ELSEVIER

Journal of Pharmacological and Toxicological Methods 44 (2000) 291–300

Journal of Pharmacological and Toxicological Methods

Genomics and proteomics
The new millennium of drug discovery and development

Mary Jane Cunningham*

Genometrix, Inc., 2700 Research Forest Drive, The Woodlands, TX 77381, USA

Received 18 September 2000; accepted 25 September 2000

Abstract

One of the most pressing issues facing the pharmaceutical and biotechnology industry is the tremendous dropout rate of lead drug candidates. Over the last two decades, several new genomic technologies have been developed in hopes of addressing the issues of target identification and lead candidate optimization. Gene expression microarray is one of these technologies and this review describes the four main formats, which are currently available: (a) cDNA; (b) oligonucleotide; (c) electrokinetic; and (d) fiberoptic. Many of these formats have been developed with the goal of screening large numbers of genes. Recently, a high-throughput array format has been developed where a large number of samples can be assayed using arrays in parallel. In addition, focusing on gene expression may be only one avenue in preventing lead candidate failure. Proteomics or the study of protein expression may also play a role. Two-dimensional polyacrylamide gel electrophoresis (2-DE) coupled with mass spectroscopy has been the most widely accepted format to study protein expression. However, protein microarrays are now being developed and modified to a high-throughput screening format. Examples of several gene and protein expression studies as they apply to drug discovery and development are reviewed. These studies often result in large data sets. Examples of how several statistical methods (principal components analysis [PCA], clustering methods, Shannon entropy, etc.) have been applied to these data sets are also described. These newer genomic and proteomic technologies and their analysis and visualization methods have the potential to make the drug discovery and development process less costly and more efficient by aiding to select better target and lead candidates.

Keywords: Microarray; Genomics; Proteomics; Toxicology; Drug development

Compound discovery and development is an intense and lengthy process. For the pharmaceutical industry, the number of years to bring a drug from discovery to market is approximately 15 years, costing up to US$500 million per individual drug (Brennan, 2000; DiMasi, 1995). The percentage of compounds, which fail or drop out of the process, is extremely high, over 99%. For every 5000 chemicals evaluated as part of discovery and preclinical testing, only 5 are allowed to proceed to human clinical trials and of these 5, only 1 is approved for the market. A total of 40% of the compounds fail due to poor pharmacokinetics and 11% due to preclinical toxicity.

To address these issues, new genomic and proteomic technologies have been developed over the last several years. These methods are aimed at (a) discovering new genes and proteins and (b) quantifying and analyzing gene and protein expression and (c) assigning functionality. Being able to compare levels of gene and protein expression between diseased and normal cells or cells treated with compounds, which vary in their efficacy and toxicity, could prove valuable in (a) identifying new drug targets and (b) optimizing the choice of lead compound candidates by more closely predicting their success or failure.

1. Genomic technologies

Genomic technologies comprise a wide-ranging group. One of the oldest methods for monitoring gene expression is by Northern blot analysis (Thomas, 1980). Newer direct methods include: (a) gene reporter assays (Lee, M. J., et al., 1997); (b) branched DNA amplification assay (Shen et al.,

* Tel.: +1-281-465-5000.
E-mail address: mcunningham@genometrix.com (M.J. Cunningham).

1056-8719/00/$ – see front matter
PII: S1056-8719(00)00111-8

1998; Todd et al., 1999); (c) PCR-based assays, including quantitative (Wang, A. M., et al., 1989), real-time (Higuchi et al., 1992, 1993), competitive (Gilliland et al., 1990), representational differences analysis (Hubank & Schatz, 1994; Lisitsyn et al., 1993), and differential display (Liang & Pardee, 1992; Martin et al., 1998); (d) scintillation proximity assay (Bosworth & Towers, 1989; Harris et al., 1996); (e) rapid analysis of gene expression (RAGE) (Wang, A., et al., 1999); (f) serial analysis of gene expression (SAGE) (Datson et al., 1999; Velculescu et al., 1995); and (g) microarrays. An indirect method consists of calculating mRNA abundance levels from expressed sequence tags (ESTs) derived from cDNA libraries (Adams et al., 1995; Lee, N. H., 1995; Okubo et al., 1992). All of these methodologies vary in the requirements necessary to perform the assay: amount of starting material, sample preparation, time to perform the assay, equipment, cost, sensitivity, specificity, reproducibility, data analysis, and final interpretation of the results. The advantages and disadvantages of some of these techniques have already been reviewed (Robinson et al., 2000).

2. Gene expression microarrays

This review will concentrate on gene expression microarrays. Microarrays have the advantage of being able to compare expression of up to 12,000 genes at a time. They are also amendable to an automated, high-throughput format, which would include all steps from sample receipt and processing to manufacturing of the arrays to analyzing the expression data. The platform is versatile and flexible; genes from any species can be arrayed and studied. Some formats require very little sample material and, as a result, new methods have been developed to amplify very small samples in order to obtain enough starting material (Phillips & Eberwine, 1996; Zhang, L., et al., 1992). In addition, microarrays, depending on the format used, have comparable sensitivity, specificity, and reproducibility to the other genomic methods listed above.

The need for screening large numbers of genes at a time arose in the 1980s. One attempt to do this was work published by Augenlicht et al. (1987). A reference cDNA library was prepared from polyA mRNA of the human colon carcinoma cell line, HT-29. The cDNA was inserted into a bacterial plasmid and over 4000 clones were isolated. These clones were replicated onto several nitrocellulose filters in the form of a grid. Radiolabeled cDNA probes from several biopsy samples (ranging from patients at low risk for colon cancer to familial adenomatous polyposis [FAP] patients to colon cancer patients) were then hybridized to the filters. The amount of radiolabel was scanned and analyzed. From this experiment, the investigators were able to visualize expression changes over a wide range of genes. A total of 2% of the genes were upregulated and 2% were downregulated in the FAP biopsies compared to the low-risk biopsies. In contrast, 20% of the FAP biopsies (in which the cells had not yet accumulated into adenomas) were upregulated and 5% were downregulated compared to the low-risk biopsies. These results suggested increased gene expression seemed to correlate with early stages of the disease. The identity of the gridded cDNAs was not necessarily known at the time of the production of the filters, but could be readily sequenced and identified subsequent to the hybridization step.

3. Sequencing by hybridization

Concurrently, various formats were being designed to improve the efficiency of DNA sequencing. These formats were collectively referred to as "sequencing by hybridization" or SBH (Bains & Smith, 1988; Drmanac et al., 1989; Fodor et al., 1991; Khrapko et al., 1989, 1991; Pease et al., 1994; Southern et al., 1992; Strezoska et al., 1991; Zhang, Y., et al., 1991). Permutations of oligonucleotides varying in sequence were designed and arrayed. Sample DNA or RNA was then isolated and hybridized onto the arrays. The resulting signal pattern from the hybridization could then be analyzed to determine the sequence of the sample. Two approaches were used: oligonucleotides were either covalently attached to the surface or were photosynthetically synthesized onto the surface. Most of these approaches used membrane substrates to attach the probes. However, Southern et al. used a method whereby oligonucleotide probes were synthesized and covalently attached to a glass surface (Maskos & Southern, 1992a, 1992b; Southern et al., 1992). Subsequently, the size of the arrays increased so that they contained over 1000 oligonucleotides (Maskos & Southern, 1993).

