

Finding New Candidate Genes for Learning and Memory

Tarek A. Leil,^{1,2} Alex Ossadtschi,³ James S. Cortes,^{1,2} Richard M. Leahy,³ and Desmond J. Smith^{1,2*}

¹Department of Molecular and Medical Pharmacology, UCLA School of Medicine, Los Angeles, California

²Crump Institute for Molecular Imaging, Los Angeles, California

³Signal and Image Processing Institute, University of Southern California, Los Angeles, California

The genetic mechanisms underlying learning and memory remain mysterious, but many of the genes are likely to be expressed in the hippocampus, a region pivotal to this process. We used a 9,000 gene microarray to examine differences in hippocampal gene expression between two F1 hybrid mouse strains that perform well on the Morris water maze and two inbred strains that perform poorly. This resulted in identification of 27 differentially expressed genes, which could be used to place the F1 hybrid and inbred strains into separate clusters based on singular value decomposition. Most of the genes have unknown function, but those with known functions may provide clues to the molecular mechanisms of learning. Using multiple strains to narrow down the number of candidate genes should be a useful general approach to genome-wide studies of behavioral and other complex traits. © 2002 Wiley-Liss, Inc.

Key words: DNA microarrays; gene expression; hippocampus; inbred strains; mice

Learning and remembering is a universal human experience and is one of the most significant properties of the nervous system. This complex phenomenon demands a scientific explanation, but despite the apparent effortlessness of learning and memory, our knowledge of this process is only in its infancy and much remains to be understood. The importance of learning and memory is painfully underscored when disordered, as occurs prominently in the congenital and acquired learning disorders. Few effective therapies have been developed for these disorders.

A number of studies have given useful insights into memory formation. These investigations have implicated events centered on neurotransmitter systems, increased synaptic strength and downstream signaling events. For example, biochemical and pharmacological studies have implicated cAMP coupled downstream signaling pathways in learning and memory (Mayford and Kandel, 1999). Classical or “forward” genetic screens in *Drosophila* have also resulted in the identification of learning and memory genes (Davis, 1996; Belvin and Yin, 1997; Dubnau and Tully, 1998), including cAMP signaling and integrin-based signaling pathways (Grotewiel et al., 1998). One

promising approach is the use of genetic mapping to find loci involved in cognitive ability (Plomin, 2001). The obstacles confronting complex trait analysis in humans, however, render this approach daunting.

It is likely that many of the pathways responsible for learning and memory will be expressed in the hippocampus, a region of the brain that plays a key role in this behavior in humans (Scoville and Milner, 2000), as well as in other mammals. In mice and rats, spatial learning on the Morris water maze task requires an intact hippocampus (Moser et al., 1993; Cho et al., 1999), and the use of genetically engineered mice has helped clarify some of the relevant molecular pathways (Chen and Tonegawa, 1997; Mayford et al., 1997). In addition, different strains of mice vary greatly in performance on learning tasks (Owen et al., 1997; Upchurch and Wehner, 1988), and differences in hippocampal protein kinase C levels have been shown to correlate with the observed performance (Bowers et al., 1995; Matsuyama et al., 1997). A broad survey of gene expression differences that may explain the widely varying learning and memory abilities of different mouse strains has yet to be performed.

The development of DNA microarray technology (Scheda et al., 1995) provides such an opportunity. We describe the use of a 9,000-cDNA microarray to compare hippocampal gene expression differences between two F1 hybrid mouse strains that learn well on the Morris water maze (B6129PF1/J and B6C3F1/J) with two inbred strains that perform poorly (129X1/SvJ, formerly 129/SvJ, and LP/J). In general, F1 hybrid strains show similar levels of good performance on the Morris water maze, a reflection of hybrid vigor. These strains efficiently acquire the position of the hidden platform, as judged by decreased latencies to find the platform and also the probe test

*Correspondence to: Desmond J. Smith, Department of Molecular and Medical Pharmacology, UCLA School of Medicine, 23-120 CHS, Box 951735, Los Angeles, CA 90095-1735.
E-mail: Dsmith@mednet.ucla.edu

Received 27 August 2001; Revised 11 December 2001; Accepted 12 December 2001

(Owen et al., 1997). In contrast, inbred strains show more variation on these tests. To leverage the maximum possible differences in our analysis, we chose two of the best learning F1 hybrids and two of the poorest learning inbred strains. By seeking consistent patterns of gene expression, the four strain comparison offers the possibility of narrowing down the number of candidate genes, eliminating those unrelated to learning and memory.

MATERIALS AND METHODS

Mice

The F1 hybrid and inbred mouse strains used for Morris water maze testing and the DNA microarray hybridization experiments (B6129PF1/J, B6C3F1/J, C57BL/6J, 129X1/SvJ, and LP/J) were purchased from the Jackson Laboratory (Bar Harbor, ME). All mice used in the experiments were 8–12-week-old adult males and were housed in the facility of the Department of Laboratory Animal Medicine at the University of California at Los Angeles. It should be noted that, unlike F2 hybrids, the F1 hybrids of a particular strain are genetically identical, as are mice from inbred strains. All mice were housed two to four per cage on a 12:12 hr light:dark cycle, with food and water given ad lib.

Morris Water Maze

The Morris water maze apparatus consisted of a circular tank made of galvanized steel and painted white (1.2 m wide \times 0.58 m high) located in a rectangular room (3.81 m \times 2.6 m). Visual cues were placed on all four walls of the room. The hidden platform was submerged 1 cm below the surface of the water, which was 28 cm deep and rendered opaque using non-toxic white paint. The movement of the mice was followed using Poly-Track video tracking system (San Diego Instruments, San Diego, CA).

The Morris water maze tests were performed as described (Smith et al., 1997). The mice were tested for 2 days to determine their ability to locate a submerged platform with a flag (the visible platform test). Naive mice were also tested for 3 days for their ability to locate an unmarked, submerged platform (the hidden platform test). At the conclusion of the third day of the hidden platform test, the mice were given a final trial of 1 min with the platform removed (probe test). The times spent in each quadrant (dwell) and the number of times the mice crossed the previous site of the platform (crossings) were recorded.

RNA Isolation and Labeling

Naive mice were euthanized by decapitation, the hippocampus was carefully dissected from the brain, placed immediately in TRIzol Reagent (Gibco BRL, Gaithersburg, MD) and homogenized by passing several times through a 26_{5/8}G needle. Subsequent steps were done according to the manufacturer's instructions. For each microarray experiment four to twelve hippocampi were pooled to obtain the quantity of RNA sufficient for hybridization. This also minimized experimental noise due to variations between different mice. The genetic homogeneity of the F1 hybrid and inbred strains further decreased noise. Cy-dye labeled first-strand cDNA was synthesized by reverse transcription (RT) using 100 μ g of total RNA in a

40 μ l reaction volume. Final concentrations of reagents during cDNA synthesis were 1 \times First Strand Buffer (Gibco BRL), 10 mM DTT, 50 nM dATP/dGTP/dTTP, 20 nM unlabeled dCTP, 100 nM FluorLink-dCTP (either Cy3 or Cy5 labeled, Amersham Pharmacia Biotechnologies, Piscataway, NJ), 1 U/ μ l RNasin, 12.5 U/ μ l Superscript II RT, and 1 μ g oligo (dT)_{12–18} (Gibco BRL). The reactions were incubated at 42°C for 1 hr, and 400 U of Superscript II RT were added and incubation continued for another hr. The reaction was stopped using 55 mM EDTA and RNA was removed by incubating in 0.2 N sodium hydroxide at 65°C for 1 hr. The reaction was then neutralized by addition of 25 μ l of 1 M Tris-HCL (pH 7.5). The Cy5 and Cy3-labeled probes were combined and purified and concentrated by washing several times using a MicroCon 30 concentrator (Amicon Corp., Beverly, MA).

