# 1 Microarrays in Systems Neurobiology and Translational Neuroscience – From Genome Research to Clinical Applications

*Jeremy A. Miller and Daniel H. Geschwind*

## 1.1 Introduction

Although microarray technology was introduced just 10 years ago, over 20 000 articles have been published using this technology as of 2006, covering areas ranging from soil ecology and yeast genomics to cancer and neurological disorders. In neuroscience, much of this represents publications since 2000, showing a remarkable trajectory as well as reflecting early skepticism that has now given way to acceptance and appreciation (Figure 1.1). Entire transcriptomes can now be assayed on a single chip at a reasonable cost, and technologies are becoming cheaper and more accurate day by day. In basic neuroscience research, microarrays have been used to assess gene expression differences across mouse strains [1], brain areas [2], cell types [3–5], and brain tumor strains [6]. They have also been used to identify genes that play an important role in neural stem cell biology [7,8], mouse models of neurodevelopmental disorders [9], and postmortem assessments of many neurodegenerative diseases such as Alzheimer's disease (AD) [2,10–15], Parkinson's disease (PD) [16], Huntington's disease (HD) [17,18], amyotrophic lateral sclerosis (ALS) [19,20], and schizophrenia [21]. In addition to providing a useful tool for basic neuroscience research, microarrays hold significant promise clinically as patient classifiers in acute and chronic neurological diseases [22].

This chapter summarizes the current state of microarray technology, presenting several clinical applications. In the next section, gene expression technologies leading up to microarray technologies are presented along with alternative high-throughput techniques. Section 1.3 provides a primer on how to design and implement a successful microarray experiment and presents challenges to the field and analytic methods that have been developed to get the most out of expression data. The last section summarizes recent microarray experiments in the field of neuroscience, highlighting key, representative papers documenting state-of-the-art experimental design, clinical uses in brain cancer, and the use of peripheral blood as a substitute for brain tissue in various neuropsychiatric conditions. Finally, genomic DNA microarrays are briefly discussed, along with speculation on the future of clinical microarray applications.

*Neural Degeneration and Repair: Gene Expression Profiling, Proteomics, and Systems Biology*
Edited by Hans Werner Müller
Copyright © 2008 WILEY-VCH Verlag GmbH & Co. KGaA, Weinheim
ISBN: 978-3-527-31707-3

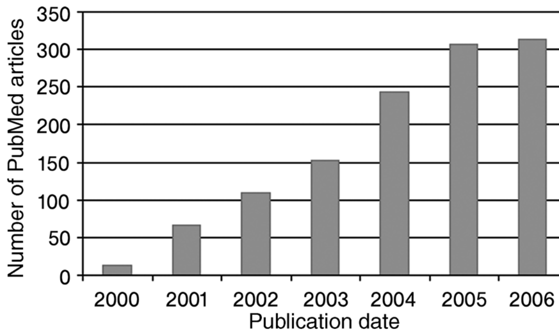


**Figure 1.1** Acceptance and use of microarrays in the twenty-first century. Since the year 2000, publications on microarrays have gained popularity in the area of neurosciences, indicating their more widespread acceptance and use as a viable tool. The *X*-axis indicates publication year, while the *Y*-axis indicates number of publications turning up in PubMed searches for "microarray" and "brain".

## 1.2 Gene Expression Before Microarrays

Since the discovery of DNA in the early 1900s and the subsequent discovery of RNA as the substrate for protein synthesis, gene expression assays have become an essential component of disease research. Gene expression approaches initially took a gene-centric view. A scientist would hypothesize a relationship between a gene and a phenotype, and then test this hypothesis using methods such as Northern blot and *in situ* hybridization. In Northern blot analysis, mRNA is denatured and separated by weight on a gel using agarose gel electrophoresis, transferred onto a membrane, and hybridized with complementary labeled probes [23]. Thus, gene expression correlates with intensity of the labeling. *In situ* hybridization, on the contrary, involves directly applying labeled probes to the tissues of interest to determine where the mRNA is expressed *in situ* [24]. Although still important for studying single genes, these high-resolution techniques are at a disadvantage with regard to the throughput now available using techniques such as RT-PCR, serial analysis of gene expression (SAGE), differential display, and microarrays. As is the case in complex, dynamic tissues such as the brain and nervous system, there is often a trade-off between scale and resolution [25].

### 1.2.1 High-Throughput Gene Expression Techniques

A paradigm shift occurred in the early 1990s, as technology improved and knowledge of the genome became widely accessible. This challenged scientists to move from a gene-by-gene study to develop methods that took into consideration the entire

system of gene expression, moving from the unimolecular to the systems level [26]. One of the earliest methods using high-throughput techniques to identify a large number of genes differentially expressed between two tissues or conditions was differential-display reverse-transcription polymerase chain reaction (DDPCR) [27]. In DDPCR, the 5′ end of mRNA is bound to anchor primers and reverse transcribed. A subset of this cDNA is then PCR amplified near its 3′ end using short arbitrary primers. The resulting amplified cDNAs from two samples are run side by side on a gel, and any differentially displayed bands of interest can be excised from the gel, reamplified, cloned, and sequenced. This method is relatively inexpensive and can test gene expression of all transcripts amplified simultaneously; however, every interesting band has to be sequenced individually, and the completion of the human genome project has rendered such time-consuming sequencing unnecessary.

Representational difference analysis (RDA) represents a more elegant genome-wide subtraction method that, unlike DDPCR, does not require sifting through an entire gel of genes to find some that are different [7,28,29]. In RDA, populations of mRNA from two separate tissues are transcribed into cDNA, digested using restriction enzymes, converted into primers, and PCR amplified. These populations are then cross-hybridized by combining an excess of one population (driver) and using that to remove identical transcripts from the less concentrated population (tester). By iteratively performing this process with each population as the driver and the tester, and then shotgun cloning the subtraction products, libraries for genes enriched in each tissue can be created. We have used this method coupled to microarray screening, which provides a powerful approach to screening genetic subtractions [7,28].

### 1.2.2
### Contemporaneous Alternatives to Microarrays

SAGE [30] is one of the several high-throughput sequencing methods that provide a powerful technique for high-resolution assessment of gene expression in a relatively small number of samples. In SAGE, cDNA is positionally anchored using restriction digestion, and short nucleotide chains around 14 base pair (bp) are removed from specific positions in each molecule, serving as tags, concatenated together into polymers of such tags, many multiples of which can be processed in a single sequencing run. Thus, small tags of each gene are present in proportion to their abundance in the starting mRNA and can be counted by efficient sequencing and bioinformatic identification of the gene from which they originate. The resolution in SAGE is limited only by the cost and time of sequencing, but it typically requires about 2000 sequencing reactions for each SAGE library to identify 50 000 tags. However, often one needs to sequence 1 million or more tags to identify low-abundance species in a complex tissue such as the CNS. To compare two tissues, several such libraries need to be prepared from each tissue, making this a high-resolution but low-throughput approach (relative to sample numbers that can be studied). In theory, this technique is sensitive enough to find any mRNA species and

has the advantage over differential display, as each sequencing run determines multiple mRNA species. In practice, however, this technique is too expensive and time consuming for massive parallelization and clinical use.

Massively parallel signature sequencing (MPSS) determines mRNA counts using a principle similar to SAGE [31]. In MPSS, fluorescently labeled cDNAs from the input sample are hybridized to a microbead cDNA library, and hybridized beads are fluorescently sorted and placed on a 2D grid. All beads are then simultaneously decoded and digested 4 bp at a time by binding unique adaptors, which can be read using a charge-coupled device (CCD). MPSS has all the advantages of SAGE and can read many more mRNA species for similar time investment (~250 000), but it requires special equipment and is expensive. Thus, for most, it remains primarily a research tool for in-depth investigation of a few specific samples of interest, although the recent advent of new sequencing technologies will significantly decrease the price of these clone and count techniques.

## 1.2.3
## Microarray Technologies

Microarrays balance sensitivity and throughput to allow efficient study of about 10 000 detected mRNA species in parallel in a large number of samples. This may not allow the maximum depth possible as with MPSS or SAGE but has the advantage of high scalability. The first high-throughput gene expression study, published in 1987, was carried out by Augenlicht *et al.* who used a nylon membrane containing 4000 cDNA sequences to examine gene expression changes in colon cancer [32]. Once solid substrates replaced nylon in the 1990s, this method provided a relatively cheap, quick, and reproducible way for high-throughput gene expression analysis. Owing to the abundant clinical and research applications of this technology, many groups and companies have created their own microarray platforms. Although an entire chapter could be devoted to describing the similarities and differences of these platforms (see [33]), there are two general categories of microarrays: one-color arrays and two-color arrays.

### 1.2.3.1 One-Color Oligonucleotide Arrays

One-color oligonucleotide (oligo) arrays (or chips) marked the first of the commercial microarray technologies [34–38] and were released by Affymetrix in 1996. These arrays required the development of two novel methodologies. Light directed chemical synthesis allows for the direct application of hundreds of thousands of nucleotides to specific positions on the chip at once, bypassing the need for PCR-amplified cDNA probes. By masking all array positions not associated with the applied nucleotide and repeating this chemical coupling for each nucleotide using multiple masks, gene-specific oligo probes, 25 bp in length, are synthetically created. After synthesis of the array, laser fluorescence microscopy can detect hybridization of fluorescently labeled cDNA (target). Expression values for each gene can then be deduced by averaging over multiple probes and using mismatch probes (where the 13th bp has been purposely changed) to account for nonspecific binding.

#### 1.2.3.2 Two-Color Arrays

Contemporaneous with the development of oligonucleotide arrays, a separate yet equally powerful method for running massively parallel gene expression experiments was created [39–43] in which thousands of cDNA probes between 0.2 and 2.5 kb in length were PCR amplified and printed onto poly-L-lysine-coated microscope glass slides using one of two printing techniques. In mechanical microspotting (or passive dispensing) – currently the more popular method – the target is loaded into a dispensing pin using capillary action and placed onto the cDNA microarray by directly contacting the slide. Drop-on-demand (or inkjet) printers use pins with piezoelectric fittings to drop a precise amount of the target onto the slide using an electrical current, without actually having the pins contact the slide. Once synthesized, these cDNA arrays, unlike their one-color counterparts, detect the differential expression between two reference samples, each of which is labeled with separate dyes (typically Cy3-dUTP and Cy5-dUTP). Hybridization fluorescence signals from each dye are detected separately with a dual-wavelength laser scanner and combined into a single pseudocolor image using computer software. Recently, most two-color platforms have shifted from cDNA probes to longer oligonucleotides (30–60 bp), as oligos are generally more customizable, potentially more target specific, and less difficult to amplify and purify than cDNAs. The Agilent platform is an example of a commercial two-color platform based on oligonucleotides [44].

#### 1.2.3.3 Bead-Based Arrays

Most current microarray systems, whether one-color (Affymetrix) or two-color (Agilent), are based on oligos attached to a solid substrate, each with a known address. Illumina universal bead arrays [45,46], however, consist of densely packed wells, $\sim$3 μm in diameter, which are randomly filled with beads containing 75 bp chimeric oligos. These wells are etched either into bundles of fiber-optic strands or onto specialized chips. Each array has an average coverage of $\sim$30 beads per feature, with the exact number variable due to the random filling of wells. For each bead, oligonucleotides consist of a 25 bp bead identifier followed by a 50 bp gene-specific probe, and $\sim$700 000 such oligos are attached to each bead. Bead types are decoded by repeated hybridization (and subsequent dehybridization) of fluorescently labeled cDNA sequences complementary to the bead identifiers. Fluorophores are chosen such that each bead has a unique sequence of fluorescent signals (e.g., red-green-none-red-red-none-green-red after eight hybridizations). After decoding, cRNA from one sample is fluorescently labeled and scanned, and the absolute abundance of transcript is determined by averaging the intensities of each bead containing that transcript.

## 1.3 Designing and Implementing a Microarray Experiment – From Start to Finish

Many articles and guides on the basic design of microarray experiments in the field of neuroscience are available [47–50]. Here, we highlight some of the key issues, starting with the basics.

### 1.3.1
### Choosing the Proper Microarray Platform

Given optimal conditions, all microarray platforms work very well; however, conditions are never optimal, and issues such as experimental assay, local expertise, cost, and gene coverage all play a role in platform selection. A two-color design is most suited for comparative assays, for example, if the experimental goal is to compare multiple tissues from a single subject (tumor versus normal tissue, cerebellum versus cortex, etc.). However, experiments seeking to correlate gene expression with phenotype (such as aging) in a single tissue tend to use one-color arrays; although two-color arrays can be used, by comparing each sample with the same reference sample [41,51]. This choice should be dictated by the statistical design of the analysis, so as to allow optimal power to detect the desired changes.