4. cDNA microarrays

From this foundation, several gene expression microarray formats have evolved: (a) cDNA arrays; (b) oligonucleotide arrays; (c) electrokinetic arrays; and (d) fiberoptic arrays. cDNA arrays were first developed by Schena et al. (1995) and Shalon et al. (1996) in P. Brown's and R. Davis' laboratories at Stanford. The initial array was made by amplifying approximately 50 cDNAs of *Arabidopsis thaliana* by PCR and printing them onto a glass microscope slide with a high-speed arrayer. The cDNAs ranged from 250 to 1500 bp in length, averaging around 1000 bp. PolyA mRNA was made from total RNA of *A. thaliana*, reverse-transcribed to cDNA, and labeled fluorescently. cDNA made from wild-type RNA labeled with fluorescein and HAT4-transgenic RNA labeled with lissamine were competitively hybridized onto the slide by mixing equal portions of each probe. The resulting data after scanning showed a 50-fold elevation of signal at the position of HAT4 in the lissamine-specific scan compared to the fluorescein-specific scan.

Further studies used this two-color hybridization format to monitor gene expression changes associated with different yeast metabolic states and to scan expression changes over the whole yeast genome (DeRisi, J. L., et al., 1997; Lashkari et al., 1997a, 1997b). In addition, other early reports showed the use of cDNA microarrays to compare gene expression in cell culture after heat shock or phorbol ester treatment, or with differing disease states, such as rheumatoid arthritis and cancer (DeRisi, J. L., et al., 1996; Heller et al., 1997; Schena et al., 1996). A major commercial vendor of these cDNA glass microarrays is Incyte Genomics.

In addition, a less expensive alternative to the glass cDNA microarrays is arraying cDNA clones onto a membrane substrate and detecting the expression difference by radioactive signal. Lennon and Lebrach (1991) first used this approach; however, being able to quantitate the signal became an early problem. Several studies provided solutions to this problem and current published results show that cDNA clones correlating to mRNA abundance of 0.01% can now be detected (Gress et al., 1992; Nguyen et al., 1995; Zhao et al., 1995). Examples of several commercial sources of the high-density filter arrays are Genome Systems, Clontech, and Research Genetics. Also, arraying equipment can be obtained commercially that enables investigators to manufacture their own microarrays.

5. Oligonucleotide microarrays

The second format of gene expression microarrays is oligonucleotide arrays. Most of the high-density oligonucleotide arrays that are currently in use are produced by the light-directed synthesis of oligomers directly onto the glass array surface (Fodor et al., 1991; McGall et al., 1996; Pease et al., 1994), such as the format used by Affymetrix. Chemical linkers capped with photoremovable protective groups are attached to a solid surface and a photolithographic mask is overlaid. A mercury lamp then activates the protective groups, removing them and leaving a 5′-hydroxy group to which another photoprotected nucleoside can attach. The cycle is then repeated until an oligomer of up to 25 bases is formed (Wodicka et al., 1997). This format allows two methods of study: monitoring gene expression over the entire gene (Chee et al., 1996; Lockhart et al., 1996; Wodicka et al., 1997) or investigating differences between wild-type and aberrant single-nucleotide polymorphisms (SNPs) (Chee et al., 1996; Cronin et al., 1996; Hacia et al., 1996; Kozal et al., 1996). Gene expression values are calculated by summing the signal from all the oligomers tiled onto the array corresponding to the same gene. Fold differences are then calculated from these expression values, thereby allowing comparisons between genes. SNPs are evaluated by the presence or absence of signal from an oligomer. If hybridization to an oligomer occurred, the signal is present; if hybridization did not occur, the signal is absent. By matching the genotype information to the absence or presence of the signal, homozygosity and heterozygosity can be inferred.

Other published approaches have covalently bound the oligonucleotides to a surface. Matson et al. (1995) used an aminated polypropylene surface and produced an array of over 4000 separate cells containing oligomers ranging in size from 8 to 16 mers. Guo et al. (1994) amino-modified both the surface and the oligomer with a phenylisothiocyanate group in between.

6. Electrokinetic microarrays

The third format of gene expression microarrays is an electrokinetic microarray, which was developed by Eggers and Hogan as part of the Genosensor Consortium (Beattie et al., 1993; Eggers & Ehrlich, 1995; Eggers, M. D., et al., 1993, 1994; Lamture et al., 1994). This format uses genosensor technology where synthetic DNA probes are immobilized onto microscopic electronic test biosites arranged in a two-dimensional fashion on a microfabricated device. Sample (or target) DNA binds to these probes by hybridization and can be detected in two different ways. A low-voltage alternating electric field may be used that discriminates the DNA dielectric relaxation frequency of hybridized probes versus the relaxation frequency of nonhybridized probes. Monitoring the signal using a CCD (charge-coupled device) detector was also developed. The CCD detector can monitor a fluorescent, chemiluminescent, or radioisotope label. Genometrix holds an exclusive license to key aspects of this technology and has sublicensed specific fields of use to Motorola.

7. Fiberoptic microarrays

Finally, a microarray format, which is beginning to show promise, utilizes fiberoptic technology (Ferguson et al., 1996; Healey et al., 1997; Michael et al., 1998). A DNA probe is immobilized on one end of a fiber and several fibers are wrapped in a bundle, which is then encased in a stainless steel tube. The other end of the tube is affixed to the imaging system. This microarray format has been shown to detect gene expression signals using cytokine targets and SNPs in the ras oncogene. Recently, a new modification has been published using molecular beacons as the immobilized targets (Steemers et al., 2000). Illumina is commercializing this microarray format.

8. Another gene expression microarray format

Another microarray format that has been described but has had limited commercial acceptability is a microarray consisting of gel pads with immobilized oligonucleotides.

First described by Khrapko et al. in 1989 and 1991, these microarrays have since been miniaturized and have been shown to be useful in discriminating between SNPs in β-thalassemia (Guschin et al., 1997; Khrapko et al., 1989, 1991; Livshits & Mirzabekov, 1996; Yershov et al., 1996).

9. High-throughput gene expression microarrays

Finally, the high-density microarrays may be useful for screening large numbers of genes, discovering new genes, or assigning functionality to unknown genes. However, there are also disadvantages. The high-density arrays can be costly and are not very practical for screening large numbers of samples. This issue has now been addressed with a microarray system of arrays in parallel, developed by Genometrix (Brignac et al., 1999; Eggers, M. D., et al., 2000; Lashkari & Gilmore, 1999; Eggers, M., 2000). The arrays are of low-to-medium density and consist of up to 250 genes. Multiple sets of 96 arrays can be fabricated and hybridized simultaneously. Hundreds to thousands of samples can be processed and analyzed at a time making this format very pertinent to screening samples derived from compound libraries.

10. Applications to drug discovery and development

Gene expression technologies may be useful in several areas of drug discovery and development, such as target identification, lead optimization, and identification of mechanisms of action, to name a few. As an aid in target identification, different approaches have been shown. One approach is to obtain genes of unknown function and analyze them using microarrays. Microarrays allow simultaneous screening of a large number of genes and a comparison of the data can point out possible new disease targets for development. Examples of early papers using microarrays in this fashion have already been cited (Chee et al., 1996; Lashkari et al., 1997b; Schena et al., 1996; Wodicka et al., 1997).

Another approach is to compare, on an extensive scale, differences between expression profiles of normal tissue and diseased tissue. Microarrays have been used in identifying genes, which are up- or downregulated in a variety of diseases. Some examples are studies involving cell lines or tissue from various cancers (alveolar rhabdomyosarcoma, breast, melanoma, ovarian, and prostate) and rheumatoid arthritis (Bittner et al., 2000; DeRisi, J., et al., 1996; Heller et al., 1997; Khan et al., 1998; Perou et al., 1999; Wang, A., et al., 1999; Wang, K., et al., 1999; Xu et al., 2000). Bittner et al. (2000) used cDNA microarrays and statistical methods to observe trends among cutaneous melanoma tissue and cell samples. Initially, there did not appear to be any correlation among the clustered gene expression results and the clinical information. However, after comparing expression information from a tightly clustered group of cutaneous melanoma samples and uveal melanoma samples (which exhibit metastatic tumor properties), a difference was observed. Genes expressed in the uveal melanoma samples strongly did not correlate with the genes expressed in the cutaneous melanoma samples, implying that the genes in the uveal melanoma samples correlate with invasive tumors. The genes in the tight cluster of cutaneous melanoma samples may be indicative of a very low-grade tumor. (Currently, there is no pathological grading system for this particular tumor.) The gene expression profiles may be helpful in setting tumor classifications as well as providing possible new gene targets in cancer.