General Microarray Strategy

Microarrays were co-hybridized with Cy3 and Cy5 labeled probes, with one of the labels acting to control for interarray variability. In the experiments described here, control RNA from the arbitrary reference strain C57BL/6J was employed for all hybridizations. This allowed the two F1 hybrid and two inbred strains to be compared after appropriate normalizations using four hybridizations per comparison rather than the six necessary if all pair-wise comparisons were performed. This strategy provided experimental efficiency, which was especially important as multiple hybridizations were performed to extract the best possible quality data. Unless otherwise noted, all results quoted represent comparison of the strains as cited, with the reference strain removed from the comparison.

Microarray Hybridization and Analysis

The DNA microarray used in all of the hybridization experiments contained 8,928 sequence verified EST clones spotted onto glass slides. The clones corresponded to both known and novel genes. Slides were printed and read at the core facilities of the departments of Molecular and Medical Pharmacology and Human Genetics at the University of California at Los Angeles. Hybridization was performed at 65°C in 3 \times SSC, 0.1% SDS, 0.1 \times Denhardt, 0.4 mg/ml polydA (Amersham), 0.2 mg/ml yeast tRNA, and 0.5 mg/ml CoT1 DNA (Gibco BRL) for 16–24 hr in humidified hybridization chambers (Corning Glass Works, Corning, NY). Slides were washed for 5 min at room temperature in 0.5 \times SSC, 0.01% SDS followed by another 5 min wash in 0.06 \times SSC. Slides were then spun briefly at 500 rpm to dry and scanned using the GenePix 4000A microarray scanner (Axon Instruments, Burlingame, CA). Images were analyzed using the GenePix Pro 3.0 software. For the F1 hybrid strain B6C3F1/J and the inbred strain 129X1/SvJ, three separate microarray hybridizations were performed for each strain using RNA isolated from independent groups of mice (Sandberg et al., 2000). For the F1 hybrid strain B6129PF1/J and the inbred strain LP/J two hybridizations were performed for each strain, again using independently isolated RNA samples.

The spot data exported from the GenePix Pro 3.0 software was processed using three types of normalization procedures. First, the data was subjected to preliminary trend removal to compensate for non-linearity due to absolute expression levels

(Yang et al., 2001). Second, spatial trends due to variations in the printing of the chips were removed by non-linear transformation of the data sets (A.O. unpublished). Third, differences in the labeling and chemical properties of the Cy3 and Cy5 dyes were compensated for by aligning the histograms of the dye signals, both within as well as between chips (Hegde et al., 2000). This non-linear procedure relied on the assumption that gene expression differences could be attributed to a few genes, and that the majority of genes do not display altered expression levels. Before further analysis, genes whose absolute signal value in the Cy5 or Cy3 channels was less than three times the background were excluded. After these procedures, a *t*-statistic was used to determine the genes that were most likely to be differentially expressed between the different strains of mice.

Real-Time QRT-PCR

Real-time quantitative reverse transcription polymerase chain reaction (QRT-PCR) was performed as described (Gibson et al., 1996; Heid et al., 1996; Freeman et al., 1999). For each strain, total RNA was extracted from the hippocampi of 8–10-week-old, 4–6 naive male mice using TRIzol reagent (Gibco BRL), following the manufacturer's instructions. RNA concentrations were determined using fluorometry. Reverse transcription used 50 ng total RNA primed with random hexamers. After reverse transcription, the amount of cDNA was quantitated using fluorometry, and 1 ng of cDNA employed in each real-time assay. The possibility of genomic contamination was excluded (Smith et al., 1997) by the use of primers that cross an intron of the housekeeping gene GdX. In addition, no reverse transcriptase controls were employed.

For three genes (AA166336, AA067119, AA170738), TaqMan reporter oligonucleotides were used, and for the other three genes (AA174533, AA389155, AA276325), the SYBR Green reporter system was employed, in each case following the manufacturer's instructions (PE Applied Biosystems, Foster City, CA). The RT-PCR was performed and analyzed using the ABI Prism 7700 Sequence Detection system (PE Applied Biosystems). For relative quantitation, a standard curve was constructed for each primer set, using total C57BL/6J mouse brain RNA. A glyceraldehyde phosphate dehydrogenase (GAPDH) control primer set was employed for normalization purposes. A balanced experimental design was employed, using six replicates for the TaqMan assays, and eight for the SYBR Green assays.

Singular Value Decomposition

The vector of 27 gene expression levels for each of the three B6C3F1/J and three 129X1/SvJ data sets were arranged as a matrix of dimension 27 by six. A set of six 27-element principal vectors was computed from this matrix using the singular value decomposition (SVD) option in Matlab (Mathworks, Natick, MA) and the two vectors corresponding to the second and third largest singular values were selected and plotted in Figure 5B. The 27-element gene expression vectors for the B6C3F1/J, 129X1/SvJ, B6129PF1/J and LP/J strains were projected onto each of these two principal vectors to produce two projection values per mouse strain that are plotted in Figure 5C. By projecting the data onto the two principal vectors we were able to demonstrate the separation of expression patterns for the

good-learning mouse strains from the poor-learning strains in the 2-D principal component subspace.

RESULTS

Morris Water Maze

Previous work has shown that F1 hybrid strains perform better in the Morris water maze test than inbred mouse strains (Owen et al., 1997). For this study we selected two F1 hybrid strains, B6129PF1/J and B6C3F1/J, and compared them with two inbred mouse strains, 129X1/SvJ and LP/J. To confirm the published results (Owen et al., 1997), the strains were tested on the hidden platform test of the Morris water maze task. In this test, mice must use spatial cues in the experimental room to learn the position of the hidden platform and escape from the water. Both of the F1 hybrid strains showed significant improvement over the nine training blocks whereas the two inbred mouse strains did not (Fig. 1A). Interestingly, even animals completely lacking the hippocampus typically show some improvement in escape latency on the hidden platform test by using non-hippocampal learning strategies (Moser et al., 1995). The more severe abnormalities in the inbred strains suggests that there may be additional extra-hippocampal deficits in these mice, although the inbred strains do not have any gross defects in vision, swimming or motivation, as judged by the visible platform test (below).

After completion of the hidden platform training, the mice were tested on the probe trial. This test determines whether the mice have acquired the spatial location of the hidden platform by allowing them to swim freely for 60 sec with the platform removed. Mice that have learned the position of the hidden platform should selectively cross the region where the platform had previously been located. The two F1 hybrid strains successfully learned the position of the hidden platform as shown by significantly higher crossings in the platform quadrant (Fig. 2A). In contrast, the two inbred mouse strains did not show a preference for the training quadrant. The preference score is calculated as the number of platform crossings in the trained quadrant minus the mean number of platform crossings in the other three quadrants (Owen et al., 1997). This score provides a single value for probe trial performance, so that direct comparisons between the different strains can be made. The two F1 hybrid strains had significantly higher preference scores than the two inbred strains (Fig. 2B), confirming that the hybrid strains are much better at learning the Morris water maze.

Because the two inbred mouse strains, 129X1/SvJ and LP/J, showed significant deficits on both the hidden platform and probe tests of the Morris water maze, naive mice from these strains were analyzed using the visible platform, to ensure that they had necessary vision, motor coordination and motivation. The F1 hybrids were also assessed using this test. Both the LP/J and 129X1/SvJ inbred strains showed effective acquisition of the visible platform task (Fig. 1B). 129X1/SvJ mice are albino, and this trait has been suggested to cause difficulties with the

