OPINION

Expressing what's on your mind: DNA arrays and the brain

David J. Lockhart and Carrolee Barlow

Questions about brain function and disease are being addressed with parallel genomic approaches. High-density DNA arrays make it possible to monitor the expression levels of thousands of genes at a time, and are being used to address old questions in new ways and to generate new hypotheses about the workings of the brain.

Studies of brain function have been approached in different ways over the past decades. The use of genetic, electrophysiological, cellular and molecular tools has made it possible to take ever more reductionist approaches. Although the molecular methods offer insight into the underlying mechanisms of brain function, they have necessarily focused on a small number of genes or on specific cellular and molecular processes. Now, with the availability of the nearly complete DNA sequence of the mouse and human genomes, and the development of new genomics tools that allow us to exploit this information, it has become possible to use 'gene-centric' approaches on a global scale. In other words, we now can afford the luxury of taking a broad, systems-based approach to the brain while, at the same time, obtaining a gene-specific molecular view.

It is sometimes assumed that because the brain is such a complex organ, experimental genomics methods are not directly applicable to neurobiological studies. However, it seems that the opposite might be true. It is precisely because the brain is complex, because brain functions are diverse and varied, and because the numbers of genes involved in neural processes are likely to be large (and largely unknown at this point), that it is important to move beyond conventional 'one-gene-at-a-time' approaches to strategies that look broadly at the activity of tens of thousands of genes in parallel. There are simply too many genes to attack these problems solely in a sequential, purely hypothesis-driven fashion. It is certainly the case that special care must be taken when using advanced genomic tools for neurobiological studies, and that they should be used in combination with other tools and traditional expertise. However, their careful, rigorous, systematic and innovative use is opening up new avenues for the study of the brain, allowing us to address old questions in ways that would have been completely impractical only a few years ago.

A primary goal of neurogenetics is to determine the genes that are responsible for specific neural phenotypes, the activities of different cell types, and the unique structures and functions of different brain regions. Given the heterogeneity of the brain, the large numbers of genes that are present in the genomes of higher mammals and the complexity of the processes that are involved in brain function, how is it possible to find the genes that determine important phenotypes and encode crucial proteins? How can we begin to understand the mechanisms that underlie various brain functions, and how can we understand what can and does go wrong in disease? How can such tasks be accomplished without being overly costly, and time- and labour-intensive?

DNA array technology

DNA arrays are among the most powerful and versatile tools for genomics and genetics research[1–6]. DNA arrays allow us to take advantage of the growing body of sequence information to make quantitative parallel measurements of gene expression (messenger RNA abundance) for tens of thousands of genes at a time. DNA arrays are passive devices that work by hybridization of DNA, or RNA, to DNA sequences that are immobilized at specific physical locations on a solid support, usually glass. The arrays interrogate complex nucleic acid samples and effectively count the number of different RNA or complementary DNA molecules that are present in a sample, which can be made directly from cellular mRNA. The process is straightforward, highly parallel (all sequences are counted simultaneously) and, if done correctly, quantitative. Depending on the array type, the concentration of a given sequence fragment or the ratio of the concentrations between two samples can be determined for tens of thousands of genes at a time based on a straightforward hybridization reaction and a post-hybridization fluorescence scan of the array. These arrays can be used to measure gene expression levels and changes in gene expression for cells in culture, dissected tissue, specific small nuclei, micro-dissected tissue, and even very small numbers of selected cells when combined with appropriate amplification steps[7–10].

There are two dominant types of DNA arrays (often called 'microarrays') that have been used for most global gene expression measurements. The first are high-density oligonucleotide arrays that are synthesized *in situ* on a glass surface using light-directed combinatorial synthesis (commercially available from Affymetrix)[11]. These oligonucleotide arrays can contain more than 400,000 sequences, typically 25-mers, in 20×20 micron features in a total area smaller than one half-inch square, with 12,000 to 13,000 genes and expressed sequence tags (ESTs) represented on commercially available designs that use

C57BL/6 (2/2)

	Cb	Cx	EC	Hp	Mb	MEF
Ag	436	52	26	56	144	2076
Cb		333	412	329	242	2159
Cx			25	27	76	2070
EC				53	175	2115
Hp					95	2081
Mb						2064

129SvEv (2/2)

	Cb	Cx	EC	Hp	Mb	MEF
Ag	413	46	23	66	164	2146
Cb		306	406	368	260	2137
Cx			20	44	106	2135
EC				69	193	2043
Hp					148	2123
Mb						2136

C57BL/6 and 129SvEv (3/4)

	Cb	Cx	EC	Hp	Mb	MEF
Ag	470	48	22	70	171	2261
Cb		356	439	388	279	2196
Cx			25	40	118	2156
EC				80	232	2237
Hp					156	2258
Mb						2206

Figure 1 | **Gene expression differences between mouse brain regions.** The number of genes that are differentially expressed (out of ~11,000) between amygdala (Ag), cerebellum (Cb), cortex (Cx), entorhinal cortex (EC), hippocampus (Hp), midbrain (Mb) and mouse embryonic fibroblasts (MEF). The numbers are based on consistent observations in independent replicates (2/2 for each strain individually or 3/4 for the combined analysis) using the criteria described in FIG. 2.