Karpf et al. (1999) used microarrays to broadly screen for genes expressed with 5-aza-2′-deoxycytidine treatment of a colon adenocarcinoma cell line. The subset of genes induced were further studied and found to be involved in another pathway linked to STAT genes. Other reports have also shown the elucidation of mechanistic pathways through the use of microarrays (Coller et al., 2000; DeRisi, J. L., et al., 1997; Roberts et al., 2000; Voehringer et al., 2000).

Microarrays may also be used in selecting lead compounds. Gray et al. (1998) designed a combinatorial library of protein kinase inhibitors used in cancer therapy and initially screened the compounds through an in vitro activity assay. Next, the authors took three of the compounds and studied their cellular effects in yeast with oligonucleotide microarrays. Diminished activity of one of the compounds in the in vitro assays was also seen in the results from the microarrays.

Another study referenced in a review by Braxton and Bedilion (1998) involved screening several lead compounds for efficacy and toxicity using cDNA microarrays. The result was that the optimized lead compound had a similar profile to a known toxin and ultimately was redesigned to a better lead compound. This latter compound was given highest priority for further development.

Weinstein et al. (1997) and Scherf et al. (2000) screened more than 70,000 compounds against 60 human cancer cell lines. After filtering the data, the authors analyzed the data using three different clustering parameters for comparison: cell line to cell line, cell line to drug activity, and genes to drug activity. In comparing the cell lines to drug activity, several colon and renal cancer cell lines clustered together. These cell lines are known to express multidrug-resistant genes; however, the overall correlation using this clustering parameter was only .21. Several examples of correlations of genes, known to be induced by certain drugs, were also observed.

One of the first reports of using cDNA microarrays to look at toxicity used an in vivo rat model to study three known hepatotoxins: benzo[*a*]pyrene (BP), acetaminophen (APAP) and clofibrate (CLO) (Cunningham et al., 1998, 2000; Fuhrman, Cunningham, Liung, Wen & Somogyi, 2000a; Fuhrman, Cunningham, Wen, Zweiger, Seilhamer, & Somogyi, 2000b). mRNA isolated from livers of Sprague–Dawley rats treated with toxic doses of the three compounds was analyzed using a cDNA microarray containing 7400 rat liver, kidney, and toxicity-related genes. All

three compounds resulted in different expression profiles. Several genes, known to be induced or repressed with APAP, BP, or CLO, were expressed according to the data obtained. Also, all three compounds induced the expression of CYP1A2, CYP2B1, CYP4A3, fatty-acid transport protein 5, a sulfotransferase (ST2A1), and at least two glutathione *s*-transferases (GST-alpha and GST-theta). However, by analyzing the induction pattern over the time points used (12 h to 28 days) in the study, the genes were induced at different times depending on the compound used.

In addition, a comparison of the highest level of differential gene expression with each compound treatment was done using three different methods of analysis: (a) *x*-fold change of gene expression; (b) expectation ratio likelihood (ERL); and (c) Shannon entropy. The latter two methods will be described in greater detail under Data analysis methods. An interesting fact emerged. BP is a genotoxic compound (i.e., acts by a DNA-damaging mechanism), while APAP and CLO are nongenotoxic compounds. In general, APAP and CLO would be expected to have the most overlap in similar genes expressed compared to BP. However, in each of the three analysis methods listed above, there was more overlap between APAP and BP than between either APAP and CLO or BP and CLO. The overlap may be explained by looking at the mechanisms of action. Both APAP and BP have their primary metabolic pathway involving cytochrome P450. The metabolism of CLO involves both cytochrome P450 and β-oxidation.

Another study analyzed the global response of yeast to the DNA damaging agent, methyl methanesulfonate (Jelinsky & Samson, 1999). Of the approximately 6200 genes on the array, 325 genes were increased fourfold upon treatment while 76 genes were decreased. Most of the gene groups, which showed a change in expression, were expected and the expression of a subset of genes was verified by Northern blotting. However, there were also unexpected changes in expression of some individual genes and further study may elucidate the reasons why these genes showed unexpected expression.

Recently, two more studies have been reported using microarrays to further analyze toxicity. Amundson et al. (1999) reported changes in gene expression changes as a result of genotoxic stress and Bartosiewicz et al. (2000) analyzed the expression of β-naphthoflavone (β-NF) in mouse liver. The latter study showed comparable CYP1A2 expression between the microarray format and Northern blotting, 5- and 10-fold, respectively. Also, it was noted that less variability was observed between array spots and between glass slides ($<15\%$) than between the animals used (18–60%).

11. Proteomic technologies

Another field, which is gaining attention, is the area of proteomics. While gene expression may be able to address the pressing issues in pharmaceutical drug discovery and development, protein expression may be equally important. Indeed, an initial paper by L. Anderson and Seilhamer (1997) showed a correlation between RNA and protein abundance in rat liver to be approximately .48. Subsequently, a study by Gygi et al. (1999) showed a correlation between RNA and protein in yeast to range between .1 and .4 for lower-abundance proteins to .94 for only higher-abundance proteins. The authors theorized that if all the yeast proteins were compared or only a random sample, the correlation would be on the order of .4 or less. The explanation for less than a 100% correlation between RNA and protein is probably due, in large part, to post-translational modifications of proteins. Therefore, by examining both RNA and protein, a clearer picture of how genes and proteins interact will arise.

Currently, protein screening and detection is being performed by two-dimensional polyacrylamide gel electrophoresis (2-DE) coupled with annotation of the detected proteins by mass spectroscopy, usually matrix-assisted laser desorption ionization mass spectroscopy (MALDI-MS) (Karas & Hillenkamp, 1988; Klose, 1975; O'Farrell, 1975). This technology platform has been used to acquire genome-wide protein information, such as a survey of proteins in yeast, rat liver, and human plasma (Anderson & Anderson, 1991; Anderson et al., 1991, 1995; Maillet et al., 1996). In addition, there are several reports where investigators have used 2-DE to (a) observe differences between normal and diseased tissues and (b) study differences between different pharmaceutical compounds and environmental agents (Anderson et al., 1996; Cunningham et al., unpublished data; Edvardsson et al., 1999; Giometti et al., 1998; Kanitz et al., 1999; Myers et al., 1997; Williams et al., 1998; Witzmann et al., 1999).

Automated methods have been devised for running several gels in parallel but the major drawback to this method is the time needed for annotating the proteins by mass spectroscopy and comparing sequences back to expressed tag sequence (EST) databases. To solve this problem, newer methods have been developed using protein microarrays. Lueking et al. (1999) described miniature protein arrays made by using crude lysates or purified expressed proteins spotted down on polyvinylidene fluoride (PVDF) filters by a stamp of spring-loaded stainless steel pins. The total area was 25×75 mm and could accommodate 4800 proteins; this area is equivalent to a glass microscope slide. Protein concentrations as low as 10 fmol/μl were detected. Another method, described by Mendoza et al. (1999) and M. Eggers et al. (2000), uses a capillary system to transfer monoclonal antibodies and antigens to a glass surface. One array contains up to 144 individual antibodies and antigens, but with the high-throughput format described, 96 arrays can be printed at one time. This format allows for 96 different samples to be analyzed simultaneously.