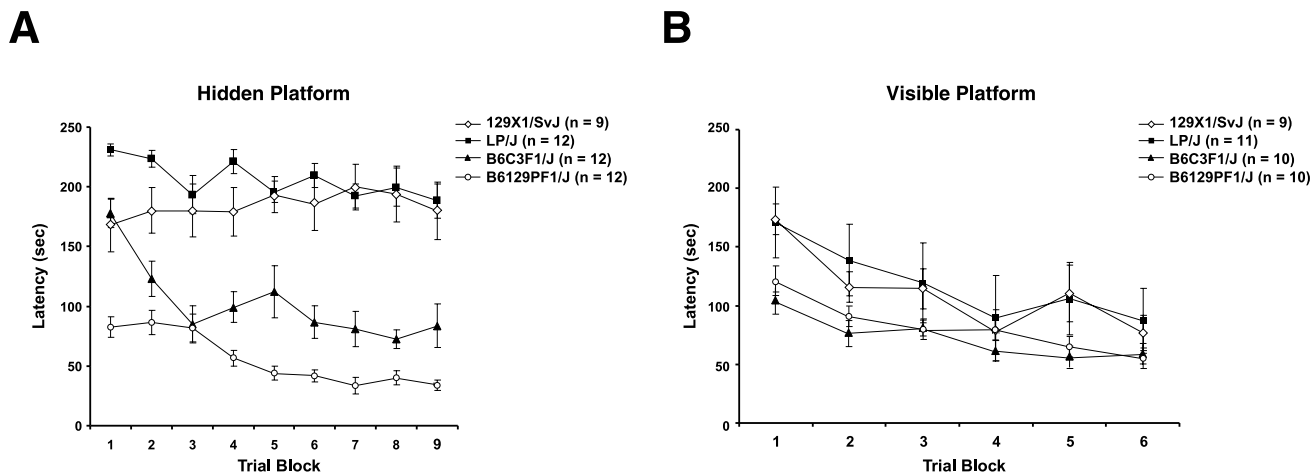


Fig. 1. Morris water maze training. **A:** Hidden platform test across the nine blocks of training for all four of the mouse strains. Both F1 hybrid strains displayed successful learning of the invisible platform test (B6129PF1/J: $F_{[8,99]} = 8.45$, $P < 0.0001$; B6C3F1/J: $F_{[8,99]} = 4.64$, $P < 0.0001$), whereas the inbred strains showed no learning (LP/J: $F_{[8,98]} = 1.84$, $P = 0.08$; 129X1/SvJ: $F_{[8,72]} = 0.22$, $P = 0.99$). In addition, the B6129PF1/J F1 hybrid was significantly faster in learning the task than the two inbred strains (B6129PF1/J vs. LP/J: $F_{[1,197]} = 1079.30$, $P < 0.0001$; B6129PF1/J vs. 129X1/SvJ: $F_{[1,171]} = 370.16$, $P < 0.0001$), as was the B6C3F1/J F1 hybrid (B6C3F1/J vs. LP/J: $F_{[1,197]} = 268.42$, $P < 0.0001$; B6C3F1/J vs. 129X1/SvJ: $F_{[1,171]} = 95.75$, $P < 0.0001$). **B:** Visible platform test. Both inbred strains

showed effective acquisition of the visible platform task (LP/J: $F_{[5,60]} = 2.84$, $P = 0.023$; 129X1/SvJ: $F_{[5,48]} = 4.24$, $P < 0.0029$), as did the F1 hybrids (B6129PF1/J: $F_{[5,54]} = 3.39$, $P < 0.0097$; B6C3F1/J: $F_{[8,99]} = 3.28$, $P < 0.0117$). Over all six training blocks, the B6129PF1/J F1 hybrid was significantly faster in acquiring the visible task than the two inbred strains (B6129PF1/J vs. LP/J: $F_{[1,114]} = 18.56$, $P < 0.0001$; B6129PF1/J vs. 129X1/SvJ: $F_{[1,102]} = 15.58$, $P < 0.0001$), as was the B6C3F1/J F1 hybrid (B6C3F1/J vs. LP/J: $F_{[1,114]} = 27.82$, $P < 0.0001$; B6C3F1/J vs. 129X1/SvJ: $F_{[1,102]} = 26.30$, $P < 0.0001$). The asymptotic performance of the four strains was equivalent, however, and by the last block there was no significant difference between them ($F_{[3,36]} = 1.62$, $P < 0.20$).

Morris water maze due to poor visual acuity (Montkowski et al., 1997). In our hands, however, both the LP/J and 129X1/SvJ inbred strains successfully acquired the visible platform task, showing that the poor performance of these strains on the hidden platform task represents specific deficits in learning and memory. The F1 hybrids also showed effective acquisition of the visible platform task, acquiring the task more quickly than the inbreds. The asymptotic performance of the two classes was equivalent, however, and by the last block there was no significant difference between them.

Microarray Hybridization and Analysis

To distinguish between variations in gene expression that are due to extraneous strain differences and those that may be causally associated with learning ability, we compared hippocampal gene expression from the two F1 hybrid strains, which perform well on the Morris water maze, with the two inbred strains that perform poorly. It was hypothesized that the F1 hybrid strains (B6129PF1/J and B6C3F1/J) may show common hippocampal gene expression differences compared to the inbred strains (129X1/SvJ and LP/J), hence restricting the number of learning and memory candidate genes. The rationale here was that it seemed likely that the basic molecular mechanisms of learning and memory would be shared between the four strains, although its efficiency might differ. The raw data from the slide hybridizations was subjected to

nonlinear transformations to normalize the data between slides and remove artifacts. Figure 3A shows a scatter plot of the average normalized log ratios of the B6C3F1/J strain compared to the 129X1/SvJ strain. The ratios are given in comparison to the reference strain, C57BL/6J which was employed in all of the slide hybridizations. Each spot on the scatter plot in Figure 3A is the result of averaging from three separate slide hybridizations. A t -statistic distribution was computed to determine which genes were differentially expressed (Fig. 3B). The distribution showed that 27 genes were differentially expressed with a cutoff t -value of >4 . Of these 27 genes, 20 of them were confirmed when compared to a hybridization where the Cy-dyes were reversed (Fig. 4A). It is important to make this comparison because it has been previously shown that dye specific labeling effects are a common problem in cDNA array hybridization (Schuchhardt et al., 2000). The probability of achieving this level of agreement (20 out of 27 genes) by chance is less than 0.01 based on binomial statistics (Fig. 4B), lending credibility to the analysis.

The 27 genes uncovered from the non-reversed experiments are listed in Table I, along with their accession numbers and proposed function or homology. Many of the genes identified are novel, but several genes were found that have known functions or have high homology to genes with known function. These include a gene involved in signal transduction (Rab2), a gene involved in intracellular scaffolding (Slc9a3r2/E3karp), a gene in-