Another issue to consider is cost versus reproducibility. Laboratory-made spotted oligo arrays cost significantly less than factory-born arrays, whether one-color or two-color, but require more effort to make. All microarrays are prone to batch effects, which can be removed by proper normalization [52,53], but may be more significant in homemade arrays. Thus, in a research-based experiment, custom arrays may be appropriate, whereas biomarker assays would more likely require factory-made arrays since thousands of identical arrays will eventually have to be made quickly. Then, local expertise has also to be taken into account. If all of the current lab personnel were trained using a specific kind of array, then the continued use of those arrays would decrease both experimental time and error. One more advantage of homemade arrays is that they are not vulnerable to changes in designs of manufacturers during the course of a series of experiments, as has been the case with every commercial platform so far.

The final issue to consider when choosing an array platform is customizability versus scalability. Both homemade and Agilent two-color arrays allow for the quick and cheap creation of arrays containing any target of interest. For example, if an experimenter aims to test the expression of multiple splice variants of a gene or to make a biomarker assay for testing the expression of 100 specific genes, such arrays would be appropriate. Nimblegen, which uses a mirror-based masking system, has the maximum synthetic flexibility and offers custom arrays on a commercial platform [54]. A wide variety of configurations are available, but the cost is far higher than homemade spotted arrays. Spotted array technologies do not lag far behind, however, as just about every microarray platform currently has an array to test the expression of every known human transcript. All of these factors have to be taken into account while choosing a proper platform for the experiment at hand.

### 1.3.2
### Preparing the Tissue for Hybridization

After selecting the microarray platform, tissue must be acquired and prepared in such a way to avoid inducing unwanted changes in gene expression. An experiment

using postmortem tissue must carefully control for gender, ethnicity, and cause of death to avoid outlier arrays [22,47,55]. Once tissue is acquired it must be properly cared for, as excessive postmortem interval, changes in pH or temperature, or improper tissue handling at any stage can lead to RNA degradation [50]. Nonlinear or excessive PCR can selectively amplify smaller segments of cDNA, leading to increased variability, while improper tissue preservation can make a sample completely unusable. Generally, an experiment should include duplicate spots or arrays to quantitatively assess variability due to human error or choice of array platform, thus, effectively determining the sensitivity of the experiment [50].

The precise steps to be taken between tissue acquisition and target hybridization depend highly on the experiment at hand and generally involve the use of a series of well-characterized procedures and commercially available kits. Generally, mRNA or total RNA samples are extracted from the tissue or cells of interest and converted into their cDNA sequences. In the case of genomic DNA assays, DNA is cut into manageable sizes. When necessary, the cDNA is then amplified. Finally, each sample is fluorescently labeled using one or two dyes, as required.

### 1.3.3
### Single-Cell Assays and Tissue Heterogeneity

Under ideal conditions, microarrays can detect mRNA in relative abundances as low as one part per 500 000 [39,43], allowing for a resolution of 3–10 copies per cell in simple tissues and cell lines. In practice, however, while these species may be detected, their detection may not be reliable enough to ascertain differential expression; so it is safer to assume reliable detection at the 1/100 000 level. In the nervous system and other complex tissues consisting of multiple cell types with uniquely expressed transcripts, resolution of cell type specific species is hampered [50]. To increase resolution, therefore, many microarray experiments now use single-cell assays to filter individual cell types of interest from heterogeneous tissue before assessing changes in gene expression [25], although this also has its costs.

Many cell purification assays, including flow cytometry, microaspiration, and laser-capture microdissection (LCM), have arisen to combat tissue heterogeneity in different situations. Flow cytometry allows thousands of cells per second to be counted, examined, and separated based on any of a number of characteristics of the cells [56]. This method is generally used to quickly obtain large quantities of a single cell type. Several studies have recently demonstrated the use of automated flow sorting [3,4] for purifying neurons from developing and adult brains [4]. In addition, fluorescence can also be applied manually [5], although it is more tedious. The use of automated sorting allows for large-scale purification of thousands of neurons in a single sort. Cells can be labeled by tracer injection [3] or in genetically modified mice [4], which are now available in many forms via the GENSAT project [57]. Microaspiration, on the contrary, involves patching onto individual cells and removing them one at a time [58]. This process is much more painstaking than flow cytometry and provides many fewer cells; however, it can provide much more accurate

dissections, and it even allows for dissection of only certain elements of the cell (soma, dendrites, etc. [59]). In LCM, a tissue slice is prepared by labeling cells of interest and placing the slice below a transfer cap on a laser-mounted microscope. Once a labeled cell is located under the microscope, a laser beam can be activated, transferring the cell to the transfer film [60]. Using LCM, several hundred cells can be collected in a relatively short time using a simple assay. Like flow cytometry, LCM allows for assaying a relatively large number of cells, and like microaspiration, it allows for the inspection of cells before acquisition, although cell extraction is not quite as precise.

### 1.3.4 Microarray Hybridization and Scanning

In two-color assays, equal amounts of differently labeled target from the two samples are combined and cohybridized, while Affymetrix array assays only require hybridization of target from a single sample. In both cases, hybridization lasts overnight, and then arrays are rinsed to remove any nonspecifically bound cDNA. A tunable laser then excites the remaining bound target at the fluorescent frequency of each dye, while a CCD or confocal microscope collects the fluorescence intensities. Most scanners come with their own preprocessing software, although stand-alone programs, such as DNA-Chip Analyzer (dChip) [61] or Imagene [62], can also be used.

### 1.3.5 Preprocessing

One important issue in microarray studies is how to go from a series of images to a table of expression values for each probe set. While methods differ between platforms, preprocessing can generally be divided into six steps: image analysis, data import, background adjustment, normalization, summarization, and quality assessment [52]. In image analysis, spot detection methods convert pixel intensities to probe intensities, and background levels are stored. Data import methods then take all image analysis files, bring them together, and convert them into a form recognized by the relevant software. Next, probe intensities are background-adjusted to account for nonspecific hybridization and noise in the scanner. All images are then normalized to account for variations among arrays. In some microarray platforms, summarization takes intensities from all spots on an array representing a single probe set and converts them into a single value representing the amount of RNA transcript. Finally, all of the data are assessed for quality, and any measurements falling outside an acceptable range of random noise are omitted.

Custom two-color arrays and Affymetrix chips use very different strategies for preprocessing. Custom two-color arrays use a much more intuitive method; for each sample, a red or a green spot on the array represents each probe set, and the ratio of these values represents the relative abundance of that transcript. In Affymetrix chips, multiple probe pairs consisting of a perfect match and a mismatch oligo represent each probe set on the array. The mismatch probe differs from the perfect

match probe only in the 13th nucleotide and is used to account for nonspecific binding. Absolute expression values are then obtained by averaging the difference of intensities between each probe pair for a given gene. Whatever the platform, once preprocessing is complete and a core table of gene expression is obtained, data analysis is ready to begin. Many methods have been used to extract signals from Affymetrix arrays and their utility varies according to the experimental design and conditions [52]. Recent comparisons of these techniques suggest that combining the methods of microarray suite (MAS) background correction, quantile normalization, MAS perfect match probe correction, and median polish expression summary offers clear improvements over the manufacturers software [53]. Additionally, the R package ''ProbeFilter'' can be used to eliminate many of the nonfunctioning and nonspecific Affymetrix probes to further improve the data quality (http://arrayanalysis.mbni.med.umich.edu/MBNIUM.html).

### 1.3.6 Gene Expression Analyses

Standard gene expression analyses involve finding all genes that significantly differ between two categories (Alzheimer's disease versus control hippocampus [12], cortex versus cerebellum [63], tumor 1 versus tumor 2 [6], etc.). The most common method for categorization is to rank each gene by ratio, and then to consider all genes with a ratio greater than $X$ (or less than $1/X$) differentially expressed, where $X$ is chosen on the basis of the level of statistical significance desired (Figure 1.2). A more elegant, higher powered method of analysis is to correlate gene expression in a single tissue with a related phenotype (age [64], neurofibrillary tangle burden [11], body weight [65], etc.). In this method, the Pearson correlation between expression and phenotype is calculated and any significantly correlated genes ($P < 0.05$, typically) are recorded. These analyses produce several to thousands of significant genes, from which a select few are generally investigated. There are many valid analytic approaches and high-quality free software, such as BioConductor (http://www.bioconductor.org/ [66]) and Multiexperiment Viewer (http://www.tm4.org/mev.html), available for analysis. It is best to consult a statistician to ensure design and analysis are optimally performed.

### 1.3.7 Analytical Challenges

One major issue in microarray analysis is determining which genes to follow up from the list of significant genes, often numbering in the thousands. A few large-scale efforts, including public efforts such as the Gene Ontology Consortium (GO; http://www.geneontology.org/) and Kyoto Encyclopedia of Genes and Genomes (KEGG; http://www.genome.jp/kegg/), and private efforts like BioCarta (http://www. biocarta.com/genes/index.asp), have gone into categorizing genes into biologically meaningful groups. GO, for instance, provides a hierarchical framework of terms related to biological processes, molecular function, and cellular components.

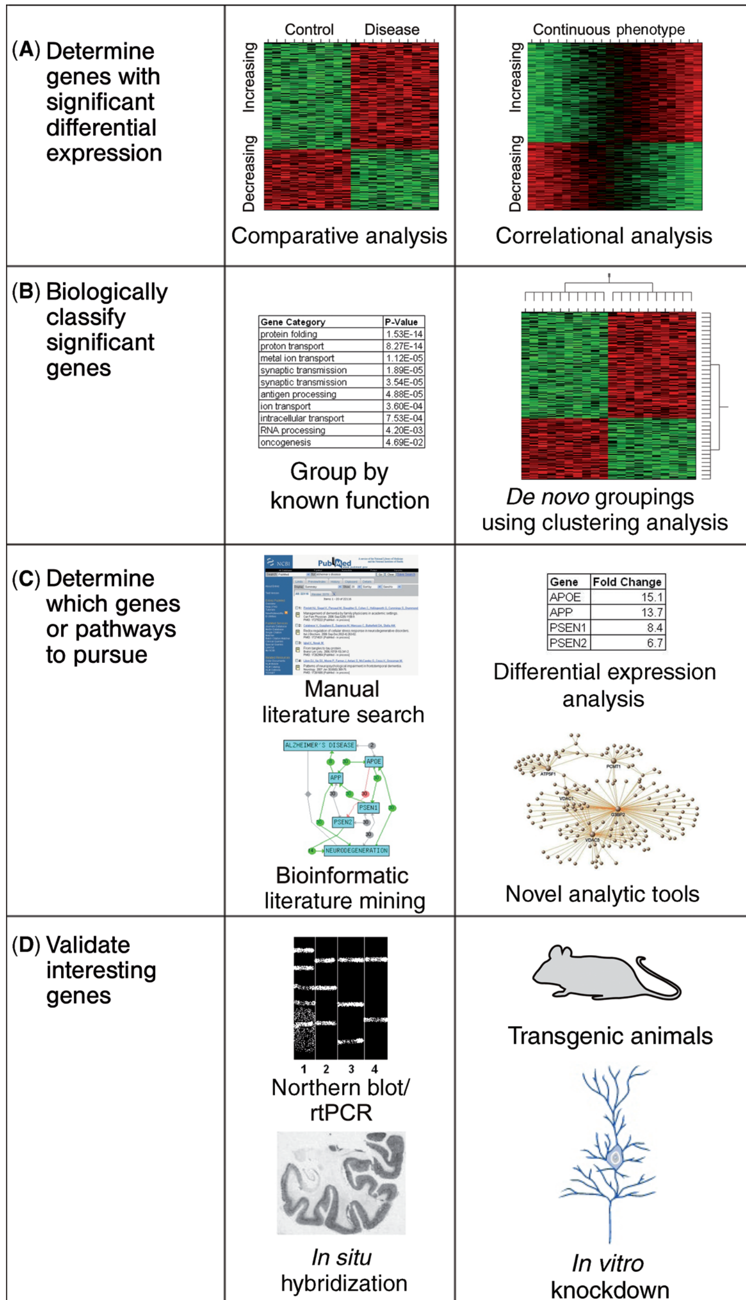


**Figure 1.2** (legend see p.11)

As the function of new genes becomes known, they are assigned relevant GO categories, which are then stored in a publicly accessible database. To navigate this database, various programs such as expression analysis systematic explorer (EASE) [67] and gene set enrichment analysis (GSEA) [68] have been created. EASE searches for overexpression of GO terms in a gene list of interest (e.g., all downregulated genes) compared to a reference list (e.g., all genes present on the array). GSEA takes all genes present on the array, ranks them by fold change, and then searches for gene ontology categories with many highly ranked genes, providing power to detect genes not identified with conventional methods. WebGestalt (http://bioinfo.vanderbilt.edu/webgestalt/) allows many of these pathway-related analyses to be performed and stored online in a user-friendly environment, and the web site keeps up-to-date links to many analytic tools [69]. Commercially available programs such as Ingenuity Pathways Analysis (IPA; Ingenuity Systems, www.ingenuity.com), which are highly curated, also help locate genes of interest relative to known pathways.