12. Data analysis methods

Gene and protein expression platforms generate an abundant amount of data. The requirement of reviewing 10^6 to 10^7 data points from microarray experiments tends to be a necessary part of the experimental design. Several statistical methods and a few newly designed algorithms may prove to be helpful in this regard. Each method takes a unique approach to analyzing the data and which method is chosen may depend on the questions being asked. Some of these methods aid in reducing the large amount of data to a more reasonable data set from which more directed queries are made. Other methods may allow a more directed approach to identify outliers in the data set. These outliers may prove to be potential drug targets or markers of efficacy or toxicity. The ultimate goal is to evaluate the data in the most meaningful manner and to be able to visualize trends for further study.

One of the first reports of applying a statistical algorithm to gene expression data was a paper by Wen et al. (1998). Using RT-PCR, the expression of 122 genes from the developing rat cervical spinal cord was studied. Agglomerative clustering (using the Euclidean distance measure) was applied to the data. Four discrete temporal waves of gene expression were observed in addition to a group of genes, which showed a constant pattern throughout the experiment. This method of clustering allowed hundreds of data points to be reduced to a set of trends, which could be easily visualized and allowed further investigation.

Eisen et al. (1998) developed another clustering approach for the analysis of gene expression data. Hierarchical clustering incorporates an average-linkage method and is visualized by a color scale of gene expression values and a dendrogram showing the similarities between genes. The color scale depicts gene upregulation as red and gene downregulation as green. In this method, genes, which cluster together, have similar function. Using data from stimulating cultured human fibroblasts with serum, genes were found to cluster into at least five functional groups by their expression patterns.

Clustering algorithms answer the question of how similar or dissimilar the data is. The methods can compare data by focusing on comparing genes or samples or both. Another clustering algorithm, which was recently reported by Alon et al. (1999), is a binary or two-way clustering method. This method is a deterministic-annealing algorithm, which compiled the data into a binary tree, and was used to compare normal and cancerous tissue samples between individuals.

Two other methods, which are helpful in reducing data sets and identifying outliers, are principal components analysis (PCA) and Shannon entropy. In a paper by Hilsenbeck et al. (1999), PCA was applied to gene expression data obtained from comparing estrogen-stimulated tumors, tamoxifen-sensitive tumors, and tamoxifen-resistant tumors in mice. A total of 15 tumor samples were hybridized to an array of about 600 genes spotted in duplicate, resulting in a total of 9000 data points. Using PCA, the data set was reduced to a set of principal components. These components identified two genes as outliers, erk2 and HSF-1. HSF-1 (known to be induced by tamoxifen) and erk2 (known to be activated by the estrogen receptor) were both found to be increased in tamoxifen-sensitive tumors by both the gene expression results as well as results from Western blotting. These genes may be potential drug targets for further study.

Shannon entropy is an older statistical method, which was used by Fuhrman, Cunningham, Liang et al., 2000a; Fuhrman, Cunningham, Wen et al., 2000b; and Cunningham et al. (2000) to analyze microarray data. The method is a measure of the information content from a series of events or dynamic pattern, such as a time course. The greater the change in pattern, the higher the entropy value. Genes with the highest entropy values over the time course are genes with the most dynamic expression patterns. For the genes significantly expressed from rat liver samples treated with APAP, BP, or CLO, only 5–8% of the total genes assayed had the highest entropy value. These genes may be potential toxicity markers.

Finally, several other analysis methods have been used to analyze gene expression data: (a) ERL; (b) self-organizing maps; (c) support vector machines; and (d) k-means (Brown et al., 2000; Cunningham et al., 2000; Fuhrman, Cunningham, Liang et al., 2000a; Fuhrman, Cunningham, Wen et al., 2000b; Golub et al., 1999; Tamayo et al., 1999; Tavazoie et al., 1999; Toronen et al., 1999). In addition, methods for visualizing the data are numerous, but the end objective is to summarize the large data sets as trends. The trends themselves may not be conclusive but can provide valuable information to refine hypotheses and further analyze subsets of data.

13. The new millennium for genomics and proteomics

The discoveries and breakthroughs in genomics and proteomics in the last decade will no doubt issue in new and exciting information and advances in this era. Gene expression microarrays are now being used as test cases for speeding up the drug discovery and development process. Newer forms of arrays, such as the high-throughput arrays in parallel and fiberoptic microarrays, may provide expression information in an efficient and cost-effective manner. Protein microarrays have the potential to do likewise. Gene and protein expression data can be analyzed faster by applying known and new statistical algorithms and visualization techniques. All of these methods could speed up the drug discovery process and/or reduce lead compound failures by providing more predictive information.

References

Adams, M. D., Kerlavage, A. R., Fleischmann, R. D., Fuldner, R. A., Bult, C. J., Lee, N. H., Kirkness, E. F., Weinstock, K. G., Gocayne, J. D., White, O., Sutton, G., Blake, J. A., Brandon, R. C., Chiu, M.-W.,

Clayton, R. A., Cline, R. T., Cotton, M. D., Earle-Hughes, J., Fine, L. D., FitzGerald, L. M., FitzHugh, W. M., Fritchman, J. L., Geoghagen, N. S. M., Glodek, A., Gnehm, C. L., Hanna, M. C., Hedblom, E., Hinkle, P. S., Kelley, J. M., Klimek, K. M., Kelley J. C., Liu, L.-I., Marmaros, S. M., Merrick, J. M., Moreno-Palanques, R. F., McDonald, L. A., Nguyen, D. T., Pellegrino, S. M., Phillips, C. A., Ryder, S. E., Scott, J. L., Saudek, D. M., Shirley, R., Small, K. V., Spriggs, T. A., Utterback, T. R., Weidman, J. F., Li, Y., Barthlow, R., Bednarik, D. P., Cao, L., Cepeda, M. A., Coleman, T. A., Collins, E.-J., Dimke, D., Feng, P., Ferrie, A., Fischer, C., Hastings, G. A., He, W.-W., Hu, J.-S., Huddleston, K. A., Greene, J. M., Gruber, J., Hudson, P., Kim, A., Kozak, D. L., Kunsch, C., Ji, H., Li, H., Meissner, P. S., Olsen, H., Raymond, L., Wei, Y.-F., Wing, J., Xu, C., Yu, G.-L., Ruben, S. M., Dillon, P. J., Fannon, M. R., Rosen, C. A., Haseltine, W. A., Fields, C., Fraser, C.M., & Venter, J. C. (1995). Initial assessment of human gene diversity and expression patterns based upon 83 million nucleotides of cDNA sequence. *Nature, 377* (Suppl. 6547), 3–174.

Alon, U., Barkai, N., Notterman, D. A., Gish, K., Ybarra, S., Mack, D., & Levine, A. J. (1999). Broad patterns of gene expression revealed by clustering analysis of tumor and normal colon tissues probed by oligonucleotide arrays. *Proceedings of the National Academy of Sciences*, *92* (12), 6745–6750.

Amundson, S. A., Bittner, M., Chin, Y., Trent, J., Meltzer, P., & Fornace, A. J. (1999). Fluorescent cDNA microarray hybridization reveals complexity and heterogeneity of cellular genotoxic stress responses. *Oncogene*, *18* (24), 3666–3672.

Anderson, N. L., & Seilhamer, J. (1997). A comparison of selected mRNA and protein abundances in human liver. *Electrophoresis*, *18* (3–4), 533–537.

Anderson, N. L., & Anderson, N. G. (1991). A two-dimensional gel database of human plasma proteins. *Electrophoresis*, *12* (11), 883–906.

Anderson, N. L., Esquer-Blasco, R., Hofmann, J. P., & Anderson, N. G. (1991). A two-dimensional gel database of rat liver proteins useful in gene regulation and drug effect studies. *Electrophoresis*, *12* (11), 907–930.

Anderson, N. L., Esquer-Blasco, R., Hofmann, J. P., Meheus, L., Raymackers, J., Steiner, S., Witzmann, F., & Anderson, N. G. (1995). An updated two-dimensional gel database of rat liver proteins useful in gene regulation and drug effect studies. *Electrophoresis*, *16* (10), 1977–1981.

