Progress in the use of microarray technology to study the neurobiology of disease

Károly Mirnics & Jonathan Pevsner

The diverse functions of the brain are mediated by neurons and glia whose phenotype is defined by a dynamically maintained set of gene transcripts, or 'transcriptome'. Large-scale analysis of gene expression in postmortem brain using microarray technology has the potential to elucidate molecular changes that occur in disease states. There are unique challenges associated with studies of postmortem brain, including limited sample sizes and variable clinical phenotypes that are typical of complex disorders. Nevertheless, recent microarray-based studies have implicated both individual dysregulated genes and abnormal patterns of gene expression in brain disorders.

Although the human genome encodes some 30,000 genes, the phenotype of each cell is specified by the selective expression of a subset of this number. Gene expression is choreographed by context: each gene produces its RNA transcript at specific times of development, in particular regions of the body, and under certain physiological conditions.

Neuroscientists are interested in characterizing gene expression in the brain for two principal reasons. First, the brain may show a more intricate pattern of gene expression than do other body regions, thus enabling the varied functions of neurons, from the collection of sensory input, to the processing of thoughts, feelings and memories, to the coordination of motor output.

Second, gene expression varies in human disease. The cause of a brain disease may be a mutation in a single gene (as in fragile X syndrome), infectious agents (viruses), environmental toxins (lead poisoning) or genomic rearrangements. Furthermore, the etiology of the most complex brain disorders may involve interaction of multiple genetic factors with environmental influences (as is probably true for schizophrenia or autism). In each of these cases, brain cells respond to the disease state by altering their transcriptional program. An examination of the transcriptional response may inform us about the pathophysiology of disease.

For essentially all human diseases, there is a major gap in our understanding of the relation between genotype and phenotype. DNA microarrays enable genome-wide measurements of transcription, and possibly a bridge between genotype and phenotype.

Karoly Mirnics is in the Department of Psychiatry, University of Pittsburgh, School of Medicine, E1453 Biomedical Science, Pittsburgh, Pennsylvania 15261, USA. Jonathan Pevsner is in the Department of Neurology, Kennedy Krieger Institute, and the Department of Neuroscience, Johns Hopkins School of Medicine, 707 N. Broadway, Baltimore, Maryland 21205, USA.
e-mail: karoly+@pitt.edu or pevsner@jhmi.edu

Microarray technology

The scientific principle underlying microarray technology is complementary hybridization between nucleic acids[1,2]. All gene expression DNA microarrays can be understood as high-throughput 'dot-blot' systems, where pieces of known DNA are anchored to a solid support, while targets are fluorescently labeled, free-floating amplified RNA or complementary DNA (cDNA) species originating from the samples[3]. When the labeled sample is hybridized to the DNA microarray, each probe binds its complementary target. Analyzing the microarray with high-resolution fluorescent scanners allows assessment of the fluorescent signal strength that originates from the probe-bound target. This signal is presumed to be directly proportional to the abundance of the RNA species present in the investigated samples.

Most commonly, one sample is hybridized to one array, and the results are standardized and mathematically compared across microarrays, uncovering fluorescent intensity differences between microarrays. These intensity differences correspond to transcript abundance differences between samples. Current microarrays contain probes corresponding to many thousands of annotated genes in the human genome, allowing 'transcriptome profiling' from each of the samples. A variety of microarray platforms are available (using oligonucleotides or cDNAs; see ref. 4). The choice of platform depends on factors like the number of genes represented on the array, cost and availability.

However, microarrays are more than a simple collection of independently performed dot blots for thousands of genes. Many expression changes are correlated in ways that suggest a causal dependence[5], and changes in relative abundances within individual samples provide valuable information about the complex pathways and cellular processes that are altered in a particular disease or condition. Furthermore, defined disease-specific expression patterns can be correlated with relevant pre-mortem information (for example, cognitive performance or medication history)[6].

Challenges in analyzing human brain tissue

Working with human postmortem brain material has a major influence on experimental design, data analysis and interpretation for several reasons[7–9]. First, DNA microarray analysis of brain tissue is highly complex. Although the brain has a limited number of primary cell types, these show immense phenotypic diversity, and gene expression changes may affect only subpopulations of cells. Consequently, even profound transcriptome changes in a small subpopulation of brain cells may not be detected: more abundant sources of transcripts can mask these changes (for review, see ref. 5). In addition, *de novo* expression, induction and repression are rarely

seen in the mature nervous system. As a result, the magnitude of expression changes found with microarrays is often only modest and hard to separate from experimental noise. Finally, as many neurons project to remote areas, gene expression changes and protein changes may occur in different brain compartments. The vast majority of transcripts are most abundant in the cell soma, whereas the protein is often localized to axonal projections or nerve terminals a considerable distance away, complicating interpretation[10]. Moreover, it is possible for the levels of a transcript to be up- or down-regulated while the protein is significantly regulated in the opposite direction. Combined study of RNA and protein levels using high-throughput approaches is desirable whenever possible, although in general these levels show limited correlation[11].