16 probe pairs per gene. The arrays are designed and synthesized on the basis of sequence information alone, and it is possible to cover tens of thousands of genes and ESTs on a single array[12]. The other main array type is made by spotting cDNAs (or more recently, pre-synthesized oligonucleotides) at specific locations on a glass slide. The cDNAs, usually polymerase chain reaction (PCR) products that are 500 to 1,000 bases in length, are spaced about 100 to 300 microns apart, allowing for more than 10,000 spots to be placed on a standard glass microscope slide.

Oligonucleotide arrays have several specific advantages. For example, they can be designed and made directly from sequence information without physical intermediates, such as cDNAs or PCR products. In addition, large numbers of probes are used to increase detection redundancy (that is, there are many probes or 'detectors' per gene). Moreover, shorter probes can be targeted to the most unique regions of genes, thereby reducing cross-hybridization and increasing specificity and the ability to discriminate between closely related members of gene families. In the case of cDNA microarrays, one of their main advantages is their versatility. Gene sets can be changed, expanded or modified to include, for example, alternatively spliced or newly discovered genes. It is also possible to make one's own custom arrays for any organism of choice, and to spot unknown cDNAs for gene discovery purposes.

A complementary approach to finding genes that are expressed in the brain is to make and sequence cDNA libraries prepared from brain mRNA. This is an excellent way to identify expressed genes, but it is not generally amenable to practical experimentation on an appropriate scale because it is serial in nature, and tends to be slow, expensive and does not produce quantitative expression information. Several non-array based methods, such as SAGE (Serial Analysis of Gene Expression)[13], TOGA (Total Gene Expression Analysis)[14] and other RNA fingerprinting approaches have also been devised to get around some of the shortcomings of direct cDNA sequencing and differential display. Nonetheless, we believe that the best combination is to use the information obtained from genomic and cDNA-sequencing approaches to make oligonucleotide or cDNA arrays that cover large numbers of genes and ESTs. Sample preparation and array hybridization methods are relatively straightforward, arrays can be made and used in significant numbers, and they are quickly becoming a standard laboratory reagent, albeit still expensive. This makes it possible to do a significant number of quantitative experiments (for example, time courses, dose-response curves, analysis of several strains or mutants) as one might do by using northern blots or reverse-transcriptase polymerase chain reactions (RT-PCRs), but for very large collections of genes at a time.

Profiles of neurological disease

Several groups have begun to apply the new genomic approaches to address questions in neurobiology and to understand the interacting array of genes that might give rise to complex behaviours and phenotypes. A few noteworthy papers specifically highlight important aspects of the technology and its proper use while, at the same time, addressing questions of importance for human health. For example, Lee *et al.* have used oligonucleotide arrays covering more than 6,000 genes to measure gene expression changes that occur in the neocortex and cerebellum of ageing mice (5-month-old compared with 30-month-old mice), and the potential mitigating effects of caloric restriction[15]. They found that, at the transcriptional level, brain ageing in mice seems to parallel changes in human neurodegenerative disorders, and that caloric restriction selectively retards many age-related effects. Several of the genes identified as differentially expressed, including ones where the change was less than a factor of two, were confirmed independently using quantitative PCR. A high level of agreement between the array-based and more conventional methods was observed. Similarly, Ginsberg *et al.* have measured expression profiles in patients with Alzheimer's disease, using cDNA microarrays that cover more than 18,000 genes and ESTs. They compared tangle-bearing with normal CA1 neurons that were aspirated from sections of brain from patients with Alzheimer's disease and controls, respectively[16]. An important technical aspect of this work is that some of the changes in mRNA abundance that were revealed by the array-based measurements were verified with immunohistochemical studies of their encoded proteins. Mirnics *et al.* have used commercially available cDNA microarrays that cover nearly 7,000 human genes and ESTs to study expression differences in the prefrontal cortex between matched pairs of people with schizophrenia and control subjects[17]. They found that genes involved in the regulation of presynaptic function were consistently decreased in the brains from the affected people, and these results were verified using *in situ* hybridization measurements in prefrontal cortex tissue sections. Furthermore, other experiments were done on monkeys that were treated with the anti-psychotic drug haloperidol to verify that the observed expression changes were not due to the effects of medication.