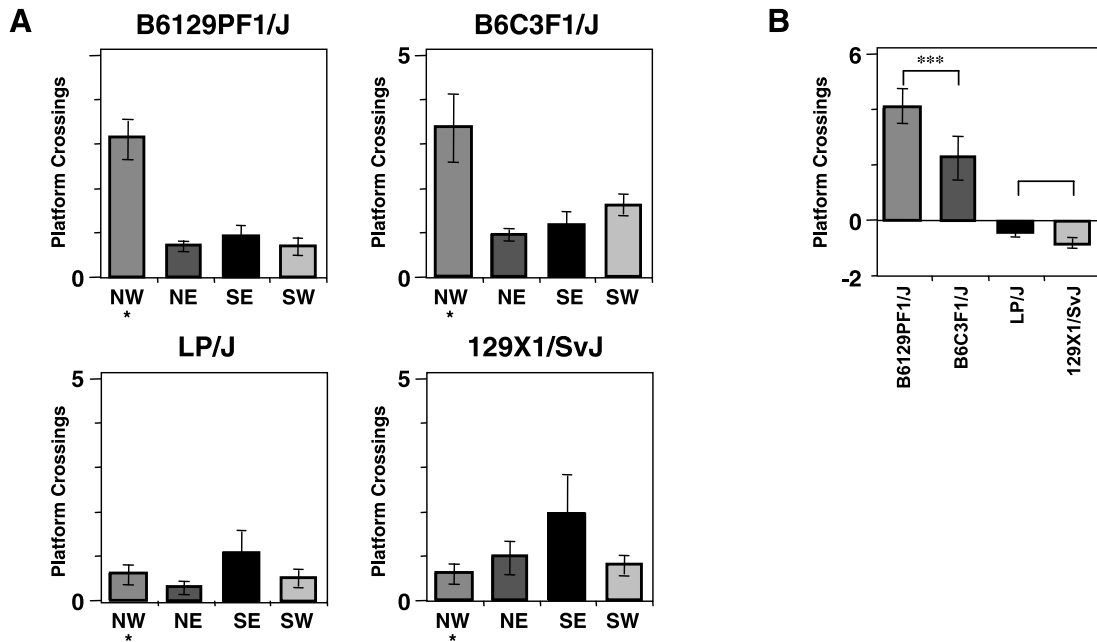


Fig. 2. Probe test. **A:** Platform crossings for the four mouse strains. F1 hybrid mice crossed the position where the hidden platform had been located more vigorously than equivalent locations in the three other quadrants (B6129PF1/J: $F_{[3,44]} = 20.03$, $P < 0.0001$; B6C3F1/J: $F_{[3,44]} = 4.19$, $P < 0.011$). In contrast, inbred strains did not display a directed search strategy (LP/J: $F_{[3,44]} = 1.91$, $P < 0.14$; 129X1/SvJ: $F_{[3,32]} = 2.43$, $P < 0.084$). Furthermore, the B6129PF1/J F1 hybrid strain showed significantly greater crossings over the correct platform location than the two inbred strains (B6129PF1/J vs. LP/J: $F_{[1,22]} = 60.70$, $P < 0.0001$; B6129PF1/J vs. 129X1/SvJ: $F_{[1,19]} = 40.82$, $P < 0.0001$), as did the B6C3F1/J F1 hybrid (B6C3F1/J vs. LP/J: $F_{[1,22]} = 8.75$, $P < 0.0073$; B6C3F1/J vs. 129X1/SvJ: $F_{[1,19]} = 5.70$, $P < 0.028$). Similar results were obtained from analysis of the quadrant dwell

times. The hidden platform was located in the NW quadrant, indicated by the asterisk. **B:** Preference scores for the strains. The B6129PF1/J F1 hybrid had a significantly greater preference score than the two inbred strains (B6129PF1/J F1 vs. LP/J: $F_{[1,22]} = 34.54$, $P < 0.0001$; B6129PF1/J F1 vs. 129X1/SvJ: $F_{[1,19]} = 35.43$, $P < 0.0001$), as did the B6C3F1/J F1 hybrid (B6C3F1/J F1 vs. LP/J: $F_{[1,22]} = 5.11$, $P < 0.034$; B6C3F1/J F1 vs. 129X1/SvJ: $F_{[1,19]} = 6.60$, $P < 0.019$). The two F1 hybrid strains were not significantly different from each other. The same was true for the two inbred strains. The mean preference scores (\pm SEM) were: B6129PF1/J F1: 4.39 ± 0.71 ; B6C3F1/J F1: 2.11 ± 0.93 ; LP/J: -0.083 ± 0.26 ; 129X1/SvJ: -0.74 ± 0.27 . *** $P < 0.0001$; comparing the grouped F1 hybrid strains to the grouped inbred strains.

involved in the ubiquitination pathway (BS4), a gene involved in purine/pyrimidine biosynthesis (PRPP synthetase), a gene involved in mitochondrial aerobic metabolism (succinate dehydrogenase), a gene for a potential seven pass transmembrane protein (AA174533), the Wolfram syndrome gene (Wfs1) and a putative transcription factor (mbFZB).

To explore the consistency of the hippocampal expression differences for the 27 genes in another F1 hybrid/inbred strain comparison, the B6129PF1/J and LP/J strains were investigated using microarrays. Figure 5A shows that there was significant correspondence of hippocampal expression differences between the F1 and inbred mouse strains. Real time QRT-PCR confirmed the differential expression between the F1 and inbred strains for five of six genes (Fig. 6). To allow simultaneous comparison of the gene expression relationships between the F1 and inbred strains, singular value decomposition (SVD) was employed. Figure 5B graphically displays the second and third principal vectors obtained from SVD of the expression levels for all 27 genes using each of the three B6C3F1/J

hybrid and three 129X1/SvJ inbred strain data sets. These vectors indicate potential patterns of co-regulation between the differentially expressed genes. They also represent a reduced subspace in which we can investigate how well the gene expression patterns cluster, not only for the two strains used to compute the SVD, but also for the other two strains, B6129PF1/J and LP/J. The projections of the data sets onto the 2-D subspace for the B6C3F1/J strain and the 129X1/SvJ strain are plotted in Figure 5C. It can be seen that the data representing the B6C3F1/J and 129X1/SvJ strains form separate clusters. We also computed the projection of the expression levels for the same set of 27 genes onto this subspace for the data from the other two strains, B6129PF1/J and LP/J. This procedure revealed that the B6129PF1/J and B6C3F1/J strains are grouped together and can be separated from the LP/J and 129X1/SvJ strains, which form their own independent cluster. Thus, the differential expression of these 27 genes effectively separates the good-learning mouse strains from the poor-learning mouse strains, agreeing with the data from the behavior experiments.

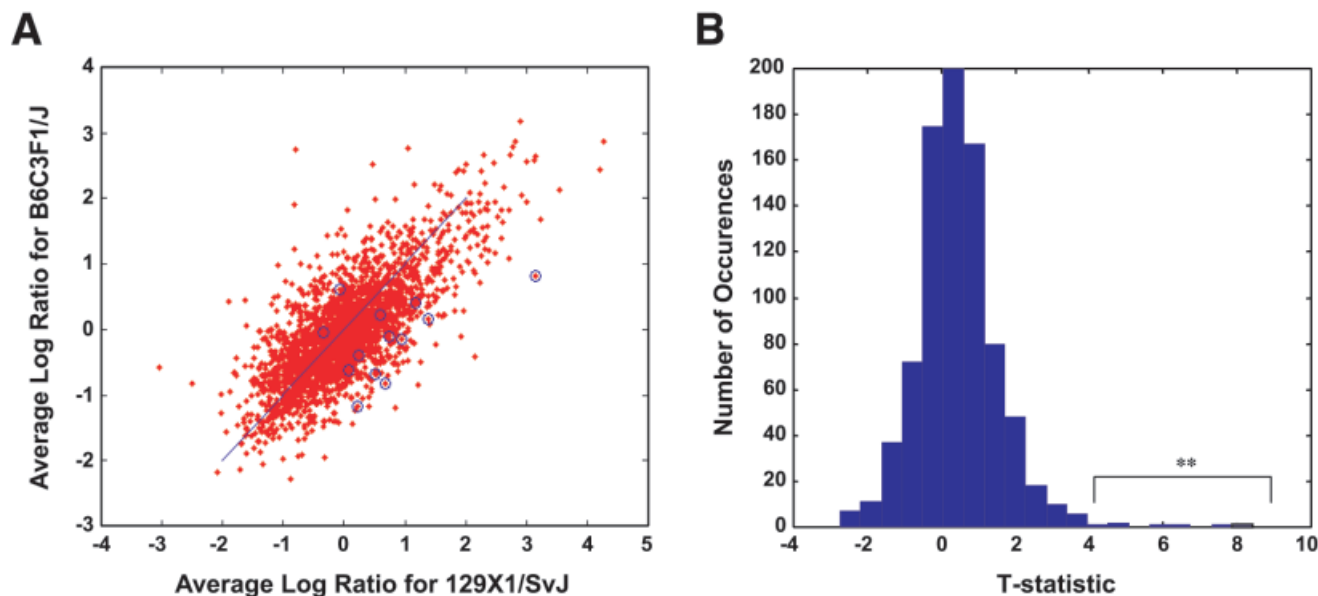


Fig. 3. Expression data. **A:** Scatterplot of selected genes. Genes that passed initial quality control were plotted against each other using the average log ratio from several slide hybridizations. The y-axis shows the average log ratio of genes from three slide hybridizations for the B6C3F1/J hybrid strain and the x-axis shows the average log ratio of three slide hybridizations from the 129X1/SvJ inbred strain. Thirteen

spots are circled, and represent typical genes found to be consistently and significantly differentially expressed between the F1 and inbred strains. **B:** *t*-Statistic distribution for the genes from the scatterplot in (A). Twenty-seven genes were selected from this distribution as being significantly outside the normal distribution (**nominal $P < 0.01$, indicated by bar).