Second, the genetic background and lifestyle of humans is diverse. This gives rise to a new level of complexity that is usually not encountered in well-controlled animal experiments. The genetic makeup of each individual is unique, and so is the transcriptome profile. Experiences also shape our brain transcriptome individually. Thus, functionally important differences in gene expression that are relevant to disease processes may be masked by other changes. Unfortunately, we have no firm estimates of combined genetic and epigenetic influences on the human brain transcriptome.

Third, the transcriptome is also shaped by disease treatment. Separating the effects of the disease from the effects of treatment remains one of the most challenging aspects of human postmortem research. Postmortem material from unmedicated patients is rarely available, making interpretation difficult. Rodent animal models and tissue culture systems are extremely valuable in assessing the effects of medication on the neural transcriptome. It is unclear, however, how well the animal data translate to drug-treatment effects in the diseased human brain. Transcriptome analyses in non-human primate models of chronic drug treatment are probably the most informative in this regard[12], although this approach is expensive and low-throughput.

Fourth, brain disorders are typically not uniform within a diagnosis class. For example, schizophrenia and Alzheimer disease show a continuum of clinical phenotypes[13]. Even for single-gene disorders such as adrenoleukodystrophy (the most common genetic disorder affecting peroxisomes), a single mutation leads to a broad range of clinical phenotypes. This disease continuum is also apparent at the level of the transcriptome.

Fifth, the availability of postmortem material is limited. Brain tissue donations from well-characterized subjects are hard to obtain[14]. As a result, without a coordinated effort, studies are done on limited sample sizes, seriously restricting their power. Furthermore, there is extreme diversity among the brains with respect to age, race, postmortem interval, medication history, lifestyle and other factors. This diversity, combined with co-morbidity with other disorders, represents a significant challenge in the interpretation. As an example, about 8% of individuals with Down syndrome also have autism.

Sixth, the age of death is not controlled. For developmental brain disorders, the available tissue is typically from older patients, long after the initial insult occurred. Thus, samples may reflect the state of the brain as it adapts to the disease state over time, rather than reflecting immediate downstream consequences of a disease-causing condition. Similarly, it is hard to separate neurodegenerative disease from the normal progression of aging.

Seventh, sample integrity is the most critical aspect in postmortem brain research. Although these samples are characterized by postmortem intervals (PMI) of up to 30 hours, most of the postmortem brains contain high-integrity mRNA that is suitable for DNA

microarray analysis. However, PMI alone is a poor predictor of sample integrity[15]. RNA integrity (measured indirectly by brain pH levels) primarily depends on the circumstances of death. Samples of pH less than 6.25 rarely contain intact RNA, and high-quality microarray analysis of these samples is usually not feasible.

Data analysis challenges

There are three main phases to microarray data analysis: preprocessing, inferential statistics and descriptive (exploratory) statistics. These phases of analysis are used to answer some of the key questions typically posed by biologists using microarrays. Which genes are significantly up- or down-regulated in the brain? Are there regional or temporal differences in gene expression? Can we use gene expression data to classify samples as derived from patients with a brain disease or from matched controls, or to subclassify patient-derived samples?

Gene expression variations between samples are attributable to a combination of biological differences and experimental artifacts. The latter include variations associated with the sample (differences in the way RNA samples are isolated and processed, or different labeling efficiencies of fluorescently tagged nucleotides), the array (uneven spotting of DNA elements) or the hardware (variable performance of fluorescence scanners). Non-biological variations in gene expression can be reduced through proper experimental design (for example, by processing experimental and control samples in parallel, using microarrays from the same lot or by using dye swap experiments). Data normalization can also be applied to remove systematic biases in the data[16]: global normalization allows the comparison of data from two microarrays, and local normalization accounts for artifactual variations that are not constant across a range of signal intensities or across the surface of a microarray[17] (www.snomad.org).

A fundamental challenge for researchers using microarrays is that there is currently no consensus for the appropriate data normalization procedures. We believe that the distribution of data should be normalized around zero, and local normalization procedures should be applied to datasets to account for gene expression values that change as a function of signal intensity. However, such corrections are not consistently applied. If three normalization procedures are applied to the same raw data set, it is likely that entirely distinct descriptions of regulated genes will be generated. Software such as the bioinformatics tools of the Bioconductor project (www.bioconductor.org) helps researchers find common approaches for microarray data analysis.