Anderson, N. L., Esquer-Blasco, R., Richardson, F., Foxworthy, P., & Eacho, P. (1996). The effects of peroxisome proliferators on protein abundances in mouse liver. *Toxicology and Applied Pharmacology*, *137* (1), 75–89.

Augenlicht, L. H., Wahrman, M. Z., Halsey, H., Anderson, L., Taylor, J., & Lipkin, M. (1987). Expression of cloned sequences in biopsies of human colonic tissue and in colonic carcinoma cells induced to differentiate in vitro. *Cancer Research*, *47* (22), 6017–6021.

Bains, W., & Smith, G. C. (1988). A novel method for nucleic acid sequence determination. *Journal of Theoretical Biology*, *135* (3), 303–307.

Bartosiewicz, M., Trounstine, M., Barker, D., Johnston, R., & Buckpitt, A. (2000). Development of a toxicological gene array and quantitative assessment of this technology. *Archives of Biochemistry and Biophysics*, *376 (1), 66–73.*

Beattie, K., Eggers, M., & Shumaker, J. (1993). Genosensor technology. *Clinical Chemistry*, *39* (4), 719–722.

Bittner, M., Meltzer, P., Chen. Y., Jiang, Y., Seftor, E., Hendrix, M., Radmacher, M., Simon, R., Yakhini, Z., Ben-Dor, A., Sampas, N., Dougherty, E., Wang, E., Marincola, F., Gooden, C., Lueders, J., Glatfelter, A., Pollock, P., Carpten, J., Gillanders, E., Leja, D., Dietrich, K., Beaudry, C., Berens, M., Alberts, D., Sondak, V., Hayward, N., & Trent, J. (2000) Molecular classification of cutaneous malignant melanoma by gene expression profiling. *Nature, 406* (6795), 536–540.

Bosworth, N., & Towers, P. (1989). Scintillation proximity assay. *Nature*, *341* (6238), 137–168.

Braxton, S., & Bedilion, T. (1998). The integration of microarray information in the drug development process. *Current Opinion in Biotechnology*, *9* (6), 643–649.

Brennan, M. B. (2000). Drug discovery: filtering out failures early in the game. *Chemical & Engineering News*, *78* (23), 63–73.

Brignac, S. J., Gangadharan, R., McMahon, M., Denman, J., Gonzales, R., Mendoza, L. G., & Eggers, M. (1999). A proximal CCD imaging system for high-throughput detection of microarray-based assays. *IEEE Engineering in Medicine and Biology Magazine*, *18* (2), 120–122.

Brown, M. P. S., Grundy, W. N., Lin, D., Cristianini, N., Sugnet, C. W., Furey, T. S., Ares, M., & Haussler, D. (2000). Knowledge-based analysis of microarray gene expression data by using support vector machines. *Proceedings of the National Academy of Sciences USA*, *97* (1), 262–267.

Chee, M., Yang, R., Hubbell, E., Berno, A., Huang, X. C., Stern, D., Winkler, J., Lockhart, D. J., Morris, M. S., & Fodor, S. P. A. (1996). Accessing genetic information with high-density DNA arrays. *Science*, *274* (5287), 610–614.

Coller, H. A., Grandori, C., Tamayo, P., Colbert, T., Lander, E. S., Eisenman, R. N., & Golub, T. R. (2000). Expression analysis with oligonucleotide microarrays reveals that MYC regulates genes involved in growth, cell cycle, signaling and adhesion. *Proceedings of the National Academy of Sciences USA*, *97* (7), 3260–3265.

Cronin, M. T., Fucini, R. V., Kim, S. M., Masino, R. S., Wespi, R. M., & Miyada, C. G. (1996). Cystic fibrosis mutation detection by hybridization to light-generated DNA probe arrays. *Human Mutation*, *7* (3), 244–255.

Cunningham, M. J., Zweiger, G., Bailey, D., Schwartz, M., Braxton, S., & Seilhamer, J. (1998). Application of a genome-based screen. *Archives of Pharmacology*, *358* (1) (Suppl. 2), R564.

Cunningham, M. J., Liang, S., Fuhrman, S., Seilhamer, J. J., & Somogyi, R. (2000). Gene expression microarray data analysis for toxicity profiling. *Annals of the New York Academy of Sciences, 919*, 52–67.

Datson, N. A., van der Perk-de Jong, J., van den Berg, M. P., de Kloet, E. R., & Vreugdenhil, E. (1999). MicroSAGE: a modified procedure for serial analysis of gene expression in limited amounts of tissue. *Nucleic Acids Research*, *27* (5), 1300–1307.

DeRisi, J. L., Penland, L., Brown, P. O., Bittner, M. L., Meltzer, P. S., Ray, Y., Chen, Y., Su, Y. A., & Trent, J. M. (1996). Use of a cDNA microarray to analyse gene expression patterns in human cancer. *Nature Genetics*, *14* (4), 457–460.

DeRisi, J. L., Iyer, V. R., & Brown, P. O. (1997). Exploring the metabolic and genetic control of gene expression on a genomic scale. *Science*, *278* (5338), 680–686.

DiMasi, J. A. (1995). Trends in drug development costs, times, and risks. *Drug Information Journal*, *29* (2), 375–384.

Drmanac, R., Labat, I., Brukner, I., & Crkvenjakov, R. (1989). Sequencing of megabase plus DNA by hybridization: theory of the method. *Genomics*, *4* (2), 114–128.

Edvardsson, U., Alexandersson, M., Brockenhuus von Lowenhielm, H., Nystrom, A. C., Ljung, B., Nilsson, F., & Dahllof, B. (1999). A proteome analysis of livers from obese (*ob*/*ob*) mice treated with the peroxisome proliferator WY14,643. *Electrophoresis*, *20* (4–5), 935–942.

Eggers, M. D., Hogan, M., Reich, R. K., Lamture, J., Ehrlich, D., Hollis, M., Kosicki, B., Powdrill, T., Beattie, K., Smith, S., Varma, R., Gangadharan, R., Mallik, A., Burke, B., & Wallace, D. (1994). A microchip for quantitative detection of molecules utilizing luminescent and radioisotope reporter groups. *BioTechniques*, *17* (3), 516–523.

Eggers, M. D., Balch, B., Brignac, S., Gilmore, J., Hogan, M., King, T., Lashkari, D., Milosavljevic, A., Powdrill, T., & Smith, A. (2000). High-throughput arrays for efficient screening and analysis. In: L. Kricka & J. Cheng. (Eds.), *Biochip technology* (in press).

Eggers, M. D., & Ehrlich, D. (1995). A review of microfabricated devices for gene-based diagnostics. *Hematology and Pathology*, *9* (1), 1–15.

Eggers, M. D., Hogan, M. E., Reich, R. K., Lamture, J. B., Beattie, K. L., Hollis, M. A., Ehrlich, D. J., Kosicki, B. B., Shumaker, J. M., Varma, R. S., Burke, B. E., Murphy, A., & Rathman, D. D. (1993). Genosensors: microfabricated devices for automated DNA sequence analysis. *Proceedings of the SPIE: Advances in DNA Sequencing Technology*, *1891,* 113–126.

Eisen, M. B., Spellman, P. T., Brown, P. O., & Botstein, D. (1998). Cluster analysis and display of genome-wide expression patterns. *Proceedings of the National Academy of Sciences USA*, *95* (25), 14863–14868.

Ferguson, J. A., Boles, T. C., Adams, C. P., & Walt, D. R. (1996). A fiber-optic DNA biosensor microarray for the analysis of gene expression. *Nature Biotechnology*, *14* (3), 1681–1684.

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* (4995), 767–773.

Fuhrman, S., Cunningham, M. J., Liang, S., Wen, X., & Somogyi, R. (2000a). The application of Shannon entropy in the identification of putative drug targets. *BioSystems*, *55* (1–3), 5–14.

