

Exploiting microarrays to reveal differential gene expression in the nervous system

Robert S Griffin, Charles D Mills, Michael Costigan and Clifford J Woolf

Address: Neural Plasticity Research Group, Department of Anesthesia and Critical Care, Massachusetts General Hospital and Harvard Medical School, Charlestown, MA 02129, USA.

Correspondence: Clifford J Woolf. E-mail: woolf.clifford@mgh.harvard.edu

Published: 29 January 2003

Genome Biology 2003, **4**:105

The electronic version of this article is the complete one and can be found online at <http://genomebiology.com/2003/4/2/105>

© 2003 BioMed Central Ltd

Abstract

Microarrays have been used in a wide variety of experimental systems, but realizing their full potential is contingent on sophisticated and rigorous experimental design and data analysis. This article highlights what is needed to get the most out of microarrays in terms of accurately and effectively revealing differential gene expression and regulation in the nervous system.

The hypothesis-driven, single-gene analytic approach has dominated molecular neurobiology research and has been very successful. In grant submissions, the US National Institutes of Health demand succinctly defined, closed hypotheses based on expected outcomes. Where does this leave experiments utilizing high-density microarrays, which do not fit this mold? The fact that microarray investigators do not know which genes or pathways will be discovered by their experiments is precisely the strength of microarrays. The ability to examine the expression profile of potentially the entire genome at once, without preconceptions, offers the possibility of completely novel and unexpected insights into entities as complex as the nervous system. But before these insights can be made, the practical problems of experimental design, data analysis, verification, and interpretation need to be addressed.

Optimizing experimental models and design

The full potential of microarrays will be realized only when the questions that are asked about a system are sophisticated, rather than seeking simple changes or differences in expression profiles. We need to avoid experiments that generate lists of genes, suitable only for archiving and future data screening. In addition, technical aspects of any experiment require attention; the speed of tissue removal to prevent RNA degradation, extraction of high quality RNA, optimal reverse

transcription, amplification, labeling and hybridization with arrays and the use of quality control measures at each step are all critical. The suitability of different array formats also needs to be considered in terms of representation, variability, and sensitivity. Before considering data analysis and verification directly, we will discuss several of the microarray studies of neurobiological interest that have been completed to date. In particular, we focus on those aspects of model choice and experimental design that allow the conclusions of greatest biological importance to be drawn.

Carefully designed microarray studies have already begun to offer new forms of biological insight. To illustrate, microarrays have been used not only to validate putative drug targets but also to identify the cellular consequences of drug treatment and to differentiate target-specific drug effects from non-specific effects. For example, a comparison has been made between wild-type yeast treated with FK506, a putative axon-regeneration-promoting drug [1] that is better known for its use as an immunosuppressive agent, and untreated yeast lacking the gene for calcineurin, the drug target [2]. By identifying genes that were regulated in the drug-treated yeast lacking the drug's therapeutic target, the investigators were able to describe genes representing the drug's transcriptional side-effect profile. Yeast studies have also pioneered the use of arrays to identify *cis*-acting control elements that regulate the transcription of many co-regulated

genes [3], a phenomenon that may be important in both neuronal development and the response of neurons to injury.

Another promising neurobiological use of microarrays has been in studying tumors of the central nervous system: here arrays have been used not only for diagnosis but also as a prognostic tool. Pomeroy and colleagues [4] collected a microarray dataset from 99 embryonic tumor samples associated with known diagnoses and patient survival outcomes. They were able to distinguish the different histopathological classes of medulloblastoma on the basis of expression profiles. In addition, they derived subsets of 21 or fewer genes whose combined expression levels could classify the medulloblastoma samples according to survival versus treatment failures. There were no known prognostic markers for medulloblastoma prior to this study.