If many significant genes fall into one or two GO categories, choosing the most significant genes in those categories would be one option. Alternatively, literature-mining tools such as Chilibot [70] allow a user to input a list of terms (genes or phenotypes), return all known relationships between all terms in the list, and suggest hypotheses for new relationships. More focused literature searches can be performed using GeneCards [71], a web site compiling links to everything that it can find about every human gene, including links to primers, RNAi products, knockout mice, and so on. Web sites such as the Allen brain atlas (http://www.brainatlas.org/aba/) and WebQTL (http://www.genenetwork.org/) provide information about gene expression patterns for thousands of genes in the mouse brain, and in some cases correlate knocking out of various genes with phenotype. A list of other useful web sites for gene expression analysis is included in Table 1.1.

**Figure 1.2** Determining genes of interest from a microarray experiment. (A) After acquiring and preprocessing microarray data, significantly expressed genes are determined using a method depending on the experimental design. In comparative analyses, with two (or more) distinct groups of samples, fold-change measurements determine differential expression, whereas in analyses with a continuous phenotype, differentially expressed genes are those with the highest Pearson correlation with the phenotype of interest. (B) These differentially expressed genes are then biologically classified, both by considering the known function of each gene in a Gene Ontology or KEGG pathway analysis and by looking at *de novo* groupings of genes using hierarchical or *k*-means clustering. (C) Of the tens to thousands of significant genes and pathways typically discovered in microarray analyses, generally only a select few are pursued, and determining relevant genes to pursue is a critical step. Genes essential to significant pathways can be determined by searching the literature either manually using PubMed or GeneCards, or though bioinformatic literature mining tools like Chilibot and IPA. However, relevant genes may be chosen based on expression alone, either by taking genes with the highest differential expression or by using novel analytic tools such as gene coexpression analysis to determine genes with high connectivity. (D) Selected genes are then validated technically using low-throughput methods such as Northern blot and *in situ* hybridization, and genes with confirmed differential expression are then functionally validated using various *in vitro* and *in vivo* genetic manipulations, such as RNAi-mediated knockdown and transgenic knockouts.

**Table 1.1** Additional web sites of interest in microarray analysis.

| Web site | Description |
|---|---|
| http://bioinformatics.org/ | Bioinformatics tools and information |
| http://brownlab.stanford.edu/ | Information on microarrays at Stanford and links to collaborating labs |
| http://derisilab.ucsf.edu/ | Information on yeast molecular biology and human infectious disease |
| http://genome.ucsc.edu/ | A comprehensive tool for scanning and analyzing the human genome |
| http://ihome.cuhk.edu.hk/%7Eb400559/ | A nearly complete list of web sites related to functional genomics |
| http://pevsnerlab.kennedykrieger.org/ | A comprehensive list of bioinformatic-related tools and databases |
| http://rana.lbl.gov/ | Useful software and web tools for genomic and gene expression analyses |
| http://www.genenetwork.org/ | A set of linked resources for systems genetics in mice |
| http://www.geneontology.org/ | The main gene ontology web site with tools for gene expression analysis |
| http://www.informatics.jax.org/ | A comprehensive mouse genomics resource |
| http://www.ncbi.nlm.nih.gov/geo/ | A repository for thousands of publicly available, MIAME compliant microarray data sets |
| http://www.ncbi.nlm.nih.gov/gquery/ | A list of all major NIH databases |
| http://www.nervenet.org/ | Several useful genetics and gene mapping databases, as well as a mouse brain library |

This list represents a comprehensive, but by no means all-inclusive, list of web sites useful for bioinformatic analysis of gene expression data. Many of these web sites have an extensive set of links to numerous other expression-related sites.

Still, moving from the gene list to function remains difficult using conventional methods. We have been very interested in moving to a systems level of understanding of gene expression data and have used the merging of network theory and systems biology, a technique of recent origin, from which extremely powerful methods for quantifying gene coexpression have developed [72–75]. These coexpression methods group genes on the basis of similarity of their expression values to each other and take advantage of three properties of biological systems: (1) gene (and protein) expression networks follow a scale-free topology, meaning that – like airports – a few ‘‘hub’’ genes share similar expression patterns with many other genes, but most connect to only a few other genes; (2) these hub genes are generally more important to cellular function than less connected genes; and (3) groups of genes with similar expression patterns are generally involved in similar biological functions. We have used this approach to identify key targets of human brain evolution [74] and therapeutic targets in cancer [76], thus proving for the first time that these techniques can revolutionize the kind of information obtained from a microarray experiment. Ultimately, a combination of these methods should be used to determine which genes to follow up, and some, if not all, of the chosen genes should be validated using other methods.

### 1.3.8
### Validating Results

Northern blot, *in situ* hybridization, and various forms of PCR are still generally used to quantitatively or qualitatively replicate the most interesting differentially expressed genes. There has been some debate in the literature about whether microarray experiments, when run properly, are more or less accurate than the techniques used to validate them. Low-stringency microarray hybridizations may show higher background, decreasing observed expression differences as compared to single-gene techniques [10]. However, small pipetting errors or incorrect choice of a housekeeping gene for normalization may lead to quantification errors in RT-PCR, which are not found when using microarray platforms [77]. In any case, successful gene expression validation, as generally occurs with highly significant results, provides even more confidence in the ability of microarrays to find phenotype-related expression changes in a highly parallel fashion. More importantly, genes discovered using microarrays and validated using alternative methods generally provide a starting point for determining therapeutic targets in the disorder being studied [47].

## 1.4
## Clinical Applications

Microarray studies generally fall into one of the two main categories. Some studies seek to identify molecular changes in disease or injury, while others aim to identify molecular biomarkers associated with disease (Figure 1.3). Studies identifying molecular changes in disease provide valuable information about pathological changes associated with disease. These studies look at tissues directly affected by disease pathology, which, with the notable exception of cancer tissue or the rare surgical specimen obtained in dementia or epilepsy, generally cannot be obtained while the subjects are alive; as a result such research cannot be directly applied in a clinical setting. However, biomarker identification studies only look at tissue readily accessible with the goal of accurately classifying patients using a simple gene expression assay. This section will focus on clinical uses of microarrays such as biomarker discovery, prognosis, diagnosis, and treatment/response studies. We will begin by presenting a short summary of notable studies in neurological disease research that have advanced the field significantly, suggesting new therapeutic targets for disease and injury treatments.

### 1.4.1
### Neurological Disease-Relevant Research

So many microarray studies have been conducted in the area of basic neuroscience research that an entire review can be written on this topic alone. However, many of them end up with a gene list useful as a database of genes only, as they do not go beyond

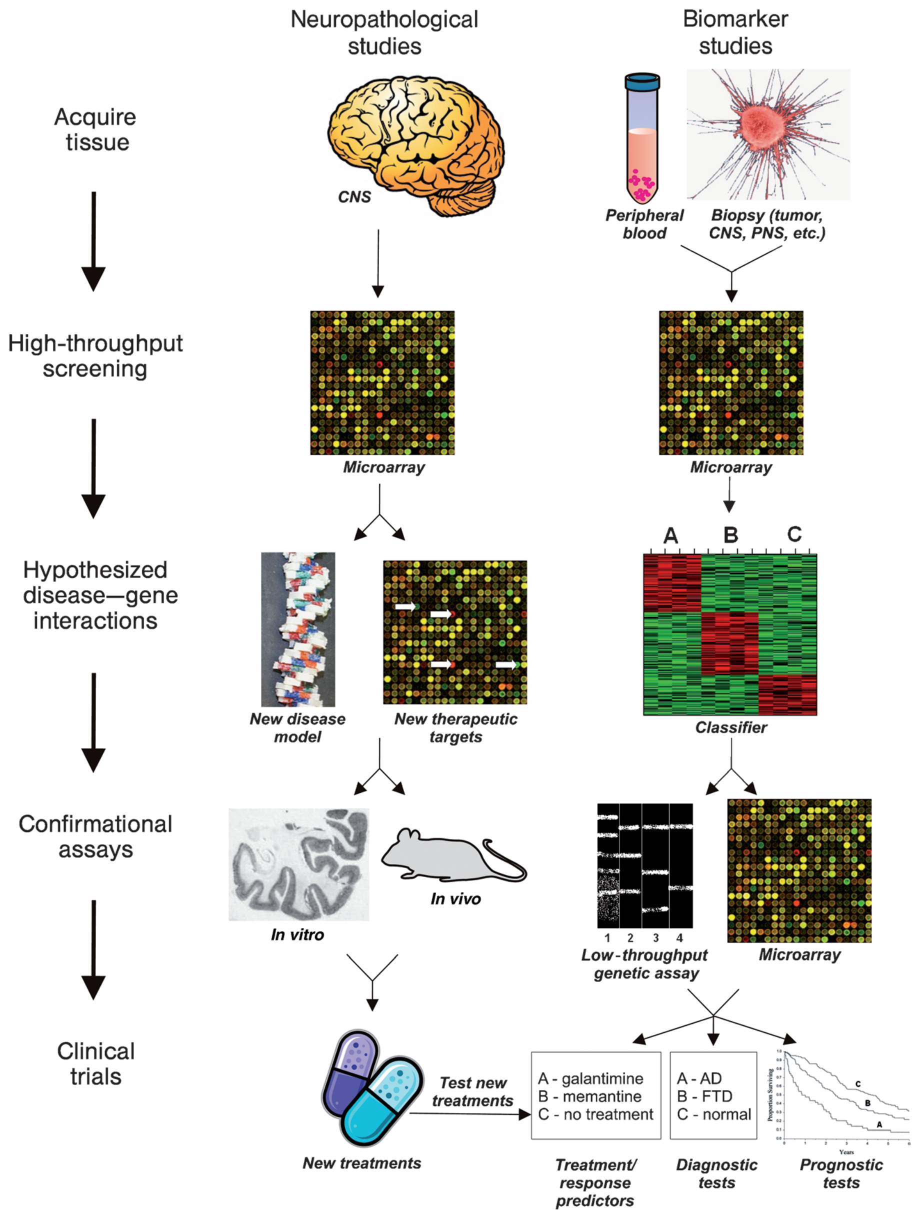


**Figure 1.3** Parallel methodologies in neuropathological and biomarker studies. Since the tissue is generally acquired postmortem in neuropathological studies (left), these studies seek to answer the question, "What went wrong in this disease and why?" However, biomarker studies (right) use peripheral blood as well as tumor and other biopsy tissues in living patients to answer the question, "How can we classify patients?" In both types of studies, high-throughput screenings are performed on acquired tissue (see Figure 1.2) to determine possible disease–gene interactions. In neuropathological studies, these interactions come in the form of new disease models and therapeutic targets. Treatments for these

that. Here we review a few significant studies because they introduce other key aspects of microarray research or suggest new therapeutic targets for neurological diseases.

#### 1.4.1.1 Incipient Alzheimer's Disease

In the most elegant study on AD published to date, Blalock *et al.* [11] used microarrays to study gene expression changes specific to early stages of Alzheimer's disease and noted which changes were sustained throughout AD progression. To do this, gene expression in samples were correlated to the cognitive status of each individual at death (MiniMental State Examination), as well as to their pathological profile (neurofibrillary tangle burden). Genes were considered AD genes if they were significantly correlated with either NFT burden or MMSE score. They were further considered early AD genes if they were also correlated when the analysis was limited to comparison of control and incipient subjects, finding an upregulation of tumor suppressors and other genes regulating cell proliferation and differentiation, as well as the expected upregulation of apoptosis and downregulation of synaptic transmission pathways.