Profiling and behaviour

Expression profiling has recently been applied to the study of cellular processes that might be related to drug addiction. Thibault *et al.* studied direct actions of ethanol on gene expression in neural cell cultures, which might be important for behaviours such as ethanol tolerance or dependence[18]. These investigators found that ethanol regulated several groups of related genes, including some involved in the synthesis or metabolism of noradrenaline. A striking overlap between the patterns of gene expression induced by ethanol or cyclic AMP was observed, indicating that cAMP signalling might be involved in a significant portion of the direct actions of ethanol on gene expression. Many of the findings in this work were

also verified with northern blots and quantitative RT-PCR measurements of mRNA levels, western blots and even measurements of noradrenaline release.

To identify the genes that are responsible for the unique structures and functions of specific brain regions, and those that might be responsible for neurobehavioural phenotypes, we measured expression profiles in specific regions of the adult mouse brain from two inbred strains of mice[19]. We found that ~1% of the genes expressed in the brain (~0.5% of genes monitored) are differentially expressed between the two strains. The overall expression profiles in different regions of the adult mouse brain are surprisingly similar, but there are a significant number of genes that are differentially expressed between regions, and a small but significant number that are uniquely expressed in one region but not others (FIG.1).

Practical considerations

On the basis of our experience and that of others, we cannot stress strongly enough the importance of great experimental care, well-characterized and rigorous analysis, and the need for appropriate follow-up and verification when doing highly parallel expression experiments, especially when using animal or human tissue. In almost all cases, experiments should be conducted at least in duplicate, with replicates done as independently as possible (for example, different mice or independent dissections of a region, independent sample preparations and independent hybridizations to physically different arrays). It is not sufficient to merely remake samples from the same extracted RNA from the same mouse or tissue sample, or to simply re-hybridize samples to other arrays, as has been done in some studies. If genetically identical, inbred mice are not used, then it is necessary to do more experiments or to pool mice to effectively average out differences due to genetic inhomogeneity. The same considerations apply when using any other animal or human tissue. At least some fraction of the genes observed to be differentially expressed should be confirmed with independent methods on independent samples (not the same RNA that was used for the array experiments), especially if less stringent analysis criteria are used or if subtle expression differences are to be interpreted. For example, we typically do northern blot or quantitative RT-PCR experiments to check particularly interesting findings, and to confirm a result that might be the basis for follow-up experiments, such as the creation of a knockout mouse. The use of western blots to measure corresponding protein levels, and immunohistochemistry and *in situ* hybridization to measure cell or region specificity of proteins and mRNAs is also highly recommended. Al though the array-based expression measurements can be made quantitative and reproducible, specific genes that are found to be differentially expressed on arrays should be viewed as high probability candidates, but not as completely confirmed. Global expression measurements should be considered a starting point for the understanding of a biological problem, and as a valuable tool for obtaining information concerning a large number of genes. They should be used in the context of other types of measurements, knowledge and information, and it should be understood that findings will need to be followed up with further experiments of various, more conventional types.