A variety of experimental models in neuroscience have now been investigated using microarrays. For example, arrays have been used to study differences in the expression profiles of two inbred strains of mice with different levels of susceptibility to seizure [5]. This study found that in a strain of mice resistant to seizure-induced hippocampal cell death (C57BL/6) the expression of many more genes was induced in the hippocampus by seizure than in a strain of mice susceptible to post-seizure cell death (129SvEv). Other array studies have examined the effects of diet and aging on the mouse brain [6,7], and the effects of environmental influences, such as exposure to an enriched environment [8]. Arrays have also been used to examine gene-expression changes in models of Parkinson's disease [9], in patients with Alzheimer's disease [10], and in schizophrenia [11]. Although these studies are so far purely descriptive or observational, they have uncovered novel genes or suggested novel mechanisms, the significance of which now needs to be explored further.

The ability to use microarrays to view changes in expression induced by null mutations or gene knockouts is also of interest. A gene-expression study of the pons and cerebellum has used such a strategy to find genes that are involved in normal target innervation [12]. In normal mice, pontine neurons develop a projection that synapses with cerebellar granule neurons. Diaz and colleagues [12] demonstrated that in *weaver* mutant mice, which lack cerebellar granule neurons, gene expression in the pons during the development of the pontocerebellar projection is altered. Expression of genes involved in axon outgrowth persists beyond their normal phase of downregulation after target innervation, whereas genes involved in synapse formation fail to be upregulated at the proper time in development. This study is noteworthy for its careful experimental design and data analysis. Statistical principles were used to identify regulated genes, and the data were examined using clustering and linear regression modeling. When microarrays begin to be used to study the effects of conditional knockout or over-expression of genes, not only can the compensatory and

downstream effects of the constitutive loss of a gene throughout development be examined, but global analysis of the response to perturbations in the levels of single genes in specific conditions will also be possible [13].

Complex tissues and diverse cell populations

Heterogeneity of the tissue being examined is a major concern when using arrays to analyze the nervous system. Zirlinger *et al.* [14] have identified region-specific gene expression within the brains of mice using microarrays. In addition, they observed that some of the genes they identified could be grouped into three spatial expression patterns, each consisting of expression within a different subset of the amygdaloid subnuclei. Other genes, however, seemed to exhibit expression patterns that did not conform to neuro-anatomical subnuclei. Lockhart and Barlow [15] have used microarrays to compare gene-expression profiles in six specific regions of the brain (amygdala, cerebellum, cortex, entorhinal cortex, hippocampus and midbrain) in human and mouse tissue; they describe 23 genes that are specific to the cerebellum, giving this region the highest number of unique genes among the regions they studied.

Averaging expression levels of an entire region of tissue, such as the brain or even a specific brain area or nucleus, will clearly minimize or even conceal large expression changes that occur in small subpopulations of cells. There are promising techniques for overcoming this obstacle, however: laser capture to microdissect single cells [16], fluorescence-activated cell sorting (FACS) [17], and cDNA synthesis from single cells [18,19]. A key issue for all these techniques is that the isolated RNA needs to be faithfully amplified. New approaches have made it possible to start with a limited number of cells and still obtain sufficient material for successful microarray experiments. For example, Affymetrix has developed a double *in vitro* transcription protocol that allows array hybridization to be based on as little as 50 ng of total RNA from the starting material [20]. These amplification methods will be particularly useful for studies of the development of the nervous system, where region-specific tissue amounts are very limited [13]. Recently, Yamagata and colleagues [21] have successfully applied single-cell RT-PCR to retinal ganglion cells, in an effort to identify those genes involved in regulating topographic mapping in the chick retinotectal projection. They were able to identify Sidekick-1 and Sidekick-2, adhesion molecules that appear to have a role in lamina-specific target recognition. The applicability of this technology to microarray studies remains to be explored, however.

Experimental replication and modes of data analysis

The analysis of data from microarrays is non-trivial. Many problems may arise in developing a means for ranking

genes, which is typically based on the degree of difference between two experimental conditions. Regardless of the particular measure of differential expression that the researcher settles on, it is now clear that methods based on replicate independent arrays are necessary for accurate gene identification (Figure 1). Even apart from variability in technique, most neurobiological experiments face such inter-individual and inter-group biological variability that it is unlikely that a single array comparison will ever provide a reliable estimate of gene-expression levels, even for pilot studies. As with any

other experiment, multiple independent trials are necessary if we are to obtain accurate data.