The goal of the second phase, inferential statistics, is to evaluate hypotheses about gene expression changes in terms of significance and confidence. We may state the null hypothesis that a given gene is not differentially regulated in five brain samples from individuals with schizophrenia relative to a similar number from matched controls, and then test whether we can reject it with a probability $P < 0.05$. However, there is little consensus on how the significance of gene expression changes should be applied.

Microarray data analysis requires understanding false-positive (type-I) and false-negative (type-II) errors. We usually do not hypothesize *a priori* that an individual gene is significantly regulated in a comparison of two conditions (normal versus brain disease), but instead we observe whether any gene is differentially expressed. In general, to reduce type-I errors, a correction for multiple measurements is applied. For example, the Bonferroni correction accounts for the number of measurements (it establishes a new significance threshold by dividing the desired significance by the number of expression measurements) and represents a very conservative approach to data analysis. However, it may also eliminate most of the

'biologically true' expression changes from further consideration, leading to a large increase in type-II errors. In contrast, recently developed microarray software analysis tools (such as Significance Analysis of Microarrays (SAM)[18] and dChip[19]) advocate a more lenient calculation of a false discovery rate of regulated genes, thus reducing type-II (but increasing type-I) error rates. These tools are useful for hypothesis testing even with small sample sizes.

Unfortunately, in many microarray publications, significance is not tested at all. This is problematic because it reflects no hypothesis testing, and such publications often involve inadequately small sample sizes. For example, a comparison of two normal to two diseased samples is likely to involve too few samples to permit an appropriate *t*-test or analysis of variance (ANOVA) that is used to generate *P* values. As a result, these publications report fold-changes for genes that may be differentially expressed. Although such information can be useful, a large fold-change can occur without statistical significance, or very small expression changes can be statistically significant without necessarily being biologically important.

The goal of exploratory statistics is to find and visualize patterns and structures in the data. These structures may be groups of genes and/or samples with some similarity in their expression profiles. The groups may be predefined (control versus diseased samples) or they may be defined by the pattern of expression data. For example, within a group with similar clinical phenotypes, one can recognize new disease subtypes based on microarray data, which may then turn out to have different clinical properties (prognosis or response to medication) that would not otherwise have been recognized. A common form of exploratory analysis is clustering[20,21]. Clustering requires a measure of the distance between objects (such as genes or samples), thus quantifying their similarity. An algorithm then generates a graph depicting these relationships. This approach has been used most successfully in classifying cell lines, including those derived from tumors[22], but it is less frequently used to analyze gene expression changes in human brain[9,23,24].

Transcriptome discoveries to date

Despite the challenges, microarray studies using human brain tissue have generated very interesting leads, which should be judged by the impact of the follow-up studies. Thus, the importance of the findings to our understanding of the disease process is hard to assess on a short-term basis.

Studies of schizophrenia implicate changes in presynaptic release[25,26], metabolic pathways[27], GABA-glutamate transcripts[25,28] and altered glial function[10,29–31]. Inferential statistics also suggest that schizophrenia is associated with complex molecular disturbances in a variety of cell types, including expression changes in genes encoding regulator of G-protein signaling 4 (RGS4), neuroserpine, AMPA-2 receptor, GAD67, N-ethylmaleimide sensitive factor, 14-3-3 isoforms, malate dehydrogenase 1, proteolipid protein 1, ErbB3, transferrin, myelin-associated glycoprotein and gelsolin. Furthermore, these studies also revealed a significant molecular diversity among subjects with schizophrenia.

Following up on microarray studies with genetic screening seems to be a promising strategy to identify disease-specific susceptibility genes. An initial microarray study on ten pairs of subjects (schizophrenia and matched control) identified *RGS4* as a gene with the most prominent expression decrease[32]. This suggested that the observed expression changes in *RGS4* are a result of a primary genetic abnormality and that some *RGS4* variants may be associated with increased risk for schizophrenia. Genetic association studies[33] on samples from three independent cohorts showed a significant transmission distortion for several *RGS4* haplotypes. Two other studies confirmed an association between *RGS4* single-nucleotide polymorphisms and schizophrenia[34,35], suggesting that *RGS4* indeed represents a new schizophrenia susceptibility gene.

Expression profiling of brains of subjects with ethanol abuse suggested disturbances in myelination, cell-cycle genes and cAMP-related transcripts[36,37], whereas expression patterns in the ventral tegmental area of cocaine overdose victims revealed significant upregulation of numerous glutamate receptors[38].