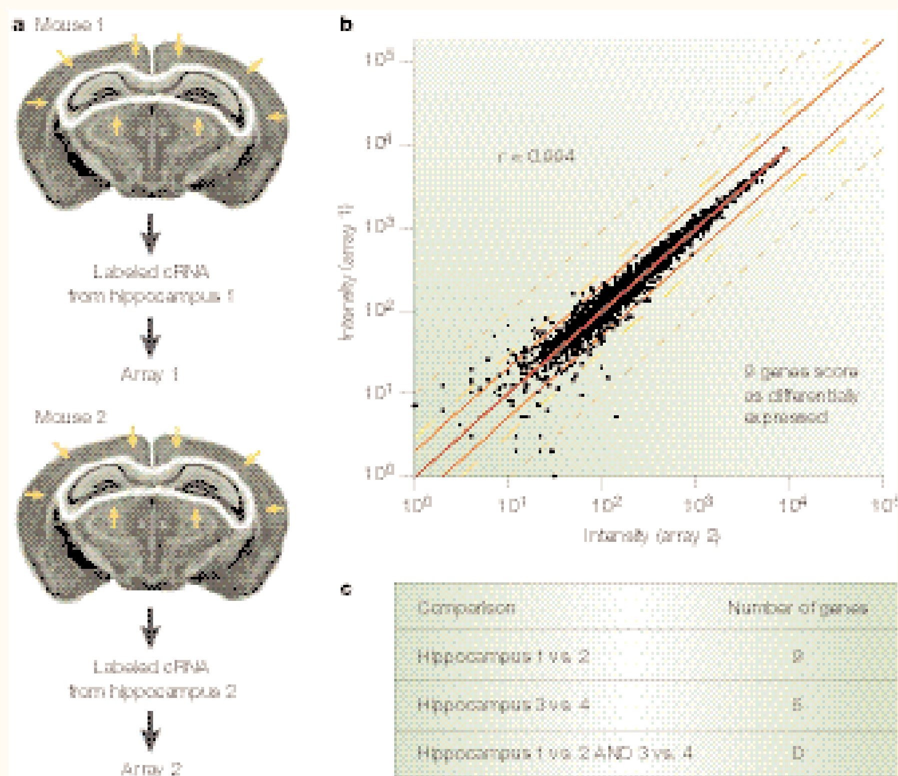

Figure 2 | **Experimental design and reproducibility. a** | General schematic of mouse brain gene expression experiments. Tissue is obtained by systematic dissection of the appropriate brain regions from at least two mice of the same age, sex and genetic background, which have been housed and handled identically. In this example, the hippocampus (arrows) was used. After obtaining tissue, total RNA is extracted and labelled complementary RNA is made for array hybridization. Each sample is hybridized to a separate array. **b** | A comparison of the quantitative results for independent replicates for two different mice (C57BL/6). The correlation coefficient is very near 1.0 (0.994), and the number of genes that score as 'differentially expressed' based on a single comparison is small (9 of a possible 6584 in this example), indicating the high degree of reproducibility of the procedures, measurements and analyses. **c** | Analysis showing the low false-positive rate achieved when using stringent analysis criteria and independent replicates. Samples were prepared from dissected hippocampus from four C57BL/6 mice. When the results for mouse 1 and mouse 2 were compared, only nine genes scored as different. When mouse 3 was compared with mouse 4, only five genes were scored as different and there were no genes that scored as different in both of the independent comparisons. The criteria used were a 1.8-fold change or greater, a qualitative call of increased, marginally increased, decreased or marginally decreased, a signal change of 50 (after scaling to an average signal of 200), and a call of present in at least one of the samples using the standard Affymetrix GeneChip algorithms and software.

We have found that to obtain the highest quality, and most meaningful and reproducible results in expression measurements in mice, all handling and brain dissections need to be conducted with great care, precision and accuracy. The extremely high level of care taken in our experiments so far[19] has made it possible to obtain accurate and reliable results while using as few mice and arrays as possible (FIG. 2). For example, only genetically identical, inbred mice have been used and the mice are always of the same age and sex. The mice are handled and housed under identical conditions, and are even dissected at the same time of day. As shown previously, this level of care

results in a low false-positive rate (between 0 and 3 genes incorrectly scored as being differentially expressed per ~13,000 genes monitored) when experiments are conducted as fully independent duplicates and conservative analysis criteria are applied. To call a gene 'increased' or 'decreased' in a pair-wise comparison, it is important to require consistency across independent comparisons of replicate results (for example, hippocampus 1 versus cerebellum 1, and hippocampus 2 versus cerebellum 2). The purpose of this conservative approach is to reduce the risk of falsely assigning a gene as differentially expressed while maintaining sensitivity to relatively subtle changes, typically 1.8-fold or greater, or as low as 1.4-fold under some circumstances[15,18]. The number of replicates required for this level of confidence is determined by an analysis of the reproducibility of the duplicate measurements (for example, hippocampus 1 versus 2 and cerebellum 1 versus 2; see FIG. 2). Because of the demonstrated low false-positive rate using independent duplicates (two mice, and two independent sample preparations, array hybridizations and analyses), it is not necessary in general to use a larger number of mice when the experiments can be controlled to this extent (FIG. 2). However, the number of independent measurements required for a given level of sensitivity and accuracy needs to be continually assessed based on the consistency of the observations made for a given strain and/or brain region. If duplicate observations are not sufficiently consistent, then further replicates need to be done. The continuing analysis of replicate data will indicate regions, strains, perturbations or protocols for which the inherent variability is greater, and for these, further repeats are necessary to maintain the same levels of confidence. In addition, a larger number of replicates are useful when attempting to detect and quantify expression differences that are even more subtle than the typical threshold of a factor of 1.8 (see REFS 15,18 for examples). As mentioned earlier, if genetically identical, inbred strains of mice are not used, then tissue from more animals needs to be used individually or in pools from five or more animals to 'average out' effects due to genetic variation. However, although they increase the efficiency of a study, pooling strategies hide the underlying variation within a group and can mask some important effects. For this reason, we recommend that genetically homogeneous, inbred strains be used whenever possible, and that follow-up measurements be conducted to characterize intra- and inter-population variation.