While many AD microarray studies have been published in the past 5 years [11–15], this study was significant for a few reasons. First, the complexity of AD was taken into consideration when designing the experiment. Instead of designating subjects as either diseased or control, each individual's cognitive ability and Tau pathology were correlated to gene expression as described above. Second, the sample size was large enough to provide decent statistical power ($N = 31$) and allow for significant correlations in early AD using only half the samples. Careful tissue handling and microarray analysis methods resulted in very low false discovery rates for microarray experiments, further increasing statistical power. Additionally, raw data from this experiment were uploaded to the Gene Expression Omnibus (GEO) repository (http://www.ncbi.nlm.nih.gov/geo/) so that other researchers could review these data using other methods. Finally, results from this microarray analysis were used to suggest a new model of incipient AD progression along myelinated axons. None of the results were validated functionally or by alternative methods, however, and no aspects of the model were tested, thereby limiting the possible clinical applications of these results.

**Figure 1.3** (*Continued*)
hypothesized therapeutic targets can then be tested *in vivo* or *in vitro*, either using known interacting compounds or by screening large libraries, with successful treatments then entering clinical trials. Disease–gene interactions in biomarker assays, however, are used to classify patients based on diagnosis, prognosis, response to treatment, or a number of other clinically relevant measures. Any promising classifications are then retested in an independent cohort of individuals using either low- or high-throughput tests, as necessary. New treatments from neuropathological studies of a disease can be tested clinically on patients predicted by biomarker assays to have poor prognosis and no predicted response to available treatments, thus providing patients for the clinical trials and possible new treatments to patients who would otherwise not be treated.

#### 1.4.1.2 **Using Animal Models to Study Neurodegeneration**

An ideal microarray experiment will make predictions that can be validated *in vitro* and *in vivo*, leading to new results with clinical applications. One early example of this was in multiple sclerosis (MS) [78], where findings in human patients were extended to functional validation in mice. Such studies are few and far between because of the effort needed to validate experiments in animal models. One such study on dementia by our group recently compared gene expression changes in mice expressing the most common FTD-causing human mutation in Tau with wild-type control mice [79]. This study found genes up- or downregulated with Tau mutation in cortex, cerebellum, brain stem, and spinal cord, suggesting that pathologically unaffected areas (like cerebellum) would show upregulation of neuroprotective genes. This hypothesis was tested using multiple *in vitro* and *in vivo* assays, focusing on puromycin-sensitive aminopeptidase (PSA) – a gene which resides very close to tau on chromosome 17, and which was upregulated in cerebellum. For example, in a fly transgenic model of tauopathies, PSA loss of function worsened the rough eye phenotype, and *in vitro*, PSA digested recombinant Tau protein. These results suggest that PSA overexpression may be a viable mechanism for treatment of many neurodegenerative disorders involving tauopathies, and have identified a major new pathway for Tau degradation *in vivo* [79].

Animal models are also useful for determining cell type specific markers. In some situations, multiple morphologically similar cell types will reside in the same cortical areas, making discrimination of these cell types difficult. Corticospinal motor neurons reside in layer V of the cortex along with other projection neurons, for example, and degenerate in upper motor neuron disorders such as ALS. Arlotta and colleagues [3] discovered markers specific to three types of cortical projection neurons in developing mouse cortex by using retrograde labeling of axonal projection fields and fluorescence activated cell sorting (FACS) to separate each cell type, and then ran microarrays on each pure cell line. Using immunohistochemistry and loss of function knockout assays, they found that CTIP2 was specific to corticospinal motor neurons and was required for axonal pathfinding in the spinal cord. While this study cannot be directly applied to adult brain, it provides evidence that functions of specific molecules in specific cell types can be determined using microarray screening and animal models.

#### 1.4.1.3 **Determining Gene Pathways Using Meta-Analysis**

Meta-analysis of microarrays provides a novel method for extracting groups of genes involved in similar biological pathways. If two genes that may show similar expression patterns in one data set by chance are coexpressed in multiple data sets, it is more likely that the two genes are functionally related. Lee *et al.* [80] looked at gene coexpression across 60 data sets, finding a substantial number of correlated expression patterns (connections) that occurred in multiple studies. Additionally, the probability that two genes were functionally related (as measured in overlapping GO terms) was directly correlated with the number of connections between those two genes. Furthermore, many of the genes could be clustered into groups related to translation, transcription, cell adhesion, electron transport, and many other biologi-

cally relevant categories based solely on their multistudy connections. With more microarray data becoming publicly available every day and with computer processor speeds increasing exponentially, such multistudy meta-analyses have the potential to provide high-throughput pathway predictions for genes expressed in the brain and other tissues [74].

### 1.4.2 Cancer Research

Cancer research represents one of the few fields of microarray study where results from the lab can be directly applied clinically (see [6] for review). Unlike most neurological disorders where diseased tissue can only be acquired from a patient postmortem, tumor samples can be removed from living patients. Tumors can then be classified based on the survival or treatment response of patients in the future. In this section, we will explore various microarray applications in cancer genomics, focusing on applications related to brain tumors.

#### 1.4.2.1 Tumor Classification and Patient Diagnosis

Traditionally, tumors are classified morphologically based on their microscopic resemblance to their CNS cell of origin or developmental progenitor cell [81]. Malignant tumors tend to look more like dedifferentiated precursor cells showing mitotic activity, tumor necrosis, and angiogenesis [6]. Less aggressive tumors, on the contrary, resemble their normal tissue counterparts, often making clinical diagnosis difficult. While this classification scheme works with reasonable accuracy, morphologically identical tumors can exhibit distinct mutational patterns and altered signaling pathways, as well as respond differently to identical drug treatments. An alternative method for categorizing tumor samples is gene expression profiling. In situations where biopsy material cannot be morphologically characterized clearly as either malignant or benign, a clinical test for gene expression patterns characteristic of specific tumor types would be quite useful. By disregarding prior knowledge and grouping tumors entirely on the basis of their transcriptional similarity, unsupervised methods can often uncover morphologically unrecognized tumor subtypes, thereby allowing for a more precise diagnosis of tumor malignancy [51].

Multiple experiments have shown that tumors from different cell types and with different levels of aggressiveness show distinct gene expression patterns [6]. One such example is glioma, where standard morphological categorizations have been useful for predicting survival but not the response to treatment [82]. Shai *et al.* performed unsupervised hierarchical clustering of various gliomas on a panel of 170 genes differentially expressed between tumor types. Glioblastomas, lower grade astrocytomas, and oligodendrogliomas, they concluded, have transcriptional signatures distinct from one another and from normal white matter tissue [82]. Additionally, secondary glioblastomas showed a wide range of expression patterning, suggesting that they constitute a very heterogeneous group. A more recent, large-scale study based on the gene expression profile of a panel of 595 genes [83] found distinct subsets of morphologically identical gliomas, confirming the hypothesis of

glioblastoma heterogeneity. More important, these tumor subtypes predicted patient survival better than the standard methods, suggesting that tumor gene expression profiling would be a viable alternative for patient prognostic tests.

#### 1.4.2.2 **Determining Patient Prognosis**

In addition to classifying tumors for diagnostic purposes, gene expression profiles can be used as biomarkers to predict patient survival. Accurate prognostic tests are useful to determine how aggressive a cancer treatment administered to a patient should be. If a malignant tumor goes untreated, a patient will likely relapse; however, aggressive treatment can lead to severe side effects, including permanent cognitive and psychological deficits [84]. Current clinical methods for risk classification in medulloblastoma fail to identify 20–30% of average-risk patients with resistant disease as well as those patients who might be overtreated [85], suggesting that more accurate methods for risk stratification are necessary. Molecular prognostic markers, as measured by microarray, RT-PCR, or Western blot, provide one such method that, as preliminary results indicate, will be more accurate than traditional methods.

An early cDNA microarray study of 60 medulloblastoma samples found that patient survival could be predicted using as few as eight genes, independent of the clinical variables traditionally used to stratify patients as average versus high risk [86]. These results suggested that not only could gene expression profiles be used as a biomarker for patient survival, but also that such classifiers would be a realistic option for the clinical setting, as so few genes were needed for classification. With the price of microarrays decreasing, repeatability is now becoming more of an issue than small gene number. As mentioned above, a recent microarray study of gliomas found that expression-based tumor classification using a panel of 595 genes is a more powerful predictor of survival than age, pathological type, or grade [83]. A subset of this classifier has been validated on an external data set, and despite the fact that a different microarray platform was used, this classifier significantly outperformed the initial histopathology grading classification. In both the initial and the validation study, samples could be classified to one of the four glioma subtypes that predicted patient survival with high accuracy. Recent experiments have led to similar findings in less common forms of brain cancer, such as ependymal neoplasms [87], suggesting that such prognostic tests may be universally applicable to many forms of brain cancer. These results indicate that a small panel of genes can be used to predict patient survival with high accuracy and to determine whether a patient needs aggressive intervention.

#### 1.4.2.3 **Predicting Therapeutic Response to Treatment**

Just as prognostic tests could be used to prevent patient overtreatment and minimize side effects, tests predicting therapeutic response would prevent ineffective treatments from being administered, thus allowing a patient to receive proper treatment as quickly as possible. Such screens are important because administration of resistant chemotherapies to a patient can lead to the development of secondary antineoplastic drug resistance, and drugs that at one point would have been useful would no longer be effective [88]. In a proof-of-principle study, Staunton *et al.* [89] developed gene expression classifiers for sensitivity or resistance of human cancer cell lines to

232 compounds. They measured expression of 6817 genes on a panel of 60 human cancer cell lines, finding classifiers of between 5 and 200 genes whose expression patterns could accurately predict chemosensitivity for approximately one third of the compounds tested. A similar, more focused study published recently tested chemosensitivity of 30 cell lines to 11 anticancer drugs at clinically achieved concentrations [88]. Prediction profiles for each chemotherapy agent were compiled and had an overall accuracy of 86%, once again demonstrating the applicability of microarray technology in drug resistance prediction. Additionally, 67 genes associated with resistance to at least four drugs were found from these profiles. Many of these associations were confirmed using RT-PCR, providing new candidate genes associated with multidrug resistance.

#### 1.4.2.4 **Determining Therapeutic Targets for Cancer**

Glioblastomas and medulloblastomas represent some of the most common forms of malignant brain tumors in adults and children, respectively. Even with aggressive treatment, patients with glioblastomas have a median survival time of only 15 months after diagnosis [76], and only 50% of children with medulloblastomas survive to adulthood, often with severe psychological and cognitive side effects [84]. Clearly, new treatment approaches for these and other cancers are necessary. One screening approach for novel therapeutic targets for cancer treatment is gene expression based high-throughput screening (GE-HTS) [90]. First, microarrays are used to find a small number of genes that serve as markers for a desired cellular state, such as differentiation. Next, thousands of compounds of interest are individually combined with a cancer cell line in a high-throughput manner. The molecular markers found using microarrays are then measured using RT-PCR for the cell lines with each combined compound to see which ones have been transformed into the desired cellular state. In a study of leukemia cells, Stegmaier *et al.* found that eight of the 1739 compounds tested caused cancer cells to differentiate, leading to growth arrest; after further validation they concluded these to be viable therapeutic targets [90]. Using this method, discovery of new therapeutic targets requires no special assays or reagents and no prior target validation, as the gene expression profiles serve as markers for the state in question.

Coexpression analysis of gene expression data has also been used to discover new molecular targets for glioblastoma treatment. In a significant recent advance, Horvath and colleagues applied a systems biology approach to study 120 patient glioblastoma samples [76]. They built a network in one set of 55 arrays and validated it in an independent group of 65 arrays. This network grouped genes into five modules using coexpression, one of which was significantly enriched in cell growth genes. They identified a set of key hub genes that were driving glioblastoma proliferation in this module, thus providing a proof-of-principle functional validation of the network predictions. By looking at the 10 most connected genes in this module, they found that most genes had already been identified as potential cancer targets, and all but one were associated more with glioblastoma survival than the traditional proliferation markers (Ki67 and PCNA). Additionally, inhibition of ASPM, a novel therapeutic target, was confirmed to inhibit glioblastoma cell growth using siRNA.

Similar results were duplicated using breast cancer samples, suggesting that these results extend across cancer types.