Rett syndrome (RTT) is a neurodevelopmental disorder that affects mainly girls, and is caused by mutation in the methyl CpG binding protein 2 (*MECP2*) gene encoding a transcriptional repressor. Microarray studies using postmortem RTT brains have reported increases in the expression of genes encoding glial transcripts, and decreased expression of genes encoding neuronal and glial proteins such as GABA and glutamate receptors[9]. However, no dramatic changes were seen in gene expression in fibroblast and lymphoblastoid cells derived from RTT patients[39] or in the brain of a mouse model[40]. The differences between these studies may simply represent intrinsic differences between different tissue types and species, further emphasizing the importance of generating datasets from the diseased human tissue.

Over 100 genes have been associated with the pathophysiology of multiple sclerosis (MS), and gene expression profiles can distinguish chronic-active and silent lesions[41]. Many of these genes encode proteins involved in inflammation/immune responses and myelination[41,42]. Importantly, the animal model experimental allergic encephalomyelitis (EAE) closely mimics a number of the expression changes seen in human subjects with MS: human brain lesions and EAE brains show a consistent expression change in 5-lipoxygenase, a key enzyme in the biosynthesis of the proinflammatory leukotrienes[43]. Furthermore, large-scale analysis of transcripts in MS lesions identify targets for therapeutic interventions that can ameliorate the course of EAE[44].

Microarray experiments with Alzheimer disease (AD) postmortem tissue highlight the complexity of the disease and the variety of experimental strategies in use (for review, see ref. 45). Most AD microarray studies focus on brain regions associated with the disease, such as the CA1 region of the hippocampus[46], amygdala and cingulate cortex[47] and superior temporal gyrus. Gene expression is also studied in individual cells, such as cholinergic neurons of the nucleus basalis[48] or hippocampal CA1 neurons and plaques[49,50]. Data from human brain have been compared with results from a transgenic mouse model with mutated amyloid precursor protein[51]. We cannot directly compare these studies because they focus on different aspects of AD. However, a variety of expression changes in individual genes have been described (for example, a consistently observed decrease in expression of nerve terminal proteins such as synapsin, synaptotagmin and synaptophysin[46–48,51,52]).

The most comprehensive microarray study of AD brains study thus far[6] reports that in the hippocampal CA1 region, 609 genes are correlated with incipient AD versus controls, and of these, changes in 89 genes correlate with both pre-mortem mental status and the neurofibrillary tangle index. Changes in transcription factors and signaling genes regulating proliferation and differentiation, particularly upregulation of tumor suppressors, oligodendrocyte growth factors and protein kinase A pathway genes seemed to be part of a co-regulated transcript network. On the basis of these findings, the investigators put forth an intriguing hypothesis of AD pathogenesis, in which early axonopathy is induced by tumor suppressor mediation and oligodendrocyte stimulation, resulting in a subsequent spreading of the disease process along myelinated nerve fibers.

Genomic disorders occur when regions of the genome undergo chromosomal rearrangements such as duplications or deletions. In trisomy 21 (Down syndrome), the expression of genes assigned to chromosome 21 is globally upregulated[53]. Comparative genomic hybridization (CGH) microarrays are used to study genome copy number variation[54], and this technology has been applied to brain tumors[55].

Unfortunately, the diagnostic specificity of the observed gene expression changes represents a great unknown in most microarray studies. Different disorders may show similar gene expression changes, and these changes can be related to a 'non-specific brain tissue malfunction'. Such changes may be present in a multitude of brain disorders, whereas others may be characteristic of a group of more closely related disorders[29]. Uncovering common expression changes across diseases will enable us to think in terms of common pathophysiological processes as well as common disease symptoms.

Sharing and comparing datasets

On the technical end, data sharing of transcriptome datasets is becoming relatively easy. Microarray data repositories (such as Gene Expression Omnibus[56] at the National Center for Biotechnology Information, and ArrayExpress[57,58] at the European Bioinformatics Institute) can accommodate even the largest datasets, and the deposited data are readily accessible by the whole scientific community. In an effort to standardize microarray data reporting, Brazma *et al.*[59] proposed a set of guidelines, Minimum Information About a Microarray Experiment (MIAME), to define parameters that uniformly describe each dataset, such as experimental design, sample preparation, hybridization procedures and use of controls[60]. Some journals, including *Nature Neuroscience*, now require public disclosure of the data in this format at the time of publication. Sharing of microarray data is also required from all researchers using the three NINDS/NIMH established microarray core facilities (http://array-consortium.cnmcresearch.org). Data generated by the consortium becomes publicly available 6 months after completion of the project.

Comparing microarray datasets is much more challenging[8]. First, we can compare outcomes of experiments. Was the expression pattern present in both experimental series? In this comparison, we rely on processed and analyzed data from different sources, accepting the data analysis that was done by the researchers who generated the datasets. This 'meta-analysis' is very useful, as replication of findings across different cohorts remains one of the critical aspects of postmortem brain research. However, negative outcomes of such comparisons are difficult to interpret: methodological differences can substantially influence the results.