The next challenge is to determine how many of the genes estimated from the array expression data are truly regulated. Hundreds, or thousands, of genes are examined in microarray experiments. A single null hypothesis (for example, the absence of regulation) is generally rejected when the probability of false rejection is 1/20. When hundreds of hypotheses are tested in the same experiment, multiple significant test

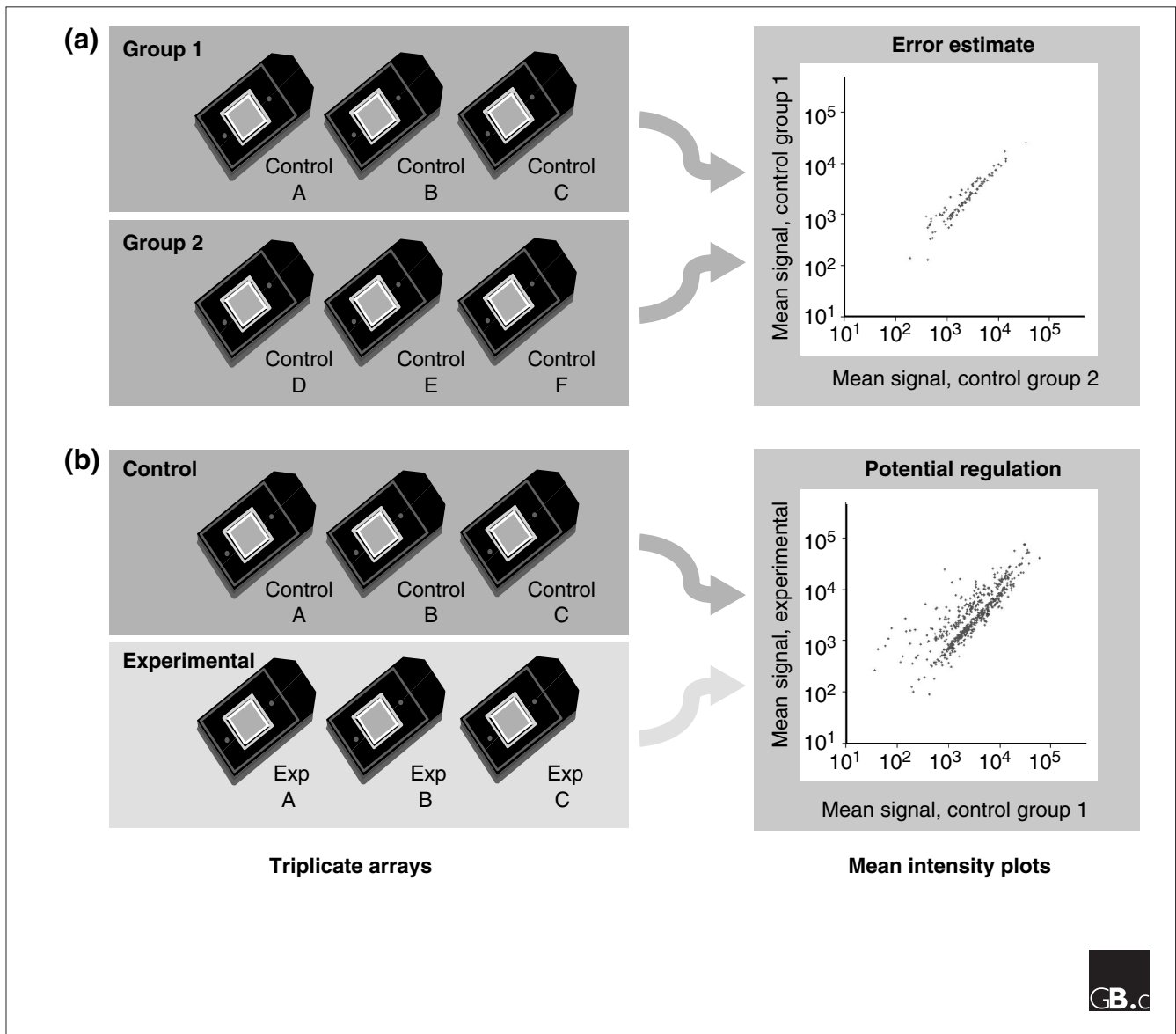


Figure 1

Filtering array data. It is essential to remove as many false positives as possible at the earliest stages of analysis. Random errors occurring in oligonucleotide array data were assessed by comparing three biologically independent control arrays to three other biologically independent control arrays. Genes with a t test $p < 0.05$ in this comparison are plotted (a). Comparison of these results with the frequency of genes differing at $p < 0.05$ between control and experimental triplicate arrays (b) were used to define a set of criteria based on present/absent, fold change, t test p value, and a signal threshold that minimized the estimated random error [31].