This study presented a new computational technique that can be applied to any cancer microarray experiment and led to insights not observable with standard analysis of differential expression [75]. With the rapidly increasing popularity of microarray and enough studies already published, new therapeutic targets should be discoverable merely by reanalyzing existing data sets using novel computational techniques, such as sophisticated network analyses.

A recent microarray study suggests that new cancer treatments may be found simply by combining old chemotherapies together in novel ways. Cheok *et al.* [91] studied single-agent versus combination therapy in patients with acute lymphocytic leukemia, by comparing gene expression profiles before and after treatment in patients who were treated with methotrexate, mercaptopurine, or both. They found that the combined effect of both chemotherapies was profoundly different from the additive effects of patients taking either drug. Thus, new treatments may be found simply by combining old treatments in novel ways and screening such combinations using microarrays.

### 1.4.3
### Using Peripheral Tissues as a Substitute for Brain Tissue

In most neurological diseases, the affected tissue that holds the key to patient survival cannot be directly assessed in a clinical setting. In these situations, alternative methods must be used for diagnosis and treatment. For example, in many cases damage to the brain can be indirectly assessed using peripheral tissue or blood. Many scientists are rightly skeptical that peripheral, nonneural tissues can provide a window into the CNS. However, as we will discuss, current results indicate that a significant proportion of CNS genes are expressed in peripheral blood or lymphoblasts, and that these tissues can be used to identify biomarkers, if not therapeutic pathways.

#### 1.4.3.1 Gene Expression Profiling of Disease

Unique blood biomarkers have been found for all kinds of neurological disorders, including stroke, hypoxia, multiple sclerosis, depression, Down's syndrome, HD, and seizure, as well as general neurological distress [92–94]. The first study to link transcriptional changes in peripheral blood to brain dysfunction measured gene expression in adult rat blood monocytes, 24 hours postinjury, for a variety of induced injuries [95]. Using several hundred genes as a biomarker, animals with ischemic strokes, hemorrhagic strokes, kainate-induced seizures, hypoxia, and insulin-induced hypoglycemia could be genetically distinguished from controls and each other. Many other genes were similarly regulated in each type of injury, suggesting they are stress related. Similar results were found when the experiment was repeated using rat cortex, confirming the link between peripheral gene expression changes and disease [96]. More recently, studies have compared gene expression in human blood following ischemic stroke [97,98], finding that only a small number of genes were required to distinguish controls from stroke patients. Such studies suggest that a blood test for stroke using either RT-PCR or even microarrays will be very possible in the near future.

One study, probably the most comprehensive to date, relating peripheral blood profiling to disease looked at genes upregulated in subjects with presymptomatic and symptomatic HD and compared age- and sex-matched controls [17]. This study was very well designed, using multiple microarray platforms and quantitative RT-PCR, as well as a test and training set assay, to confirm results. It found that gene expression profiles could separate not only HD subjects from controls, but also presymptomatic from symptomatic subjects, with early presymptomatic subjects having profiles similar to controls and late presymptomatic subjects showing profiles similar to symptomatic HD subjects. Additionally, these upregulated transcripts were less elevated in HD subjects taking the histone deacetylase inhibitor sodium phenylbutyrate, which is a potential experimental therapy. Finally, as with the animal models of acute injury, these genes upregulated in blood were also elevated in the postmortem HD caudate, suggesting that blood mRNA may reflect disease pathways in the brain. Thus, microarrays of peripheral blood can be used not only to diagnose diseases, but also to subcategorize patients with a specific disease, either by specific subphenotype or by treatment response. This provides another key point – biomarkers identified in chronic neurodegenerative diseases such as inherited ataxias and dementia may provide a more proximal and quantitative end point than clinical scales, which may be less sensitive to therapy, especially in the short term. In this manner, biomarkers of disease progression can be used to determine therapeutic efficacy early in a treatment course, increasing the power and reducing the cost of clinical trials.

#### 1.4.3.2 **Dividing Complex Phenotypes into Subtypes Using Microarrays**

Many other microarray studies in peripheral blood have also found that gene expression profiles alone can subdivide patients with complex diseases. A recent study on Tourette syndrome (TS) found that a subset of patients show overexpression of six cytotoxic T-cell/natural killer cell genes [99]. These results support the hypothesis that pediatric autoimmune neuropsychiatric disorders associated with streptococcal infections (PANDAS) can play a role in the onset of tic disorders and obsessive compulsive disorder (OCD), which is currently unproven [100]. The Sharp lab has also applied this approach to diseases with no obvious blood phenotypes, such as tuberous sclerosis complex 2, neurofibromatosis type 1, and Down's syndrome, finding in all three cases different blood profiles for patients versus controls [101]. Additionally, Down's syndrome patients with congenital heart disease showed different profiles than those without, and tuberous sclerosis patients with autism showed different profiles than those without. These studies suggest that whatever the causes of complex diseases, the resulting phenotypes can be distinguished from one another using blood biomarkers; furthermore, these blood biomarkers may also indicate the underlying mechanisms behind disease phenotypes [93].

In some cases, similar phenotypes may actually show multiple gene expression profiles. For example, a recent study of mRNA expression of leukocytes in patients with major depression disorder revealed 12 genes differentially expressed between depressed patients and controls [94]. Additionally, half of the depressed subjects showed altered expression of many more genes with no corresponding phenotypic

manifestation. As depression is a complex neuropsychiatric disease, there may be multiple pathophysiological conditions that trigger it, possibly partially explaining why only a subset of depressed patients respond to treatment. Another equally complicated phenotype is autism. We have recently used microarrays to study autism cases with single-gene etiologies and validate these pathways in idiopathic autism [115]. In this study, we compared mRNA expression between autistic males with a fragile X mental retardation 1 gene (FMR1-RM) mutation, autistic males with a 15q11–q13 duplication (dup(15q)), and controls. In addition to finding distinct gene expression profiles for each autism group, we also found a set of 68 genes commonly dysregulated in both types of autism, including a potential molecular link between the two autistic groups (CYFIP1), which we confirmed *in vitro*. These results suggest that not only can microarrays be used to subdivide complex phenotypes, but they may also be useful for determining the underlying effects of complex disorders.

#### 1.4.3.3 Applying Blood Classification to Treatments

Peripheral blood profiles can also divide patients on the basis of their responses to treatment. These effects can be seen in two ways: (1) patients taking specific drugs show specific profiles, and (2) patients who respond to treatment show different profiles from those who do not. For example, epileptic children treated with carbamazepine or valproic acid each show characteristic gene expression changes compared with controls [101]. Additionally, children who successfully responded to valproic acid had different peripheral blood profiles from those who still had seizures. The most studied disease in this area of research is multiple sclerosis (MS); reviewed in [102]. In short, MS studies of peripheral blood gene expression changes have found notable results: (1) MS patients have a different profile than controls; (2) remitting patients show different expression patterns than relapsing patients; (3) a subset of the MS profile overlaps with systemic lupus erythematosus, suggesting an ''autoimmune disease'' fingerprint; (4) both interferon-beta and methylprednisolone treatments result in unique molecular signatures; and (5) there were profiles that predicted response to certain MS drugs.

Although most microarray studies require age- and sex-matched controls, some treatment response studies can be self-controlled, removing much of the statistical variability in comparisons. Using this approach, Kalman *et al*. tested the effect of the selective serotonin and noradrenaline reuptake inhibitor, venlafaxine, on gene expression in lymphocytes [103]. Blood was taken from six otherwise healthy, elderly individuals from the same nursing home both before the start of venlafaxine treatment and 4 weeks after the start of treatment. Although this study had a very small sample size, the genetic and environmental factors were both well controlled, since the same individuals were used for both pre- and posttreatment conditions, and since nursing home life placed each subject on the same diet with similar daily routines. Fifty-seven genes related to ionic homeostasis, cell survival, neural plasticity, signal transduction, and metabolism were found to significantly change expression between pre- and posttreatment. These expression changes corresponded to a decrease in clinical levels of depression as measured by the Beck Depression

Inventory. Thus, blood expression profiling has the potential to both help uncover the physiological effects of drugs on the CNS and to separate subjects based on their responses to treatment, thereby preventing the overtreatment of individuals suffering from one of the many neurological disorders.

#### 1.4.3.4 **Advantages and Disadvantages of Using Blood Genomics for Brain Disorders**

Blood is an ideal substrate for clinical use as it can be acquired inexpensively and easily from any living patient. RNA can be stabilized immediately after extraction using vacutainer tubes (such as PAXgene and TEMPUS), reducing or eliminating many of the factors associated with postmortem tissue [92]. Additionally, unique gene expression profiles for many neurological diseases have been found in blood (as described above). Furthermore, many of the same genes are involved in the disease profiles for both blood and brain [96], with some studies showing significant correlations between gene expression changes in both substrates [17]. Thus, not only can blood provide transcriptional biomarkers for disease, subphenotypes, and treatment response, but it may also provide insight into underlying molecular mechanisms.

Brain disorders are not, however, the only factors determining gene expression in blood. In fact, just about any imaginable variable can have this effect, including, among other things, age, sex, race, diet, time of day, medication, exercise, stress, genetics, glucose levels, and time since last meal [92,94]. Conversely, some genes are expressed exclusively in the brain, and therefore will not be induced in the blood. Furthermore, any changes that do occur in the blood generally are on the order of 1.5–2.0-fold, compared with changes as high as 10-fold in cancer genomics – an issue that can be partially circumvented by using pure cell types [93]. Thus, any successful study must have a large enough sample size to account for all of these factors, with control and disease patients carefully matched. In fact, before clinical trials of any sort can be successful, some standardized method for normalizing disease samples as compared to controls must be developed.

#### 1.4.3.5 **Pure Cell Line Assays in Peripheral Blood**

As with postmortem tissue, many transcriptional changes in blood that occur in response to neurological disease or injury are cell type specific. For example, most gene expression changes in acute stroke occur in neutrophils and monocytes, while changes in Tourette syndrome occur mostly in natural killer and cytotoxic CD8 T cells, and those in migraine generally occur in platelets [92]. Most MS studies are performed on peripheral blood mononuclear cells, a subset of blood cells involved in immune response [102]. With such small magnitude gene expression changes found in blood relative to brain, pure cell assays are often essential for peripheral blood profiling. Pure cell lines can also help control for unwanted gene expression variations. Lymphoblasts in particular are useful for genetic research as they are easy to acquire and can be cryopreserved and converted into cell lines, providing an inexhaustible supply of genetic material [104]. In fact, many studies have utilized lymphoblastoid cell lines for pharmacogenetic studies. In short, pure cell lines allow for a more precise understanding of transcriptional changes due to illness or injury.

### 1.4.4 Other Types of Microarrays in Clinical Setting

While this chapter primarily emphasizes mRNA-targeting microarrays, DNA-targeting microarrays also play a significant role in clinical diagnoses in areas such as molecular karyotyping and single nucleotide polymorphism (SNP) genotyping. In general, these arrays include oligos that target specific sequences across the genome. Depending on the specific question asked, current technologies allow the entire genome to be assayed on a single chip at fairly high density.

#### 1.4.4.1 Gene Dose, Molecular Karyotyping, and Chromosomal Abnormalities

High-density DNA arrays have been used since the late 1990s to measure the gene dose effect of specific genomic regions across single chromosomes, as well as the entire genome [105–107]. These arrays allow for mapping of chromosomal copy number, for example, in diseases like Klinefelter's and Down's syndromes. High-throughput assays can simultaneously assess regional duplications, deletions, and amplifications with an average marker spacing of ~25 kb [107]. Such high-density coverage can detect microdeletions associated with developmental disorders, for example, without any need to relate phenotype to genotype [22]. Additionally, by using SNPs as oligonucleotides on microarrays, neutral chromosomal aberrations such as uniparental disomy can also be detected [107,108]. While such abnormalities commonly show no health or developmental effect, imprinting errors in chromosome 15 can lead to Prader–Willi syndrome and Angelman syndrome, and errors in chromosome 11 can lead to Beckwith–Wiedemann syndrome [109]. In short, these increasingly high-throughput arrays overcome many limitations of currently used clinical diagnostic tests, which are often subjective and low throughput, requiring relatively large amounts of DNA.