Second, we can compare RNA level changes based on the analysis of raw data generated by different laboratories. The availability of raw data permits researchers to systematically explore the changes in RNA levels using any preferred preprocessing or other data-analysis technique. Such *post-hoc* comparisons require that there be no major technical confounds between the datasets to compare, although this is almost never the case. Even if the same microarray platform and processing procedures are used, the operators, batches of reagents and microarray processing equipment differ. Furthermore, the samples are not processed in parallel. All this may introduce variability in the data and could confound the outcome of the *post-hoc* comparisons. Therefore, raw data comparison remains an important challenge.

Validity of microarray data

As there is currently no consensus concerning data analysis, data validation by an independent method will remain a critical aspect of brain disease microarray research. Quantitative RT-PCR is extremely valuable for high-throughput validation of microarray findings (for review, see ref. 61), whereas *in situ* hybridization identifies the cellular localization of expression changes.

An initial microarray experiment may lead to hypotheses regarding the specific genes (or groups of genes) that are abnormally regulated in disease samples. This pattern of gene expression can be validated in a second series of microarray experiments, preferably using a previously untested set of samples. In this way, microarrays are used as a pattern validation tool. In practice, there are only limited samples available for brain disorders, especially specimens with adequate clinical data. Thus microarray findings are often validated on RNA derived from the same samples that were used for the initial microarray studies.

However, microarray experiments cannot uncover all expression differences between two conditions. Poor performance of the microarray probes, cross-hybridization on microarrays, data normalization errors and sample amplification errors are only some of the technical obstacles that limit our ability to discover differential expression in these experiments. Hence, the lack of differential expression in a microarray dataset should be interpreted very cautiously.

Future developments

Over the next several years, we can expect rapid development of microarray technology and data analysis. Improvements in microarray platform sensitivity and labeling will allow us to measure expression changes more precisely and to use much smaller amounts of starting material with limited sample amplification. Data analysis may become more standardized, enabling meaningful comparisons across datasets from different groups of investigators. Integration of data from different sources will enable us to better understand the data. Retrieval of already existing and newly accumulated information (such as precise expression and coexpression, DNA regulatory sequences, epidemiological trends, electrophysiological properties) from different databases will become more transparent, and we will attempt to cross-correlate the combined data with the disease history of individuals. Finally, correlation of expression changes across different psychiatric disorders will be important in defining molecular and pathophysiological similarities among psychiatric disorders.

With these technological advances, we can expect that the focus of microarray experiments will shift away from analysis of bulk brain tissue. High-throughput expression profiling of postmortem material is already beginning to focus on the transcriptome profile of single cells, cell types, layers and subnuclei. Targeted harvests of cell types, preferentially affected by the disease process (for example, layer-3 projection neurons or chandelier cells in prefrontal cortex in schizophrenia) will improve the discriminative power of the microarray analysis, unmasking hidden changes that are biologically significant, but are present only in a small subpopulation of cells. Furthermore, the identification of new disease susceptibility genes will ultimately lead to genetic engineering of mice that will mimic certain aspects of the disease phenotypes[62]. Microarray analysis of these animals will be essential for sorting out the causal relationships among downstream expression changes.

Finally, we believe that administrative initiatives should greatly influence how we generate and share microarray datasets. We can and should learn from the experiences of already existing collaborative initiatives. For example, joint expression profiling at University of California at Irvine, the University of Michigan, Stanford University and the University of California at Davis (funded by the Pritzker Family Philanthropic Foundation) is a good example of successful resource pooling[63]. To determine altered gene expression patterns in

patients with schizophrenia, bipolar disorder and depression, these four centers dynamically share postmortem brain material, reagents and data analysis approaches, thus creating a truly transparent and powerful expression dataset. However, it is not clear how and when will these datasets be shared with others in the scientific community. In an independent approach, a human brain collection and material sharing effort by the Stanley Medical Research Institute provides access to postmortem brain tissue from individuals with schizophrenia, bipolar disorder and major depression, as well as from normal controls[64] (www.stanleyresearch.org). This material, freely available to researchers, provides the only source of postmortem brain tissue to many outstanding scientists who are not associated with academic brain banking centers. Autism researchers are also beginning to coordinate their material sharing efforts[65].

We believe the National Institute of Mental Health (NIMH) should take similar initiative in providing a comprehensive framework for collaborative expression profiling among all NIMH-supported human brain banks. This could be achieved by establishing a central, NIH-supported gene expression profiling core for postmortem tissue, where all the samples would undergo uniform diagnostic evaluation, be processed using rigid sample preparation standards and analyzed on the same microarray platform. The core would serve only as a data generation facility: the unprocessed microarrays results would be continuously posted to an on-line database, leaving data mining, interpretation and follow-up to thousands of scientists around the country. These data would attract many scientists whose work is directly related to mechanisms of psychiatric disorders. Furthermore, it would attract the attention of statisticians, who are very much interested in such large datasets.