DISCUSSION

DNA microarray technology permits rapid identification of candidate genes for complex mammalian phenotypes without many of the drawbacks associated with more laborious techniques, such as genetic mapping. Microarray technology has so far largely been used to study unicellular systems (Debouck and Goodfellow, 1999; Khan et al., 1999; Manger and Relman, 2000), and few investigations have attempted to associate behavioral differences with gene expression changes. In this study, a 9,000 gene microarray was used to look at differences in gene expression in the hippocampus of mouse strains that perform well in the Morris water maze compared to mouse strains that perform poorly.

Two F1 hybrid and two inbred strains were employed in the analysis, largely because they had previously been shown to display dramatic differences in learning and memory (Owen et al., 1997). Inbred strains alone were not used, as the behavioral differences between them are not as marked as those between F1 hybrids and inbreds. In addition, the four strains were chosen, insofar as possible, to minimize their relatedness given the constraints of divergent learning and memory. The idea was that molecular pathways for learning and memory were likely to be shared between strains regardless of common ancestry, but that the efficiency of the pathways must differ. Using divergent strains should increase the opportunities to find differences in gene expression profiles regardless of common genealogy. Even though the 129X1/SvJ mice are albino, all of the strains utilized showed effective acquisi-

tion of the visible platform task, suggesting that any deficits uncovered in the hidden platform task were not due to non-specific abnormalities, such as in vision or motor abilities. There is no evidence that the strains investigated have abnormalities in other sensory tests, including hearing and touch (Crawley et al., 1997).

The use of proper quality control measures when performing microarray hybridizations is very important. Although we have performed three slide hybridizations for each of the B6C3F1/J and 129X1/SvJ strains, and two hybridizations for each of the B6129PF1/J and LP/J strains, variations can often skew the data in one direction or another. Artifacts can be due to factors such as unevenness in slide surface properties, incomplete washing, dust contamination and variability in spot sizes (Sherlock, 2000). To minimize the influence of these factors on the expression analysis, we subjected the data to signal processing, in the form of non-linear normalization between slides. Many microarray expression studies determine differential expression by using an arbitrary two-fold cutoff for expression difference, but this does not take into account inherent differences in the quality of expression data between slide hybridizations. Therefore, we have subjected our normalized expression data to a *t*-statistic analysis, to determine which genes are truly outliers. This allowed us to select 27 genes that were differentially expressed between the B6C3F1/J hybrid strain and the 129S/vJ inbred strain. When these genes were looked at in the dye-flipped experiment, we were able to confirm 20 out of 27 of these genes, a significant result ($P = 0.01$,

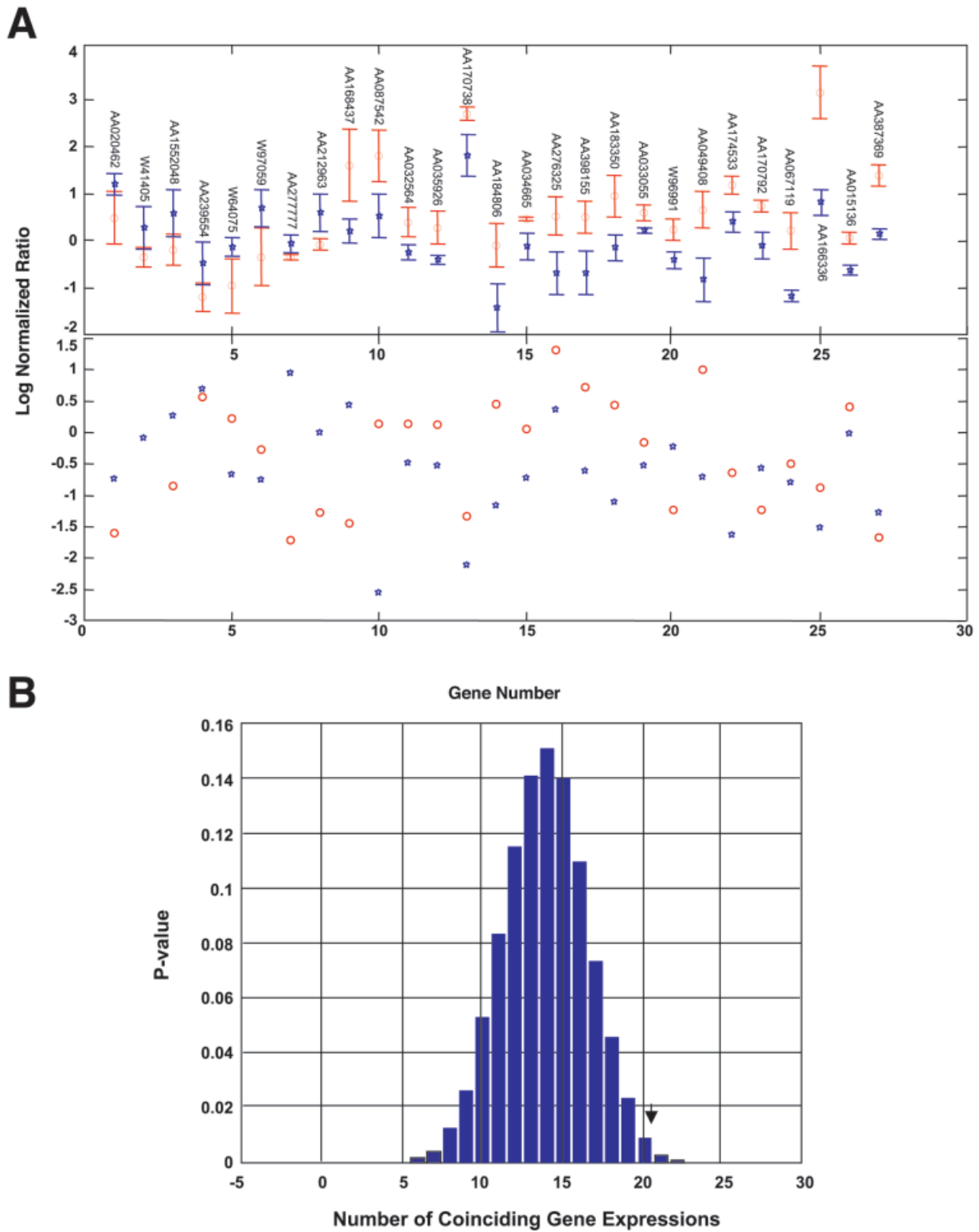


Fig. 4. Differentially expressed genes. **A:** The 27 genes found to be differentially expressed between the B6C3F1/J hybrid strain and the 129X1/SvJ inbred strain. The y-axis corresponds to normalized log ratio and the x-axis is the gene number. The expression levels from the three slide averaged experiments are shown in the top chart and the lower chart shows the expression levels detected in the dye-flipped experiment, which

confirmed 20 out of 27 genes. The blue stars indicate the expression level in the B6C3F1/J strain, whereas the red circles indicate the level of expression in the 129X1/SvJ strain. Genes are indicated by GenBank accession number. **B:** The binomial histogram shows that the probability of obtaining a confirmation of 20 out of 27 genes when the dyes are flipped is significant ($P < 0.01$, one-tailed).