#### 1.4.4.2 SNP Genotyping and Beyond

As cDNA microarrays can assess the expression of all genes in a single hybridization, whole-genome SNP arrays can assess variability across the entire genome at increasingly high densities. Current arrays from Affymetrix and Illumina can assess over 500 000 SNPs at once, allowing for large-scale association and linkage studies in increasingly short times and at low prices [110,111]. Such studies promote a gradual increase in genome-to-disease associations. Clinical studies have already begun to identify individuals at risk for cardiovascular disease, cancer, diabetes, and deep vein thrombosis, as well as adverse drug reactions [112]. In principle, genomic variation that dictates a drug response or indicates susceptibility to any disease could be optimally assessed using SNP genotyping. Unlike cDNA microarrays in which the readout is scalable, SNP arrays can be designed as binary, where the relevant spot fluoresces or not depending on which nucleotide is present. This binary quality of SNP genotyping results in a low error rate, making such arrays ideal in clinical setting. SNP band hybridization with primer exclusion arrays can also be used for gene resequencing. Affymetrix currently offers over 300 000 bp of resequencing on one array. By measuring signal intensity over a group of adjacent probes, one can

simultaneously assess changes in copy number on any of these SNP array platforms. Such copy number variation regions are a significant source of genetic variation, and this approach will likely be important [113,114].

### 1.4.5 Future Clinical Applications: Pharmacogenomics

Alone and in combination with other systems-level measurements such as neuroimaging, microarray technology provides a valuable tool for optimizing individualized therapies for neurological diseases. In individualized therapy, the patient's genetic and biological makeup is assessed and scanned for potential biomarkers to determine which, if any, treatment is necessary (Figure 1.4) [22]. SNP genotyping can be used to determine susceptibility to various neurological disorders, and then to predict treatment outcome or adverse effects in patients. Transcriptional changes in peripheral blood can be used to diagnose patients and assess the effectiveness of various treatments, both before and after drug administration. Tumors found through imaging and other techniques can be identified and patient prognosis can

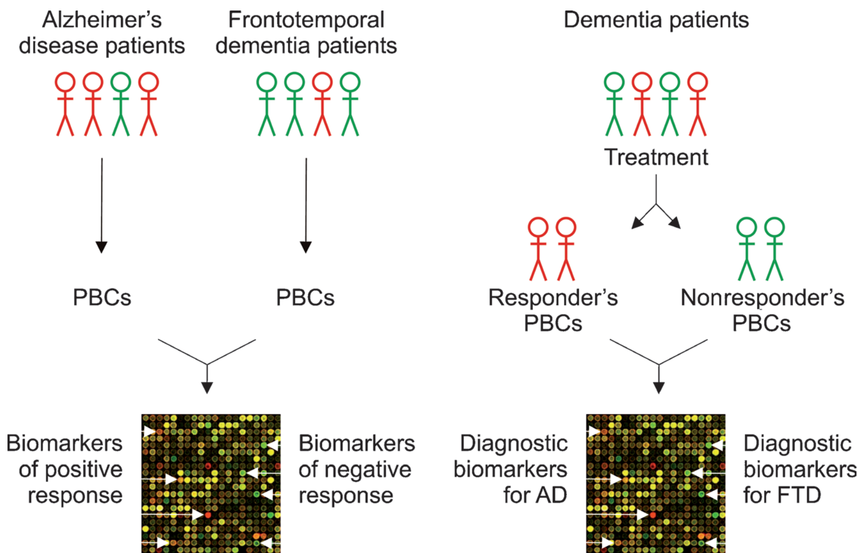


**Figure 1.4** Microarray use for individualized medicine. Left: Microarrays may be used to augment current clinical diagnostic protocols. By hybridizing mRNA from peripheral blood cells (PBCs) of dementia patients onto diagnostic microarrays, clinicians could use the expression pattern to determine whether a patient is more likely to have Alzheimer's disease or frontotemporal dementia. Right: Microarrays may be used to determine effective disease treatments. Before beginning a new treatment, gene expression or SNP microarrays may be used to predict treatment response and possible adverse effects in dementia patients. In theory, such diagnostic, prognostic, and treatment-response predictors could be implemented using microarray technology for any disease or injury with a gene expression signature in PBCs or biopsy tissue.

be determined using microarray technologies. In short, standardized microarray chips will start to be available to identify and properly treat most neurological diseases and injuries within the next decade.

Although a few such clinical tests are available and many other pharmacogenomic studies are underway, the goal of large-scale individualized therapies is still far from attained. Before microarrays reach routine clinical use, batch effects need to be decreased by standardizing the methodology and automating as much of the process as possible [112]. Although such assays have improved in quality since their introduction, there is room for additional development. Also, data mining tools need to provide integration of transcriptional and translational data, in addition to imaging, clinical, and histopathological data. With publication of many new large-scale biomarker assays and with constant advances in microarray reliability and data analysis options, individualized medicine is slowly becoming a realistic tool for the clinician.

## References


1 Sandberg, R., Yasuda, R., Pankratz, D., Carter, T., Del Rio, J., Wodicka, L., Mayford, M., Lockhart, D. and Barlow, C. (2000) Regional and strain-specific gene expression mapping in the adult mouse brain. *Proceedings of the National Academy of Sciences of the United States of America*, **97** (20), 11038–11043.

2 Brown, V., Ossadtchi, A., Khan, A., Cherry, S., Leahy, R. and Smith, D. (2002) High-throughput imaging of brain gene expression. *Genome Research*, **12** (2), 244–254.

3 Arlotta, P., Molyneaux, B., Chen, J., Inoue, J., Kominami, R. and Macklis, J. (2005) Neuronal subtype-specific genes that control corticospinal motor neuron development *in vivo*. *Neuron*, **45** (2), 207–221.

4 Lobo, M., Karsten, S., Gray, M., Geschwind, D. and Yang, W. (2006) FACS-array profiling of striatal projection neuron subtypes in juvenile and adult mouse brains. *Nature Neuroscience*, **9** (3), 443–452.

5 Sugino, K., Hempel, C., Miller, M., Hattox, A., Shapiro, P., Wu, C., Huang, J. and Nelson, S. (2005) Molecular taxonomy of major neuronal classes in the adult mouse forebrain. *Nature Neuroscience*, **9** (1), 99–107.

6 Mischel, P., Cloughesy, T. and Nelson, S. (2004) DNA-microarray analysis of brain cancer: molecular classification for therapy. *Nature Reviews. Neuroscience*, **5** (10), 782–792.

7 Geschwind, D.H., Ou, J., Easterday, M.C., Dougherty, J.D., Jackson, R.L., Chen, Z., Antoine, H., Terskikh, A., Weissman, I.L., Nelson, S.F. and Kornblum, H.I. (2001) A genetic analysis of neural progenitor differentiation. *Neuron*, **29** (2), 325–339.

8 Phillips, R., Ernst, R., Brunk, B., Ivanova, N., Mahan, M., Deanehan, J., Moore, K., Overton, G. and Lemischka, I. (2000) The genetic program of hematopoietic stem cells. *Science*, **288** (5471), 1635–1640.

9 D'Agata, V., Warren, S., Zhao, W., Torre, E., Alkon, D. and Cavallaro, S. (2002) Gene expression profiles in a transgenic animal model of fragile X syndrome. *Neurobiology of Disease*, **10** (3), 211–218.

10 Ginsberg, S., Hemby, S., Lee, V., Eberwine, J. and Trojanowski, J. (2000) Expression profile of transcripts in Alzheimer's disease tangle-bearing CA1 neurons. *Annals of Neurology*, **48** (1), 77–87.

11 Blalock, E., Geddes, J., Chen, K., Porter, N., Markesbery, W. and Landfield, P.

(2004) Incipient Alzheimer's disease: microarray correlation analyses reveal major transcriptional and tumor suppressor responses. *Proceedings of the National Academy of Sciences of the United States of America*, **101** (7), 2173–2178.

**12** Colangelo, V., Schurr, J., Ball, M., Pelaez, R., Bazan, N. and Lukiw, W. (2002) Gene expression profiling of 12633 genes in Alzheimer hippocampal CA1: transcription and neurotrophic factor down-regulation and up-regulation of apoptotic and pro-inflammatory signaling. *Journal of Neuroscience Research*, **70** (3), 462–473.

**13** Emilsson, L., Saetre, P. and Jazin, E. (2006) Alzheimer's disease: mRNA expression profiles of multiple patients show alterations of genes involved with calcium signaling. *Neurobiology of Disease*, **21** (3), 618–625.

**14** Ricciarelli, R., d'Abramo, C., Massone, S., Marinari, U., Pronzato, M. and Tabaton, M. (2004) Microarray analysis in Alzheimer's disease and normal aging. *IUBMB Life*, **56** (6), 349–354.

**15** Loring, J., Wen, X., Lee, J., Seilhamer, J. and Somogyi, R. (2001) A gene expression profile of Alzheimer's disease. *DNA and Cell Biology*, **20** (11), 683–695.

**16** Mandel, S., Grunblatt, E., Riederer, P., Amariglio, N., Jacob-Hirsch, J., Rechavi, G. and Youdim, M. (2005) Gene expression profiling of sporadic Parkinson's disease substantia nigra pars compacta reveals impairment of ubiquitin-proteasome subunits SKP1A, aldehyde dehydrogenase, and chaperone HSC-70. *Annals of the New York Academy of Sciences*, **1053**, 356–375.

**17** Borovecki, F., Lovrecic, L., Zhou, J., Jeong, H., Then, F., Rosas, H.D., Hersch, S.M., Hogarth, P., Bouzou, B., Jensen, R.V. and Krainc, D. (2005) Genome-wide expression profiling of human blood reveals biomarkers for Huntington's disease. *Proceedings of the National Academy of Sciences of the United States of America*, **102** (31), 11023–11028.

**18** Sugars, K. and Rubinsztein, D. (2003) Transcriptional abnormalities in Huntington disease. *Trends in Genetics*, **19** (5), 233–238.

**19** Dangond, F., Hwang, D., Camelo, S., Pasinelli, P., Frosch, M., Stephanopoulos, G., Stephanopoulos, G., Brown, R. and Gullans, S. (2004) Molecular signature of late-stage human ALS revealed by expression profiling of postmortem spinal cord gray matter. *Physiological Genomics*, **16** (2), 229–239.

**20** Jiang, Y.M., Yamamoto, M., Kobayashi, Y., Yoshihara, T., Liang, Y., Terao, S., Takeuchi, H., Ishigaki, S., Katsuno, M., Adachi, H., Niwa, J., Tanaka, F., Doyu, M., Yoshida, M., Hashizume, Y. and Sobue, G. (2005) Gene expression profile of spinal motor neurons in sporadic amyotrophic lateral sclerosis. *Annals of Neurology*, **57** (2), 236–251.

**21** Mirnics, K., Middleton, F., Marquez, A., Lewis, D. and Levitt, P. (2000) Molecular characterization of schizophrenia viewed by microarray analysis of gene expression in prefrontal cortex. *Neuron*, **28** (1), 53–67.

**22** Geschwind, D. (2003) DNA microarrays: translation of the genome from laboratory to clinic. *Lancet Neurology*, **2** (5), 275–282.

**23** Alwine, J., Kemp, D. and Stark, G. (1977) Method for detection of specific RNAs in agarose gels by transfer to diazobenzyloxymethyl-paper and hybridization with DNA probes. *Proceedings of the National Academy of Sciences of the United States of America*, **74** (12), 5350–5354.

**24** Gall, J. and Pardue, M. (1969) Formation and detection of RNA–DNA hybrid molecules in cytological preparations. *Proceedings of the National Academy of Sciences of the United States of America*, **63** (2), 378–383.

**25** Coppola, G. and Geschwind, D.H. (2006) Microarrays and the microscope:

balancing throughput with resolution. *The Journal of Physiology*, **575** (2), 353–359.

**26** Geschwind, D. (2000) Mice, microarrays, and the genetic diversity of the brain. *Proceedings of the National Academy of Sciences of the United States of America*, **97** (20), 10676–10678.

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

**28** Dougherty, J. and Geschwind, D. (2002) Subtraction-coupled custom microarray analysis for gene discovery and gene expression studies in the CNS. *Chemical Senses*, **27** (3), 293–298.

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

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

**31** Brenner, S., Johnson, M., Bridgham, J., Golda, G., Lloyd, D.H., Johnson, D., Luo, S., McCurdy, S., Foy, M., Ewan, M., Roth, R., George, D., Eletr, S., Albrecht, G., Vermaas, E., Williams, S.R., Moon, K., Burcham, T., Pallas, M., DuBridge, R.B., Kirchner, J., Fearon, K., Mao, J. and Corcoran, K. (2000) Gene expression analysis by massively parallel signature sequencing (MPSS) on microbead arrays. *Nature Biotechnology*, **18** (6), 630–634.