Fuhrman, S., Cunningham, M. J., Wen, X., Zweiger, G., Seilhamer, J. J., & Somogyi, R. (2000b). The application of Shannon entropy in the identification of putative drug targets. *BioSystems*, *55* (1–3), 5–14.

Gilliland, G., Perrin, S., Blanchard, K., & Bunn, H. F. (1990). Analysis of cytokine mRNA and DNA: detection and quantitation by competitive polymerase chain reaction. *Proceedings of the National Academy of Sciences USA*, *87* (7), 2725–2729.

Giometti, C. S., Tollaksen, S. L., Liang, X., & Cunningham, M. L. (1998). A comparison of liver protein changes in mice and hamsters treated with the peroxisome proliferator Wy-14,643. *Electrophoresis*, *19* (14), 2498–2505.

Golub, T. R., Slonim, D. K., Tamayo, P., Huard, C., Gaasenbeek, M., Mesirov, J. P., Coller, H., Loh, M. L., Downing, J. R., Caligiuri, M. D., Bloomfield, C. D., & Lander, E. S. (1999). Molecular classification of cancer: class discovery and class prediction by gene expression monitoring. *Science*, *286* (5439), 531–537.

Gray, N. S., Wodicka, L., Thunnissen, A.-M. W. H., Norman, T. C., Kwon, F. H., Espinoza, F. H., Morgan, D. O., Barnes, G., LeClerc, S., Meijer, S.-H., Kim, S.-H., Lockhart, D. J., & Schultz, P. G. (1998). Exploiting chemical libraries, structure and genomics in the search for kinase inhibitors. *Science*, *281* (5376), 533–538.

Gress, T. M., Hoheisel, J. D., Lennon, G. G., Zehetner, G., & Lehrach, H. (1992). Hybridization fingerprinting of high-density cDNA-library arrays with cDNA pools derived from whole tissues. *Mammalian Genome*, *3* (11), 609–619.

Guo, Z., Guilfoyle, R. A., Thiel, A. J., Wang, R., & Smith, L. M. (1994). Direct fluorescence analysis of genetic polymorphisms by hybridization with oligonucleotide arrays on glass supports. *Nucleic Acids Research*, *22* (24), 5456–5465.

Guschin, D., Yershov, G., Zeslavsky, A., Gemmell, A., Shick, V., Proudnikov, D., Arenkov, P., & Mirzabekov, A. (1997). Manual manufacturing of oligonucleotide, DNA and protein microchips. *Analytical Biochemistry*, *250* (2), 203–211.

Gygi, S. P., Rochon, Y., Franza, B. R., & Aebersold, R. (1999). Correlation between protein and mRNA abundance in yeast. *Molecular and Cellular Biology*, *19* (3), 1720–1730.

Hacia, J. G., Brody, L. C., Chee, M. S., Fodor, S. P. A., & Collins, F. S. (1996). Detection of heterozygous mutations in BRCA1 using high density oligonucleotide arrays and two-colour fluorescence analysis. *Nature Genetics*, *14* (4), 441–447.

Harris, D. W., Kenrick, M. K., Pither, R. J., Anson, J. G., & Jones, D. A. (1996). Development of a high-volume in situ mRNA by hybridization assay for the quantification of gene expression utilizing scintillating microplates. *Analytical Biochemistry*, *243* (2), 249–256.

Healey, B. G., Matson, R. S., & Walt, D. R. (1997). Fiberoptic DNA sensor array capable of detecting point mutations. *Analytical Biochemistry*, *251* (2), 270–279.

Heller, R. A., Schena, M., Chai, A., Shalon, D., Bedilion, T., Gilmore, J., Wooley, D. E., & Davis, R. W. (1997). Discovery and analysis of inflammatory disease-related genes using cDNA microarrays. *Proceedings of the National Academy of Sciences USA*, *94* (6), 2150–2155.

Higuchi, R., Dollinger, G., Walsh, P. S., & Griffith, R. (1992). Simultaneous amplification and detection of specific DNA sequences. *Biotechnology*, *10* (4), 413–417.

Higuchi, R., Fockler, C., Dollinger, G., & Watson, R. (1993). Kinetic PCR analysis: real-time monitoring of DNA amplification reactions. *Biotechnology*, *11* (9), 1026–1030.

Hilsenbeck, S. G., Friedrichs, W. E., Schiff, R., O'Connell, P., Hansen, R. K., Osborne, C. K., & Fuqua, S. A. W. (1999). Statistical analysis of array expression data as applied to the problem of tamoxifen resistance. *Journal of the National Cancer Institute*, *91* (5), 453–459.

Hubank, M., & Schatz, D. G. (1994). Identifying differences in mRNA expression by representational difference analysis of cDNA. *Nucleic Acids Research*, *22* (25), 5640–5648.

Jelinsky, S. A., & Samson, L. D. (1999). Global response of *Saccharomyces cerevisiae* to an alkylating agent. *Proceedings of the National Academy of Sciences*, *96* (4), 1486–1491.

Kanitz, M. H., Witzmann, F. A., Zhu, H., Fultz, C. D., Skaggs, S., Moorman, W. J., & Savage, R. E. (1999). Alterations in rabbit kidney protein expression following lead exposure as analyzed by two-dimensional gel electrophoresis. *Electrophoresis*, *20* (14), 2977–2985.

Karas, M., & Hillenkamp, F. (1988). Laser desorption ionization of proteins with molecular masses exceeding 10,000 daltons. *Analytical Chemistry*, *60* (20), 2299–2301.

Karpf, A. R., Peterson, P. W., Rawlins, J. T., Dalley, B. K., Yang, Q., Albertsen, H., & Jones, D. A. (1999). Inhibition of DNA methyltransferase stimulates the expression of signal transducer and activator of transcription 1, 2, and 3 genes in colon tumor cells. *Proceedings of the National Academy of Sciences USA*, *96* (24), 14007–14012.

Khan, J., Simon, R., Bittner, M., Chen, Y., Leighton, S. B., Pohida, T., Smith, P. D., Jiang, Y., Gooden, G. C., Trent, J. M., & Meltzer, P. S. (1998). Gene expression profiling of alveolar rhabdomyosarcoma with cDNA microarrays. *Cancer Research*, *58* (22), 5009–5013.

Khrapko, K. R., Lysov, Y. P., Khorlyn, A. A., Shick, V. V., Florentiev, V. L., & Mirzabekov, A. D. (1989). An oligonucleotide hybridization approach to DNA sequencing. *FEBS Letters*, *256* (1–2), 118–122.

Khrapko, K. R., Lysov, Y. P., Khorlyn, A. A., Ivanov, I. B., Yershov, G. M., Vasilenko, S. K., Florentiev, V. L., & Mirzabekov, A. D. (1991). A method for DNA sequencing by hybridization with oligonucleotide matrix. *DNA Sequence*, *1* (6), 375–388.

Klose, J. (1975). Protein mapping by combined isoelectric focusing and electrophoresis of mouse tissues. A novel approach to testing for induced point mutations in mammals. *Humangenetik*, *26* (3), 231–243.

Kozal, M. J., Shah, N., Shen, N., Yang, R., Fucini, R., Merigan, T. C., Richman, D. D., Morris, D., Hubbell, E., Chee, M., & Gingeras, T. R. (1996). Extensive polymorphisms observed in HIV-1 clade B protease gene using high-density oligonucleotide arrays. *Nature Medicine*, *2* (7), 753–759.

Lamture, J. B., Beattie, K. L., Burke, B. E., Eggers, M. D., Ehrlich, D. J., Fowler, R., Hollis, M. A., Kosicki, B. B., Reich, R. K., Smith, S. R., Varma, R. S., & Hogan, M. E. (1994). Direct detection of nucleic acid hybridization on the surface of a charge coupled device. *Nucleic Acids Research*, *22* (11), 2121–2125.