TABLE I. Genes Differentially Expressed in the Hippocampus of the B6C3F1/J and 129×1/SvJ Strains

GenBank accession number	Gene
Expression greater in B6C3F1/J hybrid mice	
AA020462	GTP Binding protein and RAS oncogene Rab2
W41405	Faim (Fas apoptotic inhibitory molecule)
AI552048	Anonymous EST
AA239554	Anonymous EST
W64075	Prtb (Proline rich protein expressed in brain)
W97059	Anonymous EST
AA277777	Wolfram syndrome 1 homolog (Wfs1)
Expression greater in 129×1/SvJ inbred mice	
AA212963	Highly similar to AF214737_1 C9orf10a [H. sapiens]
AA168437	Similar to microtubule-associated protein homolog [D. melanogaster]
AA087542	Moderately similar to Human hypothetical protein KIAA0281 [H. sapiens]
AA032564	Anonymous EST
AA035926	Mouse BS4 protein/NY-REN-18 antigen
AA170738	Highly similar to rat phosphoribosylpyrophosphate (PRPP) synthetase-associated protein
AA184806	Anonymous EST
AA034665	Anonymous EST
AA276325	Highly similar to succinate dehydrogenase [H. sapiens]
AA389155	Anonymous EST
AA183350	Weakly similar to Rat GNT5 (alpha-1,3(6)-mannosylglycoprotein beta-1,6-N-acetyl-glucosaminyltransferase V)
AA033055	Slc9a3r2 (solute carrier family 9, isoform 3 regulator 2)
W96991	rjs gene [M. musculus]
AA049408	Anonymous EST
AA174533	Weakly similar to AF154337 1 putative seven pass transmembrane protein [M. musculus]
AA170792	DNA topoisomerase I
AA067119	Anonymous EST
AA166336	Moderately similar to DRIM protein [H. sapiens]
AA015136	mbFZb (basic FGF repressed, Zic binding protein)
AA387369	PMM 1 (Phosphomannomutase I)

binomial statistics, Fig. 4B). Analysis of the 27 genes in a further study of the B6129PF1/J and LP/J strains revealed significant correspondence of F1/inbred differential expression for these genes.

Employing SVD analysis of the expression data allowed us to simultaneously compare the data from the first strain comparison (B6C3F1/J vs. 129X1/SvJ) with the data from the second comparison (B6129PF1/J vs. LP/J). We wanted to test whether these four strains could be separated based on the expression of the 27 genes that came out of the initial analysis. This would suggest potential patterns of co-regulation for the genes between the strains, providing an indication of strain similarity or difference. The fact that the two good learning and memory strains separated clearly from the two poor learning and memory strains in the principal components analysis suggests that these genes are in fact valid candidates genes for learning and memory in mice. It also supports the notion that the different strains of mice share common molecular pathways for learning and memory, but that the effectiveness of these mechanisms varies between strains. Despite these findings, there may also be other important behavioral differences between the four strains, in addition to those in learning and memory (Crawley et al., 1997). Nevertheless, our results suggest that surveying increased numbers of strains with common differences in one behavior will help identify genes responsible for that behav-

ior, independent of variations in other behaviors. Clearly, such an effort will be aided by as complete and thorough a behavioral phenotyping of the analyzed strains as possible. In this context, the mouse is a particularly favorable system for such studies, as the high degree of inter-individual genetic and environmental variability in other organisms, such as humans, renders analysis much more difficult.

Seeking stable differences in adult gene expression levels will not identify genes responsible for learning and memory variations if these variations are due to transient spatial or temporal gene expression changes during development. In addition, coding or non-coding gene polymorphisms may cause learning and memory differences independent of changes in expression levels. Microarray studies using unicellular systems have shown, however, that even single genetic perturbations result in a spreading cascade of gene expression changes that extend far beyond the initial disturbance (Brown and Botstein, 1999). Thus, it is likely that the microarray studies described here may provide entry points into novel pathways for learning and memory. This contention is supported by studies where genes involved in complex traits have been identified on the basis of expression differences (Pugliese et al., 1997; Vafiadis et al., 1997; Malhotra et al., 1998; Aitman et al., 1999; O'Connor et al., 1999; Timtchenko et al., 1999; Smithies et al., 2000).

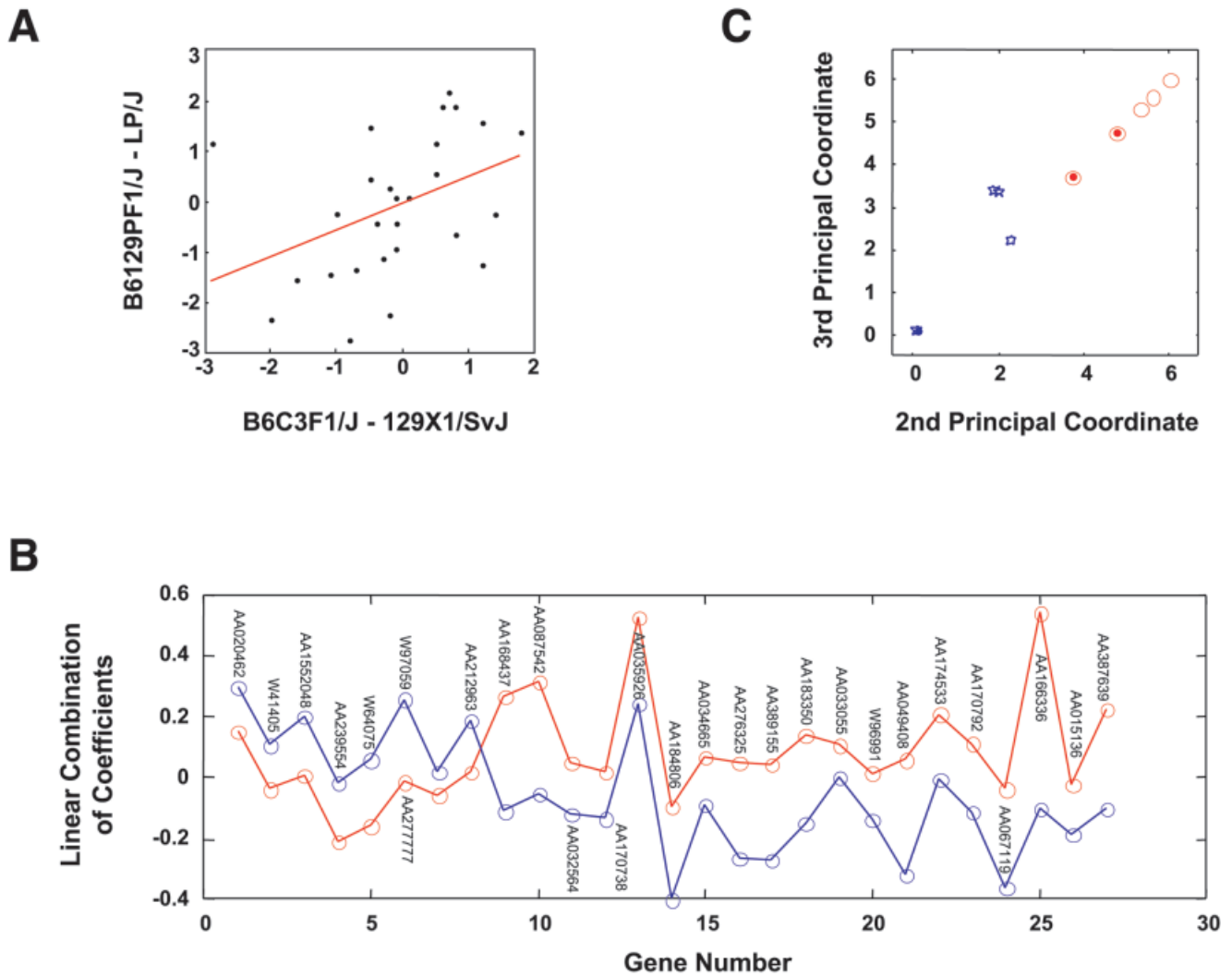


Fig. 5. Comparison of the two data sets B6C3F1/J vs. 129SvJ and B6129PF1/J vs. LP/J. **A:** Scatterplot comparing the mean expression differences between the first F1/inbred comparison (B6C3F1/J vs. 129X1/SvJ) and the second (B6129PF1/J vs. LP/J) for the 27 genes. Expression differences are shown using the logarithm (\log_2) of the gene expression ratios. There was a significant correlation between the data from the two comparisons ($r = 0.43$, $F_{[1,25]} = 5.4$, $P < 0.029$). The best fit using least squares linear regression is shown. **B:** Graphical representation of the second (red) and third (blue) principal vectors computed by singular value

decomposition of the 27 gene data from the comparison between B6C3F1/J and the 129X1/SvJ strains. These patterns show potential patterns of co-regulated variations in expression levels between the two strains. **C:** Result of projecting the 27 gene expression levels for each of the four strains (B6C3F1/J, 129X1/SvJ, B6129PF1/J, and LP/J) onto each of the two principal vectors in (B). The graph shows that the B6C3F1/J (red circles) and B6129PF1/J (filled circles) strains cluster together, as do the 129X1/SvJ (blue stars) and LP/J (filled stars) strains, and that these two clusters are mutually exclusive.