**32** Augenlicht, L., Wahrman, M., Halsey, H., Anderson, L., Taylor, J. and 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.

**33** Hardiman, G. (2004) Microarray platforms – comparisons and contrasts. *Pharmacogenomics*, **5** (5), 487–502.

**34** Fodor, S., Read, J., Pirrung, M., Stryer, L., Lu, A. and Solas, D. (1991) Light-directed spatially addressable parallel chemical synthesis. *Science*, **251** (4995), 767–773.

**35** Fodor, S., Rava, R., Huang, X., Pease, A., Holmes, C. and Adams, C. (1993) Multiplexed biochemical assays with biological chips. *Nature*, **364** (6437), 555–556.

**36** Jacobs, J. and Fodor, S. (1994) Combinatorial chemistry – applications of light-directed chemical synthesis. *Trends in Biotechnology*, **12** (1), 19–26.

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

**38** Lipshutz, R., Fodor, S., Gingeras, T. and Lockhart, D. (1999) High density synthetic oligonucleotide arrays. *Nature Genetics*, **21** (1 Suppl.), 20–24.

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

**40** Schena, M., Heller, R., Theriault, T., Konrad, K., Lachenmeier, E. and Davis, R. (1998) Microarrays: biotechnology's discovery platform for functional genomics. *Trends in Biotechnology*, **16** (7), 301–306.

**41** Iida, K. and Nishimura, I. (2002) Gene expression profiling by DNA microarray technology. *Critical Reviews in Oral Biology and Medicine*, **13** (1), 35–50.

**42** Eisen, M. and Brown, P. (1999) DNA arrays for analysis of gene expression. *Methods in Enzymology*, **303**, 179–205.

**43** Brown, P. and Botstein, D. (1999) Exploring the new world of the genome with DNA microarrays. *Nature Genetics*, **21** (1 Suppl.), 33–37.

**44** Wolber, P., Collins, P., Lucas, A., De Witte, A. and Shannon, K. (2006) The

Agilent *in situ*-synthesized microarray platform. *Methods in Enzymology*, **410**, 28–57.

**45** Fan, J., Gunderson, K., Bibikova, M., Yeakley, J., Chen, J., Wickham Garcia, E., Lebruska, L., Laurent, M., Shen, R. and Barker, D. (2006) Illumina universal bead arrays. *Methods in Enzymology*, **410**, 57–73.

**46** Gunderson, K.L., Kruglyak, S., Graige, M.S., Garcia, F., Kermani, B.G., Zhao, C., Che, D., Dickinson, T., Wickham, E., Bierle, J., Doucet, D., Milewski, M., Yang, R., Siegmund, C., Haas, J., Zhou, L., Oliphant, A., Fan, J.B., Barnard, S. and Chee, M.S. (2004) Decoding randomly ordered DNA arrays. *Genome Research*, **14** (5), 870–877.

**47** Mirnics, K. and Pevsner, J. (2004) Progress in the use of microarray technology to study the neurobiology of disease. *Nature Neuroscience*, **7** (5), 434–439.

**48** Karsten, S., Kudo, L. and Geschwind, D. (2004) Microarray platforms: introduction and application to neurobiology. *International Review of Neurobiology*, **60**, 1–23.

**49** Lockhart, D. and Barlow, C. (2001) Expressing what's on your mind: DNA arrays and the brain. *Nature Reviews. Neuroscience*, **2** (1), 63–68.

**50** Luo, Z. and Geschwind, D. (2001) Microarray applications in neuroscience. *Neurobiology of Disease*, **8** (2), 183–193.

**51** Quackenbush, J. (2006) Microarray analysis and tumor classification. *The New England Journal of Medicine*, **354** (23), 2463–2472.

**52** Huber, W., Irizarry, R. and Gentleman, R. (2005) *Bioinformatics and Computational Biology Solutions Using R and Bioconductor (Statistics for Biology and Health)*, Springer.

**53** Choe, S., Boutros, M., Michelson, A., Church, G. and Halfon, M. (2005) Preferred analysis methods for Affymetrix GeneChips revealed by a wholly defined control dataset. *Genome Biology*, **6** (2), R16.1–R16.16.

**54** Nuwaysir, E.F., Huang, W., Albert, T.J., Singh, J., Nuwaysir, K., Pitas, A., Richmond, T., Gorski, T., Berg, J.P., Ballin, J., McCormick, M., Norton, J., Pollock, T., Sumwalt, T., Butcher, L., Porter, D., Molla, M., Hall, C., Blattner, F., Sussman, M.R., Wallace, R.L., Cerrina, F. and Green, R.D. (2002) Gene expression analysis using oligonucleotide arrays produced by maskless photolithography. *Genome Research*, **12** (11), 1749–1755.

**55** Vawter, M., Crook, J., Hyde, T., Kleinman, J., Weinberger, D., Becker, K. and Freed, W. (2002) Microarray analysis of gene expression in the prefrontal cortex in schizophrenia: a preliminary study. *Schizophrenia Research*, **58** (1), 11–20.

**56** Schwartz, A. and Fernändez-Repollet, E. (2001) Quantitative flow cytometry. *Clinics in Laboratory Medicine*, **21** (4), 743–761.

**57** Heintz, N. (2004) Gene expression nervous system atlas (GENSAT). *Nature Neuroscience*, **7** (5), 483.

**58** Ginsberg, S., Elarova, I., Ruben, M., Tan, F., Counts, S., Eberwine, J., Trojanowski, J., Hemby, S., Mufson, E. and Che, S. (2004) Single-cell gene expression analysis: implications for neurodegenerative and neuropsychiatric disorders. *Neurochemical Research*, **29** (6), 1053–1064.

**59** Kacharmina, J., Crino, P. and Eberwine, J. (1999) Preparation of cDNA from single cells and subcellular regions. *Methods in Enzymology*, **303**, 3–18.

**60** Emmert-Buck, M., Bonner, R., Smith, P., Chuaqui, R., Zhuang, Z., Goldstein, S., Weiss, R. and Liotta, L. (1996) Laser capture microdissection. *Science*, **274** (5289), 998–1001.

**61** Li, C. and Wong, W. (2001) Model-based analysis of oligonucleotide arrays: model validation design issues and standard error application. *Genome Biology*, **2** (8), 1–11. research0032.0031–research0032.0011.

**62** Medigue, C., Rechenmann, F., Danchin, A. and Viari, A. (1999) Imagene: an integrated computer environment for sequence annotation and analysis. *Bioinformatics*, **15** (1), 2–15.

**63** Evans, S.J., Choudary, P.V., Vawter, M.P., Li, J., Meador-Woodruff, J.H., Lopez, J.F., Burke, S.M., Thompson, R.C., Myers, R.M., Jones, E.G. Bunney, W.E., Watson, S.J. and Akil, H. (2003) DNA microarray analysis of functionally discrete human brain regions reveals divergent transcriptional profiles. *Neurobiology of Disease*, **14** (2), 240–250.

**64** Lu, T., Pan, Y., Kao, S.-Y., Li, C., Kohane, I., Chan, J. and Yankner, B. (2004) Gene regulation and DNA damage in the ageing human brain. *Nature*, **429** (6994), 883–891.

**65** Ghazalpour, A., Doss, S., Zhang, B., Wang, S., Plaisier, C., Castellanos, R., Brozell, A., Schadt, E.E., Drake, T.A., Lusis, A.J. and Horvath, S. (2006) Integrating genetic and network analysis to characterize genes related to mouse weight. *PLoS Genetics*, **28**, 1182–1192.

**66** Reimers, M. and Carey, V. (2006) Bioconductor: an open source framework for bioinformatics and computational biology. *Methods in Enzymology*, **411**, 119–134.

**67** Hosack, D., Dennis, G., Sherman, B., Lane, H. and Lempicki, R. (2003) Identifying biological themes within lists of genes with EASE. *Genome Biology*, **4** (10), R70.1–R70.8.

**68** Subramanian, A., Tamayo, P., Mootha, V.K., Mukherjee, S., Ebert, B.L., Gillette, M.A., Paulovich, A., Pomeroy, S.L., Golub, T.R., Lander, E.S. and Mesirov, J.P. (2005) From the cover: Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proceedings of the National Academy of Sciences of the United States of America*, **102** (43), 15545–15550.

**69** Zhang, B., Kirov, S. and Snoddy, J. (2005) WebGestalt: an integrated system for exploring gene sets in various biological contexts. *Nucleic Acids Research*, **33** (Web Server issue, W741–W748).

**70** Chen, H. and Sharp, B. (2004) Content-rich biological network constructed by mining PubMed abstracts. *BMC Bioinformatics*, **5** (1), 1–13.

**71** Rebhan, M., Chalifa-Caspi, V., Prilusky, J. and Lancet, D. (1997) GeneCards: encyclopedia for genes proteins and diseases. World Wide Web URL:http://wwwgenecardsorg/.

**72** Barabasi, A. and Oltvai, Z. (2004) Network biology: understanding the cell's functional organization. *Nature Reviews. Genetics*, **5** (2), 101–113.

**73** Jeong, H., Mason, S., Barabasi, A. and Oltvai, Z. (2001) Lethality and centrality in protein networks. *Nature*, **411** (6833), 41–42.

**74** Oldham, M., Horvath, S. and Geschwind, D. (2006) Conservation and evolution of gene co-expression networks in human and chimpanzee brain. *Proceedings of the National Academy of Sciences of the United States of America*, **103**, 17973–17978.

**75** Zhang, B. and Horvath, S. (2005) A general framework for weighted gene co-expression network analysis. *Statistical Applications in Genetics and Molecular Biology*, **4** (1), 1–43.

**76** Horvath, S., Zhang, B., Carlson, M., Lu, K.V., Zhu, S., Felciano, R.M., Laurance, M.F., Zhao, W., Qi, S., Chen, Z., Lee, Y., Scheck, A.C., Liau, L.M., Wu, H., Geschwind, D.H., Febbo, P.G., Kornblum, H.I., Cloughesy, T.F., Nelson, S.F. and Mischel, P.S. (2006) Analysis of oncogenic signaling networks in glioblastoma identifies ASPM as a molecular target. *Proceedings of the National Academy of Sciences of the United States of America*, **103** (46), 17402–17407.

**77** Bustin, S. (2000) Absolute quantification of mRNA using real-time reverse transcription polymerase chain reaction assays. *Journal of Molecular Endocrinology*, **25** (2), 169–193.

**78** Lock, C., Hermans, G., Pedotti, R., Brendolan, A., Schadt, E., Garren, H., Langer-Gould, A., Strober, S., Cannella, B., Allard, J., Klonowski, P., Austin, A., Lad, N., Kaminski, N., Galli, S.J., Oksenberg, J.R., Raine, C.S., Heller, R. and Steinman, L. (2002) Gene-microarray analysis of multiple sclerosis lesions yields new targets validated in autoimmune encephalomyelitis. *Nature Medicine*, **8** (5), 500–508.

**79** Karsten, S.L., Sang, T.K., Gehman, L.T., Chatterjee, S., Liu, J., Lawless, G.M., Sengupta, S., Berry, R.W., Pomakian, J., Oh, H.S. Schulz, C., Hui, K.S., Wiedau-Pazos, M., Vinters, H.V., Binder, L.I., Geschwind, D.H. and Jackson, G.R. (2006) A genomic screen for modifiers of tauopathy identifies puromycin-sensitive aminopeptidase as an inhibitor of tau-induced neurodegeneration. *Neuron*, **51** (5), 549–560.

**80** Lee, H., Hsu, A., Sajdak, J., Qin, J. and Pavlidis, P. (2004) Coexpression analysis of human genes across many microarray data sets. *Genome Research*, **14** (6), 1085–1094.

**81** Bailey, P. and Cushing, H. (1928) *A Classification of the Tumors of the Glioma Group on a Histogenic Basis with a Correlated Study of Prognosis*, Lippincott.

**82** Shai, R., Shi, T., Kremen, T., Horvath, S., Liau, L., Cloughesy, T., Mischel, P. and Nelson, S. (2003) Gene expression profiling identifies molecular subtypes of gliomas. *Oncogene*, **22** (31), 4918–4923.

**83** Freije, W., Castro-Vargas, F., Fang, Z., Horvath, S., Cloughesy, T., Liau, L., Mischel, P. and Nelson, S. (2004) Gene expression profiling of gliomas strongly predicts survival. *Cancer Research*, **64** (18), 6503–6510.