Region-specific expression

Can global expression measurements provide insight into how brain regions or cell types function, and into the genes that are responsible for the underlying mechanisms of activity? Can we examine the regulatory elements for uniquely expressed genes to find promoters that can be used to drive expression in specific cell types or tissues in mice? We have measured gene expression patterns (more than 10,000 genes) in six different regions of the adult mouse brain to find genes that are uniquely expressed or highly enriched in one brain region compared to all the others (FIGS 1,3). In this work, we found that the cerebellum was the most unique of the six regions studied, with 23 genes that were not found to be expressed elsewhere, and 28 genes that were expressed everywhere except the cerebellum. The experimental results from these studies were validated by the observations of regional specificity for known region-specific genes (for example, the *N*-methyl-D-aspartate (NMDA) receptor NR2C subunit and Pukinje cell protein 2 (PCP-2) uniquely expressed in the cerebellum), and independent validation of the results for selected genes using northern blots and RT-PCR analysis. This study is being expanded to include many more regions of the brain in other phenotypically characterized inbred strains of mice.

We have also done similar experiments on selected brain regions from several monkeys and humans (FIG. 3) to test the generality of the observations made in mouse models (J. A. Del Rio, D.J.L. and C.B., unpublished observations). The experiments using human and monkey RNA were done on the same type of oligonucleotide array that was designed to measure the expression levels of nearly 7,000 human genes. Because of the high sequence similarity between most human and monkey genes, monkey experiments can be done using human arrays without any changes in experimental procedures or data analysis methods. The general picture concerning the number of genes that are uniquely expressed and the regions that have the most distinct expression profiles is quite consistent for humans and mice. We compared the specificity profile for homologous human and monkey genes that showed region-specificity in the mouse, and we found that many of the genes have similar patterns of expression. However, for some genes, the expression patterns are not the same across all three mammals. It is intriguing to speculate that some of these differences might give rise to the significant differences in brain function and cognitive abilities between us and our distantly related cousins.

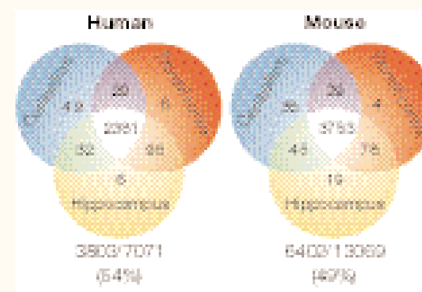

Figure 3 | **Of mouse and man.** Venn diagrams showing the number of genes with specific expression patterns in a subset of adult human and mouse brain regions. To identify genes with region-restricted gene expression, genes were classified as 'present' in a region if the gene had a call of 'present' in at least three out of four samples. Similarly, to classify genes as clearly not detected, we used a call of 'absent' in four out of four related samples (for example, two replicates from both C57BL/6 and 129SvEv mice). The total number interrogated and the number and percentage of genes that scored as present in at least one of the three brain regions is indicated below the diagrams. Note that, in the case of the cerebellum and frontal cortex, a higher percentage of genes are uniquely expressed in human than in mouse. This may be due to the fact that more finely dissected portions of these regions were used from the human brain, whereas the entire structure was used from the smaller mouse brain.

Profiling and QTL analysis

Different inbred strains of mice vary greatly in many interesting neurobehavioural phenotypes[20] and we can take advantage of the different inbred strains with characterized phenotypes that already exist. These strains are useful for neurogenetic studies because they show well-documented differences in behaviour, brain anatomy and sensitivity to environmental perturbations. As inbred strains are genetically homogenous, it is possible to determine the influences of environmental and other perturbations on gene expression without the confounding effects of differences in genetic background.

We have used gene expression profiling of several brain regions in two commonly used inbred strains (C57BL/6 and 129SvEv) to find genes that might account for the differences between inbred mouse strains that differ in their neurobehavioural phenotypes[19]. Out of more than 7,000 genes detected, only 24 were differentially expressed in all six brain regions between the two strains, and 49 other genes were differentially expressed between the strains in at least one region. This indicates that the expression levels between the strains are similar for more than 99% of the genes, and that fewer than 1% determine the differences in phenotype. We measured only about one-fifth of the genes in the mouse genome.