Many of the genes that are differentially expressed between the B6C3F1/J and the 129X1/SvJ strain have unknown function. Of the genes with known functions, however, some have intriguing prospects for a role in learning and memory. For instance the mouse homolog of the Wolfram gene, the gene involved in Wolfram syndrome, was detected as being more highly expressed in the B6C3F1/J strain. Wolfram syndrome, is an autosomal recessive disorder in which homozygotes have severe behavioral defects that often include mental retardation (Swift et al., 1998). Heterozygous carriers of the mutation are also at high risk for psychiatric disorders (Swift et al.,

1998). The evidence from the present study suggests that Wolfram may play a role in normal learning and memory. A recently discovered gene, the transcription factor mbFZB, was also identified in this study. mbFZB has high homology to the *Xenopus laevis* putative Zic3 binding protein, and the mouse gene has been shown to be strongly expressed in the developing brain as well as in the adult mouse hippocampus and piriform cortex (Vetter and Wurst, 2001).

The mouse homolog of rat phosphoribosylpyrophosphate (PRPP) synthetase associated protein was found to be more highly expressed in the 129X1/SvJ inbred

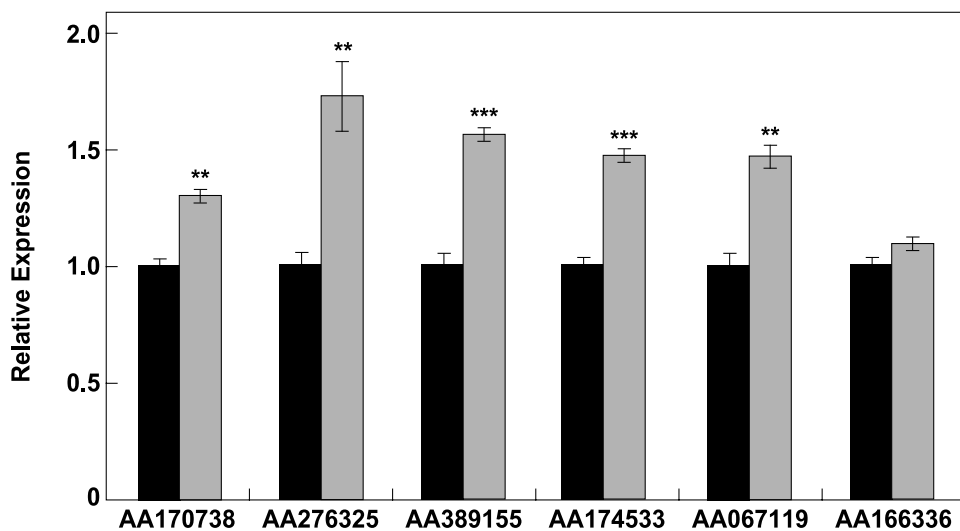


Fig. 6. Planned comparison of hippocampal gene expression differences between F1 and inbred strains using real-time QRT-PCR. Results are shown using expression relative to the F1 strains for each gene after normalization to GAPDH. Dark shading, F1 strains; light shading, inbred strains. ** $P < 0.001$; *** $P < 0.0001$, two-tailed t -test.

strain. The PRPP synthetase associated protein gene is believed to have a negative regulatory role on PRPP synthetase (Tatibana et al., 1995). Because PRPP synthetase is involved in purine metabolism, an inhibitor of this protein may affect ATP levels and thus indirectly affect learning and memory in the hippocampus. Another gene that was found to be more highly expressed in the 129X1/SvJ inbred strain is *Slc9a3r2/E3Karp*. The protein product of this gene is known to interact with the Na^+/H^+ exchanger 3 isoform (NHE3) and is required for the protein kinase A mediated inhibition of NHE3 (Yun et al., 1998). Addition of inhibitors of the Na^+/H^+ exchanger in hippocampal neurons can inhibit neuronal activity (Bonnet et al., 2000), and increased expression of an inhibitor of the Na^+/H^+ exchanger in the hippocampus might reduce learning ability.

Complex trait mapping using context and cue dependent fear conditioning in mice has found a number of loci responsible for learning and memory (Caldarone et al., 1997; Wehner et al., 1997). None of the genes uncovered in the study described here, however, appear to localize to these mapped regions. This discrepancy may be due a number of factors, including the use of different mouse strains and behavioral assays.

A recent study described the use of Affymetrix gene chips to identify differences in gene expression between two inbred mouse strains, C57BL/6J and 129SvEv (Sandberg et al., 2000), in various regions of the brain. Even though this investigation did examine gene expression in the hippocampus, the candidate genes discovered in our study did not overlap, and this may be due to differences in the genes present on the arrays, as well as the different strains examined. The present investigation looked at two F1 hybrid strains, both of which learn better than any inbred strains, and compared them with two of the poorest learning inbred strains. Using multiple strains in this way, we were able to maximize the potential for discovery of differentially expressed genes in the hippocampus that are good candidate genes for learning and memory.

An important part of our approach to identifying candidate genes for learning and memory was the use of multiple strains combined with the application of principal components analysis to the microarray expression data. This novel approach showed that the two F1 hybrid mouse strains and the two inbred strains formed two mutually exclusive clusters, which correlated well with the data from the behavior experiments. These observations suggest this approach will form a powerful strategy to leverage the maximum possible information from multiple mouse strains in the identification of candidate genes for complex behavioral and other traits. A few of the genes discovered in our analysis have known functions, such as *Wolframin*, *BS4*, *Rab2*, *PRPP synthase* and *Slc9a3r2/Karp*, and suggests that these genes may potentially play a role in hippocampal learning and memory. It will also be interesting to study the genes that have unknown function, or show only weak homology to known genes, and understand their possible role in the learning process. Although the analysis employed strains with dramatic differences in learning and memory and assayed a brain region intimately involved in this process, the uncovered genes may be involved in other behavioral processes. Knockout and transgenic strategies can be employed in the future to elucidate the function of both known and unknown candidate genes.

REFERENCES

- Aitman TJ, Glazier AM, Wallace CA, Cooper LD, Norsworthy PJ, Wahid FN, Al-Majali KM, Trembling PM, Mann CJ, Shoulders CC. 1999. Identification of *Cd36* (*Fat*) as an insulin-resistance gene causing defective fatty acid and glucose metabolism in hypertensive rats. *Nat Genet* 21:76–83.
- Belvin MP, Yin JC. 1997. *Drosophila* learning and memory: recent progress and new approaches. *Bio Essays* 19:1083–1089.
- Bonnet U, Leniger T, Wiemann M. 2000. Alteration of intracellular pH and activity of CA3-pyramidal cells in guinea pig hippocampal slices by inhibition of transmembrane acid extrusion. *Brain Res* 872:116–124.