results will be obtained. If each test was performed with a threshold for a positive result of 0.05, then of the test results identified as positives, it is likely that a subset of these will be falsely assigned. Several different statistical corrections are available to limit the accumulation of false positives, most conservatively by dividing the p value threshold by the number of hypothesis tests (Bonferroni correction). This concern will be of the greatest importance when the population under study contains rare true positives, as might occur when comparing wild-type animals to mutants that have very few transcriptional differences. Conversely, false positives will be less of a problem when there is a high prior probability of true positives. Recently, methods have been developed to limit the overall rate of falsely identified genes with the specific case of microarray experiments in mind [22]. This approach involves mathematically controlling the overall proportion of false positives in the dataset, as opposed to controlling the probability that any given test result is a false positive. But the conservative nature of most available statistical methods means that researchers often find that even well-characterized genes with distinct differential expression patterns fall short of statistical significance within array data sets. Thus, it might be appropriate to consider the values of test statistics as a means for weighting genes according to their reliability or promise for further study, not as definitive rule-in or rule-out thresholds. Researchers may also be able to define empirical statistical thresholds that are useful in their specific experimental systems by performing appropriate control experiments (Figure 1).

Microarray experiments offer the possibility of characterizing the expression of large groups of genes simultaneously using cluster analysis [23]. Briefly, clustering algorithms may be divided into supervised and unsupervised methods. Unsupervised methods are designed to group similar expression patterns in a dataset without referring to any outside information. The hierarchical, k -means, and self-organizing map clustering algorithms fall into this category, and they assist the researcher by identifying groups of co-regulated genes. In contrast, supervised methods use outside information about the experimental condition to shape the derivation of a model from the dataset. For example, a k -nearest neighbors algorithm was used by Pomeroy *et al.* [4] to derive a set of genes that predict medulloblastoma treatment outcome. In addition to being useful for the researcher in thinking about the large datasets that result from microarray experiments and in shaping future hypotheses, the groups of genes resulting from either supervised or unsupervised clustering approaches can be used as a platform for further studies of signaling systems or metabolic pathways [24].

Grouping genes into functional categories on the basis of data from other experimental systems [25] provides a useful heuristic tool for analyzing microarray data. Websites have

been set up to assist functional categorization, notably the KEGG database [26]. Grouping can be used either alone or in combination with cluster analysis to identify genes of related function that exhibit similar expression patterns over time (Figure 2) [3]. But this type of analysis has potential pitfalls when applied to the study of the nervous system: genes identified in other systems in the body frequently have entirely different functions in the nervous system. For