Therefore, by extrapolation to the entire genome, more than 350 genes might be differentially expressed between the strains in at least one region. This number of differences could account for a great deal of the phenotypic variation seen between these strains.

These results also indicate that the combination of global expression analysis with traditional mapping and positional cloning approaches might be an efficient route to identify disease-related genes or quantitative trait loci (QTL). The goal is similar for both approaches — to find genes responsible for complex traits. In the case of QTL analysis, the genes are identified positionally in the genome[21]; in expression mapping, genes are identified functionally based on measurements of gene expression. The two approaches are complementary in that standard QTL analysis identifies the genes or the loci that harbour genetic differences relevant to a phenotype, while the expression approach measures cellular consequences of any genetic variations[19,22]. Although conventional QTL analysis is a powerful tool for mapping susceptibility loci to chromosomal regions, many genes usually reside in these large regions, and more extensive work is required to identify the specific gene or genes involved in determining the phenotype. Our findings indicate that an expression-based strategy might be useful for identifying candidate genes while simultaneously identifying other genes that modify the particular trait being studied.

Array analysis from limited RNA

Brain expression profiling experiments done so far highlight the need for more complete gene coverage, the use of a larger number of different brain regions and finer dissections of inhomogeneous tissue from many brain regions. This last point is particularly important as even finely dissected brain tissue contains many cell types. Because of the cellular diversity in brain regions, important expression differences that occur for cells that make up only a fraction of the total population might not be detected in the presence of the large number of different cell types that do not show a similar effect. This is why it is so important to dissect brain regions more finely, and to minimize the effects of this cellular 'masking' or expression averaging. For even finer and more specific dissections, one can use sorted cells, cells obtained using laser capture[23], and other types of microdissection.

Fortunately, although it is more difficult, it is possible to do high-quality array-based measurements starting with RNA from a small number of cells. The key is coupling the array measurements with amplification methods that lead to a faithful and reproducible representation of the original mRNA population. In our hands, PCR-based approaches tend to produce skewing of the relative concentrations of each message because of amplification biases. But methods that lead to as much as a million-fold amplification using a multiround, linear process have been developed[7–10], and this approach has been used already in several array-based studies[24,25]. This amplification procedure makes it possible to start routinely with 100 to 1,000 cells, or with as few as 1–10, to obtain a sufficient quantity of labelled material for hybridization to gene expression arrays. The combination of reliable amplification techniques, laser capture microdissection and cell sorting opens up new areas of neurobiological studies that can take advantage of higher throughput genomics methods.

> "… to maximize the value of the data for the entire scientific community, it should be made broadly available in a usable format".

Other technological advances have increased the number of genes that can be represented on a single array, and methods for parallel amplification and sample preparation in 96-well plates and subsequent hybridization to tens of arrays at a time on a single glass wafer have also been developed[26]. It is clear that obtaining expression profiles that cover very large numbers of genes for large numbers of experimental samples is technically feasible. The main challenges at this point are not necessarily technological. What is necessary are studies based on good experimental design, good model systems, clever and systematic manipulations in animal models, and improved data analysis tools and 'knowledge-bases'[27,28] for making sense of the flood of new information[5,29–31].

Public databases

It is impossible for one group to fully analyse and understand data sets of the size that can be produced by microarray-based expression studies, and to adequately present the results in the usual journal formats. Because of this, and to maximize the value of the data for the entire scientific community, we have made, and encourage others to make, experimental data broadly available in a usable format. For example, the complete set of quantitative gene expression data from our mouse brain studies[19], along with lists of differentially expressed genes are now publicly available at our web site (see links) This practice is becoming fairly standard, and is being done on a large scale, for example, at the Cancer Genome Anatomy Project (CGAP) established and administered by the National Cancer Institute.

To maximize the scientific usefulness of the extensive gene expression data, it is important that the expression and related data be stored in a standard, useful format and that it be made broadly available to the scientific community. This task is not trivial because of the large size of the data sets (roughly 100 Megabytes per chip-based gene expression measurement, including raw and processed data files) and because of the different types of information and data represented. The quantitative gene expression data tend to be in a standardized regular format, although this format can be different for different types of arrays. However, information about mice, experiments, protocols and supplementary image data are not so simply represented. We and others have made gene expression data from arrays available in text files and spread sheets. Whereas this approach is better than no access at all, it is inefficient to use the data in this form; this format does not scale well, and it is insufficient for the needs of the broader community. What is needed are systems that can include as much of the data and information as possible in searchable fields, while providing links to other types of information (for example, free text, annotations and images) that are not as readily queried but are nonetheless important to access. It is clearly important to build large, systematic databases that allow users to retrieve, analyse, query, visualize and compare data from several sources. The interface to the database should be web-browser based so that access to the system is largely hardware- and operating-system-independent (that is, portable to most computer systems). Because the amount of data will continue to increase, the system needs to be scalable to effectively manage changes in data volume without extensive software modification. Furthermore, researchers will continue to develop new experimental protocols, which will result in acquisition of new data types. For this reason, the system should be extensible to accommodate new types of data. Finally, because different types of data and information need to be included, the database must be sufficiently flexible to accommodate features of both flat files and relational databases.