Lashkari, D. A., McCusker, J. H., & Davis, R. W. (1997a). Whole genome analysis: experimental access to all genome sequenced segments through larger-scale efficient oligonucleotide synthesis and PCR. *Proceedings of the National Academy of Sciences USA*, *94* (17), 8945–8947.

Lashkari, D. A., DeRisi, J. L., McCusker, J. H., Namath, A. F., Gentile, C., Hwang, S. Y., Brown, P. O., & Davis, R. W. (1997b). Yeast microarrays for genome wide parallel genetic and gene expression analysis. *Proceedings of the National Academy of Sciences*, *94* (24), 13057–13062.

Lashkari, D. A., & Gilmore, J. (1999). High-throughput DNA chips for efficient screening and analysis. *SigBio Newsletter*, *19*, 2–4.

Lee, M. J., Gee, P., & Beard, S. E. (1997). Detection of peroxisome proliferators using a reporter construct derived from the rat acyl-CoA oxidase promoter in the rat liver cell line H-4-II-E. *Cancer Research*, *57* (8), 1575–1579.

Lee, N. H., Weinstock, K. G., Kirkness, E. F., Earle-Hughes, J. A., Fuldner, R. A., Marmaros, S., Glodek, A., Gocayne, J. D., Adams, M. D., Kerlavage, A. R., Fraser, C. M., & Venter, J. C. (1995). Comparative

expressed-sequence-tag analysis of differential gene expression profiles in PC-12 cells before and after nerve growth factor treatment. *Proceedings of the National Academy of Sciences USA*, *92* (18), 8307–8308.

Lennon, G. G., & Lehrach, H. (1991). Hybridization analyses of arrayed cDNA libraries. *Trends in Genetics*, *7* (10), 314–317.

Liang, P., & Pardee, A. B. (1992). Differential display of eukaryotic messenger RNA by means of the polymerase chain reaction. *Science*, *257* (5072), 967–971.

Lisitsyn, N., Lisitsyn, N., & Wigler, M. (1993). Cloning the differences between two complex genomes. *Science*, *259* (5097), 946–951.

Livshits, M. A., & Mirzabekov, A. D. (1996). Theoretical analysis of the kinetics of DNA hybridization with gel-immobilized oligonucleotides. *Biophysical Journal*, *71* (5), 2795–2801.

Lockhart, D. J., Dong, H., Byrne, M. C., Folletie, M. T., Gallo, M. V., Chee, M. S., Mittmann, M., Wang, C., Kobayashi, M., Horton, H., & Brown, E. L. (1996). Expression monitoring by hybridization to high-density oligonucleotide arrays. *Nature Biotechnology*, *14* (13), 1675–1680.

Lueking, A., Horn, M., Eickhoff, H., Bussow, K., Lehrach, H., & Walter, G. (1999). Protein microarrays for gene expression and antibody screening. *Analytical Biochemistry*, *270* (1), 103–111.

Maillet, I., Lagniel, G., Perrot, M., Boucherie, H., & Labarre, J. (1996). Rapid identification of yeast proteins on two-dimensional gels. *Journal of Biological Chemistry*, *271* (17), 10263–10270.

Martin, K. J., Kwan, C.-P., O'Hare, M. J., Pardee, A. B., & Sager, R. (1998). Identification and verification of differential display cDNAs using gene-specific primers and hybridization arrays. *BioTechniques*, *24* (6), 1018–1026.

Maskos, U., & Southern, E. M. (1992a). Parallel analysis of oligodeoxyribonucleotide (oligonucleotide) interactions: I. Analysis of factors influencing oligonucleotide duplex formation. *Nucleic Acids Research*, *20* (7), 1675–1678.

Maskos, U., & Southern, E. M. (1992b). Oligonucleotide hybridizations on glass supports: a novel linker for oligonucleotide synthesis and hybridization properties of oligonucleotides synthesized in situ. *Nucleic Acids Research*, *20* (7), 1679–1684.

Maskos, U., & Southern, E. M. (1993). A study of oligonucleotide reassociation using large arrays of oligonucleotides synthesized on a glass support. *Nucleic Acids Research*, *21* (20), 4663–4669.

Matson, R. S., Rampal, J., Pentoney, S. L., Anderson, P. D., & Coassin, P. (1995). Biopolymer synthesis on polypropylene supports: oligonucleotide arrays. *Analytical Biochemistry*, *224* (1), 110–116.

McGall, G., Labadie, J., Brock, P., Wallraff, G., Nguyen, T., & Hinsberg, W. (1996). Light-directed synthesis of high-density oligonucleotide arrays using semiconductor photoresists. *Proceedings of the National Academy of Sciences USA*, *93* (24), 13555–13560.

Mendoza, L. G., McQuary, P., Mongan, A., Gangadharan, R., Brignac, S., & Eggers, M. (1999). High-throughput microarray-based enzyme-linked immunosorbent assay (ELISA). *BioTechniques*, *27* (4), 778–788.

Michael, K. L., Taylor, L. C., Schultz, S. L., & Walt, D. R. (1998). Randomly ordered addressable high-density optical sensor arrays. *Analytical Chemistry*, *70* (7), 1242–1248.

Myers, T. G., Anderson, N. L., Waltham, M., Li, G., Buolamwini, J. K., Scudiero, D. A., Paull, K. D., Sausville, E. A., & Weinstein, J. N. (1997). A protein expression database for the molecular pharmacology of cancer. *Electrophoresis*, *18* (3–4), 647–653.

Nguyen, C., Rocha, D., Granjeaud, S., Baldit, M., Bernard, K., Naquet, P., & Jordan, B. R. (1995). Differential gene expression in the murine thymus assayed by quantitative hybridization of arrayed cDNA clones. *Genomics*, *29* (1), 207–216.

O'Farrell, P. H. (1975). High resolution two-dimensional electrophoresis of proteins. *Journal of Biological Chemistry*, *250* (10), 4007–4021.

Okubo, K., Hori, N., Matoba, R., Niiyama, T., Fukushima, A., Kojima, Y., & Matsubara, K. (1992). Large scale cDNA sequencing for analysis of quantitative and qualitative aspects of gene expression. *Nature Genetics*, *2 (3), 173–179.*

Pease, A. C., Solas, D., Sullivan, E. J., Cronin, M. T., Holmes, C. P., & Fodor, S. P. A. (1994). Light-generated oligonucleotide arrays for rapid DNA sequence analysis. *Proceedings of the National Academy of Sciences USA*, *91* (11), 5022–5026.

Perou, C. M., Jeffrey, S. S., van de Rijn, M., Rees, C. A., Eisen, M. B., Ross, D. T., Pergamenschikov, A., Williams, C. F., Zhu, S. X., Lee, J. C. F., Lashkari, D., Shalon, D., Brown, P. O., & Botstein, D. (1999). Distinctive gene expression patterns in human mammary epithelial cells and breast cancers. *Proceedings of the National Academy of Sciences USA*, *96* (16), 9212–9217.

Phillips, J., & Eberwine, J. H. (1996). Antisense RNA amplification: a linear amplification method for analyzing the mRNA population from single living cells. *Methods*, *10* (3), 283–288.

Roberts, C. J., Nelson, B., Marton, M. J., Stoughton, R., Meyer, M. R., Bennett, H. A., He, Y. D., Dai, H., Walker, W. L., Hughes, T. R., Tyers, M., Boone, C., & Friend, S. H. (2000). Signaling and circuitry of multiple MAPK pathways revealed by a matrix of global gene expression profiles. *Science*, *287* (5454), 873–880.

Robinson, B. W. S., Erle, D. J., Jones, D. A., Shapiro, S., Metzger, W. J., Albelda, S. M., Parks, W. C., & Boylan, A. (2000). Recent advances in molecular biological techniques and their relevance to pulmonary research. *Thorax*, *55* (4), 329–339.