Initially, this expression profiling central core could focus on one complex disease and a limited number of brain regions. For example, if all human brain banks around the country would submit 1 gram of fresh-frozen material from the prefrontal cortex area 9 from all new harvests of subjects with schizophrenia, in five years we would generate an expression dataset on over 100 cases and matched controls. This dataset would have the power to uncover substratification within patients, defining major molecular phenotypes or genotypes of schizophrenia. Finally, each submitted sample could be accompanied by a standardized description of the disease history, which may allow scientists to correlate gene expression changes with symptoms of the disease.

This relatively small investment could fundamentally change our understanding of human brain diseases. However, the significance of microarray data will only make itself apparent through the collaboration of groups of scientists, so that leads generated by microarray research are systematically assessed in follow-up studies.

1. Schena, M., Shalon, D., Davis, R.W. & Brown, P.O. Quantitative monitoring of gene expression patterns with a complementary DNA microarray. *Science* **270**, 467–470 (1995).
2. Lockhart, D.J. *et al.* Expression monitoring by hybridization to high-density oligonucleotide arrays. *Nat. Biotechnol.* **14**, 1675–1680 (1996).
3. Cheung, V.G. *et al.* Making and reading microarrays. *Nat. Genet.* **21**, 15–19 (1999).
4. Hoffman, E. *et al.* Guidelines: Expression profiling - best practices for data generation and interpretation in clinical trials. *Nat. Rev. Genet.* **5**, 229–237 (2004).
5. Stuart, J.M., Segal, E., Koller, D. & Kim, S.K. A gene-coexpression network for global discovery of conserved genetic modules. *Science* **302**, 249–255 (2003).
6. Blalock, E.M. *et al.* Incipient Alzheimer's disease: Microarray correlation analyses reveal major transcriptional and tumor suppressor responses. *Proc. Natl. Acad. Sci. USA* **101**, 2173–2178 (2004).
7. Mirnics, K., Middleton, F.A., Lewis, D.A. & Levitt, P. Analysis of complex brain disorders with gene expression microarrays: schizophrenia as a disease of the synapse. *Trends Neurosci.* **24**, 479–486 (2001).
8. Mirnics, K. Microarrays in brain research: the good, the bad and the ugly. *Nat. Rev. Neurosci.* **2**, 444–447 (2001).
9. Colantuoni, C. *et al.* Gene expression profiling in postmortem Rett Syndrome brain: differential gene expression and patient classification. *Neurobiol. Dis.* **8**, 847–865 (2001).
10. Pongrac, J., Middleton, F.A., Lewis, D.A., Levitt, P. & Mirnics, K. Gene expression profiling with DNA microarrays: advancing our understanding of psychiatric disorders. *Neurochem. Res.* **27**, 1049–1063 (2002).
11. Greenbaum, D., Colangelo, C., Williams, K. & Gerstein, M. Comparing protein abundance and mRNA expression levels on a genomic scale. *Genome Biol.* **4**, 117 (2003).
12. Pierri, J.N., Chaudry, A.S., Woo, T.U. & Lewis, D.A. Alterations in chandelier neuron axon terminals in the prefrontal cortex of schizophrenic subjects. *Am. J. Psychiatry* **156**, 1709–1179 (1999).
13. Lewis, D.A. & Levitt, P. Schizophrenia as a disorder of neurodevelopment. *Annu. Rev. Neurosci.* **25**, 409–432 (2002).
14. Lewis, D.A. The human brain revisited: opportunities and challenges in postmortem studies of psychiatric disorders. *Neuropsychopharmacology* **26**, 143–154 (2002).
15. Harrison, P.J. *et al.* The relative importance of premortem acidosis and postmortem interval for human brain gene expression studies: selective mRNA vulnerability and comparison with their encoded proteins. *Neurosci. Lett.* **200**, 151–154 (1995).
16. Quackenbush, J. Microarray data normalization and transformation. *Nat. Genet.* **32** (Suppl.), 496–501 (2002).
17. Smyth, G. & Speed, T. Normalization of cDNA microarray data. *Methods* **31**, 265–273 (2003).
18. Tusher, V.G., Tibshirani, R. & Chu, G. Significance analysis of microarrays applied to the ionizing radiation response. *Proc. Natl. Acad. Sci. USA* **98**, 5116–5121 (2001).