**84** Sarkar, C., Deb, P. and Sharma, M. (2006) Medulloblastomas: new directions in risk stratification. *Neurology India*, **54** (1), 16–23.

**85** Gajjar, A., Hernan, R., Kocak, M., Fuller, C., Lee, Y., McKinnon, P.J., Wallace, D., Lau, C., Chintagumpala, M., Ashley, D.M., Kellie, S.J., Kun, L. and Gilbertson, R.J. (2004) Clinical, histopathologic, and molecular markers of prognosis: toward a new disease risk stratification system for medulloblastoma. *Journal of Clinical Oncology*, **22** (6), 984–993.

**86** Pomeroy, S.L., Tamayo, P., Gaasenbeek, M., Sturla, L.M., Angelo, M., Mclaughlin, M.E., Kim, J.Y., Goumnerova, L.C., Black, P.M., Lau, C., Allen, J.C., Zagzag, D., Olson, J.M., Curran, T., Wetmore, C., Biegel, J.A., Poggio, T., Mukherjee, S., Rifkin, R., Califano, A., Stolovitzky, G., Louis, D.N., Mesirov, J.P., Lander, E.S. and Golub, T.R. (2002) Prediction of central nervous system embryonal tumour outcome based on gene expression. *Nature*, **415** (6870), 436–442.

**87** Sowar, K., Straessle, J., Donson, A., Handler, M. and Foreman, N. (2006) Predicting which children are at risk for ependymoma relapse. *Journal of Neuro-Oncology*, **78** (1), 41–46.

**88** Gyorffy, B., Surowiak, P., Kiesslich, O., Denkert, C., Schäfer, R., Dietel, M. and Lage, H. (2005) Gene expression profiling of 30 cancer cell lines predicts resistance towards 11 anticancer drugs at clinically achieved concentrations. *International Journal of Cancer*, **118** (7), 1699–1712.

**89** Staunton, J.E., Slonim, D.K., Coller, H.A., Tamayo, P., Angelo, M.J., Park, J., Scherf, U., Lee, J.K., Reinhold, W.O., Weinstein, J.N., Mesirov, J.P., Lander, E.S. and Golub, T.R. (2001) Chemosensitivity prediction by transcriptional profiling. *Proceedings of the National Academy of Sciences of the United States of America*, **98** (19), 10787–10792.

**90** Stegmaier, K., Ross, K., Colavito, S., O'Malley, S., Stockwell, B. and Golub, T. (2004) Gene expression-based high-throughput screening (GE-HTS) and application to leukemia differentiation. *Nature Genetics*, **36** (3), 257–263.

**91** Cheok, M., Yang, W., Pui, C., Downing, J., Cheng, C., Naeve, C., Relling, M. and Evans, W. (2003) Treatment-specific changes in gene expression discriminate *in vivo* drug response in human leukemia cells. *Nature Genetics*, **34** (1), 85–90.

**92** Sharp, F.R., Xu, H., Lit, L., Walker, W., Apperson, M., Gilbert, D.L., Glauser, T.A., Wong, B., Hershey, A., Liu, D.Z. Pinter, J., Zhan, X., Liu, X. and Ran, R. (2006) The future of genomic profiling of neurological diseases using blood. *Archives of Neurology*, **63** (11), 1529–1536.

**93** Sharp, F.R., Lit, L., Xu, H., Apperson, M., Walker, W., Wong, B., Gilbert, D.L., Hershey, A. and Glauser, T.A. (2006) Genomics of brain and blood: progress and pitfalls. *Epilepsia*, **47** (10), 1603–1607.

**94** Ohmori, T., Morita, K., Saito, T., Ohta, M., Ueno, S. and Rokutan, K. (2005) Assessment of human stress and depression by DNA microarray analysis. *The Journal of Medical Investigation*, **52** (Suppl.), 266–271.

**95** Tang, Y., Lu, A., Aronow, B. and Sharp, F. (2001) Blood genomic responses differ after stroke seizures, hypoglycemia, and hypoxia: blood genomic fingerprints of disease. *Annals of Neurology*, **50** (6), 699–707.

**96** Tang, Y., Lu, A., Aronow, B., Wagner, K. and Sharp, F. (2002) Genomic responses of the brain to ischemic stroke intracerebral haemorrhage, kainate seizures, hypoglycemia, and hypoxia. *The European Journal of Neuroscience*, **15** (12), 1937–1952.

**97** Moore, D.F., Li, H., Jeffries, N., Wright, V., Cooper, R.A. Jr, Elkahloun, A., Gelderman, M.P., Zudaire, E., Blevins, G., Yu, H., Goldin, E. and Baird, A.E. (2005) Using peripheral blood mononuclear cells to determine a gene expression profile of acute ischemic stroke: a pilot investigation. *Circulation*, **111** (2), 212–221.

**98** Tang, Y., Xu, H., Du, X., Lit, L., Walker, W., Lu, A., Ran, R., Gregg, J.P., Reilly, M., Pancioli, A., Khoury, J.C., Sauerbeck, L.R., Carrozzella, J.A., Spilker, J., Clark, J., Wagner, K.R., Jauch, E.C., Chang, D.J., Verro, P., Broderick, J.P. and Sharp, F.R. (2006) Gene expression in blood changes rapidly in neutrophils and monocytes after ischemic stroke in humans: a microarray study. *Journal of Cerebral Blood Flow and Metabolism*, **26** (8), 1089–1102.

**99** Tang, Y., Gilbert, D., Glauser, T., Hershey, A. and Sharp, F. (2005) Blood gene expression profiling of neurologic diseases: a pilot microarray study. *Archives of Neurology*, **62** (2), 210–215.

**100** Kurlan, R. and Kaplan, E. (2004) The pediatric autoimmune neuropsychiatric disorders associated with streptococcal infection (PANDAS) etiology for tics and obsessive-compulsive symptoms: hypothesis or entity? Practical considerations for the clinician. *Pediatrics*, **113** (4), 883–886.

**101** Tang, Y., Schapiro, M.B., Franz, D.N., Patterson, B.J., Hickey, F.J., Schorry, E.K., Hopkin, R.J., Wylie, M., Narayan, T., Glauser, T.A., Gilbert, D.L., Hershey, A.D. and Sharp, F.R. (2004) Blood expression profiles for tuberous sclerosis complex 2, neurofibromatosis type 1, and Down's syndrome. *Annals of Neurology*, **56** (6), 808–814.

**102** Achiron, A. and Gurevich, M. (2006) Peripheral blood gene expression signature mirrors central nervous system disease: the model of multiple sclerosis. *Autoimmunity Reviews*, **5** (8), 517–522.

**103** Kálmán, J., Palotás, A., Juhász, A., Rimanóczy, A., Hugyecz, M., Kovács, Z., Galsi, G., Szabó, Z., Pákáski, M., Fehér, L.Z., Janka, Z. and Puskǎs, L.G. (2005) Impact of venlafaxine on gene expression profile in lymphocytes of the elderly with major depression evolution of antidepressants and the role of the neuro-immune system. *Neurochemical Research*, **30** (11), 1429–1438.

**104** Shukla, S. and Dolan, M. (2005) Use of CEPH and non-CEPH lymphoblast cell lines in pharmacogenetic studies. *Pharmacogenomics*, **6** (3), 303–310.

**105** Geschwind, D.H., Gregg, J., Boone, K., Karrim, J., Pawlikowska-Haddal, A., Rao, E., Ellison, J., Ciccodicola, A., D'Urso, M., Woods, R., Rappold, G.A., Swerdloff, R. and Nelson, S.F. (1998) Klinefelter's syndrome as a model of anomalous cerebral laterality: testing gene dosage in the X chromosome pseudoautosomal region using a DNA microarray. *Developmental Genetics*, **23** (3), 215–229.

**106** Cheung, V.G., Nowak, N., Jang, W., Kirsch, I.R., Zhao, S., Chen, X.N., Furey, T.S., Kim, U.J., Kuo, W.L., Olivier, M. Conroy, J., Kasprzyk, A., Massa, H., Yonescu, R., Sait, S., Thoreen, C., Snijders, A., Lemyre, E., Bailey, J.A., Bruzel, A., Burrill, W.D., Clegg, S.M., Collins, S., Dhami, P., Friedman, C., Han, C.S., Herrick, S., Lee, J., Ligon, A.H., Lowry, S., Morley, M., Narasimhan, S., Osoegawa, K., Peng, Z., Plajzer-Frick, I., Quade, B.J., Scott, D., Sirotkin, K., Thorpe, A.A., Gray, J.W., Hudson, J., Pinkel, D., Ried, T., Rowen, L., Shen-Ong, G.L., Strausberg, R.L., Birney, E., Callen, D.F., Cheng, J.F., Cox, D.R., Doggett, N.A., Carter, N.P., Eichler, E.E., Haussler, D., Korenberg, J.R., Morton, C.C., Albertson, D., Schuler, G., de Jong, P.J. and Trask, B.J.,BAC Resource Consortium. (2001) Integration of cytogenetic landmarks into the draft sequence of the human genome. *Nature*, **409** (6822), 953–958.

**107** Slater, H., Bailey, D., Ren, H., Cao, M., Bell, K., Nasioulas, S., Henke, R., Choo, K. and Kennedy, G. (2005) High-resolution identification of chromosomal abnormalities using oligonucleotide arrays containing 116204 SNPs. *American Journal of Human Genetics*, **77** (5), 709–726.

**108** Fan, J.-B., Chee, M. and Gunderson, K. (2006) Highly parallel genomic assays. *Nature Reviews. Genetics*, **7** (8), 632–644.

**109** Lalande, M. (1996) Parental imprinting and human disease. *Annual Review of Genetics*, **30**, 173–195.

**110** Gunderson, K.L., Kuhn, K.M., Steemers, F.J., Ng, P., Murray, S.S. and Shen, R. (2006) Whole-genome genotyping of haplotype tag single nucleotide polymorphisms. *Pharmacogenomics*, **7** (4), 641–648.

**111** Komura, D., Shen, F., Ishikawa, S., Fitch, K.R., Chen, W., Zhang, J., Liu, G., Ihara, S., Nakamura, H., Hurles, M.E. Lee, C., Scherer, S.W., Jones, K.W., Shapero, M.H., Huang, J. and Aburatani, H. (2006) Genome-wide detection of human copy number variations using high-density DNA oligonucleotide arrays. *Genome Research*, **16** (12), 1575–1584.

**112** Weeraratna, A., Nagel, J., de Mello-Coelho, V. and Taub, D. (2004) Gene expression profiling: from microarrays to medicine. *Journal of Clinical Immunology*, **24** (3), 213–224.

**113** Redon, R., Ishikawa, S., Fitch, K.R., Feuk, L., Perry, G.H., Andrews, T.D., Fiegler, H., Shapero, M.H., Carson, A.R., Chen, W., Cho, E.K., Dallaire, S., Freeman, J.L., Gonzalez, J.R., Gratacos, M., Huang, J., Kalaitzopoulos, D., Komura, D., MacDonald, J.R., Marshall, C.R., Mei, R., Montgomery, L., Nishimura, K., Okamura, K., Shen, F., Somerville, M.J., Tchinda, J., Valsesia, A., Woodwark, C., Yang, F., Zhang, J., Zerjal, T., Zhang, J., Armengol, L., Conrad, D.F., Estivill, X., Tyler-Smith, C., Carter, N.P., Aburatani, H., Lee, C., Jones, K.W., Scherer, S.W. and Hurles, M.E.,( (2006) Global variation in copy number in the human genome. *Nature*, **444** (7118), 444–454.

**114** Sebat, J., Lakshmi, B., Troge, J., Alexander, J., Young, J., Lundin, P., Maner, S., Massa, H., Walker, M., Chi, M., Navin, N., Lucito, R., Healy, J., Hicks, J., Ye, K., Reiner, A., Gilliam, T.C., Trask,

B., Patterson, N., Zetterberg, A. and Wigler, M. (2004) Large-scale copy number polymorphism in the human genome. *Science*, **305** (5683), 525–528.

**115** Nishimura, Y., Martin, C., Vazquez-Lopez, A., Spence, S., Alvarez-Retuerto, A., Sigman, M., Steindler, C., Pellegrini, S., Schanen, N., Warren, S. and Geschwind, D. (2004) Genome-wide expression profiling of lymphoblastoid cell lines distinguishes different forms of autism and reveals shared pathways. *Human Molecular Genetics*, **16** (14), 1682–1698.