Schena, M., Shalon, D., Davis, R. W., & Brown, P. O. (1995). Quantitative monitoring of gene expression patterns with a complementary DNA microarray. *Science*, *270* (5235), 467–470.

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. *Proceedings of the National Academy of Sciences USA*, *93* (20), 10614–10619.

Scherf, U., Ross, D. T., Waltham, M., Smith, L. H., Lee, J. K., Tanabe, L., Kohn, K. W., Reinhold, W. C., Myers, T. G., Andrews, D. T., Scudiero, D. A., Eisen, M. B., Sausville, E. A., Pommier, Y., Botstein, D., Brown, P. O., & Weinstein, J. N. (2000). A gene expression database for the molecular pharmacology of cancer. *Nature Genetics*, *24* (3), 236–244.

Shalon, D., Smith, S. J., & Brown, P. O. (1996). A DNA microarray system for analyzing complex DNA samples using two-color fluorescent probe hybridization. *Genome Research*, *6* (7), 639–645.

Shen, L. P., Sherindan, P., Cao, W. W., Dailey, P. J., Salazar-Gonzales, J. F., Breen, E. C., Fahey, J. L., Urdea, M. S., & Kolberg, J. A. (1998). Quantification of cytokine mRNA in peripheral blood mononuclear cells using branched DNA (bDNA) technology. *Journal of Immunological Methods*, *215* (1–2), 123–134.

Southern, E. M., Maskos, U., & Elder, J. K. (1992). Analyzing and comparing nucleic acid sequences by hybridization to arrays of oligonucleotides: evaluation using experimental models. *Genomics*, *13* (4), 1008–1017.

Steemers, F. J., Ferguson, J. A., & Walt, D. R. (2000). Screening unlabeled DNA targets with randomly-ordered fiber-optic gene arrays. *Nature Biotechnology*, *18* (1), 91–94.

Strezoska, Z., Paunesku, T., Radosavljevic, D., Labat, I., Drmanac, R., & Crkvenjakov, R. (1991). DNA sequencing by hybridization: 100 bases read by a non-gel-based method. *Proceedings of the National Academy of Sciences USA*, *88* (22), 10089–10093.

Tamayo, P., Slonim, D., Mesirov, J., Zhu, Q., Kitareewan, S., Dmitrovsky, E., Lander, E. S., & Golub, T. R. (1999). Interpreting patterns of gene expression with self-organizing maps: methods and application to hematopoietic differentiation. *Proceedings of the National Academy of Sciences USA*, *96* (6), 2907–2912.

Tavazoie, S., Hughes, J. D., Campbell, M. J., Cho, R. J., & Church, G. M. (1999). Systematic determination of genetic network architecture. *Nature Genetics*, *22* (3), 281–285.

Thomas, P. S. (1980). Hybridization of denatured RNA and small DNA fragments transferred to nitrocellulose. *Proceedings of the National Academy of Sciences*, *77* (9), 5201–5205.

Todd, M. D., Lin, X., Stankowski, L. F., Desai, M., & Wolfgang, G. H. (1999). Toxicity screening of a combinatorial library: correlation of

cytotoxicity and gene induction to compound structure. *Journal of Biomolecular Screening*, *4* (5), 259–268.

Toronen, P., Kolehmainen, M., Wong, G., & Castren, E. (1999). Analysis of gene expression data using self-organizing maps. *FEBS Letters*, *451* (2), 142–146.

Velculescu, V. E., Zhang, L., Vogelstein, B., & Kinzler, K. W. (1995). Serial analysis of gene expression. *Science*, *270* (5235), 484–487.

Voehringer, D. W., Hirschberg, D. L., Xiao, J., Lu, Q., Roederer, M., Lock, C. B., Herzenberg, L. A., Steinman, L., & Herzenberg, L. A. (2000). Gene microarray identification of redox and mitochondrial elements that control resistance or sensitivity to apoptosis. *Proceedings of the National Academy of Sciences USA*, *97* (6), 2680–2685.

Wang, A., Pierce, A., Judson-Kremer, K., Gaddis, S., Aldaz, C. M., Johnson, D. G., & MacLeod, M. C. (1999). Rapid analysis of gene expression (RAGE) facilitates universal expression profiling. *Nucleic Acids Research*, *27* (23), 4609–4618.

Wang, A. M., Doyle, M. V., & Mark, D. F. (1989). Quantitation of mRNA by the polymerase chain reaction. *Proceedings of the National Academy of Sciences*, *86* (24), 9717–9721.

Wang, K., Gan, L., Jeffery, E., Gayle, M., Gowen, A. M., Skelly, M., Nelson, P. S., Ng, W. V., Schummer, M., Hood, L., & Mulligan, J. (1999). Monitoring gene expression profile changes in ovarian carcinomas using cDNA microarray. *Gene*, *229* (1–2), 101–108.

Weinstein, J. N., Myers, T. G., O'Conner, P. M., Friend, S. H., Fornance, A. J., Kohn, K. W., Fojo, T., Bates, S. E., Rubinstein, L. V., Anderson, N. L., Buolamwini, J. K., van Osdol, W. W., Monks, A. P., Scudiero, D. A., Sauville, E. A., Zaharevitz, D. W., Bunow, B., Viswannadhan, V. N., Johnson, G. S., Wittes, R. E., & Paull, K. D. (1997). An information-intensive approach to the molecular pharmacology of cancer. *Science*, *275* (5298), 343–349.

Wen, X., Fuhrman, S., Michaels, G. S., Carr, D. B., Smith, S., Barker, J. L., & Somogyi, R. (1998). Large-scale temporal gene expression mapping of central nervous system development. *Proceedings of the National Academy of Sciences*, *95* (1), 334–339.

Williams, K., Chubb, C., Huberman, E., & Giometti, C. S. (1998). Analysis of differential protein expression in normal and neoplastic human breast epithelial cell lines. *Electrophoresis*, *19* (2), 333–343.

Witzmann, F. A., Fultz, C. D., Grant, R. A., Wright, L. S., Kornguth, S. E., & Siegel, F. L. (1999). Regional protein alterations in rat kidneys induced by lead exposure. *Electrophoresis*, *20* (4–5), 943–951.

Wodicka, L., Dong, H., Mittmann, M., Ho, M.-H., & Lockhart, D. J. (1997). Genome-wide expression monitoring in *Saccharomyces cerevisiae*. *Nature Biotechnology*, *15* (13), 1359–1367.

Xu, J., Stolk, J. A., Zhang, X., Silva, S. J., Houghton, R. L., Matsumura, M., Vedvick, T. S., Leslie, K. B., Badaro, R., & Reed, S. G. (2000). Identification of differentially expressed genes in human prostate cancer using subtraction and microarray. *Cancer Research*, *60* (6), 1677–1682.

Yershov, G., Bersky, V., Belgovskiy, A., Kirillov, E., Ivanov, I., Parinov, S., Guschin, D., Drobishev, A., Dubiley, S., & Mirzabekov, A. (1996). DNA analysis and diagnostics on oligonucleotide microchips. *Proceedings of the National Academy of Sciences*, *93* (10), 4913–4918.

Zhang, L., Cui, X., Schmitt, K., Hubert, R., Navidi, W., & Arnheim, N. (1992). Whole genome amplification from a single cell: implications for genetic analysis. *Proceedings of the National Academy of Sciences*, *89* (13), 5847–5851.

Zhang, Y., Coyne, M. Y., Levenson, C. H., & Kawasaki, E. S. (1991). Single-base mutational analysis of cancer and genetic diseases using membrane bound modified oligonucleotides. *Nucleic Acids Research*, *19* (14), 3929–3933.

Zhao, N., Hashida, H., Takahashi, N., Misumi, Y., & Sakaki, Y. (1995). High-density cDNA filter analysis: a novel approach for large-scale, quantitative analysis of gene expression. *Gene*, *156* (2), 207–213.