19. Li, C. & Hung Wong, W. Model-based analysis of oligonucleotide arrays: model validation, design issues and standard error application. *Genome Biol.* **2**, 32.1-32.11 (2001).
20. Eisen, M.B., Spellman, P.T., Brown, P.O. & Botstein, D. Cluster analysis and display of genome-wide expression patterns. *Proc. Natl. Acad. Sci. USA* **95**, 14863–14868 (1998).
21. Kerr, M.K. & Churchill, G.A. Bootstrapping cluster analysis: assessing the reliability of conclusions from microarray experiments. *Proc. Natl. Acad. Sci. USA* **98**, 8961–8965 (2001).
22. Alizadeh, A.A. *et al.* Distinct types of diffuse large B-cell lymphoma identified by gene expression profiling. *Nature* **403**, 503–511 (2000).
23. Evans, S.J. *et al.* DNA microarray analysis of functionally discrete human brain regions reveals divergent transcriptional profiles. *Neurobiol. Dis.* **14**, 240–250 (2003).
24. Mukasa, A. *et al.* Distinction in gene expression profiles of oligodendrogliomas with and without allelic loss of 1p. *Oncogene* **21**, 3961–3968 (2002).
25. Mirnics, K., Middleton, F.A., Marquez, A., Lewis, D.A. & Levitt, P. Molecular characterization of schizophrenia viewed by microarray analysis of gene expression in prefrontal cortex. *Neuron* **28**, 53–67 (2000).
26. Hemby, S.E. *et al.* Gene expression profile for schizophrenia: discrete neuron transcription patterns in the entorhinal cortex. *Arch. Gen. Psychiatry* **59**, 631–640 (2002).
27. Middleton, F.A., Mirnics, K., Pierri, J.N., Lewis, D.A. & Levitt, P. Gene expression profiling reveals alterations of specific metabolic pathways in schizophrenia. *J. Neurosci.* **22**, 2718–2729 (2002).
28. Vawter, M.P. *et al.* Microarray analysis of gene expression in the prefrontal cortex in schizophrenia: a preliminary study. *Schizophr. Res.* **58**, 11–20 (2002).
29. Tkachev, D. *et al.* Oligodendrocyte dysfunction in schizophrenia and bipolar disorder. *Lancet* **362**, 798–805 (2003).
30. Mimmack, M.L. *et al.* Gene expression analysis in schizophrenia: reproducible upregulation of several members of the apolipoprotein L family located in a high-susceptibility locus for schizophrenia on chromosome 22. *Proc. Natl. Acad. Sci. USA* **99**, 4680–4685 (2002).
31. Hakak, Y. *et al.* Genome-wide expression analysis reveals dysregulation of myelination-related genes in chronic schizophrenia. *Proc. Natl. Acad. Sci. USA* **98**, 4746–4751 (2001).
32. Mirnics, K., Middleton, F.A., Stanwood, G.D., Lewis, D.A. & Levitt, P. Disease-specific changes in regulator of G-protein signaling 4 (RGS4) expression in schizophrenia. *Mol. Psychiatry* **6**, 293–301 (2001).
33. Chowdari, K.V. *et al.* Association and linkage analyses of RGS4 polymorphisms in schizophrenia. *Hum. Mol. Genet.* **11**, 1373–1380 (2002).
34. Williams, N.M. *et al.* Support for RGS4 as a susceptibility gene for schizophrenia. *Biol. Psychiatry* **55**, 192–195 (2004).
35. Morris, D.W. *et al.* Confirming RGS4 as a susceptibility gene for schizophrenia. *Am. J. Med. Genet.* **125**, 50–53 (2004).
36. Lewohl, J.M. *et al.* Gene expression in human alcoholism: microarray analysis of frontal cortex. *Alcohol Clin. Exp. Res.* **24**, 1873–1882 (2000).
37. Lewohl, J.M., Dodd, P.R., Mayfield, R.D. & Harris, R.A. Application of DNA microarrays to study human alcoholism. *J. Biomed. Sci.* **8**, 28–36 (2001).

38. Albertson, D. *et al.* Gene expression profile of the nucleus accumbens of human cocaine abusers: evidence for dysregulation of myelin. *J Neurochem.* **88**, 1211–1219 (2004).

39. Traynor, J., Agarwal, P., Lazzeroni, L. & Francke, U. Gene expression patterns vary in clonal cell cultures from Rett syndrome females with eight different MECP2 mutations. *BMC Med. Genet.* **3**, 12 (2002).

40. Tudor, M., Akbarian, S., Chen, R.Z. & Jaenisch, R. Transcriptional profiling of a mouse model for Rett syndrome reveals subtle transcriptional changes in the brain. *Proc. Natl. Acad. Sci. USA* **99**, 15536–15541 (2002).

41. Mycko, M.P., Papoian, R., Boschert, U., Raine, C.S. & Selmaj, K.W. cDNA microarray analysis in multiple sclerosis lesions: detection of genes associated with disease activity. *Brain* **126**, 1048–1057 (2003).