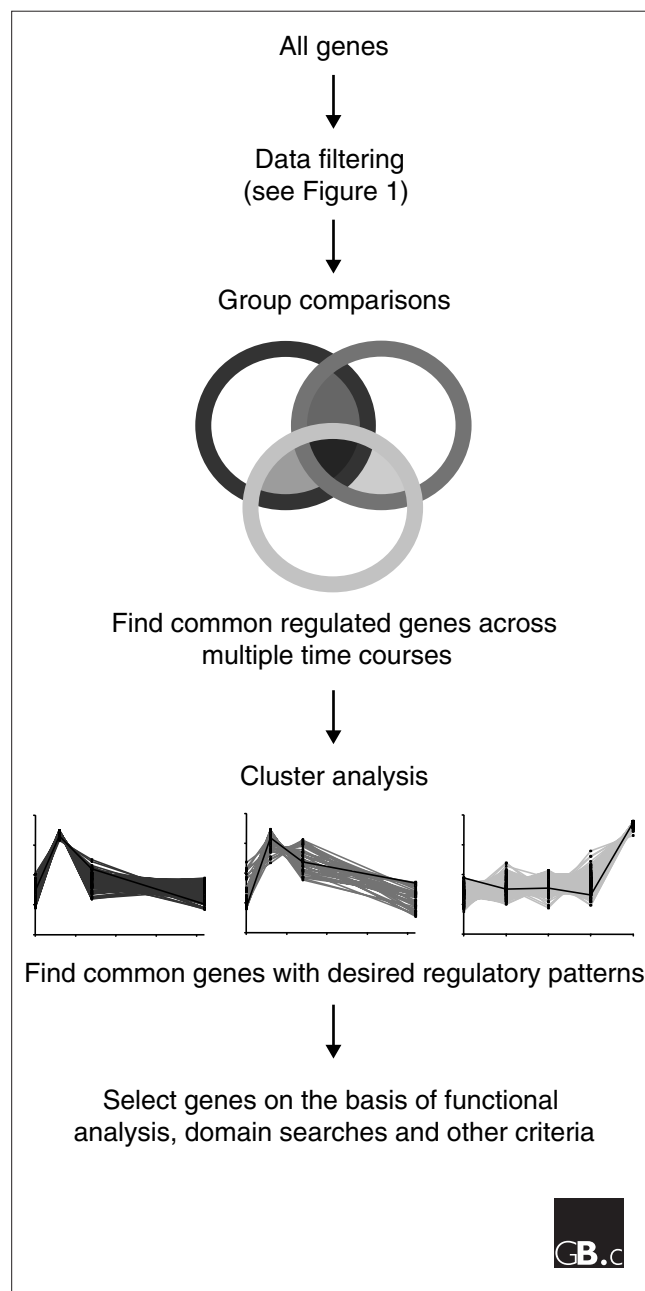


Figure 2
Candidate gene identification. Schematic of a protocol that may be used to select a few genes that obey a defined set of criteria. See text for further details.

example, the platelet-activating factor signaling system, involved in platelet aggregation and blood clotting, and the major histocompatibility (MHC) class I genes involved in cell-mediated immunity, are both involved in developmental patterning of the brain [27,28].

Experimental validation

Once the appropriate mathematical tools have been used to identify genes likely to be regulated in the manner of interest, the next step is biological characterization of these genes. The first stage of this process is validation, which requires measuring the expression of specific genes of interest using independent methods and independent RNA samples. There are important reasons why a gene identified on an array should be validated by another method. For oligonucleotide arrays, the labeled RNA used for hybridization to the chip is not the original RNA population but an amplified copy whose production requires multiple enzymatic steps, each of which may bias array results. Cross-hybridization due to sequence homology may lead to erroneous results, especially when using cDNA arrays to examine genes that have many closely related family members. If all things were equal, then different probe sets (or cDNAs) representing the same gene should behave identically on the same array, but this is often not the case [29]. To illustrate this problem, in a study of amygdala-enriched genes, the authors were able to validate expression of only approximately 60% of the transcripts identified by the microarray analysis [14].

Several possibilities are available for confirming expression patterns that are shared by multiple genes. Real-time PCR allows the expression profiles of many genes to be plotted across many RNA samples [30]. We have found that slot or dot blotting of RNA is a successful alternative option: placing RNA directly on a filter allows the use of relatively low amounts of RNA per sample (1 µg of total RNA) and reasonable scale production of filters (10-15 per hour) [31]. Visualizing changes in mRNA expression using *in situ* hybridization allows the cellular localization of regulated genes to be identified. This is critical for the nervous system, where it is the specific pattern of circuitry that is the vital information [14,32]. Finally, it should be stressed that interpretations of microarray results rely only on mRNA signals. Changes in intracellular distribution of proteins, post-transcriptional modifications, receptor sensitization and desensitization, protein dimerizations, and other protein-protein interactions are beyond the scope of microarray analysis. Substantial changes in protein levels can occur in neurons without any detectable change in mRNA, indicating that translation rate may be important in determining the amounts of some proteins [33].