Summary

It is clear that the combination of mouse genetics, whole-genome sequence information, microdissection methods, single nucleotide

polymorphism maps, genotyping methods, random and targeted mutagenesis and phenotyping, *in vivo* imaging, and tools for proteomics and parallel expression monitoring will have a significant effect on neurobiological research and its application to human health. Although these approaches are powerful, they should be used to complement the traditional methods of neuroscience, molecular biology and genetics. It is important to use them with great care and rigour to avoid interpreting a large amount of noise and false positives rather than reliable data. They should be used in the context of asking important biological questions, along with careful experimental design, rigorous data analysis, some healthy scepticism and a determination to follow up interesting observations. But it is also the case that these tools are ideal for the occasional exploratory leap. It is a good idea sometimes to take the proverbial 'fishing expedition' that might lead to new findings and the discovery of relationships between genes and processes that are totally unexpected and that would be almost impossible to discover by taking a more limited view. It is an exciting time for neuroscience, and it is exciting to see how new technologies and approaches are influencing the way important neurobiological questions are answered.

David J. Lockhart and Carrolee Barlow are at the Salk Institute for Biological Studies, Laboratory of Genetics, 10010 North Torrey Pines Road, La Jolla, California 92037, USA.
e-mails: barlow@salk.edu
lockhart@salk.edu

Links

FURTHER INFORMATION Barlow lab web site | Gene expression data from the authors | Cancer Genome Anatomy Project

1. Fodor, S. P. *et al.* Multiplexed biochemical assays with biological chips. *Nature* **364**, 555–556 (1993).
2. Southern, E. M. *et al.* Arrays of complementary oligonucleotides for analysing the hybridisation behaviour of nucleic acids. *Nucleic Acids Res.* **22**, 1368–1373 (1994).
3. 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).
4. Lockhart, D. J. *et al.* Expression monitoring by hybridization to high-density oligonucleotide arrays. *Nature Biotechnol.* **14**, 1675–1680 (1996).
5. Lockhart, D. J. & Winzeler, E. A. Genomics, gene expression and DNA arrays. *Nature* **405**, 827–836 (2000).
6. Supplement. The Chipping Forecast. *Nature Genet.* **21**, S3–S50 (1999).
7. Kwoh, D. Y. *et al.* Transcription-based amplification system and detection of amplified human immunodeficiency virus type 1 with a bead-based sandwich hybridization format. *Proc. Natl Acad. Sci. USA* **86**, 1173–1177 (1989).
8. Guatelli, J. C. *et al.* Isothermal, *in vitro* amplification of nucleic acids by a multienzyme reaction modeled after retroviral replication. *Proc. Natl Acad. Sci. USA* **87**, 7797 (1990).
9. Van Gelder, R. N. *et al.* Amplified RNA synthesized from limited quantities of heterogeneous cDNA. *Proc. Natl Acad. Sci. USA* **87**, 1663–1667 (1990).
10. Eberwine, J. *et al.* Analysis of gene expression in single live neurons. *Proc. Natl Acad. Sci. USA* **89**, 3010–3014 (1992).
11. Fodor, S. P. A. *et al.* Light-directed, spatially addressable parallel chemical synthesis. *Science* **251**, 767–773 (1991).
12. Lipshutz, R. J., Fodor, S. P., Gingeras, T. R. & Lockhart, D. J. High density synthetic oligonucleotide arrays. *Nature Genet.* **21**, 20–24 (1999).
13. Velculescu, V. E., Zhang, L., Vogelstein, B. & Kinzler, K. W. Serial analysis of gene expression. *Science* **270**, 484–487 (1995).
14. Sutcliffe, J. G. *et al.* TOGA: an automated parsing technology for analyzing expression of nearly all genes. *Proc. Natl Acad. Sci. USA* **97**, 1976–1981 (2000).
15. Lee, C. K., Weindruch, R. & Prolla, T. A. Gene-expression profile of the ageing brain in mice. *Nature Genet.* **25**, 294–297 (2000).
16. 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).
17. Mirnics, K., Middleton, A. M., Lewis, D. A. & Levitt, P. Molecular characterization of schizophrenia viewed by microarray analysis of gene expression in prefrontal cortex. *Neuron* **28**, 53–67 (2000).
18. Thibault, C. *et al.* Expresssion profiling of neural cells reveals specific patterns of ethanol-responsive gene expression. *Mol. Pharmacol.* **52**, 1593–1600 (2000).
19. Sandberg, R. *et al.* From the cover: regional and strain-specific gene expression mapping in the adult mouse brain. *Proc. Natl Acad. Sci. USA* **97**, 11038–11043 (2000).
20. Crawley, J. N. *et al.* Behavioral phenotypes of inbred mouse strains: implications and recommendations for molecular studies. *Psychopharmacology (Berl.)* **132**, 107–124 (1997).
21. Risch, N. J. Searching for genetic determinants in the new millennium. *Nature* **405**, 847–856 (2000).
22. Geschwind, D. H. Mice, microarrays, and the genetic diversity of the brain. *Proc. Natl Acad. Sci. USA* **97**, 10676–10678 (2000).
23. Emmert-Buck, M. R. *et al.* Laser capture microdissection. *Science* **274**, 998–1001 (1996).
24. Luo, L. *et al.* Gene expression profiles of laser-captured adjacent neuronal subtypes. *Nature Med.* **5**, 117–122 (1999).
25. Wang, E., Miller, L. D., Ohnmacht, G. A., Liu, E. T. & Marincola, F. M. High-fidelity mRNA amplification for gene profiling. *Nature Biotechnol.* **18**, 457–459 (2000).
26. Zarrinkar, P. *et al.* Arrays of arrays for high throughput gene expression monitoring. *Nature Biotechnol.* (submitted).
27. Bhalla, U. S. & Iyengar, R. Emergent properties of networks of biological signaling pathways. *Science* **283**, 381–387 (1999).
28. Weng, G., Bhalla, U. S. & Iyengar, R. Complexity in biological signaling systems. *Science* **284**, 92–96 (1999).
29. Cho, R. J. & Campbell, M. J. Transcription, genomes, function. *Trends Genet.* **16**, 409–415 (2000).
30. Brown, P. O. & Botstein, D. Exploring the new world of the genome with DNA microarrays. *Nature Genet.* **21**, 33–37 (1999).
31. Brent, R. Genomic biology. *Cell* **100**, 169–183 (2000).