42. Whitney, L.W. *et al.* Analysis of gene expression in mutiple sclerosis lesions using cDNA microarrays. *Ann. Neurol.* **46**, 425–428 (1999).

43. Whitney, L.W., Ludwin, S.K., McFarland, H.F. & Biddison, W.E. Microarray analysis of gene expression in multiple sclerosis and EAE identifies 5-lipoxygenase as a component of inflammatory lesions. *J. Neuroimmunol.* **121**, 40–48 (2001).

44. Lock, C. *et al.* Gene-microarray analysis of multiple sclerosis lesions yields new targets validated in autoimmune encephalomyelitis. *Nat. Med.* **8**, 500–508 (2002).

45. Marcotte, E.R., Srivastava, L.K. & Quirion, R. cDNA microarray and proteomic approaches in the study of brain diseases: focus on schizophrenia and Alzheimer's disease. *Pharmacol. Ther.* **100**, 63–74 (2003).

46. Colangelo, V. *et al.* 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. *J. Neurosci. Res.* **70**, 462–473 (2002).

47. Loring, J.F., Wen, X., Lee, J.M., Seilhamer, J. & Somogyi, R. A gene expression profile of Alzheimer's disease. *DNA Cell. Biol.* **20**, 683–695 (2001).

48. Mufson, E.J., Counts, S.E. & Ginsberg, S.D. Gene expression profiles of cholinergic nucleus basalis neurons in Alzheimer's disease. *Neurochem. Res.* **27**, 1035–1048 (2002).

49. Ginsberg, S.D. *et al.* Predominance of neuronal mRNAs in individual Alzheimer's disease senile plaques. *Ann. Neurol.* **45**, 174–181 (1999).

50. Ginsberg, S.D., Hemby, S.E., Lee, V.M., Eberwine, J.H. & Trojanowski, J.Q. Expression profile of transcripts in Alzheimer's disease tangle-bearing CA1 neurons. *Ann. Neurol.* **48**, 77–87 (2000).

51. Yao, P.J. *et al.* Defects in expression of genes related to synaptic vesicle trafficking in frontal cortex of Alzheimer's disease. *Neurobiol. Dis.* **12**, 97–109 (2003).

52. Pasinetti, G.M. Use of cDNA microarray in the search for molecular markers involved in the onset of Alzheimer's disease dementia. *J. Neurosci. Res.* **65**, 471–476 (2001).

53. Mao, R., Zielke, C.L., Zielke, H.R. & Pevsner, J. Global up-regulation of chromosome 21 gene expression in the developing Down syndrome brain. *Genomics* **81**, 457–467 (2003).

54. Pollack, J.R. *et al.* Genome-wide analysis of DNA copy-number changes using cDNA microarrays. *Nat. Genet.* **23**, 41–46 (1999).

55. Hirose, Y. *et al.* Grade II astrocytomas are subgrouped by chromosome aberrations. *Cancer Genet. Cytogenet.* **142**, 1–7 (2003).

56. Edgar, R., Domrachev, M. & Lash, A.E. Gene Expression Omnibus: NCBI gene expression and hybridization array data repository. *Nucleic Acids Res.* **30**, 207–210 (2002).

57. Brooksbank, C. *et al.* The European Bioinformatics Institute's data resources. *Nucleic Acids Res.* **31**, 43–50 (2003).

58. Brazma, A. *et al.* ArrayExpress—a public repository for microarray gene expression data at the EBI. *Nucleic Acids Res.* **31**, 68–71 (2003).

59. Brazma, A. *et al.* Minimum information about a microarray experiment (MIAME)-toward standards for microarray data. *Nat. Genet.* **29**, 365–371 (2001).

60. Causton, H.C. & Game, L. MGED comes of age. *Genome Biol.* **4**, 351 (2003).

61. Mimmack, M.L. & Bahn, S. Quantitative RT-PCR in brain research. *Biol. Psychiatry* (in press).

62. Watase, K. & Zoghbi, H.Y. Modelling brain diseases in mice: the challenges of design and analysis. *Nat. Rev. Genet.* **4**, 296–307 (2003).

63. Fell, A. The disordered mind. *UC Davis Magazine Online* **19** (1), 2001. (http://www-ucdmag.ucdavis.edu/fall01/feature_1.html).

64. Torrey, E.F., Webster, M., Knable, M., Johnston, N. & Yolken, R.H. The stanley foundation brain collection and neuropathology consortium. *Schizophr. Res.* **44**, 151–155 (2000).

65. Pickett, J. Current investigations in autism brain tissue research. *J. Autism Dev. Disord.* **31**, 521–527 (2001).