We are about to face a massive expansion of microarray studies of the nervous system [34]. But realizing the full

potential of this technology requires a clear understanding of the technical, bioinformatic and scientific issues. With attention to design, quality control and methods of analysis, the potential rewards are great.

Acknowledgements

The authors' work is supported by NIH NS3825, HD38533 NS39518 (CJW). C.D.M. is supported by NINDS grant NS45459.

References

- Lyons WE, Steiner JP, Snyder SH, Dawson TM: **Neuronal regeneration enhances the expression of the immunophilin FKBP-12.** *J Neurosci* 1995, **15**:2985-2994.
- Marton MJ, DeRisi JL, Bennett HA, Iyer VR, Meyer MR, Roberts CJ, Stoughton R, Burchard J, Slade D, Dai H, et al.: **Drug target validation and identification of secondary drug target effects using DNA microarrays.** *Nat Med* 1998, **4**:1293-1301.
- Tavazoie S, Hughes JD, Campbell MJ, Cho RJ, Church GM: **Systematic determination of genetic network architecture.** *Nat Genet* 1999, **22**:281-285.
- Pomeroy SL, Tamayo P, Gaasenbeek M, Sturla LM, Angelo M, McLaughlin ME, Kim JY, Goumnerova LC, Black PM, Lau C, et al.: **Prediction of central nervous system embryonal tumour outcome based on gene expression.** *Nature* 2002, **415**:436-442.
- Sandberg R, Yasuda R, Pankratz DG, Carter TA, Del Rio JA, Wodicka L, Mayford M, Lockhart DJ, Barlow C: **Regional and strain-specific gene expression mapping in the adult mouse brain.** *Proc Natl Acad Sci USA* 2000, **97**:11038-11043.
- Jiang CH, Tsien JZ, Schultz PG, Hu Y: **The effects of aging on gene expression in the hypothalamus and cortex of mice.** *Proc Natl Acad Sci USA* 2001, **98**:1930-1934.
- Lee CK, Weindruch R, Prolla TA: **Gene-expression profile of the aging brain in mice.** *Nat Genet* 2000, **25**:294-297.
- Rampon C, Jiang CH, Dong H, Tang YP, Lockhart DJ, Schultz PG, Tsien JZ, Hu Y: **Effects of environmental enrichment on gene expression in the brain.** *Proc Natl Acad Sci USA* 2000, **97**:12880-12884.
- Grunblatt E, Mandel S, Maor G, Youdim MB: **Gene expression analysis in N-methyl-4-phenyl-1,2,3,6-tetrahydropyridine mice model of Parkinson's disease using cDNA microarray: effect of R-apomorphine.** *J Neurochem* 2001, **78**:1-12.
- Ginsberg SD, Hemby SE, Lee VM, Eberwine JH, Trojanowski JQ: **Expression profile of transcripts in Alzheimer's disease tangle-bearing CA1 neurons.** *Ann Neurol* 2000, **48**:77-87.
- Mirnic K, Middleton FA, Marquez A, Lewis DA, Levitt P: **Molecular characterization of schizophrenia viewed by microarray analysis of gene expression in prefrontal cortex.** *Neuron* 2000, **28**:53-67.
- Diaz E, Ge Y, Yang YH, Loh KC, Serafini TA, Okazaki Y, Hayashizaki Y, Speed TP, Ngai J, Scheiffele P: **Molecular analysis of gene expression in the developing pontocerebellar projection system.** *Neuron* 2002, **36**:417-434.
- R Livesey: **Have microarrays failed to deliver for developmental biology?** *Genome Biol* 2002, **3**:comment2009.1-2009.5.
- Zirlinger M, Kreiman G, Anderson DJ: **Amygdala-enriched genes identified by microarray technology are restricted to specific amygdaloid subnuclei.** *Proc Natl Acad Sci USA* 2001, **98**:5270-5275.
- Lockhart DJ, Barlow C: **Expressing what's on your mind: DNA arrays and the brain.** *Nat Rev Neurosci* 2001, **2**:63-68.
- Emmert-Buck MR, Bonner RF, Smith PD, Chuaqui RF, Zhuang Z, Goldstein SR, Weiss RA, Liotta LA: **Laser capture microdissection.** *Science* 1996, **274**:998-1001.
- Rieseberg M, Kasper C, Reardon KF, Scheper T: **Flow cytometry in biotechnology.** *Appl Microbiol Biotechnol* 2001, **56**:350-360.
- Brady G, Billia F, Knox J, Hoang T, Kirsch IR, Voura EB, Hawley RG, Cumming R, Buchwald M, Siminovitich K, et al.: **Analysis of gene expression in a complex differentiation hierarchy by global amplification of cDNA from single cells.** *Curr Biol* 1995, **5**:909-922.
- Eberwine J, Kacharmina JE, Andrews C, Miyashiro K, McIntosh T, Becker K, Barrett T, Hinkle D, Dent G, Marciano P: **mRNA**