Acknowledgements
We would like to thank Jo A. Del Rio, Todd A. Carter, Cindy Doane, Rickard Sandberg, Dan Pankratz, Edward Callaway and members of the Barlow laboratory for their patience and help and Robert Vigorito of the Baltimore Maryland Brain Bank for assistance in sample acquisition.

LEARNING AND MEMORY

Interpretations of retrograde amnesia: old problems redux

Paula M. Millin, Erik W. Moody and David C. Riccio

Recent evidence indicates that an old memory reactivated by cueing becomes labile and vulnerable to an amnesic treatment. Although the 'reconsolidation' concept derived from these findings challenges the traditional consolidation theory, here we argue that the new concept suffers from some of the same limitations as the earlier model. We propose an alternative retrieval-based theory that accommodates the recent data, as well as other puzzling related observations.

Newly formed memories are susceptible to disruption by various post-acquisition treatments. The ability of these treatments to interfere with memory decreases as the interval between acquisition and amnesic treatment increases. This common observation prompted scientists to formulate the well-known consolidation hypothesis. According to this hypothesis, memories continue to be processed (or 'consolidated') through a complex molecular cascade of events for an unspecified period after initial learning (for reviews, see REFS 1–3). Once this process is complete, the memory is presumed to be permanently stored, impervious to future manipulations.

The idea of a permanent memory store that is no longer susceptible to alteration has been challenged by many researchers. A report by Judge and Quartermain[4], for instance, showed that the protein synthesis inhibitor anisomycin disrupted memory for old, reactivated fear conditioning, as well as for a newly acquired fear memory. Recently, Nader *et al.*[5] extended this finding by showing a similar effect when anisomycin was infused directly into the lateral and basal nuclei of the amygdala (LBA). Infusion of anisomycin shortly after memory reactivation produced amnesia for both 1 and 14 day-old consolidated fear memories. The impairment was not due to nonspecific effects, as performance was unimpaired for up to 4 hours after the infusion. Moreover, infusion of anisomycin into the LBA 6 hours after memory reactivation left the memory intact, implying a time-limited role for protein synthesis. These findings prompted the authors to conclude that an old, well-consolidated fear memory returns to a labile