeman WM, Brebner K, Patel KM, Lynch WJ, Roberts DCS, Vrana KE. Repeated cocaine self-administration causes multiple changes in rat frontal cortex gene expression. Neurochem Res 2002b;27:1181–92.

Gerhold D, Lu M, Xu J, Austin C, Caskey CT, Rushmore T. Monitoring expression of genes involved in drug metabolism and toxicology using DNA microarrays. Physiol Genomics 2001;5:161–70.

Hess KR, Zhang W, Baggerly KA, Stivers DN, Coombes KR. Microarrays: handling the deluge of data and extracting reliable information. Trends Biotechnol 2001;19:463–8.

Hossain MA, Bouton CM, Pevsner J, Laterra J. Induction of vascular endothelial growth factor in human astrocytes by lead. Involvement of a protein kinase C/activator protein-1 complex-dependent and hypoxia-inducible factor 1-independent signaling pathway. J Biol Chem 2000;275:27874–82.

Krasnova IN, McCoy MT, Ladenheim B, Cadet JL. cDNA array analysis of gene expression profiles in the striata of wild-type and Cu/Zn superoxide dismutase transgenic mice treated with neurotoxic doses of amphetamine. FASEB J 2002;16: 1379–88.

Lewohl JM, Dodd PR, Mayfield RD, Harris RA. Application of DNA microarrays to study human alcoholism. J Biomed Sci 2001;8:28–36.

Lockhart DJ, Barlow C. Expressing what's on your mind: DNA arrays and the brain. Nat Rev Neurosci 2001;2:63–8.

Lee PD, Sladek R, Greenwood CM, Hudson TJ. Control genes and variability: absence of ubiquitous reference transcripts in diverse mammalian expression studies. Genome Res 2002;12: 292–7.

Mandel S, Grunblatt E, Youdim M. cDNA microarray to study gene expression of dopaminergic neurodegeneration and neuroprotection in MPTP and 6-hydroxydopamine models: implications for idiopathic Parkinson's disease. J Neural Transm 2000;60:117–24.

Miles MF. Microarrays: lost in a storm of data? Nat Rev Neurosci 2001;2:441–3.

Norenberg MD. Hepatic encephalopathy. In: Kettenmann H, Ransom BR, editors. Neuroglia. New York: Oxford University Press; 1995.p. 950–63.

Nuwaysir EF, Bittner M, Trent J, Barrett JC, Afshari CA. Microarrays and toxicology: the advent of toxicogenomics. Mol Carcinogen 1999;24:153–9.

Raychaudhuri S, Stuart JM, Altman RB. Principal components analysis to summarize microarray experiments: application to sporulation time series. Pac Symp Biocomput 2000;5:452–63.

Sagara J, Miura K, Bannai S. Maintenance of neuronal glutathione by glial cells. J Neurochem 1993;61:1672–6.

Sonnewald U, Qu H, Aschner M. Pharmacology and toxicology of astrocyte-neuron glutamate. J Pharmacol Exp Ther 2002;301: 1–6.

Stokes AH, Freeman WM, Mitchell SG, Burnette TA, Hellman GE, Vrana KE. Induction of GADD45 and GADD153 in neuroblastoma cells by dopamine-induced toxicity. Toxicology 2002;23:675–84.

Tamayo P, Slonim D, Mesirov J, Zhu Q, Kitareewan S, Dmitrovsky E et al. Interpreting patterns of gene expression with self-organizing maps: methods and application to hematopoietic differentiation. Proc Natl Acad Sci USA 1999;96:2907–12.

Tusher VG, Tibshirani R, Chu G. Significance analysis of microarrays applied to the ionizing radiation response. Proc Natl Acad Sci USA 2001;98:5116–21.

Vesce S, Bezzi P, Volterra A. The active role of astrocytes in synaptic transmission. Cell Mol Life Sci 1999;56:991–1000.

Xie T, Tong L, Barrett T, Yuan J, Hatzidimitriou G, McCann UD et al. Changes in gene expression linked to methamphetamine-induced dopaminergic neurotoxicity. J Neurosci 2002;22:274–83.

Yang YH, Dudoit S, Luu P, Lin DM, Peng V, Ngai J et al. Normalization for cDNA microarray data: a robust composite method addressing single and multiple slide systematic variation. Nucleic Acids Res 2002;30:e15.

Zhang L, Wang L, Ravindranathan A, Miles MF. A new algorithm for analysis of oligonucleotide arrays: application to expression profiling in mouse brain regions. J Mol Biol 2002;317:225–35.

Zhou Y, Abagyan R. Match-only integral distribution (MOID) algorithm for high-density oligonucleotide array analysis. BMC Bioinformatics 2002;3:3.