- expression analysis of tissue sections and single cells. *J Neurosci* 2001, **21**:8310-8314.
20. **Affymetrix protocols**
[<http://www.affymetrix.com/technology/index.affx>]
21. Yamagata M, Weiner J, Sanes J: **Sidekicks. Synaptic adhesion molecules that promote lamina-specific connectivity in the retina.** *Cell* 2002, **110**:649-660.
22. Tusher VG, Tibshirani R, Chu G: **Significance analysis of microarrays applied to the ionizing radiation response.** *Proc Natl Acad Sci USA* 2001, **98**:5116-5121.
23. Quackenbush J: **Computational analysis of microarray data.** *Nat Rev Genet* 2001, **2**:418-427.
24. Ideker T, Thorsson V, Ranish JA, Christmas R, Buhler J, Eng JK, Bumgarner R, Goodlett DR, Aebersold R, Hood L: **Integrated genomic and proteomic analyses of a systematically perturbed metabolic network.** *Science* 2001, **292**:929-934.
25. Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, Davis AP, Dolinski K, Dwight SS, Eppig SS, et al.: **Gene ontology: tool for the unification of biology.** *Nat Genet* 2000, **25**:25-29.
26. **KEGG** [<http://www.kegg.org>]
27. Hirotsune S, Fleck MW, Gambello MJ, Bix GJ, Chen A, Clark GD, Ledbetter DH, McBain CJ, Wynshaw-Boris A: **Graded reduction of Pafah1b1 (Lis1) activity results in neuronal migration defects and early embryonic lethality.** *Nat Genet* 1998, **19**:333-339.
28. Huh GS, Boulanger LM, Du H, Riquelme PA, Brotz TM, Shatz CJ: **Functional requirement for class I MHC in CNS development and plasticity.** *Science* 2000, **290**:2155-2159.
29. Schadt EE, Li C, Su C, Wong WH: **Analyzing high-density oligonucleotide gene expression array data.** *J Cell Biochem* 2000, **80**:192-202.
30. Wang H, Sun H, Della PK, Benz R, Xu J, Gerhold D, Holder D, Koblan K: **Chronic neuropathic pain is accompanied by global changes in gene expression and shares pathobiology with neurodegenerative diseases.** *Neuroscience* 2002, **114**:529-546.
31. Costigan M, Befort K, Karchewski L, Griffin RS, Da'Urso D, Allchorne A, Sitariski J, Mannion JW, Pratt RE, Woolf CJ: **Replicate high-density rat genome oligonucleotide microarrays reveal hundreds of regulated genes in the dorsal root ganglion after peripheral nerve injury.** *BMC Neurosci* 2002, **3**:16.
32. Blackshaw S, Fraioli RE, Furukawa T, Cepko CL: **Comprehensive analysis of photoreceptor gene expression and the identification of candidate retinal disease genes.** *Cell* 2001, **107**:579-589.
33. Ji R, Samad T, Jin S, Schmoll R, Woolf C: **p38 MAPK activation by NGF in primary sensory neurons after inflammation increases TRPV1 levels and maintains heat hyperalgesia.** *Neuron* 2002, **36**:57-68.
34. **NINDS-NIMH Microarray Consortium**
[<http://arrayconsortium.cnmcresearch.org>]