- Bowers BJ, Christensen SC, Pauley JR, Paylor R, Yuva L, Dunbar SE, Wehner JM. 1995. Protein and molecular characterization of hippocampal protein kinase C in C57BL/6 and DBA/2 mice. *J Neurochem* 64:2737–2746.
- Brown PO, Botstein D. 1999. Exploring the new world of the genome with DNA microarrays. *Nat Genet* 21:33–37.
- Caldarone B, Saavedra C, Tartaglia K, Wehner JM, Dudek BC, Flaherty L. 1997. Quantitative trait loci analysis affecting contextual conditioning in mice. *Nat Genet* 17:335–337.
- Chen C, Tonegawa S. 1997. Molecular genetic analysis of synaptic plasticity, activity-dependent neural development, learning, and memory in the mammalian brain. *Ann Rev Neurosci* 20:157–184.
- Cho YH, Friedman E, Silva AJ. 1999. Ibotenate lesions of the hippocampus impair spatial learning but not contextual fear conditioning in mice. *Behav Brain Res* 98:77–87.
- Crawley JN, Belknap JK, Collins A, Crabbe JC, Frankel W, Henderson N, Hitzemann RJ, Maxson SC, Miner LL, Silva AJ. 1997. Behavioral phenotypes of inbred mouse strains: implications and recommendations for molecular studies. *Psychopharmacology* 132:107–124.
- Davis RL. 1996. Physiology and biochemistry of *Drosophila* learning mutants. *Physiol Rev* 76:299–317.
- Debouck C, Goodfellow PN. 1999. DNA microarrays in drug discovery and development. *Nature Genet* 21:48–50.
- Dubnau J, Tully T. 1998. Gene discovery in *Drosophila*: new insights for learning and memory. *Annu Rev Neurosci* 21:407–444.
- Freeman WM, Walker SJ, Vrana KE. 1999. Quantitative RT-PCR: pitfalls and potential. *Bio Tech* 26:112–125.
- Gibson UE, Heid CA, Williams PM. 1996. A novel method for real time quantitative RT-PCR. *Genome Res* 6:995–1001.
- Grotewiel MS, Beck CD, Wu KH, Zhu XR, Davis RL. 1998. Integrin-mediated short-term memory in *Drosophila*. *Nature* 391:455–460.
- Hegde P, Qi R, Abernathy K, Gay C, Dharap S, Gaspard R, Hughes JE, Snesrud E, Lee N, Quackenbush J. 2000. A concise guide to cDNA microarray analysis. *Bio Tech* 29:548–556.
- Heid CA, Stevens J, Livak KJ, Williams PM. 1996. Real time quantitative PCR. *Genome Res* 6:986–994.
- Khan J, Saal LH, Bittner ML, Chen Y, Trent JM, Meltzer PS. 1999. Expression profiling in cancer using cDNA microarrays. *Electrophoresis* 20:223–229.
- Malhotra AK, Goldman D, Mazzanti C, Clifton A, Breier A, Pickar D. 1998. A functional serotonin transporter (5-HTT) polymorphism is associated with psychosis in neuroleptic-free schizophrenics. *Mol Psychiatry* 3:328–332.
- Manger ID, Relman DA. 2000. How the host 'sees' pathogens: global gene expression responses to infection. *Curr Opin Immunol* 12:215–218.
- Matsuyama S, Namgung U, Routtenberg A. 1997. Long-term potentiation persistence greater in C57BL/6 than DBA/2 mice: predicted on basis of protein kinase C levels and learning performance. *Brain Res* 763:127–130.
- Mayford M, Kandel ER. 1999. Genetic approaches to memory storage. *Trends Genet* 15:463–470.
- Mayford M, Mansuy IM, Muller RU, Kandel ER. 1997. Memory and behavior: a second generation of genetically modified mice. *Curr Biol* 7:R580–R589.
- Montkowski A, Poettig M, Mederer A, Holsboer F. 1997. Behavioral performance in three substrains of mouse strain 129. *Brain Res* 762:12–18.
- Moser E, Moser MB, Andersen P. 1993. Spatial learning impairment parallels the magnitude of dorsal hippocampal lesions, but is hardly present following ventral lesions. *J Neurosci* 13:3916–3925.
- Moser MB, Moser EI, Forrest E, Andersen P, Morris RG. 1995. Spatial learning with a mini-slab in the dorsal hippocampus. *Proc Natl Acad Sci USA* 92:9697–9701.
- O'Connor DT, Takiyuddin MA, Printz MP, Dinh TQ, Barbosa JA, Rozansky DJ, Mahata SK, Wu H, Kennedy BP, Ziegler MG. 1999. Catecholamine storage vesicle protein expression in genetic hypertension. *Blood Press* 8:285–295.
- Owen EH, Logue SF, Rasmussen DL, Wehner JM. 1997. Assessment of learning by the Morris water task and fear conditioning in inbred mouse strains and F1 hybrids: implications of genetic background for single gene mutations and quantitative trait loci analyses. *Neuroscience* 80:1087–1099.
- Plomin R. 2001. The genetics of g in human and mouse. *Nat Rev Neurosci* 2:136–141.
- Pugliese A, Zeller M, Fernandez A Jr, Zalberg LJ, Bartlett RJ, Ricordi C, Pietropaolo M, Eisenbarth GS, Bennett ST, Patel DD. 1997. The insulin gene is transcribed in the human thymus and transcription levels correlated with allelic variation at the INS VNTR-IDD2 susceptibility locus for type 1 diabetes. *Nat Genet* 15:293–297.
- Sandberg R, Yasuda R, Pankratz DG, Carter TA, Del Rio JA, Wodicka L, Mayford M, Lockhart DJ, Barlow C. 2000. Regional and strain-specific gene expression mapping in the adult mouse brain. *Proc Natl Acad Sci USA* 97:11038–11043.
- Schena M, Shalon D, Davis RW, Brown PO. 1995. Quantitative monitoring of gene expression patterns with a complementary DNA microarray. *Science* 270:467–470.
- Schuchhardt J, Beule D, Malik A, Wolski E, Eickhoff H, Lehrach H, Herzog H. 2000. Normalization strategies for cDNA microarrays. *Nucleic Acids Res* 28:E47.
- Scoville WB, Milner B. 2000. Loss of recent memory after bilateral hippocampal lesions. *J Neuropsychiatry Clin Neurosci* 12:103–113.
- Sherlock G. 2000. Analysis of large-scale gene expression data. *Curr Opin Immunol* 12:201–205.
- Smith DJ, Stevens ME, Sudanagunta SP, Bronson RT, Makhinson M, Watabe AM, O'Dell TJ, Fung J, Weier HU, Cheng JF. 1997. Functional screening of 2 Mb of human chromosome 21q22.2 in transgenic mice implicates minibrain in learning defects associated with Down syndrome. *Nat Genet* 16:28–36.
- Smithies O, Kim HS, Takahashi N, Edgell MH. 2000. Importance of quantitative genetic variations in the etiology of hypertension. *Kidney Int* 58:2265–2280.
- Swift RG, Polymeropoulos MH, Torres R, Swift M. 1998. Predisposition of Wolfram syndrome heterozygotes to psychiatric illness. *Mol Psychiatry* 3:86–91.
- Tatibana M, Kita K, Taira M, Ishijima S, Sonoda T, Ishizuka T, Iizasa T, Ahmad I. 1995. Mammalian phosphoribosyl-pyrophosphate synthetase. *Adv Enzyme Regul* 35:229–249.
- Timchenko D, Kratzsch J, Sauerwein H, Wegner J, Souffrant WB, Schwerin M, Brockmann GA. 1999. Fat storage capacity in growth-selected and control mouse lines is associated with line-specific gene expression and plasma hormone levels. *Int J Obes Relat Metab Disord* 23:586–594.
- Upchurch M, Wehner JM. 1988. Differences between inbred strains of mice in Morris water maze performance. *Behav Genet* 18:55–68.
- Vafiadis P, Bennett ST, Todd JA, Nadeau J, Grabs R, Goodyer CG, Wickramasinghe S, Colle E, Polychronakos C. 1997. Insulin expression in human thymus is modulated by INS VNTR alleles at the IDDM2 locus. *Nat Genet* 15:289–292.
- Vetter K, Wurst W. 2001. Expression of a novel mouse gene 'mbFzB' in distinct regions of the developing nervous system and the adult brain. *Mech Dev* 100:123–125.
- Wehner JM, Radcliffe RA, Rosmann ST, Christensen SC, Rasmussen DL, Fulker DW, Wiles M. 1997. Quantitative trait locus analysis of contextual fear conditioning in mice. *Nat Genet* 17:331–334.
- Yang YH, Dudoit S, Luu P, Speed TP. 2001. Normalization for cDNA microarray data. *Proc SPIE* 4266:141–152.
- Yun CH, Lamprecht G, Forster DV, Sidor A. 1998. NHE3 kinase A regulatory protein E3KARP binds the epithelial brush border Na⁺/H⁺ exchanger NHE3 and the cytoskeletal protein ezrin. *J Biol Chem* 273:25856–25863.